STRUCTURAL AND FUNCTIONAL INVESTIGATION OF PROMOTER DISTORTION AND OPENING IN THE RNA POLYMERASE II CLEFT

DISSERTATION

For the award of the degree

"Doctor rerum naturalium"

By the Georg-August-Universität Göttingen

Within the graduate program

"Biomolecules – Structure, Function and Dynamics"

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

Submitted by

Christian Dienemann

From Jena, Germany

Göttingen 2018

MEMBERS OF THE THESIS COMMITTEE

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

Prof. Dr. Claudia Höbartner
Institute for organic Chemistry
University of Würzburg

Prof. Dr. Ralf Ficner
Institute for Microbiology and Genetics
University of Göttingen

Members of the examination board

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

Prof. Dr. Ralf Ficner (2nd referee)
Institute for Microbiology and Genetics
University of Göttingen

Other members of the examination board

Prof. Dr. Claudia Höbartner
Institute for organic Chemistry
University of Würzburg

Dr. Alex Faesen
Max Planck Institute for Biophysical Chemistry, Göttingen

Prof. Dr. Henning Urlaub
Max Planck Institute for Biophysical Chemistry, Göttingen

AFFIDAVIT

I, Christian Dienemann, hereby declare that my dissertation “Structural and functional investigation of promoter distortion and melting in the RNA polymerase II cleft” has been written independently and with no other sources and aids than quoted. This dissertation or parts of it have not been submitted elsewhere for any academic award or qualification. The electronic version of this dissertation is congruent to the printed versions both in content and in format.

Göttingen, 28th of February 2018

……

Christian Dienemann

ACKNOWLEDGEMENTS

First, I want to thank Patrick Cramer for giving me the opportunity to work with and for him on this project. The atmosphere in his lab, his interest and enthusiasm for mechanistic questions and the freedom granted by him provided the stable and fertile ground for my PhD work. From him I learned how to ask the right questions, to simplify a problem and to extract the essence of things.

Second I want to thank all the members (former and current) of the Cramer lab for the great atmosphere, the supportive spirit and the helpful advice for all kinds of things. Special thanks go to Sarah Sainsbury who introduced me to transcription initiation and the associated experimental protocols. Björn Schwalb contributed significantly to my work and I therefore thank him for all the questions he answered and all the plots he went through with me. I want to thank former and current members of the yeast Pol II lab in room 208 for lots of support and great fun during my entire PhD. You are the best! I also want to thank Simon Neyer and Hauke Hillen for the cozy beer meetings and the scientific and non-scientific discussions we had. I also thank Simon Neyer, Carrie Bernecky, Youwei Xu and Dimitry Tegunov for the EM discussions and coffee breaks. My work would have been impossible without Carina Burzinski who was helping with protein purifications. I also thank Kerstin Maier, Petra Rus and Kirsten Backs for the great lab organization.

I also want to acknowledge the Biomolecules, who are the best graduate program in the world! Thanks for all the fun. I further thank the GGNB, especially the office for their organization and help during my PhD.

I am thankful to my thesis advisory committee consisting of Prof. Ralf Ficner and Prof. Claudia Höbartner, who were very helpful and supportive throughout my PhD.

I also want to thank my family for accepting the sparseness of my visits and for being a strong support.

Lastly, I thank Theresa for her love, patience, support and always standing by my side.

PUBLICATIONS

Part of this work was published or is part of a submitted manuscript:

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

Author contributions: C.P. designed and carried out high-resolution cryo-EM structure determinations of OC1–OC4. M.H. designed and carried out Pol II-TFIIF crystallographic analysis, and cryo-EM structure determinations of OC5 and CC. C.P. and M.H. designed and carried out functional assays. C.D. cloned and purified full-length TBP and TFIIA. C.D. and C.B. assisted with protein purification. J.P. supervised electron microscopy data collection. P.C. designed and supervised research. C.P., M.H., and P.C. prepared the manuscript.

Dienemann C, Schwalb B, Schilbach S, Cramer P. Promoter distortion and opening in the RNA polymerase II cleft. *Molecular Cell,* Online 21 November 2018

Author contributions: C.D. carried out all experiments and data analysis except for the following. B.S. carried out bioinformatics analysis of Ssl2 anchor-away 4sU-sequencing data. B.S. and C.D. interpreted Ssl2 anchor-away data. S.S. carried out initial processing of the CC^{dist} data set. P.C. designed and supervised research. C.D. and P.C. wrote the manuscript.

Parts of Plaschka *et al.* are described in more detail in the following sections:

3 Methods

3.9 Expression and purification of TBP*

3.10 Expression and purification of TFIIAΔ113*

The following sections were taken from Dienemann *et al.*:

3 Methods

3.12 Ssl2 nuclear depletion and 4sU-sequencing†

3.13 Bioinformatics data analysis†

3.14 DNA opening assay†

3.15 Preparation of cPICs†

3.16 Cryo-EM and image processing of cCC1 and cCC^{dist}†

3.17 Data processing and CC^{dist} reconstruction†

3.18 Model building and refinement †

4 Results

4.2 Structural and functional investigation of promoter opening

4.2.1 TFIIH translocase is not generally required for transcription *in vivo*

4.2.2 Translocase requirement correlates with promoter stability

4.2.3 Promoter stability defines DNA meltability

4.2.4 Cryo-EM reveals distinct closed complex intermediates

4.2.5 CC^{dist} contains distorted DNA in a closed cleft

4.2.6 Complete PIC structure with distorted DNA

4.2.7 Promoters of Pol I and Pol III contain unstable IMRs

5 Discussion

5.1 A new transcription initiation intermediate

5.2 DNA distortion and initial opening

5.3 General features of transcription initiation

The following figures were taken from Dienemann *et al.* (manuscript figure number in parenthesis):

Figure 7: Ssl2-independent transcription *in vivo* (Figure 1)

Figure 8: Promoter opening depends on the initially melted region (Figure 2)

Figure 9: Cryo-EM structures of core closed complexes (cCC) (Figure 3)

Figure 10: Distortion of promoter DNA (Figure 4)
Figure 11: PIC structure with closed, distorted DNA (Figure 5)
Figure 12: IMRs of Pol I and Pol III promoters are unstable (Figure 6)
Figure 13: General model of transcription initiation (Figure 7)

Supplemental Figure 1: Validations and controls for Ssl2 anchor away (Figure S1)
Supplemental Figure 2: TFIIH independent opening of GAT1 promoter DNA (Figure S2)
Supplemental Figure 3: Data processing of cCC1 and cCCdist (Figure S3)
Supplemental Figure 4: Clamp and B-linker of cCC1 and cCCdist (Figure S4)
Supplemental Figure 5: Data processing of CCdist (Figure S5)

The following tables were taken from Dienemann *et al.* (manuscript table number in parenthesis):

Supplemental Table 1: Data collection and processing statistics (Table S1)
Supplemental Table 2: Eukaryotic closed complexes (Table S2)

The following parts of sections and elements taken from Dienemann *et al.* were adapted to match the format and structure of this work:

- Numbering of figures in the text including cross-references
- Numbering of tables in the text including cross-references
- Reference to Movie S1 was removed

Other publications:

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

TABLE OF CONTENTS

Members of the Thesis Committee II

Affidavit III

Acknowledgements IV

Publications V

Table of contents VIII

Summary XIII

1 Introduction 1

1.1 Transcription 1

1.2 RNA polymerase II 2

1.3 The RNA polymerase II transcription cycle 4

1.4 Transcription Initiation 5

1.5 Structural studies of transcription initiation 8

1.6 Promoter DNA Opening 8

1.7 The Pol I and Pol III transcription systems 9

1.8 Aim of this work 10

2 Materials 12

2.1 Chemicals 12

2.2 Antibiotics, media and additives 13

2.3 Buffers and solutions 13

2.4 Bacteria strains 15

2.5 Yeast strains 15

2.6 Plasmids and oligonucleotides 16

2.7 Enzymes 17

2.8 Kits 18

3 Methods 19

3.1 PCR 19

3.2 Site-directed mutagenesis 20

3.3 Agarose gel electrophoresis 20

3.4 In fusion cloning 20

3.5 Transformation of chemically competent *E. coli* 20

3.6 Isolation of PCR products and plasmid DNA 21

3.7 SDS-PAGE 21

3.8 Expression and solubility test 21

3.9 Expression and purification of TBP* 22

3.9.1 Expression 22

3.9.2 Purification 22

3.10 Expression and purification of TFIIAΔ113* 23

3.10.1 Expression 23

3.10.2 Purification 23

3.11 Gradient ultracentrifugation 23

3.12 Ssl2 nuclear depletion and 4sU-sequencing† 24

3.13 Bioinformatics data analysis† 24

3.14 DNA opening assay† 25

3.15 Preparation of cPICs† 25

3.16 Cryo-EM and image processing of cCC1 and CC^{dist}† 26

3.17 Data processing and CC^{dist} reconstruction† 27

3.18 Model building and refinement † 27

4 Results 29

4.1 Preparation of transcription factors and core pre-initiation complexes 29

4.1.1 Cloning, expression and purification of full length TBP 29

4.1.2 Cloning, expression and purification of TFIIAΔ113 31

4.1.3 Preparation of cPICs by gradient ultracentrifugation 33

4.2 Structural and functional investigation of promoter opening 35

4.2.1 TFIIH translocase is not generally required for transcription *in vivo* 35

4.2.2 Translocase requirement correlates with promoter stability 36

4.2.3 Promoter stability defines DNA meltability 37

4.2.4 Cryo-EM reveals distinct closed complex intermediates 38

4.2.5 CC^{dist} contains distorted DNA in a closed cleft 40

4.2.6 Complete PIC structure with distorted DNA 42

4.2.7 Promoters of Pol I and Pol III contain unstable IMRs 43

5 Discussion 45

5.1 A new transcription initiation intermediate 45

5.2 DNA distortion and initial opening 45

5.3 General features of transcription initiation 47

5.4 Roles of PIC components during promoter DNA melting 48

5.4.1 TFIIB 48

5.4.2 TFIIF subunit Tfg2 winged helix domain 49

5.4.3 TFIIE 51

5.4.4 TFIIH 52

5.4.5 RNA polymerase II elements 53

5.4.6 Other transcription initiation factors 55

5.5 Evolution of promoter DNA opening 55

5.6 Open questions and future perspective 57

5.6.1 TFIIH-independent transcription in humans cells 58

5.6.2 Transcription start site scanning by the human PIC 59

5.6.3 Promoter opening intermediates 60

5.6.4 Energy landscape of promoter opening 61

5.6.5 Equivalents of CC^{dist} and other transition states in Pol I, Pol III and bacterial PICs 62

6 Supplemental Materials 63

6.1 Supplemental Figures 63

6.2 Supplemental Tables 68

7 References 70

8 List of Figures 82

9 List of Tables 84

10 List of Abbreviations 86

11 Curriculum Vitae 87

SUMMARY

Transcription of genes is a central process of life that defines how genetic information is used. This determines the metabolic state, morphology and fate of a cell, tissue, organ and the entire organism. In eukaryotes, the most regulated step of transcription is initiation during which RNA polymerase II (Pol II) is specifically recruited to the gene promoter by assembling with transcription factors (TF) –IIA, -IIB, -IID, -IIF, -IIE, -IIH and the co-activator Mediator complex. For initial transcription, the double helix of the promoter DNA of the closed complex (CC) has to be opened in the initially melted region (IMR). The open complex (OC) then provides the single stranded template DNA that can be transcribed. Promoter opening was shown to require the DNA translocase activity of TFIIH; however, spontaneous DNA melting could also be observed. The molecular pathway of TFIIH dependent and spontaneous promoter melting is still unknown due to a lack of high resolution structural studies targeting intermediates of the CC-to-OC transition.

Here I report the structure of a new transcription initiation intermediate just before promoter DNA opening at ~5 Å resolution. This structure reveals closed distorted promoter DNA in the Pol II active center cleft. The distortion is induced by the closed Pol II clamp domain and stabilized by structural elements of TFIIF and TFIIE. The distorted DNA shows a helical axis offset and underwinding of the closed promoter DNA that weaken the DNA and prime it for melting. I further present *in vivo* and *in vitro* data showing that the DNA duplex stability of the IMR determines whether DNA spontaneously melts after distortion or whether it remains closed and requires TFIIH activity to be melted. By systematic comparison of Pol I and Pol III promoters with the Pol II system, I find that promoters of the other nuclear RNA polymerases also contain a weak DNA duplex in their IMR. This explains why they do not need a TFIIH homologue or ATP-hydrolysis to open promoter DNA. This suggests a unified mechanism of promoter DNA melting that involves DNA distortion by clamp closure and subsequent melting of the IMR. Taken together, these results greatly improve our understanding of the molecular mechanism of promoter DNA melting by nuclear RNA polymerases and for the first time provide a general mechanism for promoter opening

1 INTRODUCTION

1.1 Transcription

The blueprint of life is stored as genetic information in nucleic acids. This information defines the shape of all cells, tissues, organs and organisms. With the exception of a few life forms as viruses, the storage molecule for this information is deoxyribonucleic acid, short DNA. A central dogma of biology describes that genetic information is copied from DNA into RNA before the RNA sequence is translated into proteins, which are the molecules providing most of the biologic activity (Crick, 1970)(Figure 1). The two steps between DNA and RNA as well as between RNA and protein are major regulatory steps that influence development, growth, death and metabolism of a cell. The genetic information is stored in genes, which are defined regions of the DNA that are copied into an RNA molecule by RNA polymerases (RNAPs). This process is called transcription and starts in a region of a gene that is called promoter.

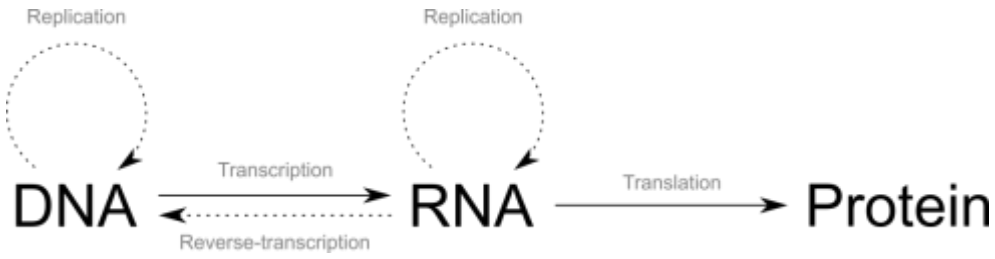


Figure 1: The central dogma of molecular biology. Adapted from Crick, 1970

The simplest transcription systems can be found in viruses and phages (Cheetham and Steitz, 2000). They often contain an RNAP composed of a single subunit and use highly specialized promoter sequences to achieve regulatory control over a gene (Cheetham et al., 1999; Rosa, 1979; Sousa et al., 1993). Prokaryotes use a more complex multi-subunit RNAP consisting of 5 subunits (Murakami, 2015). Transcriptional regulation at prokaryotic promoters depends on the so called σ-factors that recognize special DNA sequences and can be controlled by activator proteins (Feklistov et al., 2014). Eukaryotic transcription is performed by nuclear RNAPs that have specialized on certain types of

genes (Roeder and Rutter, 1969; Vannini and Cramer, 2012). The ribosomal RNA (rRNA) precursors are transcribed by Pol I. Messenger RNAs (mRNA) that encode for proteins and a number of non-coding RNAs (ncRNA) are made by Pol II. Pol III produces the transfer RNAs (tRNA) that can be coupled to amino acids and are used to translate mRNA to proteins. The number of regulatory factors used for nuclear transcription is greatly increased compared to prokaryotic transcription. Many transcriptional regulators act at the beginning of transcription (initiation) and during transcription itself (Shandilya and Roberts, 2012).

1.2 RNA polymerase II

As all DNA dependent RNA polymerases, Pol II catalyzes phosphodiester bond formation between a nascent RNA chain 3'-OH group and the 5' phosphate of a nucleoside triphosphate in a two metal-ion dependent manner (Steitz, 1998). Structural studies of Pol II in several catalytic intermediate and inhibited states elucidated the Brownian ratchet mechanism that drives the catalytic cycle of nucleotide addition (Brueckner and Cramer, 2008; Brueckner et al., 2009; Cheung and Cramer, 2012; Gnatt et al., 2001). Misincorporation of a nucleotide leads to Pol II stalling and backtracking, an off-line state that can be resolved by activating the intrinsic cleavage activity of Pol II by TFIIS (Cheung and Cramer, 2011; Kettenberger et al., 2003).

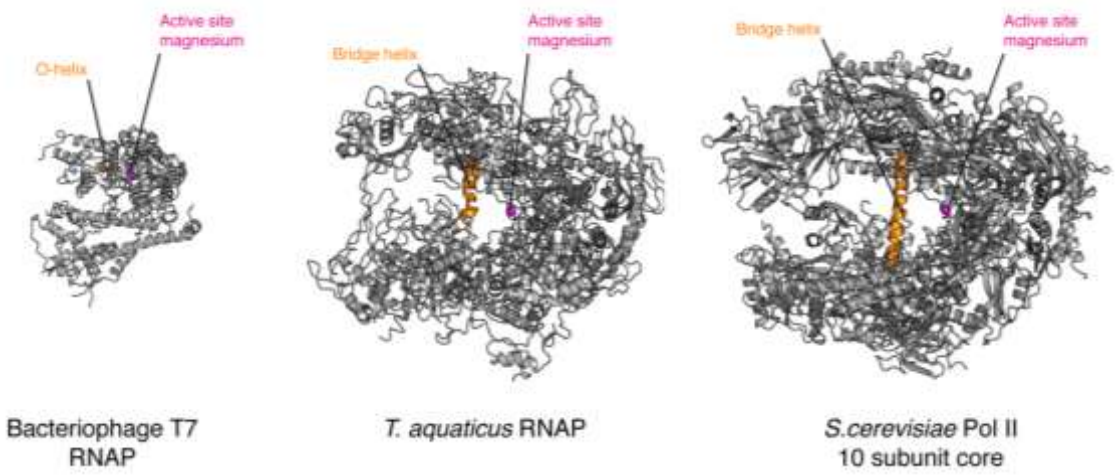


Figure 2: RNA polymerases of T7 phage, bacteria and eukaryotes. Figure adapted from PhD thesis of H. Hillen.

Pol II consists of 12 subunits of which some are shared with Pol I and Pol III (Table 1). Predicted by functional homology (Sweetser et al., 1987), sequence similarity (Lane and Darst, 2010a) and confirmed by the structures of eukaryotic Pol II (Cramer et al., 2000; Cramer et al., 2001; Gnatt et al., 2001) and bacterial RNAP (Zhang et al., 1999), the 10-subunit Pol II core is highly homologous to the bacterial RNAP (Lane and Darst, 2010b).

Table 1: Sub-units of the three nuclear RNA polymerases. Corresponding subunits of Pol I, Pol II and Pol III are shown together with their function. Functional counterparts of the Pol II system that are not Pol II subunits are shown in italic. Table was modified from Vannini and Cramer, 2012.

Pol I	**Pol II**	**Pol III**	**Function**
Polymerase Core			
A190	Rpb1	C160	Active center
A135	Rpb2	C128	Active center
AC40	Rpb3	AC40	
Rpb5	Rpb5	Rbp5	
Rpb6	Rpb6	Rpb6	
Rpb8	Rpb8	Rpb8	
A12.2 N ribbon	Rpb9	C 11 N ribbon	RNA cleavage
Rpb10	Rpb10	Rpb10	
AC19	Rpb11	AC19	
Rpb12	Rpb12	Rpb12	
Polymerase Stalk			
A14	Rpb4	C17	Initiation complex formaton
A43	Rpb7	C25	Initiation complex formaton
Additional subunits			
A12.2 C ribbon	*TFIIS*	C11 C ribbon	RNA cleavage

A49 N domain	*Tfg1*	C37	Initiation complex stabilization, start site selection
A34.5	*Tfg2*	C53	Initiation complex stabilization, start site selection
A49 C domain	*TFIIE*	C34, C82, C31	Open complex stabilization

A conservation of functional elements between single-subunit RNAPs and Pol II has also been proposed (Cramer, 2002)(Figure 2). The structures of the complete 12-subunit Pol II enzyme revealed the auxiliary Rpb4/7 stalk domain that was evolutionary acquired by Pol II in eukaryotes to increase regulatability during the process of transcription initiation (Armache et al., 2003). In higher eukaryotes, the largest subunit Rpb1 also contains a long repetitive C-terminal domain (Pol II CTD). The CTD is differentially phosphorylates throughout the transcription cycle and thereby serves as a binding hub for regulatory factors (Eick and Geyer, 2013).

The 10-subunit Pol II core can be divided into numerous functional and structural domains (Cramer et al., 2001) of which the mobile clamp domain plays an important role. The clamp opens for DNA insertion into the active center cleft during initiation and then closes over the DNA-RNA hybrid to form a stable elongation complex (Gnatt et al., 2001). The state of the mobile clamp can be modulated by regulatory factors such as by TFIIS (Kettenberger et al., 2003) and Spt4/5 (Martinez-Rucobo et al., 2011) and therefore is a central hub to regulate Pol II activity throughout the transcription cycle.

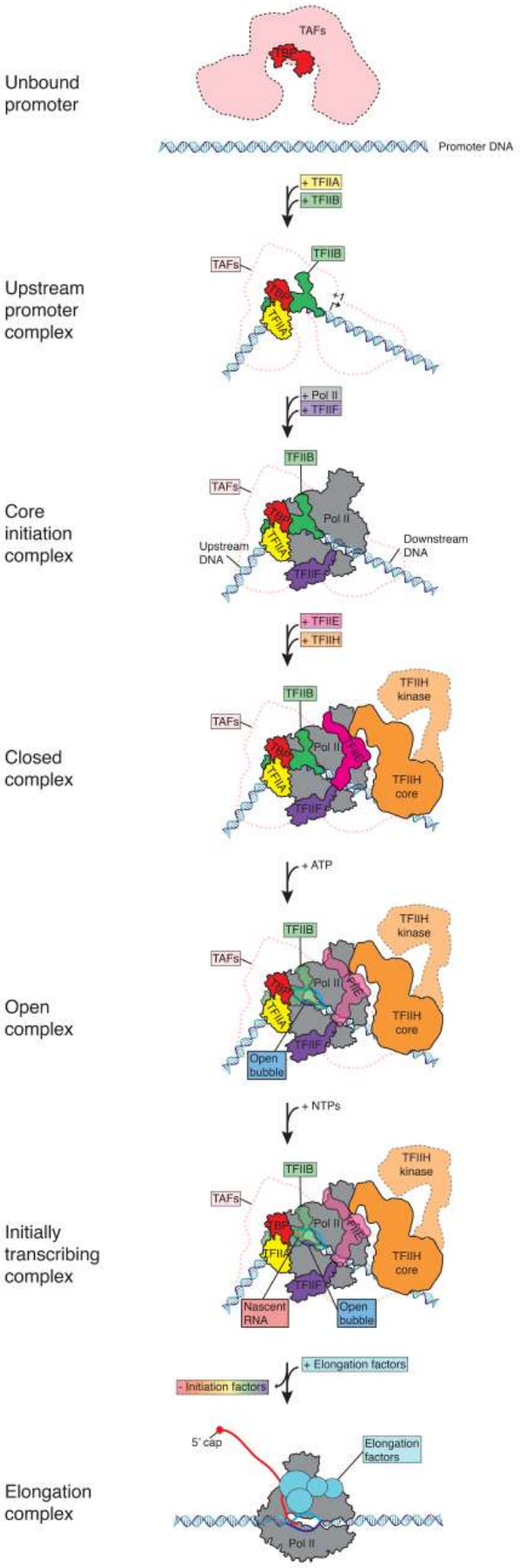
1.3 The RNA polymerase II transcription cycle

The transcription cycle of Pol II can be divided in pre-initiation, initiation, elongation, and termination and recycling (Sainsbury et al., 2015; Shandilya and Roberts, 2012). Prior to transcription initiation, basal transcription factors (TFs) assemble at the promoter site of a

gene to form a pre-initiation complex (PIC)(Sainsbury et al., 2015). This complex comprises up to 60 proteins including Pol II, basal TFs and the Mediator co-activator complex. Before initiation (i.e. RNA synthesis) can start, the DNA double helix is unwound and separated into single strands by the PIC. Abortive synthesis of short RNA products marks the point of initiation. After RNA polymerase escapes the promoter, transcription elongation proceeds through the gene body. Elongating Pol II encounters pausing sites (Adelman and Lis, 2012), nucleosomes (Kulaeva et al., 2013) and splicing events (Merkhofer et al., 2014) which are regulatory checkpoints for Pol II transcription. At the end of a gene, Pol II termination it triggered by the poly(A)-signal leading to mRNA poly-adenylation and cleavage (Mandel et al., 2008). A G/C rich stretch marks the ultimate transcription termination site (Schwalb et al., 2016) and causes Pol II to dissociate from the DNA template. Pol II is then recycled by re-initiation (Hahn, 2004; Rani et al., 2004), likely facilitated by gene looping (Hampsey et al., 2011; O'Sullivan et al., 2004).

1.4 Transcription Initiation

A set of basal transcription initiation factors is required to assemble in a stepwise manner to start transcription from a gene promoter (Sainsbury et al., 2015)(Table 2, Figure 3). TATA binding protein (TBP, as part of TFIID) binds to the TATA box sequence and bends the DNA by 90°. TFIID also binds to promoters of TATA-less genes by recognizing downstream elements or association with the first nucleosome in the gene body (Warfield et al., 2017). TFIIA aids TFIID to bind to promoter DNA. The DNA regions around the bent TATA box, the so called TFIIB recognition elements (BREs), are then bound by the cyclin domains of TFIIB. The assembly of DNA, TBP, TFIIA and TFIIB forms the upstream promoter complex that serves as a platform to specifically recruit Pol II to the promoter. The TFIIB-linker (B-linker) and zinc ribbon (B-ribbon) domains facilitate recruitment of the Pol II-TFIIF complex to the upstream promoter assembly. TFIIE binding to Pol II forms the core pre-initiation complex (cPIC). TFIIE can then recruit TFIIH by various interactions (Maxon et al., 1994; Schilbach et al., 2017), forming the full pre-initiation complex (PIC). The DNA translocase activity of the TFIIH subunit Ssl2 (Fishburn et al., 2015) opens the initially melted region (IMR) of the promoter and inserts the DNA into the Pol II active center cleft. Whereas the transcription start site (TSS) is located within the IMR in most eukaryotes (Juven-Gershon and Kadonaga, 2010), some



yeast species utilize the translocase activity of Ssl2 to scan for a TSS that can be located further downstream (Fishburn et al., 2016; Jin and Kaplan, 2014; Kuehner and Brow, 2006). The TSS is recognized by the TFIIB reader domain and Pol II (Kaplan et al., 2012; Sainsbury et al., 2013) by its conserved motif (Zhang and Dietrich, 2005). The recognition allows stable association of template DNA with the Pol II active site and makes phosphodiester bond formation possible. Pol II then starts to synthesize short RNAs. Due to unstable nature of short 2-6 nucleotides (nt) RNA-DNA hybrids, the newly made RNA is prone to dissociate which makes Pol II undergo several rounds of

Figure 3: Scheme of transcription initiation by RNA polymerase II

The canonical model for transcription initiation is shown. TBP as part of TFIID binds to TATA DNA and induces a bend. TFIIA and TFIIB bind TBP to form the upstream promoter complex. TFIIB then recruits the Pol II-TFIIF complex. Binding of TFIIE and TFIIH completes PIC formation. DNA opening is achieved by ATP hydrolysis by TFIIH. Initial transcription is followed by the dissociation of transcription factors and Pol II escaping the promoter. Figure adapted from Sainsbury *et al.*, 2015.

abortive transcription (Luse, 2013; Luse and Jacob, 1987). After overcoming abortive transcription, the growing RNA chain clashes with the TFIIB reader (B-reader) and B-ribbon domains in the Pol II active center cleft (Kostrewa et al., 2009; Sainsbury et al., 2013). This causes TFIIB to be released from Pol II and the PIC to be disassembled. Basal TFs associated with TFIID, -E and H as well as Mediator can remain bound to the promoter site to serve as a scaffold for re-initiation events (Yudkovsky et al., 2000).

Table 2: Components of the basal transcription initiation machinery and their function. The composition of different sub-complexes is indicated.

Sub-complex			Component	Subunits	Function
Pre-initiation complex (PIC)	Core pre-initiation complex (cPIC)	Upstream promoter complex	TBP	As part of TFIID	Binding of TATA DNA
			TFIIA	Toa1, Toa2	Binding of TBP
			TFIIB		Bridging between TBP and Pol II, promoter melting, start site selection
			Pol II	Rpb1 - 12	Polymerization of NTPs to an RNA chain
			TFIIF	Tfg1, Tfg2, Tfg3 (non-essential)	Start site scanning/selection, promoter opening
			TFIIE	Tfa1, Tfa2	Promoter opening, recruitment of TFIIH
			TFIIH	Tfb1, Tfb2, Tfb3, Tfb4, Tfb5, Ssl1, Ssl2, Rad3, Kin28	Promoter opening, start site scanning, CTD-phosphorylation

1.5 Structural studies of transcription initiation

Transcription initiation has been extensively studied structurally. Crystal structures of TBP (Chasman et al., 1993), TBP with TATA box DNA (Kim et al., 1993) and in complex with TFIIA (Tan et al., 1996) and the TFIIB cyclin domains (Nikolov et al., 1995) revealed the structure of the upstream promoter complex. These structures show how TBP bends the TATA DNA by ~90°. The first structure of Pol II with a TF was a complex of Pol II and TFIIB (Kostrewa et al., 2009; Sainsbury et al., 2013) suggesting a first model of PIC architecture and how the TSS is recognized by TFIIB. Crosslinking coupled to mass spectrometry extended architectural models by obtaining high-confidence distance restraints for the position of TFIIF and TFIIE (Chen et al., 2007; Eichner et al., 2010). Crosslinking of an *in-vitro* assembled cPIC revealed the detailed positions of TF domains within the cPIC (Muhlbacher et al., 2014). Electron microscopy studies of cPICs and PICs in closed (CC), open (OC) and initially transcribing (ITC) states (He et al., 2013) represent the first medium resolution structures of complete Pol II transcription initiation complexes where TFs could be resolved on the domain level. These structures enabled first mechanistic insight for the transition between CC and OC. The architecture of a core ITC with the Mediator co-activator complex (Plaschka et al., 2015) extended the model for transcriptional activation and regulation. Cryo-electron microscopy studies of closed and open cPICs and PICs at medium to high resolution (He et al., 2016; Murakami et al., 2015; Plaschka et al., 2016) resolved various states at great detail and proposed molecular mechanisms for the CC to OC transition. However, TFIIH remained at low resolution in these studies. Recent advances in sample preparation and cryo-EM data analysis improved the resolution of TFIIH dramatically providing the first near-atomic models of a complete PIC (Greber et al., 2017; Schilbach et al., 2017).

1.6 Promoter DNA Opening

Promoter DNA opening is an integral step during transcription initiation as the template DNA strand has to be single stranded to be transcribed by Pol II. The discovery that ATP stimulates promoter DNA melting in nuclear extracts (Wang et al., 1992) raised the question, which transcription factors are involved and whether enzymatic activity is required or not. TFIIH was identified to be recruited to gene promoters, possess ATPase

activity and aid promoter opening (Dvir et al., 1996). On supercoiled DNA templates, TFIIH is not required for promoter opening and to initiate transcription (Goodrich and Tjian, 1994; Holstege et al., 1995; Parvin and Sharp, 1993; Tyree et al., 1993), indicating that the torsional energy stored in DNA supercoiling is enough to facilitate DNA opening. Abortive and processive transcription from linear templates could also be observed (Goodrich and Tjian, 1994; Pan and Greenblatt, 1994), suggesting that low levels of ATP-independent promoter opening exist on linear relaxed templates. The upstream edge of the initially melted region (IMR) was found to be located 19-22 bp downstream of the TATA box in vivo (Giardina and Lis, 1993). Permanganate foot printing revealed the downstream edge of the initially melted region around 29-33 bp downstream of the TATA sequence (Holstege et al., 1997). Thus, the mature, fully opened transcription initiation bubble is 12-14 bp in size and is initially formed ~20-33 bp downstream of TATA. Preforming a mis-matched bubble from base 29 and 32 was shown to ablate TFIIH requirement completely (Holstege et al., 1996), indicating that TFIIH activity is only required to open the downstream part of the IMR. Recent mechanistic single molecule studies identified a 5-6 bp bubble intermediate during TFIIH dependent promoter opening (Tomko et al., 2017), suggesting a model in which TFIIH might be required to open a minimal DNA region at the downstream edge of the IMR. TFIIF and TFIIE strongly enhance transcription from small hetero-duplexes (Holstege et al., 1996; Pan and Greenblatt, 1994), demonstrating a stimulatory role for extending small premelted transcription bubbles. More recent studies showed that TFIIH independent promoter opening (Plaschka et al., 2016) and transcription (Alekseev et al., 2017) can occur *in vitro* and *in vivo* and raised questions about TFIIH requirement for DNA melting and it role in transcription initiation.

1.7 The Pol I and Pol III transcription systems

Besides Pol II, two other eukaryotic polymerases perform transcription in the nucleus (Roeder and Rutter, 1969). RNA polymerase I (Pol I) transcribes the 25S rRNA precursor that is processed into ribosomal RNAs (rRNA)(Moss et al., 2007). rRNA genes are present in many copies and show high transcription activity during cell growth (Moss et al., 2007). Pol I transcription is regulated by transcription factors with very little homology to the TFs in the Pol II system (Vannini and Cramer, 2012). rRNA gene transcription is mainly regulated by the core-factor and Rrn3 (Bedwell et al., 2012; Keener et al., 1998; Pilsl et al.,

2016), and the upstream activating complex (Hontz et al., 2008; Keys et al., 1996) which do not possess ATPase or helicase/translocase activity that could aid promoter DNA melting. Structural studies gave insights into the Pol I initiation mechanism (Engel et al., 2017; Han et al., 2017; Sadian et al., 2017), however, the ATP-independent DNA melting mechanism of Pol I remains poorly understood.

The third RNA polymerase (Pol III) synthesizes small RNAs that are important for various cellular functions, such as small nuclear RNAs (snRNA), transfer RNA (tRNA), the 5S ribosomal RNA (5S rRNA) and U6 splicosomal RNA. Transcriptional regulation of Pol III depends on TFs that are similar to the Pol II system (Vannini and Cramer, 2012). The TFIIIB complex, which contains TBP, Brf1 (a TFIIB paralogue) and Bdp1, is the main regulator of Pol III transcription (Kassavetis and Geiduschek, 2006). Similar to the Pol I system, Pol III does not require ATP-dependent helicase/translocase activity in order to initiate transcription. Recent structural studies on Pol III transcription initiation resolved closed, open and initially transcribing complexes (Abascal-Palacios et al., 2018; Vorländer et al., 2018). The molecular basis of the spontaneous DNA melting by Pol III could not be revealed in detail, however, indications for base-flipping similar to the bacterial system could be found (Abascal-Palacios et al., 2018). Additionally, the closed complex structures suggest that clamp opening and closing has to take place before promoter DNA loading followed by melting and initial transcription can occur (Vorländer et al., 2018).

1.8 Aim of this work

The structural basis of transcription initiation has been studied extensively over the past decades and biochemical studies gave detailed insight into the mechanism of promoter DNA melting and transcription initiation. Despite our broad understanding of these processes provided by these studies, several fundamental questions remain unanswered:

1. The region of DNA that is initially melted during promoter opening was repeatedly identified and can be now defined with high confidence to be around 20-33 bp downstream of the TATA box sequence. But what are the molecular determinants that define where promoter DNA melting occurs?

2. Promoter opening by the Pol II PIC was shown to occur in an ATP-dependent and independent manner. Do these mechanisms differ or do they follow the same molecular pathway from closed to open promoter DNA?
3. The other eukaryotic RNA polymerases do not require ATP hydrolysis for promoter opening. How do the mechanisms of promoter DNA opening by the other RNA polymerases differ from the Pol II system?

In this work, a combination of functional genomics, biochemistry and structural biology will be applied to provide answers to these questions. Yeast genetics and next-generation sequencing will be used to investigate the importance of TFIIH for transcription initiation *in vivo* and to find determinants for promoter meltability. Fluorescence based promoter melting assays will be used to investigate TFIIH-independent promoter melting *in vitro*. To determine structures of functional transition states during the CC-to-OC conversion, cryo-EM will be used.

2 MATERIALS

2.1 Chemicals

Table 3: Chemicals used in this work

Chemical	Origin
Acetic acid	Merck
Agarose	Invitrogen
Bromphenol blue	Roth
Dithiothreitol (DTT)	Roth
EDTA	Roth
Glucose	Merck
Glycerol	Roth
HCl	Merck
Hepes	Roth
KCl	Merck
KOH	Merck
NaCl	Merck
Pepton	BD Biosciences
SDS	Merck
Sucrose	Merck
SYBRsafe	Invitrogen
Tris	VWR
Trypton	BD Biosciences
Yeast extract	BD Biosciences
β-mercapto ethanol	Roth

2.2 Antibiotics, media and additives

Table 4: Antibiotics used in this work

Antibiotic	Stock (working)	Company
Ampicillin	100 mg/ml (100 µg/ml)	Roth
Chloramphenicol	34 mg/ml (34 µg/ml)	Roth

Table 5: Culture media used in this work

Medium	Composition
1x LB medium (1 liter)	10 g NaCl, 10 g Trypton, 5 g Yeast extract
1x YPD medium (1 liter)	10 g Yeast extract, 20 g Peptone, 20 g Glucose

Table 6: Culture media additives used in this work

Additive	Stock	Purpose	Company
IPTG	1 M	Induction of lacI repressed genes in transformed *E. coli*	Roth

2.3 Buffers and solutions

Table 7: Buffers used for TBP purification

Buffer	Composition
A500	25 mM Hepes-KOH, pH 7.4, 500 mM KCl, 2.5 mM DTT, protease inhibitors
A1000	25 mM Hepes-KOH, pH 7.4, 1 M KCl, 2.5 mM DTT, protease inhibitors
B	25 mM Hepes-KOH, pH 7.4, 500 mM KCl, 350 mM Imidazole, 2.5 mM DTT, protease inhibitors
A0	25 mM Hepes-KOH, pH 7.4, 2.5 mM DTT, protease inhibitors

C	10 mM Hepes-KOH, pH 7.0, 200 mM KCl, 5% Glycerol, 2 mM DTT

Table 8: Buffers used for TFIIAΔ113 purification

Buffer	**Composition**
A500	25 mM Tris-HCl, pH 8.0, 500 mM NaCl, 10% Glycerol ,2.5 mM DTT, protease inhibitors
A1000	25 mM Tris-HCl, pH 8.0, 1 M NaCl, 10% Glycerol, 2.5 mM DTT, protease inhibitors
B	25 mM Tris-HCl, pH 8.0, 500 mM NaCl, 10% Glycerol, 250 mM Imidazole, 2.5 mM DTT, protease inhibitors
A0	25 mM Tris-HCl, pH 8.0, 10% Glycerol, 2.5 mM DTT, protease inhibitors
C	10 mM Hepes-KOH, pH 7.0, 200 mM KCl, 5% Glycerol, 2 mM DTT

Table 9: Other buffers and solutions used in this work.

Buffer/Solution	**Composition/Company**	**Purpose**
2x SDS loading dye	150 mM Tris-HCl, pH 6.8, 1.2% SDS, 30% Glycerol, 2 mM β-mercapto ethanol	SDS-PAGE
SDS-PAGE running buffer	20x NuPAGE™ MES/MOPS buffer (Invitrogen)	SDS-PAGE
SDS-PAGE gel stain	InstantBlue (Expedeon)	SDS-PAGE
6x DNA loading dye	DNA Gel loading dye (Thermo Fischer)	Agarose gel electrophoresis
Agarose gel running buffer	5 mM EDTA pH 8.0 at 20°C, 250 mM Tris-acetate	Agarose gel electrophoresis
SYBRsafe DNA stain	Invitrogen	Staining od DNA in agarose gels

10x gradient buffer	100 mM Hepes-KOH, pH 7.0, 2M KCl, 50% Glycerol, 20 mM DTT

2.4 Bacteria strains

Table 10: *E. coli* strains used in this work

Strain	Genotype	Origin
BL21(DE3)pRIL	E. coli B F- ompT hsdS(rB- mB-) dcm+ Tetr E. coli gal λ (DE3) endA Hte [argU ileY leuW Camr]	Agilent
XL1 blue	recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F' proAB lacIqZΔM15 Tn10 (Tetr)]	Agilent

2.5 Yeast strains

Table 11: *S. cerevisiae* strains used in this work

Strain	Genotype	Origin
Y40343	W303; MATα ade2-1 trp1-1 can1-100 leu2-3,112 his3-11,15 ura3 GAL psi+ tor1-1 fpr1::NAT RPL13A-2xFKB12::TRP1	Haruki et al., 2008 Euroscarf
Y40343 Ssl2-FRB	W303; MATα ade2-1 trp1-1 can1-100 leu2-3,112 his3-11,15 ura3 GAL psi+ tor1-1 fpr1::NAT RPL13A-2xFKB12::TRP1 YIL143C-FRB::kanMX6	This work

2.6 Plasmids and oligonucleotides

Table 12: Vectors used in this work

Vector	Resistance	Expressed construct	Origin
pOPINE-TFIIAd95	Amp	TFIIA subunits Toa1Δ95-209 and Toa2 with C-terminal His-6 tag	This work
pOPINE-flTBP	Amp	Full length TBP with C-terminal 6x histidine tag	This work
pETDuet-TFIIA	Amp	Codon-optimized TFIIA subunits Toa1 and Toa2; includes frameshift mutation	Provided by Wolfgang Mühlbacher

Table 13: Oligonucleotides used in this work. 2AP: 2-amino purine

Oligo	Sequence	Purpose
P1_TPB_pOPINE_fwd	AGGAGATATACCATGGCCGATGAGGAACGTTTAAAG	Cloning of TBP into pOPINE
P2_TPB_pOPINE_rev	GTGATGGTGATGTTTCATTTTTCTAAATTCACTTAGCACAGGG	Cloning of TBP into pOPINE
P25_Toa1_pOPINE_F	AGGAGATATACCATGAGCAATGCCGAAGCAAG	Recloning TFIIA into pOPINE
P26_Toa2_pOPINE_R	GTGATGGTGATGTTTCTCGCTTTTTTTGCTATTGCAGG	Recloning TFIIA into pOPINE
P5_Toa1d95-209_fwd	GCAGCGAATTTAACATCAAAGAAGAGAACACCGGTAGCGCACTGCTGGATAC	Mutagenesis PCR of TFIIA

P6_Toa1d95_209_rev	TAGTTTCTTCTCTTGTGGCCATCGCGTG ACGACCTATGGCTACTTCAAC	Mutagenesis PCR of TFIIA
His4-2AP+26-nt	GCACGCTGTGTATATAATAGCTATGGA ACGTTCGATTC(2AP)CCTCCGATGTGTG TTGTACATACATAAAAATATCA	2 amino purine assay
His4-2AP+26-t	TGATATTTTTATGTATGTACAACACAC ATCGGAGGTGAATCGAACGTTCCATAG CTATTATATACACAGCGTGC	2 amino purine assay
Gat1-2AP+26-nt	CCCAGCCACATATATATAGGTGTGTGC CACTCCCGGCC(2AP)CGGTATTAGCAT GCACGTTTTCTTTCCTTTGCTTTT	2 amino purine assay
Gat1-2AP+26-t	AAAAGCAAAGGAAAGAAAACGTGCAT GCTAATACCGTGGCCGGGAGTGGCAC ACACCTATATATATGTGGCTGGG	2 amino purine assay
Gat1-nt	GCGGTGCCCGGCCCAGCCACATATATA TAGGTGTGTGCCACTCCCGGCCCCGGT ATTAGCATGCACGTTTTCTTTCCTTTGC TTT	Gradient ultracentrifugation
Gat1-t	AAAGCAAAGGAAAGAAAACGTGCATG CTAATACCGGGGCCGGGAGTGGCACA CACCTATATATATGTGGCTGGGCCGGG CACCGC	Gradient ultracentrifugation

2.7 Enzymes

Table 14: Enzymes used in this work

Enzyme	**Company**
DpnI	New England Biolabs
PmeI	New England Biolabs

NcoI	New England Biolabs
Phusion High Fidelity Polymerase	New England Biolabs

2.8 Kits

Table 15: Kits used in this work

Kit	**Purpose**	**Company**
Plasmid preparation kit	Plasmid extraction from E. coli	QIAGEN
In-fusion liquid kit	In-fusion cloning reaction	Takara Biotech
PCR Purification kit	Purification of inserts and linearized vector	QIAGEN

3 METHODS

Some of the methods described in this section were published in a reduced form in

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

These methods are marked with an asterisk (*).

Furthermore, some methods described in this section were done in collaboration with coauthors of a manuscript that has been submitted:

> Dienemann C, Schwalb B, Schilbach S, Cramer P. Promoter distortion and opening in the RNA polymerase II cleft. *Molecular Cell*, Online 21 November 2018

Methods that are part of this manuscript marked with a cross (†) and contributions of co-authors other than the author of this thesis are indicated.

A detailed list of author contributions and text and figures taken from the manuscripts is given in section Publications.

3.1 PCR

Polymerase chain reactions were set up in 50 µl using 10-50 ng template DNA, 0.5 µM of each primer, 200 µM dNTPs and 1 U Phusion polymerase. Thermal cycling was set up with 1 minute initial denaturation at 95°C. Cycling was done with 30 seconds denaturation (95°C), 30 seconds primer annealing (annealing temperature of the primer) and elongation at 72°C for 1 min/kb. The PCR was finalized with 10 minutes elongation at 72°C. The same protocol was used for mutagenesis PCR.

3.2 Site-directed mutagenesis

Site directed mutagenesis was done using the QuickChange approach. After amplification of the parental plasmid with the mutation primers, the PCR reaction was digested with 10 U DpnI for 2-3 hours at 37°C. The reaction was then transformed into chemically competent XL1 blue cells. Clones were screened and confirmed by PCR using insert specific primers and sequencing.

3.3 Agarose gel electrophoresis

Agarose gels were prepared from TAE buffer with 1% (w/v) agarose. The solution was heated in the microwave, poured to the gel chamber and left at room temperature to solidify. SYBRsafe was added before pouring for later DNA detection by UV light. Gels were run at 100 V for 20-30 minutes and visualized on a UV light imager.

3.4 In fusion cloning

The pOPINE target vector (Oxford Protein Production Facility) was linearized using 10 U NcoI/PmeI per 1 μg DNA at 37°C overnight. Linearized vectors were purified using a PCR product purification kit. In-fusion cloning was done by mixing 100 ng linearized vector with 2-fold molar excess of the purified insert. Water was added to 10 μl. The reaction was incubated for 15 minutes at 37°C and then 15 minutes at 50°C. 2.5 μl of the reaction were transformed in chemically competent *E. coli*. Colonies were grown on LB-agar plates with ampicillin. Clones were screened by colony PCR and confirmed by sequencing.

3.5 Transformation of chemically competent *E. coli*

DNA was mixed with 50 μl chemically competent E. coli and incubated for 20-30 minutes on ice. Cells were then heat shocked for 45 seconds at 42°C, kept on ice for 2 minutes and then recovered in 1 ml LB with 1% glucose. Recovery was performed at 37°C for 1-2 hours while shaking at 1000 rpm.

3.6 Isolation of PCR products and plasmid DNA

Purification of PCR products and plasmid DNA was done following the protocol of the respective kit used.

3.7 SDS-PAGE

SDS-PAGE was performed using pre-cast NuPAGE Bis-Tris 4-12% gels (Invitrogen). Samples were boiled with 2x SDS loading buffer prior to loading. Gels were run in NuPAGE MES/MOPS running buffers for 30-60 minutes at 200 V. Bands were visualized by InstantBlue staining. Destaining was done in water.

3.8 Expression and solubility test

For expression testing, 100 ml overnight culture in 1x LB supplemented with ampicillin and chloramphenicol was inoculated with a colony of transformed BL21(DE3)pRIL cells. Cells were grown at 37°C while shaking. From this overnight culture, 1 ml was used to inoculate 100 ml test expression culture in 1x LB with ampicillin and chloramphenicol. Cell growth was monitored by measuring OD at 600 nm. Once the OD reached 1.0, the culture was induced with 0.5 mM IPTG and samples were taken prior and after induction. For sample analysis, medium was removed by centrifugation and resuspended in 100 µl 2x SDS per unit OD and incubated at 95°C for 10 minutes. 10 µl of these samples were loaded on an SDS-PAGE for analysis.

For solubility testing, 15 ml samples were taken after induction. Medium was removed by centrifugation and cell pellets were resuspended in 1 ml solubility test buffer per unit OD. Cells were lysed by sonication with amplitude of 40% with 0.4 seconds on/off cycles. Lysate was cleared by centrifugation at 15,000 g to separate soluble and insoluble fractions. The insoluble pellet was resuspended in the same volume of 2x SDS loading buffer as solubility test buffer was used for cell resuspension. For SDS-PAGE analysis, 7 µl of supernatant and pellet fractions were used.

3.9 Expression and purification of TBP*

3.9.1 Expression

An overnight preculture of 200-300 ml 1x LB with ampicillin and chloramphenicol was inoculated with a single colony of transformed BL21(DE3)pRIL cells and grown at 37°C while shaking. The expression cultures were inoculated 1:25 and grown in 1x LB ampicillin and chloramphenicol at 37°C until the OD_{600} reached 0.5. Cultures were then cooled on ice for 30 minutes and induced with 0.5 mM IPTG. Expression was done for 3-4 hours at 20°C. Cells were harvested by centrifugation, pellets were flash-frozen in liquid nitrogen and stored at -80°C.

3.9.2 Purification

Cell pellets were thawed in buffer A500 at room temperature. Lysis was done by sonication using 40% amplitude, 0.4 seconds on/off cycles for 15 minutes on ice. The lysate was then cleared by centrifugation at 87,000 g for 45 minutes at 4°C. The soluble supernatant was filtered and applied on a HisTrap HP 5 ml column (GE Healthcare). The column was washed with 10 CV A500 buffer, 11 CV A1000 buffer and 10 CV A500 buffer. Before elution, the column was washed with 35% buffer B. The column was subsequently eluted with 100% buffer B. TBP containing fractions were pooled and diluted 3.3 fold with buffer A0. The sample was then centrifuged for 10 minutes at 87,000 g to remove precipitates.

The diluted sample was applied to a MonoS 5/50 column (GE Healthcare) that was equilibrated in 15% buffer A1000. After washing the column with 15% buffer A1000 for 15 CV, the sample was eluted on a linear gradient from 15 to 50% buffer A1000 over 40 CV. The column was subsequently washed with 100% buffer A1000. TBP containing fractions were pooled and concentrated using an Amicon MWCO 10 kDa filter device. As last polishing step, the concentrated sample was run on a Superose 12 10/300 (GE Healthcare) equilibrated in buffer C. TBP containing fractions were pooled, concentrated and flash-frozen at concentrations between 5 and 8 mg/ml (E0.1% = 0.535). Protein was stored at -80°C

3.10 Expression and purification of TFIIAΔ113*

3.10.1 Expression

Expression was done in BL21(DE3)pRIL cells. 200-300 ml preculture were inoculated from a single colony. Precultures were grown in 1x LB with ampicillin and chloramphenicol at 37°C overnight while shaking. Expression cultures were inoculated 1:50 and grown in 1x LB with ampicillin and chloramphenicol at 37°C until OD_{600} reached 0.5. Expression was induced with 0.5 mM IPTG, left at 37°C and harvested by centrifugation after 3-4h. Cell pellets were flash-frozen in liquid N_2 and stored at -80 °C.

3.10.2 Purification

Frozen cells were thawed and resuspended in buffer A500. Lysis was done by sonication with 40% amplitude and 0.4 seconds on/off cycles for 15 minutes on ice. The cell debris was removed by centrifugation at 87,000 g for 45 minutes. The supernatant was then applied to a HisTrap HP 5 ml column (GE Healthcare) that was equilibrated in buffer A500 before. After sample application, the column was washed with 19 CV buffer A500, 9 CV buffer A1000 and additional 9 CV buffer A500. Column bound proteins were eluted with a 0-100% gradient of buffer B over 15 CV. TFIIAΔ113 containing fractions were pooled and diluted 5-fold with buffer A0. The diluted sample was centrifuged to remove precipitation and applied to a MonoS 5/50 column (GE Healthcare). The column was washed with 15 CV 10% buffer A1000 and step eluted with 50% buffer A1000. After elution, the column was cleaned with 100% buffer A1000. Fractions containing TFIIAΔ113 were pooled and concentrated in an Amicon MWCO 10 kDa centrifugation filter. The concentrated sample was then loaded to a Superdex 75 10/300 column (GE Healthcare) that was equilibrated in buffer C. TFIIAΔ113 containing fractions were concentrated and flash-frozen in liquid N_2 at concentrations of 7-10 mg/ml (E0.1% = 1.116). Aliquots were stored at -80°C.

3.11 Gradient ultracentrifugation

Sucrose gradients from 10-30% were used for ultracentrifugation experiments. 10 ml light and heavy solution consisting of 1 ml 10x gradient buffer, 7 ml ddH2O and 2 or 6 ml 50%

sucrose, respectively. Approximately 3 ml of the light solution were filled into the centrifugation tube first using a syringe with a needle. Heavy solution was carefully filled below the light solution afterwards. The gradient was prepared using the Gardient Master (Biocomp) device. 50-150 µl of the sample were carefully pipetted on top of the mixed gradient. Ultracentrifugation was performed in a Beckmann SW60 rotor at 32,000 rpm and 4°C for 16 hours. The gradient was fractionated from the top into 200 µl fractions, which were analyzed by SDS-PAGE.

3.12 Ssl2 nuclear depletion and 4sU-sequencing[†]

The Ssl2 anchor away yeast strain was created from Y40343 (Haruki et al., 2008) by homologous recombination after PCR-amplifying a Ssl2-FRB-KanMX6 fragment with gene-specific primers. Clones were selected on G418 plates and confirmed by colony-PCR and sequencing. The Ssl2 anchor away strain shows a strong growth phenotype under depletion conditions (Supplemental Figure 1D). The anchor away was additionally validated by anti-FRB chromatin immunoprecipitation (ChIP) and qPCR with primers against specific gene promoter regions (Baejen et al., 2017). For 4sU-sequencing (4sU-seq), anchor away and metabolic 4sU-labeling, RNA extraction and library preparation were performed as described (Schulz et al., 2013).

3.13 Bioinformatics data analysis[†‡]

Data analysis was performed as described (Schulz et al., 2013), with modifications. Briefly, paired-end 50 bp reads with additional 6 bp of barcodes were obtained for labelled RNA. Reads were demultiplexed and aligned to the *S. cerevisiae* genome (sacCer3, version 64.2.1) using STAR (version 2.3.0) (Dobin et al., 2013). SAMTools was used to quality filter SAM files (Li et al., 2009). Alignments with MAPQ smaller than 7 (-q 7) were skipped and only proper pairs (-f2) were selected. Further processing of the 4sU-seq data was carried out using R/Bioconductor. We used a spike-in (RNAs) normalization strategy essentially as described (Schwalb et al., 2016) to allow observation of global shifts and

[‡] Bioinformatics were performed by Björn Schwalb

antisense bias determination (ratio of spurious reads originating from the opposite strand introduced by the RT reactions). Read counts for all features were calculated using HTSeq (Anders et al., 2015) and corrected for antisense bias using antisense bias ratios calculated as described (Schwalb et al., 2016). Gene expression fold changes upon rapamycin treatment were calculated using DESeq2 (Love et al., 2014). Differentially expressed genes were identified using a fold change of at least 1.5 and an adjusted P-value of maximal 0.1. Gene-wise DNA duplex free energies were calculated over a window of 8 nucleotides based on nearest-neighbour thermodynamics (SantaLucia, 1998).

3.14 DNA opening assay†

S. cerevisiae Pol II, TFIIB, TFIIF, TFIIE were purified as described (Plaschka et al., 2016; Schilbach et al., 2017). 2-amino purine (2-AP)-labelled promoter DNAs were synthesized (IDT) and reannealed on a slow temperature gradient from 95-20°C. *HIS4* scaffolds were based on the native sequence with A+27 replaced by 2-AP (non-template: 5'- GCA CGC TGT GTA TAT AAT AGC TAT GGA ACG TTC GAT TC**A*** CCT CCG ATG TGT GTT GTA CAT ACA TAA AAA TAT CA-3' and template: 5'- TGA TAT TTT TAT GTA TGT ACA ACA CAC ATC GGA GGT GAA TCG AAC GTT CCA TAG CTA TTA TAT ACA CAG CGT GC-3'). Base pair +27 of the native *GAT1* promoter sequence was mutated from C:G to 2-AP:T for label incorporation (non-template: 5'- CCC AGC CAC ATA TAT ATA GGT GTG TGC CAC TCC CGG CC**A*** CGG TAT TAG CAT GCA CGT TTT CTT TCC TTT GCT TTT-3' and template: 5'- AAA AGC AAA GGA AAG AAA ACG TGC ATG CTA ATA CCG TGG CCG GGA GTG GCA CAC ACC TAT ATA TAT GTG GCT GGG-3'). 25 pmol of each component was mixed in 25 uL reactions in 10 mM HEPES pH 7, 65 mM KCl, 5% glycerol, 1 mM DTT. Reactions were incubated for 90 minutes or 22 hours at 25°C and 2-AP fluorescence was measured with $\lambda_{ex}/\lambda_{em}$ of 307/370 nm. Normalization of background fluorescence from protein and nucleic acids was done with unlabelled samples. Measurements were done in triplicates.

3.15 Preparation of cPICs†

Proteins were purified as above, and TFIIA and core Mediator (cMed) were purified as described (Plaschka et al., 2016; Schilbach et al., 2017). Promoter DNA was synthesized

according to the native *GAT1* sequence (non-template: 5'- GCG GTG CCC GGC CCA GCC ACA TAT ATA TAG GTG TGT GCC ACT CCC GGC CCC GGT ATT AGC ATG CAC GTT TTC TTT CCT TTG CTT T-3' and template: 5'-AAA GCA AAG GAA AGA AAA CGT GCA TGC TAA TAC CGG GGC CGG GAG TGG CAC ACA CCT ATA TAT ATG TGG CTG GGC CGG GCA CCG C-3'). Pol II (0.32 nmol) and TFIIF (1.6 nmol) were added to a premixed complex of *GAT1* promoter DNA (0.8 nmol), TBP (1.6 nmol) and TFIIA (3.2 nmol). TFIIE (3.2 nmol) and cMed (0.38 nmol) were added and the complex was incubated for 90 minutes at 25°C in 10 mM HEPES pH 7, 200 mM KCl, 5% glycerol, 1 mM DTT.

Complex purification was carried out by sucrose-gradient centrifugation and crosslinking (Kastner et al., 2008). The gradient was prepared from a 10% sucrose light solution and a 30% sucrose heavy solution containing 0.075% glutaraldehyde. Ultracentrifugation was done for 16 h at 4°C and 175,000 g. Fractions containing the core PIC-cMed were collected and remaining cross linker was quenched with 15 mM lysine, pH 8.0. Samples were then dialyzed for 5-7 h against 10 mM HEPES pH 7.5, 150 mM KOAc, 2 mM DTT until glycerol and sucrose were removed as determined by refractive index measurements. Dialysed samples were then concentrated to 0.4 mg/ml in a Vivaspin 500 MWCO 100,000 centrifugal filter and directly used for cryo-EM grid preparation.

3.16 Cryo-EM and image processing of cCC1 and cCC^{dist†}

For cryo-EM grid preparation, 3.5 µl sample were applied to glow-discharged R1.2/1.3 UltrAuFoilTM grids (Quantifoil, Germany), blotted and plunge-frozen in liquid ethane (Vitrobot (FEI, USA) at 95% humidity, 4°C, 8.5 s blotting time, blot force 13). Data collection was performed on a Titan Krios (FEI) using a K2 direct electron detector (Gatan) in EFTEM mode. Data was collected with defocus ranging from -0.8 to -3.0 µm and at a magnification of 130,000x yielding a calibrated pixel size of 1.07 Å/px. The total electron dose was 37 e^-/A^2 distributed over 33 movie frames. About 3,800 micrographs were selected manually and dose-weighted and motion-corrected using an in-house developed software based on the MotionCorr algorithms (Li et al., 2013). The CTF was estimated by CTFFIND4 (Rohou and Grigorieff, 2015). If not stated otherwise, all data processing steps were done in RELION (Scheres, 2012).

Initial particle picking was performed using 2D projections of a yeast closed cPIC (Plaschka et al., 2016) filtered to 20 Å and particles were extracted in 300x300 pixel boxes. False positively picked particles were removed manually and by several rounds of 2D classification yielding a clean 385,000 particles data set (Supplement Figure 3B). This dataset was refined in 3D with the yeast OC filtered to 40 Å as a reference. To improve 3D classification particle polishing was performed. From this reconstruction, a mask for the DNA was isolated and used for 3D classification yielding CC and OC complexes. CC particles were pooled and reclassified using a mask containing the Tfg2 WH, TFIIE and the Pol II stalk yielding cCC1 and cCC^{dist}. Final classes were refined in 3D and B-factor sharpened with RELION. cCC1 was refined to 5.1 Å resolution (0.143 FSC) with a B-factor ranging from -80 to -150 Å^2 and cCC^{dist} was refined to 4.8 Å resolution (0.143 FSC) applying B-factors of 100-130 Å^2. Locally filtered maps were generated by a combination of local FSC-based filtering and local B-factor sharpening.

3.17 Data processing and CC^{dist} reconstruction†

A set of 255k particles was extracted from raw data used to reconstruct a previously published PIC (Schilbach et al., 2017). Closed and open complex particles were separated using a mask on promoter DNA. CC particles were reclassified using a global mask and TFIIH-containing classes were subjected to a final round of 3D classification. The two best classes were merged and refined in 3D to 6.7 Å with a fixed B-factor of 100 Å^2. Locally filtered maps were calculated as for cCC1 and cCC^{dist}, revealing a well-resolved cPIC at ~4 Å and TFIIH at lower resolution.

3.18 Model building and refinement †

The protein models for cCC1 and cCC^{dist} were build based on the previous PIC structure (Schilbach et al., 2017) and real-space fitted as rigid body domains with Phenix (Adams et al., 2010). DNA was fitted manually to the map in COOT (Emsley and Cowtan, 2004). Iterative rounds of real space refinement and geometry optimization in Phenix (Adams et al., 2010) yielded a model for cCC1 and cCC^{dist} DNA. Because the local resolution of the DNA was ~5 Å, strong B-DNA geometry restraints were used during refinement of cCC1 DNA. For cCC^{dist} DNA, deviations of helical twist and base pair shift from the standard B-

DNA values were required to obtain a good fit. CC^{dist} was also built from the existing PIC model (Schilbach et al., 2017) and DNA from the cCC^{dist} structure. Promoter DNA was then extended by B-DNA pieces generated with 3D-DART (van Dijk and Bonvin, 2009) that contained map-matching bends in the DNA. Proteins were real-space refined as rigid body modules and DNA was refined using the same strategy as for cCC1 and cCC^{dist}.

4 RESULTS

4.1 Preparation of transcription factors and core pre-initiation complexes

To structurally and functionally investigate Pol II transcription initiation, core PICs needed to be formed including TFIIA, TFIIB, TBP, TFIIF, TFIIE and Pol II. At the time this work was started, full length TBP and TFIIA were not cloned and expression and purification protocols were not established. Purifications of the other factors were well established (Kostrewa et al., 2009; Lariviere et al., 2013; Lariviere et al., 2012; Muhlbacher et al., 2014; Plaschka et al., 2015; Sainsbury et al., 2013) and are not discussed here.

To obtain an expression construct of TBP and TFIIA, their sequence has to be amplified and inserted into an expression vector. This construct then can be tested for successful expression on small scale which allows subsequent scale up and purification attempts.

4.1.1 Cloning, expression and purification of full length TBP

The genomic sequence of full length TBP was PCR amplified from genomic *S. cerevisiae* DNA (kind gift from Kristin Leike) using gene specific primers (Figure 4A). The primers also contained a 12 nt homology region to the NcoI/KpnI-linearized pOPINE vector. In-fusion cloning yielded a large number of positive clones that were confirmed by PCR and plasmid sequencing giving rise to an expression plasmid containing C-terminally His-6 tagged full length TBP. Initial expression tests showed good TBP expression after 5 hours and overnight. However, more soluble protein could be obtained after 5 hours (Figure 4B). Because the expression was visible in whole lysate, no Nickel affinity pull down was done.

The purification was then attempted from a large scale expression using similar conditions as during the expression tests. A nickel column was done as first step to capture the His-6 tagged protein. Prior to elution of the protein, the column was also washed with high salt in order to remove bound nucleic acids. TBP eluted very late from the imidazole gradient making it very pure after the first purification step (Figure 4C). Ion-exchange chromatography on a MonoS column was then performed in order to remove residual amounts of nucleic acids. The elution profile clearly showed separate peaks for DNA and

protein (Figure 4D). Very clean TBP could be obtained from the protein containing peak. A final size-exclusion chromatography step was performed to polish the sample. Since TBP seemed to stick to the Superdex resin of other columns (not shown), gel filtration needed to be performed on a Superose 12 column. TBP eluted as a double peak from the size exclusion column corresponding to monomeric TBP and concentration dependent dimers (Figure 4E). Approximately 1 mg protein could be obtained per liter expression culture.

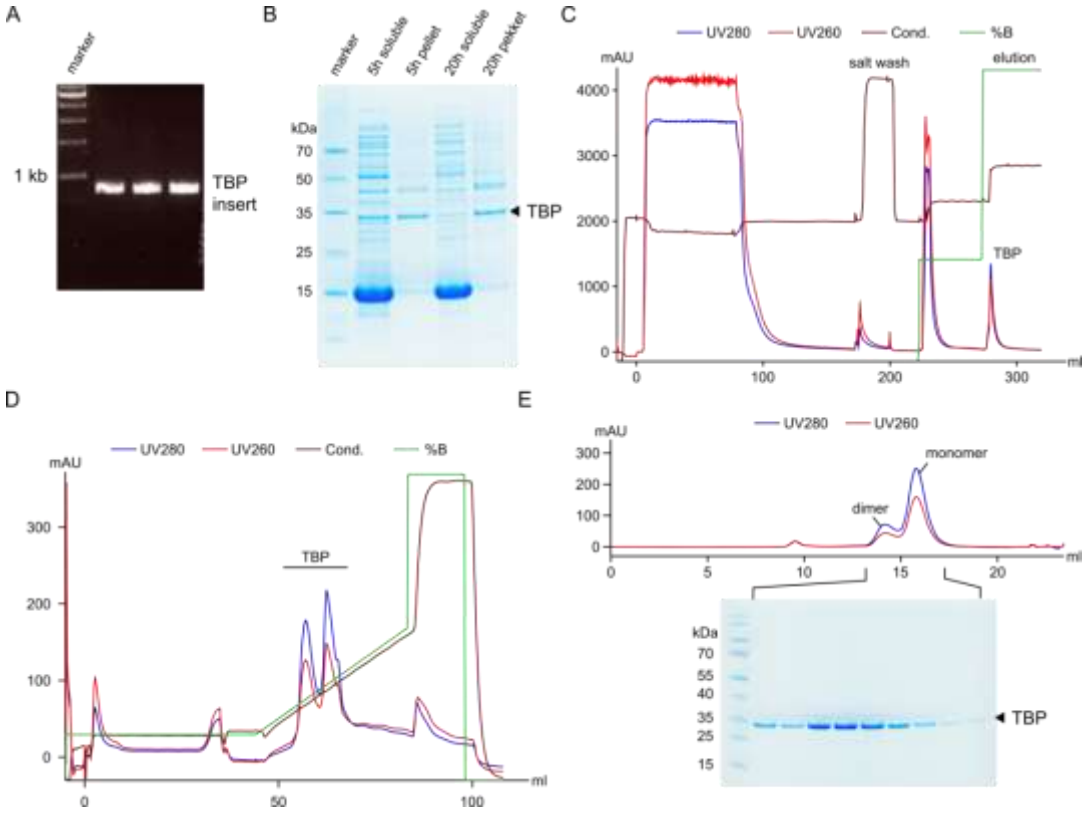


Figure 4: Cloning, expression and purification of yeast full length TBP

(**A**) TBP amplified from genomic DNA using gene specific primers. Expected fragment size is ~760 bp.

(**B**) Expression test for C-terminally tagged TBP. Soluble and pellet fractions are loaded next to one another. The protein band corresponding to TBP is marked.

(**C**) Chromatogram of a HisTrap purification of TBP. The peak corresponding to TBP is marked.

(**D**) Chromatogram showing a typical MonoS purification step of TBP. Both peaks contain TBP and were pooled for subsequent gel filtration.

(**E**) Size exclusion chromatography of TBP. Dimer and monomer peaks are indicated. The gel shows frations of the marked region. The protein band corresponding to TBP is indicated.

4.1.2 Cloning, expression and purification of TFIIAΔ113

Full length TFIIA was already present as a codon-optimized construct in a pETduet vector, however, with a frame shift mutation around the start codon of the Toa1 subunit making expression impossible. Therefore, the pETduet cassette containing Toa1 and Toa2 was re-amplified using specific primers with homology overhangs to the pOPINE vector. In-fusion cloning of that cassette into pOPINE yielded full length TFIIA with a C-terminal His6-tag on Toa2. Expression testing of the full length construct did not yield soluble protein detectable by NiNTA pulldown (not shown).

Therefore, a reduced TFIIA with a 113 amino acid deletion in Toa1 was generated. The region from amino acid 95 to 209 was predicted to be unstructured and its removal enabled large scale expression and purification before (Wang et al., 2001). The construct of TFIIAΔ113 was generated by mutagenesis PCR basically following the QuickChange approach. The resulting deletion mutant could be expressed well within 3h in initial tests as visible from NiNTA pulldowns (Figure 5A). A protocol similar to the expression test conditions was used for large scale expression.

To capture His-6 tagged TFIIAΔ113 from the lysate, a HisTrap column was used as initial purification step (Figure 5B). The sample was washed with high salt while bound to the column to remove nucleic acids that interact with the protein. During the elution using an imidazole gradient, TFIIAΔ113 eluted after unspecifically bound proteins and therefore could be obtained at good purity. In order to remove residual amounts of nucleic acids, TFIIAΔ113 was then bound to a MonoS column. The salt step used for elution clearly separated DNA from protein (Figure 5C), making TFIIA very pure. Final polishing of the protein on a size exclusion column yielded very pure and homogenous TFIIAΔ113 (Figure 5D and E) at a yield of 0.5-1 mg per liter expression culture.

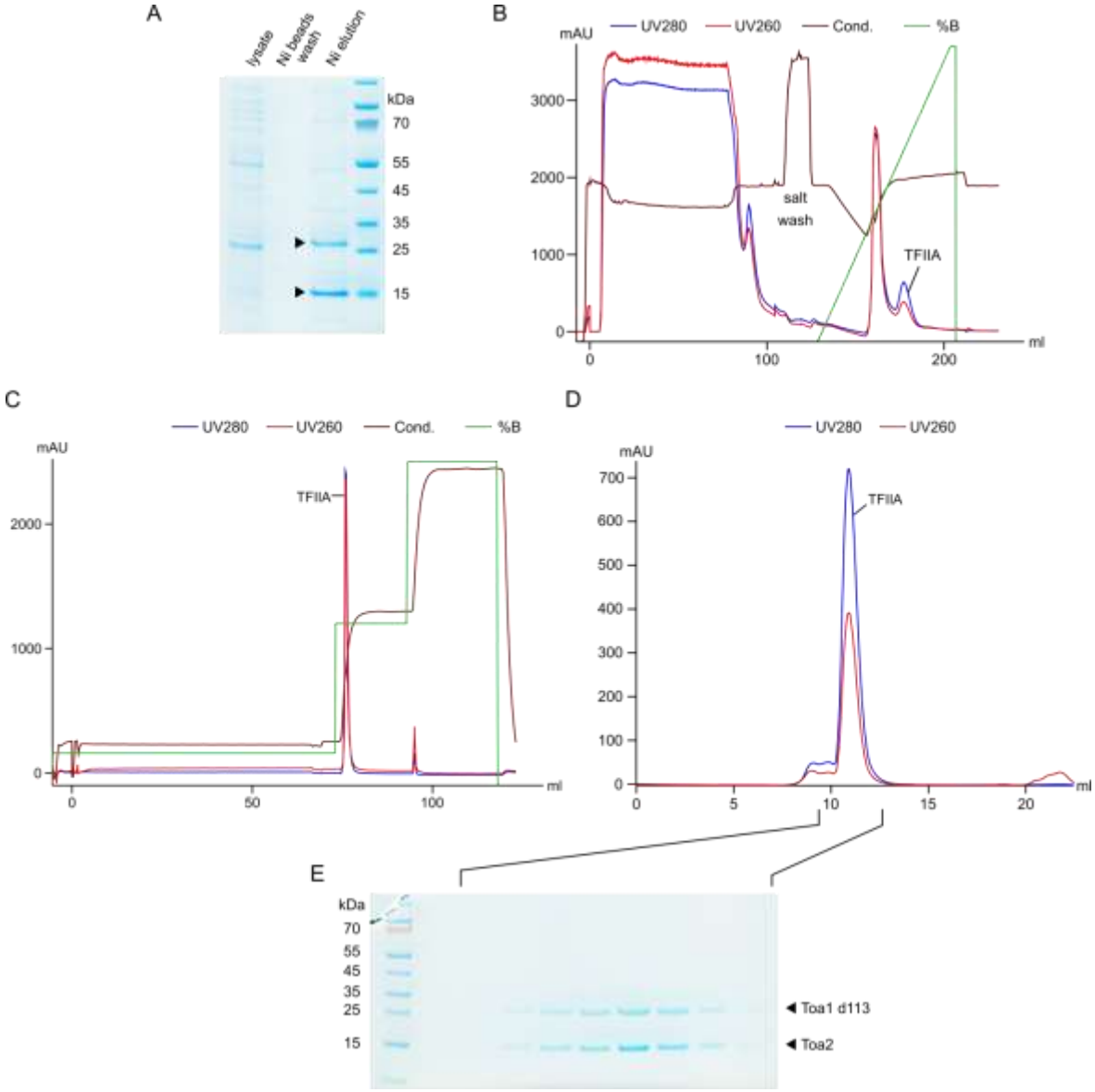


Figure 5: Expression and Purification of TFIIAΔ113

(**A**) Expression test of TFIIAΔ113. Toa1 and Toa2 band are marked. After loading the lysate to the Ni beads, the beads were washed with 20 mM imidazole. Elution was done with 250 mM imidazole.

(**B**) HisTrap purification of TFIIAΔ113. Peak containing TFIIA Δ113 is marked.

(**C**) MonoS purification step of TFIIAΔ113. TFIIA is eluted by a step increase of buffer A1000.

(**D**) Gel filtration chromatogram of TFIIAΔ113. TFIIA elutes as a single peak.

(**E**) Gel showing TFIIAΔ113 in the peak fractions of the size exclusion purification.

4.1.3 Preparation of cPICs by gradient ultracentrifugation

After obtaining pure full length TBP and TFIIAΔ113 in large amounts from recombinant expression, formation of large quantities of cPICs containing these proteins was attempted. Previous protocols used size exclusion chromatography. However, the resulting complexes were limited yields due to unspecific binding of proteins to the column resin and concentrator membranes. Therefore, gradient ultracentrifugation was used to first assemble the cPIC.

Table 16: Components used for cPIC and cPIC-cMed reconstitutions. *molar excess relative to Pol II. #cMed was not added for cPIC preparations.

Component	Mass (kDa)	nmol	Molar excess*
DNA	42	0.4	2.5
TBP	32	0.8	5
TFIIA	34	1.6	10
TFIIB	38	0.8	5
Pol II	516	0.16	1
TFIIF	128	0.8	5
TFIIE	91	3.2	10
cMed#	515	0.37	1.2

PIC components were mixed in similar molar ratios as described before (Plaschka et al., 2016)(Table 16) and applied to the 10-30% sucrose gradient. Pol II only and cPIC without Mediator served as controls. After ultracentrifugation, bands for Pol II clearly shifted to larger fraction in the cPIC sample (Figure 6A and B), indicating successful complex formation. This shift was even more pronounced in samples containing Mediator (Figure 6C), demonstrating that a cPIC including Mediator can be formed on a sucrose gradient. The same assembly protocol was then used later to prepare samples for structural studies by cryo-electron microscopy.

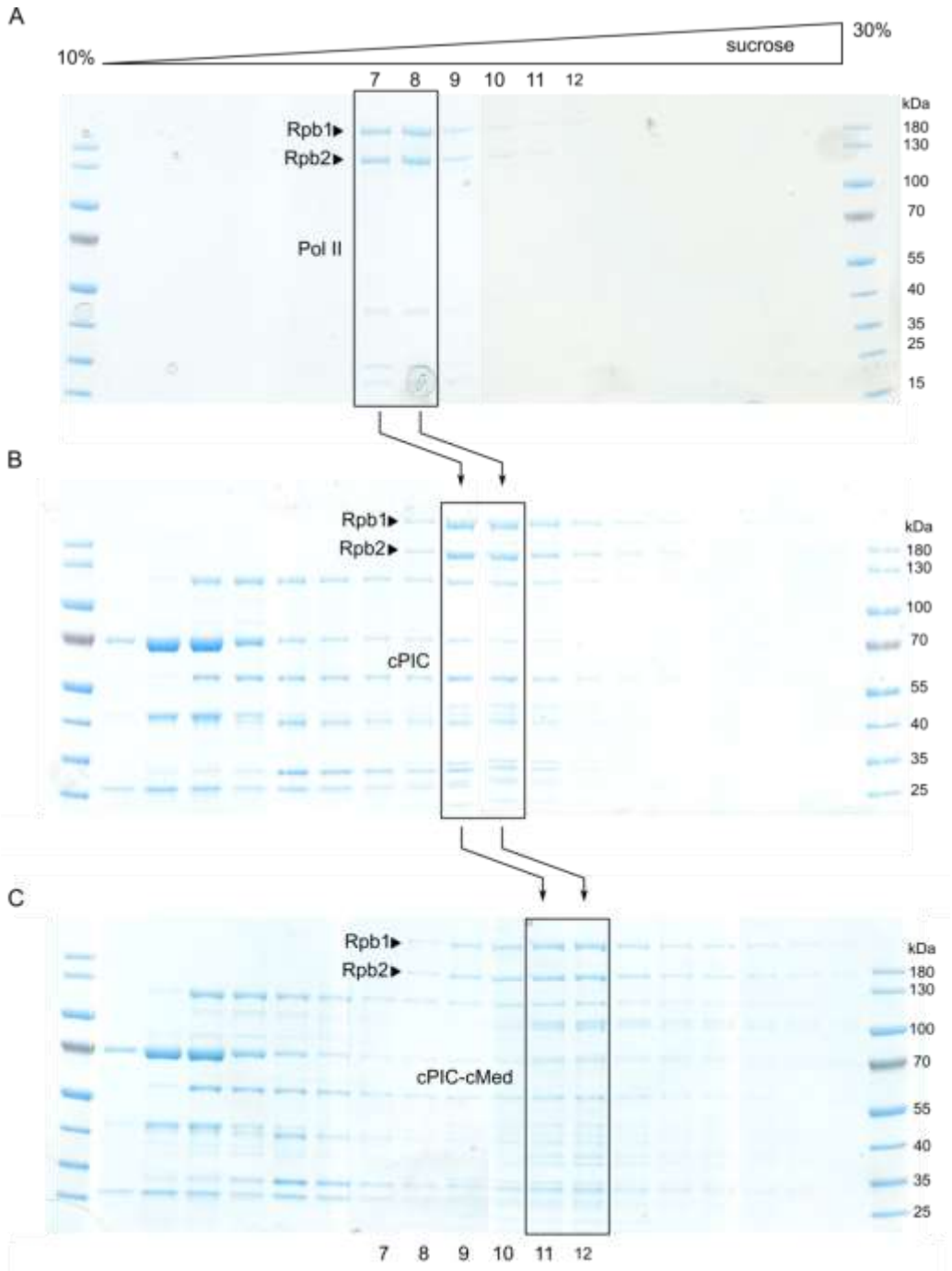


Figure 6: Formation of cPIC and cPIC-cMed by gradient ultracentrifugation.

(A) Fractions of Pol II alone run on a 10-30% sucrose gradient. Pol II localizes in fractions 7 and 8. Pol II subunits Rpb1 and Rpb2 are marked.

(B) After incubation with cPIC componends, Pol II (indicated by the bynd for Rpb1 and Rpb2) is shifted to fractions 9 and 10.

(C) Binding of cMed to the cPIC shifts the bands for Rpb1 and Rpb2 to fraction 11 and 12 indication the formation of a higher molecular weight cPIC-cMed complex.

4.2 Structural and functional investigation of promoter opening

The results described in this section were produced in collaboration with co-authors of a manuscript that has been submitted:

> Dienemann C, Schwalb B, Schilbach S, Cramer P. Promoter distortion and opening in the RNA polymerase II cleft. *Molecular Cell*, Online 21 November 2018

Contributions of coauthors other than the author of this thesis are indicated. A detailed list of text and figures taken from the manuscript is given in section Publications.

4.2.1 TFIIH translocase is not generally required for transcription *in vivo*

Our previous results obtained with the yeast system *in vitro* (Plaschka et al., 2016) suggested that the TFIIH translocase subunit Ssl2 (human XPB) is not strictly required to open promoter DNA and to initiate transcription, but this was not tested *in vivo*. To investigate the genome-wide requirement of Ssl2 for transcription initiation, we transiently depleted Ssl2 from the yeast nucleus using the anchor-away technique (Haruki et al., 2008). We verified by chromatin immunoprecipitation that Ssl2 occupancy at genes is almost entirely lost under depletion conditions (Supplemental Figure 1A). We then monitored newly synthesized RNA by 4-thiouracil sequencing (4sU-seq) (Schulz et al., 2013) and analyzed spike-in normalized sequencing reads for differential expression (see 3.13 Bioinformatics data analysis[†]) [§].

We obtained groups of responsive and non-responsive genes (Figure 7A). The majority of genes responded, demonstrating the importance of Ssl2 for transcription *in vivo*. This fraction of genes was not enriched in any Gene Ontology terms (Supplemental Figure 1B). However, RNA synthesis from 18% of genes was not significantly altered upon Ssl2 depletion, indicating that a fraction of promoters can be opened in the absence of Ssl2. These results indicate that spontaneous, Ssl2-independent DNA opening occurs for a fraction of yeast promoters *in vivo*.

[§] Bioinformatics were performed by Björn Schwalb

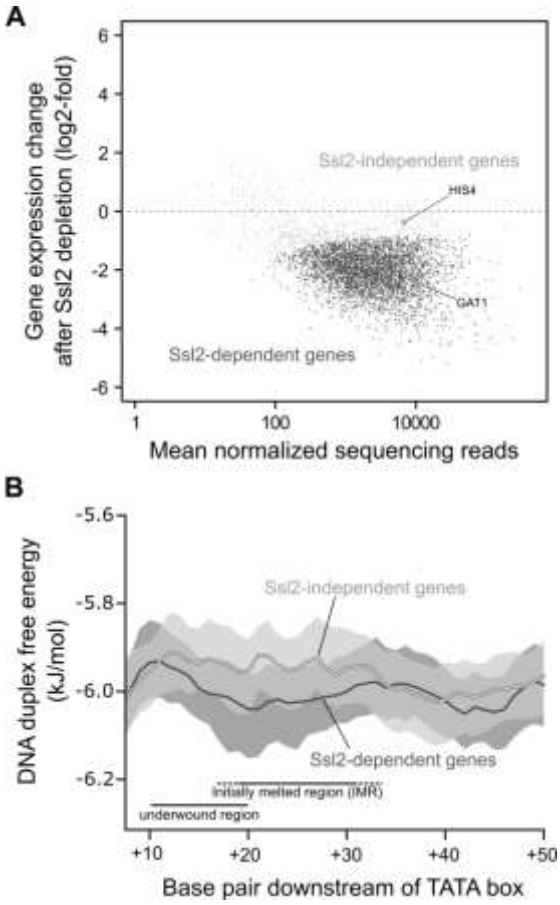


Figure 7: Ssl2-independent transcription *in vivo*

(**A**) Transcription of a subset of ~18% of protein-coding genes are not significantly affected after nuclear depletion of the TFIIH translocase subunit Ssl2. Points mark the log2-fold change of newly synthesized RNA upon Ssl2 nuclear depletion *versus* the normalized mean read count across replicates (Anders and Huber, 2010). Genes that respond significantly to Ssl2 depletion are shown in dark grey, non-responding genes without significantly altered transcription in light grey.

(**B**) DNA duplex free energy of responding and non-responding genes from (A) differs in the IMR of their promoters. The DNA duplex free energy (SantaLucia, 1998) of the promoter region is shown for responsive (dark grey) and non-responsive (light grey) TATA-containing genes. Sequences were aligned at the TATA box and base coordinates are given relative to the beginning of the TATA box. Confidence intervals are shown as areas around the traces.

4.2.2 Translocase requirement correlates with promoter stability

Because some genes can be transcribed after depletion of the translocase Ssl2 *in vivo*, the meltability of their promoters is apparently greater. To find possible determinants of promoter meltability, we analysed IMRs in promoters of two groups of genes that were either responding to Ssl2 depletion or not[**]. Neither of the two gene groups showed a bias in base pair composition (Supplemental Figure 1C). However, when calculating the free energy of the DNA duplex by the nearest-neighbour method (SantaLucia, 1998), we found that genes that were not responding to Ssl2 depletion contain less stable DNA duplexes between positions +15 and +30 downstream of the TATA box (Figure 7B). These results indicate that lower DNA duplex stability within the IMR leads to a greater promoter meltability and can circumvent a requirement for Ssl2 for transcription *in vivo*.

[**] Bioinformatics were performed by Björn Schwalb

Genes that did not respond to Ssl2 depletion included *HIS4* (Figure 7A), suggesting that the *HIS4* promoter can melt easily. This explained why in our previous structural studies of PICs with *HIS4* promoter DNA we obtained OCs although closed DNA was used and TFIIH and ATP were absent (Plaschka et al., 2016). In search for a promoter that would melt less easily, we picked from our *in vivo* data the *GAT1* promoter. *GAT1* belongs to the genes that strongly respond to Ssl2 depletion (Figure 7A) and its IMR is predicted to be much more stable that the IMR in the *HIS4* promoter (Figure 8A).

4.2.3 Promoter stability defines DNA meltability

We then tested the meltability of the *HIS4* and *GAT1* promoters in a reconstituted, TFIIH-free *in vitro* system (Figure 8). To monitor DNA opening, we incorporated 2-aminopurine (2-AP) in the IMR at the base pair located +28 positions downstream of TATA and measured the increase in fluorescence upon single strand formation (Kashkina et al., 2007). In the presence of only DNA, TBP and TFIIB, the fluorescence signal did not change after 90 minutes of incubation, providing a negative control (Figure 8B). When Pol II was added, DNA opening was detected for the *HIS4* promoter, but not for the *GAT1* promoter (Figure 8B), consistent with our *in vivo* data. The *GAT1* promoter showed DNA opening only after 22 hours of incubation (Supplemental Figure 2). Spontaneous opening required TBP and TFIIB, but not TFIIE and TFIIF, resembling the archaeal transcription system, which only requires counterparts of TBP and TFIIB for initiation (Spitalny and Thomm, 2003).

To confirm that the observed differences in promoter meltability stem from differences in the sequence of their IMRs, we mutated the IMR in the *GAT1* promoter. We changed positions +21, +24, +25 and +26 downstream of TATA, resulting in a mutant *GAT1* promoter (GAT1mut, Figure 8A). Indeed, GAT1mut DNA opened within 90 minutes, like the *HIS4* promoter (Figure 8B). In agreement with this, the DNA duplex free energy of the GAT1mut promoter IMR is very similar to that of the *HIS4* promoter (Figure 8A). Thus the sequence and the DNA duplex stability of the IMR influences spontaneous promoter opening *in vitro*. This is in agreement with our finding that an unstable IMR correlates with Ssl2 dependence *in vivo*. Thus the stability of the IMR defines promoter meltability in this system.

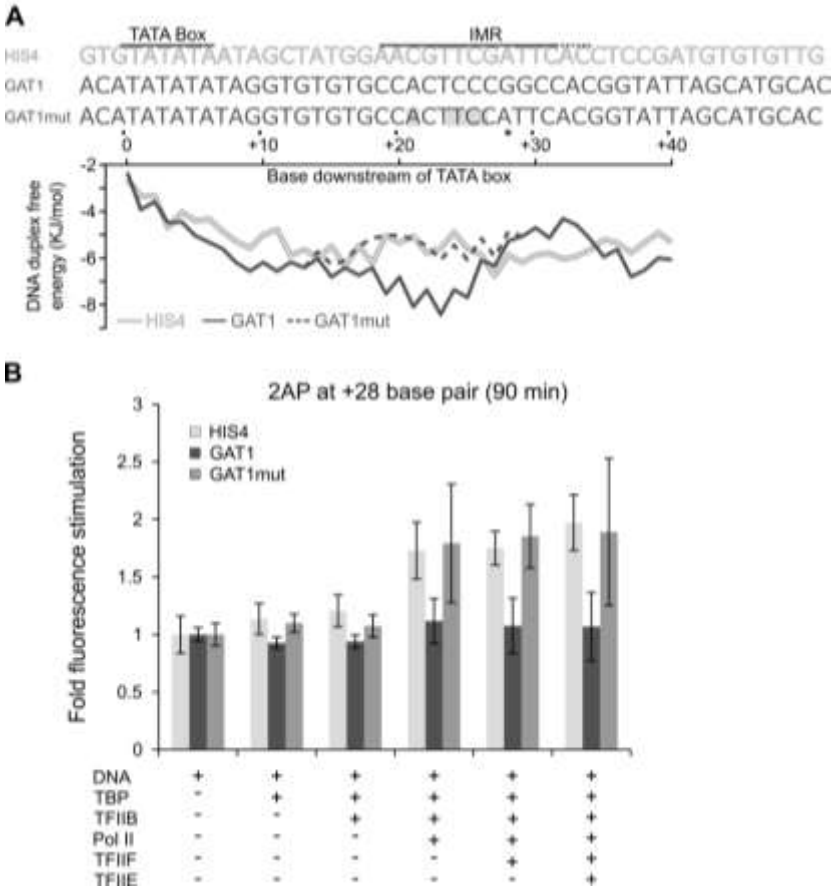


Figure 8: Promoter opening depends on the initially melted region

(**A**) Promoter DNA scaffolds used in DNA opening assays. The sequence of the non-template strand is shown aligned at the TATA box. The DNA-duplex free energy (SantaLucia, 1998) of each scaffold is shown below. The position of the 2-AP label is marked with an asterisk. Bases of the *GAT1* promoter that are mutated in *GAT1mut* are highlighted.

(**B**) The *HIS4* promoter but not the *GAT1* promoter melts spontaneously *in vitro* under conditions of our assay. Spontaneous melting of the *GAT1* promoter is achieved after its sequence was changed in the IMR to resemble the *HIS4* promoter (*GAT1mut*). Bar plots show the fluorescence increase normalized to the DNA only reaction.

4.2.4 Cryo-EM reveals distinct closed complex intermediates

Based on these results we used *GAT1* promoter DNA to form a more stable yeast CC *in vitro*. We prepared core PICs with GAT1 promoter DNA, TBP, TFIIB, TFIIE, TFIIF and Pol II for cryo-EM analysis and collected ~3800 micrograph images on a Titan Krios with a Gatan K2 direct electron detector in EFTEM mode (Supplemental Figure 3A, Supplemental Table 1). The initially picked set of particles was cleaned by 2D classification (Supplemental Figure 3B).

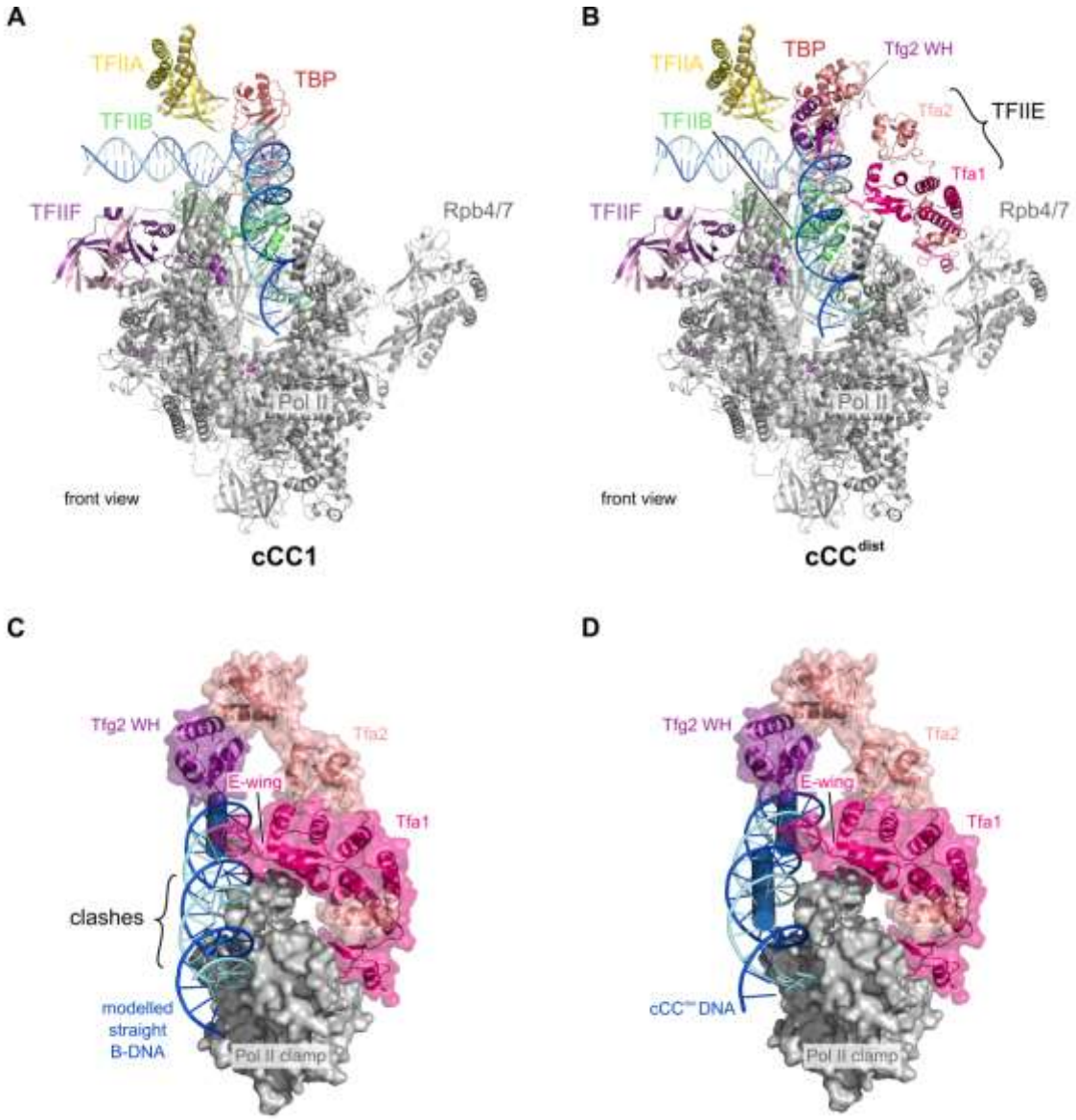


Figure 9: Cryo-EM structures of core closed complexes (cCC)

(**A**) Structure of Pol II core closed complex cCC1. Proteins are shown in cartoon representation and transcription factors are coloured according to their subunits. Pol II is shown in grey, the non-template strand in light cyan and the template strand in blue.

(**B**) Structure of Pol II core closed complex cCC^{dist}. Colouring as in panel A.

(**C**) The closed clamp in cCC^{dist} would clash with undistorted, straight B-DNA. The Pol II clamp and TFIIE are shown as transparent surface with cartoon inside. Other protein components were omitted for clarity. Straight B-DNA was extended from the yeast cOC structure (PDB 5fyw). The DNA helical axis of the OC DNA used as start for B-DNA modelling is shown as blue cylinder.

(**D**) The Tfg2 WH and TFIIE lock the DNA at the site of DNA distortion. The Pol II clamp, TFIIE and Tfg2 WH are shown in transparent surface representation with cartoon inside. Other proteins have been omitted

The remaining 380,000 particle images were classified in 3D applying a mask around the IMR of promoter DNA. This led to 24% CC and 76% OC particles (Supplemental Figure 3C). Thus, spontaneous DNA opening still occurred but was incomplete due to the use of the *GAT1* DNA that is more difficult to melt.

As a result, we obtained enough yeast CC particles for structure determination at higher resolution than previously possible (Plaschka et al., 2016). Classification of particles revealed two different structures of the core CC (cCC) (Supplemental Figure 3C, Supplemental Figure 3D). Reconstructions were obtained at nominal resolutions of 5.1 Å and 4.8 Å (Supplemental Figure 3D). Structures were built based on our previous PIC structures (Plaschka et al., 2016; Schilbach et al., 2017) and refined in real space to yield good stereochemistry (Methods).

The two cCC structures differed in the position of closed promoter DNA (Figure 9). The first structure (cCC1) contained canonical B-DNA along the upper Pol II cleft (Figure 9A). This structure resembles the previously reported yeast cCC (Plaschka et al., 2016) and lacks TFIIE and the Tfg2 WH domain (Supplemental Figure 3D), which also had very weak density in the previous structure. In the second structure, we observed a region of promoter DNA that deviated substantially from B-DNA (Figure 9B, Supplemental Figure 3E), and we therefore called it distorted cCC (cCC^{dist}). This structure revealed all components of the cPIC including TFIIE and the Tfg2 WH domain (Figure 9B). In both structures, the clamp is closed and the TFIIB B-linker is ordered (Supplemental Figure 4). Thus a single CC preparation gave rise to two different CC structures and the OC, indicating that the CC is dynamic and its structures correspond to intermediates on the path to the OC.

4.2.5 CC^{dist} contains distorted DNA in a closed cleft

The cCC^{dist} structure represents a previously unobserved initiation intermediate. In this structure, promoter DNA is loaded ~7 Å deeper into the cleft compared to cCC1 although the clamp and cleft are closed (Figure 10, Supplemental Table 2). We observe an offset of the DNA helical axis by ~5 Å around base pair +20 downstream of the TATA box (Figure 10B), which corresponds to the position of the upstream edge of the IMR (Giardina and Lis, 1993). The DNA distortion is required when the clamp is closed because modelling

straight B-form DNA results in a clash with the clamp (Figure 10C). Consistent with this, a human CC structure with straight, undistorted DNA showed an open clamp and cleft (He et al., 2013; He et al., 2016) (Supplemental Table 2).

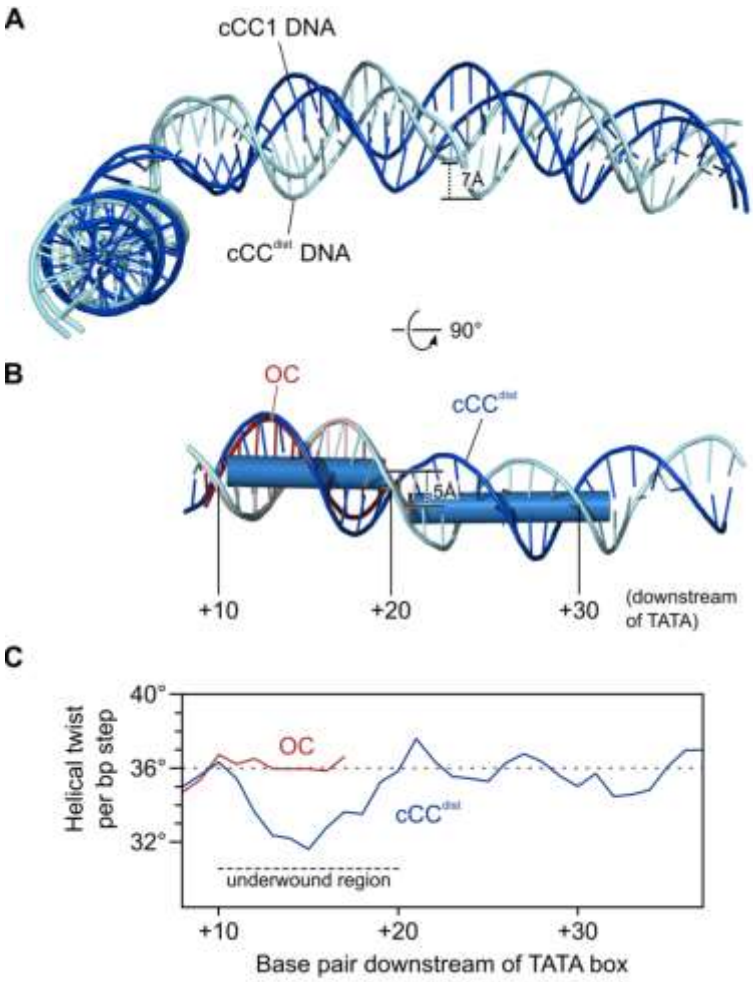


Figure 10: Distortion of promoter DNA

(**A**) Promoter DNA is loaded 7 Å deeper into the Pol II cleft in the cCC^{dist} structure compared with cCC1. The template and non-template strands are in blue and cyan, respectively.

(**B**) Distortion of promoter DNA in cCC^{dist}. The helical DNA axis was calculated by w3DNA (Zheng et al., 2009) and is indicated by blue cylinders. OC DNA (PDB 5fyw) is shown in red. OC and cCC^{dist} were superimposed with their TATA boxes. The sequence register relative to the upstream end of the TATA box is indicated for DNA in cCC^{dist}.

(**C**) The DNA region around the upstream edge of the IMR is underwound in cCC^{dist}. The DNA helical twist per base pair step as calculated with w3DNA (Zheng et al., 2009) is shown for OC (red) and cCC^{dist} (cyan/blue) structures. The average helical twist for canonical B-DNA (36°) is indicated with a dotted grey line. Values were calculated as average on a

The DNA distortion also includes an underwinding of base pairs between positions +10 and +20 downstream of TATA (Figure 10C). DNA underwinding in this region is released in an OC (Figure 10B and C), suggesting that DNA is strained in cCC^{dist} and that release of this strain facilitates DNA opening and the transition to the OC. The distorted, underwound DNA region interacts with the Tfg2 WH domain and the TFIIE extended winged helix (eWH) domain (Figure 9D), which may stabilize distorted DNA in the cleft. A hairpin

protruding from the TFIIE eWH domain, the E-wing, binds above the distorted DNA ~23 base pairs downstream of the TATA box.

4.2.6 Complete PIC structure with distorted DNA

To investigate whether the observed DNA distortions also occur in the presence of TFIIH, we further solved the structure of the CC in the presence of TFIIH. We re-classified particles from a large data set of yeast PICs comprising TFIIH and *HIS4* promoter DNA (Schilbach et al., 2017). Although this dataset contains mainly OCs, we could obtain 142,000 PIC particles with closed DNA (Supplemental Figure 5A). A complete PIC structure with closed DNA could be refined from a subset of 60,000 particles to a nominal resolution of 6.7 Å (Figure 11, Supplemental Figure 5B, Supplemental Table 1). In this structure, the distorted DNA region is observed at a local resolution of ~5 Å (Supplemental Figure 5C) and this was sufficient to define its conformation.

The structure revealed the same DNA distortion that we observed in cCC^{dist}, and we therefore refer to it as CC^{dist} (Figure 11). Despite the different promoter sequence, the DNA is again underwound in the region spanning from +10 to +20 downstream of TATA, and we again observe an offset in the helical axis around +20. Because these DNA distortions were not altered in the presence of TFIIH, they are apparently induced and stabilized by components of the cPIC, in particular the Pol II clamp, TFIIE and TFIIF (Figure 9D). Closed promoter DNA has been observed in a similar location within the yeast PIC before (Murakami et al., 2015), but at the available resolution the DNA distortion was not detected at that time.

In CC^{dist} we observe an additional ~20° DNA bend around position +33 downstream of TATA (Figure 11C) that is apparently induced by the presence of TFIIH. The region where this additional bend occurs corresponds to the nearest downstream site where transcription initiation may occur, because 30-35 nucleotides of DNA are required to reach from the upstream end of TATA to the active site of Pol II in an OC (Kostrewa et al., 2009). The DNA bend at +33 from TATA and the DNA axis offset at +20 from TATA delimit a DNA region of ~13 base pairs that corresponds to the IMR in the OC (Figure 11C). It is known that DNA underwinding (Kannan et al., 2006) and bending (Ramstein and Lavery, 1988)

weaken the DNA duplex. We thus conclude that the IMR is structurally pre-defined by two flanking DNA distortions in the complete PIC CC^{dist} intermediate.

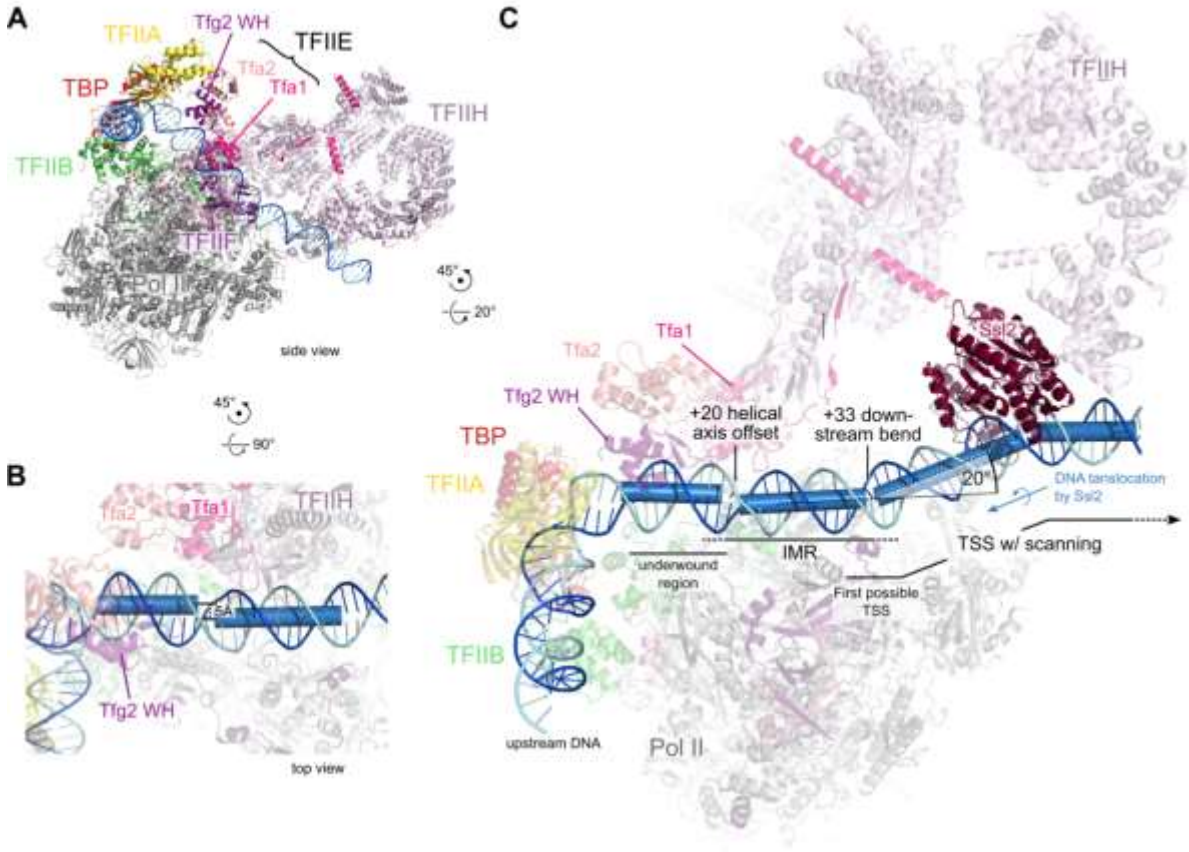


Figure 11: PIC structure with closed, distorted DNA

(**A**) Structure of Pol II PIC including TFIIH in the CC^{dist} state. The structure is shown in side view and coloured according to Figure 3A with TFIIH in pale pink.

(**B**) DNA helical axis offset. Close-up view of CC^{dist} promoter DNA with the DNA helical axis indicated as blue cylinders. Proteins around the DNA are shown in transparent cartoon.

(**C**) The presence of TFIIH induces a 20° bend at the downstream edge of the IMR.

4.2.7 Promoters of Pol I and Pol III contain unstable IMRs

The lack of Ssl2 dependence at a subset of Pol II promoters resembles the situation during transcription initiation by Pol I and Pol III, which do not require a translocase. Whereas Pol I uses unique initiation factors, the Pol III initiation machinery is closely related to that of Pol II (Vannini and Cramer, 2012). To investigate whether Pol III genes may also show

enhanced meltability in their IMR, we calculated the average DNA duplex free energy (SantaLucia, 1998) of all yeast tRNA genes aligned at their TSS. Strikingly, the IMR upstream of the TSS shows a much weaker DNA duplex than the region downstream of the TSS (Figure 12). This indicates that Pol III promoters have a less stable IMR, explaining why they are prone to open easily.

Pol I promoters were also suggested to be prone to DNA melting in their IMR (Engel et al., 2017). We therefore calculated the average DNA duplex free energy of Pol I promoters from eight different species (Moss et al., 2007) aligned at their TSS. Stability of these promoters was significantly lower in the region upstream and around the TSS (Figure 12). Thus, IMRs in Pol I promoters also show lower stability and thus greater meltability, similar to Pol III promoters and Ssl2-independent Pol II promoters. These findings suggest that the strategy for TFIIH-independent DNA opening by Pol II shows similarities to the DNA opening mechanisms employed by the Pol I and Pol III machineries.

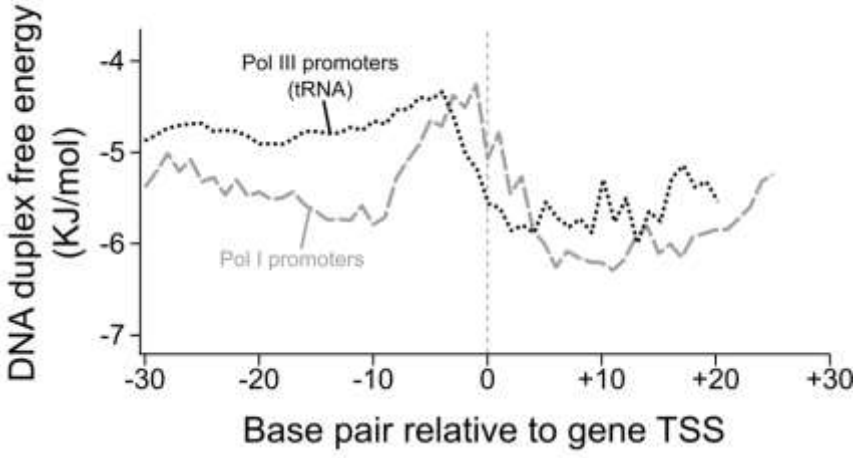


Figure 12: IMRs of Pol I and Pol III promoters are unstable

DNA duplex free energy calculated based on nearest-neighbour thermodynamics (SantaLucia, 1998) for Pol I (dashed) and Pol III (dotted) promoters. Sequences were aligned at the gene TSS. Free energy was calculated for the first base pair of the central dinucleotide of an eight base pairs running window for more stable estimates.

5 DISCUSSION

5.1 A new transcription initiation intermediate

Here we describe the structures of Pol II transcription initiation intermediates CC1 and CC^{dist} that both contain closed promoter DNA, but in different conformations. Whereas CC1 resembles known structures of the yeast CC (Plaschka et al., 2016), CC^{dist} represents a new state during transcription initiation by Pol II and was observed both in the absence and in the presence of TFIIH. CC^{dist} contains distorted promoter DNA in a closed Pol II cleft and apparently reflects the state of the PIC just before DNA opening and conversion to the OC. In contrast, the previously reported human CC structures (He et al., 2013; He et al., 2016) show an open clamp and contain canonical B-form DNA, indicating that they represent a CC state that is readily interchangeable with CC1, and therefore it may be referred to as CC2.

Together with the obtained functional data, these structures suggest a general model for promoter opening and transcription initiation (Figure 13). In the PIC, promoter DNA is first positioned above the Pol II cleft (CC1). Cleft opening then allows for DNA swinging into the cleft (CC2). Clamp closure is coupled to DNA distortion and underwinding in the cleft (CC^{dist}). In the CC^{dist} intermediate, the IMR is pre-defined, and DNA opening may now occur, either in a TFIIH-dependent or in a TFIIH-independent manner. When the IMR is stable, which is usually the case for Pol II promoters, TFIIH is required and uses its ATP-dependent translocase activity in subunit Ssl2 (human XPB) to further underwind and thereby open DNA, leading to the OC. When the IMR is unstable, DNA distortion and underwinding alone are sufficient to promote DNA opening, explaining why some genes are transcribed without TFIIH translocase activity. In both cases, the DNA template strand gets loaded into the cleft and DNA interacts more extensively with Pol II, and this stabilizes the OC and prevents re-annealing of the DNA strands.

5.2 DNA distortion and initial opening

Our results suggest that promoter opening may spontaneously nucleate at the upstream edge of the IMR in the underwound DNA region in CC^{dist}. Supporting this, TFIIH activity

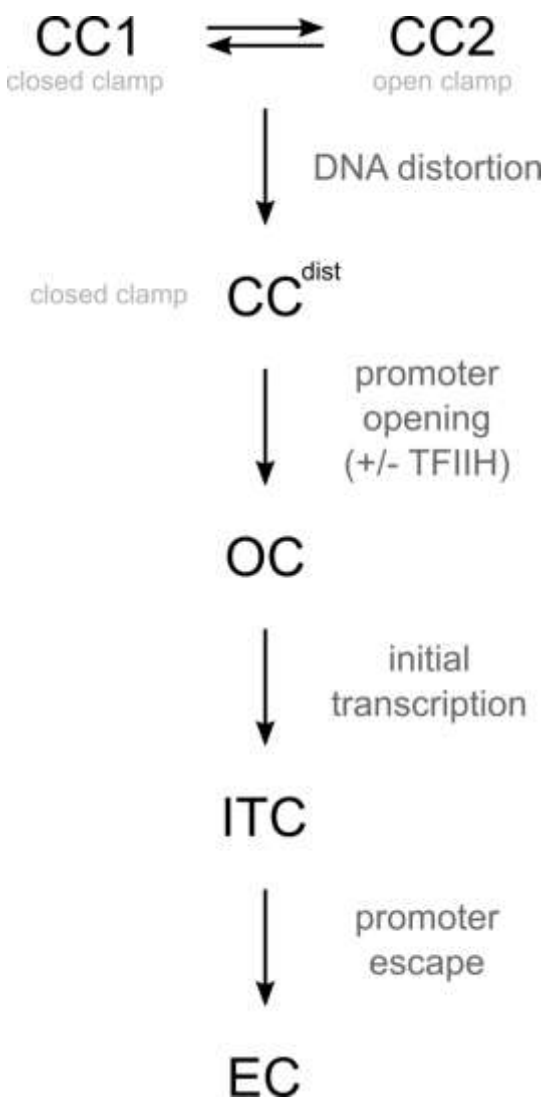


Figure 13: General model of transcription initiation

Proposed interconversion of initiation complex intermediates during DNA opening and the initiation-elongation transition. Compare text for details.

is required only for opening of the downstream part of the IMR (Holstege et al., 1996). The CC^{dist} structure also indicates why DNA opening propagates downstream, but not upstream, from the initial point of opening. Interactions of the upstream DNA duplex with the Tfg2 WH and TFIIE eWH domains are formed in CC^{dist} and may prevent propagation of the initial DNA bubble upstream towards the TATA box. However, the strain introduced by TFIIH translocase action can be released when the bubble is extended downstream towards the TSS.

The energetic cost for distorting DNA in CC^{dist} is apparently compensated by a gain of binding energy, otherwise we could not have trapped CC^{dist} *in vitro*. Conversion of CC1 to CC^{dist} may yield additional binding energy because the TFIIE eWH domain and the Tfg2 WH domain form contacts above the cleft. In addition, the B-linker element in TFIIB likely stabilizes CC^{dist}, because we observe good density for it (Figure S4) that suggests that it helps to keep the clamp in a closed state. The B-linker is known to be involved in promoter opening in the homologous archaeal system and its mutation induces a growth phenotype in yeast (Kostrewa et al., 2009). In contrast, CC2 (He et al., 2013; He et al., 2016) shows a mobile B-linker and undistorted DNA that would clash with a modelled B-linker. Thus, a transition from CC2 to CC^{dist} involves clamp closure, DNA distortion, and B-linker ordering.

5.3 General features of transcription initiation

Our results also provide insights into the evolution of the three eukaryotic transcription machineries and help explain why the two other nuclear RNA polymerases, Pol I and Pol III, do neither require energy from ATP hydrolysis nor a TFIIH-like factor for DNA opening. Recent studies on Pol III transcription initiation complexes revealed a very similar path of promoter DNA and equivalent structural elements for most of the basal Pol II TFs (Abascal-Palacios et al., 2018; Vorländer et al., 2018). Two CCs were resolved with promoter DNA along the closed cleft of Pol III (Vorländer et al., 2018) that correspond to the Pol II cCC1. It was proposed that opening of the Pol III clamp allows promoter DNA to enter the cleft and that clamp closure leads to DNA opening (Vorländer et al., 2018), also consistent with the Pol II DNA opening mechanism.

For Pol I, structural studies of initiation complexes showed that promoter DNA is positioned further inside the cleft compared to the Pol II and Pol III systems (Engel et al., 2017; Han et al., 2017; Sadian et al., 2017). Promotor DNA was observed running into the cleft directly over the Pol I wall, rather than being held well above the wall as in the Pol II and Pol III systems. A Pol I CC was not resolved, but modelling suggested that DNA loading into the cleft and subsequent promoter opening require clamp opening and closing, respectively (Engel et al., 2017), strongly suggesting a similar overall mechanism to what we describe here for Pol II. Thus, despite the differences between the three eukaryotic nuclear transcription machineries, features of all three systems include DNA loading, distortion and opening in the cleft and the necessity of opening and closing of the clamp.

Whereas Pol I and III promoters apparently melt easily, most Pol II promoters have evolved to contain stable IMRs. This may have rendered them TFIIH-dependent and more regulatable during evolution. In addition, DNA opening may be facilitated in the Pol I and Pol III systems because DNA is more extensively bound by the PICs of Pol I and Pol III and is loaded more deeply into the cleft when compared to Pol II (Supplemental Table 2). This may allow for tighter DNA binding and more severe DNA distortions during the conversion to a CC^{dist} intermediate. As a result, the efficiency of the transition from CC^{dist} to OC would be higher and DNA opening would be easier in the Pol I and Pol III systems. A lower stability of CC^{dist} may explain why this intermediate could not be trapped for Pol I and Pol III.

The archaeal initiation system is highly homologous to a minimal Pol II system and changes of the clamp state have been observed during promoter opening (Schulz et al., 2016). The initiation mechanism differs for bacterial RNA polymerase. In bacterial CC structures, promoter DNA resides further above the cleft and is contacted only by the accessory sigma factor (Glyde et al., 2017). Promoter opening is initiated by sigma factor in the upper region of the active center cleft by trapping of bases that are flipped-out (Feklistov and Darst, 2011). However, DNA bubble extension and DNA loading into the cleft require clamp opening and closing (Chakraborty et al., 2012; Feklistov and Darst, 2011). We suggest that the coupling of clamp closure to DNA distortion in the cleft is a universal feature of transcription initiation that facilitates DNA opening by multi-subunit RNA polymerases.

5.4 Roles of PIC components during promoter DNA melting

5.4.1 TFIIB

The basal transcription factor IIB (TFIIB) consists of several functional domains and was shown to be required for promoter DNA melting (Kostrewa et al., 2009). This was also confirmed for the archaeal TFIIB homologue TFB (Kostrewa et al., 2009; Spitalny and Thomm, 2003). In particular, mutation or deletion of amino acid residues located in the B-linker domain impaired promoter melting by the PIC. It was shown that the B-linker of human TFIIB intercalates between the two DNA strands in the PIC and stabilizes the DNA after OC formation suggesting a mechanism how it facilitates the promoter opening process (He et al., 2016). However, the yeast B-linker is interacting much less with the DNA after opening (Schilbach et al., 2017) and was even observed to be disordered in a yeast open complex with an overextended transcription initiation bubble (Plaschka et al., 2016). Although the B-linker is an integral part of the Pol II-TFIIB interaction (Sainsbury et al., 2013), the conservation of the B-linker on sequence and structural levels is rather weak between species (Figure 14). Therefore, it seems likely that the TFIIB linker region functions by a less specific and yet undiscovered mechanism during promoter melting.

Despite the low structural and sequence homology, the TFIIB linker occupies a relatively conserved binding site at the clamp coiled-coil (Sainsbury et al., 2013). Mutational studies confirmed this binding interface to be important for promoter DNA opening and

transcription initiation *in vitro* and *in vivo* (Kostrewa et al., 2009). As shown in this work, closure of the clamp domain and the clamp coiled-coil in particular distorts the promoter DNA. Hence, a possible function of the TFIIB linker could be to facilitate clamp closure during transcription initiation and stabilize the closed clamp state that we observed in the CC^{dist} intermediate. The human PIC structures show an open clamp state with the promoter DNA in a canonical B-DNA conformation. An explanation for the difference between the human and the yeast system could be the high mobility of the mammalian Pol II clamp domain (Bernecky et al., 2016; Kostek et al., 2006). As a consequence, spontaneous promoter DNA melting as observed in the yeast system would be much reduced or even unobservable in the human system.

5.4.2 TFIIF subunit Tfg2 winged helix domain

Transcription initiation factor IIF consists of a heterodimer formed by Tfg1 and Tfg2. Both subunits fold into a dimerization domain that functions as a binding module to the Pol II lobe domain. Both Tfg subunits possess auxiliary winged helix DNA binding domains. Deletion of the Tfg2 domain leads to severe defects in transcription initiation (Tang 1995) and confers lethality when deleted *in vivo* (Eichner et al., 2010). The underlying mechanism of Tfg2 WH function during transcription initiation, however, has not been understood yet.

In this work we describe how the Tfg2 WH domain could facilitate promoter opening. Binding of the Tfg2 WH to promoter DNA upstream of the IMR locks the DNA in the cleft where the closed clamp and TFIIE then induce DNA distortions that facilitate DNA opening. Additionally, the Tfg2 WH could define the most upstream position of an intermediate bubble edge of early promoter melting states and direct DNA melting in the downstream direction. In the mature yeast transcription initiation bubble, the upstream edge is defined by the TFIIE wing (Plaschka et al., 2016), however, in the human open complex the Tfg2 WH keeps the strands separated (He et al., 2016). Altogether, a disruption of these functions by mutation and deletion of the Tfg2 WH would explain the observed requirement of the Tfg2 WH for transcription initiation and cell survival in yeast (Eichner et al., 2010).

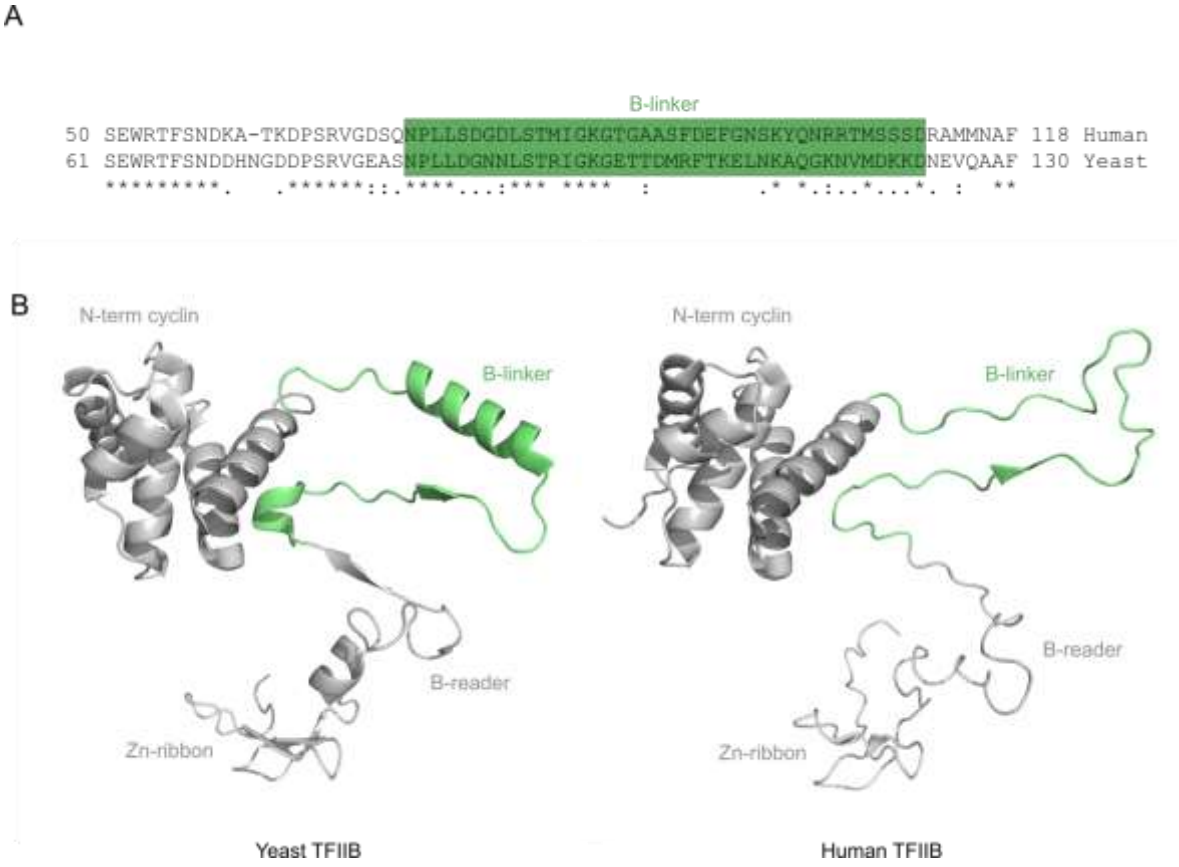


Figure 14: Sequence and structure conservation of yeast and human TFIIB

(**A**) Sequence alignment of the yeast and human TFIIB linker regions including the flanking regions. The B-linker region is marked in green.

(**B**) Structures of yeast (PDB 4BBR) and human TFIIB (PDB 5IY8). The B-linker region is colored in green, the rest of TFIIB in grey. The C-terminal cyclin domain was omitted for clarity

Since deletions and modifications of the Tfg2 WH are lethal *in vivo*, a contribution of the Tfg2 WH to promoter DNA melting would need to be tested in an *in vitro* system. Experiments similar the here described 2-AP DNA opening assays using mutant Tfg2 could elucidate a contribution of the Tfg2 WH to promoter DNA opening and transcription initiation.

5.4.3 TFIIE

TFIIE consists of the two subunits Tfa1 and Tfa2 that form a hetero dimer. Binding of TFIIE to the PIC occurs between the Pol II clamp and stalk modules (He et al., 2013; Plaschka et al., 2016)(Figure 9). Biochemical analysis suggested a role of TFIIE during promoter DNA opening (Holstege et al., 1995). Cryo-EM structures of an OC showed how the TFIIE wing of Tfa1 stabilizes the upstream DNA junction of the melted promoter DNA in yeast (Plaschka et al., 2016). In human, however, this function is taken over by the Tfg2 WH (He et al., 2016). This suggests that the Tfg2 WH and the TFIIE wing are redundant for defining the upstream bubble edge in a transcription initiation complex with melted promoter DNA. This redundancy would explain why TFIIE wing deletions on yeast only lead to very mild growth defects (Plaschka et al., 2016).

The deletion of both tandem WH domains of Tfa2 is lethal. However, when WH2 is retained WH1 is dispensable for normal growth suggesting some redundancy between the two tandem WH domains (Grunberg et al., 2012). From structural studies it is known, that the Tfg2 WH of TFIIF can be positioned correctly without TFIIE in human (He et al., 2013) but not in yeast (Plaschka et al., 2015). Although in this work I could observe promoter DNA melting in the absence of TFIIE (Figure 8), its requirement in other biochemical studies (Holstege et al., 1995; Holstege et al., 1996) suggests that one of the Tfa2 WH domains might be important to correctly position the Tfg2 WH in the PIC.

Other than the tandem WH domains, only the WH and zinc ribbon domains of Tfa1 as well as the tether (Plaschka et al., 2016) of Tfa2 which facilitates the hetero-dimerization of Tfa1 and Tfa2 are essential for yeast survival (Grunberg et al., 2012). They are likely more important for TFIIE recruitment to the PIC than for promoter DNA opening and are therefore not discussed here.

Altogether, the intricate network of interactions formed by TFIIE with TFIIF and the promoter DNA provides the structural frame for DNA distortion during promoter opening. The Tfa2 WH correctly positions the Tfg2 WH which locks the DNA in the cleft and likely prevents bubble extension further upstream towards the TATA box. The TFIIE wing helps locking the DNA and can define the upstream edge of the mature transcription initiation

bubble. Thereby, the rigid positioning of promoter DNA by TFIIF and TFIIE makes DNA distorting during the CC-to-OC transition possible.

5.4.4 TFIIH

In this work, mostly the contribution of the core initiation factors TFIIB, F and E to promoter DNA opening was investigated. However, for the vast majority of the genes, TFIIH is the main factor that facilitates promoter DNA opening *in vivo* (Figure 7). Therefore, it seems obvious that upon TFIIH binding to the core PIC it will perform its DNA translocase activity to melt the promoter DNA as suggested by cryo-EM studies of the full PIC (Schilbach et al., 2017). However, ATP is required for DNA translocase activity by Ssl2 (Fishburn et al., 2015). In contrast, without ATP, Ssl2 would just bind the downstream promoter DNA. Since Ssl2 binding to the downstream DNA strongly stabilized its position within the PIC, it might inhibit the here described spontaneous opening mechanism because it prevents loading of the downstream promoter DNA into the Pol II cleft.

Such inhibitory function would only be significant if the dissociation constant of Ssl2 with ATP lies within the normal physiological range. ATP levels within the cell can vary between 2 and 4 mM depending on the nutritional status of the cell (Yoshida et al., 2016). A binding constant for ATP to Ssl2 has not been determined yet. Structure determination of TFIIH in context of a PIC in the presence of 1 mM AMP-PNP or ADP-BeF_3 did not result in visible occupation of the Ssl2 ATP-binding site (Schilbach et al., 2017)(Schilbach S., personal communication). This implies that the affinity of Ssl2 for ATP or its hydrolysis products may lie at least within the milli-molar range. Interestingly, this is close to the physiological ATP concentration in the cell that was measured (Yoshida et al., 2016). Based on this I want to speculate, that the nutritional status of the cell could define whether TFIIH acts as an inhibitor to spontaneous promoter DNA melting or is an activator of transcription. Such hypothesis could be tested *in vitro* and *in vivo*. The fluorescence based 2-AP opening assay from this work could be used to investigate whether TFIIH binding to promoter DNA indeed causes inhibition of spontaneous DNA melting. Furthermore, the activity of genes under starvation and normal conditions could be investigated by 4sU sequencing. A comparison with the Ssl2 anchor-away data from this

work could show whether active genes under starvation conditions depend less on Ssl2 activity.

5.4.5 RNA polymerase II elements

Despite the transcription initiation factors, also RNA polymerase II itself contributes to the process of promoter DNA opening by its clamp domain that causes DNA distortion. However, other elements of multi-subunit RNA polymerases might help DNA strand separation in addition to the DNA distortion.

Regions that interact with closed promoter DNA are around the clamp head and the Pol II lobe domain as they lie in close proximity to the IMR. The lobe between contains a few basic residues which likely are important to align promoter DNA properly in the Pol II cleft. However, these residues are probably of less importance for single strand formation.

On the other side of the cleft, promoter DNA is closely aligned along the clamp head domain of Pol II. A short antiparallel β-sheet harbouring positively charged residues helps to align the promoter DNA along the clamp head. In addition, one of the usually unstructured loops of the clamp head is particularly enriched in acidic amino acids (Figure 15A, Asp 188, Asp 193, Asp 195, Glu 196 and Glu 198). This seems unexpected because these residues are in very close proximity to the negatively charged backbone of closed promoter DNA. Therefore I want to speculate, that these acidic residues might help to keep the negatively charged DNA strands separated once the DNA has undergone initial melting in that region. Supporting this hypothesis, the human Pol II clamp head domain also contains two acidic amino acids (Figure 15B, Glu 201, Glu 207, Asp 208 and Glu 211) in that short loop which could serve a similar function. However, such function remains to be very speculative and has to be experimentally tested using mutated Pol II samples in a promoter melting and transcription initiation assays.

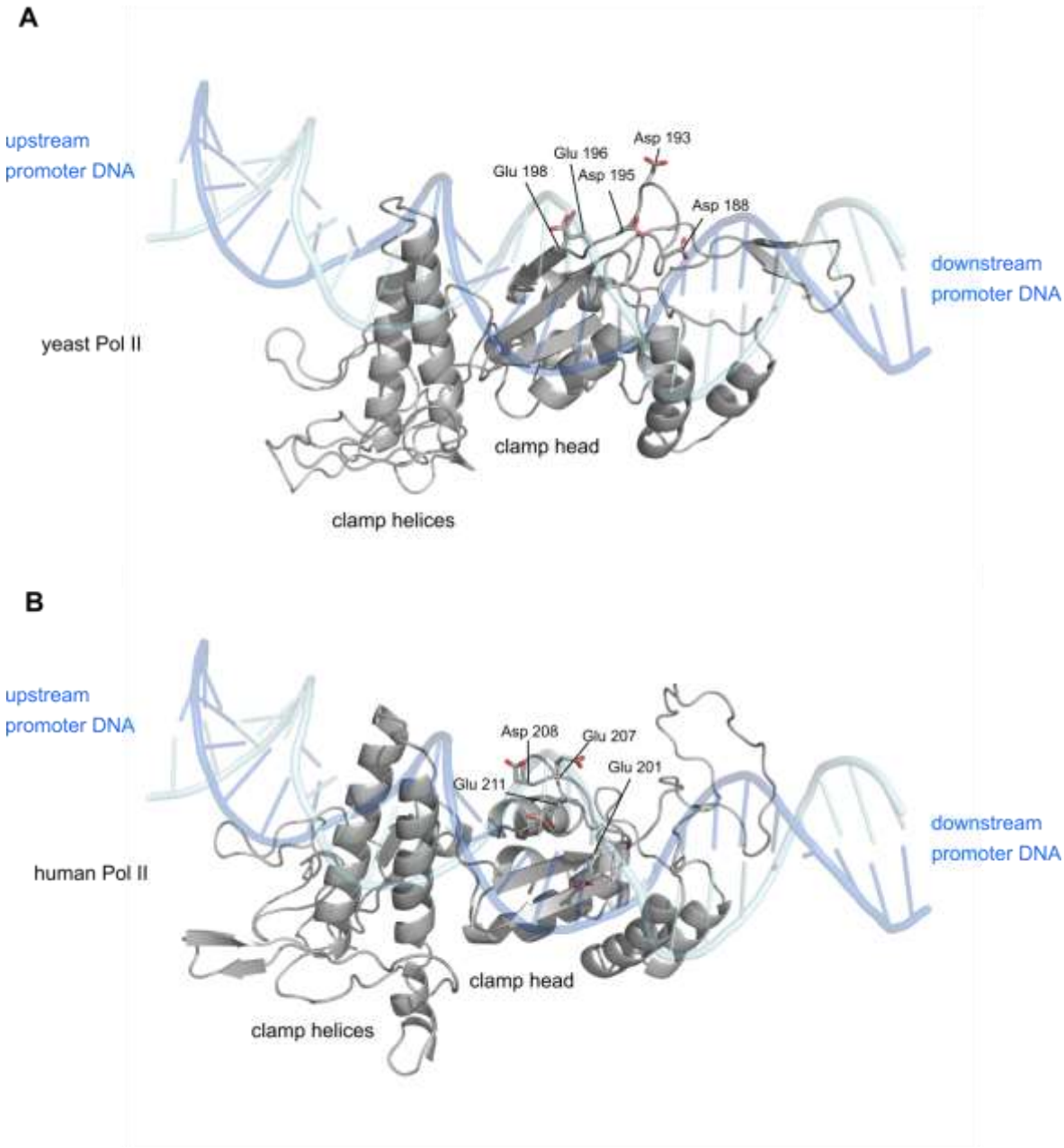


Figure 15: Acidic residues in close proximity to closed promoter DNA

(**A**) A loop in the yeast clamp head domain is enriched in acidic residues. The yeast clamp domain (PDB 1I50) is shown as cartoon. Acidic residues are shown in stick representation. CC^{dist} promoter DNA was positioned by aligning the clamp domain of the CC^{dist} structure to PDB 1I50.

(**B**) The loop within the human clamp head domain also contains many acidic residues. The clamp domain of human Pol II was generated using SWISS-model (Waterhouse et al., 2018) using Rpb1 of PDB 1I50 as a template. CC^{dist} promoter DNA was positioned by aligning the CC^{dist} structure to the Rpb1 homology model.

5.4.6 Other transcription initiation factors

In addition to the factors that obviously contribute to promoter DNA melting by direct interactions with the promoter DNA or by enzymatic activities, other components of the PIC such as TFIID and mediator could influence DNA opening as well. Such effects would be rather allosteric and could target the activity of TFIIH as TFIIE likely does (Schilbach et al., 2017). In addition, interactions of TFIID and mediator with the other factors and elements discussed before could regulate their effect on promoter DNA opening. Due to the lack of structural information on these complexes at high resolution, extend and mechanism of such effects remain elusive.

5.5 Evolution of promoter DNA opening

The complexity of gene regulation and organization differs greatly between prokaryotic and eukaryotic transcription systems that utilize multi-subunit RNAPs. Bacterial genomes are usually very compact (Lillo and Krakauer, 2007; Rogozin et al., 2002a; Rogozin et al., 2002b) and genes are often organized in operons (Jacob and Monod, 1961; Salgado et al., 2000) making co-regulation of functionally related genes easy as they are controlled by a single promoter. Archaeal genes are as well organised in operons, however, the degree of operon utilization for transcriptional regulation varies among archaeal species (Koonin and Wolf, 2008). Eukaryotes do not utilize operon structures for most of their gene regulation (Koonin, 2009). However, genes transcribed by eukaryotic Pol I (rRNA) are organized in clustered arrays (Moss et al., 2007) and transcription is regulated in an on-off fashion defining the cellular rRNA dosage by the number of active genes (Grummt, 2007; Grummt and Ladurner, 2008; Grummt and Pikaard, 2003; Lawrence and Pikaard, 2004; McKeown and Shaw, 2009). Transcription initiation at tRNA genes by Pol III is switched on or off in a similar manner by changing the ration of transcription factors IIIB and IIIC (Graczyk et al., 2018). In contrast to the relatively simple transcription systems of bacteria, archaea and the eukaryotic Pol I and III, RNA polymerase II transcription initiation is highly complex. The high number of individually transcribed genes requires varying promoter sequences (Carninci et al., 2006) and regulatory factors (Koonin and Wolf, 2008) to ensure a stable and fine-tuned transcription network.

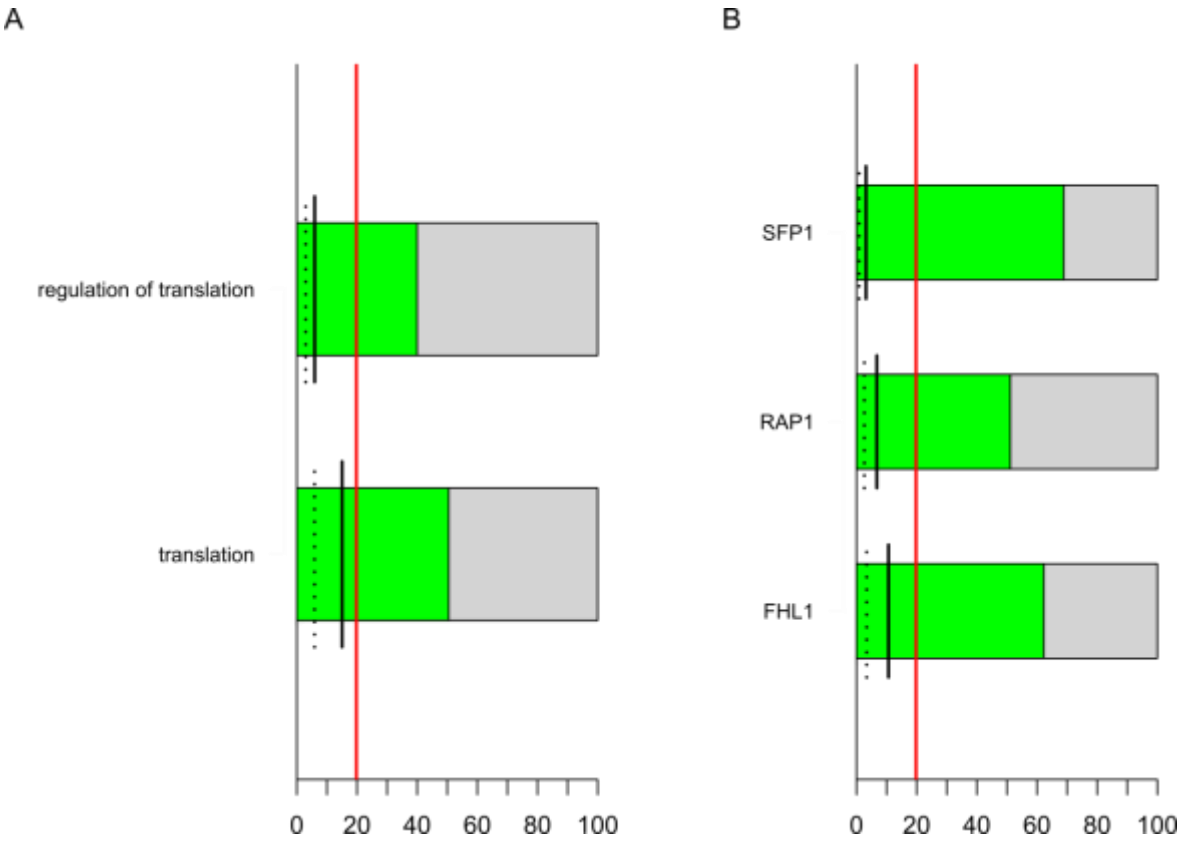


Figure 16: Ssl2 independent genes are enriched for translation related proteins

(**A**) Genes not responding to Ssl2 nuclear depletion by anchor away are enriched in translation related genes. Enrichment analysis was done using fisher tests. Top GO categories sorted by p-values. Grey bars depict the size of each GO category scaled to 100%. The red line indicates the proportion of each category that is expected by chance. Green bars show the actual enrichment of each category in percent. Black lines indicate the number of enriched genes (green bars) in relation to the number of tested genes. Dashed lines show the size of each category in proportion to the whole number of genes.

(**B**) Genes not responding to Ssl2 nuclear depletion are enriched for genes controlled by the SFP1, RAP1 and FHL1 transcription factors. Enrichment analysis was done using fisher tests. Top transcription factors are sorted by p-values. Grey bars depict the size of each transcription factor category scaled to 100%. The red line indicates the proportion of each category that is expected by chance. Green bars show the actual enrichment of each category in percent. Black lines indicate the number of enriched genes (green bars) in relation to the number of tested genes. Dashed lines show the size of each category in proportion to the whole number of genes.

Interestingly, while RNAPs of bacteria and archaea as well as Pol I and Pol III open promoter DNA without an external energy source, only the transcription initiation process of Pol II requires ATP hydrolysis for initiation. An exception are genes that do not depend

on Ssl2 during transcription initiation by Pol II. This group is strongly enriched in translation related genes (Figure 16A). Particularly, these are ribosomal proteins which are highly co-regulated by a small number of transcription factors (Figure 16B)(Marion et al., 2004; Reja et al., 2015). Transcription of ribosomal proteins is orchestrated with rRNA transcription by Pol I and tRNA transcription by Pol III (Martin et al., 2006). As promoter DNA melting is a key step of transcription initiation, it seems likely that this process has been adapted to the varying regulatory requirements during evolution of transcriptional regulation. I therefore want to speculate, that ATP-independent promoter DNA opening is sufficient for the complexity of bacterial and archaeal transcription as well as eukaryotic ribosome biogenesis. ATP-independent promoter opening might have been evolutionary maintained in these systems by a selection for IMRs with unstable DNA duplexes. With increasing regulatory requirements for the complex transcription network of Pol II genes, these promoters might have evolved towards more stable duplexes to reduce the extend of spontaneous promoter opening. On the other hand, this required utilization of the DNA repair factor TFIIH to ensure efficient promoter DNA melting and simultaneously increased the achievable regulatory complexity. Interestingly, the bacterial sigma-54 system has evolved an ATP-dependent activation step during initiation (Kustu et al., 1991) that is reminiscent of the Pol II system suggesting that the concept of ATP-dependent activated initiation is favourable concept in evolution of transcription regulation.

Altogether, it seems likely that ATP-dependent promoter opening might have evolved to ensure fine-tuned regulation of Pol II transcription.

5.6 Open questions and future perspective

The work presented here provides an extended view on how transcription is initiated. In particular, I reported a molecular model for promoter opening that unifies the mechanism for all multi-subunit polymerase transcription systems. During this work, especially the slight differences between the human and the yeast system enabled us to conclude on a general model as they revealed different intermediate steps. Nonetheless, these two systems are different in some aspects and therefore, a major research focus will be to study human promoter opening. In addition, further investigation of the Pol I and Pol III systems may ultimately prove the mechanistic similarities described here.

5.6.1 TFIIH-independent transcription in humans cells

As shown in this work, TFIIH-independent transcription is possible in yeast at promoters with weak DNA duplexes in the IMR. TFIIH-independent promoter melting and transcription also have been shown *in vitro* and *in vivo,* however, without further investigation of the required sequence context. A study on TFIIH inhibition in human cells by triptolite, THZ1and spironolactone aimed on investigating the effects of perturbing TFIIH activity and found transcription insensitive for XPB (yeast Ssl2) removal (Alekseev et al., 2017). However, the authors only looked at global cellular transcription levels, which did not allow them to observe changes of transcription for individual genes. A study aiming to observe genome wide changes of transcription would require two important experimental prerequisites: (i) a non-lethal, transient depletion or degradation of the XPB translocase (Ssl2 in yeast); and (ii) a high-resolution sequencing technique that can detect the levels of newly synthesized RNAs which accurately reflect the initiation rates of genes (Gressel et al., 2017).

The first can be solved by recent advances in mammalian genome editing techniques. The CRISPR/Cas9 (Sander and Joung, 2014) system made it possible to stably modify the genome of human cells. Although the experimental procedures are more complex and less well established to date, CRISPR/Cas9 can be used for protein tagging as it is routinely done for yeast. This allows the creation cell lines for transient and specific protein removal by the auxin-inducible degron system (Nishimura et al., 2009). The combination of these methods could be used to rapidly, transiently and quantitatively remove the human Ssl2 homologue XPB from mammalian cells.

The second experimental hurdle can be taken by using 4sU (4-thio uridine) metabolic labeling. Newly synthesized RNAs will incorporate 4sU during transcription and the 4sU label then can be used to purify these RNA molecules. In comparison to cells with intact TFIIH, next-generation sequencing will then reveal which genes lose transcriptional activity upon XPB removal. This could show whether TFIIH-independent transcription initiation occurs at human promoters *in vivo*. This furthermore could show which genes and promoters are dependent or independent of XPB activity and whether this correlates with the sequence composition of the IMR.

5.6.2 Transcription start site scanning by the human PIC

The TFIIH translocase Ssl2 is important for transcription initiation from promoters with more stable DNA duplexes in their IMR. An additional function of Ssl2 in yeast is scanning for the transcription start site (Kuehner and Brow, 2006) as the TSS may be located far downstream of the region that is initially opened and loaded to the RNA polymerase active center cleft (Hampsey, 1998). However, it seems as yeast cells can tolerate a loss of scanning since the Ssl2-dependent and independent genes presented in this work did not show any preference for the position of their TSS (data not shown).

In contrast to yeast core promoters, the human TSS is usually found 29-33 bp downstream of the TATA box sequence (Hampsey, 1998). This led to the conclusion that during transcription initiation in human cells, no TFIIH driven TSS scanning by Pol II is required, as the TSS is consistently positioned in the active site after promoter melting. However, this only holds true for so called "focused" promoters (Kadonaga, 2012). "Dispersed" promoters utilize different TSS for initiation that can be spread over regions as large as 50-100 base pairs. In addition to that, mammalian genes can have multiple TSS that can be differentially used (Reyes and Huber, 2018). This raises the question whether TFIIH-driven TSS scanning by Pol II may be used for transcription from dispersed promoters and for selection of alternative TSS. Alternatively, the initial positioning of the PIC could define the TSS that is being used.

In yeast, *in vivo* studies (Kuehner and Brow, 2006) gave rise to a unidirectional scanning model for the Pol II PIC that is now widely accepted. However, such or similar experiments were never done in mammalian cells to prove the hypothesis that the mammalian PIC does not scan for the TSS. Such could be done using a plasmid system with modified promoters that have TSS which are shifted further downstream. A reporter gene and transcript 5'-end mapping by sequencing or primer extension could reveal how these artificially shifted TSS are used *in vivo*.

To test whether the positioning of the PIC defines the TSS that is being used for initiation, already existing genome wide sequencing data could be analysed. 5'SAGE sequencing data (Hashimoto et al., 2004; Shiraki et al., 2003) accurately identified the 5'-ends of human mRNAs *in vivo*. The positions of human PICs in vivo can be retrieved from high-

resolution ChIP-exo data (Pugh and Venters, 2016). A combined analysis of such datasets could show if TSS and PIC assembly position correlate and thus suggest a model for TSS selection by PIC positioning in humans.

5.6.3 Promoter opening intermediates

In this work, the cryo-EM structure of a novel closed-to-open transition intermediate with closed distorted promoter DNA was shown adding another piece to the jigsaw puzzle of the molecular pathway of promoter DNA melting. The transition to the fully opened transcription initiation bubble may exhibit additional metastable intermediates that have not been structurally observed yet due to their unstable nature. Early biochemical studies with purified transcription initiation systems suggested the existence of defined steps during promoter opening and initial melting. One of these intermediates would be an initially opened bubble of 5 nt which was shown to ablate the requirement for TFIIH during promoter DNA melting (Holstege et al., 1996). This 5 nt bubble state was recently rediscovered as a relatively stable intermediate by single molecule experiments investigating promoter opening and scanning by TFIIH (Tomko et al., 2017). It was shown, that the DNA translocase activity of Ssl2 creates a 5-6 bp bubble that remains stable over significant time spans strongly arguing that such state might be specifically stabilized by the PIC.

The next step to extend our knowledge about the molecular mechanism of promoter DNA melting is to trap such intermediate state in a structure of the PIC. Biochemical studies suggest the position of such initial bubble to be at the downstream end of the IMR. However, stable preparation of such state is difficult because the PIC melts promoter DNA spontaneously. By chemical locking of the DNA duplex base pairs, such spontaneous opening could be prevented. The synthesis of nucleic acids with covalent base pairs was described extensively and could be adapted for a structural biology experiments (Gao and Orgel, 1999; Hattori et al., 2009; Jensen et al., 2008; Li et al., 2001a; Li et al., 2001b; Taniguchi et al., 2010). Such covalent base pairs would allow structural investigation of several transition bubble sizes on the pathway from closed to open promoter DNA.

An alternative and more natural approach would be based on a native closed DNA which is restricted to small amounts of TFIIH driven promoter DNA melting. For this, the PIC

could be formed on a closed DNA that is difficult to melt spontaneously (e.g. the GAT1 promoter sequence). The sample would then be prepared for cryo-EM and incubated with small sub-equimolar amounts of ATP for a very short time prior to flash freezing. This could lead to an ensemble of transition states between closed and open complexes. If the 5-6 nt bubble intermediate has enhanced stability, one would expect it to be enriched to a certain degree.

Both approaches could lead to a structure of an actively melting complex, providing deep insights into the mechanism of DNA opening by TFIIH.

5.6.4 Energy landscape of promoter opening

Promoter DNA melting in the absence of TFIIH and ATP hydrolysis depends on a weak enough DNA duplex in the promoter IMR. Therefore, promoter DNA stability within the IMR defines a major energy barrier during the closed-to-open transition. This suggests that small differences of the energy barrier during promoter melting might be decisive over whether a gene is transcribed or not. The height of that barrier might be modulated by protein-protein and protein-DNA interactions within the PIC allowing very fine tuning of the initiation process. However, the exact mechanism of modulating promoter DNA melting remains largely unknown.

An approach to investigate the thermodynamics of promoter DNA melting would be molecular dynamics simulations. This could involve simulation of PICs with closed promoter DNA in order to observe the dynamics of the complex before opening occurs. The comparison of very strong, weak and miss-matched DNA duplexes in the IMR of these models might provide an excellent framework to investigate the energetic properties of the closed-to-open transition. In such set up, spontaneous melting could be simulated and energetically described. Simulation of partially assembled PICs could then reveal potential mechanisms of energy modulation that regulates promoter DNA melting.

Additionally, more complex simulations with TFIIH and Ssl2 translocase activity would show how the energy barrier of a stable IMR can be overcome by ATP hydrolysis. Altogether, such molecular dynamics simulations could reveal the energy landscape of the promoter opening process.

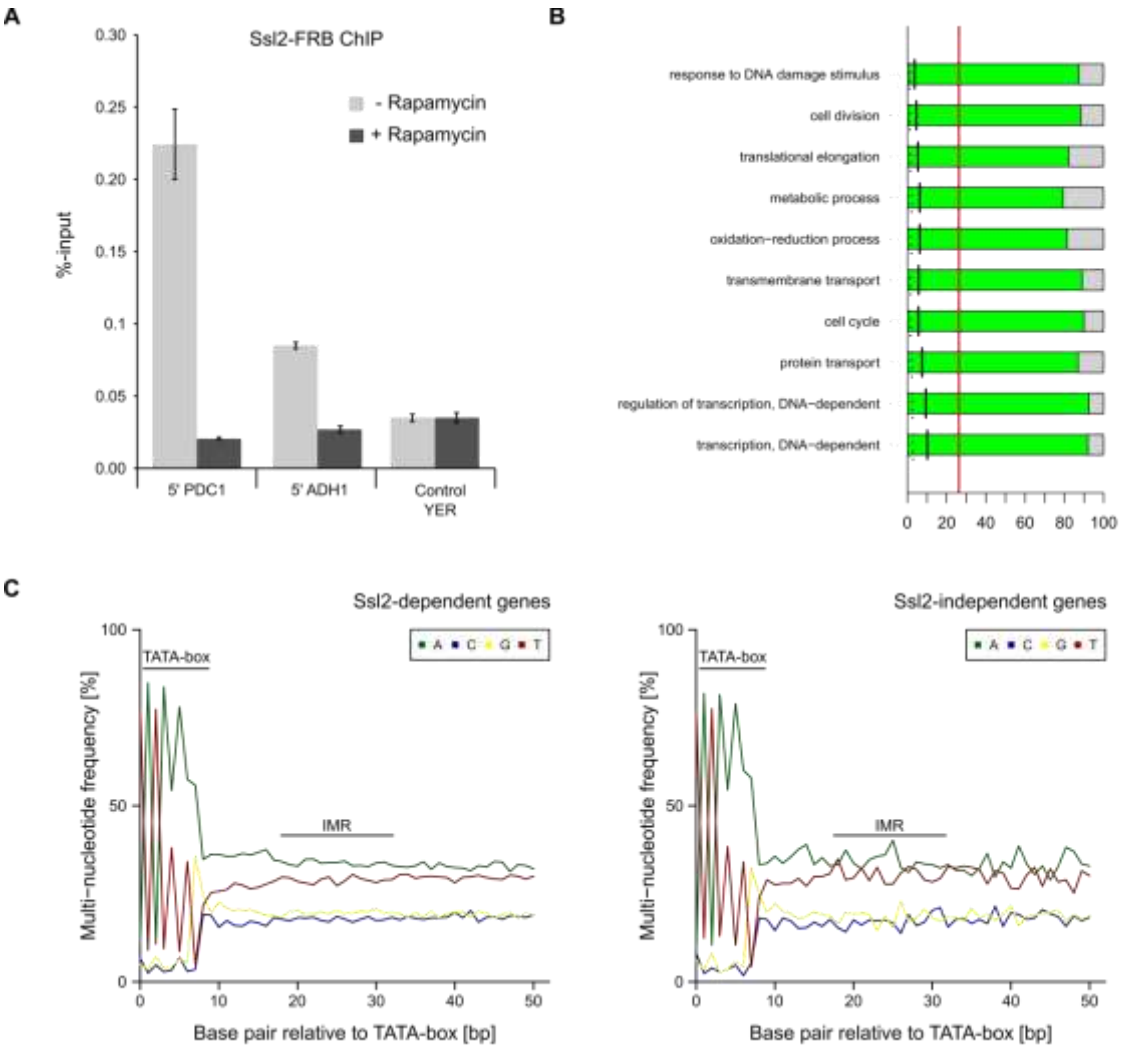
5.6.5 Equivalents of CC^{dist} and other transition states in Pol I, Pol III and bacterial PICs

In this work, a unified model for promoter opening was proposed for the Pol I, Pol II and Pol III initiation systems. Additionally, conceptual similarities with the bacterial system were discussed. In that model, the distorted closed complex (CC^{dist}) is a key intermediate reflecting a state just before promoter DNA melting. The low efficiency of spontaneous DNA melting by the Pol II PIC with a stable IMR enabled structural characterization. However, the high efficiency of OC-to-CC conversion by the Pol I, Pol III and bacterial transcription systems makes it difficult to obtain closed complexes or transition states on the pathway to the open complex.

Stable closed complexes or transition states could be forced by using very stable DNA duplexes as done for the Pol II system in this work. If such would be unsuccessful, covalent base pairs could help to keep the DNA duplex together. Closed complex structures from the Pol I, Pol III or bacterial initiation machinery that exhibit similar distortions as described in this work would potentially provide an ultimate proof for the unified mechanism proposed.

6 SUPPLEMENTAL MATERIALS

6.1 Supplemental Figures



Supplemental Figure 1: Validations and controls for Ssl2 anchor away

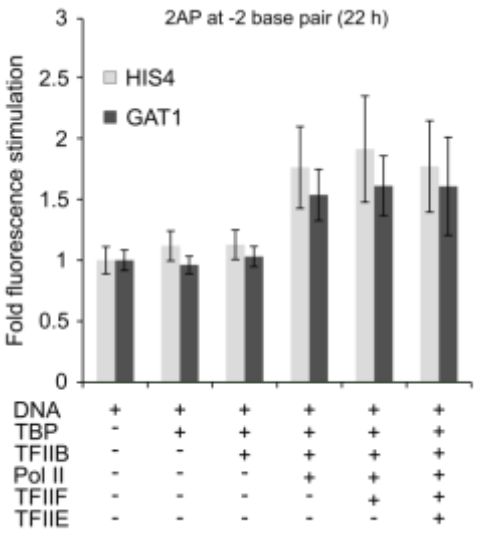
(**A**) Ssl2-FRB ChIP qPCR without and with rapamycin treatment. Primers for qPCR were chosen in the 5' region of the respective gene close to the promoter. YER is the random chromatin control.

(**B**) GO-term enrichment of Ssl2-dependent genes. Enrichment analysis was done using fisher tests. Top GO categories sorted by p-values. Grey bars depict the size of each GO category scaled to 100%. The red line indicates the proportion of each category that is expected by chance. Green bars show the actual enrichment of each category in percent. Black lines indicate the number of enriched genes (green bars) in relation to the number of tested genes. Dashed lines show the size of each category in proportion to the whole number of genes.

(**C**) Ssl2-dependent genes do not show base enrichment in their IMR. Average base content of promoter regions downstream of TATA for Ssl2-dependent genes are shown in % for each base position. Genes were aligned at their TATA box. TATA box and IMR are marked with solid lines.

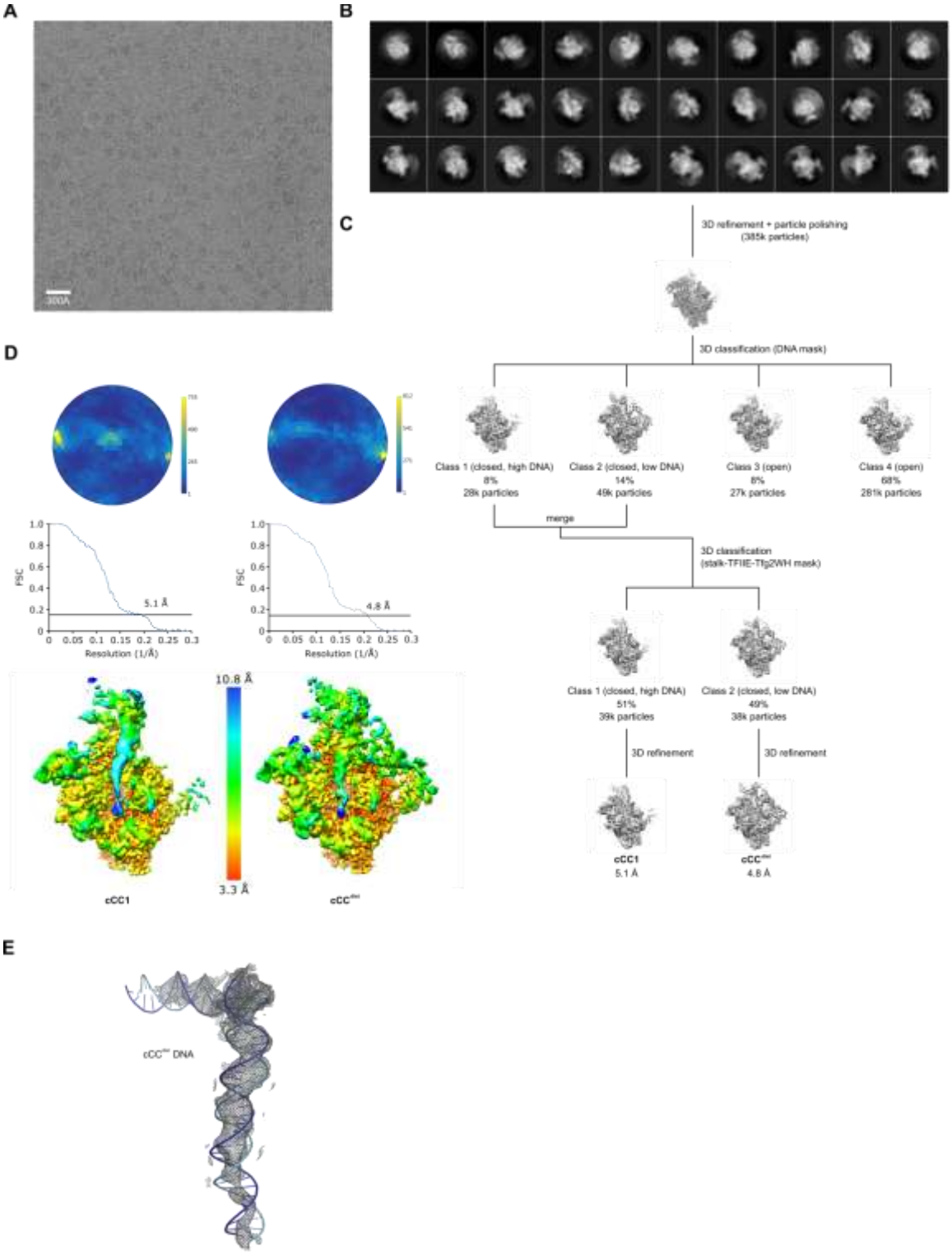
(**D**) Ssl2-independ genes do not show base enrichment in their IMR. Average base content of promoter regions downstream of TATA for Ssl2-independent genes are shown in % for each base position. Genes were aligned at their TATA box. TATA box and IMR are marked with solid lines.

(**E**) Rapamycin treatment induces growth phenotype in Ssl2 anchor away yeast. The growth curve of the Ssl2 anchor away strain is shown under normal (solid) and Ssl2 depletion (dashed) conditions. Growth curves were collected over 20 hours in YPD medium at 30°C.



Supplemental Figure 2: TFIIH independent opening of GAT1 promoter DNA

The GAT1 promoter melts spontaneously in vitro after 22 hours. Spontaneous melting of the GAT1 promoter is achieved after incubation time with cPIC components was extended to 22 hours. Bar plots show the fluorescence increase normalized to the DNA only reaction. 2-AP labelled HIS4 promoter DNA after 22 hours is shown for comparison.



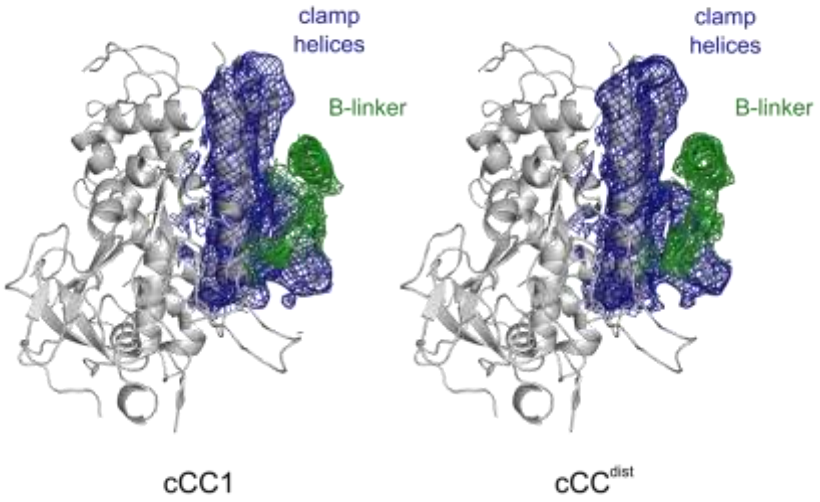
Supplemental Figure 3: Data processing of cCC1 and cCC^{dist}

(**A**) A micrograph representative for the collected data is shown. The micrographs have been contrast enhanced for clarity.

(**B**) Selected 2D classes that were used after manual and 2D-classification based particle cleanup. 2D classes are ordered by number of particles in descending order.

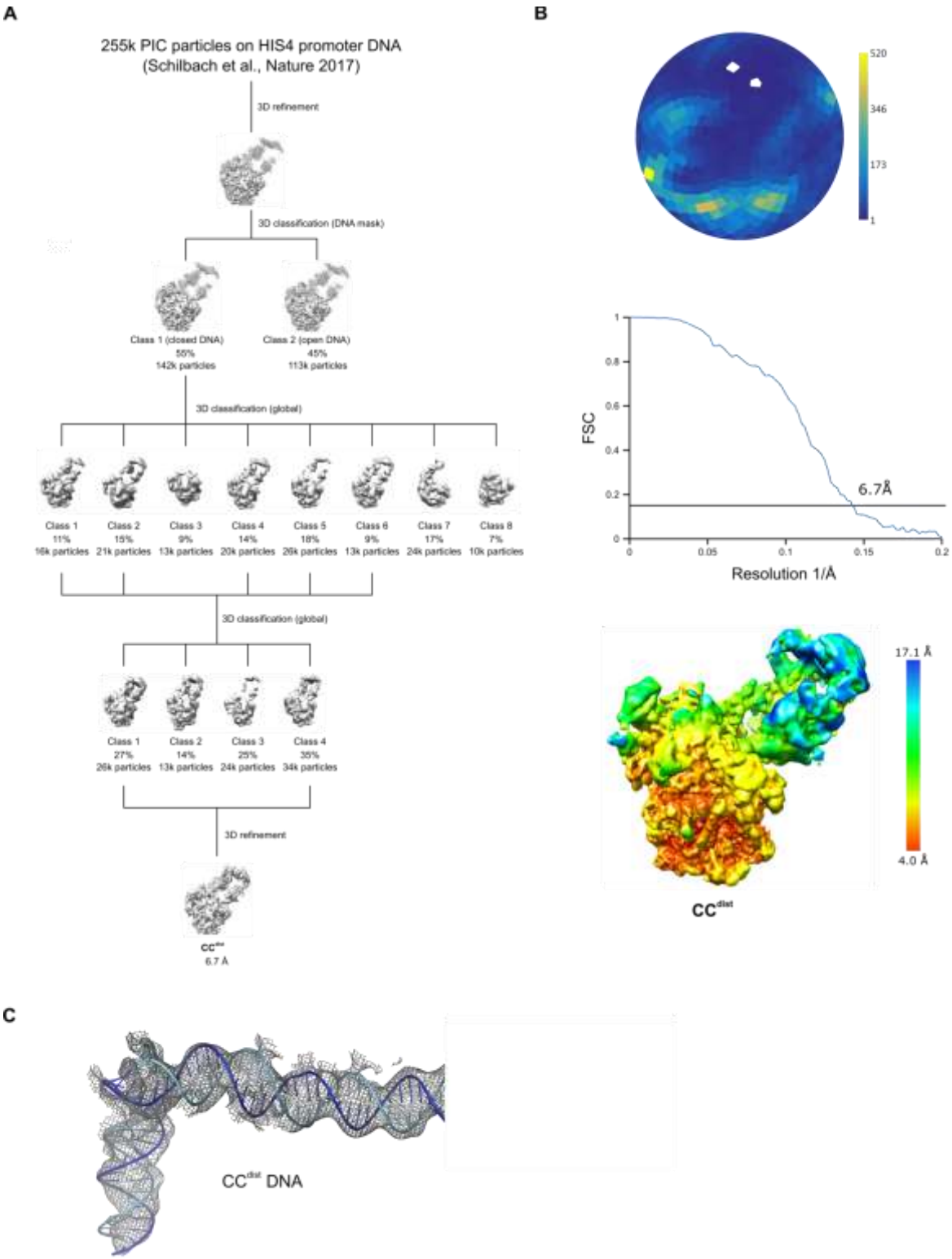
(**C**) Schematic for processing strategy of the cCC1/cCC^{dist} dataset. Reconstructions of each processing step are shown as grey surfaces. Particle numbers are rounded. Resolution is given according to the 0.143 FSC threshold and with a B-factor of -100Å^2.

(**D**) Angular distribution and Fourier-shell correlation (FSC) plots for cCC1 and cCC^{dist} reconstructions. Angular distribution was calculated in 7.5° orientation bins. The 0.143 FSC threshold is marked in FSC lots as solid line. Local resolution was calculates by a combination of local FSC weighting and B-factor sharpening. Locally filtered volumes of cCC1 and cCC^{dist} are shown in front view (**Figure 9**A and **Figure 9**B).



Supplemental Figure 4: Clamp and B-linker of cCC1 and cCC^{dist}

The B-linker is well ordered in cCC1 and cCC^{dist}.The Pol II clamp and B-linker of cCC1 and cCC^{dist} are shown in cartoon representation. The map for clamp helices and B-linker is shown as blue and green mesh, respectively.



Supplemental Figure 5: Data processing of CC^{dist}

(**A**) 3D sorting of PIC particles leads to the CC^{dist} reconstruction. A schematic of the 3D sorting strategy for the CC^{dist} reconstruction is shown. Intermediate reconstructions are shown as grey volumes. Particle numbers are rounded. Resolution is given according to the 0.143 FSC threshold and with a B-factor of -100Å2.
(**B**) Angular distribution and Fourier-shell correlation (FSC) plots for the CC^{dist} reconstruction.

Angular distribution was calculated in 7.5° orientation bins. The 0.143 FSC threshold is indicated as solid line. Local resolution was calculates by a combination of local FSC weighting and B-factor sharpening. Locally filtered volume of CC^{dist} is shown in side view (Figure 11).

6.2 Supplemental Tables

Supplemental Table 1: Data collection and processing statistics

	cCC1	cCC^{dist}	CC^{dist}
Data collection			
Acc. Voltage (kV)	300	300	300*
Detector	K2 Summit (GIF)	K2 Summit (GIF)	K2 Summit (GIF)*
Electron dose ($e^-/Å^2$)	37	37	42*
Defocus range (μm)	-0.8 to -3.0	-0.8 to -3.0	-0.5 to -5.0*
Magnification (x)	130 000	130 000	105 000*
Pixel size (Å/px)	1.07	1.07	1.37*
Particles	39 000	38 000	60 000*
Symmetry imposed	C1	C1	C1*
Reconstruction			
Map resolution	5.1	4.8	6.7
Map sharpening B-factor ($Å^2$)	-100	-100	-100
FSC threshold	0.143	0.143	0.143

* Values from Schilbach et al., 2017

Supplemental Table 2: Eukaryotic closed complexes

	Pol II						*Pol I**	*Pol III*
State	**cCC**	**cCC1**	**cCC2**	**CC2**	**cCC^{dist}**	**CC^{dist}**	**CC2**	**CC1**
Reference	Plaschka et al., 2016	this study	He et al., 2016	He et al., 2016	this study	this study	Engel et al., 2017	Vorländer et al., 2018
Resolution	7.8 Å	5.1 Å	5.4 Å	7.2 Å	4.8 Å	8.0 Å	modeled	5.5/4.2 Å
DNA distance from active site	72.9 Å	71.5 Å	55.9 Å	55.9 Å	64.3 Å	64.3 Å	44.9 Å	68.4 Å
DNA conformation	canonical	canonical	canonical	canonical	distorted	distorted	canonical	canonical
Clamp	closed	closed	open	open	closed	closed	open	closed
B-linker	weak	+	-	-	+	+		
B-reader	+	+	-	-	+	+		
TFIIE	weak	-	+	+	+	+		
Tfg2WH	weak	-	+	+	+	+		

* based on structural modeling.

7 REFERENCES

Abascal-Palacios, G., Ramsay, E.P., Beuron, F., Morris, E., and Vannini, A. (2018). Structural basis of RNA polymerase III transcription initiation. Nature *553*, 301.

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr *66*, 213-221.

Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat Rev Genet *13*, 720-731.

Alekseev, S., Nagy, Z., Sandoz, J., Weiss, A., Egly, J.M., Le May, N., and Coin, F. (2017). Transcription without XPB Establishes a Unified Helicase-Independent Mechanism of Promoter Opening in Eukaryotic Gene Expression. Mol Cell *65*, 504-514 e504.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol *11*, R106.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169.

Armache, K.J., Kettenberger, H., and Cramer, P. (2003). Architecture of initiation-competent 12-subunit RNA polymerase II. Proc Natl Acad Sci U S A *100*, 6964-6968.

Baejen, C., Andreani, J., Torkler, P., Battaglia, S., Schwalb, B., Lidschreiber, M., Maier, K.C., Boltendahl, A., Rus, P., Esslinger, S., *et al.* (2017). Genome-wide Analysis of RNA Polymerase II Termination at Protein-Coding Genes. Mol Cell *66*, 38-49 e36.

Bedwell, G.J., Appling, F.D., Anderson, S.J., and Schneider, D.A. (2012). Efficient transcription by RNA polymerase I using recombinant core factor. Gene *492*, 94-99.

Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J.M., and Cramer, P. (2016). Structure of transcribing mammalian RNA polymerase II. Nature *529*, 551-554.

Brueckner, F., and Cramer, P. (2008). Structural basis of transcription inhibition by alpha-amanitin and implications for RNA polymerase II translocation. Nat Struct Mol Biol *15*, 811-818.

Brueckner, F., Ortiz, J., and Cramer, P. (2009). A movie of the RNA polymerase nucleotide addition cycle. Curr Opin Struct Biol *19*, 294-299.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engstrom, P.G., Frith, M.C., *et al.* (2006). Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet *38*, 626-635.

Chakraborty, A., Wang, D., Ebright, Y.W., Korlann, Y., Kortkhonjia, E., Kim, T., Chowdhury, S., Wigneshweraraj, S., Irschik, H., Jansen, R., *et al.* (2012). Opening and closing of the bacterial RNA polymerase clamp. Science *337*, 591-595.

Chasman, D.I., Flaherty, K.M., Sharp, P.A., and Kornberg, R.D. (1993). Crystal structure of yeast TATA-binding protein and model for interaction with DNA. Proc Natl Acad Sci U S A *90*, 8174-8178.

Cheetham, G.M., Jeruzalmi, D., and Steitz, T.A. (1999). Structural basis for initiation of transcription from an RNA polymerase-promoter complex. Nature *399*, 80-83.

Cheetham, G.M., and Steitz, T.A. (2000). Insights into transcription: structure and function of single-subunit DNA-dependent RNA polymerases. Curr Opin Struct Biol *10*, 117-123.

Chen, H.T., Warfield, L., and Hahn, S. (2007). The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex. Nat Struct Mol Biol *14*, 696-703.

Cheung, A.C., and Cramer, P. (2011). Structural basis of RNA polymerase II backtracking, arrest and reactivation. Nature *471*, 249-253.

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

Cramer, P. (2002). Common structural features of nucleic acid polymerases. Bioessays *24*, 724-729.

Cramer, P., Bushnell, D.A., Fu, J., Gnatt, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R., and Kornberg, R.D. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. Science *288*, 640-649.

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

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

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.

Dvir, A., Garrett, K.P., Chalut, C., Egly, J.M., Conaway, J.W., and Conaway, R.C. (1996). A role for ATP and TFIIH in activation of the RNA polymerase II preinitiation complex prior to transcription initiation. J Biol Chem *271*, 7245-7248.

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

Eick, D., and Geyer, M. (2013). The RNA polymerase II carboxy-terminal domain (CTD) code. Chem Rev *113*, 8456-8490.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr *60*, 2126-2132.

Engel, C., Gubbey, T., Neyer, S., Sainsbury, S., Oberthuer, C., Baejen, C., Bernecky, C., and Cramer, P. (2017). Structural Basis of RNA Polymerase I Transcription Initiation. Cell *169*, 120-131 e122.

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

Feklistov, A., Sharon, B.D., Darst, S.A., and Gross, C.A. (2014). Bacterial sigma factors: a historical, structural, and genomic perspective. Annu Rev Microbiol *68*, 357-376.

Fishburn, J., Galburt, E., and Hahn, S. (2016). Transcription Start Site Scanning and the Requirement for ATP during Transcription Initiation by RNA Polymerase II. J Biol Chem *291*, 13040-13047.

Fishburn, J., Tomko, E., Galburt, E., and Hahn, S. (2015). Double-stranded DNA translocase activity of transcription factor TFIIH and the mechanism of RNA polymerase II open complex formation. Proc Natl Acad Sci U S A *112*, 3961-3966.

Gao, K., and Orgel, L.E. (1999). Nucleic acid duplexes incorporating a dissociable covalent base pair. Proc Natl Acad Sci U S A *96*, 14837-14842.

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

Glyde, R., Ye, F., Darbari, V.C., Zhang, N., Buck, M., and Zhang, X. (2017). Structures of RNA Polymerase Closed and Intermediate Complexes Reveal Mechanisms of DNA Opening and Transcription Initiation. Mol Cell *67*, 106-116 e104.

Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. Science *292*, 1876-1882.

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

Graczyk, D., Ciesla, M., and Boguta, M. (2018). Regulation of tRNA synthesis by the general transcription factors of RNA polymerase III - TFIIIB and TFIIIC, and by the MAF1 protein. Biochim Biophys Acta Gene Regul Mech *1861*, 320-329.

Greber, B.J., Nguyen, T.H.D., Fang, J., Afonine, P.V., Adams, P.D., and Nogales, E. (2017). The cryo-electron microscopy structure of human transcription factor IIH. Nature *549*, 414-417.

Gressel, S., Schwalb, B., Decker, T.M., Qin, W., Leonhardt, H., Eick, D., and Cramer, P. (2017). CDK9-dependent RNA polymerase II pausing controls transcription initiation. Elife *6*.

Grummt, I. (2007). Different epigenetic layers engage in complex crosstalk to define the epigenetic state of mammalian rRNA genes. Hum Mol Genet *16 Spec No 1*, R21-27.

Grummt, I., and Ladurner, A.G. (2008). A metabolic throttle regulates the epigenetic state of rDNA. Cell *133*, 577-580.

Grummt, I., and Pikaard, C.S. (2003). Epigenetic silencing of RNA polymerase I transcription. Nat Rev Mol Cell Biol *4*, 641-649.

Grunberg, S., Warfield, L., and Hahn, S. (2012). Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. Nat Struct Mol Biol *19*, 788-796.

Hahn, S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. Nat Struct Mol Biol *11*, 394-403.

Hampsey, M. (1998). Molecular genetics of the RNA polymerase II general transcriptional machinery. Microbiol Mol Biol Rev *62*, 465-503.

Hampsey, M., Singh, B.N., Ansari, A., Laine, J.P., and Krishnamurthy, S. (2011). Control of eukaryotic gene expression: gene loops and transcriptional memory. Adv Enzyme Regul *51*, 118-125.

Han, Y., Yan, C., Nguyen, T.H.D., Jackobel, A.J., Ivanov, I., Knutson, B.A., and He, Y. (2017). Structural mechanism of ATP-independent transcription initiation by RNA polymerase I. Elife *6*.

Haruki, H., Nishikawa, J., and Laemmli, U.K. (2008). The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. Mol Cell *31*, 925-932.

Hashimoto, S., Suzuki, Y., Kasai, Y., Morohoshi, K., Yamada, T., Sese, J., Morishita, S., Sugano, S., and Matsushima, K. (2004). 5'-end SAGE for the analysis of transcriptional start sites. Nat Biotechnol *22*, 1146-1149.

Hattori, K., Hirohama, T., Imoto, S., Kusano, S., and Nagatsugi, F. (2009). Formation of highly selective and efficient interstrand cross-linking to thymine without photo-irradiation. Chem Commun (Camb), 6463-6465.

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

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

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

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

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

Hontz, R.D., French, S.L., Oakes, M.L., Tongaonkar, P., Nomura, M., Beyer, A.L., and Smith, J.S. (2008). Transcription of multiple yeast ribosomal DNA genes requires targeting of UAF to the promoter by Uaf30. Mol Cell Biol *28*, 6709-6719.

Jacob, F., and Monod, J. (1961). On the Regulation of Gene Activity. Cold Spring Harbor Symposia on Quantitative Biology *26*, 193-211.

Jensen, E.A., Allen, B.D., Kishi, Y., and O'Leary, D.J. (2008). Conformational analysis of a covalently cross-linked Watson-Crick base pair model. Bioorg Med Chem Lett *18*, 5884-5887.

Jin, H., and Kaplan, C.D. (2014). Relationships of RNA polymerase II genetic interactors to transcription start site usage defects and growth in Saccharomyces cerevisiae. G3 (Bethesda) *5*, 21-33.

Juven-Gershon, T., and Kadonaga, J.T. (2010). Regulation of gene expression via the core promoter and the basal transcriptional machinery. Dev Biol *339*, 225-229.

Kadonaga, J.T. (2012). Perspectives on the RNA polymerase II core promoter. Wiley Interdiscip Rev Dev Biol *1*, 40-51.

Kannan, S., Kohlhoff, K., and Zacharias, M. (2006). B-DNA under stress: over- and untwisting of DNA during molecular dynamics simulations. Biophys J *91*, 2956-2965.

Kaplan, C.D., Jin, H., Zhang, I.L., and Belyanin, A. (2012). Dissection of Pol II trigger loop function and Pol II activity-dependent control of start site selection in vivo. PLoS Genet *8*, e1002627.

Kashkina, E., Anikin, M., Brueckner, F., Lehmann, E., Kochetkov, S.N., McAllister, W.T., Cramer, P., and Temiakov, D. (2007). Multisubunit RNA polymerases melt only a single DNA base pair downstream of the active site. J Biol Chem *282*, 21578-21582.

Kassavetis, G.A., and Geiduschek, E.P. (2006). Transcription factor TFIIIB and transcription by RNA polymerase III. Biochem Soc Trans *34*, 1082-1087.

Kastner, B., Fischer, N., Golas, M.M., Sander, B., Dube, P., Boehringer, D., Hartmuth, K., Deckert, J., Hauer, F., Wolf, E., *et al.* (2008). GraFix: sample preparation for single-particle electron cryomicroscopy. Nat Methods *5*, 53-55.

Keener, J., Josaitis, C.A., Dodd, J.A., and Nomura, M. (1998). Reconstitution of yeast RNA polymerase I transcription in vitro from purified components. TATA-binding protein is not required for basal transcription. J Biol Chem *273*, 33795-33802.

Kettenberger, H., Armache, K.J., and Cramer, P. (2003). Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. Cell *114*, 347-357.

Keys, D.A., Lee, B.S., Dodd, J.A., Nguyen, T.T., Vu, L., Fantino, E., Burson, L.M., Nogi, Y., and Nomura, M. (1996). Multiprotein transcription factor UAF interacts with the upstream element of the yeast RNA polymerase I promoter and forms a stable preinitiation complex. Genes Dev *10*, 887-903.

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

Koonin, E.V. (2009). Evolution of genome architecture. Int J Biochem Cell Biol *41*, 298-306.

Koonin, E.V., and Wolf, Y.I. (2008). Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. Nucleic Acids Res *36*, 6688-6719.

Kostek, S.A., Grob, P., De Carlo, S., Lipscomb, J.S., Garczarek, F., and Nogales, E. (2006). Molecular architecture and conformational flexibility of human RNA polymerase II. Structure *14*, 1691-1700.

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

Kuehner, J.N., and Brow, D.A. (2006). Quantitative analysis of in vivo initiator selection by yeast RNA polymerase II supports a scanning model. J Biol Chem *281*, 14119-14128.

Kulaeva, O.I., Hsieh, F.K., Chang, H.W., Luse, D.S., and Studitsky, V.M. (2013). Mechanism of transcription through a nucleosome by RNA polymerase II. Biochim Biophys Acta *1829*, 76-83.

Kustu, S., North, A.K., and Weiss, D.S. (1991). Prokaryotic transcriptional enhancers and enhancer-binding proteins. Trends Biochem Sci *16*, 397-402.

Lane, W.J., and Darst, S.A. (2010a). Molecular evolution of multisubunit RNA polymerases: sequence analysis. J Mol Biol *395*, 671-685.

Lane, W.J., and Darst, S.A. (2010b). Molecular evolution of multisubunit RNA polymerases: structural analysis. J Mol Biol *395*, 686-704.

Lariviere, L., Plaschka, C., Seizl, M., Petrotchenko, E.V., Wenzeck, L., Borchers, C.H., and Cramer, P. (2013). Model of the Mediator middle module based on protein cross-linking. Nucleic Acids Res *41*, 9266-9273.

Lariviere, L., Plaschka, C., Seizl, M., Wenzeck, L., Kurth, F., and Cramer, P. (2012). Structure of the Mediator head module. Nature *492*, 448-451.

Lawrence, R.J., and Pikaard, C.S. (2004). Chromatin turn ons and turn offs of ribosomal RNA genes. Cell Cycle *3*, 880-883.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Li, H.Y., Qiu, Y.L., and Kishi, Y. (2001a). Solution structure of n-type DNA oligomers possessing a covalently cross-linked Watson-Crick base pair model. Chembiochem *2*, 371-374.

Li, H.Y., Qiu, Y.L., Moyroud, E., and Kishi, Y. (2001b). Synthesis of DNA Oligomers Possessing a Covalently Cross-Linked Watson-Crick Base Pair Model Financial support from the National Institutes of Health (Grant: NS 12108) is gratefully acknowledged. Angew Chem Int Ed Engl *40*, 1471-1475.

Li, X., Mooney, P., Zheng, S., Booth, C.R., Braunfeld, M.B., Gubbens, S., Agard, D.A., and Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nat Methods *10*, 584-590.

Lillo, F., and Krakauer, D.C. (2007). A statistical analysis of the three-fold evolution of genomic compression through frame overlaps in prokaryotes. Biol Direct *2*, 22.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

Luse, D.S. (2013). Promoter clearance by RNA polymerase II. Biochim Biophys Acta *1829*, 63-68.

Luse, D.S., and Jacob, G.A. (1987). Abortive initiation by RNA polymerase II in vitro at the adenovirus 2 major late promoter. J Biol Chem *262*, 14990-14997.

Mandel, C.R., Bai, Y., and Tong, L. (2008). Protein factors in pre-mRNA 3'-end processing. Cell Mol Life Sci *65*, 1099-1122.

Marion, R.M., Regev, A., Segal, E., Barash, Y., Koller, D., Friedman, N., and O'Shea, E.K. (2004). Sfp1 is a stress- and nutrient-sensitive regulator of ribosomal protein gene expression. Proc Natl Acad Sci U S A *101*, 14315-14322.

Martin, D.E., Powers, T., and Hall, M.N. (2006). Regulation of ribosome biogenesis: where is TOR? Cell Metab *4*, 259-260.

Martinez-Rucobo, F.W., Sainsbury, S., Cheung, A.C., and Cramer, P. (2011). Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. EMBO J *30*, 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 Dev *8*, 515-524.

McKeown, P.C., and Shaw, P.J. (2009). Chromatin: linking structure and function in the nucleolus. Chromosoma *118*, 11-23.

Merkhofer, E.C., Hu, P., and Johnson, T.L. (2014). Introduction to cotranscriptional RNA splicing. Methods Mol Biol *1126*, 83-96.

Moss, T., Langlois, F., Gagnon-Kugler, T., and Stefanovsky, V. (2007). A housekeeper with power of attorney: the rRNA genes in ribosome biogenesis. Cellular and Molecular Life Sciences *64*, 29-49.

Muhlbacher, 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. Nat Commun *5*, 4310.

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. Proc Natl Acad Sci U S A *112*, 13543-13548.

Murakami, K.S. (2015). Structural biology of bacterial RNA polymerase. Biomolecules *5*, 848-864.

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*, 119-128.

Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T., and Kanemaki, M. (2009). An auxin-based degron system for the rapid depletion of proteins in nonplant cells. Nat Methods *6*, 917-922.

O'Sullivan, J.M., Tan-Wong, S.M., Morillon, A., Lee, B., Coles, J., Mellor, J., and Proudfoot, N.J. (2004). Gene loops juxtapose promoters and terminators in yeast. Nat Genet *36*, 1014-1018.

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

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*, 533-540.

Pilsl, M., Merkl, P.E., Milkereit, P., Griesenbeck, J., and Tschochner, H. (2016). Analysis of S. cerevisiae RNA Polymerase I Transcription In Vitro. Methods Mol Biol *1455*, 99-108.

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

Plaschka, C., Lariviere, L., Wenzeck, L., Seizl, M., Hemann, M., Tegunov, D., Petrotchenko, E.V., Borchers, C.H., Baumeister, W., Herzog, F., *et al.* (2015). Architecture of the RNA polymerase II-Mediator core initiation complex. Nature *518*, 376-380.

Pugh, B.F., and Venters, B.J. (2016). Genomic Organization of Human Transcription Initiation Complexes. PLoS One *11*, e0149339.

Ramstein, J., and Lavery, R. (1988). Energetic coupling between DNA bending and base pair opening. Proc Natl Acad Sci U S A *85*, 7231-7235.

Rani, P.G., Ranish, J.A., and Hahn, S. (2004). RNA polymerase II (Pol II)-TFIIF and Pol II-mediator complexes: the major stable Pol II complexes and their activity in transcription initiation and reinitiation. Mol Cell Biol *24*, 1709-1720.

Reja, R., Vinayachandran, V., Ghosh, S., and Pugh, B.F. (2015). Molecular mechanisms of ribosomal protein gene coregulation. Genes Dev *29*, 1942-1954.

Reyes, A., and Huber, W. (2018). Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. Nucleic Acids Res *46*, 582-592.

Roeder, R.G., and Rutter, W.J. (1969). Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. Nature *224*, 234-237.

Rogozin, I.B., Makarova, K.S., Natale, D.A., Spiridonov, A.N., Tatusov, R.L., Wolf, Y.I., Yin, J., and Koonin, E.V. (2002a). Congruent evolution of different classes of non-coding DNA in prokaryotic genomes. Nucleic Acids Res *30*, 4264-4271.

Rogozin, I.B., Spiridonov, A.N., Sorokin, A.V., Wolf, Y.I., Jordan, I.K., Tatusov, R.L., and Koonin, E.V. (2002b). Purifying and directional selection in overlapping prokaryotic genes. Trends Genet *18*, 228-232.

Rohou, A., and Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. J Struct Biol *192*, 216-221.

Rosa, M.D. (1979). Four T7 RNA polymerase promoters contain an identical 23 bp sequence. Cell *16*, 815-825.

Sadian, Y., Tafur, L., Kosinski, J., Jakobi, A.J., Wetzel, R., Buczak, K., Hagen, W.J., Beck, M., Sachse, C., and Muller, C.W. (2017). Structural insights into transcription initiation by yeast RNA polymerase I. EMBO J *36*, 2698-2709.

Sainsbury, S., Bernecky, C., and Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. Nat Rev Mol Cell Biol *16*, 129-143.

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

Salgado, H., Moreno-Hagelsieb, G., Smith, T.F., and Collado-Vides, J. (2000). Operons in Escherichia coli: genomic analyses and predictions. Proc Natl Acad Sci U S A *97*, 6652-6657.

Sander, J.D., and Joung, J.K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol *32*, 347-355.

SantaLucia, J., Jr. (1998). A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc Natl Acad Sci U S A *95*, 1460-1465.

Scheres, S.H. (2012). A Bayesian view on cryo-EM structure determination. J Mol Biol *415*, 406-418.

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

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*, 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. Proc Natl Acad Sci U S A *113*, E1816-1825.

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

Shandilya, J., and Roberts, S.G. (2012). The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling. Biochim Biophys Acta *1819*, 391-400.

Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., Kodzius, R., Watahiki, A., Nakamura, M., Arakawa, T., *et al.* (2003). Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. Proc Natl Acad Sci U S A *100*, 15776-15781.

Sousa, R., Chung, Y.J., Rose, J.P., and Wang, B.C. (1993). Crystal structure of bacteriophage T7 RNA polymerase at 3.3 A resolution. Nature *364*, 593-599.

Spitalny, P., and Thomm, M. (2003). Analysis of the open region and of DNA-protein contacts of archaeal RNA polymerase transcription complexes during transition from initiation to elongation. J Biol Chem *278*, 30497-30505.

Steitz, T.A. (1998). A mechanism for all polymerases. Nature *391*, 231-232.

Sweetser, D., Nonet, M., and Young, R.A. (1987). Prokaryotic and eukaryotic RNA polymerases have homologous core subunits. Proc Natl Acad Sci U S A *84*, 1192-1196.

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

Taniguchi, Y., Kurose, Y., Nishioka, T., Nagatsugi, F., and Sasaki, S. (2010). The alkyl-connected 2-amino-6-vinylpurine (AVP) crosslinking agent for improved selectivity to the cytosine base in RNA. Bioorg Med Chem *18*, 2894-2901.

Tomko, E.J., Fishburn, J., Hahn, S., and Galburt, E.A. (2017). TFIIH generates a six-base-pair open complex during RNAP II transcription initiation and start-site scanning. Nat Struct Mol Biol *24*, 1139-1145.

Tyree, C.M., George, C.P., Lira-DeVito, L.M., Wampler, S.L., Dahmus, M.E., Zawel, L., and Kadonaga, J.T. (1993). Identification of a minimal set of proteins that is sufficient for accurate initiation of transcription by RNA polymerase II. Genes Dev *7*, 1254-1265.

van Dijk, M., and Bonvin, A.M. (2009). 3D-DART: a DNA structure modelling server. Nucleic Acids Res *37*, W235-239.

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

Vorländer, M.K., Khatter, H., Wetzel, R., Hagen, W.J.H., and Müller, C.W. (2018). Molecular mechanism of promoter opening by RNA polymerase III. Nature *553*, 295.

Wang, S.M., Kim, G.J., and Gewirth, D.T. (2001). Structural studies of a yeast quaternary transcription-initiation complex. Acta Crystallogr D Biol Crystallogr *57*, 441-444.

Wang, W., Carey, M., and Gralla, J.D. (1992). Polymerase II promoter activation: closed complex formation and ATP-driven start site opening. Science *255*, 450-453.

Warfield, L., Ramachandran, S., Baptista, T., Devys, D., Tora, L., and Hahn, S. (2017). Transcription of Nearly All Yeast RNA Polymerase II-Transcribed Genes Is Dependent on Transcription Factor TFIID. Mol Cell *68*, 118-129 e115.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., *et al.* (2018). SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res *46*, W296-W303.

Yoshida, T., Kakizuka, A., and Imamura, H. (2016). BTeam, a Novel BRET-based Biosensor for the Accurate Quantification of ATP Concentration within Living Cells. Sci Rep *6*, 39618.

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

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

Zhang, Z., and Dietrich, F.S. (2005). Mapping of transcription start sites in Saccharomyces cerevisiae using 5' SAGE. Nucleic Acids Res *33*, 2838-2851.

Zheng, G., Lu, X.J., and Olson, W.K. (2009). Web 3DNA--a web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. Nucleic Acids Res *37*, W240-246.

8 LIST OF FIGURES

Figure 1: The central dogma of molecular biology. Adapted from Crick, 1970 1

Figure 2: RNA polymerases of T7 phage, bacteria and eukaryotes. Figure adapted from PhD thesis of H. Hillen. 2

Figure 3: Scheme of transcription initiation by RNA polymerase II 6

Figure 4: Cloning, expression and purification of yeast full length TBP 30

Figure 5: Expression and Purification of TFIIAΔ113 32

Figure 6: Formation of cPIC and cPIC-cMed by gradient ultracentrifugation 34

Figure 7: Ssl2-independent transcription *in vivo* 36

Figure 8: Promoter opening depends on the initially melted region 38

Figure 9: Cryo-EM structures of core closed complexes (cCC) 39

Figure 10: Distortion of promoter DNA 41

Figure 11: PIC structure with closed, distorted DNA 43

Figure 12: IMRs of Pol I and Pol III promoters are unstable 44

Figure 13: General model of transcription initiation 46

Figure 14: Sequence and structure conservation of yeast and human TFIIB 50

Figure 15: Acidic residues in close proximity to closed promoter DNA 54

Figure 16: Ssl2 independent genes are enriched for translation related proteins 56

Supplemental Figure 1: Validations and controls for Ssl2 anchor away 63

Supplemental Figure 2: TFIIH independent opening of GAT1 promoter DNA 64

Supplemental Figure 3: Data processing of cCC1 and cCC^{dist} 65

Supplemental Figure 4: Clamp and B-linker of cCC1 and cCC^{dist} 66

Supplemental Figure 5: Data processing of CC^{dist} 67

9 LIST OF TABLES

Table 1: Sub-units of the three nuclear RNA polymerases. Corresponding subunits of Pol I, Pol II and Pol III are shown together with their function. Functional counterparts of the Pol II system that are not Pol II subunits are shown in italic. Table was modified from Vannini and Cramer, 2012. 3

Table 2: Components of the basal transcription initiation machinery and their function 7

Table 3: Chemicals used in this work..... 12

Table 4: Antibiotics used in this work..... 13

Table 5: Culture media used in this work..... 13

Table 6: Culture media additives used in this work 13

Table 7: Buffers used for TBP purification 13

Table 8: Buffers used for TFIIAΔ113 purification..... 14

Table 9: Other buffers and solutions used in this work. 14

Table 10: *E. coli* strains used in this work..... 15

Table 11: *S. cerevisiae* strains used in this work..... 15

Table 12: Vectors used in this work 16

Table 13: Oligonucleotides used in this work 16

Table 14: Enzymes used in this work..... 17

Table 15: Kits used in this work..... 18

Table 16: Components used for cPIC and cPIC-cMed reconstitutions. *molar excess relative to Pol II. #cMed was not added for cPIC preparations. 33

LIST OF TABLES

Supplemental Table 1: Data collection and processing statistics 68

Supplemental Table 2: Eukaryotic closed complexes 69

10 LIST OF ABBREVIATIONS

Abbreviation	
bp	base pairs
BRE	TFIIB recognition element
CC	closed complex
cPIC	core pre-initiation complex
CTD	RNA polymerase II C-terminal domain
EC	elongation complex
EM	electron microscopy
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
His	histidine
IMR	initially melted region
IPTG	isopropyl-β-D-1-thiogalactopyranoside
ITC	initially transcribing complex
nt	nucleotides
NTP	nucleotide triphosphate
OC	open complex
PIC	pre-initiation complex
Pol I/II/III	RNA polymerase I/II/III
RNAP	RNA polymerase
SDS	sodium dodecyl sulfate
TF	basal transcription factor
TSS	transcription start site
UV	ultra violet

11 CURRICULUM VITAE

Personal details

Name:	Christian Dienemann
Date of birth:	February 23, 1988
Nationality:	German

Education:

Sept. 1994 – July 2007	Freie Waldorfschule Jena (comprehensive school) **Degree: Abitur (high school diploma) May 25, 2007; grade 1.7 (B+)**
Oct. 2008 – Sept. 2011	University of Jena, Germany Subject: Biochemistry (B.Sc.) **Degree: Bachelor of Science, Aug 11, 2011; grade 1.5 (A-)**
Sept. 2010 – Jan. 2011	Belarussian State University Minsk, Belarus Exchange student
Feb. 2011 – June 2011	Aarhus University, Denmark Exchange student
Oct. 2011 – Sept. 2013	University of Jena, Germany Subject: Biochemistry (M.Sc.) **Degree: Master of Science, Sept 9, 2013; grade 1.1 (A+)**

Research activities:

Feb. 2011 – Sept. 2011	Brodersen Laboratory for Protein Crystallography Centre for Structural Biology, Aarhus University, Denmark Project student (Bachelor Thesis)
April 2012 – Sept. 2013	Than Laboratory for Protein Crystallography Leibniz Institute for Age Research, Jena, Germany Master student (Master Thesis)
October 2013	Than Laboratory for Protein Crystallography Leibniz Institute for Age Research, Jena, Germany Researcher

Nov. 2013 – Jan. 2014	Cramer Laboratory Gene center, Ludwig.Maximillians-University, Munich, Germany Researcher
Since Feb. 2014	Cramer Laboratory Max-Planck-Institute for biophysical chemistry, Göttingen, Germany PhD Student
Focus areas:	Biochemistry of proteins and nucleic acids, structural biology, X-ray crystallography, cryo-EM

Other activities:

Aug. 2007 – April 2008	Compulsory community service at University Hospital Jena As alternative to military service
May 2008 – June 2008	Internship at University Hospital Jena
Oct. 2009 – March 2010	Teacher for seminars in mathematics and statistics for students of biochemistry and biology
Occasionally since 2010	Private tutor for chemistry, biology and mathematics for pupils from middle and high school

Publications:

Dienemann C, Bøggild A, Winther KS, Gerdes K, Brodersen DE. *Crystal structure of the VapBC toxin-antitoxin complex from Shigella flexneri reveals a hetero-octameric DNA-binding assembly.* J Mol Biol. 2011 Dec 16;414(5):713-22.

Xu K, Dedic E, Cob-Cantal P, **Dienemann C**, Bøggild A, Winther KS, Gerdes K, Brodersen DE. *Protein expression, crystallization, and preliminary X-ray crystallographic analysis of the isolated Shigella flexneri VapC toxin.* Acta Cryst F. 2013 Jul;69(Pt 7):762-5.

Kourosh HE, **Dienemann C**, Hoefgen S, Than ME, Hagedoorn P, Wilfred HR. *The amyloid precursor protein (APP) does not have a ferroxidase site in its E2 domain.* PLoS One. 2013 Aug 19;8(8):e72177.

Dienemann C, Coburger I, Mehmedbasic A, Andersen OM, Than ME. *Mutants of metal binding site M1 in APP E2 show metal specific differences in binding of heparin but not of sorLA.*Biochemistry. 2015 Apr 21;54(15):2490-9.

Plaschka C, Hantsche M, **Dienemann C**, Burzinski C, Plitzko J, Cramer P. *Transcription initiation complex structures elucidate DNA opening.* Nature. 2016 May 19;533(7603):353-8.

Schilbach S, Hantsche M, Tegunov D, **Dienemann C**, Wigge C, Urlaub H, Cramer P. *Structures of transcription pre-initiation complex with TFIIH and Mediator.* Nature. 2017 Nov 9;551(7679):204-209.