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

RNA polymerase I complex structures elucidate mechanisms of transcription initiation and elongation

Simon Neyer

aus

Münster, Deutschland

2016

Erklärung

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

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den 29. November 2016

Simon Neyer

Dissertation eingereicht am 01.12.2016
1. Gutachter: Prof. Dr. Patrick Cramer
2. Gutachter: PD Dr. Dietmar Martin
Mündliche Prüfung am 25.01.2017

Acknowledgement

I am very grateful to Patrick Cramer for giving me the opportunity to work in his lab, in which he provides an excellent research environment. His positive spirit and his dedication to open questions are outstanding. Through him, I had the chance to learn about many aspects of science, which even go far beyond our everyday work. Thank you, Patrick!

Special thanks go to my collaborators in Frankfurt. It was a pleasure to work with Achilleas Frangakis and his team on the elongation studies, in particular with Michael Kunz, Christian Geiss, Valentin Hodirnau, Anja Seybert and Margot Scheffer. I truly enjoyed our fruitful and open discussions. Thank you!
I further thank Tobias Gubbey and Christoph Engel for our successful collaboration that shed light on the mechanisms of transcription initiation.

Being supported by the Boehringer Ingelheim Fonds with a PhD student scholarship was an honour for me. The accompanying events, organized mainly by Claudia Walther, Sandra Schedler and Anja Hoffmann, were always highly inspiring and a great opportunity to get in touch with other BIF fellows. Thank you!

I am grateful to all past and present members of the Cramer lab, in particular to Hauke, Carrie, Felix, Tobi, Merle, Youwei, Christian, Isaac, Dimitry, Goran, Clemens, Marc, Sarah, Sara, Paulina, Sandra, Thomas, Katharina, Ania, Kristina, Björn, Wolfgang, Seychelle, Lukas, Christiane, Geli, Manu, Kirsten, Janine and Kerstin (from Göttingen); and to Kathrin, Rieke, Stefan, Jürgen, Alan, Kerstin, Fuensanta, Dietmar and Laurent (from Munich). Thank you for scientific discussions, advice and help; and also for personal exchange, coffee breaks and common leisure time activities!

I further want to take the opportunity and give credit to those who trained me before my time as a PhD student. Thank goes to Tilmann Kottke and Anna Pfeifer, who supervised me during my Bachelor thesis, and to Jenny Martin and Fabian Kurth, who supervised me during my Master thesis. Furthermore, I am grateful to Christoph Engel for excellent supervision during an undergraduate internship in the Cramer lab. All of you prepared me for my start as a PhD student. Thank you!

I thank Carrie, Merle and Hauke for critically reading this report. I further thank my examination board members: Patrick Cramer, Dietmar Martin, Achilleas Frangakis, Franz Herzog, Karl-Peter Hopfner and Roland Beckmann. Thank you for your interest in my work!

Finally, I thank my parents, my brother and my grandparents for their unconditional love and continuous support!

Summary

RNA polymerase (Pol) I, Pol II and Pol III carry out transcription in eukaryotes, a process that is essential to all cell. Ribosomal RNA synthesis by Pol I accounts for the majority of transcription products, which highlights its biological importance. During initiation Pol I is recruited to the promoter with the assistance of the general transcription factors Rrn3 and core factor (CF), a trimeric complex consisting of Rrn6, Rrn7 and Rrn11. Well-defined initiation as well as processive and accurate Pol I elongation are equally important for cell viability. However, the mechanisms of both processes are poorly understood.

Here, I report a 3.8 Å resolution cryo-electron microscopy structure of elongating *S. cerevisiae* Pol I on a 39 nucleotide DNA scaffold that embeds an 11 nucleotide mismatch bubble with a 8 nucleotide RNA-DNA hybrid. The structure reveals details of active Pol I, such as a completely helical 'bridge helix' and repulsion of the RNA cleavage stimulating domain of A12.2 from the active centre. The findings suggest a coupling of cleft states with activity. While an expanded cleft indicates inactive Pol I, it contracts upon binding of an RNA-DNA hybrid. In collaboration with the Frangakis laboratory, we further analysed Pol I on Miller spreads under *ex vivo* conditions. Cryo-electron tomography and subtomogram averaging confirmed the contracted cleft state under near physiological conditions and further allowed analysing the relative orientation of Pols.

Additionally, I report a 4.2 Å resolution cryo-electron microscopy structure of initially transcribing yeast Pol I bound to initiation factors. Collaborating with Tobias Gubbey, who solved the CF crystal structure, allowed me to generate a pseudo-atomic model. In collaboration with Christoph Engel, we further found a second CF binding interface, which blocks the Pol I cleft for DNA loading. Taken together, we show how upstream promoter DNA is bound by an interplay of Pol I and CF, which together introduce a 30° DNA bend. In addition to the promoter associated domain, CF contains another module that docks to the Pol I - Rrn3 complex. Furthermore, I modelled a closed complex, which shows that DNA is loaded on Pol I in an expanded cleft state between protrusion and clamp domains. For initial transcription the cleft then contracts and thereby allows RNA synthesis. Based on comparison with the Pol II system we suggest that the mechanism of promoter recognition in Pol I is unique among eukaryotic polymerases, relying on DNA bendability and meltability rather than on sequence recognition.

Altogether, my studies expand our mechanistic understanding of Pol I transcription for both regulatory important steps, initiation and elongation.

Publications

Part of this work has been published or is currently in the process of being published.

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

* These authors contributed equally to this work.

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

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

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

C. Engel*, T. Gubbey*, **S. Neyer***, S. J. Sainsbury, C. Oberthür, C. Bäjen, C. Bernecky, P. Cramer. (2016) Structural basis of RNA polymerase I transcription initiation. (manuscript in preparation)

* These authors contributed equally to this work.

Author contributions: CE initiated the project, planned and carried out EM structure determination of the Pol-I-Rrn3-Core-Factor complex, and planned and carried out functional assays. TG crystallized Core Factor (CF) and collected diffraction data. TG, CE and SJS analyzed crystallographic data and solved the CF crystal structure. SN planned and carried out EM structure determination of the initially transcribing complex, and contributed to functional assay design. CO, TG, SN and CE cloned, expressed and purified proteins. CBä assisted with functional assays. CBe advised on EM. PC designed, supervised and coordinated research. CE, TG, SN and PC prepared the manuscript with input from all authors.

Contents

1. Introduction

1.1. Transcription

The genetic information of all organisms, from bacteria to eukaryotes, is encoded in their DNA. While every single cell in a multicellular organism has the same genes, the subset of genes that are transcribed in a specific cell type differs. Thus, transcription plays an important role determining the cell's phenotype. Assembly of an RNA chain from individual ribonucleotides, is catalysed by DNA-dependent RNA polymerases (Pols) which, due to their fundamental function, are essential in all three kingdoms of life but differ in complexity (Werner and Grohmann, 2011).

Bacteriophages, like T7, encode for single subunit, 'right-handed' polymerases. Complexity is increased in bacteria and archea, which each possess a single type of multisubunit Pol. While Pols in eukaryotes are built from a related structural framework, these higher organisms encode for three different polymerases (Pol I, Pol II and Pol III). In plants, the additional Pol IV and Pol V predominantly transcribe non-coding RNAs. All nuclear Pol differ in function. Pol I, II and III synthesize mainly ribosomal, messenger and transfer RNA, respectively. According to the central dogma of biology, messenger RNA (mRNA) is translated to proteins by ribosomes. The main component of these ribonucleoprotein complexes is ribosomal RNA (rRNA). In yeast, the 35S precursor is transcribed by Pol I and later spliced into 28S, 18S and 5.S rRNA. During translation, rRNA functions catalytically as well as structurally.

All Pols catalyse the same chemical reaction. The release of pyrophosphate is the driving force to form a phosphodiester bond between the 3' hydroxyl group of the growing RNA chain with the incoming ribonucleotide. Consequently, the catalytic mechanism of all eukaryotic polymerases is very similar. However, they have diverged in order to be specialists for their specific tasks. This is particularly true when it comes to regulation of transcription and is also reflected in the subunit composition of the enzymes. Pol I consists of 14 subunits with a total molecular weight of 590 kDa in

S. cerevisiae (Table 1.1). The core is conserved between Pol II and III and comprises ten subunits. The five subunits Rpb5, Rpb6, Rpb8, Rpb10 and Rpb12 are common to all three polymerases. The two largest subunits A190 and A135 are unique to Pol I but are homologous to the Pol II subunits Rpb1 and Rpb2, respectively. AC40 and AC19 are shared between Pol I and Pol III and are homologous to Rpb3 and Rpb11, respectively. The N-terminal domain of A12.2 shows similarity to Rpb9 and the stalk subunits A14 and A43 can be compared to Rpb4 and Rpb7 of Pol II. The two specific subunits of Pol I, A49 and A34.5, share features of the general Pol II transcription factors TFIIF and TFIIE (Eichner et al., 2010; Vannini and Cramer, 2012).

Table 1.1. Eukaryotic (Pol I, II and III) and bacterial RNA polymerase subunits. The molecular weight of *S. cerevisiae* Pol I (in total 590 kDa) is stated in a separate column. Table adopted from Werner and Grohmann (2011).

Pol I	Pol II	Pol III	Bacteria	Pol I [kDa]
A190	Rpb1	C160	β subunit	186
A135	Rpb2	C128	β subunit	136
AC40	Rpb3	AC40	α subunit	38
AC19	Rpb11	AC19	α subunit	16
Rpb6	Rpb6	Rpb6	ω subunit	18
Rpb5	Rpb5	Rpb5		25
Rpb8	Rpb8	Rpb8		17
Rpb10	Rpb10	Rpb10		8
Rpb12	Rpb12	Rpb12		8
A14	Rpb4	C17		15
A43	Rpb7	C25		36
A12	Rpb9	C11		14
A49		C53		47
A34.5		C37		27

1.2. Organization of rDNA genes

As Pol I transcription accounts for about half of the overall eukaryotic transcription (Moss, 2004; Russell and Zomerdijk, 2005), the rDNA genes have been studied extensively. In all eukaryotes, rDNA is arranged in repeats with a conserved layout across species (Moss et al., 2007). Since most repeats are inactive under normal growth conditions, controlling the fraction of active rDNA repeats seems plausible to regulate Pol I transcription (Conconi et al., 1989). Furthermore, genome stability is suggested as a primary function of inactive rDNA repeats (Ide et al., 2010).

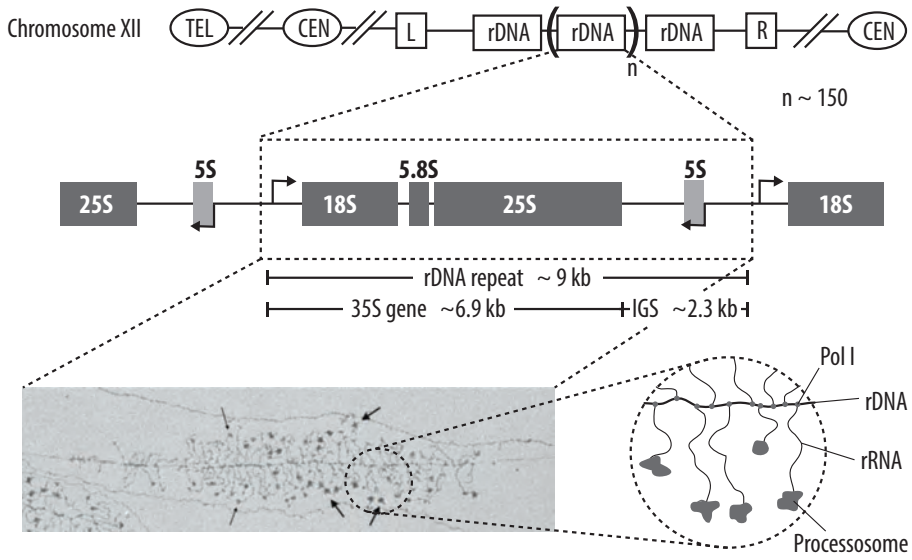


Figure 1.1. Organisation of rDNA genes. *S. cerevisiae* rDNA repeats are localised on chromosome XII and consist of the 35S gene and the intergenic spacer (IGS). The Miller spread shown was taken from Osheim et al. (2009) and the zoom-in circle visualizes Pol I together with its rDNA template and rRNA product, which is further processed within the processosome.

The genome of yeast cells contains about 150 rDNA repeats, which are embedded with upstream and downstream flanking regions (Hamperl et al., 2013) (Figure 1.1). Each repeat is built up by the 35S gene and the intergenic spacer (IGS). The IGS also hosts the 5S gene, which is transcribed by Pol III in antisense direction to the 35S gene. The high loading rate of rDNA genes is visualized in 'Miller spreads' of nucleolar chromatin (Miller and Beatty, 1969). As Pol I travels along the gene, its RNA product is extended, and the nascent rRNA is further compacted and thereby forms the 'processosome'. This terminal knob is highly dynamic and processes the 35S rRNA precursor co-transcriptionally into 18S, 25S and 5.S rRNA (Osheim et al., 2004).

1.3. Transcription cycle

The transcription cycle of all Pols involves three main stages, initiation, elongation and termination. During initiation, the polymerase has to be recruited to the promoter and the DNA duplex has to unwind. After leaving the promoter, the polymerase elongates along the gene until it reaches a termination signal.

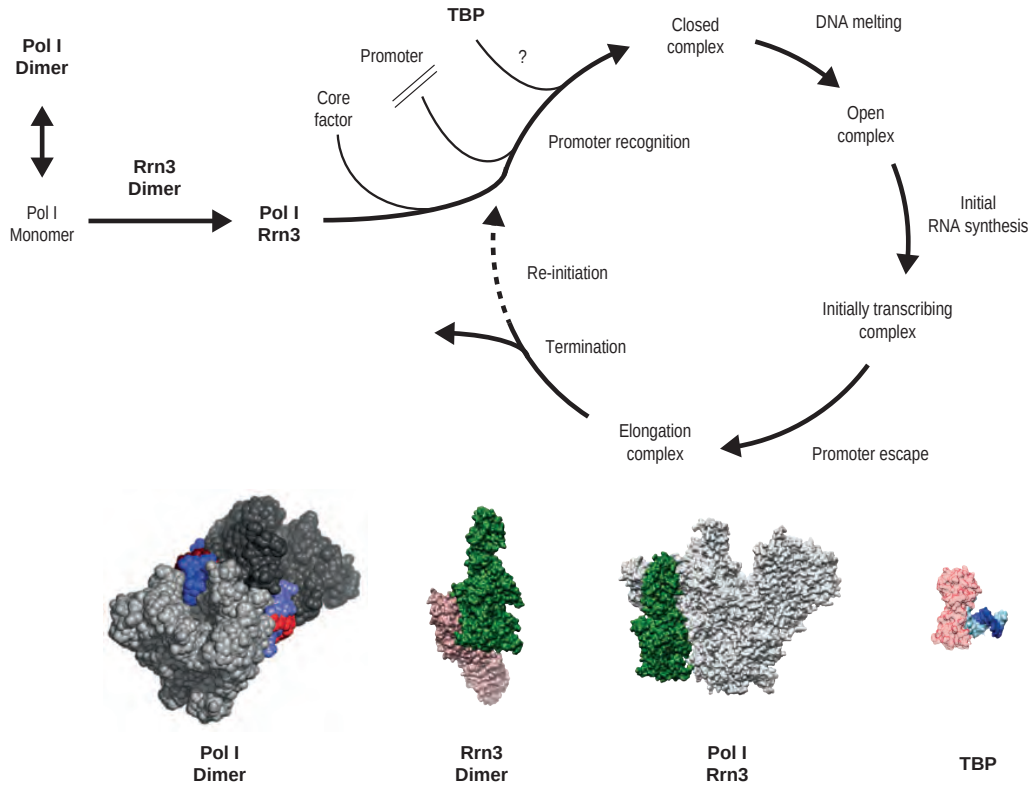


Figure 1.2. Transcription cycle. Scheme of the Pol I transcription cycle. States with known structures are marked with bold font. The image of dimeric Pol I was adopted from Engel et al. (2013)) and the Rrn3 crystal structure also revealed dimers (Blattner et al., 2011). These two homo-dimers form an initiation competent Pol I-Rrn3 hetero-dimer and this complex was solved by cryo-EM (Engel et al., 2016; Pilsl et al., 2016). Apart from Rrn3, CF is also involved in promoter recognition. The role of TBP in the Pol I system is unclear, but numerous crystal structures were solved and showed that TBP induces a bend into DNA (Gietl and Grohmann, 2013). Subsequent to promoter recognition, the transition to an open complex occurs via DNA melting. Then, the initially transcribing complex escapes the promoter and productive elongation occurs followed by either re-initiation or termination.

During the first step of the transcription cycle, Pol I is recruited to the promoter. As Pol I can dimerize into an inactive state it has to be rendered initiation competent by its recruiting factor Rrn3 (Engel et al., 2016, 2013; Fernández-Tornero et al., 2013; Pilsl et al., 2016). According to our current understanding, the hetero-dimer of Pol I and Rrn3 is further bound by the trimeric core factor (CF). While this system resembles the minimal, but specific, initiation system, the upstream activating factor (UAF), is also involved in initiation (Figure 1.3). It cooperatively supports DNA binding and enhances initiation activity. Furthermore, UAF is supposed to play a role in determining the rDNA gene promoter specificity for Pol I, but prevents Pol II and Pol III transcription at the same time. The role of TBP is currently discussed. While, TBP was found to be part of the PIC and to stimulate transcription (Aprikian et al., 2000; Steffan et al., 1998), it was also shown not to be needed for Pol I activity (Keener et al., 1998).

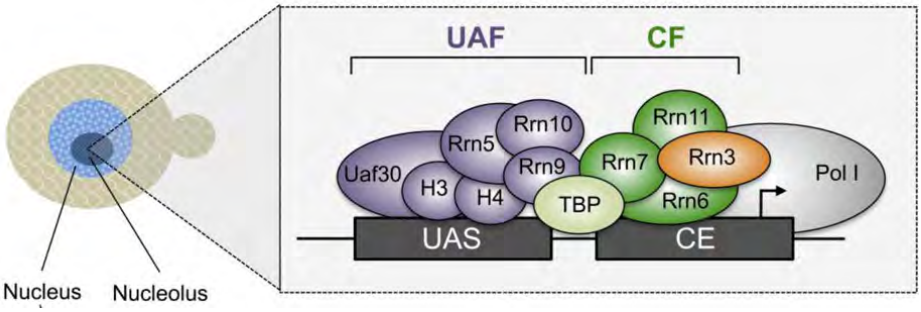


Figure 1.3. rDNA gene promoter. In addition to Pol I, the initiation factors Rrn3 and the core factor (CF) bind the promoter core element (CE) in the yeast nucleolus. The CE also harbours the transcription start site (TSS) indicated by an arrow. The upstream activating sequence (UAS) is recognised by the upstream activating factor (UAF) and both elements are bridged by TBP, although the Pol I promoter is TATA-less. The CE ranges from nucleotides -38 to +5 and the UAS from -60 to -155, with +1 indicating the TSS (Meier and Thoma, 2005). Figure adopted from Knutson and Hahn (2013).

Following successful recruitment of Pol I to the promoter, promoter opening occurs and the first ribonucleotide is incorporated, which actually initiates transcription. While Pol I moves from 3' to 5' along the template DNA, elongation proceeds in 5' to 3' direction. When the newly synthesized RNA reaches a length of 10 nucleotides, promoter clearance occurs. This process involves the removal of initiation factors and was shown to be rate limiting (Panov et al., 2001).

After promoter clearance, Pol I enters the phase of transcription elongation, which is characterized by a very dense loading of Pols on the rDNA genes at the same time. The high number of Pols on a gene was visualized in 'Miller spreads' (Miller and

Beatty, 1969). Incorporation of incorrect nucleotides can result in Pol I pausing. To rescue paused Pol, it has to backtrack and to cleave off the wrongly incorporated nucleotides from the RNA 3'-end. While Pol II relies on the cleavage factor TFIIS, Pol I has the built-in subunit A12.2, which can render the active site of the enzyme cleavage competent and thereby enables efficient elongation.

Upon termination, the elongation complex is destabilized and subsequently DNA, RNA and polymerase dissociate from each other. The Pol I termination sequence includes a 15 bp long T-rich sequence upstream of a Reb1-binding site. While this sequence together with Reb1 is sufficient to terminate Pol I *in vitro*, the process is much more complex *in vivo* and a coupling of rRNA processing as well as transcription re-initiation is discussed (Nemeth et al., 2013).

1.4. Structural studies of the RNA polymerase I system

1.4.1. RNA polymerase I crystal structure

First structural insights into the architecture of Pol were gained by electron microscopy of 2D crystals (Schultz et al., 1993), but high resolution information remained elusive for more than two decades. A working model about the position of the Pol I specific subunits was derived from negative stain immunolabelling (Bischler et al., 2002). Cryo-EM data gave rise to a model of the functional architecture of the complete enzyme (Kuhn et al., 2007), and the stalk as well as dimerisation domain sub-complexes were characterized individually (Geiger et al., 2010; Kuhn et al., 2007). Finally, well-diffracting crystals of the complete enzyme were obtained and the structure of *S. cerevisiae* Pol I was determined at 2.8 Å resolution (Engel et al., 2013; Fernández-Tornero et al., 2013). These studies not only gave insights into the structure of Pol I specific features, but also in the regulation of rDNA transcription (Figure 1.4).

Pol I was observed in a dimeric state, in which the stalk of one enzyme invades the cleft of the other and its 'connector' stabilizes the dimer. DNA binding is not possible in this conformation for three reasons and thus, the polymerase was trapped in an inactive state. Firstly, the adjacent Pol I blocks the DNA path. Secondly, the cleft is expanded and too wide to properly hold DNA and thirdly, the 'expander' is localized close to the active site and would clash with a potential DNA-RNA hybrid. Furthermore, the

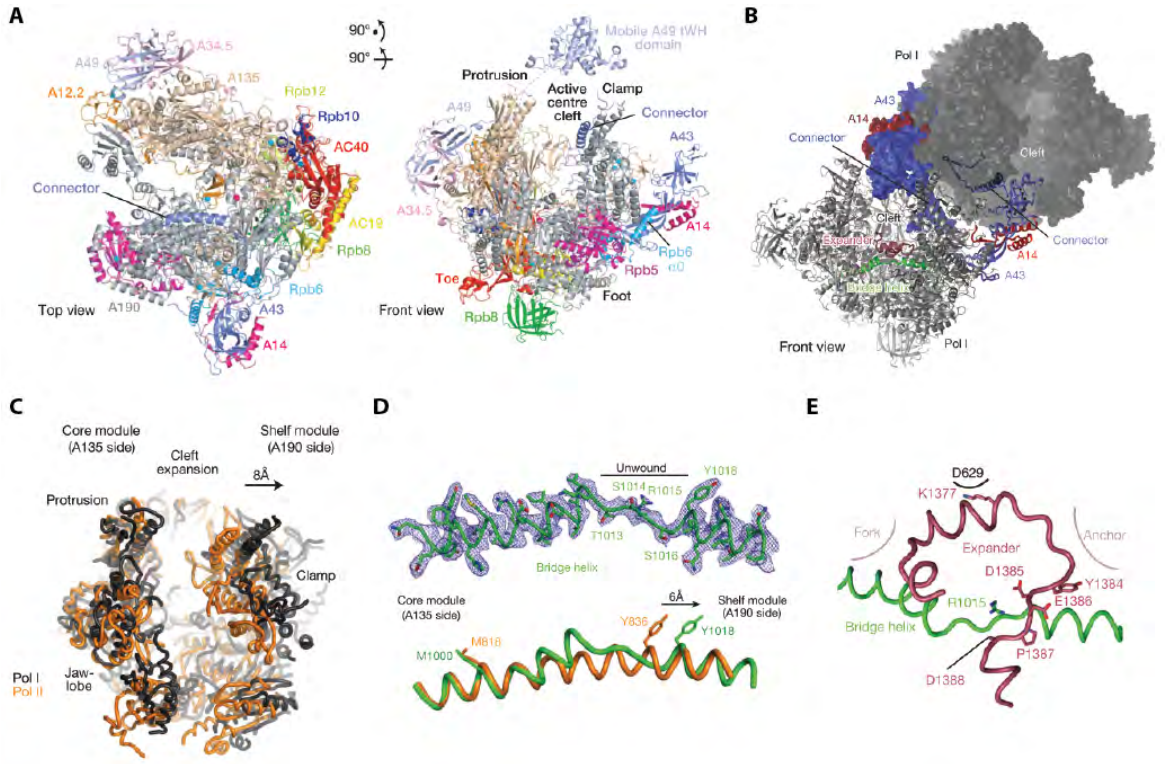


Figure 1.4. RNA polymerase I crystal structure. (A) Pol I ribbon model. All domains were resolved in the crystal structure except for the mobile tandem winged helix domain of A49. (B) Pol I dimerization is visualized with one enzyme shown in ribbon and the other in surface representation. (C) Cleft expansion of Pol I (black) compared to Pol II (orange). The models were superimposed on the second largest subunits. (D) Bridge helix superimposed with electron density mesh (upper panel). Ribbon model of Pol II (orange) and Pol I (green) bridge helix (bottom panel). (E) The Expander (red) in the Pol I crystal structure is next to the bridge helix and would clash with a DNA-RNA hybrid. Figure adopted from Engel et al. (2013).

bridge helix adopts an expanded conformation and changes in the bridge helix are generally linked to Pol functionality (Jovanovic et al., 2011).

While Pol I possess built-in proofreading activity, Pol II relies on the additional factor TFIIS (Kettenberger et al., 2003). The C-terminal domain of A12.2 fulfils this role, rendering the composite active site of Pol I active for cleavage and was indeed observed at a similar position as TFIIS with its catalytic loop reaching into the active site of the enzyme.

1.4.2. Rrn3 crystal structure

The initiation factor Rrn3 is involved in the recruitment of Pol I to the promoter and its interaction with Pol I is coupled to cell growth regulation (Grummt and Voit, 2010). Certain Rrn3 phosphorylations impair Pol I binding and reduce transcriptional activity (Mayer et al., 2005). Structure-function analysis of Rrn3 gave insights into Rrn3-regulated Pol I transcription (Blattner et al., 2011) (Figure 1.5).

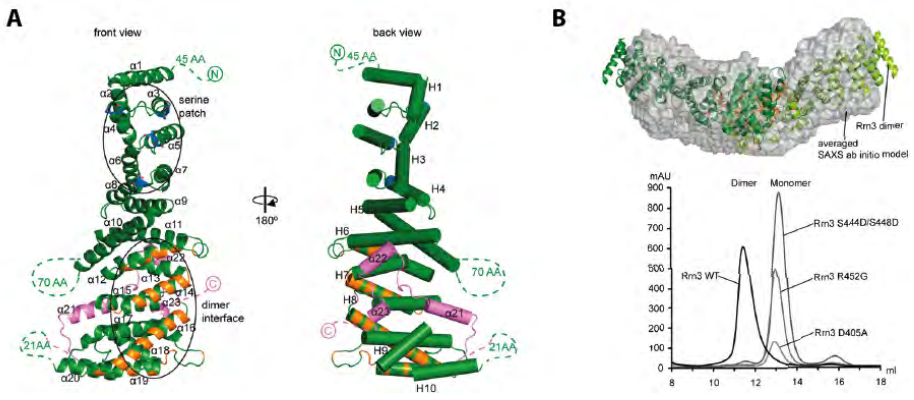


Figure 1.5. Crystal structure of Rrn3. (A) Front (left panel) and back (right panel) view of Rrn3. Each HEAT repeat (H1 to H10) consists of two anti-parallel α-helices. The N-terminal serine patch is highlighted in blue and stretches over helices α3, α4, α5 and α8. The C-terminal dimerization interface is coloured in orange. (B) SAXS confirms Rrn3 dimerization in solution (upper panel). The SAXS envelope is shown as grey semi-transparent surface and the Rrn3 dimer from the crystal structure is depicted as ribbon model. Point mutations in the dimer interface disrupt Rrn3 dimerization and consequently, these mutations are eluted at later volumes in size-exclusion chromatography experiments (bottom panel). Figure adopted from Blattner et al. (2011).

Rrn3 is a mainly helical protein of 23 α-helices, of which 20 are arranged in an anti-parallel manner forming HEAT repeats. HEAT repeats are commonly involved in protein-protein interactions and indeed, an N-terminal serine patch of Rrn3 binds

to Pol I. Phospho-mimetic mutations in this patch negatively influence binding and further impair Pol I promoter recruitment. While Rrn3 and Pol I bind each other in a 1:1 ratio, the crystals revealed dimeric Rrn3. Next to the N-terminal serine patch, Rrn3 contains an additional C-terminal protein interaction domain, the dimerisation interface. SAXS analysis confirmed that Rrn3 dimerizes not only under crystallization conditions, but also in solution and dimerization can further be disrupted by specific point mutations. A model of the Pol I-Rrn3 complex, in which Rrn3 lies on the back of the enzyme in close proximity of the RNA exit channel and the stalk, was hypothesized on the basis of lysine-lysine cross-linking.

1.4.3. RNA polymerase I - Rrn3 complex structure

While the crystallographic studies of Pol I and Rrn3 already lead to models of the Pol I-Rrn3 complex (Blattner et al., 2011; Engel et al., 2013), these models were further refined by cryo-EM analysis (Engel et al., 2016; Pilsl et al., 2016). Going from the N-terminal to the C-terminal Rrn3 helices, the N-terminal HEAT repeats form the serine patch, which binds to the stalk of Pol I. A Pol I specific clamp insertion further interacts with Rrn3. The binding interface is completed with the dock domain and two loops of the Pol I subunits AC19 and AC40. Upon Rrn3 binding, the expanded Pol I cleft also contracts partially.

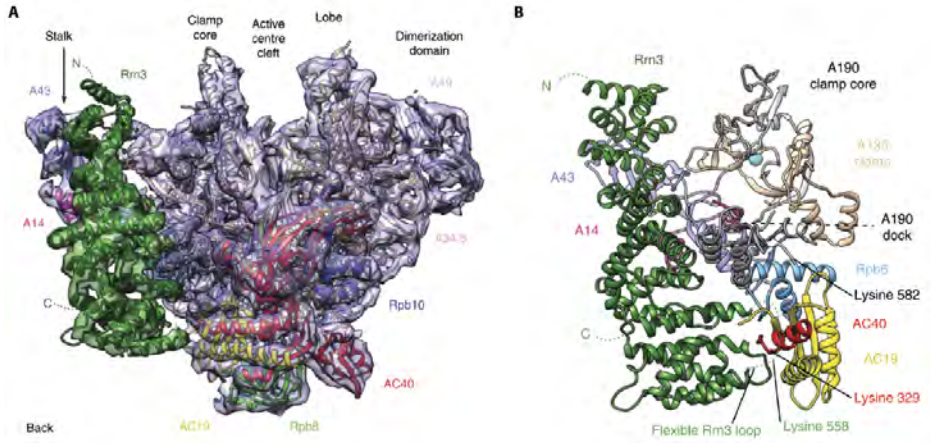


Figure 1.6. Structure of Rrn3 in complex with Pol I. (A) Ribbon model of Pol I-Rrn3 together with the EM density superimposed. Rrn3 is shown in green. (B) Protein interaction interface between Pol I and Rrn3. Figure adopted from Engel et al. (2016).

1.4.4. Core factor architecture

The core initiation factor 'CF' of Pol I consists of Rrn6, Rrn7 and Rrn11 and comprise a total molecular mass of 220 kDa. CF domains were predicted computationally and a topological model was derived from cross-linking coupled to mass spectrometry (Knutson et al., 2014) (Figure 1.7). Rrn7 is homologous to the Pol II transcription factor TFIIB and harbours a zinc ribbon, two cyclins and a C-terminal domain. Central tetratricopeptide repeats (TPRs) characterize Rrn11 and the biggest subunit Rrn6 consists of a WD40 β-propeller followed by a helical domain. Strongest contacts are made between the β-propeller and the TPRs.

Mutational studies *in vivo* and *in vitro* underline the functional relevance of CF (Knutson et al., 2014). The elements within which mutations are lethal, but retain CF integrity, will most likely be involved in interactions with Pol I or might play a role in promoter recognition. For Rrn7 this is the case for its C-terminal domain including the zinc ribbon. In the case of Rrn11 its very N-terminal part together with the C-terminal TPR elements fulfil the same criteria.

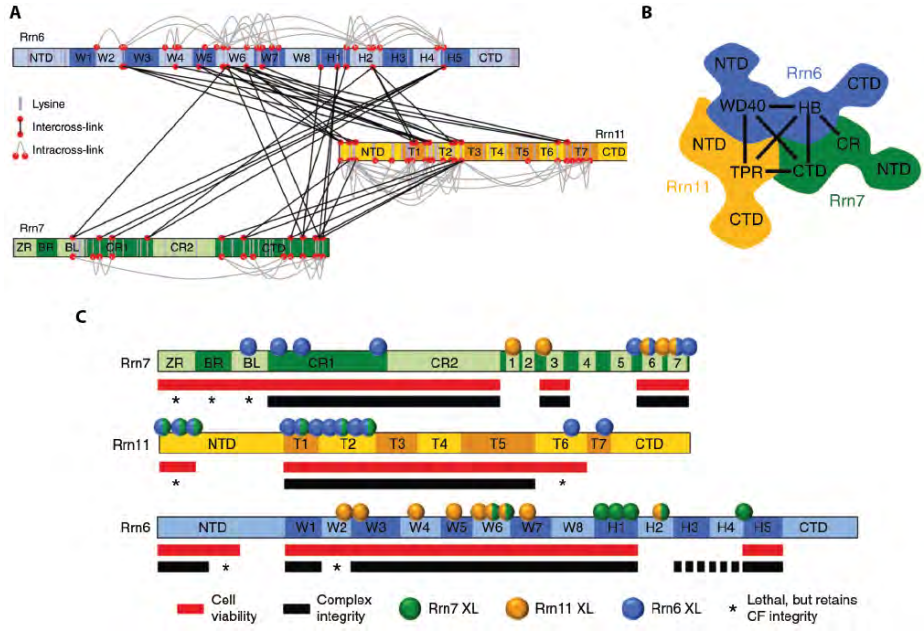


Figure 1.7. Core factor cross-linking coupled to mass spectrometry analysis. (A) CF distant restraints. (B) CF topology with the central Rrn11 TPR, Rrn6 WD40, Rrn 6 helical and Rrn7 C-terminal domains. (C) Functional CF analysis showing essential CF domains and those which are needed for CF integrity. All figures were adopted from Knutson et al. (2014).

1.5. Aims and scope

The groundbreaking crystal structure of Pol I (Engel et al., 2013; Fernández-Tornero et al., 2013) gave many insights into Pol I specific features such as the expanded cleft and the role of A12.2 as well as the novel 'expander' and 'connector' elements. As the enzyme was trapped in the inactive state of a dimer, this further raised the question how the transition to an active enzyme occurs.

Here, I determined the elongation complex to a resolution of 3.8 Å that revealed a contracted active centre cleft with bound DNA and RNA, and a narrowed pore beneath the active site that no longer holds the RNA-cleavage stimulating domain of A12.2. To investigate active Pol I under more native conditions on rDNA genes, we teamed up with the Frangakis group. In this collaboration, we not only confirmed the cleft contraction with Miller spread tomography, but also defined the relative positions of adjacent Pols.

The Pol I crystal structure provided the basis for further structural investigations on transcription regulation (Engel et al., 2016; Fernández-Tornero et al., 2013). The very first step of transcription, initiation, is most critical for regulation and thus of particular interest. How do the initiation factors Rrn3 and CF interact with Pol I, how does the transcription machinery bind to promoter DNA, and how is the start site for RNA synthesis localized?

To answer these questions, I determined the structure of the initially transcribing complex (ITC), which contains Pol I, Rrn3 and CF, as well as unwound promoter DNA and a short RNA-DNA hybrid to 4.2 Å resolution. As no high resolution information was available for the CF and *de novo* model building into the EM map was not feasible due to high CF flexibility, collaborating with Tobias Gubbey was essential. He solved the crystal structure of CF, which I then used to interpret cryo-EM maps. In collaboration with Christoph Engel we further extended our knowledge of the pre-initiation complex (PIC) in absence of promoter DNA. Furthermore, I modelled a closed and open complex. Finally, we showed how CF binds upstream promoter DNA, docks to the initiation-competent Pol I-Rrn3 complex, and loads DNA over the polymerase wall into an expanded active centre cleft. DNA opening between the protrusion and clamp domains is then coupled with contraction of the Pol I cleft resulting in an active enzyme conformation and RNA synthesis.

2. Results and Discussion

2.1. Structure of RNA polymerase I transcribing rDNA genes

Results presented in this section are published and were obtained in a collaboration with the Frangakis lab. Author contributions are stated on page VI.

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

* These authors contributed equally to this work.

While focusing on the single particle cryo-EM analysis, the following section gives a comprehensive picture of all results presented in this study. Methods are described in section 3.1. Supplemental data and methods that were predominantly obtained by the collaborators are described in the appendix B and D, respectively.

2.1.1. Abstract

RNA polymerase I (Pol I) is a highly processive enzyme that transcribes ribosomal DNA (rDNA) and regulates growth of eukaryotic cells (Goodfellow and Zomerdijk, 2013; Grummt, 2013; Novello and Stirpe, 1970; Roeder and Rutter, 1969). Crystal structures of free Pol I from the yeast *Saccharomyces cerevisiae* revealed dimers of the enzyme stabilized by a 'connector' element and an expanded cleft containing the active centre in an inactive conformation (Engel et al., 2016; Fernández-Tornero et al., 2013). The central bridge helix was unfolded and a Pol I-specific 'expander' element occupied the DNA-template-binding site. The structure of Pol I in its active transcribing conformation is yet to be determined, whereas structures of Pol II and Pol III have been solved with bound DNA template and RNA transcript (Gnatt et al., 2001; Hoffmann

et al., 2015; Kettenberger et al., 2004). Here we report structures of active transcribing Pol I from yeast solved by two different cryo-electron microscopy approaches. A single-particle structure at 3.8 Å resolution reveals a contracted active centre cleft with bound DNA and RNA, and a narrowed pore beneath the active site that no longer holds the RNA cleavage-stimulating domain of subunit A12.2. A structure at 29 Å resolution that was determined from cryo-electron tomograms of Pol I enzymes transcribing cellular rDNA confirms contraction of the cleft and reveals that incoming and exiting rDNA enclose an angle of about approximately 150°. The structures suggest a model for the regulation of transcription elongation in which contracted and expanded polymerase conformations are associated with active and inactive states, respectively.

2.1.2. Single particle cryo-EM

To determine the structure of transcribing Pol I, we performed single-particle cryo-electron microscopy (cryo-SP) with a reconstituted yeast Pol I elongation complex (EC) containing a DNA-RNA scaffold (Figure 2.2A and 2.1) similar to the one used to study transcribing mammalian Pol II (Bernecky et al., 2016). Particle classification enabled us to reconstitute the Pol I EC structure at 3.8 Å resolution from approximately 94,000 single particles (Figure 2.2C, 2.3 and 2.4). The electron density revealed the downstream DNA, the DNA-RNA hybrid (Figure 2.2B), and all Pol I domains except for the flexibly linked C-terminal domain of subunit A49 (Jennebach et al., 2012; Pilsl et al., 2016) and the C-terminal domain of subunit A12.2. An atomic model with excellent stereochemistry was obtained by fitting rigid domains of the Pol I crystal structure (Engel et al., 2013), positioning nucleic acids from the bovine Pol II EC structure (Bernecky et al., 2016), and manually rebuilding regions that were structurally altered (Table 2.1).

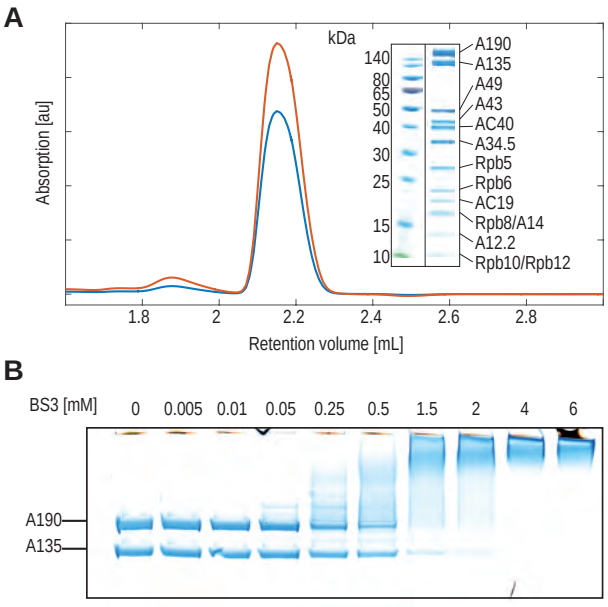


Figure 2.1. Preparation of Pol I elongation complex (EC) for cryo-SP. (A) Size-exclusion chromatogram (Superose 6 Increase 3.2/300; GE Healthcare) of reconstituted Pol I EC. Higher absorbance at 260 nm (red line) than at 280 nm (blue line) indicates presence of nucleic acids. Coomassie-stained SDS-PAGE analysis of pooled peak fractions shows the presence of all 12 Pol I subunits. **(B)** Coomassie-stained SDS-PAGE analysis of titration with BS3 cross-linker. Gel is cropped to large subunits A190 and A135. A shift to higher molecular weight is observed with increasing BS3 concentration indicating successful crosslinking. 0.9 mM BS3 was chosen for final sample preparation.

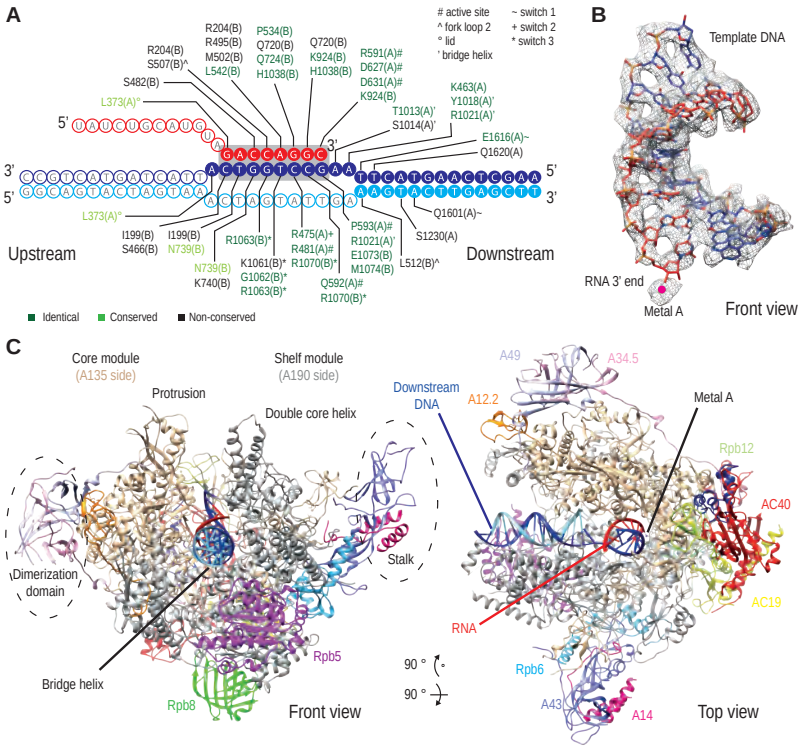


Figure 2.2. Cryo-SP structure of yeast Pol I EC at 3.8 Å resolution. (A) Nucleic acid scaffold and Pol I-nucleic acid interactions. Template DNA, non-template DNA and RNA are shown in blue, cyan, and red, respectively. Filled circles represent nucleotides that were well resolved in the cryo-EM density. Pol I residues within 4 Å distance are depicted together with the subunit identifier (A for A190, B for A135). The active site metal ion A is depicted as a magenta sphere. **(B)** Electron density for the DNA-RNA hybrid with the final model superimposed. **(C)** Ribbon model of the Pol I EC. The view is from the 'front' (Cramer et al., 2001) with the incoming downstream DNA pointing towards the reader. Subunit colour code used throughout.

Comparison of the resulting Pol I EC structure with the previous Pol I structure (Engel et al., 2013; Fernández-Tornero et al., 2013) reveals that the active centre cleft is contracted by up to 13 Å (Fig. 2.5A). Contraction occurs via relative movement of the two major polymerase modules 'core' and 'shelf' (Cramer et al., 2001) as predicted (Engel et al., 2013). The shelf module moves together with the clamp domain as a single 'shelf-clamp' unit, slightly rotating with respect to the core module (Fig. 2.5E). Another module, the 'jaw-lobe', moves closer to downstream DNA by up to 7 Å (Fig. 2.5A and D). Comparison of the Pol I EC with EC structures of Pol II (Gnatt et al., 2001; Kettenberger et al., 2004), Pol III (Hoffmann et al., 2015), and bacterial RNA polymerase (Vassylyev et al., 2007) reveal that all these polymerases adopt a similar contracted conformation in their transcribing state.

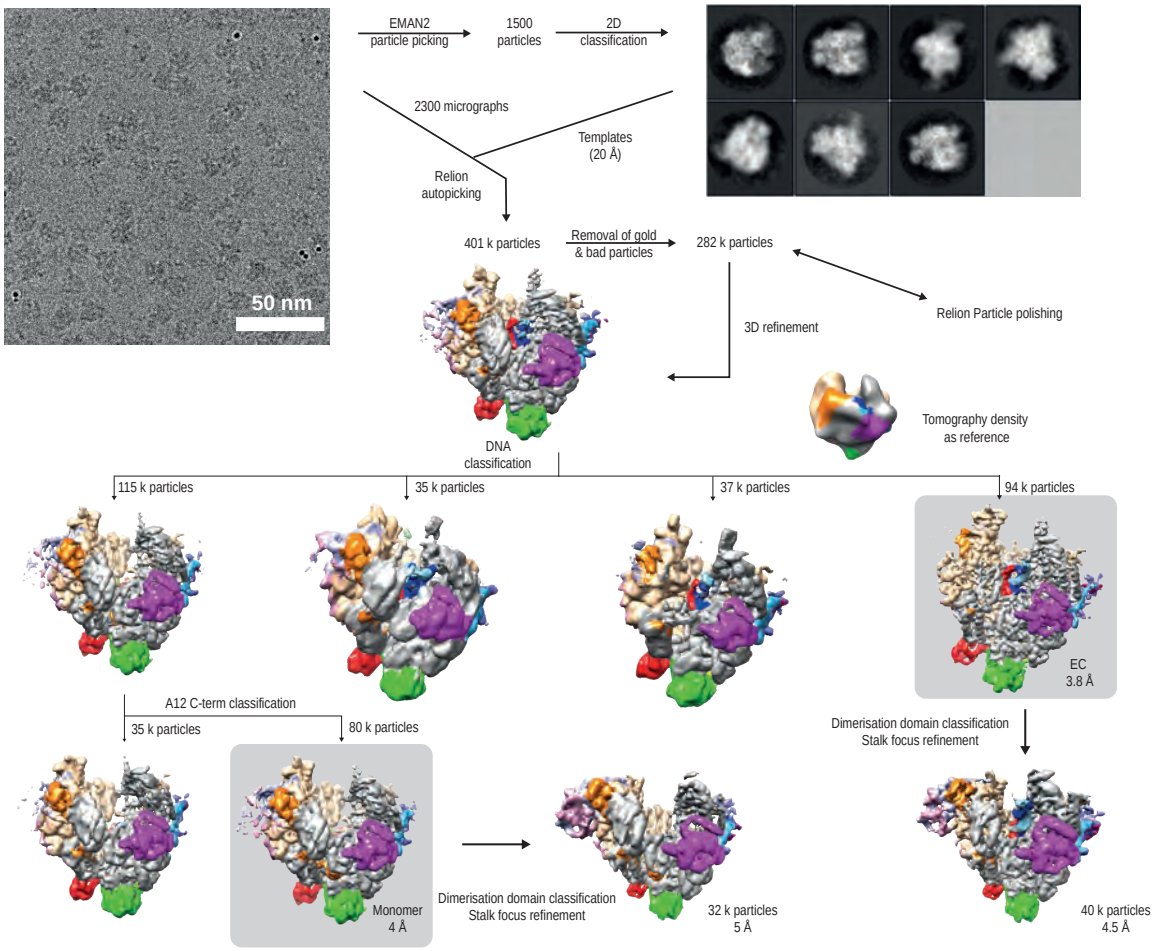


Figure 2.3. Cryo-SP particle sorting pipeline. Annotated arrows indicate the direction of processing and provide information regarding the number of particles used and the classification masks applied. A representative micrograph of the PolI EC under cryo conditions showed particles of the expected size. A set of 1500 particles was picked manually with EMAN2 (Tang et al., 2007) and used to generate initial 2D classes for template based auto-picking in Relion (Scheres, 2012). After cleaning by manual inspection and in 2D classification, per frame B-factor weighting and translational movie alignment was applied to the remaining 282 k particles. The colouring of the surfaces is according the standard polymerase subunit colouring: A190, A135, A49, A43, AC40, A34.5, Rpb5, Rpb6, AC19, Rpb8, A14, A12.2, Rpb10 and Rpb12 are coloured in grey, wheat, light blue, slate, red, pink, magenta, silver blue, yellow, green, hot pink, orange, blue and lemon, respectively. Template DNA, non-template DNA and RNA are depicted in medium blue, sky blue and red, respectively. The structures against greyed background indicate final EC and PolI monomer structures.

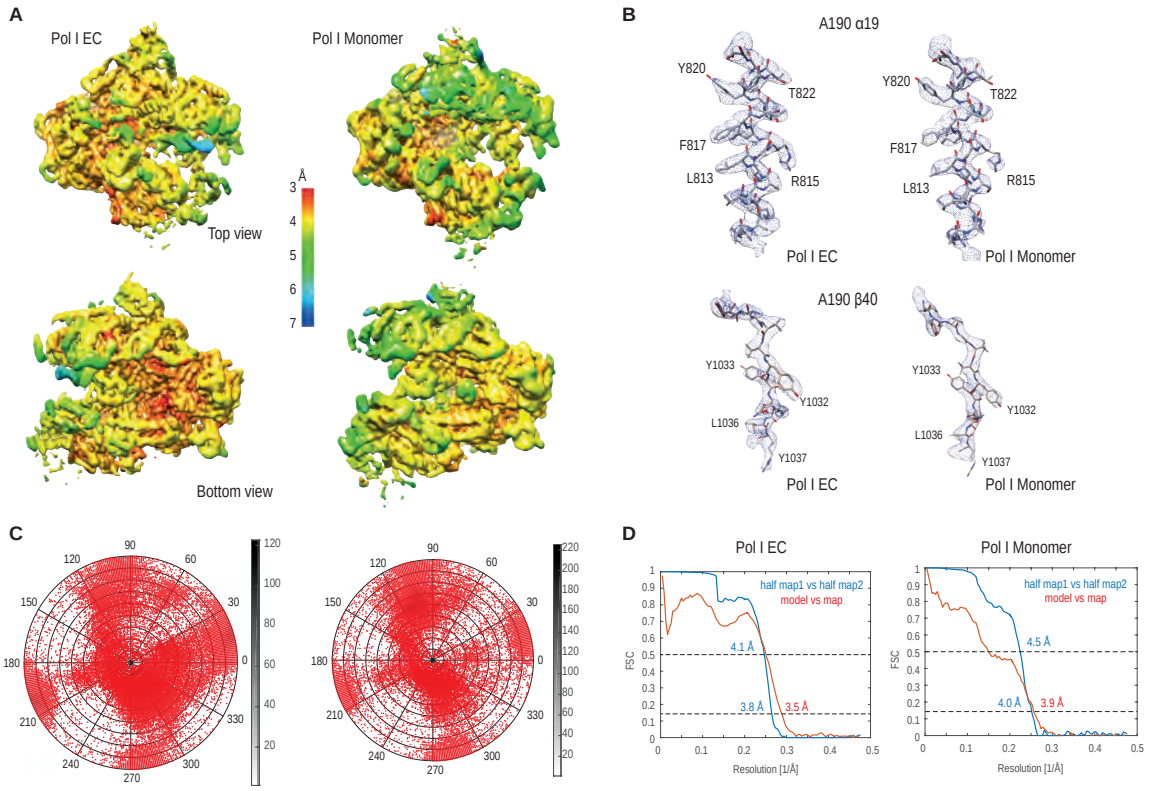


Figure 2.4. Quality of cryo-SP reconstructions. **(A)** Top and bottom view of local resolution surface maps. **(B)** Representative areas of the cryo-SP density for Pol I EC (left panel) and Pol I monomer (right panel). The A190 helix α19 (upper panel) and the A135 strand β40 (lower panel) are depicted together with the refined model superimposed. **(C)** Angular distribution of particle images. Red dots indicate views with at least one particle assigned within 1°. Black shading represents the number of particles. The orientation occupancy is similar for all three structures and covers most of the angles. **(D)** Fourier Shell Correlation (FSC) curves. Blue lines indicate the FSC between half maps of the respective reconstruction and red lines indicate FSC between the derived model against the cryo-SP map.

Table 2.1. Model refinement statistics. Statistics for the core of Pol I excluding A14/A43 and A34.5/A49 are provided in parenthesis.

	Pol I Monomer	Pol I EC
Map CC (whole unit cell)	0.850 (0.844)	0.728 (0.724)
Map CC (around atoms)	0.703 (0.739)	0.733 (0.763)
rmsd (bonds)	0.008 (0.008)	0.007 (0.007)
rmsd (angles)	1.00 (0.86)	0.929 (0.864)
All-atom clashscore	32.11 (26.42)	13.72 (10.7)
Ramachandran plot		
outliers	0.4% (0.2%)	0.4% (0.1%)
allowed	4.9% (4.6%)	6.6% (6.5%)
favored	94.7% (95.2%)	93.0% (93.4%)
Rotamer outliers	1.0% (0%)	1.1% (0%)
C-beta deviations	1 (0)	1 (0)
EMRinger score	0.65 (0.72)	2.74 (2.93)
Molprobity score	2.56 (2.24)	2.12 (1.98)

In the EC structure, the connector is detached from Pol I, as observed when Pol I is bound to the initiation factor Rrn3 (Engel et al., 2016; Pilsl et al., 2016). The expander is also displaced, enabling Pol I to form extensive interactions with the DNA-RNA hybrid (Fig. 2.2A). The enzyme contacts the DNA template at positions +4 to -9 and the RNA transcript at positions -1 to -8 (+1 represents the nucleotide addition site). Pol I generally binds nucleic acids with the same elements as Pol II (Kettenberger et al., 2004), but uses several Pol I-specific residues to contact the upstream part of the DNA-RNA hybrid. The active centre adopts a catalytically competent conformation. The bridge helix is folded throughout (Fig. 2.5B and C) and the trigger loop is largely mobile, allowing for binding of the nucleoside triphosphate substrate. The polymerase switch regions and cleft loops adopt similar positions as in the Pol II EC (Kettenberger et al., 2004) except that fork loop 1 is bent away from the hybrid (Fig. 2.6A), as in the Pol III EC (Hoffmann et al., 2015) and in a Pol II initiation intermediate (Plaschka et al., 2016).

The Pol I EC structure also provides insights into the regulation of the intrinsic RNA cleavage activity of Pol I. RNA cleavage requires subunit A12.2 (Kuhn et al., 2007; Sosunov et al., 2003), which consists of two domains. The N-terminal domain resembles that of the Pol II subunit Rpb9, whereas the C-terminal domain corresponds to the catalytic domain of the Pol II RNA cleavage factor TFIIS (Kettenberger et al., 2003; Ruan et al., 2011). In the EC structure, the N-terminal domain of A12.2 re-

mains at the outer rim of the Pol I funnel region, whereas its C-terminal domain is displaced from the pore that it occupies in the Pol I crystal structures (Engel et al., 2013; Fernández-Tornero et al., 2013; Kostrewa et al., 2015). Displacement of the A12.2 C-terminal domain from the pore apparently occurs during cleft contraction because modelling of this domain in the pore results in a clash with the contracted shelf module (Fig. 2.6B and C). Thus A12.2 can only enter the active centre when the cleft is fully or partially expanded. This predicts that Pol I adopts a partially expanded conformation during A12.2 action, which is required for RNA proofreading and polymerase reactivation after backtracking.

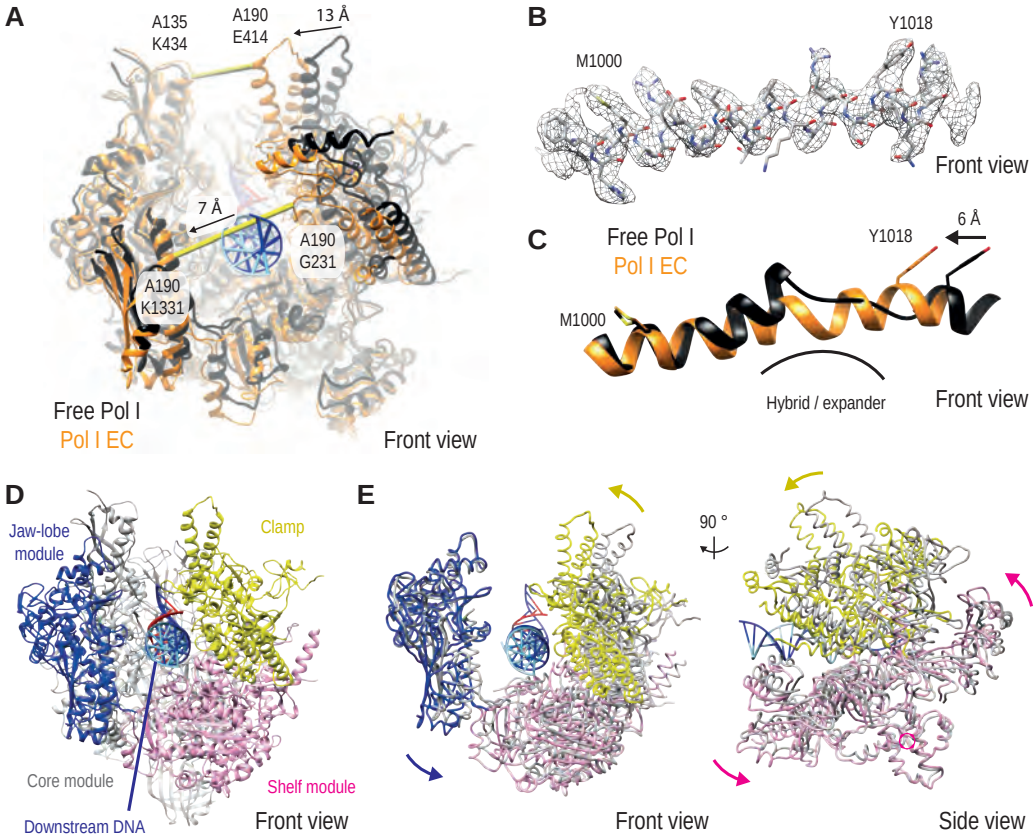


Figure 2.5. Cleft contraction and module movements. **(A)** Comparison of structures of Pol I EC (orange) and free, dimeric Pol I (PDB 4C2M, black) after superposition of their A135 subunits. Cleft width was measured between subunit A190 residue E414 and subunit A135 residue K434. For clarity, only subunits A190 and A135 are displayed. **(B)** Electron density of the folded bridge helix in the Pol I EC. **(C)** Comparison of bridge helices in the EC (orange) and free Pol I (black). **(D)** Pol I EC ribbon model coloured by four mobile modules. The peripheral subcomplexes A14/A43 and A49/A34.5 are omitted for clarity. **(E)** Movements of polymerase modules upon cleft contraction. Ribbon models of free Pol I (grey) and EC are shown after superposition of their core modules (omitted). Arrows indicate movement and rotation of the clamp-shelf and the jaw-lobe modules.

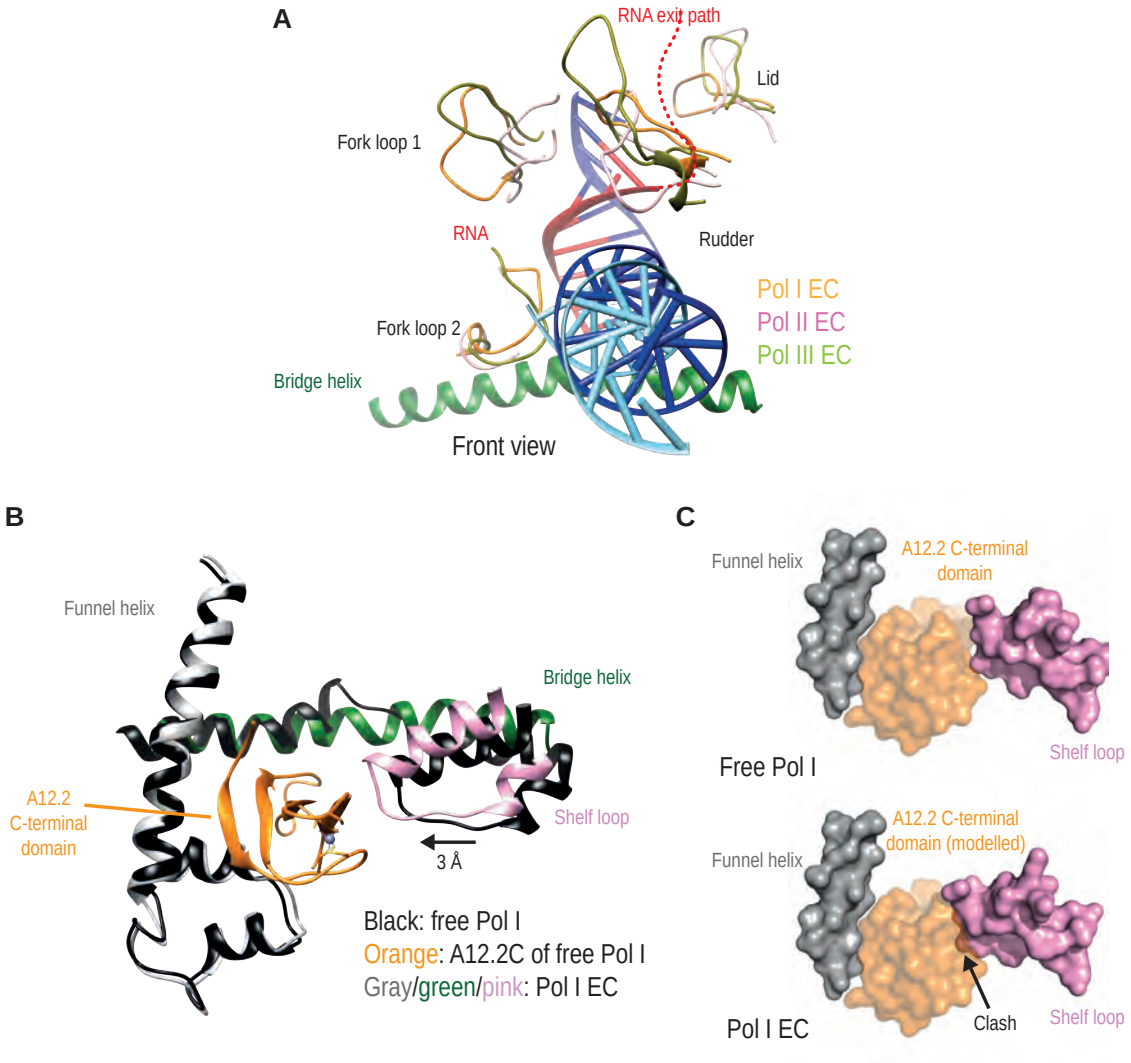


Figure 2.6. Additional details on Pol I EC. (A) Cleft loops. Ribbon model of ECs of all three *S. cerevisiae* RNA polymerases superimposed on the bridge helix. Bridge helix (green) and downstream DNA together with DNA-RNA hybrid (blue and red) are given for Pol I. **(B)** Ribbon model of free Pol I (4C2M, black and orange) superimposed on the shown inner A190 funnel helix α21 with Pol I EC (gray, green and pink). As a consequence of cleft contraction, parts of the shelf module move in and reduce the width of the pore to impair binding of the C-terminal domain of A12.2. **(C)** Modeling the A12.2 C-terminal domain into the pore of the contracted Pol I EC results in a clash. In the upper part, a surface representation of domains in free Pol I shows that the C-terminal domain of A12.2 fills the pore that is lined by the A190 funnel helix α21 and loop 1572-1579 of the A190 cleft domain in the shelf module. In the lower part, cleft contraction observed in the EC reduces the width of the pore, causing a steric clash in the model.

To investigate whether the structural differences between the Pol I EC and the free Pol I dimer arise from nucleic acid binding or from conversion of a dimer to a monomer, we also solved the structure of monomeric Pol I in the absence of nucleic acids at 4.0 Å resolution using approximately 80,000 single particles (Fig. 2.3 and 2.7A). In this structure, the connector and expander were also displaced, but the Pol I cleft was only partially contracted (approximately 5 Å), as observed in the Pol I-Rrn3 complex (Engel et al., 2016; Pilsl et al., 2016). The central bridge helix remained partially unwound, and the C-terminal domain of A12.2 remained in the pore (Fig. 2.7B and C).

Thus conversion of the Pol I dimer to a monomer leads to a partially expanded conformation, but not to the fully contracted, active conformation. The partially expanded conformation resembles the conformation observed when the bacterial RNA polymerase adopts a paused (Weixlbaumer et al., 2013) or an inhibited (Tagami et al., 2010) state. In both polymerases, movement of a rigid shelf-clamp unit allows for expansion of the cleft and a widening of the pore (called the 'secondary channel' in bacterial RNA polymerase). Available data thus suggest that RNA polymerases can adopt partially expanded and contracted conformations that are associated with inactive and active states, respectively. Binding of nucleic acids in the cleft apparently maintains the contracted conformation and excludes A12.2 from the pore, whereas rearrangements in the nucleic acids upon misincorporation or pausing could induce the partially expanded conformation that is transcriptionally inactive but enables A12.2 entry into the pore and enzyme reactivation by RNA cleavage. According to this model, transcription elongation can be regulated by allosteric coupling of nucleic acid binding with cleavage factor binding in the cleft and pore, respectively, via contraction and expansion of the polymerase.

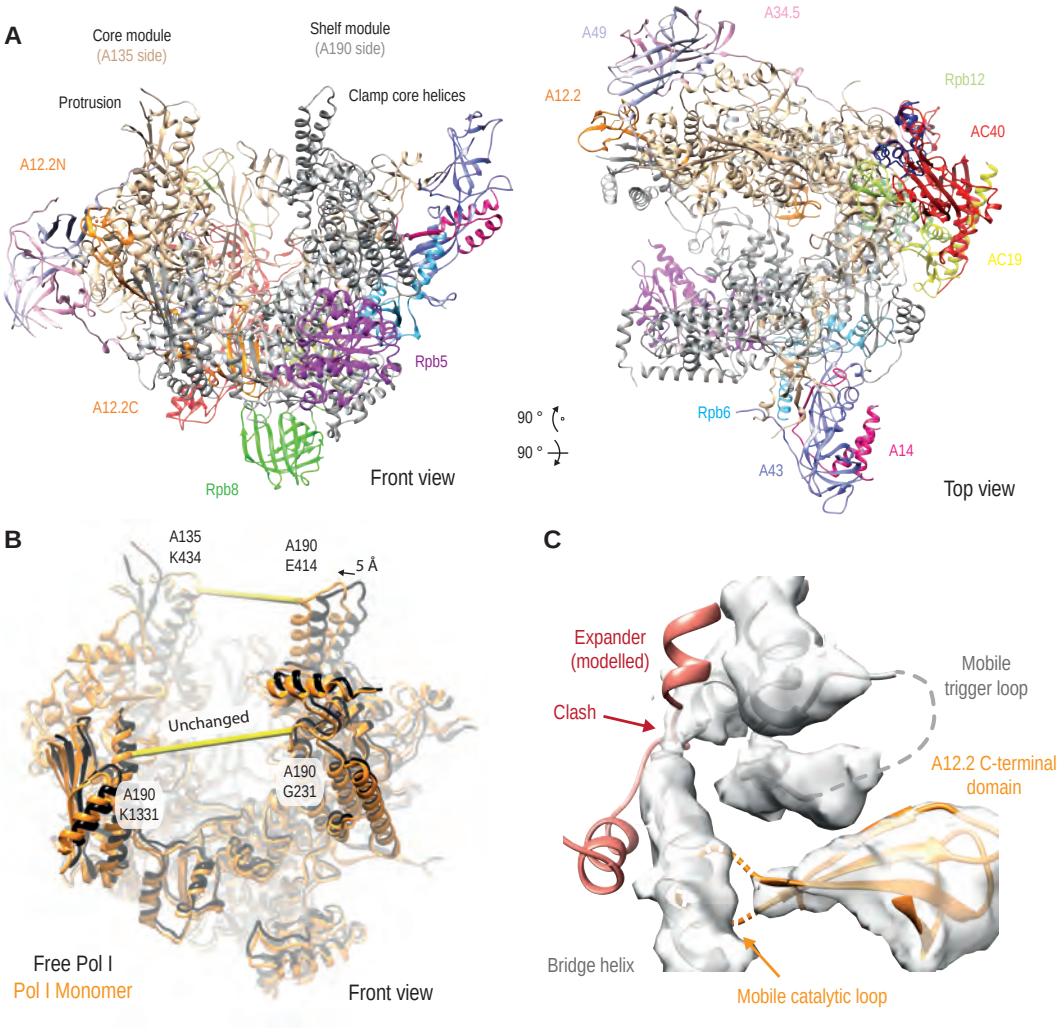


Figure 2.7. Free monomeric Pol I cryo-SP structure. **(A)** Ribbon model of free, monomeric Pol I solved by cryo-SP. The views correspond to the 'front' and 'top' views with the incoming downstream DNA pointing towards the reader. The coloring of the surfaces is according to standard polymerase coloring: subunits A190, A135, A49, A43, AC40, A34.5, Rpb5, Rpb6, AC19, Rpb8, A14, A12.2, Rpb10 and Rpb12 are colored in grey, wheat, light blue, slate, red, pink, magenta, silver blue, yellow, green, hot pink, orange, blue and lemon, respectively. **(B)** Free Pol I (PDB 4C2M, black) and Pol I Monomer (orange) were superimposed on A135. Cleft width was measured between indicated residues. For clarity, only subunits A190 and A135 are shown. **(C)** Electron microscopy density (semi-transparent grey) is shown together with models for the bridge helix, trigger loop (both grey) and the C-terminal domain of A12.2. The expander (red) is not present in this structure but modelled here based on the crystal structure of the free Pol I dimer, revealing a clash.

2.1.3. Miller spread cryo-tomography

To investigate the physiological relevance of the cryo-SP structure, we further determined the structure of the natural Pol I EC that forms in yeast cells by promoter-dependent initiation on rDNA with the use of cryo-electron tomography (cryo-ET) (Fig. 2.8). We spread active rDNA genes from exponentially growing yeast cells onto an electron microscopy grid such that they formed 'Miller trees' (Miller and Beatty, 1969) (Extended Fig. B.1). To overcome previous limitations in sample preparation, we used instant plunge-freezing to keep the sample in a close-to-native environment. The obtained images revealed the detailed arrangement of Pol I enzymes along rDNA, nascent RNA emerging from Pol I, and large densities at the RNA ends that resemble classical knobs and large small-subunit processomes (Fig. 2.8A and B) (Osheim et al., 2004). From the cryo-ET images, we selected 11 complete Miller trees and several smaller Pol I trails, each containing 10-20 Pol I enzymes with associated RNA. This yielded 993 transcribing Pol I enzymes for further analysis.

We observed that each rDNA gene is loaded with approximately 70 Pol I enzymes, which showed a median centre-to-centre distance of 18+/-10 nm (Fig. 2.8C, Extended Fig. B.2A and B), consistent with previous results (Osheim et al., 2004). Only approximately 2 % of the Pol I complexes were separated by a distance of approximately 12 nm, which would allow for interaction between enzymes. Furthermore, consecutive enzymes show random relative orientations, arguing against specific interactions that were suggested previously (Albert et al., 2011). Each triple of successive Pol I molecules created a triangle with an in-plane angle of approximately 150° (Fig. 2.8C and Extended Fig. B.2A, B and C) that was independent of the length of the DNA between enzymes. This angle could not be obtained from the cryo-SP structure, because density for upstream DNA was poor.

We used sub-tomogram averaging (n=225) to obtain a cryo-ET structure of the cellular transcribing Pol I at a resolution of approximately 2.9 nm (FSC 0.5 criterion; approximately 2.5 nm with the FSC 0.143 criterion; Fig. 2.9A). The cryo-ET structure strongly resembled the cryo-SP structure, showing an overall cross-correlation score of 0.85. An FSC plot between the cryo-ET and cryo-SP structures decreased beyond the 0.143 threshold at 3.1 nm (Extended Data Fig. 8a). The peripheral subcomplexes A14/A43 and A34.5/A49 were flexible, consistent with the weaker density observed in the cryo-SP structure (Fig. 2.3). The width of the active centre cleft was

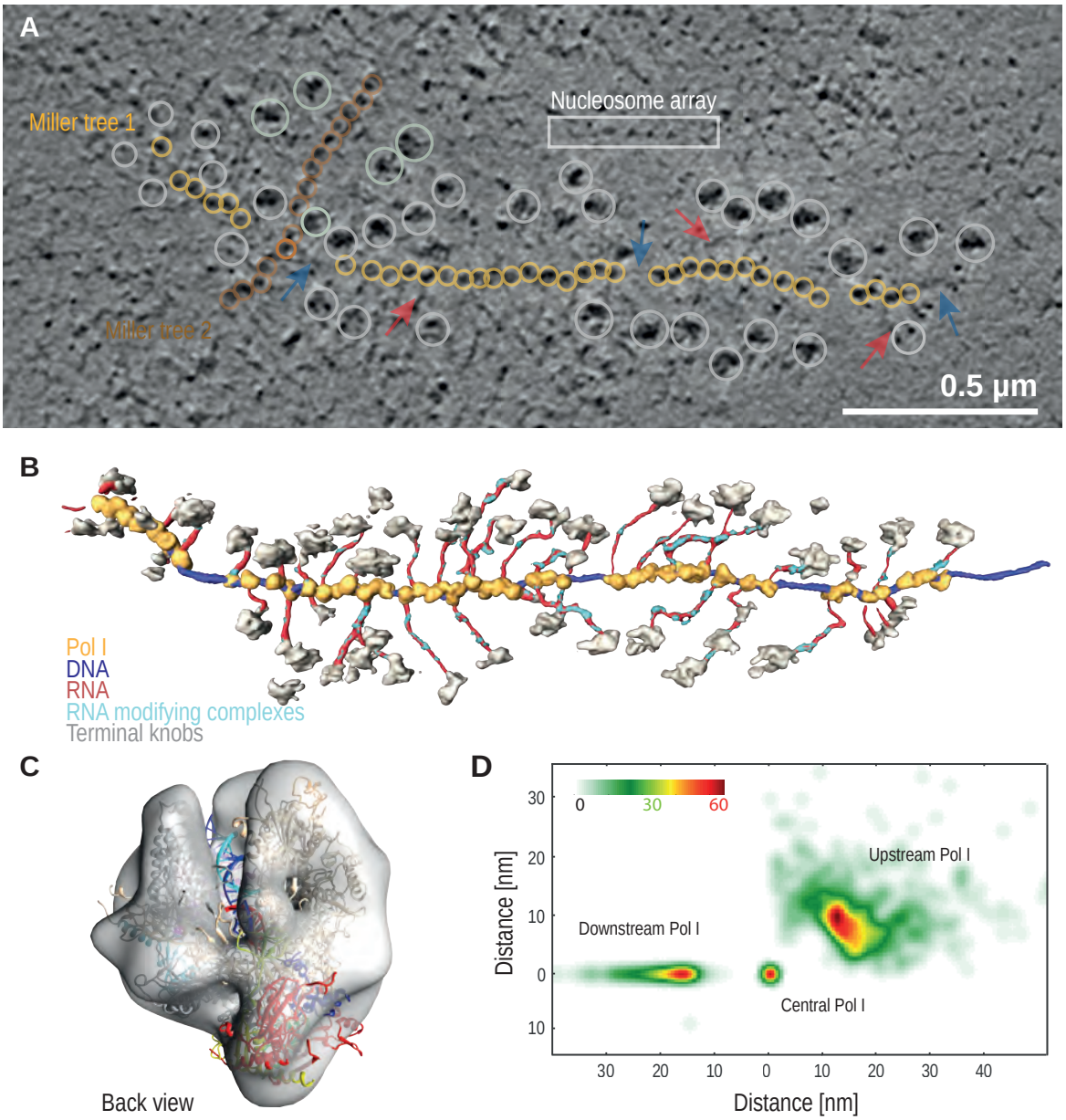


Figure 2.8. Cryo-ET analysis of Pol I transcribing rDNA genes. **(A)** 2 nm thick tomographic slice though a cryo-ET image with two of the Miller trees, showing the terminal knobs (grey circles), the DNA (typical examples marked by blue arrows), the RNA (red-pink arrows), and the Pol I enzymes (yellow and dark yellow circles for first and second Miller trees, respectively). Several nucleosomes are attached to DNA like beads on a string (white box). **(B)** Three-dimensional surface rendering of the first Miller tree in (a), showing the terminal knobs (light grey), DNA (blue), RNA (red), possible RNA-modifying complexes (cyan), and Pol I complexes (yellow). **(C)** Schematic of three consecutive Pol I enzymes together with probability density localization (heat map) of the upstream Pol I, the first downstream Pol I (in the centre), and the second downstream Pol I (on the y-axis). **(D)** Fit of the Pol I EC ribbon model from cryo-SP into the cryo-ET reconstruction in grey. The good fit observed here is not possible with the expanded conformation of Pol I (Fig. 2.9B).

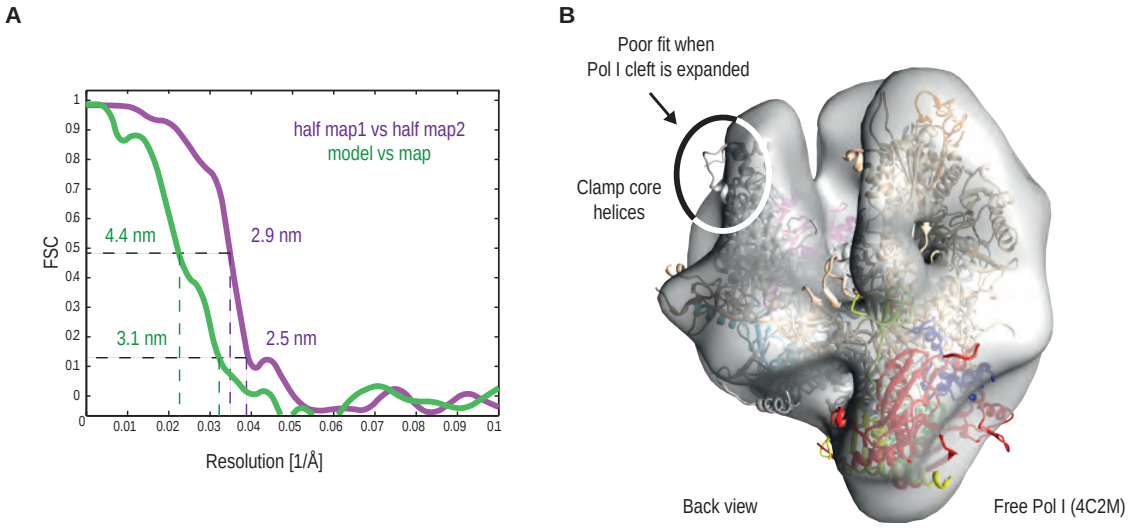


Figure 2.9. Comparisons between cryo-ET and cryo-SP structures. (A) FSC of the cryo-ET structure with a resolution of 29 Å (purple line) and mutual FSC between the cryo-ET structure and cryo-SP structure with estimated resolutions of 44 Å and 31 Å, respectively (green line). Resolutions were measured at FSC 0.5 and 0.143 criteria, respectively. **(B)** Poor fit of the expanded, free Pol I crystal structure (4C2M) to the cryo-ET density (grey). Note that the clamp domain does not fit the density unless the polymerase is in the contracted conformation (compare Fig. 2.8).

the same in both structures (Fig. 2.8D, and Fig. 2.9C), confirming that the contracted cryo-SP structure represented the natural conformation of actively transcribing Pol I.

Taken together, we used here two independent cryo-electron microscopic approaches to define the contracted Pol I conformation as the active, transcribing state of the enzyme, and provide evidence that the elongation phase of transcription is regulated by cleft contraction and expansion.

2.2. Structural basis of RNA polymerase I transcription initiation

Results presented in this section are in preparation for publication and were obtained in a collaboration with Tobias Gubbey and Christoph Engel. Author contributions are stated on page VII.

C. Engel*, T. Gubbey*, **S. Neyer***, S. J. Sainsbury, C. Oberthür, C. Bäjen, C. Bernecky, P. Cramer. (2016) Structural basis of RNA polymerase I transcription initiation. (manuscript in preparation)
* These authors contributed equally to this work.

While focusing on the analysis of the initially transcribing and the closed complex structure, the following section gives a comprehensive picture of all results presented in this study. Methods are described in section 3.2. Supplemental data and methods that were predominantly obtained by the collaborators are described in the appendix C and E, respectively.

2.2.1. Abstract

Gene class-specific transcription results from promoter-specific assembly of RNA polymerases (Pols) I, II, and III with their specific initiation factors. The molecular basis of promoter specificity remained elusive because structural studies of transcription initiation complexes were limited to Pol II. Here we use a combination of X-ray crystallography and cryo-electron microscopy to obtain a structural model for Pol I initiation. The Pol I-specific core initiation factor contains two modules, one that binds upstream promoter DNA, and one that docks to the initiation-competent Pol I-Rrn3 complex. This results in loading of DNA onto the polymerase wall and into an expanded active center cleft. DNA opening between the protrusion and clamp domains enables cleft contraction that results in an active enzyme conformation and RNA synthesis. Comparison with the Pol II system indicates that promoter specificity stems from alternative initiation factor structures and contacts with DNA and polymerase, and from differences in the 'bendability' and 'meltability' of promoter DNA.

2.2.2. Introduction

Transcription in eukaryotic cells is carried out by three different RNA polymerases, Pol I, II, and III (Roeder and Rutter, 1969). The polymerases share a conserved active center (Cramer et al., 2001) and form structurally similar complexes for transcription elongation (Gnatt et al., 2001; Hoffmann et al., 2015; Neyer et al., 2016; Tafur et al., 2016), but they synthesize different classes of RNAs. Whereas Pol I synthesizes the ribosomal RNA (rRNA) precursor (Moss et al., 2007), Pol II and Pol III mainly produce mRNAs and tRNAs, respectively. For gene class-specific transcription the polymerases assemble with different sets of initiation factors at their specific promoters. The mechanisms underlying promoter-specific initiation and gene class-specific transcription remain poorly understood.

Whereas the structure of initiation complexes of Pol I and Pol III are unknown, recent studies have elucidated the structural basis for Pol II initiation (He et al., 2016; Plaschka et al., 2016; Robinson et al., 2016). Initiation begins with the assembly of Pol II and the general initiation factors TFIIB, -D, -E, -F and -H on promoter DNA, to form a closed complex (CC) (Buratowski et al., 1989; Grunberg and Hahn, 2013; Roeder, 1996; Sainsbury et al., 2015). In the CC, promoter DNA is positioned above the active center cleft of Pol II, running along the tip of the clamp domain. Upon ATP hydrolysis by a translocase in TFIIH, DNA is unwound and the template strand is inserted into the active center, resulting in an open complex (OC) (Grunberg et al., 2012). Subsequently, the transcription start site (TSS) is recognized and RNA synthesis commences, giving rise to an initially transcribing complex (ITC). When the RNA product grows beyond a critical length, initiation factors dissociate and the elongation complex (EC) forms.

Pol II initiation critically depends on TFIIB. TFIIB bridges between Pol II and promoter DNA, and its location on Pol II determines the topology of the initiation complex (Kostrewa et al., 2009). TFIIB comprises an N-terminal zinc ribbon domain that binds the dock domain of Pol II (Bushnell et al., 2004; Chen and Hahn, 2004) and is essential for recruiting Pol II to the promoter (Ha et al., 1991; Lin et al., 1991; Malik et al., 1991). TFIIB further contains two cyclin domains (Nikolov et al., 1995) that are located on the Pol II wall to position promoter DNA above the cleft (Chen and Hahn, 2004; Kostrewa et al., 2009). The TFIIB linker and reader elements connect the TFIIB ribbon and cyclin domains, and participate in DNA opening and TSS recognition, respectively (Kostrewa et al., 2009; Sainsbury et al., 2013).

The initiation system of Pol I differs substantially from that of Pol II. Initiation by yeast Pol I requires Rrn3 and the core factor (CF), which consists of subunits Rrn6, Rrn7, and Rrn11 (Keener et al., 1998; Lalo et al., 1996). Rrn3 is an extended HEAT repeat protein (Blattner et al., 2011) that binds Pol I at the stalk subcomplex A14/A43 (Blattner et al., 2011; Engel et al., 2016; Pilsl et al., 2016), thereby stabilizing the initiation-competent, monomeric form of Pol I. Rrn6 contains a predicted WD40 β-propeller and a helical domain (Knutson et al., 2014). Rrn7 is predicted to resemble TFIIB and to contain an N-terminal ribbon, two cyclin fold domains and a C-terminal domain (Knutson and Hahn, 2011; Naidu et al., 2011), but the TFIIB reader is poorly conserved. Rrn11 is predicted to contain a tetratricopeptide repeat (TPR) domain (Knutson et al., 2014). The human counterpart of CF, selectivity factor 1 (Russell and Zomerdijk, 2006), comprises homologues to Rrn6 (TAF1C), Rrn7 (TAF1B), and Rrn11 (TAF1A), and the additional subunits TAF1D and TAF12 (Denissov et al., 2007; Gorski et al., 2007).

Here we elucidate the molecular basis of Pol I transcription initiation with the use of a structural biology hybrid approach. We report the crystal structure of yeast CF and show it strongly differs from TFIIB. We also report the cryo-electron microscopy (cryo-EM) structures of the Pol I-Rrn3-CF complex and a Pol I ITC containing Rrn3, CF and the ribosomal DNA promoter. Biochemical probing of these structures using a minimal promoter-specific initiation assay elucidates the mechanism of Pol I initiation. Comparison of our results with the Pol II system reveal major differences and provide insights into the molecular basis and evolution of gene promoter specificity of the different transcription systems.

2.2.3. Crystal structure of core factor

We obtained *Saccharomyces cerevisiae* core factor (CF) after heterologous co-expression of its three subunits in *E. coli* (Methods appendix E). For CF purification, we modified previously published protocols (Bedwell et al., 2012; Knutson et al., 2014) and could remove co-purifying chaperones (Methods appendix E). Purified CF contained the three subunits in apparently stoichiometric amounts (Fig. C.1) and could be crystallized by vapor diffusion (Methods appendix E). The crystal structure was determined by single-wavelength anomalous diffraction (SAD) using selenomethionine-derivatized crystals in space group P65 (Table C.1 and C.2). Native crystals comprised six CF complexes per asymmetric unit in space group P1 and diffracted to 3.2 Å resolution.

Building of an atomic model was supported by sequence markers, including 21 selenium sites that revealed the location of methionine residues. We additionally located 17 cysteine residues and 5 sulphate ions with the use of anomalous diffraction from sulphur atoms, and introduced six additional methionine markers in regions that were difficult to build (Methods, appendix E). The structure was refined to R/R_{free} factors of 26.0/28.9% with good stereochemistry (Table C.1).

2.2.4. Core factor comprises two modules

The crystal structure reveals that CF comprises two modules that are flexibly linked and separated by a central canyon (Fig. 2.10). Module I is formed by an N-terminal seven-bladed β-propeller domain in Rrn6 and by subunit Rrn11, which forms a superhelical TPR domain and a novel helical domain that we name 'propeller- and promoter-associated domain' (PAD). Module II is formed by a unique helical C-terminal domain in Rrn6 ('headlock domain') that wraps around Rrn7, which forms two cyclin folds as predicted (Knutson and Hahn, 2011; Naidu et al., 2011). The C-terminal Rrn7 cyclin domain contains an insertion that forms six helices (α8a-f) and reaches over to module I (Fig. 2.10B, C.1). The structure lacks only the mobile N- and C-terminal regions of Rrn6, an insertion in the β-propeller, the 93 N-terminal residues of Rrn7 which includes the flexibly linked ribbon domain, and a mobile insertion in the Rrn11 PAD. The structure enabled us to manually curate alignments of CF subunit sequences from different species (Fig. C.2, C.3). The CF structure rationalizes effects of known mutations. For example, the Rrn11 N-terminus is buried in the Rrn6 β-propeller, explaining why its deletion is lethal (Knutson et al., 2014).

2.2.5. Rrn7 differs from TFIIB

A search for structural similarity to Rrn7 using PDBeFold (Krissinel and Henrick, 2004) identified TFIIB and its archaeal homologue TFB (PDB 1VOL and 1D3U), and led to a structure-based alignment of Rrn7 with TFIIB (Fig. 2.11A-C, C.2). This is consistent with the predicted sequence homology of Rrn7 and TFIIB (Blattner et al., 2011; Knutson and Hahn, 2011; Naidu et al., 2011). However, the Rrn7 cyclin domains are imbedded in the CF structure and differ strongly from their TFIIB counterparts. In particular, the surface of cyclin domain II is altered by an insertion that bridges between the CF modules.

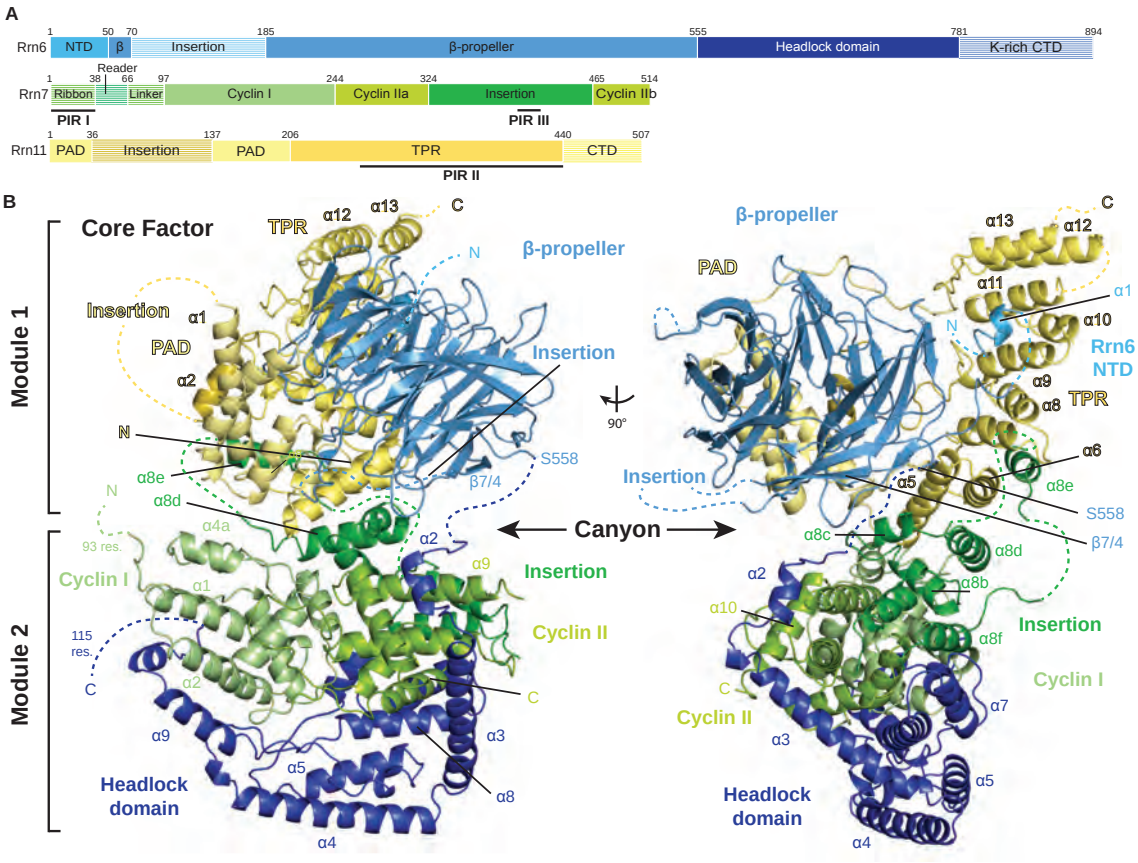


Figure 2.10. Crystal Structure of yeast Core Factor. (**A**) Domain organization of the CF subunits Rrn6 (blue), Rrn7 (green) and Rrn11 (yellow) indicated by a to-scale bar diagram. Domain boundaries, unstructured regions and polymerase interacting regions (PIRs) are indicated. (textbfB) Architecture of CF: Rrn6 forms a seven-bladed WD40 β-propeller which is connected to a helical 'headlock' domain that wraps around Rrn7. A Lysine-rich C-terminal domain is flexible in our structure. Rrn7 forms two cyclin domains, of which the second is interrupted by a large insertion. TFIIB-homologous elements on the N-terminus are flexible in the crystal. Rrn11 has a propeller/promoter-associated domain (PAD) with a flexible insertion and is followed by a TPR fold. For detailed view of each subunit see Fig. C.1.

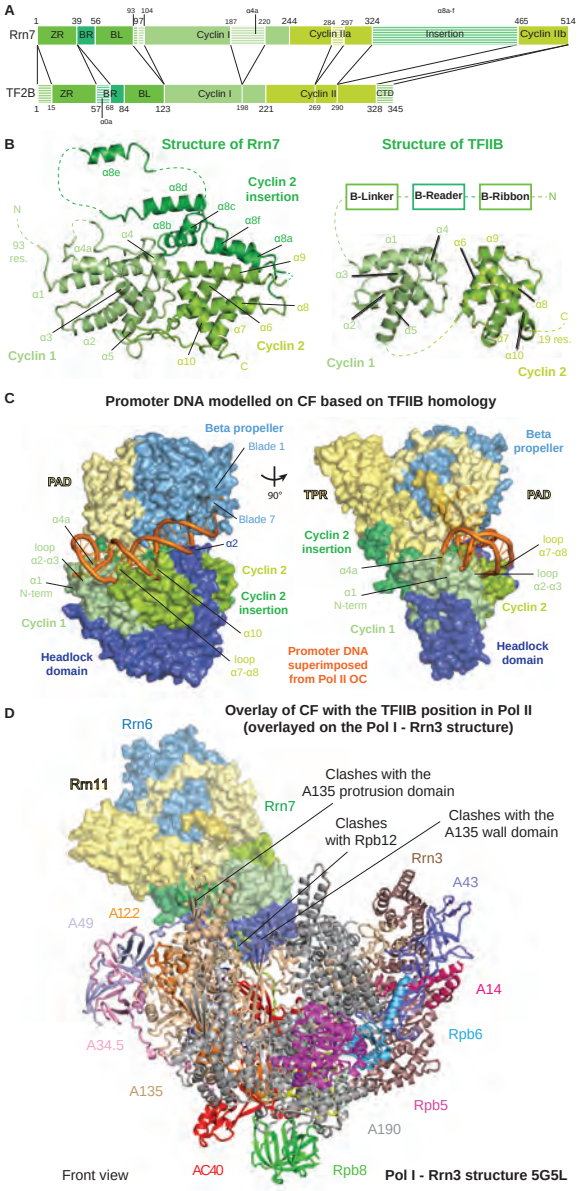


Figure 2.11. Rrn7 structurally differs from TFIIB. (**A**) Domain organization of Rrn7 and TFIIB (to scale) shows differences between the proteins. Insertions and deletions of more than 10 residues are indicated in white. (**B**) Side-by-side structure-comparison of yeast Rrn7 (left) and yeast TFIIB (right, PDB 5FYW). Among the differences are flexible loops, helix α4a, an elongated helix α9 and the large cyclin II insertion (dark green). (**C**) Overlay of Rrn7 cyclin domains with the structure of TFIIB (PDB 5FYW) in complex and promoter DNA. Clashes with parts of both cyclin domains and Rrn6 are visible. For details, see text. (**D**) Using the Pol II - TFIIB structure 4BBS as a reference, the Pol I-Rrn3 structure 5G5L is overlaid (cartoon representation) via the second largest domain (Rpb2-A135). If CF (space filling) is now positioned by superimposing Rrn7 cyclin I with its TFIIB counterpart, a number of clashes are visible, especially the A135 wall and protrusion domains and subunit Rpb12. The view is from the front (Cramer et al., 2001).

Modelling of CF onto DNA based on a superposition of Rrn7 with TFIIB in the Pol II initiation complex structure (Plaschka et al., 2016) shows that Rrn7 cannot bind to DNA in the way TFIIB does (Fig. 2.11C). Major clashes are revealed between modelled DNA and Rrn7 cyclin domains (α1, α4a, α10, and loops α2-α3 and α7-α8), the Rrn6 β-propeller and the headlock domain helix α2. Furthermore, modelling CF onto Pol I under the assumption that the N-terminal cyclin domain of Rrn7 binds to the polymerase wall like TFIIB does (Kostrewa et al., 2009; Sainsbury et al., 2013) leads to major clashes between CF and the Pol I wall and protrusion domains (Fig. 2.11D). Taken together, Rrn7 structurally differs from TFIIB and can neither bind DNA nor the polymerase like TFIIB does, questioning our current Pol II-based model of Pol I initiation complex architecture (Blattner et al., 2011; Knutson et al., 2014), and arguing that the position of CF on Pol I differs from that of TFIIB on Pol II.

2.2.6. Structure of the Pol I-Rrn3-CF complex

To determine the position of CF on the Pol I-Rrn3 complex, we formed a complex of endogenous *S. cerevisiae* Pol I with recombinant Rrn3 and CF. CF bound to Pol I in a stoichiometric manner when Rrn3 was present (Fig. C.4). The purified 18-subunit, 883-kDa complex was subjected to negative-stain EM analysis, leading to a 3D reconstruction at 16 Å resolution (Fig. C.5, Methods appendix E). Subsequent cryo-EM analysis and particle sorting led to three reconstructions at resolutions of 7.7 Å, 8.8 Å and 9.0 Å (Fig. 2.12, C.6, C.7). A model of the Pol I-Rrn3-CF complex was derived by unambiguously placing crystal structures of Pol I (PDB 4C2M), Rrn3 (PDB 3TJ1), and CF (this work) into the reconstructions and refining the position of protein domains as rigid bodies (Methods appendix E). The Pol I conformation and Rrn3 position were highly similar to those in the previously reported Pol I-Rrn3 complex structure (Engel et al., 2016; Pilsl et al., 2016). Thus Pol I remained in the partially expanded conformation that is similar to free monomeric Pol I.

2.2.7. CF contains three Pol I-interacting regions

In the Pol I-Rrn3-CF complex structure, module I of CF contacts the upstream end of the polymerase cleft (Fig. 2.12A), whereas module II protrudes outwards and is mobile, adopting different positions in the three EM reconstructions (Fig. C.7). CF interacts with the Pol I-Rrn3 complex using three polymerase-interacting regions (PIRs) that are

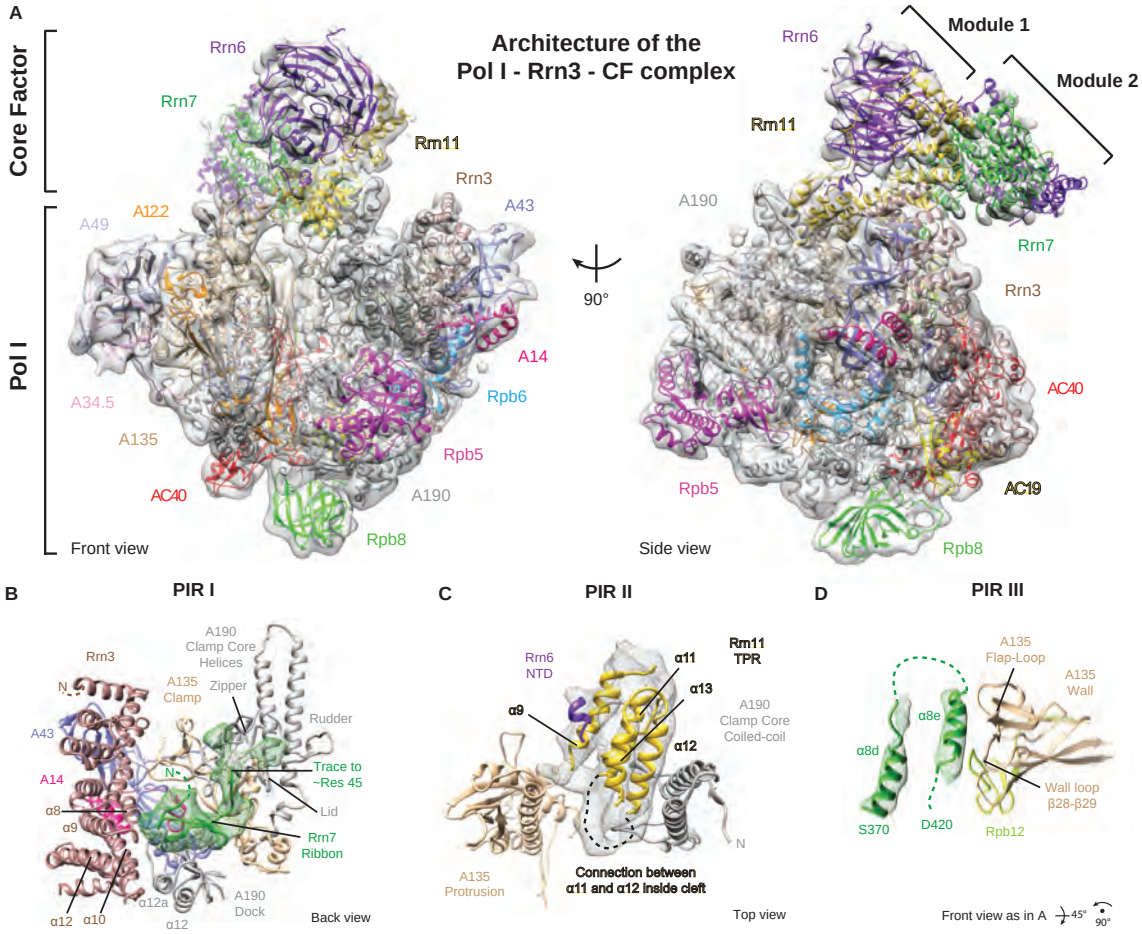


Figure 2.12. Cryo-EM reconstruction of Pol I in complex with Rrn3 and Core Factor. (**A**) Overview of the Pol I - Rrn3 - CF cryo-EM reconstruction at 7.7 Å resolution (grey envelope) which positions all Pol I subunits (cartoon), apart from the flexibly linked A49 tandem winged helix domain. CF binds to the back/top side of wall and clamp coiled-coil domains with its module I, while Rrn3 is situated at the back of the A14/43 stalk. Secondary structure elements are visible for all parts with the exception of distal CF module II domains Rrn6 headlock (dark violet) and parts of Rrn7 (dark green). (**B**) Polymerase interacting region (PIR) I between Pol I, Rrn3 and CF. As previously suggested (Engel et al., 2016) a cavity on top of the A190 dock domain binds the Rrn7 ribbon domain (modelled according to TFIIB homology, green, compare Fig. C.2). Contacts with the loop between Rrn3 helices α8 and α9 become visible. Ribbon density (green mesh) allows for additional chain tracing towards the A190 Zipper loop. The lid is either repositioned, or Rrn7 extends on top of it before exiting the polymerase and connecting with cyclin I. (**C**) PIR II. The Rrn11 TPR helices interact with the A190 clamp CC domain. Contacts to the tip of the A135 protrusion domain are visible and a loop between Rrn11 α11 and α12 becomes ordered inside the active center cleft (dashed line). (**D**) PIR III: Rrn7 helix α8e (see Fig. 1) binds the wall domain of A135, especially the Pol I specific flap loop and an A135 loop β28-β29..

defined in the cryo-EM density at the level of secondary structure (Fig. 2.12B-D). The N-terminal ribbon domain of Rrn7 (PIR I), which is mobile in the free CF structure, contacts the Pol I dock domain, like the TFIIB ribbon binds Pol II, and additionally contacts the Rrn3 loop α8-α9 (Fig. 2.12B). The C-terminal half of the Rrn11 TPR domain (PIR II) contacts the Pol I clamp and protrusion, leading to an ordering of loop α11-α12 (Fig. 2.12C). Helix α8e of the Rrn7 insertion domain (PIR III) contacts the Pol I wall (Fig. 3D). The N-terminal cyclin domain of Rrn7 is embedded in CF module II and thus positioned far away from the Pol I surface, whereas its counterpart in TFIIB contacts the Pol II wall, consistent with fundamentally different modes of DNA binding in the two systems. Provided that CF binding blocks the upstream edge of the Pol I active center cleft, major structural changes must occur for DNA loading.

2.2.8. Structure of the initially transcribing complex

To investigate how the Pol I-Rrn3-CF structure is reorganized upon promoter DNA loading and initial transcription, we solved the cryo-EM structure of an ITC that includes Pol I, Rrn3, CF, and a nucleic acid scaffold based on the natural promoter sequence (Methods section 3.2). The scaffold consisted of promoter DNA from positions -80 to +23 (+1 depicts the transcription start site), a mismatched bubble region (positions -8 to +8), and a 6-nucleotide RNA product. After selecting 567,000 particles that contained complete ITC complexes in a defined overall conformation (Fig. 2.14), we obtained a cryo-EM structure at an overall resolution of 4.2 Å (Fig. 2.13) with higher resolution in the core and lower resolution at the periphery (Fig. 2.15).

In the resulting ITC structure, Pol I was no longer in the expanded conformation but rather adopted the contracted conformation observed in the EC (Neyer et al., 2016). CF occupied an alternative position on the Pol I surface (Fig. 2.13A) that strongly differs from its position in the Pol I-Rrn3-CF complex lacking nucleic acids. CF moved by up to 90 Å and rotated by up to 30 degrees, binding the outer surface of Pol I between protrusion and wall domains. Whereas the Rrn7 ribbon (PIR I) remains bound to the polymerase dock domain, CF module I forms alternative contacts with Pol I. PIR II now contacts the outer protrusion and PIR III contacts Rpb12 and the Pol I-specific wall loop 2-3 (Fig. 2.13B). Compared with the Pol I-Rrn3-CF complex, CF employs the same PIRs to contact Pol I in the ITC. However, CF occupies enables promoter DNA loading into the cleft and cleft contraction, which are both not possible in the free Pol I-Rrn3-CF complex.

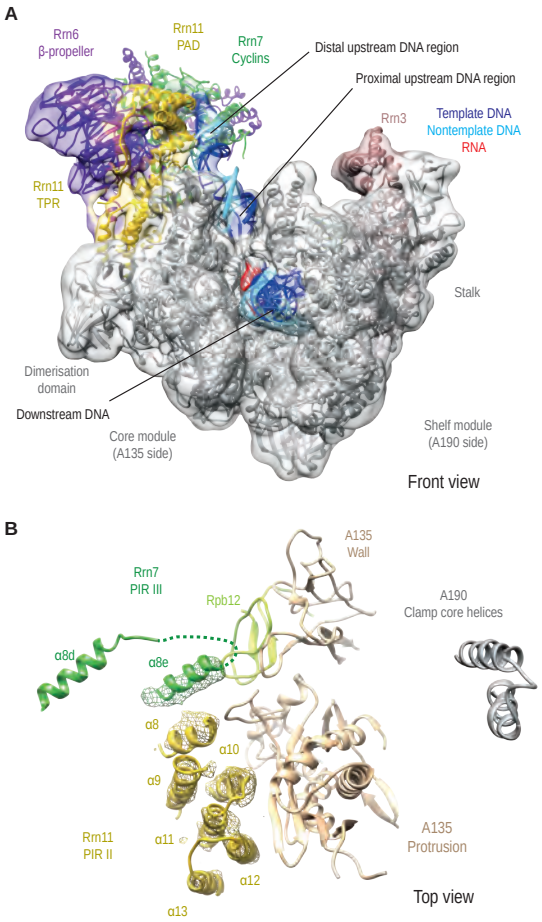


Figure 2.13. Cryo-EM reconstruction of initially transcribing Pol I. (**A**) Overview of the yeast Pol I ITC filtered cryo-EM reconstruction (transparent). Pol I, Rrn3 and DNA are shown in grey cartoon apart from protrusion and wall domains (wheat) and Rpb12 (lemon). CF and its promoter interaction and coloured as in Fig. 3. (**A**) PID II/III contacts the Pol I protrusion in the ITC. Cryo-EM density for TPR helices (yellow mesh) and cylin II insertion helix α8e (green mesh).

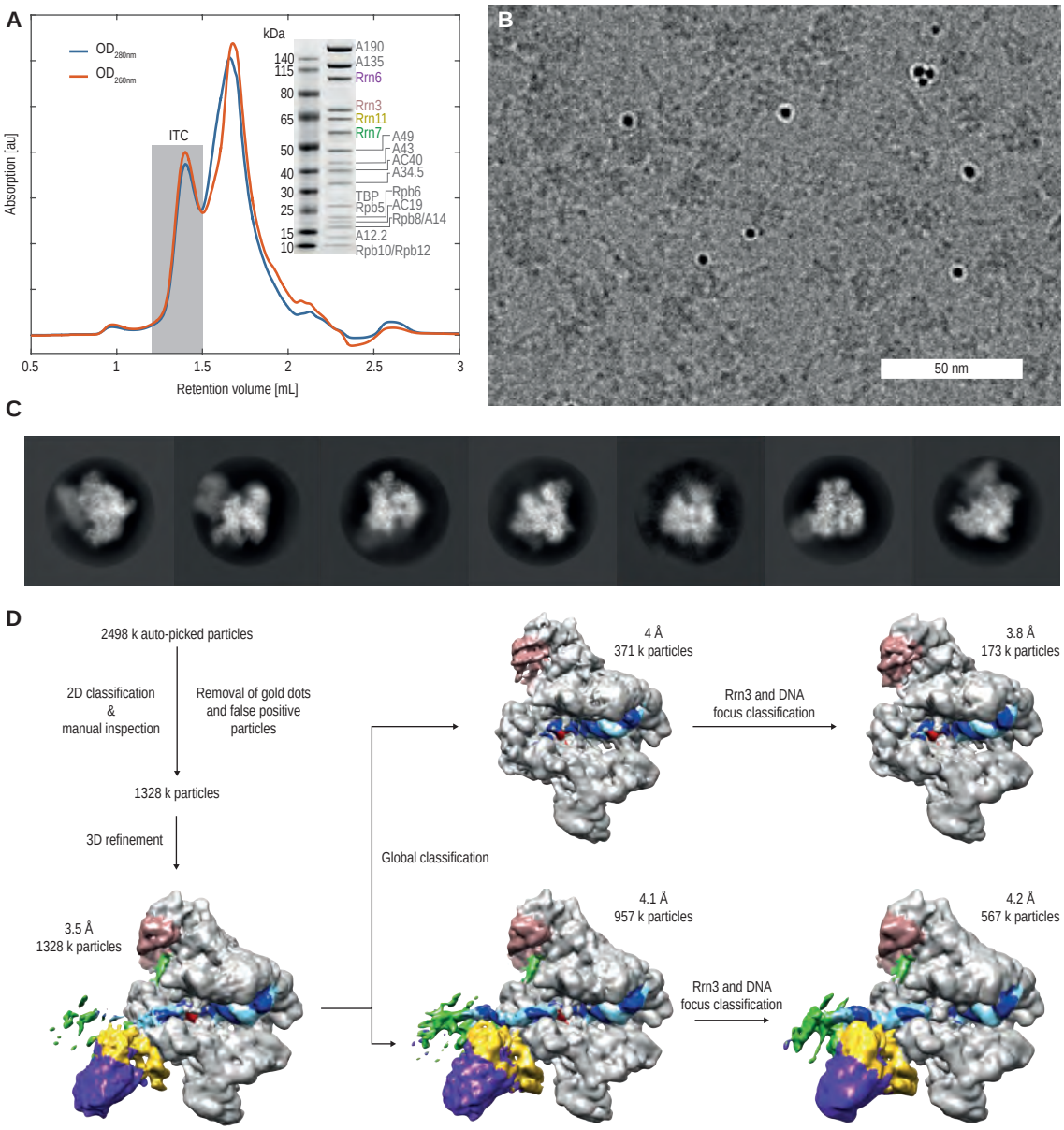


Figure 2.14. Pol I ITC sample preparation and cryo-SP data processing. (**A**) Size-exclusion chromatogram (Superose 6 Increase 3.2/300) of reconstituted Pol I ITC. Coomassie-stained SDS-PAGE of pooled peak fractions (shaded grey area) shows stoichiometric Pol I subunits, CF subunits and Rrn3 as wells as sub-stoichiometric TBP. (**B**) Representative cryo-EM micrograph showing particles of expected size as well as a high number of roundish very intense contaminations that were washed from the gold grids we used. (**C**) Examples of 2D averages after unsupervised 2D classification. (**D**) Processing tree with density elements coloured as in Fig. 2.13.

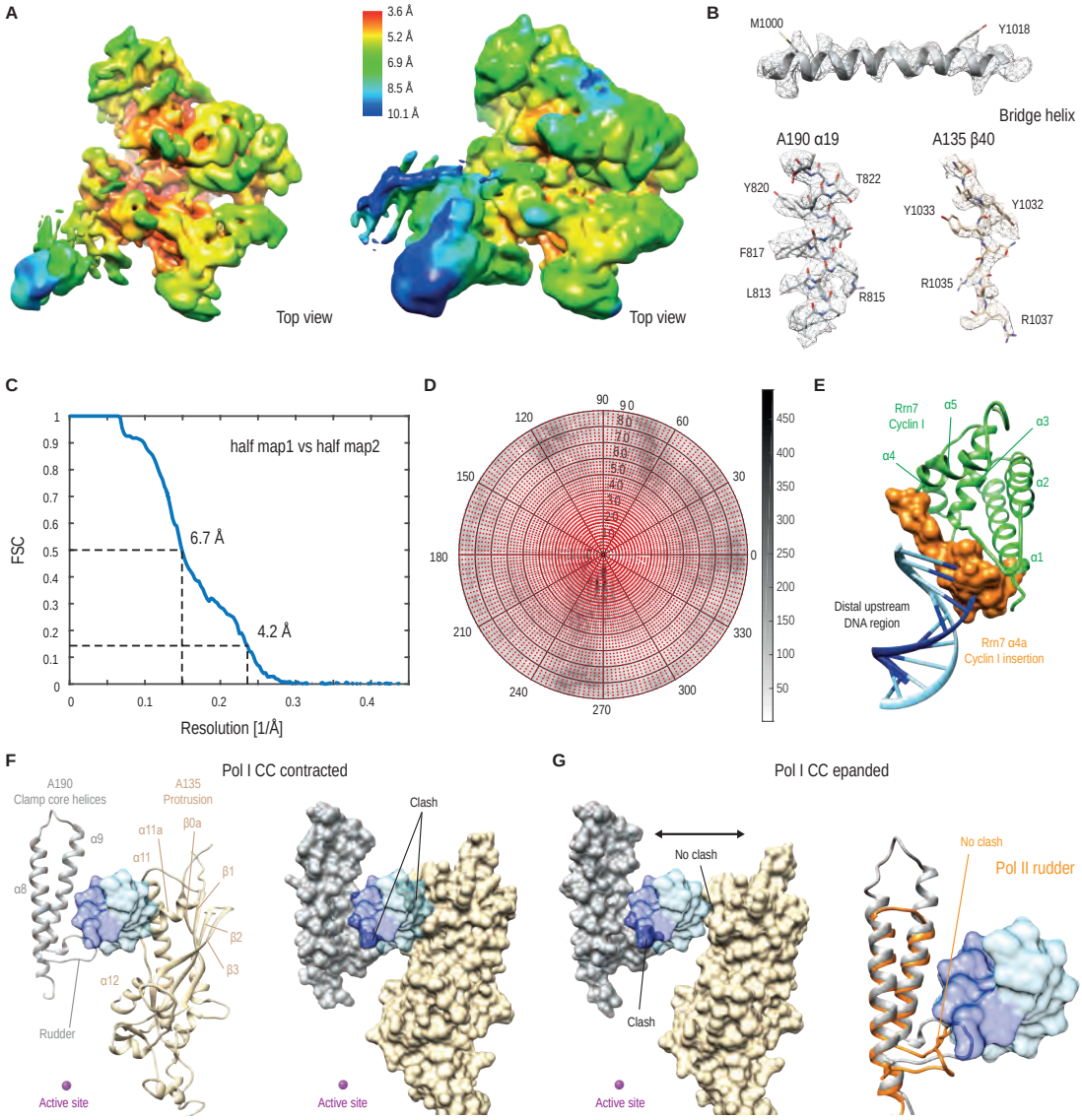


Figure 2.15. Cryo-EM reconstruction of a Pol I-Rrn3-CF initially transcribing complex. (**A**) Local resolution surface maps at normal density threshold (left panel) and lower density threshold (right panel). (**B**) Representative areas of the cryo-EM density for the Pol I core. Upper panel shows the superimposed helical bridge helix as ribbon model. The A190 helix α19 (lower left panel) and the A135 strand β40 are depicted together with the model superimposed. (**C**) Fourier shell correlation curves of the Pol I ITC density. (**D**) Angular distribution of particles used to reconstruct Pol I ITC density. (**E**) Ribbon model of distal upstream DNA region and Rrn7 Cyclin I together with the surface representation of Rrn7 helix α4a. ITC DNA binding comes together with CF re-arrangement as the Cyclin I insertion becomes flexible and would otherwise clash with DNA. (**F**) Back view of closed DNA modelled DNA on the Pol I ITC conformation displaying a contracted cleft reveals a major DNA-protrusion clash and a minor DNA-rudder clash. (**G**) Replacing contracted Pol I with the expanded Pol I-Rrn3 state (Engel et al., 2016) superimposed on the active site resolves the protrusion clash (left panel). Superposition of Pol II clamp core helices and rudder on Pol I clamp core helices visualizes the alternative rudder conformation present in Pol II. In this conformation no clash with the closed DNA is observed.

2.2.9. Promoter DNA location and contacts

The course of DNA through the ITC can be followed for most of its path from around position -32 upstream to position +20 downstream of the active site (Fig. 2.16). Upstream DNA adopts a previously unobserved position that differs from that in Pol II initiation complexes (Fig. 2.16B). Upstream DNA runs alongside the Rrn11 PAD, bends by approximately 30°, and enters the cleft between the Pol I wall and protrusion. The DNA bend apparently occurs only upon CF docking to Pol I, because regions upstream and downstream of the bend are bound by CF and Pol I, respectively.

Upstream DNA thus forms two major contacts that are likely relevant for promoter recognition (Fig. 2.16C, D). A distal upstream contact is formed between the DNA region at around registers -35 to -25 and the Rrn11 PAD helices α1 and α2, which contact the DNA major and minor groove, respectively (Fig. 2.16C), and form a positively charged patch (Fig. 2.16E). A proximal upstream contact is formed between the promoter region around registers -10 to -18 and elements in the Pol I protrusion and wall domains that are specific to Pol I and conserved among yeast (Fig. 2.16D). The polymerase region also binds to upstream DNA in an elongation complex (Tafur et al., 2016), although the orientation of the DNA differs. The downstream DNA duplex and the DNA-RNA hybrid are positioned in the Pol I cleft essentially as observed in the EC (Neyer et al., 2016), with the hybrid adopting a tilted conformation (Cheung and Cramer, 2011).

2.2.10. Pol I initiation on a minimal promoter

To test the functional significance of our structural observations, we established a minimal initiation assay. We used closed promoter DNA encompassing positions -38 to +24 of the natural *S. cerevisiae* rDNA promoter (Methods appendix E), purified Pol I and a 5-fold molar excess of recombinant Rrn3 and CF. Transcription depends on all factors, and initiates after the addition of NTPs and α32P-labelled ATP, to form a Pol I-specific full-length RNA product of 24 nucleotides in length (Fig. 2.17A). In our assay, the TATA box-binding protein (TBP) was not required, consistent with our structural observations and previous biochemistry (Keener et al., 1998; Pilsl et al., 2016).

To validate our assay, we used recombinant Rrn3 variants with inhibitory point mutations S145D or S185D that interfere with Pol I binding (Blattner et al., 2011; Engel

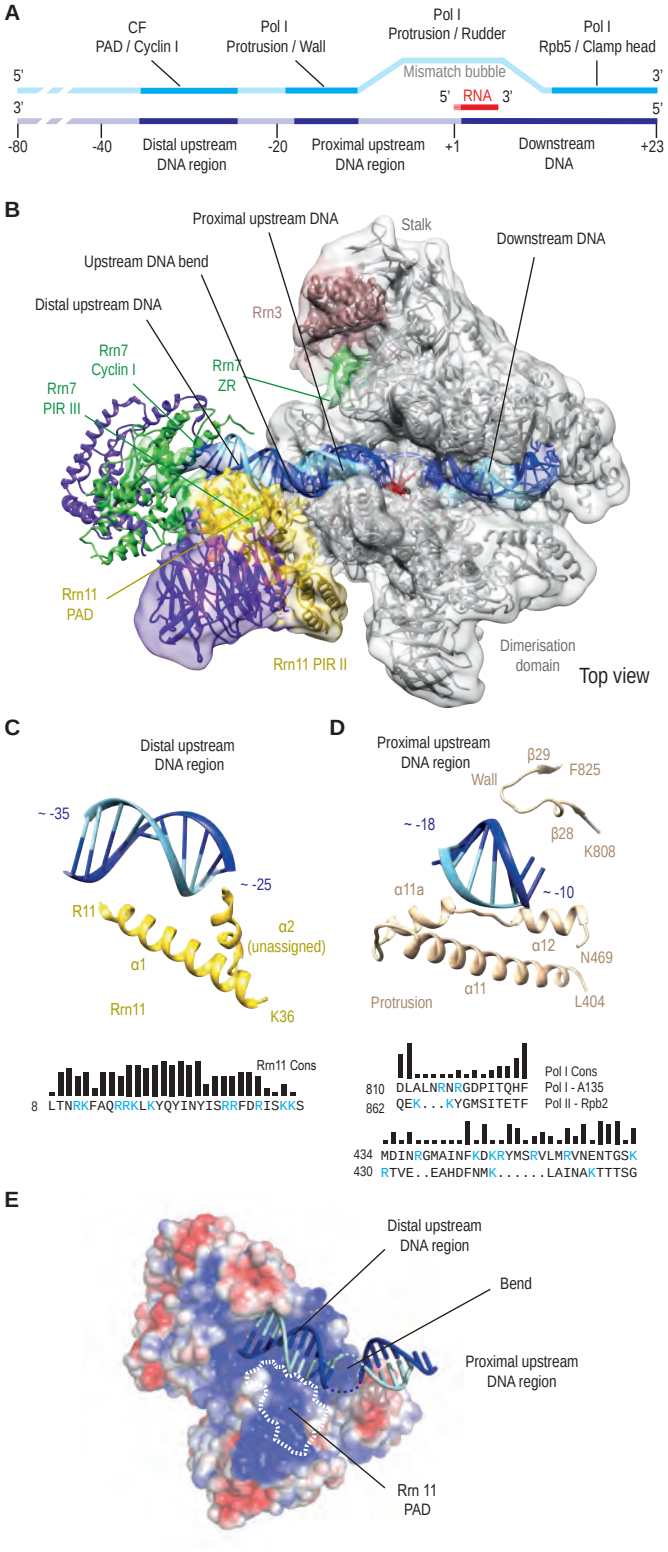


Figure 2.16. Caption on next page.

Figure 2.16. Pol I and CF contact the rDNA promoter at three distinct sites. (**A**) Schematic view of promoter contacts with the used scaffold. Semi-transparent stretches are not included in the model. Promoter- contacts are labelled according to their position relative to the TSS. (**B**) Overview of the yeast ITC in ribbon presentation with filtered electron density superimposed. Pol I (grey, apart from labelled domains), Rrn3 (brown) and DNA (blue) are shown are shown in cartoon representation. Semi-transparent density for nucleic acids (grey) is superimposed. (**C**) Distal upstream DNA extends beyond the polymerase and is recognized by PAD helices α1 and α2. Helix α2 marks the edge of straight DNA before it bends towards the polymerase. Helix α1 stretches alongside the major groove. The sequence of helix α1 is given (positively charged residues in blue) and the conservation among yeast species is indicated (Fig. C.3). (**D**) Pol I protrusion and wall domains contact the upstream promoter. A structure-based sequence alignment of *S. cerevisiae* Pol I and Pol II and conversation of Pol I residues amongst yeasts are indicated. (**E**) Surface representation of CF in a similar view as in (B) colored according to its charge as calculated with the Adaptive Poisson-Boltzmann Solver (APBS) (Baker et al., 2001) ranging from -3 kT/e (red) to +3 kT/e (blue). A positively charged patch is formed by the Rrn11 PAD and recognized promoter DNA.

et al., 2016). Indeed, these point mutations in Rrn3 abolished (S145D) or slightly reduced (S185D) Pol I initiation (Fig. 2.17B), providing a positive control. These results show that Pol I requires only recombinant Rrn3 and CF for initiation in vitro, consistent with published results (Bedwell et al., 2012; Keener et al., 1998; Pilsl et al., 2016), and demonstrating the relevance of our structural studies for understanding Pol I initiation.

2.2.11. Determinants of Pol I initiation

We next tested recombinant purified CF variants for their effect on Pol I initiation. These experiments showed that mutations in PIR I had a drastic effect, demonstrating that the Rrn7 ribbon domain is essential for initiation (Fig. 2.17C). Mutation of residues in PIR II showed weak effects, with the exception of the E351R charge reversal mutation (Fig. 2.17D), which apparently destroys salt bridges with K106 and K174 of Pol I subunit A135 that are observed in the ITC (Fig. 2.17E). A deletion of Rrn7 helix α8e of PIR III reduced initiation, but mutation of single residues in this helix had no effect (Fig. 2.17F).

We also tested which regions of the Pol I promoter were important for initiation (Fig. 2.17G, H). We inserted deletions or sequence randomizations in 10-bp windows upstream of the transcription start site. This showed that the upstream promoter region at register -1 to -30 is essential for correct initiation. The DNA region between

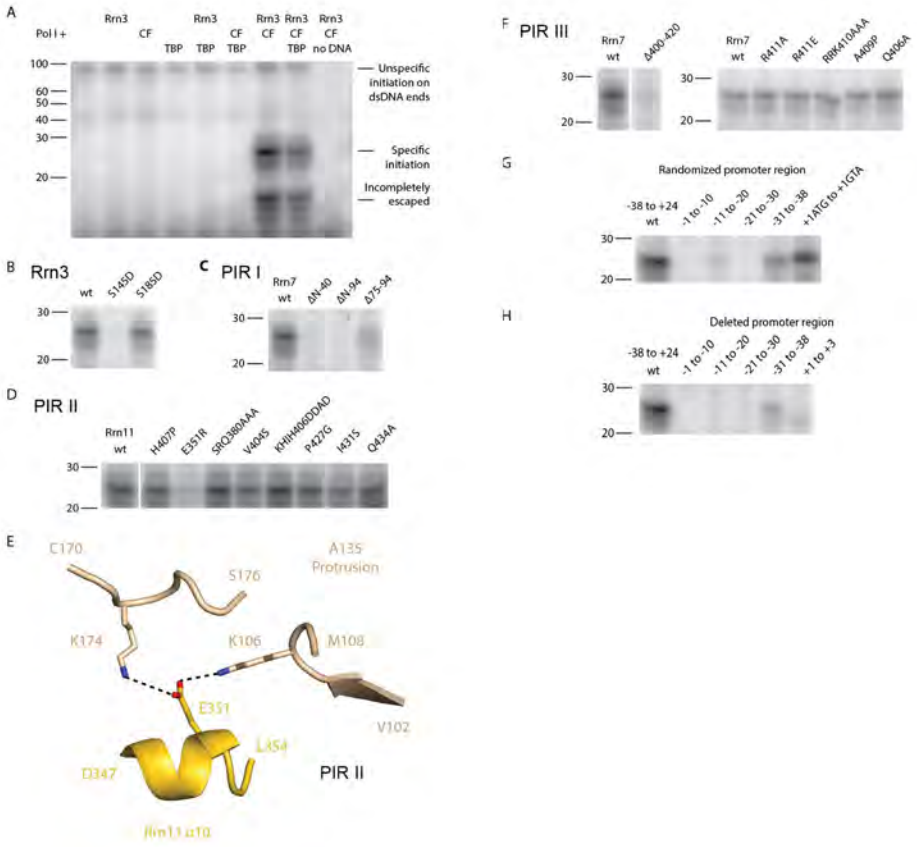


Figure 2.17. Promoter-Specific initiation outlines importance of Pol I - CF interfaces. (**A**) Urea gel resulting from Pol I-specific initiation assays (Methods appendix E). A specific initiation band at 24 nucleotides length becomes visible only if CF and Rrn3 are added. This effect is Pol I-dependent, since Pol II cannot initiate from the Pol I promoter under identical conditions (Fig. C.8). TBP slightly reduces initiation efficiency, likely due to unspecific DNA binding. RNA length in bases is displayed on the left. (**B**) Rrn3 phopho-mimetic mutants show an obstruction (S145D) or slight reduction (S185D) of initiation efficiency, likely due to interference with Pol I-Rrn3 binding (Blattner et al., 2011). (**C**) PIR I mutants show a strong reduction of initiation capabilities, underlining the importance of the ribbon contact. A partial B-linker deletion led to a reduction in initiation efficiency. (**D**) Point mutations of PIR II show little effect, with the exception of Rrn11 E351R, which reduced initiation significantly. (**E**) Environment of the mutated Rrn11 residue E351 in the ITC. Charge reversal would disturb the interface and disable interaction with protrusion lysines 106 and 174. (**F**) Point mutations in Rrn7 insertion helix α8e (PIR III) show no effect, whereas a complete deletion of the helix had an influence. (**G**) Randomization of 10 base pair stretches in the upstream region of our promoter scaffold led to an abolishment of initiation. Randomization of the -38 to -31 stretch reduces initiation, while a reversal of the first three transcribed base pairs showed only little effect.

registers -31 and -38 contributes marginally to efficient initiation. Taken together, these results strongly support our structural observations. They confirm that PIRs I-III in CF are critical for initiation, and show that the upstream DNA region that is contacted within the ITC is essential for initiation.

2.2.12. Models of the CC and OC

To obtain a complete picture of Pol I initiation, structures of the CC and OC are also required. Unfortunately, the structure of a Pol I CC cannot be obtained, because DNA spontaneously opens during complex assembly. We could however model the CC based on our ITC structure. We extended upstream promoter DNA in the ITC downstream, placing canonical B-DNA through the cleft. This leads to clashes with the protrusion and the rudder on opposite sides of the cleft. Modelling shows that these clashes can be avoided when the compacted Pol I structure in the ITC is replaced by the partially expanded structure observed in the Pol I-Rrn3 and Pol I-Rrn3-CF complexes. We therefore exchanged the compacted Pol I-Rrn3 subcomplex in the ITC against the free Pol I-Rrn3 structure (Engel et al., 2016) after superposition of polymerase active centres In the obtained CC model, DNA runs through the expanded cleft with no major clashes, and contacts Rpb5 at the downstream end of the cleft.

In the OC, the DNA template must be present in the active center to position nucleoside triphosphate substrates for initial RNA synthesis. We therefore modelled the OC by removing RNA from our ITC structure. Consistent with this approach, structures of the Pol II OC and ITC are virtually identical, except for the lack of the RNA in the OC (Plaschka et al., 2016). The obtained models of the Pol I CC and OC enabled us to rationalize the initiation-elongation transition.

2.2.13. Discussion

Here we report on a tour-de-force to elucidate the structural basis of transcription initiation by Pol I. We report the crystal structure of the heterotrimeric CF, which together with previous structures of Pol I (Engel et al., 2013; Fernández-Tornero et al., 2013) and Rrn3 (Blattner et al., 2011) completes the repertoire of high-resolution structures required to resolve Pol I initiation complexes with a structural biology hybrid approach. We then determined cryo-EM structures of the Pol I-Rrn3-CF complex and of an ITC, which additionally contained natural promoter DNA and a short initial

RNA product. The ITC structure enabled modelling of the OC, whereas a combination of the ITC and Pol I-Rrn3 structures resulted in a model of the CC. Finally, we present in vitro experimental evidence supporting the functional relevance of the structures.

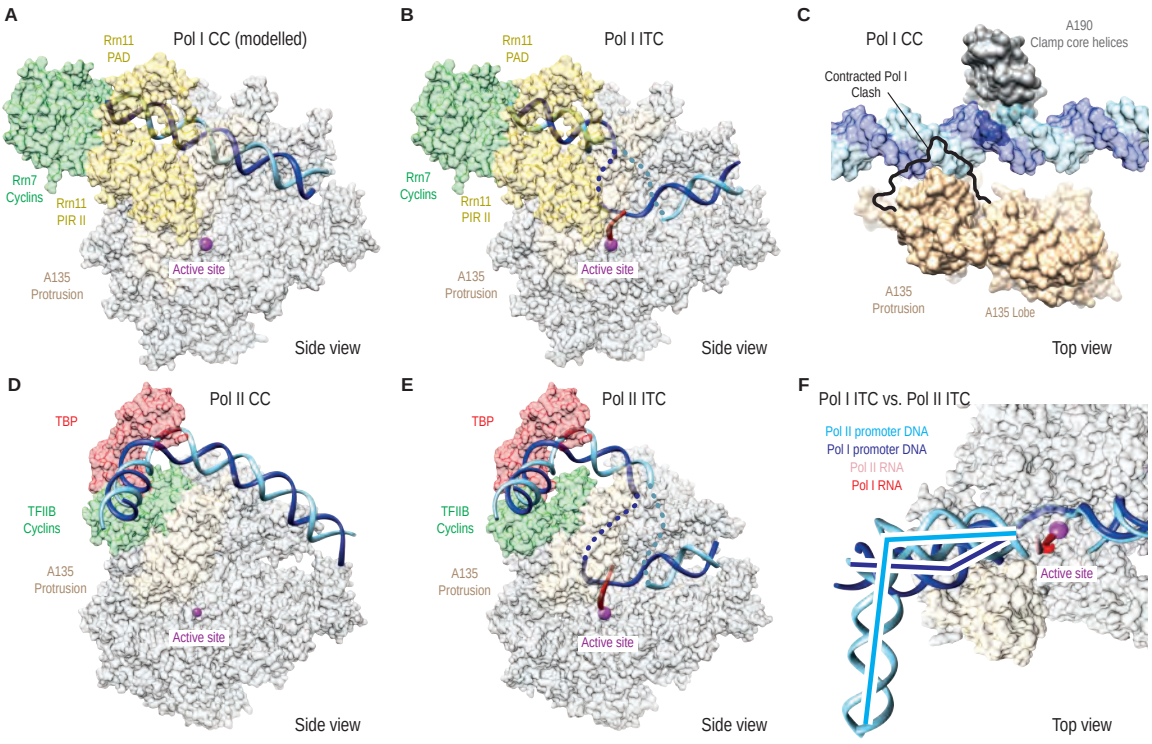


Figure 2.18. Transition from a closed to an initially transcribing Pol I complex and comparison to Pol II. (**A**) Modelled Pol I closed complex side view slice with Rrn 6 omitted for clarity. DNA (blue), protrusion (wheat), Rrn11 (yellow), Rrn7 (green) and the active site magnesium (magenta) are highlighted. Proximal and distal promoter DNA is connected with bent B-DNA generated with 3D-DART (van Dijk and Bonvin, 2009). (**B**) Pol I ITC side view slice, color code as in (A). (**C**) Top view of expanded closed complex model. A black outline marks the protrusion position assuming a contracted cleft state and shows a clash with the DNA. (**D**) Pol II closed complex (Plaschka et al., 2016). (**E**) Pol II ITC (Plaschka et al., 2015). (**F**) Superimposition of Pol I and Pol II ITC reconstructions. Pol II upstream promoter DNA (light blue) is situated on the upper rims of the cleft while its Pol I counterpart (dark blue) can enter the cleft deeper. Pol II promoter DNA (light blue trace) exits the polymerase straight and is then turned towards the core module side by TBP. In contrast, Pol I upstream promoter DNA (dark blue trace) is firstly bend around the protrusion and, secondly, kinked between Pol I and CF.

Our results converge with previous data on a model for Pol I transcription initiation and the transition to elongation (Fig. 2.18). First, Rrn3 binding stabilizes Pol I in an initiation-competent monomeric form with a largely expanded active center cleft

(Blattner et al., 2011; Engel et al., 2016; Pilsl et al., 2016). Second, CF recognizes and binds promoter DNA at a distal element upstream of the transcription start site, using its positively charged PAD domain. Third, CF uses three PIRs to dock to Pol I and to load closed promoter DNA into an expanded polymerase cleft, forming the CC. A proximal promoter region contacts Pol I-specific regions on the protrusion and wall, likely contributing to promoter recognition. Fourth, DNA opening between the protrusion and clamp enables formation of an OC with downstream DNA positioned as in the EC (Neyer et al., 2016). Spontaneous DNA opening my be rendered irreversible by trapping of the non-template DNA strand, similar to the bacterial RNA polymerase system (Feklistov and Darst, 2011; Murakami et al., 2002), because the A135 lobe can bind non-template single-stranded DNA (Tafur et al., 2016). The OC may additionally be stabilized or functionally completed by the mobile tandem winged helix domain of Pol I subunit A49 which binds proximal upstream DNA (Geiger et al., 2010; Tafur et al., 2016). Fifth, RNA synthesis commences in the OC, leading to formation of the ITC. Finally, RNA growth leads to displacement of the Rrn7 ribbon that occupies the RNA exit site, likely triggering displacement of CF and formation of the EC.

The most striking finding from our work is the unique architecture of the Pol I initiation complex, which strongly differs from its Pol II counterpart (Fig. 2.18). In the Pol I system, promoter DNA runs over the wall and through the active center cleft (Fig. 2.18A). In contrast, DNA is suspended high above the wall in the Pol II system, due to binding of TFIIB cyclin I between DNA and the wall (Fig. 2.18C). In addition, the Pol I and Pol II systems differ in the nature and direction of the bend in upstream DNA. Whereas the Pol I promoter is bend by approximately 30° around register -30 in one direction, the Pol II promoter is bend by approximately 90° at register -35 in the opposite direction (Fig. 2.18E).

These observations elucidate the molecular basis for promoter specificity and opening. We suggest that the Pol I initiation machinery can recognize a combination of the 'bendability' of upstream promoter DNA and the 'meltability' of the region around the transcription start site. Our structural observations predict that docking of a CF-DNA complex to the Pol I-Rrn3 complex is only possible when DNA can be bent in a specific way at the interface of CF and Pol I. We predict that the CF-promoter complex can only dock to Pol I if the required DNA bend can be introduced. We also predict that DNA will only open when an easy-to-melt DNA region is placed between the clamp and protrusion at the appropriate distance from the bend, providing additional promoter specificity. Only DNA that combines such bendability and meltability will be trapped

in the cleft, leading to formation of a stable OC that provides the binding energy required to compensate for the energy cost to open DNA. This model for promoter recognition based on global DNA features explains why Pol I promoter sequences are not conserved (Moss et al., 2007).

Some aspects of the Pol I initiation mechanism resemble initiation in the bacterial σ70-dependent transcription system. Although the direction and bend of upstream DNA differ in the bacterial complex (Murakami et al., 2002; Zhang et al., 2012; Zuo and Steitz, 2015), DNA opening by Pol I does not require ATP hydrolysis, may involve trapping of the non-template single strand as observed in the bacterial system (Feklistov and Darst, 2011), and may involve closure of the cleft, also as observed for bacterial RNA polymerase (Chakraborty et al., 2012). In the Pol II system, spontaneous DNA opening can occur (Plaschka et al., 2016), but may generally be impaired because DNA is suspended over the cleft, limiting interactions of downstream DNA with the polymerase, and giving rise to a requirement for TFIIH to open DNA with the use of ATP hydrolysis (Grunberg et al., 2012; Kim et al., 2000).

3. Materials and Methods

3.1. Structure of RNA polymerase I transcribing rDNA genes

3.1.1. Preparation of Pol I elongating complex

Endogenous 14-subunit Pol I was prepared from *S. cerevisiae* as described (Engel et al., 2013), with some modifications. Yeast strain CB010 expressing a C-terminal FLAG/10x histidine-tagged A190 subunit was fermented and harvested during the exponential phase. For Pol I purification, 350 g of cells were used. Proteins were precipitated overnight at 4° C with ammonium sulfate (2 M). Re-solubilized Pol I was enriched by large-scale affinity purification with Ni-NTA beads (Qiagen, Hilden, Germany). Further enrichment with anion and cation exchange chromatography yielded to pure Pol I enzyme. The sample was applied to a Superose 6 10/300 size exclusion column (GE Healthcare, Little Chalfont, UK) in 5 mM HEPES (pH 7.8), 150 mM potassium acetate, 1 mM $MgCl_2$, 10 μM $ZnCl_2$ and 10 μM β-mercaptoethanol.

DNA and RNA were purchased from IDT (Coralville, USA) and Exiqon (Vedbaek, Denmark), respectively. The nucleic acid scaffold sequences were as follows. Template DNA, 5'-AAGCTCAAGTACTTAAGCCTGGTCATTACTAGTACTGCC-3'; nontemplate DNA, 5'GGCAGTACTAGTAAACTAGTATTGAAAGTACTTGAGCTT-3'; RNA, 5'-UAUCUGCAUGUAGACCAGGC-3' (in underlined nucleosides, a methylene bridge connects the 2'-O and the 4'-C atoms of the ribose ring, thereby forming locked nucleic acids). Nucleic acids were annealed by continuously decreasing temperature from 95° C to room temperature (RT) over a period of 60 min. EC assembly was achieved by incubating Pol I (300 μg, 3.5 mg/mL) with a two-fold molar excess of scaffold for 10 min at room temperature (Fig. 2.1).

3.1.2. Single-particle cryo-electron microscopy

For single-particle cryo-electron microscopy (cryo-SP), Pol I EC complexes at a concentration of 200 μg/mL were cross-linked with 0.9 mM BS3 (Sigma Aldrich, St. Louis, USA) for 30 min at 30° C after optimization (Fig. 2.1). The reaction was stopped by adding 50 mM ammonium bicarbonate, and the sample was purified by size exclusion chromatography on a Superose 6 3.2/300 column (GE Healthcare, Little Chalfont, USA) equilibrated in 5 mM HEPES (pH 7.8), 150 mM potassium acetate, 1 mM $MgCl_2$, 10 μM $ZnCl_2$ and 10 μM β-mercaptoethanol. A 4 μL aliquot of 100 μg/mL purified sample was applied to a glow-discharged (10 s) R1.2/1.3 UltrAuFoilTM grid (Quantifoil, Grossloebichau, Germany), and plunge-frozen in liquid ethane (Vitrobot Mark IV (FEI, Hillsboro, USA) at 95 % humidity, 4° C, 8.5 s blotting time, blot force 14). Dose-fractionated movies (30 frames, 0.25 s each) were collected at a nominal magnification of 130,000x (1.05 Å/ pixel) in nanoprobe energy-filtered transmission electron microscopy (EFTEM) mode at 300 kV with a Titan Krios (FEI, Hillsboro, USA) electron microscope using a GIF Quantum SE post-column energy filter in zero loss peak mode and a K2 Summit detector (Gatan, Pleasanton, USA). The camera was operated in counting mode with a dose rate of approximately 7.5 e^-/pixel/sec and a total dose of approximately 56 $e^-/Å^2$. Defocus values ranged from -0.6 to -3 μm with marginal (<0.1 μm) astigmatism. Global motion correction was performed as described (Li et al., 2013), but cryo-SP images were not partitioned.

3.1.3. Cryo-SP image processing

Parameters of the contrast transfer function (CTF) on each micrograph were estimated with CTFFIND4 (Rohou and Grigorieff, 2015). In a first step, approximately 1500 particles were picked with the semi-automated swarm method of EMAN2's e2boxer.py (Tang et al., 2007). Relion was used for the whole-image processing workflow (Scheres, 2012) unless stated otherwise. Reference-free 2D classes were generated, seven of which were used for template-based auto-picking after filtering to 20 Å. We extracted 401,000 particles from 2300 micrographs with a 230^2 pixel box and used them for further processing. Pixels with 5 standard deviations from the mean value were replaced with values from a Gaussian distribution. All images were normalized to make the average density of the background equal to zero during pre-processing. False-positive particles showing very bright dots, which were presumably gold contamination, were removed by manual inspection or unsupervised 2D classification. The

remaining 282,000 particles were aligned on a reference generated from the PDB entry 4C2M (Engel et al., 2013) filtered to 40 Å. To correct for local motion and for radiation damage, we used the movie processing function of Relion including 'particle polishing' (Scheres, 2012). Local resolution was estimated as described (Cardone et al., 2013; Plaschka et al., 2015).

During classification of cryo-SP images (Fig. 2.3), we first separated out particles lacking nucleic acids. To this end, the Pol I cleft of the average resulting after the first round of alignment was masked. The subsequent classification led to 4 classes: (1) nucleic acid-free Pol I (115,000 particles); (2) Pol I elongation complex (94,000 particles; hereafter referred to as 'EC'); (3) Pol I elongation complex with an alternative DNA conformation (37,000 particles); and (4) other particles (35,000 particles). Among the nucleic acid-free polymerase particles, 80,000 particles displayed a defined position of the C-terminal domain of A12.2. We refer to the SP average of these particles as the 'monomer'.

In a second step, a mask around the dimerization domain was applied to remove particles from which the A49/34.5 subcomplex dissociated. This led to 32,000 and 40,000 particles in case of the Pol I monomer and Pol I EC, respectively. To visualize the mobile stalk, we then applied a mask around A14/43 during refinement allowing only local searches.

Gold-standard Fourier Shell Correlations (FSCs) were calculated during the 3D refinement in Relion between two independently refined halves of the data. According to the FSC 0.143 criterion, global resolutions of 4.0 Å and 3.8 Å were estimated for Pol I monomer and EC structures, respectively, which were sharpened with temperature factors of -146 $Å^2$ and -149 $Å^2$, respectively.

3.1.4. Structural modelling

Two separate models were built for the monomer and the Pol I EC. PDB entry 4C2M (Engel et al., 2013) was used as the starting model in both cases. Models were constructed lacking the expander, connector and, in case of the EC, the C-terminal domain of A12.2. The models were further truncated by removing the peripheral subcomplexes A49/34.5 and A14/43. The starting models were placed in densities for the monomer and the EC by fitting in UCSF Chimera (Pettersen et al., 2004), followed by rigid body fitting with a Phenix real space refinement (Adams et al., 2010). Rigid body groups were defined based on module definitions originally proposed for Pol II (Cramer et al.,

2001). A starting model for DNA and RNA was derived from bovine Pol II (Bernecky et al., 2016) and further refined. Structurally altered regions were adjusted to the density in COOT (Emsley et al., 2010) followed by real space refinement in Phenix. To generate complete models, structure of subcomplexes A49/34.5 and A14/43 were fit into the classified map in Chimera. No changes were made within the domains during model building, except for A34.5 C-terminal tail. The models were validated using the FSC between the model and the map, EMRinger (Barad et al., 2015) and Molprobity (Chen et al., 2010).

3.2. Structural basis of RNA polymerase I transcription initiation

3.2.1. Preparation of Pol I ITC complexes

Pol I, CF, Rrn3, and TBP were purified separately and reconstituted in vitro on a promoter scaffold. *S. cerevisiae* Pol I was purified from endogenous material as described before (Neyer et al., 2016). All initiation factors were expressed recombinantly in *E. coli*. CF was purified as described above and Rrn3 and TBP as described before (Blattner et al., 2011; Plaschka et al., 2016). Nucleic acids were purchased from IDT (Coralville, USA) with these sequences: Template DNA, 5'- CTTGTCTTCAACTGCTTTCGCATGAAGTACCTCCCAACTACTTTTCCTCACACTTGTACTCCATGACTAAACCCCCCCTCCCATTACAAACTAAAATCTTACT-3'; nontemplate DNA, 5'- AGTAAGATTTTAGTTTGTAATGGGAGGGGGGGTTTAGTCATGGAGTACAAGTGTGAGGAAAAGTAGTTGGGACAAGTGCTTGCATCGTGCAGTTGAAGACAAG-3'; RNA, 5'- AUGCGA-3'. Nucleic acids were annealed in water in a 1:1:1 ratio by continuously decreasing temperature from 95° C to room temperature over a period of 60 minutes. The Pol I-Rrn3 complex was pre-assembled by incubating polymerase with a 5-fold excess of Rrn3 at 4° C overnight. On the next day, 5-fold excess of CF, and 1.2-fold excess of DNA-RNA scaffold and TBP were added. The sample was incubated at room temperature for 20 minute, followed by an hour on ice. Stochiometric ITC eluted as a homogenous peak from a Superose 6 3.2/300 size exclusion column (GE Healthcare, USA) in complex preparation buffer (5 mM HEPES pH 7.8, 150 mM potassium acetate, 1 mM $MgCl_2$, 10 μM $ZnCl_2$, 5 % v/v glycerol, 10 μM β-mercaptoethanol). The peak fractions (Fig.2.14) were pooled, and cross-linked with 0.1 % glutaraldehyde for 30 minutes at 4° C before the reaction was quenched with a mixture of 2.5 mM aspartate and lysine. After 5 minutes incubation at 4° C, ammonium-bicarbonate was added to a final concentration of 100 mM. After cross-linking and quenching, the buffer was exchanged to sample buffer (complex preparation buffer lacking glycerol) using Micro P30 Bio-Spin columns (Bio-Rad, USA) and filtered with 4 mm PES Captiva syringe filters with a pore size of 0.2 μm (Agilent, USA). Final complex concentration of 0.13 mg/mL was achieved by reducing the sample volume in Vivaspin 500 with a 100 kDa cut-off (Sartorius, Germany).

3.2.2. ITC single-particle cryo-EM and image processing

A 4 μL aliquot of purified sample was applied to a glow-discharged R2/2 UltrAuFoilTM grid (Quantifoil, Germany), and plunge-frozen in liquid ethane (Vitrobot Mark IV (FEI, USA) at 95 % humidity, 4° C, 8.5 s blotting time, blot force 14). Dose-fractionated movies (24 frames, approximately 60 $e^-/Å^2$ total dose) were collected at a nominal magnification of 95,000x (1.13 Å/pixel) in nanoprobe mode at 300 kV with a Titan Krios (FEI, USA) electron microscope using a Falcon2 detector (FEI, USA). Defocus values ranged from -0.8 to -2.5 μm with marginal ($<$0.1 μm) astigmatism. Global motion correction and dose weighting was performed using Unblur (Brilot et al., 2012).

Parameters of the contrast transfer function (CTF) of each micrograph were estimated with CTFFIND4 (Rohou and Grigorieff, 2015). In a first step, approximately 20,000 particles were picked with the semi-automated swarm method of EMAN2's e2boxer.py (Tang et al., 2007). Relion was used for the whole-image processing workflow (Scheres, 2012) unless stated otherwise. Reference-free 2D classes were generated, ten of which were used for template-based auto-picking after filtering to 20 Å. We extracted 2,498,000 particles from 6,000 micrographs with a 330^2 pixel box and used them for further processing. Pixels with more than 5 standard deviations from the mean value were replaced with values from a Gaussian distribution. All images were normalized to make the average density of the background equal to zero during preprocessing. False-positive particles showing very bright dots, which were presumably gold contamination, were removed by manual inspection or unsupervised 2D classification. The remaining 1,328,000 particles were aligned to a reference generated from one Pol I molecule of the PDB entry 4C2M (Engel et al., 2013) filtered to 40 Å. According to the 0.143 FSC criteria, the initial reconstruction reached an overall resolution of 3.5 Å (Fig. 2.15).

During classification (Fig. 2.14), we first separated 371 000 particles lacking CF with or without Rrn3 and DNA. The resulting reconstruction of a Pol I-Rrn3-DNA complex was resolved to 3.8 Å (0.143 FSC criteria, not shown). 957 000 particles contained CF. This subset was also classified based on DNA and Rrn3 occupancy, resulting in a sub-set of 567,000 complete particles. The corresponding ITC reconstruction reaches an overall resolution of 4.2 Å (0.143 FSC criteria). For visualization, the density was either filtered to its nominal resolution or sharpened with a B-factor of -315 $Å^2$. Local resolution (Fig. 2.15) was estimated as described before (Plaschka et al., 2015).

3.2.3. Structural modelling of the ITC and CC

To generate a model of the Pol I ITC, we used the known crystal structures of Pol I (PDB 4C2M), Rrn3 (3TJ1) and CF (this study) as starting point (Blattner et al., 2011; Engel et al., 2013). The flexible expander, connector and C-terminal domain of A12.2 were removed from the Pol I structure. Rigid body groups were defined as before (Engel et al., 2013) and rigid body fitted using Phenix real space refinement (Adams et al., 2010). The starting coordinates for the downstream DNA and RNA were extracted from the Pol I EC (Neyer et al., 2016). Since upstream DNA did not allow one to distinguish individual nucleotides, we modelled its straight parts with B-DNA generated and rigid body-fitted in Coot (Emsley et al., 2010). We had included TBP in our sample preparation, but did not observe any corresponding cryo-EM density. Due to flexibility, density quality of CF areas varied highly. While the overall volume is large enough and has the correct shape to accommodate CF, not all parts are well resolved. In particular, density for the distal CF module II is poor, whereas density for module I is very good, revealing secondary structure elements. Therefore, we fitted CF elements as rigid bodies based on the location of well resolved domains, resulting in the model displayed in Fig. 2.13. However, the PDB coordinates were restricted to well-ordered domains presenting as defined secondary structures in our reconstruction.

For modelling of the closed promoter DNA complex (CC) of Pol I, canonical straight B-DNA was generated in COOT (Emsley et al., 2010) and placed in an ITC such that DNA superimposed with DNA in the ITC at the distal upstream PAD contact and such contacts a conserved TPSA motif in Rpb5 (Bernecky et al., 2016). We then overlayed the Pol I-Rrn3 structure (PDB code 5G5L) via its active center in order to avoid a clash between proximal upstream promoter DNA and the A135 protrusion domain by using the partially expanded cleft confirmation (Engel et al., 2016). In order to compare CCs and ITCs of Pol I and Pol II, complexes were superimposed via their active site regions using Chimera (Pettersen et al., 2004). PDB codes of the superimposed Pol II CC and ITC were 5FZ5 (Plaschka et al., 2016) and 4V1N (Plaschka et al., 2015), respectively.

4. Conclusion and Outlook

With the determination of the Pol I crystal structure (Engel et al., 2013; Fernández-Tornero et al., 2013), high-resolution structural information of a eukaryotic polymerase other than Pol II became available for the first time. The Pol I crystals contained inactive dimers suggesting that an additional regulatory event is required to yield an active enzyme. It was recently shown how Rrn3 converts Pol I dimers into an initiation competent Pol I monomer (Engel et al., 2016; Pilsl et al., 2016).

We now expanded our structural knowledge of the Pol I transcription cycle by a series of additional steps (Figure 4.1 and 1.2). The Pol I dimer is in an equilibrium with its monomeric form and we probed for conformational changes that result from this transition. The crystal structure of the CF is now available, which, together with Pol I and Rrn3, completes the set of high-resolution structures of the pre-initiation complex. The model of the closed initiation complex was derived by analysing the DNA binding interface of CF and by comparison with the Pol II system. We further solved an initially transcribing complex that allowed us to discuss DNA opening. The structure of elongating Pol I completes the set of transcribing eukaryotic polymerases. Transition from inactive to active Pol I was characterized by a change in the cleft width, which was confirmed by *ex vivo* cryo-ET data.

4.1. Completion of transcription cycle snapshots

In a next step, the models of closed and open initiation complexes need to be supported by additional experimental data. Determination of the open complex structure by cryo-EM promises to be comparable to structure determination of the ITC as sample preparation differs only by the presence of RNA. As we discovered in this study, Pol I positions the promoter DNA about 25 Å deeper in the cleft during initiation as expected. Consequently, the size of the mismatch bubble should be decreased.

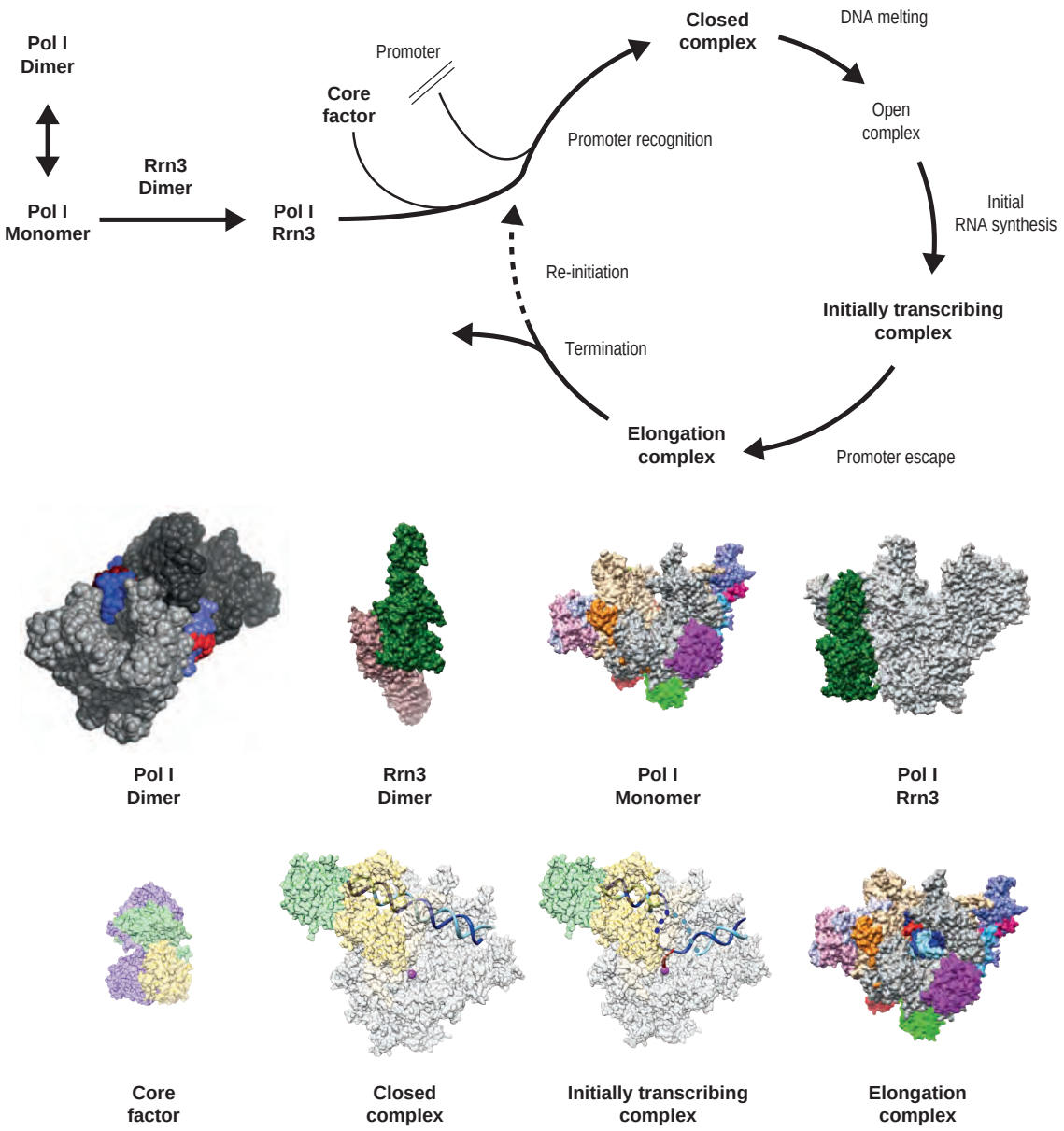


Figure 4.1. Transcription cycle. Scheme of the Pol I transcription cycle. States with known structures are marked with bold font. The crystal structure of Pol I was solved as an inactive dimer (Engel et al., 2013; Fernández-Tornero et al., 2013) ('Pol I Dimer' panel adopted from Engel et al. (2013)) and the Rrn3 crystal structure also revealed dimers (Blattner et al., 2011). These two homo-dimers form an initiation competent Pol I-Rrn3 hetero-dimer and this complex was solved be cryo-EM (Engel et al., 2016; Pilsl et al., 2016). Apart from Rrn3, CF is also involved in promoter recognition. Subsequent to promoter recognition, the transition to an open complex occurs via DNA melting. Then, the initially transcribing complex escapes the promoter and productive elongation occurs followed by either re-initiation or termination.

Additionally, a series of structures with differently sized mismatch bubbles, will allow monitoring of DNA opening and the conformational changes accompanying it. As the Pol I initiation complex is capable of opening DNA without ATP hydrolysis, such a study of sequential DNA opening faces the same challenge as the study of a closed complex. In both cases, the enzyme has to be trapped. In bacteria, rudder deletions were shown to severely impair transcription initiation (Kuznedelov et al., 2002) and our models suggest that the same might be the case in the Pol I system. *In vitro* experiments revealed that the tandem winged helix is essential for initiation but not for elongation (Pilsl et al., 2016). Thus, rudder or tandem winged helix deletions may allow sample preparation of a defined state.

Structural studies of transcriptional termination are challenging, as the elongation complex has to be destabilized in order to release its template and product. A Reb1 binding site follows the T-rich sequence at the end of rDNA genes. One model of termination assumes that Reb 1 functions as a road block on DNA and the T-stretch causes destabilization of the DNA-RNA hybrid leading to dissociation of the elongation complex (Nemeth et al., 2013). The crystal structure of Reb1 together with its termination sequence supports this hypothesis and in the same study direct interaction between Reb1 and A12.2 was discovered (Jaiswal et al., 2016). With that knowledge further structural studies of Pol I transcription termination become feasible. This is particularly promising as the Pol III subunit C11, which is homologous to A12.2, is involved in termination (Arimbasseri et al., 2013).

4.2. Upstream activating factor as part of the initiation complex

The general transcription factors Rrn3 and CF are sufficient to initiate Pol I transcription *in vitro* (Moss et al., 2007). Therefore, we limited our studies to this core initiation complexes in context of promoter DNA. *In vivo* an additional upstream activating sequence (UAS) from -60 to -150 (+1 corresponds to the transcription start site) stimulates transcription (Bordi et al., 2001; Keys et al., 1996). The hexameric upstream activating factor (UAF) consists of Rrn5, Rrn9, Rrn10, Uaf30, H3 and H4 (Keener et al., 1997). Protein interactions are mediated between UAF, CF and TBP, but not directly between UAF and Pol I (Knutson and Hahn, 2013). The UAF subunit Uaf30 facilitates UAS binding in the presence or absence of other factors (Hontz

et al., 2008), but the mechanism of how UAF stimulates transcription remains elusive. Available purification protocols for endogenous UAF (Keener et al., 1998) can be optimized for structural studies and then used to assemble a complete initiation complex. The development of baculovirus systems may also allow for recombinant UAF expression in insect cells, bearing the advantage of higher flexibility if mutations are desired (Berger et al., 2004).

Besides increasing Pol I activity, UAF also silences rDNA gene transcription by Pol II and its absence alters the chromatin architecture (Goetze et al., 2010). A structure of UAF in complex with the Pol I transcription machinery may also elucidate how the polymerase switch (Vu et al., 1999) is prevented.

4.3. In-depth elongation studies

The EC has to be highly processive in order to transcribe complete genes without premature termination. The transcription factor Spt5, NusG in bacteria, dimerizes with Spt4 and its architecture on Pol II was modelled (Martinez-Rucobo et al., 2011). The NGN domain of Spt5 binds at the clamp core helices, thereby locking the nucleic acids inside the cleft and stabilizing the EC. Spt4/5 was further shown to play a role in Pol I transcription (Schneider et al., 2006). We can now test structurally whether the conserved transcription factor Spt4/5 supports Pol I transcription in the same way as it does for Pol II activity.

During processive elongation Pol I may incorporate a wrong nucleotide that does not base-pair with its template according to Watson-Crick. To prevent release of RNA with sequence mismatches, Pol I possesses a built-in proofreading mechanism, the RNA cleavage activity of A12.2 (Kettenberger et al., 2003). Additionally, stalled and backtracked Pol I also depends on A12.2 to continue elongation. Stalling may happen as processing of the rRNA is extensively coupled to transcription and both influence each other (Woolford and Baserga, 2013). Here we showed the correlation of cleft contraction with activity states of Pol I. When Pol I binds a perfect DNA-RNA hybrid, cleft contraction also narrows the pore that accommodates the C-terminal domain of A12.2. Thus, RNA cleavage is prevented. After incorporation of an incorrect nucleotide, the hybrid will be distorted and might then cause widening of the pore to allow the catalytic domain of A12.2 to reach into the active site. As partial RNA cleavage would lead to a heterogeneous mixture of complex state, A12.2 has to be

inactivated. Point mutations at the tip of the catalytic loop of TFIIS, the Pol II homologue of A12.2, were demonstrated to prevent cleavage but maintain the structure of its cleavage activating domain (Cheung and Cramer, 2011). Yeast strains harbouring analogous mutations in A12.2 can be used to test our hypothesis.

4.4. Rationalized drug design for cancer therapy

Cancer cells are characterized by uncontrolled cell growth that requires highly increased protein synthesis. Consequently, ribosome biogenesis, and in turn also Pol I transcription, is up-regulated. Aggressive cancer cells even display enlarged nucleoli (Derenzini et al., 2000). This makes the Pol I transcription machinery a potential target for anti-cancer therapeutics (Poortinga et al., 2015). CX-5461 is such a drug in clinical trials. It prevents the interaction of SL-1, the human CF homologue, with promoter DNA and impairs Pol I activity approximately 350-fold over Pol II or Pol III activity (Drygin et al., 2011). We can test if CX-5461 is not only active in human cells but also in yeast and further determine the binding affinity between CF and CX-5461. The newly derived structure of the CF may help to discover the molecular mechanism of CX-5461 action.

Based on the ITC structure the best suited DNA construct can be designed for co-crystallization of CF with DNA. A high-resolution structure of CF in complex with DNA would give valuable insights into promoter recognition and this interaction can be disrupted in clinical applications.

4.5. Pol I transcription in its natural environment

Continuous improvements in cryo-EM image processing, data collection and sample preparation, gradually improve the resolution achievable from subtomogram averaging. With approximately 9 Å resolution, the mammalian ribosome is one of the best resolved asymmetric complexes so far (Pfeffer et al., 2015). The resolution of our subtomogram average of Pols on Miller spreads is limited by particle number. Increasing particle number is not trivial, as preparing Miller spreads suitable for data collection is very challenging. Generally, Miller spreads are most frequently prepared from *Xenopus laevis* oocytes as their high transcriptional activity eases sample preparation (Miller and Beatty, 1969). Here we used Miller spreads from *S. cerevisiae*, which allowed

us to directly compare between cryo-ET and cryo-SP reconstructions from the same organism. Switching to *X. laevis* might improve formation of Miller spreads suitable for imaging, thereby overcoming a current bottleneck.

The characteristic shape of Miller spreads allows one to correlate positions in the tomogram with its rDNA sequence, if the field of view is sufficiently large. With that information, Pols can be sorted depending on their genomic position. As Pol I transcription initiation is not only observed from the rDNA promoter but also from promoters within the intergenic spacer region (Mayer et al., 2006), one could thereby assess whether the conformation of the enzymes differ.

The cryo-ET polymerase reconstruction shown here did not reveal defined additional features accounting for accompanying transcription factors. These binding partners must either have dissociated during sample preparation or are highly dynamic. Optimizing buffer conditions or fixation by covalent cross-linking might help to trap distinct Pol I complexes on a Miller spread.

Ultimately tomographic studies can elucidate higher ordered Pol I transcription units which are disrupted upon spreading. To overcome these limitations, the nucleolus or even the cell is frozen as a whole. Focused ion beam milling then generates a thin lamella at the region of interest that allows imaging (Narayan and Subramaniam, 2015). While the thick sample is non-transparent, the area of interest has to be chosen by complementary techniques. Fluorescent labelling of Pol I would allow for correlative cryo-fluorescence light microscopy and cryo-electron tomography (Koning et al., 2014).

A. Bivalent cross-link analysis

Results presented in this section are published.

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

A.1. Abstract

Site specific cross-linking coupled to mass spectrometry (XL-MS) has the power to bridge between structures at high and low resolution. In many cases it is very challenging to determine the structure of huge biological assemblies to high resolution. Although cryo-EM improved dramatically during the last years, *de novo* building of structures is possible in no more than a few cases. Commonly, high-resolution structures of sub-domains are solved with x-ray crystallography and then fitted into the EM density. Depending on the resolution of the reference map, additional restraints are required to unambiguously place the domain. In such an integrative approach, XL-MS is frequently used (Leitner et al., 2016).

A typical XL-MS experiment results in a long list of distant restraints. To simplify cross-link interpretation, we wrote a Matlab based script that converts a list of cross-links to a vector-based diagram. Additionally, it is possible to export a list without redundant distance restraints.

Today, more sophisticated tools are available. Common examples are xVIS (Grimm et al., 2015), xTract (Walzthoeni et al., 2015) and xiNET (Combe et al., 2015).

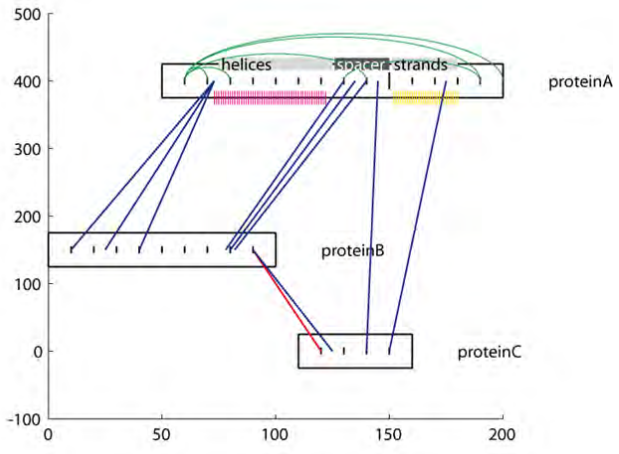


Figure A.1. Example of cross-link map. The trimeric complex proteinA/B/C is shown. Intra and inter cross-links are visualized as green and blue lines, respectively, while possible candidates are coloured in red. For proteinA, domains are highlighted. Magenta and yellow lines represent α-helices and β-strands, respectively. The axes are typically suppressed but can be used to determine the position of the boxes representing complex subunits. X- and y-values have to be provided in the *default* file. They correspond to the middle left edge of the box. It is recommended to set these values randomly in the first place. For a second round, optimized positions can be determined by moving the boxes in a vector based graphics suite.

A.2. Data preparation and input files

As input files, tab delimited *.txt files have to be placed in the folder 'input'. The names of these files should reflect what is contained in the variable *file* within the main script (BiClAn). The input files are formatted in the way that the third and fourth column state the position of cross-links of the subunits given in column one and two, respectively. The fifth column allows one to mark a cross-link, e.g. as a candidate and allows it to be drawn in a different manner (Figure A.2). Please note, that inter and intra cross-links have to be provided in separate input files. The protein names may not include spaces. This is true for all input files.

The variable *default* loads a tab delimited file which sets the name of the proteins and allows the subunits to be treated separately (Figure A.2). Please note that you have to give the name of the proteins in descending alphabetical order (as they are called in *file*). Two additional rows are needed at the end of the document. The easiest way to generate this file is to use a spreadsheet program and save as tab delimited *.txt.

demo_inter1.txt

Z0Z0Z0_protein_A	Y0Y0Y0_protein_B	23	10	
Z0Z0Z0_protein_A	Y0Y0Y0_protein_B	23	25	
Z0Z0Z0_protein_A	Y0Y0Y0_protein_B	23	40	
Z0Z0Z0_protein_A	Y0Y0Y0_protein_B	80	78	
Z0Z0Z0_protein_A	Y0Y0Y0_protein_B	85	80	
Z0Z0Z0_protein_A	Y0Y0Y0_protein_B	90	82	
Z0Z0Z0_protein_A	X0X0X0_protein_C	95	30	
X0X0X0_protein_C	Z0Z0Z0_protein_A	40	125	
Y0Y0Y0_protein_B	X0X0X0_protein_C	90	10	candidate
Z0Z0Z0_protein_A	Y0Y0Y0_protein_B	23	25	candidate

demo_inter2.txt

Y0Y0Y0_protein_B	X0X0X0_protein_C	90	15	
Y0Y0Y0_protein_B	X0X0X0_protein_C	90	15	candidate

demo_intra.txt

Z0Z0Z0_protein_A	Z0Z0Z0_protein_A	10	20
Z0Z0Z0_protein_A	Z0Z0Z0_protein_A	10	30
Z0Z0Z0_protein_A	Z0Z0Z0_protein_A	10	150
Z0Z0Z0_protein_A	Z0Z0Z0_protein_A	10	140
Z0Z0Z0_protein_A	Z0Z0Z0_protein_A	10	90
Z0Z0Z0_protein_A	Z0Z0Z0_protein_A	80	140

default.txt

	A	B	C	D	E	F	G	H	I
1	proteinC	50	1	1	0	0	0	110	0
2	proteinB	100	1	1	0	0	0	0	150
3	proteinA	150	1	1	1	1	1	50	400
4	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0
	name	*# aa*	*box?*	*inter?*	*intra?*	*SecStruc?*	*domain?*	*x*	*y*

proteinA_sec.txt

CCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCC
CCCCCCCCCCCCCCCCCCCCCCCCEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEECCCCCCCCCCCCCCCCCCCCCC

proteinA_dom.txt

alpha_helices	helices	25	75
spacer spacer	76	100	
beta-strands	strands	101	130

Figure A.2. Input files. All files required to generate the demo diagram are displayed. As with the others, *default.txt* is a tab delimited file. In this figure a screenshot from a spreadsheet program is shown to highlight the meaning of each column. However, it was saved as *.txt later on.

The number of amino acids is given in the second column. Although no axes are displayed, the diagram is an x-y-coordinate system (Figure A.1). The value in column eight and nine sets the x- and y-value, respectively. Whether an operation has ('1') or has not ('0') to be executed for a specific subunit this can be set in the residual columns.

Before drawing distance restraints, the user has to choose which input files to use. As shown in Figure A.3 , this is done by handing over the index of the input file to variable 'i'. By executing these functions one by one, cross-links from different experiments can be drawn in different colors to make them distinguishable.

```
%%%%%%%%%%%%%%%%%%%
%%% input files %%%
%%%%%%%%%%%%%%%%%%%
file = {'demo_inter1'   % 1
    'demo_inter2'       % 2
    'demo_intra'        % 3
    }

---

%%%%%%%%%%%%%%%%%%%%%%%%
%%% intra cross-links %%%
%%%%%%%%%%%%%%%%%%%%%%%%
i=[3]
param.loop='g';      %color of intra cross links
BiClIntra(param,sorted,file,intra,x,y,i)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% inter cross links from selected subunits to selected subunits only %%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
i=[]
param.candidate='r';  % color and style of candidates
param.valid='b';    % color of valid cross links
param.width=1;
BiClInter1(param,sorted,file,inter,x,y,i,conc)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% inter cross links from selected subunits to all subunits %%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
i=[1 2]
param.candidate='r';  % color and style of candidates
param.valid='b';    % color of valid cross links
param.width=1;BiClInter2(param,sorted,file,inter,x,y,i,conc)
```

Figure A.3. Script excerpt. Parts of the BiClAn script are shown to demonstrate how to choose which dataset (loaded in file) has to be drawn. The variable 'i' sets the index of the respective input file. In this example intra cross-links of 'demo_intra' are drawn. Inter cross-links from selected subunits to all subunits are drawn in case of 'demo_inter1' and 'demo_inter1'.

A.3. Domains and secondary structure

As optional features, domain architecture and secondary structure information of subunits can be displayed. In this case input files have to follow a specific nomenclature. Starting with the name as given in default, _sec or _dom is used as suffix for a tab-delimited *.txt file. H, E and C represent helices, strands and regions neither forming helices nor strands, respectively. Domain information is given in four columns. Column one will be neglected by BiClAn, but may not contain spaces. The second column states the name of the domain, whose boundaries are set in column three and four.

A.4. Output files

The generated MATLAB figure can be saved as a pixel based image (e.g. *.png or *.jpg) as well as a vector based graphic (e.g. *.eps or *.ai). By changing the size of the window displaying your diagram you also change the size of the saved image. If the figures are saved as a graphic, clipping masks might be added. It is possible to remove those in a vector graphic program. The variable 'sorted' contains (for each input file individually) a table of unique cross-links. The names of the involved subunits are represented in column one and two by the indices as indicated in default (Figure A.2). The respective amino acid number is given in column three and four. Whether a cross-link was labelled as a candidate is displayed in column three.

B. Extended Figures of the Elongation study

Figures presented in this section are published and were obtained in a collaboration with the Frangakis lab. Author contributions are stated on page VI.

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

* These authors contributed equally to this work.

The following section presents supplemental figures that were predominantly obtained by the collaborators. Results and methods are presented in chapter 2.1 and 3.1, respectively. Methods, which were predominantly used by the collaborators, are described in appendix D.

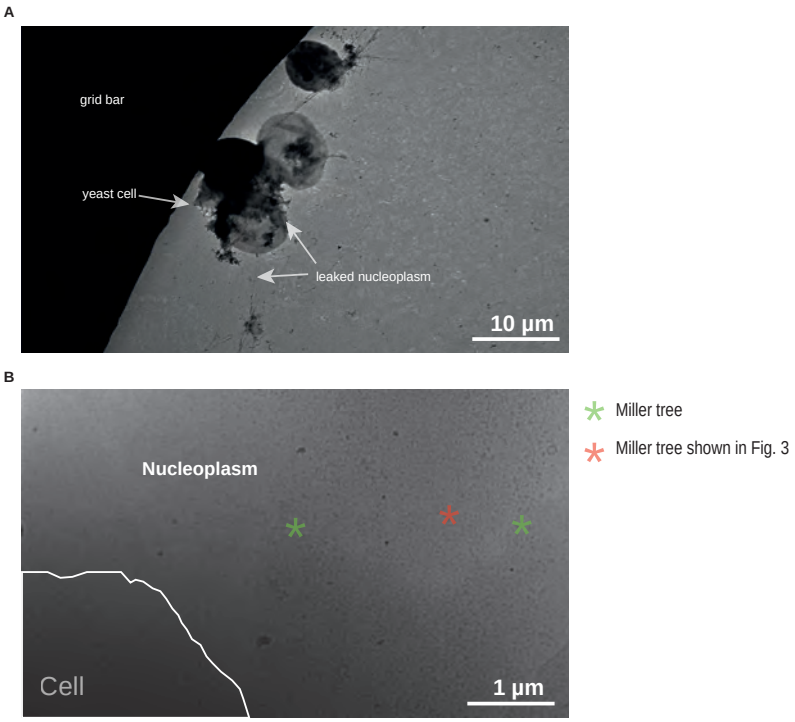


Figure B.1. Yeast cells, lysed to leak their nucleoplasm, prepared with negative stain and visualized under cryo conditions. **(a)** Electron micrograph of a negatively stained lysed yeast cell, with the nucleoplasm spread on the carbon support film. The upper left of the micrograph is occupied by the grid bar. The yeast cell has released the nuclear context on the grid, which appears as an electron-lucent leakage. **(b)** Electron micrograph of the leaked nucleoplasm of a plunge-frozen yeast cell at close-to-native conditions. In the lower left corner, the yeast cell can be seen as an electron-dense patch. The nucleoplasm is embedded in an ice layer and the asterisks indicate three Miller trees found in the vicinity of this cell. The Miller tree indicated with the red asterisk was used for recording of the tilt-series in Fig. 2.8.

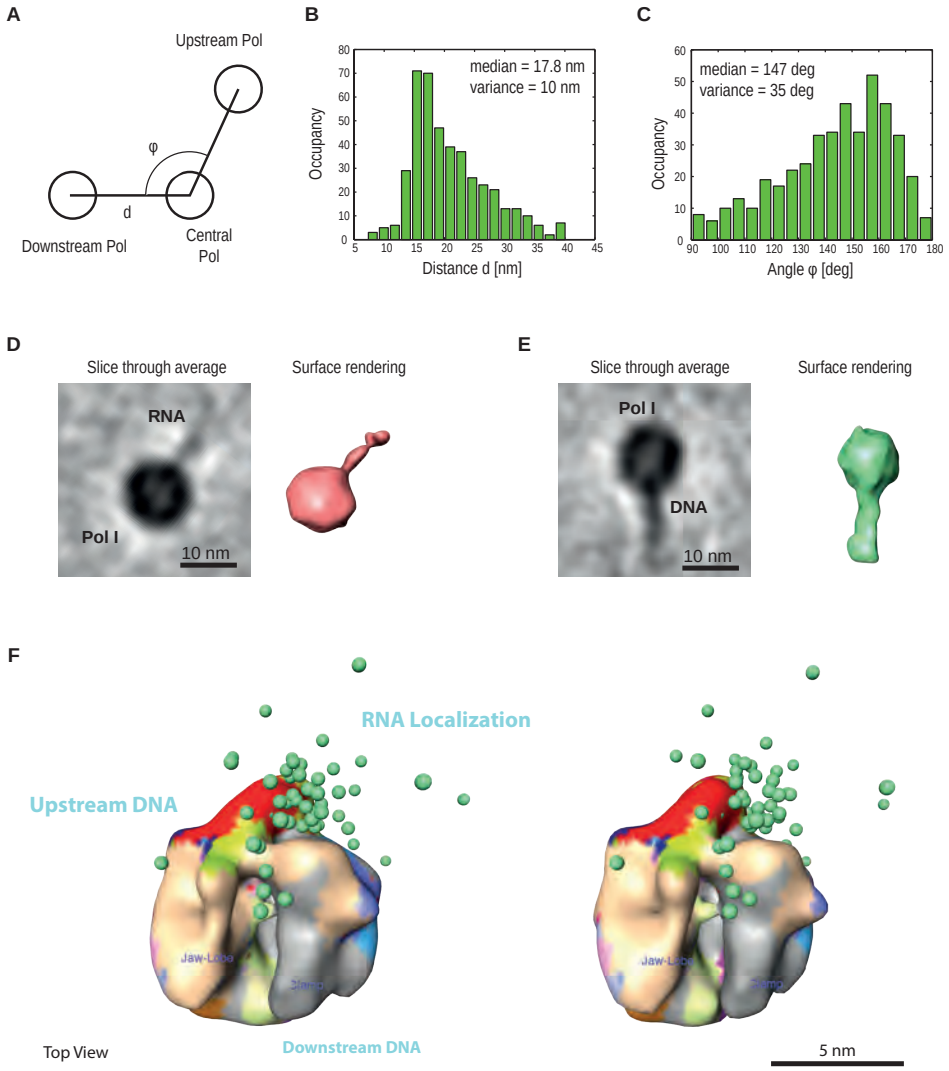


Figure B.2. Relative positions of polymerases towards each other and of protruding nucleic acids. **(a)** Schematic of three consecutive Pol I enzymes as seen in the tomogram. **(b)** Histogram of center-center distance d of two consecutive Pols as depicted in (a). **(c)** Histogram of in-plane angle ϕ spanned by three consecutive Pols as depicted in (a). **(d)** Focused sub-tomogram averaging around the RNA. The RNA exits Pol I as a approximately 10 Å thick density, both in the slice and in the isosurface representation. **(e)** Sub-tomogram average with the alignment focused on the downstream DNA. The downstream DNA is a long, straight 2 nm density, both in the slice and in the isosurface representation. In both (d) and (e), the Pol I molecule is a globular approximately 12 nm featureless density. **(f)** Stereo pair of the sub-tomogram average shows the positions of the nascent RNA chain as green balls. The positions that were manually identified by three independent users, without previous knowledge of the positions of the sub-tomogram average, correspond closely to the position of the RNA exit site that was postulated by the X-ray crystallography structure.

C. Extended Figures and Tables of the Initiation study

Figures and Tables presented in this section are in preparation for publication and were obtained in a collaboration with Tobias Gubbey and Christoph Engel. Author contributions are stated on page VII.

C. Engel*, T. Gubbey*, **S. Neyer***, S. J. Sainsbury, C. Oberthür, C. Bäjen, C. Bernecky, P. Cramer. (2016) Structural basis of RNA polymerase I transcription initiation. (manuscript in preparation)
* These authors contributed equally to this work.

The following section presents figures and tables that were predominantly obtained by the collaborators. Results and methods are presented in chapter 2.2 and 3.2, respectively. Methods, which were predominantly obtained by the collaborators are described in appendix E.

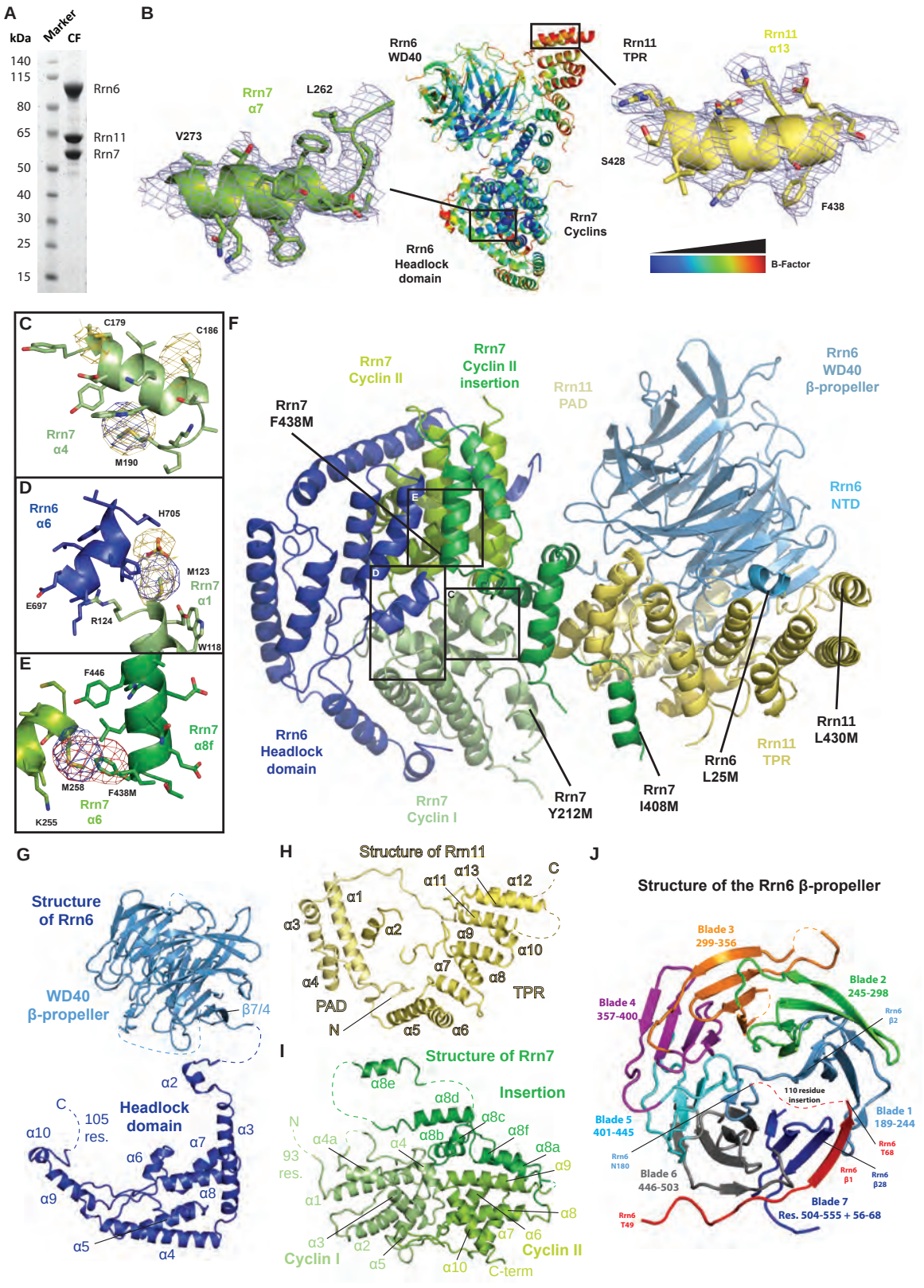


Figure C.1. Crystal structure of yeast Core Factor. Caption on next page.

Figure C.1. Crystal structure of yeast Core Factor. (**A**) Coomassie-stained SDS-PAGE of purified CF shows stoichiometric bands for the three subunits Rrn6, Rrn7 and Rrn11. (**B**) Cartoon representation of the CF crystal structure colored according to local B-factors (center), shows large differences. Central parts are well ordered and display low B-factors and a well-defined 2Fo-Fc map that allows for the unambiguous placement of side chains (eg. Rrn7 helix α7, left, map at 0.9σ). In contrast, peripheral elements, such as the Rrn11 TPR helix α13 (right; map at 0.9σ) have a high B-factor and show less well-defined density. (**C**) Cartoon representation of the Rrn7 helix α4. Anomalous difference density peaks for sulfur (map at 2.6σ, yellow) and selenium (map at 5.0σ, blue) confirm the residue placement in Rrn7. (**D**) Rrn6 helix α6 (blue cartoon) and the Rrn7 helix α1 (green cartoon) with an anomalous difference density peak for selenium (map at 5.0σ, blue) and a strong sulphate (map at 2.6σ, yellow) that is coordinated in proximity to Rrn6 H705. (**E**) Rrn7 helices α6 and α8f with the anomalous difference density for selenium. The map for the native protein (5.0σ) is shown in blue, for the F438M mutant (5.0σ) in red. (**F**) Cartoon representation of CF with domain coloring with pointing out the location and type of mutation for structure validation and helices identification. (**G**) Detailed view of the Rrn6 structure in cartoon representation shows the large distance between WD40 and headlock domains and emphasizes the synergy with Rrn7 and Rrn11. A peptide in the N-terminal region of Rrn6 (20-27) is ordered and could be assigned due to a Leucin-to-Methionine mutation of residue 25 and an anomalous sulphur signal of Cysteine 27 (Methods). This peptide interacts with Rrn11 TPR helices α9, α11 and α13 (Fig. 3C), contributing to the network of intimate CF subunits intereactions. (**H**) Rrn11: PAD and TRP domains form compact assemblies which are placed around the Rrn6 beta propeller. (**I**) Rrn7: With the 93 N-terminal residues disordered, the cyclin domains and a cyclin II insertion form a compact arrangement, from which only insertion helix α8e protrudes, forming PIR III in complex with Pol I. Additionally, helix α8d forms a contact with the first tandem repeat of Rrn11 helices α5/α6, thereby contributing to the strong association of Rrn11 with Rrn7. (**J**) Architecture of the Rrn6 β-propeller. The first sheet (7/4, red) is followed by a 110 residue insertion which is succeeded by 27 further sheets forming a 7-bladed propeller (each blade depicted in a different color) between residues 185 and 555. Striking homologies with the scaffolding protein RACK1 and the histone-binding factor WDR5 were detected by a PDBeFold homology search (Hussain et al., 2014; Schuetz et al., 2006).

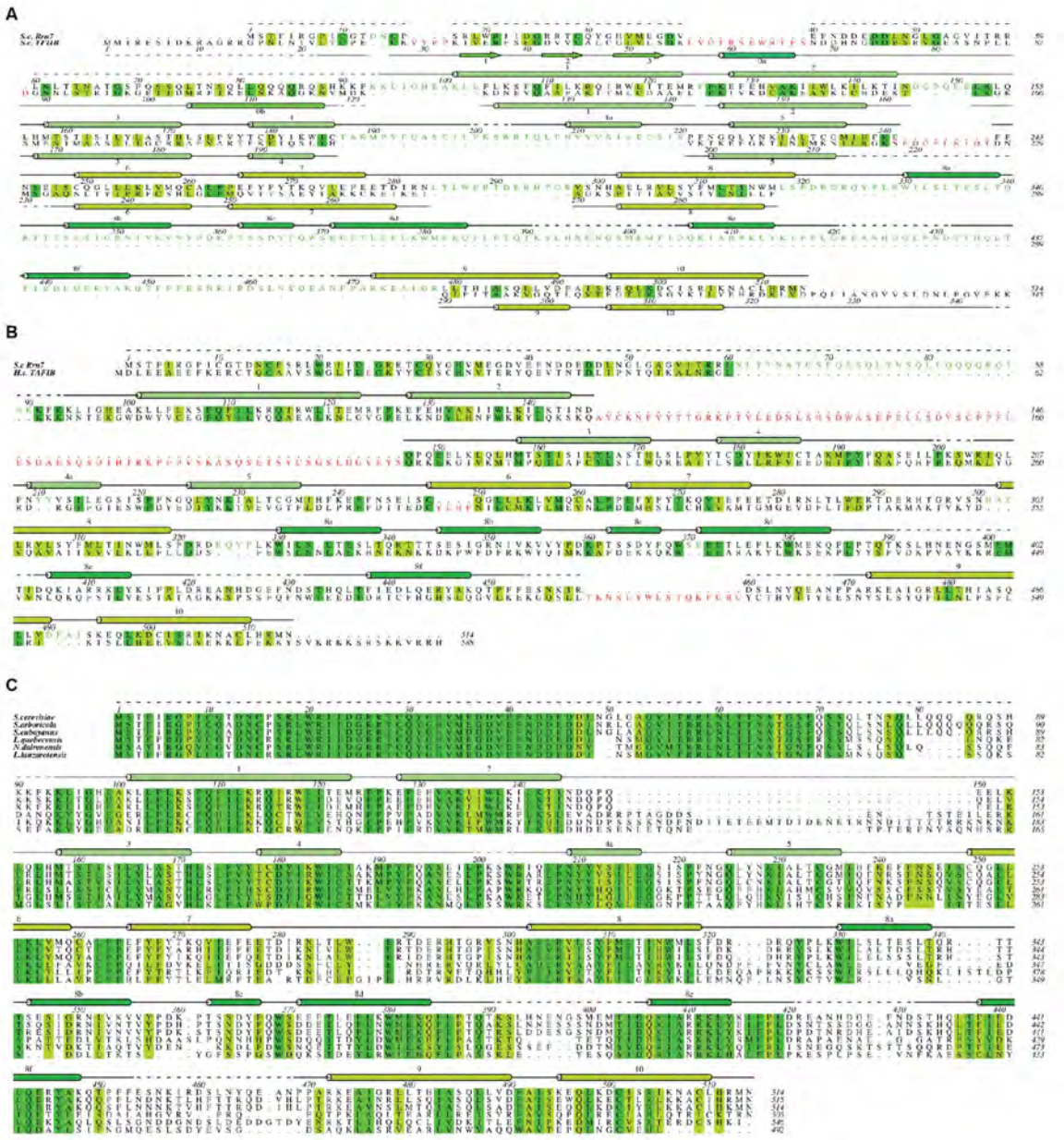


Figure C.2. Sequence alignments of Rrn7. (**A**) Structure-based sequence alignment of *S. cerevisiae* Rrn7 with TFIIB (4BBS and 5FYW). Structures were manually superposed in COOT and corresponding elements were assigned accordingly. For a 3D structure comparison see Fig. 2.11. Among minor differences, a 34 residue insertion adds helix α4a to cyclin I, that interferes with DNA binding (Fig. 2.15). The Rrn7 cyclin II insertion-helix α8e is flexibly connected to the rest of the polypeptide but embeds itself between the Rrn11 TRP helices α6 and α7, thereby protruding from the otherwise compact assembly of Rrn7 (Fig. S1I), and forming PIR III (Fig. 2.12D). (**B**) Sequence and secondary structure prediction based alignment of yeast and human Rrn7 (TAF1B). In higher organisms, the reader/linker region further truncated and an additional, apparently flexible insertion between cyclin I helices α2 and α3 is predicted. In the structure of CF (Fig. 2.10), this would position the human TAF1B insertion close to α4a and would also lie in the modelled path of PAD-bound promoter DNA. (**C**) A multiple sequence alignment of Rrn7 shows that the protein is conserved throughout species and shows high sequence similarity with other yeasts.

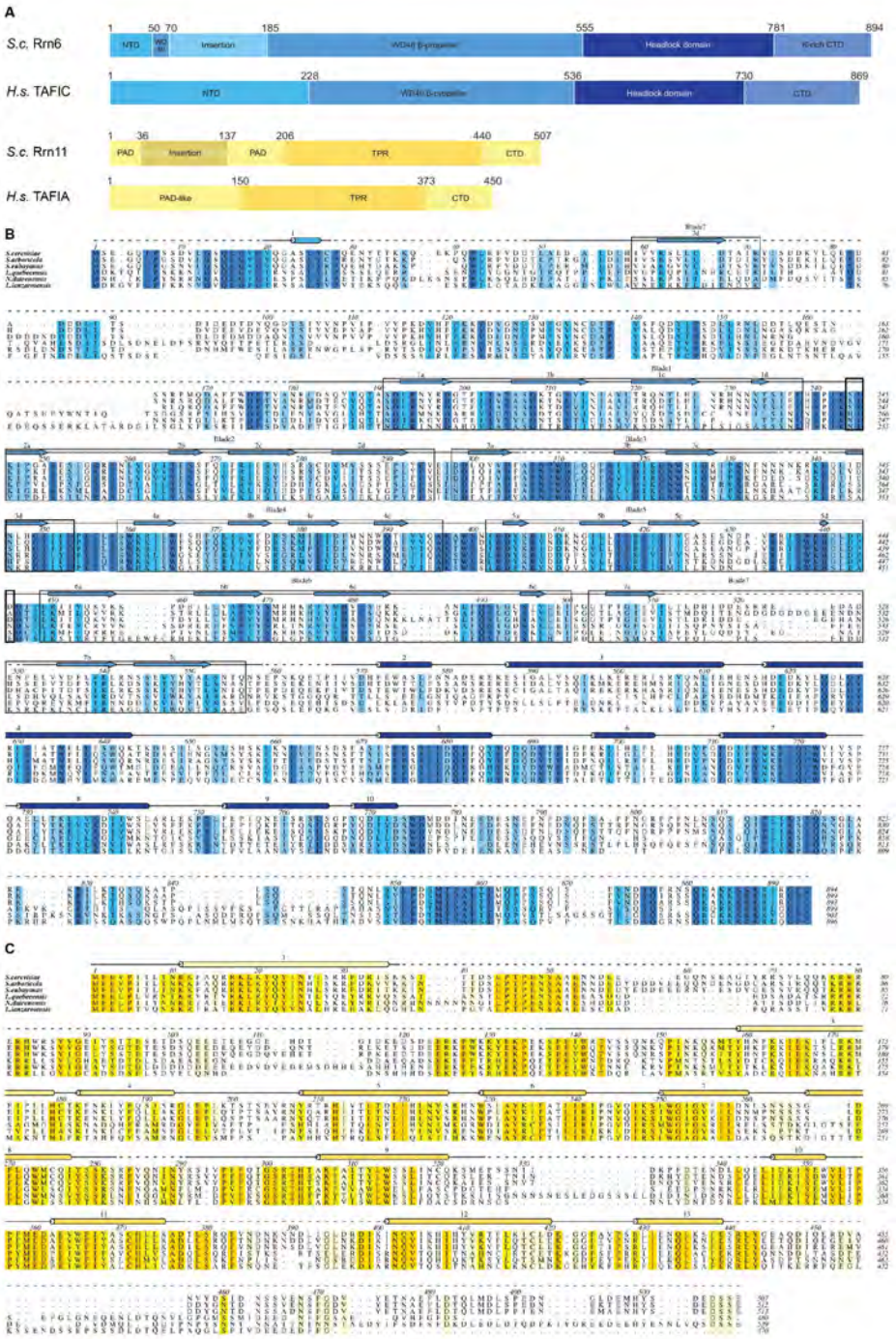
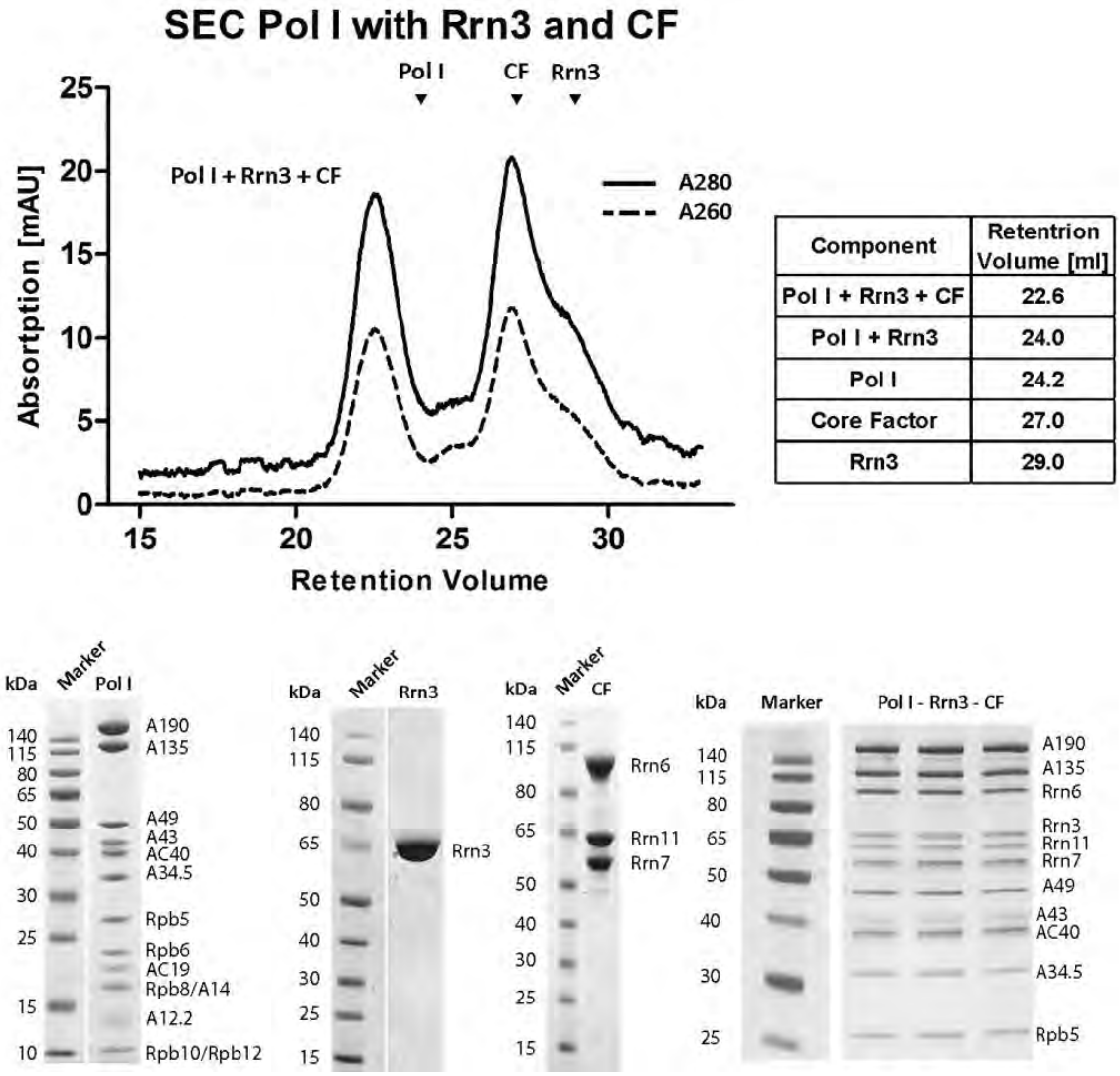


Figure C.3. Secondary structure based alignment of Rrn6 and Rrn11. (**A**) The domain architecture of Rrn6 and Rrn11 are conserved between yeast (S.c.) and human (H.s.). (**B**) A sequence and secondary structure prediction based alignment of Rrn6 between yeast species shows conserved and variable elements. (**C**) Sequence and secondary structure prediction based alignment of Rrn11 among yeast species.



Component	Retentrion Volume [ml]
Pol I + Rrn3 + CF	22.6
Pol I + Rrn3	24.0
Pol I	24.2
Core Factor	27.0
Rrn3	29.0

Figure C.4. Proteins used in this study and formation of a Pol I-Rrn3-CF complex. Size exclusion profile of a sample containing Pol I and a 5-fold molecular access of Rrn3 and CF. Retention volumes of A280 peaks of single samples are indicated. Two 10/300 Superose 6 columns (GE Healthcare) were connected. Coomassie-stained SDS-PAGE gels of Pol I, Rrn3 and Core Factor after purification and of the assembled complex peak (top) are shown.

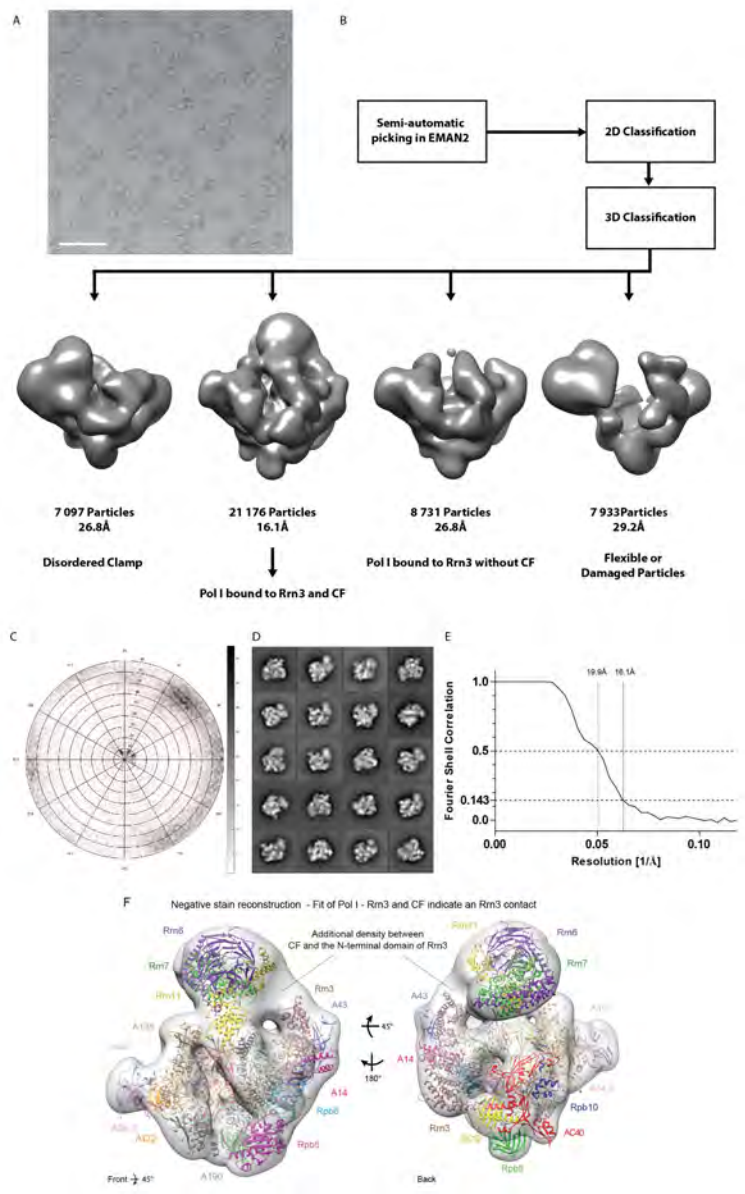


Figure C.5. Classification of the Pol I-Rrn3-CF negative-stain EM dataset. (**A**) Exemplary micrograph of the cryo-EM dataset. The scale bar is 100 nm. (**B**) Processing and classification tree. As in the cryo-EM dataset, a Pol I-Rrn3 structure and two structures of Pol I with damaged particles or clamp/stalk flexibilities are observed. The amount of particles with bound CF is higher in negative stain preparations, which is likely due to a higher stability originating from grids with continuous carbon support film (methods). (**C**) Angular distribution of single particle orientations used in cryo-EM reconstruction of the Pol I-Rrn3 complex. Shades indicate the number of particles assigned to a view; red dots indicate represented views. (**D**) Representative 2D class averages of the particles used for the Pol I-Rrn3 negative-stain EM reconstruction. (**E**) FSC plot for half-maps of the Pol I-Rrn3-CF negative stain reconstruction. 0.143 and 0.5 FSC criteria indicated. The first data point after phase randomization is omitted. (**F**) A Fit of Pol I, Rrn3 and CF into the negative stain envelop shows density for the entire CF. CF may additionally contact the N-terminal region of Rrn3, as indicated by the presence of additional density (black line) and as previously suggested (Peyroche et al., 2000).

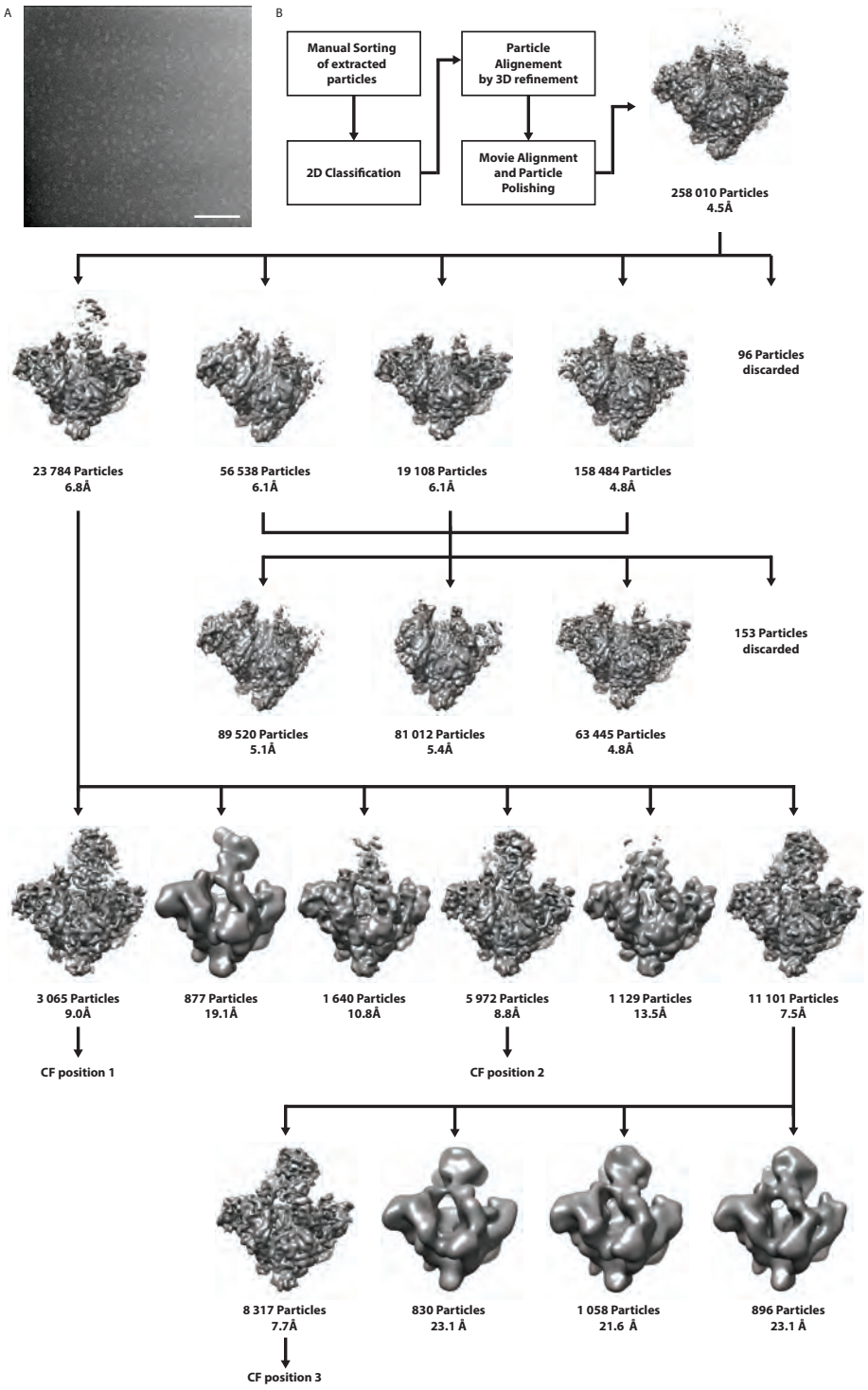


Figure C.6. Classification of the Pol I-Rrn3-CF cryo-EM dataset. (**A**) Exemplary micrograph of the cryo-EM dataset. The scale bar is 100 nm. (**B**) Processing and classification tree. A Pol I-Rrn3 structure was previously published (Engel et al., 2016) and two structures of Pol I with dissociated A49/34.5 sub-complex or disordered clamp could not be further classified and result from damaged particles or clamp/stalk flexibilities in the absence of a binding partner, such as a second Pol I molecule, Rrn3 or template DNA.

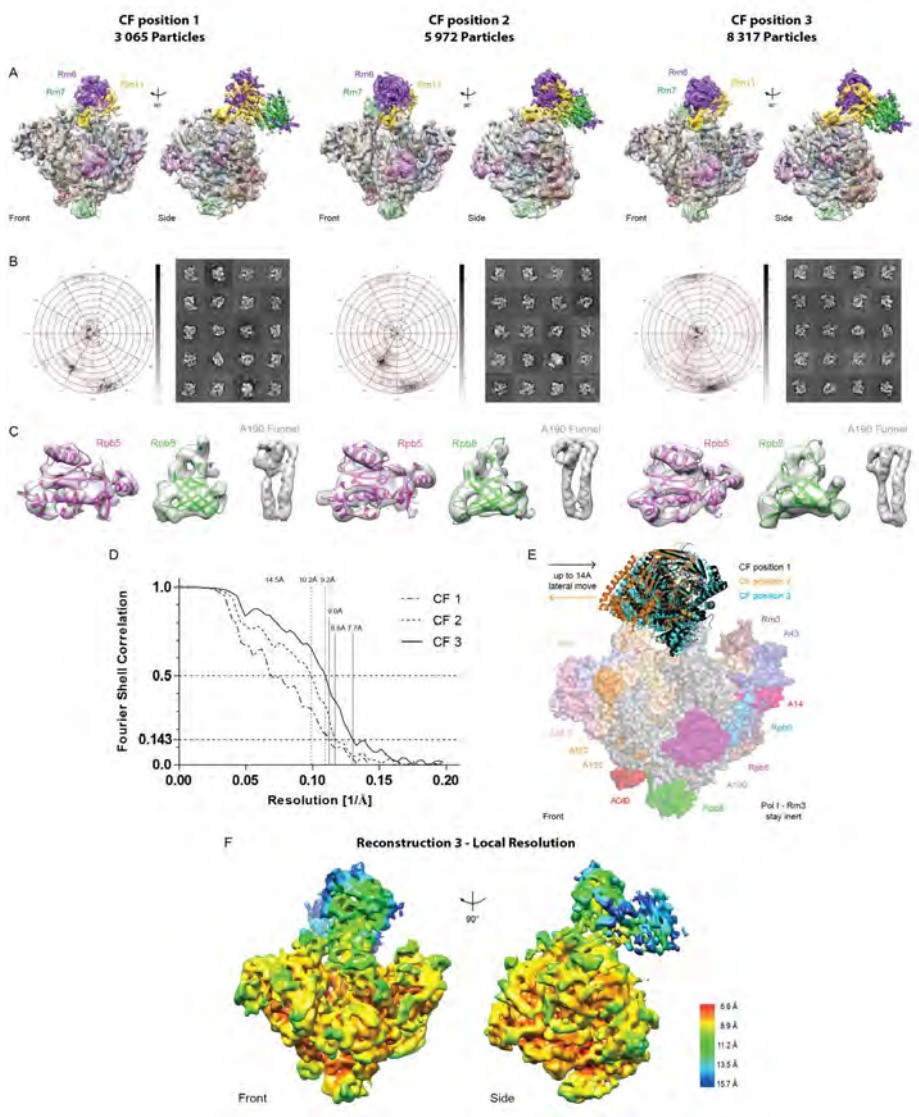


Figure C.7. Details of three Pol I-Rrn3-CF reconstructions. (**A**) Fit of Pol I, Rrn3 and CF domains in the three obtained reconstructions from front and side view. While proximal CF parts, especially the PIRs are well defined, distal CF parts appear more flexible. (**B**) Angular distribution plots (left) and 2D class averages (right) for the three reconstructions. Shades indicate the number of particles assigned to a view; red dots indicate represented views. (**C**) Example densities for the subunits Rpb5 and Rpb8 as and the A190 funnel helices indicate a high quality of fit for all reconstructions. (**D**) FSC plot for half-maps of the Pol I-Rrn3-CF cryo-EM reconstructions. 0.143 and 0.5 FSC criteria indicated. The first data point after phase randomization is omitted. (**E**) Overlay of the three reconstructions shows that Pol I and Rrn3 stay inert, while the position of CF can vary up to 14 /AA, but maintaining the same PIRs. In complex with Pol I and CF, Rrn3 adopts a conformation which is similar to the crystal structure (Blattner et al., 2011). Reconstruction 3 showed additional density spanning from the Rrn7ribbon towards the A190 zipper loop (Fig. 3). This indicates that Rrn7 residues 30-40 pass the A190 lid on its outside. In contrast, the B-reader/-linker elements of TFIIB pass the lid inside the polymerase cleft. The reader helix is apparently not present in Rrn7 (Fig. C.2) but the linker helix of TFIIB occupies the same position on the clamp coiled-coil helices (Kostrewa et al., 2009) as Rrn11 TPR helix α12 does (Fig. 2.12) in PIR II. (**F**) Local resolution of reconstruction 3 displays a well ordered, high resolution core and more flexible outer rims, especially CF module II.

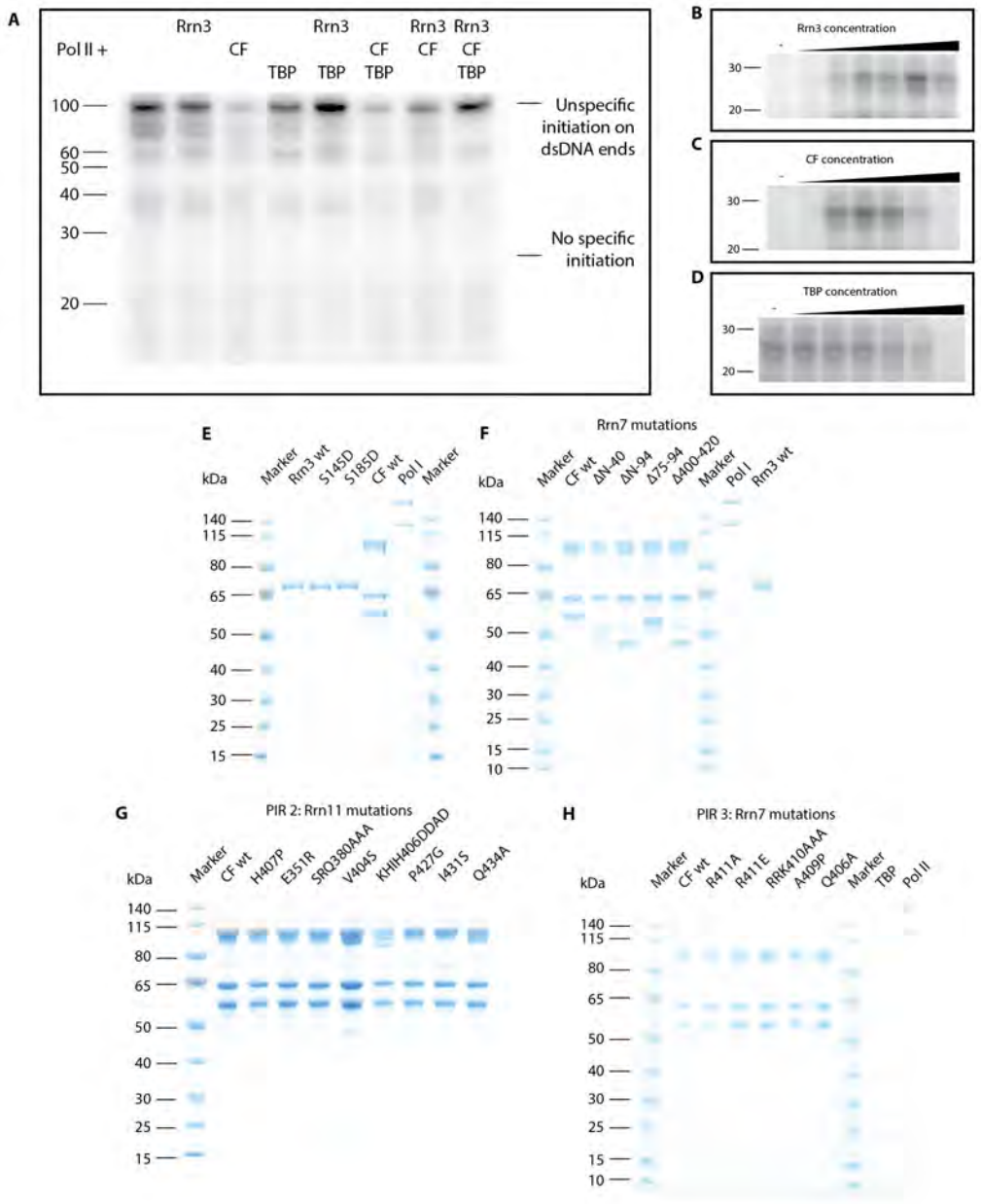


Figure C.8. Initial transcription assay shows Pol I-specificity of CF. (**A**) The initiation assay (compare Fig. 5) using Pol II instead of Pol I. Pol II shows no specific band at 24nt, but initiates from ds-breaks more efficiently than Pol I. Titrations of Rrn3 ((**B**); constant CF concentration, no TBP), CF ((**C**); constant Rrn3 concentration, no TBP) or TBP ((**D**); constant CF and Rrn3 concentration) from 0 to 50-fold molar access show that initiation is inhibited by high amounts of CF and TBP, probably due to multiple unspecific DNA binding events. (**E**)-(**H**) Coomassie-stained SDS-PAGE gels of used components and CF-mutants.

Table C.1. Data collection, phasing and refinement statistics for SAD and native structures. Values in parentheses are for highest-resolution shell.

	Native 1	Native 2	SeMet
Data collection	PX1 Pilatus M1	PX1 EIGER 16M	PX3 Pilatus 2M-F65
Space group	P65	P1	P65
Cell dimensions			
a,b,c (Å)	109.09,109.09,383.80	109.07,109.14,385.64	108.70,108.70,383.40
α,β,γ (°)	90.000,90.000,120.000	90.021,90.009,95.976	90.000,90.000,120.000
Wavelength (Å)	1.00000	0.99995	0.97941
Resolution (Å)	50.0 - 3.2	60.0 - 3.2	50.0 - 4.00
R_{merge} (%)	9.1 (282.9)	9.2 (235.1)	34.3 (227.3)
I/σ(I)	22.14 (1.38)	14.83 (1.12)	13.50 (2.08)
$CC_{1/2}$	100 (54.0)	99.9 (41.7)	99.9 (73.9)
Completeness (%)	100 (99.9)	98.9 (99.4)	99.9 (100.0)
Redundancy (%)	11.6 (11.8)	8.3 (8.2)	42.7 (42.3)
Refinement			
Resolution (Å)		54.57 - 3.20	
No. reflections		251 065	
R_{work}/R_{free} (%)		25.96 /28.90	
No. atoms		62 484	
Protein		64 328	
Sulfate		150	
B factors		147.28	
Protein		147.18	
Ligand / ion		184.83	
Water		NA	
rms deviations			
Bond lengths (Å)		0.008	
Bond angles (°)		1.190	

Table C.2. Data collection and phasing statistics for SAD datasets. SeMet mutant A, B and C correspond to Rrn11-L430M / Rrn7-I408M, Rrn6-L25M / Rrn7-F438M and Rrn11-L3M / Rrn7-V212M, respectively. Values in parentheses are for highest-resolution shell.

	Sulphur - Peak	SeMet mutant A	SeMet mutant B	SeMet mutant C
Data collection	PX1 EIGER 16M	PX1 EIGER 16M	PX1 EIGER 16M	PX1 EIGER 16M
Space group	P65	P65	P65	P65
Cell dimensions				
a,b,c (Å)	108.65, 108.65, 384.60	108.25, 108.25, 381.60	109.05, 109.05, 384.20	109.20, 109.20, 384.60
α,β,γ (°)	90,90,120	90,90,120	90,90,120	90,90,120
Wavelength (Å)	2.06640	0.97848	0.97865	0.97846
Resolution (Å)	40 - 3.4	35 - 4.0	35 - 4.0	35 - 4.2
R_{merge} (%)	53.5 (194.1)	46.5 (229.1)	34.9 (223.9)	37.4 (133.8)
I/σ(I)	29.64 (2.38)	12.4 (2.02)	12.23 (2.05)	14.61 (2.25)
$CC_{1/2}$	99.8 (37.2)	99.3 (74.1)	99.8 (78.4)	99.5 (47.1)
Completeness (%)	99.3 (92.6)	99.9 (100.0)	99.8 (100.0)	99.6 (97.6)
Redundancy	385.2 (81.5)	36.1 (24.4)	39.5 (30.1)	51.5 (19.5)

Table C.3. Cleft expansion states of Pol I. Pol I undergoes movements of its predefined domains during the steps of initiation. To correlate functional Pol I states, we compared cleft expansion in published Pol I structures (Engel et al., 2016; Engel et al., 2013; Neyer et al., 2016). In order to define cleft expansion states, we measured residue distances indicative for cleft expansion on the downstream edge (between A190 residues G231 and K1331), the upstream edge (A190 residue E414 and A135 residue K434) and compared to the crystal structure (A190 residue E414). This shows a constant cleft contraction throughout activation states, as predicted. Upon contraction, the A43-connector, the A12.2 C-terminal domain, and the A190 expander domains apparently detach and the bridge helix is rewound.

	Downstream	Clamp-Protrusion	$A190_{414}$ relative to 4C2M
Dimer (4C2M)	42 (Å)	41 (Å)	NA
Free Monomer	42 (Å)	39 (Å)	6 (Å)
Pol I - Rrn3 (5G5L)	41 (Å)	36 (Å)	6 (Å)
PIC	39 (Å)	36 (Å)	7 (Å)
ITC	35 (Å)	29 (Å)	15 (Å)
EC	34 (Å)	28 (Å)	17 (Å)

D. Extended Materials and Methods of the Elongation study

Materials and methods presented in this section are published and were obtained in a collaboration with the Frangakis lab. Author contributions are stated on page VI.

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

* These authors contributed equally to this work.

The following section presents supplemental material and methods that were predominantly obtained by the collaborators. Results and methods are presented in chapter 2.1 and 3.1, respectively. Figures, which were predominantly obtained by the collaborators, are described in chapter B.

D.1. Miller tree preparation and cryo-ET imaging

Miller chromatin spreads (Miller and Beatty, 1969) were prepared as described (Osheim et al., 2009) using the NOY1071 yeast strain with 25 copies of rDNA repeats (Machin et al., 2006), with the following modifications: yeast cells were grown to mid-log phase (A_{600}=0.4) at 30° C; sucrose was precluded from the sucrose-formalin cushion; and after depositing the yeast lysate on electron microscopy (EM) grids, the grids were transferred to 11 mM KCl pH 9 water for rapid plunge-freezing.

Prior to cryo-ET analysis, the yeast lysate was centrifuged on 300 mesh copper grids with an approximately 30 nm thick carbon support layer evaporated by a carbon-coater 208Carbon (Cressington, Watford, UK) and glow-discharged with a strong plasma for

approximately 1 min using a home-made device. The grids were immediately plunge-frozen in liquid ethane by a Vitrobot Mark IV (FEI, Eindhoven, The Netherlands) with 25 blotting force, 3 s blotting and 10-15 s draining time and the blotting chamber set to 100 % humidity at 10° C. Cryo-grids were mounted into autoloader grids with C-clippings (FEI, Eindhoven, The Netherlands) in an EM FC6 cryo-microtome (Leica, Wetzlar, Germany) that was cooled with liquid nitrogen under gaseous flow to -150° C. During mounting, grids were visually inspected to determine whether they contained an intact carbon film.

Tilt-series were recorded using DigitalMicrograph (Gatan Inc., California, USA) at a nominal magnification of 33,000x (4.0 Å/pixel) in EFTEM mode at 300 keV using a Titan Krios with a GATAN GIF Quantum SE post-column energy filter in zero loss peak mode and a K2 Summit detector. The camera was operated in counting mode with a dose rate of approximately 15 e^-/pixel/sec and a total dose of approximately 100 $e^-/Å^2$. The tilt-series ranged from -63° to +63° with an angular increment of 2° and defocus set at -5 μm. Tilted images were fiducial-less aligned (Castano-Diez et al., 2007) and reconstructed by super-sampling SART (Kunz and Frangakis, 2014). The CTF was measured and corrected in slices in 3D (Kunz and Frangakis, 2016).

D.2. Reconstruction and segmentation of Miller trees

3D reconstructions were visualized with the EMpackage in Amira (Pruggnaller et al., 2008) (FEI, Eindhoven, The Netherlands & Zuse Institute, Berlin, Germany) and analyzed by custom image processing scripts written in MATLAB (all scripts are freely accessible). Segmentation of the Miller trees was performed manually in Amira by drawing contours encompassing individual features on mildly Gaussian low-pass filtered tomograms using the high-contrast option of super-sampling SART (Kunz and Frangakis, 2014).

D.3. Sub-tomogram averaging of Pol I enzymes

For sub-tomogram averaging (i.e. the cryo-ET structure) we selected five Miller trees with an obvious transcriptional directionality (several Miller trees were not completely

in the field of view). Of those, the 225 best-resolved Pol I enzymes contributed to the sub-tomogram average, in which the orientation confirmed the transcription direction and the RNA exit site matched previous observations (Bernecky et al., 2016). Sub-tomograms containing transcribing Pol I enzymes on rDNA were manually selected. The enzymes were re-centered using a Gaussian blob of the size of Pol I. The positions of all sub-tomograms were subsequently indexed such that they were placed sequentially on the DNA. Since the DNA was visible in the reconstructions, the indexing was unambiguous.

Sub-tomogram averaging was then performed on each Miller tree individually. This was to guarantee that the directionalities of the enzymes are not mixed due to the globular shape of the enzyme, the pseudo-symmetry axis, and the varying ice thickness of the recording area leading to different signal-to-noise ratio among the enzymes. The Euler angles were determined *a priori* for each of the three consecutive Pol I enzymes per Miller tree by calculating the vector from center-to-center position. Constrained sub-tomogram averaging was performed on sub-tomograms with 64x64x64 voxels using a spherical mask (20 nm diameter). To ensure the robustness of the sub-tomogram averaging two different starting references were used (a) the average of all rotationally pre-aligned Pol I enzymes per strand, and (b) a Gaussian blob of the size of Pol I. Both converged to the approximately same density. During sub-tomogram averaging of each individual Miller tree, polymerases were low-pass filtered and the alignment was run with a translational freedom of 10 voxels around the Gaussian blob refined position, a full rotational freedom for phi and psi, and a constrained rotational freedom of +-30 ° for theta with 5° sampling increment, until the average reached convergence. The missing wedge was taken into account during the entire alignment.

The sub-tomogram averages of each Miller tree were individually inspected and the orientation of the Pol I enzymes on each Miller tree was analyzed. The 3' to 5' directionality of the enzymes on each Miller tree was analyzed. If all enzymes had the same directionality (i.e. the signal-to-noise ratio was sufficient to align them properly), their sub-tomogram average was used for further processing. If the enzymes had conflicting directionalities (including complete random directionality), their sub-tomogram average was rejected. Five Miller trees qualified for this criterion. Their enzyme directionality was visualized compared to the Miller-tree directionality, and they all conformed. Finally, the 225 enzymes (from the 993 total enzymes in the tomograms) of the five selected Miller trees were subjected to a refined sub-tomogram averaging and the resulting cryo-ET structure reached a resolution of approximately

2.9 nm with the FSC 0.5 threshold criterion (approximately 3.1 nm when compared to the cryo-SP structure).

D.4. Additional cryo-ET analysis

In the tomograms both the DNA and the RNA could be seen emanating from the enzymes. They were manually localized as close as possible to the enzyme and subsequently sub-tomogram averaging was performed around this position. To obtain evidence for the RNA exit channel visualized in the cryo-SP structure, we made three independent attempts to manually select the position of exiting RNA on Pol I in the tomogram without prior knowledge of the structure (Extended Fig. B.2F). The resulting point distribution of exiting RNA on the cryo-ET structure agreed with the location of the RNA exit channel in the cryo-SP map and further confirmed the correct superposition of the two independent structures.

The distances of consecutive Pol I enzymes were calculated as the Euclidian distance between their center-to-center positions. For plotting the probability density function, one enzyme was centered, the downstream enzyme was placed on the y-axis, and the upstream enzyme was placed on the plane. Between three consecutive neighboring enzymes, the in-plane angle was estimated.

For fitting of structures to the cryo-ET reconstruction, rigid body fitting of the cryo-ET and the cryo-SP structures of the Pol I EC was performed automatically. This resulted in a global cross-correlation value of approximately 0.8 and a FSC shown in Fig. 2.9A. The contour level for the cryo-ET structure for volume rendering of our average was calculated from the theoretical molecular mass with an average protein density of 0.8 kDa/nm^3.

E. Extended Material and Methods of the Initiation study

Materials and methods presented in this section are in preparation for publication and were obtained in a collaboration with Tobias Gubbey and Christoph Engel. Author contributions are stated on page VII.

C. Engel*, T. Gubbey*, **S. Neyer***, S. J. Sainsbury, C. Oberthür, C. Bäjen, C. Bernecky, P. Cramer. (2016) Structural basis of RNA polymerase I transcription initiation. (manuscript in preparation)
* These authors contributed equally to this work.

The following section presents materials and methods that were predominantly obtained by the collaborators. Results and methods are presented in chapter 2.2 and 3.2, respectively. Figures and Tables, which were predominantly obtained by the collaborators are described in appendix C.

E.1. Additional cryo-ET analysis

Core factor (CF) subunits from *S. cerevisiae* were co-expressed in *E. coli* BL21-CodonPlus(DE3)-RIL cells form two plasmids. Rrn6 and Rrn11 were cloned into pET-21b with a C-terminal 6xHis tag on Rrn11. Rrn7 was expressed from pET-28b with an N-terminal 6xHis tag. A single colony was used to inoculate 100 mL LB medium and incubated overnight at 37° C with ampicillin, kanamycin and chloramphenicol. The pre-culture was diluted 1:100 to inoculate 4 L of LB medium and incubated at 37° C until OD_{600} values reached 0.5-0.7. Cultures were cooled on ice for 20 min and expression was induced with 0.1 mM IPTG. Cells were grown at 18° C for 18 h and harvested by centrifugation, washed with phosphate-buffered saline (PBS) at 4° C, flash frozen in liquid nitrogen and stored at -80° C.

For expression of seleno-methionine labelled core factor, a 4 L LB starter culture was grown until OD_{600} reached 0.3-0.5. Cells were collected by centrifugation and washed three times with PBS to remove residual media. Cells were re-suspended in minimum medium depleted for methionine (Molecular Dimensions) to an OD_{600} of 0.2 - 0.3. The culture was incubated for 1.5 h at 37° C with antibiotics under shaking. After starvation, cell growth was induced by adding selenomethionine stock solution (Molecular Dimensions). After reaching an OD_{600} of 0.6, the temperature was reduced to 18° C and additional amino acids were added (25 mg/l lysine, threonine, and phenylalanine; 12.5 mg/L leucine, isoleucine, and valine). Expression was induced by adding IPTG to a concentration of 1 mM. Cells were grown for 18 h, harvested by centrifugation, frozen in liquid nitrogen and stored at -80° C.

E.2. Purification of core factor

A pellet obtained from 4 L cell culture was re-suspended in buffer A (20 mM imidazole, 350 mM NaCl, 10 mM $MgCl_2$, 10 % (v/v) Glycerol, 20 mM Hepes pH 7.8, 1 mM DTT, 1x protease inhibitor). Cells were lysed by sonication using a Branson Digital Sonifier. The lysate was cleared by centrifugation and the supernatant was filtered with a 0.22 μm filter (Millipore) to remove cell debris. Cell lysate was then applied to a Ni-NTA column (5 ml column volume (CV), GE Healthcare) and bound CF washed with 5 CV of buffer B (25 mM imidazole, 200 mM NaCl, 10 mM $MgCl_2$, 10 % (v/v) glycerol, 20 mM HEPES pH 7.8, 1 mM DTT) at 4° C. The column was transferred to room temperature, washed with 2.5 CV of buffer C (50 mM imidazole, 200 mM NaCl, 10 mM $MgCl_2$, 10 % (v/v) glycerol, 20 mM HEPES pH 7.8, 1 mM DTT, 5 mM ATP, 2 mg/ml denatured protein), incubated for 10 min, and washed again with 2.5 CV buffer C. This procedure was previously used (Rial and Ceccarelli, 2002) and aims at the removal of tightly bound chaperones. The column was transferred to 4° C and washed with 5 CV buffer D (50 mM imidazole, 200 mM NaCl, 10 mM $MgCl_2$, 10 % (v/v) glycerol, 20 mM HEPES pH 7.8, 1 mM DTT). Elution was performed with 5 CV of buffer E (350 mM imidazole, 200 mM NaCl, 10 mM $MgCl_2$, 10 % (v/v) glycerol, 20 mM HEPES pH 7.8, 1 mM DTT). Protein was then fractionated on a 5 ml heparin column (GE Healthcare). Protein was loaded in buffer F (1 mM $MgCl_2$, 10 % (v/v) glycerol, 20 mM HEPES pH 7.8, 1 mM DTT) and then eluted with a gradient ranging for 0 to 2 M NaCl, including a plateau at 550 mM NaCl of 2 CVs. CF-containing fractions were concentrated to 1 ml using a 100 kDa cut-off centrifugal filter (Millipore). Size exclusion chromatography

was carried out with a Superose 6 10/300 column (GE Healthcare) in buffer G (200 mM, 1 mM $MgCl_2$, 5 % (v/v) glycerol, 10 mM HEPES pH 7.8, 10 μM $ZnCl_2$, 1 mM DTT). CF-containing fractions were concentrated using a 100 kDa cut-off centrifugal filter (Millipore) and directly used or flash-frozen in liquid nitrogen and stored at -80° C.

E.3. Crystallization of core factor

CF was thawed and crystallized by hanging-drop vapour diffusion using a reservoir solution containing 0.5 M ammonium sulfate, 12 % (m/v) PEG 4000, 0.1 M MES pH 6.0 and 1 mM DTT. Pre-greased Crystalgen SuperClear 24-well Plates (Jena Bioscience) and siliconized cover slides (Jena Bioscience) were used to set 1 μl drops with a reservoir volume of 0.5 ml. Obtained initial crystals were improved by micro-seeding using a seed-bead kit (Hampton Research) and yielded rod-shaped crystals with a hexagonal base and a length of up to 300 μm. Crystals were harvested after 5-7 day and transferred to a cryo-protectant solution containing the reservoir condition and 25 % (v/v) glycerol in three steps. Crystals were flash-frozen and stored in liquid nitrogen until data collection.

E.4. Crystal structure determination

Data were collected at the Swiss Light Source in Villigen on the beamline PX1 with a Pilatus 6M Detector (Dectris) or an EIGER 16M detector (Dectris), and beamline PX3 using a PILATUS 2M-F detector (Dectris). Diffraction was observed to 3.2 Å resolution and data was processed with XDS (Kabsch, 2010), showing P65 symmetry, as confirmed using the program Pointless (Evans, 2006) in the CCP4 suite (Winn et al., 2011). The presence of one molecule per asymmetric unit (AU) was suggested by Matthews probability coefficient estimation (Matthews, 1968). Molecular replacement (MR) using TFIIB variants, various WD40 β-propeller domains and/or TPR domains as search models was unsuccessful. Thus, phase information was obtained by single wavelength anomalous diffraction (SAD) from selenomethionine-labelled CF crystals. Diffraction data was collected on PX3 at three different χ angels. Datasets were merged, showing a strong anomalous signal until approximately 6 Å resolution (Table C.1). A total of 21 selenium atoms were found using the SHELX C/D pipeline (Sheldrick, 2010) or HYSS, and an initial map was obtained using Phenix.autosol

(Adams et al., 2010). Phase extension to approximately 3.2 Å using the Phenix suite improved the initial map (Fig. C.1).

Structure determination was impaired by protein flexibility and a pseudo-symmetric arrangement of CF molecules. To derive additional sequence markers for model building, sulphur atom positions were determined from a sulphur-SAD dataset recorded from a native crystal at high redundancy and a wavelength of 2 Å. Using anomalous peaks for both methionine and cysteine residues as sequence markers, an initial model was built in COOT (Cowtan, 2010). To improve chain tracing and residue positioning, a total of six residues were mutated to methionine in pairs of two. Selenomethionine-labeled protein was prepared, SAD data was collected (Table C.2) and anomalous difference maps calculated using the CCP4 suite. This allowed for the unambiguous localization of residues Rrn6-L25, Rrn7-V212, Rrn7-I408, Rrn7-F438 and Rrn11-L430 (Fig. C.1). The Rrn11-L73M mutation did not yield a selenium signal, suggesting flexibility of this region. Because refinement of the structure in space group P65 using phenix.refine, Refmac (Murshudov et al., 1997) or BUSTER (Smart et al., 2012) did not result in free R-factors below 35 %, the data was reprocessed in lower symmetry space groups. MR placed six CF molecules in P1, showing a super-helical, pseudo-symmetric arrangement, but slightly deviating from translational and rotational crystallographic symmetry.

Refinement of the structure was carried out in space group P1 using phenix.refine with Cartesean non-crystallographic symmetry (NCS) restraints, 42 rigid body groups, 18 TLS groups, individual isotropic atomic displacement parameters and coordinate refinement. In the final stages, defined secondary structures were fixed as such and geometry target weights set to at a wxc scale of 0.25. This enabled refinement of our model to 3.2 Å resolution with R/Rfree factors of 26.0/28.9 % and good stereochemistry (Table C.1). At a global resolution of 3.2 Å we observed differences in the quality of the electron density map for different protein regions. While the electron density map allowed for the clear assignment of residues in most parts of Rrn7, the C-terminal (TPR) part of Rrn11 and the regions of Rrn6 are more flexible, resulting in local differences in B-factors and map quality (Fig. C.1). A single helical density could not be assigned with certainty, but apparently belongs to the flexible N-terminal region in the Rrn11 PAD. We modelled this density with poly-alanine and arbitrary residue numbers.

E.5. Preparation and cryo-EM analysis of Pol I-Rrn3-CF complex

Samples were prepared and cryo-EM data was collected as described (Engel et al., 2016). Apart from particles with disordered stalk and clamp domains, we observed particles with dissociated A49-A34.5 subcomplex, which has lower affinity to the Pol I core (Geiger et al., 2010) and Pol I in complex with Rrn3 (Engel et al., 2016). A total of 23,784 particles with localized noise were subsequently found to be in complex with CF at great heterogeneity. Those particles were subjected to focused classification using a CF mask (Fig. C.6). The largest resulting class was sub-classified using local searches. Three classes displayed CF in different orientations, of which the largest was sub-classified using local searches. The final reconstructions CF-1, CF-2 and CF-3 contained 8317, 5972 and 3065 particles, respectively. A mask encompassing Pol I, Rrn3 and CF was calculated using RELION (Scheres, 2012) and used in 3D refinement to yield reconstructions at 9.0 Å, 8.8 Å and 7.7 Å resolution, respectively. Particles showed a preferred orientation which was, however, not exclusively occupied (Fig. C.7). Resolution is based on the gold-standard FSC (0.143 criterion) and temperature factors were automatically determined and applied in RELION (-247.8 Å^2, -433.4 Å^2 and -628.2 Å^2 for the final reconstructions).

At the nominal resolution of 7.7 Å we derived a pseudo-atomic model of the Pol I-Rrn3-CF complex based on published crystal structures (PDB codes 4C2M and 3TJ1) and the structure of CF presented here. A model of a Pol I monomer lacking the expander and the connector was constructed from the PDB entry 4C2M using COOT (Emsley et al., 2010) and placed into the density using UCSF Chimera (Pettersen et al., 2004). Previously defined domains of Pol I were rigid body-fitted in real space using COOT. A Rrn3 monomer (PDB 3TJ1) was also fitted to the density with USCF Chimera and adjusted with COOT. Geometric parameters of residues located in connections between shifted domains were regularized applying standard geometrical restraints in COOT. As Pol I adopted the same confirmation in all three reconstructions, one model was built at 7.7 Å and used for all three structures. CF was globally fitted using Chimera and split into 5 rigid bodies which were separately fitted and slightly adjusted in COOT (A: Rrn11 291-440, B: Rrn11 206-282 and Rrn7 370-420, C: Rrn11 1-205, D: Rrn6 20-558, E: Rrn6 567-779 and Rrn7 94-369 and Rrn7 432-514). Figures were prepared with UCSF Chimera or PyMOL.

E.6. Negative stain EM of the Pol I-Rrn3-CF complex

Pol I-Rrn3-CF complex was prepared as before but applied to gradient fixation (GraFix) as described (Kastner et al., 2008). A 4 mL gradient from 10 % to 30 % sucrose was prepared, using a Gradient Master 108 (BioComp) with 0.025 % glutaraldehyde and 50 μL sample volume. Centrifugation for 16 h at 32,000 rpm and 4° C yielded 20 fractions of 0.2 mL. Crosslinking was quenched by adding aspartate or lysine to a final concentration of 10 mM. Fractions 14-15 from the top contained Pol I-Rrn3-CF complex and were diluted to approximately 0.10 mg mL^{-1}. Continuous carbon coated grids (Quantifoil) were glow-discharged for 20 s before applying 5 μL sample and incubated for 1 min. Grids were washed by floating on a 0.5 mL drop of distilled water for 1 min, stained for 1 min on a 50 μL drop of 2 % (w/v) uranyl formiate solution, and blotted dry. Images were acquired on a 4k x 4k CCD camera (TVIPS) at 88 k magnification (2.51 Å/pix) with a Philips CM200 FEG electron microscope operated at 160 kV. A total of 863 micrographs were collected with a defocus range of -1 to -4 μm, manually sorted and CTF corrected using CTFFIND3 (Mindell and Grigorieff, 2003). Semi-automatic particle picking using EMAN2 (Tang et al., 2007) yielded 44,937 particles which were 3D classified in a single round using a Pol I monomer from PDB entry 4C2M as reference (filtered to 40 Å). Subsequently, particles were refined using RELION 1.3, which yielded low resolution envelops of Pol I-Rrn3 and Pol I-Rrn3-CF complexes at 28 Å and 16 Å resolution, respectively (Fig. C.5).

E.7. Pol I promoter-specific transcription initiation assay

The 38 residues upstream of the Pol I transcription start site (TSS) were previously shown to be sufficient for Pol I initiation (Keener et al., 1998). Hence, we used a dsDNA scaffold comprising the region from -38 to +24 (Integrated DNA Technologies). Non-template: GAGTACAAGTGTGAGGAAAAGTAGTTGGGAGGTACTTCATGCGAAAGCAGTTGAAGACAAG; Template: CTTGTCTTCAACTGCTTTCGCATGAAGTACCTCCCAACTACTTTTCCTCACACTTGTACTC. Pol I, Pol II, CF, CF-mutants, Rrn3, and TBP were purified as described but omitting the final size exclusion step for Pol I (Blattner et al., 2011; Engel et al., 2016; Plaschka et al., 2016;

Sydow and Cramer, 2009). A 5x reaction buffer (RB) was prepared, consisting of 180 mM ammonium sulphate, 60 mM HEPES (pH 7.8), 10 mM magnesium sulphate, 30 μM $ZnCl_2$, 30 % (v/v) glycerol and 15 mM DTT. In a total volume of 10 μL, the initiation reaction was prepared on ice comprising 0.1 μM Pol I (or Pol II), 5-fold molar access of CF variants, Rrn3, and/or TBP, and 0.1 μM dsDNA scaffold and 2 μL of 5x RB. The reaction was filled to 8.75 μL volume with sterile water and 1.25 μL of a NTP mix was added which contains 20 μM GTP, 20 μM UTP, 20 μM CTP, 2 μM ATP (Thermo Scientific) and 0.8 μM [α-32P] (2.5 μCi/μl; Perkin Elmer). The reaction was well mixed and incubated at 30° C for 30 min. To stop the reaction, 10 μL of 2x TBE-Urea Sample Buffer (Thermo Scientific) were added, well mixed, incubated at 95° C for 3 min and stored on ice. A 20 % Urea-polyacrylamide gel was freshly prepared and pre-run at 500 V for 10 min in sterile 1x TBE buffer (Thermo Scientific). As a RNA-size standard, the decade marker system (Thermo Scientific) was used. Samples were loaded and run at 250 V for 1 h 45 min. The gel was incubated on a Phospho-Screen for 3-12 h at -20° C and red out using a Typhoon FLA 9500 (GE Healthcare). For titration experiments, used molar access steps for transcription factors were 0, 1x, 2.5x, 5x, 10x, 20x and 50x compared to Pol I.

References

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., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*, 66(Pt 2):213–21.

Albert, B., Leger-Silvestre, I., Normand, C., Ostermaier, M. K., Perez-Fernandez, J., Panov, K. I., Zomerdijk, J. C., Schultz, P., and Gadal, O. (2011). RNA polymerase I-specific subunits promote polymerase clustering to enhance the rRNA gene transcription cycle. *J Cell Biol*, 192(2):277–93.

Aprikian, P., Moorefield, B., and Reeder, R. H. (2000). TATA binding protein can stimulate core-directed transcription by yeast RNA polymerase I. *Mol Cell Biol*, 20(14):5269–75.

Arimbasseri, A. G., Rijal, K., and Maraia, R. J. (2013). Transcription termination by the eukaryotic RNA polymerase III. *Biochim Biophys Acta*, 1829(3-4):318–30.

Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A*, 98(18):10037–41.

Barad, B. A., Echols, N., Wang, R. Y., Cheng, Y., DiMaio, F., Adams, P. D., and Fraser, J. S. (2015). EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat Methods*, 12(10):943–6.

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(1):94–9.

Berger, I., Fitzgerald, D. J., and Richmond, T. J. (2004). Baculovirus expression system for heterologous multiprotein complexes. *Nat Biotechnol*, 22(12):1583–7.

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

Bischler, N., Brino, L., Carles, C., Riva, M., Tschochner, H., Mallouh, V., and Schultz, P. (2002). Localization of the yeast RNA polymerase I-specific subunits. *EMBO J*, 21(15):4136–44.

Blattner, C., Jennebach, S., Herzog, F., Mayer, A., Cheung, A. C., Witte, G., Lorenzen, K., Hopfner, K. P., Heck, A. J., Aebersold, R., and Cramer, P. (2011). Molecular basis of Rrn3-regulated RNA polymerase I initiation and cell growth. *Genes Dev*, 25(19):2093–105.

Bordi, L., Cioci, F., and Camilloni, G. (2001). In vivo binding and hierarchy of assembly of the yeast RNA polymerase I transcription factors. *Mol Biol Cell*, 12(3):753–60.

Brilot, A. F., Chen, J. Z., Cheng, A., Pan, J., Harrison, S. C., Potter, C. S., Carragher, B., Henderson, R., and Grigorieff, N. (2012). Beam-induced motion of vitrified specimen on holey carbon film. *J Struct Biol*, 177(3):630–7.

Buratowski, S., Hahn, S., Guarente, L., and Sharp, P. A. (1989). Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell*, 56(4):549–61.

Bushnell, D. A., Westover, K. D., Davis, R. E., and Kornberg, R. D. (2004). Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms. *Science*, 303(5660):983–8.

Cardone, G., Heymann, J. B., and Steven, A. C. (2013). One number does not fit all: mapping local variations in resolution in cryo-EM reconstructions. *J Struct Biol*, 184(2):226–36.

Castano-Diez, D., Al-Amoudi, A., Glynn, A. M., Seybert, A., and Frangakis, A. S. (2007). Fiducial-less alignment of cryo-sections. *J Struct Biol*, 159(3):413–23.

Chakraborty, A., Wang, D., Ebright, Y. W., Korlann, Y., Kortkhonjia, E., Kim, T., Chowdhury, S., Wigneshweraraj, S., Irschik, H., Jansen, R., Nixon, B. T., Knight, J., Weiss, S., and Ebright, R. H. (2012). Opening and closing of the bacterial RNA polymerase clamp. *Science*, 337(6094):591–5.

Chen, H. T. and Hahn, S. (2004). Mapping the location of TFIIB within the RNA polymerase II transcription preinitiation complex: a model for the structure of the PIC. *Cell*, 119(2):169–80.

Chen, V. B., Arendall, W. B., r., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*, 66(Pt 1):12–21.

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

Combe, C. W., Fischer, L., and Rappsilber, J. (2015). xiNET: cross-link network maps with residue resolution. *Mol Cell Proteomics*, 14(4):1137–47.

Conconi, A., Widmer, R. M., Koller, T., and Sogo, J. M. (1989). Two different chromatin structures coexist in ribosomal RNA genes throughout the cell cycle. *Cell*, 57(5):753–61.

Cowtan, K. (2010). Recent developments in classical density modification. *Acta Crystallogr D Biol Crystallogr*, 66(Pt 4):470–8.

Cramer, P., Bushnell, D. A., and Kornberg, R. D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science*, 292(5523):1863–76.

Denissov, S., van Driel, M., Voit, R., Hekkelman, M., Hulsen, T., Hernandez, N., Grummt, I., Wehrens, R., and Stunnenberg, H. (2007). Identification of novel functional TBP-binding sites and general factor repertoires. *EMBO J*, 26(4):944–54.

Derenzini, M., Trere, D., Pession, A., Govoni, M., Sirri, V., and Chieco, P. (2000). Nucleolar size indicates the rapidity of cell proliferation in cancer tissues. *J Pathol*, 191(2):181–6.

Drygin, D., Lin, A., Bliesath, J., Ho, C. B., O'Brien, S. E., Proffitt, C., Omori, M., Haddach, M., Schwaebe, M. K., Siddiqui-Jain, A., Streiner, N., Quin, J. E., Sanij, E., Bywater, M. J., Hannan, R. D., Ryckman, D., Anderes, K., and Rice, W. G. (2011). Targeting RNA polymerase I with an oral small molecule CX-5461 inhibits ribosomal RNA synthesis and solid tumor growth. *Cancer Res*, 71(4):1418–30.

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(4):706–16.

Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*, 66(Pt 4):486–501.

Engel, C., Plitzko, J., and Cramer, P. (2016). RNA polymerase I-Rrn3 complex at 4.8 A resolution. *Nat Commun*, 7:12129.

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

Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr*, 62(Pt 1):72–82.

Feklistov, A. and Darst, S. A. (2011). Structural basis for promoter-10 element recognition by the bacterial RNA polymerase sigma subunit. *Cell*, 147(6):1257–69.

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

Geiger, S. R., Lorenzen, K., Schreieck, A., Hanecker, P., Kostrewa, D., Heck, A. J., and Cramer, P. (2010). RNA polymerase I contains a TFIIF-related DNA-binding subcomplex. *Mol Cell*, 39(4):583–94.

Gietl, A. and Grohmann, D. (2013). Modern biophysical approaches probe transcription-factor-induced DNA bending and looping. *Biochem Soc Trans*, 41(1):368–73.

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(5523):1876–82.

Goetze, H., Wittner, M., Hamperl, S., Hondele, M., Merz, K., Stoeckl, U., and Griesenbeck, J. (2010). Alternative chromatin structures of the 35S rRNA genes in Saccharomyces cerevisiae provide a molecular basis for the selective recruitment of RNA polymerases I and II. *Mol Cell Biol*, 30(8):2028–45.

Goodfellow, S. J. and Zomerdijk, J. C. (2013). Basic mechanisms in RNA polymerase I transcription of the ribosomal RNA genes. *Subcell Biochem*, 61:211–36.

Gorski, J. J., Pathak, S., Panov, K., Kasciukovic, T., Panova, T., Russell, J., and Zomerdijk, J. C. (2007). A novel TBP-associated factor of SL1 functions in RNA polymerase I transcription. *EMBO J*, 26(6):1560–8.

Grimm, M., Zimniak, T., Kahraman, A., and Herzog, F. (2015). xVis: a web server for the schematic visualization and interpretation of crosslink-derived spatial restraints. *Nucleic Acids Res*, 43(W1):W362–9.

Grummt, I. (2013). The nucleolus-guardian of cellular homeostasis and genome integrity. *Chromosoma*, 122(6):487–97.

Grummt, I. and Voit, R. (2010). Linking rDNA transcription to the cellular energy supply. *Cell Cycle*, 9(2):225–6.

Grunberg, S. and Hahn, S. (2013). Structural insights into transcription initiation by RNA polymerase II. *Trends Biochem Sci*, 38(12):603–11.

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(8):788–96.

Ha, I., Lane, W. S., and Reinberg, D. (1991). Cloning of a human gene encoding the general transcription initiation factor IIB. *Nature*, 352(6337):689–95.

Hamperl, S., Wittner, M., Babl, V., Perez-Fernandez, J., Tschochner, H., and Griesenbeck, J. (2013). Chromatin states at ribosomal DNA loci. *Biochim Biophys Acta*, 1829(3-4):405–17.

He, Y., Yan, C., Fang, J., Inouye, C., Tjian, R., Ivanov, I., and Nogales, E. (2016). Near-atomic resolution visualization of human transcription promoter opening. *Nature*, 533(7603):359–65.

Hoffmann, N. A., Jakobi, A. J., Moreno-Morcillo, M., Glatt, S., Kosinski, J., Hagen, W. J., Sachse, C., and Müller, C. W. (2015). Molecular structures of unbound and transcribing RNA polymerase III. *Nature*, 528(7581):231–6.

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. *Molecular and Cellular Biology*, 28(21):6709–6719.

Hussain, T., Llacer, J. L., Fernandez, I. S., Munoz, A., Martin-Marcos, P., Savva, C. G., Lorsch, J. R., Hinnebusch, A. G., and Ramakrishnan, V. (2014). Structural changes enable start codon recognition by the eukaryotic translation initiation complex. *Cell*, 159(3):597–607.

Ide, S., Miyazaki, T., Maki, H., and Kobayashi, T. (2010). Abundance of ribosomal RNA gene copies maintains genome integrity. *Science*, 327(5966):693–6.

Jaiswal, R., Choudhury, M., Zaman, S., Singh, S., Santosh, V., Bastia, D., and Escalante, C. R. (2016). Functional architecture of the Reb1-Ter complex of Schizosaccharomyces pombe. *Proc Natl Acad Sci U S A*, 113(16):E2267–76.

Jennebach, S., Herzog, F., Aebersold, R., and Cramer, P. (2012). Crosslinking-MS analysis reveals RNA polymerase I domain architecture and basis of rRNA cleavage. *Nucleic Acids Res*, 40(12):5591–601.

Jovanovic, M., Burrows, P. C., Bose, D., Camara, B., Wiesler, S., Zhang, X., Wigneshweraraj, S., Weinzierl, R. O., and Buck, M. (2011). Activity map of the Escherichia coli RNA polymerase bridge helix. *J Biol Chem*, 286(16):14469–79.

Kabsch, W. (2010). Xds. *Acta Crystallogr D Biol Crystallogr*, 66(Pt 2):125–32.

Kastner, B., Fischer, N., Golas, M. M., Sander, B., Dube, P., Boehringer, D., Hartmuth, K., Deckert, J., Hauer, F., Wolf, E., Uchtenhagen, H., Urlaub, H., Herzog, F., Peters, J. M., Poerschke, D., Luhrmann, R., and Stark, H. (2008). GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat Methods*, 5(1):53–5.

Keener, J., Dodd, J. A., Lalo, D., and Nomura, M. (1997). Histones H3 and H4 are components of upstream activation factor required for the high-level transcription of yeast rDNA by RNA polymerase I. *Proc Natl Acad Sci U S A*, 94(25):13458–62.

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(50):33795–802.

Kettenberger, H., Armache, K. J., and Cramer, P. (2003). Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell*, 114(3):347–57.

Kettenberger, H., Armache, K. J., and Cramer, P. (2004). Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol Cell*, 16(6):955–65.

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(7):887–903.

Kim, T. K., Ebright, R. H., and Reinberg, D. (2000). Mechanism of ATP-dependent promoter melting by transcription factor IIH. *Science*, 288(5470):1418–22.

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

Knutson, B. A. and Hahn, S. (2013). TFIIB-related factors in RNA polymerase I transcription. *Biochim Biophys Acta*, 1829(3-4):265–73.

Knutson, B. A., Luo, J., Ranish, J., and Hahn, S. (2014). Architecture of the Saccharomyces cerevisiae RNA polymerase I Core Factor complex. *Nat Struct Mol Biol*, 21(9):810–6.

Koning, R. I., Celler, K., Willemse, J., Bos, E., van Wezel, G. P., and Koster, A. J. (2014). Correlative cryo-fluorescence light microscopy and cryo-electron tomography of Streptomyces. *Methods Cell Biol*, 124:217–39.

Kostrewa, D., Kuhn, C. D., Engel, C., and Cramer, P. (2015). An alternative RNA polymerase I structure reveals a dimer hinge. *Acta Crystallogr D Biol Crystallogr*, 71(Pt 9):1850–5.

Kostrewa, D., Zeller, M. E., Armache, K. J., Seizl, M., Leike, K., Thomm, M., and Cramer, P. (2009). RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature*, 462(7271):323–30.

Krissinel, E. and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr*, 60(Pt 12 Pt 1):2256–68.

Kuhn, C. D., Geiger, S. R., Baumli, S., Gartmann, M., Gerber, J., Jennebach, S., Mielke, T., Tschochner, H., Beckmann, R., and Cramer, P. (2007). Functional architecture of RNA polymerase I. *Cell*, 131(7):1260–72.

Kunz, M. and Frangakis, A. S. (2014). Super-sampling SART with ordered subsets. *J Struct Biol*, 188(2):107–15.

Kunz, M. and Frangakis, A. S. (2016). Three-dimensional CTF correction improves the resolution of electron tomograms. *J Struct Biol*.

Kuznedelov, K., Korzheva, N., Mustaev, A., and Severinov, K. (2002). Structure-based analysis of RNA polymerase function: the largest subunit's rudder contributes critically to elongation complex stability and is not involved in the maintenance of RNA-DNA hybrid length. *EMBO J*, 21(6):1369–78.

Lalo, D., Steffan, J. S., Dodd, J. A., and Nomura, M. (1996). RRN11 encodes the third subunit of the complex containing Rrn6p and Rrn7p that is essential for the initiation of rDNA transcription by yeast RNA polymerase I. *J Biol Chem*, 271(35):21062–7.

Leitner, A., Faini, M., Stengel, F., and Aebersold, R. (2016). Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines. *Trends Biochem Sci*, 41(1):20–32.

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(6):584–90.

Lin, Y. S., Ha, I., Maldonado, E., Reinberg, D., and Green, M. R. (1991). Binding of general transcription factor TFIIB to an acidic activating region. *Nature*, 353(6344):569–71.

Machin, F., Torres-Rosell, J., De Piccoli, G., Carballo, J. A., Cha, R. S., Jarmuz, A., and Aragon, L. (2006). Transcription of ribosomal genes can cause nondisjunction. *J Cell Biol*, 173(6):893–903.

Malik, S., Hisatake, K., Sumimoto, H., Horikoshi, M., and Roeder, R. G. (1991). Sequence of general transcription factor TFIIB and relationships to other initiation factors. *Proc Natl Acad Sci U S A*, 88(21):9553–7.

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(7):1302–10.

Matthews, B. W. (1968). Solvent content of protein crystals. *J Mol Biol*, 33(2):491–7.

Mayer, C., Bierhoff, H., and Grummt, I. (2005). The nucleolus as a stress sensor: JNK2 inactivates the transcription factor TIF-IA and down-regulates rRNA synthesis. *Genes Dev*, 19(8):933–41.

Mayer, C., Schmitz, K. M., Li, J., Grummt, I., and Santoro, R. (2006). Intergenic transcripts regulate the epigenetic state of rRNA genes. *Mol Cell*, 22(3):351–61.

Meier, A. and Thoma, F. (2005). RNA polymerase I transcription factors in active yeast rRNA gene promoters enhance UV damage formation and inhibit repair. *Mol Cell Biol*, 25(5):1586–95.

Miller, O. L. and Beatty, B. R. (1969). Visualization of nucleolar genes. *Science (New York, N.Y.)*, 164(3882):955–957.

Mindell, J. A. and Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol*, 142(3):334–47.

Moss, T. (2004). At the crossroads of growth control; making ribosomal RNA. *Curr Opin Genet Dev*, 14(2):210–7.

Moss, T., Langlois, F., Gagnon-Kugler, T., and Stefanovsky, V. (2007). A housekeeper with power of attorney: the rRNA genes in ribosome biogenesis. *Cell Mol Life Sci*, 64(1):29–49.

Murakami, K. S., Masuda, S., and Darst, S. A. (2002). Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 A resolution. *Science*, 296(5571):1280–4.

Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*, 53(Pt 3):240–55.

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

Narayan, K. and Subramaniam, S. (2015). Focused ion beams in biology. *Nat Methods*, 12(11):1021–31.

Nemeth, A., Perez-Fernandez, J., Merkl, P., Hamperl, S., Gerber, J., Griesenbeck, J., and Tschochner, H. (2013). RNA polymerase I termination: Where is the end? *Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms*, 1829(3-4):306–317.

Neyer, S., Kunz, M., Geiss, C., Hantsche, M., Hodirnau, V. V., Seybert, A., Engel, C., Scheffer, M. P., Cramer, P., and Frangakis, A. S. (2016). Structure of RNA polymerase I transcribing ribosomal DNA genes. *Nature*.

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

Novello, F. and Stirpe, F. (1970). Simultaneous assay of RNA polymerase I and II in nuclei isolated from resting and growing rat liver with the use of alpha-amanitin. *FEBS Lett*, 8(1):57–60.

Osheim, Y. N., French, S. L., Keck, K. M., Champion, E. A., Spasov, K., Dragon, F., Baserga, S. J., and Beyer, A. L. (2004). Pre-18S ribosomal RNA is structurally compacted into the SSU processome prior to being cleaved from nascent transcripts in Saccharomyces cerevisiae. *Molecular cell*, 16(6):943–954.

Osheim, Y. N., French, S. L., Sikes, M. L., and Beyer, A. L. (2009). Electron microscope visualization of RNA transcription and processing in Saccharomyces cerevisiae by Miller chromatin spreading. *Methods Mol Biol*, 464:55–69.

Panov, K. I., Friedrich, J. K., and Zomerdijk, J. C. (2001). A step subsequent to preinitiation complex assembly at the ribosomal RNA gene promoter is rate limiting for human RNA polymerase I-dependent transcription. *Mol Cell Biol*, 21(8):2641–9.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004). UCSF Chimera–a visualization system for exploratory research and analysis. *J Comput Chem*, 25(13):1605–12.

Peyroche, G., Milkereit, P., Bischler, N., Tschochner, H., Schultz, P., Sentenac, A., Carles, C., and Riva, M. (2000). The recruitment of RNA polymerase I on rDNA is mediated by the interaction of the A43 subunit with Rrn3. *EMBO J*, 19(20):5473–82.

Pfeffer, S., Burbaum, L., Unverdorben, P., Pech, M., Chen, Y., Zimmermann, R., Beckmann, R., and Forster, F. (2015). Structure of the native Sec61 protein-conducting channel. *Nat Commun*, 6:8403.

Pilsl, M., Crucifix, C., Papai, G., Krupp, F., Steinbauer, R., Griesenbeck, J., Milkereit, P., Tschochner, H., and Schultz, P. (2016). Structure of the initiation-competent RNA polymerase I and its implication for transcription. *Nat Commun*, 7:12126.

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

Plaschka, C., Lariviere, L., Wenzeck, L., Seizl, M., Hemann, M., Tegunov, D., Petrotchenko, E. V., Borchers, C. H., Baumeister, W., Herzog, F., Villa, E., and Cramer, P. (2015). Architecture of the RNA polymerase II-Mediator core initiation complex. *Nature*, 518(7539):376–80.

Poortinga, G., Quinn, L. M., and Hannan, R. D. (2015). Targeting RNA polymerase I to treat MYC-driven cancer. *Oncogene*, 34(4):403–12.

Pruggnaller, S., Mayr, M., and Frangakis, A. S. (2008). A visualization and segmentation toolbox for electron microscopy. *J Struct Biol*, 164(1):161–5.

Rial, D. V. and Ceccarelli, E. A. (2002). Removal of DnaK contamination during fusion protein purifications. *Protein Expr Purif*, 25(3):503–7.

Robinson, P. J., Trnka, M. J., Bushnell, D. A., Davis, R. E., Mattei, P. J., Burlingame, A. L., and Kornberg, R. D. (2016). Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. *Cell*, 166(6):1411–1422 e16.

Roeder, R. G. (1996). The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci*, 21(9):327–35.

Roeder, R. G. and Rutter, W. J. (1969). Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature*, 224(5216):234–7.

Rohou, A. and Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol*, 192(2):216–21.

Ruan, W., Lehmann, E., Thomm, M., Kostrewa, D., and Cramer, P. (2011). Evolution of two modes of intrinsic RNA polymerase transcript cleavage. *J Biol Chem*, 286(21):18701–7.

Russell, J. and Zomerdijk, J. C. (2005). RNA-polymerase-I-directed rDNA transcription, life and works. *Trends Biochem Sci*, 30(2):87–96.

Russell, J. and Zomerdijk, J. C. (2006). The RNA polymerase I transcription machinery. *Biochem Soc Symp*.

Sainsbury, S., Bernecky, C., and Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. *Nat Rev Mol Cell Biol*, 16(3):129–43.

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

Scheres, S. H. (2012). A Bayesian view on cryo-EM structure determination. *J Mol Biol*, 415(2):406–18.

Schneider, D. A., French, S. L., Osheim, Y. N., Bailey, A. O., Vu, L., Dodd, J., Yates, J. R., Beyer, A. L., and Nomura, M. (2006). RNA polymerase II elongation factors Spt4p and Spt5p play roles in transcription elongation by RNA polymerase I and rRNA processing. *Proc Natl Acad Sci U S A*, 103(34):12707–12.

Schuetz, A., Allali-Hassani, A., Martin, F., Loppnau, P., Vedadi, M., Bochkarev, A., Plotnikov, A. N., Arrowsmith, C. H., and Min, J. (2006). Structural basis for molecular recognition and presentation of histone H3 by WDR5. *EMBO J*, 25(18):4245–52.

Schultz, P., Célia, H., Riva, M., Sentenac, A., and Oudet, P. (1993). Three-dimensional model of yeast RNA polymerase I determined by electron microscopy of two-dimensional crystals. *The EMBO journal*, 12(7):2601–2607.

Sheldrick, G. M. (2010). Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr*, 66(Pt 4):479–85.

Smart, O. S., Womack, T. O., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonrhein, C., and Bricogne, G. (2012). Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr D Biol Crystallogr*, 68(Pt 4):368–80.

Sosunov, V., Sosunova, E., Mustaev, A., Bass, I., Nikiforov, V., and Goldfarb, A. (2003). Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerase. *EMBO J*, 22(9):2234–44.

Steffan, J. S., Keys, D. A., Vu, L., and Nomura, M. (1998). Interaction of TATA-binding protein with upstream activation factor is required for activated transcription of ribosomal DNA by RNA polymerase I in Saccharomyces cerevisiae in vivo. *Molecular and Cellular Biology*, 18(7):3752–3761.

Sydow, J. F. and Cramer, P. (2009). RNA polymerase fidelity and transcriptional proofreading. *Curr Opin Struct Biol*, 19(6):732–9.

Tafur, L., Sadian, Y., Hoffmann, N. A., Jakobi, A. J., Wetzel, R., Hagen, W. J., Sachse, C., and Muller, C. W. (2016). Molecular Structures of Transcribing RNA Polymerase I. *Mol Cell*.

Tagami, S., Sekine, S., Kumarevel, T., Hino, N., Murayama, Y., Kamegamori, S., Yamamoto, M., Sakamoto, K., and Yokoyama, S. (2010). Crystal structure of bacterial RNA polymerase bound with a transcription inhibitor protein. *Nature*, 468(7326):978–82.

Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I., and Ludtke, S. J. (2007). EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol*, 157(1):38–46.

van Dijk, M. and Bonvin, A. M. (2009). 3D-DART: a DNA structure modelling server. *Nucleic Acids Res*, 37(Web Server issue):W235–9.

Vannini, A. and Cramer, P. (2012). Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol Cell*, 45(4):439–46.

Vassylyev, D. G., Vassylyeva, M. N., Perederina, A., Tahirov, T. H., and Artsimovitch, I. (2007). Structural basis for transcription elongation by bacterial RNA polymerase. *Nature*, 448(7150):157–62.

Vu, L., Siddiqi, I., Lee, B. S., Josaitis, C. A., and Nomura, M. (1999). RNA polymerase switch in transcription of yeast rDNA: role of transcription factor UAF (upstream activation factor) in silencing rDNA transcription by RNA polymerase II. *Proc Natl Acad Sci U S A*, 96(8):4390–5.

Walzthoeni, T., Joachimiak, L. A., Rosenberger, G., Rost, H. L., Malmstrom, L., Leitner, A., Frydman, J., and Aebersold, R. (2015). xTract: software for characterizing conformational changes of protein complexes by quantitative cross-linking mass spectrometry. *Nat Methods*, 12(12):1185–90.

Weixlbaumer, A., Leon, K., Landick, R., and Darst, S. A. (2013). Structural basis of transcriptional pausing in bacteria. *Cell*, 152(3):431–41.

Werner, F. and Grohmann, D. (2011). Evolution of multisubunit RNA polymerases in the three domains of life. *Nat Rev Microbiol*, 9(2):85–98.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr*, 67(Pt 4):235–42.

Woolford, J. L., J. and Baserga, S. J. (2013). Ribosome biogenesis in the yeast Saccharomyces cerevisiae. *Genetics*, 195(3):643–81.

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

Zuo, Y. and Steitz, T. A. (2015). Crystal structures of the E. coli transcription initiation complexes with a complete bubble. *Mol Cell*, 58(3):534–40.

Glossary

E. coli *Escherichia coli*

S. cerevisiae *Saccharomyces cerevisiae*

X. laevis *Xenopus laevis*

ATP adenosine triphosphate

AU asymmetric unit

bp base pair

BS3 bis(sulfosuccinimidyl)suberate

CC closed complex

CE core element

CF core factor

cryo-EM cryo-electron microscopy

cryo-ET cryo-electron tomography

cryo-SP single-particle cryo-electron microscopy

CTF contrast transfer function

CTP cytidine triphosphate

CV column volume

DNA deoxyribonucleic acid

dsDNA double stranded DNA

DTT dithiothreitol

EC elongation complex

EM electron microscopy

FSC Fourier Shell Correlation

GraFix gradient fixation

GTP guanosine triphosphate

IGS Intergenic spacer

IPTG isopropyl-β-D-thiogalactopyranosid

ITC initially transcribing complex

k thousand

LB lysogeny broth

MR molecular replacement

mRNA messenger RNA

NCS non-crystallographic symmetry

Ni-NTA nickel-nitrilotriacetic acid

NTP nucleoside triphosphate

OC open complex

PAD propeller- and promoter-associated domain

PBS phosphate buffered saline

PIC pre-initiation complex

PIR polymerase interacting regions

Pol RNA polymerase

Pols RNA polymerases

RB reaction buffer

rDNA ribosomal DNA

rmsd root-mean-square deviation

RNA ribonucleic acid

rRNA ribosomal RNA

RT room temperature

SAD single-wavelength anomalous diffraction

SAXS small-angle x-ray scattering

SDS-PAGE sodium dodecyl sulfate polyacrylamide gel electrophoresis

TBP TATA-box binding protein

TPR tetratricopeptide repeat

tRNA transfer RNA

TSS transcription start site

UAF upstream activating factor

UAS upstream activating sequence

UTP uridine triphosphate

XL-MS cross-linking coupled to mass spectrometry

List of Figures

List of Tables