

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

# Crystal structure of Core Factor and its use to analyze RNA polymerase I initiation



Tobias Michael Gubbey

aus

Garmisch-Partenkirchen, Deutschland

2017

## **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 24.4.2017

Tobias Michael Gubbey

Dissertation eingereicht am 27.04.2017

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

Mündliche Prüfung am 29.05.2017

“Never, never, never give up.”

Winston Churchill

## **Acknowledgements**

I would like to thank Patrick Cramer. I'm extremely grateful to him for giving me the opportunity to do my PhD in his lab. Furthermore, for providing an outstanding scientific environment in an international and interdisciplinary research group.

I would like to thank Simon and Christoph for this really fruitful collaboration which resulted in a very nice publication. Additionally, I would like to thank Sarah, Christiane, Carlo, Carrie for their contribution.

Special thanks go Christoph for supervision during my master thesis at the Gene Center in Munich, and for supporting my transfer from a master student to a PhD student. I really appreciate the support, scientific and nonscientific discussion, advice while I was facing difficult scientific problems and especially for things which included whisky and melons.

I would like to thank the whole AG Cramer for all the nice time during my PhD period in Göttingen and a little bit in Munich for inspiration, scientific discussion, coffee breaks and keeping the lab running well. I would like to specially thank: Kirsten, Simon, Sarah, Carrie, Merle, Hauke, Katharina, Paulina, Christoph, Youwei, Anna, Simon, Saskia, Janine, Sofia, Christiane, Kerstin, Petra, Sigurd, Claudia, Stefan, Jürgen, Angelika, Manuela, and everybody I forgot. Special thanks to Dietmar for having oral master biochemistry exams together, it was always fun.

Furthermore, I would like to thank Takashi Tomizaki and all the beamline scientists from the Paul Scherrer Institut for the support during synchrotron trips. Also I would like to thank all the people I went to the synchrotron with for their cooperation and support during sleepy times at 3 am in the morning.

Very special thanks go to Merle, Sarah, Christoph and Simon for critical reading of my thesis. I also would like to thank Patrick Cramer, Dietmar Martin, Andreas Ladurner, Franz Herzog, Karl-Peter Hopfner and Karl-Klaus Conzelmann for being examiners at my PhD exam.

Mein größter Dank geht an meine Eltern, meine Schwester und meinen Großeltern die mich über all die Jahre immer unterstützt, an mich geglaubt und gefördert haben. Ohne euch wäre ich nie so weit gekommen. Ein ganz besonderes Dankeschön geht an Sarah für ihre Unterstützung, Geduld und Liebe.

## Summary

Gene transcription is carried out by three different RNA polymerases (Pols) known as Pol I, Pol II and Pol III. Up to 60% of total cellular RNA is produced by Pol I which synthesizes the 35S ribosomal RNA precursor. Therefore, tight and precise regulation of Pol I transcription is indispensable. This regulation is primarily achieved at the level of transcription initiation by the Pol I-specific initiation factors Rrn3 and Core Factor (CF). Pol I exists in an equilibrium of transcriptionally inactive dimers and monomers that can undergo activation. Rrn3 can bind to the monomeric form of Pol I, establishing an initiation-competent form of the enzyme. The Pol I-Rrn3 complex is subsequently recruited to CF, which assembles on a core element (CE) of the promoter. This is sufficient to initiate transcription from a minimal promoter template *in vitro*, thus defining a basal initiation system. Crystal structures of Pol I, and Rrn3, as well as cryo-electron microscopy (cryo-EM) reconstructions of monomeric and dimeric Pol I, actively transcribing Pol I and the Pol I-Rrn3 complex are known. However, no details of the atomic structure of CF were available. A high-resolution structure of this heterotrimeric complex consisting of Rrn6, Rrn7 and Rrn11 is the last missing piece on the way towards understanding the mechanisms of Pol I initiation. Here I present the recombinant expression, purification, crystallization and *de novo* built structure of the 220 kDa *Saccharomyces cerevisiae* CF at 3.2 Angstrom resolution. The structure reveals a bi-modular architecture with an intertwined structure of the subunits Rrn6, Rrn7, and Rrn11, including novel folds and unexpected differences to the Pol II initiation system. Rrn6 is split into two structured parts, a seven bladed N-terminal WD40 like  $\beta$ -propeller and a new helical domain which is defined as 'headlock domain'. Rrn7 is homologous to the transcription factor (TF) IIB and contains two cyclin domains, while the N-terminal part is disordered in the crystal. Cyclin I is similar to its TFIIB-equivalent whereas Cyclin II strongly differs. It contains a 141 amino acid long insertion before two final helices complete the canonical fold. Rrn11 has an N-terminal propeller/promoter-associated domain (PAD) and a C-terminal tetratricopeptide repeat domain. Cryo-EM on Pol I initiation complexes was carried out by Christoph Engel on the Pol I-Rrn3-CF complex, and by Simon Neyer on the initially transcribing Pol I-Rrn3-CF complex. The crystal structure of CF combined with cryo-EM enabled the building of molecular models for basal Pol I initiation. This shows how CF recognizes and binds the promoter DNA through interaction with the Rrn11 PAD. Furthermore, it reveals three Polymerase Interacting Regions (PIRs) of the CF. In addition our structural analysis suggests that unwinding of the dsDNA occurs between the protrusion and clamp domains of Pol I. In the initially transcribing complex a cleft contraction is visible that enables RNA synthesis. The results of this shared study elucidate the unique architecture of the Pol I initiation complex. Comparison to its Pol II counterpart shows a remarkable difference in the trajectories of the upstream DNA. Further analysis of Pol I promoters indicates that the initiation of Pol I may depend on the 'bendability' and 'meltability' of the promoter region rather than a specific sequence. Thereby, our study expands the understanding of Pol I transcription initiation by adding a novel, unique crystal structure and by expanding the in-depth understanding of the underlying molecular mechanisms.

## Publication

Christoph Engel\*, **Tobias Gubbey\***, Simon Neyer\*, Sarah Sainsbury, Christiane Oberthuer, Carlo Baejen, Carrie Bernecky, Patrick Cramer. (2017) Structural Basis of RNA Polymerase I Transcription Initiation. *Cell* Volume 169, Issue 1, Pages 120-131.

\*These authors contributed equally.

<http://dx.doi.org/10.1016/j.cell.2017.03.003>

CE initiated the project, planned and carried out EM structure determination of the Pol I-Rrn3-CF complex, and planned and carried out functional assays. TG crystallized CF and collected diffraction data. TG, CE and SS analyzed crystallographic data and solved the CF crystal structure. SN planned and carried out EM structure determination of the ITC, and contributed to functional assay design. CO, TG, SN and CE cloned, expressed and purified proteins. CBa 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.

<b>Table of Contents</b>	
<b>Erklärung</b>	<b>2</b>
<b>Eidesstattliche Versicherung</b>	<b>2</b>
<b>Acknowledgements</b>	<b>4</b>
<b>Summary</b>	<b>5</b>
<b>Publication</b>	<b>6</b>
<b>1. Introduction</b>	<b>10</b>
1.1. RNA Polymerases	10
1.2. RNA-Polymerases I transcription initiation	12
1.3. Structural Studies	13
1.3.1. Rrn3 structure	14
1.3.2. RNA Polymerase I structure	16
1.3.3. RNA-Polymerase I elongation complex	18
1.3.4. Polymerase I – Rrn3 complex	19
1.3.5. Core Factor studies	20
1.4. Aims and scope of this study	22
<b>2. Materials and Methods</b>	<b>23</b>
2.1. Materials	23
2.1.1. Bacterial strains	23
2.1.2. Media, additives and plates for cell growth	23
2.1.3. Buffers	23
2.1.4. Markers, solution and enzymes	24
2.2. General Methods	26
2.2.1. Synthesis of methionine point mutations for structure validation	26
2.2.2. Primer design for quick change polymerase chain reaction.	26
2.2.3. Quick change Polymerase chain reaction	26
2.2.4. DpnI digest	27
2.2.5. Agarose gel plasmid purification, isolation	27
2.2.6. Competent cells	27
2.2.7. Plasmid transformation and plasmid purification, isolation and verification	27

2.2.8. SDS-PAGE	28
2.2.9. Silver gel staining	28
2.2.10. Protein concentration measurement and precipitation	28
2.2.11. PCT™ Pre-Crystallization Test	29
2.3. Specific methods	30
2.3.1. Expression of recombinant Core Factor	30
2.3.2. Purification of Core Factor	31
2.3.3. Crystallization of Core Factor	32
2.3.4. Crystallization of Core Factor	32
<b>3. Results and Discussion</b>	<b>34</b>
3.1. Structural basis of RNA polymerase I transcription initiation	34
3.2. Crystal structure of Core Factor	34
3.3. Core Factor comprises two modules	40
3.4. Rrn7 differs from TFIIB	40
3.5. Structure of the Pol I-Rrn3-CF complex	45
3.6. CF contains three Pol I-interacting regions	47
3.7. Structure of the initially transcribing complex	47
3.8. Promoter DNA location and contacts	49
3.9. Basal Pol I initiation assay	51
3.10. Determinants of Pol I initiation	54
3.11. Models of the CC and OC	54
3.12. Discussion	56
<b>4. Conclusion and Outlook</b>	<b>59</b>
4.1. Future crystallization experiments	59
4.2. More complete understanding of transcription initiation	60
4.3. Targeting Pol I initiation for Cancer Therapeutics	61
<b>5. Appendix</b>	<b>62</b>
5.1. Supplemental data.	62
5.1.1. Pol I cleft expansion and contraction.	62
5.1.2. Formation and negative stain EM reconstruction of a Pol I-Rrn3-CF complex	63
5.1.3. Classification of the Pol I-Rrn3-CF cryo-EM dataset	65

5.1.4. Three Pol I-Rrn3-CF reconstructions	66
5.1.5. Pol I ITC sample preparation and cryo-EM data processing	67
5.1.6. Cryo-EM reconstructions of Pol I-Rrn3-CF ITC	68
5.1.7. Pol I-specific CF function in transcription initiation	70
5.2. Extended Material, Methods and Data	72
5.2.1. Preparation and cryo-EM analysis of Pol I-Rrn3-CF complex	72
5.2.2. Negative stain EM of the Pol I-Rrn3-CF complex	73
5.2.3. Preparation of Pol I ITC complexes	74
5.2.4. ITC single-particle cryo-EM and image processing	75
5.2.5. Structural modelling of the ITC and CC	76
5.2.6. Pol I promoter-specific transcription initiation assay	77
5.3. STAR Methods	78
5.4. Phenix Refinement	83
5.4.1. Phenix input script	83
5.4.2. NCS group parameters	83
5.4.3. Rigid body parameters	84
5.4.4. TLS parameters	84
5.4.5. Secondary structure definition	85
<b>6. References</b>	<b>90</b>
<b>7. Abbreviations</b>	<b>99</b>
<b>8. List of Figures</b>	<b>100</b>
<b>9. List of Tables</b>	<b>101</b>

# 1. Introduction

## 1.1. RNA Polymerases

The eukaryotic genome is transcribed by three different RNA polymerases (Pol) which are large multi-protein assemblies. They were discovered and described four decades ago by their chromatographic separation (Roeder and Rutter, 1969, Roeder and Rutter, 1970). Pol I, Pol II and Pol III are mainly responsible for the synthesis of the 35S ribosomal RNA (rRNA) precursor, messenger RNA and transfer RNAs, respectively (Moss et al., 2007).

In *S. cerevisiae* Pol I (589 kDa, 14 subunits), Pol II (514 kDa, 12 subunits) and Pol III (693 kDa, 17 subunits) share a structurally conserved core which consists of 10 subunits (Cramer et al., 2008). Five of the subunits are shared between Pol I, Pol II and Pol III and are Rpb5, Rpb6, Rpb8, Rpb10 and Rpb12 (Table 1). Homologous subunits except for Rpb8 exist in the archaeal polymerase as well (Minakhin et al., 2001). AC40 and AC19 are shared between Pol I and Pol III and are homologous to Rpb3 and Rpb11 in Pol II. The two largest subunits are unique to each Pol and together form the active site and cleft of the enzyme (Vannini and Cramer, 2012) (Table 1). The Pol II shows weaker intrinsic RNA cleavage activity than the Pol I and Pol III counterparts A12.2 and C11, respectively (Ruan et al., 2011). The stalk of polymerases I, II, III are formed by the subcomplexes A14/43, Rpb4/7 and C17/25, respectively (Hu et al., 2002, Peyroche et al., 2002, Armache et al., 2005, Jasiak et al., 2006, Kuhn et al., 2007). Pol I and Pol III have additional sub-complexes: Pol I contains A49/34 (Geiger et al., 2010) and Pol III the homologous C37/53 (Kassavetis et al., 2010, Landrieux et al., 2005). These sub-complexes are counterparts of the general transcription factors (GTF) of Pol II TFIIF, IIE and fulfill further functions like initiation complex stabilization (Vannini and Cramer, 2012) (Table 1). Pol III has an additional sub-complex, which consists of C82/34/31 (Wang and Roeder, 1997).

In all three Pol's initiation systems additional factors are required. TATA-binding protein (TBP) functions in all three different transcription initiation systems however its role is apparently fundamentally different in Pol I initiation (Keener et al., 1998). Rrn7 (Pol I), TFIIB (Pol II) and Brf1/Brf2 (Pol III) are structurally homologous involved in TBP/polymerase binding as well as in start site selection and DNA opening (Vannini and Cramer, 2012, Naidu et al., 2011, Knutson and Hahn, 2011).

The structural basis of Pol II transcription initiation is more fully understood compared with Pol I and Pol III due to recent cryo-EM studies (He et al., 2016, Plaschka et al., 2016, Robinson et al., 2016). In Pol II initiation the closed complex (CC) is formed by the assembly of Pol II, promoter DNA, TBP, TFIIB, TFIID, TFIIE, TFIIIF, and TFIIH (Buratowski et al., 1989, Grunberg and Hahn, 2013, Roeder, 1996, Sainsbury et al., 2015). ATP hydrolysis by TFIIH is accompanied by DNA unwinding and loading of DNA into the cleft, forming the open complex (OC) (Grunberg et al., 2012). Transcription initiation begins from the transcription start site (TSS) and leads to synthesis of an initial RNA (Initially transcribing complex – ITC).

**Table 1. RNA polymerases subunits and their relation in *S. cerevisiae*.**

Table was adapted from (Vannini and Cramer, 2012).

Polymerase part	Pol I	Pol II	Pol III	Function
<b>Polymerase Core</b>				
	A190	Rpb1	C160	Active center
	A135	Rpb2	C128	Active center
	AC40	Rpb3	AC40	
	AC19	Rpb11	AC19	
	A12.2 <sup>A</sup>	Rpb9	C11 <sup>A</sup>	RNA cleavage
	A12.2 <sup>B</sup>	TFIIS <sup>B,C</sup>	C11 <sup>B</sup>	RNA cleavage
	Rpb5	Rpb5	Rpb5	
	Rpb6	Rpb6	Rpb6	
	Rpb8	Rpb8	Rpb8	
	Rpb10	Rpb10	Rpb10	
	Rpb12	Rpb12	Rpb12	
<b>Polymerase Stalk</b>				
	A14	Rpb4	C17	Initiation and complex formation
	A43	Rpb7	C25	
<b>Subcomplex</b>	<b>A49/34.5</b>	<b>TFIIIF/TFIIE<sup>D/E</sup></b>	<b>C37/53</b>	
	A49 <sup>A</sup>		C37	Initiation complex stabilization and start site selection
	A 34.5		C53	
<b>Subcomplex</b>			<b>C82/34/21</b>	
			C34	
			C82	Open complex stabilization
			C31	
<b>Total</b>	<b>14 subunits</b>	<b>12 subunits</b>	<b>17 subunits</b>	<b>-</b>

A) N-terminal ribbon

B) C-terminal ribbon

C) TFIIS C-terminal ribbon is not part of the Pol II core but is functionally related to A12.2 in Pol I and C11 in Pol III:

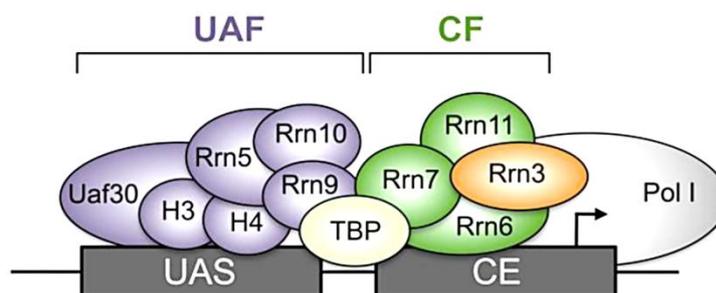
D) TFIIIF a GTF comprises Tfg1 and Tfg2.

E) TFIIE a GTF comprises Tfa1 and Tfa2.

## 1.2. RNA-Polymerases I transcription initiation

Pol I transcription is regulated by the assembly of the complete pre-initiation complex (PIC) (Aprikian et al., 2001). Pol I has two specific promoter elements. While the core element (CE) partially overlaps with the transcription start site (TSS), the upstream activating sequence (UAS) is located approximately 100 base pairs upstream of the TSS (Schneider, 2012, Drygin, 2010) (Figure 1). The complete Pol I PIC consists of four different initiation factors: the upstream activating factor (UAF) which binds the UAS, TBP, Core Factor (CF) which binds the CE, and Rrn3 (Keys et al., 1996, Keener et al., 1997, Peyroche et al., 2000). CF consists of Rrn6, Rrn7, and Rrn11 (Lalo et al., 1996, Lin et al., 1996) and UAF consists of Uaf30, Rrn9, Rrn10, Rrn5, and histones H3, and H4 (Keys et al., 1996, Keener et al., 1997) (Figure 1).

The PIC assembly is achieved in a specific way. The first step is the binding event of UAF to UAS followed by the recruitment of TBP. The two final steps are the binding of CF and the recruitment of Rrn3 and Pol I to the promoter (Aprikian et al., 2001). Pol I is transcriptionally inactive in its dimeric form (Engel et al., 2017, Fernandez-Tornero et al., 2013, Milkereit et al., 1997, Torreira et al., 2017). This dimeric complex can break up subsequently with the monomeric form being stabilized in its transcription competent state by the binding of Rrn3 at the Pol I subcomplex A14/34 known as the stalk (PilsI et al., 2016, Blattner et al., 2011, Engel et al., 2016, Torreira et al., 2017). *In vitro* data suggest that UAF stays tightly bound to the promoter DNA, while CF, TBP, and Rrn3 are released (Aprikian et al., 2001).



**Figure 1. Assembled yeast Pol I preinitiation complex.**

Illustration of the assembled Pol I pre-initiation complex on the rDNA promoter. The promoter consists of two elements which are called Upstream Activation Sequence (UAS) and Core Element (CE) with which the Upstream Activating Factor (UAF) and Core Factor (CF) are binding respectively. Figure was adapted from (Knutson and Hahn, 2013).

### 1.3. Structural Studies

Pol I is in the focus of interest for structural studies because of its importance for ribosome biogenesis and cell growth (Moss et al., 2007, Grummt, 2003). The first complete structural study in the Pol I transcription system revealed the position of the stalk (A14/43) and indicated a position of A49 above the clamp (Bischler et al., 2002). Later the crystal structure of the subcomplex A14/43 was resolved (Kuhn et al., 2007, Geiger et al., 2008) followed by the tandem winged helix domain from A49 and A34.5 dimerization module (Geiger et al., 2010). Rrn3 from *S. cerevisiae* was the first crystalized Pol I transcription initiation factor (Blattner et al., 2011). This study showed the first localization of a Pol I specific transcription factor on the Pol I based on lysine-lysine crosslinking mass spectrometry (XL-MS) data. Shortly after, XL-MS showed the basis for rRNA cleavage by the A12.2 subunit of Pol I (Jennebach et al., 2012). The atomic crystal structures of Pol I showed for the first time the details of the transcription machinery as a whole (Engel et al., 2013, Fernandez-Tornero et al., 2013). The structure of the general transcription factor TBP is known as well (Kim et al., 1993). Therefore, three of five (Chapter 1.2) components of the complete initiation complex are known, but structures of UAF and CF are still missing. The structure of CF was the next step in the line of structural investigation in the Pol I initiation system, and would complete the basal initiation system.

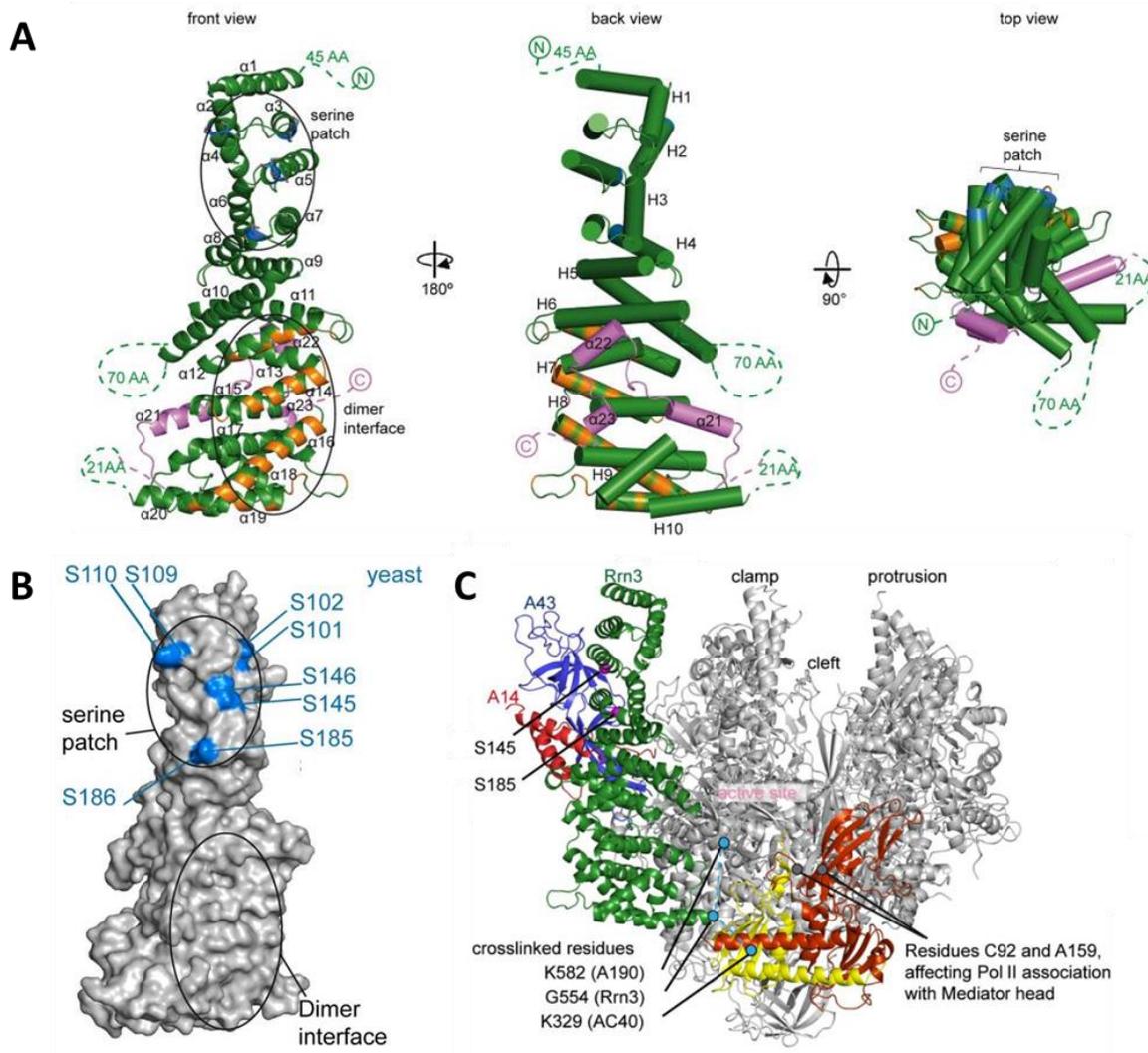
The first steps towards determining the architecture of CF were achieved by recombinant expression (Bedwell et al., 2012) and analysis of XL-MS data of the complex (Knutson et al., 2014). Furthermore, cryo-EM was recently used to determine the structures of Pol I in complex with the transcription factor Rrn3 at medium resolution (Engel et al., 2016, PilsI et al., 2016). This was directly followed by the determination of the Pol I elongation complex representing the first high resolution cryo-EM studies in the Pol I initiation system (Neyer et al., 2016, Tafur et al., 2016). The increasing number of completed structural studies over the last decade indicates the strong need of structural information on the Pol I initiation system, especially of CF and the Pol I PIC.

### 1.3.1. Rrn3 structure

Rrn3 is one of the essential initiation factors for Pol I transcription (Bodem et al., 2000, Moorefield et al., 2000, Yamamoto et al., 1996). Its complex formation with Pol I is regulated by a growth factor signaling pathway that links nutrient availability to rRNA production (Drygin, 2010, Grummt and Voit, 2010). Furthermore, signaling cascades through the target of rapamycin (TOR) pathway leads to the phosphorylation and dephosphorylation of Rrn3 and Pol I (Fath et al., 2001, Cavanaugh AH, 2002, Claypool et al., 2004, Panova et al., 2006, Gerber et al., 2008, Hoppe et al., 2009, Drygin, 2010, Clemente-Blanco et al., 2009). Only in its unphosphorylated form Rrn3 can bind Pol I (Mayer et al., 2004, Mayer et al., 2005).

The crystal structure of Rrn3 was solved and first described in 2011. Rrn3 consists of a unique HEAT repeat structure. The 10 HEAT repeats (H1-H10) are formed by 20 antiparallel  $\alpha$ -helices arranged in a superhelical fold. Additional helices pack arranged against the repeats H6-H10 (Blattner et al., 2011) (Figure 2A). Protein-protein interactions are often mediated by HEAT repeats, which are often found in transport proteins (Cingolani et al., 1999, Conti E, 2000). This structural element was not observed yet in any other general transcription factors (Blattner et al., 2011).

The interaction of Pol I subunit A43 with Rrn3 was described before (Milkereit and Tschochner, 1998, Peyroche et al., 2000, Cavanaugh et al., 2008). Rrn3 itself forms a dimer in solution through its dimer interface (Figure 2A, B) but only binds to Pol I in its monomeric form (Blattner et al., 2011) The binding is dependent on serine patch from helices 3-7. The mutation of S145D caused severe slow growth on 5-FOA plates and underlines the important of unphosphorylated Rrn3 S145 for RNA Pol I transcription initiation (Blattner et al., 2011) (Figure 2B). XL-MS further indicated that unphosphorylated Rrn3 can bind to the Pol I on the back side between the subcomplexes A14/43 and AC40/19 (Blattner et al., 2011) (Figure 2C). It was suggested as well that Rrn3 is involved in the recruitment of Pol I to the initiation site by contacting the CF subunit Rrn6 directly (Peyroche et al., 2000).



**Figure 2. Crystal structure of Rrn3 from *S. cerevisiae* and its interaction with Pol I.**

(A) Structure of Rrn3 as ribbon (left) and cylinder representations (middle and right). C-terminal parts are highlighted in pink. The residues marked in orange are involved in the dimerization of Rrn3. The conserved serine patch is represented in blue. The structure itself is illustrated in front, back and top view (left to right). (B) Surface representation of Rrn3 (grey) and position of serine (blue) residues are highlighted. (C) Representation of Pol I – Rrn3 complex modeled on the basis of cross-linking- and protein interaction data. Figure was adapted from (Blattner et al., 2011).

### 1.3.2. RNA Polymerase I structure

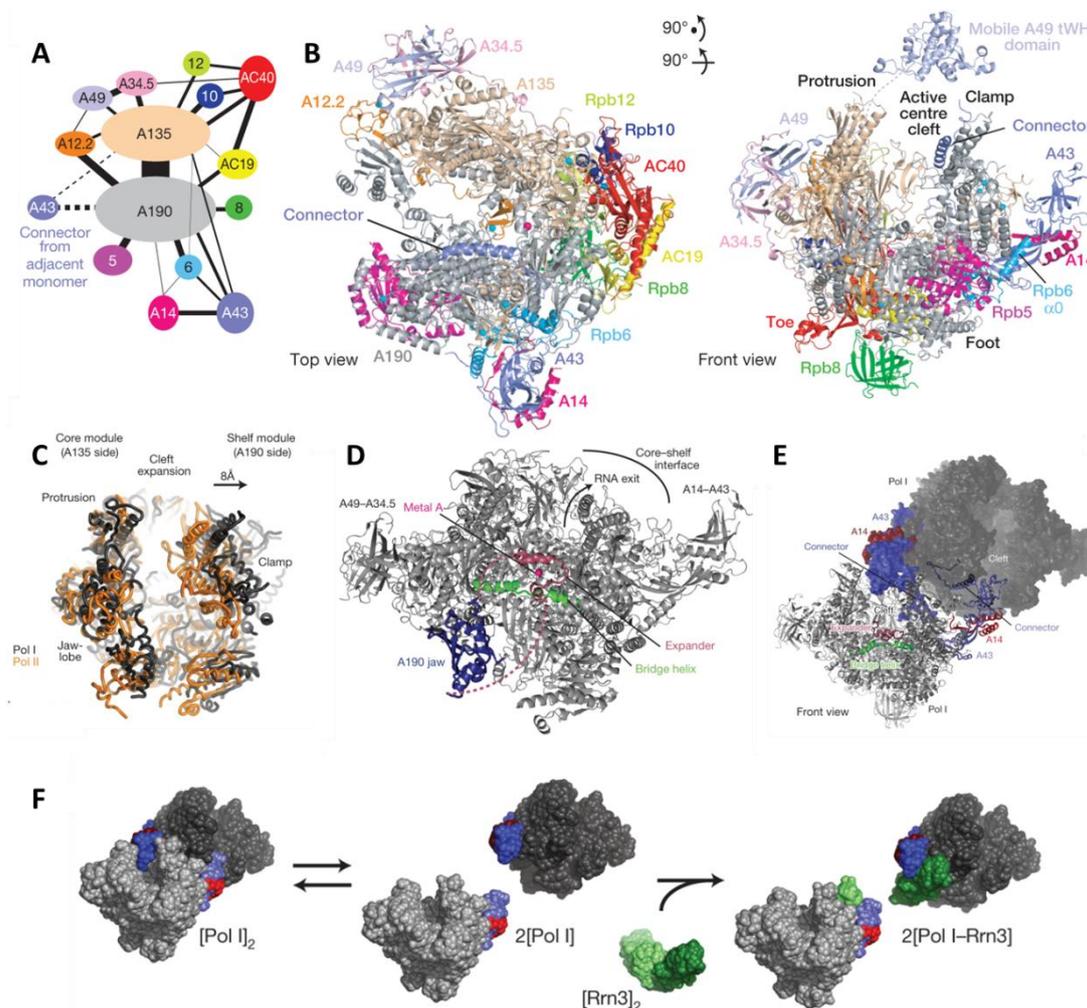
Pol I is important for ribosome biosynthesis. It synthesizes about 60% of the total cellular RNA by transcribing several copies of rRNA genes (Warner, 1999). These genes are transcribed by dozens of Pol I which can be observed as microscopic structures known as ‘Christmas trees’ (Miller, 1969). The 35S precursor rRNA transcript gets further processed to form 18S, 5.8S and 25S rRNAs in the yeast *S. cerevisiae* (Kressler et al., 2010). Pol I is therefore the key element of cell growth and determinant for the level of ribosome components (Laferté et al., 2006). Deregulation of RNA Pol I leads to cancer (Drygin, 2010).

Pol I consists of 14 subunits with a total molecular weight of 589 kDa (Engel et al., 2013). The crystal structure was solved and shows the composition of its 10 subunit core and the two subcomplexes A49/A34.5, which stabilizes A12.2, and A14/A43 which forms the stalk (Engel et al., 2013, Fernandez-Tornero et al., 2013) (Figure 3A, B). Compared to Pol II, the Pol I crystal structure displays an expanded active center cleft (Engel et al., 2013) (Figure 3C). This is likely accomplished by the movement of the two major polymerase modules termed ‘core’ and ‘shelf’ (Cramer et al., 2001) (Figure 3C). This expanded cleft correlates with the presence of an ordered expander element which is formed by an insertion in the A190 jaw domain. Two expander helices inside the active center cleft stabilize the bridge helix which is unwound in its middle (Figure 3D). Pol I was previously described as a dimer (Milkereit et al., 1997, Bischler et al., 2002), is forming a ‘handshake’ structure in its crystal form. This dimer is stabilized by the C-terminal region of A43 which invades the cleft of the adjacent polymerase and forms an alpha helix as well as a beta hairpin close to the other polymerase’s clamp core helices. This element is called the connector (Engel et al., 2013) (Figure 3E).

Pol I transcription initiation is inhibited by its dimerization because it blocks the binding site of Rrn3 (Peyroche et al., 2000, Blattner et al., 2011) (Figure 3F). Pol I dimers are in a concentration-dependent equilibrium with its monomer *in vitro* (Engel et al., 2013) but monomers are also to be expected to be inactive because the expander is blocking the DNA binding site (Engel et al., 2013, Fernandez-Tornero et al., 2013). Just recently, the hypotheses that cleft contraction correlates with transcription activity (Neyer et al., 2016, Tafur et al., 2016) and the notion that dimeric Pol represents an inactive storage form under nutrient starvation (Torreira et al., 2017) were confirmed.

Transcription initiation by Pol I involves the following steps: Firstly, the release of the connector enables Pol I monomerization results in the exposure of the Rrn3 binding site.

Secondly, release of the expander to make DNA binding possible. Thirdly, contraction of the cleft to allow rearrangements of the catalytic elements which subsequently forms a functional catalytic site which finally enables RNA synthesis (Engel et al., 2013).



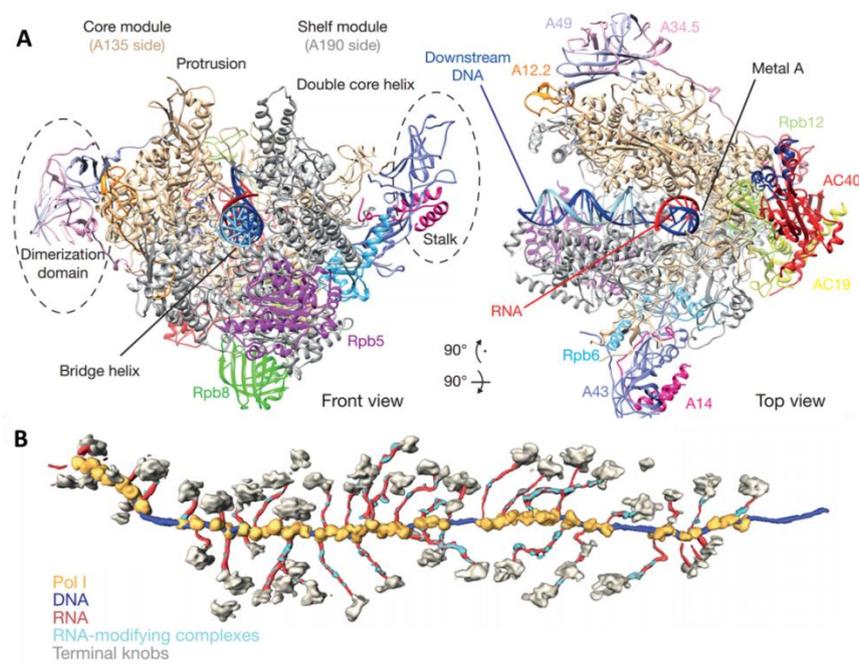
**Figure 3. Pol I crystal structures from *S. cerevisiae***

(A) Pol I subunit interactions and color key that is used in panel B as well. Width of the lines between the subunits indicate the surface area between the subunits: A190, A135, AC40, A14, Rpb5, Rpb6, A43, Rpb8, A12.2, Rpb10, AC19, Rpb12, A49, A34.5 and the A43 connector from the neighboring Pol I in the dimeric form (Engel et al., 2013). (B) Top view and front view (Cramer et al., 2001) from the Pol I structure in ribbon representation. The moveable tandem winged helix (tWH) from A49 (Geiger et al., 2010) was positioned by using crosslinking information (Jennebach et al., 2012). (C) Superimposition of Pol I (black) on Pol II (orange) to compare cleft expansion. (D) Expander element (raspberry), A190 jaw domain (blue) and unwound bridge helix (green). (E) Pol I in its ‘handshake’ dimeric form displayed as a cartoon and surface model. This dimeric form is stabilized by the A14/A43 subcomplex that invades the neighboring polymerase with the connector (blue). (F) Model for Pol I (silver, A14 (red) and A43 (blue)) activation by Rrn3 (green). Figure was adapted from (Engel et al., 2013).

### 1.3.3. RNA-Polymerase I elongation complex

The Pol I crystal structure was solved recently (Engel et al., 2013, Fernandez-Tornero et al., 2013). Therefore, the next step was to explore the conformation of Pol I in its elongation state. This structure was solved by cryo-EM (Neyer et al., 2016, Tafur et al., 2016) (Figure 4A). There the contraction of the cleft of elongating Pol I was observed (Tafur et al., 2016, Neyer et al., 2016) as was suggested before through the movement of the polymerase modules (Engel et al., 2013) ‘shelf’ and ‘core’ (Cramer et al., 2001). Furthermore this structure confirmed the role of the Pol I subunit A12.2 in RNA cleavage (Tafur et al., 2016, Neyer et al., 2016) that was predicted from functional experiments (Ruan et al., 2011) and XL-MS (Jennebach et al., 2012).

Several decades ago Pol I transcription was observed by electron microscopy. This study showed the ‘Miller tree’ for the first time (Miller, 1969), indicating the high loading rate of Pol I molecules actively transcribing a rDNA gene. Cryo-EM tomography confirmed this ‘Miller tree’ and extended this model in the third dimension. In this structure the detailed array of Pol I’s along rRNA was observed as well as RNA-modifying complexes attached to the initially transcribed RNA, confirming cleft contraction during active transcription ‘*ex vivo*’ (Neyer et al., 2016) (Figure 4B).

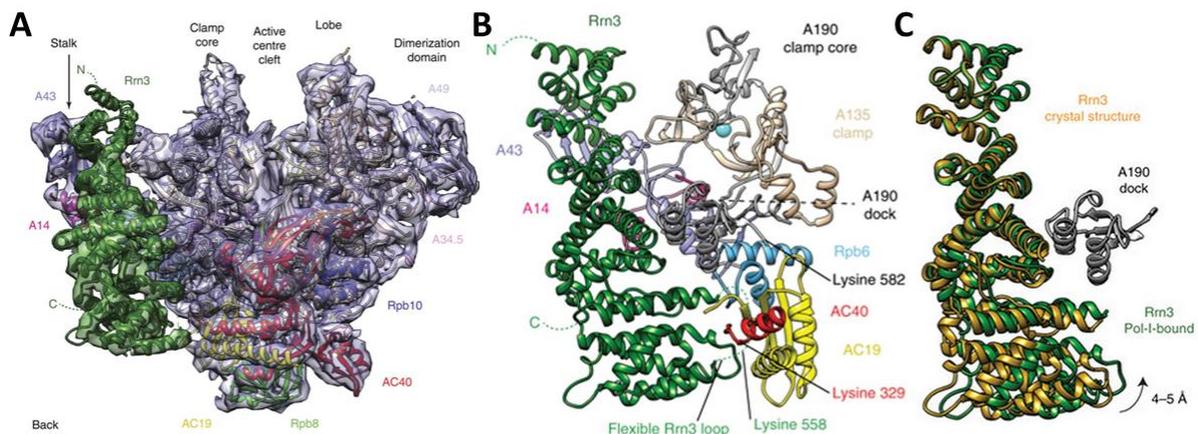


**Figure 4. Cryo-EM structure of the Pol I elongation complex of *S. cerevisiae***

(A) This figure shows the Pol I elongation complex as a ribbon model. (B) Three-dimensional surface rendering of a ‘Miller tree’. Figure was adapted from (Neyer et al., 2016)

### 1.3.4. Polymerase I – Rrn3 complex

High resolution crystal structures were already obtained from the *S. cerevisiae* Pol I (Engel et al., 2013, Fernandez-Tornero et al., 2013) and *S. cerevisiae* Rrn3 (Blattner et al., 2011). The approximate binding site of Rrn3 was located by the interpretation of crosslinking data (Blattner et al., 2011). Cryo-EM structures at medium resolution confirmed this position of Rrn3 on the Pol I and revealed the details of the interaction (PilsI et al., 2016, Engel et al., 2016) (Figure 5A). They show that Rrn3 binds Pol I on the backside of the A14-A43 stalk. The interface between Rrn3 and Pol I involves the Pol I subunits A43, A190, Rpb6, A135, AC40, and AC19 (Figure 5B). Comparison of the Rrn3 crystal structure with the cryo-EM Pol I-Rrn3 structure shows a slight shift of the C-terminal Rrn3 HEAT repeats H6-H10 towards Pol I upon binding (Engel et al., 2016) (Figure 5B, C). The expanded center of the Pol I is partially contracted Rrn3 binding and leads to formation of an initiation competent Pol I-Rrn3 complex (PilsI et al., 2016, Engel et al., 2016) (Figure 5A).



**Figure 5. Structure of the Pol I-Rrn3 complex from *S. cerevisiae*.**

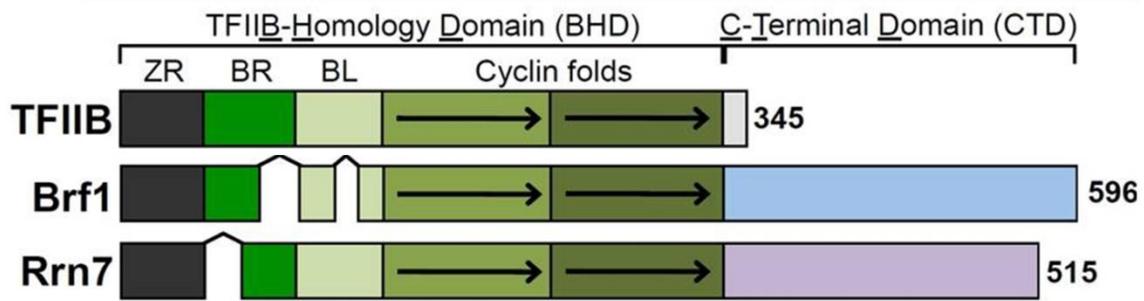
(A) Cryo-EM electron density (grey) with the fitted crystal structure of Pol I as a ribbon model with coloring as described before (Engel et al., 2013) and the fitted crystal structure of Rrn3 (green) (Blattner et al., 2011). (B) Figure shows in cartoon representation of Pol I subunits which interact with the bound Rrn3 (green) (C) The cartoon representation shows the shift of the C-terminal part from the Rrn3 crystal structure (orange) towards the Pol I in its Rrn3-Pol I complex bound form (green). Figure was adapted from (Engel et al., 2016).

### 1.3.5. Core Factor studies

The Core Factor is an essential transcription initiation factor which is a hetero trimeric complex containing the subunits Rrn6, Rrn7 and Rrn11. It forms a bridge between the UAF and Pol I. CF interacts directly with the rDNA core promoter region around the TSS (-38 to +5) and with Rrn3, as well as Rrn9 from UAF (Lalo et al., 1996, Lin et al., 1996, Steffan et al., 1996). It was reported that the C-terminus of Rrn6 contacts Rrn3 (Peyroche et al., 2000). CF is therefore the central element in Pol I transcription. *In vitro* studies indicate that Pol I transcription can start from a core promoter scaffold in presence of only Rrn3 and CF without the need for TBP and UAF (Keener et al., 1998).

The CF subunit Rrn7 shares structural similarities with transcription factors from the Pol II and Pol III systems. Rrn7 is related to TFIIB (Pol II system) and Brf1/Brf2 (Pol III system) (Knutson and Hahn, 2013, Naidu et al., 2011, Gouge et al., 2015) (Figure 6). TFIIB is crucial for Pol II initiation as it bridges Pol II and the promoter DNA bound by TBP and thus sets the topology of the Pol II initiation complex (Kostrewa et al., 2009). TFIIB binds the dock domain of Pol II through its zinc ribbon (Bushnell et al., 2004, Chen and Hahn, 2004). This is linked via the B-reader (BR) and B-linker (BL) elements that have a role in TSS selection and promoter opening respectively, to the Cyclin domains (Kostrewa et al., 2009, Sainsbury et al., 2013). The two cyclin domains are located on the wall of Pol II and help to position the promoter DNA above the cleft (Chen and Hahn, 2004, Kostrewa et al., 2009, Plaschka et al., 2015).

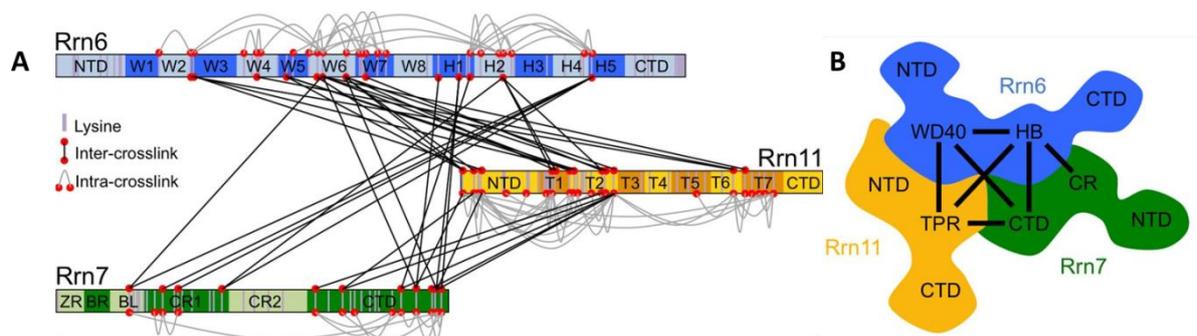
The CF subunits Rrn6 and Rrn11 do not show any obvious homologies to GTFs from Pol II and Pol III (Knutson and Hahn, 2011) (Figure 6). Furthermore, CF subunits are also conserved among eukaryotes. The human ortholog of CF is the selectivity factor (SL) 1 (Comai et al., 1992, Learned et al., 1985) which consists of TBP associated factors (TAFs) TAFIA, -B, and -C which are homologs to Rrn11, Rrn7 and Rrn6, respectively (Russell and Zomerdijk, 2006) and furthermore two additional subunits TAF12 and TAF1D (Gorski et al., 2007, Denissov et al., 2007).



**Figure 6. Domain homologies of initiation factors from Pol I, II, III**

Schematic display of the domains in TFIIB, Brf1, Rrn7 from the transcription machineries Pol II, Pol III and Pol I respectively and the homology between them. Figure was adapted from (Knutson and Hahn, 2011).

Important structural information of CF was lacking for a long time. The first structural study of CF combining XL-MS with domain prediction using HHpred gave a first glimpse of its architecture (Knutson et al., 2014, Soding et al., 2005). CF domain-architecture was predicted as follows: Rrn6 is split up into an N-terminal domain (NTD), a WD40  $\beta$ -propeller, a helical bundle, and a C-terminal domain (CTD). Rrn7 consists of the TFIIB-homologous N-terminal domain (NTD) (B-ribbon, B-reader, and B-linker), two cyclin domains and a C-terminal domain (CTD). Rrn11 is split up into a NTD, a tetratricopeptide repeat (TPR) domain and a NTD. The analyzed crosslinks indicated a possible arrangement of CF subunits (Knutson et al., 2014) (Figure 7).



**Figure 7. Overview of CF crosslinks.**

(A) Overview of the different lysine-lysine crosslinks with inter-crosslinks marked with black lines and intra-crosslinks with grey lines. Rrn7: zinc ribbon (ZR), B-reader (BR), B-linker (BL), cyclin repeat (CR), Tetratricopeptide repeat (T), WD40 (W). (B) Schematic of the interaction partners of the different subunits of CF. Figure was adapted from (Knutson et al., 2014)

## 1.4. Aims and scope of this study

Basal Pol I transcription initiation *in vitro* requires Pol I, Rrn3 and CF (Bedwell et al., 2012, Keener et al., 1998, Pilsl et al., 2016). The atomic structures of Pol I (Engel et al., 2013, Fernandez-Tornero et al., 2013) and Rrn3 are solved (Blattner et al., 2011), only CF is unknown. The atomic structure of CF is required to understand the structural basis of basal transcription initiation in Pol I.

To address the question of the CF structure, I purified, crystallized and determined the structure of the complete heterotrimeric CF from *S. cerevisiae* using X-ray crystallography. I collected native CF data, and for experimental phasing a selenomethionine single-wavelength anomalous dispersion (SAD) peak dataset. The initial experimental map was phase extended using native CF data. Subsequently, I collected sulfur SAD data, and SAD data from selenomethionine labeled protein with additional sequence markers to support structure building and validation (See author contributions in the chapter Publication). ). I determined the CF structure to a resolution of 3.2 Å providing first atomic insight into the essential initiation factor.

In a collaborative effort with Christoph Engel (CEn) and Simon Neyer (SN) the architecture of Pol I initiation complexes was studied. CEn carried out the cryo-EM analysis on the PIC consisting of Pol I-Rrn3-CF. SN carried out the cryo-EM analysis of the initially transcribing complex (ITC) which consists of Pol I-Rrn3-CF and an open bubble RNA-DNA scaffold (See author contributions in the chapter Publication).

The CF structure combined with the cryo-EM from the PIC and the ITC were analyzed by CEn, SN and me. We report how Rrn7 differs from its Pol II counterpart TFIIB and how CF interacts with Pol I by its characteristic polymerase interacting regions (PIR). Furthermore, we show the structure of the PIC and ITC of Pol I for the first time that answers the question of why CF and Rrn3 are essential. We show as well where the promoter DNA enters the Pol I and which parts of CF are interacting with the promoter.

## 2. Materials and Methods

### 2.1. Materials

#### 2.1.1. Bacterial strains

**Table 2.** *E. coli* strains used in this study

Host strain	Genotype	Source or reference
BL21-CodonPlus(DE3)RIL	<i>E. coli</i> B F <sup>-</sup> ompT hsdS(rB <sup>-</sup> mB <sup>-</sup> ) dcm <sup>+</sup> Tetr gal λ(DE3) endA Hte [argU ileYleuW Camr]	Stratagene
XL1-Blue	<i>recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac</i> [F' proAB lacIqZΔM15 Tn10 (Tetr)]	Stratagene

#### 2.1.2. Media, additives and plates for cell growth

**Table 3.** Media and plates

Name	Description and source
Lysogeny broth (LB)	1% (w/v) Tryptone , 0.5 % (w/v) yeast extract, 8.6 mM NaCl
LB plates	1% (w/v) Tryptone , 0.5 % (w/v) Yeast extract, 8.6 mM NaCl, 1.5 % (w/v) agar, 1x Antibiotic

#### 2.1.3. Buffers

**Table 4.** General buffers used in this study

Name	Description
PBS	2 mM Potassium dihydrogen phosphate pH 7.4, 4 mM Disodium hydrogen phosphate, 140 mM Sodium chloride,
MES running buffer	50 mM MES pH 7.3, 50 mM Tris Base, 0.1 % SDS, 1 mM EDTA
MOPS running buffer	50 mM MOPS pH 7.7, 50 mM Tris Base, 0.1% SDS, 1 mM EDTA,
TFB-I	30 mM Potassium acetate pH 5.8, 50 mM MnCl <sub>2</sub> , 100 mM RbCl, 10 mM CaCl <sub>2</sub> , 15% (v/v) Glycerol,
TFB-II	10 mM MOPS pH 7.0, 75 mM CaCl <sub>2</sub> , 10 mM RbCl, 15% (v/v) Glycerol,

**Table 5.** Buffers of the Pre-Crystallization Test

<b>Name</b>	<b>Description</b>
Reagent A1	0.1 M TRIS hydrochloride pH 8.5, 2.0 M Ammonium sulfate
Reagent B1	0.1 M TRIS hydrochloride pH 8.5, 1.0 M Ammonium sulfate
Reagent A2	0.1 M TRIS hydrochloride pH 8.5, 0.2 M Magnesium chloride hexahydrate, 30 % w/v Polyethylene glycol 4,000
Reagent B2	0.1 M TRIS hydrochloride pH 8.5, 0.2 M Magnesium chloride hexahydrate, 15 % w/v Polyethylene glycol 4,000

**Table 6.** Solutions for silver staining

<b>Name</b>	<b>Description</b>
Fixing solution	0.7 % HAc, 40 % Ethanol
Washing solution 1	50 % Ethanol
Washing solution 2	5 % Ethanol
Reducing solution	35 $\mu$ M DTT
Silver solution	1 mg/ml Silver nitrate, 0.001 $\mu$ M Formaldehyde
Developing solution	0.283 mM Sodium carbonate, 0.006 $\mu$ M Formaldehyde

#### 2.1.4. Markers, solution and enzymes

**Table 7.** General markers, solutions and enzymes

<b>Name</b>	<b>Description and source</b>
6x DNA loading dye	Fermentas
Gene Ruler™ 1kb DNA Ladder	Fermantas
SYBER Safe DNA gel stain, 10,000 x in DMSO	Invitrogen
100x Protease inhibitor	60 $\mu$ M Leupeptin, 200 $\mu$ M Pepstation A, 98 mM Phenylmethylsulfonylflourid (PMSF), 211 mM Benzamidinhydrochlorid, in EtOH
PageRuler Prestained Protein Ladder	Fermentas
5x Loading dye	25 % Glycerol, 7.5% (w/v) Sodiumdodecylsulfate (SDS), 250 mM Tris-HCl pH 6.8, 0.5 % Bromphenol blue, 12.5% $\beta$ -Mercatoethanol

**Table 8.** Antibiotic stock solution

Name	Description
Ampicillin 1000x	100 mg ml <sup>-1</sup> in Water
Kanamycin 1000x	30 mg ml <sup>-1</sup> in Water
Tetracyclin 1000x	12.5 mg ml <sup>-1</sup> in EtOH
Chloramphenicol 1000x	30 mg ml <sup>-1</sup> in EtOH
Streptomycin 1000x	20 mg ml <sup>-1</sup> in Water

## 2.2. General Methods

### 2.2.1. Synthesis of methionine point mutations for structure validation

The QuickChange polymerase chain reaction (PCR) method was used to introduce methionine point mutations into the expression plasmids of Rrn11, Rrn6 and Rrn7 in order to identify missing secondary structure elements in the CF structure (Chapter 3.1). For this purpose primers were designed (see Chapter 2.2.2) and the mutations introduced by PCR (See Chapter 2.2.2). After the parental template plasmid was digested with DpnI (Chapter 2.2.4) the DNA was purified (Chapter 2.2.5) and transformed into XL-1 cells. Positive clones were selected and verified as described (Chapter 2.2.7).

### 2.2.2. Primer design for quick change polymerase chain reaction.

Primers for the quick change PCR were designed with consideration to the following criteria. Firstly, there were 10-15 nucleotides 3' and 5' from the point of mutation. Secondly, the 3' and 5' end of the primer was a Guanine or a Cytosine. Thirdly, the primer melting temperature had to be between 70 °C and 85 °C. The Primers were ordered from Sigma-Aldrich Chemie GmbH.

### 2.2.3. Quick change Polymerase chain reaction

Plasmids (Chapter 5.3) were modified with the following quick change PCR protocol. The total volume of the PCR reaction was 25 µl and contained 0.25 µl Phusion<sup>®</sup> High-Fidelity DNA Polymerase (New England Biolabs), 1x Phusion HF Buffer, 200 µM dNTP's, 250 nM forward and reverse primers and 0.5 µl of plasmid and +- 5%-10% DMSO (Table 9).

**Table 9.** PCR protocol

Time	Temperature	Benefit of the temperature
1 min	98 °C	First denaturation step
15 sec	98 °C	DNA melting
1 min	55-70 °C <sup>A</sup>	Primer annealing
X <sup>B</sup> min	72 °C	Elongation
10 min	72 °C	Final elongation
∞	4 °C	Storage



A) Primer annealing was adjusted to the primer melting temperature.

B) Elongation time was calculated by the size of the Plasmid in kb times 1 kb/min.

#### **2.2.4. DpnI digest**

A total volume of 1 µl of DpnI (New England Biolabs) was added to the 25 µl PCR reaction and incubated for 3h at 37 °C.

#### **2.2.5. Agarose gel plasmid purification, isolation**

PCR products were purified by Agarose gel electrophoresis. The gels consisted of 1% (w/v) Agarose (Invitrogen) using a 1x TAE (5mM EDTA pH 8.0, 0.25 M Tris-acetate) buffer and DNA stain (Table 9). The electrophoresis was carried out in 1x TAE buffer for 40 min at 120V. In the next step the gel band was cut out and purified with the QIAquick Gel Extraction Kit from Qiagen.

#### **2.2.6. Competent cells**

Two strains of chemical competent cells were used: *E.coli* XL1-blue for plasmid amplification and the BL21-CodonPlus(DE3)-RIL for protein expression (see Table 2). Cells were made chemically competent by the Rubidium chloride method. A cell culture was inoculated and was grown to a final Optical Density 600 (OD) of 0.4. The cells were chilled on ice, centrifuged in sterile tubes in a (Eppendorf Centrifuge 5810 R) with A-4-81 rotor (Eppendorf) for 10 min at 4,000 rpm and 4°C. The supernatant was replaced by 100 ml of TFB-I buffer (Table 4) and the cells were resuspended. The resuspension was centrifuged as before and resuspended in 5 ml TFB-II (Table 4) at 4 °C. The suspension was aliquoted in 1.5 ml precooled tubes. The 50 µl cell aliquots were flash frozen in liquid nitrogen and stored at -80 °C.

#### **2.2.7. Plasmid transformation and plasmid purification, isolation and verification**

Chemically competent cells (Chapter 2.2.6) were thawed on ice. From a plasmid stock 0.5 µl and from a gel purified plasmid 2.5 µl was added to the freshly thawed cells. The cells were incubated for 14 min on ice then heat shocked for 60 s at 42 °C and chilled on ice for two min. Cells were incubated for 45 min at 37 °C after adding 500 µl of LB (Table 3). The cells were plated on LB plates (Table 3) which contained the appropriate antibiotic/s (Table 8). Plates were incubated overnight at 37 °C.

Colonies were picked from the XL1-Blue plates and used as a starter for a 5 ml overnight culture containing appropriate antibiotics. The cells were incubated overnight at 37 °C and then harvested by centrifugation. Plasmids were isolated using the Qiagen miniprep kit as described in the protocol with the modification of using water for the final elution from the spin column.

The purified plasmid was verified by sequencing by the company Seqlab (<http://www.seqlab.de/>).

### **2.2.8. SDS-PAGE**

Proteins were analyzed using a Novex<sup>®</sup> Nu PAGE<sup>®</sup> SDS-PAGE Gel System with a NuPAGE<sup>®</sup> 4-12 % Bis Tris Gels of 1.0 mm (Table 4). Gel electrophoresis was carried out with either MES or MOPS running buffer at 200 V for 40 min to 60 min. Protein subunit size was estimated by a protein marker (Table 7). The separated proteins were visualized by using the protein staining solution Instant Blue (Expedeon).

The silver staining protocol was applied for low quantities of protein (Chapter 2.2.9).

### **2.2.9. Silver gel staining**

The SDS gel was incubated for 60 min in the Fixing solution then washed with Washing solution 1 for 20 min and Washing solution 2 for 20 min. The SDS gel was incubated for 5 min in the Reducing solution and 10 min in the Silver solution. The ice-cold Developing solution was used to wash the gel twice and after the remainder was incubated with the gel until the desired intensity was reached. The development was stopped by adding solid citric acid monohydrate until it stopped bubbling.

### **2.2.10. Protein concentration measurement and precipitation**

The concentration of protein solutions were determined by using the NanoDrop 2000 (Thermo Fisher Scientific) using the protein specific extinction coefficient and molecular weight. Protein parameters were calculated by using the web interface of Protparam (Gasteiger et al., 2003). Program was used under the URL <http://web.expasy.org/protparam/>.

### **2.2.11. PCT™ Pre-Crystallization Test**

The Pre-Crystallization test was used as a guide for the protein concentration used in the crystallization trials. The sample concentration was adjusted until it produced a light precipitation. The test was used frequently for proteins expressed with the selenomethionine method (Chapter 2.3.1) because of the change in solubility by the introduction of selenomethionine into the protein. The test was carried as described by the company (Hampton Research) together with siliconized glass covers slides and pregreased 24-well plates (Jena Bioscience).

## 2.3. Specific methods

Methods described in section 2.3 were performed as part of this thesis. These methods were published and described in following publication:

Christoph Engel\*, **Tobias Gubbey\***, Simon Neyer\*, Sarah Sainsbury, Christiane Oberthuer, Carlo Baejen, Carrie Bernecky, Patrick Cramer. (2017) Structural Basis of RNA Polymerase I Transcription Initiation. *Cell* Volume 169, Issue 1, Pages 120-131.

\*These authors contributed equally.

<http://dx.doi.org/10.1016/j.cell.2017.03.003>

For details see author contributions at the chapter Publication.

Other methods performed by co-authors from this publication are listed in Extended Material, Methods and Data

### 2.3.1. Expression of recombinant Core Factor

Core Factor (CF) subunits from *S. cerevisiae* were co-expressed in *E. coli* BL21-CodonPlus(DE3)-RIL cells from 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<sub>600</sub> 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 18h 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 CF, a 4 L LB starter culture was grown from a 200 mL starter culture until OD<sub>600</sub> 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 minimal medium depleted for methionine (Molecular Dimensions) to an OD<sub>600</sub> of 0.2 - 0.3. The culture was incubated for 1.5 h at 37°C with antibiotics while shaking. After starvation, cell growth was induced by adding selenomethionine stock solution (Molecular

Dimensions). After reaching an OD<sub>600</sub> of 0.6, the temperature was reduced to 18°C and additional amino acids were added (50 mg/L lysine, threonine, and phenylalanine; 25 mg/L leucine, isoleucine, and valine). Expression was induced by adding IPTG to a concentration of 1 mM. Cells were grown for 18h, harvested by centrifugation, frozen in liquid nitrogen and stored at -80°C.

### **2.3.2. 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<sub>2</sub>, 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<sub>2</sub>, 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<sub>2</sub>, 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<sub>2</sub>, 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<sub>2</sub>, 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 (200 mM NaCl, 1 mM MgCl<sub>2</sub>, 10 % (v/v) glycerol, 20 mM HEPES pH 7.8, 1 mM DTT) and then eluted with a gradient ranging from 0.2 to 2.0 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 NaCl, 1 mM MgCl<sub>2</sub>, 5 % (v/v) glycerol, 10 mM HEPES pH 7.8, 10 µM ZnCl<sub>2</sub>, 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.

### 2.3.3. 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 at a concentration of 6.33 mg/mL. Pre-greased Crystalgen SuperClear 24-well Plates (Jena Bioscience) and siliconized cover slides (Jena Bioscience) were used to set 1  $\mu$ L 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  $\mu$ m. Crystals were harvested after 5-7 days 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.

### 2.3.4. Crystallization of Core Factor

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  $\text{\AA}$  resolution and data was processed with XDS (Kabsch, 2010), showing  $P6_5$  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  $\beta$ -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  $\chi$  angles. Datasets were merged, showing a strong anomalous signal until  $\sim 6$   $\text{\AA}$  resolution (Table S1). 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  $\sim 3.2$   $\text{\AA}$  using the Phenix suite improved the initial map (Fig. S1).

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 highly redundant sulphur-SAD dataset recorded from a native crystal at a wavelength of 2  $\text{\AA}$ . 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 S2) 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, Rrn11-W84 (weak signal) and Rrn11-L430 (Fig. S1). The Rrn11-L73M mutation did not yield a selenium signal, suggesting flexibility of this region. Refinement of the structure in space group  $P6_5$  using phenix.refine, Refmac or BUSTER did not result in free R-factors below 35%, therefore the data was reprocessed in lower symmetry space groups. MR placed six CF molecules in each asymmetric unit in space group P1, showing a super-helical, pseudo-symmetric arrangement slightly deviating from translational and rotational crystallographic symmetry.

Refinement of the structure was carried out in space group P1 using phenix.refine (Chapter 5.4) with Cartesian 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  $R/R_{\text{free}}$  factors of 26.1/28.8% and good stereochemistry (Table S1). 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 regions of Rrn6 are more flexible, resulting in local differences in B-factors and map quality (Fig. S1). Rrn11 helix  $\alpha 2$  was tentatively assigned based on a weak anomalous signal peak obtained from the SeMet-labeled W84M variant.

### 3. Results and Discussion

#### 3.1. Structural basis of RNA polymerase I transcription initiation

Results in Chapter 3 were obtained in collaboration with Christoph Engel and Simon Neyer and later published in:

Christoph Engel\*, **Tobias Gubbey\***, Simon Neyer\*, Sarah Sainsbury, Christiane Oberthuer, Carlo Baejen, Carrie Bernecky, Patrick Cramer. (2017) Structural Basis of RNA Polymerase I Transcription Initiation. *Cell* Volume 169, Issue 1, Pages 120-131.

\*These authors contributed equally.

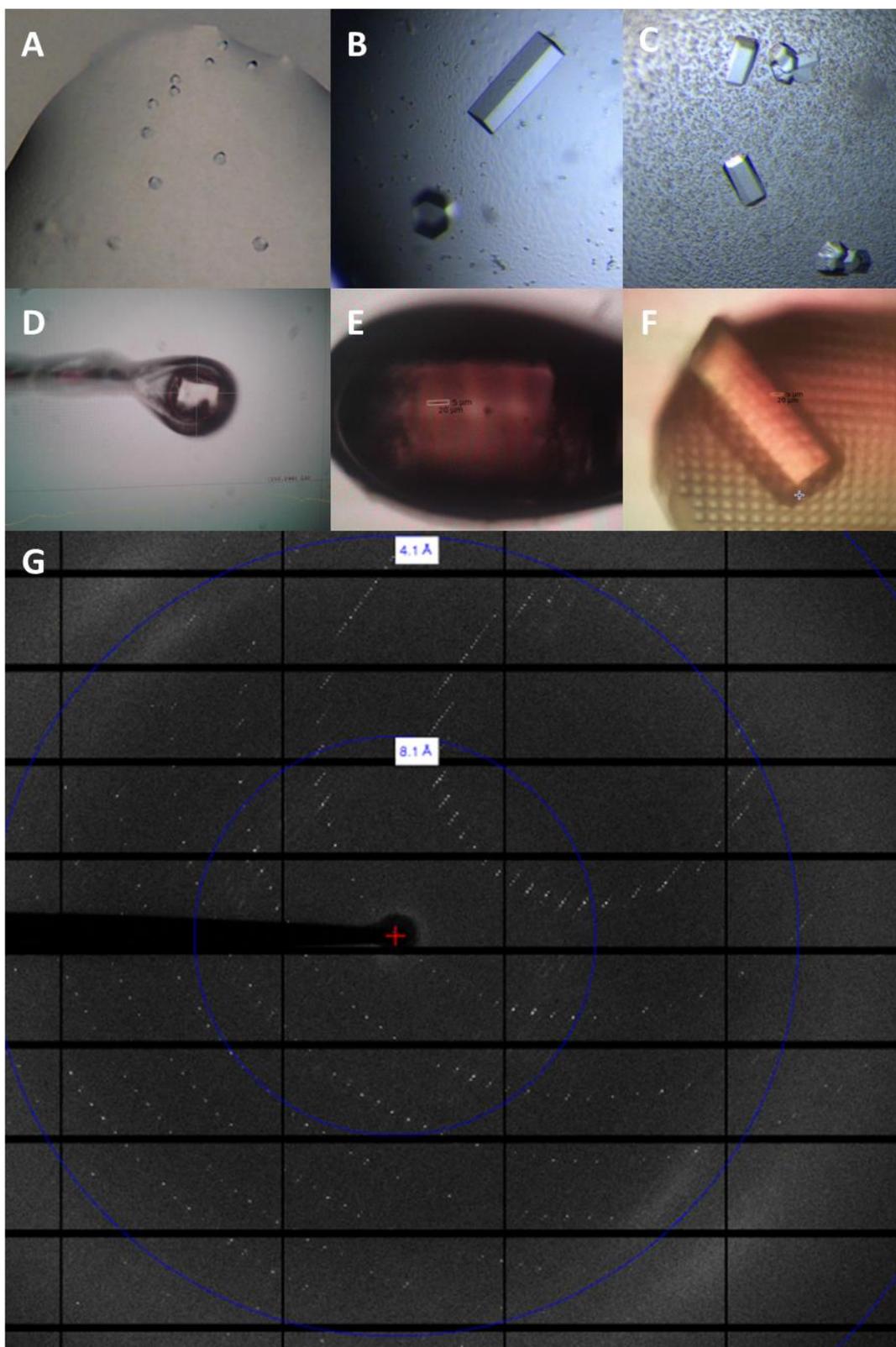
<http://dx.doi.org/10.1016/j.cell.2017.03.003>

For details see author contributions in the chapter Publication.

I focused on the crystallization and structure determination of CF which is described in chapter 3.2-3.4. Supplemental figures and tables from my collaborators are listed in the Appendix.

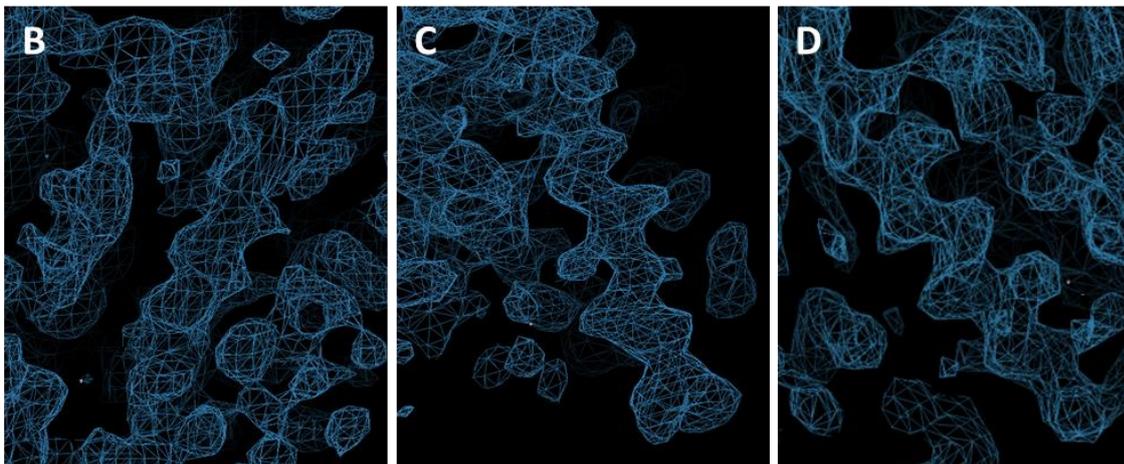
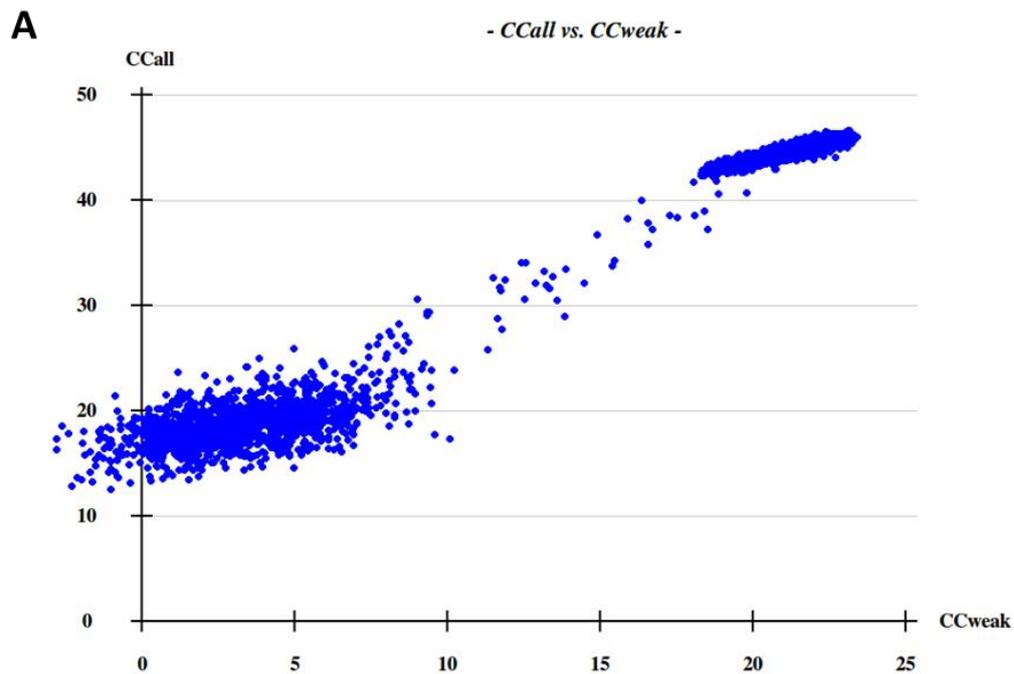
#### 3.2. Crystal structure of Core Factor

We obtained *S. cerevisiae* CF after heterologous co-expression of its three subunits in *E. coli*. For CF purification, we modified previously published protocols (Bedwell et al., 2012, Knutson et al., 2014) and could remove co-purifying chaperones (STAR Methods). Purified CF contained the three subunits in apparently stoichiometric amounts (Figure 11A) and could be crystallized by vapor diffusion. The crystal structure was determined by single-wavelength anomalous diffraction (SAD) using selenomethionine-derivatized crystals in space group P6<sub>5</sub> (Table 10) (Figure 8, Figure 9). 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 sulfate ions with the use of anomalous diffraction from sulfur atoms (Table 11), and introduced six additional methionine markers in regions that were difficult to build (Figure 11, STAR Methods). The structure was refined to R/R<sub>free</sub> factors of 26.1/28.8% with good stereochemistry (Table 10).



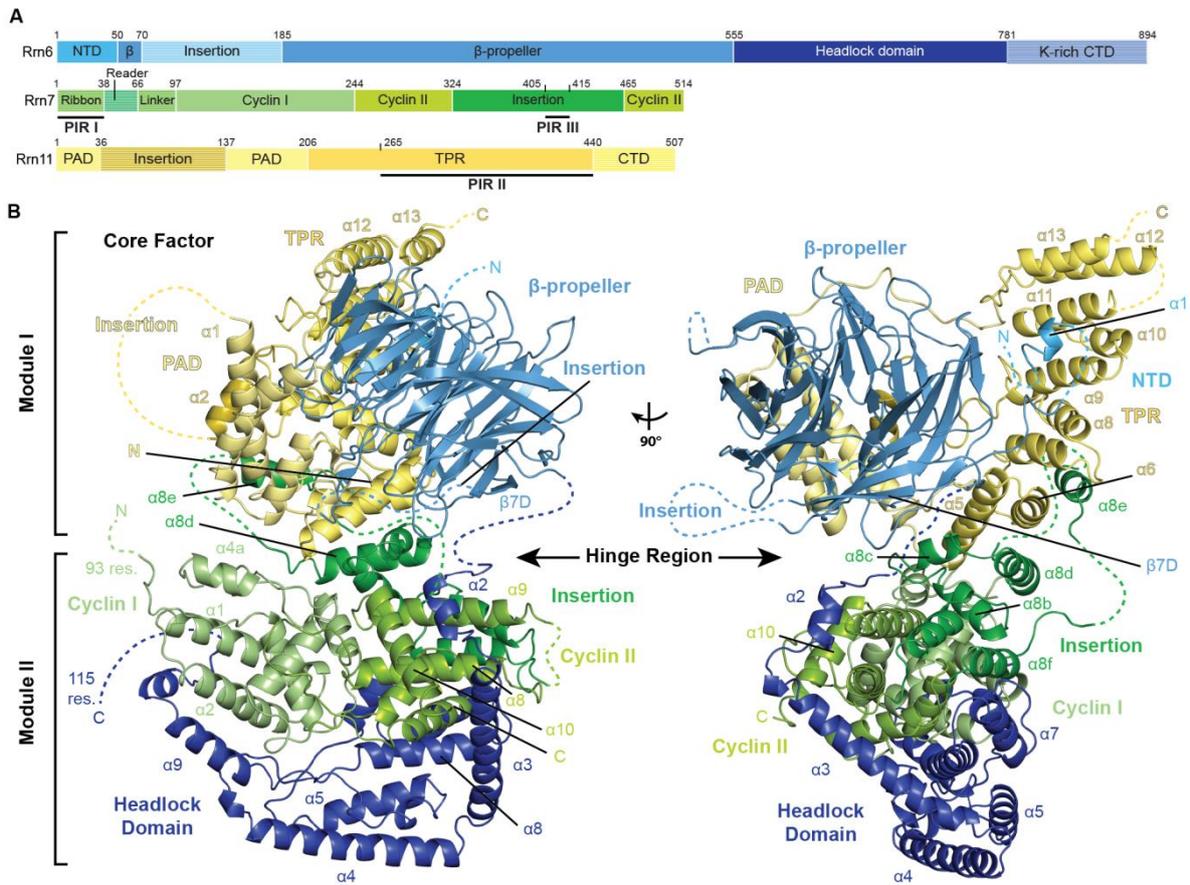
**Figure 8. CF crystallization and data collection**

(A) Initial crystallization hit of full length CF in a 96-well screening plate (B) Representative image of optimized native CF crystals in a PEG4000, ammonium sulfate condition at pH=6.0 (C) Representative image of selenomethionine CF crystals (D) Initial well diffracting native CF crystal (E) Best diffracting native crystal to 3.2 Angstrom processed in P1 (F) Crystal used for data collection at room temperature (G) Representative CF diffraction pattern.



**Figure 9. Experimental phasing of selenomethionine labeled CF**

(A) Locating heavy atoms in a SAD experiment using SHELXD. A CCall above 30% indicates a good solution which should be well separated from non-solutions (Sheldrick, 2010). (B) Representative image of a bad region of an initial experimentally phased map. (C) Representative image of a good region of an initial experimentally phased map. (D) Representative image of a good region of a phase extended initial map (Emsley and Cowtan, 2004).



**Figure 10. Crystal structure of Core Factor**

(A) Domain organization of the yeast CF subunits Rrn6 (blue), Rrn7 (green) and Rrn11 (yellow). Domain boundaries, unstructured regions and polymerase interacting regions (PIRs) are indicated. (B) Architecture of CF. Rrn7 forms two cyclin domains, the second being interrupted by a large insertion. The N-terminal ribbon domain of Rrn7 is mobile and not resolved in the CF crystal structure. The Rrn11 propeller/promoter-associated domain (PAD) contains a flexible insertion and is followed by a TPR fold. For details see Figure 11.

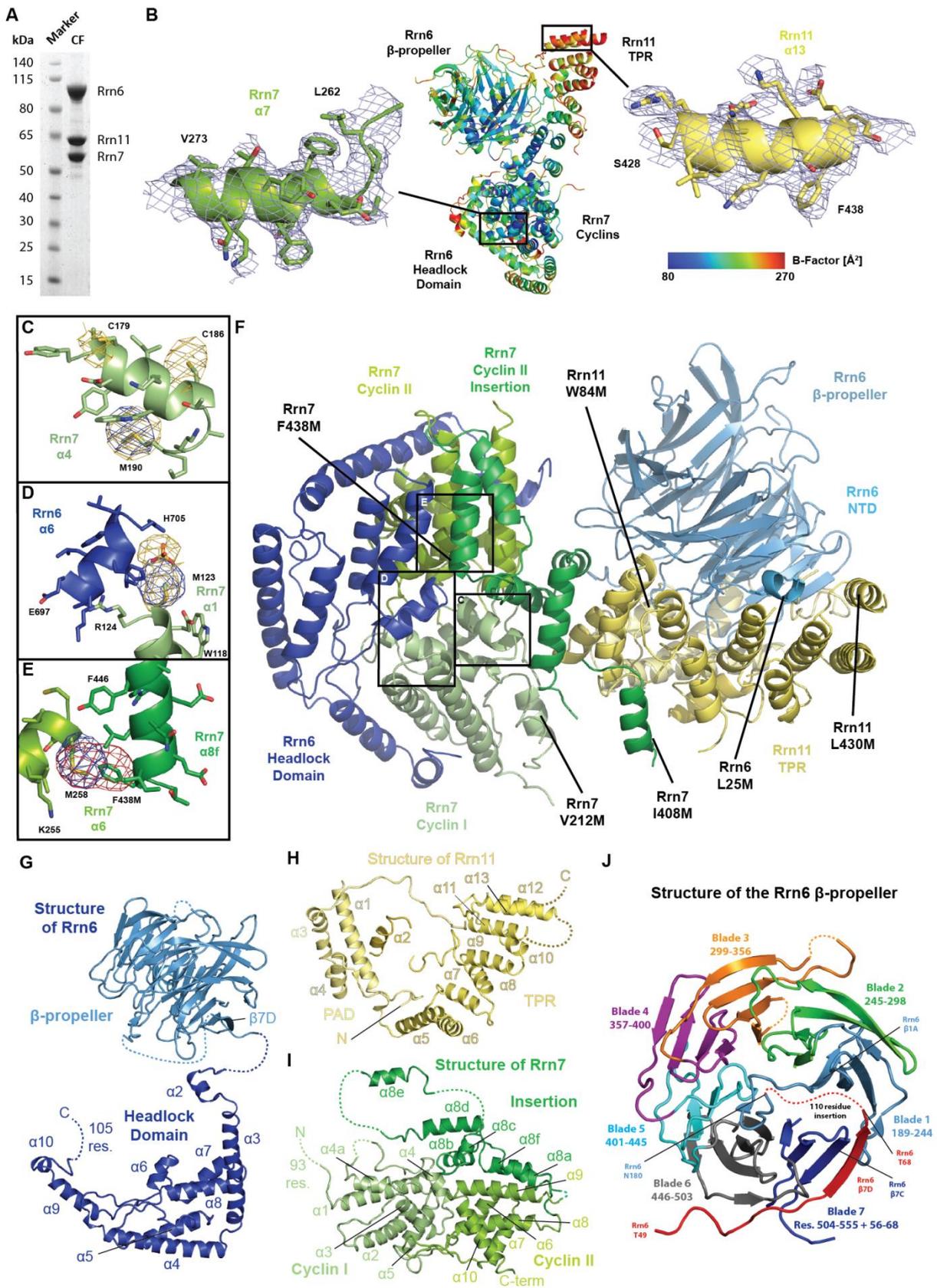


Figure 11. Crystal structure of yeast Core Factor. Figure legend on next page.

### Figure 11. Crystal structure of yeast Core Factor.

(A) Coomassie-stained SDS-PAGE of purified CF reveals that the three subunits Rrn6, Rrn7 and Rrn11 are present in stoichiometric amounts. (B) Cartoon representation of the CF crystal structure colored according to local B-factors (center). Central regions are well ordered (low B-factors, well-defined 2Fo-Fc map for Rrn7 helix  $\alpha 7$  at  $0.9\sigma$  on the left). In contrast, peripheral elements, such as the Rrn11 TPR helix  $\alpha 13$  (right; map at  $0.9\sigma$ ) show high B-factors and are less well-defined. (C) Anomalous difference density peaks for sulphur ( $2.6\sigma$ , yellow) and selenium ( $5.0\sigma$ , blue) confirm the residue placement in Rrn7 helix  $\alpha 4$ . (D) Rrn6 helix  $\alpha 6$  (blue) and the Rrn7 helix  $\alpha 1$  (green) with anomalous difference density peaks for selenium ( $5.0\sigma$ , blue) and a sulphate ion ( $2.6\sigma$ , yellow) that is coordinated in proximity to Rrn6 H705. (E) Rrn7 helices  $\alpha 6$  and  $\alpha 8f$  with the anomalous difference density for selenium in native ( $5.0\sigma$ , blue) and F438M mutant crystals ( $5.0\sigma$ , red). A total of six methionine mutations validated the residue register for most of the structure. (F) Cartoon representation of CF indicating the location of mutations. (G) The structure of Rrn6 (blue) spans a large distance from  $\beta$ -propeller to headlock domains. A part of the N-terminal region of Rrn6 (20-27) is ordered and was assigned by a L25M mutation and anomalous sulphur signal of C27. This region interacts with Rrn11 TPR helices  $\alpha 9$ ,  $\alpha 11$  and  $\alpha 13$  (Figure 15C). (H) PAD and TRP domains of Rrn11 form compact assemblies around the Rrn6  $\beta$ -propeller. The TPR domain consists of nine helices, which is uncommon (Blatch and Lasse, 1999). The N-terminal methionine 1 of Rrn11 is embedded into Rrn6 (Figure 10B) and shows an anomalous signal peak. Next to the amino terminus of Met1 in Rrn11, two sulphate ions are bound inside a positively charged pocket. (I) The structure of Rrn7 forms a compact arrangement, from which insertion helix  $\alpha 8e$  (PIR III) and the 93 N-terminal residues (disordered) protrude. Helix  $\alpha 8d$  contacts the first tandem repeat of Rrn11 helices  $\alpha 5/\alpha 6$ , contributing to the association of Rrn11 with Rrn7. (J) Architecture of the 7-bladed Rrn6  $\beta$ -propeller with each blade depicted in a different color. A 110-residue insertion is indicated.

### 3.3. Core Factor comprises two modules

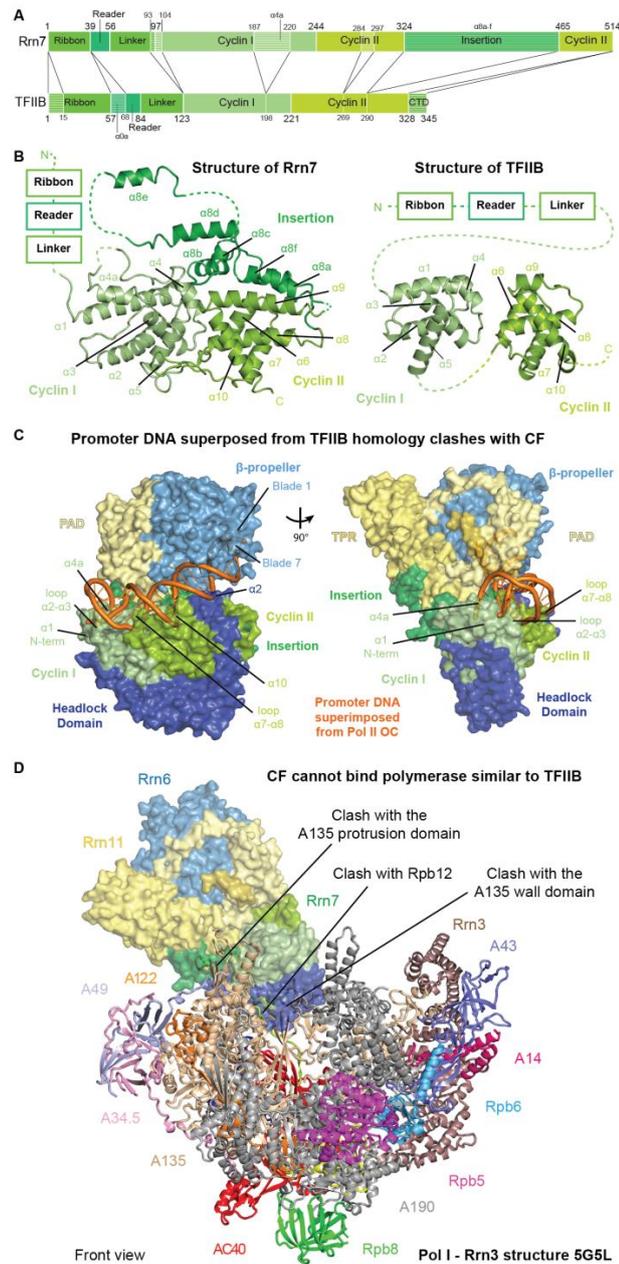
The crystal structure reveals that CF comprises two modules that are flexibly linked and separated by a central hinge region (Figure 10). Module I is formed by an N-terminal seven-bladed  $\beta$ -propeller domain in Rrn6 and by subunit Rrn11, which forms a superhelical TPR domain and a novel helical 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 ( $\alpha$ 8a-f) and reaches over to module I (Figure 10B, Figure 11). The structure lacks only the mobile N- and C-terminal regions of Rrn6, an insertion in the  $\beta$ -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 (Figure 13, Figure 14). The CF structure rationalizes effects of known mutations. For example, the Rrn11 N-terminus is buried in the Rrn6  $\beta$ -propeller, explaining why its deletion is lethal (Knutson et al., 2014).

### 3.4. Rrn7 differs from TFIIB

A search for structural similarity to Rrn7 using PDBeFold (Krissinel and Henrick, 2004) identified TFIIB and its archaeal homologue TFB, and led to a structure-based alignment of Rrn7 with TFIIB (Figure 12A, Figure 12B, Figure 13A). 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.

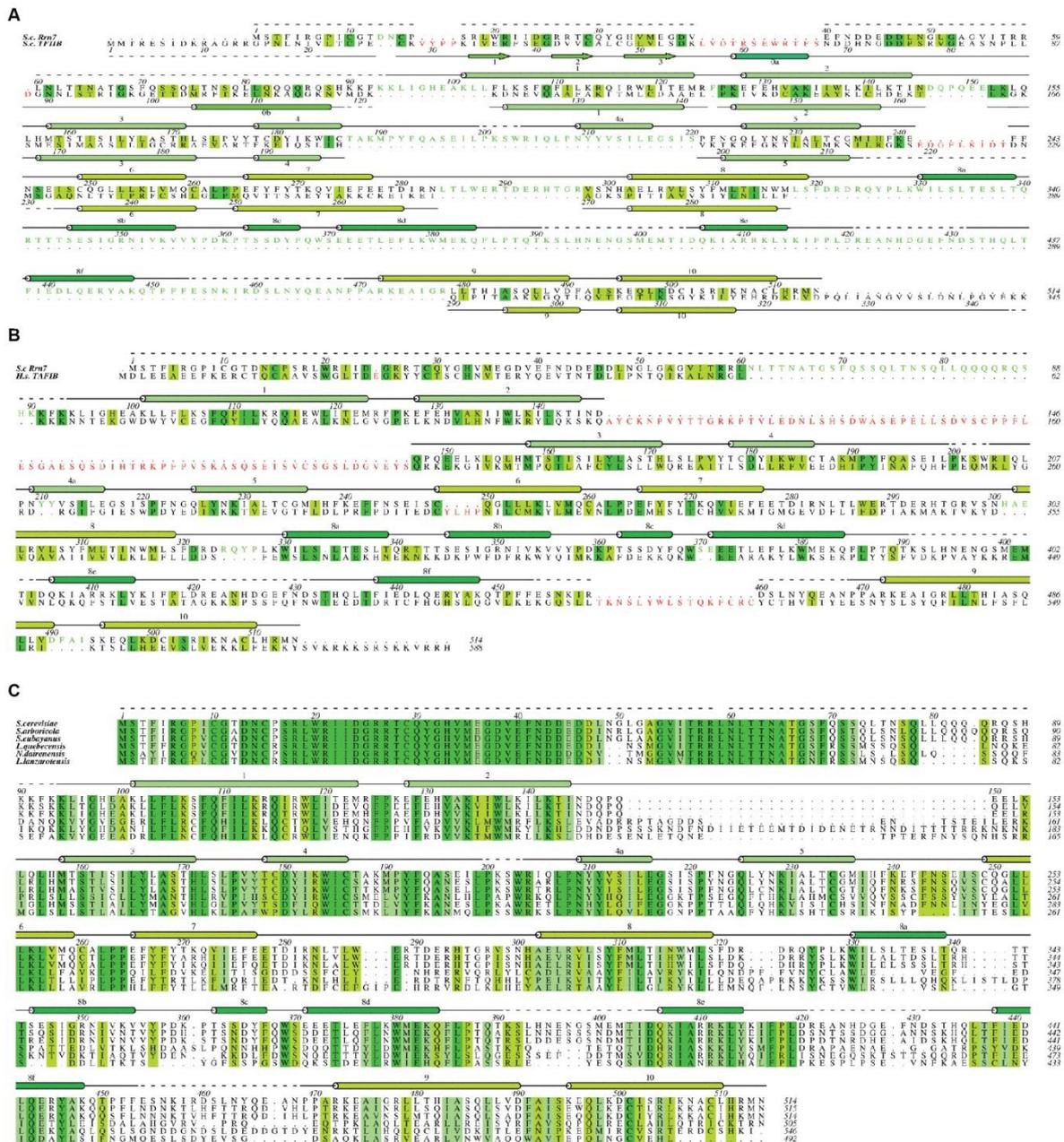
Modelling of CF onto DNA based on a superposition of Rrn7 with TFIIB in the Pol II initiation complex structure (Plaschka et al., 2016) shows that Rrn7 cannot bind to DNA in the way TFIIB does (Figure 12C). Major clashes are observed between modelled DNA and Rrn7 cyclin domains ( $\alpha$ 1,  $\alpha$ 4a,  $\alpha$ 10, and loops  $\alpha$ 2- $\alpha$ 3 and  $\alpha$ 7- $\alpha$ 8), the Rrn6  $\beta$ -propeller and the headlock domain helix  $\alpha$ 2. 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 (Figure 12D). 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.



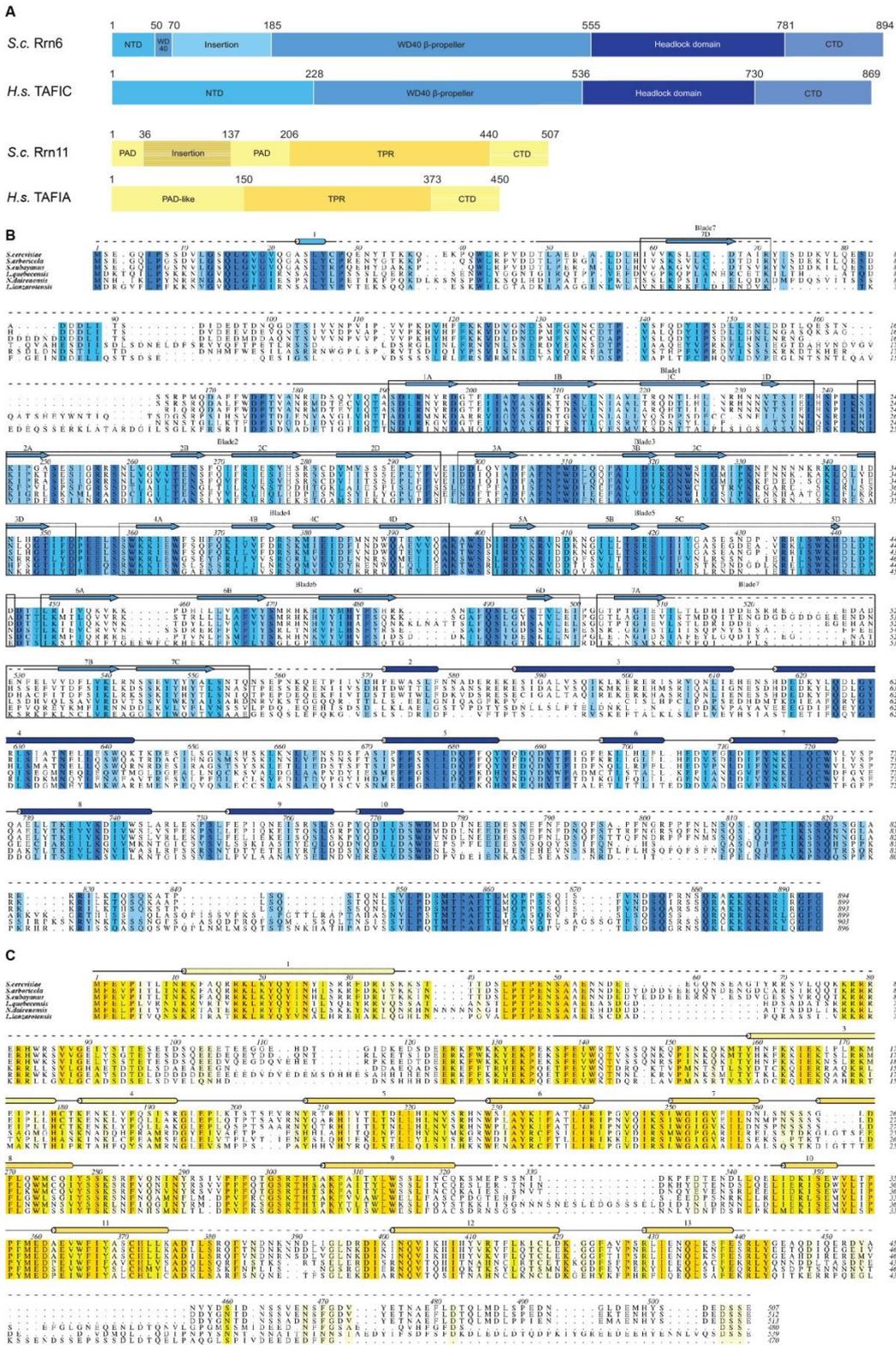
**Figure 12. Rrn7 structurally differs from TFIIB**

(A) Domain organization of Rrn7 and TFIIB depicting insertions and deletions of more than ten residues. (B) Side-by-side structural comparison of yeast Rrn7 (left) and yeast TFIIB (right, PDB 5FYW) reveals differences in cyclin II orientation. A cyclin I insertion (helix  $\alpha 4a$ ), and a large insertion in cyclin II (dark green) are visible in Rrn7, helix  $\alpha 9$  is elongated, and flexible loops change their positions. (C) Superposition of Rrn7 cyclin domains with the promoter-bound structure of TFIIB reveals clashes. Promoter DNA clashes with parts of both cyclin domains and Rrn6. TFIIB and TBP are omitted for clarity. (D) Modelling CF (space filling) onto the Pol I-Rrn3 structure 5G5L (cartoon) reveals major clashes. The Pol II-TFIIB structure (PDB 4BBS) was superimposed via Rpb2-A135 and CF positioned by superimposition of Rrn7 cyclin I with its TFIIB counterpart.



**Figure 13. 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 (Figure 12 A/B). (B) Sequence and secondary structure prediction-based alignment of yeast and human Rrn7 (TAF1B). In human Rrn7, the reader/linker region is truncated and an insertion between cyclin I helices  $\alpha 2$  and  $\alpha 3$  is predicted, likely located close to  $\alpha 4a$ , in proximity to promoter DNA. (C) Multiple sequence alignment of Rrn7 from different yeast species shows that the protein is highly conserved.



**Figure 14. Secondary structure-based alignments of Rrn6 and Rrn11 orthologs.**

(A) Domain architecture of Rrn6 and Rrn11 is conserved between yeast (*S.c.*) and human (*H.s.*). (B) 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.

**Table 10. Crystallographic data collection, phasing and refinement statistics for anomalous and native datasets collected from CF crystals.**

	Native 1	Native 2 (PDB 5N5X)	SeMet 1
<b>Data collection</b>	PX1 Pilatus M6	PX1 EIGER 16M	PX3 Pilatus 2M-F
Space group	P65	P1	P65
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	109.09, 109.09, 383.80	109.07 109.14 385.64	108.70, 108.70, 383.40
$\alpha$ , $\beta$ , $\gamma$ (°)	90.00, 90.00, 120.00	90.02 90.01 59.98	90.00, 90.00, 120.00
			<i>Peak</i>
Wavelength (Å)	1.00000	0.99998	0.97941
Resolution (Å)	50.0 - 3.2	60.0 - 3.2	50.0 - 4.0
$R_{\text{merge}}$ <sup>a</sup> (%)	9.1 (282.9) <sup>a</sup>	9.2 (235.1)	34.3 (227.3)
$I/\sigma(I)$	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)
<b>Refinement</b>			
Resolution (Å)		54.57 - 3.20	
No. reflections		251,065	
$R_{\text{work}} / R_{\text{free}}$ (%)		26.10 / 28.84	
No. atoms		62,886	
Protein		62,730	
Sulfate		150	
Magnesium		6	
<i>B</i> factors (Å <sup>2</sup> )		146.98	
Protein		146.92	
Ligand/ion		173.44	
r.m.s deviations			
Bond lengths (Å)		0.008	
Bond angles (°)		1.173	
Ramachandran			
Preferred/allowed/disallowed (%)		84.67/12.98/2.34	

<sup>a</sup> Values in parentheses are for highest-resolution shell.

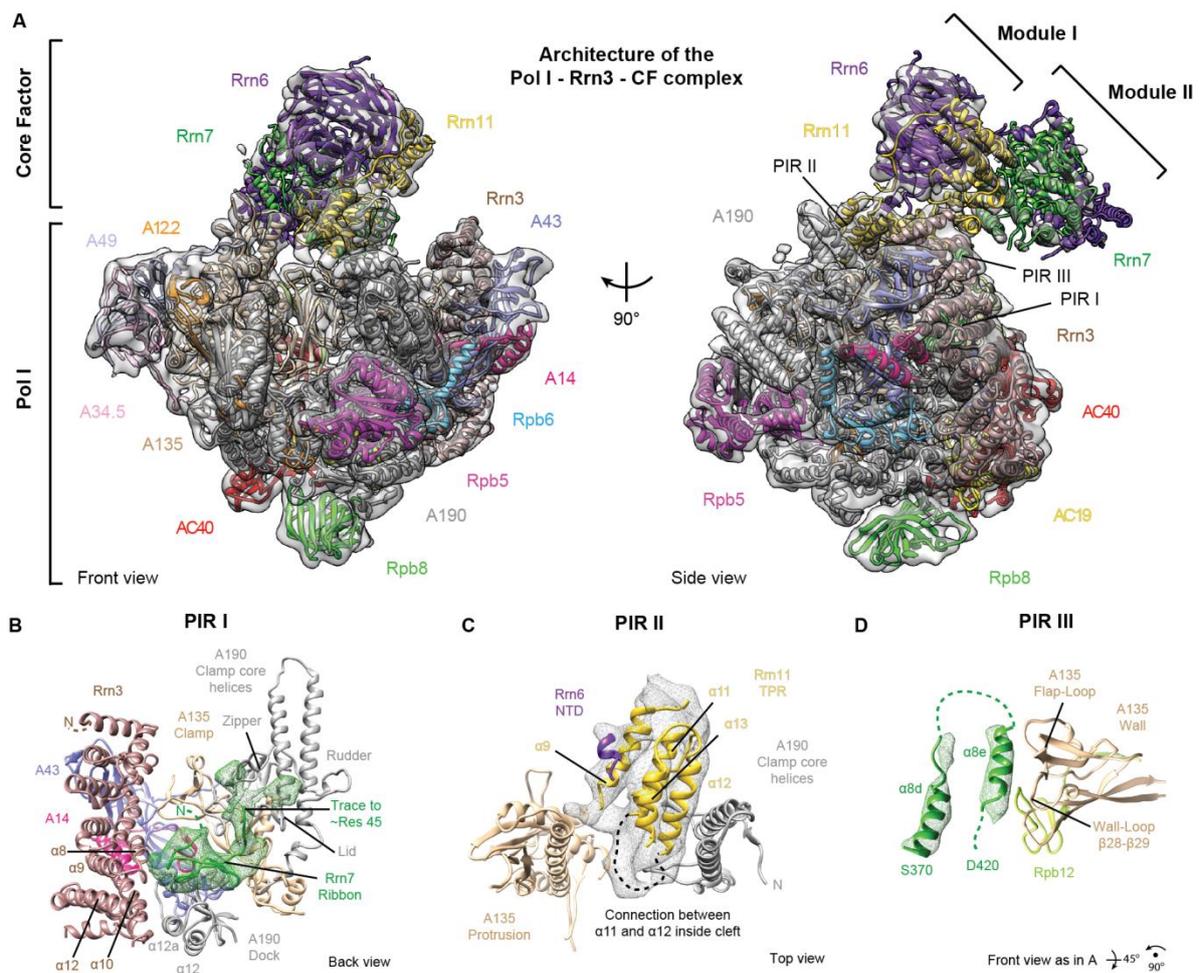
**Table 11. Crystallographic data collection and phasing statistics for anomalous datasets of native and methionine mutant CF crystals.**

	Sulphur - Peak (native crystal)	Rrn11_L430M and Rrn7_I408M	Rrn6_L25M and Rrn7_F438M	Rrn11_W84M	Rrn11_L73M and Rrn7_V212M
<b>Data collection</b>	PX1 EIGER 16M	PX1 EIGER 16M	PX1 EIGER 16M	PX1 EIGER 16M	PX1 EIGER 16M
Space group	P65	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	109.25, 109.25, 384.60	109.20, 109.20, 384.80
$\alpha, \beta, \gamma$ (°)	90.00, 90.00, 120.00	90.00, 90.00, 120.00	90.00, 90.00, 120.00	90.00, 90.00, 120.00	90.00, 90.00, 120.00
	<i>Peak</i>	<i>Peak</i>	<i>Peak</i>	<i>Peak</i>	<i>Peak</i>
Wavelength (Å)	2.06640	0.97848	0.97865	0.97986	0.97849
Resolution (Å)	40.0 - 3.4	35 - 4.0	35 - 4.0	35 - 4.5	35 - 4.2
$R_{\text{merge}}$ <sup>a</sup> (%)	53.5 (194.1) <sup>a</sup>	46.5 (229.1)	34.9 (223.9)	34.4 (128.3)	37.4 (133.8)
$I/\sigma(I)$	29.64 (2.38)	12.40 (2.02)	12.23 (2.05)	4.30 (1.33)	14.61 (2.25)
$CC_{1/2}$	99.8 (37.2)	99.3 (74.1)	99.8 (78.4)	98.4 (48.7)	99.5 (47.1)
Completeness (%)	99.3 (92.6)	99.9 (100.0)	99.8 (100.0)	99.8 (100.0)	99.6 (97.6)
Redundancy	385.2 (81.5)	36.1 (24.4)	39.5 (30.1)	6.0 (6.1)	51.5 (19.5)

<sup>a</sup> Values in parentheses are for highest-resolution shell.

### 3.5. 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 (Figure 20). The purified 18-subunit, 883-kDa complex was subjected to negative-stain EM analysis, leading to a 3D reconstruction at 16 Å resolution (Figure 20, STAR Methods). Subsequent cryo-EM analysis and particle sorting led to three reconstructions at resolutions of 7.7 Å, 8.8 Å and 9.0 Å (Figure 15, Figure 21, Figure 22). 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 (STAR Methods). 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.



**Figure 15. Cryo-EM reconstruction of the Pol I-Rrn3-CF complex**

(A) Overview of the Pol I-Rrn3-CF cryo-EM reconstruction at 7.7 Å resolution (grey envelope) with positioned Pol I subunits (cartoon). Secondary structure elements are visible except for CF module II. Rrn3 is in brown, Rrn6 in violet, Rrn7 in green and Rrn11 in yellow. (B) Polymerase-interacting region (PIR) I is located 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, Figure 13). Contacts with the loop between Rrn3 helices  $\alpha 8$  and  $\alpha 9$  become visible. Cryo-EM density (green mesh) allows for additional chain tracing towards the A190 zipper. (C) PIR II (Rrn11 TPR helices) interacts with the A190 clamp core helices and the A135 protrusion (cryo-EM density shown as grey mesh). A loop between Rrn11  $\alpha 11$  and  $\alpha 12$  becomes ordered inside the active center cleft (dashed line). (D) PIR III (Rrn7 helix  $\alpha 8e$ ) binds the wall domain of A135, including the Pol I-specific flap loop and A135 loop  $\beta 28$ - $\beta 29$  (cryo-EM density shown as green mesh).

### 3.6. CF contains three Pol I-interacting regions

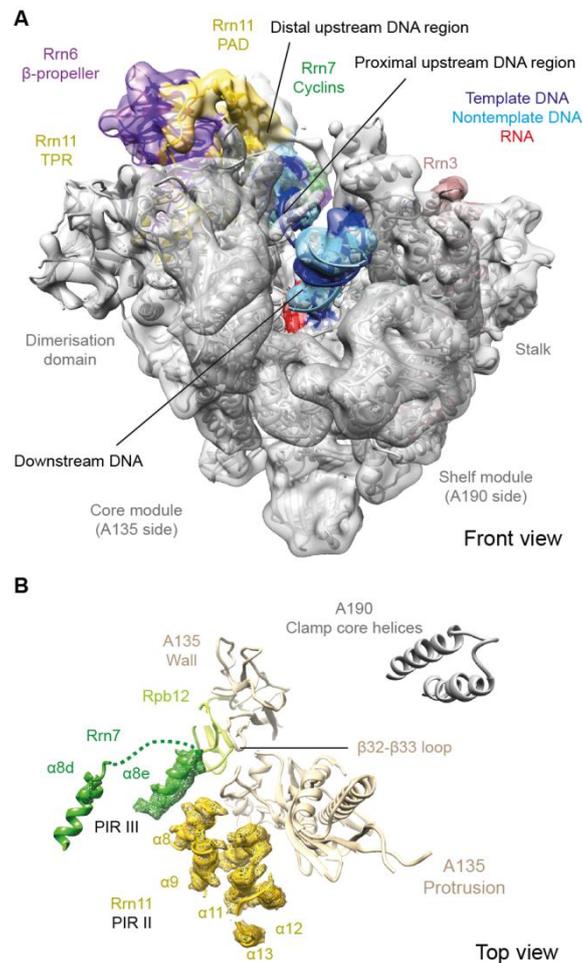
In the Pol I-Rrn3-CF complex structure, module I of CF contacts the upstream end of the polymerase cleft (Figure 15A), whereas module II protrudes outwards and is mobile, adopting different positions in the three EM reconstructions (Figure 22). CF interacts with the Pol I-Rrn3 complex using three polymerase-interacting regions ('PIRs') that are defined in the cryo-EM density at the level of secondary structure (Figure 10, Figure 15B-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  $\alpha 8$ - $\alpha 9$  (Figure 15B). The C-terminal half of the Rrn11 TPR domain (PIR II) contacts the Pol I clamp and protrusion, leading to an ordering of loop  $\alpha 11$ - $\alpha 12$  (Figure 15C). Helix  $\alpha 8e$  of the Rrn7 insertion domain (PIR III) contacts the Pol I wall (Figure 15D). 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 DNA-binding modes in the two systems. In the structure, CF blocks the upstream end of the Pol I cleft, indicating that major structural changes must occur for DNA loading into the active center.

### 3.7. 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 (STAR Methods). The scaffold consisted of promoter DNA from positions -80 to +23 (+1 depicts the transcription start site, TSS), a mismatched bubble region (positions -8 to +8), and a 6-nucleotide RNA product. After selecting 345,000 particles that contained complete ITC complexes in a defined conformation (Figure 23), we obtained a cryo-EM structure at an overall resolution of 3.4 Å (Figure 16) with higher resolution at the center and lower resolution at the periphery (Figure 24).

In the resulting ITC structure, Pol I was no longer in the expanded conformation but adopted the contracted conformation observed in the EC (Neyer et al., 2016, Tafur et al., 2016). CF occupied an alternative position on Pol I (Figure 16A) 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 60°, binding the outer surface of Pol I between the 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 contacts the outer protrusion and PIR III

contacts Rpb12 and the Pol I-specific wall loop  $\beta$ 32- $\beta$ 33 (Figure 16B). Compared with the Pol I-Rrn3-CF complex, CF employs the same PIRs to contact Pol I in the ITC. However, CF occupies a different position that enables promoter DNA loading into the cleft and cleft contraction, which is impossible in the free Pol I-Rrn3-CF complex.



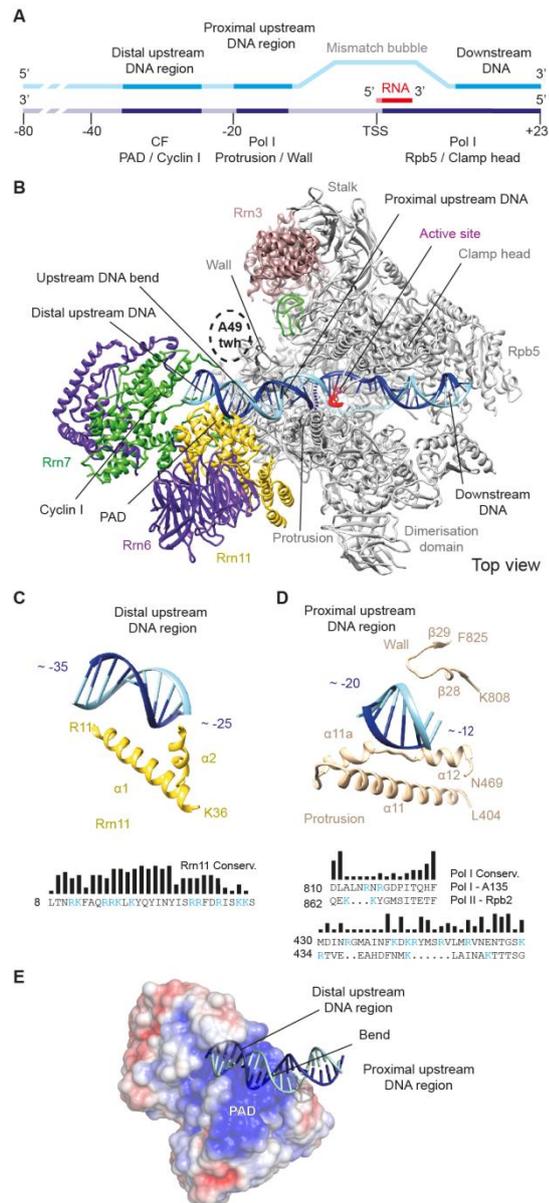
**Figure 16. Cryo-EM reconstruction of the initially transcribing complex**

(A) Overview of the yeast Pol I ITC cryo-EM reconstruction (grey envelope). Pol I is shown as grey cartoon, Rrn3 and CF subunits are colored as in Figure 15. Template DNA, nontemplate DNA and RNA are in dark blue, light blue and red, respectively. (B) PIR II/III contact the Pol I protrusion. Cryo-EM density for TPR helices (yellow mesh) and cyclin II insertion helix  $\alpha$ 8e (green mesh).

### 3.8. Promoter DNA location and contacts

The course of DNA through the ITC can be followed for most of its path from around position -35 upstream to position +20 downstream of the active site (Figure 17). Upstream DNA occupies a previously unobserved location (Figure 17B) that differs from that in Pol II initiation complexes. Upstream DNA runs alongside the Rrn11 PAD, bends by  $\sim 30^\circ$ , and enters the cleft between the Pol I wall and protrusion. Upstream DNA thus forms two major contacts that are likely relevant for promoter recognition (Figure 17C/D). A distal upstream contact is formed between the DNA region around registers -35 to -25 and the Rrn11 PAD helices  $\alpha 1$  and  $\alpha 2$ , which contact the DNA major and minor groove, respectively (Figure 17C), and form a positively charged patch (Figure 17E). A proximal upstream contact is formed between the promoter region around registers -20 to -12 and elements in the Pol I protrusion and wall domains that are specific to Pol I and conserved amongst yeast species (Figure 17D). The extend of the upstream promoter region contacted by CF and Pol I agrees with early DNase footprinting studies (Kownin et al., 1987).

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). In the EC, upstream DNA adopts a different orientation, but is also located near the polymerase protrusion and wall (Tafur et al., 2016), as in the ITC. Upstream DNA also contacts the flexible A49 tandem winged helix (tWH) domain in one EC structure (Tafur et al., 2016). We could indeed reveal the tWH domain in a similar position in the ITC after further classification (Figure 23, Figure 24). This suggests that the flexible tWH domain moves with upstream DNA during the transition from the ITC to the EC.

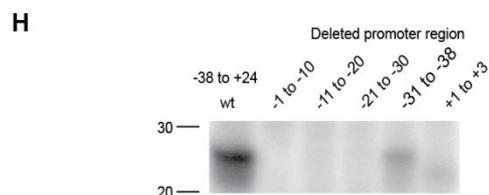
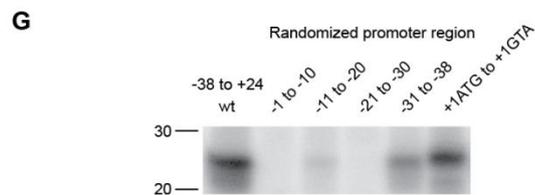
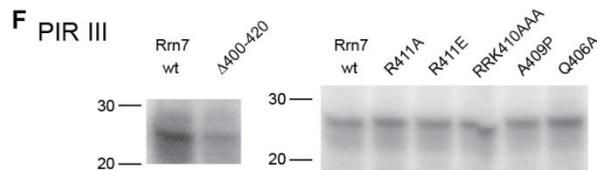
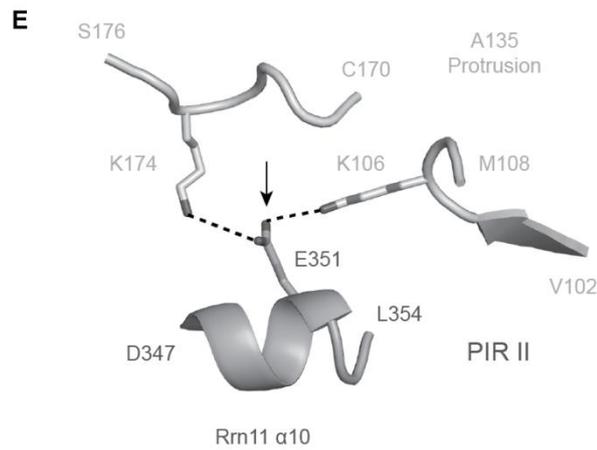
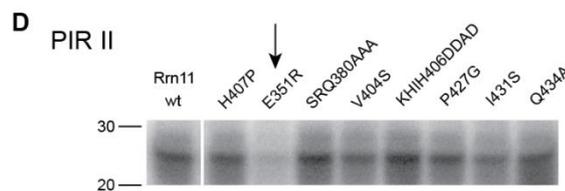
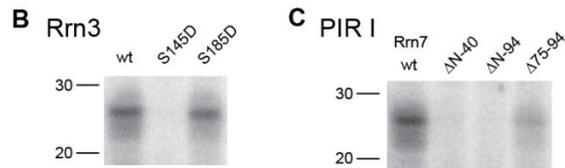
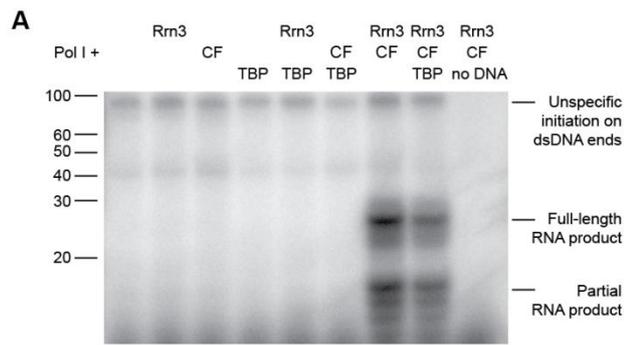


**Figure 17. Promoter DNA interactions**

(A) Schematic view of the used promoter DNA scaffold. Semi-transparent stretches are not included in the model. Promoter contacts are labelled according to their position relative to the TSS below the scheme. (B) Top view of the yeast ITC in cartoon representation depicts DNA contact sites. (C) Distal upstream DNA extends beyond the polymerase and is recognized by Rrn11 PAD helices  $\alpha 1$  and  $\alpha 2$ . Helix  $\alpha 2$  marks the edge of straight DNA before it bends towards the polymerase. Helix  $\alpha 1$  stretches along the major groove. The sequence of helix  $\alpha 1$  is given (positively charged residues in blue) and the conservation among yeast species is indicated (compare Figure 14). (D) Pol I protrusion and wall domains contact upstream promoter DNA. A structure-based sequence alignment of *S. cerevisiae* Pol I and Pol II and conservation of Pol I residues amongst yeast species are indicated. (E) Surface representation of CF (view as in (B)) colored according to electrostatic surface charge as calculated with the Adaptive Poisson-Boltzmann Solver (APBS) (Baker et al., 2001) ranging from -2 kT/e (red) to +2 kT/e (blue).

### 3.9. Basal Pol I initiation assay

To test the functional significance of our structural observations, we set up a basal initiation assay. We used closed promoter DNA encompassing positions -38 to +24 of the natural *S. cerevisiae* rDNA promoter (STAR Methods), 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  $\alpha^{32}\text{P}$ -labelled ATP, to form a Pol I-specific full-length RNA product of 24 nucleotides in length (Figure 18A). To validate our assay, we used Rrn3 variants with inhibitory point mutations S145D or S185D that interfere with Pol I binding (Blattner et al., 2011, Engel et al., 2016). Indeed, these point mutations in Rrn3 abolished (S145D) or slightly reduced (S185D) Pol I initiation (Figure 18B), providing a positive control. These results show that Pol I requires only recombinant Rrn3 and CF, but not TBP, for basal initiation *in vitro*, consistent with published results (Bedwell et al., 2012, Keener et al., 1998, Pilsl et al., 2016). The data also demonstrate the relevance of our structural studies for understanding the basic mechanism of Pol I initiation.



**Figure 18. Functional probing of structural features. Figure legend on next page.**

## Figure 18. Functional probing of structural features

(A) Pol I-specific initiation assay evaluated by urea polyacrylamide gel electrophoresis (Methods). A band at 24 nucleotides length reflects the run-off product and is only observed when both CF and Rrn3 are added to Pol I with nucleotide triphosphates (NTPs). The specific band is Pol I-specific (Figure 25A). TBP is not required, and slightly reduces initiation efficiency, likely due to unspecific DNA binding. Oligonucleotide lengths indicated on the left. (B) Phospho-mimetic Rrn3 variants abolish (S145D) or reduce (S185D) initiation, due to interference with Pol I-Rrn3 binding. (C) PIR I mutants abolish basal initiation, a linker deletion reduces initiation efficiency. (D) Point mutations of PIR II show little effect, with the exception of Rrn11 E351R (a line indicates the position of a removed lane). (E) Structural rationalization of the effect observed in panel D. Charge reversal of Rrn11 residue E351 would break salt bridges, preventing the interaction with protrusion residues K106 and K174 in the ITC. (F) Deletion of the Rrn7 insertion helix  $\alpha 8e$  (PIR III) impairs initiation, whereas point mutations have no effect. (G) Randomization of 10 base pair stretches in the upstream promoter abolishes initiation (-38 to -31 reduces initiation). Reversal of the first three transcribed base pairs showed little effect. (H) Deletion of 10 base pair stretches, as in (G), abolishes initiation (-38 to -31 deletion reduces initiation). Deletion of the first three transcribed base pairs shows a strong reduction in initiation efficiency and a shorter product. Overall scaffold length was kept constant by extending the promoter with its native sequence on its upstream edge to compensate for deleted base pairs (STAR Methods).

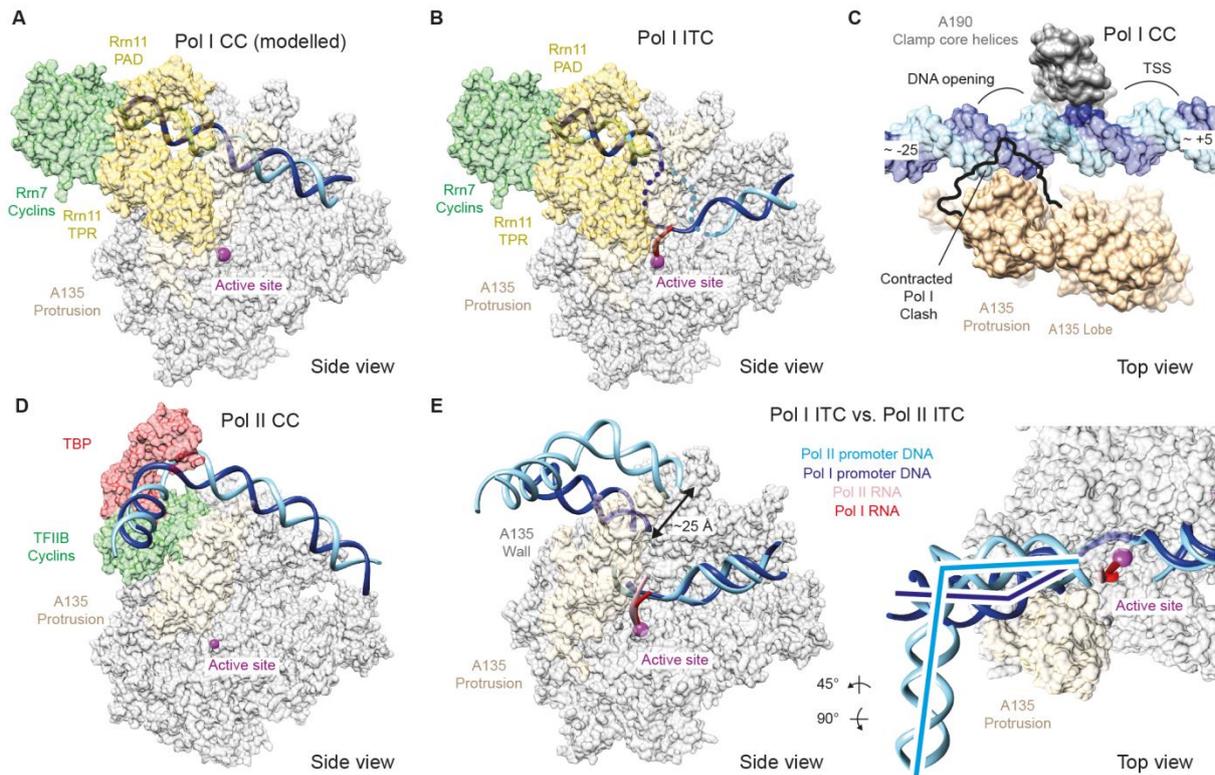
### 3.10. Determinants of Pol I initiation

We next tested recombinant purified CF variants for their effect on Pol I initiation (Figure 18). In these experiments, mutations in PIR I had a drastic effect, demonstrating that the Rrn7 ribbon domain is essential for initiation (Figure 18C). Mutation of residues in PIR II showed weak effects, with the exception of the E351R charge reversal mutation (Figure 18D), which apparently destroys salt bridges with K106 and K174 of Pol I subunit A135 in the ITC (Figure 18E). A deletion of Rrn7 helix  $\alpha$ 8e of PIR III reduced initiation, but mutation of single residues in this helix had no effect (Figure 18F).

We also tested which regions of the Pol I promoter were important for initiation (Figure 18G, Figure 18H). 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, whereas the DNA region between 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.

### 3.11. Models of the CC and OC

To understand Pol I initiation, structures of the CC and OC are also required. To obtain a model of the CC, we extended upstream promoter DNA in the ITC structure with canonical B-DNA running through the cleft. This leads to clashes with the protrusion and the rudder on opposite sides of the cleft (Figure 24I). To avoid these clashes, we exchanged the contracted Pol I-Rrn3 structure observed in the ITC against the partially expanded Pol I-Rrn3 structure (Engel et al., 2016). In the resulting CC model, DNA runs through the cleft with only minor clashes with the flexible rudder loop, and contacts Rpb5 at the downstream end of the cleft (Figure 19, Figure 24J). We could also model the OC by simply removing RNA from our ITC structure. We predict that this results in a good OC model. First, the DNA template must be present in the active center of the OC to position nucleoside triphosphate substrates for initial RNA synthesis. Second, known structures of the OC and ITC of Pol II are virtually identical, except for the lack of 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, as described below.



**Figure 19. Pol I initiation-elongation transition and comparison with Pol II**

(A) Pol I CC model viewed from the side. The core module is partially and Rm6 fully omitted for clarity. DNA (blue), the A135 protrusion (wheat), Rm11 (yellow), Rm7 (green) and the active site metal ion A (magenta) are highlighted. Straight B-DNA was fitted into the density of distal upstream DNA in the ITC and runs through the cleft contacting Rpb5 on its downstream end (see text). (B) Pol I ITC side view colored as in (A). Proximal and distal promoter DNA is connected by bent B-DNA. (C) Top view of CC model. A black outline marks the position of the protrusion assuming a contracted cleft state and shows a clash with promoter DNA. Partially expanded polymerase (PDB 5G5L, space filling) avoids the clash. (D) Pol II CC structure (PDB 5FZ5) viewed from the side as in (A). (E) Comparison of Pol I and Pol II ITCs (PDB 4V1N, DNA in cyan) shows that upstream promoter DNA trajectories differ. Pol II promoter DNA (light blue) is located on the rims of the cleft (guided by TFIIIB), whereas its Pol I counterpart (dark blue) is located at the floor of the cleft (proteins omitted for clarity). In a complete initiation complex, UAF and TBP may influence the far upstream promoter DNA (not shown), but binding to the displayed sequence is occluded by CF and Pol I itself.

### 3.12. Discussion

Here, we use an integrated structural biology approach to elucidate the structural basis of basal 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, Fernandez-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 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.

Our results converge with previous data on a model for basal Pol I initiation and the transition to elongation (Figure 19). 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, Pilsel 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 parts in the polymerase protrusion and wall, likely contributing to promoter recognition. Fourth, DNA opening can occur between the protrusion and clamp and enables formation of an OC with downstream DNA positioned as in the EC (Neyer et al., 2016). The OC may be stabilized by the mobile tandem winged helix domain of Pol I subunit A49 that binds proximal upstream DNA (Geiger et al., 2010, Tafur et al., 2016). Fifth, RNA synthesis commences, leading to formation of an ITC and complete cleft contraction (Table 12). Finally, RNA growth leads to displacement of the Rrn7 ribbon that occupies the RNA exit site, likely triggering displacement of CF and formation of an 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 (Figure 19). In the Pol I system, promoter DNA runs over the wall and through the active center cleft (Figure 19A/B). 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 (Figure 19D/E). 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  $\sim 30^\circ$  around register -20 in one direction, the Pol II promoter is bend by  $\sim 90^\circ$  at

register -35 in the opposite direction (Figure 19E). In the complete Pol I initiation complex, DNA may extend further upstream to bind UAF, but this is not expected to influence the promoter location between position -35 and the TSS, which is defined by interactions with Pol I and CF.

Our observations also provide insights into the possible molecular basis of promoter specificity and opening. The Pol I initiation machinery may recognize a combination of the ‘bendability’ of upstream promoter DNA and the ‘meltability’ of the region around the TSS. Our structural observations suggest that docking of the CF-DNA complex to the Pol I-Rrn3 complex is only possible when promoter DNA can be bent in a specific way at the CF-Pol I interface. Distal and proximal upstream promoter regions contact CF and Pol I, respectively (Figure 17C/D). These promoter contacts can only be maintained upon CF-Pol I docking if the connecting DNA adopts the observed bend. We further predict that DNA will only open when a meltable DNA region is placed between the clamp and protrusion at the appropriate distance from the upstream DNA bend. Only DNA that combines such bendability and meltability would be trapped in the cleft, leading to stabilizing interactions in an OC that compensate for the energetic cost of DNA opening.

Published data are in line with this model. First, *in vitro* initiation is largely unaffected by mutation of single bases in the Pol I promoter (Kownin et al., 1988) but is defective if insertions or deletions of >5 bp are introduced to the proximal upstream promoter region (Kownin et al., 1987). Second, a specific DNA conformation rather than a specific sequence was previously predicted to determine promoter specificity (Kownin et al., 1987). Third, bioinformatic prediction of Pol I promoter properties (Bolshoy et al., 1991, Goodsell and Dickerson, 1994) suggest that a bend is present or can be introduced around position -20 in a large range of species (Figure 25L). Our model for promoter recognition thus offers an explanation for the long-known lack of conservation of Pol I promoter sequences in different organisms (Moss et al., 2007).

Finally, our results elucidate the evolution of transcription initiation mechanisms. Several aspects of Pol I initiation resemble initiation in the bacterial  $\sigma$ 70-dependent transcription system (Murakami et al., 2002, Zhang et al., 2012, Zuo and Steitz, 2015). Pol I initiation involves contraction of the cleft, similar to the cleft closure observed in the bacterial system (Chakraborty et al., 2012). Pol I can open DNA without ATP hydrolysis, like bacterial polymerase, and this may involve trapping of the non-template single strand as in the bacterial system (Feklistov and Darst, 2011). Such spontaneous DNA opening can also occur in the Pol

II system (Plaschka et al., 2016) but may be impaired because DNA is first positioned further outside the cleft, limiting its interactions with Pol II. This may give rise to a requirement for TFIID to open DNA upon ATP hydrolysis (Grunberg et al., 2012, Kim et al., 2000). These considerations also suggest a function for the Pol I-Rrn3-CF complex structure resolved here (Figure 15). In this structure, CF is positioned above the cleft (Figure 15), where it could impair binding and spontaneous opening of non-promoter DNA and production of aberrant RNA.

## 4. Conclusion and Outlook

With this study we expand the knowledge of basal Pol I transcription (Chapter 1.3) by combining X-ray crystallography, cryo-EM and functional analyses. Here we describe the structure of CF at 3.2 Angstrom resolution which reveals a unique architecture comprising two modules linked by a hinge region. Module I contains a seven bladed  $\beta$ -propeller in Rrn6 and the entire subunit Rrn11, which contains a TPR domain and a PAD. Module II is formed from Rrn7 that contains unique insertions in its cyclin domains and a novel helical domain termed ‘headlock’, formed by the N-terminal part of Rrn6.

The CF crystal structure provides an updated domain architecture compared with a previous model obtained by a combination of XL-MS, bioinformatics prediction, and protein modeling (Compare Figure 6 and Figure 7 with Figure 13 and Figure 14). The structure of CF with unusual domain arrangements and big insertions explains some of the difficulties of modelling such complexes and shows the importance of acquiring accurate atomic models. Furthermore, we elucidated the position of CF in complex with Pol I and Rrn3 and in the ITC with DNA and RNA. These structures show how the Pol I initiation system differs from the Pol II initiation system.

Together, the structures of CF and ITC provide the basis for addressing new questions: How does the CF-promoter DNA binding take place in detail? How can we extend the analysis to include UAF in order to more fully understand Pol I transcription initiation? What is the role of TBP in Pol I initiation?

### 4.1. Future crystallization experiments

The cryo-EM structure of the ITC indicates that CF contacts promoter DNA in a very different way compared to TFIIB within the Pol II initiation system (Engel et al., 2017). However, details of the interaction between DNA and CF could not be observed because of the overall flexibility of the ITC and the limited overall resolution. The next step is to co-crystallize CF with different lengths of promoter DNA. High resolution data could then explain which amino acids of CF are involved in DNA recognition and binding. Furthermore, based on the structural information of the CF, minimal constructs could be designed that include only the structural domains relevant for DNA binding. These smaller CF constructs could lead to new crystal forms that may diffract to higher resolution.

TBP binds between the UAF/UAS and CF/CE on the DNA (Keys et al., 1996, Keener et al., 1997, Peyroche et al., 2000). A more speculative crystallization attempt is to combine the components CF, DNA and TBP directly for crystallization. These crystallization experiments have the potential to reveal the role of TBP in the Pol I initiation system. In a long term project the structure determination of the human homolog of CF, SL1, can be attempted (Comai et al., 1992, Learned et al., 1985). Crystallization of full length SL1 can be combined with domain constructs designed based on the CF structure.

## **4.2. More complete understanding of transcription initiation**

The structures of the Pol I initiation factors TBP, Rrn3 and with this study CF are known (Blattner et al., 2011, Engel et al., 2017, Kim et al., 1993). Only the structure of UAF is missing. UAF consists of UAF30, Rrn9, Rrn10, Rrn5, and histones H3, and H4 (Keys et al., 1996, Keener et al., 1997) and interacts with CF (Schneider, 2012). The structure of UAF is the next essential step to understand the full mechanism of transcription initiation in *S. cerevisiae*.

It was shown before that UAF can be purified endogenously (Keener et al., 1998). In order to get sufficient amount of pure UAF for structural studies, the recombinant expression of UAF in *E. coli* or in insect cells should be attempted. The crystallization of UAF domains or the whole complex should be tried with and without limited proteolysis, which has proved to be a powerful tool in protein crystallography (Ringel et al., 2011). Furthermore, since the minimal molecular weight required for successful structure determinations by cryo-EM is decreasing rapidly, UAF from *S. cerevisiae* with a molecular mass of 153 kDa, could be a promising target for cryo-EM in the future.

With structural information on the UAF, the ITC of Pol I, Rrn3, CF and promoter DNA can be extended to the complete ITC. Cryo-EM would be the perfect tool to analyze this complex. The result would provide an insight into the interaction of UAF with CF and its binding on the UAS. Furthermore, it is reported that TBP is located between UAF and CF (Schneider, 2012) but it's role in Pol I transcription initiation is unclear. Including TBP in the complex could provide an answer to a decades long discussion about the role of TBP in the Pol I transcription initiation process.

### 4.3. Targeting Pol I initiation for Cancer Therapeutics

In cells the rate of ribosome biogenesis is linked with cellular growth and proliferation (Ruggero and Pandolfi, 2003). Normally, it is regulated quite tightly but in tumorigenesis the regulation gets disrupted between the extracellular signaling and the ribosome biogenesis (White, 2005, White, 2008, Drygin et al., 2010, Silvera et al., 2010). Since Pol I synthesizes the rRNA precursor in eukaryotes, it is an excellent target in cancer therapy (Moss et al., 2007). The human ortholog of CF is SL1 (Comai et al., 1992, Learned et al., 1985). SL1 consists of TAFIA, -B, and -C, which are homologs of Rrn11, Rrn7 and Rrn6, respectively, and two additional subunits (Russell and Zomerdijk, 2006, Gorski et al., 2007, Denissov et al., 2007). ChIP analysis indicates that the compound CX-5461 reduces the Pol I association with the rDNA promotor by 40-60 % by disruption of the SL1-rDNA complex (Drygin et al., 2011). SL1 is a good target for cancer therapy because it is only required for Pol I transcription initiation.

Our study showed how CF contacts DNA. There is a possibility that the cancer drug CX-5461 might also bind to the CF from *S. cerevisiae* because the domains in Rrn11 and Rrn7, which bind DNA, are also likely to be partially conserved in SL1 (Figure 13, Figure 14). With our established transcription assays the effect of CX-5461 in the basal yeast Pol I initiation system could be tested and if successful the binding visualized by a single soaking/co-crystallization experiment of full length CF with CX-5461. This would show how CX-5461 binds to its target and would lead to a better understanding and maybe improvement of the anti-cancer drug. If the crystallization of SL1 is successful (see previous chapter 4.1) then the binding of CX-5461 can be visualized directly on its human target. In general, the knowledge of the mechanism of Pol I transcription initiation, and its importance in cancer cells, will lead to the prospect of new drug targets. The main focus for new Pol I specific drug targets could be the PIR, and PAD that is likely targeted by CX-5461.

## 5. Appendix

### 5.1. Supplemental data.

Table and supplemental figures in this section were prepared by Christoph Engel and Simon Neyer.

Christoph Engel\*, **Tobias Gubbey\***, Simon Neyer\*, Sarah Sainsbury, Christiane Oberthuer, Carlo Baejen, Carrie Bernecky, Patrick Cramer. (2017) Structural Basis of RNA Polymerase I Transcription Initiation. *Cell* Volume 169, Issue 1, Pages 120-131.

\*These authors contributed equally.

<http://dx.doi.org/10.1016/j.cell.2017.03.003>

For details see author contributions at the chapter Publication.

#### 5.1.1. Pol I cleft expansion and contraction.

**Table 12. Pol I cleft expansion and contraction**

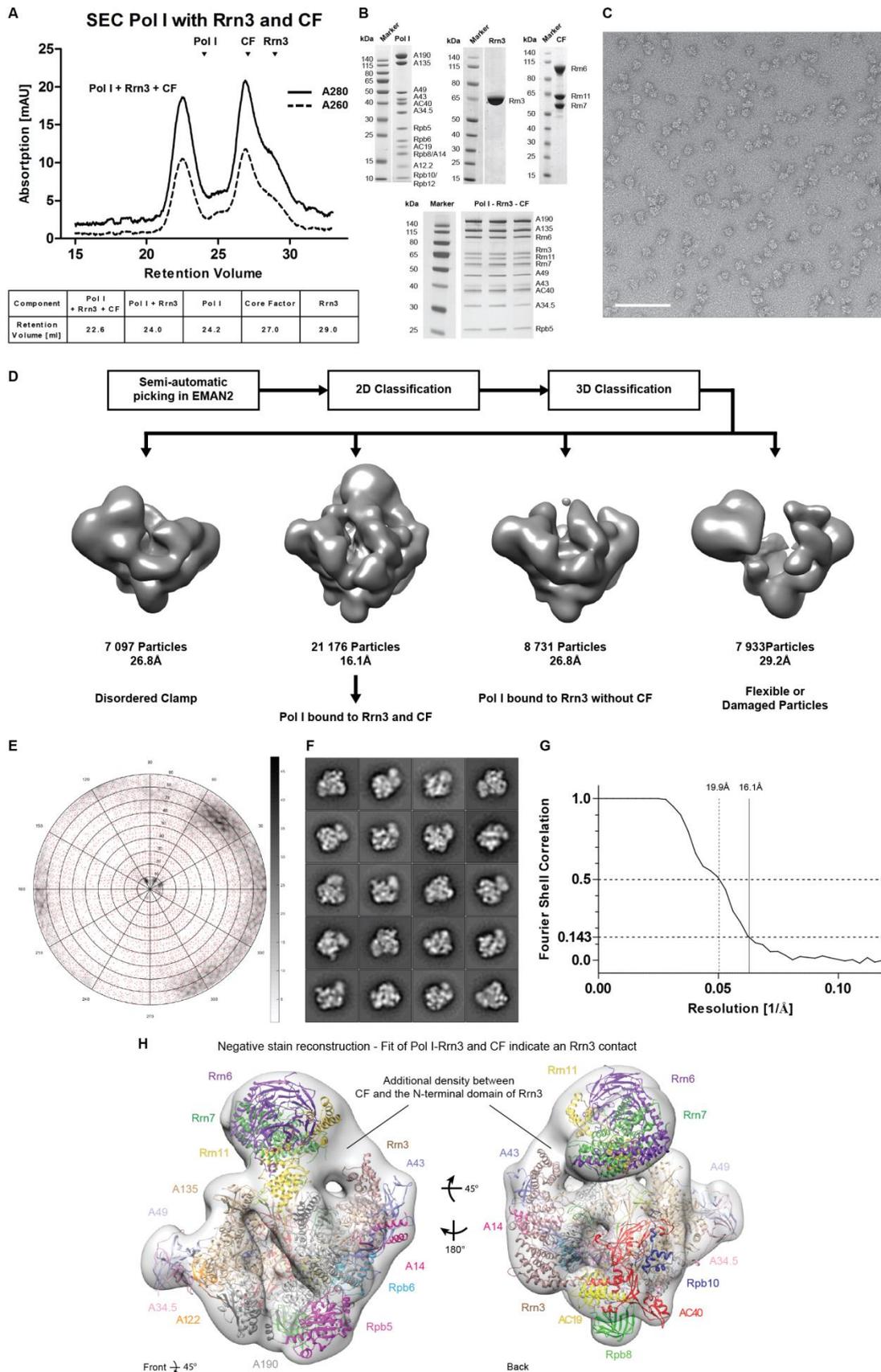
	Downstream <sup>a</sup> [Å]	Upstream <sup>b</sup> [Å]	A190 <sub>414</sub> relative to 4C2M <sup>c</sup> [Å]
Dimer (4C2M)	42	41	-
Free Monomer (5M3M)	41	39	6
Pol I - Rrn3 (5G5L)	41	36	6
Pol I-Rrn3-CF (5N5Y)	39	36	7
ITC (5N61)	35	29	15
EC (5M3F)	34	28	17

<sup>a</sup> Structures were superimposed on A135 subunits. Distances measured between C<sub>α</sub> atoms of residues G231 and K1331 in the largest Pol I subunit A190.

<sup>b</sup> Distances measured between C<sub>α</sub> atoms of residues E414 in subunit A190 and K434 in subunit A135.

<sup>c</sup> Displacement of the C<sub>α</sub> atom of the A190 clamp core residue E414 towards the protrusion in comparison to the crystal structure 4C2M.

## 5.1.2. Formation and negative stain EM reconstruction of a Pol I-Rrn3-CF complex

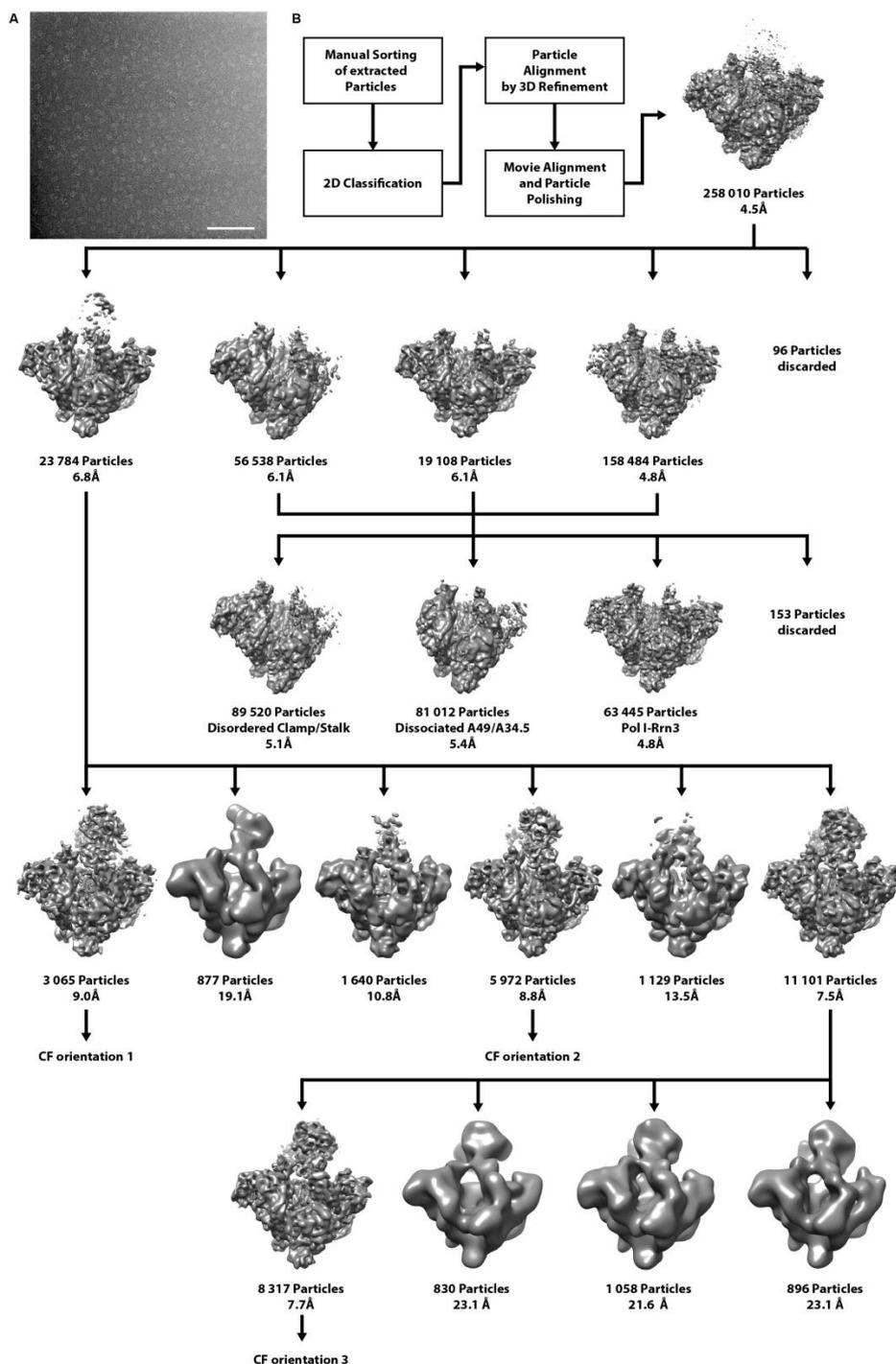


**Figure 20. Formation and negative stain EM reconstruction of a Pol I-Rrn3-CF complex. Figure legend on next page.**

**Figure 20. Formation and negative stain EM reconstruction of a Pol I-Rrn3-CF complex.**

(A) Size exclusion profile of a sample containing Pol I and a 5-fold molecular excess of Rrn3 and CF. Retention volumes of A280 peaks of single samples are indicated. Two 10/300 Superose 6 columns (GE Healthcare) were connected. (B) Coomassie-stained SDS-PAGE gels of Pol I, Rrn3 and CF after purification and of the complex peak after assembly. (C) Exemplary micrograph of the negative stain dataset. The scale bar is 100 nm. (D) Processing and classification tree. A Pol I-Rrn3 structure and two structures of Pol I with damaged particles or clamp/stalk flexibilities are observed after a single round of 3D classification. The number of particles with bound CF is high (21,176 of 44,937 particles), likely due to higher stability originating from grids with continuous carbon support film and gradient fixation (Methods). (E) Angular distribution of single particle orientations used for the Pol I-Rrn3 complex reconstruction. Shades indicate the number of particles assigned to a view; red dots indicate represented views. (F) Representative 2D class averages of the particles used for the Pol I-Rrn3 negative-stain EM reconstruction. (G) Fourier shell correlation (FSC) plot for half-maps of the Pol I-Rrn3-CF negative stain reconstruction with 0.143 and 0.5 FSC criteria indicated. The first data point after phase randomization is omitted. (H) A fit of Pol I, Rrn3 and CF into the negative stain envelope shows density for the entire CF.

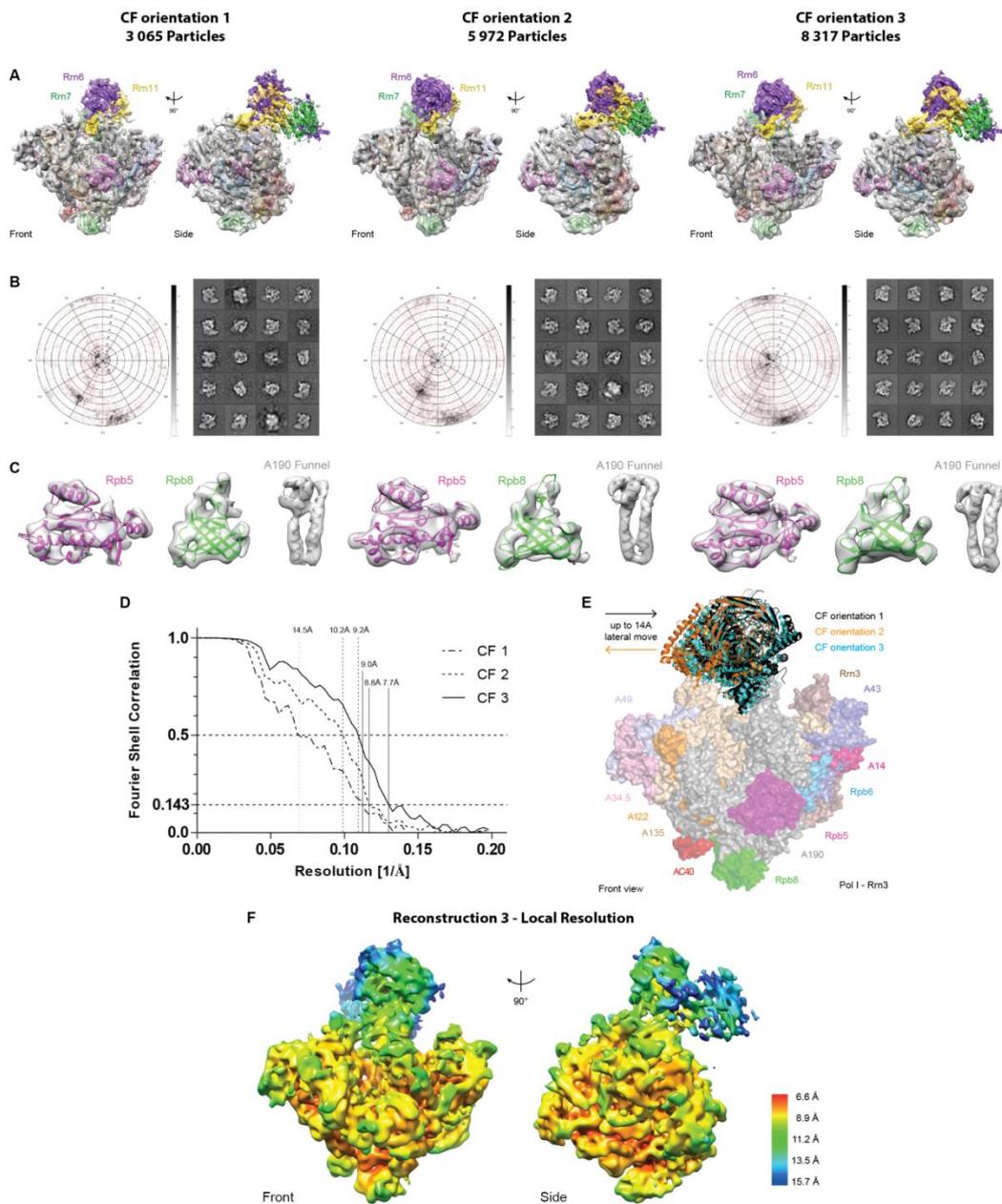
### 5.1.3. Classification of the Pol I-Rrn3-CF cryo-EM dataset



**Figure 21. Classification of the Pol I-Rrn3-CF cryo-EM dataset.**

(A) Exemplary micrograph of the Pol I-Rrn3-CF 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.

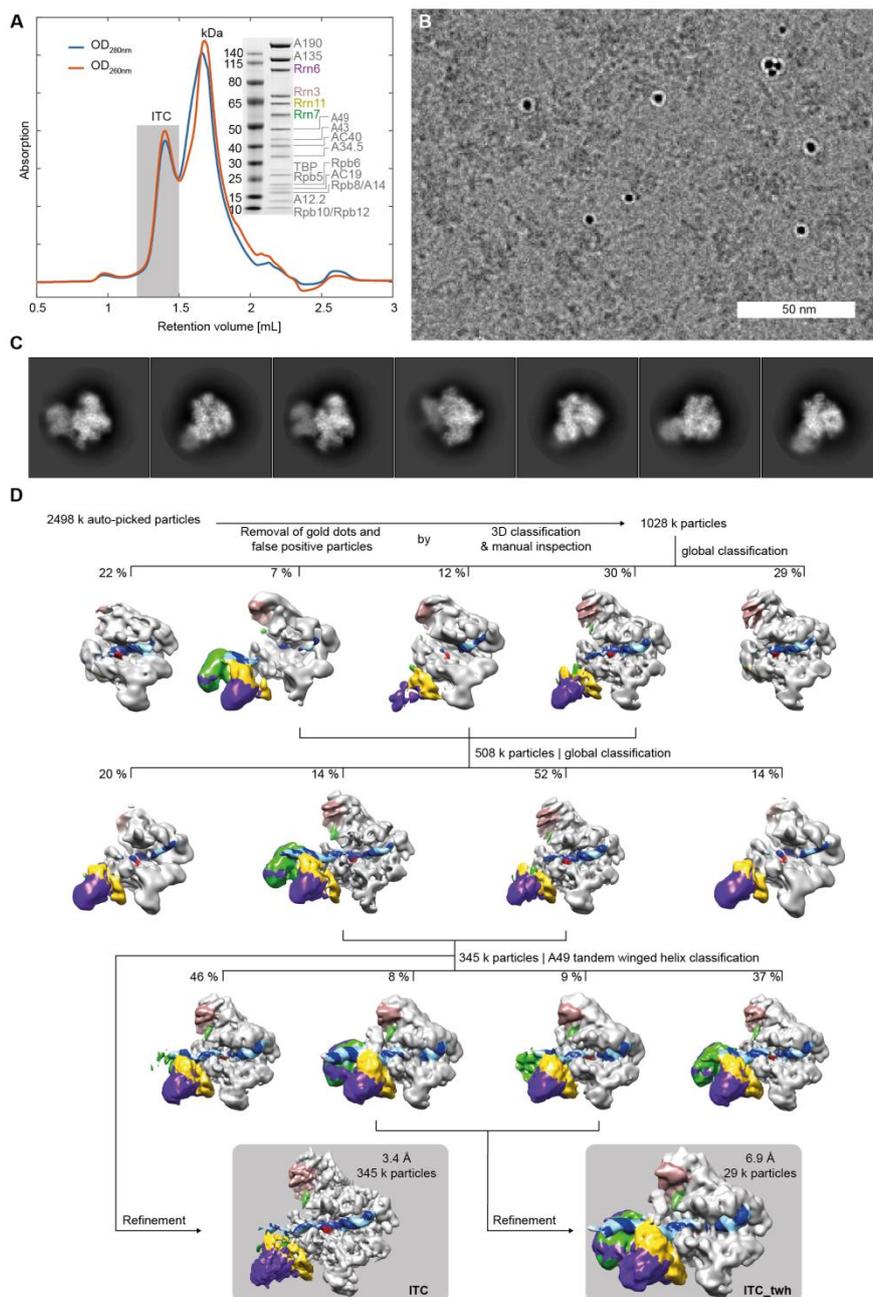
### 5.1.4. Three Pol I-Rrn3-CF reconstructions



**Figure 22. 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 are more flexible. (B) Angular distribution plots (left) and 2D class averages (right) for the reconstructions. Shades indicate the number of particles assigned to a view; red dots indicate represented views. (C) Example densities for the subunits Rpb5, Rpb8 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 exhibit a stable relative orientation, while CF moves by up to 14 Å in its periphery. (F) Local resolution of reconstruction 3 displays a well-ordered, high-resolution core and more flexible outer rims, especially CF module II.

### 5.1.5. Pol I ITC sample preparation and cryo-EM data processing



**Figure 23. Pol I ITC sample preparation and cryo-EM data processing.**

(A) Size-exclusion chromatogram (Superose 6 Increase 3.2/300) of reconstituted Pol I ITC. A Coomassie-stained SDS-PAGE of pooled peak fractions reveals an apparently stoichiometric presence of proteins except TBP. (B) Representative cryo-EM micrograph showing particles of the expected size and spherical gold particle contaminations with high contrast originating from used gold grid. (C) Exemplary 2D class averages after unsupervised classification. (D) Processing tree with density elements colored as in Figure 16. The outlined classes represent the presented ITC and an ITC reconstruction with density for the A49 tWH domain.

### 5.1.6. Cryo-EM reconstructions of Pol I-Rrn3-CF ITC

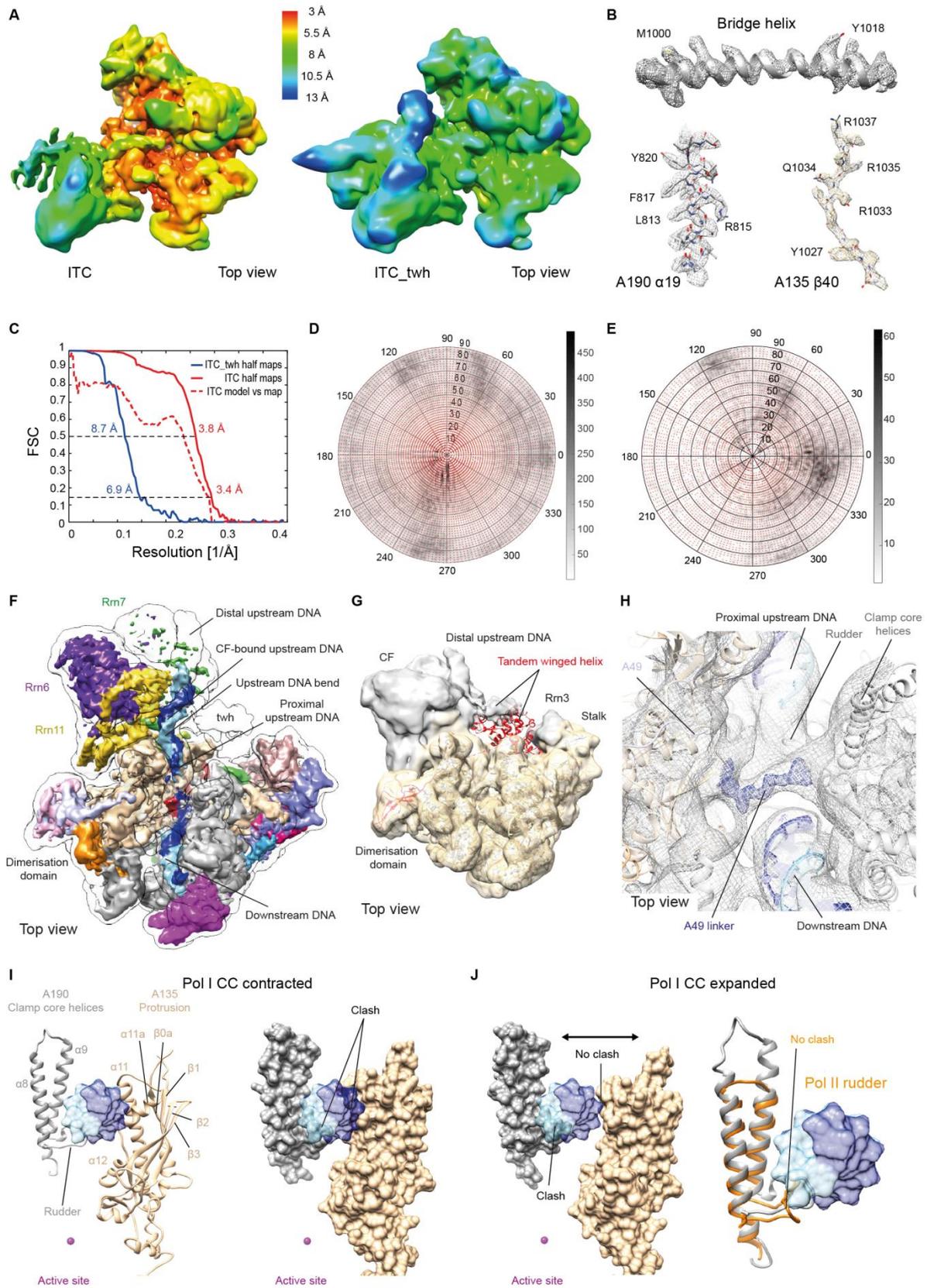


Figure 24. Cryo-EM reconstructions of Pol I-Rrn3-CF ITC. Figure legend on next page.

### Figure 24. Cryo-EM reconstructions of Pol I-Rrn3-CF ITC.

(A) Local resolution surface maps for the ITC (left panel) and an ITC reconstruction with density for the A49 tWH domain (right panel). (B) Representative areas of the cryo-EM density for the Pol I core. The upper panel shows the superimposed bridge helix as a cartoon model. Cryo-EM density (mesh) for A190 helix  $\alpha$ 19 (lower left) and the A135 strand  $\beta$ 40 (right) reveal side chains. (C) FSC curves of the Pol I ITC reconstructions. (D) Angular distribution of particles used to reconstruct the Pol I ITC. (E) Angular distribution of particles used to reconstruct a Pol I ITC with density for the A49 tWH domain. (F) Electron density map for the latter complex (filtered to 6.9 Å; black silhouette) overlaid onto the ITC map (Figure 23) filtered to 4.2 Å. This shows additional density for the A49 tWH domain. Subunit colors as in Figure 15 (G) Position of the A49 tWH domain from PDB entry 5M64 (Tafur et al., 2016) shown as a red cartoon overlaid with our ITC reconstruction showing density for the A49 tWH domain (silhouette). (H) Top view of the central cleft region shows density spanning between the A135 lobe and A190 clamp core helices that likely belongs to the A49 linker. Filtered ITC density is shown as a grey mesh at 6.1 Å and sharpened to 4.2 Å as a blue mesh. (I) Back view of closed promoter DNA modelled into the Pol I cleft using the Pol I ITC conformation. DNA placement results in a major clash with the protrusion and a minor clash with the rudder. (J) Replacing contracted Pol I with the more expanded enzyme in the Pol I-Rrn3 complex (Engel et al., 2016) avoids the clash with the protrusion (left panel). Polymerases were superimposed with their active site regions. Superposition of Pol II clamp core helices onto their Pol I counterparts shows that a Pol II-like rudder conformation would resolve the remaining clash (right).

### 5.1.7. Pol I-specific CF function in transcription initiation

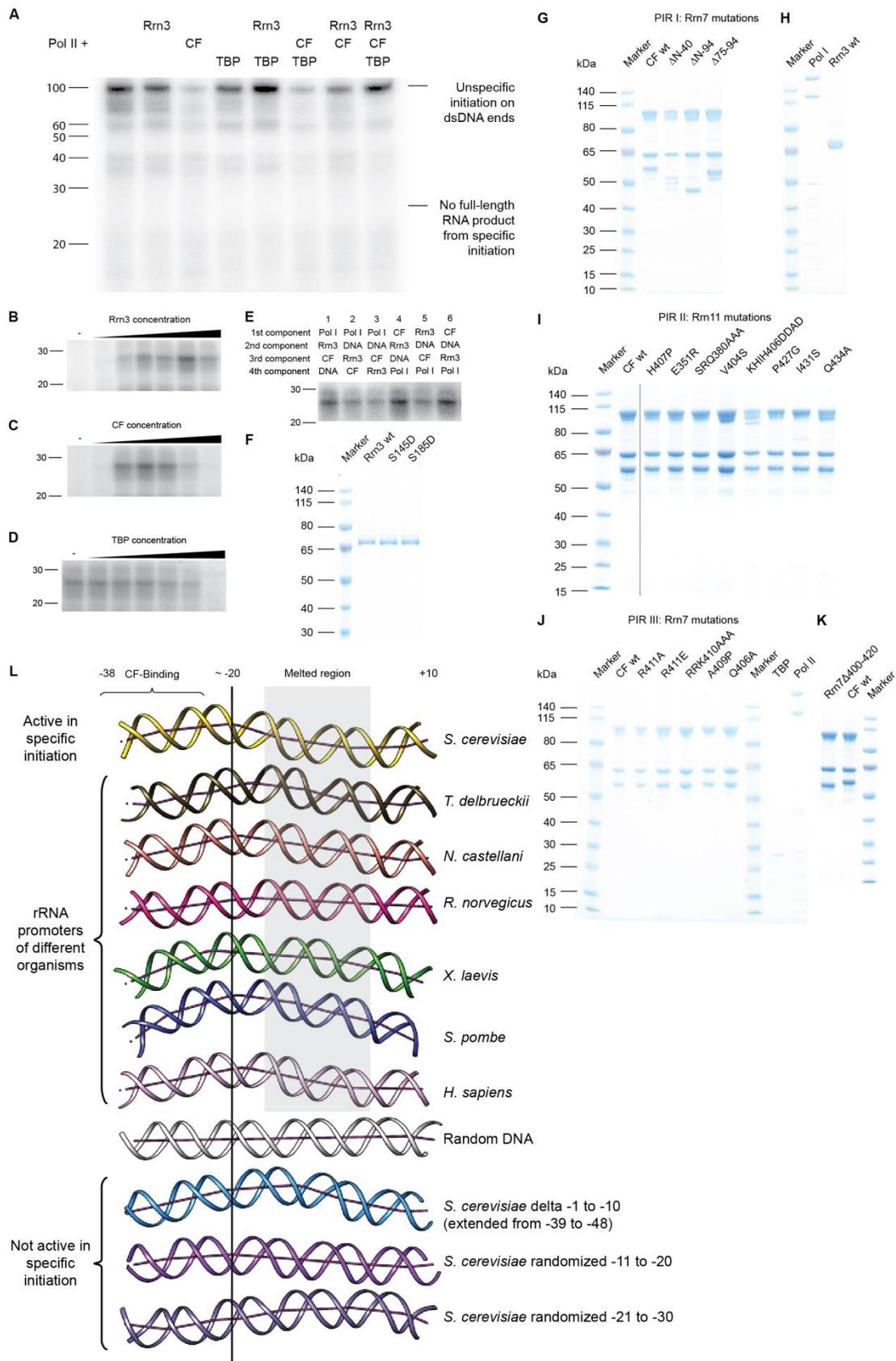


Figure 25. Pol I-specific CF function in transcription initiation. Figure legend on next page.

**Figure 25. Pol I-specific CF function in transcription initiation.**

(A) Initiation assay (Figure 17) using Pol II instead of Pol I (Methods). Pol II shows no specific initiation band at 24 nt, but initiates non-specifically from DNA ends more efficiently than Pol I. A size-marker in nucleotides is indicated on the left. (B) Titration of Rrn3 from 0 to 50-fold molar excess at constant CF concentration shows a saturation at 20x excess. (C) Titration of CF from 0 to 50-fold molar excess at constant Rrn3 concentration shows a reduced signal at 20x and 50x excess, likely due to multiple DNA binding events. (D) Titration of TBP from 0 to 50-fold molar excess at constant Rrn3 and CF concentrations shows a reduced signal at increasing concentrations, likely due to unspecific DNA binding masking CF binding sites. This shows that TBP is not required for initiation in our assay. (E) Order of addition experiments show that the passive Pol I-Rrn3-CF conformation (Figure 15) can be activated (lane 1) and initiates from preformed CF-scaffold complex to which Rrn3 and Pol I are added subsequently (lane 6). Initiation efficiency is reduced when Pol I is pre-incubated with the promoter scaffold (lanes 2 and 3). (F)-(K) Coomassie-stained SDS-PAGE gels of components and CF mutant variants used in specific initiation experiments. A single lane was removed in panel I (black line). (L) Prediction of dsDNA conformation (Bolshoy et al., 1991, Goodsell and Dickerson, 1994) of the used initiation scaffold, Pol I promoters from different organisms and tested, inactive DNAs. For most promoters, a bend around position -20 is suggested, consistent with a higher bendability in this region. CF binding sites and melted regions are indicated.

## 5.2. Extended Material, Methods and Data

Extended material, methods and data in this section were carried out by Christoph Engel and Simon Neyer.

Christoph Engel\*, **Tobias Gubbey\***, Simon Neyer\*, Sarah Sainsbury, Christiane Oberthuer, Carlo Baejen, Carrie Bernecky, Patrick Cramer. (2017) Structural Basis of RNA Polymerase I Transcription Initiation. *Cell* Volume 169, Issue 1, Pages 120-131.

\*These authors contributed equally.

<http://dx.doi.org/10.1016/j.cell.2017.03.003>

For details see author contributions at the chapter Publication.

### 5.2.1. 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 (Figure 21). 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 3,065, 5,972 and 8,317 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 (Figure 22). Resolution is based on the gold-standard FSC (0.143 criterion) and temperature factors were automatically determined and applied in RELION (-628.2 Å<sup>2</sup>, -433.4 Å<sup>2</sup> and -247.8 Å<sup>2</sup> 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 and Cowtan, 2004) 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 UCSF 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 (Pettersen et al., 2004) or PyMOL (Schrodinger, 2015).

### **5.2.2. 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 ~0.10 mg mL<sup>-1</sup>. 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 formate solution, and blotted dry. Images were acquired on a 4k x 4k CCD camera (TVIPS) at 88k 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 40Å). Subsequently, particles were refined using RELION 1.3, which yielded low resolution envelopes of Pol I-Rrn3 and Pol I-Rrn3-CF complexes at 28 Å and 16 Å resolution, respectively (Figure 20).

### 5.2.3. Preparation of Pol I ITC complexes

Pol I, CF, Rrn3, and TBP were purified separately and reconstituted *in vitro* on a promoter scaffold. *Saccharomyces 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'-CTTGTCTTCAACTGCTTTTCGCATGAAGTACCTCCCACTACTTTTCCTCACACTTGTACTCCATGACTAAACCCCCCTCCCATTACAACTAAAATCTTACT-3'; nontemplate DNA, 5'-AGTAAGATTTTAGTTTGTAAATGGGAGGGGGGGTTTAGTCATGGAGTACAAGTGTGAGGAAAAGTAGTTGGGACAAGTCTTGCATCGTGCAGTTGAAGACAAG-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 minutes, followed by an hour on ice. Stoichiometric 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<sub>2</sub>, 10 μM ZnCl<sub>2</sub>, 5 % v/v glycerol, 10 μM β-mercaptoethanol). The peak fractions (Figure 23) 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).

#### 5.2.4. ITC single-particle cryo-EM and image processing

A 4  $\mu\text{L}$  aliquot of purified sample was applied to a glow-discharged R2/2 UltrAuFoil<sup>TM</sup> 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,  $\sim 60 \text{ e}^-/\text{\AA}^2$  total dose, 1.5'' exposure) were collected at a nominal magnification of 95,000x (1.13  $\text{\AA}/\text{pixel}$ ) 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  $-2.5 \mu\text{m}$  with marginal ( $<0.1 \mu\text{m}$ ) 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,  $\sim 20,000$  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  $\text{\AA}$ . We extracted 2,498,000 particles from 6,000 micrographs with a  $330^2$  pixel box and used them for further processing. Pixels 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 pre-processing. False-positive particles showing very bright dots, which were likely gold particles, were removed by manual inspection and by 3D classification starting with a reference generated from one Pol I molecule of PDB entry 4C2M (Engel et al., 2013) filtered to 40  $\text{\AA}$ . Subsequent classifications and refinements were performed against the reconstruction of the previous round (Figure 23).

During subsequent classification (Figure 23), we merged 508,000 particles that displayed CF density and further classified globally on the presence of Rrn3. The resulting ITC reconstruction showed an overall resolution of 3.4  $\text{\AA}$  (0.143 FSC criteria, Figure 24). This subset was also classified on density between upstream DNA and the stalk, resulting in a sub-set of 29,000 particles. The corresponding ITC reconstruction revealed the A49 tWH domain and reached an overall resolution of 6.9  $\text{\AA}$  (0.143 FSC criteria). For visualization, the density was either filtered to its nominal resolution or additionally sharpened with B-factors in the range of  $-100$  to  $-215 \text{\AA}^2$ . Local resolution (Figure 24) was estimated as described (Plaschka et al., 2015).

### 5.2.5. 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) and included into the model. This subset was adjusted in Coot and refined with Phenix in real space. The upstream DNA revealed a bend which was modelled in 3D-DART (van Dijk and Bonvin, 2009). To account for the uncertainty in sequence assignment, we used repetitive ATGC stretches. Although we included TBP in our sample preparation, we 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 Figure 16. 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 and Cowtan, 2004) and placed in an ITC such that DNA superimposed with DNA in the ITC at the distal upstream PAD contact and contacts a conserved TPSA motif in Rpb5 (Bernecky et al., 2016). We then overlaid 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.

## 5.2.6. 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:

GAGTACAAGTGTGAGGAAAAGTAGTTGGGAGGTA CTTTCATGCGAAAGCAGTTGA  
AGACAAG; Template:

CTTGTCTTCAACTGCTTTCGCATGAAGTACCTCCCAACTACTTTTCCTCACACTTGT  
ACTC. 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., 2013, Plaschka et al., 2016, Sydow et al., 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  $\mu$ M ZnCl<sub>2</sub>, 30% (v/v) glycerol and 15 mM DTT. In a total volume of 10  $\mu$ L, the initiation reaction was prepared on ice comprising 0.1  $\mu$ M Pol I (or Pol II), 5-fold molar excess of CF variants, Rrn3, and/or TPB, 0.1  $\mu$ M dsDNA scaffold and 2  $\mu$ L of 5x RB. The reaction was filled to 8.75  $\mu$ L volume with sterile water and 1.25  $\mu$ L of a NTP mix was added which contained 20  $\mu$ M GTP, 20  $\mu$ M UTP, 20  $\mu$ M CTP, 2  $\mu$ M ATP (Thermo Scientific) and 0.8  $\mu$ M [ $\alpha$ -<sup>32</sup>P] (2.5  $\mu$ Ci/ $\mu$ L; Perkin Elmer). The reaction was well mixed and incubated at 30°C for 30 min. To stop the reaction, 10  $\mu$ L 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 500V 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 1h45min. The gel was incubated on a Phospho-Screen for 3-12 h at -20°C and read out using a Typhoon FLA 9500 (GE Healthcare). For titration experiments, used molar excess steps for transcription factors were 0, 1x, 2.5x, 5x, 10x, 20x and 50x compared to Pol I.

### 5.3. STAR Methods

#### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
NA		
<b>Bacterial and Virus Strains</b>		
<i>Escherichia coli</i> XL1-Blue	Agilent	#200249
<i>Escherichia coli</i> BL21-CodonPlus(DE3)-RIL	Agilent	#230245
<b>Biological Samples</b>		
NA		
<b>Chemicals, Peptides, and Recombinant Proteins</b>		
<i>S. cerevisiae</i> Rrn3 (full length; N-terminal 6xHIS tag)	(Blattner et al., 2011)	NA
<i>S. cerevisiae</i> Rrn7 (full length; N-terminal 6xHIS tag; in complex with Rrn6 and Rrn11)	Recombinant expression; this work	NA
<i>S. cerevisiae</i> Rrn6 (full length; in complex with Rrn7 and Rrn11)	Recombinant expression; this work	NA
<i>S. cerevisiae</i> Rrn11 (full length; C-terminal 6xHIS tag; in complex with Rrn7 and Rrn6)	Recombinant expression; this work	NA
<i>S. cerevisiae</i> TBP (full length; C-terminal 6xHIS tag)	(Plaschka et al., 2016)	NA
<i>S. cerevisiae</i> Rrn7 ΔN40 (ΔN40 Rrn7; N-terminal 6xHIS tag; in complex with Rrn6 and Rrn11)	Recombinant expression; this work	NA
<i>S. cerevisiae</i> Rrn7 ΔN94 (ΔN94 Rrn7; N-terminal 6xHIS tag; in complex with Rrn6 and Rrn11)	Recombinant expression; this work	NA
<i>S. cerevisiae</i> Rrn7 Δ75-94 (Δ75-94 Rrn7; N-terminal 6xHIS tag; in complex with Rrn6 and Rrn11)	Recombinant expression; this work	NA
<i>S. cerevisiae</i> Rrn7 Δ400-420 (Δ400-420 Rrn7; N-terminal 6xHIS tag; in complex with Rrn6 and Rrn11)	Recombinant expression; this work	NA
<i>S. cerevisiae</i> Rrn7 Q406A (Q406A Rrn7; N-terminal 6xHIS tag; in complex with Rrn6 and Rrn11)	Recombinant expression; this work	NA
<i>S. cerevisiae</i> Rrn7 A409P (A409P Rrn7; N-terminal 6xHIS tag; in complex with Rrn6 and Rrn11)	Recombinant expression; this work	NA
<i>S. cerevisiae</i> Rrn7 RRK410AAA (RRK410AAA Rrn7; N-terminal 6xHIS tag; in complex with Rrn6 and Rrn11)	Recombinant expression; this work	NA
<i>S. cerevisiae</i> Rrn7 R411E (R411E Rrn7; N-terminal 6xHIS tag; in complex with Rrn6 and Rrn11)	Recombinant expression; this work	NA
<i>S. cerevisiae</i> Rrn7 R411A (R411A Rrn7; N-terminal 6xHIS tag; in complex with Rrn6 and Rrn11)	Recombinant expression; this work	NA
<i>S. cerevisiae</i> Rrn11 H407P (H407P Rrn11; C-terminal 6xHIS tag; in complex with Rrn6 and Rrn7)	Recombinant expression; this work	NA
<i>S. cerevisiae</i> Rrn11 E351R (E351R Rrn11; C-terminal 6xHIS tag; in complex with Rrn6 and Rrn7)	Recombinant expression; this work	NA
<i>S. cerevisiae</i> Rrn11 SRQ380AAA (SRQ380AAA Rrn11; C-terminal 6xHIS tag; in complex with Rrn6 and Rrn7)	Recombinant expression; this work	NA
<i>S. cerevisiae</i> Rrn11 V404S (V404S Rrn11; C-terminal 6xHIS tag; in complex with Rrn6 and Rrn7)	Recombinant expression; this work	NA

<i>S. cerevisiae</i> Rrn11 KHIH406DDAD (KHIH406DDAD Rrn11; C-terminal 6xHIS tag; in complex with Rrn6 and Rrn7)	Recombinant expression; this work	NA
<i>S. cerevisiae</i> Rrn11 P427G (P427G Rrn11; C-terminal 6xHIS tag; in complex with Rrn6 and Rrn7)	Recombinant expression; this work	NA
<i>S. cerevisiae</i> Rrn11 I431S (I431S Rrn11; C-terminal 6xHIS tag; in complex with Rrn6 and Rrn7)	Recombinant expression; this work	NA
<i>S. cerevisiae</i> Rrn11 Q434A (Q434A Rrn11; C-terminal 6xHIS tag; in complex with Rrn6 and Rrn7)	Recombinant expression; this work	NA
<i>S. cerevisiae</i> Rrn6 L25M (L25M Rrn6; in complex with Rrn11 and Rrn7)	Recombinant expression; this work	NA
<i>S. cerevisiae</i> Rrn7 V212M (V212M Rrn7; N-terminal 6xHIS tag; in complex with Rrn6 and Rrn11)	Recombinant expression; this work	NA
<i>S. cerevisiae</i> Rrn7 I408M (I408M Rrn7; N-terminal 6xHIS tag; in complex with Rrn6 and Rrn11)	Recombinant expression; this work	NA
<i>S. cerevisiae</i> Rrn7 F438M (F438M Rrn7; N-terminal 6xHIS tag; in complex with Rrn6 and Rrn11)	Recombinant expression; this work	NA
<i>S. cerevisiae</i> Rrn11 L73M (L73M; C-terminal 6xHIS tag; in complex with Rrn6 and Rrn7)	Recombinant expression; this work	NA
<i>S. cerevisiae</i> Rrn11 W84M (W84M; C-terminal 6xHIS tag; in complex with Rrn6 and Rrn7)	Recombinant expression; this work	NA
<i>S. cerevisiae</i> Rrn11 L430M (L430M; C-terminal 6xHIS tag; in complex with Rrn6 and Rrn7)	Recombinant expression; this work	NA
<b>Critical Commercial Assays</b>		
NA		
<b>Deposited Data</b>		
Core Factor		PDB 5N5X
Pol I-Rrn3-CF complex (orientation 3)		PDB 5N5Y EMDB-3590
Pol I-Rrn3-CF complex (orientation 2)		PDB 5N5Z EMDB-3591
Pol I-Rrn3-CF complex (orientation 1)		PDB 5N60 EMDB-3592
Pol I ITC		PDB 5N61 EMDB-3593
Pol I ITC (tWH)		EMDB-3594
<b>Experimental Models: Cell Lines</b>		
NA		
<b>Experimental Models: Organisms/Strains</b>		
<i>Saccharomyces cerevisiae</i> CB010 (MAT $\alpha$ ; pep4::HIS3; prb1::LEU2; prc1::HISG; can1; ade2; trp1; ura3; his3; leu2-3; RPA190::RPA190-FLAG-10xHIS-KanMX)	(Engel et al., 2013)	
<b>Oligonucleotides</b>		
Initiation assay template native (dsDNA) GAGTACAAGTGTGAGGAAAAGTAGTTGGGAG GTACTTCATGCGAAAGCAGTTGAAGACAAGT	This work	
Initiation assay template randomized -1 to -10 (dsDNA) GAGTACAAGTGTGAGGAAAAGTAGTTGGGTCT AGCAGTATGCGAAAGCAGTTGAAGACAAGT	This work	
Initiation assay template randomized -11 to -20 (dsDNA) GAGTACAAGTGTGAGGAATCCGATGTAGGAG GTACTTCATGCGAAAGCAGTTGAAGACAAGT	This work	

Initiation assay template randomized -21 to -30 (dsDNA) GAGTACAAAGGCCATTGGAAGTAGTTGGGAG GTA CTT CAT GCGAAAGCAGTTGAAGACAAGT	This work	
Initiation assay template randomized -31 to -38 (dsDNA) GCTACAGAGTGTGAGGAAAAGTAGTTGGGAG GTA CTT CAT GCGAAAGCAGTTGAAGACAAGT	This work	
Initiation assay template +1ATG to +1GTA (dsDNA) GAGTACAAGTGTGAGGAAAAGTAGTTGGGAG GTA CTT CGTACGAAAGCAGTTGAAGACAAGT	This work	
Initiation assay template deleted -1 to -10 (dsDNA) TTTAGTCATGGAGTACAAGTGTGAGGAAAAGT AGTTGGATGCGAAAGCAGTTGAAGACAAGT	This work	
Initiation assay template deleted -11 to -20 (dsDNA) TTTAGTCATGGAGTACAAGTGTGAGGAAGAGG TACTTCATGCGAAAGCAGTTGAAGACAAGT	This work	
Initiation assay template deleted -21 to -30 (dsDNA) TTTAGTCATGGAGTACAAAAGTAGTTGGGAGG TACTTCATGCGAAAGCAGTTGAAGACAAGT	This work	
Initiation assay template deleted -31 to -38 (dsDNA) TTTAGTCAGTGTGAGGAAAAGTAGTTGGGAGG TACTTCATGCGAAAGCAGTTGAAGACAAGT	This work	
Initiation assay template deleted +1ATG (dsDNA) GAGTACAAGTGTGAGGAAAAGTAGTTGGGAG GTA CTT CCGAAAGCAGTTGAAGACAAGT	This work	
ITC template DNA CTTGTCTTCAACTGCTTTCGCATGAAGTACCTC CCA ACTACTTTTCTCACACTTGTACTCCATGA CTAAACCCCCCTCCATTACAACTAAAATC TTACT	This work	
ITC non-template DNA AGTAAGATTTTAGTTTGTAATGGGAGGGGGGG TTTAGTCATGGAGTACAAGTGTGAGGAAAAGT AGTTGGGACAAGTGCTTGCATCGTGCAGTTGA AGACAAG	This work	
ITC RNA AUGCGA	This work	
<b>Recombinant DNA</b>		
Plasmid pET28b-Rrn7 (N-terminal 6xHIS tag; full length <i>S. cerevisiae</i> Rrn7)	This work	
Plasmid pET21a-Rrn6-Rrn11 (C-terminal 6xHIS tag on Rrn11; full length <i>S. cerevisiae</i> sequences devided by a spacer including a second ribosome binding site)	This work	
Plasmid pET28b-Rrn3 (N-terminal 6xHIS tag; full length <i>S. cerevisiae</i> Rrn3)	(Blattner et al., 2011)	
Plasmid pOPINE-TBP (C-terminal 6xHIS tag; full length)	(Plaschka et al., 2016)	
Plasmid pET28b-Rrn3 (S145D)	(Blattner et al., 2011)	
Plasmid pET28b-Rrn3 (S185D)	(Blattner et al., 2011)	
Plasmid pET28b-Rrn7 ΔN40 (N-terminal 6xHIS tag; residues 41 to 514 of <i>S. cerevisiae</i> Rrn7)	This work	
Plasmid pET28b-Rrn7 ΔN94 (N-terminal 6xHIS tag; residues 95 to 514 of <i>S. cerevisiae</i> Rrn7)	This work	

Plasmid pET28b-Rrn7 $\Delta$ 75-94 (N-terminal 6xHIS tag; residues 75 to 94 removed from full length <i>S. cerevisiae</i> Rrn7)	This work	
Plasmid pET28b-Rrn7 $\Delta$ 400-420 (N-terminal 6xHIS tag; residues 400-420 removed from full length <i>S. cerevisiae</i> Rrn7)	This work	
Plasmid pET28b-Rrn7 Q406A (N-terminal 6xHIS tag; residue Q406 mutated to A in full length <i>S. cerevisiae</i> Rrn7)	This work	
Plasmid pET28b-Rrn7 A409P (N-terminal 6xHIS tag; residue A409 mutated to P in full length <i>S. cerevisiae</i> Rrn7)	This work	
Plasmid pET28b-Rrn7 RRK410AAA (N-terminal 6xHIS tag; residues R410, R411 and K412 mutated to A in full length <i>S. cerevisiae</i> Rrn7)	This work	
Plasmid pET28b-Rrn7 R411E (N-terminal 6xHIS tag; residue R411 mutated to E in full length <i>S. cerevisiae</i> Rrn7)	This work	
Plasmid pET28b-Rrn7 (N-terminal 6xHIS tag; residue R411 mutated to A in full length <i>S. cerevisiae</i> Rrn7)	This work	
Plasmid pET21a-Rrn6-Rrn11 H407P (C-terminal 6xHIS tag on Rrn11; full length <i>S. cerevisiae</i> Rrn6; Rrn11 residue H407 mutated to P)	This work	
Plasmid pET21a-Rrn6-Rrn11 E351R (C-terminal 6xHIS tag on Rrn11; full length <i>S. cerevisiae</i> Rrn6; Rrn11 residue E351 mutated to R)	This work	
Plasmid pET21a-Rrn6-Rrn11 SRQ380AAA (C-terminal 6xHIS tag on Rrn11; full length <i>S. cerevisiae</i> Rrn6; Rrn11 residues S380, R381 and Q382 mutated to A)	This work	
Plasmid pET21a-Rrn6-Rrn11 V404S (C-terminal 6xHIS tag on Rrn11; full length <i>S. cerevisiae</i> Rrn6; Rrn11 residue V404 mutated to S)	This work	
Plasmid pET21a-Rrn6-Rrn11 KHIH406DDAD (C-terminal 6xHIS tag on Rrn11; full length <i>S. cerevisiae</i> Rrn6; Rrn11 residues K406, H407, I408 and D409 mutated to the sequence DDAD)	This work	
Plasmid pET21a-Rrn6-Rrn11 P427G (C-terminal 6xHIS tag on Rrn11; full length <i>S. cerevisiae</i> Rrn6; Rrn11 residue P427 mutated to G)	This work	
Plasmid pET21a-Rrn6-Rrn11 I431S (C-terminal 6xHIS tag on Rrn11; full length <i>S. cerevisiae</i> Rrn6; Rrn11 residue I431 mutated to S)	This work	
Plasmid pET21a-Rrn6-Rrn11 Q434A (C-terminal 6xHIS tag on Rrn11; full length <i>S. cerevisiae</i> Rrn6; Rrn11 residue Q434 mutated to A)	This work	
Plasmid pET21a-Rrn6(L25M)-Rrn11 (C-terminal 6xHIS tag on Rrn11; Rrn6 residue L25 mutated to M)	This work	
Plasmid pET21a-Rrn6-Rrn11 L73M (C-terminal 6xHIS tag on Rrn11; full length <i>S. cerevisiae</i> Rrn6; Rrn11 residue W84 mutated to M)	This work	
Plasmid pET21a-Rrn6-Rrn11 W84M (C-terminal 6xHIS tag on Rrn11; full length <i>S. cerevisiae</i> Rrn6; Rrn11 residue W84 mutated to M)	This work	

Plasmid pET21a-Rrn6-Rrn11 L430M (C-terminal 6xHIS tag on Rrn11; full length <i>S. cerevisiae</i> Rrn6; Rrn11 residue L430 mutated to M)	This work	
Plasmid pET28b-Rrn7 (N-terminal 6xHIS tag; residue V212 mutated to M in full length <i>S. cerevisiae</i> Rrn7)	This work	
Plasmid pET28b-Rrn7 (N-terminal 6xHIS tag; residue I408 mutated to M in full length <i>S. cerevisiae</i> Rrn7)	This work	
Plasmid pET28b-Rrn7 (N-terminal 6xHIS tag; residue F438 mutated to M in full length <i>S. cerevisiae</i> Rrn7)	This work	
<b>Software and Algorithms</b>		
X-ray Detector Software (XDS)	(Kabsch, 2010)	<a href="http://xds.mpimf-heidelberg.mpg.de/">http://xds.mpimf-heidelberg.mpg.de/</a>
SHELXCD	(Sheldrick, 2010)	<a href="http://shelx.uni-ac.gwdg.de/SHELX/">http://shelx.uni-ac.gwdg.de/SHELX/</a>
Phenix suite	(Adams et al., 2010)	<a href="https://www.phenix-online.org/">https://www.phenix-online.org/</a>
CCP4 suite	(Winn et al., 2011)	<a href="http://www.ccp4.ac.uk/">http://www.ccp4.ac.uk/</a>
COOT v0.8.3	(Emsley and Cowtan, 2004)	<a href="https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/">https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/</a>
Relion v1.4	(Scheres, 2012)	<a href="https://www2.mrc-lmb.cam.ac.uk/relion/index.php/Main_Page">https://www2.mrc-lmb.cam.ac.uk/relion/index.php/Main_Page</a>
EMAN v2.1	(Tang et al., 2007)	<a href="http://blake.bcm.edu/emanwiki/EMAN2">http://blake.bcm.edu/emanwiki/EMAN2</a>
Unblur	(Brilot et al., 2012)	<a href="http://grigoriefflab.janelia.org/unblur">http://grigoriefflab.janelia.org/unblur</a>
CTFFIND4	(Rhou and Grigorieff, 2015)	<a href="http://grigoriefflab.janelia.org/ctffind4">http://grigoriefflab.janelia.org/ctffind4</a>
3D-DART	(van Dijk and Bonvin, 2009)	<a href="http://haddock.science.uu.nl/dna/dna.php">http://haddock.science.uu.nl/dna/dna.php</a>
UCSF Chimera v1.10.2	(Pettersen et al., 2004)	<a href="https://www.cgl.ucsf.edu/chimera/download.html">https://www.cgl.ucsf.edu/chimera/download.html</a>
DNA Curvature Analysis	(Bolshoy et al., 1991, Goodsell and Dickerson, 1994)	<a href="http://www.lfd.uci.edu/~gohlke/dnacurve/">http://www.lfd.uci.edu/~gohlke/dnacurve/</a>
PDBeFold	(Krissinel and Henrick, 2004)	<a href="http://www.ebi.ac.uk/msd-srv/ssm/">http://www.ebi.ac.uk/msd-srv/ssm/</a>
<b>Other</b>		
NA		

## 5.4. Phenix Refinement

### 5.4.1. Phenix input script

```
phenix.refine \  
nativeCF_rfree.mtz \  
build.pdb \  
strategy=rigid_body+individual_sites+individual_adp+tls \  
refinement.input.xray_data.r_free_flags.label=FreeRflag \  
rigid_bodies.params \  
ncs_groups.params \  
tlsmd.params \  
sec_str_manual.pdb_ss.eff \  
ncs_search.enabled=True ncs.type=cartesian \  
check_rotamer_consistency=False \  
disulfide_distance_cutoff=0.5 \  
adp.convert_to_iso=true \  
adp.set_b_iso=60 \  
refinement.target_weights.wxc_scale=0.25 \  
refinement.ncs.excessive_distance_limit=None \  
main.number_of_macro_cycles=6 \  

```

### 5.4.2. NCS group parameters

```
refinement.pdb_interpretation.ncs_group {  
  reference = chain B and resseq 90:244  
  selection = chain E and resseq 90:244  
  selection = chain H and resseq 90:244  
  selection = chain K and resseq 90:244  
  selection = chain N and resseq 90:244  
  selection = chain Q and resseq 90:244  
}  
refinement.pdb_interpretation.ncs_group {  
  reference = chain B and resseq 245:513  
  selection = chain E and resseq 245:513  
  selection = chain H and resseq 245:513  
  selection = chain K and resseq 245:513  
  selection = chain N and resseq 245:513  
  selection = chain Q and resseq 245:513  
}  
refinement.pdb_interpretation.ncs_group {  
  reference = chain A and resseq 560:650  
  selection = chain D and resseq 560:650  
  selection = chain G and resseq 560:650  
  selection = chain J and resseq 560:650  
  selection = chain M and resseq 560:650  
  selection = chain P and resseq 560:650  
}  
refinement.pdb_interpretation.ncs_group {  
  reference = chain A and resseq 651:780  
  selection = chain D and resseq 651:780  
  selection = chain G and resseq 651:780  
  selection = chain J and resseq 651:780  
  selection = chain M and resseq 651:780  
  selection = chain P and resseq 651:780  
}  
refinement.pdb_interpretation.ncs_group {  
  reference = chain C and resseq 1:197  
  selection = chain F and resseq 1:197  
  selection = chain I and resseq 1:197  
  selection = chain L and resseq 1:197  
  selection = chain O and resseq 1:197  
  selection = chain R and resseq 1:197  
}  
refinement.pdb_interpretation.ncs_group {  
  reference = chain C and resseq 198:440  
  selection = chain F and resseq 198:440  
  selection = chain I and resseq 198:440  
  selection = chain L and resseq 198:440  
  selection = chain O and resseq 198:440  
  selection = chain R and resseq 198:440  
}  
refinement.pdb_interpretation.ncs_group {  
  reference = chain A and resseq 1:560  
  selection = chain D and resseq 1:560  
}
```

```

selection = chain G and resseq 1:560
selection = chain J and resseq 1:560
selection = chain M and resseq 1:560
selection = chain P and resseq 1:560
}

```

### 5.4.3. Rigid body parameters

```

refinement.refine.sites {
rigid_body = chain A and resseq 1:560
rigid_body = chain A and resseq 561:650
rigid_body = chain A and resseq 651:780
rigid_body = chain D and resseq 1:560
rigid_body = chain D and resseq 561:650
rigid_body = chain D and resseq 651:780
rigid_body = chain G and resseq 1:560
rigid_body = chain G and resseq 561:650
rigid_body = chain G and resseq 651:780
rigid_body = chain J and resseq 1:560
rigid_body = chain J and resseq 561:650
rigid_body = chain J and resseq 651:780
rigid_body = chain M and resseq 1:560
rigid_body = chain M and resseq 561:650
rigid_body = chain M and resseq 651:780
rigid_body = chain P and resseq 1:560
rigid_body = chain P and resseq 561:650
rigid_body = chain P and resseq 651:780

rigid_body = chain B and resseq 90:244
rigid_body = chain B and resseq 245:513
rigid_body = chain E and resseq 90:244
rigid_body = chain E and resseq 245:513
rigid_body = chain H and resseq 90:244
rigid_body = chain H and resseq 245:513
rigid_body = chain K and resseq 90:244
rigid_body = chain K and resseq 245:513
rigid_body = chain N and resseq 90:244
rigid_body = chain N and resseq 245:513
rigid_body = chain Q and resseq 90:244
rigid_body = chain Q and resseq 245:513

rigid_body = chain C and resseq 1:197
rigid_body = chain C and resseq 198:440
rigid_body = chain F and resseq 1:197
rigid_body = chain F and resseq 198:440
rigid_body = chain I and resseq 1:197
rigid_body = chain I and resseq 198:440
rigid_body = chain L and resseq 1:197
rigid_body = chain L and resseq 198:440
rigid_body = chain O and resseq 1:197
rigid_body = chain O and resseq 198:440
rigid_body = chain R and resseq 1:197
rigid_body = chain R and resseq 198:440
}

```

### 5.4.4. TLS parameters

```

refinement.refine.adp {
    tls = chain A or \
        chain B or \
        chain C
    tls = chain D or \
        chain E or \
        chain F
    tls = chain G or \
        chain H or \
        chain I
    tls = chain J or \
        chain K or \
        chain L
    tls = chain M or \
        chain N or \
        chain O
    tls = chain P or \
        chain Q or \
        chain R
}

```

## 5.4.5. Secondary structure definition

Exemplary secondary structure definition of one CF subunit for refinement in Phenix.

```
# These parameters are suitable for use in e.g. phenix.real_space_refine
# or geometry_minimization. To use them in phenix.refine add
# 'refinement.' if front of pdb_interpretation.
refinement.pdb_interpretation {
  secondary_structure {
    enabled = True
    protein {
      helix {
        serial_number = 1
        helix_identifier = "1"
        selection = chain 'A' and resid 573 through 579
      }
      helix {
        serial_number = 2
        helix_identifier = "2"
        selection = chain 'A' and resid 583 through 613
      }
      helix {
        serial_number = 3
        helix_identifier = "3"
        selection = chain 'A' and resid 617 through 643
      }
      helix {
        serial_number = 4
        helix_identifier = "4"
        selection = chain 'A' and resid 672 through 686
      }
      helix {
        serial_number = 5
        helix_identifier = "5"
        selection = chain 'A' and resid 697 through 704
      }
      helix {
        serial_number = 6
        helix_identifier = "6"
        selection = chain 'A' and resid 712 through 724
      }
      helix {
        serial_number = 7
        helix_identifier = "7"
        selection = chain 'A' and resid 729 through 744
      }
      helix {
        serial_number = 8
        helix_identifier = "8"
        selection = chain 'A' and resid 752 through 764
      }
      helix {
        serial_number = 9
        helix_identifier = "9"
        selection = chain 'A' and resid 768 through 774
      }
      helix {
        serial_number = 10
        helix_identifier = "10"
        selection = chain 'B' and resid 100 through 124
      }
      helix {
        serial_number = 11
        helix_identifier = "11"
        selection = chain 'B' and resid 127 through 147
      }
      helix {
        serial_number = 12
        helix_identifier = "12"
        selection = chain 'B' and resid 158 through 172
      }
      helix {
        serial_number = 13
        helix_identifier = "13"
        selection = chain 'B' and resid 178 through 187
      }
    }
  }
}
```

```

helix {
  serial_number = 14
  helix_identifier = "14"
  selection = chain 'B' and resid 210 through 218
}
helix {
  serial_number = 15
  helix_identifier = "15"
  selection = chain 'B' and resid 223 through 237
}
helix {
  serial_number = 16
  helix_identifier = "16"
  selection = chain 'B' and resid 249 through 260
}
helix {
  serial_number = 17
  helix_identifier = "17"
  selection = chain 'B' and resid 264 through 278
}
helix {
  serial_number = 18
  helix_identifier = "18"
  selection = chain 'B' and resid 302 through 320
}
helix {
  serial_number = 19
  helix_identifier = "19"
  selection = chain 'B' and resid 329 through 340
}
helix {
  serial_number = 20
  helix_identifier = "20"
  selection = chain 'B' and resid 345 through 356
}
helix {
  serial_number = 21
  helix_identifier = "21"
  selection = chain 'B' and resid 362 through 368
}
helix {
  serial_number = 22
  helix_identifier = "22"
  selection = chain 'B' and resid 371 through 385
}
helix {
  serial_number = 23
  helix_identifier = "23"
  selection = chain 'B' and resid 405 through 416
}
helix {
  serial_number = 24
  helix_identifier = "24"
  selection = chain 'B' and resid 438 through 449
}
helix {
  serial_number = 25
  helix_identifier = "25"
  selection = chain 'B' and resid 471 through 491
}
helix {
  serial_number = 26
  helix_identifier = "26"
  selection = chain 'B' and resid 495 through 513
}
helix {
  serial_number = 27
  helix_identifier = "27"
  selection = chain 'C' and resid 11 through 36
}
helix {
  serial_number = 28
  helix_identifier = "28"
  selection = chain 'C' and resid 74 through 84
}
helix {
  serial_number = 29
  helix_identifier = "29"
  selection = chain 'C' and resid 159 through 167
}

```

```

}
helix {
  serial_number = 30
  helix_identifier = "30"
  selection = chain 'C' and resid 170 through 179
}
helix {
  serial_number = 31
  helix_identifier = "31"
  selection = chain 'C' and resid 182 through 194
}
helix {
  serial_number = 32
  helix_identifier = "32"
  selection = chain 'C' and resid 201 through 204
}
helix {
  serial_number = 33
  helix_identifier = "33"
  selection = chain 'C' and resid 208 through 226
}
helix {
  serial_number = 34
  helix_identifier = "34"
  selection = chain 'C' and resid 229 through 241
}
helix {
  serial_number = 35
  helix_identifier = "35"
  selection = chain 'C' and resid 248 through 259
}
helix {
  serial_number = 36
  helix_identifier = "36"
  selection = chain 'C' and resid 267 through 278
}
helix {
  serial_number = 37
  helix_identifier = "37"
  selection = chain 'C' and resid 307 through 324
}
helix {
  serial_number = 38
  helix_identifier = "38"
  selection = chain 'C' and resid 345 through 355
}
helix {
  serial_number = 39
  helix_identifier = "39"
  selection = chain 'C' and resid 362 through 377
}
helix {
  serial_number = 40
  helix_identifier = "40"
  selection = chain 'C' and resid 401 through 420
}
helix {
  serial_number = 41
  helix_identifier = "41"
  selection = chain 'C' and resid 428 through 440
}
helix {
  serial_number = 42
  helix_identifier = "42"
  selection = chain 'A' and resid 20 through 26
}
sheet {
  first_strand = chain 'A' and resid 60 through 65
  sheet_id = " A"
  strand {
    selection = chain 'A' and resid 547 through 553
    sense = parallel *antiparallel unknown
    bond_start_current = chain 'A' and resid 549 and name N
    bond_start_previous = chain 'A' and resid 63 and name O
  }
  strand {
    selection = chain 'A' and resid 534 through 541
    sense = parallel *antiparallel unknown
    bond_start_current = chain 'A' and resid 539 and name N
  }
}

```

```

    bond_start_previous = chain 'A' and resid 548 and name O
  }
  strand {
    selection = chain 'A' and resid 506 through 511
    sense = parallel *antiparallel unknown
    bond_start_current = chain 'A' and resid 509 and name N
    bond_start_previous = chain 'A' and resid 538 and name O
  }
}
sheet {
  first_strand = chain 'A' and resid 192 through 196
  sheet_id = " B"
  strand {
    selection = chain 'A' and resid 208 through 213
    sense = parallel *antiparallel unknown
    bond_start_current = chain 'A' and resid 211 and name N
    bond_start_previous = chain 'A' and resid 193 and name O
  }
  strand {
    selection = chain 'A' and resid 220 through 226
    sense = parallel *antiparallel unknown
    bond_start_current = chain 'A' and resid 224 and name N
    bond_start_previous = chain 'A' and resid 210 and name O
  }
  strand {
    selection = chain 'A' and resid 238 through 242
    sense = parallel *antiparallel unknown
    bond_start_current = chain 'A' and resid 240 and name N
    bond_start_previous = chain 'A' and resid 221 and name O
  }
}
sheet {
  first_strand = chain 'A' and resid 247 through 251
  sheet_id = " C"
  strand {
    selection = chain 'A' and resid 264 through 268
    sense = parallel *antiparallel unknown
    bond_start_current = chain 'A' and resid 267 and name N
    bond_start_previous = chain 'A' and resid 248 and name O
  }
  strand {
    selection = chain 'A' and resid 271 through 279
    sense = parallel *antiparallel unknown
    bond_start_current = chain 'A' and resid 275 and name N
    bond_start_previous = chain 'A' and resid 264 and name O
  }
  strand {
    selection = chain 'A' and resid 284 through 293
    sense = parallel *antiparallel unknown
    bond_start_current = chain 'A' and resid 289 and name N
    bond_start_previous = chain 'A' and resid 274 and name O
  }
}
sheet {
  first_strand = chain 'A' and resid 301 through 305
  sheet_id = " D"
  strand {
    selection = chain 'A' and resid 316 through 320
    sense = parallel *antiparallel unknown
    bond_start_current = chain 'A' and resid 317 and name N
    bond_start_previous = chain 'A' and resid 304 and name O
  }
  strand {
    selection = chain 'A' and resid 323 through 327
    sense = parallel *antiparallel unknown
    bond_start_current = chain 'A' and resid 327 and name N
    bond_start_previous = chain 'A' and resid 316 and name O
  }
  strand {
    selection = chain 'A' and resid 348 through 351
    sense = parallel *antiparallel unknown
    bond_start_current = chain 'A' and resid 349 and name N
    bond_start_previous = chain 'A' and resid 324 and name O
  }
}
sheet {
  first_strand = chain 'A' and resid 361 through 365
  sheet_id = " E"
  strand {

```



## 6. 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. & ZWART, P. H. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*, 66, 213-21.
- APRIKIAN, P., MOOREFIELD, B. & REEDER, R. H. 2001. New model for the yeast RNA polymerase I transcription cycle. *Mol Cell Biol*, 21, 4847-55.
- ARMACHE, K. J., MITTERWEGER, S., MEINHART, A. & CRAMER, P. 2005. Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J Biol Chem*, 280, 7131-4.
- BAKER, N. A., SEPT, D., JOSEPH, S., HOLST, M. J. & MCCAMMON, J. A. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A*, 98, 10037-41.
- BEDWELL, G. J., APPLING, F. D., ANDERSON, S. J. & SCHNEIDER, D. A. 2012. Efficient transcription by RNA polymerase I using recombinant core factor. *Gene*, 492, 94-9.
- BERNECKY, C., HERZOG, F., BAUMEISTER, W., PLITZKO, J. M. & CRAMER, P. 2016. Structure of transcribing mammalian RNA polymerase II. *Nature*, 529, 551-4.
- BISCHLER, N., BRINO, L., CARLES, C., RIVA, M., TSCHOCHNER, H., MALLOUH, V. & SCHULTZ, P. 2002. Localization of the yeast RNA polymerase I-specific subunits. *EMBO J*, 21, 4136-44.
- BLATCH, G. L. & LASSLE, M. 1999. The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *Bioessays*, 21, 932-9.
- BLATTNER, C., JENNEBACH, S., HERZOG, F., MAYER, A., CHEUNG, A. C., WITTE, G., LORENZEN, K., HOPFNER, K. P., HECK, A. J., AEBERSOLD, R. & CRAMER, P. 2011. Molecular basis of Rrn3-regulated RNA polymerase I initiation and cell growth. *Genes Dev*, 25, 2093-105.
- BODEM, J., DOBREVA, G., HOFFMANN-ROHRER, U., IBEN, S., ZENTGRAF, H., DELIUS, H., VINGRON, M. & GRUMMT, I. 2000. TIF-IA, the factor mediating growth-dependent control of ribosomal RNA synthesis, is the mammalian homolog of yeast Rrn3p. *EMBO Rep*, 1, 171-5.
- BOLSHOY, A., MCNAMARA, P., HARRINGTON, R. E. & TRIFONOV, E. N. 1991. Curved DNA without A-A: experimental estimation of all 16 DNA wedge angles. *Proc Natl Acad Sci U S A*, 88, 2312-6.
- BRILOT, A. F., CHEN, J. Z., CHENG, A., PAN, J., HARRISON, S. C., POTTER, C. S., CARRAGHER, B., HENDERSON, R. & GRIGORIEFF, N. 2012. Beam-induced motion of vitrified specimen on holey carbon film. *J Struct Biol*, 177, 630-7.
- BURATOWSKI, S., HAHN, S., GUARENTE, L. & SHARP, P. A. 1989. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell*, 56, 549-61.

- BUSHNELL, D. A., WESTOVER, K. D., DAVIS, R. E. & KORNBERG, R. D. 2004. Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms. *Science*, 303, 983-8.
- CAVANAUGH, A. H., EVANS, A. & ROTHBLUM, L. I. 2008. Mammalian Rrn3 Is Required for the formation of a Transcription Competent Preinitiation Complex Containing RNA Polymerase I. *Gene expression*, 14, 131-147.
- CAVANAUGH AH, H.-L. I., HU Q, DUNDR M, SMINK T, MISTELI T, ROTHBLUM LI 2002. Rrn3 phosphorylation is a regulatory checkpoint for ribosome biogenesis. *J Biol Chem*, 27423–27432.
- 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. & EBRIGHT, R. H. 2012. Opening and closing of the bacterial RNA polymerase clamp. *Science*, 337, 591-5.
- CHEN, H. T. & 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, 169-80.
- CHEUNG, A. C. & CRAMER, P. 2011. Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature*, 471, 249-53.
- CINGOLANI, G., PETOSA, C., WEIS, K. & MULLER, C. W. 1999. Structure of importin-[beta] bound to the IBB domain of importin-[alpha]. *Nature*, 399, 221-229.
- CLAYPOOL, J. A., FRENCH, S. L., JOHZUKA, K., ELIASON, K., VU, L., DODD, J. A., BEYER, A. L. & NOMURA, M. 2004. Tor pathway regulates Rrn3p-dependent recruitment of yeast RNA polymerase I to the promoter but does not participate in alteration of the number of active genes. *Mol Biol Cell*, 15, 946-56.
- CLEMENTE-BLANCO, A., MAYAN-SANTOS, M., SCHNEIDER, D. A., MACHIN, F., JARMUZ, A., TSCHOCHNER, H. & ARAGON, L. 2009. Cdc14 inhibits transcription by RNA polymerase I during anaphase. *Nature*, 458, 219-22.
- COMAI, L., TANESE, N. & TJIAN, R. 1992. The TATA-binding protein and associated factors are integral components of the RNA polymerase I transcription factor, SL1. *Cell*, 68, 965-76.
- CONTI E, K. J. 2000. Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin  $\alpha$ . *Structure*, 329–338.
- COWTAN, K. 2010. Recent developments in classical density modification. *Acta Crystallogr D Biol Crystallogr*, 66, 470-8.
- CRAMER, P., ARMACHE, K. J., BAUMLI, S., BENKERT, S., BRUECKNER, F., BUCHEN, C., DAMSMA, G. E., DENGL, S., GEIGER, S. R., JASIAK, A. J., JAWHARI, A., JENNEBACH, S., KAMENSKI, T., KETTENBERGER, H., KUHN, C. D., LEHMANN, E., LEIKE, K., SYDOW, J. F. & VANNINI, A. 2008. Structure of eukaryotic RNA polymerases. *Annu Rev Biophys*, 37, 337-52.
- CRAMER, P., BUSHNELL, D. A. & KORNBERG, R. D. 2001. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science*, 292, 1863-76.
- DENISSOV, S., VAN DRIEL, M., VOIT, R., HEKKELMAN, M., HULSEN, T., HERNANDEZ, N., GRUMMT, I., WEHRENS, R. & STUNNENBERG, H. 2007. Identification of novel functional TBP-binding sites and general factor repertoires. *EMBO J*, 26, 944-54.

- 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. & 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, 1418-30.
- DRYGIN, D., RICE, W. G. & GRUMMT, I. 2010. The RNA polymerase I transcription machinery: an emerging target for the treatment of cancer. *Annu Rev Pharmacol Toxicol*, 50, 131-56.
- DRYGIN, D., RICE, W.G., GRUMMT, I. 2010. The RNA polymerase I transcription machinery: an emerging target for the treatment of cancer. *Annu Rev Pharmacol Toxicol*, 131-156.
- EMSLEY, P. & COWTAN, K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*, 60, 2126-32.
- ENGEL, C., GUBBEY, T., NEYER, S., SAINSBURY, S., OBERTHUER, C., BAEJEN, C., BERNECKY, C. & CRAMER, P. 2017. Structural Basis of RNA Polymerase I Transcription Initiation. *Cell*, 169, 120-131.e22.
- ENGEL, C., PLITZKO, J. & CRAMER, P. 2016. RNA polymerase I-Rrn3 complex at 4.8 Å resolution. *Nat Commun*, 7, 12129.
- ENGEL, C., SAINSBURY, S., CHEUNG, A. C., KOSTREWA, D. & CRAMER, P. 2013. RNA polymerase I structure and transcription regulation. *Nature*, 502, 650-5.
- EVANS, P. 2006. Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr*, 62, 72-82.
- FATH, S., MILKEREIT, P., PEYROCHE, G., RIVA, M., CARLES, C. & TSCHOCHNER, H. 2001. Differential roles of phosphorylation in the formation of transcriptional active RNA polymerase I. *Proc Natl Acad Sci U S A*, 98, 14334-9.
- FEKLISTOV, A. & DARST, S. A. 2011. Structural basis for promoter-10 element recognition by the bacterial RNA polymerase sigma subunit. *Cell*, 147, 1257-69.
- FERNANDEZ-TORNERO, C., MORENO-MORCILLO, M., RASHID, U. J., TAYLOR, N. M., RUIZ, F. M., GRUENE, T., LEGRAND, P., STEUERWALD, U. & MULLER, C. W. 2013. Crystal structure of the 14-subunit RNA polymerase I. *Nature*, 502, 644-9.
- GASTEIGER, E., GATTIKER, A., HOOGLAND, C., IVANYI, I., APPEL, R. D. & BAIROCH, A. 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*, 31, 3784-8.
- GEIGER, S. R., KUHN, C. D., LEIDIG, C., RENKAWITZ, J. & CRAMER, P. 2008. Crystallization of RNA polymerase I subcomplex A14/A43 by iterative prediction, probing and removal of flexible regions. *Acta Crystallogr Sect F Struct Biol Cryst Commun*, 64, 413-8.
- GEIGER, S. R., LORENZEN, K., SCHREIECK, A., HANECKER, P., KOSTREWA, D., HECK, A. J. & CRAMER, P. 2010. RNA polymerase I contains a TFIIF-related DNA-binding subcomplex. *Mol Cell*, 39, 583-94.
- GERBER, J., REITER, A., STEINBAUER, R., JAKOB, S., KUHN, C.-D., CRAMER, P., GRIESENBECK, J., MILKEREIT, P. & TSCHOCHNER, H. 2008. Site specific phosphorylation of yeast RNA polymerase I. *Nucleic Acids Research*, 36, 793-802.
- GOODSELL, D. S. & DICKERSON, R. E. 1994. Bending and curvature calculations in B-DNA. *Nucleic Acids Res*, 22, 5497-503.

- GORSKI, J. J., PATHAK, S., PANOV, K., KASCIUKOVIC, T., PANOVA, T., RUSSELL, J. & ZOMERDIJK, J. C. 2007. A novel TBP-associated factor of SL1 functions in RNA polymerase I transcription. *EMBO J*, 26, 1560-8.
- GOUGE, J., SATIA, K., GUTHERTZ, N., WIDYA, M., THOMPSON, A. J., COUSIN, P., DERGAI, O., HERNANDEZ, N. & VANNINI, A. 2015. Redox Signaling by the RNA Polymerase III TFIIB-Related Factor Brf2. *Cell*, 163, 1375-87.
- GRUMMT, I. 2003. Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus. *Genes Dev*, 17, 1691-702.
- GRUMMT, I. & VOIT, R. 2010. Linking rDNA transcription to the cellular energy supply. *Cell Cycle*, 9, 225-226.
- GRUNBERG, S. & HAHN, S. 2013. Structural insights into transcription initiation by RNA polymerase II. *Trends Biochem Sci*, 38, 603-11.
- GRUNBERG, S., WARFIELD, L. & HAHN, S. 2012. Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nat Struct Mol Biol*, 19, 788-96.
- HE, Y., YAN, C., FANG, J., INOUE, C., TJIAN, R., IVANOV, I. & NOGALES, E. 2016. Near-atomic resolution visualization of human transcription promoter opening. *Nature*, 533, 359-65.
- HOPPE, S., BIERHOFF, H., CADO, I., WEBER, A., TIEBE, M., GRUMMT, I. & VOIT, R. 2009. AMP-activated protein kinase adapts rRNA synthesis to cellular energy supply. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 17781-17786.
- HU, P., WU, S., SUN, Y., YUAN, C. C., KOBAYASHI, R., MYERS, M. P. & HERNANDEZ, N. 2002. Characterization of human RNA polymerase III identifies orthologues for *Saccharomyces cerevisiae* RNA polymerase III subunits. *Mol Cell Biol*, 22, 8044-55.
- JASIAK, A. J., ARMACHE, K. J., MARTENS, B., JANSEN, R. P. & CRAMER, P. 2006. Structural biology of RNA polymerase III: subcomplex C17/25 X-ray structure and 11 subunit enzyme model. *Mol Cell*, 23, 71-81.
- JENNEBACH, S., HERZOG, F., AEBERSOLD, R. & CRAMER, P. 2012. Crosslinking-MS analysis reveals RNA polymerase I domain architecture and basis of rRNA cleavage. *Nucleic Acids Res*, 40, 5591-601.
- KABSCH, W. 2010. Xds. *Acta Crystallogr D Biol Crystallogr*, 66, 125-32.
- KASSAVETIS, G. A., PRAKASH, P. & SHIM, E. 2010. The C53/C37 Subcomplex of RNA Polymerase III Lies Near the Active Site and Participates in Promoter Opening. *The Journal of Biological Chemistry*, 285, 2695-2706.
- 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. & STARK, H. 2008. GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat Methods*, 5, 53-5.
- KEENER, J., DODD, J. A., LALO, D. & 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, 13458-62.

- KEENER, J., JOSAITIS, C. A., DODD, J. A. & NOMURA, M. 1998. Reconstitution of yeast RNA polymerase I transcription in vitro from purified components. TATA-binding protein is not required for basal transcription. *J Biol Chem*, 273, 33795-802.
- KEYS, D. A., LEE, B. S., DODD, J. A., NGUYEN, T. T., VU, L., FANTINO, E., BURSON, L. M., NOGI, Y. & NOMURA, M. 1996. Multiprotein transcription factor UAF interacts with the upstream element of the yeast RNA polymerase I promoter and forms a stable preinitiation complex. *Genes Dev*, 10, 887-903.
- KIM, T. K., EBRIGHT, R. H. & REINBERG, D. 2000. Mechanism of ATP-dependent promoter melting by transcription factor IIIH. *Science*, 288, 1418-22.
- KIM, Y., GEIGER, J. H., HAHN, S. & SIGLER, P. B. 1993. Crystal structure of a yeast TBP/TATA-box complex. *Nature*, 365, 512-20.
- KNUTSON, B. A. & HAHN, S. 2011. Yeast Rrn7 and human TAF1B are TFIIB-related RNA polymerase I general transcription factors. *Science*, 333, 1637-40.
- KNUTSON, B. A. & HAHN, S. 2013. TFIIB-related factors in RNA polymerase I transcription. *Biochim Biophys Acta*, 1829, 265-73.
- KNUTSON, B. A., LUO, J., RANISH, J. & HAHN, S. 2014. Architecture of the *Saccharomyces cerevisiae* RNA polymerase I Core Factor complex. *Nat Struct Mol Biol*, 21, 810-6.
- KOSTREWA, D., ZELLER, M. E., ARMACHE, K. J., SEIZL, M., LEIKE, K., THOMM, M. & CRAMER, P. 2009. RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature*, 462, 323-30.
- KOWNIN, P., BATEMAN, E. & PAULE, M. R. 1987. Eukaryotic RNA polymerase I promoter binding is directed by protein contacts with transcription initiation factor and is DNA sequence-independent. *Cell*, 50, 693-9.
- KOWNIN, P., BATEMAN, E. & PAULE, M. R. 1988. Effects of single-base substitutions within the *Acanthamoeba castellanii* rRNA promoter on transcription and on binding of transcription initiation factor and RNA polymerase I. *Mol Cell Biol*, 8, 747-53.
- KRESSLER, D., HURT, E. & BASSLER, J. 2010. Driving ribosome assembly. *Biochim Biophys Acta*, 1803, 673-83.
- KRISSINEL, E. & HENRICK, K. 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr*, 60, 2256-68.
- KUHN, C. D., GEIGER, S. R., BAUMLI, S., GARTMANN, M., GERBER, J., JENNEBACH, S., MIELKE, T., TSCHOCHNER, H., BECKMANN, R. & CRAMER, P. 2007. Functional architecture of RNA polymerase I. *Cell*, 131, 1260-72.
- LAFERTÉ, A., FAVRY, E., SENTENAC, A., RIVA, M., CARLES, C. & CHÉDIN, S. 2006. The transcriptional activity of RNA polymerase I is a key determinant for the level of all ribosome components. *Genes & Development*, 20, 2030-2040.
- LALO, D., STEFFAN, J. S., DODD, J. A. & 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, 21062-7.

- LANDRIEUX, E., ALIC, N., DUCROT, C., ACKER, J., RIVA, M. & CARLES, C. 2005. A subcomplex of RNA polymerase III subunits involved in transcription termination and reinitiation. *The EMBO Journal*, 25, 118.
- LEARNED, R. M., CORDES, S. & TJIAN, R. 1985. Purification and characterization of a transcription factor that confers promoter specificity to human RNA polymerase I. *Mol Cell Biol*, 5, 1358-69.
- LIN, C. W., MOOREFIELD, B., PAYNE, J., APRIKIAN, P., MITOMO, K. & REEDER, R. H. 1996. A novel 66-kilodalton protein complexes with Rrn6, Rrn7, and TATA-binding protein to promote polymerase I transcription initiation in *Saccharomyces cerevisiae*. *Mol Cell Biol*, 16, 6436-43.
- MATTHEWS, B. W. 1968. Solvent content of protein crystals. *J Mol Biol*, 33, 491-7.
- MAYER, C., BIERHOFF, H. & GRUMMT, I. 2005. The nucleolus as a stress sensor: JNK2 inactivates the transcription factor TIF-IA and down-regulates rRNA synthesis. *Genes & Development*, 19, 933-941.
- MAYER, C., ZHAO, J., YUAN, X. & GRUMMT, I. 2004. mTOR-dependent activation of the transcription factor TIF-IA links rRNA synthesis to nutrient availability. *Genes & Development*, 18, 423-434.
- MILKEREIT, P., SCHULTZ, P. & TSCHOCHNER, H. 1997. Resolution of RNA polymerase I into dimers and monomers and their function in transcription. *Biol Chem*, 378, 1433-43.
- MILKEREIT, P. & TSCHOCHNER, H. 1998. A specialized form of RNA polymerase I, essential for initiation and growth-dependent regulation of rRNA synthesis, is disrupted during transcription. *EMBO J*, 17, 3692-703.
- MILLER, O. L., JR & BEATTY, B. R. 1969. Visualization of nucleolar genes. *Science*, 955-957.
- MINAKHIN, L., BHAGAT, S., BRUNNING, A., CAMPBELL, E. A., DARST, S. A., EBRIGHT, R. H. & SEVERINOV, K. 2001. Bacterial RNA polymerase subunit  $\omega$  and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 892-897.
- MINDELL, J. A. & GRIGORIEFF, N. 2003. Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol*, 142, 334-47.
- MOOREFIELD, B., GREENE, E. A. & REEDER, R. H. 2000. RNA polymerase I transcription factor Rrn3 is functionally conserved between yeast and human. *Proc Natl Acad Sci U S A*, 97, 4724-9.
- MOSS, T., LANGLOIS, F., GAGNON-KUGLER, T. & STEFANOVSKY, V. 2007. A housekeeper with power of attorney: the rRNA genes in ribosome biogenesis. *Cell Mol Life Sci*, 64, 29-49.
- MURAKAMI, K. S., MASUDA, S. & DARST, S. A. 2002. Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science*, 296, 1280-4.
- NAIDU, S., FRIEDRICH, J. K., RUSSELL, J. & ZOMERDIJK, J. C. 2011. TAF1B is a TFIIB-like component of the basal transcription machinery for RNA polymerase I. *Science*, 333, 1640-2.
- NEYER, S., KUNZ, M., GEISS, C., HANTSCH, M., HODIRNAU, V. V., SEYBERT, A., ENGEL, C., SCHEFFER, M. P., CRAMER, P. & FRANGAKIS, A. S. 2016. Structure of RNA polymerase I transcribing ribosomal DNA genes. *Nature*.

- PANOVA, T. B., PANOV, K. I., RUSSELL, J. & ZOMERDIJK, J. C. B. M. 2006. Casein Kinase 2 Associates with Initiation-Competent RNA Polymerase I and Has Multiple Roles in Ribosomal DNA Transcription. *Molecular and Cellular Biology*, 26, 5957-5968.
- PETTERSEN, E. F., GODDARD, T. D., HUANG, C. C., COUCH, G. S., GREENBLATT, D. M., MENG, E. C. & FERRIN, T. E. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*, 25, 1605-12.
- PEYROCHE, G., LEVILLAIN, E., SIAUT, M., CALLEBAUT, I., SCHULTZ, P., SENTENAC, A., RIVA, M. & CARLES, C. 2002. The A14-A43 heterodimer subunit in yeast RNA pol I and their relationship to Rpb4-Rpb7 pol II subunits. *Proc Natl Acad Sci U S A*, 99, 14670-5.
- PEYROCHE, G., MILKEREIT, P., BISCHLER, N., TSCHOCHNER, H., SCHULTZ, P., SENTENAC, A., CARLES, C. & 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, 5473-82.
- PILSL, M., CRUCIFIX, C., PAPAI, G., KRUPP, F., STEINBAUER, R., GRIESENBECK, J., MILKEREIT, P., TSCHOCHNER, H. & SCHULTZ, P. 2016. Structure of the initiation-competent RNA polymerase I and its implication for transcription. *Nat Commun*, 7, 12126.
- PLASCHKA, C., HANTSCHKE, M., DIENEMANN, C., BURZINSKI, C., PLITZKO, J. & CRAMER, P. 2016. Transcription initiation complex structures elucidate DNA opening. *Nature*, 533, 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. & CRAMER, P. 2015. Architecture of the RNA polymerase II-Mediator core initiation complex. *Nature*, 518, 376-80.
- RIAL, D. V. & CECCARELLI, E. A. 2002. Removal of DnaK contamination during fusion protein purifications. *Protein Expr Purif*, 25, 503-7.
- RINGEL, R., SOLOGUB, M., MOROZOV, Y. I., LITONIN, D., CRAMER, P. & TEMIAKOV, D. 2011. Structure of human mitochondrial RNA polymerase. *Nature*, 478, 269-73.
- ROBINSON, P. J., TRNKA, M. J., BUSHNELL, D. A., DAVIS, R. E., MATTEI, P. J., BURLINGAME, A. L. & KORNBERG, R. D. 2016. Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. *Cell*, 166, 1411-1422 e16.
- ROEDER, R. G. 1996. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci*, 21, 327-35.
- ROEDER, R. G. & RUTTER, W. J. 1969. Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature*, 224, 234-7.
- ROEDER, R. G. & RUTTER, W. J. 1970. Specific nucleolar and nucleoplasmic RNA polymerases. *Proc Natl Acad Sci U S A*, 65, 675-82.
- ROHOU, A. & GRIGORIEFF, N. 2015. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol*, 192, 216-21.
- RUAN, W., LEHMANN, E., THOMM, M., KOSTREWA, D. & CRAMER, P. 2011. Evolution of two modes of intrinsic RNA polymerase transcript cleavage. *J Biol Chem*, 286, 18701-7.
- RUGGERO, D. & PANDOLFI, P. P. 2003. Does the ribosome translate cancer? *Nat Rev Cancer*, 3, 179-92.

- RUSSELL, J. & ZOMERDIJK, J. C. 2006. The RNA polymerase I transcription machinery. *Biochem Soc Symp*, 203-16.
- SAINSBURY, S., BERNECKY, C. & CRAMER, P. 2015. Structural basis of transcription initiation by RNA polymerase II. *Nat Rev Mol Cell Biol*, 16, 129-43.
- SAINSBURY, S., NIESSER, J. & CRAMER, P. 2013. Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature*, 493, 437-40.
- SCHERES, S. H. 2012. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol*, 180, 519-30.
- SCHNEIDER, D. A. 2012. RNA polymerase I activity is regulated at multiple steps in the transcription cycle: recent insights into factors that influence transcription elongation. *Gene*, 493, 176-84.
- SCHRODINGER, LLC 2015. The PyMOL Molecular Graphics System, Version 1.8.
- SHELDRIK, G. M. 2010. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr*, 66, 479-85.
- SILVERA, D., FORMENTI, S. C. & SCHNEIDER, R. J. 2010. Translational control in cancer. *Nat Rev Cancer*, 10, 254-66.
- SODING, J., BIEGERT, A. & LUPAS, A. N. 2005. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res*, 33, W244-8.
- STEFFAN, J. S., KEYS, D. A., DODD, J. A. & NOMURA, M. 1996. The role of TBP in rDNA transcription by RNA polymerase I in *Saccharomyces cerevisiae*: TBP is required for upstream activation factor-dependent recruitment of core factor. *Genes Dev*, 10, 2551-63.
- SYDOW, J. F., BRUECKNER, F., CHEUNG, A. C., DAMSMA, G. E., DENGL, S., LEHMANN, E., VASSYLYEV, D. & CRAMER, P. 2009. Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. *Mol Cell*, 34, 710-21.
- TAFUR, L., SADIAN, Y., HOFFMANN, N. A., JAKOBI, A. J., WETZEL, R., HAGEN, W. J., SACHSE, C. & MULLER, C. W. 2016. Molecular Structures of Transcribing RNA Polymerase I. *Mol Cell*.
- TANG, G., PENG, L., BALDWIN, P. R., MANN, D. S., JIANG, W., REES, I. & LUDTKE, S. J. 2007. EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol*, 157, 38-46.
- TORREIRA, E., LOURO, J. A., PAZOS, I., GONZALEZ-POLO, N., GIL-CARTON, D., DURAN, A. G., TOSI, S., GALLEGO, O., CALVO, O. & FERNANDEZ-TORNERO, C. 2017. The dynamic assembly of distinct RNA polymerase I complexes modulates rDNA transcription. *Elife*, 6.
- VAN DIJK, M. & BONVIN, A. M. 2009. 3D-DART: a DNA structure modelling server. *Nucleic Acids Res*, 37, W235-9.
- VANNINI, A. & CRAMER, P. 2012. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol Cell*, 45, 439-46.
- WANG, Z. & ROEDER, R. G. 1997. Three human RNA polymerase III-specific subunits form a subcomplex with a selective function in specific transcription initiation. *Genes Dev*, 11, 1315-26.

- WARNER, J. R. 1999. The economics of ribosome biosynthesis in yeast. *Trends Biochem Sci*, 24, 437-40.
- WHITE, R. J. 2005. RNA polymerases I and III, growth control and cancer. *Nat Rev Mol Cell Biol*, 6, 69-78.
- WHITE, R. J. 2008. RNA polymerases I and III, non-coding RNAs and cancer. *Trends Genet*, 24, 622-9.
- WINN, M. D., BALLARD, C. C., COWTAN, K. D., DODSON, E. J., EMSLEY, P., EVANS, P. R., KEEGAN, R. M., KRISINEL, 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. & WILSON, K. S. 2011. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr*, 67, 235-42.
- YAMAMOTO, R. T., NOGI, Y., DODD, J. A. & NOMURA, M. 1996. RRN3 gene of *Saccharomyces cerevisiae* encodes an essential RNA polymerase I transcription factor which interacts with the polymerase independently of DNA template. *EMBO J*, 15, 3964-73.
- ZHANG, Y., FENG, Y., CHATTERJEE, S., TUSKE, S., HO, M. X., ARNOLD, E. & EBRIGHT, R. H. 2012. Structural basis of transcription initiation. *Science*, 338, 1076-80.
- ZUO, Y. & STEITZ, T. A. 2015. Crystal structures of the *E. coli* transcription initiation complexes with a complete bubble. *Mol Cell*, 58, 534-40.

## 7. Abbreviations

**ATP** adenosine triphosphate

**AS** asymmetric unit

**bp** base pair

**CC** closed complex

**CE** core element

**CF** core factor

**CTD** c-terminal domain

**cryo-EM** cryo-electron microscopy

**CV** column volume

**DNA** deoxyribonucleic acid

**dsDNA** double stranded DNA

**DTT** dithiothreitol

**dNTP** deoxynucleotide

**EM** electron microscopy

**IPTG** Isopropyl- $\beta$ -D-thiogalactopyranosid

**ITC** initially transcribing complex

**k** thousand

**Da** dalton

**LB** lysogeny broth

**MR** molecular replacement

**mRNA** messenger RNA

**NCS** non-crystallographic symmetry

**Ni-NTA** nickel-nitrilotriacetic acid

**NTD** n-terminal domain

**OC** open complex

**PAD** propeller- and promoter-associated domain

**PBS** phosphate buffered saline

**PIC** pre-initiation complex +/- UAF

**PIR** polymerase interacting regions

**Pol** RNA polymerase

**Pols** RNA polymerases

**PCR** polymerase chain reaction

**PI** protease inhibitor

**PMSF** phenylmethylsulfonylflourid

**RB** reaction buffer

**rDNA** ribosomal DNA

**RNA** ribonucleic acid

**RT** room temperature

**SAD** single-wavelength anomalous diffraction

**SeMet** selenomethionine

**SDS-PAGE** sodium dodecyl sulfate polyacrylamide gel electrophoresis

**SL** selectivity factor

**TBP** Tata-box binding Protein

**TPR** tetratricopeptide repeat

**TSS** transcription start site

**TF** transcription factor

**TOR** target of rapamycin

**tWH** tandem winged helix

**TAF** TBP associated factors

**UAF** upstream activation factor

**UAS** upstream activating sequence

**UTP** uridine triphosphate

**XL-MS** cross-linking coupled to mass spectrometry

## 8. List of Figures

Figure 1 Assembled yeast Pol I preinitiation complex.....	12
Figure 2 Crystal structure of Rrn3 from <i>S. cerevisiae</i> and its interaction with Pol I.....	15
Figure 3 Pol I crystal structures from <i>S. cerevisiae</i> .....	17
Figure 4 Cryo-EM structure of the Pol I elongation complex of <i>S. cerevisiae</i> .....	18
Figure 5 Structure of the Pol I-Rrn3 complex from <i>S. cerevisiae</i> . .....	19
Figure 6 Domain homologies of initiation factors from Pol I, II, III .....	21
Figure 7 Overview of CF crosslinks.....	21
Figure 8 CF crystallization and data collection .....	35
Figure 9 Experimental phasing of CF .....	36
Figure 10 Crystal structure of core factor.....	37
Figure 11 Crystal structure of yeast core factor. ....	39
Figure 12 Rrn7 structurally differs from TFIIB .....	41
Figure 13 Sequence alignments of Rrn7. ....	42
Figure 14 Secondary structure-based alignments of Rrn6 and Rrn11 orthologs.....	43
Figure 15 Cryo-EM reconstruction of the Pol I-Rrn3-CF complex .....	46
Figure 16 Cryo-EM reconstruction of the initially transcribing complex .....	48
Figure 17 Promoter DNA interactions .....	50
Figure 18 Functional probing of structural features .....	53
Figure 19 Pol I initiation-elongation transition and comparison with Pol II.....	55
Figure 20 Formation and negative stain EM reconstruction of a Pol I-Rrn3-CF complex. ....	64
Figure 21 Classification of the Pol I-Rrn3-CF cryo-EM dataset.....	65
Figure 22 Three Pol I-Rrn3-CF reconstructions.....	66
Figure 23 Pol I ITC sample preparation and cryo-EM data processing. ....	67
Figure 24 Cryo-EM reconstructions of Pol I-Rrn3-CF ITC.....	69
Figure 25 Pol I-specific CF function in transcription initiation. ....	71

## 9. List of Tables

Table 1. RNA polymerases subunits and their relation in <i>S. cerevisiae</i> .....	11
Table 2. <i>E. coli</i> strains used in this study .....	23
Table 3. Media and plates.....	23
Table 4. General buffers used in this study .....	23
Table 5. Buffers of the Pre-Crystallization Test.....	24
Table 6. Solutions for silver staining.....	24
Table 7. General markers, solutions and enzymes .....	24
Table 8. Antibiotic stock solution .....	25
Table 9. PCR protocol.....	26
Table 10. Crystallographic data collection, phasing and refinement statistics for anomalous and native datasets collected from CF crystals. ....	44
Table 11. Crystallographic data collection and phasing statistics for anomalous datasets of native and methionine mutant CF crystals.....	45
Table 12. Pol I cleft expansion and contraction .....	62