# RNA polymerase I domain architecture and basis of rRNA cleavage

Stefan Jennebach

aus Heilbad Heiligenstadt

2011

**Erklärung**

Diese Dissertation wurde im Sinne von §13 Abs. 3 der Promotionsordnung vom 29. Januar 1998, in der Fassung der sechsten Änderungssatzung vom 16. August 2010, von Herrn Prof. Dr. Patrick Cramer betreut.


**Ehrenwörtliche Versicherung**

Diese Dissertation wurde selbstständig und ohne unerlaubte Hilfe erarbeitet. München, am 08.12.2011



Stefan Jennebach




Dissertation eingereicht am: 08.12.2011


Erstgutachter: Prof. Dr. Patrick Cramer

Zweitgutachter: PD Dr. Dietmar Martin


Tag der mündlichen Prüfung: 25.01.2012

# I  Acknowledgement

First of all I want to thank Prof. Dr. Patrick Cramer for giving me the opportunity to graduate within his lab. I am grateful for his constant support as well as his motivating attitude, especially in times when research was barely successful.

Likewise, I would like to thank all past and present members of the Cramer lab who made life much easier over the past years. Thanks to Claus for handing over his project and getting me started. Thanks to the other members of the Pol I team, Sebastian and Christoph for all the suggestions and discussions. Especially, I want to thank Claudia (Blattner) for being a great companion all the time, inside and outside the lab. Many thanks go to Stefan Benkert, without whom I had nothing to purify Pol I from. Thank you Dirk for all the help with crystallographic problems, of which we had dozens.

I also thank Claudia (Buchen) for doing a great job in organizing the lab and always taking care. Thanks to Kristin for getting me started in the lab. Many thanks of course go to you "Else" for always sharing stuff and thoughts and being the real Wikipedia. Thanks to Dengl for many silly, yet highly sophisticated conversations, sarcasm rules! Thank you Rieke for help with the purification of another polymerase than Pol I, and all the discussions on weird science and life. Thanks Alan for scientific discussions as well as introducing me to the world of (strange) British humor. Osmin, thank you for all our funny, sometimes hilarious, conversations and for your always positive view on everything.

Thank you Heidi and Dirk for being members of my thesis advisory committee and for all your help during my thesis.

Thanks to Ale for reading my thesis and for discussions.

I also wish to thank Franz Herzog from Zürich for the fruitful collaboration on crosslinking and mass spectrometry.

Thank you Alex, Niklas, and Steph for being my friends throughout all the years.

Thank you Uli for all the good times we had together, especially in the last two years.

My deepest gratitude goes to my parents for making all this possible. Without your support I wouldn't have been able to achieve anything of this.

## II  Summary

Nuclear transcription in eukaryotes is carried out by three multisubunit RNA polymerases, called Pol I, Pol II, and Pol III. Whereas Pol II transcribes mostly protein coding genes, Pol I and Pol III synthesize ribosomal RNA (rRNA) and transfer RNA (tRNA), respectively. Pol I transcription is the first step in the biogenesis of ribosomes, the cellular factories for protein synthesis. While the structure of Pol II has been studied in detail, structural information is still limited for Pol I.

In this thesis different techniques were applied to investigate structural features of Pol I. As structure solution by X-ray crystallography was impaired by a highly complex non-crystallographic symmetry, other methods had to be employed in order to gain structural information on Pol I. After analysis by cryo-EM had established an overall view of Pol I, several questions on the domain architecture remained open. These were addressed by a novel method, combining protein-protein crosslinking and mass spectrometry. These experiments revealed that subunit A12.2, which shares homologies with Rpb9 as well as with TFIIS, is located in the pore and thus complements the active site of Pol I in order to induce strong intrinsic cleavage activity. Also the position of the A49/34.5 dimerization module and the A49 tWH domain could be revealed. The dimerization module is located on the polymerase lobe, similar to the position of the TFIIF on Pol II. The A49 tWH domain resides above the central cleft of the enzyme and resembles parts of TFIIE. Crosslinking-MS experiments with Pol III support these findings, since the C37/53 dimerization module, which is homologous to that of A49/34.5, could be located at a similar position on the lobe, and C34, the TFIIE homolog in Pol III, could be positioned in a region similar to that of the A49 tWH domain. These findings indicate that Pol I, and also Pol III, are distantly related to a Pol II-TFIIS-TFIIF-TFIIE complex.

## III  Contributions

Since the results presented here could only be achieved through the collaborative work of several determined researchers, a detailed list acknowledges all major contributions.

- The Pol I purification protocol was established by Claus Kuhn, with help of Jochen Gerber from the University of Regensburg
- The Pol I ΔA49/34.5 preparation procedure was established by Stefan Jennebach with help of Claus Kuhn
- The strain for Pol I A12.2ΔC was obtained by Claus Kuhn and the resulting enzyme was purified by Stefan Jennebach
- Pol III was purified by Stefan Jennebach according to the protocol developed by Alessandro Vannini
- Initial Pol I crystals were obtained and optimized by Claus Kuhn
- Initial data processing was done by Stefan Jennebach under supervision of Dirk Kostrewa
- Molecular replacement attempts with custom-made search models were performed by Dirk Kostrewa with diffraction data acquired by Stefan Jennebach and Claus Kuhn
- Cryo-EM structures were determined by Claus Kuhn, together with Marco Gartmann and Sonja Bäumli, supervised by Roland Beckmann
- The initial Pol I homology model was created by Claus Kuhn and extended by Stefan Jennebach
- *In vitro* experiments were performed by Claus Kuhn, Stefan Jennebach, and Sebastian Geiger
- Pol I, Pol I-Rrn3, and Pol III crosslinking was performed by Stefan Jennebach according to protocols developed by Franz Herzog
- The structure of Rrn3 was solved by Claudia Blattner
- MS raw data were analyzed by Franz Herzog
- Analysis of crosslinking-MS data and molecular modeling was performed by Stefan Jennebach

Patrick Cramer designed and supervised all projects.

## IV  Publications

Parts of this work have been published or are in the process of publication.

S. Jennebach, F. Herzog, R. Aebersold, and P. Cramer
*Crosslinking-MS analysis reveals RNA polymerase I domain architecture and basis of rRNA cleavage*
(2012, in revision)

Blattner C, Jennebach S, Herzog F, Mayer A, Cheung AC, Witte G, Lorenzen K, Hopfner KP, Heck AJ, Aebersold R, Cramer P.
*Molecular basis of Rrn3-regulated RNA polymerase I initiation and growth*
Genes Dev. 2011 Sep 22 [Epub ahead of print]

P. Cramer, K.J. Armache, S. Baumli, S. Benkert, F. Brueckner, C. Buchen, G.E. Damsma, S. Dengl, S.R. Geiger, A.J. Jasiak, A. Jawhari, S. Jennebach, T. Kamenski, H. Kettenberger, C.-D. Kuhn, E. Lehmann, K. Leike, J.F. Sydow, and A. Vannini.
*Structure of Eukaryotic RNA Polymerases.*
Annu. Rev. Biophys. 37, 337-352 (2008).

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

# V  Table of contents

# 1  Introduction

## 1.1  Transcription systems in eukaryotes

### 1.1.1  Eukaryotic RNA polymerases

Nuclear transcription in eukaryotic cells is catalyzed by the three RNA polymerases (Pols) I, II and III. In *Saccharomyces cerevisiae* Pol I, II and III comprise 14, 12 and 17 subunits, respectively, and have a total molecular weight of 589 kDa, 514 kDa and 693 kDa, respectively (Cramer et al., 2008) (Table 1).

**Table 1. Subunit composition of RNA polymerases in *S. cerevisiae***

|  | **Pol I** | **MW** | **Pol II** | **Pol III** |  |
|---|---|---|---|---|---|
| **10 subunit core** | A190 | 186.4 kDa | Rpb1 | C160 | homolog |
|  | A135 | 135.7 kDa | Rpb2 | C128 | homolog |
|  | AC40 | 37.7 kDa | Rpb3 | AC40 | homolog |
|  | AC19 | 16.2 kDa | Rpb11 | AC19 | homolog |
|  | A12.2 | 13.7 kDa | Rpb9 | C11 | homolog |
|  | Rpb5 (ABC27) | 25.1 kDa | Rpb5 | Rpb5 (ABC27) | common |
|  | Rpb6 (ABC23) | 17.9 kDa | Rpb6 | Rpb6 (ABC23) | common |
|  | Rpb8 ABC14.5 | 16.5 kDa | Rpb8 | Rpb8 ABC14.5 | common |
|  | Rpb10 (ABC10α) | 8,3 kDa | Rpb10 | Rpb10 (ABC10α) | common |
|  | Rpb12 (ABC10β) | 7.7 kDa | Rpb12 | Rpb12 (ABC10β) | common |
| **Subcomplex A14/43** | A14 | 14,6 kDa | Rpb4 | C17 | counterpart |
|  | A43 | 36.2 kDa | Rpb7 | C25 | counterpart |
| **Subcomplex A49/34.5** | A49 | 46.7 kDa |  | C37 | specific |
|  | A34.5 | 26.9 kDa |  | C53 | specific |
|  |  |  |  | C82 | specific |
|  |  |  |  | C34 | specific |
|  |  |  |  | C31 | specific |
| **Total** | **14 subunits** | **589.6 kDa** | **12 subunits** | **17 subunits** |  |

RNA polymerase II transcribes all protein coding genes (Cramer, 2004), as well as many small RNAs that are involved in regulation of other genes through various mechanisms (Dye et al., 2006).

RNA polymerase III is involved in transcription of small non-translated RNA molecules such as the 5S ribosomal RNA (rRNA), transfer RNA (tRNA), 7SL RNA (an essential component of the signal-recognition particle) and RNA molecules required for post-transcriptional processing of rRNA, mRNA and tRNA. Additionally, Pol III synthesizes short interspersed nuclear elements (SINES) including for example Alu genes in human (Geiduschek and Kassavetis, 2001).

Pol I, however, exclusively synthesizes rRNA. In *S. cerevisiae* rRNA is transcribed as a precursor-rRNA, the 35S rRNA, which is then subsequently processed into 25S, 5.8S and 18S rRNA. These rRNAs are part of native ribosomes.

All three polymerases comprise five identical subunits namely Rpb5, Rpb6, Rpb8, Rpb10, and Rpb12 as well as the five subunits homologous to Rpb1, Rpb2, Rpb3, Rpb9 and Rpb11. The Rpb4/7 heterodimer at the periphery of Pol II, which is not part of the core enzyme, shares homologies with the subcomplexes A14/43 and C17/25 in Pol I and III, respectively (Geiger et al., 2008; Jasiak et al., 2006; Kuhn et al., 2007).

### 1.1.2 Specific subunits in Pol I and Pol III

For Pol II-dependent transcription the general transcription factors (GTFs) TFIIA, -B, -D, -E, -F and –H are essential (Table 2). Relationships between those GTFs and specific subunits are rare and could so far only be observed for TFIIF. All three eukaryotic polymerases consist of a 10-subunit core and a differing number of additional specific subunits (Table 1). Pol I and Pol III contain the peripheral heterodimers A14/43 and C17/25, respectively, which have been shown to be related to the Rpb4/7 subcomplex in Pol II (Geiger et al., 2008; Jasiak et al., 2006; Kuhn et al., 2007; Meka et al., 2003). For Pol I it has been shown that subunit A43 is essential for the interaction with the Pol I initiation factor Rrn3 and thus important for recruiting Pol I to the rDNA promoter (Milkereit and Tschochner, 1998; Peyroche et al., 2000).

Additionally, Pol I contains the specific subunits A49 and A34.5. Recent studies have shown that the N-termini of those subunits form a TFIIF-like dimerization module (Geiger et al., 2010). Furthermore, it was shown that A34.5 stabilizes A49 on Pol I and genetically interacts with DNA topoisomerase I (Gadal et al., 1997). Subunit A49 was shown to be important for Pol I activity (Huet et al., 1975; Liljelund et al., 1992), promoter dependent transcription (Hanada et al., 1996) as well as for recruiting Pol I and Rrn3 to the rDNA promoter (Beckouet et al., 2008).

**Table 2. General transcription factors in Pol II and known homologies to other transcription systems** (Geiger et al., 2010; Hahn, 2004; Wang and Roeder, 1997).

| GTF Pol II | Subunits | Pol I homolog | Pol III homolog | Function |
|---|---|---|---|---|
| TFIIA | 2 | | | stabilizes TBP and TFIID-DNA binding; blocks transcription inhibitors; positive and negative gene regulation |
| TFIIB | 1 | | Brf1 | binds TBP, Pol II and promoter DNA; transcription start site selection |
| TFIID/TBP | 10-16 | TBP | TBP | binds TATA-element and deforms promoter DNA; assembly platform for TFIIB and TFIIA |
| TFIIE | 2 | A49-C-term | C34 | binds promoter; may help to open or stabilize transcription bubble in open complex |
| TFIIF | 3 | A49-N-term A34.5-N-term | C37 C53 | recruitment to PIC; open complex formation |
| TFIIS | 1 | | C11 | stimulates intrinsic cleavage activity and allows for backtracking |
| TFIIH | 10 | | | ATPase/helicase necessary for promoter opening and clearance; helicase for transcription coupled DNA repair; kinase activity for phosphorylation of Pol II CTD; transition from initiation to elongation |

Apart from the Rpb4/7 homolog C17/C25, Pol III contains five specific subunits which are organized in two subcomplexes C53/37 (Landrieux et al., 2006) and C82/34/31 (Wang and Roeder, 1997). The subunits C53/37 have

been shown to adopt positions on the Pol III surface similar to that of TFIIF in the Pol II transcription machinery (Wu et al., 2011) and to play a role in promoter opening (Kassavetis et al., 2010). The trimeric subcomplex C82/34/31 was shown to interact with TFIIB and to play different roles in transcription initiation (Brun et al., 1997; Wang and Roeder, 1997).

### 1.1.3 Structural information on eukaryotic polymerases and GTFs

Most advances in understanding the molecular basis of transcription have been made in the Pol II system. Several structures of Pol II have been solved in the past ten years starting with the atomic models of the 10-subunit Pol II (Cramer et al., 2001) and the full 12-subunit enzyme (Armache et al., 2005). In addition, structures of Pol II together with its transcription factors IIB and IIS could be solved (Kettenberger et al., 2003; Kostrewa et al., 2009) (Fig. 1B). The overall architecture of Pol III has been determined by cryo-electron microscopy and native mass spectrometry (Fernandez-Tornero et al., 2010; Fernandez-Tornero et al., 2007; Lorenzen et al., 2007; Vannini et al., 2010) (Fig. 1C). Furthermore, a homology model was constructed based on the Pol II structure (Jasiak et al., 2006) and the structure of the C17/25 subcomplex was solved by X-ray crystallography (Jasiak et al., 2006). The structure of the human homolog of the Pol III specific subunit C82 was solved by X-ray crystallography (Lefevre et al., 2011).
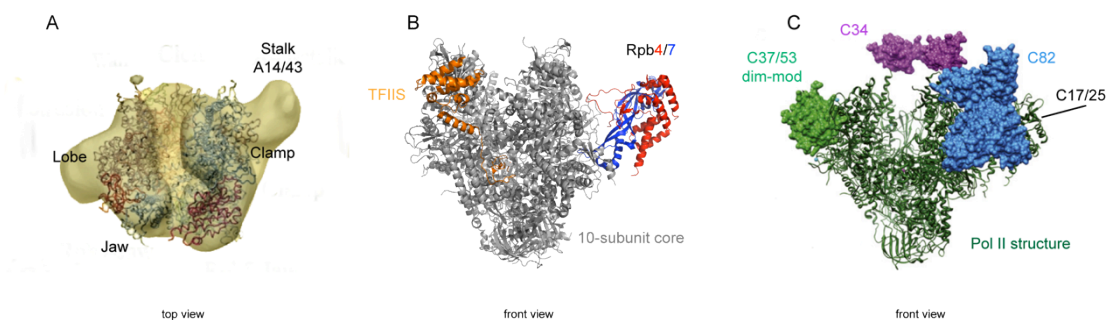


**Figure 1. Current structural models of eukaryotic RNA polymerases.** (**A**) EM model of Pol I with fitted Pol II structure (De Carlo et al., 2003). (**B**) X-ray structure of Pol II in complex with TFIIS (Cheung and Cramer, 2011). (**C**) Structural model of Pol III obtained based on cryo-EM analysis and homology modeling (Vannini et al., 2010).

First structural insights to Pol I were derived by EM analysis of two-dimensional Pol I crystals (Schultz et al., 1993). Later cryo-EM studies resulted in 3D reconstructions at 34 Å resolution and visualized the A14/43 containing stalk as well as densities for the Pol I-specific subunits A49 and A34.5 over the central cleft of the enzyme (Bischler et al., 2002; Peyroche et al., 2002). Different results were obtained by EM analysis using negative staining at 22 Å. This analysis confirmed the stalk density but not the location of A49 and A34.5 (De Carlo et al., 2003) (Fig. 1A). The atomic structure of the Rpb4/7 homolog A14/43 in Pol I was solved, confirming its relationship to Rpb4/7 (Geiger et al., 2008). The structural relationship of the specific subunits A49 and A34.5 to parts of the general transcription factors TFIIF and TFIIE was highlighted by X-ray crystallography (Geiger et al., 2010).

## 1.2   Transcription of rDNA and ribosome biogenesis

### 1.2.1   rRNA in ribosomes

Ribosomes are one of the cell's most complex molecular machines. Their importance for cell growth and proliferation arises from their central role in protein biosynthesis. Today's ribosomes most likely evolved from early protoribosomes that comprised only RNA (Poole et al., 1998). Eukaryotic ribosomes contain about two-thirds RNA and one-third ribosomal protein and the formation of peptide bonds is mostly catalyzed by ribosomal RNA. Thus rRNAs, the transcripion products of RNA polymerase I and III, are the major architectural and catalytic components of the ribosome.

The eukaryotic ribosome comprises a molecular mass of about 4 MDa. The large 60S ribosomal subunit comprises the 25S, 5.8S and 5S rRNA and the small 40S subunit contains the 18S rRNA. The 35S precursor for the 28S and 5.8S rRNA is synthesized by Pol I. Pol III, however, synthesizes the precursor of the 5S rRNA. In addition to the rRNA about 80 ribosomal proteins are part of a functional ribosome, of which in yeast 46 build up the large ribosomal subunit and 32 are located within the small subunit. The maturation of rRNA and thus the ribosome biogenesis involves several intermediate steps and requires a vast number of trans-acting factors (Henras et al., 2008). Since building ribosomes requires the combined action of all three RNA

polymerases to produce the necessary RNAs and ribosomal proteins, ribosome biogenesis is the most energy and resource-consuming process in living cells. In growing cells transcription of rRNA by Pol I accounts for up to 60% of total transcription (Warner, 1999). As a consequence of this high synthetic effort by the Pol I transcription machinery, rRNA synthesis is a central target for the regulation of metabolism and cell growth (Grummt, 2003). The effect of tumor suppressors and oncogenes on ribosome biogenesis (Fontoura et al., 1997; Moss et al., 2007) and the role of Pol I and Pol III in cancer development (White, 2005) identified the rDNA transcription machinery as a target for cancer treatment (Drygin et al., 2011; Drygin et al., 2010).

### 1.2.2  Structure of the nucleolus and organization of ribosomal DNA loci

Synthesis, processing and early stages of ribosome assembly take place in the largest subnuclear compartment, the nucleolus (Fig. 2A), which is formed during transcription of rRNA genes (Nomura, 2001). The initial assembly of ribosomes occurs co-transcriptionally and can be visualized as 'Miller spreads' on electron micrographs (Fig. 2B).
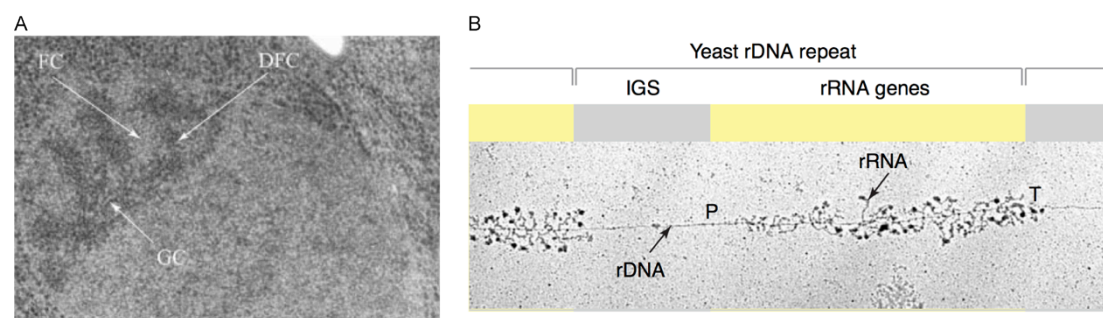


**Figure 2. Structure of the nucleolus and actively transcribed rDNA genes.** (**A**) EM image of a yeast nucleolus (Trumtel et al., 2000). Visible nucleolar elements are fibrillar center (FC), granular center (GC) and dense fibrillar center (DFC). (**B**) 'Miller spread' of a single rDNA repeat. The promoter and terminator regions are labeled with P and T, respectively. Nascent rRNA is indicated with an arrow and black globules at the end of rRNAs mark pre-ribosomes (Russell and Zomerdijk, 2005).

In *S.cerevisiae*, the rDNA locus contains approximately 150 head-to-tail repeats of the rRNA gene. Besides the coding regions, each gene contains a

5' and 3' external transcribed spacer (ETS) as well as two internal transcribed spacers (ITS). These spacer regions harbor cleavage sites important for further processing to mature rRNA (Fromont-Racine et al., 2003) (Fig. 3A). Furthermore, the rDNA locus comprises intergenic spacers (IGS). These spacers are important for efficient pre-rRNA synthesis as well as for rDNA silencing (Moss et al., 2007). The IGS regions contain multiple important sequence elements such as the rDNA promoter, enhancers, the spacer promoter and several terminator elements (Fig. 3B). Interestingly, studies on a yeast strain defective for Pol I transcription showed that neither Pol I nor tandemly arranged rDNA repeats are basal requirements for synthesis of rRNA. This study showed that Pol II could also synthesize rRNA from a high copy plasmid containing the 35S rDNA under control of the *GAL7* promoter (Nogi et al., 1991).
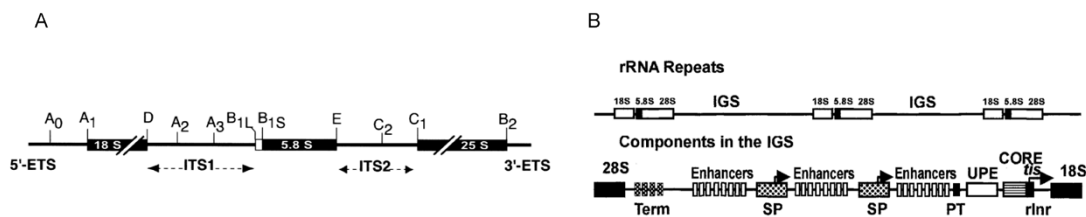


**Figure 3. Ribosomal DNA locus of *S. cerevisiae*.** (**A**) Single rDNA gene. External and internal transcribed spacers, the coding sequences for 18S, 5.8S and 25S rRNA as well as cleavage sites for rRNA processing are shown (Fromont-Racine et al., 2003). (**B**) Coding sequences are separated by intergenic spacers (IGS). IGS components are terminators (Term), enhancers, spacer promoter (SP), proximal terminator (PT), the upstream promoter element (UPE/UE) and the promoter core element (CORE/CE) with the ribosomal initiator (rInr). Sites for initiation are indicated by arrows or *tis* (Paule and White, 2000).

### 1.2.3 Epigenetics of rRNA genes

The number of rDNA repeat loci is variable between different species. Whereas *S. cerevisiae* has only one rDNA repeat locus on the right arm of chromosome XII (Venema and Tollervey, 1999), human and mouse possess five of those clusters on the short arms of different chromosomes (McStay and Grummt, 2008). Since each cluster is capable of forming a nucleolus, when its rRNA genes are actively transcribed, these regions are referred to as nucleolar organizer regions (NORs) (Nomura, 2001). EM studies revealed that

in exponentially growing yeast cells only half of the rDNA repeats are transcriptionally active (French et al., 2003). The creation of yeast strains comprising different numbers of rDNA repeats, however, revealed that the regulation of rDNA transcription is rather dependent on the Pol I loading rate than on the number of active genes (French et al., 2003). Activation and repression of rDNA is mediated by changes in DNA methylation and histone modification (Preuss and Pikaard, 2007). The epigenetic marks that characterize active, euchromatic, and rather inactive heterochromatic rDNA are similar to protein coding genes. The active chromatin state is characterized by DNA hypomethylation, histone H4 acetylation, and histone H3 dimethylation (H3K4me2), whereas the silenced state is marked by CpG methylation, histone H4 hypoacetylation, and H3K9 metyhlation (Earley et al., 2006; Lawrence et al., 2004). rDNA silencing requires sense or antisense transcripts of the non-coding intergenic spacer (IGS) (Bierhoff et al., 2010; Mayer et al., 2006). The produced non-coding RNA is processed and incorporated into the nucleolar remodeling complex (NoRC), which then associates with rDNA, dependent on the transcription termination factor I (TTF-I) (Grummt, 2007; Langst et al., 1997; Santoro et al., 2002). Activation of rDNA transcription requires ATP-dependent chromatin remodeling through CSB (Cockayne Syndrome protein B), a SWI/SNF2-like DNA-dependent ATPase, and WSTF (William syndrome transcription factor) (Bradsher et al., 2002; Felle et al., 2010). Disturbance of these epigenetic states is associated with alterations in the rRNA synthesis rates and genomic stability. It has been shown that overexpression of rRNA and thus excessive protein synthesis might be an initial step in malignant growth (Grummt, 2007; Ruggero and Pandolfi, 2003). In turn, inhibition of rDNA transcription can induce the cell's apoptotic program, which might make Pol I transcription an important target for cancer treatment (Drygin et al., 2010).

### 1.2.4 The Pol I transcription cycle

In yeast and vertebrates, the rDNA promoter is a sequence of about 140-160 base pairs (Moss et al., 2007) (Figure 4). The Pol I promoter is only poorly conserved between species consistent with the extreme species specificity of

the rDNA transcription machinery. Despite the low sequence conservation, most Pol I promoters contain two sequence elements namely the upstream promoter element (UPE/UE) and the core element (CE) (Boukhgalter et al., 2002; Hamada et al., 2001; Moss et al., 2007).
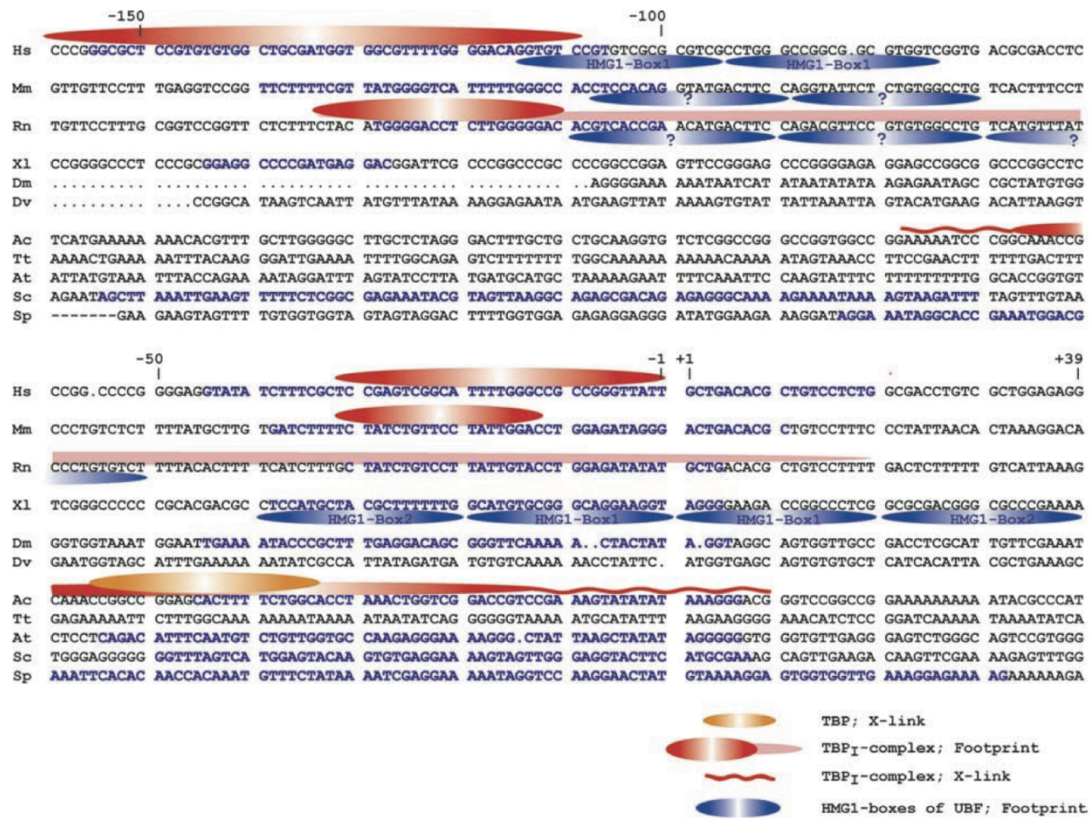


**Figure 4. Structure of Pol I promoters.** Sequences are aligned to the transcription initiation site (+1). Sites of functional importance are colored blue and sites that were shown to interact with the TBP$_I$ complex (SL-1, TIF-IB), UBF or TBP are indicated graphically. Aligned sequences are from *Homo sapiens* (Hs), *Mus musculus* (Mm), *Rattus norvegicus* (Rn), *Xenopus laevis* (Xl), *Drosophila melanogaster* (Dm), *Drosophila virilis* (Dv), *Acanthamoeba castallanii* (Ac), *Arabidopsis thaliana* (At), *Saccharomyces cerevisiae* (Sc) and *Schizosaccharomyces pombe* (Sp). Figure adapted from (Moss et al., 2007).

The spatial arrangement of the promoter elements is a prerequisite for *in vivo* function, however the core element is mostly sufficient for correct transcription initiation *in vitro*.

The formation of a pre-initiation complex (PIC) in all Pol I transcription systems requires the TATA box-binding protein (TBP) and a number of Pol I-specific TATA-box associated factors (TAFs).

In mammals, the PIC comprises the upstream binding factor (UBF), selectivity factor 1 (SL1) or TIF-IB (Bell et al., 1988). UBF is a HMG-1 box protein with specific DNA binding capabilities and has been implicated in enhancement of Pol I transcription by formation of a putative enhancesosome (Bazett-Jones et al., 1994). The mammalian initiation factor SL1 comprises the subunits $TAF_I110/95$, $TAF_I68/63$, $TAF_I48$, $TAF_I41$ and TBP (Comai et al., 1992; Gorski et al., 2007). Upon UBF binding at the promoter region SL1 is recruited through the C-terminal domain of UBF and a stable pre-initiation complex is formed (Tuan et al., 1999).
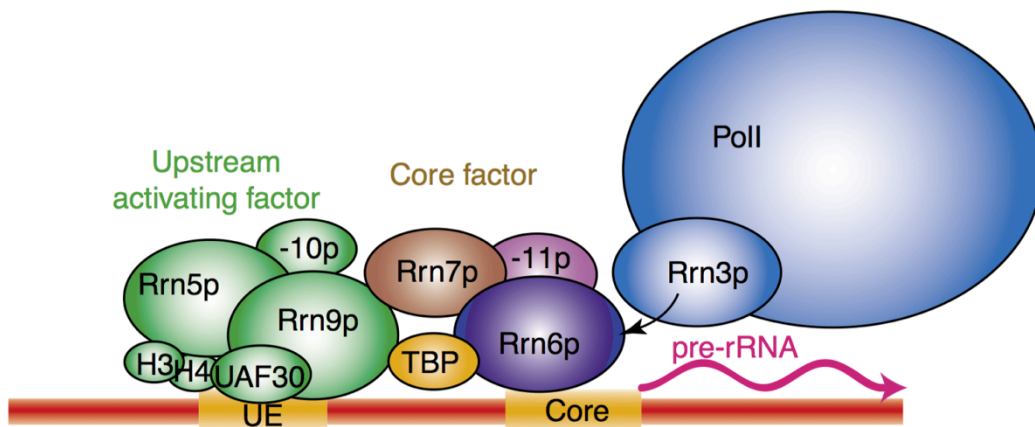


**Figure 5. The pre-initiation complex in *S. cerevisiae*.** Assembly of the Pol I PIC. DNA is shown in red, promoter elements in yellow, UAF subunits are depicted in green and the CF is labeled in brown (Moss, 2004).

The composition of the Pol I initiation complex in yeast shows major differences compared to the mammalian system (Figure 5). The upstream element of the yeast Pol I promoter is bound by the upstream activation factor (UAF) that comprises two histones H3 and H4, an uncharacterized protein UAF30 and the factors Rrn5, Rrn9 and Rrn10. The UAF interacts with TBP, the only component present in all eukaryotic transcription systems. After establishment of the UAF-UE complex, TBP is either already present or is recruited along with the CF (Aprikian et al., 2001; Lin et al., 1996; Steffan et al., 1996). The CF itself binds to the core element and comprises the subunits Rrn6, Rrn7 and Rrn11, which are homologous to $TAF_I110/95$, $TAF_I68/63$ and $TAF_I48$, respectively. The interplay of TBP with the CF is mediated by its

interaction with Rrn7 and Rrn11 (Bordi et al., 2001; Keener et al., 1998; Lalo et al., 1996; Moss et al., 2007). For TAF$_1$41, the fourth human CF subunit, no homolog has been described so far.

Once initiation is accomplished Pol I undergoes a transition to an elongating state. Pol I transcribes the 35S-precursor at a rate of 5-6 kb/min similar to elongation rates of Pol II (Darzacq et al., 2007; Dundr et al., 2002).

Transcription termination occurs at the 3' end of the transcribed region, between the IGS and the rDNA promoter (Figure 3B). Transcription termination requires TTF-I, which bends the T-rich termination site, leads to Pol I pausing and cooperation with the Pol I and transcript release factor PTRF and thus finally leads to dissociation of Pol I from the rDNA (Jansa and Grummt, 1999; Russell and Zomerdijk, 2005).

## 1.2.5 Regulation of rDNA transcription *in vivo*

Transcription of rDNA by RNA polymerase I in eukaryotes is regulated on two levels. A first level of regulation is represented by the alteration of the number of active genes through epigenetic mechanisms (see Chapter 1.2.3). Secondly, transcription regulation can be achieved through different mechanisms that alter the rate of transcription in response to cellular signals.

The regulation of initiation is mainly due to alterations in the phosphorylation pattern of the initiation factors Rrn3 and TIF-IA in yeast and mammals, respectively. The free form of the yeast initiation factor Rrn3 is mostly phosphorylated, whereas unphosphorylated Rrn3 stably binds to Pol I. Phosphorylation of specific sites on the Pol I surface seems to be a requirement for a stable association of Pol I and Rrn3 and thus efficient transcription initiation (Fath et al., 2001; Gerber et al., 2008). In addition an involvement of the target of rapamycin (TOR) kinase or kinases of the TOR signaling pathway seems plausible since rapamycin treatment decreases the amount of Pol I-Rrn3 complexes *in vivo* (Claypool et al., 2004).

In mouse, TIF-IA phosphorylation is mainly regulated through mTOR nutrient sensing pathway (Proud, 2002) and the Jun N-terminal kinase pathway (JNK) (Mayer et al., 2005; Mayer et al., 2004). In mammals, the Raf-MEK-ERK kinase pathway is also involved in modulation of TIF-IA phosphorylation, thus

effecting the formation of the TIF-IA-Pol I complex (Zhao et al., 2003). As growth factors and kinase activation of rDNA transcription do not affect the absolute number of transcribing polymerases (Stefanovsky et al., 2006a), Pol I elongation has to be regulated as well. Indeed phosphorylation of two HMG1 boxes of UBF, leading to remodeling of the potential enhancesome, facilitates transcription elongation (Stefanovsky et al., 2006b; Stefanovsky et al., 2001). Different splice variants have been recognized to be involved in differential regulation of transcription elongation by Pol I (Stefanovsky and Moss, 2008).

### 1.2.6 Ribosome assembly

Structures and function of mature cytoplasmic ribosomes of both, prokaryotes and eukaryotes have been extensively characterized (Ban et al., 2000; Ben-Shem et al., 2010; Schuwirth et al., 2005). However, the assembly pathway resulting in fully functional ribosomes is only partly understood (Fatica and Tollervey, 2002; Kressler et al., 2010; Tschochner and Hurt, 2003) (Figure 6). The synthesis of 18S rRNA in yeast involves four successive cleavages by different endonucleases, while the processes leading to 25S and 5.8S rRNA involves a large number of processing and cleavage events conducted by endo- and exonucleases (Venema and Tollervey, 1999). Terminal knobs at the 5' end of each 35S pre-rRNA, as seen in Miller spreads, indicate that rRNA processing starts co-transcriptionally (Miller and Beatty, 1969; Mougey et al., 1993) (Figure 2 B). These terminal knobs contain the nascent transcript and a number of small nucleolar ribonucleoprotein particles (snoRNPs), which catalyze modification reactions such as 2'-O-ribose methylation (Henras et al., 2008). Additionally, a subset of small subunit ribosomal proteins (Rps) and non-ribosomal factors are involved in formation of the terminal knobs. Thus, the formation of the first pre-ribosomal particle, the 90S particle occurs co-transcriptionally. The 90S particle is joined by additional S-proteins and 40S factors, which then initiate the separation of the 90S pre-ribosome into pre-60S and pre 40S ribosomal subunits. The division of the 90S particle occurs through cleavage at U3 snoRNP-dependent sites (Grandi et al., 2002; Schafer et al., 2003).
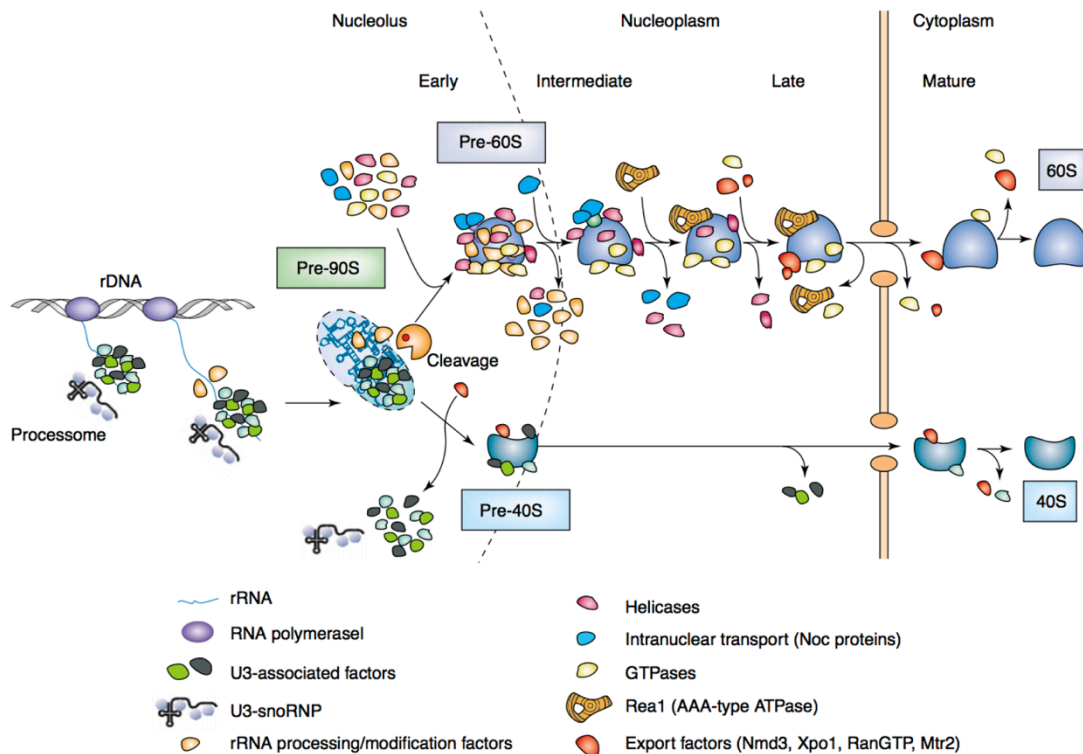
**Figure 6. Model for maturation and export of 40S and 60S ribosomal subunits.** The earliest ribosomal precursor, the 90S particle, contains the 35S pre-rRNA the large U3 snoRNP, part of the processome, and further 40S synthesis factors. Separation into pre-40S and pre-60S particles is followed by several maturation steps in the nucleus and export to the cytoplasm, where the final maturation steps take place. Factors involved in the different steps are graphically depicted. Export factors for 40S subunits are not known. (Tschochner and Hurt, 2003).

Transport and maturation of pre-60S particles involves numerous non-ribosomal factors, such as GTPases, RNA helicases, ATPases and different transport factors, that transiently interact with the pre-60S particle in time and compartment dependent manner (Tschochner and Hurt, 2003). In contrast, after separation, the pre-40S particle is accompanied by only a handful non-ribosomal components such as methyltransferases and GTPases that are involved in 40S maturation (Gelperin et al., 2001; Lafontaine et al., 1995; Lafontaine et al., 1998). Nuclear export of immature ribosomal subunits is again assisted by different export factors like Xpo1/Crm1 and the Ran GTPase (Hurt et al., 1999; Moy and Silver, 1999, 2002; Trotta et al., 2003) and is followed by the release of those factors, which requires different proteins, such as the GTPases Efl1 and Lsg1, and marks the final stage in

13

maturation of the ribosomal subunits (Kressler et al., 2010; Senger et al., 2001).

## 1.3  Aim of this work

In the past years several structures of Pol II and Pol II-specific general transcription factors have been solved (see Chapter 1.1.3). However, structural information for the Pol I transcription machinery is limited to partial structures of individual Pol I subcomplexes, namely A49/A34.5 and A14/43 (see Chapter 1.1.3). Therefore, the aim of this work was to gain insights into the structure of the whole 14-subunit RNA polymerase I. Several techniques, such as X-ray crystallography, cryo-EM and chemical crosslinking and mass spectrometry, have been employed in this work to expand our knowledge on structural properties of Pol I.

Insights into the architecture of this large, 590 kDa, macromolecular complex would possibly enable us to explain functional properties of the enzyme, which, in turn, could help to understand the structural basis of rDNA transcription and its regulation. Knowledge on structural details could also unravel possible evolutionary differences or similarities between the three different eukaryotic polymerase systems. The availability of structural information on Pol II and Pol I substructures was expected to facilitate the efforts of gaining new structural insights, especially for the structure solution by X-ray crystallography.

## 2  Materials and Methods

### 2.1  Materials

### 2.1.1  Bacterial and yeast strains

**Table 3. *E. coli* strains**

| Name | Description | Source |
|------|-------------|--------|
| BL21-CodonPlus (DE3)-RIL | *E. coli* B F$^-$ *ompT hsdS*(r$_B^-$ m$_B^-$) *dcm*$^+$ Tet$^r$ *gal* λ(DE3) *endA* Hte [*argU ileY leuW* Cam$^r$] | Stratagene |
| XL1-Blue | *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F' *proAB lacI$^q$ZΔM15* Tn10 (Tet$^r$)] | Stratagene |
| NEB Turbo Competent *E. coli* (High Efficiency) | F' *proA$^+$B$^+$ lacI$^q$* Δ *lacZ M15/ fhuA2* Δ*(lac-proAB) glnV gal R(zgb-210::Tn10)Tet$^S$ endA1 thi-1 Δ(hsdS-mcrB)5* | New England Biolabs |

**Table 4. *S.cerevisiae* strains**

| Name | Description | Source |
|------|-------------|--------|
| GPY2 | *leu2-Δ1 ade2-101 trp1-Δ63 ura3-52 his3-Δ200 lys2-801 rpa43Δ::LEU2* pAS22 (Trp1) (HA and 6-His tag) | (Kuhn et al., 2007) |
| GPY2 ΔRPA34 | *leu2-Δ1 ade2-101 trp1-Δ63 ura3-52 his3-Δ200 lys2-801 rpa34Δ::HIS3MX6 rpa43Δ::LEU2* pAS22 (Trp1) (HA and 6-His tag) | (Kuhn et al., 2007) |
| GPY2 RPA12ΔC | *leu2-Δ1 ade2-101 trp1-Δ63 ura3-52 his3-Δ200 lys2-801 rpa12Δ::KanMX rpa43Δ::LEU2* pAS22 (Trp1) (HA and 6-His tag) pRS313-*rpa12(1-234)* | (Kuhn et al., 2007) |

### 2.1.2  Media

**Table 5. Media for *E.coli* cultivation**

| Name | Description |
|------|-------------|
| LB | 1 % (w/v) Bacto tryptone, 0.5 % (w/v) yeast extract, 8.6 mM NaCl, 2.6 mM NaOH, plates contained 1.5 % (w/v) agar |
| ZY | 1 % (w/v) tryptone, 0.5 % (w/v) yeast extract, used with 5052 and NPS |
| SOB | 2% (w/v) tryptone; 0.5% (w/v) yeast extract; 8.55 mM NaCl; 2.5 mM KCl; 10 mM MgCl2 |
| SOC | SOB + 20 mM glucose[1] |

15

**Table 6. Additives for *E.coli* media**

| Name | Stock solution | Applied concentration |
| --- | --- | --- |
| 50 × 5052 | 25 % (w/v) glycerol, 2.5 % (w/v) glucose, 10 % (w/v) α-lactose | 0.5 % (w/v) glycerol, 0.05 % (w/v) glucose, 0.2 % (w/v) α-lactose |
| 20 × NPS | 0.5 M $(NH_4)_2SO_4$, 1 M $KH_2PO_4$, 1 M $Na_2HPO_4$ | 25 mM $(NH_4)_2SO_4$, 50 mM $KH_2PO_4$, 50 mM $Na_2HPO_4$ |
| Ampicillin | 30 mg/ml in $H_2O$ | 30 μg/ml |
| Kanamycin | 50 mg/ml in EtOH | 50 μg/ml |
| Chloramphenicol | 50 mg/ml in EtOH | 50 μg/ml |
| Streptomycin | 12.5 mg/ml in 70% EtOH | 12.5 μg/ml |
| IPTG | 1 M IPTG | 0.5-1 mM IPTG |

**Table 7. Media for *S.cerevisiae* cultivation**

| Name | Description |
| --- | --- |
| YEP | 1.5 % (w/v) yeast extract, 2 % (w/v) peptone |
| YEPD | 1.5 % (w/v) yeast extract, 2 % (w/v) peptone, 2 % (w/v) glucose |
| SC-Broth / 2% (w/v) glucose (Formedium) | 0.69% (w/v) nitrogen base, 2% (w/v) glucose; pH 5.6-6.0 |
| CSM-Mix –His (Formedium) | amino acid mix lacking histidine |

**Table 8. Additives and components for *S. cerevisiae* media.**

| Name | Stock solution | Applied concentration |
| --- | --- | --- |
| Ampicillin | 100 mg/ml in $H_2O$ | 0.005% (w/v) ampicillin |
| Tetracycline | 12.5 mg/ml in 70% EtOH | 0.00125% (w/v) tetracycline |
| Geneticin (G418) | 200 mg/ml in $H_2O$ | 200 μg/ml for yeast culture |
| Nourseothricin (clonNAT) | 100 mg/ml in $H_2O$ | 100 μg/ml for yeast culture |
| Antifoam C emulsion | 100% (v/v) antifoam C emulsion | 0.1% (v/v) antifoam C emulsion |

## 2.1.3 Buffers, markers, solutions and enzymes

**Table 9. General buffers, markers, dyes and solutions**

| Name | Description or source |
| --- | --- |
| $100 \times$ PI | 60 µM leupeptin, 200 µM pepstatin A, 98 mM PMSF, 211 mM benzamidine; in EtOH |
| $4 \times$ stacking gel buffer | 0.5 M Tris-HCl, 0.4 % SDS (w/v), pH 6.8 |
| $4 \times$ separation gel buffer | 3 M Tris-HCl, 0.4 % SDS (w/v), pH 8.85 |
| Tris-glycine running buffer | 25 mM Tris, 250 mM glycine, 0.1 % (w/v) SDS |
| MES running buffer | 50 mM MES, 50 mM Tris, 0.1 % (w/v) SDS, 1 mM EDTA |
| $5 \times$ SDS loading dye | 25 % (v/v) glycerol, 7.5 % (w/v) SDS, 250 mM Tris-Cl (pH 6.8 at 20 °C), 0.5 % (w/v) bromophenol blue, 12.5 % (v/v) β-mercaptoethanol |
| $1 \times$ SDS loading dye | 5 % (v/v) glycerol, 1.5 % (w/v) SDS, 50 mM Tris-Cl (pH 6.8 at 20 °C), 0.1 % (w/v) bromophenol blue, 2.5 % (v/v) β-mercaptoethanol |
| Broad range molecular weight marker | Bio-Rad |
| Coomassie stain | 50 % (v/v) ethanol, 7 % (v/v) acetic acid, 0.125 % (w/v) Coomassie Brilliant Blue R-250 |
| Destain | 7 % (v/v) acetic acid, 5 % (v/v) ethanol |
| Ethidium bromide solution 1 % | Roth |
| SYBR Safe DNA gel stain, 10,000 × concentrate in DMSO | Invitrogen |
| TBE | 90 mM Tris, 90 mM boric acid, 2 mM EDTA (pH 8.0) |
| $6 \times$ DNA loading dye | Fermentas |
| Gene Ruler 1 kb DNA ladder 0.1 µg/µl | Fermentas |

**Table 10. Buffers for Pol I purification**

| Name | Description |
| --- | --- |
| Lysis/Freezing buffer | 150 mM HEPES pH 7.8, 60 mM $MgCl_2$, 30% (v/v)glycerol, 5 mM DTT[1], 1x PI[1] |
| Dilution buffer | 100 mM HEPES pH 7.8, 20 mM MgCl2, 400 mM $(NH_4)_2SO_4$, 5 mM DTT[1], 1x PI[1] |
| Dialysis buffer (2x) | 100 mM potassium acetate (KOAc), 40 mM HEPES pH 7.8, 2 mM $MgCl_2$, 20% (v/v) glycerol, 10 mM β-mercaptoethanol[1], 1x PI[1] |
| Res/W1 buffer | 1.5 M KOAc, 20 mM HEPES pH 7.8, 1 mM $MgCl_2$, 10% (v/v) glycerol, 10 mM β-mercaptoethanol[1], 0.5x PI[1] |
| W2 buffer | 300 mM KOAc, 20 mM HEPES pH 7.8, 1 mM $MgCl_2$, 10% (v/v) glycerol, 10 mM β-mercaptoethanol[1] |
| E100 buffer | 300 mM KOAc, 20 mM HEPES pH 7.8, 1 mM $MgCl_2$, 10% (v/v) glycerol, 100 mM imidazole, 10 mM β-mercaptoethanol[1] |
| MonoQ buffer A | 10% (v/v) glycerol, 1 mM $MgCl_2$, 20 mM HEPES pH 7.8, 5 mM DTT[1] |
| MonoQ buffer B | 20 mM HEPES pH 7.8, 1 mM $MgCl_2$, 10% (v/v) glycerol, 5 mM DTT[1], 2 M KOAc |
| MonoQ buffer C | 20 mM HEPES pH7.8, 50 mM $(NH_4)_2SO_4$, 2 M urea, 1 mM $MgCl_2$, 10% (v/v) glycerol, 5 mM DTT[1] |
| MonoQ buffer D | 20 mM HEPES pH7.8, 1 M $(NH_4)_2SO_4$, 1 mM $MgCl_2$, 2 M urea, 10% (v/v) glycerol, 5 mM DTT[1] |
| Superose 6 buffer A | 60 mM $(NH_4)_2SO_4$, 5 mM HEPES pH 7.8, 1 mM $MgCl_2$, 10 μM $ZnCl_2$, 5mM DTT[1] |
| Superose 6 buffer B | 100 mM $(NH_4)_2SO_4$, 20 mM HEPES pH 7.8, 1 mM $MgCl_2$, 10 μM $ZnCl_2$, 5% (v/v) glycerol, 5 mM DTT[1] |
| Superose 6 buffer C | 20 mM HEPES pH 7.8, 1 mM $MgCl_2$, 300 mM potassium acetate, 10% glycerol, and 5 mM DTT |

[1] component added prior to usage

**Table 11. Buffers for Pol III purification**

| Name | Description |
| --- | --- |
| Buffer S500 | 200 mM Tris-HCl pH 8.0, 500 mM $(NH_4)_2SO_4$, 10 mM $MgCl_2$, 10% (v/v) glycerol, 10 mM β-mercaptoethanol[1], 1x PI[1] |
| Buffer D0-resuspension | 40 mM HEPES pH 7.8, 5 mM $MgCl_2$, 10% (v/v) glycerol, 1 mM EDTA, 10 mM β-mercaptoethanol[1], 1x PI[1] |
| Buffer D0-Äkta | 40 mM HEPES pH 7.8, 5 mM $MgCl_2$, 10% (v/v) glycerol, 10 mM imidazole, 10 mM β-mercaptoethanol[1], 1x PI[1] |
| Buffer D200 | 40 mM Hepes pH 7.8, 5 mM $MgCl_2$, 10% (v/v) glycerol, 200 mM KCl, 1 mM EDTA, 10 mM β-mercaptoethanol[1], 1x PI[1] |
| Buffer D1000 | 40 mM HEPES pH 7.8, 5 mM $MgCl_2$, 10% (v/v) glycerol, 1M KCl, 10 mM β-mercaptoethanol[1], 1x PI[1] |
| Buffer D500+10i | 40 mM HEPES pH 7.8, 5 mM $MgCl_2$, 10% (v/v) glycerol, 10 mM imidazole, 500 mM KCl, 10 mM β-mercaptoethanol[1], 1x PI[1] |
| Buffer H250+10i | 40 mM HEPES pH 7.8, 5 mM $MgCl_2$, 10% (v/v) glycerol, 10 mM imidazole, 250 mM $(NH_4)_2SO_4$ 10 mM β-mercaptoethanol[1], 1x PI[1] |
| Buffer H250+250i | 40 mM HEPES pH 7.8, 5 mM $MgCl_2$, 10% (v/v) glycerol, 250 mM imidazole, 250 mM $(NH_4)_2SO_4$, 10 mM β-mercaptoethanol[1], 1x PI[1] |
| Buffer H250 | 40 mM HEPES pH 7.8, 5 mM $MgCl_2$, 250 mM $(NH_4)_2SO_4$, 20% (v/v) glycerol, 0.5 mM EDTA, 10 mM β-mercaptoethanol[1], 1x PI[1] |
| Buffer H1000 | 40 mM HEPES pH 7.8, 5 mM $MgCl_2$, 20% (v/v) glycerol, 0.5 mM EDTA, 1 M $(NH_4)_2SO_4$, 10 mM β-mercaptoethanol[1], 1x PI[1] |
| MonoQ-0 | 40 mM Hepes pH 7.8, 5 mM $MgCl_2$, 10% (v/v) glycerol, 5 mM DTT[1] |
| MonoQ-1000 | 40 mM Hepes pH 7.8, 5 mM $MgCl_2$, 10% (v/v) glycerol, 1 M $(NH_4)_2SO_4$, 5 mM DTT[1] |
| Pol III buffer | 40 mM HEPES pH 7.8, 50 mM $(NH_4)_2SO_4$, 100 µM $MgCl_2$, 10 µM $ZnCl_2$, 5 mM DTT[1] |
| Superose6KCl | 40 mM HEPES pH 7.8, 120 mM KCl, 100 µM $MgCl_2$, 10 µM $ZnCl_2$, 5 mM DTT[1] |

[1] component added prior to usage

**Table 12. Buffers for A49/A34.5 purification**

| Name | Description |
| --- | --- |
| Buffer A | 300 mM NaCl, 50 mM Tris pH 7.5, 10 mM β-mercaptoethanol[1], 1x PI[1] |
| Buffer B | 1 M NaCl, 50 mM Tris pH 7.5, 10 mM β-mercaptoethanol[1] |
| Buffer C | 50 mM Tris pH 7.5, 10 mM β-mercaptoethanol[1] |
| Buffer D | 100 mM NaCl, 50 mM Tris pH 7.5, 5 mM DTT[1] |

[1] component added prior to usage

**Table 13. Buffers for Rrn3 purification**

| Name | Description |
| --- | --- |
| Buffer A | 50 mM HEPES pH 7.8, 200 mM NaCl, 10% (v/v) glycerol, 3 mM DTT[1] |
| Buffer B | 50 mM HEPES pH 7.8, 10% (v/v) glycerol, 5 mM DTT[1] |
| Buffer C | 50 mM HEPES pH7.8, 300 mM NaCl, 5 mM DTT[1] |

[1] component added prior to usage

**Table 14. Pol I crystallization solutions**

| Name | Description |
| --- | --- |
| Pol I buffer | 60 mM $(NH_4)_2SO_4$, 5 mM HEPES pH 7.8, 1 mM $MgCl_2$, 10 µM $ZnCl_2$, 5mM DTT[1] |
| Di-ammonium tartrate (mother solution) | 8.0, 8.5, 9.0, 9.5, 10.0% (w/v) PEG 6'000, 3% (v/v) MPD, 100 mM HEPES pH 7.5, 300 mM Di-ammoniumtartrate, 3 mM TCEP |
| Di-ammonium tartrate (cryo solution) | 9.0% (w/v) PEG 6'000, 3% (v/v) MPD, 100 mM HEPES pH 7.5, 300 mM Di-ammoniumtartrate, 3 mM TCEP, 22% (v/v) PEG 400 |

[1] component added prior to usage

**Table 15. Buffers, solutions and components for *in vitro* RNA assays**

| Name | Description |
| --- | --- |
| Transcription buffer | 20 mM HEPES pH 7.6, 60 mM $(NH_4)_2SO_4$, 8 mM $MgSO_4$, 10 µM $ZnCl_2$, 10% (v/v) glycerol, 10 mM DTT |
| 2x urea loading dye | 180 mM Tris, 180 mM boric acid, 4 mM EDTA pH 8.0, 8 M urea, 0.03% (w/v) bromophenol blue, 0.03% (w/v) xylene cyanol FF |
| 2x urea loading buffer | 180 mM Tris, 180 mM boric acid, 4 mM EDTA pH 8.0, 8 M urea |
| TBE | 90 mM Tris, 90 mM boric acid, 2 mM EDTA pH 8.0 |
| Urea gel | 10-20% acrylamide, 7 M urea, 1x TBE |

**Table 16. Buffers for mass spectrometry**

| Name | Description |
|---|---|
| SEC mobile phase buffer | 30% (v/v) acetonitrile, 0.1% (v/v) trifluoroacetic acid, 69.9 % (v/v) $H_2O$ |
| $C_{18}$ mobile phase buffer | 3% (v/v) acetonitrile, 0.1% (v/v) formic acid, 96.9% (v/v) $H_2O$ |

## 2.2 Methods

### 2.2.1 Protein expression and purification

#### 2.2.1.1 Purification of Pol I from *S.cerevisiae*

Complete 14-subunit Pol I was purified from *Saccharomyces cerevisiae* strain GPY2 (*leu2-Δ1 ade2-101 trp1-Δ63 ura3-52 his3-Δ200 lys2-801 rpa43Δ::LEU2*), carrying a pAS22 plasmid coding for a HA and hexahistidine tagged version of subunit A43. A 20 L fermenter was inoculated with culture from shaking flasks to a starting $OD_{600}$ of approximately 0.2. Cells were grown for ≈ 8h at 30 °C to an $OD_{600}$ of 1.5 – 2.0. The pre-culture was used for inoculation of a 200 L fermenter to a starting $OD_{600}$ of 0.2. Fermentation was carried out over night at 30 °C until the cells reached an $OD_{600}$ of 6-8. The cells were harvested at 20,000 rpm in a flow-through centrifuge (Padberg Z4IG) yielding approximately 2.0 – 2.5 kg of yeast cells. The pellet was resuspended in 500 ml freezing buffer per 1kg of cells and subsequently flash frozen in liquid nitrogen in 225 ml batches.

For each purification, two 225 ml batches were thawed in warm water. Ammonium sulfate and PI was added resulting in final concentrations of 400 mM and 1x, respectively. Bead beating was done using two beakers filled with 200 ml glass beads each. The cells suspension was added to the pre-cooled glass beads and the beakers were filled up with dilution buffer in order to prevent the formation of foam during the process. Cell lysis was carried out for 1.5 h at 4 °C in repetitive cycles of 30 s bead beating followed by 90 s cooling. The lysed cells were separated from the glass beads by filtration. Glass beads were washed with 50 – 100 ml dilution buffer. The lysate was centrifuged at 8,000 g in a SLA-1500 rotor (Sorvall) for 30 min at 4 °C. The supernatant was subjected to ultracentrifugation at 30,000 g for 90 min in a SW28

ultracentrifuge rotor (Beckman) in order to obtain a whole cell extract. After removing the top fat layer, the supernatant was dialyzed overnight at 4 °C in 40-50 ml batches against 2 L of 1x dialysis buffer. The dialyzed extract was centrifuged at 18,500 g for 90 min in a Ti45 rotor (Beckman). The resulting pellet was resuspended in Res/W1 buffer and incubated for 4 h with 8 ml pre-equilibrated Ni-NTA agarose (Qiagen) on a turning wheel at 4 °C. The suspension was loaded into two gravity flow columns and washed with 5 CV Res/W1 and W2 buffer, respectively. Bound proteins were eluted with 25 ml buffer E100, resulting in 50 ml of total sample. For the ensuing anion exchange chromatography a MonoQ 10/100 column (GE Healthcare) was equilibrated with 15 % MonoQ buffer B. The sample was loaded at a concentration of 15 % buffer B and eluted with buffer B by applying a multi-step gradient. Pol I containing fractions eluted at a concentration of ≈ 1.1 M KOAc (conductivity ≈ 49 mS/cm$^2$). The Pol I containing fractions were pooled and diluted 5.5-fold with MonoQ buffer A to a final KOAc concentration of 200 mM. For the final purification stage a MonoS 5/50 cation-exchange column (GE-Healthcare) was used, applying a gradient from 200 mM KOAc to 2 M KOAc with buffers A and B. Pol I eluted at a concentration of ≈ 490 mM KOAc (conductivity ≈ 49 mS/cm$^2$). In order to remove glycerol and check the sample for monodispersity a size-exclusion chromatography was performed using Superose 6 buffer A and a Superose 6 10/300 GL column (GE Healthcare). For crosslinking experiments the final gelfiltration step was performed using Superose 6 buffer C.

### 2.2.1.2 Preparation of Pol I ΔA49/A34.5

Pol I lacking the specific subunits A49 and A34.5 was purified with the help of controlled urea dissociation. Therefor Pol I containing fractions after cation-exchange chromatography were pooled and diluted 1:1 with 25% buffer B and dialyzed overnight against MonoQ buffer C. The dialyzed sample was subjected to an anion-exchange chromatography column (MonoQ 5/50 GL, GE Healthcare). Pol I ΔA49/34.5 was eluted applying a gradient from 50 mM to 1M $(NH_4)_2SO_4$ using MonoQ buffer C and D. For removal of urea and glycerol Pol I ΔA49/34.5 was further purified by size-exclusion

chromatography using Superose 6 buffer B and a Superose 6 10/300 GL column (GE Healthcare).

### 2.2.1.3 Preparation of Pol I A12.2ΔC

The Pol I variant lacking residues 79-125 of subunit A12.2 was fermented in synthetic dextrose complete (SDC) medium without histidine. The purification protocol is exactly the same as for wild-type Pol I except the final gelfiltration step was omitted.

### 2.2.1.4 Preparation of Pol I in complex with Rrn3

Purified Pol I was incubated overnight with a 9-fold molar excess of purified Rrn3. The resulting complex was finally purified by size-exclusion chromatography using a Superose 6 10/300 column (GE Healthcare) and Superose 6 buffer A containing 150 mM KOAc instead of $(NH_4)_2SO_4$.

### 2.2.1.5 Preparation of Rrn3

Rrn3 from *S. cerevisiae* was cloned into pET28b using the NheI/HindIII restriction sites resulting in an N-terminall hexahistidine tag. The protein was expressed in *E. coli* BL21 (DE3) RIL cells (Stratagene). Expression was performed in auto-inducing medium (TB) for 16 h at 24°C. Cells were harvested, washed with PBS and lysed by sonication. The lysate was centrifuged and the supernatant was loaded onto a 1 ml gravity flow Ni-NTA column (BioRad, Qiagen) equillibrated with buffer A. The column was washed with 20 CV buffer A containing 20 mM imidazole, followed by washing steps with buffer A containing 30 mM and 50 mM imidazole. Bound proteins were eluted with buffer A containing 150 mM imidazole. Rrn3 was further purified by anion exchange chromatography using a MonoQ 5/50 column (GE Healthcare). The column was equilibrated buffer B and proteins were eluted applying a linear gradient from 100 mM to 1 M NaCl over 20 CV. After concentration, the sample was applied onto a Superdex 200 10/300 GL (GE Healthcare) size-exclusion chromatography column equilibrated with buffer C. Peak fractions were pooled and concentrated.

## 2.2.1.6 Preparation of recombinant A49/A34.5

The genes for A49 and A34.5 were cloned sequentially into vector pET28b (Novagen), introducing a C-terminal hexahistidine tag on A49 and a second ribosomal binding site for bicistronic expression. The two subunits were coexpressed for 18 h at 18 °C in *E. coli* BL21 (DE3) RIL cells (Stratagene) in 4 l of LB medium. Cells were harvested by centrifugation, resuspended in 100 ml buffer A, and lysed by sonication. After centrifugation the supernatant was loaded onto a 3 ml Ni-NTA column equilibrated with buffer C. The column was washed stepwise with 15 ml of buffer B and 15 ml of buffer B containing 30 mM imidazole. The A49/34.5 heterodimer was eluted with buffer B containing 100 mM imidazole. Eluted fractions were diluted 3-fold with buffer C. A MonoS cation exchange column was equilibrated with buffer D, and proteins were eluted with a linear gradient of 18 column volumes from 100 mM to 1 M NaCl. A49/34.5 eluted at 280 mM NaCl. The sample was applied to a Superose 12 HR 10/300 gel filtration column (GE Healthcare) equilibrated with buffer D. Pooled peak fractions were concentrated to 10 mg/ml.

## 2.2.1.7 Purification of Pol III from *S.cerevisiae*

The *S.cerevisiae* strain (Lannutti et al., 1996), carrying the gene for an N-terminally His6-FLAG4-RET1-tagged C128 subunit on the parent plasmid pYE(CEN3)30 was grown to $OD_{600}$ of 6-7 at 30°C in YPD media in a 200 l fermenter. Cells were harvested by continuous flow centrifugation at 20,000 rpm in a Padberg centrifuge (Padberg Z4IG).

Cells were lysed by bead beating in ice-cold buffer S500. The subsequent steps were performed at 4 °C. The lysate was cleared by centrifugation at 8,000 g for 60 min in a SLA-1500 rotor (Sorvall). The supernatant was filtered and a whole cell extract was obtained after ultracentrifugation for 90 min at 125,000 g in a Ti45 rotor (Beckman) and separation of the clear upper-middle phase from the turbid lower phase. The supernatant was further processed by stepwise ammonium sulfate precipitation. In the first step 35% (w/v) ammonium sulfate (0.125 g crushed $(NH_4)_2SO_4$ per ml extract) were slowly added to the sample. After stirring for 30 min the sample was cleared by

centrifugation for 60 min at 8,000 g in a SLA-1500 rotor (Sorvall). The supernatant was precipitated by addition of 70% (w/v) ammonium sulfate (0.23 g/ml) and left stirring overnight. The pellet was recovered by centrifugation for 60 min at 8,000 g in rotor SLA-1500 (Sorvall) and diluted with 3 L of buffer D0-resuspension. The sample was loaded to a 250 ml Biorex 75 (Biorad) column equilibrated with buffer D200 and eluted with 500 mM KCl (1:1 mixture of buffer D0-Äkta and buffer D1000). The eluted proteins were loaded on a manually packed 12 ml Ni-NTA (Qiagen) column, equilibrated with buffer D500+10i and subsequently washed buffer H250+10i. Bound proteins were eluted using buffer H250+250i. The resulting sample was further processed by an affinity chromatography step using a 5 ml HiTrap Heparin column (GE Healthcare). Bound protein was eluted applying a gradient from 250 mM to 1M $(NH_4)_2SO_4$ using buffer H250 and H1000. Pol III elutes at ≈ 500 mM $(NH_4)_2SO_4$. Pol III containing fractions were pooled and diluted 5-fold with MonoQ-0 buffer and subjected to anion-exchange chromatography using a MonoQ 10/100 GL column (GE Healthcare) and buffers MonoQ-0 and MonoQ-1000. Applying a salt gradient from 50 mM to 1M $(NH_4)_2SO_4$ resulted in elution of Pol III at a concentration of ≈600 mM ammonium sulfate. Pol III containing fractions were pooled and diluted to 50 mM $(NH_4)_2SO_4$. In order to overcome the slight endogenous substoichiometry of subunits C53/37 in the complex (Lorenzen et al., 2007) purified Pol III was supplemented with a 10-fold molar excess of the recombinantly expressed and purified C53/37 heterodimer and incubated for 60 min on ice. The sample was concentrated to a volume of approximately 1 ml and finally purified by size-exclusion chromatography using a Superose 6 10/300 GL column (GE Healthcare) and buffer Superose6KCl. The sample was either immediately used for experiments or flash-frozen in liquid $N_2$ after the addition of 10% (v/v) glycerol.

## 2.2.2 Crystallization of Pol I and attempts on solving the structure

### 2.2.2.1 Crystallization by vapor diffusion

Pol I was purified as described and subsequently concentrated to 5.5 mg/ml (1.5, 3.0 and 4.0 mg/ml for Pol I ΔA49/A34.5) using Amicon ultra centrifugal devices with a molecular weight cutoff of 10 kDa (Millipore). Protein concentrations were determined measuring the UV absorption at 280 nm using a NanoDrop 1000 (Thermo Scientific) assuming an extinction coefficient of 422,925 $M^{-1}$ $cm^{-1}$ and 390,480 $M^{-1}$ $cm^{-1}$ (calculated with ProtParam on www.expasy.org) for Pol I and Pol I ΔA49/A34.5, respectively. For slow sample concentration the centrifuge speed was reduced to 5,000 – 6,000 rpm at 4°C. In order to remove dust and aggregated particles, the concentrated sample was centrifuged at 14,000 rpm for 30 min at 4°C before crystallization. Crystallization by vapor diffusion was carried out in hanging drops as well as in sitting drops. For hanging drops EasyXtal 24-well crystallization plates (Nextal/Qiagen) were used. Hanging drops were set using 500 µl reservoir solution and 1 µl protein + 1 µl reservoir drops. The protein drop was set prior to the addition of reservoir solution. Additionally the reservoir contained 3 mM Tris(2-carboxyethyl)phosphine (TCEP) as reducing agent. Initial crystallization screens were set up at the crystallization facility of the MPI for biochemistry using 96-well sitting drop plates and commercially available screening solutions.

### 2.2.2.2 Streak-seeding

Crystals suitable for performing X-ray diffraction experiments could only be obtained using streak-seeding (Bergfors, 2003). Hanging drop crystallization setups were allowed to equilibrate for 3-3.5 h prior to streak-seeding. Seeds were derived from source drops by the addition of 10 µl of reservoir solution containing fresh reducing agent. In order to obtain suitable seeds crystals in the source drop were shred by pipetting. Cat whiskers were used to collect seeds from the source drop by streaking the whiskers through the diluted source drop. The collected seeds were then added to at least 6 identical

drops yielding a consecutive seed dilution. The wells were immediately closed after streak-seeding.

### 2.2.2.3 Crystal harvesting and cryo-protection

Crystals that had reached their maximum size of about 500 µm x 60 µm x 10 µm were harvested using crystal manipulation and freezing tools (Hampton Research). Drops that contained suitable crystals were diluted by adding 5 µl reservoir solution. For cryo-protection crystals were transferred to to spot plates containing 100 µl reservoir solution supplemented with 6% (v/v) PEG 400 and 3% freshly prepared TCEP. After 30 min equilibration time the concentration of the cryo-protectant was increased stepwise to 12%, 18% and 22% (v/v) PEG-400 by exchanging the respective cryo-solutions. The final step was repeated for another 30 min to ensure complete exchange of the cryo-solution.

### 2.2.2.4 Heavy atom derivatization and crystal freezing

Prior to heavy atom derivatization of Pol I crystals the crystals were transferred to the final cryo-solution containing 22% PEG-400 and slowly cooled down to 8°C. Grains of different heavy atom clusters (see Table 17) were added to the crystals. The crystals were kept at 8°C for different time periods according to the heavy atom derivate that was added (see Table 17). Crystals were harvested using 20 µm CrystalCap HT equipment (loop size 0.1-0.6 mm, sample holder length 22 mm) (Hampton Research). Harvested crystals were instantly plunged into liquid nitrogen and stored at -196°C until data collection.

### 2.2.2.5 Crystallization of Pol I elongation complexes

In order to obtain different crystals or crystal forms Pol I was crystallized in the presence of nucleic acids. Nucleic acids were either co-crystallized with Pol I or soaked into previously obtained Pol I crystals.

For co-crystallization purified Pol I was incubated with a 2x molar excess of pre-annealed minimal nucleic acid elongation constructs, EC1 and EC2, for 20 min at 20°C shaking at 350 rpm in Superose 6 buffer A (EC1: non-template DNA: 5'-CGCTCGACCTCG-3', template DNA: 3'-GATCTGGTCCTGTGCGAGCTGGAGC-5', RNA: 5'-FAM-GACCAGGAC-3'; EC2: non-template DNA: 5'-CGCTCGACCTCG-3', template DNA: 3'-CTGGTCCTGTGCGAGCTGGAGC-5', RNA: 5'-FAM-AACGGAGACCAGGAC-3'). The mixture was then directly subjected to crystallization setups using commercially available screens using a Hydra II crystallization robot (BioRad).

In order to soak nucleic acid scaffolds (EC1 and EC2) into pre-existing Pol I crystals the crystals were transferred into spot plates containing Pol I cryo-solutions (see Chapter 2.2.2.3) supplemented with 10 µM nucleic acids. The cryo-solutions containing nucleic acids were exchanged every 30 min. The final step was repeated and allowed to incubate for 30 min or overnight before the crystals were flash-frozen in liquid nitrogen.

## 2.2.2.6 Data collection

All diffraction data was recorded at beamline X06SA at the Swiss Light Source (SLS) in Villigen, Switzerland using the Pilatus 6M detector (Dectris, Baden, Switzerland). In order to verify that soaked heavy metal clusters are present in the crystals, X-ray absorption scans were performed at specific absorption edges of the used heavy metal derivatives (e.g. L-III edge of tungstate at 10.21 keV) prior to the actual diffraction experiment. Because of extreme sensitivity of the crystal towards radiation damage the beam flux was kept constant at $1 \times 10^{12}$ photons/s to enable for comparison of diffraction quality between different crystals. Furthermore the beam was focused on the detector rather than on the sample to reduce radiation damage. Simulated strategies (calculated with XPLAN, part of the XDS package, (Kabsch, 1993)) for the data collection were used to minimize X-ray exposure. However, despite all precautions, several translations on a single crystal were necessary to record a full dataset at roughly the same resolution. This in turn introduced severe problems concerning data integration and scaling.

## 2.2.2.7  Data processing

Data were processed using the XDS software package (Kabsch, 1993). During integration most difficulties arose from changes in the unit cell parameters, which were induced by the severe radiation damage of Pol I crystals. This made refinement of these parameters difficult. The often observed high mosaicity of > 0.7° and many different translations impaired scaling with XSCALE (part of the XDS package). Images suitable for scaling were determined by monitoring the average I/σ and the batch-wise R-factor in the XSCALE output file. Data quality criteria of I/σ ≥ 2 and an $R_{merge}$ of < 40% were applied before phasing attempts.

Self-rotation functions were calculated using GLRF (Tong and Rossmann, 1990) and stereographic projections were calculated for κ = 51° and κ = 180°. Wilson-Plots were calculated using SFCHECK (Vaguine et al., 1999).

## 2.2.2.8  Molecular replacement

Various models based on the Pol II structure (Armache et al., 2003) as well as reconstructions derived from cryo-EM (Kuhn et al., 2007) were used for molecular replacement trials. All models were used for molecular replacement employing PHASER (McCoy et al., 2005; Read, 2001; Storoni et al., 2004).

For using cryo-EM maps in molecular replacement attempts the maps had to be converted to the CCP4 format using SPIDER (Frank et al., 1996). The spider volume was interpolated to 1 Å/pixel and the resulting volume was embedded into a 300 Å x 300 Å x 300 Å unit cell. After determination of the center of gravity the molecule was shifted  to this center and the map was finally converted to CCP4 format using CP to CCP4 in 32 bit mode. Resulting maps were used as search models in molecular replacement together with the self-rotation information calculated previously.

### 2.2.3 *In vitro* RNA assays

### 2.2.3.1 RNA extension assays using nucleic acid scaffolds

Complete Pol I, Pol I ΔA49/A34.5 or Pol I A12.2ΔC were incubated for 30 min at 20°C with 2 pmol of a pre-annealed minimal nucleic acid scaffold (template DNA: 3'-GCTCAGCCTGGTCCGCATGTGTCAGTC-5'; non-template DNA: 5'-C ACACAGTC AG-3'; RNA: 5'-FAM-UGCAUAAAGACCAGGC-3').

For complementing Pol I ΔA49/A34.5, a 5-fold molar excess of recombinant A49/A34.5 was incubated with Pol I ΔA49/A34.5 for 10 min at 20°C before assembly of the polymerase-nucleic acids complex. For elongation, complexes were incubated in the presence of 1 mM NTPs at 28°C for 20 min in transcription buffer (see Table 15). Reactions were stopped by addition of an equal volume of 2x loading buffer (Table 15) and incubation at 95°C for 5 min. 6-FAM labeled RNA extension products were separated via gel electrophoresis using polyacrylamide gels containing 8 M urea and visualized with a Typhoon 9400 imaging system (GE Healthcare). 6-FAM was excited with blue light at a wavelength of $\lambda$ = 488 nm and the fluorescent signal was recorded with a 520 BP 40 band-pass filter.

For extension assays with a complementary bubble (Kireeva et al., 2000) 6 pmol Pol I or Pol I ΔA49/A34.5 were incubated with 3 pmol of a scaffold containing template DNA and RNA (template DNA: 3'-TGCGCACCACGCTTACTGGTCCGTTCGCCTGT CCTCGACCA-5'; RNA: 5'-FAM-UGCAUUUCGACCAGGC-3') for 15 min at 20°C. The complementary bubble was made complete by addition of a 5-fold molar excess of non-template DNA (5'- TTTTTACGCGTGGTGCGAATGACCAGGCAAGCGGACA GGAGCTGGT-3') to the polymerase-RNA-template DNA complex and incubation for 15 min at 25°C. The emerged polymerase-nucleic acids complexes were incubated in the presence of 1 mM NTPs at 28°C for 1 and 5 min in transcription buffer. Reactions were stopped and analyzed by gel electrophoresis as described above.

## 2.2.3.2 RNA cleavage assays

Complete Pol I, Pol I ΔA49/A34.5 and Pol I A12.2ΔC was incubated for 30 min at 20°C with a nucleic acid scaffold that comprised an RNA with a 6-FAM fluorescent label at its 5'-end and a three-nucleotide mismatched overhang at its 3'-end (template DNA: 3'-TTACTGGTCCTTTTTCATGAACTCGA-5'; non-template DNA: 5'-TAAGTACTTGAG CT-3'; RNA: 5'-FAM-UGCAUU UCGACCAGGA<u>CGU</u>-3', overhanging nucleotides underlined). For RNA cleavage reactions, samples were incubated in transcription buffer at 28°C for up to 30 min. The resulting RNA species were analyzed by gel electrophoresis and fluorescence detection as described.

## 2.2.4 Crosslinking and mass spectrometry

### 2.2.4.1 Chemical crosslinking of multi-subunit protein complexes

Pol I, Pol I in complex with Rrn3 and Pol III were cross-linked using isotopically coded disuccimidyl suberate (DSS-$d_0$/$d_{12}$, Creative Moelcules Inc.). The optimal DSS to protein ratio was determined by mixing 8 µg aliquots of the respective protein complex with 25 mM DSS dissolved in dimethylformamide (DMF, Pierce Protein Research Products) (DSS concentration refers to the concentration of one isotope) to a final DSS concentration of 0.02 mM, 0.04 mM, 0.08 mM, 0.16 mM, 0.4 mM, 0.8 mM, 1 mM, 2mM and 4 mM, respectively. As the best concentration the concentration was chosen, which was sufficient to convert most of the individual polymerase subunits into a higher molecular weight band avoiding oligomers of the respective complexes.

Purified Pol I (100 µl containing 110 µg Pol I) was mixed with 25 mM DSS dissolved in DMF to a final concentration of 0.6 mM DSS and incubated for 30 min at 30°C and 350 rpm. The reaction was stopped by the addition of 1 M $NH_4HCO_3$ to a final concentration of 100 mM and incubation for at least 15 min at 30°C. Purified Pol I – Rrn3 complex was concentrated to 1 mg/ml (total amount 70 µg) and mixed with 25 mM DSS to a final concentration of 1.2 mM DSS. The mixture was incubated for 30 min at 30°C and 350 rpm. The cross-linking reaction was stopped by addition of 1 M $NH_4HCO_3$ to a final

concentration of 100 mM, incubating for at least 15 min at 30°C and 350 rpm. Pol III (95 μl containing 100 μg Pol III) was crosslinked accordingly, using a final DSS concentration of 1 mM.

To increase the crosslink yield, Pol I and Pol III were also crosslinked with DSS concentrations of 3.5 mM and 3 mM DSS, respectively. In order to separate potential polymerase oligomers, which might have emerged during crosslinking, a final size exclusion step was added to the crosslinking protocol. Therefore Pol I was subjected to a Superose 6 10/300 GL column using Superose 6 buffer C supplemented with 100 mM $NH_4HCO_3$. Size exclusion chromatography of Pol III after crosslinking was performed in Superose6KCL buffer containing 100 mM $NH_4HCO_3$.

### 2.2.4.2 Sample preparation for MS analysis

The cross-linked proteins were linearized by addition of two sample volumes of 8 M urea and subsequently reduced and alkylated using 5 mM Tris(2-carboxyethyl)phosphine (TCEP) and 10 mM iodoacetamide, respectively. The sample was digested using trypsin following standard protocols. Purified samples were reconstituted in 20 μl of SEC mobile phase buffer. 15 μl of reconstituted sample were applied to a Superdex Peptide PC 3.2/20 column at a flow rate of 50 μl/min. For LC-MS/MS, fractions of interest (retention volume 0.9-1.4 ml) were pooled and evaporated to dryness.

### 2.2.4.3 Mass spectrometry

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis was carried out on an Eksigent 1D-NanoLC-Ultra system connected to a LTQ Orbitrap XL mass spectrometer (Thermo Scientific) equipped with a standard nanoelectrospray source. Fractions from size exclusion chromatography (SEC) were reconstituted in mobile phase buffer. The injection volume was chosen according to the 215 nm absorption signal from SEC separation. A fraction corresponding to an estimated amount of 1 μg was loaded onto a 11 cm x 0.075mm I.D. column pre-packed with Michrom Magic $C_{18}$ material (3 μm particle size, 200 Å pore size) (Michrom Bioresources, Inc.). Peptides

were separated at a flow rate of 300 nl/min using a stepwise gradient from 0.05% (v/v) to 92% (v/v) acetonitrile.

Ion source and transmission settings of the mass spectrometer were set as follows: Spray voltage 2 kV, capillary temperature 200°C, capillary voltage 60 V and tube lens voltage 135 V. The mass spectrometer was operated in data-dependent mode, selecting up to five precursors from a MS[1] scan (resolution 60,000) in a m/z range from 350-1600 for collision induced dissociation (CID). Single and double charged precursor ions and precursors of unknown charge states were rejected. CID was performed for 30 ms using 35% normalized collision energy and an activation of q=0.25. Dynamic exclusion was activated with a repeat count of 1, exclusion duration 30 s, list size of 300 and a mass window of ±50 ppm. Ion target values were 1,000,000 (or maximum 500 ms fill time) for full scans and 10,000 (or maximum 200 ms fill time) for MS/MS scans, respectively.

### 2.2.4.4 Database searching

For data analysis, Thermo Xcalibur .raw files were converted to the open mzXML format with the help of ReAdW version 4.3.1 using default settings. The mzXML files were used as input for xQuest searches. For the following MzXML2Search (part of Trans-Proteomics Pipeline, (Keller et al., 2005)) the files were converted to the .mgf (Mascot generic file) format. MzXML2Search was executed with the option "-T10000" to export precursors with a mass above the default value of 4,2000 Da.

Unmodified peptides from the protein mix were identified by searching an in-house Mascot server (ver. 2.3.0) against the Uniprot/SwissProt data with following parameters: Maximum number of missed cleavages = 2, taxonomy = chordata, fixed modifications = carbamidomethyl-Cys, variable modification = Met oxidation, MS[1] tolerance = 15 ppm, MS[2] tolerance = 0.6 Da, instrument type = ESI-TRAP and decoy mode set to on. For validation, the peptide probability was set to p < 0.05 and additional filters were used (require bold red = yes, peptide score ≥20).

Cross-linked peptides and peptide mono-links were identified using an in-house version of the dedicated search engine xQuest applying a scoring

model as described (Walzthoeni et al., in preparation). Tandem mass spectra of precursors differing in their mass by 12.075321 Da (mass difference DSS-$d_0$ and DSS-$d_{12}$) were paired if they had a charge state of 3+ to 8+ and were triggered within 2.5 min of each other. These spectra were then searched against a pre-processed .fasta database. For the protein mixture, the database contained the UniProt/SwissProt entries of the target proteins.

xQuest search parameters were set as follows: Maximum number of missed cleavages (excluding the cross-linking site) = 2, peptide length = 4-40 amino acids, fixed modifications = carbamidomethyl-Cys (mass-shift = 57.02146 Da), mass-shift of the light cross-linker = 138.06808 Da, mass-shift of mono-links = 156.078644 and 155.096428 Da, $MS^1$ tolerance = 15 ppm, $MS^2$ tolerance = 0.2 Da for common ions and 0.3 Da for cross-link ions. The search was carried out in enumeration mode (exhaustive search). The search results were filtered ($MS^1$ tolerance window = -4 to +7 Da) and all spectra were manually validated. Identifications were only considered for the result list, when both peptides had at least four bond cleavages in total or three adjacent ones and a minimum length of six amino acids.

## 2.2.5 Figure preparation

Figures were prepared with CHIMERA (Pettersen et al., 2004) and PYMOL (DeLano Scientific).

# 3 Results and Discussion

## 3.1 Purification of Pol I variants

### 3.1.1 Purification of Pol I ΔA49/A34.5

In order to confirm the assignment of EM densities near the Pol I funnel (see Chapter 3.3.1), cryo-EM studies with a Pol I variant lacking those subunits had to be performed. As the purification of Pol I from a GPY2 strain lacking the gene for subunit A34.5, which leads to a loss of A34.5 as well as A49, failed at the cation exchange step, we decided to dissociate the Pol I specific subunits by treatment with urea (Huet et al., 1975). Therefore the Pol I preparation was performed as described (Kuhn et al., 2007). After cation exchange chromatography the sample was diluted 5-fold with 25% MonoQ buffer B and dialyzed against Mono Q buffer C overnight at 4°C. Dissociated A49 and A34.5 could be removed from the sample by anion exchange chromatography (Figure 7).
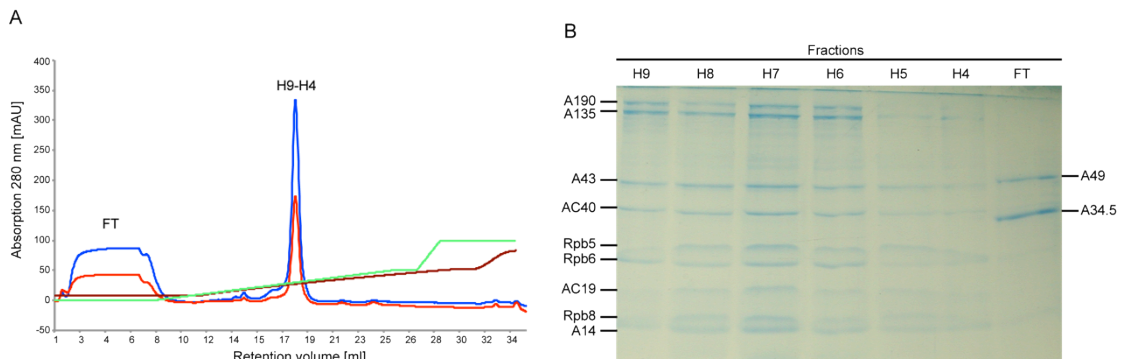


**Figure 7. Chemical dissociation of A49/A34.5.** (**A**) Chromatogram from anion exchange chromatography. Absorption at 280 and 260 nm are depicted as blue and red curve, respectively. Green and brown curves represent the concentration steps and conductivity, respectively. (**B**) SDS-PAGE analysis. The main peak (H9-H4) contains Pol I ΔA49/A34.5 while the Pol I-specific subunit stay in the flow-through (FT).

Urea was removed from the same by size exclusion chromatography and the peak fractions were concentrated and immediately used for experiments (Figure 8). The chemical dissociation yielded approximately 30-50 µg of Pol I ΔA49/A34.5.
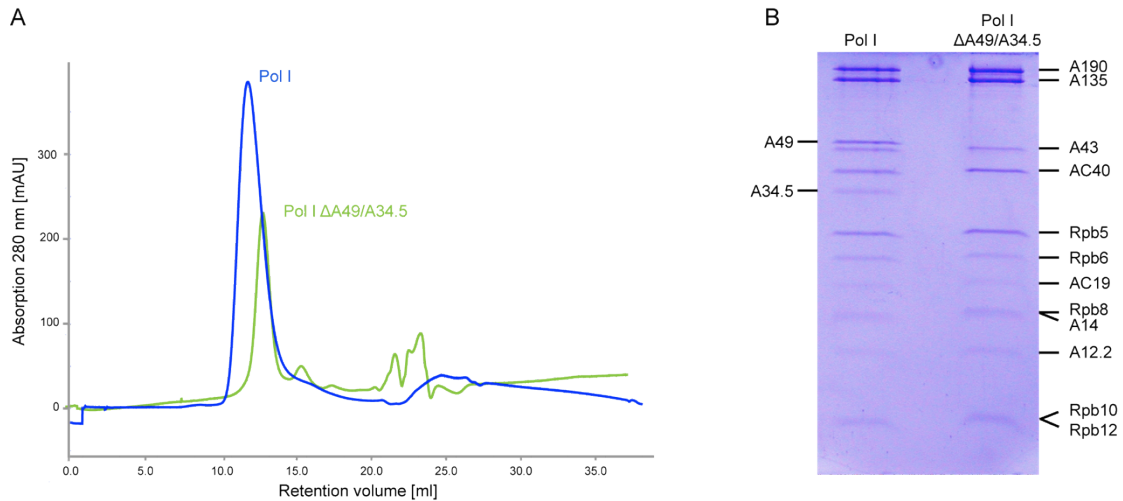
**Figure 8. Comparison of Pol I and Pol I ΔA49/A34.5.** (**A**) Size-exclusion chromatography of Pol I (blue) and Pol I ΔA49/A34.5 (lime). The peak of the Pol I variant lacking A49/A34.5 is shifted by approximately 1 ml. (**B**) SDS-PAGE analysis confirms the loss of the Pol I-specific subunits (right lane).

### 3.1.2  Purification of Pol I A12.2ΔC

Pol I lacking the C-terminal domain of subunit A12.2 was purified from a *Δrpa12* strain carrying a plasmid coding for the A12.2 N-terminus and linker region (amino acids 1-78) (see Chapter 2.1.1, Table 4). The purification was carried out exactly as for complete Pol I, except that gelfiltration was ommited (Figure 9).
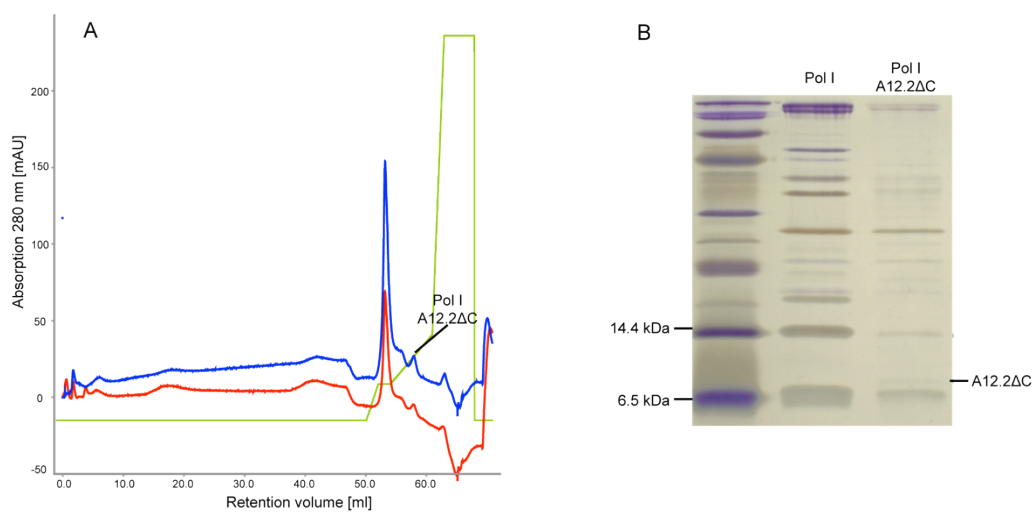


**Figure 9. Purification of Pol I A12.2ΔC.** (**A**) Cation exchange chromatography profile. The peak containing Pol I A12.2ΔC is labeled. (**B**) SDS-PAGE (silver stain) after cation exchange chromatography. A band for the truncated A12.2 variant appears at ≈ 8kDa.

The purification only yielded ≈ 10 µg Pol I A12.2ΔC from 240 g of yeast cells, which were harvested from approximately 50 L of yeast culture. The very low yield of this variant only allowed for usage in *in vitro* transcription assays. Cryo-EM and crystallization setups could not be carried out due to the lack of reasonable amounts of Pol I A12.2ΔC.

## 3.2 Crystallization of Pol I and attempts on structure solution

### 3.2.1 Crystallization of complete Pol I

Crystals of Pol I could be grown in reservoir solutions containing 100 mM HEPES pH 7.5, 3% (v/v) MPD, 9-11% PEG 6000 and 3 mM TCEP using freshly prepared Pol I at a concentration of 4-5.5 mg/ml. Crystals appeared after 3-4 days and grew to their maximum size within 7-10 days. However, those setups only yielded very small crystals with a maximum size of 50 -100 µm in their largest dimension. In order to increase the crystal size, and thus their potential diffraction power, initial crystals were used as starting material for streak-seeding experiments (Bergfors, 2003). Pre-equilibrated drops (equilibration time 3-4 h) were seeded using cat-whiskers (see Chapter 2.2.2.2) and a PEG 6000 concentration of 9% (w/v), whereas initial crystals mostly appeared at a PEG concentration of 10% (w/v). The seeding approach led to a substantial increase in crystal size up to dimensions of 500 µm x 70 µm x 20 µm for some single crystals (Figure 10A). Despite the increase in crystal size in some cases, streak-seeding did not lead to a better reproducibility. The crystal size as well as crystal morphology was dependent on a large variety of factors such as initial cell material, purification and equilibration time, quality and freshness of the seeds, the cat-whisker that was used etc. Even slightest changes in the crystallization procedure led to the disappearance of suitable crystals. Typically, crystallization setups with one Pol I batch derived from 300 g of yeast cells yielded 80-100 hanging drops, of which approximately 10 contained crystals suitable for diffraction experiments.

In order to overcome possible sub-stoichiometric binding of the Pol I specific subunits A49/A34.5 to the core, which had been observed in native MS experiments (Geiger et al., 2010), Pol I was pre-incubated with a 5-molar

excess of recombinantly expressed A49/A34.5 (see Chapter 2.2.1.6) for 1h on ice. After subsequent removal of unbound A49/A34.5 by gelfiltration crystallization experiments were set up. Ironically, this complementation led to a complete loss of Pol I crystals in initial setups as well as in drops that were subjected to streak-seeding.
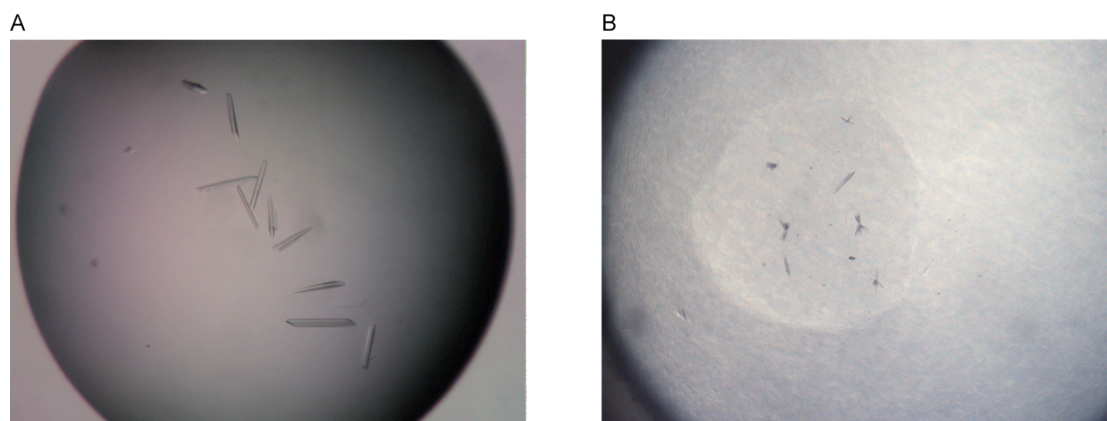


**Figure 10. Pol I crystals.** (**A**) Improved Pol I crystals after streak-seeding. (**B**) Pol I crystals obtained from screening setups at the MPI of Biochemistry. Crystals were grown in 20% PEG 3350 and 50 mM Tris pH 8.0.

Additionally, after the establishment of a state-of-the-art crystallization facility at the MPI of Biochemistry (Martinsried), Pol I was subjected to another round of initial screening using commercially available crystallization screens. Pol I crystallized in a condition containing 20% PEG 3350 and 50 mM Tris pH 8.0 (Figure 10B). However, the obtained crystals could not be reproduced in 24-well crystallization plates despite numerous optimization trials screening for different precipitant concentrations, protein concentrations, protein to precipitant ratios etc.

### 3.2.2 Crystallization of Pol I elongation complexes

Cryo-EM reconstructions (see Chapter 3.3.1) revealed a highly mobile clamp domain in Pol I, which could possibly interfere with successful crystallization and structure solution by X-ray diffraction. Therefore crystallization trials with Pol I were also carried out in the presence of nucleic acid scaffolds. To increase the rigidity of Pol I, minimal nucleic acid scaffolds were either soaked into existing Pol I crystals or co-crystallized with Pol I. For these trials two

different minimal scaffolds were used, one with a 9 bp DNA/RNA hybrid and a second one also with a 9 bp hybrid and a 6 bp overhang at the 5' end of the RNA (see Chapter 2.2.2.5).

For soaking experiments Pol I crystals were transferred from their source drops to a spot plate containing the original mother liquor supplemented with 10 µM nucleic acids (see Chapter 2.2.2.5). After back-soaking using mother liquor without nucleic acids, the soaking efficiency was checked by fluorescence microscopy showing that the 6-FAM labeled scaffolds were present in the crystals (Figure 11). However, Pol I crystals soaked with minimal nucleic acids scaffolds never showed ordered diffraction.
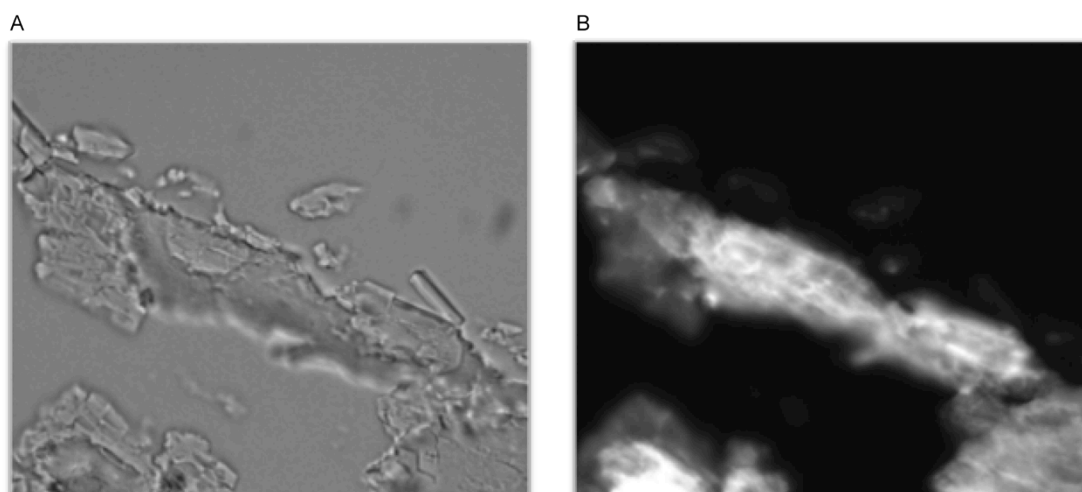


**Figure 11. Fluorescence microscopy of soaked Pol I crystals.** (**A**) Bright field image of a Pol I crystal soaked with 6-FAM labeled nucleic acids. (**B**) The same crystal viewed using an excitation filter at λ=500 nm (optimal excitation of 6-FAM at λ=494 nm), confirming that labeled nucleic acids are present in the crystal. (Crystal cracked due to use of a cover slide.)

As a second approach, Pol I was co-crystallized with the same nucleic acid scaffolds that were used for soaking experiments. Therefore Pol I was pre-incubated with a two-fold molar excess of nucleic acids (see Chapter 2.2.2.5) and subsequently used in initial screening setups. Co-crystallization experiments yielded initial crystals in different conditions containing either $CaCl_2$ or rather high concentrations of potassium acetate (Figure 12). Diffraction experiments using the PX Scanner at the MPI of Biochemistry revealed that all crystals grown in $CaCl_2$ containing conditions were salt-crystals. Crystals grown in potassium acetate containing conditions did not

exhibit any visible diffraction upon X-ray exposure, suggesting they might be bona fide Pol I crystals. However, the initial crystal could not be reproduced in 24-well hanging drop plates despite broad range screening.



0.1 M MES pH 6.5, 4M KOAc     0.1 M Bis-Tris pH 6.5, 45% MPD, 0.2 M CaCl2     0.1 M Bis-Tris pH 6.5, 30% PEG 550, 0.2 M CaCl2
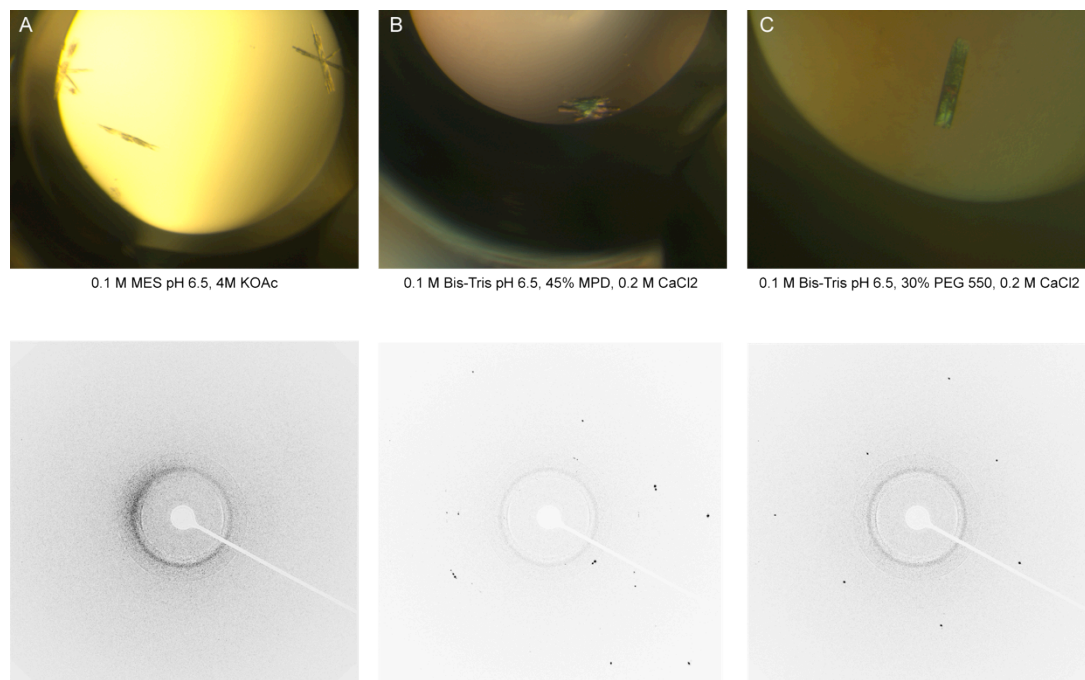
**Figure 12. Initial crystals of Pol I co-crystallized with nucleic acids.** The upper panel shows initial crystals and the lower panel their respective diffraction pattern (PX Scanner, MPI of Biochemistry). (**A**) Crystals grown in condition with high concentration of KOAc shows no diffraction on a weak X-ray source, leaving the possibility of being protein crystals. (**B**)(**C**) Crystals grown in conditions containing CaCl$_2$ show strong spots in diffraction experiments indicating salt crystals.

### 3.2.3  Crystallization of Pol I ΔA49/A34.5

In order to crystallize Pol I in a different crystal form, a few crystallization trials were set up using the Pol I variant lacking the specific subunits A49/A34.5. Drops were set up using a hand-made crystallization screen composed of known crystallization conditions for Pol II (4-7% (w/v) PEG 6000, 50 mM HEPES pH 7.0, 200 mM NH$_4$OAc, 300 mM NaOAc) and Pol I (8-12% PEG 6000 (see Chapter 3.2.1)). Some pseudo-crystals could be observed in the known Pol I crystallization condition (Figure 13) but they could not be optimized and transformed into crystals of reasonable quality for X-ray diffraction experiments. Initial large-scale screening could not be performed in

this case because the approach of chemically dissociating A49/A34.5 from Pol I only yielded 30-50 µg protein per purification.



**Figure 13. Pseudo-crystals of Pol I ΔA49/A34.5.** Formation of pseudo-crystals occured in crystallization conditions for 14-subunit Pol I at concentration of PEG 6000 of 9.5%.

### 3.2.4 Crystal manipulation and heavy atom derivatization

The diffraction quality of Pol I crystals remained one of the major issues throughout the trials of solving the Pol I structure by X-ray crystallography. Native Pol I crystals often showed only poorly ordered diffraction as well as a high degree of mosaicity and often anisotropic diffraction (Figure 14A). Crystal handling approaches such as crystal dehydration using increased PEG concentrations or cryo-annealing (Harp et al., 1999) did not improve the diffraction quality of Pol I crystals but in most cases led to a complete loss of diffraction (Table 17). Heavy atom derivatization, which had been exploited in order to gain experimental phase information, turned out to introduce order in the crystal lattice.

Various heavy atom compounds have been tested for their ability to improve diffraction quality of Pol I crystals (Table 17). The most positive effect on diffraction quality was observed when crystals were soaked with a $W_{18}$-cluster for 40-44h at 8°C (Figure 14B). All heavy atom compounds were added as grains to the final cryo-solution. The improved diffraction of crystals soaked with heavy atom compounds could be due to tighter crystal packing induced by these additives. In case of crystals soaked with $W_{18}$ the size of the unit cell was decreased by 1.5-2% in each dimension.

**Table 17. Post-crystallization manipulation of Pol I crystals.**

| Manipulation method | Component | Protocol | Result |
|---|---|---|---|
| $W_{18}$-cluster soak | $(NH_4)_6(P_2W_{18}O_{62}) \cdot 14\ H_2O$ | Soaking for 40-44 h at 8°C | Best diffraction up to 4-4.5 Å |
| $W_{18}$ back-soak | $(NH_4)_6(P_2W_{18}O_{62}) \cdot 14\ H_2O$ | Soaking for 40-44 h, backsoaking for 10-15 min in final cryo-solution | Same as normal $W_{18}$-soak, same radiation damage |
| $Ir_3$-soak | $(NH_4)_4Ir_3N(SO_4)_6$ | 24/48 h at 8°C | No diffraction |
| Cs-Os soak | $Cs_3Os_2NCl_8 \cdot 2\ H_2O$ | 24/48 h at 8°C | Diffraction to 8-9 Å |
| Cs-W soak | $Cs_6W_5P_2O_{23}$ | 24/48 h at 8°C | Non-ordered diffraction up to 7 Å |
| Pt soak | $K_2PtCl_4$ | 24/48 h at 8°C | No diffraction |
| Mersalyl acid soak | $C_{13}H_{18}HgNO_6$ | 1 week at 8°C | No diffraction |
| $W_{30}$-Cluster soak | $(NH_4)_{14}[NaP_5W_{30}O_{10}]$ | 24/48 h at 8°C | No diffration |
| Tantalum-Bromide soak | $Ta_6Br_{12}^{2+}$ | 1-3 h @ 20°C or overnight at 8°C | Very poor diffraction |
| Tungsten-Bromide soak | $W_6Br_{12}^{2+}$ | 1-2 h or overnight at 8°C | No ordered diffraction beyond 6-7 Å |
| Mercury soak | 2,4,6-Trisaeceto-(3-aminoaceto)-mercuritoluol | 30 min – 3 h @ 20°C | No diffraction |
| Cryo-annealing | --- | Crystal thawed and re-frozen in cryo stream | No diffraction |

| Manipulation method | Component | Protocol | Result |
|---|---|---|---|
| Dehydration | 25% PEG 6000 | Crystal soaked in final cryo solution containing 25% PEG 6000 for 15 min at 20°C before freezing in liquid $N_2$ | No diffraction |
| Crystal pre-cooling | | Pre-cool for 24h at 4°C before freezing | Slightly better diffraction than native crystals (5.5 – 6 Å) |
| Crystal freezing in cryo stream | | | Very weak diffraction |
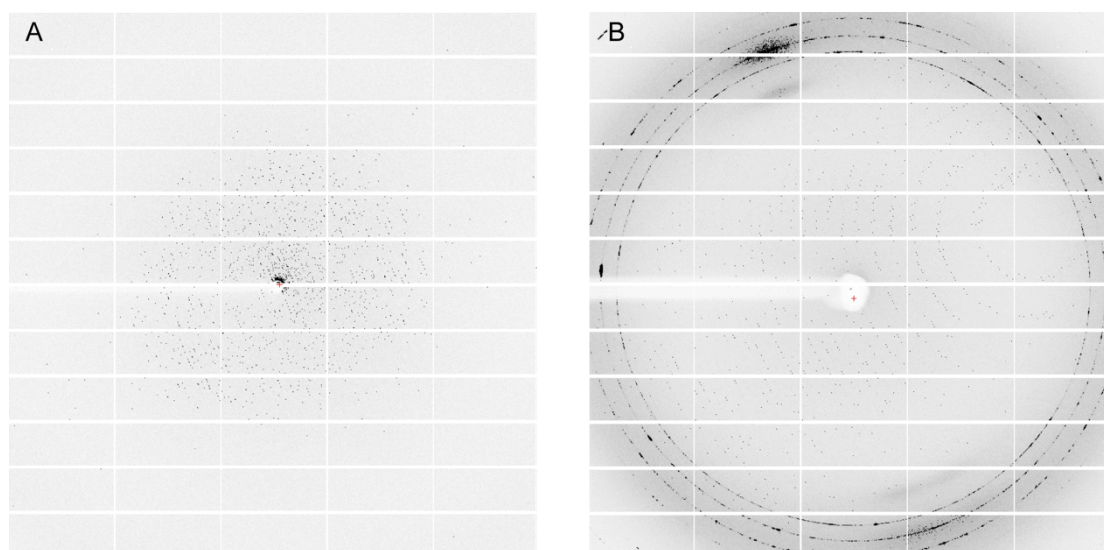


**Figure 14. Diffraction patterns of native and soaked Pol I crystals.** (**A**) Native Pol I crystal (SJ83) shows poorly ordered diffraction and anisotropy. (**B**) Pol I crystal soaked 44 h with $W_{18}$-cluster (SJ4) shows nicely ordered diffraction up to 4 Å.

### 3.2.5  Data collection and processing

All datasets were collected at the SLS (Swiss Light Source) on beamline X06SA. If the crystals were soaked with heavy atoms, X-ray absorption scans were performed for every heavy atom species to test whether the compound

was present in the crystals. Due to severe radiation damage the crystals had to be translated several times in order to record complete datasets at reasonable resolutions. Since it was shown that datasets with a high completeness at very low resolution can facilitate structure determination by molecular replacement (Jenni and Ban, 2009) datasets for Pol I crystals were also recorded at low resolution (low resolution limit 300-400 Å) (Figure 15).

Data were processed using XDS (Kabsch, 1993) and resulted in reasonable statistics to a resolution of 3.9 Å in the best case (Figure 14B). The resulting unit cell was monoclinic with a C2 symmetry and unit cell dimensions of a = 617 Å, b = 306 Å, c = 251 Å and $\beta$ = 97.3°. Anomalous signal (SigAno >1.2) of the soaked $W_{18}$-cluster could only be observed in the lowest resolution shells indicating that the heavy atom clusters were not bound to the protein but rather diffusing in the solvent channels of the crystal. This, in turn, impaired experimental phasing and only left the possibility of molecular replacement approaches in order to determine the Pol I structure (see Chapter 3.2.7).



SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= -3.0 AS FUNCTION OF RESOLUTION

| RESOLUTION LIMIT | NUMBER OF REFLECTIONS OBSERVED | UNIQUE | POSSIBLE | COMPLETENESS OF DATA | R-FACTOR observed | R-FACTOR expected | COMPARED | I/SIGMA | R-meas | Rmrgd-F |
|---|---|---|---|---|---|---|---|---|---|---|
| 200.00 | 5 | 4 | 4 | 100.0% | 11.9% | 15.2% | 2 | 4.12 | 16.8% | 12.0% |
| 150.00 | 10 | 6 | 6 | 100.0% | 25.3% | 17.4% | 7 | 4.66 | 31.6% | 25.1% |
| 75.00 | 177 | 60 | 61 | 98.4% | 16.2% | 17.8% | 166 | 6.05 | 19.2% | 10.9% |
| 50.00 | 536 | 148 | 148 | 100.0% | 17.9% | 18.5% | 530 | 6.72 | 20.9% | 8.2% |
| 25.00 | 5765 | 1470 | 1471 | 99.9% | 17.7% | 19.4% | 5759 | 6.32 | 20.4% | 9.9% |
| 20.00 | 5835 | 1590 | 1591 | 99.9% | 19.1% | 20.2% | 5822 | 5.63 | 22.2% | 13.4% |
| 15.00 | 15475 | 4364 | 4384 | 99.5% | 21.4% | 21.6% | 15405 | 5.03 | 25.1% | 17.4% |
| 10.00 | 67665 | 17892 | 17926 | 99.8% | 24.6% | 23.7% | 67454 | 4.86 | 28.6% | 19.4% |
| 9.00 | 32913 | 9359 | 9391 | 99.7% | 31.6% | 29.3% | 32755 | 3.89 | 37.3% | 30.8% |
| 8.00 | 54844 | 14715 | 14752 | 99.7% | 42.0% | 40.3% | 54639 | 3.26 | 48.9% | 38.5% |
| 7.00 | 92810 | 24255 | 24281 | 99.9% | 68.6% | 67.3% | 92591 | 2.11 | 79.7% | 69.6% |
| 6.00 | 148406 | 43163 | 43246 | 99.8% | 119.4% | 118.2% | 147657 | 1.16 | 141.4% | 146.1% |
| total | 424441 | 117026 | 117261 | 99.8% | 39.7% | 38.7% | 422787 | 2.68 | 46.4% | 51.8% |

**Figure 15. Low resolution measurements and data statistics.** (**A**) Diffraction pattern of low resolution measurements (crytsal SJ110) shows spots down to >300 Å resolution. (**B**) Processing with XDS results in a complete dataset from 310 – 7 Å.

### 3.2.6 The Pol I asymmetric unit comprises seven Pol I molecules

Calculating the Matthews coefficient (Matthews, 1968) with data observed for Pol I crystals suggested 7-9 copies of Pol I within the asymmetric unit and a solvent content of 45-55% (Figure 16A). When calculating the locked self-rotation function (Tong and Rossmann, 1990) a very strong peak is observed at $\kappa$ = 51° (Figure 16B) and 7 equally strong peaks are visible in the stereographic projection at $\kappa$ = 180° (Figure 16C). This strongly argues for an

asymmetric unit comprising seven Pol I molecules related by a 7-fold non-crystallographic symmetry axis (NCS) and 14 2-fold axes perpendicular to the 7-fold axis. The 7-fold axis is oriented along c while the 2-fold axes, one each 25.7°, lie near the ab plane. Taken together these results demonstrate that in the present crystal form the asymmetric unit likely contains 7 Pol I molecules with a total mass of almost 4.2 MDa.
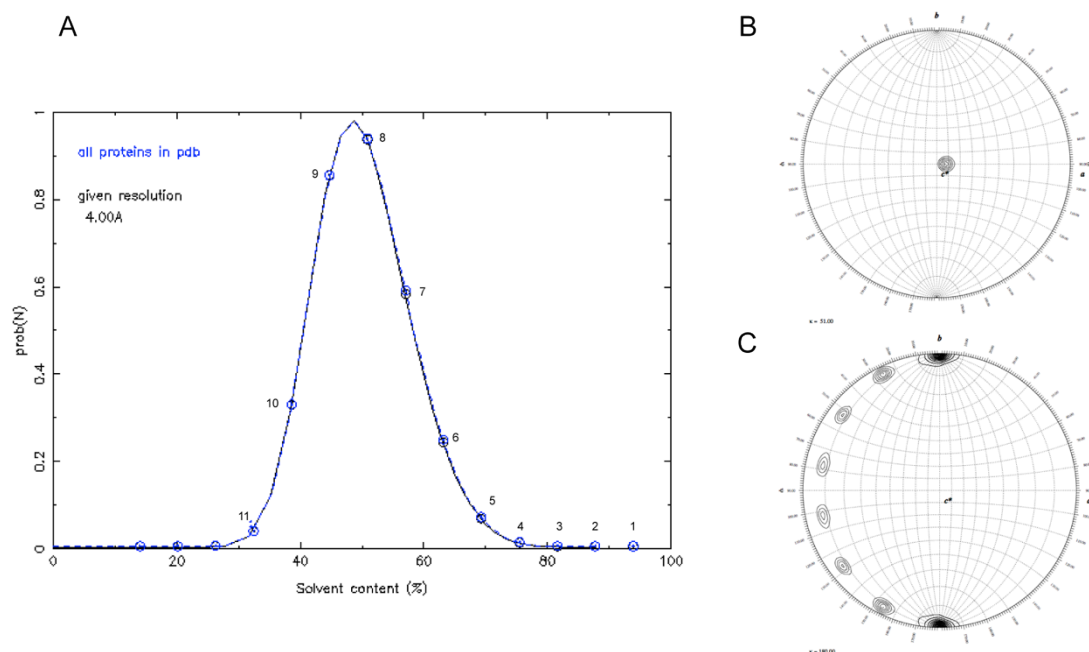


**Figure 16. Organization of the Pol I asymmetric unit.** (**A**) Calculation of the Matthews coefficient suggests 7-9 Pol I molecules in the asymmetric unit. (**B**) Stereographic projection of the self-rotation function at κ = 51° shows a 7-fold NCS axis along c. (**C**) Stereographic projection of the self-rotation function at κ = 180° shows 7 2-fold NCS axes in one polar coordinate hemisphere resulting in a total number of 14 2-fold NCS axes near the ab plane. The peaks at the poles are symmetry-equivalent.

### 3.2.7 Attempts on structure determination by molecular replacement

Initial molecular replacement attempts were carried out using Pol II-based based search models, using either the complete 12-subunit Pol II structure (Armache et al., 2005) or Pol II bound to TFIIS (Kettenberger et al., 2003). Additionally, molecular replacement trials using the Pol I EM density (see Chapter 3.3.1) were carried out employing different programs such as MOLREP (Vagin and Teplyakov, 1997), amore (Navaza, 1994), Phaser (Mccoy et al., 2007) and Replace (Tong, 1993). However, molecular

replacement attempts with crystallographic search models never resulted in groups of seven rotational solutions representing the predicted 7-fold axis in the asymmetric unit. In order to use EM maps as search model for molecular replacement, P1 unit cells containing one copy of the EM volume were generated. The structure factors for this model were calculated and used for molecular replacement. Calculation of the rotation function with this model resulted in 7 solutions related by a 7-fold axis. However, the translation function for the obtained rotation solutions could not be solved.

In subsequent molecular replacement trials, search models were introduced that already contained copies of 7 seven Pol I molecules in order to restrain the translational search (Figure 17A). These ring-shaped models were constructed by counterclockwise rotation of the Pol I EM volume based on peaks observed by applying the locked rotation function (Figure 17B).
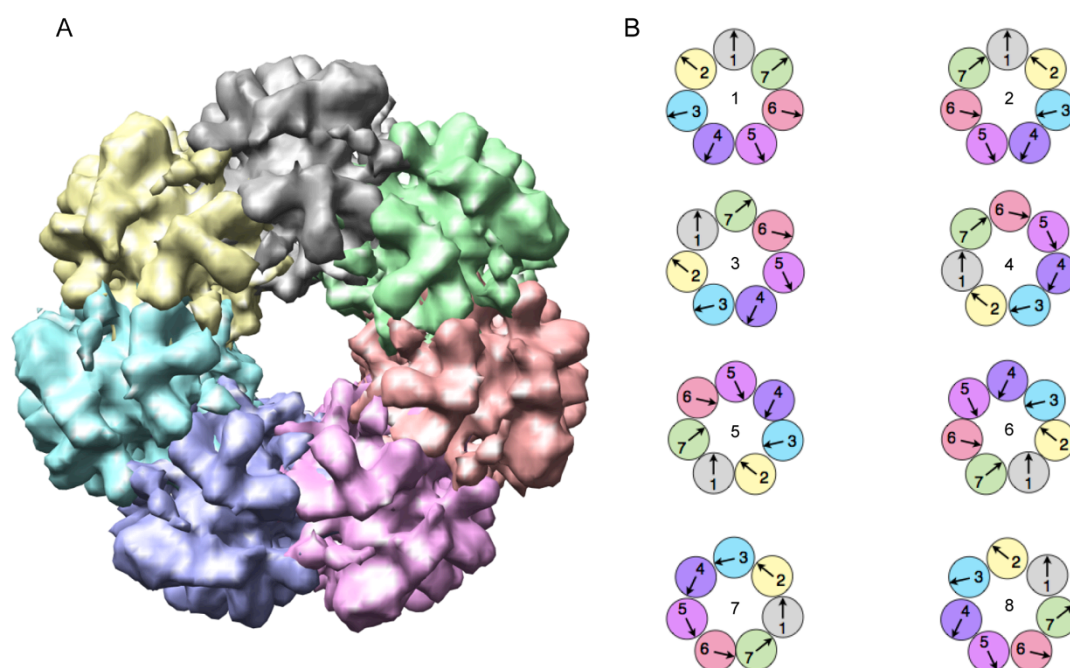


**Figure 17. Heptamer ring models for molecular replacement.** (**A**) Potential 7-mer ring based on Pol I cryo-EM reconstructions (Chapter 3.3.1). (**B**) Counterclockwise rotation of the Pol I EM density preserves the internal 7-fold NCS (1) while clockwise rotation does not (2). Schemes 3-8 show all possible start points of the 7-mer rings derived from counterclockwise rotation based on locked rotation function peaks.

Approximately 22,000 ring models varying in diameter and molecule positions were employed in extensive grid searches in order to obtain a solution of the

translation function and, thus, to gain initial phase information. Despite these exhaustive efforts, no translational solution could be obtained using available diffraction data of Pol I. An even more extensive approach of allowing internal movement (e.g. rotation) of single Pol I molecules within the heptameric search model was not pursued because the number of searchable models would grow exponentially so that this approach was limited by the available computing power. However, if it would be possible to solve this large asymmetric unit one could largely benefit from phase improvement by 7-fold NCS-averaging.

## 3.3  Cryo-electron microscopic studies on Pol I

### 3.3.1  Cryo-EM structure of Pol I at 12 Å resolution

For cryo-EM studies, Pol I was purified as described (see Chapter 2.2.1.1) and kept on ice until further usage. The optimal protein concentration for cryo-EM was determined by electron microscopy using negative stain (Figure 18A). Particles did not form aggregates and showed high particle density. Even under cryo-conditions ($\approx$ 100 K) particles could be easily identified and apparently behaved nicely during vitrification (Figure 18B). Cryo-EM reconstruction of Pol I with 46,056 particles led to a map at 11.9 Å resolution (Figure 18C) as judged by the Fourier shell correlation (FSC) function plot, applying a cutoff value of FSC = 0.5.
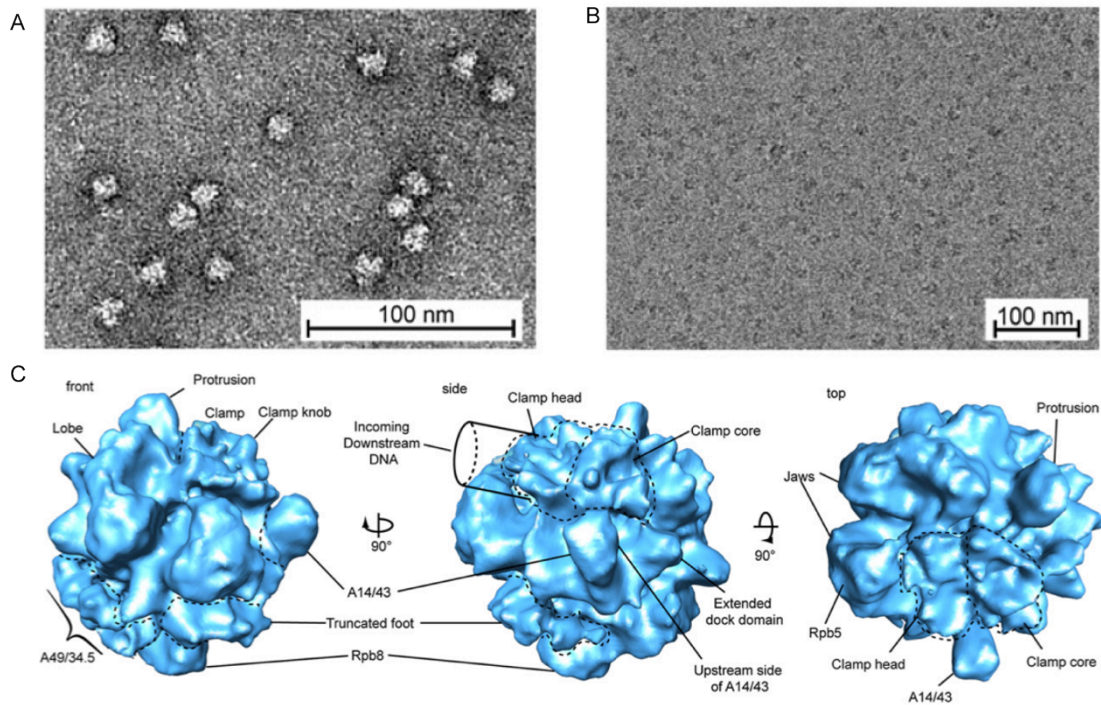
**Figure 18. EM reconstruction of Pol I.** (**A**) Negatively stained Pol I. (**B**) Pol I variant ΔA49/A34.5 under cryo conditions. (**C**) Cryo-EM reconstruction of Pol I. Views and structural regions are named according to the Pol II structure (Cramer et al., 2001).

Interpretation of the EM map was achieved by first placing the crystal structure of the Pol II 10-subunit core into the EM map as a rigid body by fitting the common five subunits, which were known to occupy similar positions on the polymerases' surface (Jasiak et al., 2006). A perfect fit of the common subunits confirmed the high quality of the map (Figure 19). Many regions of the homologous subunits fitted equally well, but strong deviations were also observed.
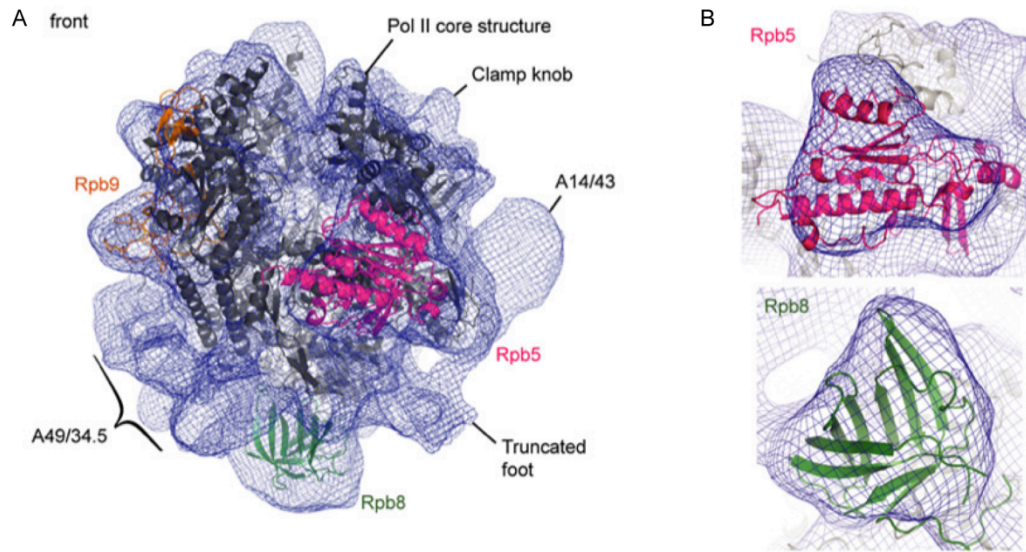
**Figure 19. Placement of the 10-subunit Pol II core into EM density.** (**A**) Placement of the Pol II core structrue (Armache et al., 2005) (grey) into the EM density (blue). Subunits Rpb5, Rpb8 and Rpb9 are highlighted in magenta, green and organge respectively. (**B**) Fit of the common subunits Rpb5 and Rpb8 into the EM map.

### 3.3.2  Cryo-EM structure of Pol I ΔA49/A34.5

After assigning EM-densities to the Pol I core, an additional large density remained on the enzyme surface that was assigned to the Pol I-specific subunits A49 and A34.5 (Figure 19A). To confirm this assignment, subunits A49 and A34.5 were dissociated from Pol I using urea (Huet et al., 1975) (see Chapter 2.2.1.2) and the resulting 12-subunit Pol I was purified and its structure was solved by cryo-EM at 25 Å resolution (Figure 20). The structure was similar to the complete Pol I, except that the density assigned to A49 and A34.5 was lacking. In addition, there was a minor change in the clamp conformation, which probably represents an average clamp position, and is unlikely to result from the absence of A49/A34.5. Density assigned to A49 and A34.5 is located near the enzyme funnel, the external domain 1, a conserved core loop with a Pol I-specific insertion (corresponding to loop α16-β20 of the Pol II pore domain), and A12.2. This is consistent with the loss of A49 when Pol I is purified from A12.2 deletion strains (Van Mullem et al., 2002).
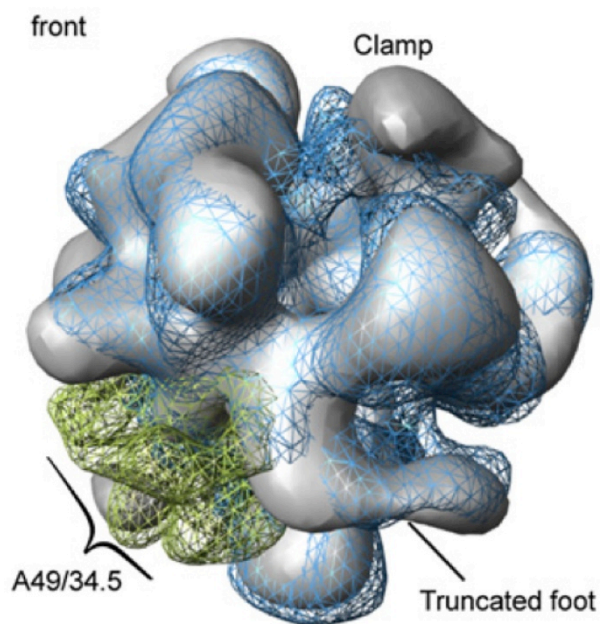
**Figure 20.** Overlay of EM structures of Pol I ΔA49/A34.5 (silver surface) and the complete Pol I (blue). The density assigned to A49/A34.5 is highlighted green.

## 3.4  Functional studies on Pol I

### 3.4.1  A49/A34.5 acts as a build in elongation factor

Bioinformatic analysis using HHPred (Soding et al., 2005) revealed homologies between A49/A34.5 and N-terminal regions of the two large subunits of human TFIIF and thus suggested that A49/34.5 has elongation stimulatory activity. Therefore, complete Pol I and Pol I ΔA49/A34.5 were compared in an RNA extension assay using a minimal DNA-RNA scaffold (see Chapter 2.2.3.1). The complete Pol I extended the RNA to the end of the template, whereas Pol I ΔA49/34.5 did not produce the run-off product (Figure 21A). Addition of recombinant A49/34.5 rescued the defect of Pol I ΔA49/34.5, and enabled elongation to the end of the template (Figure 21A, lane 4). The elongation experiments were repeated using a complete, complementary transcription bubble scaffold (Fig. 21B) (Kireeva et al., 2000). The complete Pol I produced the run-off transcript (+18), whereas Pol I ΔA49/34.5 did not, but addition of recombinant A49/34.5 heterodimer restored run-off formation (Fig. 21B, lanes 6+7). The defect was not due to differential binding of the polymerase variants to the scaffold, as it was also observed when the elongation complexes were covalently coupled to magnetic beads and

extensively washed before the reaction (not shown). Reduced elongation activity in the fully complementary system arises from a more sophisticated complex assembly, resulting in a higher proportion of RNA not bound to Pol I. Taken together, A49/34.5 is required for normal elongation activity of Pol I *in vitro.*
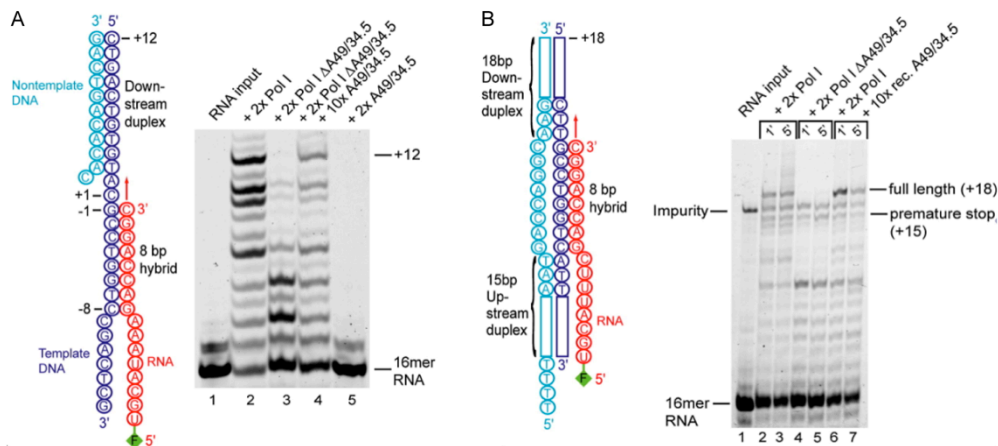


**Figure 21. Elongation stimulatory activity of A49/A34.5.** (**A**) A49/34.5 shows elongation-stimulatory activity in RNA extension assays with a minimal nucleic acid scaffold. The fluorescent label 6-carboxy-fluoresceine (FAM) on the RNA 5'-end is indicated. The times molar excess of added factors are indicated above the lanes. For lane 4, Pol I ΔA49/34.5 was complemented with a fivefold molar excess of recombinant A49/34.5 for 10 min at 20 °C prior to addition of the scaffold. (**B**) Elongation assay as in (A) but with a complete complementary bubble (Kireeva et al., 2000).

### 3.4.2 Pol I has intrinsic cleavage activity that requires A12.2

The active site of Pol II exhibits weak 3'-RNA cleavage activity that is stimulated by TFIIS (Wind and Reines, 2000). For Pol I, an RNAse H-like nuclease activity was initially described (Huet et al., 1976), but was later found to reside in a dissociable factor (Huet et al., 1977; Tschochner, 1996). To clarify whether Pol I possesses intrinsic RNA cleavage activity, a "backtracked" elongation complex was assembled using purified Pol I and a DNA-RNA scaffold that contained an RNA 3'-overhang (Figure 22 and Chapter 2.2.3.2). Incubation of the backtracked complex with 8 mM magnesium ions led to efficient shortening of the RNA from the 3'-end (Figure

22B, lanes 1-3). In more detail, Pol I mainly removed four nucleotides from the RNA, consistent with binding of the terminal hybrid base pair to the nucleotide insertion site (+1), extrusion of the RNA 3'-overhang into the polymerase pore, and cleavage of the phosphodiester bond between nucleotides at positions –1 and +1. In comparison, Pol II was unable to cleave the RNA under these conditions, but addition of TFIIS resulted in cleavage (Figure 22B, lanes 8-11). The Pol II-TFIIS complex removed three or four nucleotides, indicating that a mixture of complexes was present with the terminal hybrid base pair occupying either position -1 or +1. Taken together, Pol I has a strong intrinsic RNA cleavage activity not present in Pol II.
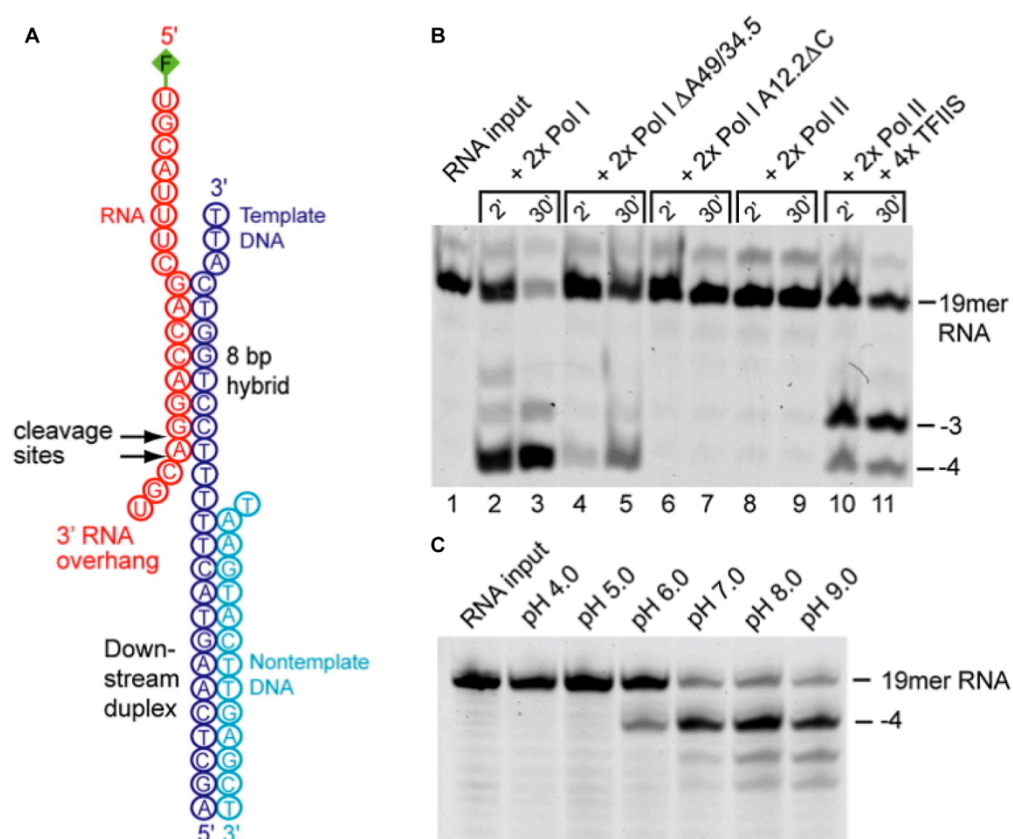


**Figure 22. Intrinsic RNA cleavage activity of Pol I.** (**A**) DNA-RNA hybrid scaffold used in cleavage assays. (**B**) Comparison of RNA cleavage by Pol I variants with Pol II and the Pol II-TFIIS complex. (**C**) pH-Dependence of Pol I cleavage activity.

The intrinsic cleavage activity likely escaped detection previously since the nucleic acid substrates used in published studies did not allow for the formation of a backtracked state, from which cleavage occurs. The previously

described dissociable factor (Huet et al., 1977; Tschochner, 1996) may not be required for cleavage *per se,* but may induce backtracking of Pol I, to create a state of the elongation complex that is prone to cleavage.

Additional cleavage assays showed that the Pol I variant ΔA49/34.5 cleaved RNA less efficiently than the complete Pol I (Fig. 22B, lanes 4+5). Cleavage stimulation by A49/34.5 is consistent with an early investigation of an RNAse H-like activity in Pol I-containing fractions (Huet et al., 1976) (Huet et al., 1976). The question arose, whether subunit A12.2 is required for cleavage, since its counterpart C11 is essential for cleavage activity of Pol III (Chedin et al., 1998). A Pol I variant lacking residues 79-125 of A12.2 (A12.2ΔC, Chapter 3.1.2) was totally inactive in RNA cleavage (Fig. 22B, lanes 6+7), but bound the nucleic acid scaffold in electrophoretic mobility shift assays, and retained elongation activity (not shown). Consistent with a function specific for the A12.2 C-terminal domain, a truncation variant remains bound to Pol I and does not show a conditional growth defect (Van Mullem et al., 2002).

The A12.2 C-terminal domain shows homology to the TFIIS C-terminal domain that inserts into the Pol II pore to stimulate RNA cleavage (Kettenberger et al., 2003). Mutation of the residues in A12.2 homologous to the catalytical D290 and E291 in TFIIS leads to a lethal phenotype, demonstrating their importance (not shown, data by Jochen Gerber, Regensburg). The conserved polymerase active site is capable of RNA cleavage in the absence of cleavage stimulatory factors, since free Pol II and the bacterial RNA polymerase can cleave RNA under mild alkaline conditions (Orlova et al., 1995; Weilbaecher et al., 2003). Consistently, the intrinsic cleavage activity of Pol I increased with increasing pH (Figure 22C).

## 3.5 Crosslinking reveals Pol I domain architecture and positions the Pol I-specific initiation factor Rrn3

### 3.5.1 Crosslinking-MS analysis of Pol I

To address unresolved questions on the functional architecture of Pol I, crosslinking-MS analysis were carried out for Pol I from the yeast *Saccharomyces cerevisiae*. Endogenous Pol I was purified as described (Gerber et al., 2008; Kuhn et al., 2007), except that the final size-exclusion chromatography step was carried out in the presence of potassium acetate (see Chapter 2.2.1.1). For crosslinking, 110 μg of Pol I were incubated with isotope-labeled disuccinimidyl suberate (DSS, Creative Molecules Inc.). DSS reacts with primary amines in lysine side chains and protein N-termini. Crosslinking was monitored by SDS-PAGE (Fig. 23A), and polymerase oligomers were removed by gelfiltration (see Chapter 2.2.4.1). After digestion with trypsin, crosslinked peptides were enriched by size exclusion chromatography, and peptides and their fragments were detected with high-resolution MS (see Chapter 2.2.4.2 - 2.2.4.4). Measurement of the four samples, of which two were Pol I-Rrn3 complexes, resulted in 1047 mass spectra that matched crosslinked peptides, an example of which is shown in Figure 23B.

### 3.5.2 Confirmation of the Pol I core model

239 unique linkage pairs were identified within nine subunits of the Pol I core (subunits A190, A135, AC40, AC19, Rpb5, Rpb6, Rpb8, Rpb10, Rpb12; excluding A12.2, A49/34.5, and A14/43) (Figure 24A). The crosslinked residues were analyzed with the atomic Pol II structure (Armache et al., 2005) and the Pol I core homology model (Kuhn et al., 2007). We assumed that the distance between Cα atoms of crosslinked lysines must be ≤30 Å, corresponding to the length of the DSS spacer (11.4 Å) plus two times the length of a lysine side chain (6.5 Å) plus an estimated coordinate error of 3 Å for flexible lysine side chain ends. Crosslinking sites that fell in regions that adopt the Pol II fold (Kuhn et al., 2007) were assigned to category A (Table 18). Sites outside these regions were assigned to category B.
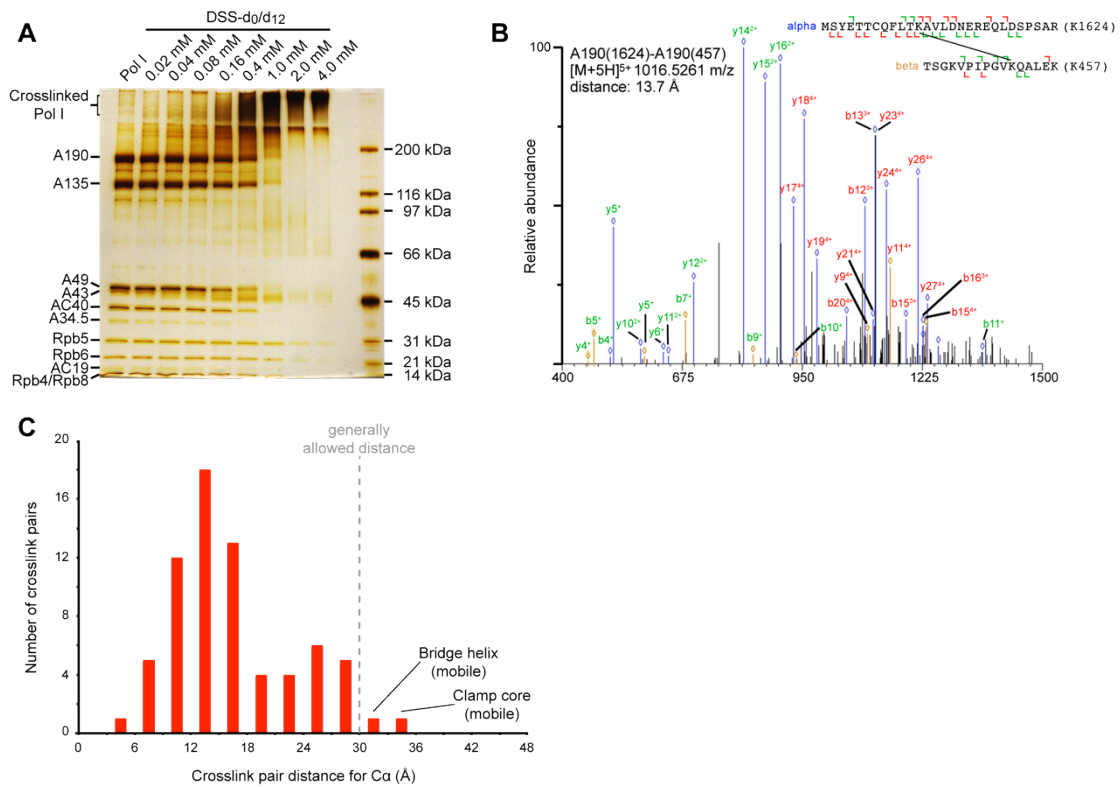
**Figure 23. Crosslining-MS analysis of Pol I.** (**A**) SDS-PAGE analysis of Pol I crosslinked with different concentrations of DSS. (**B**) Fragmentation spectrum of a crosslinked peptide. The linkage site A190 K1624-A190 K457 was observed in the crosslinked peptide MSYETTCQFLTK(xl)AVLDNEREQLDSPSAR/TSGKVPIPGVK(xl)QALEK (m/z 1016.5261, 5+; 'xl' denotes the crosslinked lysine). Extensive ion series for both peptides are observed in the fragmentation spectrum providing high confidence in the match. Peaks of the α- and β-peptide are colored blue and ocher, respectively. Common ions are labeled in green and crosslink ions are labeled in red. (**C**) Cα distance distribution for experimentally observed A-A linkage pairs within the Pol I 9-subunit core (subunits A190, A135, AC40, AC19, Rpb5, Rpb6, Rpb8, Rpb10, and Rpb12). The generally allowed distance between the Cα atoms of two crosslinked lysine residues of 30 Å is indicated by a dashed line. Observed crosslinks are in agreement with the homology model for the Pol I core as judged by analysis with the Pol II X-ray structure (PDB 1Y1V).

Of the 239 crosslink pairs, 73 (30.5%) comprised only category A sites (A-A pairs), 80 (33.5%) contained one category B site (A-B pairs), and 86 (36%) were B-B pairs. Of the 73 A-A pairs, 70 could be analyzed (Kettenberger et al., 2004), because both crosslinked residues were present in the structure. Among the 70 A-A cross-link pairs, 68 (97.1%) fell within the acceptable Cα distance of ≤30 Å (Figure 23C). The two remaining pairs exceeded the maximum distance by only 1.9 Å and 4.3 Å, respectively, and this can be

explained by structural flexibility. One crosslink involves the bridge helix, which undergoes conformational changes (Brueckner et al., 2009; Cramer et al., 2001; Gnatt et al., 2001), whereas the other involves the clamp, which is mobile in Pol I (Kuhn et al., 2007) and Pol II (Cramer et al., 2001; Gnatt et al., 2001; Kostrewa et al., 2009). We additionally obtained crosslinks within the A14/43 subcomplex, and between A14/43 and the Pol I core, which were consistent with the previously obtained structure and location of A14/43 (Geiger et al., 2008; Kuhn et al., 2007). These results demonstrate the validity of our method, and confirm the previous Pol I model.

**Table 18. Crosslinking statistics and classification**

| Crosslink pair type | Observed pairs | Both sites present in PDB 1Y1V | Distance ≤30 Å | Comment |
|---|---|---|---|---|
| A-A | 73 | 70 | 68 (97.1%) | 2 outliers involve mobile regions (bridge helix, clamp) |
| A-B | 80 | 65 | 39 (60.0%) | 22 outliers comprise altered jaw residues[1] |
| B-B | 86 | 30 | 13 (43.3%) | 13 outliers comprise altered jaw residues[1] |
| B*-B[2] | 23[2] | 17[2] | 2 (11.8%) | 11 outliers comprise altered jaw residues[1] |
| Total | 239 | 165 | 120 (72.7%) | |

[1] The jaw region comprises A190 residues 1252-1487

[2] Crosslinks are a subgroup of B-B pairs ans therefore do not contribute to total number

### 3.5.3 Model extension reveals a unique jaw

To extend the Pol I model, category A-B and B-B crosslink pairs were analyzed (Table 18, Figure 24A). A-B pairs connect residues in regions of the homology model that share the Pol II fold (category A) with residues in sequence regions with no or very weak conservation (category B). Within the nine core subunits, we observed 80 A-B crosslinks, of which 15 could not be analyzed as they contain residues within specific insertions or residues that are not present in the Pol II structure. A total of 39 A-B pairs (60%) showed Cα distances below 30 Å. These residues were reclassified as B*, and their

surrounding region (overall 125 residues within A190, A135 and AC40) was included in the Pol I homology model, as it was likely that the region containing the B* site adopts a Pol II-like fold. Of the resulting 23 B*-B crosslinks, 17 pairs contained lysine residues present in the Pol II structure and two showed a permissible Cα distance. Based on these findings  the Pol I core model was extended to parts of the clamp core, pore, funnel, foot, dock, and cleft domains of the largest subunit A190, to small parts of the lobe, fork, and wall domains of the second largest subunit A135, and to parts of domain 2, the loop domain, and the dimerization domain in subunit AC40, the counterpart of the Pol II subunit Rpb3 (Figure 24A, compare VI.1). The extended homology model relates 69.6% of the nine core Pol I subunit sequences to their Pol II counterparts, although the large Pol I subunits A190 and A135 show only 25.5% and 25.6% sequence identity to their Pol II counterparts, respectively.

Of the 166 A-B and B-B crosslinks within the Pol I core, 95 could be analyzed with the Pol II structure, of which 43 (26 A-B and 17 B-B crosslinks) did not fall within the Cα distance restraint of 30 Å (Table 18). Of these cross-links, 35 involve residues that fall in a region that may correspond to the Pol II jaw domain (A190 residues 1252-1487). This region is conserved among yeast species, but is poorly conserved in higher eukaryotes. Secondary structure prediction (Jones, 1999) indicates that this region is only partially related to the Pol II jaw. Residues 1251-1337 are predicted to form three helices and two β-strands, residues 1338-1438 are apparently unstructured, and residues 1439-1495 may form two helices and two β-strands. Crosslinks to residues K1260, K1269, K1473 and K1495 can be explained if residues in the two structured regions (1251-1337 and 1439-1495) adopt a structure and location similar to the Pol II jaw. In the unstructured region, residues K1363 and K1376 form 32 crosslinks to regions in the polymerase cleft, including the bridge helix (Figure 24B). These crosslinks indicate that Pol I has a unique jaw domain with N- and C-terminal regions similar to the jaw domain in Pol II and an additional, mobile middle part that extends along the active center cleft.
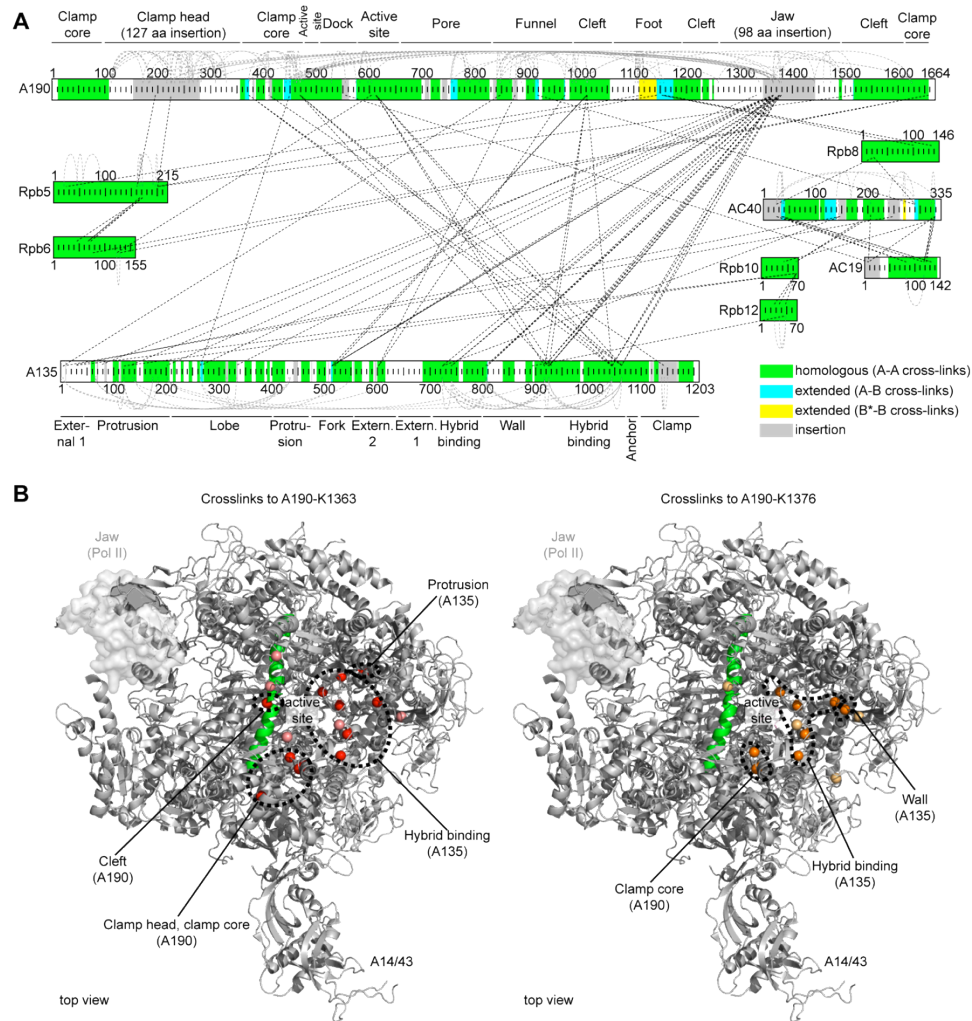
**Figure 24. Crosslink map and extended homology model for the Pol I core.** (**A**) Crosslink map of the Pol I core. The primary structure of nine core subunits are shown schematically as boxes. Regions that show a conserved fold in Pol II are in green, insertions with respect to Pol II are colored in gray, and poorly or non-conserved parts are white. Extensions of the homology model derived from crosslink data are indicated in cyan and yellow for A-B and B*-B crosslinks, respectively (compare VI.1). Black dashed lines and gray dashed arcs indicate inter-subunit and intra-subunit crosslinks, respectively.

(**B**) Crosslinks of residues K1363 and K1376 of A190. Lysine residues in the polymerase core crosslinked to K1363 are in red (category A crosslink sites) or salmon (category B* sites) and the respective Cα atoms are shown as spheres. Cα atoms of lysine residues crosslinked to K1376 are shown as orange or light orange spheres, indicating category A and category B* linkage sites, respectively. The Pol II jaw domain is shown as a molecular surface.

### 3.5.4  A12.2 C-ribbon binds pore like TFIIS

The crosslinking data provide the desired topological insights into the enigmatic A12.2 subunit, which is essential for the strong intrinsic cleavage activity of Pol I (see Chapter 3.4.2). A total of 27 cross-links was identified that contained sites in A12.2 (Figure 25A). Of those, 8 were intra-subunit cross-links and 19 were cross-links between A12.2 and other Pol I subunits, which were analyzed with the coordinates of Rpb9 and TFIIS from the Pol II–TFIIS complex structure (PDB 1Y1V) (Kettenberger et al., 2004). The intra-subunit cross-links can be explained with the Rpb9 structure, and those in the C-ribbon also with the TFIIS structure. Crosslinks between residue K46 of A12.2 and two A190 residues (K1459, K1473) in the Pol I jaw (Figure 25A) and residue K298 in the lobe of A135 are consistent with a position of the A12.2 N-ribbon similar to that of the Rpb9 N-ribbon on Pol II. Four crosslinks to K64, K68 and K73) fall within the linker between the N- and C-ribbon, which likely follows the path taken by the linker of TFIIS. The remaining 12 inter-subunit crosslinks unambiguously position the A12.2 C-ribbon in the Pol I pore. If one assumes that the A12.2 C-ribbon binds in the pore like the TFIIS C-ribbon, 11 out of the 12 cross-links fall within the maximum distance restraint, and one cross-link exceeds the limit by only 2 Å (Figure 25B). In contrast, if one assumes that the A12.2 C-ribbon binds Pol I like the Rpb9 C-ribbon binds Pol II, only one out of 12 cross-links falls within the allowed Cα distance (Figure 25C). These results show that the A12.2 N-ribbon binds the Pol I surface similar to the N-ribbon of Rpb9, whereas the A12.2 C-ribbon binds the Pol I pore, like the TFIIS C-ribbon binds Pol II.
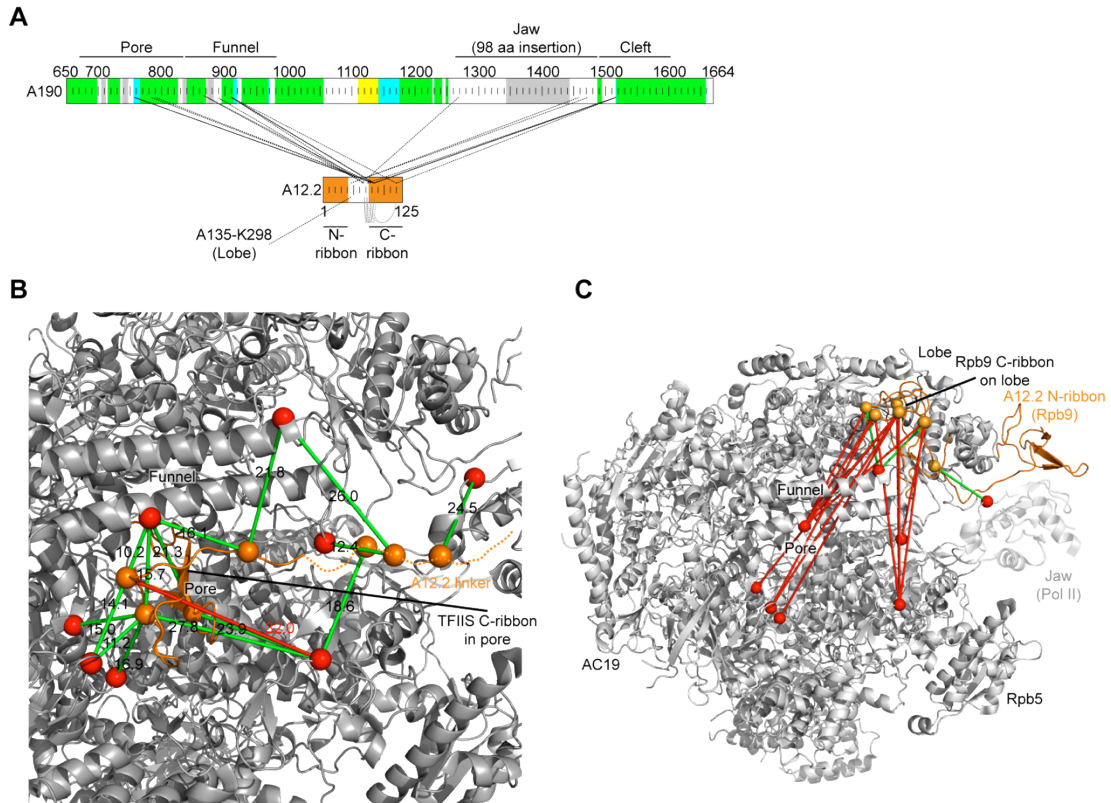
**Figure 25. The A12.2 C-ribbon binds the pore**. (**A**) Schematic representation of the observed crosslinks involving A12.2. The ribbon domains of A12.2 are in orange. The color code is as in Figure 3A. (**B**) The A12.2 C-ribbon crosslinks to the Pol I core. Crosslinks were analyzed with the Pol II-TFIIS structure ((Kettenberger et al., 2004), PDB 1Y1V). The TFIIS C-ribbon is in orange, and its linker is shown as a dashed line. The Cα atoms of crosslinked residues are shown in orange and red for residues in A12.2 and A190, respectively. (**C**) Alternative analysis of the crosslinks between A12.2 and the Pol I core shown in B assuming that A12.2 adopt the location of Rpb9. Cα atoms of crosslinked residues are depicted as orange and red spheres for lysines belonging to A12.2 and the Pol I core, respectively. Crosslinks obeying the Cα distance restraint are shown as green lines; crosslinks exceeding the restraint as red lines.

The A12.2 C-ribbon contains the charged residues R102, D105, and E106 at its tip, which correspond to TFIIS residues R287, D290, and E291, which complement the Pol II active site and induce strong RNA cleavage (Cheung and Cramer, 2011). These results are consistent with recent mutagenesis data that indicated that the corresponding Pol III subunit C11 also enters the pore with its C-ribbon to induce strong RNA cleavage (Ruan et al., 2011). The results thus explain the role of A12.2 and C11 in transcript cleavage (Chedin et al., 1998; Prescott et al., 2004), suggest a close evolutionary relationship

between Pol I and Pol III, and support the hypothesis that Pol I and Pol III differ from Pol II in their mode of RNA cleavage (Ruan et al., 2011).

### 3.5.5 A49/A34.5 binds the lobe like TFIIF

The data also position the dimerization module of the A49/34.5 subcomplex (Geiger et al., 2010) on Pol I. A total of 92 crosslinks involving the Pol I subcomplex A49/34.5 could be obtainede (Figure 26A). Within the dimerization module, 13 out of the 14 crosslinks agreed with the A49/34.5 dimerization module structure (PDB 3NFF) (Geiger et al., 2010) (Figure 26B). The distance for one crosslink pair was slightly above the limit, but this was likely due to a difference in structures, since the structure was obtained from a different species, *Candida glabrata*. Crosslinks between the dimerization module and the Pol I core indicate that the module is positioned on one side of the Pol I cleft on the lobe domain of A135. A crosslink between A49 residue K116, which is seven residues beyond the A49 C-terminal residue in the dimerization module structure (PDB 3NFF), connects to the lobe of A135. Lysines located six and 12 residues beyond the A34.5 C-terminal residue in the structure crosslink to the external domains of A135, which are adjacent to the lobe. All these cross-links can be explained when we assumed that the A49/34.5 dimerization module occupies the location of the TFIIF dimerization module on the lobe of Pol II (Chen et al., 2010), but a 30° rotation of the module structure led to an even better fit (Figure 27A).

These findings are consistent with the localization of the corresponding TFIIF dimerization module on the lobe of Pol II (Chen et al., 2010). Thus not only the structures of the dimerization modules of A49/34.5 and TFIIF are similar, but also their locations on the cores of Pol I and Pol II, respectively. Likewise, the C37/53 dimerization module binds the Pol III lobe, as shown by cryo-EM (Fernandez-Tornero et al., 2010; Vannini et al., 2010) and photo-crosslinking (Wu et al., 2011). The conserved location of the dimerization modules in all three polymerases is consistent with a similar function of the TFIIF-like subcomplex in transcription (Chen et al., 2010; Kassavetis et al., 2010). The observed stimulatory effect of the dimerization module on RNA cleavage (Geiger et al., 2010; Kuhn et al., 2007) can now be explained as an indirect effect resulting from its proximity to subunit A12.2, which is essential for

cleavage (see Chapter 3.4.2). This model is supported by the observation that deletion of C37 from Pol III leads to a loss of C11 (Landrieux et al., 2006).
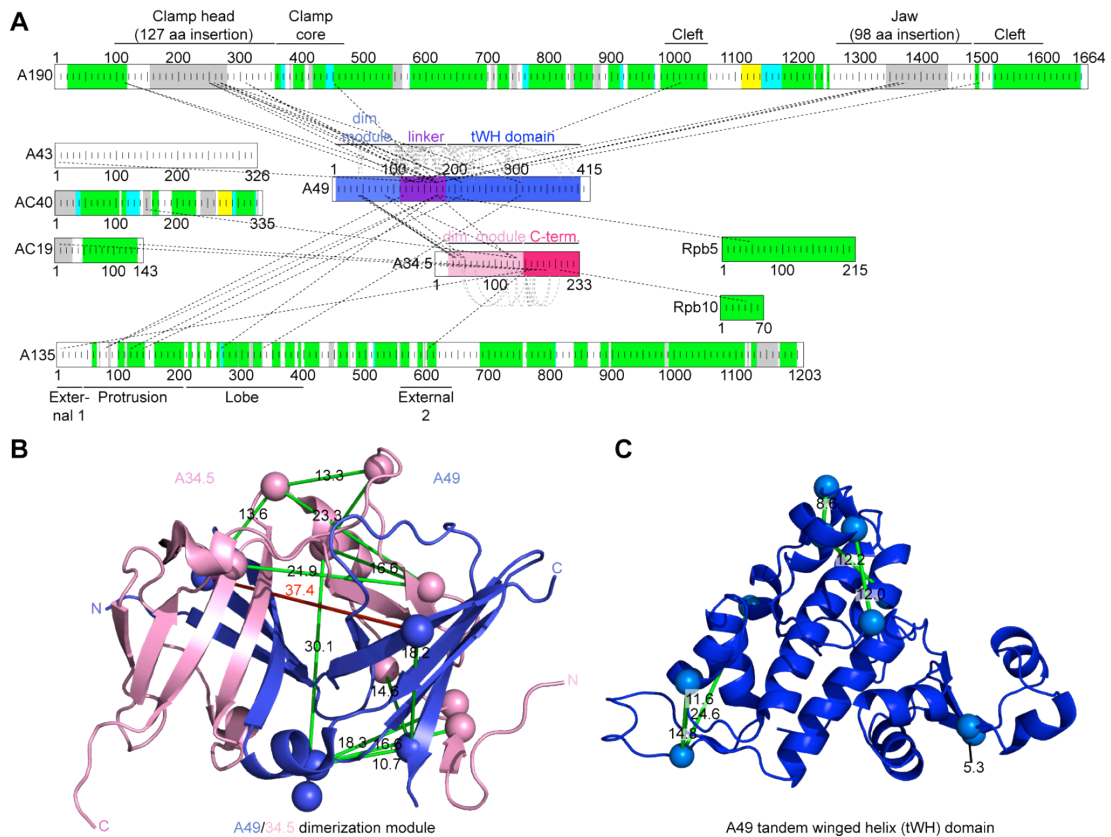


**Figure 26. Crosslinks confirm A49/A34.5 domain structures**. (**A**) Schematic overview of crosslinks involving the A49/A34.5 subcomplex. A49 and A34.5 domains of known structure are labeled. The color code is as in Figures 24A and 25A. (**B**) Crosslinks (green and red lines) are consistent with the known A49/A34.5 dimerization module structure (PDB 3NFG). Cα atoms of crosslinked residues are depicted as slate and pink spheres for A49 and A34.5, respectively. (**C**) Crosslinks are consistent with the structure of the A49 tWH domain (PDB 3NFI). Cα atoms of crosslinked residues are represented as spheres.

### 3.5.6  A49 tWH domain lies above the cleft

We obtained seven crosslinks within the tandem winged helix (tWH) domain of A49 that were consistent with the structure of the isolated domain (Geiger et al., 2010) (Figure 26C). We further observed five crosslinks between the tWH domain and the Pol I core, namely to the lobe and protrusion of A135 on one side of the cleft, and to the Pol I-specific insertion in the clamp head

domain of A190 on the other side (Figure 27B). Extended alignments (see VI.1) based on HHPred predictions (Soding et al., 2005) position the crosslinked residues K259 and K267 of A190 6 and 14 residues beyond strand β4 in a Pol I-specific 52 amino acid insertion in the β4-β5 loop of the clamp head (Cramer et al., 2001). Additionally, the A49 residue K170, which is 15 residues N-terminal of the first ordered tWH residue, cross-links to the dimerization domain of A34.5, indicating proximity between the dimerization module and the tWH domain. The location of the tWH domain over the cleft is consistent with a role of this domain in DNA binding (Geiger et al., 2010), although a repositioning of this domain is required during promoter DNA loading into the cleft. It also corresponds to the location of the Pol III subunit C34 (Vannini et al., 2010), which may be evolutionary related to the A49 tWH domain (Geiger et al., 2010), and is similar to the position of TFIIE on the Pol II clamp (Chen et al., 2007).

The A49 linker connecting the dimerization module with the tWH domain extends along the cleft, since it cross-links to a Pol I-specific insertion in the clamp head domain (residues K111/K309 and K250/K267 in the clamp head and the Pol I-specific insertion, respectively), and with residue K1012 in the bridge helix near the polymerase active center (Figure 27C). This location of the A49 linker is consistent with the location of the corresponding region in the largest TFIIF subunit on Pol II (Chen et al., 2007; Chen et al., 2010). The broad distribution of crosslinks involving the A49 linker over the central cleft and a crosslink to Rpb5 indicate that this region is mobile, consistent with NMR and circular dichroism measurements, which show that the linker is unfolded in solution (S. Geiger and P. Cramer, unpublished data). One crosslink connects the A49 linker to residue K5 in the mobile N-terminal tail of subunit A43, consistent with a genetic interaction between A49 and the N-terminus of A43 (Beckouet et al., 2011). If extended, K5 in the mobile tail could be up to 60 Å away from the first ordered residue in the A43 structure (PDB 2RF4) (Geiger et al., 2008; Kuhn et al., 2007). Only a small number of crosslinks is observed to the C-terminal extension of A34.5, including crosslinks to subunits AC40 and AC19 (Figure 27C).
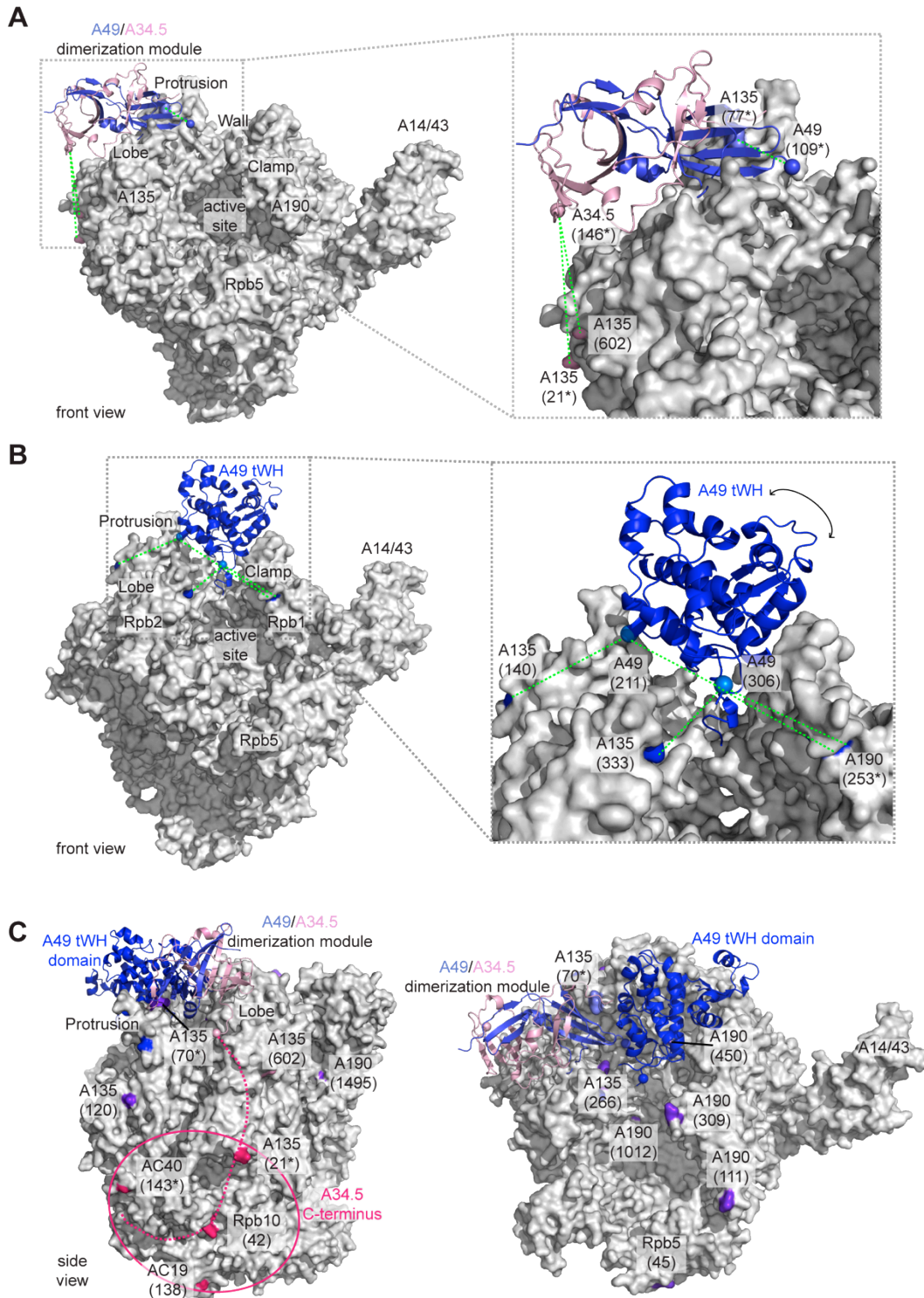
**Figure 27. Location of the A49/34.5 subcomplex on the Pol I core**. (**A**) The A49/34.5 dimerization module resides on the polymerase lobe. The A49/34.5 dimerization module structure has been placed on the Pol I surface manually based on the indicated crosslinks to the Pol I subunit A135. Crosslinks used for domain positioning are depicted as green dashed lines. Crosslink sites on the Pol I surface are highlighted in color and the crosslinked residues are labeled. (**B**) The A49 tWH domain can reside over the cleft. The A49 tWH domain has

64

been placed over the central cleft of Pol I using two crosslink pairs involving subunit A135 and two crosslinks to a Pol I-specific insertion in the clamp head domain of A190. The cross-link sites on Pol I are colored in blue and the Cα atoms of the involved lysines in the A49 tWH domain are shown as spheres and labeled with the crosslinked residue number. Crosslinks used for domain positioning are depicted as green dashed lines. The apparent mobility of the domain is indicated by an arrow. (**C**) Additional crosslinks indicate that the A49/34.5 subcomplex spans a large surface area. Crosslinks to A34.5 are colored in light pink and deep pink for the dimerization module and the A34.5 C-terminus, respectively. Crosslinks to the A49 tWH domain are depicted in blue and crosslinks to the A49 linker are shown in dark violet. Only crosslink positions of the A49 linker and the A34.5 C-terminus are labeled with their respective residue numbers. For crosslink sites that are not part of the structures, the nearest residue is colored and labeled with an asterisk.

Overall, the data indicate that the mobile A49 linker extends into the central cleft of Pol I and that the C-terminal tWH domain can be located above the cleft. This position is similar to that of C34 in the Pol III system, as revealed by cryo-EM (Fernandez-Tornero et al., 2010; Vannini et al., 2010). Bioinformatic analysis (Carter and Drouin, 2010) and homology modeling (Geiger et al., 2010) suggested an evolutionary relationship of C34 and the A49 tWH domain to the β subunit of TFIIE. Since TFIIE crosslinks to the clamp of Pol II (Chen et al., 2007), the A49 tWH domain, C34, and TFIIE can all adopt similar locations on their respective polymerase cores. Consistent with this, recent results showed that the A49 tWH domain binds single-stranded DNA and suggested a role in promoter binding and/or opening (Geiger et al., 2010). Since the Pol III subunit C82 also binds single-stranded DNA (Lefevre et al., 2011) and the archaeal TFIIE homologue TFE stabilizes an open promoter complex (Grunberg et al., 2007; Micorescu et al., 2008), we suggest that the distantly related A49 tWH domain, the Pol III subcomplex C82/34/31, and TFIIE share an old function in binding the melted DNA region above the active center cleft in an open promoter complex during initiation. Loading of the DNA into the cleft may be enabled by the observed mobility of these proteins.

### 3.5.7 Rrn3 binds Pol I near subcomplex AC40/19

To elucidate the molecular basis for binding of the Pol I-specific initiation factor Rrn3, of which the structure was solved by X-ray crystallography (Blattner et al., 2011), the Pol I–Rrn3 complex was subjected to chemical crosslinking and mass spectrometry. A pure Pol I–Rrn3 complex was crosslinked with 1.2 mM DSS, and subsequently analyzed by MS (see Chapter 2.2.4). The analysis revealed two high-confidence crosslinks between Rrn3 and Pol I, connecting Rrn3 residue K558 to Pol I residues K582 and K329 in subunits A190 and AC40, respectively (Figure 28). The crosslinked Pol I residues are located on the ''back'' of the homologous Pol II structure near the Rpb3/11 heterodimer, which corresponds to the AC40/19 heterodimer. To obtain a model for the Pol I–Rrn3 complex, the Rrn3 structure was positioned on the polymerase such that the crosslinks were explained. The crosslinked Rrn3 residue K558 is part of the short mobile loop α20–α21 that follows the ordered residue G554, which was allowed to be up to 30.9 Å from crosslinked Pol I lysines (the theoretical maximum Cα distance of 27.4 Å plus 3.5 Å for mobile residues 555–558). Only one Rrn3 orientation positioned the serine patch towards Pol I, to explain the interaction data (Blattner et al., 2011) without producing protein clashes (Figure 28). In the resulting model of the Pol I–Rrn3 complex, Rrn3 extends from the RNA exit tunnel and dock domain alongside A14/43, the counterpart of the Pol II subcomplex Rpb4/7, to AC40/19. The model explains Rrn3 binding to the OB domain of subunit A43 (Peyroche et al., 2000), an early electron microscopic projection (Peyroche et al., 2000), the observation that Rrn3 can be fused to A43 in vivo (Laferte et al., 2006), and an apparently stabilizing effect of A14 on the Rrn3–Pol I interaction (Imazawa et al., 2005)
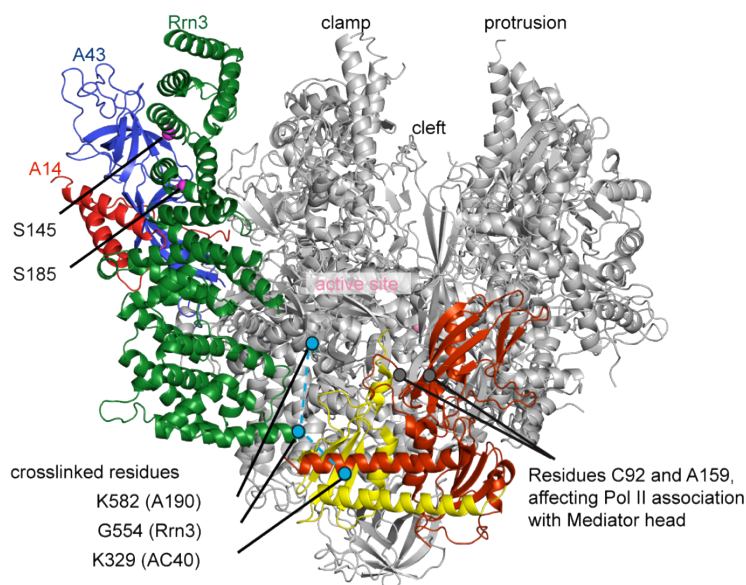
**Figure 28. Model of the Pol I-Rrn3 complex based on crosslinking and MS.** Back view of a 12-subunit Pol I model that is based on the Pol II core structure (silver) (Armache et al., 2005) and the structure of subcomplex A14/43 (red/blue) (Kuhn et al., 2007). The Pol II core heterodimer Rpb3/11 that is homologous to the Pol I heterodimer AC40/19 is highlighted in red/yellow. The positioned Rrn3 structure (Blattner et al., 2011) (green) contains a lysine residue (K558) that is part of a short mobile loop following the ordered residue G554 (cyan dot) and cross-links to two Pol I residues (cyan dots connected with dashed lines). Crosslinked positions in the Pol I core are revealed in the homologous Pol II structure; K582 in A190 corresponds to M437 in Rpb1, and K329 in AC40 corresponds to L259 in Rpb3 (cyan dots). Positions C92 and A159 in Rpb3, influencing the interaction of Pol II with the Mediator head module (Soutourina et al., 2011), are indicated as gray spheres.

### 3.5.8 Comparison of crosslinking and EM-Data

Comparison of the crosslinking data presented here with previously obtained EM data on Pol I strongly suggests that the A49/34.5 subcomplex, like its counterpart TFIIF, maintains a considerable degree of mobility on the polymerase surface. In a cryo-EM reconstruction at 12 Å resolution (see Chapter 3.3.1), densities were observed spanning from the funnel of Pol I to the AC19/40 heterodimer, consistent with some crosslinks described here for the A49 linker and the A34.5 tail, respectively (Figure 27C), but did not show densities on the lobe (Kuhn et al., 2007). An early EM investigation at lower

resolution provided evidence for A49 and A34.5 over the cleft (Bischler et al., 2002), although at that time a clear assignment was not possible. These observations can be reconciled with the mobility of A49/34.5. The two structured domains of this subcomplex are mobile but have preferred locations on the Pol I surface in solution, which are detected by crosslinking and by EM at low resolution, but not by EM at high resolution, where mobile surface structures often get blurred or disappear. Taken together, the present data derived from crosslinking an MS analysis provides the complete structural architecture of Pol I at the level of protein domains, explains the function of surface domains, and further elucidates the evolutionary relationships between the three eukaryotic RNA polymerases.

## 3.6 Crosslinking analysis of Pol III

### 3.6.1 Crosslinking confirms the Pol III homology model

Crosslinking-MS experiments were also carried out for 17-subunit Pol III in order to gain additional insights into its domain architecture. Therefore, complete Pol III from *S.cerevisiae* was purified (see Chapter 2.2.1.7) and subjected to chemical crosslinking and mass spectrometry analysis as described for Pol I (see Chapter 2.2.4.1-2.2.4.4 and 3.5.1) (Figure 29A). Measurements of two samples, crosslinked with 1.2 mM and 3.5 mM DSS, resulted in 654 mass spectra that matched crosslinked peptides.

Within the nine subunits of the Pol III core 158 unique linkage pairs were observed (subunits C160, C128, AC40, AC19, Rpb5, Rpb6, Rpb8, Rpb10, Rpb12; excluding C11, C17/C25, C53/37, and C82/34/31) (Figure 30). The crosslinked residues were analyzed with the atomic Pol II structure (Kettenberger et al., 2004) and the Pol III core homology model (Jasiak et al., 2006).
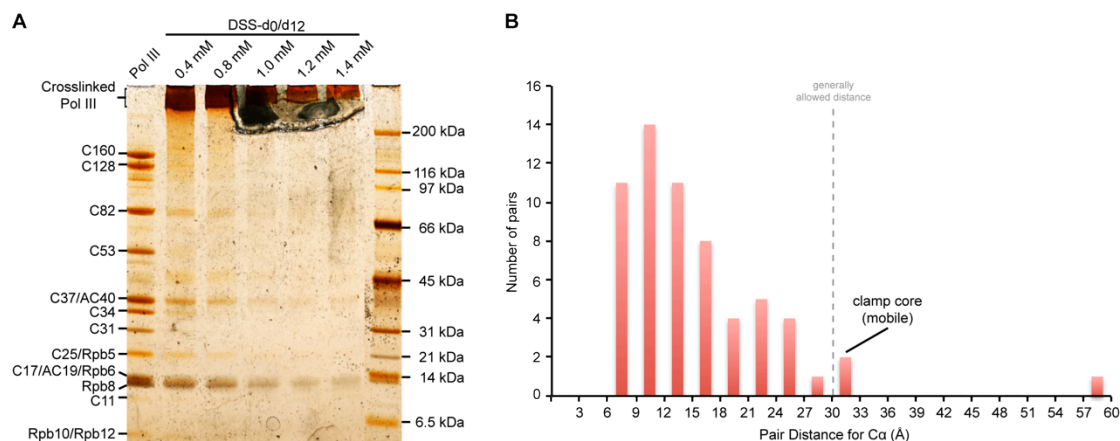
**Figure 29. Crosslinking-MS analysis of Pol III.** (**A**) SDS-PAGE of Pol III crosslinked with different concentrations of SDS. (**B**) Cα distance distribution for experimentally observed A-A linkage pairs within the 9 subunit Pol III core (subunits C160, C128, AC40, AC19, Rpb5, Rpb6, Rpb8, Rpb10, and Rpb12). The generally allowed distance between Cα atoms of two crosslinked lysine residues of 30 Å is indicated by a dashed line. Observed crosslinks are in agreement with the homology model for the Pol III core as judged by analysis with the Pol II X-ray structure (PDB 1Y1V).

The maximum allowed distance between Cα atoms of crosslinked lysins was set to ≤30 Å, similar to crosslink analysis with Pol I (see Chapter 3.5.2). Crosslinking sites that fell in regions that adopt the Pol II fold (Jasiak et al., 2006) were assigned to category A, sites outside this regions to category B (Table 19). Of the 158 crosslink pairs, 66 (41.8%) comprised only category A sites (A-A pairs), 57 (36.1%) contained one category B site (A-B pairs), and 35 (22.1%) were B-B pairs. Of the 66 A-A pairs, 61 could be analyzed, because both crosslinked residues were present in the structure (Kettenberger et al., 2004). Among the 61 A-A crosslink pairs, 58 (95.1%) fell within the acceptable Cα distance of ≤30 Å (Figure 29B). The three remaining pairs exceeded the maximum distance by 0.3 Å, 3 Å and 28.8 Å, respectively. The violation of the distance restraint, in case of the first two crosslink pairs (Cα distance 30.3 Å and 33.0 Å), can be explained by structural flexibility. Both crosslinks comprise residue K371 of the largest subunit C160, which is situated within the clamp domain. This domain has been shown to be mobile in Pol II (Cramer et al., 2001; Gnatt et al., 2001; Kostrewa et al., 2009) as well as in Pol I (Kuhn et al., 2007). Additionally, crosslinks within the C17/C25 subcomplex were obtained as well as between C17/C25 and the Pol III core

(C160 and Rpb6), which were consistent with the previously obtained structure and location of C17/C25 (Jasiak et al., 2006). Only one of 61 crosslinks appears to be false, which results in a false discovery rate of 1.5%. These results demonstrate the validity of the method and confirm the previous Pol III model.

**Table 19. Crosslinking statistics and classification**

| Crosslink pair type | Observed pairs | Both sites present in PDB 1Y1V | Distance ≤30 Å | Comment |
|---|---|---|---|---|
| A-A | 66 | 61 | 58 (95.1%) | 2 outliers involve mobile clamp domain |
| A-B | 57 | 38 | 35 (92.1%) | 1 outliers involves mobile clamp domain |
| B-B | 35 | 14 | 13 (92.9%) | |
| B*-B[1] | 13[1] | 5[1] | 5[1] (100%) | |
| Total | 158 | 113 | 106 (93.8%) | |

[1] Crosslinks are a subgroup of B-B pairs and therefore do not contribute to total number

### 3.6.2  Extension of the Pol III homology model

To extend the Pol III model, category A-B and B-B crosslink pairs were analyzed (Table 19, Figure 30). A-B pairs connect residues in regions of the homology model that share Pol II fold (category A) with residues in sequence regions with no or very weak conservation (category B). Within the nine core subunits 57 A-B crosslinks were observed, of which 19 could not be analyzed as they contain residues within specific insertions or residues that are not present in the Pol II structure. In total, 35 A-B pairs (92.1%) showed Cα distances below 30 Å. These residues were reclassified as B*, and their surrounding region (overall 150 residues within C160, C128 and AC40) was included in the Pol III homology model, as it was likely that the region containing the B* site adopts a Pol II-like fold. Of the resulting 13 B*-B crosslinks, five contained lysine residues present in the Pol II structure and all

of them showed a permissible Cα distance. Based on these findings, the Pol III homology model could be extended to parts of the clamp core, clamp head, pore, and jaw domains of the largest subunit C160, to parts of the external 1, protrusion and fork domains as well as to small parts of the lobe of subunit C128 and to parts of the zinc loop, domain 2, and loop domain of AC40, the counterpart of the Pol II subunit Rpb3 (Figure 30, compare VI.2). The extended homology model relates 85.8% of the nine-subunit Pol III core sequences to their Pol II counterparts, although the large Pol III subunits C160 and C128 show only 33.2% and 35.7% sequence identity to their Pol II counterparts, respectively.
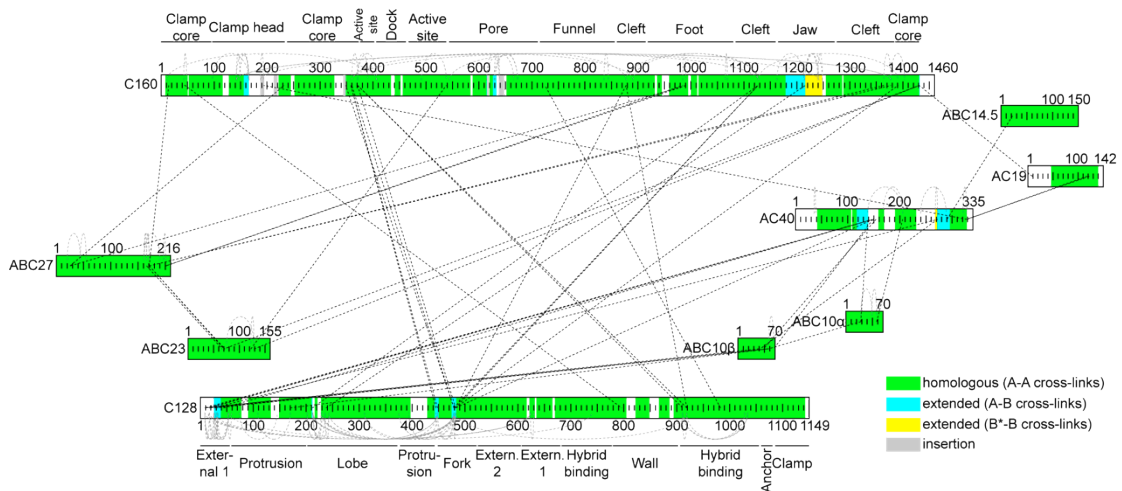


**Figure 30. Crosslink map and extended homology model for the Pol III core.** The primary structure of nine core subunits are shown schematically as boxes. Regions that show a conserved fold are in green, insertions with respect to Pol II are colored in gray, and poorly or non-conserved parts are white. Extensions of the homology model derived from crosslink data are indicated in cyan and yellow for A-B and B*-B crosslinks, respectively (compare VI.2). Black dashed lines and gray dashed arcs indicate inter-subunit and intra-subunit crosslinks, respectively.

### 3.6.3  C53/C37 binds the lobe like TFIIF and connects to C11

Bioinformatic analysis and homology modeling suggested that the Pol III-specific subunits C53/37 form a dimerization module similar to that of A49/34.5 and TFIIF in Pol I and Pol II, respectively (Geiger et al., 2010). Crosslink analysis of 17-subunit Pol III also enabled for positioning of the C53/37 dimerization module on the Pol III core. In total, a number of 129 crosslinks involving the C53/37 subcomplex could be obtained (Figure 31A).
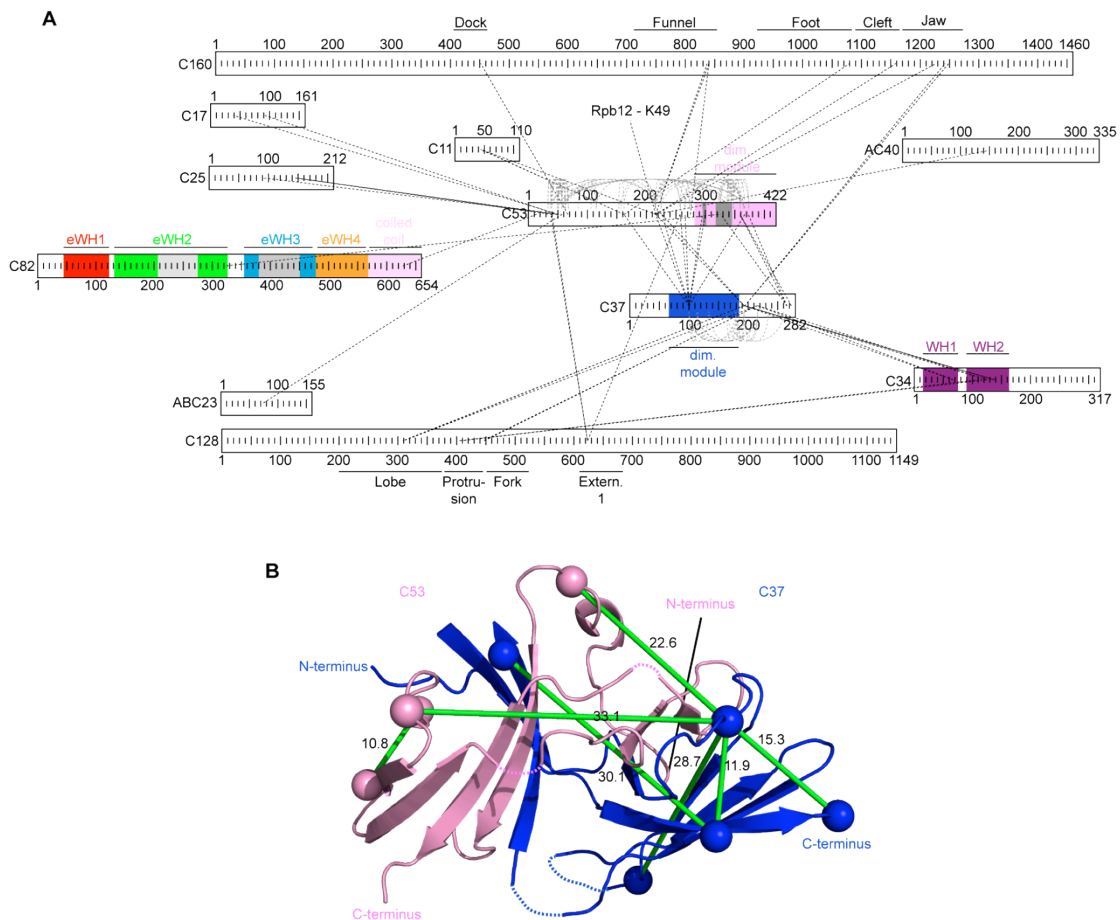


**Figure 31. Crosslinks confirm the homology model for C53/37.** (**A**) Schematic overview of crosslinks involving the C53/37 subcomplex. The proposed dimerization module formed by C53 and C37 is labeled in pink and slate blue, respectively. Insertions with respect to the A49/A34.5 dimerization module are depicted in gray. Known domains of Pol III-specific subunits C82 and C34 are colored and labeled. (**B**) Crosslinks are consistent with the homology model of the C53/C37 dimerization module (Geiger et al., 2010). Cα atoms of crosslinked residues are depicted as pink and slate spheres for C53 and C37, respectively.

Within the dimerization domains 22 intra- and inter-peptide crosslinks were obtained of which seven could be analyzed, as both crosslinked residues were present in the model (Figure 31B). Of those, six crosslinks fell within the maximum Cα distance of ≤30 Å and one exceeded the limit by only 3.1 Å, which is considered acceptable as the present model is derived from homology modeling. Crosslinks between the dimerization module and the Pol III core indicate that the C53/C37 dimerization module is positioned on one side of the Pol III cleft on the lobe domain of C128. Crosslinks of residue K194 of C37, which is ten residues beyond the last residue present in the model, connect to the lobe of C128 and the jaw of C160. A crosslink between C53 residue K274, which is 9 residues N-terminal of the first residue in the homology model, connects to the funnel of C160, which is adjacent to the lobe (Figure 32A). All crosslinks can be explained when one assumes that the C53/37 dimerization module occupies the location of the TFIIF dimerization module, but a slight shift towards the jaw domain lead to an even better fit. The location of the C34 derived from crosslinking and MS experiments is also strongly supported by cryo-EM reconstructions of Pol III (Fernandez-Tornero et al., 2010; Vannini et al., 2010) (Figure 32B) and findings from photo-crosslinking experiments (Wu et al., 2011), that place the dimerization domain at a similar position on the Pol III surface.

Furthermore, a variety of crosslinks to the Pol III core involving the N-terminal part of C53 and the C37 C-terminus could be observed (Figure 32C). The N-terminus of C53 crosslinks to the external domain of C128, to Rpb6, as well as to the dock, funnel, cleft and jaw domain of C160. Also crosslinks to the Rpb4/7 homolog C17/C25 as well as the coiled coil domain of the specfic subunit C82 were observed consistent with results from photo-crosslinking experiments (Wu et al., 2011). The broad distribution of crosslinks to the highly charged C53 N-terminus indicates a certain mobility of this part of the protein similar to the observed mobility for the charged C-terminus of A34.5, the C53 homolog in Pol I (see Chapter 3.5.6). The C-terminal part of C37 crosslinks to the protrusion of C128 as well as to the winged helix domains of C34, the potential homolog of TFIIE and the A49 tWH domain. These crosslinks are consistent with photo-crosslinking data (Wu et al., 2011) and

resemble in part the location of the A49 linker in Pol I (see Chapter 3.5.6), which connects the A49 dimerization domain with the A49 tWH domain.
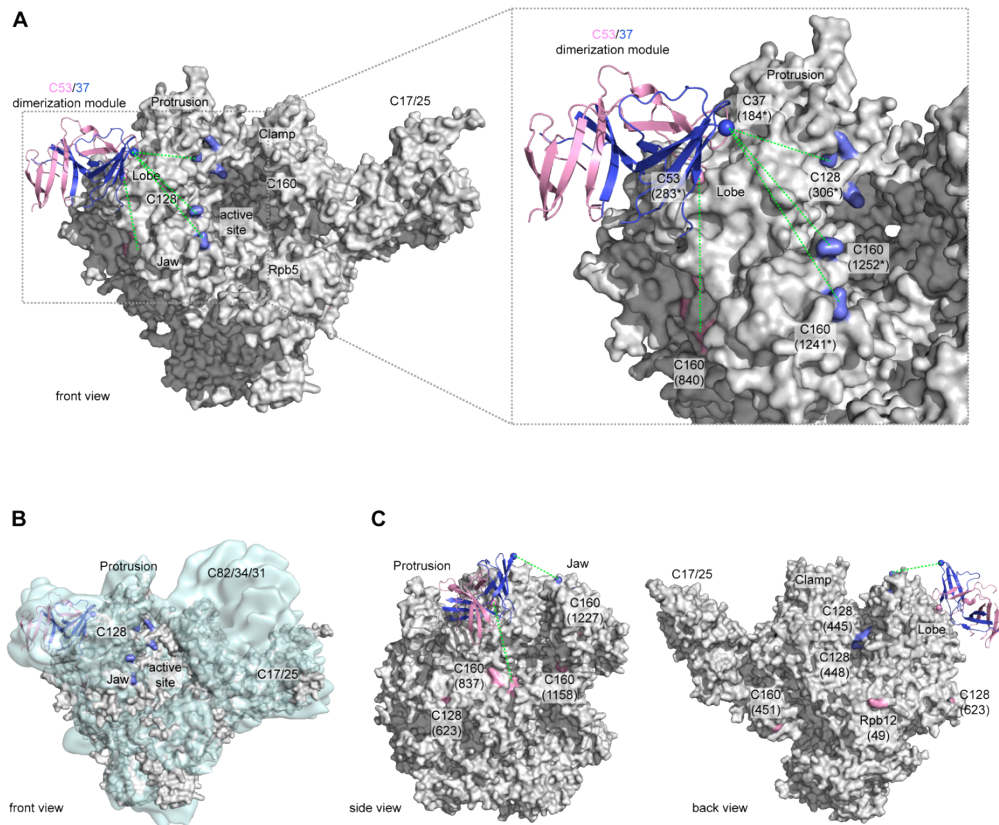


**Figure 32. Location of the C53/37 subcomplex on the Pol III core.** (**A**) The C53/37 dimerization module resides on the polymerase lobe. The model of the dimerization module has been placed on the polymerase surface manually based on the indicated crosslinks to C160 and C128. Crosslinks used for domain positioning are depicted as green dashed lines. Crosslink sites on the Pol III surface are highlighted with the respective color and the crosslinked residues are labeled. (**B**) Overlay of the model derived from crosslinking-MS analysis and a cryo-EM reconstruction at 10 Å (pale cyan) (Fernandez-Tornero et al., 2010). (**C**) Additional crosslinks to the C53 N-terminus and the C37 C-terminus show that the C53/C37 subcomplex spans a large surface area on Pol III. Crosslinks to C53 and C37 are depicted in pink and slate, respectively. Only crosslink positions to the C53 N-terminus and the C37 C-terminus are labeled with their respective residue numbers. For crosslink sites that are absent from the depicted structures, the nearest residue is colored and labeled with an asterisk.

Also crosslinks between the C53 N-terminus and the C11 linker could be observed (Figure 31A), suggesting an interaction of C53/37 with C11. This is supported by the observation that deletion of C37 leads to a loss of C53 as well as C11 (Landrieux et al., 2006). Also two crosslinks between the C11 C-ribbon and the jaw and cleft domain of C160 were observed. However, inspection with help of the structure of Pol II bound to TFIIS (PDB 1Y1V) (Kettenberger et al., 2004) did not reveal a preferred location of the C11 C-ribbon as it was the case for A12.2 on Pol I (see Chapter 3.5.4) since the observed crosslinks did not satisfy the Cα distance restraint neither assuming an Rpb9-like nor a TFIIS-like binding of the C-ribbon. This suggests that the C11 C-ribbon exhibits a higher degree of mobility than its Pol I counterpart A12.2. This mobility is consistent with X-ray studies on an Rpb9-C11 fusion protein, which was active in RNA cleavage assays, but density was only visible for the Rpb9 N-ribbon (Ruan et al., 2011).

### 3.6.4 Location of the C82/34/31 subcomplex

Crosslinking yielded 82 intra- and inter-subunit crosslinks involving the Pol III-specific subunits C82/34/31 (Figure 33A). All crosslinks were analyzed using the Pol III homology model (Jasiak et al., 2006) and homology models for C82 and C34. For C31 the intra-subunit crosslinks could not be analyzed due to a lack of structural information and no evident homologies to known structures, as judged by HHPred analysis (Soding et al., 2005).

Crosslink analysis of C82 yielded 16 intra-subunit crosslinks, of which 6 could be analyzed with a homology model derived from the structure of the human homolog C62 (Lefevre et al., 2011) using Modeller (Eswar et al., 2006). All six crosslinks fell within the Cα distance restraint of ≤30 Å and thus agreed with the calculated model (Figure 33B). The remaining crosslinks contained at least one crosslinked residue located in loops or insertions specific for *S.cerevisiae* and thus could not be analyzed.

In case of C34, two different models were calculated based on NMR structures of winged helix domains from C34 homologs in *M.musculus* and *H.sapiens* (PDB 2DK8, 2DK5, and 2YU3, He et al., unpublished). Modeling with PDB 2DK8 and 2DK5 resulted in a model, in which the winged helix

domains were connected by an extended linker (Figure 33C) whereas modeling with 2DK5 and 2YU3 (domain swapped winged helix of Rpc39) resulted in a more condensed model (Figure 33D). Inspection of both models showed that in the model comprising the extended linker only 5 out of 8 crosslinks satisfied the distance restraint whereas in the condensed model all eight crosslinks fell within the maximum Cα distance of ≤30 Å (Figure 33C,D). Thus, crosslink analysis supports a compact arrangement of the C34 winged helix domains, similar to the arrangement of the WH domains in the A49 tWH domain.
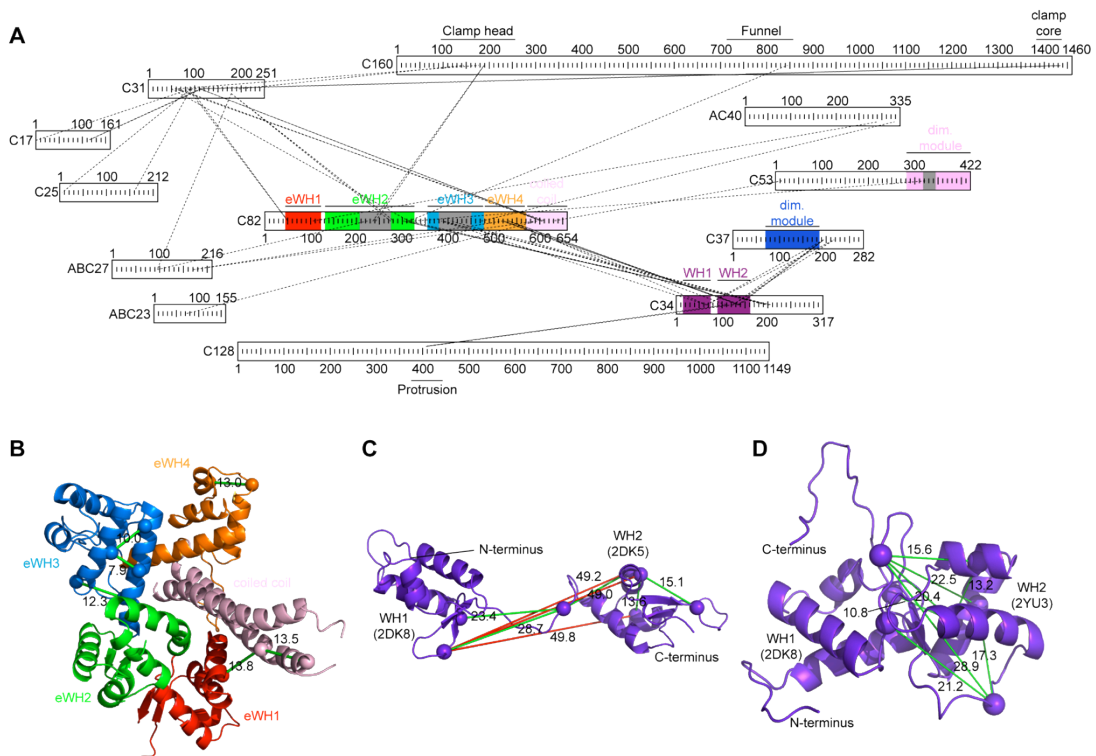


**Figure 33. Crosslinks of the C82/34/31 subcomplex.** (**A**) Schematic overview of crosslinks involving C82/34/31. The extended winged helix domains (eWH) of C82 are labeled and colored, insertions with respect to C62 are depicted as grey boxes. Winged helix domains of C34 are labeled and coloured in purple. C53/37 domains are colored and labeld as in Fig. 30. (**B**) Crosslinks (green lines) are consistent with the model of C82, which is based on the X-ray structure of Rpc62 (Lefevre et al., 2011). The extended winged helices are colored and labeled according to (A). (**C**) Crosslinks (green and red lines) deny the possibility of an extended linker between the WH domains of C34 (modeled with PDB 2DK8 and 2DK5). (**D**) Crosslinks are consistent with a C34 model that shows a dense packing of the WH domains.

Inspection of the inter-subunit crosslinks revealed distinct locations of the three specific subunits, C82/34/31, on the Pol I core (Figure 34A). C31 crosslinks to Rpb5 as well as to the clamp head and linker domains of C160. For subunit C34 crosslinks between WH2 and the protrusion of C128 could be observed. Crosslinks to C82 can be observed in the clamp head and funnel domains of C160 and in subunits Rpb5 and Rpb6. Also two crosslinks of C82 to the Rpb3 homolog AC40 were obtained but it is likely that these are false positive results since AC40 is located opposite to all other C82 crosslink sites on the polymerase core.
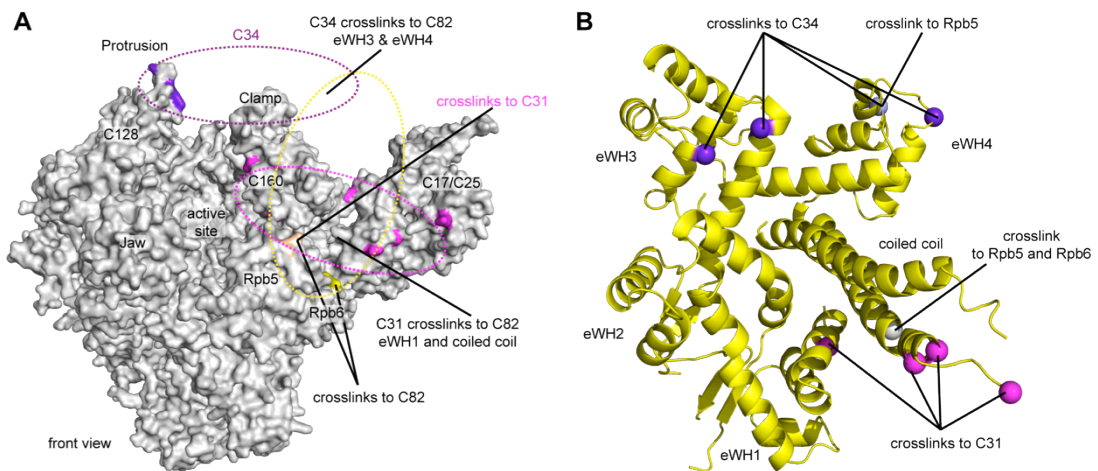


**Figure 34. Crosslinks of the Pol III specific subcomplex C82/34/31.** (**A**) Crosslinked residues on the Pol III surface. Crosslinks to C82, C34, and C31 are colored yellow, purple and magenta, respectively. One crosslink position is shared by C82 and C81 and is labeled and colored wheat. The possible locations of the C82, C34, and C31 are indicated by ellipses. Overlapping ellipses indicate that C34 and C31 crosslink to C82 in different regions. (**B**) Inter-subunit crosslinks of C82. Cα atoms of residues that crosslink to C34 and C31 are depicted as purple and magenta spheres, respectively. Cα atoms of lysines crosslinked to the Pol III core are labeled and depicted as grey spheres.

Despite several available crosslink sites on the Pol III surface, an unambiguous placement of the C82/C34/31 subcomplex was not possible for several reasons. First, the unavailability of structural information for C31 hampered the placement dramatically, since most of the crosslinks to the Pol III core involved C31 residues. Second, out of seven crosslinks to the Pol III

core, three involved C82 residues that are located in a large yeast-specific insertion in extended winged helix 2 (eWH2) (Figure 33A). Third, residues K520 and K594 of C82 both crosslink to Rpb5 K171 (Figure 34B) but are ≈55 Å apart from each other in the C82 model (and the C62 structure PDB 2XUB), which leaves only a few possibilities to place C82 in a position that does not clash with the polymerase core and is still consistent with present cryo-EM reconstructions (Fernandez-Tornero et al., 2010; Vannini et al., 2010). Such a position, in turn, does not agree with the fact that C34 and C31 crosslink to distinct sites of C82 (C34 crosslinks to eWH3 and eWH4; C31 crosslinks to eWH1 and the coiled coil) but show no crosslinks between each other (Figure 34B).

Taken together, the results from crosslinking and mass spectrometry do not allow for an accurate placement of the C82/34/31 subcomplex. However, from the present results it is likely, that C34 adopts a position similar to that of the A49 tWH domain in Pol I (see Chapter 3.5.6) and that of TFIIE (Chen et al., 2007), with its WH1 facing towards the C160 clamp domain and to C82, and the WH2 domain facing towards the C128 side of the central cleft. Furthermore, C82 is located above the C160 clamp domain and connects to C31, which spans from Rpb5 to C17/C25. These positions agree with previous cryo-EM data (Fernandez-Tornero et al., 2010; Vannini et al., 2010). However, the large insertions of C82 with respect to human C62, for which structural data is available (Lefevre et al., 2011), might lead to a rearrangement of the eWH domains which might explain the inconsistencies described above.

# 4 Conclusions and Outlook

The aim of this thesis was to characterize Pol I structurally and functionally. The use of different experimental approaches led to new insights into the architecture of Pol I, revealed the basis for different functional properties of the enzyme, and shed light on evolutionary relationships between the three eukaryotic transcription systems.

Despite exhaustive trials, the structure solution of the complete 14-subunit Pol I by X-ray crystallography remained unsuccessful for several reasons. The reproducibility of suitable crystals for diffraction experiments remained one of the major bottlenecks in structure solution by X-ray crystallography. The obtained Pol I crystals comprised a huge asymmetric unit that contained seven Pol I molecules, which prevented structure solution by molecular replacement even though various strategies, such as the use of custom-made search models derived from cryo-EM reconstructions from Claus Kuhn, were employed. Also, all attempts to solve the crystal structure by isomorphous replacement using heavy atom compounds were unsuccessful because those compounds either diminished diffraction or, in case of the $W_{18}$-cluster, could not be located at distinct positions in the crystal. Since even small changes on the surface of a protein can alter crystal form and crystal quality several strategies can be applied in order to produce crystals with a new crystal form, better diffraction quality, and less molecules within the asymmetric unit. Initial experiments by Christoph Engel showed that complete Pol I from the orthologous *Schizosaccharomyces pombe* (*S.p.*) can be purified applying a purification procedure, which is adapted from the purification protocol used for the preparation of Pol I from *S.cerevisiae*. Also, initial crystals could be obtained from *S.p.* Pol I, but yet need to be reproduced to reach dimensions suitable for X-ray diffraction experiments. Along this line, also other yeasts such as *Pichia pastoris* (*P.p.*) or the thermophilic species *Hansenula polymorpha* (*H.p.*) could be a good source to obtain crystallization grade Pol I. For *P.pastoris* genetic tools to insert affinity tags are well established. Also in case of *H. polymorpha* genome sequences are available (Ramezani-Rad et al., 2003), which enables for genetic manipulations. The use of a thermophilic organism, such as *H.p.*, could result in an enzyme, which is more rigid than

*S.c.* Pol I and therefore might lead to a different crystal form or at least better diffracting crystals. Manipulations on *S.cerevisiae* Pol I could also be of use in order to obtain a different crystal form. Crosslinking and MS analysis revealed an altered jaw domain with a large Pol I-specific insertion in A190 compared to Rpb1 in Pol II, which could interfere with potential crystal contacts. Removing this insertion as well as another specific insertion in the A190 clamp head domain could lead to different, Pol II-like, unit cell parameters and therefore facilitate structure solution by molecular replacement. Recent studies reported the heterologous expression and purification of the core factor (Bedwell et al., 2011), part of the Pol I pre-initiation complex. Co-crystallization experiments with the core factor and Pol I could also lead to a new crystal form, which could enable for the solution of the crystallographic phase problem.

After the overall architecture of Pol I was established by cryo-EM (Kuhn et al., 2007) and the structures of the peripheral subunits, namely A14/43 and A49/34.5, were solved by X-ray crystallography (Geiger et al., 2008; Geiger et al., 2010) several topological and functional questions remained. These questions were addressed by applying a novel technique of protein-protein crosslinking and mass spectrometry. Crosslinking-MS data allowed for an extension of the previously obtained homology model, confirmed the suggested location of the A14/43 subcomplex on the core, and positioned the remaining subunit A12.2 and the two domains of the peripheral subcomplex A49/34.5 on the core. The position of A12.2 in the pore explains why Pol I shows a strong intrinsic cleavage activity compared to the weak intrinsic cleavage activity of Pol II and supports the hypothesis that Pol I and Pol III differ from Pol II in their mode of RNA cleavage (Ruan et al., 2011). The locations of the dimerization module of A49/34.5 and the tWH domain of A49 are consistent with those of TFIIF and TFIIE on Pol II (Chen et al., 2007; Chen et al., 2010) implicating similar functions during transcription. From these data, the view emerges that Pol I is evolutionary related to a partial Pol II-TFIIS-TFIIF-TFIIE complex (Figure 35). The relationship extends to Pol III, which contains an A12.2-related subunit, C11, which is also required for RNA cleavage (Chedin et al., 1998), a heterodimeric subcomplex, C37/53, which is related to TFIIF and A49/34.5 (Kassavetis et al., 2010; Landrieux et al., 2006;

Wu et al., 2011), and an additional subcomplex, C82/34/31, which is related to TFIIE (Lefevre et al., 2011; Wang and Roeder, 1997). Additionally, crosslinking-MS experiments with a Pol I-Rrn3 complex could locate Rrn3 on the Pol I surface, spanning from the stalk subunits A14/43 to the dock domain of A190.
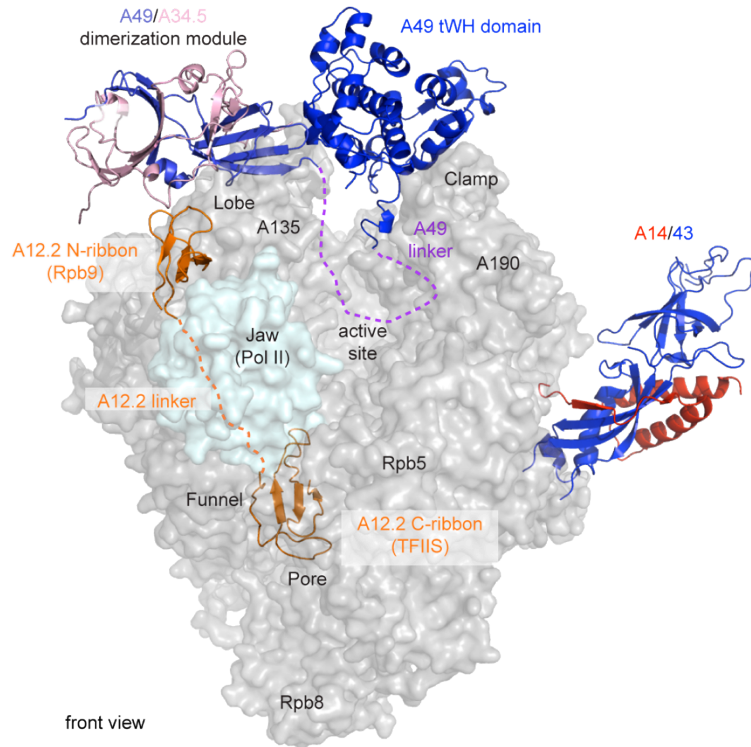


**Figure 35. Domain architecture of the complete 14-subunit Pol I.** The core enzyme is shown as transparent grey surface. The Pol II jaw domain, which is altered in Pol I, is depicted as transparent cyan surface. The A14/43 heterodimer adopts a position similar to Rpb4/7, as suggested before (Kuhn et al., 2007). The A12.2 N- and C-ribbon (orange) were modeled on the jaw/lobe and into the pore, respectively, based on the locations of the N-ribbon of Rpb9 and the C-ribbon of TFIIS. The A12.2 linker is indicated by a dashed line. The A49/A34.5 dimerization module (A49 and A34.5 colored in slate blue and light pink, respectively) is located on the lobe of A135, corresponding to the location of the TFIIF dimerization module (Chen et al., 2010). The A49 tWH domain (blue) resides over the cleft and is apparently mobile.

The results obtained by crosslinking and mass spectrometry prove that this method is very valuable to study multi-protein complexes and thus might be employed in future studies on the Pol I transcription machinery. Crosslinking-

MS, for example, could help to yield topological insights into the core factor, comprising subunits Rrn6, Rrn7 and Rrn11 (Lin et al., 1996), and its interactions with Pol I. Crosslinking-MS is also a valid tool to assist cryo-EM investigations since it enables for better domain assignments using low-resolution EM data. Also EM at high resolutions can profit from validation by crosslinking and mass spectrometry, as in these reconstructions mobile parts of proteins often get blurred or disappear. Taken together, the work described here confirms that crosslinking coupled to mass spectrometry can be a powerful tool to study protein-protein interactions of multi-protein complexes and even larger functional assemblies in cells.

# 5 References

Aprikian, P., Moorefield, B., and Reeder, R.H. (2001). New model for the yeast RNA polymerase I transcription cycle. Molecular and cellular biology *21*, 4847-4855.

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

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

Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. Science *289*, 905-920.

Bazett-Jones, D.P., Leblanc, B., Herfort, M., and Moss, T. (1994). Short-range DNA looping by the Xenopus HMG-box transcription factor, xUBF. Science *264*, 1134-1137.

Beckouet, F., Labarre-Mariotte, S., Albert, B., Imazawa, Y., Werner, M., Gadal, O., Nogi, Y., and Thuriaux, P. (2008). Two RNA polymerase I subunits control the binding and release of Rrn3 during transcription. Mol Cell Biol *28*, 1596-1605.

Beckouet, F., Mariotte-Labarre, S., Peyroche, G., Nogi, Y., and Thuriaux, P. (2011). Rpa43 and its partners in the yeast RNA polymerase I transcription complex. FEBS letters.

Bedwell, G.J., Appling, F.D., Anderson, S.J., and Schneider, D.A. (2011). Efficient transcription by RNA polymerase I using recombinant core factor. Gene.

Bell, S.P., Learned, R.M., Jantzen, H.M., and Tjian, R. (1988). Functional cooperativity between transcription factors UBF1 and SL1 mediates human ribosomal RNA synthesis. Science *241*, 1192-1197.

Ben-Shem, A., Jenner, L., Yusupova, G., and Yusupov, M. (2010). Crystal structure of the eukaryotic ribosome. Science *330*, 1203-1209.

Bergfors, T. (2003). Seeds to crystals. J Struct Biol *142*, 66-76.

Berrow, N.S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., Stuart, D.I., and Owens, R.J. (2007). A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. Nucleic acids research *35*, e45.

Bierhoff, H., Schmitz, K., Maass, F., Ye, J., and Grummt, I. (2010). Noncoding transcripts in sense and antisense orientation regulate the epigenetic state of ribosomal RNA genes. Cold Spring Harb Symp Quant Biol *75*, 357-364.

Bischler, N., Brino, L., Carles, C., Riva, M., Tschochner, H., Mallouh, V., and Schultz, P. (2002). Localization of the yeast RNA polymerase I-specific subunits. EMBO J *21*, 4136-4144.

Blattner, C., Jennebach, S., Herzog, F., Mayer, A., Cheung, A.C., Witte, G., Lorenzen, K., Hopfner, K.P., Heck, A.J., Aebersold, R.*, et al.* (2011). Molecular basis of Rrn3-regulated RNA polymerase I initiation and cell growth. Genes & development.

Bordi, L., Cioci, F., and Camilloni, G. (2001). In vivo binding and hierarchy of assembly of the yeast RNA polymerase I transcription factors. Mol Biol Cell *12*, 753-760.

Boukhgalter, B., Liu, M., Guo, A., Tripp, M., Tran, K., Huynh, C., and Pape, L. (2002). Characterization of a fission yeast subunit of an RNA polymerase I essential transcription initiation factor, SpRrn7h/TAF(I)68, that bridges yeast and mammals: association with SpRrn11h and the core ribosomal RNA gene promoter. Gene *291*, 187-201.

Bradsher, J., Auriol, J., Proietti de Santis, L., Iben, S., Vonesch, J.L., Grummt, I., and Egly, J.M. (2002). CSB is a component of RNA pol I transcription. Molecular cell *10*, 819-829.

Brueckner, F., Ortiz, J., and Cramer, P. (2009). A movie of the RNA polymerase nucleotide addition cycle. Current Opinion in Structural Biology *19*, 294-299.

Brun, I., Sentenac, A., and Werner, M. (1997). Dual role of the C34 subunit of RNA polymerase III in transcription initiation. The EMBO journal *16*, 5730-5741.

Carter, R., and Drouin, G. (2010). The increase in the number of subunits in eukaryotic RNA polymerase III relative to RNA polymerase II is due to the permanent recruitment of general transcription factors. Mol Biol Evol *27*, 1035-1043.

Chedin, S., Riva, M., Schultz, P., Sentenac, A., and Carles, C. (1998). The RNA cleavage activity of RNA polymerase III is mediated by an essential TFIIS-like subunit and is important for transcription termination. Genes & development *12*, 3857-3871.

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

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

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

Claypool, J.A., French, S.L., Johzuka, K., Eliason, K., Vu, L., Dodd, J.A., Beyer, A.L., and 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*, 946-956.

Comai, L., Tanese, N., and Tjian, R. (1992). The TATA-binding protein and associated factors are integral components of the RNA polymerase I transcription factor, SL1. Cell *68*, 965-976.

Cramer, P. (2004). Structure and function of RNA polymerase II. Adv Protein Chem *67*, 1-42.

Cramer, P., Armache, K.J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., Damsma, G.E., Dengl, S., Geiger, S.R., Jasiak, A.J.*, et al.* (2008). Structure of eukaryotic RNA polymerases. Annu Rev Biophys *37*, 337-352.

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

Darzacq, X., Shav-Tal, Y., de Turris, V., Brody, Y., Shenoy, S.M., Phair, R.D., and Singer, R.H. (2007). In vivo dynamics of RNA polymerase II transcription. Nature structural & molecular biology *14*, 796-806.

De Carlo, S., Carles, C., Riva, M., and Schultz, P. (2003). Cryo-negative staining reveals conformational flexibility within yeast RNA polymerase I. J Mol Biol *329*, 891-902.

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.*, et al.* (2011). Targeting RNA polymerase I with an oral small molecule CX-5461 inhibits ribosomal RNA synthesis and solid tumor growth. Cancer Res *71*, 1418-1430.

Drygin, D., Rice, W.G., and Grummt, I. (2010). The RNA polymerase I transcription machinery: an emerging target for the treatment of cancer. Annu Rev Pharmacol Toxicol *50*, 131-156.

Dundr, M., Hoffmann-Rohrer, U., Hu, Q., Grummt, I., Rothblum, L.I., Phair, R.D., and Misteli, T. (2002). A kinetic framework for a mammalian RNA polymerase in vivo. Science *298*, 1623-1626.

Dye, M.J., Gromak, N., Haussecker, D., West, S., and Proudfoot, N.J. (2006). Turnover and function of noncoding RNA polymerase II transcripts. Cold Spring Harb Symp Quant Biol *71*, 275-284.

Earley, K., Lawrence, R.J., Pontes, O., Reuther, R., Enciso, A.J., Silva, M., Neves, N., Gross, M., Viegas, W., and Pikaard, C.S. (2006). Erasure of histone acetylation by Arabidopsis HDA6 mediates large-scale gene silencing in nucleolar dominance. Genes & development *20*, 1283-1293.

Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M.Y., Pieper, U., and Sali, A. (2006). Comparative protein structure modeling using Modeller. Curr Protoc Bioinformatics *Chapter 5*, Unit 5 6.

Fath, S., Milkereit, P., Peyroche, G., Riva, M., Carles, C., and Tschochner, H. (2001). Differential roles of phosphorylation in the formation of transcriptional active RNA polymerase I. Proc Natl Acad Sci U S A *98*, 14334-14339.

Fatica, A., and Tollervey, D. (2002). Making ribosomes. Current opinion in cell biology *14*, 313-318.

Felle, M., Exler, J.H., Merkl, R., Dachauer, K., Brehm, A., Grummt, I., and Langst, G. (2010). DNA sequence encoded repression of rRNA gene transcription in chromatin. Nucleic acids research *38*, 5304-5314.

Fernandez-Tornero, C., Bottcher, B., Rashid, U.J., Steuerwald, U., Florchinger, B., Devos, D.P., Lindner, D., and Muller, C.W. (2010). Conformational flexibility of RNA polymerase III during transcriptional elongation. EMBO J.

Fernandez-Tornero, C., Bottcher, B., Riva, M., Carles, C., Steuerwald, U., Ruigrok, R.W., Sentenac, A., Muller, C.W., and Schoehn, G. (2007). Insights into transcription initiation and termination from the electron microscopy structure of yeast RNA polymerase III. Mol Cell *25*, 813-823.

Fontoura, B.M., Atienza, C.A., Sorokina, E.A., Morimoto, T., and Carroll, R.B. (1997). Cytoplasmic p53 polypeptide is associated with ribosomes. Molecular and cellular biology *17*, 3146-3154.

Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., and Leith, A. (1996). SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. J Struct Biol *116*, 190-199.

French, S.L., Osheim, Y.N., Cioci, F., Nomura, M., and Beyer, A.L. (2003). In exponentially growing Saccharomyces cerevisiae cells, rRNA synthesis is determined by the summed RNA polymerase I loading rate rather than by the number of active genes. Molecular and cellular biology *23*, 1558-1568.

Fromont-Racine, M., Senger, B., Saveanu, C., and Fasiolo, F. (2003). Ribosome assembly in eukaryotes. Gene *313*, 17-42.

Gadal, O., Mariotte-Labarre, S., Chedin, S., Quemeneur, E., Carles, C., Sentenac, A., and Thuriaux, P. (1997). A34.5, a nonessential component of yeast RNA polymerase I, cooperates with subunit A14 and DNA topoisomerase I to produce a functional rRNA synthesis machine. Mol Cell Biol *17*, 1787-1795.

Geiduschek, E.P., and Kassavetis, G.A. (2001). The RNA polymerase III transcription apparatus. Journal of molecular biology *310*, 1-26.

Geiger, S.R., Kuhn, C.D., Leidig, C., Renkawitz, J., and 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*, 413-418.

Geiger, S.R., Lorenzen, K., Schreieck, A., Hanecker, P., Kostrewa, D., Heck, A.J., and Cramer, P. (2010). RNA polymerase I contains a TFIIF-related DNA-binding subcomplex. Mol Cell *39*, 583-594.

Gelperin, D., Horton, L., Beckman, J., Hensold, J., and Lemmon, S.K. (2001). Bms1p, a novel GTP-binding protein, and the related Tsr1p are required for distinct steps of 40S ribosome biogenesis in yeast. Rna *7*, 1268-1283.

Gerber, J., Reiter, A., Steinbauer, R., Jakob, S., Kuhn, C.D., Cramer, P., Griesenbeck, J., Milkereit, P., and Tschochner, H. (2008). Site specific phosphorylation of yeast RNA polymerase I. Nucleic Acids Res *36*, 793-802.

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

Gorski, J.J., Pathak, S., Panov, K., Kasciukovic, T., Panova, T., Russell, J., and Zomerdijk, J.C. (2007). A novel TBP-associated factor of SL1 functions in RNA polymerase I transcription. The EMBO journal *26*, 1560-1568.

Grandi, P., Rybin, V., Bassler, J., Petfalski, E., Strauss, D., Marzioch, M., Schafer, T., Kuster, B., Tschochner, H., Tollervey, D.*, et al.* (2002). 90S pre-ribosomes include the 35S pre-rRNA, the U3 snoRNP, and 40S subunit processing factors but predominantly lack 60S synthesis factors. Molecular cell *10*, 105-115.

Grummt, I. (2003). Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus. Genes & development *17*, 1691-1702.

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

Grunberg, S., Bartlett, M.S., Naji, S., and Thomm, M. (2007). Transcription factor E is a part of transcription elongation complexes. The Journal of biological chemistry *282*, 35482-35490.

Hahn, S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. Nature structural & molecular biology *11*, 394-403.

Hamada, M., Huang, Y., Lowe, T.M., and Maraia, R.J. (2001). Widespread use of TATA elements in the core promoters for RNA polymerases III, II, and I in fission yeast. Molecular and cellular biology *21*, 6870-6881.

Hanada, K., Song, C.Z., Yamamoto, K., Yano, K., Maeda, Y., Yamaguchi, K., and Muramatsu, M. (1996). RNA polymerase I associated factor 53 binds to the nucleolar transcription factor UBF and functions in specific rDNA transcription. The EMBO journal *15*, 2217-2226.

Harp, J.M., Hanson, B.L., Timm, D.E., and Bunick, G.J. (1999). Macromolecular crystal annealing: evaluation of techniques and variables. Acta crystallographica Section D, Biological crystallography *55*, 1329-1334.

Henras, A.K., Soudet, J., Gerus, M., Lebaron, S., Caizergues-Ferrer, M., Mougin, A., and Henry, Y. (2008). The post-transcriptional steps of eukaryotic ribosome biogenesis. Cell Mol Life Sci *65*, 2334-2359.

Huet, J., Buhler, J.M., Sentenac, A., and Fromageot, P. (1975). Dissociation of two polypeptide chains from yeast RNA polymerase A. Proceedings of the National Academy of Sciences of the United States of America *72*, 3034-3038.

Huet, J., Buhler, J.M., Sentenac, A., and Fromageot, P. (1977). Characterization of ribonuclease H activity associated yeast RNA polymerase A. The Journal of biological chemistry *252*, 8848-8855.

Huet, J., Wyers, F., Buhler, J.M., Sentenac, A., and Fromageot, P. (1976). Association of RNase H activity with yeast RNA polymerase A. Nature *261*, 431-433.

Hurt, E., Hannus, S., Schmelzl, B., Lau, D., Tollervey, D., and Simos, G. (1999). A novel in vivo assay reveals inhibition of ribosomal nuclear export in ran-cycle and nucleoporin mutants. The Journal of cell biology *144*, 389-401.

Imazawa, Y., Hisatake, K., Mitsuzawa, H., Matsumoto, M., Tsukui, T., Nakagawa, K., Nakadai, T., Shimada, M., Ishihama, A., and Nogi, Y. (2005). The fission yeast protein Ker1p is an ortholog of RNA polymerase I subunit A14 in Saccharomyces cerevisiae and is required for stable association of Rrn3p and RPA21 in RNA polymerase I. J Biol Chem *280*, 11467-11474.

Jansa, P., and Grummt, I. (1999). Mechanism of transcription termination: PTRF interacts with the largest subunit of RNA polymerase I and dissociates paused transcription complexes from yeast and mouse. Molecular & general genetics : MGG *262*, 508-514.

Jasiak, A.J., Armache, K.J., Martens, B., Jansen, R.P., and 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.

Jenni, S., and Ban, N. (2009). Imperfect pseudo-merohedral twinning in crystals of fungal fatty acid synthase. Acta crystallographica Section D, Biological crystallography *65*, 101-111.

Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. Journal of molecular biology *292*, 195-202.

Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. Journal of Applied Crystallography %@ 0021-8898 *26*, 795-800.

Kassavetis, G.A., Prakash, P., and Shim, E. (2010). The C53/C37 subcomplex of RNA polymerase III lies near the active site and participates in promoter opening. J Biol Chem *285*, 2695-2706.

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

Keller, A., Eng, J., Zhang, N., Li, X.J., and Aebersold, R. (2005). A uniform proteomics MS/MS analysis platform utilizing open XML file formats. Mol Syst Biol *1*, 2005 0017.

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

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

Kireeva, M.L., Komissarova, N., Waugh, D.S., and Kashlev, M. (2000). The 8-nucleotide-long RNA:DNA hybrid is a primary stability determinant of the RNA polymerase II elongation complex. The Journal of biological chemistry *275*, 6530-6536.

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

Kressler, D., Hurt, E., and Bassler, J. (2010). Driving ribosome assembly. Biochim Biophys Acta *1803*, 673-683.

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

Laferte, A., Favry, E., Sentenac, A., Riva, M., Carles, C., and Chedin, S. (2006). The transcriptional activity of RNA polymerase I is a key determinant for the level of all ribosome components. Genes Dev *20*, 2030-2040.

Lafontaine, D., Vandenhaute, J., and Tollervey, D. (1995). The 18S rRNA dimethylase Dim1p is required for pre-ribosomal RNA processing in yeast. Genes & development *9*, 2470-2481.

Lafontaine, D.L., Preiss, T., and Tollervey, D. (1998). Yeast 18S rRNA dimethylase Dim1p: a quality control mechanism in ribosome synthesis? Molecular and cellular biology *18*, 2360-2370.

Lalo, D., Steffan, J.S., Dodd, J.A., and Nomura, M. (1996). RRN11 encodes the third subunit of the complex containing Rrn6p and Rrn7p that is essential for the initiation of rDNA transcription by yeast RNA polymerase I. J Biol Chem *271*, 21062-21067.

Landrieux, E., Alic, N., Ducrot, C., Acker, J., Riva, M., and Carles, C. (2006). A subcomplex of RNA polymerase III subunits involved in transcription termination and reinitiation. EMBO J *25*, 118-128.

Langst, G., Blank, T.A., Becker, P.B., and Grummt, I. (1997). RNA polymerase I transcription on nucleosomal templates: the transcription termination factor TTF-I induces chromatin remodeling and relieves transcriptional repression. The EMBO journal *16*, 760-768.

Lannutti, B.J., Persinger, J., and Bartholomew, B. (1996). Probing the protein-DNA contacts of a yeast RNA polymerase III transcription complex in a crude extract: solid phase synthesis of DNA photoaffinity probes containing a novel photoreactive deoxycytidine analog. Biochemistry *35*, 9821-9831.

Lawrence, R.J., Earley, K., Pontes, O., Silva, M., Chen, Z.J., Neves, N., Viegas, W., and Pikaard, C.S. (2004). A concerted DNA methylation/histone methylation switch regulates rRNA gene dosage control and nucleolar dominance. Molecular cell *13*, 599-609.

Lefevre, S., Dumay-Odelot, H., El-Ayoubi, L., Budd, A., Legrand, P., Pinaud, N., Teichmann, M., and Fribourg, S. (2011). Structure-function analysis of hRPC62 provides insights into RNA polymerase III transcription initiation. Nature structural & molecular biology *18*, 352-358.

Liljelund, P., Mariotte, S., Buhler, J.M., and Sentenac, A. (1992). Characterization and mutagenesis of the gene encoding the A49 subunit of RNA polymerase A in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A *89*, 9302-9305.

Lin, C.W., Moorefield, B., Payne, J., Aprikian, P., Mitomo, K., and 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*, 6436-6443.

Lorenzen, K., Vannini, A., Cramer, P., and Heck, A.J. (2007). Structural biology of RNA polymerase III: mass spectrometry elucidates subcomplex architecture. Structure *15*, 1237-1245.

Matthews, B.W. (1968). Solvent content of protein crystals. Journal of molecular biology *33*, 491-497.

Mayer, C., Bierhoff, H., and Grummt, I. (2005). The nucleolus as a stress sensor: JNK2 inactivates the transcription factor TIF-IA and down-regulates rRNA synthesis. Genes & development *19*, 933-941.

Mayer, C., Schmitz, K.M., Li, J., Grummt, I., and Santoro, R. (2006). Intergenic transcripts regulate the epigenetic state of rRNA genes. Molecular cell *22*, 351-361.

Mayer, C., Zhao, J., Yuan, X., and Grummt, I. (2004). mTOR-dependent activation of the transcription factor TIF-IA links rRNA synthesis to nutrient availability. Genes & development *18*, 423-434.

Mccoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J Appl Crystallogr *40*, 658-674.

McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C., and Read, R.J. (2005). Likelihood-enhanced fast translation functions. Acta crystallographica Section D, Biological crystallography *61*, 458-464.

McStay, B., and Grummt, I. (2008). The epigenetics of rRNA genes: from molecular to chromosome biology. Annu Rev Cell Dev Biol *24*, 131-157.

Meka, H., Daoust, G., Arnvig, K.B., Werner, F., Brick, P., and Onesti, S. (2003). Structural and functional homology between the RNAP(I) subunits A14/A43 and the archaeal RNAP subunits E/F. Nucleic Acids Res *31*, 4391-4400.

Micorescu, M., Grunberg, S., Franke, A., Cramer, P., Thomm, M., and Bartlett, M. (2008). Archaeal transcription: function of an alternative transcription factor B from Pyrococcus furiosus. J Bacteriol *190*, 157-167.

Milkereit, P., and 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*, 3692-3703.

Miller, O.L., Jr., and Beatty, B.R. (1969). Visualization of nucleolar genes. Science *164*, 955-957.

Moss, T. (2004). At the crossroads of growth control; making ribosomal RNA. Curr Opin Genet Dev *14*, 210-217.

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

Mougey, E.B., O'Reilly, M., Osheim, Y., Miller, O.L., Jr., Beyer, A., and Sollner-Webb, B. (1993). The terminal balls characteristic of eukaryotic rRNA transcription units in chromatin spreads are rRNA processing complexes. Genes & development *7*, 1609-1619.

Moy, T.I., and Silver, P.A. (1999). Nuclear export of the small ribosomal subunit requires the ran-GTPase cycle and certain nucleoporins. Genes & development *13*, 2118-2133.

Moy, T.I., and Silver, P.A. (2002). Requirements for the nuclear export of the small ribosomal subunit. Journal of cell science *115*, 2985-2995.

Navaza, J. (1994). Amore - an Automated Package for Molecular Replacement. Acta Crystallographica Section A *50*, 157-163.

Nogi, Y., Yano, R., and Nomura, M. (1991). Synthesis of large rRNAs by RNA polymerase II in mutants of Saccharomyces cerevisiae defective in RNA polymerase I. Proceedings of the National Academy of Sciences of the United States of America *88*, 3962-3966.

Nomura, M. (2001). Ribosomal RNA genes, RNA polymerases, nucleolar structures, and synthesis of rRNA in the yeast Saccharomyces cerevisiae. Cold Spring Harb Symp Quant Biol *66*, 555-565.

Orlova, M., Newlands, J., Das, A., Goldfarb, A., and Borukhov, S. (1995). Intrinsic transcript cleavage activity of RNA polymerase. Proceedings of the National Academy of Sciences of the United States of America *92*, 4596-4600.

Paule, M.R., and White, R.J. (2000). Survey and summary: transcription by RNA polymerases I and III. Nucleic Acids Res *28*, 1283-1298.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem *25*, 1605-1612.

Peyroche, G., Levillain, E., Siaut, M., Callebaut, I., Schultz, P., Sentenac, A., Riva, M., and Carles, C. (2002). The A14-A43 heterodimer subunit in yeast RNA pol I and their relationship to Rpb4-Rpb7 pol II subunits. Proceedings of the National Academy of Sciences of the United States of America *99*, 14670-14675.

Peyroche, G., Milkereit, P., Bischler, N., Tschochner, H., Schultz, P., Sentenac, A., Carles, C., and Riva, M. (2000). The recruitment of RNA polymerase I on rDNA is mediated by the interaction of the A43 subunit with Rrn3. EMBO J *19*, 5473-5482.

Poole, A.M., Jeffares, D.C., and Penny, D. (1998). The path from the RNA world. J Mol Evol *46*, 1-17.

Prescott, E.M., Osheim, Y.N., Jones, H.S., Alen, C.M., Roan, J.G., Reeder, R.H., Beyer, A.L., and Proudfoot, N.J. (2004). Transcriptional termination by RNA polymerase I requires the small subunit Rpa12p. Proc Natl Acad Sci U S A *101*, 6068-6073.

Preuss, S., and Pikaard, C.S. (2007). rRNA gene silencing and nucleolar dominance: insights into a chromosome-scale epigenetic on/off switch. Biochimica et biophysica acta *1769*, 383-392.

Proud, C.G. (2002). Regulation of mammalian translation factors by nutrients. European journal of biochemistry / FEBS *269*, 5338-5349.

Ramezani-Rad, M., Hollenberg, C.P., Lauber, J., Wedler, H., Griess, E., Wagner, C., Albermann, K., Hani, J., Piontek, M., Dahlems, U.*, et al.* (2003). The Hansenula polymorpha (strain CBS4732) genome sequencing and analysis. FEMS yeast research *4*, 207-215.

Read, R.J. (2001). Pushing the boundaries of molecular replacement with maximum likelihood. Acta crystallographica Section D, Biological crystallography *57*, 1373-1382.

Ruan, W., Lehmann, E., Thomm, M., Kostrewa, D., and Cramer, P. (2011). Evolution of two modes of intrinsic RNA polymerase transcript cleavage. The Journal of biological chemistry.

Ruggero, D., and Pandolfi, P.P. (2003). Does the ribosome translate cancer? Nat Rev Cancer *3*, 179-192.

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

Santoro, R., Li, J., and Grummt, I. (2002). The nucleolar remodeling complex NoRC mediates heterochromatin formation and silencing of ribosomal gene transcription. Nature genetics *32*, 393-396.

Schafer, T., Strauss, D., Petfalski, E., Tollervey, D., and Hurt, E. (2003). The path from nucleolar 90S to cytoplasmic 40S pre-ribosomes. The EMBO journal *22*, 1370-1380.

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

Schuwirth, B.S., Borovinskaya, M.A., Hau, C.W., Zhang, W., Vila-Sanjurjo, A., Holton, J.M., and Cate, J.H. (2005). Structures of the bacterial ribosome at 3.5 A resolution. Science *310*, 827-834.

Senger, B., Lafontaine, D.L., Graindorge, J.S., Gadal, O., Camasses, A., Sanni, A., Garnier, J.M., Breitenbach, M., Hurt, E., and Fasiolo, F. (2001). The nucle(ol)ar Tif6p and Efl1p are required for a late cytoplasmic step of ribosome synthesis. Molecular cell *8*, 1363-1373.

Soding, J., Biegert, A., and Lupas, A.N. (2005). The HHpred interactive server for protein homology detection and structure prediction. Nucleic acids research *33*, W244-248.

Soutourina, J., Wydau, S., Ambroise, Y., Boschiero, C., and Werner, M. (2011). Direct interaction of RNA polymerase II and mediator required for transcription in vivo. Science *331*, 1451-1454.

Stefanovsky, V., Langlois, F., Gagnon-Kugler, T., Rothblum, L.I., and Moss, T. (2006a). Growth factor signaling regulates elongation of RNA polymerase I transcription in mammals via UBF phosphorylation and r-chromatin remodeling. Mol Cell *21*, 629-639.

Stefanovsky, V.Y., Langlois, F., Bazett-Jones, D., Pelletier, G., and Moss, T. (2006b). ERK modulates DNA bending and enhancesome structure by phosphorylating HMG1-boxes 1 and 2 of the RNA polymerase I transcription factor UBF. Biochemistry *45*, 3626-3634.

Stefanovsky, V.Y., and Moss, T. (2008). The splice variants of UBF differentially regulate RNA polymerase I transcription elongation in response to ERK phosphorylation. Nucleic Acids Res *36*, 5093-5101.

Stefanovsky, V.Y., Pelletier, G., Hannan, R., Gagnon-Kugler, T., Rothblum, L.I., and Moss, T. (2001). An immediate response of ribosomal transcription to growth factor stimulation in mammals is mediated by ERK phosphorylation of UBF. Molecular cell *8*, 1063-1073.

Steffan, J.S., Keys, D.A., Dodd, J.A., and 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*, 2551-2563.

Storoni, L.C., McCoy, A.J., and Read, R.J. (2004). Likelihood-enhanced fast rotation functions. Acta crystallographica Section D, Biological crystallography *60*, 432-438.

Tong, L. (1993). Replace, a Suite of Computer-Programs for Molecular-Replacement Calculations. J Appl Crystallogr *26*, 748-751.

Tong, L.A., and Rossmann, M.G. (1990). The locked rotation function. Acta Crystallogr A *46 ( Pt 10)*, 783-792.

Trotta, C.R., Lund, E., Kahan, L., Johnson, A.W., and Dahlberg, J.E. (2003). Coordinated nuclear export of 60S ribosomal subunits and NMD3 in vertebrates. The EMBO journal *22*, 2841-2851.

Trumtel, S., Leger-Silvestre, I., Gleizes, P.E., Teulieres, F., and Gas, N. (2000). Assembly and functional organization of the nucleolus: ultrastructural analysis of Saccharomyces cerevisiae mutants. Mol Biol Cell *11*, 2175-2189.

Tschochner, H. (1996). A novel RNA polymerase I-dependent RNase activity that shortens nascent transcripts from the 3' end. Proc Natl Acad Sci U S A *93*, 12914-12919.

Tschochner, H., and Hurt, E. (2003). Pre-ribosomes on the road from the nucleolus to the cytoplasm. Trends Cell Biol *13*, 255-263.

Tuan, J.C., Zhai, W., and Comai, L. (1999). Recruitment of TATA-binding protein-TAFI complex SL1 to the human ribosomal DNA promoter is mediated by the carboxy-terminal activation domain of upstream binding factor (UBF) and is regulated by UBF phosphorylation. Molecular and cellular biology *19*, 2872-2879.

Vagin, A., and Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. J Appl Crystallogr *30*, 1022-1025.

Vaguine, A.A., Richelle, J., and Wodak, S.J. (1999). SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. Acta crystallographica Section D, Biological crystallography *55*, 191-205.

Van Mullem, V., Landrieux, E., Vandenhaute, J., and Thuriaux, P. (2002). Rpa12p, a conserved RNA polymerase I subunit with two functional domains. Mol Microbiol *43*, 1105-1113.

Vannini, A., Ringel, R., Kusser, A.G., Berninghausen, O., Kassavetis, G.A., and Cramer, P. (2010). Molecular Basis of RNA Polymerase III Transcription Repression by Maf1. Cell *143*, 59-70.

Venema, J., and Tollervey, D. (1999). Ribosome synthesis in Saccharomyces cerevisiae. Annu Rev Genet *33*, 261-311.

Wang, Z., and Roeder, R.G. (1997). Three human RNA polymerase III-specific subunits form a subcomplex with a selective function in specific transcription initiation. Genes & development *11*, 1315-1326.

Warner, J.R. (1999). The economics of ribosome biosynthesis in yeast. Trends Biochem Sci *24*, 437-440.

Weilbaecher, R.G., Awrey, D.E., Edwards, A.M., and Kane, C.M. (2003). Intrinsic transcript cleavage in yeast RNA polymerase II elongation complexes. The Journal of biological chemistry *278*, 24189-24199.

White, R.J. (2005). RNA polymerases I and III, growth control and cancer. Nature reviews Molecular cell biology *6*, 69-78.

Wind, M., and Reines, D. (2000). Transcription elongation factor SII. BioEssays : news and reviews in molecular, cellular and developmental biology *22*, 327-336.

Wu, C.C., Lin, Y.C., and Chen, H.T. (2011). The TFIIF-like Rpc37/53 Dimer Lies at the Center of a Protein Network to Connect TFIIIC, Bdp1, and the RNA Polymerase III Active Center. Molecular and cellular biology.

Zhao, J., Yuan, X., Frodin, M., and Grummt, I. (2003). ERK-dependent phosphorylation of the transcription initiation factor TIF-IA is required for RNA polymerase I transcription and cell growth. Molecular cell *11*, 405-413.

# VI Appendix

## VI.1 Alignments of subunits in Pol I and their homologs in Pol II

The alignment is taken from Kuhn et al. 2007. Regions previously assigned to be of conserved fold are underlined and marked green. Additional regions of conserved fold derived from crosslinking and MS analysis are depicted in cyan (A-B cross-links) and yellow (B*-B crosslinks). All lysin residues involved in crosslinks are bold. Additional regions of conserved fold might exist but cannot be predicted with certainty.

**A190-Rpb1 edited by hand according to 3D structure, EM density and crosslinking-MS**

```
A190    ---MDISKPVGSEITSVDFGILTAKEIRNLSAKQITNPTVLDNLG-HPVSGGLYDLALGA 56
Rpb1    MVGQQYSSAPLRTVKEVQFGLFSPEEVRAISVAKIRFPETMDETQTRAKIGGLNDPRLGS 60
                        *  **     *  *  *    *    *       *** *   **


A190    FLRNL-CSTCGLDEKFCPGHQGHIELPVPCYNPLFFNQLYIYLRASCLFCHHFRLKSVE- 114
Rpb1    IDRNLKCQTCQEGMNECPGHFGHIDLAKPVFHVGFIAKIKKVCECVCMHCGKLLLDEHNE 120
          *** * **    **** *** *  *     *          *   *    *


A190    VHRYACKLRLLQYGLIDESYKLDEITLGSLNSSMYTDDEAIEDNEDEMDGEGSKQSKDISS 175
Rpb1    LMRQALAIKDSKKRFAAIWTLCKTKMVCE----------------------------- 149
          *  * *


A190    TLLNELKSKRSEYVDMAIAKALSDGRTTERGSFTATVNDERKKLVHEFHKKLLSRGKCDN 235
Rpb1    --------------------------------------TDVPSE--DDPTQLVSRGG 166
                                                       *


A190    CGMFSPKFRKDGFTKIFETALNEKQITNNRVKGFIRQDMIKKQKQAKKLDGSNEASANDE 295
Rpb1    CGNTQPTIRKDGLKLVGS------------------------------------- 184
        **    *   ****


A190    ESFDVGRNPTTRPKTGSTYILSTEVKNILDTVFRKEQCVLQYVFHSRPNLSRKLVKADSF 355
Rpb1    ----------WKKDRATGDADEPELRVLSTEEILNIFKHISVKDFTSLGFNEVFSRPEWM 234
                                        *


A190    FMDVLVVPPTRFRLPSKLGEEVHENSQNQLLSKVLTTSLLIRDLNDDISKLQKDKVSLED 415
Rpb1    ILTCLPVPPPPVRPSISFNESQRG---EDDLTFKLADILKANISLETLEHNGAP------ 285
            * ***    *         *        *   *    *          *


A190    RRVIFSRLMNAFVTIQNDVNAFIDSTKAQG-RTSGKVPIPGVKQALEKKEGLFRKHMMGKR 475
Rpb1    --HHAIEEAESLLQFHVATYMDNDIAGQPQALQKSGRPVKSIRARLKGKEGRIRGNLMGKR 344
                                 *             *       * ***  *    ****


A190    VNYAARSVISPDPNIETNEIGVPPVFAVKLTYPEPVTAYNIAELRQAVINGPDKWPGATQ 535
Rpb1    VDFSARTVISGDPNLELDQVGVPKSIAKTLTYPEVVTPYNIDRLTQLVRNGPNEHPGAKY 404
        *    ** *** ***  *     *** *  * ***** ** *** * * * ***   ***


A190    IQNEDGSLVSLIGMSVEQRKALANQLLTPSSNVSTHTLNKKVYRHIKNRDVVLMNRQPTI 595
Rpb1    VIRDSGDRIDLR--------------YSKRAGDIQLQYGWKVERHIMDNDPVLFNRQPSL 450
             *     *                       ** ***   * ** **** *


A190    HKASMMGHKVRVLPNEKTLRLHYANTGAYNADFDGDEMNMHFPQNENARAEALNLANTDS 655
Rpb1    HKMSMMAHRVKVIPYS-TFRLNLSVTSPYNADFDGDEMNLHVPQSEETRAELSQLCAVPL 509
        ** *** * * *    * **    * *********** * ** *  *  ***   *
```

```
A190   QYLTPTSGSPVRGLIQDHISAGVWLTSKDSFFTREQYQQYIYGCIRPEDGHTTRSKIVTL 715
Rpb1   QIVSPQSNKPCMGIVQDTLCGIRKLTLRDTFIELDQVLNMLYWVPDWDG--------VIP 561
         *   * *  * *   **   **  * *   *    *            *

A190   PPTIFKPYPLWTGKQIITTVLLNVTPPDMPGINLISKNKIKNEYWGKGSLENEVLFKDGA 775
Rpb1   TPAIIKPKPLWSGKQILSVAIP----------NGIHLQRFDEGTTLLSPKDNGMLIIDGQ 611
         * * ** ***  ****             * *           *  *   **

A190   LLCGILDKSQYGASKYGIVHSLHEVYGPEVAAKVLSVLGRLFTNYITATAFTCGMDDLRL 835
Rpb1   IIFGVVEKKTVGSSNGGLIHVVTREKGPQVCAKLFGNIQKVVNFWLLHNGFSTGIGDT-- 669
          *    *   *    * *   ** * **     *   *          * * *

A190   TAEGNKWRTDILKTSVDTGREAAAEVTNLDKDTPADDPELLKRLQEILRDNNKSGILDAV 895
Rpb1   -----IADGPTMREITETIAEAKKKVLDVTKEAQAN-----------LLTAKHGMTLRES 713
              *    ** *    *   *   *              *         *

A190   TSSKVNAITSQVVSKCVPDGTMKKFPCNSMQAMALSGAKGSNVNVSQIMCLLGQQALEGR 955
Rpb1   FEDNVVRFLNEARDKAGRLAEVNLKDLNNVKQMVMAGSKGSFINIAQMSACVGQQSVEGK 773
          *          *       *      *   *  * *** * *    *** **

A190   RVPVMVSGKTLPSFKPYETDAMAGGYVKGRFYSGIKPQEYYFHCMAGREGLIDTAVKTSR 1015
Rpb1   RIAFGFVDRTLPHFSKDDYSPESKGFVENSYLRGLTPQEFFFHAMGGREGLIDTAVKTAE 833
        *        *** *        * *      * ***  ** * *************

A190   SGYLQRCLTKQLEGVHVSYDNSIRDADGTLVQFMYGGDAIDITKESHMTQFEFCLDNYYA 1075
Rpb1   TGYIQRRLVKALEDIMVHYDNTTRNSLGNVIQFIYGEDGMDAAHIEKQ-SLDTIGGSDAA 892
         ** ** *  * *   * ***   *   *   ** ** *            *

A190   LLKKY------------------------------------------------------- 1080
Rpb1   FEKRYRVDLLNTDHTLDPSLLESGSEILGDLKLQVLLDEEYKQLVKDRKFLREVFVDGEA 952
          * *                  *      *     *   * *   *        *

A190   -------NPSALIEHLDVESALKYSKKTLKYRKKHSKEPHYKQSVKYDPVLAKYNPAKYL 1133
Rpb1   NWPLPVNIRRIIQNAQQTFHIDHTKPSDLTIKDIVLGVKDLQENLLVLRGKNEIIQNAQR 1012

A190   GSVSENFQDKLESFLDKNSKLFKSSDGVNEKKFRALMQLKYMRSLINPGEAVGIIASQSV 1193
Rpb1   DAVTLFCCLLRSRLATRRVLQEYRLTKQAFDWVLSNIEAQFLRSVVHPGEMVGVLAAQSI 1072
          *                               ** *** **  * **

A190   GEPSTQMTLNTFHFAGHGAANVTLGIPRLREIVMTASAAIKTPQMTLPIWN--DVSDEQA 1251
Rpb1   GEPATQMTLNTFHFAGVASKKVTSGVPRLKEILN-VAKNMKTPSLTVYLEPGHAADQEQA 1131
        *** *************   ** * *** **    ***  *           ***

A190   DTFCKSISKVLLSEVIDKVIVTETTGTSNTAGGNAARSYVIHMRFFDNNEYSEEYDVSKE 1311
Rpb1   KLIRSAIEHTTLKSVTIASEIYYDPDPRSTVIPEDEEIIQLHFSLLDEEAEQSFDQQSPW 1191
             *     *   *         *         *    *   *        *

A190   ELQNVISNQFIHLLEAAIVKEIKKQKRTTGPDIGVAVPRLQTDVANSSSNSKRLEEDNDE 1371
Rpb1   LLRLELDRAAMNDKDLTMGQVGERIKQTFKNDLFVIWSEDNDEKLIIRCRVVRPKSLDAE 1251
          *               ** *    * *         *            *      *

A190   EQSHKKTKQAVSYDEPDEDEIETMREAEKSSDEEGIDSDKESDSDSEDEDVDMNEQINKS 1431
Rpb1   TEAEEDHMLKKIENTMLENITLR------------------------------------- 1274
                *

A190   IVEANNNMNKVQRDRQSAIISHHRFITKYNFDDESGKWCEFKLELAADTEKLLMVNIVEF 1491
Rpb1   ------------------------------------------------------GVEN 1278
                                                              **

A190   ICRKSIIRQIPHIDRCVHPEPENGKRVLVTEGVNFQAMWDQEAFIDVDGITSNDVAAVLK 1551
Rpb1   IERVVMMKYDRKVPSPTGEYVKEPEWVLETDGVNLSEVMTVPG-IDPTRIYTNSFIDIME 1337
        * *              ** * ***      ** *   *

A190   TYGVEAARNTIVNEINNVFSRYAISVSFRHLDLIADMMTRQGTYLAFNRQGMETS-TSSF 1610
Rpb1   VLGIEAGRAALYKEVYNVIASDGSYVNYRHMALLVDVMTTQGGLTSVTRHGFNRSNTGAL 1398
         * **  *       *  **  *   ** *   *  * * ** **    **  *  *

A190   MKMSYETTCQFLTKAVLDNEREQLDSPSARIVVGKLNNVGTGSFDVLAKVPNAA-- 1664
Rpb1   MRCSFEETVEILFEAGASAELDDCRGVSENVILGQMAPIGTGAFDVMIDEESLVKY 1453 ->CTD
        *  * ***  *  * *       *     *   *  *** ***  ***
```

X

**A135-Rpb2 edited by hand according to 3D structure, EM density and crosslinking-MS**

```
A135  MSKVIKPPGQARTADFRTLERESRFINPPKDKSAFPLLQEAVQPHIGSFNALTEGPDGGL 60
Rpb2  MSDLANSE-KYYDEDPYGFEDESAPITAEDSWAVISAFFREKGLVSQQLDSFNQFVDYTL 59
      **                *     * **   *                            *  *

A135  LNLGVKDIGEKVIFDGKPLNSEDEISNSGYLGNKLSVSVEQVSIAKPMSNDGVSSAVERK 120
Rpb2  QDIICEDS--TLILEQLAQHTTE----SDNISRKYEISFGKIYVTKPMVNE--SDGVTHA 111
         *       *              *        *    *     *** *    *  *

A135  VYPSESRQRLTSYRGKLLLKLKWSVNN-----GEENLFEVRD------------CGGLP 162
Rpb2  LYPQEARLRNLTYSSGLFVDVKKRTYEAIDVPGRELKYELIAEESEDDSESGKVFIGRLP 171
       ** * * *    *    *      *          * *   *              * **

A135  VMLQSNRCHLNKMSPYELVQHKEESDEIGGYFIVNGIEKLIRMLIVQRRNHPMAIIRPSF 222
Rpb2  IMLRSKNCYLSEATESDLYKLKECPFDMGGYFIINGSEKVLIAQERSAGNIVQVFKKAAP 231
       ** *  * *         *   **   ***** ** **            *

A135  ANRGASYSHYGIQIRSVRPDQTSQTNVLHYLNDGQVTFRFSWRKNEYLVPVVMILKALCH 282
Rpb2  SPISHVAEIRSALEKGSRFISTLQVKLYGREGSSARTIKATLPYIKQDIPIVIIFRALGI 291
                      *     * *         *              * * * **

A135  TSDREIFDGIIGNDVKDSFLTDRLELLLRGFKKRYPHLQNRTQVLQYLGDKFRVVFQASP 342
Rpb2  IPDGEILEHIC-YDVNDWQMLEMLKPCVEDG----FVIQDRETALDFIGR--RGTALGIK 344
        * **    *   **  *     *            * * *   * * *

A135  DQSDLEVGQEVLDRIVLVHLGKDG--SQDKFRMLLFMIRKLYSLVAGECSPDNPDATQHQ 400
Rpb2  KEKRIQYAKDILQKEFLPHITQLEGFESRKAFFLGYMINRLLLCALDRKDQDDRDHFGKK 404
             *       * *           *   *   **  *       *  *

A135  EVLLGGFLYGMILKEKIDEYLQNIIAQVRMDINRGMAINFKDKRYMSRVLMRVNENIGSK 460
Rpb2  RLDLAGPLLAQLFKTLFKKLTKDIFRYMQRTVEEAHDFNMK--------LAINAKTITSG 456
       * *   *        *       *              * *        *       * *

A135  MQYFLSTGNLVSQSGLDLQQVSGYTVVAEKINFYRFISHFRMVHRGSFFAQLKTTTVRKL 520
Rpb2  LKYALATGNWGEQK-KAMSSRAGVSQVLNRYTYSSTLSHLRRTN-TPIGRDGKLAKPRQL 514
       * * ***   *           *   *       ** *          *      * *

A135  LPESWGFLCPVHTPDGSPCGLLNHFAHKCRISTQQSDVSRIPSILYSLGVAPASHTFAAG 580
Rpb2  HNTHWGLVCPAETPEGQACGLVKNLSLMSCISVG-TDPMPIITFLSEWGMEPLEDYVPHQ 573
         **   ** ** *   *** *         **   *    *   *   *   *

A135  -PSLCCVQIDGKIIGWVSHEQGKIIADTLRYWKVEGKTPGLPIDLEIG----YVPPSTRGQ- 636
Rpb2  SPDATRVFVNGVWHGV--HRNPARLMETLRTLRRKGDINPEVSMIRDIREKELKIFTDAGRV 633
            *   *   *   *       ***   *                      *

A135  YPGLYLFGG---------HSRMLRPVRYLPLDK---------------------EDIV 662
Rpb2  YRPLFIVEDDESLGHKELKVRKGHIAKLMATEYQDIEGGFEDVEEYTWSSLLNEGLVEYI 693
       *  *                      *

A135  GPFEQVYMNIAVTPQEIQ---------------------NNVHTHVEFTPTNILSILA 701
Rpb2  DAEEEESILIAMQPEDLEPAEANEENDLDVDPAKRIRVSHHATTFTHCEIHPSMILGVAA 753
          *    ** *                          ** *  **  *

A135  NLTPFSDFNQSPRNMYQCQMGKQTMGTPGVALCHRSDNKLYRLQTGQTPIVKANLYDDYG 761
Rpb2  SIIPFPDHNQSPRNTYQSAMGKQAMGVFLTNYNVRMDTMANILYYPQKPLGTTRAMEYLK 813
        ** * ****** **  **** **      *  *       *     *   * *

A135  MDNFPNGFNAVVAVISYTGYDMDDAMIINKSADERGFGYGTMYKTEK-VDLALNRNRGDP 820
Rpb2  FRELPAGQNAIVAIACYSGYNQEDSMIMNQSSIDRGLFRSLFFRSYMDQEKKYGMSITET 873
        * ** **   * **    *   * ** *     *

A135  ITQHFGFGNDEWPKEWLEKLDEDGLPYIGTYVEEGDPICAYFDDT-------LNKTKIKT 873
Rpb2  FEKPQRTNTLRMKHGTYDKLDDDGLIAPGVRVSGEDVIIGKTTPISPDEEELGQRTAYHS 933
                        *** ***    *    *  *   *                 *

A135  YHSSEPAYIEEVNLIGDESNKFQE---LQTVSIKYRIRRTPQIGDKFSSRHGQKGVCSRK 930
Rpb2  KRDASTPLRSTENGIVDQVLVTTNQDGLKFVKVRVRTTKIPQIGDKFASRHGQKGTIGIT 993
               * * *                 * *  *        ******* *******

A135  WPTIDMPFSETGIQPDIIINPHAFPSRMTIGMFVESLAGKAGALHGIAQDSTPWIFNEDD 990
Rpb2  YRREDMPFTAEGIVPDLIINPHAIPSRMTVAHLIECLLSKVAALSGNEGDASPFT----D 1049
       ****    ** ** ****** *****     * * * ** * *          *

A135  TPADYFGEQLAKAGYNYHGNEPMYSGATGEELRADIYVGVVYYQRLRHMVNDKFQVRSTG 1050
```

```
Rpb2    ITVEGISKLLREHGYQSRGFEVMYNGHTGKKLMAQIFFGPTYYQRLRHMVDDKIHARARG 1109
              *   **   * * ** * **  * *  *  *  ********* **   *  *

A135    PVNSLTMQPVKGRKRHGGIRVGEMERDALIGHGTSFLLQDRLLNSSDYTQASVCRECGSI 1110
Rpb2    PMQVLTRQPVEGRSRDGGLRFGEMERDCMIAHGAASFLKERLMEASDAFRVHICGICGLM 1169
         *   **  *** **  * ** * ******  * **    *  **    **       *  **

A135    LTTQQSVPRIGSISTVCCRRCSMRFEDAKKLLTKSEDGEKIFIDDSQIWEDGQGNKFVGG 1170
Rpb2    TVIAKLN-----HNQFECKGCDN-----------------------------------K 1188
                        *  *

A135    NETTTVAIPFVLKYLDSELSAMGIRLRYNVEPK--- 1203
Rpb2    IDIYQIHIPYAAKLLFQELMAMNITPRLYTDRSRDF 1224
                **    * *  ** ** *   *
```

**AC40-Rpb3 edited by hand according to 3D structure, EM density and crosslinking-MS**

```
AC40    MSNIVGIEYNRVTNTTSTDFPGFSKDAENEWNVEKFKKDFEVNISSLDAREANFDLINID 60
Rpb3    --------------------------------MSEEGPQVKIREASKDNVDFILSNVD 26
                                         *  *         * *  * *

AC40    TSIANAFRRIMISEVPSVAAEYVYFFNNTSVIQDEVLAHRIGLVPLK-VDPDMLTWVDSN 119
Rpb3    LAMANSLRRVMIAEIPTLAIDSVEVETNTTVLADEFIAHRLGLIPLQSMDIEQLEYSRDC 86
          **   ** ** * *   *     ** * ** *** ** **   *   *

AC40    LPDDEKFTDENTIVLSLNVKCTRNPDAPKGSTDPKELYNNAHVYARDLKFEPQGRQSTTF 179
Rpb3    FCED--HCDKCSVVLTLQAFGESE-------------STTNVYSKDLVIVSNLMGRNIG 130
            *     ** *  *              **  **

AC40    ADCPVVPADPDILLAKLRPGQEISLKAHCILGIGGDHAKFSPVSTASYRLLPQINILQPI 239
Rpb3    HPIIQDKEGNGVLICKLRKGQELKLTCVAKKGIAKEHAKWGPAAAIEFEYDPWNKLKH-- 188
              * *** *** *    **   *** *           *

AC40    KGESARRFQKCFPPGVIGIDEGSDEAYVKDARKDTVSREVLRYEEFADK---VKLGRVRN 296
Rpb3    ---------------------TDYWYEQDSAKEWPQSKNCEYEDPPNEGDPFDYKAQAD 226
                             *  *   *   *           **

AC40    HFIFNVESAGAMTPEEIFFKSVRILKNKAEYLKNCPITQ-------------------- 335
Rpb3    TFYMNVESVGSIPVDQVVVRGIDTLQKKVASIL-LALTQMDQDKVNFASGDNNTASNMLG 282
          *  **** *              * *         **

AC40    ------------------------------ 356
Rpb3    SNEDVMMTGAEQDPYSNASQMGNTGSGGYDNAW 318
```

**AC19-Rpb11 edited by hand according to 3D structure, EM density and crosslinking-MS**

```
AC19    MTEDIEQKKTATEVTPQEPKHIQEEEEQDVDMTGDEEQEEEPDREKIKLLTQATSEDGTS 60
Rpb11   -----------------------------MNAPDRFELFLLGEGESKLKIDPDTKAPNA 30
                                      *            **

AC19    ASFQIVEEDHTLGNALRYVIMKNPDVEFCGYSIPHPSENLLNIRIQTYGETTAVDALQKG 120
Rpb11   VVITFEKEDHTLGNLIRAELLNDRKVLFAAYKVEHPFFARFKLRIQTTEGYDPKDALKNA 90
           *******  *      *  *   **    ****     ***

AC19    LKDLMDLCDVVESKFTEKIKSM-------- 142
Rpb11   CNSIINKLGALKTNFETEWNLQTLAADDAF 120
                *
```

**A12.2-Rpb9 edited by hand according to 3D structure, EM density and crosslinking-MS**

```
A12.2   MSVVGSLIFCLDCGDLLENPNAVLG---SNVECSQCKAIYPKSQFSNLKVVTTTADDAFPSSLRAKKSVVKTSL 71
Rpb9    ---MTTFRFCRDCNNMLY-PREDKENNRLLFECRTCSYVEEAGS----------PLVYRHELITNIGETAGVVQ 60
              ** **   *  *       *   ** *

A12.2   KKNELKDGATIKEKCPQCGNEEMNYHTLQLRSADEGATVFYTCTSCGYKFRTNN-------- 125
Rpb9    DIGSDPTLPRSDRECPKCHSRENVFFQSQQRRKDTSMVLFFVCLSCSHIFTSDQKNKRTQFS 122
              ** *   *      * *  *      *   **   *
```

```
A12.2 C-terminus       -----RAKKSVVKTSLKKNE---LKDGATIKEKCPQCGNEEMNYHTLQLRSADEGATVFY 52
TFIIS 3rd domain       PAPLKQKIEEIAKQNLYNAQGATIERSVTDRFTCGKCKEKKVSYYQLQTRSADEPLTTFC 60
                                   *  *        *    *  *      *  ** ***** * *

A12.2 C-terminus       TCTSCGYKFRTNN 65
TFIIS 3rd domain       TCEACGNRWKFS- 72
                       **  **
```

## VI.2 Alignments of subunits in Pol III and their homologs in Pol II

The alignment is taken from Jasiak et al. 2006. Regions previously assigned to be of conserved fold are underlined and marked green. Additional regions of conserved fold derived from crosslinking and MS analysis are depicted in cyan (A-B cross-links) and yellow (B*-B crosslinks). All lysin residues involved in crosslinks are bold. Additional regions of conserved fold might exist but cannot be predicted with certainty.

**Rpb1-C160 (Rpc1)**

```
Rpb1    MVGQQYSSAPLRTVKEVQFGLFSPEEVRAISVAKIRFPETMDETQTRAKIG-GLNDPRLG  59
C160    -MKEVVVSETPKRIKGLEFSALSAADIVAQSEVEVSTRDLFDLEKDRAPKANGALDPKMG  59
        : :    * . : :* ::*. :*. :: * * .::   : :* : **  . *  **::*

Rpb1    SIDRNLKCQTCQEGMNECPGHFGHIDLAKPVFHVGFIAKIKKVCECVCMHCGKLLLDEHN 119
C160    VSSSSLECATCHGNLASCHGHFGHLKLALPVFHIGYFKATIQILQGICKNCSAILLSETD 119
         . .*:* **: .: .* *****:.** ****:*::    :: : :* :*. :**.* :

Rpb1    ------ELMRQALAIKDSKKRFAAIWTLCKTKMVCETDVPSEDDPTQLVSRGGCGNTQPT 173
C160    KRQFLHELRRPGVDNLRRMGILKKILDQCKKQRRCLHCGALNG--VVKKAAAGAGSAALK 177
              ** * .:       : *   **.: *    . :.... : .*.*.: .

Rpb1    IRKDGLKLVG--------SWKKDRATGDADEPELR--------VLSTEEILNIFKHISVK 217
C160    IIHDTFRWVGKKSAPEKDIWVGEWKEVLAHNPELERYVKRCMDDLNPLKTLNLFKQIKSA 237
        * :* :: **        * :    *.:***.     *.. : **:**:*.

Rpb1    DFTSLGFNEVFS--RPEWMILTCLPVPPPPVRPSISFNES-QRGEDDLTFKLADILKANI 274
C160    DCELLGIDATVPSGRPETYIWRYLPAPPVCIRPSVMMQDSPASNEDDLTVKLTEIVWTSS 297
        *   **:: ... *** *   **.** :***: :::*  .*****.**::*: :.

Rpb1    SLETLEHNGAPHHAIEEAESLLQFHVATYMDNDIAGQPQALQKSG-----RPVKSIRARL 329
C160    LIKAGLDKGISINNMMEHWDYLQLTVAMYINSDSVNPAMLPGSSNGGGKVKPIRGFCQRL 357
        :::  .:* . : : *  . **: ** *::.* ... .    .*.     :*::.:  **

Rpb1    KGKEGRIRGNLMGKRVDFSARTVISGDPNLELDQVGVPKSIAKTLTYPEVVTPYNIDRLT 389
C160    KGKQGRFRGNLSGKRVDFSGRTVISPDPNLSIDEVAVPDRVAKVLTYPEKVTRYNRHKLQ 417
        ***:**:**** *******.***** ****.:*:*.**. :**.***** ** ** .:*

Rpb1    QLVRNGPNEHPGAKYVIR-DSGDRIDLRYSKRAGDIQ-LQYGWKVERHIMDNDPVLFNRQ 447
C160    ELIVNGPNVHPGANYLLKRNEDARRNLRYGDRMKLAKNLQIGDVVERHLEDGDVVLFNRQ 477
        :*: **** ****:*:::: :.. * :***..*   : ** *  ****: *.* ******

Rpb1    PSLHKMSMMAHRVKVIPYSTFRLNLSVTSPYNADFDGDEMNLHVPQSEETRAELSQLCAV 507
C160    PSLHRLSILSHYAKIRPWRTFRLNECVCTPYNADFDGDEMNLHVPQTEEARAEAINLMGV 537
        ****::*:::* .*: *: ***** .* :*****************:**:*** :* .*
```

```
Rpb1   PLQIVSPQSNKPCMGIVQDTLCGIRKLTLRDTFIELDQVLNMLYWVPD--WDGVIPTPAI  565
C160   KNNLLTPKSGEPIIAATQDFITGSYLISHKDSFYDRATLTQLLSMMSDGIEHFDIPPPAI  597
       :::::*:*.:* :. .** : *   :: :*:* :   : ::* :.*   .  **.***


Rpb1   IKPKPLWSGKQILSVAIP---NGIHLQRFDEGTTLLSP-----------KDNGMLIIDG-  610
C160   MKPYYLWTGKQVFSLLIKPNHNSPVVINLDAKNKVFVPPKSKSLPNEMSQNDGFVIIRGS  657
       :**   **:***::*: *     *.  :  .:*   ..:: *        :::*::** *


Rpb1   QIIFGVVEKKTVGSSN-GGLIHVVTREKGPQVCAKLFGNIQKVVNFWLLHNGFSTGIGDT  669
C160   QILSGVMDKSVLGDGKKHSVFYTILRDYGPQEAANAMNRMAKLCARFLGNRGFSIGINDV  717
       **: **::*..*..:  .::.: *: *** .*: :..: *:   :* :.*** **.*.


Rpb1   IADGPTMREITETIAEAKKKVLDVTKEAQANLLTAKHGMTLRESFEDNVVRFLNEARDKA  729
C160   TPADDLKQKKEELVEIAYHKCDELITLFNKGELETQPGCNEEQTLEAKIGGLLSKVREEV  777
        . .   ::  * :   * :*  ::  . : . * ::  * . .:::* :: :*.:.*::.


Rpb1   GRLAEVNLKDLNNVKQMVMAGSKGSFINIAQMSACVGQQSVEGKRIAFGFVDRTLPHFSK  789
C160   GDVCINELDNWNAPLIMATCGSKGSTLNVSQMVAVVGQQIISGNRVPDGFQDRSLPHFP K  837
       * :.  :*.: *    *. .***** :*::** * **** :.*:*:. ** **:****.*


Rpb1   DDYSPESKGFVENSYLRGLTPQEFFFHAMGGREGLIDTAVKTAETGYIQRRLVKALEDIM  849
C160   NSKTPQSKGFVRNSFFSGLSPPEFLFHAISGREGLVDTAVKTAETGYMSRRLMKSLEDLS  897
       :. :*:*****.**:: **:* **:***:.*****:***********:.***:*:***:


Rpb1   VHYDNTTRNSLGNVIQFIYGEDGMDAAHIEKQSLDTIGGSDAAFEKRYRVDLLNTDHTLD  909
C160   CQYDNTVRTSANGIVQFTYGGDGLDPLEMEGNAQPVN--FNRSWDHAYNITFNNQDKGLL  955
        :****.*.* ..::** ** **:*. .:* :: . .. : :::: *.: : * *: *


Rpb1   P-SLLESGSEILGDLKLQVLLDEEYKQLVKDRKFLREVFVDGEANWPLPVNIRRIIQNAQ  968
C160   PYAIMETANEILGPLEERLVRYDNSGCLVKREDLNKAEYVD------QYDAERDFYHSLR  1009
       * :::*:..**** *: ::: ::   *** ..: :  :**..:. . .      ..:.


Rpb1   QTFHIDHTKPSDLTIKDIVLGVKDLQENLLVLRGKNEIIQNAQRDAVTLFCCLLRSRLAT  1028
C160   EYIN----------------GKATALANLRKSRGMLGLLEPPAKELQGIDPDETVPDNVK  1053
                           *    **   **   ::: . ::  :      .  ..


Rpb1   RRVLQEYRLTKQAFDWVLSNIEAQFLRSVVHPGEMVGVLAAQSIGEPATQMTLNTFHAG  1088
C160   TSVSQLYRISEKSVRKFLEIALFKYRKARLEPGTAIGAIGAQSIGEPGTQMTLKTFHAG  1113
        * * **::::. .*.    :: :: :.** :*.:.:******.*****:******


Rpb1   VASKKVTSGVPRLKEILNVAKNMKTPSLTVYLEPGHAADQEQAKLIRSAIEHTTLKSVTI  1148
C160   VASMNVTLGVPRIKEIINASKVISTPIINAVLVN--DNDERAARVVKGRVEKTLLSDVAF  1171
       *** :** ****:***:*.:* :.** :.. * .   *:. *:::. :*:* *..*::


Rpb1   ASEIYYDPDPRSTVIPEDEEIIQLHFSLLDEEAEQSFDQQSPWLLRLELDRAAMNDKDLT  1208
C160   YVQDVYKDN------------------------LSFIQVRIDLGTIDKLQLELTIEDIA  1206
        :  *.:. ::  .....  . :   ...:. ** *    * ::  :  . :. :*:


Rpb1   MGQVGERIKQTFKNDLFVIWSEDNDEKLIIRCRVVRPKSLDAETEAEEDHMLKKIEN-TM  1267
C160   VAITRASKLKIQASDVNIIGKDRIAINVFPEGYKAKSISTSAKEPSENDVFYRMQQLRRA  1266
       :.. :   .*: :* .: ::: .  .:. * .*: :*:*:* : :
```

```
Rpb1  LENITLRGVENIERVVMMKYDRKVPSPTGEYVKEPEWVLETDGVNLSEVMTVPGIDPTRI 1327
C160  LPDVVVKGLPDISRAVINIRD------------DGKRELLVEGYGLRDVMCTDGVIGSRT 1314
      *  ::.::*: :*.*.*:     *  .  .:.:..  .: :  *  .:* .* :** . *:  :*

Rpb1  YTNSFIDIMEVLGIEAGRAALYKEVYNVIASDGSYVNYRHMALLVDVMTTQGGLTSVTRH 1387
C160  TTNHVLEVFSVLGIEAARYSIIREINYTMSNHGMSVDPRHIQLLGDVMTYKGEVLGITRF 1374
       **  .:::::.******.* :: :*:  .::..*  *: **: ** **** :* : .:**.

Rpb1  GFNRSNTGALMRCSFEETVEILFEAGASAELDDCRGVSENVILGQMAPIGTGAFDVMIDE 1447
C160  GLSKMRDSVLQLASFEKTTDHLFDAAFYMKKDAVEGVSECIILGQTMSIGTGSFKVVKG- 1433
      *:.: . ..*  .***:*.: **:*.    : *  .**** :****  .****:*.*: .

Rpb1  ESLVKYMPEQKITEIEDGQDGGVTPYSNESGLVNADLDVKDELMFSPLVDSGSNDAMAGG 1507
C160  ---------------------------------TNISEKDLVPKRCLFESLSNEAALKA 1459
                                       ::::. ** :    *.:* **:*   .

Rpb1  FTAYGGADYGEATSPFGAYGEAPTSPGFGVSSPGFSPTSPTYSPTSPAYSPTSPSYSPTS 1567
C160  N----------------------------------------------------------- 1467
```

**Rpb2-C128 (Rpc2)**

```
Rpb2    ---MSDLANSEKYYDEDPYGFEDESAPI-----------TAEDSWAVISAFFREKGLVSQ 46
C128    MVAATKRRKTHIHKHVKDEAFDDLLKPVYKGKKLTDEINTAQDKWHLLPAFLKVKGLVKQ 60
           :.  ::. : . .  .*:*    *:          *:*.* ::.**:: ****.*


Rpb2    QLDSFNQFVDYTLQDIICEDSTLILEQLAQHTTESDNISRKYEISFGKIYVTKPMVNESD 106
C128    HLDSFNYFVDTDLKKIIKAN-------------QLILSDVDPEFYLKYVDIRVGKKSSSS 107
        :***** *** *:.**  :                          ..*.


Rpb2    GVTHALYPQEARLRNLTYSSGLFVDVKKRTYEAIDVPGRELKYELIAEESEDDSESGKVF 166
C128    TKDYLTPPHECRLRDMTYSAPIYVDIEYTR-------GR----NIIMHKD--------VE 148
         :   *:*.***:***: ::**::       **    ::*  .:.          *


Rpb2    IGRLPIMLRSKNCYLSEATESDLYKLKECPFDMGGYFIINGSEKVLIAQERSAGNIVQVF 226
C128    IGRMPIMLRSNKCILYDADESKMAKLNECPLDPGGYFIVNGTEKVILVQEQLSKNRIIVE 208
        ***:*****::* * :* **.: **:***:* *****:**:***::.**: : * : *


Rpb2    KKAAPSPISHVAEIRSALEKGSRFISTLQVKLYGREGSSARTIKATLPYIKQDIPIVIIF 286
C128    ADEKK--GIVQASVTSS—THERKSKTYVITKNGK--------IYLKHNSIAEEIPIAIVL 257
         .    .    *         *         *         * :.:***.*::


Rpb2    RALGIIPDGEILEHICYDVNDWQMLEMLKPCVEDGFVIQDRETALDFIGRRGTALGIKKE 346
C128    KACGILSDLEIMQLVCGNDSSYQDIFAVNLEESSKLDIYTQQQALEYIGAKVKTMRRQKL 317
        :* **:.* **:: :* : ..:* : ::    .. : * :: **::** : .:: :*


Rpb2    KRIQYAKDILQKEFLPHITQLEGFESRKAFFLGYMINRLLLCALDRKDQDDRDHFGKKRL 406
C128    TILQEGIEAIATTVIAHLTVEALDFREKALYIAMMTRRVVMAMYNPKMIDDRDYVGNKRL 377
        . :* . : : . .:.*:*      .*:::*. * .*:::. : *  ****:.*:***


Rpb2    DLAGPLLAQLFKTLFKKLTKDIFRYMQRTVEEAH---DFNMKLAINAKT--ITSGLKYAL 461
C128    ELAGQLISLLFEDLFKKFNNDFKLSIDKVLKKPNRAMEYDALLSINVHSNNITSGLNRAI 437
        :*** *:: **: ****..*:    :::.:::.:    ::: *:**.:: *****: *:


Rpb2    ATGNWGEQKKAMSSRAGVSQVLNRYTYSSTLSHLRRTNTPIGRDGKLAKPRQLHNTHWGL 521
C128    STGNWS-LKRFKMERAGVTHVLSRLSYISALGMMTRISSQFEKSRKVSGPRALQPSQFGM 496
        :****.  *:   .****::**.* :* *:*. : * .: : :. *:: ** *: :::*:


Rpb2    VCPAETPEGQACGLVKNLSLMSCISVGTDPMPIITFLSEWGMEPLEDYVPHQSPDATRVF 581
C128    LCTADTPEGEACGLVKNLALMTHITTDDEEEPIKKLCYVLGVEDITLIDSASLHLNYGVY 556
        :*.*:****:**********:**: *:.. :  ** .:    *:* :   . .     *:
```

```
Rpb2  VNGVWHGVHRNPARLMETLRTLRRKGDINPEVSMIRDIREKELKIFTDAGRVYRPLFIVE 641
C128  LNGTLIGSIRFPTKFVTQFRHLRRTGKVSEFISIYSNSHQMAVHIATDGGRICRPLIIVS 616
      :**.  *  * *::::  :* ***.*.:.  :*:  : ::   ::* **.**: ***:**.


Rpb2  DDESLGHKELKVRKGHIAKLMATEYQDIEGGFEDVEEYTWSSLLNEGLVEYIDAEEEESI 701
C128  DGQSR------VKDIHLRKLLD-------------GELDFDDFLKLGLVEYLDVNEENDS 657
      *.:*        *:. *: **:            *   :..:*: *****:*.:**:.


Rpb2  LIAMQPEDLEPAEANEENDLDVDPAKRIRVSHHATTFTHCEIHPSMILGVAASIIPFPDH 761
C128  YIALYEKDIVP----------------------SMTHLEIEPFTILGAVAGLIPYPHH 693
       **:  :*: *                       ::** **.*  ***..*.:**:*.*


Rpb2  NQSPRNTYQSAMGKQAMGVFLTNYNVRMDTMANILYYPQKPLGTTRAMEYLKFRELPAGQ 821
C128  NQSPRNTYQCAMGKQAIGAIAYNQFKRIDTLLYLMTYPQQPMVKTKTIELIDYDKLPAGQ 753
      *********.******:*.:   *    *:**:  :: ***:*: .*:::* :.: :*****


Rpb2  NAIVAIACYSGYNQEDSMIMNQSSIDRGLFRSLFFRSYMDQEKKYGMSITETFEKPQRTN 881
C128  NATVAVMSYSGYDIEDALVLNKSSIDRGFGRCETRRKTTTVLKRYANHTQDIIGG-MRVD 812
      ** **: .****: **::::*:******: *.   *.      *:*.    : :    *.:


Rpb2  TLRMKHGTYDKLDDDGLIAPGVRVSGEDVIIGKTTPIS-PDEEELGQRTAYHSKRDASTP 940
C128  ENGDPIWQHQSLGPDGLGEVGMKVQSGQIYINKSVPTNSADAPNPNNVNVQTQYREAPVI 872
              ::.*. ***   *::*.. :: *.*::.* . .*  : .: ..  . *:*..


Rpb2  LRSTENGIVDQVLVTTNQDGLKFVKVRVRTTKIPQIGDKFASRHGQKGTIGITYRREDMP 1000
C128  YRGPEPSHIDQVMMSVSDNDQALIKVLLRQNRRPELGDKFSSRHGQKGVCGIIVKQEDMP 932
       *..*  . :***::..::.  ::** :* .: *::****:******. **  ::****


Rpb2  FTAEGIVPDLIINPHAIPSRMTVAHLIECLLSKVAALSGNEGDASPFTDITVEGISKLLR 1060
C128  FNDQGIVPDIIMNPHGFPSRMTVGKMIELISGKAGVLNGTLEYGTCFGGSKLEDMSKILV 992
      *. :*****:*:***.:******.:  *:  .*...*.*.    .: * . .:*.:**:*


Rpb2  EHGYQSRGFEVMYNGHTGKKLMAQIFFGPTYYQRLRHMVDDKIHARARGPMQVLTRQPVE 1120
C128  DQGFNYSGKDMLYSGITGECLQAYIFFGPIYYQKLKHMVLDKMHARARGPRAVLTRQPTE 1052
      ::*::   * :::*.* **: * * ***** ***:*:*** **:******  ******.*


Rpb2  GRSRDGGLRFGEMERDCMIAHGAASFLKERLMEASDAFRVHICGICGLMTVIAKLNHNQF 1180
C128  GRSRDGGLRLGEMERDCVIAYGASQLLLERLMISSDAFEVDVCDKCGLMGYSG------- 1105
      *********:*******:**:.:* **** :****.*.:*. ****   .


Rpb2  ECKGCDNKIDIYQIHIPYAAKLLFQELMAMNITPRLYTDRSRDF 1224
C128  WCTTCKSAENIIKMTIPYAAKLLFQELLSMNIAPRLRLEDIFQQ 1149
      *. *..  :* :: ***********::***:*** :   :
```

**Rpb11-AC19 (Rpc9)**

```
Rpb11   --------------------MNAPDRFELFLLGEGES---------KLKIDPDTKAPNA 30
 AC19   MTEDIEQKKTATEVTPQEPKHIQEEEEQDVDMTGDEEQEEEPDREKIKLLTQATSEDGTS 60
                          ::  :. :: : *: *.        **  :. ::  .:


Rpb11   VVITFEKEDHTLGNLIRAELLNDRKVLFAAYKVEHPFFARFKLRIQTTEGYDPKDALKNA 90
 AC19   ASFQIVEEDHTLGNALRYVIMKNPDVEFCGYSIPHPSENLLNIRIQTYGETTAVDALQKG 120
        . : : :*******:*  :::: .* *..*.: **   ::.:****  . ***::.


Rpb11   CNSIINKLGALKTNFETEWNLQTLAADDAF 120
 AC19   LKDLMDLCDVVESKFTEKIKSM-------- 142
        :.::: ..::::* : :
```


**Rpb3-AC40 (Rpc5)**

```
Rpb3    ---------------------MSEEGP------------QVKIREASKDNVDFILSNVD 26
AC40    MSNIVGIEYNRVTNTTSTDFPGFSKDAENEWNVEKFKKDFEVNISSLDAREANFDLINID 60
                            :*::.           :*:* . .  :.:* * *:*


Rpb3    LAMANSLRRVMIAEIPTLAIDSVEVETNTTVLADEFIAHRLGLIPLQSMDIEQLEYSRDC 86
AC40    TSIANAFRRIMISEVPSVAAEYVYFFNNTSVIQDEVLAHRIGLVPLK-VDPDMLTWVDSN 119
        ::**::**:**:*:*::* : * . .**:*: **.:****:**:**:::* : * :  .


Rpb3    FCED--HCDKCSVVLTLQAFGESE--------------STTNVYSKDLVIVSNLMGRNIG 130
AC40    LPDDEKFTDENTIVLSLNVKCTRNPDAPKGSTDPKELYNNAHVYARDLKFEPQGRQSTTF 179
        : :*  . *: ::**:*:.    :          ..::**::** : .:      .


Rpb3    HPIIQDKEGNGVLICKLRKGQELKLTCVAKKGIAKEHAKWGPAAAIEFEYDPWNKLKH-- 188
AC40    ADCPVVPADPDILLAKLRPGQEISLKAHCILGIGGDHAKFSPVSTASYRLLPQINILQPI 239
             . .:*:.*** ***:*.. .  **. :***:.*.:: .:.  *  :: :


Rpb3    ---------------------TDYWYEQDSAKEWPQSKNCEYEDPPNEGDPFDYKAQAD 226
AC40    KGESARRFQKCFPPGVIGIDEGSDEAYVKDARKDTVSREVLRYEEFADK---VKLGRVRN 296
                             :*  * :*: *:  . :  .**: .::....     :


Rpb3    TFYMNVESVGSIPVDQVVVRGIDTLQKKVASIL-LALTQMDQDKVNFASGDNNTASNMLG 285
AC40    HFIFNVESAGAMTPEEIFFKSVRILKNKAEYLKNCPITQ-------------------- 335
        * :****.*::. :::..:.: *::*. :  .:*** :* : :. : *..:


Rpb3    SNEDVMMTGAEQDPYSNASQMGNTGSGGYDNAW 318
```

## VI.3 Cloning and expression of Rrn7

The TFIIB homolog Rrn7 from different yeast species was cloned into vectors pET21/28 (Novagen) using restriction enzymes or by homologous recombination using the In-Fusion cloning systen (Clontech) and vectors pOPINE and pOPINF (Berrow et al., 2007). Expression was tested by induction with IPTG for 3h at 37°C or overnight at 18°C, full-length Rrn7 from *S.cerevisiae* was also expressed in auto-inducing medium and LSSB medium.

**Table S1. Cloning and expression of Rrn7**

| Organism | Construct | Domains | Vector | Expression |
|---|---|---|---|---|
| *S. cerevisiae* | 1-514 | full-length | pET21 | no expression |
| *S. cerevisiae* | 1-514 | full-length | pET28 | Ok, but massive DnaK contamination |
| *S. cerevisiae* | 63-514 | Cyclin 1+2, C-term. | pET21 | insoluble |
| *S. cerevisiae* | 63-514 | Cyclin 1+2, C-term. | pET28 | insoluble |
| *S. cerevisiae* | 63-300 | Cyclin 1+2 | pET28 | no clones |
| *S. cerevisiae* | 63-175 | Cyclin 1 | pET28 | no expression |
| *S. cerevisiae* | 175-300 | Cyclin 2 | pET28 | no expression |
| *S. cerevisiae* | 175-514 | Cyclin 2, C-term. | pET28 | insoluble |
| *S. cerevisiae* | 300-514 | C-term. | pET28 | insoluble |
| *S. cerevisiae* | 1-514 | full length | pOPINE | insoluble |
| *S. cerevisiae* | 1-514 | full length | pOPINF | insoluble |
| *S. cerevisiae* | 1-281 | Zn, Cyclin 1+2 | pOPINE | insoluble |
| *S. cerevisiae* | 1-281 | Zn, Cyclin 1+2 | pOPINF | insoluble |
| *S. cerevisiae* | 63-281 | Cyclin 1+2 | pOPINE | no expression |
| *S. cerevisiae* | 63-281 | Cyclin 1+2 | pOPINF | no expression |
| *S. cerevisiae* | 69-281 | Cyclin 1+2 | pOPINE | no expression |
| *S. cerevisiae* | 69-281 | Cyclin 1+2 | pOPINF | no expression |
| *S.pombe* | 1-537 | full-legth | pET28 | no expression |
| *S. pombe* | 1-537 | full length | pOPINE | insoluble |
| *S. pombe* | 1-537 | full length | pOPINF | no expression |
| *S. pombe* | 1-307 | Zn, Cyclin 1+2 | pOPINE | insoluble |
| *S. pombe* | 1-307 | Zn, Cyclin 1+2 | pOPINF | insoluble |
| *S. pombe* | 54-307 | Cyclin 1+2 | pOPINE | no expression |
| *S. pombe* | 54-307 | Cyclin 1+2 | pOPINF | insoluble |
| *S. pombe* | 54-537 | Cyclin 1+2, C-term. | pOPINE | insoluble |
| *S. pombe* | 54-537 | Cyclin 1+2, C-term. | pOPINF | no expression |
| *S. pombe* | 325-537 | C-terminus | pOPINE | no clones |

| | | | | |
|---|---|---|---|---|
| *S. pombe* | 325-537 | C-terminus | pOPINF | insoluble |
| *C. glabrata* | 1-543 | full-length | pET28 | no expression |
| *C. glabrata* | 1-543 | full length | pOPINE | insoluble |
| *C. glabrata* | 1-543 | full length | pOPINF | no expression |
| *C. glabrata* | 1-322 | Zn, Cyclin 1+2 | pOPINE | no clones |
| *C. glabrata* | 1-322 | Zn, Cyclin 1+2 | pOPINF | no expression |
| *C. glabrata* | 75-322 | Cyclin 1+2 | pOPINE | insoluble |
| *C. glabrata* | 75-322 | Cyclin 1+2 | pOPINF | insoluble |
| *C. glabrata* | 75-543 | Cyclin 1+2, C-term. | pOPINE | no expression |
| *C. glabrata* | 75-543 | Cyclin 1+2, C-term. | pOPINF | insoluble |
| *C. glabrata* | 340-543 | C-terminus | pOPINE | insoluble |
| *C. glabrata* | 340-543 | C-terminus | pOPINF | no expression |
| *A. gossypii* | 1-516 | full length | pOPINE | no clones |
| *A. gossypii* | 1-516 | full length | pOPINF | insoluble |
| *A. gossypii* | 1-299 | Zn, Cyclin 1+2 | pOPINE | no expression |
| *A. gossypii* | 1-299 | Zn, Cyclin 1+2 | pOPINF | insoluble |
| *A. gossypii* | 67-299 | Cyclin 1+2 | pOPINE | insoluble |
| *A. gossypii* | 67-299 | Cyclin 1+2 | pOPINF | insoluble |
| *A. gossypii* | 67-516 | Cyclin 1+2, C-term. | pOPINE | insoluble |
| *A. gossypii* | 67-516 | Cyclin 1+2, C-term. | pOPINF | insoluble |
| *A. gossypii* | 314-516 | C-terminus | pOPINE | insoluble |
| *A. gossypii* | 314-516 | C-terminus | pOPINF | no clones |

# VII Abbreviations

| | |
|---|---|
| °C | degree Celsius |
| 6-FAM | 6-carboxy-fluoresceine |
| Å | angstrom |
| Amp | ampicillin |
| ATP | adenosine triphosphate |
| bp | base pair |
| Cα | Carbon atom at α-position of the peptide chain |
| Cam | chloramphenicol |
| CE | core element |
| CF | core factor |
| CID | collision induced dissociation |
| CV | column volume |
| Da | Dalton |
| DNA | Deoxyribonucleic acid |
| DSS | di-succimidyl suberate |
| DTT | dithiothreitol |
| *E.c.* | *Escherichia coli* |
| EC | elongation complex |
| EM | electron microscopy |
| ETS | external transcribed spacer |
| eV | electron volt |
| FSC | Fourier shell correlation |
| g | g-force |
| G418 | geneticin |
| GLRF | generally locked rotation function |
| GTF | general transcription factor |
| GTP | guanosin triphosphate |
| HEPES | 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid |
| IGS | intergenic spacer |
| IPTG | isopropyl β-D-1-thiogalactopyranoside |
| Kan | kanamycin |
| LB | Luria-Bertani medium |
| LC-MS/MS | liquid chromatography tandem mass spectrometry |
| m/z | mass to charge ratio |
| mRNA | messenger RNA |
| NCS | non-crystallographic symmetry |
| NiNTA | nickel-nitrilotriacetic acid |
| NOR | nucleolar oragnizer region |

| | |
|---|---|
| NTP | nucleoside triphosphate |
| OD$_{600}$ | optical density at 600 nm wavelength |
| PAGE | polyacrylamid gel electrophoresis |
| PDB | protein data bank |
| PEG | polyethylen glycol |
| PI | protease inhibitor mix |
| Pol | RNA polymerase |
| ppm | parts per million |
| RNA | Ribonucleic acid |
| rRNA | ribosomal RNA |
| *S.c.* | *Saccharomyces cerevisiae* |
| *S.p.* | *Schizosaccharomyces pombe* |
| SDC | synthetic dextrose complete medium |
| SDS | sodium dodecyl sulfate |
| SEC | size exclusion chromatography |
| SINES | short interspersed nuclear elements |
| SLS | Swiss Light Source |
| snoRNP | small nucleolar ribonucleoprotein |
| SOB | super optimal broth medium |
| TAF | TATA-box associated factor |
| TB | terrific broth medium |
| TCEP | tris(2-carboxyethyl)phosphine |
| Tet | tetracyklin |
| TOR | target of rapamycin |
| tRNA | transfer RNA |
| UAF | upstream activating factor |
| UBF | upstream binding factor |
| UE | upstream element |
| v/v | volume per volume |
| w/v | weight per volume |
| YPD | yeast extract peptone dextrose medium |

## VIII  Curriculum vitae

Stefan Jennebach

Date of birth:         18th August 1981

Place of birth:        Heilbad Heiligenstadt

Nationality:           German

Marital status:        unmarried

## Education

1988-1992              Primary school

1992-2000              Secondary school
                       Staatliches Gymnasium "Johann-Georg Lingemann"
                       Degree: "Allgemeine Hochschulreife"

2000-2001              Studies in biology at the Georg-August-Universität
                       Göttingen

2001-2006              Studies in biology at the Ludwig-Maximilians-Universität
                       München
                       Degree: Diploma in biology

Nov. 2006-             PhD thesis at the Gene Center of the Ludwig-Maximilians
                       Universität München in the group of Prof. Dr. Patrick
                       Cramer