Functional Architecture of RNA Polymerase I

Claus-Dieter Kuhn

aus Mutlangen

2008

Acknowledgements

Summary

Part I: Introduction

Part II: Results and Discussion

Acknowledgements

First of all I would like to express my deep gratitude to Prof. Patrick Cramer for creating an excellent research environment in his group. His scientific enthusiasm infected me many times and kept my motivation on a high level even after months of research producing nothing than negative results. I hope to carry on my research in this always positive and motivated spirit, which is extremely admirable, yet hard to maintain.

Many days would not have been thinkable without coffee breaks with you guys, Dengl and Stephan. Talking about everyday's numerous scientific disappointments and rare successes and private chats about football and other important issues happening in the world made every day spent at the Gene Center enjoyable and set the basis for hopefully ongoing friendship.

Due to the interdisciplinary nature of this thesis, I had the pleasure to get in contact with lots of people. Thanks a lot, Marco, for bearing with me in front of the computer when fighting with cryo-EM scripts. Thank you, Sonja, for joining the Pol I project at the right time to finalize the EM structure and thanks to you, Roland, for strengthening this collaboration by moving to Munich.

I wish to thank you, Sebastian for enabling me to experience how it feels like to solve a structure by X-ray crystallography and for constant cooperation during years and you, Michela, for fighting with Pol I data together with me for quite a while. I am also grateful to the slowly forming Pol I team that I could initiate in this lab. I wish you good luck in finally solving Pol I by X-ray crystallography, Jenne and Dirk, and I would like to thank you both, Stefan and Kristin, for excellent technical support.

There is a myriad of people in the Gene Center that I would like to thank. Claudia, thanks for creating the enormously positive atmosphere in the Cramer lab, thank you, Karim, for struggling with the fermenter together with me in the beginning of all this work. Thanks for all crystallographic advice, Laurent and

Parts of this work have been published:

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

- Gerber, J., Reiter, A., Steinbauer, R., Jakob, S., Kuhn, C.-D., Cramer, P., Griesenbeck, J., Milkereit, P., and Tschochner, H. (2007). Site specific phosphorylation of yeast RNA polymerase I. *Nucleic Acids Res.* Advance Access published on December 15, 2007.

- Geiger, S.R., Kuhn, C.-D., Leidig, C., Renkawitz, J., and Cramer, P. Crystallization of the RNA polymerase I subcomplex A14/43 by iterative prediction, probing, and removal of multiple flexible regions. Manuscript in preparation.

Summary

Synthesis of ribosomal RNA by RNA polymerase (Pol) I is the first step in ribosome biogenesis and a regulatory switch in eukaryotic cell growth. In this thesis a reproducible large-scale purification protocol for Pol I from *S. cerevisiae* could be developed. Crystals were obtained, diffraction to < 4 Å could be recorded, however, the enormously complex non-crystallographic symmetry impeded structure solution.

Switching to cryo-electron microscopy, the structure of the complete 14-subunit enzyme could be solved to 12 Å resolution, a homology model for the core enzyme could be generated, and the crystal structure of the subcomplex A14/43 could be solved. In the resulting hybrid structure of Pol I, A14/43, the clamp, and the dock domain contribute to a unique surface interacting with promoter-specific initiation factors. The Pol I-specific subunits A49 and A34.5 form a heterodimer near the enzyme funnel that acts as a built-in elongation factor, and is related to the Pol II-associated factor TFIIF. In contrast to Pol II, Pol I has a strong intrinsic 3'-RNA cleavage activity, which requires the C-terminal domain of subunit A12.2, and apparently enables rRNA proofreading and 3'-end trimming.

Part I: Introduction

I.1 | Eukaryotic RNA polymerases

Transcription of genetic information requires specific multisubunit enzymes, RNA polymerases, that translate the information stored in DNA very reliably into RNA. In bacteria and archaea just a single RNA polymerase is synthesizing all cellular RNA. By contrast, there are 3 different types of enzymes catalyzing DNA-dependent RNA synthesis in eukaryotes (Table 1 for details):

RNA polymerase III transcribes various short non-translated RNA molecules, including the 5S ribosomal RNA (rRNA), transfer RNA (tRNA), 7SL RNA (an essential component of the signal-recognition particle) and RNA molecules required for post-translational processing of rRNA, mRNA and tRNA. In addition, Pol III synthesizes short interspersed nuclear elements (SINES), including for example over 1 million Alu genes in humans (Geiduschek and Kassavetis, 2001).

RNA polymerase II transcribes all protein-coding genes (Cramer, 2004), as well as many small RNA molecules that regulate transcription of other genes through various mechanisms (Dye et al., 2006).

RNA polymerase I (Pol I) is solely dedicated to transcribing ribosomal RNA (rRNA). In yeast rRNA is transcribed in form of a 35S precursor-rRNA, which gets subsequently processed into 25S, 5.8S and 18S rRNA and assembles into native ribosomes.

Throughout this thesis *Saccharomyces cerevisiae* serves as model organism. Most phenomena will be discussed using *S. cerevisiae* as model system, but cross-references to other organisms are given, wherever enough knowledge of that system has been acquired.

Table 1 | Subunit composition of eukaryotic RNA polymerases

Polymerase part	Pol I subunit	MW (kDa)	Corresponding Pol II subunit	Corresponding Pol III subunit	Subunit type
Core	A190	186.4	Rpb1	C160	homolog
	A135	135.7	Rpb2	C128	homolog
	AC40	37.7	Rpb3	AC40	homolog
	AC19	16.2	Rpb11	AC19	homolog
	A12.2	13.7	Rpb9	C11	homolog
	Rpb5 (ABC27)	25.1	Rpb5	Rpb5	common
	Rpb6 (ABC23)	17.9	Rpb6	Rpb6	common
	Rpb8 (ABC14.5)	16.5	Rpb8	Rpb8	common
	Rpb10 (ABC10β)	8.3	Rpb10	Rpb10	common
	Rpb12 (ABC10α)	7.7	Rpb12	Rpb12	common
Subcomplex A14/43	A14	14.6	Rpb4	C17	counterpart
	A43	36.2	Rpb7	C25	counterpart
Subcomplex A49/34.5	A49	46.7	-	C37 (?)	Pol I/III specific
	A34.5	26.9		C53 (?)	Pol I/III specific
Subcomplex C82/34/31	-	-	-	C82	Pol III specific
				C34	Pol III specific
				C31	Pol III specific
Total	14 subunits	589.6	12 subunits	17 subunits	-

I.2 | General importance of rDNA transcription

The ribosome, one of the most ancient and complex molecular machines in the cell, is composed of ~ 60% RNA and ~ 40% protein. The large subunit of the eukaryotic 80S ribosome, the 60S subunit, is composed of 3 RNA species, the 25S rRNA, the 5.8S rRNA and the 5S rRNA, and 42 proteins. The small 40S subunit contains just a single RNA species (18S rRNA) and 32 proteins.

All ribosomal RNA has to be synthesized by RNA polymerase I. Given the overwhelming emphasis paid to transcription of protein coding genes, it is astonishing that every cell has to provide 10 ribosomes per synthesized mRNA molecule. Ribosome biogenesis consumes an enormous fraction of the energy of a cell and needs therefore to be tightly regulated, mainly at the level of rDNA transcription (Grummt, 2003; Moss et al., 2007). As a consequence of this central importance, Pol I transcription accounts for up to 60% of all nuclear transcription, resulting in up 80% of total RNA in a cell (Warner, 1999).

Although highly enlarged nucleoli had been associated with cancer since 1896 (Pianese, 1896), deregulated Pol I and Pol III have just recently been implicated to have major impact on the growth potential of tumors (White, 2005). There is a growing body of evidence that Pol I transcription is one of the key regulators of cell growth and proliferation and a major signaling target after nutrient deprivation.

I.3 | Structural organization of rDNA loci

The nucleolus is the cellular compartment where rDNA transcription takes place (Fig. 1). In fact, rDNA transcription is the basis for the formation of a nucleolus, which is not separated from the nucleoplasm by a membrane. It turns out that the nucleolus hosts not only rDNA transcription, pre-rRNA processing and modification, but is also needed for snRNA- and tRNA-maturation and the biogenesis of ribonucleic particles in general (Thiry and Lafontaine, 2005).
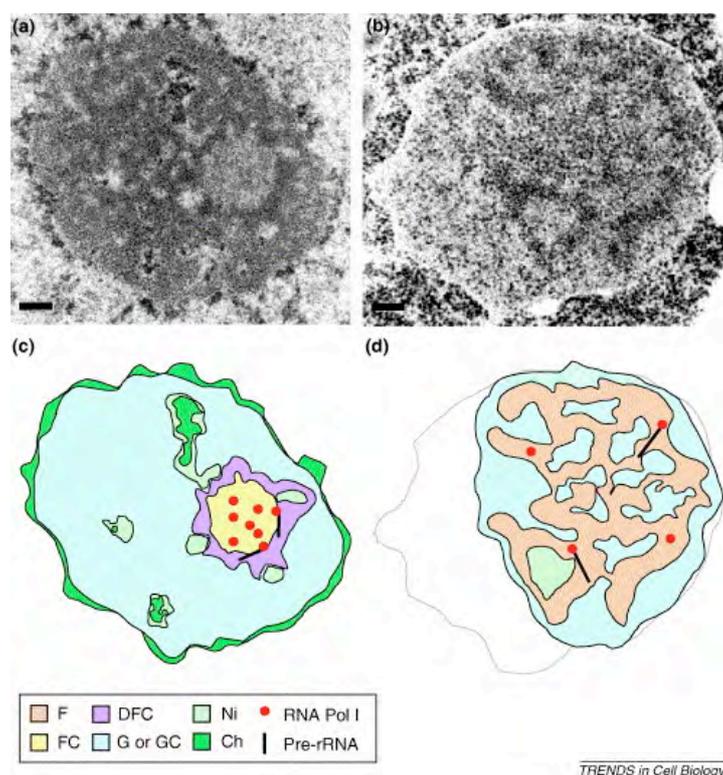
Figure 1 | Nucleolar organization in human and *S. cerevisiae* cells. (a,b) Electron micrographs of a human nucleolus and a yeast nucleus, respectively; Bars, 0.25 µm. (Note that a human nucleolus is as large as a yeast nucleus). **(c,d)** 'Blueprint' cartoons of panels (a) and (b), respectively. Key: F, fibrillar component; FC, fibrillar center; DFC, dense fibrillar component; G or GC, granular component; Ni, nucleolar interstices; Ch, condensed chromatin. In panel (d), the yeast nuclear envelope is outlined in light grey (Thiry and Lafontaine, 2005).

In eukaryotes, rDNA genes are tandemly repeated at one or a few loci. Each repeat is separated from the subsequent one by a intergenic spacer (IGS) region that is important for rDNA silencing (Chapter I.4). *S. cerevisiae* possesses ~ 150 copies of the rRNA gene coding for a 35S precursor rRNA on chromosome XII (Fig. 2). Each repeat contains important sequence elements such as the rDNA promoter, enhancer, the spacer promoter, an origin of replication and a replication fork barrier, that prevents Pol I from colliding with replication forks during S phase (Brewer et al., 1992). Among eukaryotes, *S. cerevisiae* is unique in that the gene for the 5S rRNA, transcribed by Pol III, is part of the rDNA repeats. Eukaryotic rDNA promoter sequences have

diverged significantly, which makes rDNA transcription specific to closely related species. rDNA promoter sequences are not recognized across species barriers (Heix and Grummt, 1995).

Very interestingly, neither tandemly arranged rDNA repeats nor Pol I are absolutely required for cell viability. In a yeast strain lacking the essential Pol I subunit A135, rRNA can be synthesized by Pol II from a multicopy plasmid carrying the 35S rDNA under control of a *GAL7* promoter (Nogi et al., 1991).
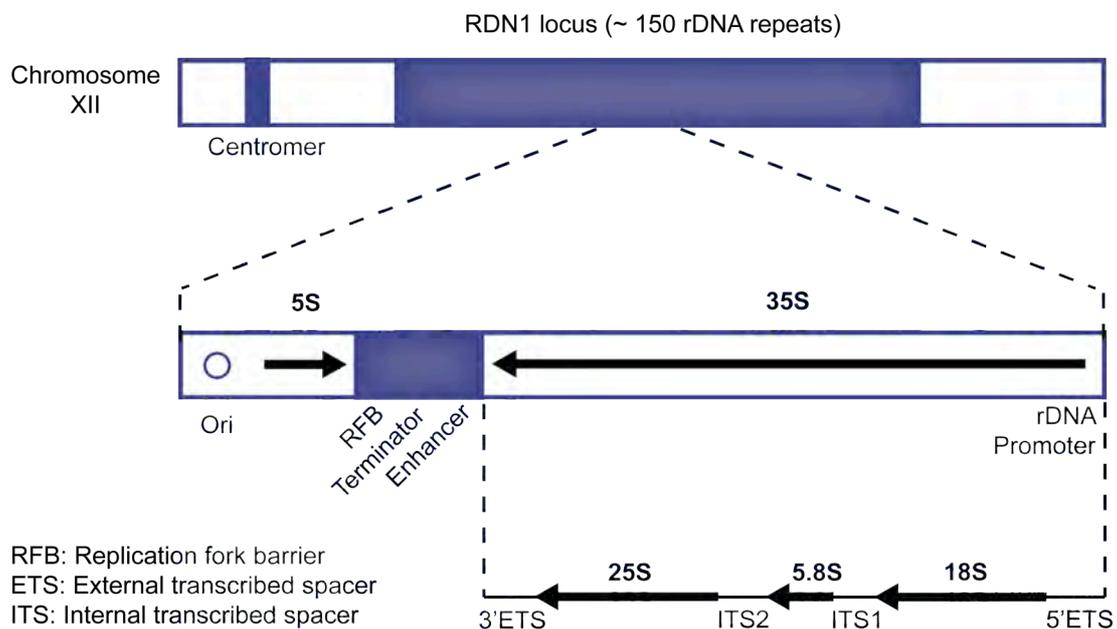
Figure 2 | Structure of the yeast rDNA locus. Figure was prepared based on Nomura, Cold Spring Harb Symp Quant Biol 2001 and Granneman & Baserga, Curr Opin Cell Biol 2005.

I.4 | Epigenetics

Each of the rDNA repeat loci (one in *S. cerevisiae*, five in humans and mouse) is capable of forming a nucleolus when rRNA genes are being transcribed and is therefore also referred to as nucleolar organizer or *NOR* (Nomura, 2001). However, even in exponentially growing cells only half of the rDNA repeats are

active, excluding the obvious possibility of transcription regulation via the number of active genes (French et al., 2003).

Epigenetic marks characterizing heterochromatic and euchromatic rDNA are very similar to protein-coding genes: DNA hypomethylation, acetylation of histone H4 and dimethylation of histone H3 at lysine 4 (H3K4me2) correlate with an 'open' or active chromatin state, whereas CpG methylation, histone H4 hypoacetylation and methylation of H3K9 correlate with transcriptional repression (Earley et al., 2006; Lawrence et al., 2004).

Silencing of rDNA apparently requires transcription of the IGS from the spacer promoter (Mayer et al., 2006). The generated non-coding RNA is processed and incorporated into the nucleolar remodeling complex (NoRC) (Grummt, 2007; Santoro et al., 2002). This complex associates with rDNA in TTF-I dependent manner (Langst et al., 1997) and leads to the recruitment of chromatin modifiers that establish heterochromatin. CSB (Cockayne Syndrome protein B), a SWI/SNF2-like DNA-dependent ATPase, and WSTF (William syndrome transcription factor) seem to be good candidates for establishing active rDNA (Bradsher et al., 2002). Perturbation of this epigenetic balance is associated with alterations in rRNA synthesis and genomic instability, ultimately leading to cell transformation and malignant growth. The most likely benefit from limiting the number of active rDNA repeats seems to be reduced DNA damage and repressed homologous recombination (Grummt, 2007).
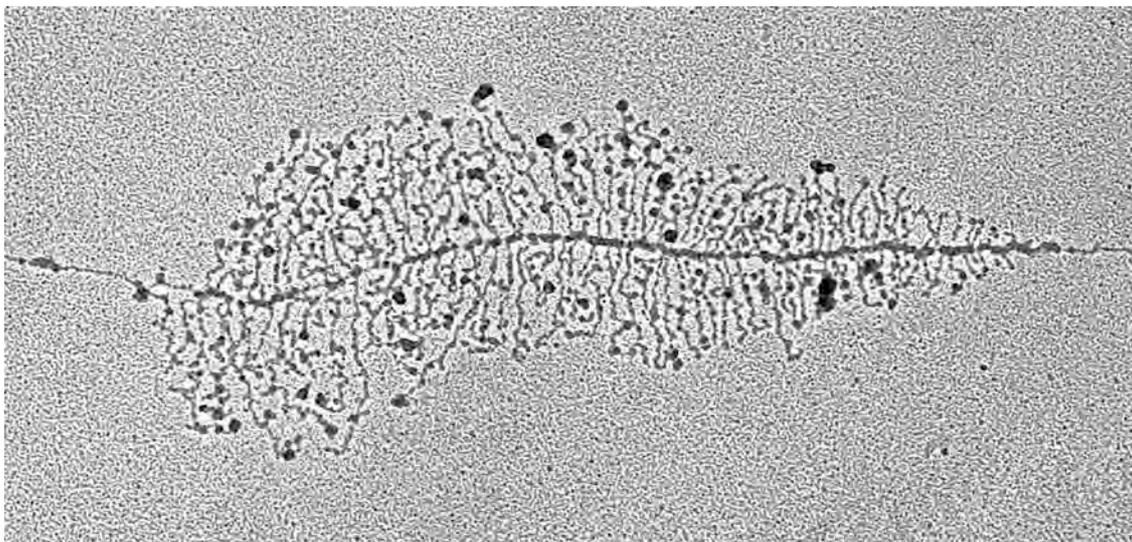
Figure 3 | Miller Spread of a single rDNA repeat in yeast. Cells were heat-shocked to slow down rRNA processing and to keep transcripts long and well defined. The horizontal linear molecule is rDNA, the branches are rRNA molecules currently being transcribed. Image courtesy of Sarah French and Ann Beyer, University of Virginia Health System.

I.5 | The RNA polymerase I transcription cycle

In vertebrates and yeasts, the rDNA promoter is a sequence of 140-160 bp, containing two functional elements, a core promoter sequence and an upstream control element (UE). The spacing of these two elements is important *in vivo*, but *in vitro* the core element is sufficient for transcription initiation.

Formation of a Pol I pre-initiation complex requires the TATA box-binding protein (TBP) and a group of Pol I specific TAFs (TATA-box associated factors), forming one or two complexes recognizing the promoter.

In human and mouse, pre-initiation complex formation requires initial recruitment of SL1 (selectivity factor 1) or TIF-IB, respectively (Bell et al., 1988). UBF (upstream binding factor) has been implicated in the enhancement of Pol I transcription via formation of a putative enhancesome (Bazett-Jones et al., 1994).

In yeast, there are two factors required for complex formation (Fig. 4) (Aprikian et al., 2001; Nomura, 2001): After establishment of the UAF (upstream activating factor) – UE complex, TBP is either already present or recruited along with the core factor. Efficient transcription requires the UAF complex, however, for low-level transcription neither the UAF and UE nor TBP are needed *in vitro* (Keener et al., 1998).

Initiation of transcription in yeast and mammals requires Rrn3 or TIF-IA, respectively. Dependent on the phosphorylation status of Pol I, Rrn3 associates with a small sub-population of Pol I (Fath et al., 2001), rendering the enzyme initiation-competent. In mammals this regulation is complicated by TIF-IA also being regulated by phosphorylation (Zhao et al., 2003a). The Pol I system apparently lacks abortive transcription prior to promoter escape (Stefanovsky et al., 2006a), but there is kinetic evidence for a rate-limiting post-initiation step (Panov et al., 2006).

Once Pol I makes the transition from initiation to elongation it transcribes the 35S-precursor with a speed of ~ 5.6 kb/min (Dundr et al., 2002), which compares well to Pol II (Darzacq et al., 2007). Whereas actively transcribing Pol II molecules are on average 4 kb apart, Pol I is tightly packed on rDNA with one polymerase every 70 bp (Fig. 3). Pausing seems to be a Pol II specific feature, since pausing of so densely packed Pol I molecules would result in catastrophic stalling (Darzacq et al., 2007).

Transcription termination sites are located at the 3' end of the transcribed region, between the spacer and rDNA promoter. TTF-I bends the T-rich termination site, forces Pol I to pause and cooperates with PTRF (Pol I and transcript release factor) to dissociate Pol I from rDNA (Jansa and Grummt, 1999; Russell and Zomerdijk, 2005).
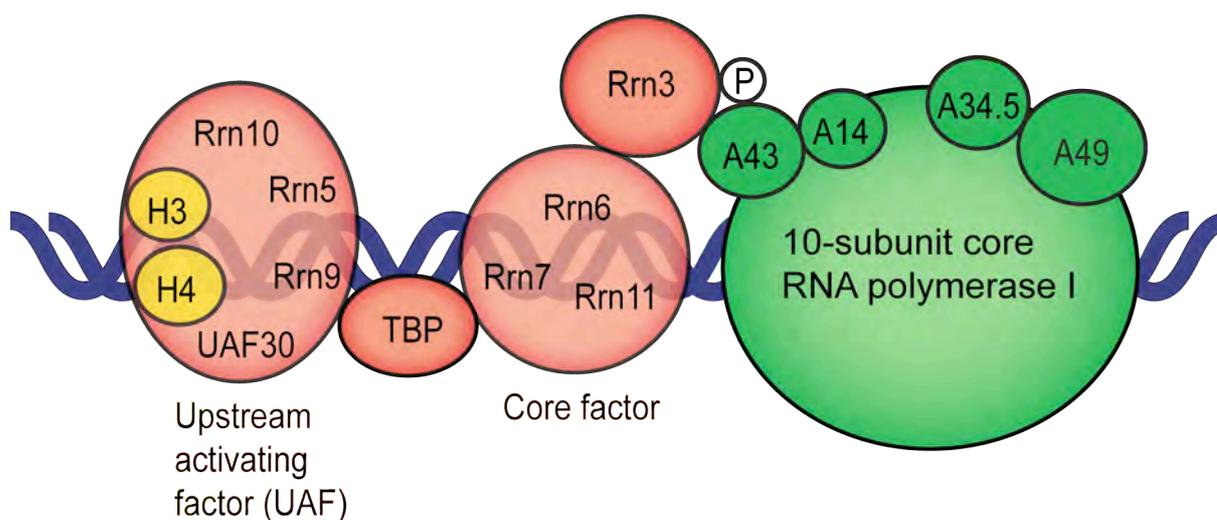
Figure 4 | The pre-initiation complex in *S. cerevisiae*. Figure was prepared based on Moss et al., Cell.Mol.Life Sci. 2007 and Grummt, Genes & Dev 2003. As Pol I possesses up to 15 different phosphorylation sites, the P-circle just indicates the general dependence of the Pol I-Rrn3 interaction on phosphorylation.

I.6 | *In vivo* regulation of rDNA transcription

Without new ribosomes, a cell cannot make protein and hence cannot grow and proliferate. In bacteria, r-protein expression is surveyed by an autoregulatory loop, in which free r-proteins negatively regulate their own synthesis. Thus, bacterial ribosome biosynthesis appears to be regulated mainly at the rRNA synthesis level (Gourse et al., 1986). In eukaryotes, both r-protein and rDNA synthesis are regulated in more sophisticated and interrelated ways, although also eukaryotic r-protein levels clearly depend on the level of rRNA synthesis (Laferte et al., 2006). Some of the eukaryotic regulation mechanisms rely on detection of intact 60S subunits (Zhao et al., 2003b). Blocking nuclear export of 60S subunits leads to a coordinated shutdown of rRNA synthesis and r-protein expression. Strangely, this seems not to be the case for the 40S subunit.

Apart from epigenetic mechanisms (Chapter I.3) almost any perturbation that slows down cell growth or interferes with protein synthesis decreases rDNA

transcription. Pol I transcription initiation seems to be regulated mainly via alterations in the phosphorylation pattern of Rrn3. In mouse, the target of rapamycin (mTOR) nutrient-sensing pathway (Proud, 2002) and the Jun N-terminal kinase pathway (JNK) regulate TIF-IA phosphorylation (Mayer et al., 2005; Mayer et al., 2004). The Raf-MEK-ERK kinase pathway (Zhao et al., 2003a) modulates phosphorylation of TIF-IA in mammals, thereby effecting formation of the TIF-IA-Pol I complex.

As growth factor and MAP kinase activation of rRNA synthesis does not increase the absolute number of transcribing Pol I complexes (Stefanovsky et al., 2006a), Pol I elongation has to be regulated as well. Phosphorylation of UBF seems to be the main tool for controlling elongation. ERK phosphorylates the two N-terminal HMG1 boxes of UBF (Stefanovsky et al., 2001), thereby altering the DNA-bending capacity of UBF. This leads to remodeling of the hypothetical enhanceosome, which facilitates transcription elongation (Stefanovsky et al., 2006b).

I.7 | Making ribosomes

The structure and function of the mature cytoplasmic ribosome is well known (Ban et al., 2000; Schuwirth et al., 2005). However, our knowledge about the pathway resulting in a fully functional ribosome is still very limited (Fatica and Tollervey, 2002; Tschochner and Hurt, 2003). The 18S synthesis pathway, involving four successive endonuclease cleavages, seems to be distinct from 25S/5.8S synthesis, which is much more complex and requires endonuclease cleavages followed by exonuclease digestion. Most RNA processing cleavage sites are used in a well-maintained order, suggesting that many proteins involved in the pathway function in assuring this specific order. The 18S rRNA probably folds into a structure close to the mature form already co-transcriptionally and assembles with its respective r-proteins already on the 35S precursor rRNA. The main portion of 60S r-proteins seems to assemble with

RNA after the 90S pre-ribosome is processed into 66S and 43S pre-ribosomes. Transport into the nucleoplasm and quality control of ribosomal subunits involves, amongst many other factors, differential heterodimeric Noc complexes (nucleolar complex associated proteins) (Milkereit et al., 2001). Export of ribosomal subunits into the cytoplasm uses the classical pathway through the nuclear pore complex involving nucleoporins, karyopherins and the Ran GTP-GDP cycle (Moy and Silver, 1999).
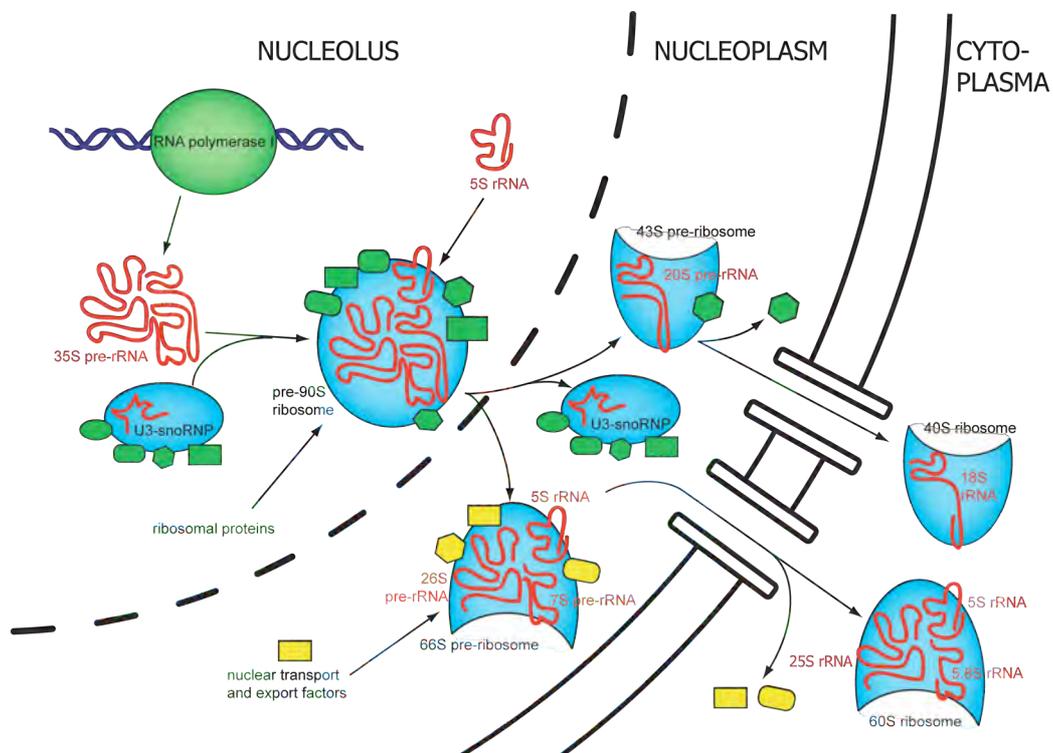
Figure 5 | Ribosome biogenesis in *S. cerevisiae*. Figure was prepared based on Tschochner & Hurt, Trends Cell Biol 2003. Figure greatly simplifies the situation to show the basic principle.

I.8 | Structural studies on eukaryotic RNA polymerases

To date most progress in structural studies was achieved for RNA polymerase II, culminating in the atomic structure of the 10-subunit core enzyme (Cramer et al., 2001) and the complete 12-subunit enzyme structure (Armache et al., 2005). Based on these groundbreaking structures functional DNA-RNA hybrid complexes could be obtained, shedding light onto the transcription mechanism and nucleotide incorporation (Kettenberger et al., 2004; Wang et al., 2006; Westover et al., 2004). Even DNA damages could be co-crystallized with the 12-subunit Pol II, leading to insights into damage recognition and lesion bypass by Pol II (Brueckner et al., 2007; Damsma et al., 2007).

For Pol III, there is no crystal structure available, yet. However, the 17-subunit enzyme has been solved by cryo-EM to 17 Å (Fernandez-Tornero et al., 2007) and the Pol III-specific subcomplex C17/25 has been solved by X-ray crystallography (Jasiak et al., 2006) to 3.2 Å resolution. In addition, a homology model for 11 subunits of the enzyme could be constructed, which demonstrated that, despite sequence homology of about 40% across all subunits, 80% of the fold seems to be conserved between Pol II and Pol III.

For Pol I, the overall shape and dimensions were first revealed by electron microscopy of 2-dimensional crystals (Schultz et al., 1993). Subsequent cryo-EM at 34 Å resolution visualized a stalk density containing the Pol I subcomplex A14/43 and densities for the Pol I-specific subunits A49 and A34.5 over the central cleft (Bischler et al., 2002; Peyroche et al., 2002). Later EM analysis with negatively stained specimen at 22 Å confirmed the stalk, but not the location of A49 and A34.5 (De Carlo et al., 2003).

I.9 | Aim of this study

As structural information to atomic resolution is limited to Pol II (Chapter I.8) and sequence identity between Pol I and Pol II is only 30% (this study), the aim of this work was to solve the structure of RNA polymerase I from *Saccharomyces cerevisiae* to atomic resolution by means of X-ray crystallography.

The structure of this huge 600 kDa macromolecular complex would enable us to possibly explain promoter specificities between the eukaryotic RNA polymerases, to understand rDNA transcription and its regulation in atomic detail and to unravel evolutionary differences between the polymerase systems. Additionally, the two Pol I-specific subunits A49 and A34.5 and the distantly related subcomplex A14/A43 were expected to provide insights into Pol I transcription, which could in return broaden our knowledge about the Pol II system. Structural information on Pol II was expected to facilitate this enormous task, especially for phasing crystallographic data.

II.1 | Large-scale purification of RNA polymerase I

Prior to crystallization, a large-scale purification protocol for Pol I had to be developed. This was based on a initially collaborative effort together with Jochen Gerber from the group of Herbert Tschochner at the University of Regensburg, Germany.

Yeast fermentation was carried out until late-log phase (OD_{600} ~ 5), but later on during this thesis it was realized that even higher OD_{600} values up to ~ 9 did not make any difference in crystallizability of the protein sample. The strain used for purifying Pol I, GPY2, contained only a few genetic modifications compared to wild-type yeast. The genomic copy of the essential subunit A43 was knocked out and placed on a yeast plasmid for introducing a hemagglutinin (HA)- and hexahistidine-tag. The engineered strain grew like wild-type yeast with a doubling time of 2 – 2.5 hours (Fig. 6). Running a 200 L fermenter yielded typically 1.7 – 2.8 kg of yeast pellet that could be used for up to 6 Pol I purifications according to the protocol described here.

GPY2 growth curve 17.12.2003
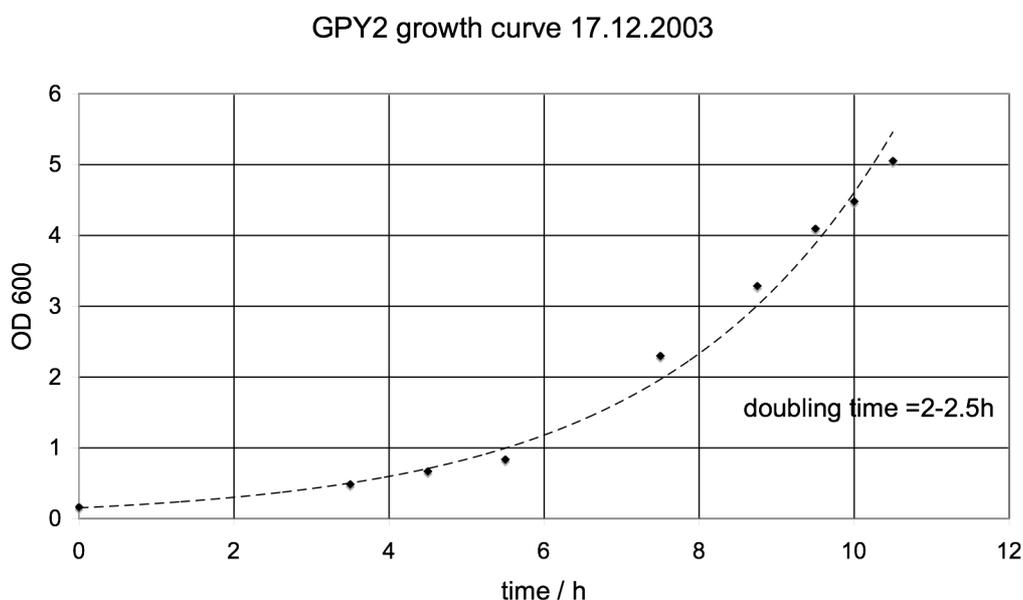
Figure 6 | Growth curve of GPY2 in the small 20 L fermenter

Developing a purification protocol for Pol I proved to be a project posing many challenges: First, the purification protocol from our collaborators in Regensburg made use of the detergent NP-40, which was incompatible with crystallization trials. Second, up-scaling this initial protocol introduced lots of problems concerning reproducibility. Third, every purification step preceding the anion exchange column had to be assessed by western blotting, which made optimization very time consuming. Taken together, obtaining enough crystallization-quality material from endogenous expression was very difficult and remained the biggest problem throughout the whole project (Chapter III.1).

For cell lysis BeadBeaters[TM] were superior to any other technique tested. Judging from cell debris versus non-broken cells, bead-beating was at least 90% effective. The salt concentration before cell lysis was adjusted to 400 mM ammonium sulfate to prevent protein aggregation. Cell debris and non-lysed cells were removed by centrifugation. Lipids and chromatin were removed by an ultracentrifugation step at 100,000 x g using two swinging bucking rotors (~ 160 mL in each rotor). Lipids above the aqueous supernatant were aspired using a vacuum gadget. Care was taken in pooling the supernatant to prevent inclusion of DNA and chromatin, which formed a huge pellet after this ultracentrifugation step.

The clear whole cell extract was dialyzed over night at 4 °C against low salt buffer (Milkereit et al., 1997; Tschochner, 1996). During this step, RNA polymerase I precipitated, whereas Pol II and Pol III stayed in the supernatant. The dialysed sample was ultra-centrifuged at 30,000 x g. By resuspending the pellet in reduced volume, Pol I could be incubated with Nickel resin in just 50 mL solution. For reasons of better reproducibility and higher protein yield, the Nickel resin was distributed between 2-4 smaller columns. Pol I was allowed to bind to the Nickel resin for 4 h at 4 °C in high salt buffer to prevent DNA and proteins from unspecific binding. Optimization of this affinity step was very difficult, as Pol I was very weakly bound to the resin with a substantial amount flowing through the column or sticking irreversibly to the column.

Pol I was eluted with 100 mM imidazole and was loaded onto an anion-exchange column, applying the gradient shown in Fig. 7 and running the column at least twice to increase protein yield.
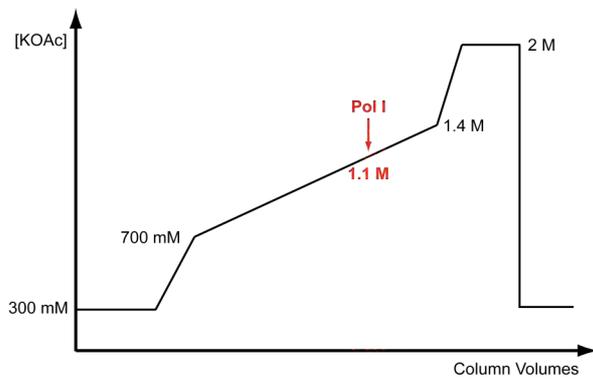
Figure 7 | KOAc-gradient used for anion-exchange chromatography. (first step to 700 mM KOAc not always performed)

As the theoretical pI value for Pol I is 6.25 it was expeted to bind to the anion exchange column. It eluted at approximately 1.1 M KOAc as the protein complex that was free of DNA. Fig. 8 shows an example of a MonoQ run. Pol I is still quite impure after this step but, nevertheless, this is the first step where it can be recognized on SDS-PAGE without western blotting.
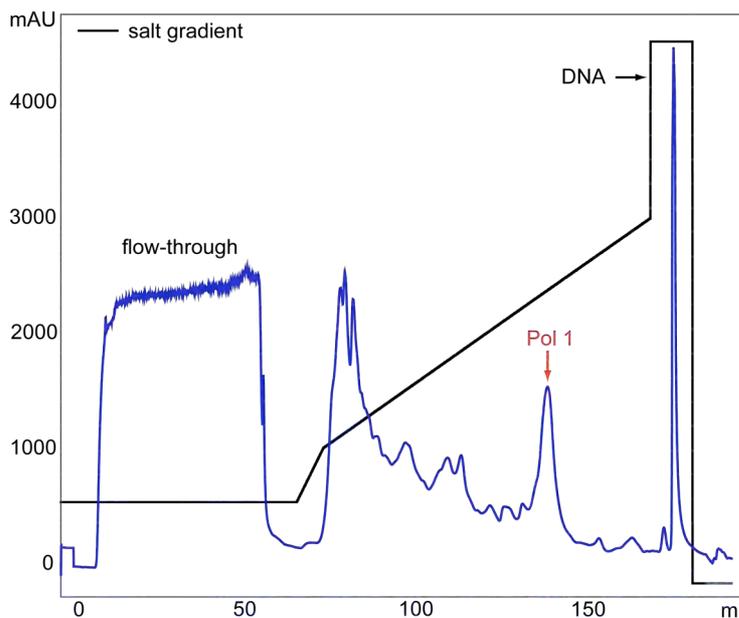
Figure 8 | Anion-exchange chromatography

The MonoQ peaks were pooled and diluted 5.5 times before they were loaded onto a small MonoS column (1 mL bed volume), using the same pH and slightly less salt as for the anion-exchange step. Attempts to use bigger columns failed. The elution gradient is shown in Fig. 9.

Figure 9 | Gradient used for cation-exchange chromatography
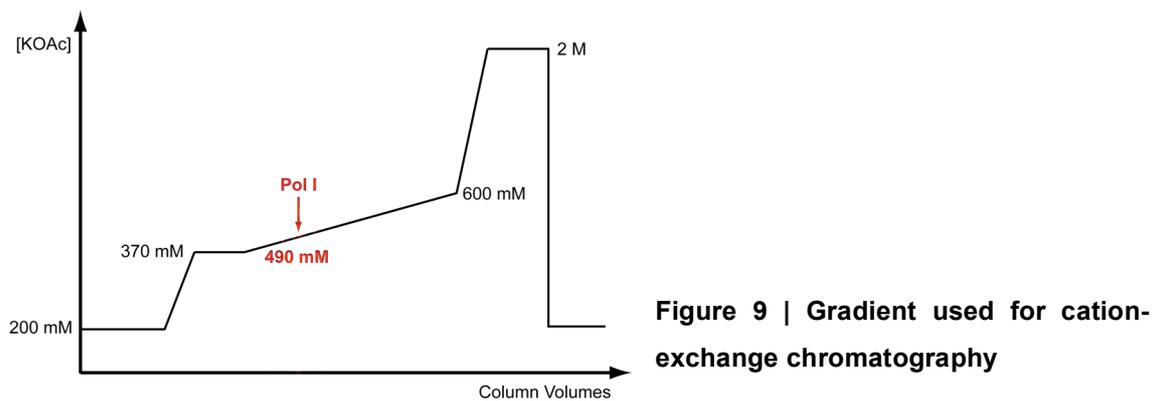
Absolutely pure Pol I eluted in a sharp peak at 490 mM KOAc, being the protein that eluted last from the column (Fig. 10).
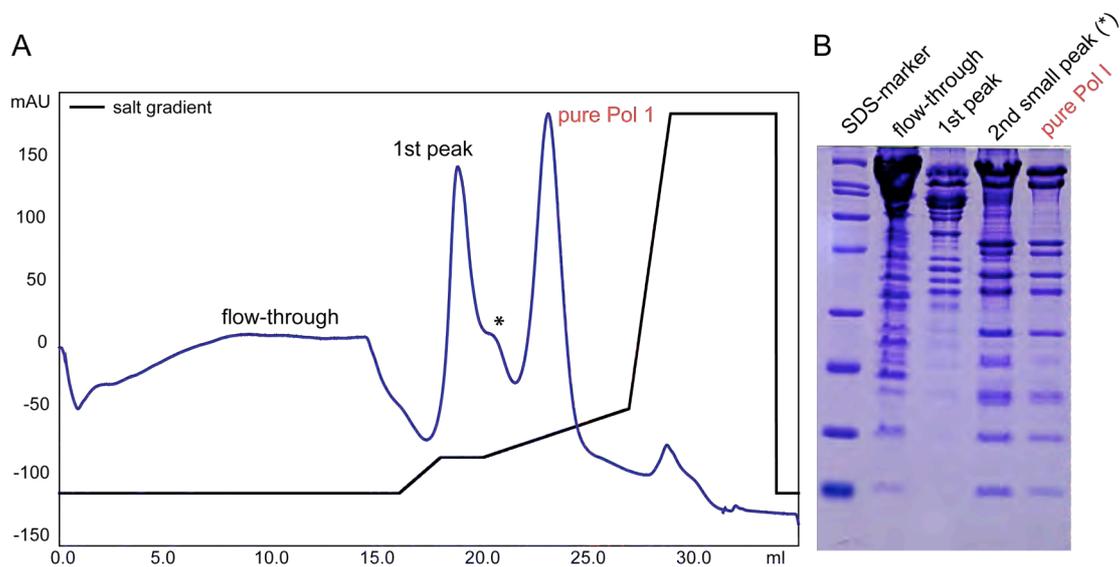
Figure 10 | Cation-exchange chromatography. (A) Chromatogram. **(B)** SDS-PAGE of selected fractions. Flow-through was TCA precipitated, SDS marker with bands of 200, 116, 97, 66, 45, 31, 21.5, 14.4 and 6.5 kDa.

Peak fractions were concentrated to 500 μL and applied to a pre-equilibrated size-exclusion column (Fig. 11). Pol I eluted at 11.9 mL, was monodisperse according to static light scattering, and was subsequently concentrated to 5.5 mg/mL for crystallization. The average yield of the purification ranged from 0.4 to 0.8 mg of pure Pol I.
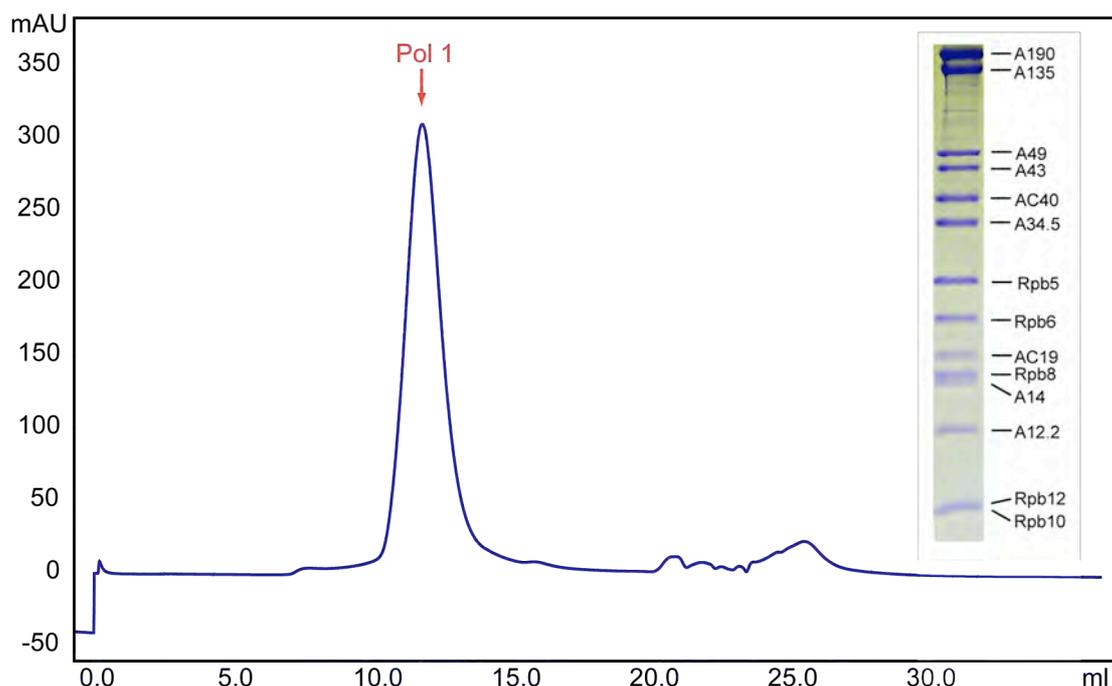
Figure 11 | Size-exclusion chromatography (and SDS-PAGE of pure Pol I)

II.2 | Crystallization of Pol I

Prior to having optimized the purification protocol, Pol I could be crystallized using a hand-made crystallization screen composed of known crystallization conditions for Pol II and a rather impure Pol I sample. Tiny crystals appeared in two conditions: The first contained 300 mM ammonium-sodium tartrate, 100 mM KSCN, 100 mM HEPES, pH 7.5, 12.5-14.5% PEG-6000 and 5mM DTT, the second contained 390 mM ammonium-sodium phosphate, pH 6.0, 50 mM dioxane, 14-15% PEG-6000 and 5 mM DTT (Fig. 12). Crystal size could only be

18

improved for the tartrate-KSCN based condition by using a protein: precipitant drop ratio of 2:1. Crystals diffracted to about 5.5 Å resolution and could be processed using DENZO from the HKL package (Otwinowski and Minor, 1997), resulting in a complete dataset in space group $C222_1$ with unit cell axis of a = 222.4 Å, b = 395.3 Å and c = 282.0 Å. Data could be phased using PHASER (McCoy et al., 2005), but unfortunately no additional density separate of the 12-subunit Pol II was visible. However, the unit cell dimensions, the space group and the crystal shape were so similar to Pol II crystals that we suspected that these crystals contained Pol II rather then Pol I, which would perfectly explain the lack of additional density. Due to the poor purity of the initial Pol I preparations, it might well be that a small Pol II impurity crystallized instead of Pol I. Although this speculation was not unambiguously confirmed, gradually improving the purification protocol led to a complete loss of the initial crystals, supporting the argument that the crystals were indeed Pol II crystals.
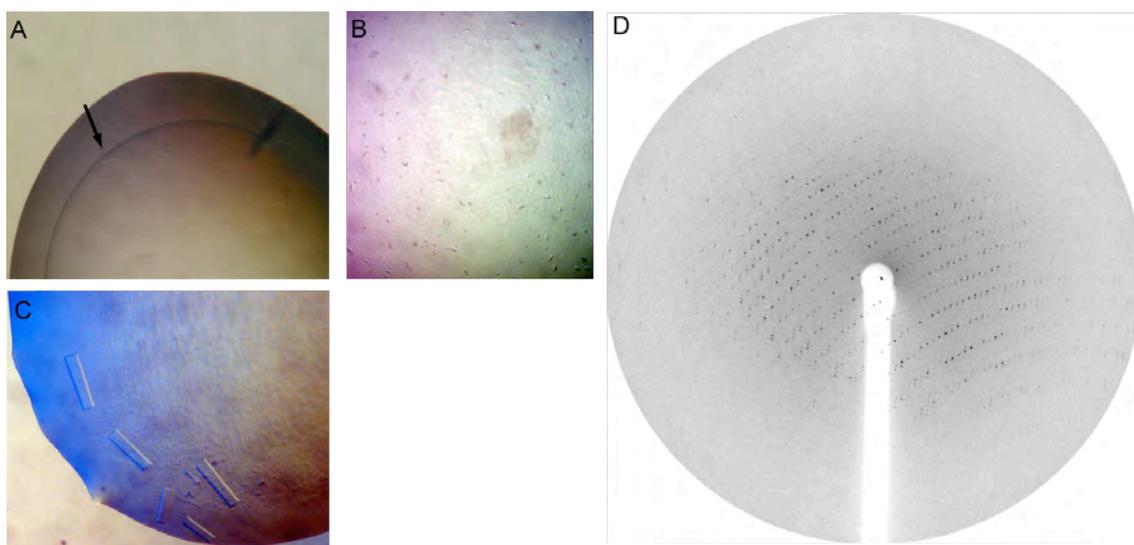
Figure 12 | Initial Pol I crystals. (A) First Pol I (?) crystals seen in a tartrate-KSCN droplet; 12/2006. **(B)** First crystals in ammonium-sodium phosphate. **(C)** Optimized tartrate-KSCN crystals. **(D)** Diffraction pattern of ck26, recorded at SLS, May 2005.

Improving the Pol I purification led to huge problems with crystallization. After introducing the cation-exchange chromatography step to increase the purity of Pol I, crystals could no longer be obtained. Huge efforts in screening and optimizing the crystallization process resulted in 3% MPD as potent additive in inducing crystallization. Additionally, potassium thiocyanate had to be left out and ammonium-sodium tartrate was replaced by di-ammonium tartrate. After optimization, the reservoir solution contained 300 mM di-ammonium tartrate, 100 mM HEPES, pH 7.5, 3% MPD, 10% PEG-6000 and 5 mM DTT or 3 mM TCEP. Even after having tried different protein : reservoir ratios, different temperatures, numerous oils and many other variables, crystals never got bigger than 100-150 μm in their largest dimension.

At this point, streak-seeding (Bergfors, 2003) was the best solution to the size problem (at least in one dimension, Fig. 13A-C). Using cat-whiskers, Pol I crystallization could be triggered in pre-equilibrated drops (equilibration time 3-4 hours), using precipitant solution with a reduced PEG-6000 concentration of 9% (initially 10%) and a protein concentration of 2 to 4 mg/mL (initially 5.5 mg/mL) (Chapter III.3.2). As seeding was very sensitive to the amount of nuclei introduced, reproducibility remained a serious problem. Crystal size seemed to be affected by a myriad of factors like the initial cell material, protein purification, equilibration time, seeds' freshness, the cat-whisker used etc. Increasing the precipitant concentration by just 0.5% resulted in no crystals or very small ones. Despite this sensitivity, single crystals reached a maximum size of 500 μm x 70 μm x 10 μm (Fig. 13D).
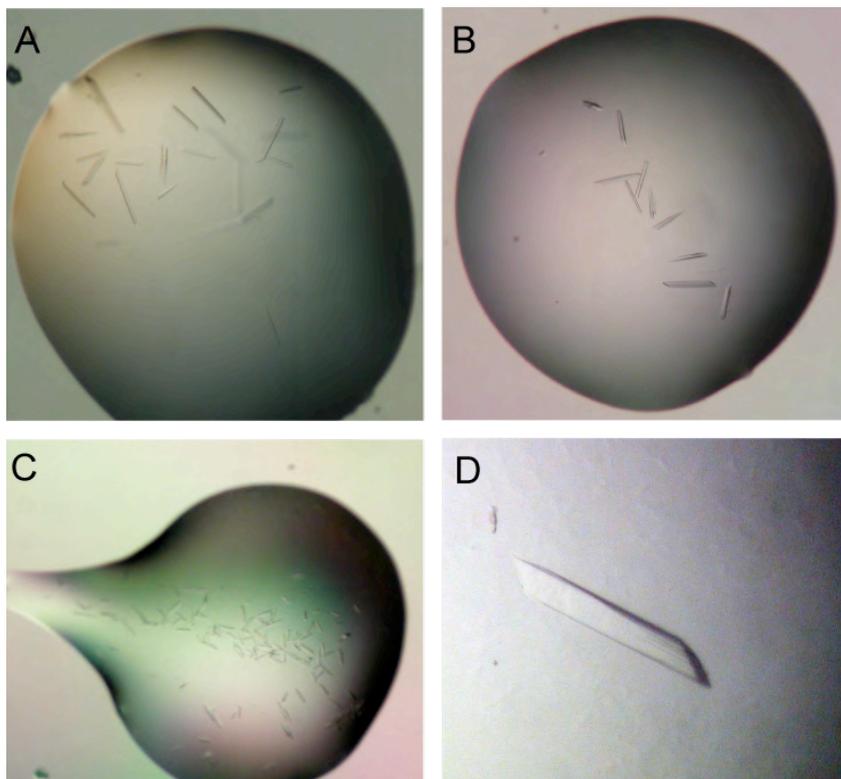
Figure 13 | Improved Pol I crystals.
(A) and **(B)** Streak-seeding examples.
(C) Nucleation after seeding too high.
(D) Crystal > 400 μm (in one dimension).

II.3 | Cryo-crystallography and heavy atom derivatization

After initial non-satisfying trials with glycerol, sucrose and L-(+)-2,3-butandiol, 22% PEG-400 was used as cryo-agent. Native crystals grown in seeded drops never showed ordered diffraction. Heavy atom derivatization, persecuted for gaining experimental phase information, proved to be essential for introducing order in the crystal lattice (Fig. 14).

Soaking of crystals in the final cryo-solution in the presence of a W_{18} cluster for ~ 2 days resulted in optimal diffraction. See Table 2 for a summary of all heavy atom clusters and manipulation techniques tried and their effects on Pol I crystal diffraction quality. Heavy atoms were always added after having transferred the crystals to the final cryo-solution.
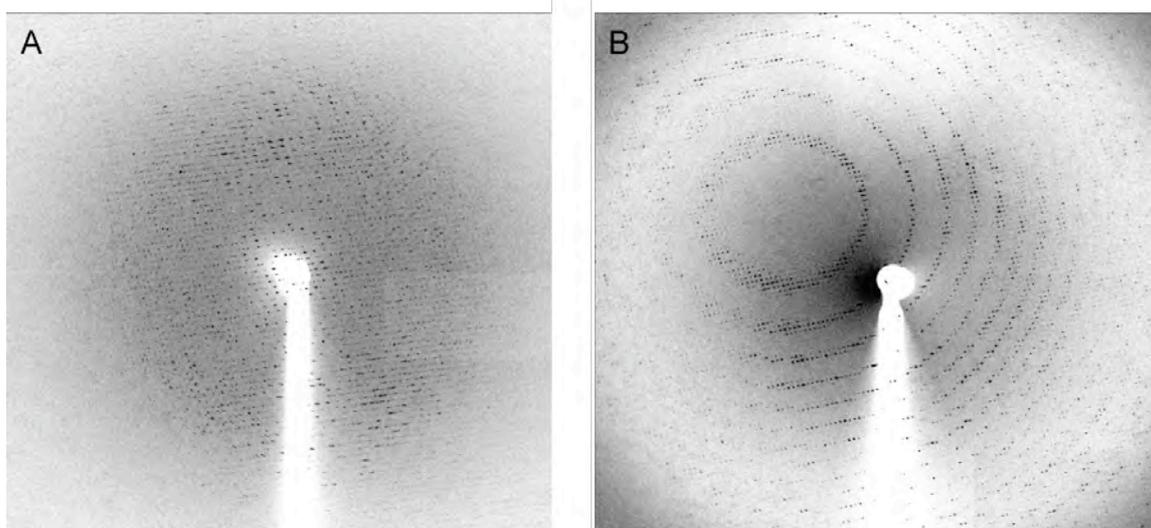
Figure 14 | Improvement of diffraction quality upon heavy atom treatment. (A) Native crystal without W_{18} treatment. **(B)** Crystal soaked for 2 days in W_{18} solution (Note: Crystals are not the same).

Table 2 | Summary of heavy atom derivatization protocols

Heavy atom	Soaking protocol	Additional manipulation	Result
-		Cross-linking using glutaraldehyde	No improvement versus native crystals.
W_{18} cluster	Overnight at 4 °C or 20 °C	-	Still diffraction like native crystals.
W_{18} cluster	**44 h at 4 °C**	**-**	**Best data quality.**
W_{18} cluster	44 h at 4 °C	Cross-linking using glutaraldehyde (optional backsoaking)	Very high resolution for first frames, but <u>too</u> much radiation damage to collect full dataset.
W_{18} cluster	44 h at 4 °C	Dehydration by increasing PEG-6000 to 20%	Diffraction quality much poorer.
W_{18} cluster	66 h at 4 °C	-	Diffraction quality diminished (compared to 44 h soaking).

W_{30} cluster	44 h at 4 °C	(Optional : cross-linking using glutaraldehyde)	Heavy atom treatment apparently destroys diffraction.
$Ta_6Br_{12}^{2+}$ cluster	1-3 h at 20 °C	-	No ordered diffraction beyond 6-7 Å.
$Ta_6Br_{12}^{2+}$ cluster	Overnight	(Optional : cross-linking using glutaraldehyde)	Very poor diffraction, most crystals dead.
Ir_3 cluster	2-3 h at 20 °C	-	Diffraction quality OK, derivatization not optimized.
$W_6Br_{12}^{2+}$	1-2 h or overnight	-	No ordered diffraction beyond 6-7 Å, derivatization not optimized.
2,4,6-Trisaceto-(3-acetamino) mercuritoluol	½ - 3 h at 20 °C	-	No diffraction at all.

II.4 | Data collection and processing

All datasets were collected at the SLS (Swiss Light Source) on beamlines X06SA and X10SA. The presence of heavy atoms in the crystals was demonstrated by performing X-ray absorption scans for every heavy atom species used (Fig. 15). Due to severe problems with radiation damage, deriving a perfect strategy for data collection was crucial (Chapter III.4).
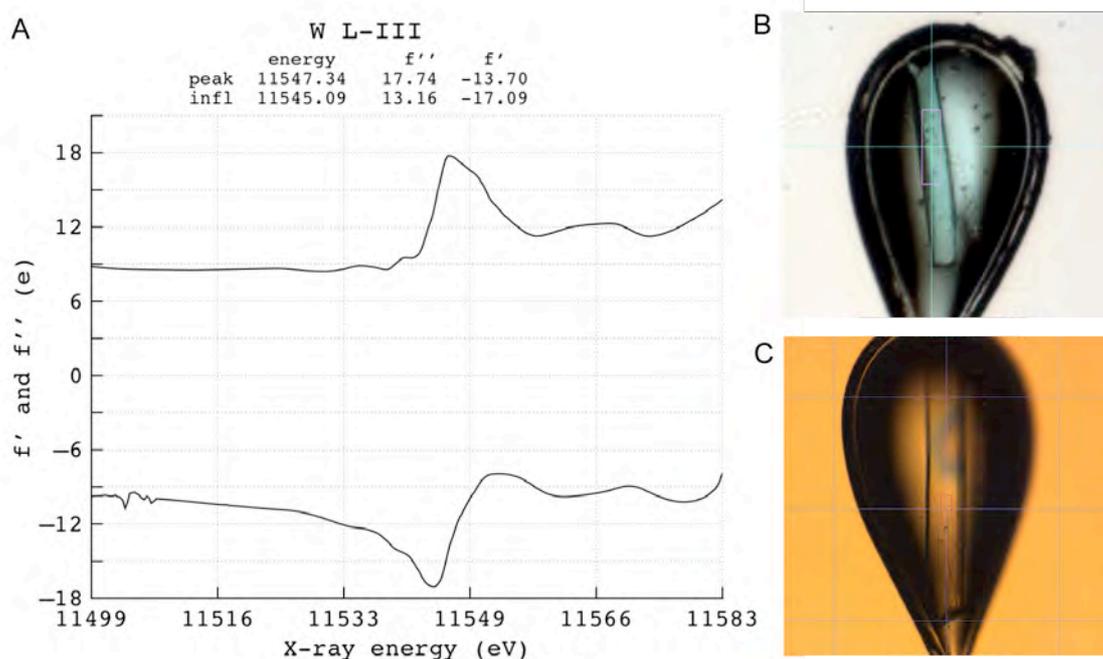
Figure 15 | Data collection of crystals soaked with heavy atoms. (A) Experimentally derived f' and f'' values for W_{18} at the L-III edge. **(B)** $Ta_6Br_{12}^{2+}$ soaked crystal. **(C)** W_{18} soaked crystal.

Processing the data with DENZO and SCALEPACK (Otwinowski and Minor, 1997) resulted in reasonable statistics to 4.8 Å (Chapter III.5). The unit cell was monoclinic C2, with unit cell dimensions of a = 615 Å, b = 304 Å, c = 253 Å and β = 97.6°. 52,205 reflections of the measured 1,237,240 were rejected (4.2%), the average redundancy being > 5. Striking was the high R-factor in the lowest resolution shell. We speculate that this is due to the anomalous signal of the W_{18} cluster, which is unfortunately not bound to the protein, but diffuses freely in the solvent channels of the crystal (Fig. 16C).
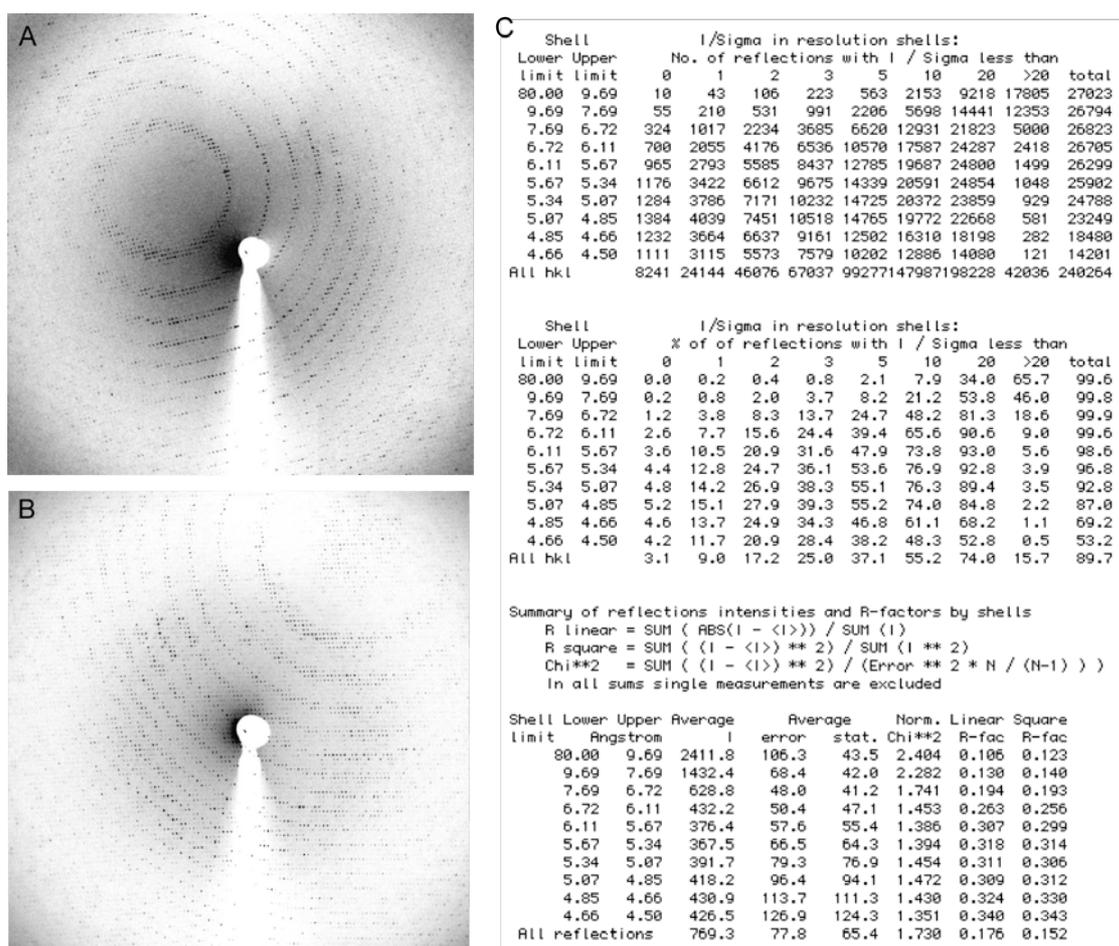
Figure 16 | Diffraction pattern of W$_{18}$ soaked crystals and data statistics. (A) Image of crystal ck284 (44 h W$_{18}$ at 4 °C). **(B)** Image of crystal ck290 (44 h W$_{18}$ at 4 °C). **(C)** Data statistics of ck290 (after SCALEPACK).

Even though heavy atom derivatization and data collection were quite sophisticated for Pol I, the main problem was a lack of reproducible crystals. Microseeding did not produce lots of suitable crystals, it was very difficult to standardize and the crystals got the bigger the fewer grew in one drop. With this small amount of crystals, choosing the right cryo-agent, heavy atom treatment or data collection strategy was very risky.

II.5 | Attempts on solving Pol I by X-ray crystallography

Calculating the Matthews coefficient for Pol I crystals suggested four copies in the asymmetric unit, when assuming the same solvent content as in Pol II crystals (75%). However, after calculating a self-rotation function with the Pol I data, the situation became even more dramatic (Chapter III.5). At κ = 52° a very strong peak could be observed, complemented by 7 equally strong peaks at κ = 180°. This strongly suggested the existence of 7 Pol I molecules in the asymmetric unit related by a 7-fold non-crystallographic symmetry (NCS) axis with 14 2-fold axes perpendicular to the 7-fold (One 2-fold axis every 25°). In standard stereographic projections, these 14 axes result in just 7 axes plotted in one polar coordinate hemisphere (Fig. 17). The 7-fold axis of this NCS ensemble is oriented along c. These findings demonstrate that the asymmetric unit of Pol I crystals contains 600 kDa x 7 = 4.2 MDa of protein (56% solvent). This is not much less than the asymmetric unit of the *E.coli* 70S ribosome, which contains 2 ribosomes (5.6 MDa) in the asymmetric unit (Schuwirth et al., 2005)!
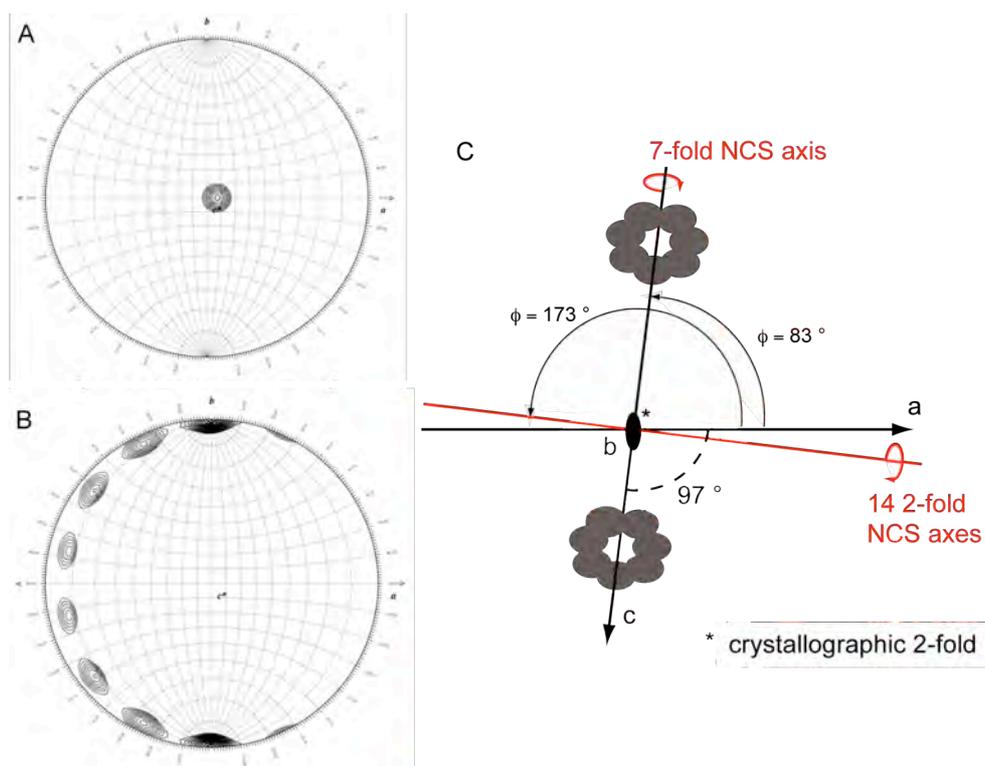
Figure 17 | Non-crystallographic organization of the Pol I asymmetric unit. (A) Stereographic projection of the self-rotation function at κ = 51° (7-fold NCS axis). **(B)** Stereographic projection at κ = 180° (2-fold NCS axes). **(C)** Sketch of the inherent symmetry of the Pol I asymmetric unit. The 14 2-fold axes lie in the ab plane, perpendicular to the page.

Two approaches for solving the Pol I crystal structure were pursued: First, heavy atom derivatization using cluster compounds for deriving experimental phase information. Second, molecular replacement using either models based on available Pol II crystal structures or structural information gained by cryo-electron microscopy.

A fortunate side effect of treating crystals with heavy atoms was the discovery that W_{18} ordered Pol I crystals in a quite unique way and eventually enabled data collection to 4.8 Å. However, binding heavy atom clusters to Pol I was fairly unsuccessful. Neither classical anomalous difference Patterson maps nor software like SOLVE (Terwilliger and Berendzen, 1999), SHELXD (Schneider and Sheldrick, 2002) or BnP (Furey and Swaminathan, 1997; Weeks et al., 1994) revealed unambiguous heavy atom sites (Chapters III.6.1 and IV.3).

The anomalous signal is greatly reduced at about 30-40 Å resolution (high R_{merge} in low resolution shell, Fig. 16C), one possible explanation for that being that W_{18} is not stably bound but rather diffused through the crystals' solvent channels. How the cluster would specifically order the crystal without specifically binding to the protein will remain a mystery.

Molecular replacement was initially carried out using models based on the Pol II structure, using either the complete 12-subunit enzyme (Armache et al., 2005) or Pol II bound to TFIIS (Kettenberger et al., 2003) as search models. Sequence elements of Pol II that were apparently divergent or missing in Pol I, were omitted according to Chapter IV.7. However, using these models in molecular replacement never resulted in groups of seven rotational solutions representing the expected 7-fold axis in the asymmetric unit. Using the known self-rotation information as restraint in the program MOLREP did at least result in some rotational solutions that were related by a 7-fold axis, but a full set of 7 clustered solutions could never be obtained using crystallographic models (Fig. 18A).

A

#Sol_		theta	phi	chi	alpha	beta	gamma	Rf	Rf/sigma	
Sol_RF	1	137.53	-30.76	117.93	188.45	70.70	69.96	0.1740E+07	6.31	*
Sol_RF	2	120.25	-8.01	79.84	239.13	67.33	75.16	0.1581E+07	5.73	*
Sol_RF	3	124.60	-11.08	77.97	234.24	62.37	76.40	0.1459E+07	5.29	
Sol_RF	4	84.59	16.23	60.85	289.40	60.55	76.94	0.1453E+07	5.27	*
Sol_RF	5	122.64	-10.69	79.28	235.23	64.99	76.62	0.1446E+07	5.24	
Sol_RF	6	48.90	47.62	77.77	345.56	56.46	70.31	0.1390E+07	5.04	*
Sol_RF	7	144.49	-51.52	160.39	140.47	69.84	63.51	0.1360E+07	4.93	*
Sol_RF	8	72.41	15.59	91.73	302.89	86.34	91.71	0.1296E+07	4.70	
Sol_RF	9	46.48	54.54	124.65	17.25	79.90	88.17	0.1200E+07	4.35	
Sol_RF	10	136.22	-20.78	113.88	201.30	70.85	62.86	0.1195E+07	4.33	
Sol_RF	11	53.67	44.17	110.53	354.69	82.91	86.34	0.1187E+07	4.30	
Sol_RF	12	135.86	98.77	163.24	290.37	87.09	272.83	0.1185E+07	4.30	
Sol_RF	13	141.98	-29.67	120.81	186.12	64.77	65.46	0.1134E+07	4.11	
Sol_RF	14	47.48	45.11	74.33	342.24	52.88	72.02	0.1120E+07	4.06	
Sol_RF	15	148.37	-51.94	166.68	135.87	62.78	59.74	0.1107E+07	4.02	
Sol_RF	16	120.09	-15.21	95.56	225.87	79.70	76.29	0.1107E+07	4.01	
Sol_RF	17	43.05	83.54	160.16	70.07	84.51	83.00	0.1100E+07	3.99	
Sol_RF	18	124.02	-18.58	99.53	217.94	78.50	75.10	0.1085E+07	3.93	
Sol_RF	19	35.82	79.14	114.43	40.95	57.69	62.67	0.1078E+07	3.91	*
Sol_RF	20	134.74	-77.47	137.31	131.57	82.84	106.51	0.1076E+07	3.90	
Sol_RF	21	33.34	114.26	156.42	100.23	65.09	51.71	0.1066E+07	3.86	
Sol_RF	22	84.83	164.78	85.34	79.53	84.91	289.97	0.1061E+07	3.85	
Sol_RF	23	32.57	106.10	156.73	92.37	63.65	60.17	0.1055E+07	3.83	*

B

#Sol_		theta	phi	chi	alpha	beta	gamma	Rf	Rf/sigma	
Sol_RF	1	129.91	-157.44	109.49	78.33	77.56	205.21	0.5679E+07	4.10	*
Sol_RF	2	55.32	-87.73	112.08	222.47	86.01	217.92	0.5667E+07	4.09	*
Sol_RF	3	141.07	173.75	141.71	17.80	72.82	210.30	0.5662E+07	4.09	*
Sol_RF	4	42.36	-63.53	144.88	273.28	79.94	220.35	0.5656E+07	4.09	*
Sol_RF	5	107.41	-132.40	89.34	121.13	84.26	205.93	0.5653E+07	4.08	*
Sol_RF	6	143.15	143.78	178.35	324.81	73.69	217.25	0.5646E+07	4.08	*
Sol_RF	7	78.26	-109.85	90.33	171.71	87.94	211.41	0.5634E+07	4.07	*
Sol_RF	8	71.37	78.34	96.53	8.05	90.00	31.36	0.5510E+07	3.98	
Sol_RF	9	93.03	49.77	90.16	316.73	90.00	37.20	0.5242E+07	3.79	
Sol_RF	10	56.29	106.51	116.43	58.35	90.00	25.33	0.4916E+07	3.55	
Sol_RF	11	37.40	-25.12	126.77	302.64	65.77	172.88	0.4735E+07	3.42	
Sol_RF	12	146.70	-144.76	154.40	50.44	64.73	159.96	0.4707E+07	3.40	
Sol_RF	13	116.04	-94.96	87.34	152.31	76.69	162.22	0.4706E+07	3.40	
Sol_RF	14	31.25	4.63	165.00	355.88	61.98	166.62	0.4702E+07	3.40	
Sol_RF	15	83.11	-73.77	77.65	201.75	76.99	169.28	0.4701E+07	3.40	
Sol_RF	16	137.01	-117.91	117.23	101.91	71.20	157.74	0.4700E+07	3.40	
Sol_RF	17	53.54	-51.60	94.09	250.95	72.12	174.15	0.4691E+07	3.39	
Sol_RF	18	70.07	54.98	75.18	339.68	69.98	49.72	0.4405E+07	3.18	
Sol_RF	19	123.51	1.22	103.22	236.36	81.62	53.91	0.4402E+07	3.18	
Sol_RF	20	101.69	26.70	76.22	287.66	74.37	54.27	0.4397E+07	3.18	

Figure 18 | Solutions of the Molecular replacement rotation function. (A) Using a crystallographic model and ck209 data. **(B)** Using an EM map (val067f_300.map) and ck290 data. Red asterisks highlight rotational solutions related by a 7-fold axis.

After having solved the Pol I structure by cryo-electron microscopy to 11.9 Å (Chapters II.6 and III.7), molecular replacement with MOLREP was tried again, using the EM density map as search model, which resulted for the first time in clusters of 7 rotation solutions that were related to each other by a 7-fold axis (Fig. 18B and Chapter IV.4). However, solving the translation function for these pairs of seven solutions was not possible.

In ongoing research we are trying to restrain the translational search by constructing search models that already consist of seven Pol I molecules (Fig. 19). These 7-mer rings are constructed by applying each rotation onto the

search model, in this case an EM density map. The resulting seven rotated molecules are easily combined to form 'common sense' rings, which still obey all restraints imposed by NCS and the unit cell dimensions. We will hopefully one day see phased Pol I density after finally having elucidated this complicated asymmetric unit or by finding a different crystallization condition with simpler non-crystallographic symmetry. However, if one could solve this complicated asymmetric unit, we could exploit the enormous power of phase improvement by 7-fold NCS-averaging.
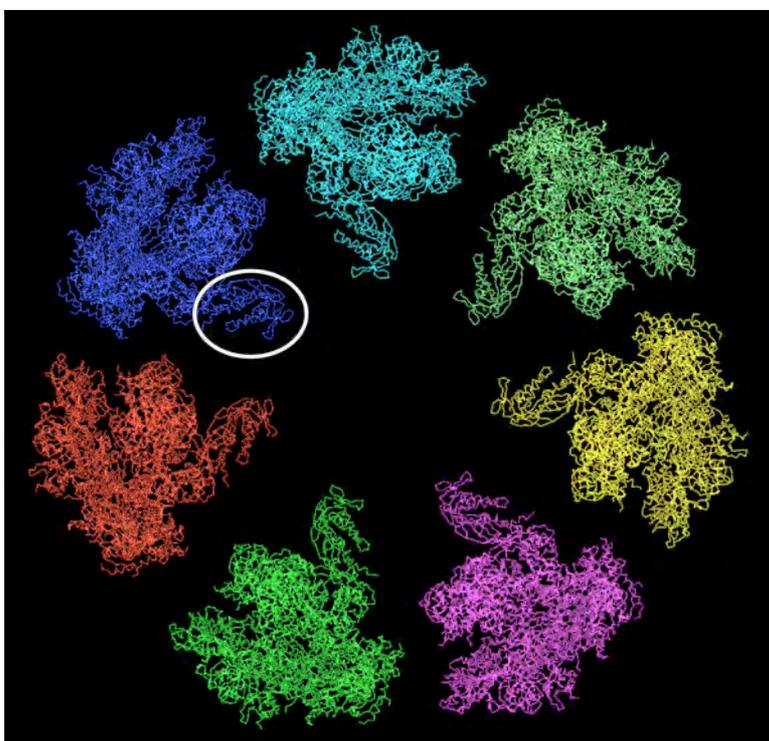
Figure 19 | Potential 7-mer ring. Molecules were rotated by applying the 7 rotational solutions onto a Pol II model fitted into the EM density. Molecules were shifted to form a 7-mer ring – in this example we chose to orient A14/43 towards the center of the ring. One A14/43 complex is marked with a white circle.

II.6 | Cryo-EM structure of Pol I at 12 Å resolution

The to-date unsuccessful crystallographic structure solution of Pol I prompted a collaboration with the group of Roland Beckmann from the FU Berlin (now Gene Center, Munich), using electron microscopy instead of X-ray crystallography for structure solution. Pol I was purified as described for crystallization (Chapter III.1) and kept on ice until further usage. The optimal protein concentration for cryo-EM, 0.1 mg/mL, was determined by electron microscopy using negative stain (Chapter III.7.1 and Fig. 20A). Particles did not form aggregates and showed high particle density. Even under cryo-conditions (100K at liquid nitrogen temperatures) particles could be easily identified and apparently behaved nicely during vitrification (Fig. 20B). Cryo-EM reconstruction of Pol I (Chapter III.7.4) with 46,056 particles led to a map at 11.9 Å resolution (Fig. 20C, D).
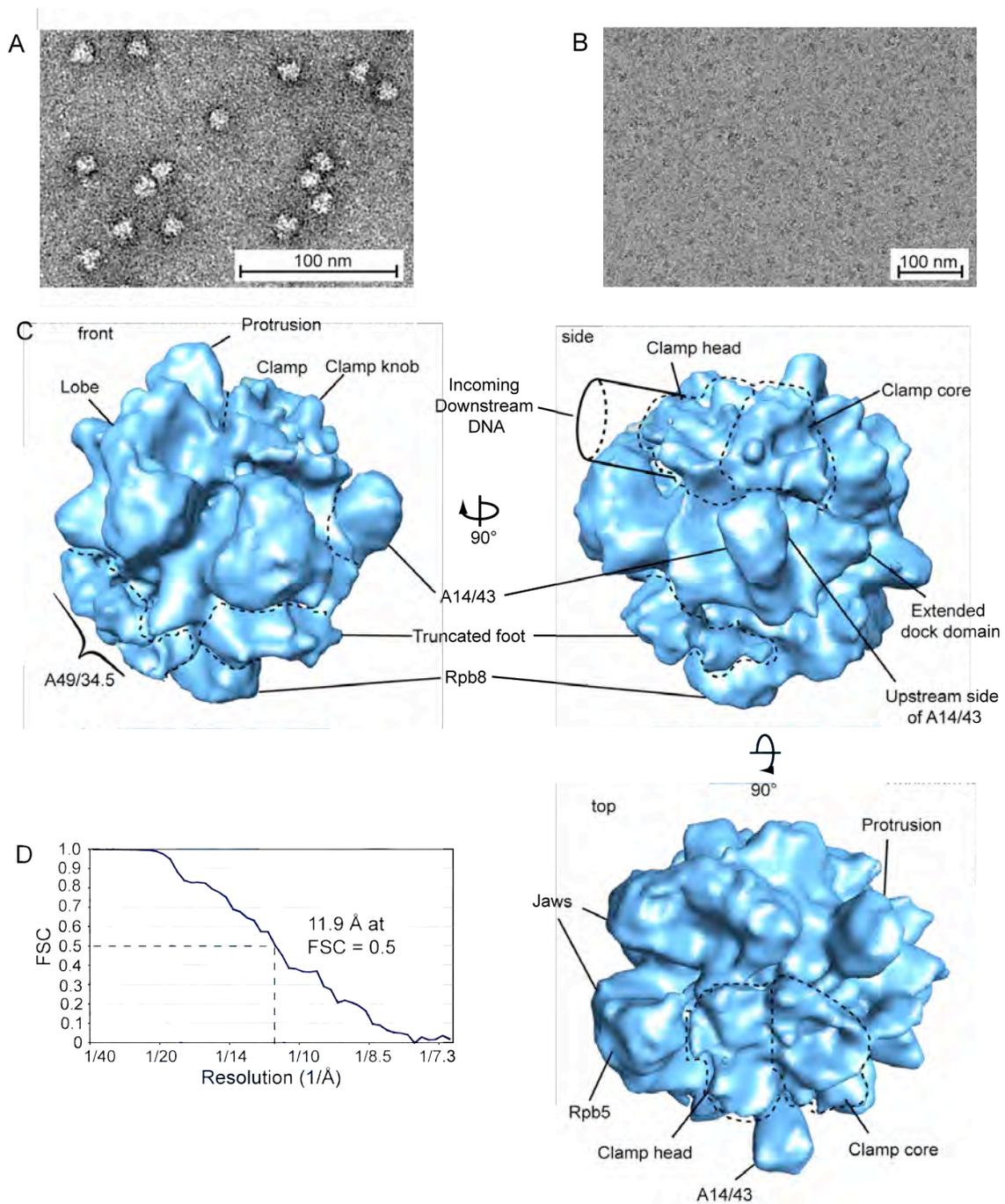
Figure 20 | EM reconstruction of Pol I. (A) Negatively stained Pol I. **(B)** Pol I variant Δ49/34.5 under cryo conditions. **(C)** Cryo-EM reconstruction of Pol I. Views and structural regions are named according to the Pol II structure (Cramer et al., 2001). **(D)** Fourier shell correlation (FSC) function plot. Based on a cutoff value of FSC=0.5, the resolution is 11.9 Å.

Interpretation of the EM map was achieved by first placing the crystal structure of the 10-subunit core into the EM map as a rigid body by fitting the five common subunits, which were known to occupy similar positions on the polymerases' surfaces (Jasiak et al., 2006). A perfect fit of the common subunits confirmed the high quality of the map (Fig. 21).
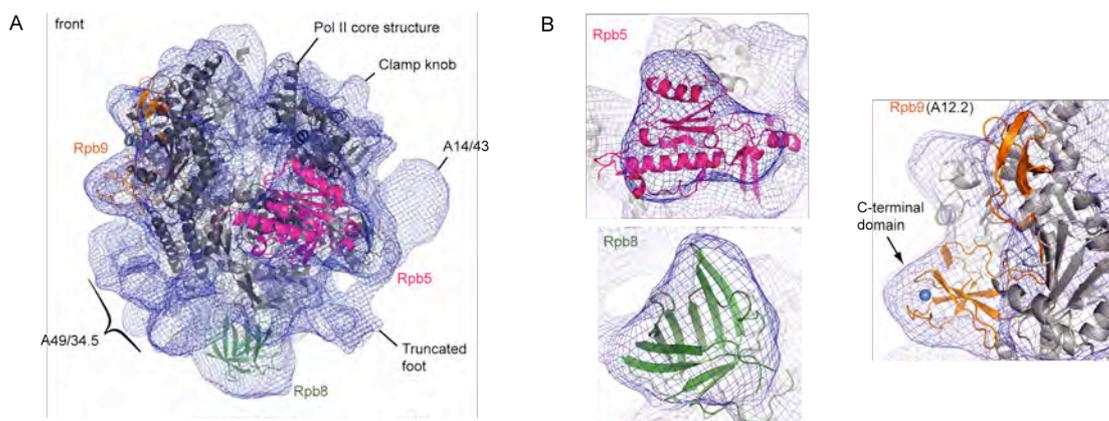
Figure 21 | Placement of 10-subunit Pol II into EM density. (A) Placement of the Pol II 10-subunit core structure (Armache et al., 2005) (grey) into the EM density (blue). The foot was deleted, and subunits Rpb5, Rpb8, and Rpb9 are highlighted in magenta, green, and orange, respectively. The clamp has been fitted as a separate rigid body. **(B)** Fit of the common subunits Rpb5 and Rpb8 to the EM map, and density for the core subunit A12.2 (the Pol II homolog Rpb9 is shown as a ribbon model).

Many regions of the homologous subunits fitted equally well, but strong deviations were also observed, in particular at the polymerase clamp and foot (Cramer et al., 2001) (Fig. 24). The clamp had swung inwards, entirely closing off the cleft (Fig. 22A). This closed clamp conformation is the predominant state of the enzyme under the used experimental conditions, as the sample contained polymerases with many different clamp conformations. In a large fraction of these particles, the clamp apparently adopted a totally closed state, which allowed for refinement of the class I volume to high resolution (Chapter III.7.4). However, several different clamp conformations were apparently present in class II, impairing refinement of the volume to high resolution.

The absence of bias during reference-based alignment could be demonstrated using a Pol II structure lacking the clamp, Rpb4/7 (except for the Rpb7 tip domain) and the foot domain. Already after initial alignment, density for the clamp reappeared and confirmed thereby the validity of our alignment approach. To exclude that the small stalk density was due to just the tip domain of Rpb7 being present, reference-based alignment was carried out again, using a model that contained the complete Rpb4/7. Again, after the first round of refinement, the density for Rpb4/7 was strongly decreased and was lacking at the outer positions (Fig. 22B). This is a clear indication that less density for the Rpb4/7 counterpart A14/43 in Pol I is not due to reference bias but reflects a high mobility of the OB domain of A43 and the absence of an HRDC domain (Chapter II.8).
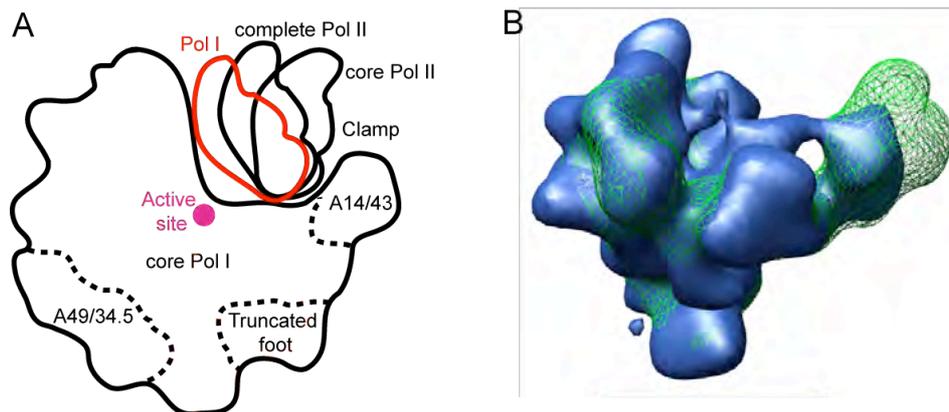
Figure 22 | Clamp flexibility and absence of reference-bias. (A) Schematic representation of the clamp positions in Pol I, the complete 12-subunit Pol II (Armache et al., 2005), and the 10-subunit core Pol II (Cramer et al., 2001). **(B)** Comparison of the reference including Rpb4/7 (green) to the volume obtained after the first round of refinement (blue). Both volumes are filtered at 20 Å.

II.7 | Homology model of the Pol I core explains EM density

To explain differences between the EM map and the Pol II core structure, we constructed a homology model for the Pol I core. Modeling was achieved as for the Pol III core (Jasiak et al., 2006), but was complicated by the weaker sequence conservation between Pol II and Pol I subunits (Fig. 23, Table 3 and Chapter IV.7). We identified regions of conserved fold in cycles of sequence alignment, model construction, detection of incorrect internal contacts, realignment of the erroneous sequence stretches, and the construction of an improved model (Fig. 23, Chapter IV.7). In the Pol I core model, well-conserved regions cluster around the active site, and peripheral regions are divergent (Chapter IV.7). However, some peripheral Pol I domains, such as the jaw and lobe, resemble in shape the Pol II domains, suggesting that their folds are conserved despite divergent sequences. The predicted conservation of Pol II folds is far less in Pol I (60.8% overall, Table 3) than in Pol III (83.4% for a 11-subunit model (Jasiak et al., 2006)).
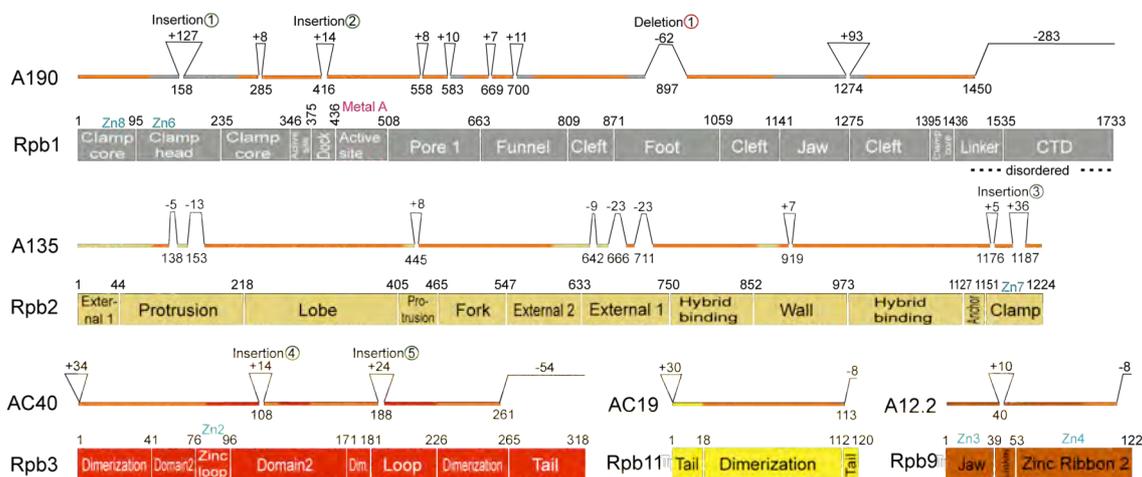
Figure 23 | Homology model of the Pol I core. Pol II structure-guided sequence alignment of the five Pol I subunits with homologs in Pol II (compare Table 3). The domain organization of Pol II subunits Rpb1, Rpb2, Rpb3, Rpb11, and Rpb9 is shown as diagrams (Cramer et al., 2001). Insertions and deletions exceeding five amino acid residues are indicated. Conserved folds are indicated by orange highlighting of the bar above the diagrams.

Table 3 | Sequence and fold conservation between Pol I and Pol II

Polymerase part	Pol I subunit	Pol II subunit	Sequence identity[1] (%)	Conserved Pol II fold[2] (%)
Core	A190	Rpb1	22.3	47.8
	A135	Rpb2	26.0	62.1
	AC40	Rpb3	21.2	53.5
	AC19	Rpb11	17.6	77.5
	A12.2	Rpb9	19.2	35.2
	Rpb5 (ABC27)	Rpb5	100	100
	Rpb6 (ABC23)	Rpb6	100	100
	Rpb8 (ABC14.5)	Rpb8	100	100
	Rpb10 (ABC10b)	Rpb10	100	100
	Rpb12 (ABC10a)	Rpb12	100	100
Subcomplex A14/43	A14	Rpb4	4.5	25.0[3]
	A43	Rpb7	8.0	78.4[3]
Subcomplex A49/34.5	A49	RAP74[4]	7.6	57.2
	A34.5	RAP30[4]	8.3	80.5
Total	-	-	29.5	60.8

[1]Number of amino acid residues in the Pol I subunit that are identical in the corresponding Pol II subunit divided by the total number of residues in the Pol I subunit. For A49/34.5, number of amino acid residues in the TFIIF RAP74/30 dimerization module structure that are identical in the A49/34.5 model divided by the total number of residues in the RAP74/30 heterodimer structure (Gaiser et al., 2000).

[2]Number of amino acid residues in the Pol II core structure that have the same fold in the Pol I homology model divided by the total number of residues in the Pol II subunit.

[3]For A43, number of amino acid residues in the Rpb7 structure that have the same fold in the A43 structure divided by the total number of residues in the Rpb7 structure. For A14, number of amino acid residues in the Rpb4 structure that have the same fold in the A14 structure divided by the number of residues of the tip-associated domain of Rpb4 (residues 1-155, HRDC domain excluded).

[4]Predicted to be partially homologous to the TFIIF subunits RAP74 and RAP30. For details see Chapter III.10.

Inspection of the EM map after placement of the core model confirmed the expected conservation of the active center, including the bridge helix, but also identified many structural features that create a Pol I-specific surface. The clamp shows two insertions near zinc site 7 ("clamp knob"), and an extended, structurally different clamp head (Figs. 20C, 24B). The dock domain shows density for a predicted (Chen and Hahn, 2003) Pol I-specific 14-residue extension (Fig. 24). In AC40, two surface elements differ from Rpb3 (Fig. 24). The foot domain has a divergent sequence, is 62 residues shorter, and has a different shape than in Pol II (Fig. 24). The jaw region contains 93 additional residues (Fig. 23), which are not conserved among fungi, and lack EM density, showing they are mobile. A12.2 occupies the location of the Pol II core subunit Rpb9, and is thus a structural counterpart of Rpb9, not TFIIS (Fig. 21B), even though the C-terminus of A12.2 can be perfectly aligned to TFIIS (Chapter IV.7).
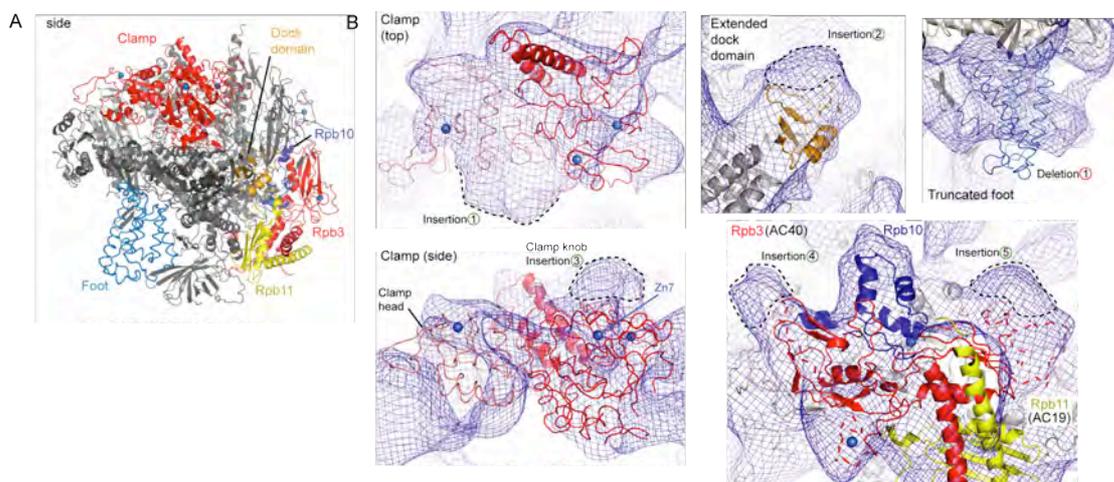
Figure 24 | Pol I specific features of the EM density. (A) View of the core Pol II structure from the side, with domains depicted in (B) highlighted. **(B)** Pol I-specific structural elements. Fitted Pol II elements are shown as ribbon models. Insertions and deletions explaining the EM density are named according to Fig. 23. The clamp head is in light red, the clamp core in red. The dock and foot domains are in beige and blue, respectively, and Rpb3, Rpb10, and Rpb11 are in red, dark blue and in yellow, respectively. Zinc ions are depicted as marine spheres.

II.8 | Crystal structure of A14/43 elucidates Pol I initiation

After assigning EM densities to the Pol I core, a stalk-like density remained at the expected location for A14/43 that was much smaller than the structure of Rpb4/7 (Figs. 20C, 22). Since the weak sequence similarity between A14/43 and Rpb4/7 or C17/25 did not allow for homology modeling, we determined in a long-term project the crystal structure of A14/43. Structure determination of A14/43 will not be part of this thesis, however, information gained from the structure explaining Pol I function will be discussed.

The overall structure of A14/43 resembles its counterparts Rpb4/7 (Armache et al., 2005), C17/25 (Jasiak et al., 2006), and the archaeal RpoF/E , except that A14 lacks the HRDC domain present in all counterparts (Fig. 25). The N-terminal tip domain of A43 shows RMS deviations in Cα atom positions of 2.2-2.5 Å, whereas the C-terminal OB domain is more divergent. A14 forms two helices that pack on the A43 tip domain (Fig. 25).
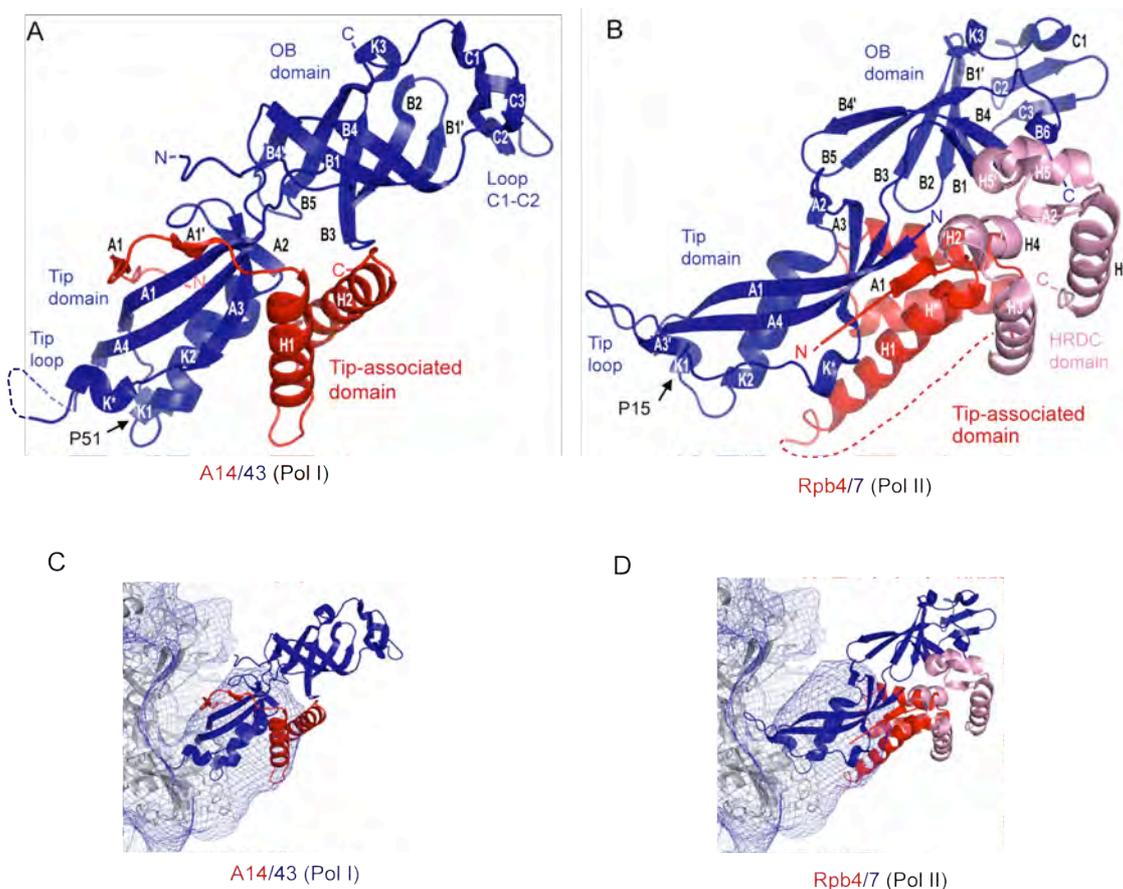
Figure 25. X-ray structure of the A14/43 subcomplex. (A) Structure of yeast A14/43 (this study). A43 is in bue, A14 in red. **(B)** Structure of yeast Rpb4/7 (Armache et al., 2005). Rpb7 is in blue and Rpb4 is in red, with the HRDC domain in light red. **(C)** Fit of the A14/43 structure into the Pol I EM density. **(D)** Fit of the Rpb4/7 structure into Pol I EM density.

In Pol II, the Rpb4/7 complex interacts with the core enzyme via two loops, the A1-K1 loop, which forms a conserved contact of Rpb4/7-like subcomplexes with their cognate core enzymes, and the tip loop, which may confer specificity to the interaction in the different RNA polymerases. To dock the A14/43 structure into the EM map, we modeled the conserved contact between an invariant proline residue in the A1-K1 loop (P51 in A43, Fig. 25A) and the common core subunit Rpb6. The tip domain and the tip-associated domain of the A14/43 structure fitted well to the EM map, and the lack of an HRDC domain could in part explain the smaller EM density (Fig. 25C+D). However, the peripheral OB

domain of A43 was not revealed in the EM density (Fig. 20C), suggesting a high degree of mobility. Consistently, the OB domain shows slightly higher B-factors in the crystal structure although it is involved in crystal contacts (not shown), and normal mode analysis of the Pol II crystal structure shows that the OB domain is the most flexible region of the enzyme. The A43 tip loop contains a specific ten-residue insertion that may confer specificity to the interaction between A14/43 and the Pol I core. The A43 tip loop is flexible in the crystal structure (Fig. 25A), but is likely folded upon binding to the Pol I core, as observed for Pol II (Armache et al., 2005).

Subunit A43 forms an essential bridge to the conserved Pol I initiation factor Rrn3 (Milkereit and Tschochner, 1998; Peyroche et al., 2000). Rrn3 was shown by EM to co-localize with A43 (Peyroche et al., 2000), and binds other initiation factors to recruit Pol I to the rDNA promoter. The A43-Rrn3 interaction is conserved in human (Yuan et al., 2002) and *S. pombe* (Imazawa et al., 2005). In a Pol I variant that is defective for Rrn3 interaction (*rpa43-6*, (Peyroche et al., 2000)), two out of three altered A43 residues map near conserved residues on the upstream surface of A14/43. Thus Rrn3 binds to A14/43 from the upstream side (Fig. 20C). Additional Pol I-specific surfaces in the vicinity include the extended dock domain and the clamp knob, which together with A14/43 create a specific upstream face for Pol I initiation complex assembly (Figs. 20, 24).

Differential initiation factor interactions and promoter-specificity of the three polymerases may generally result from differently structured dock domains, clamps, and Rpb4/7-like subcomplexes, which all constitute initiation factor binding sites. Rpb4/7 is required for Pol II initiation (Edwards et al., 1991). C17/25 binds to the Pol III initiation factor TFIIIB (Ferri et al., 2000), to the subcomplex C82/34/31 that bridges to TFIIIB (Bartholomew et al., 1993), and to the initiation factor TFIIIC (Hsieh et al., 1999). Since the surfaces, flexibility, and *in vivo* function of the HRDC domains differ in Rpb4/7 and C17/25 (Jasiak et al., 2006), the absence of an HRDC domain in A14/43 is likely to be functionally significant.

II.9 | A49 and A34.5 act as built-in, heterodimeric elongation factor

After assigning EM densities to the Pol I core and A14/43, one additional large density remained on the enzyme surface that was assigned to the Pol I-specific subunits A49 and A34.5 (Fig. 20C). To confirm this assignment, we dissociated subunits A49 and A34.5 from Pol I with the use of urea (Huet et al., 1975), purified the resulting 12-subunit variant Pol I ΔA49/34.5 (Chapter III.2), and solved its structure by cryo-EM at 25 Å resolution (Fig. 26 and Chapter III.8). The structure was similar to the complete Pol I, except that the density assigned to A49 and A34.5 was lacking (Fig. 26B). 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 A49/34.5 (Chapter III.8). 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 $\alpha16$-$\beta20$ of the Pol II pore domain), and A12.2. This is consistent with loss of A49 when Pol I is purified from *A12.2* deletion strains (Van Mullem et al., 2002).
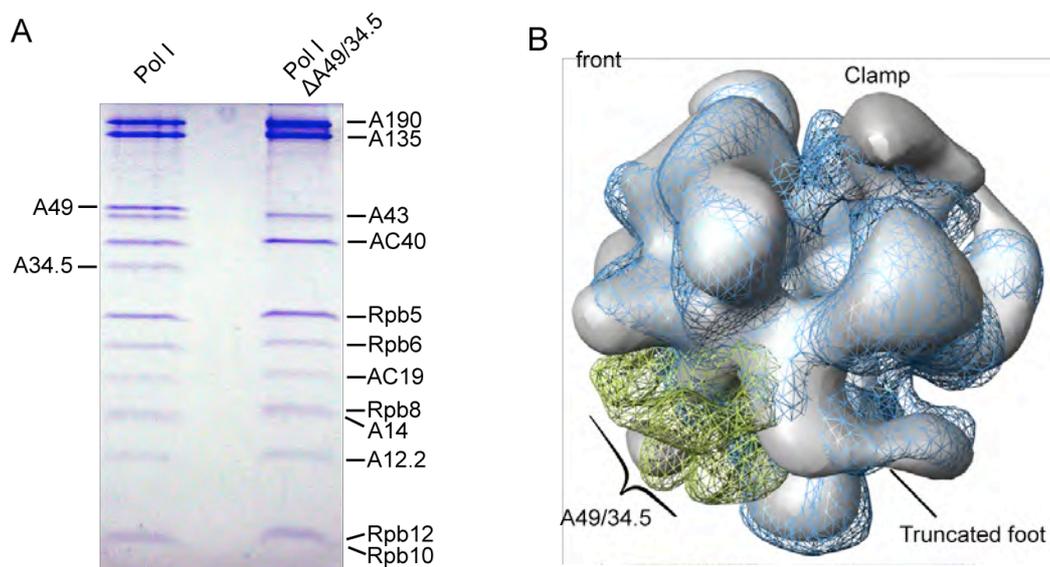
Figure 26 | Cryo-EM structure of A49/34.5. (A) SDS-PAGE analysis of the variant Pol I ΔA49/34.5 (right), obtained by urea treatment of the complete Pol I (left). **(B)** Overlay of EM structures of Pol I ΔA49/34.5 (silver surface) and the complete Pol I (blue). The density assigned to A49/34.5 is highlighted in green.

To investigate the structure and function of A49 and A34.5 we searched for weak homologies with HHpred (Soding et al., 2005). Local homologies were detected between A49 and RAP74, the large subunit of the Pol II-associated factor TFIIF, and between A34.5 and RAP30, the small TFIIF subunit (Fig. 27 and Chapter III.10). Consistently, the N-terminal regions of A49 and A34.5 were predicted to contain β-strands consistent with the fold of the RAP74-RAP30 dimerization module (Gaiser et al., 2000), and hydrophobic core residues in this fold were predicted to be conserved (Fig. 27). Consistent with these predictions, bacterial co-expression of A49 and A34.5 enabled isolation of a stoichiometric A49/34.5 heterodimer (Fig. 28A and Chapter III.11), and alanine point mutations in three different conserved hydrophobic residues in the dimerization interface (I12 and Y76 in A49, W54 in A34.5) abolished or strongly impaired A49-A34.5 co-purification (Fig. 28B-D). Thus, A49 and A34.5 form a stable TFIIF-like heterodimerization module.
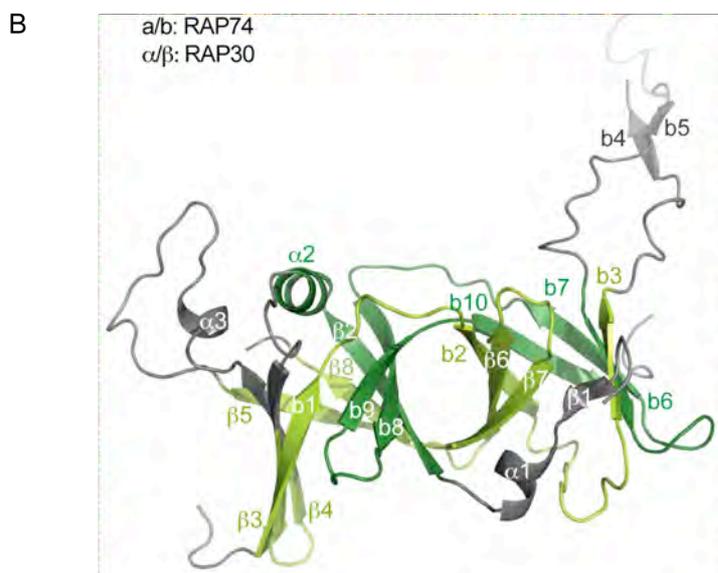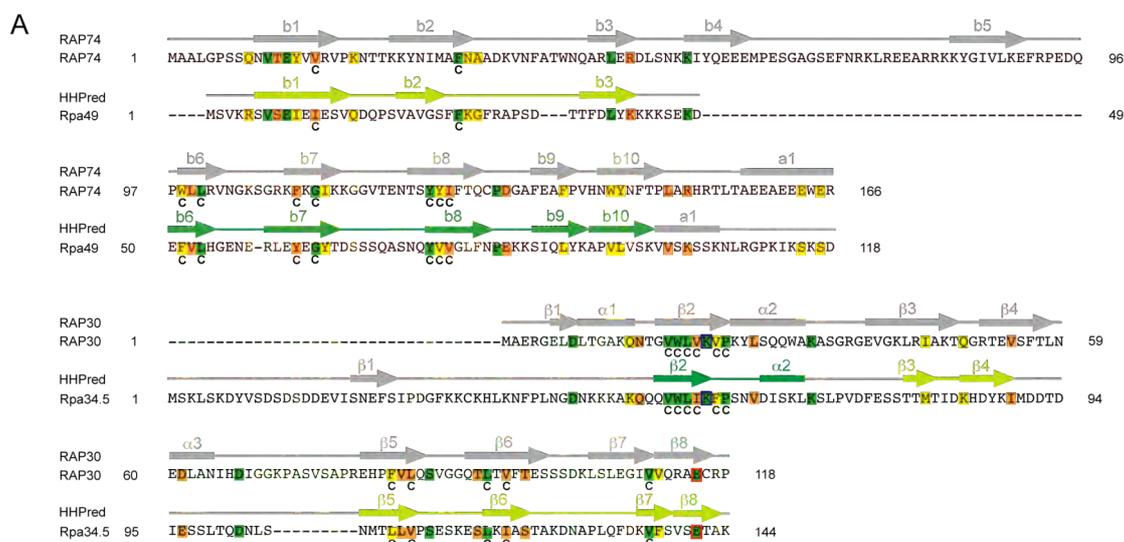
Figure 27 | (A) Sequence alignments of *S. cerevisiae* A49 and A34.5 with their putative counterparts in *H. sapiens* TFIIF (RAP74 and RAP30, respectively). Sequence similarity is only observed in the N-terminal part of both proteins (residues 1-166 in RAP74 and residues 1-118 in RAP30). Secondary structure elements are shown above the sequences (broad lines, α-helices; arrows, β-strands; lines, loops). Conserved residues are highlighted according to decreasing conservation from green, through orange, to yellow. Residues involved in a conserved core interaction are marked with a C below the sequence, while charged residues forming a salt bridge are depicted in blue and red, respectively. Secondary structure elements are depicted above the RAP74/30 sequences, according to structural information (Gaiser et al., 2000). For clarity, the symbols a/b are used in RAP74, α/β in RAP30. For A49 and A34.5, predicted secondary structure elements are depicted in dark green (aligned by HHpred), light

green (predicted to be present by secondary structure propensity) and grey (not predicted to be present). **(B)** Conservation of the TFIIF RAP74/30 dimerization module in A49/34.5. Secondary structure elements aligned to RAP74/30 are highlighted in dark and light green, respectively. For details see (A).
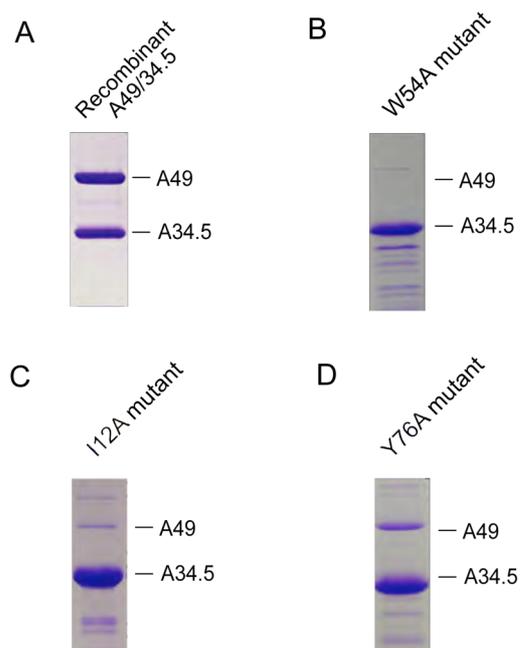
Figure 28 | Hydrohobic core point mutations. (A) Recombinant wild-type A49/34.5. **(B)** W54A mutant in A34.5. **(C)** I12A mutant in A34.5. **(D)** Y76A mutant in A49.

Heterodimerization is consistent with the observed continuous EM density, which reconciles previous EM data. Initial cryo-EM showed two separated densities over the cleft that were assigned to A49 and A34.5 (Bischler et al., 2002). EM at higher resolution did not confirm these densities, but revealed a new additional density (De Carlo et al., 2003) that was close to the location of A49/34.5 found here. The location of A49 and A34.5 distant from the DNA-binding cleft explains why neither A49 nor A34.5 could be crosslinked to DNA in Pol I initiation complexes (Bric et al., 2004).

The location of A49/34.5 at the Pol I funnel deviates from that of TFIIF on Pol II as observed by cryo-EM (Chung et al., 2003), but is more consistent with protein-protein cross-linking that maps TFIIF to the polymerase lobe and outer surface near Rpb9 (Chen et al., 2007). Discrepancies in the location of

A49/34.5 and TFIIF may be explained by different locations of a related dimerization module on the two polymerases, or by the presence of additional, unrelated domains in both factors. Sequence analysis showed that A49/34.5 and TFIIF possibly have a counterpart in Pol III, the C37/53 heterodimer, which may occupy a similar location on the Pol III surface near the lobe and funnel (Fernandez-Tornero et al., 2007).

The apparent homology of the A49/34.5 heterodimer with the N-terminal regions of the two large TFIIF subunits suggested that A49/34.5 has elongation-stimulatory activity. We therefore compared the complete Pol I with Pol I ΔA49/34.5 in an RNA extension assay using a minimal DNA-RNA scaffold (Chapter III.13.1). 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 (Fig. 29A). Addition of recombinant A49/34.5 rescued the defect of Pol I ΔA49/34.5, and enabled elongation to the end of the template (Fig. 29A, lane 4). We repeated the elongation experiments using a complete, complementary transcription bubble scaffold (Fig. 29B and Chapter III.13.2) (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. 29B, 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*.
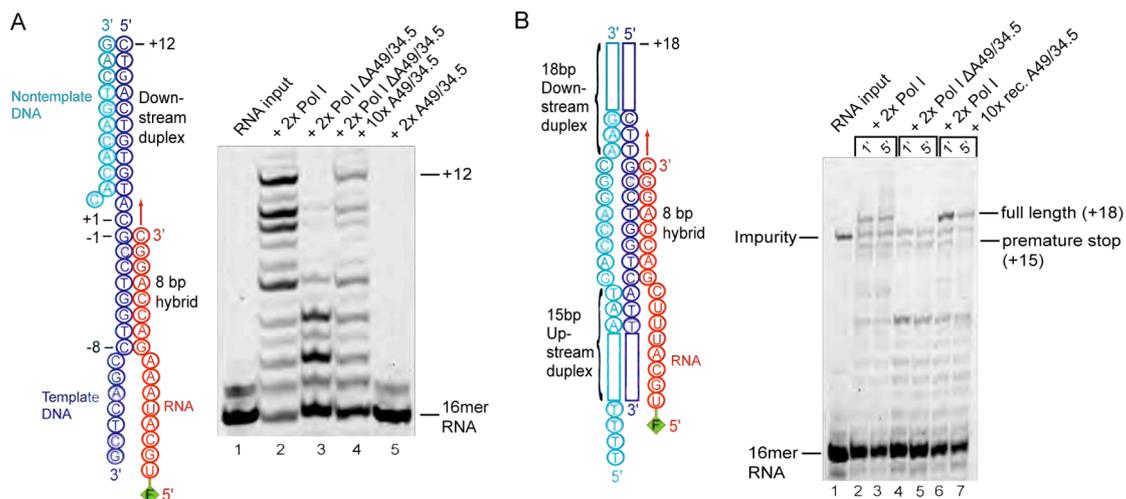
Figure 29 | Elongation-stimulatory activity of A49/34.5. (A) A49/34.5 shows elongation-stimulatory activity in RNA extension assays with a minimal nucleic acid scaffold. The fluorescent label 6-carboxy-fluoresceine (FAM) on the RNA 5'-end is indicated. The times molar excess of added factors are indicated 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).

To test whether A49/34.5 may have elongation-stimulatory function *in vivo*, we investigated if the growth phenotype of a yeast strain that lacked the gene for A34.5 (*ΔA34.5*) is affected when nucleotide supply was limited due to the presence of 6-azauracil (6AU). 6AU sensitivity is an indicator for Pol II-associated elongation factor function *in vivo*, and recently also identified a Pol I mutant defective in rRNA elongation (Schneider et al., 2007). Whereas the wild type and *ΔA34.5* strains did not show a growth difference on normal media, the *ΔA34.5* strain showed a mild slow-growth phenotype on 6AU-containing media (Fig. 30). This suggested that A49/34.5 is required for normal RNA elongation by Pol I also *in vivo*. The elongation-stimulatory activity may be due to an allosteric effect, or due to an extension from A49/34.5 into the active center, but we cannot distinguish between these possibilities with the available structural data.
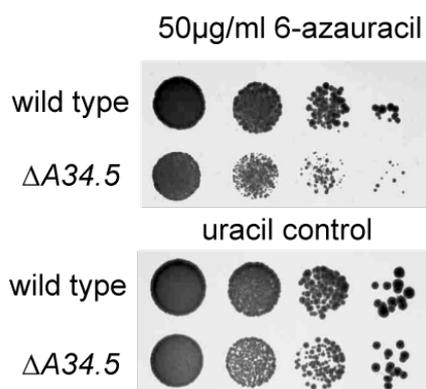
Figure 30 | Elongation activity of Pol I *in vivo*. Deletion of the gene for A34.5 leads to a 6-azauracile-sensitive phenotype. From left to right 1:10 dilution series are shown. As a control, cells were spread onto SDC plates containing uracil.

II.10 | Pol I has intrinsic RNA 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, a 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, we assembled a "backtracked" elongation complex from purified Pol I and a DNA-RNA scaffold that contained an RNA 3'-overhang (Fig. 31 and Chapter III.13.3). Incubation of the backtracked complex with 8 mM magnesium ions led to efficient shortening of the RNA from the 3'-end (Fig. 31B, lanes 1-3 and Chapter III.13.3). 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 (Fig. 31B, 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.

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.
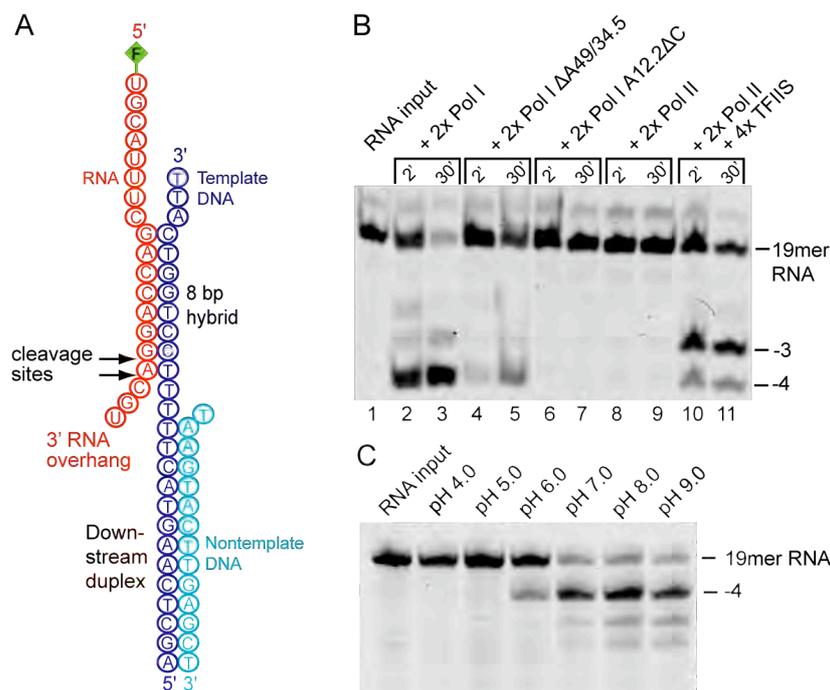
Figure 31 | Intrinsic RNA cleavage activity of Pol I. (A) DNA-RNA hybrid scaffold used in cleavage assays. **(B)** Comparison of RNA cleavage by Pol I variants with Pol II and the Pol II-TFIIS complex. **(C)** pH-Dependence of Pol I cleavage activity.

Additional cleavage assays showed that the Pol I variant ΔA49/34.5 cleaved RNA less efficiently than the complete Pol I (Fig. 31B, 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). We also asked whether subunit A12.2 is required for cleavage, since its counterpart C11 is essential for cleavage activity of Pol III (Chedin et al., 1998; Whitehall et al., 1994). A Pol I variant lacking residues 79-125 of A12.2 (A12.2ΔC, Chapter III.2)

was totally inactive in RNA cleavage (Fig. 31B, lanes 6+7), but bound the nucleic acid scaffold in electrophoretic mobility shift assays (Fig. 32A and Chapter III.13.4), and retained elongation activity (Fig. 32B). 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), but its location in Pol I corresponds to that of the Rpb9 C-terminal domain on Pol II (Fig. 21B). Although the long linker between the A12.2 N- and C-terminal domains (Chapter IV.7) could in principle allow swinging of the C-terminal domain into the pore, our results suggest that the effect of A12.2 truncation on cleavage is due to an allosteric rearrangement in the Pol I active center. This effect might be mediated by the trigger loop of Pol I, since it is in close proximity to the C-terminal domain of A12.2. Mutation of the residues in A12.2 homologous to the catalytical D53 and E54 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 (Fig. 31C). The structural basis of the effect of A12.2 truncation on RNA cleavage awaits the crystal structure of Pol I.
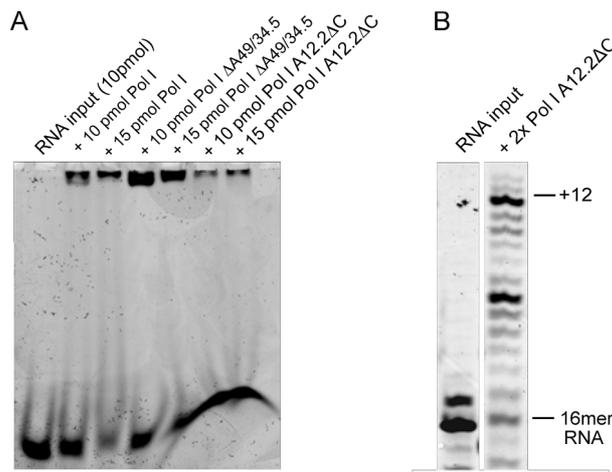
Figure 32 | Biochemical properties of Pol I A12.2ΔC. (A) Electrophoretic mobility shift assay (EMSA). **(B)** Elongation activity of the Pol I variant A12.2ΔC.

Since A12.2 is required for transcription termination (Prescott et al., 2004), Pol I cleavage activity may be involved in a termination-coupled reaction. RNA cleavage could be required for rRNA 3'-terminal trimming, a Pol I-associated RNA processing event that intimately follows termination and involves cleavage of ten nucleotides from the pre-rRNA 3'-end (Kuhn and Grummt, 1989). Consistently, Pol II can form a binary complex with RNA and cleave RNA from the 3'-end in the presence of TFIIS (Johnson and Chamberlin, 1994).

It is very likely that the intrinsic cleavage activity of Pol I also enables rRNA proofreading, to increase transcriptional fidelity. Indeed, repetition of our cleavage assay with a scaffold that contains only a single mismatch at the RNA 3'-end, mimicking the situation after a misincorporation event, induced efficient RNA cleavage (not shown). For Pol III, the intrinsic cleavage activity was recently shown to enable proofreading in a manner dependent on the A12.2 homolog C11 (Alic et al., 2007), which is required for the intrinsic cleavage activity of Pol III (Chedin et al., 1998; Landrieux et al., 2006).

II.11 | Conclusions
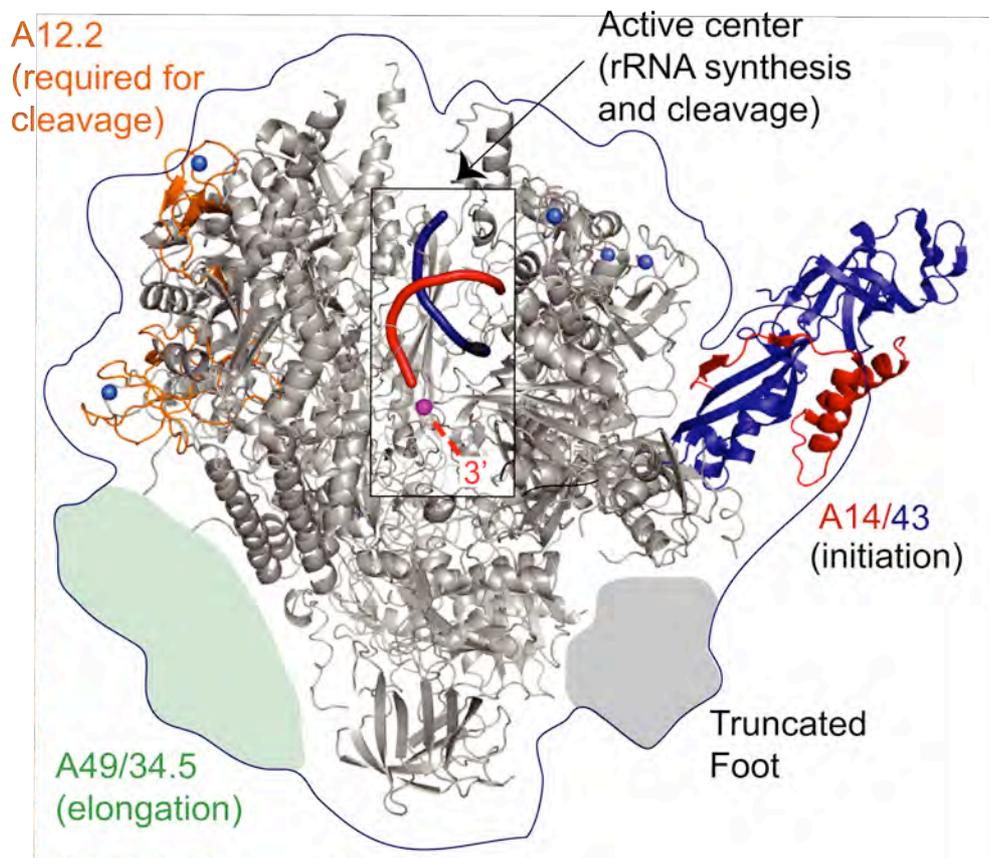
Figure 33 | Hybrid structure and functional architecture of Pol I. The EM envelope is shown as a blue line, the Pol I core ribbon model in grey with Rpb9 (A12.2) highlighted in orange, and the A14/43 crystal structure in red/blue. The window shows a cut-away view of the active center containing a modeled DNA-RNA hybrid. Red dashes indicate the RNA 3'-end extruded into the pore.

In this thesis a reproducible large-scale purification protocol for RNA polymerase I from *S. cerevisiae* was developed. Many crucial steps were completed successfully on the way to an atomic resolution X-ray structure of this huge, multi-subunit complex: Crystals were obtained by microseeding, diffraction to < 4 Å could be recorded of heavy atom soaked crystals and

complete data to 4.8 Å could be processed. However, the enormously complex non-crystallographic symmetry in the asymmetric unit of Pol I impeded structure solution.

Single-particle cryo-electron microscopy provided a way out of this dilemma: The detailed functional architecture of Pol I could be elucidated by a combination of structural biology techniques and structure-based functional analysis (Fig. 33). Comparison with the Pol II system revealed Pol I-specific features that match the unique nature of rRNA transcription. First, the distinct structure of the Pol I upstream face allows for specific initiation factor interactions and recruitment of Pol I to the rRNA promoter. Second, the built-in elongation-stimulatory Pol I-specific subcomplex A49/34.5 can explain the efficient and processive nature of rRNA transcription during cell growth. Third, the intrinsic RNA cleavage activity apparently enables rRNA 3'-trimming and proofreading, to prevent formation of erroneous rRNAs and catalytically deficient ribosomes. Finally, our results will help to unravel structural and functional relationships between the three eukaryotic transcription machineries, and form the basis for a detailed structure-function analysis of rRNA transcription and processing.

III.1 | Purification of RNA Polymerase I from *S. cerevisiae*

Buffers and media used during purification:

YPD medium (for small fermenter)

300g peptone

300g glucose

225g yeast extract

add 15 L with desalted water

pH adjusted to 6.9 with 1M NaOH

50 µg/ml ampicilin[1]

10 µg/ml tetracycline[1]

100x Protease Inhibitors (PI)

1mM phenylmethylsulphonyl fluoride (PMSF)

1mM benzamidine

200µM pepstatin

60µM leupeptin

dissolved in 100% EtOH

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 $MgCl_2$

400 mM $(NH_4)SO_4$

5 mM DTT[1]

1x PI[1]

2x Dialysis buffer

100 mM KOAc

40 mM HEPES, pH 7.8

20 mM $MgCl_2$

10% (v/v) glycerol

10 mM mercaptoethanol[1]

1 mM benzamidine[1]

1 mM PMSF[1]

Res/W1 buffer

1.5 M KOAc

20 mM HEPES, pH 7.8

1 mM $MgCl_2$

10% (v/v) glycerol

10 mM 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 mercaptoethanol[1]

E100 buffer

300 mM KOAc

20mM HEPES, pH 7.8

1 mM $MgCl_2$

100 mM imidazole

10% (v/v) glycerol

10 mM mercaptoethanol[1]

MonoQ buffer A

20 mM HEPES, pH 7.8

1 mM MgCl2

10% (v/v) glycerol

5 mM DTT[1]

MonoQ buffer B

2 M KOAc

20 mM HEPES, pH 7.8

1mM MgCl2

10% (v/v) glycerol

5 mM DTT[1]

Superose 6 buffer A

60 mM $(NH4)_2SO_4$

5 mM HEPES, pH 7.8

1 mM MgCl2

10 µM ZnCl2

5 mM DTT[1]

[1] added prior to usage

The complete 14-subunit Pol I was isolated from a modified version of the *S. cerevisiae* strain GPY2 (*ade2-101*, *trp1-Δ63*, *ura3-52*, *his3-Δ200*, *lys2-801*, *leu2*::RPA43), carrying a pAS22 plasmid coding for a HA- and hexahistidine-tagged version of A43. A 20 L fermenter (Infors ISF) was inoculated to a starting OD_{600} of 0.15-0.3 with cells cultivated in shaking flasks. Fermentation was carried out in YPD medium at 30 °C, using a stirrer speed of 650 rpm and an air flow of 8 L/min. Growth was allowed to proceed for approx. 8-9 hours until the culture reached an OD_{600} of 1.5. This pre-culture was used to inoculate a 200 L fermenter (Infors ABEC) with a starting OD_{600} of 0.15. Cells were grown over night at 30 °C until they reached an OD_{600} of 5-9 (approx. 18h, Fig. 6, Chapter II.1). Harvesting cells was achieved by flow-throw centrifugation at 20,000 rpm (Padberg Z4IG), yielding 1.7-2.8 kg of yeast pellet. Cells were re-suspended in 500 ml of freezing buffer per kg of cells and stored at -80 °C after shock-freezing in 225 mL batches in liquid nitrogen.

For each 'standard' Pol I purification two 225 mL cell batches were carefully thawed in warm water. Ammonium sulfate was added to a final concentration of 400 mM, DTT and protease inhibitors were added to final concentrations of

5 mM and 1 x respectively. To prevent foam formation every BeadBeater^{TM} (Biospec Inc.) was filled up to prevent an airspace, using dilution buffer. Yeast cells were lysed using 200 mL glass beads per BeadBeater. Lysis was carried out in repetitive cycles of 30 s bead-beating followed by 1 min of cooling. During this 1 h procedure the lysate was cooled using a salt-water mixture. Thereafter, glass beads were separated by filtration prior to clearing the lysate by centrifugation (30 min, 8000 x g, Sorvall SLA-1500). The whole cell extract was ultra-centrifuged for 90 min at 30,000 x g (Beckman SW-28). After aspiring the top fat layer, the clear supernatant was dialyzed over night at 4 °C against 1 x dialysis buffer. The dialyzed extract was centrifuged for 1 h at 18,500 x g (Beckman Ti-45), the pellet re-suspended in app. 50 mL Res/W1-buffer and incubated with 8 mL Nickel-NTA Agarose (Qiagen) for 4 h at 4 °C on a turning wheel. The resin was packed into 2 gravity flow nickel columns and washed with 5 column volumes (CV) of Res/W1 buffer and 5 CV of W2-buffer, and eluted using 50 mL E100 buffer. For anion-chromatography, a Mono-Q column (MonoQ 10/100 GL, GE Healthcare) was equilibrated with 15% MonoQ buffer B and eluted with MonoQ buffer B, using a multi-step gradient (Fig. 7, Chapter II.1), resulting in an elution peak for Pol I at 1100 mM KOAc (Fig. 8, Chapter II.1). Peak fractions were pooled (approx. 10 mL) and diluted to a final KOAc-concentration of 200 mM. A cation-exchange column (MonoS 5/50 GL, GE Healthcare) was used for the next purification stage, using the MonoQ buffers A and B and applying a gradient from 200mM KOAc to 2M KOAc (Fig. 9, Chapter II.1). Pure Pol I eluted at a salt concentration of 490 mM KOAc (Fig. 10A, Chapter II.1). To remove remaining glycerol and to check for monodispersity, the protein was finally purified on a Superose 6 HR10/30 size-exclusion column (GE Healthcare) in Superose 6 buffer A (Fig. 11, Chapter II.1).

III.2 | Purification of Pol I variants

Additional buffers for purifying Pol I variants:

<u>Urea dissociation buffer</u>

2 M urea

50 mM ammonium sulfate

1 mM magnesium chloride

20 mM HEPES, pH 7.8

10% (v/v) glycerol

5 mM DTT[1]

<u>MonoQ buffer C</u>

50 mM ammonium sulfate

20 mM HEPES, pH 7.8

1mM $MgCl_2$

10% (v/v) glycerol

5 mM DTT[1]

<u>MonoQ buffer D</u>

1 M ammonium sulfate

20 mM HEPES, pH 7.8

1mM $MgCl_2$

10% (v/v) glycerol

5 mM DTT[1]

<u>Superose 6 buffer B</u>

100 mM ammonium sulfate

20 mM HEPES, pH 7.8

5% (v/v) glycerol

1 mM $MgCl_2$

10 µM $ZnCl_2$

5 mM DTT[1]

[1]added prior to usage

Pol I lacking the A49/34.5 heterodimer (Pol I ΔA49/34.5) was prepared by controlled urea dissociation of A49/34.5 from complete Pol I. Pol I-containing fractions after cation-exchange chromatography (Chapters II.1 and III.1) were dialyzed over night against a urea dissociation buffer. A49/34.5 was separated from Pol I ΔA49/34.5 by subsequent anion-exchange chromatography, applying a linear gradient from 50 mM to 1 M ammonium sulfate, using MonoQ buffers C and D. Pol I ΔA49/34.5 was further purified by size-exclusion chromatography (Superose 6 HR10/300, GE Healthcare) using Superose 6 buffer B (Fig. 26A, Chapter II.9). For further biochemical use (Chapters II.9 and II.10) pooled fractions were concentrated to 0.5 mg/mL.

The Pol I variant lacking the C-terminal residues 79-125 of A12.2 (Pol I A12.2ΔC) was fermented in synthetic dextrose complete (SDC) medium lacking histidine and purified exactly as described for the complete enzyme (Chapter III.1), omitting the final gel filtration step. Pol I A12.2ΔC was concentrated to 0.5 mg/mL in Superose 6 buffer B. The yield for this Pol I variant was incredibly low, not more than 50 μg protein could be obtained from 200 g of cell pellet.

III.3 | Crystallization of Pol I

III.3.1 | Crystallization by vapor diffusion

Pol I purified as described in Chapter III.1, was concentrated to 5.5 mg/mL in Membra-Spin Mini columns (membraPure, Bodenheim, Germany). UV absorption at λ = 280 nm was measured and protein concentration was determined assuming an absorption coefficient of 0.656 for a 1 mg/mL protein solution (derived from ProtParam on www.expasy.ch). To allow for slow concentration of the sample centrifuge speed was reduced to 6,500 rpm at 4 °C. Before crystallization the protein was centrifuged at 14,000 rpm for 30 min at 4 °C to remove dust and aggregated particles.

Crystallization was always carried out using vapor diffusion. For hanging drops EasyXtal Tools (Nextal/Qiagen) were used, for sitting drops Linbro plates (Hampton Research). Drops were set using 500 μL reservoir solution and 1 μL protein + 1 μL reservoir drops. Protein was added prior to adding reservoir solution. The reservoir contained in all cases fresh reducing agent, either 5 mM DTT or 3 mM Tris(2-carboxyethyl)phosphine (TCEP).

III.3.2 | Streak-seeding

Crystals of suitable size for measuring X-ray diffraction data could only be obtained using streak-seeding (Bergfors, 2003). Cat whiskers were 'stolen' from Micio (Michela's cat), Lintelo (Katrin's cat) and Mia (Anette's cat). Vapor diffusion setups were allowed to equilibrate for 3-3.5 h prior to streak-seeding. Source drops, from which seeds were derived, were prepared for seeding by adding 10 μL of reservoir solution (fresh reducing agent was added prior to dilution). Seeds were collected by streaking several times with the cat whisker through the diluted source drop. Every streak-seeding trial was performed for at least 6 identical drops, diluting the seeds consecutively. Drops were closed immediately after having streaked through them (Fig. 13, Chapter II.2).

III.3.3 | Crystal harvesting and cryo-protection

Crystals grew to maximum dimensions of 500μm x 60μm x 10μm. For crystal manipulation and freezing tools from Hampton Research (Aliso Viejo, USA) were used. 5 μL of reservoir solution was added to the drop containing the crystals. For cryo-protection crystals were transferred to spot plates containing 100 μL of the crystallization condition + 6% PEG-400 (3 mM TCEP was freshly added). Crystals were allowed to equilibrate for approx. ½ h before increasing the concentration of cryo-protectant stepwise to 12, 18 and finally 22% PEG-400. Solutions were exchanged rather than crystals transferred to the new cryo-solution. The final 22% step was repeated to ensure full exchange of cryo-solutions.

III.3.4 | Heavy atom derivatization and crystal freezing

A grain of a W_{18} cluster $(NH_4)_6(P_2W_{18}O_{62}) \cdot 14H_2O$ (Dawson, 1953; Thygesen et al., 1996) was added to the final cryo-solution containing 22% PEG-400. Crystals were slowly cooled down to 8 °C using a styrofoam box, and kept at this temperature for app. 44 h. Crystals were plunged into liquid nitrogen and stored at liquid nitrogen temperature until data collection. For details on different heavy atom derivatization techniques see Table 2, Chapter II.3. Crystals were harvested using 20 μm CrystalCap HT equipment from Hampton Research (loop size 0.1 – 0.4 mm, sample holder length 22 mm).

III.4 | Data collection

During the course of improving crystal size and diffraction quality, many different ways of collecting data on Pol I crystals were tried. The following paragraph illustrates the approach for some of the best crystals collected.

All diffraction data were recorded at the beamlines X06SA and X10SA at the Swiss Light Source (SLS) in Villigen, Switzerland, using a Mar225 detector (Mar Research). To verify the W_{18} cluster in the crystal, an X-ray absorption scan was performed at the L-III edge of tungstate (10.21 keV or 1.21 Å) before measurement of reflexions (Fig. 15A, Chapter II.4). The main difficulty was the extreme sensitivity of the crystals when exposed to X-rays. Therefore the beam flux was kept constant at 1×10^{12} photons/s to enable comparison of diffraction quality between crystals. The beam was focused on the detector rather than on the crystal for reducing radiation damage. Crystals were all of monoclinic space group C2, requiring at least a 90° rotation for recording all possible anomalous pairs. Strategies to minimize X-ray exposure for collecting complete data were simulated using MOSFLM (Leslie, 2006). In most cases, an exposure time of 1 s per 0.5° oscillation was used; one image of the direct beam was recorded for every detector–crystal distance to allow best possible indexing. Even though

the unit cell dimensions were huge, no overlapping reflections were predicted by MOSFLM, an effect due to the limited resolution. However, even after following all these preventative measures, several translations on a single crystal were necessary to record a full dataset, introducing many problems concerning data integration and scaling.

III.5 | Data processing

Data were processed with HKL2000 (Otwinowski and Minor, 1997), not using the graphical interface but DENZO and SCALEPACK scripts (Chapter IV.1). SCALEPACK had to be used with its derivative SCALEPACKRIBO to account for the enormous number of measured reflections. During integration most difficulties arose from radiation-damage induced cell parameter changes, which made refinement of these parameters very difficult. Scaling often suffered from high mosaicity of > 0.7° and many different translations. Suitable images for scaling were determined by monitoring the average $I/\sigma(I)$ per frame and the batch-wise R-factor in the output log-file of SCALEPACK. The model for systematic error was stepwise adjusted including all rejected reflections in each cycle until convergence. Data quality criteria of $I/\sigma(I)$ above 2 and an R_{merge} of < 35% were applied before subsequent attempts at phasing (Fig. 16C, Chapter II.4).

Self-rotation functions were calculated using POLARRFN from the CCP4 package (CCP4, 1994) and GLRF (Tong and Rossmann, 1990). Stereographic projections were calculated for $\kappa = 52°$ and $\kappa = 180°$. See Chapter IV.2 for details.

III.6 | Attempts on structure solution

III.6.1 | Experimental phasing

For locating the W_{18} and $Ta_6Br_{12}^{2+}$ clusters in anomalous difference Patterson maps the program SOLVE was used (Terwilliger and Berendzen, 1999), see sample scripts in Chapter IV.3. SHELXD (Schneider and Sheldrick, 2002), which uses direct methods for solving sub-structures, and its graphical interface HKL2MAP (Pape and Schneider, 2004) were also tried with various input settings (resolution range, Patterson seeding, number of sites). Unfortunately, none of the various trials produced unambiguous heavy atom sites.

III.6.2 | Molecular replacement

For molecular replacement various models based on the Pol II structure (Armache et al., 2003) were constructed: Model 1 comprised Rpb1, 2, 3 and 11 of the Pol II structure, including deletions according to hand-made structure-based alignments (Chapter IV.7); the identical subunits Rpb5, 6, 8, 10 and 12 were kept, but Rpb4, 7 and 9 were excluded due to limited sequence conservation. Model 2 additionally included the tip domain of Rpb7 (amino acids 1-82) and the N-terminus of Rpb9 (amino acids 2-39), which could possibly have enhanced the molecular replacement signal due to Rpb7 protruding from of the core polymerase. Model 3 was a poly-alanine model of Model 2 (but maintained any glycine residues). Model 4 was also identical to Model 2, however, it was based on the TFIIS-bound RNA polymerase II structure (Kettenberger et al., 2004), in which large parts of Pol II are slightly shifted against each other.

All these models were used for running PHASER (McCoy et al., 2005; Read, 2001; Storoni et al., 2004). However, even after extensively examining all possible variations in the rotation and translation functions, fixing solutions or

changing parameters like the search radius, the similarity score or the included reflections, no plausible solutions could be obtained, the main problem being the presence of 7 molecules per asymmetric unit (Fig. 17, Chapter II.5). To exploit the high NCS symmetry of the apparent 7-fold ring the locked cross-rotation function of GLRF (Tong and Rossmann, 1990) and MOLREP (Vagin and Teplyakov, 1997) were used, but this did not lead to improved signals. Nevertheless, MOLREP resulted in some rotation solutions that did obey the 7-fold symmetry, but unfortunately, none of these 7-fold related solutions were successfully solved by the following translation function (Fig. 18A, Chapter II.5).

Apart from crystallographic models the cryo-EM structure of Pol I (Chapter II.6) was used for molecular replacement. For that purpose the EM map had first to be converted into CCP4-format using SPIDER (Frank et al., 1996). The SPIDER volume was interpolated to 1 Å/pixel using the command IP. The resulting volume was padded into a 300 Å x 300 Å x 300 Å unit cell using PD. After determination of the center of gravity (command CG) the molecule was shifted to this center (command SH) and finally the map was converted to CCP4-format using CP TO CCP4 in 32-bit mode. This resulting map was used as a search model in MOLREP, with the self-rotation information calculated previously. The table of rotation solutions showed for the first time clusters of 7 solutions in one plane (Figure 18B, Chapter II.5), and gave us confidence that this was representative of the 7-fold NCS. The low- and high-resolution cut-offs were set to 12 Å and 80 Å respectively, corresponding to the limits of both the crystallographic and EM data (for detailed script see Chapter IV.4).

However, the translation function could not be solved using these rotational solutions.

III.7 | Cryo-electron microscopy of Pol I

III.7.1 | Negative stain

For EM data collection, Pol I was concentrated to 5.5 mg/mL, as for crystallographic purposes (Chapter III.3.1). For determining the optimal protein concentration for cryo-EM, negative stain images were recorded with a Philips CM100 transmission electron microscope operated at 100 kV and a nominal magnification of 28,500 (defocus ranging from -300 nm to -500 nm). The sample was stained using 2% uranyl acetate. An optimal protein concentration of 0.1 mg/mL for cryo-EM could be established by assessing particle density visually (Fig. 20A, Chapter II.6).

III.7.2 | Preparation of grids

A thin carbon layer was vapor-deposited onto a mica layer (Plano GmbH, Wetzlar, Germany) using a Bench Top Turbo IV Coating System (Denton Vacuum LLC, Morestown, USA) using a vacuum of 5-10 x 10^{-6} Torr. The ultra-thin carbon layer was floated onto water before applying it to carbon holey grids (Quantifoil). For making the carbon surface hydrophilic, grids were ionized in a plasma cleaner chamber (Model PDC002, Harrick, UK). 3.5 μL of sample were applied to the grid and subsequently vitrified in liquid ethane using the half-automated Vitrobot^{TM} system (FEI, Eindhoven, Netherlands) under controlled conditions (6 °C, 100% humidity, 45 s incubation time, 7.5 s blotting) (Wagenknecht et al., 1988). Grids were transferred into liquid nitrogen for long-term storage.

III.7.3 | Cryo-EM data collection

Data were collected using a Tecnai Polara F30 field emission gun microscope operated at 300 kV and a magnification of 39,000 (Max-Planck Institute of Molecular Genetics, Berlin). Frozen grids were transferred into the specimen holder of the microscope under liquid nitrogen conditions. Meshes were screened by hand to identify suitable ones for data collection and sample images recorded on a 4k x 4k CCD camera (Fig. 20B, Chapter II.6). Micrographs were only recorded in regions of thin carbon with a low dose of 20 electrons/$\mathring{A}^2$ and an exposure time of 1 s. Micrographs were developed and scanned on a Heidelberg drum scanner with a pixel size of 1.23 $\mathring{A}$ (5334 dpi) on the object scale. Micrographs were saved as high resolution TIFF-files.

III.7.4 | Image processing for 14-subunit Pol I

All data were processed using the SPIDER software package (Frank et al., 1996). For all TIFF images the contrast transfer function and defocus values were determined using CTFFIND (*p_ctffind.rib*, (Mindell and Grigorieff, 2003)). Power spectra were visually inspected in Web (part of the SPIDER package). 59 micrographs (out of 84 recorded) that displayed very little drift and astigmatism were further considered and were 3-fold decimated to a pixel size of 3.69 $\mathring{A}$/pixel (*sig_decimate.rib*) with a box size of 60 pixels. Particles were picked automatically with SIGNATURE (*sig_pick.rib*) (Chen and Grigorieff, 2007), using 5 projections of the 12-subunit Pol II structure (Armache et al., 2005) as template. Bad particles were excluded from the dataset after visual inspection in Web (*p_window.rib*, *p_dcs_flt.rib*, *p_copygood.rib*). Selected micrographs were assigned to 29 defocus groups having similar defocus values (see Chapter IV.5).

In the first alignment step 31,600 particles from 15 micrographs were aligned to projections of the reference volume (*p_alidef.rib*). As a reference, the Pol II

structure, filtered at 20 Å resolution, was modified by deleting the clamp and foot domains of Rpb1 and Rpb4/7 except for the Rpb7 tip. Depending on the defocus value of each defocus group, the reference was distorted with the corresponding CTF function in angular increments of 15°, which resulted in 83 projections. Allowed shifts of particles in x and y directions were first kept as large as possible and successively tightened during refinement. The alignment procedure resulted in the best fitting projections (according to cross-correlation) for each particle and the shifts and rotational changes needed to match each projection.

Particles were backprojected using the parameters gained from the alignment (*p_trans.rib, p_spinnem2.rib, p_rotate.rib, p_angles.rib, bp32f_n.rib*). To determine the resolution of the reconstruction, the dataset was randomly split into two equal subsets, both were backprojected and CTF corrected. The resolution was then determined based on Fourier shell correlation (FSC), using a cut-off value of 0.5 (Fig. 20D, Chapter II.6).

In the first round of refinement particles were aligned to the volume resulting from the first backprojection. Further refinement required creation of so-called 'stack' files containing aligned particles and transformation files containing shifts and rotational parameters to fit the reference projection. Particles were iteratively aligned to new references created by the aligned particles in the preceding round. Initially, all possible reference projections were offered to each particle. Later, reference projections that are compared become more and more restricted to a defined angular and translational range. In our case, density for the clamp and foot reappeared during early refinement, confirming the absence of reference bias. To account for the many different clamp conformations, particles were sorted into two subsets according to two different clamp conformations (Penczek et al., 2006). For that purpose two different volumes were offered to the refinement algorithm. For Pol I, we offered as volume 2 the initial reference containing the clamp in a position similar to Pol II. Sorting resulted in 19,130 particles with a closed clamp conformation (volume 1, class I) and 12,546 particles with an open clamp conformation (volume 2, class II) with

3D reconstructions at a resolution of approx. 17 Å (Fig. 20C, Chapter II.6). To be able to reach higher resolution, the pixel size was decreased to 1.84 Å/pixel at this stage. Addition of more particles from the remaining 44 micrographs and further sorting against human Pol II (Kostek et al., 2006) resulted in 46,056 particles and led to a reconstruction at a resolution of 11.9 Å (0.5 FSC). During the last refinement rounds (in total 151), a better algorithm (BP RP), based on real space backprojection, was applied, which resulted in higher resolution reconstructions. Higher frequencies were corrected by multiplying the 3D-volume in Fourier space using an exponential function, similar to a crystallographic *b*-factor.

III.8 | Cryo-EM data processing for 12-subunit Pol I ∆A49/34.5

For cryo-EM structure determination of Pol I ∆A49/34.5, data processing was carried out as for the complete Pol I, and was again bias-free. 20,668 particles of high defocus values (> 3 µm, Chapter IV.6) from 13 micrographs were aligned with SPIDER (Frank et al., 1996) using the same reference as for the complete Pol I (Chapter III.7). Only spurious density fragments were observed in the region assigned to the two dissociated subunits. Sorting for the A49/34.5 density was carried out until convergence (Penczek et al., 2006). Sorting revealed once more the enormous flexibility of the clamp since, apart from the missing density for A49/34.5, there were also different clamp positions observable. The remaining 11,226 particles were backprojected using the BP32F algorithm, resulting in a volume with 25 Å resolution (Fig. 26B, Chapter II.9).

III.9 | Modeling of the Pol I core

The Rpb4/7 sub-complex was removed from the complete Pol II structure and the five common subunits were retained in the model. For the Pol II subunits Rpb1, Rpb2, Rpb3, Rpb9, and Rpb11, sequence alignments with their Pol I homologues were obtained with CLUSTAL W (Thompson et al., 1994) and were used for initial homology modeling. Side chains in these four Pol II subunits were kept when identical in the Pol I homologues, and otherwise replaced by the most common rotamer of the counterpart residues, using the rotamer library of the program O (Jones et al., 1991). Regions in Pol II subunits that were apparently not present in Pol I subunits were deleted. The resulting ten-subunit model was inspected 'residue by residue', and showed meaningful internal non-polar contacts and salt bridges in most regions. Several regions however showed steric clashes or disallowed contacts, indicating misalignment of the corresponding sequence stretches. Manual realignment of these weakly conserved stretches led to a model with good internal packing. The procedure was repeated several times until convergence (Fig. 23, Chapter II.7 and Chapter IV.7).

III.10 | Structure prediction of A49/34.5

The sequences of the two Pol I specific subunits A49 and A34.5 were sent to the HHpred server for remote protein homology detection and structure prediction(Soding et al., 2005) using default settings. For the highest scoring hit, HHpred predicted a structural similarity of the A49 N-terminal residues 52-102 to the N-terminal residues 99-150 of the large subunit of the human Pol II-associated factor TFIIF, RAP74 (P-value = 0.0023). For A34.5 the hit with the third highest score showed a similarity between the A34.5 residues 50-65 and residues 15-30 of the small subunit of TFIIF, RAP30 (P-value = 0.0003). Inspection of the predicted secondary structure elements in the apparent

regions of distant homology in A49 revealed a similar arrangement of strands as in the crystal structure of the dimerization module of RAP74/RAP30 (Gaiser et al., 2000) (PDB 1F3U) except that the two strands $\beta4$ and $\beta5$ are apparently lacking in A49, and no secondary structure corresponding to the strands $\beta6$ and $\beta7$ of RAP30 was predicted in A34.5. Strikingly, the few residues that are conserved between A49 and RAP74 and between A34.5 and RAP30 are generally part of the hydrophobic core of the heterodimer interface. Mutations of these residues led to strongly impaired co-purification of the A49/34.5 heterodimer (Fig. 28B-D, Chapter II.9).

III.11 | Purification of recombinant A49/34.5

Buffers used during purification:

A49-A
300 mM NaCl
50 mM Tris, pH 7.5
10 mM β-mercaptoethanol[1]
1x PI[1]

A49-highsalt
1 M NaCl
50 mM Tris pH 7.5
10 mM β-mercaptoethanol[1]

A49-dilution
50 mM Tris pH 7.5
10 mM β-mercaptoethanol[1]

A49-B
100 mM NaCl
50 mM Tris pH 7.5
5 mM DTT[1]

[1] added prior to usage

The genes for A49 and A34.5 were amplified from yeast genomic DNA by PCR and were cloned into vector pET28b (Novagen), resulting in a C-terminal hexahistidine tag on A49 and introducing a second ribosomal binding site for bicistronic expression. The two subunits were co-expressed for 18 hours at 18

°C in E. coli BL21 (DE3) RIL cells (Stratagene) in 4 L of LB medium (Sambrook and Russel, 2001). Cells were harvested by centrifugation, resuspended in 100 mL buffer A49-A and lysed by sonication. After centrifugation the supernatant was loaded onto a 3 mL Ni-NTA column (Qiagen) equilibrated with buffer A49-A. The column was washed stepwise with 15 mL of buffer A49-A, 15 mL of A49-highsalt buffer and 15 mL of buffer A49-A containing 30 mM imidazole. The A49/34.5 heterodimer was eluted with buffer A49-A containing 100 mM imidazole. Eluted fractions were diluted three-fold with A49-dilution buffer, and further purified by cation exchange chromatography (MonoS 10/100 GL, GE Healthcare). The MonoS column was equilibrated with buffer A49-B and proteins were eluted with a linear gradient of 18 CV 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 A49-B (Fig. 28A, Chapter II.9). Pooled peak fractions were concentrated to 1 mg/mL and glycerol was added to a final concentration of 10% (v/v). Protein aliquots were flash-frozen in liquid nitrogen and stored at -80 °C.

III.12 | Yeast genetics

III.12.1 | 6-azauracil phenotyping of GPY2 ΔRPA34

To disrupt the gene coding for A34.5, *His5*[+] from *S. pombe* (complementing *HIS3* from *S. cerevisiae*) was amplified from pFA6a-His3MX6 (Longtine et al., 1998) using PCR (Primer A: 5'AGTGAGCAGCTAGGATTCAATAAACGGGA TTAACAAAAAATTGATAGATCTGTTTAGCTTGCCTC-3'; Primer B: 5'CACA TTTTTATCTT ATGTTACACACAGTTATACGCACATACGCATGAATTCGAGCT CGTTTAAAC-3'). *S. cerevisiae* strain GPY2 was transformed by the LiAc-method (Kaiser et al., 2004), positive clones were selected using –His plates and verified by colony PCR. For testing elongation activity, GPY2 and GPY2*ΔRPA34*, both harboring the pRS316 plasmid, were spotted onto SDC

plates lacking uracil and containing 60 µg/mL 6-azauracil. Growth was monitored after 2-3 days at 30 °C.

III.12.2 | Cloning and fermentation of GPY2 RPA12ΔC

To generate a C-terminal deletion in A12.2 (ΔG79-N125), A12.2 was deleted in GPY2 essentially like described in Chapter III.12.1, using *KanMX* instead of *His5+* as genetic marker. The resulting strain GPY2 (*rpa12::KanMX*) was transformed with a plasmid (pRS313-*RPA12(aa1-78)*) coding for the N-terminus and the 'potential' linker region in A12.2 (residues 1-78). Transformed yeast cells (*rpa12::KanMX*(pRS313-*RPA12(aa1-78)*)) were selected on SDC plates – His and screened by colony PCR. A positive clone (Pol I A12.2ΔC) was grown to a maximum OD_{600} of ~ 3 in SDC medium lacking histidine (Kaiser et al., 2004) using a 20 L fermenter (Infors ISF) and following the same procedure as described in Chapter III.1. A total of 240 g of yeast pellet could be harvested from 50 L of yeast culture.

III.13 | *In vitro* RNA assays

III.13.1 | RNA extension assays using a minimal scaffold

4 pmol Pol I, Pol I ΔA49/34.5, 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 ACACAGTCAG-3'; RNA: 5'-FAM-UGCAUAAAGACCAGGC-3').

For complementing Pol I ΔA49/34.5, a fivefold molar excess of recombinant A49/34.5 was incubated with Pol I ΔA49/34.5 for 10 min at 20 °C, prior to forming the polymerase-scaffold complex. For RNA elongation, complexes were

incubated in the presence of 1 mM NTPs at 28 °C for 20 min in transcription buffer (60 mM ammonium sulfate, 20 mM HEPES pH 7.6, 8 mM magnesium sulfate, 10mM zinc chloride, 10% glycerol, 10 mM DTT). Reactions were stopped by addition of an equal volume (12 μL) 2x loading buffer (8 M urea, 2 x TBE) and incubation for 5 min at 95 °C. FAM-labeled RNA extension products were separated by denaturing gel electrophoresis (0.5 pmol RNA per lane, 0.4 mm 15-20% polyacrylamide gels containing 8 M urea, 50-55 °C) and visualized with a Typhoon 9400 phosphoimager (GE Healthcare). FAM was excited with blue light at λ = 488 nm and fluorescent signal was recorded with a 520 BP 40 band-pass filter.

III.13.2 | RNA extension assays using a complementary bubble

For RNA extension assays with a complementary bubble (Kireeva et al., 2000), 6 pmol Pol I or Pol I ΔA49/34.5 were incubated for 15 min at 20 °C with 3 pmol of a pre-annealed RNA-template DNA scaffold (template DNA: 3'-TGCGCACCACGCTTACTGGTCCGTTCGCCTGTCCTCGACCA-5'; RNA: 5'-FAM-UGCAUUUCGACCAGGC-3'). For complementing Pol I ΔA49/34.5, a fivefold molar excess of recombinant A49/34.5 (30 pmol) was incubated with Pol I ΔA49/34.5 for 15 min at 20 °C, prior to forming the polymerase-scaffold complex. Annealing to the RNA-template DNA scaffold was followed by incubation with a fivefold molar excess of non-template DNA (15 pmol; 5'-TTTTTACGCGTGGTGCGAATGACCAGGCAAGCGGACAGGAGCTGGT-3') for 15 min at 25 °C. Formed 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 in Chapter III.13.1.

III.13.3 | RNA cleavage assays

Complexes of complete Pol I, Pol I ΔA49/34.5, or Pol I A12.2ΔC were formed in transcription buffer with a nucleic acid scaffold that comprised an RNA with a 6-FAM fluorescent label at its 5'-end and a three-nucleotide non-complementary overhang at its 3'-end (template DNA: 3'-TTACTGGTCCTTTTTCATGAACTC GA-5'; non-template DNA: 5'-TAAGTACTTGAGCT-3'; RNA: 5'-FAM-UGCAUU UCGACCAGGACGU-3', overhanging nucleotides underlined). For RNA cleavage reactions, samples were incubated in transcription buffer up to 30 min at 28 °C. RNA species were revealed by electrophoresis and fluorescence detection as described in Chapter III.13.1.

III.13.4 | Electrophoretic mobility shift assay (EMSA)

10 pmol of the scaffold used for cleavage assays (Chapter III.13.3) was incubated with 10 or 15 pmol of Pol I, Pol I ΔA49/34.5 or Pol I A12.2ΔC for 30 min at 20 °C. Protein-bound scaffold was separated from unbound RNA on a native 6% TBE gel at 4 °C (0.5 x TBE as running buffer, 90V, 1 - 1.5 h). RNA was stained with 1:10,000 SYBR gold and visualized with a Typhoon scanner (Fig. 32A, Chapter II.10). SYBR gold was excited with blue light at λ = 488 nm, fluorescence was detected with a 555 BP 20 band-pass filter.

III.14 | Figure preparation

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

IV.1 | DENZO and SCALEPACK scripts

Data integration with DENZO (auto.inp):

```
[crystal rotx   37.199 roty  167.206 rotz   -8.634]
[crossfire y  0.001 x -0.007 xy -0.003]
title                 'ck290'
distance              550  [You can get the distance and wavelength]
wavelength            1.02290 [from the ASCII header of any image
file ]

[x beam y beam from the measured direct beam position at 550 mm]
x beam                108.0
y beam                109.5

air absorption length  2800  [good value for Se energies]
format   ccd unsupported-m225
goniostat aligment 0 0
goniostat single axis
monochromator         0.99

space group           c2   [ Use P1 if unknown ]
unit cell  614.24 302.58 252.83  90.00  97.51  90.00
mosaicity             0.8  [ an estimate at this point ]
weak level            5.0  [adjust value to eliminate bad peaks in
indexing]
box                   2.4 2.4
spot                  elliptical 0.35 0.35 0.0
background            elliptical 0.5 0.5 0.0
overlap               spot
profile fitting radius 20.0

raw data file
'/xtal/cr_lise2/kuhn/crystals/ck290/ck290_1_###.img' [ <== edit ]
film output file      'ck290_###.x'                  [ <== edit ]

oscillation           start -10 range 0.5
[start is phi value at image 001][range is phi width per image ]
sector                1 to 1   [ number of image, see ### in name
template ]

[fit x beam y beam cell crystal rotx roty rotz]
print statistics

longest vector        900 [ somewhat greater than longest expected
cell axis ]

peak search file      peaks.file
write predictions

resolution limits     80.0 4.5   [ <== edit ]
go
write predictions
go
```

Crystal parameter refinement during integration with DENZO (ref.dat):

```
start refinement

resolution limits 80.0 7.0
fix all
refine partiality
fit crystal rotx roty rotz
go go go
fit x beam y beam
go go go
fit cell
go go go
fit crossfire x y xy
go go go
fit cassette rotx roty
go go go go go go
fit distance
go go go

resolution limits 80.0 4.5
fix all
refine partiality
fit crystal rotx roty rotz
go go go
fit x beam y beam
go go go
fit crossfire x y xy
go go go
fit cassette rotx roty
go go go go
fit distance
go go

print profiles 3 3
list
calculate go
```

Data scaling with SCALEPACK (scale.inp):

```
[Output]
output file 'ck290.sca'
[Pretty standard stuff]
format denzo_ip
number of zones 10
estimated error
0.09   0.08   0.07   0.06   0.05
0.05   0.05   0.05   0.05   0.05
error scale factor 1.2
rejection probability 0.00005
write rejection file 0.9
scale restrain 0.02
b restrain 1.0
anomalous
[no merge original index] [<== edit for getting unmerged data, i.e.
for running SOLVE]
ignore overloads
@reject.1

[Crystal data]
space group C2
resolution 80 4.5
reference film 1

postrefine 10 [10 cycles of postrefinemet]
fit crystal a* 1 to 722
fit crystal b* 1 to 722
fit crystal c* 1 to 722
fit crystal beta* 1 to 722
fit film rotx 1 to 722
fit film roty 1 to 722
[fit batch rotz 5 to 50]
fit crystal mosaicity 1 to 60 81 to 140 173 to 334 361 to 510 551 to
634 635 to 722
[Mosaicity was fitted for each translation in this case]

add partials 1 to 60 81 to 140 173 to 334 361 to 510 551 to 634 635 to
722

[hkl matrix  0  0  1   0 -1  0   1  0  0] [<== edit for re-indexing]

sector 1 to 60
FILE 1 'ck290_###.x'

sector 81 to 140
FILE 81 'ck290_###.x'

sector 173 to 334
FILE 173 'ck290_###.x'

sector 361 to 510
FILE 361 'ck290_###.x'

sector 551 to 722
FILE 551 'ck290_###.x'
```

IV.2 | Self-rotation function scripts

Self-rotation calculation with POLARRFN (poalrrfn.com):

```
polarrfn HKLIN ../crank/ck209_FI.mtz \
MAPOUT ck209_polarrfn.map \
PLOT ck209_polarrfn.plt <<EOF
title selfrot
SELF 70.0
RESOLUTION 70 6.0
LABIN FILE 1 F=FP SIGF=SIGFP
CRYSTAL FILE 1
CRYSTAL ORTH 1
LIMITS 0 180 2 0 180 2 0 180 2
MAP
PLOT 30 5
!contour level to start - contour intervals
FIND 30 50 OUTPUT selfrotpeaks.list
!threshold for peaks - peaks to find
NOPRINT
EOF
```

Self-rotation calculation with GLRF (srf.inp):

```
title Pol1 ordinary self rotation function
!
print ck209_srf_polar.prt
!
polar xyk
euler zyz
orthog axabz
!
!locsymmetry 1 0 0 7 polar
!locsymmetry 0 1 0 2 polar
!locexpand true
!
acell 619.346 305.423 251.200 90.000 97.488 90.000
asymmetry c2
aobsfile ../ck209_noanom_noheader.sca
acutoff 1.0 1.0 0.0
aformat 3I4, 2F8.0
apower 1
origin true
!
!cutoff 0.25
!
resolution 70.0 6.2
radius 60.0
boxsize 3 3 3
geval 2
!
self true
cross false
fast true
norm false
!
sangle polar
!rcut 1 20
slimit 1 0 180 2
slimit 2 0 180 2
slimit 3 0 180 2
oangle polar xyk
!
!mapfile ck209_srf_polar.map
peak 3 50
pkfit 10 1.5
!
cntfile ck209_srf_polar.ps
cntl 400 1000 20
!
stop
```

IV.3 | SOLVE scripts

W_{18} localization using Patterson methods (solve_SAD_W18.com):

```
#COMPUTER ENVIRONMENT
#
setenv SYMINFO /xray/programs/solve/solve-2.11/lib/syminfo.lib
setenv CCP4_OPEN UNKNOWN
setenv SOLVETMPDIR /var/tmp
setenv SYMOP /xray/programs/solve/solve-2.11/lib/symop.lib
setenv SYMINFO /xray/programs/solve/solve-2.11/lib/syminfo.lib
#
unlimit
#
/xray/programs/solve/solve-2.11/bin/solve_extra_huge <<EOD

#CRYSTAL INFORMATION
resolution 70 9.0
cell 613.56   302.42   248.73  90.000  97.473  90.000
symfile /xray/programs/solve/solve-2.11/lib/c2.sym

#INPUT DATA   ! input for external phase information, here from MR
#LABIN FP=FP SIGFP=SIGFP FPH1=FPH1 SIGFPH1=SIGFPH1
#LABIN DPH1=DPH1 SIGDPH1=SIGDPH1
#HKLIN ../molrep_input.mtz
#PHASES_LABIN FC=FC PHIC=PHIC FOM=FOM
#PHASES_MTZ ../molrep.mtz

readformatted          ! readformatted/readdenzo/readtrek
                         readccp4_unmerged
unmerged               ! premerged/unmerged
read_intensities       ! read_intensities/read_amplitudes
fixscattfactors        ! fixscattfactors/refscattfactors
rawnativefile          /home2/kuhn/crystals/ck209/xds/ck209_mod.ahkl

#PSEUDO-MIR INPUT FOR W18
derivative 1
label SAD data for wclu

newatomtype wclu
clus_aval 2903 5109.4 -1197.1 -5254.3
clus_bval 509.3 -37.8 849.4 108.5
clus_cval 184 30 1.2

clus_fp_aval 0.185886 0.453782 -0.10632 -0.466651
clus_fp_bval 509.3 -37.8 849.4 108.5
clus_fp_cval 184 30 1.2

clus_fpp_aval 0.185886 0.453782 -0.10632 -0.466651
clus_fpp_bval 509.3 -37.8 849.4 108.5
clus_fpp_cval 184 30 1.2

atom wclu
fprimv -6.753
fprprv 25
```

```
#xyz  0.2161  0.0000  0.3823    ! site from visual inspection of
                                     patterson map

rawderivfile /home2/kuhn/crystals/ck209/xds/ck209_mod.ahkl

anoonly
nsolsite 7                         ! number of sites per derivative
SCALE_NATIVE                       ! scale the native dataset
SCALE_MIR                          ! scale the derivs to the native
ANALYZE_MIR                        ! analyze this MIR data and set up
                                     for SOLVE
#addsolve                          ! look for more sites then refine and
                                     phase

SOLVE
EOD
```

$Ta_6Br_{12}^{2+}$ localization using Patterson methods (solve_SAD_TaBr.com):

```
#COMPUTER ENVIRONMENT
#
setenv SYMINFO /usr/local/lib/solve/syminfo.lib
setenv CCP4_OPEN UNKNOWN
setenv SOLVETMPDIR /var/tmp
setenv SYMOP /usr/local/lib/solve/symop.lib
setenv SYMINFO /usr/local/lib/solve/syminfo.lib
#
unlimit
#
/usr/local/xtal/solve-2.10/bin/solve_giant<<EOD

#CRYSTAL INFORMATION
resolution 50 7.5
cell 615.047 305.472 251.809  90.000  97.044  90.000
symfile /usr/local/lib/solve/c2.sym

#INPUT DATA
readdenzo          ! readformatted/readdenzo/readtrek
                     readccp4_unmerged
unmerged           ! premerged/ unmerged
read_intensities   ! read_intensities/read_amplitudes
fixscattfactors    ! fixscattfactors/refscattfactors
rawnativefile      /home2/kuhn/crystals/ck209/xds/ck209_mod.sca

#PSEUDO-MIR INPUT FOR TA6BR12 CLUSTER
derivative 1
label SAD data for TaBr

newatomtype tabr
clus_aval 795.88 -757.81 908.87 127.86
clus_bval 301.24 460.86 301.37 -24.026
```

```
clus_cval -237.92 1.0 4.0

clus_fp_aval 5.565 -5.299 6.356 0.894
clus_fp_bval 301.24 460.86 301.37 -24.026
clus_fp_cval -1.516 1.0 4.0

clus_fpp_aval 5.565 -5.299 6.356 0.894
clus_fpp_bval 301.24 460.86 301.37 -24.026
clus_fpp_cval -1.516 1.0 4.0

atom tabr
fprimv -17.398
fprprv 15.780

rawderivfile /home2/kuhn/crystals/ck209/xds/ck209_mod.sca

#SAD
anoonly
nsolsite_deriv 7          ! 7 atoms max
#addsolve
SCALE_NATIVE
SCALE_MIR
ANALYZE_MIR
SOLVE
EOD
```

IV.4 | MOLREP script

Molecular replacement using an EM map (molrep_rotation.com):

```
# ------------------------------
molrep <<stop
# ------------------------------
#
_DOC Y
_SCORE Y
#
_FILE_F input/ck290_dec06.mtz
#
_F  F
_SIGF  SIGF
_END      <--- end of MTZ block
#
_FILE_M input/val067f_300.map
_DSCALEM 1
_INVERM N
_DRAD 60
_ORIGIN 0.5,0.5,0.5
#
_RESMIN 80
_RESMAX 12
#
_FUN R
_NP 20
_FILE_T  rotations_val067_300.tab
_NCSM 1
_ANISO C
_RAD 70
_SIM 0.7
_COMPL 0.14
_NMON 7
_NPT 20
#self-rotation information
_LOCK Y
_NSRF 6
_FILE_TSR input/7fold.list
_CHI 52
_END
stop
```

IV.5 | Initial cryo-EM processing for 14-subunit Pol I

Image number	Particles / image	Defocus value (μm)	Defocus group
47	1483	0,83	1
56	991	0,99	1
68	1316	1,03	2
79	1385	1,12	3
45	1484	1,12	3
48	1278	1,12	3
71	1264	1,16	4
63	1376	1,17	4
78	1367	1,18	5
38	1236	1,20	5
76	1460	1,20	5
80	1116	1,21	5
33	1001	1,26	6
70	2181	1,28	6
28	1163	1,29	7
84	1480	1,29	7
73	1429	1,31	7
34	1056	1,31	7
82	939	1,35	8
75	2329	1,36	8
42	2243	1,37	8
7	2164	1,37	9
37	1859	1,37	9
29	1011	1,38	9
43	1489	1,40	9
11	1818	1,42	10
58	1705	1,42	10
54	1139	1,43	10
18	2154	1,45	11
50	1844	1,47	11
35	1370	1,52	12
64	1332	1,53	12

59	1474	1,57	13
40	1811	1,58	13
61	1245	1,61	14
49	1500	1,63	15
44	1722	1,64	15
69	1321	1,71	16
53	1308	1,72	16
52	1266	1,72	16
16	2379	1,74	17
57	1011	1,83	18
4	1978	1,83	18
74	1436	1,86	19
25	2363	1,89	20
9	1327	1,91	20
14	2279	1,94	21
30	1325	2,00	22
24	2262	2,02	22
62	1541	2,09	23
3	1927	2,16	24
65	1944	2,19	25
15	2243	2,27	26
8	2195	2,37	27
55	1230	2,39	27
60	1957	2,41	27
6	2061	2,42	27
19	2400	2,48	28
12	2126	2,52	29

IV.6 | Initial cryo-EM processing for Pol I ΔA49/34.5

Image number	Particles/image	Defocus value (μm)	Defocus group
34	1495.0	3,02	1
46	1564.0	3,03	1
50	1582.0	3,07	2
22	1631.0	3,11	2
17	1641.0	3,31	3
33	1650.0	3,31	3
10	1709.0	3,35	3
43	1528.0	3,36	3
9	1663.0	3,44	4
49	1625.0	3,46	4
32	1552.0	3,60	5
18	1530.0	3,61	6
29	1552.0	3,61	6

IV.7 | Sequence alignments

Sequence alignments of Pol I subunits with their respective homologs in Pol II. Alignments were generated with CLUSTAL W (Thompson et al., 1994) and were then edited based on structural modeling. Regions of conserved fold are underlined. Additional regions of conserved fold likely exist but cannot be predicted with certainty.

A190-Rpb1 edited by hand according to 3D structure, EM density and secondary structure prediction

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

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

A190    VHRYACKLRLLQYGLIDESYKLDEITLGSLNSSMYTDDEAIEDNEDEMDGEGSKQSKDISS 175
Rpb1    LMRQALAIKDSKKRFAAIWTLCKTKMVCETDVPSEDDP---------------------- 158
          * * *                                 *

A190    TLLNELKSKRSEYVDMAIAKALSDGRTTERGSFTATVNDERKKLVHEFHKKLLSRGKCDN 235
Rpb1    ------------------------------------------------------------


A190    CGMFSPKFRKDGFTKIFETALNEKQITNNRVKGFIRQDMIKKQKQAKKLDGSNEASANDE 295
Rpb1    -----------------------------------------TQLVSRGGCGNTQPTI 174
                                                        *

A190    ESFDVGRNPTTRPKTGSTYILSTEVKNILDTVFRKEQCVLQYVFHSRPNLSRKLVKADSF 355
Rpb1    RKDGLKLVGSWKKDRATGDADEPELRVLSTEEILNIFKHISVKDFTSLGFNEVFSRPEWM 234
                                    *

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

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

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

A190    IQNEDGSLVSLIGMSVEQRKALANQLLTPSSNVSTHTLNKKVYRHIKNRDVVLMNRQPTL 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 1422
Rpb1   TEAEEDHMLKKIENTMLENITLR------------------------------------- 1274
                *


A190   IVEANNNMNKVQRDRQSAIISHHRFITKYNFDDESGKWCEFKLELAADTEKLLMVNIVEE 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
       * * * *  * *    *        *     *     *** ***
```

A135-Rpb2 edited by hand according to 3D structure, EM density and secondary structure prediction

```
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  LKYALATGNWGEQ-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 secondary structure prediction

```
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 secondary structure prediction

```
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 secondary structure prediction

```
A12.2    MSVVGSLIFCLDCGDLLENPNAVLG---SNVECSQCKAIYPKSQFSNLKVVTTTADDAFPSSLR 61
Rpb9     ---MTTFRFCRDCNNMLY-PREDKENNRLLFECRTCSYVEEAGSPLVYRHELITNIGETAGVVQ 60
                **  **    *   *       *    **  *                   *


A12.2    AKKSVVKTSLKKNELKDGATIKEKCPQCGNEEMNYHTLQLRSADEGATVFYTCTSCGYKFRTNN 125
Rpb9     DIGSDPTLPR----------SDRECPKCHSRENVFFQSQQRRKDTSMVLFFVCLSCSHIFTSDQ 114
            *                    ** *   *      * * *     * * **   *


A12.2    -------- 125
Rpb9     KNKRTQFS 122
```


```
A12.2 C-terminus       -----RAKKSVVKTSLKKNE---LKDGATIKEKCPQCGNEEMNYHTLQLR 32
TFIIS 3rd domain       PAPLKQKIEEIAKQNLYNAQGATIERSVTDRFTCGKCKEKKVSYYQLQTR 50
                            *  *            *      *  *        *  ** *


A12.2 C-terminus       SADEGATVFYTCTSCGYKFRTNN 65
TFIIS 3rd domain       SADEPLTTFCTCEACGNRWKFS- 72
                       ****   *  *  **   **
```

V | References

Alic, N., Ayoub, N., Landrieux, E., Favry, E., Baudouin-Cornu, P., Riva, M., and Carles, C. (2007). Selectivity and proofreading both contribute significantly to the fidelity of RNA polymerase III transcription. Proc Natl Acad Sci USA *104*, 10400-10405.

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 Nat Acad Sci USA *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. The Journal of Biological Chemistry *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.

Bartholomew, B., Durkovich, D., Kassavetis, G.A., and Geiduschek, E.P. (1993). Orientation and topography of RNA polymerase III in transcription complexes. Molecular and Cellular Biology *13*, 942-952.

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.

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.

Bergfors, T. (2003). Seeds to crystals. Journal of Structural Biology *142*, 66-76.

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. The EMBO Journal *21*, 4136-4144.

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.

Brewer, B.J., Lockshon, D., and Fangman, W.L. (1992). The arrest of replication forks in the rDNA of yeast occurs independently of transcription. Cell *71*, 267-276.

Bric, A., Radebaugh, C.A., and Paule, M.R. (2004). Photocross-linking of the RNA polymerase I preinitiation and immediate postinitiation complexes: implications for promoter recruitment. The Journal of Biological Chemistry *279*, 31259-31267.

Brueckner, F., Hennecke, U., Carell, T., and Cramer, P. (2007). CPD damage recognition by transcribing RNA polymerase II. Science *315*, 859-862.

CCP4 (1994). The CCP4 Suite: Programs for Protein Crystallography. Acta Cryst D *50*, 760-763.

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 initiation complex. Nature Structural & Molecular Biology *8*, 696-703.

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

Chen, J.Z., and Grigorieff, N. (2007). SIGNATURE: a single-particle selection system for molecular electron microscopy. Journal of Structural Biology *157*, 168-173.

Chung, W.H., Craighead, J.L., Chang, W.H., Ezeokonkwo, C., Bareket-Samish, A., Kornberg, R.D., and Asturias, F.J. (2003). RNA polymerase II/TFIIF structure and conserved organization of the initiation complex. Molecular Cell *12*, 1003-1013.

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

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.

Damsma, G.E., Alt, A., Brueckner, F., Carell, T., and Cramer, P. (2007). Mechanism of transcriptional stalling at cisplatin-damaged DNA. Nature Structural & Molecular Biology.

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.

Dawson, B. (1953). The structure of 9(18)-heteropolyanion in potassium 9(18)-tungstophosphate. Acta Crys *6*, 113-126.

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.

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 Harbor Symposia on Quantitative Biology *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.

Edwards, A.M., Kane, C.M., Young, R.A., and Kornberg, R.D. (1991). Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter *in vitro*. J Biol Chem *266*, 71-75.

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. Proceedings of the National Academy of Sciences of the United States of America *98*, 14334-14339.

Fatica, A., and Tollervey, D. (2002). Making ribosomes. Current Opinion in Cell Biology *14*, 313-318.

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. Molecular Cell *25*, 813-823.

Ferri, M.L., Peyroche, G., Siaut, M., Lefebvre, O., Carles, C., Conesa, C., and Sentenac, A. (2000). A novel subunit of yeast RNA polymerase III interacts with the TFIIB- related domain of TFIIIB70. Molecular and Cellular Biology *20*, 488-495.

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. Journal of Structural Biology *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.

Furey, W., and Swaminathan, S. (1997). PHASES-95: A Program Package for the Processing and Analysis of DIffraction Data from Macromolecules. Methods in Enzymology: Macromolecular Crystallography, Part B *277*, Chapter 31.

Gaiser, F., Tan, S., and Richmond, T.J. (2000). Novel dimerization fold of RAP30/RAP74 in human TFIIF at 1.7 A resolution. J Mol Biol *302*, 1119-1127.

Geiduschek, E.P., and Kassavetis, G.A. (2001). The RNA polymerase III transcription apparatus. J Mol Biol *310*, 1-26.

Gourse, R.L., de Boer, H.A., and Nomura, M. (1986). DNA determinants of rRNA synthesis in E. coli: growth rate dependent regulation, feedback inhibition, upstream activation, antitermination. Cell *44*, 197-205.

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. Human Molecular Genetics *16 Spec No 1*, R21-27.

Heix, J., and Grummt, I. (1995). Species specificity of transcription by RNA polymerase I. Current Opinion in Genetics & Development *5*, 652-656.

Hsieh, Y.J., Kundu, T.K., Wang, Z., Kovelman, R., and Roeder, R.G. (1999). The TFIIIC90 subunit of TFIIIC interacts with multiple components of the RNA polymerase III machinery and contains a histone-specific acetyltransferase activity. Molecular and Cellular Biology *19*, 7697-7704.

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.

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. The Journal of Biological Chemistry *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. Mol Gen Genet *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. Molecular Cell *23*, 71-81.

Johnson, T.L., and Chamberlin, M.J. (1994). Complexes of yeast RNA polymerase II and RNA are substrates for TFIIS-induced RNA cleavage. Cell *77*, 217-224.

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallographica *47 (Pt 2)*, 110-119.

Kaiser, C., Michaelis, S., and Michell, A. (2004). Methods in Yeast Genetics, 1994 Edition. Cold Spring Harbor Laboratoy Press.

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. The Journal of Biological Chemistry *273*, 33795-33802.

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

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

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

Kuhn, A., and Grummt, I. (1989). 3'-end formation of mouse pre-rRNA involves both transcription termination and a specific processing reaction. Genes & Development *3*, 224-231.

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 & Development *20*, 2030-2040.

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. The EMBO Journal *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.

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.

Leslie, A.G. (2006). The integration of macromolecular diffraction data. Acta Crystallogr D Biol Crystallogr *62*, 48-57.

Longtine, M.S., McKenzie, A., 3rd, Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., Philippsen, P., and Pringle, J.R. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast *14*, 953-961.

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., Storoni, L.C., and Read, R.J. (2005). Likelihood-enhanced fast translation functions. Acta Crystallogr D Biol Crystallogr *61*, 458-464.

Milkereit, P., Gadal, O., Podtelejnikov, A., Trumtel, S., Gas, N., Petfalski, E., Tollervey, D., Mann, M., Hurt, E., and Tschochner, H. (2001). Maturation and intranuclear transport of pre-ribosomes requires Noc proteins. Cell *105*, 499-509.

Milkereit, P., Schultz, P., and Tschochner, H. (1997). Resolution of RNA polymerase I into dimers and monomers and their function in transcription. Biol Chem *378*, 1433-1443.

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. The EMBO Journal *17*, 3692-3703.

Mindell, J.A., and Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. Journal of Structural Biology *142*, 334-347.

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.

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.

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 Harbor Symposia on Quantitative Biology *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.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray Diffraction Data in Oscillation Mode. Methods in Enzymology *276, Macromolecular Crystallography*, 307-326.

Panov, K.I., Friedrich, J.K., Russell, J., and Zomerdijk, J.C. (2006). UBF activates RNA polymerase I transcription by stimulating promoter escape. The EMBO Journal *25*, 3310-3322.

Pape, T., and Schneider, T.R. (2004). HKL2MAP: a graphical user interface for phasing with SHELX programs. Acta Crys D *37*, 843-844.

Penczek, P.A., Frank, J., and Spahn, C.M. (2006). A method of focused classification, based on the bootstrap 3D variance analysis, and its application to EF-G-dependent translocation. Journal of Structural Biology *154*, 184-194.

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. The EMBO Journal *19*, 5473-5482.

Pianese, G. (1896). Beitrag zur Histologie und Aetiologie der Carcinoma: Histologische und experimentelle Untersuchungen. Beitr Pathol Anat Allgem Pathol *142*, 1-193.

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. Proceedings of the National Academy of Sciences of the United States of America *101*, 6068-6073.

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

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

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

Sambrook, J.R., and Russel, D.W. (2001). Molecular Cloning: a laboratory manual. Cold Spring Harbor Laboratoy Press *3rd Edition*.

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.

Schneider, D.A., Michel, A., Sikes, M.L., Vu, L., Dodd, J.A., Salgia, S., Osheim, Y.N., Beyer, A.L., and Nomura, M. (2007). Transcription Elongation by RNA Polymerase I Is Linked to Efficient rRNA Processing and Ribosome Assembly. Molecular Cell *26*, 217-229.

Schneider, T.R., and Sheldrick, G.M. (2002). Substructure solution with SHELXD. Acta Crystallogr D Biol Crystallogr *58*, 1772-1779.

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.

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

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. Molecular 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., 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.

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

Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta Crystallogr D Biol Crystallogr *55*, 849-861.

Thiry, M., and Lafontaine, D.L. (2005). Birth of a nucleolus: the evolution of nucleolar compartments. Trends in Cell Biology *15*, 194-199.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res *22*, 4673-4680.

Thygesen, J., Weinstein, S., Franceschi, F., and Yonath, A. (1996). The suitability of multi-metal clusters for phasing in crystallography of large macromolecular assemblies. Structure *4*, 513-518.

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

Tschochner, H. (1996). A novel RNA polymerase I-dependent RNase activity that shortens nascent transcripts from the 3' end. Proceedings of the National Academy of Sciences of the United States of America *93*, 12914-12919.

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

Vagin, A., and Teplyakov, A. (1997). MOLREP: an Automated Program for Molecular Replacement. J Appl Cryst *30*, 1022-1025.

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.

Wagenknecht, T., Grassucci, R., and Frank, J. (1988). Electron microscopy and computer image averaging of ice-embedded large ribosomal subunits from Escherichia coli. J Mol Biol *199*, 137-147.

Wang, D., Bushnell, D.A., Westover, K.D., Kaplan, C.D., and Kornberg, R.D. (2006). Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. Cell *127*, 941-954.

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

Weeks, C.M., DeTitta, G.T., Hauptman, H.A., Thuman, P., and Miller, R. (1994). Structure solution by minimal-function phase refinement and Fourier filtering. II. Implementation and applications. Acta Crystallographica *50 (Pt 2)*, 210-220.

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.

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

White, R.J. (2005). RNA polymerases I and III, growth control and cancer. Nat Rev Mol Cell Biol *6*, 69-78.

Whitehall, S.K., Bardeleben, C., and Kassavetis, G.A. (1994). Hydrolytic cleavage of nascent RNA in RNA polymerase III ternary transcription complexes. The Journal of Biological Chemistry *269*, 2299-2306.

Wind, M., and Reines, D. (2000). Transcription elongation factor SII. Bioessays *22*, 327-336.

Yuan, X., Zhao, J., Zentgraf, H., Hoffmann-Rohrer, U., and Grummt, I. (2002). Multiple interactions between RNA polymerase I, TIF-IA and TAFI subunits regulate preinitiation complex assembly at the ribosomal gene promoter. EMBO Rep *3*, 1082-1087.

Zhao, J., Yuan, X., Frodin, M., and Grummt, I. (2003a). 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.

Zhao, Y., Sohn, J.H., and Warner, J.R. (2003b). Autoregulation in the biosynthesis of ribosomes. Molecular and Cellular Biology *23*, 699-707.

Curriculum vitae Claus-Dieter Kuhn

Date of birth: July 19, 1978

Place of birth: Mutlangen, Germany

Nationality: German

Family status: married, 1 son

Foreign languages: English, Swedish

Education

1989 - 1998	Grammar school in Kirchheim/Teck
July 1998 – April 1999	Military service at the Bundeswehr Hospital in Ulm, Division of Internal Medicine
1999 - 2002	Diploma Program in Biochemistry at the University of Regensburg
August 2002 – June 2003	Master Program at the University of Stockholm, Sweden; Department of Biochemistry and Biophysics
June 30, 2003	Degree of a Master of Science with a Major in Chemistry, University of Stockholm, Sweden
Since October 2003	PhD thesis at the Gene Center of the University of Munich, Research group Prof. Dr. Patrick Cramer

Stipends and Awards

1999 - 2003	Scholarship from the Wilhelm-Narr Fond, supporting young talented students
March 2004 – February 2006	PhD fellowship from the Federation of the German Chemical Industry (VCI)
Since 2004	Member of the International Doctorate Program Nano-Bio-Technology (part of the Elite Network Bavaria)
Since 2005	Member of the self-organized graduate program 'Protein Dynamics in Health and Disease' (part of the Elite Network Bavaria)
June 2007	Winner of the Poster Prize at RNA 2007
December 2007	Winner of the Römer Prize from the Department of Chemistry and Biochemistry, Ludwig-Maximilians-University Munich in the category dissertation