

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

# **Crystal structures of the complete 12-subunit RNA polymerase II and its subcomplex Rpb4/7, and modeling of RNA polymerases I and III**



Karim-Jean Armache  
aus Lodz, Polen  
2005

## **Erklärung**

Diese Dissertation wurde im Sinne von §13 Abs. 3 bzw. 4 der Promotionsordnung vom 29. Januar 1998 von Herrn Prof. Dr. Patrick Cramer betreut.

## **Ehrenwörtliche Versicherung**

Diese Dissertation wurde selbständig und ohne unerlaubte Hilfe erarbeitet.

München, den 31. März 2005

Karim-Jean Armache

Dissertation eingereicht am

1. Gutachter: Prof. Dr. Patrick Cramer

2. Gutachter: Prof. Dr. Michael Meisterernst

Mündliche Prüfung am 17-06-2005

## Acknowledgements

I would like to thank with a deep sense of gratitude Prof. Patrick Cramer for allowing me to carry out this research work in his group of macromolecular crystallography at the Gene Center, University of Munich. The excellent working conditions and atmosphere, his constant advice and support, were decisive in guiding this work.

Thanks to Hubert Kettenberger for the fruitful collaboration, which extended over last three years.

My special appreciation goes to Dr. Anton Meinhart who spent long hours on introducing me to complicated world of symmetry and gave me a lot of help and advice during solving the structure of Rpb4/7.

I am indebted to Claudia Buchen who introduced me to tricks and shortcuts of the biochemistry hard-core bench work which made it so much easier to carry out the technical part of my research.

I am mostly grateful to all the members of the lab especially Tomislav Kamenski, Sabine Hoepfner, Sonja Baumli and Erika Vojnic for their help, advice and support during various stages of this work.

I am also greatly thankful to Simone Mitterweger and Elisabeth Lehmann who were conducting excellent bachelor work in the lab, for all the help I got from them.

Many great thanks to Dr. Heidi Feldmann for scientific discussions and critical comments, which helped me to understand and solve many molecular biology problems, which I encountered during my work.

I would like to thank Prof. Karl-Peter Hopfner for all the help I got in crystallography and for being in my thesis committee.

I would like to thank also Prof. Ralf-Peter Jansen for critical comments on this thesis.

Thanks to my friends Luca Perabo and Karsten Beck who helped me on several occasions both socially and scientifically, *guys thanks for all the great parties we had and also all the advice I got from you.*

Many heartfelt thanks to Anja Fischer who supported me constantly in the last year and made it much easier to go through difficult times.

My family, Anna and Akram Armache, Jean-Paul Armache and Danuta Talar for being always with me and supporting me every single moment of my life: *Dziękuję wam kochani za wszystko co dla mnie zrobiliście, bez was nie byłoby tej pracy, bez waszej ciągłej pomocy przede wszystkim duchowej, ale także finansowej. Dziękuję Wam !!!*

# Table of contents

<b>Summary .....</b>	<b>1</b>
<b>Chapter I: Introduction .....</b>	<b>2</b>
Macromolecular assemblies .....	3
X-ray crystallography .....	4
RNA polymerase II .....	5
History of Pol II structure determination .....	5
General architecture of Pol II and comparison to bacterial RNA polymerase .....	9
The CTD .....	12
The subcomplex of Rpb4/7 .....	13
The mRNA transcription cycle .....	16
Initiation of transcription .....	17
<b>Chapter II: Architecture of initiation-competent 12-subunit RNA polymerase II .....</b>	<b>28</b>
Abstract .....	29
Introduction .....	30
Materials and methods .....	31
Fermentation and purification of 10-subunit Pol II .....	31
Expression and purification of full-length Rpb4/7 .....	35
Assembly of 12-subunit Pol II .....	36
Crystallization of 12-subunit Pol II .....	37
Results and Discussion .....	40
Fermentation of 10-subunit Pol II .....	40
Purification of 10-subunit Pol II .....	41
Expression and purification of full-length Rpb4/7 .....	43
Assembly of 12-subunit Pol II .....	46
Crystallization of 12-subunit Pol II .....	47
X-ray structure determination of 12-subunit Pol II .....	50
Overall structure .....	51
Rpb7 forms a conserved wedge that restrains the clamp .....	53
Implications for transcription initiation .....	54
Crystal packing in 12-subunit Pol II facilitates obtaining of higher order complexes .....	56
Conclusion .....	57
<b>Chapter III: Structures of complete RNA polymerase II and its subcomplex Rpb4/7 .....</b>	<b>58</b>
Abstract .....	59
Introduction .....	60
Materials and methods .....	61
Design of Rpb4/7 variants .....	61
Expression and purification of Rpb4/7 variants .....	63
Crystallization of Rpb4/7 variants .....	63
Results and discussion .....	64
Design of Rpb4/7 variants .....	64
Cloning of Rpb4/7 variants .....	67
Expression, purification, and crystallization of Rpb4/7 variants .....	71
Structure determinations .....	80

Rpb4/7 structure .....	82
Folding transitions upon Pol II core-Rpb4/7 interaction .....	82
Specificity of the Pol II core-Rpb4/7 interaction.....	86
Importance of the refined 12-subunit Pol II in phasing new structures .....	86
<b>Chapter IV: Modeling of Pol I and III based on the refined structure of 12-subunit Pol II.....</b>	<b>88</b>
Abstract .....	89
Introduction.....	90
Results and Discussion.....	91
Modeling of RNA polymerases I and III .....	91
Similarity of RNA polymerases and elongation mechanism.....	92
Initiation factor binding and promoter specificity .....	93
Specific assembly of Pol I, II, and III.....	97
Common subunits as molecular staples .....	98
Architectural principles .....	98
Supporting material:.....	102
Sequence alignments of homologous subunits in Pol I, Pol II and Pol III .....	102
<b>References .....</b>	<b>113</b>
<b>Curriculum Vitae .....</b>	<b>120</b>

## Publications

All parts of this work have been published or are in the process of publication:

- Armache K.-J., Kettenberger H., and Cramer P. (2003)  
**Architecture of initiation-competent 12-subunit RNA polymerase II.**  
*Proc. Natl. Acad. Sci USA* 100, 6964-6968
- Kettenberger H., Armache K.-J., and Cramer P. (2003)  
**Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage.**  
*Cell* 114, 347-357
- Kettenberger H\*., Armache K.-J\*., and Cramer P. (2004)  
**Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS.**  
*Mol. Cell* 16, 955-965  
\*(These authors contributed equally to this work)
- Armache K.-J., Mitterweger S., Meinhart A., and Cramer P. (2005)  
**Structures of complete RNA polymerase II and its subcomplex, Rpb4/7.**  
*J. Biol. Chem.*, 280(8):7131-4
- Armache K.-J., Kettenberger H., and Cramer P. (2005)  
**The dynamic mRNA elongation machinery**  
*Curr. Op. Struct. Biol.*; in press
- Armache K.-J., Mitterweger S., and Cramer P. (2005)  
**Modeling of RNA polymerases I and III based on refined structure of 12-subunit RNA polymerase.**  
Manuscript in preparation

## Summary

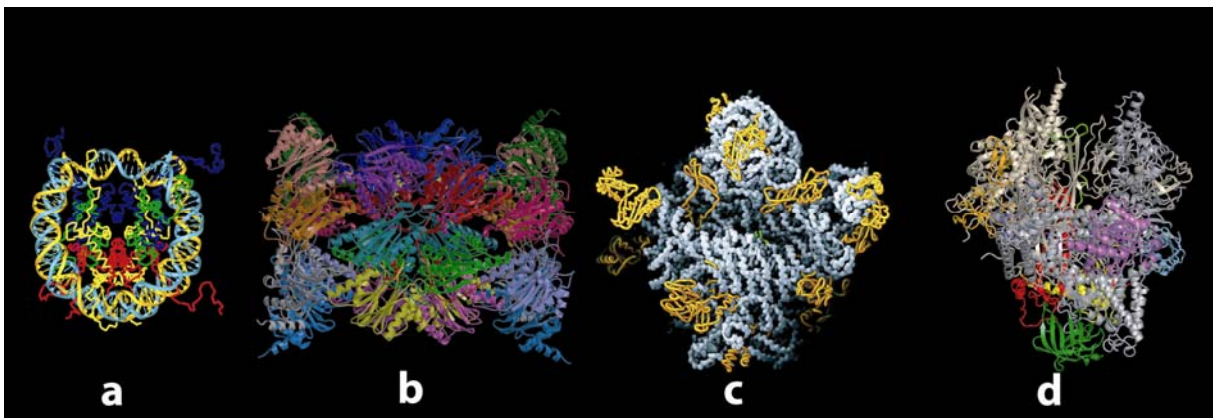
RNA polymerase II (Pol II) is the central enzyme, that synthesizes all mRNA in eukaryotic cells. In this work, I solved the structure of the complete, initiation-competent 12-subunit yeast RNA polymerase II at 3.8 Å. I also solved the structure of the Pol II subcomplex of Rpb4/7 alone at 2.3 Å resolution. These structures reveal the details of Pol II assembly from 12 subunits and give important insights into the initiation of transcription. The refined, atomic model of the complete 12-subunit Pol II enabled homology modeling of the two other nuclear RNA polymerases. In Pol I and Pol III, 65 % and 77 % of the Pol II fold are conserved, respectively. Together with a recent structure of a Pol II elongation complex, these results show that the basic mechanism of transcription applies also to the two other nuclear RNA polymerases.

# Chapter I

## *Introduction*

## Macromolecular assemblies

Most cellular processes are carried out by assemblies of proteins and not by freely diffusing and occasionally colliding proteins. These complexes are organized into networks, which cover the whole proteome, and are its functional units (Gavin et al., 2002). Specific complexes are referred to as molecular machines because of their modularity, complexity, cyclic functions and energy consumption (Nogales and Grigorieff, 2001) (Figure 1). There are currently about 12,000 known structures of assemblies from different organisms, involving two or more protein chains, and these complexes can be organized into around 3,500 groups based on sequence similarity (Sali et al., 2003). The most comprehensive information about transient and stable (cohesive) complexes exists for yeast, which has ~6,200 genes. It is estimated that in yeast around 11,000 protein-protein binary interactions occur, which corresponds to ~9 protein partners per protein (Bork et al., 2002).



**Figure 1. Macromolecular assemblies.** The figure shows four examples of macromolecular assemblies for which high resolution structures exist. In (a) the nucleosome core particle (Davey et al., 2002), in (b) 20S yeast proteasome (Groll et al., 1997), in (c) 50S subunit from the archaean *Haloarcula marismortui* (Ban et al., 2000), and in (d) structure of 10-subunit Pol II (Cramer et al., 2001). The figures do not represent the real size proportions between the structures.

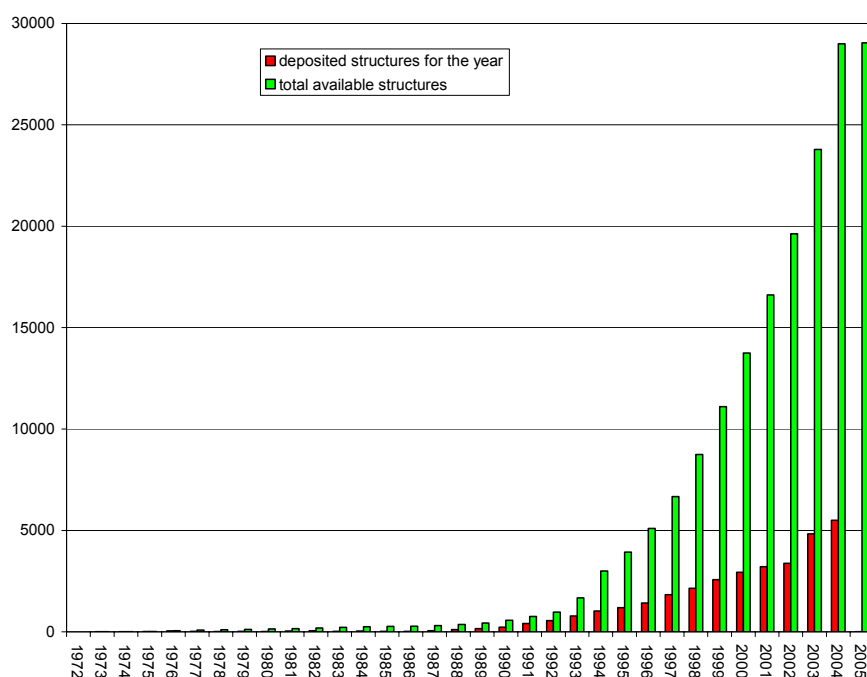
However, our knowledge of this multitude of interactions between individual proteins and macromolecular assemblies is not sufficient to describe their biochemical and cellular functions mechanistically. Such mechanistic insights can be obtained by structural characterization with a variety of techniques and approaches.

## Methods for structural characterization of macromolecular assemblies

Methods for structural characterization of protein ensembles vary in efficiency, accuracy, resolution and completeness (Russell et al., 2004). We can group them into genetic, biochemical or molecular biology methods, structural or biophysical methods, low and high resolution methods and last but not least experimental or computer-aided methods. The choice of the method is dictated by the question we want to answer, the amount of material that is required for the experiment, and several other factors.

## X-ray crystallography

This method is the most powerful method for structure determination of complexes, because it yields structures of macromolecular assemblies at atomic and near atomic resolution. The technique is based on irradiation of the crystal with x-rays and obtaining patterns produced by the diffraction of these x-rays through the closely spaced lattice of atoms in a crystal. The patterns are recorded and then analyzed to reveal the nature of that lattice. The spacings in the crystal lattice can be determined using Bragg's law. The electrons in the atoms are the entities that physically interact with the incoming x-ray photons to diffract them and so the electron density can be calculated and attributed to protein. Currently (March, 2005) there are 29,956 structures deposited to the Protein Data Bank (PDB) out of which 25544 were solved by x-ray crystallography and the remaining 4,412 by nuclear magnetic resonance (NMR) spectroscopy. Out of these deposited structures, 27,283 are of proteins, peptides and viruses, 1,242 of protein/nucleic acid complexes, 1,418 of nucleic acids and 13 of carbohydrates. Over the last decades, x-ray crystallography developed rapidly. In 1972 there were 2 structures deposited in PDB database, in 1990: 568 and in 2004: 5,501 (Figure 2).



**Figure 2. The growth of Protein Databank (PDB) from 1972-2005.**  
([www.rcsb.org/pdb](http://www.rcsb.org/pdb))

## **RNA polymerase II**

Two major steps have to be made in every cell that the protein is produced, first the genetic information, which is encoded in the DNA, has to be copied (transcribed) to mRNA, and second, this mRNA has to be translated into the protein. The process of transcription is carried out by multisubunit RNA polymerases. Bacteria and archaea have one, whereas eukaryotes have three essential, nuclear RNA polymerases, which were discovered between 1965-1970 by Robert Roeder (Roeder and Rutter, 1969a and b, 1970). RNA Pol I transcribes 18S and 28S rRNA, RNA Pol III snRNAs, tRNAs and 5.7S rRNAs and RNA Pol II transcribes some snRNAs and all mRNAs. Pol II consists of 12 subunits of total of 513 kDa. Sequence analyses showed that two big subunits are highly conserved and are the only polypeptides present in all types of cellular RNA polymerases (Allison et al., 1985). The high conservation of these subunits reflects the function, various biochemical (Cho and Kimball, 1982, Carrol and Stollar, 1983; Riva et al., 1987) and later structural studies (Gnatt et al, 2001; Cramer et al., 2001) have shown that they contain the catalytic center responsible for DNA-directed RNA transcription. It was also shown experimentally that several human Pol II subunits could substitute for their homologues in yeast (Khazak et al., 1995; McKune et al., 1995; Shpakovski et al., 1995). This suggested a high degree of both functional and structural conservation in eukaryotes and allowed for the assumption that results obtained in a yeast system can be applied to all eukaryotic RNA polymerases (Weinzierl, 1999).

Yeast Pol II is so far the best characterized out of three RNA polymerases, both biochemically and structurally. Pol II is associating with general transcription factors, coactivators (e.g. Mediator), elongation factors, processing factors, RNA export factors and others (Woychik and Hampsey, 2002; Hahn, 2004 and references therein). The regulation of Pol II underlies gene transcription and expression and therefore organisms' complexity and diversity (Levine and Tjian, 2003). To understand how this complex assembly functions, and to get insight into mechanistic properties, a major effort was put into structural studies of Pol II.

### **History of Pol II structure determination**

The history of structure determination of Pol II can be divided into few phases, which are characterized by breakthroughs. The 1<sup>st</sup> breakthrough came with the refinement of the protein purification protocol for obtaining sufficient amount of pure Pol II and application of a general method of growing single-layer crystals (Uzgiris and Kornberg, 1983). The key step for obtaining large amounts of pure Pol II was developing an immunoaffinity method based on monoclonal antibodies, (designated 8WG16) reacting with the highly conserved,

unphosphorylated heptapeptide repeat of C-terminal domain of Rpb1, the largest Pol II subunit. These antibodies allow for elution of Pol II with a gradient of polyols (Thompson et al., 1989). After successful trials with wheat germ and calf thymus Pol II, yeast Pol II was purified with the use of this method as well. The amounts and purity of prepared enzyme were much higher in comparison to older protocols that were based on several chromatography steps. The method of growing 2D crystals on charged lipid layers proved to work for Pol II. 2D crystals were analyzed by electron microscopy (EM), resulting in a resolution of 30 Å (Edwards et al., 1990). This resolution was only suitable to determine the shape of the molecule but it was a starting point for further trials. The small amounts of required material and the simplicity of the method for obtaining 2D crystals allowed for screening for larger and better diffracting crystals. It became apparent that polymerase preparations resulted in heterogeneous material, consisting of 10-subunit core enzyme and varying, substoichiometric amounts of two subunits Rpb4 and Rpb7, which form a subcomplex.

The 2<sup>nd</sup> breakthrough came with creating the deletion yeast strain (designated  $\Delta 4/7$ ) with a knockout of the gene coding for Rpb4. Polymerase purified from this deletion strain lacked Rpb7 as well. Preparation resulted in homogeneous 10-subunit core enzyme, which crystallized readily, forming 2D crystals of good shape and quality (Darst et al., 1991a). A surface representation of  $\Delta 4/7$  Pol II was derived from stained 2D crystals at 16 Å (Darst et al., 1991b and Figure 3a). This structure revealed the general shape of Pol II. Discovery of a crystal form in which the density for the part of the structure was absent, due to disorder or motion was the first proof of possible multiple conformations of Pol II. This mobile element designated as clamp was rendering Pol II in open and close conformation and was in agreement with the structural features of *E.coli* RNA polymerase (Asturias et al., 1997; 1998). 2D crystallography identified a possible location on the Pol II surface where the CTD emanates to solution (Meredith et al., 1996). This period of studies of Pol II with the use of 2D crystallography resulted in several complexes of Pol II with general transcription factors and nucleic acids with resolution limited to around 15 Å (e.g. Leuther et al., 1996).

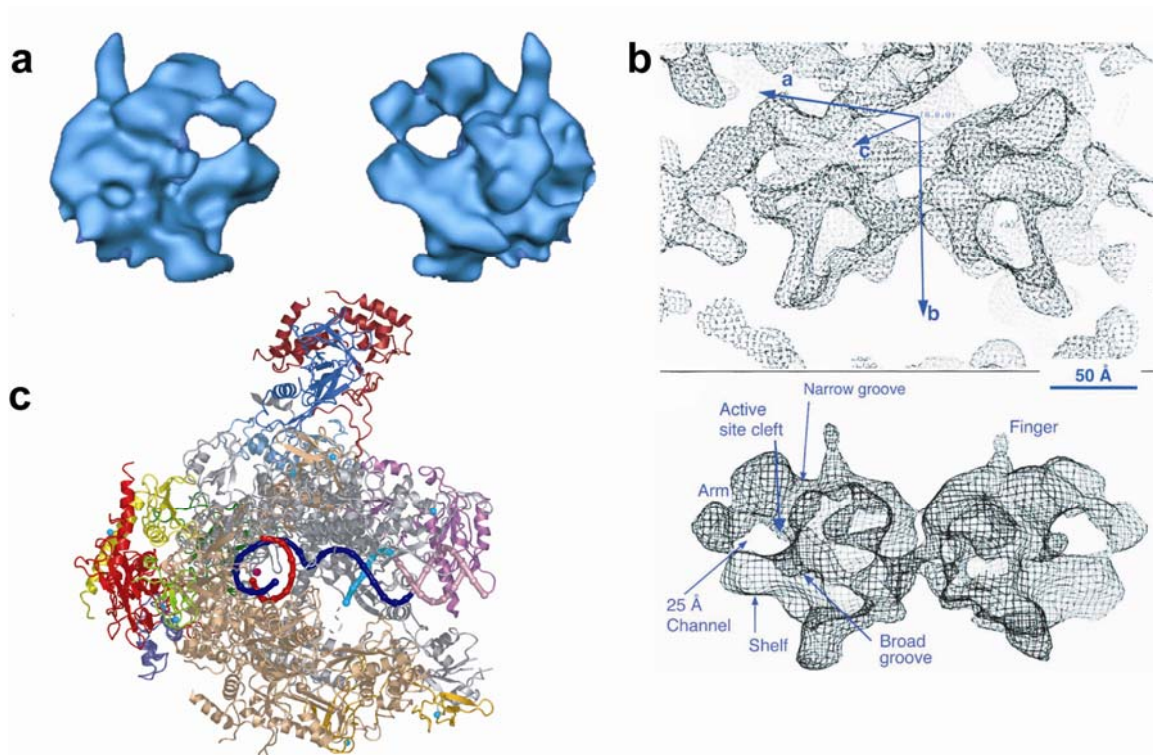
The 3<sup>rd</sup> breakthrough came with obtaining 3D crystals of 10-subunit Pol II, first by seeding with 2D crystals (Edwards et al., 1994) and then by conventional methods. Initial x-ray analysis of these crystals failed due to problems with oxidation. With the inert atmosphere and better crystals, first diffraction was measured to 3.5 Å (Fu et al., 1999). For initial phase determination 18-tungsten atom cluster was used, and the density calculated at 6 Å resolution showed a general good fit to the 16 Å molecular envelope obtained by electron crystallography (Figure 3b), except for the regions that were thought to be mobile and clamp

the DNA (Darst et al., 1991b). It proved to be very hard to obtain crystals diffracting reproducibly to  $\sim 3.5$  Å, they were nonisomorphous, with unit cell deviations of up to 10 Å and heavy metal incorporation destroyed the diffraction.

The search for new crystallization buffers, lacking phosphate, led to the 4<sup>th</sup> breakthrough when the soaking protocol which shrank the crystals of 10-subunit Pol II to minimal fixed unit cell was introduced (Cramer et al., 2000). These crystals diffracted isotropically to around 3.0 Å. Unfortunately, it was not possible to use previously obtained phases (Fu et al., 1999) because the new crystals were nonisomorphous with the old ones. The initial phase determination was based on MAD with 6-tantalum atom cluster and these phases allowed for location of other heavy atom derivatives with the use of cross-difference Fourier. Phases were obtained from 10 datasets ranging from 4.0 to 3.1 Å by the multiple isomorphous replacement with anomalous scattering (MIRAS) method. The compounds used were non-standard and contained mercury, rhenium or iridium. The electron density had a good agreement with the molecular envelope derived at 6 Å and allowed for model building of the backbone of RNA Pol II. Complete structures of yeast Rpb5 and Rpb8 together with archaeal and bacterial homologues of Rpb3, Rpb6, Rpb9, Rpb10 and Rpb11, which were determined by x-ray crystallography and NMR, facilitated model building. Subunit placement was also aided with eight structural zinc ions that were detected in anomalous difference-Fourier maps. The rest of the model was built as polyalanine. This model revealed for the first time the subunit architecture of Pol II. The atomic model of 10-subunit Pol II came with further shrinkage of one unit cell axis by  $\sim 8$  Å, which improved the diffraction to 2.6 Å (Cramer et al., 2001). Differences between this crystal form and the previous one (Cramer et al., 2000) identified four mobile modules in Pol II. Sequence markers like selenomethionine facilitated model building. Maps were improved by phase combination and the atomic model was refined at 2.8 Å. These model phases allowed for structure determination of a complex of Pol II with nucleic acids and  $\alpha$ -amanitin, an inhibitor of Pol II transcription. (Gnatt et al., 2001; Bushnell et al., 2002). The structure of the nucleic acids complex revealed new features out of which one is the clamp closure upon nucleic acids binding at the active site of polymerase. The transcribing polymerase model was then used as search model for solving the structures of other Pol II complexes, described here and elsewhere (Armache et al., 2003; Bushnell and Kornberg, 2003; Kettenberger et al., 2003; Bushnell et al., 2004; Westover et al., 2004 a and b).

The 5<sup>th</sup> breakthrough came during this work with reconstitution of the 12-subunit Pol II and a new crystal form, which are described in details in Chapter II and Chapter III. These new

crystals diffract reproducibly to 4 Å and contain ~80 % of solvent. The packing allowed for soaking of elongation factor TFIIS to preformed 12-subunit Pol II crystals and structure determination at 3.8 Å. This was the first ever structure of Pol II in complex with a transcription factor at near-atomic resolution (Kettenberger et al., 2003). Also with improvement of the crystals of 12-subunit Pol II and structure determination of the subcomplex of Rpb4/7 alone, the atomic, refined model of complete 12-subunit Pol II was determined (this work). This model was essential for phasing of the complete RNA polymerase II elongation complex (Kettenberger et al., 2004 and Figure 3c).



**Figure 3. History of Pol II structure determination.** (a) 16 Å structure of Pol II (Darst et al., 1991); (b) 6 Å electron density map (Fu et al., 1999) being the first density obtained from 3D crystals in comparison with molecular envelope derived from 16 Å EM structure (Darst et al., 1991); (c) Elongation complex of 12-subunit Pol II (this work and Kettenberger et al., 2004)

### General architecture of Pol II and comparison to bacterial RNA polymerase

Bacterial, archaeal and eukaryotic DNA dependent RNA polymerases belong to a conserved family of enzymes designated “multisubunit RNAP family”. All of these enzymes consist of conserved subunits that are termed the “core” subunits. Eukaryotic RNA polymerases also have common subunits that are shared between Pol I, II, and III. Additionally archaeal and eukaryotic RNA polymerases have several specific subunits (Table 1).

**Table 1.** Multisubunit RNA polymerases.

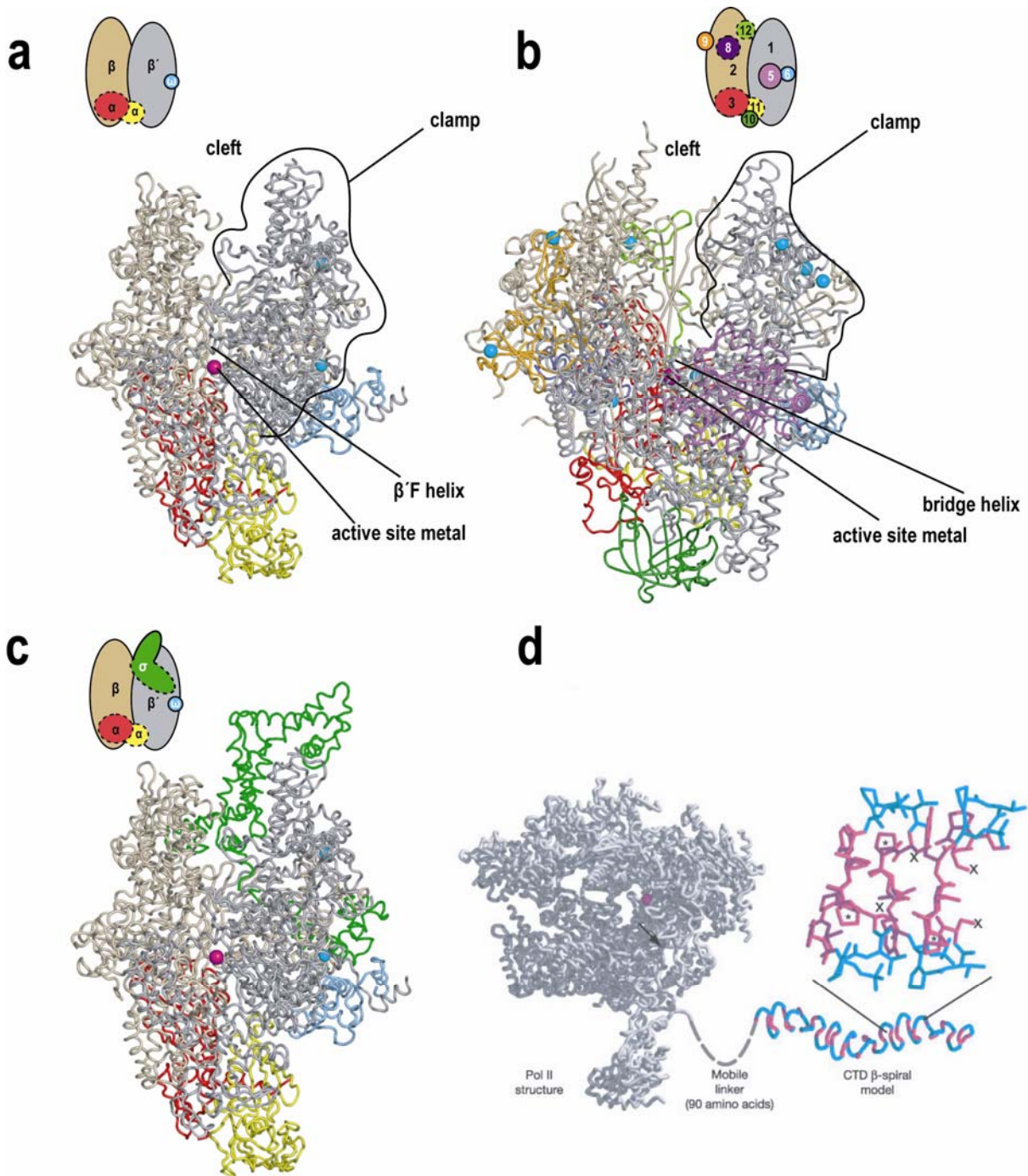
RNAP subunits					
Eukaryotes			Archaea	Bacteria	Class
Pol I	Pol II	Pol III			
A190	Rpb1	C160	A'+A''	β'	Core
A135	Rpb2	C129	B (B'+B'')	β	Core
AC40	Rpb3	AC40	D	α	Core
AC19	Rpb11	AC19	L	α	Core
Rpb6	Rpb6	Rpb6	K	ω	Core/common
Rpb5	Rpb5	Rpb5	H		Common
Rpb8	Rpb8	Rpb8	-		Common
Rpb10	Rpb10	Rpb10	N		Common
Rpb12	Rpb12	Rpb12	P		Common
(A12.2)	Rpb9	(C11)	X		
A14	Rpb4	C17	F		
A43	Rpb7	C25	E		
Specific subunits					
Pol I specific: A49, A34.5		Pol III specific: C82, C34, C31, C53, C37	Archaea specific: 1 more		

The atomic model of *S.cerevisiae* Pol II core (Cramer et al., 2001) and bacterial *T.aquaticus* and *T.thermophilus* RNA polymerase (Zhang et al., 1999; Campbell et al., 2001; Murakami et al., 2002a and b; Vassylyev et al., 2002) allowed for structural comparison of these eukaryotic and bacterial enzymes (Figure 4a and b).

The general architecture of Pol II is based on five core subunits, Rpb1, Rpb2, Rpb3, Rpb11 and Rpb6 in yeast and β', β, αα homodimer and ω in bacteria respectively. The two large subunits Rpb1/β' and Rpb2/β, form the central mass of the enzyme and lie on opposite sides of a positively charged cleft. The two large subunits are stabilized by the Rpb3-Rpb11

heterodimer in yeast and  $\alpha\alpha$  homodimer in bacteria. Rpb6/ $\omega$  further buttresses and stabilizes the large subunit. The sequential assembly of Pol II is based mainly on these “core” subunits (Chapter 4).

Comparison of two crystal forms of the Pol II core (Cramer et al., 2000, Cramer et al., 2001) revealed several mobile modules that underlie the dynamics of Pol II. On the Rpb1/ $\beta'$  side of the cleft the mobile “clamp” is located, feature which was already observed in EM studies (Asturias et al., 1997; 1998). The clamp was trapped in open and closed conformation and is believed to play an important role in initiation and transition from initiation to elongation (Chapter 2 and 3). On the other side of the cleft are situated two domains, “lobe” and “protrusion” which in bacterial polymerase are designated “ $\beta$  domain 1” and “ $\beta$  domain 2”. In both enzymes, the active site is located on the bottom of the cleft at the center of the enzyme. Before the active site, long, conserved “bridge helix” spans the cleft (in bacteria “ $\beta'$ F helix”), and below the “bridge helix” and active site “pore 1” or “secondary channel” is situated, which widens towards the bottom of the enzyme, creating a “funnel”. The rim of the pore has a loop of Rpb1, which binds  $Mg^{2+}$ , and a second catalytic metal could bind further in the pore. Beyond the active site, the cleft is blocked by a “wall” or “flap”. Other Pol II subunits are located on the outside of the enzyme, further stabilizing the assembly and constituting 12-subunit enzyme (Armache et al., 2003; Armache et al., 2005, this work).



**Figure 4. Structural comparison of eukaryotic and bacterial RNA polymerases.** (a) bacterial polymerase (Zhang et al., 1999); (b) yeast Pol II core enzyme (Cramer et al., 2000 and 2001); (c) initiation competent bacterial polymerase with  $\sigma$  (Murakami et al., 2002); (d) Pol II and proposed model of the CTD (Meinhart and Cramer, 2004).

## The CTD

The C-terminal domain of Rpb1 is a unique feature of RNA Pol II (Corden, 1993; Meininghaus et al., 2000), not present in any other RNA polymerase. It is essential for mRNA processing (Proudfoot et al., 2002), for Pol II response to enhancers (Scafe et al., 1990; Gerbe et al., 1995) and for organization of transcription foci in the nucleus (Misteli, 2000).

The CTD consists of seven amino acid repeats, whose number is species-dependent, 26 in yeast, 45 in flies and 52 in humans (Corden, 1990). The heptapeptide repeat, has the consensus sequence  $Y^1S^2P^3T^4S^5P^6S^7$  and out of these seven amino acids, five can be phosphorylated and it was shown that indeed in vivo phosphorylation is the major modification of the CTD (Bensaude et al., 1999). The phosphorylation status plays a major role in the transcription process (Dahmus, 1996) and several CTD-specific kinases and phosphatases were identified. CDK7 of TFIIF, CDK8 of Mediator and CDK of pTEFb are the CTD kinases. The phosphatases Fcp1 and Scp1 (Lin et al., 2002; Kamenski et al., 2004) as well as a number of phosphoserine/phosphothreonine phosphatases and Ssu72 (Meinhart et al., 2003b) were discovered.

It was shown that assembly of Pol II in the preinitiation complex and the interaction with Mediator are related to the unphosphorylated CTD. Upon CTD phosphorylation these interactions are being disrupted and the transition from initiation to elongation, and recruitment of pre-mRNA modifying enzymes occurs (Kobor and Greenblatt, 2002; Bentley, 2002). The variable phosphorylation status of the CTD could underlie the recruitment of different factors to Pol II. The CTD appears to be completely disordered in all Pol II structures. We only know where the linker to the CTD emerges from the Pol II surface and that in a  $\beta$ -strand conformation the CTD could span the distance of several hundreds of Å (Cramer et al., 2001). Structural studies of the CTD alone failed, indicating flexibility and a lack of secondary structure. The only structural information regarding the CTD comes from x-ray structures of CTD peptides bound to different factors such as Pcf11, a component of cleavage factor IA, a Pin1 domain, a peptidyl-prolyl isomerase, or Cgt1, a capping enzyme subunit (Meinhart and Cramer, 2004; Fabrega et al., 2003; Verdecia et al., 2000). In addition, a model of the CTD conformation was proposed (Figure 4d, Meinhart and Cramer, 2004)

**The subcomplex of Rpb4/7**

Pol II consists of 12 subunits out of which two, Rpb4 and Rpb7 form a subcomplex, which transiently interacts with the remainder, 10-subunit (core) enzyme. Whereas the 10-subunit core Pol II is sufficient for elongation and promoter-independent initiation of transcription (Edwards et al., 1991; Ruet et al., 1980; Rosenheck et al., 1998), it is not capable of promoter-directed initiation in vitro (Orlicky et al., 2001; Edwards et al., 1991), for this it needs association with the subcomplex Rpb4/7. These two subunits were identified in *S.cerevisiae* as 4<sup>th</sup> and 7<sup>th</sup> biggest subunits (Woychik and Young, 1989). The subcomplex associates with Pol II as a heterodimer (Armache et al., 2003; Bushnell and Kornberg, 2003; Edwards et al., 1991). Rpb7, like most Pol II subunits, is an essential protein (McKune et al., 1993; Mitsuzawa et al., 2003), and Rpb4 is dispensable under optimal growth conditions (Woychik and Young, 1989; Sheffer et al., 1999; Choder and Young, 1993; Farago et al., 2003; Miyao et al., 2001; Tan et al., 2000; Khazak et al., 1995).

It was very early discovered that Rpb4/7 is substoichiometric with respect to the Pol II core (Kolodziej et al., 1990), and can dissociate from the core enzyme (Edwards et al., 1991; Ruet et al., 1980). This was a major impediment for crystallization of wild-type Pol II in the early 90's, until the mutant  $\Delta$ Rpb4 strain was introduced from which stoichiometric 10-subunit polymerase core, lacking also Rpb7 could be purified. Under optimal conditions, this strain has the transcriptional capacity of a wild type strain (Choder and Young, 1993; Miyao et al., 2001; Maillet et al., 1999), whereas under suboptimal conditions transcription is abolished from most of the genes (Sheffer et al., 1999; Farago et al., 2003; Pillai et al., 2001). Although the  $\Delta$ Rpb4 strain is sensitive towards elevated temperature, ethanol and starvation, Rpb4 is not required for cell survival during oxidative or osmotic stress (Maillet et al., 1999) or nitrogen starvation (Pillai et al., 2001). The temperature-sensitive phenotype upon Rpb4 deletion can be partially suppressed by Rpb7 overexpression, which shows that Rpb7 can function to some extent independently of Rpb4 even under some moderate temperature stress (Sheffer et al., 1999; Tan et al., 2000; Maillet et al., 1999).

During exponential growth in *S.cerevisiae*, only 20 % of Pol II seems to contain Rpb4/7, whereas in stationary phase and during stress conditions 12-subunit Pol II becomes the most abundant form of the enzyme (Choder and Young, 1993). This situation seems to be different in *S.pombe* and human, where the homologues of the Rpb4/7 subcomplex are thought to be stoichiometrically associated with the core also in the exponential growth phase (Sakurai et al., 1999; Khazak et al., 1998). Structures of *S.cerevisiae* Rpb4/7 (this work, Armache et al.,

2005) and its archaeal homologue E/F subcomplex (Todone et al., 2001) were solved by x-ray crystallography at 2.3 Å and 1.7 Å respectively revealing the atomic details of the subcomplexes. Also x-ray structures of 12-subunit Pol II were determined (this work, Armache et al., 2003; Armache et al., 2005, Bushnell et al., 2003) showing Rpb4/7 in context of the whole enzyme. The structure of archaeal subcomplex of E/F showed two putative RNA binding domains harboring a conserved S1 motif, commonly seen in single-stranded nucleic acid binding proteins and forming continuous binding surface in RpoE. Experiments showed that both, RpoE and Rpa14 that is an Rpb7 homologue in the Pol I system bind RNA in vitro (Orlicky et al., 2001; Meka et al., 2003). Rpb4/7 physically interacts with TFIIF (Chung et al., 2003). In the Pol III system Rpc17, which is a homologue of Rpb4 interacts with Brf1, which is a counterpart of TFIIB (Ferri et al., 2000) and in Pol I system Rpa43, homologue of Rpb7 interacts with the general initiation factor Rrn3 (Yuan et al., 2002; Peyroche et al., 2002). Human Rpb7 was reported to interact with NOV, putative transcription factor (Perbal et al., 1999) and N-terminal portion of Edwin sarcoma oncogene (Zhou and Lee, 2001).

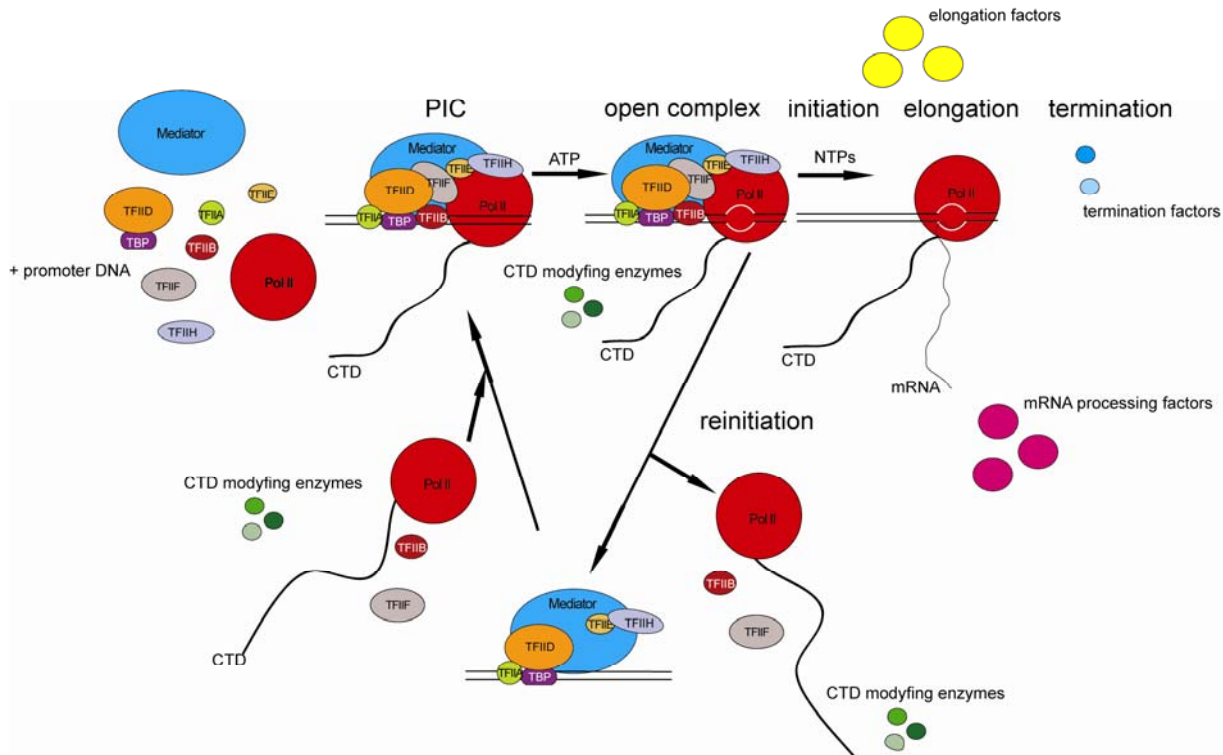
Regulated phosphorylation and dephosphorylation of the C-terminal domain of Rpb1 (CTD) plays a pivotal role in regulating Pol II transcription and the linker leading to CTD is positioned near Rpb4/7 (Armache et al., 2003). It was shown, that Rpb7 interacts genetically with ESS1, a prolyl isomerase that binds to the CTD and induces conformational changes, which affect transcription elongation and termination (Wu et al., 2003). Rpb4/7 was also shown to interact with Fcp1, a CTD phosphatase (Kamenski et al., 2004; Kimura et al., 2002). It is still not clear at which point of transcription initiation Rpb4/7 associates with the Pol II core. On the one hand 12-subunit Pol II has a higher affinity to TBP-TFIIB-promoter complex (Jensen et al., 1998), but on the other 10-subunit Pol II can also bind efficiently to a promoter in vitro (Orlicky et al., 2001). The subcomplex Rpb4/7 was also reported to participate in termination of transcription and co- and post-transcriptional processes such as mRNA export and DNA repair (Farago et al., 2003; He and Ramoter, 1999; Li and Smerdon, 2002). In *S.cerevisiae* and *S.pombe* Rpb4/7 is in excess to other Pol II subunits with Rpb4 being in 7 fold excess over Rpb7 in fission yeast. Sedimentation experiments showed that apart from Pol II, Rpb4 sediments also with small complexes (Rosenheck and Choder, 1998; Kimura et al., 2001) and it was proposed that Rpb4/7 could play additionally a role out of the Pol II context. Nrd1 in *S.cerevisiae* and its homologue in *S.pombe* are involved in RNA 3'-end formation of some genes transcribed by Pol II, and Rpb7 was shown to interact with these factors, pointing to possible Rpb4/7 involvement in mRNA processing and termination (Mitsuzawa et al., 2003). Rpb4 was shown to repress the Rpb9 mediated DNA repair pathway and to support the

Rad26 mediated DNA repair pathway (Li and Smerdon, 2002) and Rpb7 apparently plays a role in mediating repair in response to bleomycin (He and Ramoter, 1999). Rpb4/7 was shown to be localized in the nucleus (Farago et al., 2003), which is not really surprising as it is a part of Pol II but also in the cytoplasm, especially during stress conditions (Kimura et al., 2001; Khazak et al., 1995; Na et al., 2003) where it was shown to be required for efficient mRNA export (Farago et al., 2003).

Rpb4/7 is a component of transcription machinery and a great example of a protein taking part in several processes, which constitute the true interaction network underlying cell proliferation and development. It is an essential part of Pol II acts in a manner similar to general transcription factors.

### The mRNA transcription cycle

Generation of mRNA by Pol II involves multiple processes out of which some occur sequentially and some in parallel. The phases of transcript generation form transcription cycle, which can be divided into preinitiation, open complex formation, initiation, promoter clearance, elongation, termination and reinitiation. Figure 5 shows a schematic description of this process.



**Figure 5. Pol II transcription cycle.** (Adapted from Hahn, 2004)

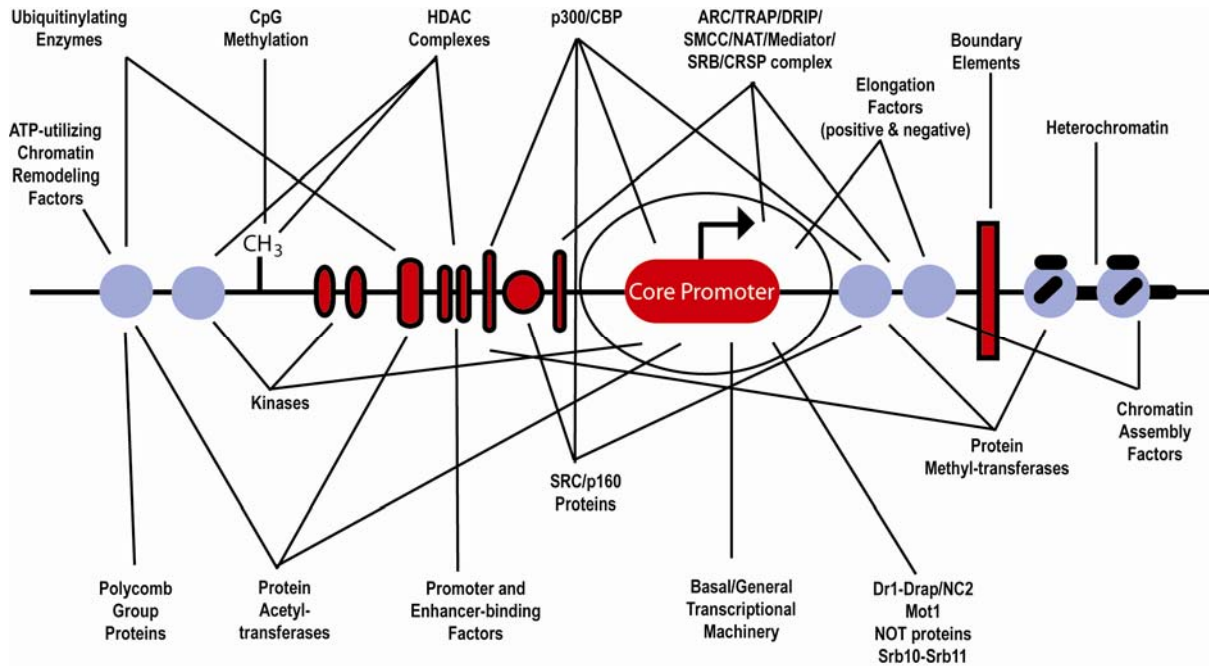
**Initiation of transcription**

12-subunit Pol II is the initiation-competent form of the enzyme but alone it cannot initiate transcription in a promoter-dependent manner. A breakthrough in understanding of the mechanism of transcription initiation came with the discovery that in a cell-free system Pol II can initiate transcription from template DNA only when supplemented with a crude extract (Weil et al. 1979; Manley et al., 1980; Dignam et al., 1983). Our understanding of the preinitiation complex (PIC) assembly came from studies of minimal set of factors required for promoter-directed transcription initiation (Van Dyke et al., 1988). In eukaryotes, these are termed general transcription factors (GTFs) and together with promoter DNA and 12-subunit RNA Pol II they constitute the so-called PIC. These and other studies established in vitro assembly pathway on the AdML promoter and identified intermediates that are stable to nuclease protection, electrophoretic mobility shift and template-challenge assays (Conaway and Conaway, 1993; Zawel and Reinberg, 1993; Buratowski et al., 1989). Intramolecular interactions and protein domains participating in them were further studied by binary interaction assays and mutagenesis techniques. These techniques were complemented by structural studies of independent components of PIC and intermediate complexes of the machinery (Figure 8). Crosslinking studies depicted position of GTFs on the promoter (Kim et al., 2000; Forget et al., 2003; Douziech et al., 2000 and Figure 6). General transcription factors enable Pol II to initiate the basal transcription but not the response to activators, for this Pol II requires additional entities like the mediator of transcription which is highly modular (Myers and Kornberg, 2000) and transmits the regulatory signals from various DNA-binding transcription factors to the Pol II initiation apparatus. Pol II general transcription factors are extremely conserved among species and Mediator complexes although variable in composition are structurally conserved in mammalian and yeast systems.

**Core promoter**

The core promoter can be understood as minimal stretch of contiguous DNA sequence that is sufficient to direct accurate initiation of the transcription machinery (Butler and Kadonaga, 2002). It usually encompasses the site of transcription initiation and extends upstream and downstream. In many cases, a core promoter comprises only ~40 nucleotides (nt). There are several commonly found sequence motifs with known specific functions in class II promoters, namely: TATA box, initiator (Inr), TFIIB recognition element (BRE) and downstream core promoter element (DPE). In addition to these core promoter elements there are cis-acting DNA sequences which regulate Pol II transcription and include the proximal promoter,

enhancers, silencers and boundary/insulator elements and can be located up to several kb from transcription start site and can either activate or repress transcription (West et al., 2002). These cis-acting elements can functionally interact with core promoter motifs. Core promoters are not only scaffolding DNA elements for Pol II and general transcription factors, but are also dynamic regulators of transcriptional activity (Figure 6).



**Figure 6. Core promoter interactions and beyond.** (Kadonaga, 2002)

The TATA box was the first identified core promoter element (Sassone-Corsi et al., 1981) and it is located 25-30 nt upstream of the transcription start site. In *S.cerevisiae* TATA box has position that is more variable and is located between 40-100nt upstream of transcription start site. Consensus sequence of TATA box is TATAAA but it can vary to various degree still promoting transcription (Singer et al., 1990).

This motif is very commonly found in class II promoters but there are a number of sequences not containing a TATA box (Kutach and Kadonaga, 2000). The TATA box is recognized predominantly by TBP or TBP-related factors ((TRFs); Berk, 2000)). The initiator element (Inr) encompasses the transcription start site (Breathnach and Chambon 1981) and the consensus sequence is TCA<sub>+1</sub>G/TTC/T (Hultmark et al., 1986) with A<sub>+1</sub> being the transcription start site. TFIID and precisely TAF<sub>II</sub>150 and TAF<sub>II</sub>250 were found to interact with Inr in a sequence-specific manner (Verrijzer et al., 1994). In addition, Pol II with TBP and TFIIF can initiate transcription in the absence of TAFs through recognition of the Inr (Carcamo et al., 1991). Downstream promoter element was found as a binding site for TFIID (Burke and Kadonaga 1996). TFIID binds cooperatively to DPE and Inr sequences and the

spacing between these two elements seems to be crucial (Kutach and Kadonaga, 2000). Mutation of only one nucleotide in the spacing reduces TFIID binding dramatically and hence reduces basal transcription (Burke and Katonaga, 1997). DPE is located precisely +28 to +32 relative to A<sub>+1</sub> transcription start site. DPE is found mainly on TATA-less promoters and does not interact with TBP. The TFIIB recognition element (BRE) is located in many cases directly upstream of some TATA boxes and facilitates incorporation of TFIIB into productive PICs (Lagrange et al., 1998). Interaction with TFIIB was visualized by x-ray crystallography (Tsai and Sigler, 2000). The sequence of BRE is GC-rich in eukaryotes in contrast to archaea, which also have BRE element, but it is rather AT-rich (Qureshi and Jackson, 1998). As mentioned above, the core promoter is also a scaffold for general transcription machinery and was studied extensively with respect to interactions with Pol II and GTFs. Figure 7 shows current understanding of these interactions based on DNA-protein crosslinks.

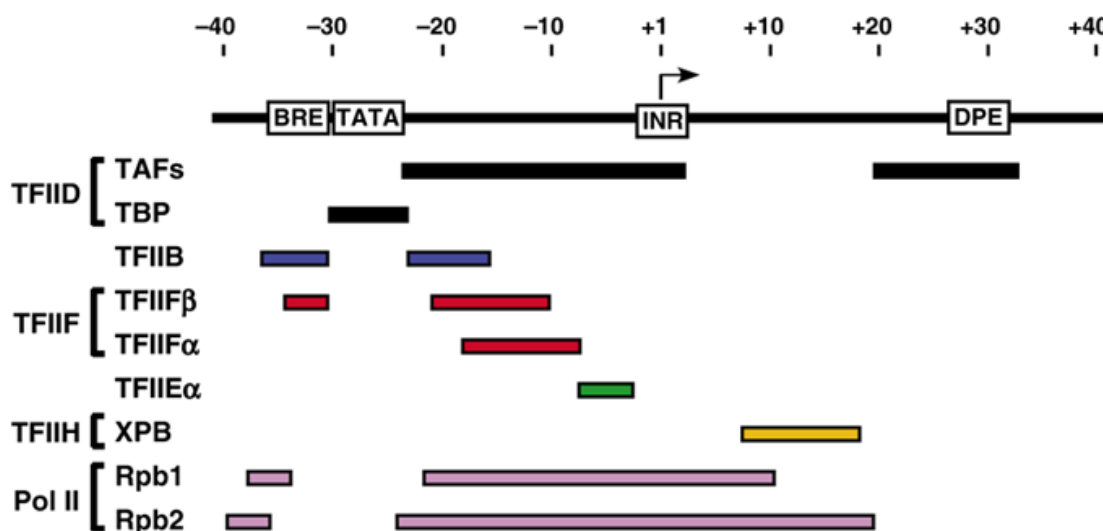


Figure 7. Interaction of initiation machinery with class II promoter. (Hahn, 2004)

## General transcription factors

### TFIIA

TFIIA, in early days sometimes referred to as AB or STF, was initially discovered as an interaction partner of TFIID (Matsui et al., 1980; Samuels et al., 1982). It is usually discussed when referring to basal transcription although substitution of TBP by intact TFIID complex eliminated requirement of TFIIA for transcription in vitro (Conaway and Conaway, 1993; Zawel and Reinberg, 1993). TFIIA becomes essential in case of non- canonical TATA promoters and basal transcription with TFIID. It is important in PIC assembly in response to some activators (Stargell et al., 2000) and TFIIA depletion abolishes response to activators in

yeast (Chou et al., 1999). TFIIA enhances TBP binding to its target site in vitro (Weideman et al., 1997; Yokomori et al., 1994). This stabilization was explained with x-ray structures of TFIIA/TBP/DNA (Tan et al., 1996; Geiger et al., 1996; Bleichenbacher et al., 2003). Due to interaction with TBP, TFIIA stabilizes also TFIID interaction with DNA and plays a major role in displacement of a TAF domain, that inhibits formation of PIC (Kokubo et al., 1998). It also derepresses basal transcription by counteracting various inhibitors of preinitiation complex assembly (Chicca et al., 1998; Mermelstein et al., 1996; Merino et al., 1993). It is indispensable for efficient transcription from snRNA promoters in vitro. Yeast TFIIA consists of two subunits, the large subunit TOA1 and small subunit TOA2, whereas in humans there are three subunits, TFIIA $\alpha$ , TFIIA $\beta$  and TFIIA $\gamma$ . These subunits have very high structural homology, share identical topology and can be superimposed with root mean square deviation (rmsd) of 0.99 Å (Bleichenbacher et al., 2003). The structure of TFIIA shows a  $\beta$ -barrel and four-helix bundle with  $\beta$ -barrel interacting with N-terminal portion of TBP and making several hydrogen bonds to DNA upstream of TATA box (Figure 8c).

### **TFIID**

TFIID, in the early literature also known as DB, BTF1 and D (Matsui et al., 1980; Samuels et al., 1982), consists of TATA-binding protein ((TBP) (Hahn et al., 1989; Horikoshi et al., 1989; Schmidt et al., 1989)) and around 14 TBP-associated factors ((TAFs) (Lewin 1990; Ptashne and Gann, 1990)). TBP is a universal factor conserved in archaeal and eukaryotic Pol I, Pol II and Pol III systems, and although TAFs differ in sizes between yeast, Drosophila and human, conserved domains are homologous reflecting evolutionary conservation. Function of TBP is specific recognition of TATA core element and directing the PIC assembly. Structures of TBP alone from different organisms were solved by x-ray crystallography. TBP has two pseudosymmetric domains bent in the middle, creating a saddle (Chasman et al., 1993; DeDecker et al., 1996; Nikolov and Burley, 1994 Nikolov et al., 1992; Koike et al., 2004), which binds the TATA element (Juo et al., 1996, Kim et al., 1993a, 1993b; Nikolov et al., 1996; Patikoglou et al., 1999). This platform created by TBP and DNA recruits then TFIIB which binds the binary complex with its two cyclin-A-like domains (Kosa et al., 1997; Nikolov et al., 1995; Littlefield et al., 1999; Tsai and Sigler, 2000 and Figure 8a). The complex of TBP bound to TATA box is stabilized by TFIIA that acts as a positive cofactor and this structure was solved by x-ray crystallography (Geiger et al., 1996; Tan et al., 1996; Bleichenbacher et al., 2003 and Figure 8c). Structure of TBP with Negative Cofactor 2 was determined by x-ray crystallography (Kamada et al., 2001) and explained the negative effect of this factor on transcription by preventing of recognition of TBP/DNA binary complex by

TFIIB. The orientation specificity of TBP for DNA sequence is not very rigorous and additional promoter elements such as BRE have to determine the orientation of PIC assembly (Cox et al., 1997). In archaea, the two TBP domains are more less symmetric so it is essential that the recognition of promoter directionality is conferred additionally by TFB binding to the BRE sequence. TFIID makes contacts not only to the TATA box (through TBP), but also additional downstream contacts, which might allow recognition of TATA-less promoters. These contacts in *Drosophila* and humans are conferred by two largest TAFs: TAF<sub>I</sub>1 and TAF<sub>I</sub>2 that can bind Inr directly (Chalkley and Verrijzer, 1999) and other two TAFs, TAF<sub>I</sub>6 and TAF<sub>I</sub>9 were reported to interact with DPE. TAFs are often regarded as coactivators, transducing the signals from activators to the core machinery (Verrijzer and Tjian, 1996). Several TAFs are essential for growth and proliferation and can have a positive or negative effect on transcription. In yeast, there is only one form of TFIID characterized, whereas in *Drosophila* and human these complexes can vary in composition. Some TAFs have a DNA-binding activity and can bind acetylated nucleosomes. Several have histone fold domains (HFDs) (Gangloff et al., 2002) and reconstitution of quaternary arrangement comparable to octamer has been reported (Selleck et al., 2001). Some complexes other than TFIID containing TAFs such as TFTC, SAGA, STAGA and PCAF have been reported (Brand et al., 1999) and TAFs in these complexes might form protein interacting cores. Low-resolution EM structures of TFIID, TFTC, and TFIID in complex with TFIIB and TFIIA were reported (Andel et al., 1999; Brand et al., 1999 and Figure 8e) and TFIID shows globular, horseshoe like organization with a central, solvent exposed concave with TBP (immunolocalized) in the middle and TFIIA and TFIIB on two sites of TBP. TFIID has a characteristic three lobes whereas TFTC has five. Surprisingly, none of the lobes could accommodate the histone-like octamer. Structures of TAFs were also investigated with x-ray crystallography especially histone-like pairs: *Drosophila* TAF<sub>I</sub>40/TAF<sub>I</sub>60 (Xie et al., 1996) and human TAF<sub>I</sub>18/TAF<sub>I</sub>28 (Birck et al., 1998) and TAF<sub>I</sub>4/TAF<sub>I</sub>12 (Werten et al., 2002 and Figure 8f) and double bromodomain module from TAF<sub>I</sub>250 (Jacobson et al., 2000) and also by NMR like TAF<sub>I</sub>230 with TBP (Liu et al., 1998).

### **TFIIB**

First characterized by Reinberg and Roeder in 1987, TFIIB plays a crucial role in PIC assembly, interacting with TBP, DNA, TFIIF and Pol II (Hampsey, 1998; Hahn, 2004). Binding studies identified this factor to be essential for TBP/DNA recognition by Pol II (Buratowski et al., 1989). TFIIB can also autoacetylate itself and this modification stabilizes interaction of TFIIB and TFIIF and activates transcription both in vitro and in vivo (Choi et

al., 2003). TFIIB consists of two domains, C-terminal domain, referred to also as core domain (TFIIBc) and N-terminal domain (TFIIBn). These domains are connected with a conserved, flexible region that forms a loop termed B-finger and a linker. The structure of TFIIBn encompassing ~30 residues was solved by NMR (Chen et al., 2000; Roberts and Green, 1994; Zhu et al., 1996; Barberis et al., 1993) and is composed of three antiparallel  $\beta$ -strands and a zinc binding motif (Bushnell et al., 2004; Chen et al., 2000; Hahn and Roberts, 2000). TFIIBn is essential for interaction of TFIIB with Pol II (Buratowski et al., 1993; Yamashita et al., 1993; Barberis et al., Ha et al., 1993; Hahn 2004). Mutations in N-terminal domain of TFIIB alter transcription start site determination (Bangur et al. 1997; Pardee et al., 1998; Hawkes et al., 1999; Bell et al., 2000, Fairley et al., 2002). It was shown by biochemical probing (Chen and Hahn, 2003) and with x-ray crystallography (Bushnell et al., 2004) to interact with the pocket formed by the wall, dock and clamp regions of Pol II. This would explain the conservation of ribbon domain in TFIIB, Brf1 (Juo et al., 2002) and archea (Kosa et al., 1997). C-terminal domain was very well characterized by several crystal structures in complex with TBP and DNA (Kosa et al., 1997; Nikolov et al., 1995; Littlefield et al., 1999; Tsai and Sigler, 2000 and Figure 8a). It consists of two cyclin A-like folds to clamp the C-terminus of TBP and interacts with DNA upstream and downstream of TATA box in a sequence-unspecific and with BRE in a sequence-specific manner (Lagrange et al., 1998). The B-finger might play multiple roles in initiation. It can stabilize the interaction of Pol II with short DNA/RNA intermediates and therefore play a role at the transition from open complex formation to actively transcribing Pol II (Bushnell et al., 2004). It was also proposed by the same group that B-finger might compete with RNA on the saddle of Pol II and account for promoter escape when RNA becomes longer than 10 nt which would disrupt TFIIB-Pol II interaction. A mechanism of TFIIB being a “molecular bridge” defining a distance between start site and TBP was proposed based on electron crystallography of TFIIB and TFIIE in complex with Pol II at 15.7 Å projection (Leuther et al., 1996) but in the light of new data (Bushnell et al., 2004 and Figure 8j) this model seems to be only partially correct.

### **TFIIF**

TFIIF was identified based on its strong, physical interaction with Pol II (Flores et al., 1990). It inhibits unspecific transcription initiation (Wang and Burton, 1995) and recruits Pol II to transcription initiation complex on promoter DNA making specific interactions with the promoter, TFIIB and TAF250 subunit of TFIID (Fang and Burton, 1996; Ruppert and Tjian, 1995; Wang and Burton, 1995). TFIIF contributes to open complex formation by enabling entry of TFIIE and TFIIH to the scaffold and therefore DNA helicase activity of TFIIH

(Maxon et al., 1994; Tirode et al., 1999). TFIIF remains bound to Pol II during RNA elongation, enhancing processivity and the polymerization rate (Lei et al., 1999; Ren et al., 1999; Yan et al., 1999). TFIIF also plays a role in the process of Pol II reinitiation by recruiting and enhancing the activity of phosphatase FCP1, which is required for dephosphorylation of the CTD of Pol II and is a prerequisite for the next round of transcription (Archambault et al., 1997). TFIIF forms a heterotetramer composed of conserved subunits, two large RAP74 in humans or  $\alpha$  in yeast and two small RAP30 or  $\beta$  in yeast. Yeast TFIIF also contains third, non-conserved subunit- $\gamma$ . X-ray studies showed the dimer of N-terminal domains of RAP74 and RA30 (Gaiser et al., 2000) and CTDs of RAP30 (Groft et al., 1998) and RAP74 (Kamada et al., 2001). N-terminal domains of these subunits form a novel, “triple barrel” dimerization fold and C-terminal domains form winged helix domains which in case of RAP30 makes unspecific DNA binding and in case of RAP74 interacts with FCP1 (Kamada et al., 2003). RAP30 through its winged helix domains might also bind DNA on both sides of TATA box and compact promoter DNA around Pol II (Forget et al., 1997; Kim et al., 1997). RAP74 can also interact with DNA. Cryo-EM of RNA Pol II in complex with TFIIF allowed for  $\sim 18$  Å reconstruction and showed RAP74 interacting with the extended surface of Pol II along the clamp and with the subcomplex of Rpb4/7 and RAP30 globular densities along the active site cleft (Chung et al., 2003 and Figure 8i).

### **TFIIE**

TFIIE was identified in early 90s' (Ohkuma et al., 1990 and 1991) and shown to recruit TFIIH to the preinitiation complex (Maxon et al., 1994). These two factors together stabilize and activate the PIC by completing the assembly and then melting the DNA around the transcription start site to create the transcription bubble, inducing the transition from closed to open PIC. TFIIE stimulates enzymatic activities of TFIIH, a helicase that is responsible for unwinding the dsDNA (Ohkuma and Roeder, 1994) and a kinase, which is responsible for phosphorylation of CTD of Pol II and transition to elongation phase of transcription (Drapkin et al., 1994). This transition is also facilitated by TFIIE, TFIIF and TFIIH, which cooperate in suppression of promoter proximal stalling. TFIIE can also melt the DNA at the supercoiled promoter. It was shown that premelted DNA around the transcription start site eliminated the requirement for TFIIE as well as TFIIH and ATP (Dvir et al., 1997; Holstege et al., 1996; Pan and Greenblatt, 1994; Tantin and Carey, 1994). It was shown that TFIIE can directly bind ssDNA (Kuldell and Buratowski, 1997) and crosslinking studies showed that it binds to promoter DNA upstream the transcription start site (Robert et al., 1996), within the transcription bubble or directly downstream of it. These contacts are consistent with electron

crystallography of TFIIE in complex with Pol II, which located it around the active center (Leuther et al., 1996). TFIIE is a heterotetramer consisting of two large subunits (Ohkuma et al., 1990) TFIIE $\alpha$  and two small TFIIE $\beta$ . NMR studies of TFIIE $\beta$  showed that it has a region with homology to the TFIIF winged helix domain (Okuda et al., 2000), which was proposed to bind DNA unspecifically in the similar manner as RAP30. The c-terminus of this small subunit of TFIIE seems to be essential for transition from initiation to elongation of transcription (Watanabe et al., 2003). NMR studies of part of TFIIE $\alpha$  showed a novel conserved zinc finger (Okuda et al., 2004), which is essential. Archaeal TFE is a homologue of TFIIE $\alpha$  and the structure of the N-terminal domain of TFE was solved (Meinhart et al., 2003 and Figure 8b). It consists of a winged helix domain which does not have DNA binding activity in vitro and was proposed to make interactions with other GTFs and Pol II.

### **TFIIH**

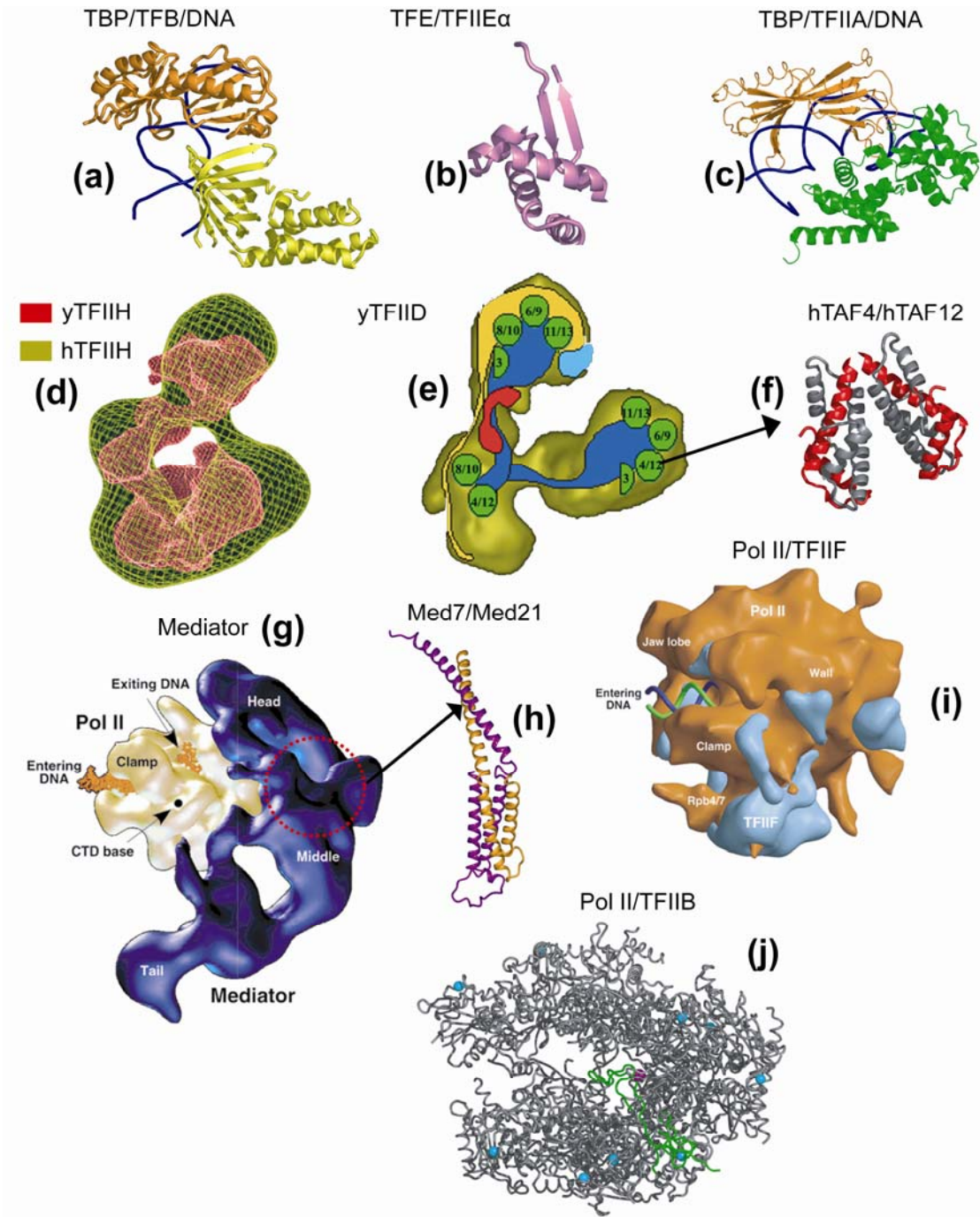
TFIIH is the largest and most complex of all GTFs. It consists of 10 subunits of total molecular weight of ~0.5 MDa and is comparable in complexity to Pol II and Mediator. Active, holo TFIIH was isolated from yeast (Svejstrup et al., 1994) and mammalian cells (Conaway et al., 1989). It is the only general transcription factor which has several enzymatic activities, including two different helicase activities of different polarities (subunit XPB/SSL2/RAD25 and subunit XPD/RAD3) and a cyclin-dependent protein kinase activity (subunit Cdk7/Kin28 and cyclin H/CCL1). It plays a central role in transcription by Pol I and Pol II, cell cycle control and DNA repair. TFIIH is composed of two subcomplexes: a core complex and the cyclin activating kinase complex (CAK). DNA dependent ATPases/helicases are essential for promoter opening by unwinding the DNA in an ATP dependent manner and this is crucial for the start of transcription. Six TFIIH subunits are sometimes referred to as repairosome (Svejstrup et al., 1995; Feaver et al., 2000), because they are involved in nucleotide excision repair and transcription coupled repair (Egly, 2001). Cdk7 phosphorylates the CTD (Gerber et al., 1995), which is a regulatory step in transition from initiation to elongation of transcription and this kinase activity is modulated by the Mediator. Cdk8, a subunit of Mediator, modulates Cdk7 through phosphorylation of cyclin H (Akoulitchiev et al., 2000). TFIIH interacts with several transcription factors (Bastien et al., 2000; Chen et al., 2000) and was also shown to be connected with the splicing machinery. Interaction with some snRNAs, like U1 snRNA stimulates the formation of first phosphodiester bond by Pol II in vitro (Kwek et al., 2002). Based on TFIIH/DNA crosslinking studies different models for DNA opening were proposed. XPB was shown to interact with DNA both upstream and downstream of the start site, resulting in tight wrapping of the template around PIC. ATP

hydrolysis would then induce conformational change in XPB, which would then open the DNA by physical interaction (Douziech et al., 2000). The second model proposes that XPB would unwind a DNA in an ATP-driven manner downstream of start site with upstream DNA fixed (Kim et al., 2000). Human TFIIH was characterized at 38 Å resolution by EM (Schultz et al., 2000 and Figure 8d) and revealed a ring-like structure with a hole in the middle, of dimensions, which could accommodate the double-stranded (ds) DNA. Diffraction from electron micrographs of the crystals in negative stain extended to about 18 Å resolution (Chang and Kornberg, 2002 and Figure 8d) and 3D reconstruction revealed several densities whose volumes corresponded well with those of individual TFIIH subunits. In yeast TFIIH there is also a ring composed of 3 subunits as in case of human structure. Crystal structures of cyclin H (Andersen et al., 1996) and Cdk7 (Lolli et al., 2004) were obtained, and the carboxy-terminal domain of P44 subunit, N-terminal domain of Mat1 and N-terminal pH/Ptb domain of P62 subunit were solved by NMR.

### **The Mediator**

The mediator is a large modular macromolecular assembly, which transmits gene signals from gene specific transcription factors to the core Pol II machinery. This huge coactivator was first discovered as a complex enabling activated transcription in vitro with purified Pol II and GTFs (Flanagan et al., 1991; Kim et al., 1994). In *Saccharomyces cerevisiae*, Mediator is composed of 25 subunits, which are organized in four subcomplexes, the head, middle, tail, and Cdk8 modules (Guglielmi et al., 2004; Blazek et al., 2005). In yeast, this coactivator is required for regulation of the majority of genes (Holstege et al., 1998) and was discovered to stimulate basal transcription (Mittler et al., 2001; Baek et al., 2002). The Mediator confers many interactions, with activators, Pol II and GTFs and can modulate the transition from initiation to elongation by stimulation of the CTD kinase activity of TFIIH (Kim et al., 1994). Mediator was also reported to remain bound near the promoter during active transcription and therefore facilitate transcription reinitiation (Liu et al., 2001; Yudkovsky et al., 2000). Mediator can transduce both positive and negative regulatory information to the core Pol II-GTFs machinery (Hampsey and Reinberg, 1999). Mediator also shows significant evolutionary conservation from yeast to human (Malik and Roeder, 2000), but the composition of different species of Mediator can vary depending on the organism and the stage of cell cycle. Mediator undergoes strong conformational changes upon binding to Pol II (Naar et al., 2002; Asturias et al., 1999; Davis et al., 2002) or activators (Taatjes et al., 2002). The resolution of these EM structures was limited to 30-35 Å and provided a gross shape of Mediator, but no structural details (Figure 8g). A high-resolution picture exists for the

conserved MED7/MED21 heterodimer, which was solved by x-ray crystallography at 3.0 Å resolution (Baumli et al., 2005 and Figure 8h).



**Figure 8. Structures of initiation machinery components.** (a) TBP/TFB/DNA (Littlefield et al., 1999, pdb entry: 1D3U); (b) TFE/TFE $\alpha$  (Meinhart et al., 2003, pdb entry: 1Q1H); (c) TBP/TFIIA/DNA (Tan et al., 1996, pdb entry: 1YTF); (d) molecular envelope of yeast ((red) Chang and Kornberg, 2000) and human ((yellow) Schultz, 2000) TFIIF; (e) molecular envelope of yeast TFIID (Brand et al., 1999) with TAFs (green) and TBP (red) positions; (f) histone pair of human Taf4 and Taf12 (Werten et al., 2002, pdb entry: 1H3O); (g) yeast holoenzyme (mediator and RNA Pol II, Davis et al., 2002); (h) Med7/Med21 structure (Baumli et al., 2005, pdb entry: 1YKH); (i) Pol II-TFIIF EM structure (Chung et al., 2003); (j) Pol II-TFIIB structure (Bushnell and Kornberg, 2004, pdb entry: 1R5U)

**Sequential assembly of PIC and interaction of GTFs with Pol II**

Based on x-ray studies, PIC assembly involves following steps.

1. Binding of TBP to TATA element through minor groove forms a stable complex, where DNA is distorted.
2. Binding of TFIIB through direct interactions with TBP and DNA.
3. Binding of TFIIA, which stabilized TBP-TFIIB-DNA complex.
4. Promoter targeting by preformed Pol II-TFIIF complex.
5. Recruitment of TFIIIE and TFIIF through interactions of TFIIIE with Pol II and DNA.

At this stage, transcription can start with promoter melting catalyzed by ATP dependent TFIIF helicase. Some structural knowledge of intermediates of these complexes exists (Figure 8) but the process of initiation will only be understood when the structural and biochemical data can be put in context of 12-subunit Pol II. At this stage, there is only one x-ray structure of GTF in complex with Pol II (Bushnell et al., 2004 and Figure 8j) and several low-resolution EM structures (e.g. Chung et al., 2003 and Figure 8i). From this data, the picture is as follows. The Zn-ribbon domain of TFIIB binds Pol II around the dock domain and the B-finger protrudes into the hybrid-binding site. It can stabilize interaction of Pol II with short DNA/RNA hybrids and therefore play a role during the transition from open complex formation to actively transcribing Pol II (Bushnell et al., 2004). This would be consistent with the action of  $\sigma$ , bacterial initiation factor that penetrates the hybrid-binding site with its 3.2 region (Cramer, 2004; Vassylyev, 2002). In the archaeal system homologues of TFIIB and TBP are sufficient for promoter loading but in the Pol II system, TFIIF is required (Tyree et al., 1993). One can expect topological similarity of promoter loading in Pol II and bacteria based on the available structures of Pol II with TFIIB and TFIIF and bacterial holoenzyme. These factors bind in the same region where  $\sigma$  was shown to bind in bacterial polymerase (Figure 4c). In this region, Mediator was also shown to bind, around the Rpb3-Rpb11 heterodimer and this region seems to be important as well for transcription regulation in bacteria (Davis et al., 2002). After assembly of TBP-TFIIB-Pol II/TFIIF complex, TFIIF with help of TFIIIE has to join and open the promoter. This step is poorly characterized with structures, electron crystallography at low resolution (Leuther et al., 1996) suggests that TFIIIE binds near the cleft, which is in agreement with studies showing that TFIIIE crosslinks DNA around the transcription start site in the PIC. There is no structural information concerning the Pol II-TFIIF complex except crosslinking (Kim et al., 2000).

## Chapter II

*Architecture of initiation-competent 12-subunit  
RNA polymerase II*

### **Abstract**

RNA polymerase II (Pol II) consists of a ten-polypeptide catalytic core and the two-subunit Rpb4/7 complex that is required for transcription initiation. Previous structures of the Pol II core revealed a “clamp,” which binds the DNA template strand via three “switch regions,” and a flexible “linker” to the C-terminal repeat domain (CTD). Here we derived a model of the complete Pol II by fitting structures of the core and Rpb4/7 to a 4.2 Å crystallographic electron density map. Rpb4/7 protrudes from the polymerase “upstream face,” on which initiation factors assemble for promoter DNA loading. Rpb7 forms a wedge between the clamp and the linker, restricting the clamp to a closed position. The wedge allosterically prevents entry of the promoter DNA duplex into the active center cleft, and induces in two switch regions a conformation poised for template strand binding. Interaction of Rpb4/7 with the linker explains Rpb4-mediated recruitment of the CTD phosphatase to the CTD during Pol II recycling. The core-Rpb7 interaction and some functions of Rpb4/7 are apparently conserved in all eukaryotic and archaeal RNA polymerases, but not in the bacterial enzyme.

## Introduction

Pol II synthesizes all eukaryotic mRNA in the course of gene transcription. The regulation of gene transcription by Pol II underlies cell proliferation and differentiation. Regulation occurs mainly at the level of transcription initiation, when Pol II assembles with the general transcription factors TFIIB, -D, -E, -F, and -H into the initiation complex on promoter DNA (Roeder, 1998; Lee and Young, 2000; Lemon and Tjian, 2000). In addition to the ten-subunit Pol II core, initiation requires the heterodimeric Rpb4/7 complex that can dissociate from core. Crystallographic structures are available for the *Saccharomyces cerevisiae* Pol II core (Cramer et al., 2000 and 2001), which is sufficient for RNA elongation (Ruet et al., 1980; Edwards et al., 1991). In the core structures, the two large subunits form opposite sides of a central “cleft,” and the eight small subunits are arrayed around the periphery. The cleft contains the active center, and is constricted at one end by a protein “wall.” One side of the cleft is formed by a mobile “clamp,” which adopts open states in two crystal forms of the Pol II core (Cramer et al., 2000 and 2001), but is closed in a further structure of a core elongation complex with bound template DNA and product RNA (Gnatt et al., 2001). Mobility of the clamp relies on five protein “switch” regions, which connect the clamp to the remainder of the enzyme (Cramer et al., 2001; Gnatt et al., 2001). A flexible “linker” emerges from the core surface below the clamp, and connects to the C-terminal repeat domain (CTD) of the largest Pol II subunit, which is disordered in all core structures. The purification of 10-subunit core was established several years ago and here the refined and optimized procedure is reported. Purification of *E.coli* overexpressed subcomplex of Rpb4/7 is shown in Chapter III and will not be presented here. The reconstitution of 12- polypeptide enzyme relies on quality of preparations of both proteins and the ultimate goal is to reconstitute stoichiometric, homogeneous species of this macromolecular assembly. It will be shown here how this reconstitution is done. Counterparts of the Rpb4/7 complex exist in the two other eukaryotic RNA polymerases (Sadhale and Woychik, 1994; Shpakovski and Shematorova, 1999; Peyroche et al., 2002; Siaut et al., 2003; Hu et al., 2002), and the archaeal RNA polymerase (Werner et al., 2000). The structure of an archaeal Rpb4/7 counterpart revealed an extended complex, spanned by the Rpb7 homolog (Todone et al., 2001). To understand Rpb4/7 function, and to elucidate the mechanism of transcription initiation, we studied the 12-subunit Pol II by X-ray crystallography.

## Materials and methods

### Fermentation and purification of 10-subunit Pol II

#### $\Delta$ Rpb4 strain

For structural biology purposes described in the introduction yeast strain CB010 $\Delta$ Rpb4 (Mata, pep4::His3/prb1::Leu2, prc1::HISG, can1, ade2, trp1, ura3, his3, leu2-3, 112) was used. This strain has a knockout of Rpb4 subunit of RNA polymerase II and since the stability of Rpb7 upon binding to Pol II is lower in absence of Rpb4, in purification stoichiometric and pure 10-subunit (lacking Rpb4/7 subcomplex) core Pol II is obtained. The other important features of the strain is a knockout of three cellular proteases namely: vacuolar proteinase B which is a serine-type endopeptidase; vacuolar carboxypeptidase Y (proteinase C) and most importantly vacuolar aspartyl protease (proteinase A), required for the posttranslational precursor maturation of vacuolar proteinases and therefore its knockout inactivates the whole cascade of proteolytic enzymes. This strain additionally has an Ade2 marker, a phosphoribosyl-aminoimidazole carboxylase that catalyzes a step in 'de novo' purine nucleotide biosynthetic pathway and the mutant cells deprived of adenine accumulate red pigment. Due to Ade2 marker, we have a good indication of the stationary phase since the cells turn red.

#### Fermentation of $\Delta$ Rpb4 strain

Since no overexpression systems are developed for obtaining RNA Pol II and as milligram quantities are required for crystallization, large scale fermentation of yeast was necessary to obtain sufficient amount of material. The amount of crystallization grade Pol II that is purified in one preparation is quite low and ranges from 6-10 mg from around 600 g of yeast pellet. For fermentation two different bioreactors ISF 20 and ISF 200 (INFORS HT, Figure 1) were used. Bioreactors have a huge advantage over shaker cultures, not only because it is possible to ferment large volumes in them but also to control several parameters like pH, temperature, amount of air and stirring. For yeast growth in ISF 20 and ISF 200 these parameters were refined experimentally. Table 1 shows parameters used in each of the fermenters. The steam and cold water flow adjust the temperature to the set value, and externally connected vessels with 2 M NaOH control the pH. Other parameters are set to fixed values and do not change during the process.

Yeast was grown at 30°C on YPD<sup>1</sup> medium with addition of 10 mg/l of Tetracyclin and 50 mg/l of Ampicilin to avoid contaminations caused by *E.coli*. First, 20 ml preculture was started from glycerol stock, grown for 16 hours and transferred to 500 ml preculture which

---

<sup>1</sup> Media, additives and buffers used during preparation of 10-subunit Pol II are listed in Table 3

was grown to OD<sub>600</sub> of 4 in the shaker. This culture was used then as inoculum for the ISF20. 15 litres of YPD medium in the fermenter were sterilized (20 minutes at 121°C) prior to inoculation and fermentation and cooled down to 30°C. For fermentation 0.003 % of Antifoam C Emulsion (Sigma) and proper amount of antibiotics were added. The starting OD in the ISF20 fermenter was usually around 0.1. The doubling time was estimated to be around 3.5-4 hours for this strain and usually yeast reached early stationary phase in approximately 23-26 hours. This culture then was pumped into ISF200 which was pre-filled with 185 litres of sterilized YPD medium and cooled down to 30°C. The antifoam and antibiotics were added and the fermentation started with the initial OD<sub>600</sub> usually around 0.1.

**Table 1.** Fermentation parameters



**Figure1.** ISF 200 (left) and ISF 20 (right)

<b>Fermenter</b>	<b>ISF 20</b>	<b>ISF 200</b>
<b>preculture; OD<sub>600</sub></b>	3-4	5-6
<b>preculture; volume (l)</b>	0.5	15
<b>starting OD<sub>600</sub></b>	0.1	0.1
<b>final OD<sub>600</sub></b>	5	6-7
<b>total volume (l)</b>	15	200
<b>temperature (°C)</b>	30	30
<b>pH</b>	7	7
<b>stirrer (RPM)</b>	600	200
<b>air (l/min)</b>	8	20
<b>fermentation time (h)</b>	23-26	22-23

Fermentation was finished after around 23 hours when yeast reached early stationary phase which was monitored with regular OD<sub>600</sub> measurements. Cells were harvested with continuous-flow Carl Padberg centrifuge and around 1.5-2 kg of yeast paste was obtained. The cells were then resuspended in 3x Pol II freezing buffer (330 ml/kg of cells), divided into 500-600 g aliquots, frozen in liquid nitrogen and stored at -80°C.

### **Lysis of yeast**

The first step before protein purification is cell lysis, which in case of yeast is quite tedious. The lysis has to be very efficient so that the majority of cells are disrupted and at the same time protein has to be intact and not degraded for example due to heating emitted during cell disruption. In case of 10-subunit Pol II yeast pellet was opened with the Bead Beaters (BioSpec Products) in which cells are disrupted by 0.45-0.5 mm borosilicate beads. In each preparation 3 Bead Beaters were used. The metal chamber was filled with 200 ml of beads and approximately 200 ml of yeast cell suspension which corresponds to approximately 200 g of dry yeast pellet. Then 1 ml of 100x protease inhibitor mix was added, mixed with a glass rod to get rid of air bubbles. The chambers were then filled to the top with buffer HSB150, and Bead Beaters assembled. The chambers were immersed in salt/ice mix and the bead beating started with a duty cycle of 60 s lysis, 90 s pause to avoid overheating. The lysis lasted for 75 minutes and every 10-15 minutes the salt/ice mix was exchanged. The efficiency of lysis was always around 90-95 %. After 75 minutes the beads were filtered from lysate with a mesh funnel. The beads were then washed with buffer HSB150 until the flowthrough became clear (the volume for three bead beaters never exceeded 600 ml of buffer HSB150). The lysate was then centrifuged twice at 12000 RPM for 30 minutes.

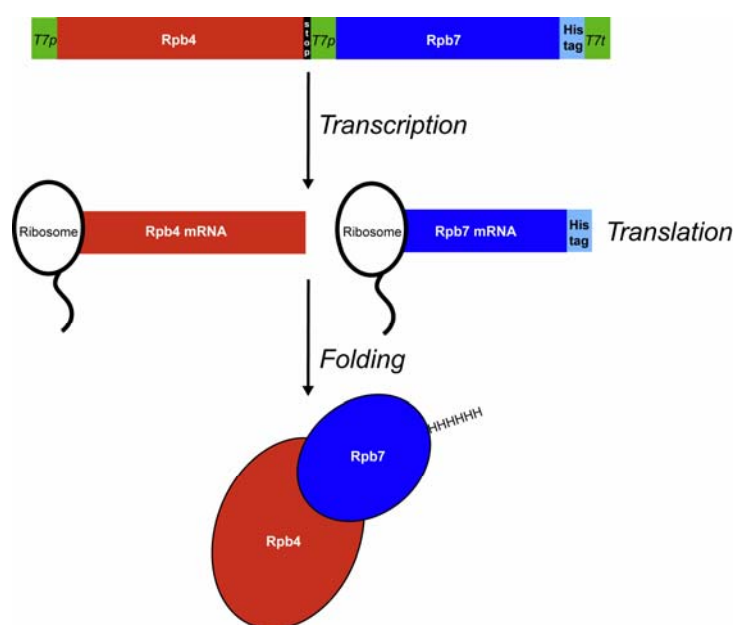
### **Purification of 10-subunit Pol II**

For 3 bead beaters preparation, 250 ml Hepharin Sepharose Fast Flow (Pharmacia, 17-0998-03) was used. The column was equilibrated with 3 column volumes (cv) of HSB150, where for last cv DTT and protease inhibitors were added. The cell extract was loaded on the column with the flow of 8 ml/min. Then the column was washed with 3 cv of buffer HSB150 and bound protein eluted with 500 ml HSB600. First 100 ml of the elution were discarded and the remaining 400 ml were precipitated for 30 minutes at 4°C with 116.8 g of solid, fine-ground ammonium sulphate. The precipitation solution was then centrifuged for 45 minutes at 12000 RPM, the supernatant decanted and pellet stored at 4°C overnight. Second day of purification started with dissolving of ammonium sulphate pellet in 80 ml of TEZ0 which corresponded to conductivity below 400µSi. 5 ml 8WG16 immunoaffinity column was equilibrated with 4 cv of buffer TEZ250 and for last column volume DTT was added to the buffer. The column was

loaded by gravity and the flowthrough was reloaded to ensure complete binding of Pol II. The wash step was done at room temperature with 5cv of TEZ500 and 10-subunit Pol II was eluted with TEZ500/glycerol buffer, 1 ml fractions were collected and checked for protein with Bradford reagent. Main fractions were pooled and stored at 4°C overnight. The 8WG16 column was cleaned with TEZ500/ethylene glycol/noDTT buffer, reequilibrated with TEZ250/noDTT buffer containing 0.02 % azide and stored at 4°C for next purification. Usually the column can be used around 10 times with high efficiency of binding. The subsequent purification step is done on AEKTA HPLC purification system (Amersham Biosciences) and is based on anion exchange chromatography. This step was changed in the course of this PhD work in comparison to old protocols. The Phenomenex Biosep DEAE 5PW which was used in a previous protocol (Cramer et al., 2000) was replaced with Biorad UnoQ 1.35 ml anion exchange column. The main fractions from immunoaffinity step were diluted with buffer UnoQ 0 until the conductivity corresponded to 90 mM ammonium sulphate (usually 1:8 dilution). The UnoQ column was washed with high salt UnoQ B buffer and preequilibrated with UnoQ A buffer. The sample was loaded and unbound protein was washed with 3 cv of buffer UnoQ A. Pol II was eluted with a linear gradient from 0-50 % over 10 cv of UnoQ B buffer and 1 ml fractions were collected. Column was then stored in water with addition of 0.02 % of azide. The peak fractions were pooled and aliquots of 500 µg of protein were precipitated with saturated ammonium sulfate for around 1h at 4°C (protein solution 1:1 with saturated ammonium sulphate). Aliquots were then centrifuged for 30 minutes at 14000 RPM, supernatant decanted (some left to keep pellet covered), frozen in liquid nitrogen and kept in -80°C ready for crystallization experiments. The yield from one purification ranged from 6 to 12 mg of 10-subunit RNA Pol II.

### Expression and purification of full-length Rpb4/7

Full length Rpb4/7 was a kind gift from M. Kimura and A. Ishihama. Neither Rpb4 nor Rpb7 could be overexpressed alone in *E.coli*. This problem was overcome by construction of double-expression system for Rpb4/7 (Sakurai et al., 1999) with T7 promoter preceding Rpb4 coding frame and second T7 promoter preceding Rpb7 coding frame followed with a His-Tag and T7 terminator. This strategy is shown in Figure 2. This strategy is a key to get both proteins overexpressed in a soluble manner, if expressed independently they do not have the interaction partner for co-folding and aggregate. Upon translation both nascent peptide chains emerge from ribosomes in physical proximity and thereby fold properly. The complex can be purified with affinity chromatography due to His-tag which is cloned C-terminally on Rpb7. Table 2 lists few important facts about both proteins.



**Figure 2. Double expression strategy.**

*Rpb4 is depicted in red and Rpb7 in blue.*

**Table 2.** Number of residues, molecular weight, pI and estimated charge of Rpb4 and Rpb7

Protein	Rpb4	Rpb7
Amino acid residues	221	171
Molecular weight (kDa)	25.415	19.059
pI	4.92	7.8
Estimated charge at pH 7.0	-15.6	1.8

The transformed *E. coli* BL21(DE3) with double expression plasmid were grown in LB medium at 37°C. When the cells reached optical density of 0.6 measured at 600nm, isopropyl- $\beta$ -D-thiogalactopyranosid (IPTG) was added to a final concentration of 0.05  $\mu$ M. The cells were then grown for 16 hours at 18°C, harvested by centrifugation, suspended in Rpb4/7 freezing buffer<sup>2</sup> and stored at -80°C. For purification of expressed protein, frozen cells were thaw at room temperature, lysed with a French Press (Gaulin) with 750 Bar pressure, centrifuged for 20 minutes at 16000 RPM two times at 4°C. The resulting supernatant was loaded on Ni<sup>2+</sup>-nitriloacetic acid agarose (Qiagen). The column was washed in a stepwise fashion with 3 times the column volume of buffer 0, 3 times buffer 0 containing 10 mM imidazole and 3 times buffer 0 containing mM imidazole. Protein was then eluted with buffer 0 containing 50 mM imidazole. To wash the column and elute the remaining bound protein 5 times column volume buffer 0 containing 200 mM imidazole was used. Fractions were checked with Bradford reagent to detect if protein was eluted, and loaded on 15 % SDS-PAGE. The main fractions were then diluted with buffer C and loaded with buffer A on the ResourceQ<sup>3</sup> 6 ml anion exchange chromatography column (Pharmacia) where they were eluted with a gradient of 0-100 % ResourceQ buffer B. Fractions were then checked with Bradford and loaded on 15 % SDS-PAGE. Main fractions were upconcentrated and loaded with Pol II buffer on Superose 12 HR 10/30 column (Pharmacia). Peak fractions were then loaded on 15 % SDS-PAGE gel and then upconcentrated to 7 mg/ml, frozen in liquid nitrogen and stored at -80°C.

### **Assembly of 12-subunit Pol II**

For reconstitution of stoichiometric 12-subunit Pol II, 500  $\mu$ g frozen ammonium sulphate pellet of 10-subunit Pol II was thaw on ice and centrifuged for 20 minutes at 14000 RPM at 4°C. The solution covering the pellet was carefully removed with a pipet and the pellet was dissolved in 300  $\mu$ l of Pol II buffer and 5 M excess of full-length Rpb4/7 was added. Total volume of the solution should not exceed 350  $\mu$ l. The solution was then incubated for 1h at 20°C and then loaded with Pol II buffer on preequilibrated Superose 6 gel filtration column. The column separates 12-subunit complex from the excess of Rpb4/7 and the samples from the peak fractions were loaded on 15 % SDS-PAGE gel to check visually the subunit composition of assembled complex.

---

<sup>2</sup> Media, additives and buffers used during preparation of Rpb4/7 are listed in Table 4

**Crystallization of 12-subunit Pol II**

For crystallization of 12-subunit RNA Pol II hanging drop method was used. Therefore 1.5  $\mu$ l of the purified protein (solutions contained 4 mg/ml protein) was set with a pipette as a drop on a plastic cover slide and mixed with 1.5  $\mu$ l of reservoir solution. The slide was set up side down over a well with reservoir solution and sealed to the reservoir edge using high vacuum grease. As the drop contains lower concentration of crystallization agent, water vapour will leave the drop and the concentration will increase until supersaturation and crystal growth is reached. Crystallization was carried out at 20°C and the plates were examined under a microscope few days after setting the drops for presence of crystals. Crystals were harvested in a mother solution and several different cryoprotectants were carefully screened. The cryoprotectants were introduced in a stepwise fashion and quick freezing at 4°C and at 20°C was tried.

**Table 3.** Media, additives and buffers used during fermentation and purification of 10-subunit Pol II

<i>Name</i>	<b>Description of additives</b>
<b>Yeast media</b>	
YPD	1.5 % yeast extract, 2 % peptone, 2 % glucose
<b>Antibiotics</b>	
Ampicillin	50 mg/l of Ampicilin
Tetracycline	10 mg/l of Tetracyclin
<b>Purification</b>	
100x protease inhibitor mix	1.42 mg Leupeptin, 6.85 mg Pepstatin A, 0.85 g PMSF, 1.65 g Benzamidine; add Ethanol up to 50 ml; stored at -20°C
3x freezing buffer	150 mM Tris-Cl, pH 7.5 at 4°C, 3 mM EDTA, 30 % glycerol, 30 µM ZnCl <sub>2</sub> , 3 % DMSO, 30 mM DTT, 3x protease inhibitor mix
<b>Heparin purification</b>	
1x HSB150	50 mM Tris-Cl pH 7.9 at 4°C, 150 mM KCl, 1 mM EDTA, 10 % glycerol, 10 µM ZnCl <sub>2</sub> , 10 mM DTT, 1x protease inhibitor mix
1x HSB600	50 mM Tris-Cl pH 7.9 at 4°C, 600 mM KCl, 1 mM EDTA, 10 % glycerol, 10 µM ZnCl <sub>2</sub> , 10 mM DTT, 1x protease inhibitor mix
<b>Antibody purification</b>	
10x TEZ0	500 mM Tris-Cl pH 7.5 at 20°C, 10 mM EDTA, 100 µM ZnCl <sub>2</sub>
1x TEZ0	5 ml 10x TEZ0, 5 mM DTT, 1x protease inhibitor mix, H <sub>2</sub> O to 50 ml
1x TEZ250	5 ml 10x TEZ0, 250 mM ammonium sulphate (3 M stock), 5 mM DTT, 1x protease inhibitor mix, H <sub>2</sub> O to 50 ml
1x TEZ500	5 ml 10x TEZ0, 500 mM ammonium sulphate (3 M stock), 5 mM DTT, 1x protease inhibitor mix, H <sub>2</sub> O to 50 ml
1x TEZ500 + glycerol	5 ml 10x TEZ0, 500 mM ammonium sulphate (3 M stock), 50 % glycerol, 5 mM DTT, no protease inhibitors, H <sub>2</sub> O to 50 ml
1x TEZ500 + ethylene glycol	5 ml 10x TEZ0, 500 mM ammonium sulphate (3 M stock), 70 % ethylene glycol, no DTT, no protease inhibitors, H <sub>2</sub> O to 50 ml
<b>Anion exchange chromatography (UnoQ)</b>	
UnoQ 0	50 mM Tris-Cl pH 7.5 at 20°C, 10 % glycerol, 1 mM EDTA, 10 µM ZnCl <sub>2</sub> , 10 mM DTT
UnoQ A	50 mM Tris-Cl pH 7.5 at 20°C, 60 mM ammonium sulphate, 10 % glycerol, 1 mM EDTA, 10 µM ZnCl <sub>2</sub> , 10mM DTT
UnoQ B	50 mM Tris-Cl pH 7.5 at 20°C, 1 M ammonium sulphate, 10 % glycerol, 1 mM EDTA, 10 µM ZnCl <sub>2</sub> , 10 mM DTT
<b>Gel filtration (Superose 6)</b>	
Pol II buffer	5 mM Hepes pH 7.25, 40 mM ammonium sulphate, 10 µM ZnCl <sub>2</sub> , 10 mM DTT

**Table 4.** Media, additives and buffers used during purification of Rpb4/7

<i>Name</i>	<b>Description of additives</b>
<b><i>E.coli</i> Media</b>	
LB	10 g Trypton, 5 g Yeast Extract, 5 g NaCl, 1.3 ml 2 M NaOH; filled with H <sub>2</sub> O to 1 litre. For making culture plates 1, 5 % Agar (w/v) was added.
<b>Antibiotics/ IPTG</b>	
Ampicillin	100 mg/ml H <sub>2</sub> O
IPTG	234.3 mg/ml H <sub>2</sub> O (1M)
<b>Purification</b>	
100x protease inhibitor mix	1.42 mg Leupeptin, 6.85 mg Pepstatin A, 0.85 g PMSF, 1.65 g Benzamidine; add Ethanol up to 50 ml; stored at -20°C
Lysis Buffer	50 mM Tris (pH 7.0), 150 mM NaCl, 10 % glycerol
<b>Ni-NTA purification</b>	
Elution buffers for Ni-NTA column	Lysis Buffer with increasing concentrations of imidazole: 10 mM, 20 mM, 50 mM, 200 mM
<b>Anion exchange chromatography (ResourceQ)</b>	
Buffer A	100 mM NaCl, 20 mM Tris pH 7.5, 1 mM EDTA, 10 mM DTT
Buffer B	1 M NaCl, 20 mM Tris pH 7.5, 1 mM EDTA, 10 mM DTT
Buffer C	20 mM Tris pH 7.5, 1 mM EDTA, 10 mM DTT
<b>Gel filtration (Superose 12)</b>	
Pol II buffer	5 mM Hepes pH 7.25, 40 mM ammonium sulphate , 10 mM ZnCl <sub>2</sub>

## Results and Discussion

### Fermentation of 10-subunit Pol II

The yeast growth was carefully monitored during fermentation. The measurements of optical density at 600 nm at defined time points were plotted on the logarithmic curve to define the phase of growth. These measurements were supported by light-microscopy examinations of cells to check for possible contaminations and look at yeast morphology.  $\Delta$ Rpb4 cells have to be harvested at late logarithmic early stationary phase to ensure that the degradation of Pol II is minimized. Pol II is mostly abundant in the cell during logarithmic phase of growth and this is another reason for which the cells have to be harvested before they reach late stationary phase. The estimate of the doubling time for this yeast strain is around 4 h. Figure 3 shows a logarithmic growth curve of  $\Delta$ Rpb4 strain.

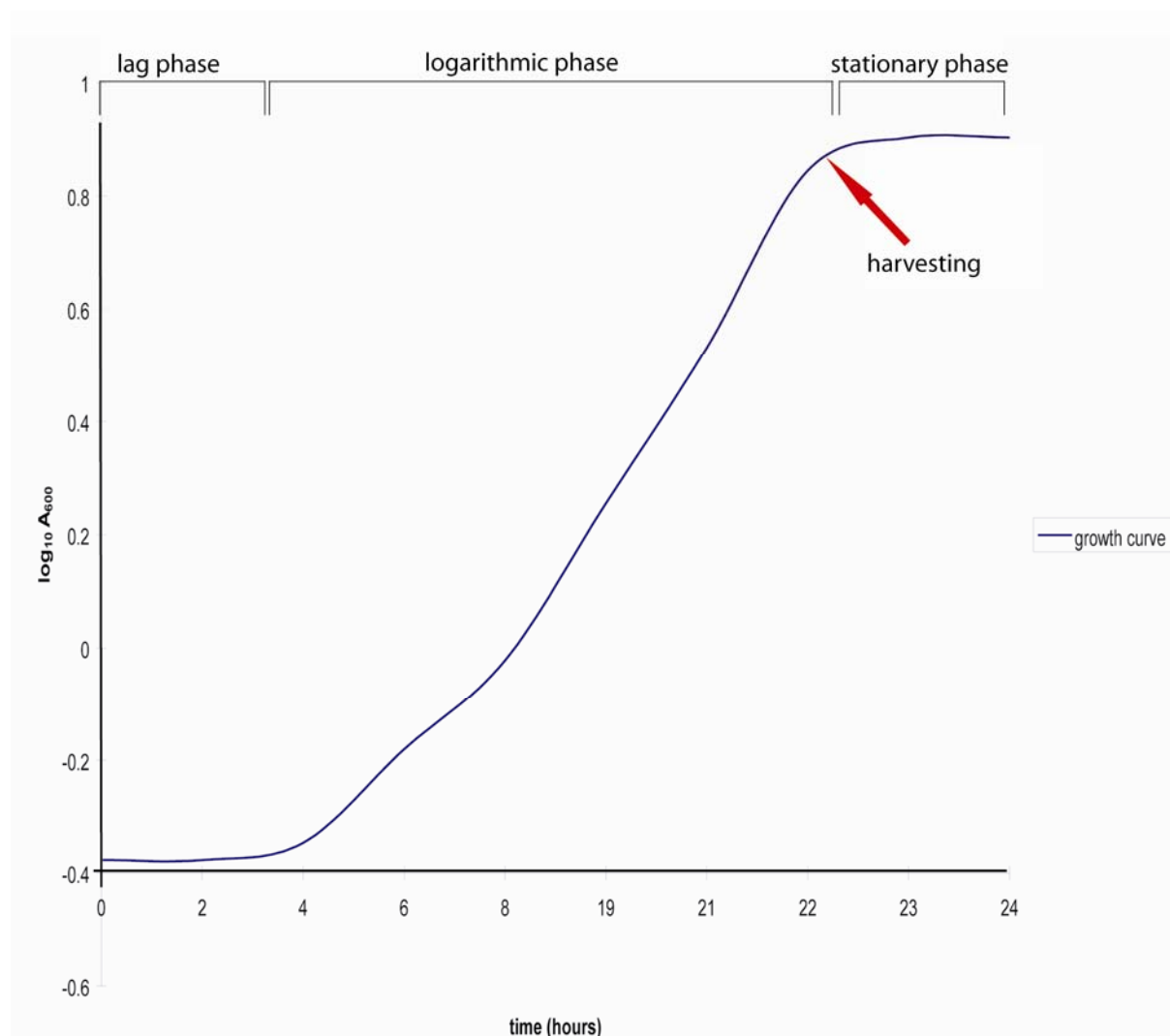
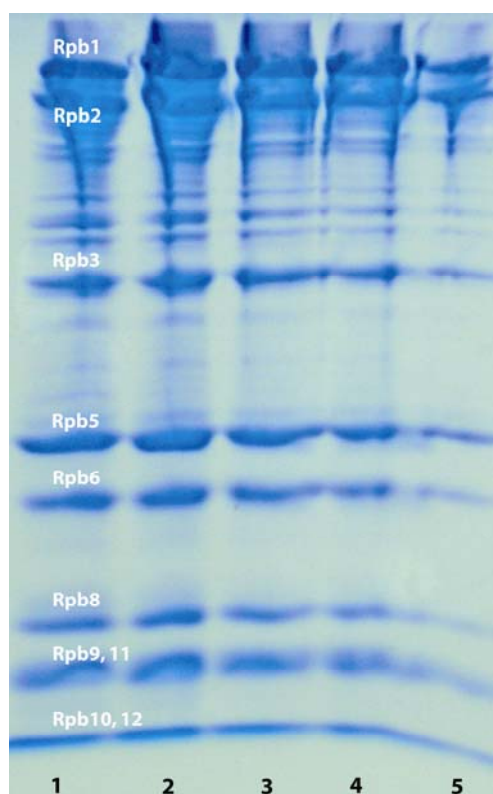


Figure 3. Logarithmic growth curve of  $\Delta$ Rpb4 strain in ISF200.

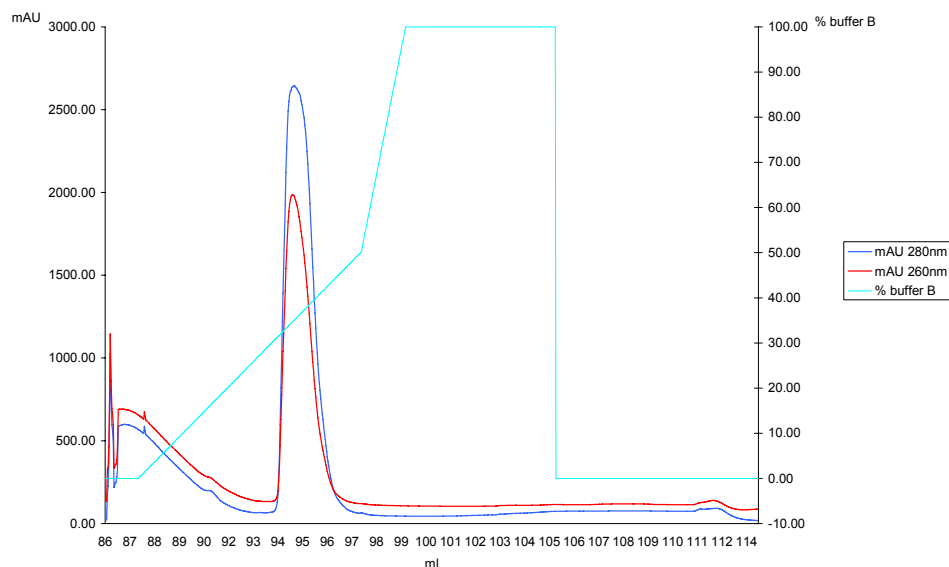
### Purification of 10-subunit Pol II

After the first step of preparation, Pol II cannot be seen on the SDS-PAGE because there is lots of other proteins which co-elute with it from Heparin column and this step is only a first crude separation based on their ionic character. The ammonium sulphate precipitation further reduces the number of other proteins in the solution. The step after which Pol II becomes distinguishable on the SDS-PAGE is the immunoaffinity 8WG16. 1 ml elution fractions from antibody column were collected, checked with Bradford reagent whether they contain protein and 20  $\mu$ l samples were loaded on 15 % SDS-PAGE gel (Figure 4).



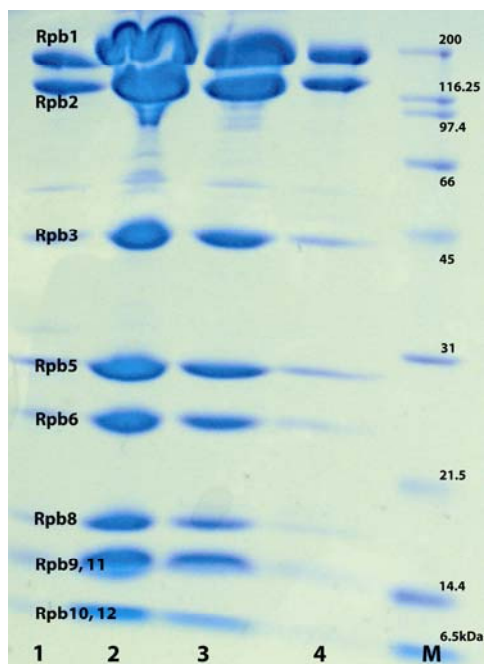
**Figure 4. 15 % SDS-PAGE of 10-subunit Pol II after antibody column.**  
Lanes 1-5 show fractions collected from column; subunits of Pol II (white).

The quality of Pol II after the antibody column was already very high but it had to be further purified because crystal setups after this step were still unsuccessful. UnoQ anion exchange chromatography was the last, refining step of purification and protein that elutes from this column is extremely pure and of crystallization quality. Figure 5 shows the UnoQ elution profile of Pol II.



**Figure 5. UnoQ anion exchange chromatography of 10-subunit RNA Pol II.**  
*Blue: absorption at 280 nm, red absorption at 260 nm, green: salt gradient*

Peak fractions were collected and loaded on SDS-PAGE gel to check for quality (Figure 6).

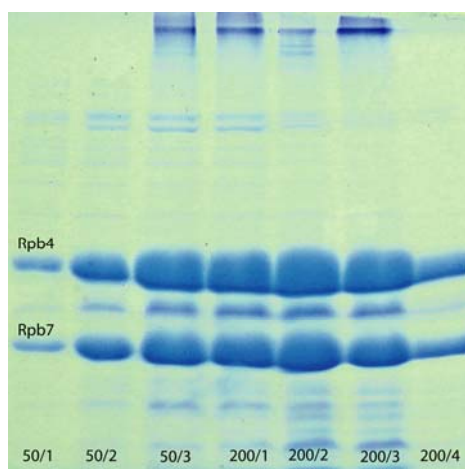


**Figure 6. 15 % SDS-PAGE of 10-subunit Pol II after UnoQ (lane 1-4).**  
*M: marker*

Pol II is then precipitated with ammonium sulphate, frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  where it is stable for 2-3 months.

### Expression and purification of full-length Rpb4/7

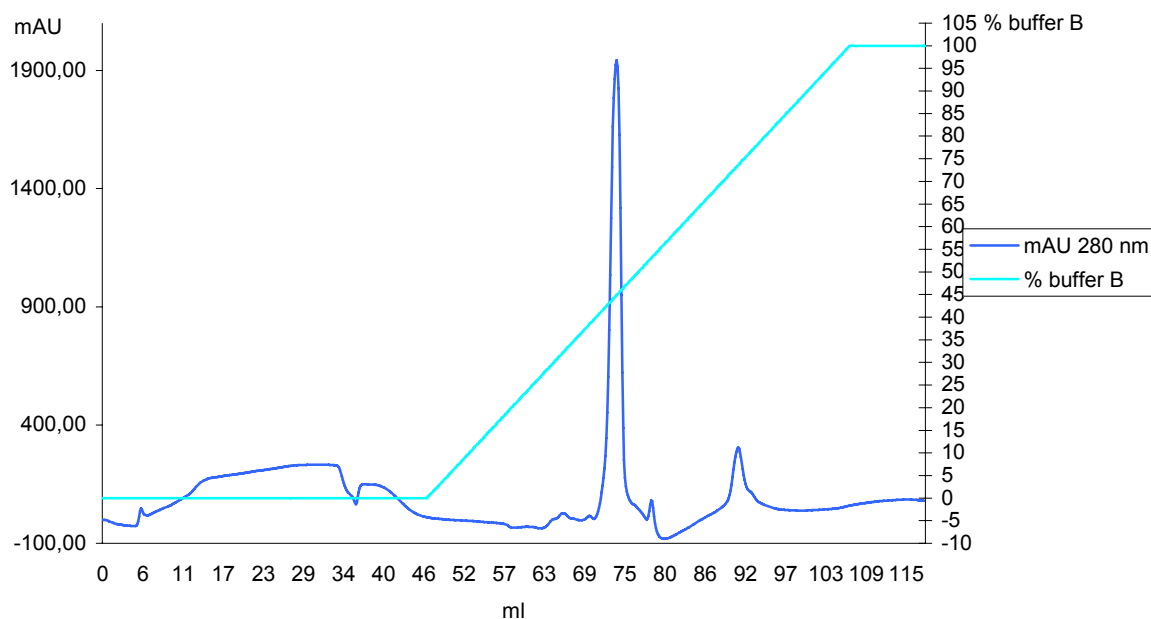
Full length Rpb4/7 was expressed in 4l of LB medium to ensure sufficient amount for further reconstitution and crystallisation of 12-subunit RNA Pol II. It was purified by Ni-NTA gravity flow affinity chromatography. The optimal imidazole concentration of the elution solution was tested and determined by a step gradient of 10, 20, 50 and 200 mM of imidazole. The elution occurred mostly at the concentration of imidazole between 20 – 200 mM. The concentration of protein in these fractions was estimated with Bradford reagent and the purity was checked on SDS-PAGE (Figure 7).



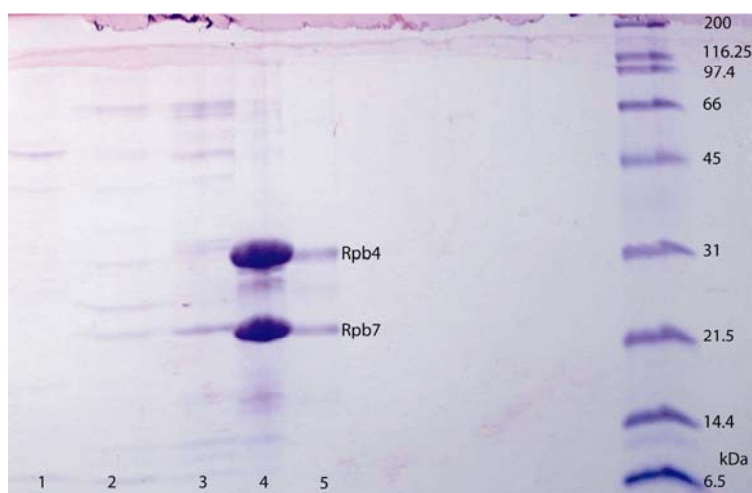
**Figure 7. 15 % SDS-PAGE of elutions of full length Rpb4/7 from Ni-NTA.**

*50-200 mM imidazole elutions of full length Rpb4/7 from left to right respectively.*

The thin band under Rpb4 is a degradation product and is removed in subsequent purification steps. The 50 and 200 mM imidazole peak fractions were pooled together, diluted 3 times with buffer C to reach the salt amount below 100mM NaCl (ResourceQ buffer A) and applied to the anion exchange chromatography ResourceQ column (Figure 8). The peak fractions were checked on SDS-PAGE (Figure 9).

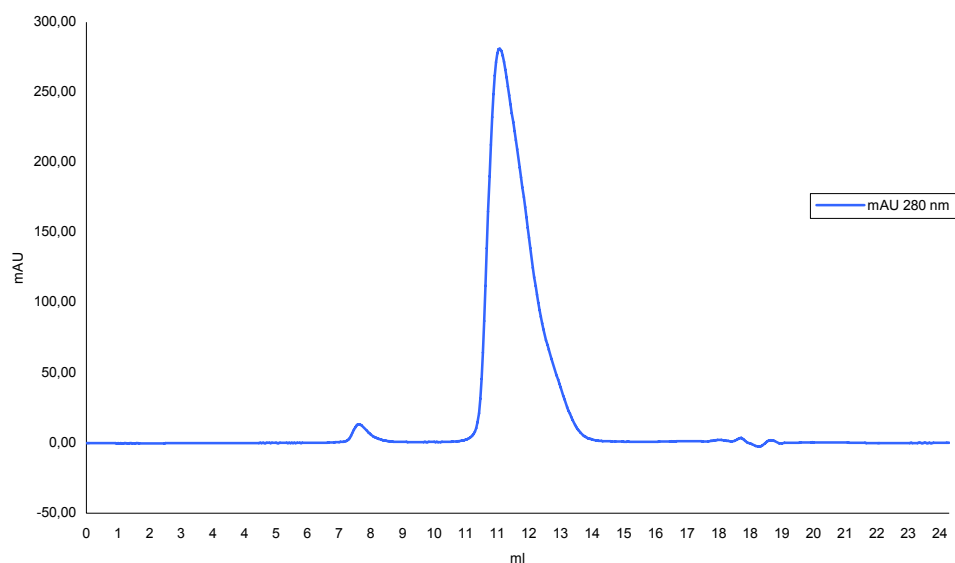


**Figure 8. ResourceQ anion exchange chromatography of full length Rpb4/7.**  
*Blue: absorption at 280 nm, green: salt gradient*

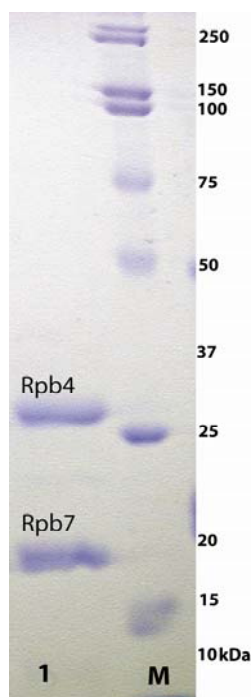


**Figure 9. 15 % SDS-PAGE with ResourceQ anion exchange fractions.**  
*Protein fractions were loaded on SDS-PAGE. Lanes 1 and 2 flowthrough fractions 3, 4, 5 peak fractions, protein marker in the last lane on the right side*

The peak fractions were pooled, concentrated to total volume of 600  $\mu$ l and subjected to next and final step of purification, Superose 12 gel filtration onto which 300  $\mu$ l of sample was loaded for one run (Figures 10 and 11).



**Figure 10. Superose 12 gel filtration chromatography of full length Rpb4/7.**  
*Blue: absorption at 280 nm*

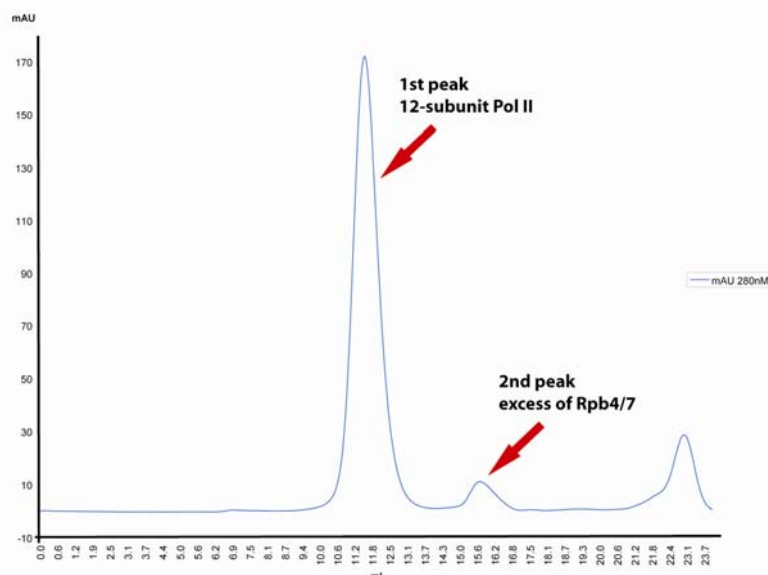


**Figure 11. 15 % SDS-PAGE with Superose 12 peak fraction.**  
*20  $\mu$ l protein from peak fraction were loaded on SDS-PAGE*  
*1: a peak fraction and M is a protein marker*

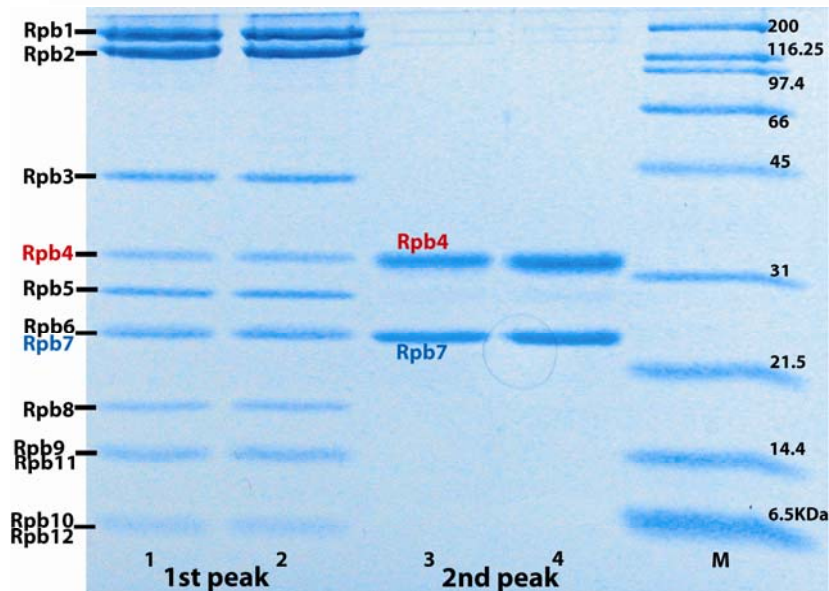
The purity of full length Rpb4/7 after gel filtration is very high, there are no additional proteins, such as degradation products or impurities visible on gel and the quality is therefore sufficient for crystallization experiments. The concentration of protein was measured with Bradford reagent, peak fractions pooled and concentrated to 7 mg/ml, flash frozen in liquid nitrogen and stored in  $-80^{\circ}\text{C}$  for further use. The total amount of this highly pure crystallization quality Rpb4/7 is around 4 mg from 4 l of expression.

### Assembly of 12-subunit Pol II

12-subunit Pol II is assembled as described in Materials and Methods. To get a homogeneous and stoichiometric sample, 10-subunit Pol II was incubated with 5 M excess of Rpb4/7. The excess of Rpb4/7 was removed with gel filtration (Figure 12 and Figure 13).



**Figure 12. Superose 6 gel filtration chromatography of 10-subunit Pol II core incubated with 5x molar excess of full length Rpb4/7 subcomplex.**  
Blue: absorption at 280 nm;  
Red arrows indicate peaks and their content



**Figure 13. 15 % SDS-PAGE of fractions from Superose 6 12-subunit Pol II assembly.**  
*1<sup>st</sup> peak from Superose 6 lane 1 and 2; 2<sup>nd</sup> peak from Superose 6 lane 3 and 4; M: marker*

After gel filtration protein was concentrated and subjected to crystallization trials.

### Crystallization of 12-subunit Pol II

Peak protein fractions from Superose 6, which contained 12-subunit RNA Pol II were concentrated to 4 mg/ml with 10 kDa cutoff concentrator. Crystallization trials were carried out with the hanging drop vapour diffusion method. Establishment of crystallization conditions was done by sparse matrix sampling using commercially available Hampton Screen and Hampton Screen1 and screens derived from conditions in which 10-subunit Pol II was crystallized. After crystals were grown, promising conditions were refined by variation of salts, pH and PEG in the reservoir. Table 5 shows the conditions in which 12-subunits Pol II crystals were obtained. Conditions which yielded crystals diffracting to 4.2 Å and 3.8 Å are marked red in Table 5.

**Table 5.** Crystallization conditions for 10- and 12-subunit Pol II

Conditions	Cryoprotocol	Comments	Max. Reso.
<b>10-subunit Pol II crystallization condition</b>			
<b>PEG/PO<sub>4</sub></b> 390 mM NH <sub>4</sub> NaPO <sub>4</sub> , pH 6.0 15-18 % PEG 6k 50 mM dioxane 5 mM DTT	shrinkage protocol (Cramer et al., 2001) 16 % PEG6K, 100 mM MES pH 6.3, 50 mM dioxane, 350 mM NaCl, 17 % PEG400 in 7 steps ; 4°C over 24 h	conditions for high-resolution 10 subunit Pol II	2.8 Å
<b>12-subunit Pol II crystallization conditions</b>			
<b>PEG20K/PO<sub>4</sub></b> 320 – 360 mM NH <sub>4</sub> NaHPO <sub>4</sub> , pH 6.0 8-12 % PEG20k 25 mM NaCl 5 mM DTT	mother solution + 0-(16-22) % glycerol in 5 steps; 8 °C over night	original 12 subunit Pol II conditions	ca. 4.5 Å
<b>Natrix #38</b> 200 mM ammonium acetate 150 mM magnesium acetate 50 mM Hepes pH 7.0 5 % PEG4k or 6k 5 mM DTT or TCEP	mother solution + 0-20 % glycerol in 5 steps; 8 °C over night		3.6 Å

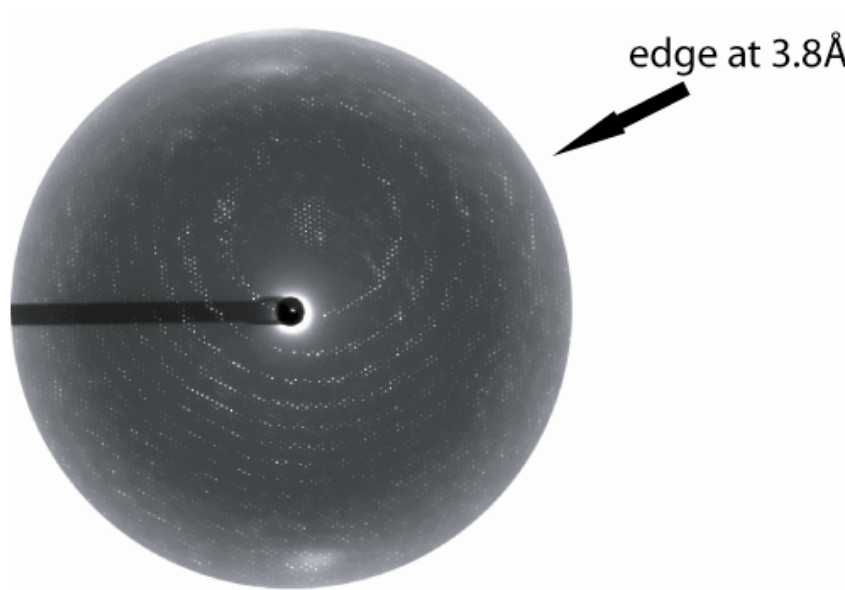
Conditions	Cryoprotocol	Comments	Max. Reso.
<b>Tartrate/Hepes</b> 750 – 950 mM $\text{NH}_4\text{NaTartrate}$ 100 mM Hepes-Na pH 7.5 5 mM DTT or TCEP	mother solution + 0-22 % glycerol in 5 steps; 8 °C over night	modified Hampton screen 1 #29	3.8 Å
<b>Tartrate/PEG/KSCN</b> 12-13 % PEG-6000 300 mM ammonium-sodium tartrate 100 mM Hepes pH 7.5 100 mM KSCN 5 mM DTT or TCEP	mother solution + 0-16 % glycerol in 5 steps; 8 °C over night	good and reproducible for bubble complexes	3.8 Å

Figure 14 shows 12-subunit crystals from initial to refined conditions



**Figure 14. Crystals of 12-subunit Pol II.** *Left: initial; Middle: primary refinement; Right: refined conditions (listed in Table3)*

Best crystals were carefully frozen with the use of different cryoprotectants and freezing protocols, and checked on the synchrotron to find out the best condition and harvesting protocol. Crystals were grown at 20° C with the hanging drop method, using a reservoir solution of 8 % PEG 20k, 360 mM  $\text{NaH}_2\text{PO}_4/(\text{NH}_4)_2\text{HPO}_4$  pH 6.0, 50 mM dioxane, 5 mM DTT. Crystals grew to a maximum size of 0.4 x 0.2 x 0.1 mm. Crystals were transferred stepwise to reservoir solutions with 5, 10, 18, 22 % ethylene glycol or glycerol, were incubated at 4° C overnight, mounted in cryo loops, and dispersed and stored in liquid nitrogen. Diffraction was observed beyond 4 Å resolution (Figure 15), but radiation damage and anisotropy limited complete data with good statistics to 4.2 Å resolution.



**Figure 15. Diffraction pattern of 12-subunit RNA Pol II crystal.**  
*Measured at SLS, X06SA beamline, 1s exposure time, oscillation increment 0.25°.*

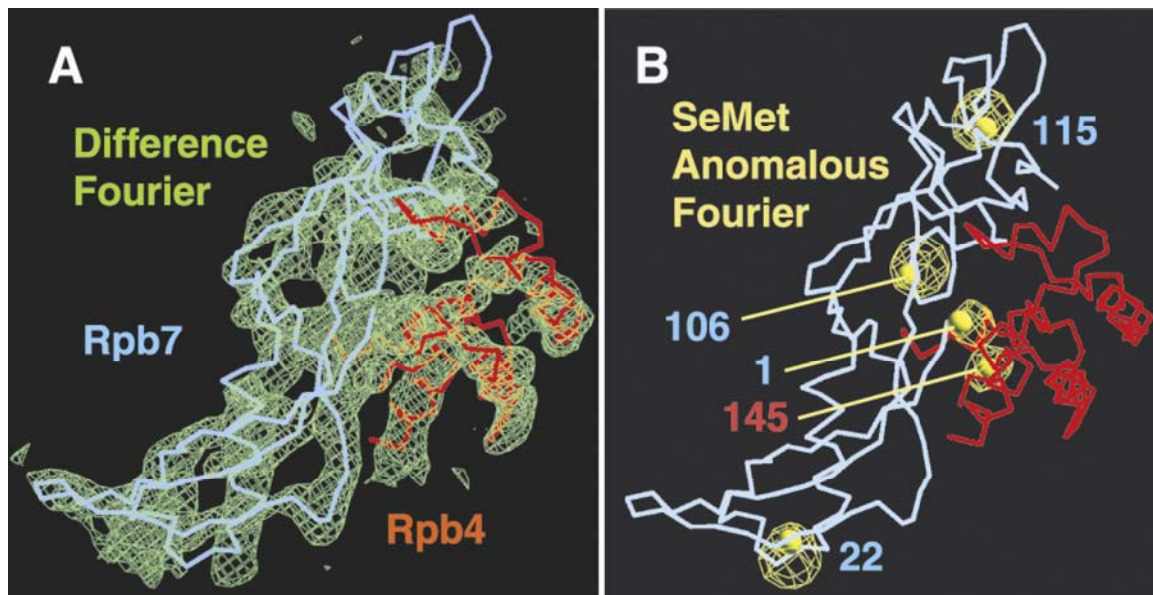
### X-ray structure determination of 12-subunit Pol II

With the use of cryo-cooling and synchrotron radiation, complete diffraction data to 4.2 Å resolution were obtained (Table 6), which allowed for structure solution by molecular replacement. The highest peak in molecular replacement with AMORE (Navaza, 1994)) was obtained using as search model the core elongation complex structure, with nucleic acids removed (correlation coefficient=45.1/18.4 for first/second peak, 10–6.5 Å). This structure was used, as a model for the core. Eight additional residues of the Rpb1 linker were included. The archaeal Rpb4/7 structure was manually fitted to  $2F_{\text{obs}}-F_{\text{calc}}$  and  $F_{\text{obs}}-F_{\text{calc}}$  electron density maps phased with the core model ((Todone et al., 2001), Figure 16A). The orientation of Rpb4/7 was additionally confirmed by labeling Rpb4/7 with selenomethionine (Figure 16B). Manual rigid body adjustments accounted for slight changes in the relative position of three domains (Rpb4; yeast Rpb7 residues 1-80; yeast Rpb7 residues 81-169). Archaeal residues 117 and 174-187 were deleted, and yeast residues 52, 53, 155, 156 were inserted. An insertion between  $\alpha$ -helices H1 and H2 in yeast Rpb4 could not be modeled. Although an additional helical density was observed, the connectivity was uncertain. No continuous density was observed for an N-terminal extension of yeast Rpb4. At the resolution of the data, no refinement was carried out and the fitted structures were reduced to C $\alpha$  backbones. Ordering of the switches was confirmed with a simulated annealing omit map (Figure 20), calculated with a model lacking switches 1-3, with the program CNS (Brunger et al., 1998). The C $\alpha$  backbone model has been deposited with the PDB (accession code 1NT9).

**Table 6.** X-ray diffraction data

Dataset <sup>1</sup>	Native	Rpb4/7 SeMet
Unit cell axis <sup>2</sup> [Å]	220.4, 391.4, 282.2	223.0, 394.5, 284.4
Wavelength [Å]	0.9185	0.9795 (Se peak)
Resolution [Å] (highest shell)	50-4.2 (4.35-4.2)	50-6.5 (6.7-6.5)
Unique reflections (highest shell)	83,431 (7,893)	25,009 (2,452)
Completeness [%] (highest shell)	93.8 (89.5)	99.9 (100.0)
Redundancy	2.9	13.4
I/ $\sigma$ I	6.5	9.4
R <sub>sym</sub> [%] (highest shell)	9.3 (36.3)	7.5 (25.8)
No. of sites	-	13 (5 Se, 8 Zn)

<sup>1</sup>Diffraction data collected at the Swiss Light Source. <sup>2</sup>Space group C222<sub>1</sub>.



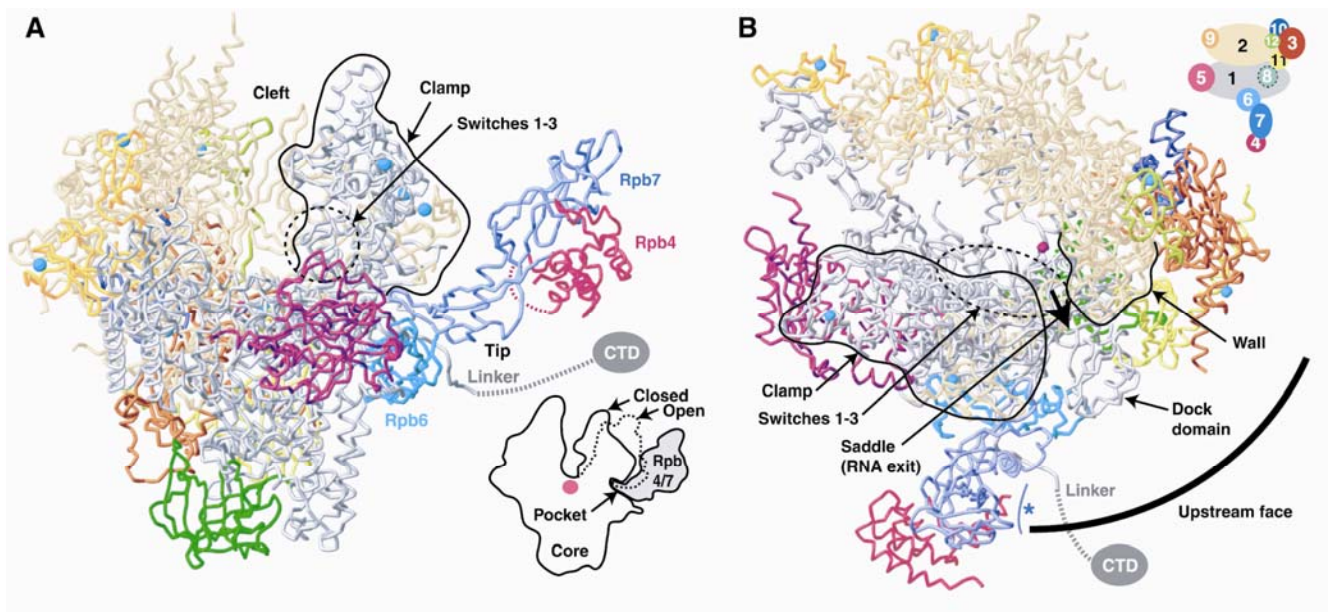
**Figure 16. Structural analysis of the 12-subunit Pol II.** (A) Fit of the Rpb4/7 counterpart structure (Todone et al., 2001) to the initial 4.2 Å difference Fourier map (green). The map was phased with the Pol II core structure, and is contoured at  $2\sigma$ . Ca backbones for Rpb4 and Rpb7 are in pink and blue, respectively. Minor adjustments were made to the structure, to account for yeast-specific sequence features. The view corresponds to the “front” view of Pol II (Cramer et al., 2000). (B) Selenomethionine anomalous difference Fourier (yellow). The Fourier was calculated with anomalous data from a crystal with selenomethionine-labeled Rpb4/7 (Table 1), and phases from the core model, and is contoured at  $6\sigma$ . Selenomethionine was incorporated as described (Budisa et al., 1995). Five selenium peaks with heights between 20.5 $\sigma$  and 10.6 $\sigma$  coincide with the location of methionine side chains. Indicated as yellow spheres are sulfur atoms in methionine side chains, after the corresponding archaeal residues were replaced. No peaks were observed for Rpb4 methionines 1 and 114, which are in flexible regions. Figure prepared with O (Jones et al., 1991).

### Overall structure

The model shows a single Rpb4/7 complex on the outside of the core, protruding from the base of the clamp (Figure 17A). The location of Rpb4/7 agrees with that in cryo-electron microscopic reconstructions of Pol II (Craighead et al., 2002) and Pol I, which reveals the Rpb4/7 counterpart (Peyroche et al., 2002; Bischler et al., 2002). Most of the Rpb4/7 surface is exposed and accessible for interactions with proteins or nucleic acids. Such interactions may influence the Rpb4/7 orientation that is fixed here by crystal packing. Exiting RNA may interact with a potential nucleic acid-binding surface of Rpb7 (Todone et al., 2001; Orlicky et al., 2001) that faces the Pol II “saddle,” from which RNA emerges (Figure 17B).

Rpb4/7 binds to core with the N-terminal ribonucleoprotein-like domain of Rpb7 (Todone et al., 2001), termed here the “tip” (Figure 17A). The remainder of Rpb4/7 extends far from the core surface, explaining why mutations in this region do not impair core binding (Orlicky et al., 2001). Consistent with the core-Rpb7 interaction, Rpb7 alone can bind to core (Sheffer et al., 1999), and Rpb7 is essential for yeast growth (McKune et al., 1993), whereas Rpb4 is not (Woychik and Young, 1989). Deletion of the *rpb4* gene in yeast facilitates dissociation of

Rpb7 from core (Edwards et al., 1991). Our model suggests that loss of the Rpb4-Rpb7 interface upon Rpb4 deletion destabilizes Rpb7 and facilitates Rpb7 dissociation. Indeed, the Rpb4-Rpb7 interface underlies Rpb4/7 integrity, and is conserved (Todone et al., 2001), such that stable chimeric heterodimers with Rpb4 and Rpb7 from various species can be formed (Werner et al., 2000; Sakurai et al., 1999; Larkin and Guilfoyle, 1998). Consistent with our model, Rpb7 overexpression can suppress Rpb4 deletion defects (Sheffer et al., 1999; Tan et al., 2000). In contrast to our findings, Rpb7 dissociation upon Rpb4 deletion was explained earlier by core interaction primarily through Rpb4 (Edwards et al., 1991; Craighead et al., 2002; Khazak et al., 1998).

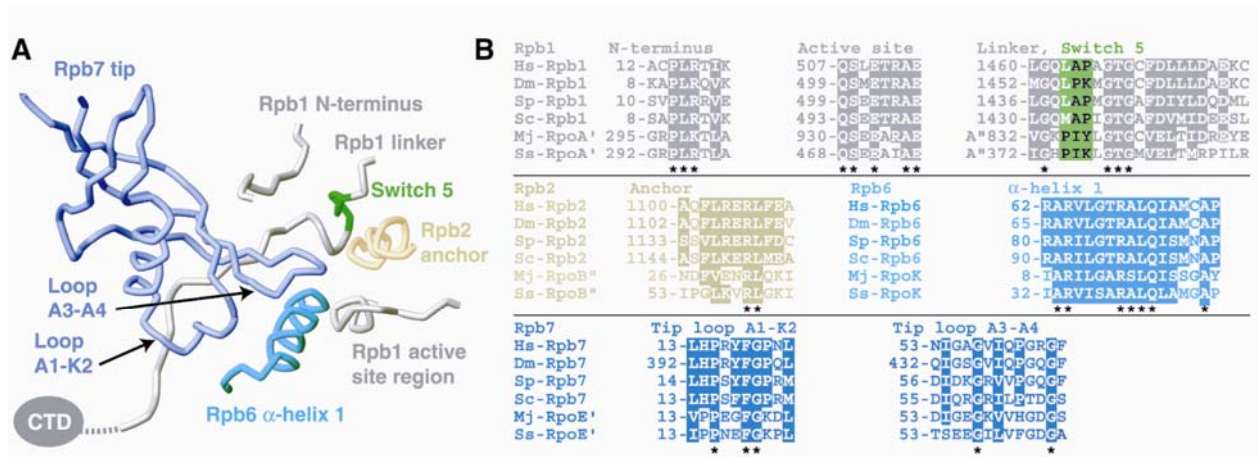


**Figure 17. Architecture of the 12-subunit Pol II, coupling of Rpb4/7 binding and clamp closure, and upstream interaction face.** (A) Ribbon model of Pol II. The view is as in Figure 16. Cyan spheres and a pink sphere depict eight zinc ions and an active site magnesium ion, respectively. A black line encircles the clamp. The linker to the C-terminal repeat domain (CTD) is indicated as a dashed line. In the lower right corner, a schematic cut-away view is shown. A dashed line indicates the open clamp position observed in form 2 of the Pol II core structure (Cramer et al., 2001). (B) Pol II upstream interaction face. View of the model in (A) from the “top” (Cramer et al., 2000). The dashed circle segment is centered at the active site and has a radius that corresponds to the minimal distance between the TATA box and the transcription start site (85 Å, about 25 base pairs). The saddle between the wall and the clamp, and the assumed direction of RNA exit are indicated. A blue asterisk indicates a potential RNA binding face of Rpb7 (Todone et al., 2001; Orlicky et al., 2001). A key to subunit color is shown in the upper right corner. Figures 17, 18 prepared with RIBBONS (Carson, 1997).

**Rpb7 forms a conserved wedge that restrains the clamp**

The Rpb7 tip forms a wedge between the clamp, the linker, and the core subunit Rpb6 (Figure 17A). The tip partially fills a surface “pocket,” which corresponds to the end of the previously identified potential RNA exit groove 1 (Cramer et al., 2000). The pocket is lined by five protein regions, three in Rpb1 and one each in Rpb2 and Rpb6 (Figure 18). Rpb4/7 binding to the pocket thus holds together three subunits, and may stabilize the Pol II subunit assembly. The pocket-Rpb7 interface is partially hydrophobic, and conserved among eukaryotes (Figure 18), explaining why human Rpb4/7 can functionally replace its yeast counterpart (Khazak et al., 1995). A corresponding pocket-tip interaction apparently exists in eukaryotic Pol I and Pol III, and in the archaeal RNA polymerase. The tip domain is conserved in the Rpb7 homologs of Pol I (Shpakovski and Shematorova, 1999; Peyroche et al., 2002), Pol III (Sadhale and Woychik, 1994; Siaut et al., 2003) and archaeal polymerase (Todone et al., 2001). Like Rpb7, the Rpb7 homolog of Pol I binds the common subunit Rpb6 (Peyroche et al., 2002), and can dissociate from core in a Pol I mutant (Smid et al., 1995). The Rpb7 homolog of Pol III is also dissociable, and disruption of its tip domain is lethal (Sadhale and Woychik, 1994). Since core subunits are either shared or homologous in all three eukaryotic enzymes (Sentenac et al., 1992), the Rpb4/7 counterparts’ most likely bind similarly to the cores. Thus the yeast 12-subunit Pol II is a good model for Pol I and Pol III, and for the enzymes in higher eukaryotes, as Pol II subunits are highly conserved in sequence and function (Woychik, 1998).

In the 12-subunit Pol II, the clamp adopts the closed conformation observed in the core elongation complex structure (Gnatt et al., 2001), consistent with electron microscopy in solution (Craighead et al., 2002). Except for the clamp closure, there are no gross changes in core structure upon Rpb4/7 binding. Modeling the clamp in the open states observed in the free Pol II core structures (Cramer et al., 2001) results in a clash with the Rpb7 tip (Figure 17A). Thus, Rpb4/7 can only bind when the clamp is closed, and the clamp can only open after Rpb4/7 has dissociated from the Pol II core. Residual space between the closed clamp and Rpb4/7 only allows for slight changes in the clamp position.



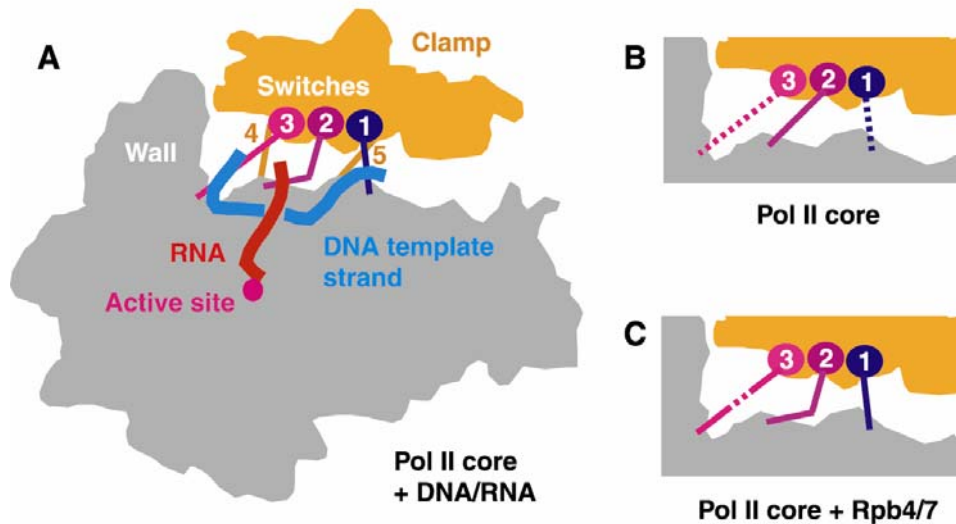
**Figure 18. Pocket-tip interaction.** (A) Ribbon model of the Rpb7 tip binding with its two outermost loops (Todone et al., 2001) to the five protein regions (Cramer et al., 2001) that line the pocket. The view is from the “back,” roughly the inverse of that in Figures 16 and 17A. Colors are as in Figure 17, except that switch region 5 is in green. (B) Sequence alignments of protein regions in (A) (Hs, *Homo sapiens*; Dm, *Drosophila melanogaster*; Sp, *Schizosaccharomyces pombe*; Sc, *Saccharomyces cerevisiae*; Mj, *Methanococcus jannaschii*; Ss, *Sulfolobus sulfataricus*). Conserved residues are highlighted. Stars below the alignments indicate invariant residues.

## Implications for transcription initiation

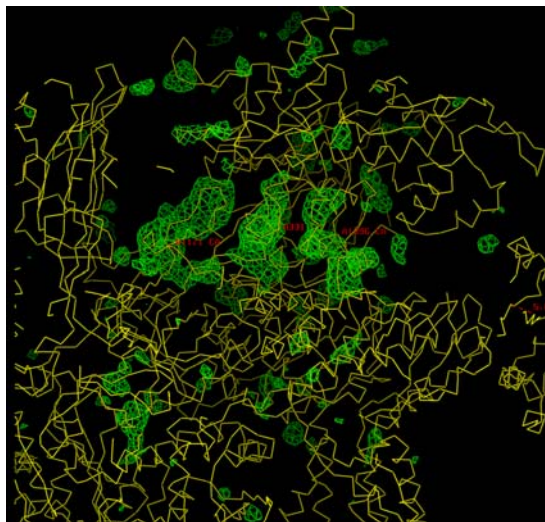
Coupling between Rpb4/7 binding and clamp closure is relevant for transcription initiation, when promoter DNA is loaded onto polymerase. Given that the DNA duplex would be loaded deeply into the cleft, between the clamp and the wall as suggested (Cramer et al., 2001), Rpb4/7 must dissociate from the Pol II core for sufficient clamp opening. However, Rpb4/7 could rebind after DNA melting and clamp closure. Although Rpb4/7 can dissociate from the Pol II core *in vitro* (Edwards et al., 1991), it is unclear if it dissociates *in vivo*. Our model reveals a small core-Rpb7 interface, consistent with a transient interaction, but the core-Rpb4/7 complex is more stable in other species (Sakurai et al., 1999; Larkin and Guilfoyle, 1998; Khazak et al., 1998).

In an alternative scenario for initiation, Rpb4/7 persistently bound to core would prevent promoter entry to the cleft, and DNA could only bind far above the active center. Upon DNA melting, the template strand could however pass the clamp, slip into the cleft, and bind to the site formed by switch regions 1–3, as observed in the core elongation complex (Gnatt et al., 2001). Except a central part in switch 3, around Rpb2 residues 1120–1127, switches 1–3 adopt a similar conformation in the 12-subunit Pol II, and are no longer flexible, as in the free core structures ((Cramer et al., 2001) (Figure 19 and Figure 20). Thus, interaction of Rpb4/7 with the Pol II core induces partial formation of the binding site for the DNA template strand. The flexible part of switch 3 may only get ordered upon template strand binding or upon formation

of an early DNA-RNA hybrid. Induced folding of the central part of switch 3 by the growing hybrid could underlie a transition that stabilizes the early elongation complex.



**Figure 19. Changes in the switch regions and binding of the DNA template strand.** (A) Schematic cut-away view of the Pol II core elongation complex (Gnatt et al., 2001). The view is related to that of Figure 17A by a 90° rotation around a vertical axis. Switches 1-3 face the cleft, are ordered, and interact with the DNA template strand. (B) In the free Pol II core structures, switches 1 and 3 are flexible, and switch 2 is in an extended conformation, allowing for clamp opening (7). (C) In the 12-subunit Pol II, switches 1-3 are ordered and in the same conformation as in (A), except a central region of switch 3 that remains flexible (Rpb2 residues 1120-1127). Ordering of the switches was confirmed with a simulated annealing omit map (CNS, Brunger et al., 1998) calculated from a model lacking switches 1-3 (Figure 20).



**Figure 20. Simulated annealing omit map.** Ordering of switches 1-3 in 12-subunit Pol II, except of a central region of switch 3 that remains flexible (Rpb2 residues 1120-1127). Ordering of the switches was confirmed with a simulated annealing omit map (CNS, Brunger et al., 1998), calculated from a model lacking switches 1-3.

It is likely that promoter loading is topologically similar during bacterial transcription initiation, which requires only the  $\sigma$  factor and the core polymerase. The structural conservation of bacterial and eukaryotic core RNA polymerases (Darst, 2001; Cramer, 2002) enables a comparison of our model with recent structures of bacterial RNA polymerase bound

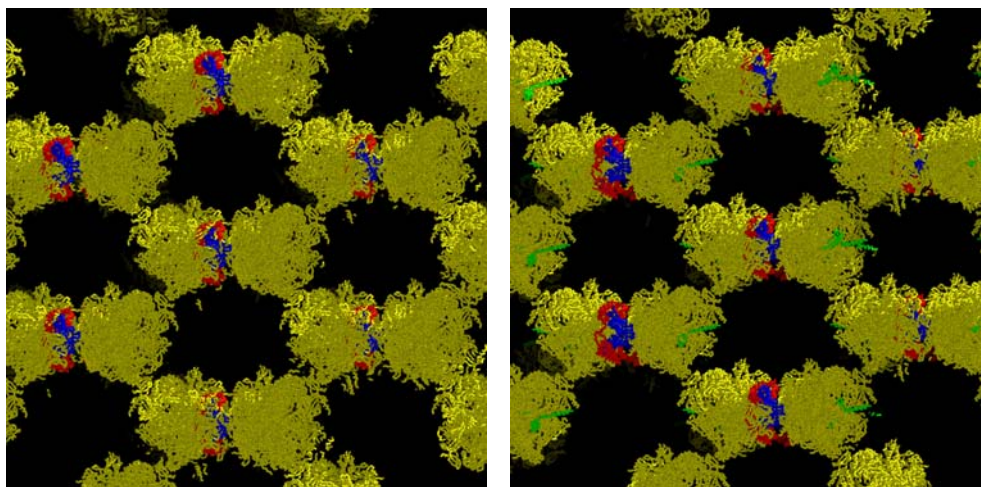
to  $\sigma$  (Murakami et al., 2002; Vassylyev et al., 2002), and bound to  $\sigma$  and upstream promoter DNA (Murakami et al., 2002). In these structures, the  $\sigma$  factor interacts with regions corresponding to the Pol II upstream face, where TFIIB, TFIIF and TBP are proposed to assemble. Although the  $\sigma$  factor shows sequence similarity to Rpb4 (Woychik and Young, 1989), it is not a functional counterpart of Rpb4/7, since Rpb4/7 and  $\sigma$  bind to different sites on the polymerase surfaces. The structure of bacterial polymerase with bound promoter shows a closed clamp and ordered switches, as observed in our model of the initiation-competent Pol II. Additionally, the bacterial complex shows promoter DNA outside the cleft, consistent with the ladder model suggested above.

The specific occurrence of Rpb4/7 in higher cells is reflected in its functions during transitions within the transcription cycle. During initiation, the Pol II CTD gets phosphorylated, and remains phosphorylated during RNA elongation. Dephosphorylation of the CTD is however required for Pol II recycling after termination, since only unphosphorylated Pol II can rejoin an initiation complex. CTD dephosphorylation is carried out by the phosphatase Fcp1 (Lin et al., 2002). Fcp1 is apparently recruited to the phosphorylated CTD by Rpb4/7, since Rpb4 binds Fcp1 (Kimura et al., 2002), and Rpb7 binds the linker to the CTD, which is disordered in our crystals (Figures 17, 18). Fcp1 inhibits initiation complex assembly (Chambers et al., 1995), maybe because Fcp1 and a general factor bind to the upstream face in a mutually exclusive manner, ensuring complete CTD dephosphorylation before transcription reinitiation.

### **Crystal packing in 12-subunit RNA Pol II facilitates obtaining of higher order complexes**

12-subunit RNA Pol II crystals show different packing in comparison to 10-subunit core. Looser packing results in unusually high solvent content of 80 % in comparison to core Pol II crystals where molecules are more tightly packed and therefore have around 50 % of solvent (Cramer et al., 2000). This packing in 12-subunit Pol II crystals results in stronger resolution/crystal size dependence due to substantially reduced number of scattering molecules per volume of the crystal. On the one hand, this dependency raised some technical problems, on the other it opened new possibilities for obtaining new, exciting, higher-order complexes of Pol II with additional factors. The structure of RNA polymerase II in complex with TFIIS (Kettenberger et al., 2003) was solved after 12-subunit Pol II crystals were incubated with recombinant, overexpressed in *E.coli* TFIIS. The very large solvent channels of the crystals allowed TFIIS entry and binding to its specific site on Pol II, which was not obstructed by crystal contacts (Figure 21). Although TFIIS binding induced large

conformational changes in crystal lattice the crystals still diffracted synchrotron radiation to 3.8 Å.



**Figure 21.**  
**Crystal lattice of**  
**12-subunit (left)**  
**and 12-subunit**  
**with TFIIS (right).**  
*Yellow: core Pol II*  
*Blue: Rpb7,*  
*Red: Rpb4,*  
*Green: TFIIS*

## Conclusion

The 12-subunit Pol II model explains and suggests functional roles of the Rpb4/7 complex, and extends our understanding of eukaryotic transcription towards the mechanism of initiation. Our work further shows that the architecture of a dissociable multiprotein complex can be determined by X-ray analysis even at moderate resolution. The methods used and developed here can be applied to complexes of Pol II with proteins and nucleic acids, to address mechanistic questions raised by the complete Pol II model.

# Chapter III

*Structures of complete RNA polymerase II and  
its subcomplex Rpb4/7*

## Abstract

I determined the x-ray structure of the RNA polymerase II (Pol II) subcomplex Rpb4/7 at 2.3 Å resolution, combined it with a previous structure of the ten-subunit polymerase core, and refined an atomic model of the complete 12-subunit Pol II at 3.8 Å resolution. Comparison of the complete Pol II structure with structures of the Pol II core and free Rpb4/7 shows that the core-Rpb4/7 interaction goes along with formation of an  $\alpha$ -helix in the linker region of the largest Pol II subunit, and with folding of the conserved Rpb7 tip loop. Details of the core-Rpb4/7 interface explain facilitated Rpb4/7 dissociation in a temperature-sensitive Pol II mutant, and specific assembly of Pol I with its Rpb4/7 counterpart A43/14. The refined atomic model of Pol II serves as the new reference structure for analysis of the transcription mechanism, and enables structure solution of complexes of the complete enzyme with additional factors and nucleic acids by molecular replacement.

## Introduction

Synthesis of mRNA in eukaryotic nuclei is carried out by RNA polymerase II (Pol II). The mechanism of Pol II transcription has been investigated in many biochemical and detailed structural studies (for recent reviews see Cramer, 2004; Hahn, 2004; Asturias, 2004; Woychik and Hampsey, 2002). Pol II consists of a ten-polypeptide catalytic core and the heterodimeric Rpb4/7 complex, which can dissociate from the yeast enzyme (Edwards et al., 1991). Refined structures were reported for the Pol II core in free form (Cramer et al., 2000; Cramer et al., 2001), in form of a minimal elongation complex (Gnatt et al., 2001), and in complex with the toxin  $\alpha$ -amanitin (Bushnell et al., 2002). Crystallographic models of the Pol II core bound to the initiation factor TFIIB, and bound to a synthetic DNA-RNA hybrid were published recently (Westover et al., 2004; Bushnell et al., 2004).

For the complete 12-subunit Pol II, backbone models were reported for the free enzyme (Armache et al., 2003; Bushnell et al., 2003), and for a complex with the elongation factor TFIIS (Kettenberger et al., 2003). Our previous Pol II model was derived by fitting structures of the core polymerase and a distantly related archaeal Rpb4/7 counterpart (Todone et al., 2001) to a 4.2 Å crystallographic electron density map, guided by selenomethionine markers (Armache et al., 2003). The model showed the overall architecture of the enzyme, but did not reveal details of the Rpb4/7 structure or its interaction with the Pol II core. Rpb4/7 was seen protruding from the polymerase upstream face, where initiation factors assemble for promoter loading. Rpb7 formed a wedge underneath the mobile Pol II clamp, apparently restricting the clamp to a closed position, and preventing entry of a DNA duplex into the active center cleft.

I present here the refined atomic model of the complete Pol II at 3.8 Å resolution. To derive this model I determined the crystal structure of the free Rpb4/7 subcomplex at 2.3 Å resolution. Comparison of the complete Pol II structure with the structures of the core enzyme and the Rpb4/7 subcomplex reveals folding events that accompany formation of Pol II. The refined complete Pol II structure provides the new reference for functional analysis of the eukaryotic transcription machinery. The structure also enables further structural studies of scarce complexes of the complete Pol II, such as a recently described complex with DNA, RNA, and the elongation factor TFIIS (Kettenberger et al., 2004).

## **Materials and methods**

### **Design of Rpb4/7 variants**

Crystallization of full length Rpb4/7 did not yield any crystals, so we designed seven deletion variants that varied in either the length of the N-terminus or in the length of deleted loop region in Rpb4. Rpb7 containing C-terminal His-tag was not mutated and remained the original sequence. This strategy is often done in protein crystallography when full length proteins contain unstructured regions which are usually flexible and can hinder crystallization. Design of deletion variants was based on results from a proteolytic digest with trypsin and chymotrypsin, on prediction of low complexity regions, on sequence alignments of yeast Rpb4/7 complex and archaeal E/F homologue, on secondary structure predictions, and finally and maybe most importantly on inspection of the electron density maps of backbone model of 12-subunit RNA polymerase II.

### **Limited proteolysis**

Full length Rpb4/7 was used as a template protein for limited proteolysis with trypsin and chymotrypsin. 150  $\mu$ l of 4 mg/ml of protein was supplemented with 1mM CaCl<sub>2</sub> and 0.5  $\mu$ l of 1 mg/ml particular protease was added. The reaction took place at 37°C and was monitored in a time course manner. Samples of 20  $\mu$ l were taken after 1, 3, 5, 10, 30, 60 and 90 minutes and immediately denatured at 92°C with SDS to stop reaction, put on 15 % SDS-PAGE gel and the fragments were analyzed with Edmann sequencing.

### **Cloning of Rpb4/7 variants**

N-terminal truncation constructs were amplified by PCR from full length Rpb4/7 (names of variants and primers contain the prefix 'Sc4' or 'Sc7' standing for Rpb4 or Rpb7 from *S. cerevisiae* (Table 1 lists primers used). Loop deletion constructs were amplified by PCR from construct Sc4/34 using 2-step PCR strategy (Table 2 lists primers used). For the ligation of each construct into the vector pET-21d all primers carried either NcoI (forward) or NotI (reverse) restriction sites and were inserted into pET21d using these restriction sites. Table 3 lists sequencing primers that were used to confirm the success of cloning.

**Table 1.** Primers used for N-terminal truncation variants of Rpb4/7

Construct	Primer Name	direction	Primer
Sc4/34	Sc4/34NcoI	forward	5'-TGCTAGCTCCCATGGCGCAGCTGAAACAGATAAAATC-3'
	Sc7/NotI	reverse	5'-CTGCTAGCTGCGGCCGCAATAGCACCCAAATAATCTTCTTTG-3'
Sc4/66	Sc4/66NcoI	forward	5'-TGCTAGCTCCCATGGCGCGTAGGAGAGCATTTAAAAGATCG-3'
	Sc7/NotI	reverse	5'-CTGCTAGCTGCGGCCGCAATAGCACCCAAATAATCTTCTTTG-3'
Sc4/83	Sc4/83NcoI	forward	5'-TGCTAGCTCCCATGGCGAAGCACGAAAACGCCAATGATG-3'
	Sc7/NotI	reverse	5'-CTGCTAGCTGCGGCCGCAATAGCACCCAAATAATCTTCTTTG-3'

**Table 2.** Primers used for loop deletion variants of Rpb4/7

Construct name	PCR		Primer Name	Primer sequence
Loop1	1	a	11p4	5'-CCCTAGTTTCAGAATGCATAAAATTGCGATCTTTTAAATGCTC-3'
			Sc4/34NcoI	5'-TGCTAGCTCCCATGGCGCAGCTGAAACAGATAAAATC-3'
		b	11p2	5'-GAGCATTTAAAAGATCGCAATTTATGCATTCTGAAACTAGGG-3'
			Sc7/NotI	5'-CTGCTAGCTGCGGCCGCAATAGCACCCAAATAATCTTCTTTG-3'
	2		Sc4/34NcoI	5'-TGCTAGCTCCCATGGCGCAGCTGAAACAGATAAAATC-3'
			Sc7/NotI	5'-CTGCTAGCTGCGGCCGCAATAGCACCCAAATAATCTTCTTTG-3'
Loop2	1	a	12p4	5'-CCCTAGTTTCAGAATGCATAAACAATGCTTCTTCTTGTTTGTGCG-3'
			Sc4/34NcoI	5'-TGCTAGCTCCCATGGCGCAGCTGAAACAGATAAAATC-3'
		b	12p2	5'-CGCAAAAAAAAAACACAAGAAGAAGCATTTGTTTATGCATTCTGAAACTAGGG-3'
			Sc7/NotI	5'-CTGCTAGCTGCGGCCGCAATAGCACCCAAATAATCTTCTTTG-3'
	2		Sc4/34NcoI	5'-TGCTAGCTCCCATGGCGCAGCTGAAACAGATAAAATC-3'
			Sc7/NotI	5'-CTGCTAGCTGCGGCCGCAATAGCACCCAAATAATCTTCTTTG-3'
Loop3	1	a	13p4	5'-GCGTTGACGTCATCTTCATCCAGCAAATGCTTCTTCTTGTTTGTGCG-3'
			Sc4/34NcoI	5'-TGCTAGCTCCCATGGCGCAGCTGAAACAGATAAAATC-3'
		b	13p2	5'-CGCAAAAAAAAAACACAAGAAGAAGCATTTGCTGGATGAAGATGACGTCAACGC-3'
			Sc7/NotI	5'-CTGCTAGCTGCGGCCGCAATAGCACCCAAATAATCTTCTTTG-3'
	2		Sc4/34NcoI	5'-TGCTAGCTCCCATGGCGCAGCTGAAACAGATAAAATC-3'
			Sc7/NotI	5'-CTGCTAGCTGCGGCCGCAATAGCACCCAAATAATCTTCTTTG-3'
Loop4	1	a	14p4	5'-GCGTTGACGTCATCTTCATCCAGTTGCGATCTTTTAAATGCTC-3'
			Sc4/34NcoI	5'-TGCTAGCTCCCATGGCGCAGCTGAAACAGATAAAATC-3'
		b	14p2	5'-GAGCATTTAAAAGATCGCAACTGGATGAAGATGACGTCAACGC-3'
			Sc7/NotI	5'-CTGCTAGCTGCGGCCGCAATAGCACCCAAATAATCTTCTTTG-3'
	2		Sc4/34NcoI	5'-TGCTAGCTCCCATGGCGCAGCTGAAACAGATAAAATC-3'
			Sc7/NotI	5'-CTGCTAGCTGCGGCCGCAATAGCACCCAAATAATCTTCTTTG-3'

**Table 3.** Primers used for sequencing of Rpb4/7 constructs

<i>Primer Name</i>		<i>Primer</i>
Sc4sequ402f	forward	5'-GGGAGGAAATAATAAAGATTTG-3'
Sc4sequ499r	reverse	5'-GAAGCTGTATAACTGCCCCGA-3'
T7r	reverse	5'-GCTAGTTATTGCTCAGCGG-3'

### Expression and purification of Rpb4/7 variants

All Rpb4/7 variants were expressed and purified exactly in the same manner like full length Rpb4/7. Test expressions were done in 1l cultures and the cells were lysed with French Press, resuspended in Freezing Buffer and centrifuged at 4°C. The pellet was treated with 6 M urea, centrifuged and supernatant was loaded on SDS-PAGE to check if and how much protein was in inclusion bodies. Expression of SeMet full length Rpb4/7 and loop3 variant was done as described previously and purification procedure did not change for these proteins. Variant loop3 crystallized in the same conditions as the native protein and the crystals were of similar size and quality.

### Crystallization of Rpb4/7 variants Sc4/34 and Loop3

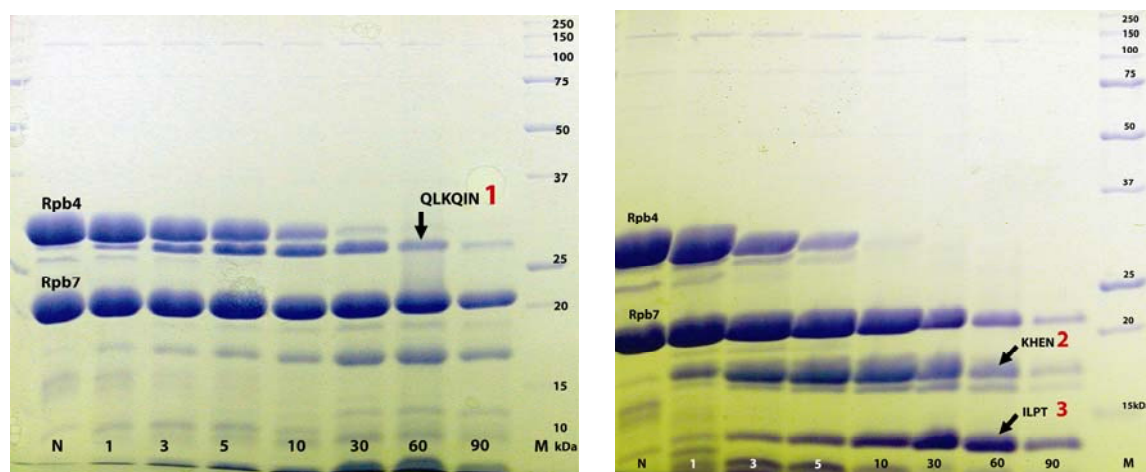
For crystallization of Sc4/34 and Loop3 sparse matrix method of trial conditions was used, which is based on known crystallization conditions (Jancarik and Kim, 1991). In this way one can test in a short time wide ranges of pH, salts and precipitants using small amounts of sample. Results from these initial trials can produce crystals or solubility information. For crystallization of the Rpb4/7 variants the hanging drop method was used. Therefore, 1 µl of the purified protein (solutions contained 9, 11 and 18 mg/ml protein in case of Sc4/34 and 3 mg/ml in case of Loop3 variant) was set with on a plastic cover slide and mixed with 1 µl of reservoir solution. The slide was set upside down over a well with reservoir solution and sealed to the reservoir edge using high vacuum grease. As the drop contains lower concentration of crystallization agent, water vapour will leave the drop and the concentration will increase until supersaturation and crystal growth is reached. Crystallization took place at 20°C and the plates were examined under a microscope few days after setting the drops for presence of crystals. Crystals were harvested with a mother solution in which they grew and several different cryoprotectants were carefully screened to give an idea which one is the most suitable. The cryoprotectants were introduced in a stepwise fashion and quick freezing at 4°C and at 20°C was tried.

## Results and discussion

### Design of Rpb4/7 variants

#### Proteolytic digest

Limited proteolysis was done by incubation of 150  $\mu$ l of 4 mg/ml solution of native Rpb4/7. 1 mM  $\text{CaCl}_2$  and 0.5  $\mu$ l of 1 mg/ml protease (trypsin or chymotrypsin) at 37°C and 10  $\mu$ l samples were taken after 1, 3, 5, 10, 30, 60 and 90 minutes and immediately denatured to inactivate the protease. After all samples were collected they were loaded onto SDS-PAGE (Figure 1) and the bands representing the parts of the protein that seemed stable were excised from the gel and analyzed by Edmann sequencing. Chymotrypsin cuts after aminoacid number 33 and trypsin after aminoacid 82 in Rpb4 and these immediately suggested the variant design for Rpb4. Trypsin also cut in Rpb7 after aminoacid 60 but as we knew this part of Rpb7 was not supposed to be manipulated with since it interacts with RNA Pol II.



**Figure 1. 15% SDS-PAGE of chymotrypsin (left) and trypsin (right) digest of Rpb4/7.** *N*: native, control. *Rpb4/7*, 1: 1 min., 2: 2 min., 5: 5 min., 10: 10 min., 30: 30 min., 60: 60 min., 90: 90 min., *M*: marker, arrows indicate bands representing cuts which were N-terminally sequenced (with sequences shown above the respective bands). These sequences are depicted with red numbers and they represent 3 distinct digestion species after proteolytic digest.

The cuts detected by proteolytic digest coincided well with the low complexity regions detected by SEG (Figure 2) and secondary structure prediction (Figure 3) for these parts of Rpb4 was indicating rather flexible loop regions.

## SEG results

```

1  MNVSTSTFQT  xxxxxxxxxxxx  xxNAATLQLG  QEFQLKQINH  QGEEEEELIAL
51  NLSEARLVIK  EALVERRRAF  KRSQxxxxxxxx  xxxxxENAxxx  xxxxxxxxxxxx
101 xxxxxxxxxxxx  xxFMHSETRE  KELESIDVLL  EQTTGGNNKD  LKNTMQYLTN
151 FSRFRDQETV  GAVIQLLKST  GLHPFEVAQL  GSLACDTADE  AKTLIPSLNN
201 KISDDELERI  LKELSNLETL  Y

```

**Figure 2. Prediction of low complexity regions in Rpb4.** (*SEG*, Wootton and Federhen, 1996)  
Red crosses mark the regions which appear to be unstructured

## PHD results

PHD sec: **H=helix**, **E=extended (sheet)**, blank=other (loop)

Rel sec: reliability index for PHDsec prediction (0=low to 9=high)

[illegible][illegible]

```

.....13.....14.....15.....16.....17.....18
AA      KELESIDVLLQTTGGNNKDLKNTMQYLTNFSRFRDQETVGAVIQLLKSTGLHPFEVAQL
PHD sec  HHHHHHHHHHHHHH HHHHHHHHHHHHHHHHHHHH HHHHHHHHHHHHHH HHHHHHHH
Rel sec  9975556898413243489999999999999998626169999999998616549999999

```

**Figure 3. Secondary structure prediction in Rpb4.** (Rost and Sander,1993)

Red H indicates helices, blue E: sheets, black: loops, below the secondary structure prediction are numbers indicating from 0-9 the reliability of prediction

**Table 4.** Rpb4/7 variants

<b>Variant name</b>	<b>Description</b>	<b>Amino acid residues</b>	<b>Total number of amino acids</b>	<b>Molecular weight kDa</b>
Sc4/34	N-terminal deletion of Rpb4 before aa No. 34	Rpb4: 34 – 221 Rpb7: full length	Rpb4: 188 aa Rpb7: 171 aa	Rpb4: 21,5 Rpb7: 19
Sc4/66	N-terminal deletion of Rpb4 before aa No. 66	Rpb4: 66 – 221 Rpb7: full length	Rpb4: 156 aa Rpb7: 171 aa	Rpb4: 17,8 Rpb7: 19
Sc4/83	N-terminal deletion of Rpb4 before aa No. 83	Rpb4: 83 – 221 Rpb7: full length	Rpb4: 139 aa Rpb7: 171 aa	Rpb4: 15,6 Rpb7: 19
Loop1	Loop deletion of Rpb4 from K75 till D112	Rpb4: 34-74, 113-221 Rpb7: full length	Rpb4: 150 aa Rpb7: 171 aa	Rpb4: 17,1 Rpb7: 19
Loop2	Loop deletion of Rpb4 from K83 till D112	Rpb4: 34-82, 113-221 Rpb7: full length	Rpb4: 158 aa Rpb7: 171 aa	Rpb4: 18,2 Rpb7: 19
Loop3	Loop deletion of Rpb4 from K83 till D100	Rpb4: 34-82, 101-221 Rpb7: full length	Rpb4: 170 aa Rpb7: 171 aa	Rpb4: 19,5 Rpb7: 19
Loop4	Loop deletion of Rpb4 from K75 till D100	Rpb4: 34-74, 101-221 Rpb7: full length	Rpb4: 162 aa Rpb7: 171 aa	Rpb4: 18,4 Rpb7: 19

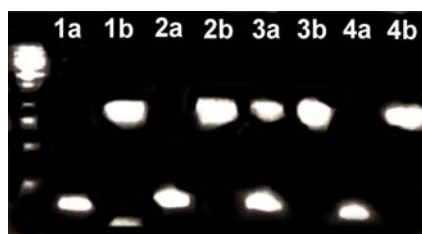
The choice of N-terminal truncations and loop deletions was clear from the start, but the proper combination of both was discovered upon expressions and purifications of individual variants.

### Cloning of Rpb4/7 variants

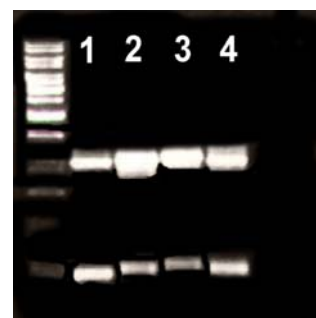
Figures 4-6 show 1 % agarose gels with PCR products.



**Figure 4. Three N-terminal deletion variants after PCR.**  
 1 = Sc4/34  
 2 = Sc4/66  
 3 = Sc4/83



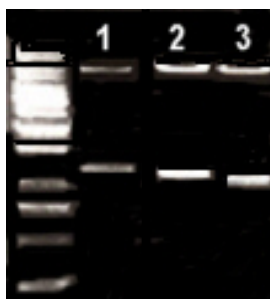
**Figure 5. Loop deletions after 1<sup>st</sup> PCR.**  
 1a = Loop1, PCRa; 1b = Loop1, PCR1b  
 2a = Loop2, PCRa; 2b = Loop2, PCR2b  
 3a = Loop3, PCRa; 3b = Loop3, PCR3b  
 4a = Loop4, PCRa; 4b = Loop4, PCR4b



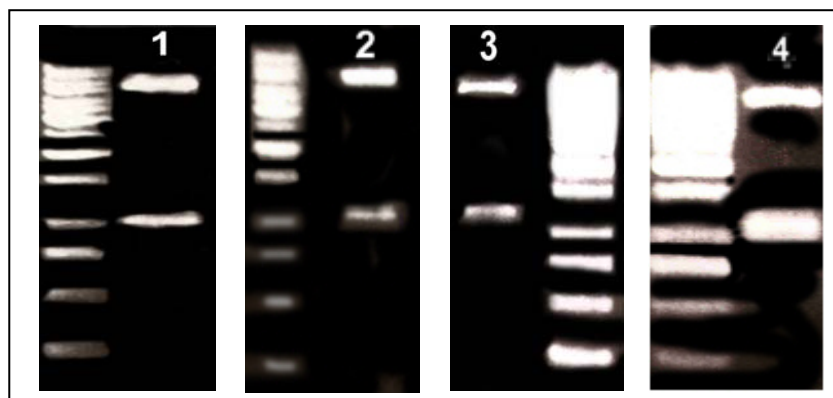
**Figure 6. Loop deletions after 2<sup>nd</sup> PCR.**  
 1 = Loop1; 2 = Loop2  
 3 = Loop3; 4 = Loop4

All variants were cloned into pET-21d vector and transformed to XL-1blue cells for DNA amplification and BL 21(DE3) for protein expression.

The incorporation of construct into vector was checked by test restriction and sequencing (Figures 7 and 8)



**Figure 7. Three N-terminal deleted variants after test restriction.**  
 Upper band = vector  
 1 = Sc4/34  
 2 = Sc4/66  
 3 = Sc4/83

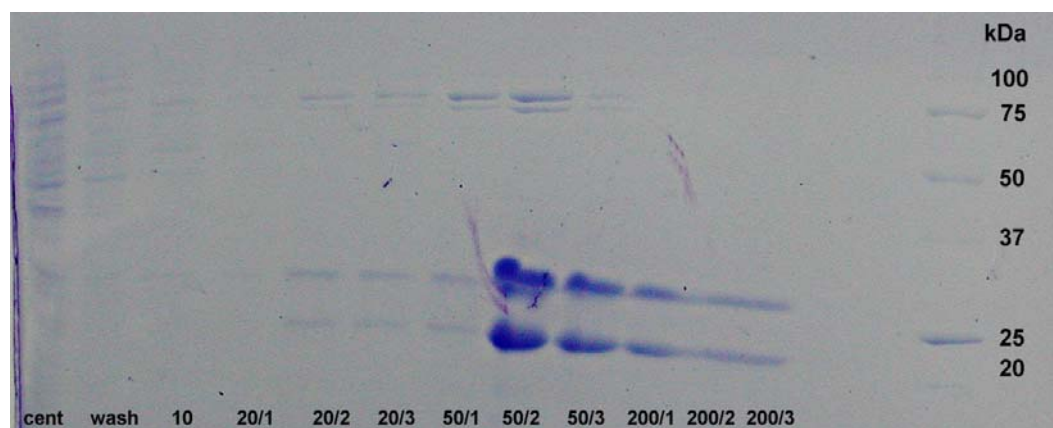


**Figure 8: Loop deletions after test restriction.**  
 Upper band = vector; 1 = Loop1; 2 = Loop2; 3 = Loop3; 4 = Loop4

Variants were cloned using the same double expression strategy, which was used for full-length Rpb4/7 construct with a T7 promoter in front of each ORF and C-terminal His-tag on Rpb7.

### Solubility test of Rpb4/7 variants

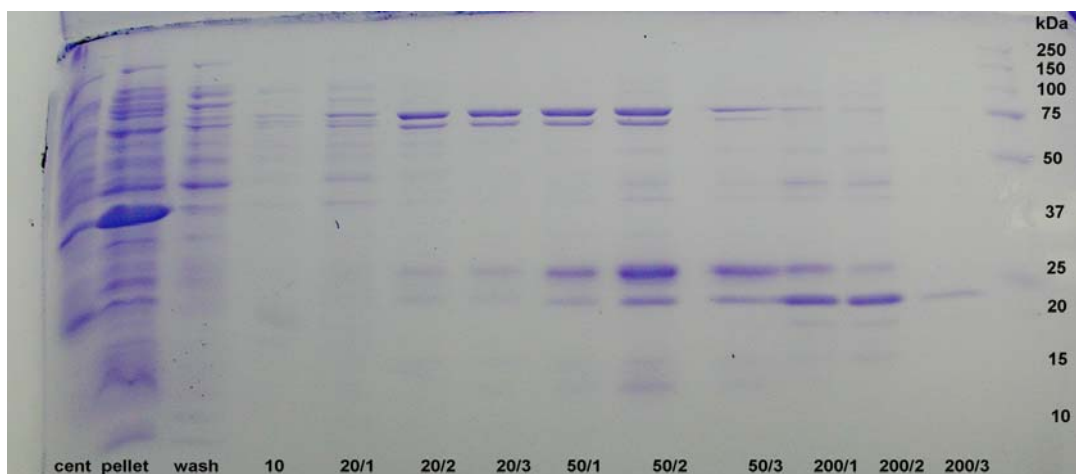
Test expressions were carried out to find the Rpb4/7 variant showing optimal expression, solubility and stoichiometry (amount of Rpb4 and Rpb7 should be equal to ensure proper behaviour during crystallization). All variants were expressed in 1l LB medium and purified over His-tag by Ni-NTA affinity chromatography. The optimal imidazole concentration of the elution mixture was tested and determined by a step-gradient of 10, 20, 50 and 200 mM imidazole. As elution occurred mostly at 50 or 200 mM, and these 2 steps (and sometimes also the elution step with 20 mM) were subdivided again. Instead of using once the full amount of elution mixture, twice or three times the half or the third, respectively, were applied onto the column, so that one could determine if the protein gets eluted early or late under this condition. The fractions were then checked on SDS-PAGE for their concentration and purity. Variant Sc4/34, lacking the first 33 amino acids at the N-terminus was well expressed and soluble, while the other two N-terminal deletions, Sc4/66 and Sc4/83 were expressed, but appeared not to be stable under high concentrations of imidazole (the amount of Rpb7 seems to increase, whereas Rpb4 is degraded). Figures 9-11 show SDS-PAGE gels of test expressions of variants Sc4/34, Sc4/66 and Sc4/83 respectively.



**Figure 9. 15 % SDS-PAGE of Sc4/34 1l expression after Ni-NTA purification.**

*45 µl protein solution were put on the gel;*

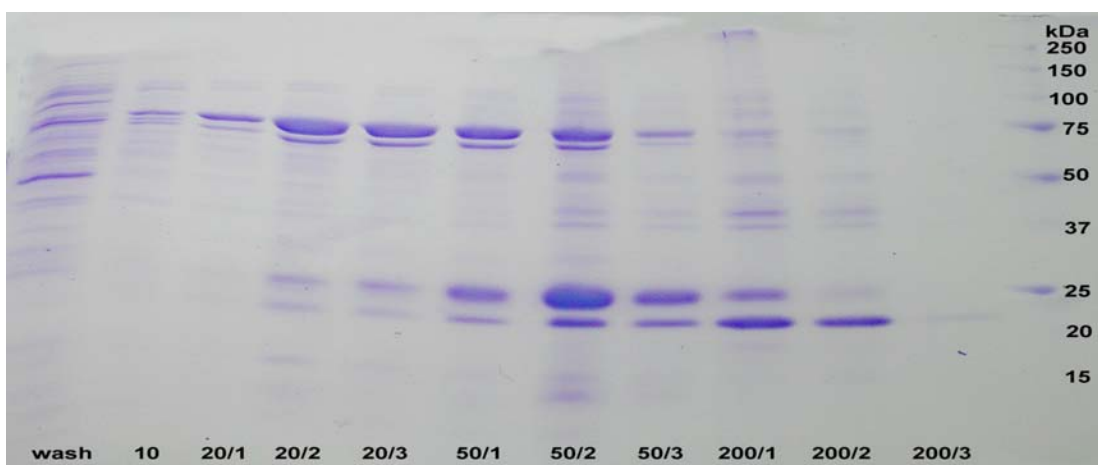
*Cent: after centrifugation (clear lysate); Numbers indicate the concentration of imidazole*



**Figure 10. 15 % SDS-PAGE of Sc4/66 II expression after Ni-NTA purification.**

*45  $\mu$ l protein solution were put on the gel;*

*Cent: after centrifugation (clear lysate); Numbers indicate the concentration of imidazole*



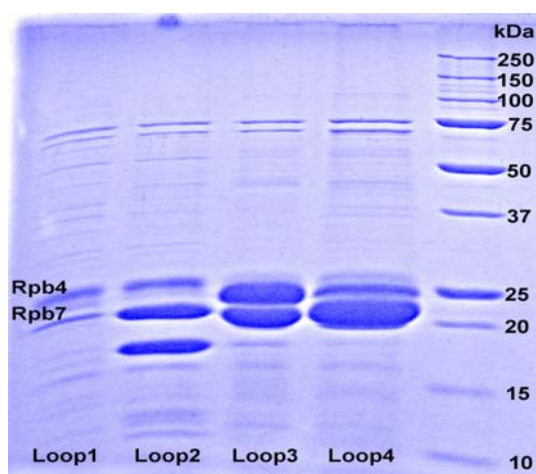
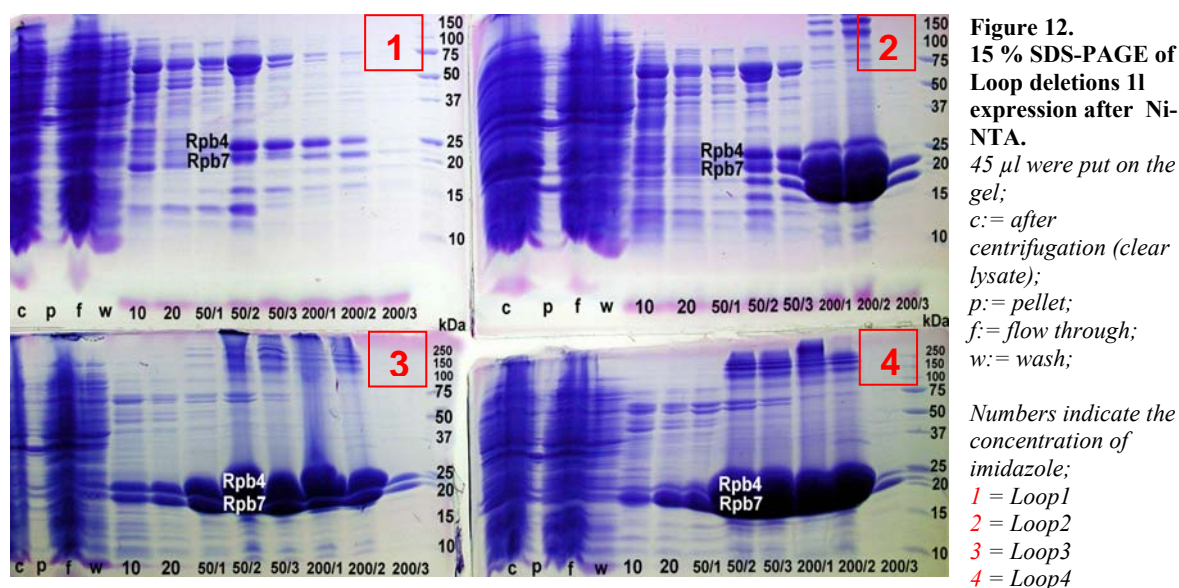
**Figure 11. 15 % SDS-PAGE of Sc4/83 after Ni-NTA purification.**

*45  $\mu$ l protein solution were put on the gel;*

*Numbers indicate the concentration of imidazole*

It was clear from these test expressions that variant Sc4/34 was very well expressed, the two subunits seemed to form a stable dimer and elute as intact stoichiometric complex. In case of Sc4/66 the dimer is apparently destabilized and what is seen on the SDS-PAGE, Rpb4 is eluted already at lower imidazole concentrations and Rpb7 is still bound to the column matrix and elutes at higher concentrations of imidazole. Exactly the same behaviour which is disfavoured for future experiments was seen for variant Sc4/83. Conclusively, from the N-terminal deletion variants, Sc4/34 was chosen for further experiments and all the loop deletion variants were based on it, so that every loop variant was shortened both at the N-terminus and at the loop.

Test expression and Ni-NTA affinity chromatography were performed in exactly the same manner for loop variants as for N-terminal truncation mutants and the fractions loaded on SDS-PAGE to check for solubility and behaviour of individual variant (Figures 12 and 13).



Following observation could be made after the analysis of test expressions and affinity chromatography purification. Loop1, containing the longest deleted loop had the lowest expression levels in comparison with the other loop deletions. Loop2 was well expressed, but showed strange shifted bands, which can be accounted for destabilization and degradation of Rpb4. Loop3 containing the shortest deleted loop, was expressed in extremely high amounts and was also soluble, and dimer showed proper stoichiometry and appeared to be stable. Loop4 was also expressed in extremely high amounts, but the dimerization of Rpb4 and Rpb7 seemed to be destabilized in this variant what can be seen on the SDS-PAGE gel where it appears that the band from Rpb7 seemed to be much stronger than the one from Rpb4 (Table

5). Consequently, from the loop deletion variants, variant Loop3 was taken for further expression and crystallization.

**Table 5.** Summary of all the information about different variants of Rpb4/7 that were created during this work.

Variant name	Expression	features
Sc4/34	++	Soluble, stable
Sc4/66	+	Not stable, higher amount of Rpb7
Sc4/83	+	Not stable, higher amount of Rpb7
Loop1	+	Soluble, stable
Loop2	++	Bands are shifted
Loop3	+++	Soluble, stable
Loop4	+++	Soluble, higher amount of Rpb7

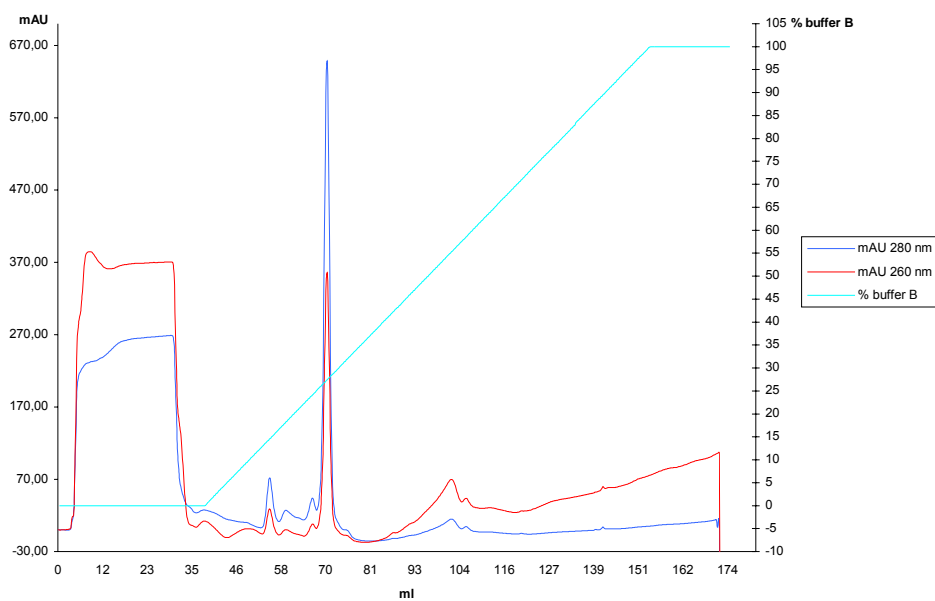
+ indicates expression

++ indicates high expression

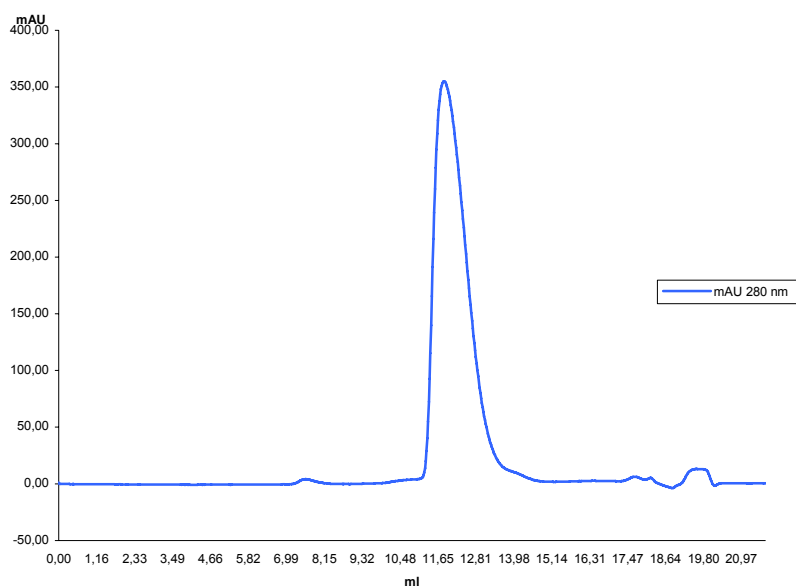
+++ indicates excellent expression

### Expression, purification, and crystallization of Rpb4/7 variants

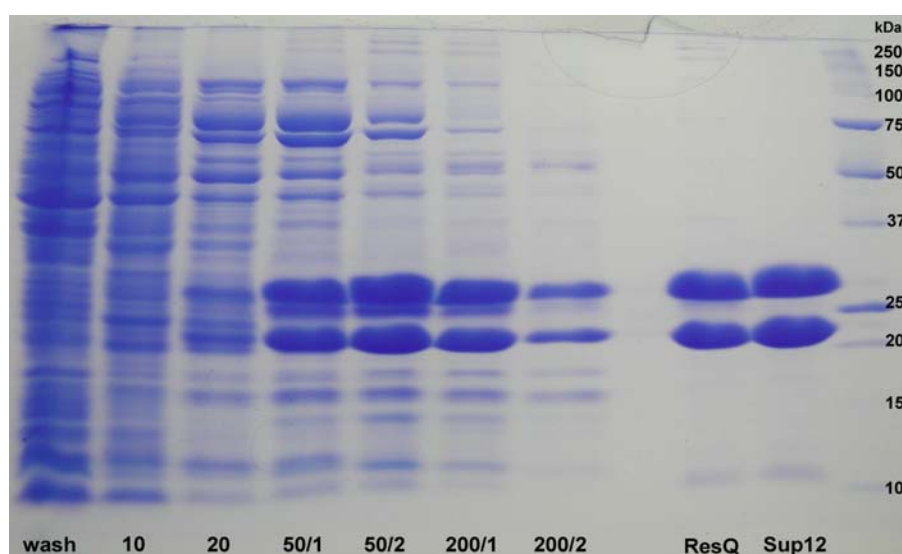
After a successful 1l test expression, Sc4/34 was expressed in 8l LB medium to get reasonable amounts for crystallization and purified by Ni-NTA gravity flow affinity chromatography. The optimal imidazole concentration of the elution solution was tested and determined by a step gradient of 10, 20, 50 and 200 mM of imidazole, and the profile was very similar to the one for full-length Rpb4/7 complex. The elution occurred mostly at the concentration of imidazole between 20–200 mM. The concentration of protein in these fractions was estimated with Bradford reagent and the purity was checked on SDS-PAGE where samples from these fractions were loaded (Figure 16). The 50 and 200 mM imidazole peak fractions were pooled together diluted and applied on the anion exchange chromatography ResourceQ column (Figure 14 and 16). The last and final step of purification was Superose 12 gel filtration and at this step Pol II buffer was used (Figure 15 and 16).



**Figure 14. ResourceQ anion exchange chromatography of Sc4/34**  
*Blue: absorption at 280 nm, red absorption at 260 nm, green: salt gradient,*



**Figure 15. Superose 12 gel filtration chromatography of Sc4/34.**  
*blue: absorption at 280 nm*



**Figure 16. 15 % SDS-PAGE of Sc4/34 after Ni-NTA, anion exchange chromatography and gel filtration.**

*Numbers indicate the concentration of imidazole*

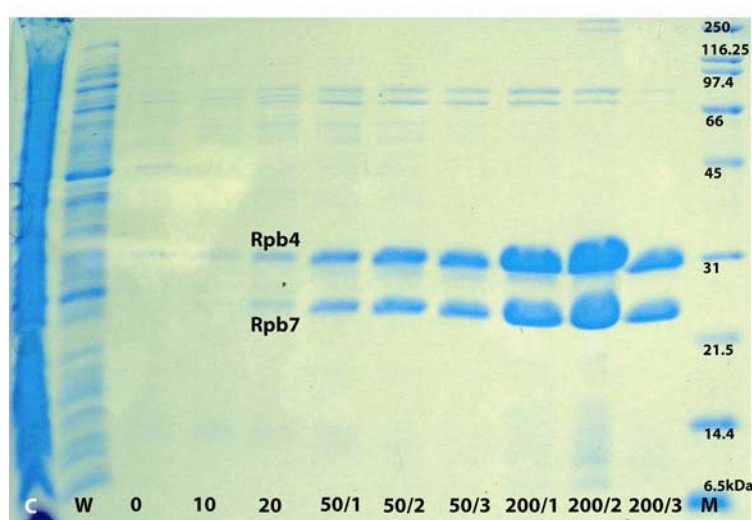
*ResourceQ:= peak fractions from ResourceQ column (anion exchanger);*

*Superose 12:= peak fractions from Superpose 12 (gel filtration)*

Crystallization with variant Sc4/34 were carried out with the hanging drop vapour diffusion method. Establishment of crystallization conditions was done by sparse matrix sampling using commercially available screens from Hampton. After crystals were grown, promising conditions were refined by variation of salts, pH, PEG and additives in the reservoir. Unfortunately the crystals were destroyed during freezing and refinement screens did not come up with satisfying results. Since I could obtain crystals by deletion of the flexible N-terminus, thus the possibility for getting crystals seemed to be even better, when a variant with additional deletion of the flexible loop would be introduced. Hence, crystallization experiments with variant Loop3 were carried out as the most promising strategy towards solving a structure of Rpb4/7.

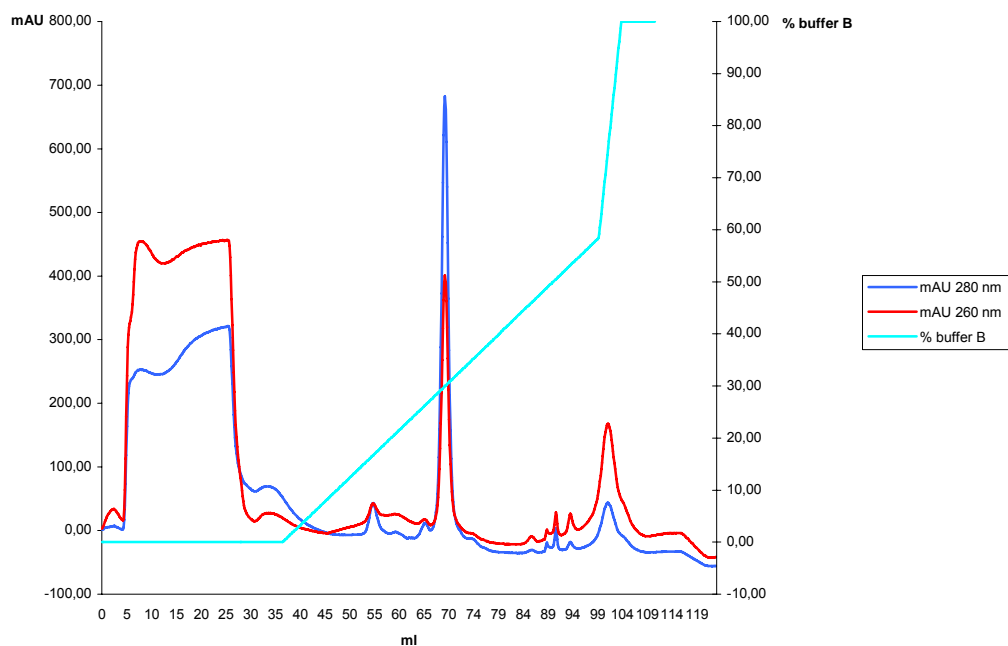
### Expression and purification of Rpb4/7 variant Loop3

Rpb4/7 variant Loop3 was expressed in 8l LB medium to get reasonable amounts for crystallization and purified by Ni-NTA gravity flow affinity chromatography. The protein is not stable in pH 7.5 in which it precipitates in few minutes after elution from Ni-NTA. The optimal pH = 8.0 and in all the subsequent purifications, the buffers were adjusted to this value. The elution occurred mostly at the concentration of imidazole between 20–200 mM. The concentration of protein in these fractions was estimated with Bradford reagent and the purity was checked on SDS-PAGE (Figure 17).

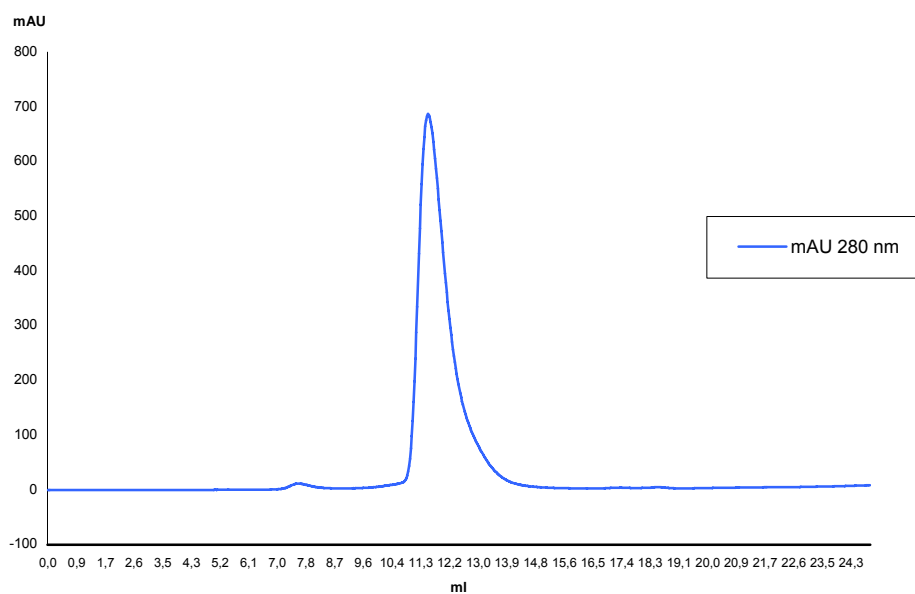


**Figure 17. 15 % SDS-PAGE of Loop3 Rpb4/7 variant after Ni-NTA purification.**  
*20  $\mu$ l protein solution were put on the gel;  
 Numbers indicate the concentration of  
 imidazole  
 M: protein marker*

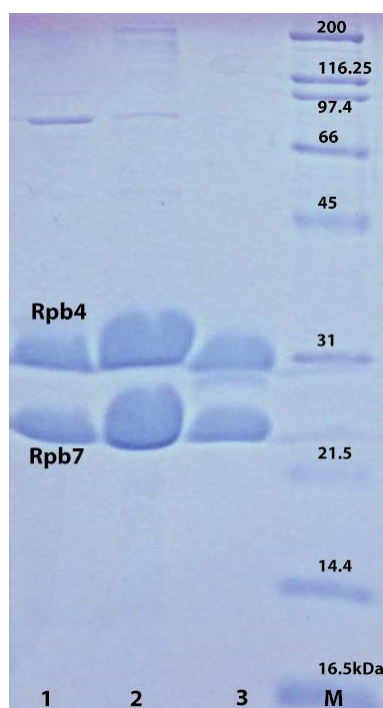
Main fractions were pooled together, diluted 3 times with buffer C to reach the salt concentration below 100 mM NaCl (ResourceQ buffer A), and applied to the anion exchange chromatography ResourceQ column (Figure 18). The final step of purification was Superose 12 gel filtration in Pol II buffer (Figure 19). The quality of the protein after this purification is very high as can be seen on the gel above (Figure 20). The protein behaves well and is easy to handle, although all the buffers need to have a proper pH, since small changes can affect Loop3 stability causing protein precipitation. Protein could be frozen in Pol II buffer in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  for the crystallization experiments.



**Figure 18. ResourceQ anion exchange chromatography of Loop3 Rpb4/7 variant.**  
*blue:absorption at 280 nm, red absorption at 260 nm, green: salt gradient*



**Figure 19. Superose 12 gel filtration chromatography of Loop3.**  
*Blue: absorption at 280 nm*



**Figure 20.**  
15 % SDS-PAGE of  
Loop3 variant after  
ResourceQ (lane 1 and  
2); and Superose 12  
(lane 3). *M*: marker

### Crystallization of Loop3 variant

Peak protein fractions from Superose 12 could be concentrated only to 3 mg/ml. Higher protein concentration was not possible to obtain due to protein precipitation which is probably the result of instability of this variant at higher concentrations. Crystallization trials with the variant Loop3 were carried out with the hanging drop vapour diffusion method. Establishment of crystallization conditions was done by sparse matrix sampling using commercially available screens from Hampton. After crystals were grown, promising conditions were refined by variation of salts, pH, PEG and additives in the reservoir (Table 6).

**Table 6.** Conditions from Hampton Screens which resulted in crystals.

Screen	Reagent	Salt	Precipitant	Buffer	Cryo	Crystals
PEG/Ion	5	0.2 M magnesium chloride hexahydrate	20 % w/v polyethylene glycol 3350	-	-	intergrown, 3D,
	13	0.2 M sodium thiocyanate	20 % w/v polyethylene glycol 3350	-	-	stacks of thin needles
	14	0.2 M potassium thiocyanate	20 % w/v polyethylene glycol 3350	-	-	stacks of thin needles
	15	0.2 M lithium nitrate	20 % w/v polyethylene glycol 3350	-	-	stacks of thin needles
	20	0.2 M magnesium formate	20 % w/v polyethylene glycol 3350	-	-	single, 3D, grow from precipitate
	<b>25</b>	0.2 M magnesium acetate tetrahydrate	20 % w/v polyethylene glycol 3350	-	-	single, 3D, fairly big, sharp edges
	33	0.2 M sodium sulfate decahydrate	20 % w/v polyethylene glycol 3350	-	-	single, 3D, high birefringence
	34	0.2 M potassium sulfate	20 % w/v polyethylene glycol 3350	-	-	single, 3D, high birefringence
	<b>36</b>	0.2 M di-sodium tartate dihydrate	20 % w/v polyethylene glycol 3350	-	-	single, 3D, fairly big, sharp edges
	37	0.2 M potassium sodium tartrate tetrahydrate	20 % w/v polyethylene glycol 3350	-	-	microcrystals
	38	0.2 M di-ammonium tartrate	20 % w/v polyethylene glycol 3350	-	-	single, long thin needles,
PEG/Ion						

	<b>44</b>	0.2 M di-ammonium hydrogen phosphate	20 % w/v polyethylene glycol 3350	-	-	intergrown, 3D, high birefringence
<b>Cryo</b>	<b>6</b>	0.2 M magnesium chloride hexahydrate	24 % w/v polyethylene glycol 4000	0.08 M tris hydrochloride pH8.5	20 % v/v glycerol anhydrous	single, 3D, sharp edges
	9	0.17 M ammonium acetate	25.5 % w/v polyethylene glycol 4000	0.085 M tri-sodium citrate dehydrate pH5.6	15 % v/v glycerol anhydrous	single, needles
	<b>22</b>	0.17 M sodium acetate trihydrate	25.5 % w/v polyethylene glycol 4000	0.085 M tris hydrochloride pH8.5	15 v/v glycerol anhydrous	single, 3D
	40	-	19 % w/v iso-propanol 19 %w/v polyethylene glycol 4000	0.095 M tri-sodium citrate dehydrate pH5.6	15 v/v glycerol anhydrous	single, needles
<b>Crystal Screen</b>	<b>12</b>	0.2 M magnesium chloride hexahydrate	30 % iso-propanol	0.1 M Sodium HEPES pH7.5	-	single, 3D, high birefringence
	<b>33</b>	-	4 M sodium formate	-	-	microcrystals

Several conditions were chosen for further refinement and they are marked red in the Table 6. Figure 21 shows a few examples of crystals obtained from initial screens (before refinement).



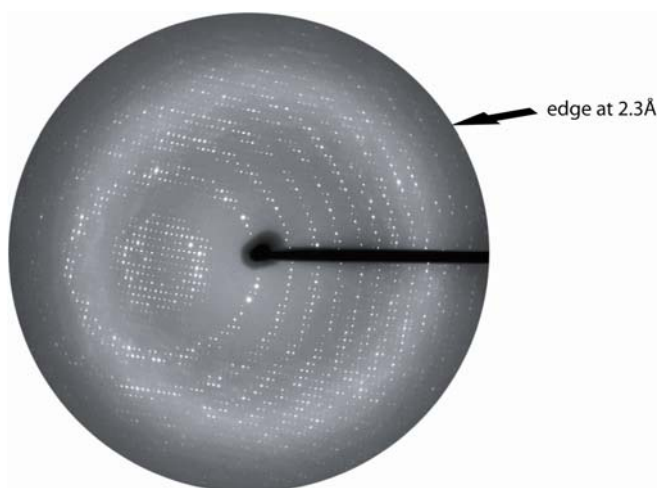
**Figure 21.** Pictures of crystals of Rpb4/7 Loop3 variant from initial screens.

The refinement screens varied the pH, the salt and the precipitant concentrations and resulted in several conditions which gave rise to crystals diffracting better than 3 Å. Figure 22 shows crystals obtained from refinement screens.



**Figure 22.** Pictures of crystals of Rpb4/7 Loop3 variant from refinement screens.

Best crystals were carefully frozen with the use of different cryoprotectants and freezing protocols and measured on the rotating anode home source to find out the best condition and harvesting protocol. The refined condition 44 from PEG/ION (0.2 M di-ammonium hydrogen phosphate pH 7.9, 20-22 % polyethylene glycol 3350, 5 mM DTT) gave rise to crystals which grew to a maximum size of 0.3 x 0.2 x 0.1 mm. Crystals were transferred stepwise to reservoir solutions containing additionally 5, 10, 18, and 22 % glycerol, and were plunged in liquid nitrogen. These crystals diffracted to around 2 Å (Figure 23) and allowed for structure determination of subcomplex of Rpb4/7.



**Figure 23.** Diffraction pattern of Rpb4/7 variant Loop3 crystal. Measured at SLS, X06SA beamline, 1s exposure time, oscillation increment 0.5°.

### Structure determinations

We could extend the resolution of the crystals of 12-subunit Pol II from 4.2 to 3.8 Å (Table 7), but not any further, likely because of the high solvent content of around 80 %. Complete diffraction data were collected in 0.5° increments at the Swiss Light Source (Table 7). Molecular replacement with AMORE and the ten-subunit Pol II (PDB code 1I6H with nucleic acids removed) as search model resulted in a unique solution (correlation coefficient = 52.4, second best solution = 30.5). An initial electron density map was improved by solvent flipping. The high quality of the map allowed us to adjust the structure of the polymerase core (Cramer et al., 2001; Gnatt et al., 2001) with program O (Jones et al., 1991) and to build parts of the Rpb4/7 as an atomic model.

To complete the model, I however had to solve the crystal structure of free Rpb4/7 at high resolution. Full-length recombinant Rpb4/7 failed to crystallize, but crystals were obtained after 34 residues were removed from the proteolytically sensitive N-terminal tail of Rpb4 (Figure 24). Additional deletion of a proteolytically sensitive loop in Rpb4 resulted in larger crystals that diffracted synchrotron radiation to 2 Å resolution. A MAD experiment was performed with selenomethionine-containing crystals at the Swiss Light Source (Table 7). Data were processed with DENZO and SCALEPACK (Otwinowski and Minor, 1996). The program SOLVE (Terwilliger, 2002) detected 14 selenium sites and was used for MAD phasing (Z-score=42.8, figure of merit=0.37). A subsequent anomalous difference Fourier identified four additional peaks. After phasing with SHARP (La Fortelle and Bricogne, 1997), six additional selenium peaks were detected, giving a total of 24 selenium peaks. The peaks stemmed from four copies of the Rpb4/7 complex in the asymmetric unit and could be assigned to methionines M1, M22, M106, and M115 in Rpb7, and M145 in Rpb4 (double peak). Phasing with all selenium peaks resulted in a figure of merit of 0.40 and in an electron density map that enabled building of an initial model with program O (Jones et al., 1991) that was partially refined with CNS (Brunger et al., 1998). The native crystal belonged to the same space group as the selenomethionine crystals (P2<sub>1</sub>2<sub>1</sub>2), but the crystallographic c-axis is half the length (Table 7), resulting in two Rpb4/7 complexes per asymmetric unit instead of four complexes. Thus a dimer of two Rpb4/7 complexes was used as a search model in AMORE (Navaza, 1994) to solve the native structure by molecular replacement (Table 7, correlation coefficient=42.7, next highest peak=36.9). The atomic Rpb4/7 structure was refined to a free R-factor of 27.4 % at 2.3 Å resolution (Table 7, Figures 24, 27). Due to the high resolution, a total of 154 water molecules could be added to the refined model. An atomic model is of very high stereochemical quality and shows 99 % of the residues in allowed and additionally

allowed regions of the Ramachandran plot, and none of the residues in disallowed regions (Table 7).

The refined Rpb4/7 structure served as a guide for completion of the Rpb4/7 model within Pol II. After removal of the water molecules, restricted refinement of atomic positions and B-factors, minor adjustments of the complete Pol II model led to a structure with high stereochemical quality with 98.5 % of the residues in allowed and additionally allowed regions of the Ramachandran plot (Table 7, Figure 28). Two factors apparently enabled refinement of an atomic Pol II model with data to 3.8 Å resolution. First, combination of two refined structures, the core Pol II structure and the Rpb4/7 structure, resulted in a very good starting model. Second, the high solvent content allowed for powerful electron density modification, and resulted in a favorable ratio of observed reflections to refinement parameters. The 3.8 Å electron density map has high quality, and side chains are generally visible (Figure 27).

**Table 7.** X-ray diffraction data

Crystal	Rpb4/7 SeMet			Rpb4/7 Native	Complete Pol II
<i>Data collection</i> <sup>1</sup>					
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2			P2 <sub>1</sub> 2 <sub>1</sub> 2	C222 <sub>1</sub>
Wavelength [Å]	0.97946 peak	0.93930 remote	0.97973 inflection	0.93200	0.97924
Unit cell axis [Å]	103.39, 114.69, 161.42			103.65, 114.81, 80.48	222.7, 395.1, 284.3
Resolution [Å]	20-2.6 (2.69-2.60)	20-2.6 (2.69-2.60)	20-2.6 (2.69-2.60)	50-2.3 (2.4-2.3)	50-3.8 (3.95-3.8)
Completeness [%]	99.6 (99.5)	99.7 (99.6)	99.7 (99.6)	95.2 (92.0)	99.3 (97.0)
Unique reflections	59,077 (5,811)	59,144 (5,816)	59,129 (5,815)	41,315 (4,872)	121,866 (13,089)
Redundancy	4.5	4.5	4.3	3.2	3.2
R <sub>sym</sub> [%]	7.9 (21.6)	7.4 (29.4)	8.3 (40.3)	4.9 (31.0)	7.3 (35.3)
<I/σI>	15.0	13.3	11.7	16.0	11.5
<i>f</i> <sup>''</sup>	-7.8	-2.28	-9.6		
<i>f</i> <sup>'''</sup>	5.0	3.53	2.5		
<b>Refinement</b>					
Residues				583	3902
Zn <sup>2+</sup> /Mg <sup>2+</sup>				-/-	8/1
Waters molecules				154	-
RMSD bonds [Å]				0.007	0.009
RMSD angles [°]				1.36	1.53
R <sub>cryst</sub> [%]				22.8	25.7
R <sub>free</sub> [%]				27.4	28.5

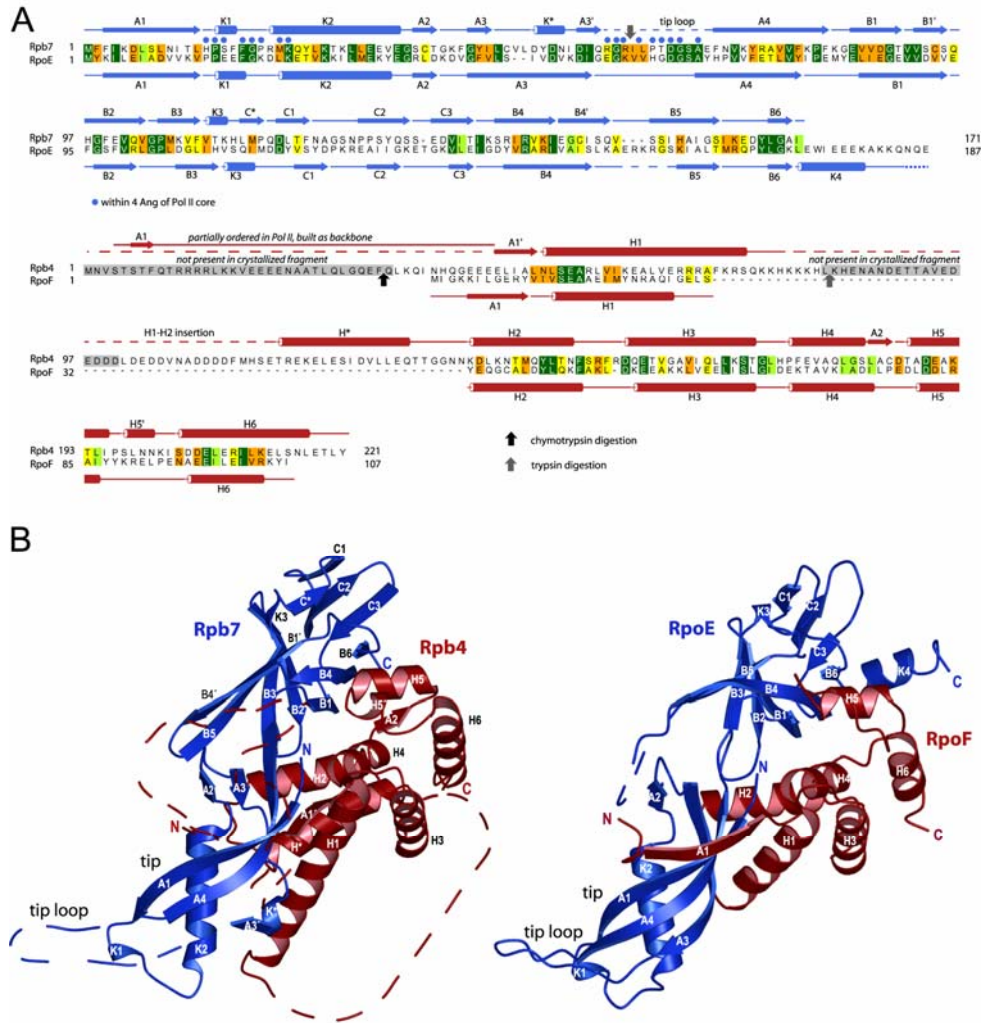
<sup>1</sup>All data were collected at beamline X06SA at the Swiss Light Source, Villigen, Switzerland.

**Rpb4/7 structure**

The overall structure of the Rpb4/7 complex is similar to its archaeal counterpart (Todone et al., 2002). Rpb4 binds between two Rpb7 domains, the N-terminal “tip” domain and a C-terminal domain (Figure 24B). Compared with the archaeal protein, Rpb7 lacks the C-terminal helix K4, and has an additional helix inserted into strand A3 (K\*, Figure 24). About half of the Rpb4 residues (48-57, 138-221) fold as in the archaeal counterpart (Figure 25). Rpb4 additionally contains a non-conserved N-terminal extension (residues 1-46), a longer helix H1, and an insertion between helices H1 and H2 (H1-H2 insertion), which comprises a long disordered loop and an additional helix (H\*, Figure 24). The Rpb4 N-terminal extension is flexible in free Rpb4/7, but forms a partially ordered loop in the Pol II structure. This Rpb4 N-terminal extension is apparently associated weakly with Rpb2 in the complete Pol II structure (Figures 28 and 29). Residues involved in Rpb2-Rpb4 contacts are however not conserved, and the 34 N-terminal residues of Rpb4 are not required for binding to the Pol II core in vitro (not shown). Three N-terminal residues of Rpb4 associate with the sheet of the Rpb7 tip, similar to the N-terminal half of strand A1 in the archaeal counterpart (Figures 24B, 28). Charge distribution on both archaeal and yeast Rpb4/7 is consistent with putative role of dimer in binding of exiting, transcribed RNA (Figure 26).

**Folding transitions upon Pol II core-Rpb4/7 interaction**

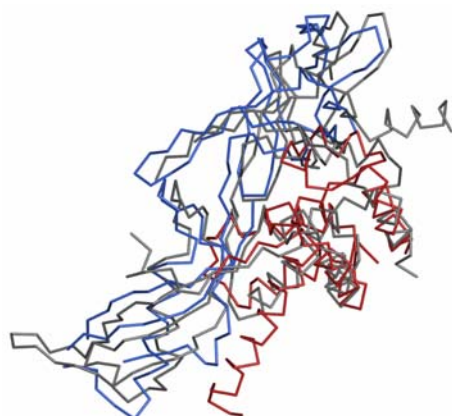
In addition to the Rpb4 N-terminal extension, other protein elements fold upon Rpb4/7-core interaction. The Rpb7 tip domain binds to the Pol II core with the helical turn K1, and with the highly conserved “tip loop,” which is disordered in the free Rpb4/7 structure, but folds upon core binding (Figure 29). There are no obvious changes in the conformation of the H1-H2 insertion upon core binding. Rpb4/7 binding however induces formation of an  $\alpha$ -helix in the previously disordered residues 1445-1455 in the linker region of the largest Pol II subunit (Rpb1 helix  $\alpha$ 50, Figure 29). Folding transitions upon interaction with Pol II are also observed in the elongation factor TFIIS (Kettenberger et al., 2003) and the initiation factor TFIIB (Bushnell et al., 2004), and appear to be a common theme in interactions of the Pol II core with accessory factors.



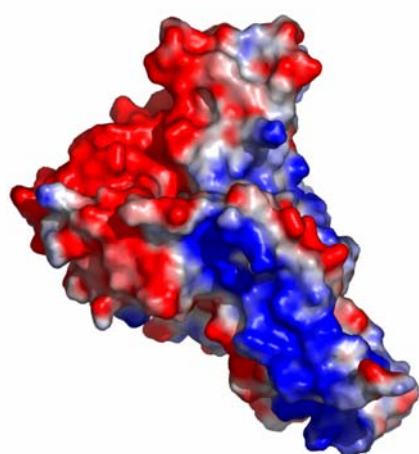
**Figure 24. Structure of the Rpb4/7 complex.**

(A) Primary and secondary structure. Structure-based alignments of amino acid sequences of *S. cerevisiae* Rpb7 with *M. jannaschii* RpoE (top) and of *S. cerevisiae* Rpb4 with *M. jannaschii* RpoF (bottom). Secondary structure elements are shown above and beneath the respective sequences (cylinders,  $\alpha$ -helices; arrows,  $\beta$ -strands; lines, loops; dashed lines, disordered; grey shading, absent from the crystallized variant), and are labeled according to (Todone et al., 2001). Conserved residues are highlighted according to decreasing conservation from dark green, through light green and orange, to yellow. Cleavage sites revealed by limited proteolysis are indicated with arrows. Residues that are within 4 Å from residues of the Pol II core in the complete polymerase structure are indicated with a blue dot. Figure prepared with ALSRIPT (Barton, 1993).

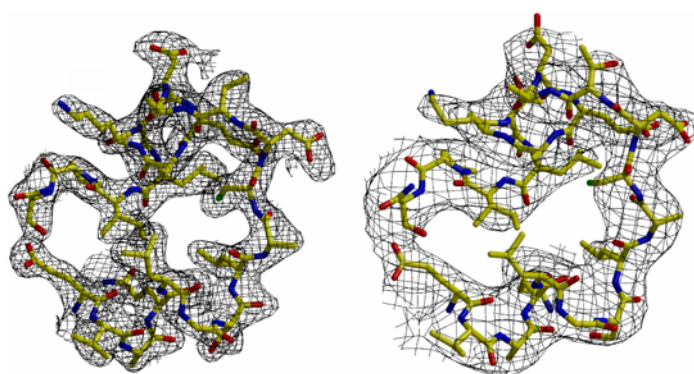
(B) Comparison of the structures of yeast Rpb4/7 (this study, left) and archaeal RpoF/E (15). Rpb7/RpoE are in blue and Rpb4/RpoF are in red. Disordered in the Rpb4/7 structure are the Rpb7 tip loop (residues 57-68), and Rpb4 residues 35-46, 77-81, and 101-118. Rpb4 residues 1-34 and 82-100 were not present in the crystallized variant. Figures were prepared with MOLSCRIPT (Kraulis, 1991) and PYMOL (DeLano Scientific).



**Figure 25. Superposition of Rpb4/7 and E/F.** RMSD for 45 C-alpha atoms is around 3.2Å; Rpb4/7 in red and blue respectively and E/F in grey



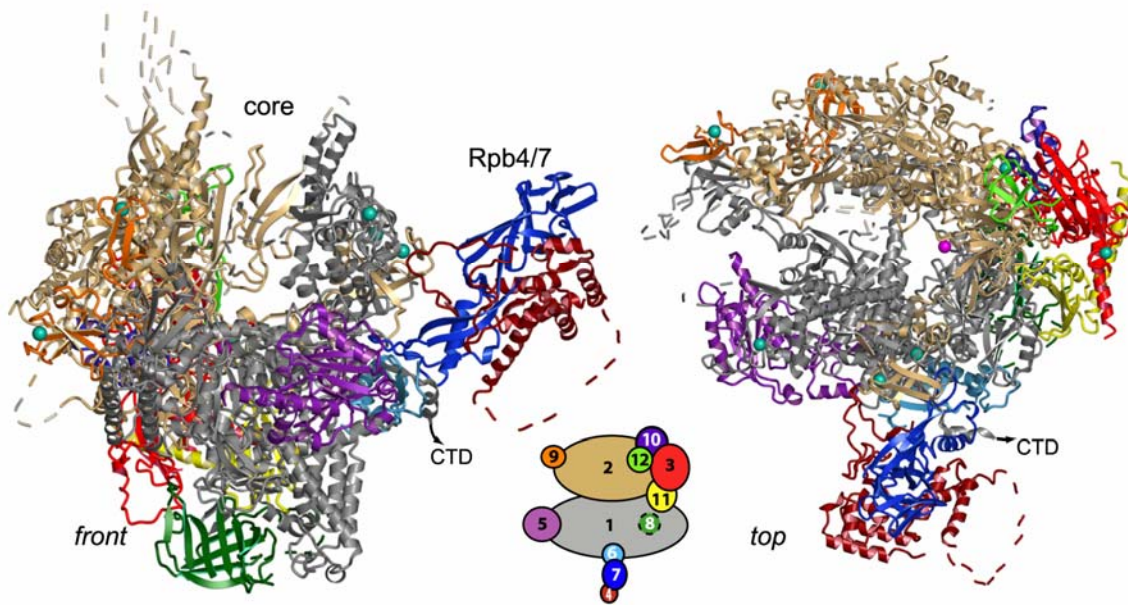
**Figure 26. Electrostatic potential of Rpb4/7.** This view corresponds to the classical view rotated by 180° around y-axis and tilted towards us about 20° around x-axis. Potential was calculated with Pymol. Blue positive, red: negative



Rpb4/7 at 2.3 Å

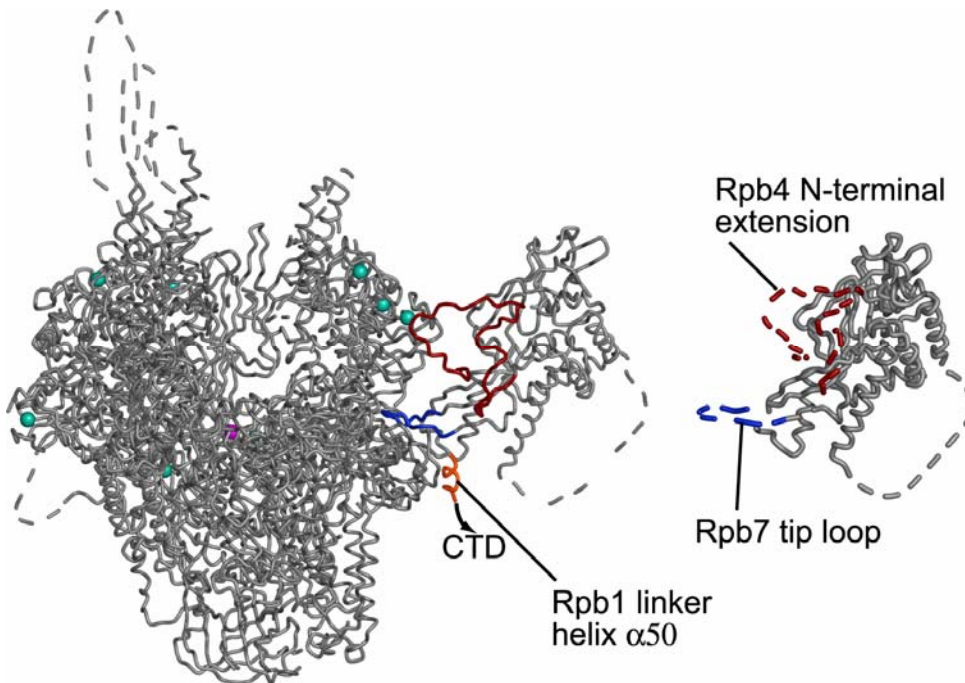
Pol II at 3.8 Å

**Figure 27. Electron density maps.**  $2F_o - F_c$  Electron density maps around the final model of Rpb4 residues 176-196 are shown for free Rpb4/7 at 2.3 Å resolution (left) and for the complete Pol II at 3.8 Å resolution (right). The maps are contoured at  $1\sigma$ . The figure was prepared with BOBSCRIPT (Esnouf, 1997).



**Figure 28. Complete RNA polymerase II structure.**

*Ribbon diagram. Two standard views “front” and “top” are shown (Cramer et al., 2000 and 2001). The 12 subunits Rpb1-Rpb12 are colored according to the key below the views. Dashed lines represent disordered loops. Eight zinc ions and the active site magnesium ion are depicted as cyan spheres and a pink sphere, respectively. Secondary structure assignments for the Pol II core taken from (Cramer et al., 2001).*



**Figure 29. Folding transitions upon Rpb4/7 binding to the Pol II core.**

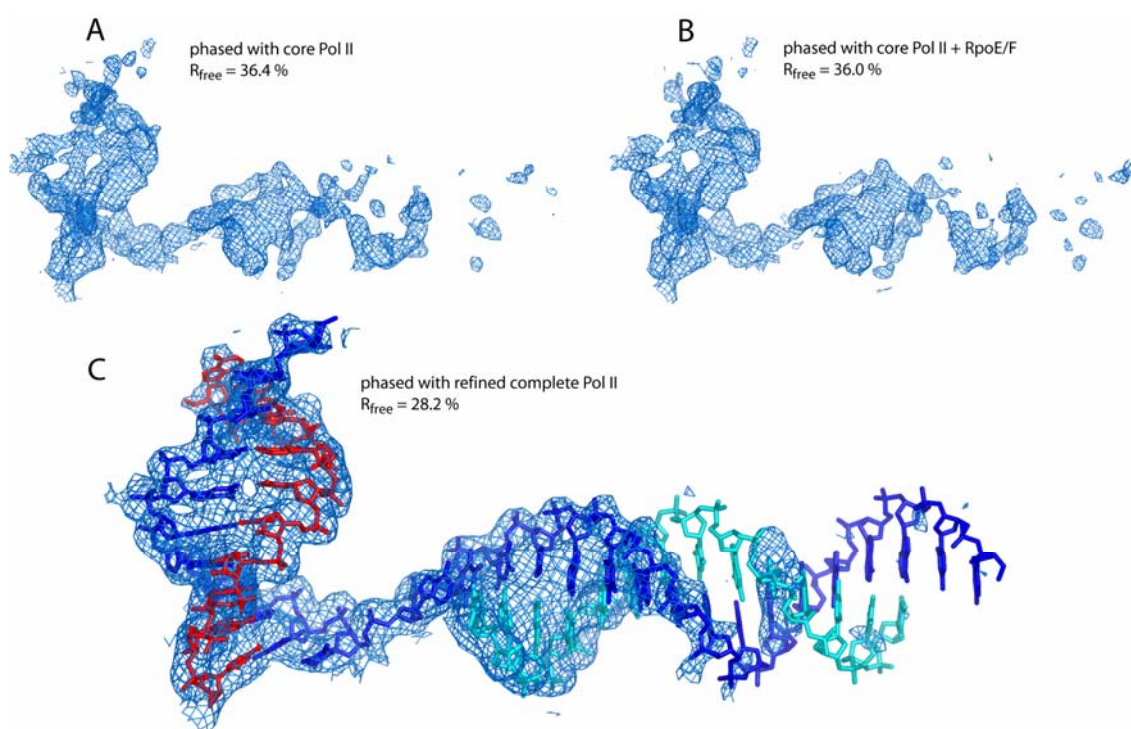
*The complete Pol II structure (left) and free Rpb4/7 structure (right) are represented as grey coils. Elements that fold upon the interaction are colored in blue (Rpb7 tip loop), red (Rpb4 N-terminal extension), and orange (additional helix  $\alpha 50$  in the Rpb1 linker to the carboxyl-terminal domain).*

**Specificity of the Pol II core-Rpb4/7 interaction**

The complete Pol II structure reveals details of the Pol II core-Rpb4/7 interaction, which refine our previous description of the involved protein elements (Armache et al., 2003). For example, the structure reveals a hydrogen bond between residue G66 of Rpb7 and residue Q100 of Rpb6 in the Pol II core. This hydrogen bond explains why mutation of residue Q100 in Rpb6 leads to facilitated Rpb4/7 dissociation and to reduced Pol II activity upon a shift to non-permissive temperatures (Tan et al., 2003). The detailed information on the core-Rpb4/7 interaction also helps to rationalize why another nuclear RNA polymerase, Pol I, binds specifically to A43/14, the Rpb4/7 counterpart. Specificity of A43/14 for the Pol I core may result from an ionic contact between a lysine that replaces the Rpb7 residue P63 in A43 and a patch of three glutamates that replace the Rpb1 residues T13, M1433, and G1439 in the largest Pol I subunit. I could however not find a straightforward explanation for specific interaction of the core of the third nuclear RNA polymerase, Pol III, with its Rpb4/7 counterpart C25/17. It is possible that additional, Pol III-specific subunits are involved in defining the specificity in subunit assembly.

**Importance of the refined 12-subunit Pol II in phasing new structures**

The most important factor in a successful molecular replacement is a good search model. The refined, atomic model of 12-subunit RNA polymerase II at 3.8 Å is crucial for phasing new structures of the complete Pol II in complex with additional factors (Kettenberger et al., 2004). The complex of 12-subunit Pol II with nucleic acids was initially phased with 10-subunit core Pol II model and with 12-subunit 4.2 Å model but the density for nucleic acids became clear enough to build them after the refined 12-subunit 3.8 Å model was used for phasing (Kettenberger et al., 2004; Figure 30).



**Figure 30. 2Fo-Fc of nucleic acids bound to Pol II calculated using different Pol II models for phasing.** (Kettenberger et al., 2004); (A) 10-subunit Pol II (Cramer et al., 2001) was used for phasing; (B) 12-subunit 4.2Å model (Armache et al., 2003) was used for phasing; (C) 3.8Å refined atomic model of 12-subunit Pol II (Armache et al., 2004) was used for phasing;  $R_{\text{free}}$  after rigid body refinement (CNS)

## **Chapter IV**

*Modeling of Pol I and III based on the refined  
structure of 12-subunit Pol II*

### **Abstract**

The refined atomic model of the complete 12-subunit RNA polymerase II at 3.8 Å resolution enabled homology modeling of the two other nuclear RNA polymerases. In Pol I and Pol III, 65 % and 77 % of the Pol II fold are conserved, respectively, much more than anticipated from canonical alignments. Conservation of the nucleic acid-binding cleft predicts that the mechanism of RNA elongation is essentially the same in these enzymes. Structural differences occur on the upstream surfaces, which bind different initiation factors, accounting for promoter specificity. Differential assembly of the three polymerases relies on four major specific interfaces between homologous subunits. The homologous subunits also contribute invariant residues to interfaces with the five subunits that are common to all three polymerases, and that form molecular staples at equivalent surface positions.

## **Introduction**

The refined structure of the complete Pol II (Armache et al., 2004) enabled homology modeling of Pol I and Pol III, which helps to understand the specific assembly of the three eukaryotic, nuclear RNA polymerases. Additionally analysis of these models showed that surface regions involved in RNA elongation are conserved, whereas regions involved in initiation factor binding are divergent. Since the five common subunits are identical in all three polymerases, they cannot contribute to specific enzyme assembly. Instead, they appear to stabilize the homologous subunits and their specific interactions and serve as “molecular staples”. The Pol I and Pol III models show that the common subunits can bind at equivalent locations in these enzymes, despite the divergent interfaces, by forming a limited set of key contacts with invariant residues of the homologous subunits (“hot spots”). The refined complete Pol II structure, together with the models of Pol I and Pol III, provide a reference for functional analysis of the eukaryotic transcription machineries.

## Results and Discussion

### Modeling of RNA polymerases I and III

The complete Pol II structure enabled homology modeling of Pol I and Pol III, which provides insights into similarities and differences in the structure and function of the three nuclear RNA polymerases in eukaryotes. Subunits Rpb5, 6, 8, 10, and 12 are common to all three polymerases and were included in the models (Table 1). The Pol II subunit Rpb9 was excluded from modeling because its putative counterparts A12.2 and C11 also show homology to the elongation factor TFIIS in their C-terminal domains, resulting in an ambiguity. The Rpb4/7 complex has counterparts in Pol I and Pol III (Hu et al., 2002; Peyroche et al., 2002; Sadhale and Woychik, 1994; Shpakovski and Shematorova, 1999). Rpb7 shows clear homology, but Rpb4 shows very weak sequence conservation, and thus was excluded from the models. The two large subunits, Rpb1 and Rpb2, and subunits Rpb3 and Rpb11, have homologues in Pol I and Pol III (Table 1).

**Table 1.** Subunits in Pol II structure and Pol I/III models

Pol II			Pol I				Pol III		
Subunit	Type	Residues in sequence	Residues in refined structure	Subunit	Sequence identity [%]	Conserve d fold [%]	Subunit	Sequence identity [%]	Conserved fold [%]
Rpb1	homolog	1733	1416	A190	22.1	63.6	C160	28.4	83.2
Rpb2	homolog	1224	1097	A135	25.4	71.0	C129	35.8	87.2
Rpb3	homolog	318	266	AC40	25.8	60.2	AC40	25.8	60.2
Rpb4	homolog	221	177	A14	excluded	excluded	C17	Excluded	excluded
Rpb5	common	215	214	Rpb5	100	100	Rpb5	100	100
Rpb6	common	155	84	Rpb6	100	100	Rpb6	100	100
Rpb7	homolog	171	171	A43	21.6	46.8	C25	24.6	46.8
Rpb8	common	146	133	Rpb8	100	100	Rpb8	100	100
Rpb9	unclear	122	119	A12.2	excluded	excluded	C11	Excluded	excluded
Rpb10	common	70	65	Rpb10	100	100	Rpb10	100	100
Rpb11	homolog	120	114	AC19	20.8	81.6	AC19	20.8	81.6
Rpb12	common	70	46	Rpb12	100	100	Rpb12	100	100
Specific subunits									
-	-	-	-	<b>Pol I specific: A49, A34.5</b>	-	-	<b>Pol III specific: C82/C34 /C31, C53, C37</b>	-	-
Overall		4562	3902		33.6	65.4		38.8	77.1

Initial sequence alignments of the Pol II subunits Rpb1, 2, 3, 7, and 11 with their counterparts in Pol I and Pol III (Clustal W, Thompson et al., 1994) contained many obvious errors, which resulted from locally weak or absent conservation, and from misplaced gaps. We corrected the alignments in an iterative procedure based on the Pol II structure (Figure 2). Residues that lacked counterparts in Pol I or Pol III were removed from the structure, identical residues were kept, and conserved residues were replaced by their counterparts. Inspection of the resulting models for the chemical environment of each residue revealed reasonable packing in many regions. Key conserved residues often formed hydrophobic cores or salt bridges with residues that are distant in sequence, confirming the alignment in regions with weak sequence similarity. Steric clashes or disallowed contacts identified sequence stretches that were incorrectly aligned. These stretches were realigned manually, resulting in improved homology models that were inspected again. Iteration of this procedure converged on reliable models for large portions of Pol I and Pol III (Figure 3). Insertions and deletions in the two large subunits of Pol I and Pol III locate to the enzyme surfaces, and do not disturb the core folds (Figure 3). Since doubtful parts were excluded from the models, further regions of conserved fold may exist.

### **Similarity of RNA polymerases and elongation mechanism**

The modeling showed that at least 65 % and 77 % of the Pol II fold is conserved in Pol I and Pol III, respectively (Table 2). In the two dominating large subunits of Pol III, even 83 % and 87 % of the Pol II fold are conserved. This is surprising given that the sequence identity in the two large subunits in Pol I and Pol III ranges from 22 % to 36 % only (Table 2). The structurally and functionally essential core of Pol II is thus fully conserved in the two other polymerases (Figure 2). Of the two large subunits, only the folds of the Rpb1 jaw domain and the Rpb2 external domain 1 are not conserved (Figure 2). In Pol I, the foot domain in the largest subunit is also likely to have a different fold, or it could be highly diverged in sequence. Rpb4/7 is structurally different from its counterparts, except for the Rpb7 tip domain that binds the polymerase core.

Conserved surfaces on homologous subunits of the three polymerases are found around the incoming DNA, in the active site and the binding-site for the DNA-RNA hybrid, and at the RNA exit tunnel. This suggests that the basic mechanism of RNA elongation is the same in all three polymerases. Also conserved is the overall surface charge distribution of the enzymes, with a positively charged active center cleft (not shown). Surfaces outside the cleft are not conserved (Figure 3), except for a few small patches, reflecting differential interactions with transcription factors and polymerase-specific subunits. In addition to the 12 subunits that are

either identical or homologous to the Pol II subunits, Pol I contains the specific subunits A34.5 and A49, and Pol III contains the specific subcomplex C82/C34/C31, and the specific subunits C53 and C37.

### **Initiation factor binding and promoter specificity**

During initiation, the polymerases contact promoter DNA indirectly, via specific initiation factors (Figure 1). Initiation factors bind to the enzyme upstream face, which is mainly formed by the clamp, wall, dock domain, and the Rpb4/7 complex (Figure 3, (Armache et al., 2003)). The upstream faces of the three polymerases are diverse, due to insertions and deletions on the surfaces of the clamp, wall and dock domains, due to structural differences between Rpb4/7 and its counterparts, and due to specific Pol I and Pol III subunits bound in this region. Structural differences in the upstream face of the three polymerases can account for specific initiation factor binding, explaining recognition of different promoter classes.

The Pol II dock domain binds TFIIB (Bushnell et al., 2004; Chen and Hahn, 2003). A similar dock domain in Pol III could bind the TFIIB-related factor Brf1 (Figure 3), and a 14-residue insertion on the dock domain surface in Pol I reflects the absence of a TFIIB-like factor in this system (Figure 3, Bushnell et al., 2004; Chen and Hahn, 2003). The Rpb4/7 complex binds TFIIF (Chung et al., 2003), whereas its structurally different counterparts A43/14 and C25/17 bind Rrn3 (Peyroche et al., 2000; Yuan et al., 2002), and Brf1 (Ferri et al., 2000), respectively. C25/17 also binds the Pol III-specific subunit C31, which resides in a subcomplex C82/C34/C31 that also binds to the initiation factors TBP and TFIIC (Flores et al., 1999; Geiduschek and Kassavetis, 2001; Schramm and Hernandez, 2002; Wang and Roeder, 1997; and references therein), and to the outside of the clamp near zinc site Zn8 (Figure 3). The second largest Pol III subunit C129 contains a seven-residue deletion near the adjacent zinc site Zn7 (Figures 2 and 3). The outside of the clamp in Pol I is also structurally different, since it shows insertions of five and 36 residues near zinc site Zn7 (Figure 3). The flap loop on top of the wall of Pol I is seven residues shorter than in Pol II. The coiled-coil at the inner side of the clamp is seven residues longer (Figure 3), and apparently binds subunit A49 (Bischler et al., 2002). A corresponding region interacts with the  $\sigma$  factor in bacterial RNA polymerase (Vassylyev et al., 2002; Murakami et al., 2002), and is near TFIIF in the Pol II system (Chung et al., 2003). Below the dock domain is the Rpb3/11 heterodimer that binds the Mediator complex that is required for regulated Pol II transcription (Davis et al., 2002). An exposed zinc-binding loop in Rpb3 is not present in AC40, and may contribute to specific association of Mediator with Pol II.

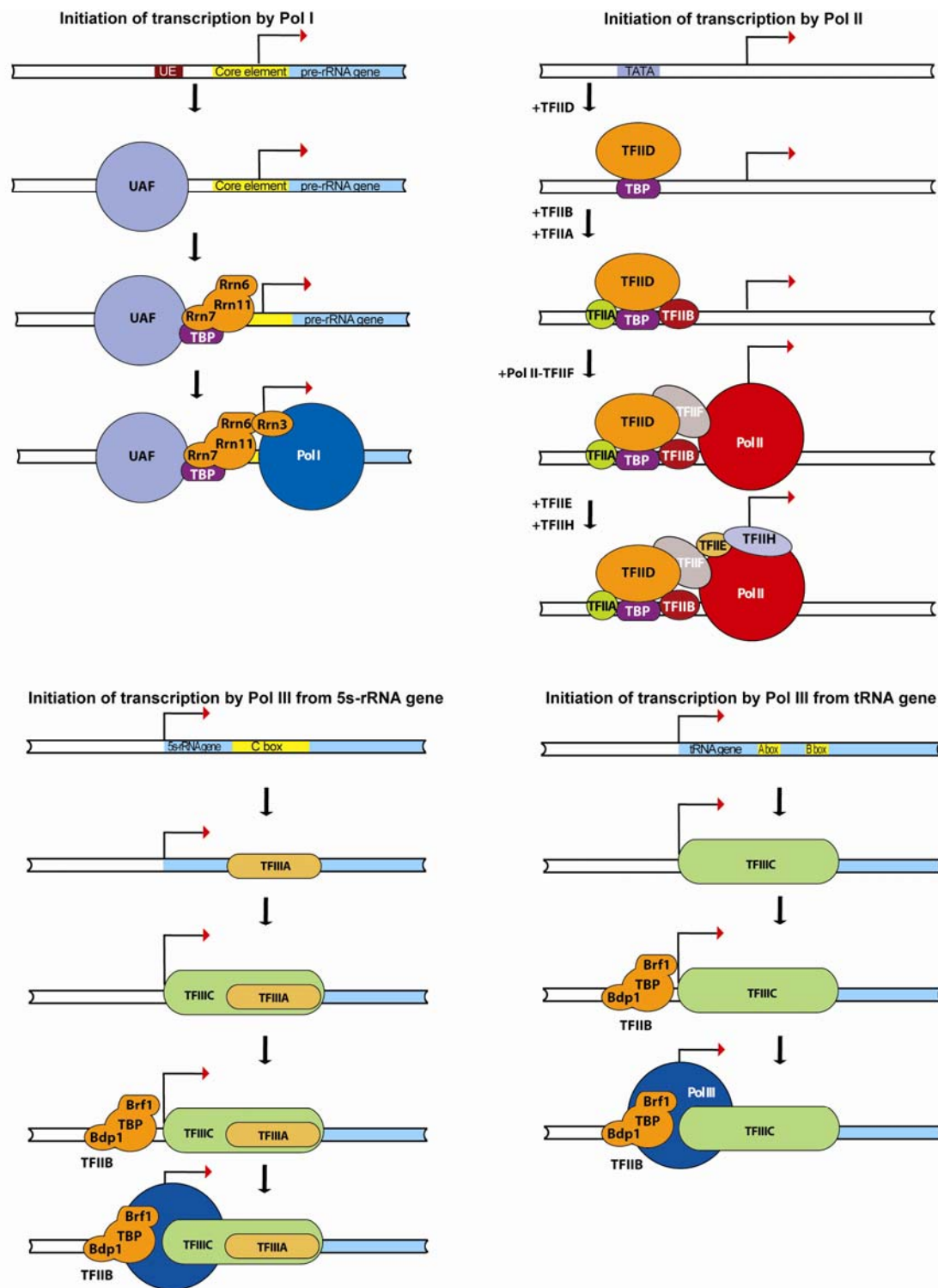
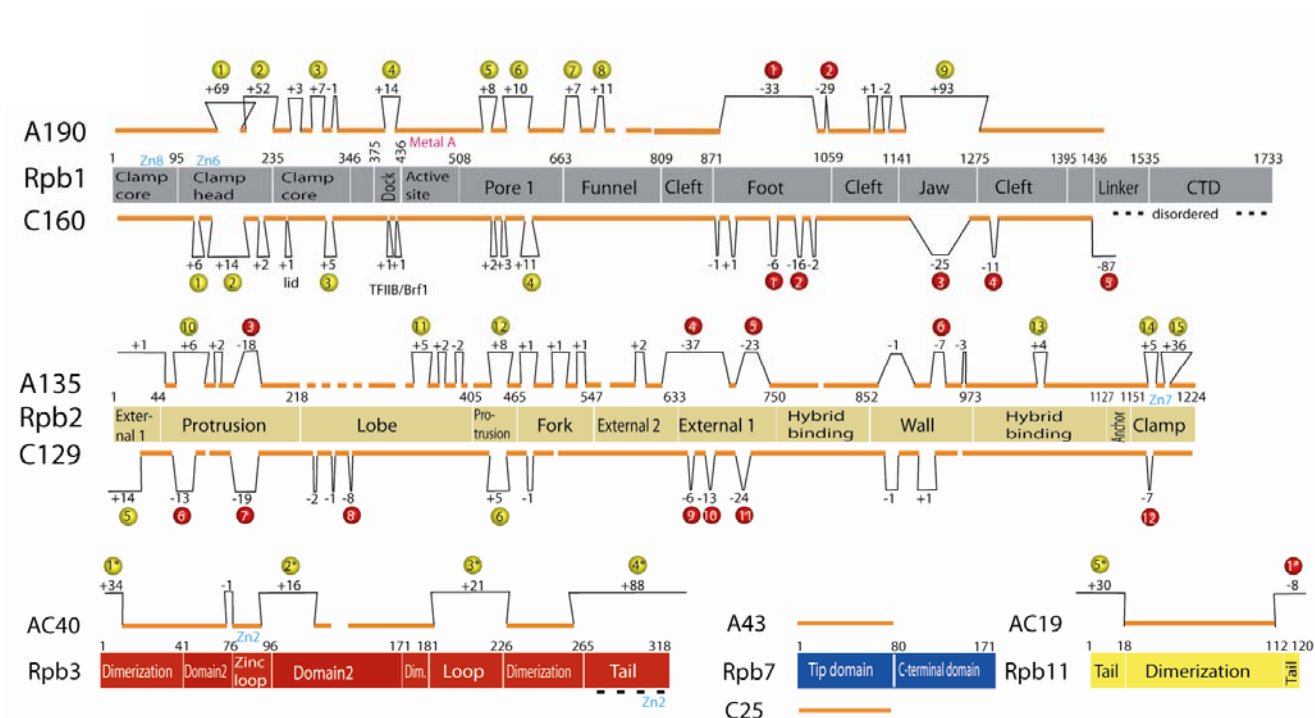
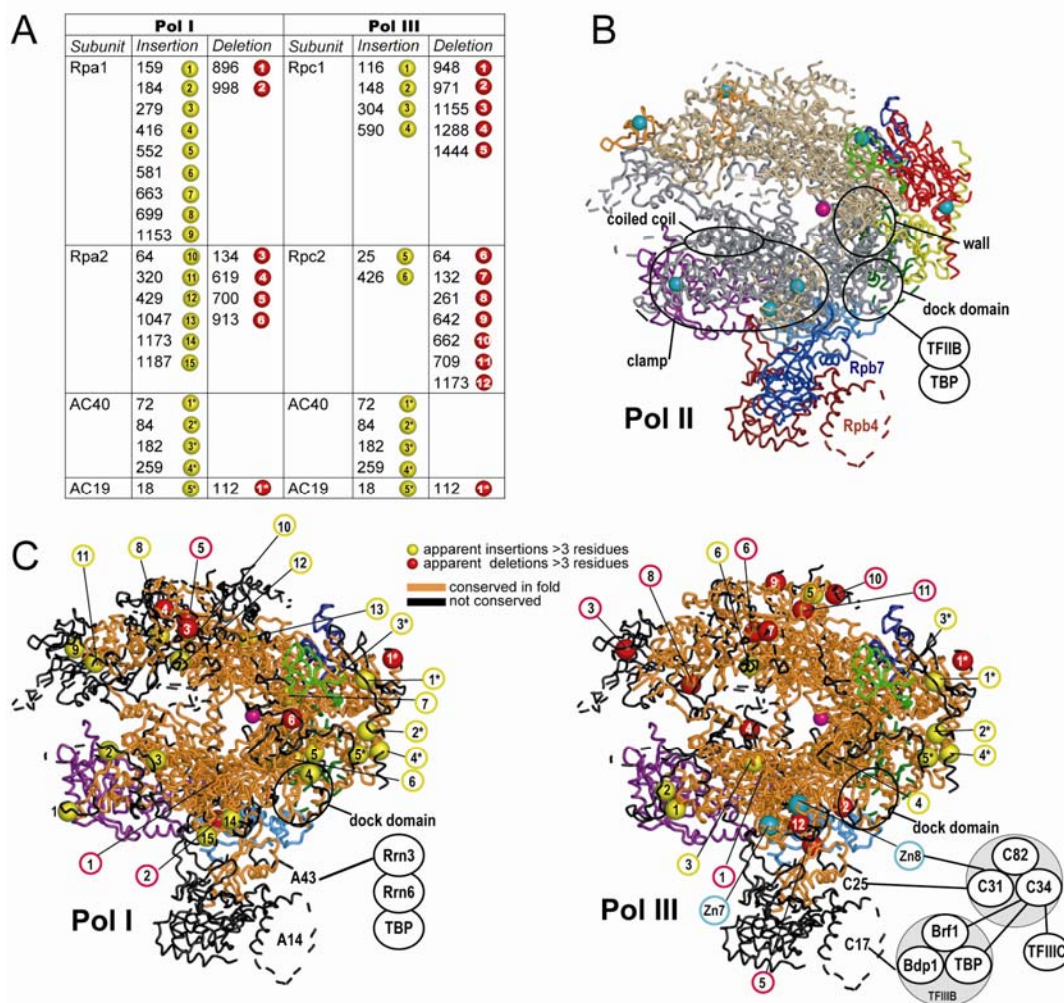


Figure 1. Initiation of transcription by 3 eucaryotic, nuclear RNA polymerases.



**Figure 2. Modeling of RNA polymerases I and III.** *Pol II structure-guided sequence alignments of subunits homologous in nuclear RNA polymerases. The domain organization of Pol II subunits Rpb1, 2, 3, 7 and 11 is shown as diagrams (Cramer et al., 2001). Above and below the diagrams, regions conserved in fold in the homologous subunits of Pol I and Pol III are indicated with orange bars. Regions that apparently adopt a different structure are indicated with black brackets. The number of residues apparently inserted or deleted in the Pol I/III subunits as compared to the Pol II subunits are indicated. Yellow and red spheres represent insertions and deletions of more than three residues, respectively, in the Pol I/III sequences*



**Figure 3. Modeling of RNA polymerases I and III.** (A) Exact positions of apparent insertions and deletions in the homologous subunits. The numbers correspond to the last aligned Pol II residue before the insertion/deletion in Pol I and Pol III. (B) Backbone coil representation of Pol II colored according to Fig. 2A (top). Elements on the upstream face are encircled (clamp, clamp coiled-coil, dock domain, wall). (C) Conservation of the Pol II fold in Pol I (left) and Pol III (right). The Pol II representation in (B) has been colored as follows. Regions in homologous subunits that are conserved in fold are in orange (C). Deletions and insertions in the Pol I and Pol III amino acid sequences as compared with Pol II are shown as red and yellow spheres, respectively, if they exceed three amino acid residues in length (numbers defined in (A)). Additional subunits and initiation factors are indicated as black circles and their physical interactions with black lines (compare text).

### **Specific assembly of Pol I, II, and III**

The Pol II structure and the models for Pol I and Pol III suggest that differential assembly of the three enzymes from sets of common, homologous, and specific subunits can be achieved in five major specific steps (Figure 4). First, stable heterodimers are formed by the Pol II subunits Rpb3 and Rpb11, or the corresponding subunits AC40 and AC19 in both Pol I and Pol III. Second, the Rpb3/11 heterodimer binds Rpb2, and the AC40/19 heterodimer binds either A135 or C129, the second largest subunits of Pol I and Pol III, respectively. The common subunits Rpb10 and Rpb12 bridge between the heterodimer and the second largest subunit, stabilizing the interaction. Third, the second largest subunit specifically binds its cognate largest subunit, which is stabilized by the common subunits Rpb5, 6, and 8. Fourth, the two large subunits specifically bind the Rpb4/7 complex or its counterparts. Finally, additional specific subunits associate with the surface of Pol I and Pol III. Biochemical data generally support this assembly model (Werner and Weinzierl, 2002), but the order of the assembly steps remains unclear.

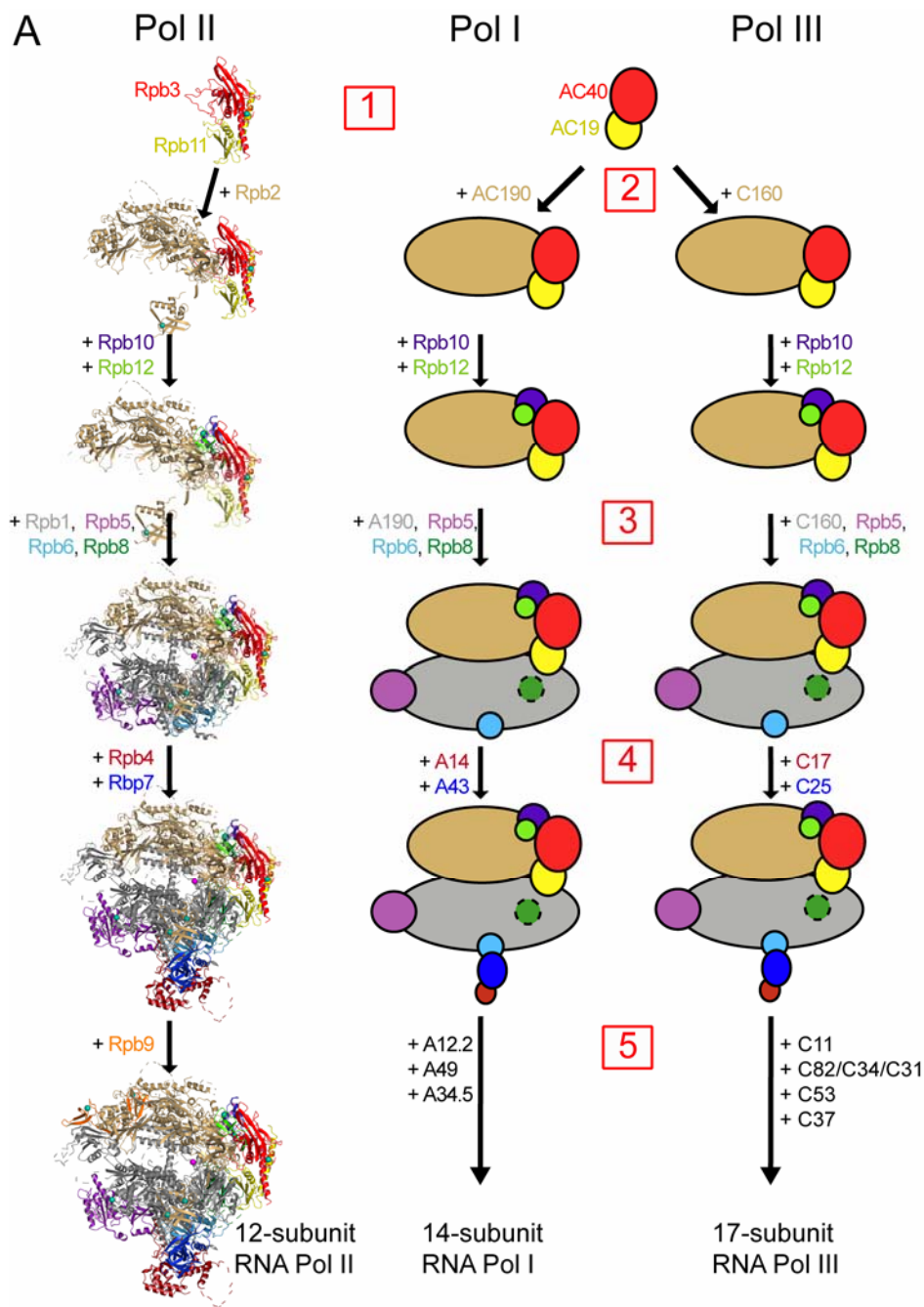
Detailed analysis of the subunit-subunit interfaces in the three polymerases suggests the structural basis for specific enzyme assembly. Several hydrophobic residues in the heterodimeric interfaces of Rpb3/11 and AC40/19 differ in size, resulting in specific heterodimerization (step 1 in Figure 4). For example, residues Rpb3 L33 and Rpb11 A48 are replaced by phenylalanine and tyrosine, respectively, in the AC40/19 heterodimer. Either A135 or C129 can interact with AC40/19, because they contribute three invariant residues to the interface (Rpb2 residues Y833, F1001, and Y1073, step 2 in Figure 4). Association of AC40/19 with either A135 or C129 may lead to formation of a new salt bridge between a glutamate that replaces Rpb2 residue K1079 in both A135 and C129, and an arginine that replaces the Rpb3 residue A225 in AC40. Rpb2 association with AC40/19 is therefore disfavored because two positively charged residues would approach each other in the interface. The two large subunits form a very extended contact surface that is strongly conserved, and specificity in the interaction between cognate large subunits can thus not be attributed to a defined region (step 3 in Figure 4). Specific binding of Rpb4/7 and its counterparts to the cores of the three polymerases can partially be rationalized (step 4 in Figure 4). Specificity of the Rpb4/7 counterpart A43/14 for the Pol I core may result from an ionic contact between a lysine that replaces the Rpb7 residue P63, and a patch of three glutamates that replace the Rpb1 residues T13, M1433, and G1439.

**Common subunits as molecular staples**

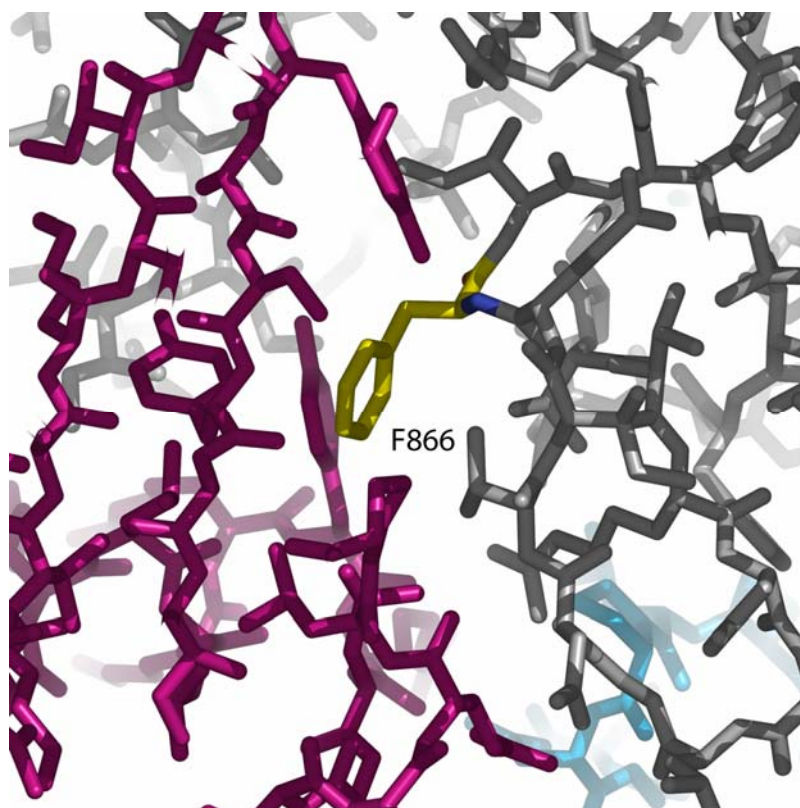
Since the five common subunits are identical in all three polymerases, they cannot contribute to specific enzyme assembly. Instead they appear to stabilize the homologous subunits and their specific interactions. Rpb5 binds to a region in Rpb1 that is predicted to be disordered in free Rpb1 (FOLDINDEX, Uversky et al., 2000), Rpb6 has been shown to stabilize Rpb1 (Nouraini et al., 1996), Rpb10 and Rpb12 bridge between Rpb2 and Rpb3, and Rpb12 additionally bridges Rpb2 regions that are distant in sequence. The Pol I and Pol III models show that the common subunits can bind at equivalent locations in these enzymes, despite the divergent interfaces, by forming a limited set of key contacts with invariant residues of the homologous subunits (“hot spots,” Table 3). For example, the invariant phenylalanine F866 in Rpb1 inserts into a hydrophobic pocket of Rpb5, whereas the rest of the interface between Rpb5 and the largest polymerase subunits is strongly divergent (Figure 5). The two acidic Rpb2 residues E116 and D896 form salt bridges with two basic residues in Rpb12, and are invariant in Pol I and Pol III (Table 3).

**Architectural principles**

Comparison of the structures of Pol II and the ribosome (Ban et al., 2000; Ramakrishnan, 2002; Schlutzen et al., 2000; Wimberly et al., 2000) reveals principles underlying the architecture of these multi-macromolecular machines. In Pol II, two large subunits and an additional heterodimer form the central mass of the enzyme and the active center, whereas five small surface subunits stabilize the complex. In the ribosome, large RNA chains form the central mass and the active site of the ribozyme, whereas small protein subunits on the surface seem to stabilize the complex like molecular staples. The number of these stabilizing protein subunits increased during evolution. Eukaryotic ribosomes contain more small protein subunits than bacterial ribosomes, and eukaryotic RNA polymerases contain five stabilizing common subunits, of which only one has a homologue in the bacterial enzyme (Rpb6, (Minakhin et al., 2001; Mukherjee and Chatterji, 1997).



**Figure 4. Specific assembly of eukaryotic nuclear RNA polymerases.** Model for assembly of Pol II, Pol I, and Pol III (from left to right). Subunit colors and the view (top) are as in Fig. 2A. Five specific subunit-subunit interactions may direct differential assembly of the three enzymes (numbers in red squares).



**Figure 5. Example of a prominent contact residue (“hot spot”).** Invariant phenylalanine F866 (yellow) in Rpb1 (grey) interacts with a hydrophobic pocket of Rpb5 (magenta), whereas the rest of the interface between Rpb5 and the largest polymerase subunit is strongly divergent.

**Table 2.** Invariant contacts between common and homologous subunits

Common subunit	Residues	Contacted homologous subunit in Pol I, II, III	Invariant “hot spot” residues
Rpb5	Y168	Rpb1, A190 C160	R857
	Y168, Y211		F866
Rpb6	Y88, A91	Rpb1, A190, C160	L504
	E89, T81, R136		Y852
	G95,		A499
	N104,		T381
	I120		Y383
	Q100	Rpb7, A43, C25	G66
Rpb8	Y20, N21, K22, V23	Rpb1, A190, C160	D538
	N43		F540
	L46		L571
	N43, L46, D94, Y95,		K567
	V96		P563
	W79		W572
	W79		P568
	D94		I565
	V96		I566
Rpb10	I2	Rpb3, AC40	L143
	E19		Y114
	D55, E58, I57		K146
	F60, L61		V57
	L61, R62, Y63, N64	Rpb2, A135, C129	R148
	Y44		I1006
	R48		D1009
Rpb12	R54	Rpb2, A135, C129	E116
	K58		D896
	R60, F67	Rpb3, AC40	D60

# Supporting material

## Sequence alignments of homologous subunits in Pol I, Pol II and Pol III

*Underlined-conserved in fold*

*Star- invariant*

### Rpa1-Rpb1

```
sp|P10964|RPA1_YEAST    --MDISKPVGSEITSVDFGILTAKEIRNLSAKQITNPTVLDNLG-HPVSGGLYDLALGA 56
sp|P04050|RPB1_YEAST    MVGQQYSSAPLRTVKEVQFGFLSPEEVRAISVAKIRFPETMDETQTRAKIGGLNDPRLGS 60
                        : *. .      : ..*:***:::..*: * :. * * .*: :.   *** *  **:

sp|P10964|RPA1_YEAST    FLRNL-CSTCGLDEKFCPGHQGHELPVCYNPLFFNQLYIYLRASCLFCHHFRLKSVE- 114
sp|P04050|RPB1_YEAST    IDRNLKQTCQEGMNECPGHFGHIDLAKPVFHVGFIAKIKKVCECVMCHCGKLLDEHNE 120
                        : *** *.** . : **** ***:*. * :: * : :: . . *:.* :: *.

sp|P10964|RPA1_YEAST    VHRYACKLRLLQYGLIDESYKLDEITLGSLNSSMYTDDEAIEDNEDEMDGEGSKQSKDISS 175
sp|P04050|RPB1_YEAST    LMRQALAIKDSKRFAAIWTLCKTKMVCETDVPSEDDP-----

sp|P10964|RPA1_YEAST    TLLNELKSKRSEYVDMAIAKALSDGRTTERGSFTATVNDERKKLVHEFHKLLSRGKCDN 235
sp|P04050|RPB1_YEAST    -----TQLVSRGG 166

sp|P10964|RPA1_YEAST    CGMFSPKFRKDGFTKIFETALNEKQITNRVKGFIRQDMIKKQKQAKKLDGSNEASANDE 295
sp|P04050|RPB1_YEAST    CGNTQPTIRKDGLKLVGS-----WKKDRATGDADEPELRLVLSTE 205
                        ** .*.:***:. : .      ***:: : . * .: . *

sp|P10964|RPA1_YEAST    ESFDVGRNPTTRPKTGSTYILSTEVKNILDTVFRKEQCVLQYVFHSRPNLSRKLVKADSF 355
sp|P04050|RPB1_YEAST    EILNIFKHISVKDFTS-----LGFNEVFSRPEWM 234
                        * ::: : :. : *.      .:. : :. :

sp|P10964|RPA1_YEAST    FMDVLVVPPTRFRLPSKLGEEVHENSQNQLLSKVLTTSLLIRDLNDDLSKLQDKVSLED 415
sp|P04050|RPB1_YEAST    ILTCLPVPPPPVRPSISFNESQRG---EDDLTFKLADILKANISLETLEHNGAP----- 291
                        :: * ***. *. . .:*. : :.: * : * * . : *.: : :*:

sp|P10964|RPA1_YEAST    RRVIFSRLMNAFVTIQNDVNAFIDSTKAQG-RTSGKVPIPGVKQALEKKEGLFRKHMMGKR 475
sp|P04050|RPB1_YEAST    --HHAIEEAESLLQFHVATYMDNDIAGPQALQKSGRPVKSIRARLKGKEGRIRGNLMGKR 344

sp|P10964|RPA1_YEAST    VNYAARSVISPDNINETNEIGVPPVFAVKLTYPEPVTAYNIAELRQAVINGPDKWPGATQ 535
sp|P04050|RPB1_YEAST    VDFSARTVISGDPNLELDQVGVPKSIAKTLTYPEVVTPYNIDRLTQLVRNGPNEHPGAKY 404
                        *::*:*** ***: * ::*** :* .***** **.* ** * * ***: ***.

sp|P10964|RPA1_YEAST    IQNEDGSLVSLIGMSVEQRKALANQLLPSSNVSTHTLNKKVYRHIKNRDVVLMNRQPTL 595
sp|P04050|RPB1_YEAST    VIRDSGDRIDLR-----YSKRAGDIQLQYGWKVERHIMDNDPVLFNRQPSL 450
                        : .:.*. :.*      : :. . ** *** :.* ***:***:*
```

sp P10964 RPA1_YEAST	<u>HKASMMGHKVRVLPNEKTLRLHYANTGAYNADFDGDEMNMHFPQENENARAEALNLANTDS</u>	655
sp P04050 RPB1_YEAST	HKMSMMAHRVKVIPYS-TFRLNLSVTSYPNADFDGDEMNLHVPQSEETRAELSQCVAVPL	509
	** ***,*::**:* . *::** : *..*****:*..*::** :* .	
sp P10964 RPA1_YEAST	<u>QYLTPTSQSPVRGLIQDHISAGVWLTSKDSFFTREQYQQYIYGCI</u>	715
sp P04050 RPB1_YEAST	QIVSPQSNKPCMGIQDTLCGIRKLTLRDTFIELDQVLNMLYWVPDWDG-----VIP	561
	* ::* *..* *::** :.. ** *::** :* ::* :. *	
sp P10964 RPA1_YEAST	<u>PPTIFKPYPLWTGKQIITTVLLNVTTPDMPGINLISKNIKNEYWGKGSLENEVLFKDGA</u>	775
sp P04050 RPB1_YEAST	TPAI IKPKPLWSGKQILSVAIP-----NGIHLQRFDEGTLLSPKDNGLIIDGQ	611
	.*:*** ***:*****:..: * * :::: .. :* :*: **	
sp P10964 RPA1_YEAST	<u>LLCGILDKSQYGASKYGIVHSLHEVYGPEVAAKVLVSLGRLEFTNYITATAFTCGMDDLRL</u>	835
sp P04050 RPB1_YEAST	IIFGVVEKKTVGSSNGGLIHVVTRKGPQVCAKLFNGIQKVVNFWLLHNGFSTGI----	671
	:: *::** .*: *::* . **:*..*::: : :.. :. :*: *::*	
sp P10964 RPA1_YEAST	<u>TAEGNKWRTDILKTSVDTGREAAAETNLDKDTPADDP</u>	895
sp P04050 RPB1_YEAST	--GDTIADGPTMREITETIAEAKKKVLDVTKEAQAN-----LLTAKHGMTLRES	713
	. :.* ** :* :*: * : * :. *	
sp P10964 RPA1_YEAST	<u>TSSKVNAITSQVSKCVPDGTMKKFPNCMSQAMALSGAKGSNVNVSQIMCLLGQQA</u>	955
sp P04050 RPB1_YEAST	FEDNVVRFLNEARDKAGRLAEVNLKDLNNVKQVMAGSKGSFINIAQMSACVQQSVEGK	773
	..* : .. .* . : *::* *::** *::** : *::**:	
sp P10964 RPA1_YEAST	<u>RVPVMVSGKTLPSFKPYETDAMAGGYVKGRFYSGIKPQEYFHC</u>	1015
sp P04050 RPB1_YEAST	RIAFGFVDRTLPHFSKDDYSPESKGFVENSYLRLTPQEFFFHAMGGREGLIDTAVKTAE	833
	*:.. . :*** *. : .. : *::* . : *::*::**.*.*****:..	
sp P10964 RPA1_YEAST	<u>SGYLQRCLTKQLEGVHVSVDNSIRDADGTLVQFMYGGDAIDITKESHMTQFEFCLDNYYA</u>	1075
sp P04050 RPB1_YEAST	TGYIQRLVKALEDIMVHYDNTTRNSLGNVIQFIYGEDGMDAAHIEKQ-SLDTIGGS	892
	::** *.* *:: * ***: *::* *::** *::* :. :. :. *	
sp P10964 RPA1_YEAST	<u>LLKKY-----NPSALIEHLDVESALKYSKKTLYRKKHS-----KEP</u>	1112
sp P04050 RPB1_YEAST	FEKRYRVDLNTDHTLDPSSLSESGSEILGDLKLQVLLDEEYKQLVKDRKFLREVFDGEA	952
	: ** . * :. *.:: *..* * :*	
sp P10964 RPA1_YEAST	<u>HYKQSVKYDPVLAKYNPAKYL-----GSVSENFQDKLESFLDKN-----</u>	1151
sp P04050 RPB1_YEAST	NWPLPVNIRRIIQNAQQTFFHIDHTKPSDLTIKDIVLGVKDLQENLLVLRGKNEIIQNAQR	1012
	:: .*: : : : :. . :::** : **	
sp P10964 RPA1_YEAST	<u>-----SKLFKSSDGVNEKKFR--ALMQLKYMRLINPGEAVGIIASQSV</u>	1193
sp P04050 RPB1_YEAST	DAVTLFCCLLRSLATRRVLQYRLTKQAFDWVLSNIEAQFLRSVVHPGEMVGLAAQSI	1072
	:. . . :::* --- :::**::** *::**:	
sp P10964 RPA1_YEAST	<u>GEPSTQMTLNTFHFAGHGAANVTLGIPRLREIVMTASAAIKTPQMTLPIWN--DVSDEQA</u>	1251
sp P04050 RPB1_YEAST	GEPATQMTLNTFHFAGVASKKVTSGVPRLKEILN-VAKNMKTPSLTVYLEPGHAADQEQA	1131
	***:***** : :** *::** : : ***: :. :.***	
sp P10964 RPA1_YEAST	<u>DTFCKSISKVLLSEVIDKVIVTETTGTSTAGGNAARSYVIHMRFFDNNEYSEEYDVSKE</u>	1311
sp P04050 RPB1_YEAST	KLIRSAIEHTTLKSVTIASEIYYDPDRSTVIPEDDEIIQLHFSLLDEAEQSFDDQSPW	1191
	. : .*:..*.* : ... .* : . :*: :::: .. : *	

```

sp|P10964|RPA1_YEAST      ELQNVISNQFIHLLEAAIVKEIKKQKRTTGPDIGVAVP-----RLQTDVANSSNS 1362
sp|P04050|RPB1_YEAST      LLRLELDRAAMNDKDLTMGQVGERIKQTFKNDLFVIWSEDNDEKLIIRCVVRPKSLDAE 1251
                             *:  :..  ::  : :: :  :: *: *   *: *   .           * :.  .:*  . .

sp|P10964|RPA1_YEAST      KRLEEDNDEEQSHKKTQAVSYDEPDEDEIETMREAEKSSDEEGIDSDKESDSDSEDEDV 1422
sp|P04050|RPB1_YEAST      TEAEEDHMLKKIENTMLENITLR
                             .. ***:  :: :..  : ::

sp|P10964|RPA1_YEAST      DMNEQINKSIVEANNMNKVVQRDRQSAIISHHRFITKYNFDDESGKWCEFKLELAADTEKLLMVN 1487


sp|P10964|RPA1_YEAST      IVEEICRKSIIRQIPHIDRCVHPEPENGKRVLVTEGVNFQAMWDQEAFDVDGITSNDVA  --
sp|P04050|RPB1_YEAST      GVENIERVVMKYDRKVPSPTEGYVKEPEWVLETDGVNLSEVMTVPG-IDPTRIYTNSFI 1333
                             *** *   :::  ::  .   :: : ** *::*: . :   . **   * :*.

sp|P10964|RPA1_YEAST      AVLPTYGVEAARNTIVNEINNVSRYAISVSFRHLDLIADMMTRQGTYLAFNRQGMETS-  --
sp|P04050|RPB1_YEAST      DIMEVLGIEAGRAALYKEVYNVIASDGSYVNYRHMALLVDVMTTQGGLTSVTRHGFNRSN 1393
                             :::  *::*. * :: :*: **:  .  *::*: *:.*::* **   :..*:::  *

sp|P10964|RPA1_YEAST      TSSFMKMSYETTCQFLTKAVLDNEREQLDSPSARIVVGKLNNVGTGSFVDLAKVPNAA--  --
sp|P04050|RPB1_YEAST      TGALMRCSFEETVEILFEAGASAELDDCRGVSENVILGQMAPIGTGAFDVMIDEESLVKY ->CTD
                             *.::::  *:: *  :: *  .  *  ::  .  *  .::*:  :*****:  .  . .

```

**Rpb1-Rpc1**

```

sp|P04050|RPB1_YEAST      MVGQQYSSAPLRTVKEVQFGLFSPEEVRAISVAKIRFPETMDETQTRAKIG-GLNDPRLG  59
sp|P04051|RPC1_YEAST      -MKEVVVSETPKRIKGLEFSALSAADIVAQSEVEVSTRDLFDLEKDRAPKANGALDPKMG  59
                          : : * . : : * : * . : : * : * . * * : : *

sp|P04050|RPB1_YEAST      SIDRNLKCQTCQEGMNECPGHFGHIDLAKPVFHVGFIAKIKKVCEVCVMHCGLLLDEHN 119
sp|P04051|RPC1_YEAST      VSSSSLECATCHGNLASCHGHFGLKLALPVFHIGYFKATIQILQGICKNCSAILLSETD 119
                          . . * : * * : . : * * : : * * : : : : : : : : * : * : * : * :

sp|P04050|RPB1_YEAST      -----ELMRQALAIKDSKKRFAAIWTLCKTKMVCETDVPSEDDPTQLVSRGGCGNTQPT 173
sp|P04051|RPC1_YEAST      KRQFLHELRRPGVDNLRMGILKKILDQCKKQRRCLHCGALNG--VVKKAAAGAGSAALK 177
                          * * * . : : * * : * . : . . . : . * . : .

sp|P04050|RPB1_YEAST      IRKDGLKLVG-----SWKKDRATGDADEPELR-----VLSTEEILNIFKHISVK 217
sp|P04051|RPC1_YEAST      IIHDTFRWVGKKSAPKEDIWVGWKEVLAHNPELERYVKRCMDDLNPLKTLNLFKQIKSA 237
                          * : * : * * * : * : * : * : * . : * : * : * :

sp|P04050|RPB1_YEAST      DFTSLGFNEVFS--RPEWMILTCLPVPPPVRPSISFNES-QRGEDDLTFKLADILKANI 274
sp|P04051|RPC1_YEAST      DCELLGIDATVPSGRPETIYIWRYPAPVPCIRPSVMMQDSPASNEDDLTVKLTEIVWTSS 297
                          * * : : . . * * * * : * : * : * : * : * : * : * : * :

sp|P04050|RPB1_YEAST      SLETLEHNGAPHHAIEEAESLLQFHVATYMDNDIAGQPQALQKSG----RPVKSIRARL 329
sp|P04051|RPC1_YEAST      LIKAGLDKGISINNMMEHWDYLQTLVAMYINSDSVNPAMLPGSSNGGGKVKPIRGFCQRL 357
                          : : * . : : * . * : * * : * . . . . * . : : : : *

sp|P04050|RPB1_YEAST      KGKEGRIRGNLMGKRVDFSARTVISGDPNLELDQVGVPKSIAKTLTYPEVVTYPYNIDRLT 389
sp|P04051|RPC1_YEAST      KGKQGRFRGNLSGKRVDFSGRTVISDPNLSIDEVAVPDRVAKVLTTYPEKVTRYNRHKLQ 417
                          * : * : * : * : * : * : * : * : * : * : * : * : * : * :

sp|P04050|RPB1_YEAST      QLVRNGPNEHPGAKYVIR-DSGDRIDLRYSKRAGDIQ-LQYGWKVERHIMDNDPVLFNRRQ 447
sp|P04051|RPC1_YEAST      ELIVNGPNVHPGANYLLKRNEDARRNLRYGDRMKLAKNLQIGDVVERHLEDGDVVLFNRRQ 477
                          : * : * : * : * : : : . . * : * : * . : * * * * : * * : * : * :

sp|P04050|RPB1_YEAST      PSLHKMSMMAHRVKVIPIYSTFRLNLSVTSPYNADFDGDEMNLHVPQSEETRAELSQLCAV 507
sp|P04051|RPC1_YEAST      PSLHRLSILSHYAKIRPWRTFRLNECVCTPYNADFDGDEMNLHVPQTEEARAEAINLMGV 537
                          * : * : * : * : * : * : * : * : * : * : * : * : * : * :

sp|P04050|RPB1_YEAST      PLQIVSPQSNKPCMGIVQDTLCGIRKLTLRDTFIELDQVLNMLYWVPD--WDGVIPTPAI 565
sp|P04051|RPC1_YEAST      KNNLLTPKSGEPIIAATQDFITGSYLISHKDSFYDRATLTQLLSMMSDGIHFDPPIPAI 597
                          : : : : * : * : . * : * : : * : * : : * : * : * : * :

sp|P04050|RPB1_YEAST      IKPKPLWSGKQILSVAIP--NGIHLQRFDEGTTLSP-----KDNGLIIDG- 610
sp|P04051|RPC1_YEAST      MKPYLWTGKQVFSLLIKPNHNSPVVINLDAKNKVFVPPKSKSLPNEMSQNDGFVIIRGS 657
                          : * * * : * : * : * . : : * . : : * : : : * : : : * : *

sp|P04050|RPB1_YEAST      QIIFGVVEKKTVGSSN-GGLIHVVTREKGPQVCAKLFGNIQKVNVFWLLHNGFSTGIGDT 669
sp|P04051|RPC1_YEAST      QILSGVMDKSVLGDGKKHSVFYITILRDYGPQEAANAMNRMAKLCARFLGNRGSIGINDV 717
                          * : * : * : * : . : : : * : * * : * : : * : * : * : * : * :

```

sp P04050 RPB1_YEAST	<u>IADGPTMREITETIAEAKKKVLDVTKEAQANLLTAKHGMTLRESFEDNVVRFLNEARDKA</u>	729
sp P04051 RPC1_YEAST	TPADDLKQKKEELVEIAYHKCDELITLFNKGELTQPGCNEEQTLEAKIGGLLSKVREEV	777
	. . : : * : * : * : : . : . * : : * . : : : * : : : * : : : .	
sp P04050 RPB1_YEAST	<u>GRLAEVNLKDLNNVKQMVMAGSKGSFINIAQMSACVGGQSSVEGKRIAFGFVDRTLPHFSK</u>	789
sp P04051 RPC1_YEAST	GDVCINELDNWNAPLIMATCGSKGSTLNVSQMVAVVGQIIISGNRVPDGFQDRSLPHFPK	837
	* : . : * : : * * . : : : * : : : * : : : * : : : * : : : *	
sp P04050 RPB1_YEAST	<u>DDYSPESKGFVENSYLRLTPQEFFFFHAMGGREGLIDTAVKTAETGYIQRRLVKALEDIM</u>	849
sp P04051 RPC1_YEAST	NSKTPQSKGFVRNSFFSGLSPPEFLFHAISGREGLVDTAVKTAETGYMSRRLMKSLEDLS	897
	: . : : : * : : : * : : * : : : * : : : * : : : * : : : * : : : *	
sp P04050 RPB1_YEAST	<u>VHYDNTTRNSLGNVIQFIYGEDGMDAAHIEKQSLDTIGGSDAAFEKRYRVDLLNTDHTLD</u>	909
sp P04051 RPC1_YEAST	CQYDNTVRTSANGIVQFTYGGDGLDPLEMEGNAQPVN--FNRSWDHAYNITFNNQDKGLL	955
	: * : : * . * . : : : * * : : * . : : * : : . : : : * : : : * : : *	
sp P04050 RPB1_YEAST	<u>P-SLLESGSEILGDLKLQVLLDEEYKQLVKDRKFLREVFDGEANWPLPVNIRRIIQNAQ</u>	968
sp P04051 RPC1_YEAST	PYAIMETANEILGPLEERLVRYDNSGCLVKREDLNKAEYVD----QYDAERDFYHSLR	996
	* : : : * : . * : : * : : : : * : : : : * : : : : * : : : : * : : *	
sp P04050 RPB1_YEAST	<u>QTFHIDHTKPSDLTIKDIVLGVKDLQENLLVLRGKNEIIQNAQRDAVTLFCCLLRSLAT</u>	1028
sp P04051 RPC1_YEAST	EYIN-----GKATALANLRKSRGMLGLEPPAKELQGIDPDETVPDNVK	1053
	----- * * * : : . : : : . .	
sp P04050 RPB1_YEAST	<u>RRVLQEYRLTKQAFDWVLSNIEAQFLRSVHPGEMVGVLAQAQSIGEPATQMTLNTFHFAG</u>	1088
sp P04051 RPC1_YEAST	TSVSQLYRISEKSVRKFLFIALFKYRKARLEPGTAIGAIGAQSIGEPGTQMTLTKTFHFAG	1113
	* * * : : : : . * . : : : : * * : : : * : : : * : : : * : : *	
sp P04050 RPB1_YEAST	<u>VASKKVTSGVPRLKEILNVAKNMKTPSLTVYLEPGHAADQEQAKLIRSAIEHTTLKSVTI</u>	1148
sp P04051 RPC1_YEAST	VASMNVTLGVPRIKEIINASKVISTPIINAVLVN--DNDEARARVVKGRVEKTLSDVAF	1171
	* : : * * : : : * : : : * : : : * : : : * : : : * : : : * : : *	
sp P04050 RPB1_YEAST	<u>ASEIYYDPDRSTVIPEDDEEIIQLHFSLLDEEAEQSFDQQSPWLLRLELDRAAMNDKDLT</u>	1208
sp P04051 RPC1_YEAST	YVQDVYKDN-----LSFIQVRIDLGTIDKLQLELTIEDIA	1206
	: * . : . : : . . . . . : . . . : * * * : : : : : : : : : helixPHD	
sp P04050 RPB1_YEAST	<u>MGQVGERIKQTFKNDLFVIWSEDNDEKLIIRCRVVRPKSLDAETEAEDHMLKKIEN-TM</u>	1267
sp P04051 RPC1_YEAST	VAITRASKLKIQASDVNIIGKDRIAINVFPEGYKAKSISTSAKEPSENDVFYRMQQLRRA	1266
	: . . : : * : : * : : : : . . . * . * : : * : : : : :	
sp P04050 RPB1_YEAST	<u>LENITLRGVENIERVMMKYDRKVPSPTEGYVKEPEWVLETDGVNLSEVMTVPGIDPTRI</u>	1327
sp P04051 RPC1_YEAST	LPDVVVKGLPDISRAVINIRD-----DGKRELLVEGYGLRDMCTDGVIGSRT	1314
	* : : : : : : * . * : * . : : : . : : * . * . * : * : * : *	
sp P04050 RPB1_YEAST	<u>YTNSFIDIMEVLGIEAGRAALYKEVYNVIASDGSYVNYRHMALLVDVMTTQGGTTSVTRH</u>	1387
sp P04051 RPC1_YEAST	TTNHVLEVFVSLGIEAARYSIIREINYTMSNHGMSVDPRIQLLGDVMTYKGEVLGITRF	1374
	* * : : : : * : : * : : : : . : . * : * : * * : : : : * : *	
sp P04050 RPB1_YEAST	<u>GFNRSNTGALMRCSFEETVEILFEAGASAELEDDCRGVSENVILGQMAPIGTGAFDVMIDE</u>	1447
sp P04051 RPC1_YEAST	GLSKMRDSVLQLASFEKTTDHLFDAAFYMKKDAVEGVSECIILGQTMSIGTGSFKVVKG-	1433
	* : . : . . * . * : : * : * : * . * : * : * : * : * : * : *	

```

sp|P04050|RPB1_YEAST    ESLVKYMPEQKITEIEDGQDGGVTPYSNESGLVNADLDVKDELMFSPLVDSGSNDAMAGG 1507
sp|P04051|RPC1_YEAST    -----TNISEKDLVPKRCLFESLSNEAALKA 1459
                          .: . . . . :. . . . . :. :.:. . :.:. ** :   *.:* **:*.
                          .

sp|P04050|RPB1_YEAST    FTAYGGADYGEATSPFGAYGEAPTSPGFGVSSPGFSPTSPTYSPTSPAYSPTSPPSYSPTS 1567
sp|P04051|RPC1_YEAST    NCLUSTAL----- 1467
                          . * .:.:. .: .:.:. . :. . :.:. :.:. :.:. :.:.

```

**Rpb2-Rpa2**

sp P08518 RPB2_YEAST	MSDLANSE-KYYDEDPYGFEDESAPITAEDSWAVISAFFREKGLVSQLDSFNQFVDYTL 59
sp P22138 RPA2_YEAST	MSKVIKPPGQARTADFRTERESRFINPPKDKSAFPLLQEAVQPHIGSFNALTEGPDGGL 60
	**.: :. : * : ** * *.. .. :.:. : . :.:.: : * *
sp P08518 RPB2_YEAST	QDIICEDS--TLILEQLAQHTTE---SDNISRKYEISFGKIYVTKPMVNE--SDGVTHA 111
sp P22138 RPA2_YEAST	LNLGVKDIGEKVIFDGKPLNSEDEISNSGYLGKLSVSVEQVSIAPKPSNDGVSSAVERK 120
	:: :* .:.: : . : : * . :.* .:*. : :**** *: *. * :
sp P08518 RPB2_YEAST	LYPQEARLRNLTYS SGLFVDVKKRTYEADVPGRELKYELIAEESEDDSESGKVFIGRLP 171
sp P22138 RPA2_YEAST	VYPSESQRQLTSYRGKLLKLKWSVNN-----GEENLFEVRD-----CGGLP 162
	:**.*:* * : * . *.:.* . : : . . *.* : : . :. :. :. : * *
sp P08518 RPB2_YEAST	IMLRSKNCYLSEATESDLYKLKECPFDMGGYFIINGSEKVLIAQERSAGNIVQVFKKAAP 231
sp P22138 RPA2_YEAST	VMLQSNRCHLNKMSPYELVQHKEESDEIGGYFIVNGIEKLIRMLIVQRRNHPMAIIRPSF 222
	:***:.**: : : * : ** . :*****:** *: : . * . : : :
sp P08518 RPB2_YEAST	SPISHVAEIRSALEKGSRFISTLQVKLYGREGSSARTIKATLPYIKQDIPIVIIIFRALGI 291
sp P22138 RPA2_YEAST	ANRGASYSHYGIQIRSVRPDQTSQTNLHYLNDGQVTFRFSWRKNEYLPVVMILKALCH 282
	: . . . :. * . * .: : . . * : : : : : : : : : : : *
sp P08518 RPB2_YEAST	IPDGEILEHIC-YDVNDWQMLEMLKPCVEDG---FVIQDRETALDFIGR--RGTALGIK 344
sp P22138 RPA2_YEAST	TSDREIFDGIIGNDVKDSFLTDRLELLLRGFKKRYPHLQNRQVLQYLGDKFRVVFQASP 342
	. * **:: * **:* : : * : .. :*** .:*** * . .
sp P08518 RPB2_YEAST	KEKRIQYAKDILQKEFLPHITQLEGFESRKAFFLGYMINRLLLCALDRKDQDDRDHFGKK 404
sp P22138 RPA2_YEAST	DQSDLEVQGQEVLDRIVLVHLGKDG--SQDKFRMLLFMIRKLYSLVAGECSPDNPDATQHQ 400
	.: . : : : : : * * : : . . * : * :***.* . . . * : * : :
sp P08518 RPB2_YEAST	RDLLAGPLLAQLFKTLFKKLTKDIFRYMQRTVEEAHDFNMK-----LAINAKTITSG 456
sp P22138 RPA2_YEAST	EVLLGGFLYGMILKEKIDEYLQNIIAQVRMDINRGMAINFKDKRYMSRVLMRVNENIGSK 460
	.: *.* * . : : * : : : : : : : : : : : : : : * :.* *
sp P08518 RPB2_YEAST	LKYALATGNWGEQK-KAMSSRAGVSQVLNRYTYSSTLSHLRRTN-TPIGRDGKLAKPRQL 514
sp P22138 RPA2_YEAST	MQYFLSTGNLVSQSGLDLQVSGYTVVAEKINFYRFISHFRMVHRGSFFAQLKTTTVRKL 520
	: : * * :*** .*. :. : * : * : : : : : :***. : . : : * . : * :
sp P08518 RPB2_YEAST	HNTHWGLVCPAETPEGQACGLVKNL SLMSCISVG-TDPMPIITFLSEWGMEPLEDYVPHQ 573
sp P22138 RPA2_YEAST	LPESWGFLCPVHTPDGSPCGLLNHFHAKCRISTQQSDVSRIPSILYSLGVAPASHTFAAG 580
	**::**..**:*..**::: : . ** . : * * : : * . * : * ..
sp P08518 RPB2_YEAST	SPDATRVFVNGVWHGV--HRNPARLMETLRTLRRKGDINPEVSMIRDIREKELKIFTDAGRV 633
sp P22138 RPA2_YEAST	-PSLCCVQIDGKIIGWSHSEQGKIADTLRYWKVEGKTPGLPIDLEIG---YVPPSTRGQ- 636
sp P08518 RPB2_YEAST	YRPLFIVEDDES LGHKELKVRKGHIAKL MATEYQDIEGGFEDVEEYTWSSLLNEGLVEYI 693
sp P22138 RPA2_YEAST	YPGLYLFGG-----HSRMLRPVRYLPLDK-----EDIV 662
	* * : : . . : . . : * : : : . : :*** * **:* : : : . . ---

SUPPORTING MATERIAL: SEQUENCE ALIGNMENTS OF HOMOLOGOUS SUBUNITS IN POL I, POL II AND POL III

sp P08518 RBP2_YEAST	<u>DAEEEE</u> SILIAMQPEDLEPAEANEENDLDVDPAKRIRVSHHATTFTHCEIHPSMILGVAA 753
sp P22138 RPA2_YEAST	GPFEQVYMNIAVTPQEIQ-----NNVHTHVEFTPTNLSILA 701
	----- **: *:::..:..... . .: : ...** *: *: **: *
sp P08518 RBP2_YEAST	<u>SIIPFPDHNQSPRNTYQS</u> AMGKQAMGVFLTNYNVRMDTMANILYYPKPLGTTAMEYLYK 813
sp P22138 RPA2_YEAST	NLTPFSDFNQSPRNMYYCQMGKQTMGTPGVALCHRSDNKLYRLQTGQTPIVKANLYDDYG 761
	.: **. *.***** **. *****:.. . * * * * *: .: :
sp P08518 RBP2_YEAST	<u>FREL</u> PAGQNAIVAIACYSGYNQEDSMIMNQSSIDRGLFRSLFFRSYMDQEKKYGMSITET 873
sp P22138 RPA2_YEAST	MDNFPNGFNNAVVAVISYGYDMDAMIINKSADERGFGYGTMYKTEK-VDLALNRNRGDP 820
	: ::* * **::: .*:::: :*::::***: :*: . :::: . : . . :.
sp P08518 RBP2_YEAST	FEKPQRTNTRLRMKHGTYD <u>KLDDGLI</u> APGVRVSGEDVIIGKTTPISPDEEELGQRTAYHS 933
sp P22138 RPA2_YEAST	ITQHFGFGNDEWPKEWLEKLDDEGLPYIGTYVEEGDPICAYFDDT-----LNKTKIKT 873
	: : . . : :***:** * *-----*----- :*: :
sp P08518 RBP2_YEAST	KRDASTPLRSTENGIVDQVLVTTNQDGLKFVKVRVRTTKIPQIGDKFASRHGQKGTIGIT 993
sp P22138 RPA2_YEAST	YHSSEPAYIEEVNLIGEDSNKFQE--LQTVSIKYRIRRTQPIGDKFSSRHGQKGVCSRK 930
	:.:... . * * *: :.: : *: *.: * : *****:*****. . .
sp P08518 RBP2_YEAST	<u>YRREDMPFTAEGIVPDLI</u> INPHAIPSRMTVAHLIECLLSKVAALSGNEGDAFPFT---D 1049
sp P22138 RPA2_YEAST	WPTIDMPFSETGIQPDIIINPHAFPSRMTIGMFVESLAGKAGALHGIAQDSTPWIFNEDD 990
	: *****: ** **::*****:*****.: :*. * *.** * **::: *
sp P08518 RBP2_YEAST	<u>ITVEGISKLLREHGYS</u> RGFEVYNGHTGKKLMAQIFFGPTYQRLRHMVDDKI HARARG 1109
sp P22138 RPA2_YEAST	TPADYFGEQLAKAGYNYHGNEPMYSGATGEELRADIYGVVYYQRLRHMVNDKFQVRSTG 1050
	..: :.: * : ** : * * **.* **::* **::.* .*****:*****:..*: *
sp P08518 RBP2_YEAST	<u>PMQVLTRQPVEGRSRD</u> GGLRFGEMERDCMIAHGAASFLKERLMEASDAFRVHICGICGLM 1169
sp P22138 RPA2_YEAST	PVNSLTMQPVKGRKRHGIRVGEMERDALIGHGTSFLLQDRLLNSSDYTQASVCRECGSI 1110
	*.: ** **::**.*.*.*:..*****.:*.**:: :*::::***: .: :* ** :
sp P08518 RBP2_YEAST	<u>TVIAKLN</u> ----HNQFECKGCDN-----K 1188
sp P22138 RPA2_YEAST	LTTQQSVPRIGSISTVCCRRCSMRFEDAKKLLTKSEDEKIFIDDSQIWEDGQGNKFVGG 1170
	. : . . *: *
sp P08518 RBP2_YEAST	<u>IDIQIHIPIYA</u> AKLLFQELMAMNITPRLYTDRSRDF 1224
sp P22138 RPA2_YEAST	NETTTVAIPFVLKYLDSELSAMGIRLRYNVEPK--- 1203
	: : **:. * * ** **.* * .: . .

**Rpb2-Rpc2**

sp P08518 RPB2_YEAST	---MSDLANSEKYYDEDPYGFEDESAPI----- <u>TAEDSWAVISAFFREKGLVSQ</u>	46
sp P22276 RPC2_YEAST	MVAATKRRKTHIHKHVKDEAFDDLKPVYKGKKLTDEINTAQDKWHLPAFLKVKGLVKQ	60
	:. :. : . . .*:* * : * : * : * : *	
sp P08518 RPB2_YEAST	<u>QLDSFNQFVDYTLQDIICEDSTLILEQLAQHTTESDNISRKYEISFGKIYVTKPMVNESD</u>	106
sp P22276 RPC2_YEAST	HLDSFNQFVDTLKKIIKAN----- <u>QLILSDVDPEFYLYVDIRVGKKSSSS</u>	107
	:***** * :.* : : ..*	
sp P08518 RPB2_YEAST	GVTHALYPQEARLRNLTYSGLFVDVKKRTYEAIDVPGRELKYELIAEESDDSESGKVF	166
sp P22276 RPC2_YEAST	TKDYLTTPPHECLRDMTYSAPIYVDIEYTR-----GR----NIIMHKD-----VE	148
	: * :.* :***** : : : : * : : * : *	
sp P08518 RPB2_YEAST	<u>IGRLPIMLRSKNCYLSEATESDLYKLKECPFDMGGYFIINGSEKVLIAQERSAGNIVQVF</u>	226
sp P22276 RPC2_YEAST	IGRMPIMLRSNKCILYDADESKMAKLNCEPLDPGGYFIVNGTEKVVILVQEQLSKNRIIVE	208
	***:*****: * * : * * : * : * : * : * : * : * : * : *	
sp P08518 RPB2_YEAST	<u>KKAAPSPISHVAEIRSALEKGSRFISTLQVKLYGREGSSARTIKATLPYIKQDIPIVIIF</u>	286
sp P22276 RPC2_YEAST	ADEKK--GIVQASVTSS--THERKSKTYVITKNGK-----IYLKHNSIAEEIPIAIVL	257
	. . ----*-----*----- * * : : * : * : *	
sp P08518 RPB2_YEAST	<u>RALGIIPDGEILEHICYDVNDWQMLEMLKPCVEDGFVIQDRETALDFIGRRGTALGIKKE</u>	346
sp P22276 RPC2_YEAST	KACGILSDLEIMQLVCGNDSSYQDIFAVNLEESSKLDIYTQQQALEYIGAKVKTMRRQKL	317
	: * * : * * : : * : : * : : * : : * : : * : : *	
sp P08518 RPB2_YEAST	<u>KRIQYAKDILQKEFLPHITQLEGFESRKAFFLGYMINRLLLCALDRKDQDDRDHFGKKRL</u>	406
sp P22276 RPC2_YEAST	TILQEGIEAIIATTVIAHLTVEALDFREKALYIAMMTRRVVMAMYNPKMIDDRDYVGNKRL	377
	. : * . : : . . : * : * : : . * : : : . * * : : * : *	
sp P08518 RPB2_YEAST	<u>DLAGPLLAQLFKTLFKKLTKDIFRYMQRTVEEAH---DFNMKLAINAKT--ITSGLKYAL</u>	461
sp P22276 RPC2_YEAST	ELAGQLISLLFEDLFKKFNNDFKLSIDKVLKKNRAMEYDALLSINVHNSNITSGLNRAI	437
	: * * * : : * : * : * : : : : : : : * : * : * : *	
sp P08518 RPB2_YEAST	<u>ATGNWGEQKKAMSSRAGVSQVLNRYTYSSTLSHLRRTNTPIGRDGKLAKPRQLHNTHWGL</u>	521
sp P22276 RPC2_YEAST	STGNWS-LKRFKMERAGVTHVLSRLSYISALGMMTRISSQFEKSRKVSQPRALQPSQFGM	496
	: * * * . * : . * : : * : * : * : * : * : * : *	
sp P08518 RPB2_YEAST	<u>VCPAETPEGQACGLVKNLMSLMSISVGTDPMPITFLSEWGMEPLEDYVPHQSPDATRVF</u>	581
sp P22276 RPC2_YEAST	LCTADTPEGEACGLVKNLALMTHITTDDEEPIKKLCYVLGVEDITLIDSASLHLNMGVY	556
	: * . * : * : * : * : * : * : * : * : * : *	
sp P08518 RPB2_YEAST	<u>VNGVWHGVHRNPARMETLRTLRRKGDINPEVSMIRDIREKELKIFTDAGRVRPLFIVE</u>	641
sp P22276 RPC2_YEAST	LNGTLIGSIRFPFKFTVQFRHLRRTGKVSEFISISNSHQMAVHIATDGRICRPLIIVS	616
	: * * . * * * : : : : * * * : * : : : : * * : * : *	
sp P08518 RPB2_YEAST	<u>DDESLGHKELKVRKGHIAKLMATEYQDIEGGFEDVEEYTWSSLLNEGLVEYIDAEESIESI</u>	701
sp P22276 RPC2_YEAST	DGQSR-----VKDIHLRKLDD-----GELDFDDFLKLGLVEYLDVNEENDS	657
	* : * * : * : * : * : * : * : * : *	
sp P08518 RPB2_YEAST	<u>LIAMQPEDLEPAEANEENDLDVDPKRIRVSHHATTFTHCEIHPSMILGVAASIIPFPDH</u>	761
sp P22276 RPC2_YEAST	YIALYEKDIVP-----SMTHLEIEPFTILGAVAGLIPIPHH	693

[illegible]

**Rpb11/AC19**

```

sp|P38902|RPBY_YEAST -----MNAPDRFELFLLGEGES-----KLKIDPDTKAPNA 30
sp|P28000|RPC9_YEAST MTEDIEQKKTATEVTPQEPKHIQEEEEQDVDMTGDEEQEEEPDREKIKLLTQATSEDGTS 60
                        ::  .  :: :  *  *  .  **  .  ::  .  :

sp|P38902|RPBY_YEAST VVITFEKEDHTLGNLIRAELLNDRKVLFAAYKVEHPFFARFKLRIQTTEGYDPKDALKNA 90
sp|P28000|RPC9_YEAST ASFQIVEEDHTLGNALRYVIMKNPDVEFCGYSIPHPSENLLNIRIQTYGETTAVDALQKG 120
.  :  :  :*****  :*  ::::  .*  *.  .  :  **  ::*****  .  ****:.

sp|P38902|RPBY_YEAST CNSIINKLGALKTNFETEWNLQTLAADDAF 120
sp|P28000|RPC9_YEAST LKDLMDLCDVVESKFTTEKIKSM----- 142
                        :.:::  .:::.*  :  :  :  :.:.:

```

**Rpb3/AC40**

```

sp|P16370|RPB3_YEAST -----MSEEGP-----QVKIREASKDNVDFILSNVD 26
sp|P07703|RPC5_YEAST MSNIVGIEYNRVINTTSTDTPFGFSKDAENENWVEKFKKDFEVNISSLDAREANFDLINID 60
                        ::::.  :*:  .  .  :.:  *  *  *:

sp|P16370|RPB3_YEAST LAMANSLRRVMIAEIPTLAIDSVEVETNTTVLADEFIAHRLGLIPLQSMDIEQLEYSRDC 86
sp|P07703|RPC5_YEAST TSIANAFRRIMISEVPSVAAEYVYFFNNTSVIQDEVLAHRIGLVPLK-VDPDMLTWVDSN 119
:****:****:****:  :  *  .  .  ****:  **.:****:****:****:  :  *  :  .

sp|P16370|RPB3_YEAST FCED--HCDKCSVVLTLQAFGESE-----STTNVYSKDLVIVSNLMGRNIG 130
sp|P07703|RPC5_YEAST LPDDEKFTDENTIVLSLVNCKTRNPDAPKGSTDPKELYNNAHVYARDLKFEPPQGRQSTTF 179
:  :  *  .  *:  ::****:  .  :  .::****:  :  :  .

sp|P16370|RPB3_YEAST HP I IQDKEGNGVLICKLRKGQELKLTCAKKGIAKEHAKWGPAAAIEFEYDPWNKLKH-- 188
sp|P07703|RPC5_YEAST ADCPVVPADPDILLAKLRPGQEISLKAHCILGIGGDHAKFSVPVSTASYRLLPQINILQPI 239
.  .:*.***  **:.*.  .  **.  :****:*.::  .:  *  :  :

sp|P16370|RPB3_YEAST -----TDYWYEQDSAKEWPQSKNCEYEDPPNEGDPFDYKAQAD 226
sp|P07703|RPC5_YEAST KGESARRFQKCFPPGVIGIDEGSDEAYVKDARKDTSREVLRYEEFADK---VKLGRVRN 296
                        :  *  *:  *:  .  :  .*:  .::.....  :

sp|P16370|RPB3_YEAST TFYMNVESVGSIPVDQVVVRGIDTLQKKVASIL-LALTQMDQDKVNFASGDNNTASN--- 282
sp|P07703|RPC5_YEAST HFIFNVESAGAMTPEEIFFKSVRILKNKAEYLNKCPITQMTEDIEQKKTATEVTPQEPKH 356
*  :****.***:  .:::..:  :  *:  *  .  :  .:***  :  :  :  :  *  .:

sp|P16370|RPB3_YEAST -MLGSNEDVMMTG---AEQDPYSNASQMNTGSG--GYDNAW----- 318
sp|P07703|RPC5_YEAST IQEEEEQDVDMTGDEEQEEEPDREKIKLLTQATSEDGTSASFQIVEEDHTLGNALRYVIM 416
.::**  ***  *:  :  :  .  .:  *  .  :

sp|P16370|RPB3_YEAST -----
sp|P07703|RPC5_YEAST KNPDVEFCGYSIPHPSENLLNIRIQTYGETTAVDALQGLKDLMDLCDVVESKFTTEKIKS 476

sp|P16370|RPB3_YEAST -
sp|P07703|RPC5_YEAST M 47

```

## References

1. Adams, M. D., et al. (2000). *Science* **287**(5461): 2185-95.
2. Akoulitchhev, S., et al. (2000). *Nature* **407**(6800): 102-6.
3. Allison, L. A., et al. (1985). *Cell* **42**(2): 599-610.
4. Andel, F., 3rd, et al. (1999). *Science* **286**(5447): 2153-6.
5. Andersen, G., et al. (1996). *FEBS Lett* **397**(1): 65-9.
6. Archambault, J., et al. (1997). *Proc Natl Acad Sci U S A* **94**(26): 14300-5.
7. Armache, K. J., et al. (2003). *Proc Natl Acad Sci U S A* **100**(12): 6964-8.
8. Armache, K. J., Kettenberger H., and Cramer P. (2005). *Curr Opin Struct Biol*.
9. Armache, K. J., et al. (2005). *J Biol Chem* **280**(8): 7131-4.
10. Asturias, F. J. (2004). *Curr Opin Struct Biol* **14**(2): 121-9.
11. Asturias, F. J., et al. (1998). *Ultramicroscopy* **70**(3): 133-43.
12. Asturias, F. J., et al. (1999). *Science* **283**(5404): 985-7.
13. Asturias, F. J., et al. (1997). *J Mol Biol* **272**(4): 536-40.
14. Baek, H. J., et al. (2002). *Mol Cell Biol* **22**(8): 2842-52.
15. Baker, D. and A. Sali (2001). *Science* **294**(5540): 93-6.
16. Ban, N., et al. (2000). *Science* **289**(5481): 905-20.
17. Ban, N., et al. (2000). *Science* **289**(5481): 905-20.
18. Bangur, C. S., et al. (1997). *Mol Cell Biol* **17**(12): 6784-93.
19. Barberis, A., et al. (1993). *Proc Natl Acad Sci U S A* **90**(12): 5628-32.
20. Barton, G. J. (1993). *Protein Eng* **6**: 37-40.
21. Bastien, J., et al. (2000). *J Biol Chem* **275**(29): 21896-904.
22. Baumli, S., et al. (2005). *J Biol Chem*.
23. Bell, S. D. and S. P. Jackson (2000). *J Biol Chem* **275**(17): 12934-40.
24. Bensaude, O., et al. (1999). *Biochem Cell Biol* **77**(4): 249-55.
25. Bentley, D. (2002). *Curr Opin Cell Biol* **14**(3): 336-42.
26. Berk, A. J. (2000). *Cell* **103**(1): 5-8.
27. Birck, C., et al. (1998). *Cell* **94**(2): 239-49.
28. Bischler, N., et al. (2002). *Embo J* **21**(15): 4136-44.
29. Blazek, E., et al. (2005). *Chromosoma*.
30. Bleichenbacher, M., et al. (2003). *J Mol Biol* **332**(4): 783-93.
31. Bork, P. (2002). *Bioinformatics* **18 Suppl 2**: S64.
32. Brand, M., et al. (1999). *Science* **286**(5447): 2151-3.
33. Brunger, A. T., et al. (1998). *Acta Crystallogr D Biol Crystallogr* **54 ( Pt 5)**:
34. Budisa, N., et al. (1995). *Eur J Biochem* **230**(2): 788-96.
35. Buratowski, S., et al. (1989). *Cell* **56**(4): 549-61.
36. Buratowski, S. and H. Zhou (1993). *Proc Natl Acad Sci U S A* **90**(12): 5633-7.
37. Burke, T. W. and J. T. Kadonaga (1996). *Genes Dev* **10**(6): 711-24.
38. Burke, T. W. and J. T. Kadonaga (1997). *Genes Dev* **11**(22): 3020-31.
39. Bushnell, D. A., et al. (2002). *Proc Natl Acad Sci U S A* **99**(3): 1218-22.
40. Bushnell, D. A. and R. D. Kornberg (2003). *Proc Natl Acad Sci U S A* **100**(12): 7131-4.
41. Bushnell, D. A., et al. (2004). *Science* **303**(5660): 983-8.
42. Campbell, E. A., et al. (2001). *Cell* **104**(6): 901-12.
43. Carcamo, J., et al. (1991). *Proc Natl Acad Sci U S A* **88**(18): 8052-6.
44. Carroll, S. B. and B. D. Stollar (1983). *J Mol Biol* **170**(3): 777-90.
45. Carson, M. (1997). *Meth. Enzym.* **277**: 493-505.
46. Chalkley, G. E. and C. P. Verrijzer (1999). *Embo J* **18**(17): 4835-45.
47. Chambers, R. S., et al. (1995). *J Biol Chem* **270**(25): 14962-9.

48. Chang, W. H. and R. D. Kornberg (2000). *Cell* **102**(5): 609-13.
49. Chasman, D. I., et al. (1993). *Proc Natl Acad Sci U S A* **90**(17): 8174-8.
50. Chen, D., et al. (2000). *Mol Cell* **6**(1): 127-37.
51. Chen, H. T. and S. Hahn (2003). *Mol Cell* **12**(2): 437-47.
52. Chen, H. T., et al. (2000). *Protein Sci* **9**(9): 1743-52.
53. Chicca, J. J., 2nd, et al. (1998). *Mol Cell Biol* **18**(3): 1701-10.
54. Cho, J. M., et al. (1982). *Biochem Pharmacol* **31**(16): 2583-9.
55. Choder, M. and R. A. Young (1993). *Mol Cell Biol* **13**(11): 6984-91.
56. Choi, C. H., et al. (2003). *Nature* **424**(6951): 965-9.
57. Chou, S., et al. (1999). *Genetics* **153**(4): 1573-81.
58. Chung, W. H., et al. (2003). *Mol Cell* **12**(4): 1003-13.
59. Conaway, R. C. and J. W. Conaway (1989). *Proc Natl Acad Sci U S A* **86**(19): 6174-8.
60. Conaway, R. C. and J. W. Conaway (1993). *Annu Rev Biochem* **62**: 161-90.
61. Corden, J. L. (1990). *Trends Biochem Sci* **15**(10): 383-7.
62. Corden, J. L. (1993). *Curr Opin Genet Dev* **3**(2): 213-8.
63. Cox, J. M., et al. (1997). *Proc Natl Acad Sci U S A* **94**(25): 13475-80.
64. Craighead, J. L., et al. (2002). *Structure (Camb)* **10**(8): 1117-25.
65. Cramer, P. (2002). *Curr Opin Struct Biol* **12**(1): 89-97.
66. Cramer, P. (2004). *Adv Protein Chem* **67**: 1-42.
67. Cramer, P., et al. (2000). *Science* **288**(5466): 640-9.
68. Cramer, P., et al. (2001). *Science* **292**(5523): 1863-76.
69. Dahmus, M. E. (1996). *Methods Enzymol* **273**: 185-93.
70. Darst, S. A. (2001). *Curr Opin Struct Biol* **11**(2): 155-62.
71. Darst, S. A., et al. (1991). *Cell* **66**(1): 121-8.
72. Darst, S. A., et al. (1991). *J Mol Biol* **221**(1): 347-57.
73. Darst, S. A., et al. (1989). *Nature* **340**(6236): 730-2.
74. Davey, C. A. and T. J. Richmond (2002). *Proc Natl Acad Sci U S A* **99**(17): 10880-5.
75. Davis, J. A., et al. (2002). *Mol Cell* **10**(2): 409-15.
76. DeDecker, B. S., et al. (1996). *J Mol Biol* **264**(5): 1072-84.
77. Dignam, J. D., et al. (1983). *Methods Enzymol* **101**: 582-98.
78. Douziech, M., et al. (2000). *Mol Cell Biol* **20**(21): 8168-77.
79. Drapkin, R., et al. (1994). *Nature* **368**(6473): 769-72.
80. Dvir, A., et al. (1997). *J Biol Chem* **272**(45): 28175-8.
81. Ebright, R. H. (1998). *Cold Spring Harb Symp Quant Biol* **63**: 11-20.
82. Edwards, A. M., et al. (1990). *Proc Natl Acad Sci U S A* **87**(6): 2122-6.
83. Edwards, A. M., et al. (1994). *Nat Struct Biol* **1**(3): 195-7.
84. Edwards, A. M., et al. (1991). *J Biol Chem* **266**(1): 71-5.
85. Egly, J. M. (2001). *FEBS Lett* **498**(2-3): 124-8.
86. Esnouf, R. M. (1997). *J Mol Graph Model* **15**(2): 132-4, 112-3.
87. Fabrega, C., et al. (2003). *Mol Cell* **11**(6): 1549-61.
88. Fairley, J. A., et al. (2002). *Mol Cell Biol* **22**(19): 6697-705.
89. Fang, S. M. and Z. F. Burton (1996). *J Biol Chem* **271**(20): 11703-9.
90. Farago, M., et al. (2003). *Mol Biol Cell* **14**(7): 2744-55.
91. Feaver, W. J., et al. (2000). *J Biol Chem* **275**(8): 5941-6.
92. Ferri, M. L., et al. (2000). *Mol Cell Biol* **20**(2): 488-95.
93. Flanagan, P. M., et al. (1991). *Nature* **350**(6317): 436-8.
94. Flores, O., et al. (1990). *J Biol Chem* **265**(10): 5629-34.
95. Forget, D. and B. Coulombe (2003). *Methods Enzymol* **370**: 701-12.
96. Forget, D., et al. (1997). *Proc Natl Acad Sci U S A* **94**(14): 7150-5.
97. Fu, J., et al. (1999). *Cell* **98**(6): 799-810.
98. Gaiser, F., et al. (2000). *J Mol Biol* **302**(5): 1119-27.

- 
99. Gangloff, Y. G., et al. (2000). *Mol Cell Biol* **20**(1): 340-51.
  100. Gavin, A. C., et al. (2002). *Nature* **415**(6868): 141-7.
  101. Geiduschek, E. P. and G. A. Kassavetis (2001). *J Mol Biol* **310**(1): 1-26.
  102. Geiger, J. H., et al. (1996). *Science* **272**(5263): 830-6.
  103. Gerber, H. P., et al. (1995). *Nature* **374**(6523): 660-2.
  104. Ghaemmaghami, S., et al. (2003). *Nature* **425**(6959): 737-41.
  105. Gnatt, A. L., et al. (2001). *Science* **292**(5523): 1876-82.
  106. Goffeau, A., et al. (1996). *Science* **274**(5287): 546, 563-7.
  107. Groft, C. M., et al. (1998). *Proc Natl Acad Sci U S A* **95**(16): 9117-22.
  108. Groll, M., et al. (1997). *Nature* **386**(6624): 463-71.
  109. Grummt, I. (2003). *Genes Dev* **17**(14): 1691-702.
  110. Guglielmi, B., et al. (2004). *Nucleic Acids Res* **32**(18): 5379-91.
  111. Ha, I., et al. (1993). *Genes Dev* **7**(6): 1021-32.
  112. Hahn, S. (2004). *Nat Struct Mol Biol* **11**(5): 394-403.
  113. Hahn, S., et al. (1989). *Cell* **58**(6): 1173-81.
  114. Hahn, S. and S. Roberts (2000). *Genes Dev* **14**(6): 719-30.
  115. Hampsey, M. (1998). *Microbiol Mol Biol Rev* **62**(2): 465-503.
  116. Hampsey, M. and D. Reinberg (1999). *Curr Opin Genet Dev* **9**(2): 132-9.
  117. Hawkes, N. A. and S. G. Roberts (1999). *J Biol Chem* **274**(20): 14337-43.
  118. He, C. H. and D. Ramotar (1999). *Biochem Cell Biol* **77**(4): 375-82.
  119. Ho, Y., et al. (2002). *Nature* **415**(6868): 180-3.
  120. Holstege, F. C., et al. (1998). *Cell* **95**(5): 717-28.
  121. Holstege, F. C., et al. (1996). *Embo J* **15**(7): 1666-77.
  122. Horikoshi, M., et al. (1989). *Nature* **341**(6240): 299-303.
  123. Hu, P., et al. (2002). *Mol Cell Biol* **22**(22): 8044-55.
  124. Huh, W. K., et al. (2003). *Nature* **425**(6959): 686-91.
  125. Hultmark, D., et al. (1986). *Cell* **44**(3): 429-38.
  126. Ito, T., et al. (2001). *Proc Natl Acad Sci U S A* **98**(8): 4569-74.
  127. Jacobson, R. H., et al. (2000). *Science* **288**(5470): 1422-5.
  128. Jensen, G. J., et al. (1998). *Embo J* **17**(8): 2353-8.
  129. Jones, T. A., et al. (1991). *Acta Crystallogr A* **47** ( Pt 2): 110-9.
  130. Juo, Z. S., et al. (1996). *J Mol Biol* **261**(2): 239-54.
  131. Juo, Z. S., et al. (2003). *Nature* **422**(6931): 534-9.
  132. Kamada, K., et al. (2001). *Proc Natl Acad Sci U S A* **98**(6): 3115-20.
  133. Kamada, K., et al. (2001). *Cell* **106**(1): 71-81.
  134. Kamenski, T., et al. (2004). *Mol Cell* **15**(3): 399-407.
  135. Kettenberger, H., et al. (2003). *Cell* **114**(3): 347-57.
  136. Kettenberger, H., et al. (2004). *Mol Cell* **16**(6): 955-65.
  137. Khazak, V., et al. (1998). *Mol Cell Biol* **18**(4): 1935-45.
  138. Khazak, V., et al. (1995). *Mol Biol Cell* **6**(7): 759-75.
  139. Killeen, M., et al. (1992). *J Biol Chem* **267**(14): 9463-6.
  140. Kim, J. L., et al. (1993). *Nature* **365**(6446): 520-7.
  141. Kim, M., et al. (2004). *Nature* **432**(7016): 517-22.
  142. Kim, T. K., et al. (2000). *Science* **288**(5470): 1418-22.
  143. Kim, T. K., et al. (1997). *Proc Natl Acad Sci U S A* **94**(23): 12268-73.
  144. Kim, Y., et al. (1993). *Nature* **365**(6446): 512-20.
  145. Kim, Y. J., et al. (1994). *Cell* **77**(4): 599-608.
  146. Kimura, M., et al. (2001). *Eur J Biochem* **268**(3): 612-9.
  147. Kimura, M., et al. (2002). *Mol Cell Biol* **22**(5): 1577-88.
  148. Kobor, M. S. and J. Greenblatt (2002). *Biochim Biophys Acta* **1577**(2).
  149. Koike, H., et al. (2004). *Structure (Camb)* **12**(1): 157-68.

150. Kokubo, T., et al. (1998). *Mol Cell Biol* **18**(2): 1003-12.
151. Kolodziej, P. A., et al. (1990). *Mol Cell Biol* **10**(5): 1915-20.
152. Kosa, P. F., et al. (1997). *Proc Natl Acad Sci U S A* **94**(12): 6042-7.
153. Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**: 946-950.
154. Kuldell, N. H. and S. Buratowski (1997). *Mol Cell Biol* **17**(9): 5288-98.
155. Kumar, A., et al. (2002). *Genes Dev* **16**(6): 707-19.
156. Kutach, A. K. and J. T. Kadonaga (2000). *Mol Cell Biol* **20**(13): 4754-64.
157. Kwek, K. Y., et al. (2002). *Nat Struct Biol* **9**(11): 800-5.
158. La Fortelle, E. D., Bricogne, G (1997). *Meth. Enzym* **B**: 472-494.
159. Lagrange, T., et al. (1998). *Genes Dev* **12**(1): 34-44.
160. Larkin, R. M. and T. J. Guilfoyle (1998). *J Biol Chem* **273**(10): 5631-7.
161. Lee, T. I. and R. A. Young (2000). *Annu Rev Genet* **34**: 77-137.
162. Lei, L., et al. (1999). *Mol Cell Biol* **19**(12): 8372-82.
163. Lemon, B. and R. Tjian (2000). *Genes Dev* **14**(20): 2551-69.
164. Leuther, K. K., et al. (1996). *Cell* **85**(5): 773-9.
165. Levine, M. and R. Tjian (2003). *Nature* **424**(6945): 147-51.
166. Lewin, B. (1990). *Cell* **61**(7): 1161-4.
167. Li, S. and M. J. Smerdon (2002). *Embo J* **21**(21): 5921-9.
168. Lin, P. S., et al. (2002). *J Biol Chem* **277**(48): 45949-56.
169. Lin, P. S., et al. (2002). *Prog Nucleic Acid Res Mol Biol* **72**: 333-65.
170. Littlefield, O., et al. (1999). *Proc Natl Acad Sci U S A* **96**(24): 13668-73.
171. Liu, D., et al. (1998). *Cell* **94**(5): 573-83.
172. Liu, Y., et al. (2001). *J Biol Chem* **276**(10): 7169-75.
173. Lolli, G., et al. (2004). *Structure (Camb)* **12**(11): 2067-79.
174. Magill, C. P., et al. (2001). *J Biol Chem* **276**(50): 46693-6.
175. Maillet, I., et al. (1999). *J Biol Chem* **274**(32): 22586-90.
176. Malik, S. and R. G. Roeder (2000). *Trends Biochem Sci* **25**(6): 277-83.
177. Manley, J. L., et al. (1980). *Proc Natl Acad Sci U S A* **77**(7): 3855-9.
178. Manley, J. L., et al. (1983). *Methods Enzymol* **101**: 568-82.
179. Matsui, T., et al. (1980). *J Biol Chem* **255**(24): 11992-6.
180. Maxon, M. E., et al. (1994). *Genes Dev* **8**(5): 515-24.
181. McKune, K., et al. (1993). *Yeast* **9**(3): 295-9.
182. Meinhart, A., et al. (2003). *J Biol Chem* **278**(48): 48267-74.
183. Meinhart, A. and P. Cramer (2004). *Nature* **430**(6996): 223-6.
184. Meinhart, A., et al. (2003). *J Biol Chem* **278**(18): 15917-21.
185. Meininghaus, M., et al. (2000). *J Biol Chem* **275**(32): 24375-82.
186. Meka, H., et al. (2003). *Nucleic Acids Res* **31**(15): 4391-400.
187. Meredith, G. D., et al. (1996). *J Mol Biol* **258**(3): 413-9.
188. Merino, A., et al. (1993). *Nature* **365**(6443): 227-32.
189. Mermelstein, F., et al. (1996). *Genes Dev* **10**(8): 1033-48.
190. Minakhin, L., et al. (2001). *Proc Natl Acad Sci U S A* **98**(3): 892-7.
191. Misteli, T. (2000). *J Cell Sci* **113** ( Pt 11): 1841-9.
192. Mitsuzawa, H., et al. (2003). *Nucleic Acids Res* **31**(16): 4696-701.
193. Mittler, G., et al. (2001). *EMBO Rep* **2**(9): 808-13.
194. Miyao, T., et al. (2001). *J Biol Chem* **276**(49): 46408-13.
195. Mukherjee, K. and D. Chatterji (1997). *Eur J Biochem* **247**(3): 884-9.
196. Murakami, K. S., et al. (2002). *Science* **296**(5571): 1285-90.
197. Murakami, K. S., et al. (2002). *Science* **296**(5571): 1280-4.
198. Murray, S., et al. (2001). *Mol Cell Biol* **21**(13): 4089-96.
199. Myers, L. C. and R. D. Kornberg (2000). *Annu Rev Biochem* **69**: 729-49.
200. Na, X., et al. (2003). *Embo J* **22**(16): 4249-59.

- 
201. Naar, A. M., et al. (2002). *Genes Dev* **16**(11): 1339-44.
  202. Navaza, J. (1994). *Acta Crystallogr A*: 157-163.
  203. Nikolov, D. B. and S. K. Burley (1994). *Nat Struct Biol* **1**(9): 621-37.
  204. Nikolov, D. B., et al. (1996). *Proc Natl Acad Sci U S A* **93**(10): 4862-7.
  205. Nikolov, D. B., et al. (1995). *Nature* **377**(6545): 119-28.
  206. Nikolov, D. B., et al. (1992). *Nature* **360**(6399): 40-6.
  207. Nogales, E. and N. Grigorieff (2001). *J Cell Biol* **152**(1): F1-10.
  208. Nouraini, S., et al. (1996). *Mol Cell Biol* **16**(11): 5985-96.
  209. Ohkuma, Y. and R. G. Roeder (1994). *Nature* **368**(6467): 160-3.
  210. Ohkuma, Y., et al. (1990). *Proc Natl Acad Sci U S A* **87**(23): 9163-7.
  211. Okuda, M., et al. (2004). *J Biol Chem* **279**(49): 51395-403.
  212. Okuda, M., et al. (2000). *Embo J* **19**(6): 1346-56.
  213. Orlicky, S. M., et al. (2001). *J Biol Chem* **276**(13): 10097-102.
  214. Orphanides, G., et al. (1996). *Genes Dev* **10**(21): 2657-83.
  215. Otwinowski, Z., and Minor, W. (1997). *Meth. Enzym.* **276**: 307-326.
  216. Pal, M. and D. S. Luse (2003). *Proc Natl Acad Sci U S A* **100**(10): 5700-5.
  217. Pan, G. and J. Greenblatt (1994). *J Biol Chem* **269**(48): 30101-4.
  218. Pardee, T. S., et al. (1998). *J Biol Chem* **273**(28): 17859-64.
  219. Patikoglou, G. A., et al. (1999). *Genes Dev* **13**(24): 3217-30.
  220. Perbal, B. (1999). *Mol Pathol* **52**(2): 84-91.
  221. Peyroche, G., et al. (2002). *Proc Natl Acad Sci U S A* **99**(23): 14670-5.
  222. Peyroche, G., et al. (2000). *Embo J* **19**(20): 5473-82.
  223. Pillai, B., et al. (2001). *J Biol Chem* **276**(33): 30641-7.
  224. Poglitsch, C. L., et al. (1999). *Cell* **98**(6): 791-8.
  225. Proudfoot, N. J., et al. (2002). *Cell* **108**(4): 501-12.
  226. Ptashne, M. and A. A. Gann (1990). *Nature* **346**(6282): 329-31.
  227. Qureshi, S. A. and S. P. Jackson (1998). *Mol Cell* **1**(3): 389-400.
  228. Ramakrishnan, V. (2002). *Cell* **108**(4): 557-72.
  229. Reed, R. and E. Hurt (2002). *Cell* **108**(4): 523-31.
  230. Reinberg, D. and R. G. Roeder (1987). *J Biol Chem* **262**(7): 3310-21.
  231. Ren, D., et al. (1999). *Mol Cell Biol* **19**(11): 7377-87.
  232. Riva, M., et al. (1987). *J Biol Chem* **262**(30): 14377-80.
  233. Roberts, S. G. and M. R. Green (1994). *Nature* **371**(6499): 717-20.
  234. Roberts, S. G. and M. R. Green (1996). *Methods Enzymol* **273**: 110-8.
  235. Roeder, R. G. (1998). *Cold Spring Harb Symp Quant Biol* **63**: 201-18.
  236. Roeder, R. G. and W. J. Rutter (1969). *Nature* **224**(216): 234-7.
  237. Roeder, R. G. and W. J. Rutter (1970). *Biochemistry* **9**(12): 2543-53.
  238. Roeder, R. G. and W. J. Rutter (1970). *Proc Natl Acad Sci U S A* **65**(3): 675-8.
  239. Rosenheck, S. and M. Choder (1998). *J Bacteriol* **180**(23): 6187-92.
  240. Ruet, A., et al. (1980). *J Biol Chem* **255**(13): 6450-5.
  241. Ruppert, S. and R. Tjian (1995). *Genes Dev* **9**(22): 2747-55.
  242. Russell, R. B., et al. (2004). *Curr Opin Struct Biol* **14**(3): 313-24.
  243. Sadhale, P. P. and N. A. Woychik (1994). *Mol Cell Biol* **14**(9): 6164-70.
  244. Sakurai, H., et al. (1999). *Mol Cell Biol* **19**(11): 7511-8.
  245. Sali, A., et al. (2003). *Nature* **422**(6928): 216-25.
  246. Samuels, M., et al. (1982). *J Biol Chem* **257**(23): 14419-27.
  247. Sassone-Corsi, P., et al. (1981). *Nucleic Acids Res* **9**(16): 3941-58.
  248. Scafe, C., et al. (1990). *Nature* **347**(6292): 491-4.
  249. Schlutzenzen, F., et al. (2000). *Cell* **102**(5): 615-23.
  250. Schmidt, M. C., et al. (1989). *Proc Natl Acad Sci U S A* **86**(20): 7785-9.
  251. Schramm, L. and N. Hernandez (2002). *Genes Dev* **16**(20): 2593-620.

- 
252. Schultz, P., et al. (2000). *Cell* **102**(5): 599-607.
253. Selleck, W., et al. (2001). *Nat Struct Biol* **8**(8): 695-700.
254. Sentenac et al., (1992). *Cold Spring Harb Laboratory Press*.
255. Shatkin, A. J. and J. L. Manley (2000). *Nat Struct Biol* **7**(10): 838-42.
256. Sheffer, A., et al. (1999). *Mol Cell Biol* **19**(4): 2672-80.
257. Shpakovski, G. V., et al. (1995). *Mol Cell Biol* **15**(9): 4702-10.
258. Shpakovski, G. V. and E. K. Shematorova (1999). *Curr Genet* **36**(4): 208-14.
259. Siaut, M., et al. (2003). *Mol Cell Biol* **23**(1): 195-205.
260. Singer, V. L., et al. (1990). *Genes Dev* **4**(4): 636-45.
261. Smid, A., et al. (1995). *J Biol Chem* **270**(22): 13534-40.
262. Stargell, L. A., et al. (2000). *J Biol Chem* **275**(17): 12374-80.
263. Svejstrup, J. Q., et al. (1994). *J Biol Chem* **269**(45): 28044-8.
264. Svejstrup, J. Q., et al. (1995). *Cell* **80**(1): 21-8.
265. Taatjes, D. J., et al. (2002). *Science* **295**(5557): 1058-62.
266. Tan, Q., et al. (2000). *Mol Cell Biol* **20**(21): 8124-33.
267. Tan, Q., et al. (2003). *Mol Cell Biol* **23**(9): 3329-38.
268. Tan, S., et al. (1996). *Nature* **381**(6578): 127-51.
269. Tantin, D. and M. Carey (1994). *J Biol Chem* **269**(26): 17397-400.
270. Teixeira, A., et al. (2004). *Nature* **432**(7016): 526-30.
271. Terwilliger, T. C. (2002). *Acta Crystallogr D Biol Crystallogr* **58**(Pt 11): 1937-40.
272. Thompson, J. D., et al. (1994). *Nucleic Acids Res* **22**(22): 4673-80.
273. Thompson, N. E., et al. (1989). *J Biol Chem* **264**(19): 11511-20.
274. Tirode, F., et al. (1999). *Mol Cell* **3**(1): 87-95.
275. Todone, F., et al. (2001). *Mol Cell* **8**(5): 1137-43.
276. Tong, A. H., et al. (2001). *Science* **294**(5550): 2364-8.
277. Tsai, F. T. and P. B. Sigler (2000). *Embo J* **19**(1): 25-36.
278. Tyree, C. M., et al. (1993). *Genes Dev* **7**(7A): 1254-65.
279. Uetz, P. and R. E. Hughes (2000). *Curr Opin Microbiol* **3**(3): 303-8.
280. Uversky, V. N., et al. (2000). *Proteins* **41**(3): 415-27.
281. Uzgiris, E. E. and R. D. Kornberg (1983). *Nature* **301**(5896): 125-9.
282. Van Dyke, M. W., et al. (1988). *Science* **241**(4871): 1335-8.
283. Vassilyev, D. G., et al. (2002). *Nature* **417**(6890): 712-9.
284. Venter, J. C., et al. (2001). *Science* **291**(5507): 1304-51.
285. Verdecia, M. A., et al. (2000). *Nat Struct Biol* **7**(8): 639-43.
286. Verrijzer, C. P. and R. Tjian (1996). *Trends Biochem Sci* **21**(9): 338-42.
287. Verrijzer, C. P., et al. (1994). *Science* **264**(5161): 933-41.
288. Wang, B. Q. and Z. F. Burton (1995). *J Biol Chem* **270**(45): 27035-44.
289. Wang, Z. and R. G. Roeder (1997). *Genes Dev* **11**(10): 1315-26.
290. Watanabe, T., et al. (2003). *Mol Cell Biol* **23**(8): 2914-26.
291. Waterston, R. H., et al. (2002). *Nature* **420**(6915): 520-62.
292. Weideman, C. A., et al. (1997). *J Mol Biol* **271**(1): 61-75.
293. Weil, P. A., et al. (1979). *Cell* **18**(2): 469-84.
294. Werner, F., et al. (2000). *Nucleic Acids Res* **28**(21): 4299-305.
295. Werner, F. and R. O. Weinzierl (2002). *Mol Cell* **10**(3): 635-46.
296. Werten, S., et al. (2002). *J Biol Chem* **277**(47): 45502-9.
297. West, A. G., et al. (2002). *Genes Dev* **16**(3): 271-88.
298. West, S., et al. (2004). *Nature* **432**(7016): 522-5.
299. Westover, K. D., et al. (2004). *Cell* **119**(4): 481-9.
300. Westover, K. D., et al. (2004). *Science* **303**(5660): 1014-6.
301. Wimberly, B. T., et al. (2000). *Nature* **407**(6802): 327-39.
302. Woychik, N. A. (1998). *Cold Spring Harb Symp Quant Biol* **63**: 311-7.

- 303. Woychik, N. A. and M. Hampsey (2002). *Cell* **108**(4): 453-63.
- 304. Woychik, N. A. and R. A. Young (1989). *Mol Cell Biol* **9**(7): 2854-9.
- 305. Wu, X., et al. (2003). *Genetics* **165**(4): 1687-702.
- 306. Xiao, H., et al. (1994). *Mol Cell Biol* **14**(11): 7507-16.
- 307. Xie, X., et al. (1996). *Nature* **380**(6572): 316-22.
- 308. Yamashita, S., et al. (1993). *Science* **261**(5120): 463-6.
- 309. Yan, Q., et al. (1999). *J Biol Chem* **274**(50): 35668-75.
- 310. Yokomori, K., et al. (1994). *Genes Dev* **8**(19): 2313-23.
- 311. Yuan, X., et al. (2002). *EMBO Rep* **3**(11): 1082-7.
- 312. Yudkovsky, N., et al. (2000). *Nature* **408**(6809): 225-9.
- 313. Zawel, L. and D. Reinberg (1993). *Prog Nucleic Acid Res Mol Biol* **44**: 67-108.
- 314. Zhang, G., et al. (1999). *Cell* **98**(6): 811-24.
- 315. Zhou, H. and K. A. Lee (2001). *Oncogene* **20**(12): 1519-24.
- 316. Zhu, W., et al. (1996). *Nat Struct Biol* **3**(2): 122-4.

# Curriculum Vitae

Name: Karim-Jean Armache

Date of birth: January 15, 1977

Place of birth: Lodz, Poland

Nationality: Polish

## Education:

2001-2005 Graduate studies in macromolecular crystallography at the Gene Center of the LMU Munich in the group of Prof. Dr. Patrick Cramer

2000 University of Lodz  
Institute of Physiology, Cytology and Cytogenetics, Department of Cytophysiology  
Visiting student

1995-2000 Technical University of Lodz,  
Faculty of Food Chemistry and Biotechnology  
Undergraduate course of Technical Biochemistry.

1991-1995 IV Liceum Ogólnokształcące im. Emilii Sczanieckiej w Łodzi  
High school

## Conferences and courses:

2005 German Society of Crystallography Meeting, invited speaker

2004 EMBO conference on Structures in Biology, Heidelberg, poster presentation

2004 6<sup>th</sup> EMBL Transcription Meeting, Heidelberg, poster presentation

2003 Transregio 5 Symposium, Chromatin, Assembly and Inheritance of Functional States, poster presentation

2003 Summer School of Crystallography, St. Andrews, Scotland, participant.

2002 EMBO Young Investigator Programme, PhD course, participant.

## Publications:

- Armache K.-J., Kettenberger H., and Cramer P. (2003)  
**Architecture of initiation-competent 12-subunit RNA polymerase II**  
*Proc. Natl. Acad. Sci USA* 100, 6964-6968
- Kettenberger H., Armache K.-J., and Cramer P. (2003)  
**Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage.**  
*Cell* 114, 347-357
- Kettenberger H\*, Armache K.-J\*, and Cramer P. (2004)  
**Complete RNA polymerase II Elongation Complex Structure and Its Interactions with NTP and TFIIS.**  
*Mol Cell* 16, 955-965  
**\*(These authors contributed equally to this work)**
- Armache K.-J., Mitterweger S., Meinhart A., and Cramer P. (2005)  
**Structures of complete RNA polymerase II and its subcomplex, Rpb4/7.**  
*J. Biol. Chem.*, 280(8):7131-4
- Armache K.-J., Kettenberger H., and Cramer P. (2005)  
**The dynamic mRNA elongation machinery**  
*Curr. Op. Struct. Biol.*; in press