

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

# **Structure and function of RNA polymerase I subunits**



Sebastian R. Geiger  
aus München

2010

## **Erklärung**

Diese Dissertation wurde im Sinne von §13 Abs. 3 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, am \_\_. Juni 2010

---

Sebastian Geiger

Dissertation eingereicht am 11. Juni 2010

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

Mündliche Prüfung am 16. Juli 2010

## Acknowledgements

First of all, I would like to thank Professor Doctor Patrick Cramer for the opportunity to perform this fascinating research in his group and within the rich scientific environment of the Gene Center. Moreover I thank Patrick for his constant and strong support and his respectful and friendly attitude. His high motivation and never-ending passion for science inspired me throughout my thesis.

I would like to thank all present and former members of the Cramer laboratory for their help, all the stimulating discussions and the enjoyable atmosphere in the lab. Thanks to Claus for his pioneering work in the RNA Polymerase I field, which paved the way for the projects of this thesis. Many thanks to you Amelie and to you Patrizia for more than eight months of excellent performance, being my students. Thanks to Stefan and Claudia for all the scientific exchange within the Pol I team, thanks to Dirk for his mentoring in X-ray crystallography, thanks to Anselm for the joined Pol I cryo-electron microscopy studies and thanks to Martin for helping with all EMSA-related questions. Thanks to Anass for his support and all the inspiring discussions we shared. Thanks to Stefan for his help with yeast fermentation and to Claudia for doing a great job as lab manager.

I would like to thank Kristina Lorenzen and Albert Heck from the University of Utrecht, The Netherlands, for their sophisticated mass spectrometry experiments, Sabine Wenzel and Paul Roesch from the University of Bayreuth for their NMR and CD measurements, as well as Robert Steinbauer and Herbert Tschochner from the University of Regensburg for their help with RNA polymerase I-specific assays. Thanks to Johannes Söding and Lucia Puchbauer from the Gene Center for their help with homology predictions. Many thanks to Jérôme Basquin and Karina Valer-Saldaña from the crystallization facility at the MPI in Martinsried for their amazing support and the huge number of crystal plates that they prepared throughout the years. Thanks to all the helpful scientists at the various synchrotrons, which I have visited for more than a dozen times, including the Swiss Light Source in Villigen, Switzerland, the European Synchrotron Radiation Facility in Grenoble, France, and the BESSY in Berlin.

Many thanks to Joachim Rädler, Marie and Marilena for their constant support and the wonderful and very creative time, which I could spend as the student representative of the International Doctorate Program NanoBioTechnology. It was a real pleasure to work together with all the motivated and very nice fellow students.

Thanks to Philip Milkereit and Daniel Wilson for their help and support as my thesis co-advisors and thanks to Dietmar Martin, Karl-Peter Hopfner, Jens Michaelis, Klaus Förstemann and Roland Beckmann for being my PhD examiners.

A lot of thanks to Andi, Tobi, Fabi, Mai and Daniel for all the nice and enjoyable time in the lab and for having a lot of fun together during our shared lunch breaks.

Many thanks to my parents for their generous and unwavering support during my whole life, and for making my education possible.

## Summary

Three multisubunit RNA polymerases (Pol) I, II and III catalyze DNA-dependent RNA synthesis in the eukaryotic nucleus. Synthesis of ribosomal RNA (rRNA) by RNA polymerase (Pol) I is the first step in ribosome biogenesis and a regulatory switch in eukaryotic cell growth. While the structure of Pol II has been studied in detail, structural information is still limited for Pol I.

This thesis describes three new crystal structures for the peripheral Pol I subcomplexes A14/43 and A49/34.5. A14/A43 could be crystallized with the use of an iterative procedure of predicting flexible regions, experimentally testing and improving these predictions and combining deletions of flexible regions in a stepwise manner. This strategy enabled the crystallization of two additional domains of A49/34.5 and could be applied to other subcomplexes with multiple flexibilities in the future, as required for structure solution of large macromolecular assemblies with hybrid methods. The structure of A14/43 allowed us, together with a cryo-electron microscopic structure of the complete 14-subunit Pol I and a homology model for the core enzyme, to obtain a Pol I hybrid structure. In this model, A14/43, the clamp, and the dock domain contribute to a unique surface, interacting with promoter-specific initiation factors.

The Pol I-specific subunits A49 and A34.5 form a heterodimer that stimulates Pol I processivity and is related to the Pol II initiation factors TFIIF and TFIIE. The N-terminal regions of A49 and A34.5 form a dimerization module with a fold that resembles the TFIIF core. The C-terminal region of A49 resembles TFIIE, forming a novel ‘tandem winged helix’ domain that binds DNA with a preference for the upstream promoter non-template strand. Similar domains are predicted in Pol III-specific subunits. Thus Pol I/III subunits that have no counterparts in the Pol II core enzyme are evolutionary related to Pol II initiation factors, and may have evolved to mediate to promoter specificity.

In contrast to Pol II, Pol I has a strong intrinsic 3'-RNA cleavage activity, which depends on the C-terminal domain of subunit A12.2 and the A49/34.5 dimerization module and, apparently, enables ribosomal RNA proofreading and 3'-end trimming.

## Contributions

Since the scientific achievements presented in this study were only possible by the collaborative work of several determined researchers, a detailed list acknowledges all major contributions.

Experiments described in chapter II ‘Crystallization of RNA polymerase I subcomplex A14/A43 by iterative prediction, probing, and removal of flexible regions’ were performed and completed by Sebastian Geiger. Claus Kuhn contributed with advice, Christoph Leidig and Jörg Renkawitz helped with protein purification.

Chapter III ‘Structure of A14/43 in functional context of RNA Polymerase I’ was based on several studies, which are listed as follows.

- RNA polymerase I preparation protocol was established by Claus Kuhn, with the help of Jochen Gerber and Herbert Tschochner from the University of Regensburg
- Cryo-electron microscopic structures were determined by Claus Kuhn, together with Sonja Baumli and Marco Gartmann, supervised by Roland Beckmann
- Crystal structure determination of subcomplex A14/43 was done by Sebastian Geiger
- Positioning of A14/43 in the Pol I cryo-electron microscopic map was done by Sebastian Geiger, supplemented by Claus Kuhn
- A49/34.5 purification protocol and TFIIF homology model was created by Sebastian Geiger
- RNA extension assays *in vitro* were performed by Claus Kuhn, supplemented by Sebastian Geiger. Pol I *in vivo* experiments were done by Claus Kuhn
- Intrinsic RNA cleavage activity of Pol I was investigated by Sebastian Geiger and Claus Kuhn, with the help of Stefan Jennebach

Experiments described in chapter IV ‘RNA polymerase I contains a TFIIF-related promoter-binding subcomplex’ were performed and completed by Sebastian Geiger, with contributions as follows. Kristina Lorenzen and Albert Heck from the University of Utrecht, The Netherlands, performed mass spectrometry, while Sabine Wenzel and Paul Roesch from the University of Bayreuth did NMR and Circular dichroism analysis. Amelie Schrieck helped with protein purification and crystallization, Patrizia Hanecker with protein purification and subcomplex delineation. Dirk Kostrewa gave advice during structure determination.

Patrick Cramer supervised all projects.

## Publications

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

1. **S.R. Geiger**, K. Lorenzen, A. Schreieck, P. Hanecker, D. Kostrewa, A. Heck and P. Cramer.  
*RNA polymerase I contains a TFIIIF-related promoter-binding subcomplex.*  
Mol Cell - accepted, (2010).
2. **S.R. Geiger**, C.-D. Kuhn, C. Leidig, J. Renkawitz and P. Cramer.  
*Crystallization of RNA polymerase I subcomplex A14/A43 by iterative prediction, probing and removal of flexible regions.*  
Acta Cryst. F64, 413–418 (2008).
3. P. Cramer, K.-J. Armache, S. Baumli, S. Benkert, F. Brueckner, C. Buchen, G.E. Damsma, S. Dengl, **S.R. Geiger**, A.J. Jasiak, A. Jawhari, S. Jennebach, T. Kamenski, H. Kettenberger, C.-D. Kuhn, E. Lehmann, K. Leike, J. Sydow and A. Vannini.  
*Structure of Eukaryotic RNA Polymerases.*  
Annu. Rev. Biophys. 37, 337-352 (2008).
4. C.-D. Kuhn, **S.R. Geiger**, S. Baumli, M. Gartmann, J. Gerber, S. Jennebach, T. Mielke, H. Tschochner, R. Beckmann and P. Cramer.  
*Functional Architecture of RNA Polymerase I.*  
Cell 131, 1260-1272 (2007).

# Table of Contents

|  |           |
|--|-----------|
| <b>Acknowledgements</b> .....  | <b>3</b>  |
| <b>Summary</b> .....   | <b>5</b>  |
| <b>Contributions</b> .....   | <b>6</b>  |
| <b>Publications</b> .....  | <b>7</b>  |
| <b>Chapter I: General introduction</b> .....   | <b>10</b> |
| <b>1 Introduction</b> .....  | <b>11</b> |
| 1.1 DNA-dependent RNA polymerases .....  | 11        |
| 1.2 Ribosome biogenesis .....  | 12        |
| 1.3 Structure of the nucleolus and ribosomal DNA loci .....  | 12        |
| 1.4 Ribosomal DNA promoter .....   | 14        |
| 1.5 RNA Polymerase I pre-initiation complex .....  | 15        |
| 1.6 RNA polymerase I-specific subunits.....  | 16        |
| 1.7 Structural studies on eukaryotic RNA polymerases.....  | 17        |
| <b>2 Scope of this study</b> .....   | <b>18</b> |
| <b>Chapter II: Crystallization of RNA polymerase I subcomplex A14/A43<br/>by iterative prediction, probing and removal of flexible regions</b> ..... | <b>19</b> |
| <b>1 Introduction</b> .....  | <b>20</b> |
| <b>2 Results and Discussion</b> .....  | <b>23</b> |
| 2.1 Prediction of unstructured regions.....  | 23        |
| 2.2 Identification of unstructured regions in A43 .....  | 24        |
| 2.3 Identification of unstructured regions in A14 .....  | 26        |
| 2.4 Additional changes to obtain diffraction-quality crystals .....  | 27        |
| 2.5 Conclusions .....  | 28        |
| <b>3 Experimental procedures</b> .....   | <b>30</b> |
| 3.1 Bioinformatic analysis.....  | 30        |
| 3.2 Cloning, expression, purification and limited proteolysis of protein variants .....  | 30        |
| 3.3 Crystallization and crystal screening.....   | 31        |
| <b>Chapter III: Structure of A14/43 in functional context of RNA polymerase I</b> .....  | <b>32</b> |
| <b>1 Introduction</b> .....  | <b>33</b> |
| <b>2 Results and Discussion</b> .....  | <b>34</b> |
| 2.1 Cryo-EM structure of Pol I at 12 Å resolution .....  | 34        |
| 2.2 Homology model of the Pol I core .....   | 34        |
| 2.3 Crystal structure of the A14/43 subcomplex .....   | 35        |
| 2.4 Interaction of A14/43 with the Pol I core .....  | 40        |
| 2.5 Specific interactions with initiation factors.....   | 42        |
| 2.6 A49 and A34.5 form a TFIIF-like heterodimer near the funnel.....   | 43        |
| 2.7 A49/A34.5 is a built-in processivity factor .....  | 46        |
| 2.8 Pol I has intrinsic RNA cleavage activity.....   | 48        |
| 2.9 Pol I cleavage activity requires A12.2.....  | 50        |
| 2.10 Possible functions of the cleavage activity.....  | 50        |
| 2.11 Conclusions .....   | 51        |

|   |            |
|---|------------|
| <b>3 Experimental procedures.....</b>   | <b>52</b>  |
| 3.1 Pol I preparation .....   | 52         |
| 3.2 Cryo-EM structure determinations .....  | 52         |
| 3.3 Modeling of the Pol I core.....   | 52         |
| 3.4 A14/43 preparation.....   | 53         |
| 3.5 A14/43 crystal structure determination .....  | 53         |
| 3.6 Preparation of A49/34.5 .....   | 54         |
| 3.7 A49/34.5 structure prediction .....   | 54         |
| 3.8 Yeast genetic manipulations and 6-azauracil phenotyping .....                       | 55         |
| 3.9 Preparation of Pol I variants .....   | 55         |
| 3.10 RNA extension assays .....   | 55         |
| 3.11 RNA cleavage assays .....  | 56         |
| 3.12 Figure preparation .....   | 56         |
| 3.13 Accession numbers.....   | 56         |
| <b>Chapter IV: RNA polymerase I contains a TFIIF-related DNA-binding subcomplex....</b> | <b>57</b>  |
| <b>1 Introduction .....</b>   | <b>58</b>  |
| <b>2 Results.....</b>   | <b>59</b>  |
| 2.1 Pol I subcomplex A49/34.5 forms a heterodimer with two distinct domains .....       | 59         |
| 2.2 The A49/34.5 dimerization module resembles TFIIF .....                              | 61         |
| 2.3 A49 contains a novel tandem winged helix domain.....                                | 65         |
| 2.4 The tWH domain binds double-stranded DNA with a conserved surface .....             | 68         |
| 2.5 The tWH domain binds single-stranded promoter DNA.....                              | 72         |
| 2.6 A49/34.5 domains have distinct functions .....                                      | 74         |
| <b>3 Discussion .....</b>   | <b>79</b>  |
| 3.1 Structure and DNA-binding function of Pol I subcomplex A49/34.5.....                | 79         |
| 3.2 Pol I resembles a minimal Pol II-TFIIF-TFIIIE complex.....                          | 79         |
| 3.3 Pol III contains TFIIF- and TFIIIE-like regions .....                               | 80         |
| 3.4 Polymerase evolution and promoter specificity .....                                 | 81         |
| <b>4 Experimental procedures.....</b>   | <b>82</b>  |
| 4.1 Mass spectrometry.....  | 82         |
| 4.2 Recombinant proteins.....   | 82         |
| 4.3 Crystal structure determinations.....   | 83         |
| 4.4 RNA cleavage and extension assays .....   | 84         |
| 4.5 Electrophoretic mobility shift assay .....  | 84         |
| 4.6 Protein interaction assay.....  | 86         |
| <b>Final conclusions and outlook .....</b>  | <b>87</b>  |
| <b>Appendix (unpublished results) .....</b>   | <b>91</b>  |
| 1 Functional analysis of A49 tWH variants .....   | 91         |
| 2 Interaction studies of A49/34.5 domains.....  | 92         |
| 3 Additional Pol I cryo-EM studies.....   | 92         |
| 4 Pol I crosslinking experiments .....  | 93         |
| 5 Crystal structure of ClpP from <i>Staphylococcus aureus</i> .....                     | 94         |
| <b>Abbreviations.....</b>   | <b>95</b>  |
| <b>References .....</b>   | <b>98</b>  |
| <b>Curriculum Vitae .....</b>   | <b>110</b> |

## **Chapter I:**

# **General introduction**

# 1 Introduction

## 1.1 DNA-dependent RNA polymerases

Transcription of genetic information from DNA into RNA is performed by specific multisubunit enzymes, DNA-dependent RNA polymerases (Pol). While in bacteria and archaea all cellular RNA is synthesized by a single RNA polymerase, the eukaryotic genome is transcribed by three different Pols, as illustrated in Table 1.

**Table 1. RNA polymerase subunits**

| Polymerase part     | Pol I subunit                 | MW (kDa) | Corresponding Pol II subunit    | Corresponding Pol III subunit | Subunit type       |
|---------------------|-------------------------------|----------|---------------------------------|-------------------------------|--------------------|
| Core                | A190                          | 186.4    | Rpb1                            | C160                          | homolog            |
|                     | A135                          | 135.7    | Rpb2                            | C128                          | homolog            |
|                     | AC40                          | 37.7     | Rpb3                            | AC40                          | homolog            |
|                     | AC19                          | 16.2     | Rpb11                           | AC19                          | homolog            |
|                     | A12.2                         | 13.7     | Rpb9                            | C11                           | homolog            |
|                     | Rpb5 (ABC27)                  | 25.1     | Rpb5                            | Rpb5                          | common             |
|                     | Rpb6 (ABC23)                  | 17.9     | Rpb6                            | Rpb6                          | common             |
|                     | Rpb8 (ABC14.5)                | 16.5     | Rpb8                            | Rpb8                          | common             |
|                     | Rpb10 (ABC10 $\beta$ )        | 8.3      | Rpb10                           | Rpb10                         | common             |
|                     | Rpb12 (ABC10 $\alpha$ )       | 7.7      | Rpb12                           | Rpb12                         | common             |
| Subcomplex A14/43   | A14                           | 14.6     | Rpb4                            | C17                           | counterpart        |
|                     | A43                           | 36.2     | Rpb7                            | C25                           | counterpart        |
| Subcomplex A49/34.5 | A34.5 <sup>1</sup>            | 26.9     | (RAP30) <sup>1</sup>            | C53 <sup>1</sup>              | Pol I/III specific |
|                     | A49 (N-terminal) <sup>2</sup> | 46.7     | (RAP74) <sup>2</sup>            | C37 <sup>2</sup>              | Pol I/III specific |
|                     | A49 (C-terminal) <sup>3</sup> |          | (TFIIE- $\beta$ ) <sup>3</sup>  | C34 <sup>3</sup>              | Pol I/III specific |
| -                   | -                             | -        | (TFIIE- $\alpha$ ) <sup>4</sup> | C82 <sup>4</sup>              | Pol III specific   |
| -                   | -                             | -        | -                               | C31                           | Pol III specific   |
| Total               | 14 subunits                   | 589.6    | 12 subunits                     | 17 subunits                   | -                  |

<sup>1</sup>The Pol I subunit A34.5 is homologous to the Pol II initiation factor TFIIF- $\beta$  (Rap30) and to the Pol III-specific subunit C53. For details see chapters III and IV.

<sup>2</sup>The N-terminal domain of Pol I subunit A49 is homologous to the Pol II initiation factor TFIIF- $\alpha$  (Rap74) and to the Pol III-specific subunit C37. For details see chapters III and IV.

<sup>3</sup>The C-terminal domain of Pol I subunit A49 is homologous to the Pol II initiation factor TFIIE- $\beta$  and to the Pol III-specific subunit C34. For details see chapter IV.

<sup>4</sup>The Pol III-specific subunit C82 is homologous to the Pol II initiation factor TFIIE- $\alpha$ . For details see chapter IV.

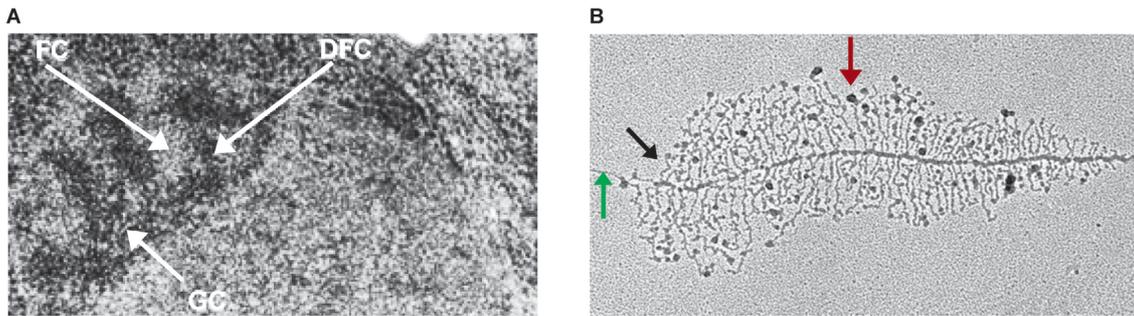
RNA polymerase III (Pol III) transcribes various non-coding RNA molecules, including transfer RNA (tRNA), the 5S ribosomal RNA (rRNA), the splicosomal U6 snRNA, the signal recognition particle 7SL RNA and short regulatory RNAs (Paule and White, 2000). RNA polymerase II (Pol II) transcribes protein-coding genes (Cramer, 2004), as well as many small regulatory RNAs (Dye et al., 2006). RNA polymerase I (Pol I) exclusively synthesizes ribosomal RNA (rRNA). In yeast, the rRNA is transcribed as a 35S primary transcript, which is modified and processed cotranscriptionally into the 25S, 5.8S and 18S rRNA (Kos and Tollervey, 2010), constituting the major building blocks of ribosomes.

## **1.2 Ribosome biogenesis**

The eukaryotic 80S eukaryotic ribosome is composed of approximately 65% RNA and 35% protein with a molecular mass of about 4 MDa. The large 60S ribosomal subunit consists of the 25S, 5.8S and 5S rRNA, while the small 40S subunit comprises the 18S rRNA. Besides the rRNA, around 80 additional ribosomal proteins are incorporated to complete the functional ribosome. Ribosome biogenesis is a very complex process, with several intermediate steps in rRNA maturation that involve a large number of trans-acting factors (Henras et al., 2008). Since each cell provides around 10 ribosomes for every synthesized mRNA (Warner, 1999), ribosome synthesis occupies a substantial fraction of the cellular resources. Therefore, Pol I transcription accounts for up to 60% of all nuclear transcription, which results in up to 80% of the total RNA in a cell (Warner, 1999). As a consequence, rRNA synthesis is a focal point for the regulation of cell metabolism and cell growth (Grummt, 2003). Consistent with the effect of tumor suppressors and oncogenes on ribosome biogenesis (Fontoura et al., 1997; Moss et al., 2007), Pol I and Pol III influence cancer development (White, 2005).

## **1.3 Structure of the nucleolus and ribosomal DNA loci**

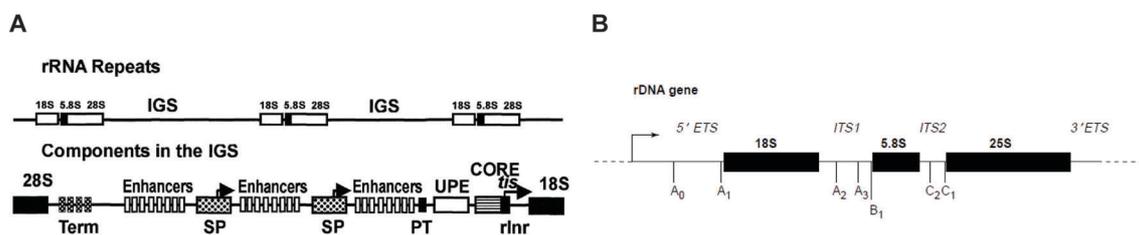
rRNAs are synthesized, processed and assembled into ribosomes within the largest sub-nuclear compartment, the nucleolus (Fig. 1A), which is formed when rRNA genes are transcribed (Nomura, 2001). Initial assembly of ribosomes occur cotranscriptionally, which can be visualized in ‘Miller spread’ electron micrographs (Fig. 1B)



**Figure 1. Structure of the nucleolus and active ribosomal DNA genes**

(A) Electron microscopy of yeast nuclei (Trumtel et al., 2000). Depicted nucleolar elements are a fibrillar centre (FC), a granular center (GC) and a dense fibrillar center (DFC). (B) ‘Miller spread’ of a single rDNA repeat in yeast. The rDNA (green arrow), the transcribed rRNAs (black arrow) and pre-ribosomes (brown arrow) are indicated. Courtesy of Sarah French and Ann Beyer, University of Virginia.

In yeast, the ribosomal genes are organized in a tandem array, which consists of approximately 150 identical repeats of the rDNA gene. Besides the rDNA coding regions, each gene contains an intergenic spacer (IGS), important for efficient pre-rRNA synthesis, as well as rDNA silencing (Moss et al., 2007). The IGS comprises important sequence elements such as the rDNA promoter, multiple enhancers, the spacer promoter and several terminator elements (Fig. 2A). The rDNA coding region consists of an external transcribed spacer (ETS) and internal transcribed spacer (ITS) sequences, which are removed during pre-rRNA processing (Fig. 2B). It is interesting that neither Pol I nor tandemly arranged rDNA repeats are absolutely required for rRNA synthesis. In a yeast strain, which is defective for Pol I transcription, rRNA can be synthesized by Pol II from a high-copy plasmid with the 35S rDNA under control of the *GAL7* promoter (Nogi et al., 1991).

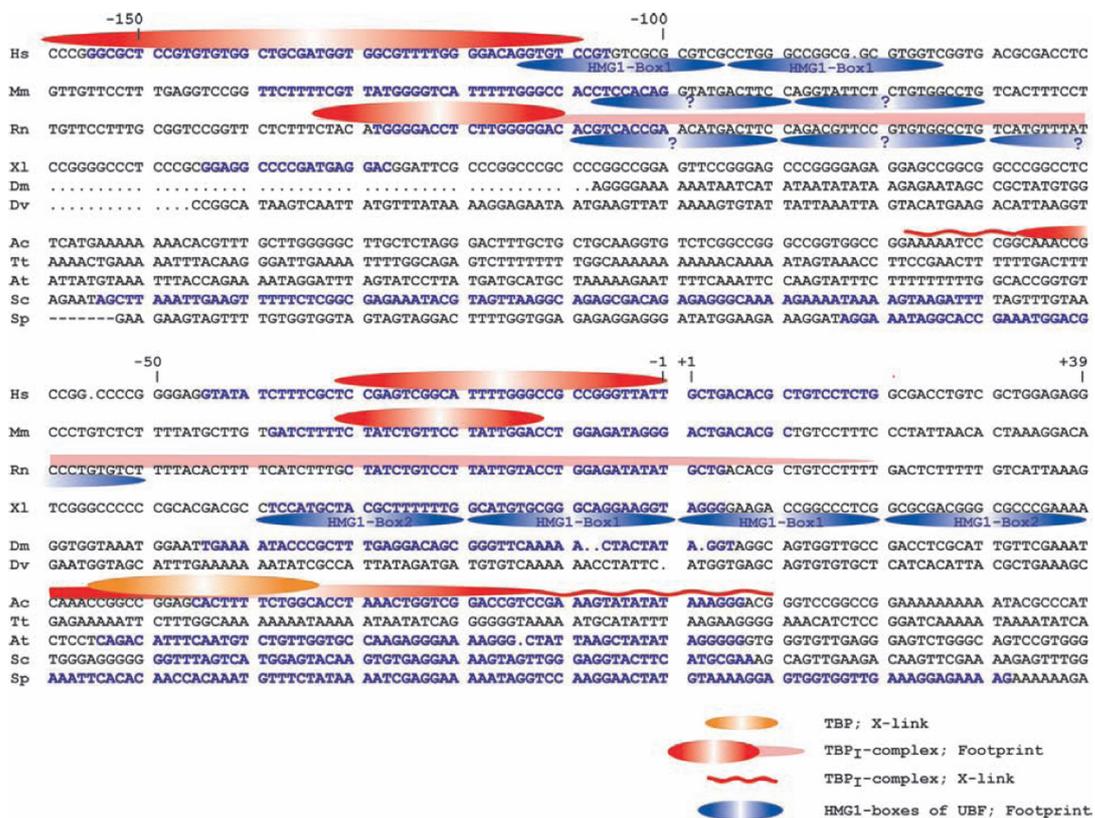


**Figure 2. Yeast ribosomal DNA locus**

(A) rRNA coding units are separated by intergenic spacers (IGS), containing terminators (Term), enhancers, spacer promoter (SP), proximal terminator (PT), the upstream promoter element (UPE/UE) and the promoter core element (CORE/CE), with the ribosomal initiator (rInr). Initiation sites are indicated by arrows or *tis* (Paule and White, 2000). (B) Single rDNA gene (Granneman and Baserga, 2005). External and internal transcribed spacers (ETS, ITS), 18S, 5.8S and 25S rRNA coding sequences, and specific cleavage sites are shown.

## 1.4 Ribosomal DNA promoter

In yeasts and vertebrates, the rDNA promoter consists of 140 to 160 base pairs, located mostly upstream of the mapped pre-rRNA initiation site (Fig. 3). The poor conservation of various Pol I promoter sequences is consistent with the extreme species specificity of the Pol I transcription system, but the general architecture of promoter elements is highly conserved from yeast to human. Most rDNA promoters contain two distinct sequence elements, a core promoter element (CORE/CE) and an upstream promoter element (UPE/UE). A more detailed analysis revealed three distinct domains within the promoter, CE domain 1 (positions -28 to +8), UE domain 2 (-70 to -51) and UE domain 3 (-146 to -76) (Choe et al., 1992; Kulkens et al., 1991; Musters et al., 1989).



**Figure 3. RNA polymerase I promoter sequences**

Promoter sequences are shown according to the mapped initiation site (+1). Functional sequence elements are labeled in blue and regions shown to interact with the TBP<sub>I</sub> complex (SL-1/TIF-IB), UBF or with TBP are indicated graphically (see below). Species are *Homo sapiens* (Hs), *Mus musculus* (Mm), *Rattus norvegicus* (Rn), *Xenopus laevis* (Xl), *Drosophila melanogaster* (Dm), *Drosophila virilis* (Dv), *Acanthamoeba castellanii* (Ac), *Thermus thermophilus* (Tt), *Arabidopsis thaliana* (At), *Saccharomyces cerevisiae* (Sc) and *Schizosaccharomyces pombe* (Sp). Adapted from (Moss et al., 2007).

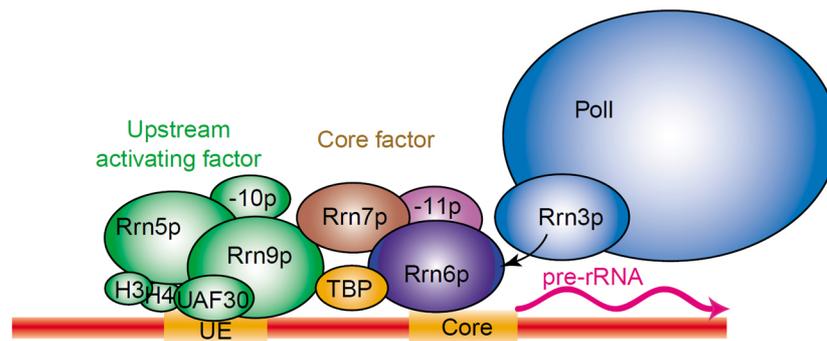
## 1.5 RNA Polymerase I pre-initiation complex

Formation of the Pol I pre-initiation complex (PIC) requires the TATA-box-binding protein (TBP) and a group of Pol I-specific TBP-associated factors (TAF), to recognize the promoter (Table 2). In human and mouse, the PIC includes transcription initiation factor IB (TIF-IB/SL1), which consists of TBP and three TAFs, as well as a second non-specific DNA-binding protein, the upstream binding factor (UBF). In yeast, the PIC requires the core factor (CF) and the upstream activating factor (UAF), which could be related to their mammalian counterparts TIF-IB and UBF, respectively (Boukhgalter et al., 2002) and bind to the CE and UE of the Pol I promoter (Fig. 4).

**Table 2. Subunits from yeast CF (corresponding to mammalian TIF-IB/SL1) and UAF (Moss et al., 2007).**

| Yeast Core Factor (CF) | (=>) Mammalian TIF-IB (SL1) | Yeast upstream activating factor (UAF) |
|------------------------|-----------------------------|--|
| Rrn6                   | TAF <sub>I</sub> 110/95     | Rrn5                                   |
| Rrn7                   | TAF <sub>I</sub> 63/68      | Rrn9                                   |
| Rrn11                  | TAF <sub>I</sub> 48         | Rrn10                                  |
| TBP                    | TBP                         | UAF30                                  |
|                        |                             | H3                                     |
|                        |                             | H4                                     |

After the UAF-UE complex is formed, TBP is either already present or is recruited along with the CF (Steffan et al., 1998). The UAF complex is needed for efficient rDNA transcription, although low-level specific transcription initiation *in vitro* does not require the UE of the promoter, UAF or TBP (Keener et al., 1998). Initiation in yeast and mammals requires the additional Pol I-associated factor Rrn3 (TIF-IA), which is essential for functional recruitment of Pol I to the PIC (Keener et al., 1998; Milkereit and Tschochner, 1998). Rrn3 is normally found associated with a small fraction of phosphorylated Pol I (Fath et al., 2001) and is released after the polymerase initiates transcription (Bier et al., 2004). Rrn3 itself is also phosphorylated and its phosphorylation state is regulated by the mTOR kinase nutrient-sensing pathway and the Jun N-terminal kinase (Mayer et al., 2005).



**Figure 4. RNA polymerase I pre-initiation complex**

Assembly of the Pol I pre-initiation complex in yeast. DNA is shown in red and promoter elements in yellow, UAF is in green, Pol I and Rrn3 are in blue, TBP in yellow and CF in brown, purple and magenta (Moss, 2004).

## 1.6 RNA polymerase I-specific subunits

Pol II utilizes a set of additional initiation factors TFIIA, -B, -D, -E, -F, and -H, which are important for PIC formation (Table 3) and are apparently not related to Pol III and Pol I transcription factors. All three eukaryotic RNA polymerases consist of a conserved 10-subunit core, plus an additional number of peripheral subunits (Table 1). Pol III contains five extra subunits, which are organized in two subcomplexes C37/C53 (Landrieux et al., 2006) and C82/34/31 (Wang and Roeder, 1997). In Pol I there are four additional subunits, the Pol II-related A14/43 heterodimer (Meka et al., 2003) and the Pol I-specific subunits A34.5 and A49. Subunit A43 forms an essential bridge to the conserved Pol I initiation factor Rrn3 (Milkereit and Tschochner, 1998; Peyroche et al., 2000), thereby recruiting Pol I to the rDNA promoter. A34.5 stabilizes A49 on Pol I and was shown to genetically interact with DNA topoisomerase 1 (Gadal et al., 1997). A49 is important for Pol I activity (Huet et al., 1975; Liljelund et al., 1992), promoter-dependent transcription (Hanada et al., 1996), as well as for recruitment of Pol I and Rrn3 to the rDNA promoter (Beckouet et al., 2008).

**Table 3. Pol II initiation factors in yeast (Hahn, 2004).**

| Factor        | Function   |
|---------------|--|
| TFIIA         | Stabilizes TBP and TFIID-DNA binding. Blocks transcription inhibitors. Positive and negative gene regulation.  |
| TFIIB         | Binds TBP, Pol II, and promoter DNA. Helps to fix transcription start site.  |
| TFIID/<br>TBP | Binds TATA element and deforms promoter DNA. Assembly platform for TFIIB, TFIIA and TAFs.  |
| TFIIE         | Binds promoter near transcription start. Stabilization of transcription bubble in the Open Complex.  |
| TFIIF         | Binds Pol II and is involved in Pol II recruitment to PIC and in Open Complex formation  |
| TFIIH         | Functions in transcription and DNA repair. Kinase and two helicase activities. Essential for Open Complex formation. Mutations in IIH can cause human disease (e.g. Xeroderma pigmentosum) |

## 1.7 Structural studies on eukaryotic RNA polymerases

Most structural advances were achieved with Pol II, namely the crystal structures of the 10-subunit core enzyme (Cramer et al., 2001), the complete 12-subunit enzyme (Armache et al., 2005), the Pol II-TFIIS complex (Kettenberger et al., 2003) and the Pol II-TFIIB initiation complex (Kostrewa et al., 2009). Since all three polymerases share a similar core structure and active center (Cramer et al., 2008), their basic mechanism of DNA-dependent RNA elongation is similar and well understood at a structural level (Brueckner et al., 2009).

The Pol III architecture was determined by cryo-electron microscopy (EM) (Fernandez-Tornero et al., 2007) (Vannini and Cramer, unpublished), a homology model for the core enzyme was constructed (Jasiak et al., 2006), and the structure of the subcomplex C17/25 was solved by X-ray crystallography (Jasiak et al., 2006).

For Pol I, a three-dimensional surface model and approximate dimensions were first revealed by EM analysis of two-dimensional crystals (Schultz et al., 1993). A subsequent cryo-EM map at 34 Å resolution visualized a stalk containing the Pol I subcomplex A14/43 and densities for the Pol I-specific subunits A49 and A34.5 over the central cleft (Bischler et al., 2002; Peyroche et al., 2002). Later EM analysis with negative staining at 22 Å confirmed the stalk density, but not the location of A49 and A34.5 (De Carlo et al., 2003).

## **2 Scope of this study**

In contrast to the well-studied Pol II system, which includes atomic structures of the complete 12-subunit enzyme in complex with nucleic acid scaffolds and transcription factors, detailed structural information for Pol I is still limited.

Therefore the overall aim of this work was to obtain crystal structures for all four Pol I-specific subunits A14, A43, A34.5 and A49 and to complete the structural information with biochemical data. Such information would enable us to better understand evolutionary differences between the three eukaryotic RNA polymerases.

A combined approach was pursued, which first included the development of a protein engineering strategy to obtain diffracting crystals, suitable for accelerated X-ray structure determination, as described in chapter II. This strategy was applied to obtain the atomic structure of the A14/43 subcomplex and, together with a Pol I cryo-EM map, to construct a hybrid structure for the 12-subunit enzyme. The structural data was combined with biochemical analysis of the RNA cleavage activity and processivity of the complete enzyme, to obtain a model of the functional architecture of Pol I, as summarized in chapter III. Finally, two crystal structures of subunits A49 and A34.5 were solved, completing the structural analysis of all peripheral Pol I domains and resulting in a more detailed understanding of novel functional properties of the subcomplex, such as DNA binding and a possible role in Pol I promoter recognition, as described in chapter IV.

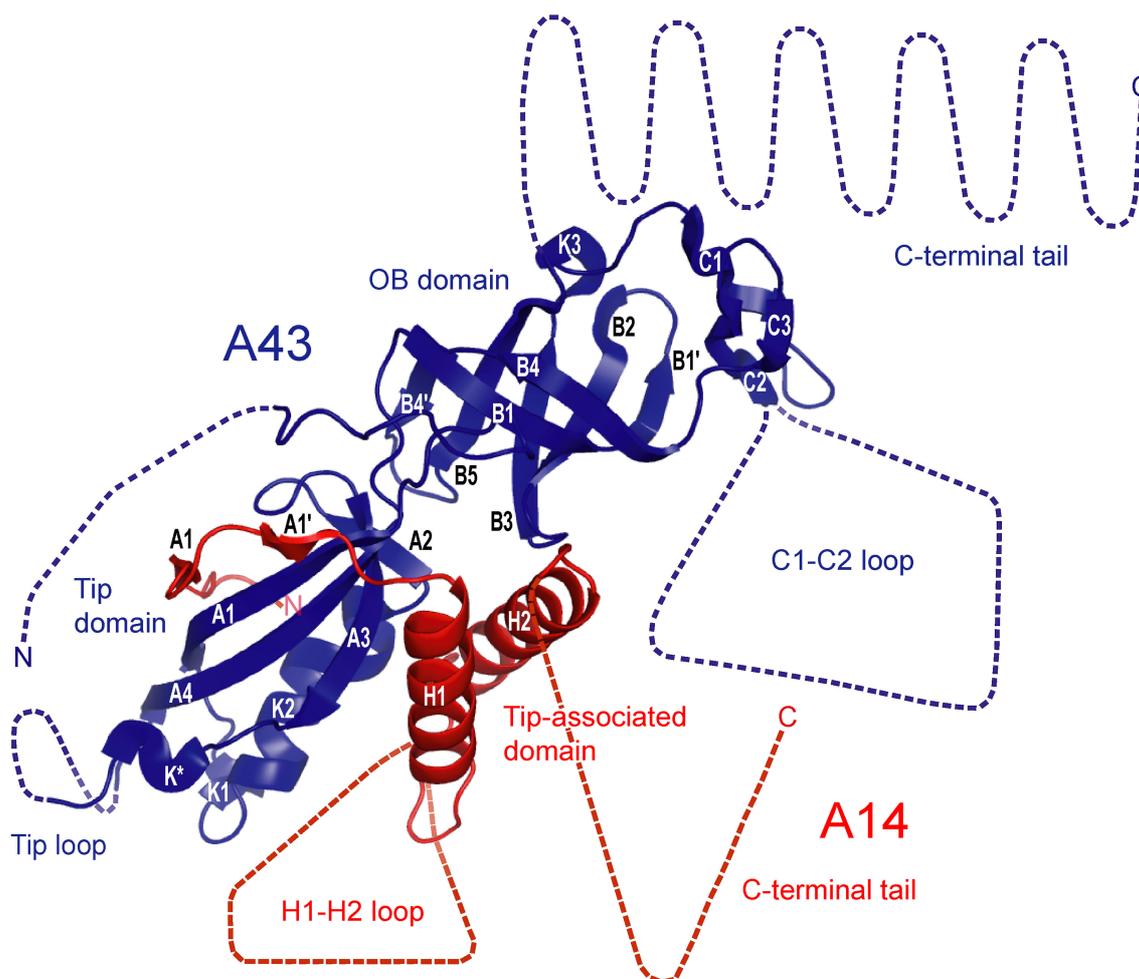
## **Chapter II:**

# **Crystallization of RNA polymerase I subcomplex A14/A43 by iterative prediction, probing and removal of flexible regions**

## 1 Introduction

Structure determination of large multi-component complexes often requires hybrid approaches that combine electron microscopy (EM) with X-ray crystallography. Whereas EM can establish the structure of the complex at medium resolution, the subunit architecture may be resolved by interpreting the EM density using high-resolution structures of subcomplexes obtained by X-ray crystallography. However, such subcomplexes often fail to crystallize because they contain flexible regions, some of which result from removing the subcomplex from the intact native complex. In chapter III, we report the hybrid structure of yeast RNA polymerase (Pol) I, a 14-subunit 600 kDa complex. Hybrid analysis of the complex relied on the crystal structure of the heterodimeric polymerase subcomplex A14/A43, which contained several extended flexible regions. Here, we show how iterative cycles of structure prediction and experiments were used to identify and remove flexible protein regions, which led to the crystallization of the A14/A43 subcomplex. This approach was additionally employed to crystallize two domains of Pol I-specific subcomplex A49/34.5 (chapter IV) and could be applied to comparable crystallization challenges in the future.

The A14/A43 crystal structure, described in chapter III (2.3), revealed an overall similarity to its counterparts Rpb4/7 in Pol II, C17/25 in Pol III, and RpoF/E in the archaeal Pol (Hu et al., 2002; Kuhn et al., 2007; Meka et al., 2003; Peyroche et al., 2002; Sadhale and Woychik, 1994; Shpakovskii and Shematorova, 1999; Siaux et al., 2003; Todone et al., 2001) (Fig. 5 and Table 4). A14 contains the ‘tip-associated domain’, which includes a flexible loop (H1-H2), but lacks the C-terminal helicase and RNAse D C-terminal (HRDC) domain present in all A14 counterparts, instead having a long flexible C-terminal tail. A43 consists of an N-terminal ‘tip domain’ and a C-terminal OB domain that is present in all its counterparts. The overall structural similarity of A14/A43 with its counterparts is surprising, as there is no sequence similarity between A14 and its counterparts and only 8% of the A43 residues are identical in Rpb7 (Table 4). Nevertheless, 78% of the Rpb7 residues have the same fold as in A43 (Table 4 and chapter III, 2.3). A43 differs from Rpb7 by having flexible N- and C-terminal tails, an extended loop (C1-C2) within the C-terminal OB domain and a ten-residue insertion in the ‘tip loop’ (chapter III, 2.4). Thus, A14/A43 contains at least four very extended flexible regions (Fig. 5), which had to be identified and at least partially removed for crystallization.



**Figure 5. Crystal structure of yeast A14/A43**

A43 and A14 are colored blue and red, respectively. Disordered regions and regions which are not present in the protein variant are drawn to scale, assuming extended folds, as dashed lines. Adapted from chapter III (Fig. 13).

Here we present our approach to obtain diffraction-quality crystals of the A14/A43 subcomplex, and an evaluation of this approach in light of the structure that is now available. We show how repetitive cycles of predictions and experiments were used to obtain a minimal A14/A43 variant that lacks most of the flexible residues and forms well-ordered crystals. The described approach is superior to standard sequence analysis and structure prediction. For example, there were conflicting predictions for the A14 structure, one suggesting that a HRDC domain is present (Meka et al., 2003), and another that it is absent (Peyroche et al., 2002). Our approach revealed the absence of the HRDC domain, which was key to obtaining crystals. The presented strategy may be adapted to comparable crystallogenesis challenges that will be faced more frequently in the future.

**Table 4. Similarity of A14 and A43 to homologous and orthologous polymerase subunits**

| Pol I subunit | Homologous subunit         | Sequence identity <sup>1</sup><br>(%) | Conserved fold <sup>2</sup><br>(%) | Sequence similarity <sup>3</sup><br>(%) |
|---------------|----------------------------|---------------------------------------|------------------------------------|---|
| A14           | Rpb4 (Pol II)              | 4.5                                   | 25.0                               |   |
|               | C17 (Pol III)              | 6.6                                   | 52.9                               |   |
|               | RpoF (archaeal Pol)        | 3.7                                   | 76.6                               |   |
|               | A14 ( <i>C. glabrata</i> ) | 25.6                                  |                                    | 31.4                                    |
|               | A14 ( <i>A. gossypii</i> ) | 36.5                                  |                                    | 52.6                                    |
|               | A14 ( <i>K. lactis</i> )   | 42.3                                  |                                    | 55.5                                    |
| A43           | Rpb7 (Pol II)              | 8.0                                   | 78.4                               |   |
|               | C25 (Pol III)              | 8.6                                   | 64.2                               |   |
|               | RpoE (archaeal Pol)        | 6.1                                   | 69.5                               |   |
|               | A43 ( <i>C. glabrata</i> ) | 62.9                                  |                                    | 73.6                                    |
|               | A43 ( <i>A. gossypii</i> ) | 55.5                                  |                                    | 69.0                                    |
|               | A43 ( <i>K. lactis</i> )   | 56.4                                  |                                    | 69.0                                    |

<sup>1</sup>Number of amino acid residues in A14 or A43 that are identical in the homologous polymerase subunit divided by the total number of residues in A14 or A43. For orthologues, the number of amino acid residues in *S. cerevisiae* A14 or A43 that are identical in the orthologous polymerase subunit divided by the total number of residues in *S. cerevisiae* A14 or A43.

<sup>2</sup>Number of amino acid residues in the homologous polymerase subunit that have the same fold in the A14/A43 structure divided by the total number of residues in the homologous polymerase subunit. For A14, the number of amino acid residues in the homologous polymerase subunit that have the same fold in the A14 structure divided by the number of residues of the tip-associated domain of the homologous polymerase subunit (HRDC domain excluded).

<sup>3</sup>The number of amino acid residues in *S. cerevisiae* A14 or A43 that are similar (e.g. hydrophobic, acidic or basic) in the orthologous polymerase subunit divided by the total number of residues in *S. cerevisiae* A14 or A43.

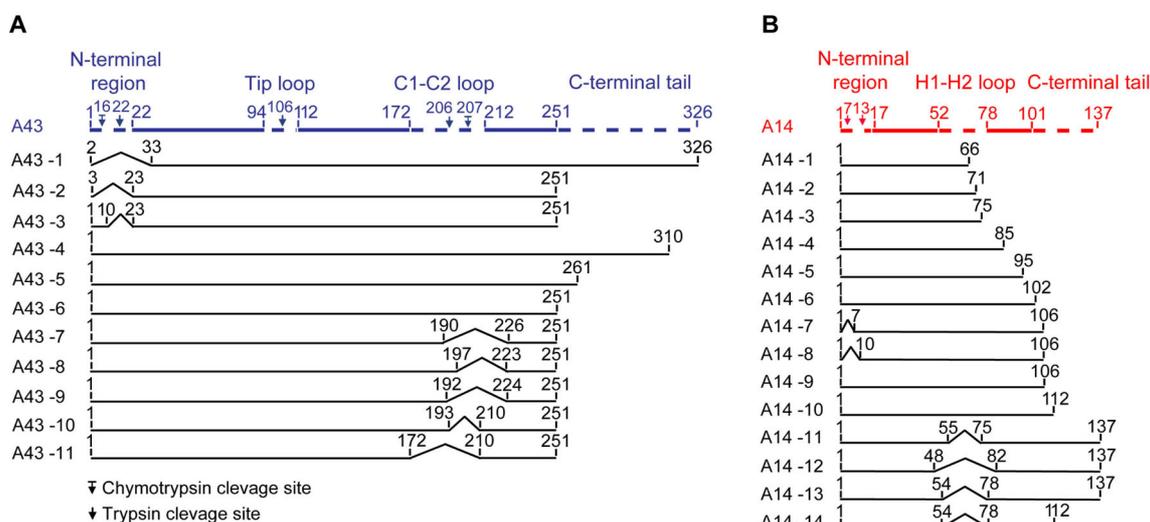
## 2 Results and Discussion

### 2.1 Prediction of unstructured regions

Recombinantly expressed full-length A14/A43 (Experimental procedures, 3.2) did not crystallize, indicating the presence of unstructured flexible regions within the complex. To aid in the design of protein variants that were suitable for crystallization, we used bioinformatics tools to predict the structured and unstructured regions in both subunits. Firstly, we predicted the secondary-structure elements in both subunit sequences using PredictProtein (Rost et al., 2004). Secondly, we prepared a sequence alignment with closely related yeast orthologues detected by BLAST (Altschul et al., 1997). This alignment was used to judge the reliability of the secondary structure predictions, assuming that structured elements must be conserved among closely related species. Thirdly, a preliminary alignment of A14 and A43 sequences was obtained by matching the predicted secondary structure elements with the precisely located secondary structure elements present in the crystal structures of the A14/A43 counterparts in other RNA polymerases. Since sequence homology was extremely weak (A43) or essentially absent (A14), the obtained alignments obviously contained errors, but they still allowed an initial prediction of the location of structured and unstructured regions. The final alignment of the secondary-structure elements observed in the A14/43 crystal structure with predicted secondary-structure elements is shown in Figure 8. The preliminary alignment between A43 and its counterparts in the Rpb7 family suggested that A43 contains a long flexible C-terminal tail and a disordered internal loop. However, the precise length of the tail and the exact position of the loop were unclear owing to a lack of conservation and possible misalignment in the C-terminal region. For A14, two generally possible alignments were obtained. The first comprised a C-terminal HRDC domain, consistent with one published model (Meka et al., 2003), yet not all of the helices required to form this domain were predicted. The second model postulated that the HRDC domain was absent and A14 contained extended flexible regions, consistent with another model (Peyroche et al., 2002).

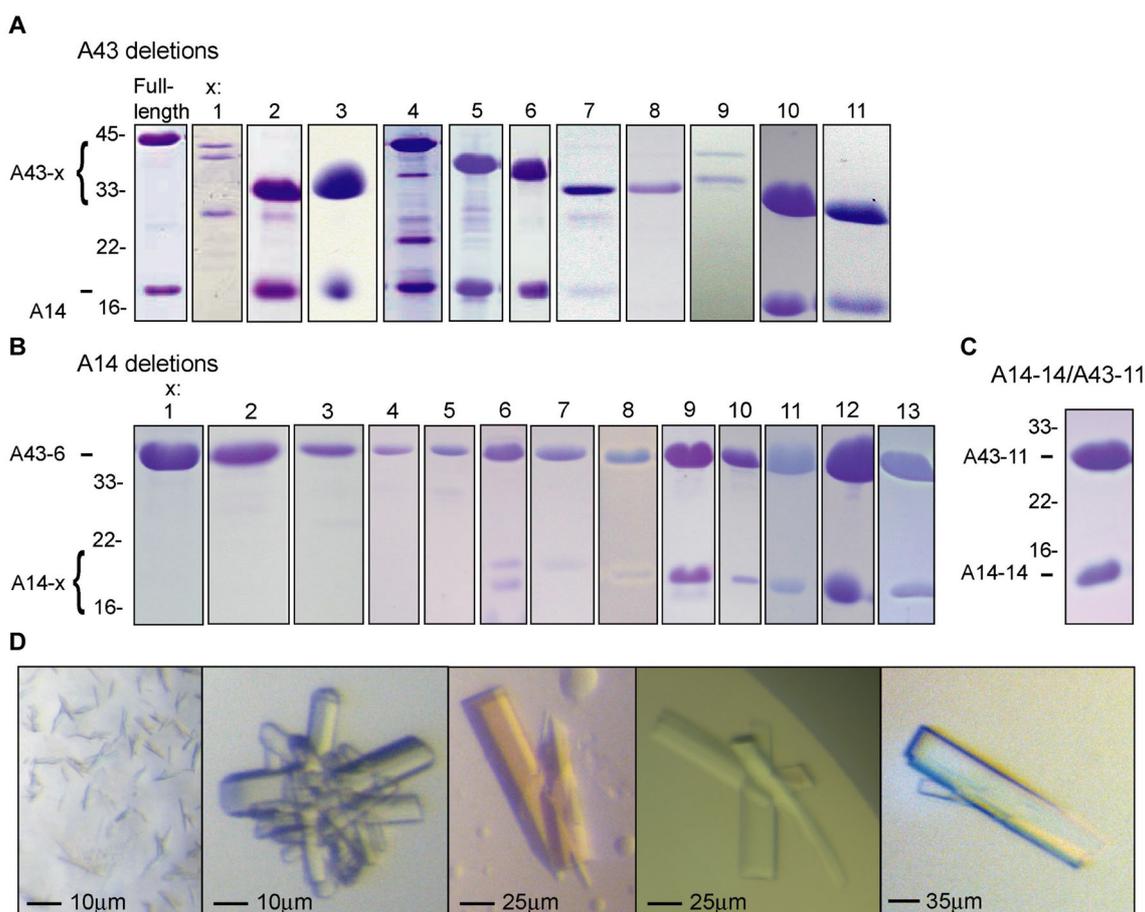
## 2.2 Identification of unstructured regions in A43

We tested the predictions of flexible regions in the A14/A43 heterodimer by preparing deletion variants of A14 and A43, examining their interaction after coexpression in *E. coli*, and assessing the solubility of the obtained subcomplexes. We first investigated the nature of a possible N-terminal tail in A43, since previously published data suggested that 32 N-terminal residues of A43 can be removed from the protein (Meka et al., 2003). However, our variant A14/A43-1, which also lacks 32 N-terminal residues, could only be expressed as a substoichiometric complex (Figs. 6A and 7). The A14/A43 structure subsequently showed that residues 24-27 of A43 interact with residues 28-30 of A14. Limited proteolysis of full-length A14/A43 revealed two cleavage sites in the N-terminal region of A43 after residues 16 and 22 (Fig. 6A). Variants A14/A43-2 and A14/A43-3, containing two different deletions before residue 23 at the N-terminus, could be purified as stable stoichiometric subcomplexes (Figs. 6A and 7). Thus, the first 22 residues at the N-terminus can be removed without affecting the stability of the complex, while residues 23-32 were important for stable binding of A14 to A43. Owing to the remaining uncertainties we decided not to remove any residues at the N-terminus and instead focused on more extended loops in A43.



**Figure 6. Schematic representation of yeast A43 and A14 variants**

The full-length proteins A43 (**A**) and A14 (**B**) are shown as colored bars (blue and red, respectively). Disordered regions that are not present in the crystal structure are depicted as dashed lines. Arrows indicate proteolytic cleavage sites. Variants are numbered A43-1 to A43-11 for A43 and A14-1 to A14-14 for A14. All residue numbers correspond to residues present in the variant. The minimal A43 and A14 variants used for final crystallization are shown at the bottom (variants A43-11 and A14-14).



**Figure 7. Purification and crystallization of A14/A43 variants**

(A) SDS-PAGE analysis (Coomassie staining) of A14 purification after coexpression of different A43 deletion variants A43-x (Fig. 6A). (B) SDS-PAGE analysis (Coomassie staining) of A43-6 purification after coexpression of different A14 deletion variants A14-x (Fig. 6B). (C) SDS-PAGE analysis (Coomassie-staining) of the A14-14/A43-11 variant. (D) Crystals obtained with the A14-14/A43-11 variant. The protein was crystallized using as protein buffer 100 mM NaCl, 20 mM Tris-HCl pH 8.0 and 5 mM DTT. Reservoir solutions: small microcrystals, 200 mM NaOAc, 30% (w/v) PEG 4000, 100 mM Tris-HCl pH 8.5 (left panel); intergrown crystals, 50 mM potassium phosphate, 20% (w/v) PEG 8000 (second panel from the left); larger crystals, 200 mM potassium acetate, 20% (w/v) PEG 3350 (third panel from the left); clusters of single crystals, 220 mM potassium acetate, 26% (w/v) PEG 3350 (fourth panel from the left); crystal obtained after microseeding, 300 mM potassium acetate, 20% (w/v) PEG 3350 (right panel).

Limited proteolysis of the A14/A43 subcomplex showed an additional cleavage site within the predicted ‘tip loop’ region (Fig. 6A). However, we did not create a deletion variant lacking the tip loop because it is involved in a key interaction with the core of Pol I. During purification of the A14/A43-1 variant, an additional protein variant was detected by SDS-PAGE. Edman sequencing revealed a truncated A43 variant starting at residue 81, preceded by an additional methionine residue. Analysis of the A43 nucleotide sequence revealed a

Shine-Dalgarno-Sequence within the A43 open reading frame. To prevent expression of the truncated variant, two silent point mutations (nucleotide changes A228G and G243T) were inserted and were retained for all subsequent A14/A43 variants.

Consistent with the prediction that A43 contains a long flexible C-terminal tail, a 46 amino acid C-terminal deletion variant (Peyroche et al., 2002) could be expressed without affecting the solubility or stoichiometry of the complex. Screening of variants that were shortened from the C-terminus demonstrated that up to 75 residues of A43 could be removed without affecting the solubility or stability of the subcomplex (variants A14/A43-4, A14/A43-5 and A14/A43-6, Figs. 6A and 7). After the C-terminal tail had been successfully delineated and removed, variant A14/A43-6 was subjected to limited proteolysis, to identify the location of the major loop. Comparison with Rpb7 revealed approximately 35 additional amino acids that were predicted to form a major loop in A43. In order to locate the loop, limited proteolysis experiments were performed, which revealed two cleavage sites after residue K206 and F207 of A43 (Fig. 6A). However, variants A14/A43-7, A14/A43-8 and A14/A43-9, which contained different A43 loop deletions including the proteolytic cleavage sites, showed strongly impaired solubility (Figs. 6A and 7), indicating that the loop was smaller. Indeed, a A43 variant lacking residues 194-209 (variant A14/A43-10) could be expressed as a soluble stoichiometric heterodimer (Figs. 6A and 7). The loop deletion could be extended to residues 173-209 (A14/A43-11, Figs. 6A and 7). These results suggested that the loop was located between  $\beta$ -strands C1 and C2 within the OB domain (Figs. 5 and 6A), as  $\beta$ -strand C1 was predicted to lie between residues 165 and 170, immediately before the loop deletion. Combining this loop deletion with the C-terminal truncation resulted in a variant that lacks residues 173-209 and 252-326 and was well expressed and soluble (variant A14/A43-11, Figs. 6A and 7).

### **2.3 Identification of unstructured regions in A14**

Next we tested experimentally the two models for the A14 structure that differed by the presence or absence of a C-terminal HRDC domain. We checked the presence of a HRDC domain by coexpressing A43 with A14 variants with various C-terminal deletions and assessed the solubility and stoichiometry of the A14/A43 heterodimer. Since in the counterpart structures subunit interaction does not rely on the HRDC domain, we expected that the HRDC domain could be removed without impairing the solubility of the subcomplex, as observed for the C17/25 subcomplex (Jasiak et al., 2006). Multiple C-terminal deletion

variants were designed, lacking 71, 66 and 62 residues, respectively (variant A14-1/A43-6, A14-2/A43-6 and A14-3/A43-6) (Figs. 6B and 7), consistent with the previously predicted HRDC domain (Meka et al., 2003). Purification of the hexahistidine-tagged A43 did not result in copurification of the truncated A14 protein, indicating that the deletion variant lacked parts of the A43-binding domain. We therefore reduced the length of the A14 C-terminal deletion until a minimal stoichiometric A14/A43 subcomplex was obtained (A14-4/A43-6 to A14-10/A43-6, Figs. 6B and 7). The A14-10/A43-6 complex could be purified at stoichiometric levels. This result argued against the presence of a HRDC domain in A14. Instead, the insolubility of the shorter A14 variants suggested that these contained an incomplete tip-associated domain.

We therefore tested the second model for the A14 structure, which suggested that a predicted  $\alpha$ -helix between residues 82 and 96 corresponds to helix H2 of the tip-associated domain (Peyroche et al., 2002). This model proposed a major loop in A14 between helices H1 and H2, ranging approximately from residue 50 to 80. Consistent with this model, coexpression of several A14 variants with various loop deletions (variants A14-11/A43-6 to A14-13/A43-6, Figs. 6B and 7) led to a minimal A14 variant, which forms a soluble and stoichiometric complex with the A43 variant (variant A14-14/A43-11, Figs. 6 and 7C).

## **2.4 Additional changes to obtain diffraction-quality crystals**

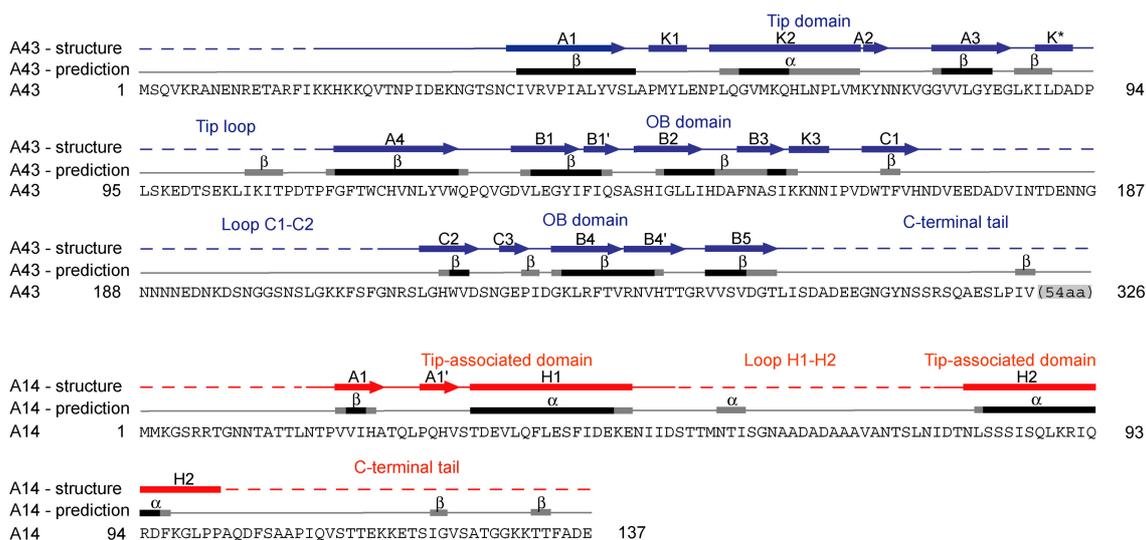
By combining a total of four precise deletions, two each in A14 and A43, we obtained a minimal A14/A43 variant (A14-14/A43-11, Figs. 6 and 7C). This variant was purified and subjected to various crystallization screens, varying protein concentration, protein buffer concentration and pH, reducing agents and additives. Although the variant was highly soluble, not sensitive to limited proteolysis, and apparently contained stoichiometric amounts of subunits, only very small microcrystals and clusters of small crystals were obtained under varying conditions and using different crystallization methods (Fig. 7D, left panel). We therefore changed the position of the affinity-purification tag from the C-terminus of A43 to the N-terminus of A14 and removed it by thrombin cleavage after affinity purification, finally leading to the formation of single protein crystals. However, over-nucleation resulted in crystals of medium size that were often clustered and intergrown (Fig. 7D, middle three panels). Only very few crystals showed diffraction beyond 4 Å resolution. With the use of microseeding (Bergfors, 2003), we could reduce nucleation and obtain well shaped, large single crystals when using 300 mM potassium acetate and 20% (w/v) PEG 3350 as reservoir

solution. These crystals were suitable for data collection to high resolution and enabled structure determination (Fig. 7D, right panel).

## 2.5 Conclusions

The removal of flexible protein regions has emerged as a general tool to achieve crystallization of proteins and their complexes. In the future, the number of crystallization projects hampered by multiple protein flexibilities will increase, in particular since there is a need to crystallize subcomplexes of larger assemblies that are subjected to hybrid analysis, combining EM of the entire assembly with X-ray crystallography of its subcomplexes. However, if more than one flexible region needs to be removed from a protein or a protein complex, the number of variants that have to be cloned, expressed, purified and subjected to crystallization trials becomes very large, as several variants for each deletion must be tested in all combinations. If four extended flexible regions exist, as in the example described here, only ten variants for each flexible region will lead to a total of 10000 variants with different combinations of deletions. In addition, if only one flexible region is not predicted correctly, all variants will be insoluble and useless for the identification of a suitable variant. Furthermore, high-throughput techniques can be applied to delineate flexible terminal tails, but are not suited to finding and removing flexible internal loops, which is often crucial for crystallogenesis.

Here, we present an alternative approach to engineering a protein with multiple extended flexible regions such that it will form well-ordered crystals. The novelty of this approach is that it successfully integrates structure prediction and structure-guided variant design and that it involves cycles of predicting flexible regions and experimental verification or falsification of the predictions. In particular, flexible regions are detected stepwise and also combined in a stepwise manner, rather than in the parallel, high-throughput approach that is often followed in structural proteomics. This stepwise detection and removal of flexible regions dramatically reduces the number of variants to be made and tested. The key to success of this procedure is to iteratively improve the prediction of unstructured regions and then use the improved prediction to design new variants for additional experiments. As more and more protein structures become available, the likelihood of finding a structure of a protein that is distantly related to the target protein to be crystallized increases. As a consequence, the number of crystallization projects that are able to use this approach will increase in the future.



**Figure 8. Comparison of secondary structure elements observed in the A14/A43 crystal structure with predicted secondary structure elements.**

Predicted secondary structure elements ( $\alpha$ -helices and  $\beta$ -strands) are shown above the amino acid sequences of A43 and A14 as black or grey bars for elements predicted with high (9-5) or low (4-1) reliability index (Experimental procedures). Secondary structure elements taken from the crystal structure (chapter III, Figs. 13, 14 and 15) are shown above the predicted secondary structure elements (bars,  $\alpha$  helices; arrows,  $\beta$  strands; lines, loops; dashed lines, disordered regions).

### 3 Experimental procedures

#### 3.1 Bioinformatic analysis

ClustalW (Chenna et al., 2003) was used to align A14 (GeneID 851734) and A43 (GeneID 854518) with their counterparts Rpb4, C17, RpoF and Rpb7, C25, RpoE, respectively. Alignments were manually edited using SEAVIEW (Galtier et al., 1996). In addition, A14 (137 residues, molecular weight of 14.6 kDa) and A43 (326 residues, molecular weight 36.2 kDa) were aligned with related yeast orthologues (*Candida glabrata*, *Candida albicans*, *Yarrowwia lipolica*, *Debaryomyces hanseii*, *Ashbya gossypii*, *Kluyveromyces lactis* and *Pichia stipitis*). The secondary structures of A14 and A43 were predicted with PredictProtein (Rost et al., 2004). The sequences of A14 and A43 were submitted to the HHpred server (Soding et al., 2005) using default settings.

#### 3.2 Cloning, expression, purification and limited proteolysis of protein variants

Cloning of all A14/A43 variants (Fig. 6) was carried out as described in chapter III (3.4), except that genes were cloned sequentially into vector pET21b, resulting in a C-terminal hexahistidine tag on A43. Additional residues from the polylinker and tag sequence, which remain on A43, were AAALHHHHHH. All A14/A43 variants were expressed recombinantly (Maniatis et al., 1982) and purified as described in chapter III (3.4), except for the thrombin-cleavage step and some individual changes, as follows. Full-length A14/A43 and variants A14/A43-1, A14/A43-2, A14/A43-4, A14/A43-5 and A14/A43-6 were purified by immobilized metal-affinity chromatography (IMAC) using as buffer 150 mM NaCl, 50 mM Tris pH 7.5, 5% glycerol, 10 mM  $\beta$ -mercaptoethanol, 1 mM PMSF, 1 mM benzamidine, 200  $\mu$ M pepstatin, and 60  $\mu$ M leupeptin. Full-length A14/A43, A14/A43-2 and A14/A43-6 were further purified on a Mono Q 10/100 GL anion exchange column (GE Healthcare), using as buffer 150 mM NaCl, 50 mM Tris pH 7.5, 5 mM DTT and were subsequently applied onto various gel filtration columns (Superose 6 HR 10/300, Superose 12 HR 10/300 and Superdex 75 FPLC, GE Healthcare). For limited proteolysis (Hubbard, 1998), protein was diluted to a concentration of 1 mg/ml. 100  $\mu$ l protein solution was incubated with 1  $\mu$ l of chymotrypsin or trypsin (1 mg/ml) and incubated for 1 h at 37 °C. 15  $\mu$ l samples were taken at different time points ranging from 1 to 60 min. The reaction was stopped by adding 5  $\mu$ l of 4x SDS sample

buffer (50 mM Tris pH 7.0, 14% 1,4-dithiothreitol, 10% glycerol, 1%  $\beta$ -mercaptoethanol, 0.1% bromophenol blue, 0.1% sodium lauryl sulfate) and incubated for 5 min at 95 °C. Samples were analyzed by SDS-PAGE.

### **3.3 Crystallization and crystal screening**

Initial crystallization experiments of all A14/A43 variants were performed at 20 °C in sitting drops using the Hydra II Plus One crystallization robot (Thermo Fisher Scientific). Commercially available screening kits were used for initial screens (Qiagen, Jena Bioscience, Hampton Research). The A14/A43 variants were concentrated to 10 to 30 mg/ml. Details on the crystallization of native and selenomethionine-derivatized A14-14/A43-11 can be found in chapter III (3.5).

## **Chapter III:**

# **Structure of A14/43 in functional context of RNA polymerase I**

## 1 Introduction

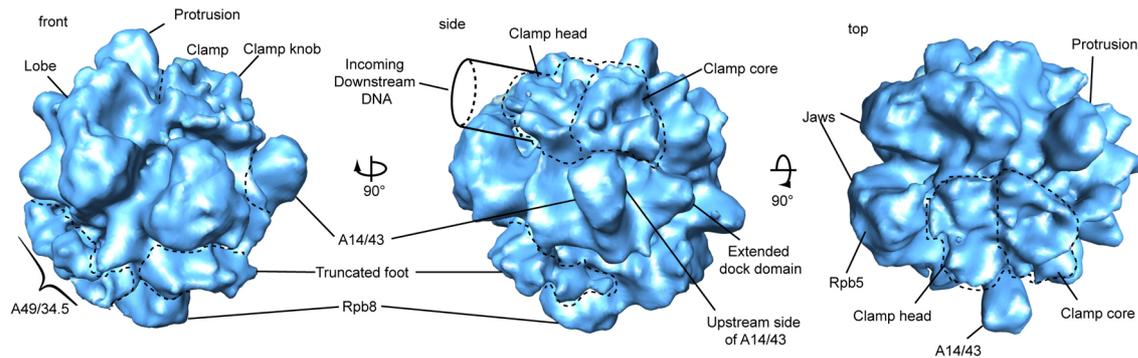
Pol I has a molecular weight of 590 kDa and comprises 14 subunits (Russell and Zomerdijk, 2006) (chapter I, Table 1). Subunits Rpb5, Rpb6, Rpb8, Rpb10, and Rpb12 are identical in all three polymerases. The two large Pol I subunits A190 and A135 contain regions homologous to the Pol II subunits Rpb1 and Rpb2, respectively. Subunits AC40 and AC19 are identical in Pol I and Pol III, and homologous to the Pol II subunits Rpb3 and Rpb11, respectively. Subunits A14 and A43 form the heterodimer A14/43 (chapter II), which is distantly related to Rpb4/7 in Pol II and C17/25 in Pol III (Hu et al., 2002; Jasiak et al., 2006; Meka et al., 2003; Peyroche et al., 2002; Sadhale and Woychik, 1994; Siaut et al., 2003). Subunits A49 and A34.5 are Pol I-specific. Subunit A12.2 is homologous to subunit Rpb9 in Pol II and C11 in Pol III, but its C-terminal domain is also related to the Pol II transcript cleavage factor TFIIS. To date most progress in structural studies of nuclear RNA polymerases has been made for Pol II, culminating in the refined atomic structures of the 10-subunit core (Cramer et al., 2001) and the complete enzyme (Armache et al., 2005). For Pol III, electron microscopy (EM) structures (Fernandez-Tornero et al., 2007) (Vannini and Cramer, unpublished), and a homology model for the core enzyme and the crystal structure of the C17/25 subcomplex are available (Jasiak et al., 2006). For Pol I, the overall shape and dimensions were first revealed by EM analysis of two-dimensional crystals (Schultz et al., 1993). Subsequent cryo-EM at 34 Å resolution visualized a stalk containing A14/43, and densities for A49 and A34.5 over the central cleft (Bischler et al., 2002; Peyroche et al., 2002). Later EM analysis with cryo-negative staining at 22 Å confirmed the stalk, but not the location of A49 and A34.5 (De Carlo et al., 2003).

Here we integrate structural biology methods to establish the complete subunit architecture and domain organization of Pol I. In addition, we define functional roles for subunits A49, A34.5, and A12.2, and report an intrinsic RNA cleavage activity of Pol I. The results uncover conserved and specific structural and functional principles in eukaryotic RNA polymerases and enable a detailed structure-function analysis of rRNA transcription.

## 2 Results and Discussion

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

Preparation protocol for the Pol I enzyme was established by Claus Kuhn and Jochen Gerber (Experimental procedures, 3.1), (Gerber et al., 2008; Kuhn et al., 2007). Cryo-EM map of Pol I was reconstructed by Claus Kuhn and Sonja Baumli (Fig. 9) (Kuhn et al., 2007).

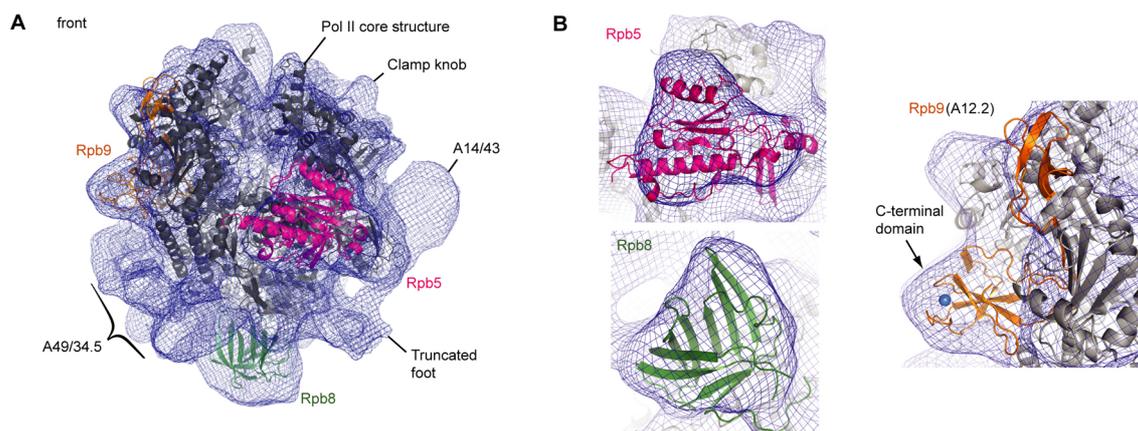


**Figure 9. Cryo-EM reconstruction of Pol I**

Views and structural regions are named as in the Pol II structure (Cramer et al., 2001), as in (Kuhn et al., 2007).

### 2.2 Homology model of the Pol I core

A Pol I homology model was created by Claus Kuhn (Fig. 10) (Kuhn et al., 2007).

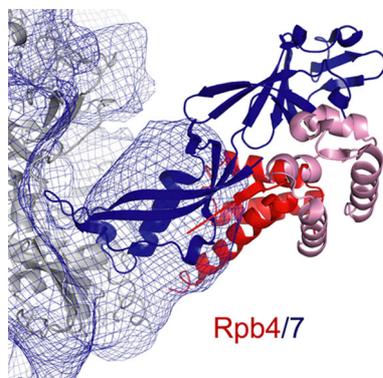


**Figure 10. Model and EM features of the Pol I core**

(A) Placement of the Pol II 10-subunit core structure (Armache et al., 2005) (grey) into the EM density (blue). The foot was deleted, and subunits Rpb5, Rpb8, and Rpb9 are highlighted in magenta, green, and orange, respectively. The clamp has been fitted as a separate rigid body. (B) Fit of the common subunits Rpb5 and Rpb8 to the EM map, and density for the core subunit A12.2 (the Pol II homolog Rpb9 is shown as a ribbon model), as in (Kuhn et al., 2007).

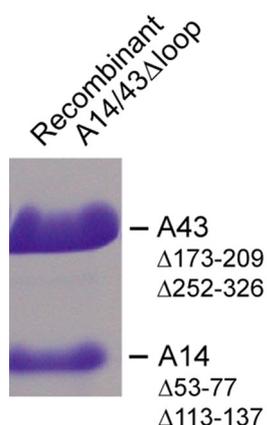
### 2.3 Crystal structure of the A14/43 subcomplex

After assigning EM densities to the Pol I core, a stalk-like density remained at the expected location for A14/43 that was much smaller than the structure of Rpb4/7 (Figs. 9, 10 and 11). Since the weak sequence similarity between A14/43 and Rpb4/7 or C17/25 did not allow for homology modeling, we determined the crystal structure of A14/43 (Experimental procedures 3.5). Partial proteolysis of a recombinant A14/43 heterodimer and bioinformatics revealed four mobile regions in A14/43 that were dispensable for dimerization (Figs. 12 and 15), as described in chapter II. An A14/43 variant lacking the mobile regions crystallized and enabled structure determination at 3.1 Å resolution (Figs. 13, 16 and Table 5).



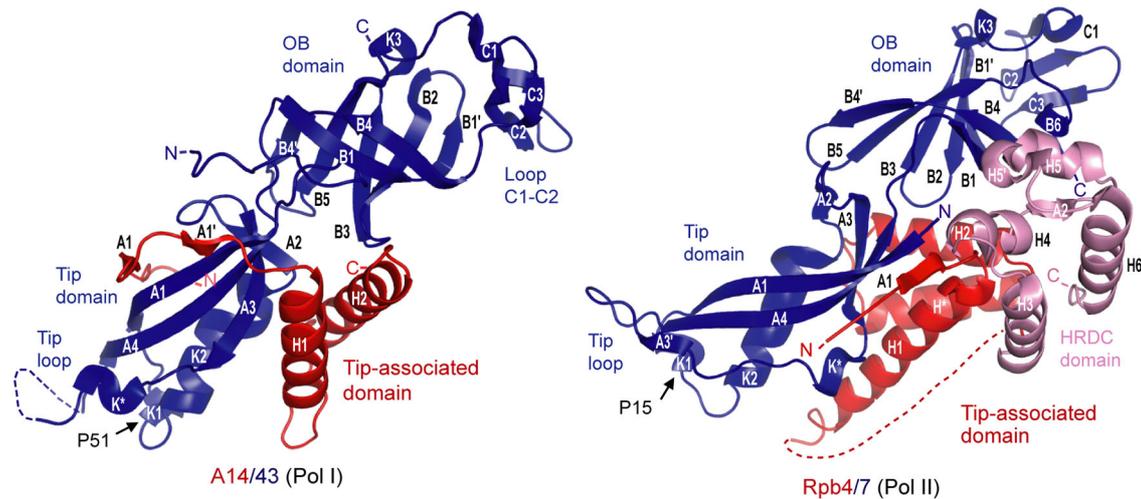
**Figure 11. EM density of Pol I stalk**

Fit of yeast Rpb4/7 (Armache et al., 2005) to part of the Pol I cryo-EM density, obtained by superposition with the tip domain of A43.



**Figure 12. Preparation of A14/43Δloop**

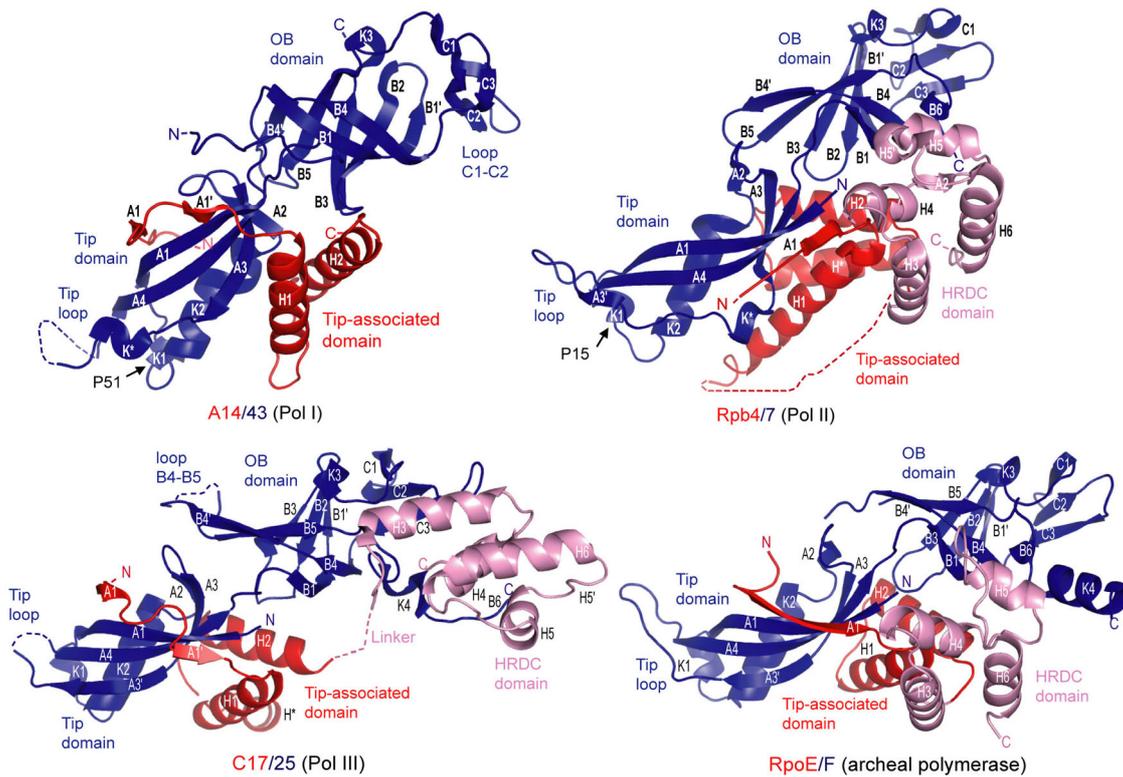
SDS-PAGE analysis of purified yeast A14/43Δloop used for crystallization (Experimental procedures).



**Figure 13. X-ray structure of the A14/43 subcomplex**

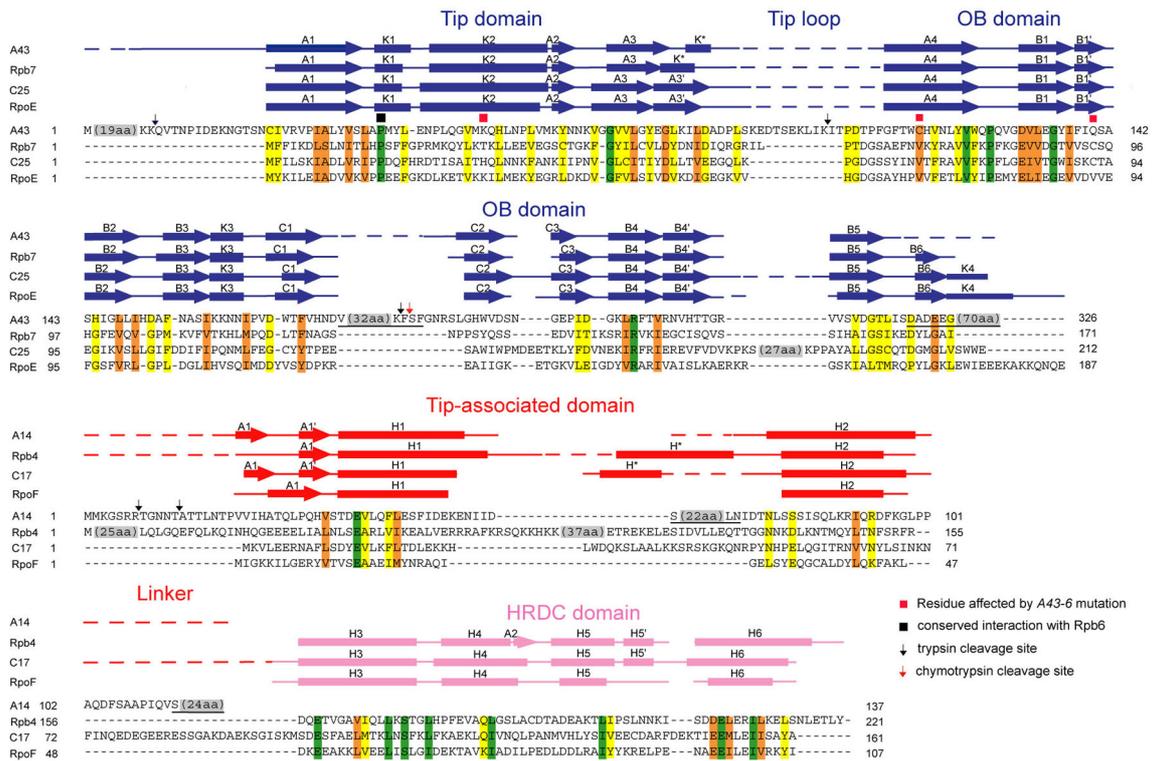
Structure of yeast A14/43 (this study) (left) and Rpb4/7 (Armache et al., 2005) (right). A43 and Rpb7 are in blue and A14 and Rpb4 are in red, with the HRDC domain in light red. For details see Figures 14 and 15. The view is as in Figures 9 and 10.

The overall structure of A14/43 resembles its counterparts Rpb4/7 (Armache et al., 2005), C17/25 (Jasiak et al., 2006), and the archaeal RpoF/E (Todone et al., 2001), except that A14 lacks the HRDC domain present in all counterparts (Figs. 13, 14 and 15). The N-terminal tip domain of A43 shows RMS deviations in C $\alpha$  atom positions of 2.2-2.5 Å, whereas the C-terminal OB domain is more divergent. A14 forms two helices that pack on the A43 tip domain (Figs. 13, 14 and 15).



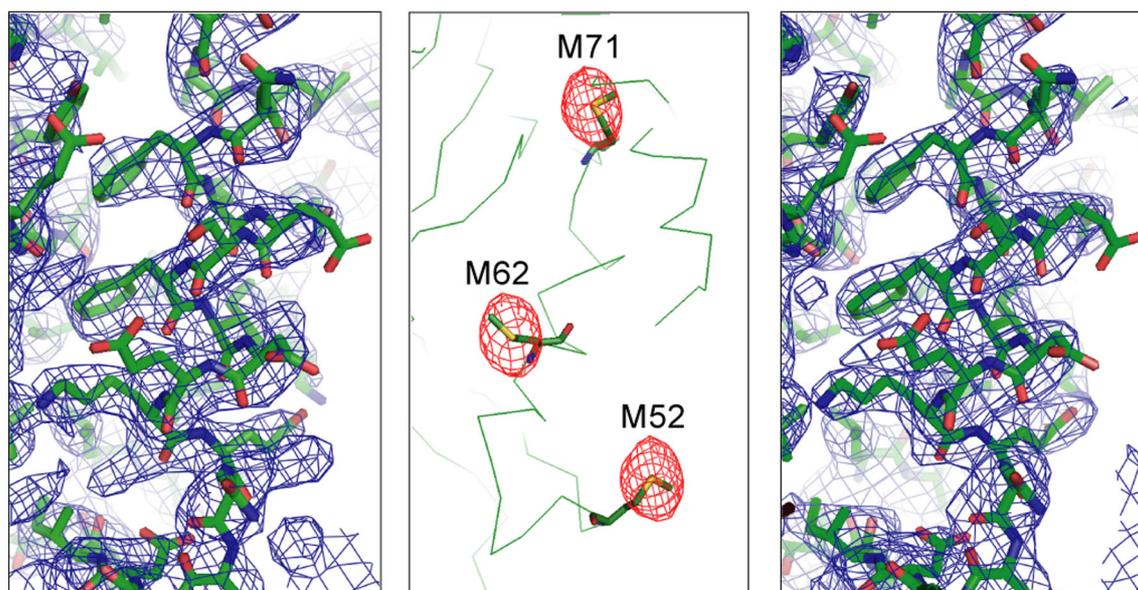
**Figure 14. Comparison of crystal structures of A14/43 and its counterparts in other RNA polymerases**

Structural comparison of yeast A14/43 (this study, upper left), Rpb4/7 (Armache et al., 2005) (upper right), C17/25 (Jasiak et al., 2006) (lower left) and archaeal RpoE/F (Todone et al., 2001) (lower right). A43, Rpb7, C25 and RpoE are in blue. A14, Rpb4, C17 and RpoF are in red; the HRDC domains are in light red. Disordered regions in the A14/43 structure include the A43 N-terminus (residues M1-H20), the A43 tip loop (residues S96-T111), the A43 loop C1-C2 (residues V173-F209), the A43 C-terminus (residues S251-D326), the A14 N-terminus (residues M1-T15), the A14 loop H1-H2 (residues S53-N77) and the A14 C-terminus (residues A102-E137).



**Figure 15. Structure-based alignment of A14/43 and its counterparts in other RNA polymerases**

Primary and secondary structure. Structure-based alignments of amino acid sequences of *S. cerevisiae* A43 (top) and A14 (bottom) with their counterparts in Pol II (Rpb7 and Rpb4), Pol III (C25 and C17) and an archaeal RNA polymerase (*M. Jannaschii* RpoE and RpoF). Secondary structure elements are shown above the sequences (broad lines,  $\alpha$ -helices; arrows,  $\beta$ -strands; lines, loops; dashed lines, disordered regions). Conserved residues are highlighted according to decreasing conservation from green, through orange, to yellow. Cleavage sites revealed by limited proteolysis are indicated with arrows. The invariant proline residue that is predicted to contribute to the A43-Rpb6 interface is indicated with a black square. Residues involved in A43-Rrn3 interaction (Peyroche et al., 2000) are indicated with red squares. Residues not present in the crystallized variant are underlined.



**Figure 16. Electron density for A14/43**

**(left)** Initial unbiased electron density map calculated from observed amplitudes and phases derived from single anomalous diffraction (blue, contoured at  $1.0\sigma$ ). The final model is superimposed. **(middle)** Anomalous difference Fourier map calculated with phases from the final model, revealing the selenium atoms (red, contoured at  $4.0\sigma$ ). The backbone of the final model is shown, selenomethionine side chains are depicted. **(right)** Final model with final 2Fo-Fc map superimposed (blue, contoured at  $1.0\sigma$ ).

**Table 5. A14/43 X-ray diffraction data and refinement statistics**

| Crystal A14/43 SeMet               |                                  |                        |       |
|------------------------------------|----------------------------------|------------------------|-------|
| <b>Data collection<sup>a</sup></b> |                                  | <b>Refinement</b>      |       |
| Space group                        | P2 <sub>1</sub> 2 <sub>1</sub> 2 | Amino acid residues    | 702   |
| Wavelength (Å)                     | 0.97848                          | RMSD bonds (Å)         | 0.009 |
| Unit cell axis (Å)                 | 229.9 63.9 65.4                  | RMSD angles (°)        | 1.9   |
| Resolution (Å)                     | 30-3.1 (3.25-3.1) <sup>b</sup>   | R <sub>cryst</sub> (%) | 25.3  |
| Completeness (%)                   | 99.6 (99.6)                      | R <sub>free</sub> (%)  | 28.5  |
| Unique reflections                 | 33,670 (4,457)                   |                        |       |
| Redundancy                         | 5.5 (5.4)                        |                        |       |
| R <sub>sym</sub> (%)               | 7.7 (39.9)                       |                        |       |
| <I/σI>                             | 16.1 (5.0)                       |                        |       |

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

<sup>b</sup>Numbers in parenthesis refer to the highest resolution shell.

## 2.4 Interaction of A14/43 with the Pol I core

In Pol II, the Rpb4/7 complex interacts with the core enzyme via two loops, the A1-K1 loop, which forms a conserved contact of Rpb4/7-like subcomplexes with their cognate core enzymes, and the tip loop, which may confer specificity to the interaction in the different RNA polymerases (Armache et al., 2005; Jasiak et al., 2006). To dock the A14/43 structure into the EM map, we modeled the conserved contact between an invariant proline residue in the A1-K1 loop (P51 in A43, Figs. 13 and 19) and the common core subunit Rpb6 (Armache et al., 2005; Jasiak et al., 2006). The tip domain and the tip-associated domain of the A14/43 structure fitted well to the EM map, and the lack of an HRDC domain could in part explain the smaller EM density (Fig. 17). However, the peripheral OB domain of A43 was not revealed in the EM density (Fig. 17), suggesting a high degree of mobility. Consistently, the OB domain shows slightly higher B-factors in the crystal structure although it is involved in crystal contacts (Fig. 18), and normal mode analysis of the Pol II crystal structure shows that the OB domain is the most flexible region of the enzyme (not shown). The A43 tip loop contains a specific ten-residue insertion that may confer specificity to the interaction between A14/43 and the Pol I core (Figs. 14 and 15). The A43 tip loop is flexible in the crystal structure (Figs. 13, 14 and 15), but is likely folded upon binding to the Pol I core, as observed for Pol II (Armache et al., 2005).

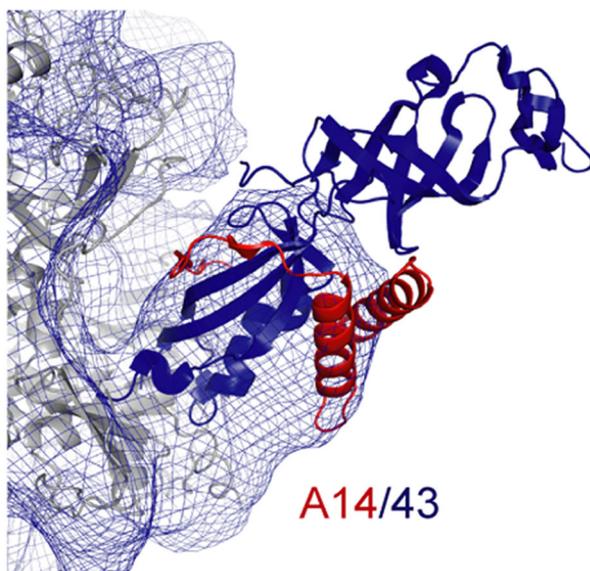
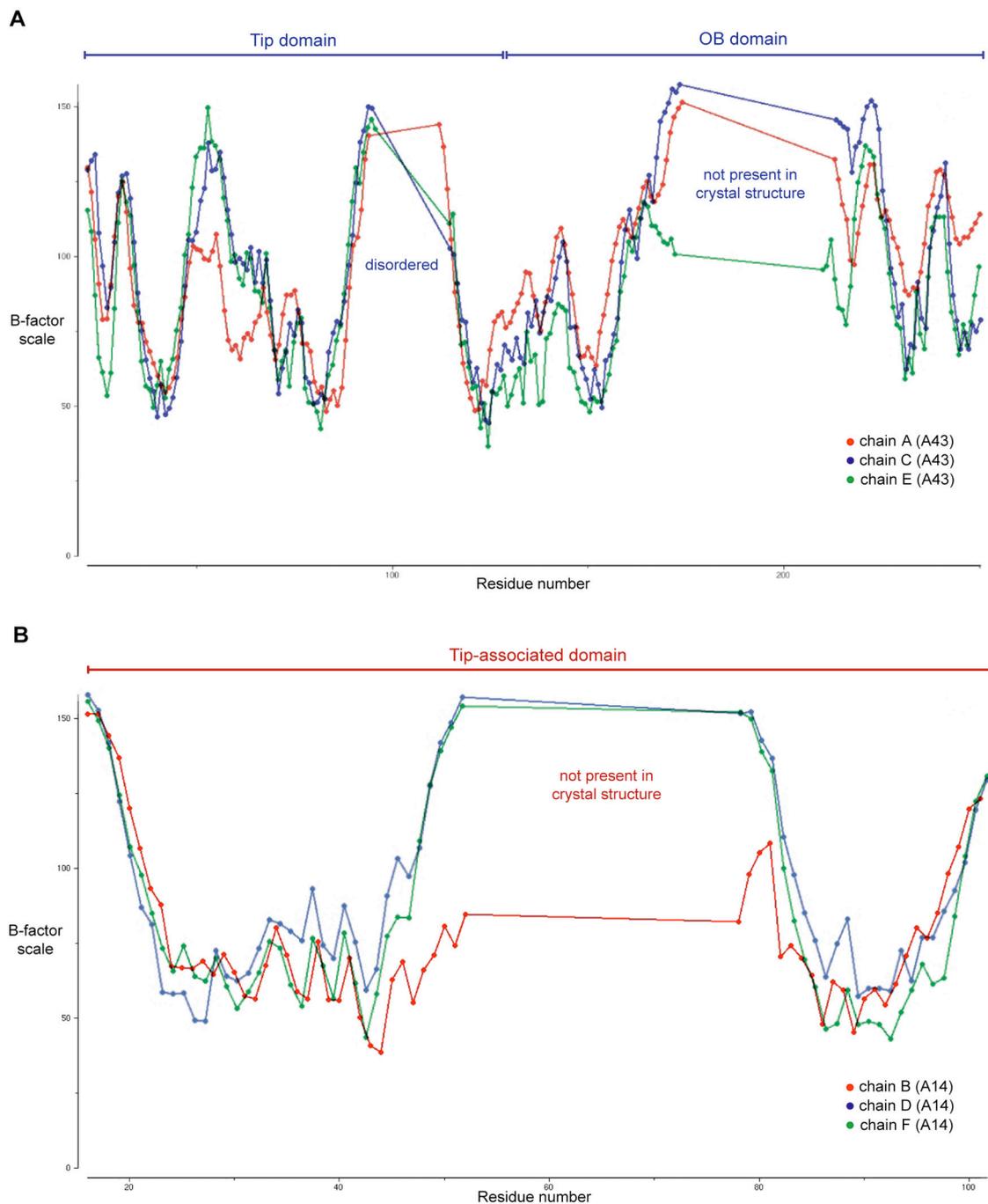


Figure 17. Fit of the A14/43 structure into the Pol I EM density



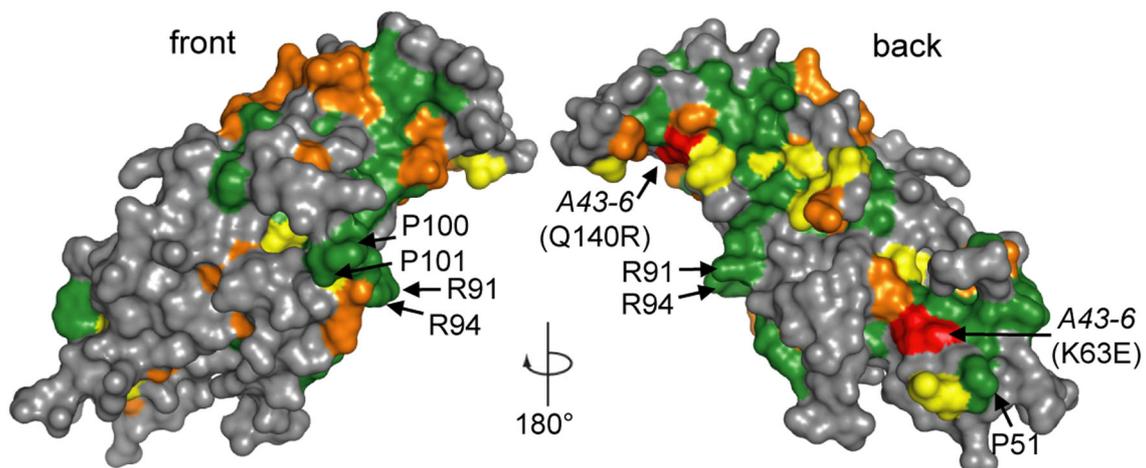
**Figure 18. Distribution of crystallographic B-factors in the A14/43 structure**

The graph shows the distribution of crystallographic B-factors according to the residue number in A43 (A) and A14 (B). The three complexes in the asymmetric unit are colored in red (chain A and chain B), blue (chain C and chain D) and green (chain E and chain F).

## 2.5 Specific interactions with initiation factors

Subunit A43 forms an essential bridge to the conserved Pol I initiation factor Rrn3 (Milkereit and Tschochner, 1998; Peyroche et al., 2000). Rrn3 was shown by EM to co-localize with A43 (Peyroche et al., 2000) and binds other initiation factors to recruit Pol I to the rDNA promoter. The A43-Rrn3 interaction is conserved in human (Yuan et al., 2002) and *S. pombe* (Imazawa et al., 2005). In a Pol I variant that is defective for Rrn3 interaction (*rpa43-6*) (Peyroche et al., 2000), two out of three altered A43 residues map near conserved residues on the upstream surface of A14/43 (Fig. 19). Thus Rrn3 binds to A14/43 from the upstream side (Fig. 9). Additional Pol I-specific surfaces in the vicinity include the extended dock domain and the clamp knob, which together with A14/43 create a specific upstream face for Pol I initiation complex assembly (Figs. 9 and 10).

Differential initiation factor interactions and promoter-specificity of the three polymerases may generally result from differently structured dock domains, clamps, and Rpb4/7-like subcomplexes, which all constitute initiation factor binding sites. Rpb4/7 is required for Pol II initiation (Edwards et al., 1991). C17/25 binds to the Pol III initiation factor TFIIB (Ferri et al., 2000), to the subcomplex C82/34/31 that bridges to TFIIB (Bartholomew et al., 1993; Brun et al., 1997; Thuillier et al., 1995), and to the initiation factor TFIIC (Hsieh et al., 1999). Since the surfaces, flexibility, and *in vivo* function of the HRDC domains differ in Rpb4/7 and C17/25 (Jasiak et al., 2006), the absence of an HRDC domain in A14/43 is likely functionally significant.

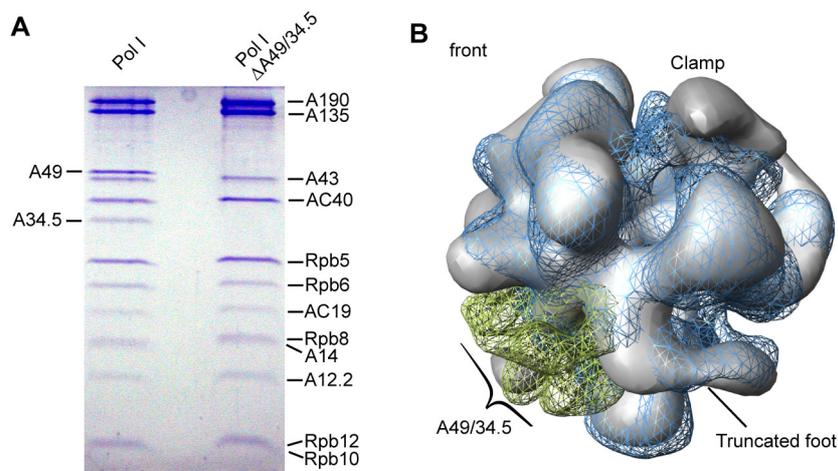


**Figure 19. Surface representation of the A14/43 subcomplex**

Residues conserved among eight, selected *Saccharomycotinae* are colored in green, orange, and yellow, according to decreasing conservation (compare Figure 15). Residues affected by the *A43-6* mutations (Peyroche et al., 2000) are in red.

## 2.6 A49 and A34.5 form a TFIIIF-like heterodimer near the funnel

After assigning EM densities to the Pol I core and A14/43, one additional large density remained on the enzyme surface that was assigned to the Pol I-specific subunits A49 and A34.5 (Fig. 9). To confirm this assignment, subunits A49 and A34.5 were dissociated from Pol I with the use of urea (Huet et al., 1975), the resulting 12-subunit variant Pol I  $\Delta$ A49/34.5 (Pol I $\Delta$ , previously called Pol A\* (Huet et al., 1975)) was purified, and its structure was solved by cryo-EM at 25 Å resolution (Fig. 20). The Pol I $\Delta$  structure was similar to the complete Pol I, except that the density assigned to A49 and A34.5 was lacking (Fig. 20B). In addition, there was a minor change in the clamp conformation, which however represents an average clamp position and is unlikely to result from the absence of A49/34.5. Density assigned to A49 and A34.5 is located near the enzyme funnel, the external domain 1, a conserved core loop with a Pol I-specific insertion (equals loop  $\alpha$ 16- $\beta$ 20 of Pol II pore domain), and A12.2, which is consistent with loss of A49 when Pol I is purified from A12.2 deletion strains (Van Mullem et al., 2002).

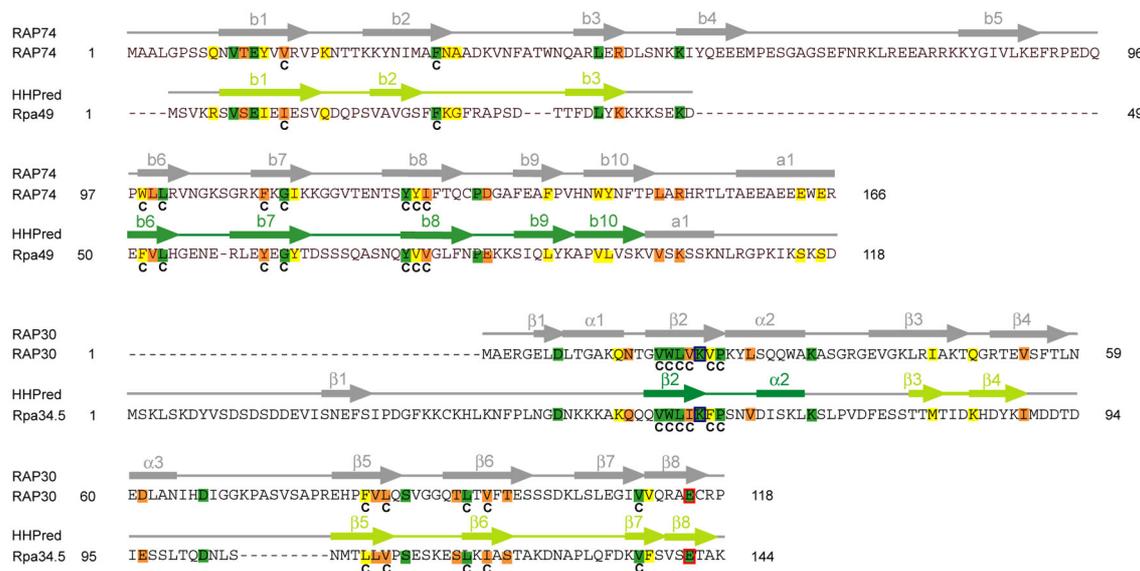


**Figure 20. Structural features of A49/34.5**

(A) SDS-PAGE of Pol I $\Delta$  (right), obtained by urea treatment of the complete Pol I (left). (B) EM structures of Pol I  $\Delta$  (silver surface) and the complete Pol I (blue). The density assigned to A49/34.5 is highlighted in green, as in (Kuhn et al., 2007).

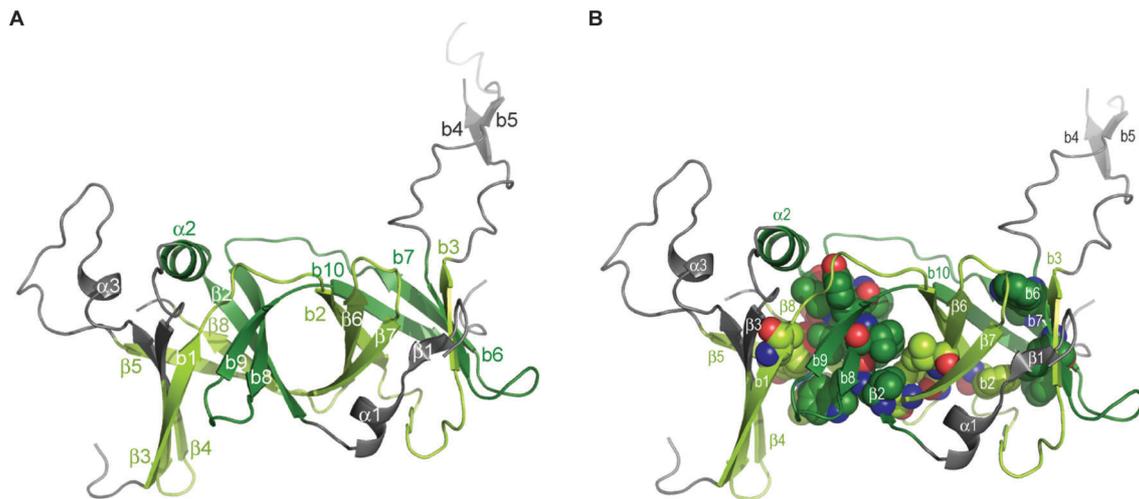
- Pol I $\Delta$  preparation and structure determination was performed by Claus Kuhn, with the help of Marco Gartmann (Kuhn et al., 2007) -

To investigate the structure and function of A49 and A34.5, we searched for weak homologies with HHpred (Soding et al., 2005). Local homologies were detected between A49 and RAP74, the large subunit of the Pol II-associated factor TFIIIF, and between A34.5 and RAP30, the small TFIIIF subunit (Fig. 21, Experimental procedures). Consistently, the N-terminal regions of A49 and A34.5 were predicted to contain  $\beta$ -strands consistent with the fold of the RAP74-RAP30 dimerization module (Gaiser et al., 2000), and hydrophobic core residues in this fold were predicted to be conserved (Figs. 21 and 22).



**Figure 21. Sequence alignment of A49/34.5 with TFIIIF Rap74/30**

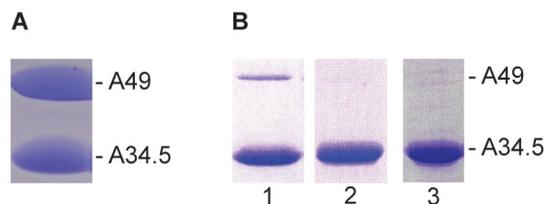
Sequence alignments of amino acid sequences of *S. cerevisiae* A49 (top) and A34.5 (bottom) with their putative counterparts in *H. sapiens* TFIIIF (RAP74 and RAP30, respectively). Sequence similarity is only observed in the N-terminal part of both proteins (residues 1-166 in RAP74 and residues 1-118 in RAP30). Secondary structure elements are shown (bars,  $\alpha$ -helices; arrows,  $\beta$ -strands; lines, loops), according to structural information (Gaiser et al., 2000). For clarity, the symbols a/b are used in RAP74,  $\alpha/\beta$  in RAP30. For A49 and A34.5, predicted secondary structure elements are depicted in dark green (aligned by HHpred), light green (predicted by secondary structure propensity) and grey (not predicted to be present). Conserved residues are highlighted according to decreasing conservation from green, through orange, to yellow. Residues involved in a conserved core interaction are marked with a C below the sequence, while charged residues forming a salt bridge are depicted in blue and red, respectively.



**Figure 22. Structural similarity between A49/34.5 and TFIIIF Rap74/30**

Conservation of the TFIIIF RAP74/30 dimerization module (Gaiser et al., 2000) in A49/34.5. RAP74 residues 2-172 and RAP30 residues 2-119 form a triple barrel  $\beta$ -structure, which is likely to be conserved in A49/34.5. **(A)** Secondary structure elements aligned to RAP74/30 are highlighted in dark and light green, respectively. For details see Figure 21. **(B)** Residues involved in conserved hydrophobic core interactions are shown as spheres and colored according to their atom types. Residues involved in a putative salt bridge are shown as spheres and colored in blue (K22, RAP30) and red (E115, RAP30).

Consistent with these predictions, bacterial coexpression of A49 and A34.5 enabled isolation of a stoichiometric A49/34.5 heterodimer (Fig. 23A), and alanine point mutations in three different conserved hydrophobic residues in the dimerization interface (I12 and Y76 in A49, W54 in A34.5) abolished or strongly impaired A49/A34.5 copurification (Fig. 23B). Thus, A49 and A34.5 form a stable TFIIIF-like heterodimerization module.



**Figure 23. Preparation of A49/34.5 variants**

**(A)** SDS-PAGE analysis of recombinant A49/34.5 heterodimer (wildtype). **(B)** A49/34.5 variants, with His<sub>6</sub>-tag on A34.5, A49 Y76A (lane 1), A49 I12A (lane 2) and A34.5 W54A (lane 3).

Heterodimerization of A49 and A34.5 explains why the two subunits dissociate together from Pol I upon urea treatment (Huet et al., 1975), why Pol I purified from a yeast strain lacking the gene for A34.5 also lacks A49 (Gadal et al., 1997), and why two distantly related mammalian Pol I subunits bind each other (Yamamoto et al., 2004). It is also consistent with the observed continuous EM density, which reconciles previous EM data. Initial cryo-EM showed two separated densities over the cleft that were assigned to A49 and A34.5 (Bischler et al., 2002). EM at higher resolution did not confirm these densities, but revealed a new additional density (De Carlo et al., 2003) that was close to the location of A49/34.5 found here.

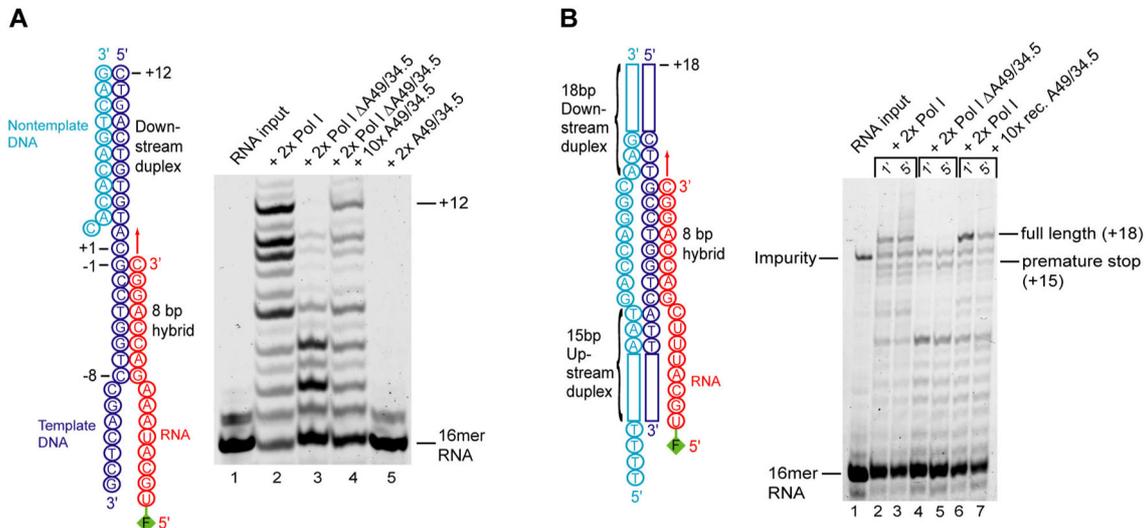
The location of A49/34.5 at the Pol I funnel deviates from that of TFIIF on Pol II as observed by cryo-EM (Chung et al., 2003), but is more consistent with protein-protein cross-linking that maps TFIIF to the polymerase lobe and outer surface near Rpb9 (Chen et al., 2007; Chen et al., 2010). Discrepancies in the location of A49/34.5 and TFIIF may be explained by different locations of a related dimerization module on the two polymerases, or by the presence of additional, unrelated domains in both factors. Sequence analysis showed that A49/34.5 and TFIIF have a counterpart in Pol III, the C37/53 heterodimer (chapter IV, Fig. 52), which occupies a similar location on the Pol III surface near the lobe and funnel (Fernandez-Tornero et al., 2007).

## **2.7 A49/A34.5 is a built-in processivity factor**

The apparent homology of the A49/34.5 heterodimer with the N-terminal regions of the two large TFIIF subunits suggested that A49/34.5 could stimulate Pol I processivity. This prediction is consistent with previous reports that deletion of A49 or A49 and A34.5 reduce Pol I activity (Huet et al., 1975; Liljelund et al., 1992), and that A34.5 genetically interacts with DNA topoisomerase I (Gadal et al., 1997). We therefore compared the complete Pol I with Pol I $\Delta$  in an RNA extension assay using a minimal DNA-RNA scaffold (Brueckner et al., 2007). The complete Pol I extended the RNA to the end of the template, whereas Pol I $\Delta$  did not produce the runoff product (Fig. 24A). Addition of recombinant A49/34.5 rescued the defect of Pol I $\Delta$  and enabled synthesis of the run-off product (Fig. 24A, lane 4).

We repeated the extension experiments using a complete, complementary transcription bubble scaffold (Kireeva et al., 2000). The complete Pol I produced the run-off transcript (+18), whereas Pol I $\Delta$  did not, but addition of recombinant A49/34.5 heterodimer restored run-off formation (Fig. 24B). The defect was not due to differential binding of the polymerase

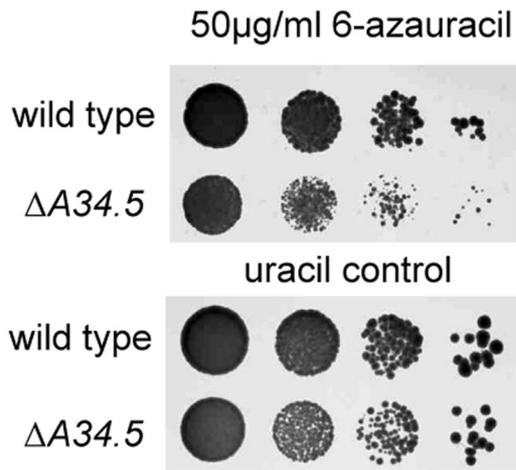
variants to the scaffold, as it was also observed when the elongation complexes were covalently coupled to magnetic beads and extensively washed before the reaction (not shown). Thus, A49/34.5 is required for normal processivity of Pol I *in vitro*.



**Figure 24. A49/34.5 extension assays**

(A) Minimal nucleic-acid scaffold used in RNA extension assays (left). The fluorescent label 6-carboxyfluoresceine (FAM) on the RNA 5' end is indicated. A49/34.5 stimulates Pol I processivity (right). The times molar excess of added factors are indicated above the lanes. For lane 4, Pol I  $\Delta$  was complemented with a 5-fold molar excess of recombinant A49/34.5 for 10 min at 20 °C prior to addition of the scaffold. (B) RNA extension assay as in A, but with a complete complementary bubble (Kireeva et al., 2000), as in (Kuhn et al., 2007).

To test whether A49/34.5 may stimulate Pol I processivity *in vivo*, we investigated if the growth phenotype of a yeast strain that lacked the gene for A34.5 ( $\Delta$ A34.5) is affected when nucleotide supply was limited due to the presence of 6-azauracil (6AU). 6AU sensitivity is an indicator for Pol II-associated elongation factor function *in vivo* and recently also identified a Pol I mutant defective in rRNA elongation (Schneider et al., 2007). Whereas the wildtype and  $\Delta$ A34.5 strains did not show a growth difference on normal media, the  $\Delta$ A34.5 strain showed a mild slow-growth phenotype on 6AU-containing media (Fig. 25). This suggests that A49/34.5 is required for normal processivity of RNA synthesis by Pol I also *in vivo*. Stimulation of processivity is probably due to an allosteric effect (chapter IV) and not due to an extension from A49/34.5 into the active center (see also Appendix 4), but we cannot definitely exclude one possibility with the available structural data.



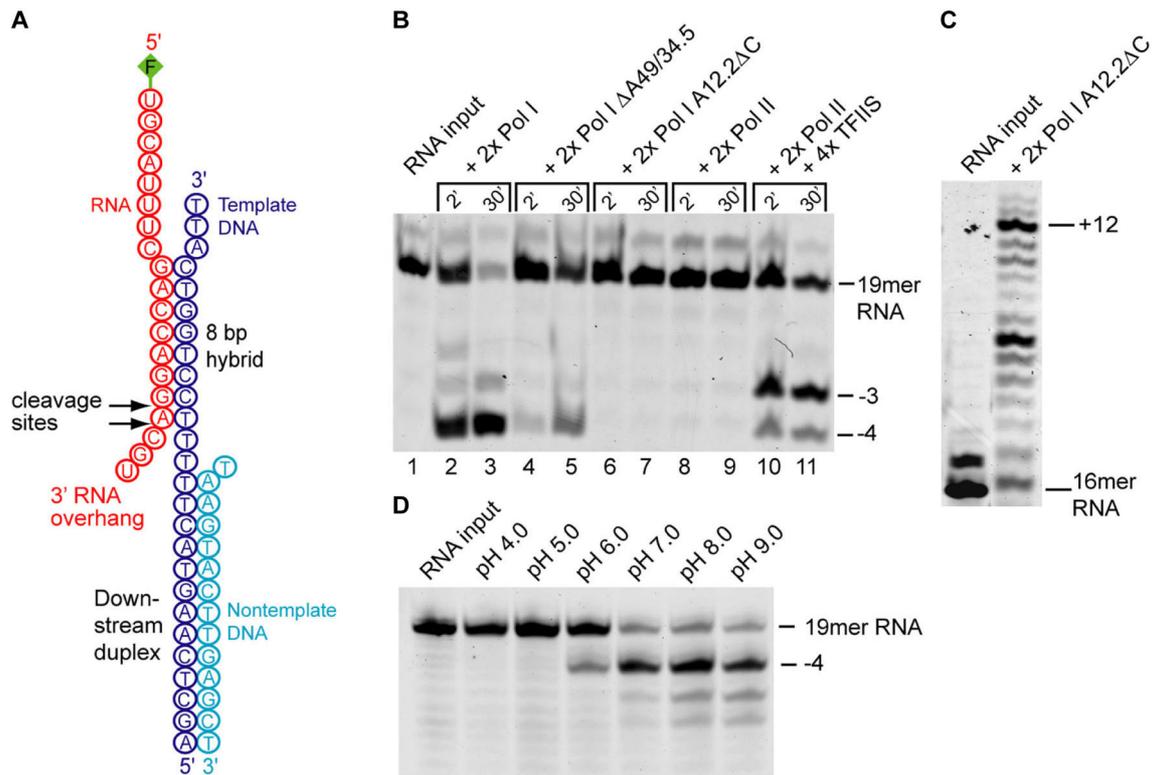
**Figure 25. A49/34.5 deletion strains *in vivo***

Deletion of the gene for A34.5 leads to a 6-azauracile-sensitive phenotype. From left to right 1:10 dilution series are shown. As a control, cells were spread onto SDC plates containing uracil, as in (Kuhn et al., 2007).

- *In vitro* assays were performed by Claus Kuhn, supplemented by Sebastian Geiger. *In vivo* assays were performed by Claus Kuhn (Kuhn et al., 2007) -

## 2.8 Pol I has intrinsic RNA cleavage activity

The active site of Pol II exhibits weak 3'-RNA cleavage activity that is stimulated by TFIIS (Wind and Reines, 2000). For Pol I, a RNase H-like nuclease activity was initially described (Huet et al., 1976), but was later found to reside in a dissociable factor (Huet et al., 1977; Iborra et al., 1979; Labhart, 1997; Tschochner, 1996). To clarify whether Pol I possesses intrinsic RNA cleavage activity, we assembled a 'backtracked' elongation complex from purified Pol I and a DNA-RNA scaffold that contained an RNA 3'-overhang (Fig. 26A, Experimental procedures, 3.11). Incubation of the backtracked complex with 8 mM magnesium ions led to efficient shortening of the RNA from the 3'-end (Fig. 26B, lanes 1-3). In comparison, Pol II was unable to cleave the RNA under these conditions, but addition of TFIIS resulted in cleavage (Fig. 26B, lanes 8-11). Thus, Pol I has a strong intrinsic RNA cleavage activity.



**Figure 26. Intrinsic RNA cleavage activity of Pol I**

(A) DNA-RNA hybrid scaffold used in cleavage assays. (B) Comparison of RNA cleavage by Pol I variants with Pol II and the Pol II-TFIIS complex. Pol I mainly removed four nucleotides from the RNA, consistent with binding of the terminal hybrid base pair to the nucleotide insertion site (+1), extrusion of the RNA 3' overhang into the polymerase pore and cleavage of the phosphodiester bond between nucleotides at positions -1 and +1 (Fig. 26A). The Pol II-TFIIS complex removed three or four nucleotides, indicating that a mixture of complexes was present with the terminal hybrid base pair occupying either position -1 or +1. (C) Processivity of the Pol I variant A12.2ΔC (D) pH dependence of Pol I cleavage activity, as in (Kuhn et al., 2007).

The intrinsic cleavage activity likely escaped detection previously since the nucleic acid substrates used in published studies did not allow for the formation of a backtracked state, from which cleavage occurs. The previously described dissociable factor (Huet et al., 1977; Iborra et al., 1979; Labhart, 1997; Tschochner, 1996) may not be required for cleavage per se, but may induce backtracking of Pol I, to create a state of the elongation complex that is prone to cleavage.

- Cleavage assays, as described in 2.8 and 2.9, were performed by Sebastian Geiger and Claus Kuhn, with the help of Stefan Jennebach (Kuhn et al., 2007) -

## 2.9 Pol I cleavage activity requires A12.2

Additional cleavage assays showed that Pol I $\Delta$  cleaved RNA less efficiently than the complete Pol I (Fig. 26B, lanes 4-5). Cleavage stimulation by A49/34.5 is consistent with an early investigation of an RNase H-like activity in Pol I-containing fractions (Huet et al., 1976). We also asked whether subunit A12.2 is required for cleavage, since its counterpart C11 is essential for cleavage activity of Pol III (Chedin et al., 1998; Whitehall et al., 1994). A Pol I variant lacking residues 79–125 of A12.2 (A12.2 $\Delta$ C, Experimental procedures 3.9) was totally inactive in RNA cleavage (Fig. 26B, lanes 6-7), but bound the nucleic-acid scaffold in electrophoretic mobility shift assays (not shown) and retained processivity (Fig. 26D). Consistent with a function specific for the A12.2 C-terminal domain, a truncation variant remains bound to Pol I and does not show a conditional growth defect (Van Mullem et al., 2002).

The A12.2 C-terminal domain shows homology to the TFIIS C-terminal domain that inserts into the Pol II pore to stimulate RNA cleavage (Kettenberger et al., 2003), but its location in Pol I corresponds to that of the Rpb9 C-terminal domain on Pol II (Fig. 10B). Although the long linker between the A12.2 N- and C-terminal domains could in principle allow swinging of the C-terminal domain into the pore, our results suggest that the effect of A12.2 truncation on cleavage is due to an allosteric rearrangement in the Pol I active center. The conserved polymerase active site is capable of RNA cleavage in the absence of cleavage stimulatory factors, since free Pol II and the bacterial RNA polymerase can cleave RNA under mild alkaline conditions (Orlova et al., 1995; Weilbaecher et al., 2003). Consistently, the intrinsic cleavage activity of Pol I increased with increasing pH (Fig. 26C). The structural basis of the effect of A12.2 truncation on RNA cleavage awaits the crystal structure of Pol I.

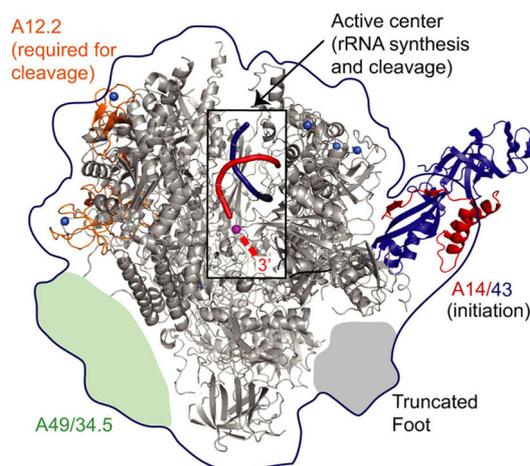
## 2.10 Possible functions of the cleavage activity

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

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

## 2.11 Conclusions

Here we analyzed the detailed functional architecture of Pol I by a combination of structural biology techniques and structure-based functional analysis (Fig. 27). A comparison with the Pol II system revealed Pol I-specific features that match the unique nature of rRNA transcription. First, the distinct structure of the Pol I upstream face allows for specific initiation factor interactions and recruitment of Pol I to the rRNA promoter. Second, the built-in processivity-stimulatory Pol I-specific subcomplex A49/34.5 can explain the efficient and processive nature of rRNA transcription during cell growth. Third, the intrinsic RNA cleavage activity apparently enables rRNA 3'-trimming and proofreading, to prevent formation of erroneous rRNAs and catalytically deficient ribosomes. Finally, our results unravel structural and functional relationships between the three eukaryotic transcription machineries, and form the basis for a detailed structure-function analysis of rRNA transcription and processing.



**Figure 27. Functional architecture of Pol I**

Hybrid structure and functional architecture of Pol I. The EM envelope is shown as a blue line, the Pol I core ribbon model in grey with Rpb9 (A12.2) highlighted in orange, and the A14/43 crystal structure in red/blue. The window shows a cut-away view of the active center containing a modeled DNA-RNA hybrid. Red dashes indicate the RNA 3'-end extruded into the pore, as in (Kuhn et al., 2007).

### 3 Experimental procedures

#### 3.1 Pol I preparation

The *S. cerevisiae* strain GPY2, carrying a pAS22 plasmid coding for HA- and hexahistidine-tagged A43 (Fath et al., 2000), was grown in a 200 L fermenter overnight at 30 °C, and cells were harvested at an OD600 of 5-6. Cells (30% slurry) were lysed by bead beating in 400 mM ammonium sulfate, 60 mM MgCl<sub>2</sub>, 150 mM HEPES pH 7.8, 30% glycerol, 5 mM DTT, 1 mM PMSF, 1 mM benzamidine, 200 μM pepstatin, and 60 μM leupeptin. After filtration, the lysate was cleared by centrifugation (30 min, 8000 g), and ultracentrifugation (90 min; 30,000 g). The supernatant was dialyzed overnight at 4 °C against 50 mM potassium acetate (KOAc), 20 mM HEPES at pH 7.8, 1 mM MgCl<sub>2</sub>, 10% glycerol, and 10 mM β-mercaptoethanol. After centrifugation, (1 h; 18,500 g), the pellet was resuspended in 1.5 M KOAc, 1 mM MgCl<sub>2</sub>, 10% glycerol, and 10 mM β-mercaptoethanol and incubated with 8 ml Nickel-NTA agarose for 4 h at 4 °C. Bound protein was washed with 5 column volumes of 300 mM KOAc, 1 mM MgCl<sub>2</sub>, 10% glycerol, and 10 mM β-mercaptoethanol and eluted with the same buffer containing 100 mM imidazole. Eluted protein was bound to an anion exchange column (MonoQ 10/100 GL) and eluted with a gradient from 300 mM to 2 M KOAc. Pol I eluted at 1100 mM KOAc. Pol I was bound to a cation exchange column (MonoS 5/50 GL), and eluted at 490 mM KOAc, using a gradient from 200 mM to 2 M KOAc. Final purification over a Superose 6 HR 10/30 size-exclusion column equilibrated with 60 mM ammonium sulfate, 1 mM MgCl<sub>2</sub>, 5 mM HEPES at pH 7.8, 10 μM ZnCl<sub>2</sub> and 5 mM DTT, resulted in 0.5 mg of pure polymerase from 350–400 g of cells. Preparation protocol was established by Claus Kuhn and Jochen Gerber, for further details see (Gerber et al., 2008).

#### 3.2 Cryo-EM structure determinations

Experiments were done by Claus Kuhn, Sonja Baumli and Marco Gartmann (Kuhn et al., 2007).

#### 3.3 Modeling of the Pol I core

Experiments were performed by Claus Kuhn (Kuhn et al., 2007).

### 3.4 A14/43 preparation

The genes for A14 and A43 were cloned sequentially into vector pET28b (Novagen), resulting in a thrombin-cleavable N-terminal hexahistidine tag on A14. A ribosomal binding site was introduced before A43 to enable bicistronic expression. The deletion construct A14<sub>Δ53-77</sub>Δ113-137/43<sub>Δ173-209</sub>Δ252-326 (A14/43Δloop) was generated by PCR overlap extension (Higuchi et al., 1988). Only regions that were proteolytically sensitive and/or not predicted to form secondary structure elements were deleted (chapter II). A14/43 Δloop was expressed for 18 h at 18 °C in *E. coli* BL21 (DE3) RIL cells (Stratagene) in 4 l of LB medium (Maniatis et al., 1982). Cells were harvested by centrifugation, resuspended in 100 ml buffer A (100 mM NaCl, 20 mM Tris pH 8.0, 10 mM β-mercaptoethanol, 1 mM protease inhibitor mix: 1 mM PMSF, 1 mM benzamidine, 200 μM pepstatin, and 60 μM leupeptin), and lysed by sonication. After centrifugation, the supernatant was loaded onto a 3 ml Ni-NTA column equilibrated with buffer A. After washing, proteins were eluted with buffer A containing 100 mM imidazole. Eluted fractions were diluted three-fold with buffer A lacking NaCl, and incubated with thrombin (1U protease/1 mg protein) for 16 h at 4° C. A Mono Q 10/10 GL anion exchange column was equilibrated with buffer B (100 mM NaCl, 20 mM Tris pH 8.0, 5 mM DTT), and proteins were eluted with a linear gradient from 100 mM to 1 M NaCl. A14/43 eluted at 220 mM NaCl. After concentration, the sample was applied to a Superose 12 HR 10/300 gel filtration column (GE Healthcare) equilibrated with buffer B. Pooled peak fractions were concentrated to 10 mg/ml.

### 3.5 A14/43 crystal structure determination

Crystals of A14/43 Δloop were grown at 20 °C in hanging drops, using 22% PEG 3350 and 240 mM potassium acetate as reservoir solution. Crystals were harvested in reservoir solution, cryo-protected by stepwise transfer to reservoir solution containing 7-20% PEG 400, and flash-cooled by plunging into liquid nitrogen. Selenomethionine-labeled protein was prepared as described (Budisa et al., 1995; Meinhart et al., 2003) and crystallized at 20 °C with the use of microseeding and a reservoir solution of 18% PEG 3350 and 350 mM potassium acetate. Crystals reached a size of 250 μm x 120 μm x 80 μm and were cryo-preserved as above. A SAD experiment was performed at the Swiss Light Source (Table 5). Data were processed with XDS (Kabsch, 1993). Programs SHELXD/HKL2MAP (Pape and Schneider, 2004; Schneider and Sheldrick, 2002) detected nine selenium sites, which stemmed from three

A14/43 $\Delta$ loop complexes in the asymmetric unit. SHARP (La Fortelle and Bricogne, 1997) was used for refining heavy-atom positions, SAD phasing, and density modification. The model was built with COOT (Emsley and Cowtan, 2004), and refined with CNS (Brunger et al., 1998) to a free R-factor of 28.3% (Table 5). Ninety-eight and a half percent of the residues fall in allowed and additionally allowed regions of the Ramachandran plot, and no residues fall in disallowed regions (Laskowski et al., 1993).

### **3.6 Preparation of A49/34.5**

The genes for A49 and A34.5 were cloned sequentially into vector pET28b (Novagen), introducing a C-terminal hexahistidine tag on A49 and a second ribosomal binding site for bicistronic expression. The two subunits were coexpressed for 18 h at 18 °C in *E. coli* BL21 (DE3) RIL cells (Stratagene) in 4 l of LB medium. Cells were harvested by centrifugation, resuspended in 100 ml buffer C (300 mM NaCl, 50 mM Tris pH 7.5, 10 mM  $\beta$ -mercaptoethanol, and 1mM protease inhibitor mix), and lysed by sonication. After centrifugation the supernatant was loaded onto a 3 ml Ni-NTA column equilibrated with buffer C. The column was washed stepwise with 15 ml of buffer C containing 1 M NaCl and 15 ml of buffer C containing 30 mM imidazole. The A49/34.5 heterodimer was eluted with buffer C containing 100 mM imidazole. Eluted fractions were diluted 3-fold with 50 mM Tris pH 7.5 and 10 mM  $\beta$ -mercaptoethanol. A MonoS cation exchange column was equilibrated with buffer D (100 mM NaCl, 50 mM Tris pH 7.5, 5 mM DTT), and proteins were eluted with a linear gradient of 18 column volumes from 100 mM to 1 M NaCl. A49/34.5 eluted at 280 mM NaCl. The sample was applied to a Superose 12 HR 10/300 gel filtration column (GE Healthcare) equilibrated with buffer D. Pooled peak fractions were concentrated to 10 mg/ml.

### **3.7 A49/34.5 structure prediction**

The sequences of A49 and A34.5 were submitted to the HHpred server (Soding et al., 2005). As the hit with the highest score, HHpred predicted a structural similarity of the A49 N-terminal residues 52-102 to the N-terminal residues 99-150 of human RAP74 (p value = 0.0023). For A34.5 the hit with the third highest score was a similarity between the A34.5 residues 50-65 and residues 15-30 of human RAP30 (p value = 0.0003). Inspection of the predicted secondary structure elements in A49 and A34.5 revealed a similar arrangement of

strands as in the RAP74/RAP30 dimerization module crystal structure (Gaiser et al., 2000) (PDB 1F3U) except that strands  $\beta$ 4 and  $\beta$ 5 were apparently lacking in A49, and no secondary structure corresponding to strands  $\beta$ 6 and  $\beta$ 7 of RAP30 was predicted in A34.5. The few residues conserved between A49 and RAP74 and between A34.5 and RAP30 are generally part of the hydrophobic heterodimer interface (Figs. 21 and 22).

### **3.8 Yeast genetic manipulations and 6-azauracil phenotyping**

Experiments were performed by Claus Kuhn (Kuhn et al., 2007).

### **3.9 Preparation of Pol I variants**

Pol I  $\Delta$ A49/34.5 was prepared by dialyzing Pol I-containing fractions after cation-exchange chromatography overnight against 2 M urea, 50 mM ammonium sulfate, 1 mM magnesium chloride, 20 mM HEPES pH 7.8, 10% glycerol, and 5 mM DTT. A49/34.5 was separated from Pol I  $\Delta$ A49/34.5 by subsequent anion-exchange chromatography, applying a gradient from 50 mM to 1 M ammonium sulfate (Mono Q 5/50 GL, GE Healthcare). Pol I  $\Delta$ A49/34.5 eluted at 250 mM ammonium sulfate and was further purified by size-exclusion chromatography (Superose 6 HR10/300) with buffer E (100 mM ammonium sulfate, 1 mM  $MgCl_2$ , 20 mM HEPES pH 7.8, 5% glycerol, 5 mM DTT). Pooled fractions were concentrated to 0.5 mg/ml.

Pol I variant A12.2  $\Delta$ C was purified by Stefan Jennebach and Claus Kuhn (Kuhn et al., 2007).

### **3.10 RNA extension assays**

Four picomoles of polymerase was incubated for 30 min at 20 °C with 2 pmol of a preannealed minimal nucleic-acid scaffold (template DNA: 3'-GCTCAGCCTGGTCCGCATGTGTCAGTC-5'; nontemplate DNA: 5'-CACACAGTCAG-3'; RNA: 5'-FAM-UGCAUAAAGACCAGGC-3'). For RNA elongation, complexes were incubated in the presence of 1 mM NTPs at 28 °C for 20 min in transcription buffer (60 mM ammonium sulfate, 20 mM HEPES pH 7.6, 8 mM magnesium sulfate, 10 mM zinc chloride, 10% glycerol, and 10 mM DTT). Reactions were stopped by addition of an equal volume 2x loading buffer (8 M urea, 2x TBE) and incubation for 5 min at 95 °C. The FAM-labeled RNA extension products were separated by denaturing gel electrophoresis (0.5 pmol RNA per lane,

0.4 mm 15–20% polyacrylamide gels containing 8M urea, (50 °C) and visualized with a Typhoon 9400 phosphoimager (GE Healthcare). For RNA extension assays with a complementary bubble (Kireeva et al., 2000), 6 pmol Pol I or Pol I  $\Delta$ A49/34.5 were incubated for 15 min at 20 °C with 3 pmol of a preannealed template DNA-RNA scaffold (template DNA: 3'-TGCGCACCACGCTTACTGGTCCGTTTCGCCTGTCCTCGACCA-5'; RNA: 5'-FAM-UGCAUUUCGACCAGGC-3'), followed by incubation with a 5-fold molar excess of nontemplate DNA (15 pmol; 5'-TTTTTACGCGTGGTGCGAATGACCAGGCAAGCGGACAGGAGCTGGT-3') for 15 min at 25 °C. Complexes were incubated in the presence of 1 mM NTPs at 28 °C for 1 and 5 min in transcription buffer. Reactions were stopped and analyzed by gel electrophoresis as above.

### **3.11 RNA cleavage assays**

Complexes of complete Pol I and polymerase variants were formed in transcription buffer with a nucleic acid scaffold that comprised an RNA with a 6-FAM fluorescent label at its 5'-end and a three-nucleotide non-complementary overhang at its 3'-end (template DNA: 3'-TTACTGGTCCTTTTTTCATGAACTCGA-5'; non-template DNA: 5'-TAAGTACTTGAGCT-3'; RNA: 5'-FAM-UGCAUUUCGACCAGGACCGU-3', overhanging nucleotides underlined). Samples were incubated in transcription buffer up to 30 min at 28 °C, and reaction products were analyzed as above.

### **3.12 Figure preparation**

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

### **3.13 Accession numbers**

The Pol I EM map was deposited in the European Bioinformatics Institute EM data bank under accession code EMD-1435. Coordinates and structure factors of the A14/43 crystal structure have been deposited with the protein data bank under accession code 2RF4.

## **Chapter IV:**

# **RNA polymerase I contains a TFIIF-related DNA-binding subcomplex**

## 1 Introduction

The structure of Pol II has been studied in detail (Cramer et al., 2008), but structural information is still limited for Pol I (chapter III) and Pol III (Fernandez-Tornero et al., 2007; Jasiak et al., 2006). All three polymerases contain a 10-subunit catalytic core of shared or homologous subunits, and a peripheral heterodimeric subcomplex of conserved structure, called A14/43, Rpb4/7, and C17/25 in Pol I, II, and III, respectively (chapter III, 2.3) (Armache et al., 2005; Jasiak et al., 2006). In addition to these 12 related subunits, Pol I contains the specific subcomplex A49/34.5 (chapter III, 2.6 and 2.7), and Pol III contains the specific subcomplexes C37/53 (Kassavetis et al., 2010; Landrieux et al., 2006) and C82/34/31 (Wang and Roeder, 1997), which contains the stable C82/34 dimer (Lorenzen et al., 2007).

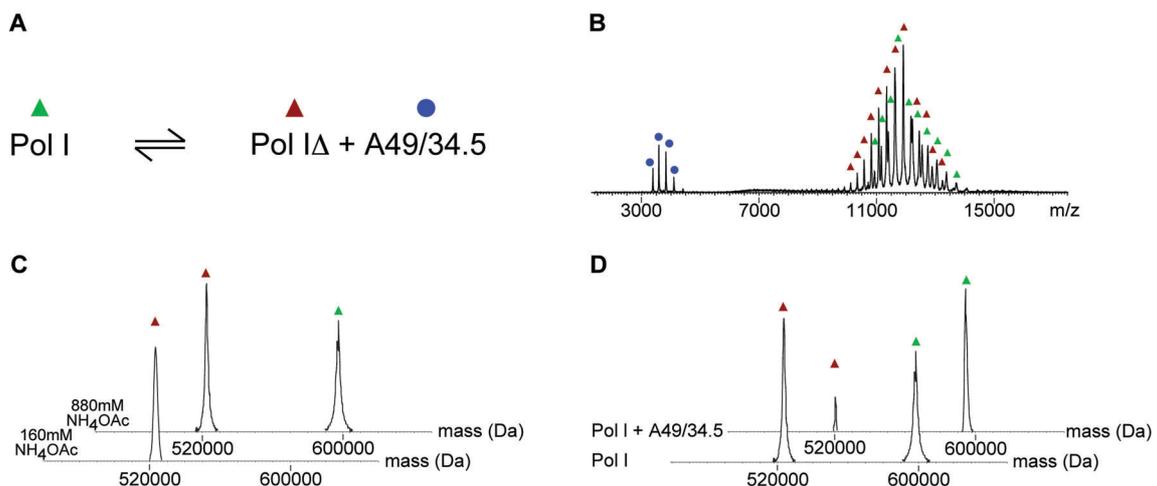
Since all three polymerases share a similar core structure and active center, their mechanism of DNA-dependent RNA elongation is similar and well understood at a structural level (Brueckner et al., 2009). In contrast, the structural basis for promoter-specific transcription initiation remains unclear. To recognize different promoters, the polymerases use different initiation factors. Pol II uses the initiation factors TFIIA, -B, -D, -E, -F, and -H, which are apparently not related to Pol I initiation factors (Moss et al., 2007).

Here we show that the Pol I subcomplex A49/34.5 consists of two distinct domains, and provide the crystal structures of these domains. We demonstrate that the fold of the combined N-terminal dimerization module is similar to that of TFIIF, whereas the A49 C-terminal domain contains a novel tWH domain that has a previously unknown DNA-binding activity and is predicted in TFIIE. Our results complete the structural information on Pol I subunits, demonstrate that Pol I, in contrast to Pol II, contains a DNA-binding surface domain, and show how the three RNA polymerases are evolutionarily related.

## 2 Results

### 2.1 Pol I subcomplex A49/34.5 forms a heterodimer with two distinct domains

It was previously shown that the Pol I subunits A49 and A34.5 interact and can dissociate from Pol I (chapter III, 2.6) (Huet et al., 1976). To investigate whether these two subunits form a heterodimer we used native mass spectrometry (Benesch et al., 2007; Heck, 2008). We purified the complete Pol I as described (chapter III, 3.1), and obtained a mass of 593 kDa (Fig. 28). This agrees well with the theoretical molecular weight of 592 kDa for Pol I and indicates that one copy of each Pol I subunit is present, including A49 and A34.5. Pol I preparations also gave rise to peaks with masses of 520 and 74 kDa, which correspond to the theoretical masses of Pol I lacking the A49/34.5 heterodimer (Pol I $\Delta$ , 518 kDa) and the free A49/34.5 heterodimer (74 kDa), respectively (Fig. 28A, B).

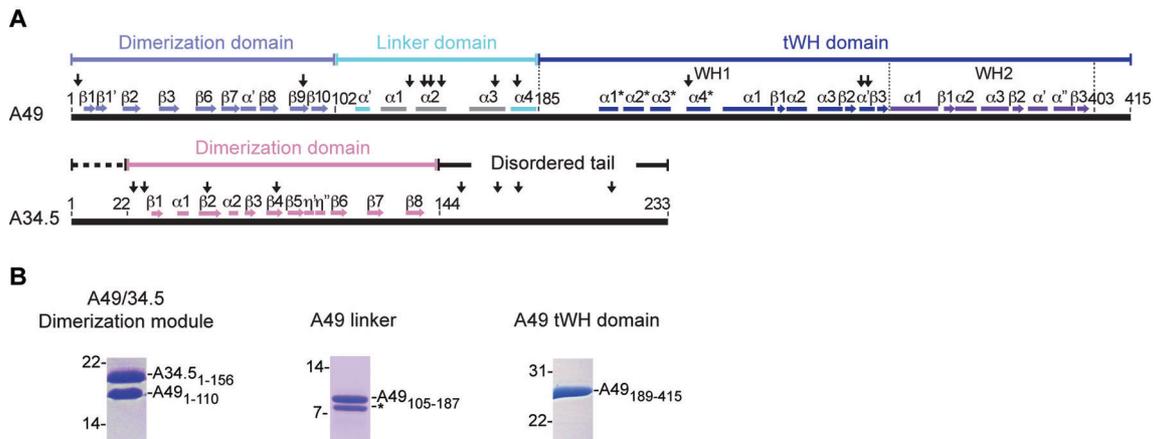


**Figure 28. Native mass spectrometry of Pol I.**

(A) Dissociation of the A49/34.5 subcomplex from the complete, 14-subunit Pol I generates the 12-subunit complex Pol I $\Delta$ . (B) Native mass spectrum of Pol I, electrosprayed from aqueous ammonium acetate (880 mM). In the high  $m/z$  region two different charge distributions are present, corresponding to Pol I (green triangles), and Pol I $\Delta$  (brown triangles). The ‘free’ A49/34.5 heterodimer is also detected in the low  $m/z$  region (blue circles). (C) Pol I stability depends on ionic strength. Two transformed zero-charge mass spectra are shown, recorded at 880 mM and 160 mM ammonium acetate. (D) A49/34.5 can be associated with Pol I $\Delta$ . The lower spectrum was recorded with a sample containing Pol I in 880 mM ammonium acetate, whereas the upper spectrum was obtained after pre-incubation of the sample with a five-fold molar excess of recombinant A49/34.5.

The stability of Pol I was dependent on the ionic strength. At 160 mM ammonium acetate, Pol I was largely dissociated into Pol IA and A49/34.5, whereas at 880 mM ammonium acetate, less dissociation was observed (Fig. 28C). When an excess of recombinant A49/34.5 was added, higher relative amounts of complete Pol I were detected (Fig. 28D). These results show that Pol I contains a single copy of each of its 14 subunits and that the A49/34.5 subcomplex forms a stable heterodimer.

To probe for flexible protein regions and to delineate domains, we subjected recombinant A49/34.5 (chapter III, 3.6) to limited proteolysis. The C-terminal tail of A34.5 and the central region of A49 were particularly sensitive to proteases (Fig. 29A). Consistent with this finding, these regions are predicted to be largely unstructured (Biegert et al., 2006). Based on these results we designed minimal variants for subunit co-expression and performed several rounds of solubility studies and variant optimization (related to chapter II). This revealed that the A49/34.5 heterodimer forms two structured domains connected by a flexible linker. The N-terminal regions of both subunits constitute a dimerization module, whereas the C-terminal region of A49 forms an independently folded domain (Fig. 29A, B).

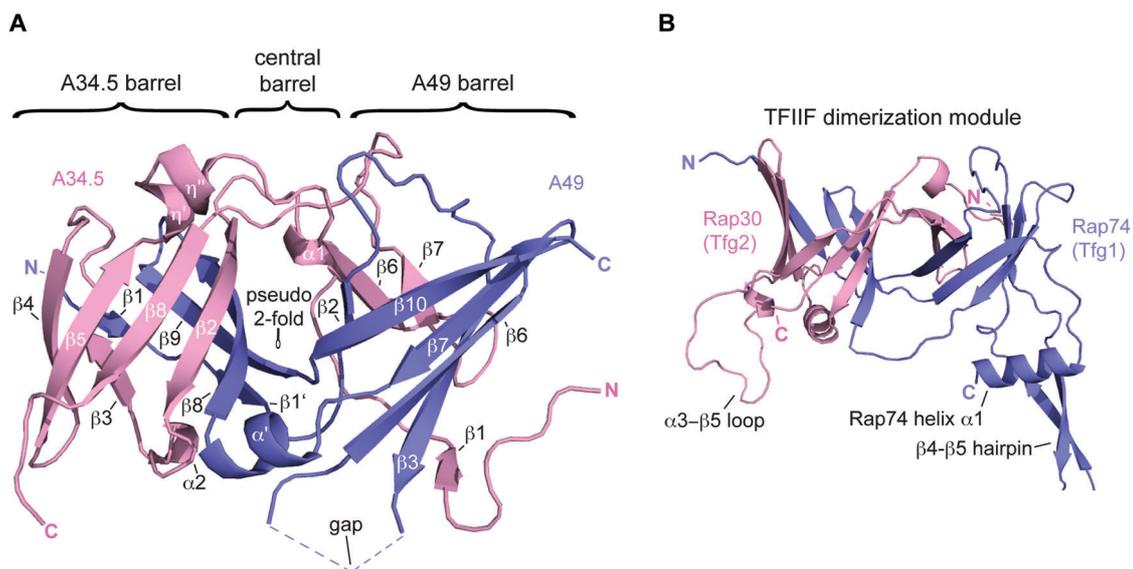


**Figure 29. A49/34.5 domain organization.**

(A) Schematic representation of A49 and A34.5. The A49 N-terminal dimerization domain, interdomain linker, and C-terminal tandem winged helix (tWH) domain are shown in light blue, cyan, and dark blue, respectively. The A34.5 dimerization domain and C-terminal tail are in magenta and black, respectively. Secondary structure elements are indicated as bars ( $\alpha$ -helices) or arrows ( $\beta$ -strands). Dashed vertical lines delineate A49 tWH subdomains WH1 and WH2. Proteolytic cleavage sites are marked with vertical arrows. (B) Preparation of recombinant A49/34.5 domains. SDS-PAGE analysis of the dimerization module, A49 linker and A49 tWH domain. An asterisk indicates an N-terminal degradation product.

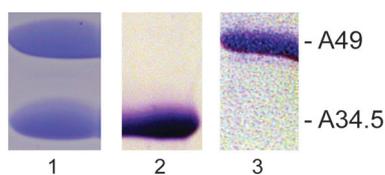
## 2.2 The A49/34.5 dimerization module resembles TFIIIF

We prepared the A49/34.5 dimerization module from various species, and obtained crystals for the *Candida glabrata* variant A49<sub>1-99</sub>/A34.5<sub>25-143</sub>. The structure was determined by single wavelength anomalous diffraction at 2.9 Å resolution (Table 6). It revealed three interconnected β-barrels (Fig. 30A). A total of 17 interwoven β-strands and three α-helices contribute to a conserved hydrophobic core. Consistent with the structure, subunit dimerization is impaired by mutation of key core residues (chapter III, Fig. 23) or deletion of A49 strands β1, β1' and β2 (Fig. 31, lane 2). The structure resembles that of the TFIIIF Rap74/30 dimerization module (Gaiser et al., 2000), with a DALI Z-score of 4.3 (Holm and Sander, 1995) and an RMS deviation in 117 Cα atom positions of 2.6 Å. It however lacks three elements of the Rap74/30 module, the β4-β5 hairpin, the Rap74 C-terminal helix α1, and the Rap30 loop α3-β5 (Figs. 30 and 32).



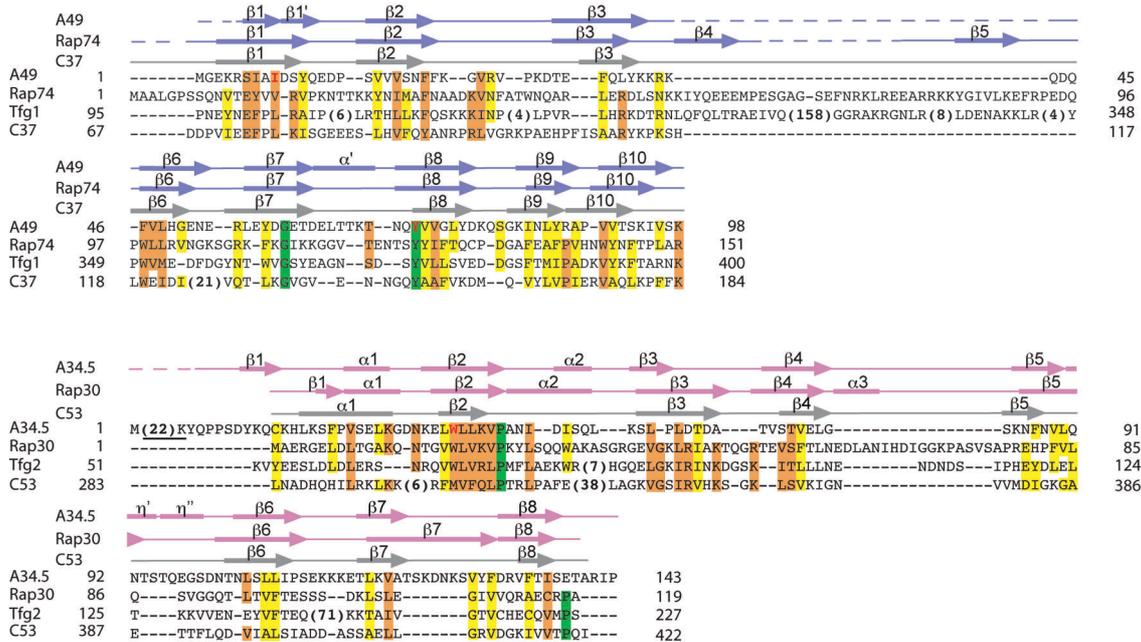
**Figure 30. A49/34.5 resembles a TFIIIF-like dimerization module**

(A) Ribbon model of the dimerization module crystal structure, which is part of a hetero-tetramer (Fig. 34). A49 and A34.5 are depicted in light blue and magenta, respectively. A pseudo 2-fold axis is indicated. Figures were prepared with PyMOL (DeLano Scientific). (B) Structure of the TFIIIF dimerization module (Gaiser et al., 2000). Rap74 and Rap30 are depicted in light blue and magenta, respectively. TFIIIF-specific features are indicated.



**Figure 31. Preparation of A49/34.5 variants**

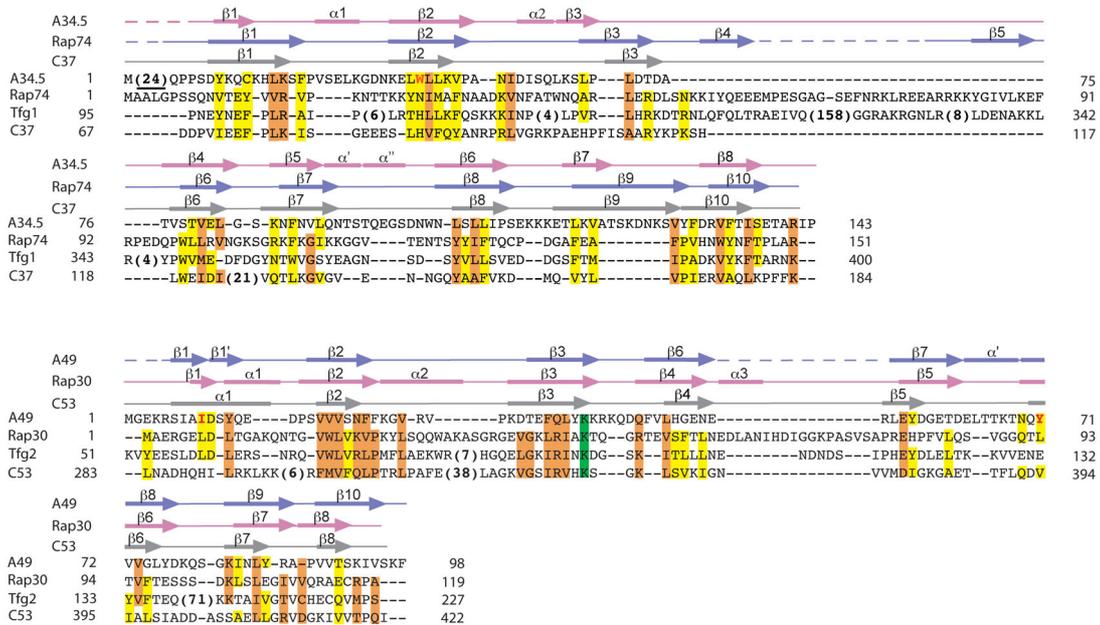
SDS-PAGE analysis of wildtype A49/34.5 (lane 1), an A49/34.5 variant lacking A49 residues 2-29 (lane 2, hexahistidine tag on A34.5) and individually expressed A49 (lane 3, thanks to Yvonne Krauss).



**Figure 32. Alignment of dimerization module sequences from Pol I, II and III**

Structure-based alignments of sequences of *C. glabrata* A49 and A34.5 with their counterparts in Pol II (*H. sapiens* Rap74 and Rap30, *S. cerevisiae* Tfg1 and Tfg2), and in Pol III (*S. cerevisiae* C37 and C53). Secondary structure elements observed in the A49/34.5 and TFIIF crystal structures are shown (bars,  $\alpha$ -helices; arrows,  $\beta$ -strands; lines, loops; dashed lines, disordered regions). For C37 and C53, predicted secondary structure elements are in grey. As the hit with the highest score, HHpred (Biegert et al., 2006; Soding et al., 2005) identified similarity of the *S. cerevisiae* C53 C-terminal residues 300-421 to the N-terminal residues 13-118 of human Rap30 (p-value = 0.0006). For *S. pombe* C37 the hit with the highest score was detected between C37 residues 26-145 and Rap74 residues 10-151 (p-value = 0.0037). The  $\beta 4$ - $\beta 5$  hairpin is apparently lacking in C37. Conserved residues are highlighted according to decreasing conservation from green, through orange, to yellow. Residues located in the hydrophobic heterodimer interface are generally conserved between A49/34.5, Rap74/30 and C37/53. Insertions are marked with numbers in the sequence. Mutated residues that abolish proper folding (chapter III, Fig. 23) are highlighted in red. Residues not present in the crystallized variant are underlined.

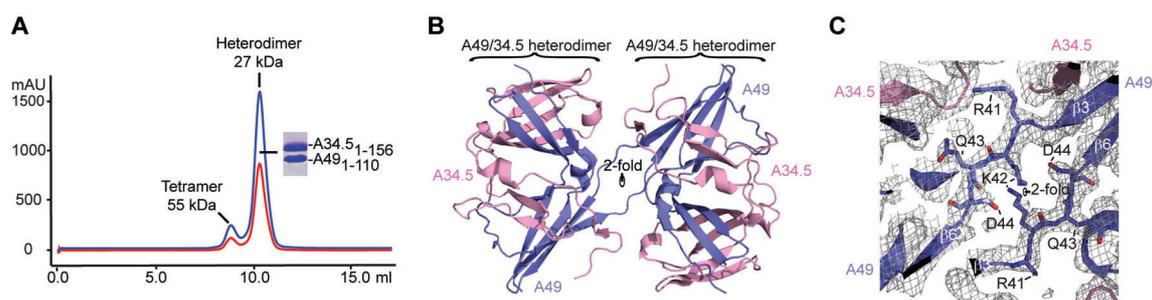
Because of a pseudo-twofold axis (Fig. 30A), the A49/34.5 dimerization module may alternatively be superimposed with the TFIIF Rap74/30 module such that subunits A49 and A34.5 overlay with Rap30 and Rap74, respectively (Fig. 33).



**Figure 33. Alternative alignment of dimerization domain sequences from Pol I, II and III.**

Structure-based alignment of amino acid sequences of *C. glabrata* A34.5 and A49 with their alternative counterparts in Pol II (*H. sapiens* Rap74 and Rap30, *S. cerevisiae* Tfg1 and Tfg2), and their alternative Pol III counterparts *S. cerevisiae* C37 and C53, respectively. All other labels are as in Figure 32.

The similarity of the A49 and A34.5 dimerization domains is consistent with the formation of soluble homodimers after their expression in isolation (Fig. 31). Although native A49/34.5 is heterodimeric (Fig. 28), tetramers are observed in the crystal lattice (Fig. 34B, C). Tetramerization occurs via an unusual exchange of  $\beta$ -strands, is an artifact of the high protein concentrations required for crystallization (Fig. 34A), and may explain previous observations of TFIIF tetramers (Flores et al., 1990).



**Figure 34. A49/34.5 tetramer formation.**

(A) Size-exclusion chromatography of the A49/34.5 dimerization module. Absorption at 280 nm and 260 nm is shown in blue and red, respectively. Molecular masses were determined by static light scattering. The protein was visualized by SDS-PAGE. (B) Ribbon model of the A49/34.5 tetramer, composed of two A49/34.5 heterodimers. A49 is depicted in light blue and A34.5 in magenta. A pseudo 2-fold symmetry axis is indicated. (C)  $\beta$ -strand exchange in the tetramer observed in crystals. View as in B. Secondary structure elements and A49 residues are indicated. A  $2F_o - F_c$  map is shown as a grey mesh and contoured at 1.0 sigma.

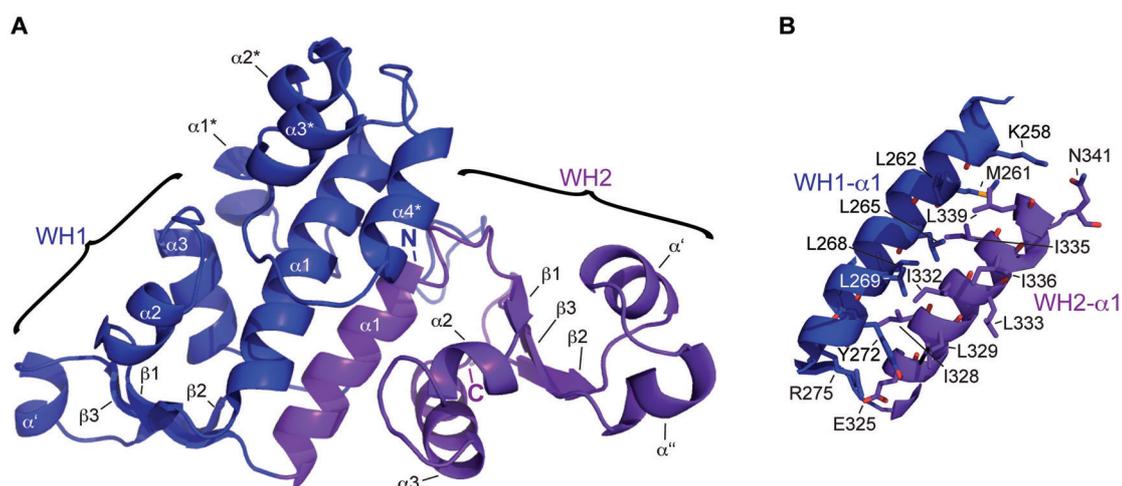
**Table 6.** Diffraction data and refinement statistics

| A49/34.5 variant            | Dim. Module<br>A49 <sub>1-99</sub> / 34.5 <sub>25-143</sub> | Dim. Module (+linker helix $\alpha'$ )<br>A49 <sub>1-119</sub> / 34.5 <sub>25-143</sub> |
|-----------------------------|---|---|
| <b>Data collection</b>      |   |   |
| Space group                 | P2 <sub>1</sub>   | C222 <sub>1</sub>   |
| Unit cell axes (Å)          | 69.3<br>113.9<br>118.0                                      | 108.9<br>221.8<br>129.2   |
| Unit cell $\beta$ angle (°) | 102.9   | 90  |
| Wavelength (Å)              | 0.9792  | 0.9330  |
| Resolution range (Å)        | 50-2.9 (3.0-2.9) <sup>2</sup>                               | 80-4.0 (4.2-4.0) <sup>2</sup>   |
| Unique reflections          | 88,325 <sup>1</sup> (8,594) <sup>2</sup>                    | 13,572 <sup>3</sup> (1,801) <sup>2</sup>  |
| Completeness (%)            | 97.8 (97.6) <sup>2</sup>                                    | 99.8 (99.5) <sup>2</sup>  |
| Redundancy                  | 4.9 (5.1) <sup>2</sup>                                      | 8.0 (10.7) <sup>2</sup>   |
| Mosaicity (°)               | 0.19  | 0.34  |
| R <sub>sym</sub> (%)        | 3.7 (65.1) <sup>2</sup>                                     | 20.7 (79.2) <sup>2</sup>  |
| I/ $\sigma$ (I)             | 24.2 (3.3) <sup>2</sup>                                     | 8.9 (3.7) <sup>2</sup>  |
| <b>Refinement</b>           |   |   |
| Nonhydrogen atoms           | 13,794  | 6,539   |
| RMSD bonds                  | 0.010   | 0.010   |
| RMSD angles                 | 1.28  | 1.25  |
| R <sub>cryst</sub> (%)      | 19.4  | 28.7  |
| R <sub>free</sub> (%)       | 24.3  | 36.1  |
| Preferred (%)               | 95.2  | 83.0  |
| Allowed (%)                 | 4.3   | 11.1  |
| Disallowed (%)              | 0.5   | 5.9   |

<sup>1</sup> Friedel pairs not merged. <sup>2</sup> Values in parentheses are for highest resolution shell. <sup>3</sup> Friedel pairs merged

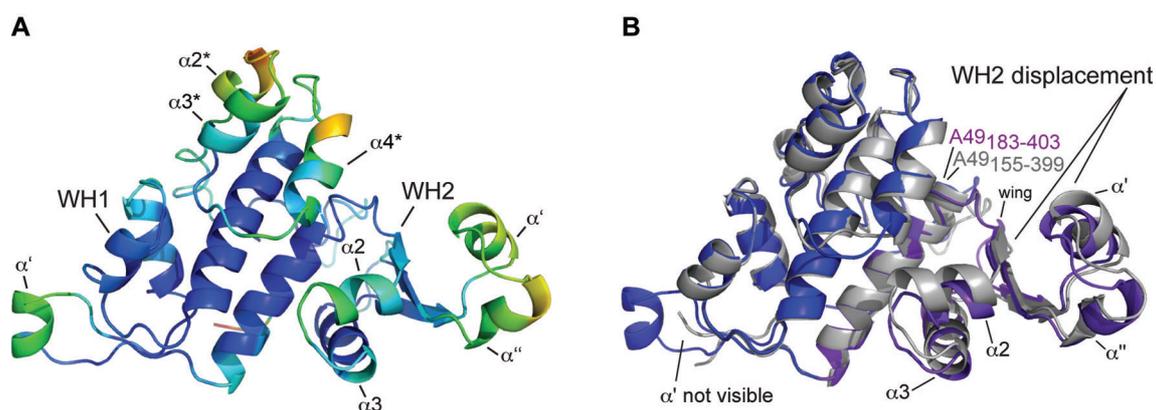
### 2.3 A49 contains a novel tandem winged helix domain

The C-terminal domain of *S. cerevisiae* A49 could also be crystallized (Experimental procedures) and its structure was determined by multi-wavelength anomalous diffraction at 2.0 Å resolution (Table 7). The structure revealed a single domain (Fig. 35A) with a fold that is not found in the database (Berman et al., 2000). The core of the domain consists of two tightly packed subdomains with a winged helix-turn-helix (winged helix, WH) fold (WH1, residues 254-321; WH2, residues 322-403) (Fig. 35A). The WH fold is present in transcription factors and nucleic acid-binding proteins (Brennan, 1993; Gajiwala and Burley, 2000; Kenney, 2002), and consists of three  $\alpha$ -helices and three  $\beta$ -strands, in the order  $\alpha_1$ ,  $\beta_1$ ,  $\alpha_2$ ,  $\alpha_3$ ,  $\beta_2$ , and  $\beta_3$  (Figs. 29A and 35A). Subdomains WH1 and WH2 additionally contain one or two short  $\alpha$ -helices, respectively, inserted in the  $\beta_2$ - $\beta_3$  loop (Fig. 35A). In WH1, helix  $\alpha_1$  is extended and surrounded with four additional N-terminal helices ( $\alpha_1^*$ - $\alpha_4^*$ , Fig. 35A). This N-terminal helical bundle and the  $\beta_2$ - $\beta_3$  insertions are flexible, since they have higher B-factors (Fig. 36A), and are partially disordered in an alternative crystal form (Fig. 36B and Table 7). The WH subdomains form an interface with conserved hydrophobic residues on their  $\alpha_1$ -helices that pack in an antiparallel fashion (Figs. 35B and 37). Because the WH subdomains are packed into a single domain, we refer to this structure as tandem winged helix (tWH) domain.



**Figure 35. A novel tandem winged helix domain structure in A49**

(A) Ribbon diagram of the A49 tWH domain crystal structure. Subdomains WH1 and WH2 are displayed in blue and purple, respectively. (B) The WH1-WH2 subdomain interface is formed by conserved residues.



**Figure 36. A49 tandem winged helix domain contains flexible regions**

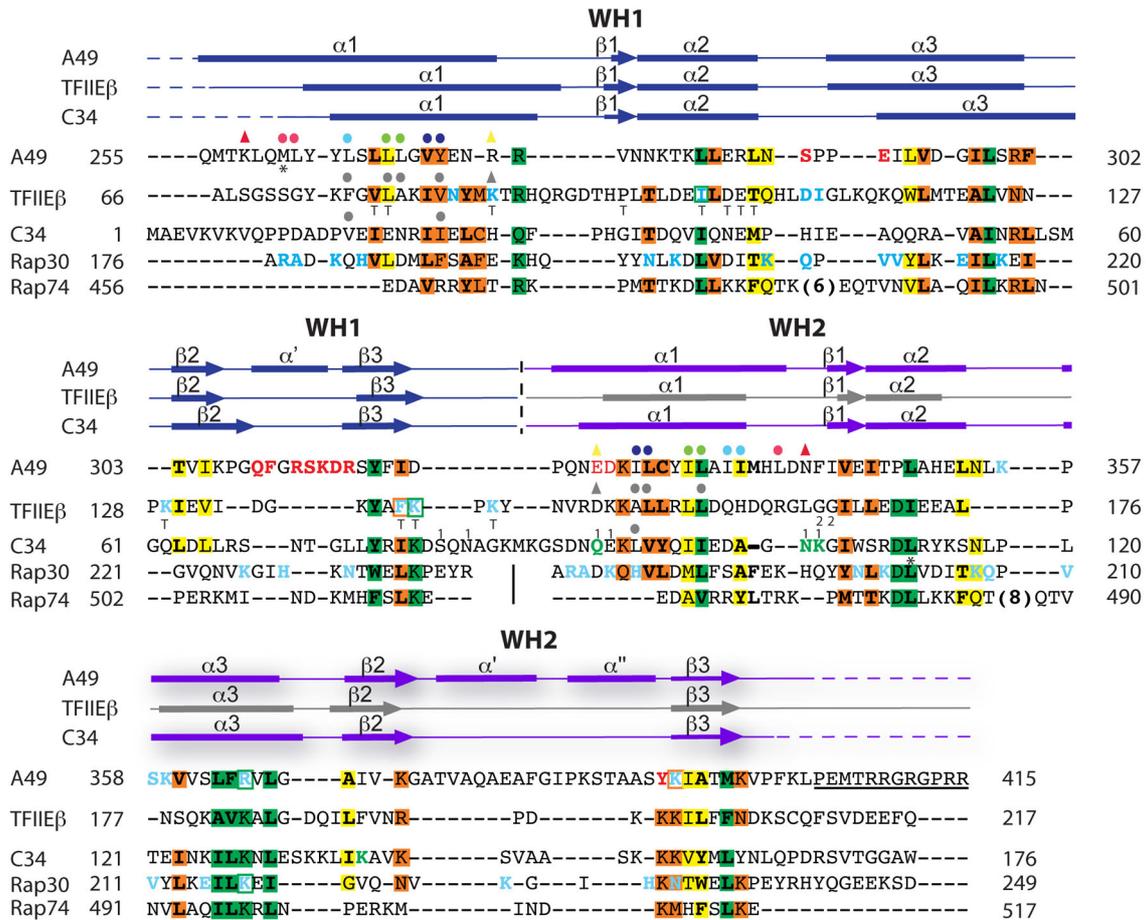
(A) Structure of the A49 tWH domain, colored according to B-factors. B-factor distribution is shown in rainbow colors ranging from blue (low B-factors) to red (high B-factors). Secondary structure elements with high B-factors are indicated. (B) Comparison of two related A49 tWH structures. A49 tWH<sub>183-403</sub> is colored in blue (WH1) and purple (WH2), while A49 tWH<sub>155-399</sub> is in grey. WH1 helix  $\alpha'$  is not ordered in the structure of A49 tWH<sub>155-399</sub>, and parts of WH2 are shifted.

**Table 7. Diffraction data and refinement statistics**

| A49/34.5 variant            | tWH domain A49 <sub>171-403</sub> |  |  | tWH domain A49 <sub>155-399</sub>        |
|-----------------------------|-----------------------------------|--|--|--|
|                             | Peak                              | Inflection                               | Remote                                   |  |
| <b>Data collection</b>      |                                   |  |  |  |
| Space group                 | P2 <sub>1</sub>                   | P2 <sub>1</sub>                          | P2 <sub>1</sub>                          | P4                                       |
| Unit cell axes (Å)          | 85.1                              | 78.1                                     | 100.7                                    | 96.7                                     |
|                             |                                   |  | 113.4                                    | 96.7                                     |
| Unit cell $\beta$ angle (°) |                                   |  |  | 54.6                                     |
| Wavelength (Å)              | 0.9790                            | 0.9790                                   | 0.9180                                   | 90                                       |
| Resolution range (Å)        | 80-2.0 (2.08-2.0) <sup>2</sup>    | 50-2.8 (2.9-2.8) <sup>2</sup>            | 50-2.5 (2.6-2.5) <sup>2</sup>            | 80-2.35 (2.45-2.35) <sup>2</sup>         |
| Unique reflections          | 159,335 <sup>1</sup> (17,800)     | 58,751 <sup>1</sup> (5,918) <sup>2</sup> | 82,833 <sup>1</sup> (9,221) <sup>2</sup> | 21,259 <sup>3</sup> (2,466) <sup>2</sup> |
| Completeness (%)            | 99.0 (99.8) <sup>2</sup>          | 99.2 (99.5) <sup>2</sup>                 | 99.3 (99.7) <sup>2</sup>                 | 99.8 (99.8) <sup>2</sup>                 |
| Redundancy                  | 3.6 (3.8) <sup>2</sup>            | 3.8 (3.8) <sup>2</sup>                   | 3.8 (3.8) <sup>2</sup>                   | 11.7 (12.5) <sup>2</sup>                 |
| Mosaicity (°)               | 0.16                              | 0.17                                     | 0.30                                     | 0.09                                     |
| R <sub>sym</sub> (%)        | 6.6 (60.6) <sup>2</sup>           | 8.1 (69.7) <sup>2</sup>                  | 7.7 (75.7) <sup>2</sup>                  | 8.0 (59.7) <sup>2</sup>                  |
| I/ $\sigma$ (I)             | 16.8 (3.4) <sup>2</sup>           | 16.8 (3.4) <sup>2</sup>                  | 18.3 (3.5) <sup>2</sup>                  | 27.5 (6.3) <sup>2</sup>                  |
| <b>Refinement</b>           |                                   |  |  |  |
| Nonhydrogen atoms           |                                   | 9,698                                    |  | 3,587                                    |
| RMSD bonds                  |                                   | 0.010                                    |  | 0.010                                    |
| RMSD angles                 |                                   | 1.04                                     |  | 1.08                                     |
| R <sub>cryst</sub> (%)      |                                   | 18.9                                     |  | 18.8                                     |
| R <sub>free</sub> (%)       |                                   | 22.4                                     |  | 22.5                                     |
| Preferred (%)               |                                   | 98.5                                     |  | 99.0                                     |
| Allowed (%)                 |                                   | 1.5                                      |  | 1.0                                      |
| Disallowed (%)              |                                   | 0.0                                      |  | 0.0                                      |

<sup>1</sup> Friedel pairs not merged. <sup>2</sup> Values in parentheses are for highest resolution shell. <sup>3</sup> Friedel pairs merged.

<sup>4</sup> Ramachandron plot statistics from PROCHECK (Laskowski et al., 1993).



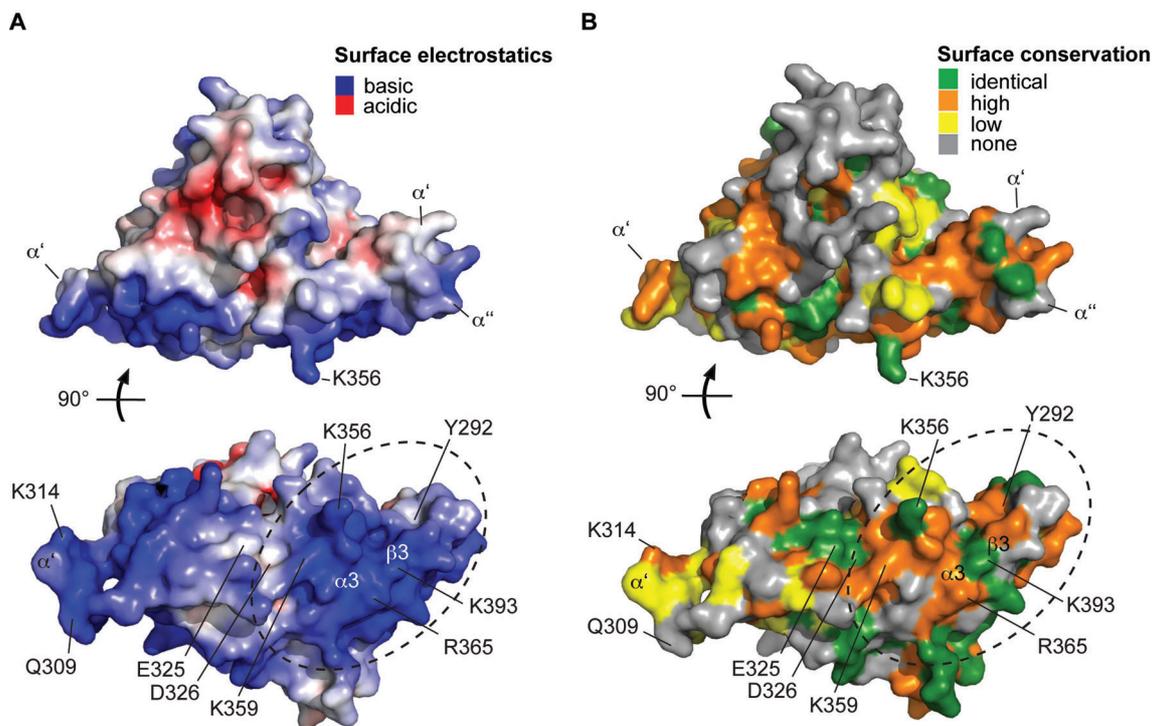
**Figure 37. Homologous WH domains in Pol I/III subunits and Pol II-associated factors**

Structure-based alignment of amino acid sequences of the *S. cerevisiae* A49 tWH domain (top) with their proposed human Pol II and Pol III counterparts TFIIIE-β (PDB 1D8K (Okuda et al., 2004)) and C34 (PDB 2DK8 and 2DK5), respectively. Related individual WH domains in human TFIIF proteins Rap30 (PDB 1BBY (Groft et al., 1998)) and Rap74 (PDB 1I27 (Kamada et al., 2001)) are aligned for both WH1 and WH2 of A49.

For WH2 of TFIIIE-β, predicted secondary structure elements are depicted in grey, when aligned by HHpred (p-value < 0.0001) (Soding et al., 2005) and predicted to be present by secondary structure propensity (Biegert et al., 2006). Residues predicted in the hydrophobic cores of individual WH domains are highlighted in black. Residues in Rap30 (Groft et al., 1998), TFIIIE-β (Tanaka et al., 2009) and A49 (this study) that interact with dsDNA are in cyan. Additionally mutated A49 residues are in red. For TFIIIE-β, mutated residues affecting *in vitro* transcription (Tanaka et al., 2009) are indicated with a 'T' below the sequence. For C34, mutated residues are colored in dark green when displaying a cryo-sensitive phenotype, or are indicated with numbers for a lethal hexa- (1) or double-mutant (2), respectively (Brun et al., 1997). A49 tWH subdomain interface residues are indicated with dots (hydrophobic) or triangles (hydrophilic), with matched colors indicating interactions. For TFIIIE-β and C34, residues predicted to contribute to a putative tWH interface are indicated with grey dots. For TFIIIE-β, grey triangles indicate two residues that are predicted to form an interface salt bridge. A residue in A49 and a residue in C34 that deviate in the structures from database entry sequences are marked with an asterisk below the protein sequence. All other labels are as in Figure 32.

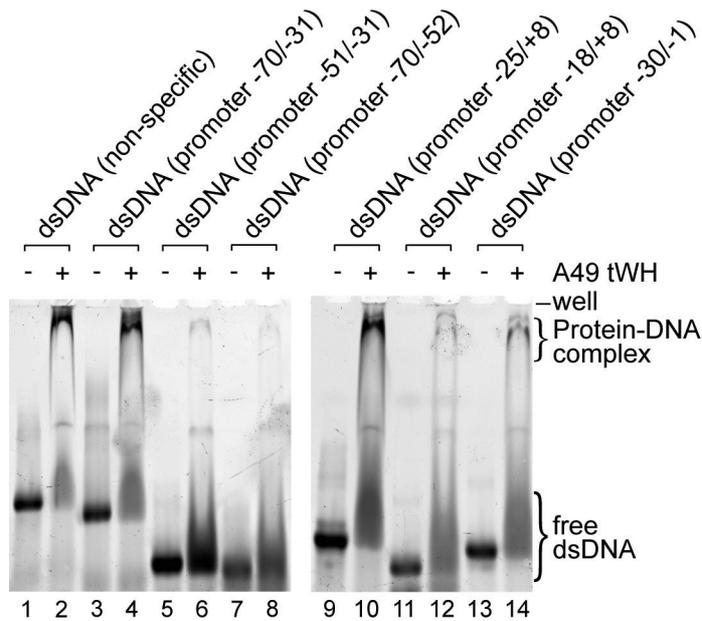
## 2.4 The tWH domain binds double-stranded DNA with a conserved surface

The tWH domain contains an extended positively charged surface (Fig. 38A) suggesting it binds nucleic acids. We investigated this using an electrophoretic mobility shift assay (EMSA, Experimental procedures 4.5). A double-stranded (ds) DNA of random sequence caused a shifted band indicating a tWH-dsDNA complex (Fig. 39, lanes 1-2). Similar results were obtained with dsDNA corresponding to parts of the Pol I promoter (Fig. 39, lanes 3-4, 9-10). When duplexes of less than 30 base pairs were used, dsDNA binding was impaired (Fig. 39, lanes 5-8, 11-14). The dimerization module did not bind DNA (Fig. 41, lane 2), providing a negative control. Consistent with these findings, a DNA affinity of A49 was previously noticed, but not investigated (Liljelund et al., 1992). These results unravel a DNA-binding function of A49/34.5 located in its tWH domain.



**Figure 38. A49 tWH domain contains a conserved basic surface**

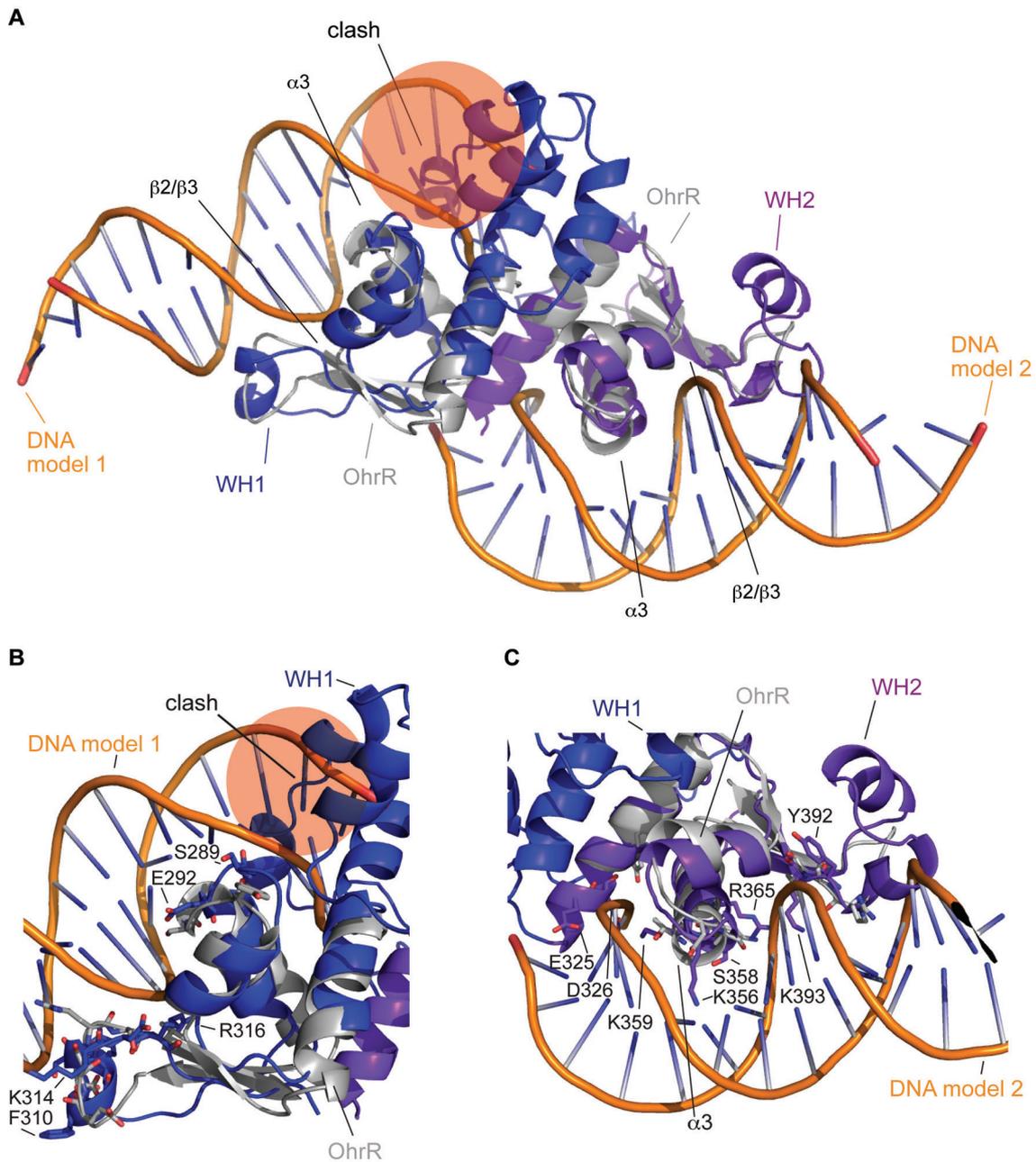
(A) Surface charges. Red and blue color indicates negative and positive charges, respectively, as calculated with APBS (Baker et al., 2001). A dashed circle indicates a positively charged conserved region. The top view is as in Figure 35A, the bottom view is rotated around a horizontal axis by 90 degrees. (B) Surface conservation. Decreasing conservation among seven different *Saccharomycotinae* species is indicated in green, orange, yellow and grey. The views are as in A.



**Figure 39. A49 tWH domain binds DNA**

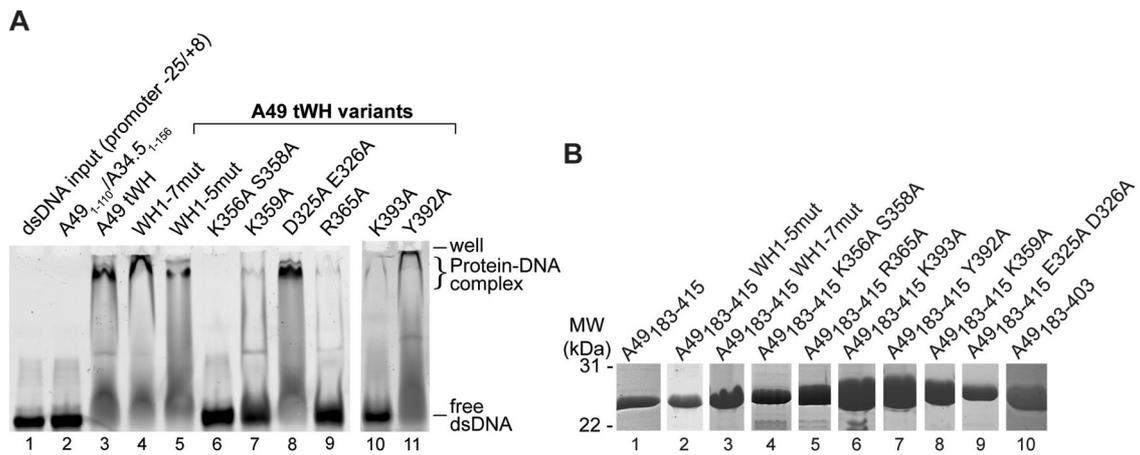
EMSA analysis of dsDNA binding. dsDNAs included a 47 base pair (bp) DNA of unrelated sequence (non-specific) and regions of the Pol I promoter (Choe et al., 1992; Kulkens et al., 1991; Musters et al., 1989) as indicated by positions relative to the transcription start site +1. The latter included extended domain 2 of upstream promoter element (-70 to -31) and parts of it (-70 to -52, -51 to -31), and the core element domain 1 (-25 to +8) and parts of it (-18 to +8, -30 to -1).

To predict the DNA-binding surface of the tWH domain, we superimposed the WH domain of the transcription factor OhrR-DNA complex (Hong et al., 2005) with both A49 WH subdomains. Superposition with the A49 WH1 subdomain resulted in a clash between DNA and protein, whereas superposition with the WH2 subdomain did not, suggesting that  $\alpha 3$ ,  $\beta 2$  and  $\beta 3$  of WH2 bind DNA (Fig. 40A). The putative DNA-binding surface is positively charged and conserved (Fig. 38). Consistent with these results, multiple point mutations on WH1 (Fig. 40B) did not influence DNA binding (Fig. 41, lanes 4-5), whereas mutation of WH2 residues K356/S358, K359, R365, or K393 (Fig. 40C) impaired binding (Fig. 41, lanes 6-7, 9-10). Further consistent with the model, a polyethylene glycol molecule is bound in the crystal at the position of the modeled DNA, indicating affinity of this site for nucleic acids (Fig. 42). Thus, DNA-binding of the tWH domain requires WH2 residues in  $\alpha 3$  and  $\beta 3$ , consistent with a typical WH-DNA interaction.



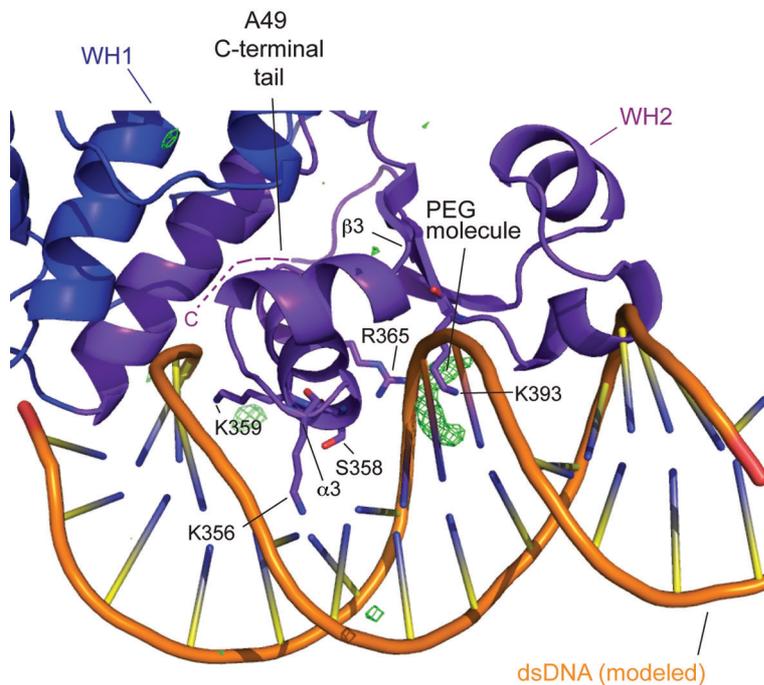
**Figure 40. Two DNA-binding models for the A49 tWH domain**

(A) Alternative superpositions of OhrR-DNA complex structure (PDB 1Z9C (Hong et al., 2005)) onto A49 tWH subdomains WH1 (blue) or WH2 (purple). OhrR is depicted in grey and DNA in orange. (B) Modeled interaction between A49 WH1 and DNA. Putative DNA-interacting residues in A49 are shown as sticks and labeled. (C) Modeled interaction between A49 WH2 and DNA.



**Figure 41. A49 tWH domain variants do not bind dsDNA**

(A) Subdomain WH2 surface mutations impair dsDNA binding. The interaction of tWH variants with promoter dsDNA -25 to +8 (Fig. 39) was analyzed by EMSA. In lane 2, the dimerization module was used as negative control. The tWH domain A49<sub>183-415</sub> and its variants were used in lane 3 and lanes 4-11, respectively (WH1-7mut, Q309A/F310A/R312A/S313A/K314A/D315A/R316A; WH1-5mut, S289A/E292A/Q309A/S313A/K314A; point mutant variants as indicated). (B) SDS-PAGE analysis of various purified A49 tWH variants.

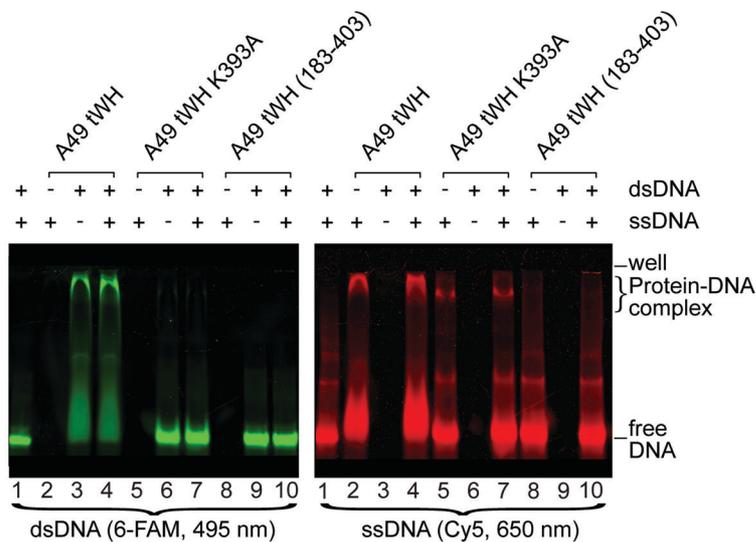


**Figure 42. DNA-binding of the A49 tWH domain**

Modeled interaction between A49 WH2 and dsDNA based on superposition in Figure 40C. Residues important for DNA binding are depicted. A polyethylene glycol molecule was observed in a difference electron density map (green mesh, contoured at  $3.8\sigma$ ).



To investigate whether the tWH domain can simultaneously bind ssDNA and dsDNA, we labeled ssDNA and dsDNA with a red and green fluorescent dye, respectively. The tWH domain bound ssDNA in the presence of bound dsDNA (Fig. 44, lanes 2-4). Since the fluorescence intensities of the migrating protein-DNA complexes were similar, a competitive DNA binding is not likely. Moreover, a tWH variant defective in dsDNA binding still bound ssDNA (Fig. 44, lanes 5-7). However, a tWH variant lacking 12 C-terminal residues did not bind DNA at all (Fig. 44, lanes 8-10). Thus the DNA-binding WH2 surface is required for dsDNA binding, whereas the basic C-terminal tail is required for both ssDNA and dsDNA binding (Fig. 45). These results show that the tWH domain binds ssDNA with a preference for the non-template strand in the upstream promoter region and that this activity is not mutually exclusive with its dsDNA-binding activity.



**Figure 44. The A49 tWH domain binds double- and single-stranded DNA**

Binding of dsDNA and ssDNA is not mutually exclusive. 6-FAM-labeled dsDNA (-25 to +8) and Cy5-labeled ssDNA (-70 to -31) were visualized within the same gel at 495 nm (green, left) and 650 nm (red, right), respectively. tWH variants included wildtype A49<sub>183-415</sub> and its variant K393A which is defective in dsDNA binding and its variant A49<sub>183-403</sub> that lacks the C-terminal basic tail.

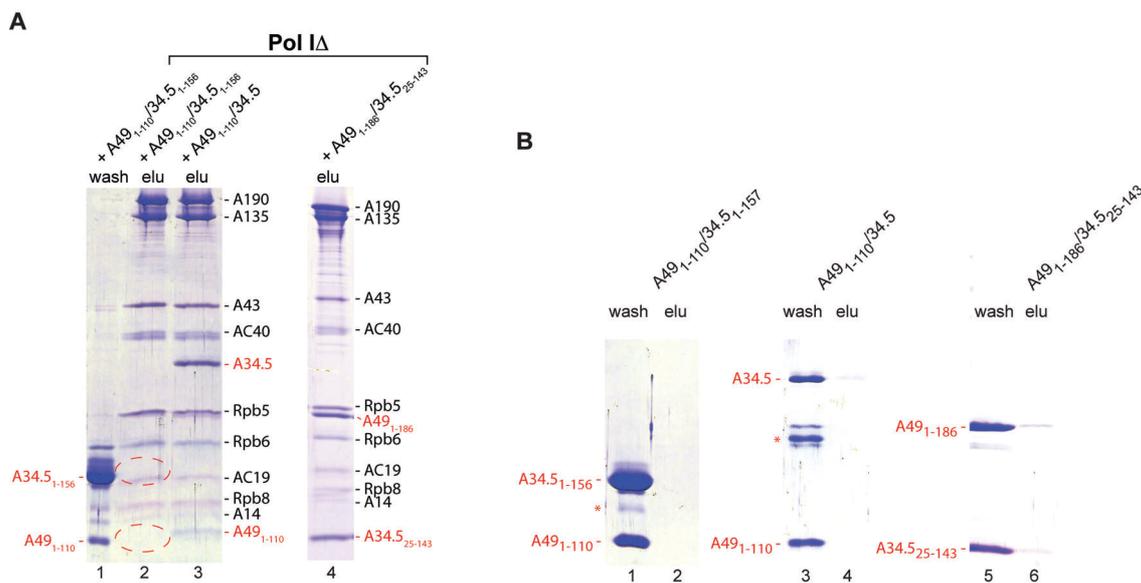
|            | α <sup>2</sup> -2                   | β3-2           | C-terminal tail |     |
|------------|-------------------------------------|----------------|-----------------|-----|
| Scer-A49   | IPKSTAASYKIATMKVPFKL-PEMTRRGR-GPRR- |                |                 | 415 |
| Dhan-A49   | IPKVSASTYKVATLKVPFKL-PEMSKVRKGGR--  |                |                 | 415 |
| Calb-A49   | IPKSAVGTYKIATLKVPFKL-PELTRRGR--R--  |                |                 | 393 |
| Cgla-A49   | IPKSAATYKIASLKVPFKL-PELTRRGR---R--  |                |                 | 407 |
| Spom-A49   | LNKTDAKNHKRAVLKIPLEF-PKP-RRGRA--RN- |                |                 | 425 |
| Mmus-PAF53 | -----EESHRLGTLVPLPP-AQNSDRQS        | KRRKM          |                 | 434 |
| Hsap-PAF53 | -----EEDHKLGTLSLPLPP-AQTSDRLA       | KRRKIT         |                 | 419 |
| Xlae-PAF53 | -----EEGHKIGLLTIPLV-YKPSGDEL        | KRRKM-         |                 | 419 |
|            | -----                               | .....*:-:----- |                 |     |

**Figure 45. A49 contains a C-terminal tail**

Sequence alignment of the C-terminal tail from yeast and mammalian A49. Specific conserved basic residues and secondary structure elements are indicated.

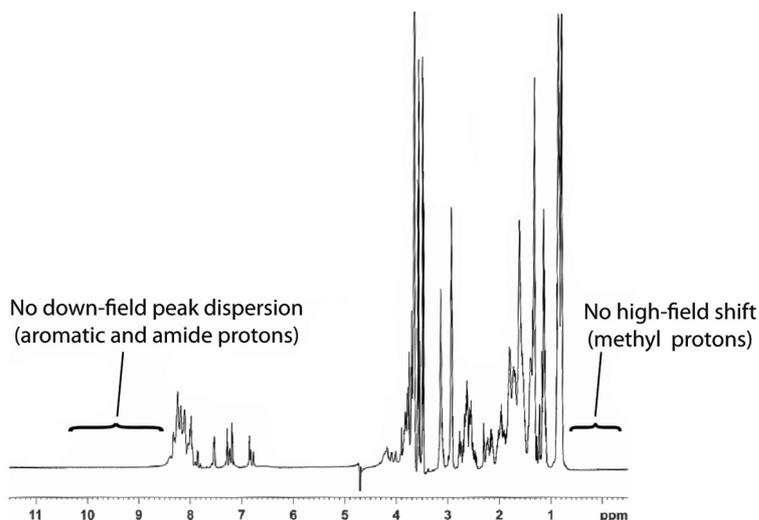
## 2.6 A49/34.5 domains have distinct functions

To map protein regions that anchor A49/34.5 to the Pol I core, we tested A49/34.5 variants for binding to Pol I $\Delta$ . The dimerization module did not bind Pol I $\Delta$ , but its extension by the A49 linker, or by the A34.5 tail, enabled binding (Fig. 46A, lanes 1-4). The A49 linker is proteolytically sensitive (Fig. 29A), and a purified linker variant (Fig. 29B) was unfolded in solution (1D-NMR, Fig. 47). A Circular Dichroism (CD) measurement confirmed this result, but showed that the linker could adopt a partly  $\alpha$ -helical fold, when increasing amounts of trifluorethanol were added (Fig. 48). Consistent with this result, the linker is predicted to contain five helices, two of which were observed in crystals. A part of the N-terminal linker helix (residues 100-106) was observed in a poorly diffracting crystal of the extended dimerization module variant A49<sub>1-119</sub>/34.5<sub>25-143</sub> (Fig. 49A and Table 6). The C-terminal linker helix (residues 172-181) was observed in the tWH crystal in one molecule in the asymmetric unit (Fig. 49B), and in a different position in another crystal form (not shown, Table 7). These results show that A49/34.5 is anchored on Pol I with its domain linker, and suggests that the A49/34.5 domains maintain some mobility on the Pol I surface.



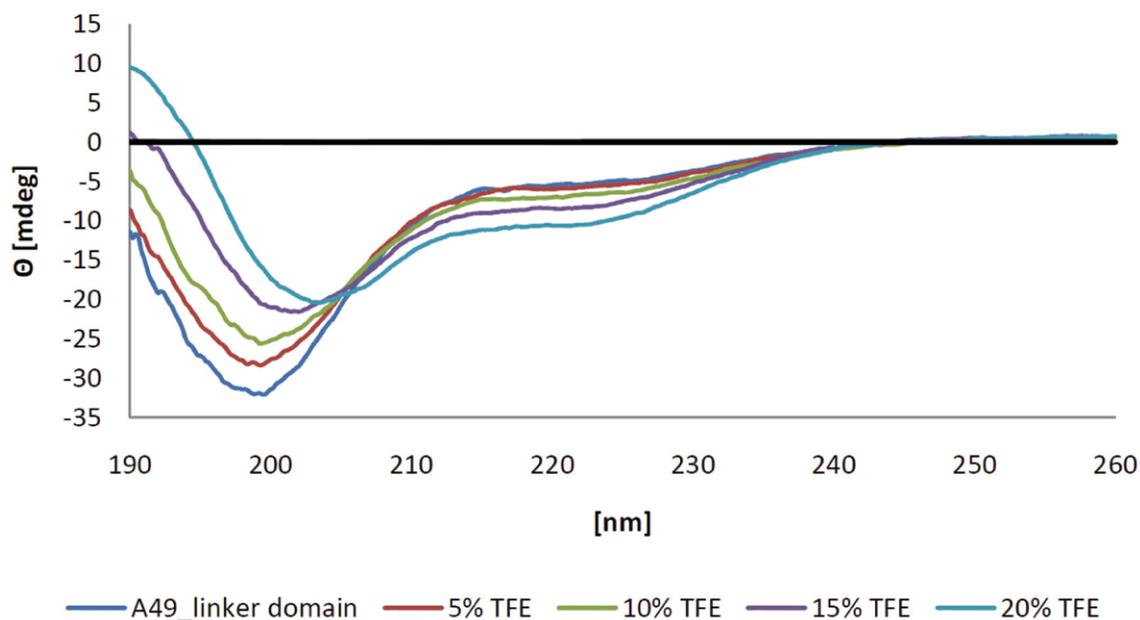
**Figure 46. Different A49/34.5 regions stabilize the dimerization module on Pol I**

(A) The A49 linker or the A34.5 C-terminal tail are required to anchor A49/34.5 on Pol I. Wash and elution (elu) fractions of pulldown experiments were analyzed by SDS-PAGE. Pol I $\Delta$  subunits are indicated in black, A49 and A34.5 variants are in red. The dimerization module A49<sub>1-110</sub>/34.5<sub>1-156</sub> did not bind Pol I $\Delta$  (lanes 1-2), but binding was enabled by its extension with the A49 linker (A49<sub>1-186</sub>/34.5<sub>25-143</sub>, lane 4), or with the A34.5 tail (A49<sub>1-110</sub>/34.5, lane 3). Wash samples were TCA-precipitated (lane 1). (B) Pulldown controls for various A49/34.5 variants. Degradation products are indicated with an asterisk. A49<sub>1-110</sub>/34.5<sub>1-156</sub> (lanes 1-2), A49<sub>1-110</sub>/34.5 (lanes 3-4) and *C. glabrata* A49<sub>1-186</sub>/34.5<sub>25-143</sub> (lanes 5-6).



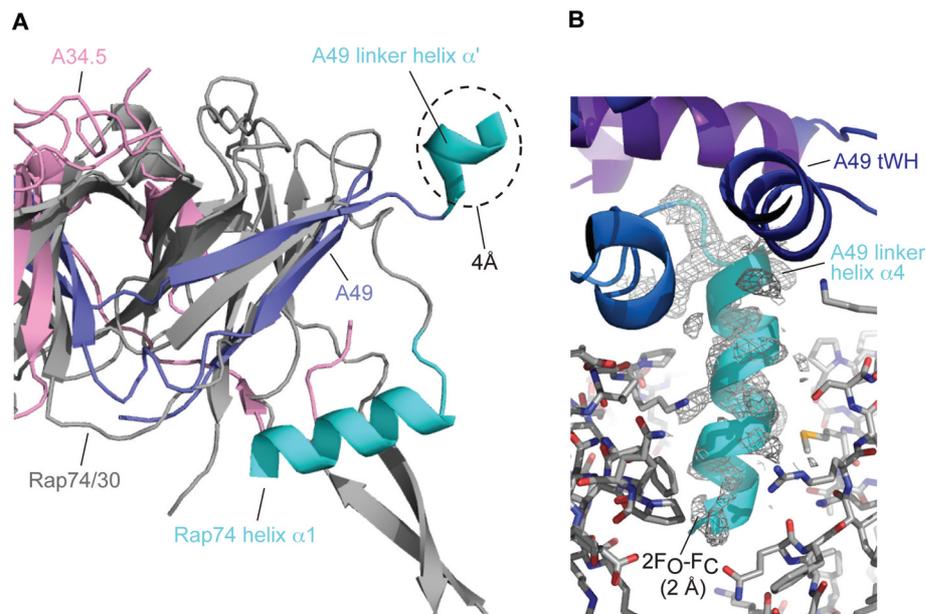
**Figure 47. A49 linker domain is unfolded in solution**

1D- $^1\text{H}$ -NMR spectrum of A49 linker domain, with chemical shifts of  $^1\text{H}$  protons indicated in ppm. Aromatic and amide protons were only visible between 6.7 and 8.4 ppm, with no observed peak dispersion (down-field chemical shift,  $> 9$  ppm). High-field shifted methyl protons ( $< 1$  ppm) were not detected, too. These results show that the A49 linker domain is unfolded in solution. Data were collected by Sabine Wenzel (University of Bayreuth).



**Figure 48. A49 linker domain can adopt a  $\alpha$ -helical fold**

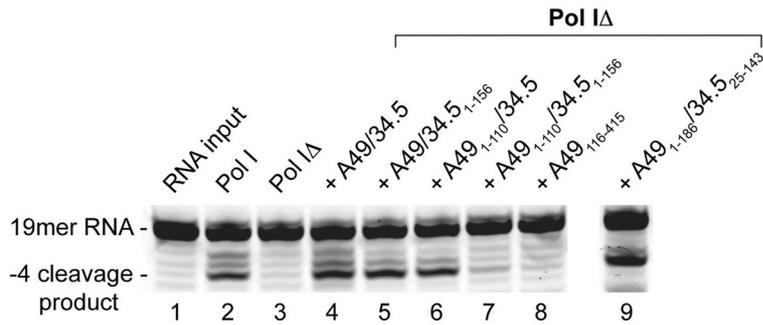
Circular dichroism (CD) analysis of A49 linker domain. The wavelength (nm) and molar ellipticity ( $\theta$ ) are indicated. The A49 linker domain is unfolded in solution (blue curve), but titration with increasing amounts of trifluoroethanol (TFE) induced a partly  $\alpha$ -helical fold (brown, green, purple, and light blue curve, respectively). Data were collected by Sabine Wenzel (University of Bayreuth).



**Figure 49. Structure of the A49 linker domain**

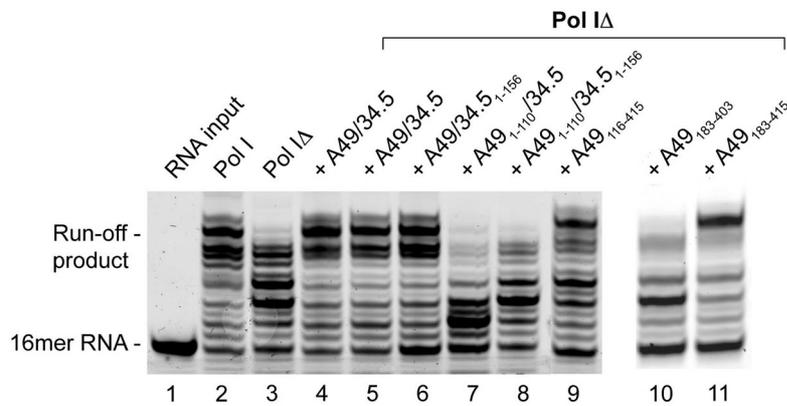
(A) Structural superposition of the extended A49/34.5 dimerization module with the TFIIF Rap74/30 dimerization module structure (grey). Linker helices  $\alpha'$  (A49) and  $\alpha 1$  (Rap74) are in cyan. (B) Observation of A49 linker helix  $\alpha 4$  in one tWH molecule of the asymmetric unit of the structure in Figure 35. The linker helix is in cyan, and neighboring domains in the crystal are in grey. Other colors are as in Figure 29A. The  $2F_o - F_c$  electron density map is contoured at  $0.8\sigma$  (grey mesh).

We previously showed that A49/34.5 is required for normal Pol I RNA cleavage activity and processivity on synthetic templates (chapter III, 2.7). To map regions required for these functions, A49/34.5 variants were added to Pol I $\Delta$  and the activities were tested *in vitro* (Experimental procedures). These experiments revealed that normal RNA cleavage requires the dimerization module and the A49 linker or the A34.5 tail (Fig. 50). Thus the dimerization module mediates full RNA cleavage activity, but this requires anchoring the domain on Pol I. Processivity, as measured by the ability of Pol I to synthesize RNA to the end of the DNA template, required the tWH domain, including the basic C-terminal tail (Fig. 51). Taken together our results reveal the functional architecture of A49/34.5 (Fig. 52) and enabled homology modeling for the related mammalian subcomplex Paf53/49.



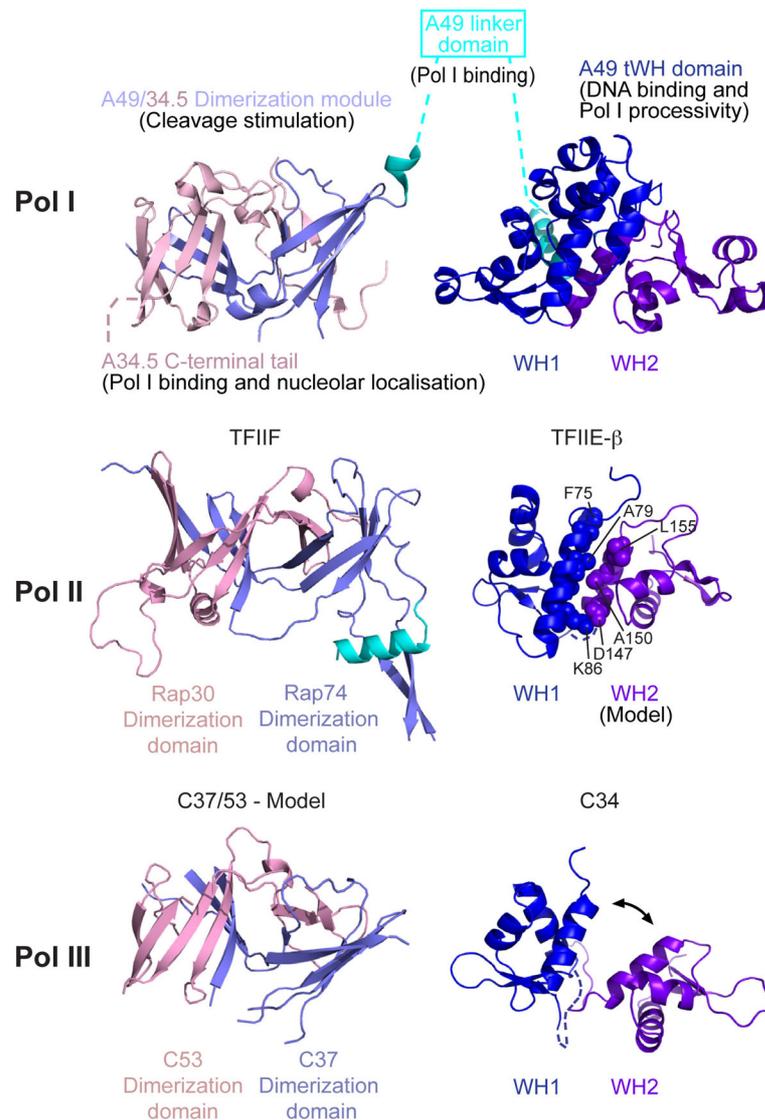
**Figure 50. The dimerization module stimulates RNA cleavage if it is anchored to Pol I**

Pol I removes four nucleotides from the RNA (lane 2). This activity depends on A49/34.5, as Pol IΔ is defective in cleavage (lane 3). Cleavage can be restored by addition of A49/34.5 (lane 4) or its variants containing the dimerization module and either the A49 linker (lane 9) or the A34.5 tail (lane 6).



**Figure 51. A49 tWH domain is required for Pol I processivity**

Pol I processivity on a minimal DNA-RNA scaffold (chapter III, Fig. 24A) requires the A49 tWH domain. Pol I elongates RNA (lane 1) by 12 nucleotides to the run-off product (lane 2) whereas Pol IΔ can not (lane 3). Processivity is restored by addition of recombinant A49/34.5 (lanes 4-5) or variants that contain the tWH domain (lanes 6, 9 and 11) and depends on the tWH C-terminal basic tail (compare lanes 10-11).



**Figure 52. Evolutionary relationship between Pol I/III-specific subunits and Pol II factors**

On the top, a model of the Pol I subcomplex A49/34.5 based on the new structural information is shown. Colors are as follows: A49 dimerization domain, light blue; A34.5 dimerization domain, magenta; A49 linker, cyan; A49 WH1, blue; A49 WH2, purple. For functional assignments compare text. The A34.5 C-terminal tail is important for nucleolar localization (Ushijima et al., 2008). In the middle, the A49/34.5-related domains in Pol II initiation factors are shown. The structure of the TFIIF Rap74/30 dimerization module (Gaiser et al., 2000) and the proposed TFIIE- $\beta$  WH1 subdomain (Okuda et al., 2000) are known. The structure of the proposed TFIIE- $\beta$  WH2 was modeled (Sali et al., 1995) based on the tWH structure. Some of the residues that potentially contribute to the tWH subdomain interface (F75, L78, A79, V82, A150, L151, and L155) are indicated with spheres. On the bottom, models for A49/34.5-related domains in Pol III-specific subunits are shown. The C37/53 dimerization module was modeled based on the A49/34.5 dimerization module structure, using a secondary structure-based alignment (Fig. 32). Structures for C34 WH1 and WH2 have been determined (2DK8, 2DK5). Based on the modeling it is unlikely that the WH1-WH2 interface is stable, and two domains may move with respect to each other.

### 3 Discussion

#### 3.1 Structure and DNA-binding function of Pol I subcomplex A49/34.5

The main difference between Pol I and Pol II is the presence of the additional subcomplex A49/34.5 in Pol I. We show here that A49/34.5 forms a heterodimer that consists of two structured domains connected by a flexible linker. The N-terminal regions of both subunits form a dimerization module with a TFIIF-like fold, whereas the C-terminal region of A49 forms a tWH domain with two WH subdomains. The dimerization module mediates normal RNA cleavage activity of Pol I, whereas the tWH domain stimulates Pol I processivity and binds DNA. The tWH domain binds both ssDNA and dsDNA, and these interactions are apparently not mutually exclusive. This observation is consistent with a previously reported nonspecific dsDNA binding affinity of Pol I (Bric et al., 2004). The tWH domain binds preferentially to the non-template single strand in the upstream promoter region. These results suggest that the tWH domain is involved in promoter opening and/or stabilization of the early transcription bubble by trapping the non-template DNA strand. The transcription bubble was mapped for the protozoan amoebae *Acanthamoeba castellanii* around the transcription start site (Kahl et al., 2000), but the location of the initial Pol I transcription bubble in yeast remains unknown.

Published results are consistent with a DNA-binding function of the tWH domain that may be relevant during transcription initiation and elongation *in vivo*. In *S. cerevisiae*, tWH residues 367-415 are responsible for a cold-sensitive phenotype, whereas A49 residues 1-119 in the dimerization domain and A34.5 are entirely dispensable *in vivo* (Beckouet et al., 2008). A yeast strain lacking residues 367-415 of the tWH domain was sensitive to 6-azauracile and mycophenolate, showed compromised Pol I processivity, and a reduction of Pol I occupancy in the rDNA promoter and transcribed region (Beckouet et al., 2008). Mouse A49 is required for promoter-dependent transcription (Hanada et al., 1996). This was confirmed for *S. pombe* A49 and it was additionally shown that the region forming the tWH domain is responsible for a cold-sensitive phenotype (Nakagawa et al., 2003).

#### 3.2 Pol I resembles a minimal Pol II-TFIIF-TFIIE complex

Since TFIIF contains the dimerization module, but not the tWH domain, we searched for a tWH domain in another Pol II initiation factor. Two consecutive WH domains were found in the TFIIE subunit  $\beta$  (Table 8), one that was structurally resolved (Okuda et al., 2000) and one

that could be predicted (Fig. 37) with HHPred (Soding et al., 2005). Modeling the arrangement of these two WH domains within a putative tWH structure revealed conserved hydrophobic residues in the subdomain interface, and putative conserved salt bridge between residues K86 and D147 (Figs. 37 and 52). Consistent with these results, the mutation K86E impairs transcription (Tanaka et al., 2009). This suggests that TFIIE- $\beta$  contains a tWH domain and that A49/34.5 combines features of TFIIF and TFIIE, which both contribute to Pol II promoter binding. TFIIF binds dsDNA (Groft et al., 1998), prevents non-specific DNA binding to Pol II (Killeen and Greenblatt, 1992), and is required for initiation complex formation (Tan et al., 1995). TFIIE binds ssDNA (Okamoto et al., 1998) and dsDNA (Okuda et al., 2000; Tanaka et al., 2009). The ssDNA-binding activity requires TFIIE- $\beta$  residues 257-277 (Okamoto et al., 1998), which correspond to the A49 C-terminal tail that is required for ssDNA binding to the A49 tWH domain *in vitro* (Figs. 43, 44 and 45).

**Table 8.** A49 tWH homology statistics

|   | A49 WH1               |                                | A49 WH2               |                                |
|---|-----------------------|--------------------------------|-----------------------|--------------------------------|
|   | RMSD (Å) <sup>1</sup> | Fold identity (%) <sup>1</sup> | RMSD (Å) <sup>1</sup> | Fold identity (%) <sup>1</sup> |
| C34 WH1 (2DK8) <sup>2</sup>               | <u>2.0</u>            | <u>77.6</u>                    | 2.9                   | 59.0                           |
| C34 WH2 (2DK5) <sup>2</sup>               | 3.3                   | 64.2                           | <u>2.4</u>            | <u>60.2</u>                    |
| TFIIE- $\beta$ WH1 (1D8K) <sup>2</sup>    | <u>2.2</u>            | <u>68.7</u>                    | 3.7                   | 61.5                           |
| TFIIE- $\beta$ WH2 (modeled) <sup>2</sup> | 3.0                   | 62.4                           | <u>2.5</u>            | <u>65.1</u>                    |

<sup>1</sup>Deviations in C $\alpha$ -positions (RMSD) and fold identities were analyzed by superposing the corresponding atomic structures with LSQKAB (Kabsch, 1976).

<sup>2</sup> PDB-codes are indicated. For TFIIE- $\beta$ , the modeled structure (2D1H) was used. The homologous structural pairs are highlighted in bold and underlined letters.

### 3.3 Pol III contains TFIIF- and TFIIE-like regions

There is also evidence that Pol III subunits resemble parts of TFIIF and TFIIE. The Pol III subunit heterodimer C37/53 apparently forms a TFIIF-like dimerization module (chapter III, 2.6), and could be modeled with the A49/34.5 dimerization module structure (Fig. 52). All  $\beta$ -strands of a triple-barrel were predicted (Fig. 32) and most hydrophobic core residues were conserved (Fig. 32). The related dimerization modules in A49/34.5, TFIIF, and C37/53 bind at similar locations on the Rpb2 side of Pol II (chapter III, 2.6)(Chen et al., 2010; Fernandez-Tornero et al., 2007). Unfortunately we could not investigate this further since the structures

presented here could not be placed unambiguously into the previous Pol I electron microscopic density (chapter III, Figs. 9 and 10). Subunit C34 contains two consecutive WH domains (PDB-codes 2DK8, 2DK5) that could form a tWH domain, although the putative interface contains only three hydrophobic residues (V16, I23, L94, Figs. 37 and 52). Subunit C82 was also predicted to contain subsequent WH domains, but these include an additional N-terminal helix, as in the ‘extended WH domain’ of TFIIE- $\alpha$  (Meinhart et al., 2003). Thus the C82/34 heterodimer (Lorenzen et al., 2007) may be distantly related to the TFIIE- $\alpha/\beta$  heterodimer. Consistent with this homologies, C37/53 and C82/34/31 function in transcription initiation (Brun et al., 1997; Kassavetis et al., 2010; Landrieux et al., 2006; Wang and Roeder, 1997). While this work was about to be completed, a bioinformatic analysis also suggested some similarities between Pol III subunits and Pol II initiation factors (Carter and Drouin, 2010). These results indicate that Pol III is also related to a Pol II-TFIIF-TFIIE complex.

### **3.4 Polymerase evolution and promoter specificity**

Here we provide evidence that Pol I- and Pol III-specific subcomplexes contain domains that are structurally related to parts of the Pol II initiation factors TFIIF and TFIIE (Fig. 52). Our results and published data indicate that TFIIF- and TFIIE-related subcomplexes are involved in promoter binding. These subcomplexes however show substantial differences in their domain organization, suggesting that they evolved to allow for gene class-specific promoter recognition, and apparently also for additional functions during elongation such as processivity.

## 4 Experimental procedures

### 4.1 Mass spectrometry

A49/34.5, Pol I and Pol I $\Delta$  were prepared as described in chapter III (3.1, 3.6 and 3.9). For native MS the sample buffer was exchanged to a solution containing 160 mM or 880 mM ammonium acetate using centrifugal filter units (Millipore) and sample concentration was adjusted to 2  $\mu$ M. MS was carried out on a Q-ToF I instrument (Lorenzen, 2007; van den Heuvel et al., 2006). The cone voltage was 150 V and the needle voltage was 1.3 kV. The pressure in the source region was 10 mbar. Xenon was used as a collision gas with a pressure of 20  $\mu$ bar (Lorenzen, 2007). Data were analyzed with MassLynx (Waters).

### 4.2 Recombinant proteins

The genes for A49 and A34.5 were cloned sequentially into vector pET28b (Merck), resulting in a thrombin-cleavable N-terminal hexahistidine tag on A49. For bicistronic expression, a second ribosomal binding site was inserted before A34.5. The *S. cerevisiae* variant A49<sub>1-110</sub>/34.5<sub>1-156</sub> was expressed for 18 h at 18 °C in *E. coli* BL21 (DE3) RIL cells (Agilent). Cells were harvested by centrifugation, resuspended in 50 ml buffer A (50 mM MES pH 6.3, 300 mM NaCl, 10 mM  $\beta$ -mercaptoethanol, 1 mM protease inhibitor mix containing 1 mM PMSF, 1 mM benzamidine, 200  $\mu$ M pepstatin, and 60  $\mu$ M leupeptin) and lysed by sonication. After centrifugation, the supernatant was loaded onto a 3 ml Ni-NTA column (Qiagen) equilibrated with buffer A. The column was washed stepwise with 25 ml of buffer A containing 1 M NaCl and 15 ml of buffer A containing 10 mM imidazole. Protein was eluted with buffer A containing 200 mM imidazole. Eluted fractions were diluted three-fold with buffer A lacking NaCl and incubated with thrombin (1 U protease/1 mg protein) for 16 h at 4 °C. A MonoQ column (GE Healthcare) was equilibrated with buffer B (50 mM MES pH 6.3, 100 mM NaCl, 5 mM DTT). Protein was obtained in the flow-through, applied to a MonoS column (GE Healthcare) equilibrated with buffer B, and eluted with a linear gradient from 100 mM to 1M NaCl. The sample was concentrated and applied to a Superose 12 HR column (GE Healthcare) equilibrated with 50 mM Tris pH 7.5, 150 mM NaCl, and 5 mM DTT. Pooled peak fractions were concentrated to 20 mg/ml. Purification of other dimerization module variants was as above except A49/34.5<sub>1-156</sub>, which was purified as described in chapter III (3.6).

The A49 linker variant A49<sub>105-187</sub> was expressed for 4 h at 37 °C in *E. coli* BL21 (DE3) RIL cells and purified as above. Purified linker was concentrated to approximately 10 mg/ml.

A49 tWH domain variants were cloned, expressed, and purified essentially as above, except that the MonoQ step was skipped and gel filtration was carried out in 50 mM Tris pH 7.0, 150 mM NaCl, and 5 mM DTT. Pooled peak fractions were concentrated to 16 mg/ml. Limited proteolysis was performed as described (Geiger et al., 2008; Hubbard, 1998), using trypsin and chymotrypsin as proteases.

### 4.3 Crystal structure determinations

Crystals for *C. glabrata* variant A49<sub>1-99</sub>/34.5<sub>25-143</sub> were grown at 20 °C in hanging drops using as reservoir solution 22% PEG 3350 and 250 mM sodium fluoride. Crystals were cryo-protected by a stepwise transfer to reservoir solution containing 7-20% PEG 400, and flash cooled by plunging into liquid nitrogen. Two additional methionines were introduced as non-disruptive point mutations (A49-V72M, A34.5-L55M). Selenomethionine-labeled variant was prepared as described (Budisa et al., 1995; Meinhart et al., 2003) and crystallized at 20 °C using as reservoir solution 20% PEG 3350 and 50 mM Tris pH 7.5. Crystals reached a size of 300 µm x 60 µm x 60 µm, and were cryo-preserved in reservoir solution containing 23% glycerol. SAD diffraction data were obtained at the Swiss Light Source and processed with XDS (Kabsch, 1993) (Table 6). Programs SHELXD/HKL2MAP (Pape and Schneider, 2004; Schneider and Sheldrick, 2002) detected 16 selenium sites, two sites in each of the eight heterodimers in the asymmetric unit. SHARP (de La Fortelle and Bricogne, 1997) was used for SAD phasing, and SOLOMON (Abrahams and Leslie, 1996) for density modification. NCS averaging was performed with DM (Cowtan, 1994). The model was built with COOT (Emsley and Cowtan, 2004) and refined with PHENIX (Afonine et al., 2005) and BUSTER (Blanc et al., 2004) (Table 6). Crystals for *C. glabrata* variant A49<sub>1-119</sub>/34.5<sub>25-143</sub> were grown using as reservoir solution 20% PEG 3350 and 50 mM Tris pH 7.0. Diffraction data were collected at the European Synchrotron Radiation Facility ESRF (Table 6). The structure was solved by molecular replacement with PHASER (McCoy et al., 2005), using the dimerization module structure as a search model. A NCS-averaged map was generated with PARROT (Cowtan, 2010). The model was manually extended and refined at 4 Å resolution (Table 6).

Crystals for *S. cerevisiae* A49<sub>171-403</sub> were grown at 4 °C in hanging drops, using as reservoir solution 25% PEG 3350 and 100 mM Tris pH 8.5. Crystals reached a size of 400 µm x 80 µm x 80 µm, were harvested in reservoir solution containing 10-20% PEG 400 and flash cooled.

In addition to present methionine residues (M170, M256, M397), two additional methionines were introduced (L178M, L261M) and selenomethionine-labeled protein was prepared as above and crystallized at 4 °C using as reservoir solution 24% PEG 3350 and 50 mM Tris pH 8.0. MAD data were obtained at ESRF, processed as above (Table 7), and the structure was determined essentially as above, except that MAD phasing was used. 20 selenium sites were detected, four in each of five tWH domains in the asymmetric unit. The model was built automatically with ARP/wARP (Langer et al., 2008), manually adjusted, and refined (Table 7). Crystals for variant A49<sub>155-399</sub> were grown at 4 °C in hanging drops, using as reservoir solution 100 mM Tris pH 8.5, 26% PEG 4000 and 200 mM lithium sulfate monohydrate, reached a size of 80 µm x 30 µm x 30 µm and were cryo-preserved in 20% ethylene glycol. The structure was solved by molecular replacement using the A49 tWH structure as a search model and was refined (Table 7).

#### **4.4 RNA cleavage and extension assays**

RNA cleavage assays were performed with a nucleic-acid scaffold containing a three-nucleotide non-complementary overhang at its 3' end as described in chapter III (3.11), except that the reaction buffer contained 20 mM MES pH 6.0. RNA extension assays were performed with a pre-annealed minimal nucleic-acid scaffold as described in chapter III (3.10).

#### **4.5 Electrophoretic mobility shift assay**

EMSA assays were performed essentially as described ([www.labs.fhcrc.org/hahn/Methods](http://www.labs.fhcrc.org/hahn/Methods)) except that the buffer was 20 mM Tris pH 8.0, 150 mM NaCl, 1 mM DTT, 40 µg/ml Heparin, 5 mM MgCl<sub>2</sub>, 4% glycerol. dsDNA was prepared by annealing equimolar amounts of complimentary oligonucleotides (sequences, Table 9). For dual wavelength experiments, 6-FAM labeled dsDNA and Cy5-labeled ssDNA were used. 1 pmol of DNA were incubated with 100 pmol of protein for 15 min at 4 °C. The DNA-protein complexes were resolved by gel-electrophoresis (0.2 pmol per lane) in an 8/0.08% acrylamid/bisacrylamid gel (25 mM Tris pH 8.2, 19 mM glycine, 2.5% glycerol, 0.5 mM DTT) at 150 Volt for 50 min and visualized by a Typhoon 9400 phosphoimager (GE Healthcare).

**Table 9.** Nucleic acid sequences used in EMSA analysis

| Name                        | Promoter position <sup>1</sup> | Label (5') <sup>2</sup> | Nucleotide sequence (5'-3', non-template strand)<br>Reverse complementary sequence (3'-5', template strand) <sup>3</sup> |
|-----------------------------|--------------------------------|-------------------------|--|
| Non sequence specific dsDNA | (47nt)                         | 6-FAM                   | 5'-TGAGTAAGCCGTTTTGCAAGGA<br>AGTGTGTTTCTTAAGTACTTGAGCT<br>5'-AGCTCAAGTACTTAAGAAACA<br>CACTTCCTTGCAAAACGGCTTACTCA         |
| Ex. Upst. Dom. 2            | -70 to -31                     | 6-FAM                   | 5'-TAGTTTGTAATGGGAGGGGGG GTTTAGTCATGGAGTACAA<br>5'-TTGTAAGTCCATGACTAAACCC CCCTCCATTACAAACTA                              |
| Core El. linker             | -51 to -31                     | 6-FAM                   | 5'-GGGTTTAGTCATGGAGTACAA-3'<br>5'-TTGTAAGTCCATGACTAAACCC-3'  |
| Upst. Dom. 2                | -70 to -52                     | 6-FAM                   | 5'-TAGTTTGTAATGGGAGGGG-3'<br>5'-CCCCTCCATTACAAACTA-3'  |
| Core El. Dom. 1             | -25 to +8                      | 6-FAM                   | 5'-AGGAAAAGTAGTTGGGAGGTACTTCATGCGAAA-3'<br>5'-TTTCGCATGAAGTACCTCCCAACTACTTTTCCT-3'                                       |
| Tr. Core El. a              | -18 to +8                      | 6-FAM                   | 5'-GTAGTTGGGAGGTACTTCATGCGAAA-3'<br>5'-TTTCGCATGAAGTACCTCCCAACTAC-3'   |
| Tr. Core El. b              | -30 to -1                      | 6-FAM                   | 5'-GTGTGAGGAAAAGTAGTTGGGAGGTACTTC-3'<br>5'-GAAGTACCTCCCAACTACTTTTCCTCACAC-3'   |
| ss-upst.-prom. (a)          | -110 to -71                    | Cy5                     | 5'-AGAGCGACAGAGAGGGCAAAGAAAATAAAAAGTAAGATTT<br>-   |
| ss-upst.-prom. (b)          | -100 to -61                    | Cy5                     | 5'-AGAGGGCAAAGAAAATAAA AGTAAGATTTTAGTTTGTA<br>-  |
| ss-upst.-prom. (c)          | -90 to -51                     | Cy5                     | 5'-AGAAAATAAAAAGTAAGATTTTAGTTTGTAATGGGAGGGGG<br>-  |
| ss-upst.-prom. (d)          | -80 to -40                     | Cy5                     | 5'-AGTAAGATTTTAGTTTGTAATGGGAGGGGGGGTTTAGTCAT<br>-  |
| ss-upst.-prom. (e)          | -74 to -35                     | Cy5                     | 5'-GATTTTAGTTTGTAATGGGAGG GGGGGTTTAGTCATGGAG<br>-  |
| ss-upst.-prom. (f)          | -70 to -31                     | Cy5                     | 5'-TAGTTTGTAATGGGAGGGGGGGT TTAGTCATGGAGTACAA<br>-  |
| ss-upst.-prom. (g)          | -65 to -26                     | Cy5                     | 5'-TGTAATGGGAGGGGGGGTTAGT CATGGAGTACAAGTGTG<br>-   |
| ss-upst.-prom. (h)          | -60 to -20                     | Cy5                     | 5-TGGGAGGGGGGGTTTAGTCATGGAGTACAAGTGTGAGGAAA<br>-   |
| ss-upst.-prom. (i)          | -49 to -10                     | Cy5                     | 5'-GTTTAGTCATGGAGTACAAGTGT GAGGAAAAGTAGTTGGG<br>-  |
| ss-upst.-pr.-r (a)          | -100 to -61                    | Cy5                     | 5'-TTACAAACTAAAATCTTACTTTTATTTTCTTTTGCCCTCT<br>-   |
| ss-upst.-pr.-rev(b)         | -90 to -50                     | Cy5                     | 5'-CCCCCTCCATTACAAACTAAAATCTTACTTTTATTTTCT<br>-  |
| ss-upst.-pr.-rev(c)         | -80 to -40                     | Cy5                     | 5'-ATGACTAAACCCCCCTCCATTACAAACTAAAATCTTACT<br>-  |
| ss-upst.-pr.-rev(d)         | -70 to -31                     | Cy5                     | 5'-TTGTAAGTCCATGACTAAACCCCCCTCCATTACAAACTA<br>-  |
| ss-upst.-pr.-rev(e)         | -60 to -20                     | Cy5                     | 5'-TTTCCTCACACTTGTACTCCATGACTAAACCCCCCTCCA<br>-  |
| ss-upst.-pr.-rev(f)         | -49 to -10                     | Cy5                     | 5'-CCCAACTACTTTTCCTCACACTTGTACTCCATGACTAAAC<br>-   |

<sup>1</sup> Pol I promoter positions are indicated according to the mapped (+1) initiation site (Moss et al., 2007).<sup>2</sup> DNA-fragments are labeled at the 5'-end with either 6-FAM and Cy5 for double- or single-stranded DNA.<sup>3</sup> Nucleotide sequence is shown for non-template strand (5'-3') and template strand (3'-5') of the Pol I promoter.

#### **4.6 Protein interaction assay**

15  $\mu\text{g}$  of Pol I $\Delta$  was incubated with 5  $\mu\text{g}$  of A49/34.5 variant in buffer C (5 mM Hepes pH 7.8, 60 mM ammonium sulfate, 1 mM  $\text{MgCl}_2$ , 10  $\mu\text{M}$   $\text{ZnCl}_2$ , 10  $\mu\text{M}$   $\beta$ -mercaptoethanol) for 25 min at 15  $^\circ\text{C}$ . 40  $\mu\text{l}$  of Ni-NTA beads (Qiagen) equilibrated with buffer C were added and the mixture was incubated for 1 h at 4  $^\circ\text{C}$ . Pol I bound to the beads via a hexahistidine tag on its A43 subunit. Beads were centrifuged and washed four times with buffer C containing 5 mM imidazole and 0.2% NP-40. Protein complexes were eluted with buffer C containing 200 mM imidazole, and were incubated at 95  $^\circ\text{C}$  for 5 min in SDS-PAGE loading buffer.

## Final conclusions and outlook

During this thesis it was possible to study RNA polymerase I subunits both at a structural and functional level. This included comprehensive biochemical investigations and led to three, new crystal structures of individual Pol I subunits, which enabled the construction of the hybrid structure of the 12-subunit Pol I enzyme.

As described in chapter II, a comprehensive protein engineering strategy was developed to delineate protein subcomplexes. This approach was applied to predict, locate and remove multiple flexible regions in a stepwise manner and was essential to obtain crystals for the Pol I subcomplex A14/43.

In chapter III, the resulting atomic structure was positioned in a cryo-EM map of Pol I, obtained by Claus Kuhn and colleagues, leading to a homology model of the Pol I core and a hybrid structure of the 12-subunit polymerase. This structural information was complemented with functional data, such as the discovery of the intrinsic RNA cleavage activity of Pol I, which was dependent on the A12.2 subunit. The Pol I-specific subcomplex A49/34.5 was purified and predicted to be homologous to the Pol II initiation factor TFIIF. This finding was consistent with *in vitro* and *in vivo* experiments by Claus Kuhn, which illustrated that A49/34.5 is a built-in Pol I processivity factor. These various structural and functional results were combined, culminating in a novel model of the functional architecture of Pol I.

In chapter IV, the focus was to investigate the peripheral subcomplex A49/34.5 in detail. Two new crystal structures of the A49/34.5 dimerization module and the A49 tandem winged helix (tWH) domain, homologous to the Pol II initiation factors TFIIF and TFIIE and to the Pol III-specific subunits C37/C53 and C34, illustrate an evolutionary relationship between the three RNA polymerases. Functional investigations revealed that the A49/34.5 dimerization module and the A49 tWH were required for full Pol I cleavage activity and processivity, respectively. The A49 tWH domain binds dsDNA, which could be modeled on a conserved surface. Interestingly, a promoter-specific ssDNA binding affinity of the A49 tWH domain was observed, which indicates that the subcomplex may participate in Pol I promoter recognition.

Results presented in this thesis revealed new structural and functional relationships between the three eukaryotic polymerases and raised several questions that require further investigation.

The A14/43 subcomplex structure was fitted in the Pol I stalk, in a very similar position compared to the other polymerases. However, A14/43 is the only Rpb4/7-like subcomplex discovered so far without a HRDC domain, which is consistent with a higher flexibility of the A43 OB domain. A14/43 mutations and additional Pol I-specific surfaces, including the extended dock domain and the clamp knob, indicate that Pol I initiation factor Rrn3 binds to the polymerase from the upstream side. Pol I is phosphorylated when bound to Rrn3 (Fath et al., 2001) and A43 contains phosphorylation sites in disordered regions (Gerber et al., 2008). Therefore, we could gain further insights into Pol I recruitment to the promoter, if detailed structural information of the A14/43-Rrn3 interaction will be available.

The catalytic center of all polymerases is capable of RNA cleavage, however there are major differences between Pol I, II, and III. Pol II can only cleave RNA at mild alkaline conditions or with assistance of the transcript cleavage factor TFIIS. This is different to Pol I and III, which both have an intrinsic cleavage activity, dependent on their subunits A12.2 and C11 (Chedin et al., 1998), respectively. The Pol III subcomplex C37/53 is important for termination and reinitiation (Landrieux et al., 2006), similar to the Pol I A49/34.5 dimerization module, which is important for full cleavage activity of Pol I, indicating a related mechanism for Pol I and Pol III. The cryo-EM structure of Pol I suggested that the location of A12.2 corresponds to Rpb9 in Pol II (chapter III, Fig. 10), which argues for an allosteric reorganization of the Pol I active center and not a direct interaction, as observed in the Pol II-TFIIS complex (Kettenberger et al., 2003). Therefore, it would be interesting to perform a detailed structure-function analysis of the intrinsic RNA cleavage activities of Pol I and III.

The A49/34.5 subcomplex consists of two separate domains. The A49 tWH domain by itself was sufficient for full Pol I processivity. This is in contrast to the A49/34.5 dimerization module, which showed a mild repressive effect in RNA extension assays (chapter IV, Fig. 51, lane 7). These results indicate that both A49/34.5 domains have contrary roles within the Pol I enzyme, either stimulating Pol I transcription or enhancing cleavage and repressing Pol I processivity. Since Pol I has a tunable active site, which is capable of both RNA synthesis and cleavage, a putative allosteric rearrangement of Pol I by the two A49/34.5 domains is a possible explanation for their differing functions. However, a more detailed understanding of the intrinsic cleavage mechanism and the stimulation of RNA synthesis will likely require the crystal structure of the entire Pol I.

The fact that A49 tWH is essential for full Pol I processivity could be related to its ability to interact with nucleic acids. A49 tWH binds dsDNA, which was modeled on a highly conserved surface. EMSA experiments showed that the dsDNA binding affinity is not sequence-specific and that a stable protein-DNA complex required a minimal length of approximately 30 base pairs. This indicates that the complete bottom surface of the A49 tWH domain may be involved in the protein-DNA interaction interface. A published study showed that disruption of the A49 tWH domain leads to an overall reduction of the Pol I occupancy on the complete rDNA gene (Beckouet et al., 2008). Therefore, it is possible that the A49 tWH domain acts as a mobile ‘anchor’, which keeps the Pol I enzyme on track during rRNA synthesis. Consistent with this assumption, Pol I could be crosslinked to an extended DNA template, indicating a nonspecific DNA interaction (Bric et al., 2004).

Additionally, A49 binds ssDNA with a sequence preference for the nontemplate strand of the upstream region of the Pol I promoter. These results indicate that the tWH domain is involved in promoter opening and/or stabilizing of the early transcription bubble by trapping the nontemplate strand. Since a transcription bubble was mapped around the transcription start site, a possible explanation could be that Pol I opens the promoter upstream and then scans for the transcription start site downstream, as it is known for Pol II. To combine these findings into a comprehensive model, a more detailed analysis of Pol I transcription initiation is needed. The functional data has to be complemented with structural information of the Pol I initiation complex, to understand the different steps from initial promoter recognition, through promoter opening towards open complex formation.

Another important finding is that factors thought to be Pol I- and Pol III-specific, resemble domains, which are present in Pol II transcription factors. It could be that the essential functions of TFIIF and TFIIE in the Pol II system are incorporated in the Pol I and the Pol III enzymes to obtain both minimal and highly effective RNA polymerases. These enzymes would be capable of promoter recognition, initiation complex formation, highly processive RNA synthesis and even proofreading, without the additional degree of complexity, which the recruitment of multiple transcription factors would implicate.

To integrate all data into a comprehensive model, detailed structural information for the complete Pol I enzyme is missing. This does not just include the 14-subunit Pol I, but additionally Pol I complexes with various nucleic acid scaffolds and/or initiation factors. Initial 14-subunit Pol I crystals could be obtained, which enabled the collection of several 4 Å diffraction data sets. However due to the lack of initial phase information, the crystal structure could not be solved so far (Claus Kuhn, Dissertation 2008). Nevertheless obtaining an atomic structure of Pol I is an extremely important goal for the future. Since this effort will be a long-term project, alternative strategies to obtain structural information need to be explored. Cryo-electron microscopy enabled the assembly of a 12-subunit hybrid structure of Pol I, but also showed the limits of this technique. It did not allow to unambiguously position the two domains of the transiently bound and partly flexible A49/34.5 subcomplex in the EM density (Appendix 3). An alternative approach could be to perform crosslinking studies, to determine interactions between residues in A49/34.5 and the Pol I surface. Moreover, native mass spectrometry was shown to be very efficient to study the binding of A49/34.5 to Pol I. A recent publication of the Pol II-bound factor TFIIF showed that the combination of crosslinking together with mass spectrometry is an extremely powerful tool to study the architecture of multi-component complexes (Chen et al., 2010), which could be applied to the Pol I system.

To summarize, the results presented in this thesis pave the way for additional studies, to unravel the mechanism of RNA polymerase cleavage and proofreading activity, as well as gene-class specific promoter recognition by the different eukaryotic polymerases.

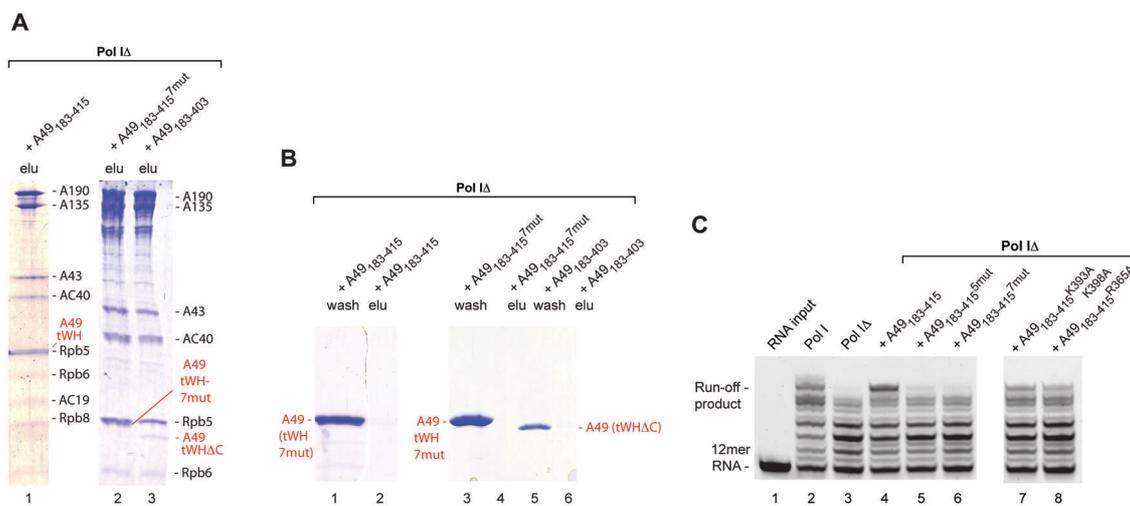
## Appendix (unpublished results)

### 1 Functional analysis of A49 tWH variants

Since dsDNA-binding was impaired by the mutation of several residues in the A49 tWH domain (chapter IV, 2.4), it was interesting to investigate the effect of these mutations on Pol I binding and processivity (chapter IV, 2.6).

Two mutations in subdomain WH2 that impair dsDNA binding of the tWH domain (K393A and R365A, chapter IV, Fig. 41A) do not disturb Pol I processivity in this assay (Fig. 53C, lanes 7-8). In contrast, multiple point mutations on WH1, which do not affect dsDNA (chapter IV, Fig. 41A) and Pol I binding (Fig. 53A, lane 2), displaying a minor effect on Pol I processivity, which leads to a reduction of the run-off product (Fig. 53C, lane 6).

Interestingly, an A49 tWH variant, lacking the C-terminal tail of A49 (chapter IV, Fig. 45), which could not stimulate full Pol I processivity (chapter IV, Fig. 51), binds to Pol I in a substoichiometric way (Fig. 53A, lane 3). Therefore, the functional defect of this variant, could also be related to its impaired stability on the Pol I surface.

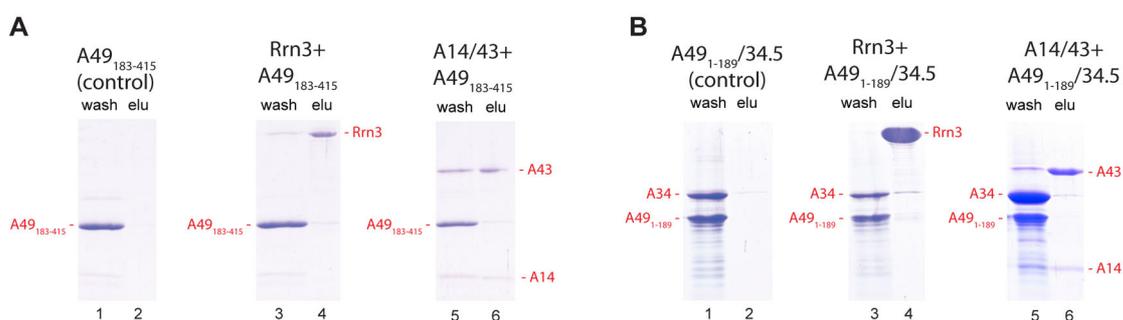


**Figure 53. Functional studies of A49 tWH variants**

(A) Pull-down experiments as described in chapter IV, Figure 46. A49 tWH (lane 1) and A49 WH1-7mut, Q309A/F310A/R312A/S313A/K314A/D135A/R316A (lane 2) bind to Pol I. Both proteins migrate close to the Rpb5 subunit of Pol I. A49<sub>183-403</sub> binds to Pol I in a substoichiometric way (lane 3). (B) Pull-down controls as described in chapter IV, Figure 46B. (C) RNA extension assays as described in chapter IV, Figure 51. Processivity is restored by addition of A49 tWH (lane 4), A49 WH2 K393A K398A (lane 7) and A49 WH2 R365A (lane 8). A49 WH1-5mut, S289A/E292A/Q309A/S313A/K314A (lane 5) and A49 WH1-7mut, Q309A/F310A/R312A/S313A/K314A/D135A/R316A (lane 6) display a minor defect in stimulation of Pol I processivity.

## 2 Interaction studies of A49/34.5 domains

Recent experiments described a functional connection between A49/34.5 and Pol I subcomplex A14/43, as well as Pol I initiation factor Rrn3 (Beckouet et al., 2008). To test for a direct interaction, pulldown experiments were performed for both A49 tWH (Fig. 54A) and the A49/34.5 Dimerization module with the A49 linker and the A34.5 C-terminal tail (Fig. 54B). Hexahistidine-tagged Rrn3 or A14/43 did not interact with the two A49/34.5 domains, suggesting the functional dependence may be indirect or that the interaction may be weak and could not be detected with this experiment.



**Figure 54. Pulldown experiments of A49/34.5 domains**

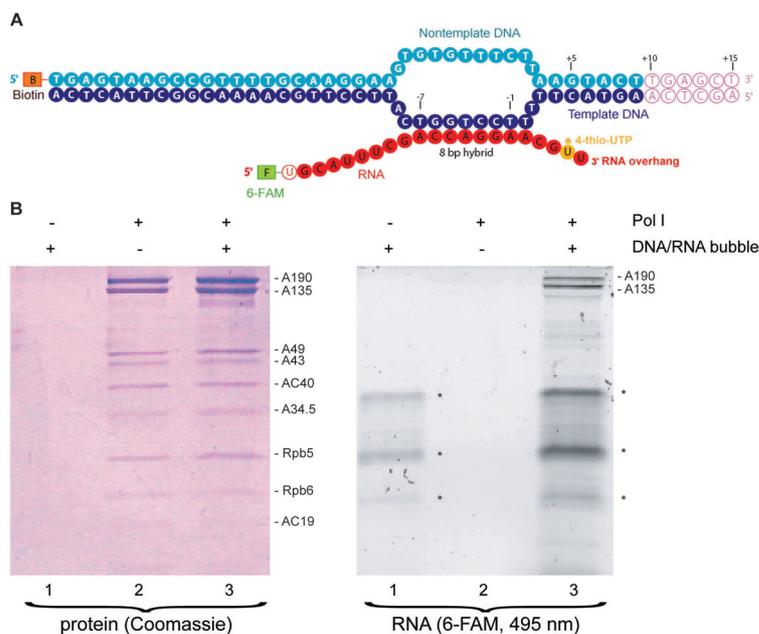
Pulldown experiments for A49/34.5 domains are described in chapter IV, Figure 46. **(A)** A49 tWH did not bind to beads (lanes 1-2, control), His<sub>6</sub>-tagged Rrn3 (lane 3-4) and His<sub>6</sub>-tagged A14/43 (lane 5-6). **(B)** The extended A49/34.5 dimerization module (A49<sub>1-189</sub>/34.5), as in A. Rrn3 and A14/43 samples were kindly provided by Claudia Blattner.

## 3 Additional Pol I cryo-EM studies

Since it was not possible to unambiguously position the two A49/34.5 crystal structures in the Pol I electron microscopic density (chapter IV, 3.3), additional cryo-EM data was collected and processed by Anselm Kusser and Sebastian Geiger. To increase the likelihood of success, two different experimental approaches were pursued. First, a minimal DNA/RNA hybrid scaffold (chapter III, Fig. 24A) was assembled with Pol I and Pol I $\Delta$ , to stabilize the highly flexible A49/34.5 domains on the Pol I surface. Second, the individual A49/34.5 domains were assembled separately with Pol I $\Delta$ , which results in two different, minimal Pol I complexes. Unfortunately, both strategies were not successful. It was not possible to obtain Pol I cryo-EM maps, which allowed for a convincing positioning of the A49/34.5 dimerization module or the A49 tWH domain structure.

## 4 Pol I crosslinking experiments

Since a major part of the EM density assigned to A49/34.5 was located near the enzyme funnel (chapter III, 2.6), flexible regions of A49/34.5 could be theoretically inserted into the polymerase pore, in proximity to the Pol I active center. To investigate this further, photocrosslinking experiments were performed. Pol I samples were assembled with biotinylated and fluorescence labeled DNA/RNA hybrid scaffolds, containing a 4-thio UTP labeled RNA, which could be crosslinked to active center residues upon UV exposure (Temiakov et al., 2000) (Fig. 55A). Polymerase-nucleic acid complexes were assembled, bound to magnetic, streptavidin-coated beads (Dengl and Cramer, 2009), extensively washed, and exposed to UV irradiation. SDS PAGE analysis of crosslinked Pol I complexes showed that A190 and A135, which together comprise the Pol I active center, are fluorescently labeled (Fig. 55B, lane 3), but no significant fluorescence signal is detected for A49 or A34.5. This preliminary experiment needs further investigation, but indicates that A49/34.5 is not in close proximity to the polymerase active center.

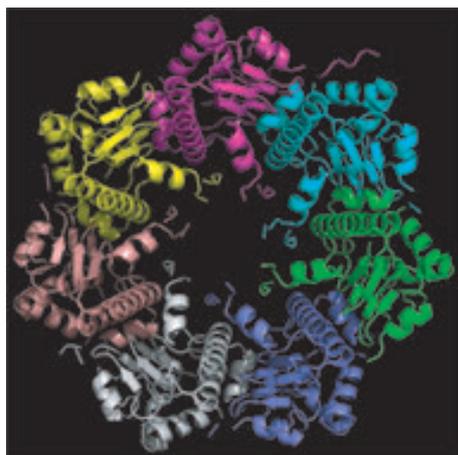


**Figure 55. Pol I crosslinking experiments**

(A) A biotinylated and 6-FAM labeled DNA/RNA hybrid scaffold, containing a 4-thio-UTP base close to the RNA 3' end. (B) Pol I-nucleic acid complexes were assembled, extensively washed, UV-exposed and analyzed by SDS PAGE (lane 3). A crosslinked DNA/RNA scaffold (lane 1) and Pol I (lane 2) are shown as controls. DNA/RNA crosslinking adducts are marked with asterisks. To prevent cleavage of the RNA 3' overhang, complexes were assembled in a low-pH transcription buffer (chapter III, 3.10), using sodium acetate pH 5.0.

## 5 Crystal structure of ClpP from *Staphylococcus aureus*

Methicillin-resistant *Staphylococcus aureus* has become a major re-emerging worldwide pathogen (Kennedy et al., 2008). The caseinolytic protein protease (ClpP) was shown to be important for the virulence of *S. aureus* (Frees et al., 2003). Chemically modified  $\beta$ -lactones were found to be potent ClpP inhibitors, reducing its hemolytic and proteolytic activities (Böttcher and Sieber, 2008). Thomas Böttcher and Sebastian Geiger developed a new ClpP preparation protocol, which lead to diffracting protein crystals. The crystal structure was determined by Sebastian Geiger (Fig. 56) and gave new insights how the protease active site is composed, enabling a new design of structurally optimized  $\beta$ -lactone inhibitors. The chemical synthesis of inhibitors and the analysis of proteolytic and hemolytic activities of ClpP were performed by Thomas Böttcher and Evelyn Zeiler. A comprehensive publication including crystal structure, inhibitor synthesis and *in vitro* studies is in preparation.



**Figure 56.** Crystal structure of ClpP from *Staphylococcus aureus*

Ribbon model of the 7-mer ClpP ring structure (Top view).

## Abbreviations

|                    |  |
|--------------------|--|
| 6AU                | 6-Azaauracil   |
| bp                 | base pairs   |
| C $\alpha$         | Carbon atom at $\alpha$ -position of the peptide chain |
| CD                 | Circular Dichroism                                     |
| CE/CORE            | core promoter element                                  |
| CF                 | core factor  |
| <i>C. glabrata</i> | <i>Candida glabrata</i>                                |
| Cy5                | cyanine-5  |
| Da                 | Dalton   |
| DFC                | dense fibrillar center                                 |
| DNA                | deoxyribonucleic acid                                  |
| ds                 | double-stranded  |
| DTT                | dithiothreitol   |
| <i>E. coli</i>     | <i>Escherichia coli</i>                                |
| EM                 | electron microscopy                                    |
| EMSA               | electrophoretic mobility shift assay                   |
| ESRF               | European Synchrotron Radiation Facility                |
| ETS                | external transcribed spacer                            |
| FAM                | 6-Carboxyfluorescein                                   |
| FC                 | fibrillar center                                       |
| GC                 | granular center  |
| h                  | hour   |
| Hepes              | 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid     |
| HRDC               | helicase and RNase D C-terminal domain                 |
| <i>H. sapiens</i>  | <i>Homo sapiens</i>                                    |
| IGS                | intergenic spacer                                      |
| IMAC               | Immobilized metal ion affinity chromatography          |
| ITS                | internal transcribed spacer                            |
| LB                 | Luria-Bertani media                                    |
| MAD                | multi-wavelength anomalous diffraction                 |
| MES                | 2-(N-morpholino)ethanesulfonic acid                    |
| min                | minute   |

|                      |  |
|----------------------|--|
| mRNA                 | messenger RNA  |
| mTOR                 | mammalian Target of Rapamycin                        |
| MW                   | molecular weight                                     |
| NaCl                 | sodium chloride                                      |
| NCS                  | noncrystallographic symmetry                         |
| Ni-NTA               | nickel nitrilotriacetic acid                         |
| NMR                  | nuclear magnetic resonance                           |
| NP-40                | nonyl phenoxy polyethoxy ethanol 40                  |
| NTP                  | nucleotide triphosphate                              |
| OB                   | oligosaccharide-binding                              |
| OD                   | optical density                                      |
| PAGE                 | polyacrylamide gel electrophoresis                   |
| PCR                  | Polymerase Chain Reaction                            |
| PDB                  | protein data bank                                    |
| PEG                  | polyethylene glycol (average molecular weight in Da) |
| pH                   | measure of acidity                                   |
| PIC                  | pre-initiation complex                               |
| PMSF                 | Phenylmethylsulfonylfluorid                          |
| Pol                  | RNA polymerase                                       |
| PT                   | proximal terminator                                  |
| rINR                 | ribosomal initiator element                          |
| RMSD                 | root mean square deviation                           |
| rRNA                 | ribosomal RNA  |
| RNA                  | ribonucleic acid                                     |
| Rpb                  | Pol II subunit (RNA polymerase B)                    |
| S                    | Svedberg (sedimentation coefficient)                 |
| SAD                  | single-wavelength anomalous diffraction              |
| <i>S. cerevisiae</i> | <i>Saccharomyces cerevisiae</i>                      |
| SDS                  | sodium dodecyl sulfate                               |
| SP                   | spacer promoter                                      |
| <i>S. pombe</i>      | <i>Schizosaccharomyces pombe</i>                     |
| SL-1                 | selectivity factor 1                                 |
| SLS                  | Swiss Light source                                   |
| ss                   | single-stranded                                      |

|        |                                 |
|--------|---------------------------------|
| TAF    | TBP-associated factor           |
| TBP    | TATA-box binding protein        |
| Term   | Terminator                      |
| TFE    | triflouroethanol                |
| TFII   | Pol II transcription factor     |
| TIF    | transcription initiation factor |
| Tris   | Trishydroxymethylaminomethan    |
| tRNA   | tRNA                            |
| tWH    | tandem winged helix-turn-helix  |
| UAF    | upstream activating factor      |
| UBF    | upstream binding factor         |
| UE/UPE | upstream promoter element       |
| WH     | winged helix-turn-helix         |
| w/v    | weight per volume               |

---

## References

- Abrahams, J.P., and Leslie, A.G. (1996). Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr D Biol Crystallogr* *52*, 30-42.
- Afonine, P.V., Grosse-Kunstleve, R.W., and Adams, P.D. (2005). A robust bulk-solvent correction and anisotropic scaling procedure. *Acta Crystallogr D Biol Crystallogr* *61*, 850-855.
- Alic, N., Ayoub, N., Landrieux, E., Favry, E., Baudouin-Cornu, P., Riva, M., and Carles, C. (2007). Selectivity and proofreading both contribute significantly to the fidelity of RNA polymerase III transcription. *Proc Natl Acad Sci U S A* *104*, 10400-10405.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* *25*, 3389-3402.
- Armache, K.J., Mitterweger, S., Meinhart, A., and Cramer, P. (2005). Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J Biol Chem* *280*, 7131-7134.
- Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* *98*, 10037-10041.
- Bartholomew, B., Durkovich, D., Kassavetis, G.A., and Geiduschek, E.P. (1993). Orientation and topography of RNA polymerase III in transcription complexes. *Mol Cell Biol* *13*, 942-952.
- Beckouet, F., Labarre-Mariotte, S., Albert, B., Imazawa, Y., Werner, M., Gadai, O., Nogi, Y., and Thuriaux, P. (2008). Two RNA polymerase I subunits control the binding and release of Rrn3 during transcription. *Mol Cell Biol* *28*, 1596-1605.
- Benesch, J.L., Ruotolo, B.T., Simmons, D.A., and Robinson, C.V. (2007). Protein complexes in the gas phase: technology for structural genomics and proteomics. *Chem Rev* *107*, 3544-3567.
- Bergfors, T. (2003). Seeds to crystals. *J Struct Biol* *142*, 66-76.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. *Nucleic Acids Res* *28*, 235-242.
- Biegert, A., Mayer, C., Remmert, M., Soding, J., and Lupas, A.N. (2006). The MPI Bioinformatics Toolkit for protein sequence analysis. *Nucleic Acids Res* *34*, W335-339.
- Bier, M., Fath, S., and Tschochner, H. (2004). The composition of the RNA polymerase I transcription machinery switches from initiation to elongation mode. *FEBS Lett* *564*, 41-46.

- Bischler, N., Brino, L., Carles, C., Riva, M., Tschochner, H., Mallouh, V., and Schultz, P. (2002). Localization of the yeast RNA polymerase I-specific subunits. *Embo J* *21*, 4136-4144.
- Blanc, E., Roversi, P., Vonrhein, C., Flensburg, C., Lea, S.M., and Bricogne, G. (2004). Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr D Biol Crystallogr* *60*, 2210-2221.
- Bottcher, T., and Sieber, S.A. (2008). Beta-lactones as specific inhibitors of ClpP attenuate the production of extracellular virulence factors of *Staphylococcus aureus*. *J Am Chem Soc* *130*, 14400-14401.
- Boukhgalter, B., Liu, M., Guo, A., Tripp, M., Tran, K., Huynh, C., and Pape, L. (2002). Characterization of a fission yeast subunit of an RNA polymerase I essential transcription initiation factor, SpRrn7h/TAF(I)68, that bridges yeast and mammals: association with SpRrn11h and the core ribosomal RNA gene promoter. *Gene* *291*, 187-201.
- Brennan, R.G. (1993). The winged-helix DNA-binding motif: another helix-turn-helix takeoff. *Cell* *74*, 773-776.
- Bric, A., Radebaugh, C.A., and Paule, M.R. (2004). Photocross-linking of the RNA polymerase I preinitiation and immediate postinitiation complexes: implications for promoter recruitment. *J Biol Chem* *279*, 31259-31267.
- Brueckner, F., Hennecke, U., Carell, T., and Cramer, P. (2007). CPD damage recognition by transcribing RNA polymerase II. *Science* *315*, 859-862.
- Brueckner, F., Ortiz, J., and Cramer, P. (2009). A movie of the RNA polymerase nucleotide addition cycle. *Curr Opin Struct Biol* *19*, 294-299.
- Brun, I., Sentenac, A., and Werner, M. (1997). Dual role of the C34 subunit of RNA polymerase III in transcription initiation. *EMBO J* *16*, 5730-5741.
- Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., *et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* *54*, 905-921.
- Budisa, N., Steipe, B., Demange, P., Eckerskorn, C., Kellermann, J., and Huber, R. (1995). High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in *Escherichia coli*. *Eur J Biochem* *230*, 788-796.
- Carter, R., and Drouin, G. (2010). The increase in the number of subunits in eukaryotic RNA polymerase III relative to RNA polymerase II is due to the permanent recruitment of general transcription factors. *Mol Biol Evol* *27*, 1035-1043.

- Chedin, S., Riva, M., Schultz, P., Sentenac, A., and Carles, C. (1998). The RNA cleavage activity of RNA polymerase III is mediated by an essential TFIIS-like subunit and is important for transcription termination. *Genes Dev* *12*, 3857-3871.
- Chen, H.-T., Warfield, L., and Hahn, S. (2007). The positions of TFIIF and TFIIE in the RNA polymerase II transcription initiation complex. *Nat Struct Mol Biol*.
- Chen, Z.A., Jawhari, A., Fischer, L., Buchen, C., Tahir, S., Kamenski, T., Rasmussen, M., Lariviere, L., Bukowski-Wills, J.C., Nilges, M., *et al.* (2010). Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *EMBO J* *29*, 717-726.
- Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G., and Thompson, J.D. (2003). Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* *31*, 3497-3500.
- Choe, S.Y., Schultz, M.C., and Reeder, R.H. (1992). In vitro definition of the yeast RNA polymerase I promoter. *Nucleic Acids Res* *20*, 279-285.
- Chung, W.H., Craighead, J.L., Chang, W.H., Ezeokonkwo, C., Bareket-Samish, A., Kornberg, R.D., and Asturias, F.J. (2003). RNA polymerase II/TFIIF structure and conserved organization of the initiation complex. *Mol Cell* *12*, 1003-1013.
- Cowtan, K. (1994). Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography *31*, 34-38.
- Cowtan, K. (2010). Recent developments in classical density modification. *Acta Crystallogr D Biol Crystallogr* *66*, 470-478.
- Cramer, P. (2004). Structure and function of RNA polymerase II. *Adv Protein Chem* *67*, 1-42.
- Cramer, P., Armache, K.J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., Damsma, G.E., Dengl, S., Geiger, S.R., Jasiak, A.J., *et al.* (2008). Structure of eukaryotic RNA polymerases. *Annu Rev Biophys* *37*, 337-352.
- Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* *292*, 1863-1876.
- De Carlo, S., Carles, C., Riva, M., and Schultz, P. (2003). Cryo-negative staining reveals conformational flexibility within yeast RNA polymerase I. *J Mol Biol* *329*, 891-902.
- de La Fortelle, E., and Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol* *276*, 472-494.
- Dengl, S., and Cramer, P. (2009). Torpedo nuclease Rat1 is insufficient to terminate RNA polymerase II in vitro. *J Biol Chem* *284*, 21270-21279.

- Dye, M.J., Gromak, N., Haussecker, D., West, S., and Proudfoot, N.J. (2006). Turnover and function of noncoding RNA polymerase II transcripts. *Cold Spring Harb Symp Quant Biol* *71*, 275-284.
- Edwards, A.M., Kane, C.M., Young, R.A., and Kornberg, R.D. (1991). Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter *in vitro*. *J. Biol. Chem.* *266*, 71-75.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* *60*, 2126-2132.
- Fath, S., Milkereit, P., Peyroche, G., Riva, M., Carles, C., and Tschochner, H. (2001). Differential roles of phosphorylation in the formation of transcriptional active RNA polymerase I. *Proc Natl Acad Sci U S A* *98*, 14334-14339.
- Fath, S., Milkereit, P., Podtelejnikov, A.V., Bischler, N., Schultz, P., Bier, M., Mann, M., and Tschochner, H. (2000). Association of yeast RNA polymerase I with a nucleolar substructure active in rRNA synthesis and processing. *J Cell Biol* *149*, 575-590.
- Fernandez-Tornero, C., Bottcher, B., Riva, M., Carles, C., Steuerwald, U., Ruigrok, R.W., Sentenac, A., Muller, C.W., and Schoehn, G. (2007). Insights into Transcription Initiation and Termination from the Electron Microscopy Structure of Yeast RNA Polymerase III. *Mol Cell* *25*, 813-823.
- Ferri, M.L., Peyroche, G., Siaux, M., Lefebvre, O., Carles, C., Conesa, C., and Sentenac, A. (2000). A novel subunit of yeast RNA polymerase III interacts with the TFIIB-related domain of TFIIB70. *Mol Cell Biol* *20*, 488-495.
- Flores, O., Ha, I., and Reinberg, D. (1990). Factors involved in specific transcription by mammalian RNA polymerase II. Purification and subunit composition of transcription factor IIF. *J Biol Chem* *265*, 5629-5634.
- Fontoura, B.M., Atienza, C.A., Sorokina, E.A., Morimoto, T., and Carroll, R.B. (1997). Cytoplasmic p53 polypeptide is associated with ribosomes. *Mol Cell Biol* *17*, 3146-3154.
- Frees, D., Qazi, S.N., Hill, P.J., and Ingmer, H. (2003). Alternative roles of ClpX and ClpP in *Staphylococcus aureus* stress tolerance and virulence. *Mol Microbiol* *48*, 1565-1578.
- Gadal, O., Mariotte-Labarre, S., Chedin, S., Quemeneur, E., Carles, C., Sentenac, A., and Thuriaux, P. (1997). A34.5, a nonessential component of yeast RNA polymerase I, cooperates with subunit A14 and DNA topoisomerase I to produce a functional rRNA synthesis machine. *Mol Cell Biol* *17*, 1787-1795.
- Gaiser, F., Tan, S., and Richmond, T.J. (2000). Novel dimerization fold of RAP30/RAP74 in human TFIIF at 1.7 Å resolution. *J Mol Biol* *302*, 1119-1127.

- Gajiwala, K.S., and Burley, S.K. (2000). Winged helix proteins. *Curr Opin Struct Biol* *10*, 110-116.
- Galtier, N., Gouy, M., and Gautier, C. (1996). SEAVIEW and PHYLO\_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Comput Appl Biosci* *12*, 543-548.
- Geiger, S.R., Kuhn, C.D., Leidig, C., Renkawitz, J., and Cramer, P. (2008). Crystallization of RNA polymerase I subcomplex A14/A43 by iterative prediction, probing and removal of flexible regions. *Acta Crystallogr Sect F Struct Biol Cryst Commun* *64*, 413-418.
- Gerber, J., Reiter, A., Steinbauer, R., Jakob, S., Kuhn, C.D., Cramer, P., Griesenbeck, J., Milkereit, P., and Tschochner, H. (2008). Site specific phosphorylation of yeast RNA polymerase I. *Nucleic Acids Res* *36*, 793-802.
- Granneman, S., and Baserga, S.J. (2005). Crosstalk in gene expression: coupling and co-regulation of rDNA transcription, pre-ribosome assembly and pre-rRNA processing. *Curr Opin Cell Biol* *17*, 281-286.
- Groft, C.M., Uljon, S.N., Wang, R., and Werner, M.H. (1998). Structural homology between the Rap30 DNA-binding domain and linker histone H5: implications for preinitiation complex assembly. *Proc Natl Acad Sci U S A* *95*, 9117-9122.
- Grummt, I. (2003). Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus. *Genes Dev* *17*, 1691-1702.
- Hahn, S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol* *11*, 394-403.
- Hanada, K., Song, C.Z., Yamamoto, K., Yano, K., Maeda, Y., Yamaguchi, K., and Muramatsu, M. (1996). RNA polymerase I associated factor 53 binds to the nucleolar transcription factor UBF and functions in specific rDNA transcription. *EMBO J* *15*, 2217-2226.
- Heck, A.J. (2008). Native mass spectrometry: a bridge between interactomics and structural biology. *Nat Methods* *5*, 927-933.
- Henras, A.K., Soudet, J., Gerus, M., Lebaron, S., Caizergues-Ferrer, M., Mouglin, A., and Henry, Y. (2008). The post-transcriptional steps of eukaryotic ribosome biogenesis. *Cell Mol Life Sci* *65*, 2334-2359.
- Higuchi, H., Yoshioka, T., and Maruyama, K. (1988). Positioning of actin filaments and tension generation in skinned muscle fibres released after stretch beyond overlap of the actin and myosin filaments. *J Muscle Res Cell Motil* *9*, 491-498.
- Holm, L., and Sander, C. (1995). Dali: a network tool for protein structure comparison. *Trends Biochem Sci* *20*, 478-480.

- Hong, M., Fuangthong, M., Helmann, J.D., and Brennan, R.G. (2005). Structure of an OhrR-ohrA operator complex reveals the DNA binding mechanism of the MarR family. *Mol Cell* 20, 131-141.
- Hsieh, Y.J., Kundu, T.K., Wang, Z., Kovelman, R., and Roeder, R.G. (1999). The TFIIC90 subunit of TFIIC interacts with multiple components of the RNA polymerase III machinery and contains a histone-specific acetyltransferase activity. *Mol Cell Biol* 19, 7697-7704.
- Hu, P., Wu, S., Sun, Y., Yuan, C.C., Kobayashi, R., Myers, M.P., and Hernandez, N. (2002). Characterization of human RNA polymerase III identifies orthologues for *Saccharomyces cerevisiae* RNA polymerase III subunits. *Mol Cell Biol* 22, 8044-8055.
- Hubbard, S.J. (1998). The structural aspects of limited proteolysis of native proteins. *Biochim Biophys Acta* 1382, 191-206.
- Huet, J., Buhler, J.M., Sentenac, A., and Fromageot, P. (1975). Dissociation of two polypeptide chains from yeast RNA polymerase A. *Proc Natl Acad Sci U S A* 72, 3034-3038.
- Huet, J., Buhler, J.M., Sentenac, A., and Fromageot, P. (1977). Characterization of ribonuclease H activity associated yeast RNA polymerase A. *J Biol Chem* 252, 8848-8855.
- Huet, J., Wyers, F., Buhler, J.M., Sentenac, A., and Fromageot, P. (1976). Association of RNase H activity with yeast RNA polymerase A. *Nature* 261, 431-433.
- Iborra, F., Huet, J., Breant, B., Sentenac, A., and Fromageot, P. (1979). Identification of two different RNase H activities associated with yeast RNA polymerase A. *J Biol Chem* 254, 10920-10924.
- Imazawa, Y., Hisatake, K., Mitsuzawa, H., Matsumoto, M., Tsukui, T., Nakagawa, K., Nakadai, T., Shimada, M., Ishihama, A., and Nogi, Y. (2005). The fission yeast protein Ker1p is an ortholog of RNA polymerase I subunit A14 in *Saccharomyces cerevisiae* and is required for stable association of Rrn3p and RPA21 in RNA polymerase I. *J Biol Chem* 280, 11467-11474.
- Jasiak, A.J., Armache, K.J., Martens, B., Jansen, R.P., and Cramer, P. (2006). Structural biology of RNA polymerase III: subcomplex C17/25 X-ray structure and 11 subunit enzyme model. *Mol Cell* 23, 71-81.
- Johnson, T.L., and Chamberlin, M.J. (1994). Complexes of yeast RNA polymerase II and RNA are substrates for TFIIS-induced RNA cleavage. *Cell* 77, 217-224.
- Kabsch, W. (1976). A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr A* 32, 922-932.
- Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Crystallogr* 26, 795-800.

- Kahl, B.F., Li, H., and Paule, M.R. (2000). DNA melting and promoter clearance by eukaryotic RNA polymerase I. *J Mol Biol* 299, 75-89.
- Kamada, K., De Angelis, J., Roeder, R.G., and Burley, S.K. (2001). Crystal structure of the C-terminal domain of the RAP74 subunit of human transcription factor IIF. *Proc Natl Acad Sci U S A* 98, 3115-3120.
- Kassavetis, G.A., Prakash, P., and Shim, E. (2010). The C53/C37 subcomplex of RNA polymerase III lies near the active site and participates in promoter opening. *J Biol Chem* 285, 2695-2706.
- Keener, J., Josaitis, C.A., Dodd, J.A., and Nomura, M. (1998). Reconstitution of yeast RNA polymerase I transcription in vitro from purified components. TATA-binding protein is not required for basal transcription. *J Biol Chem* 273, 33795-33802.
- Kennedy, A.D., Otto, M., Braughton, K.R., Whitney, A.R., Chen, L., Mathema, B., Mediavilla, J.R., Byrne, K.A., Parkins, L.D., Tenover, F.C., *et al.* (2008). Epidemic community-associated methicillin-resistant *Staphylococcus aureus*: recent clonal expansion and diversification. *Proc Natl Acad Sci U S A* 105, 1327-1332.
- Kenney, L.J. (2002). Structure/function relationships in OmpR and other winged-helix transcription factors. *Curr Opin Microbiol* 5, 135-141.
- 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.
- Killeen, M.T., and Greenblatt, J.F. (1992). The general transcription factor RAP30 binds to RNA polymerase II and prevents it from binding nonspecifically to DNA. *Mol Cell Biol* 12, 30-37.
- Kireeva, M.L., Komissarova, N., Waugh, D.S., and Kashlev, M. (2000). The 8-nucleotide-long RNA:DNA hybrid is a primary stability determinant of the RNA polymerase II elongation complex. *J Biol Chem* 275, 6530-6536.
- Kos, M., and Tollervey, D. (2010). Yeast pre-rRNA processing and modification occur cotranscriptionally. *Mol Cell* 37, 809-820.
- Kostrewa, D., Zeller, M.E., Armache, K.J., Seizl, M., Leike, K., Thomm, M., and Cramer, P. (2009). RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature* 462, 323-330.
- Kuhn, A., and Grummt, I. (1989). 3'-end formation of mouse pre-rRNA involves both transcription termination and a specific processing reaction. *Genes Dev* 3, 224-231.
- Kuhn, C.D., Geiger, S.R., Baumli, S., Gartmann, M., Gerber, J., Jennebach, S., Mielke, T., Tschochner, H., Beckmann, R., and Cramer, P. (2007). Functional architecture of RNA polymerase I. *Cell* 131, 1260-1272.

- Kulkens, T., Riggs, D.L., Heck, J.D., Planta, R.J., and Nomura, M. (1991). The yeast RNA polymerase I promoter: ribosomal DNA sequences involved in transcription initiation and complex formation in vitro. *Nucleic Acids Res* 19, 5363-5370.
- La Fortelle, E.d., and Bricogne, G. (1997). *Meth. Enzym.* 276, 472-494.
- Labhart, P. (1997). Transcript cleavage in an RNA polymerase I elongation complex. Evidence for a dissociable activity similar to but distinct from TFIIS. *J Biol Chem* 272, 9055-9061.
- Landrieux, E., Alic, N., Ducrot, C., Acker, J., Riva, M., and Carles, C. (2006). A subcomplex of RNA polymerase III subunits involved in transcription termination and reinitiation. *EMBO J* 25, 118-128.
- Langer, G., Cohen, S.X., Lamzin, V.S., and Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc* 3, 1171-1179.
- Laskowski, R.A., Moss, D.S., and Thornton, J.M. (1993). Main-chain bond lengths and bond angles in protein structures. *J Mol Biol* 231, 1049-1067.
- Liljelund, P., Mariotte, S., Buhler, J.M., and Sentenac, A. (1992). Characterization and mutagenesis of the gene encoding the A49 subunit of RNA polymerase A in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* 89, 9302-9305.
- Lorenzen, K. (2007). Optimizing macromolecular tandem mass spectrometry of large non-covalent complexes using heavy collision gases. *Int J Mass Spectrometry* 268, 198-206.
- Lorenzen, K., Vannini, A., Cramer, P., and Heck, A.J. (2007). Structural biology of RNA polymerase III: mass spectrometry elucidates subcomplex architecture. *Structure* 15, 1237-1245.
- Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982). *Molecular Cloning, a Laboratory Manual* (Cold Spring Harbor Laboratory).
- Mayer, C., Bierhoff, H., and Grummt, I. (2005). The nucleolus as a stress sensor: JNK2 inactivates the transcription factor TIF-IA and down-regulates rRNA synthesis. *Genes Dev* 19, 933-941.
- McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C., and Read, R.J. (2005). Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr* 61, 458-464.
- Meinhart, A., Blobel, J., and Cramer, P. (2003). An extended winged helix domain in general transcription factor E/II $\epsilon$  alpha. *J Biol Chem* 278, 48267-48274.

- Meka, H., Daoust, G., Arnvig, K.B., Werner, F., Brick, P., and Onesti, S. (2003). Structural and functional homology between the RNAP(I) subunits A14/A43 and the archaeal RNAP subunits E/F. *Nucleic Acids Res* 31, 4391-4400.
- Milkereit, P., and Tschochner, H. (1998). A specialized form of RNA polymerase I, essential for initiation and growth-dependent regulation of rRNA synthesis, is disrupted during transcription. *EMBO J* 17, 3692-3703.
- Moss, T. (2004). At the crossroads of growth control; making ribosomal RNA. *Curr Opin Genet Dev* 14, 210-217.
- Moss, T., Langlois, F., Gagnon-Kugler, T., and Stefanovsky, V. (2007). A housekeeper with power of attorney: the rRNA genes in ribosome biogenesis. *Cell Mol Life Sci* 64, 29-49.
- Musters, W., Knol, J., Maas, P., Dekker, A.F., van Heerikhuizen, H., and Planta, R.J. (1989). Linker scanning of the yeast RNA polymerase I promoter. *Nucleic Acids Res* 17, 9661-9678.
- Nakagawa, K., Hisatake, K., Imazawa, Y., Ishiguro, A., Matsumoto, M., Pape, L., Ishihama, A., and Nogi, Y. (2003). The fission yeast RPA51 is a functional homolog of the budding yeast A49 subunit of RNA polymerase I and required for maximizing transcription of ribosomal DNA. *Genes Genet Syst* 78, 199-209.
- Nogi, Y., Yano, R., and Nomura, M. (1991). Synthesis of large rRNAs by RNA polymerase II in mutants of *Saccharomyces cerevisiae* defective in RNA polymerase I. *Proc Natl Acad Sci U S A* 88, 3962-3966.
- Nomura, M. (2001). Ribosomal RNA genes, RNA polymerases, nucleolar structures, and synthesis of rRNA in the yeast *Saccharomyces cerevisiae*. *Cold Spring Harb Symp Quant Biol* 66, 555-565.
- Okamoto, T., Yamamoto, S., Watanabe, Y., Ohta, T., Hanaoka, F., Roeder, R.G., and Ohkuma, Y. (1998). Analysis of the role of TFIIE in transcriptional regulation through structure-function studies of the TFIIEbeta subunit. *J Biol Chem* 273, 19866-19876.
- Okuda, M., Tanaka, A., Arai, Y., Satoh, M., Okamura, H., Nagadoi, A., Hanaoka, F., Ohkuma, Y., and Nishimura, Y. (2004). A novel zinc finger structure in the large subunit of human general transcription factor TFIIE. *J Biol Chem* 279, 51395-51403.
- Okuda, M., Watanabe, Y., Okamura, H., Hanaoka, F., Ohkuma, Y., and Nishimura, Y. (2000). Structure of the central core domain of TFIIEbeta with a novel double-stranded DNA-binding surface. *EMBO J* 19, 1346-1356.
- Orlova, M., Newlands, J., Das, A., Goldfarb, A., and Borukhov, S. (1995). Intrinsic transcript cleavage activity of RNA polymerase. *Proc Natl Acad Sci U S A* 92, 4596-4600.
- Pape, T., and Schneider, T.R. (2004). HKL2MAP: a graphical user interface for phasing with SHELX programs. *Acta Crystallogr D Biol Crystallogr* 37, 843-844.

- Paule, M.R., and White, R.J. (2000). Survey and summary: transcription by RNA polymerases I and III. *Nucleic Acids Res* 28, 1283-1298.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25, 1605-1612.
- Peyroche, G., Levillain, E., Siaux, M., Callebaut, I., Schultz, P., Sentenac, A., Riva, M., and Carles, C. (2002). The A14-A43 heterodimer subunit in yeast RNA pol I and their relationship to Rpb4-Rpb7 pol II subunits. *Proc Natl Acad Sci U S A* 99, 14670-14675.
- Peyroche, G., Milkereit, P., Bischler, N., Tschochner, H., Schultz, P., Sentenac, A., Carles, C., and Riva, M. (2000). The recruitment of RNA polymerase I on rDNA is mediated by the interaction of the A43 subunit with Rrn3. *EMBO J* 19, 5473-5482.
- Prescott, E.M., Osheim, Y.N., Jones, H.S., Alen, C.M., Roan, J.G., Reeder, R.H., Beyer, A.L., and Proudfoot, N.J. (2004). Transcriptional termination by RNA polymerase I requires the small subunit Rpa12p. *Proc Natl Acad Sci U S A* 101, 6068-6073.
- Rost, B., Yachdav, G., and Liu, J. (2004). The PredictProtein server. *Nucleic Acids Res* 32, W321-326.
- Russell, J., and Zomerdijk, J.C. (2006). The RNA polymerase I transcription machinery. *Biochem Soc Symp* 73, 203-216.
- Sadhale, P.P., and Woychik, N.A. (1994). C25, an essential RNA polymerase III subunit related to the RNA polymerase II subunit RPB7. *Mol Cell Biol* 14, 6164-6170.
- Sali, A., Potterton, L., Yuan, F., van Vlijmen, H., and Karplus, M. (1995). Evaluation of comparative protein modeling by MODELLER. *Proteins* 23, 318-326.
- Schneider, D.A., Michel, A., Sikes, M.L., Vu, L., Dodd, J.A., Salgia, S., Osheim, Y.N., Beyer, A.L., and Nomura, M. (2007). Transcription elongation by RNA polymerase I is linked to efficient rRNA processing and ribosome assembly. *Mol Cell* 26, 217-229.
- Schneider, T.R., and Sheldrick, G.M. (2002). Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr* 58, 1772-1779.
- Schultz, P., Celia, H., Riva, M., Sentenac, A., and Oudet, P. (1993). Three-dimensional model of yeast RNA polymerase I determined by electron microscopy of two-dimensional crystals. *Embo J* 12, 2601-2607.
- Shpakovskii, G.V., and Shematorova, E.K. (1999). [Characteristics of the cDNA of the *Schizosaccharomyces pombe* rpa43+ gene: structural similarity of the Rpa43 subunit of RNA-polymerase I with the Rpc25 subunit of RNA-polymerase III]. *Bioorg Khim* 25, 791-796.

- Siaut, M., Zaros, C., Levivier, E., Ferri, M.L., Court, M., Werner, M., Callebaut, I., Thuriaux, P., Sentenac, A., and Conesa, C. (2003). An Rpb4/Rpb7-like complex in yeast RNA polymerase III contains the orthologue of mammalian CGRP-RCP. *Mol Cell Biol* *23*, 195-205.
- Soding, J., Biegert, A., and Lupas, A.N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* *33*, W244-248.
- Steffan, J.S., Keys, D.A., Vu, L., and Nomura, M. (1998). Interaction of TATA-binding protein with upstream activation factor is required for activated transcription of ribosomal DNA by RNA polymerase I in *Saccharomyces cerevisiae* in vivo. *Mol Cell Biol* *18*, 3752-3761.
- Tan, S., Conaway, R.C., and Conaway, J.W. (1995). Dissection of transcription factor TFIIIF functional domains required for initiation and elongation. *Proc Natl Acad Sci U S A* *92*, 6042-6046.
- Tanaka, A., Watanabe, T., Iida, Y., Hanaoka, F., and Ohkuma, Y. (2009). Central forkhead domain of human TFIIIE beta plays a primary role in binding double-stranded DNA at transcription initiation. *Genes Cells* *14*, 395-405.
- Temiakov, D., Montesana, P.E., Ma, K., Mustaev, A., Borukhov, S., and McAllister, W.T. (2000). The specificity loop of T7 RNA polymerase interacts first with the promoter and then with the elongating transcript, suggesting a mechanism for promoter clearance. *Proc Natl Acad Sci U S A* *97*, 14109-14114.
- Thuillier, V., Stettler, S., Sentenac, A., Thuriaux, P., and Werner, M. (1995). A mutation in the C31 subunit of *Saccharomyces cerevisiae* RNA polymerase III affects transcription initiation. *Embo J* *14*, 351-359.
- Todone, F., Brick, P., Werner, F., Weinzierl, R.O., and Onesti, S. (2001). Structure of an archaeal homolog of the eukaryotic RNA polymerase II RPB4/RPB7 complex. *Mol Cell* *8*, 1137-1143.
- Trumtel, S., Leger-Silvestre, I., Gleizes, P.E., Teulieres, F., and Gas, N. (2000). Assembly and functional organization of the nucleolus: ultrastructural analysis of *Saccharomyces cerevisiae* mutants. *Mol Biol Cell* *11*, 2175-2189.
- Tschochner, H. (1996). A novel RNA polymerase I-dependent RNase activity that shortens nascent transcripts from the 3' end. *Proc Natl Acad Sci U S A* *93*, 12914-12919.
- Ushijima, R., Matsuyama, T., Nagata, I., and Yamamoto, K. (2008). Nucleolar targeting of proteins by the tandem array of basic amino acid stretches identified in the RNA polymerase I-associated factor PAF49. *Biochem Biophys Res Commun* *369*, 1017-1021.
- van den Heuvel, R.H., van Duijn, E., Mazon, H., Synowsky, S.A., Lorenzen, K., Versluis, C., Brouns, S.J., Langridge, D., van der Oost, J., Hoyes, J., and Heck, A.J. (2006). Improving the

performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. *Anal Chem* *78*, 7473-7483.

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

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

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

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

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

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

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

Yamamoto, K., Yamamoto, M., Hanada, K., Nogi, Y., Matsuyama, T., and Muramatsu, M. (2004). Multiple protein-protein interactions by RNA polymerase I-associated factor PAF49 and role of PAF49 in rRNA transcription. *Mol Cell Biol* *24*, 6338-6349.

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

# Curriculum Vitae

**Sebastian R. Geiger**



## Personal Data

Date of birth: 8<sup>th</sup> January 1980  
Place of birth: Munich, Germany  
Nationality: German  
Marital status: not married  
Email: geiger@lmb.uni-muenchen.de

## Research Experience

---

- Since 01/2006      **Doctoral thesis**  
Gene Center of the Ludwig-Maximilians-University, Munich  
Advisor:          Prof. Dr. Patrick Cramer  
Topic:              Structure and function of RNA polymerase I subunits
- 06/2005 – 12/2005      **Diploma thesis**  
Gene Center of the Ludwig-Maximilians-University, Munich  
Advisor:          Prof. Dr. Patrick Cramer  
Topic:              Structural studies of the Med20/Med18 dimer and other components of the Mediator head module
- 08/2003 – 03/2004      **Research internship**  
University of California, San Francisco, USA  
Advisor:          Prof. Dr. James H. McKerrow  
Topics:             Laboratory projects focused on the analysis of enzymes involved in the host-parasite interaction of *Plasmodium berghei*, *Fasciola hepatica* and *Trypanosoma brucei*
- 09/2001 – 12/2005      **Research assistant**  
GPC Biotech AG, Munich  
Tasks:              Laboratory work in the section biotechnical automation, Scientific literature research

## Education

---

- 10/2000 – 12/2005      **Diploma studies in Chemistry and Biochemistry**  
Ludwig-Maximilians-University, Munich  
Title:                Dipl. chem.
- 07/1999 – 04/2000      **Military service**  
1<sup>st</sup> Mountain division, Murnau
- 09/1990 – 06/1999      **High school**  
Christoph-Probst-Gymnasium, Gilching (Abitur)

## Publications

---

- **S.R. Geiger**, K. Lorenzen, A. Schreieck, P. Hanecker, D. Kostrewa, A. Heck and P. Cramer, RNA polymerase I contains a TFIIF-related promoter-binding subcomplex, *Mol Cell*, *accepted*, (2010).
- **S.R. Geiger**, C.-D. Kuhn, C. Leidig, J. Renkawitz and P. Cramer, Crystallization of RNA polymerase I subcomplex A14/A43 by iterative prediction, probing and removal of flexible regions, *Acta Cryst. F64*, 413–418 (2008).
- P. Cramer, K.-J. Armache, S. Baumli, S. Benkert, F. Brueckner, C. Buchen, G.E. Damsma, S. Dengl, **S.R. Geiger**, A.J. Jasiak, A. Jawhari, S. Jennebach, T. Kamenski, H. Kettenberger, C.-D. Kuhn, E. Lehmann, K. Leike, J. Sydow and A. Vannini, Structure of Eukaryotic RNA Polymerases, *Annu. Rev. Biophys.* 37, 337-352 (2008).
- C.M. Stack, C.R. Caffrey, S.M. Donnelly, A. Seshadri, J. Lowther, J.F. Tort, P.R. Collins, M.W. Robinson, W. Xu, J.H. McKerrow, C.S. Craik, **S.R. Geiger**, R. Marion, L.S. Brinen and J.P. Dalton, Structural and functional relationships in the virulence-associated cathepsin L proteases of the parasitic liver fluke, *Fasciola hepatica*, *J. Biol. Chem.* 283, 9896-9908 (2008).
- C.-D. Kuhn, **S.R. Geiger**, S. Baumli, M. Gartmann, J. Gerber, S. Jennebach, T. Mielke, H. Tschochner, R. Beckmann and P. Cramer, Functional Architecture of RNA Polymerase I, *Cell* 131, 1260-1272 (2007).
- L. Larivière, **S. Geiger**, S. Hoepfner, S. Röther, K. Sträßer and P. Cramer, Structure and TBP binding of the Mediator head subcomplex Med8–Med18–Med20, *Nat. Struct. Mol. Biol.* 13, 895 – 901 (2006).

## Stipends and Awards

---

- Member and Student representative of the International Doctorate Program NanoBioTechnology (2006-2009) - [www.cens.de/doctorate-program](http://www.cens.de/doctorate-program)
- PhD scholarship of the Elite Network of Bavaria
- Winner of the CeNS Publication Award 2008

## Languages

---

German (native), English (fluent), French (basic)

## Conferences

---

- 7<sup>th</sup> International Biennial Conference on RNA Polymerase I & III, Warrenton, USA, 3<sup>th</sup> to 7<sup>th</sup> June 2010. Lecture: RNA polymerase I contains TFIIE- and TFIIF-like domains
- 8<sup>th</sup> International Conference on Ribosome Synthesis, Regensburg, Germany, 26<sup>th</sup> to 30<sup>th</sup> August 2009. Lecture: Functional architecture of RNA polymerase I
- 25<sup>th</sup> European Crystallography Meeting, Istanbul, Turkey, 16<sup>th</sup> to 21<sup>th</sup> August 2009. Oral presentation: Functional architecture of RNA polymerase I
- CCP4 study weekend, Leeds, UK, 3<sup>rd</sup> to 5<sup>th</sup> January 2008.
- 57<sup>th</sup> Nobel Laureate Meeting in Physiology and Medicine, Lindau, Germany, 1<sup>st</sup> to 6<sup>th</sup> July 2007.
- CeNS conference ‘Emerging Nanosystems’, Venice, Italy, 25<sup>th</sup> to 29<sup>th</sup> September 2006. Poster presentation: Structural studies on RNA polymerase I