

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

Structure of the Complete RNA Polymerase II Elongation Complex and its Interaction with the Elongation Factor TFIIIS



Hubert Kettenberger
aus Burghausen

2005

Erklärung

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

Ehrenwörtliche Versicherung

Diese Dissertation wurde selbständig und ohne unerlaubte Hilfe erarbeitet.

München, den 18. Februar 2005

A handwritten signature in black ink that reads "Hubert Kettenberger". The signature is written in a cursive style with a prominent loop at the end of the last name.

Hubert Kettenberger

Dissertation eingereicht am 25. Februar 2005

1. Gutachter: Prof. Dr. Patrick Cramer
2. Gutachter: Prof. Dr. Karl-Peter Hopfner

Mündliche Prüfung am 27. April 2005

ACKNOWLEDGEMENTS

This work is the result of a very enjoyable period in the Gene Centre in the group of Prof. Patrick Cramer. I am most grateful to Patrick for introducing me to the art and science of crystallography and for giving me a challenging and most exciting project to work on in an excellent environment. Patrick's cheerful and encouraging attitude made working in his group a very pleasant experience.

Substantial parts of this work were made possible by work of Karim Armache. He pioneered in establishing key techniques in the young group, and his structure of the complete RNA polymerase II was indispensable for solving larger structures such as those in my thesis.

Moreover, I am grateful for both scientific and social support from all the members of the Cramer lab. It was a precious fortune to work in a group of people who readily collaborate, share ideas and are great folks to work and hang out with. Special thanks to Claudia Buchen for getting me started with biochemical labwork and to Sabine Höppner, Sonja Baumli, Erika Vojnic, Karim Armache and Tomislav Kamenski for their friendship. Many thanks to Florian Brückner for continuing on the Pol II project.

I would also like to thank Prof. Michael Famulok and his group for very good collaboration, as well as Prof. Karl-Peter Hopfner for help with computing issues and for being in my thesis committee, together with Prof. Dirk Eick.

I am most indebted to Felix Hartlepp for being a central part of my life. He was the one who endured my „biochemical questions“ at almost any time of the day.

My parents are the ones who deserve my deepest gratefulness. They supported me mentally, materially and cake-wise during the past nine years of my studies and thus contributed vitally to this episode of my life.

MEINEN ELTERN

SUMMARY

PUBLICATIONS

PART 1: INTRODUCTION

1	Flow of genetic information	1
2	Discovery of RNA polymerases	2
3	Three classes of cellular RNA polymerases	2
4	Mechanism of DNA-dependent RNA polymerases	3
5	Transcription factors regulate RNA polymerases	4
6	Composition of multi-subunit polymerases.....	4
7	Structure determination of RNA polymerases	6
7.1	Single-subunit RNA polymerases.....	6
7.2	Multi-subunit RNA polymerases.....	9
8	Structure of Pol II and comparison with other multi-subunit RNA polymerases.....	14
8.1	Core Pol II and the subcomplex Rpb4/Rpb7.....	17
9	The C-terminal domain of Pol II	19
10	The eukaryotic transcription cycle	20
10.1	Preinitiation and initiation	20
10.2	Promoter clearance and early elongation	21
10.3	Elongation and the nucleotide addition cycle.....	22
10.4	Termination and mRNA processing in eukaryotes	23
10.5	Pol II recycling	24
11	Elongation factors that regulate Pol II activity	24
12	Discovery and function of TFIIIS.....	27
13	Transcription <i>in vitro</i>	30
14	Structure determination of multi-subunit complexes.....	30
15	Open questions and scope of this work.....	33

PART 2: RESULTS AND DISCUSSION

16	RNA polymerase II-nucleic acid complexes.....	34
16.1	Assembly of complete Pol II-bubble-RNA complexes	34
16.2	Characterisation of Pol II-DNA/RNA complexes	37
16.3	Crystallisation of Pol II-nucleic acid complexes	39
16.4	Characterisation of complete Pol II-DNA/RNA cocrystals	43
16.5	Data collection	44
16.6	Phasing and refinement	47
16.7	Overview of nucleic acid structure.....	50
16.8	DNA unwinding	53
16.9	RNA displacement and exit	55
16.10	Initiation-elongation transition	57
16.11	NTP binding, selection, and incorporation	58
16.12	Comparison with a single-subunit RNA polymerase	62
17	The RNA polymerase II-TFIIS complex.....	64
17.1	Purification of TFIIS.....	64
17.2	Structural analysis.....	66
17.3	Induced folding of TFIIS	68
17.4	TFIIS extends from a polymerase jaw to the active site	72
17.5	TFIIS opens a crevice in the funnel.....	73
17.6	TFIIS domain III inserts into the pore	73
17.7	The TFIIS acidic hairpin complements the polymerase active site.....	74
17.8	RNA cleavage.....	76
17.9	Switching between polymerisation and cleavage.....	77
17.10	Proofreading.....	77
17.11	Pore restriction and RNA backtracking	78
17.12	TFIIS remodels the polymerase active centre	78
17.13	TFIIS realigns RNA	80
17.14	Functional conformations of Pol II.....	82
17.15	Conservation of transcript cleavage factors.....	82
18	Assays for the composition of complex crystals.....	85

18.1	Fluorescence-based detection of nucleic acids in the crystals.....	85
18.2	Monitoring protein soaking into crystals.....	89
19	Isolation and crystallisation of Pol II without CTD.....	91
20	Conclusion and outlook.....	95

PART 3: MATERIALS AND METHODS

21	Isolation of RNA Pol II from yeast.....	96
21.1	Yeast fermentation	96
21.2	Purification of RNA Pol II.....	97
21.2.1	Buffers.....	97
21.2.2	Harvesting and storage of yeast	98
21.2.3	Purification - day 1 (lysis and heparin column).....	98
21.2.4	Purification - day 2 (immunoaffinity column).....	99
21.2.5	Purification - day 3 (anion exchange column).....	99
21.2.6	Preparation of Pol II immunoaffinity resin.....	99
22	Purification of the subcomplex Rpb4/7	100
23	Electrophoretic mobility shift assay.....	101
24	Transcription assay	102
25	Introduction of a TEV proteolysis site into Rpb1.....	102
25.1	Construct design and mutagenesis	102
25.2	Yeast transformation.....	104
25.3	Isolation of Pol II Δ CTD	106
26	Cloning, mutagenesis and purification of TFIIS.....	107
26.1	Cloning and mutagenesis.....	107
26.2	Expression and purification	107
27	Preparation of Pol II-nucleic acid complexes.....	108
28	Crystal growth and treatment.....	109
28.1	Preparation of substrate-containing Pol II cocrystals	109
28.2	Preparation Pol II-TFIIS complex crystals.....	110
29	Fluorescence-based assays for complex crystals	110
29.1	Detection of nucleic acids in crystals.....	110
29.2	Monitoring TFIIS soaking into RNA polymerase II crystals.....	111
30	X-Ray analysis	112
30.1	Data collection	112

TABLE OF CONTENTS

30.2	Data Reduction	112
30.3	Molecular replacement and phasing.....	113
30.4	Anomalous Maps	113
30.5	Model Building.....	113
30.5.1	Pol II-TFIIS complex	113
30.5.2	Pol II-nucleic acid complexes.....	114
30.5.3	Refinement.....	114

REFERENCES

CURRICULUM VITAE

Summary

This thesis describes crystal structures of complete, 12-subunit yeast RNA polymerase II (Pol II) in complex with a synthetic transcription bubble and product RNA, with an NTP substrate analogue, and in complex with the transcription elongation factor TFIIS. The structure of the Pol II-transcription bubble-RNA complex reveals incoming template and non-template DNA, a seven base-pair DNA-RNA hybrid, and three nucleotides each of separating DNA and RNA. Based on this structure, those parts of Pol II were identified which are involved in separating template DNA from non-template DNA before the active site, and DNA from product RNA at the upstream end of the DNA-RNA hybrid. In both instances, strand separation can be explained by Pol II-induced duplex distortions. Only parts of the complete transcription bubble present in the complexes are ordered in the crystal structure, explaining the way in which high processivity of Pol II is reconciled with rapid translocation along the DNA template. The presence of an NTP substrate analogue in a conserved putative pre-insertion site was unveiled in a Pol II-transcription bubble-RNA complex crystal soaked with the substrate analogue GMPCPP. The structure of the Pol II-TFIIS complex was obtained from Pol II crystals soaked with TFIIS. TFIIS extends from the Pol II surface to the active site and complements the active site with two essential and invariant acidic residues for hydrolytic RNA cleavage. TFIIS also induces extensive structural changes in Pol II that reposition nucleic acids, in particular RNA, near the active centre. These results support the idea that Pol II contains a single tuneable active site for RNA polymerisation and cleavage. The technical obstacles imposed by crystal structure determination of large, transient protein-DNA-RNA complexes were overcome by two novel, fluorescence-based assays to monitor and optimise the composition of the crystals. Both assays are not limited to Pol II complexes, but can serve as a general tool for the crystallographic community.

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

- Armache K. J., Kettenberger H., and Cramer P. (2003)
Architecture of initiation-competent 12-subunit RNA polymerase II.
Proc. Natl. Acad. Sci. USA 100, 6964-6968.
- Kettenberger H., Armache K. J., and Cramer P. (2003)
Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage.
Cell 114, 347-357.
- Kettenberger H., Armache K. J., and Cramer P. (2004)
Complete RNA Polymerase II Elongation Complex Structure and Its Interactions with NTP and TFIIS.
Mol Cell 16, 955-965.
- Armache K. J., Kettenberger H., and Cramer P. (2005)
The dynamic mRNA elongation machinery.
Curr. Op. Struct. Biol.; in press
- Kettenberger H. and Cramer P. (2005)
Sensitive and selective detection of nucleic acids in crystals of macromolecular complexes.
Manuscript in preparation
- Kettenberger H., Famulok M. and Cramer P. (2005)
Crystal structure of RNA Polymerase II bound to an RNA inhibitor.
Manuscript in preparation

Introduction

1 FLOW OF GENETIC INFORMATION

Genes specify the kinds of proteins that are made by cells. However, DNA is not the direct template for protein synthesis. Rather, the templates for protein synthesis are RNA molecules. A special class of RNA, messenger RNA (mRNA), is the information-carrying intermediate in protein synthesis (Brenner *et al.*, 1961). Other classes of RNA include transfer RNA (tRNA) and ribosomal RNA (rRNA) which are involved in protein biosynthesis, and small nuclear RNAs (snRNAs) which play a key role in directing the alignment of splice sites and in mediating the splicing process (see chapter 10.4). The “molecular machines” producing all forms of cellular RNA are RNA polymerases, which take instructions from DNA templates (“DNA dependent RNA polymerases”). In the case of mRNAs, this process of transcription is followed by translation, the synthesis of proteins according to instructions given by mRNA templates. Thus, the flow of genetic information in normal cells, enunciated by F. Crick in 1958 as “The central dogma of molecular biology” (Crick, 1970), proceeds as illustrated in Figure 1.



Figure 1: The central dogma of molecular biology

The concept of mRNA as the information-carrying intermediate in protein synthesis was proposed by F. Jacob and J. Monod (Jacob, 1961) and experimentally confirmed by S. Brenner, F. Jacob and M. Meselson (Brenner *et al.*, 1961). By that time it was evident that an intermediate is necessary, which transports the information stored by DNA in the nucleus to the protein-synthesising ribosomes in the cytoplasm of eukaryotic cells. From studies of time-dependent expression of the inducible protein β -galactosidase, they hypothesised that the messenger molecule must be a very short-lived polynucleotide. Jacob and Monod proposed, based on infection studies of *E. coli* with T2 bacteriophages, that the base composition of the messenger is analogous to the base composition of the DNA that codes for it (Jacob, 1961). Variable lengths of messenger molecules reflect variable lengths of proteins occurring in a cell. The rapidly synthesised and degraded messenger molecules were proposed to be transiently associated with ribosomes. Jaques Monod and François Jacob received the 1965 Nobel Prize for medicine together with André Lwoff for “their discoveries concerning genetic control of enzyme and virus synthesis”.

2 DISCOVERY OF RNA POLYMERASES

The novel concept of mRNA fuelled the search for an enzyme that produces this molecule in a DNA-templated manner. Such an enzyme was discovered independently by Jerard Hurwitz (Hurwitz *et al.*, 1961) and Samuel Weiss (Weiss and Nakamoto, 1961). Both termed this enzyme “RNA polymerase”. The earliest evidence for the existence of DNA dependent RNA polymerases was the finding that the base composition of newly synthesised RNA is complementary to that of the DNA template. Hybridisation experiments revealed that also the RNA sequence is complementary to its DNA template (Hall and Spiegelman, 1961).

3 THREE CLASSES OF CELLULAR RNA POLYMERASES

While in prokaryotes all RNA is synthesised by a single kind of polymerase, the work is shared by three types of RNA polymerases in the nucleus of eukaryotes. These three polymerases differ in promoter specificity, localisation and susceptibility to inhibitors. The three types of nuclear RNA polymerases were identified in crude cell lysate fractionated by chromatography over an anion exchange (DEAE) column. At three different salt concentrations, fractions were obtained which catalysed RNA synthesis. These fractions were classified according to their sensitivity to α -amanitin, a toxic cyclic octapeptide from the mushroom *Amanita phalloides*. (Roeder, 1974a; Roeder, 1974b; Schwartz and Roeder, 1974).

RNA polymerase (Pol) I is insensitive against α -amanitin, Pol III needs 10 $\mu\text{g}/\text{mL}$ of α -amanitin for complete inhibition, while Pol II is inhibited by concentrations as low as 0.5 $\mu\text{g}/\text{mL}$ of α -amanitin (Schwartz *et al.*, 1974). RNA polymerase I transcribes the tandem array of genes for 18S, 5.8S and 28S ribosomal RNA and is located in the nucleolus. The other ribosomal RNA molecule, 5S rRNA, as well as several small RNAs and all transfer RNA are synthesised by RNA polymerase III in the nucleoplasm (reviewed in Lassar *et al.*, 1983). RNA polymerase II is also located in the nucleoplasm, possibly clustered in “transcriptional hotspots” (Cook, 1999; Iborra *et al.*, 1996; Lodish, 1996) and transcribes all mRNA, as well as several snRNAs.

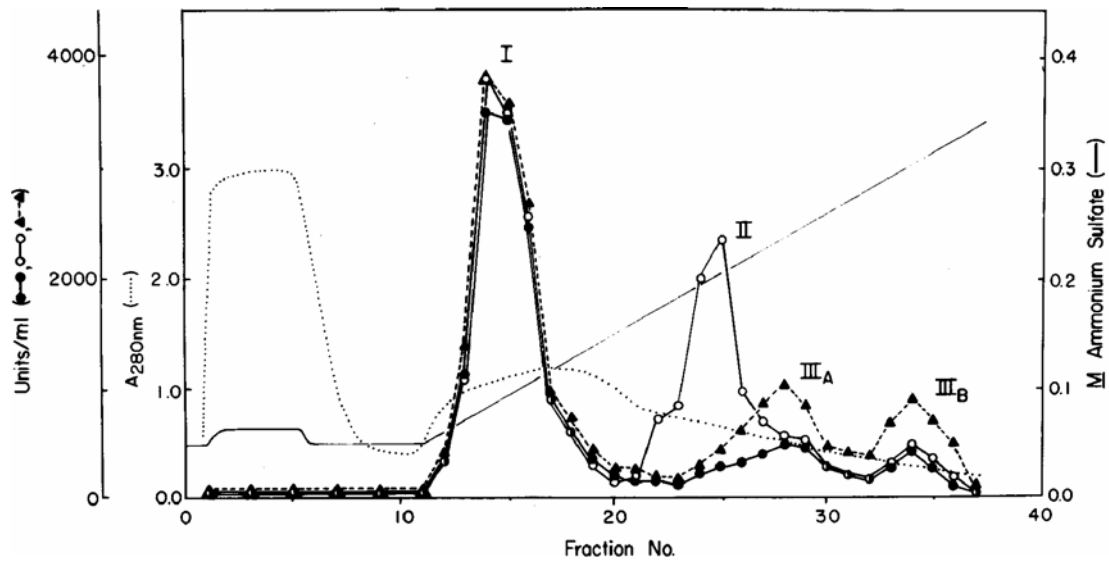


Figure 2: DEAE anion exchange chromatography of RNA polymerases from mouse myeloma cells (Schwartz and Roeder, 1974). Fractions were collected and the RNA elongation activity was measured with calf thymus DNA in the absence (○) or presence (●) of 0.5 $\mu\text{g}/\text{mL}$ of α -amanitin or with poly[(d(AT))] DNA in the presence of 0.5 $\mu\text{g}/\text{mL}$ of α -amanitin (▲). —, ammonium sulphate concentration;, absorbance at 280 nm.

4 MECHANISM OF DNA-DEPENDENT RNA POLYMERASES

All DNA-dependent RNA polymerases are able to generate RNA from a DNA template in 5'→3' direction independent of a primer. At the site of transcription, the double-stranded DNA is locally melted to form a “transcription bubble” (Figure 3). RNA polymerases proceed along the DNA, opening the downstream end of the transcription bubble. Upstream of the polymerase, template and non-template DNA re-anneal. The RNA transcript is separated from the template after a short DNA-RNA hybrid region. RNA synthesis has to be processive, i.e. RNA polymerases must not dissociate before synthesis of transcripts of up to several thousand bases in length is complete. The fidelity of RNA polymerases is several orders of magnitude lower than the fidelity of DNA polymerases (Jeon and Agarwal, 1996; Thomas *et al.*, 1998), likely because an erratic RNA is rapidly turned over and a misincorporation is not further propagated. The basis of discrimination between the correct and incorrect base is the slow addition of the next nucleotide to the mismatched terminus (Thomas *et al.*, 1998). The chain elongation reaction proceeds in an S_N2 -type fashion and depends on two metal ions (see chapter 10.3). RNA synthesis can be interrupted by pausing and/or arrest, which co-occur with a backward movement of the polymerase, termed “backtracking” (Figure 3).

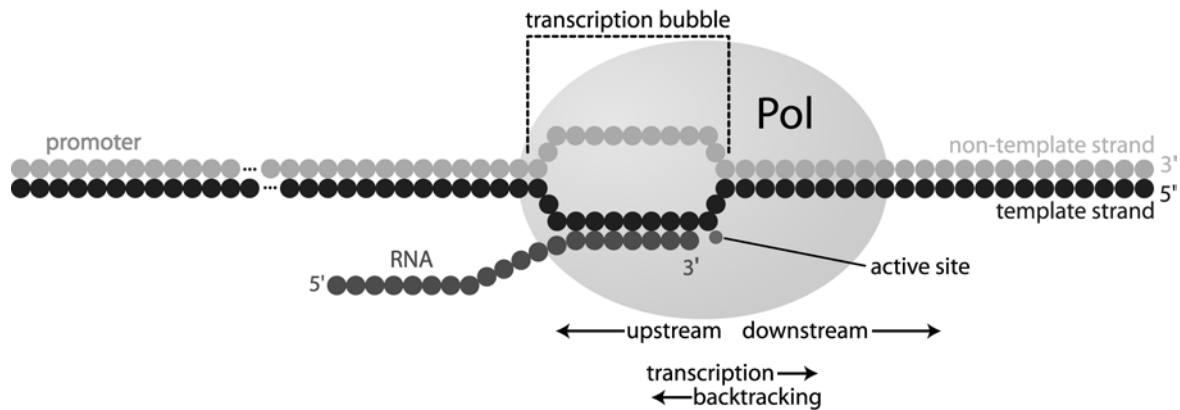


Figure 3: Schematic view of a transcribing DNA-dependent RNA polymerase.

5 TRANSCRIPTION FACTORS REGULATE RNA POLYMERASES

With all the information of tens of thousands of genes being transcribed by a single enzyme, the necessity for precise regulation is obvious. Therefore, a plethora of transcription regulatory factors exists to activate or repress the synthesis, processing and degradation of a particular set of mRNAs at a given time. Regulation at the level of transcription is indeed the major regulatory principle of protein biosynthesis. A large number of gene-specific transcription factors regulate the expression of certain proteins at a certain time. At these factors, a manifold of cellular signalling pathways converge and influence each other. Both the temporal and quantitative pattern of gene expression is subject to stringent control, which is essential for all cell processes such as metabolism, proliferation, differentiation and the response to environmental signals. Defective regulation of transcription is responsible for a number of serious diseases including cancer (Gnatt, 2002; Hurst, 1996).

6 COMPOSITION OF MULTI-SUBUNIT POLYMERASES

In contrast to single-subunit RNA polymerases like those found in bacteriophages or in mitochondria, prokaryotic and eukaryotic RNA polymerases consist of multiple subunits (Table 1). The 12 subunits of *S. cerevisiae* RNA Pol II, as a representative for eukaryotic RNA polymerases, sum up to a mass of about 500 kDa. There is evidence that multi-subunit RNA polymerases evolved from a single-subunit ancestor, based on the following set of observations: The active site of multi-subunit RNA polymerases is formed by elements from the two large subunits (Figure 7), (Cramer *et al.*, 2000; Cramer *et al.*, 2001).

Table 1: Subunit composition of multi-subunit RNA polymerases.
Adapted from (Cramer, 2002b)

Eukaryotes			Archaea	Bacteria	Class*
Pol I	Pol II	Pol III			
A190	Rpb1	C160	A'+A''	β'	core
A135	Rpb2	C128	B (B'+B'')	β	core
AC40	Rpb3	AC40	D	α	core
AC19	Rpb11	AC19	L	α	core
Rpb6	Rpb6	Rpb6	K	ω	core/common
Rpb5	Rpb5	Rpb5	H	—	common
Rpb8	Rpb8	Rpb8	—	—	common
Rpb10	Rpb10	Rpb10	N	—	common
Rpb12	Rpb12	Rpb12	P	—	common
A12.2	Rpb9	C11	X	—	
A14**	Rpb4	—	F		
A43**	Rpb7	C25	E		
+ two others		+ four others	+one other	—	

*Core, sequence partially homologous in all RNA polymerases; common, shared by all eukaryotic RNA polymerases. ** Potential paralogues of Pol II subunits Rpb4 and Rpb7 (Cramer, 2002b).

In archaea and eubacteria, the genes for the two large RNA polymerase subunits are arranged in tandem, with the second largest subunit always preceding the largest subunit (Severinov *et al.*, 1997). Accordingly, the multi-subunit RNA polymerase structures show that the C-terminus of the second largest subunit is in close proximity of the N-terminus of the largest subunit. Indeed, the two large subunits are fused in *Helicobacter pylori* (Zakharova *et al.*, 1998). Further, a functional polypeptide can be engineered by fusing the two large subunits of the *E. coli* enzyme. (Severinov *et al.*, 1997). Analogously, the two largest subunits are responsible for substrate

binding and catalysis, while the smaller subunits adopt mainly structural roles (see below). The two largest subunits are split into three polypeptides in most archaea.

7 STRUCTURE DETERMINATION OF RNA POLYMERASES

7.1 Single-subunit RNA polymerases

The smaller size of 99 kDa (883 residues) and the composition of only one polypeptide chain of bacteriophage T7 RNA polymerase fostered structure determination seven years before a detailed atomic model of multi-subunit polymerases was possible. Although unrelated to multi-subunit RNA polymerases in both sequence and structure, catalytic principles are shared between these enzymes. T7 RNA polymerase has however modest sequence homology to the single-subunit Pol I family of DNA polymerases.

First structures of T7 RNA polymerase as free enzyme and in complex with a natural inhibitor, T7 lysozyme, were obtained already in 1993 (Sousa *et al.*, 1993) and in 1998 (Jeruzalmi and Steitz, 1998), respectively. An “early initiation complex” (Cheetham and Steitz, 1999) of T7 RNA polymerase bound to a 17 base-pair promoter DNA containing a five nucleotide single-strand template extension and a trinucleotide RNA showed an unexpected rearrangement of large portions of the enzyme (Figure 4A). Yet another drastic structural rearrangement, namely alternative folding of about 300 residues was observed in the structure of a T7 elongation complex (Tahirov *et al.*, 2002; Yin and Steitz, 2002) in 2002. Here, formation of an eight base-pair DNA-RNA hybrid in the active site induces this rearrangement (Figure 4B). In this conformation, the DNA-RNA hybrid is buried in a deep cavity and RNA emerges from a newly formed exit pathway.

DNA-RNA strand separation after a 7 bp DNA/RNA hybrid is achieved by a hydrophilic loop, the “RNA displacement loop”, that projects towards the axis of the hybrid and has a net negative charge (Tahirov *et al.*, 2002; Jiang *et al.*, 2004; Yin and Steitz, 2002). The RNA is then redirected into a positively charged RNA exit pore, which is only present in the elongation complex. An intact non-template strand, which can base-pair with the template strand upstream of the transcription bubble was shown to be essential for efficient DNA-RNA strand separation. DNA-DNA strand separation at the downstream end of the transcription bubble is achieved by rotation of the so-called “O-helix” upon pyrophosphate release. This structure also lead to a

concept of the transition from the initiation phase to processive elongation, because the major refolding results in a loss of the promoter binding site and goes along with promoter clearance.

Most recent structural data on nucleotide incorporation and translocation have been published by two independent groups and lead to complementary results (Temiakov *et al.*, 2004; Yin and Steitz, 2004). These structures include complexes of T7 Pol with template DNA, product RNA and substrate NTP, as well as the reaction product pyrophosphate. These two models form the structural basis for substrate selection and for the discrimination of ribose versus deoxyribose substrates. Dissociation of the product pyrophosphate after nucleotide addition was proposed to drive the conformational change resulting in translocation and strand separation. Thus, valuable insights into the mechanism of regulated, template-directed RNA synthesis came from studies of T7 RNA polymerase.

Taking together structural and biochemical data, each round of nucleotide addition can be divided into at least four steps (Steitz, 2004):

1. The NTP binds to the polymerase in a pre-insertion site.
2. The NTP occupies the NTP-binding site (N-site) and is basepaired to the template base.
3. Catalysis: pyrophosphate is formed and the RNA transcript is extended by one nucleotide.
4. The 3' end of the RNA translocates from the N-site to the priming site (P-site).

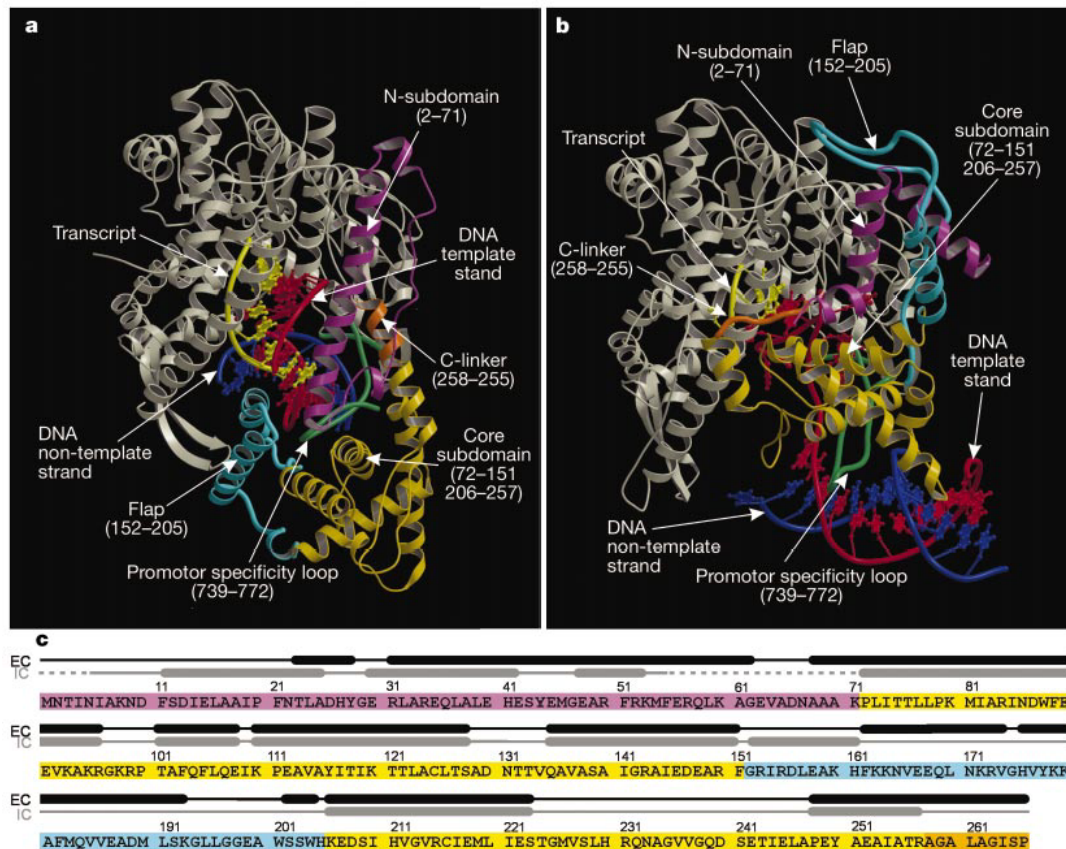


Figure 4: Transition from initiation to elongation in T7 RNA polymerase.

Ribbon diagrams of the elongation complex (a) and the early initiation complex (b) are shown in the same orientation with respect to the unaltered C-terminal portion (white). Nucleic acids are shown as ribbon (phosphate backbone) and ball-and-stick (bases) models. (c) Amino acid sequence of the N-terminal domain; secondary structures for the elongation complex and the early initiation complex are shown above (cylinders, α -helices; thin lines, coils; dashed lines, disordered regions). The colour scheme is the same as in (a) and (b). Adapted from (Tahirov *et al.*, 2002).

The set of T7 polymerase structures elucidated structural changes accompanying these four steps. The substrate analogue adenosine-5'-[(α,β)-methyleno]-triphosphate (AMPCPP) was found to bind to the catalytically inactive “open” conformation, where it is base paired with the acceptor template base and bound two conserved basic residues (R627 and K631) in the O helix. Specificity for NTPs versus dNTPs is achieved by a conserved tyrosine (Y639) which interacts with the 2'-OH of incoming NTPs (Svetlov *et al.*, 2004). The binding of the correct NTP leads to the closing conformational change. During this transition, helices O and O' move, displacing a tyrosine residue, which blocked the nucleotide addition site. Two highly conserved aspartate residues (D537 and D812) coordinate two metal ions that are required during the phosphoryl transfer reaction. The end of a nucleotide addition cycle is accompanied by an opening

conformational change, which is the reversal of the O/O' helix rotation described above. This rotation is coupled to the release of PP_i and is responsible for DNA-DNA strand separation, DNA/RNA separation and translocation (Temiakov *et al.*, 2004; Yin and Steitz, 2004).

7.2 Multi-subunit RNA polymerases

By the time this work was started, atomic-resolution structures of 10-subunit core RNA polymerase II (Pol II) from yeast at various resolutions and in the presence or absence of a “tailed template”-DNA/RNA complex were available (Table 2). These structures were the result of a series of attempts carried out over more than a decade to obtain structural information of this complex.

Table 2: Structure determination of yeast Pol II

Model	Resolution (Å)	PDB-ID	Reference
Electron crystallographic model from lipid layer crystals	≈30	—	(Edwards <i>et al.</i> , 1990)
Electron crystallographic model from 2D crystals	≈16	—	(Darst <i>et al.</i> , 1991)
Electron crystallographic model of open conformation	—	—	(Asturias and Kornberg, 1995)
Localisation of the C-terminal domain of Rpb1 by electron microscopy	16	—	(Meredith <i>et al.</i> , 1996)
Electron crystallographic model of closed conformation	—	—	(Asturias <i>et al.</i> , 1997)
Electron microscopic reconstitution (2D crystals and multi-particle averaging)	≈12-20	—	(Asturias <i>et al.</i> , 1998)
First x-ray crystallographic characterisation of Pol II (self-rotation function, similar folds of Rpb1 and 2)	≈4 (unphased)	—	(Fu <i>et al.</i> , 1998)
Electron crystallographic model from lipid layer crystals	16	—	(Asturias and Kornberg, 1999)
Electron microscopic image of a Pol II-Mediator complex	40	—	(Asturias <i>et al.</i> , 1999)

Model	Resolution (Å)	PDB-ID	Reference
Tantalum heavy-atom cluster-phased Pol II electron density map	≈6.0	—	(Fu <i>et al.</i> , 1999)
Electron microscopic reconstitution of an elongation complex	16-24	—	(Poglitsch <i>et al.</i> , 1999)
Backbone model of core Pol II (form 1)	3.1	1I3Q	(Cramer <i>et al.</i> , 2000)
Backbone model of core Pol II (form 2)	2.8	1I50	(Cramer <i>et al.</i> , 2001)
X-ray crystal structure of core Pol II - DNA/RNA minimal elongation complex	3.3	1I6H	(Gnatt <i>et al.</i> , 2001)
X-ray crystal structure of a Pol II- α -amanitin complex	2.8	1K83	(Bushnell <i>et al.</i> , 2002)
Electron microscopic reconstitution of complete Pol II	18	—	(Craighead <i>et al.</i> , 2002)
Backbone model of complete Pol II	4.2	1NT9	(Armache <i>et al.</i> , 2003)
Backbone model of complete Pol II	4.1	1NIK	(Bushnell and Kornberg, 2003)
Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage	3.8	1Y1V	(Kettenberger <i>et al.</i> , 2003), this study
Backbone model of the complete Pol II-TFIIB complex	4.5	1R5U	(Bushnell <i>et al.</i> , 2004)
Electron microscopy model of a complete Pol II-TFIIF complex	≈18	—	(Chung <i>et al.</i> , 2003)
Core Pol II-“Strand Separated Elongation Complex”	3.6	1SFO	(Westover <i>et al.</i> , 2004b)
Structural basis of substrate selection by Pol II	4.3 3.5 2.3	1R9S 1R9T 1TWF	(Westover <i>et al.</i> , 2004a)
Complete RNA Polymerase II Elongation Complex Structure and Its Interactions with NTP and TFIIS	4.0 4.0 4.5	1Y1W 1Y1Y 1Y77	(Kettenberger <i>et al.</i> , 2004), this study
Atomic structure of Rpb4/7 and refined model of complete Pol II	3.8	1WCM	(Armache <i>et al.</i> , 2004)

Two-dimensional crystals for electron crystallography had been available as early as in 1990 (Edwards *et al.*, 1990). Electron microscopic studies of Pol II were published in 1990, 91, 95, 97, 98 and 99 (Table 2). In one of these crystal forms, the conformation of the enzyme appears “open”, allowing entry of DNA, as required for the initiation of transcription (Asturias and Kornberg, 1995). By contrast, another crystal form contained the enzyme in a “closed” conformation, appropriate for retention of DNA during RNA chain elongation (Asturias *et al.*, 1997). Interaction with two polymerase subunits, the heterodimer Rpb4/Rpb7, were shown to trigger the open-closed transition. Biochemical evidence was available by that time which led to the conclusion that the open-closed transition is a crucial step in the transcription initiation process. In 1997, Gnatt *et al.* reported the purification and crystallisation of a “minimal elongation complex” (Gnatt *et al.*, 1997). This complex comprised the 10-subunit Pol II core and a “tailed template” DNA with the product RNA synthesised in the pre-assembled Pol II/DNA complex. However, structure solution of this complex had to await the year 2002 (see below). In 1998, the first electron microscopic multi-particle reconstruction of Pol II was described (Asturias *et al.*, 1998) to a nominal resolution of 12 Å. Between 12 and 20 Å, data quality was rather weak, indicating conformational flexibility of large parts of the enzyme. This flexibility was later ascribed to a movement of the “clamp” in the absence of nucleic acids (Cramer, 2002b). The first extension of the Pol II core machinery was reported in 1999 (Asturias *et al.*, 1999). In this publication, the multi-particle electron microscopic reconstitution of a Pol II “Holoenzyme”, i.e. the complex of Pol II with the Mediator, was reported. It showed the multi-subunit Mediator complex wrapped around Pol II at a resolution of about 40 Å. In 1999, two other electron crystallographic models were published. One model showed Pol II in complex with nucleic acids corresponding to a transcription bubble and product RNA (Poglitsch *et al.*, 1999). The other electron crystallographic model revealed Pol II in complex with TFIIE and TFIIIF (Asturias and Kornberg, 1999). This model revised the view of DNA entering the polymerase at one face and leaving it at the opposite face. In contrast, nucleic acids were shown to interact with only one face (i.e. the cleft, Figure 7) of Pol II.

In the same year, a breakthrough in structure determination using x-ray crystallography could be achieved: A 6 Å crystallographic map of core Pol II was obtained with phases derived from a Ta₆Br₁₂²⁺ heavy atom cluster (Fu *et al.*, 1999; Figure 6). For the first time, more robust information of the cleft and the mobile clamp could be observed. The crystallographic map was

reported to be in good accordance with the electron crystallographic map derived earlier (Asturias and Kornberg, 1999).

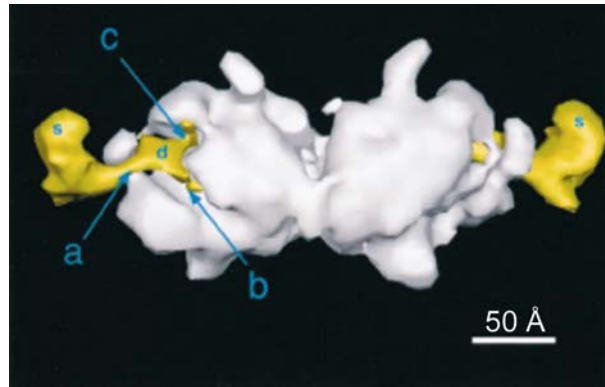


Figure 5: Three-dimensional positive differences between streptavidin-labelled, paused elongation complexes and RNA polymerase II dimer, superimposed on the 3D reconstruction of RNA polymerase II. Differences 2.5σ above the mean are shown in yellow, and the surface of RNA polymerase II is shown in white. Isolated, small difference densities less than about 15 \AA in largest dimension were omitted for clarity. A portion of this density, labelled d, appears in the cleft, directly connected by density a to that corresponding to streptavidin, labelled s. Two other protrusions of density, labelled b and c, emanate from the cleft. Adapted from (Poglitsch *et al.*, 1999).

The low-resolution structural data were excelled by a backbone model of core Pol II at 3.1 \AA resolution (Cramer *et al.*, 2000). For the first time, the arrangement and secondary structure elements of the 10 subunits in the complex could be determined, and functional domains like jaws, clamp, wall and the active site could be identified. Additionally, a suggested path for substrate NTP entry and for the extrusion of backtracked RNA (see chapter 10.3) was presented. Finally, atomic models of core Pol II and a Pol II-DNA/RNA complex could be obtained at 2.8 and 3.3 \AA resolution, respectively (Cramer *et al.*, 2001; Gnatt *et al.*, 2001). This breakthrough in structure determination was obtained by a special dehydration and shrinkage protocol, which improved diffraction substantially. The phase problem was solved by MIRAS, using a set of 12 datasets to obtain phase information from zinc, tantalum, iridium, mercury and rhenium heavy atom derivatives (Cramer *et al.*, 2000). These structures allowed the detailed analysis of the proteins involved, they explained a plethora of biochemical observations, and fuelled structure-guided analysis of the transcription machinery.

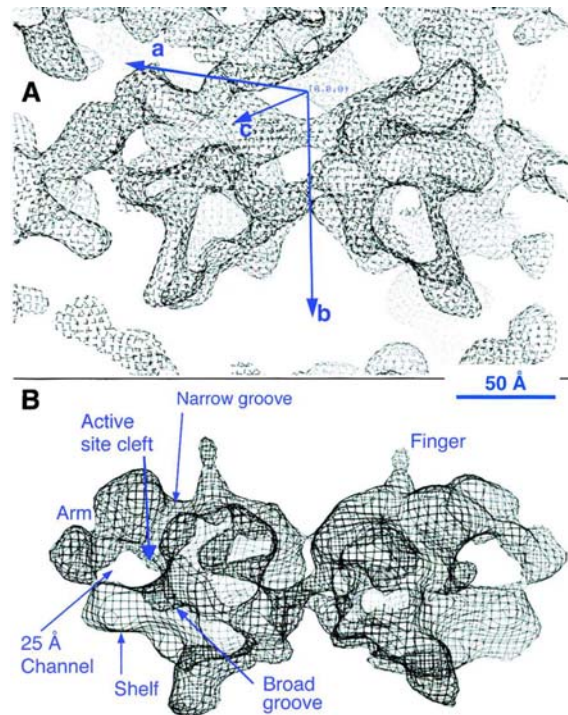


Figure 6: Molecular Envelopes of RNA Pol II. (A) Protein envelope obtained by MIRAS at 6 Å resolution (Fu *et al.*, 1999). The envelope is displayed as a single isocontour representing about 90% of the molecular volume. a, b, and c denote the unit cell axes of the 3D crystal. (B) The molecular envelope from electron crystallography (Asturias and Kornberg, 1999) with a contour level representing about 70% of the molecular volume. Adapted from (Fu *et al.*, 1999).

Table 3: Structures of bacterial RNA polymerase

Structure	Resol. (Å)	PDB-ID	Reference
Crystal structure of a <i>Thermus aquaticus</i> core RNA polymerase	3.3	1HQM	(Zhang <i>et al.</i> , 1999)
<i>Thermus aquaticus</i> core RNA polymerase-Rifampicin Complex	3.3	1I6V	(Campbell <i>et al.</i> , 2001)
Crystal structure of a <i>Thermus aquaticus</i> RNA polymerase σ -subunit fragment	2.9	1KU2	(Campbell <i>et al.</i> , 2002)
<i>Thermus aquaticus</i> RNA polymerase holoenzyme	4.0	1L9U	(Murakami <i>et al.</i> , 2002b)
<i>Thermus aquaticus</i> RNA polymerase holoenzyme/fork-junction promoter DNA complex	6.5	1L9Z	(Murakami <i>et al.</i> , 2002a)
Crystal structure of the RNA polymerase holoenzyme from <i>Thermus thermophilus</i>	2.6	1IW7	(Vassylyev <i>et al.</i> , 2002)
Regulation of RNA polymerase by the Alarmone ppGpp	2.7	1SMY	(Artsimovitch <i>et al.</i> , 2004)

8 STRUCTURE OF POL II AND COMPARISON WITH OTHER MULTI-SUBUNIT RNA POLYMERASES

The multi-subunit cellular RNA polymerases of eukaryotes, bacteria and archaea exhibit a high degree of sequence homology (Cramer *et al.*, 2001). A particular high degree of conservation is observed for a core around the active site (Minakhin *et al.*, 2001; Sweetser *et al.*, 1987). Comparing the structures of bacterial RNA polymerase from *Thermus aquaticus* (Zhang *et al.*, 1999) and yeast Pol II (Cramer *et al.*, 2000; Cramer *et al.*, 2001), it is evident that the five core subunits (Table 1) share a common overall architecture. The central mass and opposite sides of a central cleft is formed by two large subunits (Rpb1 and 2 in Pol II; β and β' in bacterial Pol; Figure 7). Two smaller subunits are involved in assembly and fastening the two large subunits of both RNA polymerases (Rpb3-Rpb11 in Pol II; α homodimer in bacterial Pol). Rpb1, but not the β subunit, has a C-terminal repeat domain (CTD) of 26-50 heptapeptide repeats (see chapter 9). Another core subunit (Rpb6 in Pol II; ω in bacterial Pol) further buttresses and stabilises the large subunits (Minakhin *et al.*, 2001; Mukherjee and Chatterji, 1997; Nouraini *et al.*, 1996; Sweetser *et al.*, 1987).

A striking feature which was already evident in the early electron density maps (Fu *et al.*, 1999) is a mobile “clamp”. In two crystal forms of free Pol II, the clamp was trapped in two different open states. In the first Pol II elongation complex, the clamp was found to be in a closed state, apparently induced by nucleic acid binding (Gnatt *et al.*, 2001). The transition from the open to the closed state goes along with a 30° rotation, paralleled by ordering of five mobile “switch” regions upon nucleic acid binding (Cramer *et al.*, 2001; Gnatt *et al.*, 2001; Figure 7). The high stability of elongation complexes, which is required for efficient transcription of very long genes, can be rationalised by this clamp closure. The switches apparently couple clamp closure to the presence of the DNA-RNA hybrid. The dissociable subcomplex Rpb4/7 (chapter 8.1) further restricts the clamp to its closed conformation (Armache *et al.*, 2003; Bushnell and Kornberg, 2003). The active site is located at the floor of the cleft above a funnel, which narrows down to a pore (Pol II) or secondary channel (bacterial Pol), respectively. The active site is surrounded by a wall (Pol II)/flap (bacterial Pol) and a bridge helix/ β' F-helix. In the EM model and in the structure of the minimal elongation complex, nucleic acids were observed in a multi-subunit RNA polymerase. The DNA-RNA hybrid is situated above the active site and the pore in the cleft (Gnatt *et al.*, 2001; Poglitsch *et al.*, 1999). Substrate NTPs are likely to enter the active site

via the funnel and the pore, although very recent kinetic data suggest the existence of an alternative route: Incoming NTPs are hypothesised to base-pair with the +2 base in the template strand, entering through the cleft. In the translocation step (see chapters 4 and 10.3), this NTP is shuttled to the catalytic site. Thus, two NTP molecules are bound simultaneously at the +1 and +2 position and “line up” for incorporation in this model (Gong *et al.*, 2004; Nedialkov *et al.*, 2003; Zhang and Burton, 2004). To date, there is no published structure of a bacterial RNA polymerase-DNA/RNA complex.

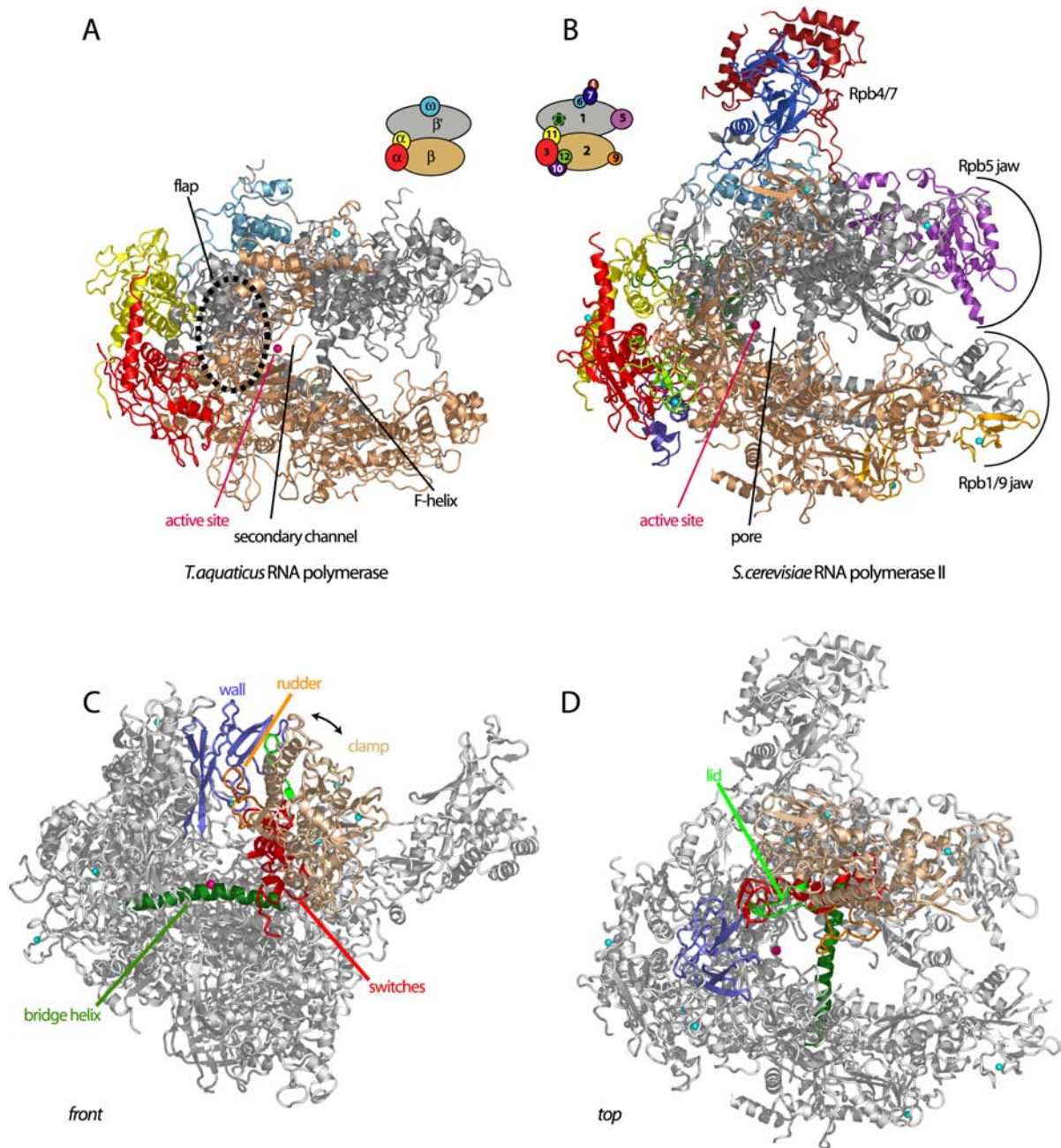


Figure 7: Overview of multi-subunit RNA polymerases. (A) RNA polymerase from *T. aquaticus* (Zhang *et al.*, 1999) and (B) complete Pol II from *S. cerevisiae* (Armache *et al.*, 2004) in similar orientations (related to the “top” view in Cramer *et al.*, 2000). Subunits are coloured according to the scheme above the structures. Homologous subunits are coloured identically. In panels (C) and (D), two different orientations of *S. cerevisiae* Pol II shown in (B) are coloured to highlight functional elements discussed in the text. This figure and all other structure figures were prepared with PyMol (www.pymol.org) unless otherwise indicated.

8.1 Core Pol II and the subcomplex Rpb4/Rpb7

Pol II consists of the ten-subunit catalytic core and the dissociable heterodimer Rpb4/Rpb7 (Rpb4/7). Assembly of Pol II with general transcription factors to the pre-initiation complex (PIC, see chapter 10) requires the presence of Rpb4/7 (Edwards *et al.*, 1991; Ruet *et al.*, 1980), whereas it is dispensable for transcription elongation. In yeast, a knockout of Rpb4 leads to facilitated dissociation of the (essential) Rpb7. Therefore Pol II isolated from a Δ Rpb4 strain lacks both Rpb4 and Rpb7. In the course of this project, the structure of the complete Pol II including Rpb4/7 was derived independently by two groups (Armache *et al.*, 2003; Bushnell and Kornberg, 2003). Bushnell *et al.* isolated endogenous complete Pol II by means of an affinity tag on Rpb4 from yeast cells in the stationary phase. Armache *et al.* established an alternative approach: Endogenous purified core Pol II was supplemented with an excess of recombinant Rpb4/7 and purified by gel filtration. The advantage of the latter strategy lies in the fact that mutations and heavy atom derivatisation can be performed more easily in the recombinant system.

The subcomplex Rpb4/7 was first localised by electron microscopy (Craighead *et al.*, 2002). The Rpb4/7-Pol II interaction via Rpb4 proposed by these authors was disproved in the x-ray models (Armache *et al.*, 2003; Bushnell and Kornberg, 2003). The early x-ray data at ≈ 4.2 Å resolution (Armache *et al.*, 2003; Bushnell *et al.*, 2002) allowed to reconstruct the 12-subunit complex from atomic models of the 10-subunit Pol II core and the distantly related archaeal orthologues of Rpb4/7, subunits RpoE and RpoF from *Methanococcus jannaschii* (Todone *et al.*, 2001). A backbone model for most parts of Rpb4/7 could be derived this way, since archaeal RpoE and RpoF exhibit comparable overall structures (Armache *et al.*, 2003). Eventually, a refined atomic model of complete 12-subunit Pol II could be derived at 3.8 Å by an improved alternative strategy (Armache *et al.*, 2004): First, an atomic model of Rpb4/7 alone was derived and fitted, together with an atomic model of core Pol II, into a 3.8 Å electron density map. Full-length Rpb4/7 failed to crystallise, but crystals were obtained after 34 residues were removed from the proteolytically sensitive N-terminal tail of Rpb4. Additional deletion of a proteolytically sensitive loop in Rpb4 resulted in larger crystals, suitable for structure determination at 2.3 Å resolution (Figure 8). Starting from two refined structures, refinement was possible even at 3.8 Å resolution and converged at a free R-factor of 28.5 %. This structure of complete Pol II was used for phasing diffraction data of Pol II complexes presented in this work (see Results and Discussion).

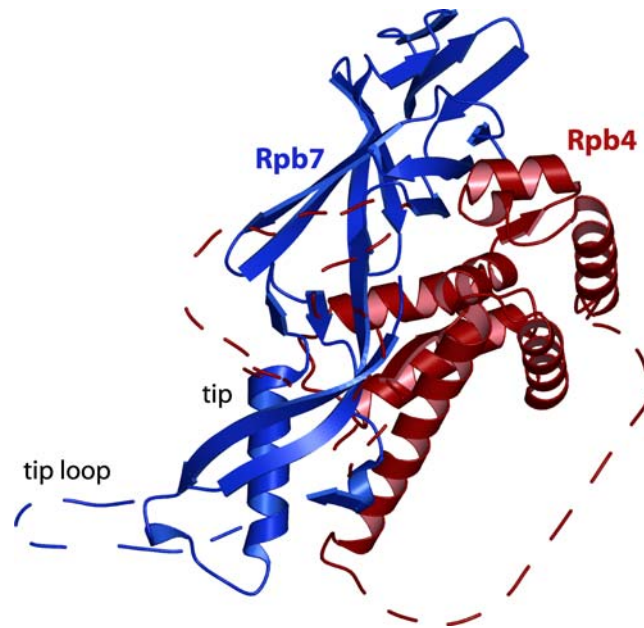


Figure 8: High-resolution structure of the Rpb4/Rpb7 heterodimer (Armache *et al.*, 2004). Disordered in the structure are the Rpb7 tip loop (residues 57-68), Rpb4 residues 35-46, 77-81, and 101-118. Rpb4 residues 1-34 and 82-100 were not present in the crystallised variant. This high-resolution structure was docked in a 3.8 Å crystallographic map of complete Pol II to obtain an atomic structure of the whole complex.

Rpb4/7 protrudes from the polymerase “upstream face” on which initiation factors are thought to assemble for promoter DNA loading and preinitiation complex formation (Armache *et al.*, 2003; Bushnell *et al.*, 2002). Since Rpb7 contains a ribonucleoprotein (RNP) fold, and an oligonucleotide-binding (OB) fold, a possible involvement of Rpb7 in transcript binding and its redirection towards the CTD has been suggested (Armache *et al.*, 2003; Bushnell *et al.*, 2002). A very small interaction surface is consistent with the dissociable nature of Rpb4/7. The presence of Rpb4/7 has pronounced effects on the structure of core Pol II: Rpb7 forms a wedge between the clamp and the linker which restricts the clamp to a closed position. Binding of Rpb4/7 induces a conformation in two switch regions, which is indistinguishable from the conformation with bound template strand. A closed clamp prevents the entry of double-stranded promoter DNA into the active centre cleft, and switch regions poised for template strand binding might facilitate transcription initiation. Rpb4/7 also plays a role in Pol II recycling (see chapter 10.5). CTD dephosphorylation is carried out by the phosphatase Fcp1 (Lin *et al.*, 2002), which is apparently recruited to the phosphorylated CTD by Rpb4/7, since Rpb4 binds Fcp1 (Chambers *et al.*, 1995).

9 THE C-TERMINAL DOMAIN OF POL II

The C-terminal domain (CTD) of Pol II is comprised of a number of repeats of the heptapeptide YSPTSPS which seems to increase with organismic complexity from 26 in yeast, 32 in *C. elegans*, 45 in *Drosophila* to 52 in mammals (Hampsey, 1998). Unique to Pol II, the CTD is essential for cell viability and plays a central role in the regulation of mRNA synthesis *in vivo*.

An unstructured, extended yeast CTD would span a distance of about 650 Å (Cramer *et al.*, 2001) and serves as a platform for the assembly and recruitment of various factors involved in initiation, elongation, mRNA maturation, surveillance, and export (Hirose and Manley, 2000; Orphanides and Reinberg, 2002; Proudfoot *et al.*, 2002). The CTD is extensively phosphorylated and dephosphorylated during the transcription cycle. Phosphorylation in the preinitiation complex (see chapter 20) is performed by a kinase-module of TFIIF. Generally, CTD phosphorylation at serine-5 coincides with transcription initiation and early elongation, whereas serine-2 phosphorylation is associated with transcribing Pol II further away from the promoter (Sims *et al.*, 2004).

Dephosphorylation of the CTD, a prerequisite for Pol II recycling, is accomplished by the phosphatases Fcp1 (Ser2) and Ssu72 (Ser5) (Kamenski *et al.*, 2004; Meinhart *et al.*, 2003b; Yeo *et al.*, 2003). Fcp1 was shown to interact directly with TFIIF (Chambers *et al.*, 1995; Kamada *et al.*, 2003; Nguyen *et al.*, 2003) and with Pol II via the CTD (Krishnamurthy *et al.*, 2004). A family of Ser2-specific small CTD phosphatases (SCPs) were found in mammals (Yeo *et al.*, 2003).

Structure determination of the free CTD using both NMR spectroscopy (Cagas and Corden, 1995) and x-ray crystallography failed, indicating that the CTD is devoid of a well-defined secondary structure. In all crystal structures of Pol II reported so far, the CTD and parts of the linker connecting it to Pol II are disordered. Nevertheless, crystal structures of CTD peptides bound to the capping enzyme Cgt1, to the peptidyl-proline isomerase Pin1, and to the CTD-interacting domain of the RNA-processing factor Pcf11, have been solved (Fabrega *et al.*, 2003; Meinhart and Cramer, 2004; Verdecia *et al.*, 2000). These structures indicate a highly versatile nature of CTD folding, depending on its phosphorylation status and binding partner. The CTD can be thought of as a nucleation centre for the multitude of factors that regulate mRNA

processing events concomitant with elongation. In addition, the CTD could mediate factor recruitment in general, using a code comparable to the histone code (Sims *et al.*, 2004).

10 THE EUKARYOTIC TRANSCRIPTION CYCLE

The generation of mature mRNA by RNA polymerase involves multiple processes, some of which occur sequentially and others in parallel. The primary phases of transcript generation form a so-called transcription cycle and include preinitiation, initiation, elongation, termination and reinitiation (Figure 9; Hahn, 2004).

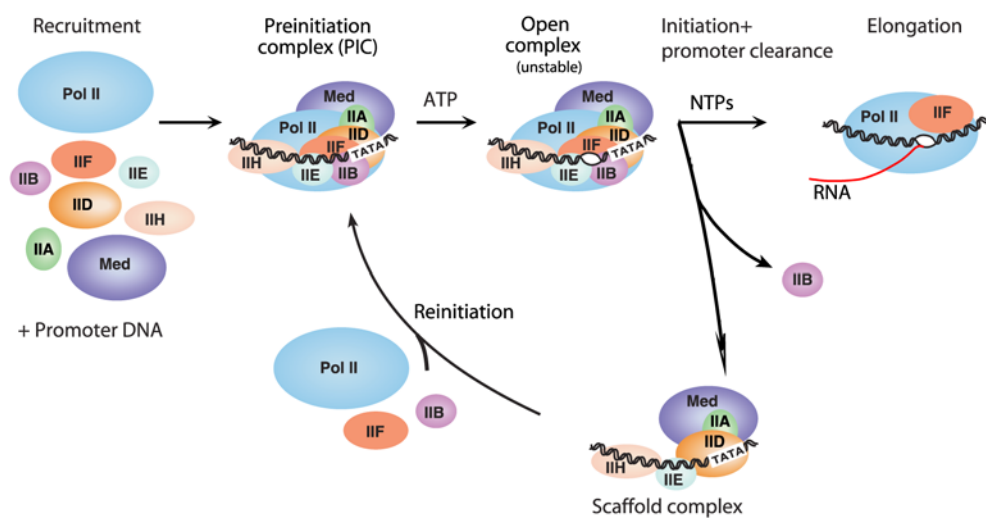


Figure 9: The pathway of transcription initiation and re-initiation by Pol II. See Table 4 for a description of the general transcription factors. Adapted from (Hahn, 2004).

10.1 Preinitiation and initiation

The transcription cycle starts with the assembly of a preinitiation complex (PIC) at the promoter. The eukaryotic PIC includes the general transcription factors TFIID, TFIIB, TFIIE, TFIIIF, and TFIIH as well as RNA polymerase II and several additional cofactors (Orphanides *et al.*, 1996; Orphanides and Reinberg, 2002; Table 4). In bacterial RNA polymerase, the σ factor performs this function of promoter recognition and may hold the flap in a defined position (Arthur *et al.*, 2000; Arthur and Burgess, 1998) during promoter loading. In Pol II, Rpb4/7 restricts the clamp to a closed position, potentially preventing duplex DNA from entering the cleft (Armache *et al.*, 2003; Bushnell and Kornberg, 2003). Thus, duplex DNA might be suspended over the cleft before melting (Armache *et al.*, 2003). Melting of DNA at the initiation site and formation of a

transcription bubble is a prerequisite for initiation. This process is ATP dependent and requires the action of TFIIE and TFIIH (Goodrich and Tjian, 1994; Holstege *et al.*, 1996; Kim *et al.*, 2000) in eukaryotes. Initiation factors are believed to dock to other sites on the Pol surface, too, in order to align the transcription start site with the active centre (Cramer, 2002b).

Table 4: *S. cerevisiae* general transcription factors and coactivators

Factor	Subunits	Function (Hahn, 2004)	
TFIIA	2	stabilises TBP and TFIID-DNA binding; positive and negative gene regulation	
TFIIB	1	binds TBP, Pol II and promoter DNA; helps fix transcription start site	
TFIID	TBP	1	binds TATA element and bends promoter DNA; platform for assembly of TFIIB, TFIIA and TAFs
	TAFs	14	bind INR and downstream promoter elements; regulatory target
Mediator	24	target of regulatory factors; stimulates basal and activated transcription	
TFIIF	3	binds Pol II and is involved in Pol II recruitment to the PIC	
TFIIE	2	binds promoter DNA near transcription start; may help open or stabilise the transcription bubble	
TFIIH	10	links transcription and DNA repair; kinase and helicase activities; essential for open complex formation	

10.2 Promoter clearance and early elongation

Once polymerase has assembled with template DNA to form a so-called “open complex”, transcription initiation occurs upon addition of the two initiating nucleoside triphosphates dictated by the DNA sequence leading to the formation of a first phosphodiester bond. The earliest stages of transcription are marked by instability of the transcription complex and a pronounced tendency to release short pieces of RNA (2-5mers) in a phase, which is termed “abortive initiation”. Stabilisation of early elongation complexes is observed when the DNA-RNA hybrid reaches 8 bases in length (Kireeva *et al.*, 2000a; Kireeva *et al.*, 2000b). Synthesis of about 23 bases renders the elongation complex maximally stable and allows RNA polymerase II to clear the promoter (Pal and Luse, 2003). The transition to elongation goes along with partial PIC disassembly: A subset of general transcription factors and the Mediator remain at the

promoter, serving as a scaffold for the formation of the next transcription initiation complex (Yudkovsky *et al.*, 2000; Zawel *et al.*, 1995). Therefore, re-initiation of transcription is a much faster process compared to the initial round and is responsible for the bulk of transcription in the cell (Jiang and Gralla, 1993; Orphanides and Reinberg, 2000; Ranish *et al.*, 1999). TFIIF plays an integral role in facilitating promoter clearance (Dvir *et al.*, 1996; Goodrich and Tjian, 1994; Kumar *et al.*, 1998). Of all the general transcription factors, only TFIIF can be found bound to Pol II during the elongation phase (Sims *et al.*, 2004). Most recent structural data of a Pol II-TFIIB complex indicate an impediment of the exit-path of newly formed RNA by the fingers domain of TFIIB. Only dissociation of this TFIIB domain would allow the synthesis of longer transcripts (Bushnell *et al.*, 2004).

10.3 Elongation and the nucleotide addition cycle

A manifold of biochemical data and analogies to T7 RNA polymerase and DNA polymerases led to a working model with a catalytic step proceeding via a “unified two metal ion mechanism” (Steitz, 1993). According to this mechanism, one divalent metal ion (metal B) enters the active site together with an NTP. A permanently bound divalent metal A coordinates the α -phosphate group of an NTP and activates it by implying a more positive charge on the P atom. The 3'-OH group of the RNA attacks this P atom in an S_N2 -type reaction, generating an elongated RNA plus a pyrophosphate ion, which leaves the active site, presumably together with metal B.

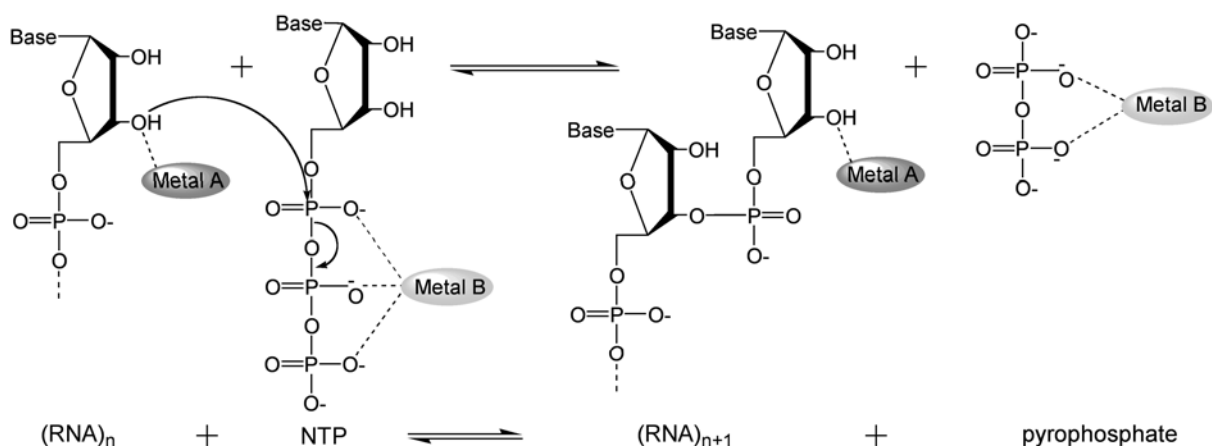


Figure 10: Reaction mechanism of RNA chain elongation

There is an ongoing debate to clarify the mechanisms of translocation by single and multi-subunit RNA polymerases. The “Brownian ratchet mechanism” assumes that RNA polymerase distributes itself by diffusional sliding between all accessible positions on the template with

relative occupancy determined by relative free energy. Subsequent NTP binding locks polymerase in the post-translocated state (Guajardo and Sousa, 1997). Various “powerstroke” models, in which a protein conformational change drives forward translocation have also been suggested (Steitz, 2004; Wang and Oster, 2002). In the case of T7 RNA polymerase, the powerstroke can be rationalised as rotation of the O-helix, driven by release of pyrophosphate after the catalytic step, since coupling between translocation and pyrophosphate release was observed in the recent structures of T7 RNA polymerase (Yin and Steitz, 2004). In yeast RNA polymerase II, the powerstroke may be accompanied by bending or local unwinding of the bridge helix toward the DNA/RNA hybrid, as suggested by several groups (Cramer *et al.*, 2001; Gnatt *et al.*, 2001; Holmes and Erie, 2003; Vassylyev *et al.*, 2002). The powerstroke model is further supported by recent kinetic data (Zhang and Burton, 2004). However, no alternative bridge-helix conformations have been observed in crystal structures of bacterial and yeast Pol so far, which would support this model. Landick put forward the idea of the “Brownian ratchet mechanism”, in which RNA polymerases oscillate between a pre-translocated and a post-translocated state due to Brownian motion (Landick, 2004). Further evidence for a ratchet mechanism of translocation comes from investigations of *E. coli* polymerase mutants (Bar-Nahum *et al.*, 2005). The authors suggest two independent “ratchet domains”, namely the incoming NTP and the flexible F-/bridge helix, which act in concert to drive forward translocation. Whether a powerstroke model or a ratchet mechanism, or a combination of both, best describe translocation of RNA polymerase along its template seems to remain a matter of discussion.

10.4 Termination and mRNA processing in eukaryotes

The choreography of termination and mRNA processing is intimately linked in eukaryotes. Before a gene transcript leaves the nucleus as mature mRNA, it has to undergo three major processing events: During capping, a GMP molecule is fused to the 5'-terminus of nascent RNA via an unusual 5'-5' triphosphate linkage and methylated at position N7. This cap protects the nascent RNA from exonucleolytic 5'→3' degradation and is recognised during nuclear export of mature mRNA. The splicing process removes non-coding sequences (introns), which are present in most mammalian genes and some yeast genes. The 3'-terminus of mRNAs is comprised of a poly(A) tail of about 200 nucleotides. This poly(A) tail is attached to pre-mRNA after cleavage at certain sequences by a nuclease in the polyadenylation machinery or by an intrinsic ribozyme

(Teixeira *et al.*, 2004). All three processes affect each other's efficiency as well as the efficiency of initiation and termination (Proudfoot, 2004; Proudfoot *et al.*, 2002).

Compared to the extensive information on transcriptional termination in bacteria, control of eukaryotic termination is less well understood. Purified Pol III terminates *in vitro* after the synthesis of a series of uracils automatically, without the need for hairpin-structures or auxiliary termination factors. Pol I requires a Pol I-specific termination factor which binds to a sequence at the end of the rRNA-coding DNA (reviewed in Reeder and Lang, 1994). During termination, both the transcript is released from the site of transcription and Pol II is detached from the DNA template. While the pre-mRNA is protected from rapid degradation by the 5'-cap and the 3'-poly(A) tail, the downstream cleavage product is very short-lived. Thus, a so-called "torpedo" model suggested that rapid degradation of the downstream product with the RNase catching up with still elongating Pol II is responsible for termination at a site 900-1600 bases downstream of the cleavage site (Connelly and Manley, 1988; Kim *et al.*, 2004; Proudfoot, 1989; Teixeira *et al.*, 2004; West *et al.*, 2004). An alternative "antiterminator" or "allosteric" model (Logan *et al.*, 1987) assumes a conformational change in the transcribing Pol II upon passing through the poly(A) signal that makes it termination competent. Further research will have to prove the one or the other model or reconcile parts of both models (reviewed in Luo and Bentley, 2004).

10.5 Pol II recycling

Pol II is recycled through dephosphorylation of serine 2 by the phosphatase activity of Fcp1 (Cho *et al.*, 1999) and through dephosphorylation of serine 5 by the Ser5-specific phosphatase Ssu72 (Ganem *et al.*, 2003; Krishnamurthy *et al.*, 2004; Meinhart *et al.*, 2003b; see chapter 9). A polymerase "reset" this way to its initiation-competent state can engage in another round of transcription, maybe as part of a "transcriptional hotspot" (Cook, 1999; Iborra *et al.*, 1996). Recent observations proclaim close spatial proximity of the end and the start of a gene, coupling termination and re-initiation (Proudfoot, 2004).

11 ELONGATION FACTORS THAT REGULATE POL II ACTIVITY

Purified Pol II is capable of elongating RNA chains *in vitro* at rates of about 100 to 300 nucleotides per minute (Izban and Luse, 1992a). In contrast, mRNA is synthesised in cells, on chromatin templates, at rates approaching 2000 nucleotides per minute (Shermoe and Ofarrell,

1991; Thummel *et al.*, 1990; Ucker and Yamamoto, 1984) These observations prompted the search for activities that would enhance the elongation rate of the purified enzyme on both chromatin templates and naked DNA. On the other hand, factors were identified that slow down the turnover number of elongating Pol II. Table 5 gives an overview of transcription factors which modulate the elongation rate of Pol II. Many of these elongation factors act by reducing the likelihood for transcriptional pausing (see chapter 12; Takagi *et al.*, 1995; Sims *et al.*, 2004).

Table 5: Elongation factors which modulate the elongation rate of Pol II
(Sims *et al.*, 2004).

Elongation factor	Homologues in yeast/mammals	Properties
TFIIS	yes/yes	stimulates transcript cleavage to alleviate arrest (see chapter 12)
TFIIF	yes/yes	Essential for pre-initiation complex formation; stimulates Pol II by reduced pausing
Elongins	no/yes	stimulates Pol II by reduced pausing
ELL	no/yes	stimulates Pol II by reduced pausing
DSIF	yes/yes	stimulates elongation, suppresses early transcript termination, stimulates capping
NELF	yes/yes	halts Pol II to allow capping checkpoint control
P-TEFb	yes/yes	relieves NELF-mediated pausing, phosphorylates the CTD of Pol II and Spt5
CSB1	yes/yes	stimulates elongation, modulates TFIIS, rescues Pol II at DNA lesions, involved in transcription-coupled DNA repair

TFIIF plays an integral role in recruiting Pol II during PIC formation (Conaway *et al.*, 1991; Flores *et al.*, 1991; Orphanides *et al.*, 1996). Aside of this function, TFIIF diminishes the time Pol II is paused and thereby stimulates the overall elongation rate (Bengal *et al.*, 1991; Flores *et al.*, 1989; Izban and Luse, 1992a). TFIIF is one of the strongest interacting proteins with Pol II. After PIC formation, TFIIF is released from the transcription complex but re-associates with stalled elongation complexes (Zawel *et al.*, 1995).

The Elongin complex was initially isolated as a stimulatory factor (Bradsher *et al.*, 1993a; Bradsher *et al.*, 1993b) which exerts its function in a manner similar to TFIIF. Unlike TFIIF however, Elongin is not required for PIC formation or initiation (Bradsher *et al.*, 1993b). Since Elongin is not found in yeast, it may be a higher eukaryotic gene-specific factor.

ELL (eleven-nineteen lysine rich leukemia) is the third member in the group of transcription elongation factors, that act by preventing pausing of elongating Pol II (Shilatifard *et al.*, 1997; Shilatifard *et al.*, 1996). ELL was initially identified by its involvement in acute myeloid leukemia (Thirman *et al.*, 1994). *Drosophila* ELL (dELL) is an essential factor during development and dELL mutants show defects in the transcription of larger genes (Eissenberg *et al.*, 2002; Shilatifard, 2004)

DSIF (DRB sensitivity inducing factor; DRB = 5,6-dichloro-1-D-ribofuranosylbenzimidazole, a nucleotide analogue) was reported to antagonise the negative effects of transcriptional pausing caused by the chromatin remodelling factor Isw1 (Morillon *et al.*, 2003). DSIF genetically and physically interacts with a set of factors such as TFIIF, TFIIS, FACT and others (Sims *et al.*, 2004) and could therefore be an adaptor that connects other modulators to elongating Pol II.

The NELF (negative elongation factor) complex promotes Pol II pausing (Yamaguchi *et al.*, 1999) in the presence of DSIF. DSIF binds directly to Pol II and NELF preferentially binds to the assembled DSIF/Pol II complex, but not to DSIF or Pol II alone (Yamaguchi *et al.*, 2002). Elongation is resumed when NELF dissociates from the paused elongation complex when the CTD is being phosphorylated by P-TEFb (Ivanov *et al.*, 2000; Kim and Sharp, 2001).

P-TEFb (positive transcription elongation factor b) is capable to stimulate DRB-sensitive transcription of long transcripts *in vitro*. (Marshall and Price, 1992). P-TEFb has kinase activity which targets serine 2 of the CTD repeats of Pol II at actively transcribed genes (Ni *et al.*, 2004). The elongation activity of P-TEFb is dependent on this kinase activity (Sims *et al.*, 2004).

The elongation factor CSB (cockayne sndrome complementation group B) is a DNA-dependent ATPase which directly binds to Pol II and affects the activity of TFIIS (Selby and Sancar, 1997; Tantin *et al.*, 1997). CSB links transcription by Pol II to transcription-coupled nucleotide repair

(Gregory and Sweder, 2001; Licht *et al.*, 2003) and to rescue of stalled Pol II at DNA lesion sites (Woudstra *et al.*, 2002).

12 DISCOVERY AND FUNCTION OF TFIIS

TFIIS (also referred to as SII) was the first Pol II-associated factor to be purified (Sekimizu *et al.*, 1976), based on its ability to promote efficient synthesis of long transcripts *in vitro*. Biochemical studies subsequently revealed that this function relies on the ability of preventing Pol II from arrest. Elongating RNA polymerases (bacterial, archaeal and eukaryotic Pol I, II and III) appear to oscillate between inactive and active conformations at each step of nucleotide addition (Erie *et al.*, 1993; Gu and Reines, 1995; Rhodes and Chamberlin, 1974). Consequently, the overall RNA polymerase elongation rate is not determined solely by the rate of phosphodiester bond formation, but also by the fraction of time polymerase dwells in inactive conformations. Transcriptional pausing occurs when Pol II halts the addition of NTPs to the nascent RNA transcript for a time, before resuming productive elongation on its own. The paused state slowly isomerises to an arrested state that can be reversed only by the action of the elongation factor TFIIS or its bacterial counterparts GreA and GreB (Gu and Reines, 1995; Wind and Reines, 2000). Reasons for transcriptional pausing and arrest include intrinsic DNA sequences, which code for unstable A-U rich DNA-RNA hybrids, and proteins or drugs bound to DNA, which block the forward movement of transcribing polymerases (Erie, 2002; Fish and Kane, 2002; Nudler *et al.*, 1997; Shilatifard *et al.*, 2003).

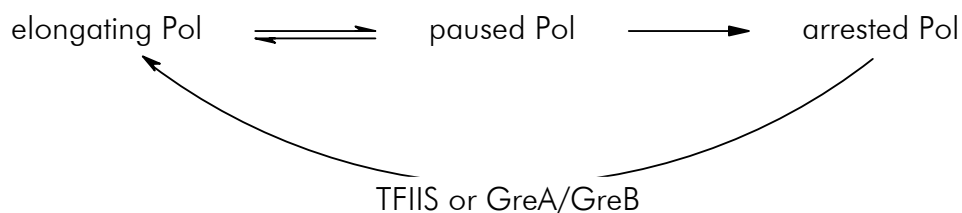


Figure 11: During transcription, RNA polymerases undergo frequent pausing, which can lead to permanent arrest. The action of TFIIS or GreA/GreB is required to reactivate such stalled RNA polymerases

Transcriptional pausing and arrest are coupled to a reverse movement (backtracking) of the polymerase relative to the DNA template, thereby disengaging the active site and the 3'-OH

group of the nascent RNA. Backtracking by a few nucleotides results in pausing, a temporary block to elongation, from which Pol II can escape by itself. More extensive backtracking however can lead to transcriptional arrest. The backtracked 3'-terminus is most likely extruded through the funnel and pore (Cramer *et al.*, 2000). Thus, arrest may result from irreversible trapping of the 3'-end of the displaced transcript at a binding site in the pore (Cramer *et al.*, 2000; Cramer *et al.*, 2001; Gnatt *et al.*, 2001; Westover *et al.*, 2004a). TFIIS or GreA/B then promote cleavage of short pieces of the 3' terminus of the backtracked RNA (preferentially dinucleotides and 7-9mers, respectively) (Gu and Reines, 1995), which is a necessary and sufficient prerequisite for the reactivation of arrested polymerases (Fish and Kane, 2002; Shilatifard *et al.*, 2003). TFIIS strongly enhances a weak RNA nuclease activity that is intrinsic to Pol II (Reines *et al.*, 1992; Wang and Hawley, 1993). The nuclease activity represents the second enzymatic function of the enzyme beside RNA polymerisation (and its reverse reaction, pyrophosphorolysis). (Surratt *et al.*, 1991). This demonstrates that RNA polymerase-mediated transcript cleavage is an evolutionarily conserved process. In addition to stimulating transcription by overcoming arrest, TFIIS was shown to reduce the likelihood of transient pausing of Pol II (Bengal *et al.*, 1991; Rappaport *et al.*, 1987; Sluder *et al.*, 1989). The latter stimulatory function is different from the function of elongation factors such as TFIIF, Elongin and ELL, which increase the rate of nucleotide incorporation (Zhang *et al.*, 2003).

Studies using an artificial DNA-RNA hybrid demonstrated that Pol II is able to incorporate a mismatched nucleotide, and that TFIIS enhances the preferential removal of the misincorporated base (Jeon and Agarwal, 1996). This study and additional kinetic studies (Thomas *et al.*, 1998) demonstrated the proofreading activity of Pol II, which is dramatically enhanced by TFIIS *in vitro*. Under normal circumstances, yeast TFIIS null mutants do not show an apparent phenotype, thus the relevance of mRNA proofreading is unclear. Under oxidising conditions, however, transcriptional fidelity was reported to be ninefold higher in the wild type compared to a TFIIS knockout strain (Koyama *et al.*, 2003).

Chromatin immunoprecipitation studies have shown TFIIS to localise with the coding regions of genes (Pokholok *et al.*, 2002). A knockout of *PPR2*, the gene encoding for TFIIS in yeast, is not lethal but leads to a sensitivity to 6-azauracil (Hubert *et al.*, 1983; Nakanishi *et al.*, 1985). The substrate analogue 6-azauracil interferes with the pyrimidine biosynthesis pathway and lowers the

cellular concentrations of CTP and UTP, thus altering the relative abundance of the four NTPs. A sensitivity to 6-azauracil is therefore attributed to defects in transcription elongation.

TFIIS is a modular factor that comprises an N-terminal domain I, a central domain II, and a C-terminal domain III, which are connected by flexible inter-domain linkers. The weakly conserved domain I forms a four-helix bundle (Booth *et al.*, 2000) and is not required for TFIIS activity in the context of elongation (see below). Domain II forms a three-helix bundle (Awrey *et al.*, 1998; Morin *et al.*, 1996), and domain III adopts a zinc ribbon fold with a thin protruding β -hairpin (Olmsted *et al.*, 1998; Qian *et al.*, 1993b). Domain II and the linker between domains II and III are required for Pol II binding, whereas domain III is essential for stimulation of RNA cleavage. The linker between domains II and III confers species selectivity and was thought to position and orient domains II and III (Shimasaki and Kane, 2000). Solution structures of all three domains were solved individually (Booth *et al.*, 2000; Morin *et al.*, 1996; Qian *et al.*, 1993b) and proved to be useful for driving structure determination in this work. In addition to the major, ubiquitously expressed form of TFIIS in higher eukaryotes, there are multiple tissue- or developmental stage-specific members of the TFIIS family (Kanai *et al.*, 1991; Xu *et al.*, 1994). In Pol III, the zinc finger domain of Rpc11 is the functional homologue of TFIIS domain III (Huang *et al.*, 2005), and is essential for 3'-cleavage, tRNA maturation and termination. The Pol I subunit A12.2 comprises a zinc ribbon domain which is homologous to the zinc ribbon of TFIIS (Figure 34).

TFIIS genetically interacts with a variety of components of the transcription machinery. Several genes were identified that are necessary for either cell survival or robust growth when the gene encoding TFIIS has been disrupted. These include amongst others genes for the Swi/Snf chromatin remodelling complex. SWI1 and SNF2 disruptions were also synthetically lethal with a deletion of the TFIIS-encoding gene PPR2, suggesting that the reduced ability to remodel chromatin confers the synthetic phenotype. Detection of these genetic interactions provides another functional link between the Swi-Snf complex and the elongation machinery (Davie and Kane, 2000). The RNA cleavage activity of TFIIS is required for the ultra-rapid initiation of heat shock genes such as Hsp70 (Adelman *et al.*, 2005), where elongating Pol II is stalled early beyond the promoter, poised to be released for transcription by TFIIS upon heat-shock.

13 TRANSCRIPTION *IN VITRO*

Elongation complexes can be readily assembled *in vitro*. These complexes are stable for at least a week under various conditions and can be resolved by gel electrophoresis and chromatography. (Daube and von Hippel, 1994; Gnatt *et al.*, 1997). Functional elongation complexes to study RNA polymerases can be assembled on natural promoters in the presence of general transcription factors. Another technique suitable for obtaining large quantities relies on RNA polymerases binding to the 3' end of short RNAs annealed to a single-strand template DNA (DNA-RNA hybrid, see chapter 16.1 in Results and Discussion). (Daube and von Hippel, 1992). A complementary template strand can be subsequently incorporated into elongation complexes assembled this way. Alternatively, Pol II was shown to start *de novo* RNA synthesis on double-stranded DNA bearing a poly(dC) tail of > 15 bases at one 3' end ("tailed template") (Gnatt *et al.*, 1997). Elongation complexes can be separated from free RNA polymerase by chromatography on heparin resin, which specifically retains the free enzyme (Komissarova *et al.*, 2003).

14 STRUCTURE DETERMINATION OF MULTI-SUBUNIT COMPLEXES

Proteins most often interact with partner proteins or other types of (macro)molecules in their cellular context (Patrick Aloy, EBI). Current estimates assume an average of about five interacting partners per protein in the proteome of the yeast *S. cerevisiae*. (Grigoriev, 2003). Protein interactions and their associated networks are the key to understanding cellular processes in the post-genome era. Meanwhile, about 60 % of nearly 30000 entries in the protein data base (www.rcsb.org/pdb) consist of more than one biopolymer chain, but most of these structures have been deposited over the last few years (Figure 12).

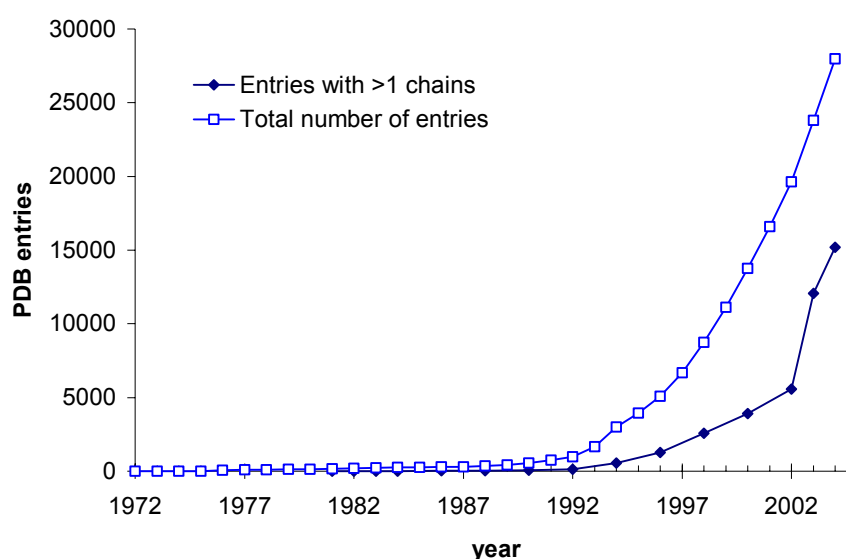


Figure 12: Entries in the protein data base (www.rcsb.org) by December 31, 2004. The graph shows the total number of deposited x-ray-derived structures (□) and structures consisting of more than one biopolymer chain (protein/DNA/RNA/carbohydrates) (◆).

Among all PDB entries, there are 1220 protein-nucleic acid complexes (i.e. 4 %) and only eleven (i.e. 0.04 %) ternary complexes of protein, DNA and RNA. Prevalent experimental difficulties and the requirement for reliable methods still render structure determination of large complexes a formidable challenge.

All methods for structure determination rely on highly purified, homogeneous material. Due to the transient nature of many complexes, purification by standard techniques is often impossible. Therefore, carefully developed and optimised strategies are necessary to obtain purified complexes. One way is affinity chromatography, e.g. with a TAP (tandem affinity purification) tag under mild conditions. (Rigaut *et al.*, 1999). Alternatively, a multi-subunit complex can be dissected in subcomplexes, which can be produced recombinantly (“*in vivo* reconstitution”) (Tan, 2001). Since isolated components of complexes are often dependent on an interaction partner for correct folding to a soluble protein, the art of this strategy lies in finding the correct combination of complex components (Tan, 2001). Yet another method is based on *in vitro* reconstitution of a complex from separately purified components. In this approach, complex assemblies, e.g. a fully recombinant archaeal RNA polymerase (Ouhammouch *et al.*, 2004; Werner and Weinzierl, 2002) or complete Pol II (Armache *et al.*, 2003) have been obtained. The protein/DNA/RNA

complexes of Pol II reported in this thesis have been assembled from endogenous, recombinant and synthetic material. In very rare cases, i.e. in the case of a Initiation Factor 1-(IF1)-ribosome complex (Carter *et al.*, 2001) and in this thesis, *in vitro* assembly was performed in pre-formed crystals by soaking these crystals in a solution of an interacting partner protein (see chapter 17).

The next hurdle in structure determination by crystallographic means lies in obtaining diffraction-quality crystals from the purified material. Buffer conditions have to be found under which the complex is monodisperse and stable enough over crystallographic time scales (several days to weeks). If initial crystallisation trials do not lead to promising results, the subunit composition or the nature of the constructs used offers an over-abundance of parameters to vary for subsequent experiments. Moreover, the length and the nature of single-stranded overhangs of co-crystallised DNA or RNA can be varied systematically, since in most nucleic-acid complexes, DNA and/or RNA is involved in crystal contacts.

The composition of a putative complex crystal can be analysed by gel electrophoresis, which has the disadvantage of sacrificing at least one, but typically several large crystals. Both the presence and the size of protein and nucleic acid components can be detected this way. A novel fluorescence-based assay, which allows rapid, reliable detection of nucleic acids was developed during the course of this thesis.

Structural data of multi-subunit complexes need not only come from crystallography. Lower-resolution information can also come from techniques such as small angle x-ray scattering (SAXS), which can achieve a resolution up to $\approx 10 \text{ \AA}$, (Koch *et al.*, 2003). SAXS proved to be particularly useful in cases where no crystals could be obtained or where additional information about conformational dynamics in solution was desired. In principle, SAXS can lead to *ab initio* models of biomolecules from several kilodaltons up to several megadaltons. A second technique complementary to x-ray crystallography is electron microscopy (EM) in its manifestations of electron crystallography, electron tomography and single-particle EM. In fortunate cases, a resolution of $\approx 10 \text{ \AA}$ can be obtained by multi-particle averaging techniques.

15 OPEN QUESTIONS AND SCOPE OF THIS WORK

Various crystal structures of Pol II and a minimal elongation complex (Cramer *et al.*, 2000; Cramer *et al.*, 2001; Gnatt *et al.*, 2001) gave substantial insights into the structural basis of the transcription process. However, these structures, as well as extensive biochemical work, left several questions unanswered, which were addressed in this work. The mechanism of duplex DNA separation and re-annealing before and after the transcription bubble, and the process of DNA/RNA strand separation had not been fully understood. This question was addressed by cocrystallising Pol II with a complete, synthetic transcription bubble plus product RNA. Moreover, no structural evidence how substrates are bound and selected by Pol II had existed. Therefore, structure determination of an NTP analogue bound to an elongation complex was another aim of this thesis. Finally, there were no structures of Pol II in complex with transcription factors. Such factors can bind transiently to Pol II and enhance or reduce the transcription rate by hitherto unknown mechanisms. The structure of TFIIS, the most prominent of these factors, bound to Pol II was solved in this work.

X-ray crystallographic studies of large Pol II complexes (molecular weight > 500 kDa) require enhanced techniques for complex assembly, crystallisation, data collection and analysis. The standard crystal form of 12-subunit Pol II has a very high solvent content, which impedes collection of high-resolution diffraction data. Therefore, one aim of this thesis was to construct a variant of the large Pol II subunit, which allows to proteolytically remove the unstructured C-terminal domain, leading to an improved crystal form. Two novel fluorescence-based assays were developed in order to investigate the complex crystal composition prior to x-ray analysis.

Results and Discussion

16 RNA POLYMERASE II-NUCLEIC ACID COMPLEXES

16.1 Assembly of complete Pol II-bubble-RNA complexes

The first prerequisite for any kind of structure determination is isolation of highly pure, homogeneous material. During this thesis, an existing purification protocol (Patrick Cramer, personal communication; Edwards *et al.*, 1990) was re-established and slightly modified in order to maximise reproducibility and yield (Armache *et al.*, 2003; see chapter 21.2). In essence, the previous DEAE chromatography step was replaced by a Uno-Q anion exchange column. Notably, Pol II purified in two days rather than three days, i.e. with the immunoaffinity and the UnoQ step performed in one day, does not yield crystallisable Pol II. For unclear reasons, this material would merely form showers of small crystals in the usual crystallisation conditions.

Complete yeast Pol II elongation complexes were reconstituted *in vitro* from endogenous 10-subunit core Pol II, recombinant Rpb4/7, and synthetic oligonucleotides. These are comprised of a 41-mer DNA duplex with an eleven nucleotide mismatched bubble region, annealed to an RNA 20-mer with eight complementary bases (see Table 6 and Materials and Methods). Besides this “standard” transcription bubble/RNA construct, a set of different DNA and RNA oligonucleotides was designed and used for purposes described in Table 6. Essentially, substitutions of uracils by 5-bromouracils as sequence markers at various positions are realised in a set of template and non-template DNA oligonucleotides. A set of RNA oligonucleotides with various 5'-overhangs was designed, as well as RNAs with and without a 3'-OH group for substrate studies (see Table 6).

Table 6: Overview of oligonucleotides used for crystallisation**Hybrid 1**

RNA	5' -	CCAGGC	-3'
TEMPLATE	3' -	CTGGTCCGAAA	-5'

☞ not detected in core Pol II crystals.

Hybrid 2

RNA	5' -	GACCAGGC*	-3'
TEMPLATE	3' -	CTGGTCCGAAA	-5'

☞ not detected in core Pol II crystals.

Hybrid 3

RNA	5' -	CAGCACGACCAGGC*	-3'
TEMPLATE	3' -	CTGGTCCGAAA	-5'

☞ not detected in core Pol II crystals.

Hybrid 4

RNA	5' -	GACCAGGC	-3'
TEMPLATE	3' -	CTGGTCCGAAA	-5'
		ACGCAGCACG	

☞ not detected in core Pol II crystals.

Hybrid 5

RNA	5' -	GACCAGGC	-3'
TEMPLATE	3' -	CCGTCATGATCAAACTGGTCCGAAA	-5'
NONTEMPLATE	5' -	CCGGCAGTACTAGT	-3'

☞ not detected in core Pol II crystals.

Bubble 1-Br

RNA	5' -	GACCAGGC	-3'
TEMPLATE	3' -	CCGTCATGATCAAACTGGTCCGAAAACGAACCGAACC	-5'
NONTEMPLATE	5' -	CCGGCAGTACTAGTAAACTAGTATTGAAAGTACTTGAGCTT	-3'

☞ complete transcription bubble.

Bubble 2-Br

RNA	5' -	UAUAUGCAUAAA	
RNA		GACCAGGC	-3'
TEMPLATE	3' -	CCGTCATGATCAAACTGGTCCGAAAACGAACCGAACC	-5'
NONTEMPLATE	5' -	CCGGCAGTACTAGTAAACTAGTATTGAAAGTACTTGAGCTT	-3'

☞ Complete transcription bubble. Structure solved with this construct

Backtracked Bubble

```

                                ACGCAGCACG          -3'
RNA          5' -                GACCAGGC
TEMPLATE    3' -    CCGTCATGATCAAAACTGGTCCGGAAACATGAACTCGAACC -5'
NONTEMPLATE 5' -    CCGGCAGTACTAGTAAACTAGTATTGAAAGTACTTGAGCTT -3'

```

➡ No density for backtracked RNA. Poor density for hybrid region.

Nontemplate-Br 1

```

                    UUAUUGCAUAAA
RNA          5' -                GACCAGGC          -3'
HK-DNA4 :   3' -    CCGTCATGATCAAAACTGGTCCGGAAACATGAACTCGAACC -5'
HK-BRDNA6A: 5' -    CCGGCAGTACTAGTAAACTAGTATTGAAAGTACTTGAGCTT -3'

```

➡ Br labels in non-template strand (1 of 2)

Nontemplate-Br 2

```

                    UUAUUGCAUAAA
RNA          5' -                GACCAGGC          -3'
HK-DNA4 :   3' -    CCGTCATGATCAAAACTGGTCCGGAAACATGAACTCGAACC -5'
HK-BRDNA6B: 5' -    CCGGCAGTACTAGTAAACTAGTATTGAAAGTACTTGAGCTT -3'

```

➡ Br labels in non-template strand (2 of 2)

RNA aptamer for nucleic acid staining assay (Thomas *et al.*, 1997)

```

5' - GGGCCGCUAAGUCCUCGCUCACGAUACAGCACUGAUUGCGG...
    ...UCGAGGUAGCUUGAUGGGCCACGCGCGACUCGGAUCCAC -3'

```

T = 5-Bromo uracil; **bubble region**; *: 3'-dideoxy cytidine

The complexes could readily be isolated from the excess of Rpb4/7 and nucleic acids by gel filtration over a Superose6 10/300HR column (Figure 13). Heparin-derivatised chromatography media retain apo-Pol II, whereas Pol II complexed with nucleic acids is not bound (Gnatt *et al.*, 1997). Consequently, an additional heparin column following the gel filtration binds less than 5 % of the Pol II-bubble/RNA complex after gel filtration (not shown), indicating satisfactory stability of elongation complexes.

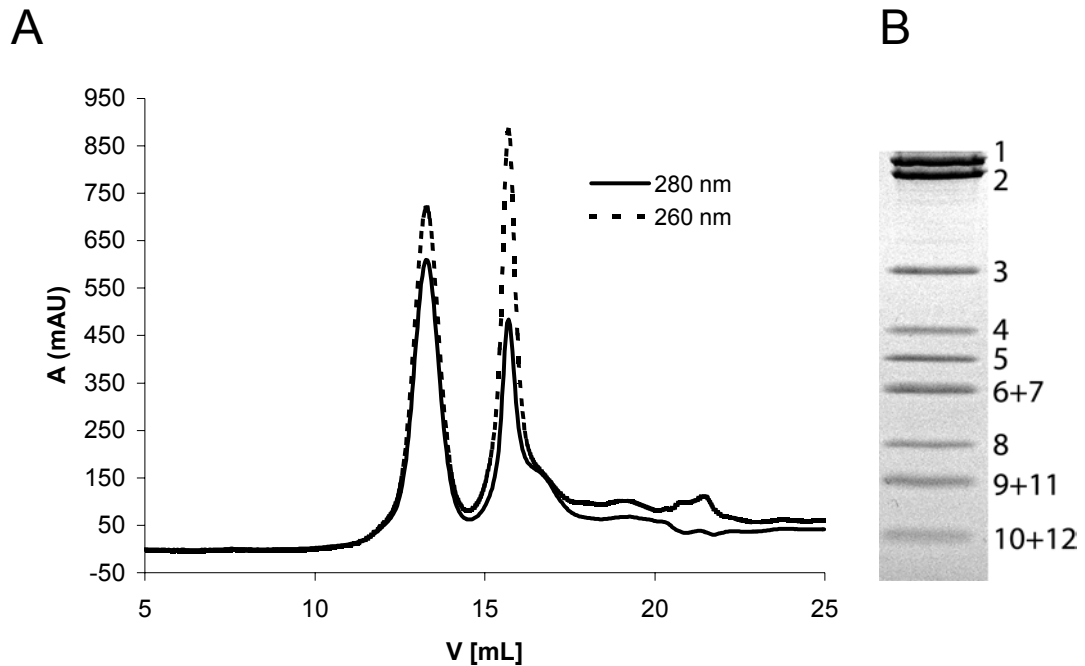


Figure 13: Purification of complete Pol II-DNA/RNA complexes.

(A) Size-exclusion chromatogram. The left peak corresponds to the Pol II-DNA/RNA complex and is clearly separated from the right peak, which contains the surplus of DNA/RNA and Rpb4/7.

(B) Coomassie-stained SDS-polyacrylamide gel of complete Pol II isolated as described in Materials and Methods. The subunits Rpb1-Rpb12 are indicated.

16.2 Characterisation of Pol II-DNA/RNA complexes

Binding of synthetic transcription bubble constructs to Pol II was assayed by electrophoretic mobility shift assay (EMSA). It was known that double-stranded DNA with a single-stranded 5'-overhang (tailed templates) bind to and can be elongated by purified Pol II *in vitro*. (Gnatt *et al.*, 1997). However, it was unclear whether a pre-annealed transcription bubble and a synthetic RNA product would assemble correctly with Pol II, especially when the Rpb4/7 subcomplex forces the clamp to a closed position.

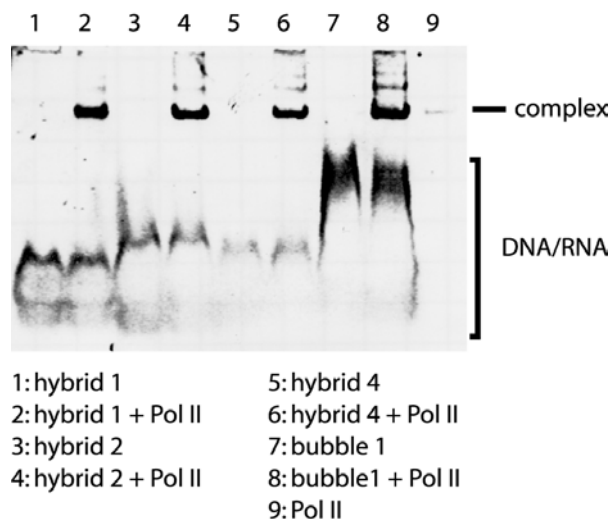


Figure 14: Synthetic transcription bubble/RNA constructs bind to complete Pol II.

100 pmol of nucleic acids described in Table 6 were incubated with 20 pmol of Pol II and the complex was separated from free nucleic acids by native gel electrophoresis as described in Materials and Methods. Nucleic acids were stained with SYBR-Gold.

Figure 14 shows an EMSA of several different DNA/RNA constructs, all of which strongly bind to Pol II. Remarkably, DNA/RNA hybrids number 1, 2 and 4 (Table 6) bind to Pol II, but are not visible in the electron density of core Pol II cocrystals. These hybrids most likely dissociate during crystallisation due to their short length. Thus, further studies were carried out with complete transcription bubble and product RNA constructs. Notably, nucleic acid binding does not depend on the order of addition of Rpb4/7 and nucleic acids to core Pol II within the scope of the bandshift assay. Presumably, the clamp conformation imposed by binding of Rpb4/7 has enough flexibility to allow entry of a DNA/RNA hybrid.

When NTPs are added to these Pol II elongation complexes, template-directed elongation of the RNA is observed (Figure 15). Thus, the elongation complexes are catalytically active.

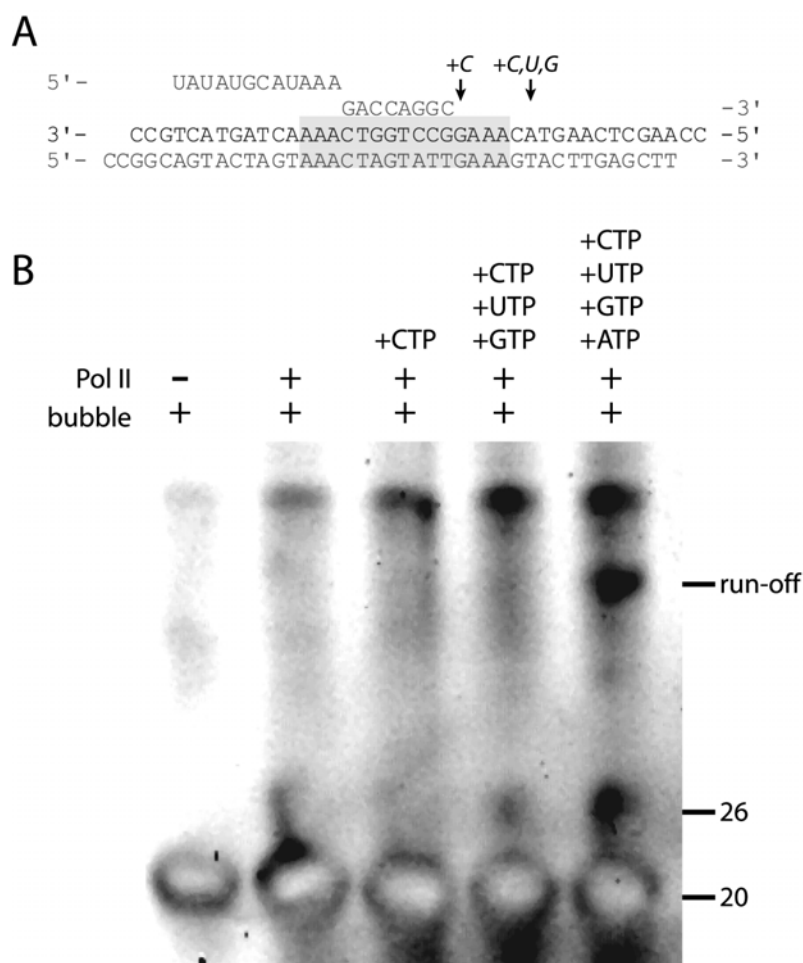


Figure 15: Elongation Assay.

(A) Complete Pol II was mixed with a synthetic transcription bubble of the given sequence. (T): template DNA; (N): non-template DNA. Addition of a subset of NTPs leads to elongation of RNA as indicated by arrows. The shaded box denotes the mismatched “transcription bubble” region.

(B) Denaturing gel electrophoresis of RNAs isolated from the transcription assay and stained with SYBR-Gold. The RNA length is indicated at the right. With the full set of NTPs (rightmost lane), the run-off product and a second, shorter product is observed.

16.3 Crystallisation of Pol II-nucleic acid complexes

The initial strategy of this project was to co-crystallise core Pol II with DNA/RNA hybrids #1 – 5 (Table 6), representing only the RNA and a template strand. Although such DNA/RNA oligonucleotides bound strongly to Pol II (Figure 14), no electron density was visible in crystals grown in the original PEG-phosphate conditions (row 1 in Table 7). These initial crystals resembled the fully shrunk form of core apo-Pol II (Cramer *et al.*, 2001) in both morphology and unit cell parameters.

Open questions (see chapter 15), unanswered by previous crystal structures of Pol II elongation complexes (Gnatt *et al.*, 2001; Kettenberger *et al.*, 2004; Westover *et al.*, 2004b) and T7 RNAP (see Table 2) prompted us to crystallise Pol II together with a whole transcription bubble. Core Pol II/bubble/RNA complexes refused to give rise to three-dimensional crystals, and a refined crystallographic model of the complete 12-subunit apo-Pol II became available during this project (Armache *et al.*, 2003). Therefore, the focus of this project was re-oriented towards achieving a model of a complete transcription bubble and product RNA in the context of complete, 12-subunit Pol II.

Well-diffracting crystals of core Pol II were obtained under conditions described earlier (Cramer *et al.*, 2001), and complete Pol II was initially crystallised in a modified version of these conditions (Armache *et al.*, 2003). In contrast to core Pol II crystals, diffraction of complete Pol II crystals could not be improved by dehydration described by (Cramer *et al.*, 2001). For Pol II-nucleic acid complexes, these conditions were not ideal due to their high phosphate concentration which generally fosters dissociation of complexes. Thus, complete Pol II and its nucleic acid complexes were subjected to commercial screens (Crystal screen 1+2, Natrrix[®], PEG/Ion screen; Hampton Research). Three-dimensional crystals were obtained in condition #38 of the Natrrix[®] screen (200 mM ammonium acetate, 150 mM magnesium acetate, 50 mM Hepes pH 7.0, 5% PEG4k, 5 mM TCEP) and in condition # 29 of Crystal Screen 1 (800 mM potassium-sodium tartrate; 100 mM HEPES pH 7.5). These initial conditions were refined to result in the conditions described in Table 7. The modified Natrrix[®] condition #29 turned out to give rise to reproducible, large crystals with satisfactory occupancy of nucleic acids (see chapter 16.4). The TFIIS-soak (chapter 17.1) is possible in both the Natrrix[®] condition and in the ammonium-sodium-tartrate condition (Table 7). TFIIS is however insoluble in the original phosphate/PEG conditions (Table 7).

Table 7: Overview of Pol II crystallisation conditions

Conditions	10 su Pol II	12 su Pol II	12su + bubble	cryoprotocol	comments	max. resolution
original PEG/PO₄ 390 mM NH ₄ NaPO ₄ , pH 6.0 15-18 % PEG 6k 50 mM dioxane 5 mM DTT	+	-	-	shrinkage protocol (Cramer <i>et al.</i> , 2001): 16% PEG6K, 100 mM MES pH 6.3, 50 mM dioxane, 350 mM NaCl, 17% PEG400 in 7 steps ; 4°C over 24 h	conditions for high- resolution 10 subunit Pol II	2.8 Å
PEG20K/PO₄ 320 – 360 mM NH ₄ NaHPO ₄ , pH 6.0 8-12 % PEG20k 25 mM NaCl 5 mM DTT	-	+	n.d.	mother solution + 0-(16-22)% glycerol in 5 steps; 8 °C over night	original 12 subunit Pol II conditions (Armache <i>et al.</i> , 2003)	4.2 Å
Natrix #38 200 mM ammonium acetate 150 mM magnesium acetate 50 mM Hepes pH 7.0 5 % PEG4k or 6k 5 mM DTT or TCEP	n.d.	+	+	mother solution + 0-20% glycerol in 5 steps; 8 °C over night		3.8 Å
Citrate/PEG 12 % PEG-6000 80-100 mM Na ₃ - citrate 100 mM Hepes pH 7.5 100 mM ammonium sulphate 5 mM DTT or TCEP	n.d.	n.d.	+	mother solution + 0-16% glycerol in 5 steps; 8 °C over night	2D and 3D crystals mixed. max. 120 by 80 by 30 µm	n.d.
Tartrate/Hepes 750 – 950 mM NH ₄ NaTartrate 100 mM Hepes-Na pH 7.5 5 mM DTT or TCEP	-	+	+	mother solution + 0-22% glycerol in 5 steps; 8 °C over night	modified Hampton screen 1 #29	3.8 Å

Conditions	10 su Pol II	12 su Pol II	12su + bubble	cryoprotocol	comments	max. resolution
Tartrate/PEG/ KSCN 12-13 % PEG-6000 300 mM ammonium- sodium tartrate 100 mM HEPES pH 7.5 100 mM KSCN 5 mM DTT or TCEP	n.d.	+	+	mother solution + 0-16% glycerol in 5 steps; 8 °C over night	good and reproducible for bubble complexes	3.8 Å
Tartrate/PEG additive 12-13 % PEG-6000 300 mM ammonium- sodium tartrate 100 mM HEPES pH 7.5 10 mM spermidine or 50 mM ethylene diamine or 50 mM LiCl 5 mM DTT or TCEP	n.d.	n.d.	+	mother solution + 0-16% glycerol in 5 steps; 8 °C over night	additives give bigger crystals, esp. if bubble is present in excess.	n.d.

+: 3D crystals obtained; -: no 3D crystals obtained; n.d.: not determined

A small number of larger crystals is typically achieved with a protein : reservoir ratio of 2 : 1, whereas use of sitting drop plates or microdialysis crystal set-ups did not lead to improved crystal size. At a later stage of this work, crystal size could be improved by covering the reservoir solution with 200 – 400 μ L of Al's oil (paraffin oil : silicon oil 1 : 1; Hampton Research). Crystals grew to an ideal size of about 0.4 x 0.3 x 0.2 mm within 1-2 weeks at 20 °C. Larger crystals often showed damage due to cracking during cryo treatment.

Early electron density maps showed poor electron density for nucleic acid regions (see chapter 16.7). A drastic improvement was achieved by adding a slight excess (1-2 μ M) of nucleic acids to the crystal setup in order to achieve full occupancy of nucleic acid sites (see Materials and Methods). Figure 16 shows typical representatives of crystals obtained under conditions described in Table 7.

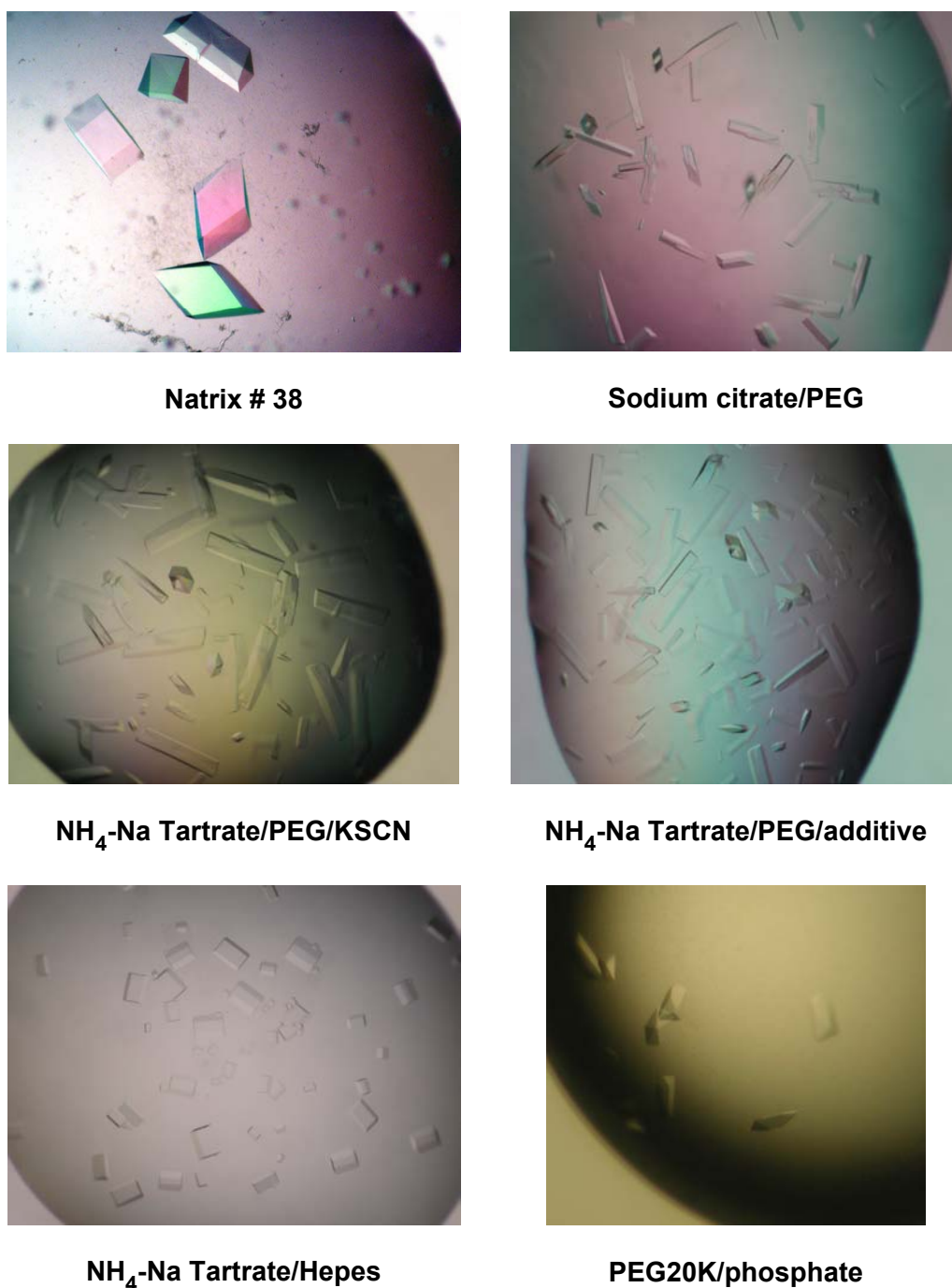


Figure 16: Typical representatives of crystals of the complete Pol II/nucleic acid complex, obtained under various conditions described in Table 7.

16.4 Characterisation of complete Pol II-DNA/RNA cocrystals

Crystals of the 12-subunit Pol II are characterised by their unusually high solvent content of 80 % (Matthews coefficient of 6.2). This is the highest solvent content of all Pol II crystal forms described so far (Table 8). Consequently, only a slow, stepwise procedure allowed adding the

necessary amounts of glycerol as cryoprotectant to the mother solution surrounding the crystals. Other cryoprotectants such as PEG400, sucrose, 2-methyl-propanediol (MPD), ethylene glycol, propylene glycol, and glucose destroyed diffraction. Crystals are extremely sensitive to any kind of osmotic stress. Diffraction quality must be further enhanced from about 6 Å up to 3.8 Å by slowly cooling crystals to 4 – 8 °C in the last cryo-solution. Cryo-cooling in liquid nitrogen was performed after storage over night in the last cryo-solution. Further lowering the temperature to –20 °C before cryocooling did not improve diffraction, possibly because of the high amounts of glycerol (30 %) necessary to avoid ice formation at –20 °C. Likewise, diffraction was hampered by cryocooling in mother solution without additives using supercooled liquid propane.

Table 8: Crystal forms of core and complete Pol II and Pol II complexes

Crystal Form	Unit Cell (Å)	Space group	Solvent content	Reference
Free core Pol II, form 1	131, 225, 369	I222	58 %	(Cramer <i>et al.</i> , 2000)
Free core Pol II, shrunk form	123, 223, 376	I222	56 %	(Cramer <i>et al.</i> , 2001)
Minimal elongation complex ¹	157, 221, 191 $\beta=97.5^\circ$	C2	65 %	(Gnatt <i>et al.</i> , 2001; Westover <i>et al.</i> , 2004b)
Complete Pol II	220, 391, 282	C222 ₁	80 %	(Armache <i>et al.</i> , 2003) and this study

¹ All complexes of core Pol II and nucleic acids exhibit a very similar crystall lattice (Westover *et al.*, 2004a; Westover *et al.*, 2004b).

All DNA/RNA oligonucleotides cocrystallised with complete Pol II resulted in the same crystal form, regardless the length and nature of dsDNA ends and overhangs of all constructs tested.

16.5 Data collection

Having harvested a total of about 700 crystals of the 12-subunit Pol II crystal form, of which about 300 crystals were screened at the synchrotron, a great deal of experience in judging the quality and in optimising data collection could be gained. Mosaicities < 0.5 greatly facilitate data

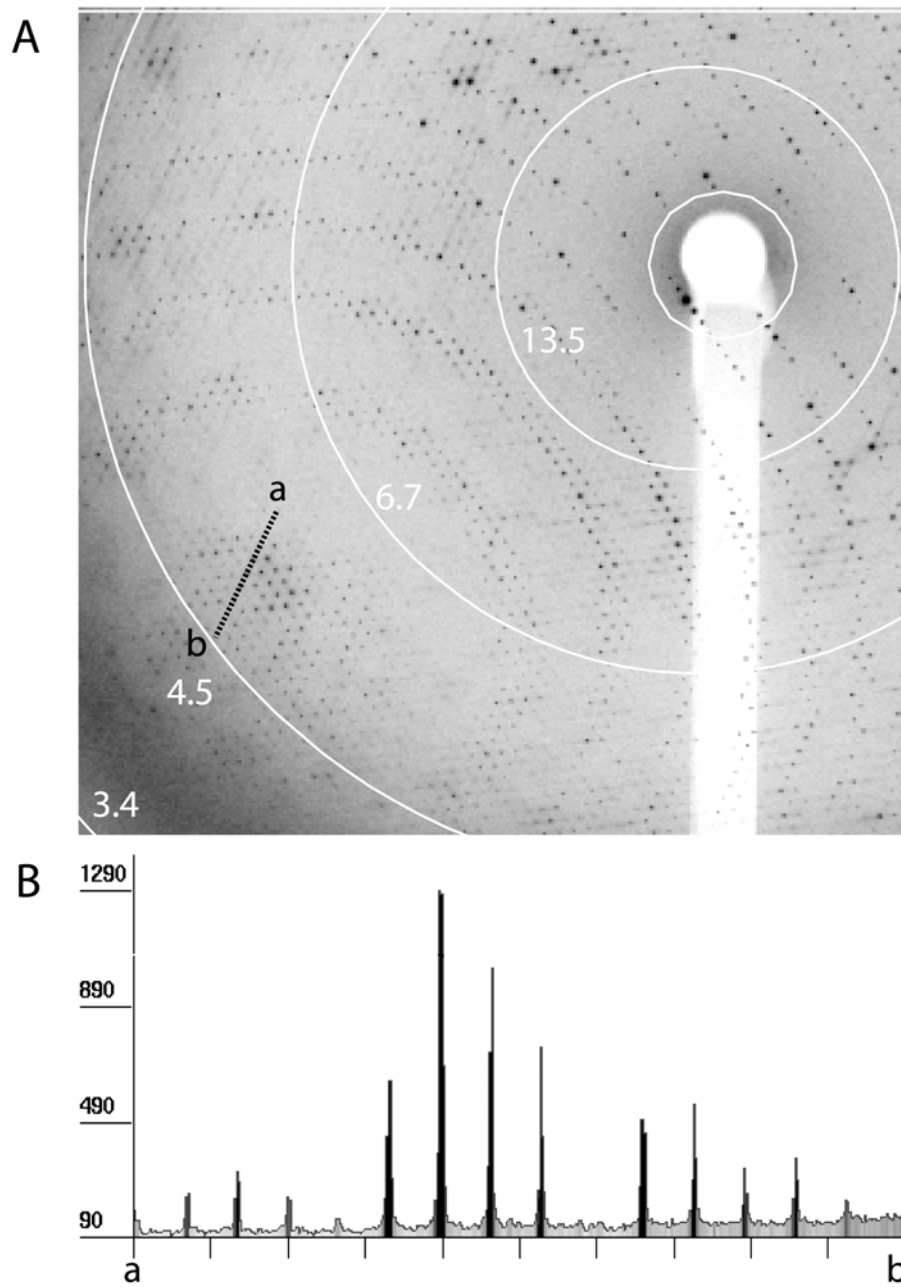


Figure 17: Diffraction pattern of a complete Pol II-transcription bubble-RNA cocrystal.

(A) The diffraction pattern was recorded with a Mar MosaicCCD 225 detector at beamline X06SA of the SLS, at a wavelength of 0.98 Å and a crystal-detector distance of 400 mm. White circles indicate the resolution and a dashed line marks the path of the linescan in panel (B).

(B) Linescan between points a and b in panel (A). The vertical axis represents the diffraction intensity (arbitrary units).

processing and improve statistics. Radiation damage is the factor, which limits the resolution and quality of diffraction data of the 12-subunit crystal form. Thus, particular care was taken to reduce both the dose and the background due to diffuse scattering. Both aims were achieved by

routinely taking diffraction images of small 0.25° oscillation increments at the longest reasonable crystal-detector distance and an x-ray beam focused on the detector. “Light” mother solutions, i.e. those, which do not contain phosphorous or other heavier elements, were preferred to reduce diffuse scattering from the solvent.

A considerable improvement was observed when the detector at the SLS beamline X06SA was changed from a MarCCD165 with 165 mm diameter to a Mar225 MOSAIC CCD detector with 225 mm diameter. This larger detector allows longer crystal-detector distance at a given resolution cut-off. Since background due to diffuse scattering falls off with $1/d^2$ (d being the crystal-detector distance), any increase in the distance reduces the background on diffraction images and thus increases the signal-to-background ratio.

Recently, the use of radical scavengers such as ascorbic acid was shown to reduce radiation sensitivity in some cases (Garman, 2003; Garman and Nave, 2002; Murray and Garman, 2002). First experiments with Pol II crystals indicate that 100 mM of ascorbate or 5 mM of TEMPO (2,2,6,6-tetramethylpiperidiny-N-oxid) might slightly prevent radiation damage. However, the availability of sufficiently large crystals and a reliable data collection strategy largely bypassed the need for additional stabilisation by radical scavengers. A dose necessary to obtain $I/\sigma(I)$ values greater than about 4 in the highest resolution shell was best judged by visual inspection of diffraction images during data collection. Typical exposure times at the SLS beamline X06SA ranged between 1.0 and 1.5 seconds of unattenuated beam per 0.25° at a wavelength of 0.91 \AA and a ring current of 350 mA. Whenever possible, an oscillation range $\geq 100^\circ$ was collected per dataset and the crystal was translated when the mosaicity increased or the intensities of high-resolution spots decreased noticeably. Radiation damage did not lead to a reproducible increase of unit cell volume as described (Murray and Garman, 2002). However, a sometimes dramatic increase of mosaicity during data collection could be observed (Figure 18).

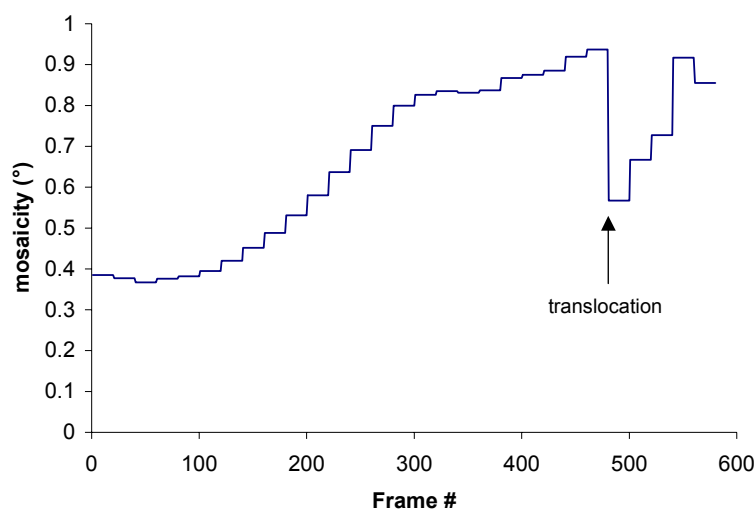


Figure 18: Typical mosaicity increase during data collection.

In most Pol II datasets collected during this work, mosaicity increased during data collection, presumably due to radiation damage. The arrow indicates a translocation of the crystal in the x-ray beam, so that a “fresh” portion of the crystal is irradiated. Together with increasing mosaicity, decreasing $I/\sigma(I)$ and disappearing high-resolution reflections are observed.

16.6 Phasing and refinement

The phase problem was solved by molecular replacement with the core Pol II elongation complex (PDB-ID 1I50; Gnatt *et al.*, 2001; $R_{\text{free}} = 36.8\%$), with the model of core Pol II and the archaeal homologues of Rpb4/7 (PDB-ID 1NT9; Armache *et al.*, 2003; $R_{\text{free}} = 36.0\%$) or, from the time it was available, with the refined model of 12-subunit Pol II (PDB-ID 1WCM; Armache *et al.*, 2004; $R_{\text{free}} = 27.6\text{--}29.7\%$). The resulting electron density strikingly depends on the nature of the phasing model (Figure 20). Only the atomic model of the complete Pol II results in maps that show excellent electron density for the nucleic acid regions. Density modification (CNS), however, annihilated much of the additional electron density and was not used further on.

Backbone phosphate groups and individual nucleotides were resolved, but the identity of the nucleotides was not revealed at the given resolution of 3.8–4.5 Å. Therefore, 5-bromouracil was introduced in various positions of the DNA template and non-template strand. A bromine atom in the nucleotide at position -4 of the template strand was evident in an anomalous difference Fourier map, defining the register of nucleic acids (Figure 20B). However, no anomalous signals above noise were detected for the other bromine labels at positions +7, +12, -7, -14 and -20 in

the template strand, and at positions +6, -1, -2, -4, -5, -7, -9, -10, and -12 in the non-template strand, indicating high mobility of DNA before and after the hybrid.

Nucleic acids were built into unbiased electron density. Starting with the refined 12-subunit Pol II model, positional and B-factor refinement (20 cycles each, target: maximum likelihood function) converged readily at a free R-factor of 27.6 % (Table 9).

Table 9: X-ray diffraction data and refinement statistics

Dataset number	1	2	3
Crystal composition	Pol II-bubble-RNA ^a	Pol II-bubble-RNA (with Br-dU) ^{a,b}	Pol II-bubble-RNA + GMPCPP ^{a,c}
Unit cell axes (Å) ^e	221.4, 392.5, 283.2	220.6, 393.1, 282.0	222.5, 393.1, 283.7
Wavelength (Å)	0.9200	0.9195 (Br peak)	0.97960
Resolution range (Å) ^f	50-4.0 (4.14-4.0)	50-4.2 (4.35-4.20)	50-4.5 (4.66-4.50)
Unique reflections ^f	102020 (10148)	69541 (6559)	58178 (5134)
Completeness (%) ^f	99.3 (99.5)	77.5 (73.8)	78.4 (69.8)
Redundancy ^f	4.0 (4.0)	4.9 (4.3)	2.8 (2.5)
Mosaicity (°)	0.35-1.23 ^d	0.51	0.20-0.54 ^d
R _{sym} (%) ^f	9.0 (27.9)	7.6 (18.6)	8.1 (23.1)
I/σ(I) ^f	13.7 (4.8)	16.6 (7.2)	7.2 (3.0)
Peaks in anomalous Fourier	-	8 Zn in Pol II (15.1-8.1 σ) 1 Br in DNA (5.4 σ)	-
R _{free} after rigid body refinement (%)		28.6	29.7
Refinement			
Number of residues	3926 amino acid residues, 36 nucleotides		
RMSD bonds (Å)	0.010		
RMSD angles (°)	1.55		
R _{cryst} (%)	25.3		
R _{free} (%)	27.6		
PDB-ID	1Y1W		1Y77

^a Diffraction data were collected at beamline X06SA at the Swiss Light Source.

^b The complex contained a DNA template strand in which thymine at position -4 was replaced by the isosteric bromouracil, and the G at position +1 was exchanged to C.

^c The complex was cocrystallized with the nonreactive NTP analogue GMPCPP.

^d Due to radiation damage, mosaicity increased during data collection, and was refined in segments of five frames.

^e Crystals belong to the space group C222₁.

^f Values in parentheses correspond to the highest resolution shell.

16.7 Overview of nucleic acid structure

The final model comprises all 12 Pol II subunits, and 19, 7 and 10 residues of template DNA, non-template DNA, and RNA, respectively (Figure 19).

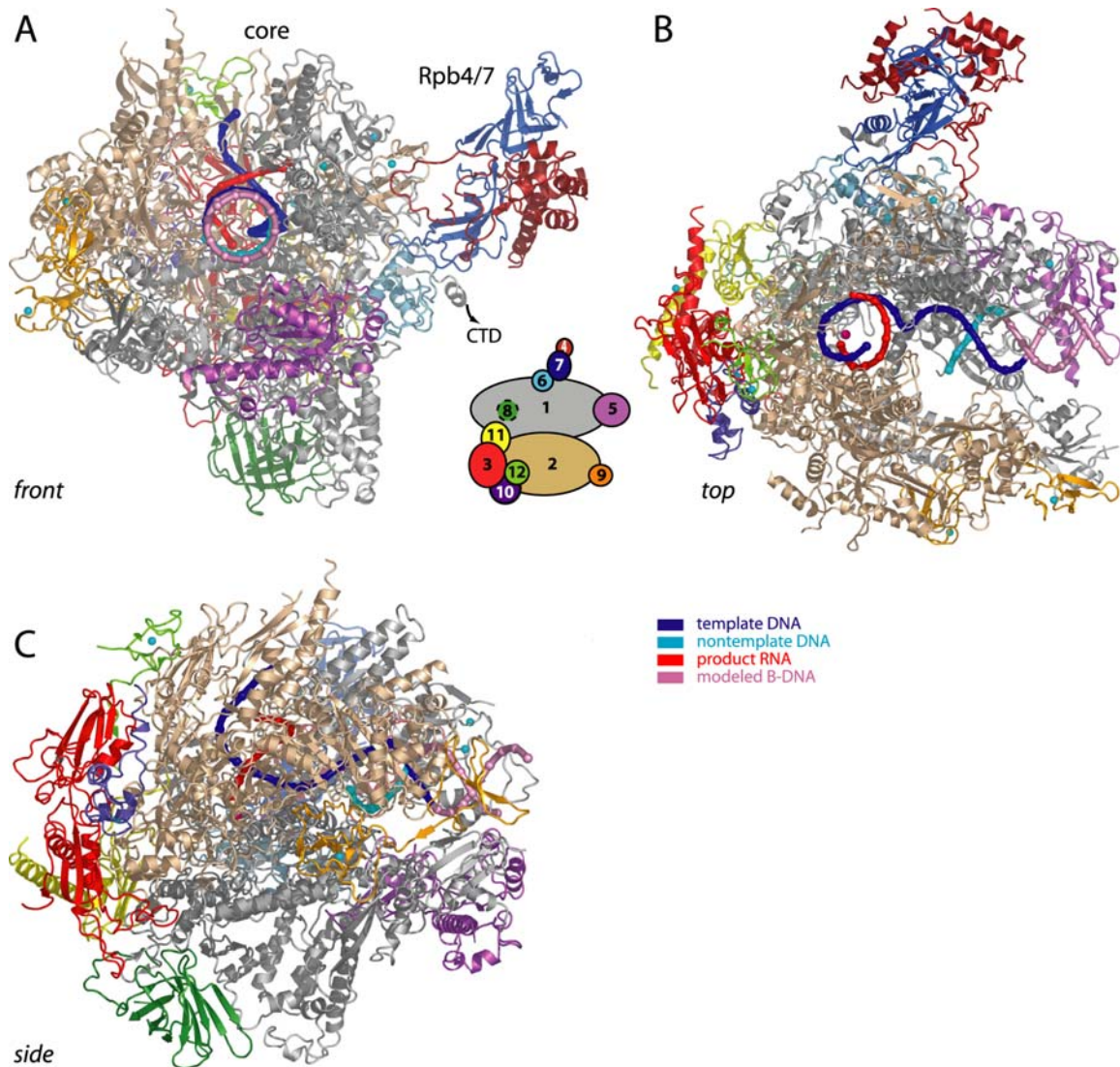


Figure 19: Structure of the complete Pol II elongation complex.

(A)-(C): Three views of a ribbon model of the protein subunits and nucleic acids. The polymerase subunits Rpb1-Rpb12 are colored according to the key between the views. Template DNA, nontemplate DNA and product RNA are shown in blue, cyan and red, respectively. Phosphorous atoms are indicated as spheres and extrapolated B-form downstream DNA is coloured in light pink. Eight zinc ions and the active site magnesium ion are depicted as cyan spheres and a magenta sphere, respectively. This color code is used throughout. Secondary structure assignments for Pol II are according to (Cramer *et al.*, 2001) and (Armache *et al.*, 2004).

Strong continuous electron density is observed for seven base-pairs of downstream DNA before the active site, for the DNA-RNA hybrid, for the template strand connecting between the downstream duplex and the hybrid, and for three unpaired nucleotides of both DNA and RNA upstream of the hybrid (Figure 20). The complex adopts the post-translocation state, with no density in the presumed NTP-binding site (position +1; Figure 20A, D). The nucleic acids are not involved in crystal packing, and their location is thus determined by interactions with the polymerase, providing an unbiased view of the elongation complex.

Although all nucleic acids of a natural elongation complex are present in the crystals, only selected regions are bound by the polymerase and thus visible in the electron density. DNA entering the cleft is highly mobile, but is bound for strand separation closer to the active site (Figure 20D). Downstream DNA runs along the Rpb1-face of the cleft (Figure 19). Upstream of the point of downstream DNA strand separation, the DNA template strand projects the coding base in the active site. The DNA-RNA hybrid forms numerous polymerase contacts (Figure 20A, C) which are generally in accordance with earlier observations (Gnatt *et al.*, 2001), and confer stability and thus processivity to the elongation complex (Kireeva *et al.*, 2000b). Upstream of the hybrid, the polymerase continues to contact the RNA and the DNA template strand, and these contacts may be mainly responsible for strand separation and proper RNA exit. The upstream DNA duplex and the non-template strand in the bubble are not observed, and they do not contribute significantly to elongation complex stability (Kireeva *et al.*, 2000b).

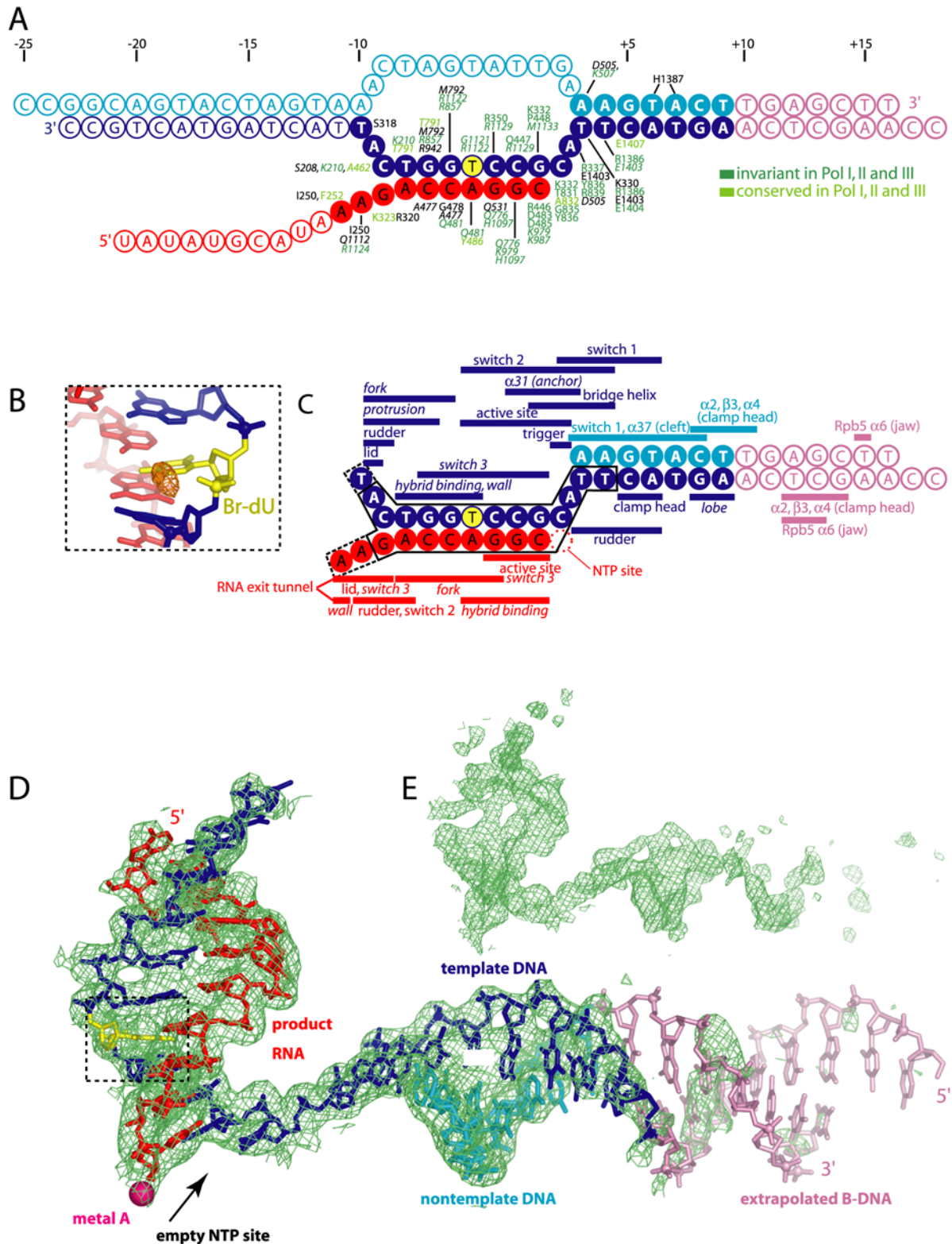


Figure 20: Nucleic acids in the elongation complex.

(A) Nucleic acid constructs and Pol II interactions. Filled and open circles denote nucleotides that are ordered and disordered, respectively, in the electron density of the cocrystals. Pol II residues that are within 4 Å distance of nucleic acids are depicted. Residues that are invariant, conserved, and differing among the three yeast nuclear RNA polymerases are in dark green, light green, and black, respectively. Italics are used to distinguish Rpb2 from Rpb1 residues. The yellow circle marks the thymine substituted

Figure 20 (continued)

by 5-bromo uracil and localised by anomalous diffraction. Nucleotide positions are numbered with respect to the nucleotide addition site (position +1). Upstream and downstream nucleotides/positions have negative and positive numbers, respectively.

(B) Bromine anomalous difference Fourier map. The map was calculated from data set 2 (Table 9), taken at the bromine absorption maximum, and phases from the complete Pol II structure without nucleic acids. It is contoured at 4.0σ . The observed peak has a height of 5.4σ and coincides with the location of the bromine atom in a 5-bromouracil residue incorporated at position -4 of the DNA template strand, thereby defining the register of nucleic acids. The view is related to that in Fig. 2A by a 180° rotation about a vertical axis.

(C) Long-range Pol II–nucleic acid interactions. Bars indicate Pol II domains and elements within a distance of 8 \AA from the nucleic acids (Cramer *et al.*, 2001). The extent and color of the bars define the DNA or RNA nucleotides contacted by Pol II. Rpb2 elements are in italics. The encircled nucleotide positions have previously been observed in Pol II–nucleic acid complex structures [solid black line (Gnatt *et al.*, 2001); dashed black line, (Westover *et al.*, 2004b)].

(D) Electron density map. A $2F_o-F_c$ map was calculated from data set 1 (Table 9), using phases from the final model with nucleic acids omitted. The map is contoured at 1.0σ . The final model for the nucleic acids is superimposed. The region within the dashed box is magnified in 1B. Note the absence of electron density in the substrate-binding (+1) site.

(E) Initial electron density map. The $2F_o-F_c$ map has been calculated with the same experimental data as in (D), but was phased with the incomplete 10-subunit core Pol II structure (PDB-ID 1I6H). The map is contoured at 1.0σ .

16.8 DNA unwinding

In contrast to previous structures of Pol II–nucleic acid complexes (Gnatt *et al.*, 2001; Westover *et al.*, 2004b), seven base pairs of downstream DNA are observed, at positions +3 to +9 (Figure 19, Figure 20). The downstream double-stranded DNA forms direct contacts with Pol II, and is surrounded by several polymerase elements (see Figure 20C and Figure 21 for details). The ordered part of the downstream DNA occupies a DNA-binding site identified previously in bacterial polymerase (Nudler, 1999; Nudler *et al.*, 1996; Nudler *et al.*, 1998), and its sequence can influence polymerase pausing (Palangat *et al.*, 2004; Palangat *et al.*, 1998). The overall position of downstream DNA agrees with those inferred from electron microscopy (Poglitsch *et al.*, 1999) and from scattered electron density features in a crystallographic map (Gnatt *et al.*, 2001). The observed conformation of downstream DNA suggests how the template strand is removed from the non-template strand before reaching the active site. DNA at register +5 to +9 is essentially in canonical B-form, but the template strand deviates from this conformation at position +2 to +4 (Figure 21A). Apparently, this deviation is achieved by three positively charged residues in switch 2 (R326, K330, R337 of Rpb1), which cause the template strand to deviate from the duplex axis. Three negatively charged residues in switch 1 (E1403, E1404, E1407 of Rpb1) repel the DNA strand (Figure 21A). Thus, an enzyme-induced distortion and destabilisation of the incoming DNA duplex may drive strand unwinding. In addition, the

polymerase fork loop 2 sterically blocks duplex binding, interferes with the non-template strand upstream of position +3, and prevents re-association of separated strands (Figure 21).

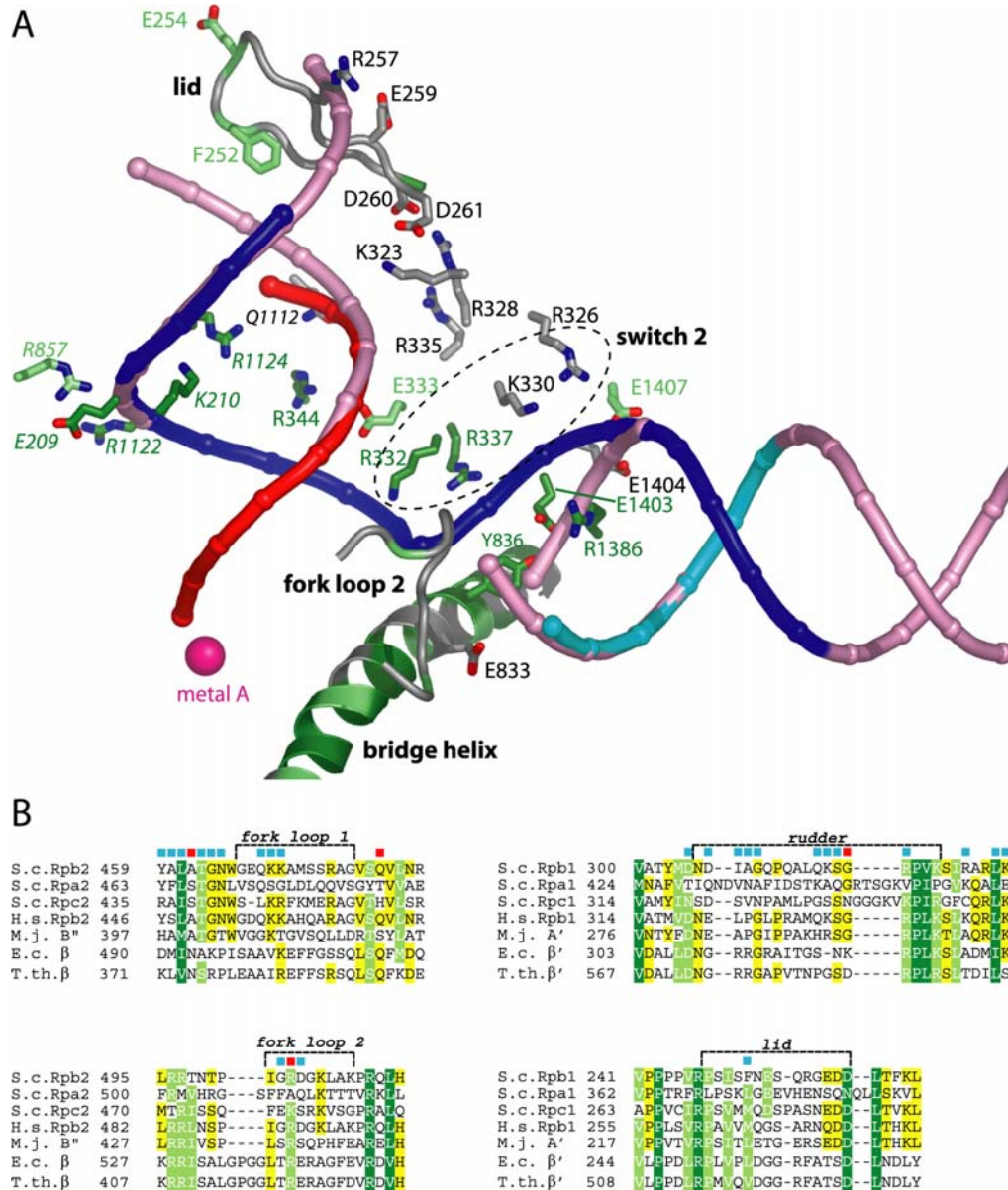


Figure 21: Nucleic acid strand separation.

(A) DNA unwinding and DNA-RNA strand separation. Nucleic acid backbones are colored as in Figure 20. Spheres represent the location of phosphorous atoms. Downstream DNA that has been extrapolated as a canonical B-form duplex is in light magenta. The DNA/RNA hybrid at positions -1 to -5 has been used to extrapolate the hybrid structure upstream and is also shown in light magenta. The active site metal ion A is shown as a pink sphere. Pol II residues that are apparently involved in nucleic acid strand separation are colored according to their conservation in Pol I, II and III as in Fig. 2A. The bridge helix, fork loop 2, and the lid are depicted. Basic amino acids in switch 2 that may pull the template strand upwards are encircled with a dashed line.

Figure 21 (continued)

(B) Sequence alignments of loops that partition the cleft. Amino acid alignments of the two fork loops in Rpb2 and the rudder and lid in Rpb1 are colored according to conservation between eukaryotic, archaeal, and bacterial RNA polymerases. Invariant, conserved, and weakly conserved residues are colored in dark green, light green, and yellow, respectively. Red and blue squares above the alignments indicate residues within 4 Å and 8 Å distance from nucleic acids, respectively. Sequences in subunits of Pol II of *S. cerevisiae* (S.c.), human (H.s.) and *Thermus thermophilus* (T.th.) are aligned with their homologues in S.c. Pol I and III, and *E. coli* (E.c.) and *Methanococcus janaschii* (M.j.) RNA polymerase.

16.9 RNA displacement and exit

The hybrid comprises seven base pairs in the post-translocation state, and would be eight base pairs long upon nucleotide incorporation. At register -8 , the Watson-Crick edges of the bases are ≈ 4 Å apart, indicating that the RNA begins to separate from the DNA template strand. At position -9 and -10 , the RNA deviates from an extrapolated course beyond the point of strand separation (Figure 21A), suggesting that Pol II induces a distortion of the stable heteroduplex structure. The separating strands form backbone contacts with Pol II residues that are generally conserved among nuclear RNA polymerases (Figure 20A, Figure 21A, and K. Armache [unpublished data]), suggesting that these enzymes make use of a common mechanism. Contacts of Pol II to RNA positions -8 to -10 may constitute a previously characterized RNA-binding site that contributes to elongation complex stability (Nudler, 1999). The contacts to the separating strands were not observed in the tailed-template structure (Gnatt *et al.*, 2001), apparently due to formation of an overextended hybrid. Lack of these contacts explains the negative effect of an overextended hybrid on elongation complex stability (Kireeva *et al.*, 2000a).

In principle, it is possible that the design of the nucleic acids used in this project, and in particular the A-A mismatch at position -9 , result in an artificial situation, and that the hybrid in a natural elongation complex may be one base pair longer. Indeed it cannot be excluded that a different nucleic acid sequence at the point of DNA-RNA strand separation would allow for Watson-Crick base-pair formation at position -8 , since only a 1 Å decrease in the distance between the bases would be required. Overall, however, the location of the separating upstream strands detected here is very similar to that seen in the Pol II core complex with a synthetic hybrid (Westover *et al.*, 2004b), although the sequence of nucleic acids differs, strongly arguing for the sequence-independent nature of the nucleic acid locations and the polymerase-nucleic acid interactions observed here.

Maintenance of the upstream end of the hybrid also involves the lid, a prominent loop that protrudes from the edge of the clamp (Figure 21, Figure 22). Modelling shows that an extrapolated template strand clashes with the lid (Figure 21A).

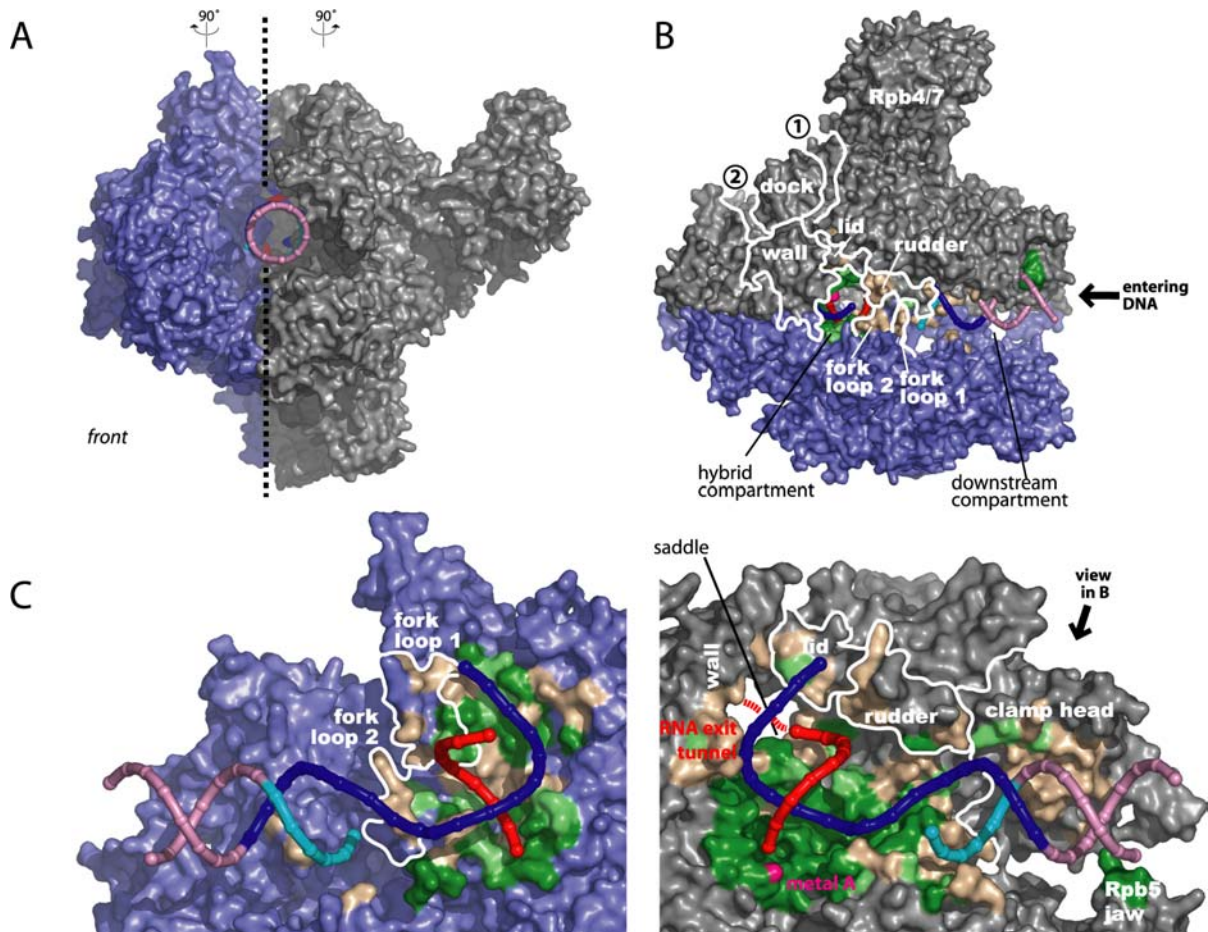


Figure 22: Conservation and compartmentalization of the polymerase cleft.

(A) Overall structure of the Pol II-bubble-RNA complex. The complete Pol II is shown as a molecular surface, and nucleic acid backbones are drawn as ribbons. The view corresponds to the front view (Cramer *et al.*, 2000; Cramer *et al.*, 2001). A dashed line indicates the slice plane used to create the views in (C).

(B) Top view of the model in (A). Pol II loops are outlined that partition the enzyme cleft. During transcription elongation, DNA enters from the right. Previously proposed RNA exit grooves are labeled ① and ②.

(C) Conservation of nucleic-acid interaction surfaces. The model in (A) was intersected along the plane indicated in (A), and the resulting halves were rotated by 90° around a vertical axis in opposite directions. The molecular surface of residues within 8 Å distance from nucleic acids is colored in beige. Residues that are invariant and conserved between Pol I, II and III are highlighted in dark and light green, respectively. Pol II elements that contact nucleic acids and partition the enzyme cleft are outlined.

Except phenylalanine 252, the lid residues do not contact separating nucleic acids, and are weakly conserved between the three nuclear RNA polymerases (Figure 21B). Hence, the lid apparently plays a steric role in DNA-RNA strand separation. The lid and the saddle between the wall and the clamp (Cramer *et al.*, 2001) form a narrow RNA exit tunnel that only allows passage of a single strand (Figure 22C). A similar ordering of the loops occurs in the Pol II core bound to a DNA-RNA hybrid (Westover *et al.*, 2004b). The RNA 5'-end is located at the entrance to this tunnel, and its extension through the tunnel leads to two potential RNA-binding grooves that flank the Pol II dock domain (Cramer *et al.*, 2000; Cramer *et al.*, 2001; Figure 22B). There is no electron density in the grooves although the RNA in the complex crystals is long enough to reach them, indicating that RNA does not interact stably with the grooves in the absence of other factors.

16.10 Initiation-elongation transition

After transcription initiation, growing RNA roving along the identified path may trigger structural transitions that lead to a stable elongation complex. Growing RNA would displace the initiation factor TFIIB, which binds to the dock domain (Chen and Hahn, 2003), and extends over the saddle into the hybrid site (Bushnell *et al.*, 2004). The transition from the closed pre-initiation complex to an open complex includes movement of downstream DNA, which lies inside the cleft during elongation, but is apparently suspended above the cleft in the initiation complex (Armache *et al.*, 2003; Bushnell and Kornberg, 2003; Craighead *et al.*, 2002). Upstream DNA is mobile during elongation, but during initiation it is bound by several factors, and may lie above the cleft (Chen and Hahn, 2004), as in a bacterial RNA polymerase–DNA complex (Murakami *et al.*, 2002a).

The transition also goes along with an almost complete embedding of the DNA-RNA hybrid, the hallmark of the elongation complex (Figure 22). Two loops that extend from opposite sides of the cleft, the rudder and fork loop 1, are mobile in free Pol II (Cramer *et al.*, 2001), but are ordered and contact each other in the elongation complex, creating two compartments in the cleft for holding downstream DNA and the hybrid (Figure 22; Westover *et al.*, 2004b). Given the apparent importance of the lid, fork loops, and rudder for separation of the nucleic acids, it is surprising that these loops are not well conserved in Pol I and Pol III, and even deviate in length (K. Armache [unpublished results], Figure 22B). Apparently, these loops are only responsible for a passive topological partitioning of the cleft.

The initiation-elongation transition was structurally characterized for the single subunit RNA polymerase of phage T7, which does not require initiation factors. A large portion of this polymerase refolds during the transition, to create a hybrid-binding site and an RNA exit tunnel (see Figure 4; Steitz, 2004; Tahirov *et al.*, 2002; Yin and Steitz, 2002). In Pol II, the hybrid site and exit tunnel pre-exist in the initiation-competent form of the enzyme, but are partially filled during initiation with TFIIB (Bushnell *et al.*, 2004), and possibly also with the general transcription factors TFIIE (Forget *et al.*, 2004; Kim *et al.*, 2000; Meinhart *et al.*, 2003a) and TFIIIF (Chung *et al.*, 2003). These factors may all contribute to the stabilisation of an early transcribing complex, but must liberate the hybrid site and RNA exit tunnel during formation of a stable elongation complex. Whereas the initiation complexes of the single-subunit polymerase from T7 bacteriophage and Pol II are very different, the arrangement of nucleic acids in the elongation complexes is similar (Figure 25).

16.11 NTP binding, selection, and incorporation

The electron density shows an empty space at the 3'-end of the RNA, which corresponds to the presumed binding site for substrate NTP (Figure 20D). The DNA template cytidine at position +1 points towards this site, ready for binding a complementary GTP. This provided an incentive to co-crystallise the complex with the non-reactive GTP analog GMPCPP (see chapter 28.1). The resulting small co-crystals diffracted weakly and were very radiation sensitive, but complete diffraction data to 4.5 Å resolution were obtained by merging datasets from five crystals (Table 9).

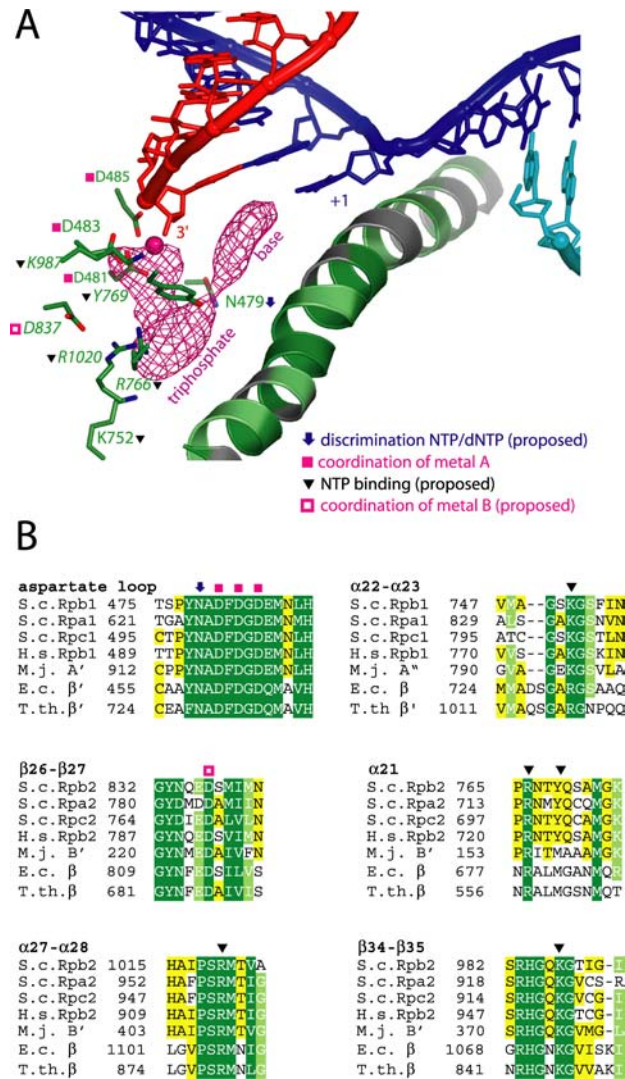


Figure 23: NTP-binding site.

(A) Fourier difference electron density for a GTP substrate analog. The F_0-F_c map was calculated from data of the Pol II–bubble–RNA complex after cocrystallization with GMPCPP (Table 9), was phased with the final elongation complex structure, and is contoured at 2.8σ . Two prominent lobes of electron density are seen in the presumed NTP-binding site that is empty in the Pol II–bubble–RNA structure. One lobe can represent the nucleotide base near the template DNA base at position +1 (the “coding” base). The other lobe may be attributed to the triphosphate moiety of the NTP. Invariant Pol II residues in the vicinity of the density peak are shown. Between the two lobes, the ribose moiety would be located, at a position where its 2'-OH group could form a hydrogen bond with Rpb1 residue N479, consistent with a role of this residue in NTP over dNTP selection.

(B) Sequence alignment of residues surrounding the NTP site. Sequences in subunits of Pol II of *S. cerevisiae* (S.c.), human (H.s.) and *Thermus thermophilus* (T.th.) are aligned with their homologues in S.c. Pol I and III, and *E. coli* (E.c.) and *Methanococcus janaschii* (M.j.) RNA polymerase.

A difference Fourier map revealed electron density in the Pol II pore near the presumed NTP site between the RNA 3'-end and the bridge helix (Figure 23A). Two density lobes can be ascribed to the nucleotide base and triphosphate moiety (Figure 23A). The nucleotide base is slightly further away from the DNA template base at position +1, as required for Watson-Crick

base pairing. The NTP ribose would be close to the side chain of Rpb1 residue N479, which could bind the ribose 2'-OH group, selecting for ribonucleotides (Gnatt *et al.*, 2001). Indeed NTP versus desoxy-NTP discrimination in bacterial RNA polymerase involves the residue corresponding to N479 (Svetlov *et al.*, 2004). Polymerase residues lining the NTP site and the active site are highly conserved among species and among nuclear RNA polymerases (Figure 23B), suggesting that the mechanism of NTP binding, selection and incorporation is the same in all these enzymes. The NTP triphosphate moiety could be bound by four conserved basic residues (Figure 23), but is not in a position that allows nucleotide incorporation into RNA. The triphosphate is too far away from the catalytic metal ion A, and from Rpb2 residue D837, which may bind a second catalytic metal ion (Cramer *et al.*, 2001; Sosunov *et al.*, 2003), and is required for full activity of an archaeal RNA polymerase (Werner and Weinzierl, 2002). Together, these results indicate that the observed NTP site is apparently not identical to the insertion site occupied during catalysis, but may correspond to a pre-insertion site.

Indeed, in November 2004, Kornberg and coworkers presented crystallographic models of an elongation complex with both a matched and a mismatched NTP in the +1-site (Westover *et al.*, 2004a). In this model, the “matched”, i.e. templated NTP is about 1 Å closer to the Watson-Crick edge of the template base. The alpha-phosphate group is in a position poised for a metal A-assisted catalysis and incorporation to the RNA 3'-end. These apparent differences might be explained by slightly different experimental conditions: While in this study the non-hydrolysable GMPCPP and an intact RNA 3'-end was present, Kornberg *et al.* made use of a 3'-deoxy-RNA and an unmodified substrate base. The resolution is similar in both studies.

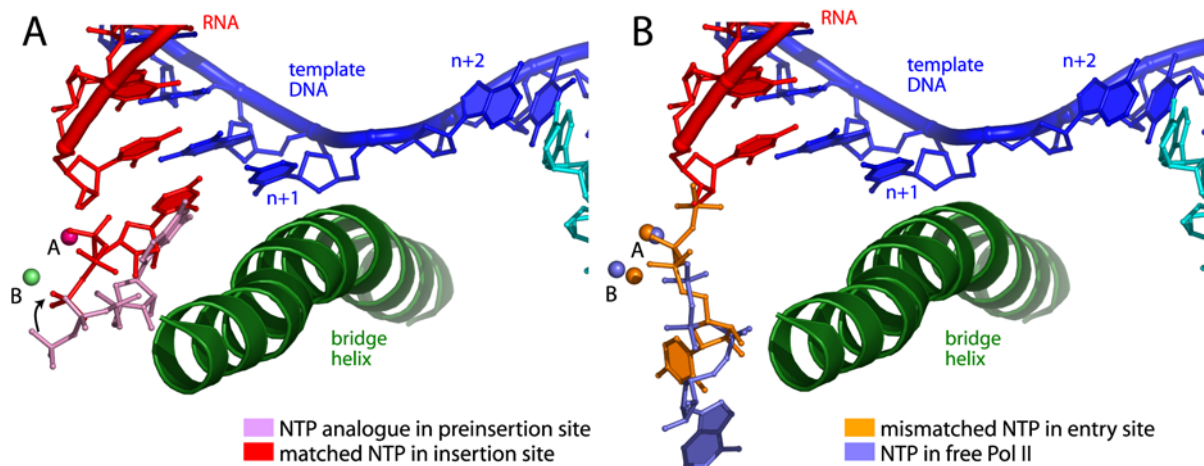


Figure 24: NTP sites in the Pol II active center.

Parts of the DNA and RNA strands and the bridge helix of the complete EC structure are depicted (Kettenberger *et al.*, 2004), together with the two catalytic metal ions A and B (Westover *et al.*, 2004a). The bridge helix and the aspartate loop (residues 478-489 of the largest subunit) were used for superposition of the complete EC and core EC structures. Note that only the nucleic acid strands in the complete EC are depicted.

(A) Two binding sites for templated (“matched”) NTPs. NTPs in the insertion/A-site (Westover *et al.*, 2004a; PDB 1R9S), and in the putative preinsertion site (Kettenberger *et al.*, 2004; PDB 1Y77) are shown.

(B) Two inverted NTP sites (Westover *et al.*, 2004a). A mismatched nucleotide bound to core EC is shown in light blue (PDB 1R9T). A UTP bound to core Pol II in the absence of DNA and RNA is shown in orange (PDB 1TWF).

When a “mismatched” NTP is bound to an elongation complex, Kornberg and coworkers observed an unexpected rotation of the NTP around metal B (Westover *et al.*, 2004a). The NTP orientation is similar to an NTP bound to Pol II in the absence of nucleic acids (Figure 24). Considering these structures in the light of biochemical data, a possible pathway of NTP incorporation emerges. NTPs may enter through the funnel and pore, and may bind to the entry site (Figure 24B). This interaction may help overcome restricted NTP diffusion in the pore as inferred by simulation (Batada *et al.*, 2004). The NTP may then rotate into the preinsertion site (Figure 24B) for sampling of correct pairing with the template base, and for discrimination against dNTPs (Figure 23A). Only correctly paired NTPs may dwell long enough in the insertion site, allowing catalytic phosphodiester bond formation. In an alternative model, NTPs pair with downstream template bases and enter from the Pol II cleft (Epshtein *et al.*, 2002; Nedialkov *et al.*, 2003). Indeed, the template base at position +2 is unpaired in the complete EC structure, and could bind an incoming NTP, but this requires conformational changes in fork loop 2. Position +2 corresponds to the point of strand separation. Hence, further line-up of template-paired

NTPs before the active site, including positions +3 and +4, disagrees with the structural observations.

The active site in the free Pol II core structure revealed a metal ion A, bound by aspartates D481, D483 and D485 of the largest Pol II subunit, and a metal ion B, located 5-6 Å from metal A, near residues D836 and E837 of the second largest subunit (Cramer *et al.*, 2001). Metal B was mobile and thought to enter with the NTP substrate (Cramer *et al.*, 2001). The new core EC study supports and extends this view. Two metal ions are bound by the NTP in the insertion site (Westover *et al.*, 2004a). Whereas metal A remains at the previously detected location, metal B binds closer to metal A, interacting with D481 and D483 of the largest subunit, with D836 of the second largest subunit, and with two NTP phosphates. Details of the mechanisms of catalysis or NTP incorporation are however not revealed, since the core EC and complete EC structures with matched NTPs have been determined at 4.2 Å and 4.5 Å resolution, respectively (Westover *et al.*, 2004a) and (Figure 24). At higher resolution, structures of the free Pol II core with bound NTP and of a bacterial RNA polymerase bound to ppGpp (guanosine 3'-diphosphate 5'-diphosphate) revealed two metal ions at the same location, supporting the two metal ion mechanism (Artsimovitch *et al.*, 2004).

The overall location of the NTP with respect to the active site, DNA and RNA, is similar to recent substrate complexes of T7 RNA polymerase (Landick, 2004; Temiakov *et al.*, 2004; Yin and Steitz, 2004). Cocrystallisation of a T7 RNA polymerase elongation complex with the NTP analog AMPCPP can result in NTP binding to a pre-insertion site, and in an open conformation at the active site (Temiakov *et al.*, 2004), as observed here for a Pol II elongation complex cocrystallized with GMPCPP. Analysis of the pre-insertion site in T7 RNA polymerase suggested that the NTP triphosphate moiety is bound by three basic residues in bacterial RNA polymerase (Temiakov *et al.*, 2004), which correspond exactly to the basic residues found near the difference density in the Pol II complex.

16.12 Comparison with a single-subunit RNA polymerase

Comparison of the new Pol II EC structures with EC structures of the unrelated single-subunit RNA polymerase from phage T7 (Temiakov *et al.*, 2004; Yin and Steitz, 2004) shows that two catalytic metal ions bind in the same relative orientation, supporting a general two-metal ion mechanism for all nucleic acid polymerases (Steitz, 1998). The comparison with T7 elongation

complex structures further reveals a similar overall arrangement of DNA, RNA and NTP substrate, but also clear differences. The length of the DNA-RNA hybrid is the same in both elongation complexes. There is also a similar 90° bend between axes of the incoming DNA and the exiting DNA-RNA hybrid (Cramer, 2002a), but the two duplex axes intersect in the Pol II EC, whereas they are offset in the T7 EC (Figure 25). Further, there is a similar 90° twist between template bases at positions +2 and +1. The distance between the insertion and the pre-insertion sites is small in the Pol II, but much larger in the T7 elongation complexes. The entry site in Pol II (Figure 24B) does apparently not exist in the T7 enzyme. The Pol II elongation complex structures did not reveal conformation changes upon NTP binding, but the T7 structures showed structural changes in the active centre that accompany NTP incorporation and nucleic acid translocation (Landick, 2004; Steitz, 2004).

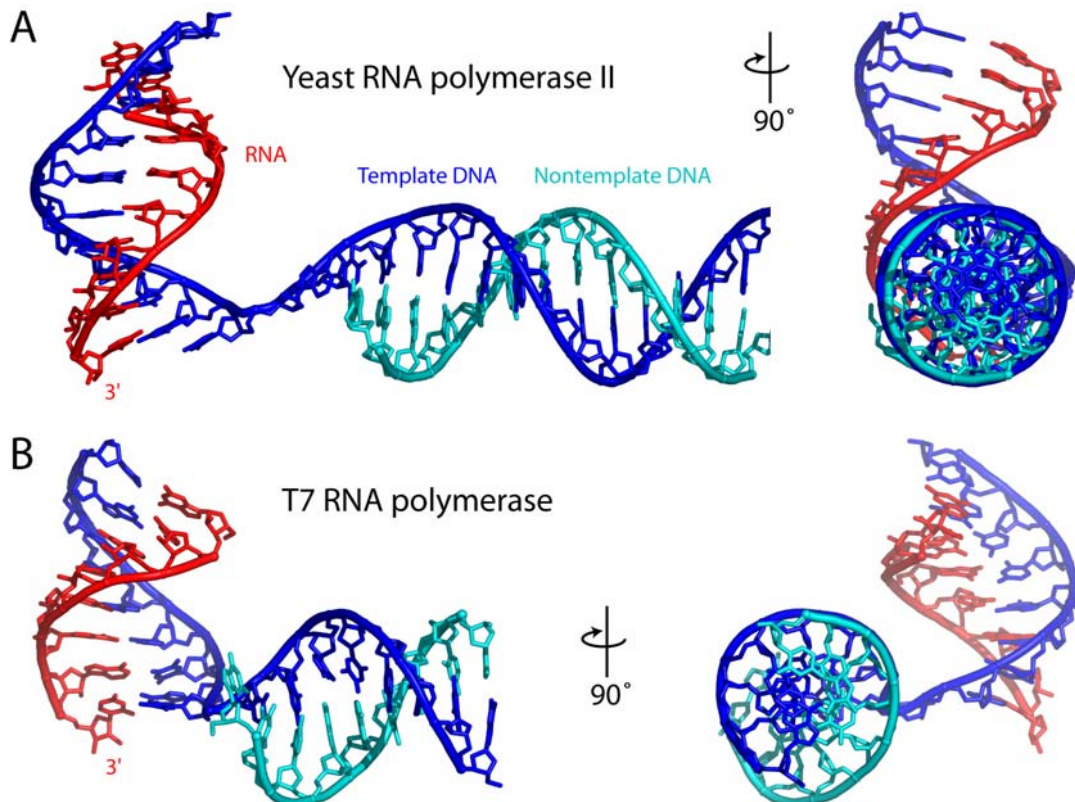


Figure 25: Nucleic acid conformation in elongation complexes of Pol II and T7 RNA polymerase.

(A) Complete yeast Pol II elongation complex (PDB 1Y1W, this study). The two views are related by a 90° turn about a vertical axis.

(B) Bacteriophage T7 RNA polymerase elongation complex (Yin and Steitz, 2004; PDB 1S0V).

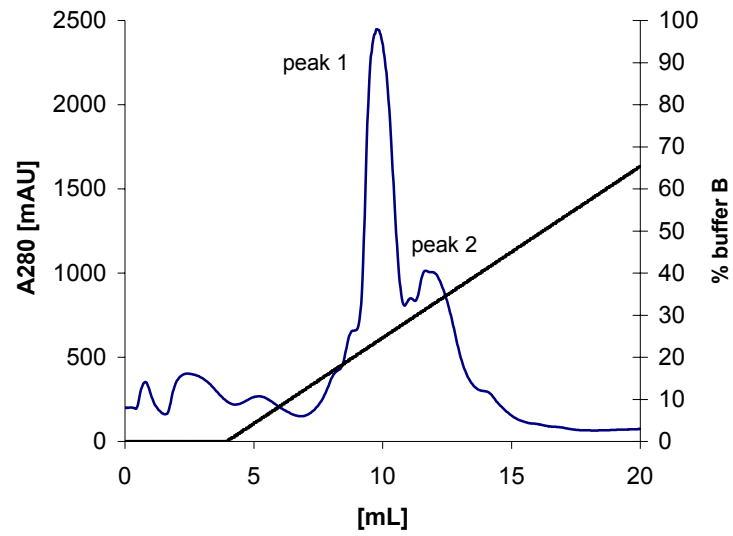
17 THE RNA POLYMERASE II-TFIIS COMPLEX

17.1 Purification of TFIIS

The TFIIS variant comprising domains II and III (residues 131-309, Figure 29A), as well as the double mutant D290A/E291A could be readily overexpressed in *E. coli* and purified via an N-terminal hexahistidine tag, essentially as described (Awrey *et al.*, 1997). Yields ranged around 20 mg of purified TFIIS variant per litre of culture. Purification by a nickel-NTA affinity column followed by a cation exchange column (Mono-S) resulted in two peaks which were both identified as the same species by SDS gel electrophoresis, mass spectroscopy and circular dichroism (Figure 26 and chapter 26).

Overexpression in selenomethionine-containing minimal medium and an analogous purification strategy resulted in a protein which was indistinguishable from the underivatised protein, except a molecular weight difference of 47 Da (ESI-MS). This difference is consistent with the incorporation of a single selenomethionine residue. Unexpectedly, despite its modular nature, the TFIIS variant turns out to be stable towards degradation for weeks at 4 °C.

A



B

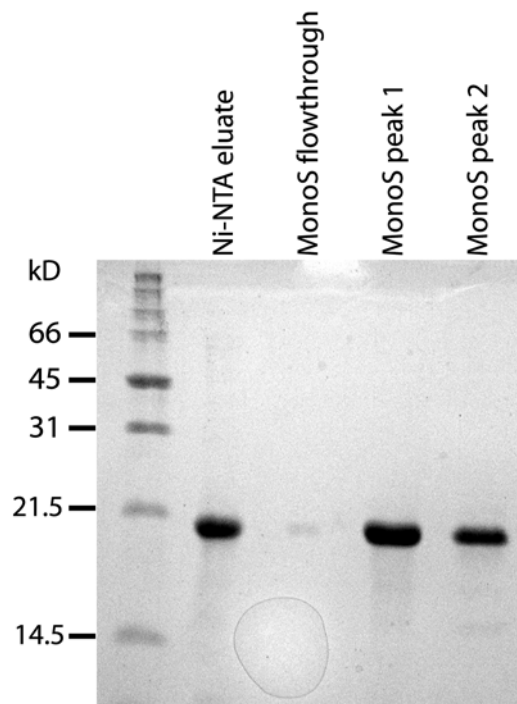


Figure 26: TFIIS purification. (A) Chromatogram of the Mono-S column. TFIIS elutes in two peaks (peaks A and B) in the salt gradient. (B) The Coomassie-stained SDS-polyacrylamide gel shows peak fractions of the two-step purification protocol described in chapter 26.

17.2 Structural analysis

The 12-subunit yeast Pol II was prepared as described in chapter 28. Harvested crystals were incubated with recombinant TFIIS (see chapter 17.1). The very large solvent channels of the crystals allowed TFIIS entry and binding to its specific site on Pol II, which is not obstructed by crystal contacts (Figure 39).

Table 10: X-ray structural analysis of the RNA polymerase II–TFIIS complex

Crystal	Pol II+TFIIS	Pol II+TFIIS SeMet	Pol II+DNA+RNA+TFIIS ⁵
	Data collection ¹		
Unit cell dimensions (Å) ²	218.9 x 395.3 x 281.0	219.7 x 394.9 x 282.3	220.2, 395.7, 282.1
Wavelength (Å)	0.9790	0.9795	0.9919
Resolution range (Å) ³	50-3.8 (3.94-3.8)	50-6.5 (6.73-6.5)	50-4.0 (4.14-4.0)
Unique reflections ³	115 508 (9618)	24 571 (2407)	103460 (10312)
Completeness (%) ³	96.2 (81.0)	100 (100)	98.7 (99.2)
Redundancy ³	2.9 (1.2)	6.1 (6.1)	4.3 (4.3)
Mosaicity (°)	0.21-0.53 ⁴	0.40	0.29-0.37 g
R _{sym} (%) ³	8.9 (35.5)	7.1 (32.6)	7.0 (27.9)
I/σ(I) ³	9.3 (1.9)	11.6 (4.6)	20.3 (5.4)
Peaks in anomalous Fourier map	8 Zn in Pol II (10.9-5.1 σ) 1 Zn in TFIIS (9.1σ)	8 Zn in Pol II (11.6-7.4 σ) 1 Zn in TFIIS (4.5σ) 1 Se in TFIIS (7.3σ)	—
Refinement			
Number of residues	4001		
RMSD bonds (Å)	0.011		
RMSD angles (°)	1.62		
R _{cryst} (%)	28.2		
R _{free} (%)	29.4		28.1 (after rigid body refinement)
PDB-ID	1Y1V		1Y1Y

¹Diffraction data were collected at beamline X06SA at the Swiss Light Source.

²Crystals belong to the space group C222₁.

³Values in parentheses correspond to the highest resolution shell.

⁴Due to radiation damage, the mosaicity increased during data collection, and was refined in segments of 10°.

⁵The complex crystals were soaked with the inactive TFIIS variant D290A/E291A

⁶ Due to radiation damage, mosaicity increased during data collection, and was refined in segments of five frames.

It is intriguing that the crystal lattice can accommodate extensive structural changes induced by TFIIIS (Figure 32). The resulting 13-polypeptide asymmetric complex has a molecular weight of 536 kDa. Diffraction data to 3.8 Å resolution could be obtained at the synchrotron SLS and data set statistics are shown in Table 10.

Initial electron density maps were phased with the core Pol II structure and were improved by solvent flipping (see chapter 30). These maps showed positive difference density on the Pol II surface, which could to a large extent be accounted for with the rigid cores of the NMR structures of TFIIIS domains II and III (Morin *et al.*, 1996; Qian *et al.*, 1993a), (Figure 28A, Figure 29A). The positioning of the TFIIIS domains was confirmed by locating a selenomethionine residue in domain II and the zinc ion in domain III by means of anomalous signals (Figure 28B, Table 9). The first electron density map also revealed a long linker between the two TFIIIS domains, and extensive structural changes in Pol II.

17.3 Induced folding of TFIIS

Structure analysis and model building was assisted by the availability of NMR structures of domains II and III (Morin *et al.*, 1996; Qian *et al.*, 1993a; Figure 27). A striking feature of these structures is the apparent flexibility, particularly of the inter-domain linker, which flanks a rigid core of domain II. However, in the context of Pol II, most of the flexible parts of TFIIS undergo induced folding and adopt a well-defined conformation (Figure 28). Induced folding upon Pol II interaction could indeed be a recurring motif in many transcription factors (Chen and Hahn, 2003; Chung *et al.*, 2003; Hahn, 2004). This would explain the prevalent difficulties in crystallising isolated factors such as TFIIF or TFIIB.

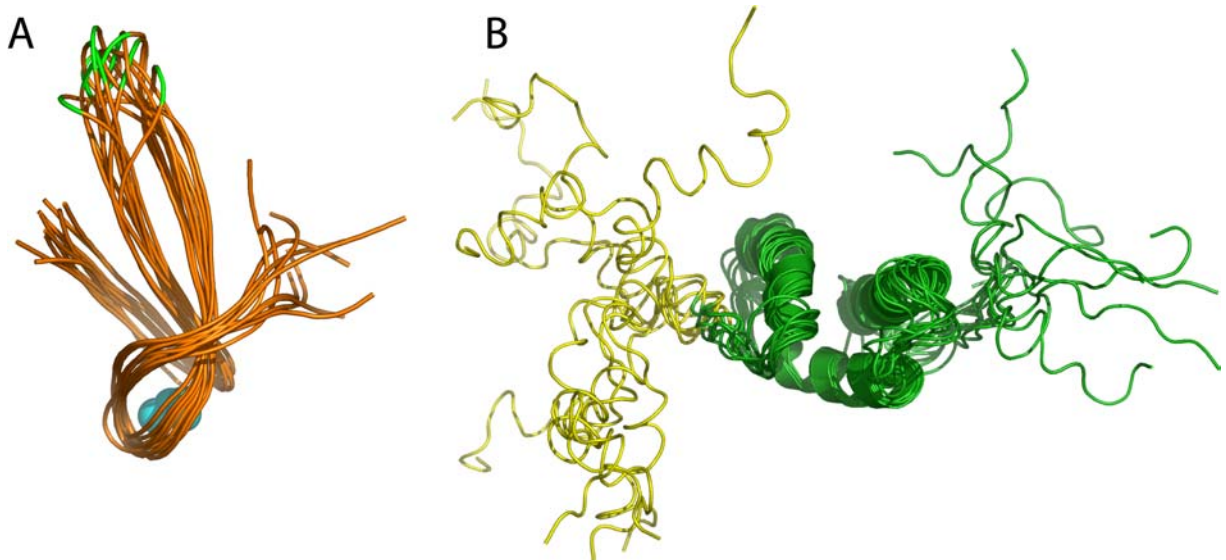


Figure 27: NMR ensembles of isolated TFIIS domains.

(A) Ribbon model of domain III according to PDB-ID 1ENW (Qian *et al.*, 1993b). The acidic hairpin residues D290 and E291 are coloured in green. Magenta spheres represent zinc ions.

(B) Domain II (green) and part of the domain II-III inter-domain linker (yellow) according to PDB-ID 1TFI (Morin *et al.*, 1996).

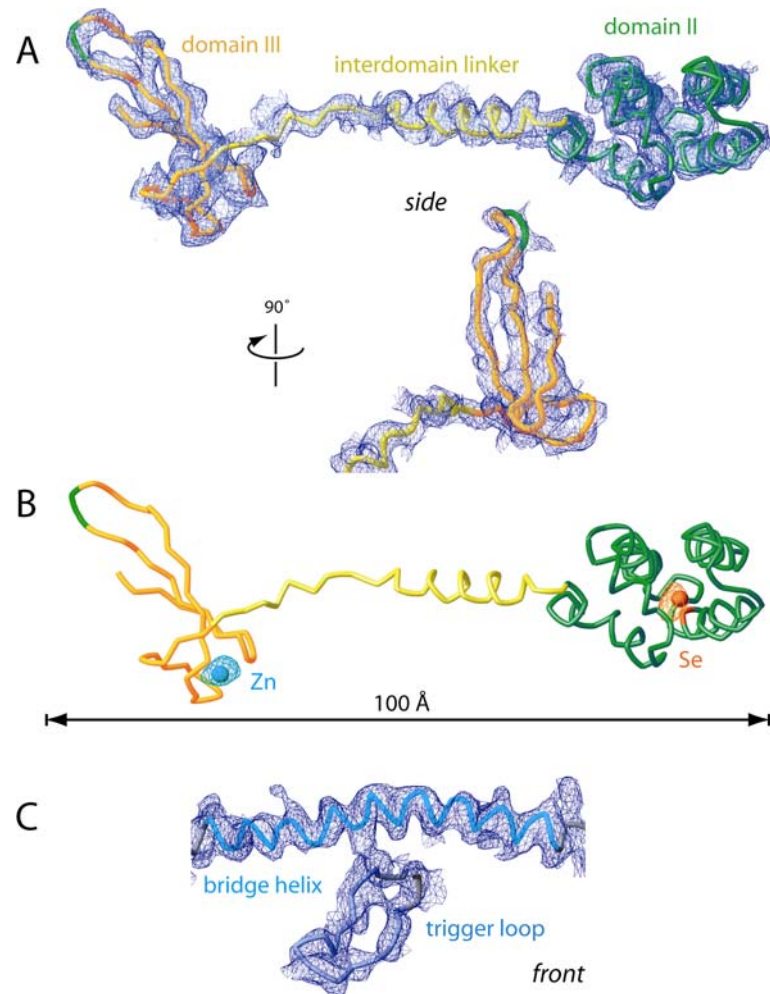


Figure 28: Structural analysis of the Pol II-TFIIS complex.

(A) Electron density omit map. The map is contoured at 1.0σ and was calculated after solvent flipping using phases from the positioned atomic model of the 10-subunit Pol II. Omitted from map calculation were the TFIIS model, the Rpb4/7 complex, the bridge helix, and newly ordered elements of Pol II such as the trigger loop. TFIIS is depicted as a ribbon model. TFIIS domain II, the inter-domain linker and domain III are coloured in green, yellow, and orange, respectively, and this colour code is used throughout. The light green ribbon fragment in domain III corresponds to the acidic hairpin residues D290 and E291.

(B) Anomalous difference Fourier maps were calculated from anomalous differences in native (cyan mesh) and TFIIS-SeMet (orange mesh) data, and were contoured at 2.5σ and 3.0σ , respectively. Anomalous peaks in the anomalous coincide with the position of a zinc ion in domain III (cyan sphere) and a selenium atom in SeMet 182 of domain II (orange sphere).

(C) Electron density omit map described in (A). Depicted are the Pol II bridge helix and trigger loop, which were omitted from map calculation. Figures were prepared with RIBBONS (Carson, 1997).

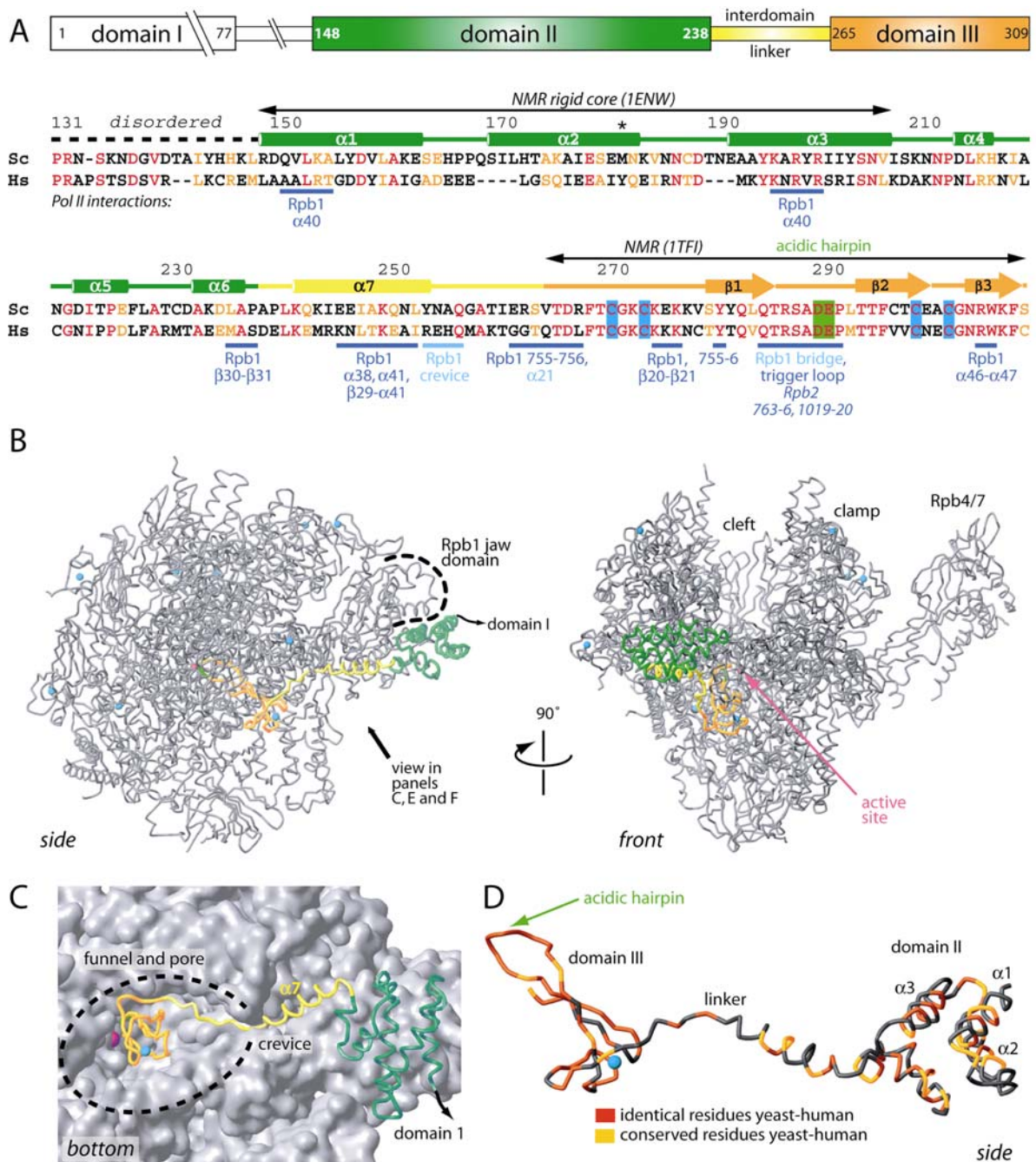


Figure 29: Architecture of the Pol II–TFIIIS complex.

(A) Domain organisation and secondary structure of TFIIIS. The diagram shows TFIIIS domain borders. Secondary structure elements (cylinders, α -helices; arrows, β -strands) are indicated above an alignment of TFIIIS sequences from the yeast *Saccharomyces cerevisiae* (Sc) and human (Hs). Identical and conserved residues are printed in red and light orange, respectively. The colour code for TFIIIS domains is the same as described in the legend of Figure 28. Black double-headed arrows span regions for which available NMR structures (PDB accession codes in parentheses) were placed in the experimental electron density map. Residue M182 in domain II, which was observed in the anomalous Fourier map shown in Figure 28B, is indicated with an asterisk. The four cysteine residues that co-ordinate the zinc ion in domain III are highlighted in cyan, and the acidic hairpin residues D290 and E291 in light green. Regions in TFIIIS that are involved in Pol II interactions are underlined with blue bars. TFIIIS-interacting structural elements in Pol II subunits Rpb1 and Rpb2 are indicated below the blue bars.

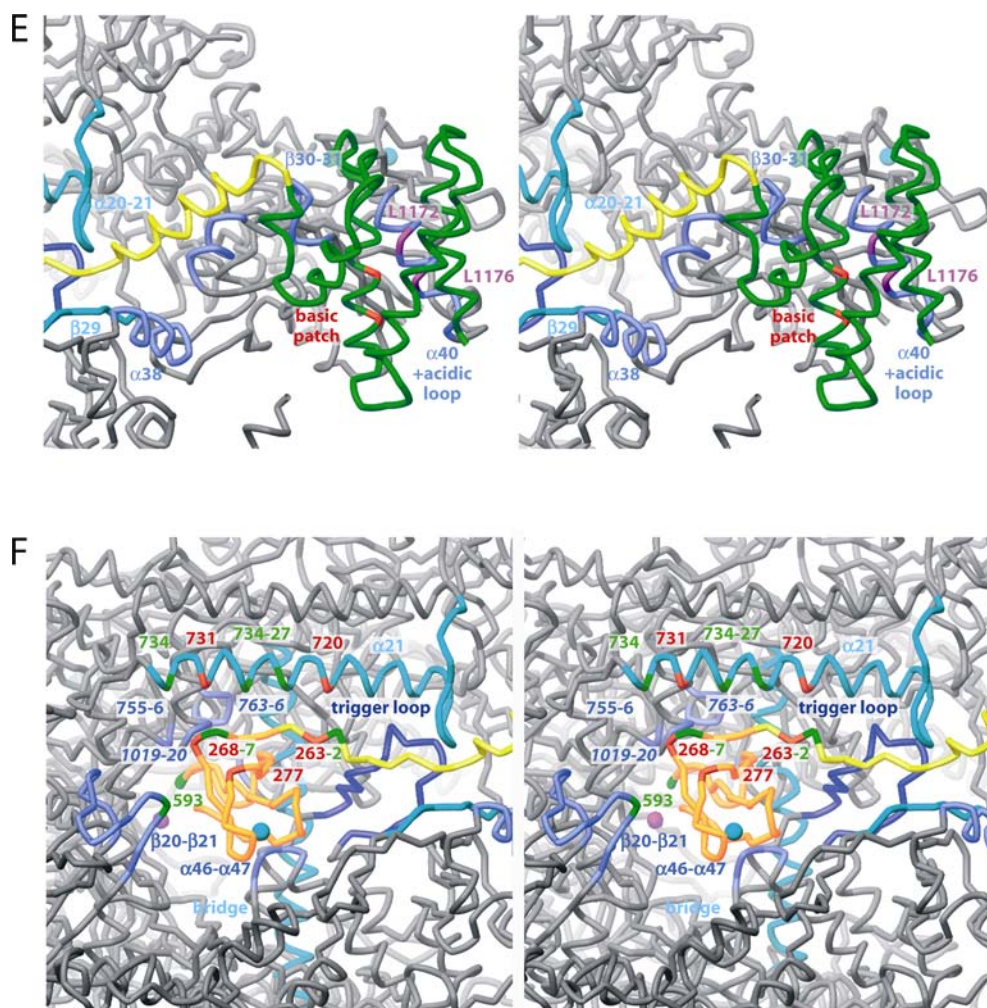


Figure 29 (continued)

(B) Ribbon diagram of the Pol II–TFIIS complex backbone model. The 12 subunits of Pol II are coloured in grey. A pink sphere marks the location of the active site metal ion A (Cramer *et al.*, 2000; Cramer *et al.*, 2001). Eight structural zinc ions in Pol II and one zinc ion in TFIIS are depicted as cyan spheres. The two views are related by a 90° rotation around a vertical axis.

(C) Binding of TFIIS to the crevice, funnel and pore. TFIIS is shown as a ribbon model on the molecular surface of Pol II. The view is from the bottom face, as indicated in panel (B).

(D) Conservation of TFIIS viewed from the side as in panel (B). Residues that are identical or conserved between yeast and human TFIIS are in red and orange, respectively.

(E-F) Details of polymerase–TFIIS interaction. Shown are two stereo views of the part of the model shown in (C), with the Pol II surface removed. Regions of Pol II that interact with TFIIS are labelled with plain and italic lettering for subunits Rpb1 and Rpb2, respectively.

Pol II regions that had shifted were omitted from electron density calculation and were manually docked to the resulting map. After rigid body refinement, the quality of the map was very high, and large protein side chains were generally visible. Newly ordered Pol II regions, and TFIIS regions that were not accounted for by the NMR structures, were modelled as C_{α} -backbones. Revision of molecular replacement with an improved model of complete Pol II (Armache *et al.*,

2003), followed by positional and B-factor refinement further improved the electron density maps. Into these maps, side-chains of TFIIIS residues 240-309 could be built and refined. Refinement of all Pol II residues and residues 240-309 of TFIIIS converged at a final free R-factor of 29.4 %.

17.4 TFIIIS extends from a polymerase jaw to the active site

The structure shows that TFIIIS extends along the Pol II surface, spanning a distance of 100 Å (Figure 28B, Figure 29). Based on visual inspection of the structure, the domain borders of TFIIIS were re-defined (Figure 29A). To describe interactions of TFIIIS with Pol II, the nomenclature of Pol II domains and secondary structure elements introduced previously (Cramer *et al.*, 2001) is used throughout. TFIIIS domain II docks to the exposed Rpb1 jaw domain of Pol II. The TFIIIS inter-domain linker extends from domain II along the Pol II surface into the “funnel” (Figure 29B, C). Domain III inserts into the Pol II “pore”, and approaches the polymerase active site from the bottom face of the enzyme as predicted (Cramer *et al.*, 2000). The binding sites for domain III and α -amanitin overlap, explaining why this toxin interferes with TFIIIS activity, but not with the intrinsic nuclease activity of Pol II (Izban and Luse, 1992b; Rudd and Luse, 1996; Weilbaecher *et al.*, 2003).

TFIIIS domain II comprises a bundle of three helices (α 1- α 3, Figure 29A, D), which had been observed for the isolated domain by NMR (Morin *et al.*, 1996), and three short helices, which form upon Pol II interaction (α 4- α 6). TFIIIS helices α 1 and α 3 pack against helix α 40 in the Rpb1 jaw, to bury leucines L1172 and L1176 in helix α 40 (Figure 29E). An acidic loop following Rpb1 helix α 40 interacts with a basic patch on TFIIIS helix α 3, which includes residues required for TFIIIS–Pol II interaction (Awrey *et al.*, 1998). In addition, TFIIIS helix α 6 contacts Rpb1 loop β 30- β 31 (Figure 29E). Consistently, mutation of residue E1230 in this loop weakens Pol II binding to TFIIIS (Wu *et al.*, 1996). The location of TFIIIS domain II at the Rpb1/9 jaw, next to the point of DNA entry to the cleft, may be relevant for interaction of TFIIIS with the chromatin remodelling complex Swi/Snf, revealed in a genetic screen (Davie and Kane, 2000). The N-terminal domain I of TFIIIS stays on the outer surface of Pol II, where it is potentially available for interaction with other components of the transcription machinery. Domain I was shown to engage in specific interactions with Spt8, a part of the SAGA co-activator, and with Med13, a subunit of the Mediator kinase module (Shimasaki and Kane, 2000).

17.5 TFIIIS opens a crevice in the funnel

The TFIIIS inter-domain linker is unstructured in free TFIIIS (Olmsted *et al.*, 1998). Upon Pol II binding, the linker forms an α -helix ($\alpha 7$), which runs along the Pol II bottom face, near loop $\beta 29$ - $\alpha 41$ and helix $\alpha 38$ in Rpb1 (Figure 29E). At the end of helix $\alpha 7$, the linker passes through a narrow crevice into the funnel (Figure 29C). The crevice is lined by loop $\alpha 20$ - $\alpha 21$ in the Rpb1 funnel domain on one side, and by strand $\beta 32$ in the Rpb1 cleft domain on the other side (Figure 29E, F). The crevice is closed in previous Pol II structures, with its two sides contacting each other, and opening of the crevice is partly responsible for major structural changes in Pol II (see chapter 17.12). The induced folding and polymerase interactions of the TFIIIS linker are important for its function: Linker residues contribute to TFIIIS activity (Awrey *et al.*, 1998), and confer species-specificity to the Pol II–TFIIIS interaction (Shimasaki and Kane, 2000) and isolated domains II and III are not functional (Awrey *et al.*, 1998). Since the attachment of TFIIIS on Pol II is accomplished by domain II, an important function of the linker is the correct positioning of domain III in the funnel and pore. Accordingly, mutations that change the linker length abolish TFIIIS activity (Awrey *et al.*, 1998).

17.6 TFIIIS domain III inserts into the pore

TFIIIS domain III is bound with its zinc-binding base at the entrance to the pore on the bottom face of Pol II (Figure 29B, C, F). The thin β -hairpin of domain III extends along one side of the pore, and approaches the active site. Although dispensable for Pol II binding, domain III is ordered in the electron density due to many contacts with Pol II (Figure 29A, F). The base of domain III makes a hydrophobic contact to Rpb1 residues 755-756 (Figure 29F). Several salt bridges can be formed between charged residues at the base of domain III and helix $\alpha 21$ and loop $\beta 20$ - $\beta 21$ of Rpb1 (Figure 29F). The β -hairpin of domain III is highly flexible in free TFIIIS (Qian *et al.*, 1993b), but is fixed here by contacts with polymerase regions that line the pore, including Rpb1 residues 823-830 in the bridge helix and residues 1078-1080 in the trigger loop, as well as several Rpb2 residues (Figure 29A, F). The high conservation of the pore-lining regions concurs with the high conservation of TFIIIS domain III and the Pol II-interacting residues of domain II.

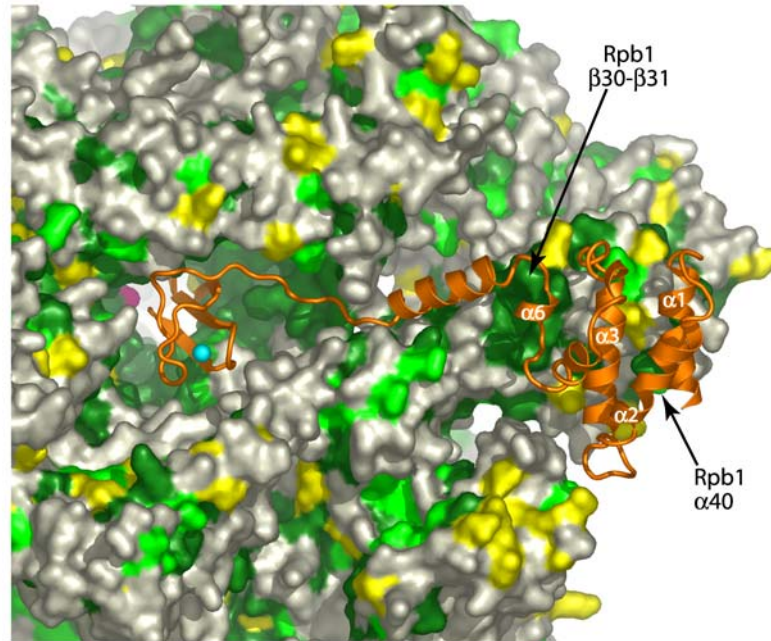


Figure 30: Inter-species conservation of the TFIIIS interaction surface. The view is similar to the view in Figure 29C. TFIIIS is shown as orange ribbons on a molecular surface of Pol II. Pol II residues which are conserved between *S. cerevisiae*, *S. pombe*, *C. elegans*, *D. melanogaster* and *H. sapiens* are coloured in dark green, light green and yellow, corresponding to 100, 80, and 60 % conservation, respectively. Major binding sites (Awrey *et al.*, 1998) on Rpb1 are indicated with arrows.

17.7 The TFIIIS acidic hairpin complements the polymerase active site

TFIIIS domain III reaches the Pol II active site with the highly conserved loop of the protruding β -hairpin (Figure 31). Two invariant acidic residues in this loop, D290 and E291, are in close proximity of the Pol II catalytic metal ion A. The two acidic hairpin residues are essential for TFIIIS activity, and even conservative mutations of these residues render TFIIIS inactive (Jeon *et al.*, 1994). The domain III hairpin therefore complements the polymerase active site with acidic groups that are essential for TFIIIS function.

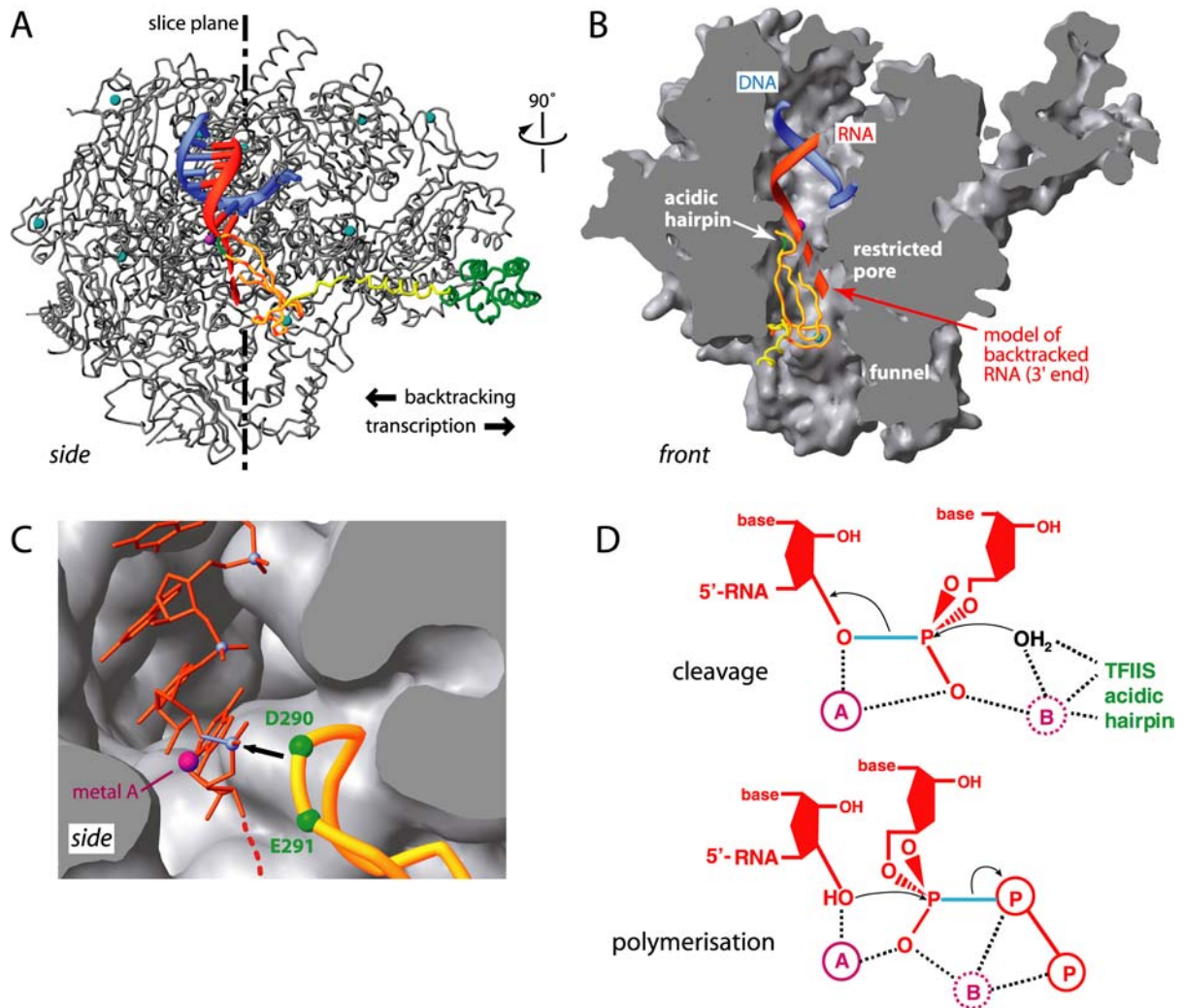


Figure 31: Model of a Pol II-TFIIS-nucleic acid complex.

(A) Side view of the Pol II-TFIIS-nucleic acid complex model. The DNA template strand (blue) and the RNA transcript (red) were placed onto the ribbon model of Fig. 2B according to their location in the elongation complex structure (Gnatt *et al.*, 2001). The Rpb2 protrusion, fork, and external domains were omitted for clarity. The presumed location of backtracked RNA is indicated as a dashed red ribbon. The arrows indicate movement of Pol II relative to the nucleic acids.

(B) Cut-away view of the model in (A) from the front. TFIIS and nucleic acids are shown as ribbon models on the molecular surface of Pol II, which is cut along the vertical slice plane indicated in (A). The presumed path of backtracked RNA through the restricted pore is drawn as a dashed ribbon. This portion of RNA would be cut during TFIIS-induced RNA cleavage.

(C) Proximity of the TFIIS acidic hairpin to the potential scissile RNA phosphodiester bond. The view is as in (A). RNA was placed according to the location in the Pol II elongation complex structure (Gnatt *et al.*, 2001), and is shown as a stick model with phosphorous atoms highlighted as blue spheres. The black arrow indicates the direction of a possible S_N2 -type nucleophilic in-line attack of the scissile bond (blue).

(D) Reaction mechanism of cleavage and polymerisation. During TFIIS-mediated RNA cleavage, a coordinated nucleophilic water molecule attacks the P atom of the phosphodiester bond, leading to a break of the scissile bond (blue). During RNA polymerisation, the 3'-OH group of the RNA terminus acts as the nucleophile to attack the P_α atom of a substrate NTP. By breaking the scissile bond (blue), pyrophosphate is liberated. A and B indicates the two catalytic metal ions; encircled letters "P" symbolise phosphate groups.

Thus, Pol II apparently contains a single tuneable active site for both RNA polymerisation and cleavage, instead of two catalytic sites with distinct locations. It had been suggested previously that the active sites for RNA polymerisation and cleavage are close together or even identical (Powell *et al.*, 1996; Rudd *et al.*, 1994; Wang and Hawley, 1993).

17.8 RNA cleavage

The approximate location of the TFIIS hairpin loop with respect to substrate RNA is revealed by superposition of the presented model with the previous Pol II elongation complex structure (Figure 31). The superposition shows that TFIIS domain III comes in close proximity of the RNA, explaining why the RNA 3'-end can be crosslinked to TFIIS (Powell *et al.*, 1996). The two acidic residues in the TFIIS hairpin approach the RNA from the side of the sugar-phosphate backbone, as required for catalytic RNA cleavage. The RNA phosphodiester bond that is potentially cleaved is observed in the elongation complex structure, since this complex is apparently trapped after nucleotide incorporation, but before translocation, adopting the pre-translocation state (Gnatt *et al.*, 2001). The potential scissile phosphodiester bond (Figure 31C, D) connects the 3'-oxygen in the ribose of the penultimate nucleotide with the α -phosphorous of the terminal RNA nucleotide, since RNA cleavage leaves a 3'-OH group on the RNA, while liberating 5'-phosphonucleotides (Izban and Luse, 1993b).

The exact location of chemical groups in the active site cannot be determined at the given resolution. However, the location of the TFIIS hairpin and metal ion A with respect to the potential scissile RNA bond in the model suggests that the mechanism of RNA cleavage could resemble that of DNA cleavage by the Klenow DNA polymerase. In the Klenow exonuclease active site, two catalytic metal ions are co-ordinated by acidic residues (Beese and Steitz, 1991; Joyce and Steitz, 1994). A first metal ion binds the 3' oxygen of the leaving ribose. A second metal ion and one of the acidic residues position a water molecule for an S_N2 -type nucleophilic attack of the phosphorous atom from the side opposite the 3' oxygen, in-line with the scissile phosphodiester bond. In Pol II, metal A coordinates the RNA 3'-oxygen. One of the TFIIS acidic hairpin residues could place a metal ion B nearby, and could help position a water molecule for hydrolytic RNA cleavage. Such a mechanism would be consistent with the requirement for divalent metal ions in RNA cleavage by the Pol II-TFIIS complex (Izban and Luse, 1992b;

Weilbaeher *et al.*, 2003). It is also consistent with evidence for a nucleophilic water molecule, based on dramatic increase in cleavage activity at high pH (Weilbaeher *et al.*, 2003).

17.9 Switching between polymerisation and cleavage

A two-metal-ion mechanism could thus underlie both RNA cleavage (see chapter 17.8) and RNA polymerisation by Pol II (Cramer *et al.*, 2001). Both polymerisation and cleavage require metal A, that is persistently bound by the Rpb1 aspartate loop. Switching from polymerisation to cleavage may however involve differential positioning of a mobile metal ion B. For polymerisation, metal B could be recruited by the substrate nucleoside triphosphate (Cramer *et al.*, 2001). For cleavage in the presence of TFIIS, the TFIIS acidic hairpin could contribute to metal B coordination. Cleavage in the absence of TFIIS, however, would likely require additional coordination partners for metal B. Based on analysis of the nuclease activity in bacterial RNA polymerase, Goldberg and co-workers suggested that during cleavage, metal B could be coordinated by the phosphates of an unpaired nucleotide bound to a site in the pore (Sosunov *et al.*, 2003). They further suggest a unified two-metal-ion mechanism for RNA synthesis and cleavage that is consistent with implications of the present Pol II–TFIIS model. The unique feature of Pol II is its capability to switch between two distinct catalytic activities at a single active site. This is in contrast to a mere activation of an enzyme by complementation of an active site by an external factor, which was observed e.g. in the Ras-RasGAP complex (Scheffzek *et al.*, 1997).

17.10 Proofreading

The available Pol II structures and many biochemical observations are consistent with the following model for mRNA proofreading: Incorporation of the correct nucleotide drives rapid forward translocation (Nedialkov *et al.*, 2003). However, upon incorporation of an incorrect nucleotide, forward translocation is disfavoured, opening a time window for hydrolytic RNA cleavage and removal of the misincorporated nucleotide. Alternatively, misincorporation can trigger backtracking by one nucleotide and subsequent cleavage of a dinucleotide. Cleavage of mononucleotides (from the pretranslocation state) and of dinucleotides (from a backtracked state) both result in a new RNA 3'-end at metal A, from which polymerisation can continue. Proofreading may involve mainly cleavage of mono- and dinucleotides, so that extensive backtracking and RNA extrusion into the pore would not be required. Proofreading reactions can be stimulated by TFIIS *in vitro* (Jeon and Agarwal, 1996; Thomas *et al.*, 1998), but a

contribution of TFIIS to proofreading *in vivo* does not necessarily have to be significant (Shaw *et al.*, 2002).

17.11 Pore restriction and RNA backtracking

The present model shows that TFIIS domain III does not completely block the pore, but restricts it (Figure 31B). The restricted pore leaves enough space for NTP entry, as required for RNA polymerisation in the presence of TFIIS (Horikoshi *et al.*, 1984). Modelling shows that RNA backtracking by more than two nucleotides (relative to the post-translocated state) would result in a clash with the TFIIS hairpin, unless RNA base stacking is given up, and the RNA is redirected and threaded into the restricted pore. Restriction of the pore may thus explain why TFIIS preferably releases dinucleotides (Izban and Luse, 1992b; Izban and Luse, 1993a; Izban and Luse, 1993b).

While transient pausing of Pol II is associated with minimal backtracking (0-2 nucleotides, see Introduction), arrest occurs after more extensive backtracking by about 7-9 nucleotides (Fish and Kane, 2002; Wind and Reines, 2000). The restricted pore is wide enough to accommodate an RNA single strand, and likely corresponds to an RNA-binding site defined previously in Pol II–RNA binary complexes that undergo TFIIS-induced cleavage (Johnson and Chamberlin, 1994). Extensively backtracked RNA and TFIIS can thus bind simultaneously to the Pol II pore, as required for rescue of arrested Pol II complexes. Consistently, the backtracked RNA 3'-end in an arrested bacterial RNA polymerase complex was crosslinked to a protein fragment that lines the restricted pore (Markovtsov *et al.*, 1996).

17.12 TFIIS remodels the polymerase active centre

In addition to a complementation of the Pol II active site, TFIIS induces structural changes in the active centre (Figure 32A). These changes are revealed by a comparison of the Pol II–TFIIS complex with the free 12-subunit Pol II (Armache *et al.*, 2004). Since both models were obtained from the same crystal form, and since TFIIS binding does not induce significant changes in crystal packing or unit cell dimensions, the detected structural changes can be attributed to TFIIS binding.

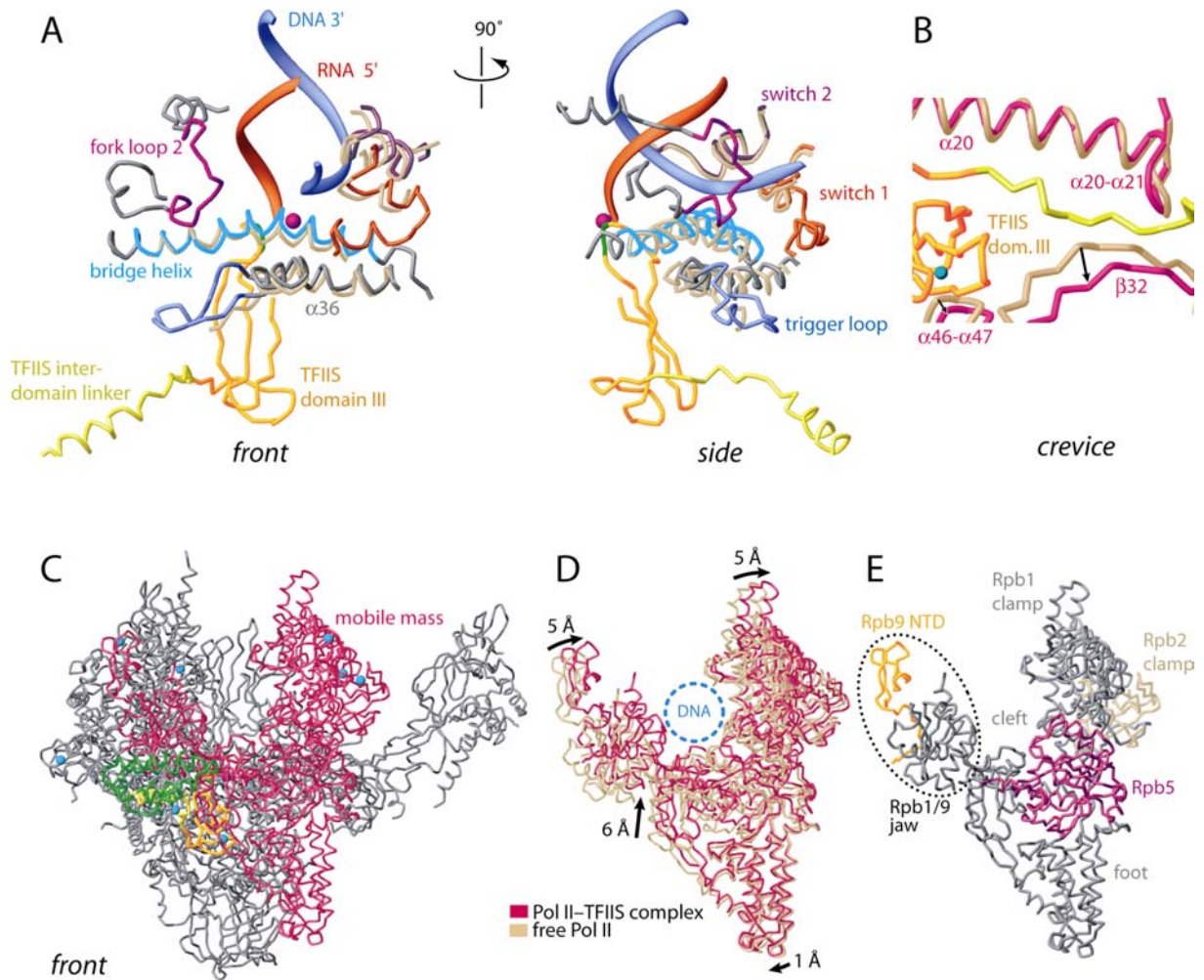


Figure 32: TFIIS-induced structural changes in Pol II

(A) Local remodelling of the Pol II active centre. Structural elements of the active centre in the Pol II–TFIIS complex are shown as ribbons in different colours. The corresponding elements in the free 12-subunit Pol II structure are shown superimposed in beige. In the Pol II–TFIIS complex, the trigger loop (blue) and fork loop 2 (pink) are folded, parts of the bridge helix (cyan) are shifted upwards, and switches 1 and 2 (red and purple, respectively) moved outwards. The DNA template strand (blue) and product RNA have been placed according to the Pol II elongation complex structure (Gnatt *et al.*, 2001). A pink sphere marks the location of metal ion A.

(B) Opening of the Pol II crevice by insertion of the TFIIS linker. Detailed view of the TFIIS linker (yellow) passing through the crevice (magenta). The location of the two crevice-forming elements in free Pol II is shown in beige. The view is from the bottom, as in Figure 29C, E and F.

(C–E) Global repositioning of the Pol II mobile mass. In (C), regions of the model in Figure 29B that are repositioned upon TFIIS binding (mobile mass) are highlighted in magenta. In (D), changes in the mobile mass between free Pol II (beige) and the Pol II–TFIIS complex (magenta) are shown. The models were superimposed based on the unchanged regions in the Pol II core module (Cramer *et al.*, 2001). Arrows indicate the direction and magnitude of movements at outer positions. The location of the incoming DNA duplex during transcription elongation is indicated as a dashed blue circle. In (E), the contents of the mobile mass are shown. Subunits that contribute to the mobile mass are coloured according to the colour code used previously (Armache *et al.*, 2003; Cramer *et al.*, 2000; Cramer *et al.*, 2001). Pol II subunits and domains that contribute to the mobile mass are labelled.

Binding of TFIIS domain III induces folding of Rpb1 residues 1082-1091 (Figure 28C, Figure 29F, Figure 32A), which correspond to the trigger loop in bacterial RNA polymerase (Vassilyev *et al.*, 2002). TFIIS makes potential contacts to residues 1078-1080 preceding the trigger loop. The folded trigger loop closes a previously observed second perforation in the polymerase cleft (pore 2, (Cramer *et al.*, 2000). Structural changes in the trigger loop, and in the preceding helix $\alpha 36$, are propagated to the bridge helix (Figure 32A). TFIIS can form several potential contacts to Rpb1 residues 820-830, resulting in a movement of the C-terminal half of the bridge helix towards the top face of Pol II by 2–3 Å (Figure 28C, Figure 32A). Changes in the bridge helix are further propagated to switch regions 1 and 2, which move slightly outwards, resulting in a widening of the DNA-RNA hybrid binding site (Figure 32A), which may influence the strength of hybrid-polymerase interaction. Since both switches and the C-terminal half of the bridge helix all interact with the DNA template strand in the elongation complex (Gnatt *et al.*, 2001), changes in their position might be responsible for repositioning nucleic acids in the active centre, maybe to facilitate RNA cleavage (see chapter 17.13). Structural changes further include ordering of fork loop 2 (Rpb2 residues 503-508, Figure 32A), which restricts the cleft to a diameter of 15 Å, consistent with the proposal that this loop removes the DNA non-template strand from the template strand before the active site (Cramer *et al.*, 2001; Gnatt *et al.*, 2001).

17.13 TFIIS realigns RNA

In chapter 17.12, remodelling of about one third of Pol II in the context of TFIIS was discussed. This conformational change also affects the active site region and the DNA/RNA binding site. DNA and RNA discussed in Figure 31, Figure 32 and chapter 17.8 represent merely a model, based on previously determined elongation complex structure. Only a genuine Pol II-transcription bubble-RNA complex could disclose the effects of TFIIS on bound DNA and RNA. To this end, crystals (see chapter 16.3) were soaked with a TFIIS variant that is inactive in RNA cleavage stimulation, due to a D290A E291A double mutation in the acidic hairpin (see chapter 26). A dataset was collected to a resolution of 4.0 Å, and for molecular replacement, the refined model of the Pol II-TFIIS complex described in chapter 17.2, was used. The resulting difference electron density in the Pol II-bubble-RNA-TFIIS complex was noisy but allowed modelling of four base pairs of the DNA-RNA hybrid (positions –1 to –4, Figure 33).

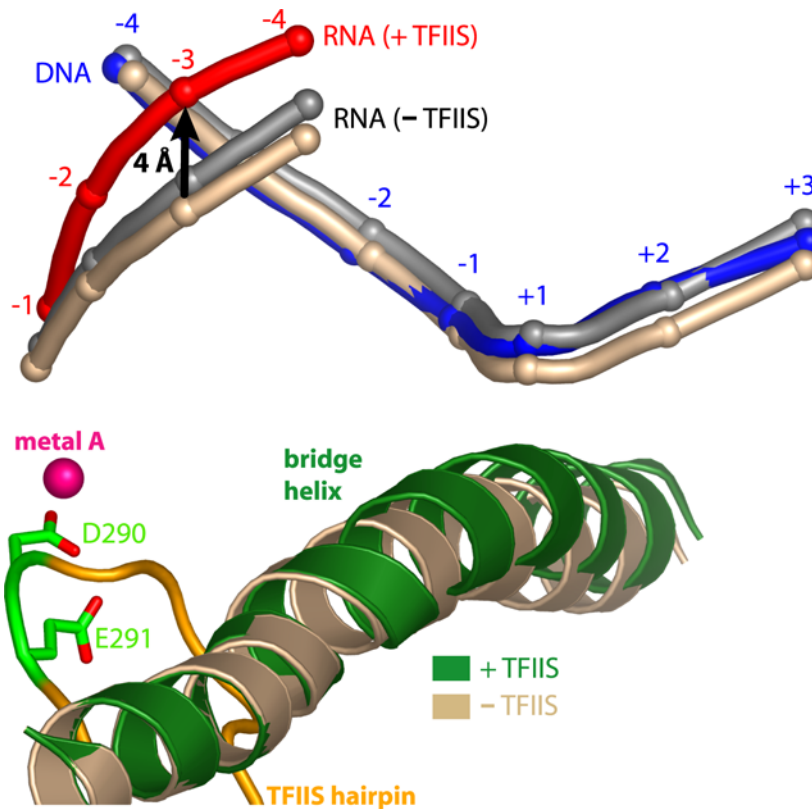


Figure 33: TFIIS-induced RNA realignment

Selected elements in the Pol II active center that move upon TFIIS binding are shown. The bridge helix, DNA, and RNA in the Pol II-bubble-RNA-TFIIS complex are in green, blue and red, respectively. The TFIIS hairpin is in orange with the two acidic functionally essential and invariant residues in green. Nucleic acids in the Pol II-bubble-RNA complex structure after superposition of residues in the active site aspartate loop or in switch 2 are shown in beige and grey, respectively. Switch 2 moves slightly upon TFIIS binding (Kettenberger *et al.*, 2003), explaining the difference in the two superpositions.

TFIIS induces an unexpected shift of the RNA strand, with phosphate groups moving up to 4 Å (Figure 33). This realignment of the RNA in the active site may advantage an optimised in-line attack by a nucleophilic water molecule during RNA cleavage (compare chapter 17.8 and Figure 32). TFIIS does not simply reposition the DNA-RNA hybrid as a rigid body, but its exact conformation is not revealed at the limited resolution of the data. These new structural data suggest alternative explanations for how TFIIS facilitates readthrough. First, the TFIIS-induced repositioning of RNA may result in a different elongation complex conformation that is less prone to pausing. Recent kinetic data are consistent with an alternative TFIIS-induced state of the elongation complex (Zhang and Burton, 2004). Second, the acidic residues in the TFIIS hairpin could assist in NTP binding. The TFIIS hairpin overlaps the NTP site described above,

but small changes in the hairpin conformation may promote metal-mediated stabilising interactions with the NTP triphosphate.

17.14 Functional conformations of Pol II

In addition to a local remodelling of the active centre, TFIIS induces a co-ordinated repositioning of about one third of the polymerase mass (Figure 32C, D). The repositioned mass includes the jaws, the clamp, and the Rpb1 cleft and foot domains, and corresponds essentially to three previously identified mobile polymerase modules (Cramer *et al.*, 2001). The mobile mass is tilted towards the top face of Pol II, by up to 6 Å at the jaws. The repositioning seems to be caused by insertion of TFIIS into the Pol II funnel and pore. In particular, the TFIIS linker opens the crevice, and TFIIS domain III forms a wedge between helix $\alpha 21$ and loop $\alpha 46$ - $\alpha 47$ in the Rpb1 funnel and cleft domains, respectively (Figure 29C, E, F). The mobile polymerase mass is connected to the remainder of the enzyme by the bridge helix, the switches, and the linker between the two Rpb9 domains, which all undergo structural changes to accommodate the repositioning.

A major conformational change in the polymerase upon TFIIS binding was suggested by functional studies of TFIIS and Pol II mutants (Cipres-Palacin and Kane, 1994; Hemming *et al.*, 2000). There is also evidence for a conformational isomerisation of Pol II upon the transition from initiation to elongation, and for distinct conformational states of elongating Pol II (Erie, 2002; Palangat and Landick, 2001). Since the mobile Pol II mass surrounds nucleic acids (Figure 32D), its repositioning could influence Pol II elongation properties, and the observed Pol II conformation may correspond to one of the functional states of the enzyme. Consistently, deletion of a region in bacterial RNA polymerase that corresponds to the Rpb1/9 jaw (Figure 32E) affects transcriptional pausing (Ederth *et al.*, 2002).

17.15 Conservation of transcript cleavage factors

Eukaryotic Pol III is capable of RNA cleavage, and its C11 subunit and magnesium ions are required for this activity (Chedin *et al.*, 1998). C11 consists of two zinc ribbon domains. The C-terminal zinc ribbon shows high sequence similarity to domain III of TFIIS, and comprises the two acidic hairpin residues (Figure 34). Since mutation of these two residues is lethal to yeast (Chedin *et al.*, 1998), the C-terminal domain of C11 may function like domain III of TFIIS, with a corresponding acidic hairpin playing an essential role. The N-terminal domain of C11 is related

in sequence to the N-terminal domain of Rpb9, and may occupy the same position, since only few amino acid residues would be required to connect it to its C-terminal domain that would be located in the pore. This model may also apply to the archaeal transcript cleavage factor TFS, which shows sequence similarity to C11, and contains the acidic hairpin residues (Hausner *et al.*, 2000). The role of the zinc finger of A12.2 in transcript cleavage remains unclear since it is completely dispensible in yeast (Van Mullem *et al.*, 2002) and transcript cleavage in Pol I is mediated by a TFIIS-like factor (Tschochner, 1996).

Zinc finger		**
TFIIS 274 :	CKEKKVSY	YQLQTRSADEPLTTFCTCEAC
A12.2 89 :	CGNEEMNYHTL	QLRSADEGATVFYTCCTSC
C11 75 :	CGGESAYF	FQLQIRSADEPMTTFYKCVNC

Figure 34: Alignment of the zinc finger domains of TFIIS and its homologues in RNA Pol I (A12.2) and Pol III (C11).

Invariant and conserved residues are shaded in dark grey and light grey, respectively. The asterisks mark the acidic hairpin residues D290 and E291 in TFIIS.

Bacteria do not contain a TFIIS homolog, but several lines of evidence have suggested that the bacterial transcript cleavage factors GreA and GreB function essentially like TFIIS. First, the binding site for GreB on bacterial polymerase corresponds to the rim of the Pol II funnel where TFIIS binds (Korzheva *et al.*, 2000; Koulich *et al.*, 1997). Second, the structure of GreA, although different from that of TFIIS, shows a coiled-coil protrusion with acidic residues at the tip (Stebbins *et al.*, 1995), which could reach the polymerase active site just like the TFIIS acidic hairpin does. Third, like TFIIS, Gre factors can be crosslinked to the RNA 3' end (Stebbins *et al.*, 1995).

Darst and coworkers published a manuscript back-to-back with the present results, that describes docking of structures of GreB and bacterial RNA polymerase to a 15 Å molecular envelope of the polymerase-GreB complex derived from electron microscopy (Opalka *et al.*, 2003). Strikingly, Opalka *et al.* find that the coiled coil of GreB binds in the secondary channel of bacterial polymerase, which corresponds to the Pol II pore, and reaches the active site with its acidic tip. The authors further show that the acidic tip residues are critical for Gre factor function. These findings are fully consistent with the findings discussed above and demonstrate in a powerful way the conserved strategies employed for RNA cleavage stimulation by the structurally unrelated bacterial and eukaryotic RNA polymerase cleavage factors.

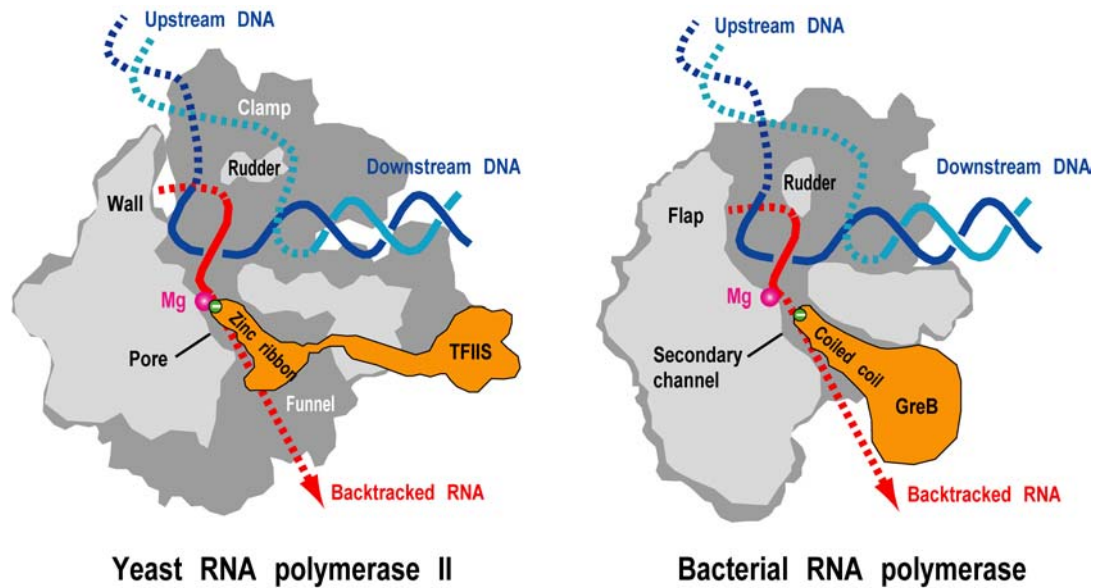


Figure 35: Common features of TFIIS-Pol II complex (left, this study) and the bacterial GreB-Pol complex (right), (Opalka *et al.*, 2003). Shown are major features of the complexes, including the RNA polymerase II pore and bacterial RNA polymerase channel, TFIIS, and GreB (orange), and the catalytic magnesium ion (metal A, pink). The positions of the conserved acidic residues in TFIIS and GreB are indicated by the green circles. The nascent transcript is shown in red, and the DNA template is blue. The presumed locations of backtracked RNAs are indicated by the dashed arrows. Adapted from (Conaway *et al.*, 2003).

18 ASSAYS FOR THE COMPOSITION OF COMPLEX CRYSTALS

18.1 Fluorescence-based detection of nucleic acids in the crystals

During the efforts to co-crystallise Pol II with synthetic nucleic acids, an assay was developed for the detection of nucleic acids in presumed complex crystals. The assay involves incubation of putative co-crystals with the nucleic acid-specific fluorescent dye SYBR-Gold, and their observation under a fluorescence microscope with standard FITC filters (see Mat&Meth). Crystals are incubated in mother solution containing trace amounts of SYBR-Gold, an asymmetric cyanine dye that binds to any class of nucleic acids, irrespective if single or double-stranded, DNA or RNA (Cosa *et al.*, 2001; Tuma *et al.*, 1999). Binding to nucleic acids enhances the fluorescence yield of SYBR-Gold dramatically (> 1000-fold), whereas proteins have essentially no effect (Tuma *et al.*, 1999). Nucleic acid-containing Pol II crystals show intense green fluorescence after 1-2 hours of incubation with SYBR-Gold, whereas crystals of the free polymerase remained essentially non-fluorescent under the same conditions (Figure 36).

To demonstrate the broad applicability of the assay for the detection of nucleic acids in crystals, it was applied to three different protein-nucleic acid complexes. One complex was the restriction enzyme NgoMIV with and without an 11 bp DNA duplex (a gift from Saulius Grazulis, Vilnius/Lithuania), for which the structure is known (Deibert *et al.*, 2000). Crystals of the restriction enzyme-DNA complex were compared side-by-side with free enzyme crystals under identical conditions. The NgoMIV-DNA complex crystals exhibited strong fluorescence, whereas crystals of the free enzyme remained essentially dark (Figure 36A). Crystals shown in Figure 36B were assembled from core Pol II and an 80mer synthetic RNA oligonucleotide (Table 6). RNA binding could meanwhile be confirmed by a crystallographic electron density map (not shown). The third crystal form investigated (Figure 36C) was comprised of complete Pol II with and without a transcription bubble and product RNA (Bubble 1 in Table 6)

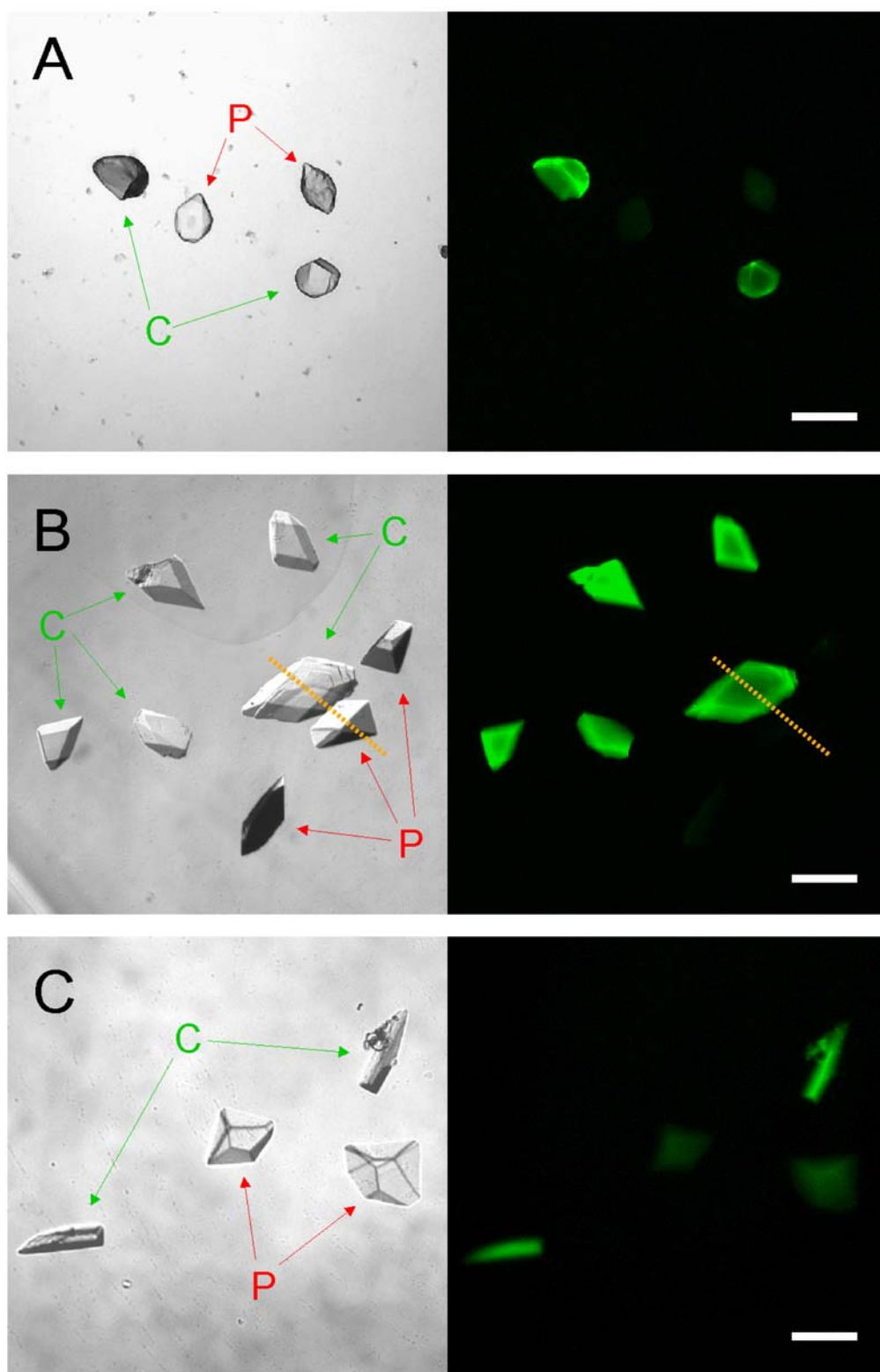


Figure 36: Selective fluorescence staining of nucleic acid-containing crystals.

Three comparisons of protein crystals (P) and protein-nucleic acid complex crystals (C) are shown. Transmission light microscopic images (left) reveal all crystals and fluorescence images (right) show selective staining of nucleic acid complex crystals after incubation with SYBR-Gold.

Figure 36 (continued)

(A) Crystals of the restriction enzyme NgoMIV, in free form and in complex with an 11 base pair DNA duplex (Deibert *et al.*, 2000).

(B) Crystals of Pol II, in free form and in complex with an 80-mer inhibitory RNA (Thomas *et al.*, 1997). The yellow line marks the path of the intensity-profile shown in Figure 37.

(C) Crystals of Pol II, in free form and in complex with a partially mismatched 41-mer DNA duplex and an 11-mer RNA.

The white scale bar corresponds to 0.1 mm.

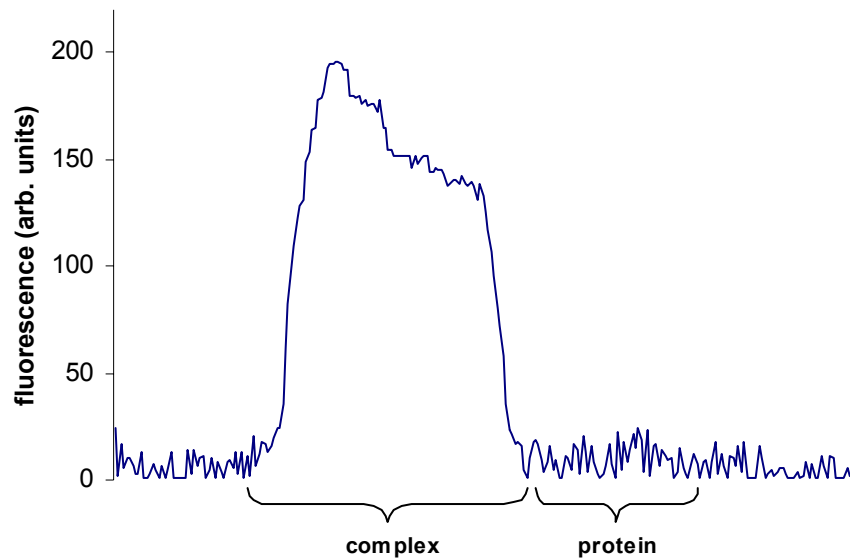


Figure 37: Intensity profile of the fluorescence micrograph shown in Figure 36B. Regions along the yellow line in Figure 36B, which correspond to the Pol II-RNA complex crystal and to the protein crystal, are indicated.

For all three nucleic acid complex crystals, fluorescence was directly visible under the fluorescence microscope, without the need for digital image processing. Quantified fluorescence intensities for complex crystals were at least tenfold, and up to 50-fold, higher than intensities observed for free protein crystals of similar size, demonstrating the specificity of the assay.

Since crystals of the free proteins remained essentially non-fluorescent, even after prolonged incubation with the dye overnight, it can be inferred that this method can reliably detect nucleic acid complex crystals, bypassing the need for control crystals of the free protein if these are unavailable. The sensitivity of the assay is remarkable, considering the low relative amounts of the nucleic acids in the crystallised complexes. The DNA in the NgoMIV-DNA complex contributes

19 % to the mass of the complex, whereas the nucleic acids in the Pol II-DNA-RNA complex and the Pol II-RNA complex account for only 7 % and 5 % of the complex mass, respectively.

In order to investigate the kinetics of crystal staining, a time course (Figure 38) was recorded. For observing a defined layer in the crystal during dye diffusion and accumulation, the experiment was done under a confocal microscope.

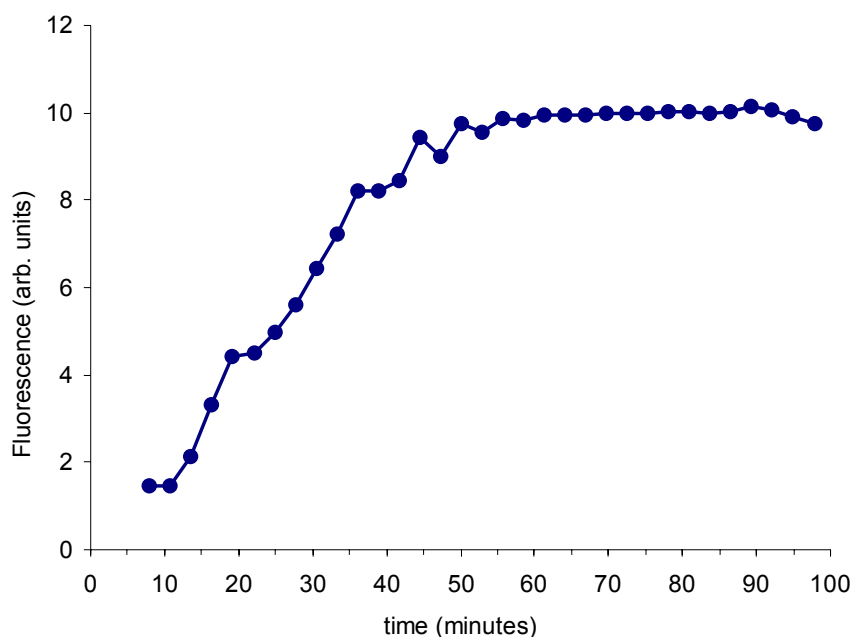


Figure 38: Time course of staining with SYBR-Gold. The time course was started with the immersion of a crystal (size approx. 100 by 80 by 80 μm) in mother solution containing SYBR-Gold. The increasing fluorescence intensity in a thin slice (3.9 μm), integrated in a polygon encircling the crystal, was followed under a confocal microscope.

Under the experimental conditions used, saturation of the crystal with the dye was reached after about one hour. Since SYBR-Gold binds to nucleic acids within seconds (Tuma *et al.*, 1999), the slow staining apparently reflects dye diffusion through the solvent channels of the crystals. This implies that small molecules can take hours to penetrate a crystal, and to occupy all binding sites. Soaking times of several hours may therefore be beneficial for derivatisation of crystals with small ligands or for cryoprotection. By means of this assay, those conditions could be identified, which lead to maximum fluorescence, i.e. where the occupancy of nucleic acids was as complete as possible. Optimising occupancy was particularly critical because both apo-Pol II and the nucleic acid complexes crystallise under the same conditions and give rise to the same crystal form (Table 8). Thus, both the complex and free Pol II could be present in the same crystal. However, a slight

excess of free nucleic acids shifts the equilibrium to the side of the complex and low-salt crystallisation conditions were found to diminish the dissociation of elongation complexes (see chapter 16.4).

18.2 Monitoring protein soaking into crystals

The initial inspiration to soak additional proteins into preformed crystals of the complete 12-subunit RNA polymerase II resulted from the crystallographic analysis. It revealed a solvent content of 80 % (see chapter 16.4) and wide solvent channels of up to 140 Å in diameter (Figure 39).

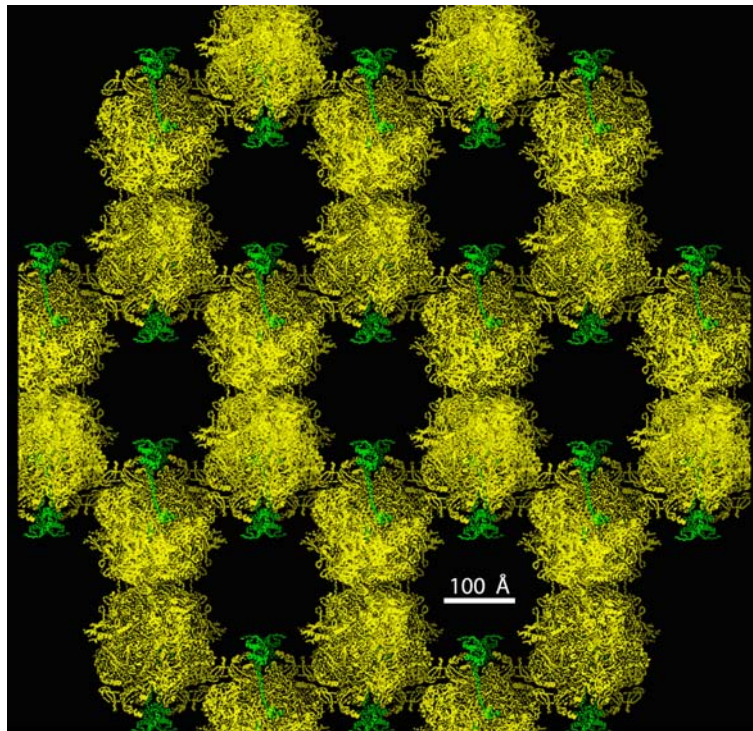


Figure 39: Crystal packing in complete Pol II crystals soaked with TFIIS. The view is along the crystallographic *c*-axis. 12-subunit Pol II is shown in yellow and TFIIS bound to Pol II is shown in green.

Although a 20 kDa polypeptide like TFIIS should easily fit through the solvent channels, it cannot be predicted whether a specific target site on the polymerase surface is accessible in the crystal packing. There is only one example in the literature where a multi-component complex was assembled by soaking a protein into pre-formed crystals (Carter *et al.*, 2001). In that case, the translation initiation factor IF1, a 8.1 kDa protein, was diffused into crystals of the 30S ribosomal subunit and the structure was determined at 3.1 Å resolution.

To determine whether TFIIS can indeed diffuse into polymerase crystals, and whether it can bind to its specific site on the polymerase surface, accumulation of TFIIS in the crystals was monitored by fluorescence microscopy. A Cy3 label was covalently attached to TFIIS. Crystals were incubated in a solution of the labelled TFIIS, and were observed under a fluorescence microscope. The fluorescence intensity within the crystals increased up to around 2-fold over background after 4 hours (Figure 40).

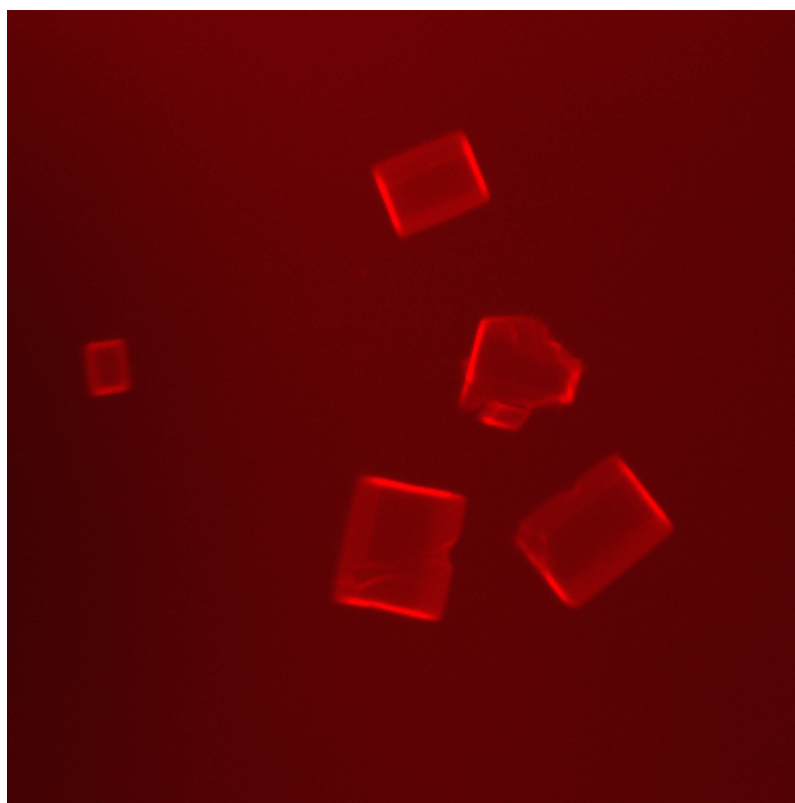


Figure 40: Fluorescence micrograph of complete Pol II crystals soaked for 4 hours in Cy3-labelled TFIIS. TFIIS accumulates within the crystals, since mere diffusion into crystals would not lead to an increase of fluorescence over background.

This indicated that TFIIS indeed bound to the polymerase and accumulated in the crystals, since mere diffusion of TFIIS through the crystals would not have resulted in an increase of fluorescence intensity over background levels. This assay allowed testing protein soaking in various buffers, and with crystals grown under different conditions. Specifically, the original phosphate-PEG20K crystallisation conditions (Table 7) caused most of TFIIS to precipitate before it could penetrate the crystals and TFIIS. The result was a protocol that led to maximum

fluorescence intensity, and thus to maximum TFIIS occupancy (see Materials and Methods). It was found that the crystals withstand a high concentration of TFIIS, up to 2 mg/mL, which corresponds to a concentration of 0.1 mM or $\approx 1000 \cdot K_D$ (Awrey *et al.*, 1998; Sijbrandi *et al.*, 2002; Wu *et al.*, 1996), suggesting that the protocol is applicable to weak and transient protein interactions that have dissociation constants in the micromolar range.

Attempts to derivatise crystals of the 10-subunit RNA polymerase II core (Cramer *et al.*, 2000; Cramer *et al.*, 2001; Fu *et al.*, 1999) with TFIIS under similar conditions failed and the fluorescence assay showed labelled TFIIS not to accumulate in the crystals. Compared to the complete polymerase crystals, this crystal form (see Table 8) has a much lower solvent content of around 56 %, and a much larger area of the polymerase surface, including the TFIIS binding site, is involved in crystal contacts (not shown). Thus, the fluorescence assay could correctly predict the outcome of a later crystallographic analysis of protein-soaked polymerase crystals, and allowed to optimise the soaking procedure to achieve maximum occupancy of the additional protein on the polymerase surface, dramatically accelerating the crystallographic analysis. Nevertheless, such diffusion and binding of a protein into a crystal is limited to exceptional cases, since many prerequisites have to be fulfilled: The preformed crystals need to have large solvent channels that allow the diffusion of the additional protein into the crystal lattice. Moreover, the specific binding site of the soaked protein on the molecular surface of the target must be accessible and not involved in crystal contacts, and possible conformational changes accompanying binding must be tolerated by the crystal lattice. Soaking of Pol II crystals with labelled subunits of TFIIF, with TFIIE, Spt4/5 and the Mediator subcomplex MED7/MED21 did not lead to any detectable accumulation within the crystals.

19 ISOLATION AND CRYSTALLISATION OF POL II WITHOUT CTD

In all crystal structures of Pol II solved so far, the C-terminal repeat domain (CTD) is apparently disordered. Nevertheless, these disordered 281 amino acids comprise a mass of about 30 kD. Crystals of 12-subunit Pol II are characterised by an unusually high solvent content (see chapter 17.2), i.e. by a loose packing of Pol II molecules in the crystal lattice. Thus, proteolytic removal of the (essential) CTD prior to crystallisation was thought to facilitate the formation of a more closely packed crystal form, eventually leading to higher resolution. It has been shown that the initiation and chain elongation properties of a CTD-less Pol II are unchanged compared to the full-length enzyme (Li and Kornberg, 1994). Different from a previously described factor Xa

protease site immediately preceding the first heptapeptide repeat (Li and Kornberg, 1994), the present strategy aimed to remove all of the CTD including the disordered part of the linker.

Mutagenesis in the yeast strain CB010 Δ Rpb4 as described in chapter 25 of Materials and Methods resulted in the strain (CB010) Δ rpb4 RPB1-TEV, which carried a TEV protease cleavage site in the linker region of Rpb1. Hence, the C-terminal domain could be cleaved off during Pol II purification (see Materials and Methods chapter 25.3).

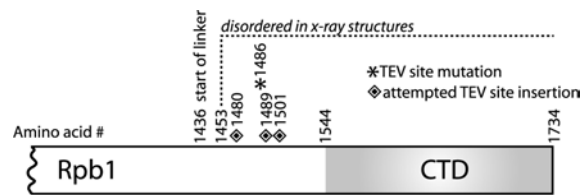


Figure 41: Attempted insertions and mutations of the TEV protease site into Rpb1. Diamonds indicate the three sites in the disordered linker region of Rpb1 at which the heptapeptide TEV protease consensus sequence was inserted. An asterisk shows the position of the (successful) mutation of *RPB1* into a TEV protease cleavage site, leaving the linker length unchanged. Amino acid numbers of the respective sites are given below the bar.

Interestingly, only a mutation but not an insertion of the heptapeptide TEV protease consensus sequence gave rise to a viable strain in the context of the *rpb4* knockout strain. This indicates that, at least in the context of this strain, alterations in the linker length are detrimental for yeast viability. The resulting strain is characterised by slow growth (doubling time in the fermenter at 25 °C is about 3.5 hours) and a clustering phenotype (Figure 42). A mutation of five residues plus an insertion of another residue in the linker-CTD junction was however not reported to show an apparent growth defect in an otherwise wildtype strain (Li and Kornberg, 1994).

Treating this purified Pol II variant with TEV protease lead to quantitative cleavage of the CTD (Figure 43), so that an additional step in the purification protocol, namely dialysis and TEV treatment after the immunoaffinity step (see Materials and Methods chapter 21.2), resulted in crystallisable amounts of Pol II- Δ CTD. However, compared to the wild-type variant, overall yields are about 50 % lower.

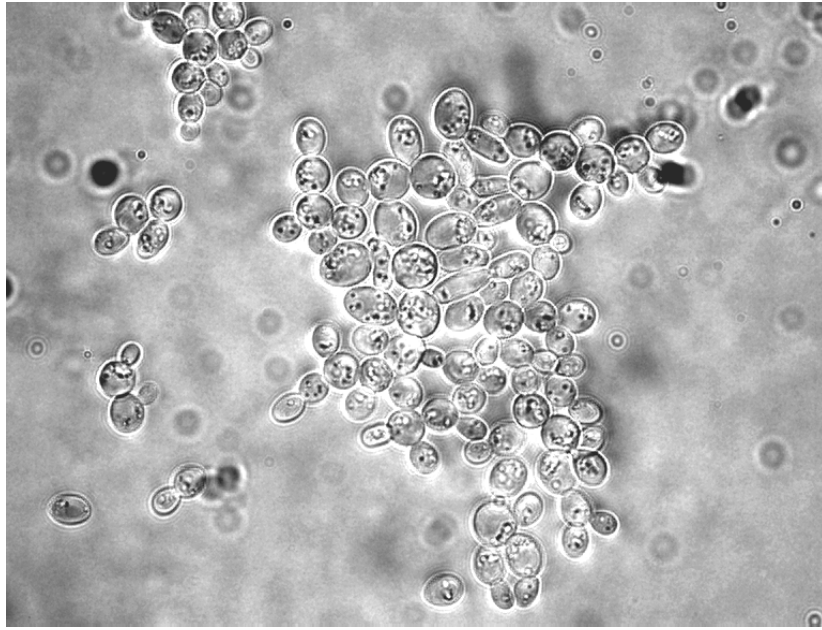
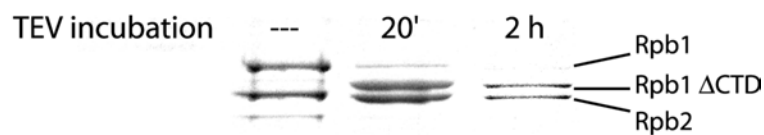


Figure 42: Micrograph of yeast strain (CB010) Δ rpb4 RPB1-TEV (80-fold magnification). Cells were grown in the small fermenter at 25 °C to late logarithmic phase.

A



B

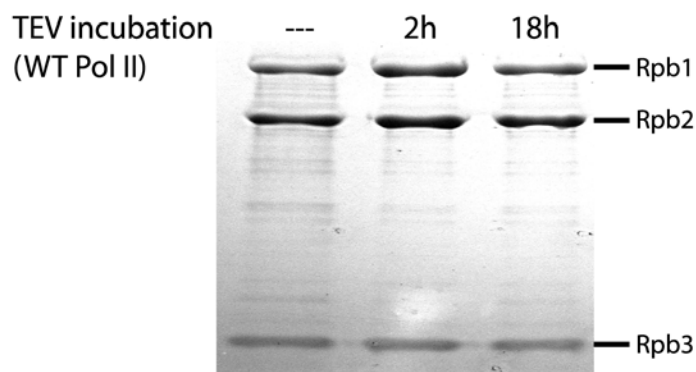


Figure 43: TEV protease treatment of Pol II.

The silver-stained SDS-polyacrylamide gel shows Rpb1 and Rpb2 of Pol II after the immunoaffinity step (see chapter 21.2). Samples were treated with 0.2 $\mu\text{g}/\mu\text{L}$ TEV protease in buffer UnoQ-A (see Materials and Methods) and incubated at 25 °C for the time indicated. (A) Timecourse to show rapid cleavage of Rpb1 by TEV protease. (B) Wild-type Pol II does not undergo proteolysis when treated with TEV protease.

Nucleic acid complexes of CTD-less core Pol II have been subjected to a set of commercial crystal screens (Crystal Screen 1 and 2, PEG-Ion Screen, Natrix [Hampton Research]), but no immediate hits could be obtained (Florian Brückner, personal communication). Nevertheless, 12-subunit Pol II lacking the CTD gives rise to large, three-dimensional crystals. These form more readily and more reproducibly as crystals of the full-length Pol II. Crystallographically, these crystals are indistinguishable from full-length Pol II in terms of unit cell dimensions and space group (data not shown). Generally, larger size enhances diffraction intensity and reduces the susceptibility to radiation damage. Indeed, the first complete dataset to 3.6 Å resolution was obtained with a CTD-less Pol II crystal (Florian Brückner, personal communication).

20 CONCLUSION AND OUTLOOK

Details of the mechanisms of RNA elongation, substrate selection, maintenance of the transcription bubble and RNA exit were revealed by the results presented in this thesis. The structural analysis of the Pol II-TFIIS complex gave experimental evidence for a tuneable active site of Pol II, which can perform both RNA elongation and cleavage. The model system of complete, 12-subunit Pol II, together with a complete transcription bubble, product RNA, and the extrinsic transcription factor TFIIS approximates the *in vivo* situation of a regulated elongation complex as closely as possible at the moment.

Further research will have to clarify the relevance of proposed conformational changes during the nucleotide addition cycle and during pausing and arrest. Higher resolution data are necessary to reveal chemical details responsible for the selection of correct nucleotides and the consequences of transcription through DNA lesion sites. Structures of other Pol II-transcription factor complexes will shed light onto the way in which these factors dynamically regulate Pol II. An initiated structural investigation of an RNA inhibitor bound to Pol II is the first step into an investigation of the way Pol II transcription is being regulated by RNAs.

Results and Discussion

21 ISOLATION OF RNA POL II FROM YEAST

21.1 Yeast fermentation

Core RNA polymerase II was isolated from the *Saccharomyces cerevisiae* strain CB010ΔRpb4 (MATa *pep4::HIS3/prb1::LEU2, prc1::HISG, can1, ade2, trp1*) (Edwards *et al.*, 1990; Fu *et al.*, 1999). This strain carries knockouts of several cellular proteases and of RPB4. In the absence of Rpb4, Rpb7 dissociates from core Pol II during purification, giving rise to homogeneous 10-subunit core Pol II. Two types of fermenters were available for producing up to 1.5 kg of yeast pellet per batch. The small fermenter (ISF200, Infors) has a nominal volume of 20 L and is ideally run with up to 15 L of media. The large fermenter (ABEC, Infors) has a nominal volume of 200 L and can be filled with up to 160 L of media. Table 11 shows the media composition and the culture parameters of both fermenters:

Table 11: Conditions for the fermentation of yeast

	small fermenter (20 L)	large fermenter (200 L)
YPD media	300 g peptone	3200 g peptone
	300 g glucose	3200 g glucose
	222 g yeast extract	2370 g yeast extract
	15 L desalted water	160 L desalted water
Antibiotics ¹	0.75 g ampicillin	8.0 g ampicillin
	0.15 g tetracycline·HCl	1.6 g tetracycline·HCl
typical inoculate volume	0.3 L @ OD ₆₀₀ ≈ 2	4-5 L @ OD ₆₀₀ ≈ 2
air flow	8 L/min	20 L/min
stirrer speed	800 rpm	200 rpm
typical growth time	12 – 15 hours	12 – 15 hours

¹added after sterilisation, prior to inoculation

21.2 Purification of RNA Pol II

21.2.1 Buffers

<u>3x freezing buffer</u>	<u>100 x protease inhibitor mix (p.i.)</u>
150 mM Tris-Cl, pH 7.9 @ 4 °C	1.42 mg Leupeptin
3 mM EDTA	6.85 mg Pepstatin A
30 % glycerol	850 mg PMSF
30 µM ZnCl ₂	1650 mg benzamidine
3 % DMSO	dry ethanol to 50 mL
30 mM DTT	
3 x protease inhibitor mix	stored at -20 °C; added immediately before use
<u>1 x HSB150</u>	<u>1 x HSB600</u>
50 mM Tris-Cl, pH 7.9 @ 4 °C	50 mM Tris-Cl, pH 7.9 @ 4 °C
150 mM KCl	600 mM KCl
1 mM EDTA	1 mM EDTA
10 % glycerol	10 % glycerol
10 µM ZnCl ₂	10 µM Zn Cl ₂
10 mM DTT	10 mM DTT
1 x protease inhibitor mix	1 x protease inhibitor mix
<u>1x TEZ</u>	<u>UnoQ buffer</u>
50 mM Tris-Cl, pH 7.5 @ 20 °C	50 mM Tris-Cl, pH 7.5 @ 20 °C
1 mM EDTA	1 mM EDTA
10 µM ZnCl ₂	10 µM ZnCl ₂
5 mM DTT	10 % (v/v) glycerol
1 x protease inhibitor mix	10 mM DTT
	no protease inhibitors
<u>1 x Pol II buffer</u>	<u>Acetate buffer</u>
5 mM Hepes pH 7.25 @ 20 °C	100 mM sodium acetate pH 4.0
40 mM ammonium sulphate	500 mM sodium chloride
10 µM ZnCl ₂	
10 mM DTT	
<u>PBS</u>	<u>Coupling buffer</u>
4.3 mM Na ₂ HPO ₄	100 mM sodium bicarbonate pH 8.3
1.4 mM KH ₂ PO ₄	500 mM sodium chloride
137 mM sodium chloride	
2.7 mM potassium chloride	
pH 7.4	

21.2.2 Harvesting and storage of yeast

Yeast was harvested at late logarithmic/early stationary phase, monitored by OD_{600} measurement. Cells were pelleted by centrifugation for 20 min. at 5000 rpm in a SLC6000 rotor (small fermenter) or by a continuous flow centrifuge (Padberg Z4IG, 20000 rpm) in case of the large fermenter. The cell pellet was resuspended in 330 mL of 3x freezing buffer per kg cells and stirred at 4 °C for 30 min., before shock-freezing in liquid nitrogen and stored at – 80 °C.

21.2.3 Purification - day 1 (lysis and heparin column)

For three bead-beaters (BioSpec), up to 600 mL of cell suspension were thawed in warm water. Each bead-beater was filled with 200 mL of borosilicate glass beads (0.45-0.50 mm diameter), 1 mL of protease inhibitor mix and 200 mL of the cell suspension. HSB150 was added to fill the bead-beater completely, taking care to avoid any remaining air bubbles. Lysis was achieved within one hour of bead-beating (30 s on/90 s off) while the beater chambers were submersed in a salt/ice mixture.

Glass beads were removed by filtration through a mesh funnel. The beads were washed with HSB150 until the flowthrough was clear. The lysate was cleared by two rounds of centrifugation (30 min. at 12000 rpm in a GS3 rotor). Lipids were then removed by filtration of the supernatant through two layers of paper filter discs underneath a dressing cloth.

The cleared lysate was applied onto a column packed with 250 mL of Heparin Sepharose 6 FF (Amersham), pre-equilibrated with 750 mL of HSB150. Elution was accomplished with 500 mL of HSB600. Proteins in the eluate were precipitated by adding 291 g of fine-ground ammonium sulphate per litre of eluate (= 50 % saturation), followed by 20 min. of stirring at 4 °C and centrifugation (45 min. at 12000 rpm in a GS3 rotor). The pellet was stored over night at 4 °C. The heparin column was restored by washing with 1 L of 6 M urea, followed by water, and stored in 5 mM potassium acetate and 20 % (v/v) ethanol. Every five runs, the heparin column was regenerated by a brief wash with 500 mL of 0.1 M NaOH, followed by water and 5 mM potassium acetate in 20 % (v/v) ethanol.

21.2.4 Purification - day 2 (immunoaffinity column)

The ammonium sulphate pellet of day 1 was dissolved in 50 mL of buffer TEZ. More TEZ was added to adjust the conductivity below the conductivity of TEZ containing additionally 400 mM ammonium sulphate (TEZ400). This sample was loaded by gravity flow onto the immunoaffinity column (see chapter 21.2.6). The column was pre-equilibrated with 20 mL of TEZ containing 250 mM ammonium sulphate (TEZ250). The flowthrough was re-loaded to yield 10-20 % more Pol II. The column was brought to room temperature, washed with 25 mL of TEZ250 at room temperature and Pol II was eluted in 1 mL fractions with TEZ500 containing additionally 50 % (v/v) glycerol (ca. 15 mL). Directly afterwards, the column was washed with 5 mL of TEZ500 containing 70 % (v/v) ethylene glycol but no DTT, and re-equilibrated with 25 mL of TEZ250 containing 0.02 % sodium azide. The column can be used > 10 times before recovery decreases.

21.2.5 Purification - day 3 (anion exchange column)

Peak fractions from day 2 were diluted six-fold and loaded onto a UnoQ column (BioRad, column volume 1.35 mL), pre-equilibrated with buffer UnoQ containing 60 mM ammonium sulphate (UnoQ-A). The column was washed with 3 column volumes of this buffer, and Pol II was eluted with a linear gradient over 10 column volumes from 0 – 50 % buffer UnoQ containing 1 M ammonium sulphate (UnoQ-B). Pol II elutes at about 25 % buffer UnoQ-B. The column was restored by washing with 5 column volumes of UnoQ-B.

Peak fractions were pooled and split into aliquots of 500 µg Pol II. The aliquots were mixed 1 : 1 with ammonium sulphate solution saturated at room-temperature, incubated for > 1 hour at 4 °C and centrifuged for 45 min. at 4 °C in a table-top centrifuge at 15000 rpm. Most of the supernatant was decanted so that the pellet was still covered with supernatant, before it was shock-frozen in liquid nitrogen and stored at –80 °C. Pol II stored this way is stable for about 3 months. From 600 g yeast pellet, a yield of 5-8 mg of highly purified Pol II can be expected.

21.2.6 Preparation of Pol II immunoaffinity resin

The monoclonal antibody 8WG16 (NeoClone, Madison/USA), described in (Thompson and Burgess, 1996) is specific for the unphosphorylated CTD of Pol II and optimised to release Pol II upon treatment with glycerol or ethylene glycol at room temperature (“polyol responsive

antibody“). The antibodies were purified from mouse ascites and immobilised on activated chromatography media according to the following procedure:

Lyophilised ascites was dissolved in PBS to its original volume and filtered through 0.2 μm membrane filters. The solution was passed > 3 times through a protein-A sepharose column (5 mL column volume, Sigma), pre-equilibrated in PBS. The column was washed with 50 mL PBS and antibodies were eluted with 20 mL of 0.75 M acetic acid. Fractions of 1 mL were collected into tubes containing 200 μL of 2 M Hepes (pH 7.9) to neutralise the acid. Peak fractions were pooled and the protein-A sepharose column was regenerated by washing for 5 min. with 1 M acetic acid, followed by PBS with 0.02 % sodium azide.

The matrix for immuno-affinity columns was cyanogen bromide (CNBr)-activated sepharose 4 B (Sigma), which reacts with free amines, e.g. accessible $-\text{NH}_2$ groups on proteins. Care was taken to avoid other sources of free amines (e.g. Tris) and to use a sealed bottle of activated sepharose. For each immunoaffinity column, 5 mL of gel was prepared by suspending 1.43 g of CNBr-sepharose in several mL of 1 mM HCl in a disposable gravity-flow column. The suspended CNBr-sepharose was washed with 20 mL of coupling buffer. Coupling was performed with 10 mg of purified antibodies per column for 2 hours at 20 $^{\circ}\text{C}$ or over night at 4 $^{\circ}\text{C}$. When the coupling reaction was completed, no protein was detectable in the supernatant. The column was then washed with 25 mL of 1 M Tris, pH 8 and incubated for 2 hours at room temperature. Finally, the column was washed with 20 mL of coupling buffer, followed by acetate buffer and coupling buffer. Columns were stored at 4 $^{\circ}\text{C}$ in TEZ60 with 0.02 % sodium azide and could be used > 10 times, if DTT exposure was reduced to a minimum.

22 PURIFICATION OF THE SUBCOMPLEX RPB4/7

Buffer 1

150 mM NaCl
5 % (v/v) glycerol
50 mM Tris pH 7.5
10 mM β -mercaptoethanol
protease inhibitors
(see chapter 26.2)

Buffer 2

50 mM Tris pH 7.5
5 mM DTT
1 mM EDTA

Recombinant yeast Rpb4/7 was cloned and overexpressed in *E. coli* using a bicistronic vector, as described (Sakurai *et al.*, 1999). Cells from 2 L of culture were resuspended in buffer 1 and lysed by means of a French Press. The lysate was cleared by centrifugation (30 min. at 15000 rpm in a SS34 rotor) and applied onto a NiNTA column (Quiagen; 1 mL column volume). The column was washed with 5 mL of buffer 1 containing additionally 20 mM imidazole. Elution was performed with 6 mL of buffer 1 containing additionally 200 mM imidazole. Peak fractions were pooled, diluted 1:3 with buffer 2 and applied on a ResourceQ column (Amersham, 6 mL column volume), pre-equilibrated in buffer 2. Rpb4/7 was eluted with a linear gradient from 0-1000 mM NaCl in buffer 2. Peak fractions were concentrated and applied on a Superose12 HR10/30 gel filtration column (Amersham), pre-equilibrated in Pol II buffer (see chapter 21.2). The purified Rpb4/7 heterodimer was concentrated to 10 mg/mL and aliquots were stored at -80°C .

23 ELECTROPHORETIC MOBILITY SHIFT ASSAY

TGEM-buffer

25 mM Tris pH 7.5
190 mM glycine
1 mM EDTA
8 mM magnesium acetate
10 % (v/v) glycerol
20 g/L PEG6000.

Sample buffer

TGEM buffer containing
additionally 40 % (v/v)
glycerol

Nucleic acids were analysed for their capability to bind Pol II by electrophoretic mobility shift assays, essentially as described (Daube and von Hippel, 1992). In 10 μL of Pol II buffer (see chapter 21.2.1), 100 pmol of nucleic acids were incubated with 20 pmol of Pol II for 20 minutes at 20°C . 4 μL of sample buffer was added before the samples were loaded on a 5 % acrylamide gel in TGEM buffer. Electrophoresis was carried out at 4°C at a voltage of 70-90 V for 90-120 minutes. Nucleic acids were stained for 30 minutes with 0.01 % SYBR gold in TBE buffer and visualized with 310 nm UV light. Thereafter, protein bands were stained with Coomassie Brilliant Blue.

24 TRANSCRIPTION ASSAY

This assay was utilised as a quality control for the purified Pol II complex, showing that it is transcriptionally active on the template used for crystallisation.

Transcription buffer

20 mM Hepes pH 7.5
10 % glycerol
8 mM magnesium chloride
40 mM ammonium sulphate

Stop solution

24 μ L of 10 % (w/v) SDS
26 μ L proteinase K (20 mg/mL)
750 μ L RNase-free water (Sigma)

To 560 pmol of purified Pol II in 120 μ L of transcription buffer, 1.4 equivalents of an artificial transcription bubble (Table 6) and 1 mM each of CTP, UTP, ATP and GTP was added. The mixture was incubated at 28 °C for 20 minutes. The reaction was quenched by adding 800 μ L of stop solution. This mixture was incubated for 1 hour at 40 °C before 600 μ L Phenol/Chloroform (1:1) (Rotiphenol, Roth) were added and RNA was extracted. To precipitate RNA, 6 μ L glycogen (20 mg/mL), 60 μ L 5 M NaCl and 2.4 mL cold ethanol were added and precipitation was allowed to proceed over night at -20 °C. RNA was isolated by centrifugation at maximum speed at 0 °C. RNAs were separated on a denaturing 3 % acrylamide-urea gel (Sambrook, 2001), stained for 30 minutes with 0.01 % SYBR gold in TBE buffer, followed by 30 minutes of destaining with water. RNA bands were visualised with 310 nm UV light.

25 INTRODUCTION OF A TEV PROTEOLYSIS SITE INTO RPB1

25.1 Construct design and mutagenesis

The consensus sequence of the cleavage site of tobacco etch virus (TEV) NIa protease (Carrington and Dougherty, 1988; Parks *et al.*, 1992) was introduced into the linker region of Rpb1 by mutating the residues DELMFSP (residues 1488-1497) to ENLYFQG. This site was chosen because it required a minimal number of unconservative mutations. Previous experiments to insert the consensus sequence into three different positions in the linker, rather than keeping the linker length constant, failed (see chapter 19). To create a construct for homologous recombination in yeast, the sequence encoding the C-terminus of Rpb1 (residues 1271-1734) and 440 bases of the 3'-UTR was amplified from wild-type yeast genomic DNA by the primer pair A_forw and A_rev.

Table 12: Primers for introduction of a TEV cleavage site in Rpb1

name	sequence (5' → 3')
A_forw	TCCCCCGGGGGGACATTACGTGGTGTAGAGAACATCGAG
A_rev	GCGTATTTCCGTGATCTGGTAATGCGCGGATCCGCG
mut_forward	GAAAATTTGTACTTTCAAGGTGATCTTGACGTTAAAGATGAGC
mut_reverse	ACCTTGAAAGTACAAATTTTCGTTACTGTATGGTGTGACGCC
458_fo	GATTGTCGTGGTGTTCG
1400_rev	CCCAGGCTTTACACTTTATG
trp_rev	CGTCTGCAAGCCGCAAACCTTTC

shading marks the sequence coding for the TEV-site

Mutagenesis was performed with the overlapping pair of primers mut_forward and mut_reverse (Table 12). This fragment was cloned into the vector pBS II KS(+) (Stratagene) using the restriction sites XmaI and BamHI. A *TRP1* auxotrophy marker cassette was inserted at position 5408 of the 3'-UTR. For this purpose, the plasmid described above was cut with SmaI (NEB). The *TRP1* cassette was excised from the plasmid YDp-W (Berben *et al.*, 1991) using BamHI. Blunt ends of the excised fragment were created with Pfu DNA polymerase (New England Biolabs), before it was ligated into the linearised vector using T4 ligase (Fermentas). The correct incorporation of the *TRP1* gene was confirmed by PCR using primer pair 458_fo and trp_rev, that anneal in the *TRP1* insert (Table 12, Figure 44) and in the 3'-UTR before the insertion. The correct sequence of the whole construct was confirmed by DNA sequencing.

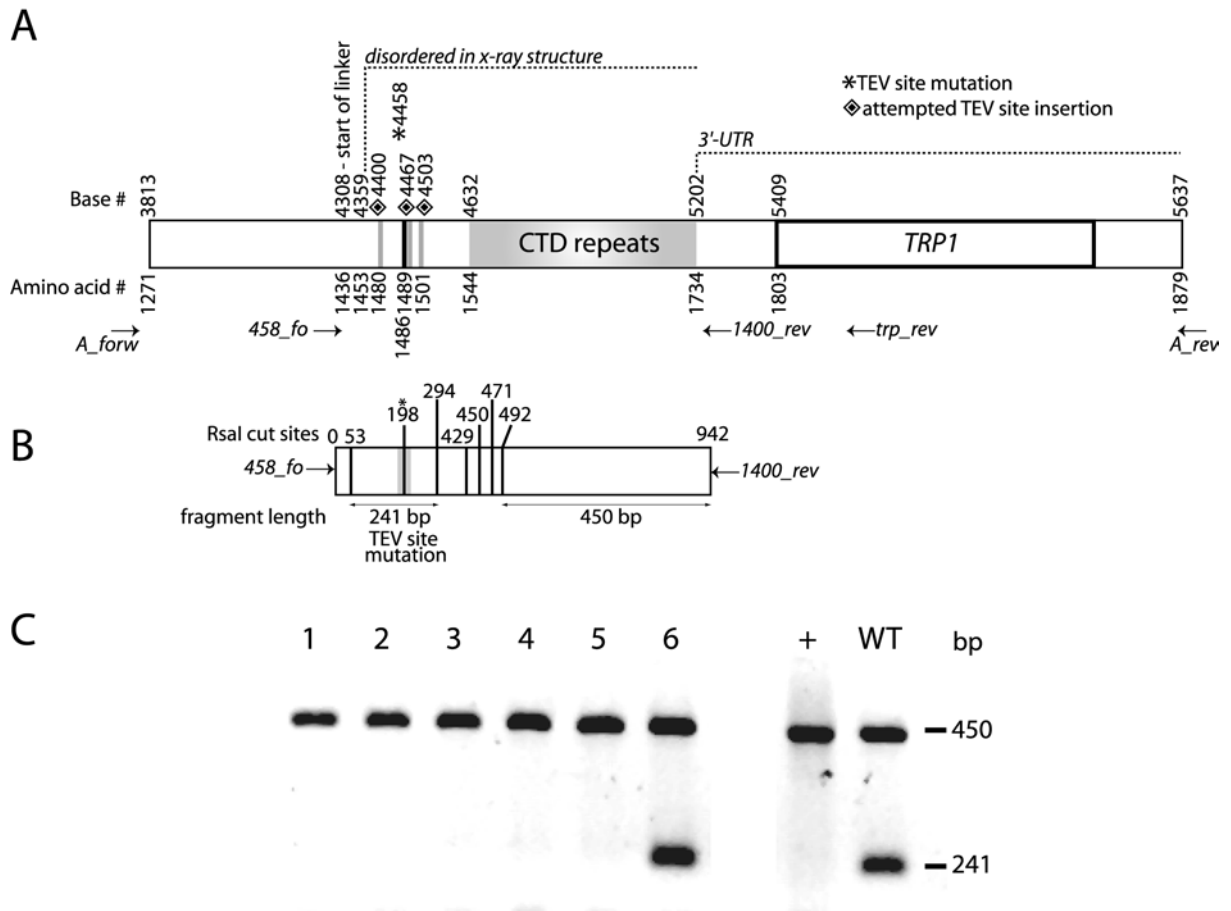


Figure 44: Cloning strategy for the inclusion of a TEV protease site in the Rpb1 “linker” region.

(A) Schematic representation of the C-terminus and the 3' UTR of *RPB1*. Sites of attempted but unsuccessful TEV site insertions are indicated with grey lines. An asterisk and a black line marks the site of mutation yielding viable clones. Numbers refer to the base or the amino acid followed by the mutation. The position of the C-terminal repeat domain (CTD) and *TRP1* cassette in the transformed fragment are indicated. Primers used for cloning and test-PCRs are indicated with arrows.

(B) A 942 bp DNA sequence surrounding the desired site of mutation was amplified with the primer pair indicated with arrows. *RsaI* cut sites are depicted as black lines. An additional *RsaI* site, marked with an asterisk, is present in the sequence coding for the TEV site.

(C) Agarose gel (1 %) of six clones tested for the presence of the TEV-site by *RsaI* digest. The restriction profile shown in (B) leads to a 450 bp and a 241 bp fragment (and to shorter fragments not shown on this gel). The 241 bp fragment is absent in the positive case. 1-7: clone numbers; +: positive control; WT: wild-type. All clones except #6 carry the desired mutation.

25.2 Yeast transformation

For the transformation, the construct described above was cut from the plasmid using *XmaI* and *BamHI*. Yeast transformation was performed according to the lithium acetate-PEG procedure (Ito *et al.*, 1983):

<p><u>Buffers:</u> <u>10 x TE</u> 100 mM Tris pH 7.4 10 mM EDTA (sterile filtered)</p> <p><u>solution 1</u> 0.5 mL 10 x TE 0.5 mL 10 x LiAc 4 mL water (sterile filtered)</p> <p><u>Media:</u> <u>YPD medium</u> 20 g peptone 20 g glucose 14.8 g yeast extract H₂O ad 1 L</p>	<p><u>10 x LiAc</u> 1 M lithium acetate pH 7.5 (sterile filtered)</p> <p><u>solution 2</u> 0.5 mL 10 x TE 0.5 mL 10 x LiAc 4 mL 50 % (w/v) PEG4000</p> <p><u>SDC-Trp medium</u> 6.75 g Yeast nitrogen base (Formedia) 20 g glucose 50 mg arginine HCl 80 mg aspartate 50 mg isoleucine 50 mg lysine HCl 20 mg methionine 50 mg phenylalanine 100 mg threonine 50 mg tyrosine 140 mg valine 20 mg adenine 20 mg histidine 100 mg leucine 20 mg uracil H₂O ad 1 L pH 5.5</p>
---	--

For five transformations, 50 mL of yeast culture were grown to an OD₆₀₀ of 0.5-0.8 and pelleted by centrifugation for 5 min. at 3600 rpm. As a host strain, the BJ926 strain described in chapter 21.1 was used which proved to be suitable for producing crystallisable core Pol II. The pellet was washed with 10 mL of sterile water, centrifuged again and resuspended in 500 µL of solution 1. Cells were washed once with solution 1 and resuspended in 250 µL of solution 1. For each transformation, 50 µL of the cell suspension was mixed with 1-4 µg of the linearised DNA in < 4 µL TE buffer and mixed with 300 µL of solution 2 and 5 µL of carrier DNA. The mixture was incubated at room temperature for 30 min. and heat-shocked for 10 min. at 42 °C, followed by a 3 min. incubation on ice. Cells were then washed twice with water, resuspended in 100 µL of water and spread on SDC-Trp selective media plates.

Colonies that grew within 3 days at 25 °C on SDC-Trp plates were re-plated and grown for another 3 days at 25 °C. From 19 colonies and a wild type control, genomic DNA was isolated and the region of interest (1824 bp, Figure 44) was amplified by PCR using the primer pair A_forw and A_rev (Table 12). The presence of one correctly oriented copy of *TRP1* in the 3'-UTR of *RPB1* was confirmed by PCR using the primer pair 458_fo and trp_rev. These primers were designed to yield fragments of 1298 bp in case one correctly oriented copy of *TRP1* is present in the template.

In order to assay for the presence of the TEV cleavage site, 942 bp of DNA including the site of mutagenesis was amplified by PCR using the primers 458_fo and 1400_rev. The presence of the mutation encoding for the TEV cleavage site introduces an additional RsaI restriction site. Therefore, clones, carrying the desired mutation can be identified by a distinct restriction pattern (Figure 44). From the 19 clones tested, 17 clones had the *TRP1* gene in the correct position. From these 17 clones, the fastest growing six clones were further analyzed. Eventually, five out of six clones carried the desired TEV site mutation. No other mutation was present in the gene coding for Rpb1p as confirmed by DNA sequencing.

It is noteworthy that an alternative strategy to insert rather than to mutate the TEV consensus sequence in three positions in the linker (after residues 1480, 1489 and 1501, respectively, Figure 44) following an analogous procedure did not lead to any viable clones carrying the mutation, although about 60 clones per site were analysed.

The resulting yeast strain carrying the TEV site mutation in *RPB1* is characterized by a clumpy phenotype and a prolonged doubling time of about 4 hours, compared to about 2.5 hours for the original BJ926 strain. This phenotype makes cell density determination by measuring the optical density imprecise. Therefore, cells were counted directly using a Neubauer counting chamber.

25.3 Isolation of Pol II Δ CTD

The first steps of RNA Pol II preparation, namely chromatography via a heparin column and an immunoaffinity column were performed as described above (chapter 21.2.4). No protease inhibitors were present in the wash and elution buffer of the immunoaffinity column. The eluate was mixed with 400 μ g of recombinant TEV protease and dialysed against UnoQ buffer A in a

dialysis membrane (5 kDa cutoff) over night at 4 °C. The dialysed sample was centrifuged and applied onto a UnoQ column, precipitated and shock-frozen as described above (chapter 21.2.5). Quantitative TEV cleavage of Rpb1 was confirmed by SDS-PAGE (Figure 43).

26 CLONING, MUTAGENESIS AND PURIFICATION OF TFIIS

26.1 Cloning and mutagenesis

The TFIIS variants were expressed in *E. coli* as a fusion protein containing an amino-terminal hexahistidine tag, essentially as described (Awrey *et al.*, 1997). The DNA encoding for yeast TFIIS domains II and III (residues 131-309) was amplified by PCR from *S. cerevisiae* genomic DNA using primers IIS_fo and IIS_rev (Table 13), and subcloned into vector pET28a(+) (Novagen) *via* the restriction sites BamHI and NdeI. A double mutant D290A E291A was constructed by site-directed mutagenesis using two-step PCR with the primers DE_fo and DE_rev.

Table 13: Primers for cloning and mutagenesis of TFIIS

name	sequence (5' → 3')
IIS_fo	GGGAATTCATATGCCAAGAAATAGTAAGAACGATGG
IIS_rev	CGCGGATCCCTAAGAGAATTTCATCTGTTACC
DE_fo	GCAAACAAGATCTGCGGCTGCACCATTGACC
DE_rev	GGACAATGGTGCAGCCGCAGATCTTGTTC

DE→AA mutation

26.2 Expression and purification

buffer A

50 mM Hepes pH 7.5
 300 mM NaCl
 5% glycerol
 10 μM ZnCl₂
 1 x protease inhibitors
 10 mM β-mercaptoethanol

10 x protease inhibitors

3 mg/L leupeptin
 14 mg/L pepstatin A
 1.7 g/L PMSF
 3.3 g/L benzamidine
 dissolved in ethanol

For the expression of TFIIS variants, transformed BL21(DE3) RIL (Stratagene) cells were grown in LB medium, supplemented with chloramphenicol (30 mg/L) and kanamycin (30 mg/L), to an OD₆₀₀ of 0.7 and cooled down to 20 °C before expression was induced with 1 mM IPTG, and continued over night. For selenomethionine incorporation, the TFIIS variant was introduced to the methionine auxotroph *E. coli* strain B834(DE3) (Budisa *et al.*, 1995). Cells were grown in one litre of LB medium, supplemented with kanamycin (30 mg/L) to an OD₆₀₀ of 0.6, harvested by centrifugation (10 min, 5000 rpm, GS-3 rotor), and were resuspended in one litre of selenomethionine-containing minimal medium (Budisa *et al.*, 1995). The cell suspension was agitated until growth resumed (2-3 hours) at 20°C, before expression was induced with 1 mM IPTG, and continued over night.

Cells from 1 L of culture were lysed by sonication in 50 mL of buffer A. The lysate was cleared by centrifugation (30 min, 16000 rpm, SS34 rotor) and was applied to a Ni-NTA agarose column (1.5 mL; Quiagen). The column was washed with buffer A containing 500 mM NaCl, and the protein was eluted with a gradient of 0 mM to 500 mM imidazole in buffer A containing 500 mM NaCl. Peak fractions were diluted 5-fold in buffer A and loaded onto a Mono-S 5/50HR anion exchange column (Amersham), pre-equilibrated with buffer A containing 100 mM NaCl. The TFIIS variant was eluted over a total of 15 column volumes with a gradient of 100-500 mM NaCl in buffer A. This yielded two peaks, which both contained the TFIIS variant, as confirmed by mass spectroscopy. TFIIS in the two fractions shows the same apparent molecular weight in gel electrophoresis and gives rise to the same CD spectra. Fractions of both peaks were pooled, concentrated to 8 mg/mL, shock-frozen in liquid nitrogen and stored at -80°C.

27 PREPARATION OF POL II-NUCLEIC ACID COMPLEXES

Assembly buffer

50 mM Hepes pH 7.5
40 mM ammonium sulphate
5 µM ZnCl₂
5% glycerol
10 mM DTT

Pol II buffer

5 mM Hepes pH 7.25
40 mM ammonium sulphate
10 µM ZnCl₂
10 mM DTT

Endogenous yeast ten-subunit Pol II core and recombinant Rpb4/7 heterodimer were purified as described in chapters 21 and 22. DNA bubble-RNA complexes were annealed by mixing equimolar amounts of synthetic template DNA, non-template DNA (biomers.net) and RNA

oligonucleotides (biomers.net) (Table 6) in RNase-free TE buffer (Fluka) at a final concentration of 100 μ M. The mixture was heated to 90 °C in a preheated thermo block, and slowly cooling to 20 °C over night.

The Pol II-bubble-RNA complex was assembled by incubating core Pol II for 5 minutes at 20 °C with a 1.5-fold molar excess of the bubble-RNA construct in assembly buffer. To this complex, a 5-fold molar excess of recombinant Rpb4/7 was added, followed by incubation for another 10 minutes at 20 °C. The complex was purified by gel filtration (Superose 6 HR, Amersham, flowrate 0.3 mL/min) in Pol II buffer. The complex was concentrated to 3.5-4.5 mg/mL using a MembraSpin centrifugal concentrator (100 kDa cutoff). Prior to crystallisation, an additional amount of bubble-RNA complex was added to a final concentration of 2 μ M in order to ensure full occupancy. The light-sensitive bromine-derivatised template DNA contained 5-bromo-uracil at positions -4, -7 and -10 and was treated with minimal light exposure.

28 CRYSTAL GROWTH AND TREATMENT

Generally, crystals were grown at 20 °C with the hanging drop vapour diffusion method by mixing 2-4 μ L of sample solution with 1-2 μ L of reservoir solution. Crystals typically reached their maximum size after about 10 days. For cryoprotection, the mother solution was exchanged in five steps to mother solution containing 20, 40, 60, 80 and 100 % of the final glycerol concentration (Table 7). Incubation time between these steps was 1 hour. Subsequently, crystals in the final cryo-protecting solution were slowly cooled down to 4-8 °C in a styrofoam box, incubated for another 14-24 h, mounted into nylon loops (Hampton Research) and plunged into liquid nitrogen.

28.1 Preparation of substrate-containing Pol II cocrystals

Crystals containing additionally the substrate analogue GMPCPP (Guanosine-5'-[[α,β]-methylene]-triphosphate) were grown under modified Natrix condition #38 (Table 7) in the presence of 3 mM GMPCPP (Jena Biosciences). To ensure maximum occupancy, GMPCPP was also present in the last cryo solution at a concentration of 3 mM.

28.2 Preparation Pol II-TFIIS complex crystals

Crystals of the Pol II-TFIIS complex (Kettenberger *et al.*, 2003) were prepared by soaking preformed cryo-protected crystals of 12-subunit Pol II for 24 hours at 4 °C in the last cryo solution of the ammonium-sodium tartrate condition (Table 7), containing additionally 1 mg/mL of TFIIS. Crystals of the Pol II-bubble-RNA-TFIIS complex were obtained analogously by soaking crystals grown in the modified Natrix condition #38 (Table 7) for less than 12 hours in the last cryo solution, containing 1 mg/mL TFIIS double mutant D290A/E291A. Additionally, the RNA-derivative present in this complex is less prone to hydrolysis because it contains only thiophospho-diester groups in order to reduce cleavage by TFIIS.

29 FLUORESCENCE-BASED ASSAYS FOR COMPLEX CRYSTALS

29.1 Detection of nucleic acids in crystals

Mother solutions: **NgoMIV and its DNA complex**

4.0 M sodium formate

Core Pol II and its complexes

390 mM $(\text{NH}_4)_2\text{HPO}_4/\text{NaH}_2\text{PO}_4$ pH 6.0

16 % PEG-6000

50 mM dioxane

5 mM DTT

Complete 12-subunit Pol II

150 mM magnesium acetate

200 mM ammonium acetate

50 mM HEPES pH 7.0

5% PEG4000

Crystals were harvested in their respective mother solutions. To assure complete removal of non-specifically associated nucleic acids, crystals were back-soaked in harvest solution at 20° C overnight. To allow for direct comparison under identical conditions, crystals of free protein and the protein-nucleic acid complex were placed side by side on a glass spot plate in 50 µL of the harvest solution. To this drop, 50 µL of SYBR-Gold (Molecular Probes, Eugene/USA), diluted 1:5000 with harvest solution were added. The excitation and emission maxima of SYBR-Gold, 497 nm and 520 nm, respectively (Molecular Probes, www.probes.com), are close to the values for fluorescein (498 nm and 520 nm, respectively; Molecular Probes, www.probes.com).

Therefore, filters that are routinely used for fluorescein could be employed. Maximal fluorescence was typically reached within one hour, although prolonged incubation for up to 12 hours did not significantly change the results, nor did it lead to increased staining of crystals containing only protein. Fluorescence microscopic images were collected with a Zeiss Axioskop microscope at 50-fold magnification. Digital images were recorded with a VisiCam (Visitron Systems) black-and-white CCD camera and processed with ScionImage (Scion Corporation)

A time-course for quantitative kinetics was recorded on a confocal microscope (Leica DM IRE2) at 64-fold magnification. The time course was started by placing a back-soaked crystal (size ≈ 100 by 80 by $80 \mu\text{m}$) in $20 \mu\text{L}$ of harvest solution containing 0.01% of SYBR gold. Images of a $3.9 \mu\text{m}$ thick slice were recorded every 168 seconds. Integration times were chosen to stay within the dynamic range of the cameras and image processing was carried out with ScionImage.

29.2 Monitoring TFIIS soaking into RNA polymerase II crystals.

<u>Carbonate buffer</u>	<u>Storage buffer</u>
100 mM NaHCO_3	5 mM HEPES
100 mM NaCl	40 mM ammonium sulphate
5 mM DTT pH 9.0	5 mM DTT, pH 7.25

Cy3-labelled TFIIS was prepared following the manufacturer's protocol. To this end, 1.0 mg of purified TFIIS in 1 mL of carbonate buffer was incubated with one aliquot of Cy3 antibody labelling kit (Amersham) at 20°C for 1 hour. The reaction was quenched by adding $100 \mu\text{L}$ of 1 M TRIS pH 8, followed by incubation for another hour. Labelled protein was separated from unreacted dye by chromatography over a PD10 desalting column (Amersham), pre-equilibrated with storage buffer. Labelled protein was concentrated to 10 mg/mL . During concentration, no free dye was detectable in the flowthrough by absorption spectroscopy. Labelling resulted in a molar dye-to-protein ratio of 0.6. Cryo-protected crystals were then incubated for 18-24 hours in the final cryo solution containing additionally 2 mg/mL TFIIS. Only for the fluorescence assay, labelled TFIIS was used and fluorescence micrographs were recorded with the equipment described in chapter 29.1. The excitation and emission maxima of Cy3 are at 550 and 570 nm, respectively (Amersham Biosciences), thus a standard Cy3 filter pair was used here.

30 X-RAY ANALYSIS

30.1 Data collection

All data were collected at the protein crystallography beamline X06SA at the Swiss Light Source (SLS) in Villigen/Aargau. Because the crystals were radiation sensitive and diffraction power was rather weak, the best $I/\sigma(I)$ values possible had to be obtained with the lowest possible x-ray dose. To this end, fine-slicing ($0.25^\circ/\text{frame}$) was applied routinely, together with a beam focused on the detector rather than on the sample. Since the background level due to diffuse scattering decreases with $1/d^2$ (d = crystal to detector distance), the maximum reasonable crystal to detector distance was chosen for data collection. This strategy notably benefited from a larger Mar225 detector (Mar Research) with an active surface of 225 by 225 mm², which became available at the SLS from November 2004.

The total oscillation range of 90° , which yields maximum completeness was determined prior to data collection with MOSFLM (Powell, 1999). In case of premature decay, crystals were translated up to four times in order to irradiate an undamaged zone of the crystal. When only small crystals were available, partial datasets of up to five crystals could be merged at the step of scaling (Table 9, dataset 3). This was possible by selecting the combination of five out of eight crystals that were most isomorphous and gave rise to the best statistics of the merged dataset.

30.2 Data Reduction

Raw data were indexed and integrated with DENZO (vers. 1.97.2) (Otwinowski and Minor, 1997) and (from the time it was available, September 2003) with HKL2000. Three-dimensional profile fitting, implemented in HKL2000, slightly enhanced data quality, because it better accounts for mosaicities greater than the oscillation width. Due to radiation damage, mosaicity increased, typically from 0.3° to about 1.2° . Therefore it was refined in segments of 5 frames (HKL2000) or in segments of 20° (DENZO). Choosing a small spot size (0.18-0.22 mm diameter) is beneficial for stable refinement. Scalepack Vers. 1.97.2 or its derivatives Scalepack8m, Scalepackbig or Scalepack_ribo (Otwinowski and Minor, 1997) were used when necessary to allow for a high amount of measured reflections. A total of 10 cycles of global refinement was carried out and the maximum resolution was set to the shell in which R_{merge} was $< 32\%$ or $I/\sigma(I)$ was < 3 .

30.3 Molecular replacement and phasing

Intensity data were converted to structure factor amplitudes using scalepack2mtz (CCP4 4-5.0.2) (Collaborative Computational Project, 1994). Further conversion to CNS format was performed by mtz2various (Collaborative Computational Project, 1994). To obtain model phases, the 12-subunit apo-Pol II model (Armache *et al.*, 2004) was subjected to molecular replacement, using the program Phaser (Storoni *et al.*, 2004), and the orientation was further improved by rigid-body refinement with CNS (Brunger *et al.*, 1998). The clamp (residues 2-346 and 1395-1436 of Rpb1 and residues 1151-1224 of Rpb2) and the rest of the enzyme were defined as 2 independent rigid bodies in the case of the transcription bubble structure. Further subdivision into 6 rigid bodies was necessary for the TFIIIS complex. Initial examination of the electron density of the Pol II-TFIIIS complex showed that about one third of the Pol II mass underwent a conformational rearrangement. Therefore, following rigid bodies were defined: foot (Rpb1 877-1081), Rpb5-jaw (2-137), Rpb2 N-terminus (1-852), Rpb1-jaw (1092-1330), Rpb9-jaw (2-51), clamp (Rpb1 2-346 and 1395-1436, and Rpb2 1151-1224)

Rigid-body refinement was carried out in 2 cycles of 30 steps against e^2e^2 as a refinement target. Electron density for the TFIIIS complex could be further improved by solvent flipping (CNS), assuming a solvent content of 70 %. Solvent flipping could albeit not improve the quality of the electron density for the nucleic acid complexes, because initially observed density in the nucleic acid region was flattened out by solvent flipping.

30.4 Anomalous Maps

Anomalous Fourier maps were calculated by CNS using measured anomalous structure factor amplitude differences plus model phases derived as described above.

30.5 Model Building

30.5.1 Pol II-TFIIIS complex

After rigid body refinement and solvent flipping, clear electron density (3.8 Å resolution) was visible in a $2f_o - f_c$ map for the Pol II regions which moved upon TFIIIS binding and for parts of TFIIIS. Hence, these regions were fitted manually and by a second round of rigid-body refinement to match the electron density. Model maps calculated from this updated model gave

rise to consecutive difference electron density for most parts of the cocrystallised TFIIS variant. Since NMR models for domain II and III were available (Figure 27), these models were docked into the electron density maps as rigid bodies. Slight deviations from these NMR models were accounted for by manual rebuilding. A selenomethionine label in M182 of domain II served as a sequence marker. A newly ordered loop (trigger loop, residues 1078-1080 in Rpb1) was built manually into this electron density map, as well as parts of the TFIIS inter-domain linker, which is unstructured in solution (Figure 27).

30.5.2 Pol II-nucleic acid complexes

Modelling of nucleic acids started from the high-resolution coordinates of the minimal elongation complex (Gnatt *et al.*, 2001) and from coordinates of a canonical B-DNA. At the given resolution, the phosphate backbone and most bases showed clear density, thus adjustments to the initial base positions were made, followed by positional and B-factor refinement. During this refinement, base-planarity and Watson-Crick base-pairing of paired bases was restrained.

30.5.3 Refinement

Positional refinement was carried out starting with a set of coordinates and B-factors derived as described (Armache *et al.*, 2004). Starting B-factors for nucleic acids and for the refined part of TFIIS (residues 240-309) were set to 60 Å². The set of reflections not included in the refinement was the same than the set used to refine the apo-Pol II structure (Armache *et al.*, 2004). The maximum likelihood function was used as a target for 20 steps each of positional and B-factor refinement.

References

- Adelman, K., Marr, M. T., Werner, J., Saunders, A., Ni, Z., Andrulis, E. D., and Lis, J. T. (2005). Efficient release from promoter-proximal stall sites requires transcript cleavage factor TFIIS. *Mol Cell* *17*, 103-112.
- Armache, K. J., Kettenberger, H., and Cramer, P. (2003). Architecture of initiation-competent 12-subunit RNA polymerase II. *Proc Natl Acad Sci U S A* *100*, 6964-6968.
- Armache, K. J., Mitterweger, S., Meinhart, A., and Cramer, P. (2004). Structures of complete RNA polymerase II and its subcomplex Rpb4/7. *J Biol Chem*.
- Arthur, T. M., Anthony, L. C., and Burgess, R. R. (2000). Mutational analysis of beta '260-309, a sigma 70 binding site located on Escherichia coli core RNA polymerase. *J Biol Chem* *275*, 23113-23119.
- Arthur, T. M., and Burgess, R. R. (1998). Localization of a sigma70 binding site on the N terminus of the Escherichia coli RNA polymerase beta' subunit. *J Biol Chem* *273*, 31381-31387.
- Artsimovitch, I., Patlan, V., Sekine, S., Vassylyeva, M. N., Hosaka, T., Ochi, K., Yokoyama, S., and Vassylyev, D. G. (2004). Structural basis for transcription regulation by alarmone ppGpp. *Cell* *117*, 299-310.
- Asturias, F. J., Chang, W., Li, Y., and Kornberg, R. D. (1998). Electron crystallography of yeast RNA polymerase II preserved in vitreous ice. *Ultramicroscopy* *70*, 133-143.
- Asturias, F. J., Jiang, Y. W., Myers, L. C., Gustafsson, C. M., and Kornberg, R. D. (1999). Conserved structures of mediator and RNA polymerase II holoenzyme. *Science* *283*, 985-987.
- Asturias, F. J., and Kornberg, R. D. (1995). A novel method for transfer of two-dimensional crystals from the air/water interface to specimen grids. EM sample preparation/lipid-layer crystallization. *J Struct Biol* *114*, 60-66.
- Asturias, F. J., and Kornberg, R. D. (1999). Protein crystallization on lipid layers and structure determination of the RNA polymerase II transcription initiation complex. *J Biol Chem* *274*, 6813-6816.
- Asturias, F. J., Meredith, G. D., Poglitsch, C. L., and Kornberg, R. D. (1997). Two conformations of RNA polymerase II revealed by electron crystallography. *J Mol Biol* *272*, 536-540.
- Awrey, D. E., Shimasaki, N., Koth, C., Weilbaecher, R., Olmsted, V., Kazanis, S., Shan, X., Arellano, J., Arrowsmith, C. H., Kane, C. M., and Edwards, A. M. (1998). Yeast transcript elongation factor (TFIIS), structure and function. II: RNA polymerase binding, transcript cleavage, and read-through. *J Biol Chem* *273*, 22595-22605.
- Awrey, D. E., Weilbaecher, R. G., Hemming, S. A., Orlicky, S. M., Kane, C. M., and Edwards, A. M. (1997). Transcription elongation through DNA arrest sites. A multistep process involving both RNA polymerase II subunit RPB9 and TFIIS. *J Biol Chem* *272*, 14747-14754.
- Bar-Nahum, G., Epshtein, V., Ruckenstein, A. E., Rafikov, R., Mustaev, A., and Nudler, E. (2005). A Ratchet Mechanism of Transcription Elongation and Its Control. *Cell* *120*, 183-193.
- Batada, N. N., Westover, K. D., Bushnell, D. A., Levitt, M., and Kornberg, R. D. (2004). Diffusion of nucleoside triphosphates and role of the entry site to the RNA polymerase II active center. *Proc Natl Acad Sci U S A* *101*, 17361-17364.
- Beese, L. S., and Steitz, T. A. (1991). Structural basis for the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism. *Embo J* *10*, 25-33.
- Bengal, E., Flores, O., Krauskopf, A., Reinberg, D., and Aloni, Y. (1991). Role of the mammalian transcription factors IIF, IIS, and IIX during elongation by RNA polymerase II. *Mol Cell Biol* *11*, 1195-1206.

-
- Berben, G., Dumont, J., Gilliquet, V., Bolle, P. A., and Hilger, F. (1991). The YDp plasmids: a uniform set of vectors bearing versatile gene disruption cassettes for *Saccharomyces cerevisiae*. *Yeast* *7*, 475-477.
- Booth, V., Koth, C. M., Edwards, A. M., and Arrowsmith, C. H. (2000). Structure of a conserved domain common to the transcription factors TFIIIS, elongin A, and CRSP70. *J Biol Chem* *275*, 31266-31268.
- Bradsher, J. N., Jackson, K. W., Conaway, R. C., and Conaway, J. W. (1993a). RNA polymerase II transcription factor SIII. I. Identification, purification, and properties. *J Biol Chem* *268*, 25587-25593.
- Bradsher, J. N., Tan, S., McLaury, H. J., Conaway, J. W., and Conaway, R. C. (1993b). RNA polymerase II transcription factor SIII. II. Functional properties and role in RNA chain elongation. *J Biol Chem* *268*, 25594-25603.
- Brenner, S., Meselson, M., and Jacob, F. (1961). Unstable Intermediate Carrying Information from Genes to Ribosomes for Protein Synthesis. *Nature* *190*, 576-&.
- 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* (Pt 5), 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.
- Bushnell, D. A., Cramer, P., and Kornberg, R. D. (2002). Structural basis of transcription: alpha-amanitin-RNA polymerase II cocrystal at 2.8 Å resolution. *Proc Natl Acad Sci U S A* *99*, 1218-1222.
- Bushnell, D. A., and Kornberg, R. D. (2003). Complete, 12-subunit RNA polymerase II at 4.1 Å resolution: implications for the initiation of transcription. *Proc Natl Acad Sci U S A* *100*, 6969-6973.
- Bushnell, D. A., Westover, K. D., Davis, R. E., and Kornberg, R. D. (2004). Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Å resolution. *Science* *303*, 983-988.
- Cagas, P. M., and Corden, J. L. (1995). Structural Studies of a Synthetic Peptide Derived from the Carboxyl-Terminal Domain of RNA Polymerase II. *Proteins-Structure Function and Genetics* *21*, 149-160.
- Campbell, E. A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A., and Darst, S. A. (2001). Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell* *104*, 901-912.
- Campbell, E. A., Muzzin, O., Chlenov, M., Sun, J. L., Olson, C. A., Weinman, O., Trester-Zedlitz, M. L., and Darst, S. A. (2002). Structure of the bacterial RNA polymerase promoter specificity sigma subunit. *Mol Cell* *9*, 527-539.
- Carrington, J. C., and Dougherty, W. G. (1988). A viral cleavage site cassette: identification of amino acid sequences required for tobacco etch virus polyprotein processing. *Proc Natl Acad Sci U S A* *85*, 3391-3395.
- Carson, M. (1997). Ribbons. *Meth Enzym* *277*, 493-505.
- Carter, A. P., Clemons, W. M., Jr., Brodersen, D. E., Morgan-Warren, R. J., Hartsch, T., Wimberly, B. T., and Ramakrishnan, V. (2001). Crystal structure of an initiation factor bound to the 30S ribosomal subunit. *Science* *291*, 498-501.

-
- Chambers, R. S., Wang, B. Q., Burton, Z. F., and Dahmus, M. E. (1995). The activity of COOH-terminal domain phosphatase is regulated by a docking site on RNA polymerase II and by the general transcription factors IIF and IIB. *J Biol Chem* *270*, 14962-14969.
- 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 TFIIIS-like subunit and is important for transcription termination. *Genes Dev* *12*, 3857-3871.
- Cheetham, G. M., and Steitz, T. A. (1999). Structure of a transcribing T7 RNA polymerase initiation complex. *Science* *286*, 2305-2309.
- Chen, H. T., and Hahn, S. (2003). Binding of TFIIB to RNA polymerase II: Mapping the binding site for the TFIIB zinc ribbon domain within the preinitiation complex. *Mol Cell* *12*, 437-447.
- Chen, H. T., and Hahn, S. (2004). Mapping the location of TFIIB within the RNA polymerase II transcription preinitiation complex: a model for the structure of the PIC. *Cell* *119*, 169-180.
- Cho, H., Kim, T. K., Mancebo, H., Lane, W. S., Flores, O., and Reinberg, D. (1999). A protein phosphatase functions to recycle RNA polymerase II. *Genes Dev* *13*, 1540-1552.
- 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.
- Cipres-Palacin, G., and Kane, C. M. (1994). Cleavage of the nascent transcript induced by TFIIIS is insufficient to promote read-through of intrinsic blocks to elongation by RNA polymerase II. *Proc Natl Acad Sci U S A* *91*, 8087-8091.
- Collaborative Computational Project, N. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* *50*, 760-763.
- Conaway, J. W., Hanley, J. P., Garrett, K. P., and Conaway, R. C. (1991). Transcription Initiated by Rna Polymerase-II and Transcription Factors from Liver - Structure and Action of Transcription Factor-Epsilon and Factor-Tau. *Journal of Biological Chemistry* *266*, 7804-7811.
- Conaway, R. C., Kong, S. E., and Conaway, J. W. (2003). TFIIIS and GreB: Two Like-Minded Transcription Elongation Factors with Sticky Fingers. *Cell* *114*, 272-274.
- Connelly, S., and Manley, J. L. (1988). A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev* *2*, 440-452.
- Cook, P. R. (1999). The organization of replication and transcription. *Science* *284*, 1790-1795.
- Cosa, G., Focsaneanu, K. S., McLean, J. R. N., McNamee, J. P., and Scaiano, J. C. (2001). Photophysical properties of fluorescent DNA-dyes bound to single- and double-stranded DNA in aqueous buffered solution. *Photochemistry and Photobiology* *73*, 585-599.
- Craighead, J. L., Chang, W. H., and Asturias, F. J. (2002). Structure of yeast RNA polymerase II in solution: implications for enzyme regulation and interaction with promoter DNA. *Structure (Camb)* *10*, 1117-1125.
- Cramer, P. (2002a). Common structural features of nucleic acid polymerases. *Bioessays* *24*, 724-729.
- Cramer, P. (2002b). Multisubunit RNA polymerases. *Curr Opin Struct Biol* *12*, 89-97.
- Cramer, P., Bushnell, D. A., Fu, J., Gnat, A. L., Maier-Davis, B., Thompson, N. E., Burgess, R. R., Edwards, A. M., David, P. R., and Kornberg, R. D. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* *288*, 640-649.
- 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.
- Crick, F. (1970). Central Dogma of Molecular Biology. *Nature* *227*, 561.

-
- Darst, S. A., Edwards, A. M., Kubalek, E. W., and Kornberg, R. D. (1991). Three-dimensional structure of yeast RNA polymerase II at 16 Å resolution. *Cell* *66*, 121-128.
- Daube, S. S., and von Hippel, P. H. (1992). Functional Transcription Elongation Complexes from Synthetic RNA-DNA Bubble Duplexes. *Science* *258*, 1320-1324.
- Daube, S. S., and von Hippel, P. H. (1994). RNA displacement pathways during transcription from synthetic RNA-DNA bubble duplexes. *Biochemistry* *11*, 340-347.
- Davie, J. K., and Kane, C. M. (2000). Genetic interactions between TFIIIS and the Swi-Snf chromatin-remodeling complex. *Mol Cell Biol* *20*, 5960-5973.
- Deibert, M., Grazulis, S., Sasnauskas, G., Siksnys, V., and Huber, R. (2000). Structure of the tetrameric restriction endonuclease NgoMIV in complex with cleaved DNA. *Nature Structural Biology* *7*, 792-799.
- Dvir, A., Conaway, R. C., and Conaway, J. W. (1996). Promoter escape by RNA polymerase II. A role for an ATP cofactor in suppression of arrest by polymerase at promoter-proximal sites. *J Biol Chem* *271*, 23352-23356.
- Ederth, J., Artsimovitch, I., Isaksson, L. A., and Landick, R. (2002). The downstream DNA jaw of bacterial RNA polymerase facilitates both transcriptional initiation and pausing. *J Biol Chem* *277*, 37456-37463.
- Edwards, A. M., Darst, S. A., Feaver, W. J., Thompson, N. E., Burgess, R. R., and Kornberg, R. D. (1990). Purification and lipid-layer crystallization of yeast RNA polymerase II. *Proc Natl Acad Sci U S A* *87*, 2122-2126.
- 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.
- Eissenberg, J. C., Ma, J., Gerber, M. A., Christensen, A., Kennison, J. A., and Shilatifard, A. (2002). dELL is an essential RNA polymerase II elongation factor with a general role in development. *Proc Natl Acad Sci U S A* *99*, 9894-9899.
- Epshtein, V., Mustaev, A., Markovtsov, V., Bereshchenko, O., Nikiforov, V., and Goldfarb, A. (2002). Swing-gate model of nucleotide entry into the RNA polymerase active center. *Mol Cell* *10*, 623-634.
- Erie, D. A. (2002). The many conformational states of RNA polymerase elongation complexes and their roles in the regulation of transcription. *Biochim Biophys Acta* *1577*, 224-239.
- Erie, D. A., Hajiseyedi, O., Young, M. C., and von Hippel, P. H. (1993). Multiple RNA polymerase conformations and GreA: control of the fidelity of transcription. *Science* *262*, 867-873.
- Fabrega, C., Shen, V., Shuman, S., and Lima, C. D. (2003). Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Mol Cell* *11*, 1549-1561.
- Fish, R. N., and Kane, C. M. (2002). Promoting elongation with transcript cleavage stimulatory factors. *Biochim Biophys Acta* *1577*, 287-307.
- Flores, O., Lu, H., Killeen, M., Greenblatt, J., Burton, Z. F., and Reinberg, D. (1991). The Small Subunit of Transcription Factor IIF Recruits RNA Polymerase II into the Preinitiation Complex. *Proceedings of the National Academy of Sciences of the United States of America* *88*, 9999-10003.
- Flores, O., Maldonado, E., and Reinberg, D. (1989). Factors Involved in Specific Transcription by Mammalian RNA Polymerase-II - Factor-IIE and Factor-IIF Independently Interact with RNA Polymerase-II. *Journal of Biological Chemistry* *264*, 8913-8921.
- Forget, D., Langelier, M. F., Therien, C., Trinh, V., and Coulombe, B. (2004). Photo-cross-linking of a purified preinitiation complex reveals central roles for the RNA polymerase II mobile clamp and TFIIIE in initiation mechanisms. *Mol Cell Biol* *24*, 1122-1131.

-
- Fu, J., Gerstein, M., David, P. R., Gnatt, A. L., Bushnell, D. A., Edwards, A. M., and Kornberg, R. D. (1998). Repeated tertiary fold of RNA polymerase II and implications for DNA binding. *J Mol Biol* 280, 317-322.
- Fu, J., Gnatt, A. L., Bushnell, D. A., Jensen, G. J., Thompson, N. E., Burgess, R. R., David, P. R., and Kornberg, R. D. (1999). Yeast RNA polymerase II at 5 Å resolution. *Cell* 98, 799-810.
- Ganem, C., Devaux, F., Torchet, C., Jacq, C., Quevillon-Cheruel, S., Labesse, G., Facca, C., and Faye, G. (2003). Ssu72 is a phosphatase essential for transcription termination of snoRNAs and specific mRNAs in yeast. *Embo Journal* 22, 1588-1598.
- Garman, E. (2003). 'Cool' crystals: macromolecular cryocrystallography and radiation damage. *Current Opinion in Structural Biology* 13, 545-551.
- Garman, E., and Nave, C. (2002). Radiation damage to crystalline biological molecules: current view. *Journal of Synchrotron Radiation* 9, 327-328.
- Gnatt, A. (2002). Elongation by RNA polymerase II: structure-function relationship. *Biochim Biophys Acta* 1577, 175-190.
- Gnatt, A., Fu, J., and Kornberg, R. D. (1997). Formation and crystallization of yeast RNA polymerase II elongation complexes. *J Biol Chem* 272, 30799-30805.
- Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A., and Kornberg, R. D. (2001). Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* 292, 1876-1882.
- Gong, X. Q., Nedialkov, Y. A., and Burton, Z. F. (2004). Alpha-amanitin blocks translocation by human RNA polymerase II. *J Biol Chem* 279, 27422-27427.
- Goodrich, J. A., and Tjian, R. (1994). Transcription factors IIE and IIIH and ATP hydrolysis direct promoter clearance by RNA polymerase II. *Cell* 77, 145-156.
- Gregory, S. M., and Sweder, K. S. (2001). Deletion of the CSB homolog, RAD26, yields Spt(-) strains with proficient transcription-coupled repair. *Nucleic Acids Res* 29, 3080-3086.
- Grigoriev, A. (2003). On the number of protein-protein interactions in the yeast proteome. *Nucleic Acids Res* 31, 4157-4161.
- Gu, W., and Reines, D. (1995). Identification of a decay in transcription potential that results in elongation factor dependence of RNA polymerase II. *J Biol Chem* 270, 11238-11244.
- Guajardo, R., and Sousa, R. (1997). A model for the mechanism of polymerase translocation. *Journal of Molecular Biology* 265, 8-19.
- Hahn, S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol* 11, 394-403.
- Hall, B. D., and Spiegelman, S. (1961). Sequence complementarity of T2-DNA and T2-specific RNA. *Proc Natl Acad Sci U S A* 47, 137-163.
- Hampsey, M. (1998). Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol Mol Biol Rev* 62, 465-503.
- Hausner, W., Lange, U., and Musfeldt, M. (2000). Transcription factor S, a cleavage induction factor of the archaeal RNA polymerase. *J Biol Chem* 275, 12393-12399.
- Hemming, S. A., Jansma, D. B., MacGregor, P. F., Goryachev, A., Friesen, J. D., and Edwards, A. M. (2000). RNA polymerase II subunit Rpb9 regulates transcription elongation in vivo. *J Biol Chem* 275, 35506-35511.
- Hirose, Y., and Manley, J. L. (2000). RNA polymerase II and the integration of nuclear events. *Genes Dev* 14, 1415-1429.

-
- Holmes, S. F., and Erie, D. A. (2003). Downstream DNA sequence effects on transcription elongation. Allosteric binding of nucleoside triphosphates facilitates translocation via a ratchet motion. *J Biol Chem* 278, 35597-35608.
- Holstege, F. C., van der Vliet, P. C., and Timmers, H. T. (1996). Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIIH. *Embo J* 15, 1666-1677.
- Horikoshi, M., Sekimizu, K., and Natori, S. (1984). Analysis of the stimulatory factor of RNA polymerase II in the initiation and elongation complex. *J Biol Chem* 259, 608-611.
- Huang, Y., Intine, R. V., Mozlin, A., Hasson, S., and Maraia, R. J. (2005). Mutations in the RNA polymerase III subunit Rpc11p that decrease RNA 3' Cleavage activity increase 3'-terminal oligo(U) length and La-dependent tRNA processing. *Mol Cell Biol* 25, 621-636.
- Hubert, J. C., Guyonvarch, A., Kammerer, B., Exinger, F., Liljelund, P., and Lacroute, F. (1983). Complete sequence of a eukaryotic regulatory gene. *Embo J* 2, 2071-2073.
- Hurst, H. C. (1996). Transcription factors as drug targets in cancer. *Eur J Cancer* 32A, 1857-1863.
- Hurwitz, J., Bresler, A., and Diring, R. (1961). The enzymic incorporation of ribonucleotides into polyribonucleotides and the effect of DNA. *Biochemical and Biophysical Research Communications* 3, 15-19.
- Iborra, F. J., Pombo, A., Jackson, D. A., and Cook, P. R. (1996). Active RNA polymerases are localized within discrete transcription "factories" in human nuclei. *J Cell Sci* 109 (Pt 6), 1427-1436.
- Ito, H., Fukuda, Y., Murata, K., and Kimura, A. (1983). Transformation of intact yeast cells treated with alkali cations. *J Bacteriol* 153, 163-168.
- Ivanov, D., Kwak, Y. T., Guo, J., and Gaynor, R. B. (2000). Domains in the SPT5 protein that modulate its transcriptional regulatory properties. *Mol Cell Biol* 20, 2970-2983.
- Izban, M. G., and Luse, D. S. (1992a). Factor-stimulated RNA polymerase II transcribes at physiological elongation rates on naked DNA but very poorly on chromatin templates. *J Biol Chem* 267, 13647-13655.
- Izban, M. G., and Luse, D. S. (1992b). The RNA polymerase II ternary complex cleaves the nascent transcript in a 3'-5' direction in the presence of elongation factor SII. *Genes & Development* 6, 1342-1356.
- Izban, M. G., and Luse, D. S. (1993a). The increment of SII-facilitated transcript cleavage varies dramatically between elongation competent and incompetent RNA polymerase II ternary complexes. *J Biol Chem* 268, 12874-12885.
- Izban, M. G., and Luse, D. S. (1993b). SII-facilitated transcript cleavage in RNA polymerase II complexes stalled early after initiation occurs in primarily dinucleotide increments. *J Biol Chem* 268, 12864-12873.
- Jacob, F. M., J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* 3, 318-356.
- Jeon, C., and Agarwal, K. (1996). Fidelity of RNA polymerase II transcription controlled by elongation factor TFIIS. *Proc Natl Acad Sci U S A* 93, 13677-13682.
- Jeon, C., Yoon, H., and Agarwal, K. (1994). The transcription factor TFIIS zinc ribbon dipeptide Asp-Glu is critical for stimulation of elongation and RNA cleavage by RNA polymerase II. *Proc Natl Acad Sci U S A* 91, 9106-9110.
- Jeruzalmi, D., and Steitz, T. A. (1998). Structure of T7 RNA polymerase complexed to the transcriptional inhibitor T7 lysozyme. *Embo J* 17, 4101-4113.
- Jiang, M., Ma, N., Vassilyev, D. G., and McAllister, W. T. (2004). RNA displacement and resolution of the transcription bubble during transcription by T7 RNA polymerase. *Mol Cell* 15, 777-788.

-
- Jiang, Y., and Gralla, J. D. (1993). Uncoupling of initiation and reinitiation rates during HeLa RNA polymerase II transcription in vitro. *Mol Cell Biol* *13*, 4572-4577.
- 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.
- Joyce, C. M., and Steitz, T. A. (1994). Function and structure relationships in DNA polymerases. *Annu Rev Biochem* *63*, 777-822.
- Kamada, K., Roeder, R. G., and Burley, S. K. (2003). Molecular mechanism of recruitment of TFIIF-associating RNA polymerase C-terminal domain phosphatase (FCP1) by transcription factor IIF. *Proc Natl Acad Sci U S A* *100*, 2296-2299.
- Kamenski, T., Heilmeyer, S., Meinhart, A., and Cramer, P. (2004). Structure and mechanism of RNA polymerase II CTD phosphatases. *Mol Cell* *15*, 399-407.
- Kanai, A., Kuzuhara, T., Sekimizu, K., and Natori, S. (1991). Heterogeneity and tissue-specific expression of eukaryotic transcription factor S-II-related protein mRNA. *J Biochem (Tokyo)* *109*, 674-677.
- Kettenberger, H., Armache, K. J., and Cramer, P. (2003). Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell* *114*, 347-357.
- Kettenberger, H., Armache, K. J., and Cramer, P. (2004). Complete RNA Polymerase II Elongation Complex Structure and Its Interactions with NTP and TFIIS. *Mol Cell* *16*, 955-965.
- Kim, J. B., and Sharp, P. A. (2001). Positive transcription elongation factor B phosphorylates hSPT5 and RNA polymerase II carboxyl-terminal domain independently of cyclin-dependent kinase-activating kinase. *J Biol Chem* *276*, 12317-12323.
- Kim, M., Krogan, N. J., Vasiljeva, L., Rando, O. J., Nedeá, E., Greenblatt, J. F., and Buratowski, S. (2004). The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* *432*, 517-522.
- Kim, T. K., Ebright, R. H., and Reinberg, D. (2000). Mechanism of ATP-dependent promoter melting by transcription factor IIIH. *Science* *288*, 1418-1422.
- Kireeva, M. L., Komissarova, N., and Kashlev, M. (2000a). Overextended RNA:DNA hybrid as a negative regulator of RNA polymerase II processivity. *J Mol Biol* *299*, 325-335.
- Kireeva, M. L., Komissarova, N., Waugh, D. S., and Kashlev, M. (2000b). 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.
- Koch, M. H., Vachette, P., and Svergun, D. I. (2003). Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Q Rev Biophys* *36*, 147-227.
- Komissarova, N., Kireeva, M. L., Becker, J., Sidorenkov, I., and Kashlev, M. (2003). Engineering of elongation complexes of bacterial and yeast RNA polymerases. *Methods Enzymol* *371*, 233-251.
- Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A., and Darst, S. A. (2000). A structural model of transcription elongation. *Science* *289*, 619-625.
- Koulich, D., Orlova, M., Malhotra, A., Sali, A., Darst, S. A., and Borukhov, S. (1997). Domain organization of *Escherichia coli* transcript cleavage factors GreA and GreB. *J Biol Chem* *272*, 7201-7210.
- Koyama, H., Ito, T., Nakanishi, T., Kawamura, N., and Sekimizu, K. (2003). Transcription elongation factor S-II maintains transcriptional fidelity and confers oxidative stress resistance. *Genes Cells* *8*, 779-788.
- Krishnamurthy, S., He, X. Y., Reyes-Reyes, M., Moore, C., and Hampsey, M. (2004). Ssu72 is an RNA polymerase II CTD phosphatase. *Molecular Cell* *14*, 387-394.

-
- Kumar, K. P., Akoulitchev, S., and Reinberg, D. (1998). Promoter-proximal stalling results from the inability to recruit transcription factor IIIH to the transcription complex and is a regulated event. *Proc Natl Acad Sci U S A* *95*, 9767-9772.
- Landick, R. (2004). Active-site dynamics in RNA polymerases. *Cell* *116*, 351-353.
- Lassar, A. B., Martin, P. L., and Roeder, R. G. (1983). Transcription of class III genes: formation of preinitiation complexes. *Science* *222*, 740-748.
- Li, Y., and Kornberg, R. D. (1994). Interplay of Positive and Negative Effectors in Function of the C-Terminal Repeat Domain of Rna-Polymerase-II. *Proceedings of the National Academy of Sciences of the United States of America* *91*, 2362-2366.
- Licht, C. L., Stevnsner, T., and Bohr, V. A. (2003). Cockayne syndrome group B cellular and biochemical functions. *Am J Hum Genet* *73*, 1217-1239.
- Lin, P. S., Marshall, N. F., and Dahmus, M. E. (2002). CTD phosphatase: role in RNA polymerase II cycling and the regulation of transcript elongation. *Prog Nucleic Acid Res Mol Biol* *72*, 333-365.
- Lodish, H. F. (1996). *Molekulare Zellbiologie*, 2 edn (New York, Walter de Gruyter).
- Logan, J., Falck-Pedersen, E., Darnell, J. E., Jr., and Shenk, T. (1987). A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. *Proc Natl Acad Sci U S A* *84*, 8306-8310.
- Luo, W., and Bentley, D. (2004). A ribonucleolytic rat torpedo RNA polymerase II. *Cell* *119*, 911-914.
- Markovtsov, V., Mustaev, A., and Goldfarb, A. (1996). Protein-RNA interactions in the active center of transcription elongation complex. *Proc Natl Acad Sci U S A* *93*, 3221-3226.
- Marshall, N. F., and Price, D. H. (1992). Control of formation of two distinct classes of RNA polymerase II elongation complexes. *Mol Cell Biol* *12*, 2078-2090.
- Meinhart, A., Blobel, J., and Cramer, P. (2003a). An extended winged helix domain in general transcription factor E/IIIE alpha. *J Biol Chem* *278*, 48267-48274.
- Meinhart, A., and Cramer, P. (2004). Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature* *430*, 223-226.
- Meinhart, A., Silberzahn, T., and Cramer, P. (2003b). The mRNA transcription/processing factor Ssu72 is a potential tyrosine phosphatase. *J Biol Chem* *278*, 15917-15921.
- Meredith, G. D., Chang, W. H., Li, Y., Bushnell, D. A., Darst, S. A., and Kornberg, R. D. (1996). The C-terminal domain revealed in the structure of RNA polymerase II. *Journal of Molecular Biology* *258*, 413-419.
- Minakhin, L., Bhagat, S., Brunning, A., Campbell, E. A., Darst, S. A., Ebricht, R. H., and Severinov, K. (2001). Bacterial RNA polymerase subunit omega and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. *Proc Natl Acad Sci U S A* *98*, 892-897.
- Morillon, A., Karabetsov, N., O'Sullivan, J., Kent, N., Proudfoot, N., and Mellor, J. (2003). Isw1 chromatin remodeling ATPase coordinates transcription elongation and termination by RNA polymerase II. *Cell* *115*, 425-435.
- Morin, P. E., Awrey, D. E., Edwards, A. M., and Arrowsmith, C. H. (1996). Elongation factor TFIIS contains three structural domains: solution structure of domain II. *Proc Natl Acad Sci U S A* *93*, 10604-10608.
- Mukherjee, K., and Chatterji, D. (1997). Studies on the omega subunit of Escherichia coli RNA polymerase--its role in the recovery of denatured enzyme activity. *Eur J Biochem* *247*, 884-889.
- Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O., and Darst, S. A. (2002a). Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* *296*, 1285-1290.

-
- Murakami, K. S., Masuda, S., and Darst, S. A. (2002b). Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science* 296, 1280-1284.
- Murray, J., and Garman, E. (2002). Investigation of possible free-radical scavengers and metrics for radiation damage in protein cryocrystallography. *Journal of Synchrotron Radiation* 9, 347-354.
- Nakanishi, Y., Horikoshi, M., Sekimizu, K., and Natori, S. (1985). [Regulation of eukaryotic transcription by RNA polymerase II stimulatory factors]. *Tanpakushitsu Kakusan Koso* 30, 1659-1672.
- Nedialkov, Y. A., Gong, X. Q., Hovde, S. L., Yamaguchi, Y., Handa, H., Geiger, J. H., Yan, H., and Burton, Z. F. (2003). NTP-driven translocation by human RNA polymerase II. *J Biol Chem* 278, 18303-18312.
- Nguyen, B. D., Abbott, K. L., Potempa, K., Kobor, M. S., Archambault, J., Greenblatt, J., Legault, P., and Omichinski, J. G. (2003). NMR structure of a complex containing the TFIIF subunit RAP74 and the RNA polymerase II carboxyl-terminal domain phosphatase FCP1. *Proc Natl Acad Sci U S A* 100, 5688-5693.
- Ni, Z., Schwartz, B. E., Werner, J., Suarez, J. R., and Lis, J. T. (2004). Coordination of transcription, RNA processing, and surveillance by P-TEFb kinase on heat shock genes. *Mol Cell* 13, 55-65.
- Nouraini, S., Archambault, J., and Friesen, J. D. (1996). Rpo26p, a subunit common to yeast RNA polymerases, is essential for the assembly of RNA polymerases I and II and for the stability of the largest subunits of these enzymes. *Mol Cell Biol* 16, 5985-5996.
- Nudler, E. (1999). Transcription elongation: structural basis and mechanisms. *J Mol Biol* 288, 1-12.
- Nudler, E., Avetisova, E., Markovtsov, V., and Goldfarb, A. (1996). Transcription processivity: Protein-DNA interactions holding together the elongation complex. *Science* 273, 211-217.
- Nudler, E., Gusarov, I., Avetisova, E., Kozlov, M., and Goldfarb, A. (1998). Spatial organization of transcription elongation complex in *Escherichia coli*. *Science* 281, 424-428.
- Nudler, E., Mustaev, A., Lukhtanov, E., and Goldfarb, A. (1997). The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell* 89, 33-41.
- Olmsted, V. K., Awrey, D. E., Koth, C., Shan, X., Morin, P. E., Kazanis, S., Edwards, A. M., and Arrowsmith, C. H. (1998). Yeast transcript elongation factor (TFIIS), structure and function. I: NMR structural analysis of the minimal transcriptionally active region. *J Biol Chem* 273, 22589-22594.
- Opalka, N., Chlenov, M., Chacon, P., Rice, W. J., Wriggers, W., and Darst, S. A. (2003). Structure and function of the transcription elongation factor GreB bound to bacterial RNA polymerase. *Cell* 114, 335-345.
- Orphanides, G., Lagrange, T., and Reinberg, D. (1996). The general transcription factors of RNA polymerase II. *Genes Dev* 10, 2657-2683.
- Orphanides, G., and Reinberg, D. (2000). RNA polymerase II elongation through chromatin. *Nature* 407, 471-475.
- Orphanides, G., and Reinberg, D. (2002). A unified theory of gene expression. *Cell* 108, 439-451.
- Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Macromolecular Crystallography, Pt A* 276, 307-326.
- Ouhammouch, M., Werner, F., Weinzierl, R. O., and Geiduschek, E. P. (2004). A fully recombinant system for activator-dependent archaeal transcription. *J Biol Chem* 279, 51719-51721.
- Pal, M., and Luse, D. S. (2003). The initiation-elongation transition: lateral mobility of RNA in RNA polymerase II complexes is greatly reduced at +8/+9 and absent by +23. *Proc Natl Acad Sci U S A* 100, 5700-5705.

-
- Palangat, M., Hittinger, C. T., and Landick, R. (2004). Downstream DNA selectively affects a paused conformation of human RNA polymerase II. *J Mol Biol* *341*, 429-442.
- Palangat, M., and Landick, R. (2001). Roles of RNA:DNA hybrid stability, RNA structure, and active site conformation in pausing by human RNA polymerase II. *J Mol Biol* *311*, 265-282.
- Palangat, M., Meier, T. I., Keene, R. G., and Landick, R. (1998). Transcriptional pausing at +62 of HIV-1 nascent RNA modulates formation of the TAR RNA structure. *Mol Cell* *1*, 1033-1042.
- Parks, T. D., Smith, H. A., and Dougherty, W. G. (1992). Cleavage profiles of tobacco etch virus (TEV)-derived substrates mediated by precursor and processed forms of the TEV NIa proteinase. *J Gen Virol* *73* (Pt 1), 149-155.
- Poglitsch, C. L., Meredith, G. D., Gnatt, A. L., Jensen, G. J., Chang, W. H., Fu, J., and Kornberg, R. D. (1999). Electron crystal structure of an RNA polymerase II transcription elongation complex. *Cell* *98*, 791-798.
- Pokholok, D. K., Hannett, N. M., and Young, R. A. (2002). Exchange of RNA polymerase II initiation and elongation factors during gene expression in vivo. *Mol Cell* *9*, 799-809.
- Powell, H. R. (1999). The Rossmann Fourier autoindexing algorithm in MOSFLM. *Acta Crystallogr D Biol Crystallogr* *55* (Pt 10), 1690-1695.
- Powell, W., Bartholomew, B., and Reines, D. (1996). Elongation factor SII contacts the 3'-end of RNA in the RNA polymerase II elongation complex. *J Biol Chem* *271*, 22301-22304.
- Proudfoot, N. (2004). New perspectives on connecting messenger RNA 3' end formation to transcription. *Curr Opin Cell Biol* *16*, 272-278.
- Proudfoot, N. J. (1989). How RNA polymerase II terminates transcription in higher eukaryotes. *Trends Biochem Sci* *14*, 105-110.
- Proudfoot, N. J., Furger, A., and Dye, M. J. (2002). Integrating mRNA processing with transcription. *Cell* *108*, 501-512.
- Qian, X., Gozani, S. N., Yoon, H., Jeon, C. J., Agarwal, K., and Weiss, M. A. (1993a). Novel zinc finger motif in the basal transcriptional machinery: three-dimensional NMR studies of the nucleic acid binding domain of transcriptional elongation factor TFIIS. *Biochemistry* *32*, 9944-9959.
- Qian, X., Jeon, C., Yoon, H., Agarwal, K., and Weiss, M. A. (1993b). Structure of a new nucleic-acid-binding motif in eukaryotic transcriptional elongation factor TFIIS. *Nature* *365*, 277-279.
- Ranish, J. A., Yudkovsky, N., and Hahn, S. (1999). Intermediates in formation and activity of the RNA polymerase II preinitiation complex: holoenzyme recruitment and a postrecruitment role for the TATA box and TFIIB. *Genes Dev* *13*, 49-63.
- Rappaport, J., Reinberg, D., Zandomeni, R., and Weinmann, R. (1987). Purification and functional characterization of transcription factor SII from calf thymus. Role in RNA polymerase II elongation. *J Biol Chem* *262*, 5227-5232.
- Reeder, R. H., and Lang, W. (1994). The mechanism of transcription termination by RNA polymerase I. *Mol Microbiol* *12*, 11-15.
- Reines, D., Ghanouni, P., Li, Q. Q., and Mote, J., Jr. (1992). The RNA polymerase II elongation complex. Factor-dependent transcription elongation involves nascent RNA cleavage. *J Biol Chem* *267*, 15516-15522.
- Rhodes, G., and Chamberlin, M. J. (1974). Ribonucleic acid chain elongation by *Escherichia coli* ribonucleic acid polymerase. I. Isolation of ternary complexes and the kinetics of elongation. *J Biol Chem* *249*, 6675-6683.
- Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* *17*, 1030-1032.

-
- Roeder, R. G. (1974a). Multiple forms of deoxyribonucleic acid-dependent ribonucleic acid polymerase in *Xenopus laevis*. Isolation and partial characterization. *J Biol Chem* *249*, 241-248.
- Roeder, R. G. (1974b). Multiple forms of deoxyribonucleic acid-dependent ribonucleic acid polymerase in *Xenopus laevis*. Levels of activity during oocyte and embryonic development. *J Biol Chem* *249*, 249-256.
- Rudd, M. D., Izban, M. G., and Luse, D. S. (1994). The active site of RNA polymerase II participates in transcript cleavage within arrested ternary complexes. *Proc Natl Acad Sci U S A* *91*, 8057-8061.
- Rudd, M. D., and Luse, D. S. (1996). Amanitin greatly reduces the rate of transcription by RNA polymerase II ternary complexes but fails to inhibit some transcript cleavage modes. *J Biol Chem* *271*, 21549-21558.
- Ruet, A., Sentenac, A., Fromageot, P., Winsor, B., and Lacroute, F. (1980). A mutation of the B220 subunit gene affects the structural and functional properties of yeast RNA polymerase B in vitro. *J Biol Chem* *255*, 6450-6455.
- Sakurai, H., Mitsuzawa, H., Kimura, M., and Ishihama, A. (1999). The rpb4 subunit of fission yeast *Schizosaccharomyces pombe* RNA polymerase II is essential for cell viability and similar in structure to the corresponding subunits of higher eukaryotes. *Molecular and Cellular Biology* *19*, 7511-7518.
- Sambrook, J. R. D. W. (2001). *Molecular Cloning*, 3 edn (Cold Spring Harbour, NY, Cold Spring Harbour Laboratory Press).
- Scheffzek, K., Ahmadian, M. R., Kabsch, W., Wiesmuller, L., Lautwein, A., Schmitz, F., and Wittinghofer, A. (1997). The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants. *Science* *277*, 333-338.
- Schwartz, L. B., and Roeder, R. G. (1974). Purification and subunit structure of deoxyribonucleic acid-dependent ribonucleic acid polymerase I from the mouse myeloma, MOPC 315. *J Biol Chem* *249*, 5898-5906.
- Schwartz, L. B., Sklar, V. E., Jaehning, J. A., Weinmann, R., and Roeder, R. G. (1974). Isolation and partial characterization of the multiple forms of deoxyribonucleic acid-dependent ribonucleic acid polymerase in the mouse myeloma, MOPC 315. *J Biol Chem* *249*, 5889-5897.
- Sekimizu, K., Kobayashi, N., Mizuno, D., and Natori, S. (1976). Purification of a factor from Ehrlich ascites tumor cells specifically stimulating RNA polymerase II. *Biochemistry* *15*, 5064-5070.
- Selby, C. P., and Sancar, A. (1997). Human transcription-repair coupling factor CSB/ERCC6 is a DNA-stimulated ATPase but is not a helicase and does not disrupt the ternary transcription complex of stalled RNA polymerase II. *J Biol Chem* *272*, 1885-1890.
- Severinov, K., Mooney, R., Darst, S. A., and Landick, R. (1997). Tethering of the large subunits of *Escherichia coli* RNA polymerase. *J Biol Chem* *272*, 24137-24140.
- Shaw, R. J., Bonawitz, N. D., and Reines, D. (2002). Use of an in vivo reporter assay to test for transcriptional and translational fidelity in yeast. *J Biol Chem* *277*, 24420-24426.
- Shermoen, A. W., and Ofarrell, P. H. (1991). Progression of the Cell-Cycle through Mitosis Leads to Abortion of Nascent Transcripts. *Cell* *67*, 303-310.
- Shilatifard, A. (2004). Transcriptional elongation control by RNA polymerase II: a new frontier. *Biochim Biophys Acta* *1677*, 79-86.
- Shilatifard, A., Conaway, R. C., and Conaway, J. W. (2003). The RNA polymerase II elongation complex. *Annu Rev Biochem* *72*, 693-715.
- Shilatifard, A., Haque, D., Conaway, R. C., and Conaway, J. W. (1997). Structure and function of RNA polymerase II elongation factor ELL. Identification of two overlapping ELL functional domains

-
- that govern its interaction with polymerase and the ternary elongation complex. *J Biol Chem* 272, 22355-22363.
- Shilatifard, A., Lane, W. S., Jackson, K. W., Conaway, R. C., and Conaway, J. W. (1996). An RNA polymerase II elongation factor encoded by the human ELL gene. *Science* 271, 1873-1876.
- Shimasaki, N. B., and Kane, C. M. (2000). Structural basis for the species-specific activity of TFIIS. *J Biol Chem* 275, 36541-36549.
- Sijbrandi, R., Fiedler, U., and Timmers, H. T. M. (2002). RNA polymerase II complexes in the very early phase of transcription are not susceptible to TFIIS-induced exonucleolytic cleavage. *Nucleic Acids Research* 30, 2290-2298.
- Sims, R. J., 3rd, Belotserkovskaya, R., and Reinberg, D. (2004). Elongation by RNA polymerase II: the short and long of it. *Genes Dev* 18, 2437-2468.
- Sluder, A. E., Greenleaf, A. L., and Price, D. H. (1989). Properties of a Drosophila RNA polymerase II elongation factor. *J Biol Chem* 264, 8963-8969.
- Sosunov, V., Sosunova, E., Mustaev, A., Bass, I., Nikiforov, V., and Goldfarb, A. (2003). Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerase. *Embo J* 22, 2234-2244.
- Sousa, R., Chung, Y. J., Rose, J. P., and Wang, B. C. (1993). Crystal structure of bacteriophage T7 RNA polymerase at 3.3 Å resolution. *Nature* 364, 593-599.
- Stebbins, C. E., Borukhov, S., Orlova, M., Polyakov, A., Goldfarb, A., and Darst, S. A. (1995). Crystal structure of the GreA transcript cleavage factor from *Escherichia coli*. *Nature* 373, 636-640.
- Steitz, T. A. (1993). DNA-Dependent and RNA-Dependent DNA-Polymerases. *Current Opinion in Structural Biology* 3, 31-38.
- Steitz, T. A. (1998). A mechanism for all polymerases. *Nature* 391, 231-232.
- Steitz, T. A. (2004). The structural basis of the transition from initiation to elongation phases of transcription, as well as translocation and strand separation, by T7 RNA polymerase. *Curr Opin Struct Biol* 14, 4-9.
- Storoni, L. C., McCoy, A. J., and Read, R. J. (2004). Likelihood-enhanced fast rotation functions. *Acta Crystallogr D Biol Crystallogr* 60, 432-438.
- Surratt, C. K., Milan, S. C., and Chamberlin, M. J. (1991). Spontaneous cleavage of RNA in ternary complexes of *Escherichia coli* RNA polymerase and its significance for the mechanism of transcription. *Proc Natl Acad Sci U S A* 88, 7983-7987.
- Svetlov, V., Vassilyev, D. G., and Artsimovitch, I. (2004). Discrimination against deoxyribonucleotide substrates by bacterial RNA polymerase. *J Biol Chem* 279, 38087-38090.
- Sweetser, D., Nonet, M., and Young, R. A. (1987). Prokaryotic and eukaryotic RNA polymerases have homologous core subunits. *Proc Natl Acad Sci U S A* 84, 1192-1196.
- Tahirov, T. H., Temiakov, D., Anikin, M., Patlan, V., McAllister, W. T., Vassilyev, D. G., and Yokoyama, S. (2002). Structure of a T7 RNA polymerase elongation complex at 2.9 Å resolution. *Nature* 420, 43-50.
- Takagi, Y., Conaway, J. W., and Conaway, R. C. (1995). A Novel Activity Associated with Rna-Polymerase-II Elongation-Factor Siii - Siii Directs Promoter-Independent Transcription Initiation by Rna-Polymerase-II in the Absence of Initiation-Factors. *Journal of Biological Chemistry* 270, 24300-24305.
- Tan, S. (2001). A modular polycistronic expression system for overexpressing protein complexes in *Escherichia coli*. *Protein Expr Purif* 21, 224-234.
- Tantin, D., Kansal, A., and Carey, M. (1997). Recruitment of the putative transcription-repair coupling factor CSB/ERCC6 to RNA polymerase II elongation complexes. *Mol Cell Biol* 17, 6803-6814.

-
- Teixeira, A., Tahiri-Alaoui, A., West, S., Thomas, B., Ramadass, A., Martianov, I., Dye, M., James, W., Proudfoot, N. J., and Akoulitchev, A. (2004). Autocatalytic RNA cleavage in the human beta-globin pre-mRNA promotes transcription termination. *Nature* *432*, 526-530.
- Temiakov, D., Patlan, V., Anikin, M., McAllister, W. T., Yokoyama, S., and Vassilyev, D. G. (2004). Structural basis for substrate selection by $\tau 7$ RNA polymerase. *Cell* *116*, 381-391.
- Thirman, M. J., Levitan, D. A., Kobayashi, H., Simon, M. C., and Rowley, J. D. (1994). Cloning of ELL, a gene that fuses to MLL in a t(11;19)(q23;p13.1) in acute myeloid leukemia. *Proc Natl Acad Sci U S A* *91*, 12110-12114.
- Thomas, M., Chedin, S., Carles, C., Riva, M., Famulok, M., and Sentenac, A. (1997). Selective targeting and inhibition of yeast RNA polymerase II by RNA aptamers. *J Biol Chem* *272*, 27980-27986.
- Thomas, M. J., Platas, A. A., and Hawley, D. K. (1998). Transcriptional fidelity and proofreading by RNA polymerase II. *Cell* *93*, 627-637.
- Thompson, N. E., and Burgess, R. R. (1996). Immunoaffinity purification of RNA polymerase II and transcription factors using polyol-responsive monoclonal antibodies. *Methods Enzymol* *274*, 513-526.
- Thummel, C. S., Burtis, K. C., and Hogness, D. S. (1990). Spatial and Temporal Patterns of E74 Transcription During *Drosophila* Development. *Cell* *61*, 101-111.
- 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.
- Tschochner, H. (1996). A novel RNA polymerase I-dependent RNase activity that shortens nascent transcripts from the 3' end. *Proceedings of the National Academy of Sciences of the United States of America* *93*, 12914-12919.
- Tuma, R. S., Beaudet, M. P., Jin, X. K., Jones, L. J., Cheung, C. Y., Yue, S., and Singer, V. L. (1999). Characterization of SYBR gold nucleic acid gel stain: A dye optimized for use with 300-nm ultraviolet transilluminators. *Analytical Biochemistry* *268*, 278-288.
- Ucker, D. S., and Yamamoto, K. R. (1984). Early Events in the Stimulation of Mammary-Tumor Virus-Rna Synthesis by Glucocorticoids - Novel Assays of Transcription Rates. *Journal of Biological Chemistry* *259*, 7416-7420.
- Van Mullem, V., Landrieux, E., Vandenhaute, J., and Thuriaux, P. (2002). Rpa12p, a conserved RNA polymerase I subunit with two functional domains. *Molecular Microbiology* *43*, 1105-1113.
- Vassilyev, D. G., Sekine, S., Laptenko, O., Lee, J., Vassilyeva, M. N., Borukhov, S., and Yokoyama, S. (2002). Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* *417*, 712-719.
- Verdecia, M. A., Huang, H. K., Dutil, E., Kaiser, D. A., Hunter, T., and Noel, J. P. (2000). Structure of the human anti-apoptotic protein survivin reveals a dimeric arrangement. *Nature Structural Biology* *7*, 602-608.
- Wang, D., and Hawley, D. K. (1993). Identification of a 3' - 5' exonuclease activity associate with human RNA polymerase II. *Proc Natl Acad Sci USA* *90*, 843-847.
- Wang, H., and Oster, G. (2002). Ratchets, power strokes, and molecular motors. *Applied Physics a- Materials Science & Processing* *75*, 315-323.
- Weillbaecher, 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*.
- Weiss, S. B., and Nakamoto, T. (1961). Net synthesis of ribonucleic acid with a microbial enzyme requiring deoxyribonucleic acid and four ribonucleoside triphosphates. *J Biol Chem* *236*, PC18-20.
- Werner, F., and Weinzierl, R. O. (2002). A recombinant RNA polymerase II-like enzyme capable of promoter-specific transcription. *Mol Cell* *10*, 635-646.

-
- West, S., Gromak, N., and Proudfoot, N. J. (2004). Human 5' → 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* *432*, 522-525.
- Westover, K. D., Bushnell, D. A., and Kornberg, R. D. (2004a). Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell* *119*, 481-489.
- Westover, K. D., Bushnell, D. A., and Kornberg, R. D. (2004b). Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science* *303*, 1014-1016.
- Wind, M., and Reines, D. (2000). Transcription elongation factor SII. *Bioessays* *22*, 327-336.
- Woudstra, E. C., Gilbert, C., Fellows, J., Jansen, L., Brouwer, J., Erdjument-Bromage, H., Tempst, P., and Svejstrup, J. Q. (2002). A Rad26-Def1 complex coordinates repair and RNA pol II proteolysis in response to DNA damage. *Nature* *415*, 929-933.
- Wu, J., Awrey, D. E., Edwards, A. M., Archambault, J., and Friesen, J. D. (1996). In vitro characterization of mutant yeast RNA polymerase II with reduced binding for elongation factor TFIIS. *Proc Natl Acad Sci U S A* *93*, 11552-11557.
- Xu, Q., Nakanishi, T., Sekimizu, K., and Natori, S. (1994). Cloning and identification of testis-specific transcription elongation factor S-II. *J Biol Chem* *269*, 3100-3103.
- Yamaguchi, Y., Inukai, N., Narita, T., Wada, T., and Handa, H. (2002). Evidence that negative elongation factor represses transcription elongation through binding to a DRB sensitivity-inducing factor/RNA polymerase II complex and RNA. *Mol Cell Biol* *22*, 2918-2927.
- Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., and Handa, H. (1999). NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* *97*, 41-51.
- Yeo, M., Lin, P. S., Dahmus, M. E., and Gill, G. N. (2003). A novel RNA polymerase II C-terminal domain phosphatase that preferentially dephosphorylates serine 5. *J Biol Chem* *278*, 26078-26085.
- Yin, Y. W., and Steitz, T. A. (2002). Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase. *Science* *298*, 1387-1395.
- Yin, Y. W., and Steitz, T. A. (2004). The structural mechanism of translocation and helicase activity in t7 RNA polymerase. *Cell* *116*, 393-404.
- Yudkovsky, N., Ranish, J. A., and Hahn, S. (2000). A transcription reinitiation intermediate that is stabilized by activator. *Nature* *408*, 225-229.
- Zakharova, N., Hoffman, P. S., Berg, D. E., and Severinov, K. (1998). The largest subunits of RNA polymerase from gastric helicobacters are tethered. *J Biol Chem* *273*, 19371-19374.
- Zawel, L., Kumar, K. P., and Reinberg, D. (1995). Recycling of the general transcription factors during RNA polymerase II transcription. *Genes Dev* *9*, 1479-1490.
- Zhang, C., and Burton, Z. F. (2004). Transcription factors IIF and IIS and nucleoside triphosphate substrates as dynamic probes of the human RNA polymerase II mechanism. *J Mol Biol* *342*, 1085-1099.
- Zhang, C., Yan, H., and Burton, Z. F. (2003). Combinatorial control of human RNA polymerase II (RNAP II) pausing and transcript cleavage by transcription factor IIF, hepatitis delta antigen, and stimulatory factor II. *J Biol Chem* *278*, 50101-50111.
- Zhang, G., Campbell, E. A., Minakhin, L., Richter, C., Severinov, K., and Darst, S. A. (1999). Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* *98*, 811-824.

CURRICULUM VITAE

HUBERT KETTENBERGER

PERSONAL DETAILS

- *Date and Place of Birth:*
October 4th, 1973,
Burghausen/Germany
- *Marital Status:*
unmarried
- *Nationality:*
German

SCHOOLS AND VOCATIONAL TRAINING

- September 1980 – July 1990
Grund- und Realschule in Tittmoning and Trostberg
- September 1990 – July 1993
Vocational training as a laboratory assistant at Wacker-Chemie,
Burghausen
- September 1994 – July 1996
Berufsoberschule in Altötting and Regensburg,
graduated with “Fachgebundene Hochschulreife” university entrance
qualifications

COMMUNITY SERVICE

- September 1993 – October 1994
Caregiver in a student’s residence for the physically disabled in
Regensburg

ACADEMIC TRAINING

- November 1994 – September 1999 and April 2000 – December 2001
Study of chemistry at the University of Regensburg, graduated with
“Diploma” exam.
Scholarship for the Gifted from the Bavarian State during the whole
period.
- October 1999 – March 2000
Study abroad semester at the University of Sussex at Brighton/U.K.
- February 2002 – April 2005
PhD thesis in macromolecular crystallography with Prof. Patrick Cramer
at the Gene Center, University of Munich.