

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

**Structural basis of transcription:  
RNA polymerase II fidelity mechanisms  
and RNA 3' fraying**



Jasmin F. Sydow  
aus München

2009

## **Erklärung**

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

## **Ehrenwörtliche Versicherung**

Diese Dissertation wurde selbständig und ohne unerlaubte Hilfe erarbeitet.

München, am

---

Jasmin Sydow

Dissertation eingereicht am 20. August 2009

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

Mündliche Prüfung am 30. September 2009

## Acknowledgements

First and most of all, I want to thank Professor Doctor Patrick Cramer, not only for giving me the chance to let me work on this exciting project, but also for his personal and scientific support throughout the time of my PhD. I want to thank him especially for his great enthusiasm for science and his ability to infect me with it constantly, for his motivation and for always being open for scientific discussion and answering questions, thus creating an excellent work environment. Moreover I thank Patrick for his skill to have recruited a fantastic group of people to be members of his laboratory which altogether made it the most enjoyable workplace I have been at so far.

I particularly thank Florian for having me introduced to the world of Pol II, for all-encompassing help and support not only at the beginning but also later on, for his never-ending capability for examining scientific issues but especially for his great and true interest in science and for most interesting discussion on every observed result, which made me look forward to every forthcoming experiment. Thank you for the great time together in the lab.

Very special thanks to my other Pol II coworkers for the outstanding team work, for sharing ideas and helping out in every aspect. I thank Alan for having become another Pol II expert, for help with all crystallographic questions, for Pol II discussions and all kinds of other truly inspiring conversations. Thank you, Elisabeth, for most efficient team work in the lab and for most enjoyable, sometimes endless conversations, not only at the synchrotron. I thank Gerke for having been my personal Illustrator mentor, and for her enjoyable and pleasant calmness that she radiates constantly.

Moreover, I would like to thank Dengl, not only for his wide-ranging all-round skills in the lab and thus helping out on many issues, but most for his very unusual way to make it impossible for me to not constantly laugh in every single conversation with him or when I just see him. I wish to thank Anselm for a lot of support in all sorts of things related to IT, but especially for his coolness and freestyle in everything he does. I thank Laurent for spreading his impressive knowledge and for his support and help. Thanks to Anass for his enthusiasm, and for scientific and also entertaining and philosophical discussions. I am thankful to Dirk and formatting-Ale for support

concerning crystallography. I thank Claudia Buchen, Stefan Benkert and Kristin for having the lab under control and creating an efficient work environment.

Thank you for your friendship, Jenne and Claudi, and for the good times, even when things were sometimes not going so well. For the great time in but also outside the lab, I want to thank Rieke, Elmar, Martin, Larissa, Michi, Christian, Tobias and Erika. I would like to extend my thanks also to the rest of the lab for the enjoyable time during my PhD.

Thanks to Dietmar Martin, Roland Beckmann, Dirk Eick, Klaus Förstemann and Karl-Peter Hopfner for being my PhD examiners.

Allen voran danke ich von Herzen und unendlich meinen Eltern für all die selbstlose Unterstützung, nicht nur während meiner Doktorarbeit, sondern auch während meines Studiums, ohne die ich nie so weit gekommen wäre, wie ich es jetzt bin. Auch danke ich Nic, für Dein Interesse und alle Unterstützung und für die gute Zeit während fast der Hälfte meines Lebens.

## Summary

Gene transcription is the first step in the decoding of genetic information. RNA polymerase II is the eukaryotic enzyme catalyzing transcription of all protein-coding genes into a complementary chain of ribonucleotides, the messenger RNA (mRNA). High fidelity during this process is of essential importance for every cell as it is thought to prevent formation of erroneous mRNAs and mutant proteins with impaired function. This thesis describes recent advances of our understanding of RNA polymerase fidelity, which stem from structural and functional studies of RNA polymerase II. To study the molecular mechanisms underlying transcription fidelity, we reconstituted complete yeast RNA polymerase II ECs and carried out a systematic, quantitative analysis of the three reactions that determine fidelity: misincorporation, mismatch extension, and cleavage of mismatched RNA 3' ends. The work of this thesis shows that RNA polymerase II prevents erroneous transcription *in vitro* with different strategies that depend on the type of DNA•RNA base mismatch. Certain mismatches are efficiently formed, but impair RNA extension. Other mismatches allow for RNA extension, but are inefficiently formed and efficiently proofread by RNA cleavage. Exemplary erroneous transcription events are rationalized with X-ray structures of T•U mismatch-containing ECs. These studies show accommodation of a T•U wobble base pair (bp) at the active center that dissociates the catalytic metal ion and misaligns the RNA 3' end. Thereby, they explain mismatch-induced disruption of the catalytic site. The mismatch can also stabilize a paused state of RNA polymerase II with a frayed RNA 3' nucleotide. The frayed nucleotide binds in the RNA polymerase II pore either parallel or perpendicular to the DNA-RNA hybrid axis (fraying sites I and II, respectively), and overlaps the nucleoside triphosphate (NTP) site, explaining how it halts transcription during proofreading, before backtracking and RNA cleavage.

## Publications

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

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

Brueckner, F., Armache, K. J., Cheung, A., Damsma, G. E., Kettenberger, H., Lehmann, E., Sydow, J. F., Cramer, P. (2009). Structure-function studies of the RNA polymerase II elongation complex. *Acta Crystallogr D Biol Crystallogr.* **65**, 112-120.

Sydow, J. F., Brueckner, F., Cheung, A. C., Damsma, G. E., Dengl, S., Lehmann, E., Vassylyev, D., Cramer, P. (2009). Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. *Mol Cell* **34**(6): 710-21.

Sydow, J. F., and Cramer, P. (2009). Error prevention, recognition, and removal by cellular RNA Polymerases. *Curr Opin Struct Biol.* *submitted*

## Table of contents

<b>Erklärung</b>	<b>II</b>
<b>Ehrenwörtliche Versicherung</b>	<b>II</b>
<b>Acknowledgements</b>	<b>III</b>
<b>Summary</b>	<b>V</b>
<b>Publications</b>	<b>VI</b>
<b>Table of contents</b>	<b>VII</b>
<b>1. INTRODUCTION</b>	<b>1</b>
1.1 The eukaryotic transcription machinery	1
1.2 DNA-dependent RNA polymerases	4
1.3 Structure of RNA polymerase II	6
1.4 The elongation complex and the nucleotide addition cycle	9
1.5 Fidelity mechanisms of DNA polymerases	13
1.6 Fidelity mechanisms of RNA polymerases	17
1.6.1 Error prevention: substrate loading and selection of rNTPs over dNTPs	17
1.6.2 Error prevention: selection of the correct complementary NTP	20
1.7 Scope of this work	21
<b>2. MISMATCH SPECIFICITY OF RNA POLYMERASE II</b>	<b>23</b>
2.1 Misincorporation efficiency is mismatch-specific	23
2.2 Transcript extension efficiency is mismatch-specific	25
<b>3. STRUCTURAL BASIS OF MISMATCH SPECIFICITY</b>	<b>28</b>
3.1 RNA polymerase II accommodates a T•U wobble pair	28
3.2 Active site disruption explains impaired RNA extension	31
<b>4. RNA POLYMERASE PAUSING</b>	<b>34</b>
4.1 Mismatch extension and RNA 3' fraying	34
4.2 Two RNA fraying sites	38
<b>5. RNA POLYMERASE II PROOFREADING</b>	<b>42</b>
5.1 Nucleotide-specific cleavage of mismatched RNA ends	42

---

5.2 Impaired mismatch accommodation	43
<b>6. DISCUSSION</b>	<b>45</b>
6.1 Mismatch-specific transcription fidelity mechanisms	45
6.2 Mechanistic insights into pausing	46
6.3 Error recognition: mismatches induce off-line states	49
6.4 Error removal: RNA cleavage	50
6.5 A model for RNA proofreading	54
<b>7. EXPERIMENTAL PROCEDURES</b>	<b>57</b>
7.1 Measurement of protein concentration	57
7.2 Isolation of 10-subunit core RNA polymerase II from yeast	57
7.2.1 Yeast fermentation	57
7.2.2 Purification of 10-subunit core RNA polymerase II	59
7.3 Purification of the subcomplex Rpb4/7	65
7.4 Purification of His-tagged RNA polymerase II	67
7.5 EC assembly	69
7.6 Bead-based RNA extension and cleavage assays	69
7.7 Crystal structure determinations	70
<b>8. CONCLUSIONS AND OUTLOOK</b>	<b>72</b>
<b>9. APPENDIX</b>	<b>74</b>
9.1 Unpublished RNA polymerase II EC structures	74
<b>10. ABBREVIATIONS</b>	<b>76</b>
<b>11. REFERENCES</b>	<b>78</b>
<b>12. CURRICULUM VITAE</b>	<b>91</b>



## 1. INTRODUCTION

### 1.1 The eukaryotic transcription machinery

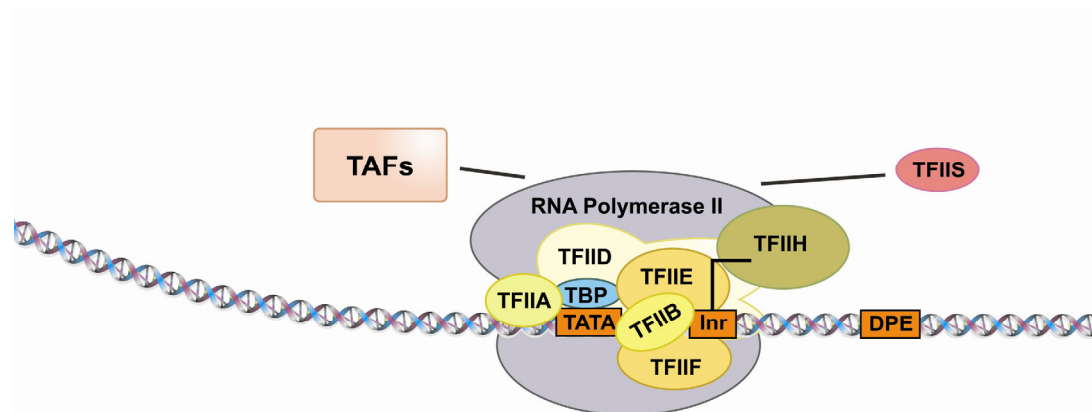
The process of DNA transcription into messenger RNA (mRNA) is catalyzed by DNA-dependent RNA polymerases. The mRNA transcription cycle consists broadly of three stages: initiation, elongation and termination, where the initiation phase is subject to the most regulation. Appropriate modification of chromatin at the promoter region is essential to allow initiation (Li et al, 2007). RNA polymerase has to be then recruited to the promoter. In eukaryotes, the core promoter is the basis for the assembly of the transcription preinitiation complex (PIC). Additionally, regulatory factors, namely activators and repressors, bind to enhancer and silencer elements on the DNA respectively, to allow transmission of regulatory signals via the coactivators. The PIC comprises the general transcription factors TFIIA, TFIIB, TFIID, TFII E, TFII F, TFII H, and RNA polymerase II (Thomas & Chiang, 2006) (Table 1). These factors function collectively to initiate transcription at the transcription start site (Fig. 1). PIC formation begins with the binding of transcription factor TFIID to the TATA box, to the initiator and/or to the downstream promoter element (DPE). The entry of other general transcription factors follows by one of two possible pathways, which is either a sequential assembly pathway or a preassembled RNA polymerase II holoenzyme pathway. The promoter-bound complex is sufficient for a basal level of transcription. However, general cofactors are required to transmit regulatory signals between gene-specific activators and the general transcription machinery in the case of regulated, activator-dependent transcription (Thomas & Chiang, 2006). There exist three classes of general cofactors: the TBP-associated factors (TAFs), the Mediator, and the upstream stimulatory activity-derived positive cofactors and negative cofactor 1. Promoter activity in a gene-specific or cell-type-specific manner is usually adjusted by the independent or combined function of the general cofactors.

**Table 1. Components of the human general transcription machinery (Thomas & Chiang, 2006).**

<b>Factor</b>	<b>Function</b>
TFIIA	Antirepressor. Stabilization of TBP-TATA complex. Coactivator.
TFIIB	Start site selection. Stabilization of TBP-TATA complex. Recruitment of RNA polymerase II/TFIIF.
TFIID	Core promoter-binding factor. Coactivator. Protein kinase. Ubiquitin-activating/conjugating activity. Histone acetyltransferase.
TFIIE	Recruits TFIIH. Facilitates formation of an initiation-competent RNA polymerase II. Involved in promoter clearance.
TFIIF	Binds RNA polymerase II and facilitates RNA polymerase II recruitment to the promoter. Recruits TFIIE and TFIIH. Functions with TFIIB and RNA polymerase II in start site selection. Facilitates RNA polymerase II promoter escape. Enhances the efficiency of RNA polymerase II elongation.
TFIIH	ATPase activity for transcription initiation and promoter clearance. Helicase activity for promoter opening. Transcription-coupled nucleotide excision repair. Kinase activity for phosphorylating RNA polymerase II CTD. E3 ubiquitin ligase activity.
RNA polymerase II	Transcription initiation, elongation, termination. Recruitment of mRNA capping enzymes. Transcription-coupled recruitment of splicing and 3' end processing factors. CTD phosphorylation, glycosylation, and ubiquitination.

Studies for over a decade led to the model of a sequential assembly pathway resulting in a productive PIC assembly at the promoter region (Buratowski et al, 1989; Fire et al, 1984; Hawley & Roeder, 1985; Samuels & Sharp, 1986; Van Dyke et al, 1989). It includes binding of TFIID to the promoter region, followed by entry of TFIIA and TFIIB which help to stabilize promoter-bound TFIID. Recruitment of RNA polymerase II with TFIIF leads to formation of a stable TFIID-TFIIA-TFIIB-RNA polymerase II/TFIIF-promoter complex. After this, TFIIE is recruited, followed by entry of TFIIH. An alternative is the RNA polymerase II holoenzyme pathway which was revealed when several groups discovered that RNA polymerase II could be purified as a preassembled holoenzyme complex containing RNA polymerase II and SRBs

(suppressors of RNA polymerase B mutations (Kim et al, 1994; Koleske & Young, 1994)), in combination or without several general transcription factors, and other proteins involved in chromatin remodeling, DNA repair, and mRNA processing (Cairns et al, 1996; Chao et al, 1996; Cho et al, 1998; Liu et al, 2001; Maldonado et al, 1996; McCracken et al, 1997; Nakajima et al, 1997; Ossipow et al, 1995; Wilson et al, 1996; Wu & Chiang, 1998; Wu & Hampsey, 1999; Yuryev et al, 1996). The RNA polymerase II holoenzyme complex was proposed to contain RNA polymerase II, TFIIB, TFIIE, TFIIF, TFIIH, GCN5 histone acetyltransferase, SWI/SNF chromatin remodeling factor, and SRBs, but to lack TFIID and TFIIA (Thomas & Chiang, 2006). The fact that in this RNA polymerase II holoenzyme complex TFIID is missing, suggests that TFIID may facilitate entry of RNA polymerase II holoenzyme to the promoter region. This would be analogous to the mechanism in the prokaryotic system where the dissociable  $\sigma$  factor recruits core RNA polymerase to the promoter region for PIC assembly. It was proposed that both assembly pathways exist *in vivo*, but that either pathway may be employed in response to different environmental cues (Thomas & Chiang, 2006).



**Figure 1. The eukaryotic transcription machinery.** After formation of the PIC at one of the eukaryotic core promoters, coactivator complexes serve as an interface between the general RNA polymerase II machinery and transcriptional activators and repressors.

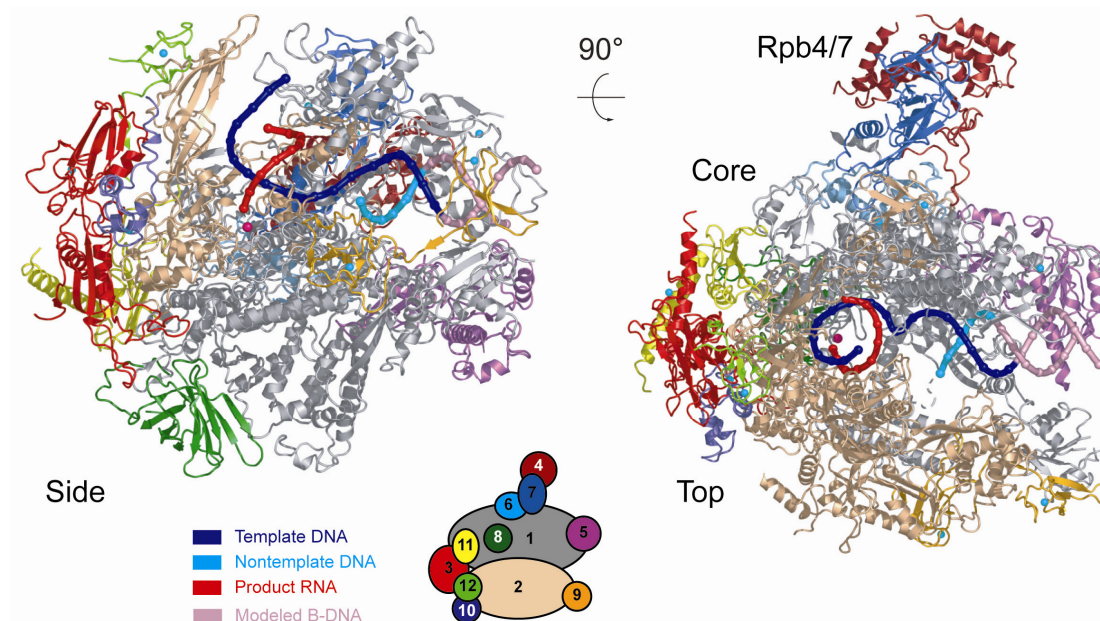
In the open complex a transcription bubble has formed. After initiation, RNA polymerase II enters the elongation phase, leaving most initiation factors behind. The produced pre-mRNA is co-transcriptionally processed by capping and splicing. Elongation factors such as TFIIS are involved in modulation of the catalytic activity, pausing, and transcriptional arrest of RNA polymerase II (Sims et al, 2004). Transcription through chromatin requires additional factors (Armstrong, 2007; Kulaeva et al, 2007; Li et al, 2007). At the end of a gene, transcription is terminated (Gilmour & Fan, 2008) and upon a signal on the transcript, RNA is cleaved and polyadenylated. Thereupon RNA polymerase II can be removed from the DNA and recycled to start with another transcription cycle. The C-terminal domain of RNA polymerase II (CTD) integrates nuclear events by binding proteins involved in mRNA biogenesis (Buratowski, 2003; Hirose & Manley, 2000; Meinhart et al, 2005). The CTD is flexibly linked to the core enzyme and consists of heptapeptide repeats of the consensus sequence YSPTSPS. CTD-binding proteins recognize a specific CTD phosphorylation pattern, which changes during the transcription cycle. Structural and functional studies of CTD-binding and CTD-modifying proteins and their complexes with CTD peptides elucidated CTD structure and revealed some of the mechanisms underlying CTD function.

## **1.2 DNA-dependent RNA polymerases**

There are generally two families of DNA-dependent RNA polymerases, comprising single-subunit and multisubunit RNA polymerases (Cramer, 2002).

Single-subunit RNA polymerases are those of bacteriophages and the mitochondrial RNA polymerase. Similar to DNA polymerases, these enzymes possess one protein subunit which resembles the shape of a right hand, including a palm, thumb, and a finger domain.

Multisubunit RNA polymerases comprise enzymes of all three domains of life, archaea, bacteria, and eukaryotes (Cramer, 2002). Archaea and bacteria possess only one type of RNA polymerase. In contrast, in eukaryotic cells, three different DNA dependent RNA polymerases – RNA polymerase I, RNA polymerase II, and RNA polymerase III – are responsible for gene transcription. Production of ribosomal RNA is carried out by RNA polymerase I, synthesis of messenger RNAs and small nuclear RNAs by RNA polymerase II, and production of transfer RNAs and other small RNAs by RNA polymerase III. Plants have two additional nuclear RNA polymerases that have been recently discovered, RNA polymerase IV (Herr et al, 2005; Kanno et al, 2005; Onodera et al, 2005) and RNA polymerase V (Pontier et al, 2005; Wierzbicki et al, 2008). They play nonredundant roles in siRNA-directed DNA methylation and gene silencing (Ream et al, 2009). RNA polymerase I, II, and III comprise 14, 12, and 17 subunits, respectively, and have a total molecular weight of 589, 514, and 693 kDa, respectively (Cramer et al, 2008). RNA polymerase IV and V are composed of subunits that are paralogous or identical to the 12 subunits of RNA polymerase II and are thus thought to be RNA polymerase II-like enzymes that evolved specialized roles in the production of noncoding transcripts for RNA silencing and genome defense (Ream et al, 2009). Ten subunits form a structurally conserved core, and additional subunits are located on the periphery (Fig. 2).



**Figure 2. Complete RNA polymerase II EC structure (Kettenberger et al, 2004).** Side view and top view of a ribbon model of RNA polymerase II structure and a schematic representation of the 12 subunits in the according color demonstrate the complexity of the structure. The catalytic metal ion A is shown as a pink sphere, the eight zinc ions as cyan spheres.

### 1.3 Structure of RNA polymerase II

RNA polymerase II consists of a 10-subunit core enzyme and a peripheral heterodimer of subunits Rpb4 and Rpb7 (Fig. 2). The core enzyme comprises subunits Rpb1, Rpb2, Rpb3, and Rpb11, which contain regions of sequence and structural similarity in RNA polymerase I, RNA polymerase III, bacterial RNA polymerases (Vassylyev et al, 2002; Zhang et al, 1999), and the archaeal RNA polymerase (Hirata et al, 2008; Korkhin et al, 2009; Kusser et al, 2008). The RNA polymerase II core also comprises subunits Rpb5, Rpb6, Rpb8, Rpb10, and Rpb12, which are shared between RNA polymerase I, II, and III (common subunits, Table 2). In the archaeal polymerase, counterparts of these common subunits exist, except Rpb8, but only a counterpart of Rpb6 exists in the bacterial enzyme (Minakhin et al, 2001). Finally, homologues of the core subunit Rpb9 exist in RNA polymerase I and RNA polymerase III, but not in the archaeal or bacterial enzyme. Initial electron

microscopic studies of RNA polymerase II revealed the overall shape of the enzyme (Darst et al, 1991). The core RNA polymerase II could subsequently be crystallized and led to an electron density map at 6 Å resolution (Fu et al, 1999). A backbone model of the RNA polymerase II core resulted from crystal improvement by controlled shrinkage and phasing at 3 Å resolution (Cramer et al, 2000). This revealed that Rpb1 and Rpb2 form opposite sides of a positively charged active center cleft, whereas the smaller subunits are arrayed around the periphery. Refined atomic structures of the core RNA polymerase II were obtained in two different conformations and revealed domain-like regions within the subunits, as well as surface elements predicted to have functional roles (Cramer et al, 2001) (Fig. 3). The active site and the bridge helix, which spans the cleft, line a pore in the floor of the cleft. The Rpb1 side of the cleft forms a mobile clamp.

The clamp was trapped in two different open states in the free core structures (Cramer et al, 2001) but was closed in the structure of a core complex that included DNA and RNA (Gnatt et al, 2001). The mobile clamp is connected to the body of the polymerase by five switch regions that show conformational variability. The Rpb2 side of the cleft consists of the lobe and protrusion domains. Rpb2 also forms a protein wall that blocks the end of the cleft. The RNA polymerase II core structures lacked subunits Rpb4 and Rpb7, which can dissociate from the yeast enzyme (Edwards et al, 1991). A structure of the archaeal homologue of the Rpb4/7 heterodimer showed that Rpb7 contains an N-terminal domain, later called the tip domain, and a C-terminal domain that includes an oligosaccharide-binding fold (Todone et al, 2001). The approximate location of Rpb4/7 on the core polymerase was first determined by electron microscopy (EM) of two-dimensional crystals (Jensen et al, 1998).

Table 2. RNA polymerase subunits

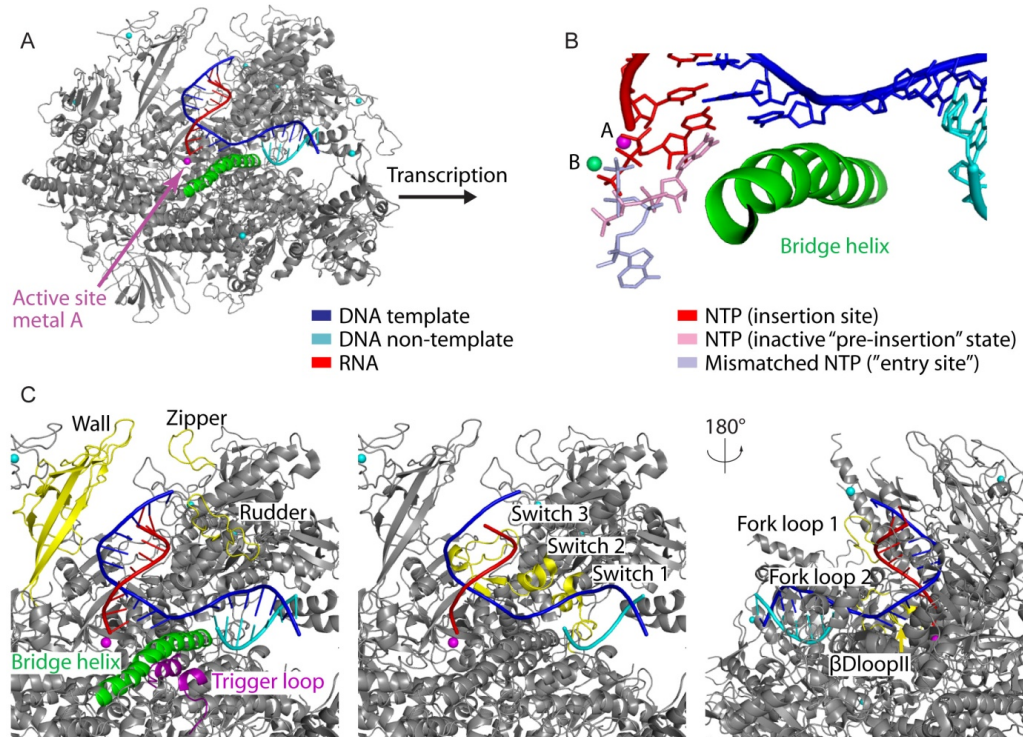
RNA polymerase	RNA polymerase I	RNA polymerase II	RNA polymerase III
Ten-subunit core	A190	Rpb1	C160
	A135	Rpb2	C128
	AC40	Rpb3	AC40
	AC19	Rpb11	AC19
	A12.2	Rpb9	C11
	Rpb5 (ABC27)	Rpb5	Rpb5
	Rpb6 (ABC23)	Rpb6	Rpb6
	Rpb8 (ABC14.5)	Rpb8	Rpb8
	Rpb10 (ABC10 $\beta$ )	Rpb10	Rpb10
Rpb12 (ABC10 $\alpha$ )	Rpb12	Rpb12	
Stalk	A14	Rpb4	C17
	A43	Rpb7	C25
TFIIF-like subcomplex <sup>a</sup>	A49	(Tfg1/Rap74)	C37
	A34.5	(Tfg2/Rap30)	C53
Pol III-specific subcomplex	-	-	C82
	-	-	C34
	-	-	C31
Number of subunits	14	12	17

<sup>a</sup>The two subunits in Pol I and Pol III are predicted to form heterodimers that resemble part of the pol II initiation/elongation factor TFIIF, which is composed of subunits Tf1, Tfg2, and Tfg3 in *Saccharomyces cerevisiae*, and of subunits Rap74 and Rap30 in human.

Later, EM analysis of single particles revealed a closed clamp and showed that the Rpb4/7 subcomplex protrudes from outside the core enzyme below the clamp (Craighead et al, 2002). A different open-closed transition that involved the polymerase jaws was observed by EM of two-dimensional crystals (Asturias et al, 1997). Crystallographic backbone models of the complete RNA polymerase II then revealed the exact position and orientation of Rpb4/7 and showed that it formed a wedge between the clamp and the linker to the unique tail-like C-terminal repeat



domain (CTD) of the polymerase (Armache et al, 2003; Bushnell & Kornberg, 2003). Refinement of a complete atomic model of RNA polymerase II was finally possible with the crystal structure of free Rpb4/7, together with an improved resolution of the complete RNA polymerase II crystals (Armache et al, 2005).



**Figure 3. Structure of the RNA polymerase II EC (Kettenberger et al, 2004; Wang et al, 2006; Westover et al, 2004a)** (A) Overview of the EC structure. (B) superposition of NTP-binding sites (red, insertion site; violet, entry site; pink, inactive pre-insertion-like state). (C) Functional RNA polymerase II surface elements in the EC.

#### 1.4 The elongation complex and the nucleotide addition cycle

During the process of gene transcription, RNA polymerase II is moving along a DNA template and synthesizes a complementary mRNA chain. This is achieved by repetitive cycles of adding a substrate nucleotide. The EC is characterized by the transcription bubble, an unwound DNA region. The transcription bubble contains a short hybrid duplex formed between the DNA template strand and the RNA product

emerging from the active site. The mechanism of RNA elongation was elucidated by structural studies of RNA polymerase II – nucleic acid complexes (Fig. 3).

The point of DNA entry to the RNA polymerase II cleft was first revealed by EM (Poglitsch et al, 1999). The core RNA polymerase II transcribing a tailed template DNA, which allows for promoter-independent transcription initiation, was the first crystal structure of an RNA polymerase II – nucleic acid complex (Gnatt et al, 2001). This structure revealed downstream DNA entering the cleft and an 8 to 9 base pair DNA-RNA hybrid in the active center. Protein surface elements predicted to play functional roles were revealed by comparison with the high-resolution core RNA polymerase II structure (Cramer et al, 2001). Later, polymerase EC structures utilized synthetic DNA-RNA scaffolds (Kettenberger et al, 2004; Westover et al, 2004b) and could thereby show the exact location of the downstream DNA and several nucleotides upstream of the hybrid (Figure 2). It was suggested how RNA polymerase II unwinds downstream DNA and how it separates the RNA product from the DNA template at the end of the hybrid. Usually RNA polymerase II uses DNA as a template, but there is also evidence that the enzyme can use RNA templates. Structures of an RNA template-product duplex showed that this RNA-RNA hybrid can bind to the site normally occupied by the DNA-RNA hybrid and provided the structural basis for the phenomenon of RNA dependent RNA synthesis by RNA polymerase II (Lehmann et al, 2007). Additional structures of RNA polymerase II ECs included the NTP substrate (Kettenberger et al, 2004; Wang et al, 2006; Westover et al, 2004a). Mechanisms for correct NTP selection and nucleotide incorporation into RNA were suggested.

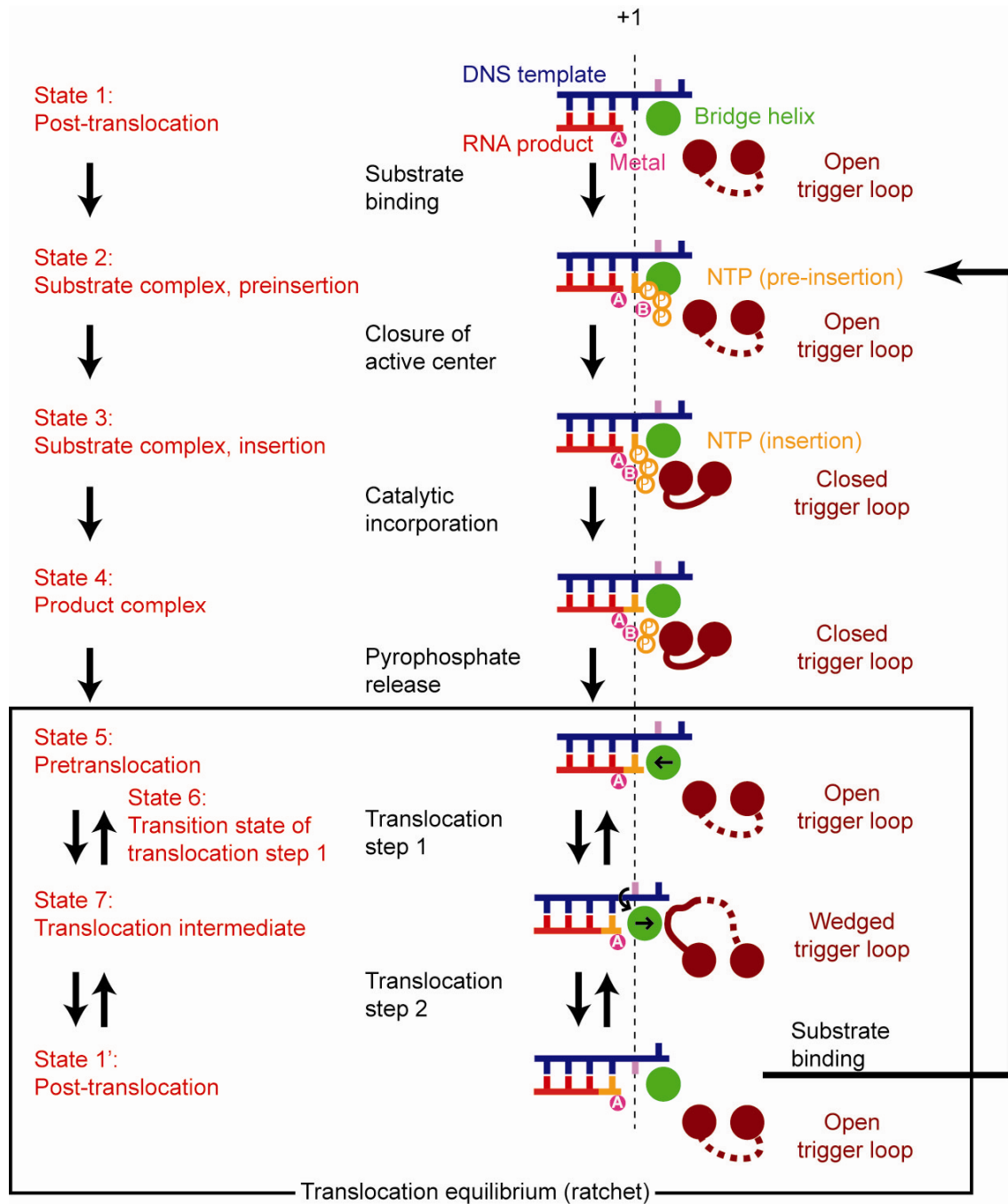
A conserved nucleotide addition cycle mechanism for all three kingdoms of life has been recently proposed (Brueckner & Cramer, 2008) (Fig. 4) and has also been visualized as a movie (Brueckner et al, 2009). It is based on a brownian ratchet model which assumes that the ground state of the EC is an equilibrium between inter-converting pre-translocation and post-translocation states (Bar-Nahum et al,

2005). This oscillation is temporarily stopped by substrate binding and resumes around the next template position after nucleotide addition. The cycle begins with the binding of an NTP substrate to the EC. The growing 3' end of the RNA chain becomes elongated by catalytic addition of the nucleotide which results in formation of a pyrophosphate ion. The release of pyrophosphate leads to the pre-translocation state whereas the incorporated 3' terminal nucleotide remains in the substrate site. After translocation of DNA and RNA, the EC is in a post-translocation state, with a free substrate site for binding of the next incoming NTP. The nucleotide addition cycle can then be repeated.

In eukaryotic RNA polymerase II, an NTP substrate has been crystallographically trapped in the insertion site (Wang et al, 2006; Westover et al, 2004a) (Fig. 3B) which is formed by closure of the active site. Insertion site substrate binding results in a complete folding of the trigger loop (Wang et al, 2006), a mobile part of the active center first observed in free bacterial RNA polymerase (Vassylyev et al, 2002), and in the RNA polymerase II-TFIIS complex (Kettenberger et al, 2003). The NTP has also been trapped in an overlapping site (Kettenberger et al, 2004), termed the pre-insertion site. NTPs bound to either site form Watson-Crick interactions with a base in the DNA template strand although only the insertion site NTP base is co-planar with the templating base. Recent studies of functional complexes of the bacterial RNA polymerase revealed the close conservation of the EC structure (Vassylyev et al, 2007a).

Structural studies are consistent with a two-metal ion mechanism for all polymerases (Steitz, 1998). The NTP binds two catalytic metal ions (Cramer et al, 2001; Westover et al, 2004a) named metal A and B. Whereas metal A is persistently bound to the active site, the second mobile metal B enters with the NTP, bound to its triphosphate moiety (Westover et al, 2004a). Metal A is held by three invariant aspartate side chains and binds the RNA 3' end (Cramer et al, 2001). Catalysis is not permitted in the pre-insertion state, as the NTP triphosphate and metal B are too far

from metal A (Brueckner et al, 2009). A working model of the RNA polymerase active center is based on the prerequisite that the polymerase and the nuclease reactions are performed by the same active site, that both reactions are based on the substitution nucleophilic bimolecular ( $S_N2$ ) mechanism operating in opposite directions, and that the reactions' geometry requires the two  $Mg^{2+}$  ions to be situated at equal distances from the non-bridging oxygen of the scissile phosphate collinearly to the axis of the attack by an activated water molecule (Sosunov et al, 2003). Recent studies have proposed a two-step mechanism of translocation via a trigger loop-stabilized EC intermediate with an altered structure of the central bridge helix (Brueckner & Cramer, 2008). Translocation includes a first step, during which the hybrid moves from the pre- to the post-translocation position (Fig. 4, boxed region). Downstream DNA translocates until the next DNA template base reaches the pre-templating position above the bridge helix. The template base twists by  $90^\circ$ , to reach its templating position in the active center during the second step. A flipping of the phosphate backbone group between DNA template bases +1 and +2 and sliding of downstream DNA to the post-translocation position accompanies the template base twisting.



**Figure 4. Schematic representation of the nucleotide addition cycle (Brueckner & Cramer, 2008; Brueckner et al, 2009).** The seven states of the nucleotide addition cycle are indicated on the left. The vertical dashed line indicates register +1.

## 1.5 Fidelity mechanisms of DNA polymerases

The general mechanism of replication and transcription is the same, as both RNA and DNA polymerases translocate along a DNA template to produce a duplex of

nucleic acids. High fidelity of replication by DNA polymerases is crucial for maintaining the genetic integrity. Transcription fidelity is thought to prevent formation of erroneous mRNAs and mutant proteins with impaired function (Saxowsky & Doetsch, 2006). Because of the similarity of the catalyzed processes of replication and transcription and of the existence of numerous extensive studies on the fidelity of DNA polymerases, the known replicative fidelity mechanisms will be described here.

DNA polymerase fidelity can vary between low and high levels, dependent on their biological function and on the organism (reviewed in (Bebenek & Kunkel, 1995; Echols & Goodman, 1991; Kunkel & Bebenek, 2000)). Whereas viral enzymes (as well as some bacterial ones) have the selection pressure for low fidelity to increase mutation rates in the virus, eukaryotic enzymes have to ensure low mutation rates and thus possess replicative enzymes with very high fidelity. However, eukaryotic cells need several different DNA polymerases and some have very low fidelity. Thus, DNA polymerase nucleotide insertion fidelities range from  $10^{-2}$  to  $10^{-5}$  (meaning one error in  $10^2$  to  $10^5$  insertions is made) in viral systems to  $10^{-3}$  to  $10^{-5}$  in most bacterial and high-fidelity eukaryotic systems to  $10^0$  to  $10^{-2}$  in low-fidelity eukaryotic enzymes (summarized in (Kool, 2002)).

The major contribution to high replication fidelity is the high selectivity against incorporation of a wrong nucleotide, which is achieved partly by enthalpy-entropy compensation and partly by the exquisite shape complementarity of the four canonical Watson-Crick base pairs in the binding pockets (McCulloch & Kunkel, 2008) and the selection of the correct sugar. To accomplish the correct selection, there is evidence for other important influences other than base-pair hydrogen bonding such as active site geometry, the size and shape of the base pairs and minor groove interactions (Kool, 2002). After misincorporation, the efficiency of extension, proofreading and DNA repair mechanisms contribute to the overall fidelity.

In the *Bacillus stearothermophilus* DNA polymerase I fragment, five sites on the enzyme are important for fidelity, the (i) insertion site, in which the cognate

nucleotide pairs with the template base, (ii) the catalytic site directly adjacent to the insertion site in which the 3' hydroxyl of the primer strand and the coordination sphere for two  $Mg^{2+}$  ions are located, forming the catalytic center, (iii) the pre-insertion site, which houses the template base in a step prior to incorporation, (iv) the post-insertion site in which the growing 3' end of the duplex DNA is located and (v) the DNA duplex binding-region in which a four bp duplex DNA segment is bound (Johnson & Beese, 2004).

Extensive structural studies on mismatch-containing DNA polymerase - DNA structures have proposed four broad categories of mismatch-induced disruptions of the active site, followed by the assumption that a consequence of each of these mechanisms is disruption of the insertion site: (i) displacement of the template strand and disruption of the pre-insertion site; (ii) disruption of the primer strand and the assembly of the catalytic site; (iii) disruption of both the template and primer strands; and (iv) fraying of the DNA at the insertion site (Johnson & Beese, 2004). DNA polymerases select against rNTPs by forming a "steric gate" between two amino acid side chains (Glu and Phe) which sandwich the substrate sugar moiety and exclude the 2' OH (Astatke et al, 1998; Boyer et al, 2000; Gao et al, 1997; Joyce, 1997; Yang et al, 2002). Thereby, the substrate is in the insertion site in an active, closed conformation.

Results of extensive analysis of each of the 12 possible misincorporation events and their extension in *E. coli* DNA polymerase I (Klenow fragment) demonstrated that the polymerase discriminates between mismatches mainly on the basis of the mismatch identity with the surrounding sequence context playing a significant, but secondary role (Joyce et al, 1992). It was possible to summarize observations with the following simple rules; (i) at template pyrimidine positions, misinsertion of the non-complementary purine is favored over pyrimidine insertion, (ii) at template A positions, dATP insertion is preferred over dCTP and dGTP, (iii) at template G, misinsertion rates of the three dNTPs are more similar than at the other

template bases, (iv) in the insertion reaction, dGTP incorporation opposite T and dATP incorporation opposite A are the most rapid reactions. For the extension reactions, they reasoned that (i) purine-pyrimidine (or pyrimidine-purine) and pyrimidine-pyrimidine mispairs are much more readily extended than purine-purine mispairs, (ii) of the latter, G•G and A•A tend to be preferred over G•A and A•G, (iii) in the extension reactions T•G and T•C are the preferred mispairs, (iv) for pyrimidine-pyrimidine mispairs, the exact nature of the bases has a strong influence, with T•G much more readily formed and extended than C•A. Further studies on DNA polymerase fidelity generally agree with the described results (Bebenek et al, 1990; Joyce et al, 1992; Kwok et al, 1990; Lai & Beattie, 1988; Mendelman et al, 1989; Mendelman et al, 1990; Perrino & Loeb, 1989; Perrino et al, 1989).

Additional contribution to the overall replicative fidelity is provided after nucleotide selection, such as control of extension after erroneous incorporation and proofreading as well as various DNA repair mechanisms. Also additional subunits of multi-component enzymes, like the sliding clamp, lead to increased fidelity (Kool, 2002). To proofread replicative errors, the 3'-5' exonuclease activity of DNA polymerases comes into play. Biochemical, structural and genetic experiments have demonstrated that the polymerase and 3'-5' exonuclease activities of *E. coli* DNA polymerase I reside on different domains of its large proteolytic fragment (Klenow fragment) (Beese & Steitz, 1989; Derbyshire et al, 1988; Freemont et al, 1986; Joyce & Steitz, 1987; Ollis et al, 1985). Klenow fragment can bind a second DNA substrate and carry out exonucleolytic cleavage, even under conditions where the polymerase active site is occupied by a duplex of DNA strands (Catalano & Benkovic, 1989). In DNA polymerase III holoenzyme from *E. coli*, the DNA polymerase and 3'-5' exonuclease activities are also located in different active sites, residing on separate subunits rather than on two domains of the same subunit (Maki & Kornberg, 1985; Scheuermann & Echols, 1984).



## 1.6 Fidelity mechanisms of RNA polymerases

In comparison to DNA polymerases, the investigation of RNA polymerase fidelity has been less well studied, but several biochemical and structural results are available (Alic et al, 2007; Holmes et al, 2006; Kireeva et al, 2008; Svetlov et al, 2004; Thomas et al, 1998; Wang et al, 2006). These studies showed that misincorporation leads to slow addition of the next nucleotide, and that a mismatched RNA 3' end can be removed with factors that stimulate the polymerase cleavage activity. In a bacterial EC, a mismatched RNA 3' nucleotide induces an unactivated state, and is removed by cleavage-stimulatory Gre factors (Erie et al, 1993). In human RNA polymerase II, a mismatched RNA 3' nucleotide causes slow addition of the next nucleotide, and RNA cleavage is stimulated by TFIIS (Thomas et al, 1998). The accuracy of transcription is relatively high, with an estimated error rate of less than  $10^{-5}$  for bacterial and eukaryotic RNA polymerases (Blank et al, 1986; de Mercoyrol et al, 1992; Rosenberger & Hilton, 1983). As well as in DNA polymerases, it is generally achieved by two main fidelity-determining mechanisms, the discrimination against the wrong nucleotide, and recognition and removal of a mismatched nucleotide (proofreading). ECs of transcribing RNA polymerases exist in four translocation states: a post-translocation state, a pre-translocation state, a translocation intermediate between pre- and post-translocation states (Brueckner & Cramer, 2008), and a backtracked state. The pre-translocated state can exist in a paused conformation, for instance after a misincorporation event, interrupting the process of catalytic nucleotide addition.

### 1.6.1 Error prevention: substrate loading and selection of rNTPs over dNTPs

Although concentrations of 2'-deoxy NTPs are at least 10-fold lower than those of rNTPs *in vivo* (Albert & Gudas, 1985; Kornberg & Baker, 1992; Reichard, 1985), RNA

polymerases must have evolved mechanisms to positively select only nucleotides with the correct sugar.

In single-subunit T7 RNA polymerase, the selection process starts in the pre-insertion site, where the substrate is located (Temiakov et al, 2004). Positioning of the substrate in a pre-insertion site in the T7 enzyme represents part of a two-step mechanism of substrate loading, including a template base dependent binding of the substrate NTP to an inactive pre-insertion conformation as a first step, followed by the second step of isomerization of the EC into an active insertion state by closure of the active site (Temiakov et al, 2004; Yin & Steitz, 2004). In single-subunit T7 RNA polymerase, it has been shown that the hydroxyl group of Tyr639 forms a hydrogen bond with the 2' OH of an incoming rNTP (Briebe & Sousa, 2000; Huang et al, 1997; Sousa & Padilla, 1995).

In multisubunit RNA polymerases, studies showed that two different conformations of bound NTP existed (Erie et al, 1993). After later structural studies, the mechanism of substrate selectivity became highly debated. Apart from the postulation of other sites to which substrate can bind simultaneously with the insertion site based on kinetic analyses (Foster et al, 2001; Gong et al, 2005), structural findings proposed binding of the NTP to a template dependent pre-insertion site in the eukaryotic RNA polymerase II (Kettenberger et al, 2004). These findings were complicated by the detection of a template-independent site of entry (Westover et al, 2004a). The template-dependent model of NTP binding in multisubunit RNA polymerases (Kettenberger et al, 2004) was later confirmed by high-resolution structural work on a bacterial enzyme in the presence of the antibiotic streptolydigin, which supports a two-step mechanism of substrate loading, similar to the T7 system (Vassylyev et al, 2007b). First, the NTP substrate adopts an inactive pre-insertion intermediate state and binds to an open active center conformation, whereas the NTP works as a ratchet to stabilize the post-translocated EC. The pre-insertion intermediate may serve as a first sieve for substrate selection, passing into a

catalytically active insertion intermediate. This active intermediate is stabilized by two trigger helices emerging of refolding of the trigger loop and represents a second, finer sieve for substrate selection. Second, folding of the trigger loop and thus closure of the active center leads to all contacts required for catalysis (Vassylyev et al, 2007b; Wang et al, 2006). After delivery of the correct NTP to the insertion site, subsequent catalysis results in RNA extension and pyrophosphate formation. The resulting release of pyrophosphate is thought to destabilize the closed conformation of the active center and to lead to trigger loop unfolding. The alternative model for nucleotide addition, which involves binding of the NTP to a putative entry site in the pore and in which the nucleotide base is oriented away from the DNA template (Westover et al, 2004a) was proposed as an additional substrate intermediate en route to the pre-insertion state (Vassylyev et al, 2007b). The pre-insertion site of eukaryotic RNA polymerase differs from the one in the T7 system. In the T7 EC, the substrate gets anchored by the folded O/O' helices in the pre-insertion state, positioning it far from the active site whereas the trigger helices in the pre-insertion bacterial EC are probably unfolded and bind the substrate only after their folding in the insertion complex (Temiakov et al, 2004).

In the *T. thermophilus* RNA polymerase, residue Asn  $\beta$ '737 permits discrimination against non-cognate dNTPs (Svetlov et al, 2004; Vassylyev et al, 2007b) by forming hydrogen bonds with the O3' and O2' of the substrate ribose. This differs from the eukaryotic analogous residue, which is interacting only with the O3' atom of the substrate NTP (Wang et al, 2006). However, the role of Arg446 or Arg $\beta$ '704 in discrimination between rNTP and dNTP is the same in eukaryotic RNA polymerase II (Wang et al, 2006) and the bacterial system (Vassylyev et al, 2007b), respectively. A dramatic increase of dNTP incorporation was demonstrated for mutation of  $\beta$ 'Asn458 of the *E. coli* enzyme (Svetlov et al, 2004). Nevertheless, mutation of the corresponding residue in the RNA polymerase II system did not have any effect on transcription fidelity (Wang et al, 2006). It could rather be shown, that

mutation of Rpb1 residue Glu1103 decreases selectivity against dNTPs in RNA polymerase II (Kireeva et al, 2008). Structural data proposed His1085 of the RNA polymerase II trigger loop to interact amongst others with the NTP sugar, explaining the basis for incorporation of an NTP into RNA several 100-fold more rapidly than a dNTP (Wang et al, 2006).

### **1.6.2 Error prevention: selection of the correct complementary NTP**

Biochemical studies on the single-subunit wild-type T7 RNA polymerase and of several T7 RNA polymerase point mutants demonstrated that the wild-type enzyme selects strongly against incorporation of an incorrect nucleotide and that RNAs bearing 3' mismatches are extended more slowly than correctly paired 3' termini (Huang et al, 2000). The point mutant His784 resulted in increased misincorporation and mismatch extension whereas point mutation of Gly640, Phe644, and Gly645 lead only to an increase of misincorporation, but not of mismatch extension. In comparison to other RNA polymerases, a post-misincorporation proofreading mechanism could not be detected. Thus, T7 RNA polymerase fidelity depends entirely on discrimination against misincorporation events.

In the multisubunit enzyme of *E. coli*, mutation of  $\beta$ Asp675 leads to a dramatically increased incorporation of incorrect nucleotides and was thus proposed to affect the transfer from the proposed E to the A site (Holmes et al, 2006). In RNA polymerase II, recognition of a correct NTP was suggested to be coupled with catalysis, ensuring the fidelity of transcription (Wang et al, 2006). The trigger loop was shown to swing beneath a correct NTP in the proposed A site, positioning its residue His1085 to form an interaction network with the NTP base, sugar, phosphates, and additional RNA polymerase II residues (Wang et al, 2006). Thus, the trigger loop was suggested to detect the topology of a correct RNA-DNA hybrid base pair and to exclude not only a dNTP, but also purine-purine and pyrimidine-

pyrimidine mismatches. This was accounted for by different helix parameters between downstream DNA-DNA (B form) and upstream RNA-DNA (A form) helices. On the basis of the effects of two mutations in the trigger loop of the *E. coli* RNA polymerase on misincorporation and pausing, the two-pawl ratchet mechanism of transcription elongation suggested that the trigger loop to play a role in transcription fidelity (Bar-Nahum et al, 2005). Indirect involvement of Rpb1 residue Glu1103 in trigger loop closure was suggested by the promotion of transition-type and transversion-type misincorporation, when mutated (Kireeva et al, 2008).

Misincorporation can also be the result of transcription of DNA lesions (Brueckner et al, 2007; Damsma et al, 2007). Biochemical studies illustrated that template misalignment can be another reason for nucleotide misincorporation (Kashkina et al, 2006; Pomerantz et al, 2006).

## **1.7 Scope of this work**

Prior to this work, a systematic, quantitative analysis of the fidelity determining mechanisms of RNA polymerase II was still missing. It was unknown which misincorporation events were efficient, how RNA polymerase II would handle the situation of an existing mismatch in the active center by means of its elongation or cleavage, and how these mechanisms were combined to lead to the known high transcriptional fidelity. To further elucidate these issues, the scope of this work was to reconstitute complete yeast RNA polymerase II ECs and to investigate the three reactions that determine fidelity: misincorporation, mismatch extension, and cleavage of mismatched RNA 3' ends.

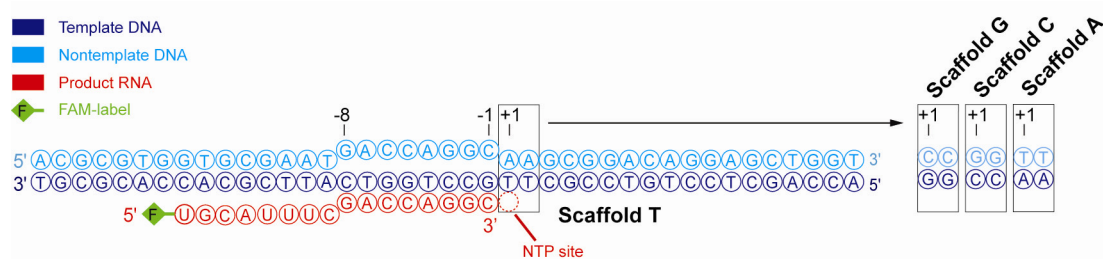
Moreover, the structural basis for effects of DNA•RNA mismatches on the conformation of the RNA polymerase II active center was unknown. To rationalize exemplary erroneous transcription events, we introduced mismatches in the nucleic acid scaffolds at several positions in the polymerase active site. The co-crystallized

ECs were structurally studied to obtain explanations for mismatch accommodation, impaired mismatch extension and possibly RNA cleavage.

## 2. MISMATCH SPECIFICITY OF RNA POLYMERASE II

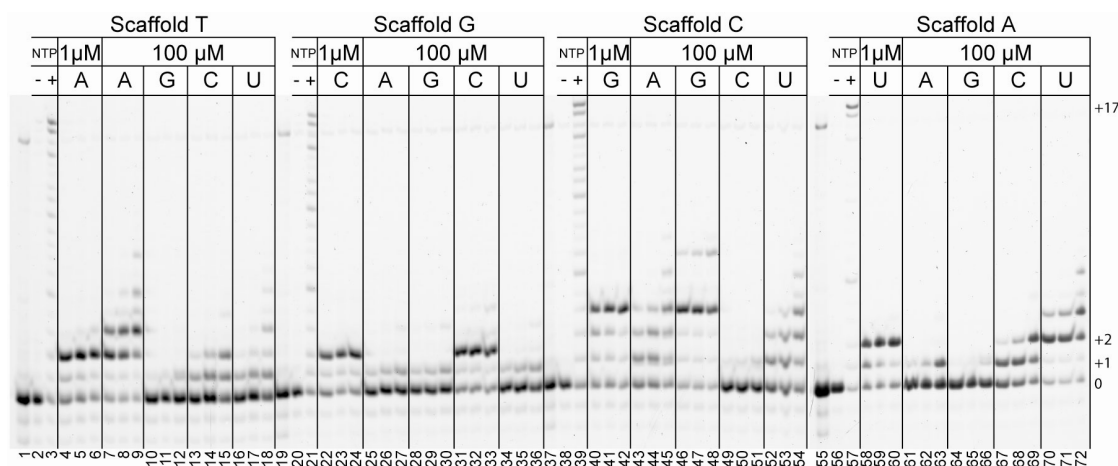
### 2.1 Misincorporation efficiency is mismatch-specific

To determine the efficiency of misincorporation by RNA polymerase II, we performed RNA extension assays with reconstituted ECs (Brueckner et al, 2007; Kireeva et al, 2003). The nucleic acid scaffolds contained fully complementary DNA strands, 18 bps of downstream DNA, 15 bps of upstream DNA, an eight bp DNA-RNA hybrid, and eight nucleotides of exiting RNA labeled with 6-carboxyfluoresceine (FAM) at its 5'-end (Fig. 5). The scaffolds T, G, C and A differed in the +1 nucleotide opposite the NTP site (Fig. 5).



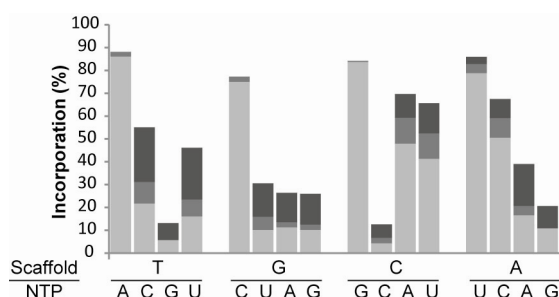
**Figure 5. Nucleic acid scaffolds for reconstitution of RNA polymerase II ECs.** Scaffold T and related scaffolds G, C and A were used in incorporation assays.

The +1 and +2 nucleotides were identical, to prevent misincorporation by template misalignment (Kashkina et al, 2006). To compare the efficiency of all 16 incorporation events (four correct incorporations and 12 misincorporations), the four scaffolds were assembled with RNA polymerase II into ECs that were incubated with 0.1 mM of each NTP. Reactions were stopped at 0.5, 1, or 5 minutes, and product RNAs were separated by gel electrophoresis and quantified with a fluorimager (Fig. 6, Experimental procedures).



**Figure 6. Systematic quantitative analysis of misincorporation, RNA extension, and RNA cleavage** Representative gel electrophoresis separation of RNA products obtained in incorporation assays. Lanes 1, 19, 37 and 55 show the fluorescently labeled reactant RNA. ECs of samples shown in lanes 2, 20, 38 and 56 were incubated for 5 min in transcription buffer without addition of NTPs (Experimental procedures). Run-off controls after incubation with 100  $\mu$ M NTPs for 5 min are shown in lanes 3, 21, 39 and 57. In the other lanes, the scaffolds were incubated with the indicated NTPs for 0.5, 1, and 5 min (left to right).

The relative amounts of misincorporation with respect to correct incorporation are provided in figure 7.

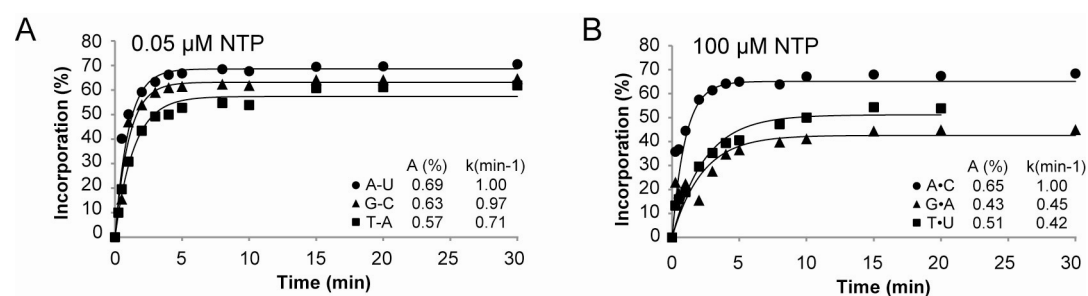


**Figure 7. Systematic quantitative analysis of match and mismatch incorporation.** Summary of incorporation efficiencies determined by addition of 100  $\mu$ M of the indicated NTP to the EC. Light grey, grey, and dark grey bars represent the 0.5 min, 1 min and 5 min time points, respectively. Average values are shown for two independent experiments that generally resulted in very similar values, indicating the high reproducibility.

DNA•RNA mismatches are indicated with a dot throughout. Misincorporations generating a purine•purine mismatch occurred with low efficiency, whereas those generating a pyrimidine•pyrimidine mismatch were more efficient, except for the C•C mismatch (Fig 7). No general rule could be derived for misincorporations resulting in



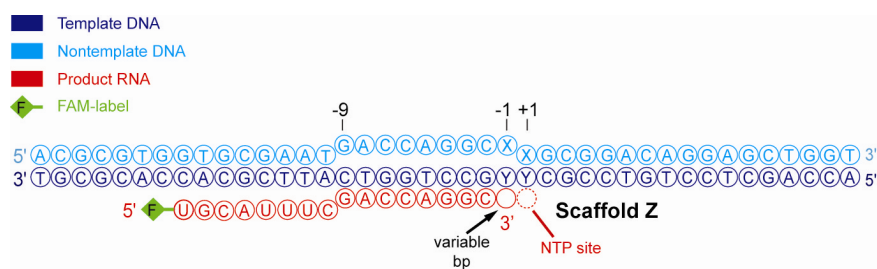
purine•pyrimidine and pyrimidine•purine mismatches. Misincorporations resulting in T•G or G•U mismatches were inefficient, but those resulting in C•A or A•C mismatches were efficient. To determine first order rate constants, we performed time course experiments for three types of misincorporations that were representative for low (G•A), medium (T•U), and high (A•C) efficiencies (Fig. 8B), and for their corresponding correct incorporations (Fig. 8A). Compared to correct incorporations, the misincorporations leading to G•A, T•U, and A•C mismatches was 4300-, 3400-, and 2000-fold slower, respectively (Experimental procedures). Thus, RNA polymerase II misincorporation efficiencies depend on the type of the resulting mismatch.



**Figure 8. Time course experiments for selected incorporation reactions.** For correct incorporations, 0.05 μM NTPs were used (A). For misincorporations, 100 μM NTPs were used (B). The pre-exponential factor A and the rate constant k were calculated with the program OriginPro 8 (ADDITIVE GmbH) using the equation  $c(t) = A \cdot (1 - \exp(-k \cdot t))$ . For comparison of rate constants of correct incorporation and misincorporation, a dilution factor of 2000 was applied, assuming that reduction of NTP concentration (from 100 to 0.05 μM) leads to equivalent reduction of the rate constant, as described (Alic et al, 2007).

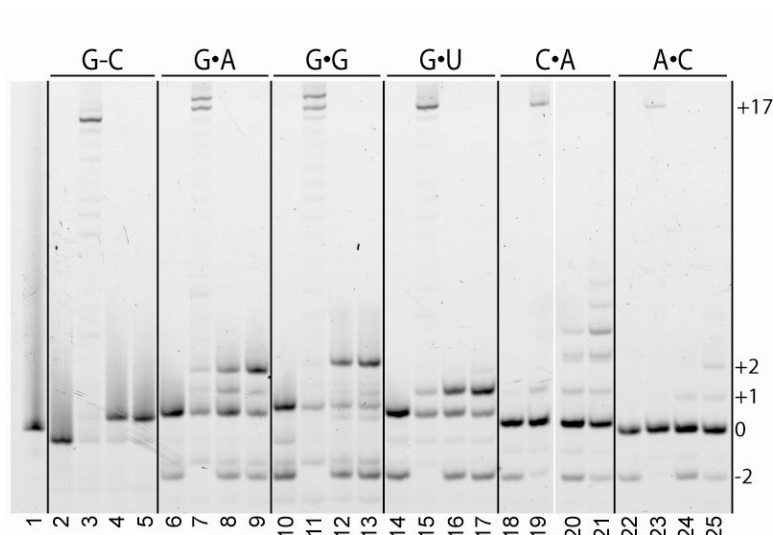
## 2.2 Transcript extension efficiency is mismatch-specific

To investigate the efficiency of RNA extension after misincorporation, ECs were reconstituted that contained the 12 different mismatches at position -1 (scaffold Z, Fig. 9).



**Figure 9. Nucleic acid scaffold for reconstitution of RNA polymerase II Ecs.** Scaffold Z is variable and was used for extension and cleavage assays. The variable bp (black arrow) was one of the sixteen different matched or mismatched bps, to mimick the result of all 16 possible (mis)incorporation events obtained with scaffolds T, G, C and A (Fig. 5).

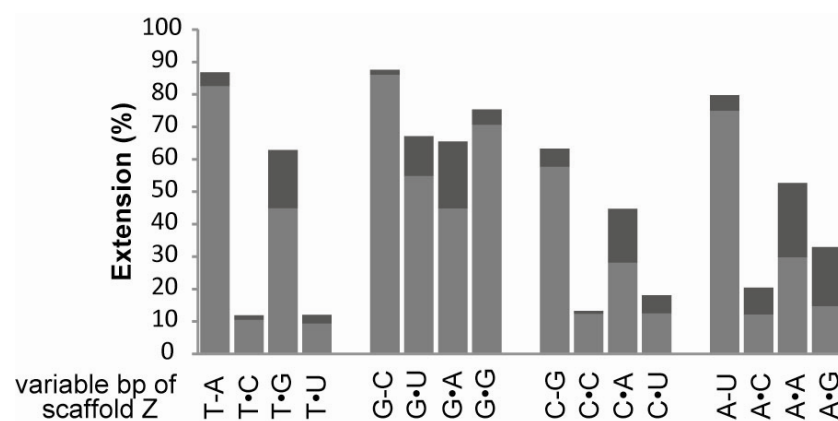
These ECs mimic the situation after misincorporation, and allow monitoring the addition of the next nucleotide. For RNA extension, we added the next complementary NTP, and stopped reactions at 1 or 5 minutes (Fig. 10, Experimental procedures).



**Figure 10. Representative electrophoretic separation of RNA products resulting from RNA extension and cleavage.** Six examples are shown for which the bp at position -1 of scaffold Z (Fig. 9) is given. Lane 1 shows the fluorescently labeled reactant RNA. Each block of four lanes shows from left to right the cleavage experiment, the run-off control, and extension experiments stopped after 1 and 5 min of incubation. For RNA extension, ECs were incubated with 100  $\mu$ M of the corresponding next correct NTP.

To prevent extension after RNA dinucleotide cleavage as a side reaction, the nucleotides at -2 and +1 were different. Incorporation of the next nucleotide was

always less efficient when a mismatch was present at -1 instead of a match (Fig. 10, 11). Purine•purine mismatches were more efficiently extended than pyrimidine•pyrimidine mismatches (Fig. 11). Extensions with a mismatched guanine in the RNA were all efficient, generally consistent with results for a bacterial RNA polymerase (Zenkin et al, 2006). Amongst the pyrimidine•purine and purine•pyrimidine mismatches, extension was more efficient for T•G and G•U, and less efficient for C•A and A•C. Extensions with a guanine at template position -1 were all efficient. Control experiments showed that the efficiency of incorporating a nucleotide following a matched bp was very similar for the different bps (Fig. 11). Thus, the efficiency of RNA extension is always lower in the presence of a mismatch, but varies with the type of mismatch.

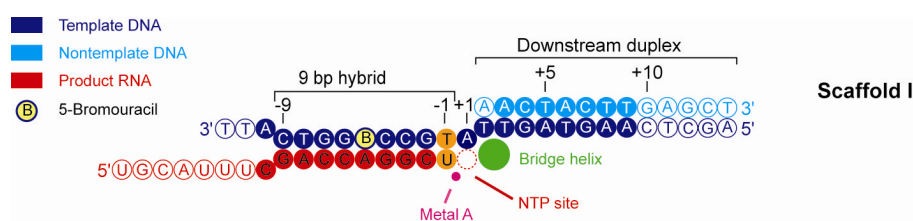


**Figure 11. Summary of RNA extension efficiencies.** Grey and dark grey bars represent 1 min and 5 min time points, respectively. Average values for two independent experiments are shown.

### 3. STRUCTURAL BASIS OF MISMATCH SPECIFICITY

#### 3.1 RNA polymerase II accommodates a T•U wobble pair

To unravel the molecular basis of fidelity mechanisms for one type of mismatch, we determined structures of complete RNA polymerase II ECs containing a T•U mismatch. Complete RNA polymerase II was co-crystallized with a scaffold containing the mismatch at position  $-1$  (scaffold I, Fig. 12).



**Figure 12. Nucleic acid scaffold I for reconstitution of an RNA polymerase II EC.** Scaffold I contains a T•U mismatch at position  $-1$  (orange) and was used for structural analysis. Filled circles denote nucleotides with interpretable electron density.

For the resulting EC I, diffraction data of very high quality were obtained, and the register of nucleic acids was defined by bromine labeling (Table 3, Fig. 13, Experimental procedures). With the use of zonal scaling (Vassylyev et al, 2007a), the structure was refined to a free R-factor of 25.2% at 3.2 Å resolution, the highest resolution for a complete RNA polymerase II structure (Fig. 13A-C, Table 3).

The structure showed that EC I adopts the post-translocation state and accommodates the T•U mismatch at the active center at position  $-1$  (Fig. 13A, C and Fig. 14). The mismatch adopts a wobble bp that is stabilized by two hydrogen bonds formed between the N3 and the O2 atoms of the uracil and the O4 and N3 atoms, respectively, of the template thymine (Fig. 14). The accommodation of a wobble pair may explain why uridine misincorporation opposite a template thymine is efficient

(Fig. 7), and supports our previous proposal that uridine misincorporation opposite a thymine within a DNA photolesion results from wobble formation (Brueckner et al, 2007).

**Table 3. Diffraction data and refinement statistics**

	EC I	EC II	EC III	EC IV	EC V	EC VI
<b>Data collection</b>						
Space group	C2	C222 <sub>1</sub>	C222 <sub>1</sub>	C222 <sub>1</sub>	C222 <sub>1</sub>	C222 <sub>1</sub>
Unit cell axes (Å)	394.3, 221.6, 283.4	222.3, 393.4, 283.1	222.7, 396.0, 283.5	221.4, 393.8, 281.8	221.6, 393.7, 282.6	222.1, 392.7, 282.4
Unit cell β angle (°)	90.9	90	90	90	90	90
Wavelength (Å)	0.9189	0.9190	0.9188	0.9188	0.9188	0.9177
Resolution range (Å)	40-3.20	50-3.50	50-3.60	50-3.65	50-3.65	50-3.40
Unique reflections	372,166 <sup>1</sup> (32,852) <sup>2</sup>	155,150 (21,507)	144,009 (19,441)	135,977 (18,105)	136,470 (18,185)	168,339 (24,019)
Completeness (%)	95.6 (84.7)	99.9 (100)	99.9 (99.9)	99.9 (100)	99.9 (100)	99.9 (100)
Redundancy	3.0 (2.2)	7.3 (7.2)	7.3 (7.3)	7.6 (7.8)	7.5 (7.4)	7.5 (7.9)
Mosaicity (°)	0.38-0.72 <sup>3</sup>	0.11	0.14	0.12	0.09	0.08
R <sub>sym</sub> (%)	7.5 (37.5)	9.5 (52.9)	9.2 (75.0)	7.7 (63.6)	6.6 (52.0)	6.4 (50.4)
I/σ(I)	20.7 (2.6)	15.8 (4.7)	17.6 (3.2)	22.0 (3.7)	22.9 (4.3)	24.1 (5.0)
<b>Refinement</b>						
Nonhydrogen atoms	63,666	31,778	31,877	31,962	31935	31,804
RMSD bonds	0.010	0.010	0.011	0.010	0.011	0.010
RMSD angles	1.60	1.59	1.65	1.61	1.65	1.61
R <sub>cryst</sub> (%)	23.3	21.0	21.4	21.0	21.2	21.6
R <sub>free</sub> (%)	25.2	22.6	25.4	25.3	25.0	25.4
Br peak in anom. Fourier (σ)	8.6	8.6	10.9	8.1	8.7	9.6
<b>Ramachandran statistics</b>						
Core (%)	71.8 <sup>4</sup> / 72.0 <sup>5</sup>	71.1	72.7	72.6	70.6	74.0
Allowed (%)	23.4/23.3	24.1	22.6	22.9	24.6	21.7
Generally allowed (%)	3.1/3.0	3.2	3.2	2.6	3.2	3.1
Disallowed (%)	1.7/1.7	1.6	1.5	1.9	1.6	1.2

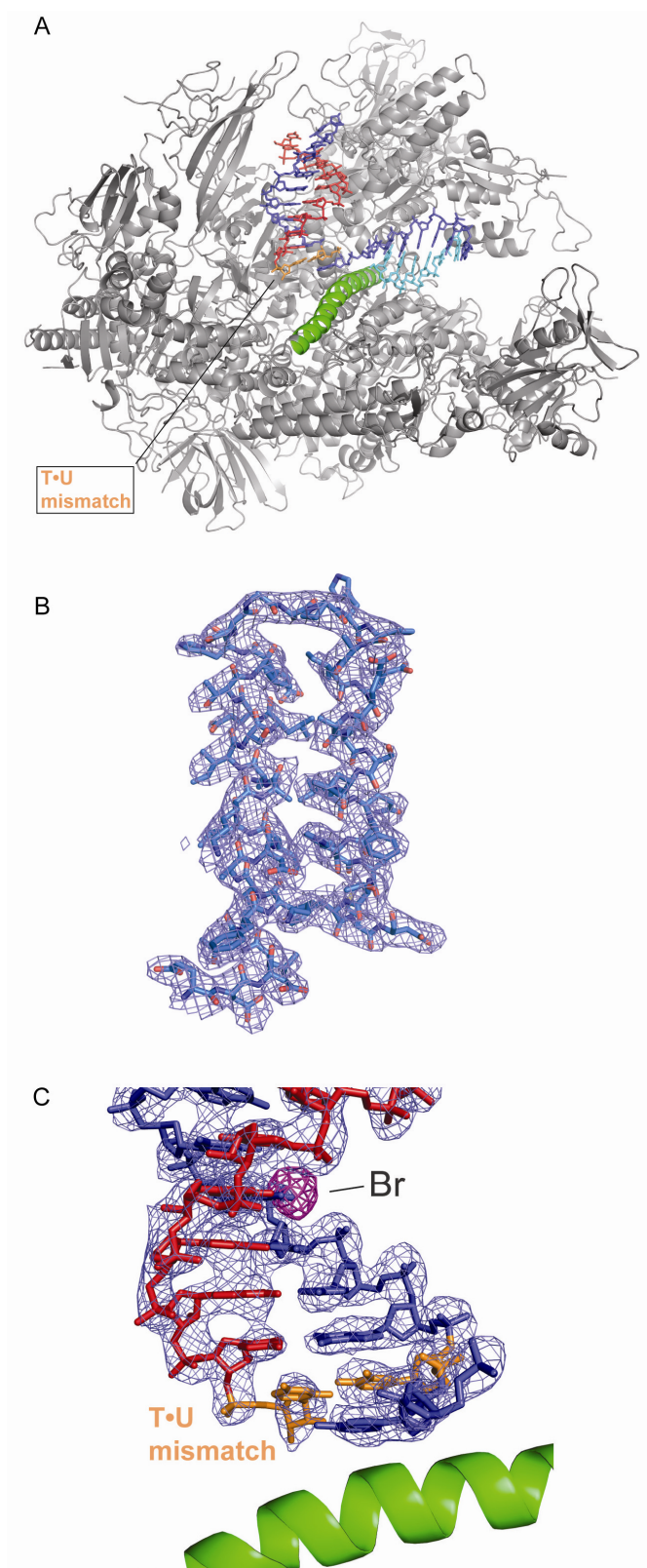
<sup>1</sup> Friedel pairs are merged

<sup>2</sup> Values in parentheses are for highest resolution shell

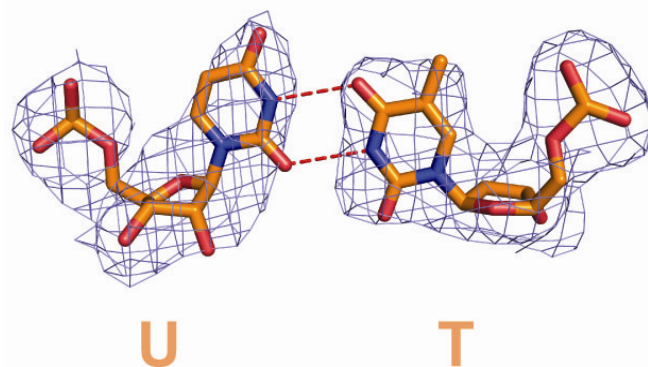
<sup>3</sup> Refined for batches of images

<sup>4</sup> Molecule 1 of the asymmetric unit

<sup>5</sup> Molecule 2 of the asymmetric unit



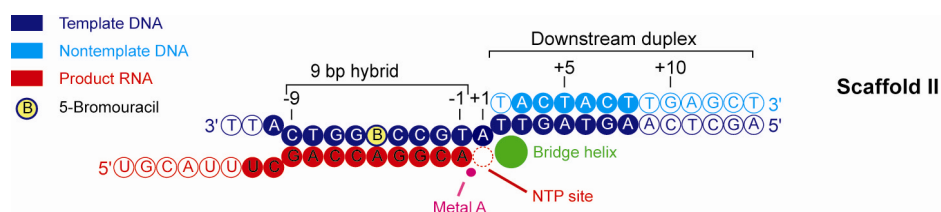
**Figure 13. Structure of EC I reveals a T•U wobble pair at 3.2 Å resolution** (A) Structure of the T•U mismatch-containing RNA polymerase II EC I. RNA polymerase II is shown from the side as a ribbon model in silver, with the bridge helix highlighted in green, and a portion omitted for clarity. The nucleic acids are shown as stick models using the same color code as in Fig. 12. The T•U mismatch is shown in orange throughout. (B) Representative protein electron density. The final 2F<sub>o</sub>-F<sub>c</sub> density is shown as a blue mesh, contoured at 1.1σ. Depicted is the clamp coiled-coil, an exposed part of subunit Rpb1. (C) Electron density of part of the DNA-RNA hybrid (2F<sub>o</sub>-F<sub>c</sub> map contoured at 1.8σ). A peak in the anomalous difference Fourier map (magenta, contoured at 4.3σ) reveals the location of the bromine atom at position -5 of the template strand, defining the post-translocated state.



**Figure 14. T•U wobble base pair in the RNA polymerase II active center.** The final  $2F_o - F_c$  electron density map is shown in blue, contoured at  $1.0\sigma$ . Hydrogen bonds are indicated by red dashed lines.

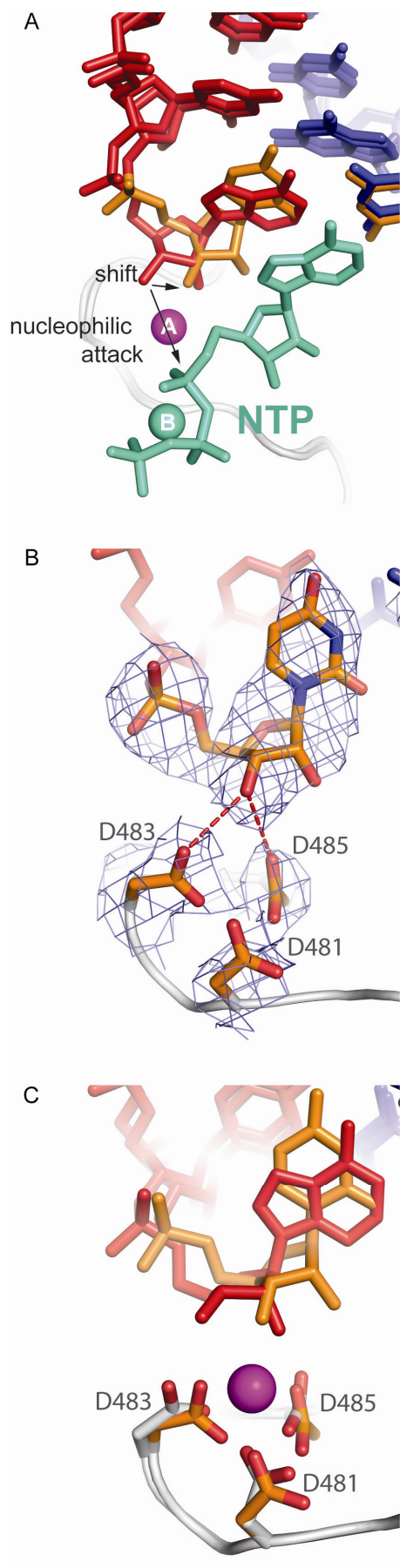
### 3.2 Active site disruption explains impaired RNA extension

To detect the structural changes in EC I that result from the T•U mismatch, we solved a reference structure that contained a matched T-A bp at position -1 (scaffold II, EC II, Fig. 15, Table 3).



**Figure 15. Nucleic acid scaffold II for reconstitution of an RNA polymerase II EC.** Scaffold II contains a T-A match at position -1. Filled circles denote nucleotides with interpretable electron density.

The overall structures of ECs I and II did not deviate, but in the mismatched EC I the 3' terminal RNA nucleotide at position -1 and its 5'-flanking phosphate were shifted away from the active site by over 2 Å (Fig. 16A). Thus, the T•U wobble triggers misalignment of the nucleophilic RNA 3' end with the catalytic site and NTP, and a deviation from the optimum geometry for catalysis, a collinear in-line attack during an  $S_N2$  reaction (Fig. 16A).



**Figure 16. Active site disruption by accommodation of a T•U mismatch.** (A) Superposition of the mismatched EC I with the matched EC II (at 3.5 Å resolution, Table 3) reveals a 2 Å shift of the RNA 3' hydroxyl (horizontal arrow, the mismatched terminal RNA U residue is shown in orange). As a consequence, the nucleophilic RNA 3' end is no longer in a position suited for an in-line nucleophilic attack (vertical arrow) of the phosphodiester bond between the  $\alpha$  and  $\beta$  phosphates of the incoming NTP substrate (green cyan, taken from PDB-code 2O5J (Vassylyev et al, 2007b)). The structures EC I and 2O5J were superimposed by least squares fitting of Rpb1 residues A478-A487 to  $\beta'$  residues D745-D736 and RNA residues in positions -1, -2 and -3. Metal ion A is from EC II and metal ion B is from 2O5J. For NTP modeling, we used the bacterial NTP complex structure rather than the yeast core RNA polymerase II NTP complex since it contains an intact RNA 3'hydroxyl group. (B) Loss of metal ion A in the active site of EC I. The final  $2F_o - F_c$  electron density map is contoured at  $1.0\sigma$ . (C) Comparison of the RNA 3' nucleotide and the catalytic aspartate loop in EC I (orange) and EC II (grey). Metal A (pink sphere) is only present in EC II.

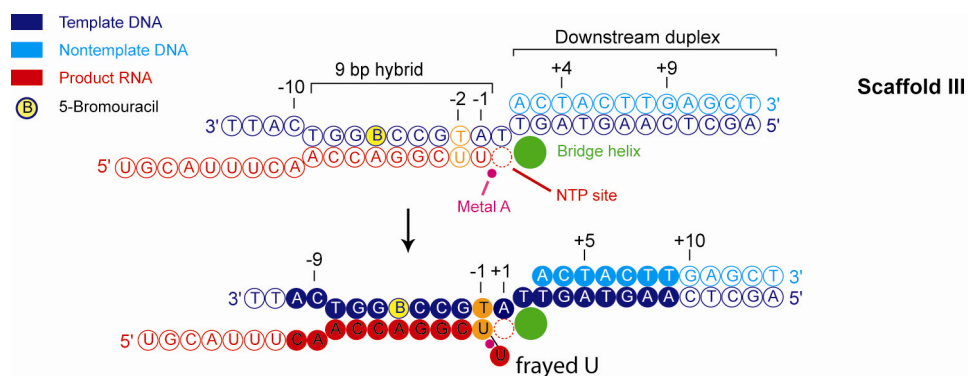


In addition, the active site aspartate loop lost the catalytic metal ion A (Fig. 16B). The three metal-binding aspartate side chains in Rpb1 changed conformation (Fig. 16C). The D481 carboxylate is mobile and the side chains of D483 and D485 could both form a hydrogen bond with the RNA 3' hydroxyl (Fig. 16B). Metal A is apparently lost due to the disruption of the active site by the wobble bp since it is observed in EC II and in a published EC structure obtained under the same conditions (Brueckner & Cramer, 2008). Thus, the low efficiency of RNA extension after a T•U mismatch can be explained by disruption of the catalytic site that involves loss of the catalytic metal A and a shift of the RNA 3' end.

## 4. RNA POLYMERASE PAUSING

### 4.1 Mismatch extension and RNA 3' fraying

To investigate RNA extension past the mismatch, we prepared a scaffold with the T•U mismatch at position -2 and an A-U bp at position -1 (scaffold III, Fig. 17).

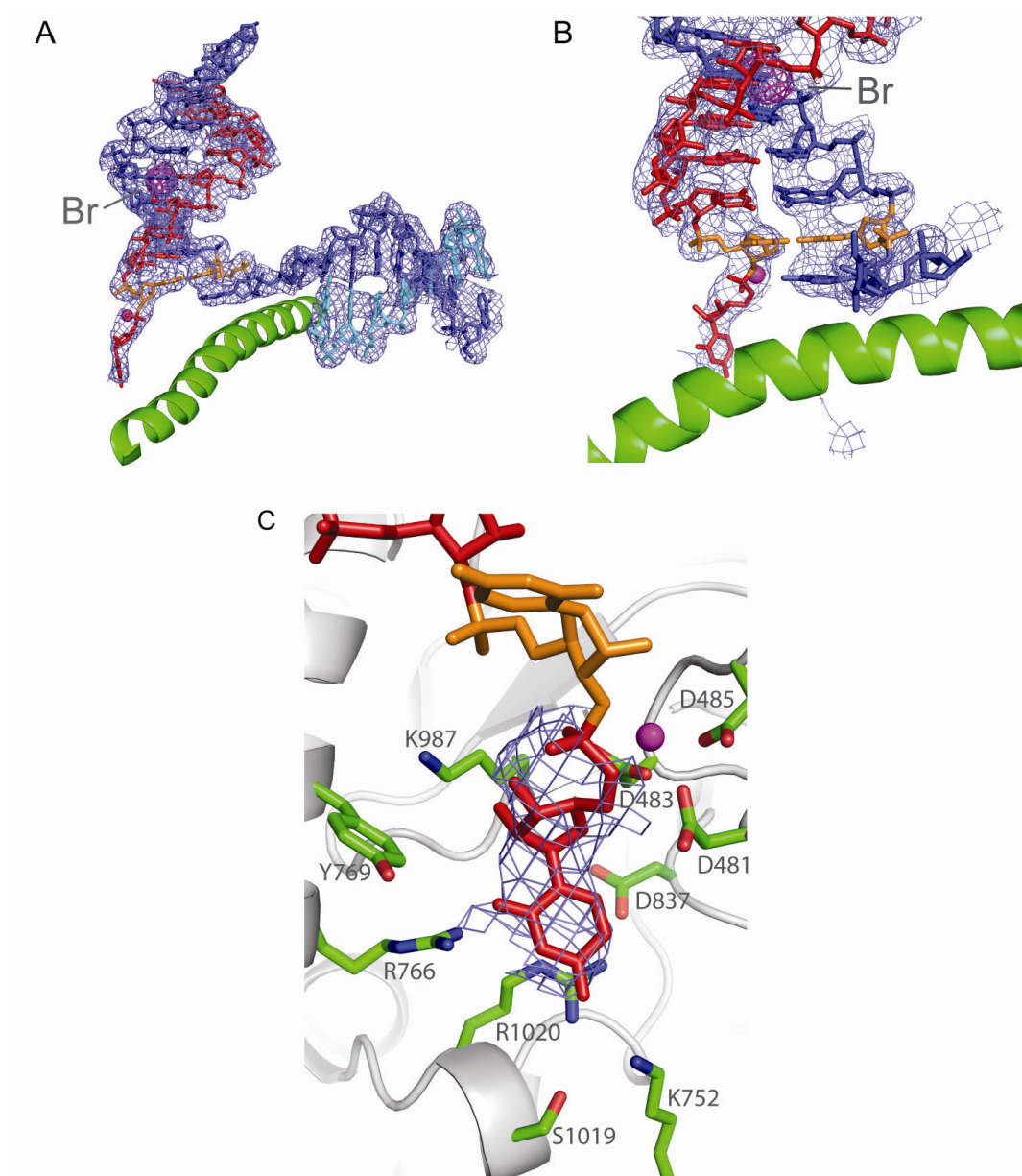


**Figure 17. Nucleic acid scaffold III for reconstitution of an RNA polymerase II EC.** Scaffold III as designed (top), containing a T•U mismatch at position -2 (orange) and an A-U match at position -1, and as observed in the crystal (bottom) with a frayed 3' uridine. Filled circles denote nucleotides with interpretable electron density.

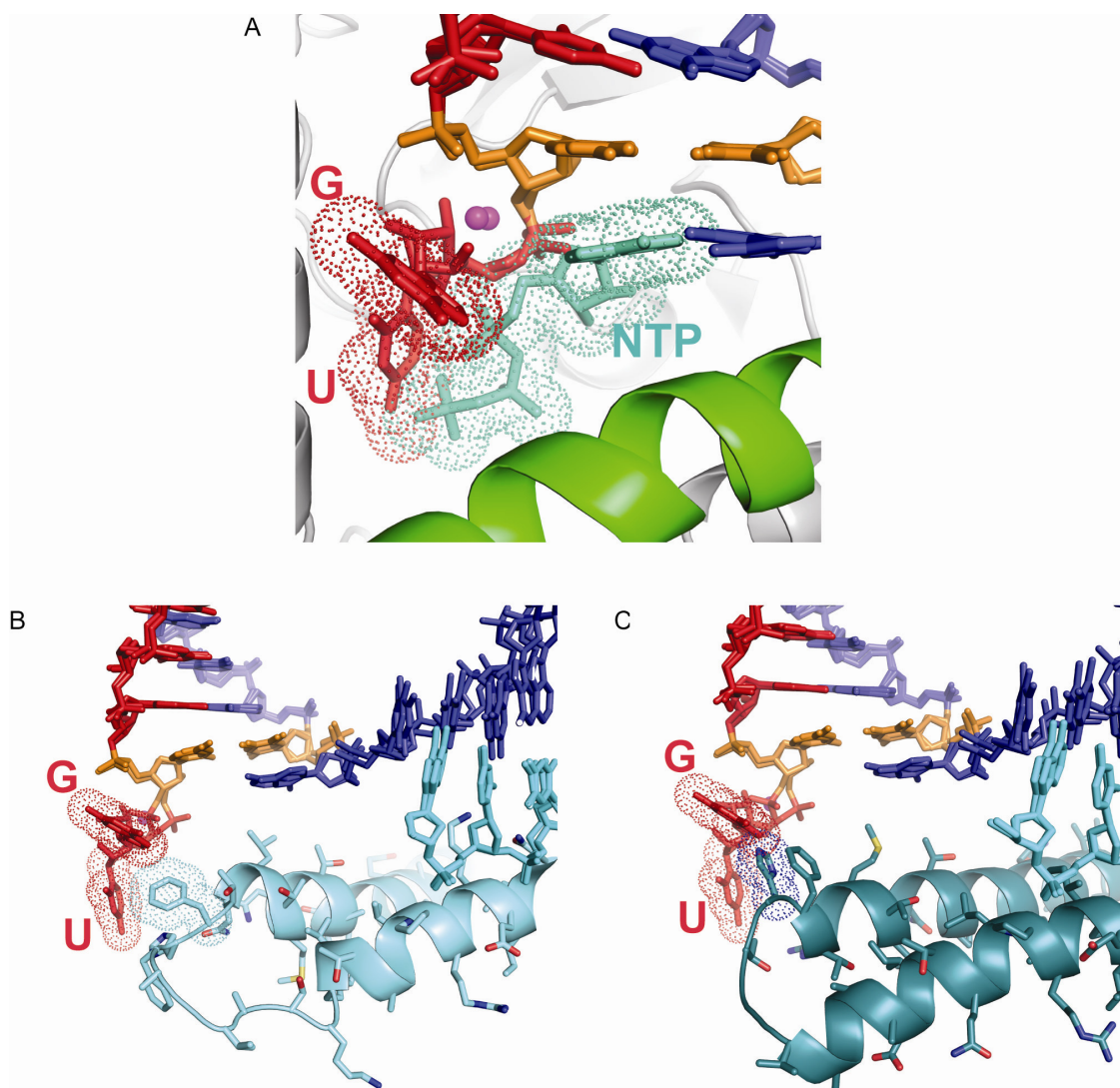
In the resulting EC III structure (Table 3), the hybrid was similar to that in EC I, including the T•U wobble bp at position -1, and downstream DNA was slightly shifted as previously observed (Brueckner et al, 2007). The 3' terminal RNA uridine however did not form a bp with the template adenine as designed, but was flipped away from the template, creating a frayed RNA end (Fig. 18A, B). The frayed uracil was oriented parallel to the axis of the DNA-RNA hybrid, and occupied a site in the pore (“fraying site I”, Fig. 18C).

A frayed RNA 3' nucleotide was shown biochemically to be the hallmark of a common elongation intermediate, the elemental pause, that occurs during polymerase pausing, and before transcription arrest and termination (Artsimovitch & Landick, 2000; Chan et al, 1997; Touloukhonov et al, 2007). The frayed nucleotide

overlaps the tip of the closed trigger loop and the NTP in the insertion site (Figs. 18C and 19), and contacts Rpb2 residues R766 and R1020, which also bind the NTP triphosphate (Table 4). This explains how the frayed RNA end interferes with nucleotide binding and incorporation.



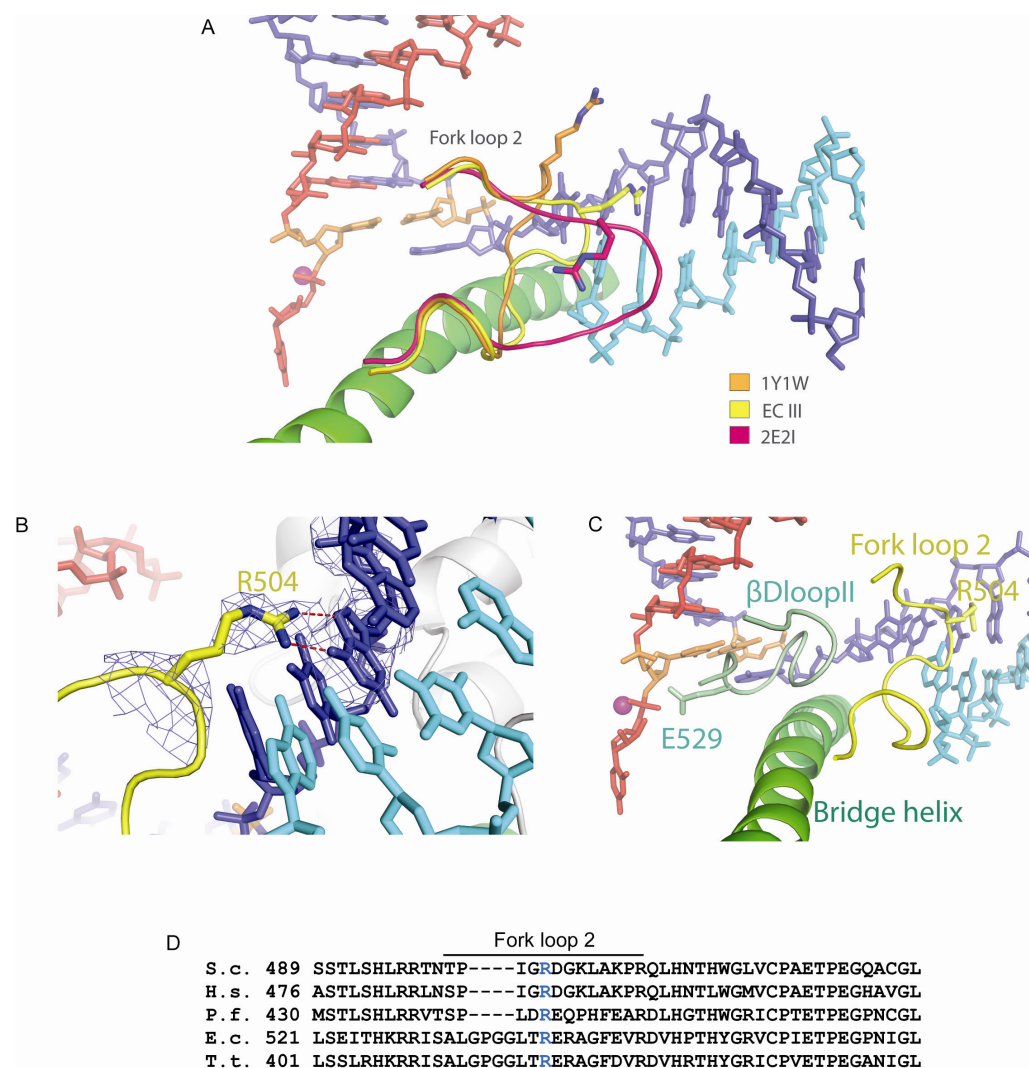
**Figure 18. Structure of RNA polymerase II EC III reveals a frayed 3' terminal RNA uridine at 3.6 Å resolution.** (A) The final  $2F_o - F_c$  electron density of the nucleic acids is shown as a blue mesh, contoured at  $1.2\sigma$ . The location of the bromine atom at position -5 defines the register (the anomalous difference Fourier map is shown in magenta, contoured at  $4.2\sigma$ ). (B) Detailed view of the electron density map in (A) near the active center. (C) Fraying site I. Depicted are RNA polymerase II residues contacting the frayed 3' terminal RNA uridine. The final  $2F_o - F_c$  density is shown for the frayed nucleotide, contoured at  $0.9\sigma$ .



**Figure 19. Frayed nucleotides overlap the NTP and the closed trigger loop.** (A) Frayed nucleotides (ECs III and IV) overlap the NTP bound to the insertion site (green cyan, taken from bacterial RNA polymerase EC, PDB-code 2O5J (Vassilyev et al, 2007b)). Van der Waals radii are represented by colored dots. All structures were superimposed with their active site regions. (B-C) Frayed nucleotides overlap the closed trigger loop (cyan) at residue F1084 (B, taken from the RNA polymerase II EC, PDB-code 2E2H (Wang et al, 2006)) and/or at residue H1242 (C, bacterial RNA polymerase EC, PDB-code 2O5J (Vassilyev et al, 2007b)).

In EC III, fork loop 2 adopts a new conformation (Fig. 20). Fork loop 2 residues have moved by up to 6 Å towards the DNA nontemplate strand at the downstream edge of the transcription bubble (Fig. 20A). The guanidinium head group of Rpb2 residue R504 forms two hydrogen bonds to N7 and O6 of the template guanine at +4 (Fig. 20B). R504 is invariant among RNA polymerase II enzymes and bacterial and

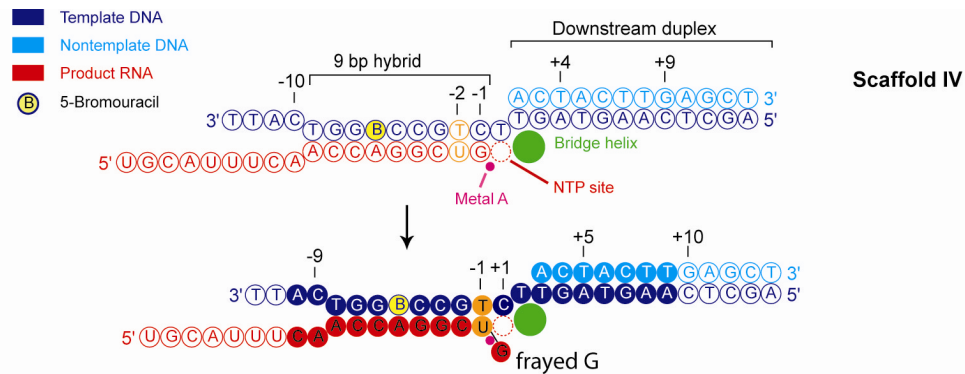
archaeal RNA polymerases, but not conserved in RNA polymerase I and III (Fig. 20D) (Jasiak et al, 2006; Kuhn et al, 2007; Naji et al, 2007). This arginine is important for promoter-dependent transcription and normal elongation (Naji et al, 2007). It is possible that the observed fork loop 2-downstream DNA interaction, or alternative contacts of the flexible arginine (Fig. 20C) with other nearby bases in DNA, contribute to the stability of the paused state as suggested (Touloukhonov et al, 2007).



**Figure 20. Fork loop 2-downstream DNA contact.** (A) Comparison of the conformation of fork loop 2 in EC III with that in previous RNA polymerase II EC structures (PDB-codes 1Y1W (Kettenberger et al, 2004) and 2E2I (Wang et al, 2006)). (B) Interaction of the side chain of fork loop 2 Rpb2 residue R504 with the guanine base at position +4 of downstream DNA. The final  $2F_o - F_c$  electron density is shown in blue, contoured at  $0.7\sigma$ . (C) Interaction of regions in EC III that may be involved in pausing, including the frayed nucleotide,  $\beta$ DloopII, the bridge helix, fork loop 2, and downstream DNA. (D) Multiple sequence alignment of fork loop 2 and surrounding Rpb2 residues from *S. cerevisiae*, *H. sapiens*, *P. furiosus*, *E. coli* and *T. thermophilus* (CLUSTAL W). The conserved R504 from *S. cerevisiae* is highlighted in blue.

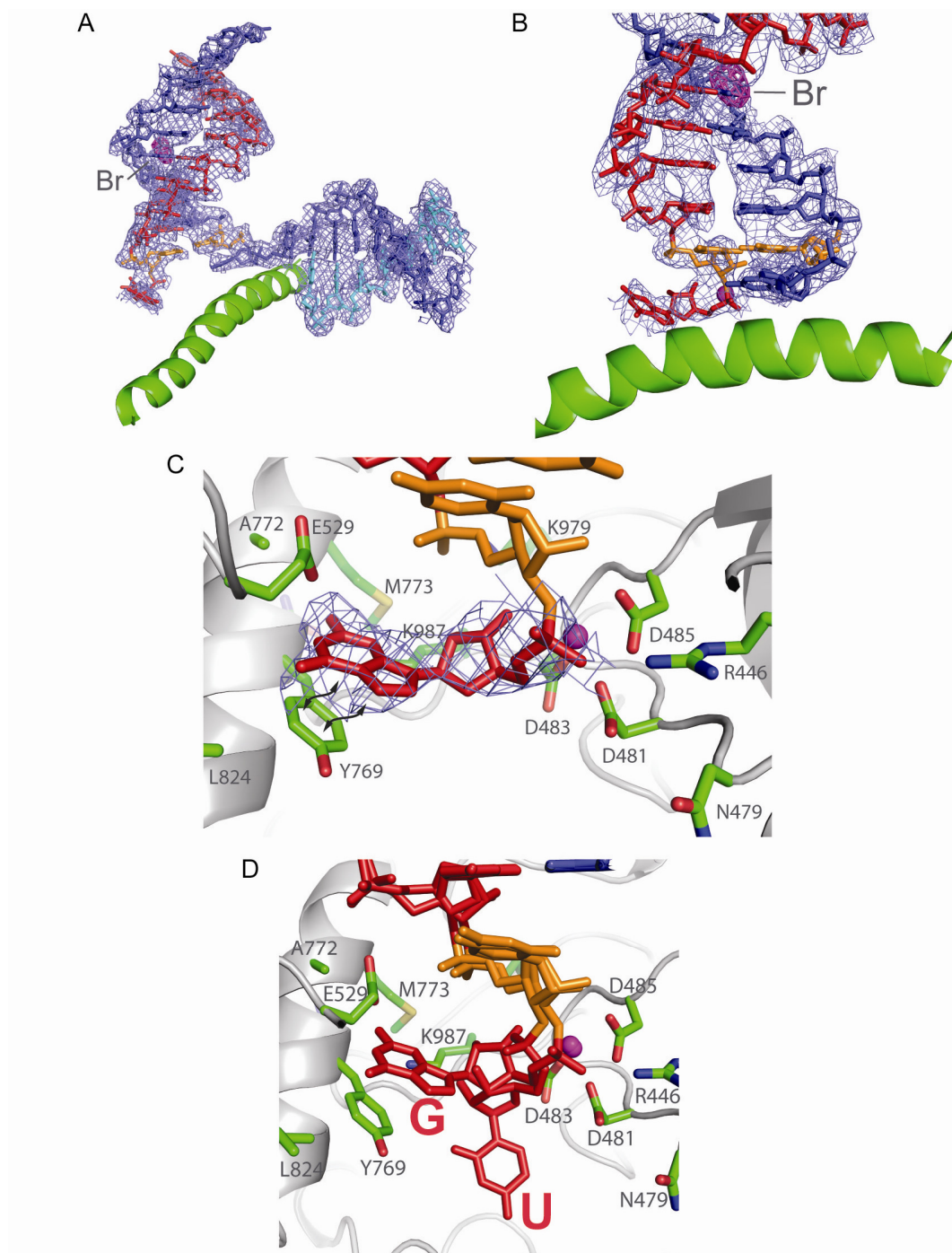
## 4.2 Two RNA fraying sites

To test whether the fraying was dependent on the stability of the bp at the end of the hybrid, we replaced the A-U bp in scaffold III with a C-G bp (scaffold IV, Fig. 21).



**Figure 21. Nucleic acid scaffold IV for reconstitution of an RNA polymerase II EC.** Scaffold IV as designed (top), containing a T•U mismatch at position -2 (orange) and a G-C match at position -1, and as observed in the crystal (bottom) with a frayed 3' guanine. Filled circles denote nucleotides with interpretable electron density.

The resulting EC IV structure was very similar to that of EC III, including the T•U wobble bp (Fig. 22A, B). The RNA 3' nucleotide was again frayed, but was oriented perpendicular to the hybrid axis, occupying a different site in the pore (“fraying site II,” Fig. 22C). Fraying sites I and II are both lined by Rpb1 residues K987 and D483, but are separated by Rpb2 residue Y769, which stacks against the frayed guanine (Fig. 22D). The frayed guanine contacts Rpb2 residue E529 in a region called  $\beta$ DloopII in bacterial RNA polymerase (Table 4). Thus, the RNA 3' nucleotide can occupy at least three alternative sites, the pre-translocated position, which preserves base pairing with the template, and two alternative fraying sites in the pore, in which this base pairing is disrupted. The frayed nucleotide is either oriented along the hybrid axis and approaches the NTP triphosphate-binding site (fraying site I), or it is oriented perpendicular to the hybrid axis and approaches  $\beta$ DloopII (fraying site II) (Fig. 22D).



**Figure 22. Two frayed states of the RNA 3' nucleotide (D)** Nucleic acids structure of EC IV reveals a frayed 3' terminal RNA guanine at 3.65 Å resolution. The final  $2F_o - F_c$  is shown as a blue mesh, contoured at  $1.0\sigma$ . The bromine peak at position -5 defines the register (anomalous difference Fourier, magenta, contoured at  $4.5\sigma$ ). (E) Detailed view of the electron density map in (D) near the active center. (F) Fraying site II. Depicted are RNA polymerase II residues contacting the frayed 3' terminal RNA guanine. The final  $2F_o - F_c$  density is shown for the frayed nucleotide, contoured at  $1.0\sigma$ . Stacking interactions are indicated by two-headed arrows. (D) Superposition of the structures of ECs III and IV allows for comparison of the two frayed RNA 3' nucleotides that are either oriented parallel (U, fraying site I) or perpendicular (G, fraying site II) to the axis of the DNA-RNA hybrid (vertical in this view).

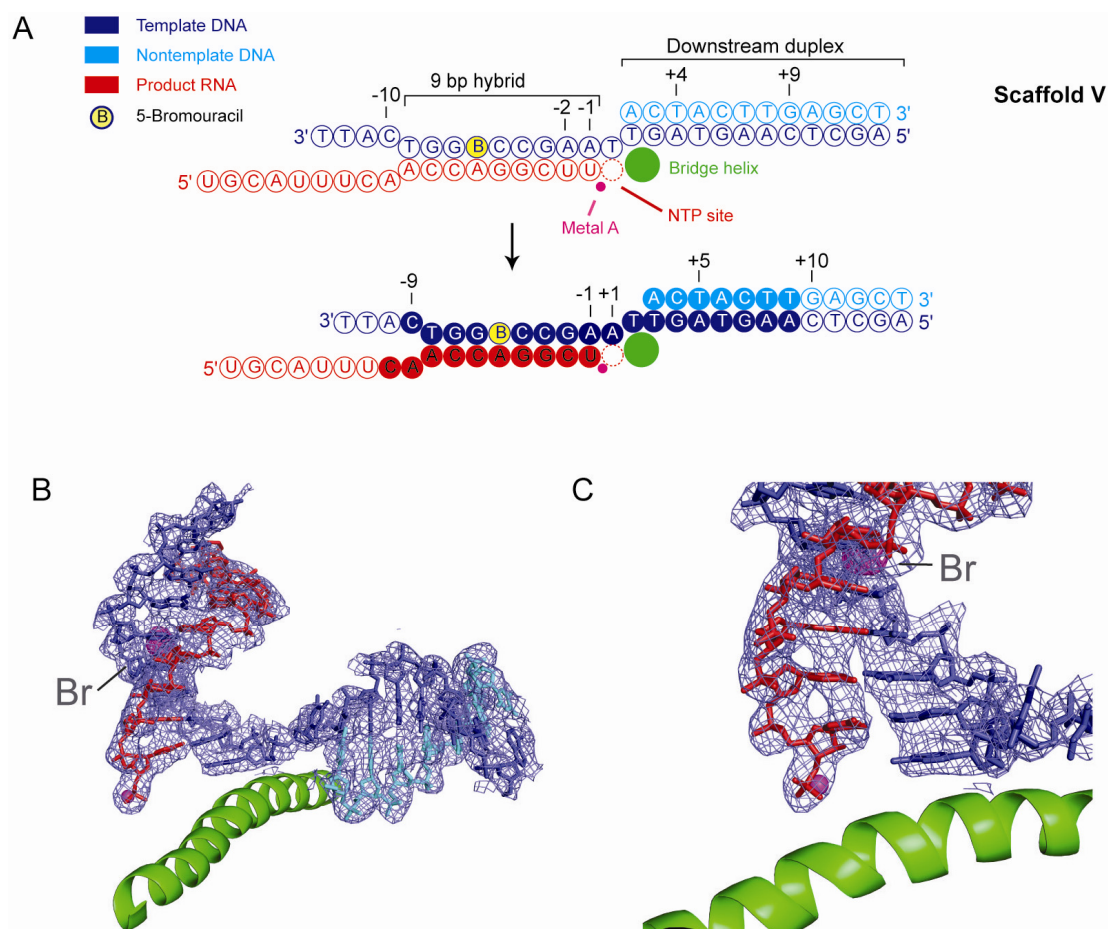
**Table 4. Amino acid residues contacting the frayed nucleotides in fraying site I (EC III) and fraying site II (EC IV) and their conserved counterparts in human RNA polymerase II (Pol II) and bacterial RNA polymerase**

<b>Fraying site I</b>				
<b>Pol II atom</b>	<b>RNA atom</b>	<b>Distance (Å)</b>	<b><i>H. sapiens</i></b>	<b><i>T. thermophilus</i></b>
RPB2 R766 NH1	Frayed U C4	3.0	R721	R557
RPB1 D483 OD2	Frayed U O4'	2.4	D497	D741
RPB2 R1020 CZ	Frayed U C4	3.8	R975	R879
RPB1 D483 CG	Frayed U C1'	3.5	D497	D741
RPB2 K987 CE	Frayed U C2'	3.5	K942	K846
RPB2 K987 CD	Frayed U C2'	3.8	K942	K846
RPB1 D483 CG	Frayed U C4'	3.7	D497	D741
RPB1 D481 CG	Frayed U C5'	3.8	D495	D743
<b>Fraying site II</b>				
<b>Pol II atom</b>	<b>RNA atom</b>	<b>Distance (Å)</b>	<b><i>H. sapiens</i></b>	<b><i>T. thermophilus</i></b>
RPB1 R446 NH2	Frayed G O1P	3.2	R460	R704
RPB2 E529 OE1	Frayed G N1	2.8	E516	E445
Penultimate RNA base U C5'	Frayed G C2'	3.8	-	-
Penultimate RNA base U C5'	Frayed G C3'	3.1	-	-
RPB2 K987 CE	Frayed G C1'	3.4	K942	K846
RPB2 K987 CE	Frayed G C2'	3.4	K942	K846
RPB2 Y769 CZ	Frayed G C6	3.8	Y724	M560
RPB2 Y769 CZ	Frayed G C5	3.8	Y724	M560
RPB2 Y769 CE2	Frayed G C6	3.2	Y724	M560
RPB2 Y769 CE2	Frayed G C5	3.7	Y724	M560
RPB2 Y769 CD2	Frayed G C6	3.4	Y724	M560
RPB2 Y769 CD2	Frayed G C2	3.5	Y724	M560
RPB2 Y769 CG	Frayed G C2	3.4	Y724	M560
RPB2 Y769 CB	Frayed G C2	3.6	Y724	M560
RPB2 Y769 CA	Frayed G C2	3.8	Y724	M560
RPB2 E529 CB	Frayed G C6	3.9	E516	E445
RPB1 D483 CG	Frayed G C4'	3.7	D497	D741

To test whether fraying was due to the T•U mismatch, we replaced the mismatch in scaffold III by a correct A-U bp (Fig. 23A). The resulting EC V structure (Table 3, Fig. 23B) revealed electron density for the RNA -1 uridine and for the phosphate of the



RNA nucleotide at position +1 (Fig. 23C), but not for the terminal uracile base and ribose at register +1, which are mobile. These observations suggest that two uridine residues at the RNA 3' terminus, which are present at canonical pause sites, destabilize the bp at +1 and favor a frayed state, which can be stabilized at specific locations by a T·U mismatch at position -1 and can then be observed crystallographically.

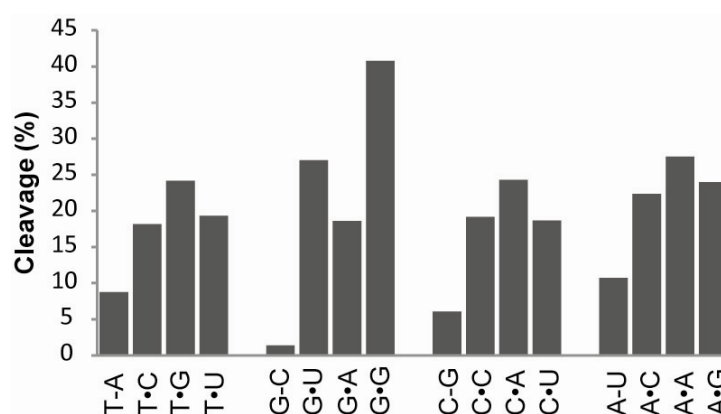


**Figure 23. Structure of EC V at 3.65 Å resolution.** (A) Nucleic acid scaffold V for reconstitution of an RNA polymerase II EC. Scaffold V as designed (top), containing A-U matches at positions -2 and -1, and as observed in the crystal (bottom). Filled circles denote nucleotides with interpretable electron density. (B) Structure of EC V reveals a mobile 3' terminal nucleotide. The final  $2F_o - F_c$  electron density is contoured at  $1.0\sigma$  (blue) and shows only the phosphate group of the 3' nucleotide. The location of the bromine atom at position -5 defines the register (anomalous difference Fourier, magenta, contoured at  $4.3\sigma$ ). (C) Detailed view of the electron density map in (B).

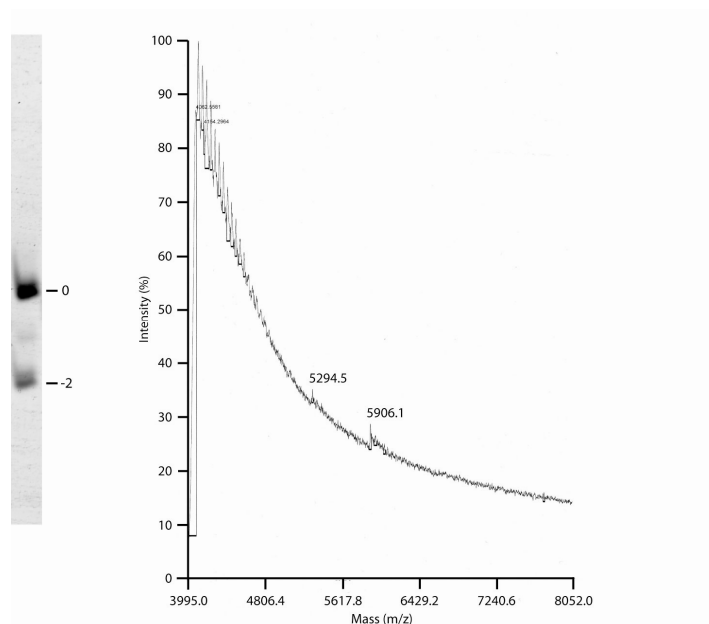
## 5. RNA POLYMERASE II PROOFREADING

### 5.1 Nucleotide-specific cleavage of mismatched RNA ends

The above results rationalize slow mismatch extension, which is a prerequisite for RNA cleavage during proofreading (Erie et al, 1993; Thomas et al, 1998). To investigate RNA polymerase II cleavage efficiency for different mismatches, we incubated the ECs used for extension assays with standard transcription buffer containing 8 mM magnesium ions (Fig. 10 and 24). Cleavage of dinucleotides was generally observed (Fig. 10), and confirmed by MALDI mass spectrometry of the RNA products (Fig. 25). Most efficient cleavage was observed for G•G, A•A, G•U, T•G, A•G, A•C, and G•A mismatches. RNA in the mismatched ECs was always more efficiently cleaved than in the matched ECs, and cleavage was very efficient for those mismatches that support extension (Fig. 11 and 24).



**Figure 24. Summary of RNA cleavage efficiencies.** For these experiments, ECs were incubated in transcription buffer for 5 min. Average values for two independent experiments are shown.

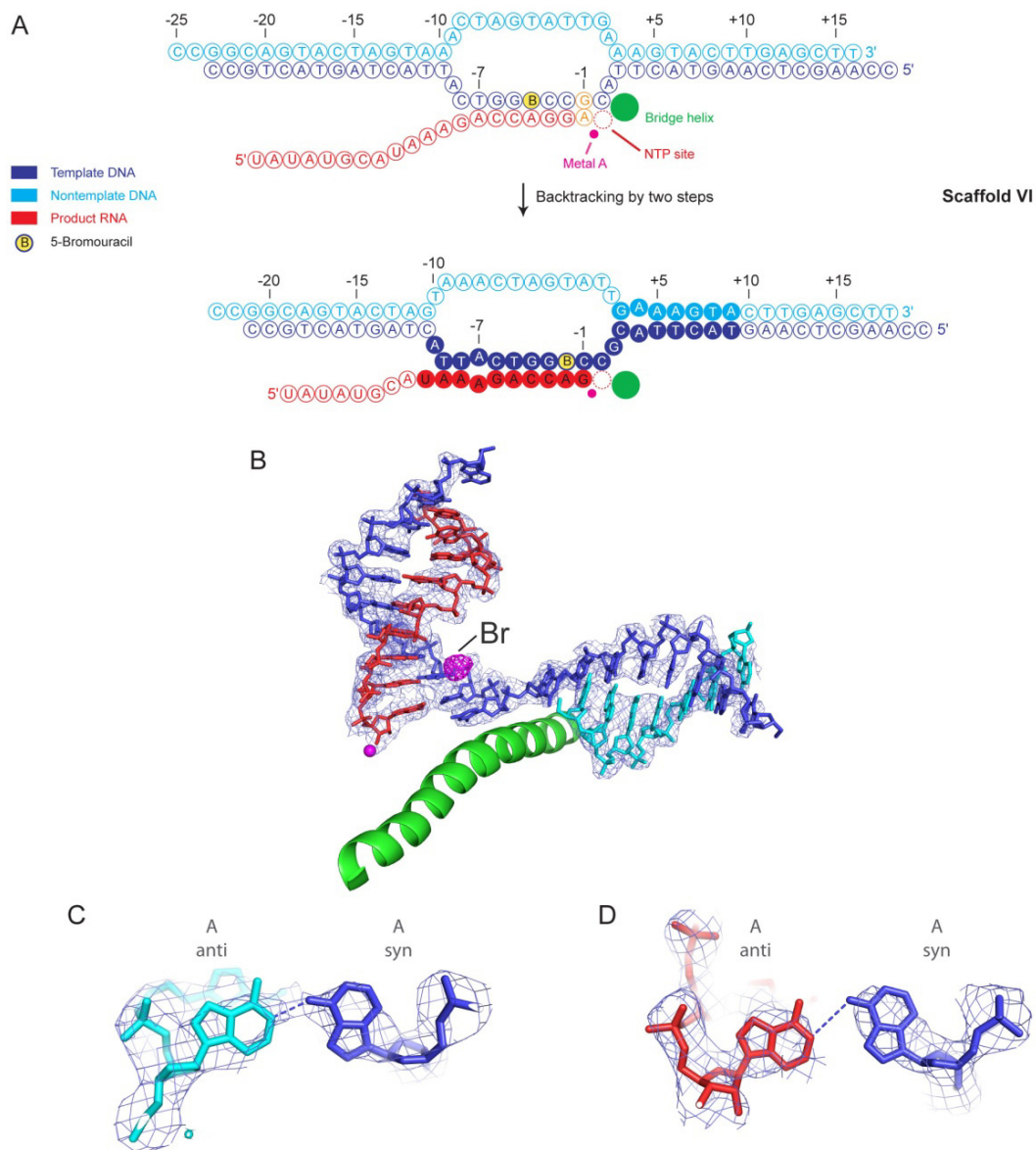


**Figure 25. Dinucleotide cleavage by RNA polymerase II.** Two signals observed on a gel electrophoresis separation correspond to the original RNA and to the product obtained in a cleavage experiment. Cleavage in a dinucleotide step was verified by MALDI-TOF analysis (Experimental procedures) whereas the original RNA (5901 Da) and the dinucleotide cleavage product (5289 Da) could be identified.

## 5.2 Impaired mismatch accommodation

To further investigate efficient cleavage of a mismatch that is efficiently extended, we included a G•A mismatch at the end of the hybrid and solved the structure of the resulting EC VI (scaffold VI, Fig. 26, Table 3). The overall structure was similar to the RNA polymerase II EC that contains the same nucleic acid scaffold except a matched G-C bp at position -1 (Kettenberger et al, 2004). However, a bromine label revealed that RNA polymerase II had apparently backtracked by two steps, although this required accommodation of A•A mismatches at positions -7 and +4 (Fig. 26C and D). Backtracking moved the templating G of the G•A mismatch from the designed position -1 to the downstream position +2. There was only fragmented electron density for the two backtracked terminal RNA nucleotides, indicating that dinucleotide cleavage had occurred prior to crystal analysis. Thus, impaired accommodation of the purine•purine mismatch in the active center apparently favors backtracking and

creates the state of the EC that is prone to dinucleotide RNA cleavage, which is observed in functional assays (Fig. 10 and 24).

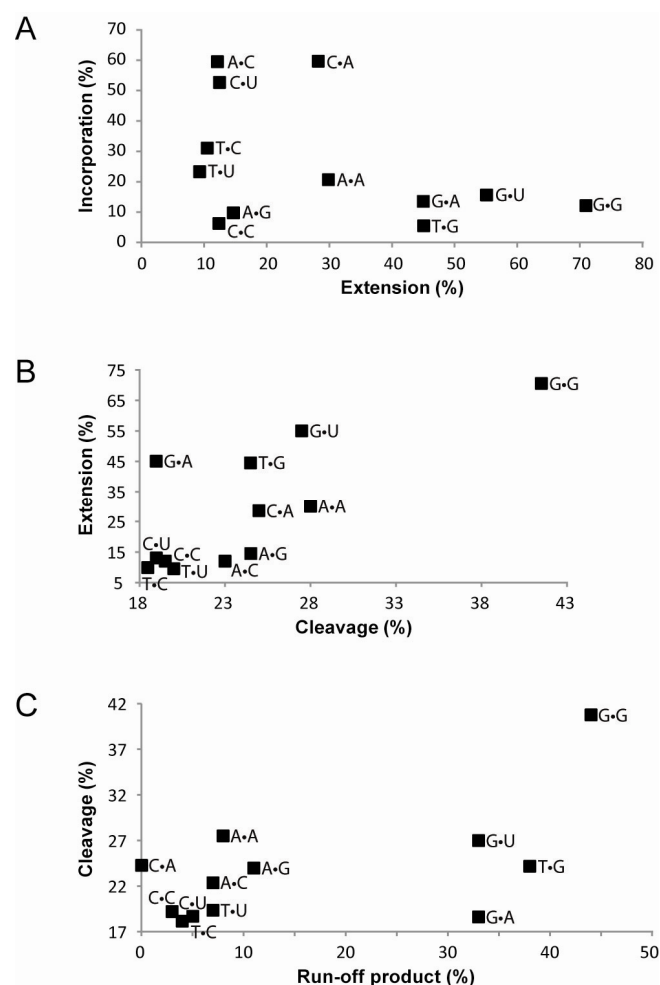


**Figure 26. Structure of EC VI at 3.4 Å resolution** (A) Scaffold VI containing a G•A mismatch at the designed position -1 (top) and as observed in the crystal after backtracking by two steps (bottom). (B) Structure of nucleic acids. The final  $2F_o-F_c$  electron density map is shown as a blue mesh, contoured at  $1.0\sigma$ . A peak in the anomalous difference Fourier map (magenta, contoured at  $4.4\sigma$ ) reveals the location of the bromine atom at position -2 of the template strand, indicating that RNA polymerase II had backtracked by two steps. (C) Detailed view of the  $2F_o-F_c$  map in (A) around the A•A *anti-syn* bps at position +4 in the downstream DNA duplex. A putative hydrogen bond is indicated by a dashed line. (D) Detailed view of the  $2F_o-F_c$  map in (A) contoured at  $1.2\sigma$  around the A•A *anti-syn* bp at position -7 in the RNA-DNA hybrid. A putative hydrogen bond is indicated by a dashed line.

## 6. DISCUSSION

### 6.1 Mismatch-specific transcription fidelity mechanisms

In this thesis, we analyzed the ability of RNA polymerase II to select the correct nucleotide for incorporation, to impair RNA extension beyond a mismatch, and to cleave a mismatched RNA 3' end in a systematic and quantitative way. We show that RNA polymerase II evolved mismatch-specific fidelity mechanisms. Mismatches that are efficiently formed, impair RNA elongation, and mismatches that do not strongly impair RNA elongation are not formed efficiently (Fig. 27A). Mismatches that are efficiently extended are also cleaved efficiently (Fig. 27B), and this can be followed by efficient re-extension (Fig. 27C), providing the basis for proofreading. Our misincorporation efficiencies are consistent with those reported recently (Kireeva et al, 2008) and with misincorporation opposite a template cytosine by RNA polymerase III (Alic et al, 2007). The efficiencies of misincorporation, mismatch extension, and cleavage are apparently dominated by the type of mismatch, and sequence context has a minor influence, as seen for a DNA polymerase (Joyce et al, 1992).



**Figure 27. Correlations between fidelity reaction efficiencies** (A) Correlation between misincorporation and mismatch extension efficiencies. Mean values of 1 min time points shown in Figs. 7 and 24 are plotted. (B) Correlation between efficiencies of mismatch extension (1 min time point mean values, Fig. 11) and RNA cleavage (Fig. 24). (C) Correlation between efficiencies of RNA cleavage and run-off product formation. For run-off experiments, mismatch-containing ECs with scaffolds Z (Fig. 9) were incubated for 5 min with 100  $\mu$ M of a mixture of all NTPs and the bands corresponding to run-off products were quantified (see Experimental procedures).

## 6.2 Mechanistic insights into pausing

Once detected, misincorporation of nucleotides may induce the transition of the still pre-translocated EC to an off-line state when RNA elongation is temporarily stalled (Fig. 29). It was suggested that misincorporation might directly result in transcriptional pausing (Toulokhonov et al, 2007). Additionally, downstream DNA sequences affect pausing (Artsimovitch & Landick, 2000; Chan et al, 1997; Holmes &

Erie, 2003; Landick, 1997; Lee et al, 1990; Palangat et al, 2004; Palangat & Landick, 2001; Wang et al, 1995), where the specific contact of residue R504 of ECIII fork loop 2 with downstream DNA suggests an explanation of this influence on a structural level (Fig. 20).

Transcriptional pausing is a frequent event, occurring on average every 100 bases of DNA (Neuman et al, 2003), mediating the regulation of RNA chain elongation. It is known to be involved in regulation of gene expression (Artsimovitch & Landick, 2002; Bailey et al, 1997; Donahue & Turnbough, 1994; Landick, 2006; Palangat et al, 1998; Ring et al, 1996; Tang et al, 2000; Winkler & Yanofsky, 1981), to be essential for RNA folding, synchronization of transcription and translation (Pan et al, 1999; Yakhnin et al, 2006) to allow splicing and polyadenylation in eukaryotes (de la Mata et al, 2003; Yonaha & Proudfoot, 1999), and to be the first step in backtracking, proofreading, and termination (Artsimovitch & Landick, 2000; Gusarov & Nudler, 1999; Komissarova & Kashlev, 1997a; Landick, 2006; Nudler et al, 1997; Palangat et al, 1998; Park et al, 2004).

Single-molecule studies classified pauses dependent on their life-time (Herbert et al, 2006). Only a small fraction of pauses have a life-time more than 20 seconds, defined as long-lived pauses, such as those stabilized by hairpins and backtracking. The vast majority of pauses, which have been termed ubiquitous pauses (related to the elemental pause), are short-lived pauses, occurring about 1 time per 100 base pairs. Results of those single-molecule studies have proposed a two-tiered mechanism, according to which a long-lived regulatory pause would be comprised of two components acting in succession. First, a common sequence element that triggers a temporary (elemental) pause state, followed by additional sequence elements that convert the elemental pause into a long-lived pause (Herbert et al, 2006).

By means of the underlying mechanism, pausing can be divided broadly into 2 classes: backtrack pausing, in which rearward movement of RNA polymerase leads

to displacement of the RNA 3' end and thus blocks RNA elongation, and non-backtrack pausing, in which nucleotide addition is inhibited by active site rearrangements (Landick, 2009). Non-backtracked pauses can also be stabilized by a hairpin which is formed by the nascent RNA transcript (Herbert et al, 2006). It is involved in attenuation control (Artsimovitch & Landick, 2000; Henkin & Yanofsky, 2002). The other class of defined pauses is stabilized by backtracking of RNA polymerase (Artsimovitch & Landick, 2000; Komissarova & Kashlev, 1997b; Palangat & Landick, 2001). Backtracking has been observed at several pause sites (Artsimovitch & Landick, 2000; Palangat & Landick, 2001; Samkurashvili & Luse, 1996) and was thought previously to be the main source of transcriptional pausing (Epshtein & Nudler, 2003; Komissarova & Kashlev, 1997b; Landick, 1997; Nudler, 1999; Nudler et al, 1997). Subsequent applied force studies on *E. coli* ECs proposed that ubiquitous pausing is independent of backtracking (Neuman et al, 2003) and is rather related to a conformational change within the enzyme, representing the earlier described unactivated intermediate (Erie et al, 1993). Also recent biochemical characterization of previously-reported pauses of *E. coli* RNA polymerase showed that these pauses were not associated with backtracking but contained the 3' end of the transcript in the active center, being capable of binding the next NTP, characterized by much slower bond formation (Kireeva & Kashlev, 2009). In addition, it was shown that sequence-specific pausing is not conserved between prokaryotic RNA polymerases and yeast RNA polymerase II. It became clear that ubiquitous pausing by bacterial RNA polymerase must include non-backtrack pause sites. Structures of the paused ECs III and IV (Fig. 18 and 22) represented in this work, confirm a non-backtracked translocation state in the eukaryotic elemental pause.

Active-site rearrangement in the elemental pause was postulated to include a trigger loop conformation located close to the RNA 3' nucleotide and a conformation of  $\beta$ DloopII that allows fraying of the 3' nucleotide away from the DNA template (Toulokhonov et al, 2007). Such a frayed RNA 3' OH away from the catalytic center



was suggested earlier as a common feature of all pause, arrest, and termination signals (Artsimovitch & Landick, 2000). The existence of a frayed RNA 3' terminal nucleotide in the elemental pause state was revealed by site-directed crosslinking and mutagenesis (Artsimovitch & Landick, 2000; Chan et al, 1997; Touloukhonov et al, 2007). Compared to the previously proposed "two-pawl ratchet" model which suggests bending of the bridge helix into the A site, leading to inhibition of substrate loading or translocation due to fraying of the template DNA base (Bar-Nahum et al, 2005), the "trigger loop based" pausing model can explain coupling of NTP loading and translocation without unfolding of the bridge helix and is accompanied by the RNA 3' nucleotide fraying into the E site (Touloukhonov et al, 2007).

Structural insights into the elemental pause state are obtained from ECs III and IV. A frayed RNA 3' nucleotide binds in two different sites in the pore that are lined by conserved residues (Table 4). Both sites overlap the NTP site, and the tip of the closed trigger loop (Fig. 19), explaining how pausing prevents NTP-coupled translocation and nucleotide addition. The EC V structure further suggests that two A-U bps at the end of the hybrid, which are obtained by transcription of a canonical pause sequence, result in a non-translocated, non-backtracked paused state. All structures reveal a mobile trigger loop and do not elucidate the proposed paused conformation of the trigger loop (Touloukhonov et al, 2007).

### **6.3 Error recognition: mismatches induce off-line states**

Once an error has occurred, RNA polymerases have to recognize it and react appropriately. In this work, we report structures of mismatch-containing RNA polymerase II ECs, which suggest three mechanisms of how misincorporation impairs RNA extension. First, a mismatch may stably bind to RNA polymerase II and disrupt the catalytically competent active site conformation. For example, a T•U mismatch can bind to the -1 position and cause loss of the catalytic metal ion A and

misalignment of the RNA 3' end. Second, a mismatch may facilitate backtracking and RNA cleavage. For example, a G•A mismatch results in a backtracked state in a crystal, and in RNA dinucleotide cleavage *in vitro*. Third, misincorporation may result in an off-line state of the EC with a frayed RNA 3' end.

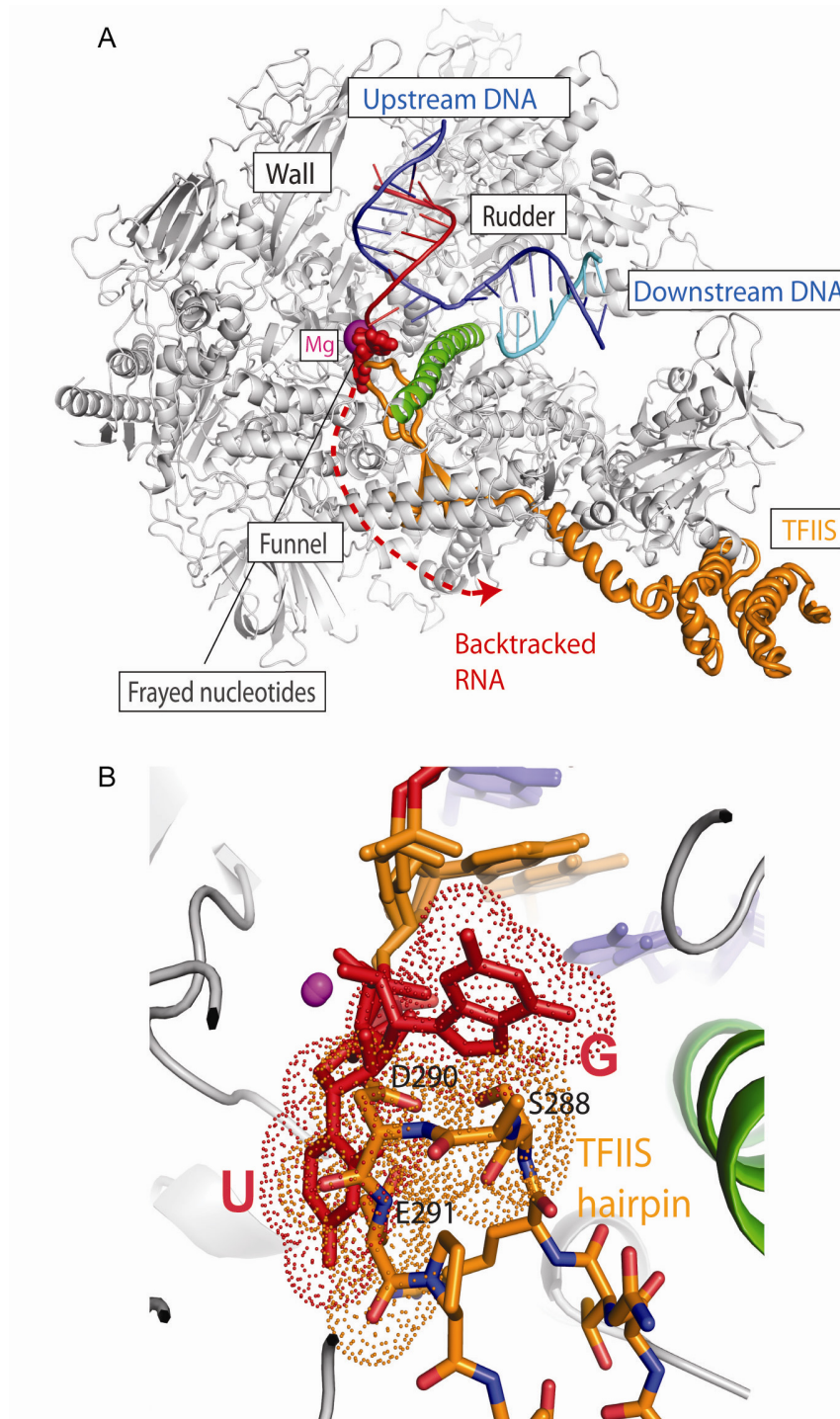
#### **6.4 Error removal: RNA cleavage**

Mismatch removal by RNA cleavage is the third mechanism contributing to transcription fidelity, besides selection of the correct nucleotide and impairment of mismatch extension. Internal hydrolytic cleavage leads to a new 3' OH at the end of the RNA chain, permitting renewed RNA synthesis (Borukhov et al, 1993; Izban & Luse, 1992; Surratt et al, 1991). The general phenomenon of RNA cleavage preferentially in dinucleotide units was observed in bacterial, eukaryotic and archaeal RNA Polymerases (Guo & Price, 1993; Izban & Luse, 1992; Izban & Luse, 1993; Lange & Hausner, 2004; Thomas et al, 1998; Whitehall et al, 1994). The cleavage mechanism including the exonuclease and endonuclease activity was explained by a unified two metal ion mechanism of RNA synthesis and degradation (Sosunov et al, 2003). In a bacterial RNA polymerase, the 3' terminal nucleotide of the mismatched transcript itself stimulates hydrolysis of the penultimate phosphodiester bond and thus its own removal (Zenkin et al, 2006). This represents a "product-assisted catalysis", in contrast to the previously proposed "substrate-assisted catalysis" describing transcript cleavage stimulation by non-complementary NTP (Sosunov et al, 2003). RNA polymerase I and RNA polymerase III both possess a strong intrinsic RNA cleavage activity (Chedin et al, 1998; Kuhn et al, 2007). In RNA polymerase III, it is indeed so strong that misincorporation studies could only be performed with a cleavage-deficient isoform of the enzyme (Alic et al, 2007). In contrast, T7 RNA polymerase does not possess any detectable endogenous RNase activity in ECs

(Huang et al, 2000). Only pyrophosphorolysis was detected but does not appear to contribute to proofreading.

Escape from transcriptional arrest, characterized by extensive backtracking, requires RNA cleavage by the elongation factor TFIIS in the RNA polymerase II system (reviewed in (Fish & Kane, 2002; Wind & Reines, 2000)). Also efficient release from promoter-proximal stall sites requires TFIIS (Adelman et al, 2005), a factor which strongly stimulates the weak intrinsic RNA polymerase II nuclease activity (Izban & Luse, 1992; Reines, 1992; Wang & Hawley, 1993). Addition of TFIIS to paused and arrested ECs releases RNA dinucleotides and 7-8-mers, respectively (Gu & Reines, 1995). *In vitro* studies showed higher transcriptional fidelity in the presence of TFIIS (Thomas et al, 1998). Transcription fidelity *in vivo* does not depend on TFIIS though (Nesser et al, 2006; Shaw et al, 2002), but needs the RNA polymerase II subunit Rpb9 (Nesser et al, 2006). In this work, we show that cleavage rates of mismatch-containing ECs are higher than for those with matched complexes in the absence of TFIIS. These results are consistent with previous studies on RNA cleavage (Zenkin et al, 2006). The same effect was observed in the presence of TFIIS (Jeon & Agarwal, 1996; Wang, 2009). Such preferential removal of misincorporated RNA residues has also been shown in eukaryotic RNA polymerase III (Alic et al, 2007), for GreA stimulated cleavage in *E. coli* (Erie et al, 1993) and TFS stimulated cleavage in the archaeal system (Hausner et al, 2000; Lange & Hausner, 2004). Structural data helped to understand the mechanism of mRNA cleavage and supported the idea of one single tunable RNA polymerase II active site that can switch between RNA synthesis and cleavage modes (Kettenberger et al, 2003). A recent RNA polymerase II TFIIS EC in a backtracked state proposed a regional rearrangement to allow accommodation of TFIIS and backtracked RNA in the interior of the enzyme (Wang, 2009). Superimposition of our structures of EC III and IV with the TFIIS-RNA polymerase complex reveals an overlap of the TFIIS acidic hairpin with the frayed nucleotides (Fig. 28). We therefore postulate TFIIS to trigger the

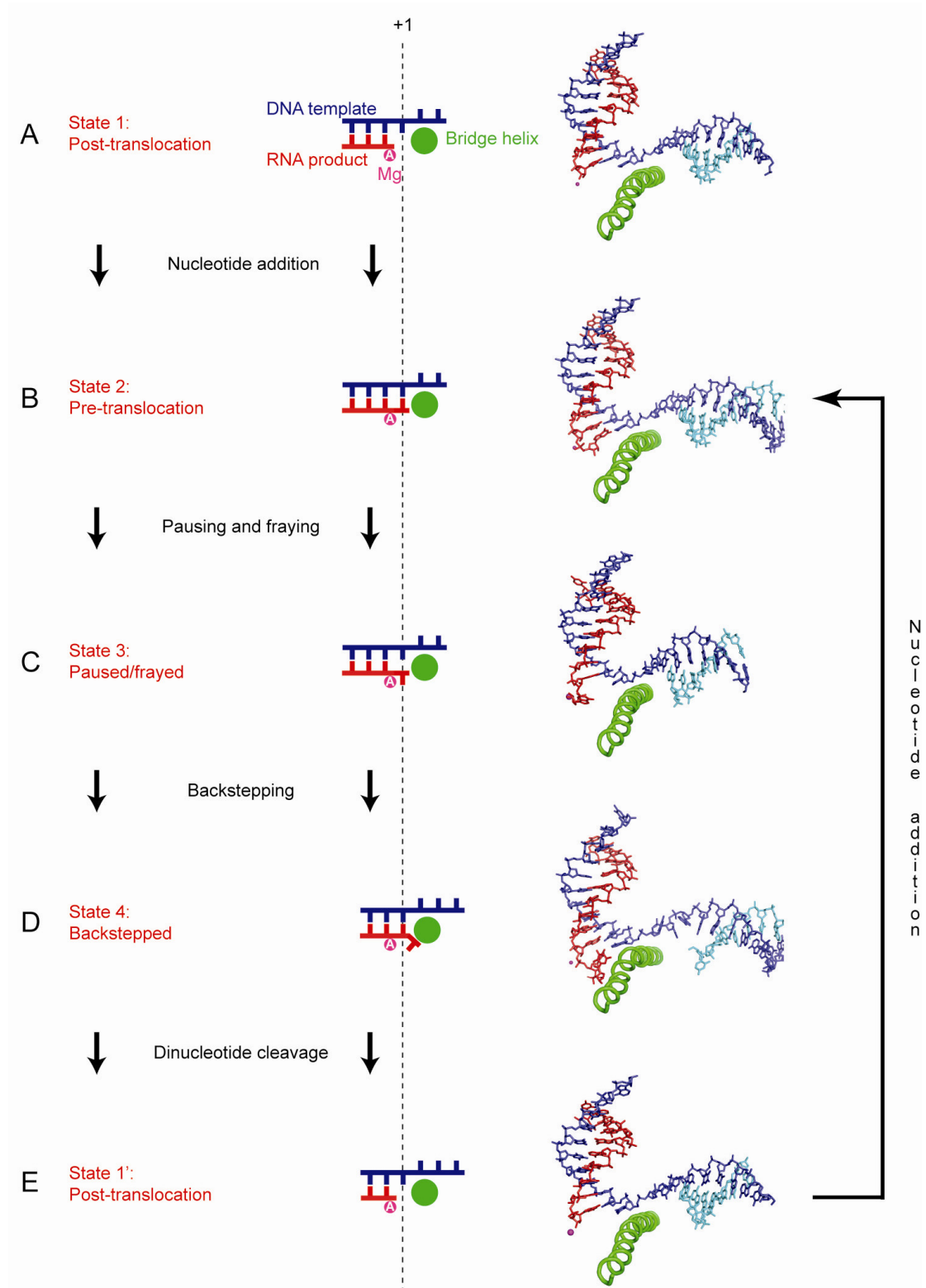
release of a frayed nucleotide. It may also suppress fraying, and thus prevent pausing by keeping the EC in the pre-translocated on-line state. Also in the bacterial system, additional factors for removal of terminal RNA nucleotides are not essential as the enzymes possess intrinsic cleavage activity, although transcript cleavage in *E. coli* is facilitated by the transcription factors GreA and GreB (Borukhov et al, 1992; Borukhov et al, 1993) which merely enhance the intrinsic transcript cleavage activity of a bacterial RNA polymerase (Orlova et al, 1995). Despite having similar function as eukaryotic TFIIIS, bacterial GreA and GreB proteins lack sequence or structural similarity. GreA and GreB stimulate cleavage in different ways. Cleavage of 3' RNA fragments of two to three nucleotides in length are induced by GreA and are only able to prevent arrested complex formation, whereas GreB can rescue preexisting arrested complexes by inducing cleavage of fragments up to 18 nucleotides in length (Opalka et al, 2003).



**Figure 28. Frayed nucleotides overlap the TFIIIS hairpin** (A) Frayed nucleotides overlap the tip of the hairpin of the cleavage-stimulatory factor TFIIIS. The structures of EC III, EC IV, and the RNA polymerase II-TFIIIS complex (PDB-code 1PQV, (Kettenberger et al, 2003)) were superimposed with their active center regions. TFIIIS is shown in orange. The canonical side view is used. (B) Detailed view of the superposition in (D) around the active site, revealing a potential clash of the TFIIIS acidic hairpin with the frayed nucleotides.

## 6.5 A model for RNA proofreading

In a DNA polymerase, the 3' cleavage rate is governed by the rate of fraying (Morales & Kool, 2000), suggesting that RNA fraying occurs during transcriptional proofreading. Since RNA cleavage generally occurs in dinucleotide steps (Izban and Luse, 1993), the polymerase must backtrack by one step after fraying. Backtracking allows the terminal nucleotide to contribute catalytic groups to the active site for cleavage stimulation (Zenkin et al, 2006). The structural studies of our work and of backtracked RNA polymerase II ECs (Wang, 2009) together with published biochemical work (Alic et al, 2007; Thomas et al, 1998; Zenkin et al, 2006) converge on the following proposed mechanism for RNA proofreading. After misincorporation (Fig. 29B), polymerase recognizes the resulting mismatch, such that further RNA extension is slowed down or prevented. The mismatch may lead to distortion of the on-line pre-translocated state. This active site distortion might cause a stabilization of a paused off-line state with a frayed nucleotide (Fig. 29C). Polymerase then backtracks by one position (back-stepping) and places the RNA end into a proofreading site that is identical or very similar to a possible fraying site (Fig. 29D). This forms the basis for dinucleotide cleavage, resulting in an empty NTP binding site (Fig. 29E). By doing so, the EC re-accesses the online-state that allows for continuation of transcription (Fig. 29 A). Cleavage stimulatory factors may help recruit the water molecule required for catalysis and/or position the substrates and may remobilize backtracked RNA bound to certain non-productive sites.



**Figure 29. Possible model of the RNA polymerase II proofreading cycle.** The vertical dashed line indicates register +1. (A) RNA polymerase II in the post-translocation state (PDB code 1Y1W), (B) in the pre-translocation state (PDB code 1I6H, the DNA nontemplate strand was modeled from 1Y1W (Brueckner et al, 2009)), (C) in a paused conformation with a frayed 3' terminal guanine (ECIII), (D) in a backstepped state (PDB code 3GTJ), (E) and after dinucleotide cleavage (EC VI, PDB code 3HOY).

The model presented here can correspond either to a misincorporation event that results in mismatch accommodation directly in the active center at position -1, not leading to translocation, or to mismatch incorporation with subsequent translocation, leading to its positioning at register -2. In both cases, RNA 3' fraying could occur, which results in either a mismatched or a matched nucleotide in the fraying site, respectively, and after subsequent backstepping in the proofreading site. In both cases, the mismatch that is either at the terminal or the penultimate position relative to the RNA 3' end, respectively, can be removed as cleavage occurs in dinucleotide steps.



## 7. EXPERIMENTAL PROCEDURES

### 7.1 Measurement of protein concentration

For the determination of protein concentrations the Bradford protein assay was used (Bradford, 1976). Dye reagent was purchased from Biorad and the assay was performed according to the manufacturer's instructions. For each new batch of dye reagent a calibration curve was generated using Bovine serum albumin (Fraktion V, Roth).

### 7.2 Isolation of 10-subunit core RNA polymerase II from yeast

#### 7.2.1 Yeast fermentation

Isolation of core RNA polymerase II was carried out from the *Saccharomyces cerevisiae* strain CB010 $\Delta$ Rpb4 (MATa *pep4::HIS3/prb1::LEU2*, *prc1::HISG*, *can1*, *ade2*, *trp1*) (Edwards et al, 1990; Fu et al, 1999). The strain is characterized by knockouts of several cellular proteases and of RPB4. Homogeneous 10-subunit core RNA polymerase II could be obtained in the absence of Rpb4, as Rpb7 dissociates from core RNA polymerase II during purification. For production of up to 1.5 kg of yeast pellet per batch, two types of fermentors were available. The small fermentor (ISF200, Infors) has a nominal volume of 20 l and should be run with up to 15 l of media. The large fermentor (ABEC, Infors) has a larger capacity, with a nominal volume of 200 l and can be filled with up to 160 l of media. The media composition and the culture parameters of both fermentors were as follows:

**Small fermentor (20 l)****YPD media**

300 g peptone

300 g glucose

222 g yeast extract

15 l desalted water

**Antibiotics**

0.75 g ampicillin

0.15 g tetracycline · HCl

**Typical inoculate volume**

0.3 l @ OD600 ≈ 2

**Air flow**

8 l/min

**Stirrer speed**

800 rpm

**Typical growth time**

12 – 15 hours

**Large fermentor (200 l)****YPD media**

3200 g peptone

3200 g glucose

2370 g yeast extract

160 l desalted water

**Antibiotics**

8.0 g ampicillin

1.6 g tetracycline · HCl

**Typical inoculate volume**

4-5 l @ OD600 ≈ 2

**Air flow**

20 l/min

**Stirrer speed**

200 rpm

**Typical growth time**

12 – 15 hours

## 7.2.2 Purification of 10-subunit core RNA polymerase II

### 7.2.2.1. Buffers

#### 100 x protease inhibitor mix

1.42 mg Leupeptin

6.85 mg Pepstatin A

850 mg PMSF

1650 mg benzamidine

dry ethanol to 50 ml

stored at  $-20\text{ }^{\circ}\text{C}$ ; added immediately before use

#### 3x freezing buffer

150 mM Tris-Cl, pH 7.9 @  $4\text{ }^{\circ}\text{C}$

3 mM EDTA

30 % glycerol

30  $\mu\text{M}$   $\text{ZnCl}_2$

3 % DMSO

30 mM DTT

3 x protease inhibitor mix

#### 1 x HSB150 buffer

50 mM Tris-Cl, pH 7.9 @  $4\text{ }^{\circ}\text{C}$

150 mM KCl

1 mM EDTA

10 % glycerol

10  $\mu\text{M}$   $\text{ZnCl}_2$

10 mM DTT

1 x protease inhibitor mix

**1 x HSB600 buffer**

50 mM Tris-Cl, pH 7.9 @ 4 °C

600 mM KCl

1 mM EDTA

10 % glycerol

10  $\mu$ M ZnCl<sub>2</sub>

10 mM DTT

1 x protease inhibitor mix

**TEZ buffer**

50 mM Tris-Cl, pH 7.5 @ 20 °C

1 mM EDTA

10  $\mu$ M ZnCl<sub>2</sub>

1 mM DTT

1 x protease inhibitor mix

**UnoQ buffer**

50 mM Tris-Cl, pH 7.5 @ 20 °C

1 mM EDTA

10  $\mu$ M ZnCl<sub>2</sub>

10 % (v/v) glycerol

10 mM DTT

no protease inhibitors

**1 x RNA polymerase II buffer**

5 mM HEPES pH 7.25 @ 20 °C

40 mM ammonium sulfate

10  $\mu$ M ZnCl<sub>2</sub>

10 mM DTT

### **Acetate buffer**

100 mM sodium acetate pH 4.0

500 mM sodium chloride

### **PBS**

4.3 mM Na<sub>2</sub>HPO<sub>4</sub>

1.4 mM KH<sub>2</sub>PO<sub>4</sub>

137 mM sodium chloride

2.7 mM potassium chloride

pH 7.4

### **Coupling buffer**

100 mM sodium bicarbonate pH 8.3

500 mM sodium chloride

#### **7.2.2.2. Harvesting and storage of yeast**

The yeast cells were harvested at late logarithmic/early stationary phase, monitored by OD600 measurement. Cells were pelleted by centrifugation for 20 min at 5000 rpm in a SLC6000 rotor (small fermentor) or by a continuous flow centrifuge (Padberg Z4IG, 20000 rpm) in case of the large fermentor. Subsequently, 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.

### **7.2.2.3. Purification - day 1 (lysis and heparin column)**

Up to 600 ml of cell suspension were thawed in warm water for three bead-beaters (BioSpec). 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 were filled in each bead-beater. The bead-beater was filled completely with HSB150, whilst avoiding any remaining air bubbles. The cell lysis was achieved within 60-75 min of bead-beating (30 s on/90 s off) while the beater chambers were submersed in a salt/ice mixture. Removal of glass beads was achieved by filtration through a mesh funnel. Subsequently, the beads were washed with HSB150 until the flowthrough was clear. Two rounds of centrifugation cleared the lysate (45 min at 9000 rpm in a GS3 rotor or 30 min at 12000 rpm in a SLA1500 rotor). Filtration of the supernatant through two layers of paper filter discs underneath a dressing cloth led to removal of lipids. The cleared lysate was loaded onto a column packed with 250 ml of Heparin Sepharose 6 FF (GE Healthcare) (flow rate: 6-8 ml/min) and pre-equilibrated with 750 ml of HSB150. Proteins were eluted with 500 ml of HSB600 (flow rate: 6-8 ml/min) and then precipitated by adding 291 g of fine-ground ammonium sulfate per litre of eluate (= 50 % saturation), followed by 60 min of stirring at 4 °C, over-night incubation at 4 °C and finally centrifugation (45 min. at 12000 rpm in a SLA1500 rotor). Washing with 1 l of 6 M urea and water was applied to restore the heparin column. For storage, 5 mM potassium acetate and 20 % (v/v) ethanol were applied. 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.

### **7.2.2.4. Purification - day 2 (immunoaffinity column)**

On the second day, the ammonium sulfate pellet of day 1 was dissolved in 50 ml of TEZ buffer. The conductivity was adjusted below the conductivity of TEZ containing additionally 400 mM ammonium sulfate (TEZ400) by addition of more TEZ buffer.

Centrifugation of the sample (15 min at 14000 rpm in a SLA1500 rotor) was done to remove undissolved constituents. Afterwards, it was loaded by gravity flow onto the immunoaffinity column at 4 °C which was pre-equilibrated before with 20 ml of TEZ containing 250 mM ammonium sulfate (TEZ250). To increase the yield of RNA polymerase II, the flowthrough was loaded onto a second column. The columns were mounted at room temperature, washed with 25 ml of TEZ500 at room temperature and RNA polymerase II was eluted in 1 ml fractions with TEZ500 containing additionally 50 % (v/v) glycerol (ca. 15 ml). 9 mM DTT was added to the elution fractions containing RNA polymerase II (monitored with the Bradford assay) directly afterwards, and they were stored at 4 °C over night. Columns were 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. Generally, the recovery of RNA polymerase II decreased with each use of the column starting already from the first use. One reason is probably the sensitivity of the antibody towards DTT.

#### **7.2.2.5. Purification - day 3 (anion exchange chromatography or buffer exchange)**

After six-fold dilution, peak fractions from day 2 were loaded onto a UnoQ column (BioRad, column volume 1.35 ml) and pre-equilibrated with buffer UnoQ containing 60 mM ammonium sulfate (UnoQ-A). This buffer was used to wash the the column with 3 column volumes, and RNA polymerase II was eluted with a linear gradient over 10 column volumes from 0-50 % buffer UnoQ containing 1 M ammonium sulfate (UnoQ-B). RNA polymerase II eluted at about 25 % buffer UnoQ-B. Washing with 5 column volumes of UnoQ-B was used to restore the column. An increase of the final yield of RNA polymerase II could be achieved by replacement of the anion exchange step by a simple buffer exchange procedure, which didn't affect the suitability of the purified RNA polymerase II for structural or functional experiments. Centrifugal ultrafiltration devices (MWCO 100,000 Da, Millipore Amicon Ultra-15) served for

buffer exchange from the buffer in the antibody column elution fractions to 1x RNA polymerase II buffer. By measuring the conductivity of the flowthrough, the completeness of the buffer exchange was monitored. Subsequently, RNA polymerase II was concentrated to 1-2 mg/ml. After anion exchange chromatography or after buffer exchange, the RNA polymerase II sample was split into aliquots of 100-500 µg protein. The aliquots after anion exchange chromatography were mixed with an equal volume of ammonium sulfate solution saturated at room-temperature, the aliquots in 1 x RNA polymerase II buffer with 1.13 times the volume. The mixture was incubated for at least 1 hour at 4 °C and centrifuged for 45 min at 4 °C in a table-top centrifuge at 13000 rpm. Most of the supernatant was decanted so that the pellet was still covered with supernatant. Finally, it was shock-frozen in liquid nitrogen and stored at -80 °C. RNA polymerase II stored this way is stable for about 3 months. A yield of 0.5-4 mg of highly purified RNA polymerase II was achieved from 600 g yeast pellet.

#### **7.2.2.6. Preparation of RNA polymerase II immunoaffinity resin**

A monoclonal antibody, 8WG16 (NeoClone, Madison/USA) (described in (Thompson & Burgess, 1996)) is specific for the unphosphorylated CTD of RNA polymerase II and optimized to release RNA polymerase II upon treatment with glycerol or ethylene glycol at room temperature ("polyol responsive antibody"). These antibodies were purified from mouse ascites. They were immobilized on activated chromatography media according to the following procedure:

PBS was used to dissolve the lyophilized ascites to their original volume which were then filtered through 0.2 µm membrane filters. The solution was passed more than 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. 1 ml fractions were collected



into tubes containing 200  $\mu$ l of 2 M HEPES (pH 7.9) to neutralize the acid. The 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.

Immunoaffinity columns had a cyanogen bromide (CNBr)-activated sepharose 4 B (Sigma) matrix, which reacts with free amines, e.g. accessible  $-\text{NH}_2$  groups on proteins. Other sources of free amines (e.g. Tris) were avoided and a sealed bottle of activated sepharose was used. 5 ml of gel were prepared for each immunoaffinity column 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 first washed with 100 ml of 1 mM HCl, then with 20 ml of coupling buffer. 10 mg of purified antibodies per column were coupled for 2 hours at 20  $^{\circ}\text{C}$  or over night at 4  $^{\circ}\text{C}$ . No protein was detectable in the supernatant when the coupling reaction was completed. The column was then washed with 25 ml of 1 M Tris, pH 8 and incubated for 2 hours at room temperature or over night at 4  $^{\circ}\text{C}$ . Subsequently, the column was washed with 20 ml of coupling buffer, followed by acetate buffer and coupling buffer. The columns were stored at 4  $^{\circ}\text{C}$  in TEZ60 with 0.02 % sodium azide. It was possible to use the columns several times if DTT exposure was reduced to a minimum, but in general a decrease in RNA polymerase II yield was observed already after the first and even further after subsequent uses.

### **7.3 Purification of the subcomplex Rpb4/7**

#### **Buffer 1**

150 mM NaCl

5 % (v/v) glycerol

50 mM Tris pH 7.5

10 mM  $\beta$ -mercaptoethanol

protease inhibitors (see chapter 4.5)

**Buffer 2**

50 mM Tris pH 7.5

5 mM DTT

1 mM EDTA

*E. coli* BL21(DE3) RIL (Stratagene) with a bicistronic vector was used to express recombinant yeast Rpb4/7. 2 x 2 L cultures of cells were grown in auto-induction medium (Studier, 2005). After about 4 h at an OD600  $\approx$  0.6, the temperature was shifted from 30 °C to 20 °C. After 11 h cell growth, cells were harvested by centrifugation (15 min at 5000 rpm in a SLC6000 rotor), resuspended in buffer 1 and lysed using a French Press. Before application to a NiNTA column (Quiagen; 1 ml column volume), the lysate was cleared by centrifugation (30 min at 15000 rpm in a SS34 rotor). Subsequently, the column was washed with 3 ml of buffer 1, 3 ml buffer 1 containing additionally 10 mM imidazole and 3 ml of buffer 1 containing additionally 20 mM imidazole. Proteins were eluted with 3 ml of buffer 1 containing additionally 50 mM imidazole and 3 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 (GE Healthcare, 6 ml column volume) which was pre-equilibrated in buffer 2. A linear gradient from 0-1000 mM NaCl in buffer 2 was applied to elute Rpb4/7. The peak fractions were concentrated and applied on a Superose12 10/300 GL gel filtration column (GE Healthcare) which were pre-equilibrated in Pol II buffer. The resulting purified Rpb4/7 heterodimer was concentrated to 10 mg/ml and was stored in aliquots at -80 °C.

## 7.4 Purification of His-tagged RNA polymerase II

### Ni buffer

20 mM Tris-HCl pH 7.9

150 mM KCl

10  $\mu$ M ZnCl<sub>2</sub>

10% v/v glycerol

10 mM DTT

1x protease inhibitors

### High salt buffer

20 mM Tris-HCl pH 7.9

1000 mM KCl

7 mM imidazole

10  $\mu$ M ZnCl<sub>2</sub>

10% v/v glycerol

10 mM DTT

1x protease inhibitors

### Ni7 buffer

20 mM Tris-HCl pH 7.9

150 mM KCl

7 mM imidazole

10  $\mu$ M ZnCl<sub>2</sub>

10 mM DTT

1x protease inhibitors)

**MonoQ buffer**

20 mM Tris-acetate pH 7.9

0.5 mM EDTA

10  $\mu$ M ZnCl<sub>2</sub>

10% v/v glycerol

10 mM DTT

*S. cerevisiae* RNA polymerase II containing a hexahistidine-tagged Rpb3 subunit (strain kindly provided by the laboratory of M. Kashlev) was purified as described (Kireeva et al, 2003), but with several significant modifications. Briefly, 150 gram of cell pellet were resuspended in freezing buffer and were lyzed by bead beating for 80 min using intervals of 30 seconds followed by 90 second pauses. The lysate was cleared by centrifugation and ultracentrifugation. RNA polymerase II was precipitated by the addition of 50% saturated ammonium sulphate solution. The pellet was dissolved in Ni buffer and subjected to Ni-NTA affinity chromatography (2 x 8 ml fresh Ni-NTA) using gravity flow. After washing with high salt buffer and with Ni7 buffer, the protein was eluted with Ni7 buffer containing 100 mM imidazole and no protease inhibitors. The eluted protein was diluted with MonoQ buffer and subjected to anion exchange chromatography (MonoQ, GE healthcare) using a gradient from 150 mM to 1500 mM KOAc. The last elution peak (at a conductivity of 50 mS/cm) was collected and concentrated. The concentrated RNA polymerase II was precipitated by the addition of 50% ammonium sulfate, and the pellets were stored at -80 °C.

## 7.5 EC assembly

### Transcription buffer (TB)

20 mM HEPES pH 7.6

60 mM  $(\text{NH}_4)_2\text{SO}_4$

8 mM  $\text{MgSO}_4$

10  $\mu\text{M}$   $\text{ZnCl}_2$

10% v/v glycerol

10 mM DTT

For the bead-based assays, the ECs containing complete complementary scaffolds were assembled essentially as described (Kireeva et al, 2003). Briefly, the DNA non-template was 5'-end-labeled with Biotin with the use of a TTTTT linker. The RNA was 5'-end-labeled with 6-carboxyfluoresceine (FAM). For EC assembly, RNA polymerase II was incubated with a hybrid of the DNA template strand annealed to the RNA (2-fold excess) in TB for 15 min at 20 °C, subsequently with the biotinylated non-template DNA strand (4-fold excess) for 10 min at 25 °C, and then with recombinant Rpb4/7 (5-fold excess) for 10 min at 25 °C.

## 7.6 Bead-based RNA extension and cleavage assays

Bead-based assays were carried out as described with some modifications (Dengl & Cramer, 2009). Briefly, beads (Dynabeads MyOne™ Streptavidin T1 from Invitrogen) were added to ECs for assembly and incubated for 30 min at 25 °C. Beads were subsequently washed with TB containing 0.1% Triton-X, TB containing 0.2 M  $(\text{NH}_4)_2\text{SO}_4$ , and with TB. Beads were re-suspended in TB. For RNA extension assays including time course experiments, different amounts of NTPs (Jena Bioscience) were added, the mixture was incubated at 28 °C and reactions were stopped at

different time points by addition of an equal volume of 100 mM EDTA, essentially as described (Brueckner et al, 2007). For cleavage assays, the bead-coupled ECs were incubated at 28 °C in TB for 5 min and stopped as described for extension assays. The beads were transferred into urea loading buffer, samples were heated to 95 °C and loaded on a 20% polyacrylamid gel containing 7 M Urea. The FAM 5'-labeled RNA products were visualized with a Typhoon 9400 scanner (GE Healthcare). Gel bands were quantified using ImageQuant (GE healthcare). In case more than one product was observed (A), the amounts of different RNA products were added up. For MALDI-TOF analysis, the reaction was incubated, stopped and analyzed as described (Brueckner et al, 2007). NTP samples were analyzed by reverse phase HPLC analysis and no cross-contamination with other NTPs was detected. We also requested analytic data from the supplier, which showed that the NTPs are 99.8% pure and the remaining impurities are NDPs and NMP of the same kind, but not other types of NTPs. Since the NTPs are synthesized *de novo*, and are not derived from fractionation of an NTP pool, cross-contamination cannot occur. We are therefore certain that misincorporation took place.

## 7.7 Crystal structure determinations

The match and mismatch-containing scaffolds were cocrystallized and the structures were determined essentially as described (Brueckner et al, 2007), with minor changes. The crystallization solution lacked magnesium ions (200 mM ammonium acetate, 300 mM sodium acetate, 50 mM Hepes pH 7.0, 4-6 % w/v PEG 6000, 5 mM TCEP). Diffraction data of EC I were collected at the beamline X06A of the Swiss Light Source using a mar225 CCD detector, whereas data of ECs II-VI were collected using a PILATUS 6M pixel detector (Broennimann et al, 2006) (Table 3). Raw data of EC I were processed with HKL2000, data of ECs II-VI with XDS (Kabsch, 1993). The structure of EC I and VI were solved by molecular replacement with the program

PHASER (McCoy et al, 2005), using the structure of the complete 12-subunit RNA polymerase II without nucleic acids as a search model (PDB 1Y1W) (Kettenberger et al, 2004). The higher resolution of the EC I crystal produced a superior model of the protein compared with 1Y1W, as judged by the quality of the Ramachandran plot and  $R_{\text{cryst}}/R_{\text{free}}$  values. When used as a search model for molecular replacement, EC I resulted in better quality models in the determination of the EC II, III, IV and V structures. The molecular-replacement solution was subjected to rigid-body refinement with CNS version 1.2 (Brünger et al, 1998). Model-building was done with Coot (Emsley & Cowtan, 2004) and Moloc (Gerber Molecular Design, Switzerland, <http://www.moloc.ch>). The nucleic acids were built stepwise into unbiased  $F_o-F_c$  electron density. The register of the nucleic acids was unambiguously defined by bromine labeling as described (Brueckner et al, 2007). Refinement of ECs II-VI was monitored with the free R-factor calculated from the same set of excluded reflections as in the refinement of the complete RNA polymerase II complex (Armache et al, 2005), and the complete RNA polymerase II EC (Brueckner et al, 2007; Damsma et al, 2007; Kettenberger et al, 2004). Due to the different space group and higher resolution of EC I, a new test set of reflections was generated.

## 8. CONCLUSIONS AND OUTLOOK

The studies of this thesis have led to a more detailed understanding of RNA polymerase fidelity mechanisms. They revealed similarity and differences to DNA polymerase fidelity mechanisms. Taken together, NTP selectivity is likely governed by similar mechanisms in DNA and RNA polymerases, involving an induced-fit mechanism and a conformational change in the enzyme active center induced by an accurate base pairing of the NTP to the template base. We show that RNA polymerase II prevents erroneous transcription *in vitro* by combination of pre- and post-incorporation fidelity mechanisms (Fig. 27). Efficient mismatch formation is combined with impaired RNA extension, or efficient mismatch extension with inefficient mismatch formation beforehand and/or efficient mismatch removal by RNA cleavage afterwards during proofreading. The different strategies depend on the type of DNA•RNA mismatch. Furthermore, our reported mismatch-containing structures suggest three mechanisms of how misincorporation impairs RNA extension; (i) by disruption of a catalytically competent active site conformation by stable mismatch binding; (ii) by induction of an offline-state of the EC with frayed RNA 3' nucleotides, which also gave unexpected insights into pausing; (iii) by facilitated backtracking and RNA cleavage. These results together with other biochemical and structural studies propose a model for RNA proofreading. First, a misincorporation event leads to distortion of the on-line pre-translocated state. Second, this distortion might lead to stabilization of a paused off-line state with a frayed nucleotide. Third, the polymerase might then backtrack by one step, positioning the RNA 3' end into a proofreading site. Fourth, this forms the basis for dinucleotide cleavage. The resulting empty NTP binding site leads to re-accession of the online-state and to resumption of transcription.



In the future, the molecular basis of nucleotide selectivity may be analyzed with EC structures containing mismatched NTP substrates, although such structures are very difficult to obtain for RNA polymerase II (Brueckner et al, 2007), and also for single-subunit DNA polymerases (Batra et al, 2008). The closed RNA polymerase II active center may accommodate small template-NTP pyrimidine•pyrimidine mismatches, explaining the facilitated formation of T•U, T•C, and C•U mismatches (Fig. 7). Other misincorporations may however occur via an EC intermediate that lacks a DNA base in the templating site, as suggested by recent structures of a DNA polymerase with a mismatched NTP (Batra et al, 2008). Indeed, an empty templating site was observed in a recent RNA polymerase II EC intermediate structure (Brueckner & Cramer, 2008), and RNA polymerase II can misincorporate opposite an abasic template site (Damsma et al, 2007), and likely also opposite an empty templating site that results from a failure to translocate a bulky DNA lesion into the active site (Brueckner et al, 2007; Damsma et al, 2007).

Further questions concerning RNA polymerase fidelity remain unanswered. How much influence does TFIIIS have on the fidelity-determining mechanisms observed in this work? How much would DNA•RNA mismatches influence RNA chain elongation at a position further upstream in the DNA-RNA hybrid, i.e. at positions -2, -3 or -4, and when would they stop to affect RNA elongation efficiencies? Can the model for RNA proofreading be confirmed *in vitro*? Would the reconstituted RNA polymerase II ECs with frayed nucleotides be subject to dinucleotide cleavage? Would other types of mismatches in the polymerase active center (position +1) directly lead to RNA fraying, instead of inducing fraying only after extension by one nucleotide (ECs III and IV)? We could gain further insights in mismatch-induced active site disruption and/or impaired RNA extension by trapping ECs with other types of mismatches in the active center and at different positions.

## 9. APPENDIX

### 9.1 Unpublished RNA polymerase II EC structures

Complete 12-subunit RNA polymerase II was co-crystallized with each of the following nucleic acid scaffolds.

#### Scaffold FB-4-TA-x

```

FB-3RNA+A:      5'-          UG CAU UUC
                  GAC CAG GCA
FB-4tem:        3'-          TTA CTG GBC CGT ATT GAT GAA CTC GA  -3'
FB-3ntem:      5'-          AA CTA CTT GAG CT  -3'
  
```

#### Scaffold JS9-MM-GA-1-x

```

JS-RNA9-MM-1-GA-x:5'-      UG CAU UUC
                          GAC CAG GCA
JS-temG:          3'-      TTA CTG GBC CGG GCG CCT GTC CTC GA  -3'
JS-12ntem:       5'-      C GGA CAG GAG CT  -3'
  
```

#### Scaffold JS9-GC-x

```

JS-RNA9+C-x:     5'-      UG CAU UUC
                          GAC CAG GCC
JS-temG:         3'-      TTA CTG GBC CGG GCG CCT GTC CTC GA  -3'
JS-12ntem:      5'-      C GGA CAG GAG CT  -3'
  
```

#### Scaffold JS8-MM-GA-1-x

```

JS-RNA8-MM1-GA-x:5'-      UG CAU UUC A
                          AC CAG GCA
JS-temG:         3'-      TTA CTG GBC CGG GCG CCT GTC CTC GA  -3'
JS-12ntem:      5'-      C GGA CAG GAG CT  -3'
  
```

#### Scaffold JS9-MM-TC-1-x

```

JS-RNA9+C-x:     5'-      UG CAU UUC
                          GAC CAG GCC
FB-4tem:         3'-      TTA CTG GBC CGT ATT GAT GAA CTC GA  -3'
FB-3ntem:      5'-      AA CTA CTT GAG CT  -3'
  
```

#### Scaffold JS8-AC-x

```

JS-RNA8+C-x:     5'-      UG CAU UUC A
                          AC CAG GCC
JS-temA:         3'-      TTA CTG GBC CGA ACG CCT GTC CTC GA  -3'
JS-12ntem:      5'-      C GGA CAG GAG CT  -3'
  
```

**JS-AC mismatch bubble**

```

JS-RNA8+C-x:      5' -           UG CAU UUC A
                   AC CAG GCC
JS-temAn-MMB:    3' -   CCG TCA TGA TCA TTA CTG GBC CGA ACG CCT GTC CTC GAA CC -3'
JS-ntem-MMB:     5' -CC GGC AGT ACT AGT AAA CTA GTA TTG ATC GGA CAG GAG CTT -3'

```

**JS-AU mismatch bubble**

```

JS-RNA8+U-x:     5' -           UG CAU UUC A
                   AC CAG GCU
JS-temAn-MMB:    3' -   CCG TCA TGA TCA TTA CTG GBC CGA ACG CCT GTC CTC GAA CC -3'
JS-ntemN-MMB:   5' -CC GGC AGT ACT AGT AAT CTA GTA TTG ATC GGA CAG GAG CTT -3'

```

**Scaffold JS9-TA-2-x**

```

JS-RNA9-TA-2-x:  5' -           UG CAU UUC A
                   AC CAG GCA G
JS-2tem:         3' -           TTA CTG GBC CGT CTT GAT GAA CTC GA -3'
JS-2ntem:       5' -           A CTA CTT GAG CT -3'

```

Despite high diffraction data quality, all RNA polymerase II EC structures crystallized with the above listed nucleic scaffolds were in mixed translocation states, defined by multiple bromine peaks in the anomalous maps and could therefore not be refined.

## 10. ABBREVIATIONS

bp	Base pair
CTD	C-terminal domain of Rpb1 of Pol II
DNA	Deoxyribonucleic acid
dNTP	2' Deoxyribonucleoside triphosphate
DPE	Downstream promoter element
DTT	Dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EC	Elongation complex
EDTA	Ethylene diamine tetraacetic acid
EM	Electron microscopy
GTF	General transcription factor
GTP	Guanosine triphosphate
Hepes	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
Inr	Initiator element
IPTG	Isopropyl $\beta$ -D-1-thiogalactopyranoside
MALDI-TOF	Matrix-assisted laser desorption ionization with time-of-flight analyser mass spectrometry
min	Minutes
mRNA	Messenger RNA
MWCO	Molecular weight cutoff
NTP	Nucleotide triphosphate
PDB	Protein data bank
PEG	Polyethylene glycol (number indicates average molecular weight in Da)
PIC	Preinitiation complex
RMSD	Root mean square deviation

---

RNA	Ribonucleic acid
rNTP	Ribonucleoside triphosphate
Rpb	Subunit of Pol II (=RNA polymerase B)
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
TAF	TBP-associated factor
TBP	TATA binding protein
TCEP	Tris (2-carboxyethyl) phosphine
TFII	Transcription factor of Pol II transcription
Tris	Trishydroxymethylaminomethane
<i>T. thermophilus</i>	<i>Thermus thermophilus</i>

## 11. REFERENCES

- Adelman K, Marr MT, Werner J, Saunders A, Ni ZY, Andrulis ED, Lis JT (2005) Efficient release from promoter-proximal stall sites requires transcript cleavage factor TFIIS. *Molecular Cell* **17**(1): 103-112
- Albert DA, Gudas LJ (1985) Ribonucleotide reductase activity and deoxyribonucleoside triphosphate metabolism during the cell cycle of S49 wild-type and mutant mouse T-lymphoma cells. *J Biol Chem* **260**(1): 679-684
- Alic N, Ayoub N, Landrieux E, Favry E, Baudouin-Cornu P, Riva M, Carles C (2007) Selectivity and proofreading both contribute significantly to the fidelity of RNA polymerase III transcription. *Proc Natl Acad Sci USA* **104**(25): 10400-10405
- Armache K-J, Kettenberger H, Cramer P (2003) Architecture of the initiation-competent 12-subunit RNA polymerase II. *Proc Natl Acad Sci USA* **100**: 6964-6968
- Armache K-J, Mitterweger S, Meinhart A, Cramer P (2005) Structures of complete RNA polymerase II and its subcomplex Rpb4/7. *J Biol Chem* **280**: 7131-7134
- Armstrong JA (2007) Negotiating the nucleosome: factors that allow RNA polymerase II to elongate through chromatin. *Biochem Cell Biol* **85**(4): 426-434
- Artsimovitch I, Landick R (2000) Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc Natl Acad Sci U S A* **97**(13): 7090-7095
- Artsimovitch I, Landick R (2002) The transcriptional regulator RfaH stimulates RNA chain synthesis after recruitment to elongation complexes by the exposed nontemplate DNA strand. *Cell* **109**(2): 193-203
- Astatke M, Ng K, Grindley ND, Joyce CM (1998) A single side chain prevents *Escherichia coli* DNA polymerase I (Klenow fragment) from incorporating ribonucleotides. *Proc Natl Acad Sci U S A* **95**(7): 3402-3407
- Asturias FJ, Meredith GD, Poglitsch CL, Kornberg RD (1997) Two conformations of RNA polymerase II revealed by electron crystallography. *J Mol Biol* **272**(4): 536-540.
- Bailey PJ, Dowhan DH, Franke K, Burke LJ, Downes M, Muscat GE (1997) Transcriptional repression by COUP-TF II is dependent on the C-terminal domain and involves the N-CoR variant, RIP13delta1. *J Steroid Biochem Mol Biol* **63**(4-6): 165-174
- Bar-Nahum G, Epshtein V, Ruckenstein AE, Rafikov R, Mustaev A, Nudler E (2005) A ratchet mechanism of transcription elongation and its control. *Cell* **120**(2): 183-193
- Batra VK, Beard WA, Shock DD, Pedersen LC, Wilson SH (2008) Structures of DNA polymerase beta with active-site mismatches suggest a transient abasic site intermediate during misincorporation. *Mol Cell* **30**(3): 315-324
- Bebenek K, Joyce CM, Fitzgerald MP, Kunkel TA (1990) The fidelity of DNA synthesis catalyzed by derivatives of *Escherichia coli* DNA polymerase I. *J Biol Chem* **265**(23): 13878-13887

- Bebenek K, Kunkel TA (1995) Analyzing fidelity of DNA polymerases. *Methods Enzymol* **262**: 217-232
- Beese LS, Steitz TA (1989) *Nucleic Acids and Molecular Biology Springer-Verlag, Berlin* **3**: 28-43
- Blank A, Gallant JA, Burgess RR, Loeb LA (1986) An RNA polymerase mutant with reduced accuracy of chain elongation. *Biochemistry* **25**(20): 5920-5928
- Borukhov S, Polyakov A, Nikiforov V, Goldfarb A (1992) GreA protein: a transcription elongation factor from *Escherichia coli*. *Proc Natl Acad Sci U S A* **89**(19): 8899-8902
- Borukhov S, Sagitov V, Goldfarb A (1993) Transcript cleavage factors from *E. coli*. *Cell* **72**(3): 459-466
- Boyer PL, Sarafianos SG, Arnold E, Hughes SH (2000) Analysis of mutations at positions 115 and 116 in the dNTP binding site of HIV-1 reverse transcriptase. *Proc Natl Acad Sci U S A* **97**(7): 3056-3061
- Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**: 248-254
- Briebe LG, Sousa R (2000) Roles of histidine 784 and tyrosine 639 in ribose discrimination by T7 RNA polymerase. *Biochemistry* **39**(5): 919-923
- Broennimann C, Eikenberry EF, Henrich B, Horisberger R, Huelsen G, Pohl E, Schmitt B, Schulze-Briese C, Suzuki M, Tomizaki T, Toyokawa H, Wagner A (2006) The PILATUS 1M detector. *J Synchrotron Radiat* **13**(Pt 2): 120-130
- Brueckner F, Cramer P (2008) Structural basis of transcription inhibition by alpha-amanitin and implications for RNA polymerase II translocation. *Nat Struct Mol Biol* **15**(8): 811-818
- Brueckner F, Hennecke U, Carell T, Cramer P (2007) CPD damage recognition by transcribing RNA polymerase II. *Science* **315**(5813): 859-862
- Brueckner F, Ortiz J, Cramer P (2009) A movie of the RNA polymerase nucleotide addition cycle. *Curr Opin Struct Biol* **19**(3): 294-299
- Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**(Pt 5): 905-921
- Buratowski S (2003) The CTD code. *Nat Struct Biol* **10**(9): 679-680
- Buratowski S, Hahn S, Guarente L, Sharp PA (1989) Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* **56**: 549-561
- Bushnell DA, Kornberg RD (2003) Complete RNA polymerase II at 4.1 Å resolution: implications for the initiation of transcription. *Proc Natl Acad Sci U S A* **100**: 6969-6972

Cairns BR, Lorch Y, Li Y, Lacomis L, Erdjument-Bromage H, Tempst P, Laurent B, Kornberg RD (1996) RSC, an abundant and essential chromatin remodeling complex. *Cell* **87**: 1249-1260

Catalano CE, Benkovic SJ (1989) Inactivation of DNA polymerase I (Klenow fragment) by adenosine 2',3'-epoxide 5'-triphosphate: evidence for the formation of a tight-binding inhibitor. *Biochemistry* **28**(10): 4374-4382

Chan CL, Wang D, Landick R (1997) Multiple interactions stabilize a single paused transcription intermediate in which hairpin to 3' end spacing distinguishes pause and termination pathways. *J Mol Biol* **268**(1): 54-68

Chao DM, Gadbois EL, Murray PJ, Anderson SF, Sonu MS, Parvin JD, Young RA (1996) A mammalian SRB protein associated with an RNA polymerase II holoenzyme. *Nature* **380**: 82-85

Chedin S, Riva M, Schultz P, Sentenac A, 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**(24): 3857-3871

Cho H, Orphanides G, Sun X, Yang XJ, Ogryzko V, Lees E, Nakatani Y, Reinberg D (1998) A human RNA polymerase II complex containing factors that modify chromatin structure. *Mol Cell Biol* **18**(9): 5355-5363

Craighead JL, Chang WH, Asturias FJ (2002) Structure of yeast RNA polymerase II in solution: implications for enzyme regulation and interaction with promoter DNA. *Struct Fold Des* **10**(8): 1117-1125.

Cramer P (2002) Common structural features of nucleic acid polymerases. *Bioessays* **24**(8): 724-729

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

Cramer P, Bushnell DA, Fu J, Gnatt AL, Maier-Davis B, Thompson NE, Burgess RR, Edwards AM, David PR, Kornberg RD (2000) Architecture of RNA polymerase II and implications for the transcription mechanism [see comments]. *Science* **288**(5466): 640-649

Cramer P, Bushnell DA, Kornberg RD (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**(5523): 1863-1876.

Damsma GE, Alt A, Brueckner F, Carell T, Cramer P (2007) Mechanism of transcriptional stalling at cisplatin-damaged DNA. *Nat Struct Mol Biol* **14**(12): 1127-1133

Darst SA, Edwards AM, Kubalek EW, Kornberg RD (1991) Three-dimensional structure of yeast RNA polymerase II at 16Å resolution. *Cell* **66**: 121-128

de la Mata M, Alonso CR, Kadener S, Fededa JP, Blaustein M, Pelisch F, Cramer P, Bentley D, Kornblihtt AR (2003) A slow RNA polymerase II affects alternative splicing in vivo. *Mol Cell* **12**(2): 525-532



- de Mercoyrol L, Corda Y, Job C, Job D (1992) Accuracy of wheat-germ RNA polymerase II. General enzymatic properties and effect of template conformational transition from right-handed B-DNA to left-handed Z-DNA. *Eur J Biochem* **206**(1): 49-58
- Dengl S, Cramer P (2009) Torpedo Nuclease Rat1 Is Insufficient to Terminate RNA Polymerase II in Vitro. *J Biol Chem* **284**(32): 21270-21279
- Derbyshire V, Freemont PS, Sanderson MR, Beese L, Friedman JM, Joyce CM, Steitz TA (1988) Genetic and crystallographic studies of the 3',5'-exonucleolytic site of DNA polymerase I. *Science* **240**(4849): 199-201
- Donahue JP, Turnbough CL, Jr. (1994) Nucleotide-specific transcriptional pausing in the pyrBI leader region of Escherichia coli K-12. *J Biol Chem* **269**(27): 18185-18191
- Echols H, Goodman MF (1991) Fidelity mechanisms in DNA replication. *Annu Rev Biochem* **60**: 477-511
- Edwards AM, Darst SA, Feaver WJ, Thompson NE, Burgess RR, Kornberg RD (1990) Purification and lipid layer crystallization of yeast RNA polymerase II. *Proc Natl Acad Sci USA* **87**: 2122-2126
- Edwards AM, Kane CM, Young RA, Kornberg RD (1991) Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter *in vitro*. *J Biol Chem* **266**: 71-75
- Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**(Pt 12 Pt 1): 2126-2132
- Epshtein V, Nudler E (2003) Cooperation between RNA polymerase molecules in transcription elongation. *Science* **300**(5620): 801-805
- Erie DA, Hajiseyedjavadi O, Young MC, von Hippel PH (1993) Multiple RNA polymerase conformations and GreA: Control of the fidelity of transcription. *Science* **262**: 867-873
- Fire A, Samuels M, Sharp PA (1984) Interactions between RNA polymerase II, factors, and template leading to accurate transcription. *J Biol Chem* **259**(4): 2509-2516
- Fish RN, Kane CM (2002) Promoting elongation with transcript cleavage stimulatory factors. *Biochim Biophys Acta* **1577**(2): 287-307
- Foster JE, Holmes SF, Erie DA (2001) Allosteric binding of nucleoside triphosphates to RNA polymerase regulates transcription elongation. *Cell* **106**(2): 243-252
- Freemont PS, Ollis DL, Steitz TA, Joyce CM (1986) A domain of the Klenow fragment of Escherichia coli DNA polymerase I has polymerase but no exonuclease activity. *Proteins* **1**(1): 66-73
- Fu J, Gnatt AL, Bushnell DA, Jensen G, J., Thompson NE, Burgess RR, David PR, Kornberg RD (1999) Yeast RNA polymerase II at 5 Å resolution. *Cell* **98**: 799-810

- Gao G, Orlova M, Georgiadis MM, Hendrickson WA, Goff SP (1997) Conferring RNA polymerase activity to a DNA polymerase: a single residue in reverse transcriptase controls substrate selection. *Proc Natl Acad Sci U S A* **94**(2): 407-411
- Gilmour DS, Fan R (2008) Derailing the locomotive: transcription termination. *J Biol Chem* **283**(2): 661-664
- Gnatt AL, Cramer P, Fu J, Bushnell DA, Kornberg RD (2001) Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* **292**(5523): 1876-1882.
- Gong XQ, Zhang C, Feig M, Burton ZF (2005) Dynamic error correction and regulation of downstream bubble opening by human RNA polymerase II. *Mol Cell* **18**(4): 461-470
- Gu W, Reines D (1995) Variation in the size of nascent RNA cleavage products as a function of transcript length and elongation competence. *J Biol Chem* **270**: 30441-30447
- Guo H, Price DH (1993) Mechanism of DmS-II-mediated pause suppression by Drosophila RNA polymerase II. *J Biol Chem* **268**(25): 18762-18770
- Gusarov I, Nudler E (1999) The mechanism of intrinsic transcription termination. *Mol Cell* **3**(4): 495-504
- Hausner W, Lange U, Musfeldt M (2000) Transcription factor S, a cleavage induction factor of the archaeal RNA polymerase. *J Biol Chem* **275**(17): 12393-12399
- Hawley DK, Roeder RG (1985) Separation and partial characterization of three functional steps in transcription initiation by human RNA polymerase II. *J Biol Chem* **260**(13): 8163-8172
- Henkin TM, Yanofsky C (2002) Regulation by transcription attenuation in bacteria: how RNA provides instructions for transcription termination/antitermination decisions. *Bioessays* **24**(8): 700-707
- Herbert KM, La Porta A, Wong BJ, Mooney RA, Neuman KC, Landick R, Block SM (2006) Sequence-resolved detection of pausing by single RNA polymerase molecules. *Cell* **125**(6): 1083-1094
- Herr AJ, Jensen MB, Dalmay T, Baulcombe DC (2005) RNA polymerase IV directs silencing of endogenous DNA. *Science* **308**(5718): 118-120
- Hirata A, Klein BJ, Murakami KS (2008) The X-ray crystal structure of RNA polymerase from Archaea. *Nature* **451**(7180): 851-854
- Hirose Y, Manley JL (2000) RNA polymerase II and the integration of nuclear events. *Genes Dev* **14**(12): 1415-1429
- Holmes SF, Erie DA (2003) Downstream DNA sequence effects on transcription elongation. Allosteric binding of nucleoside triphosphates facilitates translocation via a ratchet motion. *J Biol Chem* **278**(37): 35597-35608
- Holmes SF, Santangelo TJ, Cunningham CK, Roberts JW, Erie DA (2006) Kinetic investigation of Escherichia coli RNA polymerase mutants that influence nucleotide

discrimination and transcription fidelity. *Journal of Biological Chemistry* **281**(27): 18677-18683

Huang J, Brieba LG, Sousa R (2000) Misincorporation by Wild-Type and Mutant T7 RNA Polymerases: Identification of Interactions That Reduce Misincorporation Rates by Stabilizing the Catalytically Incompetent Open Conformation. *Biochemistry* **39**(38): 11571-11580

Huang Y, Eckstein F, Padilla R, Sousa R (1997) Mechanism of ribose 2'-group discrimination by an RNA polymerase. *Biochemistry* **36**(27): 8231-8242

Izban MG, Luse DS (1992) 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 MG, Luse DS (1993) SII-facilitated transcript cleavage in RNA polymerase II complexes stalled early after initiation occurs in primarily dinucleotide increments. *J Biol Chem* **268**: 12864-12873

Jasiak AJ, Armache KJ, Martens B, Jansen RP, Cramer P (2006) Structural biology of RNA polymerase III: subcomplex C17/25 X-ray structure and 11 subunit enzyme model. *Mol Cell* **23**(1): 71-81

Jensen GJ, Meredith G, Bushnell DA, Kornberg RD (1998) Structure of wild-type yeast RNA polymerase II and location of Rpb4 and Rpb7. *EMBO J* **17**(8): 2353-2358

Jeon C-J, Agarwal K (1996) Fidelity of RNA polymerase II transcription controlled by elongation factor TFIIS. *Proc Natl Acad Sci USA* **93**: 13677-13682

Johnson SJ, Beese LS (2004) Structures of mismatch replication errors observed in a DNA polymerase. *Cell* **116**(6): 803-816

Joyce CM (1997) Choosing the right sugar: how polymerases select a nucleotide substrate. *Proc Natl Acad Sci U S A* **94**(5): 1619-1622

Joyce CM, Steitz TA (1987) DNA polymerase I: from crystal structure to function via genetics. *Trends Biochem Sci* **12**(288-292)

Joyce CM, Sun XC, Grindley ND (1992) Reactions at the polymerase active site that contribute to the fidelity of Escherichia coli DNA polymerase I (Klenow fragment). *J Biol Chem* **267**(34): 24485-24500

Kabsch W (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *Journal of Applied Crystallography* **26**(6): 795-800

Kanno T, Huettel B, Mette MF, Aufsatz W, Jaligot E, Daxinger L, Kreil DP, Matzke M, Matzke AJ (2005) Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat Genet* **37**(7): 761-765

Kashkina E, Anikin M, Brueckner F, Pomerantz RT, McAllister WT, Cramer P, Temiakov D (2006) Template misalignment in multisubunit RNA polymerases and transcription fidelity. *Mol Cell* **24**(2): 257-266

- Kettenberger H, Armache K-J, 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, Cramer P (2004) Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol Cell* **16**: 955-965
- Kim YJ, Bjorklund S, Li Y, Sayre MH, Kornberg RD (1994) A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77**: 599-608
- Kireeva ML, Kashlev M (2009) Mechanism of sequence-specific pausing of bacterial RNA polymerase. *Proc Natl Acad Sci U S A* **106**(22): 8900-8905
- Kireeva ML, Lubkowska L, Komissarova N, Kashlev M (2003) Assays and affinity purification of biotinylated and nonbiotinylated forms of double-tagged core RNA polymerase II from *Saccharomyces cerevisiae*. *Methods Enzymol* **370**: 138-155
- Kireeva ML, Nedialkov YA, Cremona GH, Purtov YA, Lubkowska L, Malagon F, Burton ZF, Strathern JN, Kashlev M (2008) Transient reversal of RNA polymerase II active site closing controls fidelity of transcription elongation. *Mol Cell* **30**(5): 557-566
- Koleske AJ, Young RA (1994) An RNA polymerase II holoenzyme responsive to activators. *Nature* **368**: 466-469
- Komissarova N, Kashlev M (1997a) RNA polymerase switches between inactivated and activated states by translocating back and forth along the DNA and the RNA. *J Biol Chem* **272**: 15329-15338
- Komissarova N, Kashlev M (1997b) Transcription arrest: *Escherichia coli* RNA polymerase translocates backward, leaving the 3' end of the RNA intact and extruded. *Proc Natl Acad Sci USA* **94**: 1755-1760
- Kool ET (2002) Active site tightness and substrate fit in DNA replication. *Annu Rev Biochem* **71**: 191-219
- Korkhin Y, Unligil UM, Littlefield O, Nelson PJ, Stuart DI, Sigler PB, Bell SD, Abrescia NG (2009) Evolution of Complex RNA Polymerases: The Complete Archaeal RNA Polymerase Structure. *PLoS Biol* **7**(5): e102
- Kornberg A, Baker TA (1992) *DNA Replication Freeman, New York*
- Kuhn CD, Geiger SR, Baumli S, Gartmann M, Gerber J, Jennebach S, Mielke T, Tschochner H, Beckmann R, Cramer P (2007) Functional architecture of RNA polymerase I. *Cell* **131**(7): 1260-1272
- Kulaeva OI, Gaykalova DA, Studitsky VM (2007) Transcription through chromatin by RNA polymerase II: histone displacement and exchange. *Mutat Res* **618**(1-2): 116-129
- Kunkel TA, Bebenek K (2000) DNA replication fidelity. *Annu Rev Biochem* **69**: 497-529
- Kusser AG, Bertero MG, Naji S, Becker T, Thomm M, Beckmann R, Cramer P (2008) Structure of an archaeal RNA polymerase. *J Mol Biol* **376**(2): 303-307

- Kwok S, Kellogg DE, McKinney N, Spasic D, Goda L, Levenson C, Sninsky JJ (1990) Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. *Nucleic Acids Res* **18**(4): 999-1005
- Lai MD, Beattie KL (1988) Influence of DNA sequence on the nature of mispairing during DNA synthesis. *Biochemistry* **27**(5): 1722-1728
- Landick R (1997) RNA polymerase slides home: pause and termination site recognition. *Cell* **88**(6): 741-744
- Landick R (2006) The regulatory roles and mechanism of transcriptional pausing. *Biochem Soc Trans* **34**(Pt 6): 1062-1066
- Landick R (2009) Transcriptional pausing without backtracking. *Proc Natl Acad Sci U S A* **106**(22): 8797-8798
- Lange U, Hausner W (2004) Transcriptional fidelity and proofreading in Archaea and implications for the mechanism of TFS-induced RNA cleavage. *Mol Microbiol* **52**(4): 1133-1143
- Lee DN, Phung L, Stewart J, Landick R (1990) Transcription pausing by *Escherichia coli* RNA polymerase is modulated by downstream DNA sequences. *J Biol Chem* **265**: 15145-15153
- Lehmann E, Brueckner F, Cramer P (2007) Molecular basis of RNA-dependent RNA polymerase II activity. *Nature* **450**(7168): 445-449
- Li B, Carey M, Workman JL (2007) The role of chromatin during transcription. *Cell* **128**(4): 707-719
- Liu Y, Ranish JA, Aebersold R, Hahn S (2001) Yeast nuclear extract contains two major forms of RNA polymerase II mediator complexes. *J Biol Chem* **276**(10): 7169-7175
- Maki H, Kornberg A (1985) The polymerase subunit of DNA polymerase III of *Escherichia coli*. II. Purification of the alpha subunit, devoid of nuclease activities. *J Biol Chem* **260**(24): 12987-12992
- Maldonado E, Shiekhhattar R, Sheldon M, Cho H, Drapkin R, Rickert P, Lees E, Anderson CW, Linn S, Reinberg D (1996) A human RNA polymerase II complex associated with SRB and DNA-repair proteins. *Nature* **381**: 86-89
- McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ (2005) Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr* **61**(Pt 4): 458-464
- McCracken S, Fong N, Yankulov K, Ballantyne S, Pan G, Greenblatt J, Patterson SD, Wickens M, Bentley DL (1997) The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**(6614): 357-361
- McCulloch SD, Kunkel TA (2008) The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res* **18**(1): 148-161
- Meinhart A, Kamenski T, Hoepfner S, Baumli S, Cramer P (2005) A structural perspective of CTD function. *Genes Dev* **19**(12): 1401-1415

- Mendelman LV, Boosalis MS, Petruska J, Goodman MF (1989) Nearest neighbor influences on DNA polymerase insertion fidelity. *J Biol Chem* **264**(24): 14415-14423
- Mendelman LV, Petruska J, Goodman MF (1990) Base mispair extension kinetics. Comparison of DNA polymerase alpha and reverse transcriptase. *J Biol Chem* **265**(4): 2338-2346
- Minakhin L, Bhagat S, Brunning A, Campbell EA, Darst SA, Ebright RH, 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**(3): 892-897.
- Morales JC, Kool ET (2000) Importance of terminal base pair hydrogen-bonding in 3'-end proofreading by the Klenow fragment of DNA polymerase I. *Biochemistry* **39**(10): 2626-2632
- Naji S, Bertero MG, Spitalny P, Cramer P, Thomm M (2007) Structure function analysis of the RNA polymerase cleft loops elucidates initial transcription, DNA unwinding and RNA displacement. *Nucleic Acids Res*
- Nakajima T, Uchida C, Anderson SF, Parvin JD, Montminy M (1997) Analysis of a cAMP-responsive activator reveals a two-component mechanism for transcriptional induction via signal-dependent factors. *Genes Dev* **11**(6): 738-747
- Nesser NK, Peterson DO, Hawley DK (2006) RNA polymerase II subunit Rpb9 is important for transcriptional fidelity in vivo. *Proc Natl Acad Sci U S A* **103**(9): 3268-3273
- Neuman KC, Abbondanzieri EA, Landick R, Gelles J, Block SM (2003) Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking. *Cell* **115**(4): 437-447
- Nudler E (1999) Transcription elongation: structural basis and mechanisms. *J Mol Biol* **288**(1): 1-12
- Nudler E, Mustaev A, Lukhtanov E, Goldfarb A (1997) The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell* **89**: 38-41
- Ollis DL, Brick P, Hamlin R, Xuong NG, Steitz TA (1985) Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. *Nature* **313**: 762-
- Onodera Y, Haag JR, Ream T, Nunes PC, Pontes O, Pikaard CS (2005) Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120**(5): 613-622
- Opalka N, Chlenov M, Chacon P, Rice WJ, Wriggers W, Darst SA (2003) Structure and function of the transcription elongation factor GreB bound to bacterial RNA polymerase. *Cell* **114**(3): 335-345
- Orlova M, Newlands J, Das A, Goldfarb A, Borukhov S (1995) Intrinsic transcript cleavage activity of RNA polymerase. *Proc Natl Acad Sci U S A* **92**(10): 4596-4600

- Ossipow V, Tassan J-P, Nigg EA, Schibler U (1995) A mammalian RNA polymerase II holoenzyme containing all components required for promoter-specific transcription initiation. *Cell* **83**: 137-146
- Palangat M, Hittinger CT, Landick R (2004) Downstream DNA selectively affects a paused conformation of human RNA polymerase II. *J Mol Biol* **341**(2): 429-442
- Palangat M, 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**(2): 265-282
- Palangat M, Meier TI, Keene RG, Landick R (1998) Transcriptional pausing at +62 of HIV-1 nascent RNA modulates formation of the TAR RNA structure. *Mol Cell* **1**: 1033-1042
- Pan T, Artsimovitch I, Fang XW, Landick R, Sosnick TR (1999) Folding of a large ribozyme during transcription and the effect of the elongation factor NusA. *Proc Natl Acad Sci U S A* **96**(17): 9545-9550
- Park NJ, Tsao DC, Martinson HG (2004) The two steps of poly(A)-dependent termination, pausing and release, can be uncoupled by truncation of the RNA polymerase II carboxyl-terminal repeat domain. *Mol Cell Biol* **24**(10): 4092-4103
- Perrino FW, Loeb LA (1989) Differential extension of 3' mispairs is a major contribution to the high fidelity of calf thymus DNA polymerase-alpha. *J Biol Chem* **264**(5): 2898-2905
- Perrino FW, Preston BD, Sandell LL, Loeb LA (1989) Extension of mismatched 3' termini of DNA is a major determinant of the infidelity of human immunodeficiency virus type 1 reverse transcriptase. *Proc Natl Acad Sci U S A* **86**(21): 8343-8347
- Poglitsch CL, Meredith GD, Gnatt AL, Jensen GJ, Chang WH, Fu J, Kornberg RD (1999) Electron crystal structure of an RNA polymerase II transcription elongation complex. *Cell* **98**: 791-798
- Pomerantz RT, Temiakov D, Anikin M, Vassilyev DG, McAllister WT (2006) A mechanism of nucleotide misincorporation during transcription due to template-strand misalignment. *Mol Cell* **24**(2): 245-255
- Pontier D, Yahubyan G, Vega D, Bulski A, Saez-Vasquez J, Hakimi MA, Lerbs-Mache S, Colot V, Lagrange T (2005) Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. *Genes Dev* **19**(17): 2030-2040
- Ream TS, Haag JR, Wierzbicki AT, Nicora CD, Norbeck AD, Zhu JK, Hagen G, Guilfoyle TJ, Pasa-Tolic L, Pikaard CS (2009) Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II. *Mol Cell* **33**(2): 192-203
- Reichard P (1985) Ribonucleotide reductase and deoxyribonucleotide pools. *Basic Life Sci* **31**: 33-45
- Reines D (1992) Elongation factor-dependent transcript shortening by template-engaged RNA polymerase II. *J Biol Chem* **267**(6): 3795-3800

- Ring BZ, Yarnell WS, Roberts JW (1996) Function of *E. coli* RNA polymerase sigma factor sigma 70 in promoter-proximal pausing. *Cell* **86**(3): 485-493
- Rosenberger RF, Hilton J (1983) The frequency of transcriptional and translational errors at nonsense codons in the *lacZ* gene of *Escherichia coli*. *Mol Gen Genet* **191**(2): 207-212
- Samkurashvili I, Luse DS (1996) Translocation and transcriptional arrest during transcript elongation by RNA polymerase II. *J Biol Chem* **271**(38): 23495-23505
- Samuels M, Sharp PA (1986) Purification and characterization of a specific RNA polymerase II transcription factor. *J Biol Chem* **261**(5): 2003-2013
- Saxowsky TT, Doetsch PW (2006) RNA polymerase encounters with DNA damage: transcription-coupled repair or transcriptional mutagenesis? *Chem Rev* **106**(2): 474-488
- Scheuermann RH, Echols H (1984) A separate editing exonuclease for DNA replication: the epsilon subunit of *Escherichia coli* DNA polymerase III holoenzyme. *Proc Natl Acad Sci U S A* **81**(24): 7747-7751
- Shaw RJ, Bonawitz ND, Reines D (2002) Use of an in vivo reporter assay to test for transcriptional and translational fidelity in yeast. *J Biol Chem* **277**(27): 24420-24426
- Sims RJ, 3rd, Belotserkovskaya R, Reinberg D (2004) Elongation by RNA polymerase II: the short and long of it. *Genes Dev* **18**(20): 2437-2468
- Sosunov V, Sosunova E, Mustaev A, Bass I, Nikiforov V, Goldfarb A (2003) Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerase. *Embo J* **22**(9): 2234-2244
- Sousa R, Padilla R (1995) A mutant T7 RNA polymerase as a DNA polymerase. *EMBO J* **14**: 18
- Steitz TA (1998) A mechanism for all polymerases. *Nature* **391**(6664): 231-232
- Studier FW (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41**(1): 207-234
- Surratt CK, Milan SC, Chamberlin MJ (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 USA* **88**: 7983-7987
- Svetlov V, Vassylyev DG, Artsimovitch I (2004) Discrimination against deoxyribonucleotide substrates by bacterial RNA polymerase. *J Biol Chem* **279**(37): 38087-38090
- Tang H, Liu Y, Madabusi L, Gilmour DS (2000) Promoter-proximal pausing on the *hsp70* promoter in *Drosophila melanogaster* depends on the upstream regulator. *Mol Cell Biol* **20**(7): 2569-2580
- Temiakov D, Patlan V, Anikin M, McAllister WT, Yokoyama S, Vassylyev DG (2004) Structural basis for substrate selection by t7 RNA polymerase. *Cell* **116**(3): 381-391



Thomas MC, Chiang CM (2006) The general transcription machinery and general cofactors. *Crit Rev Biochem Mol Biol* **41**(3): 105-178

Thomas MJ, Platas AA, Hawley DK (1998) Transcriptional fidelity and proofreading by RNA polymerase II. *Cell* **93**(4): 627-637

Thompson NE, Burgess RR (1996) Immunoaffinity purification of RNA polymerase II and transcription factors using polyol-responsive monoclonal antibodies. *Methods Enzymol* **274**: 513-526

Todone F, Brick P, Werner F, Weinzierl RO, Onesti S (2001) Structure of an archaeal homolog of the eukaryotic RNA polymerase II RPB4/RPB7 complex. *Mol Cell* **8**(5): 1137-1143.

Toulokhonov I, Zhang J, Palangat M, Landick R (2007) A central role of the RNA polymerase trigger loop in active-site rearrangement during transcriptional pausing. *Mol Cell* **27**(3): 406-419

Van Dyke MW, Sawadogo M, Roeder RG (1989) Stability of transcription complexes on class II genes. *Mol Cell Biol* **9**(1): 342-344

Vassylyev DG, Sekine S, Laptenko O, Lee J, Vassylyeva MN, Borukhov S, Yokoyama S (2002) Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* **417**(6890): 712-719.

Vassylyev DG, Vassylyeva MN, Perederina A, Tahirov TH, Artsimovitch I (2007a) Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* **448**(7150): 157-162

Vassylyev DG, Vassylyeva MN, Zhang J, Palangat M, Artsimovitch I, Landick R (2007b) Structural basis for substrate loading in bacterial RNA polymerase. *Nature* **448**(7150): 163-168

Wang D, Bushnell DA, Westover KD, Kaplan CD, Kornberg RD (2006) Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell* **127**(5): 941-954

Wang D, Bushnell, D.A., Huang, X., Westover, K.D., Levitt, M., Kornberg, R.D. (2009) Structural Basis of Transcription: Backtracked RNA Polymerase II at 3.4 Å Resolution. *Science* **324**(5931): 1203-1206

Wang D, Hawley DK (1993) Identification of a 3' - 5' exonuclease activity associated with human RNA polymerase II. *Proc Natl Acad Sci USA* **90**: 843-847

Wang D, Meier TI, Chan CI, Feng G, Lee DN, Landick R (1995) Discontinuous movements of DNA and RNA in RNA polymerase accompany formation of a paused transcription complex. *Cell* **81**: 341-350

Westover KD, Bushnell DA, Kornberg RD (2004a) Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell* **119**(4): 481-489

Westover KD, Bushnell DA, Kornberg RD (2004b) Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science* **303**(5660): 1014-1016

- Whitehall SK, Bardeleben C, Kassavetis GA (1994) Hydrolytic cleavage of nascent RNA in RNA polymerase III ternary transcription complexes. *J Biol Chem* **269**(3): 2299-2306
- Wierzbicki AT, Haag JR, Pikaard CS (2008) Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* **135**(4): 635-648
- Wilson CJ, Chao DM, Imbalzano AN, Schnitzler GR, Kingston RE, Young RA (1996) RNA polymerase II holoenzyme contains SWI/SNF regulators involved in chromatin remodeling. *Cell* **84**(2): 235-244
- Wind M, Reines D (2000) Transcription elongation factor SII. *Bioessays* **22**(4): 327-336
- Winkler ME, Yanofsky C (1981) Pausing of RNA polymerase during in vitro transcription of the tryptophan operon leader region. *Biochemistry* **20**(13): 3738-3744
- Wu S-Y, Chiang C-M (1998) Properties of PC4 and an RNA polymerase II complex in directing activated and basal transcription *in vitro*. *J Biol Chem* **273**: 12492-12498
- Wu WH, Hampsey M (1999) An activation-specific role for transcription factor TFIIIB *in vivo*. *Proc Natl Acad Sci USA* **96**: 2764-2769
- Yakhnin AV, Yakhnin H, Babitzke P (2006) RNA polymerase pausing regulates translation initiation by providing additional time for TRAP-RNA interaction. *Mol Cell* **24**(4): 547-557
- Yang G, Franklin M, Li J, Lin TC, Konigsberg W (2002) A conserved Tyr residue is required for sugar selectivity in a Pol alpha DNA polymerase. *Biochemistry* **41**(32): 10256-10261
- Yin YW, Steitz TA (2004) The structural mechanism of translocation and helicase activity in t7 RNA polymerase. *Cell* **116**(3): 393-404
- Yonaha M, Proudfoot NJ (1999) Specific transcriptional pausing activates polyadenylation in a coupled in vitro system. *Mol Cell* **3**(5): 593-600
- Yuryev A, Patturajan M, Litingtung Y, Joshi RV, Gentile C, Gebara M, Corden JL (1996) The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. *Proc Natl Acad Sci U S A* **93**(14): 6975-6980
- Zenkin N, Yuzenkova Y, Severinov K (2006) Transcript-assisted transcriptional proofreading. *Science* **313**(5786): 518-520
- Zhang G, Campbell EA, Minakhin L, Richter C, Severinov K, Darst SA (1999) Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* **98**(6): 811-824

## 12. CURRICULUM VITAE

### JASMIN F. SYDOW

Date of birth: September 29, 1978

Place of birth: Munich, Germany

Nationality: German

Family status: unmarried



### Educational background

---

#### *since Jan 2005: PhD thesis*

Gene Center, University of Munich (LMU), Munich, Germany. Group of Prof. Dr. Patrick Cramer

Topic: *Structural basis of transcription: RNA polymerase II fidelity mechanisms and RNA 3' fraying*

#### *Jun 2005 – May 2007: PhD fellowship „Cusanuswerk–Bischöfliche Studienförderung“*

#### *Jan 2005 – Feb 2007: Scientific Work in Structural Biology towards PhD degree*

Institute of Molecular Biology of Barcelona, CSIC, Parc Científic de Barcelona, Barcelona, Spain. Group of Prof. Dr. Miquel Coll

Topic: *Biochemical and structural analysis of mitochondrial regulatory proteins: Studies on the mitochondrial transcription factor A in complex with DNA*

#### *Jul 2003 – Jun 2004: Diploma Thesis in Microbiology*

Department of Microbiology and Genetics, University of Munich (LMU), Munich, Germany. Group of Prof. Dr. Volker Müller

Topic: *Biosynthesis of compatible solutes in the moderately halophilic bacterium Halobacillus halophilus: Discovery and characterisation of a salt-dependent and Cl<sup>-</sup>-stimulated glutamine synthetase*

#### *Nov 1998 – Jul 2004: Diplom-Biol. Univ.*

University of Munich (LMU), Munich, Germany

Main subjects: Microbiology, Biochemistry, Genetics and Immunobiology

#### *Sep 1989 – Jun 1998: Abitur*

Ernst-Mach-Gymnasium (Secondary School), Haar, Germany

Majors: Biology, French

## Publications

---

Sydow, J. F., and Cramer, P. (2009). Error prevention, recognition, and removal by cellular RNA Polymerases. *Curr Opin Struct Biol. In press.*

Sydow, J.F., Brueckner, F., Cheung, A.C., Damsma, G.E., Dengl, S., Lehmann, E., Vassylyev, D., Cramer, P. (2009). Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. *Mol Cell* **34**(6): 710-21.

Brueckner, F., Armache, K.-J., Cheung, A.C., Damsma, G.E., Kettenberger, H., Lehmann, E., Sydow, J.F., Cramer, P. (2009). Structure-function studies of the RNA polymerase II elongation complex. *Acta Crystallogr D Biol Crystallogr.* **65**, 112-120.

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

Saum, S.H., Roeßler, M., Koller, C., Sydow, J.F., Müller, V. (2007). Glutamate restores growth but not motility in the absence of chloride in the moderate halophile *Halobacillus halophilus*. *Extremophiles* **11** (5), 711-717.

Saum, S.H.\*, Sydow, J.F.\*, Pfeiffer, F., Oesterhelt, D., Müller, V. (2006). Biochemical and molecular characterization of the biosynthesis of glutamine and glutamate, two major compatible solutes in the moderately halophilic bacterium *Halobacillus halophilus*. *Journal of Bacteriology* **188** (19), 6808-6815. \*contributed equally.

## Conferences and Courses

---

**Mechanisms of Eukaryotic Transcription**, Cold Spring Harbor, USA, August 2009

**BioCrys2008 – Fundamentals of Modern Methods in Biocrystallography**, Oeiras, Portugal, October 2008

**Murnau Conference on Structural Biology of Disease Mechanisms**, Murnau, Germany, September 2007

**International School on Biological Crystallization**, Granada, Spain, May 2006