

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

Evolution of two modes of intrinsic RNA polymerase transcript cleavage



Wenjie Ruan
aus Anhui, P.R.China

2011

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

Evolution of two modes of intrinsic RNA polymerase transcript cleavage

Wenjie Ruan
aus Anhui, P.R.China

2011

Erklärung

Diese Dissertation wurde im Sinne von §13 Abs. 3 der Promotionsordnung vom 29. Januar 1998 (in der Fassung der vierten Änderungssatzung vom 26. November 2004) von Herrn Prof. Dr. Patrick Cramer betreut.

Ehrenwörtliche Versicherung

Diese Dissertation wurde selbständig und ohne unerlaubte Hilfe erarbeitet.

München, den 06. April 2011

Wenjie Ruan

Dissertation eingereicht am 07. April 2011

1. Gutachter: Prof. Dr. Patrick Cramer

2. Gutachter: Prof. Dr. Dietmar Martin

Mündliche Prüfung am 11. Mai 2011

Acknowledgements

Five years ago, on the beautiful fall of 2006, when I first set foot on this land, colorful leaves, blue sky, smiling and courteous people, were the first impressions Deutschland gave me. This was my first time coming abroad, touching a completely different world and culture. During the last years, I harvested a lot, both on academic life, and on mentality, grown up to be a strong person. The long journey would not have been possible without the help of many people. I wish to give them my sincere thanks here.

Prof. Patrick Cramer, you are the first and most important person I want to thank. As a foreign student, huge differences on culture and language once gave me a lot of pressure. You understand me, tolerate my complaints and always support and encourage me patiently. Your contagious enthusiasm for science and life will be an inspiration for me forever. You are the most excellent professor that I can imagine. I feel so lucky and honored for coming here and being your student!

I thank Elisabeth for an efficient collaboration and discussions.

I specially thank Dirk for teaching me on the structure determination of the Pol II variant hand by hand and never being annoyed for my endless small questions. Your excellent crystallography knowledge really impresses me.

Moreover, I would like to thank Anselm for collaborating in the project about archaeal preinitiation complex. I enjoyed the time working with you because of your optimistic presence and consideration. I wish you happiness and success.

Both the projects about archaeal transcription preinitiation complex and intrinsic cleavage activity were strongly supported by Prof. Michael Thomm in Regensburg, who provided me with many carefully prepared *P.fu* cells, and helpful discussions about the projects.

I am thankful to all the present and former members of the Cramer lab for their help and nice atmosphere in the lab. Many thanks to Dr. Michaela Bertero for leading me to the fantastic archaea world. Jasmin, thank you for your kind and patient help in the beginning of my cleavage project. Gerke, you are a so kind-hearted person and I sincerely thank you for all the conversations and editing my thesis.

I wish to give a big thank to Alan for your endless warm-hearted help. I not only learnt so much helpful scientific expertise from you, but also gained familiar

eastern friendship. I really thank you for editing my thesis so kindly.

I am very thankful to the members of my Thesis Advisory Committee, Dr. Dierk Niessing and Dr. Andreas Bracher, for their advice and support. I also want to thank Dr. Hans-Joerg Schaeffer at the coordination office of IMPRS, who helped me so much during last years.

亲爱的爸爸妈妈，感谢你们，感谢你们如深海般的包容，感谢你们四年来如涓涓细流般绵延不绝的关切与爱，今天，希望我的成功能带给你们幸福与喜悦，愿这快乐伴随我们一生！

亲爱的姥爷，没能见上你最后一面是我今生最大的遗憾，但我知道，在我戴上“小桌子”那一天，你会在天堂里对我微笑。姥爷，你是我求学路上永远的灯塔，我想念你，愿你在天堂里永远幸福……

I wish to dedicate this work to my most beloved parents.

Summary

DNA-dependent RNA polymerases synthesize RNA transcripts according to the information carried on the DNA template. During gene transcription, the RNA polymerase (Pol) active center can also catalyze RNA cleavage. This intrinsic cleavage activity is strong for Pol I and Pol III, but very weak for Pol II. Accessory factor SII/TFIIS is required for rapid and effective cleavage in Pol II. The reason for this difference is unclear since the active centers of the polymerases are virtually identical. Work in this thesis shows that Pol II gains strong cleavage activity when the C-terminal zinc ribbon domain (C-ribbon) of subunit Rpb9 is replaced by its counterpart from the Pol III subunit C11. X-ray analysis shows that the C-ribbon has detached from its site on the Pol II surface and is mobile. Mutagenesis indicates that the C-ribbon transiently inserts into the Pol II pore to complement the active center. This mechanism is also used by SII/TFIIS, the factor that can bind Pol II and induce strong RNA cleavage. Together with published data, these results indicate that Pol I and Pol III contain catalytic C-ribbons that complement the active center, whereas Pol II contains a non-catalytic C-ribbon that is immobilized on the enzyme surface. Evolution of the Pol II system may have rendered mRNA transcript cleavage controllable by the dissociable factor TFIIS, to enable promoter-proximal gene regulation and elaborate 3'-processing and transcription termination.

Publication

Ruan, W., Lehmann, E., Thomm, M., Kostrewa, Cramer P. (2011)
Evolution of two modes of intrinsic RNA polymerase transcript cleavage.
J Biol Chem. *Paper in Press*.
doi: 10.1074/jbc.M111.222273.

Table of contents

Erklärung	II
Ehrenwörtliche Versicherung	II
Acknowledgements	III
Summary	IV
Publication	VII
Chapter I: General Introduction	1
1. Transcription mechanism and machinery	1
2. Three boundary theory and archaea	5
3. Archaeal and eukaryotic transcription machinery.....	6
Chapter II: Evolution of two modes of intrinsic RNA polymerase transcript cleavage	8
1. Introduction	8
1.1. Chemical mechanism of transcriptional cleavage.....	8
1.2. Weak intrinsic cleavage activity of RNA polymerases	9
1.3. Extrinsic factors induce transcriptional cleavage	10
1.4. Eukaryotic RNA polymerase with strong intrinsic cleavage activity.....	16
1.5. Cleavage activity in archaeal RNA polymerase.....	21
1.6. Aims and objectives of this thesis.....	22
2. Results	24
2.1. A Pol II variant with strong intrinsic RNA cleavage.....	24
2.2. The cleavage-active Pol II variant contains a mobile C-ribbon.....	31
2.3. Evidence that the C-ribbon is catalytic and binds the Pol II pore.....	35
2.4. The C-ribbon could reach the pore and active center through a long linker	37
2.5. The C11 C-ribbon functions in the Pol II pore	38
2.6. Catalytic C-ribbons are conserved between archaea and eukarya	38
3. Discussion	39
3.1. Two cleavage models in RNA polymerases	39
3.2. Evolution of the cleavage activity	43
4. Conclusions and Outlook.....	46

5. Experimental procedures	48
5.1. Isolation of yeast genomic DNA	48
5.2. Cloning, Expression and purification of cleavage factors	49
5.3. Assembly of transcription elongation complexes	51
5.4. Transcription cleavage assay	52
5.5. Crystallization and structure determination.....	53
Chapter III: Structure study of archaeal transcription pre-initiation complex(PIC) (unpublished)	56
1. Introduction	56
1.1. Transcription initiation in eukarya	56
1.2. Transcription initiation in archaea	59
1.3. Aim of this study	65
2. Results	66
2.1. Assembly of archaeal PIC	66
2.2. A tailed scaffold with RNA improves formation of single crystals	68
2.3. Truncation of the acidic C-terminus of TBP to improve crystals	71
2.4. TFE did not help the crystallization on bubble scaffold	73
2.5. Trial using a histone promoter	73
2.6. Different crystallization techniques and post-crystallization trials	74
3. Discussion and outlook	76
4. Experimental procedures	80
4.1. Oligonucleotides and cloning.....	80
4.2. Expression and purification of transcription initiation factors	81
4.3. Purification of <i>P.fu</i> RNA polymerase	84
4.4. Assembly of PIC	87
4.5. Crystallization	88
Abbreviations	90
References	92
Curriculum vitae – Wenjie Ruan	110

Chapter I: GENERAL INTRODUCTION

1. Transcription mechanism and machinery

1.1 Transcription mechanism

Transcription is a fundamental cellular process in which genetic information is transferred from DNA to RNA. Based on the information carried on RNA, polypeptide chains of protein are further synthesized by another key process called translation. This flow is known as the central dogma of molecular biology (Crick, 1970). Transcription of a eukaryotic protein-coding gene can be divided into five stages: pre-initiation, initiation, promoter clearance, elongation and termination (figure 1).

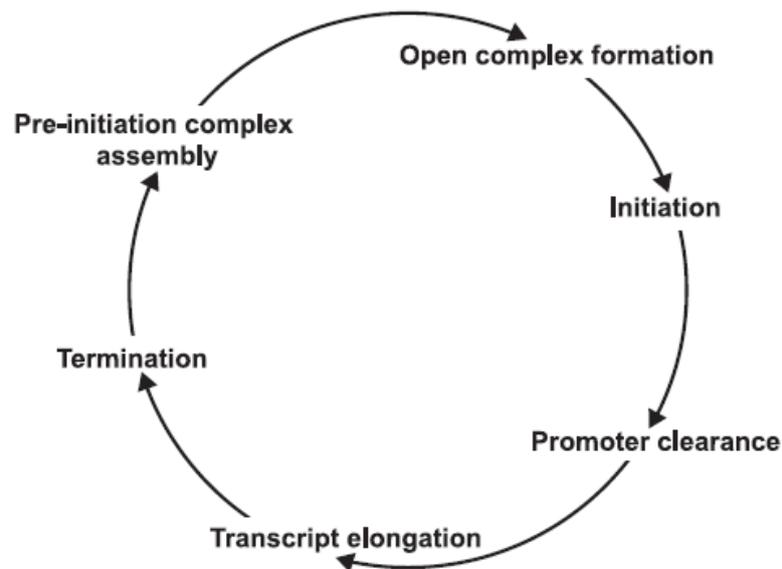


Figure 1. The transcription cycle. Adapted from (Svejstrup, 2004).

In the early stages of the transcription cycle, specific DNA elements in the core promoter are recognized by initiation factors. The common core promoter elements includes the TATA-box, TFIIB recognition upstream and downstream elements (BRE^u and BRE^d), the downstream promoter element (DPE) and the initiator element (Inr) (Baumann, *et al.*, 2010; Smale and Kadonaga, 2003).

Other cis-acting elements, including enhancers, silencers and insulators, are involved in regulating gene expression. Assembly of the pre-initiation complex (PIC) is the first step of transcription initiation. The PIC is composed of RNA polymerase II (Pol II) and general transcription factors (GTFs): TFIIA, TFIIB, TFIID, TFIIE, TFIIIF, TFIIH. Two assembly pathways are possible: a sequential assembly pathway, and the RNAPII holoenzyme pathway. TFIIH contribute to the promoter melting, which can be stimulated by TFIIE (Thomas and Chiang, 2006). After the open complex forms, RNA polymerase can initiate phosphodiester bond synthesis. Then, phosphorylation of CTD disrupts the binding to the initiation specific factors and interacts with elongation specific factors, leading to promoter clearance (Svejstrup, 2004). Pol II thus escapes from the promoter and enters the elongation phase. The pre-mRNA is synthesized accompanied by co-transcriptional processing factors involved in capping and splicing (Bentley, 2002). These processing factors are recruited through the CTD which is kept phosphorylated during elongation. Besides processing, many other elongation specific factors contribute to passage through chromatin, phosphorylation of CTD, regulation of the elongating rate and efficiency, proofreading and packaging of RNA (Shilatifard, 1998; Shilatifard, *et al.*, 2003; Svejstrup, 2004). Incorporation of NTP to the nascent RNA chain follows a two-metal ion mechanism in a well organized nucleotide addition cycle (NAC) (Brueckner and Cramer, 2008; Steitz, 1998). The transcription termination stage includes the release of transcript and polymerase from the DNA template. Compared to bacterial RNA polymerase (RNAP) (Henkin, 2000) and RNA polymerases I (Pol I) and III (Pol III) (Paule and White, 2000), termination of Pol II is less well known. The mRNA 3' end processing is indicated to regulate termination (Proudfoot, *et al.*, 2002). After displacement from the DNA template, the CTD is dephosphorylated and can rebind the initiation factors. Pol II then recycle for a new round of transcription.

1.2 DNA-dependent RNA polymerases

Two main classes of RNA polymerases can be summarized: single-subunit and multiple-subunit RNA polymerases (Cramer, 2002). The structures of single-subunit RNA polymerases show similarities with DNA polymerases, with a hand-like architecture. T7 RNAP is a best studied single-subunit RNAP. Multiple-subunit RNAPs include those from plant chloroplasts, bacteria, archaea and eukarya. Bacteria have only one RNA polymerase transcribing all different genes and high resolution X-ray structures from *Thermus aquaticus* and *Thermus thermophilus* were determined in the past few years (Murakami, *et al.*, 2002; Vassylyev, *et al.*, 2002; Zhang, *et al.*, 1999). The overall structure consists of a core enzyme, including five subunits: α_2 , β , β' , ω , with a total molecular weight of 400kDa, and an additional subunit, σ , which is part of the holoenzyme, and only required for initiation. The polymerase is shaped like a crab claw with an internal channel. The active site is located on the back wall of the channel, having an essential Mg^{2+} . Archaea also contain only one RNA polymerase. Two X-ray structures are also available, both from the *Sulfolobus* genus (*Sulfolobus solfataricus* and *Sulfolobus shibatae*) (Hirata, *et al.*, 2008; Korkhin, *et al.*, 2009). In eukarya, different kinds of RNA polymerases are responsible for different kinds of genes. RNA pol I transcribes only ribosomal DNA, and is located in nucleolus. A single precursor transcript is then processed to mature 5.8S, 18S and 28S rRNA. RNA pol II transcripts include all the protein-coding genes, small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs), and micro RNAs (miRNAs). RNA pol III transcribes a diverse group of genes to synthesize very short RNAs including transfer RNAs (tRNAs), 5S ribosomal RNA (5S rRNA) and U6 small nuclear RNA (U6 snRNA). Both RNA pol II and III are located in the nucleus. RNA pol I, II and III comprise 14, 12 and 17 subunits and a molecular weight of 589, 514, 693kDa respectively (Table 1). Five core subunits are conserved in all three kingdoms of life, forming the core enzyme. The two largest subunits are homologous to the subunits β and β' of bacterial RNAP. Other subunits are on the periphery of

the core enzyme. RNA pol II is a key focus of many research papers after the emergence of high resolution X-ray structures from *Saccharomyces cerevisiae* (Cramer, *et al.*, 2001; Kettenberger, *et al.*, 2004). X-ray structures for RNA pol I and III are still lacking, only EM structures could be obtained (Fernandez-Tornero, *et al.*, 2010; Fernandez-Tornero, *et al.*, 2007; Kuhn, *et al.*, 2007). Plants also have another two RNA polymerases: Pol IV and V (Ream, *et al.*, 2009).

Table 1. Subunit composition of multisubunit RNA polymerases.

RNA polymerase	Pol I	Pol II	Pol III	Archaea ^b	Bacteria
core	A190	Rpb1	C160	A' + A''	β'
core	A135	Rpb2	C128	B	β
core	AC40	Rpb3	AC40	D	α
core	AC19	Rpb11	AC19	L	α
core/common	Rpb6 (ABC23)	Rpb6	Rpb6	K	ω
common	Rpb5 (ABC27)	Rpb5	Rpb5	H	-
common	Rpb8 (ABC14.5)	Rpb8	Rpb8	G	-
common	Rpb10 (ABC10β)	Rpb10	Rpb10	N	-
common	Rpb12 (ABC10α)	Rpb12	Rpb12	P	-
	A12.2	Rpb9	C11	-	-
Rpb4/7	A14	Rpb4	C17	F	-
complexes	A43	Rpb7	C25	E'	-
TFIIF-like subcomplex ^a	A49	(Tfg1/Rap74)	C37	-	-
	A34.5	(Tfg2/Rap30)	C53	-	-
Pol III-specific subcomplex	-	-	C82	-	-
	-	-	C34	-	-
	-	-	C31	-	-

^aThe 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 Tfg1, Tfg2, and Tfg3 in *Saccharomyces cerevisiae*, and of subunits Rap74 and Rap30 in human.

^bArchaea RNAP has another subunit Rpo13, which does not have a homolog in eukaryotic polymerases, but architecturally corresponds to an insertion in bacterial β' subunit.

2. Three boundary theory and archaea

In 1977, a landmark discovery was that Woese and Fox proposed a previously unrecognized group of bacteria: archaeobacteria, as a third form of life, based on comparisons of 16S/18S ribosomal RNA of organisms from bacteria, archaea and eukarya (Woese and Fox, 1977). The distinct nature of archaea resulted in the three boundary theory that organisms are divided to three kingdoms: bacteria, archaea and eukarya (Woese, *et al.*, 1990). Archaea look similar to bacteria in phenotype: most metabolic pathways, cellular size and morphology, the absence of nucleus or cytoskeleton and a circular genome. But in the central information transfer processes such as transcription and translation, archaea appear to be more related to eukarya than bacteria (Bell and Jackson, 1998; Huet, *et al.*, 1983; Olsen and Woese, 1997; Rivera, *et al.*, 1998). Archaea have a RNA polymerase which contains subunits that are homologous to eukaryotic polymerases and absent in bacteria. Especially, archaea have extremely similar mechanism of transcription initiation, both in the promoter character and initiation factors. However, the regulation of archaeal transcription is more the bacterial-like. In the translation process, initiation factors are also homologous to eukaryotic factors. Recruitment of initiator tRNA-IF2-GTP complex happens before the mRNA binding, and archaea use methionine instead of formyl methionine in the initiating tRNA. But archaeal ribosome recognizes the start codon in a bacterial way: based on the complementary of "Shine-Dalgarno sequence" and 16S rRNA. Thus, archaea have a mosaic of eukaryotic and bacterial features in transcription and translation processes. Because of its simplicity and similarity of the transcription/translation apparatus, the archaeal system now receives more attention as a model to study the fundamental mechanisms of eukaryotic molecular central processes.

3. Archaeal and eukaryotic transcription machinery

Purified archaeal transcription machinery is found to be far more complex than bacterial RNAP, but closely resembles that of eukarya (Zillig, *et al.*, 1979). Archaeal RNAP contains at least 12 subunits with a total molecular weight of about 370 kDa, named alphabetically (Table 1). EM structures and X-ray structures of archaeal RNAP were obtained recently, revealing a crab claw molecule (Hirata, *et al.*, 2008; Korkhin, *et al.*, 2009; Kusser, *et al.*, 2008)(figure 2). Structural elements are rather similar in the archaeal and eukaryotic transcription machinery, while eukaryotic polymerase is more complicated for advanced functions. Subunits A' and A'' represent the bacteria β' and eukaryotic Rpb1 and subunit B represents bacteria β and eukaryotic Rpb2, all contributing to form the catalytic center. Subunit A'' doesn't have the Pol II CTD domain. Subunits D/L, homolog of eukaryotic Rpb3/Rpb11 forms a heterodimer, similar to the α_2 homodimer counterpart in bacteria. Subunit D has a 4Fe-4S cluster-binding domain which is unique for some archaea species as well as eukaryotic Pol I and III, and is indicated to support the D-subunit folding. Subunit E'/F form a highly mobile heterodimer which is homolog of eukaryotic Rpb4/7 heterodimer and its structure was solved separately (Todone, *et al.*, 2001). Archaeal E'/F dimer associates with the core enzyme tightly, and induces a closed conformational clamp in the solved structures (Armache, *et al.*, 2003; Edwards, *et al.*, 1991; Grohmann, *et al.*, 2009). *In vitro* studies using reconstituted archaeal RNAP indicated that subunit E' was required for transcription at low temperatures and stimulates open complex formation (Naji, *et al.*, 2007). Subunit H is homolog of Rpb5, but lacking the N-terminal jaw domain. Subunit K, homolog of Rpb6 also lacks the unconserved disordered N-terminal domain. RpoG which is the homolog of Rpb8 was also found in the newest structure. RpoG was thought to be absent in archaeal RNAP for a long time since it's lost during purification of polymerase (Hirata, *et al.*, 2008; Werner, 2007). Just recently, RpoG was

demonstrated to be unique in *Korarchaea* and *Crenarchaea* (Koonin, *et al.*, 2007; Kwapisz, *et al.*, 2008). It has an OB fold and positions peripherally like Rpb8. Subunits N and P together with D/L heterodimer form an extended platform that is required for efficient assembly of RNAP. Strikingly, a novel subunit Rpo13 was reported, which does not have any counterpart in eukarya, but architecturally corresponds to an insertion in the bacterial β' subunit. Rpo13 was suggested to facilitate the open bubble formation in the initiation stage. In general, archaeal RNA polymerase is like a truncated version of RNA pol II, with differences only in the periphery subunits. Simple additions of pol II-specific subunits like Rpb9 or domains such as CTD or the Rpb5 jaw, to the archaeal polymerase can lead to RNAP II that does not need any change in the core enzyme.

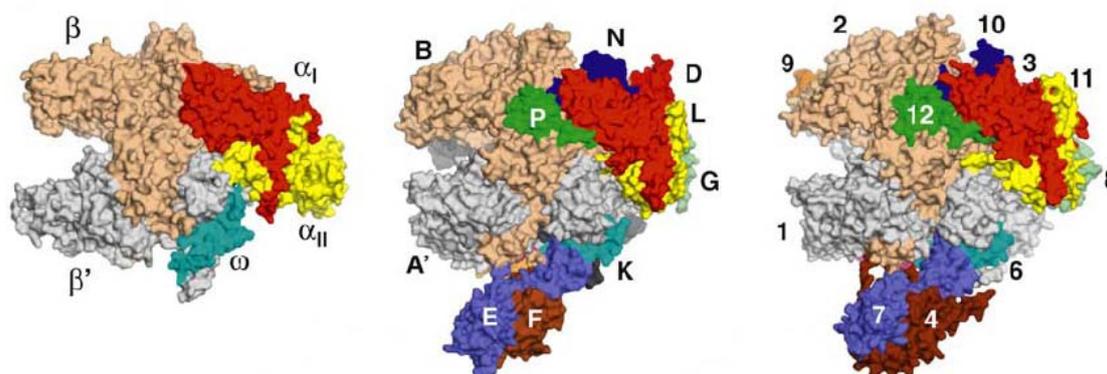


Figure 2. RNAP structures from bacteria (left, *T.aquaticus* core enzyme), archaea (center, *S.Solfataricus* RNAP) and eukarya (right, *S.cerevisiae* Pol II). Each subunit is denoted by a unique color and labeled. Orthologous subunits are depicted with the same color. Adapted from (Hirata and Murakami, 2009).

Chapter II: Evolution of two modes of intrinsic RNA polymerase transcript cleavage

1. Introduction

1.1 Chemical mechanism of transcriptional cleavage

RNA polymerases possess a conserved active center with two Mg^{2+} ions (Cramer, *et al.*, 2001; Vassylyev, *et al.*, 2002). One ion is persistently bound, while the second is not and can be recruited and stabilized *ad hoc* for different kinds of catalysis (Cramer, *et al.*, 2001; Wang, *et al.*, 2006; Westover, *et al.*, 2004). Three Asp residues coordinate with Mg I, and partly interact with Mg II (Sosunov, *et al.*, 2005). Both the nucleotide incorporation and transcript cleavage were suggested to follow a unified two-metal-ion mechanism (figure 3) (Kettenberger, *et al.*, 2003; Sosunov, *et al.*, 2003; Sosunova, *et al.*, 2003; Steitz, 1998; Wang, *et al.*, 2006; Zenkin, *et al.*, 2006). Two Mg ions are involved to stabilize the pentacovalent transition state. The single, tunable active site of RNA polymerase operates various modes. When RNAP functions as a polymerase catalyzing RNA synthesis, Mg II is stabilized by the β and γ phosphates in nucleoside triphosphate (figure 3A). Transcript 3' terminal hydroxyl group attacks α -phosphorus atom of the NTP, following substitution nucleophilic bimolecular (S_N-2) mechanism and forms phosphodiester bond. Intrinsic RNAP cleavage could resemble the DNA cleavage by the Klenow DNA polymerase (Beese and Steitz, 1991). A water molecule attacks the phosphorous atom in the scissile phosphodiester bond. A backtracked nucleotide can also stabilize and orient Mg II and the active water molecule, a mechanism termed transcript-assisted cleavage (Zenkin, *et al.*, 2006). A non-complementary NTP is also indicated to be able to coordinate Mg II to stimulate RNA cleavage, named as substrate-assisted cleavage, and the stimulatory effect was indeed observed in some RNA polymerases (Hagler and Shuman, 1993; Sosunov, *et al.*, 2003; Westover, *et al.*, 2004; Zenkin, *et al.*, 2006). When an external transcription cleavage factor such as Gre or TFIIS

participates, it can also coordinate Mg II and orient the active water molecule and then stabilize the penta-covalent transition state (figure 3B).

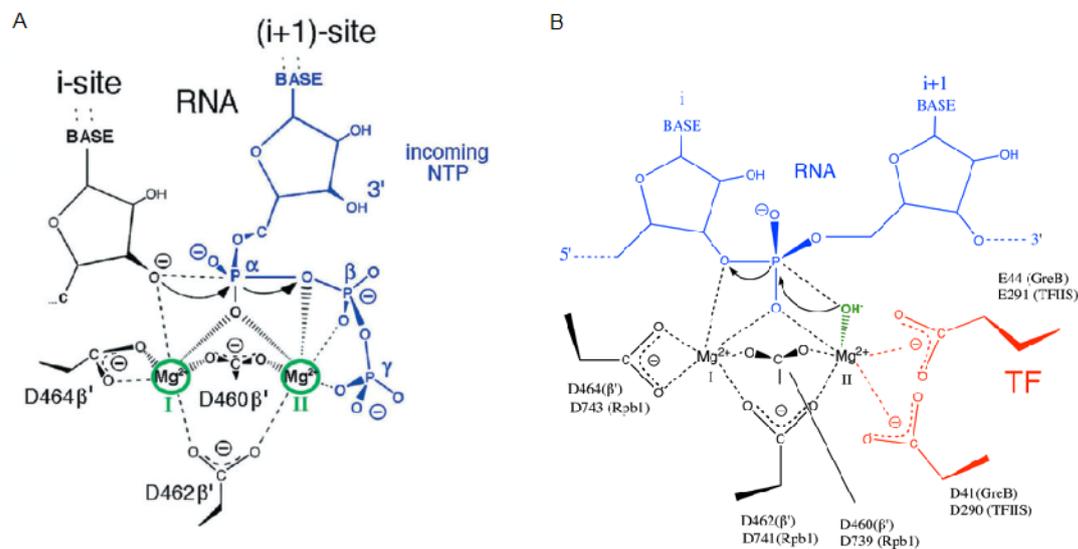


Figure 3. Schematic diagrams of polymerase active center during (A) polymerization reaction, (B) factor assisted transcription cleavage reaction. Adapted from (Poole and Logan, 2005; Sosunov, *et al.*, 2005)

1.2 Weak intrinsic cleavage activity of RNA polymerases

Intrinsic cleavage activity in ternary elongation complexes was first observed in *E. coli* RNA polymerase independently of external factors (Orlova, *et al.*, 1995; Surratt, *et al.*, 1991). Soon after, it was found to be a common feature in other polymerases, including T7 polymerase, vaccinia virus polymerase and Pol II (Hagler and Shuman, 1993; Izban and Luse, 1992; Reines, 1992; Sastry and Ross, 1997; Wang and Hawley, 1993; Weilbaecher, *et al.*, 2003). Intrinsic RNA cleavage activity, which is at a low level without stimulatory factors at physiological pH, requires divalent metal ions, and cleaved transcript is able to be elongated if NTPs are added. α -amanitin, a specific inhibitor of Pol II, impairs intrinsic cleavage and TFIIS stimulated cleavage to different extents, reducing the rate of intrinsic cleavage, but completely abolishing TFIIS stimulated cleavage (Izban and Luse, 1992; Rudd and Luse, 1996; Weilbaecher, *et al.*, 2003). Additionally, Sarkosyl which can remove dissociable factors fails to abolish intrinsic cleavage (Wang and Hawley, 1993). Therefore, intrinsic cleavage resides in RNA polymerase, and is likely to follow

a different mechanism to factor-stimulated cleavage. The cleavage activity is suggested to be carried out by the polymerization site itself (Rudd, *et al.*, 1994). Remarkably, alkaline pH substantially stimulates intrinsic cleavage, as proved in bacterial RNAP and Pol II (Awrey, *et al.*, 1997; Orlova, *et al.*, 1995; Weillbaecher, *et al.*, 2003). Based on the chemistry mechanism presumed before, increased deprotonation of active water molecules by alkaline pH could be the reason for cleavage stimulation (Sosunov, *et al.*, 2003).

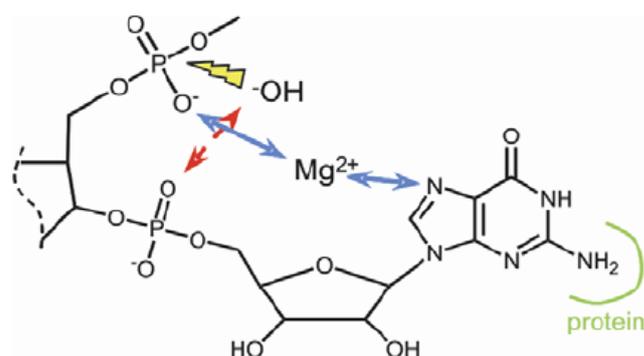


Figure 4. Role of the nascent transcript 3'-terminal nucleotide in cleavage. Adapted from (Zenkin, *et al.*, 2006)

Mechanism of intrinsic RNA cleavage was proposed recently by structural and biochemical data (Wang, *et al.*, 2009; Yuzenkova and Zenkin, 2010; Zenkin, *et al.*, 2006). They argue that RNA polymerase is able to backtrack by one nucleotide, an action especially favored after misincorporation, and the mismatched nucleotide is bound to a stable backtracked site. One non-esterified oxygen in the ultimate phosphodiester bond orients the active water by a hydrogen bond and a nitrogen in the base of backtracked nucleotide coordinates with Mg II. Cleavage occurs on the penultimate phosphodiester bond and produces a dinucleotide (figure 4). A flexible domain, the trigger loop, is proved to be required for intrinsic cleavage in RNA polymerase from *Thermus aquaticus* (Yuzenkova and Zenkin, 2010).

1.3 Extrinsic factors induce transcriptional cleavage

1.3.1 Cleavage factors in bacteria

Intrinsic cleavage activity in RNA polymerase is dramatically stimulated by

exogenous stimulation factors. In bacteria, two factors Gre A and Gre B stimulate cleavage activity (Borukhov, *et al.*, 1992; Borukhov, *et al.*, 1993). The Gre family contain homologous proteins with a molecular weight of 19 kDa. Gre-homologs exist ubiquitously in over 60 organisms, including *Mycoplasma genitalium* which has the smallest known genome, indicating the importance of their biological function (Hutchison, *et al.*, 1999). Deletion of either the greA or the greB gene had no effect on *E.coli* cell phenotype, but the double deletion strain is temperature sensitive (Orlova, *et al.*, 1995). Crystal structure and CD spectra of *E.coli* GreA and GreB, and homology modeling of GreB show that both factors comprise a N-terminal anti-parallel α -helical coiled-coil domain linked to a globular C-terminal domain by a short loop (Koulich, *et al.*, 1997; Stebbins, *et al.*, 1995; Vassylyeva, *et al.*, 2007) (figure 5A). Although highly related in structure, they stimulate cleavage in different ways: GreA induces cleavage of mostly di- and tri- nucleotides and can only prevent transcriptional arrest while GreB induces cleavage of fragments of various lengths from 2 to 18 nucleotides and is able to rescue arrested polymerases (Borukhov, *et al.*, 1993; Feng, *et al.*, 1994). Their distinct functions derive from the difference of an essential basic patch on the surface of the coiled-coil domain (figure 5B) (Koulich, *et al.*, 1997; Kulish, *et al.*, 2000). Additionally, GreB binds polymerase with an affinity of about two orders of magnitude higher than GreA.

The C-terminal domain of Gre factor contains an α -helix and a four/five-strand β -sheet, forming an open hydrophobic cavity (Vassylyeva, *et al.*, 2007). It doesn't directly stimulate cleavage but participates in binding to the RNA polymerase and is required for full stimulatory activity (Koulich, *et al.*, 1998; Koulich, *et al.*, 1997; Polyakov, *et al.*, 1998). The C-terminal domain binds RNA polymerase near the secondary channel, while its particular binding site was proposed by conflicting models with opposite orientations (Laptenko, *et al.*, 2003; Opalka, *et al.*, 2003; Polyakov, *et al.*, 1998; Sosunova, *et al.*, 2003; Vassylyeva, *et al.*, 2007).

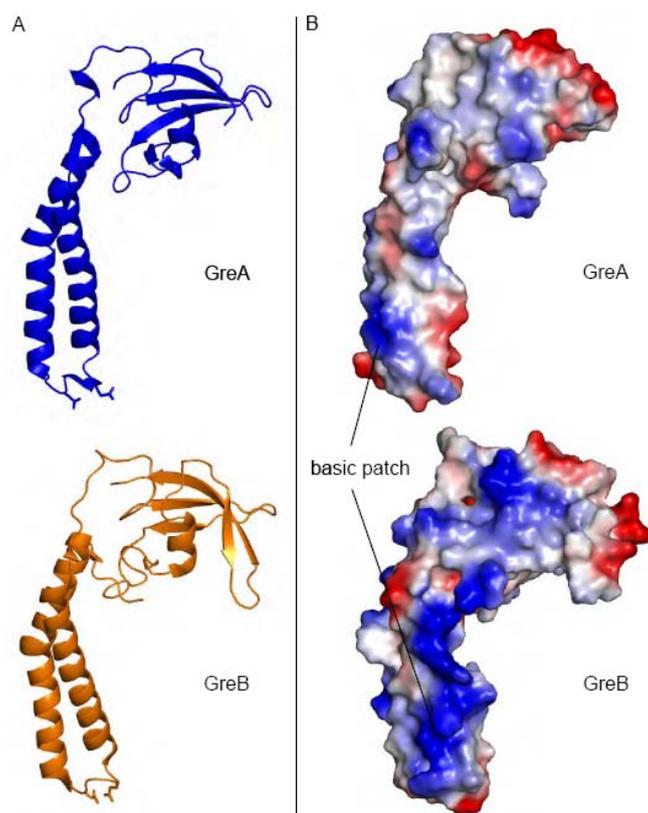


Figure 5. A. Structure of *E.coli* GreA and GreB. Two acidic side chains are shown in blue and orange for GreA and B respectively. B. Charge distribution of GreA and GreB. The surface is colored by the electrostatic potential. White, uncharged; red, negative; blue, positive. Basic patch is shown.

The N-terminal domain can not bind polymerase on its own, but is the domain that is responsible for inducing cleavage, and can stimulate cleavage when added to RNA polymerase in saturating amounts (Koulich, *et al.*, 1998; Polyakov, *et al.*, 1998). This domain is also responsible for transcriptional readthrough (Koulich, *et al.*, 1998; Koulich, *et al.*, 1997). Two elements on the N-terminal domain are crucial for their function. Crosslinking results limits the interaction region of Gre and RNA to the end tip of the coiled-coil domain (Koulich, *et al.*, 1997; Stebbins, *et al.*, 1995). Mutagenesis and crosslinking results demonstrate that the whole coiled-coil domain inserts to the secondary channel and two conserved acidic residues on the tip can be put into the active center and position the Mg II as well as the active water (Laptenko, *et al.*, 2003; Sosunova, *et al.*, 2003) (figure 3B, 5A). Another important element is so-called basic patch. GreA has just two Arg residues making up a short basic patch of

approximately 7Å long, while GreB has an extended basic region across the whole surface of the protein of about 35Å long containing several basic residues (figure 5B). The basic patch was proposed to anchor negatively charged nascent transcripts and thus determine the length of RNA to be cleaved as a “molecular ruler” (Kulish, *et al.*, 2000).

1.3.2 Cleavage factors in eukarya

Cleavage factor TFIIIS stimulates transcript cleavage in Pol II (Izban and Luse, 1992; Reines, 1992; Rudd, *et al.*, 1994). It can also stimulate cleavage on binary complex composed of RNAP and RNA alone (Johnson and Chamberlin, 1994). In yeast, only one TFIIIS gene has been identified, but in vertebrates like humans, multiple genes were identified including a more widely expressed form and some tissue specific forms, and all the isoforms can stimulate transcript cleavage *in vitro* (Labhart and Morgan, 1998; Plant, *et al.*, 1996; Williams and Kane, 1996). TFIIIS is not essential for cell viability in yeast, but the deletion mutant shows sensitivity to oxidants like menadione and drugs like 6-azauracil (Koyama, *et al.*, 2003; Koyama, *et al.*, 2007; Nakanishi, *et al.*, 1995). Interestingly, TFIIIS was also detected as a RNAP III transcription factor, indicating a general contribution of this protein (Ghavi-Helm, *et al.*, 2008). TFIIIS induces cleavage in two different ways. Stalled ternary complex generates primarily dinucleotides with the assistance of TFIIIS while in an arrested complex, oligonucleotides of up to 17 nt can be released (Izban and Luse, 1992; Izban and Luse, 1993; Izban and Luse, 1993). TFIIIS has a molecular weight of 35 kDa. Limited proteolysis revealed that TFIIIS was composed of three domains (Morin, *et al.*, 1996). NMR or X-ray structures were solved for all three domains, separately and for domain II and III combined, from yeast to human (Booth, *et al.*, 2000; Kettenberger, *et al.*, 2003; Morin, *et al.*, 1996; Qian, Gozani, *et al.*, 1993) (figure 6).

Domain I covers the N-terminal residues 1-130 which are variable in TFIIIS homologs (Labhart and Morgan, 1998). It contains sequence homologous to

another two elongation factors, elongin A and CRSP70 (Booth, *et al.*, 2000). The function of domain I is poorly understood as it is dispensable for the cleavage stimulatory function of TFIIS, but required for efficient interaction with RNAP holoenzyme which includes several general initiation factors and promotes active preinitiation complex formation (Kim, *et al.*, 2007; Pan, *et al.*, 1997). Domain I can be phosphorylated, and this TFIIS form can't stimulate polymerase activity (Hirai, *et al.*, 1988; Horikoshi, *et al.*, 1985), indicating that this domain also participates on the regulation of TFIIS activity by (de)phosphorylation. A NMR structure shows a four helix bundle structure (figure 6). A basic patch was also found on the top of the helix bundle and the face formed by helix 1 and 3.

TFIIS domains II and III and the linker between them are fully sufficient for binding to polymerase (Awrey, *et al.*, 1998). NMR and X-ray structure reveal that it contains a stably folded three-helix bundle and some helical secondary structure which can only be seen upon binding to Pol II (Kettenberger, *et al.*, 2003; Morin, *et al.*, 1996; Olmsted, *et al.*, 1998) (figure 6). Several positively charged residues built up a basic patch on the third helix and the loop after. Mutations on this domain, especially on the basic patch, severely reduce Pol II binding without interfering with the stimulatory activity (Awrey, *et al.*, 1998; Cipres-Palacin and Kane, 1995). The structure of Pol II-TFIIS shows that helix 1 and 3 pack against the Rpb1 jaw domain, and the basic patch interacts with two acidic loops (Cheung and Cramer, 2010; Kettenberger, *et al.*, 2003; Wu, *et al.*, 1996). The linker domain is flexible in free TFIIS, but forms a helix upon polymerase binding (figure 6). After the helix, the linker passes through a narrow crevice into the secondary channel, and crevice opening is induced by the linker binding. Mutations, five-residue deletion or insertion in the linker affect TFIIS activity indicating the importance of the residues and spacing (Awrey, *et al.*, 1998). Moreover, the linker domain determines species-specificity probably through orienting the domain II and III (Shimasaki and Kane, 2000).

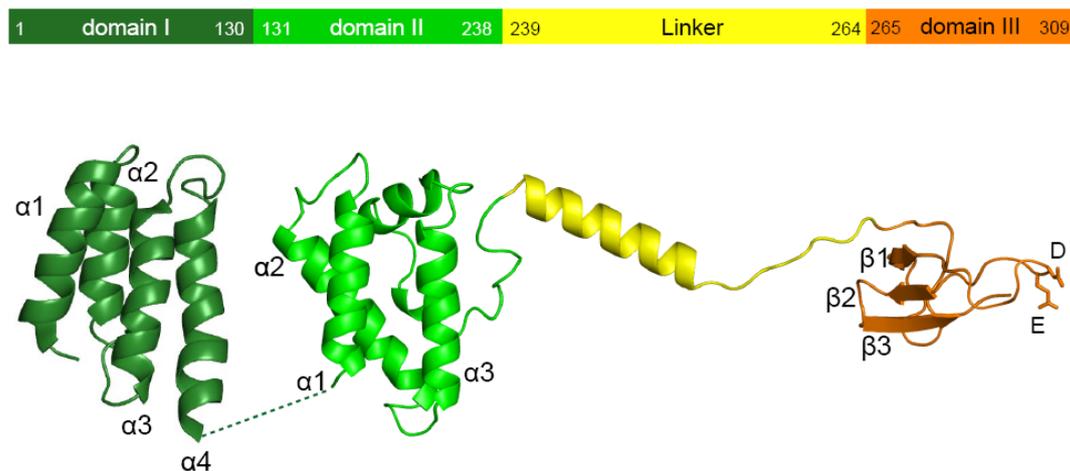


Figure 6. Domain organization and structure of TFIIIS. Secondary structure elements are shown. TFIIIS domain I, II, linker, III are colored in dark green, light green, yellow and orange, respectively. Two principal acidic side chains are shown as sticks.

Domain III (zinc ribbon) is the cleavage-stimulatory domain, but cannot induce cleavage separately even when added to Pol II in saturating amounts (Awrey, *et al.*, 1998). It is highly conserved and functionally exchangeable between TFIIIS orthologs *in vitro* and *in vivo* (Shimasaki and Kane, 2000). It contains three antiparallel β -sheets stabilized by four cysteines chelating a zinc ion (figure 6) (Olmsted, *et al.*, 1998; Qian, Jeon, *et al.*, 1993). It binds polymerase through many hydrophobic contacts and salt bridges. An acidic tip in the β -hairpin reaches pol II active center. Two conserved residues, Asp and Glu, are essential for activity. Mutations or exchange completely abolish TFIIIS function (Jeon, *et al.*, 1994).

In the pol II-TFIIIS complex, TFIIIS binds to the Rpb1 jaw through domain II and linker helix, then extends into the funnel, inserts domain III into the pore, and positions the tip into the active center (Kettenberger, *et al.*, 2003). The mechanism of cleavage stimulation by TFIIIS is proposed to be analogous to that by Gre factors (Cramer, 2004), indicating that although sequence and structure of these factors diverged during evolution, their functions were maintained.

1.3.3 Effects of cleavage factors to transcription

Cleavage factors induced cleavage has multiple biological roles in transcription. A well studied function is to rescue arrested complexes, allowing readthrough and productive elongation. RNAP frequently pauses on DNA templates. Blocks can be from primary DNA sequences (e.g. A/T rich), DNA lesions (e.g. those caused by oxidative damage) or from DNA-binding proteins (e.g. histones, “roadblock”) (Izban and Luse, 1993; Kireeva, *et al.*, 2005). Long lifetime paused complexes falls into a state (“arrested state”) that is unable to continue elongation although the enzyme is intact and NTPs are supplied (Fish and Kane, 2002; Wind and Reines, 2000). The arrested state is severely disruptive to gene expression and can cause cell death. In the presence of cleavage factors, the nascent transcript is cleaved generating a new 3' end at the active site of RNA polymerase, and can continue to be elongated. However, in pol II, cleavage itself is not enough to allow readthrough (Cipres-Palacin and Kane, 1994). Thus, besides its ability to induce cleavage, TFIIIS is suggested to also induce conformational changes in polymerase that allow readthrough. A second proposed role is to increase fidelity (Erie, *et al.*, 1993; Jeon and Agarwal, 1996; Koyama, *et al.*, 2003; Koyama, *et al.*, 2007; Thomas, *et al.*, 1998). This role is apparently achieved by the ability of TFIIIS to stimulate excision of misincorporated nucleotides. A third proposed role is to facilitate the transition from initiation to elongation by rescuing arrested polymerase and suppressing early RNA release (Adelman, *et al.*, 2005; Hsu, *et al.*, 1995; Malagon, *et al.*, 2004).

1.4 Eukaryotic RNA polymerase with strong intrinsic cleavage activity

1.4.1 RNA polymerase I

RNA Pol I was initially found to have RNase H cleavage activity with “broad base specificity” degrading RNA from RNA-DNA hybrids as mono- and dinucleotides (Huet, *et al.*, 1976). The activity associates with two subunits A49 and A34.5 which dissociate from Pol I in the presence of urea (Huet, *et al.*,

1977). Later on, RNase H activity was proposed to reside in dissociable factors which can be removed by high concentrations of Sarkosyl (Iborra, *et al.*, 1979; Labhart, 1997; Tschochner, 1996). Kuhn *et al.* clarified recently that, Pol I indeed possesses intrinsic cleavage (Kuhn, *et al.*, 2007). And the cleavage activity is based on the A12.2 subunit, in particular the C-terminus, whose full activity requires a heterodimer made by subunits A49/A34.5, at least, the A49/A34.5 dimerization module together with either A49 linker or A34.5 tail are required (Geiger, *et al.*, 2010).

A12.2 has the sequence homologue to a factor TFS of archaeal RNA polymerase, subunit Rpb9 of polymerase II and to subunit C11 of Pol III, containing two potential zinc binding motifs: $CX_2CX_nCX_2C$. The sequence of the C-terminal zinc binding domain has an identity of 40% to TFIS zinc ribbon containing an invariant motif Q.RSADE..T.F. Indeed, A12.2 was found to strongly bind zinc through a radioactive zinc binding technique (Treich, *et al.*, 1991). Deletion of A12.2 makes cells heat sensitive (Nogi, *et al.*, 1993). Although the C-terminus is more conserved, deletion of this part has no effect on the cell growth in elevated temperatures and no sensitivity to drugs like 6-azauracil or mycophenolate, neither affects the interaction with the second largest subunit A135. In contrast, these functions require the N-terminal part, which is poorly conserved (Gadal, *et al.*, 1997; Van Mullem, Landrieux, *et al.*, 2002). This indicates that N-terminal A12.2 is required for stability and the conformational change in polymerase which is enough for cell viability while C-terminal A12.2 specifically induces cleavage. Beside its functions in stabilization/cleavage, A12.2 is also important for transcription termination (Prescott, *et al.*, 2004). Deletion of A12.2 resulted in a significant read-through of the terminator to the spacer sequence. Actually, a 3' trimming processing event happens immediately after termination, cleaving 10 nt from pre-rRNA in a stretch of uridines (Kuhn and Grummt, 1989).

Localization of A12.2 is suggested by low resolution EM structures and immunolabelling (Bischler, *et al.*, 2002; Chedin, *et al.*, 1998; De Carlo, *et al.*,

2003; Kuhn, *et al.*, 2007). A12.2 is indicated to localize the same as subunit Rpb9 of Pol II which has a very low intrinsic cleavage activity. Since this position is over 30Å towards the active center (figure 7), it is difficult to understand how A12.2 induces high cleavage from there.

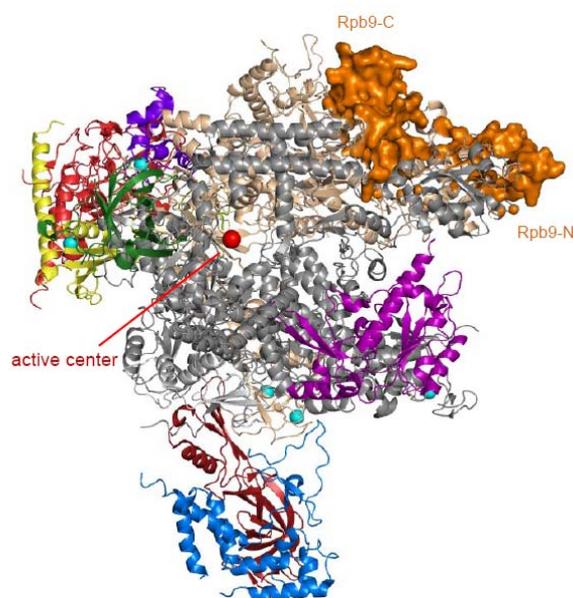


Figure 7. Structure of RNA polymerase II and localization of Rpb9 domains. PDB code is 1WCM. Active center is shown with a red Mg ion. RNA pol II is shown in bottom view.

1.4.2 RNA polymerase III

Hydrolytic activity is proved to be universal for RNA polymerases. RNA Pol III stalled on a SUP4 tRNA^{Tyr} gene template is also able to cleave nascent RNA from the 3' end and this is factor-independent (Whitehall, *et al.*, 1994). Similarly, Pol III cleavage products appear predominantly as di-nucleotide (and some mono- or tri-nucleotides also appear dependant on the sequence) from 3' end of nascent RNA, and long cleavage products were not found for RNA polymerase III (Bobkova and Hall, 1997; Whitehall, *et al.*, 1994). Cleavage positions were determined by the RNA sequence, especially by the secondary structure of the RNA-DNA heteroduplex. Pol III is prone to cleave 5' to an internal uridine positions where a weak rU:dA hybrid is present (Martin and Tinoco, 1980). Divalent cations are also required to allow cleavage to occur. Pol III has similar cleavage activity characteristics with factor-independent vaccinia virus RNA polymerase in which subunit rpo30 was proposed to induce

cleavage (Hagler and Shuman, 1993). First, non-cognate NTPs can stimulate the rate of cleavage in polymerase III. Second, Sarkosyl which is able to separate dissociable factors has no effect on cleavage activity by RNA polymerase III, indicating that it is not dependent on exogenous factors.

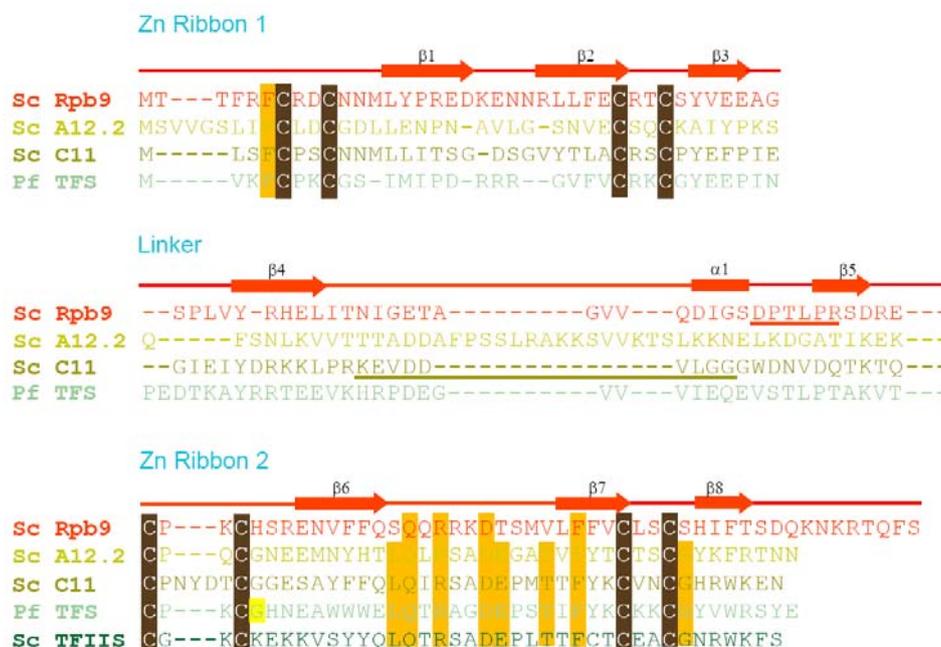


Figure 8. Sequence alignment of *S.cerevisiae* Rpb9, A12.2, C11, TFIIS-C-terminal zinc ribbon, *P.furiosus* TFS. Cysteines of the zinc-binding domains are boxed in brown. Amino acids identical in sequence of at least A12.2, C11, TFS and TFIIS are boxed in orange. Archaeal TFS specific Gly is boxed in yellow. Polymerase specific conserved sequences in the linker domains are underlined in red and dark yellow for Rpb9 and C11 respectively. Secondary structures of Rpb9 are shown in red.

An endogenous Pol III subunit encoding a small size protein (~110 residues, 11 kDa) in yeast was found to be responsible for transcript cleavage, and was particularly efficient in cleaving misincorporated nucleotides (Alic, *et al.*, 2007; Chedin, *et al.*, 1998; Landrieux, *et al.*, 2006). Orthologs in *S.pombe*, human and zebrafish were also reported (Chedin, *et al.*, 1998; Huang, *et al.*, 2005; Yee, *et al.*, 2007). Sequence alignment indicates that C11 is homologous to Rpb9 and A12.2, containing two potential zinc binding domains, separated by a nonconserved linker. The C-terminal zinc-binding domain is most conserved, sharing especially high similarity with that of TFIIS (~67%) (figure 8). Similarly, C11 also has two invariant residues Asp and Glu in the C-terminal zinc-binding

domain and they are essential for cleavage activity (Alic, *et al.*, 2007; Landrieux, *et al.*, 2006). Additionally, disruption of the potential C-terminal zinc binding domain also decreases cleavage activity indicating that active C11 requires an intact structure (Huang, *et al.*, 2005). The linker region is conserved only among C11 orthologs, containing a stretch of residues which were suggested to be necessary for the assembly of C11 to Pol III (underlined in figure 8). However, C11 (and especially the invariant acidic residues DE) is essential for cell viability (Chedin, *et al.*, 1998), whereas A12.2 and Rpb9 which are not essential under normal growth conditions (Nogi, *et al.*, 1993; Woychik, *et al.*, 1991). Besides its cleavage activity, C11 is also important for termination and reinitiation (Huang, *et al.*, 2005; Landrieux, *et al.*, 2006; Yee, *et al.*, 2007).

Interestingly, some mutations in the two largest subunits in Pol III, C160 and C128 were shown to be able to increase or decrease cleavage activity of pol III (Bobkova, *et al.*, 1999; Thuillier, *et al.*, 1996). These results, together with observation of a faraway position of its homolog Rpb9 from the active center in Pol II, led to the view that C11 is involved indirectly for the cleavage activity which was embodied in polymerase itself (Geiduschek and Kassavetis, 2001; Walmacq, *et al.*, 2009). However, none of these mutants are devoid of cleavage activity. Since they locate either in the vicinity of C11, or in the mobile domains which contact RNA-DNA duplex or the downstream DNA, it is possible that those mutations affect cleavage activity indirectly by affecting the movement of DNA or RNA-DNA duplex, regulating the pausing time during which cleavage can happen, or affecting C11 activity.

There are very few investigations of C11 so far. Therefore, the mechanism of cleavage activity and localization of C11 remain unclear. Because of the conserved C-terminal sequence with TFIIIS, C11 was assumed to interact with polymerase in the same manner, by insertion of the C-terminal domain to the pore and direct stimulation, although no evidence was given. Crosslinking between C11 and RNA also was unsuccessful (Kassavetis, *et al.*, 2010). Yeast

two-hybrid screening proposed that C11 interacts with early N-terminus of C128 “protrusion” region, so does its homolog A12.2 in Pol I (Flores, *et al.*, 1999; Van Mullem, Landrieux, *et al.*, 2002). EM structures of Pol III indicated that the N-terminal domain of the C11 is positioned similarly to its homolog Rpb9 in Pol II, but the C-terminal is missing (Fernandez-Tornero, *et al.*, 2010; Fernandez-Tornero, *et al.*, 2007). The EM structure of Pol I reveals that the C-terminal domain of A12.2, which is the functional and sequence homolog of C11, is positioned at the same place as the Rpb9 C-terminus (Kuhn, *et al.*, 2007) (figure 7). These results indicate a fuzzy localization speculation of C11 which will lead to significantly different mechanism. Due to the lack of high resolution atomic structure of complete Pol III, it is still an argumentative question.

1.5 Cleavage activity in archaeal RNA polymerase

As mentioned in Chapter I, archaea are more similar to eukarya than bacteria in transcription and translation machinery and are thought to be closer to the common ancestor. Since cleavage activity is essential for cells, archaeal RNAP should be expected to have cleavage activity as well (Sigurdsson, *et al.*, 2010). An ORF immediately after the gene coding for L subunit in the same gene cluster in archaea was found to code a small protein homologous to the A12.2/Rpb9 subunit of eukaryotic RNAPs (Langer, *et al.*, 1995; Langer and Zillig, 1993). Sequence analysis found that this protein shares similarity with eukaryotic subunits even more than the eukaryotic subunits do with each other. In other words, eukaryotic subunits diverged more from each other than from the archaeal homolog (Kaine, *et al.*, 1994). Since it is also homologous to eukaryotic TFIIS, this protein was named as subunit M or TFS. Later on, this protein was indeed proved to induce di-nucleotide cleavage in archaeal RNAP (Hausner, *et al.*, 2000). TFS is also able to induce cleavage of misincorporated nucleotides and reduce misincorporation, and can therefore improve transcription fidelity (Lange and Hausner, 2004). Although TFS shows high

similarity with Rpb9/C11/A12.2, it doesn't bind or binds weakly to archaeal pol, indicating that TFS is not an intrinsic subunit (Hausner, *et al.*, 2000). Thus in the X-ray structures of archaeal RNAP, TFS is not included. Sequence alignment shows that TFS possesses two potential zinc binding domains as its eukaryotic homologs (figure 8). Although a NMR structure of the C-terminal domain was solved, structure of the whole protein is still lacking (Wang, *et al.*, 1998). The mechanism of TFS induced cleavage is also not understood.

1.6 Aims and objectives of this thesis

Although the active centers of eukaryotic RNA Pol I, II, and III are conserved, the strength of their intrinsic cleavage activity greatly differs with each other. Whereas the cleavage activity is very strong for Pol I (Kuhn, *et al.*, 2007) and Pol III (Alic, *et al.*, 2007; Thuillier, *et al.*, 1996), it is very weak for Pol II, which needs stimulation by TFIIIS, containing a Pol II-binding domain and a C-terminal Zn-ribbon (hereafter referred to as C-ribbon). Because of the lacking of high resolution structural information of Pol I and Pol III, the molecular basis for this phenomenon remains unknown. Intrinsic cleavage requires the Rpb9 homologous subunits TFS, A12.2, and C11 in archaeal pol, eukaryotic Pol I, and Pol III, respectively (Chedin, *et al.*, 1998; Hausner, *et al.*, 2000; Kuhn, *et al.*, 2007). Rpb9 resides on the Pol II surface, where its N-ribbon forms part of the Rpb1/9 jaw and its C-ribbon binds on between the Rpb1 and Rpb2 domains, distinct from the active center. The C-ribbon of TFIIIS is also homolog of Rpb9/A12.2/C11. However, the C-ribbon of TFIIIS binds the pore and reaches the active site with a hairpin containing the invariant residues D290 and E291 that are required for function. Whereas A12.2 and C11 contain these hairpin residues, Rpb9 lacks the residue corresponding to E291. The aim of this work was to study the molecular basis for the different intrinsic RNA cleavage activities between Pol I, II, III and answer the question how the C-ribbon domains are related evolutionarily and mechanistically, and how this may result in different cleavage activities. This objective was achieved

by using a combination of mutagenesis, cleavage assays, and X-ray crystallography techniques. The results unravel the molecular basis for differential intrinsic RNA cleavage activities of Pol II and Pol III, and suggest how the C-ribbon domains are related evolutionarily and how different cleavage activities arose during evolution.

2. Results

2.1 A Pol II variant with strong intrinsic RNA cleavage.

2.1.1 Rpb9 is required for weak intrinsic cleavage activity.

Previous studies indicate the importance of Rpb9 for the fidelity of Pol II transcription *in vivo* (Nesser, *et al.*, 2006). Before nucleotide incorporation, the incoming NTP was discriminated by polymerase and only the correct NTP is then sequestered by a mobile element called trigger loop (TL) after which a phosphoryl group transfers and a phosphodiester bond forms (Sydow and Cramer, 2009). Pre-incorporation fidelity requires Rpb9, and was suggested to do so by delaying TL closure (Walmacq, *et al.*, 2009). Post-incorporation fidelity in pol II requires both Rpb9 and TFIIS (Awrey, *et al.*, 1997; Hemming and Edwards, 2000; Koyama, *et al.*, 2007). Rpb9 has never been demonstrated to contribute to the Pol II intrinsic cleavage activity. Actually, *in vitro* experiments propose that Rpb9 is dispensable for intrinsic cleavage activity (Awrey, *et al.*, 1997; Weilbaecher, *et al.*, 2003) by using assays done under elevated pH and suggest that Rpb9 just transfers the signal from TFIIS to polymerase. However, elevated pH can increase the amount of active water which attacks phosphodiester bond and thus would provide an artificial and inaccurate conclusion. I aimed to investigate the basis for intrinsic RNA cleavage *in vitro* by using Pol II lacking Rpb9 (Pol II Δ 9), and complementing the Pol II Δ 9 enzyme with Rpb9 variants, and investigating the resulting Pol II variants for their cleavage activity. Pol II Δ 9 was prepared from a yeast strain lacking the *rpb9* gene (Janke, *et al.*, 2004). As expected, Pol II Δ 9 was inactive in cleaving the RNA 3'-end in reconstituted elongation complexes with a 3'-RNA-DNA G-G mismatch (figure 9, 10, 11) (Kuhn, *et al.*, 2007; Sydow, *et al.*, 2009), whereas addition of Rpb9 led to mild cleavage stimulation (figure 11A lanes 2-3). Since the cleavage reaction I used is under physiological pH, this *in*

in vitro result proves that Rpb9 indeed contributes for the weak intrinsic cleavage activity of Pol II.

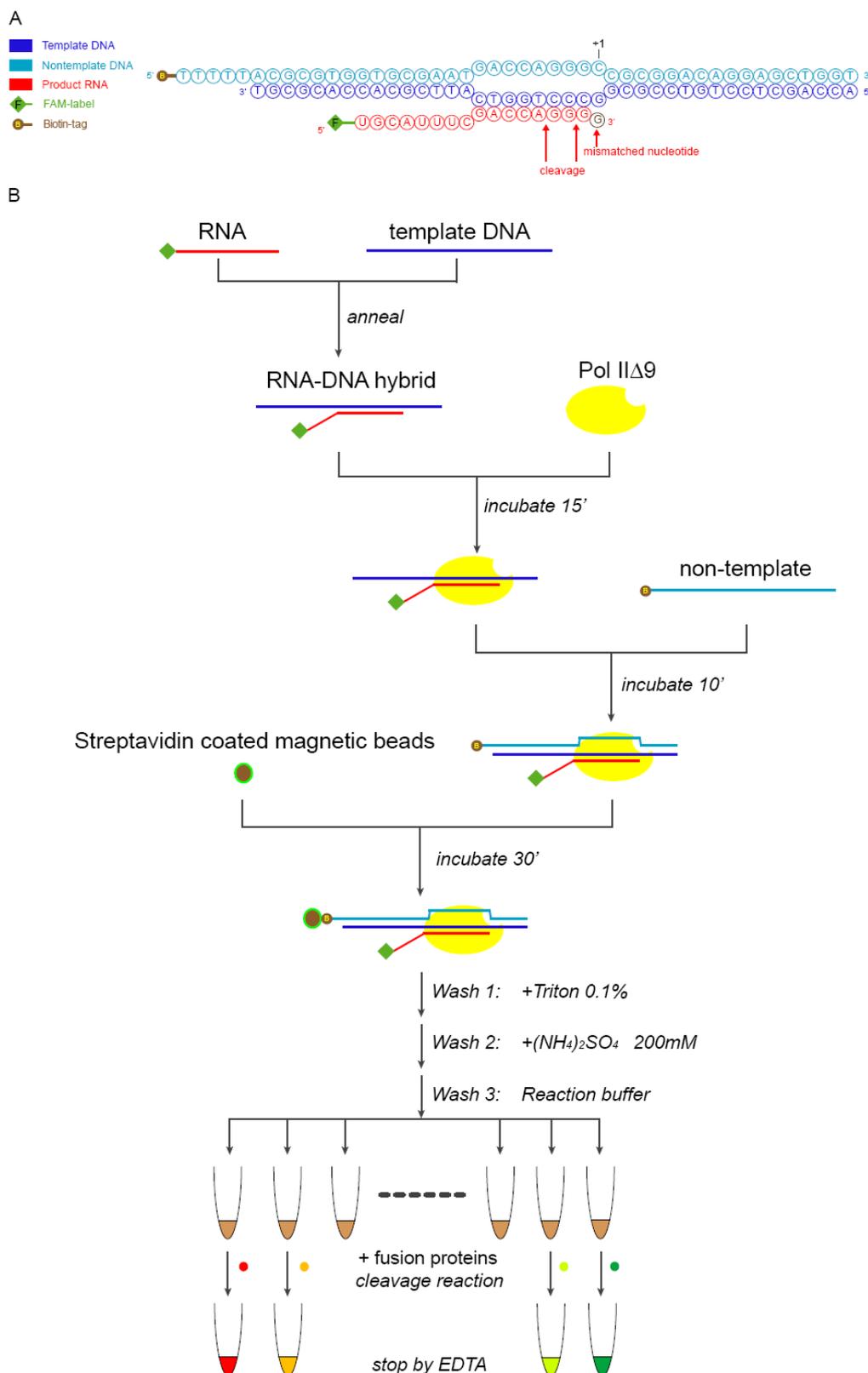


Figure 9. Nucleic acid scaffold used for the assays(A) and experiment strategy(B). Non-template and template DNA are shown in cyan and blue, respectively, RNA is in red. A mismatched nucleotide at the RNA 3'-end is shown in brown. Red, orange, light green and dark green spheres represent different Rpb9 constructs.

2.1.2 A Pol II variant with strong intrinsic RNA cleavage.

To explore whether the pol III subunit C11 may replace Rpb9 function in pol II, I prepared a Rpb9-C11 fusion protein that contains the Rpb9 N-ribbon fused to the C11 C-ribbon. Since the Rpb9 N-ribbon and the linker between the two ribbon domains, including the conserved residues 65-70, interact with pol II (Hemming and Edwards, 2000), I fused Rpb9 residues 1-74 to C11 residues 69-110 (protein variant Rpb9-C11-0, figure 10) Surprisingly, this fusion protein conferred very strong RNA cleavage activity to pol II (figure 11A-B, lane 4).

To investigate this interesting gain-of-function mutation, I prepared and functionally analyzed a total of 27 fusion protein variants (variants Rpb9-C11-0 to -24, Rpb9-C11-26, -27, figure 10, figure 12). These experiments revealed that the minimal C11 region required to transfer strong cleavage to Pol II comprised C11 residues 84-110 (figure 11, variants Rpb9-C11-1, -2, -3, -4, -7, -8, lanes 5-8, 11-12). This region forms the core of the zinc-binding C-ribbon fold, suggesting that the C-ribbon domain must be structurally intact to induce strong cleavage. Indeed, the N- and C-ribbon are both required for strong cleavage (figure 11, variants Rpb9-C11-5, -6, -18, -19, lanes 9-10, 22-23) consistent with the finding that individual zinc domains of Rpb9 can't bind to pol II (Hemming and Edwards, 2000).

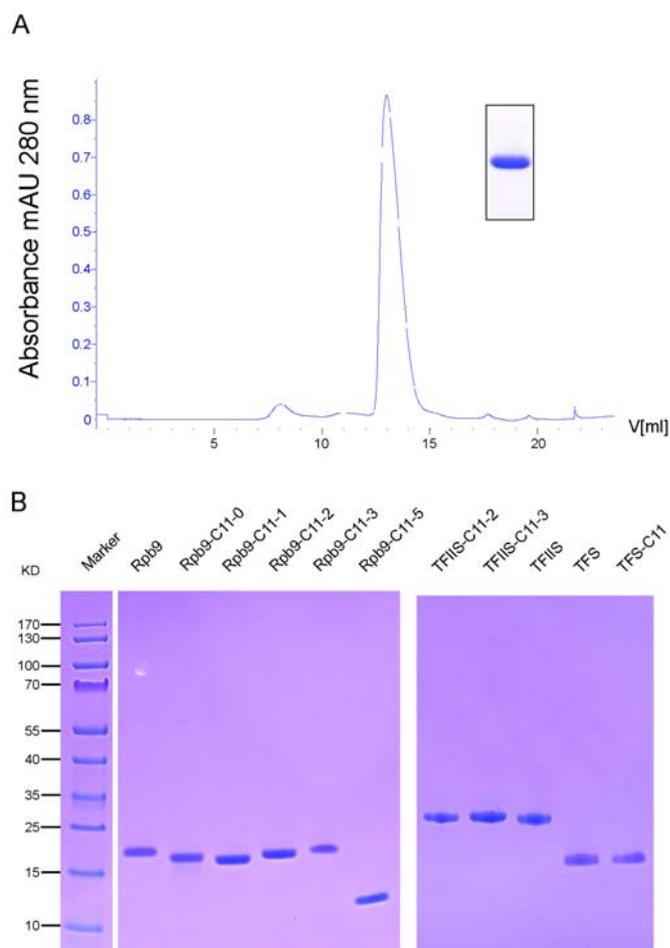


Figure 10. Protein variants used in functional and structural analysis. On the top, an alignment of amino acid sequences of the C-ribbons in *S. cerevisiae* (Sc) Rpb9, A12.2, C11, and TFII5, and *P. furiosus* (Pfu) TFS is shown. Secondary structure elements in Rpb9 and TFII5 are in orange and green, respectively. Below the alignment, the C-terminal sequences of the fusion protein variants are shown.

reaction, the amounts of uncleaved RNA and -2 and -4 cleavage products were quantified. The cleavage activity was calculated as the percentage of -2 and -4 cleavage products with respect to total RNA observed. Reaction times of 10 and 60 min are indicated as red and dark green bars, respectively. Average values for two independent experiments are shown. Experiments were highly reproducible.

2.1.3 Purification of 36 protein variants.

36 protein variants were expressed and purified as described in Experiment Procedures in this Chapter. After cell lysis, *E.coli* lysate was centrifuged and the supernatant was loaded onto Histrap Ni-affinity column. The column was washed with buffer containing 2M NaCl and 40mM imidazol. *E.coli* proteins which bind the chromatographic resin unspecifically and nucleic acids could be removed by the high salt. Low concentration of imidazol removed most of *E.coli* protein contaminants. Then wash the column with buffer containing 100mM imidazol. Protein variants washed at 300mM imidazol already show high purity. Subsequent gel filtration resulted in a single peak and highly pure protein confirmed by SDS-PAGE analysis (figure 12A). Cell lysate containing archaeal TFS variants was additionally treated by high temperature(90°C) for 20min before affinity column which removed almost all of the *E.coli* contaminant. Subsequent purification steps were the same as pol II protein variants. Figure 12 A gives a purification example using Rpb9-C11-1. In figure 12 B some examples of purified protein variants are given showing their final purity by SDS-PAGE analysis.

**Figure 12.**

A. Chromatogram of the Superdex75 gel filtration. Absorbance at 280nm is measured to detect protein elution. SDS-PAGE analysis of the peak fraction is shown next to it.

B. SDS-PAGE analysis of some purified proteins from three variants groups:

Rpb9-C11/TFIIIS fusion proteins comprise all the three domains, or individual N-/C-domains;

TFIIIS-C11 fusion proteins;

TFS-C11 fusion proteins.

2.2 The cleavage-active Pol II variant contains a mobile C-ribbon.

2.2.1 The variant binds weakly to the RNA polymerase II

To investigate the structural basis for the observed strong RNA cleavage, I crystallized the Pol II variant containing the fusion protein Rpb9-C11-1, which has the minimal replacement of Rpb9-C-terminus by residues from C11. In order to get a homogenous Pol II variant, after incubating 5-fold molar excess of Rpb9-C11-1 with Pol II Δ 9 for 20min, a gel filtration run was performed to remove excess of Rpb9-C11-1. Surprisingly, after TCA precipitation and SDS-PAGE analysis, Rpb9-C11-1 can not be found in the expected complex fractions (figure 13). However, by using Rpb9 wild-type, a complete Pol II can be found after gel filtration, indicating a stable and tight binding of Rpb9 to Pol II Δ 9 (figure 13B). Apparently, even after making a minimal replacement, the residues from C11 broke some interactions with Pol II Δ 9, and resulted in weak

binding.

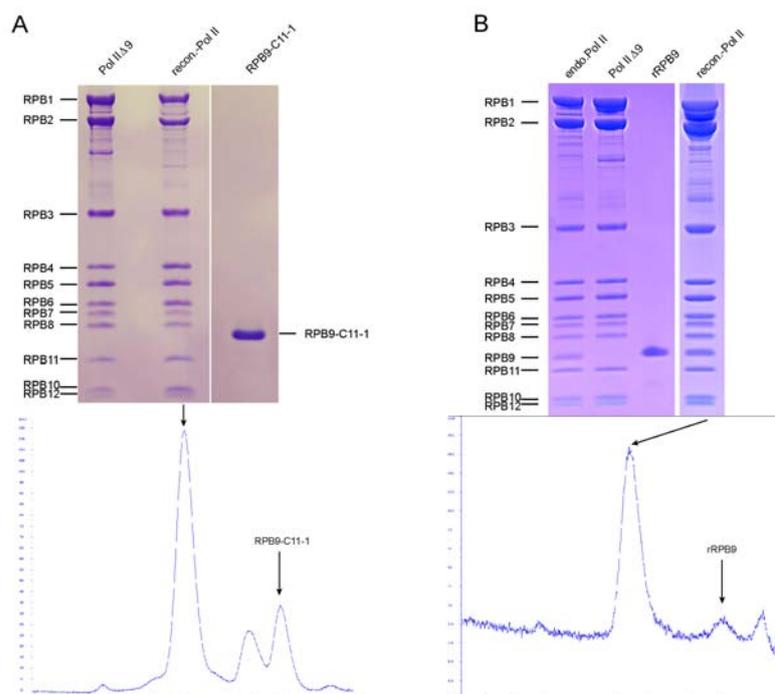


Figure 13. A. SDS-PAGE analysis of reconstitution result of PolIIΔ9 and Rpb9-C11-1. Chromatogram of the Superose6 gel filtration after assembly is shown in the bottom panel. Interested peaks are marked by an arrow. B. SDS-PAGE analysis of reconstitution result of PolIIΔ9 and recombinant Rpb9. Chromatogram of the Superose6 gel filtration after assembly is shown in the bottom panel.

2.2.2 X-ray analysis of the cleavage-active Pol II variant

Because of the weak binding of the cleavage-active variant, a mixture solution of 11-subunit Pol IIΔ9 and Rpb9-C11-1 was subjected to crystallization trials. Equal molar amounts of Pol IIΔ9 and Rpb9-C11-1 were mixed and incubated for 30 min resulting a final protein mixture with a concentration of 4.5 mg/ml. Crystals were grown by mixing 2.5 μl of the protein mixture solution with 1 μl of reservoir solution (Experimental procedures). Crystals grew to a maximum size around $0.1 \times 0.1 \times 0.1 \text{ mm}^3$. Despite extensive efforts, only poorly diffracting crystals could be obtained, but eventually I solved the structure at 4.3 Å resolution (Table 2). The structure revealed that the conformation of Pol II around the active center was unchanged, with the bridge helix straight and the trigger loop open and mobile. The Rpb9-C11-1 N-ribbon and the Rpb9 linker strand β4 (residues 1-48) were located at the Rpb1 jaw as in wild-type Pol II.

However, the C-ribbon was mobile and did not occupy the position of the Rpb9 C-ribbon on the surface (figures 14, 15). These results indicated that strong cleavage was not due to enhanced allostery.

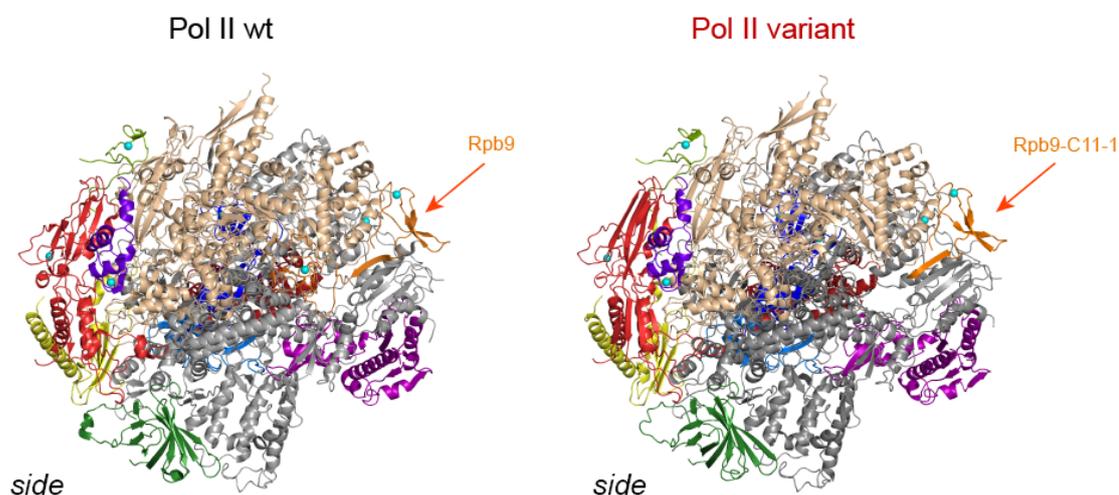


Figure 14. Structure comparison of free RNA polymerase II wild type(left) and Pol II variant(right) which shows high cleavage activity. PDB code of Pol II wt is 1WCM. Both structures are shown as ribbon and side view. The main difference from Rpb9/variant colored as orange is shown by arrows.

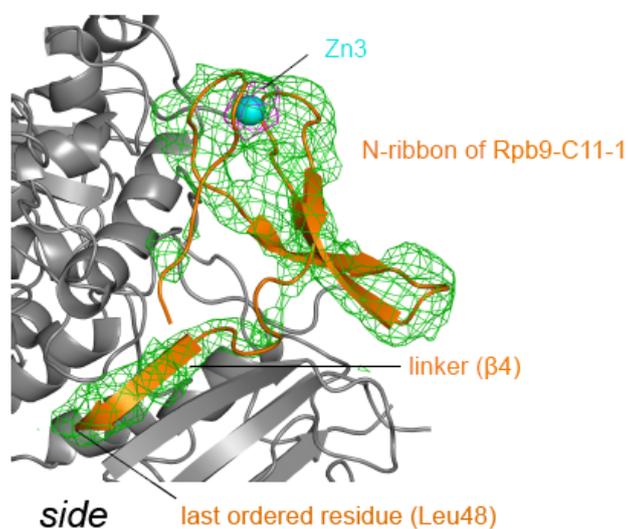


Figure 15. Crystallographic analysis of the highly cleavage-active Pol II variant containing Rpb9-C11-1. Shown is the difference electron density map contoured at 2.5σ (green mesh) for the N-ribbon of Rpb9-C11-1 (orange ribbon model). A peak in the anomalous difference electron density map (magenta mesh) coincides with the position of the N-ribbon zinc ion Zn3 (cyan sphere).

Table 2. Crystallographic data and refinement statistics for the cleavage-inducing Pol II variant containing the fusion protein Rpb9-C11-1.

Data collection	
Space group	C222 ₁
Cell dimensions <i>a, b, c</i> (Å)	222.4, 393.4, 281.4
Resolution (Å)	48.6-4.3 (4.4-4.3) ^a
<i>R</i> _{sym} (%)	10.6 (103.3)
<i>I</i> / $\sigma(I)$	8.6 (2.1)
Completeness (%)	98.4 (99.0)
Redundancy	3.8 (3.9)
Wavelength (Å)	1.2664
Refinement	
Resolution (Å)	48.6-4.3 (4.4-4.3)
No. reflections	82,532 (6065)
<i>R</i> _{work} / <i>R</i> _{free} (%)	23.5 / 28.1
No. atoms	30,544
R.m.s. deviations	
Bond lengths (Å)	0.011
Bond angles (°)	1.099
7 Zn peaks in anomalous difference Fourier (σ)	9.4 11.7 9.9 ^b 9.2 8.2 13.0 11.6

^aValues in parenthesis correspond to the highest resolution shell.

^bPeak for zinc ion in the N-ribbon of Rpb9-C11-1.

2.2.3 Weakening or loss of several contacts allow the detachment of C-ribbon

Inspection of the Pol II structure suggested that detachment of the C-ribbon from the lobe requires weakening or loss of several contacts (figure 16). First, a contact of Rpb9 residue R92 with the lobe residues E262 and D391 is lost in the Rpb9-C11 variant because the arginine is replaced by a serine. Second, Rpb9 residue K93 is within contact distance with Rpb2 residue D391 in the lobe, but this lysine is not present in C11, leading to loss of a potential salt bridge. Third, the C-ribbon residue R91 forms a salt bridge with Rpb1 residue D781 (figure 16). This arginine is invariant in all C-ribbons, thus the salt bridge could in principle be maintained. However, the preceding Rpb9 residues S88 and Q90 buttress C-ribbon residues that interact with the Pol II surface. In

particular, residue Q90 buttresses Rpb9 residues 50-52, which bind the Pol II lobe, and the preceding residue Q87 interacts with the Pol II funnel domain (figure 16). Since the counterparts of S88 and Q90 are hydrophobic in C11 (L85 and I87) and also in the C-ribbons of TFIS, A12.2 and TFS (figure 10), C-ribbon binding to the polymerase surface is apparently weakened. Consistent with this prediction, replacing the two hydrophobic residues in the cleavage-inducing variant Rpb9-C11-1 with valines retained strong cleavage (figures 10, 11, variant Rpb9-C11-9). These results help rationalize why the C-ribbon is detached from the polymerase surface and mobile in the cleavage-active Pol II variant.

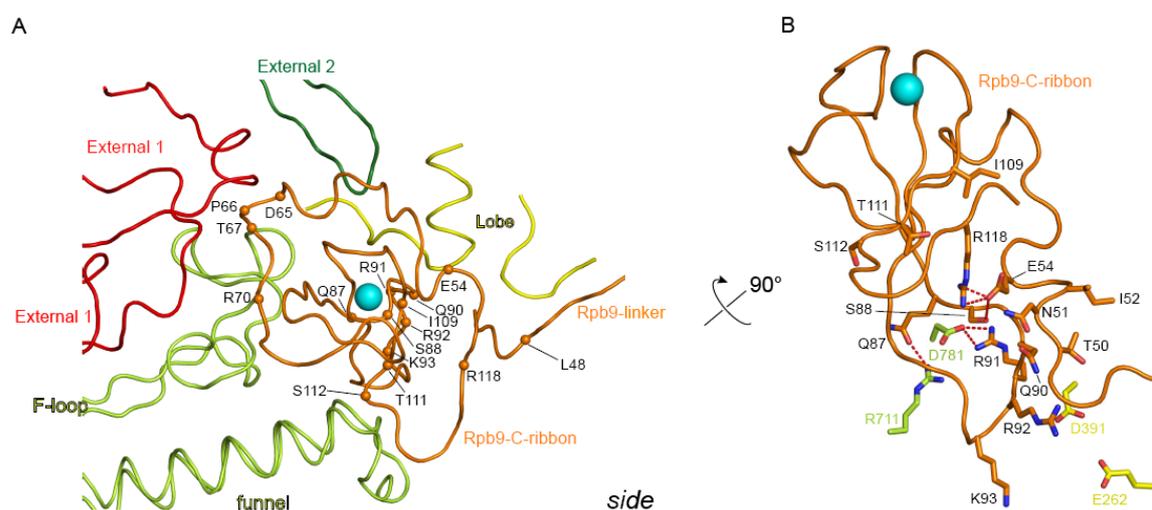


Figure 16. Interface between the Rpb9 C-ribbon and linker (orange) and Pol II domains in different colors as indicated (PDB 1WCM). (A) Side view as in Figure 2 and 4. Orange spheres indicate the location of Rpb9 amino acid residues referred to in the results. (B) View rotated by 90 degrees with respect to that in A as indicated. Important interface residues in Rpb9 and Rpb1 are depicted. Dashed lines indicate salt bridges.

2.3 Evidence that the C-ribbon is catalytic and binds the Pol II pore

The above results suggested that cleavage stimulation by the Rpb9-C11 fusion protein is not due to enhanced allostery, but that a switch in cleavage mechanism occurred, and the C-ribbon transiently inserted into the pore, to

directly stimulate cleavage by complementation of the active center with catalytic residues in the hairpin. This model predicted that the hairpin residues are required for cleavage stimulation, just like the corresponding catalytic residues in TFIIIS. Indeed, mutation of C11 residues D91 and E92 in the β_6 - β_7 hairpin of the fusion protein or just residue D91 to alanine abolished cleavage (figures 10, 11, variants Rpb9-C11-16/17, lanes 20/21). The model also predicted that the residue K108 in the C-ribbon forms a salt bridge with Rpb1 pore residue D1359, as observed in the Pol II-TFIIIS complex structure (Kettenberger, *et al.*, 2003). Indeed, mutation of K108 leads to a strong reduction in cleavage stimulation (figures 10, 11, 17, variant Rpb9-C11-20, -21, -24, lane 24, 25, 28). In addition, the conserved residue E109 in the C-ribbon forms a salt bridge with the Rpb1 residue K619 that is also located in the pore, and is invariant in Pol III enzymes. Consistent with this proposal, deletion of the C-terminal C11 residue E109 leads to a strong reduction in cleavage stimulation (figures 10, 11, 17, variant Rpb9-C11-22, -23, lanes 26, 27). Variants that do not contain this residue also lost activity (figures 10, 11, variant Rpb9-C11-13, -14, lanes 17, 18).

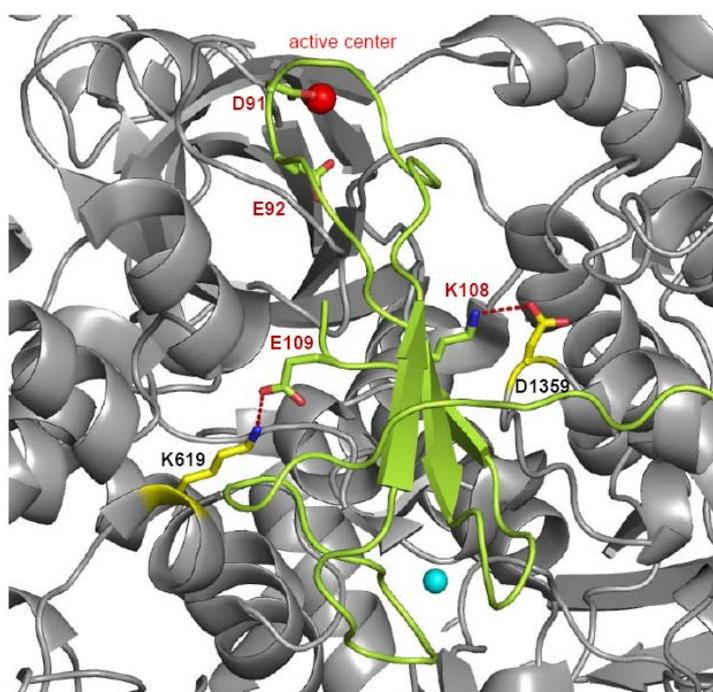


Figure 17. Interactions between the Rpb9-C11-1 C-ribbon model and Pol II in the secondary channel. Side view as in Figure 14-16. Mg in the active center is shown as red sphere. Zn in the Rpb9-C11-1 C-ribbon is shown as cyan sphere. Dashed lines indicate salt bridges.

2.4 The C-ribbon could reach the pore and active center through a long linker

We next investigated whether and how the C-ribbon could reach the pore and active center. Modelling showed that the Rpb9 linker residues 48-53 are not long enough to link the Rpb9-C11 N-ribbon located on the jaw with a C-ribbon located in the pore. However, residues 54-85 could additionally be used to link the domains if their limited interactions with the Rpb9 C-ribbon would be broken in the Rpb9-C11-1 variant. This is apparently achieved in the variant because Rpb9 C-ribbon residues I109, and T111, which interact with linker residues, are replaced with arginine and lysine, respectively, in the fusion protein, which apparently breaks the hydrophobic contacts between the linker and C-ribbon. Consistent with this proposal, mutations in the cleavage-inducing variant Rpb9-C11-1 that were predicted to prevent detachment of the Rpb9 linker from the C-ribbon could not stimulate strong cleavage (figure 10, 11, Rpb9-C11-3, -7, -8, -10, -11, -12, -20, -21, -24, lanes 7, 11-12, 14-16, 24-25, 28).

I also tested whether shortening of the linker between the two ribbons would abolish cleavage because the C-ribbon could not reach the active center. Indeed, variants with shorter linkers did not induce strong RNA cleavage (figures 10, 11, variants Rpb9-C11-26, -27). In addition, Rpb9 contains a salt bridge between the linker residue E54 and R118 in the C-ribbon (figure 16), but this is lost in cleavage-inducing variants that lack the C-terminal arginine. In variants that could form this interaction, strong cleavage activity was lost (figures 10, 11, variant Rpb9-C11-3, -7, -10, -11, -12, -15, lanes 7, 11, 14-16, 19).

2.5 The C11 C-ribbon functions in the Pol II pore

All the above results support the model that in the Rpb9-C11-1 variant the N-ribbon remains on the jaw, whereas the C-ribbon transiently occupies the pore to induce strong RNA cleavage. This requires that the C11 C-ribbon can function in the Pol II pore. To test this, I prepared TFIIIS variants in which the TFIIIS C-ribbon is replaced by the C11 C-ribbon. Indeed, such fusion proteins were as active as wild-type TFIIIS (figures 10, 11, variant TFIIIS-C11-2, -3). Further, the model predicted that replacing the C11 C-ribbon in the Rpb9-C11-1 variant by the TFIIIS C-ribbon should also induce strong RNA cleavage. This was indeed observed, although cleavage was weaker when I replaced the TFIIIS linker with the Rpb9 linker (figures 10, 11, variants Rpb9-TFIIIS-1-4). Weaker cleavage induction by the variants Rpb9-TFIIIS-3/4 compared to the variants Rpb9-TFIIIS-1/2 can however be explained by a loss of TFIIIS residues D267 and R268 that form salt bridges with Pol II at the entrance to the pore (Kettenberger, *et al.*, 2003) and the loss of E109 in the C-ribbon which is also predicted to form a salt bridge in the pore. These results show that the C11 C-ribbon can bind the Pol II pore and induce strong RNA cleavage, and that a cleavage-inducing C-ribbon can reach the pore if tethered to the Rpb9 N-ribbon located on the jaw.

2.6 Catalytic C-ribbons are conserved between archaea and eukaryotes

The above analysis suggested a simple evolutionary relationship between A12.2, Rpb9, C11, and TFIIIS (figures 10, 11, 15). First, A12.2 and C11 correspond to the archaeal TFS. In A12.2, C11, and TFS, the N-ribbon corresponds to that of Rpb9, whereas the C-ribbon corresponds to that in TFIIIS. To test this prediction, we performed cleavage assays with the archaeal RNA polymerase from *Pyrococcus furiosus* (*Pfu*). The polymerase alone could not induce cleavage, but addition of recombinant *Pfu* TFS enabled strong cleavage (figure 18) consistent with previous reports (Hausner, *et al.*, 2000;

Lange and Hausner, 2004). Mutagenesis revealed that cleavage required the TFS hairpin residues D90 and E91 as predicted (figures 10, 18). Addition of a fusion protein in which the TFS N-ribbon was fused to the C11 C-ribbon (figure 10, 18) also enabled cleavage, strongly arguing that the pore-binding cleavage-inducing function of the C-ribbon was conserved between archaea and eukaryotes during evolution, and supporting our model for the domain relationships.

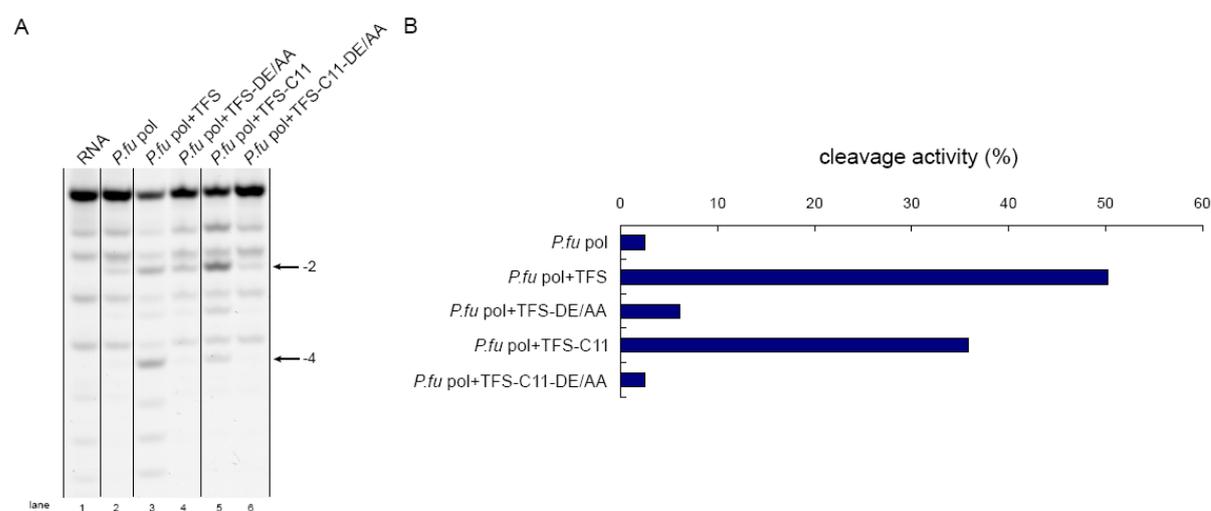


Figure 18. The C11 C-ribbon functions in the archaeal system. (A) Electrophoretic analysis of RNA products in a cleavage assay with different protein variants (Figure 11). RNA bands obtained after cleavage of mainly two or four nucleotides are indicated by arrows (-2 and -4, respectively). Lane 1 shows the reactant RNA. (B) Quantification of cleavage activities.

3. Discussion

3.1 Two cleavage models in RNA polymerases.

Work in this thesis unravels the molecular basis for the difference in RNA cleavage activities of Pol II and Pol III. I show that replacement of the Rpb9 C-ribbon by the C11 C-ribbon confers strong intrinsic cleavage to Pol II. This unexpected gain of function stems from a switch in the cleavage mechanism, as suggested by X-ray crystallography and mutagenesis. Whereas the Rpb9

C-ribbon acts allosterically from the polymerase surface, the C11 C-ribbon acts directly by binding the pore and complementing the active center with its catalytic hairpin. Thus two modes exist for polymerase-intrinsic RNA cleavage, an allosteric, weak mode used by Rpb9, and a direct, strong mode used by C11 and TFIIIS.

3.1.1 Weak cleavage activity in Pol II via an allosteric model.

After Surratt et al. (Surratt, *et al.*, 1991) first discovered that polymerases possess intrinsic cleavage, further work unravelled that this activity is conserved in many DNA-dependent RNA polymerases and is important in the control of elongation. In pol II, subunit Rpb9 was studied a lot for its contribution for transcription fidelity, however, because of experimental strategy they applied, Rpb9 has never been demonstrated to contribute to the intrinsic cleavage activity (Awrey, *et al.*, 1997; Weilbaecher, *et al.*, 2003). In this thesis, by using a bead assay under physiological pH, the result indicates that Rpb9 is indeed involved in the weak cleavage activity of pol II. Because Rpb9 locates on the surface of pol II, it must stimulate the activity through an allosteric model. However, structure of free polymerase II didn't give any information about how Rpb9 contribute to the intrinsic cleavage activity. There are two possibilities as follows that do not contradict each other.

The first possibility is that Rpb9 affects the dynamics of an element called trigger loop (TL) indirectly (figure 19). The trigger loop is strictly conserved from bacteria to archaeal and eukarya, based on its role to transcription. Trigger loop is also a very flexible structure which can have at least five conformations: closed, opened, wedged, backtrack-intermediate, trapped (Brueckner and Cramer, 2008; Cheung and Cramer, 2010; Cramer, *et al.*, 2001; Wang, *et al.*, 2009; Wang, *et al.*, 2006). The flexibility is important for regulating transcription. When the trigger loop is in the backtrack-intermediate conformation, residues N1082 and Q1078 contact the phosphate group between backtracked residue +2 and residue +1. Together with other residues

of polymerase, backtracked residues can be stabilized and coordinate and orient Mg II and active water for nucleophilic attack during cleavage reaction (Zenkin, *et al.*, 2006). The trigger loop was indeed proved to be essential for intrinsic cleavage *in vitro* in *Thermus aquaticus* (Yuzenkova and Zenkin, 2010). Rpb9 doesn't appear to have direct interaction with trigger loop at this conformation, and, Rpb9 interacts with another two elements: the F-loop (FL) and funnel, and they in turn, interact and stabilize TL through several hydrogen bonds as well as hydrophobic interactions (figure 19). Most of the interactions are from the funnel. Although in *Taq* RNAP, FL was demonstrated to affect intrinsic cleavage activity only slightly (Miropolskaya, *et al.*, 2009), it is very short in bacteria, while having a large insertion in eukaryotic FL which enables FL to interact with TL.

The second possibility is that Rpb9 modifies the overall structure of RNA polymerase. The N-terminus of Rpb9 connects pol II "lobe" domain from Rpb2 and "jaw" domain from Rpb1 like a clip, thus the role of Rpb9 to intrinsic cleavage would be to maintain the conformation of whole polymerase which provides optimal chemistry environment for cleavage reaction to happen. Consistently, strains which have only N-terminus of Rpb9 are not sensitive to high temperature or nucleotide-depleting drug mycophenolate (Van Mullem, Wery, *et al.*, 2002).

Rpb9 may also affect other regions of pol II like Bridge helix which is also important for transcription. However, there is no evidence so far that bridge helix is required for pol II cleavage activity.

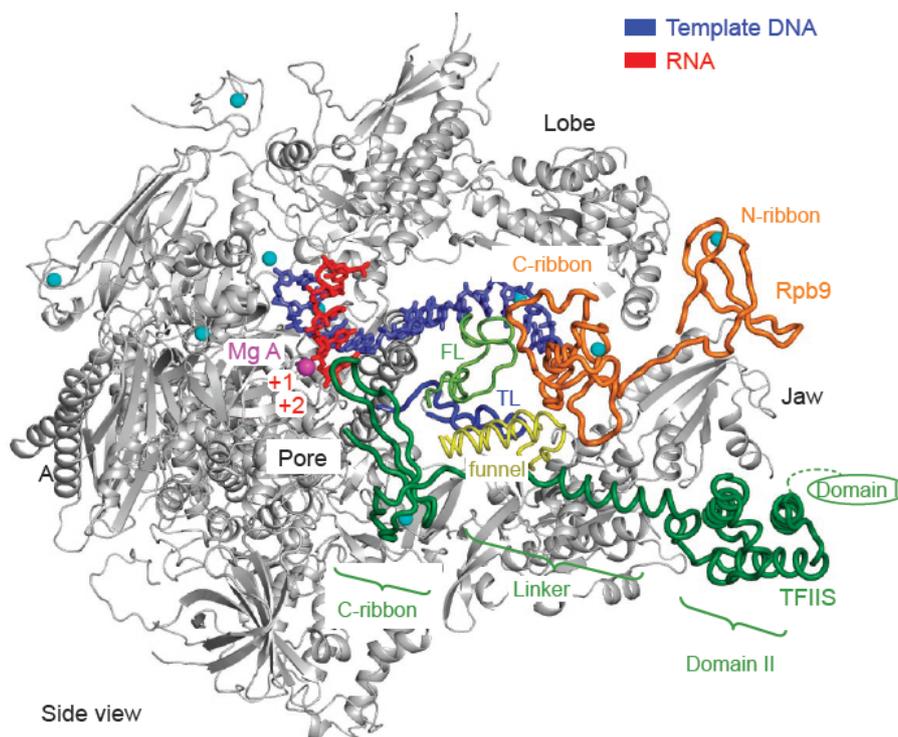


Figure 19. Structure of the Pol II-TFIIS complex with nucleic acids viewed from the side.

3.1.2 Strong cleavage activity in Pol I, III and archaeal RNAP via a direct model

In eukarya, RNA pol II produces protein-coding RNAs which will be translated to protein to function, while pol I and III produce ribosomal RNAs and tRNAs which are directly used. Thus, accuracy of transcription in these two polymerases appears to be more crucial. The most efficient way to increase cleavage activity is to retain a second Mg ion and active water directly and stimulate cleavage reaction. Work of this thesis indicates the similar mechanism between C11 and TFS/TFIIS and also indicates the flexibility of C11 C-ribbon. This mechanism represents another model in intrinsic cleavage activity via a direct way (figure 19) and suggests a scenario for proofreading process of pol I and III: as an intrinsic subunit of these polymerases, A12.2/C11 N-terminus associates with polymerase during transcription, allowing their fast and efficient interactions with other regions of polymerase through their flexible C-ribbon; as soon as a mistake happens, forward translocation is disfavored, opening a time window for A12.2/C11 C-ribbon to

go to the secondary pore, interact with polymerase and correctly position the catalytic acidic residues; these residues coordinate second Mg ion, activate and orient a water molecule for nucleophilic attack of the phosphodiester bond; after cleavage, A12.2/C11 C-ribbon leaves the pore, and elongation continues. Since A12.2 has a much longer linker than C11/Rpb9, this linker might be able to contribute to affinity for RNA Pol I, just like the TFIIIS linker for the RNA Pol II, and thus make A12.2 stimulate Pol I cleavage more efficiently.

An interesting indication from this work is that C-ribbon of these subunits/factors cleavage determines the mechanism of cleavage stimulation. Rpb9 C-ribbon replaced by C11 C-ribbon switches the mechanism to the direct model, stimulating high intrinsic cleavage activity in Pol II. Rpb9 C-ribbon replaced by TFIIIS C-ribbon also switches to corresponding mechanism of TFIIIS.

3.2 Evolution of the cleavage activity.

My results also suggest a model how polymerase cleavage activities evolved (figure 20). Pol I and Pol III have strong intrinsic cleavage activities because they contain homologues of archaeal TFS (A12.2 and C11, respectively) that contain C-ribbons with catalytic hairpins that can enter the pore to directly stimulate cleavage. In the Pol II system, the two domains are however part of two different polypeptides. Whereas the N-ribbon is part of Rpb9, the C-ribbon is part of TFIIIS. During evolution, the C-ribbon likely duplicated and was altered in Rpb9 to attach the domain to the surface and to allow only for weak, allosteric cleavage induction. Corresponding to this model, the N-terminal domain of TFS probably positions itself on archaeal RNAP in a position similar to the Rpb9 N-terminal domain, contacting the polymerase lobe and jaw domain while its C-terminal domain positions itself in a similar fashion to the TFIIIS C-terminal domain in the pore. A12.2, C11 are predicted to inherit the same localization on RNA Pol I and III from archaeal TFS(figure 19, figure 20).

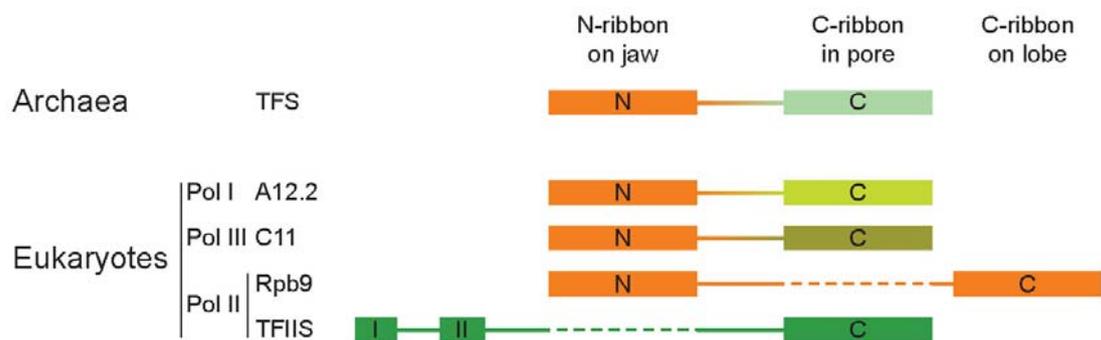


Figure 20. Proposed evolutionary relationship between zinc ribbon domains in TFS, A12.2, C11, Rpb9, and TFIIIS.

These results are consistent with published data. First, mutation of the C11 hairpin residues is lethal (Chedin, *et al.*, 1998). Second, the C11 C-ribbon is not observed on the surface in a recent electron microscopic structure of Pol III, consistent with transient binding to the pore (Fernandez-Tornero, *et al.*, 2010). Third, C11 and A12.2 are required for transcription termination by Pol I and Pol III (Chedin, *et al.*, 1998; Prescott, *et al.*, 2004), and the termination mechanism likely resembles that of archaeal polymerase (Spitalny and Thomm, 2008), but is different in Pol II. Fourth, the A12.2 and C11 C-ribbon domains may be able to swing between surface and pore locations, because some density for the A12.2 C-ribbon was observed near the lobe in a Pol I EM reconstruction (Kuhn, *et al.*, 2007), and because the strong Pol III cleavage can be even further enhanced by a mutation of the largest subunit that is predicted to disrupt a salt bridge between the Pol III counterpart of Rpb1 residue D781 in the funnel domain F-loop and the C11 residue R88 (corresponding to Rpb9 R91) (Thuillier, *et al.*, 1996). It remains to be confirmed that A12.2 uses the same mechanism as C11. Unfortunately, replacing the Rpb9 C-ribbon with the A12.2 C-ribbon did not confer strong cleavage to Pol II (not shown), likely because Pol I has diverged much more from Pol II than Pol III.

These results unveil the exceptional nature of Pol II, in contrast to Pol I, Pol III, and the archaeal polymerase, with respect to RNA cleavage. In the Pol II

system, implementation of allosteric and direct cleavage stimulatory modes on two different proteins may have enabled new mechanisms of transcription regulation such as regulation by release of promoter-proximally stalled Pol II (Adelman, *et al.*, 2005; Cheung and Cramer, 2010; Nechaev, *et al.*, 2010; Palangat, *et al.*, 2005). It may also have prevented premature Pol II termination at sites that would terminate Pol III, and may have enabled elaborate 3'-end processing of Pol II transcripts. The weak intrinsic cleavage activity of Pol II may however suffice for proofreading after ubiquitous misincorporation events.

4. Conclusions and Outlook

Work of this thesis elucidates an important problem in the enzymatic action of eukaryotic RNA polymerases, namely the difference in the intrinsic transcript cleavage activity of the eukaryotic RNA Pol II with Pol I and Pol III. Given that the three polymerases have similar components, why is intrinsic cleavage slow and essentially factor-dependent uniquely for Pol II? Previous work indicated that this difference may lie in the A12.2/Rbp9/C11 subunits. I created a large collection of hybrid molecules between these two subunits, primarily between the N-terminus of Rbp9 and the C-terminal ribbon domain of C11; many new Rbp9-C11 fusions were then made Pol II efficient in transcript cleavage, comparable to the normal enzyme with TFIIS. Structural data indicated that in Pol II with the stimulating-active Rbp9-C11-1 chimeric protein, the C11 segment was not located over the jaw as with intact Rbp9, indicating that the C-terminal segment in the fusion subunits was directly accessing the active site, as is the case with TFIIS and its highly flexibility which is required for the function. Cleavage results with many variants of the fusion subunit fully support the model.

In future, some experiments can still be tried to further confirm this model. First, this model suggests that C11 C-ribbon passes through the pore, arriving at the active center. Therefore, any block on this way will impede its stimulatory function. This aim can be achieved by using α -amanitin, which is a toxin that binds to a pocket formed by F-loop and funnel, and sits in the secondary channel (Brueckner and Cramer, 2008). Published data prove that α -amanitin slows but does not stop the intrinsic cleavage of Pol II, in contrast, it completely blocks the TFIIS-stimulated cleavage, which apparently because TFIIS can't reach the active center (Izban and Luse, 1992; Rudd and Luse, 1996; Weilbaecher, *et al.*, 2003). Therefore, α -amanitin is an excellent candidate for directly pointing out the pathway of C11 stimulated cleavage.

Second, crosslinking technique can also directly indicate the mechanism,

although photochemical RNA-protein crosslinking using 5-[N'-(p-azidobenzoyl)-3-aminoallyl]UTP (ABUTP) incorporated at 3' end of RNA failed, since C11 was found to strongly bind to free-RNA and caused high non-specific background (Kassavetis, *et al.*, 2010). However, site-specific protein-protein crosslinking can be used instead. By using Mg-absent buffer and mismatched transcript substrate, it might induce C11 to bind the pore more frequently which can then be crosslinked to specific positions, which can then be identified by MS.

This work also indicated that Rpb9 contributed to the intrinsic cleavage. But how it works allosterically is unknown. The trigger loop is a potential target for its stimulatory function, and the negative effect of α -amanitin might also suggest a contribution of the trigger loop which is trapped by α -amanitin through just His1085 (Brueckner and Cramer, 2008), that is therefore consistent with the observation that intrinsic cleavage can be slowed but not inhibited. But since Rpb9 doesn't have direct interactions with trigger loop, how does it affect it? Are there any other targets? This could be an important and interesting project. The answer might be achieved by making Rpb9 variants and test the intrinsic cleavage activity. Once an interesting mutant is found, mutations in trigger loop, funnel, or F-loop (see discussion) which can interact with trigger loop, can be further studied to investigate the allosteric pathway. Crystallography can also be done jointly.

This work also pointed out an evolutionary connection between TFS and A12.2/Rpb9/C11, and suggested that TFS N-ribbon binds to the jaw/lobe, and C-ribbon goes to the pore. Since *in vitro* reconstitution of fully recombinant archaeal polymerases have been developed (Naji, *et al.*, 2007; Werner and Weinzierl, 2002), it would be convenient to find out the interaction sites between TFS N-terminus and archaeal polymerase. Thus this would be another promising project related to this thesis.

5. Experiment procedures

5.1 Isolation of yeast genomic DNA

5.1.1 Buffers

Breaking buffer

2% TritonX-100

1% SDS

100mM NaCl

10mM Tris-HCl pH8.0

1mM EDTA pH8.0

PCI(4°C)

Phenol:Chloroform:Isoamylalcohol 25:24:1

RNase A stock

10mg/ml in H₂O, boil for 5min at 100°C

5.1.2 Isolation of yeast genomic DNA procedure

Collect 10ml overnight yeast culture. Pellet cells by centrifugation at 4000rpm, RT for 5 minutes. Resuspend pellets in 0.5ml H₂O, transfer to a safe lock tube. Centrifuge for 5sec. at 14000rpm and remove the supernatant. Add 200µl breaking buffer to the pellet, resuspend then add 0.3g Glass beads and 200µl PCI. Strong vortex for 3~4min in cold room. Add 400µl TE, shortly vortex. Centrifuge at 14000rpm for 5min. Transfer aqueous phase to a new tube, and add 1ml 96% EtOH followed by centrifugation at 14000rpm for 3min. Remove all the supernatant by quick spin. Then dry the pellet at 37°C for 3min. Then resuspend the pellet in 200µl H₂O or 10mM Tris. For RNase digestion, 7µl RNase A was added and incubated at 37°C for 15min. Mix by inverting after adding 20µl 3M NaAC and 500µl 96% EtOH. Centrifuge at 14000rpm for 3min and remove the supernatant. Dry the pellet at 37°C for 3min. Finally, resuspend pellet in 50µl H₂O.

5.2 Cloning, Expression and purification of cleavage factors

5.2.1 Design and cloning of different cleavage factor variants

Fragments containing wild-type Rpb9 coding regions or C11 coding regions were amplified by PCR using yeast genomic DNA as a template. Fusion proteins were constructed with two-step PCR. The first PCR reaction generated two fragments, one fragment containing the N-terminal sequence of Rpb9 and a short sequence of C-terminal C11, and a second fragment containing the C-terminal sequence of C11 and the N-terminal part of Rpb9. These fragments were used as templates in the second PCR reaction that produced the coding region for the fusion protein containing the required mutations, as well as NdeI and XhoI restriction sites at the 5'- and 3'-terminus, respectively. The same strategy was used for other fusion proteins. Each PCR product was digested with restriction endonucleases and inserted into the pET 28b(+) expression vector (Novagen) to generate a fusion protein with an N-terminal hexahistidine tag. All plasmids were verified by sequencing.

5.2.2 Expression and purification of eukaryotic cleavage factor variants

5.2.2.1 Buffers

PBS buffer

8 g NaCl

0.2 g KCl

1.44 g Na₂HPO₄

0.24 g KH₂PO₄

Adjust the pH to 7.4 with HCl. Add H₂O to 1 liter

Purification buffer

50 mM Hepes-Na, pH 7.5

10 μM ZnCl₂

10% glycerol

Transcription buffer

20 mM Hepes-Na, pH 7.6

60 mM (NH₄)₂SO₄

8 mM MgCl₂

10 μM ZnCl₂

10% glycerol

10 mM DTT

5.2.2.2 Expression and purification procedures

About 12ml overnight preculture was added to fresh LB medium supplemented with antibiotics and shaken at 37°C in 140rpm until the culture reached log phase (OD 0.6-0.8). After cooling the *E.coli* suspension on ice, isopropyl-1-thio-β-D-galactopyranoside (IPTG) was added to a final concentration of 40μM. Cells were grown over night at 18°C, 140rpm. Then collect cells by centrifugation at 4°C for 15min in 5000rpm (SLS6000 rotor).

Cells were sonicated in PBS buffer containing 10 μM ZnCl₂. All subsequent steps were performed at 4°C. The supernatant of the resulting crude extract was collected by centrifugation at 16,000rpm for 15min and applied to a 1 ml HisTrap column (GE Healthcare), charged with 100 mM NiSO₄ and equilibrated in purification buffer containing 100 mM NaCl. The column was washed with 10 column volumes of purification buffer containing 2 M NaCl and 40 mM imidazole followed by another 10 column volumes of purification buffer containing 100 mM NaCl and 100 mM imidazole. The fusion protein was eluted from the column with purification buffer containing 100 mM NaCl and 300 mM imidazole. The eluate was concentrated and further purified on a gel filtration column (Superdex 75, GE Healthcare), equilibrated with transcription buffer and stored at -80°C. The final preparation for each protein had a protein concentration between 0.2 and 0.5 mg/ml.

5.2.2 Expression and purification of archaeal cleavage factor variants

Expression of archaeal cleavage factor variants (TFS/TFS-AA) was the same as the procedure for eukaryotic cleavage factors. Cells were collected by centrifugation at 4°C for 15min in 5000rpm (SLS6000 rotor).

Cells were sonicated in PBS buffer containing 10 µM ZnCl₂. The supernatant of the resulting crude extract was collected by centrifugation at 16,000rpm for 15min. Then heat treatment was performed for 20min at 90°C. After ultracentrifugation for 1h at 30,000rpm, the proteins remained in the supernatant and were applied to a 1 ml HisTrap column (GE Healthcare). All subsequent purification steps were as above and performed at 4°C. The final preparation for each protein had a protein concentration between 0.2 and 0.5 mg/ml.

5.3 Assembly of transcription elongation complexes

Pol II Δ9 was provided by Elisabeth Lehmann. It was created from strain BJ5464 Rpb3 His-Bio (Cheung and Cramer, 2010; Kireeva, *et al.*, 2000) by homologous recombination. The natNT2 cassette was amplified by polymerase chain reaction (PCR) from pFA6a-natNT2 (Janke, *et al.*, 2004) and was used to replace the RPB9 ORF. The gene replacement was verified by PCR. Core pol IIΔ9 were purified essentially as described (Sydow, *et al.*, 2009). Recombinant Pol II subcomplex Rpb 4/7 was purified essentially as described (Armache, *et al.*, 2005).

Purification of *Pyrococcus furiosus* RNA polymerase will be described in Chapter III.

Assembly procedure for *in vitro* bead-based assays as described (Komissarova, *et al.*, 2003) was also shown in figure 9. First, equal molar amounts of template DNA and RNA whose 5' was labeled by fluorescence FAM group were mixed in RNase-free TE buffer. The oligonucleotides were annealed by heating to 95°C for 2 minutes and slowly cooling to room

temperature. Then 5 picomol of core Pol II Δ 9 was incubated with a two-fold molar excess of the annealed hybrid in transcription buffer for 15 minutes at 20°C shaking in 350rpm. A four-fold molar excess of non-template DNA, containing Biotin at the 5' end, was added and the mixture was incubated for 10 minutes at 25°C shaking in 350rpm. Then 5-fold molar excess purified recombinant Rpb4/7 was added and incubated for 10 min at 25°C shaking in 350rpm.

5.4 Transcription cleavage assay

5.4.1 Buffers

Beads breaking buffer

50 mM Tris-HCl, pH 8.0

150 mM NaCl

0.1% TritonX-100

5% Glycerol

0.5 mM DTT

Beads blocking buffer

50 mM Tris-HCl, pH 8.0

150 mM NaCl

2 mM EDTA pH 8.0

0.1% TritonX-100

5% Glycerol

0.5% BSA

200 μ l/ml Insulin

0.1 mg/ml heparin

0.5 mM DTT

5.4.2 Transcription cleavage assay procedures

The bead-based cleavage assay was carried out as described shown in figure 9 (Dengl and Cramer, 2009; Sydow, *et al.*, 2009). Streptavidin-coated

magnetic beads (Dyna-beads MyOne Streptavidin T1, Invitrogen) were washed twice with beads breaking buffer. Then incubate in 500 μ l beads blocking buffer overnight at 4°C. Before the assay, wash twice again with transcription buffer containing 0.1% TritonX-100 and resuspend the beads in the original volume(20 μ l beads per reaction). The beads were added to assembled transcription elongation complexes followed by incubation for 30min at 25°C shaking in 550rpm. Unbound complexes were removed by three steps washing with transcription buffer containing 0.1% TritonX-100, 200mM Ammonium sulfate and only transcription buffer respectively (figure 9). Beads were resuspended in transcription buffer and 5-fold molar excess purified recombinant proteins were added followed by incubation at 28°C shaking in 550rpm. For assays with *Pyrococcus furiosus* RNA polymerase, purified TFS and its variants (5-fold molar excess) were added to the reaction mixture followed by incubation at 70°C for 10 min. The reactions were stopped by adding an equal volume of 100 mM EDTA. Samples were 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). All gel bands were quantified using ImageQuant7 (GE Healthcare).

5.5 Crystallization and structure determination

5.5.1 TCA precipitation

TCA precipitate samples by adding 100% TCA to protein solution to reach a final concentration of 15%. Then mix well and incubate from 30 min to several hours on ice. Centrifuge for 1 hour in 13,000rpm at 4°C. After carefully removing and discarding the supernatant, wash the pellet by 1ml ice cold acetone. Then centrifuge again for 5min and discard supernatant. Let the pellets dry in a fume hood for 1 hour. Add 1X SDS-PAGE loading Buffer. If there is still some acid left in the precipitate, it turns yellow since loading buffer contains bromphenol blue. Neutralize the sample with NH₃.

5.5.2 Assembly of 11-subunit RNAP and cleavage factors and crystallization

5.5.2.1 Buffer

Pol II buffer

5 mM HEPES-Na, pH 7.25

40 mM (NH₄)₂SO₄

10 μM ZnCl₂

5 mM DTT

5.5.2.2 Assembly and analysis procedures

To investigate whether the recombinant Rpb9 and variants can still bind tightly to Pol IIΔ9, 50μg purified core Pol IIΔ9 was incubated with 5-fold molar excess Rpb9 or variants and Rpb4/7 for 30min at 20°C. Then assembled polymerase was loaded onto a gel filtration column equilibrated with Pol II buffer (Superose 6, GE Healthcare). Collect fractions from peaks and perform TCA precipitation. Denaturing gel electrophoresis was adapted to separate complex proteins. Both discontinuous Laemmli system and NuPAGE® Bis-Tris system (Invitrogen) were used for analysis. Protein samples were totally unfolded by adding β-mercaptoethanol to the loading dye. Gels were then stained with Coomassie (SIGMA) solution. While Rpb9 showed tight binding to Pol IIΔ9 and formed a complete 12-subunit complex, Rpb9 variants bound weakly and separated during gel filtration (see results).

5.5.2.3 Crystallization of Pol IIΔ9 and variant Rpb9-C11-1

Pol IIΔ9 was incubated with 5-fold molar excess of recombinant Rpb4/7 for 20 min at 20°C. The complex was purified on a Superose 6 gel filtration column (GE Healthcare), which was equilibrated with pol II buffer to generate the 11-subunit Pol IIΔ9. For buffer exchange, a single gel filtration run was performed with Rpb9-C11-1. Equal molar amounts of 11-subunit Pol IIΔ9 and

Rpb9-C11-1 were mixed and incubated for 30 min at 20°C. After centrifugation the final protein mixture had a concentration of 4.5 mg/ml. Crystals were grown at 20°C using the hanging drop vapor diffusion method by mixing 2.5 µl of the protein mixture solution with 1 µl of reservoir solution (50 mM Hepes, pH 7.0, 200 mM NH₄Ac, 300 mM NaAc, 5-6% w/v PEG6000, 5 mM TCEP). Crystals grew to a maximum size $\approx 0.1 \times 0.1 \times 0.1$ mm³. For cryo-protection, crystals were transferred stepwise over 5h to the reservoir solution containing additionally 0-22% (v/v) glycerol. After incubation at 8°C for 24h in the presence of cryo-protectant, crystals were flash-cooled by plunging them into liquid nitrogen.

5.5.3 Structure determination

Diffraction data were collected as consecutive series of 0.25° rotation images at cryo-temperature at the beamline X06SA at the Swiss Light Source. Diffraction data were processed with XDS and scaled with XSCALE (Kabsch, 1988). The structure was solved by molecular replacement with PHASER (McCoy, *et al.*, 2005) using the structure of the 12-subunit Pol II (PDB code 3HOU, (Sydow, *et al.*, 2009)) with Rpb9 omitted as a search model. The structure was refined with PHENIX (Afonine, *et al.*, 2005), using additional hydrogen-bond distance restraints for secondary structure elements (Kostrewa, *et al.*, 2009), against the observed data that were sharpened with a B-factor of -80 \AA^2 .

Chapter III: Structural study of archaeal transcription pre-initiation complex (PIC) (unpublished)

1. Introduction

1.1 Transcription initiation in eukarya

Eukaryotic transcription initiation on protein-coding genes is a rather complicated event compared to bacteria. It requires not only polymerase itself, but also many cofactors. Several elements in DNA are recognized by these proteins and allow transcription initiation and its regulation. An central element is called core promoter, which provides a platform for assembly of a group of general transcription factors(GTF) as well as RNA polymerase. This huge complex is the preinitiation complex, which has the intrinsic ability to accurately transcribe through core promoter and allows basal transcription on naked DNA *in vitro*. However, a remarkable feature of eukarya is that basal transcription machinery always acts in conjunction with diverse cofactors, regulating the expression of different genes. These cofactors either modify chromatin structure, or directly regulate formation or function of the preinitiation complex, positively, or negatively (Li, *et al.*, 2007; Roeder, 2005; Sikorski and Buratowski, 2009).

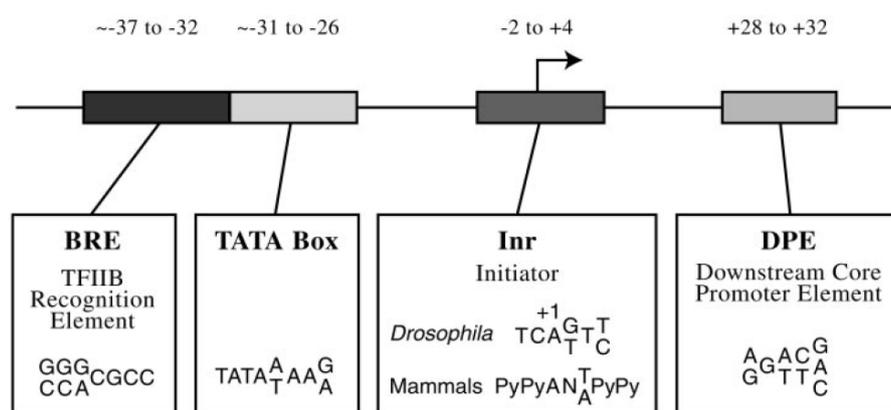


Figure 21. Eukaryotic core promoter motifs. Consensus sequences are shown for each element. The DPE consensus was determined with *Drosophila*. This diagram depicts some elements that can contribute to basal transcription from a core promoter. A particular core promoter may contain some, all, or none of these elements. Adapted from (Smale and Kadonaga, 2003)

Core promoter contains variant elements (Smale and Kadonaga, 2003) (figure 21). The first identified element is the TATA box, an A/T rich sequence which is 25-30bp upstream of the transcription start site. TATA box has a consensus sequence TATA(A/T)AA(G/A) and can be recognized by TATA-binding protein(TBP), a subunit of TFIID. TBP-TATA structure reveals that two quasi-symmetrical domains fold like a saddle, sitting in the minor groove of the duplex DNA through hydrophobic interactions(Nikolov, *et al.*, 1996). TBP also induces kinks on DNA and partially unwinds the duplex through the insertion of Phe residues. The second element is the transcription start site (Inr), most containing an A at the start site (+1) and a C at the -1 position within a consensus sequence PyPyAN(T/A)PyPy. Inr is recognized by TATA-associated-factor (TAF) 1 and TAF2 which are also subunits of TFIID. The third element is the downstream promoter element (DPE) which locates at +28 to +32 downstream of the start site. DPE is required for TATA-less promoters and acts in conjunction with Inr. Similarly, DPE is also recognized by TAF6 and TAF9 of TFIID. The fourth element is TFIIB-recognize-element (BRE). Structure of TFIIB-TBP-TATA complex reveals that TFIIB interacts with a major groove upstream of TATA box and a minor groove downstream of the TATA box(Nikolov, *et al.*, 1995). Other elements include proximal sequence element (PSE) (for snRNA transcription), downstream core element (DCE), motif ten element (MTE), also contribute to the core promoter activity such as promoter strength, efficiency of PIC assembly, or targeting regions for enhancer activity. GTFs contain TFIIA, IIB, IID, IIE, IIF, and IIH. TFIIIE and TFIIF are essential for promoter melting and the transition from initiation to elongation. TFIIF also phosphorylates pol II CTD. After TFIID binds the core promoter, PIC begins to form either through a sequential pathway or via a preassembled pol II holoenzyme pathway. For the first one, the order is: TFIID first binds to the promoter, then TFIIA and TFIIB stabilize TFIID-DNA binding, and then preassembled pol II and TFIIF are recruited by TFIIB, and finally, TFIIIE and TFIIF bind. The other pathway was discovered when purifying pol II. Pol II was

found to be purified as a preassembled holoenzyme containing pol II, a subset of GTFs, some components of multisubunit mediator such as SRB or Mediator, and other proteins involved in chromatin remodeling, DNA repair and mRNA processing (Chao, *et al.*, 1996; Maldonado, *et al.*, 1996; Ossipow, *et al.*, 1995; Wilson, *et al.*, 1996). Both pathways may exist *in vivo* (Thomas and Chiang, 2006).

Transcription cofactors include four classes: TAFs of TFIID, Mediator, upstream stimulatory activity (USA)-positive/negative cofactors, chromatin remodeling and histone modifying enzymes. TAFs act as coactivators bridging the gene-specific activator and basal transcription machinery. Mediator also transmits regulator signals. In yeast, Mediator is an at least 1MDa complex consisting of about 20 polypeptides with an extended shape which can be divided to three domains: head which binds pol II, middle, and tail which binds activators. Upon forming a complex with activators, these domains can be adapt to distinct conformation and the conformational shifts control PIC structure and function (Taatjes, 2010; Thomas and Chiang, 2006). USA includes positive factors: PC1-4 and a negative factor NC1. They have a dual role: in the presence of activators, as a coactivator, either through enhancing PIC assembly, or through interactions bridging activator and basal transcription machinery; in the absence of activators, as a repressor inhibiting basal transcription. Besides bridging function, another crucial mechanism for these three classes of cofactors is posttranslational modification on transcriptional components, such as phosphorylation, acetylation, glycosylation, ubiquitination and poly(ADP-ribosyl)ation. Dependant on the targets and consequences of these activities, cofactors play different regulatory functions. Besides these PIC formation/function regulatory cofactors, activators also recruit the forth class cofactors: chromatin remodeling cofactors, such as Swi/Snf, RSC, modification cofactors such as HATs (Li, *et al.*, 2007), since the chromatin structure is an obstacle to transcription. These cofactors either utilize ATP-hydrolysis or directly modify histones to alter histone-DNA interactions,

leading to histone eviction and making the nucleosomal DNA more accessible to GTFs.

1.2 Transcription initiation in archaea

1.2.1 Promoter architecture

Cell-free systems, analyzing tRNA promoter of *Methanococcus vannielii*, rRNA promoter of *Sulfolobus shibatae* and GDH promoter of *Pyrococcus furiosus* have been used to identify archaeal promoter signals *in vitro* (Hain, *et al.*, 1992; Hausner, *et al.*, 1991; Hethke, *et al.*, 1996; Reiter, *et al.*, 1990). *In vivo* analysis on tRNA promoter of *Haloferas volcanii* was also reported (Palmer and Daniels, 1995). Archaeal promoters contain several elements: two major elements, TATA-box and TFB recognition element (BRE) and another two less-understood elements: initiator element (IE) and promoter proximal element (PPE) (figure 22).

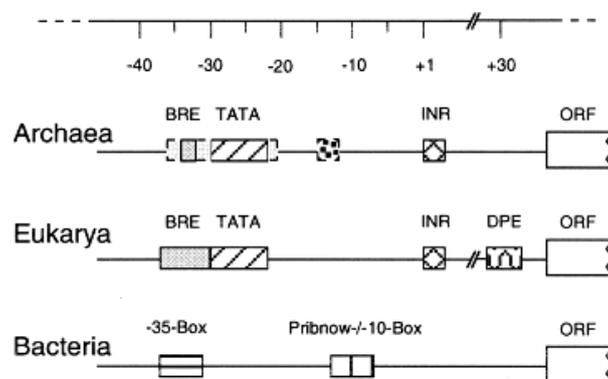


Figure 22. Comparison of promoter elements from archaea, eukarya, bacteria. The promoter elements are shown aligned to the scale shown on the top. Inr is initiator element(IE). Adapted from(Soppa, 1999).

TATA-box, also called box A, is an AT rich 8bp sequence, locating approximately 25nt upstream of the transcription start site, and is essential for transcription initiation. Deletions, or single point mutations in this region abolished or dramatically reduced initiation efficiency. TATA-box also strictly defines the start site to 22 to 27bp downstream. TATA box doesn't have a strict sequence requirement, and its consensus sequences are different in archaeal

subsets (Soppa, 1999). BRE is a purine rich 7bp sequence upstream of the TATA box. BRE increases promoter strength by recruiting TFB to stabilize the interactions between TBP and promoter (Qureshi and Jackson, 1998), and defines the direction of transcription (Bell, Kosa, *et al.*, 1999). Initiator element (IE) contains a pyrimidine-purine dinucleotide near the start site, with initiation at the purine, usually a G. As a special case, IE in *S.shibatae* viral T6 promoter contributes to the promoter strength and start site selection as well (Qureshi, 2006). However, in other promoters, TATA box and BRE are the dominant elements. PPE comprises A/T rich sequences, locating around -10. Mutations in this region affected initiation at *S.shibatae* rRNA promoter, but it's not clear which initiation functions were affected (Hain, *et al.*, 1992). Generally, TATA box and BRE are the major promoter elements in archaea, determining the promoter activity and transcription orientation, while IE and PPE are only used in special subsets.

1.2.2 Basic transcription initiation factors

1.2.2.1 TBP

Archaeal TATA-binding protein (TBP) recognizes TATA box and is a homolog of eukaryotic TBP. Most archaeal species have only one TBP while in *Halobacterium* NRC-1, six TBP homologs and seven TFB homologs are found by genome sequencing (Baliga, *et al.*, 2000). Structures of TBP reveal a similar conformation with eukaryotic TBP (DeDecker, *et al.*, 1996; Kosa, *et al.*, 1997; Littlefield, *et al.*, 1999) (figure 23). A conserved core contains a direct repeat sequence, with 36%-41% similarity to eukaryotic TBP, but forming a more symmetrical saddle shape conformation. The N-terminus extends by four to five residues, and the C-terminus has a highly acidic tail in some archaeal species. TBP binds to the minor groove of TATA box, inducing a distortion of DNA about 65°. Two pairs Phe also penetrate between the first two and last two base pairs, inducing two kinks at either end of TATA element. Most hydrophobic interactions are preserved between archaeal TBP and TATA box.

However, eukaryotic TBP forms many salt bridges, water-mediated hydrogen bonds with backbone atoms while in archaea these interactions are replaced by van der Waals contacts with the sugar, which explains high stability of archaeal complex in elevated salt concentration and temperature (DeDecker, *et al.*, 1996; Kim, *et al.*, 1993).

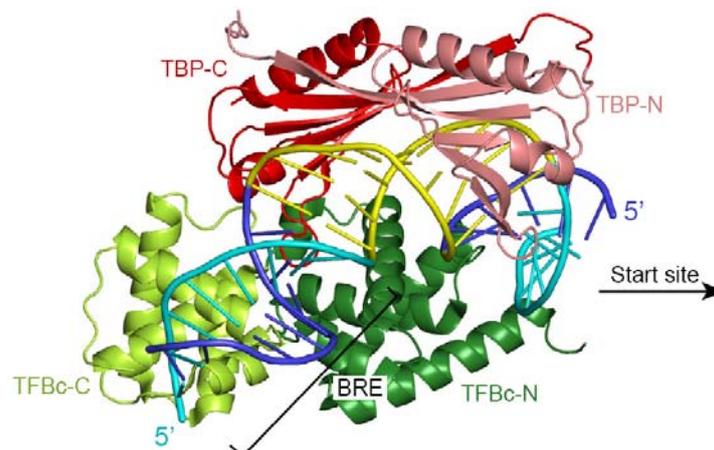


Figure 23. Crystal structure of TBP-TFBC-DNA containing TATA box-BRE ternary complex. Non-template DNA is colored in cyan and template DNA in blue. TATA element is colored in yellow. TBP is shown as ribbon colored in pink and red for N-, C-terminal domain respectively. TFBC is shown as ribbon colored in dark green and light green for N-, C-terminal domain respectively. PDB code is 1D3U.

1.2.2.2 TFB

The BRE element recognizing factor is TFB, which can be divided into two regions: N-terminal region containing a zinc ribbon motif, a conserved B-finger motif with a linker following; C-terminal direct repeats forming a globular domain. TFB_N is required for RNAP recruitment and NMR structure of the zinc ribbon is solved (Zhu, *et al.*, 1996). Photocrosslinking shows that TFB interacts from BRE to transcription start site, and since TFB_C binds only around TATA box, the downstream interactions which exist or strengthen when RNAP is in the complex, are predicted to be from TFB_N (Renfrow, *et al.*, 2004). Although TFB is shown to have weak interactions with subunit K (Rpb6 in pol II), reconstituted RNAP lacking K is functional in factor-dependent transcription initiation. EMSA shows that outside surface of BDLNP subcomplex has the important interaction site of TFB, and different from TFIIB, archaeal dock

domain is not essential for TFB binding (Goede, *et al.*, 2006; Werner and Weinzierl, 2005). However, unlike pol II whose recruitment to TFIIB is facilitated by TFIIF, archaeal pol has only TFB. Thus it's a difficult and interesting question that how TFB efficiently recruits archaeal pol. B-finger motif is indicated to be important to the late stages of initiation, probably during promoter clearance (Bell and Jackson, 2000). The structure of the TFB_C-TBP-DNA complex is available (Littlefield, *et al.*, 1999) (figure 23). The structure reveals that TFB_C binds the major groove of the BRE element sequence-specifically through a C-end HTH motif, C-terminal TBP, and also nonspecifically binds the minor groove downstream of TATA box, therefore this asymmetric contact determines the direction of transcription (Bell, Kosa, *et al.*, 1999). The structure of the TFB homolog in eukarya, TFIIB complexed with pol II is solved recently (Kostrewa, *et al.*, 2009). It indicates that the C-terminal globular domain binds the pol II "wall"; DNA is opened with the help of the linker after B-finger motif; B-finger approaches the active center and helps scan the transcription start site (figure 24).

1.2.2.3 TFE

TFE is the third basal archaeal transcription factor. Different from eukaryotic TFIIE which is a heterotetramer formed by two subunits, α and β , archaeal TFE contains a homolog of only the N-terminal half of TFIIE subunit α , having a N-terminal helix-turn-helix motif within a Leu rich region, and a C-terminal zinc finger which is essential for the function (Hanzelka, *et al.*, 2001). Archaeal TFE is not strictly essential for transcription initiation *in vitro*, but stimulates transcription activity by 2-3 fold on weak promoters, through increasing the sub-optimal interactions between TBP and TATA box (Bell, *et al.*, 2001). TFE also stabilizes open promoter formation through binding to the non-template strand upstream of the open bubble, in a manner dependent on subunit E'/F (Grunberg, *et al.*, 2007; Micorescu, *et al.*, 2008; Ouhammouch, *et al.*, 2004). Structure of TFE N-terminal winged helix-turn-helix has been solved, revealing

a basic and aromatic wing which may play a functional role, and two hydrophobic patches which may serve as protein interaction sites (Meinhart, *et al.*, 2003). Although the localization of TFE on archaeal pol is not clear, photocrosslinking and hydroxyl-radical cleavage assay indicate that its eukaryotic homolog TFIIE positions near Rpb1 clamp domain (Chen, *et al.*, 2007), which is consistent with the result that N-terminal TFE is crosslinked to the non-template DNA at position -9 to -11 (Grunberg, *et al.*, 2007)(figure 24). Further structural study is required to reveal the accurate position of TFE and the interactions with archaeal polymerase.

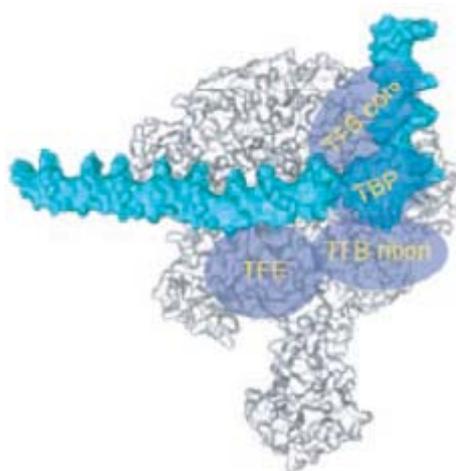


Figure 24. Archaeal PIC components. Promoter DNA is colored as cyan. Positions of general transcription factors are represented as ellipses. Adapted from (Jun, *et al.*, 2011).

1.2.3 Transcription initiation regulation

Except TBP, TFB and TFE, archaeal genomes don't have homologs of any other eukaryotic general initiation factors. During the initiation phase, TBP nucleates the formation of TBP-TFB-DNA ternary complex, and in some cases TFE is also included. Then, N-terminal TFB recruits and correctly positions RNAP, which can be facilitated by TFE. With the contribution of TFB-B-linker region, E' subunit and TFE, DNA duplex is melted from approximately -10 to +3, and the template strand slides into the DNA binding channel and the active site.

Subsequently, TFB B-finger and RNAP scan for the correct initiation sites, and transcription initiates.

Although archaeal transcription machinery resembles that of eukarya, transcription regulation utilizes pathways in both bacteria and eukarya.

Bacteria-pathway is the main regulation in archaea. Most archaeal genomes contain homologs of bacterial regulatory protein, who negatively regulate through competitively binding to the promoter and preventing the recruitment of initiation factors or RNAP; who can also positively regulate through facilitating the PIC formation. For example, metal-dependent repressor (MDR1) and Lrp A bind overlapping the start site, thus blocking the RNAP recruitment (Bell, Cairns, *et al.*, 1999; Dahlke and Thomm, 2002); Lrp14, who binds the TATA/BRE region, inhibits the formation of TBP-TFB-DNA ternary complex (Bell and Jackson, 2000). Activators bind upstream or downstream of the promoter region (Ouhammouch, 2004).

Like eukarya, archaea do have histones in *Euryarchaea* and some nucleoid DNA-binding protein. Archaeal histones don't have N-tails which are the targets of histone modification, and archaeal genomes don't encode histone modification/modeling machinery as well. However, archaeal histones can form both heterodimers and homodimers, which have different affinity to DNA sequences. Therefore, regulating the combination of histone proteins could control the accessibility of initiation factors to promoters (Reeve, 2003). Additionally, a positively charged nucleoid protein Alba (Acetylation lowers binding affinity) binding minor grooves of DNA can be (de)acetylated, and it may resemble the origin of the eukaryotic histone (Reeve, 2003).

Another pathway utilizes the multiplicity of initiation factors. Many archaeal genomes encodes at least two TBPs or TFBs, some of which have more (Baliga, *et al.*, 2000). Like bacteria which use different σ factors for recognizing different promoters, this multiplicity may also regulate transcription under certain conditions.

1.3 Aim of this study

Regulation of gene expression happens at many stages from transcription initiation to protein degradation. Transcription initiation is an important step for regulation. Elucidation of the structure of the initiation machinery becomes crucial to understand its regulation. In eukarya, transcription initiation involves basal transcription machinery which contains a series of initiation factors. Because of the complexity, it is difficult to investigate the eukaryotic initiation mechanism. Archaea possesses a rather similar RNA polymerase as eukarya, and is also thought to mandate a similar initiation mechanism (Bell and Jackson, 1998). Because archaeal preinitiation complex contains mainly just two factors, TBP and TFB, studies on archaeal PIC will provide a wealth of data on the mechanism of initiation. The aim of this work is to determine the structure of *Pyrococcus furiosus* transcription preinitiation complex (PIC) by crystallography. Previous work obtained needle shaped PIC crystals with bad diffraction (9 to 10 Å). EM revealed very weak density at proposed regions of TBP and TFB indicating their high flexibility in PIC. Therefore, this work attempts to improve the stability of PIC complex and facilitate crystallization. New crystals having hexagonal shape were achieved by utilizing initiation factor variants, changing nucleic acid substrate and crystallization conditions. However, they diffracted similar as before indicating that there are still quite flexible and not well ordered.

2. Results

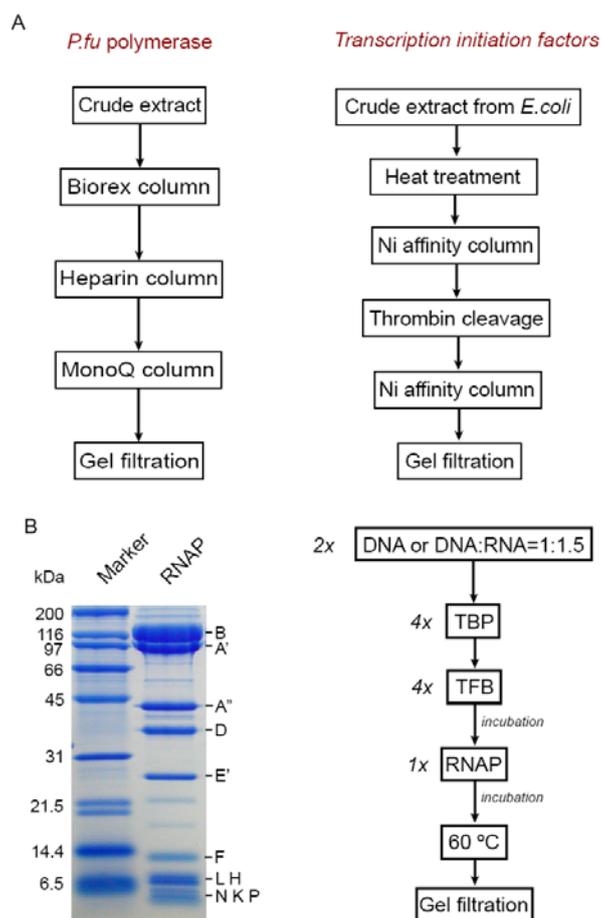
2.1 Assembly of archaeal PIC

In this work, I used PIC from *Pyrococcus furiosus*. Endogenous *P.fu* polymerase was purified from *P.fu* cells. Recombinant TFB, TBP and their variants were expressed and purified from *E.coli*. The overall purification and assembly scheme of the various subunits is summarized in figure 25. *P.fu* RNAP was purified by four different chromatography steps: a weak cation-exchange column (Bioresx70), an affinity column (Heparin), a strong anion-exchange column (MonoQ), and a size-exclusion column (Superose 6). Yield of purified endogenous *P.fu* RNAP was 0.05~0.1mg protein/g cells. Recombinant initiation factors were overexpressed as fusion proteins carrying a His₆ tag at the N-terminus and were purified by Ni-affinity chromatography. Thrombin was used to cleave N-terminal his-tag to maintain more natural interactions with RNAP after assembly. A second Ni-affinity chromatography and subsequent size-exclusion chromatography removed uncleaved factors and other contaminants. Properties of the *P.fu* RNAP subunits and initiation factors are listed in Table 3. RNA polymerase, nucleic acids substrate, initiation factors were assembled with molar ratio 1:2:4. TBP who recognizes TATA box was added first, then TFB who interacts with TBP and BRE was added subsequently, because without TBP, TFB can't form a complex with DNA(Bell and Jackson, 2000). Purified RNAP, whose recruitment requires TFB, was then added. After incubation, proteins were heated in 60°C to activate the PIC complex and form an open transcription bubble. After gel filtration, excess of initiation factors and nucleic acid substrates are removed.

Table 3. Properties of the ORFs that encode *P.fu* RNAP subunits and initiation factors.

RNAP SU	Protein length (amino acids)	Protein mass (kDa)	pI
B	1117	127.0	6.7
A'	910	103.4	6.8
A''	397	44.4	5.8
D	261	29.8	4.7
E'	189	21.7	7.7
F	120	14.1	4.6
L	95	11.1	5.1
H	82	9.2	9.7
N	70	8.2	6.5
K	57	6.2	10.3
P	49	5.8	10.7

TFs			
TBP	191	21.3	5.0
TFB	300	34.1	9.5

**Figure 25. A. A schematic representation of *P.fu* polymerase and transcription factors purification. B. SDS-PAGE of commassie stained RNAP. Subunits were confirmed by MALDI-TOF. Right panel shows schematic representation of the assembly experiment.**

2.2 A tailed scaffold with RNA improves formation of single crystals

RNA is proved to form a hybrid with DNA and induce the ordering of several flexible regions of polymerase and formation of a highly stable complex (Cramer, *et al.*, 2001; Gnatt, *et al.*, 2001). In order to get a more stable complex with consistent conformation, I attempted to induct RNA to the PIC. In the elongation complex, the ordered RNA was observed as 9 bases (Gnatt, *et al.*, 2001). Biochemical results demonstrates that 8 bp DNA-RNA hybrid is necessary for forming a stable pol II elongation complex (Kireeva, *et al.*, 2000). Therefore, 8 bases RNA were designed complementary to DNA from -7 to +1. Once 8 bp hybrid forms, it was shown to clash with the B-finger of TFIIB, and trigger the release of TFIIB from pol II, which contributes to the promoter escape (Bushnell, *et al.*, 2004). Thus corresponding B-finger of *P.fu* TFB was truncated from residue 43 to 62 in this work (figure 26). Non-template DNA is mobile and can't observed in the pol II elongation structure (Kettenberger, *et al.*, 2004). Therefore, in designed "tail" scaffold, the non-template DNA was also truncated (figure 27).

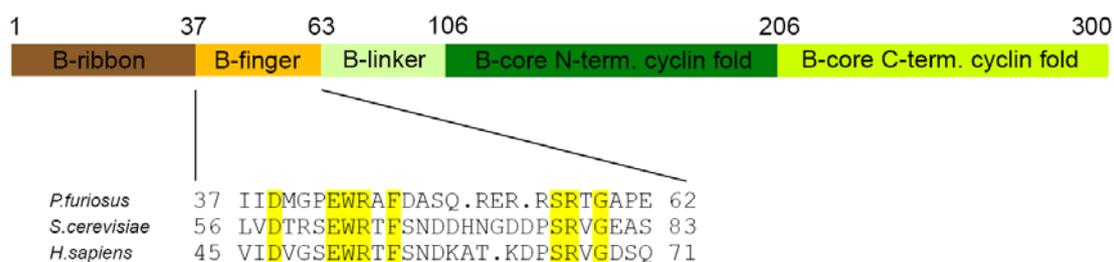


Figure 26. *P.fu* TBP domain organization and sequence alignment in the B-finger region from *P.furiosus*, *S.cerevisiae* and *H.sapiens*. Yellow highlighting indicates conserved residues.

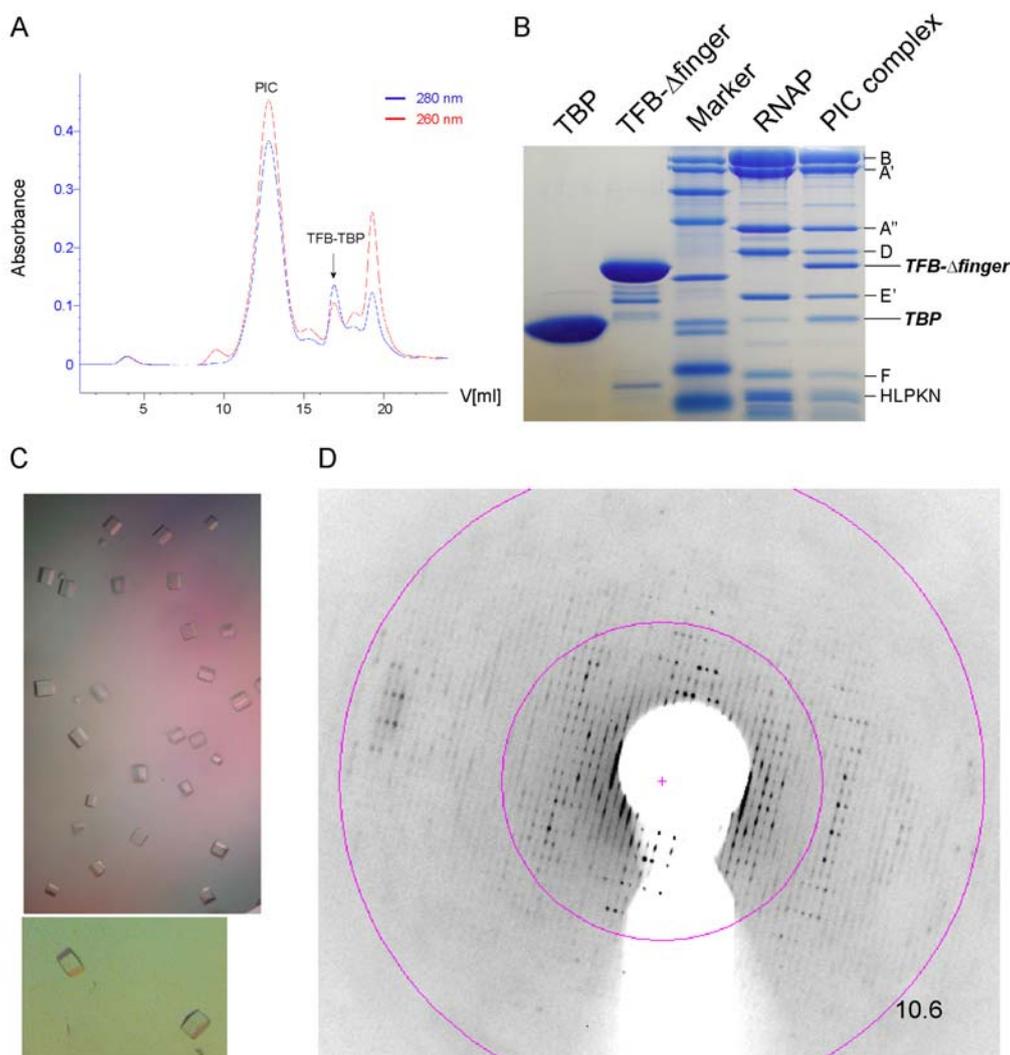


Figure 28. A. Size-exclusion chromatography profile of assembled PIC complex. B. SDS-PAGE analysis shown PIC components: TBP, TFB- Δ finger, RNAP. C. PIC Crystals. D. Diffraction pattern of PIC crystals.

After annealing, RNA-DNA substrate was assembled with TFB- Δ finger, TBPwt forming an artificial PIC (figure 28). PIC complexes with different substrates (natural closed, bubble, fork, tail) were purified and concentrated to 4 mg/ml, then were screened using commercial conditions including Nextal Classic, MPD, Anions and Cations Suite (Qiagen), Hampton Index, Natrix, PEG/Ion screens (Hampton Research), Jena JBScreen classic HTS 1 and 2 (Jena Bioscience) performed with Hydra II semi-automatic protein crystallization robot (Matrix Technologies Apogent Discoveries) and polymerase screens from which pol II crystals were achieved successfully, performed manually.

However, using commercial conditions, only a few conditions yielded microcrystals which can't be optimized manually. Using polymerase screens, PIC only with the "tail" type scaffold achieved nice single crystals (figure 28). Using "tail" 6 and 7 with incomplete BRE, no crystal could be obtained, indicating the importance of keeping interactions between BRE and TFB. "Tail" 5 led to the best crystals which might be because of additional contacts between the extended downstream template DNA and the polymerase. However, although the morphology of these crystals changed dramatically and looked promising, they were quite small (maximum size $\approx 100\mu\text{m} \times 100\mu\text{m} \times 100\mu\text{m}$) and still diffracted badly (figure 28), indicating that the molecules were still not well organized in the crystals.

2.3 Truncation of the acidic C-terminus of TBP to improve crystals

Since *P.fu* polymerase alone didn't yield any crystals, other components in the PIC such as TBP/TFB/nucleic acids, must be involved for the crystal packing. Thus slight alteration of these components may dramatically affect the packing of molecules. As mentioned in the introduction, *P.fu* TBP contains a highly mobile C-terminal acidic tail formed by six continuous Glu and several Leu, which is conserved in most archaeal TBP but absent in eukaryotic TBP (DeDecker, *et al.*, 1996) (figure 29). Since DNA is negatively charged, this mobile tail may affect the conformation of nucleic acids which leads to conformational heterogeneity in the crystal. Indeed, it's demonstrated that for TBP-TFB-DNA ternary complex, truncation of this tail is crucial for the high quality of crystals (Kosa, *et al.*, 1997; Littlefield, *et al.*, 1999). Therefore, I attempted to reduce this potential impact by truncating the TBP acidic tail as well. Assembly using TBP- Δ tail which removed residues 182-191, TFB- Δ finger, RNAP, DNA-RNA led to nice chromatography profile and crystals with similar morphology (figure 30). However, although the spots became clearer, the resolution is still very low.

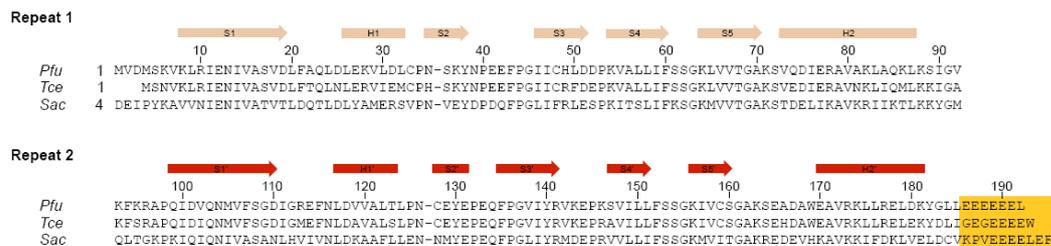


Figure 29. Sequences alignment of archaeal TBPs from *Pyrococcus furiosus*, *Thermococcus celer*, *Sulfolobus acidocaldarius*. Numbering and secondary structure (S = beta sheet, H = alpha helix) labeled according to the *P.woesei* TBP structure (DeDecker, *et al.*, 1996). Conserved C-terminal acidic tails are boxed in orange.

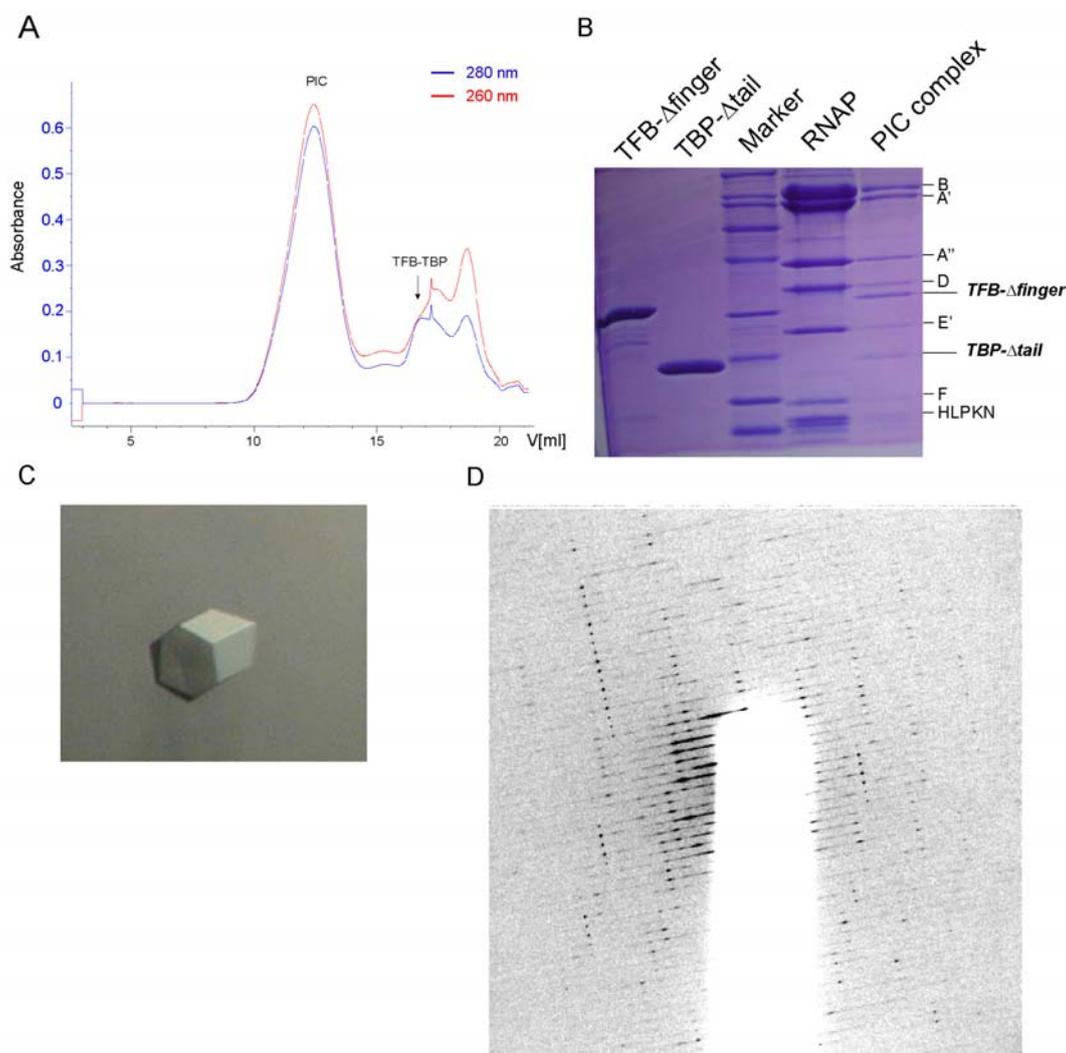


Figure 30. A. Size-exclusion chromatography profile of assembled PIC complex. B. SDS-PAGE analysis showed modified PIC components: TBP- Δ tail, TFB- Δ finger, RNAP. C. PIC Crystals with size $\approx 130\mu\text{m} \times 130\mu\text{m} \times 130\mu\text{m}$. D. Diffraction pattern of PIC crystals. Resolution is approximately 10Å.

2.4 TFE did not help the crystallization on bubble scaffold

Diffraction pattern showed low resolution as well as smearing spots, indicating the complex is still flexible in some regions which can't be improved by inducing DNA-RNA hybrids or TBP/TFB modifications. Since TFE can stabilize PIC by increasing TBP-TATA-box interaction and stabilizing the open promoter through binding to the non-template strand (Bell, *et al.*, 2001; Grunberg, *et al.*, 2007), it was used for assembly and crystallization in this work. Although full-length TFE alone can't be crystallized successfully (Meinhart, *et al.*, 2003), interactions with RNAP and nucleic acids may induce the ordering of its C-terminal structure. Additionally, C-terminal zinc finger is crucial for the function of TFE (Hanzelka, *et al.*, 2001). Therefore, in the PIC crystallization both N- and C-terminal domains were kept. "Bubble" scaffolds without RNA were used to keep the interactions between TFE and non-template strand. Since the stable interaction between RNAP with template DNA is strictly dependent on the B-finger of TFB when RNA is absent (Werner and Weinzierl, 2005). Therefore TFBwt and TBPwt were chosen to maintain potential contacts. The complex was screened using commercial conditions including Nextal Classic (Qiagen), Hampton PEG/Ion screens (Hampton Research) and polymerase screens. However, this more natural PIC didn't yield any crystals.

2.5 Trials using a histone promoter

From previous and current work on this project, promoter character was found to be crucial for the crystallization. Using GDH promoter, although "tail" scaffold was found to improve the crystallization, poor diffraction property suggested that slight modification of the scaffold might not be able to affect the crystal property fundamentally. Thus another histone promoter was used for PIC assembly (figure 31). "Fork" and "tail" types of scaffolds were designed. However, using this promoter crystals didn't form in previous crystallization conditions.

compounds were also used: $(\text{NH}_4)_{14}[\text{NaP}_5\text{W}_{30}\text{O}_{10}]$, W_6Br_{12} , $\text{Ta}_6\text{Br}_{12}^{2+}$. For PIC crystals achieved in this work, ethylene glycol yielded best protection effect while others led to phase separation or crystal cracking. These soaking techniques however didn't improve the diffraction quality.

Cross-linking agent glutaraldehyde (GA) can react with ϵ -amino groups of lysine residues through aldehyde groups and covalently crosslink them both inter- and intra-molecularly. Adjacent macromolecules in the lattice are crosslinked and then crystals can be stabilized. Therefore crosslinking using glutaraldehyde has been revived as a useful procedure. PIC crystals were crosslinked using three different concentrations including 0.001%, 0.01%, 0.1% GA and crosslinked for 2 or 3 hrs (Jacobson, *et al.*, 1996). Higher concentration or longer time would lead to yellow-colored crystals. Crosslinked crystals yielded worse diffraction which might be because of excessive crosslinking although the concentration of GA was already very low.

Crystal annealing techniques were also used to reduce potential mosaic spread caused by flash-cooling. Macromolecular crystal annealing (MCA), flash-annealing (FA) method and annealing on the loop (AL) were also tried (Heras and Martin, 2005), but none of these techniques led to better diffraction.

3. Discussion and outlook

Photocrosslinking and X-ray structure of subset of eukaryotic and archaeal PIC reflects the general conserved geometry of interaction with transcribed DNA (Bartlett, *et al.*, 2004; Chen and Hahn, 2004; Kostrewa, *et al.*, 2009; Littlefield, *et al.*, 1999; Nikolov, *et al.*, 1995) and suggests the position of initiation factors and DNA in PIC (figure 32). In eukaryotic PIC, TFIIB zinc ribbon was shown binding to a concave formed by Rpb1 dock, clamp domain and Rpb2 wall (Kostrewa, *et al.*, 2009) which however might be a modified case for archaeal TFB (Goede, *et al.*, 2006). Subcomplex of RNAP subunits B-D-L-N-P binds TFB efficiently indicating that interaction with dock domain is not essential for archaeal TFB. In fact, archaeal TFB N-terminal domain recruits RNAP itself without the facility of other factors such as TFIIF in eukarya (Bell and Jackson, 2000) suggesting a different arrangement of TFB. TFIIB then extends from the RNA-exit channel to DNA-RNA hybrid binding site and active center by B-finger motif, then to the rudder and clamp coiled-coil by B-linker motif. Density in these regions was also weak because of their flexible nature for their functions. These motifs in archaeal TFB might position similarly since they show homologous functions (Kostrewa, *et al.*, 2009). TFIIB C-terminal contains two cyclin domains which contacts the polymerase “wall” and is suggested to have only three salt bridges between Helix2 from Cyclin domain1 and the Rpb2 wall, leading to weak density of N-terminal cyclin domain. Additionally, non density of C-terminal cyclin domain could be observed, indicating weak or no interactions between this domain and polymerase and it may be highly flexible (Kostrewa, *et al.*, 2009). Supposed interacting residues in TFB and RNAP are conserved in archaea (Hirata, *et al.*, 2008), indicating similar interactions in archaea. Archaeal TBP binds to DNA and TFB and can't form a complex with RNAP alone (Bell and Jackson, 2000). Therefore, in archaeal PIC, specific positions of TBP, TFB-C-terminal-cyclin-domain and upstream DNA may be influenced by

tenuous interaction network between DNA duplex and polymerase subunits Rpb1/Rpb2/Rpb5, DNA duplex and TFB-N-cyclin-domain and linker, TFB-N-cyclin-domain and polymerase.

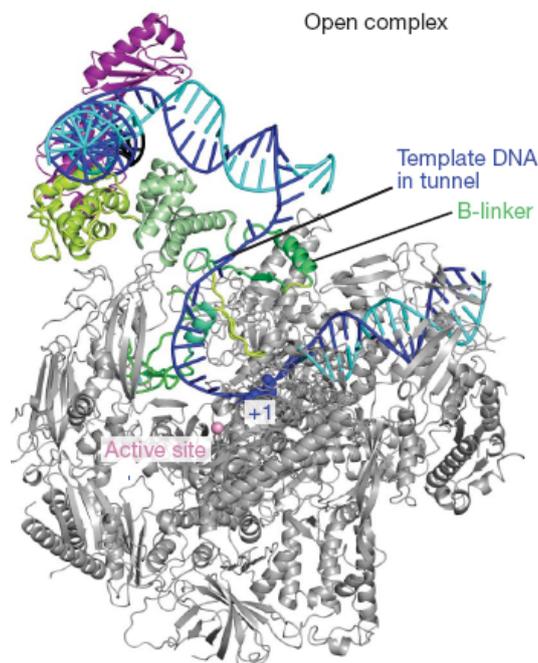


Figure 32. Models of opened initiation complex showing the position of initiation factors and DNA strands. DNA template and non-template strands are in blue and cyan, respectively. The TATA element is in black. Views are from the side. Adapted from (Kostrewa, *et al.*, 2009).

Since the formation of transcription initiation complex is a transient, unstable stage, this would hamper the crystallization. DNA is predicted to have crucial contribution both to the stability of PIC complex and the specific positions of initiation factors. Consistently, previous work tested large amount of DNA scaffold with various length and bubble style, and demonstrated the important contribution of DNA residues to the crystallization. In the work of this thesis, nucleic acid substrates were also shown to severely affect the formation and quality of crystals. Many crystallization conditions were applied for three kinds of DNA-RNA hybrid scaffolds shown in figure 27. Replacing “bubble” or “Fork” type scaffolds with “tail” effectively allowed the formation of reproducible and nice single crystals which could be because of the deletion of flexible non-template strand in the vicinity of B subunit of polymerase that might

impede the crystal packing (Bartlett, *et al.*, 2004). Truncation of only two base pairs in BRE completely prevented the formation of crystals. This should be because, the second G in the BRE can form a hydrogen bond with the TFB helix-turn-helix motif and is essential for TFB binding (Littlefield, *et al.*, 1999). Deletion of this nucleotide led to the disassociation of TFB from the promoter and subsequent disassembly of PIC complex. Several strategies were occupied jointly to improve the crystallization of *P.fu* PIC complex including the induction of RNA which forms a hybrid with DNA to stabilize conformation of polymerase, and modification on TBP acidic tail which could affect the crystal ordering potentially. New crystals were achieved with a different form indicating a hexagonal unit cell (figure 28, 30) and provided a promising start point for obtaining better quality crystals. Diffraction property of these crystals were still not good. Since the induction of RNA is able to stabilize DNA template and the conformation of polymerase, I speculate that this is because of an unfixed location of TBP and/or TFB-C-terminal cyclin domain and they may not be involved in the intermolecular interaction in the crystals. In some trials TFE was crystallized together with PIC with “bubble” scaffold, because of its function of binding to non-template strand and stabilizing the complex. No crystals were achieved in the two commercial conditions and the conditions from which “tail” shape crystals grown, although it might be because of the negative effect of non-template strand in the “Bubble” scaffold. Thus TFE didn't facilitate the crystallization in these conditions. However, TFE interacts with both RNAP and TBP, enhancing the TBP-DNA binding (Bell, *et al.*, 2001), so it might be helpful to fix the localization of TBP. Therefore, it's still worthwhile to screen for new crystal forms using PIC containing TFE in future. Using histone promoter as the substrate, although the scaffold was designed as a “tail” shape that is the same as that of GDH promoter, it led to a dramatic difference and no crystals formed using conditions for the GDH promoter. Since histone promoter is also a strong promoter which binds TBP and TFB efficiently, their difference at region from -15 to -12 might impede the crystallization. DNA

opening site in PIC is from -9 to +5 (Spitalny and Thomm, 2003), thus heat treatment might cause over-melting of this region when activation of the PIC complex, since it's AT rich in histone promoter, leading to weaken duplex interactions. Post-crystallization annealing operation didn't help the diffraction property, indicating that the poor resolution was not the result of cooling process.

In this work, the length of RNA for stabilizing PIC designed as 8 nt and B-finger deletion strategy were based on the initial structure of pol II-TFIIB which determined B-finger as a loop shown to clash with RNA beyond 5 nt (Bushnell, *et al.*, 2004) and biochemical analysis on stability of elongation complex with different RNA lengths (Kireeva, *et al.*, 2000). Recent structures indicate that when the +1 nucleotide positions at the pre-translocation site, -7 nt just adjacent to the B-finger, at the border of clashing with it (Kostrewa, *et al.*, 2009). Therefore, in future work, 7nt or shorter RNA for example 5 nt could be designed which can bind and stabilize the B-finger. In this case, TFB wt should be used. Co-crystallization of PIC-TFE with/without RNA would also be promising to yield crystals with better quality. But abundant screening for crystallization conditions will be necessary for new crystal form. PIC-TFE complex will also be a good choice for EM study. Although recently an EM structure of *P.fu* PIC was published, their resolution is extremely low (25Å) and some extra density suggest the rough positions of TFB/TBP which is however too weak to fit the X-ray structures and demonstrate the correct positions, consistent with their flexibility (De Carlo, *et al.*, 2010). Therefore, TFE might facilitate the EM analysis. In addition, it would be interesting to crystallize TFB N-terminal domain together with RNAP to investigate how TFIIB recruits polymerase independently.

4. Experiment procedures

4.1 Oligonucleotides and cloning

All oligonucleotides listed in figure 27 were purchased from Biomers.net (Ulm, Germany) with HPLC purified. Primers for cloning recombinant TFB and TBP mutants were purchased from Thermo Electron Corporation.

Primers for cloning TFB- Δ finger (delete residue 43-62):

Fwd: 5' -GAACATAATTGATATGGGTCCTAGTATTCTTCTTCATGACAAG-3'

Rev: 5' -CTTGTCATGAAGAAGAATACTAGGACCCATATCAATTATGTTC-3'

Primers for cloning TBP- Δ tail (delete residue 182-191):

Rev: 5' -CGACGGAGCTCGAATTCTCAATACTTATCCAATTCTCTCAAC-3'

TFB- Δ finger was constructed by two-PCR techniques from plasmids carrying wild type TFB gene used in previous work. In round 1, T7 primer (forward or reverse) and TFB variant primer (Rev or Fwd) flanking the internal deletion sequence were used in two separate reactions. PCR reaction generated two PCR products carrying the sequence flanking the deletion region. Purified products were added jointly in round 2. T7 primers (forward and reverse) were used. PCR product was then digested by restriction endonucleases NdeI and BamHI and inserted into the pET 28b(+) expression vector (Novagen) to generate a fusion protein with an N-terminal hexahistidine tag. Plasmid was verified by sequencing.

To construct TBP- Δ tail, T7 forward and TBP Rev primers were used for PCR. Plasmid carrying wild type TBP gene was used as the template. PCR product was then digested by restriction endonucleases NdeI and XhoI and inserted into the pET 28b(+) expression vector (Novagen). Plasmid was verified by sequencing.

4.2 Expression and purification of transcription initiation factors

4.2.1 Expression and purification of TFB wt and Δ finger

4.2.1.1 Buffers

Lysis buffer:

50 mM Tris-HCl pH 8.0
300 mM NaCl
10% glycerol
1x Protease Inhibitor

Ni-NTA binding buffer:

50 mM Tris-HCl pH 8.0
300 mM NaCl
40 mM imidazole
10% glycerol

Ni-NTA washing buffer:

50 mM Tris-HCl pH 8.0
2M NaCl
40 mM imidazole
10% glycerol

Ni-NTA elution buffer:

50 mM Tris-HCl pH 8.0
300 mM NaCl
300 mM imidazole
10% glycerol

Dilution buffer:

50 mM Tris-HCl pH 8.0
300 mM NaCl

Gel filtration buffer:

20mM Hepes-Na pH 7.0
400mM KCl
10% Glycerol

4.2.1.2 Expression and purification procedures

Inoculate 25ml LB in the presence of antibiotics (Kanamycin and Chloramphenicol) with one fresh clone and shake at 37°C overnight. Then dilute preculture in 2l LB with Kanamycin and Chloramphenicol and shake at 37°C until OD₆₀₀=0.6~0.7. Induce the cells with 0.5 mM isopropyl- β -D-thiogalactopyranoside for 3 hrs at 30°C. Cells were collected by centrifugation at 5000rpm for 15min at 4°C (SLS6000 rotor). Pellet was resuspended in lysis

buffer and sonicated for 10min at 40W with intervals. After centrifugation at 16,000rpm for 15min at 4°C, the supernatant was incubated at 65°C for 15min. Then ultracentrifuge at 30,000rpm for 40min at 4°C (rotor Ti45, Beckman). Supernatant was then loaded to 1ml Histrap column (GE Healthcare) which was pre-equilibrated with binding buffer. The column was washed with 10 column volumes of washing buffer. Protein was eluted from the column with elution buffer. After SDS-PAGE analysis, samples were collected and exchanged to dilution buffer by centrifugation using Amicon® Ultra (MWCO 10kDa, Millipore). After concentrating protein to 2mg/ml, thrombin was added to cleave his₆-tag with 2 unit/per mg protein. Protein was digested for 2 hrs or overnight at 4°C, and then put to 65°C for 15min to denature the thrombin. After centrifugation at 14,000rpm for 20min, supernatant was collected and re-loaded to Histrap pre-equilibrated with dilution buffer. Protein without his-tag was collected in the flowthrough. Concentrated sample was further purified on a gel filtration column (Superose 12, GE Healthcare) equilibrated with gel filtration buffer. Fractions were analyzed by SDS-PAGE, concentrated to 3mg/ml and stored at -80°C.

4.2.2 Expression and purification of TBP wt and Δ tail

TBPwt and Δ tail variant were expressed and purified essentially the same as described for TFB. Observable precipitate always formed after thrombin cleavage, however after heat treatment at 65°C for 15min, proteins refolded correctly and yielded a clear protein solution. Proteins were analyzed by SDS-PAGE, concentrated to 2mg/ml and stored at -80°C.

4.2.3 Expression and purification of TFE

4.2.3.1 Buffers

Lysis buffer:

50 mM Hepes-Na pH 7.8

300 mM NaCl

10 μ M ZnCl₂

10 mM imidazole

10% glycerol

1x Protease Inhibitor

Ni-NTA binding buffer:

50 mM Hepes-Na pH 7.8

300 mM NaCl

10 μ M ZnCl₂

10 mM imidazole

10% glycerol

Ni-NTA washing buffer:

50 mM Hepes-Na pH 7.8

2 M NaCl

10 μ M ZnCl₂

40 mM imidazole

10% glycerol

Ni-NTA elution buffer:

50 mM Hepes-Na pH 7.8

300 mM NaCl

10 μ M ZnCl₂

300 mM imidazole

10% glycerol

Gel filtration buffer:

20 mM Hepes-Na pH 7.0

300 mM KCl

10% Glycerol

5 mM DTT

4.2.3.2 Expression and purification procedures

TFEwt was expressed as described for TFB. Cells were collected by centrifugation at 5000rpm for 15min at 4°C (SLS6000 rotor). Pellet was resuspended in lysis buffer and sonicated for 10min at 40W with intervals. After centrifugation at 16,000rpm for 15min at 4°C, the supernatant was incubated at 80°C for 20min. Then ultracentrifuge at 30,000rpm for 1hr at 4°C

(rotor Ti45, Beckman). Supernatant was then loaded to 1ml Histrap column (GE Healthcare) which was pre-equilibrated with binding buffer. The column was washed with 10 column volumes of washing buffer. Protein was eluted from the column with elution buffer. After SDS-PAGE analysis, samples were collected, concentrated, and further purified on a gel filtration column (Superdex 75, GE Healthcare) equilibrated with gel filtration buffer. Fractions were analyzed by SDS-PAGE, concentrated to 1mg/ml and stored at -80°C.

4.3 Purification of *P.fu* RNA polymerase

4.3.1 Buffers

Lysis buffer

50 mM Tris-HCl pH8.0

100mM KCl

10mM MgCl₂

20% Glycerin

10mM β-mercaptoethanol

1x Protease Inhibitor

Biorex +0/+1000/+100 Buffer

50 mM Tris-HCl pH8.0

0/1000/100mM KCl

2.5mM MgCl₂

1mM EDTA

10% Glycerin

10mM β-mercaptoethanol

Dialysis Buffer

50 mM Tris-HCl pH8.0

100mM KCl

2.5mM MgCl₂

10% Glycerin

10mM β-mercaptoethanol

Heparin +100/+1000 Buffer

50 mM Tris-HCl pH7.5

100/1000mM KCl

2.5mM MgCl₂

10% Glycerin

10mM β-mercaptoethanol

MonoQ +0/+100/+1000 Buffer

50 mM Tris-HCl pH8.0

0/100/1000mM KCl

2.5mM MgCl₂

10% Glycerin

10mM β-mercaptoethanol

Superose 6 Buffer

10mM Hepes pH7.0

150mM KCl

2.5mM MgCl₂

5mM DTT

10% Glycerin

4.3.2 Purification – day 1 (Preparing the cells)

P.fu cells were grown in a synthetic sea-water-medium under anaerobic conditions at 95°C and harvested during exponential growth phase. Cells were kindly provided by Prof. Michael Thomm's lab. Up to 60g frozen cells were thawed and resuspended in 200ml Lysis buffer overnight with stirring at 4°C.

P.fu cells are very sticky and are black.

4.3.3 Purification – day 2 (lysis and biorex column)

Cells were disrupted by EmulsiFlex-C5 (2 cycles at 1000-1500bar). Lysate was centrifuged for 30min at 10,000rpm (4°C). Supernatant was then ultracentrifuged at 40,000rpm for 90min at 4°C (rotor Ti45, Beckman). Then supernatant was collected which still had a black color and loaded to Biorex column (400ml Biorex 70 resin packed in a XK50 column) which was pre-equilibrated with Biorex+100 Buffer. Sample loading and washing used peristaltic pump (2-3 ml/min). Then wash biorex column with 2 column volumes of Biorex+100 Buffer in the presence of 1x Protease inhibitor (3ml/min, 2nd column volume washing on ÄKTA system). Proteins were eluted with a gradient from 100 to 1000mM KCl in 3 column volumes at 1-1.5ml/min on ÄKTA (GE Healthcare).

4.3.4 Purification – day 3 (Dialysis and heparin column)

Around 160ml black fractions were collected and dialysis against 5l Heparin+100 Buffer for at least 4 hrs at 4°C. Then protein solution was centrifuged at 12,000rpm for 30min at 4°C and loaded to heparin column(25ml HiPrep™ 16/10 Heparin FF) pre-equilibrated with Heparin+100 Buffer. Sample loading still used peristaltic pump(1.2 ml/min). After washing with 2 column volumes Heparin+100 Buffer, proteins were eluted with a gradient from 100 to 1000mM KCl in 10 column volumes at 1ml/min on ÄKTA(GE Healthcare). Around 60ml fractions were collected and dialysis again 2l MonoQ+100 Buffer overnight at 4°C.

4.3.5 Purification – day 4 (MonoQ column and gel filtration)

Samples were centrifuged at 12,000rpm for 30min at 4°C and loaded to MonoQ column pre-equilibrated with MonoQ+100 Buffer, at 0.5ml/min on ÄKTA(GE Healthcare). Column was washed with MonoQ+100 for 2 column

volumes. Proteins were eluted with a gradient from 150mM to 400mM KCl in 20 column volumes. Fractions were analysed with SDS-PAGE and the protein containing fractions were concentrated and loaded on a Superose 6 gel filtration column. The protein was concentrated to 3mg/ml and frozen to -80°C.

4.4 Assembly of PIC

4.4.1 Assembly of RNAP, TFB Δ finger, TBPwt/ Δ tail and DNA-RNA

DNA non-template, template, and RNA were annealed as a molar ratio 1:1:1.5 by heating to 95°C for 2 minutes and slowly cooling to room temperature. First TBPwt/ Δ tail, then TFB Δ finger were added to DNA-RNA and incubated for 1hr at 15°C. Then *P.fu* RNAP was added followed by another incubation for 1hr at 15°C. PIC complex was activated by incubation at 60°C for 15min. RNAP, TFB Δ finger, TBPwt/ Δ tail, DNA-RNA scaffold were assembled as a molar ratio 1:4:4:2. Gel filtration pre-equilibrated with Superose 6 Buffer was finally used to remove excess nucleic acids and factors.

4.4.2 Assembly of RNAP, TFBwt, TBPwt, TFEwt and DNA

DNA non-template, template from "Bubble" scaffolds were annealed as a molar ratio 1:1 by heating to 95°C for 2 minutes and slowly cooling to room temperature. First TBPwt, then TFBwt were added to DNA scaffold and incubated for 1hr at 15°C. Then *P.fu* RNAP was added followed by TFEwt and another incubation for 1hr at 15°C. PIC complex was activated by incubation at 60°C for 15min. RNAP, TFBwt, TBPwt, TFEwt, DNA scaffold were assembled as a molar ratio 1:4:4:4:2. Gel filtration pre-equilibrated with Superose 6 Buffer was finally used to remove excess nucleic acids and factors.

4.5 Crystallization

4.5.1 PIC crystallization and optimization

PIC formed by RNAP-TFB Δ finger-TBPwt/ Δ tail-DNA-RNA-“tail”-scaffold were concentrated to 4~5mg/ml. Initial crystallization setups using commercial screens were performed with the Hydra II semi-automatic protein crystallization robot (Matrix Technologies Apogent Discoveries) by sitting drop vapor diffusion method using 96-well crystallization plates (Corning). Equal amount of protein and crystallization solution drops (0.5 μ l) with 50 μ l reservoir solution were set. Initial crystallization setups using polymerase screens and subsequent optimization were performed manually using 24-well plates EasyXtal Tools (Qiagen) and the hanging drop method was applied. PIC crystals were grown at 20°C by using 2 μ l protein + 1 μ l drops from a reservoir solution: 12.5% PEG6000, 340mM NH₄NaTart, 100mM KSCN, 100mM HEPES pH7.5, 5mM DTT. Crystals were transferred stepwise to crystallization solution containing 15% PEG6000 and additionally 20% Ethylene Glycol and plunged into liquid nitrogen.

4.5.2 Crystallization techniques and post-crystallization methods

For macroseeding, small crystals (size around 40 μ m) were transferred to reservoir by capillary and washed. Then they were transferred to a new protein solution using a Cryoloop for crystal growth. For microseeding, several crystals were crushed using vortex to produce a seed stock. A dilution series was made to test for optimal seed concentration.

Cryo-protectants were tested by either direct transfer or stepwise transfer to the final solution. MPD (15%~25%), PEG400 (20%~30%), Glycerol (15%~25%), Ethylene Glycol (20%~30%), Melonate (3.5M~5M) and sucrose (25%~30%) were tested in combination with varying the concentration of precipitant PEG6000 (from concentration in reservoir up to 20%).

Three crystal annealing techniques were occupied as described (Heras and

Martin, 2005). For macromolecular crystal annealing (MCA), cryocooled crystal was removed from the cryostream and placed in cryosolution for 2~3min. The crystal was recooled in the cryostream. For the flash-annealing (FA) method, the cold-stream was blocked for 1.5~2s three times with intervals of 6s between each thawing step. For the annealing on the loop (AL) method, the cryo-stream was blocked until the drop became clear.

Abbreviations

ABC	subunits common for Pol I, II and III
BRE	TF(II)B recognition element
C	subunit of Pol III
CRSP	cleavage/polyadenylation specificity factor
CTD	C-terminal domain of Rpb1 of Pol II
DNA	deoxyribonucleic acid
DPE	downstream promoter element
DTT	dithiothreitol
<i>E.coli</i>	<i>Escherichia coli</i>
EDTA	ethylene diamine tetraacetic acid
EM	electron microscopy
EMSA	electrophoretic mobility shift assay
EtOH	ethanol
GDH	glutamate dehydrogenase
GTF	general transcription factor
HAT	histone acetyltransferase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HTH	helix turn helix
H.sapiens	Homo sapiens
Ihr/IE	initiator element
MALDI-TOF	matrix-assisted laser desorption ionization-time of flight
MPD	2-methyl-2,4-pentanediol
NaAC	sodium acetate
NMR	nuclear magnetic resonance

NTP	nucleotide triphosphate
ORF	open reading frame
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol
<i>P.furiosus/P.fu</i>	<i>Pyrococcus furiosus</i>
PIC	preinitiation complex
Pol	RNA polymerase
PPE	promoter proximal element
<i>P.woesei</i>	<i>Pyrococcus woesei</i>
R.m.s.deviation	root mean square deviation
RNA	ribonucleic acid
RNAP	RNA polymerase
Rpb	subunit of Pol II
RSC	remodel the structure of chromatin
<i>S.cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
S _N -2	substitution nucleophilic bimolecular
<i>S.pombe</i>	<i>Schizosaccharomyces pombe</i>
<i>S.solfataricus</i>	<i>Sulfolobus solfataricus</i>
<i>T.aquaticus/Taq</i>	<i>Thermus aquaticus</i>
TBP	TATA binding protein
TCA	trichloroacetic acid
TCEP	tris(2-carboxyethyl)phosphine
TFII	transcription factor of Pol II transcription
Tris	trishydroxymethylaminomethane
tRNA	transfer RNA
USA	upstream stimulatory activity

References

- Adelman, K., M. T. Marr, J. Werner, A. Saunders, Z. Ni, E. D. Andrulis, and J. T. Lis. 2005.** Efficient release from promoter-proximal stall sites requires transcript cleavage factor TFIIIS. *Mol Cell* **17**:103-12.
- Afonine, P. V., R. W. Grosse-Kunstleve, and P. D. Adams. 2005.** A robust bulk-solvent correction and anisotropic scaling procedure. *Acta Crystallogr D Biol Crystallogr* **61**:850-5.
- Alic, N., N. Ayoub, E. Landrieux, E. Favry, P. Baudouin-Cornu, M. Riva, and C. Carles. 2007.** Selectivity and proofreading both contribute significantly to the fidelity of RNA polymerase III transcription. *Proc Natl Acad Sci U S A* **104**:10400-5.
- Armache, K. J., H. Kettenberger, and P. Cramer. 2003.** Architecture of initiation-competent 12-subunit RNA polymerase II. *Proc Natl Acad Sci U S A* **100**:6964-8.
- Armache, K. J., S. Mitterweger, A. Meinhart, and P. Cramer. 2005.** Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J Biol Chem* **280**:7131-4.
- Awrey, D. E., N. Shimasaki, C. Koth, R. Weilbaecher, V. Olmsted, S. Kazanis, X. Shan, J. Arellano, C. H. Arrowsmith, C. M. Kane, and A. M. Edwards. 1998.** Yeast transcript elongation factor (TFIIS), structure and function. II: RNA polymerase binding, transcript cleavage, and read-through. *J Biol Chem* **273**:22595-605.
- Awrey, D. E., R. G. Weilbaecher, S. A. Hemming, S. M. Orlicky, C. M. Kane, and A. M. Edwards. 1997.** Transcription elongation through DNA arrest sites. A multistep process involving both RNA polymerase II subunit RPB9 and TFIIS. *J Biol Chem* **272**:14747-54.
- Baliga, N. S., Y. A. Goo, W. V. Ng, L. Hood, C. J. Daniels, and S. DasSarma. 2000.** Is gene expression in Halobacterium NRC-1 regulated by multiple TBP and TFB transcription factors? *Mol Microbiol* **36**:1184-5.
- Bartlett, M. S., M. Thomm, and E. P. Geiduschek. 2004.** Topography of the euryarchaeal transcription initiation complex. *J Biol Chem* **279**:5894-903.
- Baumann, M., J. Pontiller, and W. Ernst. 2010.** Structure and basal transcription complex of RNA polymerase II core promoters in the mammalian genome: an overview. *Mol Biotechnol* **45**:241-7.
- Beese, L. S., and T. A. Steitz. 1991.** Structural basis for the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism. *EMBO J* **10**:25-33.

- Bell, S. D., A. B. Brinkman, J. van der Oost, and S. P. Jackson. 2001.** The archaeal TFIIE α homologue facilitates transcription initiation by enhancing TATA-box recognition. *EMBO Rep* **2**:133-8.
- Bell, S. D., S. S. Cairns, R. L. Robson, and S. P. Jackson. 1999.** Transcriptional regulation of an archaeal operon in vivo and in vitro. *Mol Cell* **4**:971-82.
- Bell, S. D., and S. P. Jackson. 1998.** Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. *Trends Microbiol* **6**:222-8.
- Bell, S. D., and S. P. Jackson. 2000.** Mechanism of autoregulation by an archaeal transcriptional repressor. *J Biol Chem* **275**:31624-9.
- Bell, S. D., and S. P. Jackson. 2000.** The role of transcription factor B in transcription initiation and promoter clearance in the archaeon *Sulfolobus acidocaldarius*. *J Biol Chem* **275**:12934-40.
- Bell, S. D., P. L. Kosa, P. B. Sigler, and S. P. Jackson. 1999.** Orientation of the transcription preinitiation complex in archaea. *Proc Natl Acad Sci U S A* **96**:13662-7.
- Bentley, D. 2002.** The mRNA assembly line: transcription and processing machines in the same factory. *Current Opinion in Cell Biology* **14**:336-342.
- Bischler, N., L. Brino, C. Carles, M. Riva, H. Tschochner, V. Mallouh, and P. Schultz. 2002.** Localization of the yeast RNA polymerase I-specific subunits. *EMBO J* **21**:4136-44.
- Bobkova, E. V., N. Habib, G. Alexander, and B. D. Hall. 1999.** Mutational analysis of the hydrolytic activity of yeast RNA polymerase III. *J Biol Chem* **274**:21342-8.
- Bobkova, E. V., and B. D. Hall. 1997.** Substrate specificity of the RNase activity of yeast RNA polymerase III. *J Biol Chem* **272**:22832-9.
- Booth, V., C. M. Koth, A. M. Edwards, and C. H. Arrowsmith. 2000.** Structure of a conserved domain common to the transcription factors TFIIS, elongin A, and CRSP70. *J Biol Chem* **275**:31266-8.
- Borukhov, S., A. Polyakov, V. Nikiforov, and A. Goldfarb. 1992.** GreA protein: a transcription elongation factor from *Escherichia coli*. *Proc Natl Acad Sci U S A* **89**:8899-902.
- Borukhov, S., V. Sagitov, and A. Goldfarb. 1993.** Transcript cleavage factors from *E. coli*. *Cell* **72**:459-66.

- Brueckner, F., and P. Cramer. 2008.** Structural basis of transcription inhibition by alpha-amanitin and implications for RNA polymerase II translocation. *Nature Structural & Molecular Biology* **15**:811-818.
- Bushnell, D. A., K. D. Westover, R. E. Davis, and R. D. Kornberg. 2004.** Structural basis of transcription: an RNA polymerase II-TFIIB cocystal at 4.5 Angstroms. *Science* **303**:983-8.
- Chao, D. M., E. L. Gadbois, P. J. Murray, S. F. Anderson, M. S. Sonu, J. D. Parvin, and R. A. Young. 1996.** A mammalian SRB protein associated with an RNA polymerase II holoenzyme. *Nature* **380**:82-5.
- Chedin, S., M. Riva, P. Schultz, A. Sentenac, and C. Carles. 1998.** The RNA cleavage activity of RNA polymerase III is mediated by an essential TFIIIS-like subunit and is important for transcription termination. *Genes Dev* **12**:3857-71.
- Chen, H. T., and S. Hahn. 2004.** Mapping the location of TFIIB within the RNA polymerase II transcription preinitiation complex: a model for the structure of the PIC. *Cell* **119**:169-80.
- Chen, H. T., L. Warfield, and S. Hahn. 2007.** The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex. *Nat Struct Mol Biol* **14**:696-703.
- Cheung, A., and P. Cramer. 2010.** Structural basis of RNA polymerase II backtracking, arrest, and reactivation. *Nature in revision*.
- Cipres-Palacin, G., and C. M. Kane. 1994.** Cleavage of the nascent transcript induced by TFIIIS is insufficient to promote read-through of intrinsic blocks to elongation by RNA polymerase II. *Proc Natl Acad Sci U S A* **91**:8087-91.
- Cipres-Palacin, G., and C. M. Kane. 1995.** Alanine-scanning mutagenesis of human transcript elongation factor TFIIIS. *Biochemistry* **34**:15375-80.
- Cramer, P. 2002.** Common structural features of nucleic acid polymerases. *Bioessays* **24**:724-9.
- Cramer, P. 2004.** RNA polymerase II structure: from core to functional complexes. *Curr Opin Genet Dev* **14**:218-26.
- Cramer, P., D. A. Bushnell, and R. D. Kornberg. 2001.** Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**:1863-76.
- Crick, F. 1970.** Central dogma of molecular biology. *Nature* **227**:561-3.

- Dahlke, I., and M. Thomm. 2002.** A Pyrococcus homolog of the leucine-responsive regulatory protein, LrpA, inhibits transcription by abrogating RNA polymerase recruitment. *Nucleic Acids Res* **30**:701-10.
- De Carlo, S., C. Carles, M. Riva, and P. Schultz. 2003.** Cryo-negative staining reveals conformational flexibility within yeast RNA polymerase I. *J Mol Biol* **329**:891-902.
- De Carlo, S., S. C. Lin, D. J. Taatjes, and A. Hoenger. 2010.** Molecular basis of transcription initiation in Archaea. *Transcr* **1**:103-111.
- DeDecker, B. S., R. O'Brien, P. J. Fleming, J. H. Geiger, S. P. Jackson, and P. B. Sigler. 1996.** The crystal structure of a hyperthermophilic archaeal TATA-box binding protein. *J Mol Biol* **264**:1072-84.
- Dengl, S., and P. Cramer. 2009.** Torpedo nuclease Rat1 is insufficient to terminate RNA polymerase II in vitro. *J Biol Chem* **284**:21270-9.
- Edwards, A. M., C. M. Kane, R. A. Young, and R. D. Kornberg. 1991.** Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. *J Biol Chem* **266**:71-5.
- Erie, D. A., O. Hajiseyedjavadi, M. C. Young, and P. H. von Hippel. 1993.** Multiple RNA polymerase conformations and GreA: control of the fidelity of transcription. *Science* **262**:867-73.
- Feng, G. H., D. N. Lee, D. Wang, C. L. Chan, and R. Landick. 1994.** GreA-induced transcript cleavage in transcription complexes containing Escherichia coli RNA polymerase is controlled by multiple factors, including nascent transcript location and structure. *J Biol Chem* **269**:22282-94.
- Fernandez-Tornero, C., B. Bottcher, U. J. Rashid, U. Steuerwald, B. Florchinger, D. P. Devos, D. Lindner, and C. W. Muller. 2010.** Conformational flexibility of RNA polymerase III during transcriptional elongation. *EMBO J* **29**:3762-72.
- Fernandez-Tornero, C., B. Bottcher, M. Riva, C. Carles, U. Steuerwald, R. W. Ruigrok, A. Sentenac, C. W. Muller, and G. Schoehn. 2007.** Insights into transcription initiation and termination from the electron microscopy structure of yeast RNA polymerase III. *Mol Cell* **25**:813-23.
- Fish, R. N., and C. M. Kane. 2002.** Promoting elongation with transcript cleavage stimulatory factors. *Biochim Biophys Acta* **1577**:287-307.
- Flores, A., J. F. Briand, O. Gadal, J. C. Andrau, L. Rubbi, V. Van Mullem, C. Boschiero, M. Goussot, C. Marck, C. Carles, P. Thuriaux, A. Sentenac, and M. Werner. 1999.** A

- protein-protein interaction map of yeast RNA polymerase III. *Proc Natl Acad Sci U S A* **96**:7815-20.
- Gadal, O., S. Mariotte-Labarre, S. Chedin, E. Quemeneur, C. Carles, A. Sentenac, and P. Thuriaux. 1997.** A34.5, a nonessential component of yeast RNA polymerase I, cooperates with subunit A14 and DNA topoisomerase I to produce a functional rRNA synthesis machine. *Mol Cell Biol* **17**:1787-95.
- Geiduschek, E. P., and G. A. Kassavetis. 2001.** The RNA polymerase III transcription apparatus. *J Mol Biol* **310**:1-26.
- Geiger, S. R., K. Lorenzen, A. Schrieck, P. Hanecker, D. Kostrewa, A. J. Heck, and P. Cramer. 2010.** RNA polymerase I contains a TFIIIF-related DNA-binding subcomplex. *Mol Cell* **39**:583-94.
- Ghavi-Helm, Y., M. Michaut, J. Acker, J. C. Aude, P. Thuriaux, M. Werner, and J. Soutourina. 2008.** Genome-wide location analysis reveals a role of TFIIIS in RNA polymerase III transcription. *Genes Dev* **22**:1934-47.
- Gnatt, A. L., P. Cramer, J. Fu, D. A. Bushnell, and R. D. Kornberg. 2001.** Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* **292**:1876-82.
- Goede, B., S. Naji, O. von Kampen, K. Ilg, and M. Thomm. 2006.** Protein-protein interactions in the archaeal transcriptional machinery: binding studies of isolated RNA polymerase subunits and transcription factors. *J Biol Chem* **281**:30581-92.
- Grohmann, D., A. Hirtreiter, and F. Werner. 2009.** RNAP subunits F/E (RPB4/7) are stably associated with archaeal RNA polymerase: using fluorescence anisotropy to monitor RNAP assembly in vitro. *Biochem J* **421**:339-43.
- Grunberg, S., M. S. Bartlett, S. Naji, and M. Thomm. 2007.** Transcription factor E is a part of transcription elongation complexes. *J Biol Chem* **282**:35482-90.
- Hagler, J., and S. Shuman. 1993.** Nascent RNA cleavage by purified ternary complexes of vaccinia RNA polymerase. *J Biol Chem* **268**:2166-73.
- Hain, J., W. D. Reiter, U. Hudepohl, and W. Zillig. 1992.** Elements of an archaeal promoter defined by mutational analysis. *Nucleic Acids Res* **20**:5423-8.
- Hanzelka, B. L., T. J. Darcy, and J. N. Reeve. 2001.** TFE, an archaeal transcription factor in *Methanobacterium thermoautotrophicum* related to eucaryal transcription factor TFIIIEalpha. *J Bacteriol* **183**:1813-8.

- Hausner, W., G. Frey, and M. Thomm. 1991.** Control regions of an archaeal gene. A TATA box and an initiator element promote cell-free transcription of the tRNA(Val) gene of *Methanococcus vannielii*. *J Mol Biol* **222**:495-508.
- Hausner, W., U. Lange, and M. Musfeldt. 2000.** Transcription factor S, a cleavage induction factor of the archaeal RNA polymerase. *J Biol Chem* **275**:12393-9.
- Hemming, S. A., and A. M. Edwards. 2000.** Yeast RNA polymerase II subunit RPB9. Mapping of domains required for transcription elongation. *J Biol Chem* **275**:2288-94.
- Henkin, T. M. 2000.** Transcription termination control in bacteria. *Current Opinion in Microbiology* **3**:149-153.
- Heras, B., and J. L. Martin. 2005.** Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta Crystallogr D Biol Crystallogr* **61**:1173-80.
- Hethke, C., A. C. Geerling, W. Hausner, W. M. de Vos, and M. Thomm. 1996.** A cell-free transcription system for the hyperthermophilic archaeon *Pyrococcus furiosus*. *Nucleic Acids Res* **24**:2369-76.
- Hirai, H., K. Sekimizu, M. Horikoshi, Y. Nakanishi, and S. Natori. 1988.** Stimulation of transcription from accurate initiation sites by purified S-II. *FEBS Lett* **238**:119-22.
- Hirata, A., B. J. Klein, and K. S. Murakami. 2008.** The X-ray crystal structure of RNA polymerase from Archaea. *Nature* **451**:851-4.
- Hirata, A., and K. S. Murakami. 2009.** Archaeal RNA polymerase. *Curr Opin Struct Biol* **19**:724-31.
- Horikoshi, M., K. Sekimizu, S. Hirashima, Y. Mitsuhashi, and S. Natori. 1985.** Structural relationships of the three stimulatory factors of RNA polymerase II from Ehrlich ascites tumor cells. *J Biol Chem* **260**:5739-44.
- Hsu, L. M., N. V. Vo, and M. J. Chamberlin. 1995.** Escherichia coli transcript cleavage factors GreA and GreB stimulate promoter escape and gene expression in vivo and in vitro. *Proc Natl Acad Sci U S A* **92**:11588-92.
- Huang, Y., R. V. Intine, A. Mozlin, S. Hasson, and R. J. Maraia. 2005.** Mutations in the RNA polymerase III subunit Rpc11p that decrease RNA 3' cleavage activity increase 3'-terminal oligo(U) length and La-dependent tRNA processing. *Mol Cell Biol* **25**:621-36.
- Huet, J., J. M. Buhler, A. Sentenac, and P. Fromageot. 1977.** Characterization of ribonuclease H activity associated yeast RNA polymerase A. *J Biol Chem*

252:8848-55.

- Huet, J., R. Schnabel, A. Sentenac, and W. Zillig. 1983. Archaeobacteria and eukaryotes possess DNA-dependent RNA polymerases of a common type. *EMBO J* 2:1291-4.
- Huet, J., F. Wyers, J. M. Buhler, A. Sentenac, and P. Fromageot. 1976. Association of RNase H activity with yeast RNA polymerase A. *Nature* 261:431-3.
- Hutchison, C. A., S. N. Peterson, S. R. Gill, R. T. Cline, O. White, C. M. Fraser, H. O. Smith, and J. C. Venter. 1999. Global transposon mutagenesis and a minimal Mycoplasma genome. *Science* 286:2165-9.
- Iborra, F., J. Huet, B. Breant, A. Sentenac, and P. Fromageot. 1979. Identification of two different RNase H activities associated with yeast RNA polymerase A. *J Biol Chem* 254:10920-4.
- Izban, M. G., and D. S. Luse. 1992. The RNA polymerase II ternary complex cleaves the nascent transcript in a 3'----5' direction in the presence of elongation factor SII. *Genes Dev* 6:1342-56.
- Izban, M. G., and D. S. Luse. 1993. The increment of SII-facilitated transcript cleavage varies dramatically between elongation competent and incompetent RNA polymerase II ternary complexes. *J Biol Chem* 268:12874-85.
- Izban, M. G., and D. S. Luse. 1993. SII-facilitated transcript cleavage in RNA polymerase II complexes stalled early after initiation occurs in primarily dinucleotide increments. *J Biol Chem* 268:12864-73.
- Jacobson, E. M., P. Li, M. G. Rosenfeld, and A. K. Aggarwal. 1996. Crystallization and preliminary X-ray analysis of Pit-1 POU domain complexed to a 28 base pair DNA element. *Proteins* 24:263-5.
- Janke, C., M. M. Magiera, N. Rathfelder, C. Taxis, S. Reber, H. Maekawa, A. Moreno-Borchart, G. Doenges, E. Schwob, E. Schiebel, and M. Knop. 2004. A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. *Yeast* 21:947-62.
- Jeon, C., and K. Agarwal. 1996. Fidelity of RNA polymerase II transcription controlled by elongation factor TFIIIS. *Proc Natl Acad Sci U S A* 93:13677-82.
- Jeon, C., H. Yoon, and K. Agarwal. 1994. The transcription factor TFIIIS zinc ribbon dipeptide Asp-Glu is critical for stimulation of elongation and RNA cleavage by RNA polymerase II. *Proc Natl Acad Sci U S A* 91:9106-10.

- Johnson, T. L., and M. J. Chamberlin. 1994.** Complexes of yeast RNA polymerase II and RNA are substrates for TFIIIS-induced RNA cleavage. *Cell* **77**:217-24.
- Jun, S. H., M. J. Reichlen, M. Tajiri, and K. S. Murakami. 2011.** Archaeal RNA polymerase and transcription regulation. *Crit Rev Biochem Mol Biol* **46**:27-40.
- Kabsch, W. 1988.** Automatic indexing of rotation diffraction patterns. *J. Applied Crystallography* **21**:67-71.
- Kaine, B. P., I. J. Mehr, and C. R. Woese. 1994.** The sequence, and its evolutionary implications, of a *Thermococcus celer* protein associated with transcription. *Proc Natl Acad Sci U S A* **91**:3854-6.
- Kassavetis, G. A., P. Prakash, and E. Shim. 2010.** The C53/C37 subcomplex of RNA polymerase III lies near the active site and participates in promoter opening. *J Biol Chem* **285**:2695-706.
- Kettenberger, H., K. J. Armache, and P. Cramer. 2003.** Architecture of the RNA polymerase II-TFIIIS complex and implications for mRNA cleavage. *Cell* **114**:347-57.
- Kettenberger, H., K. J. Armache, and P. Cramer. 2004.** Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIIS. *Molecular Cell* **16**:955-965.
- Kim, B., A. I. Nesvizhskii, P. G. Rani, S. Hahn, R. Aebersold, and J. A. Ranish. 2007.** The transcription elongation factor TFIIIS is a component of RNA polymerase II preinitiation complexes. *Proc Natl Acad Sci U S A* **104**:16068-73.
- Kim, J. L., D. B. Nikolov, and S. K. Burley. 1993.** Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365**:520-7.
- Kireeva, M. L., B. Hancock, G. H. Cremona, W. Walter, V. M. Studitsky, and M. Kashlev. 2005.** Nature of the nucleosomal barrier to RNA polymerase II. *Mol Cell* **18**:97-108.
- Kireeva, M. L., N. Komissarova, D. S. Waugh, and M. Kashlev. 2000.** The 8-nucleotide-long RNA:DNA hybrid is a primary stability determinant of the RNA polymerase II elongation complex. *J Biol Chem* **275**:6530-6.
- Komissarova, N., M. L. Kireeva, J. Becker, I. Sidorenkov, and M. Kashlev. 2003.** Engineering of elongation complexes of bacterial and yeast RNA polymerases. *Methods Enzymol* **371**:233-51.
- Koonin, E. V., K. S. Makarova, and J. G. Elkins. 2007.** Orthologs of the small RPB8 subunit of the eukaryotic RNA polymerases are conserved in hyperthermophilic

- Crenarchaeota and "Korarchaeota". *Biol Direct* **2**:38.
- Korkhin, Y., U. M. Unligil, O. Littlefield, P. J. Nelson, D. I. Stuart, P. B. Sigler, S. D. Bell, and N. G. A. Abrescia. 2009.** Evolution of Complex RNA Polymerases: The Complete Archaeal RNA Polymerase Structure. *Plos Biology* **7**:-
- Kosa, P. F., G. Ghosh, B. S. DeDecker, and P. B. Sigler. 1997.** The 2.1-A crystal structure of an archaeal preinitiation complex: TATA-box-binding protein/transcription factor (II)B core/TATA-box. *Proc Natl Acad Sci U S A* **94**:6042-7.
- Kostrewa, D., M. E. Zeller, K. J. Armache, M. Seizi, K. Leike, M. Thomm, and P. Cramer. 2009.** RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature* **462**:323-30.
- Koulich, D., V. Nikiforov, and S. Borukhov. 1998.** Distinct functions of N and C-terminal domains of GreA, an Escherichia coli transcript cleavage factor. *J Mol Biol* **276**:379-89.
- Koulich, D., M. Orlova, A. Malhotra, A. Sali, S. A. Darst, and S. Borukhov. 1997.** Domain organization of Escherichia coli transcript cleavage factors GreA and GreB. *J Biol Chem* **272**:7201-10.
- Koyama, H., T. Ito, T. Nakanishi, N. Kawamura, and K. Sekimizu. 2003.** Transcription elongation factor S-II maintains transcriptional fidelity and confers oxidative stress resistance. *Genes Cells* **8**:779-88.
- Koyama, H., T. Ito, T. Nakanishi, and K. Sekimizu. 2007.** Stimulation of RNA polymerase II transcript cleavage activity contributes to maintain transcriptional fidelity in yeast. *Genes Cells* **12**:547-59.
- Kuhn, A., and I. Grummt. 1989.** 3'-end formation of mouse pre-rRNA involves both transcription termination and a specific processing reaction. *Genes Dev* **3**:224-31.
- Kuhn, C. D., S. R. Geiger, S. Baumli, M. Gartmann, J. Gerber, S. Jennebach, T. Mielke, H. Tschochner, R. Beckmann, and P. Cramer. 2007.** Functional architecture of RNA polymerase I. *Cell* **131**:1260-72.
- Kulish, D., J. Lee, I. Lomakin, B. Nowicka, A. Das, S. Darst, K. Normet, and S. Borukhov. 2000.** The functional role of basic patch, a structural element of Escherichia coli transcript cleavage factors GreA and GreB. *J Biol Chem* **275**:12789-98.
- Kusser, A. G., M. G. Bertero, S. Naji, T. Becker, M. Thomm, R. Beckmann, and P. Cramer. 2008.** Structure of an archaeal RNA polymerase. *J Mol Biol* **376**:303-7.

- Kwapisz, M., F. Beckouet, and P. Thuriaux. 2008.** Early evolution of eukaryotic DNA-dependent RNA polymerases. *Trends Genet* **24**:211-5.
- Labhart, P. 1997.** Transcript cleavage in an RNA polymerase I elongation complex. Evidence for a dissociable activity similar to but distinct from TFIIIS. *J Biol Chem* **272**:9055-61.
- Labhart, P., and G. T. Morgan. 1998.** Identification of novel genes encoding transcription elongation factor TFIIIS (TCEA) in vertebrates: conservation of three distinct TFIIIS isoforms in frog, mouse, and human. *Genomics* **52**:278-88.
- Landrieux, E., N. Alic, C. Ducrot, J. Acker, M. Riva, and C. Carles. 2006.** A subcomplex of RNA polymerase III subunits involved in transcription termination and reinitiation. *EMBO J* **25**:118-28.
- Lange, U., and W. Hausner. 2004.** Transcriptional fidelity and proofreading in Archaea and implications for the mechanism of TFS-induced RNA cleavage. *Mol Microbiol* **52**:1133-43.
- Langer, D., J. Hain, P. Thuriaux, and W. Zillig. 1995.** Transcription in archaea: similarity to that in eucarya. *Proc Natl Acad Sci U S A* **92**:5768-72.
- Langer, D., and W. Zillig. 1993.** Putative tflIs gene of *Sulfolobus acidocaldarius* encoding an archaeal transcription elongation factor is situated directly downstream of the gene for a small subunit of DNA-dependent RNA polymerase. *Nucleic Acids Res* **21**:2251.
- Laptenko, O., J. Lee, I. Lomakin, and S. Borukhov. 2003.** Transcript cleavage factors GreA and GreB act as transient catalytic components of RNA polymerase. *EMBO J* **22**:6322-34.
- Li, B., M. Carey, and J. L. Workman. 2007.** The role of chromatin during transcription. *Cell* **128**:707-19.
- Littlefield, O., Y. Korkhin, and P. B. Sigler. 1999.** The structural basis for the oriented assembly of a TBP/TFB/promoter complex. *Proc Natl Acad Sci U S A* **96**:13668-73.
- Malagon, F., A. H. Tong, B. K. Shafer, and J. N. Strathern. 2004.** Genetic interactions of DST1 in *Saccharomyces cerevisiae* suggest a role of TFIIIS in the initiation-elongation transition. *Genetics* **166**:1215-27.
- Maldonado, E., R. Shiekhattar, M. Sheldon, H. Cho, R. Drapkin, P. Rickert, E. Lees, C. W. Anderson, S. Linn, and D. Reinberg. 1996.** A human RNA polymerase II complex associated with SRB and DNA-repair proteins. *Nature* **381**:86-9.
- Martin, F. H., and I. Tinoco, Jr. 1980.** DNA-RNA hybrid duplexes containing oligo(dA:rU)

sequences are exceptionally unstable and may facilitate termination of transcription. *Nucleic Acids Res* **8**:2295-9.

- McCoy, A. J., R. W. Grosse-Kunstleve, L. C. Storoni, and R. J. Read. 2005.** Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr* **61**:458-64.
- Meinhart, A., J. Blobel, and P. Cramer. 2003.** An extended winged helix domain in general transcription factor E/IIIE alpha. *J Biol Chem* **278**:48267-74.
- Micorescu, M., S. Grunberg, A. Franke, P. Cramer, M. Thomm, and M. Bartlett. 2008.** Archaeal transcription: function of an alternative transcription factor B from *Pyrococcus furiosus*. *J Bacteriol* **190**:157-67.
- Miropolskaya, N., I. Artsimovitch, S. Klimasauskas, V. Nikiforov, and A. Kulbachinskiy. 2009.** Allosteric control of catalysis by the F loop of RNA polymerase. *Proc Natl Acad Sci U S A* **106**:18942-7.
- Morin, P. E., D. E. Awrey, A. M. Edwards, and C. H. Arrowsmith. 1996.** Elongation factor TFIIS contains three structural domains: solution structure of domain II. *Proc Natl Acad Sci U S A* **93**:10604-8.
- Murakami, K. S., S. Masuda, and S. A. Darst. 2002.** Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science* **296**:1280-4.
- Naji, S., S. Grunberg, and M. Thomm. 2007.** The RPB7 orthologue E' is required for transcriptional activity of a reconstituted archaeal core enzyme at low temperatures and stimulates open complex formation. *J Biol Chem* **282**:11047-57.
- Nakanishi, T., M. Shimoaraiso, T. Kubo, and S. Natori. 1995.** Structure-function relationship of yeast S-II in terms of stimulation of RNA polymerase II, arrest relief, and suppression of 6-azauracil sensitivity. *J Biol Chem* **270**:8991-5.
- Nechaev, S., D. C. Fargo, G. dos Santos, L. Liu, Y. Gao, and K. Adelman. 2010.** Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in *Drosophila*. *Science* **327**:335-8.
- Nesser, N. K., D. O. Peterson, and D. K. Hawley. 2006.** RNA polymerase II subunit Rpb9 is important for transcriptional fidelity in vivo. *Proc Natl Acad Sci U S A* **103**:3268-73.
- Nikolov, D. B., H. Chen, E. D. Halay, A. Hoffman, R. G. Roeder, and S. K. Burley. 1996.** Crystal structure of a human TATA box-binding protein/TATA element complex. *Proc Natl Acad Sci U S A* **93**:4862-7.

- Nikolov, D. B., H. Chen, E. D. Halay, A. A. Usheva, K. Hisatake, D. K. Lee, R. G. Roeder, and S. K. Burley. 1995.** Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature* **377**:119-28.
- Nogi, Y., R. Yano, J. Dodd, C. Carles, and M. Nomura. 1993.** Gene RRN4 in *Saccharomyces cerevisiae* encodes the A12.2 subunit of RNA polymerase I and is essential only at high temperatures. *Mol Cell Biol* **13**:114-22.
- Olmsted, V. K., D. E. Awrey, C. Koth, X. Shan, P. E. Morin, S. Kazanis, A. M. Edwards, and C. H. Arrowsmith. 1998.** Yeast transcript elongation factor (TFIIS), structure and function. I: NMR structural analysis of the minimal transcriptionally active region. *J Biol Chem* **273**:22589-94.
- Olsen, G. J., and C. R. Woese. 1997.** Archaeal genomics: an overview. *Cell* **89**:991-4.
- Opalka, N., M. Chlenov, P. Chacon, W. J. Rice, W. Wriggers, and S. A. Darst. 2003.** Structure and function of the transcription elongation factor GreB bound to bacterial RNA polymerase. *Cell* **114**:335-45.
- Orlova, M., J. Newlands, A. Das, A. Goldfarb, and S. Borukhov. 1995.** Intrinsic transcript cleavage activity of RNA polymerase. *Proc Natl Acad Sci U S A* **92**:4596-600.
- Ossipow, V., J. P. Tassan, E. A. Nigg, and U. Schibler. 1995.** A mammalian RNA polymerase II holoenzyme containing all components required for promoter-specific transcription initiation. *Cell* **83**:137-46.
- Ouhammouch, M. 2004.** Transcriptional regulation in Archaea. *Curr Opin Genet Dev* **14**:133-8.
- Ouhammouch, M., F. Werner, R. O. Weinzierl, and E. P. Geiduschek. 2004.** A fully recombinant system for activator-dependent archaeal transcription. *J Biol Chem* **279**:51719-21.
- Palangat, M., D. B. Renner, D. H. Price, and R. Landick. 2005.** A negative elongation factor for human RNA polymerase II inhibits the anti-arrest transcript-cleavage factor TFIIS. *Proc Natl Acad Sci U S A* **102**:15036-41.
- Palmer, J. R., and C. J. Daniels. 1995.** In vivo definition of an archaeal promoter. *J Bacteriol* **177**:1844-9.
- Pan, G., T. Aso, and J. Greenblatt. 1997.** Interaction of elongation factors TFIIS and elongin A with a human RNA polymerase II holoenzyme capable of promoter-specific initiation and responsive to transcriptional activators. *J Biol Chem* **272**:24563-71.

- Paule, M. R., and R. J. White. 2000.** Transcription by RNA polymerases I and III. *Nucleic Acids Research* **28**:1283-1298.
- Plant, K. E., A. Hair, and G. T. Morgan. 1996.** Genes encoding isoforms of transcription elongation factor TFIIS in *Xenopus* and the use of multiple unusual RNA processing signals. *Nucleic Acids Res* **24**:3514-21.
- Polyakov, A., C. Richter, A. Malhotra, D. Koulich, S. Borukhov, and S. A. Darst. 1998.** Visualization of the binding site for the transcript cleavage factor GreB on *Escherichia coli* RNA polymerase. *J Mol Biol* **281**:465-73.
- Poole, A. M., and D. T. Logan. 2005.** Modern mRNA proofreading and repair: clues that the last universal common ancestor possessed an RNA genome? *Mol Biol Evol* **22**:1444-55.
- Prescott, E. M., Y. N. Osheim, H. S. Jones, C. M. Alen, J. G. Roan, R. H. Reeder, A. L. Beyer, and N. J. Proudfoot. 2004.** Transcriptional termination by RNA polymerase I requires the small subunit Rpa12p. *Proc Natl Acad Sci U S A* **101**:6068-73.
- Proudfoot, N. J., A. Furger, and M. J. Dye. 2002.** Integrating mRNA processing with transcription. *Cell* **108**:501-12.
- Qian, X., S. N. Gozani, H. Yoon, C. J. Jeon, K. Agarwal, and M. A. Weiss. 1993.** Novel zinc finger motif in the basal transcriptional machinery: three-dimensional NMR studies of the nucleic acid binding domain of transcriptional elongation factor TFIIS. *Biochemistry* **32**:9944-59.
- Qian, X., C. Jeon, H. Yoon, K. Agarwal, and M. A. Weiss. 1993.** Structure of a new nucleic-acid-binding motif in eukaryotic transcriptional elongation factor TFIIS. *Nature* **365**:277-9.
- Qureshi, S. A. 2006.** Role of the *Sulfolobus shibatae* viral T6 initiator in conferring promoter strength and in influencing transcription start site selection. *Can J Microbiol* **52**:1136-40.
- Qureshi, S. A., and S. P. Jackson. 1998.** Sequence-specific DNA binding by the *S. shibatae* TFIIB homolog, TFB, and its effect on promoter strength. *Mol Cell* **1**:389-400.
- Ream, T. S., J. R. Haag, A. T. Wierzbicki, C. D. Nicora, A. D. Norbeck, J. K. Zhu, G. Hagen, T. J. Guilfoyle, L. Pasa-Tolic, and C. S. Pikaard. 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**:192-203.
- Reeve, J. N. 2003.** Archaeal chromatin and transcription. *Mol Microbiol* **48**:587-98.

- Reines, D. 1992.** Elongation factor-dependent transcript shortening by template-engaged RNA polymerase II. *J Biol Chem* **267**:3795-800.
- Reiter, W. D., U. Hudepohl, and W. Zillig. 1990.** Mutational analysis of an archaebacterial promoter: essential role of a TATA box for transcription efficiency and start-site selection in vitro. *Proc Natl Acad Sci U S A* **87**:9509-13.
- Renfrow, M. B., N. Naryshkin, L. M. Lewis, H. T. Chen, R. H. Ebright, and R. A. Scott. 2004.** Transcription factor B contacts promoter DNA near the transcription start site of the archaeal transcription initiation complex. *J Biol Chem* **279**:2825-31.
- Rivera, M. C., R. Jain, J. E. Moore, and J. A. Lake. 1998.** Genomic evidence for two functionally distinct gene classes. *Proc Natl Acad Sci U S A* **95**:6239-44.
- Roeder, R. G. 2005.** Transcriptional regulation and the role of diverse coactivators in animal cells. *FEBS Lett* **579**:909-15.
- Rudd, M. D., M. G. Izban, and D. S. Luse. 1994.** The active site of RNA polymerase II participates in transcript cleavage within arrested ternary complexes. *Proc Natl Acad Sci U S A* **91**:8057-61.
- Rudd, M. D., and D. S. Luse. 1996.** Amanitin greatly reduces the rate of transcription by RNA polymerase II ternary complexes but fails to inhibit some transcript cleavage modes. *J Biol Chem* **271**:21549-58.
- Sastry, S. S., and B. M. Ross. 1997.** Nuclease activity of T7 RNA polymerase and the heterogeneity of transcription elongation complexes. *J Biol Chem* **272**:8644-52.
- Shilatifard, A. 1998.** The RNA polymerase II general elongation complex. *Biological Chemistry* **379**:27-31.
- Shilatifard, A., R. C. Conaway, and J. W. Conaway. 2003.** The RNA polymerase II elongation complex. *Annual Review of Biochemistry* **72**:693-715.
- Shimasaki, N. B., and C. M. Kane. 2000.** Structural basis for the species-specific activity of TFIIIS. *J Biol Chem* **275**:36541-9.
- Sigurdsson, S., A. B. Dirac-Svejstrup, and J. Q. Svejstrup. 2010.** Evidence that transcript cleavage is essential for RNA polymerase II transcription and cell viability. *Mol Cell* **38**:202-10.
- Sikorski, T. W., and S. Buratowski. 2009.** The basal initiation machinery: beyond the general transcription factors. *Curr Opin Cell Biol* **21**:344-51.

- Smale, S. T., and J. T. Kadonaga. 2003.** The RNA polymerase II core promoter. *Annu Rev Biochem* **72**:449-79.
- Soppa, J. 1999.** Normalized nucleotide frequencies allow the definition of archaeal promoter elements for different archaeal groups and reveal base-specific TFB contacts upstream of the TATA box. *Mol Microbiol* **31**:1589-92.
- Soppa, J. 1999.** Transcription initiation in Archaea: facts, factors and future aspects. *Mol Microbiol* **31**:1295-305.
- Sosunov, V., E. Sosunova, A. Mustaev, I. Bass, V. Nikiforov, and A. Goldfarb. 2003.** Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerase. *EMBO J* **22**:2234-44.
- Sosunov, V., S. Zorov, E. Sosunova, A. Nikolaev, I. Zakeyeva, I. Bass, A. Goldfarb, V. Nikiforov, K. Severinov, and A. Mustaev. 2005.** The involvement of the aspartate triad of the active center in all catalytic activities of multisubunit RNA polymerase. *Nucleic Acids Res* **33**:4202-11.
- Sosunova, E., V. Sosunov, M. Kozlov, V. Nikiforov, A. Goldfarb, and A. Mustaev. 2003.** Donation of catalytic residues to RNA polymerase active center by transcription factor Gre. *Proc Natl Acad Sci U S A* **100**:15469-74.
- Spitalny, P., and M. Thomm. 2003.** Analysis of the open region and of DNA-protein contacts of archaeal RNA polymerase transcription complexes during transition from initiation to elongation. *J Biol Chem* **278**:30497-505.
- Spitalny, P., and M. Thomm. 2008.** A polymerase III-like reinitiation mechanism is operating in regulation of histone expression in archaea. *Mol Microbiol* **67**:958-70.
- Stebbins, C. E., S. Borukhov, M. Orlova, A. Polyakov, A. Goldfarb, and S. A. Darst. 1995.** Crystal structure of the GreA transcript cleavage factor from Escherichia coli. *Nature* **373**:636-40.
- Steitz, T. A. 1998.** Structural biology - A mechanism for all polymerases. *Nature* **391**:231-232.
- Surratt, C. K., S. C. Milan, and M. J. Chamberlin. 1991.** Spontaneous cleavage of RNA in ternary complexes of Escherichia coli RNA polymerase and its significance for the mechanism of transcription. *Proc Natl Acad Sci U S A* **88**:7983-7.
- Svejstrup, J. Q. 2004.** The RNA polymerase II transcription cycle: cycling through chromatin. *Biochimica Et Biophysica Acta-Gene Structure and Expression* **1677**:64-73.

- Sydow, J. F., F. Brueckner, A. C. Cheung, G. E. Damsma, S. Dengl, E. Lehmann, D. Vassilyev, and P. Cramer. 2009.** Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. *Mol Cell* **34**:710-21.
- Sydow, J. F., and P. Cramer. 2009.** RNA polymerase fidelity and transcriptional proofreading. *Curr Opin Struct Biol* **19**:732-9.
- Taatjes, D. J. 2010.** The human Mediator complex: a versatile, genome-wide regulator of transcription. *Trends Biochem Sci* **35**:315-22.
- Thomas, M. C., and C. M. Chiang. 2006.** The general transcription machinery and general cofactors. *Critical Reviews in Biochemistry and Molecular Biology* **41**:105-178.
- Thomas, M. J., A. A. Platas, and D. K. Hawley. 1998.** Transcriptional fidelity and proofreading by RNA polymerase II. *Cell* **93**:627-37.
- Thuillier, V., I. Brun, A. Sentenac, and M. Werner. 1996.** Mutations in the alpha-amanitin conserved domain of the largest subunit of yeast RNA polymerase III affect pausing, RNA cleavage and transcriptional transitions. *EMBO J* **15**:618-29.
- Todone, F., P. Brick, F. Werner, R. O. Weinzierl, and S. Onesti. 2001.** Structure of an archaeal homolog of the eukaryotic RNA polymerase II RPB4/RPB7 complex. *Mol Cell* **8**:1137-43.
- Treich, I., M. Riva, and A. Sentenac. 1991.** Zinc-binding subunits of yeast RNA polymerases. *J Biol Chem* **266**:21971-6.
- Tschochner, H. 1996.** A novel RNA polymerase I-dependent RNase activity that shortens nascent transcripts from the 3' end. *Proc Natl Acad Sci U S A* **93**:12914-9.
- Van Mullem, V., E. Landrieux, J. Vandenhaute, and P. Thuriaux. 2002.** Rpa12p, a conserved RNA polymerase I subunit with two functional domains. *Mol Microbiol* **43**:1105-13.
- Van Mullem, V., M. Wery, M. Werner, J. Vandenhaute, and P. Thuriaux. 2002.** The Rpb9 subunit of RNA polymerase II binds transcription factor TFIIE and interferes with the SAGA and elongator histone acetyltransferases. *J Biol Chem* **277**:10220-5.
- Vassilyev, D. G., S. Sekine, O. Laptenko, J. Lee, M. N. Vassilyeva, S. Borukhov, and S. Yokoyama. 2002.** Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* **417**:712-9.
- Vassilyeva, M. N., V. Svetlov, A. D. Dearborn, S. Klyuyev, I. Artsimovitch, and D. G.**

- Vassilyev. 2007.** The carboxy-terminal coiled-coil of the RNA polymerase beta'-subunit is the main binding site for Gre factors. *EMBO Rep* **8**:1038-43.
- Walmacq, C., M. L. Kireeva, J. Irvin, Y. Nedialkov, L. Lubkowska, F. Malagon, J. N. Strathern, and M. Kashlev. 2009.** Rpb9 subunit controls transcription fidelity by delaying NTP sequestration in RNA polymerase II. *J Biol Chem* **284**:19601-12.
- Wang, B., D. N. M. Jones, B. P. Kaine, and M. A. Weiss. 1998.** High-resolution structure of an archaeal zinc ribbon defines a general architectural motif in eukaryotic RNA polymerases. *Structure with Folding & Design* **6**:555-569.
- Wang, D., D. A. Bushnell, X. H. Huang, K. D. Westover, M. Levitt, and R. D. Kornberg. 2009.** Structural Basis of Transcription: Backtracked RNA Polymerase II at 3.4 Angstrom Resolution. *Science* **324**:1203-1206.
- Wang, D., D. A. Bushnell, K. D. Westover, C. D. Kaplan, and R. D. Kornberg. 2006.** Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell* **127**:941-54.
- Wang, D., and D. K. Hawley. 1993.** Identification of a 3'→5' exonuclease activity associated with human RNA polymerase II. *Proc Natl Acad Sci U S A* **90**:843-7.
- Weilbaecher, R. G., D. E. Awrey, A. M. Edwards, and C. M. Kane. 2003.** Intrinsic transcript cleavage in yeast RNA polymerase II elongation complexes. *J Biol Chem* **278**:24189-99.
- Werner, F. 2007.** Structure and function of archaeal RNA polymerases. *Mol Microbiol* **65**:1395-404.
- Werner, F., and R. O. Weinzierl. 2002.** A recombinant RNA polymerase II-like enzyme capable of promoter-specific transcription. *Mol Cell* **10**:635-46.
- Werner, F., and R. O. Weinzierl. 2005.** Direct modulation of RNA polymerase core functions by basal transcription factors. *Mol Cell Biol* **25**:8344-55.
- Westover, K. D., D. A. Bushnell, and R. D. Kornberg. 2004.** Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell* **119**:481-9.
- Whitehall, S. K., C. Bardeleben, and G. A. Kassavetis. 1994.** Hydrolytic cleavage of nascent RNA in RNA polymerase III ternary transcription complexes. *J Biol Chem* **269**:2299-306.
- Williams, L. A., and C. M. Kane. 1996.** Isolation and characterization of the

- Schizosaccharomyces pombe gene encoding transcript elongation factor TFIIIS. *Yeast* **12**:227-36.
- Wilson, C. J., D. M. Chao, A. N. Imbalzano, G. R. Schnitzler, R. E. Kingston, and R. A. Young. 1996.** RNA polymerase II holoenzyme contains SWI/SNF regulators involved in chromatin remodeling. *Cell* **84**:235-44.
- Wind, M., and D. Reines. 2000.** Transcription elongation factor SII. *Bioessays* **22**:327-36.
- Woese, C. R., and G. E. Fox. 1977.** Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* **74**:5088-90.
- Woese, C. R., O. Kandler, and M. L. Wheelis. 1990.** Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* **87**:4576-9.
- Woychik, N. A., W. S. Lane, and R. A. Young. 1991.** Yeast RNA polymerase II subunit RPB9 is essential for growth at temperature extremes. *J Biol Chem* **266**:19053-5.
- Wu, J., D. E. Awrey, A. M. Edwards, J. Archambault, and J. D. Friesen. 1996.** In vitro characterization of mutant yeast RNA polymerase II with reduced binding for elongation factor TFIIIS. *Proc Natl Acad Sci U S A* **93**:11552-7.
- Yee, N. S., W. Gong, Y. Huang, K. Lorent, A. C. Dolan, R. J. Maraia, and M. Pack. 2007.** Mutation of RNA Pol III subunit *rpc2/polr3b* Leads to Deficiency of Subunit *Rpc11* and disrupts zebrafish digestive development. *PLoS Biol* **5**:e312.
- Yuzenkova, Y., and N. Zenkin. 2010.** Central role of the RNA polymerase trigger loop in intrinsic RNA hydrolysis. *Proc Natl Acad Sci U S A* **107**:10878-83.
- Zenkin, N., Y. Yuzenkova, and K. Severinov. 2006.** Transcript-assisted transcriptional proofreading. *Science* **313**:518-20.
- Zhang, G., E. A. Campbell, L. Minakhin, C. Richter, K. Severinov, and S. A. Darst. 1999.** Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* **98**:811-24.
- Zhu, W., Q. Zeng, C. M. Colangelo, M. Lewis, M. F. Summers, and R. A. Scott. 1996.** The N-terminal domain of TFIIIB from *Pyrococcus furiosus* forms a zinc ribbon. *Nat Struct Biol* **3**:122-4.
- Zillig, W., K. O. Stetter, and D. Janekovic. 1979.** DNA-dependent RNA polymerase from the archaeobacterium *Sulfolobus acidocaldarius*. *Eur J Biochem* **96**:597-604.

CURRICULUM VITAE – Wenjie Ruan

PERSONAL DETAILS

Name: Wenjie Ruan
Date of birth: September 21, 1984
Place of Birth: Anhui. P.R. China
Nationality: Chinese

EDUCATION

1989-1994: Primary School in Changfeng. Anhui. P.R. China
1994-2000: Junior and senior Middle School in Changfeng. Anhui. P.R. China
2000-2004: University of Science and Technology of China(USTC)
09/2004: **Bachelor degree**
2004-2006: Graduate University of Chinese Academy of Sciences
Institute of Biophysics, Chinese Academy of Sciences
2006- present: **Ph.D studies**, Prof. Patrick Cramer Group
Gene Center
Ludwig-Maximilians University Munich. Germany
Topic: “Evolution of two modes of intrinsic RNA polymerase transcript cleavage”
2006- present: **Member** of the International Max-Planck Research
School of Life Science(IMPRS)
Munich, Germany
05/2011 **Ph.D thesis defense**

WORKSHOPS AND CONFERENCE

2007 2nd Transregio 5 Symposium Chromatin “Assembly and Inheritance of Functional States”
Munich, Germany
participant
2007 IMPRS workshop “Effective scientific writing” Avril
Arthur Goettig
2008 IMPRS workshop “High performance Presentation”
Ruth Willmott

PUBLICATION

Ruan, W., Lehmann, E., Thomm, M., Kostrewa, Cramer P. (2011)
Evolution of two modes of intrinsic RNA polymerase transcript cleavage.
J Biol Chem. *Paper in Press*. doi: 10.1074/jbc.M111.222273.