

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

Architecture of the RNA polymerase II- Paf1C-TFIIS transcription elongation complex



Youwei Xu

aus

Gaoyou, Jiangsu Provinz, P.R.China

2016

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

Architecture of the RNA polymerase II- Paf1C-TFIIS transcription elongation complex



Youwei Xu

aus

Gaoyou, Jiangsu Provinz, P.R.China

2016

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Patrick Cramer betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

Göttingen, den 28.11.2016

Youwei Xu

Dissertation eingereicht am 01.12.2016

1. Gutachter: Prof. Dr. Patrick Cramer

2. Gutachter: PD Dr. Dietmar Martin

Mündliche Prüfung am 25.01.2017

"My apologies to great questions for small answers."

- Wislawa Szymborska

Acknowledgements

Firstly, I would like to express my sincere gratitude to Patrick Cramer for giving me such a great opportunity to study in this lab. This is an excellent lab circumstanced by outstanding scientific atmosphere. I still remember that you asked me whether I had further questions for you at the end of our interview via the phone, I asked how about the weather in Munich. I feel very lucky being your student. You always gave me fantastic suggestions, sufficient support, and enough patience. I learned a lot from you, and your enthusiasm for science and life inspires me.

I am most indebted to my past and present colleagues of the Cramer group from Munich and Göttingen for the great assistance, stimulating discussions, advice, encouragement, and joyful life, especially the following people: Carrie Bernecky, Christoph Engel, Simon Neyer, Christian Dienemann, Clemens Plaschka, Sarah Sainsbury, Merle Hantsche, Anna Sawicka, Jinmi Choi, Carina Demel, Margaux Michel, Tobias Gubbey, Dimitry Tengunov, and Sara Osman. It is a precious fortune to work with you.

In particular, I also want to thank Mai Sun, Claudia Buchen, and Thomas Wild for their patient help at the beginning of my PhD, Kerstin C. Maier and Björn Schwalb for the 4tU-Seq.

Thanks to the collaborators, Chung-Tien Lee and Henning Urlaub for the mass spectrometry analysis.

Thanks to Jürgen Plitzko from the Baumeister department at the Max Plank Institute in Martinsried for the supervision and technical support during the cryo-EM data collection.

Thanks to the members of my thesis committee: Dr. Dietmar Martin, Prof. Dr. Andreas Ladurner, Prof. Dr. Karl-Peter Hopfner, Dr. Franz Herzog, and Prof. Dr. Roland Beckmann.

I am greatly thankful to my friends, Rui Gong, Jingdong Cheng, Ping Wang, and Yi Liu for their scientific news discussion.

I am obliged to my parents and sister for their warmth and unwavering support.

Last but not least, I don't know how to thank my lovely girlfriend, Qianmin Wang for her patience, understanding, kindness, and grace. I will always love you, Qianmin.

Summary

Transcription of cellular genomes is the first step of gene expression, which is carried out by DNA-dependent RNA polymerases (Pols). Pol II transcribes protein-coding genes into pre-mRNA with the assistance of accessory factors. The conserved Paf1 complex (Paf1C) plays multiple roles in chromatin transcription and genomic regulation. Paf1C consists of the five subunits Paf1, Leo1, Ctr9, Cdc73, and Rtf1, and binds to the Pol II transcription elongation complex (EC).

In this thesis I establish preparation of pure and recombinant Paf1C from the yeast *Saccharomyces cerevisiae* after co-expression of its subunits in *Escherichia coli*. Using a combination of Edman sequencing and iterative truncation, I obtained a structured core of Paf1C. The purified Paf1C binds to Pol II in a substoichiometric manner *in vitro*, and another transcription elongation factor TFIIS enhances this binding via its N-terminal domain.

Here, I report the cryo-electron microscopic structure of the ~1 megadalton yeast Pol II-Paf1C-TFIIS EC. Protein crosslinking data, which together revealed that Paf1C extends over the Pol II surface from lobe domain of Rpb2 to subunit Rpb3. The Paf1-Leo1 heterodimer and Cdc73 contact Rpb2 and Rpb3, respectively, and from opposite ends of Paf1C, whereas Ctr9 bridges between them. My collaborators also show that Paf1C is globally required for mRNA transcription in yeast. I used 4tU-Seq to monitor the newly synthesized RNA. Comparing RNA synthesis in the wild-type strain and in strains lacking Paf1 or Rtf1, we found that Paf1C affects Pol II transcription through the gene body. Moreover, in the absence of the subunit Rtf1, Paf1C binding to RNA was dramatically reduced *in vitro*, indicating that Paf1C regulates transcription elongation also through protein-RNA interactions. These results further indicate why Paf1C binds only after disassembly of the transcription initiation complex; the initiation factor TFIIF impairs Paf1C binding to Pol II.

These results provide the first three-dimensional framework for analyzing Paf1C function in chromatin transcription and transcription-coupled events.

Publication

Parts of this thesis have been submitted and are in the process of being published:

Architecture of the RNA polymerase II-Paf1C-TFIIS transcription elongation complex

Youwei Xu, Carrie Bernecky, Chung-Tien Lee, Kerstin C. Maier, Björn Schwalb, Dimitry Tegenov, Jürgen M. Plizko, Henning Urlaub, and Patrick Cramer

Author contributions: Y.X. designed and performed experiments, unless stated otherwise. C.B. helped with EM data collection. J.M.P. provided access to the EM facility and supervised EM data collection. C.L. performed MS analysis. H.U. supervised MS analysis. K.C.M. carried out the 4tU-seq experiment. B.S. analyzed 4tU-Seq data. C.B. and D.T. helped with EM data processing. P.C. designed and supervised research. Y.X. and P.C. prepared the manuscript, with the input from all authors.

Contents

Erklärung	II
Eidesstattliche Versicherung.....	III
Acknowledgements.....	VI
Summary	VII
Publication	VIII
Contents	IX
1. Introduction.....	1
1.1 DNA-dependent RNA polymerases	1
1.2 Regulation of RNA polymerase II.....	3
1.2.1 The RNA polymerase II transcription cycle.....	3
1.2.2 General transcription factors in yeast.....	4
1.2.3 The Carboxy-terminal domain (CTD).....	6
1.2.4 General transcription factor TFIIIS	7
1.2.5 Transcription elongation factors.....	8
1.2.6 Structural studies on Pol II.....	10
1.3 Transcription factor Paf1 complex	11
1.3.1 Multiple functions of Paf1 complex in chromatin transcription	11
1.3.2 The recruitment of Paf1 complex during transcription	12
1.3.3 Structural studies on Paf1 complex	12
1.4 Aims and scope.....	14
2. Materials and Methods.....	15
2.1 Materials	15
2.1.1 Bacterial strains.....	15
2.1.2 Yeast strains	15
2.1.3 Plasmids	16
2.1.4 Synthetic Genes and oligonucleotides.....	22
2.1.5 Growth medium and additives	24
2.1.6 Buffers and solutions.....	25
2.2 Methods	30
2.2.1 Molecular cloning	30
2.2.2 General protein methods	32

2.2.3 Specific methods	34
2.2.4 Bioinformatics tools	43
3. Results and Discussion	45
3.1 Recombinant Paf1 complex (Paf1C)	45
3.2 Paf1C structural core and flexible periphery	46
3.3 Paf1C and TFIIS bind Pol II synergistically.....	48
3.4 Cryo-EM analysis	50
3.5 Architecture of Pol II-Paf1C-TFIIS elongation complex	55
3.6 Crosslinking analysis.....	57
3.7 Paf1C subcomplex architecture	60
3.8 Competitive Pol II binding of Paf1C and TFIIIF.....	61
3.9 Paf1C is globally required for Pol II transcription	63
3.10 Paf1C binds RNA independent of Rtf1	64
4. Conclusion and Outlook	65
References.....	69
Appendix.....	81
Abbreviations.....	99
List of Figures.....	103
List of Tables	105

1. Introduction

1.1 DNA-dependent RNA polymerases

All forms of life on earth use nucleic acids (NA) to store their genetic information. With the exception of some viruses that utilize ribonucleic acids (RNA), the common biomolecule for information storage is deoxyribonucleic acid (DNA). However, the information of DNA sequences cannot be converted to proteins directly, but depends on interpretative RNA to transport the information. The ‘Central Dogma of Molecular Biology’ defines this flow of genetic information from DNA via messenger RNA (mRNA) to protein involving the processes of transcription and translation. Transcription of genetic information from DNA to RNA is catalyzed by DNA-dependent RNA polymerases (Pols) and can be found in all three kingdoms of life. Bacteria and archaea have one Pol whereas eukaryotes use at least three different types of Pols that synthesize distinct subsets of RNA molecules. Pol I is located in nucleoli and sequentially transcribes 28S, 5.8S, and 18S ribosomal RNAs (rRNAs), ensuring synthesis of equimolar amounts of these three rRNAs. Pol II and Pol III are located in the nucleoplasm. Pol II transcribes all protein-coding messenger RNAs (mRNAs) as well as several small RNAs. Pol III synthesizes all the transfer RNAs (tRNAs), 5S rRNA, U6 small nuclear RNAs (snRNA) and other small cellular RNAs. In plants, two additional Pols called Pol IV and Pol V exist that transcribe small interfering RNAs (siRNAs) required for heterochromatic silencing (Lahmy *et al.*, 2010; Matzke *et al.*, 2009). Whereas the Pols mentioned above are multisubunit enzymes (Table1), Pols in mitochondria and bacteriophages are single-subunit enzymes. Different from mitochondria, chloroplasts transcription in higher plants is carried by two types of Pols. Plastid-encoded Pol (PEP) is a bacteria-type multisubunit enzyme and transcribes photosynthesis genes whereas nuclear-encoded Pol is a bacteriophage-type single-subunit enzyme and transcribes a few house-keeping genes (Borner *et al.*, 2015; Kanamaru & Tanaka, 2004).

Table 1: Subunit composition of multisubunit RNA polymerases. Adapted from (Ream *et al.*, 2009; Vannini & Cramer, 2012; Weinzierl, 2013).

	Pol I	Pol II	Pol III	Pol IV	Pol V	Archaea	Bacteria	PEP
Core	A190	Rpb1	C160	Nrpd1	Nrpe1	A'+A''	β'	$\beta' + \beta''$
	A135	Rpb2	C128	Nrpd2	Nrpd2	B'+B''	β	β
	AC40	Rpb3	AC40	Nrpb3	Nrpb3/ Nrpe3b	D	α	α
	AC19	Rpb11	AC19	Nrpd11	Nrpd11	L	α	α
	A12.2	Rpb9	C11	Nrpb9b	Nrpb9a/ Nrpb9b	-	-	-
	Rpb5	Rpb5	Rpb5	Nrpb5	Nrpe5	H	-	-
	Rpb6	Rpb6	Rpb6	Nrpb6	Nrpb6	K	ω	-
	Rpb8	Rpb8	Rpb8	Nrpb8	Nrpb8	(G)	-	-
	Rpb10	Rpb10	Rpb10	Nrpb10	Nrpb10	N	-	-
	Rpb12	Rpb12	Rpb12	Nrpb12	Nrpb12	P	-	-
Stalk	A14	Rpb4	C17	Nrpd4	Nrpd4	E	-	-
	A43	Rpb7	C25	Nrpd7	Nrpe7	F	-	-
TFIIF-like	A49	(Tfg1)	C37	?	?	-	-	-
	A34.5	(Tfg2)	C53	?	?	-	-	-
Pol III specific	-	-	C82	-	-	-	-	-
	-	-	C34	-	-	-	-	-
	-	-	C31	-	-	-	-	-
Subunits	14	12	17	12	12	12 (13)	5	5

Among multisubunit complexes, eukaryotic Pol I, Pol II, and Pol III contain 14, 12, and 17 subunits, respectively (Table 1) (Vannini & Cramer, 2012). Five of core subunits are shared between Pol I and Pol II and two more between Pol I and Pol III (Table 1). The structure and function of eukaryotic Pol I, Pol II, and Pol III has been well studied in the last decades (Cramer *et al.*, 2008). The first crystal structure of the 10-subunit Pol II core from *Saccharomyces cerevisiae* (*S. cerevisiae*) was solved in 2000 (Cramer *et al.*, 2000) and was later extended to a complete 12 subunits with the peripheral Rpb4/7 stalk subcomplex (Armache *et al.*, 2003; Bushnell & Kornberg, 2003). The crystal structure of yeast Pol I was solved in two independent studies (Engel *et al.*, 2013; Fernandez-Tornero *et al.*, 2013) and most recently, cryo-electron

microscopy (cryo-EM) structures of yeast Pol III were determined at medium-resolution (Hoffmann *et al.*, 2015). Single particle cryo-EM also allowed the first visualization of mammalian Pol II at high resolution (Bernecky *et al.*, 2016). Although eukaryotic Pol I, Pol II, and Pol III have different composition, molecular weight and transcribe distinct subsets of genes, the structural core and enzymatic mechanism is greatly conserved apparent from biochemical studies and comparison of the structural information available (Cramer, 2002; Vannini & Cramer, 2012)

1.2 Regulation of RNA polymerase II

1.2.1 The RNA polymerase II transcription cycle

The transcription cycle of Pol II can be intuitively divided into five steps: pre-initiation, initiation, elongation, termination and recycling (Hahn, 2004; Svejstrup, 2004) (Figure 1). During pre-initiation Pol II assembles with numerous general transcription factors (GTFs) to form a giant pre-initiation complex (PIC) localized at promoter sequences comprising a TATA box on closed, double-stranded DNA. Once assembly of the PIC is complete, the closed, double-stranded DNA is melted and the transcription bubble is formed. The single-stranded template DNA is inserted into the Pol II active site where Pol II initiates transcription by synthesis of a nascent RNA (X. Liu *et al.*, 2013; Luse, 2013; Sainsbury *et al.*, 2015). When the initially transcribed RNA reaches at a length of ~13 nucleotides, transcription switches to elongation by clearing the promoter and exchanging initiation factors with elongation factors (Pal *et al.*, 2005). During the elongation step, the transcription bubble is retained (Martinez-Rucobo & Cramer, 2013). When the transcription reaches the 3' termini of genes, transcript cleavage and poly-adenylation (pA) occur (Proudfoot *et al.*, 2002; Svejstrup, 2004). Newly synthesized RNAs are released from Pol II at the pA site while transcribing Pol II is terminated shortly after (Arndt & Reines, 2015; Schwalb *et al.*, 2016). Already co-transcriptionally, the pre-mRNA undergoes further processing (splicing) while terminated Pol II can re-initiate for another transcription event.

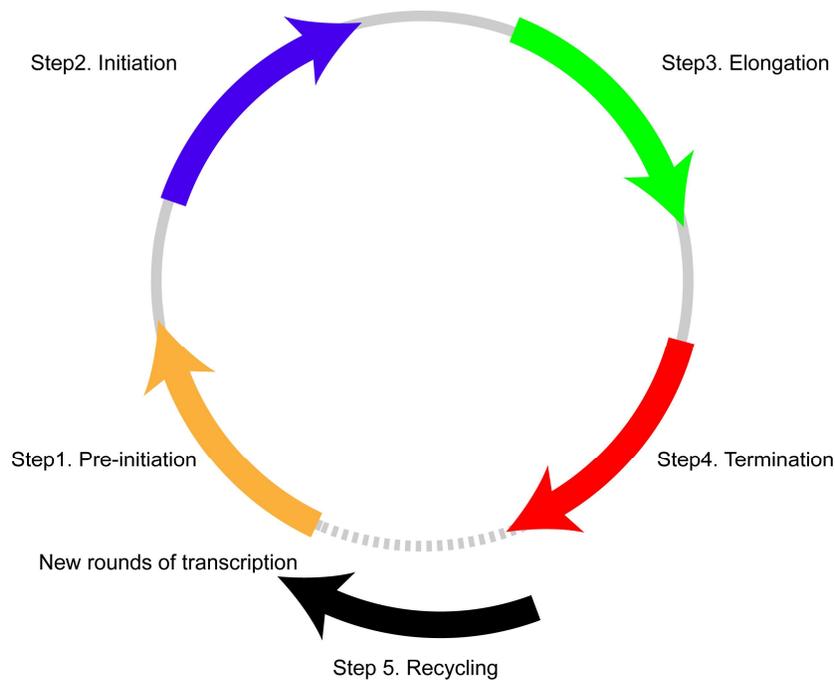


Figure 1: Pol II transcription cycle. Adapted from (Blombach *et al.*, 2013). The five main steps are in different colors. The transition from one step to the next is always accompanying the exchanges of factors.

Each step of the Pol II transcription cycle is highly regulated by a large number of factors with many types of interactions. Furthermore, specific secondary structures of nucleic acids (Kassube *et al.*, 2013; Lehmann *et al.*, 2007) and post-translational modifications (PTMs) of the Rpb1 carboxy-terminal domain (CTD) (Bowman & Kelly, 2014; Buratowski, 2009; Eick & Geyer, 2013) play critical roles in transcription regulation.

1.2.2 General transcription factors in yeast

General transcription factors TFIIA, -B, -D, -E, -F, -H tightly regulate transcription initiation *in vivo* (Table 2) (Sainsbury *et al.*, 2015). The first step of pre-initiation complex assembly is the binding of the saddle-shaped TFIID subunit TATA binding protein (TBP) to the TATA box region of promoter DNA, which is located around 30-120 bp upstream of the transcription start site (TSS). The minor groove of the TATA-box is bound by the concave surface of TBP, which bends the DNA double helix by nearly 90° (J. L. Kim *et al.*, 1993; Y. Kim *et al.*, 1993). This

allows the C-terminal core domain of transcription factor IIB (TFIIB) to bind to TATA-box and surrounding upstream and downstream sequences in a cooperative manner (Kostrewa *et al.*, 2009; Littlefield *et al.*, 1999; Sainsbury *et al.*, 2013). The dispensable transcription factor TFIIA is not required for basal transcription but can stabilize the TBP-DNA complex via binding to the upstream region of the TATA box (Geiger *et al.*, 1996). The N-terminal domain of TFIIB can now recruit Pol II to the promoter. Binding of TFIIF to Pol II stabilizes the interaction with TFIIB and prevents non-specific interactions with DNA (Cabart *et al.*, 2011; Fishburn & Hahn, 2012). In later steps of transcription initiation, TFIIF stabilizes the transcription bubble and helps the early RNA synthesis (Sainsbury *et al.*, 2015). TFIIE is then recruited to the Pol II-DNA-TBP-TFIIB-TFIIF complex, stabilizes it and promotes DNA melting. TFIIE also enhances the affinity of TFIIH to the pre-initiation complex. Whereas TFIIE alone can already open certain promoter DNA (Plaschka *et al.*, 2016), the Ssl2 subunit of TFIIH unwinds the promoter DNA in an ATP-dependent manner (Holstege *et al.*, 1997; Holstege *et al.*, 1996). Another important function of TFIIH is phosphorylating the CTD via its Kin28 subunit. TBP, TFIIB, and TFIIF but not TFIIE or TFIIH are minimally required for the transcription initiation *in vitro* (Fishburn & Hahn, 2012).

Table 2: General transcription factors in yeast. Adapted from (Sainsbury *et al.*, 2015; Sikorski & Buratowski, 2009)

Protein complex	Subunits	Function
TFIIA	2	stabilizes the TBP-DNA complex; counteracts repressive effects of negative co-factors
TFIIB	1	stabilizes the TBP-DNA complex; aids in the recruitment of Pol II and TFIIF; directs accurate start site selection
TFIID	14	including TBP and TBP Associated Factors (TAFs); recognizes the promoters; recruits Pol II
TFIIE	2	stabilizes the DNA opening; recruits TFIIH
TFIIF	3	stabilizes TFIIB; directs accurate start site selection
TFIIH	10	promoters opening; DNA repair; phosphorylates CTD
TFIIS	1	stimulates the transcription elongation; induces mRNA cleavage and resume transcription after arrest

1.2.3 The Carboxy-terminal domain (CTD)

The regulation of Pol II transcription by GTFs is only the initial point of transcriptional regulation. Another important regulator for the transition between different stages is the phosphorylation state of specific residues within the extended CTD of the largest Pol II subunit Rpb1. The CTD is composed of dozens of tandem heptapeptide repeats with a consensus sequence of tyrosine-serine-proline-threonine-serine-proline-serine (Y₁-S₂-P₃-T₄-S₅-P₆-S₇). Although very important in the Pol II system, Pol I and Pol III do not have a repetitive CTD in their largest subunits Rpa1 and Rpc1, respectively. The number of hepta-repeats in the CTD of Pol II may vary between different species. While yeast Pol II only has 26 repeats, mammalian Pol II possesses a CTD with 52 repeats. However, the length of CTD seems not to be related to the complexity of the organism (Eick & Geyer, 2013). Throughout the entire transcription cycle

including pre-initiation, initiation, pausing, capping, elongation, selection of the RNA 3' processing site and termination, the CTD is involved in regulatory events. Depending on the functional context within the transcription process, all residues of the CTD repeats can be dynamically modified and thereby regulate the recruitment of specific factors that recognize the modifications. Tyrosine (Y₁), threonine (T₄), and serines (S₂, S₅, and S₇) can be phosphorylated, while the prolines (P₃ and P₇) can be isomerized. In addition to the phosphorylation, S₅ and S₇ can be *O*-glycosylated. Of these reversible modifications, S₂ and S₅ phosphorylation are the most frequent and best-studied modifications. Briefly, phosphorylation on S₅ was found in proximity to promoter regions of genes showing its function in transcription initiation, whereas S₂ was found within the gene body according to productive elongation (Komarnitsky *et al.*, 2000; Mayer *et al.*, 2010). The enzymes leading to the dynamic changes of those residues are kinases and phosphatases, the so-called 'writers' and 'erasers', respectively. Factors that have a CTD interaction domains (CID), which can recognize the modification patterns, are then 'readers' of the established CTD code (Jeronimo *et al.*, 2013). Furthermore, the CTD crosstalks with the processing of epigenetic marks in chromatin by recruiting chromatin modifying proteins (Eick & Geyer, 2013).

1.2.4 General transcription factor TFIIS

Different from the GTFs described above, which are mainly required for transcription initiation, another GTF TFIIS is required during the transcription elongation (Table 2). TFIIS was first discovered in 1972 (Natori *et al.*, 1973) and is a single subunit protein, which consists of Domain I (amino acid 1-77), Domain II (amino acid 148-238), Domain III (amino acid 265-309) and a linker (amino acid 238-265) between Domain II and III (Kettenberger *et al.*, 2003). Multiple functions are described for TFIIS including binding to Pol II (Shimoaraiso *et al.*, 1997) and stimulation of the transcription elongation (Agarwal *et al.*, 1991)

Most importantly, TFIIS triggers cleavage of nascent, backtracked mRNA when the enzyme arrests during transcription (Kettenberger *et al.*, 2003). Pol II arrests and backtracks either because of nucleotide misincorporation or at sequences that are difficult to transcribe. In such state, TFIIS is recruited, inserts its Domain III into the Pol II pore and cleaves the displaced backtracked RNA. Cleavage activity of TFIIS is facilitated by two acidic residues D290 and

E291 that can coordinate metal ion B in the Pol II active site and thereby stimulate transcript cleavage (Cheung & Cramer, 2011; Kettenberger *et al.*, 2003).

TFIIS is a specific factor for Pol II transcription, but the Pol I/III subunits A12.2 and C11 are functionally and structurally related to TFIIS (Arimbasseri *et al.*, 2013; Chedin *et al.*, 1998; Engel *et al.*, 2013; Hausner *et al.*, 2000; Jennebach *et al.*, 2012; Lisica *et al.*, 2016).

1.2.5 Transcription elongation factors

Both structural and functional basis of transcription initiation have been well studied in the past two decades. However, elongation is becoming increasingly considered as a key stage in eukaryotic transcriptional regulation. The transition initiation to elongation occurs once the nascent RNA transcript is extended beyond ~13 nucleotides (Sainsbury *et al.*, 2015). This transcript length is predicted to clash with the TFIIB ribbon domain at the RNA exit tunnel ultimately inducing promoter clearance and the transition to elongation. As during initiation, many factors are involved in regulating transcription elongation (Table 3). Before Pol II goes into processive elongation, the early RNA synthesis steps are always accompanied by events as promoter proximal pausing, backtracking, arrest, and reactivation, which are rate limiting steps for transcription of a gene. Promoter proximal pausing happens shortly after initiation and was found in metazoans. 5,6-dichloro-1- β -D-ribofuranosylbenzimidazole (DRB) sensitivity inducing factor (DSIF) and negative elongation factor (NELF) were implicated in pausing (C. H. Wu *et al.*, 2003; Yamaguchi *et al.*, 2013). Furthermore, Gdown1 is considered to stabilize poised Pol II in the pausing state (Cheng *et al.*, 2012). The paused transcription machinery can be released by the positive transcription elongation factor b (P-TEFb), a subunit of the super elongation complex (SEC) by phosphorylating NELF, DSIF and the CTD.

Another member of the SEC, the factor eleven-nineteen lysine-rich leukemia (ELL) also increases the elongation rate. Other subunits of the SEC regulate elongation by direct interactions or/and by serving as a docking platform for more factors (Luo *et al.*, 2012).

Nucleosomes residing within gene bodies may impede Pol II during transcription elongation and need to be overcome in order to maintain high elongation rates (~ 4 kb/min). The tight complex of DNA wound around the core histone octamer (two molecules of each H2A, H2B, H3, and H4) can be resolved by many different chromatin remodelers. Chd1, IWS1, Swi, and Ino80, use the energy from ATP hydrolysis to change the nucleosome position, which

increases accessibility for the transcription machinery. Histone chaperones as FACT and Spt6 directly bind to H2A-H2B and H3-H4, respectively, to disassemble histones from the nucleosome particle and thereby enable Pol II to transcribe through without displacing the entire nucleosome (Belotserkovskaya & Reinberg, 2004; Belotserkovskaya *et al.*, 2004). These histone chaperones also regulate the enzymatic activity of histone tail modifying enzymes or serve as interaction bridges between histones and PTM adding enzymes. The dynamically added and removed post-translational modifications of histone tails like methylation, phosphorylation, sumoylation, ubiquitination, acetylation, and glycosylation also influence transcription elongation rate positively or negatively (Kwak & Lis, 2013; W. H. Liu & Churchill, 2012; Petesch & Lis, 2012; Tessarz & Kouzarides, 2014). Elongation factors also can be recruited or dissociated from elongating Pol II facilitating modified histone tails (Guccione *et al.*, 2007; Steward *et al.*, 2006; J. Wu & Xu, 2012).

Table 3: A list of transcription elongation factors. Adapted from (Kwak & Lis, 2013).

Protein complex	Subunits	Function
NELF	4	stabilizes Pol II pausing
DSIF	2	stabilizes Pol II pausing; facilitates elongation
P-TEFb	2	phosphorylates NELF, DSIF, and CTD of Pol II to release the elongation from pause
ELL	1	increases the elongation rate
SEC	5-10	including P-TEFb and ELL; serves as a platform for other elongation factors
CE	2	capping at RNA 5' end
Gdown1	1	stabilizes paused Pol II
FACT	2	H2A-H2B chaperone
Spt6	1	H3-H4 chaperone
Chd1	1	chromatin remodeler
Paf1C	5	binds to Pol II; serves as a platform for other elongation factors; facilitates the chromosomal transcription elongation, <i>et.al.</i> (see below)

1.2.6 Structural studies on Pol II

The first crystal structure of the 10-subunit Pol II core was observed in 2000 and provided an initial architecture of the enzyme (Cramer *et al.*, 2000). Later, this was extended to a complete 12-subunit crystal structure, with the additional Rpb4-Rpb7 stalk sub-complex (Armache *et al.*, 2003; Bushnell & Kornberg, 2003). X-ray crystallographic structures of complete Pol II in complex with factors or/and nucleic acids provide detailed insights in the mechanisms of RNA synthesis by Pol II and how various factors assist and regulate this process. Structures of Pol II-TFIIS in complex with a scaffold including a transcription bubble and a short RNA product show how TFIIS binds to Pol II and facilitates RNA cleavage to release Pol II from arrested and backtracked states (Cheung & Cramer, 2011; Kettenberger *et al.*, 2003, 2004). The crystal structure of Pol II-TFIIB with DNA and very short RNA transcripts, which mimic an initially transcribing complex, revealed how TFIIB interacts with Pol II, positions DNA in the active site, assists in transcription start site selection, helps DNA opening at the promoter of genes and finally triggers promoter clearance. Another crystal structure of Pol II in complex with Bye1, a homolog to TFIIS, shows that it binds to Pol II in a similar manner as TFIIS (Kinkelin *et al.*, 2013). However, different from TFIIS, Bye1 has a plant homeodomain (PHD), which can recognize histone H3 containing trimethylated lysine 4 (H3K4) (Shi *et al.*, 2007), and suggests Bye1 to be involved in transcribing packed chromatin in yeast (Kinkelin *et al.*, 2013).

Recently, cryo-electron microscopy (cryo-EM) is becoming more and more powerful for determining structures of biological macromolecules and in particular further elucidating the mechanism of Pol II transcription at near-atomic resolution.

Medium resolution structures revealed the overall architectures of large transcription related assemblies (He *et al.*, 2013; Martinez-Rucobo *et al.*, 2015; Plaschka *et al.*, 2015; Robinson *et al.*, 2016). Bernecky *et al.* first determined the structure of a mammalian Pol II at 3.4 Å resolution using single particle cryo-EM, which ended the low-resolution electron microscopy analysis of mammalian Pol II and its complexes (Bernecky *et al.*, 2016). High resolution studies of different yeast and human initiation complexes revealed conformational changes accompanying the assembly of the initiation complex and suggested mechanisms for promoter opening (He *et al.*, 2016; Plaschka *et al.*, 2016).

1.3 Transcription factor Paf1 complex

The polymerase-associated factor 1 (Paf1) complex (Paf1C) is a general and conserved RNA polymerase (Pol) II transcription elongation factor (Tomson & Arndt, 2013). Paf1C was first identified through its co-purification with Pol II from yeast cells (Shi *et al.*, 1997; Shi *et al.*, 1996; Wade *et al.*, 1996). Yeast Paf1C comprises the subunits Paf1, Leo1, Ctr9, Cdc73, and Rtf1 (Koch *et al.*, 1999; Krogan *et al.*, 2002; Mueller & Jaehning, 2002; Squazzo *et al.*, 2002). Paf1C shows genetic interactions with the yeast transcription elongation factors Spt4-Spt5 and Spt16-Pob3, the counterparts of human DSIF and FACT, respectively (Squazzo *et al.*, 2002). Paf1C also associates with transcribed regions *in vivo* (Pokholok *et al.*, 2002), suggesting that it is a transcription elongation factor. Paf1C subunits are required for efficient transcription *in vivo* (Rondon *et al.*, 2004).

1.3.1 Multiple functions of Paf1 complex in chromatin transcription

Paf1C has multiple roles in chromatin transcription. Yeast Paf1C functions in methylation of histone H3 by Set1 and Dot1, thus linking transcription elongation to chromatin methylation (Krogan *et al.*, 2003). In yeast, Rtf1 binds the chromatin remodeler Chd1 (Simic *et al.*, 2003) and is required for ubiquitination of histone H2B (Ng *et al.*, 2003; Wood *et al.*, 2003; Xiao *et al.*, 2005) and histone methylation (Warner *et al.*, 2007). *Drosophila* Rtf1 also functions in histone methylation, gene expression, and Notch signaling (Tenney *et al.*, 2006). Human Paf1C binds to histone H3 tails with dimethylated histone H3 arginine17 (J. Wu & Xu, 2012).

Paf1C also has important functions that are not directly related to chromatin. Paf1C is required for co-transcriptional RNA 3'-processing (Penheiter *et al.*, 2005; Sheldon *et al.*, 2005). Human Cdc73 physically interacts with protein complexes required for 3'-processing (Rozenblatt-Rosen *et al.*, 2009). Paf1C also represses cryptic transcription (Y. Chu *et al.*, 2007) and is implicated in cellular differentiation (Kubota *et al.*, 2014) and human cancer (Chaudhary *et al.*, 2007; Takahashi *et al.*, 2011; Tan *et al.*, 2010). Paf1C represses gene silencing by small RNAs in *S. pombe* (Kowalik *et al.*, 2015), and Leo1 is involved in heterochromatin spreading (Verrier *et al.*, 2015). Paf1C has recently been found to regulate Pol II phosphorylation, promoter-proximal pausing, and release into gene bodies (F. X. Chen *et al.*, 2015; Yu *et al.*,

2015). Paf1C is also involved in the resolution of transcription-replication conflicts (Poli *et al.*, 2016).

1.3.2 The recruitment of Paf1 complex during transcription

Paf1 is generally recruited to transcribed units, apparently entering the Pol II elongation complex (EC) downstream of the transcription start site, and exiting at the polyadenylation (pA) site (Mayer *et al.*, 2010). There is evidence that Paf1C recruitment to Pol II requires direct contacts with Pol II and additional contacts with Pol II-associated factors. Paf1C recruitment *in vivo* requires the Bur1-Bur2 kinase (Larabee *et al.*, 2005), and is aided by Spt4 (Qiu *et al.*, 2006). Paf1C and its Cdc73 subunit bind the phosphorylated CTD of Pol II and the phosphorylated C-terminal repeat region (CTR) of Spt5, which is also a general elongation factor (Qiu *et al.*, 2012). A Plus3 domain in Rtf1 can bind the Spt5 CTR (Mayekar *et al.*, 2013). Rtf1 is however not stably associated with Paf1C in all species and is not required for Paf1C recruitment in human cells, where it has non-overlapping functions (Cao *et al.*, 2015). In fission yeast, Rtf1 also has other functions (Mbogning *et al.*, 2013). The C-terminal GTPase-like domain of Cdc73 is important for chromatin association of Paf1C (Amrich *et al.*, 2012). Leo1 is also important for Paf1C recruitment and binds RNA (Dermody & Buratowski, 2010).

1.3.3 Structural studies on Paf1 complex

Structural studies have revealed that Paf1C is a modular and flexible complex with several structured regions. The crystal structure of a complex of regions in Paf1 and Leo1 revealed antiparallel beta-sheets for heterodimerization (Figure 2) (X. Chu *et al.*, 2013). This study also showed that Ctr9 is a scaffold for Paf1C onto which the Paf1-Leo1 heterodimer and Cdc73 assemble. The structure of the GTPase-like domain in the C-terminal region of yeast Cdc73 was also solved (Figure 2) (Amrich *et al.*, 2012; H. Chen *et al.*, 2012). In addition, structures were reported for the Plus3 domain of human Rtf1 (de Jong *et al.*, 2008; Wier *et al.*, 2013), and for the Plus3 domain in complex with a phosphorylated Spt5 CTR repeat (Figure 2) (Wier *et al.*, 2013). There is however no structural information on Ctr9, the largest Paf1C subunit.

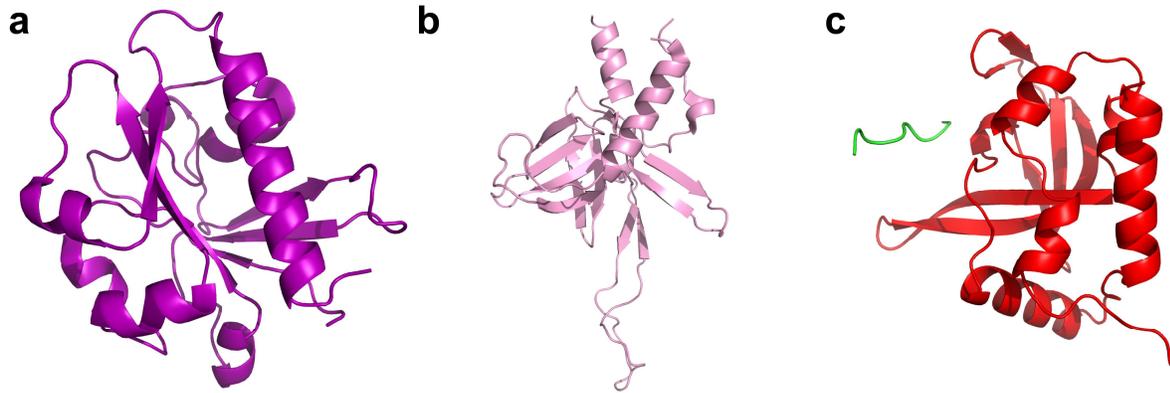


Figure 2: Crystal structures of Paf1C subunits. The ribbon models of solved domains within Paf1C (**a**, GTPase-like C-terminal domain of Cdc73 from yeast; **b**, Paf1-Leo1 heterodimer from human; **c**, the Plus3 domain (red) in complex with phosphorylated Spt5 CTR peptide (green) in human). Adapted from PDB entry: 3V46 (Amrich *et al.*, 2012), 4M6T (X. Chu *et al.*, 2013), and 4L1U (Wier *et al.*, 2013), respectively.

1.4 Aims and scope

Although Paf1C has critical and ubiquitous roles in Pol II transcription, little is known about its architecture, how it associates with Pol II and its way of function as an elongation factor. The primary goal of this thesis was to determine the structure of the entire Paf1C by crystallographic methods and further investigate Paf1C function with biochemical assays. The entire complex could be recombinantly expressed in *Escherichia coli* (*E. coli*) and reconstituted *in vitro* with high quality and excellent yields. However, crystallization remained unsuccessful. Limited proteolysis indicated a high degree of flexibility for the entire complex leading to the design of truncated constructs based on results from limited proteolysis, Edman sequencing and cross-linking coupled with mass spectrometry. Crystals from a Paf1C variation could be obtained by *in situ* proteolysis with chymotrypsin that diffracted only to 7 Å.

Then I aimed to determine the Paf1C structure together with Pol II. Simultaneously, a complex of Paf1C with Pol II and TFIIS was assembled because TFIIS was shown to enhance Pol II-Paf1C interactions. The architecture of the Pol II-Paf1C-TFIIS transcription elongation complex was then studied by cryo-EM and crosslinking coupled to mass spectrometry. The determined architecture reveals a trilobal architecture of Paf1C and the location of Paf1C on elongating Pol II. Existing crystal structures and crosslinking coupled to mass spectrometry data allowed assignment of several Paf1C subunits to the EM density. Comparing this structure to a Pol II-TFIIF initiation complex structure, it was apparent, that part of Paf1C occupies a similar binding site as the TFIIF dimerization domain. Complementary analytical sucrose gradient ultracentrifugation assays proved competitive binding of the two.

I also aimed to know if Paf1C affects the global transcription in yeast. Our 4tU-Seq reveals subunits Paf1 and Rtf1 are key co-transcriptional factors during Pol II transcription.

Finally, the interaction of Paf1C to RNA in the presence or absence of Rtf1 was studied, showing that removal of Rtf1 decreases affinity of Paf1C for RNA. It suggests Paf1C regulates the transcription not only rely on the protein-protein interaction, but also on protein-RNA. All together, this work provides a basis for mechanistic investigation of a central player in transcription elongation.

2. Materials and Methods

2.1 Materials

2.1.1 Bacterial strains

Table 4: *E. coli* strains used in this study.

Strain	Description	Source
XL1-Blue	endA1 gyrA96(nalR) thi-1 recA1 relA1 lac glnV44 F[::Tn10 proAB+ lacI ^q Δ(lacZ)M15] hsdR17(rK ⁻ mK ⁺) nalidixic acid resistant, tetracycline resistant (carried on the F plasmid)	Stratagene
BL21-CodonPlus(DE3)-RIL	<i>E. coli</i> B F ⁻ ompT hsdS(r _B ⁻ m _B ⁻) dcm ⁺ Tetr gal λ(DE3) endA Hte [argU ileY leuW Cam ^r] for high-level protein expression and easy induction in T7 expression systems	Stratagene

2.1.2 Yeast strains

Table 5: Yeast strains used in this study.

Strain	Description	Source
BJ5464 Rpb3 His-Bio	BJ5464 Rpb3 His-Bio tag introduced at 5' end of Rpb3 gene, use of URA3 selection marker for endogenous Pol II purification	(Kireeva <i>et al.</i> , 2003)
BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 for 4tU-Seq	Euroscarf

Strain	Description	Source
Δ Paf1	MATa his3 Δ 1 leu2 Δ 0 met15 Δ 0 ura3 Δ 0 paf1::kanMX for 4tU-Seq	Euroscarf
Δ Rtf1	MATa his3 Δ 1 leu2 Δ 0 met15 Δ 0 ura3 Δ 0 rtf1::kanMX for 4tU-Seq	Euroscarf

2.1.3 Plasmids

Table 6: Plasmids used in this study. The genes chemically synthesized by GeneArt™ (Thermo Fisher Scientific) are labeled in up case.

Vector	Insert	Type	Tag	Res.	Source
YXU001	Cdc73	pET21b	C-6xHis	Amp	this work
YXU002	Leo1	pET21b	C-6xHis	Amp	this work
YXU003	Paf1	pET21b	C-6xHis	Amp	this work
YXU004	Ctr9	pET21b	C-6xHis	Amp	this work
YXU005	Rtf1	pET21b	C-6xHis	Amp	this work
YXU006	Cdc73 (184-393)	pET21b	C-6xHis	Amp	this work
YXU007	Cdc73 (201-393)	pET21b	C-6xHis	Amp	this work
YXU008	Cdc73 (235-393)	pET21b	C-6xHis	Amp	this work
YXU009	Yaf9 (YEATS, 8-176)	pGEX4T-1	N-GST- Thrombin	Amp	this work
YXU010	Yaf9 (YEATS, 8-176)	pETDuet1	N-6xHis	Amp	this work

Vector	Insert	Type	Tag	Res.	Source
YXU011	Paf1-Yaf9	pETDuet1	Paf1, N-6xHis	Amp	this work
YXU012	Paf1-Yaf9 (YEATS,8-176)	pETDuet1	Paf1, N-6xHis	Amp	this work
YXU013	Ctr9-Cdc73	pETDuet1	Ctr9, N-6xHis	Amp	this work
YXU014	Paf1-Leo1	pETDuet1	Paf1, N-6xHis	Amp	this work
YXU015	Leo1-Paf1	pETDuet1	Leo1, N-6xHis	Amp	this work
YXU016	Ctr9-Leo1	pETDuet1	Ctr9, N-6xHis	Amp	this work
YXU017	Cdc73-Paf1	pETDuet1	Cdc73, N-6xHis	Amp	this work
YXU018	Paf1-Cdc73	pETDuet1	Paf1, N-6xHis	Amp	this work
YXU019	Cdc73	pETDuet1		Amp	this work
YXU020	Cdc73 (229-393)	pGEX4T-1	N-GST-Thrombin	Amp	this work
YXU021	Ctr9 (56-251)	pETDuet1	N-6xHis	Amp	this work
YXU022	Ctr9 (56-371)	pETDuet1	N-6xHis	Amp	this work
YXU023	Ctr9 (56-462)	pETDuet1	N-6xHis	Amp	this work
YXU024	Ctr9 (56-534)	pETDuet1	N-6xHis	Amp	this work
YXU025	Ctr9 (56-811)	pETDuet1	N-6xHis	Amp	this work

Vector	Insert	Type	Tag	Res.	Source
YXU026	Ctr9 (56-967)	pETDuet1	N-6xHis	Amp	this work
YXU027	Ctr9 (56-1077)	pETDuet1	N-6xHis	Amp	this work
YXU028	Cdc73	pGEX4T-1	N-GST- Thrombin	Amp	this work
YXU029	Cdc73 (1-228)	pGEX4T-1	N-GST- Thrombin	Amp	this work
YXU030	Ctr9 (461-967)	pETDuet1	N-6xHis	Amp	this work
YXU031	Ctr9 (461-1077)	pETDuet1	N-6xHis	Amp	this work
YXU032	Ctr9 (731-967)	pETDuet1	N-6xHis	Amp	this work
YXU033	Ctr9 (731-1077)	pETDuet1	N-6xHis	Amp	this work
YXU034	Ctr9 (811-967)	pETDuet1	N-6xHis	Amp	this work
YXU035	Ctr9 (811-1077)	pETDuet1	N-6xHis	Amp	this work
YXU036	Ctr9 (811-967)-Cdc73	pETDuet1	N-6xHis	Amp	this work
YXU037	Rtf1-Cdc73	pRSFDuet1	Rtf1, N- 6xHis	Kan	this work
YXU038	Leo1	pCDFDuet1		Sm	this work
YXU039	Cdc73	pET24b		Kan	this work
YXU040	Ctr9 (888-1077)	pET21b	C-6xHis	Amp	this work
YXU041	Ctr9 (888-1077)	pET24b		Kan	this work
YXU042	Paf1 (150-445)	pET21b	C-6xHis	Amp	this work
YXU043	Paf1 (150-445)	pET24b		Kan	this work

Vector	Insert	Type	Tag	Res.	Source
YXU044	Paf1 (150-445)	pET21b		Amp	this work
YXU045	Paf1 (230-445)	pET21b	C-6xHis	Amp	this work
YXU046	Paf1	pET24b		Kan	this work
YXU047	Ctr9	pET24b	C-6xHis	Kan	this work
YXU048	Paf1 (277-445)	pET21b	C-6xHis	Amp	this work
YXU049	Rtf1 (246-558)	pET21b	C-6xHis	Amp	this work
YXU050	Paf1 (277-445)	pET24b		Kan	this work
YXU051	Rtf1-Cdc73	pET21b		Amp	this work
YXU052	Ctr(801-1011)	pET24b		Kan	this work
YXU053	Paf1 (150-445)-Cdc73	pET24b		Kan	this work
YXU054	Paf1 (150-445)- Yaf9(YEATS, 8-176)	pET24b		Kan	this work
YXU055	Paf1 (178-445)	pET21b	C-6xHis	Amp	this work
YXU056	Paf1 (194-445)	pET21b	C-6xHis	Amp	this work
YXU057	Paf1 (207-445)	pET21b	C-6xHis	Amp	this work
YXU058	Paf1 (241-445)	pET21b	C-6xHis	Amp	this work
YXU059	Paf1 (264-445)	pET21b	C-6xHis	Amp	this work
YXU060	Leo1 (49-464)	pCDFDuet1		Sm	this work
YXU061	Leo1 (107-464)	pCDFDuet1		Sm	this work
YXU062	Leo1 (165-464)	pCDFDuet1		Sm	this work
YXU063	Leo1 (213-464)	pCDFDuet1		Sm	this work

Vector	Insert	Type	Tag	Res.	Source
YXU064	Paf1 (178-445)	pET24b	C-6xHis	Kan	this work
YXU065	Paf1 (1-276)	pET21b	C-6xHis	Amp	this work
YXU066	Paf1 (90-276)	pET21b	C-6xHis	Amp	this work
YXU067	Paf1 (90-445)	pET21b	C-6xHis	Amp	this work
YXU068	Leo1 (107-310)	pCDFDuet1		Sm	this work
YXU069	Leo1 (1-437)	pCDFDuet1		Sm	this work
YXU070	Leo1 (107-391)	pCDFDuet1		Sm	this work
YXU071	Leo1 (107-437)	pCDFDuet1		Sm	this work
YXU072	Rtf1 (397-558)	pET24b	C-6xHis	Kan	this work
YXU073	PAF1	pET21b	C-6xHis	Amp	this work
YXU074	Leo1-PAF1	pCDFDuet1		Sm	this work
YXU075	PAF1 (150-445)	pET21b	C-6xHis	Amp	this work
YXU076	LEO1-PAF1	pCDFDuet1		Sm	this work
YXU077	Ctr9 (1-810)	pET24b	C-His	Kan	this work
YXU078	Ctr9 (1-867)	pET24b	C-His	Kan	this work
YXU079	Ctr9 (1-887)	pET24b	C-His	Kan	this work
YXU080	Ctr9 (1-913)	pET24b	C-His	Kan	this work
YXU081	Ctr9 (1-966)	pET24b	C-His	Kan	this work
YXU082	LEO1 (93-464)-PAF1	pCDFDuet1		Sm	this work
YXU083	Rtf1 (397-558)-Cdc73	pET21b		Amp	this work

Vector	Insert	Type	Tag	Res.	Source
YXU084	Rtf1 (442-558)-Cdc73	pET21b		Amp	this work
YXU085	LEO1 (118-376)-PAF1 (1-360)	pCDFDuet1		Sm	this work
YXU086	Rtf1-Cdc73	pGEX4T-1	Rtf1, N-GST-Thrombin	Amp	this work
YXU087	SPT5	pET21b		Amp	this work
YXU088	Spt4	pOPINK	N-6xHis-GST-3C	Kan	this work
2097	Spt5 (CTR)	pET28a	N-6xHis-GST	Kan	Amelie Schrieck
224	TFIIS	pET28a	N-6xHis	Kan	Hubert Kettenberger
225	TFIIS (131-309)	pET28a	N-6xHis	Kan	Hubert Kettenberger
SS191	TFIIS (D290A/E291A) (TFIIS (DE-AA))	pET28a	N-6xHis-GST-3C	Kan	Sarah Sainsbury

2.1.4 Synthetic Genes and oligonucleotides

Table 7: Synthetic genes used in this study.

Name	Organism	Company	Comment
PAF1	<i>S. cerevisiae</i>	GeneArt	Codon optimized for <i>E. coli</i>
LEO1	<i>S. cerevisiae</i>	GeneArt	Codon optimized for <i>E. coli</i>
SPT5	<i>S. cerevisiae</i>	GeneArt	Codon optimized for <i>E. coli</i>

Table 8: Synthetic oligonucleotides in this study.

Name	Sequence (5'-3')	Usage
Template DNA	AAGCTCAAGTACTTAAGCCTGGTCATTACTAGT ACTGCC	Pol II elongation complex
Non-template DNA	GGCAGTACTAGTAACTAGTATTGAAAGTACTT GAGCTT	Pol II elongation complex
RNA-20	UAUAUGCAUAAAGACCAGGC	Pol II elongation complex
RNA-50	GAACGAGAUCAUAAACAUUUGAACAAAGAAUUAU AUAUACAUAAA	Pol II elongation complex
Coding RNA-33	FAM- AAUAUUCAAGACGAUUUAGACGAUAAUAUCA UA	Fluorescence anisotropy assays
Random RNA-33	FAM- AAUAUUCAAGCAGUAUUAGCAGUAAAUAUCA UA	Fluorescence anisotropy assays
FAM-RNA-20	FAM-UAUAUGCAUAAAGACCAGGC	Transcription elongation assay

Name	Sequence (5'-3')	Usage
TA-1-tempalte DNA	ACAAATTACTGGGAAGTCGACTATGCAATACA GGCATCATTGATCAAGCTCAAGTACTTAAGCC TGGTCATTACTAGTACTGCC	Transcription elongation assay
TA-1-nontemplate DNA	GGCAGTACTAGTAATGACCAGGCTTAAGTACTT GAGCTTGATCAAATGATGCCTGTATTGCATAGT CGACTTCCCAGTAATTTGT	Transcription elongation assay
FAM-GK-RNA-20	FAM-UAUACAUA AUGGAGUAGGGU	Transcription elongation assay
GK-TA-pML5 template DNA	AGGAAGCTAAAATGGAGAAAAAATCACTGGA TATACCACCGTTGATATATTACCCTACTCCATA ACTTCAACAACC	Transcription elongation assay
GK-TA-pML5 nontemplate DNA	GGTTGTTGAAGTTATGGAGTAGGGTAATATATC AACGGTGGTATATCCAGTGATTTTTTCTCCATT TTAGCTTCCT	Transcription elongation assay

2.1.5 Growth medium and additives

Table 9: Growth medium for *E. coli* and *S. cerevisiae*.

Medium	Application	Description
Lysogeny broth (LB)	<i>E. coli</i>	1 % (w/v) tryptone, 0.5 % (w/v) yeast extract, 0.5 % (w/v) NaCl
LB plates	<i>E. coli</i>	1 % (w/v) tryptone, 0.5 % (w/v) yeast extract, 0.5 % (w/v) NaCl, 1.5% (w/v) agar
Yeast Extract-Peptone-Dextrose Broth (YPD)	<i>S. cerevisiae</i>	2% (w/v) peptone, 2% (w/v) glucose, 1% (w/v) yeast extract
YPD plates	<i>S. cerevisiae</i>	2% (w/v) peptone, 2% (w/v) glucose, 1% (w/v) yeast extract, 1.5% (w/v) agar

Table 10: Medium additives for *E. coli* and *S. cerevisiae*.

Additive	Stock solution	Applied concentration	Application
Ampicillin	100 (w/v) mg/ml in H ₂ O	0.1 mg/L	Antibiotic
Kanamycin	50(w/v) mg/ml in H ₂ O	0.05 mg/L	Antibiotic
Chloramphenicol	30 (w/v) mg/mL in Ethanol	0.03 mg/L	Antibiotic
Streptomycin	50(w/v) mg/ml in H ₂ O	0.05 mg/L	Antibiotic
IPTG	1 M in H ₂ O	0.5 mM	Protein expression induction

2.1.6 Buffers and solutions

Table 11: General buffers, dyes and solutions.

Name	Description	Application
1x TE	10 mM Tris-HCl pH 8.0, 1 mM EDTA	Various
10x DNA loading dye	Thermo Fisher Scientific	Agarose gel electrophoresis
GeneRuler 1 kb DNA Ladder	Thermo Fisher Scientific	Agarose gel electrophoresis
10x TAE	50 mM EDTA pH 8.0, 2.5 M Tris-acetate	Agarose gel electrophoresis
SYBR Safe	Thermo Fisher Scientific	DNA gel stain
Electrophoresis buffer	20x MOPS NuPAGE buffer (Life Technologies/Thermo Fisher Scientific)	SDS-PAGE
5X SDS loading dye	10% (w/v) SDS, 10 mM DTT, 20% (v/v) glycerol, 0.2 M Tris-HCl pH 6.8, 0.05% (w/v) bromophenol blue	SDS-PAGE
Instantblue	Expedeon	SDS-PAGE
SDS-PAGE Coomassie staining solution	50% (v/v) ethanol, 7% (v/v) acetic acid, 0.125% (w/v) Coomassie Brilliant Blue R-250	SDS-PAGE
Destain solution	5% (v/v) ethanol, 7.5% (v/v) acetic acid	SDS-PAGE
Broad-Range SDS-PAGE Standards	Bio-Rad Laboratories	SDS-PAGE
Prestained Protein Ladder	Thermo Fisher Scientific	SDS-PAGE
Silver nitrate solution	0.1% (w/v) silver nitrate, 0.01% (v/v) formalin (37% formaldehyde)	Silver staining

Name	Description	Application
Developing solution	3% (w/v) sodium carbonate, 0.05% (v/v) formalin (37% formaldehyde)	Silver staining
Transfer buffer (commercial)	NuPAGE TransferBuffer (Life Technologies/Thermo Fisher Scientific)	Western blot
Transfer buffer	25mM Tris, 192mM glycine, 20% ethanol	Western blot
100x Protease inhibitor (PI)	0.028 mg/mL Leupeptin, 0.137 mg/mL Pepstatin A, 0.017 mg/mL PMSF, 0.33 mg/mL Benzamidine in ethanol	Protein purification
1X PBS	137mMNaCl, 2.7mMKCl, 10mMNa ₂ HPO ₄ , 1.76mM KH ₂ PO ₄ pH 7.4	Western blot

Table 12: Buffers used for preparation of competent *E. coli* cells.

Name	Description
TFB-I	30 mM KOAc, 50 mM MnCl ₂ , 100 mM RbCl, 10 mM CaCl ₂ , 15 % (v/v) glycerol pH set with HOAc at 5.8 at 4 °C
TFB-II	10 mM MOPS, pH 7.0 at 4 °C, 75 mM CaCl ₂ , 10 mM RbCl, 15 % (v/v) glycerol

Table 13: Enzymes, buffers, and components used for cloning.

Name	Source
2x Phusion High-Fidelity PCR Master Mix	New England Biolabs
Dimethyl Sulfoxide (DMSO)	Sigma-Aldrich
Restriction enzymes	New England Biolabs, Fermentas
FastaAP	Fermentas
DpnI	New England Biolabs
QIAquick PCR purification kit	Qiagen
QIAquick gel extraction kit	Qiagen
T4 DNA Ligase	New England Biolabs
10x T4 DNA ligase reaction buffer	New England Biolabs
Quick T4 DNA Ligase	New England Biolabs
2x Quick ligation buffer	New England Biolabs
QIAprep spin miniprep kit	Qiagen

Table 14: Buffers used for Paf1C purification, crosslinking and pull-down.

Name	Description	Application
Buffer A	50 mM Tris pH 8.7, 600 mM potassium acetate (KOAc), 2 mM dithiothreitol (DTT), 2 mM MgCl ₂ , 10 mM imidazole, 10 μM ZnCl ₂	Purification
Buffer B	50 mM Tris pH 8.7, 70 mM KOAc, 2 mM MgCl ₂ , 2 mM DTT, 10 μM ZnCl ₂	Purification

Name	Description	Application
Buffer C	10 mM Tris pH 8.5, 200 mM NaCl, 2 mM MgCl ₂ , 2 mM DTT, 10 μM ZnCl ₂	Purification
Buffer D	10 mM Hepes pH 7.5, 100 mM NaCl, 2 mM MgCl ₂ , 2 mM DTT, 10 μM ZnCl ₂	Crosslinking and Pol II-Paf1C-TFIIS elongation complex preparation
Buffer P	50 mM Hepes pH 7.5, 0.1% NP-40, 150 mM KOAc, 2 mM DTT	Pull-down assays

Table 15: Buffers used for sucrose gradient ultracentrifugation and fixation.

Name	Description	Application
Light solution	10 mM Hepes pH 7.5, 100 mM NaCl, 2 mM DTT, 2 mM MgCl ₂ , 10 μM ZnCl ₂ , 10% (v/v) sucrose	Gradient fixation (with heavy solution 1), and gradient ultracentrifugation (with heave solution 2)
Heavy solution 1	10 mM Hepes pH 7.5, 100 mM NaCl, 2 mM DTT, 2 mM MgCl ₂ , 10 μM ZnCl ₂ , 30% (v/v) sucrose, 0.075% (v/v) glutaraldehyde	Gradient fixation
Heavy solution 2	10 mM Hepes pH 7.5, 100 mM NaCl, 2 mM DTT, 2 mM MgCl ₂ , 10 μM ZnCl ₂ , 30% (v/v) sucrose	Gradient ultracentrifugation

Table 16: Buffers used for fluorescence anisotropy assays.

Name	Description
Buffer F	200 mM NaCl, 10 mM Tris pH 8.0, 1 mM TCEP, 2.4 μ l, and 2 mM MgCl ₂
Buffer G	10 mM Tris pH 8.0, 1 mM TCEP, 2 mM MgCl ₂ , 8% glycerol, 0.02 mg/ml BSA, 10 μ g/ml yeast tRNA

Table 17: Buffers used for transcription elongation assays.

Name	Description
10x TBE	Sigma-Aldrich
1x Transcription buffer (TB)	100 mM NaCl, 30 mM HEPES pH 7.5, 3 mM MgCl ₂ , 2 mM DTT, 4% (v/v) glycerol
2x Stop buffer	1x TBE buffer, 50 mM EDTA pH 8.0, 6.5 M urea, 0.33 μ g/ μ L proteinase K
20% TBE urea gel	8M urea, 1x TBE, 20% 19:1 acrylamide 40% solution, 0.5% (v/v) APS stock (10% (w/v)), 0.1% TEMED

2.2 Methods

2.2.1 Molecular cloning

Cloning strategy

The gene sequences encoding for Ctr9, Rtf1, Paf1C, Leo1, and Cdc73 were amplified from *Saccharomyces cerevisiae* (*S. cerevisiae*) genomic DNA by polymerase chain reaction (PCR). In addition, open reading frames of Paf1 and Leo1 were chemically synthesized by GeneArt™ (Thermo Fisher Scientific) to optimize codon usage for more efficient expression in bacteria (Table 7). Each coding region was preceded by a ribosome binding sequence (RBS). Multiple proteins were cloned either in a vector containing two multiple cloning sites (MCS) or in in-house modified vectors. To increase the expression yield, different tags, C-terminal or N-terminal hexa-histidine tag, or untagged together with the sequential coding genes were as indicated in the plasmids list (Table 6). Plasmids with different antibiotic resistances were used for co-expression.

Polymerase Chain Reaction (PCR)

Oligonucleotide PCR primers were designed generally with a length of ~39 nucleotides containing ~21 nucleotides complementary to the target sequences with a GC content of 40-60%, and also with two restriction enzymes sites simultaneously. There were minor changes for amplifying some difficult targets. PCR was carried out by using 2x Phusion High Fidelity PCR Master Mix (NEB), both forward and reverse primers, and DNA templates. The reactions were generally performed in a Biometra T3000 Thermocycler with ~30 thermocycles. The annealing temperature and extension time was dependent on the GC content of the complementary nucleotides and the length of the coding genes. PCR products were visualized by agarose gel electrophoresis with Sybr Safe stain (Thermo Fisher Scientific). 1 µl of DpnI (NEB) was added to the PCR products and incubated at 37 °C for 1 hour to digest the template plasmids. The PCR products were purified using QIAquick PCR purification kit (Qiagen, Table 11).

Restriction digest and dephosphorylation

Purified PCR products and vectors were double digested using appropriate restriction enzymes and reaction buffer (NEB) at 37 °C for 2 hours. Linearized vectors were further

dephosphorylated with FastaAP (NEB). Dephosphorylated vectors and cleaved coding gene fragments were purified using agarose gel electrophoresis and subsequently the QIAquick Gel Extraction Kit (Qiagen). 30 ng of vector and 4-fold molar excess of target gene were ligated using the Quick ligation kit (NEB) in a 20- μ l reaction volume at room temperature for 5 min as recommended by the manufacturer.

Chemically competent cells preparation

Two strains of chemically competent *E. coli* cells were used in this study for transformation (Table 4). Ligation products were transformed in to XL 1-Blue cells. This strain of *E. coli* cells was also used for plasmids amplification. The other strain, BL21 CodonPlus(DE3)RIL cells were used for protein over-expression.

To prepare chemically competent cells, 5 ml LB medium culture containing appropriate antibiotics and *E. coli* cells were grown at 37 °C at 140 rpm overnight. 500 ml LB medium with appropriate antibiotics (Table 10) was inoculated by the overnight culture in a ratio of 1:100 and cells were grown at 37 °C with 140 rpm shaking for ~3 hours to an OD_{600nm} of ~0.3. The culture was put on ice for 10 min followed by centrifugation at 4,000 x g at 4 °C for 5 min to harvest the cells. From this step on, the cells were kept on ice. After resuspending in 100 ml pre-chilled Transformation buffer 1 (TFB-1) (Table 12), the cells were incubated on ice for 5 min again subsequently cell were harvested by centrifugation at 4,000 x g at 4 °C for 5 min. The supernatant was discarded, and the pellet was resuspended in 10 ml Transformation buffer 2 (TFB-2) (Table 12). After another incubation on ice for 15 min, 50 μ l aliquots were flash frozen in liquid nitrogen and stored at -80 °C.

Transformation, plasmid verification and sequencing

The competent cells were thawed on ice and 10 μ l out of 20 μ l ligation product or 50 ng pure plasmid was added and incubated on ice for 20 min. The competent cells were heat shocked at 42 °C for 1 minute followed by an incubation on ice for 2 min. 450 μ l LB medium was added and cells were recovered at 37 °C for 45 min with 900 rpm sharking in a thermomixer (eppendorf). Cells were then plated on LB agar plates containing the appropriate antibiotic, and incubated at 37 °C overnight. 5 ml LB medium containing the appropriate antibiotic was inoculated with a single colony and incubated at 37 °C with 150 rpm shaking overnight. 0.2 μ l of the culture was

used as template for colony PCR. After analyzing the PCR products by agarose gel electrophoresis, plasmids DNA of positive colonies was extracted using the Miniprep purification kit (Qiagen). Plasmid sequence was further confirmed by sequencing (GATC). For protein co-expression, two or three plasmids were co-transformed into the same competent cells at the same time and plated on LB agar plates containing multiple antibiotics.

Determination of nucleic acid concentration

All the nucleic acid concentrations were determined using a NanoDrop-2000 spectrophotometer (Thermo Fisher Scientific) by measuring the absorption at 260 nm.

2.2.2 General protein methods

Protein expression and purification

All the proteins except yeast endogenous RNA polymerase II used in this study were expressed in *E. coli*. Protein purification usually includes affinity purification, ion exchange and size exclusion chromatography (see 2.2.3 Specific methods).

Protein concentrating, determination of concentration, and storage

Proteins were concentrated using AMICON Ultra centrifugal concentrators (Millipore) or Vivaspin concentrators (GE) with a proper molecular weight cutoff. All the protein concentrations were determined using a NanoDrop-2000 spectrophotometer (Thermo Fisher Scientific) by measuring the absorption at 280 nm. The extinction coefficients were calculated using ProtParam (<http://web.expasy.org/protparam/>). The aliquots of purified proteins were flash-frozen in liquid nitrogen and stored at -80 °C.

SDS-Polyacrylamide gel electrophoresis (SDS-PAGE)

Protein samples were mixed with 5 x SDS-PAGE loading dye (Table 11) followed by an electrophoretic separation using precast NuPAGE Novex 4-12% Bis-Tris gels (Invitrogen). Protein marker (PageRuler Prestained Protein Ladder, Thermo Fisher Scientific) was loaded as a size control. Gels were run in 1 x MOPS running buffer (Invitrogen) at 200 V for 45-90 min and stained with InstantBlue (Expedeon) before imaging. In case of low protein concentration,

trichloroacetic acid (TCA) precipitation was used before loading samples on the gels, or silver staining was used to detect proteins after electrophoretic separation (see below).

Trichloroacetic acid (TCA) protein precipitation

TCA was added to a final concentration of 10% to the protein sample and incubated at -20 °C for 10 min. The solution was centrifuged for 10 min at 14,000 rpm at 4 °C and the supernatant was carefully removed. The pellet was washed with 1 ml pre-chilled (-20 °C) acetone. Centrifugation was carried out as before and the supernatant was carefully removed before incubating for 5 min at 95 °C to evaporate the residual acetone. The pellet was resuspended in 10-15 µl water and 5 µl 5 x SDS loading dye. If the sample turned yellow, indicating a pH change, Tris pH 8 was added until the sample turned blue.

Silver staining

The acrylamide gel was first rinsed twice in water to remove the MOPS buffer, and then soaked in 40 ml 50% (v/v) ethanol for 20 min and then 40 ml 5% (v/v) ethanol for 20 min, and then in 40 ml water containing 35 µM DTT for 5 min followed by soaking the gel in 40 ml silver nitrate solution (Table 11) for 10 min. The gel was rinsed twice with 40 ml water each time and 25 ml ice-chilled developing solution (Table 11) before soaked in 50 ml ice-chilled developing solution to the desired darkness of the protein bands. Solid citric acid monohydrate was added to neutralize the pH and stop the reaction. The gel was transferred to water and imaged.

Protein identification

Mass spectrometry was carried out for protein identification by the protein analysis core facility of Adolf-Butenandt-Institute at LMU, Munich or Thomas Fröhlich at the Genzentrum, LMU, Munich or by Monika Raabe from Urlaub laboratory at the MPIbpc, Göttingen.

Western blot

Protein samples were applied to SDS-PAGE and transferred to polyvinylidene difluoride (PVDF) membranes (GE Healthcare) using the Trans-Blot Turbo Blotting system (Bio-Rad). The transfer was done in ice-cold 1x transfer buffer (Table 11) at 150 V for 90 min on ice. The membrane was either stained with Ponceau S for Edman sequencing (see 2.2.3 Specific methods)

or for immunostaining. Immunostaining against His antibody was mainly used in this study. The membrane was first blocked with 2 % milk for 30 min at room temperature in 1x PBS (Table 11) containing 0.02% Tween 20 (PBS-T). Then the membrane was incubated with the primary antibody in 2% milk in PBS-T at 4 °C overnight. The membrane was washed three times with 2% milk in PBS-T for 5 min for each wash before incubating with the secondary antibody for 60 min at room temperature. The membrane was washed as described before developing using the Super SignalWest Pico Chemiluminescent Substrate (Thermo Fisher Scientific) and imaged with an Advanced Fluorescence Imager (Intas).

2.2.3 Specific methods

Pol II, TFIIIS, and TFIIIF purification

Endogenous *S. cerevisiae* Pol II, recombinant TFIIIF, and recombinant TFIIIS and its variants were purified as described (Kettenberger *et al.*, 2003, 2004; Plaschka *et al.*, 2016; Sydow *et al.*, 2009).

Scaffold preparation

To anneal DNA-RNA scaffolds for elongation complex preparation and *in vitro* transcription elongation assays, synthetic oligonucleotides were dissolved in TE (Table 11) to a final concentration of 400 μ M. Equimolar amounts of template DNA and RNA were mixed in a PCR tube. The mixed DNA-RNA scaffolds were incubated at 95 °C for 5 min and then annealed by reducing the temperature from 95 °C to 20 °C in 1 °C/min steps using the Thermocycler (Biometra). The annealed scaffolds were stored at -80 °C.

Paf1C expression and purification

The gene sequences encoding for Ctr9, Rtf1, Cdc73 were amplified from the *S. cerevisiae* genomic DNA by PCR. Open reading frames of Paf1 and Leo1 were chemically synthesized by GeneArtTM (Thermo Fisher Scientific) to optimize codon usage for more efficient bacterial expression. Full-length Ctr9 was cloned into pET24b (Novagen) resulting in a non-cleavable hexa-histidine tag fused to the Ctr9 carboxyl-terminus. The PCR products of full-length Rtf1 or Cdc73 were cloned into an in-house modified version of the pET21b vector (Novagen). Ribosome binding sequence (RBS)-Cdc73, which contains an RBS on the amino-terminus of

Cdc73 was inserted sequentially after Rtf1. Additionally, Cdc73 was cloned into multiple cloning sites 2 (MCS2) of pETDuet-1 (Novagen) for the co-expression of a four-subunit Paf1 complex lacking Rtf1 (Paf1C- Δ Rtf1). Leo1 and Paf1 were PCR-amplified and cloned into two multiple cloning sites of pCDFDuet-1 vector (Novagen) separately and sequentially. The five-subunit Paf1C and its variants were heterologously co-expressed in *E. coli* BL21 CodonPlus (DE3)RIL cells (Stratagene). Cultures of bacteria transformed with pET24b-Ctr9 and pCDFDuet-1-Leo1-Paf1 were grown in LB medium at 37 °C to an OD_{600nm} of ~0.8. The temperature was reduced to 18 °C, and protein over-expression was induced by addition of 1 mM isopropyl-b-D-thiogalactoside (IPTG) and with continued growth at 18 °C overnight. The same strategy was used for the co-expression of Rtf1 and Cdc73.

Cells were harvested and co-lysed by sonication in buffer A (Table 14) containing a 1:100 dilution of protease inhibitor cocktail (1 mM leupeptin, 2 mM pepstatin A, 100 mM phenylmethylsulfonyl fluoride, 280 mM benzamidine), 1,000 units benzonase, and 0.4 μ g/mL DNaseI. The extract was cleared by centrifugation (20,000 x g, 30 min) and the supernatant was loaded onto a 2-ml Ni-NTA agarose bead column (QIAGEN), equilibrated in buffer A. The column was washed extensively with buffer A containing 20 mM imidazole. The complex was eluted with buffer A containing 150 mM imidazole. The eluted protein was diluted 10-fold and was further purified by cation exchange chromatography using a 1-ml HiTrap SP HP column (GE Healthcare). The column was equilibrated in buffer B (Table 14) and proteins were eluted with a linear gradient from 70 mM to 2 M KOAc in buffer B. Fractions containing the protein of interest were concentrated and loaded onto a Superose 6 10/300 (GE Healthcare) size-exclusion chromatography column equilibrated with buffer C (Table 14) or buffer D (Table 14). The protein complex was concentrated by centrifugation in Amicon Ultra 4-ml concentrators (MWCO=50 kDa, Millipore) to 3 mg/mL. Protein was aliquoted, flash-frozen in liquid nitrogen and stored at -80 °C. The similar strategy was used for purifying the five-subunit Paf1C variants.

The four-subunit Paf1 complex (Paf1C- Δ Rtf1) and its variants were expressed and purified as described for the five-subunit complex (Paf1C) except that we co-transformed pET24b-Ctr9, pCDFDuet-1-Leo1-Paf1 and pETDuet-1-Cdc73 plasmids together in *E. coli* BL21 CodonPlus(DE3) RIL cells. All the components were separated by SDS-PAGE and confirmed by mass spectrometry.

Limited proteolysis and Edman sequencing

Limited proteolysis experiments were performed in buffer C containing 1.6 mg/mL Paf1C and different concentration of respective protease. 0.04-0.4 ng/ μ L subtilisin, 1-10 ng/ μ L trypsin, and 2 ng/ μ L ArgC were incubated with Paf1C at room temperature for 2 min, 2 min, and a 5-20 min time course, respectively. The reactions were stopped using SDS loading buffer. Samples were analyzed by SDS-PAGE. The separated limited proteolysis products on the SDS gel were blotted onto a PVDF membrane. The membrane was stained with Ponceau S before the fragments were excised and sequenced using a Procise cLC (Applied Biosystems, Foster City, CA, USA).

Pull-down assay

In each pull-down assay, 5.8 μ g purified Pol II was biotinylated on the Rpb3 subunit as previously described (Kireeva *et al.*, 2003) and immobilized on 20 μ L Dynabeads M-280 Streptavidin resin (Thermo Fisher Scientific), equilibrated in buffer P (Table 14). Five fold molar excess of purified Paf1C or/and TFIIS were incubated with immobilized Pol II or control resin at 4 °C for 1 hour. Beads were washed 5 times. Input and the bound proteins were subjected to SDS-PAGE analysis.

Elongation complex preparation

The nucleic acid scaffold (Integrated DNA Technologies) used for transcribing mammalian RNA polymerase II (Bernecky *et al.*, 2016), which contains an 11 nucleotide mismatch transcription bubble and 20 nucleotide RNA (bubble-RNA) was used to assemble Pol II-Paf1C-TFIIS EC (template DNA sequence 5'-AAGCTCAAGTACTTAAGCCTGGTCATTACTAGTACTGCC-3', non-template DNA sequence 5'-GGCAGTACTAGTAACTAGTATTGAAAGTACTTGAGCTT-3', and RNA sequence 5'-UAUAUGCAUAAAGACCAGGC-3', Table 8). 250 pmol purified Pol II was mixed with equimolar bubble-RNA as described (Kettenberger *et al.*, 2004). A 1.8-fold molar excess of Paf1C-Ctr9- Δ C913 and a 1.8-fold molar excess of TFIIS (DE-AA) inactive mutant were incubated with Pol II-bubble-RNA in assembly buffer D for 15 min at 20 °C in a 65 μ L reaction volume. We then centrifuged the reaction for 10 min at 4°C at 15,000 rpm and kept the supernatant for size-exclusion chromatography or gradient fixation (GraFix) before XL-MS or EM, respectively.

Crosslinking and mass spectrometry analysis

The assembled Pol II-Paf1C-TFIIS complex was injected onto size-exclusion chromatography to obtain a homogeneous complex. The fractions containing target complex were collected and crosslinked at various concentrations of bis(sulfosuccinimidyl) suberate (BS3, Thermo Fisher Scientific) to empirically determine the optimal reaction conditions. The best condition, 0.5 mM BS3, was sufficient to convert most of individual component into a high molecular weight band in SDS-PAGE and was chosen for final sample preparation. The Pol II-Paf1C-TFIIS at a concentration of 425 µg/mL was crosslinked with 0.5 mM BS3 and incubated for 30 min at 30 °C. The reaction was quenched by adding 50 mM ammonium bicarbonate. The crosslinked sample was re-purified by size-exclusion chromatography on a Superose 6 PC 3.2/300 column (GE Healthcare) equilibrated in buffer D. Crosslinked sample was digested with trypsin as described before (Rauhut *et al.*, 2016).

Crosslinked peptides were enriched and divided into two parts. Both halves of the sample were measured on an Orbitrap Fusion LC-MS/MS instrumentation platform (Thermo Fisher Scientific) and the datasets were analyzed with pLink 1.23 (Yang *et al.*, 2012) against a database containing the sequences of the proteins components in the complex separately. An initial false discovery rate (FDR) cutoff of 1% was set. E-value was calculated in this processing. To visualize this score better, the negative logarithm of E-value was used. The final set of crosslinks were required to satisfy 3 criteria: 1) appeared in both replicates; 2) the max score value from each dataset was higher than 5; 3) each crosslink must have a minimal spectral count of 2 in each dataset. The final result was subsequently visualized using the xiNET online server (Combe *et al.*, 2015). The same strategy was used for Paf1C XL-MS, except that the samples were measured on an AB Sciex Triple-ToF instrument (AB SCIEX).

Gradient fixation

To reconstitute a homogeneous Pol II-Paf1C-TFIIS EC, the sucrose Gradient Fixation (GraFix) was carried out as described (Kastner *et al.*, 2008; Stark, 2010). Each sucrose gradient for GraFix was generated by mixing equal volumes of light solution (Table 15) and heavy solution 1 (Table 15) using a gradient mixer (Gradient Master 108, BioComp Instruments). This resulted in a dual gradient of 10-30% sucrose and 0-0.075% glutaraldehyde in an 11 x 60 mm ultracentrifuge tube (Beckman Coulter). Next 60 µL of the *in vitro* reconstituted Pol II-Paf1C-

TFIIS EC were applied on top of the gradient. After ultracentrifugation at 32,000 rpm in a SW60 swinging bucket rotor (Beckman Coulter) for 16 h at 4 °C, 200 μ L fractions of the gradient were collected by pipetting carefully from top to bottom of the tube. Parallel sucrose gradient fractions of samples applied to gradients either containing or lacking glutaraldehyde showed the same sedimentation profile when analyzed by SDS-PAGE. The crosslinking reaction was quenched by adding 0.5 M pH 7.8 aspartate to a final concentration of 12.5 mM. Micro spin chromatography columns (Bio-Rad) were used for buffer exchange to remove sucrose, glutaraldehyde and aspartate. The samples were concentrated using a GE concentrator (MWCO=100 kDa, GE Healthcare) and immediately used for cryo-EM.

Cryo-EM specimen preparation and data acquisition

An FEI Vitrobot Mark IV plunger (FEI) was used for preparation of frozen-hydrated specimens. 4 μ L of sample was placed onto Quantifoil Cu R3.5/1 and Cu R2/1 glow-discharged 200 mesh holey carbon grids, which were then blotted for 8.5 s with blot force 13 to remove the excess solution before they were flash-frozen in liquid ethane. The Vitrobot chamber was operated at constant 4 °C and 100% humidity during blotting. The grids were transferred and stored in liquid nitrogen before data acquisition.

Two cryo-EM datasets were acquired on a 300 keV FEI Titan Krios electron microscope equipped with a K2 Summit direct electron counting camera (Gatan) positioned post a GIF Quantum® energy filter (Gatan) to increase the signal-noise ratio. Automated data collection was carried out using the TOM toolbox (Korinek *et al.*, 2011). Movie images were recorded at a nominal magnification of 37,000x (corresponding to a calibrated sampling of 1.35 Å per physical pixel) in super-resolution mode, thus yielding a pixel size of 0.675 Å per pixel. For the first dataset, two movies were acquired in each hole and a total of 595 movie stacks with a defocus range of -0.7 μ m to -4.2 μ m were collected from Quantifoil Cu R3.5/1 grids at a dose rate of 7.6 electrons per pixel per second. Each movie encompassed a total dose of \sim 33 electrons per Å² with a total exposure time of 10.8 s fractionated into 27 frames. Each frame had an exposure time of 0.4 s. For the second dataset collected from the Quantifoil Cu R2/1 grids, one movie was acquired in each hole and 2,146 movie stacks were recorded with an exposure time of 12 s fractionated into 30 frames, a dose rate of 4.2 electrons per pixel per second. Each frame had an exposure time of 0.4 s, resulting in a total accumulated dose of approximate 28 electrons per Å²

per stack. Defocus values ranged from $-0.6\ \mu\text{m}$ to $-4.2\ \mu\text{m}$. Movie stacks from two datasets were aligned and binned as previously described with the frame-based motion-correction algorithm to generate drift-corrected micrographs for further processing (Li, Mooney, *et al.*, 2013; Li, Zheng, *et al.*, 2013; Plaschka *et al.*, 2015) except images were not partitioned into quadrants.

Cryo-EM image processing

Contrast transfer function (CTF) was estimated using CTFFIND3 (Mindell & Grigorieff, 2003) and CTFFIND4 (Rohou & Grigorieff, 2015) for the datasets from Quantifoil Cu R3.5/1 grids (R3.5) and Quantifoil Cu R2/1 grids (R2), respectively. For the R3.5 dataset, 123 aligned micrographs were excluded because of contaminations or bad ice quality. After removing these micrographs, we used `e2boxer.py` from EMAN2 package (Tang *et al.*, 2007) to semi-automatically pick 84,362 particles with a box size of 240^2 pixels from the remaining micrographs. Two-dimensional (2D) reference-free classification within RELION 1.4 (Scheres, 2012) was used to remove micelles or other false positives. After this step 79,024 particles remained. We deleted 512 aligned micrographs from the R2 dataset and semi-automatically picked 21,301 particles using `e2boxer.py` from EMAN2 package with a box size of 240^2 pixels.

Reference-free 2D classes were generated, and nine representative classes were low-pass filtered to $20\ \text{\AA}$ and used as templates for autopicking (Scheres, 2015). The resulting 863,235 particles were screened manually and subjected to reference-free 2D classification, yielding 291,817 particles. We then merged particles from two datasets. 370,841 particles were used as an input for the subsequent three-dimensional (3D) reconstruction performed with RELION 1.4 (Scheres, 2012), unless noted otherwise. A published reconstruction of bovine Pol II (Bernecky *et al.*, 2016) filtered to $50\ \text{\AA}$ was used as an initial reference for 3D refinement. The aligned particles were subjected to particle polishing using RELION 1.4 (Scheres, 2014) to reduce the noise and correct the local motion and radiation damage.

3D classifications were carried out without image alignment. The first 3D classification was performed to separate out particles lacking the Paf1C density. This led to the dismissal of 158,422 particles. We refined the remaining 212,419 Paf1C-containing particles. Gold-standard Fourier Shell Correlation (FSC) was calculated after the 3D refinement in RELION as described (S. Chen *et al.*, 2013). A $5.5\ \text{\AA}$ average resolution map was observed at the gold-standard FSC criteria of 0.143. In the second step, a soft mask encompassing Pol II, TFIIS, Paf1C and

DNA/RNA was generated using the ‘volume erase’ option in UCSF Chimera (Pettersen *et al.*, 2004) and RELION (Scheres, 2012). After 3D classification within this mask, we obtained three Pol II-Paf1C-TFIIS maps resolving different parts of Paf1C density (parts A, B, C). The individual classes were auto-refined using the respective masks. The complex containing part A was refined from 114,672 particles to an average resolution of 5.7 Å. The complex containing part B was refined from 54,722 particles to an average resolution of 5.9 Å. The complex containing part C was refined from 43,025 particles to an average resolution of 6.2 Å. All three maps were unsharpened and filtered according to local resolution (Cardone *et al.*, 2013). All structural figures were generated using UCSF Chimera (Pettersen *et al.*, 2004) and PyMOL (Schrödinger, LLC.).

Structural modeling

The X-ray crystallographic structure of the Pol II-TFIIS complex (PDB entry: 3PO3 (Cheung & Cramer, 2011)) was used as the starting reference model and placed into the cryo-EM map using UCSF Chimera (Pettersen *et al.*, 2004). Amino acids 68-89 and 132-168 of the protrusion domain of Rpb2 were replaced by the coordinates from the PDB entry 5C4J (Barnes *et al.*, 2015). Template DNA (nucleotides 1-33), non-template DNA (nucleotides 7-39) and RNA (nucleotides 7-20) were derived from the PDB entry 5FLM (Bernecky *et al.*, 2016). The models were fitted as rigid bodies into the density map using ‘Fit in Map’ in UCSF Chimera (Pettersen *et al.*, 2004). The C-terminal GTPase-like domain of Cdc73 (PDB entry: 3V46) was fitted into the map of the complex containing part C using UCSF Chimera (Pettersen *et al.*, 2004) and adjusted in COOT (Emsley *et al.*, 2010) based on the XL-MS results. We also calculated the approximate molecular weight of the density, which is present in addition to Pol II, with Chimera using TFIIS as a reference (Pettersen *et al.*, 2004).

Binding assays involving analytical sucrose gradient ultracentrifugation

To investigate the binding of TFIIF and Paf1C to Pol II, analytical sucrose gradient ultracentrifugation was carried out. A sucrose gradient was generated by mixing equal volumes of light solution and heavy solution 2 (Table 15) using a gradient mixer (Gradient Master 108, BioComp Instruments).

Reconstituted Pol II-Paf1C-TFIIS EC was divided into portions. The first portion was incubated with a 1.8-fold molar excess of TFIIF (Plaschka *et al.*, 2016) while the other portion was incubated with buffer D as a control. Next the reconstituted complexes were applied on top of the gradient. After ultracentrifugation at 32,000 rpm in a SW60 swinging bucket rotor (Beckman Coulter) for 16 h at 4 °C, 200 μ L fractions were collected by pipetting carefully from top to bottom of the tube before analyzing them by SDS-PAGE. The assay was repeated but starting from a Pol II-TFIIF complex and then incubating with excess Paf1C-TFIIS.

4tU-Seq data collection

4-thiouracil (4tU) labeling of cellular RNA was performed as described (Sun *et al.*, 2012) with minor changes. 40 mL of each replicate culture were used for metabolic RNA labeling. Yeast cells were grown in YPD medium overnight, diluted to an OD_{600nm} of 0.1, and grown to an OD_{600nm} of 0.8. 4tU (Sigma, 2 M in DMSO) was added to the media at a final concentration of 5 mM, and cells were harvested after 6 min of labeling by centrifugation at 2,465 g and 30 °C for 1 min. The supernatant was discarded and the pellet resuspended in RNAlater solution (Ambion/Applied Biosystems) and then flash frozen in liquid nitrogen. Cell number was determined from an aliquot with a Cellometer N10 (Nexus) cell counter. Total RNA was extracted with phenol chloroform. RNA spike-ins were added to cell pellets at the first step of RNA purification (Schwalb *et al.*, 2016). Amount of spike-ins was adjusted to the cell number of each sample (120 ng of spike-in mix to 2.5×10^8 cells for all samples). Labeled RNA was chemically biotinylated and purified using streptavidin-coated magnetic beads as described (Dolken *et al.*, 2008). Sequencing libraries were prepared according to the manufacturer's recommendations using the Ovation Universal RNA-Seq System (NuGEN). Libraries were quantified with Qubit 1.0. Libraries were pooled and sequenced on an Illumina HiSeq 2500 sequencer.

4tU-Seq data analysis

Data analysis was performed as described (Schulz *et al.*, 2013), with minor modifications. Briefly, paired-end 50 base pair reads with additional 6 base pair of barcodes were obtained for labeled RNA. Reads were demultiplexed and aligned to the *S. cerevisiae* genome (sacCer3, version 64.2.1) using STAR (version 2.3.0) (Dobin *et al.*, 2013). SAMTools was used to quality

filter SAM files (Dobin *et al.*, 2013). Alignments with MAPQ smaller than 7 (-q 7) were skipped and only proper pairs (-f99, -f147, -f83, -f163) were selected. Further processing of the RNA-Seq data was carried out using the R/Bioconductor environment. Piled-up counts for every genomic position were summed up over replicates, using physical coverage i.e. counting both sequenced bases covered by reads and unsequenced bases spanned between proper mate-pair reads. We used a spike-in normalization strategy (Schwalb *et al.*, 2016) to allow observation of global shifts in RNA synthesis rates.

Fluorescence anisotropy assays with Paf1C

Paf1C-Ctr9- Δ C913 and Paf1C-Ctr9- Δ C913- Δ Rtf1 proteins were expressed and purified as described above. 5' 6- fluorescein (FAM) labeled coding RNA-33 (Table 8) were obtained from Integrated DNA Technologies and dissolved in water to 100 μ M. Sequences were designed as ssRNA, which is according to the natural sequence of exposed coding strand, and random RNA-33 (see Table 8) refers to an artificial single strand RNA, which mimics the natural ssRNA with the same GC content. Their sequences are 5'-FAM-AAUAUUCAAGACGAUUUAGACGAUAAUAUCAUA-3' and 5'-FAM-AAUAUUCAAGCAGUAUUAGCAGUAAUAUCAUA-3', respectively. Protein was serially diluted in 2-fold steps in buffer F (Table 16). Protein (12.5 μ L, 23 nM- 12 μ M), RNA (2.5 μ L, 96 nM) and 15 μ L 2 times assay buffer G (Table 16) were mixed gently and incubated on ice for 10 min. The assay was brought to a final volume of 30 μ L system within 8 nM florescent labeled RNA and 10 nM- 5 μ M protein. To note, we added 5 μ g/mL yeast tRNA (Sigma) as a competitor to get ride of non-specific binding. The reaction was incubated for 20 min at room temperature in the dark. 20 μ L of each solution was transferred to a Greiner 384 Flat Bottom Black Small volume plate.

Fluorescence anisotropy was measured at 30 °C with an Infinite® M1000Pro reader (Tecan) with an excitation wavelength of 470 nm (\pm 5 nm), an emission wavelength of 518 nm (\pm 20 nm) and a gain of 72. All measurements were repeated three times and results were averaged and subsequently analyzed with GraphPad Prism Version 6. Error bars are representative of the standard deviation from the mean of triplicates.

Transcription elongation assay

For transcription elongation assay, tailed templates as listed and indicated were used (Table 16). RNA was labeled with 5' 6- fluorescein (FAM). Tailed template DNA and FAM-RNA were annealed as described. Purified Pol II was mixed with 2-fold excess template DNA-RNA and incubated at 25 °C at 600 rpm for 10 min. Then 2-fold excess of non-template DNA was added and incubated at 25 °C at 600 rpm for 10 min, the Pol II-templates was diluted to 500 nM using TB buffer (Table 17). Four types of ribonucleotide (NTPs) were mixed in equimolar amount in water to a concentration of 25 mM and diluted to 2.5 mM with TB buffer. Each reaction was performed in a 10- μ L system, containing 100 nM Pol II-templates. Additional proteins or equivalent volume of TB was added. 2 μ L of 2.5 mM NTPs was added to a final concentration of 0.5 mM to star the reaction and incubated in a pre-heated Thermocycler (Biometra) for 10 min at 37 °C. The reaction was quenched by adding 2x stop buffer (Table 17) and further incubated at 37 °C for 1 h and 95 °C for 5 min. The 20% TBE urea gel was pre-run in 1x TBE buffer (Table 17) for 15 min at 300 V before the sample was loaded. The gel was run for 2 h at 300 V in dark and imaged with Typhoon FLA 9500 (GE Healthcare).

2.2.4 Bioinformatics tools

Gene and protein sequences were obtained from the NCBI or *S. cerevisiae* genome (SGD) databases. ClustalW (Larkin *et al.*, 2007) and Clustal Omega (Sievers *et al.*, 2011) were used for sequence alignments. Psipred was used to analysis the secondary structure predictions(Jones, 1999). HHpred (Alva *et al.*, 2016; Hildebrand *et al.*, 2009; Meier & Soding, 2015; Remmert *et al.*, 2011; Soding, 2005; Soding *et al.*, 2005) and SMART (Letunic *et al.*, 2015) were used to predict the motifs and low complexity regions. ProtParm online software was used to predict the protein specific parameters.

3. Results and Discussion

3.1 Recombinant Paf1 complex (Paf1C)

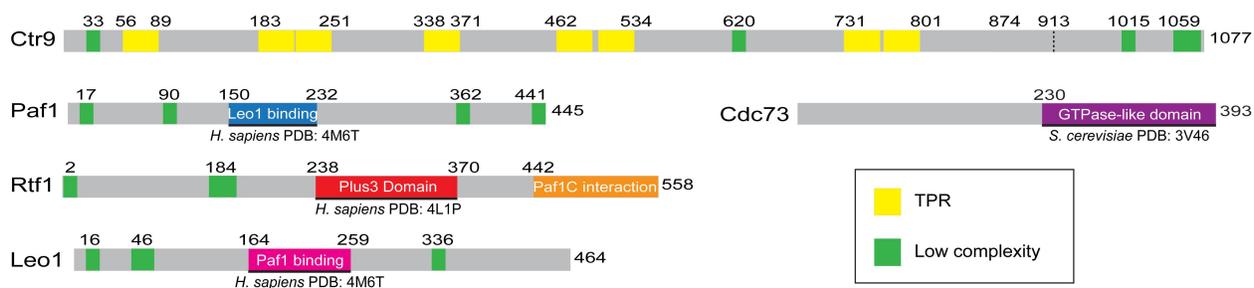


Figure 3: Predicted domain structure of the five Paf1C subunits from *S. cerevisiae*. The predicted TPR motifs (yellow) and low complexity regions (green) were defined using SMART (Letunic *et al.*, 2015). Structurally resolved regions are indicated with black lines. The Plus3 domain in Rtf1 (red) binds single-stranded DNA (de Jong *et al.*, 2008) and the Spt5 CTR (Wier *et al.*, 2013). A Paf1C-interacting region (orange) in Rtf1 (Warner *et al.*, 2007) was confirmed in this work.

Paf1C from the yeast *S. cerevisiae* has a molecular weight of 340 kDa and consists of five subunits, Paf1, Leo1, Ctr9, Cdc73, and Rtf1 (Figure 3). Analysis of the primary and secondary structures of these subunits (Alva *et al.*, 2016; Hildebrand *et al.*, 2009; Letunic *et al.*, 2015; Meier & Soding, 2015; Remmert *et al.*, 2012; Soding, 2005; Soding *et al.*, 2005) predicted known structured domains (Amrich *et al.*, 2012; H. Chen *et al.*, 2012; X. Chu *et al.*, 2013; de Jong *et al.*, 2008; Wier *et al.*, 2013) and eight tetratricopeptide repeats (TPR) in Ctr9 (Figure 3). In addition, multiple regions of low sequence complexity were detected in all Paf1C subunits except Cdc73, consistent with the known flexibility of Paf1C.

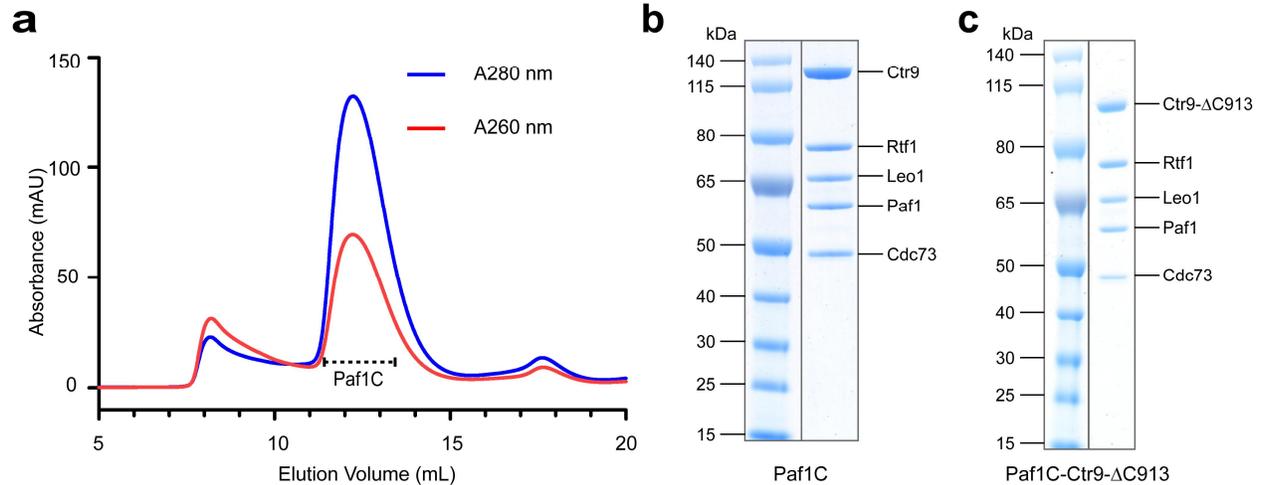


Figure 4: Preparation of recombinant Paf1C. **a**, Size-exclusion chromatogram (Superose 6 10/300; GE Healthcare) of recombinant full-length Paf1C indicates a pure complex free of nucleic acids. **b**, Coomassie-stained SDS-PAGE analysis of recombinant five-subunit Paf1C after size-exclusion chromatography reveals the presence of all five subunits in apparently stoichiometric amounts. Molecular weight markers are indicated on the left. The identity of subunits was confirmed by mass spectrometry. **c**, Same as **b** but for purified complex Paf1C-Ctr9- Δ C913.

To study Paf1C structurally, we established preparation of pure recombinant Paf1C after co-expression of its subunits in *E. coli* (Figure 4a, 4b, and also see 2.2.3 Specific methods). The five Paf1C subunits were co-expressed from three vectors in *E. coli* and the complex purified using chromatographic methods (Figure 4a). We obtained approximately 0.4 mg of pure Paf1C per liter of *E. coli* cell culture. Purified Paf1C contained all five subunits in apparently stoichiometric amounts (Figure 4b).

3.2 Paf1C structural core and flexible periphery

To map a stable structural core of Paf1C, we identified a Ctr9 variant that lacked 164 amino acids at its C-terminus (Ctr9- Δ C913) but was sufficient to form a stable complex with the other subunits after co-expression (Paf1C-Ctr9- Δ C913) (Figure 4c).

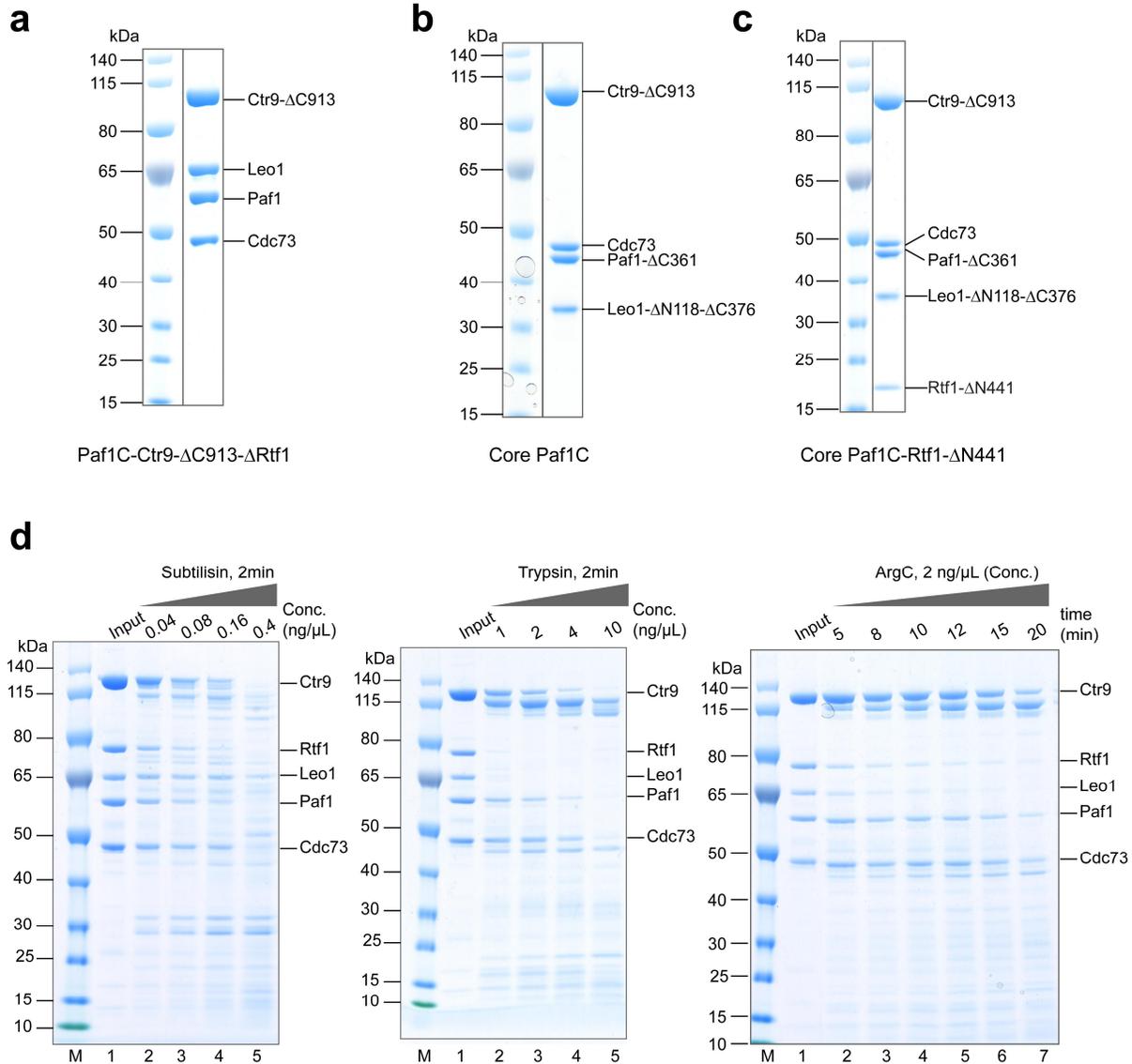


Figure 5: Preparation of various Paf1C recombinant proteins. **a-c**, Coomassie-stained SDS-PAGE analysis of (a) Paf1C-Ctr9- Δ C913- Δ Rtf1, (b) core Paf1C, and (c) core Paf1C-Rtf1 Δ N441 after size-exclusion chromatography (Superose 6 10/300; GE Healthcare) reveals four-subunit and five-subunit Paf1C variants. Molecular weight markers are indicated on the left. The identity of subunits was confirmed by mass spectrometry. **d**, Limited proteolysis using subtilisin, trypsin, and ArgC under either various concentrations or reaction times shows Paf1C contains many flexible regions.

We also obtained a recombinant Paf1C variant that additionally lacked subunit Rtf1 (Paf1C-Ctr9- Δ C913- Δ Rtf1, Figure 5a). To further delineate the core of Paf1C, we used limited proteolysis and Edman sequencing (Figure 5d and see 2.2.3 Specific methods) and designed deletion mutants of the remaining three subunits. Using iterative truncation, co-expression, and co-purification, we defined a structural core of Paf1C that additionally lacked the C-terminal

region of Paf1 and both terminal regions of Leo1. The resulting Paf1C core contained Ctr9- Δ C913, Paf1- Δ C361, Leo1- Δ N118- Δ C376, and Cdc73 (Figure 5b). The defined Paf1C core comprises 1,926 amino acid residues out of a total of 2,937 residues, i.e. 65% of the total protein mass. Thus one third of Paf1C forms flexible regions on the periphery of the complex.

3.3 Paf1C and TFIIS bind Pol II synergistically

To investigate whether purified Paf1C binds to yeast Pol II *in vitro*, we performed pull-down assays using biotinylated Pol II coupled to streptavidin beads (see 2.2.3 Specific methods). Paf1C interacted with Pol II, albeit in a substoichiometric manner (Figure 6a, lane 4). Considering the known interaction of human Paf1C with TFIIS (Kim *et al.*, 2010), we tested whether TFIIS enhances Pol II-Paf1C binding. Indeed, TFIIS strongly enhanced binding of Paf1C to Pol II in our assay (Figure 6a, lane 5). In contrast, an N-terminal deletion variant of TFIIS lacking 130 residues (TFIIS- Δ N130) could not enhance Pol II-Paf1C interaction (Figure 6b, lane 7 and 8). These results suggest that the N-terminal domain I of TFIIS, which is mobile in the Pol II-TFIIS structure (Kettenberger *et al.*, 2003), interacts with Paf1C to increase its affinity to Pol II. In addition, Paf1C lacking the C-terminal domain of Ctr9 and Rtf1 retained Pol II binding (Figure 6c, lane 6 and 8) and TFIIS-enhanced binding (Figure 6c, lane 7 and 9). These results show that recombinant Paf1C interacts with Pol II *in vitro*, that this interaction does not require Rtf1, and that it is strongly enhanced by TFIIS.

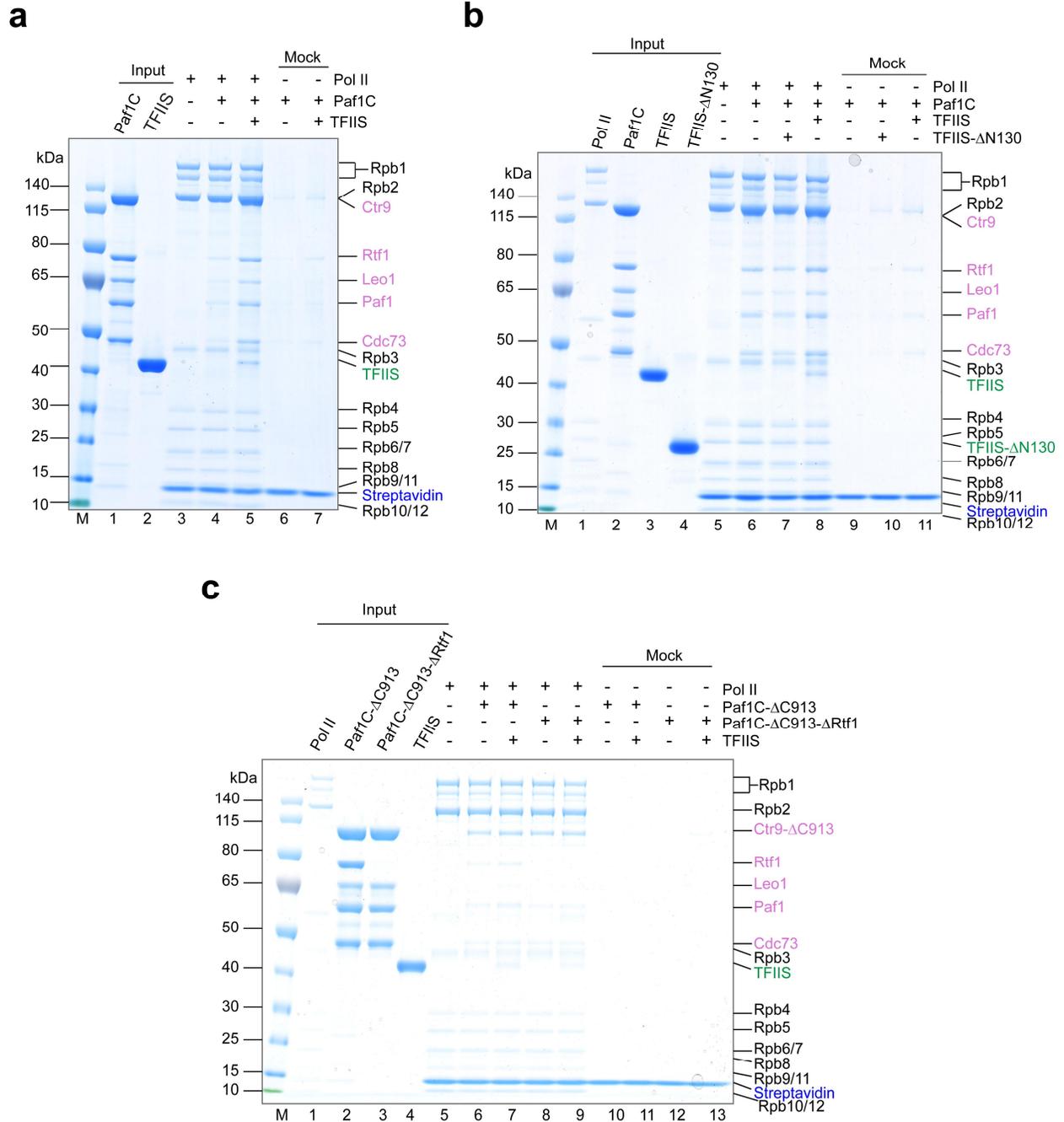


Figure 6: Analysis of Pol II-Paf1C-TFIIS EC interactions. SDS-PAGE analysis of Pol II-Paf1C-TFIIS binding assays. Subunits from Pol II, Paf1C, and TFIIS are labeled in black, pink, and green, respectively. **a**, TFIIS enhances Pol II-Paf1C binding. Lane 4 and 5 show Pol II-Paf1C and Pol II-Paf1C-TFIIS interactions, respectively. When TFIIS is present, Pol II and Paf1C bind better. **b**, TFIIS, but not TFIIS-ΔN130, cooperatively binds to Pol II-Paf1C (lane 7 and 8). **c**, Paf1C lacking Rtf1 and the C-terminal domain of Ctr9 retains Pol II binding (lane 6 and 8) and TFIIS-enhanced binding (lane 7 and 9).

3.4 Cryo-EM analysis

To determine the structure of Paf1C bound to a Pol II EC, we prepared a complex containing the complete 12-subunit Pol II (Sydow *et al.*, 2009), a DNA-RNA scaffold (Bernecky *et al.*, 2016), full-length TFIIS carrying two point mutations that render it inactive in RNA cleavage stimulation (Cheung & Cramer, 2011; Kettenberger *et al.*, 2003) (see 2.2.3 Specific methods), and recombinant Paf1C lacking the Ctr9 C-terminus (Paf1C-Ctr9- Δ C913), which is more stable than Paf1C with full-length Ctr9. The complex contained all 18 polypeptides in apparently stoichiometric amounts after sucrose gradient ultracentrifugation (Figure 7a). After gradient fixation (Kastner *et al.*, 2008; Stark, 2010) (Figure 7b), the sample contained single particles, as revealed by EM in negative stain, and was subjected to cryo-EM data collection. A total of 2,741 movie stacks were collected on a Titan Krios equipped with a Gatan K2 direct electron detection device (Figure 8).

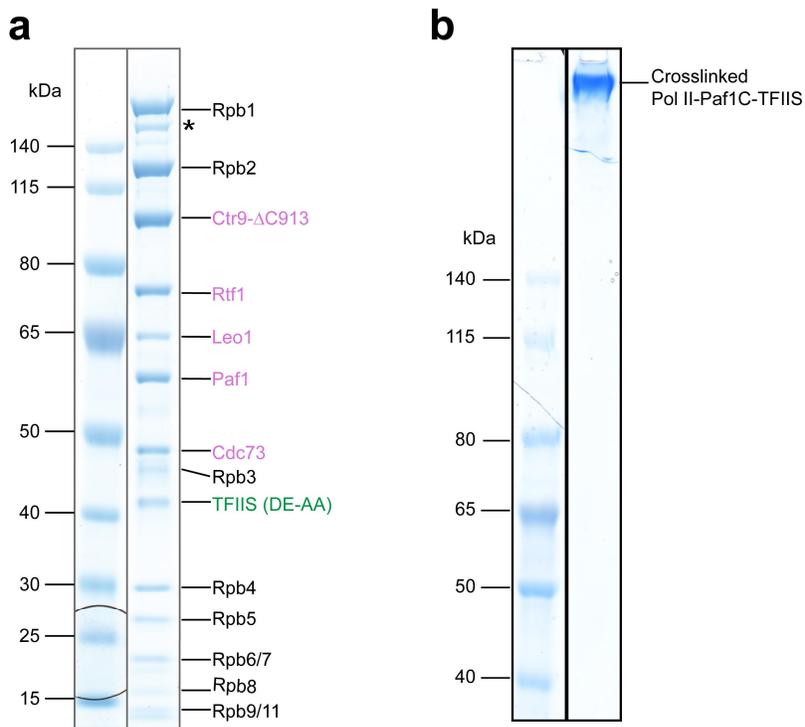


Figure 7: Pol II-Paf1C-TFIIS sample preparation for cryo-EM. **a**, SDS-PAGE analysis (Coomassie staining) of Pol II-Paf1C-TFIIS EC after sucrose gradient ultracentrifugation. Subunits from Pol II and Paf1C are labeled in black and pink, respectively. The inactive variant of TFIIS is labeled in green. The asterisk marks a degradation product of Rpb1. **b**, SDS-PAGE analysis of crosslinked Pol II-Paf1C-TFIIS after sucrose gradient fixation. A shift to higher molecular weight indicates successful crosslinking.

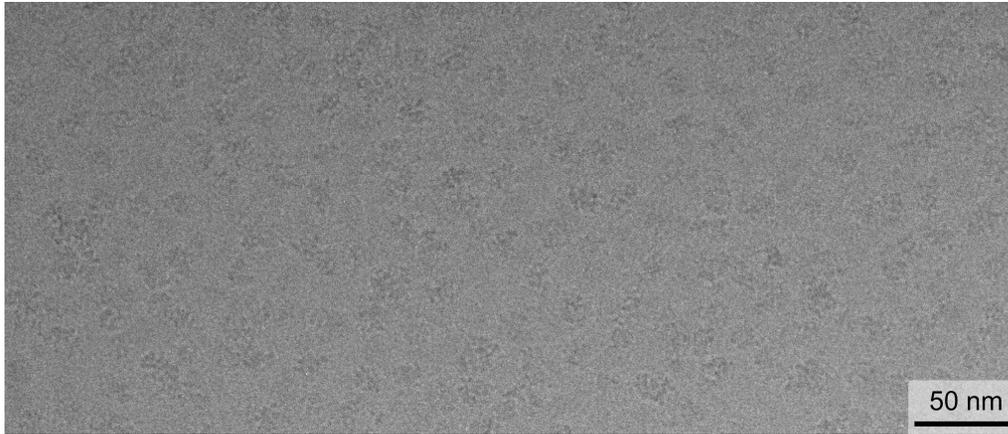


Figure 8: A representative cryo-EM micrograph of Pol II-Paf1C-TFIIS. 50 nm scale bar is included as a size reference.

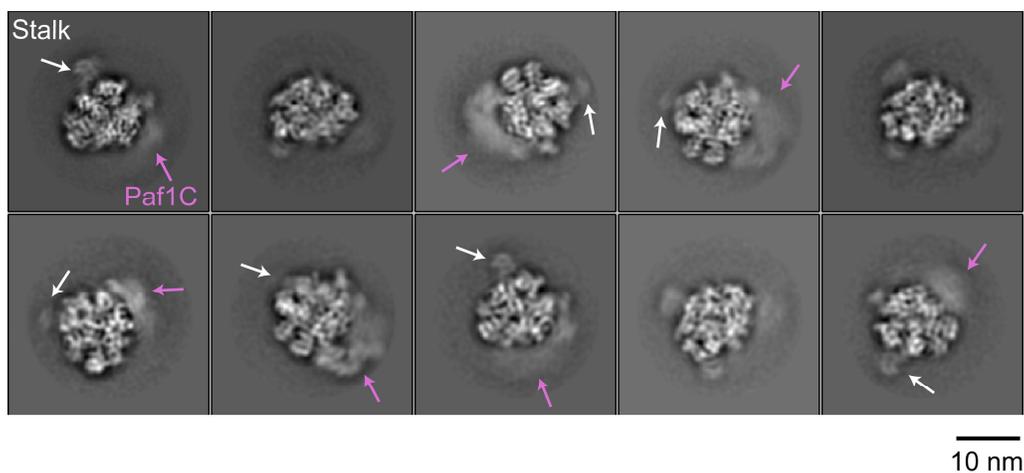


Figure 9: 2D cryo-EM class averages of Pol II-Paf1C-TFIIS. Ten representative reference-free two-dimensional (2D) cryo-EM class averages reveal the known flexibility of the Pol II stalk subcomplex Rpb4/7 (white arrow) and density for Paf1C (pink arrow).

From the cryo-EM micrographs, a total of 947,597 particles were extracted. Particle images were processed and subjected to reference-free two-dimensional (2D) classification in RELION (Scheres, 2012), yielding 370,841 particles after clearance. In several 2D classes, smeared densities for the peripheral Paf1C and the Pol II stalk subcomplex Rpb4-Rpb7 were observed on the polymerase surface (Figure 9). As a reference we used a reconstruction of bovine Pol II (Bernecky *et al.*, 2016) filtered to low resolution (50 Å). After particle polishing in

RELION, 3D-classification was used to separate out 158,422 particles that contained Pol II-TFIIS EC that lost Paf1C. We refined the remaining 212,419 Paf1C containing particles, which led to a structure at 5.5 Å global resolution that contained the Pol II-TFIIS EC that closely resembled the crystal structure (Figure 10, Figure 11c, 11d, and Supplementary Figure 3). Further sorting resulted in three different Pol II-Paf1C-TFIIS EC reconstructions that we refer to as A, B, and C (Figure 10 and see 2.2.3 Specific methods). These reconstructions showed average resolutions of 5.7 Å, 5.9 Å and 6.2 Å, respectively (FSC=0.143, Figure 11a, 11b, and see Specific methods) and enabled rigid-body fitting of the Pol II-TFIIS crystal structure (Barnes *et al.*, 2015; Cheung & Cramer, 2011) and the DNA-RNA model (Bernecky *et al.*, 2016) (Figure 11c, 11d, and see 2.2.3 Specific methods). The reconstructions did not reveal structural changes within the Pol II-TFIIS complex compared to the known crystal structure.

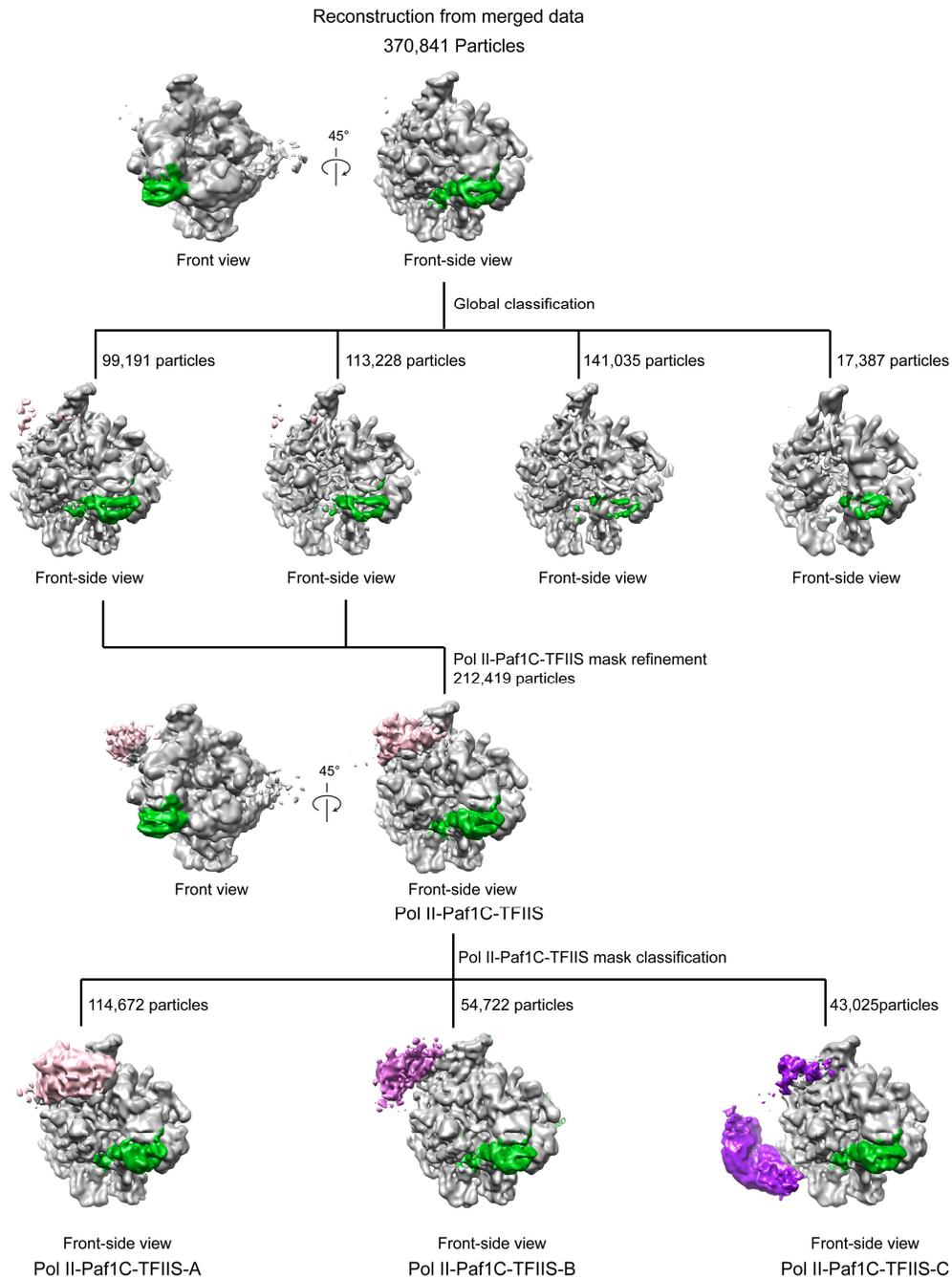


Figure 10: Computational sorting of cryo-EM particle images. Particles were 3D-classified in RELION to reveal additional features of Pol II-Paf1C-TFIIS after 2D classification. Classifications were performed without image alignment by global 3D classification using a spherical mask. After global classification, we separated Pol II-Paf1C-TFIIS and Pol II-TFIIS complexes. We further merged two classes with additional density near the protrusion. After auto-refinement and subsequent 3D classification with a mask encompassing Pol II-Paf1C-TFIIS, we obtained three structures displaying different parts of Paf1C. Pol II and scaffold are colored in grey, whereas TFIIS is in green. Parts A, B, and C of Paf1C are in pink, orchid and purple, respectively. To visualize the density from Paf1C, a front-side view is related by one rotation from the front view (Cramer *et al.*, 2000) as indicated.

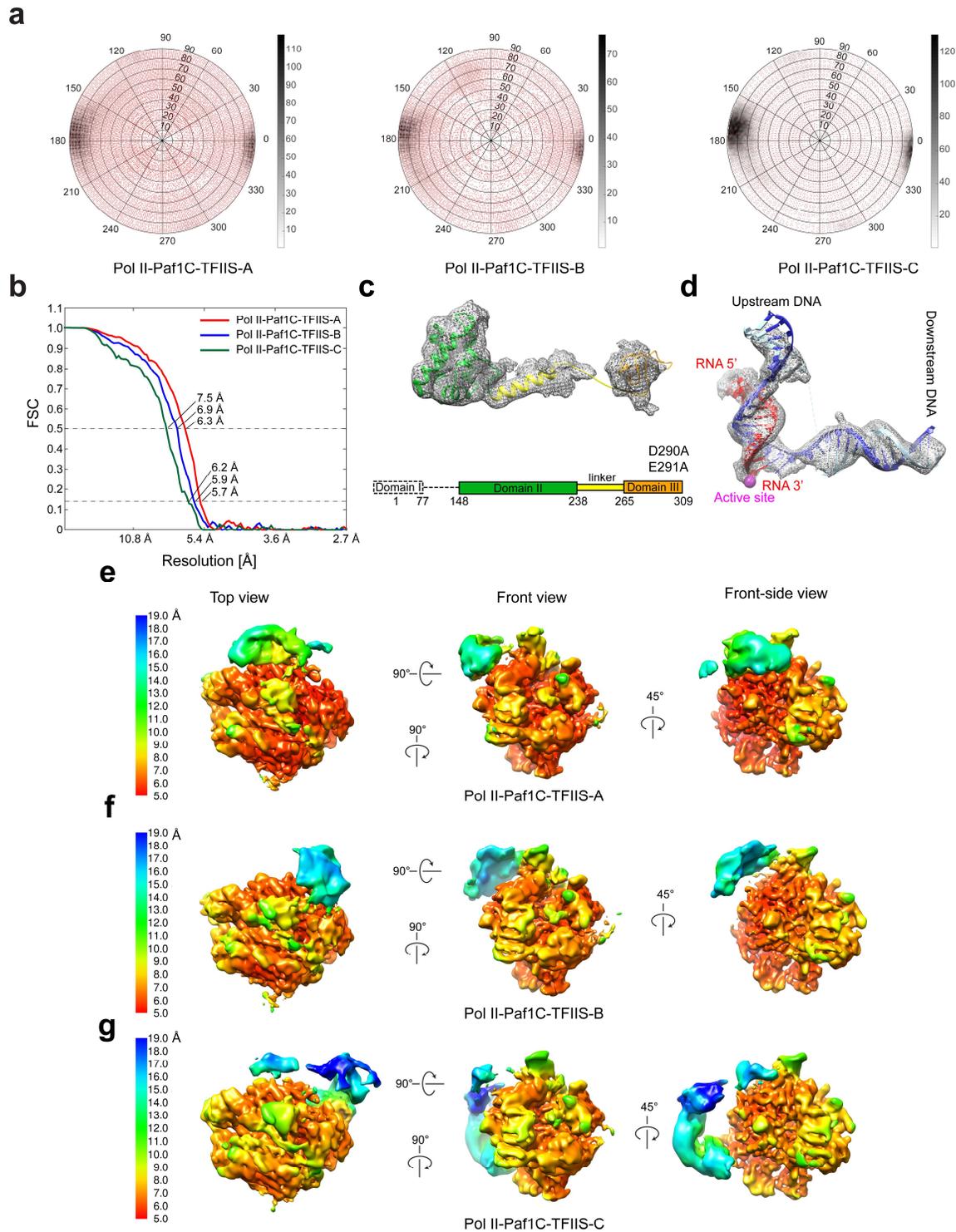


Figure 11: Cryo-EM reconstruction of Pol II-Paf1C-TFIIS subclasses. **a**, Angular distribution of Pol II-Paf1C-TFIIS reconstructions A, B, and C from cryo-EM single particle reconstructions. Red dots indicate at least one particle was assigned within 1° of the point. Black shading indicates the number of

particles assigned to a given view. **b**, The FSC curves for the Pol II-Paf1C-TFIIS subclasses. Color scheme: Pol II-Paf1C-TFIIS reconstruction A, red; reconstruction B, blue; reconstruction C, dark green. The resolution of the reconstructions using the Fourier shell cutoff at 0.5 and 0.143 is shown. **c**, TFIIS EM density (grey mesh) from the unsharpened Pol II-Paf1C-TFIIS. Crystal structure derived from 3PO3 is fitted. A schematic diagram of TFIIS domains is shown below. Domain II, linker, and domain III are represented in green, yellow, and orange boxes. Point mutations of two functionally essential acidic residues are indicated. No density was observed for TFIIS residues 1-148 containing domain I and a linker, which are labeled with dashed lines due to their flexibility. **d**, EM density for the DNA-RNA hybrid and up- and downstream DNA of Pol II-Paf1C-TFIIS (grey mesh) with the final model superimposed. Color scheme: non-template DNA, cyan; template DNA, blue; RNA, red; active site, magenta. **e-g**, Top, front and front-side views of unsharpened (**e**) Pol II-Paf1C-TFIIS reconstruction A, (**f**) reconstruction B, and (**g**) reconstruction C colored by local resolution. All the densities are shown at the same threshold level and locally normalized. Resolution bars are shown on the left.

3.5 Architecture of Pol II-Paf1C-TFIIS elongation complex

Reconstructions A, B, and C revealed different portions of Paf1C on the Pol II surface that we refer to as Paf1C parts A, B, and C, respectively. The observation that these three parts of Paf1C could only be observed in different cryo-EM reconstructions shows that their relative orientation is flexible. Due to this mobility on the Pol II surface, the local resolution of Paf1C was low (Figure 11e-g). When contoured at a level where no noise peaks are observed, the volumes of parts A, B, and C could account for approximately 50, 45, and 90 kDa of folded protein. Because part C overlaps with part A by approximately 10 kDa, a total of ~175 kDa of Paf1C is visible by cryo-EM, corresponding to ~50% of the total mass of Paf1C, suggesting that most of the structured core (65%) is revealed. We combined reconstructions A-C into a composite map that is further interpreted below.

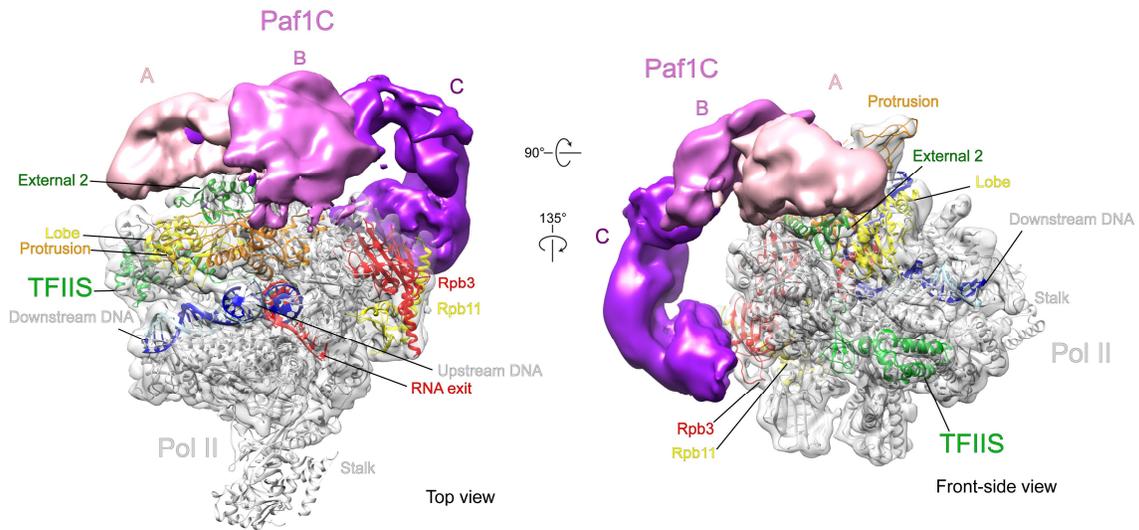


Figure 12: Cryo-EM structure of Pol II-Paf1C-TFIIS EC. Two views of the EM density for the Pol II-Paf1C-TFIIS EC and the fitted structure of the Pol II-TFIIS EC. The view on the left is from the top (Cramer *et al.*, 2000), and the view on the right is related by two rotations as indicated. Density for the Pol II-TFIIS EC subcomplex is shown as a silver semi-transparent surface. The composite electron density for Paf1C is shown as a solid surface colored in pink, magenta, and purple for Paf1C parts A, B, and C, respectively. The same threshold level in Chimera was used to contour densities. The structure of the Pol II-TFIIS EC is shown in ribbon representation (non-template DNA, cyan; template DNA, blue; RNA, red, TFIIS, green; Rpb2 lobe, yellow; Rpb2 protrusion, orange; Rpb2 external 2, green; Rpb3, red; Rpb11, yellow).

Paf1C binds to the outer surface of Pol II on the Rpb2 side, spanning from the Rpb2 lobe, external 2, and protrusion domains to subunit Rpb3 and reaching near the rim of the funnel opposite the active center cleft (Figure 12). Paf1C part A contacts mainly the external 2 domain and likely the protrusion domain of Rpb2. Part B contacts the protrusion and part C contacts Rpb11 and Rpb3 around its helix $\alpha 3$. Part B also bridges between parts A and C, which otherwise do not contact each other. Taken together, cryo-EM revealed that the Paf1C core may be divided in three parts, with parts A and C contacting Pol II at the Rpb2 lobe and Rpb3, respectively, and part B bridging between parts A and C.

3.6 Crosslinking analysis

To confirm the contacts of Paf1C with Pol II, and to assign Paf1C subunits to the three parts of Paf1C, we subjected the Pol II-Paf1C-TFIIS EC to chemical crosslinking coupled to mass spectrometry (XL-MS). We used the crosslinking reagent bis(sulfosuccinimidyl) suberate (BS3), which reacts with lysine side chains and N-termini. In each crosslinking dataset, two replicates were measured. All the spectra of the crosslinks were filtered at a false discovery rate (FDR) cutoff of 1% and the maximum score value (negative logarithm of E-value) higher than 5. Each crosslink was required to have a minimal spectral count of 2 in the two replicates of each dataset (see 2.2.3 Specific methods). We obtained 239 unique inter-subunit crosslinks (Figure 13, Table 18, and Supplementary Table 13).

Table 18: Summary of BS3 crosslinks in Pol II-Paf1C-TFIIS EC. The numbers of inter-protein crosslinks are given. Inter-protein crosslinks between Pol II subunits are not shown here for clarity. Each unique crosslink is identified by more than one crosslink. BS3 is a crosslinker with 11.4 Å spacer arm and reacts efficiently with amino groups (lysine and N-terminus).

Protein 1	Protein 2	Unique Crosslinks
		BS3
Paf1C	Pol II	49
Ctr9	Rpb1	1
	Rpb3	1
Rtf1	Rpb2	1
Leo1	Rpb1	2
	Rpb2	35
Paf1	Rpb2	3
Cdc73	Rpb11	4
	Rpb2	1
	Rpb3	1
TFIIS	Paf1C	2
	Leo1	2
Paf1C	Paf1C	136
Ctr9	Rtf1	36
	Leo1	7
	Paf1	20
	Cdc73	19
	Rtf1	6
Leo1	Paf1	19
	Paf1	25
	Cdc73	4
Pol II	Pol II	47
Pol II	TFIIS	5
	Total	239

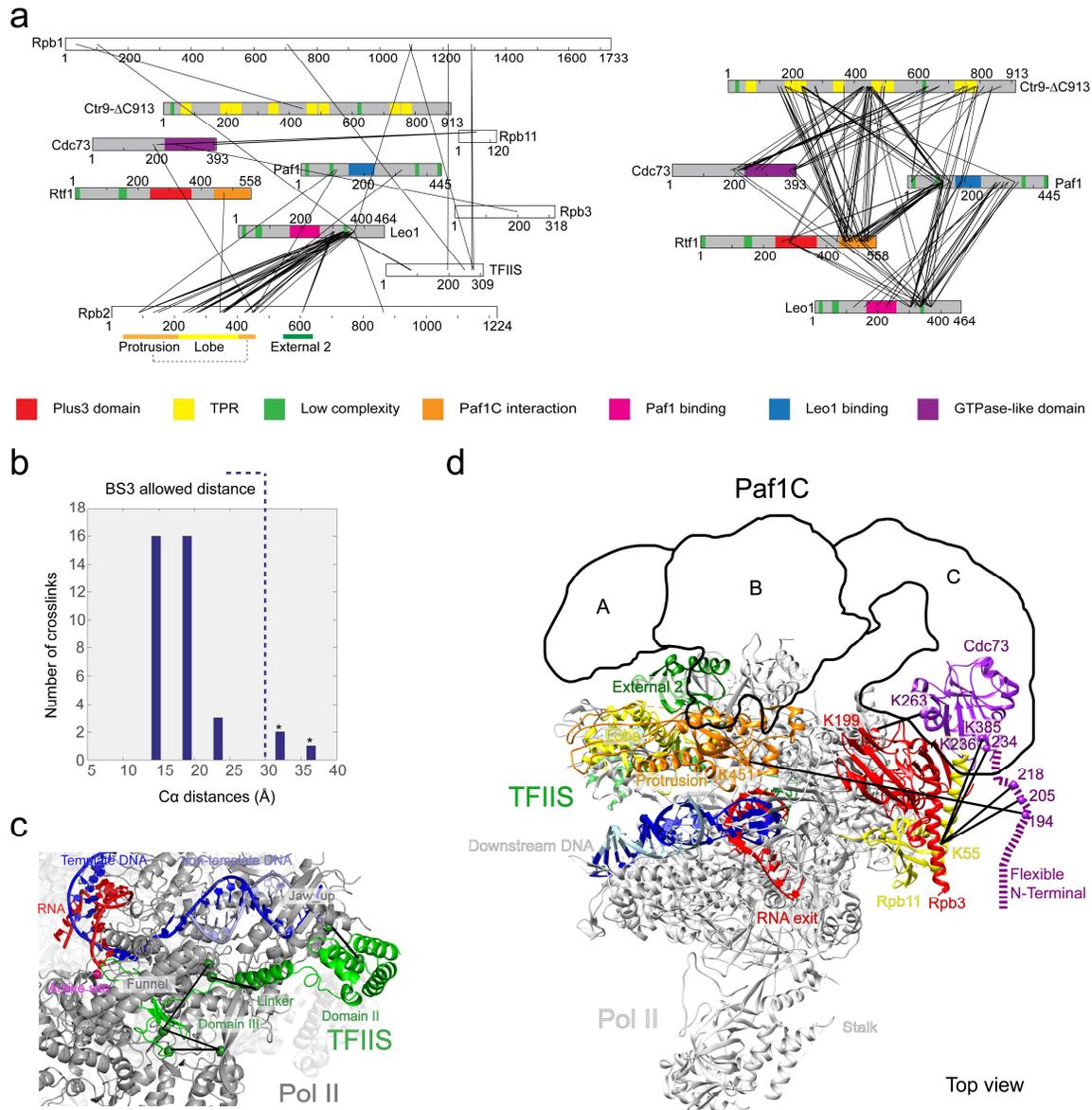


Figure 13: Crosslinking analysis of Pol II-Paf1C-TFIIS EC. **a**, Network diagrams of pairwise crosslinks (black lines) obtained after incubation of Pol II-Paf1C-TFIIS EC with BS3. The left diagram depicts inter-subunit crosslinks between Pol II and Paf1C, Pol II and TFIIS, and between Paf1C and TFIIS. The right diagram depicts crosslinks between Paf1C subunits. Crosslinks between Pol II subunits were excluded for clarity. Paf1C subunits colored as in Figure 3. In addition, the regions of protrusion, lobe, and external 2 in Rpb2 are indicated with orange, yellow, and green bars below Rpb2 diagram, respectively. **b**, Ca-Ca distance distribution of observed Pol II-Pol II and Pol II-TFIIS crosslinks for lysine residues that were resolved in the Pol II-TFIIS crystal structure. These crosslinks serve as a positive control for our crosslinking data. Asterisks indicate crosslinks to flexible protein regions. **c**, Crosslinks between Pol II and TFIIS (black lines) are consistent with the known Pol II-TFIIS complex crystal structure. Color coding as in Figure 12. **d**, Position of the Cdc73 C-terminal GTPase-like domain (purple) contacting Rpb3. The crystal structure of the GTPase-like domain was docked into density of the

correct size and oriented according to crosslinks to Rpb3 and Rpb11 (black lines). Paf1C density is shown as an outline in the same orientation as in Figure 12. View is from the top.

We first mapped crosslinks between Pol II subunits and between Pol II and TFIIS onto the Pol II-TFIIS EC crystal structure (Cheung & Cramer, 2011). The C α -C α distances between crosslinked residues were within the allowed distances of 30 Å for 35 out of 38 inter-subunit pairs. The other 3 crosslinks that showed longer C α distances fell in regions with structural flexibility and higher crystallographic B-factors (Figure 13b, 13c, and Supplementary Table 3). In addition, we observed that Leo1 was crosslinked to two lysine residues (K78, K80) in domain I of TFIIS, consistent with our pull down assays (see above). These results provided an internal, positive control for our crosslinking approach.

3.7 Paf1C subcomplex architecture

The data also contained 136 inter-subunit crosslinks within Paf1C (Figure 13a, Table 18, and Supplementary Table 3). These included 25 crosslinks between Paf1 and Leo1, consistent with the known dimerization of Paf1 and Leo1 subunits. Crosslinks of the Leo1 residue K225 to Paf1 residues K167 and K170 can be rationalized with the Paf1-Leo1 X-ray structure. A total of 27 crosslinks were observed between the Paf1-Leo1 dimer and Ctr9, revealing their proximity. The subassembly Paf1-Leo1-Ctr9 crosslinked to the C-terminal regions of Cdc73 and Rtf1. Consistent with this, the C-terminal region of Rtf1 suffices to bind Paf1C, because its co-expression with the Paf1C core resulted in a stable complex (Figure 5c). We also crosslinked free Paf1C and the obtained crosslinking pattern was similar (Supplementary Figure 2 and Supplementary Table 2).

We also obtained 49 inter-subunit crosslinks between Paf1C and Pol II (Figure 13a, Table 18, and Supplementary Table 3). 44 of these crosslinks map to the Rpb2 lobe, protrusion, and external 2 domains, and subunits Rpb3 and Rpb11 (Supplementary Table 3), consistent with the location of Paf1C observed by cryo-EM. Extensive crosslinking between the Paf1-Leo1 dimer and the Pol II protrusion and lobe show that the Paf1-Leo1 heterodimer resides in part A. Crosslinks between Cdc73 and Pol II subunits Rpb3 and Rpb11 reveal that Cdc73 resides in part C. Indeed, part C contains a globular density that contacts Rpb3 and corresponds in size to the crystal structure of the C-terminal GTPase-like domain of Cdc73 (Amrich *et al.*, 2012) (Figure 13d). We could fit this crystal structure to the globular density (Pettersen *et al.*, 2004) such that

the crosslinks between Cdc73 and Pol II were explained. Three crosslinks occurred between the fitted structure and Rpb3, whereas Cdc73 residues in the adjacent N-terminal tail crosslinked to the nearby Rpb11 subunit (Figure 13d). Finally, Ctr9 can be assigned to part B because it crosslinks to both, Paf1-Leo1 in part A and Cdc73 in part C. Thus Ctr9 bridges between Paf1-Leo1 and Cdc73, which both contact Pol II. Ctr9 did not crosslink efficiently to Pol II, consistent with the bridging density B that does not reach the Pol II surface.

Combining the crosslinking data with the cryo-EM results and our mapping of flexible regions in Paf1C enabled us to derive the overall architecture of Paf1C bound to Pol II. The sizes of the EM densities can be reconciled with the Paf1C subunit molecular weights as follows. Part A reveals only about half of the Paf1-Leo1 heterodimer (~50 out of 105 kDa), the other half is flexible. Part B reflects about half of the mass (~45 kDa) of Ctr9- Δ C913 (105 kDa), whereas the other half of Ctr9 resides in part C, which is much larger than the Cdc73 GTPase-like domain alone (~20 kDa). Finally, Rtf1 (65 kDa) remains flexible.

3.8 Competitive Pol II binding of Paf1C and TFIIF

Superposition of the Pol II-Paf1C-TFIIS EC onto our previously reported Pol II initiation complex structure (Plaschka *et al.*, 2016) reveals a clash between Paf1C part A and the dimerization domain of the initiation factor TFIIF bound to the outer lobe of Pol II at an overlapping position (Figure 14a). This suggested that binding of Paf1C and TFIIF to Pol II is competitive or mutually exclusive. To investigate this, we performed binding assays using analytical sucrose gradient ultracentrifugation (see 2.2.3 Specific methods). We incubated preformed Pol II-Paf1C-TFIIS complex with a 1.8-fold molar excess of TFIIF and separated the resulting complexes on a 10-30% sucrose gradient. Based on subsequent SDS-PAGE analysis, we estimate that TFIIF replaced Paf1C in about half of the Pol II complexes (Figure 14b). When we instead incubated preformed Pol II-TFIIF complex with Paf1C and TFIIS, some TFIIF was displaced, TFIIS was able to join the complex, and Paf1C bound to a low extent (Figure 14c). These results indicate that Paf1C and TFIIF bind to Pol II in a competitive manner, consistent with the cryo-EM reconstruction.

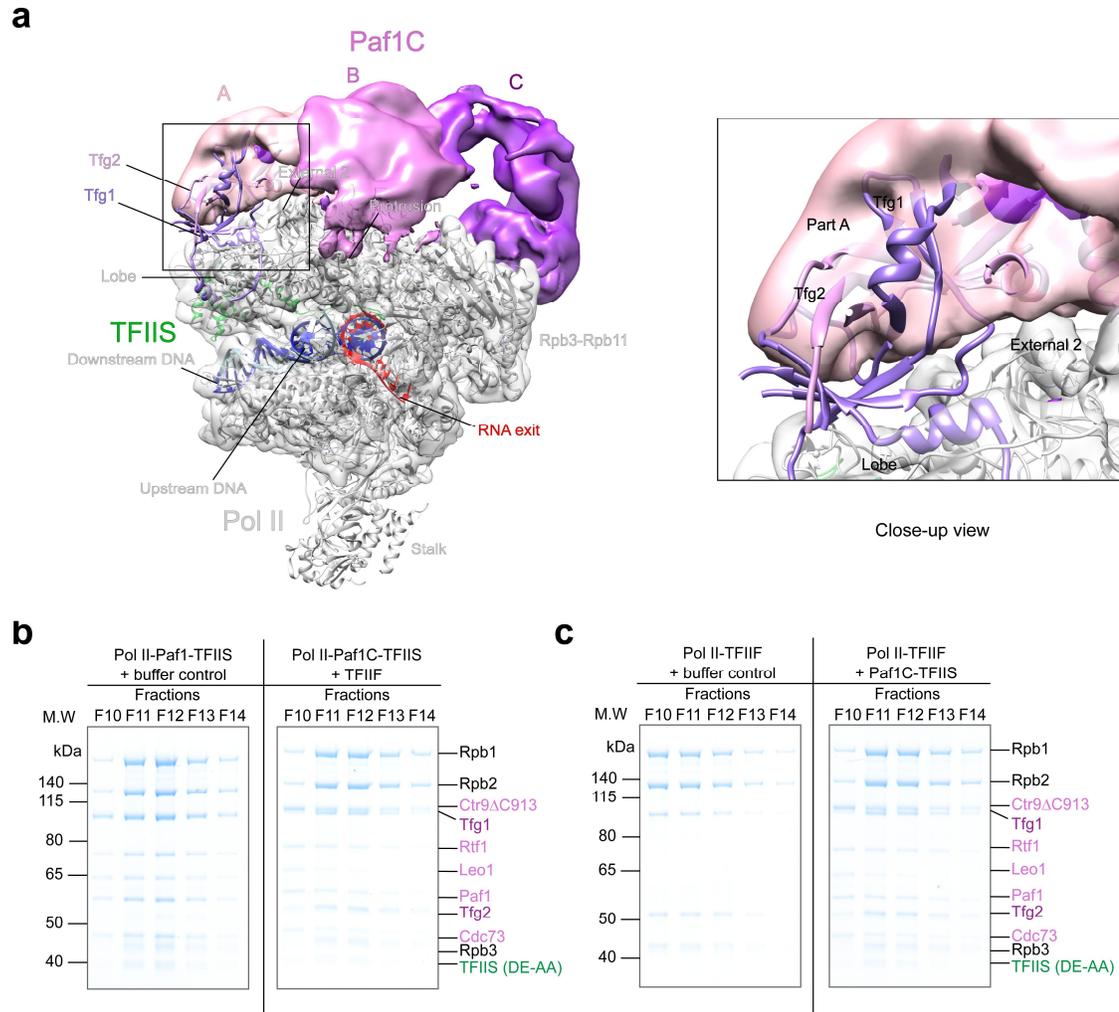


Figure 14: Paf1C and TFIIF compete for Pol II binding. **a**, Positioning the initiation factor TFIIF dimerization domain as in the Pol II initiation complex (PDB accession: 5FYW) (Plaschka *et al.*, 2016) onto the Pol II-Paf1C-TFIIS EC cryo-EM reconstruction results in a clash of TFIIF with Paf1C (top view). Colors as in Figure 12. The TFIIF dimerization domain is colored in medium purple (TFIIF subunit Tfg1) and plum (TFIIF subunit Tfg2). A close-up view of the clashing region is shown on the right. **b-c**, SDS-PAGE analysis of TFIIF and Paf1C competitive binding to Pol II after analytical sucrose gradient ultracentrifugation. In each panel, the gels on the left depict complexes without competitor protein. The gels on the right represent complexes formed after competitor addition. In panel **b**, a 1.8-fold molar excess of TFIIF was added to preformed Pol II-Paf1C-TFIIS EC. In panel **c**, Paf1C-TFIIS was added to preformed Pol II-TFIIF complex.

3.9 Paf1C is globally required for Pol II transcription

Thus far it had not been demonstrated whether Paf1C is a general transcription factor in yeast. To investigate whether Paf1C is globally required for transcription *in vivo*, or whether it has gene-specific functions, we monitored RNA synthesis with 4tU-Seq in yeast (Eser *et al.*, 2016; Miller *et al.*, 2011; Schulz *et al.*, 2013). This method uses metabolic RNA labeling with 4-thiouracil (4tU) coupled to strand-specific sequencing of labeled, newly synthesized RNA. We used 4tU-Seq to monitor RNA synthesis in strains lacking either Paf1 (Δ Paf1) or Rtf1 (Δ Rtf1), and compared this to a wild-type (WT) strain using global normalization based on spike-in probes (Schwalb *et al.*, 2016). Two biological replicates were measured. We found that knockout of Paf1 or Rtf1 led to a strong, global decrease in synthesis of mRNA transcripts (Figure 15), showing that Paf1C is globally required for normal Pol II transcription.

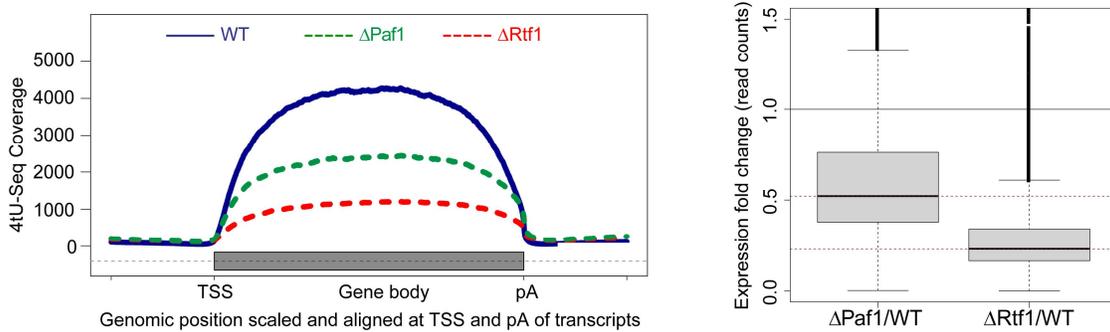


Figure 15: Paf1C is globally required for Pol II transcription in yeast. Left panel: coverage of newly synthesized RNA measured by 4tU-seq in WT (solid blue line), Δ Rtf1 (dashed red line), and Δ Paf1 yeast cells (dashed green line). 4tU-seq signals were globally normalized using spike-in probes. The gene body is defined as the region spanning from the transcription start site (TSS) to the polyadenylation (pA) site and is depicted as a grey box. Right panel: box plot showing the expression fold change of read counts for Δ Paf1 and Δ Rtf1 vs. the WT strain. Dashed lines indicate median fold (0.52 for Δ Paf1 and 0.23 for Δ Rtf1). The majority of mRNAs are below 1 (indicated by the solid black line). Decreased mRNAs make up 85% for Δ Paf1 and 94% for Δ Rtf1 of all mRNAs with a coverage of at least 2 in the WT strain.

3.10 Paf1C binds RNA independent of Rtf1

Considering that Leo1 subunit and Rtf1 subunit both have RNA binding capacity (Dermody & Buratowski, 2010), we investigated whether Paf1C lacking Rtf1 influences the binding to RNA. We used fluorescent anisotropy titration assays. We first analyzed Paf1C-Ctr9- Δ C913 and Paf1C-Ctr9- Δ C913- Δ Rtf1 with a 33 nucleotide natural coding RNA sequence bearing a 5' 6-FAM label. We found Paf1C binds the RNA with a K_d of $\sim 0.43 \pm 0.02 \mu\text{M}$, whereas Paf1C-Ctr9- Δ C913- Δ Rtf1 bound more weakly with a K_d of $\sim 1.17 \pm 0.10 \mu\text{M}$. Our data reveals Paf1C binds to RNA, and that Rtf1 subunit plays a key role. To gain further insights, we designed a random RNA sequence, in which a few nucleotides were modified but with the same GC content compared to the natural coding RNA. The result we observed indicates Paf1C also binds strongly to random sequence RNA, with Paf1C- Δ Rtf1 binding weaker, suggesting a universal function (Figure 16). These results also suggest Paf1C binding to RNA does not only rely on Rtf1. Unfortunately, we couldn't purify Leo1 depletion or Paf1 depletion variants and test their binding to RNA, because they are crucial subunits for forming Paf1C.

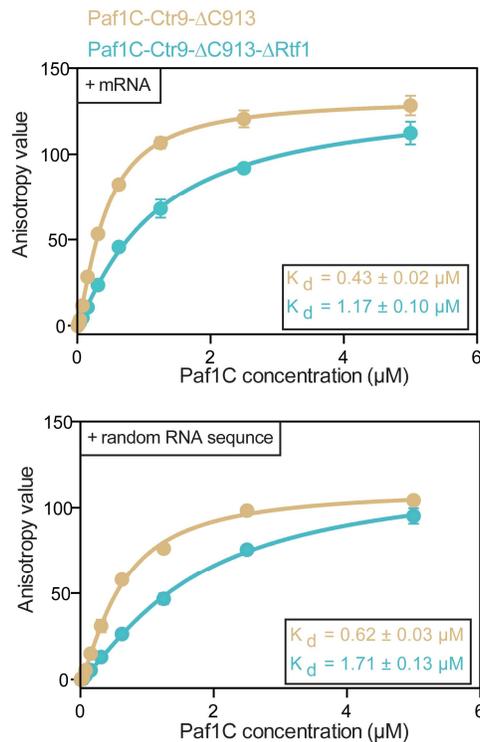


Figure 16: Rtf1 is very important for Paf1C binding to RNA. Purified Paf1C-Ctr9- Δ C913 and Paf1C-Ctr9- Δ C913- Δ Rtf1 variants were used in this fluorescent anisotropy assay to test the binding to natural coding RNA (up) and random RNA (bottom), respectively. The color code shows the different binding of Paf1 complexes to RNA with binding affinities (K_d).

4. Conclusion and Outlook

From this thesis described here, I showed that we were able to purify the 340 kDa Paf1C recombinantly, and used a combination of cryo-EM and crosslinking to show that Paf1C forms a trilobal architecture with three parts. Parts A and C contain Paf1-Leo1 and Cdc73, respectively, and contact Pol II near the external domain 2 and Rpb3, respectively. Ctr9 forms part B and extends into part C, and forms a flexible bridge between parts A and C that is not directly contacting Pol II. These results provide a framework for integrating published data. Ctr9 is known to form a scaffold that bridges between the Paf1-Leo1 heterodimer and Cdc73 as predicted (X. Chu *et al.*, 2013). The Cdc73 GTPase-like domain binds Pol II, consistent with the requirement for this domain to recruit Paf1C to chromatin (Amrich *et al.*, 2012). Most of Rtf1 is flexible, including the Plus3 domain that binds the phosphorylated C-terminal region (CTR) of Spt5 (Wier *et al.*, 2013), which is also flexibly connected to the EC (Figure 17). My results also explain how Paf1C and TFIIS can cooperatively bind to Pol II (Kim *et al.*, 2010). Pull down assays show that TFIIS facilitates Pol II-Paf1C binding via its domain I, I observe crosslinking between Leo1 and TFIIS domain I, and this interaction apparently occurs near the jaw-lobe module of Pol II. We arrive at an overall architecture of the Paf1C on the Pol II elongation complex.

The results elucidate the general transcription cycle, because Paf1C is generally required for transcription *in vivo*. In particular, they provide insights into the transition from transcription initiation to elongation, when initiation factors are replaced by elongation factors on the Pol II surface (Guo & Price, 2013; Kwak & Lis, 2013; Mayer *et al.*, 2010). In the initiation complex, association of Paf1C with Pol II is impaired because TFIIF blocks the site for binding of Paf1C part A. Similarly, TFIIE blocks the site for binding the elongation factor Spt4-Spt5 on the clamp (Grohmann *et al.*, 2011; Martinez-Rucobo *et al.*, 2011). These observations indicate why Paf1C and Spt4-Spt5 bind the EC only upon disassembly of the initiation complex. The individual interactions of these elongation factors with Pol II appear to be weak, but may be facilitated by interactions between them (Mayekar *et al.*, 2013; Qiu *et al.*, 2012), and between Paf1C and TFIIS (Kim *et al.*, 2010). I note that the second contact of Paf1C with Pol II, formed by part C, in particular the GTPase domain of Cdc73, with Rpb3, appears unique as it does not overlap with known factor positions on Pol II. This contact may therefore help to retain Paf1C in the

elongation complex when TFIIIF reassembles during elongation (Zawel *et al.*, 1995). These studies provide a framework for further dissection of the multiple functions of Paf1C in transcription.

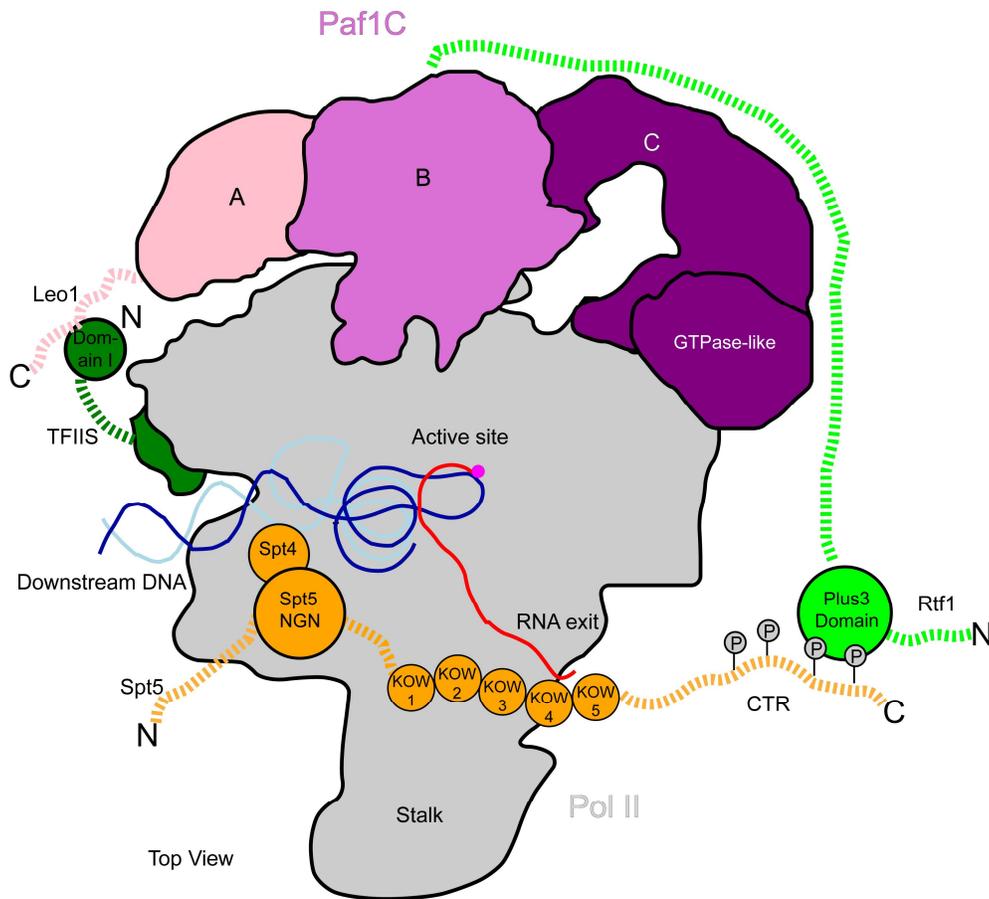


Figure 17: Model of the Pol II EC with bound elongation factors. Summary of our current understanding of elongation factor location and interactions on the Pol II EC surface. In addition to direct Pol II interactions by domains in TFIIIS, Spt5, and Paf1C, elongation factors contain many flexible domains and regions that interact. In particular, the Paf1-Leo1 heterodimer may contact the N-terminal domain I of TFIIIS and the C-terminal region of Rtf1 is anchored to Paf1C mainly by Paf1 and Ctr9 subunits shown in XL-MS, whereas its Plus3 domain binds the flexible C-terminal repeat domain (CTR) of Spt5. The N-terminal tail of Rtf1 is important for recruiting the chromatin remodeler Chd1 (Simic *et al.*, 2003). Dash lines in this model indicate flexible regions.

Although an initial understanding of the highly flexible Pol II-Paf1C-TFIIIS EC emerges in this study, there are still many unanswered questions.

For example, what is the near-atomic structure of Paf1C? So far, the crystal structural studies on this complex have been limited to a few domains. After much effort towards getting

this complex, attempts to crystallize the Paf1C and its variants failed, mainly due to its many flexible regions. Even the solid core of Paf1C didn't lead to a good crystal. We obtained crystals when we used Paf1C-Ctr9- Δ C913- Δ Rtf1 with chymotrypsin in situ proteolysis. The crystals diffracted to 7 Å resolution, which was insufficient to resolve the structure. We expect that x-ray free electron laser (XFEL) could be helpful. Also, as better imaging hardware, phase-plates, and other technologies are developing, solving this 'flexible' complex, with a molecular weight less than 0.5 MDa, to a near-atomic resolution will be possible using cryo-EM. Moreover, a higher-resolution structure of Pol II-Paf1C-TFIIS would reveal the exact contacting residues between Pol II and Paf1C, and will make clear whether there are any conformational changes between free Paf1C and Pol II-bound Paf1C, which will help us better understand its mechanism.

To get high-resolution structures, more transcriptional factors besides TFIIS may be needed to stabilize Paf1C, especially the flexible regions. This leads to another question; what are the factors responsible for productive transcription elongation? We already knew that Spt4-Spt5 is a key complex in regulating transcription elongation, and the CTR domain of Spt5 physically interacts with Paf1C via the plus3 domain of the Rtf1 subunit (Wier *et al.*, 2013). Pol II provides vital links between them acting as a bridge. In the human system, the regulation of transcription elongation is more sophisticated. P-TEFb not only phosphorylates the CTR to facilitate the recruitment of Paf1C, but also phosphorylates NELF to release Pol II from the paused state (Kwak & Lis, 2013). Studies on a yeast Pol II-Paf1C-TFIIS-Spt4-Spt5 complex will be instructive for understanding these mechanisms in human.

Does Paf1C play any role in backtracking? TFIIS has at least a dual role: to help normal transcription elongation, and to stimulate the cleavage activity of Pol II (Cheung & Cramer, 2011; Kettenberger *et al.*, 2003, 2004). In our study, we observed that TFIIS enhances the Pol II-Paf1C binding most likely via its domain I. What role does Paf1C play together with TFIIS? Does it enhance the elongation rate or influence the cleavage activity? From our in vitro transcription elongation assays, we hardly see any effect of Paf1C during transcription elongation in the presence or absence of TFIIS. However, so far we focused only on in vitro analysis, how Paf1C functions together with TFIIS *in vivo* remains unclear.

Another question is about the composition of the Paf1C in yeast. What role does Ski8 play? Ski8 is a WD repeat protein. Ski8 is a 44-kDa protein (PDB entry: 1SQ9) (Madrona & Wilson, 2004). In yeast, Ski8 is reported as a component of Ski complex and to primarily

function in the regulation of exosome-mediated, 3' to 5' degradation of damaged mRNA and meiosis (Halbach *et al.*, 2013; Madrona & Wilson, 2004). As we know, Paf1C is conserved from yeast to human. Ski8 is a component of human Paf1C instead of Rtf1 (Cao *et al.*, 2015; X. Chu *et al.*, 2013; Rozenblatt-Rosen *et al.*, 2009; Zhu *et al.*, 2005). In contrast, in yeast Ski8 has not been reported to be a Paf1C subunit, but it nevertheless shows genetic interaction with Cdc73 (Wilmes *et al.*, 2008). Our study already suggested that Rtf1 is not an essential subunit for forming Paf1C in agreement with that Rtf1 has an independent role during the transcription elongation (Cao *et al.*, 2015). Why is the composition of Paf1C from the two species different? Is there any evolutionary significance to this difference? We could expand our research to include Ski8, to better understand its functions in the two species, and compare them. Furthermore, we could also investigate Paf1C from other species to see if there are any evolutionary trends.

The transition from initiation to elongation is poorly understood. The transition is always accompanied by a series of dynamic processes, including the exchange of factors. Learning what the intermediate complexes are will greatly help us comprehend this process.

All these lead to an advanced question; what are the mechanisms Paf1C uses to facilitate chromatin transcription? The reason for this is that we have addressed here the interface between Paf1C and the transcribing polymerase, whereas insights into the mechanisms of chromatin transcription will only come from a structural analysis of the interface of Paf1C with nucleosomes and chromatin-associated factors. This poses a formidable challenge for the future.

References

- Agarwal, K., Baek, K. H., Jeon, C. J., Miyamoto, K., Ueno, A., & Yoon, H. S. (1991). Stimulation of transcript elongation requires both the zinc finger and RNA polymerase II binding domains of human TFIIS. *Biochemistry*, *30*(31), 7842-7851.
- Alva, V., Nam, S. Z., Soding, J., & Lupas, A. N. (2016). The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. *Nucleic Acids Res*, *44*(W1), W410-415. doi:10.1093/nar/gkw348
- Amrich, C. G., Davis, C. P., Rogal, W. P., Shirra, M. K., Heroux, A., Gardner, R. G., . . . VanDemark, A. P. (2012). Cdc73 subunit of Paf1 complex contains C-terminal Ras-like domain that promotes association of Paf1 complex with chromatin. *J Biol Chem*, *287*(14), 10863-10875. doi:10.1074/jbc.M111.325647
- Arimbasseri, A. G., Rijal, K., & Maraia, R. J. (2013). Transcription termination by the eukaryotic RNA polymerase III. *Biochim Biophys Acta*, *1829*(3-4), 318-330. doi:10.1016/j.bbagr.2012.10.006
- Armache, K. J., Kettenberger, H., & Cramer, P. (2003). Architecture of initiation-competent 12-subunit RNA polymerase II. *Proc Natl Acad Sci U S A*, *100*(12), 6964-6968. doi:10.1073/pnas.1030608100
- Arndt, K. M., & Reines, D. (2015). Termination of Transcription of Short Noncoding RNAs by RNA Polymerase II. *Annu Rev Biochem*, *84*, 381-404. doi:10.1146/annurev-biochem-060614-034457
- Barnes, C. O., Calero, M., Malik, I., Graham, B. W., Spahr, H., Lin, G., . . . Calero, G. (2015). Crystal Structure of a Transcribing RNA Polymerase II Complex Reveals a Complete Transcription Bubble. *Mol Cell*, *59*(2), 258-269. doi:10.1016/j.molcel.2015.06.034
- Belotserkovskaya, R., & Reinberg, D. (2004). Facts about FACT and transcript elongation through chromatin. *Curr Opin Genet Dev*, *14*(2), 139-146. doi:10.1016/j.gde.2004.02.004
- Belotserkovskaya, R., Saunders, A., Lis, J. T., & Reinberg, D. (2004). Transcription through chromatin: understanding a complex FACT. *Biochim Biophys Acta*, *1677*(1-3), 87-99. doi:10.1016/j.bbexp.2003.09.017
- Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J. M., & Cramer, P. (2016). Structure of transcribing mammalian RNA polymerase II. *Nature*, *529*(7587), 551-554. doi:10.1038/nature16482
- Blombach, F., Daviter, T., Fielden, D., Grohmann, D., Smollett, K., & Werner, F. (2013). Archaeology of RNA polymerase: factor swapping during the transcription cycle. *Biochem Soc Trans*, *41*(1), 362-367. doi:10.1042/BST20120274
- Borner, T., Aleynikova, A. Y., Zubo, Y. O., & Kusnetsov, V. V. (2015). Chloroplast RNA polymerases: Role in chloroplast biogenesis. *Biochim Biophys Acta*, *1847*(9), 761-769. doi:10.1016/j.bbabi.2015.02.004
- Bowman, E. A., & Kelly, W. G. (2014). RNA polymerase II transcription elongation and Pol II CTD Ser2 phosphorylation: A tail of two kinases. *Nucleus*, *5*(3), 224-236. doi:10.4161/nucl.29347
- Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. *Mol Cell*, *36*(4), 541-546. doi:10.1016/j.molcel.2009.10.019
- Bushnell, D. A., & Kornberg, R. D. (2003). Complete, 12-subunit RNA polymerase II at 4.1-A resolution: implications for the initiation of transcription. *Proc Natl Acad Sci U S A*, *100*(12), 6969-6973. doi:10.1073/pnas.1130601100

- Cabart, P., Ujvari, A., Pal, M., & Luse, D. S. (2011). Transcription factor TFIIF is not required for initiation by RNA polymerase II, but it is essential to stabilize transcription factor TFIIB in early elongation complexes. *Proc Natl Acad Sci U S A*, *108*(38), 15786-15791. doi:10.1073/pnas.1104591108
- Cao, Q. F., Yamamoto, J., Isobe, T., Tateno, S., Murase, Y., Chen, Y., . . . Yamaguchi, Y. (2015). Characterization of the Human Transcription Elongation Factor Rtf1: Evidence for Nonoverlapping Functions of Rtf1 and the Paf1 Complex. *Mol Cell Biol*, *35*(20), 3459-3470. doi:10.1128/MCB.00601-15
- Cardone, G., Heymann, J. B., & Steven, A. C. (2013). One number does not fit all: mapping local variations in resolution in cryo-EM reconstructions. *J Struct Biol*, *184*(2), 226-236. doi:10.1016/j.jsb.2013.08.002
- Chaudhary, K., Deb, S., Moniaux, N., Ponnusamy, M. P., & Batra, S. K. (2007). Human RNA polymerase II-associated factor complex: dysregulation in cancer. *Oncogene*, *26*(54), 7499-7507. doi:10.1038/sj.onc.1210582
- Chedin, S., Riva, M., Schultz, P., Sentenac, A., & Carles, C. (1998). The RNA cleavage activity of RNA polymerase III is mediated by an essential TFIIS-like subunit and is important for transcription termination. *Genes Dev*, *12*(24), 3857-3871.
- Chen, F. X., Woodfin, A. R., Gardini, A., Rickels, R. A., Marshall, S. A., Smith, E. R., . . . Shilatifard, A. (2015). PAF1, a Molecular Regulator of Promoter-Proximal Pausing by RNA Polymerase II. *Cell*, *162*(5), 1003-1015. doi:10.1016/j.cell.2015.07.042
- Chen, H., Shi, N., Gao, Y., Li, X., Teng, M., & Niu, L. (2012). Crystallographic analysis of the conserved C-terminal domain of transcription factor Cdc73 from *Saccharomyces cerevisiae* reveals a GTPase-like fold. *Acta Crystallogr D Biol Crystallogr*, *68*(Pt 8), 953-959. doi:10.1107/S09074444912017325
- Chen, S., McMullan, G., Faruqi, A. R., Murshudov, G. N., Short, J. M., Scheres, S. H., & Henderson, R. (2013). High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy*, *135*, 24-35. doi:10.1016/j.ultramic.2013.06.004
- Cheng, B., Li, T., Rahl, P. B., Adamson, T. E., Loudas, N. B., Guo, J., . . . Price, D. H. (2012). Functional association of Gdown1 with RNA polymerase II poised on human genes. *Mol Cell*, *45*(1), 38-50. doi:10.1016/j.molcel.2011.10.022
- Cheung, A. C., & Cramer, P. (2011). Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature*, *471*(7337), 249-253. doi:10.1038/nature09785
- Chu, X., Qin, X., Xu, H., Li, L., Wang, Z., Li, F., . . . Long, J. (2013). Structural insights into Paf1 complex assembly and histone binding. *Nucleic Acids Res*, *41*(22), 10619-10629. doi:10.1093/nar/gkt819
- Chu, Y., Simic, R., Warner, M. H., Arndt, K. M., & Prelich, G. (2007). Regulation of histone modification and cryptic transcription by the Bur1 and Paf1 complexes. *EMBO J*, *26*(22), 4646-4656. doi:10.1038/sj.emboj.7601887
- Combe, C. W., Fischer, L., & Rappsilber, J. (2015). xiNET: cross-link network maps with residue resolution. *Mol Cell Proteomics*, *14*(4), 1137-1147. doi:10.1074/mcp.O114.042259
- Cramer, P. (2002). Multisubunit RNA polymerases. *Curr Opin Struct Biol*, *12*(1), 89-97.
- Cramer, P., Armache, K. J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., . . . Vannini, A. (2008). Structure of eukaryotic RNA polymerases. *Annu Rev Biophys*, *37*, 337-352. doi:10.1146/annurev.biophys.37.032807.130008

- Cramer, P., Bushnell, D. A., Fu, J., Gnatt, A. L., Maier-Davis, B., Thompson, N. E., . . . Kornberg, R. D. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. *Science*, *288*(5466), 640-649.
- de Jong, R. N., Truffault, V., Diercks, T., Ab, E., Daniels, M. A., Kaptein, R., & Folkers, G. E. (2008). Structure and DNA binding of the human Rtf1 Plus3 domain. *Structure*, *16*(1), 149-159. doi:10.1016/j.str.2007.10.018
- Dermody, J. L., & Buratowski, S. (2010). Leo1 subunit of the yeast paf1 complex binds RNA and contributes to complex recruitment. *J Biol Chem*, *285*(44), 33671-33679. doi:10.1074/jbc.M110.140764
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., . . . Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15-21. doi:10.1093/bioinformatics/bts635
- Dolken, L., Ruzsics, Z., Radle, B., Friedel, C. C., Zimmer, R., Mages, J., . . . Koszinowski, U. H. (2008). High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA*, *14*(9), 1959-1972. doi:10.1261/rna.1136108
- Eick, D., & Geyer, M. (2013). The RNA polymerase II carboxy-terminal domain (CTD) code. *Chem Rev*, *113*(11), 8456-8490. doi:10.1021/cr400071f
- Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*, *66*(Pt 4), 486-501. doi:10.1107/S0907444910007493
- Engel, C., Sainsbury, S., Cheung, A. C., Kostrewa, D., & Cramer, P. (2013). RNA polymerase I structure and transcription regulation. *Nature*, *502*(7473), 650-655. doi:10.1038/nature12712
- Eser, P., Wachutka, L., Maier, K. C., Demel, C., Boroni, M., Iyer, S., . . . Gagneur, J. (2016). Determinants of RNA metabolism in the *Schizosaccharomyces pombe* genome. *Mol Syst Biol*, *12*(2), 857. doi:10.15252/msb.20156526
- Fernandez-Tornero, C., Moreno-Morcillo, M., Rashid, U. J., Taylor, N. M., Ruiz, F. M., Gruene, T., . . . Muller, C. W. (2013). Crystal structure of the 14-subunit RNA polymerase I. *Nature*, *502*(7473), 644-649. doi:10.1038/nature12636
- Fishburn, J., & Hahn, S. (2012). Architecture of the yeast RNA polymerase II open complex and regulation of activity by TFIIF. *Mol Cell Biol*, *32*(1), 12-25. doi:10.1128/MCB.06242-11
- Geiger, J. H., Hahn, S., Lee, S., & Sigler, P. B. (1996). Crystal structure of the yeast TFIIA/TBP/DNA complex. *Science*, *272*(5263), 830-836.
- Grohmann, D., Nagy, J., Chakraborty, A., Klose, D., Fielden, D., Ebright, R. H., . . . Werner, F. (2011). The initiation factor TFE and the elongation factor Spt4/5 compete for the RNAP clamp during transcription initiation and elongation. *Mol Cell*, *43*(2), 263-274. doi:10.1016/j.molcel.2011.05.030
- Guccione, E., Bassi, C., Casadio, F., Martinato, F., Cesaroni, M., Schuchlantz, H., . . . Amati, B. (2007). Methylation of histone H3R2 by PRMT6 and H3K4 by an MLL complex are mutually exclusive. *Nature*, *449*(7164), 933-937. doi:10.1038/nature06166
- Guo, J., & Price, D. H. (2013). RNA polymerase II transcription elongation control. *Chem Rev*, *113*(11), 8583-8603. doi:10.1021/cr400105n
- Hahn, S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol*, *11*(5), 394-403. doi:10.1038/nsmb763

- Halbach, F., Reichelt, P., Rode, M., & Conti, E. (2013). The yeast ski complex: crystal structure and RNA channeling to the exosome complex. *Cell*, *154*(4), 814-826. doi:10.1016/j.cell.2013.07.017
- Hausner, W., Lange, U., & Musfeldt, M. (2000). Transcription factor S, a cleavage induction factor of the archaeal RNA polymerase. *J Biol Chem*, *275*(17), 12393-12399.
- He, Y., Fang, J., Taatjes, D. J., & Nogales, E. (2013). Structural visualization of key steps in human transcription initiation. *Nature*, *495*(7442), 481-486. doi:10.1038/nature11991
- He, Y., Yan, C., Fang, J., Inouye, C., Tjian, R., Ivanov, I., & Nogales, E. (2016). Near-atomic resolution visualization of human transcription promoter opening. *Nature*, *533*(7603), 359-365. doi:10.1038/nature17970
- Hildebrand, A., Remmert, M., Biegert, A., & Soding, J. (2009). Fast and accurate automatic structure prediction with HHpred. *Proteins*, *77 Suppl 9*, 128-132. doi:10.1002/prot.22499
- Hoffmann, N. A., Jakobi, A. J., Moreno-Morcillo, M., Glatt, S., Kosinski, J., Hagen, W. J., . . . Muller, C. W. (2015). Molecular structures of unbound and transcribing RNA polymerase III. *Nature*, *528*(7581), 231-236. doi:10.1038/nature16143
- Holstege, F. C., Fiedler, U., & Timmers, H. T. (1997). Three transitions in the RNA polymerase II transcription complex during initiation. *EMBO J*, *16*(24), 7468-7480. doi:10.1093/emboj/16.24.7468
- Holstege, F. C., van der Vliet, P. C., & Timmers, H. T. (1996). Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. *EMBO J*, *15*(7), 1666-1677.
- Jennebach, S., Herzog, F., Aebersold, R., & Cramer, P. (2012). Crosslinking-MS analysis reveals RNA polymerase I domain architecture and basis of rRNA cleavage. *Nucleic Acids Res*, *40*(12), 5591-5601. doi:10.1093/nar/gks220
- Jeronimo, C., Bataille, A. R., & Robert, F. (2013). The writers, readers, and functions of the RNA polymerase II C-terminal domain code. *Chem Rev*, *113*(11), 8491-8522. doi:10.1021/cr4001397
- Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol*, *292*(2), 195-202. doi:10.1006/jmbi.1999.3091
- Kanamaru, K., & Tanaka, K. (2004). Roles of chloroplast RNA polymerase sigma factors in chloroplast development and stress response in higher plants. *Biosci Biotechnol Biochem*, *68*(11), 2215-2223.
- Kassube, S. A., Fang, J., Grob, P., Yakovchuk, P., Goodrich, J. A., & Nogales, E. (2013). Structural insights into transcriptional repression by noncoding RNAs that bind to human Pol II. *J Mol Biol*, *425*(19), 3639-3648. doi:10.1016/j.jmb.2012.08.024
- Kastner, B., Fischer, N., Golas, M. M., Sander, B., Dube, P., Boehringer, D., . . . Stark, H. (2008). GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat Methods*, *5*(1), 53-55. doi:10.1038/nmeth1139
- Kettenberger, H., Armache, K. J., & Cramer, P. (2003). Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell*, *114*(3), 347-357.
- Kettenberger, H., Armache, K. J., & Cramer, P. (2004). Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol Cell*, *16*(6), 955-965. doi:10.1016/j.molcel.2004.11.040
- Kim, J., Guermah, M., & Roeder, R. G. (2010). The human PAF1 complex acts in chromatin transcription elongation both independently and cooperatively with SII/TFIIS. *Cell*, *140*(4), 491-503. doi:10.1016/j.cell.2009.12.050

- Kim, J. L., Nikolov, D. B., & Burley, S. K. (1993). Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature*, *365*(6446), 520-527. doi:10.1038/365520a0
- Kim, Y., Geiger, J. H., Hahn, S., & Sigler, P. B. (1993). Crystal structure of a yeast TBP/TATA-box complex. *Nature*, *365*(6446), 512-520. doi:10.1038/365512a0
- Kinkelin, K., Wozniak, G. G., Rothbart, S. B., Lidschreiber, M., Strahl, B. D., & Cramer, P. (2013). Structures of RNA polymerase II complexes with Bye1, a chromatin-binding PHF3/DIDO homologue. *Proc Natl Acad Sci U S A*, *110*(38), 15277-15282. doi:10.1073/pnas.1311010110
- Kireeva, M. L., Lubkowska, L., Komissarova, N., & Kashlev, M. (2003). Assays and affinity purification of biotinylated and nonbiotinylated forms of double-tagged core RNA polymerase II from *Saccharomyces cerevisiae*. *Methods Enzymol*, *370*, 138-155. doi:10.1016/S0076-6879(03)70012-3
- Koch, C., Wollmann, P., Dahl, M., & Lottspeich, F. (1999). A role for Ctr9p and Paf1p in the regulation G1 cyclin expression in yeast. *Nucleic Acids Res*, *27*(10), 2126-2134.
- Komarnitsky, P., Cho, E. J., & Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev*, *14*(19), 2452-2460.
- Korinek, A., Beck, F., Baumeister, W., Nickell, S., & Plitzko, J. M. (2011). Computer controlled cryo-electron microscopy--TOM(2) a software package for high-throughput applications. *J Struct Biol*, *175*(3), 394-405. doi:10.1016/j.jsb.2011.06.003
- Kostrewa, D., Zeller, M. E., Armache, K. J., Seizl, M., Leike, K., Thomm, M., & Cramer, P. (2009). RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature*, *462*(7271), 323-330. doi:10.1038/nature08548
- Kowalik, K. M., Shimada, Y., Flury, V., Stadler, M. B., Batki, J., & Buhler, M. (2015). The Paf1 complex represses small-RNA-mediated epigenetic gene silencing. *Nature*, *520*(7546), 248-252. doi:10.1038/nature14337
- Krogan, N. J., Dover, J., Wood, A., Schneider, J., Heidt, J., Boateng, M. A., . . . Shilatifard, A. (2003). The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. *Mol Cell*, *11*(3), 721-729.
- Krogan, N. J., Kim, M., Ahn, S. H., Zhong, G., Kobor, M. S., Cagney, G., . . . Greenblatt, J. F. (2002). RNA polymerase II elongation factors of *Saccharomyces cerevisiae*: a targeted proteomics approach. *Mol Cell Biol*, *22*(20), 6979-6992.
- Kubota, Y., Tsuyama, K., Takabayashi, Y., Haruta, N., Maruyama, R., Iida, N., & Sugimoto, A. (2014). The PAF1 complex is involved in embryonic epidermal morphogenesis in *Caenorhabditis elegans*. *Dev Biol*, *391*(1), 43-53. doi:10.1016/j.ydbio.2014.04.002
- Kwak, H., & Lis, J. T. (2013). Control of transcriptional elongation. *Annu Rev Genet*, *47*, 483-508. doi:10.1146/annurev-genet-110711-155440
- Lahmy, S., Bies-Etheve, N., & Lagrange, T. (2010). Plant-specific multisubunit RNA polymerase in gene silencing. *Epigenetics*, *5*(1), 4-8.
- Laribee, R. N., Krogan, N. J., Xiao, T., Shibata, Y., Hughes, T. R., Greenblatt, J. F., & Strahl, B. D. (2005). BUR kinase selectively regulates H3 K4 trimethylation and H2B ubiquitylation through recruitment of the PAF elongation complex. *Curr Biol*, *15*(16), 1487-1493. doi:10.1016/j.cub.2005.07.028

- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., . . . Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, *23*(21), 2947-2948. doi:10.1093/bioinformatics/btm404
- Lehmann, E., Brueckner, F., & Cramer, P. (2007). Molecular basis of RNA-dependent RNA polymerase II activity. *Nature*, *450*(7168), 445-449. doi:10.1038/nature06290
- Letunic, I., Doerks, T., & Bork, P. (2015). SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res*, *43*(Database issue), D257-260. doi:10.1093/nar/gku949
- Li, X., Mooney, P., Zheng, S., Booth, C. R., Braunfeld, M. B., Gubbens, S., . . . Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat Methods*, *10*(6), 584-590. doi:10.1038/nmeth.2472
- Li, X., Zheng, S. Q., Egami, K., Agard, D. A., & Cheng, Y. (2013). Influence of electron dose rate on electron counting images recorded with the K2 camera. *J Struct Biol*, *184*(2), 251-260. doi:10.1016/j.jsb.2013.08.005
- Lisica, A., Engel, C., Jahnel, M., Roldan, E., Galburt, E. A., Cramer, P., & Grill, S. W. (2016). Mechanisms of backtrack recovery by RNA polymerases I and II. *Proc Natl Acad Sci U S A*, *113*(11), 2946-2951. doi:10.1073/pnas.1517011113
- Littlefield, O., Korkhin, Y., & Sigler, P. B. (1999). The structural basis for the oriented assembly of a TBP/TFB/promoter complex. *Proc Natl Acad Sci U S A*, *96*(24), 13668-13673.
- Liu, W. H., & Churchill, M. E. (2012). Histone transfer among chaperones. *Biochem Soc Trans*, *40*(2), 357-363. doi:10.1042/BST20110737
- Liu, X., Bushnell, D. A., & Kornberg, R. D. (2013). RNA polymerase II transcription: structure and mechanism. *Biochim Biophys Acta*, *1829*(1), 2-8. doi:10.1016/j.bbagr.2012.09.003
- Luo, Z., Lin, C., & Shilatifard, A. (2012). The super elongation complex (SEC) family in transcriptional control. *Nat Rev Mol Cell Biol*, *13*(9), 543-547. doi:10.1038/nrm3417
- Luse, D. S. (2013). Promoter clearance by RNA polymerase II. *Biochim Biophys Acta*, *1829*(1), 63-68. doi:10.1016/j.bbagr.2012.08.010
- Madrona, A. Y., & Wilson, D. K. (2004). The structure of Ski8p, a protein regulating mRNA degradation: Implications for WD protein structure. *Protein Sci*, *13*(6), 1557-1565. doi:10.1110/ps.04704704
- Martinez-Rucobo, F. W., & Cramer, P. (2013). Structural basis of transcription elongation. *Biochim Biophys Acta*, *1829*(1), 9-19. doi:10.1016/j.bbagr.2012.09.002
- Martinez-Rucobo, F. W., Kohler, R., van de Waterbeemd, M., Heck, A. J., Hemann, M., Herzog, F., . . . Cramer, P. (2015). Molecular Basis of Transcription-Coupled Pre-mRNA Capping. *Mol Cell*, *58*(6), 1079-1089. doi:10.1016/j.molcel.2015.04.004
- Martinez-Rucobo, F. W., Sainsbury, S., Cheung, A. C., & Cramer, P. (2011). Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO J*, *30*(7), 1302-1310. doi:10.1038/emboj.2011.64
- Matzke, M., Kanno, T., Daxinger, L., Huettel, B., & Matzke, A. J. (2009). RNA-mediated chromatin-based silencing in plants. *Curr Opin Cell Biol*, *21*(3), 367-376. doi:10.1016/j.ceb.2009.01.025
- Mayekar, M. K., Gardner, R. G., & Arndt, K. M. (2013). The recruitment of the *Saccharomyces cerevisiae* Paf1 complex to active genes requires a domain of Rtf1 that directly interacts with the Spt4-Spt5 complex. *Mol Cell Biol*, *33*(16), 3259-3273. doi:10.1128/MCB.00270-13

- Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Soding, J., & Cramer, P. (2010). Uniform transitions of the general RNA polymerase II transcription complex. *Nat Struct Mol Biol*, *17*(10), 1272-1278. doi:10.1038/nsmb.1903
- Mbogning, J., Nagy, S., Page, V., Schwer, B., Shuman, S., Fisher, R. P., & Tanny, J. C. (2013). The PAF complex and Prf1/Rtf1 delineate distinct Cdk9-dependent pathways regulating transcription elongation in fission yeast. *PLoS Genet*, *9*(12), e1004029. doi:10.1371/journal.pgen.1004029
- Meier, A., & Soding, J. (2015). Automatic Prediction of Protein 3D Structures by Probabilistic Multi-template Homology Modeling. *PLoS Comput Biol*, *11*(10), e1004343. doi:10.1371/journal.pcbi.1004343
- Miller, C., Schwalb, B., Maier, K., Schulz, D., Dumcke, S., Zacher, B., . . . Cramer, P. (2011). Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Mol Syst Biol*, *7*, 458. doi:10.1038/msb.2010.112
- Mindell, J. A., & Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol*, *142*(3), 334-347.
- Mueller, C. L., & Jaehning, J. A. (2002). Ctr9, Rtf1, and Leo1 are components of the Paf1/RNA polymerase II complex. *Mol Cell Biol*, *22*(7), 1971-1980.
- Natori, S., Takeuchi, K., Takahashi, K., & Mizuno, D. (1973). DNA dependent RNA polymerase from Ehrlich ascites tumor cells. II. Factors stimulating the activity of RNA polymerase II. *J Biochem*, *73*(4), 879-888.
- Ng, H. H., Dole, S., & Struhl, K. (2003). The Rtf1 component of the Paf1 transcriptional elongation complex is required for ubiquitination of histone H2B. *J Biol Chem*, *278*(36), 33625-33628. doi:10.1074/jbc.C300270200
- Pal, M., Ponticelli, A. S., & Luse, D. S. (2005). The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. *Mol Cell*, *19*(1), 101-110. doi:10.1016/j.molcel.2005.05.024
- Penheiter, K. L., Washburn, T. M., Porter, S. E., Hoffman, M. G., & Jaehning, J. A. (2005). A posttranscriptional role for the yeast Paf1-RNA polymerase II complex is revealed by identification of primary targets. *Mol Cell*, *20*(2), 213-223. doi:10.1016/j.molcel.2005.08.023
- Petes, S. J., & Lis, J. T. (2012). Overcoming the nucleosome barrier during transcript elongation. *Trends Genet*, *28*(6), 285-294. doi:10.1016/j.tig.2012.02.005
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*, *25*(13), 1605-1612. doi:10.1002/jcc.20084
- Plaschka, C., Hantsche, M., Dienemann, C., Burzinski, C., Plitzko, J., & Cramer, P. (2016). Transcription initiation complex structures elucidate DNA opening. *Nature*, *533*(7603), 353-358. doi:10.1038/nature17990
- Plaschka, C., Lariviere, L., Wenzek, L., Seizl, M., Hemann, M., Tegunov, D., . . . Cramer, P. (2015). Architecture of the RNA polymerase II-Mediator core initiation complex. *Nature*, *518*(7539), 376-380. doi:10.1038/nature14229
- Pokholok, D. K., Hannett, N. M., & Young, R. A. (2002). Exchange of RNA polymerase II initiation and elongation factors during gene expression in vivo. *Mol Cell*, *9*(4), 799-809.
- Poli, J., Gerhold, C. B., Tosi, A., Hustedt, N., Seeber, A., Sack, R., . . . Gasser, S. M. (2016). Mec1, INO80, and the PAF1 complex cooperate to limit transcription replication

- conflicts through RNAPII removal during replication stress. *Genes Dev*, 30(3), 337-354. doi:10.1101/gad.273813.115
- Proudfoot, N. J., Furger, A., & Dye, M. J. (2002). Integrating mRNA processing with transcription. *Cell*, 108(4), 501-512.
- Qiu, H., Hu, C., Gaur, N. A., & Hinnebusch, A. G. (2012). Pol II CTD kinases Bur1 and Kin28 promote Spt5 CTR-independent recruitment of Paf1 complex. *EMBO J*, 31(16), 3494-3505. doi:10.1038/emboj.2012.188
- Qiu, H., Hu, C., Wong, C. M., & Hinnebusch, A. G. (2006). The Spt4p subunit of yeast DSIF stimulates association of the Paf1 complex with elongating RNA polymerase II. *Mol Cell Biol*, 26(8), 3135-3148. doi:10.1128/MCB.26.8.3135-3148.2006
- Rauhut, R., Fabrizio, P., Dybkov, O., Hartmuth, K., Pena, V., Chari, A., . . . Luhrmann, R. (2016). Molecular architecture of the *Saccharomyces cerevisiae* activated spliceosome. *Science*, 353(6306), 1399-1405. doi:10.1126/science.aag1906
- Ream, T. S., Haag, J. R., Wierzbicki, A. T., Nicora, C. D., Norbeck, A. D., Zhu, J. K., . . . Pikaard, C. S. (2009). Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II. *Mol Cell*, 33(2), 192-203. doi:10.1016/j.molcel.2008.12.015
- Remmert, M., Biegert, A., Hauser, A., & Soding, J. (2011). HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods*, 9(2), 173-175. doi:10.1038/nmeth.1818
- Remmert, M., Biegert, A., Hauser, A., & Soding, J. (2012). HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods*, 9(2), 173-175. doi:10.1038/nmeth.1818
- Robinson, P. J., Trnka, M. J., Bushnell, D. A., Davis, R. E., Mattei, P. J., Burlingame, A. L., & Kornberg, R. D. (2016). Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. *Cell*, 166(6), 1411-1422 e1416. doi:10.1016/j.cell.2016.08.050
- Rohou, A., & Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol*, 192(2), 216-221. doi:10.1016/j.jsb.2015.08.008
- Rondon, A. G., Gallardo, M., Garcia-Rubio, M., & Aguilera, A. (2004). Molecular evidence indicating that the yeast PAF complex is required for transcription elongation. *EMBO Rep*, 5(1), 47-53. doi:10.1038/sj.embor.7400045
- Rozenblatt-Rosen, O., Nagaike, T., Francis, J. M., Kaneko, S., Glatt, K. A., Hughes, C. M., . . . Meyerson, M. (2009). The tumor suppressor Cdc73 functionally associates with CPSF and CstF 3' mRNA processing factors. *Proc Natl Acad Sci U S A*, 106(3), 755-760. doi:10.1073/pnas.0812023106
- Sainsbury, S., Bernecky, C., & Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. *Nat Rev Mol Cell Biol*, 16(3), 129-143. doi:10.1038/nrm3952
- Sainsbury, S., Niesser, J., & Cramer, P. (2013). Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature*, 493(7432), 437-440. doi:10.1038/nature11715
- Scheres, S. H. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol*, 180(3), 519-530. doi:10.1016/j.jsb.2012.09.006
- Scheres, S. H. (2014). Beam-induced motion correction for sub-megadalton cryo-EM particles. *Elife*, 3, e03665. doi:10.7554/eLife.03665
- Scheres, S. H. (2015). Semi-automated selection of cryo-EM particles in RELION-1.3. *J Struct Biol*, 189(2), 114-122. doi:10.1016/j.jsb.2014.11.010

- Schulz, D., Schwalb, B., Kiesel, A., Baejen, C., Torkler, P., Gagneur, J., . . . Cramer, P. (2013). Transcriptome surveillance by selective termination of noncoding RNA synthesis. *Cell*, *155*(5), 1075-1087. doi:10.1016/j.cell.2013.10.024
- Schwalb, B., Michel, M., Zacher, B., Fruhauf, K., Demel, C., Tresch, A., . . . Cramer, P. (2016). TT-seq maps the human transient transcriptome. *Science*, *352*(6290), 1225-1228. doi:10.1126/science.aad9841
- Sheldon, K. E., Mager, D. M., & Arndt, K. M. (2005). A Requirement for the *Saccharomyces cerevisiae* Paf1 complex in snoRNA 3' end formation. *Mol Cell*, *20*(2), 225-236. doi:10.1016/j.molcel.2005.08.026
- Shi, X., Chang, M., Wolf, A. J., Chang, C. H., Frazer-Abel, A. A., Wade, P. A., . . . Jaehning, J. A. (1997). Cdc73p and Paf1p are found in a novel RNA polymerase II-containing complex distinct from the Srbp-containing holoenzyme. *Mol Cell Biol*, *17*(3), 1160-1169.
- Shi, X., Finkelstein, A., Wolf, A. J., Wade, P. A., Burton, Z. F., & Jaehning, J. A. (1996). Paf1p, an RNA polymerase II-associated factor in *Saccharomyces cerevisiae*, may have both positive and negative roles in transcription. *Mol Cell Biol*, *16*(2), 669-676.
- Shi, X., Kachirskaia, I., Walter, K. L., Kuo, J. H., Lake, A., Davrazou, F., . . . Gozani, O. (2007). Proteome-wide analysis in *Saccharomyces cerevisiae* identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36. *J Biol Chem*, *282*(4), 2450-2455. doi:10.1074/jbc.C600286200
- Shimoaraiso, M., Nakanishi, T., Kubo, T., & Natori, S. (1997). Identification of the region in yeast S-II that defines species specificity in its interaction with RNA polymerase II. *J Biol Chem*, *272*(42), 26550-26554.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., . . . Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*, *7*, 539. doi:10.1038/msb.2011.75
- Sikorski, T. W., & Buratowski, S. (2009). The basal initiation machinery: beyond the general transcription factors. *Curr Opin Cell Biol*, *21*(3), 344-351. doi:10.1016/j.ceb.2009.03.006
- Simic, R., Lindstrom, D. L., Tran, H. G., Roinick, K. L., Costa, P. J., Johnson, A. D., . . . Arndt, K. M. (2003). Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes. *EMBO J*, *22*(8), 1846-1856. doi:10.1093/emboj/cdg179
- Soding, J. (2005). Protein homology detection by HMM-HMM comparison. *Bioinformatics*, *21*(7), 951-960. doi:10.1093/bioinformatics/bti125
- Soding, J., Biegert, A., & Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res*, *33*(Web Server issue), W244-248. doi:10.1093/nar/gki408
- Squazzo, S. L., Costa, P. J., Lindstrom, D. L., Kumer, K. E., Simic, R., Jennings, J. L., . . . Hartzog, G. A. (2002). The Paf1 complex physically and functionally associates with transcription elongation factors in vivo. *EMBO J*, *21*(7), 1764-1774. doi:10.1093/emboj/21.7.1764
- Stark, H. (2010). GraFix: stabilization of fragile macromolecular complexes for single particle cryo-EM. *Methods Enzymol*, *481*, 109-126. doi:10.1016/S0076-6879(10)81005-5
- Steward, M. M., Lee, J. S., O'Donovan, A., Wyatt, M., Bernstein, B. E., & Shilatifard, A. (2006). Molecular regulation of H3K4 trimethylation by ASH2L, a shared subunit of MLL complexes. *Nat Struct Mol Biol*, *13*(9), 852-854. doi:10.1038/nsmb1131

- Sun, M., Schwalb, B., Schulz, D., Pirkl, N., Etzold, S., Lariviere, L., . . . Cramer, P. (2012). Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. *Genome Res*, 22(7), 1350-1359. doi:10.1101/gr.130161.111
- Svejstrup, J. Q. (2004). The RNA polymerase II transcription cycle: cycling through chromatin. *Biochim Biophys Acta*, 1677(1-3), 64-73. doi:10.1016/j.bbaexp.2003.10.012
- Sydow, J. F., Brueckner, F., Cheung, A. C., Damsma, G. E., Dengl, S., Lehmann, E., . . . Cramer, P. (2009). Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. *Mol Cell*, 34(6), 710-721. doi:10.1016/j.molcel.2009.06.002
- Takahashi, A., Tsutsumi, R., Kikuchi, I., Obuse, C., Saito, Y., Seidi, A., . . . Hatakeyama, M. (2011). SHP2 tyrosine phosphatase converts parafibromin/Cdc73 from a tumor suppressor to an oncogenic driver. *Mol Cell*, 43(1), 45-56. doi:10.1016/j.molcel.2011.05.014
- Tan, J., Muntean, A. G., & Hess, J. L. (2010). PAFc, a key player in MLL-rearranged leukemogenesis. *Oncotarget*, 1(6), 461-465. doi:10.18632/oncotarget.181
- Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I., & Ludtke, S. J. (2007). EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol*, 157(1), 38-46. doi:10.1016/j.jsb.2006.05.009
- Tenney, K., Gerber, M., Ilvarsonn, A., Schneider, J., Gause, M., Dorsett, D., . . . Shilatifard, A. (2006). Drosophila Rtf1 functions in histone methylation, gene expression, and Notch signaling. *Proc Natl Acad Sci U S A*, 103(32), 11970-11974. doi:10.1073/pnas.0603620103
- Tessarz, P., & Kouzarides, T. (2014). Histone core modifications regulating nucleosome structure and dynamics. *Nat Rev Mol Cell Biol*, 15(11), 703-708. doi:10.1038/nrm3890
- Tomson, B. N., & Arndt, K. M. (2013). The many roles of the conserved eukaryotic Paf1 complex in regulating transcription, histone modifications, and disease states. *Biochim Biophys Acta*, 1829(1), 116-126. doi:10.1016/j.bbagr.2012.08.011
- Vannini, A., & Cramer, P. (2012). Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol Cell*, 45(4), 439-446. doi:10.1016/j.molcel.2012.01.023
- Verrier, L., Tagliani, F., Barrales, R. R., Webb, S., Urano, T., Braun, S., & Bayne, E. H. (2015). Global regulation of heterochromatin spreading by Leo1. *Open Biol*, 5(5). doi:10.1098/rsob.150045
- Wade, P. A., Werel, W., Fentzke, R. C., Thompson, N. E., Leykam, J. F., Burgess, R. R., . . . Burton, Z. F. (1996). A novel collection of accessory factors associated with yeast RNA polymerase II. *Protein Expr Purif*, 8(1), 85-90. doi:10.1006/prev.1996.0077
- Warner, M. H., Roinick, K. L., & Arndt, K. M. (2007). Rtf1 is a multifunctional component of the Paf1 complex that regulates gene expression by directing cotranscriptional histone modification. *Mol Cell Biol*, 27(17), 6103-6115. doi:10.1128/MCB.00772-07
- Weinzierl, R. O. (2013). The RNA polymerase factory and archaeal transcription. *Chem Rev*, 113(11), 8350-8376. doi:10.1021/cr400148k
- Wier, A. D., Mayekar, M. K., Heroux, A., Arndt, K. M., & VanDemark, A. P. (2013). Structural basis for Spt5-mediated recruitment of the Paf1 complex to chromatin. *Proc Natl Acad Sci U S A*, 110(43), 17290-17295. doi:10.1073/pnas.1314754110

- Wilmes, G. M., Bergkessel, M., Bandyopadhyay, S., Shales, M., Braberg, H., Cagney, G., . . . Krogan, N. J. (2008). A genetic interaction map of RNA-processing factors reveals links between Sem1/Dss1-containing complexes and mRNA export and splicing. *Mol Cell*, 32(5), 735-746. doi:10.1016/j.molcel.2008.11.012
- Wood, A., Schneider, J., Dover, J., Johnston, M., & Shilatifard, A. (2003). The Paf1 complex is essential for histone monoubiquitination by the Rad6-Bre1 complex, which signals for histone methylation by COMPASS and Dot1p. *J Biol Chem*, 278(37), 34739-34742. doi:10.1074/jbc.C300269200
- Wu, C. H., Yamaguchi, Y., Benjamin, L. R., Horvat-Gordon, M., Washinsky, J., Enerly, E., . . . Gilmour, D. (2003). NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in *Drosophila*. *Genes Dev*, 17(11), 1402-1414. doi:10.1101/gad.1091403
- Wu, J., & Xu, W. (2012). Histone H3R17me2a mark recruits human RNA polymerase-associated factor 1 complex to activate transcription. *Proc Natl Acad Sci U S A*, 109(15), 5675-5680. doi:10.1073/pnas.1114905109
- Xiao, T., Kao, C. F., Krogan, N. J., Sun, Z. W., Greenblatt, J. F., Osley, M. A., & Strahl, B. D. (2005). Histone H2B ubiquitylation is associated with elongating RNA polymerase II. *Mol Cell Biol*, 25(2), 637-651. doi:10.1128/MCB.25.2.637-651.2005
- Yamaguchi, Y., Shibata, H., & Handa, H. (2013). Transcription elongation factors DSIF and NELF: promoter-proximal pausing and beyond. *Biochim Biophys Acta*, 1829(1), 98-104. doi:10.1016/j.bbagr.2012.11.007
- Yang, B., Wu, Y. J., Zhu, M., Fan, S. B., Lin, J., Zhang, K., . . . Dong, M. Q. (2012). Identification of cross-linked peptides from complex samples. *Nat Methods*, 9(9), 904-906. doi:10.1038/nmeth.2099
- Yu, M., Yang, W., Ni, T., Tang, Z., Nakadai, T., Zhu, J., & Roeder, R. G. (2015). RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. *Science*, 350(6266), 1383-1386. doi:10.1126/science.aad2338
- Zawel, L., Kumar, K. P., & Reinberg, D. (1995). Recycling of the general transcription factors during RNA polymerase II transcription. *Genes Dev*, 9(12), 1479-1490.
- Zhu, B., Mandal, S. S., Pham, A. D., Zheng, Y., Erdjument-Bromage, H., Batra, S. K., . . . Reinberg, D. (2005). The human PAF complex coordinates transcription with events downstream of RNA synthesis. *Genes Dev*, 19(14), 1668-1673. doi:10.1101/gad.1292105

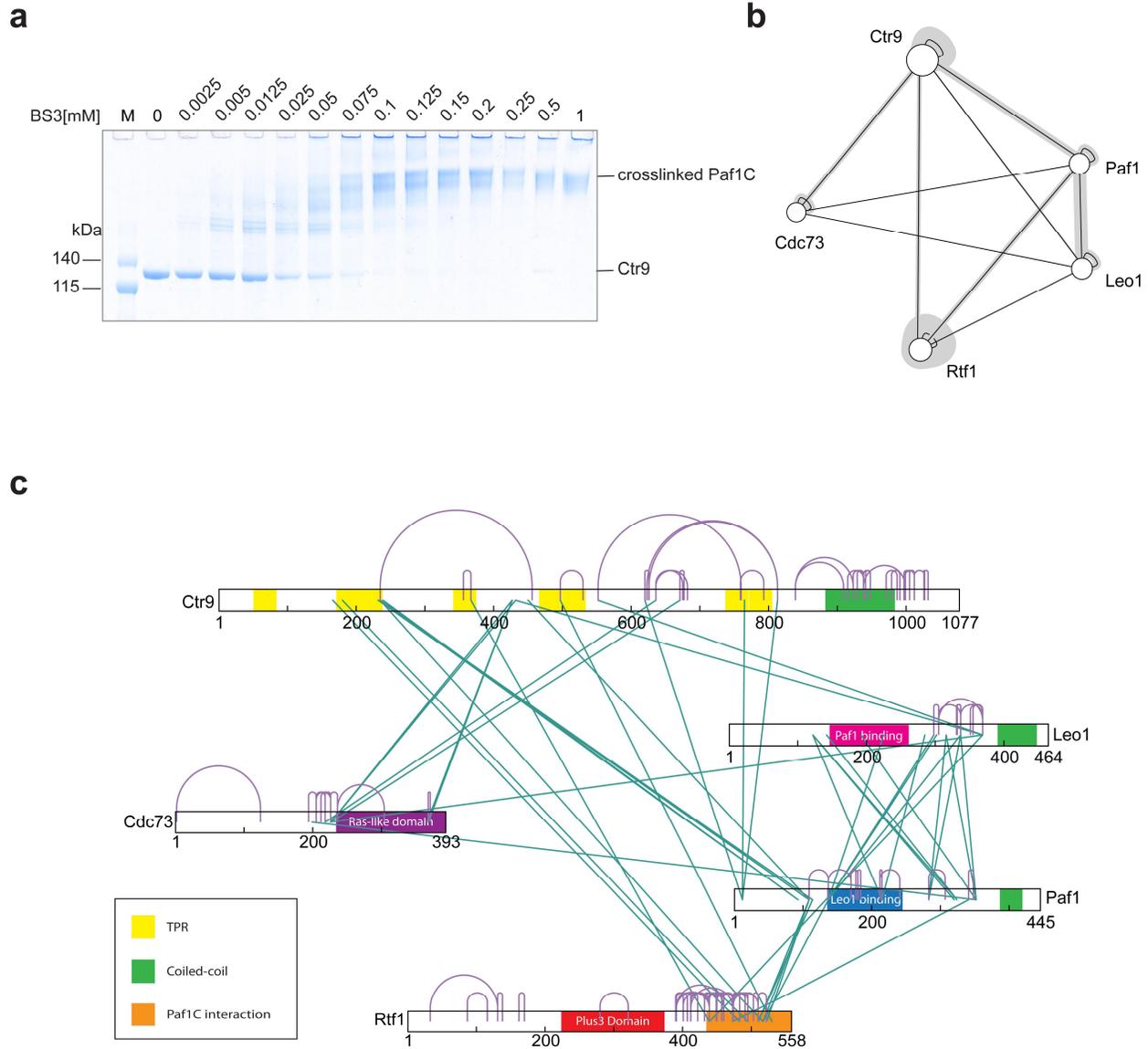
Appendix

Supplementary Material 1

Transcription elongation assay

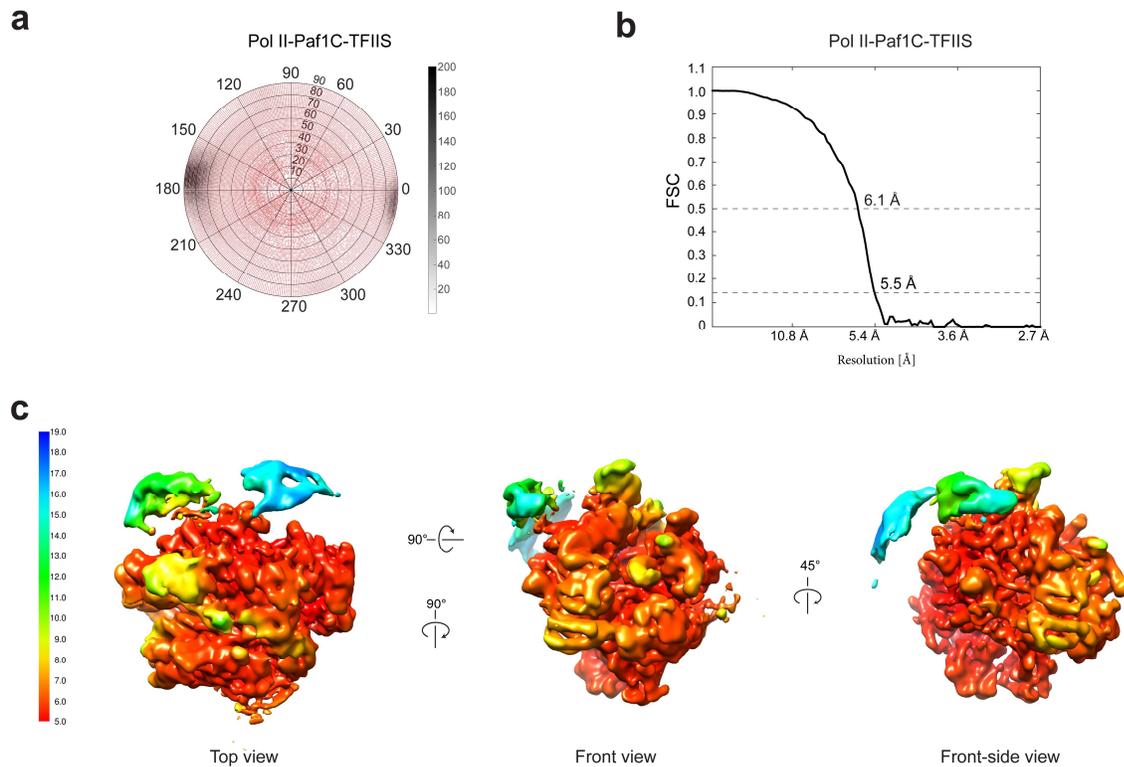
To analyze whether Paf1C influences the Pol II transcription, the in vitro transcription elongation assay was performed. Two types of tailed template were used in this study (Table 8). One was designed based on *His4* gene (Table 8, RNA: FAM-RNA-20, template DNA: TA-1-tempalte DNA, non-template DNA: TA-1-nontemplate DNA), and we refer it to 'elongation template'. The other one template used was GK-TA-pML5 template DNA buried AAAAAAA sequence, which could induce TFIIS cleavage activity (Table 8, RNA: FAM-GK-RNA-20, template DNA: GK-TA-pML5 template DNA, non-template DNA: GK-TA-pML5 nontemplate DNA), and we refer this to 'backtracked template'. Neither Paf1C nor Paf1C together with TFIIS alter the Pol II transcription using both templates, even Paf1C was at high concentration (Supplementary Figure 1).

Supplementary Figure 2



Supplementary Figure 2: Crosslinking analysis of Paf1C. **a**, Coomassie-stained SDS-PAGE analysis of BS3 cross-linker titration. A fixed amount of purified Paf1 complex was mixed with increasing amounts of BS3 cross-linker. Gel was cropped to the largest subunit Ctr9. Successfully cross-linked Paf1 complex migrates at a higher molecular weight. 0.1 mM BS3 was chosen for final sample preparation. **b**, Model of interactions within Paf1C. The circles represent the size of subunits. The solid lines show the interaction obtained from XL-MS results. The grey shading indicates the numbers of observed cross-links. **c**, Network diagrams of unique cross-links of Paf1 complex. Observed inter-links and intra-links between are colored in dark cyan and purple, respectively.

Supplementary Figure 3



Supplementary Figure 3: Cryo-EM reconstitution of Pol II-Paf1C-TFIIS. **a**, Angular distribution of Pol II-Paf1C-TFIIS Cryo-EM single particle reconstruction. Red dots indicate at least one particle was assigned within 1° of the point. Black shading indicates the number of particles assigned to a given view. **b**, The overall resolution is estimated to be 5.5 Å on the basis of the gold-standard Fourier Shell Correlation (FSC) criteria of 0.143. **c**, Top, front and front-side view of unsharpened Pol II-Paf1C-TFIIS reconstruction colored by local resolution.

Supplementary Table 1

Supplementary Table 1: Summary of various recombined Paf1 complexes.

Name \ Subunit	Ctr9 (1-1077 aa)	Leo1 (1-464 aa)	Paf1 (1-445 aa)	Cdc73 (1-393 aa)	Rtf1 (1-558 aa)
Paf1C	1-1077 aa	1-464 aa	1-445 aa	1-393 aa	1-558 aa
Paf1C-Ctr9-ΔC913	1-913 aa	1-464 aa	1-445 aa	1-393 aa	1-558 aa
Paf1C-Ctr9-ΔC913- ΔRtf1	1-913 aa	1-464 aa	1-445 aa	1-393 aa	-
core Paf1C- Rtf1ΔN441	1-913 aa	118-376 aa	1-360 aa	1-393 aa	442-558 aa
core Paf1C	1-913 aa	118-376 aa	1-360 aa	1-393 aa	-

Supplementary Table 2

Supplementary Table 2: High-confidence crosslinks of Paf1C.

Link type	Protein a	Protein b	Residue protein a	Residue protein b	Max of Score
Inter	Ctr9	Rtf1	166	483	9.690369833
Inter	Ctr9	Rtf1	180	495	7.782516056
Inter	Ctr9	Rtf1	235	517	7.4867824
Inter	Ctr9	Rtf1	239	510	7.954677021
Inter	Ctr9	Rtf1	497	530	9.801342913
Inter	Ctr9	Leo1	432	369	10.06803389
Inter	Ctr9	Leo1	552	369	9.774690718
Inter	Ctr9	Paf1	235	93	11.24488773
Inter	Ctr9	Paf1	239	93	9.879426069
Inter	Ctr9	Paf1	232	109	22.06098022
Inter	Ctr9	Paf1	232	114	25.02227639
Inter	Ctr9	Paf1	235	109	5.559090918
Inter	Ctr9	Paf1	450	109	8.03574037
Inter	Ctr9	Paf1	620	12	12.11918641
Inter	Ctr9	Paf1	626	12	5.725842151
Inter	Ctr9	Paf1	765	12	12.79048499
Inter	Ctr9	Paf1	813	12	8.074687909
Inter	Ctr9	Cdc73	427	226	9.076238039
Inter	Ctr9	Cdc73	427	366	5.029188389
Inter	Ctr9	Cdc73	432	226	14.27490548
Inter	Ctr9	Cdc73	636	218	5.272458743
Inter	Ctr9	Cdc73	671	226	8.798602876
Inter	Rtf1	Leo1	439	369	6.754487332
Inter	Rtf1	Paf1	473	341	13.79048499
Inter	Rtf1	Paf1	510	109	10.80134291
Inter	Rtf1	Paf1	510	114	10.57186521
Inter	Rtf1	Paf1	517	109	17.85387196
Inter	Rtf1	Paf1	517	137	11.56066731
Inter	Rtf1	Paf1	522	109	8.761953897
Inter	Leo1	Paf1	122	215	18.60032628
Inter	Leo1	Paf1	122	241	12.25806092
Inter	Leo1	Paf1	122	320	12.74232143
Inter	Leo1	Paf1	122	324	23.45593196
Inter	Leo1	Paf1	142	324	13.34678749
Inter	Leo1	Paf1	192	352	22.35359627
Inter	Leo1	Paf1	225	137	6.448550002

Link type	Protein a	Protein b	Residue protein a	Residue protein b	Max of Score
Inter	Leo1	Paf1	285	215	20.83564714
Inter	Leo1	Paf1	298	137	6.014573526
Inter	Leo1	Paf1	316	283	15.40560745
Inter	Leo1	Paf1	316	352	19.19382003
Inter	Leo1	Paf1	337	283	13.8068754
Inter	Leo1	Paf1	337	352	7.879426069
Inter	Leo1	Cdc73	369	226	16.1700533
Inter	Paf1	Cdc73	352	199	13.67778071
Intra	Ctr9	Ctr9	356	367	14.82390874
Intra	Ctr9	Ctr9	456	235	7.91721463
Intra	Ctr9	Ctr9	497	530	20.02181948
Intra	Ctr9	Ctr9	552	760	15.31785492
Intra	Ctr9	Ctr9	620	626	5.543633967
Intra	Ctr9	Ctr9	671	677	6.170053304
Intra	Ctr9	Ctr9	682	636	5.218244625
Intra	Ctr9	Ctr9	793	759	15.97469413
Intra	Ctr9	Ctr9	813	626	7.184422252
Intra	Ctr9	Ctr9	909	839	6.8569852
Intra	Ctr9	Ctr9	923	938	5.718966633
Intra	Ctr9	Ctr9	929	938	5.555955204
Intra	Ctr9	Ctr9	941	948	15.73282827
Intra	Ctr9	Ctr9	999	938	6.350665141
Intra	Ctr9	Ctr9	1005	997	5.752026734
Intra	Ctr9	Ctr9	1012	999	6.104025268
Intra	Ctr9	Ctr9	1012	1005	8.040005162
Intra	Ctr9	Ctr9	1012	1027	9.653647026
Intra	Rtf1	Rtf1	33	131	11.87942607
Intra	Rtf1	Rtf1	87	116	10.72815839
Intra	Rtf1	Rtf1	131	138	12.89619628
Intra	Rtf1	Rtf1	138	131	6.829738285
Intra	Rtf1	Rtf1	162	170	5.496209317
Intra	Rtf1	Rtf1	391	407	9.349692477
Intra	Rtf1	Rtf1	391	460	11.95860731
Intra	Rtf1	Rtf1	401	390	10.36552273
Intra	Rtf1	Rtf1	401	412	17.00524306
Intra	Rtf1	Rtf1	412	401	12.79588002
Intra	Rtf1	Rtf1	412	432	9.129011186
Intra	Rtf1	Rtf1	422	407	8.258848401
Intra	Rtf1	Rtf1	439	391	13.63264408

Link type	Protein a	Protein b	Residue protein a	Residue protein b	Max of Score
Intra	Rtf1	Rtf1	439	401	16.78515615
Intra	Rtf1	Rtf1	439	412	5.872895202
Intra	Rtf1	Rtf1	439	450	13.32148162
Intra	Rtf1	Rtf1	439	457	6.485452247
Intra	Rtf1	Rtf1	439	460	10.38510278
Intra	Rtf1	Rtf1	449	432	6.943095149
Intra	Rtf1	Rtf1	449	457	5.298432015
Intra	Rtf1	Rtf1	449	460	20.13846559
Intra	Rtf1	Rtf1	449	474	13.5654311
Intra	Rtf1	Rtf1	457	460	15.37882372
Intra	Rtf1	Rtf1	457	473	10.18442225
Intra	Rtf1	Rtf1	460	450	9.596879479
Intra	Rtf1	Rtf1	460	473	11.47886192
Intra	Rtf1	Rtf1	474	450	12.39147397
Intra	Rtf1	Rtf1	474	460	19.86646109
Intra	Rtf1	Rtf1	483	473	9.91721463
Intra	Rtf1	Rtf1	495	473	5.002613616
Intra	Rtf1	Rtf1	503	510	13.73518218
Intra	Rtf1	Rtf1	510	473	10.27572413
Intra	Rtf1	Rtf1	510	522	5.823908741
Intra	Leo1	Leo1	298	305	12.51855737
Intra	Leo1	Leo1	331	337	12.32148162
Intra	Leo1	Leo1	337	305	10.21824463
Intra	Leo1	Leo1	337	332	5.578396073
Intra	Leo1	Leo1	355	369	10.97469413
Intra	Leo1	Leo1	369	305	6.709965389
Intra	Leo1	Leo1	369	337	10.1739252
Intra	Paf1	Paf1	137	109	6.850780887
Intra	Paf1	Paf1	137	170	9.463441557
Intra	Paf1	Paf1	170	180	7.677780705
Intra	Paf1	Paf1	241	212	5.614393726
Intra	Cdc73	Cdc73	124	2	6.026872146
Intra	Cdc73	Cdc73	194	205	21.03011836
Intra	Cdc73	Cdc73	226	205	15.24336389
Intra	Cdc73	Cdc73	226	218	7.540607512
Intra	Cdc73	Cdc73	236	226	15.61978876
Intra	Cdc73	Cdc73	236	304	14.31425826
Intra	Cdc73	Cdc73	368	371	11.7212464

Supplementary Table 3

Supplementary Table 3: High-confidence crosslinks of Pol II-Paf1C-TFIIS complex.

Type of crosslinking	Protein a	Protein b	Residue protein a	Residue protein b	Count of Spectrum 1	Max of Score 1	Count of Spectrum 2	Max of Score 2	Distance (Å)/Domains
Pol II-Pol II	Rpb1	Rpb11	368	37	6	11.54060751	4	11.17198494	20.2
Pol II-Pol II	Rpb1	Rpb11	644	26	53	12.15428198	51	13.92445304	18.5
Pol II-Pol II	Rpb1	Rpb11	644	62	4	14.31785492	4	15.97061622	18.4
Pol II-Pol II	Rpb1	Rpb2	2	886	2	7.785156152	2	11.59006688	30.9
Pol II-Pol II	Rpb1	Rpb2	2	934	4	13.47366072	4	15.01818139	39.2
Pol II-Pol II	Rpb1	Rpb2	34	1183	3	5.534617149	5	5.772113295	14.8
Pol II-Pol II	Rpb1	Rpb2	323	934	4	7.946921557	3	8.744727495	34.8
Pol II-Pol II	Rpb1	Rpb2	1093	227	7	16.92811799	7	16.86327943	20.9
Pol II-Pol II	Rpb1	Rpb2	1093	228	3	9.876148359	2	13.1260984	21
Pol II-Pol II	Rpb1	Rpb2	1093	507	6	14.77469072	5	15.03715732	21.2
Pol II-Pol II	Rpb1	Rpb2	1102	507	21	11.60730305	19	12.77989191	17.6
Pol II-Pol II	Rpb1	Rpb2	1112	507	7	12.84466396	11	13.01412464	18.9
Pol II-Pol II	Rpb1	Rpb2	1262	270	6	14.23210238	7	15.22257318	21.8
Pol II-Pol II	Rpb1	Rpb4	15	1 (unstructured)	4	19.73754891	3	16.55129368	
Pol II-Pol II	Rpb1	Rpb4	129	1 (unstructured)	4	14.00921731	5	12.79317412	
Pol II-Pol II	Rpb1	Rpb4	129	2 (unstructured)	2	13.08039898	2	12.77989191	
Pol II-Pol II	Rpb1	Rpb5	15	171	2	11.69897	4	10.82390874	20.6
Pol II-Pol II	Rpb1	Rpb5	129	161	5	8.066006836	7	7.958607315	21.8
Pol II-Pol II	Rpb1	Rpb5	129	171	28	12.75448733	26	12.26841123	15
Pol II-Pol II	Rpb1	Rpb5	938	152	3	5.987162775	6	7.288192771	19.8
Pol II-Pol II	Rpb1	Rpb5	938	201	3	6.448550002	5	9.057495894	15.5
Pol II-Pol II	Rpb1	Rpb5	941	152	4	7.353596274	6	7.583359493	18.3
Pol II-Pol II	Rpb1	Rpb5	941	201	7	8.07007044	7	6.642065153	13.3

Type of crosslinking	Protein a	Protein b	Residue protein a	Residue protein b	Count of Spectrum 1	Max of Score 1	Count of Spectrum 2	Max of Score 2	Distance (Å)/Domains
Pol II-Pol II	Rpb1	Rpb5	1003	197	4	10.89962945	4	11.87614836	20.3
Pol II-Pol II	Rpb1	Rpb6	15	67 (unstructured)	3	3.798602876	4	6.187755303	
Pol II-Pol II	Rpb1	Rpb6	15	72	15	8.052566278	14	8.560667306	17.8
Pol II-Pol II	Rpb1	Rpb6	129	67 (unstructured)	7	7.835647144	6	8.184422252	
Pol II-Pol II	Rpb1	Rpb6	1003	72	2	9.07211659	2	10.26280736	19.4
Pol II-Pol II	Rpb1	Rpb6	1003	76	3	8.288192771	3	7.232844134	13.9
Pol II-Pol II	Rpb1	Rpb8	977	136	3	7.844663963	2	8.522878745	13.3
Pol II-Pol II	Rpb1	Rpb9	1093	93	7	16.31069114	6	14.32422166	16.2
Pol II-Pol II	Rpb2	Rpb10	191	68 (unstructured)	4	6.698970004	4	7.735182177	
Pol II-Pol II	Rpb2	Rpb10	813	59	12	9.91721463	8	13.51427857	14
Pol II-Pol II	Rpb2	Rpb3	191	149	3	7.111259039	4	8.89279003	13.5
Pol II-Pol II	Rpb2	Rpb9	227	93	8	11.03668449	3	8.305394801	16.2
Pol II-Pol II	Rpb3	Rpb10	149	68 (unstructured)	4	7.73754891	5	6.505845406	
Pol II-Pol II	Rpb3	Rpb11	253	37	2	8.075720714	2	9.321481621	14.5
Pol II-Pol II	Rpb3	Rpb12	149	37	4	12.79048499	3	12.24260397	15.2
Pol II-Pol II	Rpb4	Rpb5	1 (unstructured)	171	3	15.4723701	4	14.84466396	
Pol II-Pol II	Rpb4	Rpb5	2 (unstructured)	171	3	14.81247928	2	14.54060751	
Pol II-Pol II	Rpb4	Rpb6	1 (unstructured)	72	2	7.806875402	2	7.714442691	
Pol II-Pol II	Rpb5	Rpb6	171	67 (unstructured)	9	6.177178355	7	5.123205024	
Pol II-Pol II	Rpb5	Rpb6	171	70 (unstructured)	7	12.23062267	10	11.43415218	
Pol II-Pol II	Rpb5	Rpb6	171	72	16	12.76700389	19	12.83564714	16.1
Pol II-Pol II	Rpb6	Rpb7	67 (unstructured)	73	7	13.86327943	6	7.59345982	
Pol II-Pol II	Rpb6	Rpb7	70	73	9	13.08565684	9	13.16052195	

Type of crosslinking	Protein a	Protein b	Residue protein a	Residue protein b	Count of Spectrum 1	Max of Score 1	Count of Spectrum 2	Max of Score 2	Distance (Å)/Domains
			(unstructured)						
Pol II-Pol II	Rpb6	Rpb7	72	73	24	16.89619628	24	15.8096683	17.6
Pol II-TFIIS	TFIIS	Rpb1	196	1217	12	22.77728353	9	20.98716278	16.4
Pol II-TFIIS	TFIIS	Rpb1	250	705	2	5.300162274	2	6.364516253	13.9
Pol II-TFIIS	TFIIS	Rpb1	273	1290	2	9.077793723	2	8.508638306	15.9
Pol II-TFIIS	TFIIS	Rpb1	278	1093	2	9.735182177	2	14.01954211	25.2
Pol II-TFIIS	TFIIS	Rpb1	278	1290	4	16.87942607	4	18.91721463	19.4
Pol II-Paf1C	Ctr9	Rpb1	443	34	2	5.032452024	3	6.378823718	Clamp_Core
Pol II-Paf1C	Ctr9	Rpb3	749	199	4	19.91009489	4	18.52287875	Rpb3
Pol II-Paf1C	Rtf1	Rpb2	473	344	2	19.8728952	4	9.565431096	Lobe
Pol II-Paf1C	Leo1	Rpb1	369	101	2	11.69464863	2	7.145693958	Clamp_Head
Pol II-Paf1C	Leo1	Rpb1	369	1102	4	7.821023053	5	7.358525889	Cleft
Pol II-Paf1C	Leo1	Rpb2	298	606	4	6.920818754	3	10.23657201	External 2
Pol II-Paf1C	Leo1	Rpb2	302	606	5	7.171984936	8	6.632644079	External 2
Pol II-Paf1C	Leo1	Rpb2	305	246	3	12.00086946	3	10.66354027	Lobe
Pol II-Paf1C	Leo1	Rpb2	305	606	8	9.343901798	5	9.619788758	External 2
Pol II-Paf1C	Leo1	Rpb2	332	87	6	14.59345982	3	8.866461092	Protrusion
Pol II-Paf1C	Leo1	Rpb2	332	353	5	8.785156152	3	7.354577731	Lobe
Pol II-Paf1C	Leo1	Rpb2	337	87	8	13.45222529	9	16.16052195	Protrusion
Pol II-Paf1C	Leo1	Rpb2	337	133	2	12.69680394	2	8.407823243	Protrusion
Pol II-Paf1C	Leo1	Rpb2	337	148	3	9.528708289	4	12.4828041	Protrusion
Pol II-Paf1C	Leo1	Rpb2	337	228	5	13.36552273	6	11.82102305	Lobe
Pol II-Paf1C	Leo1	Rpb2	337	246	4	10.61261017	3	11.82390874	Lobe
Pol II-Paf1C	Leo1	Rpb2	337	270	2	9.93930216	2	10.24412514	Lobe
Pol II-Paf1C	Leo1	Rpb2	337	277	2	7.448550002	5	9.13430394	Lobe
Pol II-Paf1C	Leo1	Rpb2	337	344	4	11.90657831	5	12.32148162	Lobe
Pol II-Paf1C	Leo1	Rpb2	337	353	4	10.4710833	4	8.576754126	Lobe
Pol II-Paf1C	Leo1	Rpb2	337	358	3	9.838631998	3	10.08196966	Lobe

Type of crosslinking	Protein a	Protein b	Residue protein a	Residue protein b	Count of Spectrum 1	Max of Score 1	Count of Spectrum 2	Max of Score 2	Distance (Å)/Domains
Pol II-Paf1C	Leo1	Rpb2	337	422	2	6.838631998	2	8.628932138	Protrusion
Pol II-Paf1C	Leo1	Rpb2	337	426	2	7.696803943	2	7.801342913	Protrusion
Pol II-Paf1C	Leo1	Rpb2	337	445	3	12.25181197	4	12.86327943	Protrusion
Pol II-Paf1C	Leo1	Rpb2	337	451	4	11.03105032	3	11.39685563	Protrusion
Pol II-Paf1C	Leo1	Rpb2	351	87	2	5.943095149	2	5.474955193	Lobe
Pol II-Paf1C	Leo1	Rpb2	351	344	4	5.46470588	5	5.66756154	Lobe
Pol II-Paf1C	Leo1	Rpb2	353	353	3	8.07007044	4	9.903089987	Lobe
Pol II-Paf1C	Leo1	Rpb2	355	228	2	15.46852108	4	17.04383157	Lobe
Pol II-Paf1C	Leo1	Rpb2	355	277	3	18.22402567	3	21.03668449	Lobe
Pol II-Paf1C	Leo1	Rpb2	355	344	5	6.312471039	4	7.764471553	Lobe
Pol II-Paf1C	Leo1	Rpb2	369	87	3	14.8096683	2	18.65364703	Protrusion
Pol II-Paf1C	Leo1	Rpb2	369	228	13	26.33535802	12	24.61083392	Lobe
Pol II-Paf1C	Leo1	Rpb2	369	270	13	18.4100504	13	19.22112553	Lobe
Pol II-Paf1C	Leo1	Rpb2	369	277	9	19.6716204	13	22.58169871	Lobe
Pol II-Paf1C	Leo1	Rpb2	369	344	7	14.86646109	9	13.25727487	Lobe
Pol II-Paf1C	Leo1	Rpb2	369	353	2	7.560667306	2	6.235823868	Lobe
Pol II-Paf1C	Leo1	Rpb2	369	426	7	12.83564714	10	11.95467702	Protrusion
Pol II-Paf1C	Leo1	Rpb2	369	471	3	9.943095149	2	10.0259491	Fork
Pol II-Paf1C	Leo1	Rpb2	369	865	3	23.09582563	2	18.44977165	Wall
Pol II-Paf1C	Paf1	Rpb2	109	87	4	17.20273246	5	13.99567863	Protrusion
Pol II-Paf1C	Paf1	Rpb2	114	451	3	11.19859629	2	9.344861565	Protrusion
Pol II-Paf1C	Paf1	Rpb2	320	353	2	9.681936665	3	7.540607512	Lobe
Pol II-Paf1C	Cdc73	Rpb11	205	55	4	12.87942607	4	11.79860288	Rpb11
Pol II-Paf1C	Cdc73	Rpb11	218	55	10	10.79860288	10	10.9788107	Rpb11
Pol II-Paf1C	Cdc73	Rpb11	236	55	2	6.080398976	2	17.87942607	Rpb11
Pol II-Paf1C	Cdc73	Rpb11	385	55	2	6.126098402	3	6.809668302	Rpb11
Pol II-Paf1C	Cdc73	Rpb2	194	451	3	17.27164622	5	14.58169871	Protrusion
Pol II-Paf1C	Cdc73	Rpb3	263	199	3	15.28399666	3	14.34775366	Rpb3

Type of crosslinking	Protein a	Protein b	Residue protein a	Residue protein b	Count of Spectrum 1	Max of Score 1	Count of Spectrum 2	Max of Score 2	Distance (Å)/Domains
Paf1C-TFIIS	Leo1	TFIIS	331	78	7	7.42945706	6	6.730487056	Domain I
Paf1C-TFIIS	Leo1	TFIIS	331	80	3	7.844663963	3	6.8569852	Domain I
Paf1C-Paf1C	Ctr9	Cdc73	427	226	2	6.027334408	3	9.510041521	
Paf1C-Paf1C	Ctr9	Cdc73	427	263	4	8.211124884	4	9.567030709	
Paf1C-Paf1C	Ctr9	Cdc73	427	385	5	12.21041929	4	12.63451202	
Paf1C-Paf1C	Ctr9	Cdc73	432	194	7	10.40011693	6	13.9625735	
Paf1C-Paf1C	Ctr9	Cdc73	432	282	3	11.82390874	3	11.07987667	
Paf1C-Paf1C	Ctr9	Cdc73	432	385	3	9.749579998	3	8.638272164	
Paf1C-Paf1C	Ctr9	Cdc73	443	263	4	7.500312917	4	7.853871964	
Paf1C-Paf1C	Ctr9	Cdc73	636	212	7	7.336299075	6	10.01099538	
Paf1C-Paf1C	Ctr9	Cdc73	636	218	7	12.30451832	6	10.45842076	
Paf1C-Paf1C	Ctr9	Cdc73	636	226	7	7.410050399	5	11.62342304	
Paf1C-Paf1C	Ctr9	Cdc73	636	385	3	20.66354027	4	17.4202164	
Paf1C-Paf1C	Ctr9	Cdc73	671	205	4	12.46852108	3	10.94692156	
Paf1C-Paf1C	Ctr9	Cdc73	671	212	3	7.294992041	5	6.732828272	
Paf1C-Paf1C	Ctr9	Cdc73	671	218	2	6.779891912	2	8.484126156	
Paf1C-Paf1C	Ctr9	Cdc73	671	385	5	9.991399828	5	9.841637508	
Paf1C-Paf1C	Ctr9	Cdc73	760	185	4	10.86012091	3	10.57839607	
Paf1C-Paf1C	Ctr9	Cdc73	784	385	4	6.549750892	4	5.375717904	
Paf1C-Paf1C	Ctr9	Cdc73	793	371	2	14.75202673	2	10.51286162	
Paf1C-Paf1C	Ctr9	Cdc73	867	385	3	9.718966633	3	10.61083392	
Paf1C-Paf1C	Ctr9	Leo1	443	337	3	9.400116928	2	9.049148541	
Paf1C-Paf1C	Ctr9	Leo1	450	369	2	13.8728952	2	15.60906489	
Paf1C-Paf1C	Ctr9	Leo1	749	305	3	7.612610174	5	8.728158393	
Paf1C-Paf1C	Ctr9	Leo1	759	305	2	10.87614836	2	12.82102305	
Paf1C-Paf1C	Ctr9	Leo1	784	337	2	8.519993057	2	5.623423043	
Paf1C-Paf1C	Ctr9	Leo1	793	302	3	8.59345982	5	7.716698771	
Paf1C-Paf1C	Ctr9	Leo1	839	337	2	8.089909454	2	8.183096161	

Type of crosslinking	Protein a	Protein b	Residue protein a	Residue protein b	Count of Spectrum 1	Max of Score 1	Count of Spectrum 2	Max of Score 2	Distance (Å)/Domains
Paf1C-Paf1C	Ctr9	Paf1	180	81	14	15.47495519	15	14.24795155	
Paf1C-Paf1C	Ctr9	Paf1	180	341	2	5.338187314	2	5.009661145	
Paf1C-Paf1C	Ctr9	Paf1	232	109	21	16.92445304	22	19.82973828	
Paf1C-Paf1C	Ctr9	Paf1	232	114	8	13.07314329	9	13.91721463	
Paf1C-Paf1C	Ctr9	Paf1	232	137	3	5.806875402	2	6.504455662	
Paf1C-Paf1C	Ctr9	Paf1	235	109	3	7.463441557	3	7.728158393	
Paf1C-Paf1C	Ctr9	Paf1	239	93	23	9.066512712	23	9.950781977	
Paf1C-Paf1C	Ctr9	Paf1	239	109	3	12.6716204	3	14.67985371	
Paf1C-Paf1C	Ctr9	Paf1	443	109	2	9.188424994	2	10.37365963	
Paf1C-Paf1C	Ctr9	Paf1	443	137	4	11.60906489	3	11.20342567	
Paf1C-Paf1C	Ctr9	Paf1	450	109	11	18.3419886	12	16.02826041	
Paf1C-Paf1C	Ctr9	Paf1	450	114	4	16.39147397	4	15.81247928	
Paf1C-Paf1C	Ctr9	Paf1	450	137	5	10.76700389	5	15.57348874	
Paf1C-Paf1C	Ctr9	Paf1	456	109	5	10.37263414	6	11.21609642	
Paf1C-Paf1C	Ctr9	Paf1	456	114	4	13.84771166	3	13.22621356	
Paf1C-Paf1C	Ctr9	Paf1	620	12	19	10.38510278	15	9.311580178	
Paf1C-Paf1C	Ctr9	Paf1	626	12	35	7.596879479	25	8.160521953	
Paf1C-Paf1C	Ctr9	Paf1	793	170	6	8.663540266	5	8.978810701	
Paf1C-Paf1C	Ctr9	Paf1	813	12	5	8.085656843	5	11.09799711	
Paf1C-Paf1C	Ctr9	Paf1	839	180	2	5.732828272	3	5.498940738	
Paf1C-Paf1C	Ctr9	Rtf1	156	457	3	9.675717545	3	8.218244625	
Paf1C-Paf1C	Ctr9	Rtf1	156	473	3	11.02826041	2	11.4698003	
Paf1C-Paf1C	Ctr9	Rtf1	166	460	5	14.83564714	5	16.81815641	
Paf1C-Paf1C	Ctr9	Rtf1	166	473	6	8.55129368	6	10.82681373	
Paf1C-Paf1C	Ctr9	Rtf1	166	483	3	13.79860288	2	12.96657624	
Paf1C-Paf1C	Ctr9	Rtf1	166	495	4	8.056505484	4	7.787812396	
Paf1C-Paf1C	Ctr9	Rtf1	180	457	5	11.8096683	2	10.81247928	
Paf1C-Paf1C	Ctr9	Rtf1	180	495	11	17.72584215	5	15.10679325	

Type of crosslinking	Protein a	Protein b	Residue protein a	Residue protein b	Count of Spectrum 1	Max of Score 1	Count of Spectrum 2	Max of Score 2	Distance (Å)/Domains
Paf1C-Paf1C	Ctr9	Rtf1	196	473	3	14.61978876	3	17.01188716	
Paf1C-Paf1C	Ctr9	Rtf1	232	457	4	13.95078198	4	12.59516628	
Paf1C-Paf1C	Ctr9	Rtf1	232	460	5	18.79317412	5	19.97061622	
Paf1C-Paf1C	Ctr9	Rtf1	232	473	6	15.66354027	5	14.17457388	
Paf1C-Paf1C	Ctr9	Rtf1	232	474	4	14.37882372	4	17.51855737	
Paf1C-Paf1C	Ctr9	Rtf1	235	517	3	7.43062609	4	6.360513511	
Paf1C-Paf1C	Ctr9	Rtf1	239	508	4	16.00392635	4	12.67985371	
Paf1C-Paf1C	Ctr9	Rtf1	239	510	4	13.77469072	3	12.91009489	
Paf1C-Paf1C	Ctr9	Rtf1	274	530	7	14.10402527	4	13.80134291	
Paf1C-Paf1C	Ctr9	Rtf1	277	530	7	12.49485002	6	12.74232143	
Paf1C-Paf1C	Ctr9	Rtf1	367	457	2	12.326058	1	7.602059991	
Paf1C-Paf1C	Ctr9	Rtf1	443	280	2	7.600326279	2	7.223298816	
Paf1C-Paf1C	Ctr9	Rtf1	443	432	2	10.91009489	2	9.274905479	
Paf1C-Paf1C	Ctr9	Rtf1	443	450	3	10.39147397	2	9.519993057	
Paf1C-Paf1C	Ctr9	Rtf1	443	460	4	13.22257318	4	13.92081875	
Paf1C-Paf1C	Ctr9	Rtf1	443	473	2	10.96657624	2	12.43533394	
Paf1C-Paf1C	Ctr9	Rtf1	450	284	2	7.360513511	3	9.05207638	
Paf1C-Paf1C	Ctr9	Rtf1	450	457	3	10.94692156	4	11.86966623	
Paf1C-Paf1C	Ctr9	Rtf1	450	460	2	12.35359627	2	10.85078089	
Paf1C-Paf1C	Ctr9	Rtf1	450	473	4	19.05650548	4	21.48545225	
Paf1C-Paf1C	Ctr9	Rtf1	450	510	2	10.80410035	2	9.343901798	
Paf1C-Paf1C	Ctr9	Rtf1	456	517	7	15.41680123	7	12.13727247	
Paf1C-Paf1C	Ctr9	Rtf1	456	522	4	8.772113295	4	9.37675071	
Paf1C-Paf1C	Ctr9	Rtf1	456	530	5	19.20342567	5	18.26520017	
Paf1C-Paf1C	Ctr9	Rtf1	487	530	4	14.92811799	3	16.23732144	
Paf1C-Paf1C	Ctr9	Rtf1	496	530	3	11.26280736	4	15.48017201	
Paf1C-Paf1C	Ctr9	Rtf1	497	530	61	19.66756154	56	18.12959609	
Paf1C-Paf1C	Ctr9	Rtf1	749	284	5	8.774690718	4	8.657577319	

Type of crosslinking	Protein a	Protein b	Residue protein a	Residue protein b	Count of Spectrum 1	Max of Score 1	Count of Spectrum 2	Max of Score 2	Distance (Å)/Domains
Paf1C-Paf1C	Rtf1	Leo1	280	302	2	5.288192771	2	6.298432015	
Paf1C-Paf1C	Rtf1	Leo1	280	337	2	7.029188389	3	6.53313238	
Paf1C-Paf1C	Rtf1	Leo1	280	369	3	15.82973828	3	16.627088	
Paf1C-Paf1C	Rtf1	Leo1	450	337	2	9.88941029	3	8.239577517	
Paf1C-Paf1C	Rtf1	Leo1	460	369	3	17.00656377	3	17.09151498	
Paf1C-Paf1C	Rtf1	Leo1	510	337	3	13.12205305	3	11.30364361	
Paf1C-Paf1C	Rtf1	Paf1	256	109	2	11.60906489	2	8.809668302	
Paf1C-Paf1C	Rtf1	Paf1	256	114	3	8.231361899	3	7.534617149	
Paf1C-Paf1C	Rtf1	Paf1	457	114	3	20.75202673	4	19.19246497	
Paf1C-Paf1C	Rtf1	Paf1	460	90	2	12.56066731	3	9.832682665	
Paf1C-Paf1C	Rtf1	Paf1	460	109	3	15.38404995	3	16.28650946	
Paf1C-Paf1C	Rtf1	Paf1	460	114	3	10.53910216	2	15.10679325	
Paf1C-Paf1C	Rtf1	Paf1	460	137	2	6.463441557	3	8.935542011	
Paf1C-Paf1C	Rtf1	Paf1	473	109	3	17.8827287	3	17.15552282	
Paf1C-Paf1C	Rtf1	Paf1	474	90	2	16.99567863	2	18.73992861	
Paf1C-Paf1C	Rtf1	Paf1	495	90	3	14.9625735	3	12.22475374	
Paf1C-Paf1C	Rtf1	Paf1	495	114	3	7.616184634	2	9.124360063	
Paf1C-Paf1C	Rtf1	Paf1	508	90	11	14.47495519	8	10.25727487	
Paf1C-Paf1C	Rtf1	Paf1	510	109	13	15.75945075	13	15.11861534	
Paf1C-Paf1C	Rtf1	Paf1	510	114	7	14.20065945	6	15.26520017	
Paf1C-Paf1C	Rtf1	Paf1	517	109	19	12.7212464	21	15.79048499	
Paf1C-Paf1C	Rtf1	Paf1	517	114	7	12.24565166	7	12.1857524	
Paf1C-Paf1C	Rtf1	Paf1	522	109	2	10.83268267	3	9.016373713	
Paf1C-Paf1C	Rtf1	Paf1	530	109	2	10.14508698	2	9.621602099	
Paf1C-Paf1C	Rtf1	Paf1	530	114	3	14.10568394	4	14.22767829	
Paf1C-Paf1C	Leo1	Cdc73	305	194	3	10.74472749	3	10.1713401	
Paf1C-Paf1C	Leo1	Cdc73	337	385	3	14.14327111	3	13.4828041	
Paf1C-Paf1C	Leo1	Cdc73	369	205	3	13.4723701	3	11.13489603	

Type of crosslinking	Protein a	Protein b	Residue protein a	Residue protein b	Count of Spectrum 1	Max of Score 1	Count of Spectrum 2	Max of Score 2	Distance (Å)/Domains
Paf1C-Paf1C	Leo1	Cdc73	369	371	2	13.10513034	2	16.97061622	
Paf1C-Paf1C	Leo1	Paf1	122	320	2	11.2873503	6	12.73282827	
Paf1C-Paf1C	Leo1	Paf1	156	295	6	18.99567863	9	15.70553377	
Paf1C-Paf1C	Leo1	Paf1	192	283	2	9.160521953	2	6.410050399	
Paf1C-Paf1C	Leo1	Paf1	225	167	11	14.31247104	11	15.34103516	
Paf1C-Paf1C	Leo1	Paf1	225	170	3	10.86012091	4	12.93930216	
Paf1C-Paf1C	Leo1	Paf1	225	295	5	11.2873503	5	13.70553377	
Paf1C-Paf1C	Leo1	Paf1	302	137	10	10.43179828	13	8.716698771	
Paf1C-Paf1C	Leo1	Paf1	302	283	2	5.89279003	4	6.872895202	
Paf1C-Paf1C	Leo1	Paf1	305	114	2	5.590066877	2	8.609064893	
Paf1C-Paf1C	Leo1	Paf1	305	137	9	10.61261017	11	11.64975198	
Paf1C-Paf1C	Leo1	Paf1	305	283	4	7.434152181	4	6.872895202	
Paf1C-Paf1C	Leo1	Paf1	316	283	8	18.27083521	11	14.93554201	
Paf1C-Paf1C	Leo1	Paf1	331	341	3	7.2915791	4	9.048176965	
Paf1C-Paf1C	Leo1	Paf1	332	283	2	8.54515514	4	7.847711656	
Paf1C-Paf1C	Leo1	Paf1	332	352	2	6.899629455	2	5.279840697	
Paf1C-Paf1C	Leo1	Paf1	337	109	7	15.84163751	6	15.92445304	
Paf1C-Paf1C	Leo1	Paf1	337	114	2	15.9788107	3	13.91009489	
Paf1C-Paf1C	Leo1	Paf1	337	137	4	14.05749589	4	13.55284197	
Paf1C-Paf1C	Leo1	Paf1	337	283	6	17.13846559	6	17.3990271	
Paf1C-Paf1C	Leo1	Paf1	337	307	4	7.70333481	3	9.872895202	
Paf1C-Paf1C	Leo1	Paf1	337	341	4	13.71669877	5	11.47495519	
Paf1C-Paf1C	Leo1	Paf1	337	352	18	12.88605665	11	14.36251027	
Paf1C-Paf1C	Leo1	Paf1	369	109	2	12.64975198	2	13.81530857	
Paf1C-Paf1C	Leo1	Paf1	369	320	3	9.707743929	5	16.03011836	
Paf1C-Paf1C	Leo1	Paf1	369	341	2	10.19654288	3	10.73048706	

Abbreviations

°C	degree celsius
Δ	deletion
Å	ångström
aa	amino acid(s)
APS	ammonium persulfate
ATP	adenosine 5'-triphosphate
bp	base pair(s)
BS3	bis(sulfosuccinimidyl) suberate
BSA	bovine serum albumine
Cdc73	cell division cycle 73
CID	CTD interaction domain
cryo-EM	cryo electron microscopy
CTD	carboxy-terminal domain
CTR	C-terminal repeat region of Spt5
Ctr9	CLN3 requiring 9
Da	Dalton
DMSO	Dimethyl sulfoxide
DRB	5,6-dichloro-1-b -D-ribofuranosylbenzimidazole
DSIF	DRB sensitivity inducing factor
DTT	dithiothreitol elongation
<i>E. coli</i>	<i>Escherichia coli</i>
EC	elongation complex
EDTA	ethylene diamine tetraacetate
EM	electron microscopy
FAM	5' 6- fluorescein
FDR	false discovery rate
g	gram or gravitational acceleration on earth's surface
GraFix	gradient fixation
GTF	transcription factor

HCl	hydrochloric acid
HEPES	2-[4-(2-hydroxyethyl)-1-piperazinyl]-ethanesulfonic acid
His	Histidine
IPTG	isopropyl β -D-1-thiogalactopyranoside
k	thousand
L	liter
LB	lysogeny broth
Leo1	left open reading frame 1
M	molar
min	minute(s)
MOPS	3-(N-morpholino)propanesulfonic acid
MW	molecular weight
NA	nucleic acid(s)
NCBI	National Center for Biotechnology Information
NELF	negative elongation factor
nt	nucleotide
NTP	nucleoside triphosphate open
OD _{600nm}	optical density at 600 nm
ORF	open reading frame
P-TEFb	positive transcription elongation factor b
pA	poly-adenylation
Pa1fC	Paf1 complex
Paf1	polymerase-associated factor 1
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEP	plastid-encoded Pol
PHD	plant homeodomain
PI	protease inhibitor pre-initiation
PIC	pre-initiation complex
Pol	DNA-dependent RNA polymerase

PTM	post-translational modification
RNA	ribonucleic acid(s)
rpm	revolutions per minute
rRNA	ribosomal RNA
Rtf1	restores TBP function 1
<i>S. pombe</i>	<i>Saccharomyce pombe</i>
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
SDS	sodium dodecyl sulfates
SEC	super elongation complex
siRNA	interfering RNA
snRNA	small nuclear RNA
TAE	tris-acetate-EDTA
TBE	tris borate EDTA buffer
TBP	TATA binding protein
TE	tris-EDTA
TEMED	N,N,N',N'-tetramethylethylenediamine
TFII	transcription factor II
Tris	tris-(hydroxymethyl)-aminomethane
tRNA	transfer RNA
TSS	transcription start site
V	volt
v/v	volume per volume
w/v	weight per volume
XFEL	x-ray free electron laser
XL-MS	chemical crosslinking coupled to mass spectrometry
YPD	yeast extract peptone dextrose

List of Figures

1: Pol II transcription cycle.....	4
2: Crystal structures of Paf1C subunits.....	13
3: Predicted domain structure of the five Paf1C subunits from <i>S. cerevisiae</i>	45
4: Preparation of recombinant Paf1C.....	46
5: Preparation of various Paf1C recombinant proteins.....	47
6: Analysis of Pol II-Paf1C-TFIIS EC interactions.....	49
7: Pol II-Paf1C-TFIIS sample preparation for cryo-EM.....	50
8: A representative cryo-EM micrograph of Pol II-Paf1C-TFIIS.....	51
9: 2D cryo-EM class averages of Pol II-Paf1C-TFIIS.....	51
10: Computational sorting of cryo-EM particle images.....	53
11: Cryo-EM reconstruction of Pol II-Paf1C-TFIIS subclasses.....	54
12: Cryo-EM structure of Pol II-Paf1C-TFIIS EC.....	56
13: Crosslinking analysis of Pol II-Paf1C-TFIIS EC.....	59
14: Paf1C and TFIIF compete for Pol II binding.....	62
15: Paf1C is globally required for Pol II transcription in yeast.....	63
16: Rtf1 is very important for Paf1C binding to RNA.....	64
17: Model of the Pol II EC with bound elongation factors.....	66
Supplementary Figure 1: Transcription elongation assay <i>in vitro</i>	82
Supplementary Figure 2: Crosslinking analysis of Paf1C.....	83
Supplementary Figure 3: Cryo-EM reconstitution of Pol II-Paf1C-TFIIS.....	84

List of Tables

1: Subunit composition of multisubunit RNA polymerases	2
2: General transcription factors in yeast	6
3: A list of transcription elongation factors	9
4: <i>E. coli</i> strains used in this study	15
5: Yeast strains used in this study	15
6: Plasmids used in this study	16
7: Synthetic genes used in this study	22
8: Synthetic oligonucleotides in this study	22
9: Growth medium for <i>E. coli</i> and <i>S. cerevisiae</i>	24
10: Medium additives for <i>E. coli</i> and <i>S. cerevisiae</i>	24
11: General buffers, dyes and solutions	25
12: Buffers used for preparation of competent <i>E. coli</i> cells.....	26
13: Enzymes, buffers, and components used for cloning	27
14: Buffers used for Paf1C purification, crosslinking and pull-down	27
15: Buffers used for sucrose gradient ultracentrifugation and fixation.....	28
16: Buffers used for fluorescence anisotropy assays	29
17: Buffers used for transcription elongation assays	29
18: Summary of BS3 crosslinks in Pol II-Paf1C-TFIIS EC	58
Supplementary Table 1: Summary of various recombined Paf1 complexes	85
Supplementary Table 2: High-confidence crosslinks of Paf1C	86
Supplementary Table 3: High-confidence crosslinks of Pol II-Paf1C-TFIIS complex ..	89