

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

Cryo-electron microscopic studies of RNA polymerase complexes



Anselm G. Kusser
aus München

2010

Erklärung

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

Ehrenwörtliche Versicherung

Diese Dissertation wurde selbständig und ohne unerlaubte Hilfe erarbeitet.

München, den 14.05.2010

Anselm Kusser

Dissertation eingereicht am 14.05.2010

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

Mündliche Prüfung am 21.03.2011

Acknowledgements

I am deeply grateful to my parents, and I want to address this directly to them. You brought me to life. You supported me in all situations of life, no matter what the concern was: personal, intellectual or financial. You taught me life's basics of being sincere, honest and straightforward and many more of the characteristics and little details that make me the person I am today. I am proud of what I accomplished so far and I look forward to what lies ahead.

I love you, and I wish to dedicate my PhD Thesis to you.

I want to thank Patrick Cramer, who gave me the opportunity to graduate in his lab. His open-mindedness and seemingly never-ceasing endurance kept me motivated over all these years. The door of his office was always open for me to enter and directly raise any concern I would have: scientific, personal or merely bureaucratic; I took quite a leaf out of his book.

A great captain selects a great crew! I really enjoyed my four years in Patrick's group. It was a very collaborative and friendly environment, which is the basis for cutting-edge research, like we are able to conduct it here. I especially want to give a shout-out to Anass for getting me started in the beginning of my PhD and a high-five to Ale for the fun the Cryo-team had in- and outside of the lab, at the microscope or just partying/DJing. Many thanks also to Wenjie and Rieke for the fruitful and very enjoyable collaborations. A special thank you goes to Alan's native tongue and I hope I will get to do an EM-project together with him before he leaves our group. Finally, thanks to all the other members of the Cramer lab that I could not address directly here.

A great deal of energy was needed to complete this Thesis and I am grateful to Erich Sämmer, my Aikido teacher, for all the practice and mental strength I received through him. Arigatō gozaimasu Erich-san!

Summary

Related RNA polymerases (RNAPs) carry out gene transcription all three domains of life. This thesis deals with the structure determination of RNAPs and their functional complexes from different species. Protein complexes were preserved in their native state in aqueous solution, imaged by cryo-transmission electron microscopy and structural models were obtained using the single particle reconstruction method. New and physiologically relevant insights into RNAPs subunit architecture, the general transcription mechanism and its regulation were gained.

The structure of an archaeal RNA polymerase identified similarities to its eukaryotic counterpart, RNA polymerase II. The conservation of the overall enzyme architecture as well as the close resemblance of structural elements and functional surfaces needed for basic transcription mechanisms underlines the evolutionary relationship between archaeal and eukaryotic RNAPs.

The comprehensive study of RNA polymerase III and its regulation by Maf1 gave profound insights into the molecular basis of how eukaryotic transcription is shutdown under stress conditions to ensure cell survival. Maf1 binds RNAP III at its clamp domain and rearranges a specific subcomplex needed for interaction with the initiation factor Brf1. This specifically impairs binding of RNAP III to its promoters and inhibits transcription initiation. Furthermore, it was demonstrated that Maf1 binds to RNAP III that is already engaged in transcription elongation, thus leaving activity intact but preventing re-initiation. Taken altogether, these results converge on the essential mechanism of RNAP III-specific transcription repression by Maf1.

Publications

Part of this work has been published or is in the process of being published.

Vannini, A., Ringel, R., Kusser, A.G., Berninghausen, O., Kassavetis, G., Cramer, P. (2010) Molecular basis of RNA polymerase III transcription repression by Maf1. *Submitted*.

Kusser, A.G., Bertero, M.G., Naji, S., Becker, T., Thomm, M., Beckmann, R., Cramer, P. (2008) Structure of an Archaeal RNA Polymerase. *J. Mol. Biol.* **376**, 303–307.

Table of contents

Erklärung	I
Ehrenwörtliche Versicherung	II
Acknowledgements	III
Summary	IV
Publications	V
Table of contents	VI
Chapter I: General introduction	1
1 Life and the flow of genetic information	2
2 Transcription by DNA dependent RNA polymerases	3
2.1 Subunit composition of RNA polymerases	4
2.2 Structural information on RNA polymerases	5
3 Regulation of transcription	7
4 Dynamics of transcription	11
5 Open questions and scope of this work	12
Chapter II: Cryo-EM structure of an archaeal RNA polymerase	14
1 Introduction and aim of this study	15
1.1 <i>Pyrococcus furiosus</i>	15
1.2 Archaeal transcription	16
1.3 Aim of this study	17
2 Results and discussion	18
2.1 Cryo-EM analysis of the <i>Pyrococcus furiosus</i> RNA polymerase	18
2.2 Conclusions	21
3 Outlook	22
4 Experimental procedures	23
4.1 Measurement of protein concentration	23
4.2 Purification of <i>Pyrococcus furiosus</i> RNA polymerase	23
4.2.1 Chemicals and buffers used for purification	23
4.2.2 Purification of endogenous <i>P.fu</i> RNA polymerase	26
4.3 Cryo-electron microscopic structure determination	27
Chapter III: Molecular basis of RNA polymerase III transcription repression by Maf1	30
1 Introduction	31

2 Results and discussion	33
2.1 Pol III EM structure reveals C82/34/31 mobility	33
2.2 Nucleic acid binding restricts C82/34/31 mobility	36
2.3 Maf1 structure is globular, not modular	38
2.4 Regulated Maf1 localization	41
2.5 Maf1 binds the Pol III clamp and rearranges C82/34/31	43
2.6 Maf1 impairs closed promoter complex formation	45
2.7 Maf1 does not inhibit Pol III activity	47
3 Conclusions and outlook	49
4 Experimental procedures	50
4.1 Purification of yeast RNA polymerase	50
4.1.1 Chemicals and buffers used for Pol III purification	50
4.1.2 Purification of endogenous yeast Pol III	54
4.2 Maf1 purification	56
4.2.1 Chemicals and buffers used for Maf1 purification	56
4.2.2 Purification of Maf1	58
4.3 Proteolytic cleavage of Maf1	59
4.4 Maf1 crystal structure determination	59
4.5 Cryo-EM structure determination	60
4.6 Nanogold™ labelling	62
4.7 Initiation factor-independent transcription assay	63
4.8 RNA-extension assay	63
Appendix: Unpublished results	65
1 <i>Pyrococcus furiosus</i> RNAP minimal pre-initiation complex studies	66
1.1 Sample preparation and cryo-EM data collection	66
1.2 Cryo-EM data processing and classification trials	68
2 Yeast RNAP III minimal pre-initiation complex studies	73
2.1 Sample preparation and cryo-EM data collection	73
2.2 Cryo-EM data processing and results	73
3 Technical issues addressed during this work	79
3.1 Electron acceleration voltage and image contrast	79
3.2 CCD vs film data	81
3.3 Attempts to resolve heterogeneity	82
3.3.1 Unsupervised classification	82
3.3.2 A “hybrid-approach”	88
4 Conclusions and outlook	93
Final conclusions and outlook	95
Abbreviations	96
References	98
Curriculum vitae	108

Chapter I

General Introduction



"Inochi" ("Life")

(Calligraphy, K. Kuwahara)

1 Life and the flow of genetic information

A living organism can be defined in biology as follows (Nealson and Conrad, 1999): it has a structure and a unique chemistry associated with that structure, it strives to replicate with fidelity, it evolves, it has a metabolism by which it consumes energy from its environment, thus creating metabolic products and it has a means for escaping these products.

Replication with fidelity and evolution are parts of a plan implemented in an organisms DNA, which is the carrier of genetic information in all known living beings on Earth (Berg, 2007). The flow of genetic information from DNA via RNA to protein is termed the central dogma of molecular biology (Crick, 1970), a subdivision of biology concerned with studying the chemistry of life. The pathway in this dogma divides up into two main parts: the rewriting of genetic information from DNA into RNA, a process called transcription, and the subsequent translation of RNA into proteins. Transcription is carried out by DNA dependent RNA polymerases. Translation at the ribosome builds a protein product that can then be further modified post-translationally to become fully functional (Berg, 2007).

Knowledge of a biological macromolecule's structure can lead to a detailed understanding of its function as it was already demonstrated in the early 1950s, when Watson & Crick postulated the DNA double helix structure and its implications for the transmission mechanism of genetic information (Watson, 1953). DNA is an example for the direct relationship of form and function in biological macromolecules: the sequential arrangement of bases on a strand of DNA allows to code for genes and regulatory elements and thus for storage of genetic information. Furthermore, it is the ability of bases to specifically pair that lays the foundations for DNA replication and transcription. A double-stranded parental DNA can be divided up into two single DNA strands, from which by base pairing it is possible to reconstruct two double-stranded daughter DNAs carrying the same information as the parental DNA. This process of replication allows for transmission of genetic information from a cell to its daughter cells and it is carried out by DNA dependent DNA

polymerases (Berg, 2007). In a similar fashion it is possible to transcribe a region of DNA into a RNA copy. Transcription allows for expression of individual genes needed in a certain cellular state or context and is carried out by DNA dependent RNA polymerases.

2 Transcription by DNA dependent RNA polymerases

Since all organisms are dependent on the usage of genetic information stored in their DNA, they all possess RNA polymerases (RNAPs) to transcribe genes (Berg, 2007). Bacteria and archaea possess only one type of RNAP that transcribe all types of RNA in the cell whereas in eukaryotes, there are at least three different nuclear RNAPs (Berg, 2007; Carey, 2009). Eukaryotic RNAP I is located in the nucleoli and is specialized in production of precursors for 5.8S, 18S and 28S ribosomal RNA (rRNA). RNAP II is located in the nucleoplasm and is responsible for transcription of messenger RNA (mRNA) from all protein coding genes as well as for production of small nucleolar RNA (snoRNA) and small nuclear RNA (snRNA). RNAP III is mainly located in the nucleoplasm and specialized in transcription of transfer RNA (tRNA), 5S RNA, U6 small nuclear RNA and other small catalytic RNAs (Berg, 2007; Carey, 2009). Cytoplasmic localization of RNAP III was only recently discovered and linked to RNAP III being involved in immunological responses to pathogenic bacterial DNA (Chiu et al., 2009). In plants, two additional RNAP II related polymerases have lately been reported (RNAP IV and V) that are connected to the RNA interference machinery and generate noncoding RNAs responsible for epigenetic regulation (Matzke et al., 2009; Ream et al., 2009). All these RNAPs are protein complexes that differ in subunit composition (Cramer, 2002).

2.1 Subunit composition of RNA polymerases

RNA polymerase	RNAP I	RNAP II	RNAP III	Archaeal RNAP	Bacterial RNAP
Ten-subunit core	A190	Rpb1	C160	A' + A''	β'
	A135	Rpb2	C128	B (B' + B'')	β
	AC40	Rpb3	AC40	D	α
	Rpb5	Rpb5	Rpb5	H	-
	Rpb6	Rpb6	Rpb6	K	ω
	Rpb8	Rpb8	Rpb8	-	-
	A12.2	Rpb9	C11	-	-
	Rpb10	Rpb10	Rpb10	N	-
	AC19	Rpb11	AC19	L	α
Rpb12	Rpb12	Rpb12	P	-	
Rbp4/7-like subcomplex	A14	Rpb4	C17	F	-
	A43	Rbp7	C25	E'	-
TFIIF-like subcomplex	A49 (N-ter.)	-	C37	-	-
	A34.5	-	C53	-	-
TFIIE-like subcomplex	-	-	C82	-	-
	A49 (C-ter.)	-	C34	-	-
	-	-	C31	-	-
Number of subunits	14	12	17	11 (12)	5
Molecular weight in kDa (species)	589 (<i>S.cerevisiae</i>)	514 (<i>S.cerevisiae</i>)	693 (<i>S.cerevisiae</i>)	380 (<i>P.furiosus</i>)	375 (<i>T.aquaticus</i>)

Table 1: Subunit composition of RNAPs, adapted from (Brueckner, 2008; Cramer et al., 2008; Kusser et al., 2008)

Almost all RNAPs are multi-subunit protein complexes except for some bacteriophage RNAPs like T7 (Sousa et al., 1993; Steitz, 2009) and the mitochondrial RNAPs (Gaspari et al., 2004). All three domains of life, however, utilize multi-subunit RNAPs for transcribing their genomic DNA (Table 1). The bacterial RNAPs comprise four subunits: α , β , β' and ω with stoichiometry $\alpha_2\beta\beta'\omega$, archaeal RNAPs consist of up to 13 subunits (Korkhin et al., 2009) while in eukaryotes, RNAPs can have as much as 17 subunits (Werner et al., 2009). Interestingly, the archaeal RNAP bears much closer resemblance to eukaryotic than to the bacterial RNAP, structurally as well as regarding subunit composition, suggesting that it is a more recent evolutionary ancestor of the eukaryotic transcription apparatus than the bacterial RNAP (Hirata and Murakami, 2009). This phylogenetic relationship is supported by

other studies like comparison of 16S ribosomal RNA of different representative species (Woese et al., 1990).

RNAP II consists of a 10-subunit core enzyme and the peripheral heterodimer Rpb4/7. The core enzyme comprises subunits Rpb1, 2, 3, and 11, which contain regions of sequence and structural similarity to the other multisubunit RNAPs (Korkhin et al., 2009; Hirata et al., 2008; Kusser et al., 2008; Vassylyev et al., 2002; Zhang et al., 1999), and subunits Rpb5, 6, 8, 10, and 12, which are shared between RNAP I, II and III (Table 1).

2.2 Structural information on RNA polymerases

As exemplified above, the understanding of a macromolecules function generally benefits tremendously from knowledge about its structure. There is a variety of structural information available, ranging from high-resolution X-ray crystallographic structures of RNAPs (Cramer, 2002; Hirata and Murakami, 2009; Vassylyev et al., 2002) single molecule FRET (Chen et al., 2009; Tang et al., 2009) and mass spectrometry studies of RNAPs (Chen et al., 2010; Lorenzen et al., 2007) to low-resolution cryo-electron microscopic structures of eukaryotic RNAPs (Fernandez-Tornero et al., 2007; Kostek et al., 2006; Kuhn et al., 2007).

High-resolution crystal structures have been determined for RNAPs from all three domains of life: bacteria (Vassylyev et al., 2002; Zhang et al., 1999), archaea (Hirata et al., 2008; Korkhin et al., 2009) and eukaryotes (Cramer et al., 2001). These structures showed that all the investigated multi-subunit RNAPs share a common core and a highly conserved active center region formed by Rpb1 and Rpb2 or their corresponding subunits (Table 1, Figure 1). The smaller subunits are arrayed around the periphery of the enzyme, forming functional domains (Figure 1).

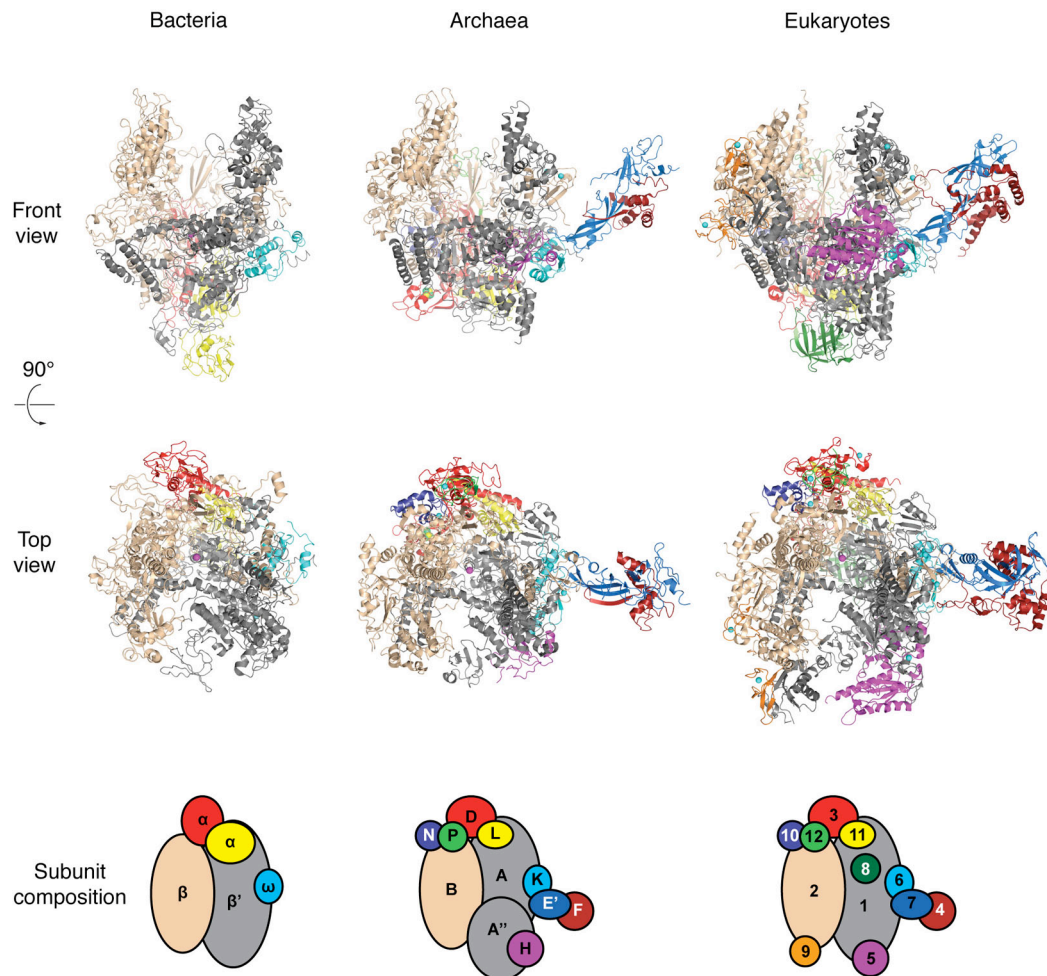


Figure 1: Structural comparison of RNAPs from the three domains of life

A five subunit bacterial RNAP (*T.Aquaticus*, Zhang et al., 1999, left column), a 12-subunit archaeal RNAP (*S.Solfataricus*, Hirata et al., 2008, middle column) and a 12-subunit eukaryotic RNAP (*S.Cerevisiae*, Armache et al., 2005, right column) are shown in front view (top row), top view (middle row) and in a top view schematic representation of their subunit composition and arrangement (bottom row, compare Table 1).

Most structural data have been accumulated for RNAP II and its functional complexes. These include structures providing insight into the nucleotide addition cycle (Brueckner et al., 2009), how RNAP II deals with damages on the DNA template (Brueckner et al., 2007; Damsma et al., 2007; Damsma and Cramer, 2009), transcriptional fidelity and pausing (Sydow et al., 2009), backtracking of the product RNA (Wang et al., 2009) as well as the transcription initiation mechanism (Kostrewa et al., 2009; Liu et al., 2009). It was, however, not yet possible to obtain high-resolution structural information

for the specialized eukaryotic RNAPs I and III: they have resisted X-ray crystallographic studies so far (Cramer et al., 2008). Nonetheless, it was possible to obtain atomic models of subcomplexes A14/43 of RNAP I (Geiger et al., 2008; Kuhn et al., 2007) and C17/25 of RNAP III (Jasiak et al., 2006), subunits that are homologous to the Rpb4/7 subcomplex from RNAP II (Table 1). Structures of subcomplexes specific to either one of the RNAPs I and III were solved only recently (Geiger and Cramer, unpublished; Fribourg, unpublished). Furthermore, it has recently been found that subunits specific to RNAP I and III are homologous to transcription factors from the RNAP II system (Table 1; Carter and Drouin, 2009; Kassavetis et al., 2010; Kuhn et al., 2007; Geiger and Cramer, unpublished). Transcription factors (TFs) are proteins that regulate RNAP in the various stages of the transcription cycle (Berg, 2007).

3 Regulation of transcription

Expression of a gene embraces transcription of its DNA sequence into RNA, and for most genes, translation of the RNA into a protein product. A genome usually contains thousands of genes, many of which are expressed almost constantly (Berg, 2007; Carey, 2009). Other genes are merely expressed in specific situations or cellular states. These facultative genes may encode for proteins being needed under stress conditions or for an immune response to an infection. All gene expression, however, is regulated: it may for example be necessary to down regulate cell growth when there is only little food source available. Yet for facultative genes, tight regulation is even more crucial to ensure expression at adequate time points and spatial locations, as in the case of an immune response (Berg, 2007).

Transcription regulation has a profound impact on gene expression, as it is one of the first major steps that entails, and may consequently indirectly regulate all subsequent steps (Berg, 2007; Carey, 2009). Regulation of transcription in prokaryotes is relatively well understood and relies mainly on

two principles: the repression of genes that are not needed in the current state of the cell as well as the activation of genes by stimulation of RNAP activity (Berg, 2007). The most famous example is probably the *lac*-operon, its repression by the *lac*-repressor and its activation by CAP (Balaeff et al., 2004; Bell and Lewis, 2001; Lewis et al., 1996).

In eukaryotes, regulation of transcription is much more complex (Venters and Pugh, 2009). Eukaryotes have their genomic DNA condensed into chromatin, such as to be able to store it within the limited space of the cellular nucleus. The DNA is wrapped around histones, thus forming nucleosomes, which are in turn assembled into higher-order structures with different properties, depending on the regulatory context (Carey, 2009). Chromatin also maintains genes in an inactive state by restricting access to RNAPs and accessory factors. To activate a gene, the chromatin encompassing that gene and its regulatory regions must to be unpacked or “remodelled” to permit transcription (Berg, 2007; Carey, 2009).

Assuming that the gene is spatially accessible, RNAP needs to engage in transcription, a process that can be divided up into three main stages: transcription initiation, elongation and termination (Svejstrup, 2004). Initiation terms finding of the right start site (within the promoter region), loading and opening of double stranded DNA, synthesis of the first phosphodiester bonds between single RNA nucleotides and ultimately escaping the promoter region to elongate the nascent RNA chain. Elongation is the stage of processive RNA production, a task that is linked to post-transcriptional RNA processing (e.g. splicing). Termination finally stops transcription at the correct site and is linked to RNA transcript release, allowing RNAP to be released from the template and to re-engage in another round of transcription (compare Figure 2A). All these steps are performed by RNAP in a complex and most dynamic interplay with scores of regulatory and catalytic molecules (Hager et al., 2009; Venters and Pugh, 2009), a subset of which is referred to as the general transcription machinery (GTM). In the RNAP II system, the general transcription factors TFIIA, TFIIB, TFIID, TFIIE, TFIIIF and TFIIH as well as a major coactivator termed the Mediator form the GTM (compare Figure 2,

Table 2), together with RNAP II itself (Carey, 2009). It is strongly believed that all these proteins, together with other coactivators, are needed to properly initiate eukaryotic transcription *in vivo* (Kelleher et al., 1990; Kornberg, 2007; Reinberg et al., 1998; Takagi and Kornberg, 2006).

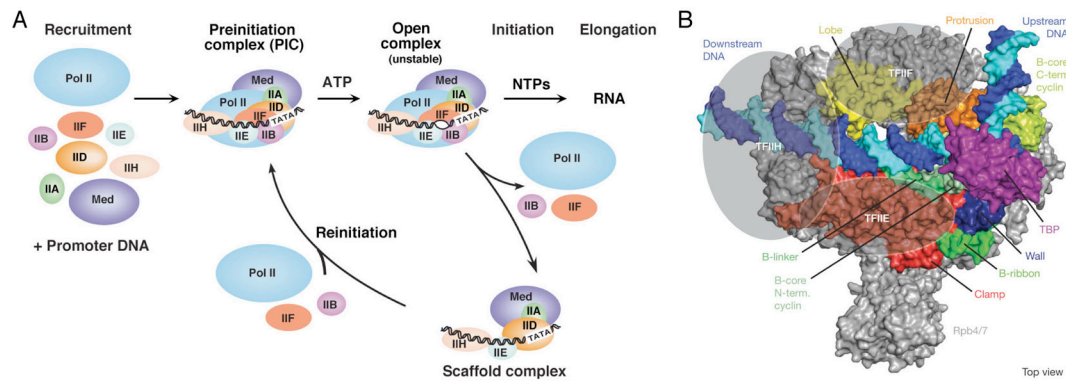


Figure 2: Transcription initiation and RNAP II PIC architecture

(A) Pathway of transcription initiation and RNAP II PIC subunit composition (compare Table 2, adapted from Hahn, 2004)

(B) Current model of the RNAP II PIC architecture (without Mediator, compare Table 2, adapted from Kostrewa et al., 2009)

A term that is commonly used in the context of transcription initiation is the pre-initiation complex (PIC). The yeast RNAP II PIC will be defined here as the assembly of RNAP II together with the GTFs TFIIB, TFIID, TFIIE, TFIIF and TFIIH (compare Figure 2, Table 2). A subset of these, namely the complex of RNAP II together with TFIIB, TFIIF and the TATA-binding protein TBP, which is a part of TFIID (compare Table 2), will be referred to as the yeast RNAP II minimal PIC (mPIC) since it is sufficient to initiate transcription at artificial promoters *in vitro* (Moqtaderi et al., 1996; Tyree et al., 1993).

For the specialized eukaryotic RNAPs I and III, the PIC and mPIC definition differs. Recent results suggest, that in case of the RNAPs I and III, modified versions of transcription factors from the RNAP II system have been permanently incorporated into the enzymes to provide for their higher specificity to the target genes of this highly determined machines. Subunits A49 and A34.5 of RNAP I and subunits C53 and C37 of RNAP III show homology to TFIIF (Geiger and Cramer, unpublished; Kassavetis et al., 2010)

while the C-terminus of A49 as well as regions of C34 and C82 structurally resemble TFIIE (Geiger and Cramer, unpublished). This incorporation of transcription factors during evolution is most certainly a means of optimization, or “becoming better” at performing a specific task.

Factor		Nr. of subunits	Function
TFIIA		2	Stabilizes TBP and TFIID-DNA binding. Blocks transcription inhibitors. Positive and negative gene regulation.
TFIIB		1	Binds TBP, Pol II and promoter DNA. Helps fix transcription start site.
TFIID	TBP	1	Binds TATA element and deforms promoter DNA. Platform for assembly of TFIIB, TFIIA and TAFs.
	TAFs	14	Binds INR and DPE promoter elements. Target of regulatory factors.
TFIIF		3	Binds Pol II and is involved in Pol II recruitment to PIC and in open complex formation.
TFIIE		2	Binds promoter near transcription start. May help open or stabilize the transcription bubble in the open complex.
TFIIH		10	Functions in transcription and DNA repair. Kinase and two helicase activities. Essential for open complex formation.
Mediator		24	Binds cooperatively with Pol II. Kinase and acetyltransferase activity. Stimulates basal and activated transcription. Target of regulatory factors.

Table 2: The yeast general transcription factors, adapted from (Hahn, 2004)

Available structural data regarding transcriptional regulation is largely limited to the yeast RNAP II system. Several structures of isolated general transcription factors (GTFs) or (sub-) complexes of these (including some from other systems) were reported (Elmlund et al., 2009; Geiger et al., 1996; Jawhari et al., 2006; Juo et al., 2003; Nikolov et al., 1995; Okuda et al., 2000; Schultz et al., 2000; Tan et al., 1996; Tsai, 2000). Data from comprehensive studies of the PIC architecture encompass X-ray studies of RNAP II-TFIIB complexes (Kostrewa et al., 2009; Liu et al., 2009) and site-specific biochemical probing (Chen and Hahn, 2003) as well as research concerned with the transcription factors TFIIE, TFIIF, TFIIH and TFIIS (Chen et al., 2007; Chen et al., 2010; Kim et al., 2007; Kim et al., 2000) and the path of promoter

DNA in the PIC (Miller and Hahn, 2006). Combination of these data resulted in the current model of the eukaryotic PIC architecture (Figure 2B, Kostrewa et al., 2009). TBP and TFIIB bind close the wall domain of RNAP II and position DNA on top of the active center cleft. TFIIH binds near the polymerase jaws and is required for DNA opening (Kim et al., 2000). TFIIIE binds to the clamp domain where it may stabilize open complex DNA (Chen et al., 2007; Werner and Weinzierl, 2005), while TFIIIF binds on the opposite side where it may contact downstream and upstream DNA (Chen et al., 2010; Chen et al., 2007; Kim et al., 2000).

It should, however, be kept in mind that the *in vivo* situation could differ to a dramatic extent. Many of the known factors can occur in multiple splice variants in the cell (Carey, 2009) increasing the combinatorial intricacy of these multi-component complexes even further. It is also not clear if a complete PIC assembles at all active promoters: it was recently suggested that RNAP and other components assemble stochastically, meaning that transcription complex or possibly even RNAP II assembly happens as a progressive series of transient subcomplexes that form and disintegrate. Increasing evidence suggests that such transcription factor dynamics are indeed a physiologically relevant mechanism of regulating transcription (Hager et al., 2009).

4 Dynamics of transcription

Most, if not all aspects of transcription and its regulation involve dynamic events: the basal transcription machinery is dynamically directed to its target genes, regulatory factors dynamically activate and repress transcription according to the cellular condition and transcription factor interaction with RNAP during the transcription cycle is highly dynamic as well (Hager et al., 2009). These dynamics involve movements that are hard to investigate by X-ray crystallographic studies since these provide only single snapshots of complexes trapped in a crystalline form; how states transition into one another

remains unclear for the time being. Combination of a plethora of different X-ray crystallographic snapshots can provide information about conformational dynamics (Brueckner et al., 2009) but crystals must be painstakingly grown and large amounts of pure protein are needed.

Cryo-EM does not *per se* limit the possibility to investigate complexes bearing high intrinsic flexibility: it allows to image single molecules preserved in their native state in aqueous solution, which is the basis for subsequent image analysis with respect to conformational heterogeneity (Leschziner and Nogales, 2007). One approach towards the understanding of transcription regulation and dynamics on a molecular level will thus be creating high-resolution snapshots of (sub-) complexes or just single modules by X-ray crystallography and interconnecting these snapshots using cryo-EM of larger assemblies, ideally including the study of intrinsic flexibility. This combination of miscellaneous complementary principles to form one unified new technique is another means of evolving, of “becoming better”.

5 Open questions and scope of this work

Transcription is one of the very first steps of gene expression and, in a similar fashion, can transcription initiation be viewed as the first step of transcription (Berg, 2007). Regulation of an early step has dramatic effects, because all subsequent steps can be indirectly co-regulated: e.g. when there is no initiation, there is no elongation, no termination and no RNA maturation (plus, in the case of mRNA, no translation and hence no functional protein product): the gene is not expressed. It is therefore commonly assumed that most transcriptional control works through regulating initiation (Juven-Gershon et al., 2008) and understanding transcription initiation and its regulation should give profound insights into regulation of gene expression in general.

Since diseases like developmental disorders or certain cancers have been found closely linked to misregulated gene expression (D'Alessio et al., 2009; Ho and Crabtree, 2010; Ransom et al., 2010; Sahar and Sassone-

Corsi, 2009; Shor et al., 2010), getting to the bottom of these regulation processes will help to understand the origins of such diseases and thereby improve adequate treatment of patients.

Prior to this work, only little was known about pre-initiation complex (PIC) formation, architecture and regulation. Even when taking into consideration the latest insights, the positions of transcription factors other than the sub-domains of TFIIB observed in the crystal structures remain only vaguely defined. TBP for example, is an essential component for initiation at many promoters that is highly conserved between archaea and eukaryotes. Its position within the mPIC can so far only be modelled by superimposition of common parts from two distinct structures: one of a TFIIB-TBP-TATA-element ternary complex (Nikolov et al., 1995) and the actual structure of the RNAP II-TFIIB complex (Kostrewa et al., 2009). Furthermore, there was only very little data available for regulation of PIC complex formation and PIC architecture in other systems like RNAP I and III or even just about the basic RNAP architecture in the archaeal system.

The scope of this work was therefore to investigate by cryo-electron microscopy and single particle analysis, firstly, the core RNAP and mPIC architecture in the archaeal system to gain insights about evolutionary conservation on a structural level and secondly, the regulation of mPIC complex formation and the general mPIC architecture in the yeast RNAP III system to draw parallels to the RNAP II and other systems, thus bringing forward our understanding of transcription in general.

Chapter II

Cryo-EM structure of an archaeal RNA polymerase

"Marvellous enough that there are life forms withstanding temperatures above 100°C, but hyperthermophiles which thrive and prosper under these conditions seem beyond any comprehension."

(Prof. Karl Otto Stetter, discoverer of *P.fu*)

1 Introduction and aim of this study

Hyperthermophiles, a group of superheat-loving life forms that include many members of the archaeal domain, seem to be our most early ancestors (Stetter, 2006). Their traces go back as far as 3.5×10^9 years, to the era of the Early Archaean (Schopf and Packer, 1987). Archaea are classified into four kingdoms: Euryarchaeota, Crenarchaeota, Nanoarchaeota and Korarchaeota (Ouhammouch, 2004). In a phylogenetic tree based on comparison of small subunit ribosomal RNA (Woese et al., 1990), it becomes obvious that archaea are more closely related to eukaryotes than bacteria are (Stetter, 2006). Many archaea are extremophiles, thriving in fierce environments with extreme temperature, acidity, salinity or pressure, like the Euryarchaeon *Pyrococcus furiosus*.

1.1 *Pyrococcus furiosus*

The archaeon *Pyrococcus furiosus* (*P.fu*) was first discovered in 1986, when it could be extracted from geothermally heated marine sediment taken by Professor Karl Otto Stetter at the isle of Volcano, Italy (Fiala, 1986). It is spherically shaped with a size of about 0.8 to 2.5 μm and exhibits monopolar polytrichous flagellation used for locomotion, adherence to surfaces and for building cell-cell-connections (Nather et al., 2006). *P.fu* is strictly anaerobic, sulfur-dependent and heterotrophic, and is capable of growing at a pH range from 5 to 9 and at temperatures ranging from 70° to 103°C with an optimal growth at 100°C (Fiala, 1986). Its doubling time is among the shortest found in Archaea – only 37 min under optimal conditions – which gave rise to its name *Pyrococcus furiosus*, meaning “rushing fireball”. The genome of *P.fu* has been sequenced and found to be approximately 1.9 Mbp in size containing 2,192 open reading frames (Poole et al., 2005). These properties put forward *P.fu* to study it on a molecular level. Furthermore, *P.fu* can serve as a simplified model system for eukaryotes: comprehending its basal molecular biology, like

the process of transcription, might underline what is most essential in higher organisms and provide insights into their evolution (Werner, 2007).

1.2 Archaeal transcription

Transcription in archaea is carried out by a basal machinery that bears much closer resemblance to the eukaryotic transcription apparatus rather than to the prokaryotic one (Langer et al., 1995). Transcription factors TBP and TFB are homologous to the core initiation factors of eukaryotic RNAP II and III (Thomm, 1996). TBP, TFB and RNAP are necessary and sufficient for open complex formation at many archaeal promoters *in vitro* whereas some promoters require TFE, the archaeal homologue of the α -subunit of TFIIE (Ouhammouch, 2004). No evidence has been found so far for homologues of eukaryotic transcription factors IIA, IIF, IIH or the β -subunit of TFIIE (Ouhammouch, 2004) but the fact that several archaeal species possess multiple paralogues of TBP and TFB, suggests the possibility of differential gene expression through the formation of alternative TBP-TFB-RNAP complexes (Baliga et al., 2000).

Archaeal RNA polymerases contain 10 to 13 subunits, most of which show direct homology to a corresponding eukaryotic RNAP II subunit (Hirata and Murakami, 2009; Werner, 2007). The RNAP from *P.fu* is an 11-subunit enzyme that displays high sequence identity to the eukaryotic RNAP II and its subunit composition is much alike: direct homologs for nearly all eukaryotic RNAP II subunits exist, however, A' and A'' together form the homolog of Rpb1, while homologs for Rpb8 and Rpb9 are missing (Table 3, Goede et al., 2006). The archaeal RNAP can thus be seen as a simplified version of the eukaryotic RNAP (Hirata and Murakami, 2009; Kusser et al., 2008). Another substantial advantage of *P.fu* RNAP is, that all subunits can be expressed recombinantly and assembled *in vitro* to form a catalytically active enzyme (Naji et al., 2007). This allows for systematic functional testing of mutant variants to a degree, not possible for any eukaryotic RNAP to date.

Furthermore, the close resemblance of archaeal and eukaryotic RNA polymerases enables transfer of structure-function relationships from one system over to the other, to certain extent. This approach has been successfully applied in case of the eukaryotic RNAP II-TFIIB complex X-ray structure (Kostrewa et al., 2009).

High-resolution structures of archaeal RNA polymerases have been solved from *Sulfolobus shibatae* (Korkhin et al., 2009) and *Sulfolobus solfataricus* (Hirata et al., 2008). These are similar to the eukaryotic RNAP II in overall size and shape but the *S.shibatae* RNAP contains a 13th subunit, absent in the eukaryotic enzyme while the RNAP from *S. solfataricus* contains an additional iron-sulfur cluster (compare Figure 1). However, no high-resolution structure of a complete *P.fu* RNAP exists so far, and, at the time when the research described in this chapter was initiated, no structural information on a complete archaeal RNA polymerase was available whatsoever.

1.3 Aim of this study

Prior to this work, attempts were being made to obtain the crystal structure of *P.fu* RNAP in its free form and in complex with transcription initiation factors and nucleic acids (Patrick Cramer, personal communication). Initial crystals could be acquired but were only poorly diffracting and did not allow for solving the corresponding structures. Since Cryo-EM single particle analysis does not rely on protein crystal formation, the scope of this study was, to apply that method for obtaining structural data at lower resolution that would reflect the enzyme's general architecture and would hopefully allow for visualization of flexible regions. This information may then assist the optimization of constructs for crystallization trials to ultimately obtain structural information at atomic resolution.

2 Results and discussion

2.1 Cryo-EM analysis of the *Pyrococcus furiosus* RNA polymerase

To tackle the archaeal RNAP structure, we established a large-scale purification protocol for the endogenous polymerase from *Pyrococcus furiosus* (*P.fu*) based on a previously published method (Experimental procedures, Hethke et al., 1996). The improved protocol enabled purification of five milligrams of RNAP from 60 g of *P.fu* cells. Pure RNAP comprised all subunits in an apparently stoichiometric manner (Figure 3A). RNAP preparations were monodisperse and catalytically active (data not shown), contained single particles according to EM with negative staining, and enabled collection of high quality cryo-EM data (Figure 3B). The reconstruction from 22,240 conformationally uniform particles led to a map at 13 Å resolution (Figure 3).

Polymerase part	<i>P.fu</i> subunit	MW (kDa)	Corresponding RNAP II subunit	Sequence identity ¹ (%)
Core	A'	103.1	Rpb1 N-term. Part	42.2
	A''	44.4	Rpb1 C-term. Part	37.0
	B	127.0	Rpb2	43.2
	D	29.8	Rpb3	26.8
	H	9.2	Rpb5	40.2
	K	6.2	Rpb6	42.1
	-	-	Rpb8	-
	-	-	Rpb9	-
	N	7.8	Rpb10	52.3
	L	11.1	Rpb11	30.5
	P	5.8	Rpb12	36.7
Subcomplex Rpb4/7	E'	21.7	Rpb7	22.2
	F	14.1	Rpb4	5.0
Total	-	380.2	-	38.2

¹Number of amino acid residues in the *P.fu* RNAP subunit that are identical in the corresponding RNAP II subunit divided by the total number of residues in the *P.fu* RNAP subunit.

Table 3: *P.fu* RNAP subunits in comparison to RNAP II.

The detailed EM map for *P.fu* RNAP enabled a unique fit of the crystal structure of the complete 12-subunit *S.cerevisiae* RNAP II (Armache et al., 2005) (Figure 3D). Comparison of the EM map with the RNAP II structure confirmed the overall conservation of the enzyme architecture and active center, including the polymerase bridge helix, the pore, and the clamp. Outside the active center, at least one functional surface is also conserved. The dock domain and its surrounding regions are highly similar in *P.fu* RNAP and RNAP II, reflecting a conserved interaction with the initiation factor TFB/TFIIB (Bushnell et al., 2004; Chen and Hahn, 2003; Kostrewa et al., 2009; Liu et al., 2009). No density was present at the locations for the RNAP II subunits Rpb8 and Rpb9 (Figures 3D and F), consistent with a lack of archaeal homologues for these subunits (Table 3).

To explain the observed differences between the EM map and the RNAP II structure, we constructed a RNAP II-based homology model for *P.fu* RNAP (Figure 4). In the model, the regions of highest sequence conservation cluster around the active center cleft. Many peripheral regions are also conserved, and only several surface domains are divergent. In particular, the archaeal subunit A', which is homologous to the N-terminal part of the largest RNAP II subunit Rpb1, contains a deletion in the clamp head, an extended pore domain, and it largely lacks the foot domain (Figures 3F and 4). In subunit A'', which is homologous to the C-terminal part of Rpb1, several loops in the jaw domain are shorter, and the C-terminal repeat domain is missing. The C-terminus of A' and the N-terminus of A'' form residual structure in the region of the Rpb1 foot (Figures 3F and 4). Subunit B contains an eight-residue insertion in the Rpb2 protrusion domain (Figures 3F and 4).

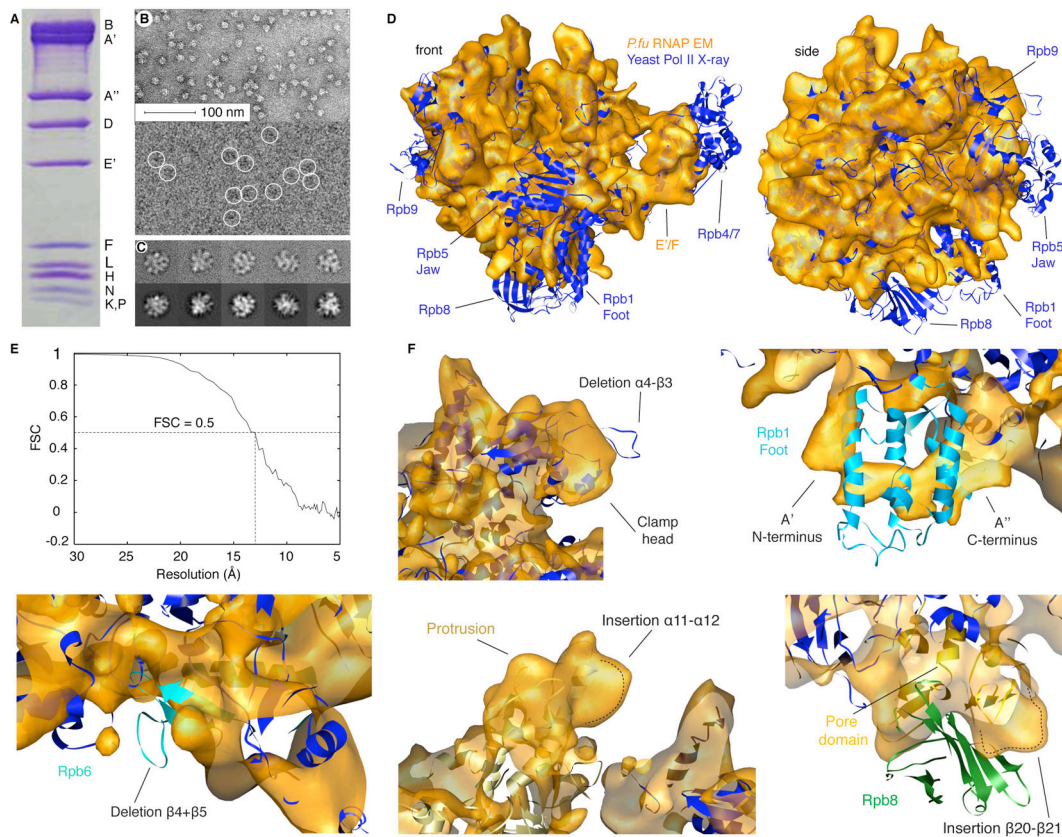


Figure 3: Cryo-EM structure of *P.fu* RNAP.

(A) SDS-PAGE analysis of purified endogenous *P.fu* RNAP.

(B) Negative stain image of *P.fu* RNAP particles (top), and cryo-EM micrograph with particles indicated by circles (bottom).

(C) Five representative projection averages (out of a total of 83, top) and corresponding re-projections (bottom).

(D) Cryo-EM density of *P.fu* RNAP (orange) with fitted X-ray structure of *S.cerevisiae* RNAP II (blue ribbon model, Armache et al., 2005). The views are from the front (left) and from the side (right, Cramer et al., 2001).

(E) Fourier shell correlation (FSC) function curve. The resolution is estimated to be around 13 Å based on the 0.5 FSC cut-off criterion.

(F) Gallery of regions of the *P.fu* RNAP structure deviating from yeast RNAP II (compare Figure 4). Depicted domains in RNAP II are highlighted using the original RNAP II colour code (Armache et al., 2005; Cramer et al., 2001).

Among the small archaeal subunits, D and L resemble the RNAP counterparts Rpb3 and Rpb11, respectively, except that several surface loops are missing, including the zinc loop in Rpb3 (Figure 4). Subunit H lacks the N-terminal jaw domain present in its eukaryotic counterpart Rpb5 (Figure 3D). Subunit K lacks the N-terminal region of the counterpart Rpb6, and the last two β -strands in the C-terminal assembly domain of Rpb6 (Figure 3F). Consequently, subunit K closely resembles the bacterial RNAP subunit

homolog ω (Minakhin et al., 2001). The heterodimeric subcomplex F/E' is situated at the expected surface position that is occupied by its counterpart Rpb4/7 in RNAP II (Figure 3D). However, EM density is largely lacking for the outermost domains of the subcomplex F/E', the OB and HRDC domains (Figure 3D), indicating their mobility. Indeed, the OB domain is the most flexible region of RNAP II according to normal mode analysis, and the HRDC domain is mobile in RNAP III (Fernandez-Tornero et al., 2007; Jasiak et al., 2006).

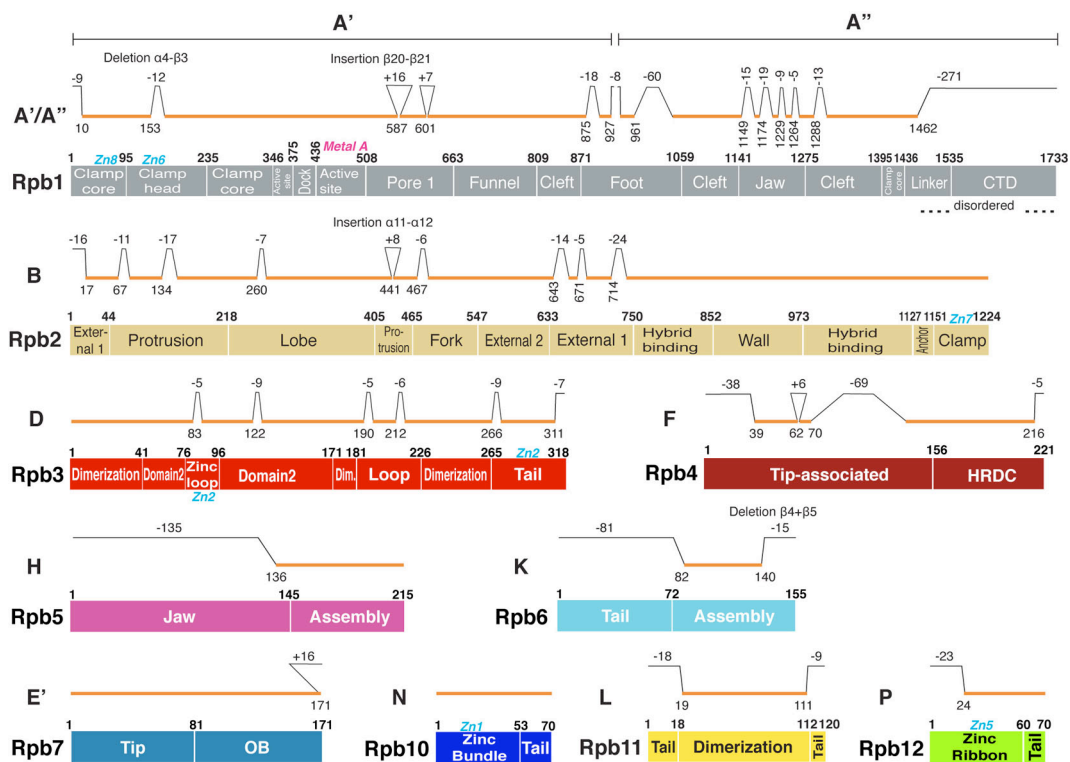


Figure 4: Conservation of RNAP II regions in *P.fu* RNAP.

The domain organization of RNAP subunits is shown as diagrams compared to RNAP II (Cramer et al., 2001). An orange bar above the diagrams indicates conserved regions. Insertions and deletions exceeding four amino acid residues are depicted. Yeast RNAP II subunits Rpb8 and Rpb9 have no homologue in *P.fu* RNAP (Table 3).

2.2 Conclusions

The first structure of an archaeal RNAP identified similarities to the related structure of eukaryotic RNAP II, but also revealed unique deviations.

The main finding is that the archaeal enzyme can largely be regarded as a truncated version of RNAP II. *P.fu* RNAP mainly lacks parts that are peripheral in RNAP II, including two small subunits and several surface domains and loops. Finally, our results provide the basis for structural studies of archaeal RNAP complexes that should provide significant mechanistic insights into eukaryotic transcription as well.

3 Outlook

With a good model of the free *P.fu* RNAP at hand, the next obvious steps will be to investigate functional complexes of RNAP and its regulatory factors. Again, the rather simplistic archaeal system should be beneficial for the reconstitution of active and physiologically relevant complexes and thus for obtaining results that provide insights into the basics and the evolution of the transcription mechanism (Hirata and Murakami, 2009). A candidate of high interest is of course the archaeal minimal pre-initiation complex that has, however, already shown to be difficult to tackle (compare Appendix, paragraph 1). Additionally, TFE or TFS could be incorporated into the mPIC, in the hope that either of them would further stabilize the interaction network between the components. Recent insights into eukaryotic PIC architecture (Kostrewa et al., 2009; Liu et al., 2009) and into transcription initiation dynamics (Patrick Cramer, personal communication) will guide optimization of constructs and nucleic acid scaffolds to finally overcome the obvious issue of intrinsic flexibility. If the problem remains, mild cross-linking could be tested to reduce any putative compositional heterogeneity (Kastner et al., 2008).

A presumably more stable assembly might be the complex of RNAP with NusG, an elongation factor that is exceptional for its unique conservation in all three domains of life (Hirtreiter et al., 2010; Zhou et al., 2009). First attempts towards solving the architecture of an archaeal RNAP-NusG complex are already underway (Patrick Cramer, personal communication).

4 Experimental procedures

Purification of *P.fu* RNA polymerase was initially supervised by Dr. Michela Bertero.

4.1 Measurement of protein concentration

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

4.2 Purification of *Pyrococcus furiosus* RNA polymerase

Pyrococcus furiosus RNA polymerase was purified from its endogenous source, resulting in approximately 5 mg of pure protein per purification that contained all subunits in an apparently stoichiometric manner.

4.2.1 Chemicals and buffers used for purification

100 x protease inhibitor (PI)

1.42 mg Leupeptin

6.85 mg Pepstatin A

850 mg PMSF

1650 mg benzamidine

add dry ethanol to 50 ml

store at $-20\text{ }^{\circ}\text{C}$; add immediately before use

Lysis buffer

50 mM Tris-HCl pH 8.0
100 mM KCl
10 mM MgCl₂
20% Glycerol
10 mM β-mercaptoethanol
1x PI

Biorex100 buffer

50 mM Tris-HCl pH 8.0
100 mM KCl
2.5 mM MgCl₂
1 mM EDTA
10% Glycerol
10 mM β-mercaptoethanol
1x PI

Biorex1000 buffer

50 mM Tris-HCl pH 8.0
1 M KCl
2.5 mM MgCl₂
1 mM EDTA
10% Glycerol
10 mM β-mercaptoethanol
1x PI

Dialysis buffer

50 mM Tris-HCl pH 8.0
100 mM KCl
2.5 mM MgCl₂
10% Glycerol
10 mM β-mercaptoethanol

Heparin100 buffer

50 mM Tris-HCl pH 7.5
100 mM KCl
2.5 mM MgCl₂
10% Glycerol
1 mM β-mercaptoethanol

Heparin1000 buffer

50 mM Tris-HCl pH 7.5
1 M KCl
2.5 mM MgCl₂
10% Glycerol
1 mM β-mercaptoethanol

MonoQ0 buffer

50 mM Tris-HCl pH 8.0
100 mM KCl
2.5 mM MgCl₂
10% Glycerol
10 mM β-mercaptoethanol

MonoQ1000 buffer

50 mM Tris-HCl pH 8.0
1 M KCl
2.5 mM MgCl₂
10% Glycerol
10 mM β-mercaptoethanol

Sup150 buffer

10 mM Hepes pH 7.0

150 mM KCl

2.5 mM MgCl₂

5mM DTT

4.2.2. Purification of endogenous *P.fu* RNAP

Pyrococcus furiosus cells were obtained frozen (-80 °C) from the Archaeenzentrum, University of Regensburg. To thaw the cells, 60g were resuspended in 200 ml lysis buffer at 4°C over night, while stirring. To disrupt the cells an Emulsiflex (Avestin Emulsiflex C5) was used applying 2 pressure cycles at 1000 - 1500 bar. Cell lysates were applied via a sieve to prevent clogging of the Emulsiflex tubes. To ensure cell lysis, a small aliquot was checked by phase-contrast microscopy. The lysate was centrifuged (30 min, 10,000 rpm, Sorvall SLA-1500) and then ultracentrifuged (1 h 30 min, 40,000 rpm, Beckman Ti-45), recovering the supernatant after both steps.

The final supernatant was loaded on a Biorex column (GE Healthcare XK50) using a peristaltic pump at 2-3 ml/min. The Biorex column was packed with ~400 ml Biorex resin that was previously equilibrated with Biorex100 buffer. The loaded column was washed with 1000 ml of Biorex100 buffer to remove protein that was bound non-specifically. The *P.fu* RNAP was eluted with 1500 ml Biorex1000 buffer at 1 ml/min using a KCl gradient from 0.1 M to 1 M. *P.fu* RNAP eluted between 55% and 80% Biorex1000 buffer. The fractions were checked by NTCA (non specific transcriptional activity) test. The Biorex column was finally washed with of Biorex1000 buffer, then 2M KCl, H₂O and finally 20 % Ethanol using 1000 ml in each case. When treated like this, the resin lasts for 3-4 purifications.

The fractions containing *P.fu* RNAP were pooled, dialysed for 4 h against 2 l Dialysis buffer while stirring, then loaded at 2 ml/min on a Heparin column (20 ml HiPrep 16/10 Heparin FF) that was pre-equilibrated with Heparin100

buffer. The polymerase was eluted with 200 ml Heparin1000 buffer with a gradient from 0.1 M to 1 M KCl. The fractions were checked by NTCA test. *P.fu* RNAP eluted between 17% and 42% Heparin1000 buffer. The Heparin column was subsequently washed with 100 ml of pure Heparin1000 buffer.

The fractions containing *P.fu* RNAP were centrifuged (30 min, 12,000 rpm, Sorvall SLA-1500), diluted to a final salt concentration of 150 mM KCl and loaded at 1 ml/min onto a MonoQ Column (MonoQ 10/100, GE Healthcare) that was previously equilibrated with MonoQ0 buffer, collecting the flowthrough. The column was washed with 90% MonoQ1000 buffer collecting the wash-flowthrough. The two flowthrough fractions were dialysed against 10% MonoQ1000 buffer and loaded to a MonoQ Column (MonoQ 10/100, GE Healthcare) that was prepared as described above. The column was then washed with 20 ml of 20% MonoQ1000 buffer and eluted with a gradient from 15% to 40% MonoQ1000 buffer. *P.fu* RNAP was eluted at 1 ml/min with MonoQ1000 buffer using a 160 ml gradient from 0.15 M to 1 M KCl. *P.fu* RNAP eluted between 25% and 30% MonoQ1000 buffer. Fractions containing *P.fu* RNAP were pooled and concentrated to a final concentration of 1 mg/ml. For long-term storage (2-3 months at -80°C), Glycerol was added to a final concentration of 20% and samples were flash frozen in liquid nitrogen.

Approximately 1 mg of *P.fu* RNAP at a concentration of 1 mg/ml were centrifuged for 15 min at 12,000 rpm (Sorvall SLA-1500) and loaded to a Superose 6 size exclusion column (GE Healthcare) pre-equilibrated in Sup150 buffer and eluted at 0.3 ml/min. The fractions containing *P.fu* RNAP were checked by 15% SDS-PAGE for purity and all the bands were double checked by mass spectrometry.

4.3 Cryo-electron microscopic structure determination

Purified *P.fu* RNAP was diluted to ~0.1 mg/ml using Sup150 buffer and was applied to freshly glow discharged carbon coated holey grids (Quantifoil

R3/3 + 2nm Carbon on top). Negative-stain samples were treated with 2% uranyl acetate and subsequently dried at RT. Cryo samples were flash-frozen in liquid ethane using a semi-automated controlled-environment system (Vitrobot, FEI Company) at 4° C, 95% humidity and stored in liquid nitrogen until transfer to the microscope.

Micrographs were recorded on Kodak SO-163 Electron film under low-dose conditions of ~ 25 electrons/Å² on a Tecnai Polara F30 field emission gun microscope at 300 kV using a pre-exposure of 100 ms. Negative film images were scanned on a Heidelberg drum scanner resulting in a pixel size of 1.23 Å on the object scale.

The contrast transfer function was determined using CTFFIND and SPIDER (Frank et al., 1996). Particles were picked semi-automatically with SIGNATURE (Chen and Grigorieff, 2007) and verified visually. Data were processed using SPIDER. Initially 41,669 particles images from 23 micrographs were decimated three-fold by pixel binning and aligned to reference projections of a version of the complete RNAP II structure (Armache et al., 2005) that was modified by deleting the clamp (residues 10-340 of subunit Rpb1) and the foot (residues 871-1059 of subunit Rpb1) and that was Gaussian filtered to a resolution of 25 Å. During early refinement, density for the clamp as well as for a weaker foot reappeared, substantiating the fact that the obtained volumes represented experimental data and not a reproduction of the reference. Particles were sorted into subsets according to different clamp conformations by providing two reference volumes: the reconstruction obtained from all particles and a manually created “collapsed-clamp” version of the complete RNAP II (Armache et al., 2005; Kuhn et al., 2007). All particles were aligned against each one of the references and a cross-correlation value was calculated for each particle and each corresponding reference projection. Particles were assigned to the reference volume bearing the higher cross-correlation value. Particle groups were then back-projected separately and the resulting reconstructions were used as input references for the next round of sorting. This procedure was repeated until the distribution of particles converged and resulted in 22,240 particles with an open clamp

conformation (class I) and 19,429 particles with other conformations (Vassylyev et al., 2002). At this stage the resolution for the class I volume was estimated to 16.8 Å based on a cut-off value of 0.5 for the Fourier shell correlation. The class I volume was further refined to a resolution of 14.5 Å using two-times decimated particle images. The final back-projection was performed in real space using the BP RP algorithm of SPIDER and non-decimated particle images. Using the command TH M in SPIDER, a mask of the obtained volume was generated and used for final calculation of the resolution estimate of 13 Å.

Chapter III

Molecular basis of RNA polymerase III transcription repression by Maf1

"I believe that we cannot live better than in seeking to become better."

(Socrates)

1 Introduction

The eukaryotic nuclear genome is transcribed by the multisubunit enzymes RNA polymerase (RNAP or Pol) I, II and III, which catalyze DNA-dependent RNA synthesis. Pol III transcribes genes encoding short, untranslated RNAs such as tRNAs, 5S rRNA, the spliceosomal U6 snRNA, the signal recognition particle 7SL RNA, and short regulatory RNAs. Pol III genes are essential in all cells and involved in fundamental processes such as ribosome and protein biogenesis, RNA processing, and protein transport. Pol III transcription is tightly co-regulated with Pol I activity, accounting together for up to 80% of nuclear gene transcription in growing cells (Grummt, 2003; Paule and White, 2000; Willis et al., 2004).

Pol III, with its 17 subunits and nearly 700 kDa mass, is the most complex nuclear RNA polymerase (Schramm and Hernandez, 2002). Five subunits, Rpb5, 6, 8, 10, and 12, are common to Pol I, II and III. Subunits AC40 and AC19 are common to Pol I and III, and homologous to Pol II subunits Rpb3 and Rpb11, respectively. The two largest Pol III subunits C160 and C128 are homologous to Pol II subunits Rpb1 and Rpb2, respectively, and encompass the active center of the enzyme. Subunits C17 and C25 form a heterodimeric subcomplex with homology to the Pol II subcomplex Rpb4/7 (Ferri et al., 2000; Jasiak et al., 2006; Sadhale and Woychik, 1994). Subunit C11 shares homology with Pol II subunit Rpb9. The Pol III-specific subunits C82, C53, C37, C34 and C31 form two subcomplexes. The C53/37 subcomplex shows limited homology to the Pol II initiation factor TFIIF and is involved in promoter opening, elongation, correct termination and re-initiation (Carter and Drouin, 2010; Cramer et al., 2008; Kassavetis et al., 2010; Landrieux et al., 2006). The subcomplex C82/34/31 is involved in promoter recognition and initiation. C34 interacts with TFIIB, the initiation factor that recruits Pol III to promoters (Thuillier et al., 1995; Wang and Roeder, 1997; Werner et al., 1993) and plays a subsequent role in open complex formation (Brun et al., 1997). Because of the central role of Pol III transcripts in basal cellular processes, the level of Pol III transcription is a critical determinant of

cell growth. In yeast, the ability to rapidly shut off synthesis of tRNAs and rRNAs during environmental stress conditions ensures cell survival (Warner, 1999). Stress conditions lead to Pol III repression by Maf1, a phosphoprotein that is phylogenetically conserved from yeast to human (Pluta et al., 2001; Upadhyaya et al., 2002). Maf1-dependent Pol III repression occurs during DNA damage, oxidative stress, growth to stationary phase, treatment with rapamycin and chlorpromazine, and upon blocking the secretory pathway (Upadhyaya et al., 2002; Willis et al., 2004).

In growing yeast, Maf1 is phosphorylated and localized mainly in the cytoplasm. Stress conditions lead to rapid Maf1 dephosphorylation and import into the nucleus (Oficjalska-Pham et al., 2006; Roberts et al., 2006). Nuclear import of yeast Maf1 is directed by two independent nuclear localization signal (NLS) sequences and phosphorylation impairs nuclear import (Lee et al., 2009; Moir et al., 2006). Once in the nucleus, dephosphorylated Maf1 binds to Pol III, blocking its interaction with TFIIIB, thus preventing recruitment of Pol III to promoters (Desai et al., 2005; Moir et al., 2006; Roberts et al., 2006). Maf1 also inhibits TFIIIB assembly at Pol III promoters by binding Brf1, a subunit of TFIIIB that resembles TFIIIB in its N-terminal half but also contains a Pol III-specific C-terminal domain (Desai et al., 2005). Maf1 leads to reduced genome-wide occupancy of Pol III genes by Brf1 and Pol III (Oficjalska-Pham et al., 2006; Roberts et al., 2006). Similar results have been obtained in human cells, thus establishing Maf1 as a conserved global repressor of Pol III transcription (Reina et al., 2006).

In this study we elucidate the mechanism of Pol III repression by Maf1. To date, structural information on Pol III is limited to a cryo-electron microscopic (cryo-EM) map that revealed the general location of the two Pol III-specific subcomplexes (Fernandez-Tornero et al., 2007), a homology model for the 10-subunit core enzyme, and the crystal structure of C25/17 (Jasiak et al., 2006). Here we report cryo-EM structures of Pol III in its free form and in complex with a DNA-RNA scaffold, assign the locations of Pol III subunits, present the Maf1 crystal structure, and combine the resulting information with a cryo-EM structure of a Pol III-Maf1 complex. Together with

functional studies, these results suggest the mechanism of Pol III repression by Maf1.

2 Results and discussion

2.1 Pol III EM structure reveals C82/34/31 mobility

We established a large-scale purification protocol for tagged Pol III from the yeast *Saccharomyces cerevisiae* (Experimental procedures). Pure Pol III samples comprised all 17 subunits, were monodisperse, and appeared homogeneous in EM with negative stain (Figure 5B). We collected high-quality cryo-EM data after vitrification under native conditions. A reconstruction of Pol III from 20,480 single particles led to a map at 21 Å resolution (Figures 5E and 6A, Experimental procedures). The cryo-EM map showed good agreement with the previously published map (Fernandez-Tornero et al., 2007).

The core Pol III homology model plus the C25/17 crystal structure (Jasiak et al., 2006) or the 12-subunit Pol II crystal structure (Armache et al., 2005) could both be fitted unambiguously to the map (Figures 5E and 7). After this fitting, two extended additional densities remained, one on top of the clamp adjacent to C25/17, and one at the lobe near Rpb9 (Figure 5E). Densities at the lobe and clamp were attributed to subcomplexes C53/37 and C82/34/31, respectively (Fernandez-Tornero et al., 2007). The density at the lobe was fitted with a homology model of the C53/37 dimerization module based on the crystal structure of the A49/34.5 module in Pol I (Geiger et al., submitted) (Figure 8). Adjacent densities protruding towards the C160/C11 jaw and the funnel were explained by extensions from the dimerization module. The location of C53/37 agrees with the previously reported association of C53/37 with C11 (Chedin et al., 1998) and with the location of the dimerization domain of TFIIIF, the distant homologue of C53/37 (Cramer et al., 2008) on the Pol II lobe (Chen et al., 2010; Eichner et al., 2010). The additional density at the

clamp could only account for a small part of the 138 kDa subcomplex C82/34/31, indicating intrinsic flexibility (Figure 5E).

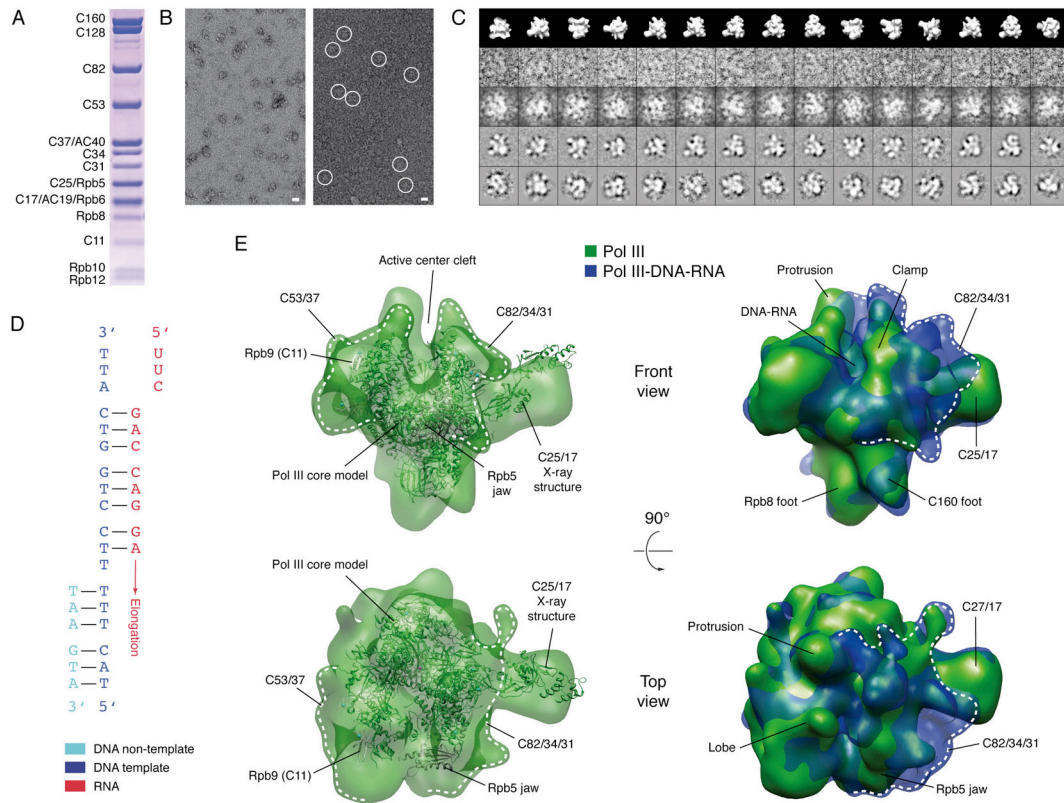


Figure 5: Cryo-EM reconstruction of Pol III and Pol III-DNA-RNA complexes

(A) SDS-PAGE gel of purified yeast Pol III. The 17 subunits are indicated. The identity of the bands was confirmed by mass spectrometry (not shown).

(B) EM micrographs of Pol III in negative stain (left) and vitrified ice (right). White circles indicate particles in vitrified ice. Scale bars are 10 nm in length.

(C) Different views of the Pol III reconstruction (first row) with corresponding raw single particle images (second row), low-pass filtered single particle images (middle row), class averages (forth row) and reference free averages (bottom row).

(D) DNA-RNA scaffold used in the Pol III-DNA-RNA complex (Brueckner et al., 2007).

(E) Cryo-EM reconstruction of Pol III (green) and Pol III-DNA-RNA complex (blue). The 10-subunit Pol III core homology model and the C25/17 crystal structure (Jasiak et al., 2006) were fitted to the map and are shown as ribbon models.

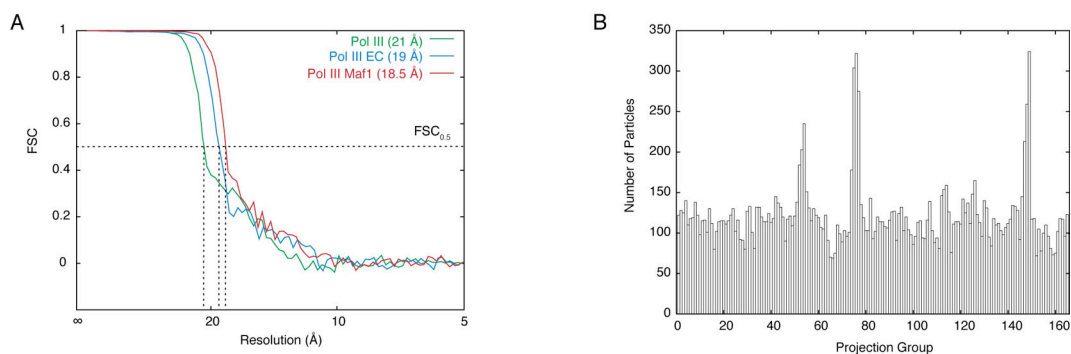


Figure 6. Single particle analysis of RNA Pol III

(A) Fourier shell correlation (FSC) function plot. Reported resolution is based on a cut-off value of FSC = 0.5 (dashed horizontal line).

(B) Distribution of particles among 166 different projection groups, sampled over the eularian angular space with equal distances. Overrepresented outliers were limited to prevent predominant views.

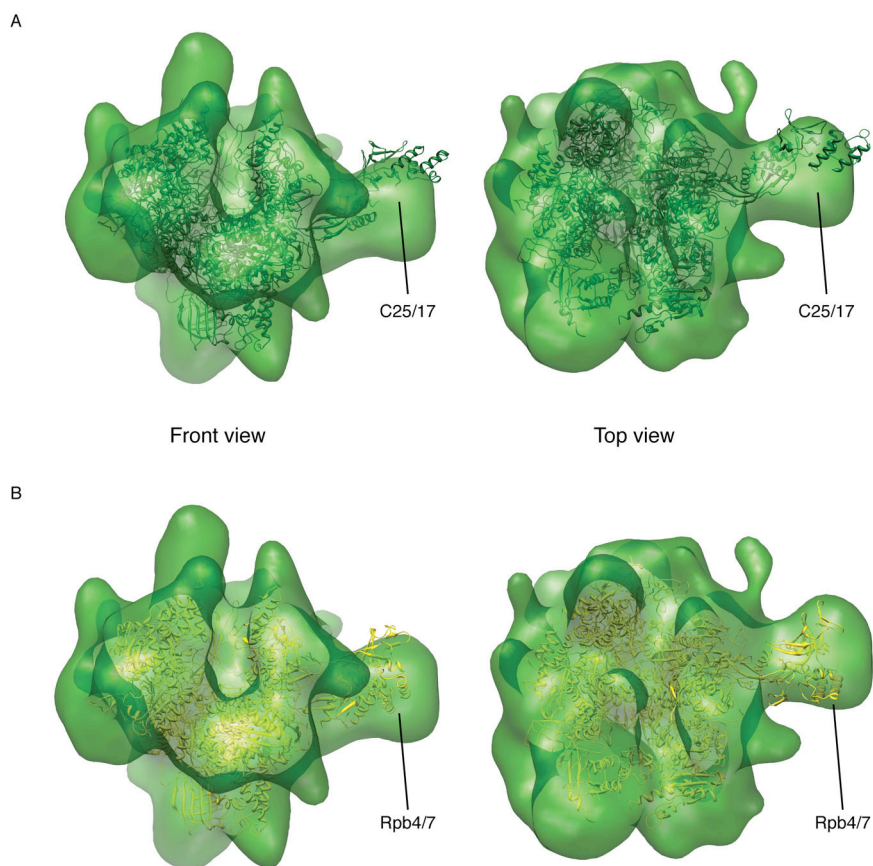


Figure 7: Fit of Pol III homology model and Pol II crystal structure to the cryo-EM map

(A) Fit of the 12-subunit Pol III homology model (green ribbon model, Jasiak et al., 2006) to the free Pol III EM map (green surface) reveals that the HDRC domain of the C17/25 subcomplex in solution adopts a conformation different to the one observed in the crystal structure.

(B) Fit of the 12-subunit RNA Pol II crystal structure (yellow ribbon model, Armache et al., 2005). The Rpb4/7 subunits (paralogues of Pol III subunits C17/25), as observed in the Pol II crystal structure, are more consistent with the cryo EM map, suggesting that the HDRC domain of C25/17 is indeed mobile.

2.2 Nucleic acid binding restricts C82/34/31 mobility

To see how nucleic acid binding influences the Pol III structure, we determined the cryo-EM structure of a Pol III complex with a minimal DNA-RNA scaffold (Figure 5D, Experimental Procedures). This complex mimics an active elongation complex (Brueckner et al., 2007). A reconstruction at 19 Å resolution was obtained from 11,965 single particles (Figures 5E and 6, Experimental Procedures). As expected, the reconstruction revealed density for nucleic acids in the cleft, but also a structural ordering of the C82/34/31 subcomplex, giving rise to an extended density between the top of the clamp, the Rpb5 jaw, and C25/17 (Figure 5E), revealing all the mass of the 138 kDa subcomplex.

A long continuous density between the clamp and the jaw (Figure 5E) could be fitted with the crystal structure of the human C82 homologue (Sebastien Fribourg, unpublished data) (Figure 8A). A prominent density remained between the clamp and the protrusion, forming a suspension over the cleft (Figures 5E, 8A and B). We assigned this density to subunit C34 since its two lobes fitted the structures of two winged helix (WH) domains in the N-terminal region of C34 (PDB codes 2dk5, 2dk8), and since C34 cross-links to promoter DNA around position -21 (Bartholomew et al., 1993) that is nearby in the homologous Pol II closed and open promoter complex models (Kostrewa et al., 2009). The remaining globular density near zinc site Zn8 in C160 between the clamp and C25/17 (Figure 8) was assigned to C31 since this position explains the known interactions with subunits C160, C82, C34, and C17 (Chedin et al., 1998; Geiduschek and Kassavetis, 2001; Schramm and Hernandez, 2002), the requirement of the zinc site for C82/34/31 binding (Werner et al., 1992), and association of C31 with Pol III after dissociation of the C82/34 heterodimer (Lorenzen et al., 2007). We note that alternatively one C34 domain could be placed into the density assigned to C31 (Figure 9), but this assignment could not explain the published biochemical data. Taken together, we could assign all Pol III subunits to EM densities in a way that is consistent with known interactions.

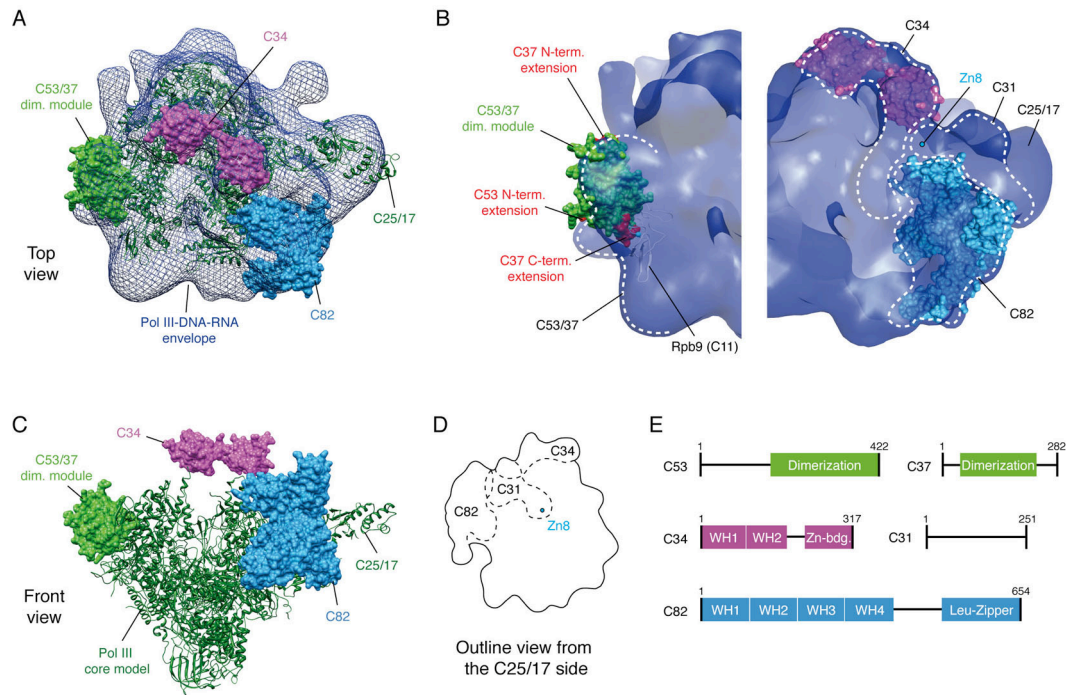


Figure 8: Subunit architecture of Pol III

(A) Pol III-specific subunits placed into the cryo-EM envelope of the Pol III-DNA-RNA complex. A homology model of the C53/37 dimerization domain (green, Geiger and Cramer, unpublished), the human C82 homolog crystal structure (cyan, S. Fribourg, unpublished) and the two C34 WH domain crystal structures (magenta) are shown as molecular surfaces. The 10-subunit Pol III core homology model and the C25/17 crystal structure (Jasiak et al., 2006) are shown as green ribbon models.

(B) Close-up view of Pol III-specific subunits fitted in the cryo-EM envelope of the Pol III-DNA-RNA complex. Terminal extensions of the C53/37 dimerization module are highlighted in red.

(C) Localization of Pol III-specific subunits on the core homology model and the C25/17 crystal structure (Jasiak et al., 2006).

(D) Schematic representation of C82/34/31 subcomplex organization on Pol III.

(E) Domain organization of Pol III specific subunits. Indicated domains were either revealed by homology modelling (C53, C37), X-ray crystallography (C34), or HHPred and secondary structure prediction (C82).

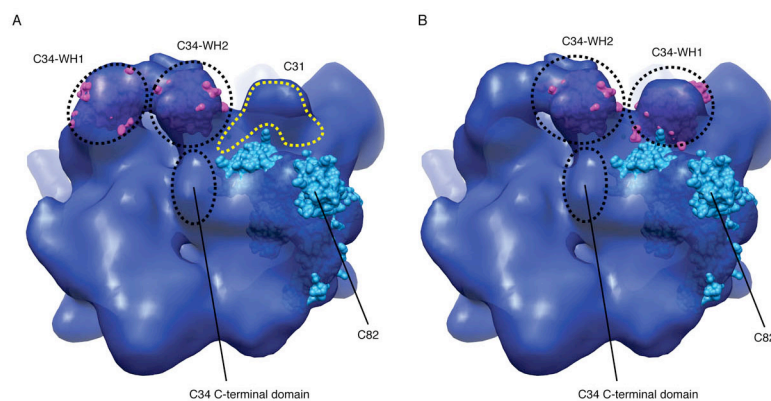


Figure 9: Location of the C82/34/31 supcomplex

(A) Subunits C82 (human C62, S. Fribourg, personal communication) and C34 (two N-terminal winged helix domains) are displayed as cyan and magenta molecular surfaces, respectively, and fitted in the Pol III-DNA-RNA cryo-EM map, as described in the text. With this fit, C31 (indicated by a yellow dashed line) could be placed in the globular density on top of the N-terminal C160 Zn-binding site.

(B) An alternative model for fitting the C34 winged helix domains (magenta molecular surface representation) is presented. As described in the text, this model seems less probable.

2.3 Maf1 structure is globular, not modular

To elucidate Pol III repression by Maf1, we determined the Maf1 structure by X-ray crystallography. We subjected recombinant *S. cerevisiae* and human Maf1 to limited proteolysis (Experimental procedures). Two flexible regions were revealed, a mobile insertion (human residues 50-82, yeast residues 50-224), and an acidic C-terminal tail (Figure 10). Based on cleavage sites and secondary structure predictions we constructed Maf1 variants for crystallization trials. Crystals were obtained for a human variant that lacked both mobile regions. The structure was solved by bromide phasing and refined to a free R-factor of 21.2% at 1.55 Å resolution (Table 4, Experimental procedures).

Data set		NaBr soak		Native
<i>Data collection</i>				
Space group		P212121		P212121
Unit cell axis <i>a, b, c</i> (Å)		48.07, 48.34, 80.50		48.39, 48.81, 79.32
	Peak	Remote	Inflection	
Wavelength (nm)	0.9196	0.9211	0.9200	0.91870
Resolution (Å)	26.83-1.9	26.83-1.9	26.83-1.9	25.974-1.55
R _{meas} (%)	7.7 (50.7)	6.0 (39.2)	7.0 (51.3)	5.2 (75.7)
I/σ (I)	22.0 (2.5)	22.7 (3.0)	22.0 (2.4)	22.3 (1.2)
Completeness (%)	99.0 (99.5)	98.8 (99.4)	98.9 (99.5)	90.4 (87.4)
Redundancy	3.9 (4.0)	3.8 (3.9)	3.8 (4.0)	3.2 (1.9)
<i>Refinement</i>				
Resolution (Å)				1.55-25.97
No. reflections				26,137
R _{work} (%)				18.81
R _{free} (%)				21.15
No. atoms				
Protein				1313
Water				142
B-factors (Å ²)				
Protein				33.64
Water				43.95
r.m.s.d. from ideal				
Bond lengths (Å)				0.006
Bond angles (°)				0.959

Table 4: Maf1 X-ray diffraction and refinement statistics

Maf1 forms a globular structure with a central five-stranded antiparallel β -sheet that is flanked by a single helix on one side, and by three helices on the other (Figure 10B). The Maf1 fold is frequently found in proteins, but not in proteins involved in transcription, as revealed by DALI (Holm and Park, 2000; Krissinel and Henrick, 2004). The Maf1 structure is apparently conserved among eukaryotes, since hydrophobic core residues are conserved from yeast to human (Figure 10A). The Maf1 structure shows that the previously defined conserved sequence boxes A, B, and C (Desai et al., 2005; Pluta et al., 2001; Reina et al., 2006), do not correspond to structural modules or defined surface patches (Figure 10C). Thus, previous functional analyses of Maf1 deletion constructs, lacking one or more of these boxes, must be re-evaluated taking into account that the variants may adopt non-native structures.

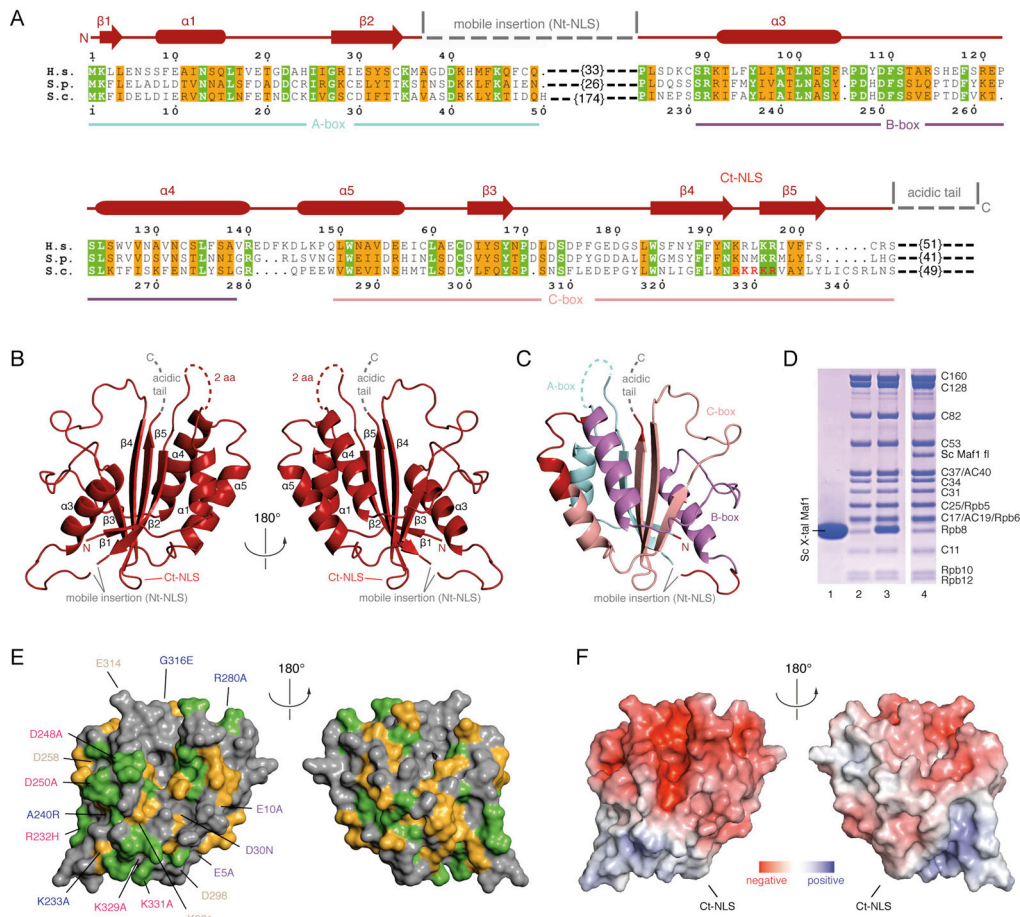


Figure 10: Maf1 crystal structure

(A) Multiple sequence alignment and structural conservation of Maf1 from *Homo sapiens* (H.s.), *Schizosaccharomyces pombe* (S.p.) and *Saccharomyces cerevisiae* (S.c.). Secondary structure elements are indicated above the sequence (cylinders for α -helices, arrows for β -strands). Red elements are included in the structure. Identical and conserved residues are highlighted in green and orange, respectively. The mobile region includes proteolytic cleavage sites (this work), phosphorylation sites (Dephoure et al., 2008; Lee et al., 2009; Moir et al., 2006), and the N-terminal NLS (Nt-NLS). The C-terminal NLS (Ct-NLS) is also indicated. Regions that are not present in the crystal structure are indicated by dashed lines. The crystallized protein was a human Maf1 variant comprising residues 1-35 and 83-205.

(B) Two views of a ribbon model of the human Maf1 crystal structure. Secondary structure elements are labelled according to A.

(C) Ribbon model of human Maf1 with the conserved A-, B- and C-boxes highlighted in blue, purple, and rose, respectively.

(D) Purification of Pol III-Maf1 complexes. 200 μ g of Pol III and a 5-fold molar excess of full-length yeast Maf1 or the yeast Maf1 variant comprising residues 1-35 and 225-345 (lane 1, equivalent to the crystallized human protein) were incubated for 20 min at 20°C, loaded on a Superose 6 gel filtration column and analysed on SDS-PAGE gels. Lane 2 shows Pol III, lane 3 the Pol III complex with the Maf1 variant, and lane 4 shows the Pol III complex with full-length Maf1.

(E) Surface conservation of Maf1. Identical and conserved residues are highlighted in green and yellow, respectively. Residues labelled in blue were analyzed for a role in Pol III interaction (Figure 11). Residues labelled in red, pink, and wheat show severe, mild, or no phenotypes, respectively (Dephoure et al., 2008; Moir et al., 2006; Roberts et al., 2006).

(F) Surface charge distribution of Maf1. Red, blue, and white areas indicate negative, positive, and neutral charges, respectively.

2.4 Regulated Maf1 localization

The Maf1 crystal structure reveals that the two NLS sequences (yeast residues 205-208 and 328-332; Moir et al., 2006) are accessible on the domain surface (Figure 10B). The C-terminal NLS (Ct-NLS) is located between strands β 4 and β 5, and the N-terminal NLS (Nt-NLS) is part of the directly adjacent mobile region (Figure 10B). The adjacent location suggests that phosphorylation of the mobile insertion regulates nuclear localization by masking the NLS sequences (Lee et al., 2009; Moir et al., 2006). This mechanism is apparently conserved from yeast to human, although the exact phosphorylation sites within the mobile insertion differ (Dephore et al., 2008; Lee et al., 2009; Moir et al., 2006; Shor et al., 2010). The Ct-NLS and the residues at the mobile insertion form the only positively charged region on Maf1 (Figure 10F). Point mutants that led to defects in phosphorylation, growth on glycerol at 37°, or Pol III repression (Moir et al., 2006; Roberts et al., 2006) (Figure 10E), cluster in this surface region. Consistent with a role of this region in phosphorylation-regulated nuclear import, point mutations in this region did not abolish Pol III binding (Figure 11).

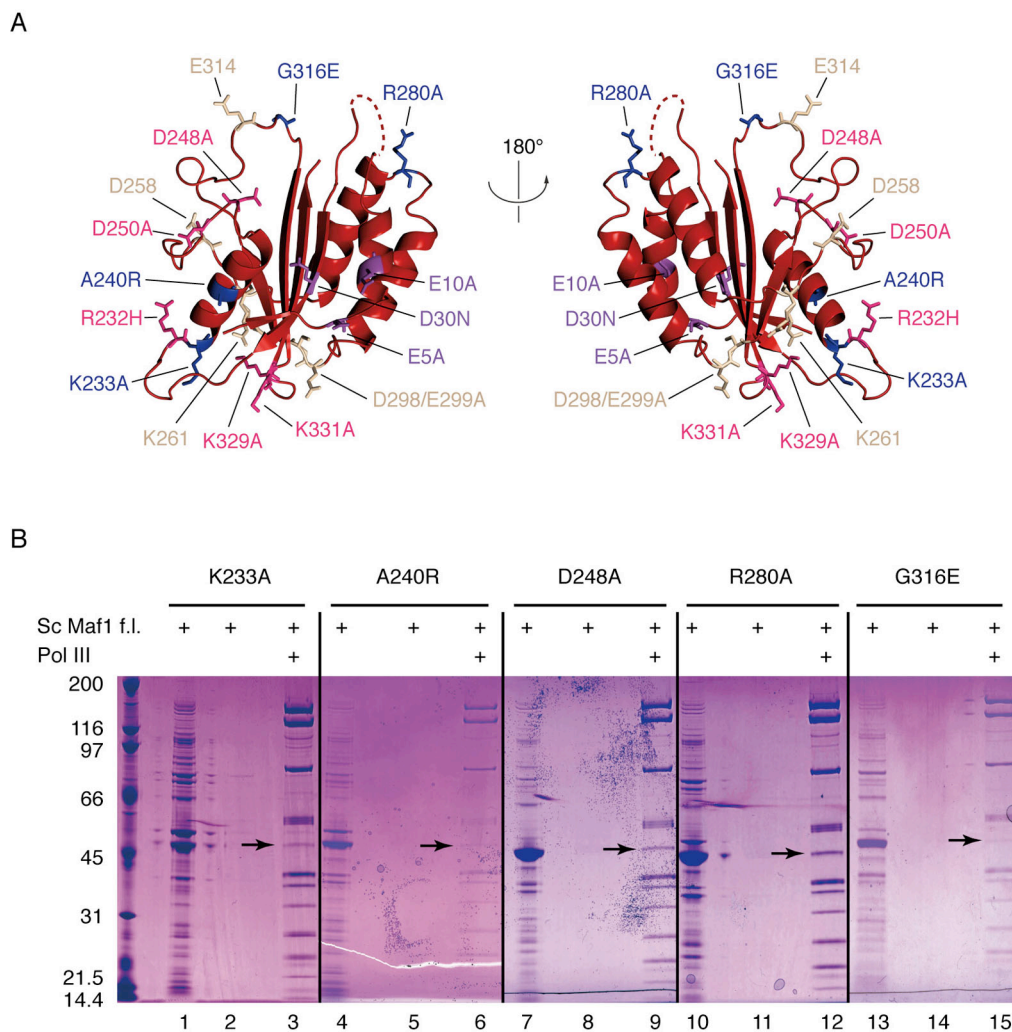


Figure 11: Maf1 mutational analysis

(A) Ribbon model of Maf1 structure (residues 1-35;83-205). Investigated residues are shown in ball and stick representation. Residues labeled in blue were analyzed by pulldowns in this study and had no phenotypic effect. Other residues were described in literature (Roberts et al., 2006; Moir et al., 2006): residues depicted in red showed a severe phenotype in vivo while the other labeled residues showed mild (pink) or no phenotypic effect (wheat) in vivo.

(B) Binding properties of point mutants. Described point mutants of *S.cerevisiae* Maf1 full length (Sc Maf1 f.l.) were analyzed in pulldowns for its binding ability to Pol III. Therefore 15 μ g of Pol III were preincubated with 5 molar excess of Maf1 mutants in buffer M and loaded on Ni-NTA beads, washed with 30 mM imidazol and eluted with 500mM imidazol in buffer M. Gels show partly purified Maf1 mutants used for binding (lanes 1, 4, 7, 10 and 13), negative control of unspecific Maf1 binding to the resin (lanes 2, 5, 8, 11 and 14) and the eluted fractions, indicating that all point mutants can still bind Pol III (lanes 3, 6, 9, 12 and 15).

2.5 Maf1 binds the Pol III clamp and rearranges C82/34/31

To investigate how Maf1 binds yeast Pol III, we prepared full-length yeast Maf1 as a recombinant protein and a variant that lacked both mobile regions and corresponded to the crystallized human protein. Both variants formed a complex with Pol III that could be purified by size-exclusion chromatography (Figure 10D, lanes 3 and 4). Maf1 binding was specific, as human Maf1 did not bind yeast Pol III (not shown). Thus, the two mobile regions are not required for Pol III binding, and the human Maf1 crystal structure is relevant for the understanding of the Pol III-Maf1 interaction. We collected cryo-EM data of the pure Pol III-Maf1 full-length complex and used 16,974 particles to obtain a reconstruction at 18.5 Å resolution (Figures 6 and 12, Experimental procedures). The structure revealed a continuous density for C82/34/31, similar to the density in the Pol III-DNA-RNA complex (Figure 12C).

Maf1 was assigned to a new density on top of the clamp, with the help of calculated difference maps. The Maf1 X-ray structure fitted this density well (Figures 12A and C). To provide additional support for the Maf1 location, we labelled the C-terminal hexahistidine tags on Maf1 and the Pol III subunit C128 with Ni-NTA-Nanogold™ and located the labels by 2D cryo-EM image analysis (Experimental procedures). The locations of the labels were consistent with Maf1 binding on top of the clamp domain (Figure 12B). This location also agreed with published biochemical and genetic interactions of Maf1 with the N-terminal region of C160 that forms most of the clamp (Boguta et al., 1997; Oficjalska-Pham et al., 2006; Reina et al., 2006) (Figure 12D). Further consistent with this location, C160, C82, and C34 are the top interacting partners of Maf1 in the yeast interactome (Gavin et al., 2006).

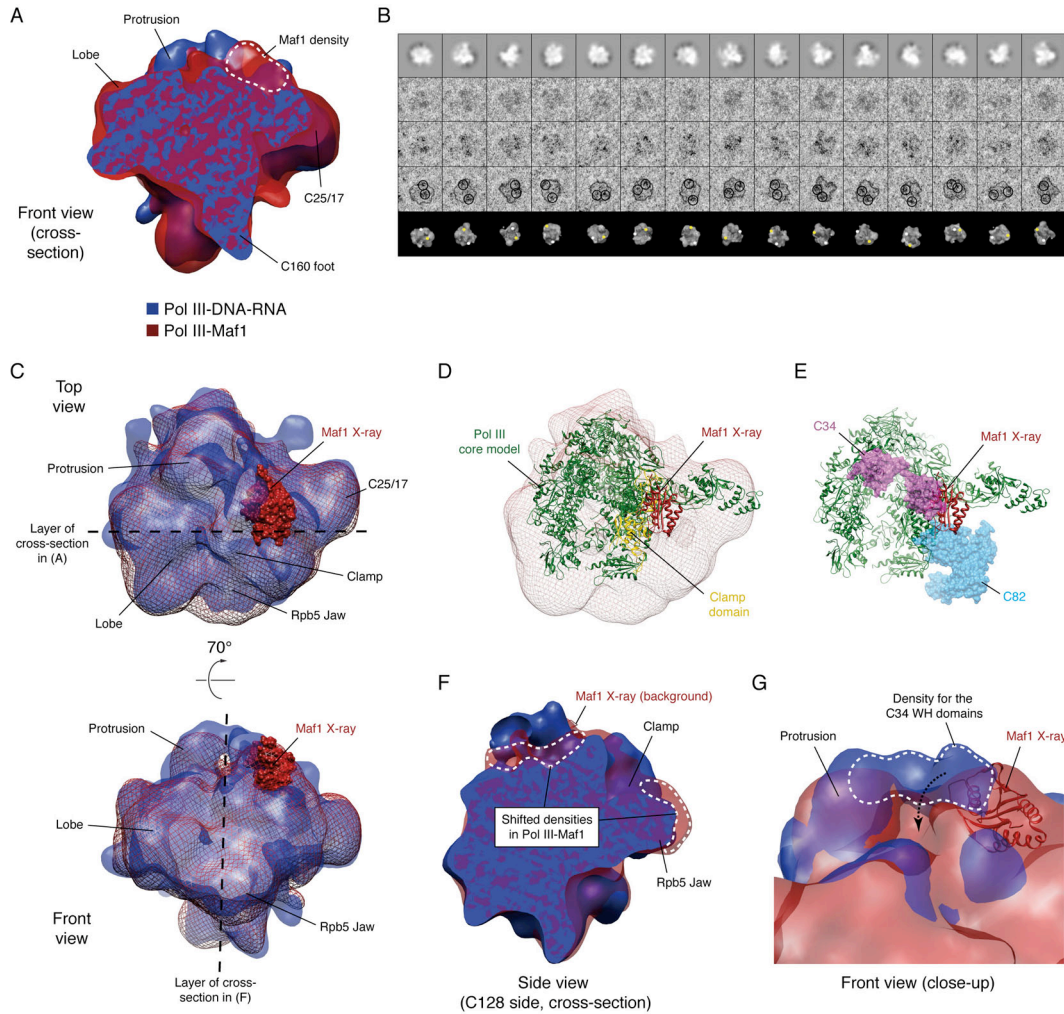


Figure 12: Cryo-EM structure of the Pol III-Maf1 complex

(A) Comparison of cross sections of EM structures of the Pol III-DNA-RNA complex (blue) and the Pol III-Maf1 complex (red) reveals an additional density for Maf1.

(B) Different views of reference projections of the Pol III-Maf1 reconstructions (top row), corresponding thresholded Nanogold-labeled Pol III-Maf1 particles used for alignment (second row), raw Nanogold-labeled particles (middle row), Nanogold particles with outline of Pol III-Maf1 structure and circles indicating the location of the Nanogold labels and the expected variability (fourth row, compare text), and surface representations of the Pol III-Maf1 reconstructions with the N-terminus of C128 and the location of Maf1 indicated by white and yellow dots, respectively (bottom row). The Nanogold signals are consistent with the location of Maf1 shown in A.

(C) Fit of the Maf1 X-ray structure (red molecular surface) to the Pol III-Maf1 EM map (red mesh). The cryo-EM map of the Pol III-DNA-RNA complex is also shown (blue surface).

(D) Ribbon representation of the Pol III-Maf1 complex. The Pol III homology model is depicted in green while the Maf1 X-ray structure is depicted in red. The clamp domain of C160 (res. 1-245) is highlighted in yellow. The Pol III-Maf1 cryo-EM map is shown as red mesh.

(E) Maf1 (red ribbon) sterically clashes with C34 (purple) and C82 (cyan) as positioned in the Pol III-DNA-RNA complex.

(F) Comparison of cross sections of EM structures of the Pol III-DNA-RNA complex (blue) and the Pol III-Maf1 complex (red) reveals a shift of the density attributed to the C82/34/31 subcomplex upon Maf1 binding.

(G) Close-up view of the region over the clamp. Most of the density attributed to the two C34 WH domains in the Pol III-DNA-RNA complex (blue) is absent in the Pol III-Maf1 complex (red).

A detailed comparison of the EM structures revealed that the C82/34/31 density observed in the Pol III-Maf1 complex differed from that in the Pol III-DNA-RNA complex. In particular, it appears that most of the density assigned to the C34 WH domains in the Pol III-DNA-RNA complex is absent in the Pol III-Maf1 complex, as an effect of a Maf1-dependent displacement of these domains. Probably these domains are shifted and become partially mobile, as suggested by the presence of a shifted residual density, attributed to the second C34 WH domain (Figure 12F). The densities assigned to C31 and C82 undergo a similar change in location towards the Rpb5 jaw domain, giving rise to additional density in this region (Figures 12C and F). Consistently, Maf1 overlaps with the assigned locations of the C34 second WH domain and with C82 and C31 in the Pol III-DNA-RNA complex (Figures 12C and E).

2.6 Maf1 impairs closed promoter complex formation

To analyze how the structural changes induced by Maf1 binding could repress Pol III transcription, we constructed a model for the Pol III-Brf1-TBP closed promoter complex. Brf1 resembles the Pol II initiation factor TFIIB in its N-terminal region, but contains a specific C-terminal extension that binds TBP (Figure 13; Khoo et al., 1994). We therefore combined the Pol II-TFIIB-TBP closed promoter complex model (Kostrewa et al., 2009) with the structure of TBP bound to the Brf1 C-terminal residues 437-507 (Juo et al., 2003). Comparison of the resulting Pol III closed promoter complex model with the EM densities revealed that C34 was well positioned for interacting with both the Brf1 N- and C-terminal regions (Figure 14A), consistent with both regions interacting with C34 (Andrau et al., 1999; Brun et al., 1997; Khoo et al., 1994). In the Pol III-Maf1 complex, C34 is locked in a different position that is apparently incompatible with Brf1 interaction, suggesting that Maf1 impairs Pol III recruitment to Brf1-containing promoters (Figures 14A and B).

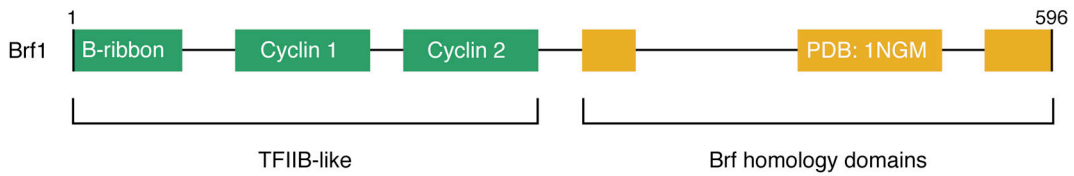


Figure 13. Brf1 domain organization

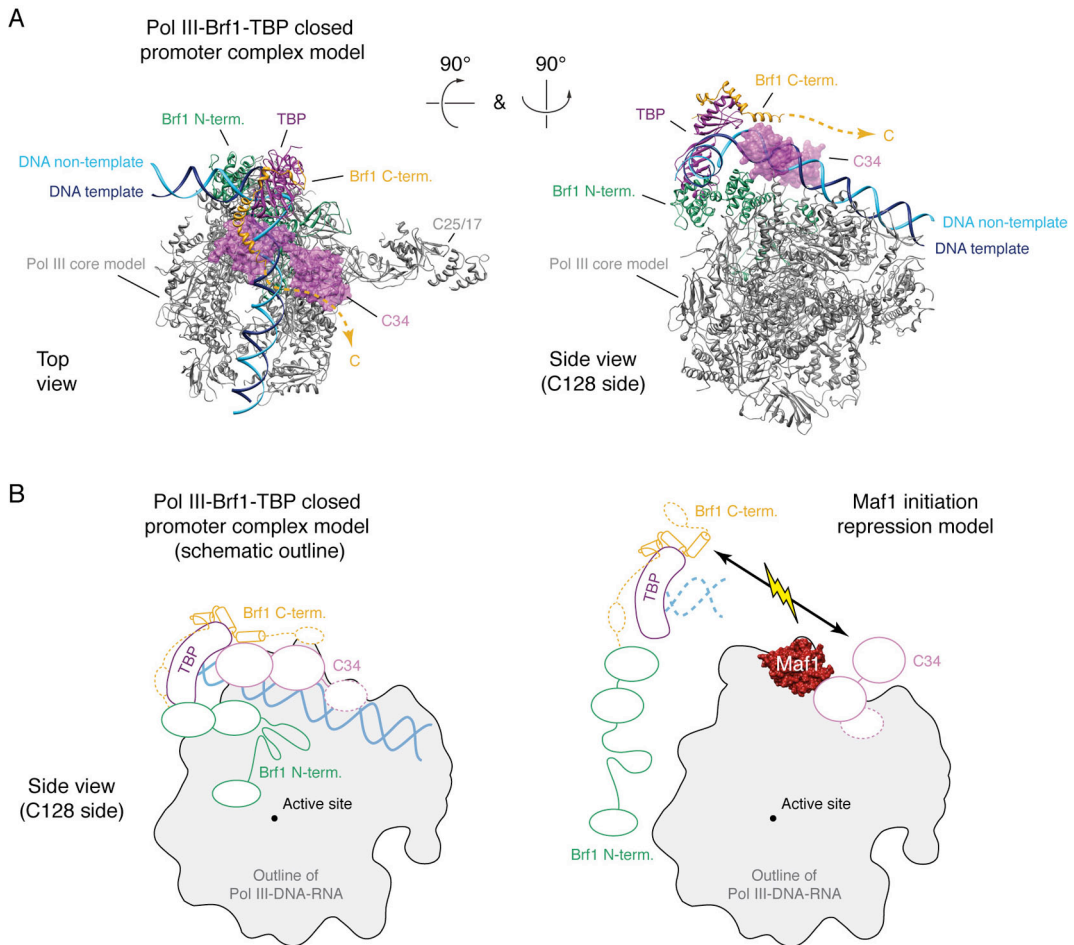


Figure 14: Mechanism of Pol III transcription repression by Maf1

(A) Model of Pol III-Brf1-TBP-DNA closed promoter complex. The Pol III core homology model and C25/17 crystal structure (Jasiak et al., 2006) are depicted as gray ribbons. The C34 WH domains are depicted as magenta surfaces. The Brf1 N-terminal domain (green ribbon) and the closed promoter DNA (cyan and blue ribbons) are based on the Pol II-TFIIB-TBP-DNA closed promoter complex model (Kostrewa et al., 2009). TBP (dark purple) and the Brf1 C-terminal domain (orange) are depicted as ribbons and are based on the Brf1-TBP-DNA crystal structure (Juo et al., 2003).

(B) Schematic model for Maf1-dependent repression of Pol III-Brf1-TBP-DNA closed promoter complex formation, coloured as in A.

To test this model for the Maf1 repression mechanism, we investigated by size-exclusion chromatography whether the Pol III-Maf1 complex can bind to a pre-assembled, transcriptionally functional, Brf1-TBP fusion protein-DNA promoter complex (Kassavetis et al., 2005). We used U6 snRNA promoter DNA from position -40 to +20 relative to the transcription start site +1. Whereas free Pol III formed a stable closed promoter complex (Figure 15E, lane 3), the Pol III-Maf1 complex did not bind the Brf1-TBP-DNA complex, even when a five-fold molar excess was used (lane 5). When we repeated the experiment with a mismatched bubble region at positions -11 to +2 (Figure 15A), the same result was obtained (Figure 15E, lanes 6 and 7). Further, pre-assembled Pol III-Brf1-TBP promoter complexes were unable to bind Maf1, even when a five-fold molar excess was used (lane 4). These experiments indicate that the interaction of Pol III with Maf1 and a Brf1-TBP-DNA complex are mutually exclusive, showing that Maf1 impairs formation of a closed promoter complex. This is consistent with evidence that Maf1 binds Pol III and prevents promoter interaction (Desai et al., 2005; Moir et al., 2006; Roberts et al., 2006).

2.7 Maf1 does not inhibit Pol III activity

The above model predicts that Maf1 inhibits binding of promoter DNA over the active center cleft but not in the cleft. To test this, we compared pure Pol III and Pol III-Maf1 complexes in an initiation factor-independent transcription assay using a 3'-tailed DNA template and a priming RNA dinucleotide (Bardeleben et al., 1994, Experimental procedures) that bind directly in the cleft. Consistent with the model, both complexes were equally active in RNA synthesis, and an excess of Maf1 or DNA did not change activity (Figure 15C). We additionally performed RNA extension assays using a minimal DNA-RNA scaffold (Damsma and Cramer, 2009). The presence of Maf1 neither prevented scaffold binding nor elongation to the end of the

template, and this was independent of the order of factor addition (Figure 15B).

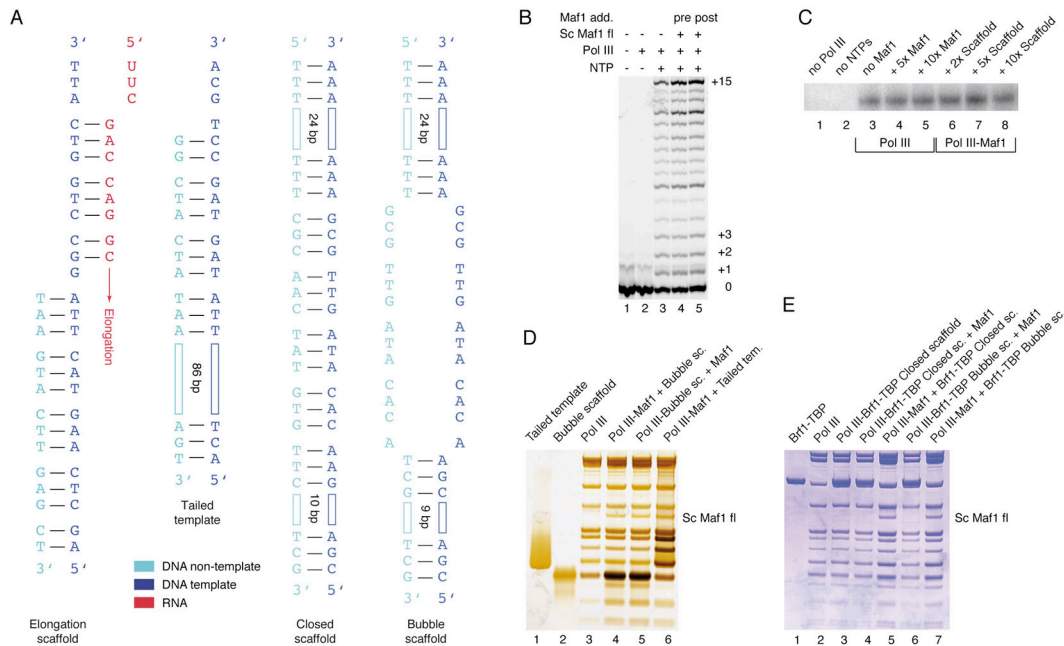


Figure 15: Maf1 impairs closed promoter complex formation but not Pol III activity

(A) Nucleic acid scaffolds.

(B) RNA extension assay with the minimal DNA-RNA elongation scaffold shown in A.

(C) Pol III factor-independent transcription assay with the tailed template scaffold shown in A.

(D) Competition assays show that binding of Maf1 and nucleic acids to Pol III are not mutually exclusive. Pol III-Maf1 and Pol III-DNA complexes were incubated as indicated with 5-fold molar excesses of DNA or Maf1, respectively, loaded onto a Superose 6 gel-filtration column, and the peak fraction was analyzed by silver-staining. Staining of a Pol III-Maf1 complex (without DNA) is identical to that in lanes 4, 5 and 6 (data not shown).

(E) Competition assays reveal that Maf1 impairs binding of Pol III to a Brf1-TBP-DNA complex. Pre-assembled Pol III-Brf1-TBP-DNA or Pol III-Maf1 complexes were incubated with 5-fold molar excesses of competing factor or complex as indicated, loaded onto a gel-filtration Superose 6 column (Superose 6 10/300 GL, GE Healthcare), and the resulting peak fraction was analyzed by SDS-PAGE. In lanes 3, 4 and 6, the presence of DNA was deduced by the high A260/A280 ratio (~ 1) compared to the A260/280 ratio (~ 0.6) in lanes 2, 5 and 7.

To rule out that nucleic acids displace Maf1 from Pol III or prevent its binding, we tested by size-exclusion chromatography whether Pol III is capable of binding Maf1 and nucleic acids simultaneously. Pol III-Maf1 complexes with 3'-tailed template or bubble scaffold could be purified, independent of the order of addition (Figure 15D). Thus, Maf1 neither prevents nucleic acid binding in the active center cleft nor DNA-dependent RNA synthesis. The observation that Pol III can simultaneously bind Maf1 and

nucleic acids suggests that the increased Maf1 occupancy at Pol III genes under repressive conditions (Geiduschek and Kassavetis, 2006; Oficjalska-Pham et al., 2006; Roberts et al., 2006) is due to Maf1 binding to Pol III elongation complexes. Such Maf1-containing Pol III elongation complexes would be unable to re-initiate, explaining the observation that Maf1 represses multiple-round but not single-round transcription by Pol III (Cabart et al., 2008).

3 Conclusions and outlook

Our results converge with published data on the mechanism of Pol III-specific transcription repression by Maf1. Cellular stress leads to dephosphorylation of a mobile surface region in Maf1 that unmask adjacent NLS sequences, leading to nuclear import of Maf1. In the nucleus, Maf1 binds Pol III at its clamp domain and rearranges the C82/34/31 subcomplex. This impairs Pol III binding to a TBP-Brf1-promoter complex and specifically abolishes initiation from Pol III promoters, which require Brf1. Maf1 also binds Pol III that is engaged in transcription elongation, leaving activity intact, but preventing re-initiation. Since Pol III genes are short and elongation is fast, this leads to rapid shutdown of all Pol III transcription.

The obtained insights into how transcription initiation in the Pol III system is repressed, together with the background information about the subunit arrangement in the free enzyme form a good basis to analyze a minimal Pol III PIC using cryo-EM. Recent discoveries about how promoter bubble design affects mPIC conformational variability should allow for additional stabilization of such complexes (Patrick Cramer, personal communication). Preliminary investigations on a Pol III mPIC have already been made (compare Appendix, paragraph 2) and can add some experience to the mixture. Again, it will be of critical importance to improve the resolution of the reconstructions by utilization of better instruments and optimization of conditions during sample preparation and data collection (see Appendix, paragraph 3). Dramatic

improvements in observed densities upon addition of nucleic acids in the case of the Pol III-DNA-RNA complex has demonstrated that a stable and conformationally well-defined assembly already bears high potential for a meaningful reconstruction. Remaining flexibility is most likely still restricting the possibilities to create models of better detail. Therefore, powerful classification methods need to be sought. These will rely on an ideal SNR to be able to classify single images, hopefully in an unsupervised, thus unbiased manner.

4 Experimental procedures

Purification of yeast RNA polymerase III and assembly of complexes for cryo-EM was done by Dr. Alessandro Vannini. Expression and purification as well as proteolytic cleavage, crystallization and structure determination of Maf1 was done by Rieke Ringel and partly supervised by Dr. Alessandro Vannini. Biochemical assays were conducted by Rieke Ringel and Dr. Alessandro Vannini.

4.1 Purification of yeast RNA polymerase III

Yeast RNA polymerase (Pol) III was purified from its endogenous source, resulting in approximately 1 mg of pure protein per purification.

4.1.1 Chemicals and buffers used for Pol III purification

Pol III lysis buffer

200 mM Tris-HCl pH 8.0

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

10 mM MgCl_2

10% Glycerol
10 mM β -mercaptoethanol
1 mM PMSF
1 mM benzamidine
200 mM pepstatin
60 mM leupeptin

Biorex0 buffer

40 mM Hepes pH 7.8
5 mM $MgCl_2$
10% Glycerol
1 mM EDTA
10 mM β -mercaptoethanol
1 mM PMSF
1 mM benzamidine
200 mM pepstatin
60 mM leupeptin

Biorex500 buffer

40 mM Hepes pH 7.8
500 mM KCl
5 mM $MgCl_2$
10% Glycerol
5 mM imidazole pH 8.0
1 mM EDTA
10 mM β -mercaptoethanol
1 mM PMSF
1 mM benzamidine
200 mM pepstatin
60 mM leupeptin

Ni-NTA-wash-A buffer

40 mM Hepes pH 7.8
500 mM KCl
5 mM MgCl₂
10% Glycerol
10 mM imidazole pH 8.0
1 mM EDTA
10 mM β-mercaptoethanol
1 mM PMSF
1 mM benzamidine
200 mM pepstatin
60 mM leupeptin

Ni-NTA-wash-B buffer

40 mM Hepes pH 7.8
5 mM MgCl₂
250 mM (NH₄)₂SO₄
10% Glycerol
10 mM imidazole
10 mM β-mercaptoethanol
1 mM PMSF
1 mM benzamidine
200 mM pepstatin
60 mM leupeptin

Ni-NTA-elution buffer

40 mM Hepes pH 7.8
5 mM MgCl₂
250 mM (NH₄)₂SO₄
10% Glycerol
250 mM imidazole
10 mM β-mercaptoethanol

1 mM PMSF
1 mM benzamidine
200 mM pepstatin
60 mM leupeptin

Pol III Heparin0 buffer

40 mM Hepes pH 7.8
5 mM MgCl₂
20% Glycerol
0.5 mM EDTA
10 mM β-mercaptoethanol
1 mM PMSF
1 mM benzamidine
200 mM pepstatin
60 mM leupeptin

Pol III Heparin1000 buffer

40 mM Hepes pH 7.8
5 mM MgCl₂
1000 mM (NH₄)₂SO₄
20% Glycerol
0.5 mM EDTA
10 mM β-mercaptoethanol
1 mM PMSF
1 mM benzamidine
200 mM pepstatin
60 mM leupeptin

Pol III MonoQ0 buffer

40 mM Hepes pH 7.8

1 mM MgCl₂

10% Glycerol

5 mM DTT

Pol III MonoQ1000 buffer

40 mM Hepes pH 7.8

1 mM MgCl₂

1000 mM (NH₄)₂SO₄

10% Glycerol

5 mM DTT

Sup50 buffer

20 mM Hepes pH 7.8

50 mM (NH₄)₂SO₄

0.1 mM MgCl₂

0.01 mM ZnCl₂

5 mM DTT

4.1.2 Purification of endogenous yeast Pol III

The *Saccharomyces cerevisiae* strain NZ16 (Lannutti et al., 1996), carrying the gene for an N-terminally His₆-FLAG₄-RET1-tagged C128 subunit on the parent plasmid pYE(CEN3)30 was grown to OD₆₀₀ = 6-7 at 30°C in YPD media in a 200 l fermenter (Infors ABEC). The cells were harvested by continuous flow centrifugation (Padberg) and lysed by bead beating in ice-cooled Pol III lysis buffer. All subsequent steps were performed at 4°C. Glass beads were separated by filtration prior to clearing the lysate by centrifugation (60 min, 8000 g, Sorvall SLA-1500). A whole cell extract was obtained after centrifugation at 125,000 g for 90 min (Beckman Ti45) by separating the clear

upper-middle phase from the turbid lower phase. The supernatant was further processed by step-wise ammonium sulfate precipitation. 35% $(\text{NH}_4)_2\text{SO}_4$ was slowly added to the sample, allowed to stir for 30 min and cleared by centrifugation (60 min, 8000 g, Sorvall SLA-1500). The supernatant was precipitated by adding 70% $(\text{NH}_4)_2\text{SO}_4$ and left stirring over night. The pellet was recovered by centrifugation (60 min, 8000 g, Sorvall SLA-1500) and resuspended in 3 l of Biorex0 buffer. The sample was applied to a 250 ml Biorex resin column (Biorad). Bound proteins were eluted with Biorex500 buffer. The eluting proteins were loaded onto a 12 ml Ni-NTA Agarose (Qiagen) column. Subsequent washing steps were performed with Ni-NTA-wash-A buffer and Ni-NTA-wash-B buffer. Proteins were eluted with Ni-NTA-elution buffer and loaded onto a HiTrap Heparin 5 ml column (GE Healthcare) and fractionated by applying a salt gradient from 250 mM to 1000 mM $(\text{NH}_4)_2\text{SO}_4$ with Pol III Heparin0 buffer and Pol III Heparin1000 buffer. The pooled Pol III-containing fractions (eluting at 500 mM $(\text{NH}_4)_2\text{SO}_4$) were diluted 5-fold with Pol III Heparin0 buffer, loaded onto a Mono Q anion exchange column (Mono Q 10/100 GL, GE Healthcare), and fractionated by applying a salt gradient from 50 mM to 1000 mM $(\text{NH}_4)_2\text{SO}_4$ with Pol III MonoQ0 buffer and Pol III MonoQ1000 buffer. The pooled Pol III-containing fractions (eluting at 600 mM $(\text{NH}_4)_2\text{SO}_4$) were diluted to reach a final concentration of 50 mM $(\text{NH}_4)_2\text{SO}_4$ supplemented with a 10-fold molar excess of recombinant full-length C53/37 heterodimer (preparation will be described elsewhere) and incubated for 60 min to overcome the endogenous sub-stoichiometry of this subcomplex (Lorenzen et al., 2007). The sample was concentrated to a volume of 1 ml using an Amicon Ultra-4 centrifugal filter unit (MWCO 10kDA, Millipore) and applied to gel-filtration chromatography on a Superose 6 column (Superose 6 10/300 GL, GE Healthcare) using Sup50 buffer. Pol III-containing fractions were pooled, concentrated to 1mg/ml using an Amicon Ultra-4 centrifugal filter unit (MWCO 10kDA, Millipore) and flash-frozen in liquid N_2 upon addition of 10% Glycerol.

4.2 Maf1 purification

4.2.1 Chemicals and buffers used for Maf1 purification

Maf1-lysis buffer

50 mM Hepes pH 7.8

0.5 M NaCl

10 mM Imidazole

5 mM MgCl₂

10 μM EDTA

10% Glycerol

10 mM β-mercaptoethanol

Maf1-Ni-NTA-loading buffer

50 mM Hepes pH 7.8

0.5 M NaCl

20 mM Imidazole

5 mM MgCl₂

10 μM EDTA

10% Glycerol

10 mM β-mercaptoethanol

Maf1-Ni-NTA-elution buffer

50 mM Hepes pH 7.8

0.5 M NaCl

300 mM Imidazole

5 mM MgCl₂

10 μM EDTA

10% Glycerol

10 mM β-mercaptoethanol

Maf1-MonoQ0 buffer

50 mM Hepes pH 7.8

5 mM MgCl₂

100 μM EDTA

10% Glycerol

10 mM β-mercaptoethanol

Maf1-MonoQ1000 buffer

50 mM Hepes pH 7.8

5 mM MgCl₂

1000 mM NaCl

100 μM EDTA

10% Glycerol

10 mM β-mercaptoethanol

Maf1-Superdex25 buffer

25 mM Hepes pH 7.0

25 mM NaCl

5 mM DTT

Maf1-Superdex40 buffer

50 mM Hepes 7.8

40 mM (NH₄)₂SO₄

100 μM MgCl₂

10 μM ZnCl₂

5 mM DTT

Maf1-Strep-loading buffer

100 mM TrisHCl 8.0

150 mM NaCl

1 mM EDTA

Maf1-Strep-elution buffer

100 mM TrisHCl 8.0

150 mM NaCl

1 mM EDTA

2.5 mM Desthiobiotin

4.2.2 Purification of Maf1

DNA encoding *S.cerevisiae* or human Maf1 was PCR-amplified from genomic DNA and cloned into pET-28b vector (Novagen) using the NdeI/NotI restriction sites, resulting in a N-terminal hexahistidine tag. *E.coli* BL21 (DE3) RIL cells (Stratagene) were transformed with the plasmid and grown in LB medium at 37°C to an OD₆₀₀ of 0.6. Expression was induced with 0.5 mM IPTG for 16 h at 18° C. Cells were lysed by sonification in Maf1-lysis buffer. After centrifugation the supernatant was loaded onto a 3 ml Ni-NTA column (Qiagen) equilibrated with Maf1-Ni-NTA-loading buffer. The column was washed with 20 column volumes (CVs) of Maf1-Ni-NTA-loading buffer and eluted with Maf1-Ni-NTA-elution buffer. Proteins were purified by anion exchange chromatography (Mono Q, GE healthcare). The column was equilibrated with Maf1-MonoQ0 buffer and proteins were eluted with a linear gradient of 20 CVs from 10 mM to 1 M NaCl using Maf1-MonoQ0 buffer and Maf1-MonoQ1000 buffer. After concentration, the sample was applied to a Superdex75 size exclusion column (GE Healthcare) equilibrated with Maf1-Superdex25 for crystallization experiments or Maf1-Superdex40 for binding experiments.

For purification of Strep-tagged *S. cerevisiae* Maf1 variants, a streptavidin tag was inserted via the 5' primer in the PCR reaction and the DNA was cloned into a pET21b vector (Novagen), resulting in a N-terminal streptavidin tag. Expression and purification was done as described above, except that a 1 ml Strep-Tactin Macro Prep (IBA) column was used instead of the Ni-NTA column. Avidin was added to a final concentration of 3 nmol per liter *E. coli*

culture to the supernatant and loaded to the Strep-Tactin column equilibrated with Maf1-Strep-loading buffer. The column was washed with 10 column volumes (CVs) of Maf1-Strep-loading buffer and eluted with Maf1-Strep-elution buffer. For purification of untagged *S.cerevisiae* Maf1 variants, DNA was cloned into a pET21b vector (Novagen). Expression and purification was done as described above, except that a 40% ammonium sulfate precipitation step was used instead of the Ni-NTA column. Maf1 was precipitated by the addition of saturated ammonium sulfate solution.

4.3 Proteolytic cleavage of Maf1

For trypsin and chymotrypsin treatment of purified protein samples, 100 μ l of purified Maf1 at a concentration of 1 mg/ml were mixed with 1 μ g of the corresponding protease. The reaction was carried out at 37° C in size-exclusion buffer supplemented with 1 mM CaCl₂. Aliquots of 10 μ l were taken at 1, 3, 5, 10, 30, and 60 min, and the reaction was stopped by addition of 5 \times SDS sample buffer and incubation at 95° C for 5 min. Samples were analyzed by SDS-PAGE. The N-termini of digestion products were analyzed by Edman sequencing.

4.4 Maf1 crystal structure determination

For crystallization, pure human Maf1 variant 1-35;83-205 was concentrated to 40 mg/ml. Crystals were grown at 20° C in hanging drops over a reservoir solution containing 50 mM MES pH 6.0 and 175 mM sodium oxalate. Crystals grew after 2 days and to a maximum length of 500 μ m. Native crystals were transferred into reservoir solution containing 25% Glycerol and were flash-cooled in liquid nitrogen. Crystals were soaked for 0.5-2 minutes in a reservoir solution containing 25% Glycerol and 0.5 M NaBr and flash-frozen in liquid nitrogen. Diffraction data were collected at 100 K on

a PILATUS 6M detector at the Swiss Light Source (SLS), Villigen, Switzerland (Table 4). A native data was collected to a resolution of 1.55 Å. Three-wavelength anomalous diffraction data were collected from a Bromide-labelled crystal to a resolution of 1.9 Å. Diffraction data were processed using MOSFLM (Leslie, 1986). Data scaling and merging was done with SCALA (Evans, 1993) and data quality was assessed with Phenix.Xtriage (Adams et al.). The program Phenix.HySS (Adams et al.) identified six bromide sites that were used for phasing with the program SOLVE (Terwilliger and Berendzen, 1999). Subsequently, density modification was carried out with RESOLVE (Terwilliger, 2003). The resulting electron density map was of very high quality and was used to build a model in COOT (Emsley and Cowtan, 2004). The resulting model was then refined against the native data set with the program Phenix.Refine (Adams et al.) to a free R factor of 21% (Table 4).

4.5 Cryo-EM structure determination

Purified Pol III complexes were diluted to ~0.1 mg/ml using Sup50 buffer and applied to glow-discharged pre-coated carbon holey grids (Quantifoil R3/3, 2nm carbon on top). Negative-stain samples were treated with 2% uranyl acetate and subsequently dried at room temperature. Cryo samples were flash-frozen in liquid ethane using a semi-automated controlled-environment system (Vitrobot, FEI Company) at 4° C, 95% humidity, and stored in liquid nitrogen until transfer to the microscope.

Micrographs were recorded under low dose conditions of ~15 electrons/Å² on a FEI Tecnai Spirit microscope operating at 120 kV, equipped with a LaB₆ filament and a Gatan side entry cryo-holder. Images were acquired at underfocus values in the range of 1.5-4 μm on a 2k x 2k FEI Eagle CCD camera applying a pre-exposure of 100 ms at a magnification of 90.000x resulting in a pixel size of 3.31 Å/px on the object scale. All subsequent

image-processing operations were carried out using the SPIDER software package, if not stated otherwise (Frank et al., 1996).

Initial particle selection and windowing was performed semi-automatically using the boxer program from the EMAN software package (Ludtke et al., 1999). Reference particles were picked manually on every micrograph to avoid discrepancies due to defocus and ice differences and the resulting automatically selected particles were verified visually. Windowed particles were aligned to 83 projections of the Pol II X-ray structure (1Y1W, gaussian low-pass filtered to 35 Å), which was modified by removing the OB and HRDC domains from subcomplex Rpb4/7 since they are deemed to conformational flexibility. Particle assignment to the reference projections was evenly distributed barring only few over-represented outliers that were limited to prevent predominant views (Figure 6). Backprojection of the particle images using the angles from reference-based alignment resulted in a first reconstruction that showed additional densities at the clamp and C25/17 regions and was then used as a reference for 20 rounds of angular refinement. The images were back-projected in real space using the refined angles. The resulting reconstruction was Gaussian low-pass filtered to 25 Å and used as reference for another round of alignment and refinement as described above. This procedure was iterated until no changes in growth of additional densities were observed.

A 21 Å reconstruction of Pol III from a data set of 20,480 particles was obtained (Figures 5 and 6). During early stage of the refinement process, density for a complete C25/17 complex and other additional densities appeared that could be confirmed by an independent 23 Å reconstruction from 12,174 particles (not shown). Projections of the 20,480 particle Pol III reconstruction, as well as their corresponding particle averages, were compared to averages resulting from a reference-free 2D alignment method using the program refine2d from the EMAN software package (Ludtke et al., 1999). A good portion of significantly similar averages supported the fact that the alignment and refinement procedure based on the Pol II X-ray structure was not reference-biased (Figure 5C). A cryo-EM data set of Pol III incubated

with a three-fold molar excess of a minimal DNA/RNA scaffold was collected and a reconstruction at 19 Å resolution was obtained with 11,965 particles (Figures 5E and 6A). Finally, an 18.5 Å reconstruction of a size-exclusion purified Pol III-Maf1 complex was obtained from 16,974 particles (Figures 12 and 6A).

4.6 Nanogold™ labelling

Size-exclusion purified Pol III was incubated for 60 min at 4° C in Maf1-lysis buffer with a 10-fold molar excess of recombinant full-length Maf1. The complex was then incubated with a 20-fold molar excess of Ni-NTA-Nanogold™ (Nanoprobes, INC, www.nanoprobes.com) particles for 30 min. The sample was concentrated to 1 ml using an Amicon Ultra-4 centrifugal filter unit (MWCO 10kDA, Millipore) and applied to gel-filtration chromatography on a Superose 6 column (Superose 6 10/300 GL, GE Healthcare) with Sup50 buffer. The Pol III-Maf1 Nanogold™ labelled containing fractions were pooled and prepared for cryo-EM as described above. Cryo-EM CCD data were collected as described for the free Pol III but at an underfocus range of 3-4 μm to obtain high image contrast. A large portion of particles showed both His-tags bound with Nanogold™ clusters. These were picked from the micrographs and aligned to projections of the Pol III-Maf1 reconstruction. The strong signal of the Nanogold™ was dampened in the images by manually applying a threshold to the image histograms. The in-plane rotation parameters resulting from the alignment were applied to the original images and the rotated images were compared to corresponding 2D surface views with the location of Maf1 and the N-terminus of C128 indicated (Figure 12B). The length of the hexa-histidine tag and the ~0.9 nm linker between the gold cluster and the nickel-nitrilotriacetic acid group of the Nanogold™ reagent give an expected mean variability of ~2 nm radius that was taken into account for interpretation of the Nanogold™ locations on the EM images. The gold signal on the N-terminus of C128 displayed more

apparent variability, which can be explained by the presence of four additional tandem FLAG sequences.

4.7 Initiation factor-independent transcription assay

1.5 pmol of Pol III or Pol III-Maf1 complex were incubated for 30 min at 20° C with 2 pmol (or variable amounts) of a pre-annealed tailed-template scaffold (Figure 15) in the presence of 0.5 mM GpG dinucleotide primer. For *in vitro* transcription, complexes were incubated for 30 min at 20° C in the presence of 0.3 mM NTPs (ATP, GTP, CTP and [α -³²P]-UTP) in 20 μ l reaction mixtures, containing 40 mM Tris-HCl pH 8.0, 60 mM NaCl, 7 mM MgCl₂, 7% Glycerol and 5 mM DTT. Reactions were stopped by addition of an equal volume of 2 \times loading buffer (8 M urea, 2 \times TBE) and incubation for 5 min at 95° C. The [α -³²P]-UTP-labelled RNA products were separated by denaturing gel electrophoresis (20 μ l of the stopped reaction mixtures, 1 mm 15% polyacrylamide gels containing 7 M urea) and visualized with a Typhoon 9400 phosphoimager (GE Healthcare).

4.8 RNA-extension assays

Five pmol of Pol III or Pol III pre-incubated (10 min at 20° C) with a five-fold molar excess of yeast Maf1 was incubated for 30 min at 20° C with 5 pmol of a pre-annealed minimal nucleic-acid scaffold (Figure 15). Maf1 was added at a five-fold molar excess followed by incubation for 5 min at 20° C. For RNA elongation, complexes were incubated for 10 min in the presence of 1 mM NTPs at 28° C in transcription buffer (60 mM ammonium sulfate, 20 mM Hepes at pH 7.6, 8 mM magnesium sulfate, 10 μ M zinc chloride, 10% Glycerol, and 10 mM DTT). Reactions were stopped by addition of an equal volume of 2 \times loading buffer (8 M urea, 2 \times TBE) and incubation for 5 min at 95° C. The FAM-labelled RNA extension products were separated by

denaturing gel electrophoresis (0.5 pmol RNA per lane, 0.4 mm 15%–20% polyacrylamide gels containing 8 M urea, 50° C), and visualized with a Typhoon 9400 phosphoimager (GE Healthcare).

Appendix: Unpublished results

Solving the architecture of the RNA polymerase pre-initiation complex (PIC) has been tried for over a decade now, but only little steps forward could be made so far (Cramer, 2007; Kornberg, 1999; Roeder, 1996). PIC complexes are highly transient; naturally, because in their physiological context, they represent a short lived intermediate state (Cramer, 2007). A PIC complex has a tremendously strong goal at its determination: to initiate transcription, with RNA polymerase trying to escape the promoter as well as its initiation factors and to engage in transcription elongation. In an *in vitro* assembled PIC, when there are no nucleotides present, it is conceivable that the complex oscillates between different states, urging to proceed. This hypothesis of the oscillating states has only recently been substantiated further by single molecule FRET studies of RNAP II mPIC complexes (Patrick Cramer, personal communication). When performing structural studies, this oscillation introduces additional conformational heterogeneity. However, the artificial situation of the complex being trapped in an intermediate state is imperative to study the assembly in the first place, to maintain it long enough as to carry out experiments. It has proven extremely complicated to find conditions that are the right trade-off between stabilizing the complex as much as it is needed for studies, while preserving it as natural as possible. It will, in all likelihood, be essential to combine available information, to implement all the resulting conclusions and to improve existing techniques to ultimately unravel the mechanism of transcription initiation and its regulation on a molecular level.

The following paragraphs are meant to document trials and accomplishments regarding mPIC architecture and optimization of methods during the time of this research. Parts of this Chapter present new results, others sum up or confirm what was already known before, as to provide a comprehensive overview for any potential reader new to the field of RNA polymerase structure determination by cryo-electron microscopy and single particle analysis.

1 *Pyrococcus furiosus* RNAP minimal pre-initiation complex studies

The *P.fu* PIC complexes that were studied during this work were minimal PICs consisting of *P.fu* RNAP, TBP, TFB and a short stretch of a promoter DNA scaffold containing an artificial bubble and a tailed overhang. Complex assembly done by Dr. Michela Bertero and Wenjie Ruan. Cloning, expression and purification of recombinant factors TBP and TFB was done by Dr. Michela Bertero and Wenjie Ruan and detailed protocols for these steps will be described elsewhere. A variant of TFB, the so-called “delta-B-finger” mutant was also probed together with a DNA/RNA scaffold (data not shown).

1.1 Sample preparation and cryo-EM data collection

A *P.fu* mPIC was investigated containing RNAP, TBP, TFB and an open bubble *gdh* promoter DNA scaffold with a short 5'-template (T) overhang as depicted below (TATA-region underlined).

Open bubble *gdh* scaffold:

```

                                     AAAAATGTATCGCC
5' ACCGAAAGCTTTATATAGGCTATTGCCCTTACACCTAATTTGGAG 3' (NT-DNA)
3' TGGCTTTCGAAATATATCCGATAACGGGTTAGTGGATTAAACCTCCCC 5' (T-DNA)
                                     AAAAATGTATCGCC

```

To assemble the mPIC complex, 2 fold excess of TBP was added to the DNA scaffold and mixed at 15°C, then 2 fold excess of TFB was added. After incubation at 15°C for 1 hour, purified *P.fu* RNAP was added such that the ratio between RNAP and DNA was 1:2, incubated at 15°C for another 1 hour, and then heated to 60°C and incubated for 15 minutes. The assembled mPIC complex was purified by size-exclusion chromatography on a Superose6 column (GE Healthcare) using Sup150 buffer as described in the experimental procedures section of Chapter II. The peak fractions showed an A260/A280 ratio close to 1, a clear indication for binding of nucleic acids (Figure 16). Fractions were collected, concentrated to ~1 mg/ml, supplied with 20%

Glycerol and flash frozen in liquid nitrogen for transfer to the MPI Berlin, where data was collected at that time. The samples were sent there in dry ice, thawed gently at RT upon arrival, diluted to ~ 0.1 mg/ml and applied to freshly glow discharged carbon coated grids (Quantifoil R3/3 + 2nm Carbon on top). After 45 seconds of incubation, excess of buffer was blotted for 10 seconds using a semi-automated controlled-environment system (Vitrobot, FEI Company) at 4°C, 95% humidity. Samples were then flash-frozen by plunging them into liquid ethane and subsequently stored in liquid nitrogen until transfer to the microscope.

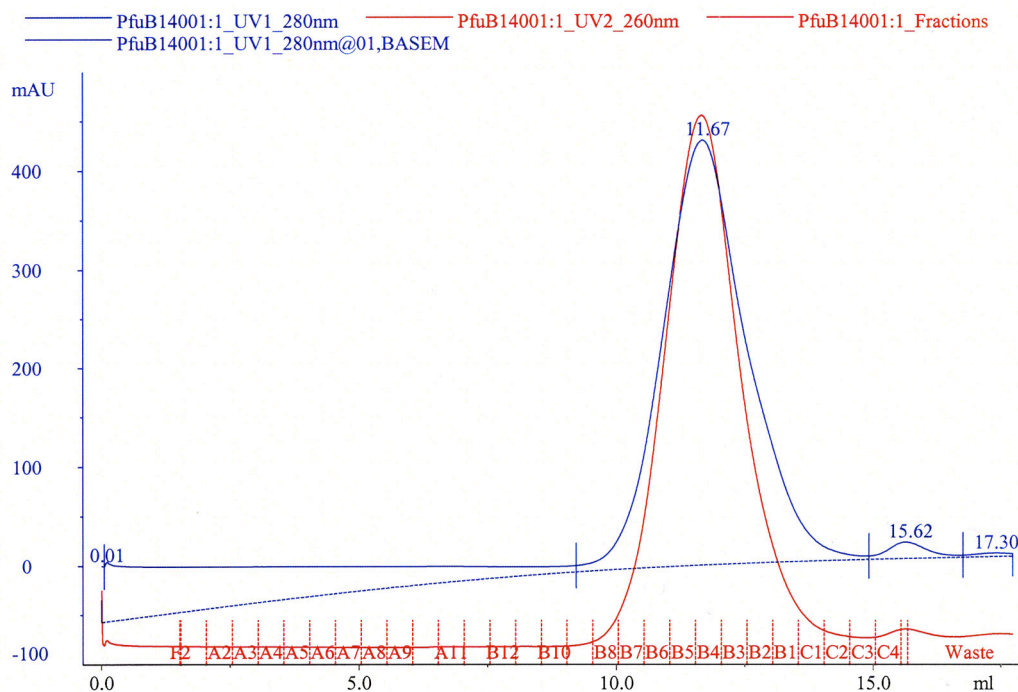


Figure 16: Superose6 size-exclusion chromatography profile of *P.fu* mPIC

Data were collected under low dose conditions of ~ 20 electrons/ \AA^2 at a Tecnai Polara F30 field emission gun microscope operating at 300 kV. Micrographs were recorded at underfocus values of 1.5-3.5 μm on Kodak SO-163 Electron film applying a pre-exposure of 100 ms at a magnification of 67.000x.

1.2 Cryo-EM data processing and classification trials

Negative film images were developed and scanned on a Heidelberg drum scanner resulting in a pixel size of 1.23 Å/px on the object scale. The contrast transfer function was determined using CTFFIND and SPIDER (Frank et al., 1996). Initially, 125,159 particles images from 56 micrographs were picked semi-automatically with SIGNATURE (Chen and Grigorieff, 2007) and verified visually. Data were further processed using SPIDER. Windowed particle images were decimated three-fold by pixel binning and aligned to the reconstruction of the free *P.fu* RNAP (Chapter II) that was Gaussian low-pass filtered to a resolution of 25 Å. After 25 rounds of angular refinement, the resolution improved to 19.5 Å (Figure 17).

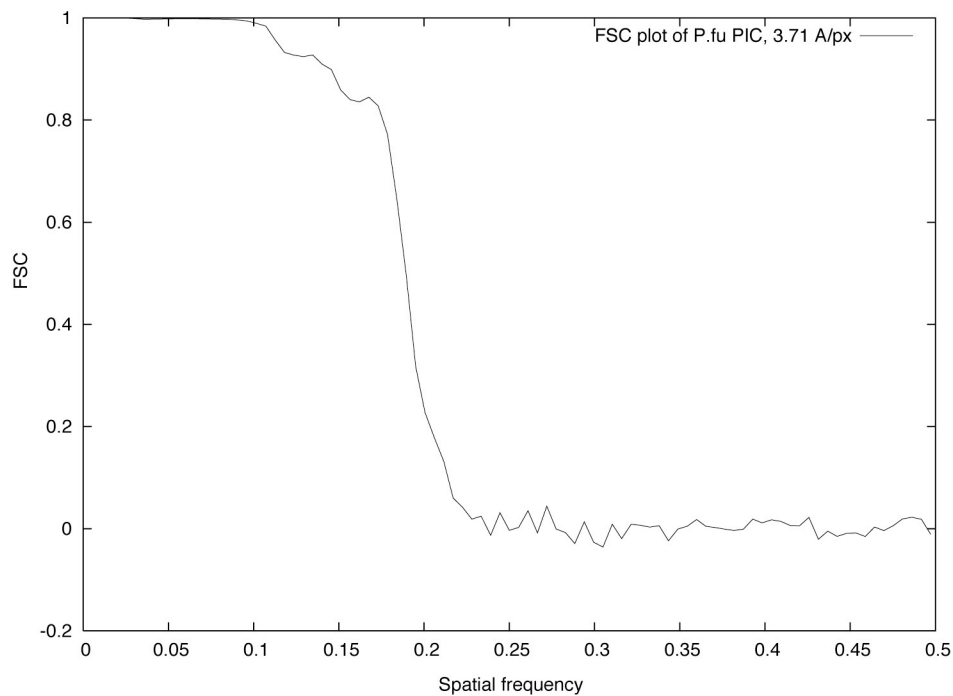


Figure 17: Fourier shell correlation plot for the *P.fu* mPIC reconstruction

After 25 round of angular refinement, the resolution was estimated to be around 19.5 Å according to the $FSC_{0.5}$ criterion.

Weak additional densities could be observed in the dock domain region of RNAP, consistent with the expected binding location of the N-terminal part of TFB, and close to the wall domain, again consistent with the expected binding location of the C-terminal part of TFB (Figure 18).

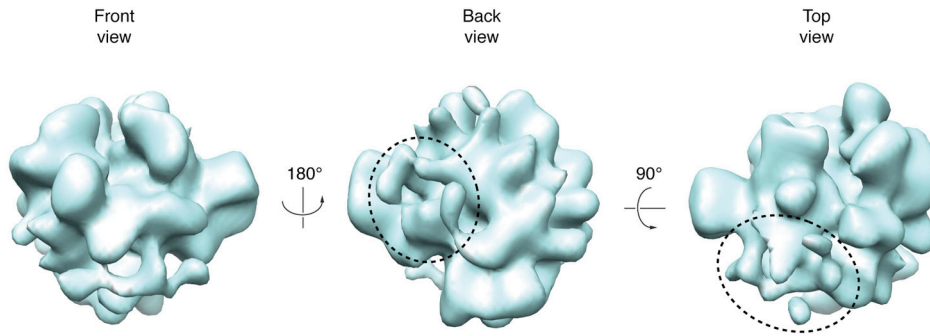


Figure 18: *P.fu* mPIC reconstruction

The model had an estimated resolution of about 19.5 Å ($FSC_{0.5}$, Figure 17) and is shown in front, back and top view. Additional densities are indicated by dashed ellipses.

However, the densities were not well defined enough to clearly attribute them to domains of TFB. Furthermore, no density for nucleic acids and TBP could be observed (Figure 18). The first explanation was, that maybe not all particle images represented a fully assembled mPIC, but some represented RNAP from which factors had dissociated between thawing the complex and sample vitrification. One well-known problem is the charged carbon surface of the grid, towards which particles are pressed during blotting, which might cause disassembly of transient complexes (Hans Elmlund, personal communication). It was consequently tried to classify images using the supervised, reference based classification approach based on cross-correlation values (sorting), as described in the experimental procedures section of Chapter II. The refined volume at that stage (Class I) and the original reference volume (Class II, free *P.fu* RNAP, compare Chapter II) were provided as references for 25 rounds of sorting that resulted in a rather stable separation of the dataset into 72,785 Class I and 52,374 Class II particles. However, resolution and additional densities did not improve in the Class I volume and upon comparison of the two reconstructions, it became clear that the densities of interest seemed to cease (Figure 19). It was concluded that the sorting procedure was not able to give a reasonable separation of the dataset, in the sense that the two classes represent two distinct conformations or complexes of different composition.

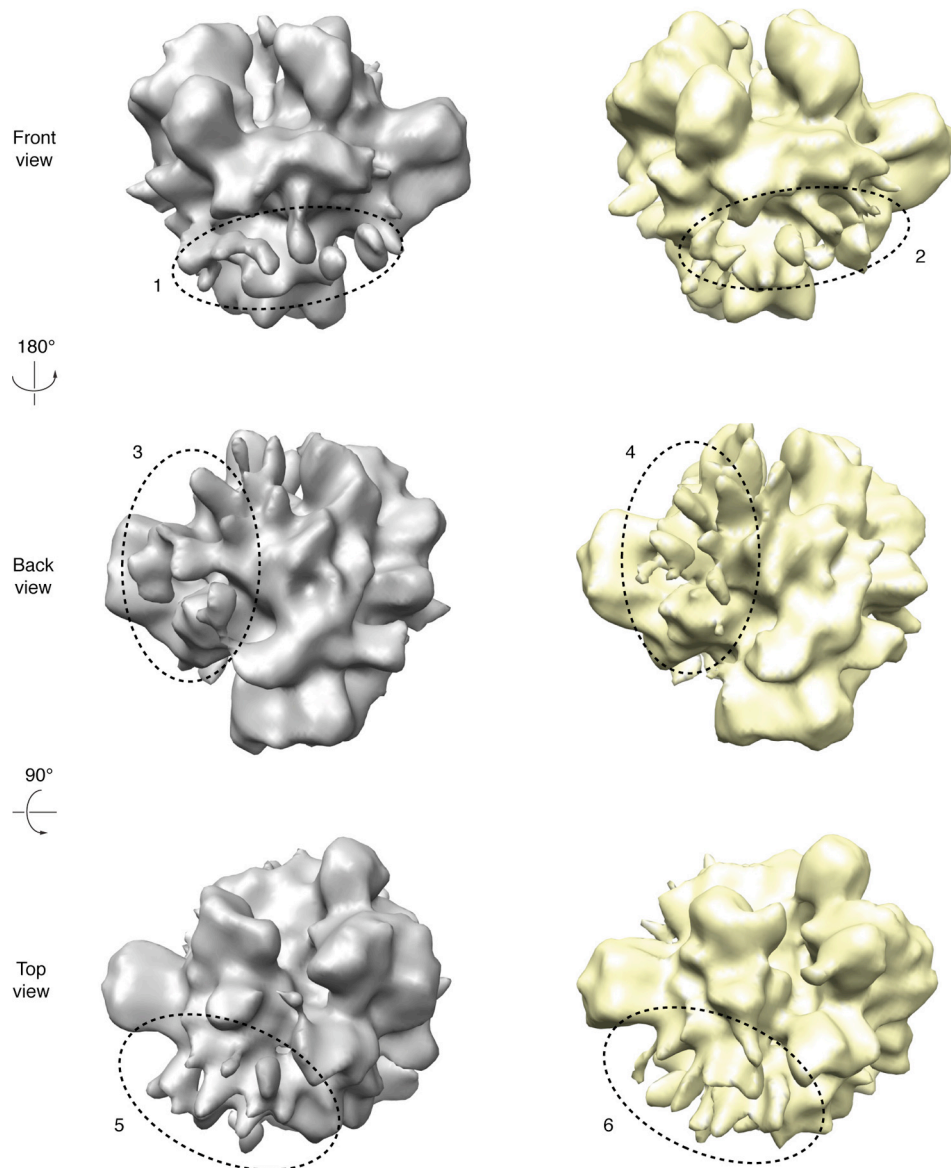


Figure 19: Two *P.fu* mPIC reconstructions from a sorting trial of the data set

The resolution of the two models did not improve upon sorting (data not shown) and additional densities of interest weakened for the region where TFB was expected (dashed ellipses 4 to 6). Densities for the foot region (dashed ellipses 1 and 2) showed a similar behaviour.

Furthermore, at some regions it even appeared as the images were “overaligned” (spiky densities in the top view of Figure 20), meaning that noise dominated the alignment process in these regions (Roland Beckmann, personal communication). Looking at the single particle images (Figure 21) it

was conceivable that classification on these was too demanding a task because of the low signal-to-noise ratio (SNR), even though, at this stage, three-fold decimation by pixel binning was applied to the images. The same procedure was then tried using only higher contrast particle images with defocus values below $-3 \mu\text{m}$, but again, no successful separation of the dataset could be generated. To rule out that there could have been a problem with complex assembly in the first place, another dataset was collected from a newly prepared sample and data were processed in a similar fashion but results were comparable. Another obstacle for classification was the lack of information about where to expect conformational variability. For the free RNAP it was possible to obtain a rather high-resolution model by sorting of the dataset with the use of *a priori* information from crystallographic studies, namely that the clamp domain was expectedly the most mobile element (Cramer et al., 2001; Kuhn et al., 2007). Nonetheless, it seemed reasonable that, by sufficiently improving the SNR of the single images, it would be feasible to improve supervised classification. At that time, research on yeast RNAP III had already been initiated, and first results seemed promising. It was therefore decided to use the gained experience and put the newly devised ideas and approaches to the test with this project (compare Chapter III and Appendix, paragraph 2).

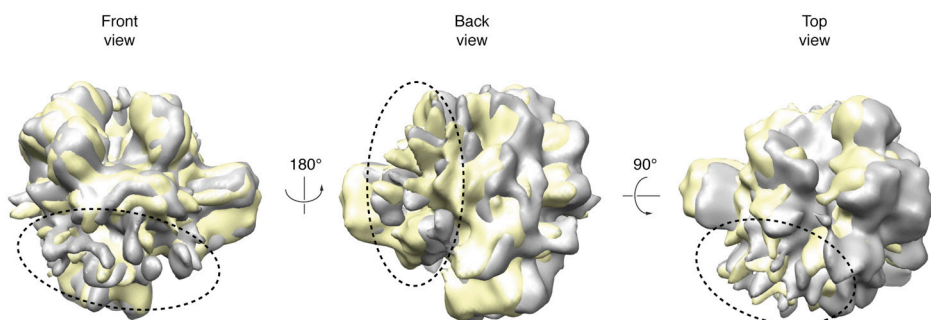


Figure 20: Overlap representation of the two volumes shown in Figure 19
Indications for overallignment are visible (spiky densities, dashed ellipses)

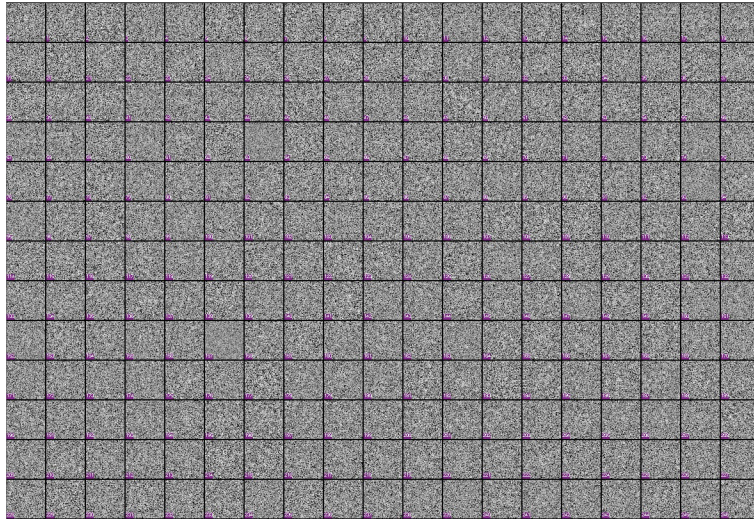


Figure 21: Single particle images of the *P.fu* mPIC

2 Yeast RNAP III minimal pre-initiation complex studies

The complex that was studied during this work was a minimal PIC consisting of *S.cerevisiae* RNAP III, a Brf-TBP fusion protein (Kassavetis et al., 2005) and a short stretch of a promoter DNA scaffold containing an artificial bubble. Expression and purification of the Brf-TBP fusion protein as well as mPIC assembly was done by Dr. Alessandro Vannini and detailed protocols will be described elsewhere.

2.1 Sample preparation and cryo-EM data collection

A yeast mPIC was investigated containing RNAP III, a Brf-TBP fusion protein (Kassavetis et al., 2005) and an open bubble DNA scaffold (Figure 15A, bubble scaffold). To assemble the complex, RNAP III was purified until after the MonoQ column as described in the experimental procedures section in Chapter III and then diluted to ~50 mM $(\text{NH}_4)_2\text{SO}_4$ using Sup50 buffer. DNA was pre-incubated with a 2 fold excess of Brf-TBP for 5 minutes at 20°C. Diluted RNAP III was added such that the ratio of RNAP III to DNA was 1:2 and incubated for 20 minutes at 20°C. The protein was then concentrated and subjected to size-exclusion chromatography on a Superose6 column (GE Healthcare) using Sup50 buffer as described in Chapter III. The peak fractions showed a reasonable 260/280-absorbance ratio, which supported binding of nucleic acids (data not shown). Fractions were collected and the protein concentration was measured as described in Chapter II. The sample was then diluted to a concentration of ~0.1 mg/ml and applied to freshly glow discharged carbon coated grids (Quantifoil R3/3 + 2nm Carbon on top). After 45 seconds of incubation, excess of buffer was blotted for 10 seconds using a semi-automated controlled-environment system (Vitrobot, FEI Company) at 4°C, 95% humidity. Samples were then flash-frozen by plunging them into liquid ethane and subsequently stored in liquid nitrogen until transfer to the microscope.

Micrographs were recorded under low dose conditions of ~ 15 electrons/ \AA^2 on a FEI Tecnai Spirit microscope operating at 120 kV, equipped with a LaB6 filament and a Gatan side entry cryo-holder. Images were acquired at underfocus values in the range of 1.5-7 μm on a 2k x 2k FEI Eagle CCD camera applying a pre-exposure of 100 ms at a magnification of 90,000x resulting in a pixel size of 3.31 $\text{\AA}/\text{px}$ on the object scale. A wide range of underfocus values was chosen to obtain a good portion of the dataset at high defocus and maximum image contrast for classification, taking into account the rather limited high spatial frequency information of these images.

2.2 Cryo-EM data processing and results

Initial particle selection and windowing was performed semi-automatically using the boxer program from the EMAN software package (Ludtke et al., 1999). Reference particles were picked manually on every micrograph to avoid discrepancies due to defocus and ice differences and the resulting automatically selected particles were verified visually. All subsequent image-processing operations were carried out using the SPIDER software package (Frank et al., 1996), if not stated otherwise. 9,803 windowed particles were aligned to 83 projections of the Pol II X-ray structure (1Y1W, Gaussian low-pass filtered to 35 \AA), which was modified by removing the OB and HRDC domains from subcomplex Rpb4/7 since they are deemed to conformational flexibility. After 16 rounds of angular refinement, clear additional densities appeared at the lobe and funnel as well as the clamp head domain (Figure 22, dashed ellipses 1 and 2, respectively), as it could be observed in the free RNAP III and the RNAP III-DNA-RNA complex described in Chapter III. Moreover, surplus densities became apparent at the dock and the wall domain of RNAP III (Figure 22, dashed ellipses 3 and 4), consistent with earlier results of *P.fu* mPIC studies and recent crystallographic results (Kostrewa et al., 2009). The resolution of the reconstruction at that stage was

determined by the $FSC_{0.5}$ criterion on the unmasked, unfiltered volumes to be around 27.5 Å (Figure 23).

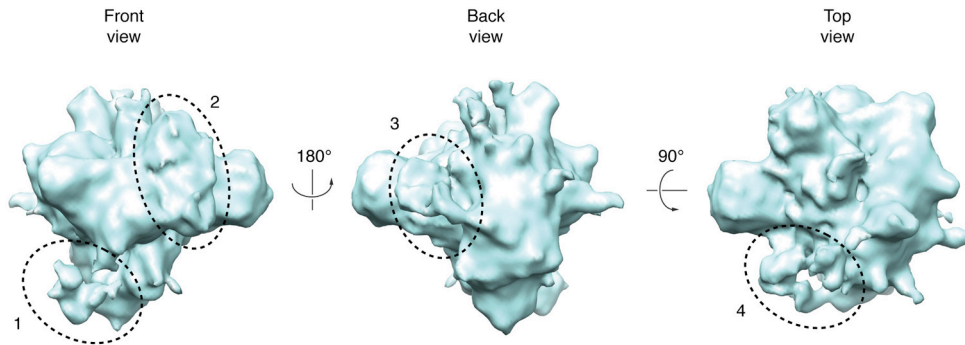


Figure 22: Reconstruction of the RNAP III mPIC at 27.5 Å

Clear additional densities appeared at the funnel and the clamp domain (dashed ellipses 1 and 2, respectively) and close to the dock and the wall domain (dashed ellipses 3 and 4).

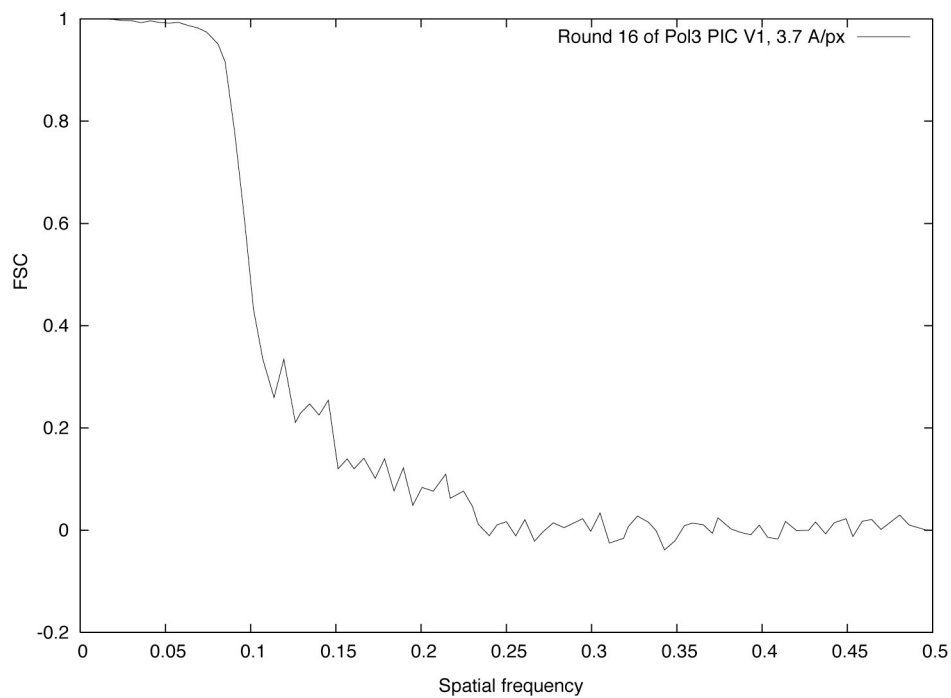


Figure 23: Fourier shell correlation plot for the RNAP III mPIC reconstruction

After 16 rounds of angular refinement, the resolution of the unmasked, unfiltered volume was estimated to be around 27.5 Å according to the $FSC_{0.5}$ criterion.

It was then tried to take advantage of the presumably better SNR and image phases in the lower spatial frequency of the CCD images (compare Appendix, paragraph 3) for supervised classification of the single particle images by sorting. The cryo-EM reconstruction of the free RNAP III (compare Chapter III) was used as second reference for another 20 rounds of angular refinement that resulted in two models with apparently different densities at the funnel domain (Figure 24, dashed ellipses 1 and 3) and on the clamp (Figure 24, dashed ellipses 2 and 4), presumably indicating different conformations of C53/37 and C82/34/31, respectively. The additional densities in the back of RNAP III did not show significant improvements, neither did the resolution of either of the two models (Figure 25). An explanation could be that there were too many dimensions of variability as to be able to separate them by only two references, one of which was rather unrelated to this variability, the other being an average from all conformations present in the data set. Henceforth, it was attempted to resolve heterogeneity with the use of more sophisticated methods.

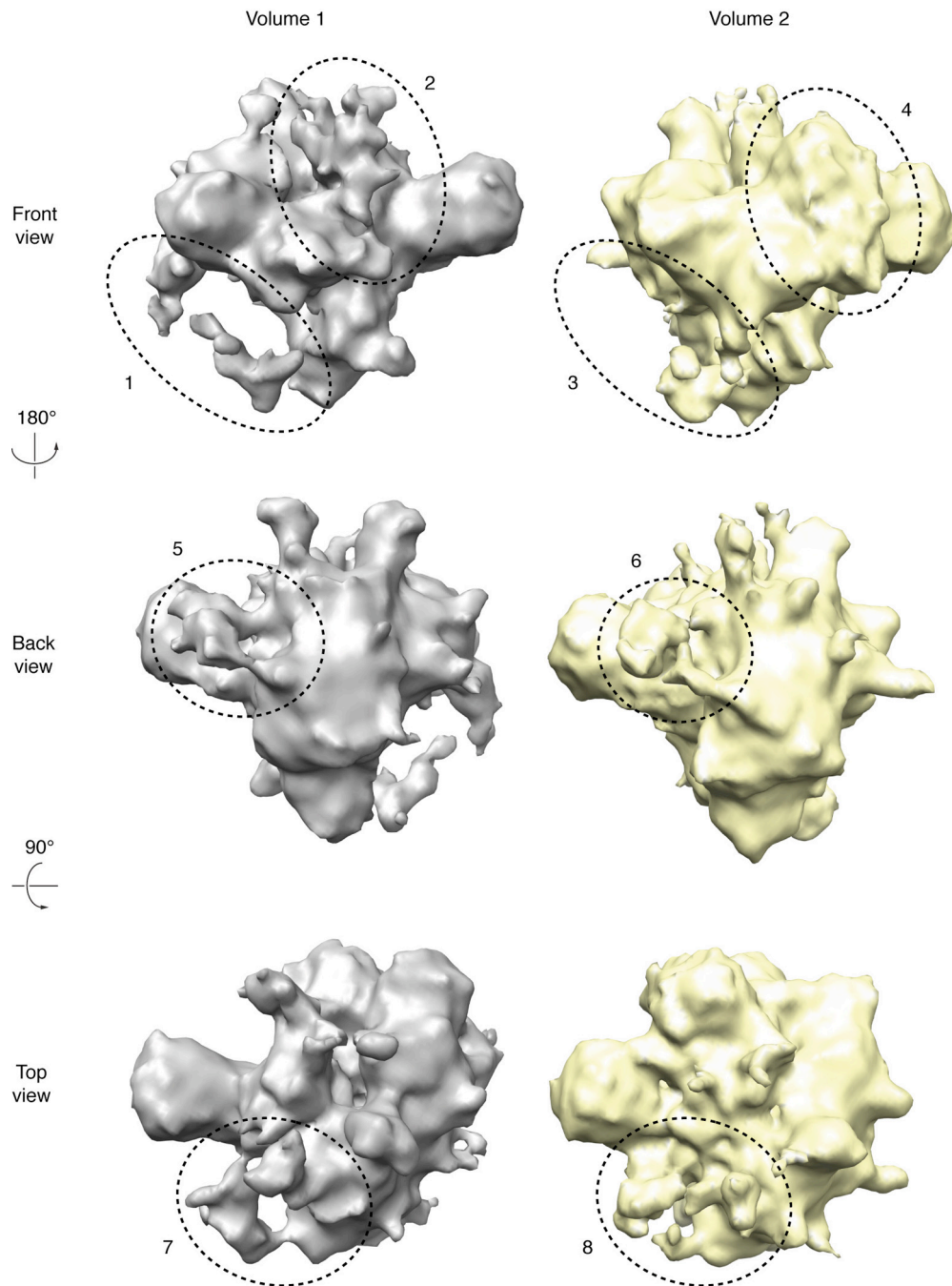


Figure 24: Two reconstructions of the RNAP III mPIC at $\sim 27.5 \text{ \AA}$

Differences in observed additional densities suggest conformational variability of subcomplexes C53/37 (dashed ellipses 1 and 3) and C82/34/31 (dashed ellipses 2 and 4). Similar changes are visible for densities in the region where Brf1 is expected (dashed ellipses 5 to 8).

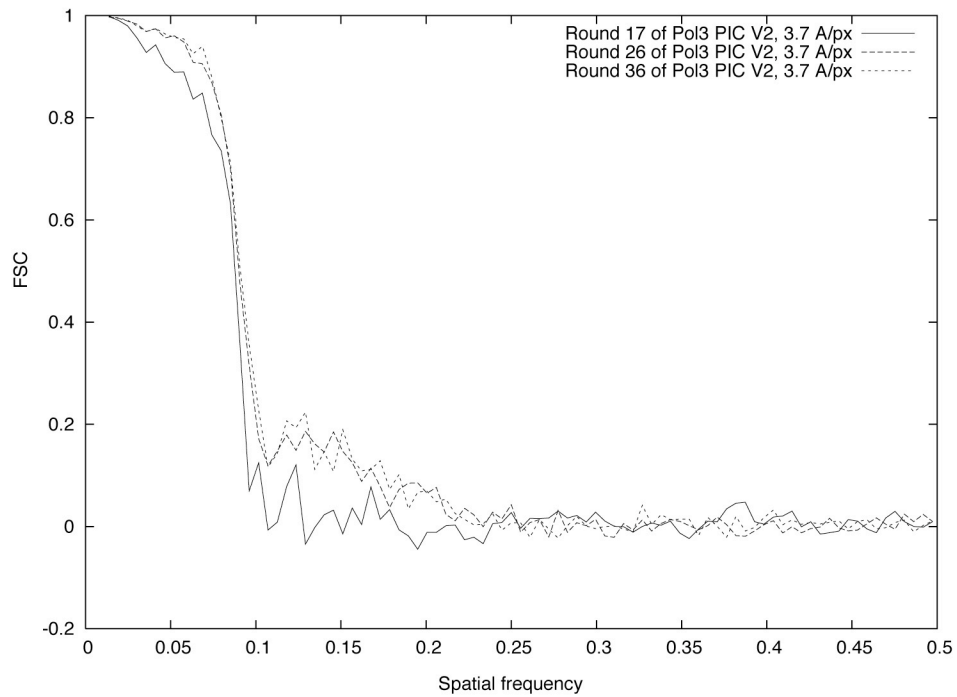
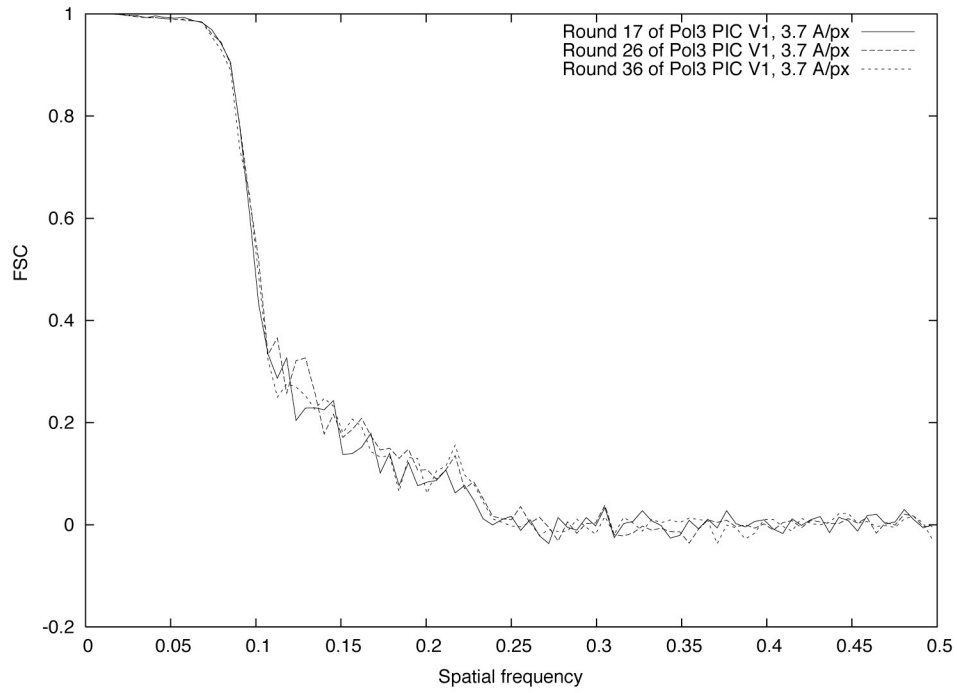


Figure 25: Fourier shell correlation plot for the two RNAP III mPIC reconstructions

3 Technical issues addressed during this work

Several technical issues were addressed during this work and this paragraph is meant to be a comprehensive protocol and maybe a starting point for a reader new to the field of cryo-electron microscopic analysis of RNAP complexes.

3.1 Electron acceleration voltage and image contrast

In theory, higher electron acceleration voltage and the consequential smaller electron wavelength allows for higher resolution imaging of the specimen. In practice, especially when dealing with biological specimens, the wavelength of the electron is not the limiting factor as it becomes obvious when expressing its numeric value in Å: considering them as waves, 120 keV electrons have a de Broglie wavelength of approximately 0.033 Å, which is two orders of magnitude smaller than what can be resolved from biological specimens by a state-of-the-art instrument like the Titan Krios (Frank, 2006). Most limiting to resolution on the instrument side are usually specimen or sample stage drift, aberrations of the electro-magnetic lenses and sub-optimal microscope alignment (Frank, 2006; Reimer and Kolh, 2008; Williams and Carter, 2009). There is an important aspect to the choice of acceleration voltage when dealing with potentially heterogeneous biological specimens, though. The higher the acceleration voltage and the resulting velocity of the electrons (considering them as particles in this case), the lower the so-called scattering cross-section, which is a measure of the probability of interaction of an electron with the specimen (Reimer and Kolh, 2008). “The target gets smaller as the bullets get faster”. This means that higher energy electrons indeed interact less with the specimen and give rise to lower signal on the image while lower energy electrons interact stronger, and give better signal or image contrast for that matter, at the possible cost of higher radiation damage (Reimer and Kolh, 2008). Figure 26 shows particle images of RNAP III

imaged on film at 300 kV while Figure 27 show particle images of RNAP III imaged at 200 kV at the same electron dose (the smaller image size in Figure 26 is due to different boxing parameters, not image decimation factors).

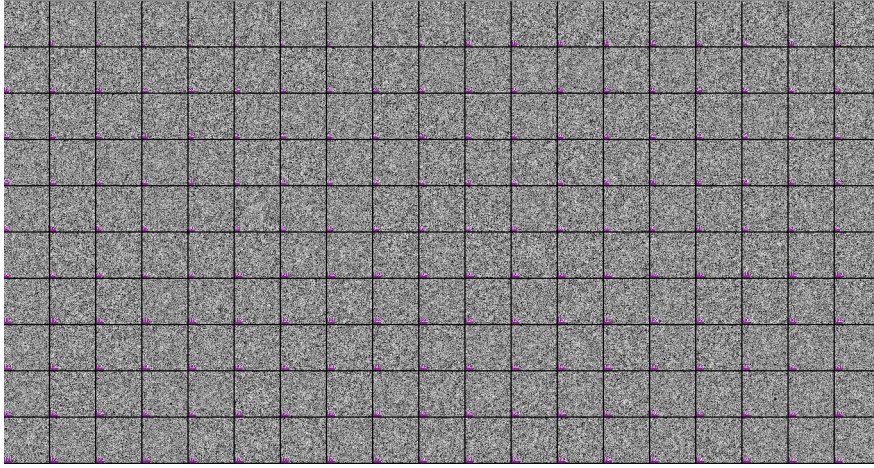


Figure 26: RNAP III particles collected at 300 kV

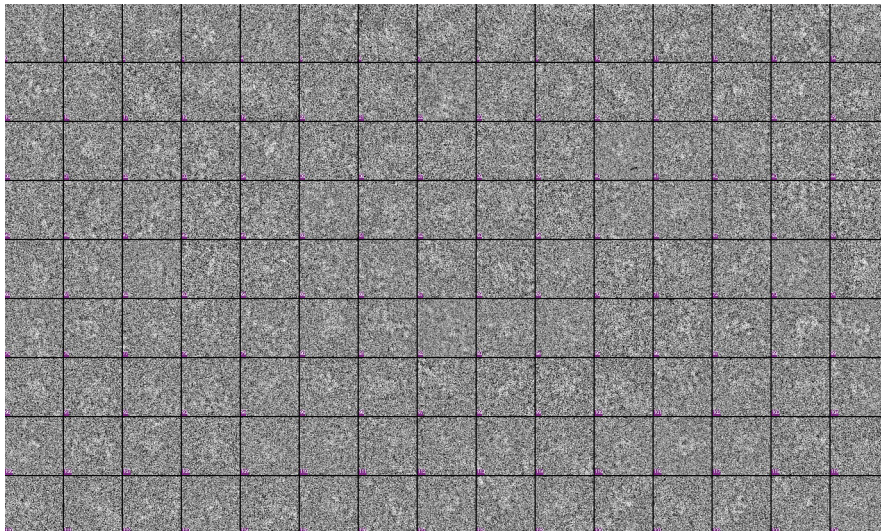


Figure 27: RNAP III particles collected at 200 kV

It is in principle possible to balance out for the loss of electron-specimen interaction at higher voltage by using a higher electron dose, thus having more electrons that effectively deliver signal to the recorded image, but a general conclusion has not yet been reached about what electron dose an average biological sample in vitrified ice can take before radiation damage

hampers image processing and reconstruction (Hans Elmlund, personal communication). It is conventional to be conservative regarding electron dose: imaging for single particle 3D reconstruction of biological specimens is usually done by exposing for about 1 s under low-dose conditions of 10-30 electrons/Å² in an electron energy range of 100-300 keV, respectively (Otto Berninghausen, personal communication). It was however suggested that vitrified biological samples might be able to take as much as 60 electrons/Å² at 300 keV for an exposure time of 1 s (Hans Elmlund, personal communication). The majority of data collection during the time of this work was done at 120 kV and under low-dose conditions of ~15 electrons/Å².

3.2 CCD vs film data

When determining structures of protein complexes that have not been extensively studied already, and for which no or only little *a priori* information about existing conformational states is available, high SNR ratios are a critical point. It has been shown that in the lower spatial frequency ranges of ~10-20 Å, and using oversampling, CCD images can have up to two-fold better spectral signal-to-noise ratios and that image phases on CCD in the 20 Å range are more reliable (Sander et al., 2005). CCD data has other advantages: when collecting data manually, an experienced user can already assess the quality of the images taken, by judging ice thickness as well as particle distribution and appearance on the images. Also, the period from data collection to a first reconstruction is dramatically shortened because time consuming tasks like developing and scanning of the micrographs are gratuitous. Moreover, image artefacts coming from dust during the scan or developer stain are eliminated or at least minimized.

In conclusion, CCD data can be considered superior when data of new specimens is analyzed, where workflows or sample preparation techniques need to be optimized or when it comes to rather low-resolution initial structure determination of potentially heterogeneous specimens that have to be

analyzed with respect to this heterogeneity. If classification is not crucial because good models for segregation of the data sets exist already, it is advisable to collect film data for obtaining reconstructions in the sub-nanometer resolution range.

3.3 Attempts to resolve heterogeneity

RNAP reconstructions published so far failed to reach a nominal resolution better than 11 Å. When almost 100,000 RNAP III particles from a high-end microscope collected at 200 kV on film were used and the nominal resolution determined by $FSC_{0.5}$ was stuck at 19 Å, it became evident that averaging of images representing an inhomogeneous particle population was a major issue. Ribosome reconstructions can reach 5-6 Å if appropriate separation of the data set can be applied (Seidelt et al., 2009; Roland Beckmann, personal communication). For RNA polymerase, the lack of *a priori* information about conformational dynamics within the enzyme evidently limits the possibilities to create good models for supervised classification approaches. Therefore, methods for unsupervised classification were tested.

3.3.1 Unsupervised classification

For unsupervised classification, a comprehensive software package called SIMPLE (paper submitted) was tested. During that period, the package was still under development and installation and execution of code was done with kind support from Hans and Dominika Elmund. SIMPLE combines an *ab initio* reconstructions technique (Elmlund et al., 2008) and a continuous angular space refinement method based on spectrally self-adapting common lines (Elmlund and Elmlund, 2009). It was successfully applied to obtain two reconstructions of the general transcription factor TFIID at a resolution below 10 Å (Elmlund et al., 2009).

Creation of an *ab initio* model using the common lines method generally needs data with an optimal SNR. Such data is usually obtained from negative stain or cryo-negative stain specimens. Since such data was not available at the time, and a reasonable first low-resolution model existed (compare Chapter III) we decided to use this model as initial reference.

Initially, 25,242 CTF corrected (Wiener-filtered) particle images from two merged data sets of free RNAP III (collected as described in Chapter III) were aligned to the 21 Å reconstruction of free RNAP III (Figure 28, compare Chapter III). The obtained inplane rotational parameters were applied to the particle images, allowing for subsequent principal component analysis and hierarchical clustering of the dataset. The resulting averaged particle classes (Figure 29) were used as input for a state-assignment algorithm using a point normalized cross common line correlation coefficient as distance function and simulated annealing as an optimization procedure (Elmlund et al., 2009). Euler angles are determined from the averages by a modified common line method.

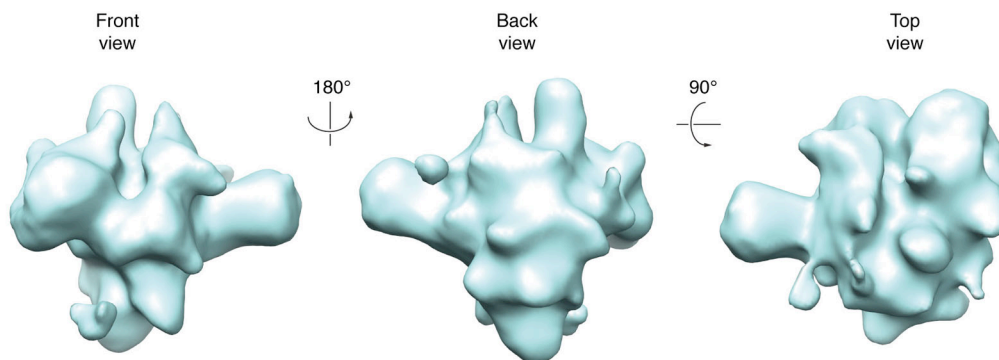


Figure 28: RNAP III reference volume (as described in Chapter III)

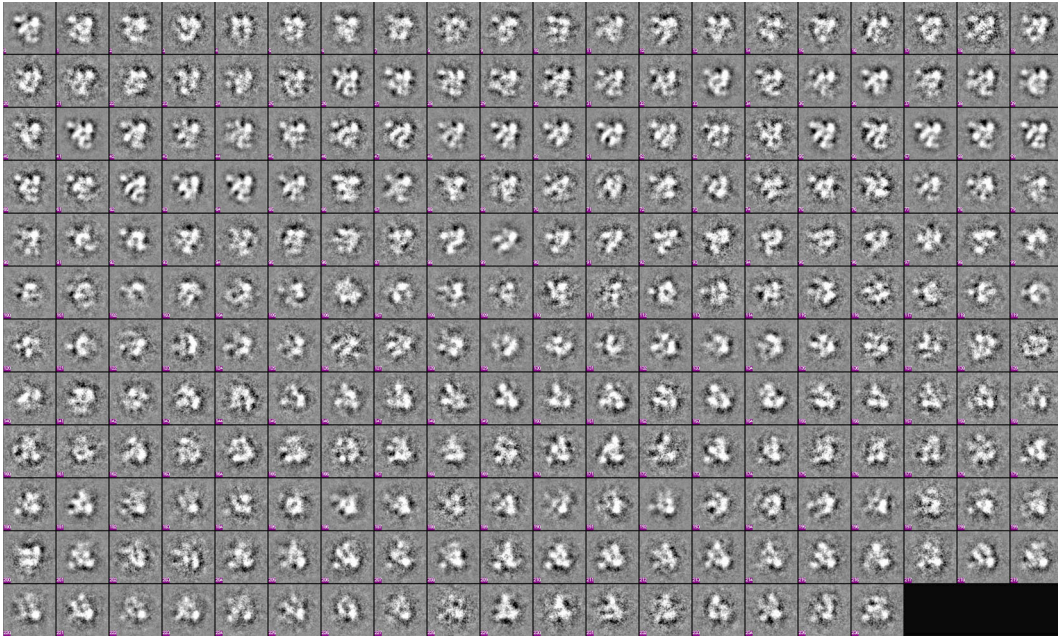


Figure 29: Class averages

Image classes were obtained after inplane alignment to reference projections of the free RNAP III, subsequent principal component analysis and hierarchical clustering of the data set.

The combination of these methods resulted in three models (Figure 30) apparently displaying different conformations of the stalk, the clamp and the lobe domain where the C82/34/31 and the C53/37 subcomplexes are located, respectively (compare Chapter III). These models were used for subsequent refinement in continuous angular space and resulted in three first reconstructions (Figure 31). These models showed some discontinuity as it can be seen in the top views of reconstruction 1 and 3 and discontinuities aggravated after another round of refinement (Figure 32), suggesting that the initial models did not yet represent the complete conformational variability present within the data set, or that the prior CTF correction was based on incorrect parameters (Hans Elmlund, personal communication).

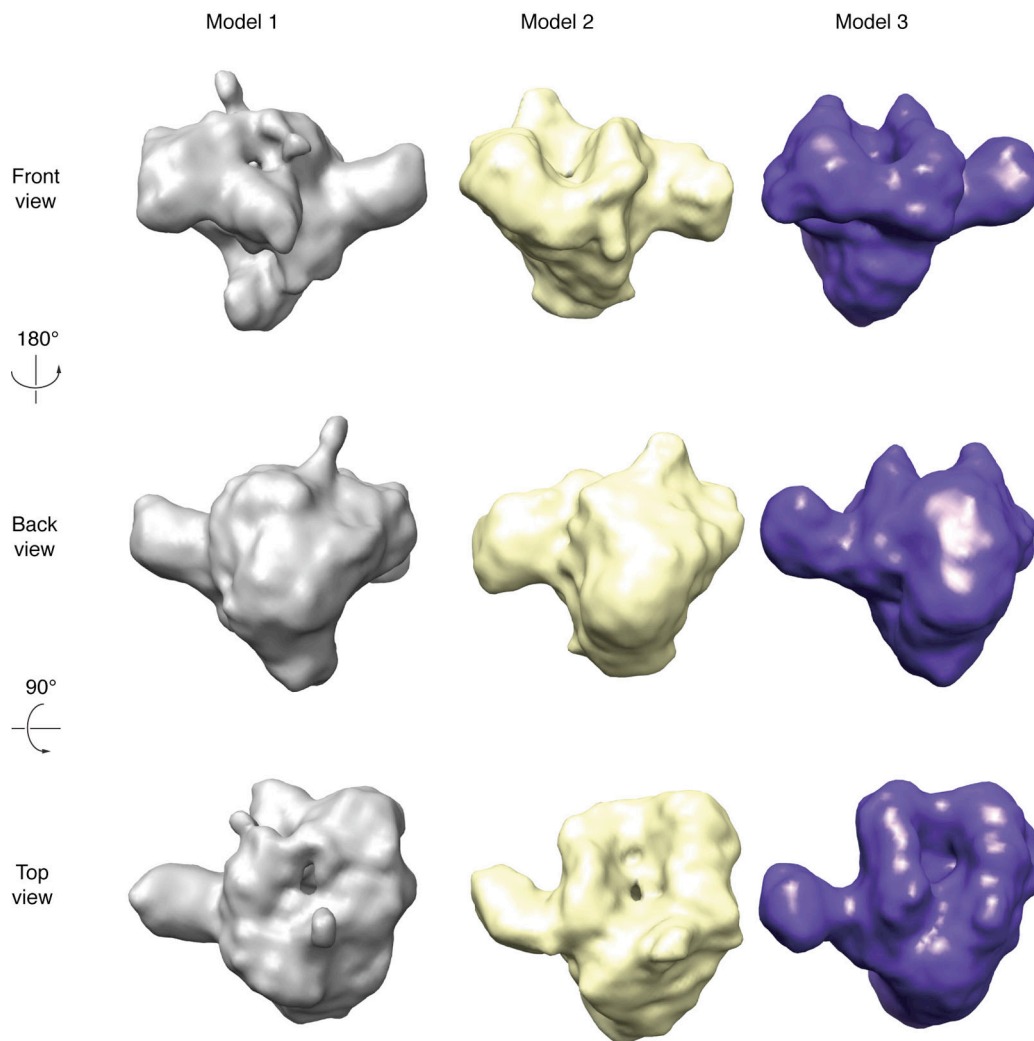


Figure 30: Initial models generated from the class averages shown in Figure 29
Different stalk, clamp and lobe conformations seem present.

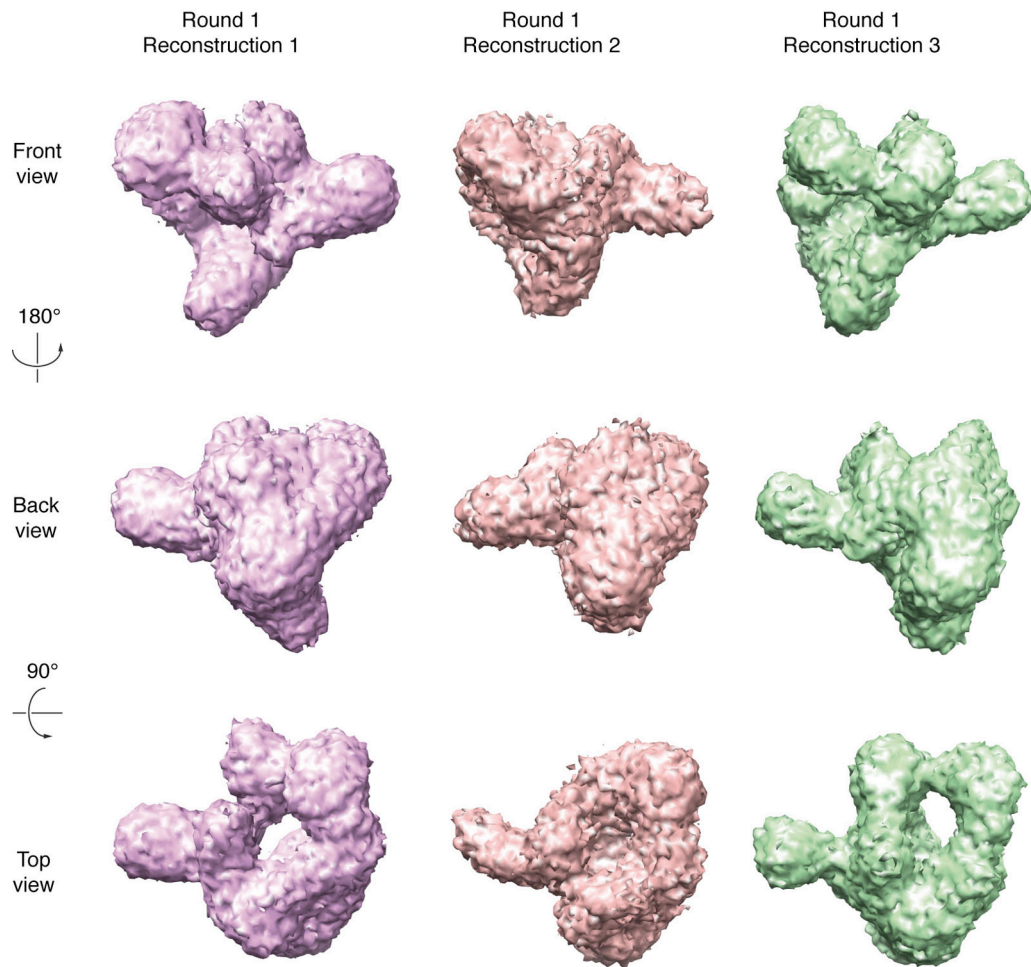


Figure 31: Models after first refinement round

Discontinuities can be observed in the top views of reconstructions 1 and 3. The figure shows threshold masked, unfiltered volumes since this is required for the self-adapting common line algorithm that is used in this software.

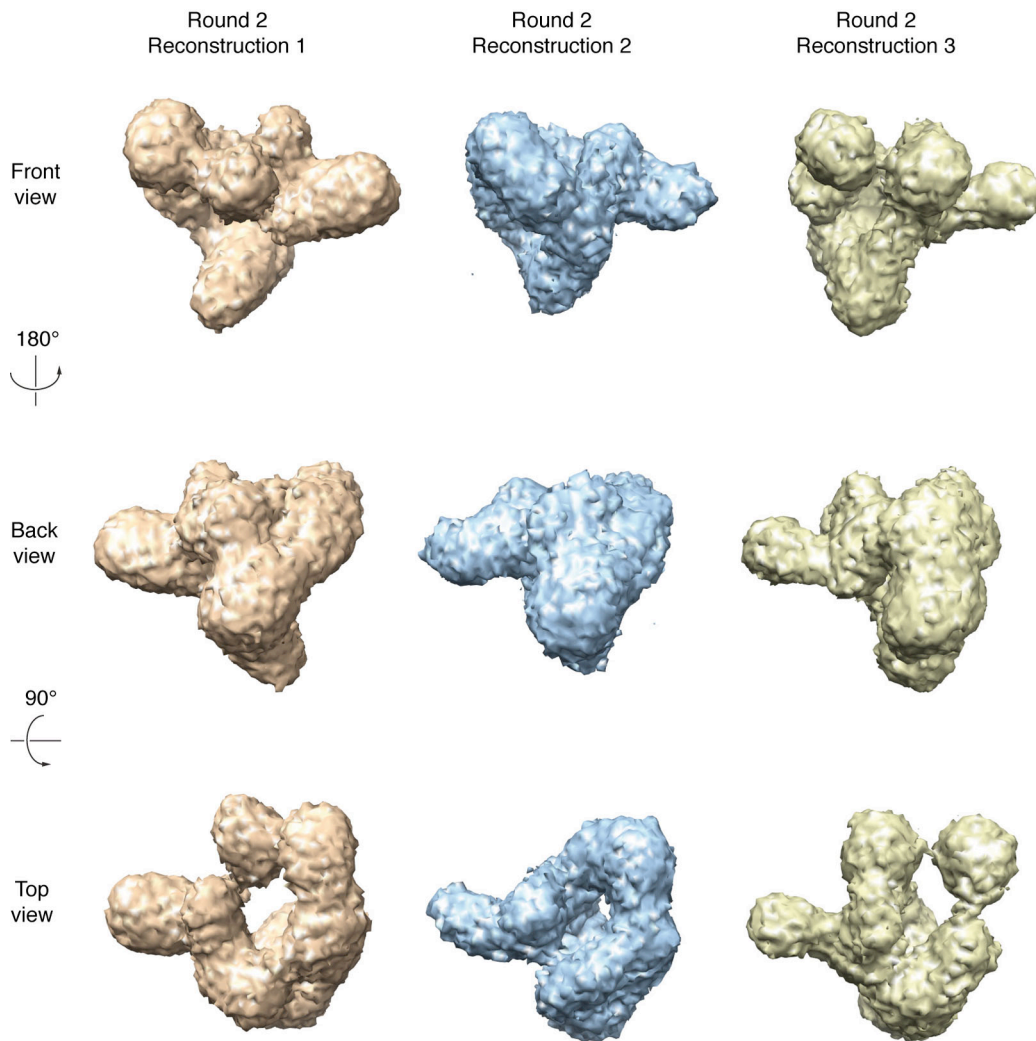


Figure 32: Models after second refinement round

Discontinuities aggravated in all three reconstructions. The figure shows threshold masked, unfiltered volumes since this is required for the self-adapting common line algorithm that is used in this software.

The 120 kV CCD data used in this case made it necessary to parameterize the CTF manually, which could well have been a source of error. A faulty CTF correction could however be rendered unlikely by the fact that similar results were obtained using 200 kV film data. For this data, the CTF parameterization was possible in an automated and thus more reliable and reproducible fashion. This is partly due to the fact that more high-frequency information is present in the corresponding micrographs and the fact that a single micrograph covers a larger area, making the results of the sub-area sampling procedure used in the CTF parameter determination algorithm more

accurate (Frank, 2006). Another general problem could be the determination of euler angles by the common line method, which is based on the assumption that the images are projections of the same object, only taken from different viewing points, or copies of the same object in different orientations relative to the single point of view, for that matter. This assumption is obviously questionable since we are actually trying to resolve heterogeneity in the first place.

Furthermore, the refinement procedure of the SIMPLE software was still under development at that time. Hence it was tried to combine the euler angle determination by reference-based alignment with the unsupervised state-assignment method for determining heterogeneity and the resulting models were used as input for the supervised classification technique described in Chapter II to form a “hybrid-approach”.

3.3.2 A “hybrid approach”

It seems reasonable to use any available *a priori* information for aiding the separation of different conformational states if necessary. Since it is known that RNAP III subunits C53/37 are present in the endogenous enzyme in a sub-stoichiometric manner (Lorenzen et al., 2007), it was tried to make use of that information by directing (supervising) the classification accordingly.

Initial euler angle assignment to the particle images was performed by reference-based alignment as described in Chapter II and III using as a reference the model of the free RNAP III at a resolution of 21 Å. Angles were refined for 10 rounds and the refined angles were used for state-assignment to the images (Elmlund et al., 2009). Backprojection of the segregated particle images resulted in three states, two of which were promising to use them as starting models for further supervised classification (Figure 33).

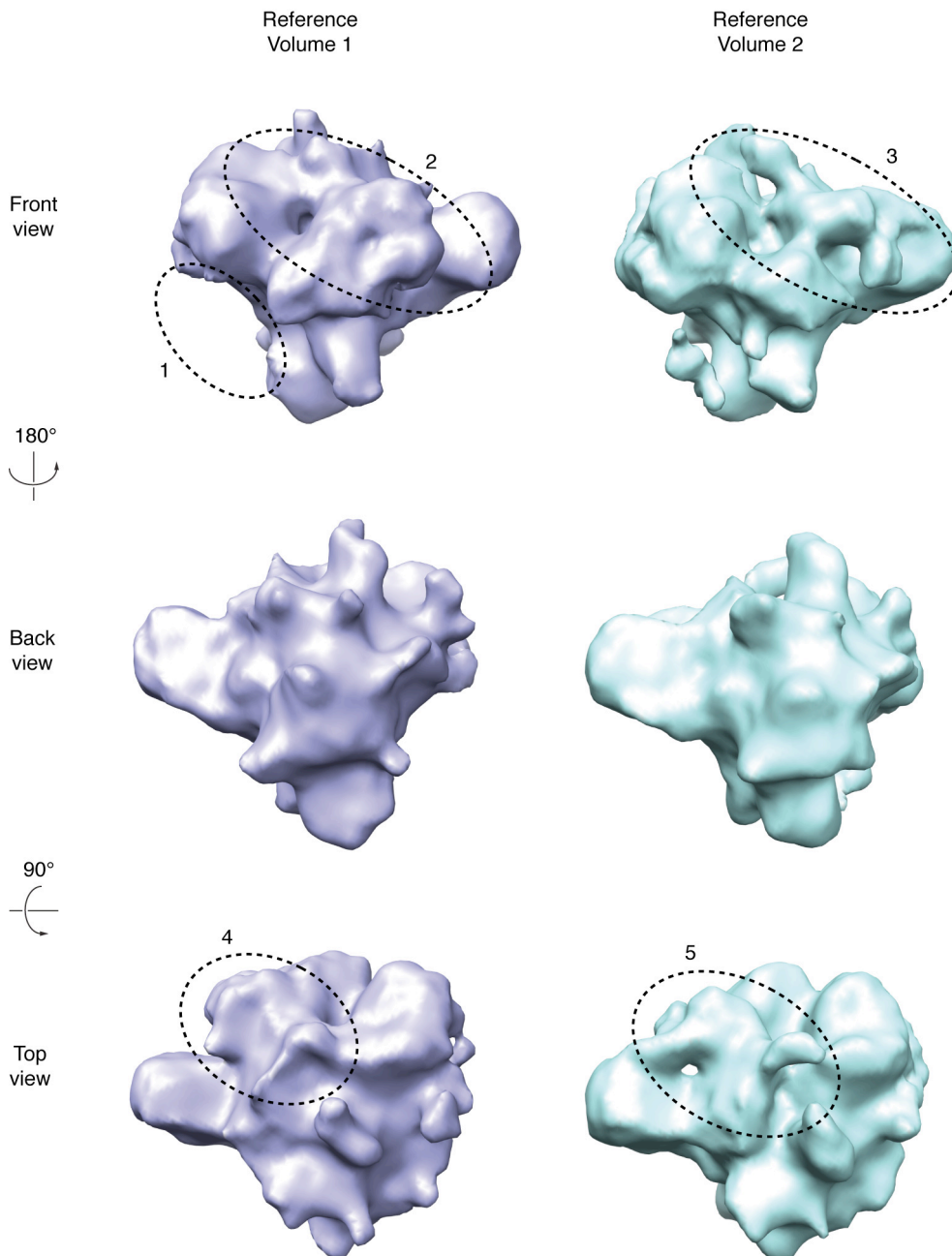


Figure 33: Initial models after state-assignment, backprojection and manual density modification (compare text)

Residual densities at the funnel region were removed manually (dashed ellipse 1, compare text). Densities on top of the clamp (dashed ellipses 2 to 5) suggested conformational variability of the C82/34/31 subcomplex as observed before.

These two models showed apparent conformational variability in the region of the C82/34/31 subcomplex (Figure 33, dashed ellipses 2, 3, 4 and 5). One of these models showed stronger density in the region of the RNAP funnel and lobe, where the C53/37 subcomplex is situated. The model

displaying weaker density was modified by manually removing any residual density from that region (Figure 33, dashed ellipse 1), assuming that this volume represented a subset of particles having only little amount of C53/37 bound (Lorenzen et al., 2007). The two models were then used as references for the whole data set, applying the sorting procedure described in Chapter II.

The first round of sorting already improved the resolution of the two reconstructions as it can be seen in Figure 34 (dashed graph labelled “Round 10”) compared to the last round of unsorted refinement (graph labelled “Round 9”). After 11 rounds of sorting, the resolution did not show significant improvements any more (graph labelled “Round 21”).

The resulting two final models showed more defined conformations of the C82/34/31 subcomplex (Figure 35, dashed ellipses 2 and 4) and clear details for that region in volume 2. A little stronger density for the C53/37 subcomplex could also be observed in volume 2 (dashed ellipse 3). A small additional density appeared in volume 1 (dashed ellipses 1 and 5). Most interestingly, a density that was previously attributed to the N-terminal WH domain of C34 (compare Chapter III) seems in a flapped out conformation in volume 2 (dashed ellipses 6 and 9), such that it would allow for interaction with an incoming Brf1-TBP-DNA complex in the back of RNAP III (compare Chapter III). Another rather small density protruding out over the cleft in volume 2 (dashed ellipse 8) from a location close to the clamp coiled-coil (Cramer et al., 2001) can possibly be assigned to the C-terminal part of C34. It would also appear that this region is in a more closed conformation in volume 1 (dashed ellipse 7). This could give further clues for interpretation of the function of C34 in the context of transcription initiation and elongation.

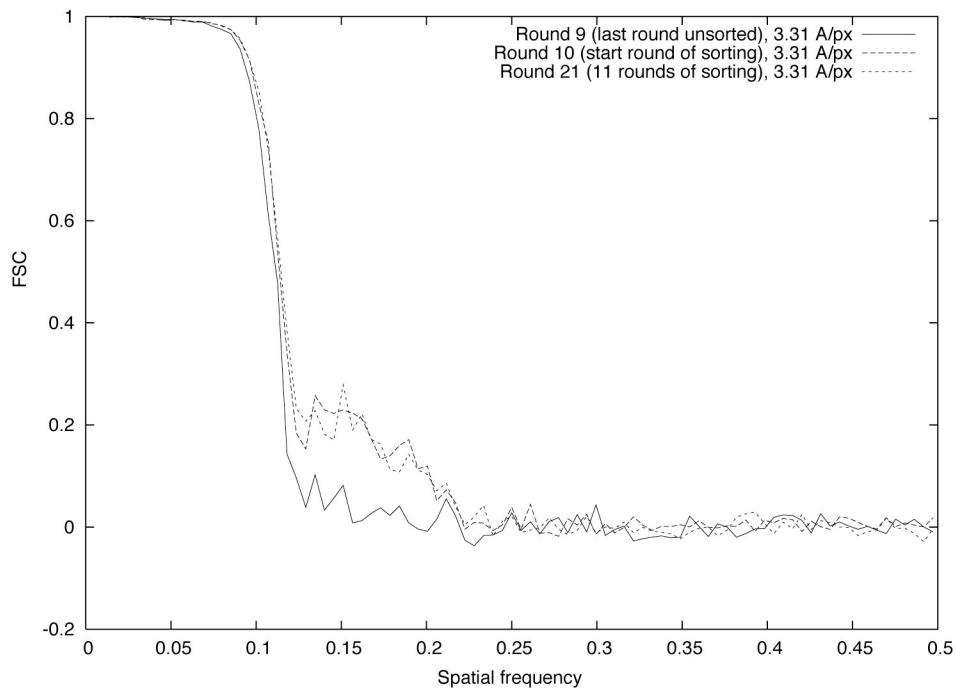
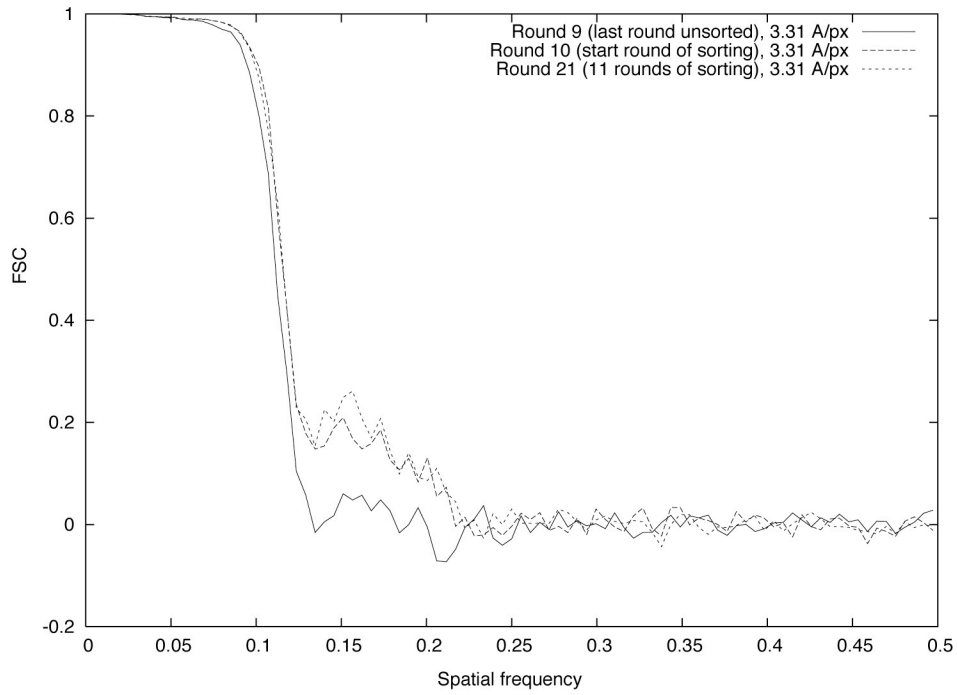


Figure 34: Fourier shell correlation plot for two RNAP III mPIC reconstructions
After the first round of sorting, the resolution improved in the spatial frequency range between 0.12 and 0.2. Subsequent rounds did not give any significant further improvements.

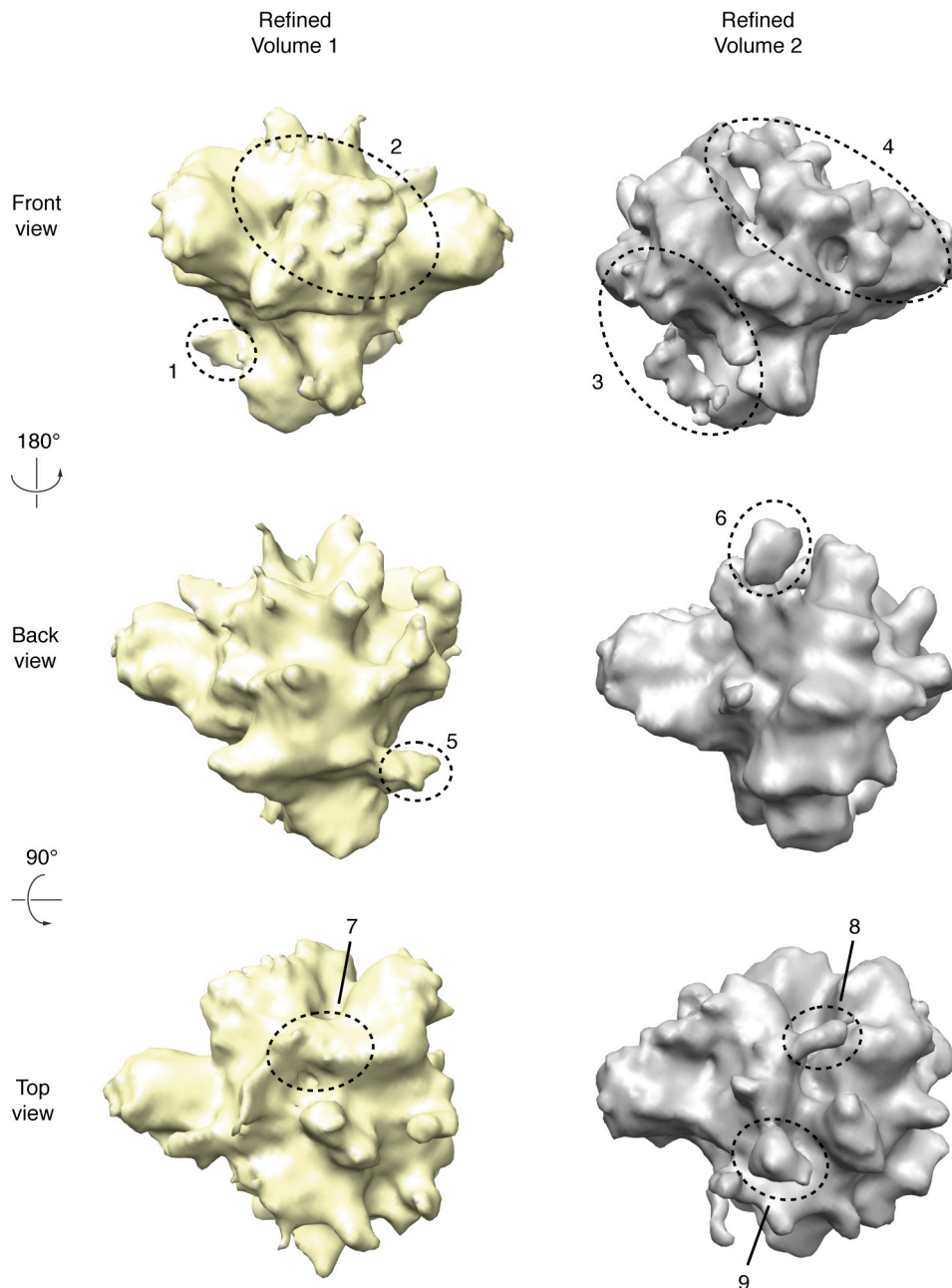


Figure 35: Refined models of RNAP III mPIC after classification

The final models showed more defined conformations of the C82/34/31 subcomplex (dashed ellipses 2 and 4) and clear details for that region in volume 2. A little stronger density for the C53/37 subcomplex (dashed ellipse 3) could also be observed in volume 2. A small additional density appeared in volume 1 (dashed ellipses 1 and 5). A density that was previously attributed to the N-terminal WH domain of C34 seems in a flapped out conformation in volume 2 (dashed ellipses 6 and 9), such that it would allow for interaction with an incoming Brf1-TBP-DNA. Another rather small density protruding out over the cleft in volume 2 (dashed ellipse 8) can possibly be assigned to the C-terminal part of C34. This region seems in a more closed conformation in volume 1 (dashed ellipse 7).

4 Conclusions and outlook

The comparison of the two mPIC studies discussed in paragraph 1 and 2 already point towards the importance of high SNR for reliable classification of single particle images. The 300 kV data set of the archaeal mPIC resulted in partly “overaligned” volumes with even weaker additional densities in regions of interest when trying to classify the images. The 120 kV data set of the RNAP III mPIC showed more promising results. Taken together with the results of paragraph 3.3.2, the technical advances leading to enhanced image contrast seem to pay off. However, unexpectedly weak densities in regions where transcription factors are expected point out that there remains room for improvement.

New state-of-the art instruments such as the Titan Krios are already being installed. The Titan Krios is developed bottom-up with the application of cryo-transmission electron microscopy in mind and will allow imaging of biological macromolecules at improved signal-to-noise ratios. This is accomplished by permitting a change of acceleration voltage down to 80 kV, by the use of constant-current lenses that increase thermal and electromagnetic stability, by the utilization of a third condenser lens improving beam parallelism, by the so-called “Auto-loader” system allowing for contamination-free change of up to 12 samples and by a soon-to-be-fully-automated data acquisition system that allows faster collection and subsequent processing of cryo-EM data.

That increases the importance of reliable methods on the data processing side of things, no doubt, but in the same time opens the door to extensive testing of new algorithms for classification on data sets of already established and well-known samples. With the improved quality, higher SNR and better resolution of the images, *ab initio* reconstruction and classification could be near at hand. New and improved alignment algorithms as well as a powerful 3D variance technique based on bootstrapping, both implemented in the latest version of the software package SPARX (Hohn et al., 2007), are close to be ready for first trials. Moreover, as soon as a final release of SIMPLE is

available, it will also be tested thoroughly, since high potential is attributed to the method of self-adapting common lines and refinement in continuous angular space (Elmlund et al., 2009). As soon as methods are assessed and optimized, challenging specimens like larger PIC or even Mediator complexes are going to be tackled.

Final conclusions and outlook

In conclusion, this thesis revealed new aspects of RNAP subunit architecture, the general transcription mechanism and its regulation in different domains of life. The detailed structure of an archaeal RNA polymerase underlined the evolutionary relationship between archaeal and eukaryotic RNAPs and was instrumental for initial studies of archaeal minimal pre-initiation complexes. The insights into RNA polymerase III and its regulation by Maf1 allowed modelling of the essential mechanism of RNAP III-specific transcription repression and improved our understanding of the RNAP III transcription initiation mechanism. The obtained structural models represented groundwork for studies of RNAP III minimal pre-initiation complexes and aided the interpretation of preliminary results.

In the future, these results will provide the basis for further structural and functional studies of RNA polymerase complexes. An archaeal RNAP-NusG complex is already under investigation and in depth studies of archaeal and eukaryotic pre-initiation complexes will be pursued. The technical experience gained during this thesis will be of great benefit for tackling these most challenging tasks. If the problem to resolve conformational and compositional heterogeneity can be solved, the consequential revelations might conduct optimization of crystallization construct design, eventually leading to groundbreaking atomic models, similar to what has been accomplished for RNAP II. The insights into subunit and transcription factor interactions from the studies on RNP III could be used to devise X-ray crystallographic projects concerning a Brf1-TBP complex together with nucleic acids and a region of C34 presumably stabilizing the C-terminus of Brf1. Furthermore, the presented model of RNAP III transcription initiation repression by Maf1 can help to bring reasons for misregulated RNAP III transcription to light, a dysfunctionality that has been found related to cancer.

Abbreviations

ABC	subunit prefix for RNAP I (A), II (B) and III (C)
AC	subunit prefix for RNAP I (A) and III (C)
C	subunit prefix for RNAP III (C)
CAP	catabolite gene activator protein
CCD	charge-coupled device
Cryo-EM	cryo-electron microscopy
C-terminus	carboxy-terminus
Ct-NLS	C-terminal NLS
CTF	contrast transfer function
CV	column volume
DNA	deoxyribonucleic acid
DTT	dithiothreitol
<i>E.coli</i>	<i>Escherichia coli</i>
EDTA	ethylene diamine tetraacetic acid
EM	electron microscopy
FRET	fluorescence resonance energy transfer
FSC	Fourier shell correlation
GTF	general transcription factor
GTM	general transcription machinery
Hepes	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HRDC	helicase and RNase D C-terminal domain
<i>H.sapiens</i>	Homo sapiens
<i>lac</i>	lactose
LB	lysogeny broth
MES	2-(N-morpholino)ethanesulfonic acid
mPIC	minimal pre-initiation complex
mRNA	messenger RNA
N-terminus	amino-terminus
Ni-NTA	Nickel-nitrilotriacetic acid
NLS	nuclear localization signal

NTCA	non specific transcriptional activity
Nt-NLS	N-terminal NLS
NTP	nucleotide triphosphate
PDB	Protein Data Bank
<i>P.fu</i>	<i>Pyrococcus furiosus</i>
PI	protease inhibitor
PIC	pre-initiation complex
PMSF	phenylmethylsulfonyl fluoride
Pol	DNA dependent RNA polymerase
RNA	ribonucleic acid
RNAP	DNA dependent RNA polymerase
Rpb	prefix of RNAP II subunits
rRNA	ribosomal RNA
RT	room temperature
<i>S.cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
SDS-PAGE	sodium dodecylsulfate polyacrylamide gel electrophoresis
SLS	Swiss Light Source
SNR	signal-to-noise ratio
snRNA	small nuclear RNA
snoRNA	small nucleolar RNA
<i>S.shibatae</i>	<i>Sulfolobus shibatae</i>
<i>S.solfataricus</i>	<i>Sulfolobus solfataricus</i>
TBE	Tris/Borate/EDTA
TBP	TATA-box binding protein
TF	transcription factor
Tris	trishydroxymethylaminomethane
tRNA	transfer RNA
UTP	uridine triphosphate
WH	winged helix
YPD	yeast peptone dextrose

References

- Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* *66*, 213-221.
- Andrau, J.C., Sentenac, A., and Werner, M. (1999). Mutagenesis of yeast TFIIIB70 reveals C-terminal residues critical for interaction with TBP and C34. *J Mol Biol* *288*, 511-520.
- Armache, K.-J., Mitterweger, S., Meinhart, A., and Cramer, P. (2005). Structures of complete RNA polymerase II and its subcomplex Rpb4/7. *J Biol Chem* *280*, 7131-7134.
- Balaeff, A., Mahadevan, L., and Schulten, K. (2004). Structural basis for cooperative DNA binding by CAP and lac repressor. *Structure* *12*, 123-132.
- Baliga, N.S., Goo, Y.A., Ng, W.V., Hood, L., Daniels, C.J., and DasSarma, S. (2000). Is gene expression in *Halobacterium* NRC-1 regulated by multiple TBP and TFB transcription factors? *Mol Microbiol* *36*, 1184-1185.
- Bardeleben, C., Kassavetis, G.A., and Geiduschek, E.P. (1994). Encounters of *Saccharomyces cerevisiae* RNA polymerase III with its transcription factors during RNA chain elongation. *J Mol Biol* *235*, 1193-1205.
- Bartholomew, B., Durkovich, D., Kassavetis, G.A., and Geiduschek, E.P. (1993). Orientation and topography of RNA polymerase III in transcription complexes. *Mol Cell Biol* *13*, 942-952.
- Bell, C.E., and Lewis, M. (2001). The Lac repressor: a second generation of structural and functional studies. *Curr Opin Struct Biol* *11*, 19-25.
- Berg, J.M., Tymoczko, J. L., Stryer, L. (2007). *Biochemistry*, 6 edn (New York, W. H. Freeman and Company).
- Boguta, M., Czerska, K., and Zoladek, T. (1997). Mutation in a new gene MAF1 affects tRNA suppressor efficiency in *Saccharomyces cerevisiae*. *Gene* *185*, 291-296.
- Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* *72*, 248-254.
- Brueckner, F. (2008). Molecular basis of translocation, alpha-amanitin inhibition, and CPD damage recognition by RNA polymerase II. In *Gene Center* (Munich, LMU).
- Brueckner, F., Hennecke, U., Carell, T., and Cramer, P. (2007). CPD damage recognition by transcribing RNA polymerase II. *Science* *315*, 859-862.
- Brueckner, F., Ortiz, J., and Cramer, P. (2009). A movie of the RNA polymerase nucleotide addition cycle. *Curr Opin Struct Biol* *19*, 294-299.
- Brun, I., Sentenac, A., and Werner, M. (1997). Dual role of the C34 subunit of RNA polymerase III in transcription initiation. *EMBO J* *16*, 5730-5741.
- Bushnell, D.A., Westover, K.D., Davis, R.E., and Kornberg, R.D. (2004). Structural basis of transcription: an RNA polymerase II-TFIIIB cocrystal at 4.5 Angstroms. *Science* *303*, 983-988.

- Cabart, P., Lee, J., and Willis, I.M. (2008). Facilitated recycling protects human RNA polymerase III from repression by Maf1 in vitro. *J Biol Chem* *283*, 36108-36117.
- Carey, M.F., Peterson, C.L., Smale, S.T. (2009). *Transcriptional Regulation in Eukaryotes* (Cold Spring Harbor Laboratory Press).
- Carter, R., and Drouin, G. (2009). The increase in the number of subunits in eukaryotic RNA polymerase III relative to RNA polymerase II is due to the permanent recruitment of general transcription factors. *Mol Biol Evol*.
- Carter, R., and Drouin, G. (2010). The increase in the number of subunits in eukaryotic RNA polymerase III relative to RNA polymerase II is due to the permanent recruitment of general transcription factors. *Mol Biol Evol* *27*, 1035-1043.
- Chedin, S., Riva, M., Schultz, P., Sentenac, A., and Carles, C. (1998). The RNA cleavage activity of RNA polymerase III is mediated by an essential TFIIIS-like subunit and is important for transcription termination. *Genes Dev* *12*, 3857-3871.
- Chen, C.Y., Chang, C.C., Yen, C.F., Chiu, M.T., and Chang, W.H. (2009). Mapping RNA exit channel on transcribing RNA polymerase II by FRET analysis. *Proc Natl Acad Sci U S A* *106*, 127-132.
- Chen, H.-T., Warfield, L., and Hahn, S. (2007). The positions of TFIIIF and TFIIIE in the RNA polymerase II transcription initiation complex. *Nat Struct Mol Biol* *8*, 696-703.
- Chen, H.T., and Hahn, S. (2003). Binding of TFIIIB to RNA polymerase II: Mapping the binding site for the TFIIIB zinc ribbon domain within the preinitiation complex. *Mol Cell* *12*, 437-447.
- Chen, J.Z., and Grigorieff, N. (2007). SIGNATURE: a single-particle selection system for molecular electron microscopy. *J Struct Biol* *157*, 168-173.
- Chen, Z.A., Jawhari, A., Fischer, L., Buchen, C., Tahir, S., Kamenski, T., Rasmussen, M., Lariviere, L., Bukowski-Wills, J.C., Nilges, M., *et al.* (2010). Architecture of the RNA polymerase II-TFIIIF complex revealed by cross-linking and mass spectrometry. *EMBO J* *29*, 717-726.
- Chiu, Y.H., Macmillan, J.B., and Chen, Z.J. (2009). RNA polymerase III detects cytosolic DNA and induces type I interferons through the RIG-I pathway. *Cell* *138*, 576-591.
- Cramer, P. (2002). Multisubunit RNA polymerases. *Curr Op Struct Biol* *12*, 89-97.
- Cramer, P. (2007). Finding the right spot to start transcription. *Nat Struct Mol Biol* *14*, 686-687.
- Cramer, P., Armache, K.-J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., Damsma, G.E., Dengl, S., Geiger, G., Jasiak, A., *et al.* (2008). Structure of eukaryotic RNA polymerases. *Annu Rev Biophys* *37*, 337-352.
- Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* *292*, 1863-1876.
- Crick, F. (1970). Central dogma of molecular biology. *Nature* *227*, 561-563.
- D'Alessio, J.A., Wright, K.J., and Tjian, R. (2009). Shifting players and paradigms in cell-specific transcription. *Mol Cell* *36*, 924-931.
- Damsma, G.E., Alt, A., Brueckner, F., Carell, T., and Cramer, P. (2007). Mechanism of transcriptional stalling at cisplatin-damaged DNA. *Nat Struct Mol Biol* *14*, 1127-1133.

- Damsma, G.E., and Cramer, P. (2009). Molecular basis of transcriptional mutagenesis at 8-oxoguanine. *J Biol Chem* *284*, 31658-31663.
- Dephoure, N., Zhou, C., Villen, J., Beausoleil, S.A., Bakalarski, C.E., Elledge, S.J., and Gygi, S.P. (2008). A quantitative atlas of mitotic phosphorylation. *Proc Natl Acad Sci U S A* *105*, 10762-10767.
- Desai, N., Lee, J., Upadhyaya, R., Chu, Y., Moir, R.D., and Willis, I.M. (2005). Two steps in Maf1-dependent repression of transcription by RNA polymerase III. *J Biol Chem* *280*, 6455-6462.
- Eichner, J., Chen, H.T., Warfield, L., and Hahn, S. (2010). Position of the general transcription factor TFIIF within the RNA polymerase II transcription preinitiation complex. *EMBO J* *29*, 706-716.
- Elmlund, D., and Elmlund, H. (2009). High-resolution single-particle orientation refinement based on spectrally self-adapting common lines. *J Struct Biol* *167*, 83-94.
- Elmlund, H., Baraznenok, V., Linder, T., Szilagy, Z., Rofougaran, R., Hofer, A., Hebert, H., Lindahl, M., and Gustafsson, C.M. (2009). Cryo-EM reveals promoter DNA binding and conformational flexibility of the general transcription factor TFIID. *Structure* *17*, 1442-1452.
- Elmlund, H., Lundqvist, J., Al-Karadaghi, S., Hansson, M., Hebert, H., and Lindahl, M. (2008). A new cryo-EM single-particle ab initio reconstruction method visualizes secondary structure elements in an ATP-fueled AAA+ motor. *J Mol Biol* *375*, 934-947.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* *60*, 2126-2132.
- Evans (1993). SCALA. Proceedings of CCP4 study weekend.
- Fernandez-Tornero, C., Bottcher, B., Riva, M., Carles, C., Steuerwald, U., Ruigrok, R.W., Sentenac, A., Muller, C.W., and Schoehn, G. (2007). Insights into Transcription Initiation and Termination from the Electron Microscopy Structure of Yeast RNA Polymerase III. *Mol Cell* *25*, 813-823.
- Ferri, M.L., Peyroche, G., Siaux, M., Lefebvre, O., Carles, C., Conesa, C., and Sentenac, A. (2000). A novel subunit of yeast RNA polymerase III interacts with the TFIIB-related domain of TFIIB70. *Mol Cell Biol* *20*, 488-495.
- Fiala, G., Stetter, K.O. (1986). *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaeobacteria growing optimally at 100°C. *Arch Microbiol* *145*, 56-61.
- Frank, J. (2006). Three-Dimensional Electron Microscopy of Macromolecular Assemblies.
- Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., and Leith, A. (1996). SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J Struct Biol* *116*, 190-199.
- Gaspari, M., Larsson, N.G., and Gustafsson, C.M. (2004). The transcription machinery in mammalian mitochondria. *Biochim Biophys Acta* *1659*, 148-152.
- Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dumpelfeld, B., *et al.* (2006). Proteome survey reveals modularity of the yeast cell machinery. *Nature* *440*, 631-636.
- Geiduschek, E.P., and Kassavetis, G.A. (2001). The RNA polymerase III transcription apparatus. *J Mol Biol* *310*, 1-26.

- Geiduschek, E.P., and Kassavetis, G.A. (2006). Transcription: adjusting to adversity by regulating RNA polymerase. *Curr Biol* *16*, R849-851.
- Geiger, J.H., Hahn, S., Lee, S., and Sigler, P.B. (1996). Crystal structure of the yeast TFIIA/TBP/DNA complex. *Science* *272*, 830-836.
- Geiger, S.R., Kuhn, C.D., Leidig, C., Renkawitz, J., and Cramer, P. (2008). Crystallization of RNA polymerase I subcomplex A14/A43 by iterative prediction, probing and removal of flexible regions. *Acta Crystallogr Sect F Struct Biol Cryst Commun* *64*, 413-418.
- Goede, B., Naji, S., von Kampen, O., Ilg, K., and Thomm, M. (2006). Protein-protein interactions in the archaeal transcriptional machinery: binding studies of isolated RNA polymerase subunits and transcription factors. *J Biol Chem* *281*, 30581-30592.
- Grummt, I. (2003). Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus. *Genes Dev* *17*, 1691-1702.
- Hager, G.L., McNally, J.G., and Misteli, T. (2009). Transcription dynamics. *Mol Cell* *35*, 741-753.
- Hahn, S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol* *11*, 394-403.
- Hethke, C., Geerling, A.C., Hausner, W., de Vos, W.M., and Thomm, M. (1996). A cell-free transcription system for the hyperthermophilic archaeon *Pyrococcus furiosus*. *Nucleic Acids Res* *24*, 2369-2376.
- Hirata, A., Klein, B.J., and Murakami, K.S. (2008). The X-ray crystal structure of RNA polymerase from Archaea. *Nature* *451*, 851-854.
- Hirata, A., and Murakami, K.S. (2009). Archaeal RNA polymerase. *Curr Opin Struct Biol* *19*, 724-731.
- Hirtreiter, A., Damsma, G.E., Cheung, A.C., Klose, D., Grohmann, D., Vojnic, E., Martin, A.C., Cramer, P., and Werner, F. (2010). Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. *Nucleic Acids Res*.
- Ho, L., and Crabtree, G.R. (2010). Chromatin remodelling during development. *Nature* *463*, 474-484.
- Hohn, M., Tang, G., Goodyear, G., Baldwin, P.R., Huang, Z., Penczek, P.A., Yang, C., Glaeser, R.M., Adams, P.D., and Ludtke, S.J. (2007). SPARX, a new environment for Cryo-EM image processing. *J Struct Biol* *157*, 47-55.
- Holm, L., and Park, J. (2000). DaliLite workbench for protein structure comparison. *Bioinformatics* *16*, 566-567.
- Jasiak, A.J., Armache, K.J., Martens, B., Jansen, R.P., and Cramer, P. (2006). Structural biology of RNA polymerase III: subcomplex C17/25 X-ray structure and 11 subunit enzyme model. *Mol Cell* *23*, 71-81.
- Jawhari, A., Uhring, M., De Carlo, S., Crucifix, C., Tocchini-Valentini, G., Moras, D., Schultz, P., and Poterszman, A. (2006). Structure and oligomeric state of human transcription factor TFIIE. *EMBO Rep* *7*, 500-505.
- Juo, Z.S., Kassavetis, G.A., Wang, J., Geiduschek, E.P., and Sigler, P.B. (2003). Crystal structure of a transcription factor IIIB core interface ternary complex. *Nature* *422*, 534-539.

- Juven-Gershon, T., Hsu, J.Y., Theisen, J.W., and Kadonaga, J.T. (2008). The RNA polymerase II core promoter - the gateway to transcription. *Curr Opin Cell Biol* 20, 253-259.
- Kassavetis, G.A., Prakash, P., and Shim, E. (2010). The C53/C37 subcomplex of RNA polymerase III lies near the active site and participates in promoter opening. *J Biol Chem* 285, 2695-2706.
- Kassavetis, G.A., Soragni, E., Driscoll, R., and Geiduschek, E.P. (2005). Reconfiguring the connectivity of a multiprotein complex: fusions of yeast TATA-binding protein with Brf1, and the function of transcription factor IIIB. *Proc Natl Acad Sci U S A* 102, 15406-15411.
- Kastner, B., Fischer, N., Golas, M.M., Sander, B., Dube, P., Boehringer, D., Hartmuth, K., Deckert, J., Hauer, F., Wolf, E., *et al.* (2008). GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat Methods* 5, 53-55.
- Kelleher, R.J., 3rd, Flanagan, P.M., and Kornberg, R.D. (1990). A novel mediator between activator proteins and the RNA polymerase II transcription apparatus. *Cell* 61, 1209-1215.
- Khoo, B., Brophy, B., and Jackson, S.P. (1994). Conserved functional domains of the RNA polymerase III general transcription factor BRF. *Genes Dev* 8, 2879-2890.
- Kim, B., Nesvizhskii, A.I., Rani, P.G., Hahn, S., Aebersold, R., and Ranish, J.A. (2007). The transcription elongation factor TFIIS is a component of RNA polymerase II preinitiation complexes. *Proc Natl Acad Sci U S A* 104, 16068-16073.
- Kim, T.K., Ebright, R.H., and Reinberg, D. (2000). Mechanism of ATP-dependent promoter melting by transcription factor IIH. *Science* 288, 1418-1422.
- Korkhin, Y., Unligil, U.M., Littlefield, O., Nelson, P.J., Stuart, D.I., Sigler, P.B., Bell, S.D., and Abrescia, N.G. (2009). Evolution of Complex RNA Polymerases: The Complete Archaeal RNA Polymerase Structure. *PLoS Biol* 7, e102.
- Kornberg, R.D. (1999). Eukaryotic transcriptional control. *Trends Cell Biol* 9, M46-49.
- Kornberg, R.D. (2007). The molecular basis of eukaryotic transcription. *Proceedings of the National Academy of Sciences of the United States of America* 104, 12955-12961.
- Kostek, S.A., Grob, P., De Carlo, S., Lipscomb, J.S., Garczarek, F., and Nogales, E. (2006). Molecular architecture and conformational flexibility of human RNA polymerase II. *Structure* 14, 1691-1700.
- Kostrewa, D., Zeller, M.E., Armache, K.J., Seizl, M., Leike, K., Thomm, M., and Cramer, P. (2009). RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature* 462, 323-330.
- Krissinel, E., and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* 60, 2256-2268.
- Kuhn, C.D., Geiger, S.R., Baumli, S., Gartmann, M., Gerber, J., Jennebach, S., Mielke, T., Tschochner, H., Beckmann, R., and Cramer, P. (2007). Functional architecture of RNA polymerase I. *Cell* 131, 1260-1272.
- Kusser, A.G., Bertero, M.G., Naji, S., Becker, T., Thomm, M., Beckmann, R., and Cramer, P. (2008). Structure of an archaeal RNA polymerase. *J Mol Biol* 376, 303-307.

- Landrieux, E., Alic, N., Ducrot, C., Acker, J., Riva, M., and Carles, C. (2006). A subcomplex of RNA polymerase III subunits involved in transcription termination and reinitiation. *Embo J* *25*, 118-128.
- Langer, D., Hain, J., Thuriaux, P., and Zillig, W. (1995). Transcription in archaea: similarity to that in eucarya. *Proc Natl Acad Sci U S A* *92*, 5768-5772.
- Lannutti, B.J., Persinger, J., and Bartholomew, B. (1996). Probing the protein-DNA contacts of a yeast RNA polymerase III transcription complex in a crude extract: solid phase synthesis of DNA photoaffinity probes containing a novel photoreactive deoxycytidine analog. *Biochemistry* *35*, 9821-9831.
- Lee, J., Moir, R.D., and Willis, I.M. (2009). Regulation of RNA polymerase III transcription involves SCH9-dependent and SCH9-independent branches of the target of rapamycin (TOR) pathway. *J Biol Chem* *284*, 12604-12608.
- Leschziner, A.E., and Nogales, E. (2007). Visualizing flexibility at molecular resolution: analysis of heterogeneity in single-particle electron microscopy reconstructions. *Annu Rev Biophys Biomol Struct* *36*, 43-62.
- Leslie, A.G.W., Brick, P., and Wonacott, A. T. (1986). *Protein Crystallogr* *18*, 33-39.
- Lewis, M., Chang, G., Horton, N.C., Kercher, M.A., Pace, H.C., Schumacher, M.A., Brennan, R.G., and Lu, P. (1996). Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* *271*, 1247-1254.
- Liu, X., Bushnell, D.A., Wang, D., Calero, G., and Kornberg, R.D. (2009). Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism. *Science* *327*, 206-209.
- Lorenzen, K., Vannini, A., Cramer, P., and Heck, A.J. (2007). Structural biology of RNA polymerase III: mass spectrometry elucidates subcomplex architecture. *Structure* *15*, 1237-1245.
- Ludtke, S.J., Baldwin, P.R., and Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* *128*, 82-97.
- Matzke, M., Kanno, T., Daxinger, L., Huettel, B., and Matzke, A.J. (2009). RNA-mediated chromatin-based silencing in plants. *Curr Opin Cell Biol* *21*, 367-376.
- Miller, G., and Hahn, S. (2006). A DNA-tethered cleavage probe reveals the path for promoter DNA in the yeast preinitiation complex. *Nat Struct Mol Biol* *13*, 603-610.
- Minakhin, L., Bhagat, S., Brunning, A., Campbell, E.A., Darst, S.A., Ebright, R.H., and Severinov, K. (2001). Bacterial RNA polymerase subunit omega and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. *Proc Natl Acad Sci U S A* *98*, 892-897.
- Moir, R.D., Lee, J., Haeusler, R.A., Desai, N., Engelke, D.R., and Willis, I.M. (2006). Protein kinase A regulates RNA polymerase III transcription through the nuclear localization of Maf1. *Proc Natl Acad Sci U S A* *103*, 15044-15049.
- Moqtaderi, Z., Bai, Y., Poon, D., Weil, P.A., and Struhl, K. (1996). TBP-associated factors are not generally required for transcriptional activation in yeast. *Nature* *383*, 188-191.
- Naji, S., Grunberg, S., and Thomm, M. (2007). The RPB7 orthologue E' is required for transcriptional activity of a reconstituted archaeal core enzyme at low temperatures and stimulates open complex formation. *J Biol Chem* *282*, 11047-11057.

- Nather, D.J., Rachel, R., Wanner, G., and Wirth, R. (2006). Flagella of *Pyrococcus furiosus*: multifunctional organelles, made for swimming, adhesion to various surfaces, and cell-cell contacts. *J Bacteriol* *188*, 6915-6923.
- Nealson, K.H., and Conrad, P.G. (1999). Life: past, present and future. *Philos Trans R Soc Lond B Biol Sci* *354*, 1923-1939.
- Nikolov, D.B., Chen, H., Halay, E.D., Usheva, A.A., Hisatake, K., Lee, D.K., Roeder, R.G., and Burley, S.K. (1995). Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature* *377*, 119-128.
- Oficjalska-Pham, D., Harismendy, O., Smagowicz, W.J., Gonzalez de Peredo, A., Boguta, M., Sentenac, A., and Lefebvre, O. (2006). General repression of RNA polymerase III transcription is triggered by protein phosphatase type 2A-mediated dephosphorylation of Maf1. *Mol Cell* *22*, 623-632.
- Okuda, M., Watanabe, Y., Okamura, H., Hanaoka, F., Ohkuma, Y., and Nishimura, Y. (2000). Structure of the central core domain of TFIIEbeta with a novel double-stranded DNA-binding surface. *Embo J* *19*, 1346-1356.
- Ouhammouch, M. (2004). Transcriptional regulation in Archaea. *Curr Opin Genet Dev* *14*, 133-138.
- Paule, M.R., and White, R.J. (2000). Survey and summary: transcription by RNA polymerases I and III. *Nucleic Acids Res* *28*, 1283-1298.
- Pluta, K., Lefebvre, O., Martin, N.C., Smagowicz, W.J., Stanford, D.R., Ellis, S.R., Hopper, A.K., Sentenac, A., and Boguta, M. (2001). Maf1p, a negative effector of RNA polymerase III in *Saccharomyces cerevisiae*. *Mol Cell Biol* *21*, 5031-5040.
- Poole, F.L., 2nd, Gerwe, B.A., Hopkins, R.C., Schut, G.J., Weinberg, M.V., Jenney, F.E., Jr., and Adams, M.W. (2005). Defining genes in the genome of the hyperthermophilic archaeon *Pyrococcus furiosus*: implications for all microbial genomes. *J Bacteriol* *187*, 7325-7332.
- Ransom, M., Dennehey, B.K., and Tyler, J.K. (2010). Chaperoning histones during DNA replication and repair. *Cell* *140*, 183-195.
- Ream, T.S., Haag, J.R., Wierzbicki, A.T., Nicora, C.D., Norbeck, A.D., Zhu, J.K., Hagen, G., Guilfoyle, T.J., Pasa-Tolic, L., and 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*, 192-203.
- Reimer, R., and Kolh, H. (2008). Transmission Electron Microscopy.
- Reina, J.H., Azzouz, T.N., and Hernandez, N. (2006). Maf1, a new player in the regulation of human RNA polymerase III transcription. *PLoS One* *1*, e134.
- Reinberg, D., Orphanides, G., Ebright, R., Akoulitchev, S., Carcamo, J., Cho, H., Cortes, P., Drapkin, R., Flores, O., Ha, I., *et al.* (1998). The RNA polymerase II general transcription factors: past, present, and future. *Cold Spring Harb Symp Quant Biol* *63*, 83-103.
- Roberts, D.N., Wilson, B., Huff, J.T., Stewart, A.J., and Cairns, B.R. (2006). Dephosphorylation and genome-wide association of Maf1 with Pol III-transcribed genes during repression. *Mol Cell* *22*, 633-644.
- Roeder, R. (1996). The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci* *21*, 327-335.

- Sadhale, P.P., and Woychik, N.A. (1994). C25, an essential RNA polymerase III subunit related to the RNA polymerase II subunit RPB7. *Mol Cell Biol* *14*, 6164-6170.
- Sahar, S., and Sassone-Corsi, P. (2009). Metabolism and cancer: the circadian clock connection. *Nat Rev Cancer* *9*, 886-896.
- Sander, B., Golas, M.M., and Stark, H. (2005). Advantages of CCD detectors for de novo three-dimensional structure determination in single-particle electron microscopy. *J Struct Biol* *151*, 92-105.
- Schopf, J.W., and Packer, B.M. (1987). Early Archean (3.3-billion to 3.5-billion-year-old) microfossils from Warrawoona Group, Australia. *Science* *237*, 70-73.
- Schramm, L., and Hernandez, N. (2002). Recruitment of RNA polymerase III to its target promoters. *Genes Dev* *16*, 2593-2620.
- Schultz, P., Fribourg, S., Poterszman, A., Mallouh, V., Moras, D., and Egly, J.M. (2000). Molecular structure of human TFIIH. *Cell* *102*, 599-607.
- Seidelt, B., Innis, C.A., Wilson, D.N., Gartmann, M., Armache, J.P., Villa, E., Trabuco, L.G., Becker, T., Mielke, T., Schulten, K., *et al.* (2009). Structural insight into nascent polypeptide chain-mediated translational stalling. *Science* *326*, 1412-1415.
- Shor, B., Wu, J., Shakey, Q., Toral-Barza, L., Shi, C., Follettie, M., and Yu, K. (2010). Requirement of the mTOR kinase for the regulation of Maf1 phosphorylation and control of RNA polymerase III-dependent transcription in cancer cells. *J Biol Chem*.
- Sousa, R., Chung, Y.J., Rose, J.P., and Wang, B.-C. (1993). Crystal structure of bacteriophage T7 RNA polymerase at 3.3Å resolution. *Nature* *364*, 593-599.
- Steitz, T.A. (2009). The structural changes of T7 RNA polymerase from transcription initiation to elongation. *Curr Opin Struct Biol* *19*, 683-690.
- Stetter, K.O. (2006). Hyperthermophiles in the history of life. *Philos Trans R Soc Lond B Biol Sci* *361*, 1837-1842; discussion 1842-1833.
- Svejstrup, J.Q. (2004). The RNA polymerase II transcription cycle: cycling through chromatin. *Biochim Biophys Acta* *1677*, 64-73.
- Sydow, J.F., Brueckner, F., Cheung, A.C., Damsma, G.E., Dengl, S., Lehmann, E., Vassylyev, D., and Cramer, P. (2009). Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. *Mol Cell* *34*, 710-721.
- Takagi, Y., and Kornberg, R.D. (2006). Mediator as a general transcription factor. *J Biol Chem* *281*, 80-89.
- Tan, S., Hunziker, Y., Sargent, D.F., and Richmond, T.J. (1996). Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature* *381*, 127-151.
- Tang, G.Q., Roy, R., Bandwar, R.P., Ha, T., and Patel, S.S. (2009). Real-time observation of the transition from transcription initiation to elongation of the RNA polymerase. *Proc Natl Acad Sci U S A* *106*, 22175-22180.
- Terwilliger, T.C. (2003). SOLVE and RESOLVE: automated structure solution and density modification. *Methods Enzymol* *374*, 22-37.
- Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. *Acta Crystallogr D Biol Crystallogr* *55* (Pt 4), 849-861.

- Thomm, M. (1996). Archaeal transcription factors and their role in transcription initiation. *FEMS Microbiol Rev* 18, 159-171.
- Thuillier, V., Stettler, S., Sentenac, A., Thuriaux, P., and Werner, M. (1995). A mutation in the C31 subunit of *Saccharomyces cerevisiae* RNA polymerase III affects transcription initiation. *Embo J* 14, 351-359.
- Tsai, F.T.F., Sigler, P.B. (2000). Structural basis of preinitiation complex assembly on human Pol II promoters. *EMBO Journal* 19, 25-36.
- Tyree, C.M., George, C.P., Lira-DeVito, L.M., Wampler, S.L., Dahmus, M.E., Zawel, L., and Kadonaga, J.T. (1993). Identification of a minimal set of proteins that is sufficient for accurate initiation of transcription by RNA polymerase II. *Genes Dev* 7, 1254-1265.
- Upadhyya, R., Lee, J., and Willis, I.M. (2002). Maf1 is an essential mediator of diverse signals that repress RNA polymerase III transcription. *Mol Cell* 10, 1489-1494.
- Vassilyev, D.G., Sekine, S., Laptenko, O., Lee, J., Vassilyeva, M.N., Borukhov, S., and Yokoyama, S. (2002). Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* 417, 712-719.
- Venters, B.J., and Pugh, B.F. (2009). How eukaryotic genes are transcribed. *Crit Rev Biochem Mol Biol* 44, 117-141.
- Wang, D., Bushnell, D.A., Huang, X., Westover, K.D., Levitt, M., and Kornberg, R.D. (2009). Structural basis of transcription: backtracked RNA polymerase II at 3.4 angstrom resolution. *Science* 324, 1203-1206.
- Wang, Z., and Roeder, R.G. (1997). Three human RNA polymerase III-specific subunits form a subcomplex with a selective function in specific transcription initiation. *Genes Dev* 11, 1315-1326.
- Warner, J.R. (1999). The economics of ribosome biosynthesis in yeast. *Trends Biochem Sci* 24, 437-440.
- Watson, J.D., Crick, F.H.C. (1953). Molecular Structure of Nucleic Acids. *Nature* 171.
- Werner, F. (2007). Structure and function of archaeal RNA polymerases. *Mol Microbiol* 65, 1395-1404.
- Werner, F., and Weinzierl, R.O. (2005). Direct modulation of RNA polymerase core functions by basal transcription factors. *Mol Cell Biol* 25, 8344-8355.
- Werner, M., Chaussivert, N., Willis, I.M., and Sentenac, A. (1993). Interaction between a complex of RNA polymerase III subunits and the 70-kDa component of transcription factor IIIB. *J Biol Chem* 268, 20721-20724.
- Werner, M., Hermann-Le Denmat, S., Treich, I., Sentenac, A., and Thuriaux, P. (1992). Effect of mutations in a zinc-binding domain of yeast RNA polymerase C (III) on enzyme function and subunit association. *Mol Cell Biol* 12, 1087-1095.
- Werner, M., Thuriaux, P., and Soutourina, J. (2009). Structure-function analysis of RNA polymerases I and III. *Curr Opin Struct Biol* 19, 740-745.
- Williams, D.B., and Carter, C.B. (2009). *Transmission Electron Microscopy*.
- Willis, I.M., Desai, N., and Upadhyya, R. (2004). Signaling repression of transcription by RNA polymerase III in yeast. *Prog Nucleic Acid Res Mol Biol* 77, 323-353.

Woese, C.R., Kandler, O., and Wheelis, M.L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* 87, 4576-4579.

Zhang, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K., and Darst, S.A. (1999). Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* 98, 811-824.

Zhou, H., Liu, Q., Gao, Y., Teng, M., and Niu, L. (2009). Crystal structure of NusG N-terminal (NGN) domain from *Methanocaldococcus jannaschii* and its interaction with rpoE". *Proteins* 76, 787-793.

Curriculum vitae

Anselm Kusser

Personal details:

Name: Gregor Anselm Kusser
 Date of birth: 17th September 1979
 Place of birth: Munich
 Nationality: German
 Eye color: brown
 Marital status: unmarried
 Foreign languages: English, French



Curriculum vitae:

1986 - 1990: Montessori elementary school, Munich-Grosshadern
 1990 - 1999: St.-Anna-Gymnasium, Munich-Lehel
 5/1999: Qualification of University matriculation
 9/1999 - 8/2000: Civilian service at the Montessori elementary school, Munich-Grosshadern
 10/2000 - 4/2006: Diploma course in Bioinformatics (LMU/TU Munich)
 5/2006: Diploma degree in Bioinformatics (passed with merit)
 Since 7/2006: PhD Thesis at the Gene Center Munich, group of Prof. Dr. Patrick Cramer
 7/2010: PhD defense

Courses and conferences:

11/2007: Workshop: Advanced topics in 3D Cryo-EM, Scripps, USA
Poster: Cryo-EM structure of the P.fu RNAP

06/2008: EMBO-Meeting on Transcription in Yeast, St. Feliu, Spain
Poster: Cryo-EM structure of the P.fu RNAP

09/2009: 3d repertoire final meeting, Warsaw, Poland

06/2010: 3D EM Gordon Conference, Il Ciocco, Italy
Poster: Molecular basis of RNA polymerase III transcription repression by Maf1