

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

**Structure and Mechanism of the RNA
Polymerase II CTD Phosphatase Scp1
and Large-scale Preparation of the RNA
Polymerase II-TFIIF Complex**



Tomislav Kamenski
aus Zagreb, Kroatien

2006

Erklärung

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

Ehrenwörtliche Versicherung

Diese Dissertation wurde selbständig, ohne unerlaubte Hilfe erarbeitet.

München, den 2. Mai 2006

Dissertation eingereicht am 4. Mai 2006

1. Gutachter: Prof. Dr. Patrick Cramer
2. Gutachter: Prof. Dr. Ralf-Peter Jansen

Mündliche Prüfung am 29. Mai 2006

Acknowledgements

First of all I owe my gratitude to my PhD supervisor Prof. Patrick Cramer for providing me excellent conditions for independent research work and for constant support.

I am most grateful to Claudia, for being a great companion on the long and difficult way of TFIIIF project and for loud laughs at the time when the project did not seem promising. Toni Meinhart, thank you for being patient with me, while learning basics of crystallography. I would also like to thank Stefan Benkert for great effort in fermenting yeast, Marian Kalocay, for providing me with useful vectors and tips.

Karim, thank you for accompanying me to fast food restaurants (willingly) after long days in the lab and Michela for good laughs and late evening jogging rounds. To Hubert, Florian, Laurent and other lab members, I am grateful for being great colleagues, always willing to help, and for nice atmosphere in the laboratory. I would like to thank especially Kathrin Gmelch for being an excellent diploma student, as well as Susanna Heilmeier for introducing me to phosphatase project. I am grateful to Anass for great collaboration, exchange of scientific views and I wish him all the luck in the continuation of the project.

I would like to thank Katja Sträßer and Heidi Feldmann for providing great advice and being ready to teach me about yeast.

I would also like to thank members of my PhD thesis committee, Alexander Brehm and Ralf-Peter Jansen for a lot of advice.

I wish to thank Ita Gruić, for her expertise on RNA biochemistry, as well as Prof. Ivana Weygand-Đurašević for support even while being abroad.

I thank Boehringer Ingelheim Fonds for generous financial support and possibility to meet other young motivated scientists with similar interests.

Finally, I am deeply indebted to my family for all their help and support during my education.

Summary

1 Introduction

1.1	Eukaryotic RNA polymerases	1
1.2	Structure of Pol II	2
1.3	Eukaryotic transcription cycle.....	3
1.4	TFIIF is a unique eukaryal general transcription factor	9
1.4.1	Roles of TFIIF during PIC formation and promoter escape	13
1.4.2	Roles of TFIIF in elongation	14
1.4.3	TFIIF interactions with Pol II	14
1.5	Carboxy-terminal domain of Pol II (CTD)	16
1.6	CTD phosphorylation and transcriptional regulation	17
1.7	CTD kinases	18
1.8	CTD phosphatase Ssu72	19
1.9	CTD phosphatases of the Fcp1 family	21
1.10	Aims of this work.....	24

2 Materials and Methods

2.1	Bacterial and yeast strains	25
2.2	Plasmids	26

Table of Contents

2.3	Media and supplements	26
2.4	Buffers and solutions.....	27
2.5	Bioinformatic tools.....	29
2.6	Molecular cloning techniques.....	29
2.7	Preparation of competent cells.....	31
2.8	Electrophoretic methods	31
2.9	Preparation of samples for Edman sequencing	32
2.10	Protein expression in <i>E. coli</i> and selenomethionine labeling.....	32
2.11	Cloning and protein purification of CTD phosphatases	33
2.12	Limited Proteolysis.....	34
2.13	Crystallization and structure determination	35
2.14	Phosphatase kinetics and inhibitor studies	36
2.15	Activity assay with active site mutants	37
2.16	Purification of Pol II core enzyme.....	37
2.17	Purification of Rpb4/7 subcomplex.....	39
2.18	Reconstitution of a Pol II-Fcp1 complex.....	40
2.19	Purification of TFIIF variants expressed in <i>E. coli</i>	41
2.20	Generation of a yeast strain overexpressing TFIIF	44
2.21	Isolation of genomic DNA from yeast.....	45

Table of Contents

2.22	Yeast transformation	45
2.23	Small-scale Pol II-TFIIF tandem affinity purification (TAP)	46
2.24	Dot blot assay	47
2.25	Large-scale isolation of Pol II-TFIIF complex from yeast	48
2.25.1	Yeast fermentation.....	48
2.25.2	Harvesting and storage of yeast	49
2.25.3	Large-scale purification of the Pol II-TFIIF complex.....	49
2.26	Purification of TFIIB.....	50
2.27	Purification of TBP core	51
2.28	Electrophoretic mobility shift assay	52
2.29	Assembly of an initially transcribing complex	52
2.30	Isolation of RNA from protein preparations	53
2.31	Isolation of RNA from polyacrylamide gel	53
2.32	Identification of RNA	54
3	Results and Discussion	
3.1	Structure and mechanism of Pol II CTD phosphatases.....	56
3.1.1	Domain organization and structure determination	56
3.1.2	Overall structure	60
3.1.3	Phosphatase activity.....	62

Table of Contents

3.1.4	Specific inhibition.....	62
3.1.5	Active center.....	63
3.1.6	Catalytic mechanism	65
3.1.7	CTD specificity.....	67
3.1.8	The Pol II subcomplex Rpb4/7 recruits Fcp1	68
3.2	Preparation of the Pol II-TFIIF complex	72
3.2.1	Expression of TFIIF in <i>E. coli</i>	72
3.2.2	Isolation of TFIIF from yeast.....	74
3.2.3	Overexpression of TFIIF in yeast.....	75
3.2.4	Assembly of an initially transcribing complex (ITC)	79
3.2.5	Presence of 5.8S rRNA in Pol II-TFIIF preparations.....	81
	Conclusions.....	83
	References.....	85

Summary

TFIIF is the only general transcription factor that has been implicated in the preinitiation complex assembly, open complex formation, initiation and transcription elongation. In addition, TFIIF stimulates Fcp1, a central phosphatase needed for recycling of RNA polymerase II (Pol II) after transcription by dephosphorylation of the Pol II C-terminal domain (CTD). This thesis reports the X-ray structure of the small CTD phosphatase Scp1, which is homologous to the Fcp1 catalytic domain. The structure shows a core fold and an active center similar to phosphotransferases and -hydrolases that solely share a DXDX(V/T) signature motif with Fcp1/Scp1. It was further demonstrated that the first aspartate in the signature motif undergoes metal-assisted phosphorylation during catalysis, resulting in a phosphoaspartate intermediate that was structurally mimicked with the inhibitor beryllofluoride. Specificity may result from CTD binding to a conserved hydrophobic pocket between the active site and an insertion domain that is unique to Fcp1/Scp1. Fcp1 specificity may additionally arise from phosphatase recruitment near the CTD via the Pol II subcomplex Rpb4/7, which is shown to be required for Fcp1 binding to the polymerase *in vitro*. Until now, the main impediment in the high resolution crystallographic studies of TFIIF in complex with Pol II and other members of transcription machinery was unavailability of soluble, stoichiometric TFIIF complex in sufficient amounts. This thesis reports on the development of the overexpression system in yeast and a purification protocol that enabled for the first time to isolate milligram amounts of a pure and soluble, 15-subunit (~0,7 MDa) stoichiometric Pol II-TFIIF complex. Such complex together with the promoter DNA, RNA, TBP and TFIIB assembles *in vitro* into the yeast initially transcribing complex, which can now be studied structurally.

Parts of this work have been published:

- Kamenski, T., S. Heilmeyer, A. Meinhart, and P. Cramer. 2004. Structure and mechanism of RNA polymerase II CTD phosphatases. *Mol Cell* **15**: 399-407.
- Meinhart, A., T. Kamenski, S. Hoepfner, S. Baumli, and P. Cramer. 2005. A structural perspective of CTD function. *Genes Dev* **19**: 1401-15.

1 Introduction

1.1 Eukaryotic RNA polymerases

In bacteria and archaea, a single kind of RNA polymerase synthesizes all of the cell's RNA except the RNA primers employed in DNA replication. Eukaryotes, however, generally possess three different nuclear RNA polymerases, each responsible for synthesis of a different class of RNAs.

RNA polymerase I (Pol I) is located in the nucleoli and synthesizes precursors of most ribosomal RNAs (rRNAs). RNA polymerase II (Pol II) occurs in the nucleoplasm and synthesizes mRNA precursors from protein-coding genes. RNA polymerase III (Pol III) also occurs in the nucleoplasm and synthesizes the precursors of 5S ribosomal RNA, the tRNAs, and a variety of other small nuclear and cytosolic RNAs. Recently discovered in plants, RNA polymerase IV (Pol IV) helps produce siRNAs that target de novo cytosine methylation events that play role in heterochromatin formation (Onodera *et al.*, 2005).

The central enzyme of eukaryotic transcription of all protein-coding genes is Pol II. It interacts with general transcription factors within the preinitiation complex (PIC), breaks these interactions upon initiation and promoter clearance and associates with a new set of factors during elongation, termination, and mRNA processing. Pol II is a multisubunit complex which consists of 12 subunits, which can be divided into three categories (Hahn, 2004). These are subunits of the core domain having homologous counterparts in bacterial Pol (Rpb1, 2, 3 and 11), subunits shared between all three nuclear polymerases (Rpb5, 6, 8, 10 and 12) and subunits specific to Pol II but not essential for transcription elongation (Rpb4, 7 and 9). The subunits that make the core of every RNA polymerase investigated thus far are homologous to subunits from all cellular RNA polymerases implying same basic structure and mechanism (Ebright, 2000).

1.2 Structure of Pol II

A wealth of information on the transcription mechanism was provided over the past years by high quality X-ray structures of bacterial polymerase (Zhang *et al.*, 1999) and its complexes with GreB (Opalka *et al.*, 2003) and σ -subunit (Mekler *et al.*, 2002; Murakami *et al.*, 2002; Vassylyev *et al.*, 2002). The hallmark of eukaryotic transcription, however, was the structure of the 10-subunit core Pol II (Cramer *et al.*, 2001). This structure served as a starting point for further structures of the two transcribing complexes (Gnatt *et al.*, 2001; Westover *et al.*, 2004), the 12-subunit Pol II structure (Armache *et al.*, 2003; Bushnell and Kornberg, 2003, Armache *et al.*, 2005), as well as its complexes with transcription factors TFIIB (Bushnell *et al.*, 2004) and TFIIS (Kettenberger *et al.*, 2003). Cryo electron microscopy complemented this data with the overall architecture of Pol II in complex with its coactivator Mediator (Davis *et al.*, 2002) and with general transcription factor TFIIF (Chung *et al.*, 2003).

Pol II consists of a folded region that is responsible for mRNA synthesis, and the mobile C-terminal domain (CTD) not seen in the electron density (Cramer *et al.*, 2001; Figure 1). The core of the Pol II enzyme comprises Rpb3, 10, 11, 12 as well as the regions of Rpb1 and 2 which form the active center (Cramer *et al.*, 2001). These subunits account for half of the mass of Pol II and are either shared or homologous between all cellular RNA polymerases. The center of the enzyme represents a deep cleft where incoming DNA enters from one side and the active site is buried at the base. The cleft is formed by four mobile elements of the enzyme, namely the core, the clamp, the shelf and the jaw-lobe that move relative to each other. It was observed in both open and closed conformation due to the clamp movement. The clamp is connected to the core through the set of flexible switches and can move by a swinging motion of up to 30 Å. In the structure of the 12-subunit Pol II (Figure 1), the clamp is locked by Rpb7 into a closed conformation (Armache *et al.*, 2003; Bushnell *et al.*, 2003). This however implies that during initiation double-stranded DNA template does not enter the active site cleft but rather the single-stranded DNA template strand is inserted deep into the cleft to reach the active site during open

complex formation. Rbp4 and 7 subunits additionally provide a binding surface for other factors and possibly for RNA exiting the elongating Pol II.

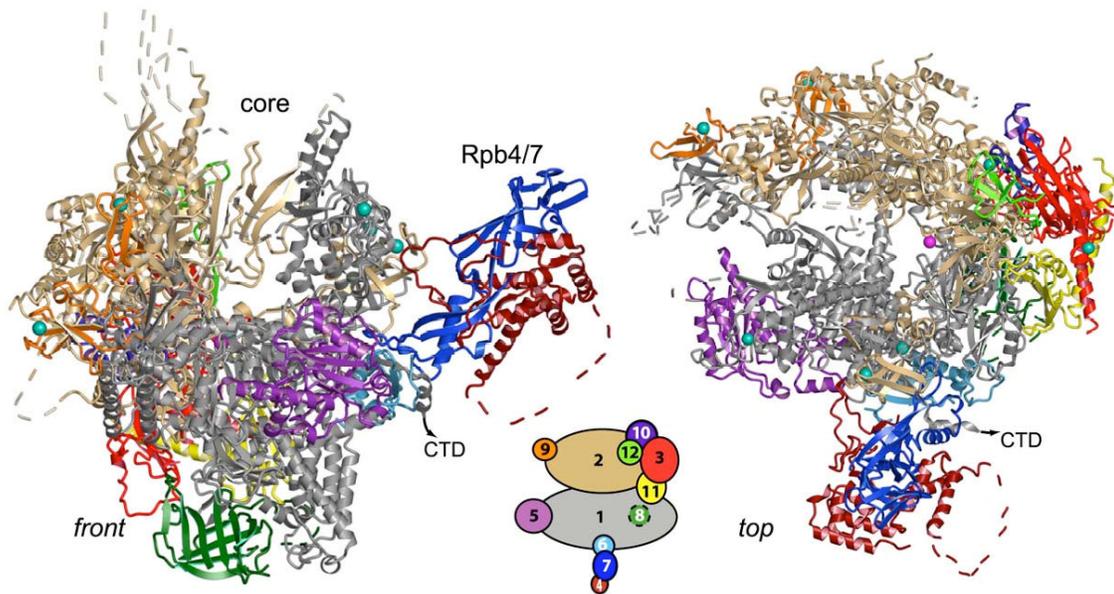


Figure 1: Complete RNA polymerase II structure. Two standard views “front” and “top” are shown. The 12 subunits Rpb1-Rpb12 are colored according to the key below the views. Dashed lines represent disordered loops. Eight zinc ions and the active site magnesium ion are depicted as cyan spheres and a pink sphere, respectively. Adopted from (Armache *et al.*, 2005).

1.3 Eukaryotic transcription cycle

The control of gene expression is achieved by a common mechanism that underlines selective binding of proteins to specific control sequences elements in order to modulate the rate of transcription initiation. Pol II, unlike bacterial RNA polymerases (Pol), possesses little if any inherent ability to bind to its promoters. In bacteria a single polypeptide, the σ subunit recognizes promoter sequences, promotes

conformational changes in the Pol-DNA complex upon initiation and directly interacts with some transcription activators. In contrast, eukaryotic Pol II and five general transcription factors (GTFs), called TFIIB, -D, -E, -F, and -H, which substitute for σ functions, position onto the core promoter in a state termed the preinitiation complex (PIC, Figure 2). Additionally TFIIA can contribute stabilizing interactions (Pugh, 2000; Bleichenbacher *et al.*, 2003). Archaea rely only on two essential general factors, TATA-binding protein (TBP) and TFB (related to the TFIIB in Pol II).

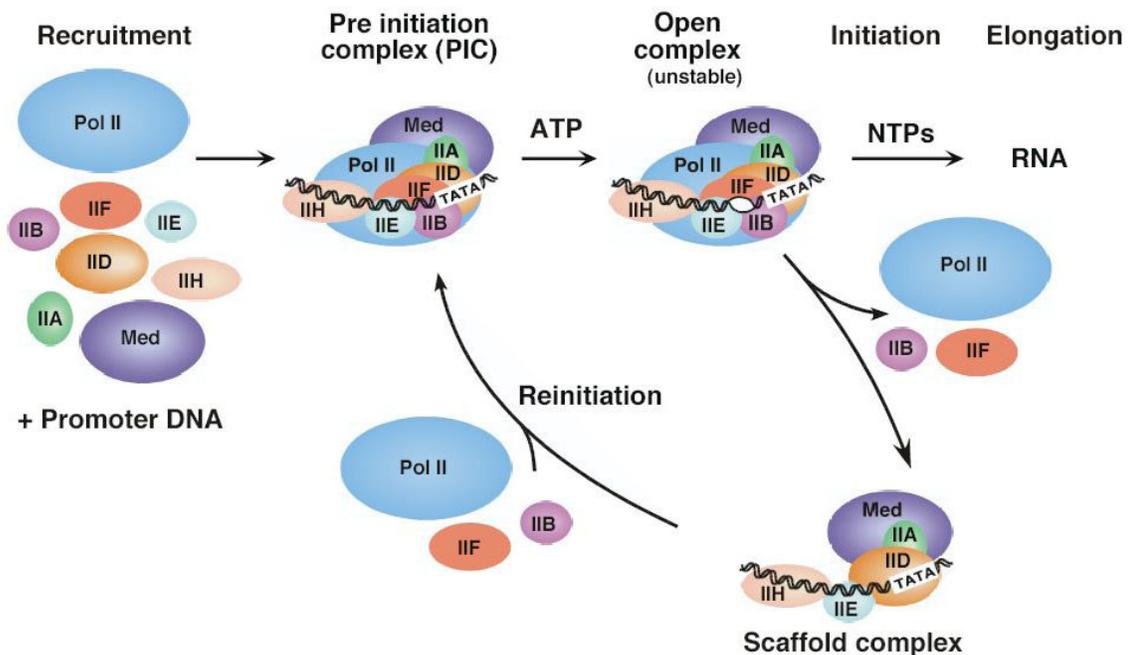


Figure 2: Pol II pathway of transcription initiation and reinitiation. Adopted from (Hahn, 2004).

TBP together with 14 TBP-associated factors (TAFs) forms the multisubunit TFIID complex. However, at some promoters TBP is sufficient for promoter recognition (Kuras *et al.*, 2000; Li *et al.*, 2000). TBP has a core domain consisting of two imperfect repeats which form a saddle-shaped molecule that binds the widened minor groove of an 8-bp TATA element of TATA-containing promoters. Binding of TBP causes unwinding of about a third of a helical turn and bending the DNA about

80 Å towards the major groove (Kim *et al.*, 1993a; Kim *et al.*, 1993b). The role and mode of possible binding in TATA-less promoters is not clear. In addition to TBP, higher eukaryotes have one or two copies of genes encoding TBP-related factors (TRFs) (Hochheimer and Tjian, 2003). Their role is to promote transcription from a subset of protein-coding genes in a cell type-specific fashion. The TBP-DNA complex creates a large asymmetric interface, a platform for binding other components of the transcription machinery.

Nearly all of TAFs have been conserved through evolution (Albright and Tjian, 2000; Green, 2000; Tora, 2002). TAFs function in promoter recognition and in positive and negative regulation of transcription. In higher eukaryotes TFIID have alternative subunits that change the composition of TFIID in a cell type- and development-specific fashion. Some TAFs are also subunits of complexes lacking TBP involved in covalent chromatin modification and transcriptional coactivation (Green, 2000). Current structural information on TFIID includes X-ray structures of several individual subunits which reveal histone fold domains (Xie *et al.*, 1996; Werten *et al.*, 2002, Gangloff *et al.*, 2001). The general architecture of the whole TFIID complex was revealed by electron microscopy and showed three lobes of a horseshoe in both closed and open configurations (Andel *et al.*, 1999; Brand *et al.*, 1999). According to immune localization experiments, TBP is positioned in the center lobe on the inside of the horseshoe. Overall TFIID protects 40-60 bp of DNA from DNase I cleavage (Sanders *et al.*, 2002; Chi *et al.*, 1995). However TFIID is not universally required at all promoters (Freiman *et al.*, 2001; Shen *et al.*, 2003) but rather some TAFs are important for gene regulation.

TFIIA and TFIIB specifically and independently interact with TBP and DNA (Nikolov *et al.*, 1995; Geiger *et al.*, 1996; Tan *et al.*, 1996). TFIIA stabilizes TBP-DNA binding (Weideman *et al.*, 1997) and strongly promotes binding of TFIID to DNA through competition with the TAF1 N-terminal domain that occludes the DNA-binding surface of TBP when TFIID is not bound to DNA (Kokubo *et al.*, 1998; Liu *et al.*, 1998; Sanders *et al.*, 2002). Moreover, it was demonstrated that TFIIA can stimulate basal transcription by interaction with both TFIIF and TFIIE (Langelier *et al.*, 2001).

The general transcription factor TFIIB consists of two domains conserved in Pol III and archaea factors Brf1 and TFB. A flexible linker connects N-terminal Zn-ribbon domain to the C-terminal core domain (TFIIBc) that binds the TBP-DNA subcomplex (Chen and Hahn, 2003; Pardee *et al.*, 1998). The functional surface of the ribbon domain is conserved in TFIIB, Brf1 and TFB and is essential for recruitment of Pol II to the PIC (Hahn and Roberts, 2000). The linker connecting the ribbon and core domains contains a short conserved block of sequence that forms a loop termed the B-finger, which is positioned in the active site of Pol II (Bushnell *et al.*, 2004).

TFIIE acts at a late stage in the PIC formation, interacting with TFIIB, TFIIH, Pol II, as well as with the promoter DNA (Maxon *et al.*, 1994; Orphanides *et al.*, 1996, Roeder, 1996). Together with TFIIH, TFIIE participates in promoter melting and open complex formation. In higher eukaryotes TFIIE consists of two highly charged subunits, referred to as α and β , respectively. The acidic C-terminus of TFIIE α and conserved zinc-finger domain were mapped to interact with TFIIH (Okamoto *et al.*, 1998). The central core region of the β -subunit is essential for basal and activated transcription and binds to double-stranded DNA. The C-terminal half of TFIIE β contains two basic stretches that interact with TFIIB, RAP30 subunit of TFIIIF and single-stranded DNA (Okamoto *et al.*, 1998). NMR studies have shown that the TFIIE β region with homology to TFIIIF forms a winged helix domain (Okuda *et al.*, 2000). Structural information on TFIIE α includes the structure of the archaeal homolog, TFE that adopts a winged helix-turn-helix fold (Meinhart *et al.*, 2003a) and the zinc-finger domain, composed of one α -helix and five β -strands (Okuda *et al.*, 2004). Three-dimensional envelopes of the α/β particles were analyzed by electron microscopy and showed an elongated structure composed of three distinct modules (Jawhari *et al.*, 2006).

TFIIH is a 10-subunit complex, by size and complexity comparable to Pol II itself. It contains helicase activities that unwind DNA and form a transcription bubble (Svejstrup *et al.*, 1996; Coin and Egly, 1998). The overall architecture of human TFIIH reveals a ring-like structure that could easily accommodate double-stranded

DNA (Schultz *et al.*, 2000). TFIIF also comprises a kinase that phosphorylates the Pol II CTD during the transition from initiation to elongation (see chapter 1.7). Both enzymatic activities of TFIIF are stimulated by TFIIE (Ohkuma and Roeder, 1994).

The core promoter is a minimal DNA sequence needed to specify nonregulated or basal transcription and serves as a platform for recruitment of the transcription machinery (Cosma, 2002). Promoters of protein-coding genes contain one or more of the following DNA elements: TATA (TBP-binding site), BRE (TFIIB-recognition element), Inr (Initiator element) and DPE (downstream promoter element) (Smale and Kadonaga, 2003; Figure 3). Their role is to unidirectionally guide the transcription machinery to the promoter. However, only 30% of mRNA genes analyzed in *Drosophila melanogaster* have promoters with a recognizable TATA element (Ohler *et al.*, 2002). In archaea the primary determinant of transcription orientation is the BRE (Bell *et al.*, 1999; Littlefield *et al.*, 1999). Inr and DPE serve as binding sites for certain TAFs, subunits of TFIID (Burke and Kadonaga, 1997).

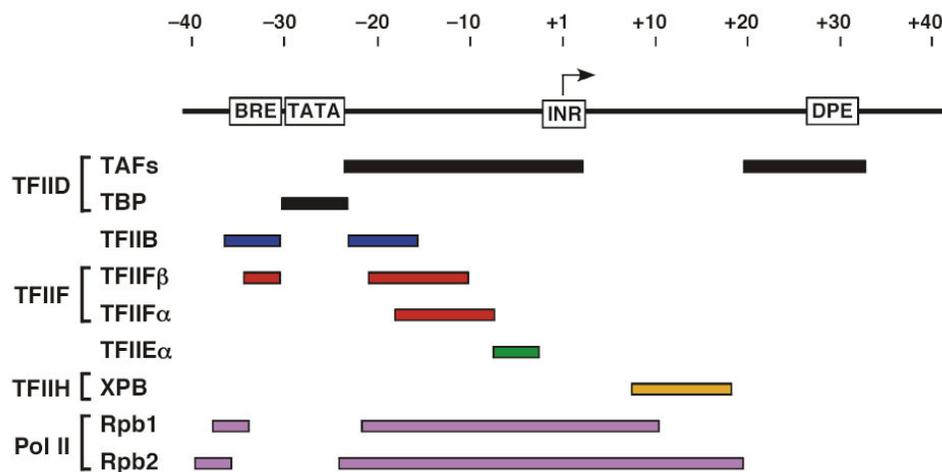


Figure 3: Summary of human general transcription factor protein-DNA crosslinks at a promoter. Top line represents promoter DNA with the position of functional elements indicated. Arrow represents the transcription start site. Adopted from (Hahn, 2004).

During initiation, 11-15 base pairs of DNA surrounding the transcription start site are melted and the template strand of the promoter is positioned within the active site cleft of Pol II to form the open complex (Wang *et al.*, 1992). Upon synthesis of first phosphodiester bond of RNA, many of the 3-10 nucleotides long abortive products are formed before Pol II initiates productive synthesis of full-length RNAs (Luse and Jacob, 1987; Holstege *et al.*, 1997). After the RNA reaches a length of 30 nucleotides, Pol II is thought to release its contacts with the core promoter and enter the stage of transcription elongation. Many of the general transcription factors remain at the promoter after initiation in the form of the so-called scaffold complex (Yudkovsky *et al.*, 2000), which can rapidly recruit the remaining general factors to promote transcription reinitiation.

In the transcription elongation complex (TEC), incoming DNA is unwound before the polymerase active site and is rewound beyond it to form the exiting duplex. In the unwound region, known as the transcription bubble, the DNA template strand forms a hybrid duplex with the nascent pre-mRNA. Pol II maintains the bubble, selects nucleoside triphosphates in a template-directed manner, synthesizes RNA, translocates along the DNA and separates RNA from DNA at the upstream end of the hybrid. The elongating Pol II recruits many factors that promote productive RNA chain elongation (Table 1), by enhancing the elongation rate, or by regulating RNA processing, RNA export and chromatin modification (Bentley, 2002).

Far less is known on the mechanism of transcription termination in eukaryotes. Recently, however, a 5'-3' riboexonuclease (yeast Rat1 and human Xrn2) has been implicated in the process (Kim *et al.*, 2004; Teixeira *et al.*, 2004; West *et al.*, 2004). The exonuclease travels with Pol II and gains access to the nascent RNA after endonucleolytic cleavage site at the poly(A) site or at a second cotranscriptional cleavage site (CoTC). However, this degradation can occur even without eliciting termination, implying that it is not sufficient to cause Pol II release (Luo *et al.*, 2006). Rather, Rat1 is critical for recruitment of 3'-processing factors and for correct 3'-end formation, but is not a dedicated transcription factor (Luo *et al.*, 2006).

Table 1: Human transcription elongation factors that modulate rate of elongation (Sims *et al.*, 2004)

Elongation factor	Function
TFIIF	Alleviates pausing, stimulates the rate of Pol II, modulates TFIIIS
TFIIIS	Stimulates Pol II-mediated cleavage of nascent transcript to alleviate arrest
Elongins	Alleviates pausing, stimulates the rate of Pol II, Ub-related events?
ELL	Alleviates pausing, stimulates the rate of Pol II
DSIF	Stimulates elongation, suppresses early transcript termination, stimulates capping
NELF	Halts Pol II to allow timely 5'-capping, checkpoint control
CSB	Stimulates elongation, modulates TFIIIS, has a role in rescuing Pol II at DNA lesions and transcription-coupled nucleotide excision repair
Fcp1	Stimulates elongation, recycles Pol II, role in capping
Spt6	Stimulates elongation, modulates chromatin structure, histone chaperone activity
HDAg	Stimulates elongation, binds Pol II, displaces NELF, functionally distinct from TFIIF

1.4 TFIIF is a unique eukaryal general transcription factor

TFIIF is the only general transcription factor that has been implicated in the formation of the PIC, open complex formation, initiation as well as elongation. It was initially purified as a factor that directly binds immobilized Pol II (Sopta *et al.*, 1985). In yeast cells about 50% of Pol II is found associated with the general transcription factor TFIIF (Rani *et al.*, 2004).

In humans TFIIF forms a heterodimer of RAP30 and RAP74 subunits, named after their apparent electrophoretic mobility in SDS gels. The human RAP74 polypeptide

(517 amino acids; calculated mass 58 kDa) comprises a globular N-terminal domain that binds RAP30 (Wang and Burton, 1995), a highly charged central linker region, and a globular C-terminal domain involved in the interaction with Pol II (Fang and Burton, 1996), TFIIB (Wang and Burton, 1995) and Fcp1 (Chambers *et al.*, 1995; Kobor *et al.*, 2000). The human RAP30 polypeptide (249 amino acids, 28 kDa) is composed of an N-terminal domain that binds RAP74 and TFIIB (Fang and Burton, 1996), a central domain that interacts with Pol II (McCracken *et al.*, 1991; Sopta *et al.*, 1989), and a cryptic C-terminal DNA-binding domain with similarity to that found in members of the σ^{70} family of bacterial sigma factors (Garret *et al.*, 1992; Tan *et al.*, 1994).

The yeast TFIIF homolog consists of three polypeptides named Tfg1, Tfg2, and Tfg3 (Henry *et al.*, 1994). The Tfg1 polypeptide is homologous to the human RAP74 but significantly larger (735 versus 517 amino acids). Similarly, the Tfg2 polypeptide is homologous to, but significantly larger than, the human RAP30 (400 versus 249 amino acids). Tfg3 has no counterpart in mammalian TFIIF. Additionally known as Anc1, Swp29, TAF30, or TAF14, it belongs to a newly described family of proteins containing a conserved YEATS domain, the function of which is unknown (Cairns *et al.*, 1996; Henry *et al.*, 1994; John *et al.*, 2000). In yeast, members of this family include components of complexes involved in chromatin silencing, histone acetyltransferase complex and chromatin remodeling complex (Masson *et al.*, 2003; Poon *et al.*, 1995; Zhang, 2004). Accordingly, Tfg3 is part of chromatin remodeling complexes RSC and INO80, and of the histone H3-acetyltransferase complex NuA3. Interactions with the Tfg1 subunit of TFIIF, and the TAF2 subunit of the general transcription factor TFIID were reported, suggesting a common regulatory function (Kabani *et al.*, 2005). Strains devoid of Tfg3 are viable but thermosensitive, suggesting that Tfg3 plays important yet dispensable roles in the cell.

For RAP74, structural information is available for the human N-terminal interaction domain, as well as a short C-terminal RAP74 domain in free and Fcp1 interacting helix bound state (Table 2, Figure 4). Additionally, the structure of the C-terminal DNA-binding domain of RAP30 was solved by NMR (Figure 4B). The X-ray structure

of the RAP30/RAP74 interaction domain heterodimer at 1.7 Å resolution reveals a single core structure consisting of three interwoven β-barrels, a novel “triple barrel” dimerization fold (Gaiser *et al.*, 2000) (Figure 4A). Together with mutational data, the structure suggests that interactions with the transcription apparatus are mediated not only by this triple β-barrel, but also via flexible loops and α- and β-structures extending from it (Gaiser *et al.*, 2000). The RAP30 C-terminal domain that contains a cryptic DNA-binding motif, was demonstrated by NMR to be similar to the “winged” helix-turn-helix DNA-binding domains of linker histone H5 and hepatocyte nuclear transcription factor HNF3/forkhead (Groft *et al.*, 1998) (Figure 4B).

Table 2: Available structural information on TFIIF

Structure	Source	Experimental technique (PDB code)	Reference
RAP30 DNA-binding domain	<i>H. sapiens</i>	NMR (1bby, 2bby)	Groft <i>et al.</i> , 1998
RAP30/RAP74 interaction domain	<i>H. sapiens</i>	X-ray (1f3u)	Gaiser <i>et al.</i> , 2000
RAP74 subunit C-terminal domain	<i>H. sapiens</i>	X-ray (1i27); NMR (1nha)	Kamada <i>et al.</i> , 2001; Nguyen <i>et al.</i> , 2003a
RAP74 C-terminal domain complexed with Fcp1 C-terminal peptide	<i>H. sapiens</i>	X-ray (1j2x); NMR (1onv)	Kamada <i>et al.</i> , 2003; Nguyen <i>et al.</i> , 2003b
Pol II-TFIIF architecture	<i>S. cerevisiae</i>	Cryo-EM	Chung <i>et al.</i> , 2003

Like the C-terminal domain of RAP30, which may be responsible for nonspecific DNA binding, the C-terminal domain of RAP74 possesses a canonical winged-helix fold common to HNF-3γ, the linker histones H1 and H5, and a large family of so-called forkhead transcription factors (Figure 4C; Kamada *et al.*, 2001; Nguyen *et al.*, 2003a). However, the surface electrostatic properties of this compact domain differ significantly from other winged-helix DNA-binding domains. RAP74 has been shown to interact with the TFIIF-associated C-terminal domain phosphatase, Fcp1, and a putative phosphatase binding site has been identified within the RAP74 winged-helix domain. A cocrystal structure of the winged-helix domain of human RAP74 bound to

the alpha-helical C-terminus of human Fcp1 (residues 944-961) was reported (Figure 4D; Kamada *et al.*, 2001; Nguyen *et al.*, 2003b). Similarly to the free RAP74 C-terminal domain, when in the complex, it also forms a winged-helix domain consisting of three consecutive α -helices followed by an antiparallel β -sheet. Interestingly, the free C-terminal Fcp1 domain is devoid of any stable structural element but adopts a 17-residue α -helix upon interaction with RAP74 (Figure 4D).

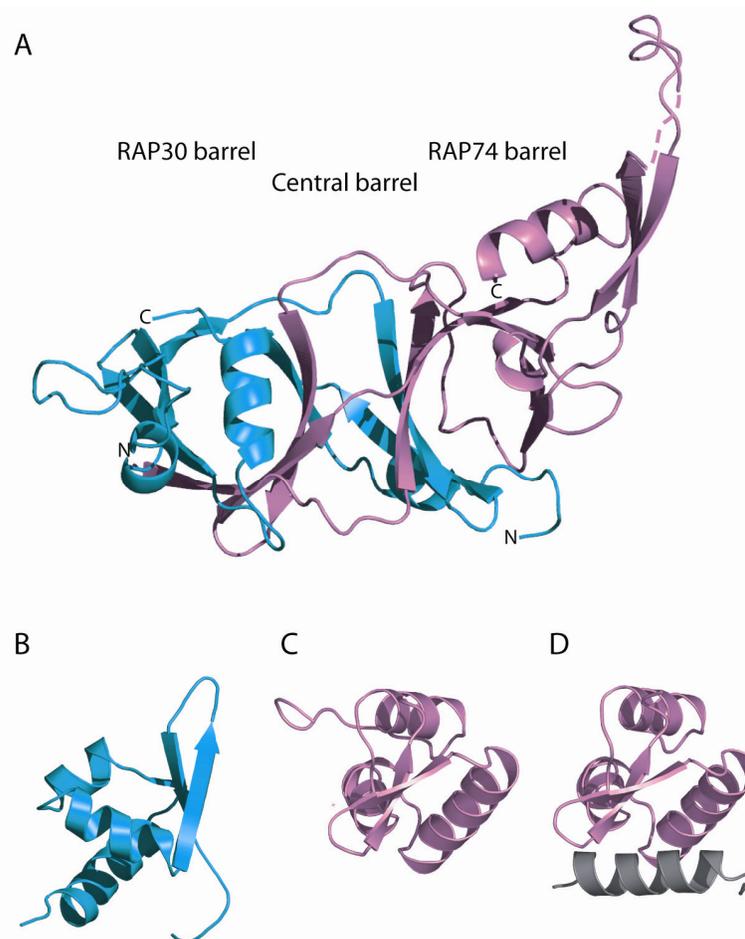


Figure 4: Available structures of human TFIIF. (A) X-ray structure of the RAP30/RAP74 interaction domain at 1.7 Å (PDB code 1i27, Gaiser *et al.*, 2000); (B) NMR structure of the RAP30 DNA binding domain (PDB code 1bby, Groft *et al.*, 1998). (C) X-ray structure of the RAP74 subunit C-terminal domain (PDB code 1i27, Kamada *et al.*, 2001); (D) X-ray structure of the RAP74 subunit C-terminal domain complexed with Fcp1 C-terminal peptide (PDB code 1j2x, Kamada *et al.*, 2003). Figures were prepared with PyMol (www.pymol.org).

1.4.1 Roles of TFIIF during PIC formation and promoter escape

TFIIF recruits Pol II to promoter DNA during PIC formation (Conaway *et al.*; 1991; Flores *et al.*, 1991; Orphanides *et al.*, 1996). Chromatin immunoprecipitation (ChIP) experiments localized TFIIF predominantly at the promoter (Krogan *et al.*, 2002; Pokholok *et al.*, 2002) region. Photo-cross-linking experiments showed that TFIIF binds both upstream of the TATA element and downstream of the transcription start site (Forget *et al.*, 2004) suggesting that promoter DNA wraps around the mobile clamp of Pol II. It seems that the location of RAP74 is centered downstream of the first nucleotide to be transcribed whereas that of RAP30 is centered on the TATA box and immediately downstream of it (see chapter 1.3, Figure 3).

A role for TFIIF in start site selection was uncovered in a genetic screen for suppressors of the cold-sensitive growth defect associated with the *sua7-1*-encoded TFIIB (Sun and Hampsey, 1995). Analyses of double mutant strains demonstrated functional interactions between the Tfg1 mutations and mutations in Tfg2, TFIIB, and Pol II that also confer alterations in start site utilizations (Ghazy *et al.*, 2004). Thus TFIIB and TFIIF are critical determinants of start site selection in *S. cerevisiae*. However, the mechanisms by which altered forms of these factors affect start site selection remain unknown.

TFIIF is further required for the entry of TFIIE and TFIIH into the PIC (Conaway *et al.*, 1991; Flores *et al.*, 1992; Conaway *et al.*, 1990a). Together with TFIIE, TFIIF participates in DNA strand separation by inducing the wrapping of DNA around Pol II (Coulombe and Burton, 1999). It displays dual roles in promoter escape, by cooperating with TFIIH to modulate premature arrest of early elongation intermediates, and, in a reaction dependent on TFIIF elongation activity, by increasing the processivity of very early elongation intermediates (Yan *et al.*, 1999a).

1.4.2 Roles of TFIIIF in elongation

Following its release from the transcription initiation complex, TFIIIF reassociates with the TEC in particular when Pol II complex has stalled (Zawel *et al.*, 1995). ChIP experiments localized TFIIIF also in the coding and in 3'-untranslated region (Krogan *et al.*, 2002; Pokholok *et al.*, 2002). TFIIIF appears to represent an active elongation factor, but does not remain associated with the actively moving Pol II. TFIIIF associates with multiple elongation factors including Spt5 of DSIF (Lindstrom *et al.*, 2003), and components of the PAF complex (Shi *et al.*, 1997). Moreover TFIIIF appears to influence the TFIIIS cleavage factor (Elmendorf *et al.*, 2001; Zhang *et al.*, 2003). In the absence of TFIIIF, TFIIIS supports entry of Pol II into the backtracking, RNA cleavage and restart pathway (Zhang *et al.*, 2003). TFIIIF diminishes the time Pol II is paused and stimulates the rate of Pol II transcription elongation (Flores *et al.*, 1989; Price *et al.*, 1989; Bengal *et al.*, 1991; Izban and Luse, 1992; Tan *et al.*, 1994). Consequently, TFIIIF is an established elongation factor which has important role on the regulation of Pol II TEC.

The initiation and elongation activities of TFIIIF are regulated mostly by the phosphorylation of RAP74 (Kitajima *et al.*, 1994) largely by the TAF250 subunit of TFIID, and TFIIH (Ohkuma and Roeder, 1994; Dikstein *et al.*, 1996; Yankulov and Bentley, 1997; Yonaha *et al.*, 1997). In addition an autophosphorylation activity was reported (Rossignol *et al.*, 1999). Mutational analysis strongly suggested that autophosphorylation regulates the transcription elongation process. In accordance with its roles in initiation and elongation, TFIIIF has the ability to associate with both hypo- and hyperphosphorylated form of Pol II (Sims *et al.*, 2004).

1.4.3 TFIIIF interactions with Pol II

The general architecture of the Pol II-TFIIIF complex at low resolution (18 Å) determined by cryo-electron microscopy and single particle analysis showed density

due to TFIIF not concentrated in one area, but rather widely distributed across the surface of polymerase (Chung *et al.*, 2003; Figure 5).

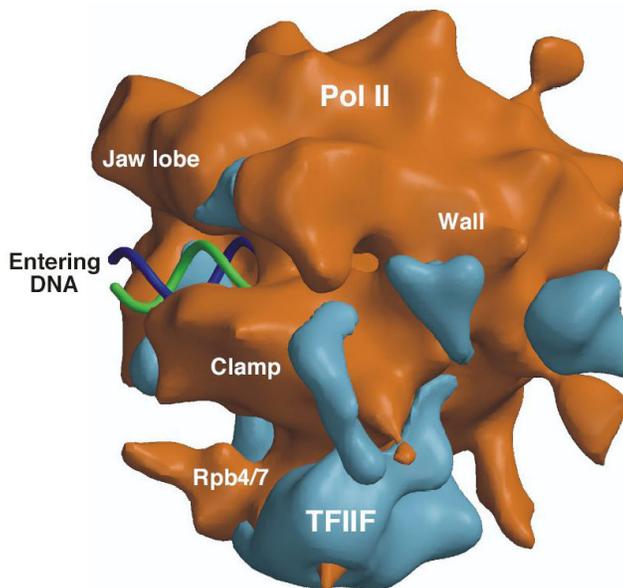


Figure 5: EM architecture of the Pol II-TFIIF complex (Chung *et al.*, 2003). Pol II is an orange surface with TFIIF density in blue. The complex is shown in "top" orientation (as that in Figure 1, *right*). Also shown is a model for downstream double-stranded DNA entering the active site cleft (blue and green helix). Figure was adopted from (Hahn, 2004).

The distribution of Tfg2 was very similar to that reported for the σ subunit in the bacterial RNA polymerase holoenzyme, consisting of a series of discrete globular domains extending along the polymerase active site cleft connected by extended linkers (Murakami *et al.*, 2002; Vassylyev *et al.*, 2002). The authors (Chung *et al.*, 2003) postulate Tfg2 as a true structural homolog of the bacterial σ factor. The bacterial σ factor serves 2 distinct roles: recognizing the -10 and -35 regions of the promoter and facilitating the formation of an unwound region of DNA surrounding the transcription start site (Gross *et al.*, 1998). The role in promoter binding recognition entails not only promoter binding but also the capacity of σ to disrupt a nonspecific RNA polymerase – DNA complex. In eukaryotes this role is carried out by TFIIF

(Conaway and Conaway, 1990; Killeen and Greenblatt, 1992; Tan *et al.*, 1995). Additional TFIIIF density was located along the length of the active site cleft of polymerase, starting behind the wall, near subunit Rpb12, next to the protrusion, along the top of the clamp, past the downstream end of the active site cleft near Rpb5, and into the Rpb1/Rpb8 feet (Figure 5). This is consistent with the previously reported interaction of the middle region of RAP30 and Rpb5 by GST pull-down assay (Wei *et al.*, 2001). A significant amount of TFIIIF density probably not related to Tfg2 is closely associated with the Rpb4/7 subcomplex of Pol II, which has a well-established tendency to dissociate from the rest of the enzyme.

1.5 Carboxy-terminal domain of Pol II (CTD)

The CTD is a unique tail-like feature of the largest Pol II subunit that consists of 26 (yeast) and 52 (human) heptapeptide repeats of the consensus sequence Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7. However yeast requires at least eight repeats for viability (Nonet *et al.*, 1987; West and Corden, 1995). The CTD integrates a whole range of nuclear events by binding proteins involved in mRNA biogenesis. Through its interaction with the Mediator complex, it plays a critical role in transcription activation (Thompson *et al.*, 1993; Kim *et al.*, 1994; Gerber *et al.*, 1995). Furthermore, the CTD is required for efficient capping, splicing, cleavage and polyadenylation of mRNAs *in vivo*, and binds to processing factors *in vitro* (McCracken *et al.*, 1997a; Hirose and Manley, 2000; Proudfoot *et al.*, 2002). CTD-binding proteins recognize a specific CTD phosphorylation pattern, which changes during the transcription cycle due to the action of CTD-modifying enzymes.

The CTD protrudes from an 80-residue linker bridging it to the core polymerase. The crystal structures of yeast Pol II do not reveal the CTD due to the mobility (Cramer *et al.*, 2001; Armache *et al.*, 2005). The free CTD is currently considered to be a flexible polypeptide chain with some residual structure and propensity to form β -turns. This view is mostly based on available NMR studies of a single consensus repeat

(Harding, 1992). Further studies by NMR and circular dichroism on the CTD polypeptide comprising of eight consensus repeats confirmed small population of β -turn structures in water, increasing to 75% in trifluoroethanol (Cagas and Corden, 1995; Bienkiewicz *et al.*, 2000). Recent studies of a two-repeat CTD peptide with a central phosphorylated S2 residue, however, revealed a dynamic disordered ensemble (Noble *et al.*, 2005). Disordered nature of the free CTD would not be surprising considering a great variety of interacting proteins. Assuming an extended β -strand conformation, the yeast CTD and linker would be ~ 650 Å and ~ 250 Å long, respectively, potentially reaching anywhere on the surface of Pol II (~ 150 Å in diameter; Meinhart *et al.*, 2005). However, electron micrographs revealed some weak density attributed to the CTD which measured only 100 Å pointing out that CTD is mostly compact, at least in its unphosphorylated form (Meredith *et al.*, 1996). Phosphorylation of the CTD results in a far more extended and more protease-sensitive structure (Laybourn and Dahmus, 1989; Zhang and Corden, 1991). Over the last two decades, several different models for CTD structure have been proposed, ranging from a compact random coil model (Cramer *et al.*, 2001) to variety of β -spiral models (Matsushima *et al.*, 1990; Suzuki, 1990; Cagas and Corden, 1995; Meinhart and Cramer, 2004). Current evidence suggest that the CTD polypeptide adopts secondary structure elements in a portion-wise fashion leading to an overall compaction, strongly dependent on the level of phosphorylation.

1.6 CTD phosphorylation and transcriptional regulation

In the course of the transcription cycle, the CTD undergoes dynamic phosphorylation and dephosphorylation (Orphanides and Reinberg, 2002). Transcription initiation requires an unphosphorylated CTD, whereas transcription elongation is carried out by Pol II with a hyperphosphorylated CTD that binds mRNA processing factors for transcription-coupled mRNA maturation. For recycling of Pol II and reinitiation of transcription, the CTD must be dephosphorylated.

There are five potential phosphorylation sites in a CTD consensus repeat (Y1, S2, T4, S5, and S7), out of which only S2 and S5 are being commonly phosphorylated (Corden *et al.*, 1985; Zhang and Corden, 1991). These two serine positions are not equivalent in function (West and Corden, 1995; Zhang and Corden, 1991) and the phosphorylation pattern changes during the transcription cycle due to the coordinated action of CTD-kinases and -phosphatases. S5 phosphorylation occurs in promoter-proximal regions, and leads to recruitment of the capping enzyme (Cho *et al.*, 1997; McCracken *et al.*, 1997b; Ho *et al.*, 1998; Komarnitsky *et al.*, 2000). S2 phosphorylation predominates in regions that are more distal from the promoter, and triggers binding of the 3'-RNA processing machinery (Komarnitsky *et al.*, 2000; Cho *et al.*, 2001). In 2003 Buratowski proposed that a "CTD code" specifies the position of Pol II within the transcription cycle. The CTD code would take into account possible combinations of different phosphorylation states (S2 and S5) and proline (P3 and P6) conformations (cis/trans) within a single consensus peptide and would give rise to 16 different states of a CTD repeat (Buratowski, 2003).

1.7 CTD kinases

In humans three cyclin-dependent kinases, CDK7, CDK8, and CDK9, associated with cyclins H, C and T, respectively, phosphorylate the CTD (Dynlacht, 1997; Bregman *et al.*, 2000; Murray, 2004).

The CDK7/cyclin H is a part of subcomplex of the ten-subunit general transcription factor TFIIF, which phosphorylates the CTD at S5 during transcription initiation (Coin and Egly, 1998). In yeast this role is conferred on Kin28, which is essential for viability and required for normal transcript levels *in vivo* (Valay *et al.*, 1995; Holstege *et al.*, 1998; Komarnitsky *et al.*, 2000; Schroeder *et al.*, 2000; Liu *et al.*, 2004).

The CDK8/cyclin C pair (Srb10/Srb11 in yeast) together with MED12 and MED13 forms a module of the Mediator complex that phosphorylates the CTD at S5 residues and is therefore associated with transcription initiation complexes. This module is

conserved among eukaryotes, and is a target of signal transduction pathways (Liu *et al.*, 2001; Borggrefe *et al.*, 2002; Boubé *et al.*, 2002; Samuelson *et al.*, 2003). The CDK8/cyclin C pair is thought to be mainly implicated in transcriptional repression (Hengartner *et al.*, 1998). The mode of this repression remains, however, unknown.

The CDK9/cyclin T pair forms the core of the positive transcription elongation factor P-TEFb (Price, 2000) and phosphorylates S2 of the CTD. P-TEFb was isolated by its ability to overcome arrest of Pol II complexes during early elongation, a function that requires the CTD (Marshall and Price, 1995; Marshall *et al.*, 1996). Ctk1 and Bur1 are two putative homologs of CDK9 found in yeast (Prelich and Winston, 1993; Murray *et al.*, 2001; Prelich, 2002; Guo and Stiller, 2004) thought to phosphorylate both CTD and elongation factor Spt5 (Keogh *et al.*, 2003).

Presently crystal structures are available of free CDK7 (Lolli *et al.*, 2004), cyclin H (Andersen *et al.*, 1997) and cyclin C (Hoepfner *et al.*, 2005). Still however the basis for specificity of a kinase for the CTD and for recognition of a particular CTD residue remains unknown and awaits the structure solution of CDK/CTD complexes.

1.8 CTD phosphatase Ssu72

The yeast Ssu72 is a highly conserved and essential protein, playing roles in all three phases of the transcription cycle. It was initially identified based on genetic and physical interactions with the general transcription factor TFIIB (Sun and Hampsey, 1996; Wu *et al.*, 1999; Pappas and Hampsey, 2000; Dichtl *et al.*, 2002). In addition, it is a component of the CPF 3'-end processing machinery and is required for pre-mRNA and snoRNA 3'-end formation (Gavin *et al.*, 2002; Dichtl *et al.*, 2002; He *et al.*, 2003; Steinmetz and Brow, 2003; Nedeá *et al.*, 2003).

Primary sequence of Ssu72 contains the CX₅RS signature motif indicative of protein tyrosine phosphatases (PTPases) (Figure 6A; Meinhart *et al.*, 2003b; Ganem *et al.*, 2003). Phosphatase activity was confirmed by cleavage of the synthetic substrate p-

nitrophenyl phosphate (Meinhart *et al.*, 2003; Ganem *et al.*, 2003). In PTPases, the conserved cysteine and arginine residues of the signature motif form part of the active site (Figure 6A; Ramponi and Stefani, 1997). The cysteine attacks the substrate phosphorus atom, leading to formation of a phosphocysteinyl intermediate (Figure 6B). The arginine stabilizes the transition state (Burke and Zhang, 1998).

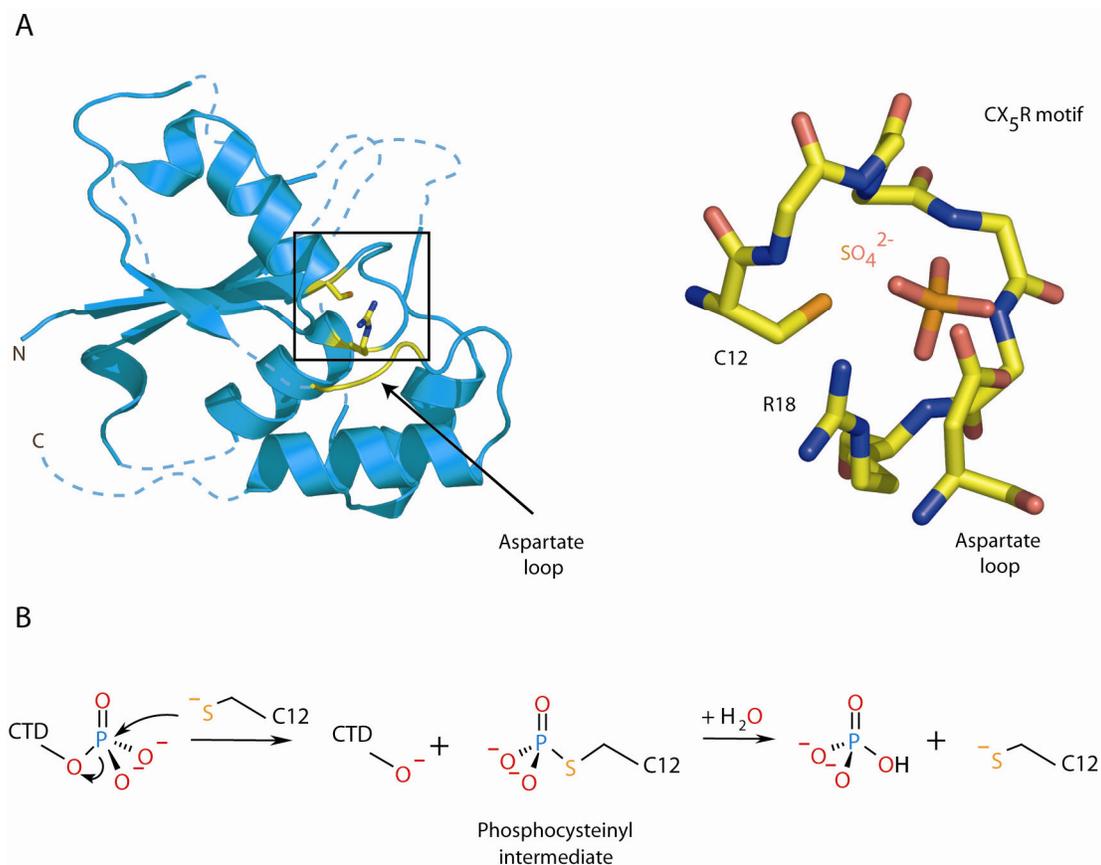


Figure 6: Ssu72 model (A) and proposed mechanism of catalysis (B). (A) Ssu72 has been modeled by homology based on the structure of the bovine low-molecular-weight phosphotyrosine phosphatase (PDB code 1pfr) (Su *et al.*, 1994). Loop regions that are uncertain are shown as dashed lines. The detailed view on the right depicts the active-site region (marked by a rectangle in the structure on the left). A sulfate ion trapped in the structure is depicted in orange/red. The phosphorylated side chain of cysteine residue C12 is shown. Figure was prepared with PyMol (www.pymol.org). (B) The reaction mechanism involves the formation of a phosphocysteinyl intermediate (Ssu72), followed by recycling of the catalytic side chain with the use of a nucleophilic water molecule. Mechanism was drawn with MDL ISIS/Draw.

Although there is no apparent sequence homology between Ssu72 and PTPases outside the signature motif, secondary structure prediction suggested that Ssu72 adopts the fold of the low-molecular-weight family of PTPases (Figure 6A; Meinhart *et al.*, 2003b; Meinhart *et al.*, 2005). In PTPases, an aspartate in a distant loop serves as a general acid/base (Barford *et al.*, 1995; M. Zhang *et al.*, 1998). Ssu72 may also contain a somewhat modified aspartate loop that includes two catalytically important aspartates instead of one (Meinhart *et al.*, 2003b). This and other distinguishing features suggest that Ssu72 is the founding member of a new phosphatase subfamily that is related to low-molecular weight PTPases.

Recently it was demonstrated that Ssu72 is a CTD phosphatase with specificity towards S5 phosphorylation (Krishnamurthy *et al.*, 2004). Depletion of Ssu72 *in vivo* results in an increase of S5-phosphorylated Pol II, and Ssu72 dephosphorylates recombinant CTD that had been phosphorylated at S5 by TFIIF *in vitro*. The essential role of Ssu72 in 3'-processing does not, however, depend on its phosphatase activity.

1.9 CTD phosphatases of the Fcp1 family

Dephosphorylation of CTD serine residues is catalyzed by the CTD phosphatase Fcp1 (Archambault *et al.*, 1997; Archambault *et al.*, 1998; Chambers and Dahmus, 1994; Chambers and Kane, 1996; Cho *et al.*, 1999; Kobor *et al.*, 1999; Lin *et al.*, 2002a). Fcp1 is the founding member of a new phosphatase family with no sequence similarity to other phosphatases, except for the signature motif DXDX(T/V) that is shared within a superfamily of phosphotransferases and -hydrolases (Collet *et al.*, 1998). The Fcp1 sequence comprises two conserved regions. The N-terminal Fcp1 homology (FCPH) region includes the DXDX(T/V) signature motif with residues important for catalysis. The C-terminal BRCT (breast cancer protein related carboxy-terminal) domain binds to the phosphorylated CTD (Yu *et al.*, 2003).

Fcp1 is conserved throughout eukaryotes and is essential for cell viability. Genome-wide expression analysis of a temperature-sensitive Fcp1 mutant shows a nearly global defect in Pol II transcription (Kobor *et al.*, 1999). In humans, partial deficiency of Fcp1 is associated with a strong neurodegenerative disorder known as congenital cataracts facial dysmorphism neuropathy syndrome (Varon *et al.*, 2003). Dephosphorylation of the Pol II CTD by Fcp1 can facilitate recycling of the hyperphosphorylated form of the polymerase for a new round of transcription (Cho *et al.*, 1999). Fcp1 dephosphorylates the CTD in solution (Chambers *et al.*, 1995; Cho *et al.*, 1999) and when associated with a transcription elongation complex (Cho *et al.*, 1999; Lehman and Dahmus, 2000). The phosphatase activity of Fcp1 is stimulated by the general transcription factor TFIIIF, and the general factor TFIIIB inhibits this stimulation (Chambers *et al.*, 1995). Taken together, Fcp1 is a central regulator of the mRNA transcription cycle, but its structure and mechanism remained unclear.

In higher eukaryotes a second CTD phosphatase was described, the small CTD phosphatase Scp1 (Yeo *et al.*, 2003). Scp1 shows homology to the catalytic FCPH domain of Fcp1, but lacks the BRCT domain. Scp1 is a transcriptional regulator that silences neuronal genes in nonneuronal tissue (Yeo *et al.*, 2005). Plants contain CTD phosphatase-like proteins (CPLs), which also comprise a catalytic domain with a DXDXT motif, and also lack a BRCT domain (Koiwa *et al.*, 2004). Microsporidian parasite *Encephalitozoon cuniculi*, however, encodes a minimized 411-amino acid Fcp1-like protein (EcFcp1), which consists of a DXDXT phosphatase domain and a BRCT domain, but lacks the large N- and C-terminal domains found in fungal and metazoan Fcp1 enzymes (Hausmann *et al.*, 2004).

Fcp1 and Scp1 were reported to dephosphorylate S5 and S2 (Table 3; Hausmann and Shuman, 2002; Lin *et al.*, 2002a; Yeo *et al.*, 2003). Highly purified Fcp1 was recently shown to dephosphorylate S5, but not S2 (Kong *et al.*, 2005). Plant CPLs were also shown to specifically dephosphorylate S5 (Koiwa *et al.*, 2004). EcFcp1 dephosphorylates CTD positions S2 and S5 with similar efficacy *in vitro* (Hausmann *et al.*, 2004). An overview of the CTD phosphatases and their preferential targets is given in the Table 3.

Table 3: Specificity of CTD phosphatases

CTD phosphatases	S2 dephosphorylation	S5 dephosphorylation	References
Fcp1 (<i>S. cerevisiae</i>)	+		Cho <i>et al.</i> , 2001a
Fcp1 (<i>S. cerevisiae</i>)		+	Kong <i>et al.</i> , 2005
Fcp1 (<i>S. pombe</i>)	+		Hausmann and Shuman, 2002
Fcp1 (<i>H. sapiens</i>)	+	+	Lin <i>et al.</i> , 2002a
EcFcp1	+	+	Hausmann <i>et al.</i> , 2004
Scp1	(+)	+	Yeo <i>et al.</i> , 2003
Ssu72		+	Krishnamurthy <i>et al.</i> , 2004
CPLs		+	Koiwa <i>et al.</i> , 2004

1.10 Aims of this work

Several years ago, the first structures of CTD interacting domains (CIDs) were brought to scientific attention. These included an N-terminal WW domain of Pin1/Ess1 prolyl isomerase (Verdecia *et al.*, 2000), the FF domain in FBP11 (Allen *et al.*, 2002) and a domain of the capping enzyme subunit Cgt1 (Fabrega *et al.*, 2003). A striking observation was a variety of the emerging CTD-binding folds even though they all have the “same” binding partner. However, no structural information was available on any of the CTD-modifying enzymes at the time. It was only logical to pursue the structure of the Fcp1-family CTD phosphatases, with Fcp1 by then being the only reported CTD phosphatase that stands at the center of Pol II recycling. Therefore, one aim of this work was to elucidate the mechanism of Pol II dephosphorylation by structure-function analysis of Pol II CTD phosphatases.

Contrary to Fcp1, substantial structural information on portions of human TFIIF was already available (Groft *et al.*, 1998; Gaiser *et al.*, 2000; Kamada *et al.*, 2001; Kamada *et al.*, 2003; Nguyen *et al.*, 2003a; Nguyen *et al.*, 2003b). However, the mode of interaction with Pol II, aside of biochemical data, was not known. The main obstacle towards that goal was of technical nature. Yeast TFIIF, unlike other smaller general transcription factors, could not be overexpressed in bacteria, and could only be purified from yeast in insufficient amounts for crystallographic study. In the course of this PhD project, however, the general topology of Pol II-TFIIF complex was elucidated by electron microscopy (Chung *et al.*, 2003), a technique that requires far less protein material compared to X-ray crystallography. The limitations of this technique still do not allow higher resolution studies, and the details of Pol II-TFIIF interaction remained unknown. Therefore, the second aim of this work was to prepare Pol II-TFIIF complex as a major step towards the structure of the initially transcribing complex.

2 Materials and Methods

2.1 Bacterial and yeast strains

Table 4: *E. coli* strains

Strain	Description	Source or reference
XL-1 blue strain	recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac[F ⁺ proAB lacI ^q ΔM15Tn10(Tet ^r)]	Stratagene
BL21-CodonPlus (DE3)-RIL	B F ⁻ ompT hsdS(r _B ⁻ m _B ⁻) dcm ⁺ Tet ^r gal λ(DE3) endA Hte [argU ileY leuW Cam ^r]	Stratagene
B834	<i>E. coli</i> (DE3) (hsd metB)	(Budisa <i>et al.</i> , 1995)

Table 5: *S. cerevisiae* strains

Strain	Description	Source or reference
CB010	<i>MATa pep4::HIS3/prb1::LEU2, prc1::HISG, can1, ade2, trp1, ura3, his3, leu2-3</i>	Edwards <i>et al.</i> , 1990
CB010ΔRpb4	identical to CB010 except <i>RPB4</i> deletion	Edwards <i>et al.</i> , 1990; Fu <i>et al.</i> , 1999
CB010Tfg1TAP	identical to CB010 except <i>TFG1-TAP::TRP1</i>	K. Sträßer
CB010Tfg2TAP	identical to CB010 except <i>TFG2-TAP::TRP1</i>	K. Sträßer
DSY5	<i>MATa leu2 trp1 ura3-52 his3 pep4 prb1</i>	Dualsystems Biotech
DSY5-Int1	<i>MATa leu2 ura3-52 his3 pep4 prb1; TRP1::P_{ADH1}-TFG2-TAP-t_{ADH1}::TRP1</i>	this study
DSY5-Int2	<i>MATa leu2 his3 pep4 prb1; TRP1::P_{ADH1}-TFG2-TAP-t_{ADH1}::TRP1; URA3-52::P_{ADH1}-TFG3-t_{ADH1}::URA3-52</i>	this study
DSY5-Int3	<i>MATa his3 pep4 prb1; TRP1::P_{ADH1}-TFG2-TAP-t_{ADH1}::TRP1; URA3-52::P_{ADH1}-TFG3-t_{ADH1}::URA3-52; LEU2::P_{ADH1}-TFG1-t_{ADH1}::LEU2</i>	this study

2.2 Plasmids

The pET-21a-d(+) vectors carry an N-terminal T7-Tag sequence plus an optional C-terminal His-Tag sequence. These vectors differ from pET-24a-d(+) only by their selectable marker (ampicillin vs. kanamycin resistance). Target genes are cloned under control of strong bacteriophage T7 transcription and translation signals, and expression is induced by providing source of T7 RNA polymerase in the host cell. Additionally, these vectors possess origin of replication (*ori*) from pBR322 and f1 origin.

YI-type (integrating) *E. coli*-yeast shuttle vectors permit visual detection of recombinants by beta-galactosidase. Multiple cloning site (MCS) stems from pUC19 with all its 10 unique cloning sites. YIplac vectors contain no replication of origin and are used to integrate DNA fragments into the yeast chromosomes by homologous recombination. These vectors possess ampicillin bacterial selectable marker, but differ in a yeast selectable marker: *LEU2* (YIplac128), *URA3* (YIplac211) or *TRP1* (YIplac204). The vectors were kindly provided by Marian Kalocay (Jentsch Laboratory).

2.3 Media and supplements

Table 6: Supplements and antibiotics

Supplement or antibiotic	Stock solution	Applied
Ampicillin	100 g/L in H ₂ O	100 µg/mL for <i>E. coli</i> ; 50 µg/mL for yeast
Kanamycin	30 g/L in H ₂ O	30 µg/mL
Tetracycline	12.5 g/L	125 µg/mL
Chloramphenicol	50 g/L in ethanol	50 µg/mL
IPTG	1 M in H ₂ O	0.5 mM

Table 7: Growth media

Media	Description	Source or reference
LB	1% (w/v) tryptone; 0.5% (w/v) yeast extract; 0.5% (w/v) NaCl (+ 1.5 (w/v) agar for selective media plates)	Sambrook and Russel, 2001
SOC	2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl ₂ , 20 mM MgSO ₄ , 20 mM glucose	Sambrook and Russel, 2001
Minimal media	7.5 mM (NH ₄) ₂ SO ₄ , 8.5 mM NaCl, 55 mM KH ₂ PO ₄ , 100 mM K ₂ HPO ₄ , 1 mM MgSO ₄ , 20 mM glucose, 1 µg/L trace elements (Cu ²⁺ , Mn ²⁺ , Zn ²⁺ , Mo ₄ ²⁻), 10 mg/L thiamine, 10 mg/L biotine, 1 mg/L Ca ²⁺ , 1 mg/L Fe ²⁺ , 100 mg/L amino acids (A, C, D, E, F, G, H, I, K, L, N, P, Q, R, S, T, V, W, Y), 100 mg/L selenomethionine.	Budisa <i>et al.</i> , 1995; Meinhart <i>et al.</i> , 2003a
YPD	2% (w/v) peptone, 2% (w/v) glucose, 1.48% (w/v) yeast extract (+ 1.8% (w/v) agar for selective media plates)	

2.4 Buffers and solutions

The following table contains the buffers and solutions used in this work. Buffers for specific protein purifications are given separately along with purification protocols.

Table 8. Buffers, dyes and solutions

SDS-PAGE		
4x stacking gel buffer		0.5 M Tris; 0.4% (w/v) SDS; pH 6.8 (25 °C)
4x separation gel buffer		3 M Tris; 0.4% (w/v) SDS; pH 8.9 (25 °C)
electrophoresis buffer		250 mM glycine; 25 mM Tris; 0.1% (w/v) SDS
5x sample suffer		10% (v/v) glycerol, 50 mM Tris-HCl pH 7.0 (25 °C), 0.1% (w/v) bromophenol blue, 0.1% (w/v) lauryl sulfate, 1% (w/v) β-mercaptoethanol, 14% (w/v) 1,4-dithiothreitol

Table 8. (continuation)

Coomassie staining	
Staining solution	50% (v/v) ethanol, 7% (v/v) acetic acid, 0.125% (w/v) Coomassie Brilliant Blue R-250
Distaining solution	5% (v/v) ethanol, 7.5% (v/v) acetic acid
Silver staining (Bloom <i>et al.</i> , 1987)	
Fixing solution	50% (v/v) ethanol, 12.5% (v/v) acetic acid, 0.01% (w/v) formaldehyde
Rinse solution I	50% (v/v) ethanol
Rinse solution II	30% (v/v) ethanol
Sensitizer	0.2 g/L sodium thiosulfate
Staining solution	0.1% (w/v) silver nitrate, 0.015% (w/v) formaldehyde
Developer	6% (w/v) sodium carbonate, 0.01% (w/v) formaldehyde, 0.4 ng/ μ L sodium thiosulfate
Stop solution	5% (v/v) acetic acid
Dot blots	
Washing buffer, 1xPBS	137 mM NaCl, 2.7 mM KCl, 10 mM Na ₂ HPO ₄ , 2 mM KH ₂ PO ₄
Blocking buffer	5% (w/v) skimmed milk powder in 1xPBS
Gel electrophoresis of nucleic acids	
Electrophoresis buffer, 1xTBE	8.9 mM Tris, 8.9 mM boric acid, 2 mM EDTA (pH 8.0, 25 °C)
6x DNA loading dye	1.5 g/L bromophenol blue, 1.5 g/L xylene cyanol, 50% (v/v) glycerol
2x RNA loading dye (Fermentas)	95% formamide, 0.025% (w/v) SDS, 0.025% (w/v) bromophenol blue, 0.025% (w/v) xylene cynol, 0.5 mM EDTA
Sample preparation for Edman sequencing	
Blotting buffer	10% (v/v) methanol in ddH ₂ O
Rehydration (swelling) buffer	200 mM Tris-HCl pH 8.5 (25 °C), 2% (w/v) SDS
Buffers for the preparation of chemically competent cells	
TFB-1	30 mM KOAc, 50 mM MnCl ₂ , 100 mM RbCl, 10 mM CaCl ₂ , 15% (v/v) glycerol, pH 5.8 (25 °C)
TFB-2	10 mM MOPS (pH 7.0, 25 °C), 10 mM RbCl, 75 mM CaCl ₂ , 15% (v/v) glycerol

2.5 Bioinformatic tools

DNA and protein sequences of yeast proteins were found using Saccharomyces Genome Database (<http://www.yeastgenome.org>). Sequences from other organisms were found using NCBI database (<http://www.ncbi.nlm.nih.gov/>). Homolog searches were performed using the NCBI Basic Local Alignment Search Tool (BLAST) server (<http://www.ncbi.nlm.nih.gov/BLAST/>). Multiple sequence alignments were done with ClustalW of the program MaC Vector (Accelrys) using default settings and displayed by the programs Amas and Alscript (Barton, 1993).

Secondary structure predictions were performed by the PredictProtein server (<http://cubic.bioc.columbia.edu/predictprotein>). Domain organization was predicted by SMART (a Simple Modular Architecture Research Tool) that allows the identification and annotation of genetically mobile domains and the analysis of domain architectures (<http://smart.embl-heidelberg.de/>; Schultz *et al.*, 1998).

A structural homolog search was performed by Dali server (Holm and Sander, 1995; <http://www.ebi.ac.uk/dali/>). Dali compares submitted coordinates of a query protein structure against those in the Protein Data Bank (www.rcsb.org). In favorable cases, comparing 3D structures may reveal biologically interesting similarities that are not detectable by comparing sequences.

2.6 Molecular cloning techniques

Oligonucleotide design and polymerase chain reaction (PCR). All PCR primers for cloning of genes were constructed in a following way. After an overhang of several nucleotides (nt) to assure efficiency of cleavage (as recommended by NEB catalogue), corresponding restriction sites were introduced followed by 20 to 25 nt of the gene sequence of interest. Point mutations were introduced by the PCR overlap extension method where two overlapping PCR products are produced carrying the

desired mutation in the primer. The resulting products serve as template in a second PCR round.

Template for PCR was *S. cerevisiae* genomic DNA unless specified otherwise. DNA was amplified with Herculase (Stratagene), where higher fidelity was essential, or *Thermus aquaticus* (Taq) DNA polymerase (Fermentas). Generally, 50 μ l reactions contained 200 μ M of each of the four dNTPs, 0.25 μ M of forward and reverse primer, and approximately 1 ng of template DNA. Additionally, in PCR reactions with Herculase, DMSO was varied between 0 and 6%, and with Taq Polymerase, 2 to 4 mM MgCl₂ was added. Thermocycling program contained 30-35 cycles. Times and temperatures of denaturation, annealing and elongation were varied to meet the special requirements of the polymerase and primer-template pairs used in different amplifications. PCR products were subsequently purified with QIAquick PCR Purification Kit Protocol (Qiagen).

Restriction cleavage and ligation. DNA was digested using restriction endonucleases (New England Biolabs (NEB)) as recommended by the producer. Digested products were purified over an agarose gel using QIAquick Gel Extraction Kit Protocol (Qiagen). Ligation of the digested DNA fragments and linearized vectors was conducted in 10 μ l total volume for one hour at room temperature using T4 ligase (NEB) and corresponding buffer. In most cases linearized vector was incubated with a large (approx. tenfold) excess of insert to improve efficiency of ligation.

Transformation of *E. coli* and isolation of plasmid DNA. Transformation of plasmid DNA into chemically competent *E. coli* was performed with standard heat shock protocols. After thawing a 50 μ l aliquot of chemically competent cells on ice, 1 μ l of ligation mixture or purified plasmid DNA was added. Cells were incubated on ice for 5 minutes, heat shocked at 42 °C for 45 seconds and incubated on ice for another two minutes. Finally, 0.5 mL of LB medium was added to the cells and the mixture was incubated for 30 minutes in a 37 °C shaker before plating.

E. coli cells from 5 ml of an overnight culture, grown from a single clone, were sedimented by centrifugation. Plasmid DNA was extracted from the pellet using the QIAprep Spin Miniprep Kit Protocol (Qiagen). Positive clones were verified by restriction analysis and DNA sequencing.

2.7 Preparation of competent cells

200 mL LB media were inoculated with 5 ml of an overnight culture of the desired bacterial strain. The cells were grown at 37 °C until an OD₆₀₀ of 0.4-0.55 was reached. After 10 minutes incubation on ice the cells were centrifuged for 10 minutes at 5,000 rpm (4 °C). All following steps were performed at 4 °C. The pellet was washed with 50 ml of TFB 1 buffer (see chapter 2.4, Table 8) and centrifuged again. This pellet was resuspended in 4 ml of TFB 2 buffer (see chapter 2.4, Table 8), before aliquoting and plunging into liquid nitrogen.

2.8 Electrophoretic methods

Electrophoretic separation of DNA. DNA was separated in horizontal 1x TBE, 1% agarose (w/v) gels containing ethidium bromide (0.7 µg/mL). Samples were mixed with 6x DNA loading dye (see chapter 2.4, Table 8). DNA was visualized using standard ultraviolet transilluminator ($\lambda = 254$ nm, Eagle Eye, Stratagene).

Electrophoretic separation of RNA. Electrophoresis was performed in 8% polyacrylamide gel (acrylamide:bisacrylamide (19:1)). The gel contained 7M urea and 10x TBE buffer (see chapter 2.4, Table 8). Immediately before pouring the gel, 0.1% APS (v/v) and 0.03% (v/v) TEMED were added. RNA samples were mixed with 2x RNA-loading dye (see chapter 2.4, Table 8) and denatured for 2 minutes at 65 °C. Electrophoresis was performed in 1x TBE buffer. The RNA was stained with 0.02% (v/v) SYBR-Gold Nucleic Acid Stain (Molecular Probes) and visualized at excitation

maxima for dye-nucleic acid complexes ($\lambda = 495$ nm) with Typhoon Variable Mode Imager (GE Healthcare).

Protein separation by SDS-PAGE. For protein samples glycine-SDS-PAGE with 12.5%-15% acrylamide gels (acrylamide:bisacrylamide (37.5:1)) (Laemmli, 1970) was performed. Gels were routinely stained with Coomassie staining solution. If very low amount of protein had to be visualized, silver staining was performed (1 h fixing, 2x 10 min washing, 1 min sensitizing, 3x 30 sec H₂O wash, 20 sec staining, 2x 30 sec H₂O wash, development, reaction stop).

2.9 Preparation of samples for Edman sequencing

For N-terminal sequencing proteins were separated on SDS-PAGE and stained with Coomassie staining solution. The protein band of interest was excised and dried in a speed vac. The dried gel piece of gel was rehydrated in the swelling buffer (see chapter 2.4, Table 8). Subsequently, 150 μ L of distilled water was added to set up a concentration gradient together with a small piece of pre-wet (methanol) polyvinylidene difluoride (PVDF) membrane (Schleicher & Schuell). Once the solution turned blue 10 μ L of methanol was added as a catalyst. After 1 to 2 days the membrane turned blue, indicating a complete transfer. The membrane was washed 5 times with 10% methanol. Protein of interest was N-terminally sequenced from the dry membrane in a PROCISE 491 sequencer (Applied Biosystems).

2.10 Protein expression in *E. coli* and selenomethionine labeling

Proteins were expressed recombinantly in *E. coli* if not stated otherwise. Plasmids containing the desired protein variants were transformed into competent *E. coli* BL21

DE3 cells (Stratagene). The cells were grown at 37 °C in LB medium supplemented with the antibiotic corresponding to the resistance cassette of the plasmid. Once the cells grew to an OD₆₀₀~0.5 cultures were cooled to 20 °C. Protein expression was induced by addition of 0.5 mM IPTG and cells were grown overnight. Cells were harvested by centrifugation (5000 rpm, SLC6000 rotor, 30 minutes) at 4 °C, resuspended in lysis buffer (see corresponding purifications) and flash frozen in liquid nitrogen. Cell pellets were stored at -80 °C.

Selenomethionine incorporation was essentially performed as described in (Meinhart *et al.*, 2003a). Plasmid DNA containing gene of interest was transformed into the methionine auxotroph *E.coli* strain B834 (DE3) (Stratagene). Bacteria were grown in LB medium supplemented with the appropriate antibiotic at 37 °C to an OD₆₀₀~0.5. Cells were harvested and resuspended in the same amount of minimal medium (Table) supplemented with selenomethionine (100 mg/l) and antibiotics. Cells were grown till the OD₆₀₀ increased by 0.2 at 37 °C to deplete the medium of any residual methionine. Cultures were cooled to 18 °C and protein expression was induced by the addition of 0.5 mM IPTG. Protein was expressed overnight.

2.11 Cloning and protein purification of CTD phosphatases

Table 9: Buffers for purification of CTD Phosphatases

Buffer	Description
Buffer A	300 mM NaCl, 50 mM Tris pH 8.0 (25 °C), 5% glycerol, 10 mM β-mercaptoethanol
Buffer B	50 mM NaCl, 50 mM Tris pH 7.5 (25 °C), 5% glycerol, 3 mM DTT
Buffer C	150 mM NaCl, 10 mM Hepes pH 7.5 (25 °C), 3 mM DTT

The gene for human Scp1 (or CTDSP1) was amplified from HUVEC cDNA (provided by G. A. Vlastos) and subcloned into pET21b vector (Stratagene). Variants of human Scp1 were expressed for 15 h at 20 °C in *E. coli* BL21 (DE3) RIL (Stratagene). Cells were harvested by centrifugation and resuspended in buffer A (Table 9). Cells were lysed by sonication. After centrifugation, the supernatant was loaded onto a 2 mL Ni-NTA column (Qiagen), equilibrated with buffer A. After washing with 10 mL of buffer A, bound protein was eluted with buffer A containing 200 mM imidazole, and in the case of full-length Scp1 was further purified by anion exchange chromatography (MonoQ, GE Healthcare). The column was equilibrated with buffer B (Table 9), and the protein was eluted with a gradient of 10 column volumes from 50 mM to 1 M NaCl. After concentration, the sample was applied to a Superose-12 HR gel filtration column (GE Healthcare) equilibrated with buffer C (Table 9). Pooled peak fractions were concentrated for crystallization to 20 mg mL⁻¹. Variants of *Saccharomyces cerevisiae* Fcp1 were obtained by PCR-amplification of the corresponding regions in the gene from yeast genomic DNA and were subcloned into pET21b (S. Heilmeyer). Expression and purification of variant Fcp1c (residues 168-606) was essentially as for Scp1. The variant Scp1ΔN comprising residues 77-261, as well as the active site mutants (D96N, D96A, D98N, T152S, T152V, T152A, K190A and N207D), with a C-terminal HisTag were subcloned into pET21b (Stratagene), expressed in *E. coli*, and purified as above, but without the anion exchange step. For MAD phasing, the Scp1ΔN mutant L165M/L205M, which contains two additional methionines at positions of conserved hydrophobic residues, was constructed with the PCR overlap extension method and selenomethionine was incorporated as described in chapter 2.10.

2.12 Limited Proteolysis

For chymotrypsin and trypsin treatment 1 µg of corresponding protease was added to 200 µg of purified protein. Digests were done in the buffers used for gel filtration and supplemented with CaCl₂ to a final concentration of 1 mM. The reaction mixture was

incubated at 37 °C for 1, 2, 3, 10, 30, and 60 minutes. The reactions were stopped by the addition of 5x sample buffer (see chapter 2.4, Table 8) and were heated to 95 °C for 5 minutes.

For proteinase K treatment 1 µl of dilutions of proteinase K (400 ng/µL, 40 ng/µL, 4 ng/µL, and 0.4 ng/µL) were added to the protein samples. The mixtures were incubated on ice for 1 h. The reactions were stopped by the addition of sample buffer and boiled as above. All samples were analyzed by SDS-PAGE. Bands of interest were passively transferred to PVDF membrane and analyzed by Edman sequencing as described in chapter 2.9.

2.13 Crystallization and structure determination

Samples were crystallized at 20 °C with the hanging drop method, using as reservoir solution 20-30% PEG 3300 or PEG 6000 or PEG 8000, 200 mM NH₄OAc, 100 mM citrate buffer pH 5.6, 5 mM DTT. The thin plate-like crystals grew to a maximum size of 400x150x5 µm. To the drops containing crystals, 100 µL cryo solution (25% PEG 3300 or PEG 6000 or PEG 8000, 15% PEG 400, 200 mM NH₄OAc, 100 mM citrate buffer pH 5.6, 5 mM DTT) were added, and after 10 min crystals were flash-cooled by plunging into liquid nitrogen. A MAD experiment was performed on crystals from the selenomethionine-labeled L165M/L205M double mutant at the Swiss Light Source (see chapter 3.1.1, Table 18). Data were processed with DENZO and SCALEPACK (Mutant L165M/L205M SeMet MAD data, Otwinowski and Minor, 1996), or with XDS (wild type beryllofluoride data, Kabsch, 1993). The crystals belong to space group P2₁2₁2. Program SOLVE (Terwilliger, 2002) was used for detection of 5 selenium sites (M76, M142, M165, M205, M229) and for MAD phasing (Z-score=43.3, FOM=0.62). Phases were further improved by SHARP (La Fortelle and Bricogne, 1997). The resulting electron density map allowed building of an atomic model, which was refined with CNS (Brunger *et al.*, 1998) against the remote wavelength dataset to 2.3 Å resolution (see chapter 3.1.1, Table 18). To trap the

beryllofluoride inhibitor in the active site, preformed wild type crystals were soaked with 5 mM BeCl₂/24 mM NaF for 10 min before flash-cooling and data collection at the Swiss Light Source. After phasing with the refined model, clear difference density at residue D96 indicated the presence of beryllofluoride. Except for minor changes in the water structure in the active site, the structure is essentially identical and was refined to 2.1 Å resolution (see chapter 3.1.1, Table 18). In both refined structures, 99% of the residues fall in allowed and additionally allowed regions of the Ramachandran plot, and none of the residues are in disallowed regions.

2.14 Phosphatase kinetics and inhibitor studies

A discontinuous indirect spectrophotometric assay for *p*-nitrophenol (pNP) was used to determine enzymatic parameters. Reactions were carried out at 37 °C in a total volume of 100 μL and contained 50 mM Tris-acetate pH 5.5, 10 mM MgCl₂, 0.19 μM Scp1ΔN (or 0.1 μM Fcp1c) and varying concentrations of *p*-nitrophenylphosphate (pNPP). Reactions were initiated by addition of the enzyme and quenched subsequently in 30 s or 60 s increments by adding 900 μL of 1 M sodium carbonate. Prior to enzyme addition, reaction mixtures were incubated for 5 min to allow for temperature equilibration. Rates of pNP release were determined by monitoring absorbance at 405 nm and interpolating the value to a pNP standard curve. A minimum of five colinear data points for each substrate concentration were used to carry out linear fits to obtain the initial reaction velocity (not shown). Velocities were plotted against substrate concentration to obtain standard Michaelis-Menten curves (not shown). To derive K_M and k_{cat} , velocities were plotted against velocity divided by substrate concentration according to Eadie-Hofstee (see Figure 11A). To obtain the molar extinction coefficient ϵ , a standard curve was determined by measuring the absorbance at 405 nm of known pNP concentrations (Sigma) in the assay mixture. This plot was linear between 5 and 150 mM pNP, and yielded $\epsilon = 18,300 \text{ (M} \times \text{cm)}^{-1}$. To study the effect of various inhibitors on phosphatase reactions, 100 μL reaction mixtures containing 10 mM pNPP and additionally different phosphatase inhibitors

were incubated for 30 min at 37 °C, quenched with 900 µL 1 M sodium carbonate, and the amount of released pNP was determined by measuring the absorbance at 405 nm with the use of the pNP standard curve (compare Fig. 11B).

2.15 Activity assay with active site mutants

To study the effect of individual active site amino acid replacements on phosphatase reactions, 100 µL reaction mixtures containing 10 mM pNPP, 50 mM Tris-acetate pH 5.5, 10 mM MgCl₂, and additionally 0.1 µM of different active site mutants (D96N, D96A, D98N, T152S, T152V, T152A, K190A and N207D) were incubated for 30 min at 37 °C. The reactions were quenched with 900 µL 1 M sodium carbonate, and the amount of released pNP was determined by measuring the absorbance at 405 nm.

2.16 Purification of Pol II core enzyme

Table 10: Buffers for Pol II core enzyme purification

Buffer	Description
3x freezing buffer	150 mM Tris-HCl, pH 7.9 (4 °C), 3 mM EDTA, 30% glycerol, 30 µM ZnCl ₂ , 3% DMSO, 30 mM DTT, 3x protease inhibitor mix
100x protease inhibitor mix (p.i.)	1.42 mg leupeptin, 6.85 mg pepstatin A, 850 mg PMSF, 1650 mg benzamidine, dry ethanol to 50 mL, stored at -20 °C; added immediately before use
1x HSB150	50 mM Tris-HCl, pH 7.9 (4 °C), 150 mM KCl, 1 mM EDTA, 10% glycerol, 10 µM ZnCl ₂ , 10 mM DTT, 1x protease inhibitor mix
1x HSB600	50 mM Tris-HCl, pH 7.9 (4 °C), 600 mM KCl, 1 mM EDTA, 10% glycerol, 10 µM ZnCl ₂ , 10 mM DTT, 1x protease inhibitor mix
1x TEZ	50 mM Tris-HCl, pH 7.5 (20 °C), 1 mM EDTA, 10 µM ZnCl ₂ , 5 mM DTT, 1x protease inhibitor mix
UnoQ buffer	50 mM Tris-HCl, pH 7.5 (20 °C), 1 mM EDTA, 10 µM ZnCl ₂ , 10% (v/v) glycerol, 10 mM DTT, no protease inhibitors
Pol II buffer	5 mM Hepes pH 7.3 (20 °C), 40 mM ammonium sulfate, 10 µM ZnCl ₂ , 10 mM DTT

Pol II was essentially purified as described in (Armache *et al.*, 2003). For three bead-beaters (BioSpec), up to 600 mL of cell suspension were thawed in warm water. Each bead-beater was filled with 200 mL of borosilicate glass beads (0.45-0.50 mm diameter), 1 mL of protease inhibitor mix (Table 10) and 200 mL of the cell suspension. HSB150 (Table 10) was added to fill the bead-beater completely, taking care to avoid any remaining air bubbles. Lysis was achieved within one hour of bead-beating (30 s on/90 s off) while the beater chambers were submersed in a salt/ice mixture. Glass beads were removed by filtration through a mesh funnel. The beads were washed with HSB150 until the flowthrough was clear. The lysate was cleared by two rounds of centrifugation (30 minutes at 12000 rpm in a GS3 rotor). Lipids were then removed by filtration of the supernatant through two layers of paper filter discs underneath a dressing cloth.

The cleared lysate was applied onto a 250 mL of Heparin Sepharose 6 FF (GE Healthcare) column, pre-equilibrated with 750 mL of HSB150. Elution was accomplished with 500 mL of HSB600 (Table 10). Proteins in the eluate were precipitated by adding 291 g of fine-ground ammonium sulfate per liter of eluate (~50% saturation), followed by 20 minutes of stirring at 4 °C and centrifugation (45 minutes at 12000 rpm in a GS3 rotor). The pellet was stored over night at 4 °C. The heparin column was restored by washing with 1 L of 6 M urea, followed by water, and stored in 5 mM potassium acetate and 20 % (v/v) ethanol.

The ammonium sulfate pellet was dissolved in 50 mL of TEZ buffer (Table 10). More TEZ was added to set the conductivity below the conductivity of TEZ containing additionally 400 mM ammonium sulfate (TEZ400). This sample was loaded by gravity flow onto the immunoaffinity column with coupled 8WG16 monoclonal antibodies (NeoClone, Madison/USA), specific for the unphosphorylated CTD of Pol II and optimized to release Pol II upon treatment with glycerol or ethylene glycol at room temperature (“polyol responsive antibody“, Thompson and Burgess, 1996). The column was pre-equilibrated with 20 mL of TEZ containing 250 mM ammonium sulfate (TEZ250). The flowthrough was re-loaded to yield 10-20% more Pol II. The column was brought to room temperature, washed with 25 mL of TEZ250 at room

temperature and Pol II was eluted in 1 mL fractions with TEZ500 containing additionally 50% (v/v) glycerol (ca. 15 mL). Directly afterwards, the column was washed with 5 mL of TEZ500 containing 70% (v/v) ethylene glycol but no DTT, and re-equilibrated with 25 mL of TEZ250 containing 0.02% sodium azide. Protein containing fractions were diluted six-fold and loaded onto a UnoQ column (BioRad, column volume 1.35 mL), pre-equilibrated with UnoQ buffer (Table 10) containing 60 mM ammonium sulfate (UnoQ-A). The column was washed with 3 column volumes of this buffer, and Pol II was eluted with a linear gradient over 10 column volumes from 0-50% buffer UnoQ containing 1 M ammonium sulfate (UnoQ-B). Pol II elutes at about 25% buffer UnoQ-B. The column was restored by washing with 5 column volumes of UnoQ-B.

Peak fractions were pooled and split into aliquots of 500 µg Pol II. The aliquots were mixed 1:1 with ammonium sulfate solution saturated at room-temperature, incubated for ~1 hour at 4 °C and centrifuged for 30 minutes at 4 °C in a table-top centrifuge at 14000 rpm. Most of the supernatant was decanted so that the pellet was still covered with supernatant, before it was shock-frozen in liquid nitrogen and stored at -80 °C. From 600 g yeast pellet, a yield of 5-8 mg of highly purified Pol II can be expected.

2.17 Purification of Rpb4/7 subcomplex

Table 11: Buffers for Rpb4/7 purification

Buffer	Description
Buffer A	150 mM NaCl, 50 mM Tris pH 7.5 (25 °C), 5% glycerol, 10 mM β-mercaptoethanol, protease inhibitors
Buffer B	50 mM Tris pH 7.5 (25 °C), 1 mM EDTA, 5 mM DTT

Recombinant yeast Rpb4/7 was cloned and overexpressed in *E. coli* using a vector having both subunits under the control of separate T7 promoters (Sakurai *et al.*, 1999). Cells from 2 L of culture were resuspended in buffer A (Table 11) and lysed by sonication. The lysate was cleared by centrifugation (30 minutes at 15000 rpm in a SS34 rotor) and applied onto a Ni-NTA column (Qiagen; 1 mL column volume). The column was washed with 5 mL of buffer A containing additionally 20 mM imidazole. Elution was performed with 6 mL of buffer A containing 200 mM imidazole. Peak fractions were pooled, diluted 1:3 with buffer B (Table 11) and applied on a ResourceQ column (GE Healthcare, 6 mL column volume), pre-equilibrated in buffer B. Rpb4/7 subcomplex was eluted with a linear gradient from 0-1 M NaCl in buffer B. Peak fractions were concentrated and applied on a Superose 12 HR10/30 gel filtration column (GE Healthcare), pre-equilibrated in Pol II buffer (see chapter 2.16, Table 10). The purified Rpb4/7 heterodimer was concentrated to 10 mg/mL and aliquots were stored at -80 °C.

2.18 Reconstitution of a Pol II-Fcp1 complex

For reconstitution of a Pol II-Fcp1 complex, a total of 0.5 mg of 10-subunit core Pol II in Pol II buffer (see chapter 2.16, Table 10) was incubated with a 5-fold excess of recombinant Rpb4/7 subcomplex at 20 °C for 45 min, followed by a 5-fold excess of Fcp1 with respect to Rpb4/7 and an additional incubation of 2 h. The Pol II-Fcp1 complex was separated from free Rpb4/7 and Fcp1 by gel filtration on a Superose 6 column (GE Healthcare).

2.19 Purification of TFIF variants expressed in *E. coli*

Table 12: Overview of TFIF expression trials in *E. coli*. Tfg1 and Tfg2 were cloned into pET21b vector and Tfg3 into pET24d. (“::” indicates deletion)

Bicistronic expression of Tfg1-His and Tfg2 (+ co-transformation of Tfg3)						
Tfg1 (735 aa, 82.2 kDa)		Tfg2 (400 aa, 46.6 kDa)		Tfg3 (244 aa, 27.4 kDa)		Outcome
Amino acid	MW (kDa)	Amino acid	MW (kDa)	Amino acid	MW (kDa)	
88-345	28.1	46-273	27.5	-		Tfg1(+), Tfg2(-)
56-661 (F1r1)	67.8	46-273	27.5	-		No overexpression
56-510 (F1r2)	50.8	46-273	27.5	-		No overexpression
56-442 (F1r3)	43	46-273	27.5	-		No overexpression
234-661 (F2r1)	48	46-273	27.5	-		No overexpression
234-442 (F2r3)	23.1	46-273	27.5	-		Tfg1(+poor exp., Edman), Tfg2(-)
56-661 (F1r1)	67.8	46-144:: 192-273	21.2	-		No overexpression
56-510 (F1r2)	50.8	46-144:: 192-273	21.2	-		Tfg1(+), Tfg2(+)
56-442 (F1r3)	43	46-144:: 192-273	21.2	-		Tfg1(+), Tfg2(+)
234-661 (F2r1)	48	46-144:: 192-273	21.2	-		Poor expression
234-442 (F2r2)	31	46-144:: 192-273	21.2	-		Tfg1(+), Tfg2(+, Edman)
234-442 (F2r3)	23.1	46-144:: 192-273	21.2	-		Tfg1(+), Tfg2(+, Edman)
234-735 (F2r0)	56.3	46-144:: 192-273	21.2	-		Tfg1(+), Tfg2(-)
56-735 (F1r0)	56.3	46-144:: 192-273	21.2	+	27.4	Edman:Tfg1(+), Tfg2(-), Tfg3(+)

Table 12 (continuation)

56-735 (F1r0)*	56.3	46-144:: 192-273	21.2	+	27.4	Tfg1(+), Tfg2(-), Tfg3(+)
234-735 (F2r0)*	56.3	46-144:: 192-273	21.2	+	27.4	Tfg1(+), Tfg2(Edman), 3(+)
56-735 (F1r0)*	56.3	46-273	27.5	+	27.4	Tfg1(+), Tfg2(?), Tfg3(+)
234-735 (F2r0)*	56.3	46-273	27.5	+	27.4	Tfg1(+), Tfg2(?), Tfg3(+)
56-510 (F1r2)*	50.8	46-144:: 192-273	21.2	+	27.4	Tfg1(+), Tfg2(?), Tfg3(+)
Expression of Tfg2-ht as single subunit						
Amino acids			MW (kDa)		Conclusion	
46-144::192-370 (FR1,Dloop)			32.6		Soluble	
46-370 (FR1)			38.9		Soluble	
209-370 (F1R1)			18.8		Soluble	
Expression of Tfg1 and Tfg2, each under the control of separate T7 promoter (+co-transformation with Tfg3)						
Tfg1 (735 aa, 82.2 kDa)		Tfg2 (400 aa, 46.6 kDa)		Tfg3 (244 aa, 27.4 kDa)		Conclusion
Amino acids	MW (kDa)	Amino acids	MW (kDa)	Amino acids	MW (kDa)	
234-735 (F2r0)His	56.3	46-370 (FR1)	38.9	-		Tfg1(+), Tfg2(-)
306-661 (F3r0)His	40.7	46-370 (FR1)	38.9	-		Tfg1(+), Tfg2(-)
234-735 (F2r0)His	56.3	46-144::192-370 (FR1,Dloop)	32.6	-		Tfg1(+), Tfg2(-)
306-661 (F3r0)His	40.7	46-144::192-370 (FR1,Dloop)	32.6	-		Tfg1(+poor exp.), 2(-)
234-735 (F2r0)	56.3	46-370 (FR1)His	38.9	-		No overexpression
306-661 (F3r0)	40.7	46-370 (FR1)His	38.9	-		No overexpression
234-735 (F2r0)His	56.3	46-370 (FR1)	38.9	+	27.4	Tfg1(+), Tfg2(-,MALDI), Tfg3(+)
306-661 (F3r0)ht	40.7	46-370 (FR1)	38.9	+	27.4	Tfg1(+), Tfg2(-), Tfg3(+)
234-735 (F2r0)His	56.3	46-144::192-370 (FR1,Dloop)	32.6	+	27.4	Tfg1(+), Tfg2(-), Tfg3(+)
306-661 (F3r0)His	40.7	46-144::192-370 (FR1,Dloop)	32.6	+	27.4	Tfg1(+), Tfg2(-), Tfg3(+)

Table 12 (continuation)

234-735 (F2r0)His	56.3	46-370 (FR1)His	38.9	+	27.4	Tfg1(+), Tfg2(-), Tfg3(+)
306-661 (F3r0)His	40.7	46-370 (FR1)His	38.9	+	27.4	Tfg1(+), Tfg2(-), Tfg3(+)

Table 13: Buffers for purification of TFIIIF variants expressed in *E. coli*

Buffer	Description
Buffer A	150 mM NaCl, 50 mM Tris pH 8.0 (25 °C), 10 mM β -mercaptoethanol
Buffer B	20 mM Hepes pH 7.0 (25 °C), 3 mM DTT
Buffer C	150 mM NaCl, 20 mM Hepes pH 7.5 (25 °C), 3 mM DTT

Different recombinant yeast TFIIIF variants were cloned and over expressed in *E. coli* (see Table 12). Cells from 2 L of culture were resuspended in buffer A (Table 13) and lysed by sonication. The lysate was cleared by centrifugation (30 minutes at 15000 rpm in a SS34 rotor) and applied onto a Ni-NTA column (Qiagen; 1 mL column volume). The column was washed with 5 mL of buffer A containing additionally 10 mM imidazole. Elution was performed with 3 mL of buffer A containing 200 mM imidazole. Peak fractions were pooled, diluted 1:3 with buffer B (Table 13) and applied on a Heparin column (GE Healthcare, 1 mL column volume), pre-equilibrated in buffer B containing 50 mM NaCl. TFIIIF variants were eluted with a linear gradient from 50 mM-1 M NaCl in buffer B. Peak fractions were concentrated and applied on a Superose 6 HR10/30 gel filtration column (GE Healthcare), pre-equilibrated in buffer C (Table 13).

2.20 Generation of a yeast strain overexpressing TFIIIF

An overexpression cassette containing *ADH1* promoter, multiple cloning site (MCS) and *ADH1* terminator sequence was subcloned into *E. coli*-yeast shuttle integrative (YIplac) vectors (YIplac128, YIplac204 and YIplac211) at *EcoRI* and *HindIII* restriction sites (see chapter 3.2.3, Figure 17). These vectors contain markers (*TRP1*, *URA2* and *LEU2*) which complement specific auxotrophic mutations in yeast DSY5 strain (Dualsystems Biotech) and allow selection of transformants containing the desired plasmids. Genes coding for Tfg1, Tfg2 carrying a C-terminal TAP-tag, and Tfg3, amplified from yeast genomic DNA, were subcloned into *XbaI* and *SalI* sites within the MCS of the overexpression cassette in YIplac128, YIplac204 and YIplac211, respectively. The YIplac204 plasmid carrying a gene for Tfg2-TAP under the control of the *ADH1* promoter was linearized with *EcoRV* restriction endonuclease within the *TRP1* gene and used for transformation of DSY5 yeast strain (Dualsystems Biotech). The linear ends are recombinogenic and direct integration to the site in the yeast genome that is homologous to these ends. A resulting yeast strain (DSY5-Int1) was recovered on a YPD selective plate lacking tryptophan. From a single clone, a yeast culture was grown and centrifuged. Resulting yeast pellet served for the transformation with YIplac211 plasmid carrying a gene for Tfg3 under the control of the *ADH1* promoter and linearized with *StuI* restriction enzyme within the *URA3* gene. A resulting yeast strain (DSY5-Int2) was recovered on a YPD selective plate lacking uracil and used as described for the transformation with the YIplac128 plasmid carrying a gene for Tfg1 under the control of the *ADH1* promoter and linearized with *EcoRV* restriction enzyme within the *LEU2* gene. Finally, resulting yeast strain (DSY5-Int3), containing all three subunits of TFIIIF, each under the control of the *ADH1* promoter, was recovered on a YPD selective plate lacking leucine.

2.21 Isolation of genomic DNA from yeast

2-5 mL of a dense yeast culture was centrifuged. The pellets were resuspended in 200 μ L of the buffer containing 100 mM Tris-HCl pH 8.0 (25 $^{\circ}$ C), 100 mM NaCl, 1 mM EDTA, 1% (w/v) SDS, 2% Triton X-100. 200 μ L of phenol-chloroform-isoamyl mixture (Roth) was added and 300 mg of glass beads. The suspension was agitated in the vibraxer at maximum speed for 5 minutes. 200 μ L of 1x TE buffer (see chapter 2.22, Table 14) was added and the suspension was centrifuged for 5 minutes at 14000 rpm. Upper layer was extracted, followed by chloroform wash and re-extraction. The genomic DNA in the upper phase was precipitated with 1 mL of pure ethanol followed by 1 min centrifugation. Additional wash with 70% ethanol was performed. The pellets were air-dried and dissolved in 100 μ L of TE buffer, containing RNase.

2.22 Yeast transformation

Yeast transformation was performed according to the lithium acetate-PEG procedure (Ito *et al.*, 1983). 50 mL of yeast culture (adequate for 5-6 transformations) were grown to OD₆₀₀ between 0.5 and 0.8. The culture was centrifuged for 5 minutes at 3600 rpm. The pellets were washed with 10 mL water and resuspended in 500 μ L of solution I (Table 14). Resulting suspension was centrifuged for 3 minutes at 3600 rpm. The pellet was resuspended in 250 μ L of solution I. For each transformation 50 μ L of the cell suspension was mixed with 1-4 μ g (3 μ L) of the linearized DNA, 10 μ g of carrier DNA (DNA from salmon testes, Sigma) and 300 μ L solution II (Table 14). The mixture was incubated at room temperature for 30 minutes. Subsequently, a heat shock was performed for 10 minutes at 42 $^{\circ}$ C followed by 3 minute incubation on ice. The cells were washed twice with water, centrifuged for 3 minutes at 3600 rpm. Resulting pellet was resuspended in 500 μ L of water and 100 μ L was spread on a selective media plate.

Table 14: Buffers and solutions for yeast transformation

Buffer	Description
10x TE	100 mM Tris pH 7.5 (25 °C), 10 mM EDTA
10x LiOAc	1 M LiOAc pH adjusted to 7.5 (25 °C)
Solution I	0.5 mL 10x TE, 0.5 mL 10x LiOAc, 4 mL H ₂ O (sterile filtered)
Solution II	0.5 mL 10x TE, 0.5 mL 10x LiOAc, 4 mL 50% PEG3350 (sterile filtered)
Selective plate media, pH 5.5 (25 °C)	0.67% (w/v) yeast nitrogen base (Formedia), 2% (w/v) glucose, 20 mg/L arginine (HCl), 80 mg/L aspartate, 100 mg/L glutamate, 50 mg/L isoleucine, 50 mg/L lysine (HCl), 20 mg/L methionine, 50 mg/L phenylalanine, 375 mg/L serine, 100 mg/L threonine, 50 mg/L tyrosine, 140 mg/L valine, 20 mg/L adenine, 20 mg/L histidine, 100 mg/L leucine, 40 mg/L tryptophan, 20 mg/L uracil
YPD-Trp ⁻	Selective plate media minus tryptophan
YPD-Leu ⁻	Selective plate media minus leucine
YPD-Ura ⁻	Selective plate media minus uracil

2.23 Small-scale Pol II-TFIIF tandem affinity purification (TAP)

Yeast pellet from 2 L of culture was thawed and filled with the lysis buffer (Tris 25 mM pH 7.9 (25 °C), 200 mM potassium acetate, 0.1% Igepal CA-630 (Sigma), 5 mM EDTA, 5% glycerol) to a total volume of 25 mL. The same volume of borosilicate beads, followed by the suspension, was added to a cold lysis container. Lysis was performed in bead-beater (Planeten-Monomühle, Fritsch) (2x 4 min on/1 min off). Borosilicate beads were removed over a 50 mL syringe with an inserted metal filter. The beads were washed by adding 10 mL of lysis buffer. The lysate was centrifuged at 4000 rpm for 10 minutes (4 °C). Supernatant was ultracentrifuged for 1 hour at 40000 rpm (TI45 rotor). The upper phase containing lipid flakes was removed with water pump. Glycerol was added to the clear lysate to a total concentration of 5%

prior to freezing. The clear lysate was submitted to fractional ammonium sulfate precipitation and fractions were analyzed by dot blot assay. Pol II-TFIIF complex completely precipitated at 45% ammonium sulfate saturation, the pellet was dissolved and incubated with 0.4 mL of pre-washed IgG slurry (GE Healthcare) for 1 hour at 4 °C. Following centrifugation (1800 rpm, 3 minutes), supernatant was removed and beads were thoroughly washed with 10 mL of lysis buffer. 150 µL of lysis buffer and 4 µg of TEV protease were added to the IgG beads. Subsequently, suspension was incubated for 1 hour at 20 °C. The protein was recovered and either loaded on the SDS-gel or loaded on the Superose 6 gel filtration column (GE Healthcare) and finally precipitated with trichloroacetic acid.

2.24 Dot blot assay

Dilution series of early steps of the TAP-purification were prepared in 96-well plates. Following transfer from a 96-well plate to a prewashed PVDF membrane (Schleicher & Schuell) with a 96-well vacuum dot-blotter (Schleicher & Schuell), the membrane was washed with 1x PBS buffer (see chapter 2.4, Table 8), blocked with 5%-milk powder solution (w/v), and subjected to a one-step immunoaffinity procedure with a peroxidase-anti-peroxidase conjugate (1.2 µg Ab/mL, Sigma). Bound IgG was detected by chemiluminescence (ECL, GE Healthcare). The detection limit of this method (1 fmol of protein) allows the quantization of proteins present at very low levels and processing of many samples rapidly and in parallel (Borgreffe *et al.*, 2001).

2.25 Large-scale isolation of Pol II-TFIIF complex from yeast

2.25.1 Yeast fermentation

Pol II-TFIIF complex was isolated from the *Saccharomyces cerevisiae* DSY5-Int3 strain (*MATa his3 pep4 prb1; TRP1::P_{ADH1}-TFG2-TAP-t_{ADH1}::TRP1; URA3-52::P_{ADH1}-TFG3-t_{ADH1}::URA3-52; LEU2::P_{ADH1}-TFG1-t_{ADH1}::LEU2*), produced during course of this work (see chapter 2.20). This strain carries three TFIIF subunits, each under the control of the *ADH1* promoter. Doubling time of DSY5-Int3 yeast strain is roughly 1h 45 min.

A fermenter (ABEC, Infors) of a nominal volume of 200 L was available for producing of up to 1.1 kg of a yeast pellet per batch. Table 15 shows the media composition and the culture parameters used for the fermentation of yeast.

Table 15: Conditions for the fermentation of yeast

YPD media	4 kg peptone 4 kg glucose 3 kg yeast extract 195 L water
antibiotics ¹	10 g ampicillin 2 g tetracycline
typical inoculate volume	6-7 L shaker culture, OD ₆₀₀ ~ 3
air flow	20 L/min
stirrer speed	220 rpm
pH	6.9 (30 °C)
typical growth time	8-10 hours
¹ added after sterilization, prior to inoculation	

2.25.2 Harvesting and storage of yeast

Yeast was harvested at $OD_{600} \sim 3.0$. Cells were pelleted by centrifugation flow centrifuge (Padberg Z4IG, 20000 rpm). From 200 liters of yeast culture 1.1-1.3 kg of the yeast cell pellet could be obtained. The cell pellet was resuspended in 350 mL of buffer A (50 mM Tris, 200 mM KOAc, 1 mM EDTA, 5% glycerol, 0,18 mM dodecyl-beta-D-maltosid, 0.5 mM DTT, 1x p.i. mix, pH adjusted to 6.9 at 25 °C) per kg cells. Resuspended cells were shock-frozen as 200 mL aliquots in liquid nitrogen and stored at -80 °C.

2.25.3 Large-scale purification of the Pol II-TFIIF complex

For three bead-beaters (BioSpec), up to 750 mL of cell suspension (corresponding to 100 L of yeast culture) were thawed in warm water. Each bead-beater was filled with 200 mL of borosilicate glass beads (0.45-0.50 mm diameter), 1 mL of protease inhibitor mix and ~250 mL of the cell suspension. Buffer A (see 2.25.2) was added to fill the bead-beater completely, taking care to avoid any remaining air bubbles. Lysis was achieved within one hour of bead-beating (30 s on/90 s off) while the beater chambers were submersed in a salt/ice mixture. Glass beads were removed by filtration through a mesh funnel. The beads were washed with buffer A until the flowthrough was clear. The volumes were kept as small as possible.

The lysate was somewhat clarified by centrifugation (30 min at 6000 rpm, SLC6000) or alternatively (12000 rpm, 2 x 30 min, GS3 rotor). Lipids were then removed by filtration of the supernatant through two layers of paper filter discs underneath a cheese cloth. The lysate was further centrifuged for additional 1 hr at 12000 rpm (SLA1000 rotor) and 1 hr at 15000 rpm (SS34 rotor). Consequential rounds of centrifugation can be replaced with 1 hr ultracentrifugation (40000 rpm, TI45 rotor).

IgG beads were washed three times with buffer A with no protease inhibitors and recovered by centrifugation at 1800 rpm. The clear lysate was divided into 3 batches

containing 5 mL of IgG beads and incubated overnight at 4 °C. The IgG beads were recovered by centrifugation and transferred into 5 Mobicol columns (5 mL, MoBiTec), followed by an intensive wash with Buffer A. To each Mobicol, 2 mL of buffer A (without protease inhibitors) and 25 µg of TEV protease were added and incubated on a rotating wheel (4 hr, 20 °C). Mobicols were eluted by gravity flow, and subsequently centrifuged to collect residual eluate. The elution fractions were concentrated at 4000 rcf in the Amicon Ultra Centrifugal Filter Devices (15 mL, 100 kDa cut-off, Millipore) to the volume of 2 mL (4000 rcf). Concentration was accompanied by significant protein precipitation. Concentrated eluate was loaded on a Superose 6 gel filtration column (GE Healthcare) previously equilibrated with buffer A. The fractions were collected and checked by SDS-electrophoresis.

The peak fractions containing Pol II-TFIIF stoichiometric complex were pooled and concentrated using small centricons (4 mL, 100 kDa cut-off, Millipore). Resulting sample was loaded on Superose 6 column (GE Healthcare) equilibrated with Pol II buffer, not containing any detergents (see chapter 2.16, Table 10). In such manner up to ~1 - 1.5 mg of Pol II-TFIIF complex could be obtained from 100 liters of yeast culture. Peak fractions were concentrated appropriately for intended use. The concentrated protein complex is stable for 2 days at 4 °C.

2.26 Purification of TFIIB

Table 16: Buffers for TFIIB purification

Buffer	Description
Buffer A	50 mM TRIS-HCl, pH 7.5 (25 °C), 150 mM NaCl, 10 mM β-mercaptoethanol, 1x protease inhibitors
Buffer B	50 mM Tris-HCl pH 7.5 (25 °C), 5 mM DTT

Full-length TFIIB with a C-terminal His-tag was expressed for 15 h at 20 °C in *E. coli* BL21 (DE3) RIL (plasmid kindly provided by K. Armache). Cells were harvested by centrifugation and resuspended in buffer A (Table 16). Cells were lysed by sonication. After centrifugation, the supernatant was loaded onto a 2 mL Ni-NTA column (Qiagen), equilibrated with buffer A. After washing with 10 mL of buffer A containing 1 M NaCl, bound protein was eluted with buffer A containing 250 mM imidazole. The protein was diluted 1:3 with buffer B (Table 16) and was further purified by cation exchange chromatography (MonoS 5/5, GE Healthcare). The column was equilibrated with buffer B containing 50 mM NaCl, and the protein was eluted with a gradient from 50 mM to 1 M NaCl. After concentration, the sample was applied to a Superose 6 HR gel filtration column (GE Healthcare) equilibrated with Pol II buffer (see chapter 2.16, Table 10).

2.27 Purification of TBP core

TBP core with an N-terminal His-tag (amino acids 61-240, Z. S. Juo) was expressed for 15 h at 20 °C in *E. coli* BL21 (DE3) RIL (Stratagene). Cells were harvested by centrifugation and resuspended in buffer A (25 mM Tris-HCl, pH 8.0 (25 °C), 500 mM NaCl, 50 mM ammonium acetate, 10% glycerol). Cells were lysed by sonication. After centrifugation, the supernatant was loaded onto a 2 mL Ni-NTA column (Qiagen), equilibrated with buffer A. After washing with 10 mL of buffer A containing 1 M NaCl, bound protein was eluted with buffer A containing 250 mM imidazole. The protein was purified by heparin sepharose column (GE Healthcare). The column was equilibrated with buffer A, and the protein was eluted with a gradient from 500 mM to 1 M NaCl. After concentration, the sample was applied to a Superose 6 HR gel filtration column (GE Healthcare) equilibrated with Pol II buffer (see chapter 2.16, Table 10).

2.28 Electrophoretic mobility shift assay

Table 17: Oligonucleotide sequences for assembly of the initially transcribing complex

Oligonucleotide	Sequence
Non-template DNA strand	5'-ACGGGCGCCTATAAAAGGCAGTACTAGTAACTAGTATTGAAAGTACTT GAGCTT-3'
Template DNA strand	5'-AAGCTCAAGTACTTACGCCTGGTCATTACTAGTACTGCCTTTTATAGGC GCCCGT-3'
RNA	5'-AAAGACCAGGC-3'

Nucleic acids corresponding to promoter region and early transcription bubble were analyzed for their capability to bind Pol II-TFIIF complex and general transcription factors TBP and TFIIB by electrophoretic mobility shift assay. In 15 μ L of Pol II buffer (see chapter 2.16, Table 10), 20 pmol of nucleic acids were incubated with 20 pmol of TBP, TFIIB and 10 pmol of Pol II-TFIIF complex in various combinations (see chapter 3.2.4, Figure 20) for 20 minutes at 20 °C. The samples were loaded on a 6% native polyacrylamide gel (Novex DNA Retardation Gels) in 1x TBE buffer (see chapter 2.4, Table 8). Electrophoresis was carried out at 4 °C at a voltage of 70-90 V for 2 hours. Nucleic acids were stained for 30 minutes with 0.01 % SYBR gold in 1x TBE buffer and visualized by Typhoon Variable Mode Imager (GE Healthcare) at 495 nm.

2.29 Assembly of an initially transcribing complex

TBP (230 μ g) was mixed with DNA/RNA (8 nmol) which lead to the formation of a reversible precipitate. After 10 minute incubation at 20 °C , TFIIB (560 μ g) was added and the incubation was continued. Finally Pol II-TFIIF (750 μ g) complex was added and altogether was agitated for 20 minutes and centrifuged for 5 minutes at 11000

rpm. The sample was loaded on a Superose 6 gel filtration column (GE Healthcare) using Pol II buffer (see chapter 2.16, Table 10) and peak fractions were analyzed by SDS-PAGE.

2.30 Isolation of RNA from protein preparations

Protein preparations after gel filtration were treated with 1 μ L (~23 U, Fluka) of Proteinase K in the presence of 5 mM CaCl_2 . Incubation was performed for 1 hour at 4 $^\circ\text{C}$ on a steering wheel. In order to extract RNA, phenol/chloroform/isoamyl alcohol (125:24:1, pH 4) was added to protein preparation in 1:1 ratio (v:v). Such suspension was vortexed, followed by 10 minutes long centrifugation at 13000 rpm (4 $^\circ\text{C}$). The upper aqueous phase containing the extracted RNA was transferred into a fresh Eppendorf tube. 20 μ g of glycogen (Fermentas) per milliliter of nucleic acid solution were added as a carrier for better precipitation and visualization of RNA pellets. 2-3 volumes of cold ethanol with addition of sodium acetate to a final concentration of 0.3 M, pH 5.2 (25 $^\circ\text{C}$) were added. This suspension was vortexed and stored for half an hour at -80 $^\circ\text{C}$ (or 10 minutes on dry ice). After 20 minutes of centrifugation at 13000 rpm (4 $^\circ\text{C}$), the supernatant was decanted. The precipitate was washed with cold 80% ethanol and centrifuged again for 10 minutes. The supernatant was discarded and the pellet was air-dried. Finally the RNA pellets were dissolved in 100 μ L of water. The integrity and size of RNA was checked on RNA denaturing polyacrylamide gel (see chapter 2.8).

2.31 Isolation of RNA from polyacrylamide gel

RNA bands were cut out from the polyacrylamide gel and immersed into an elution buffer (0.5 M Tris-HCl pH 7.0 (25 $^\circ\text{C}$), 0.1% (w/v) SDS, 0.1 mM EDTA, 10 mM MgCl_2). The samples were agitated overnight and in the morning buffer was exchanged and agitation continued for another 6 hours. Subsequently, the RNA

contained in the elution buffer was extracted and precipitated as described in chapter 2.30.

2.32 Identification of RNA

For the identification of isolated RNA in protein extracts, a modification of RLM-RACE (rapid amplification of cDNA ends) technique (GeneRacer Kit, Invitrogen) was used (Figure 7). Individual reactions were performed according to the protocols provided by producer unless stated here. All the reactions were performed in the presence of RNase inhibitors (RNaseOut, 40 U) and only in diethylpyrocarbonate (DEPC) treated water. After each enzymatic reaction with RNA, RNA was precipitated essentially as described in section 2.30 and pellets were dissolved in 8 μ L DEPC treated water.

The isolated RNA was subjected to RNase-free DNase I (NEB) digestion in order to remove contaminating genomic DNA. Precipitated RNA was treated with Poly A polymerase (5 U, Invitrogen) in the presence of 40 mM TRIS-HCl (pH 8.0 at 25 °C), 10 mM MgCl₂, 2.5 mM MnCl₂, 250 mM NaCl, 1 mM ATP (Fermentas), 50 μ g/mL BSA (NEB) for 30 minutes. Precipitated poly-A RNA was treated with tobacco acid pyrophosphatase in order to remove possible CAP structures which could hinder 5'-ligation of the RNA-oligo by RNA ligase in the next step. Such RNA with 5'-oligo of known sequence and 3'-poly-A tail was subjected to SuperScript III reverse transcriptase at 50 °C in the presence of oligo-dT primer (24 nt) for first-strand cDNA synthesis. The reaction was inactivated at 70 °C for 15 minutes, cooled and RNase H was added to degrade the RNA. Such RACE-ready cDNA with known priming sites on each end was used as a template for amplification by PCR. Taq polymerase (Fermentas) was used to produce 3' A-overhang-containing products for cloning into pCR4-TOPO vectors. Resulting plasmids were propagated in *E.coli* competent cells (One Shot TOP10) and isolated by QIAprep Spin Miniprep Kit Protocol (Qiagen). Finally the identity of RNA was determined by DNA-sequencing of the isolated plasmids.

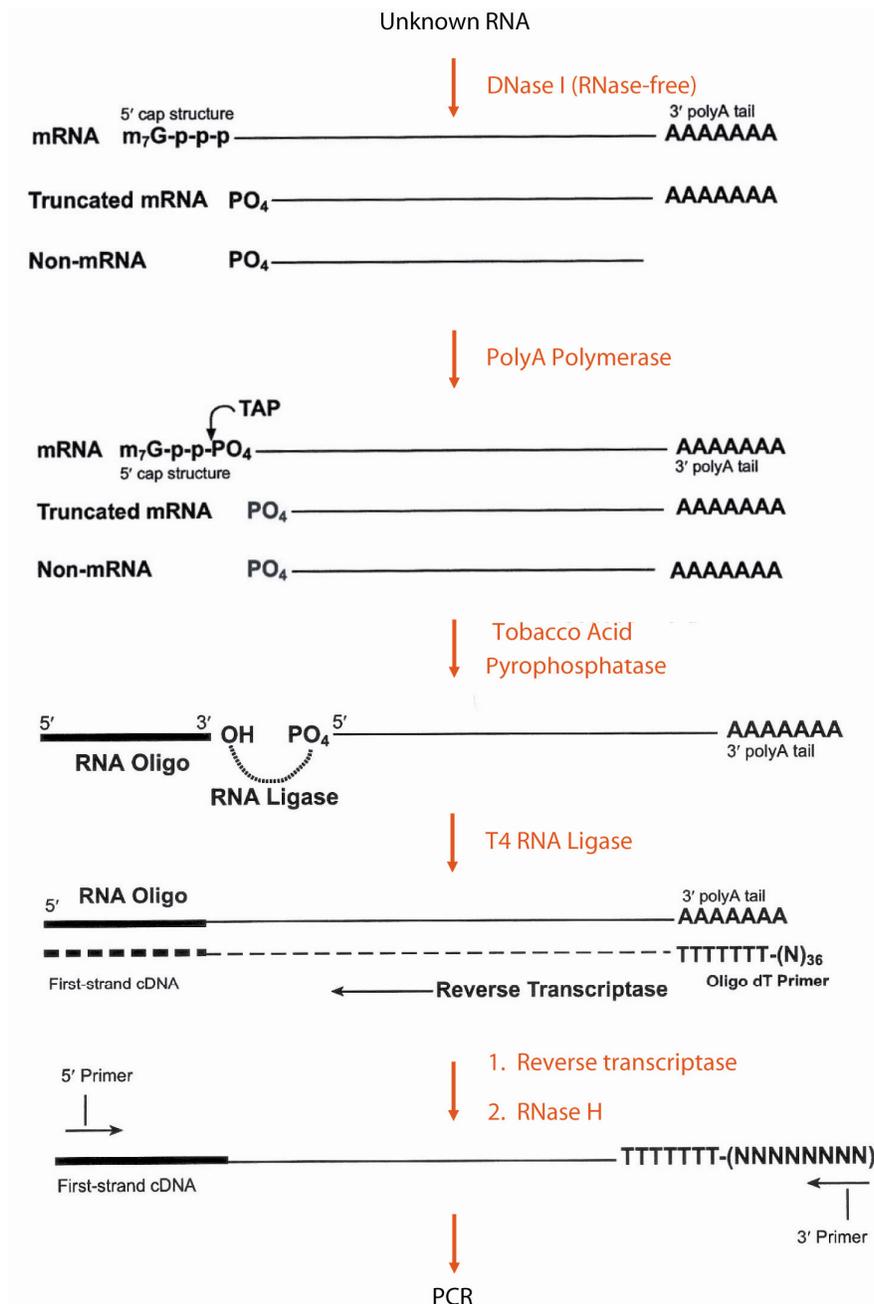


Figure 7: Identification of RNA by modified RLM-RACE procedure. RNA was isolated from purified Pol II-TFIIF protein sample by phenol-chloroform extraction and ethanol precipitation. Final product of the RACE procedure was amplified by PCR and cloned into pCR4-TOPO vectors for subsequent sequencing.

3 Results and Discussion

3.1 Structure and mechanism of Pol II CTD phosphatases

3.1.1 Domain organization and structure determination

A soluble core variant of the yeast *Saccharomyces cerevisiae* Fcp1, comprising the FCPH and BRCT regions could be expressed (Fcp1c, residues 168-606, Figure 8A). The protein was purified by Ni²⁺-NTA affinity chromatography, followed by an anion-exchange column (MonoQ) (Figure 8B) and a final gel filtration step (Superose12) (Figure 8B). The purified protein crystallized (Figure 8C); however, Fcp1c crystals were not suited for structural studies. Limited proteolysis of Fcp1c suggested that the FCPH and BRCT regions form structured domains connected by a partially exposed linker (Figure 8A). Since the catalytic FCPH domain was generally protected from proteolytic cleavage, Fcp1 variants comprising only the FCPH domain for crystallization were prepared, but these were not soluble under the conditions tested.

Scp1, which corresponds to the FCPH domain with an N-terminal extension, seemed better suited for crystallization (Figure 8A). When the full-length protein was expressed in *E. coli*, the 63 N-terminal residues of the extension were cleaved during expression (Fig. 8A), and consequently several Scp1 variants with N-terminal truncations were prepared. A variant lacking 76 N-terminal residues (Scp1 Δ N) was highly soluble. The protein was purified using a single gel filtration step after Ni²⁺-NTA affinity chromatography (Fig. 9A) and crystallized in a plate-like morphology (Figure 9B). Although these crystals were only about 5 μ m thick, they diffracted synchrotron radiation to better than 2 Å resolution. The X-ray structure was determined by multiwavelength anomalous diffraction with the selenomethionine-substituted double methionine mutant L165M/L205M of Scp1 Δ N (Table 18). The structure was refined at 2.3 Å resolution (Table 18). The refined structure has very good stereochemistry, and reveals chemical details.

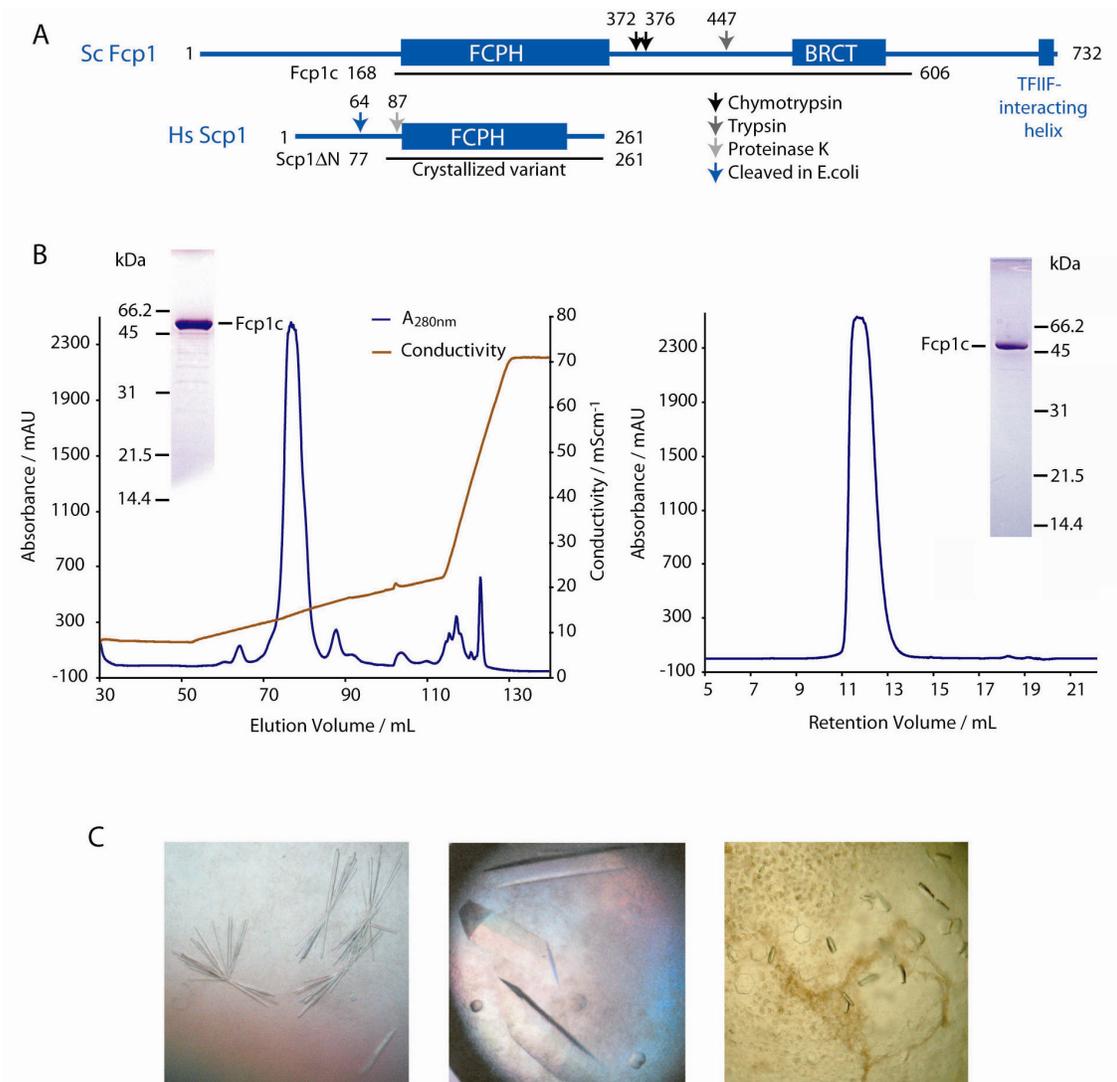


Figure 8: Purification and crystallization of Fcp1.

(A) Domain organization of *Saccharomyces cerevisiae* Fcp1 (Sc Fcp1), and human Scp1 (Hs Scp1). FCPH and BRCT denote the Fcp1 homology regions and the breast cancer protein related carboxy-terminal domains, respectively. The amino acids C-terminal of protease cleavage sites are depicted with arrows. Protein variants used in this study are indicated below the diagrams.

(B) Purification of Fcp1c by anion exchange chromatography (MonoQ) and gel filtration (Superose 12). Peak fractions are shown on accompanying SDS-gels.

(C) Crystallization of Fcp1c. Crystals were obtained in 20% PEG4000, 0.5 M Li_2SO_4 , 0.1 M MES pH 6.0 (*left*); 1.6 M MgSO_4 , 0.1 M NaMES pH 6.5 (S. Heilmeyer, *middle*); and 22% PEG4000, 0.7 M Li_2SO_4 , 0.1 M Hepes pH 7.5 (*right*).

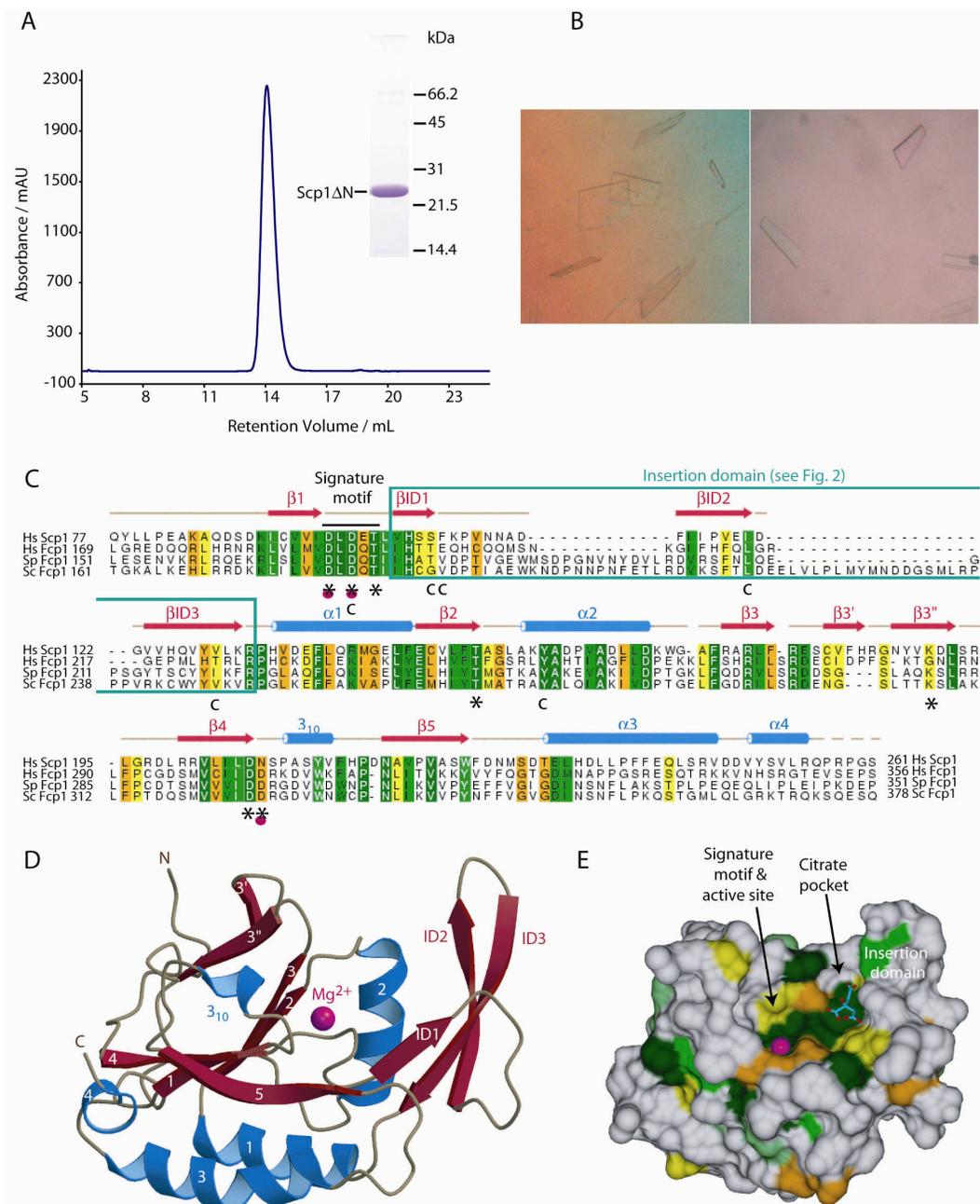


Figure 9: Structure of the CTD phosphatase Scp1.

(A) Purification of Scp1ΔN by gel filtration (Superose 12). Peak fraction is shown on accompanying SDS-gel.

(B) Crystallization of Scp1ΔN. Crystals were obtained in 20-30% PEG 3300 or PEG 6000 or PEG 8000, 200 mM NH₄OAc, 100 mM citrate buffer pH 5.6, 5 mM DTT. The thin plate-like crystals grew to a maximum size of 400x150x5 μm.

Figure 9 (continuation)

(C) Alignment of human Scp1 (Hs Scp1) with the catalytic FCPH domains of Fcp1 from human (Hs), *Schizosaccharomyces pombe* (Sp), and *Saccharomyces cerevisiae* (Sc). α -helices and β -strands are indicated as cylinders and arrows, respectively. Active site residues are marked with an asterisk below the alignment. Residues involved in metal ion and citrate binding are marked below the alignment with a pink dot and with a “c,” respectively. Conserved residues are highlighted with the degree of conservation decreasing from dark green to yellow.

(D) Ribbon model of Scp1. Secondary structure elements are colored according to (B). The catalytic metal ion is shown as a pink sphere.

(E) Surface conservation and putative specificity pocket. The molecular surface of Scp1 is colored according to amino acid residue conservation as in (B). A citrate ion that binds to the largely conserved putative specificity pocket is shown as a stick model.

Table 18: Structure determination and refinement

Crystal	Mutant (L165M/L205M) SeMet MAD			Wild type + BeF ₃ ⁻
Data collection				
Space group	P2 ₁ 2 ₁ 2			P2 ₁ 2 ₁ 2
Unit cell dimensions (Å)	117.6 x 47.2 x 40.0			117.8 x 47.2 x 40.1
Wavelength (Å)	0.9795 peak	0.9797 inflection	0.9686 remote	0.9919
Resolution (Å)	20-2.3 (2.38-2.3)	20-2.3 (2.38-2.3)	20-2.3 (2.38-2.3)	20-2.1 (2.2-2.1)
Completeness (%)	88.7 (78.7)	90.8 (81.9)	90.9 (80.1)	97.7 (87.5)
Unique reflections	9,386 (815)	9,639 (857)	9,635 (840)	13,356 (1,520)
Redundancy	17.4	17.2	17.0	3.8
R _{sym} (%)	6.3 (9.8)	5.2 (10.2)	5.2 (8.5)	5.7 (6.2)
Mean I/σ I	21.8 (15.5)	21.1 (12.7)	22.7 (14.0)	19.0 (16.8)
f'	-7.0	-10.0	-2.3	
f''	5.1	2.5	3.5	
Refinement				
Residues	181 (M76-Q255)			181 (M76-Q255)
Water molecules	131			148
Magnesium ion	1			1
Citrate ion	1			1
RMSD bonds (Å)	0.006			0.006
RMSD angles (°)	1.29			1.25
R _{cryst} (%)	18.7			20.7
R _{free} (%)	23.4			22.7

3.1.2 Overall structure

Scp1 forms a central 5-stranded parallel β -sheet with the strand order 3-2-1-4-5, flanked by two α -helices on one side and a two-stranded β -sheet and a short 3_{10} -helix on the other side (Figure 9D). The signature motif is located at the end of strand β 1, and together with other conserved residues lines a central depression that forms the active site of the enzyme. Although there is no sequence similarity outside the signature motif, a DALI search (Holm and Sander, 1995) revealed that the core fold of Scp1 largely corresponds to that of phosphoserine phosphatase (PSP, Wang *et al.*, 2001) and β -phosphoglucomutase (β -PGM, Lahiri *et al.*, 2003), both members of the DXDX(T/V) superfamily of phosphohydrolases and -transferases, and to 1, 2-L-haloacid dehalogenase (Ridder *et al.*, 1999) (Figure 10). The major structural difference between these enzymes lies in a large insertion after strand β 1, which may determine substrate specificity (“insertion domain,” Figures 9, 10). Within the core domain, Scp1 lacks a sixth β -strand that is present in the central β -sheet of the other enzymes (Figure 10).

Homology modeling of the FCPH domain of human Fcp1 showed that residues forming the hydrophobic core are conserved, although the overall sequence identity in the FCPH domains of human Scp1 and Fcp1 is only 21% (Figure 9C). This indicates that this structure is a valid model for the Fcp1 FCPH domain, and that mutational data on Fcp1 can be interpreted based on the Scp1 structure.

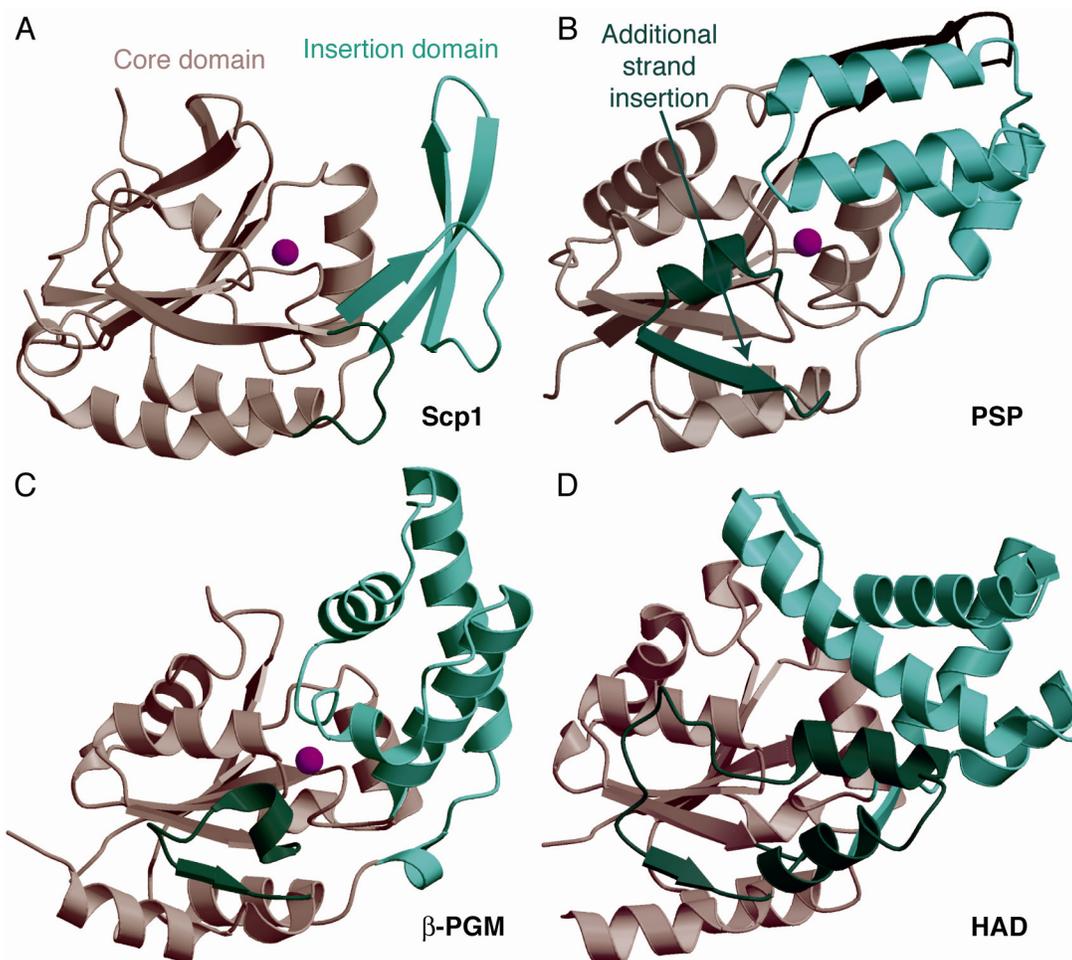


Figure 10: Comparison of DXDX(T/V) superfamily enzyme structures.

(A-D) Ribbon models of the DXDX(T/V) superfamily phosphotransferases Scp1 (A) (this study), phosphoserine phosphatase PSP (B) (PDB code 1F5S, Wang *et al.*, 2001), β -phosphoglucomutase β -PGM (C) (PDB code 1O03, Lahiri *et al.*, 2003), and 1, 2-haloacid dehalogenase HAD (D) (PDB code 1QQ5, Ridder *et al.*, 1999). Structural similarities were found with the DALI server (Holm and Sander, 1995), which detects fold similarities by superposing a query structure onto representative structures in the database and calculating the RMS deviation of carbon- α atom positions. The catalytic core domain of the DXDX(T/V) enzymes, which includes a central parallel β -sheet and its flanking regions, is shown in brown. The insertion domain of Scp1 and the corresponding regions in the other enzymes are shown in turquoise. The catalytic metal ion is depicted as a pink sphere.

3.1.3 Phosphatase activity

To investigate whether the structural resemblance of Fcp1 and Scp1 is reflected in comparable enzymatic properties, a highly purified recombinant human Scp1 Δ N and yeast Fcp1c were subjected to a spectrophotometric assay based on cleavage of the nonspecific substrate *para*-nitrophenylphosphate (pNPP), and the catalytic parameters were determined (see chapter 2.14). Both proteins readily cleaved pNPP in a pH-dependent manner, with maximum activity at pH 5.5. Michaelis-Menten kinetic analysis revealed K_M values of 6 mM and 32 mM for human Scp1 Δ N and yeast Fcp1c, respectively (Figure 11A). These K_M values are comparable to published values for endogenous *S. cerevisiae* Fcp1 (60 mM, Kobor *et al.*, 1999) and full-length recombinant *Schizosaccharomyces pombe* Fcp1 (19 mM, Hausmann and Shuman, 2002). The turnover numbers k_{cat} were 17 s⁻¹ and 33 s⁻¹ for human Scp1 Δ N and yeast Fcp1c, respectively. Thus the recombinant proteins used in these structural studies are active and the structural similarity of the FCPH domains of Scp1 and Fcp1 is reflected in similar enzymatic properties, although Fcp1 has slightly higher catalytic activity.

3.1.4 Specific inhibition

To investigate the catalytic mechanism, the effect of various inhibitors of phosphoryl-transfer reactions on the activity of Scp1 Δ N and Fcp1c were tested (Figure 11B). Phosphatase activity of both enzymes was essentially abolished by the beryllofluoride anion BeF₃⁻, produced *in situ* from BeCl₂ and NaF. In contrast, BeCl₂ or NaF alone did not have a strong effect on activity. The inhibitory effect of AlF₄⁻ was less severe, and sodium vanadate only inhibited activity slightly. Magnesium ions were essential for activity. The data indicate the presence of a catalytic aspartate residue, as BeF₃⁻ can form a stable tetrahedral adduct with catalytic aspartate side chains, mimicking a labile phosphoaspartate intermediate (Cho *et al.*, 2001b; Yan *et al.*, 1999), whereas AlF₄⁻ tends to mimic a phosphate (Chabre, 1990). The response of Fcp1 and Scp1 to

the inhibitors was very similar, indicating that these enzymes share a common catalytic mechanism.

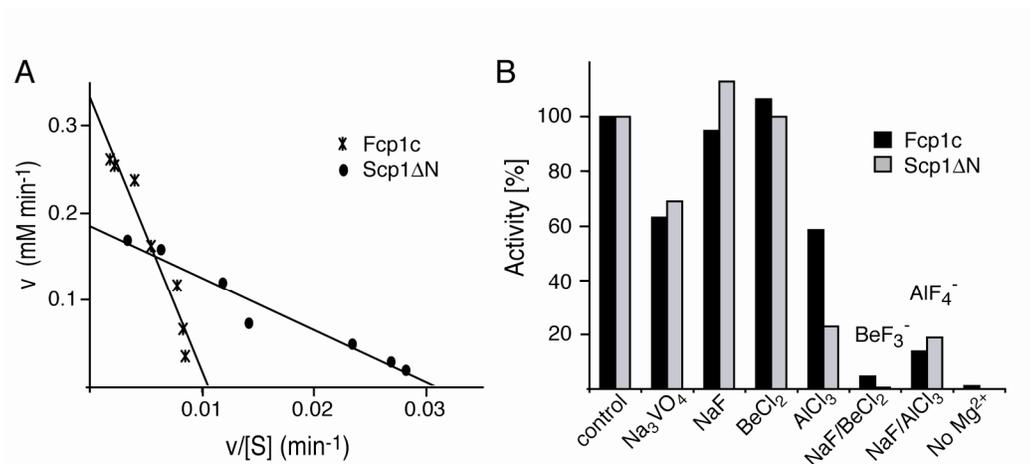


Figure 11: Catalytic activity and specific inhibition of CTD phosphatases.

(A) Determination of catalytic parameters. Catalytic activity of Scp1ΔN and Fcp1c was determined as described in chapter 2.14. To derive K_M and k_{cat} , the initial velocities v at seven different substrate concentrations were plotted according to Eadie-Hofstee against the velocities divided by the substrate concentration. The plot was fitted to the equation $v = V_{max} - K_M(v/[S])$.

(B) Specific inhibition. The inhibitors were added to reaction mixtures (see chapter 2.14) in the following concentrations: 1 mM sodium vanadate (Na_3VO_4), 1 mM sodium fluoride (NaF), 100 μM beryllium chloride (BeCl_2), 100 μM aluminum chloride (AlCl_3), a mixture of 1 mM sodium fluoride and 100 μM beryllium chloride (to produce the berylliofluoride anion BeF_3^- in situ), and a mixture of 1 mM sodium fluoride 100 μM aluminum chloride (to produce AlF_4^- in situ).

3.1.5 Active center

The enzymatic and inhibitory data are reflected in the active site structure. Mutational analysis of Fcp1 from *S. pombe* has defined 11 amino acid residues in the FCPH domain that are important for catalytic activity (Hausmann *et al.*, 2004; Hausmann

and Shuman, 2003). The Scp1 structure shows that seven out of these residues, including three residues in the signature motif, cluster in the central depression and form the active site (Table 19, Figure 12). The remaining four residues are not found in the active center depression, and are involved in salt bridges or hydrophobic core interactions, predicting that their mutation disrupts the domain structure (*S. pombe* Fcp1 residues R223, Y237, Y249, and D258, corresponding to Scp1 residues R132, F164, Y158, and D167, respectively, Figure 9C). The seven active site residues generally superimpose well with corresponding residues in structures of other enzymes of the DXDX(T/V) superfamily (Figure 12C). In the case of K190, the counterpart residues in PSP and PGM protrude from different locations of the protein backbones, but functional head groups occupy the same position. This is also true for the counterpart of D206 in PSP. Consistent with the requirement for metal ions in catalysis, the active site comprises a metal ion that is coordinated by residues D96, D98, N207, and two water molecules. The metal ion shows the highest peak in a difference Fourier map phased with the final model lacking any water molecules or ions (9.8σ).

Table 19: Active site residues in DXDX(T/V) phosphotransferases and –hydrolases¹

H.s. Scp1	S.c. Fcp1	S.p. Fcp1	PSP	β -PGM	Functional or structural role
D96	D180	D170	D11	D8	Phosphoryl acceptor, metal coordination
D98	D182	D172	D13	D10	General acid/base, metal coordination
T100	T184	T174	T15	V12	Positions side chains of D96 and N207
T152	T270	T243	S99	S114	Transition state stabilization
K190	K307	K280	K144	K145	Transition state charge stabilization
D206	D324	D297	D171	E169	Salt bridge with K190 in Scp1
N207	D325	D298	D167	D170	Metal coordination

¹Scp1, small CTD phosphatase 1; Fcp1, TFIIIF-dependent CTD phosphatase 1; PSP, phosphoserine phosphatase (PDB codes 1F5S, 1J97); β -PGM, β -phosphoglucomutase (PDB codes 1O08, 1O03).

²These residues protrude from different sites of the protein backbones but their functional head groups are at positions equivalent to those in the Scp1 structure (see Figure 12C).

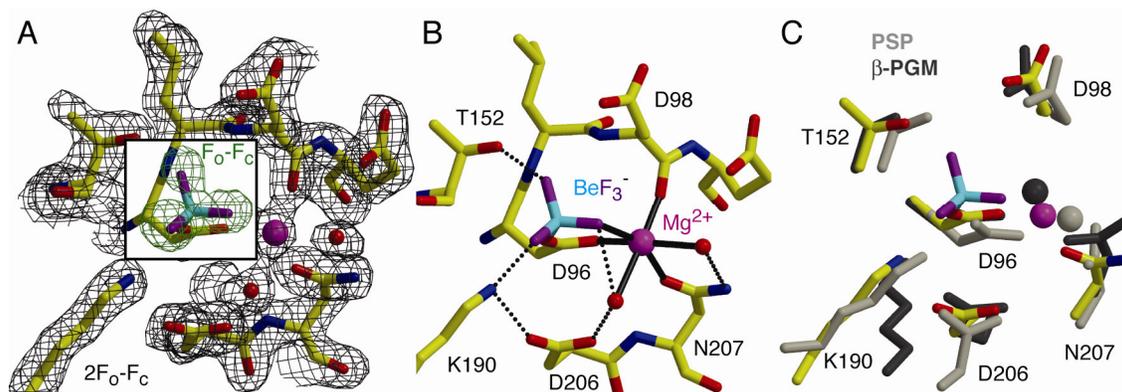


Figure 12: Active site and mimicry of the phosphoaspartate intermediate.

(A) Electron density. The final $2F_o-F_c$ density is shown for residues in the active site (black, contoured at 1.4σ). For the berylllofluoride anion, which mimics the phosphoaspartate-96 intermediate, the F_o-F_c density calculated from the model lacking the berylllofluoride is shown (green, contoured at 3.0σ), and the $2F_o-F_c$ density is omitted.

(B) Active site interactions. Hydrogen bonds are shown as dotted lines, and metal ion-ligand interactions as solid black lines. The metal ion is coordinated by the D96 side chain oxygen (2.3 Å), by the D98 carbonyl oxygen (2.1 Å), by the N207 side chain oxygen (2.3 Å), by a fluorine atom in the berylllofluoride (2.2 Å), and by two water molecules (2.4 Å, 2.5 Å).

(C) Superposition of active site residues in Scp1 (yellow), phosphoserine phosphatase (light grey, PDB code 1F5S, Wang *et al.*, 2001), and β -phosphoglucosmutase (dark grey, PDB code 1O08, Lahiri *et al.*, 2003). Six out of seven active site residues are shown.

3.1.6 Catalytic mechanism

The structural and functional data strongly suggest that the catalytic mechanism of Scp1 and Fcp1 involves the metal-dependent formation of a phosphoaspartate intermediate. This mechanism was suggested for *S. pombe* Fcp1 based on biochemical data (Hausmann and Shuman, 2003), and was shown for other DXDX(T/V) superfamily enzymes, which use the N-terminal aspartate in the signature motif as the phosphoryl acceptor (Cho *et al.*, 2001b; Lahiri *et al.*, 2003).

To directly show that the first aspartate in the signature motif acts as the phosphoryl acceptor in Scp1, the crystals were soaked with trifluoroberyllate (beryllofluoride). A difference Fourier map revealed a berylliofluoride anion bound to one oxygen atom of the D96 side chain carboxylate (Figure 12A). The metal ion is bound to the other D96 side chain oxygen atom, and shows an approximate octahedral coordination with metal-ligand distances of 2.0-2.3 Å (Figure 12B). A metal ion is found at an equivalent position in structures of PSP (Wang *et al.*, 2001) and β -PGM (Lahiri *et al.*, 2003). Thus berylliofluoride mimics the labile phosphoaspartate intermediate formed at residue D96 in the active center of Scp1. Consistently, mutation of this aspartate in Scp1 Δ N (Figure 13A) or Fcp1c (not shown) to alanine or asparagine abolished activity. The same mutation abolishes activity of full-length *S. cerevisiae* Fcp1 (Kobor *et al.*, 1999) and *S. pombe* Fcp1 (Hausmann and Shuman, 2002). These data establish D96 as the phosphoryl acceptor in Scp1 (Figure 13B). The conservative Scp1 mutation D96E retains some activity (Yeo *et al.*, 2003), apparently because the carboxylate group can still act as the phosphoryl acceptor.

The functional role of the other active site residues can be inferred from their relative location and from comparison with other enzymes of the DXDX(T/V) superfamily. In addition to metal ion binding, D98 may act as a general acid/base, donating a proton to the leaving group (the CTD serine side chain). D98 may also position and activate a water molecule for the second step, dephosphorylation of the phosphoaspartate and regeneration of the free D96 side chain. The metal ion and the head groups of the side chains of T152 and K190 all lie in a plane slightly above the position of the beryllium atom, and thus within a presumed equatorial plane of a pentavalent trigonal bipyramidal transition state. The metal ion and residues T152 and K190 are therefore expected to stabilize the geometry of the transition state and to partially neutralize its charge. Mutational analysis supports the proposed roles of the individual residues that form the active center (Figure 13A). While alanine replacements of T152 and K190 completely abolish enzymatic activity, T152S and N207D retain some activity due to the similar chemical properties.

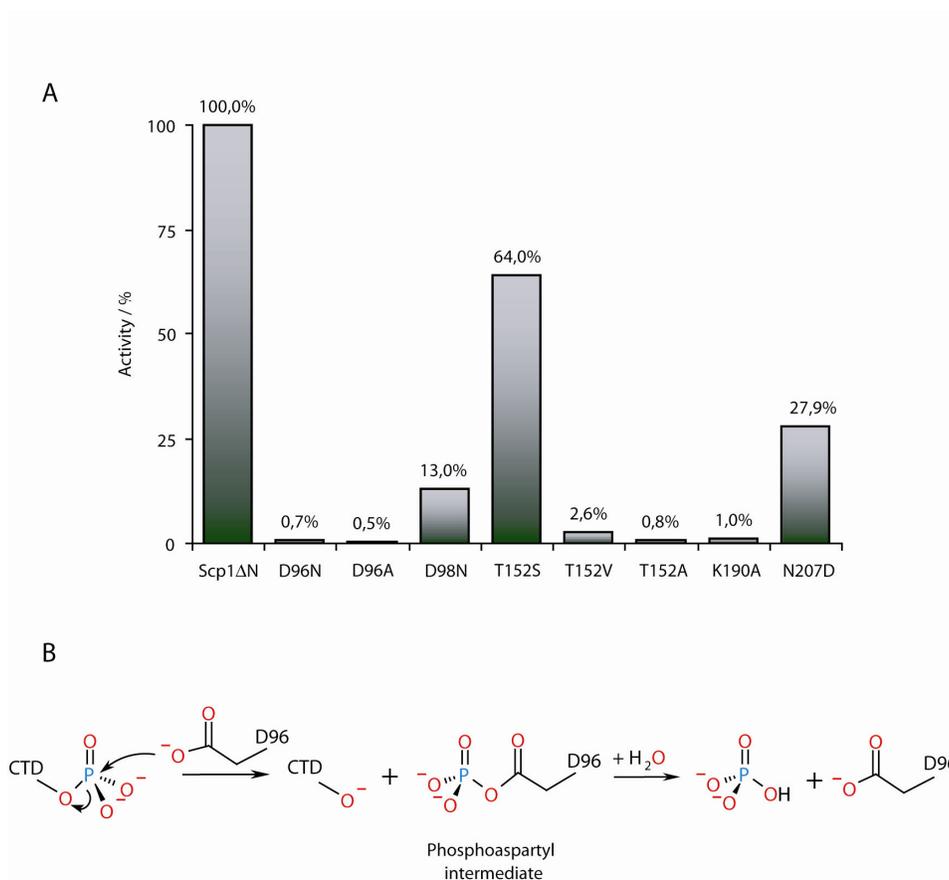


Figure 13: Mechanism of CTD phosphatases of the Fcp1 family.

(A) Mutational analysis of the active site mutants. Each mutant was tested for ability to cleave nonspecific substrate pNPP as described in chapter 2.15.

(B) Proposed mechanism of CTD phosphatases Scp1/Fcp1. The reaction mechanism involve the formation of a phosphoaspartyl intermediate, followed by recycling of the catalytic side chain with the use of a nucleophilic water molecule. Mechanism was drawn with MDL ISIS/Draw.

3.1.7 CTD specificity

Specificity for the CTD may result from recognition of CTD residues neighboring the phosphorylated target side chain. There is evidence that CTD recognition involves the specific insertion domain of Scp1 and Fcp1, which directly follows the signature motif (Figures 9C, 10A). The insertion domain and the core fold line a deep pocket, which binds a citrate ion that was present in the crystallization solution (Figure 9E). The citrate is only 8 Å away from the berylliofluoride. The pocket is largely hydrophobic,

with many of the lining residues conserved between Scp1 and Fcp1, including several residues that contact the citrate (Figure 9C). The insertion domain of Scp1 forms a 3-stranded β -sheet, and sequence conservation indicates that it must be similar in Fcp1, except for the two loops of the sheet that are longer in yeast Fcp1 (Figure 9C). The insertion domain of PSP, β -PGM and HAD is involved in substrate binding, but has a totally different structure (Figure 10), as expected for different substrate specificities. Together these findings suggest that the Fcp1/Scp1 pocket between the insertion domain and the active site binds the CTD and confers substrate specificity. Mutational analysis of the presumed substrate pocket could confirm this proposal.

Consistent with CTD binding to the specificity pocket, *S. pombe* Fcp1 activity requires at least four N-terminal and two C-terminal CTD residues flanking phosphoserine 2, and single alanine mutations of the flanking Tyr1 and Pro3 decrease activity sixfold (Hausmann *et al.*, 2004). However, Fcp1 and Scp1 can dephosphorylate both Ser5 and Ser2, with some preference for either of these serines (Hausmann and Shuman, 2002; Lin *et al.*, 2002a; Yeo *et al.*, 2003). Since both serines are flanked on the C-terminal side by a proline, it is likely that the CTD phosphatases bind the adjacent prolines Pro3 or Pro6, and preferential dephosphorylation of Ser2 and Ser5 is achieved by binding to other nearby residues. Indeed, the Pro3 side chain binds to a hydrophobic pocket in the known CTD peptide complex structures (Fabrega *et al.*, 2003; Verdecia *et al.*, 2000; Meinhart and Cramer, 2004) and specific recognition of a flanking proline is consistent with Fcp1 inhibition by the prolyl isomerase Pin1 (Xu *et al.*, 2003). A better understanding of CTD specificity however requires structure determination of Fcp1 or Scp1 in complex with a phosphorylated CTD peptide.

3.1.8 The Pol II subcomplex Rpb4/7 recruits Fcp1

Specificity of Fcp1 for the CTD may not only result from direct recognition of the CTD residues but also from binding of Fcp1 to a docking site on Pol II distinct from the CTD (Chambers *et al.*, 1995). To investigate if the docking site is on the ten-subunit

core of Pol II or on the heterodimeric polymerase subcomplex Rpb4/7, reconstitution of Fcp1 with Pol II core complexes and complete Pol II (Pol II core plus recombinant Rpb4/7, Armache *et al.*, 2003) was tried. To this end pure endogenous yeast Pol II core with an excess of recombinant Fcp1c was incubated, either in the presence or the absence of an additional excess of recombinant Rpb4/7 complex, and the samples were subjected to size exclusion chromatography. In both cases, two separated peaks were obtained. In the absence of Rpb4/7, the first peak corresponded to the Pol II core alone, and the second peak contained Fcp1c. In the presence of Rpb4/7, however, the first peak contained all 12 subunits of Pol II and Fcp1c, representing a reconstituted Pol II-Fcp1c complex (Figure 14). Since formation of a Pol II-Fcp1c complex relied on the presence of Rpb4/7, Fcp1 interacts with Pol II mainly via the Rpb4/7 subcomplex.

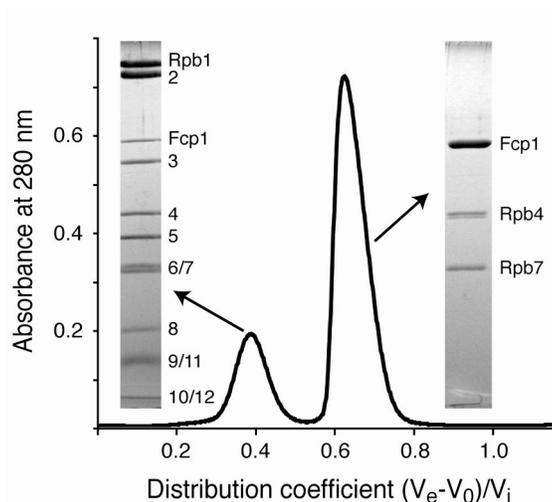


Figure 14: Reconstitution of the Pol II-Fcp1 complex

Purified endogenous yeast Pol II core enzyme was incubated with an excess of recombinant Rpb4/7 subcomplex and Fcp1c and the resulting mixture was separated on a Superose 6 gel filtration column. The elution profile was normalized by the distribution coefficient ($V_e =$ elution volume, $V_0 =$ void volume, $V_i =$ inner volume as determined by retention of acetone). The two peak fractions were analyzed by SDS-PAGE. The first peak corresponds to a reconstituted Pol II-Fcp1 complex.

Recently an additional interaction between Fcp1 and Pol II was reported, distinct from both the CTD and Rpb4/7 subcomplex (Suh *et al.*, 2005). The authors speculated that this interaction might mediate the stimulatory effect of Fcp1 on transcription elongation. These multiple interactions may play a role in regulating Fcp1 activity during the transcription cycle, and may also allow for dephosphorylation of the repetitive CTD in a pseudo-processive manner, although isolated Fcp1 is a distributive enzyme, apparently associating with the CTD and dissociating from it for each catalytic cycle (Hausmann *et al.*, 2004).

3.2 Preparation of the Pol II-TFIIF complex

3.2.1 Expression of TFIIF in *E. coli*

Initiation complex assembly requires several general transcription factors, among them multisubunit factors (TFIIF and –H), which for a long time have not been available in large quantities by overexpression but had to be obtained in lengthy multi-step purification procedures with very low yields. Whereas yeast TBP and the small general factors TFIIB and TFIIA can be produced in large quantities and pure form by overexpression in bacteria (Chasman *et al.*, 1993; Geiger *et al.*, 1996), large-scale preparation of the yeast general transcription factors TFIIF and –H is challenging since these are multiprotein complexes. Expression of single subunits of these complexes in most cases leads to insoluble samples.

Bacterially expressed human RAP30 and RAP74 can associate *in vitro*, but very inefficiently even after denaturation of the two proteins with urea and subsequent renaturation (Burton *et al.*, 1988; Wang *et al.*, 1993). This observation is explained by the structure of the RAP74/RAP30 dimerization domain (see chapter 1.4, Figure 4A). Since β -strands from both the RAP30 and RAP74 subunits contribute to the three intertwined β -barrels (Gaiser *et al.*, 2000), it seems that formation of the triple barrel requires co-folding of subunits, as opposed to docking of stable monomers (Wei *et al.*, 2001). Co-expression of human RAP30 and RAP74 (Tan *et al.*, 1994) as well as *S. pombe* TFIIF in insect cells (Spahr *et al.*, 2003) yielded an assembled form of active TFIIF. Co-expression of human RAP30/74 in bacteria was also possible (Tan *et al.*, 1995).

Biochemical analysis of yeast TFIIF has been hampered by the inability to produce recombinant yeast TFIIF due to the reported toxicity of the Tfg1 subunit in *E. coli* (Henry *et al.*, 1994). Only one successful trial to express yeast Tfg1-Tfg2 complex that decreased toxicity/instability of the expression plasmid was reported. It included

the use of a low copy broad host range vector for cloning of Tfg1 and Tfg2 under the control of the adjacent T7 promoters, a specific combination of different tags, as well as the induction of expression by infection with bacteriophage λ in a recombinant *E. coli* strain lacking T7 RNA polymerase (Ziegler *et al.*, 2003).

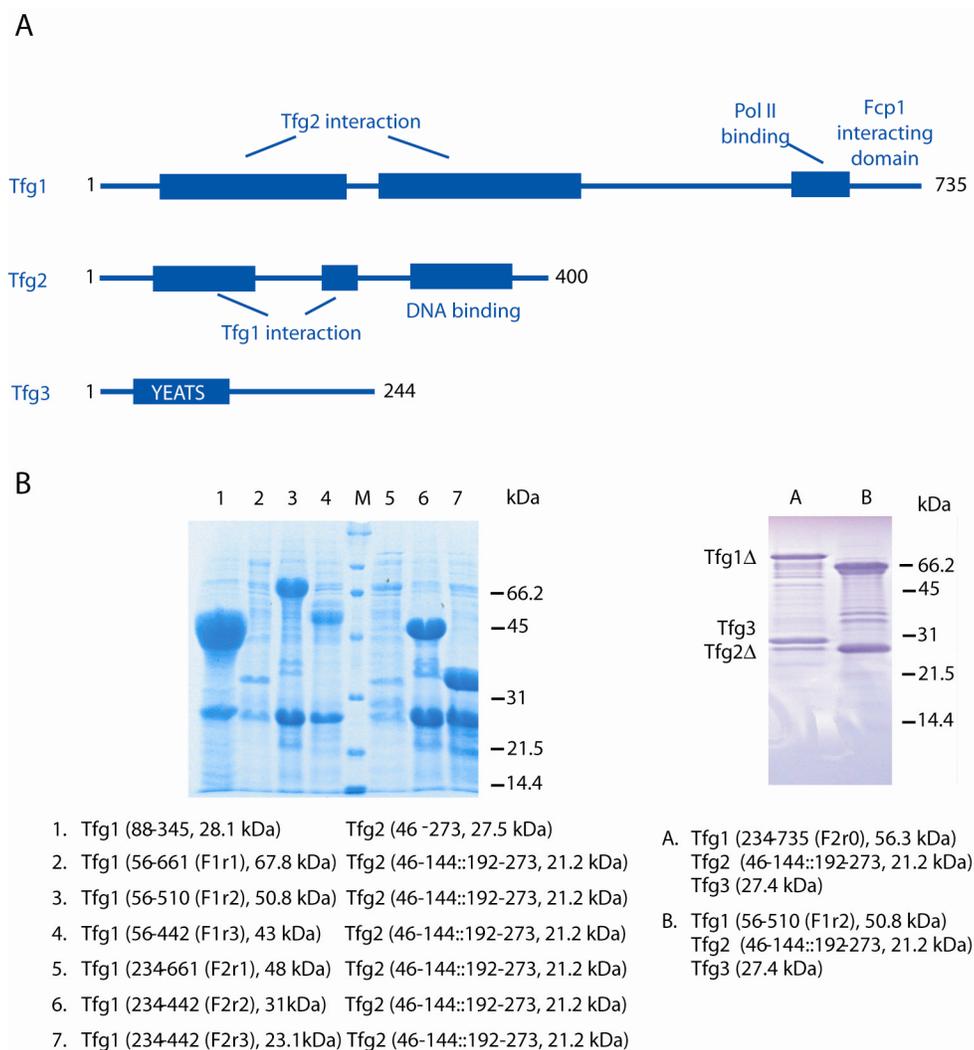


Figure 16: Domain organization of TFIIIF and *E. coli* expression trials.

(A) Domain organization of yeast TFIIIF.

(B) Variants of TFIIIF expressed in *E. coli*. Truncated variants of Tfg1 and Tfg2 could be coexpressed using a bicistronic vector, as shown on the SDS gel after Ni-NTA affinity purification (*left*). Truncated Tfg1 (234-735) and Tfg2 (46-144::192-273) could be additionally coexpressed from bicistronic vector by co-transformation with full-length Tfg3. SDS gel after a Superose 6 gel filtration column is shown (*right*). “::” indicates deletion.

Two possible routes for obtaining large quantities of the yeast TFIIIF in free form and bound to Pol II were explored during this work. Together with the research technician in our laboratory, Claudia Buchen, expression of the yeast 3-subunit TFIIIF in *E. coli* was tried. Many recombinant variants were produced (chapter 2.19, Table 12), and different expression strategies were tried. Those included co-transformation of plasmids carrying individual subunits, bicistronic expression, and expression of Tfg1 and Tfg2 subunits under the control of adjacent T7 promoters on a single vector and co-transforming with Tfg3 in pET expression system. While TFIIIF comprised of truncated subunits could be expressed (Figure 16), a 3-subunit, full-length initiation competent TFIIIF could not be expressed in *E. coli*.

Moreover, a reconstitution of the TFIIIF complex was tried by mixing the singly expressed subunits in variety of combinations. None of these trials, however, resulted in a complete and stoichiometric three-subunit complex, capable of binding Pol II. This leads to the conclusion, that it is not possible to produce a complete yeast TFIIIF in *E. coli* under the conditions tested. Variants of Tfg2 were expressed alone in *E. coli*, but did not bind the 12-subunit Pol II. The overview of different trials of bacterial expression of TFIIIF is given in the chapter 2.19, Table 12.

3.2.2 Isolation of TFIIIF from yeast

Consequently we switched to the yeast system with the help of Dr. Katja Sträßer (Gene Center) who produced a strain, with introduced tandem affinity purification (TAP)-tag on either Tfg1 or Tfg2. TAP is based on two successive affinity chromatography steps (Rigaut *et al.*, 1999; Puig, *et al.*, 2001). The tag fused to a target protein is composed of protein A having very high affinity for IgG, a TEV protease cleavage site, and the calmodulin-binding peptide having high affinity for calmodulin. TAP purification enables isolation of native protein complexes as well as co-purifications and identification of *in vivo* interacting proteins.

For TAP purification, yeast extract containing the TAP-tagged TFIIF was mixed with the IgG affinity resin. Bound TFIIF was released by TEV protease. This eluate was further used for gel filtration. Soluble stoichiometric TFIIF complex was obtained. Due to the high affinity of TFIIF for Pol II, a significant portion of Pol II stayed bound to TFIIF even after more stringent salt washes. However, after extensive purification optimization and up-scaling trials, the amount of the soluble material obtained was not sufficient for structural studies. Major obstacles were high losses during gel filtration and during the concentration step. These trials established, however, the importance of a non-ionic detergent for successful isolation of the complex from yeast extract, its solubility and enrichment.

The initial trials included Nonidet P-40 (Igepal CA-630), a non-ionic detergent which has a high absorption at 280 nm making a high background, and therefore it is unsuitable for chromatographic purification. Further, this impaired concentration measurements with standard protein concentration measurement assays. In addition, it is a very heterogeneous compound highly inappropriate for structural studies that require crystallization. The new detergent of choice was dodecyl- β -D-maltoside. In buffer containing this detergent, the protein eluted with a Gaussian-shaped peak. Dodecyl- β -D-maltoside is often used in structural studies of membrane-associated proteins and does not absorb at the wavelengths which are routinely used for protein concentration determination.

3.2.3 Overexpression of TFIIF in yeast

In order to produce larger amounts of TFIIF, a new strategy was introduced. Genes for each of the three subunits were subcloned in the separate Ylplac yeast-*E. coli* shuttle vectors carrying different *in vitro* mutagenized yeast genes (*LEU2*, *TRP1* and *URA3*, respectively) (Gietz and Sugino, 1988). Each subunit was placed under the control of the relatively strong alcohol dehydrogenase 1 (*ADH1*) promoter followed by a corresponding terminator (figure 17). To the middle subunit, Tfg2, a TAP tag was

added to the C-terminal end. Such plasmids were linearized and subsequently used for yeast transformation. Due to the homologous recombination, such linearized plasmids containing *ADH1* promoter and individual TFIIF subunits integrate into the yeast chromosomes.

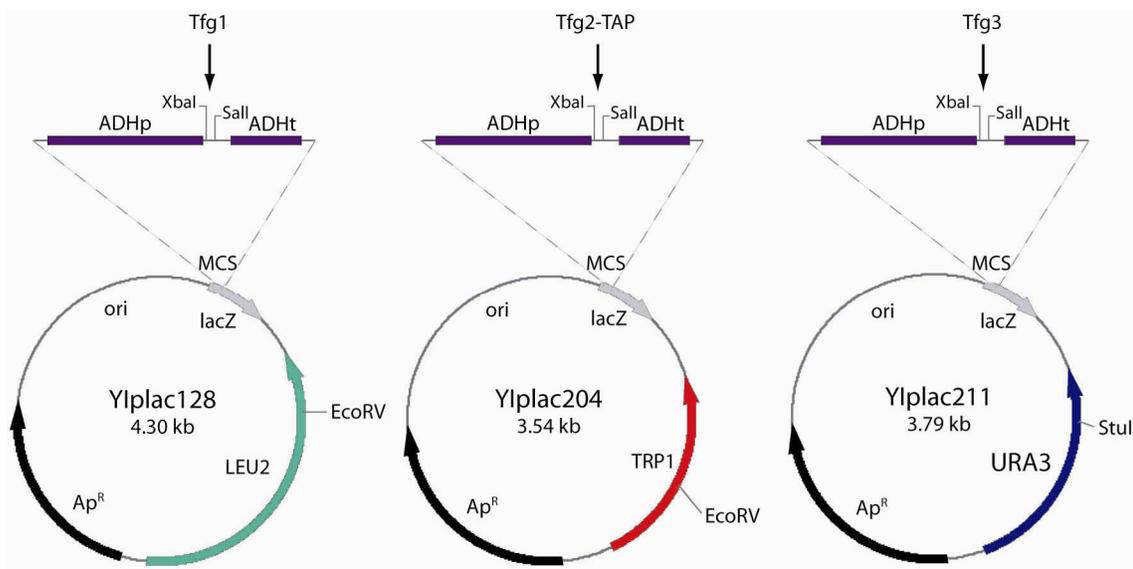


Figure 17: Generation of yeast TFIIF overexpression strain. Each of the TFIIF subunits was cloned into *E. coli*-yeast shuttle vector under the control of strong *ADH1* promoter as described in chapter 2.20. The plasmids were consequently linearized and used for yeast transformation (DSY5 strain). Due to the homologous recombination, plasmids were integrated into yeast genome.

In order to test the newly created strain for overexpression and compare it with the TFIIF yield resulting from expression under the control of the endogenous promoter, the protein complex from both strains was purified in an analogous manner. Relative comparison of the TFIIF yield by dot blot analysis (Borggreffe *et al.*, 2001) revealed substantially higher expression of TFIIF in the yeast strain where the protein expression was under the control of the *ADH1* promoter (Figure 18A, B). However, although the yield of TFIIF was raised, the amounts gained even from 16 liters of yeast culture, after all the purification and concentration steps, still did not exceed several micrograms and did not suffice for structural studies of the complex.

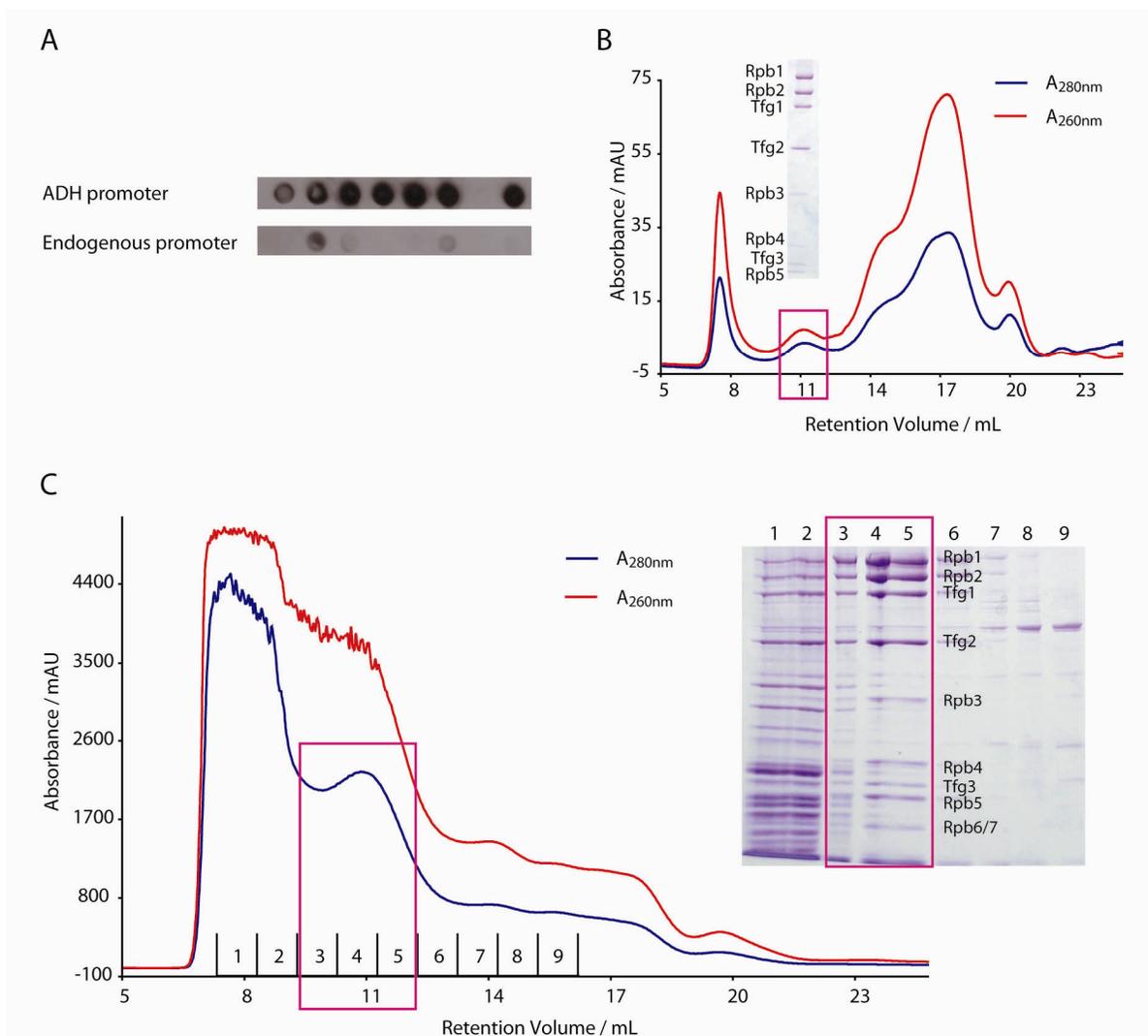


Figure 18. Overexpression of TFIIF in yeast

(A) Comparison of TFIIF-TAP yield, expressed under control of *ADH1* and endogenous promoter. Protein fractions were analyzed by dot blots (chapter 2.24). In order to enable relative comparison, proteins were purified identically from the same amount of yeast culture (2L of yeast culture harvested at OD₆₀₀~3).

(B) Small-scale TFIIF purification from overexpression strain. In this case 2L of yeast culture of OD₆₀₀~3 was used. After TEV cleavage, protein sample was loaded directly onto a gel filtration column (Superose 6).

(C) Large-scale TFIIF purification from overexpression strain. First gel filtration step in the presence of detergent (dodecyl- β -D-maltoside) corresponding to 100 L of yeast culture (OD₆₀₀~3) is shown. Elution profile was analyzed by SDS-PAGE. Fractions 3-5 contain stoichiometric Pol II-TFIIF complex.

In order to up-scale significantly, a substantial amount of the input yeast culture had to be used. Initially as much as 200 liters of yeast culture resulted only in 70 μg of crystallography-grade 15-subunit ~ 0.7 kDa stoichiometric Pol II-TFIIF complex (Figure 18C, corresponding to 100 liters). It seems that the binding of TFIIF on the Pol II surface prevents the dissociation of Rpb4/7 subcomplex as previously reported (Edwards *et al.*, 1990). This implies binding of TFIIF over Rpb4/7 surface which agrees with the low resolution architecture of Pol II-TFIIF complex by electron microscopy. By increasing the amount of IgG resin (30 mL per 1.1 kg of cell pellet), more Pol II-TFIIF complex could be isolated which allowed omitting the detergent in the final gel filtration step. The input yeast culture could be reduced down to 100 liters per a single purification and the yield was increased dramatically, reaching 1 to 1.5 mg of pure material, suitable for crystallographic studies and assembly of the PIC and early transcribing complexes (Figure 19, 200 micrograms loaded on the column).

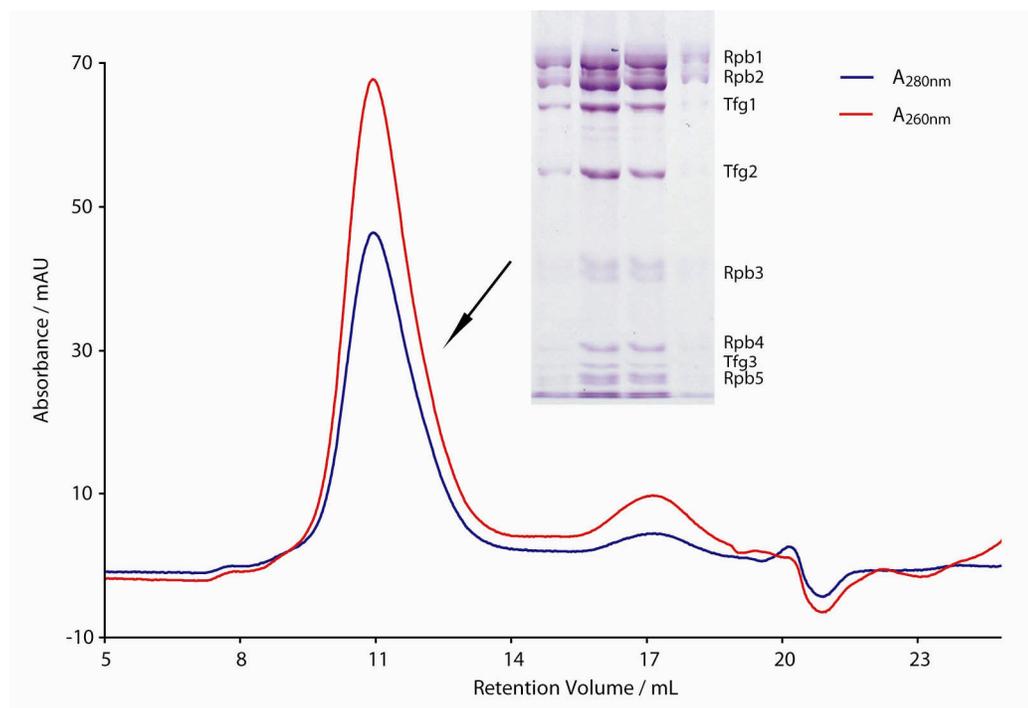


Figure 19: Optimized Pol II-TFIIF complex purification. Second gel filtration step without presence of detergents is shown. In this case 200 μg of Pol II-TFIIF complex after first gel filtration was loaded on the column. Peak fractions were analyzed by SDS-PAGE.

Crystallization trials of Pol II-TFIIF complex were performed and yielded crystals in the conditions characteristic for the growth of the Pol II crystals. Synchrotron diffraction data were collected to 4.2 Å resolution. After molecular replacement with the refined model of 12-subunit Pol II (Armache *et al.*, 2005), electron density maps did not reveal any additional density that could be attributed to TFIIF. Additionally, the crystals were washed, dissolved in SDS-loading dye and loaded on an SDS-gel. The individual bands were cut out, analyzed by MALDI peptide mass fingerprints and did not reveal any TFIIF subunits. Thus the Pol II-TFIIF complex had apparently dissociated during crystallization, and the free Pol II was crystallized. Possibly, TFIIF binds Pol II only by a few distinct globular domains connected through extensive linkers essential for biological function, and is only sufficiently stabilized by other general transcription factors of the transcription machinery.

3.2.4 Assembly of an initially transcribing complex (ITC)

With the Pol II-TFIIF complex being available for the first time in sufficient amounts, new routes towards structural studies of the preinitiation and early transcribing complexes are open. In cooperation with A. Jawhari from the laboratory, it was shown that the Pol II-TFIIF complex preparations can be used to assemble an ITC. An electrophoretic mobility shift assay was used to monitor the step-wise assembly of recombinantly expressed general transcription factors TBP and TFIIB and Pol II-TFIIF complex on the DNA promoter region in the presence of 11-nucleotide long RNA, mimicking an initially transcribed pre-mRNA (figure 20C). Both TFIIB and TBP separately bound promoter DNA, (Figure 20C, lanes 2 and 3). When mixed together, a TBP-TFIIB-DNA/RNA complex was formed (Figure 20C, lane 5). As expected, Pol II-TFIIF complex alone (Figure 20C, lane 4) as well as with TFIIB-TBP complex (lane 6) shifted dramatically the promoter DNA band. These two complexes, however, cannot be distinguished due to their size and consequently their poor resolution on the native gel used.

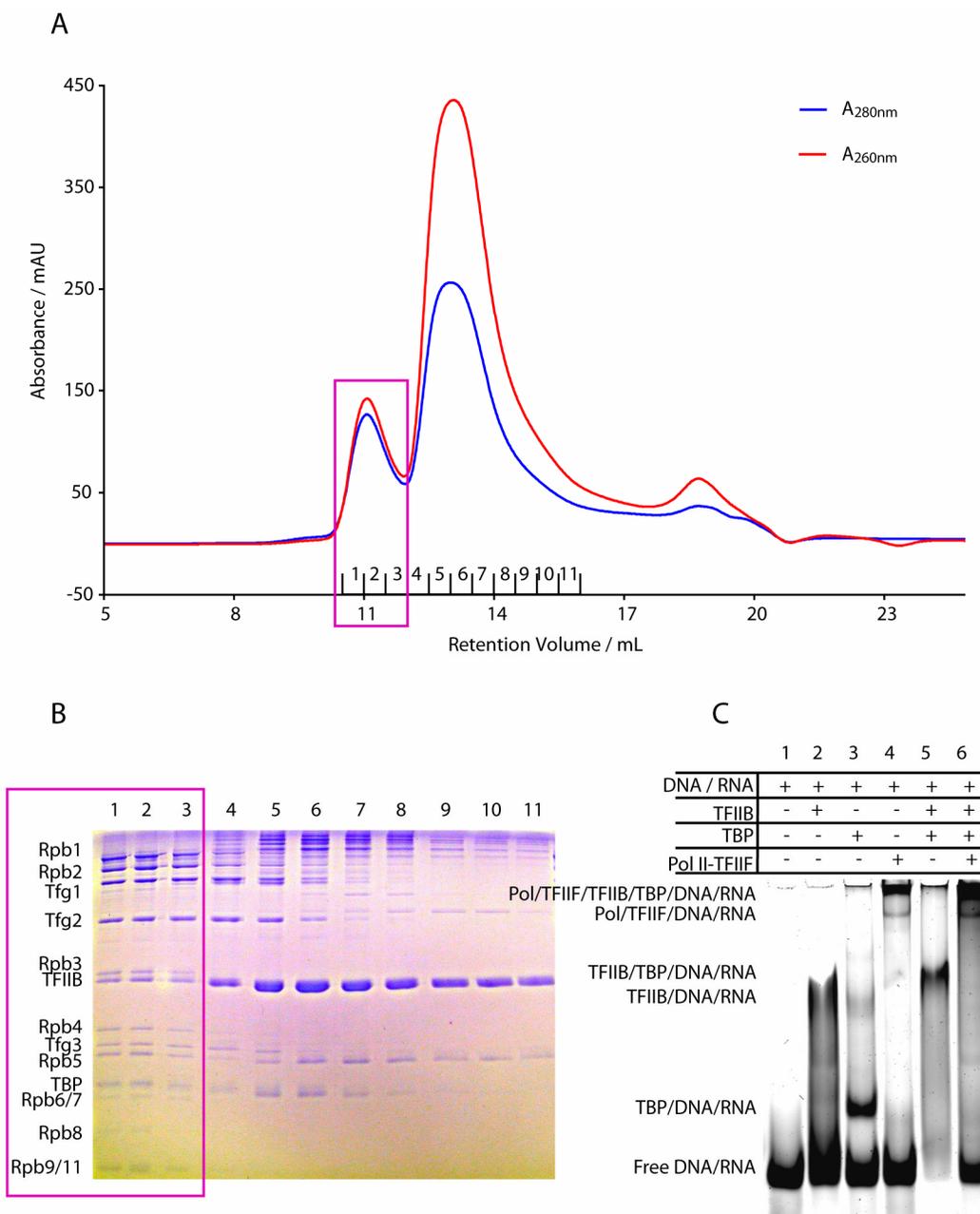


Figure 20: Assembly of the initially transcribing complex.

(A) Reconstitution of ITC complex. ITC was assembled *in vitro* as described in chapter 2.29 and separated on a Superose 6 gel filtration column.

(B) The two peak fractions were analyzed by SDS-PAGE. The first peak corresponds to a reconstituted ITC.

(C) Stepwise assembly of ITC monitored by electrophoretic mobility shift assay (EMSA). Lane 6 represents reconstituted ITC.

Additionally, ITC was analogously assembled *in vitro* and applied to a gel filtration column (Figure 20A). The first peak fraction contained all 17 polypeptides used for the assembly representing the yeast initially transcribing complex assembled *in vitro* that can be used for structural studies (Figure 20A, B). Preliminary crystallization trials resulted in small crystals, but their contents remain to be determined (A. Jawhari, personal communication).

To conclude, the large-scale preparation of the Pol II-TFIIF complex is a milestone towards structural studies of the yeast PIC and early transcribing complexes that will ensure new insight into the mechanism of eukaryotic gene expression.

3.2.5 Presence of 5.8S rRNA in Pol II-TFIIF preparations

The gel filtration profile of the Pol II-TFIIF complex showed that the absorption at 260 nm is higher than the absorption at 280 nm (Figures 18C, 19) which is indicative of nucleic acid binding. Such relation was kept even during fractional ammonium sulfate precipitation which includes a very high salt concentration. Selective digestion of the protein preparation with both RNase and DNase (RNase-free) independently, helped to identify the nucleic acid. While DNase I (RNase-free) digestion did not significantly influence the gel filtration profile, the RNase I digestion resulted in significantly reduced absorption at 260 nm at the retention volume of the Pol II-TFIIF complex (data not shown). This implies the presence of an RNA species bound to the complex. Upon phenol/chloroform extraction and ethanol precipitation, the nucleic acid sample was loaded on an 8% urea denaturing polyacrylamide gel. The gel showed a prominent band of ~160 nucleotides, resistant to DNase I cleavage, confirming it to be RNA. In order to establish the RNA identity, a modification of the RACE method (chapter 2.32, Figure 7) was employed to synthesize cDNA which was then further amplified by PCR and cloned into pCR4-TOPO vector for sequencing.

Sequencing of the vector surprisingly identified 5.8S rRNA to be present in the Pol II-TFIIF complex preparation. 5.8S rRNA is a component of the large ribosomal

subunit, which is synthesized together with 18S and 25S rRNA as a single polycistronic rRNA precursor by RNA Pol I (Kressler *et al.*, 1999; Venema and Tollervey, 1999; Fatica and Tollervey, 2002). In order to investigate the potential specificity of the interaction between Pol II-TFIIF complex and 5.8S rRNA, our collaborators (S. Ferreira-Cerca and H. Tschochner) immunoprecipitated the Pol II-TFIIF complex and searched for different RNA species by Northern blotting. However they did not observe any significant association of 5.8S rRNA when compared to the background under conditions tested. Hence, the observed interaction of 5.8S rRNA with the Pol II-TFIIF complex is apparently nonspecific and could result from the nucleic acid binding affinity of the Pol II-TFIIF complex and huge abundance of 5.8S rRNA in the cells. Additionally, interactions of 5.8S rRNA with 28S rRNA and ribosomal protein are sufficiently weak to permit a reversible dissociation of the 5.8S rRNA molecule even under relatively mild conditions (Nazar, 1978; Lo *et al.*, 1987). Finally, many ribosomal proteins are often co-purified during the immunoaffinity purification using TAP-tags.

Conclusions

In this thesis the high-resolution crystal structure of human Scp1 was reported. The structure of the CTD phosphatase Scp1 reveals a core fold similar to that of other enzymes of the DXDX(T/V) superfamily that are essentially unrelated in sequence, and is a good model for the catalytic FCPH domain of all Fcp1 and Scp1 enzymes. Biochemical and further structural studies of recombinant Scp1 and Fcp1 revealed the catalytic mechanism of these enzymes. The signature motif is part of a central depression that forms the active site and binds a metal ion. Catalysis involves the metal-assisted phosphorylation of the first aspartate in the DXDX(T/V) signature motif. Magnesium ions are essential for Fcp1 and Scp1 activity, and the trifluoroberyllate anion inhibits activity by forming a stable tetrahedral adduct with the catalytic aspartate side chain, mimicking a labile phosphoaspartyl intermediate. A conserved hydrophobic pocket formed between the active site and the Fcp1/Scp1-specific insertion domain is likely involved in CTD recognition. Specificity of Fcp1 for the Pol II CTD may additionally arise from multiple interactions with the Pol II machinery, including docking of Fcp1 to the Rpb4/7 subcomplex, which is located adjacent to a protein linker connecting to the CTD. Whereas the catalytic mechanism of Fcp1/Scp1 phosphatases is now well understood, the basis for their CTD specificity remains to be fully established. Open questions on CTD phosphatases and Pol II recycling include details of the interactions between the phosphatase, Pol II, and the CTD, and the timing of these transient protein-protein interactions during the transcription cycle. In addition, the question on communication between the CTD phosphatases, kinases and other CTD-interacting proteins has yet to be answered.

Whereas yeast TBP and other small general transcription factors can be produced in large amounts in bacteria, TFIIF was the last obstacle towards structural studies of the minimal preinitiation complex and early transcribing complexes. This thesis describes different attempts to overexpress and purify TFIIF from bacteria. None, however, resulted in a full-length 3-subunit stoichiometric complex. Pol II-TFIIF

complex from yeast could be isolated using a TAP purification in scarce amounts, not amenable for crystallographic studies. Consequently, an overexpression system in yeast has been developed by placing individual TFIIF subunits under the control of the *ADH1* promoter. Using up to 200 liters of yeast culture it was possible to establish a purification protocol that allowed for isolation of milligram quantities of pure stoichiometric Pol II-TFIIF complex. This complex was further shown to assemble, together with the DNA containing promoter and initial coding region, a short RNA stretch mimicking a transcribed pre-mRNA, TBP and TFIIB, into ITC. Only pieces of this puzzle have so far been visualized either by crystallography or NMR, or at very low resolution by electron microscopy. The exact interplay of those subcomplexes and concomitant conformational changes are still out of our reach. The results presented here should pave the way towards the higher resolution structure of a complete PIC and initially transcribing complexes in the future.

References

1. Albright, S.R. and R. Tjian. 2000. TAFs revisited: more data reveal new twists and confirm old ideas. *Gene* **242**: 1-13.
2. Allen, M., A. Friedler, O. Schon, and M. Bycroft. 2002. The structure of an FF domain from human HYPA/FBP11. *J Mol Biol* **323**: 411-6.
3. Andel, F., 3rd, A.G. Ladurner, C. Inouye, R. Tjian, and E. Nogales. 1999. Three-dimensional structure of the human TFIID-IIA-IIB complex. *Science* **286**: 2153-6.
4. Andersen, G., D. Busso, A. Poterszman, J.R. Hwang, J.M. Wurtz, R. Ripp, J.C. Thierry, J.M. Egly, and D. Moras. 1997. The structure of cyclin H: common mode of kinase activation and specific features. *Embo J* **16**: 958-67.
5. Archambault, J., R.S. Chambers, M.S. Kobor, Y. Ho, M. Cartier, D. Bolotin, B. Andrews, C.M. Kane, and J. Greenblatt. 1997. An essential component of a C-terminal domain phosphatase that interacts with transcription factor IIF in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* **94**: 14300-5.
6. Archambault, J., G. Pan, G.K. Dahmus, M. Cartier, N. Marshall, S. Zhang, M.E. Dahmus, and J. Greenblatt. 1998. FCP1, the RAP74-interacting subunit of a human protein phosphatase that dephosphorylates the carboxyl-terminal domain of RNA polymerase II. *J Biol Chem* **273**: 27593-601.
7. Armache, K.J., H. Kettenberger, and P. Cramer. 2003. Architecture of initiation-competent 12-subunit RNA polymerase II. *Proc Natl Acad Sci U S A* **100**: 6964-8.
8. Armache, K.J., S. Mitterweger, A. Meinhart, and P. Cramer. 2005. Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J Biol Chem* **280**: 7131-4.
9. Barford, D., Z. Jia, and N.K. Tonks. 1995. Protein tyrosine phosphatases take off. *Nat Struct Biol* **2**: 1043-53.
10. Barton, G.J. 1993. ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng* **6**: 37-40.

11. Bell, S.D., P.L. Kosa, P.B. Sigler, and S.P. Jackson. 1999. Orientation of the transcription preinitiation complex in archaea. *Proc Natl Acad Sci U S A* **96**: 13662-7.
12. Bengal, E., O. Flores, A. Krauskopf, D. Reinberg, and Y. Aloni. 1991. Role of the mammalian transcription factors IIF, IIS, and IIX during elongation by RNA polymerase II. *Mol Cell Biol* **11**: 1195-206.
13. Bentley, D. 2002. The mRNA assembly line: transcription and processing machines in the same factory. *Curr Opin Cell Biol* **14**: 336-42.
14. Bienkiewicz, E.A., A. Moon Woody, and R.W. Woody. 2000. Conformation of the RNA polymerase II C-terminal domain: circular dichroism of long and short fragments. *J Mol Biol* **297**: 119-33.
15. Bleichenbacher, M., S. Tan, and T.J. Richmond. 2003. Novel interactions between the components of human and yeast TFIIA/TBP/DNA complexes. *J Mol Biol* **332**: 783-93.
16. Bloom H., H. Beier, H.J. Gross. 1987 *Electrophoresis* 8:93-9
17. Borggreffe, T., R. Davis, A. Bareket-Samish, and R.D. Kornberg. 2001. Quantitation of the RNA polymerase II transcription machinery in yeast. *J Biol Chem* **276**: 47150-3.
18. Borggreffe, T., R. Davis, H. Erdjument-Bromage, P. Tempst, and R.D. Kornberg. 2002. A complex of the Srb8, -9, -10, and -11 transcriptional regulatory proteins from yeast. *J Biol Chem* **277**: 44202-7.
19. Boube, M., L. Joulia, D.L. Cribbs, and H.M. Bourbon. 2002. Evidence for a mediator of RNA polymerase II transcriptional regulation conserved from yeast to man. *Cell* **110**: 143-51.
20. Brand, M., C. Leurent, V. Mallouh, L. Tora, and P. Schultz. 1999. Three-dimensional structures of the TAFII-containing complexes TFIID and TFTC. *Science* **286**: 2151-3.
21. Bregman, D.B., R.G. Pestell, and V.J. Kidd. 2000. Cell cycle regulation and RNA polymerase II. *Front Biosci* **5**: D244-57.
22. Brunger, A.T., P.D. Adams, G.M. Clore, W.L. DeLano, P. Gros, R.W. Grosse-Kunstleve, J.S. Jiang, J. Kuszewski, M. Nilges, N.S. Pannu, R.J. Read, L.M.

- Rice, T. Simonson, and G.L. Warren. 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**: 905-21.
23. Budisa, N., B. Steipe, P. Demange, C. Eckerskorn, J. Kellermann, and R. Huber. 1995. High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in *Escherichia coli*. *Eur J Biochem* **230**: 788-96.
24. Buratowski, S. 2003. The CTD code. *Nat Struct Biol* **10**: 679-80.
25. Burke, T.R., and Z.Y. Zhang. 1998. Protein-tyrosine phosphatases: structure, mechanism, and inhibitor discovery. *Biopolymers* **47**: 225-41.
26. Burke, T.W. and J.T. Kadonaga. 1997. The downstream core promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by TAFII60 of *Drosophila*. *Genes Dev* **11**: 3020-31.
27. Burton, Z.F., M. Killeen, M. Sopta, L.G. Ortolan, and J. Greenblatt. 1988. RAP30/74: a general initiation factor that binds to RNA polymerase II. *Mol Cell Biol* **8**: 1602-13.
28. Bushnell, D.A. and R.D. Kornberg. 2003. Complete, 12-subunit RNA polymerase II at 4.1-A resolution: implications for the initiation of transcription. *Proc Natl Acad Sci U S A* **100**: 6969-73.
29. Bushnell, D.A., K.D. Westover, R.E. Davis, and R.D. Kornberg. 2004. Structural basis of transcription: an RNA polymerase II-TFIIB cocystal at 4.5 Angstroms. *Science* **303**: 983-8.
30. Cagas, P.M. and J.L. Corden. 1995. Structural studies of a synthetic peptide derived from the carboxyl-terminal domain of RNA polymerase II. *Proteins* **21**: 149-60.
31. Cairns, B.R., N.L. Henry, and R.D. Kornberg. 1996. TFG/TAF30/ANC1, a component of the yeast SWI/SNF complex that is similar to the leukemogenic proteins ENL and AF-9. *Mol Cell Biol* **16**: 3308-16.
32. Chabre, M. 1990. Aluminofluoride and beryllifluoride complexes: a new phosphate analogs in enzymology. *Trends Biochem Sci* **15**: 6-10.
33. Chambers, R.S. and M.E. Dahmus. 1994. Purification and characterization of

- a phosphatase from HeLa cells which dephosphorylates the C-terminal domain of RNA polymerase II. *J Biol Chem* **269**: 26243-8.
34. Chambers, R.S. and C.M. Kane. 1996. Purification and characterization of an RNA polymerase II phosphatase from yeast. *J Biol Chem* **271**: 24498-504.
 35. Chambers, R.S., B.Q. Wang, Z.F. Burton, and M.E. Dahmus. 1995. The activity of COOH-terminal domain phosphatase is regulated by a docking site on RNA polymerase II and by the general transcription factors IIF and IIB. *J Biol Chem* **270**: 14962-9.
 36. Chasman, D.I., K.M. Flaherty, P.A. Sharp, and R.D. Kornberg. 1993. Crystal structure of yeast TATA-binding protein and model for interaction with DNA. *Proc Natl Acad Sci U S A* **90**: 8174-8.
 37. Chen, H.T. and S. Hahn. 2003. Binding of TFIIB to RNA polymerase II: Mapping the binding site for the TFIIB zinc ribbon domain within the preinitiation complex. *Mol Cell* **12**: 437-47.
 38. Chi, T., P. Lieberman, K. Ellwood, and M. Carey. 1995. A general mechanism for transcriptional synergy by eukaryotic activators. *Nature* **377**: 254-7.
 39. Cho, E.J., M.S. Kobor, M. Kim, J. Greenblatt, and S. Buratowski. 2001a. Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. *Genes Dev* **15**: 3319-29.
 40. Cho, E.J., T. Takagi, C.R. Moore, and S. Buratowski. 1997. mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev* **11**: 3319-26.
 41. Cho, H., T.K. Kim, H. Mancebo, W.S. Lane, O. Flores, and D. Reinberg. 1999. A protein phosphatase functions to recycle RNA polymerase II. *Genes Dev* **13**: 1540-52.
 42. Cho, H., W. Wang, R. Kim, H. Yokota, S. Damo, S.H. Kim, D. Wemmer, S. Kustu, and D. Yan. 2001b. BeF(3)(-) acts as a phosphate analog in proteins phosphorylated on aspartate: structure of a BeF(3)(-) complex with phosphoserine phosphatase. *Proc Natl Acad Sci U S A* **98**: 8525-30.
 43. Chung, W.H., J.L. Craighead, W.H. Chang, C. Ezeokonkwo, A. Bareket-Samish, R.D. Kornberg, and F.J. Asturias. 2003. RNA polymerase II/TFIIF

- structure and conserved organization of the initiation complex. *Mol Cell* **12**: 1003-13.
44. Coin, F. and J.M. Egly. 1998. Ten years of TFIIH. *Cold Spring Harb Symp Quant Biol* **63**: 105-10.
 45. Collet, J.F., V. Stroobant, M. Pirard, G. Delpierre, and E. Van Schaftingen. 1998. A new class of phosphotransferases phosphorylated on an aspartate residue in an amino-terminal DXDX(T/V) motif. *J Biol Chem* **273**: 14107-12.
 46. Conaway, J.W. and R.C. Conaway. 1990. An RNA polymerase II transcription factor shares functional properties with Escherichia coli sigma 70. *Science* **248**: 1550-3.
 47. Conaway, J.W., D. Reines, and R.C. Conaway. 1990. Transcription initiated by RNA polymerase II and purified transcription factors from liver. Cooperative action of transcription factors tau and epsilon in initial complex formation. *J Biol Chem* **265**: 7552-8.
 48. Conaway, R.C., K.P. Garrett, J.P. Hanley, and J.W. Conaway. 1991. Mechanism of promoter selection by RNA polymerase II: mammalian transcription factors alpha and beta gamma promote entry of polymerase into the preinitiation complex. *Proc Natl Acad Sci U S A* **88**: 6205-9.
 49. Corden, J.L., D.L. Cadena, J.M. Ahearn, Jr., and M.E. Dahmus. 1985. A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. *Proc Natl Acad Sci U S A* **82**: 7934-8.
 50. Cosma, M.P. 2002. Ordered recruitment: gene-specific mechanism of transcription activation. *Mol Cell* **10**: 227-36.
 51. Coulombe, B. and Z.F. Burton. 1999. DNA bending and wrapping around RNA polymerase: a "revolutionary" model describing transcriptional mechanisms. *Microbiol Mol Biol Rev* **63**: 457-78.
 52. Cramer, P., D.A. Bushnell, and R.D. Kornberg. 2001. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**: 1863-76.
 53. Davis, J.A., Y. Takagi, R.D. Kornberg, and F.A. Asturias. 2002. Structure of the yeast RNA polymerase II holoenzyme: Mediator conformation and

- polymerase interaction. *Mol Cell* **10**: 409-15.
54. Dichtl, B., D. Blank, M. Ohnacker, A. Friedlein, D. Roeder, H. Langen, and W. Keller. 2002. A role for SSU72 in balancing RNA polymerase II transcription elongation and termination. *Mol Cell* **10**: 1139-50.
 55. Dikstein, R., S. Ruppert, and R. Tjian. 1996. TAFII250 is a bipartite protein kinase that phosphorylates the basal transcription factor RAP74. *Cell* **84**: 781-90.
 56. Dynlacht, B.D. 1997. Regulation of transcription by proteins that control the cell cycle. *Nature* **389**: 149-52.
 57. Ebright, R.H. 2000. RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *J Mol Biol* **304**: 687-98.
 58. Edwards, A. M., Darst, S. A., Feaver, W. J., Thompson, N. E., Burgess, R. R., and Kornberg, R. D. 1990. Purification and lipid-layer crystallization of yeast RNA polymerase II. *Proc Natl Acad Sci U S A* **87**: 2122-6.
 59. Elmendorf, B.J., A. Shilatifard, Q. Yan, J.W. Conaway, and R.C. Conaway. 2001. Transcription factors TFIIF, ELL, and Elongin negatively regulate SII-induced nascent transcript cleavage by non-arrested RNA polymerase II elongation intermediates. *J Biol Chem* **276**: 23109-14.
 60. Fabrega, C., V. Shen, S. Shuman, and C.D. Lima. 2003. Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Mol Cell* **11**: 1549-61.
 61. Fang, S.M. and Z.F. Burton. 1996. RNA polymerase II-associated protein (RAP) 74 binds transcription factor (TF) IIB and blocks TFIIB-RAP30 binding. *J Biol Chem* **271**: 11703-9.
 62. Fatica, A. and D. Tollervey. 2002. Making ribosomes. *Curr Opin Cell Biol* **14**: 313-8.
 63. Flores, O., H. Lu, M. Killeen, J. Greenblatt, Z.F. Burton, and D. Reinberg. 1991. The small subunit of transcription factor IIF recruits RNA polymerase II into the preinitiation complex. *Proc Natl Acad Sci U S A* **88**: 9999-10003.
 64. Flores, O., H. Lu, and D. Reinberg. 1992. Factors involved in specific transcription by mammalian RNA polymerase II. Identification and

- characterization of factor IIH. *J Biol Chem* **267**: 2786-93.
65. Flores, O., E. Maldonado, and D. Reinberg. 1989. Factors involved in specific transcription by mammalian RNA polymerase II. Factors IIE and IIF independently interact with RNA polymerase II. *J Biol Chem* **264**: 8913-21.
66. Forget, D., M.F. Langelier, C. Therien, V. Trinh, and B. Coulombe. 2004. Photo-cross-linking of a purified preinitiation complex reveals central roles for the RNA polymerase II mobile clamp and TFIIE in initiation mechanisms. *Mol Cell Biol* **24**: 1122-31.
67. Freiman, R.N., S.R. Albright, S. Zheng, W.C. Sha, R.E. Hammer, and R. Tjian. 2001. Requirement of tissue-selective TBP-associated factor TAFII105 in ovarian development. *Science* **293**: 2084-7.
68. Fu, J., Gnatt, A. L., Bushnell, D. A., Jensen, G. J., Thompson, N. E., Burgess, R. R., David, P. R., and Kornberg, R. D. 1999. Yeast RNA polymerase II at 5 Å resolution. *Cell* **98**: 799-810.
69. Gaiser, F., S. Tan, and T.J. Richmond. 2000. Novel dimerization fold of RAP30/RAP74 in human TFIIF at 1.7 Å resolution. *J Mol Biol* **302**: 1119-27.
70. Ganem, C., F. Devaux, C. Torchet, C. Jacq, S. Quevillon-Cheruel, G. Labesse, C. Facca, and G. Faye. 2003. Ssu72 is a phosphatase essential for transcription termination of snoRNAs and specific mRNAs in yeast. *Embo J* **22**: 1588-98.
71. Gangloff, Y.G., C. Romier, S. Thuault, S. Werten, and I. Davidson. 2001. The histone fold is a key structural motif of transcription factor TFIID. *Trends Biochem Sci* **26**: 250-7.
72. Garrett, K.P., H. Serizawa, J.P. Hanley, J.N. Bradsher, A. Tsuboi, N. Arai, T. Yokota, K. Arai, R.C. Conaway, and J.W. Conaway. 1992. The carboxyl terminus of RAP30 is similar in sequence to region 4 of bacterial sigma factors and is required for function. *J Biol Chem* **267**: 23942-9.
73. Gavin, A.C., M. Bosche, R. Krause, P. Grandi, *et al.* 2002. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**: 141-7.
74. Geiger, J.H., S. Hahn, S. Lee, and P.B. Sigler. 1996. Crystal structure of the

- yeast TFIIA/TBP/DNA complex. *Science* **272**: 830-6.
75. Gerber, H.P., M. Hagmann, K. Seipel, O. Georgiev, M.A. West, Y. Litingtung, W. Schaffner, and J.L. Corden. 1995. RNA polymerase II C-terminal domain required for enhancer-driven transcription. *Nature* **374**: 660-2.
76. Ghazy, M.A., S.A. Brodie, M.L. Ammerman, L.M. Ziegler, and A.S. Ponticelli. 2004. Amino acid substitutions in yeast TFIIIF confer upstream shifts in transcription initiation and altered interaction with RNA polymerase II. *Mol Cell Biol* **24**: 10975-85.
77. Gietz, R.D. and A. Sugino. 1988. New yeast-Escherichia coli shuttle vectors constructed with *in vitro* mutagenized yeast genes lacking six-base pair restriction sites. *Gene* **74**: 527-34.
78. Gnatt, A.L., P. Cramer, J. Fu, D.A. Bushnell, and R.D. Kornberg. 2001. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* **292**: 1876-82.
79. Green, M.R. 2000. TBP-associated factors (TAFIIIs): multiple, selective transcriptional mediators in common complexes. *Trends Biochem Sci* **25**: 59-63.
80. Groft, C.M., S.N. Uljon, R. Wang, and M.H. Werner. 1998. Structural homology between the Rap30 DNA-binding domain and linker histone H5: implications for preinitiation complex assembly. *Proc Natl Acad Sci U S A* **95**: 9117-22.
81. Gross, C.A., C. Chan, A. Dombroski, T. Gruber, M. Sharp, J. Tupy, and B. Young. 1998. The functional and regulatory roles of sigma factors in transcription. *Cold Spring Harb Symp Quant Biol* **63**: 141-55.
82. Guo, Z. and J.W. Stiller. 2004. Comparative genomics of cyclin-dependent kinases suggest co-evolution of the RNAP II C-terminal domain and CTD-directed CDKs. *BMC Genomics* **5**: 69.
83. Hahn, S. 2004. Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol* **11**: 394-403.
84. Hahn, S. and S. Roberts. 2000. The zinc ribbon domains of the general transcription factors TFIIIB and Brf: conserved functional surfaces but different

- roles in transcription initiation. *Genes Dev* **14**: 719-30.
85. Harding, M.M. 1992. NMR studies on YSPTSPSY: implications for the design of DNA bisintercalators. *J Med Chem* **35**: 4658-64.
86. Hausmann, S., H. Erdjument-Bromage, and S. Shuman. 2004. Schizosaccharomyces pombe carboxyl-terminal domain (CTD) phosphatase Fcp1: distributive mechanism, minimal CTD substrate, and active site mapping. *J Biol Chem* **279**: 10892-900.
87. Hausmann, S., B. Schwer, and S. Shuman. 2004. An *encephalitozoon cuniculi* ortholog of the RNA polymerase II carboxyl-terminal domain (CTD) serine phosphatase Fcp1. *Biochemistry* **43**: 7111-20.
88. Hausmann, S. and S. Shuman. 2002. Characterization of the CTD phosphatase Fcp1 from fission yeast. Preferential dephosphorylation of serine 2 versus serine 5. *J Biol Chem* **277**: 21213-20.
89. Hausmann, S. and S. Shuman. 2003. Defining the active site of Schizosaccharomyces pombe C-terminal domain phosphatase Fcp1. *J Biol Chem* **278**: 13627-32.
90. He, X., A.U. Khan, H. Cheng, D.L. Pappas, Jr., M. Hampsey, and C.L. Moore. 2003. Functional interactions between the transcription and mRNA 3' end processing machineries mediated by Ssu72 and Sub1. *Genes Dev* **17**: 1030-42.
91. Hengartner, C.J., V.E. Myer, S.M. Liao, C.J. Wilson, S.S. Koh, and R.A. Young. 1998. Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases. *Mol Cell* **2**: 43-53.
92. Henry, N.L., A.M. Campbell, W.J. Feaver, D. Poon, P.A. Weil, and R.D. Kornberg. 1994. TFIIF-TAF-RNA polymerase II connection. *Genes Dev* **8**: 2868-78.
93. Hirose, Y. and J.L. Manley. 2000. RNA polymerase II and the integration of nuclear events. *Genes Dev* **14**: 1415-29.
94. Ho, C.K., V. Sriskanda, S. McCracken, D. Bentley, B. Schwer, and S. Shuman. 1998. The guanylyltransferase domain of mammalian mRNA capping enzyme binds to the phosphorylated carboxyl-terminal domain of

- RNA polymerase II. *J Biol Chem* **273**: 9577-85.
95. Hochheimer, A. and R. Tjian. 2003. Diversified transcription initiation complexes expand promoter selectivity and tissue-specific gene expression. *Genes Dev* **17**: 1309-20.
 96. Hoepfner, S., S. Baumli, and P. Cramer. 2005. Structure of the mediator subunit cyclin C and its implications for CDK8 function. *J Mol Biol* **350**: 833-42.
 97. Holm, L. and C. Sander. 1995. Dali: a network tool for protein structure comparison. *Trends Biochem Sci* **20**: 478-80.
 98. Holstege, F.C., U. Fiedler, and H.T. Timmers. 1997. Three transitions in the RNA polymerase II transcription complex during initiation. *Embo J* **16**: 7468-80.
 99. Holstege, F.C., E.G. Jennings, J.J. Wyrick, T.I. Lee, C.J. Hengartner, M.R. Green, T.R. Golub, E.S. Lander, and R.A. Young. 1998. Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**: 717-28.
 100. Izban, M.G. and D.S. Luse. 1992. Factor-stimulated RNA polymerase II transcribes at physiological elongation rates on naked DNA but very poorly on chromatin templates. *J Biol Chem* **267**: 13647-55.
 101. Ito, H., Y. Fukuda, K. Murata, and A. Kimura. 1983. Transformation of intact yeast cells treated with alkali cations. *J Bacteriol* **153**: 163-168.
 102. Jawhari, A., M. Uhring, S. De Carlo, C. Crucifix, G. Tocchini-Valentini, D. Moras, P. Schultz, and A. Poterszman. 2006. Structure and oligomeric state of human transcription factor TFIIE. *EMBO Rep*.
 103. John, S., L. Howe, S.T. Tafrov, P.A. Grant, R. Sternglanz, and J.L. Workman. 2000. The something about silencing protein, Sas3, is the catalytic subunit of NuA3, a yTAF(II)30-containing HAT complex that interacts with the Spt16 subunit of the yeast CP (Cdc68/Pob3)-FACT complex. *Genes Dev* **14**: 1196-208.
 104. Kabani, M., K. Michot, C. Boschiero, and M. Werner. 2005. Anc1 interacts with the catalytic subunits of the general transcription factors TFIID and TFIIIF, the chromatin remodeling complexes RSC and INO80, and the histone acetyltransferase complex NuA3. *Biochem Biophys Res Commun* **332**: 398-

- 403.
105. Kabsch, W. 1993. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Crystallogr* **26**: 795–800.
 106. Kamada, K., J. De Angelis, R.G. Roeder, and S.K. Burley. 2001. Crystal structure of the C-terminal domain of the RAP74 subunit of human transcription factor IIF. *Proc Natl Acad Sci U S A* **98**: 3115-20.
 107. Kamada, K., R.G. Roeder, and S.K. Burley. 2003. Molecular mechanism of recruitment of TFIIF- associating RNA polymerase C-terminal domain phosphatase (FCP1) by transcription factor IIF. *Proc Natl Acad Sci U S A* **100**: 2296-9.
 108. Keogh, M.C., V. Podolny, and S. Buratowski. 2003. Bur1 kinase is required for efficient transcription elongation by RNA polymerase II. *Mol Cell Biol* **23**: 7005-18.
 109. Kettenberger, H., K.J. Armache, and P. Cramer. 2003. Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell* **114**: 347-57.
 110. Killeen, M.T. and J.F. Greenblatt. 1992. The general transcription factor RAP30 binds to RNA polymerase II and prevents it from binding nonspecifically to DNA. *Mol Cell Biol* **12**: 30-7.
 111. Kim, J.L., D.B. Nikolov, and S.K. Burley. 1993. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365**: 520-7.
 112. Kim, M., N.J. Krogan, L. Vasiljeva, O.J. Rando, E. Nedeá, J.F. Greenblatt, and S. Buratowski. 2004. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* **432**: 517-22.
 113. Kim, Y., J.H. Geiger, S. Hahn, and P.B. Sigler. 1993. Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**: 512-20.
 114. Kim, Y.J., S. Bjorklund, Y. Li, M.H. Sayre, and R.D. Kornberg. 1994. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77**: 599-608.
 115. Kimura, M., H. Suzuki, and A. Ishihama. 2002. Formation of a carboxy-

- terminal domain phosphatase (Fcp1)/TFIIF/RNA polymerase II (pol II) complex in *Schizosaccharomyces pombe* involves direct interaction between Fcp1 and the Rpb4 subunit of pol II. *Mol Cell Biol* **22**: 1577-88.
116. Kitajima, S., T. Chibazakura, M. Yonaha, and Y. Yasukochi. 1994. Regulation of the human general transcription initiation factor TFIIF by phosphorylation. *J Biol Chem* **269**: 29970-7.
117. Kobor, M.S., J. Archambault, W. Lester, F.C. Holstege, O. Gileadi, D.B. Jansma, E.G. Jennings, F. Kouyoumdjian, A.R. Davidson, R.A. Young, and J. Greenblatt. 1999. An unusual eukaryotic protein phosphatase required for transcription by RNA polymerase II and CTD dephosphorylation in *S. cerevisiae*. *Mol Cell* **4**: 55-62.
118. Kobor, M.S., L.D. Simon, J. Omichinski, G. Zhong, J. Archambault, and J. Greenblatt. 2000. A motif shared by TFIIF and TFIIB mediates their interaction with the RNA polymerase II carboxy-terminal domain phosphatase Fcp1p in *Saccharomyces cerevisiae*. *Mol Cell Biol* **20**: 7438-49.
119. Koiwa, H., S. Hausmann, W.Y. Bang, A. Ueda, N. Kondo, A. Hiraguri, T. Fukuhara, J.D. Bahk, D.J. Yun, R.A. Bressan, P.M. Hasegawa, and S. Shuman. 2004. Arabidopsis C-terminal domain phosphatase-like 1 and 2 are essential Ser-5-specific C-terminal domain phosphatases. *Proc Natl Acad Sci U S A* **101**: 14539-44.
120. Kokubo, T., M.J. Swanson, J.I. Nishikawa, A.G. Hinnebusch, and Y. Nakatani. 1998. The yeast TAF145 inhibitory domain and TFIIA competitively bind to TATA-binding protein. *Mol Cell Biol* **18**: 1003-12.
121. Komarnitsky, P., E.J. Cho, and S. Buratowski. 2000. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev* **14**: 2452-60.
122. Kong, S.E., M.S. Kobor, N.J. Krogan, B.P. Somesh, T.M. Sogaard, J.F. Greenblatt, and J.Q. Svejstrup. 2005. Interaction of Fcp1 Phosphatase with Elongating RNA Polymerase II Holoenzyme, Enzymatic Mechanism of Action, and Genetic Interaction with Elongator. *J Biol Chem* **280**: 4299-306.
123. Kressler, D., P. Linder, and J. de La Cruz. 1999. Protein trans-acting factors

- involved in ribosome biogenesis in *Saccharomyces cerevisiae*. *Mol Cell Biol* **19**: 7897-912.
124. Krishnamurthy, S., X. He, M. Reyes-Reyes, C. Moore, and M. Hampsey. 2004. Ssu72 is an RNA polymerase II CTD phosphatase. *Mol Cell* **14**: 387-94.
125. Krogan, N.J., M. Kim, S.H. Ahn, G. Zhong, M.S. Kobor, G. Cagney, A. Emili, A. Shilatifard, S. Buratowski, and J.F. Greenblatt. 2002. RNA polymerase II elongation factors of *Saccharomyces cerevisiae*: a targeted proteomics approach. *Mol Cell Biol* **22**: 6979-92.
126. Kuras, L., P. Kosa, M. Mencia, and K. Struhl. 2000. TAF-Containing and TAF-independent forms of transcriptionally active TBP *in vivo*. *Science* **288**: 1244-8.
127. La Fortelle, E.D., and G. Bricogne. 1997. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**: 472-94.
128. Lahiri, S.D., G. Zhang, D. Dunaway-Mariano, and K.N. Allen. 2003. The pentacovalent phosphorus intermediate of a phosphoryl transfer reaction. *Science* **299**: 2067-71.
129. Langelier, M.F., D. Forget, A. Rojas, Y. Porlier, Z.F. Burton, and B. Coulombe. 2001. Structural and functional interactions of transcription factor (TF) IIA with TFIIE and TFIIIF in transcription initiation by RNA polymerase II. *J Biol Chem* **276**: 38652-7.
130. Laybourn, P.J. and M.E. Dahmus. 1989. Transcription-dependent structural changes in the C-terminal domain of mammalian RNA polymerase subunit Ila/o. *J Biol Chem* **264**: 6693-8.
131. Le Masson, I., D.Y. Yu, K. Jensen, A. Chevalier, R. Courbeyrette, Y. Boulard, M.M. Smith, and C. Mann. 2003. Yaf9, a novel NuA4 histone acetyltransferase subunit, is required for the cellular response to spindle stress in yeast. *Mol Cell Biol* **23**: 6086-102.
132. Lehman, A.L. and M.E. Dahmus. 2000. The sensitivity of RNA polymerase II in elongation complexes to C-terminal domain phosphatase. *J Biol Chem* **275**:

- 14923-32.
133. Li, X.Y., S.R. Bhaumik, and M.R. Green. 2000. Distinct classes of yeast promoters revealed by differential TAF recruitment. *Science* **288**: 1242-4.
134. Lin, P.S., M.F. Dubois, and M.E. Dahmus. 2002. TFIIIF-associating carboxyl-terminal domain phosphatase dephosphorylates phosphoserines 2 and 5 of RNA polymerase II. *J Biol Chem* **277**: 45949-56.
135. Lindstrom, D.L., S.L. Squazzo, N. Muster, T.A. Burckin, K.C. Wachter, C.A. Emigh, J.A. McCleery, J.R. Yates, 3rd, and G.A. Hartzog. 2003. Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins. *Mol Cell Biol* **23**: 1368-78.
136. Littlefield, O., Y. Korkhin, and P.B. Sigler. 1999. The structural basis for the oriented assembly of a TBP/TFB/promoter complex. *Proc Natl Acad Sci U S A* **96**: 13668-73.
137. Liu, D., R. Ishima, K.I. Tong, S. Bagby, T. Kokubo, D.R. Muhandiram, L.E. Kay, Y. Nakatani, and M. Ikura. 1998. Solution structure of a TBP-TAF(II)230 complex: protein mimicry of the minor groove surface of the TATA box unwound by TBP. *Cell* **94**: 573-83.
138. Liu, Y., J.A. Ranish, R. Aebersold, and S. Hahn. 2001. Yeast nuclear extract contains two major forms of RNA polymerase II mediator complexes. *J Biol Chem* **276**: 7169-75.
139. Lo, A.C., W.Y. Liu, D.E. Culham, and R.N. Nazar. 1987. Effects of ribosome dissociation on the structure of the ribosome-associated 5.8S RNA. *Biochem Cell Biol* **65**: 536-42.
140. Lolli, G., E.D. Lowe, N.R. Brown, and L.N. Johnson. 2004. The crystal structure of human CDK7 and its protein recognition properties. *Structure* **12**: 2067-79.
141. Luo, W., A.W. Johnson and D.L. Bentley. 2006. The role of rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a united allosteric-torpedo model. *Genes Dev* **20**: 954-965.
142. Luse, D.S. and G.A. Jacob. 1987. Abortive initiation by RNA polymerase II *in vitro* at the adenovirus 2 major late promoter. *J Biol Chem* **262**: 14990-7.

143. Marshall, N.F., J. Peng, Z. Xie, and D.H. Price. 1996. Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *J Biol Chem* **271**: 27176-83.
144. Marshall, N.F. and D.H. Price. 1995. Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *J Biol Chem* **270**: 12335-8.
145. Matsushima, N., C.E. Creutz, and R.H. Kretsinger. 1990. Polyproline, beta-turn helices. Novel secondary structures proposed for the tandem repeats within rhodopsin, synaptophysin, synexin, gliadin, RNA polymerase II, hordein, and gluten. *Proteins* **7**: 125-55.
146. Maxon, M.E., J.A. Goodrich, and R. Tjian. 1994. Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance. *Genes Dev* **8**: 515-24.
147. McCracken, S., N. Fong, E. Rosonina, K. Yankulov, G. Brothers, D. Siderovski, A. Hessel, S. Foster, S. Shuman, and D.L. Bentley. 1997b. 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes Dev* **11**: 3306-18.
148. McCracken, S., N. Fong, K. Yankulov, S. Ballantyne, G. Pan, J. Greenblatt, S.D. Patterson, M. Wickens, and D.L. Bentley. 1997a. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**: 357-61.
149. McCracken, S. and J. Greenblatt. 1991. Related RNA polymerase-binding regions in human RAP30/74 and Escherichia coli sigma 70. *Science* **253**: 900-2.
150. Meinhart, A., J. Blobel, and P. Cramer. 2003a. An extended winged helix domain in general transcription factor E/IIE alpha. *J Biol Chem* **278**: 48267-74.
151. Meinhart, A. and P. Cramer. 2004. Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature* **430**: 223-6.
152. Meinhart, A., T. Kamenski, S. Hoepfner, S. Baumli, and P. Cramer. 2005. A structural perspective of CTD function. *Genes Dev* **19**: 1401-15.
153. Meinhart, A., T. Silberzahn, and P. Cramer. 2003b. The mRNA

- transcription/processing factor Ssu72 is a potential tyrosine phosphatase. *J Biol Chem* **278**: 15917-21.
154. Mekler, V., E. Kortkhonjia, J. Mukhopadhyay, J. Knight, A. Revyakin, A.N. Kapanidis, W. Niu, Y.W. Ebright, R. Levy, and R.H. Ebright. 2002. Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. *Cell* **108**: 599-614.
155. Meredith, G.D., W.H. Chang, Y. Li, D.A. Bushnell, S.A. Darst, and R.D. Kornberg. 1996. The C-terminal domain revealed in the structure of RNA polymerase II. *J Mol Biol* **258**: 413-9.
156. Murakami, K.S., S. Masuda, E.A. Campbell, O. Muzzin, and S.A. Darst. 2002. Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* **296**: 1285-90.
157. Murray, A.W. 2004. Recycling the cell cycle: cyclins revisited. *Cell* **116**: 221-34.
158. Murray, S., R. Udupa, S. Yao, G. Hartzog, and G. Prelich. 2001. Phosphorylation of the RNA polymerase II carboxy-terminal domain by the Bur1 cyclin-dependent kinase. *Mol Cell Biol* **21**: 4089-96.
159. Nazar, R.N. 1978. The release and reassociation of 5.8 S rRNA with yeast ribosomes. *J Biol Chem* **253**: 4505-7.
160. Nedeia, E., X. He, M. Kim, J. Pootoolal, G. Zhong, V. Canadien, T. Hughes, S. Buratowski, C.L. Moore, and J. Greenblatt. 2003. Organization and function of APT, a subcomplex of the yeast cleavage and polyadenylation factor involved in the formation of mRNA and small nucleolar RNA 3'-ends. *J Biol Chem* **278**: 33000-10.
161. Nguyen, B.D., K.L. Abbott, K. Potempa, M.S. Kobor, J. Archambault, J. Greenblatt, P. Legault, and J.G. Omichinski. 2003. NMR structure of a complex containing the TFIIF subunit RAP74 and the RNA polymerase II carboxyl-terminal domain phosphatase FCP1. *Proc Natl Acad Sci U S A* **100**: 5688-93.
162. Nguyen, B.D., H.T. Chen, M.S. Kobor, J. Greenblatt, P. Legault, and J.G. Omichinski. 2003. Solution structure of the carboxyl-terminal domain of

- RAP74 and NMR characterization of the FCP1-binding sites of RAP74 and human TFIIB. *Biochemistry* **42**: 1460-9.
163. Nikolov, D.B., H. Chen, E.D. Halay, A.A. Usheva, K. Hisatake, D.K. Lee, R.G. Roeder, and S.K. Burley. 1995. Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature* **377**: 119-28.
164. Noble, C.G., D. Hollingworth, S.R. Martin, V. Ennis-Adeniran, S.J. Smerdon, G. Kelly, I.A. Taylor, and A. Ramos. 2005. Key features of the interaction between Pcf11 CID and RNA polymerase II CTD. *Nat Struct Mol Biol* **12**: 144-51.
165. Nonet, M., D. Sweetser, and R.A. Young. 1987. Functional redundancy and structural polymorphism in the large subunit of RNA polymerase II. *Cell* **50**: 909-15.
166. Ohkuma, Y. and R.G. Roeder. 1994. Regulation of TFIIH ATPase and kinase activities by TFIIIE during active initiation complex formation. *Nature* **368**: 160-3.
167. Ohler, U., G.C. Liao, H. Niemann, and G.M. Rubin. 2002. Computational analysis of core promoters in the Drosophila genome. *Genome Biol* **3**: RESEARCH0087.
168. Okamoto, T., S. Yamamoto, Y. Watanabe, T. Ohta, F. Hanaoka, R.G. Roeder, and Y. Ohkuma. 1998. Analysis of the role of TFIIIE in transcriptional regulation through structure-function studies of the TFIIIEbeta subunit. *J Biol Chem* **273**: 19866-76.
169. Okuda, M., A. Tanaka, Y. Arai, M. Satoh, H. Okamura, A. Nagadoi, F. Hanaoka, Y. Ohkuma, and Y. Nishimura. 2004. A novel zinc finger structure in the large subunit of human general transcription factor TFIIIE. *J Biol Chem* **279**: 51395-403.
170. Okuda, M., Y. Watanabe, H. Okamura, F. Hanaoka, Y. Ohkuma, and Y. Nishimura. 2000. Structure of the central core domain of TFIIIEbeta with a novel double-stranded DNA-binding surface. *Embo J* **19**: 1346-56.
171. Onodera, Y., J.R. Haag, T. Ream, P.C. Nunes, O. Pontes, and C.S. Pikaard. 2005. Plant nuclear RNA polymerase IV mediates siRNA and DNA

- methylation-dependent heterochromatin formation. *Cell* **120**: 613-22.
172. Opalka, N., M. Chlenov, P. Chacon, W.J. Rice, W. Wriggers, and S.A. Darst. 2003. Structure and function of the transcription elongation factor GreB bound to bacterial RNA polymerase. *Cell* **114**: 335-45.
173. Orphanides, G., T. Lagrange, and D. Reinberg. 1996. The general transcription factors of RNA polymerase II. *Genes Dev* **10**: 2657-83.
174. Orphanides, G. and D. Reinberg. 2002. A unified theory of gene expression. *Cell* **108**: 439-51.
175. Otwinowski, Z., and W. Minor. 1996. Processing of X-ray diffraction data collected in oscillation mode. *Meth Enzym* **276**: 307–326.
176. Pappas, D.L., Jr. and M. Hampsey. 2000. Functional interaction between Ssu72 and the Rpb2 subunit of RNA polymerase II in *Saccharomyces cerevisiae*. *Mol Cell Biol* **20**: 8343-51.
177. Pardee, T.S., C.S. Bangur, and A.S. Ponticelli. 1998. The N-terminal region of yeast TFIIB contains two adjacent functional domains involved in stable RNA polymerase II binding and transcription start site selection. *J Biol Chem* **273**: 17859-64.
178. Pokholok, D.K., N.M. Hannett, and R.A. Young. 2002. Exchange of RNA polymerase II initiation and elongation factors during gene expression *in vivo*. *Mol Cell* **9**: 799-809.
179. Poon, D., Y. Bai, A.M. Campbell, S. Bjorklund, Y.J. Kim, S. Zhou, R.D. Kornberg, and P.A. Weil. 1995. Identification and characterization of a TFIIID-like multiprotein complex from *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* **92**: 8224-8.
180. Prelich, G. 2002. RNA polymerase II carboxy-terminal domain kinases: emerging clues to their function. *Eukaryot Cell* **1**: 153-62.
181. Prelich, G. and F. Winston. 1993. Mutations that suppress the deletion of an upstream activating sequence in yeast: involvement of a protein kinase and histone H3 in repressing transcription *in vivo*. *Genetics* **135**: 665-76.
182. Price, D.H. 2000. P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. *Mol Cell Biol* **20**: 2629-34.

183. Price, D.H., A.E. Sluder, and A.L. Greenleaf. 1989. Dynamic interaction between a Drosophila transcription factor and RNA polymerase II. *Mol Cell Biol* **9**: 1465-75.
184. Proudfoot, N.J., A. Furger, and M.J. Dye. 2002. Integrating mRNA processing with transcription. *Cell* **108**: 501-12.
185. Pugh, B.F. 2000. Control of gene expression through regulation of the TATA-binding protein. *Gene* **255**: 1-14.
186. Puig, O., F. Caspary, G. Rigaut, B. Rutz, E. Bouveret, E. Bragado-Nilsson, M. Wilm, and B. Seraphin. 2001. The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* **24**: 218-29.
187. Ramponi, G. and M. Stefani. 1997. Structure and function of the low Mr phosphotyrosine protein phosphatases. *Biochim Biophys Acta* **1341**: 137-56.
188. Rani, P.G., J.A. Ranish, and S. Hahn. 2004. RNA polymerase II (Pol II)-TFIIF and Pol II-mediator complexes: the major stable Pol II complexes and their activity in transcription initiation and reinitiation. *Mol Cell Biol* **24**: 1709-20.
189. Ridder, I.S., H.J. Rozeboom, K.H. Kalk, and B.W. Dijkstra. 1999. Crystal structures of intermediates in the dehalogenation of haloalkanoates by L-2-haloacid dehalogenase. *J Biol Chem* **274**: 30672-8.
190. Rigaut, G., A. Shevchenko, B. Rutz, M. Wilm, M. Mann, and B. Seraphin. 1999. A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* **17**: 1030-2.
191. Roeder, R.G. 1996. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci* **21**: 327-35.
192. Rossignol, M., A. Keriél, A. Staub, and J.M. Egly. 1999. Kinase activity and phosphorylation of the largest subunit of TFIIF transcription factor. *J Biol Chem* **274**: 22387-92.
193. Sakurai, H., H. Mitsuzawa, M. Kimura, and A. Ishihama. 1999. The Rpb4 subunit of fission yeast *Schizosaccharomyces pombe* RNA polymerase II is essential for cell viability and similar in structure to the corresponding subunits of higher eukaryotes. *Mol Cell Biol* **19**: 7511-8.
194. Sambrook J., D.W. Russel. 2001. *Molecular Cloning: a laboratory manual*, 3rd

- ed. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press.
195. Samuelsen, C.O., V. Baraznenok, O. Khorosjutina, H. Spahr, T. Kieselbach, S. Holmberg, and C.M. Gustafsson. 2003. TRAP230/ARC240 and TRAP240/ARC250 Mediator subunits are functionally conserved through evolution. *Proc Natl Acad Sci U S A* **100**: 6422-7.
 196. Sanders, S.L., K.A. Garbett, and P.A. Weil. 2002. Molecular characterization of *Saccharomyces cerevisiae* TFIID. *Mol Cell Biol* **22**: 6000-13.
 197. Schroeder, S.C., B. Schwer, S. Shuman, and D. Bentley. 2000. Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes Dev* **14**: 2435-40.
 198. Schultz, J., F. Milpetz, P. Bork and C.P. Ponting. 1998. SMART. A simple modular architecture research tool: identification of signalling domains. *Proc Natl Acad Sci U S A* **85**: 5857-64.
 199. Schultz, P., S. Fribourg, A. Poterszman, V. Mallouh, D. Moras, and J.M. Egly. 2000. Molecular structure of human TFIIH. *Cell* **102**: 599-607.
 200. Shen, W.C., S.R. Bhaumik, H.C. Causton, I. Simon, X. Zhu, E.G. Jennings, T.H. Wang, R.A. Young, and M.R. Green. 2003. Systematic analysis of essential yeast TAFs in genome-wide transcription and preinitiation complex assembly. *Embo J* **22**: 3395-402.
 201. Shi, X., M. Chang, A.J. Wolf, C.H. Chang, A.A. Frazer-Abel, P.A. Wade, Z.F. Burton, and J.A. Jaehning. 1997. Cdc73p and Paf1p are found in a novel RNA polymerase II-containing complex distinct from the Srbp-containing holoenzyme. *Mol Cell Biol* **17**: 1160-9.
 202. Sims, R.J., 3rd, R. Belotserkovskaya, and D. Reinberg. 2004. Elongation by RNA polymerase II: the short and long of it. *Genes Dev* **18**: 2437-68.
 203. Smale, S.T. and J.T. Kadonaga. 2003. The RNA polymerase II core promoter. *Annu Rev Biochem* **72**: 449-79.
 204. Sopta, M., Z.F. Burton, and J. Greenblatt. 1989. Structure and associated DNA-helicase activity of a general transcription initiation factor that binds to RNA polymerase II. *Nature* **341**: 410-4.
 205. Sopta, M., R.W. Carthew, and J. Greenblatt. 1985. Isolation of three proteins

- that bind to mammalian RNA polymerase II. *J Biol Chem* **260**: 10353-60.
206. Spahr, H., O. Khorosjutina, V. Baraznenok, T. Linder, C. O. Samuelson, D. Hermand, T.P. Mäkelä, S. Holmberg and C.M. Gustafsson. 2003. Mediator influences *Schizosaccharomyces pombe* RNA Polymerase II-dependent transcription *in vitro*. *J Biol Chem* **278**: 51301-6.
207. Steinmetz, E.J. and D.A. Brow. 2003. Ssu72 protein mediates both poly(A)-coupled and poly(A)-independent termination of RNA polymerase II transcription. *Mol Cell Biol* **23**: 6339-49.
208. Su, X.D., N. Taddei, M. Stefani, G. Ramponi, and P. Nordlund. 1994. The crystal structure of a low-molecular-weight phosphotyrosine protein phosphatase. *Nature* **370**: 575–578.
209. Suh, M.H., P. Ye, M. Zhang, S. Hausmann, S. Shuman, A.L. Gnatt, and J. Fu. 2005. Fcp1 directly recognizes the C-terminal domain (CTD) and interacts with a site on RNA polymerase II distinct from the CTD. *Proc Natl Acad Sci U S A* **102**: 17314-9.
210. Sun, Z.W. and M. Hampsey. 1995. Identification of the gene (SSU71/TFG1) encoding the largest subunit of transcription factor TFIIF as a suppressor of a TFIIB mutation in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* **92**: 3127-31.
211. Sun, Z.W. and M. Hampsey. 1996. Synthetic enhancement of a TFIIB defect by a mutation in SSU72, an essential yeast gene encoding a novel protein that affects transcription start site selection *in vivo*. *Mol Cell Biol* **16**: 1557-66.
212. Suzuki, M. 1990. The heptad repeat in the largest subunit of RNA polymerase II binds by intercalating into DNA. *Nature* **344**: 562-5.
213. Svejstrup, J.Q., P. Vichi, and J.M. Egly. 1996. The multiple roles of transcription/repair factor TFIIF. *Trends Biochem Sci* **21**: 346-50.
214. Tan, S., T. Aso, R.C. Conaway, and J.W. Conaway. 1994. Roles for both the RAP30 and RAP74 subunits of transcription factor IIF in transcription initiation and elongation by RNA polymerase II. *J Biol Chem* **269**: 25684-91.
215. Tan, S., R.C. Conaway, and J.W. Conaway. 1995. Dissection of transcription factor TFIIF functional domains required for initiation and elongation. *Proc Natl*

- Acad Sci U S A* **92**: 6042-6.
216. Tan, S., K.P. Garrett, R.C. Conaway, and J.W. Conaway. 1994. Cryptic DNA-binding domain in the C terminus of RNA polymerase II general transcription factor RAP30. *Proc Natl Acad Sci U S A* **91**: 9808-12.
217. Tan, S., Y. Hunziker, D.F. Sargent, and T.J. Richmond. 1996. Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature* **381**: 127-51.
218. Teixeira, A., A. Tahiri-Alaoui, S. West, B. Thomas, A. Ramadass, I. Martianov, M. Dye, W. James, N.J. Proudfoot, and A. Akoulitchev. 2004. Autocatalytic RNA cleavage in the human beta-globin pre-mRNA promotes transcription termination. *Nature* **432**: 526-30.
219. Terwilliger, T.C. 2002. Automated structure solution, density modification and model building. *Acta Crystallogr D Biol Crystallogr* **58**: 1937-40.
220. Thompson, C.M., A.J. Koleske, D.M. Chao, and R.A. Young. 1993. A multisubunit complex associated with the RNA polymerase II CTD and TATA-binding protein in yeast. *Cell* **73**: 1361-75.
221. Thompson, N.E. and R.R. Burgess. 1996. Immunoaffinity purification of RNA polymerase II and transcription factors using polyol-responsive monoclonal antibodies. *Methods Enzymol* **274**: 513-26.
222. Tora, L. 2002. A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription. *Genes Dev* **16**: 673-5.
223. Valay, J.G., M. Simon, M.F. Dubois, O. Bensaude, C. Facca, and G. Faye. 1995. The KIN28 gene is required both for RNA polymerase II mediated transcription and phosphorylation of the Rpb1p CTD. *J Mol Biol* **249**: 535-44.
224. Varon, R., R. Gooding, C. Steglich, L. Marns, *et al.* 2003. Partial deficiency of the C-terminal-domain phosphatase of RNA polymerase II is associated with congenital cataracts facial dysmorphism neuropathy syndrome. *Nat Genet* **35**: 185-9.
225. Vassylyev, D.G., S. Sekine, O. Laptenko, J. Lee, M.N. Vassylyeva, S. Borukhov, and S. Yokoyama. 2002. Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* **417**: 712-9.

-
226. Venema, J. and D. Tollervey. 1999. Ribosome synthesis in *Saccharomyces cerevisiae*. *Annu Rev Genet* **33**: 261-311.
227. Verdecia, M.A., M.E. Bowman, K.P. Lu, T. Hunter, and J.P. Noel. 2000. Structural basis for phosphoserine-proline recognition by group IV WW domains. *Nat Struct Biol* **7**: 639-43.
228. Wang, B.Q. and Z.F. Burton. 1995. Functional domains of human RAP74 including a masked polymerase binding domain. *J Biol Chem* **270**: 27035-44.
229. Wang, B.Q., C.F. Kostrub, A. Finkelstein, and Z.F. Burton. 1993. Production of human RAP30 and RAP74 in bacterial cells. *Protein Expr Purif* **4**: 207-14.
230. Wang W., M. Carey, J.D. Gralla. 1992. Polymerase II promoter activation: closed complex formation and ATP-driven start opening. *Science* **255**: 450-3.
231. Wang, W., R. Kim, J. Jancarik, H. Yokota, and S.H. Kim. 2001. Crystal structure of phosphoserine phosphatase from *Methanococcus jannaschii*, a hyperthermophile, at 1.8 Å resolution. *Structure (Camb)* **9**: 65-71.
232. Wei, W., D. Dorjsuren, Y. Lin, W. Qin, T. Nomura, N. Hayashi, and S. Murakami. 2001. Direct interaction between the subunit RAP30 of transcription factor IIF (TFIIF) and RNA polymerase subunit 5, which contributes to the association between TFIIF and RNA polymerase II. *J Biol Chem* **276**: 12266-73.
233. Weideman, C.A., R.C. Netter, L.R. Benjamin, J.J. McAllister, L.A. Schriedekamp, R.A. Coleman, and B.F. Pugh. 1997. Dynamic interplay of TFIIA, TBP and TATA DNA. *J Mol Biol* **271**: 61-75.
234. Werten, S., A. Mitschler, C. Romier, Y.G. Gangloff, S. Thuault, I. Davidson, and D. Moras. 2002. Crystal structure of a subcomplex of human transcription factor TFIID formed by TATA binding protein-associated factors hTAF4 (hTAF(II)135) and hTAF12 (hTAF(II)20). *J Biol Chem* **277**: 45502-9.
235. West, M.L. and J.L. Corden. 1995. Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. *Genetics* **140**: 1223-33.
236. West, S., N. Gromak, and N.J. Proudfoot. 2004. Human 5' → 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites.

- Nature* **432**: 522-5.
237. Westover, K.D., D.A. Bushnell, and R.D. Kornberg. 2004. Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science* **303**: 1014-6.
238. Williams, R.S., R. Green, and J.N. Glover. 2001. Crystal structure of the BRCT repeat region from the breast cancer-associated protein BRCA1. *Nat Struct Biol* **8**: 838-42.
239. Wu, W.H., I. Pinto, B.S. Chen, and M. Hampsey. 1999. Mutational analysis of yeast TFIIB. A functional relationship between Ssu72 and Sub1/Tsp1 defined by allele-specific interactions with TFIIB. *Genetics* **153**: 643-52.
240. Xie, X., T. Kokubo, S.L. Cohen, U.A. Mirza, A. Hoffmann, B.T. Chait, R.G. Roeder, Y. Nakatani, and S.K. Burley. 1996. Structural similarity between TAFs and the heterotetrameric core of the histone octamer. *Nature* **380**: 316-22.
241. Xu, Y.X., Y. Hirose, X.Z. Zhou, K.P. Lu, and J.L. Manley. 2003. Pin1 modulates the structure and function of human RNA polymerase II. *Genes Dev* **17**: 2765-76.
242. Yan, D., H.S. Cho, C.A. Hastings, M.M. Igo, S.Y. Lee, J.G. Pelton, V. Stewart, D.E. Wemmer, and S. Kustu. 1999. Beryllium fluoride mimics phosphorylation of NtrC and other bacterial response regulators. *Proc Natl Acad Sci U S A* **96**: 14789-94.
243. Yan, Q., R.J. Moreland, J.W. Conaway, and R.C. Conaway. 1999a. Dual roles for transcription factor IIF in promoter escape by RNA polymerase II. *J Biol Chem* **274**: 35668-75.
244. Yankulov, K.Y. and D.L. Bentley. 1997. Regulation of CDK7 substrate specificity by MAT1 and TFIIH. *Embo J* **16**: 1638-46.
245. Yeo, M., S.K. Lee, B. Lee, E.C. Ruiz, S.L. Pfaff, and G.N. Gill. 2005. Small CTD phosphatases function in silencing neuronal gene expression. *Science* **307**: 596-600.
246. Yeo, M., P.S. Lin, M.E. Dahmus, and G.N. Gill. 2003. A novel RNA polymerase II C-terminal domain phosphatase that preferentially

- dephosphorylates serine 5. *J Biol Chem* **278**: 26078-85.
247. Yonaha, M., T. Tsuchiya, and Y. Yasukochi. 1997. Cell-cycle-dependent phosphorylation of the basal transcription factor RAP74. *FEBS Lett* **410**: 477-80.
248. Yu, X., C.C. Chini, M. He, G. Mer, and J. Chen. 2003. The BRCT domain is a phospho-protein binding domain. *Science* **302**: 639-42.
249. Yudkovsky, N., J.A. Ranish, and S. Hahn. 2000. A transcription reinitiation intermediate that is stabilized by activator. *Nature* **408**: 225-9.
250. Zawel, L., K.P. Kumar, and D. Reinberg. 1995. Recycling of the general transcription factors during RNA polymerase II transcription. *Genes Dev* **9**: 1479-90.
251. Zhang, C., H. Yan, and Z.F. Burton. 2003. Combinatorial control of human RNA polymerase II (RNAP II) pausing and transcript cleavage by transcription factor IIF, hepatitis delta antigen, and stimulatory factor II. *J Biol Chem* **278**: 50101-11.
252. Zhang, G., E.A. Campbell, L. Minakhin, C. Richter, K. Severinov, and S.A. Darst. 1999. Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* **98**: 811-24.
253. Zhang, H., D.O. Richardson, D.N. Roberts, R. Utlej, H. Erdjument-Bromage, P. Tempst, J. Cote, and B.R. Cairns. 2004. The Yaf9 component of the SWR1 and NuA4 complexes is required for proper gene expression, histone H4 acetylation, and Htz1 replacement near telomeres. *Mol Cell Biol* **24**: 9424-36.
254. Zhang, J. and J.L. Corden. 1991. Phosphorylation causes a conformational change in the carboxyl-terminal domain of the mouse RNA polymerase II largest subunit. *J Biol Chem* **266**: 2297-302.
255. Zhang, M., C.V. Stauffacher, D. Lin, and R.L. Van Etten. 1998. Crystal structure of a human low molecular weight phosphotyrosyl phosphatase. Implications for substrate specificity. *J Biol Chem* **273**: 21714-20.
256. Ziegler, L.M., D.A. Khapersky, M.L. Ammerman, and A.S. Ponticelli. 2003. Yeast RNA polymerase II lacking the Rpb9 subunit is impaired for interaction with transcription factor IIF. *J Biol Chem* **278**: 48950-6.

CURRICULUM VITAE

TOMISLAV KAMENSKI

PERSONAL DETAILS

Date of birth: September 21, 1977

Place of birth: Zagreb, Croatia

Nationality: Croatian

Family status: single, no children

EDUCATION

- | | |
|-------------------------|---|
| October, 2002 - present | University of Munich, Department of Chemistry and Biochemistry, Gene Center, Munich, Germany
PhD Studies in Biochemistry. Supervisor: Prof. Patrick Cramer |
| 1996 - 2002 | University of Zagreb, Faculty of Science, Department of Chemistry, Zagreb, Croatia
Chemiae Ingeniarius Diplomate Probatus (Diplomirani inženjer kemije). |
| 1992 – 1996 | Secondary education: Srednja Škola Krapina, Krapina, Croatia |
| 1984 – 1992 | Primary education: Osnovna škola Ljudevit Gaj, Krapina, Croatia |

FELLOWSHIPS AND AWARDS

- | | |
|-------------|---|
| 2003 - 2005 | Boehringer Ingelheim Fonds (B.I.F.) predoctoral fellowship |
| 2004 - 2005 | Member of the International Graduate School “Nano-Bio-Technology” at the Center for NanoScience, LMU München |
| 2001 | University of Zagreb Rector’s Award (“Rektorova nagrada”); for the best student work in the field of chemistry in academic year 2000/2001 |