Structure and requirement of the Spt6 SH2 domain and an *in vitro* system to test the "torpedo model" of transcription termination

Stefan Dengl

aus

München

2009

Erklärung

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

Ehrenwörtliche Versicherung

Diese Dissertation wurde selbständig und ohne unerlaubte Hilfe erarbeitet.

München, am

_____

Stefan Dengl

Dissertation eingereicht am

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

Mündliche Prüfung am 25. März 2009

Acknowledgements

First of all I would like to thank my supervisor Prof. Dr. Patrick Cramer for giving me the opportunity to work on challenging projects in a highly interdisciplinary manner. The freedom to realize own ideas in combination with your continuous advice and motivation is invaluable for the enthusiasm of a young scientist.

I want to thank my collaborators Andreas Mayer, Mai Sun, Kristin Leike, Matthias Siebert and Johannes Soeding for their work and the discussions, but also for a lot fun and reflief from drawbacks in the projects. It was really a pleasure to work with you guys!

Very special thanks go out to Claus Kuhn and Stephan Jellbauer, not only for teaching me crystallography and yeast work, respectively, but most important for your friendship in good and in hard times. The legendary "coffee-at-11.30" was often enough an escape from scientific misery and sometimes gave birth to new ideas. I will never forget the time we spent together, also outside of the lab.

I wish to thank Elisabeth, Jasmin, Gerke and Florian for their help with elongation complexes. Laurent, Ale and Dirk for support concerning crystallography. Ania for advice on the ChIP-on-chip protocol. Claudia and Stefan Benkert for help with everything, without you the lab would be a battlefield. Many thanks to Heidi Feldmann for help with yeast work and for not killing me for some of my jokes. Thank you Gregor for illuminating me with your fluorescent expertise. I want to thank the former members of the lab, Hubert, Karim, Michela, Sabine, Sonja, Toni, and Tomislav for a lot of support especially in the beginnig. I want to thank you, Erika, for being an accomplice in "chromatin-related" issues that were otherwise quite exotic in the lab.
Of course I want to thank everybody else in the Cramer group for help, media, buffers, DNA and protein standards, gels, coffee, cake, meaningful discussions and silly conversations.

Von ganzem Herzen danke ich meinen Eltern und meiner Oma Elisabeth. Ohne eure moralische, aber auch finanzielle Unterstützung wäre das Studium und so viel mehr undenkbar gewesen.

Dir Moni danke ich für das schöne letzte Jahr und dass du mir immer gute Laune verordnet hast auch wenn die Batterie mal leer war.

Summary

During the biogenesis of messenger RNA, RNA polymerase II (RNAP II, Pol II) associates with numerous proteins and multiprotein complexes. These factors regulate the correct progression of the transcription cycle, which can be divided into three major phases: initiation, elongation and termination. This thesis describes the characterization of two important factors involved in two different phases of the transcription cycle in a highly interdisciplinary approach. Spt6 is an essential, modular protein involved in transcription elongation. It was characterized as a histone chaperone that binds to histones and assembles nucleosomes onto DNA after the passage of Pol II. In addition, it is linked to processes like splicing, mRNA processing and export, and histone modification. This functional versatility makes it a central player in the elongation process. The first part of this work shows the high resolution structure of the C-terminal SH2 domain of Spt6. The domain was shown previously to interact with the C-terminal domain (CTD) of Pol II, phosphorylated at Ser_2 residues. Thus it links Spt6 functions directly to the transcription machinery. The domain has an unconventional binding specifity in contrast to the numerous SH2 domains involved in cellular signaling pathways. It binds phosphoserine instead of phosphotyrosine. In addition, it is the only SH2 domain encoded in the yeast genome, opposing a multitude of these domains in the genomes of higher eukaryotes. The X-ray structure gives insight into the peculiarities of this domain, with implications on its substrate specificity and molecular evolution. In this light, a model for the interaction with the CTD, as well as a deep analysis of the evolutionary relationship to other SH2 domains is presented. Microarray gene expression analysis shows the impact of a deletion of the SH2 domain on the transcription of the yeast genome. A genome wide localization map of Spt6, obtained by ChIP-on-chip experiments, is presented and compared to the localization of Pol II.

The nuclear exoribonuclease complex Rat1/Rai1 plays a role in transcription termination. Two models explain the events at the end of a protein coding gene. The so-called „torpedo model" states that Rat1/Rai1 processively degrades 3' nascent RNA that is still attached to elongating Pol II after the mRNA product is cleaved by cleavage/polyadenylation factors. Upon contact with Pol II, Rat1/Rai1 is thought to disrupt the elongation complex and terminate transcription. There is however also data supporting the "allosteric model" that predicts the recruitment of a termination factor or the dissociation of an anti-termination factor. The second part of this thesis describes the establishment of a highly defined biochemical assay to test the "torpedo model" *in vitro*. A protocol for the expression and purification of the recombinant and active exonuclease complex and an additional interacting protein, Rtt103, was established. In combination with an improved *in vitro* elongation assay it is shown that Rat1 is not the dedicated termination factor that is predicted by the torpedo model.

Publications

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

Dengl S., Mayer, A., Sun, M., Cramer, P. (2009). Structure and widespread requirement of the Spt6 SH2 domain, the only SH2 domain in yeast. *Journal of molecular biology,* in revision.

Dengl S., Cramer P. (2009). The torpedo nuclease Rat1 is insufficient to terminate RNA polymerase II *in vitro. The Journal of biological chemistry,* in revision.

Sydow, J. F. , Brueckner, F., Dengl, S., Cramer P. (2009). Molecular basis of RNA polymerase II fidelity: mismatch specificity and RNA fraying. Manuscript in preparation.

Cramer P, Armache KJ, Baumli S, Benkert S, Brueckner F, Buchen C, Damsma GE, Dengl S, Geiger SR, Jasiak AJ, Jawhari A, Jennebach S, Kamenski T, Kettenberger H, Kuhn CD, Lehmann E, Leike K, Sydow JF, Vannini A (2008) Structure of eukaryotic RNA polymerases. *Annual review of biophysics* **37:** 337-352

Table of contents

1 General introduction

1.1 Structure of RNA polymerase II and the elongation complex

In eukaryotic cells, the transcription of all protein-coding genes, as well as small nucleolar and small nuclear RNA (snoRNA and snRNA, respectively) is carried out by RNA polymerase II (Pol II) in the nucleoplasm.

Polymerase II consists of a 10 subunit core and a peripheral heterodimer of the subunits Rpb4 and Rpb7 . The crystal structures of core Pol II revealed, that the two largest subunits, Rpb1 and Rpb2, are on opposite sides of a positively charged cleft that bears the active site (Cramer et al, 2000). Together with the bridge helix, which spans the cleft, the active site lines a pore in the floor of the cleft. The smaller subunits are placed around the "jaw-like" structure of Rpb1 and Rpb2. Indeed, these two subunits work like a jaw, as the Rpb1 side of the cleft forms a mobile clamp, which was trapped in two different open states in the free core structures (Cramer et al, 2001a). In the structure of a core complex that included DNA and RNA the clamp was closed (Gnatt et al, 2001; Kettenberger et al, 2004b). This mobile clamp is connected to the body of the polymerase by five switch regions that are variable in conformation. The Rpb2 side of the cleft consists of the lobe and protrusion domains and Rpb2 also forms a protein wall that blocks the end of the cleft. Additional crystal structures at a resolution around 4 Å revealed the position of the two missing subunits, Rpb4 and Rpb7, which can dissociate from the yeast enzyme (Edwards et al, 1991). The heterodimer binds like a wedge between the clamp and the linker of the long C-terminal domain (CTD, see 1.2) of Pol II (Armache et al, 2003; Armache et al, 2005b; Bushnell & Kornberg, 2003). In the crystal structures of the 12 subunit Pol II, the clamp was always in a closed conformation. A complete atomic model of Pol II could be refined, when the crystal structure of free Rpb4/7 together with an improved resolution of the complete Pol II was available (Armache et al, 2005b).

The interactions of Pol II with its substrate was revealed in structural studies of Pol II-nucleic acid complexes. The point of DNA entry into the polymerase cleft was shown by EM (Poglitsch et al, 1999). The first crystal structure of the core Pol II transcribing a tailed template DNA revealed downstream DNA entering the cleft and a 8 to 9 base pair DNA-RNA hybrid in the active center (Gnatt et al, 2001). In addition, comparison with the structure of the core polymerase (Cramer et al, 2001a) suggested protein surface elements for functional roles. With the use of syntethic nucleic acid scaffols containing mismatch transcription bubbles, the location of the downstream DNA duplex with RNA annealed to the central mismatched bubble region could be detected, as well as a part of the upstream region (Fig. 1, Kettenberger et al., 2004b).
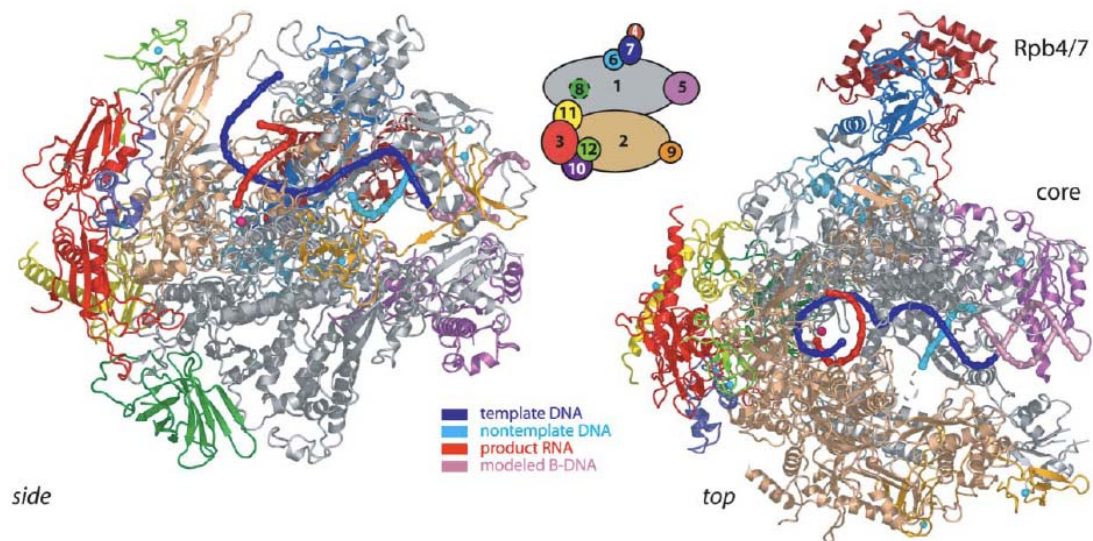
Figure 1: Structure of complete 12-subunit RNA polymerase II elongation complex (Kettenberger et al, 2004a).

Two views of a ribbon model of the protein subunits and nucleic acids, side view (left) and top view (right). The polymerase subunits Rpb1–Rpb12 are colored according to the key between the views. Template DNA, nontemplate DNA, and product RNA are shown in blue, cyan, and red, respectively. Phosphorous atoms are indicated as spheres and extrapolated B-form downstream DNA is colored in light pink. Eight zinc ions and the active site magnesium ion are depicted as cyan spheres and a magenta sphere, respectively. Secondary structure assignments for pol II are according to (Cramer et al, 2001b) and (Armache et al, 2005a). This figure is adapted from Kettenberger et al, (2004a).

The 3' end of the RNA is centered at the active site of Pol II that contains two magnesium ions which are needed for the catalysis of nucleotide incorporation (Steitz, 1998). One metal ion is permanently bound inside the Pol II active site (Cramer et al, 2001a), the second metal is binding the tri-phosphate moiety of the incoming NTP (Westover et al, 2004). Nucleotide insertion involves the trigger loop, a mobile part of the active center (Kettenberger et al, 2003; Vassylyev et al, 2002), that folds upon binding of the new NTP (Kettenberger et al, 2003; Vassylyev et al, 2002; Wang et al, 2006). Crystal structures including the NTP substrate suggested how Pol II selects the correct NTP for incorporation (Kettenberger et al, 2004b; Wang et al, 2006; Westover et al, 2004). The NTP was trapped in two slightly different but overlapping conformations, termed the insertion site (active) and the preinsertion site (inactive). In both states the NTP forms Watson-Crick interactions with a base in the DNA template.

In one model, the NTP to be incorporated first binds to an open active center conformation in the preinsertion site, where the two magnesium ions are not close enough for catalysis. Folding of the trigger loop then closes the active center, delivers the NTP to the insertion site

and enables catalysis (Kettenberger et al, 2004b; Vassylyev et al, 2007b). An alternative model for nucleotide addition involves binding of the NTP to a putative entry site in the pore, in which the nucleotide base is oriented away from the DNA template. Rotation of the NTP around its bound metal ion brings it into the insertion site (Westover et al, 2004).

The bridge helix, that spans the cleft of Pol II, is also involved in the nucleotide addition cycle. Straight and bent conformations of this highly conserved helix were observed in different crystal structures of different polymerases (Cramer et al, 2001a; Vassylyev et al, 2002; Zhang et al, 1999) and this movement in concert with movement of the trigger loop apparently directs nucleotide addition to the growing RNA-chain and translocation of Pol II relative to the template RNA (Brueckner & Cramer, 2008; Tuske et al, 2005).

1.2 The C-terminal domain (CTD) of Pol II subunit Rpb1

The CTD is a unique feature of the largest subunit of Pol II, Rpb1. It consists of multiple heptapeptide repeats of the consensus sequence Tyr_1-Ser_2-Pro_3-Thr_4-Ser_5-Pro_6-Ser_7 (Y_1-S_2-P_3-T_4-S_5-P_6-S_7). The length of the CTD increases with organismic complexity, for example 26 repeats in yeast and 52 in human cells. Yeast needs at least eight repeats for viability (Nonet et al, 1987; West & Corden, 1995). The heptads are not uniformly distributed over the length of the CTD, which can be divided into 3 regions: a short N-terminal linker that attaches it to the body of the polymerase (Cramer et al, 2001a), a region of continuously repeated consensus heptads and a rather diverse part at the C-terminus where the repeats diverge from the consensus sequence (Chapman et al, 2008). The CTD serves as a platform for the integration of nuclear events by binding proteins that are involved in mRNA biogenesis and other transcription-coupled reactions. These interactions are timed and allocated to certain phases of the transcription cycle by the dynamic phosphorylation and dephosphorylation of the heptapeptide residues (Orphanides & Reinberg, 2002). Potential phosphorylation sites in the consensus sequence are Tyr_1, Ser_2, Thr_4, Ser_5, Ser_7. Phosphorylation of Ser_2 and Ser_5 predominates the regulative events that are described so far (Corden et al, 1985; Dahmus, 1996; Zhang & Corden, 1991). This phosphorylation pattern is established and maintained by the concerted action of CTD kinases and phosphatases (Meinhart et al, 2005). Ser_5-phosphorylation occurs proximal to the promoter during initiation and early elongation, which is essential for the recruitment of capping enzymes (Cho et al, 1997; Komarnitsky et al, 2000; McCracken et al, 1997; Schroeder et al, 2000). Following transcription towards the 3' end, Ser_5-phosphorylation decreases, whereas Ser_2-phosphorylation increases, with an intermittent overlap of both phosphorylations (Ser_5-Ser_2) (Cho et al, 2001; Schroeder et al, 2000). This leads to the recruitment of mRNA processing, polyadenylation and termination factors (Ahn et al, 2004; McCracken et al, 1997; Proudfoot et al, 2002). Both modifications are independently essential for cell viability (West & Corden, 1995). Recently, it was shown

that phosphorylation of Ser_7 is involved in the expression of snRNA genes (Chapman et al, 2007; Egloff et al, 2007), which further increases the complexity of CTD-directed regulation.

A further layer of complexity is introduced by the combination of the specific phosphorylations combined with *cis-/trans*-interconversions of peptide bonds N-terminal of prolines in the heptad repeats, which are regulated by peptidyl-prolyl isomerases (Hani et al, 1999; Lu et al, 1999). Considering only Ser_2 and Ser_5 phosphorylation, this gives a total of 16 different CTD states. This defines the "CTD code" which specifies the position of Pol II within the transcription cycle (Buratowski, 2003).

Due to its flexibility, the CTD cannot be seen in Pol II structures (Armache et al, 2003; Cramer et al, 2001a), but it is not entirely unstructured. Studies by NMR and circular dichroism showed an overall structural plasticity with residual structure and the propensity to form β-turns (Meinhart et al, 2005). A leap forward in gaining information about CTD-structure were crystal-structures of CTD-binding domains in conjunction with synthetic CTD peptides. The structure of the WW domain of Pin1, a CTD-specific Peptidylprolyl-cis/trans-isomerase, binding to a single heptad repeat with phosphate moieties on Ser_2 and Ser_5 shows the CTD binding as an extended coil with both phosphoserine-proline peptide bonds in the *trans* configuration (Verdecia et al, 2000). A different conformation was seen in the structure of guanylyltransferase Cgt1 with Ser_5-phosphorylated repeats. There, an extended surface of Cgt1 binds the repeats by anchoring both ends by electrostatic interactions with Ser_5-P. Additional van der Waals contacts between Cgt1 and CTD residues contribute to the binding (Fabrega et al, 2003). The first structure of a protein binding a Ser_2-phosphorylated CTD-peptide was the CID-domain (CTD-interacting domain) of Pcf11, which is involved in pre-mRNA 3'-end processing and transcription termination (Amrani et al, 1997). There, a Ser_2P-Pro_3-Thr_4-Ser_5 is recognized not directly via the phosphate moiety. The four central amino-acids form a β-turn, whereas the flanking residues are in a extended conformation. The phospho-group points away from the domain-surface but seems to stabilize the β-turn by a hydrogen-bond (Meinhart & Cramer, 2004). The NMR structure of the histone H3 methyltransferase Set2 SRI-domain shows a novel CTD-binding fold, a left handed three-helix bundle. The interaction with two heptapeptide repeats of a Ser_2-/Ser_5-phosphorylated peptide was shown by NMR titration experiments and biacore binding studies, and revealed the binding of both tyrosine residues by the SRI-domain (Li et al, 2005; Vojnic et al, 2006).

Taken together, the interactions of the CTD with the various factors that assemble on it are structurally diverse and best characterized by an "induced fit" mechanism.

1.3 The mRNA transcription cycle

Unlike bacterial RNA polymerases, where a single subunit called the σ-factor is sufficient to assist the polymerase to start transcription, the situation in eukaryotic organisms is much more complex. Pol II is guided by numerous factors to start transcription at the promoters of genes and every subsequent step is highly regulated by a large arsenal of different proteins and regulative nucleic acid sequences. This interplay of factors can be described as the transcription cycle, which can be divided into the three major stages of initiation, elongation and termination. A further subdivision of these stages is shown in Fig. 2.
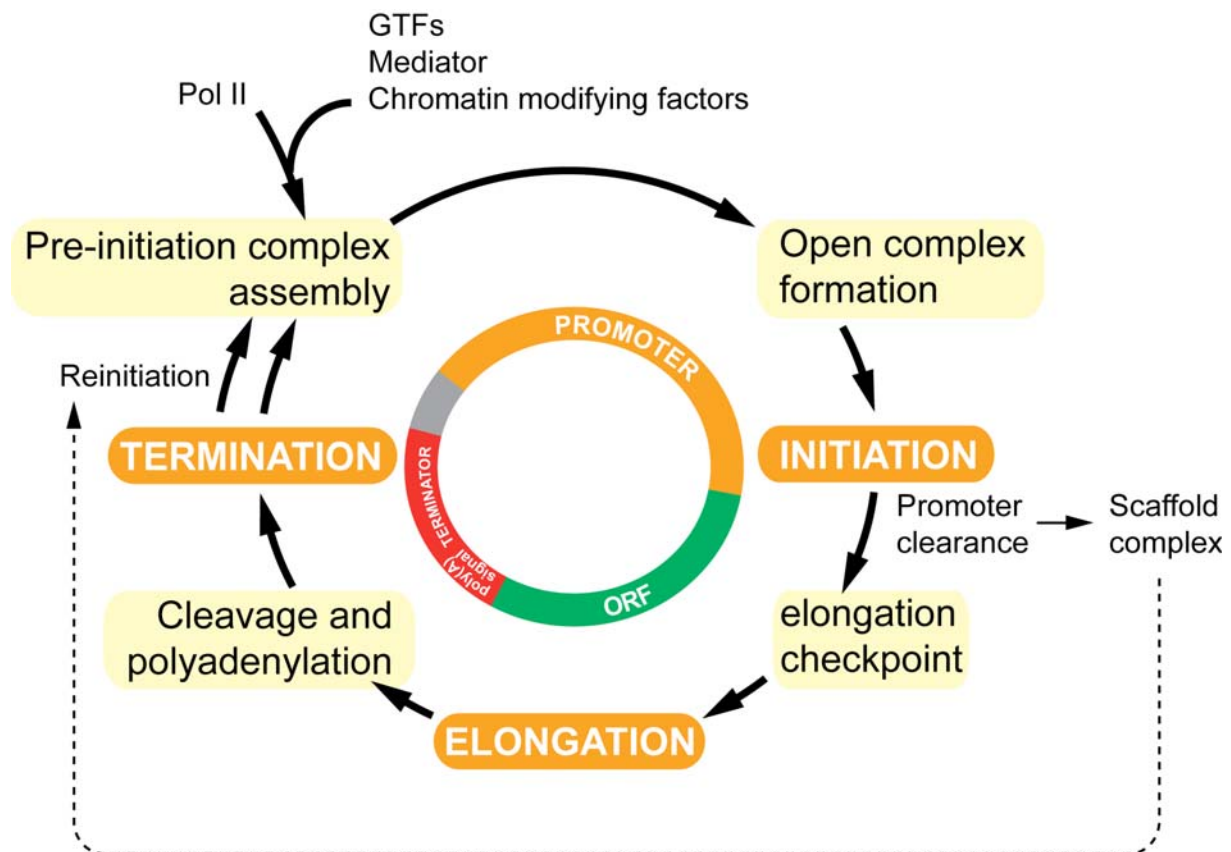
Figure 2: The mRNA transcription cycle

Main phases of the transcription cycle are colored orange, important events of regulation are coloured in yellow. The circle in the middle depicts the occurrence of the events in relation to the gene. GTFs = general transcription factors; ORF = open reading frame.

Numerous factors regulate the initiation of gene transcription. The general transcription factors (GTFs) guide Pol II to the promoter and open the DNA duplex. The large multi-subunit complex Mediator is essential for basal and activated transcription and transmits signals from regulatory factors to Pol II (Hahn, 2004; Lee & Young, 2000). Factors that interact with chromatin (ATP-dependent chromatin remodeling enzymes, histone acetylases, histone methylases) have to regulate the structure of chromatin in promoter regions to facilitate access of the transcription machinery to DNA to initiate transcription. Likewise the CTD seems to play a role in the formation of a stable initiation complex and its transformation into a elongation competent form (Lux et al, 2005). During promoter clearance, the pre-initiation complex is partially disassembled. A subset of GTFs remains at the promoter, leaving behind a scaffold-complex (Fig. 2) for the facilitated reinitiation of transcription of the same gene (Hahn, 2004).

Shortly after initiation in the early elongation phase, the elongation complex is not yet stably formed and Pol II tends to release RNA, resulting in the appearance of short RNA products (abortive initiation). This is accompanied by the tendency of Pol II to slip laterally on the template (Pal & Luse, 2002). This tendency diminishes when the RNA is synthesized to a length of 8-9 nt and becomes undetectable when the RNA is ~23 nt long. This comes together with an increased stability of the transcription complex (Kireeva et al, 2000; Pal & Luse, 2003). Starting early elongation, the transcription complex enters a checkpoint to ensure proper capping of the 5' end of freshly generated RNA. A major player in this checkpoint regulation is the heterodimeric factor Spt4/Spt5 (DSIF in human cells) that binds to Pol II shortly after initiation (Wada et al, 1998). Subsequently, this complex recruits the negative elongation factor (NELF), that traps the transcription machinery at promoter proximal sites (Yamaguchi et al, 1999) (see also section 3.1.2). During this time window, the CTD is phosphorylated at Ser_5-residues and capping enzymes are recruited (Pei & Shuman, 2002). After capping, the kinase P-TEFb binds to Pol II and phosphorylates Spt5 within a region called the CTR (C-terminal region) and the CTD on Ser_2-residues (Yamada et al, 2006). This abolishes the repressive nature of the Pol II-Spt4/5-NELF interaction and releases the transcription machinery into the phase of productive transcript elongation.

1.3.1 Transcription elongation

At first sight, the phase of transcript elongation appears to be a straight-forward addition of nucleotides to the 3' end of the RNA-chain. This central enzymatic reaction, catalyzed by Pol II, is quite well understood, as outlined in section 1.1. However, also this productive phase is highly regulated. This is necessary because Pol II encounters diverse obstacles on its way along the gene. One category of these impediments to elongation is coming from the DNA template itself, including drug-induced or sequence dependent pause and arrest. DNA

lesions can hinder Pol II from proceeding with nucleotide addition. A UV light induced thymine-thymine cyclobutane pyrimidine dimer (CPD) or a guanine-guanine intrastrand cross-link induced by the anticancer drug cisplatin can form such a block. Recent structural studies revealed the molecular basis of how the transcription apparatus copes with these situations (Brueckner et al, 2007; Damsma et al, 2007). Pol II stalls at these lesions and the components of the transcription coupled repair (TCR) machinery are assembled, the lesion-containing DNA fragment is removed and the gap in the DNA is repaired. Importantly, the detailed mechanism of transcriptional stalling at the two lesions differ.

Pol II often encounters pause-sites that are template-intrinsic. Those are often A-T rich and Pol II moves backwards at these sites, which extrudes the RNA 3' end through the polymerase pore beneath the active site (Cramer et al, 2000; Nudler et al, 1997). Elongation factor TFIIS assists Pol II in overcoming such an arrest by cleavage of the extruded RNA (Fish & Kane, 2002). TFIIS thereby enhances a weak nuclease activity that is intrinsic to Pol II (Izban & Luse, 1992; Reines, 1992; Wang & Hawley, 1993). Cleavage is accomplished by insertion of TFIIS into the pore, modifying the Pol II active site, thus triggering nuclease activity (Kettenberger et al, 2003).

In the cell, Pol II is not transcribing naked DNA but a chromatinized template. The basic building block of chromatin is a nucleosome, consisting of an octamer of four histone proteins wrapped in 147 base pairs of DNA. The histone proteins H2A, H2B, H3 and H4 are each composed of a globular domain and an unstructured tail domain (Luger et al, 1997). These histone tails can be modified in various ways, including acetylation of lysines, methylation of lysines and arginines, phosphorylation of serines and threonines, ubiquitination of lysines, sumoylation of lysines, and ADP-ribosylation of glutamic acids. These modifications as a whole write down the "histone code" which is defining epigenetic regulation that preserves genomic information past the genetic code (Jenuwein & Allis, 2001; Strahl & Allis, 2000). Nucleosomes initially pose a block to transcription by Pol II and the chromatin has to be "prepared" by specific factors for access of the transcription machinery to do its job. Two especially important elongation factors that are maintaining chromatin structure – histone chaperones Spt6 and FACT – are described in more detail in section 3.1.2. Numerous other factors are involved in this process, regulating modifications like histone-acetylation, histone-methylation, ATP-dependent chromatin remodeling and histone composition (Hartzog et al, 2002; Sims et al, 2004). These and other important elongation factors are summarized in Table 1.

When elongating polymerase II reaches the 3' regions of a gene and transcribes the poly(A) site, the transcription cycle enters its final phases: cleavage/polyadenylation and termination.

Table 1: Elongation factors of RNA polymerase II

Elongation factor	Function
TFIIF	involved in PIC formation, alleviates pausing, stimulates rate of Pol II, modulates TFIIS
Elongins	alleviate pausing, stimulate rate of Pol II
ELL	alleviates pausing, stimulates rate of Pol II
Spt4/5 (DSIF)	stimulates elongation, suppresses early transcript termination, stimulates capping
NELF	Pol II checkpoint control
CSB	Stimulates elongation, modulates TFIIS, has a role in rescuing RNAP II at DNA lesions and transcription-coupled nucleotide excision repair
FCP1	Stimulates elongation, recycles RNAP II through dephosphorylation of Ser_5 CTD, role in capping
TFIIS	Stimulates RNAP II-mediated cleavage of nascent transcript to alleviate arrest
Spt6	Stimulates elongation, modulates chromatin structure, histone chaperone activity
HDAg	Stimulates elongation, binds RNAP II, displaces NELF, functionally distinct from TFIIF
19S Proteosome	Recruited by H2B monoubiquitination, involved in H3-K4 methylation
P-TEFb	Relieves NELF-mediated pausing, phosphorylates Ser_2 on the CTD and DSIF (Spt5)
Ssu72	Dephosphorylation of Ser_5 CTD, role in 3'-end processing events
SWI/SNF	Remodels chromatin in an ATP-dependent fashion
Isw1	Regulates Ser_5 and Ser_2 phosphorylation, H3-K4 and H3-K36 methylation, and the recruitment of 3'-end processing factors
Chd1	Genetically and physically interacts with elongation factors, localized within coding regions
FACT	Facilitates elongation through chromatin, modulates chromatin structure, histone chaperone activity
Set1	Methylates histone H3-K4, localized to promoter and coding regions
Set2	Methylates histone H3-K6, localized to coding regions
Paf	Modulates H2B monoubiquitination and H3-K4 methylation, recruits Set1 and Set2, involved in modulating mRNA maturation
THO	Required for transcription of long transcripts, or those with high GC content
TREX	Links elongation to pre-mRNA splicing and export, surveillance, interaction with Spt6
Iws1/Spn1	Associates with Spt6, localizes to coding regions

The table is taken and modified from (Sims et al, 2004)

1.3.2 Transcription termination

All the preceding phases of the transcription cycle were regulated by a large number of different proteins in a highly cooperative manner. This is also true for the final stages of transcription, cleavage/polyadenylation and termination. Transcription termination is an important process in gene regulation, allowing elongation complexes (ECs) to dissociate from template DNA. Defects in termination lead to interference of non-terminated EC with the initiation events of genes downstream on the DNA, leading to major disturbances of proper gene expression (Bateman & Paule, 1988; Henderson et al, 1989).

Termination is best described in prokaryotes. In these organisms, termination sites can be basically classified as intrinsic or factor-dependent. Intrinsic termination is dependent on DNA sequences consisting of a palindromic region followed by a run of T residues (d'Aubenton Carafa et al, 1990). When transcribed, the RNA forms a stable hairpin structure followed by a stretch of U-residues. There are basically three regions of protein-nucleic acids interactions within the bacterial RNAP: the double-stranded DNA binding site (DBS), the RNA/DNA heteroduplex binding site (HBS), and the single-stranded RNA binding site (RBS) (Korzheva et al, 1998). The formation of the hairpin leads to a destabilization of these interactions. The poly-U-stretch forms an unstable A-U-hybrid and occupies the HBS, while the hairpin binds the RBS and displaces the RNA. Partial melting of the hybrid is a prerequisite for hairpin formation. Subsequently, these changes in elongation complex structure lead to destabilization and thus disintegration (Gusarov & Nudler, 1999; Yarnell & Roberts, 1999). RNAP can terminate independently by this mechanism, but *in vivo* it is stimulated by additional factors, such as NusA (Farnham et al, 1982; Schmidt & Chamberlin, 1987; Ward & Gottesman, 1981).

Rho-dependent termination depends on the homo-hexameric RNA-translocase Rho (Roberts, 1969). Termination is induced by a sequence in the nascent transcript called RUT (Rho-utilization site), to which Rho is binding (Ceruzzi et al, 1985). After binding, cycles of ATP-hydrolysis induce a conformational change that leads to a 5' to 3' directed translocation of Rho along the RNA. Upon contact, Rho disintegrates the elongation complex and induces termination (Nudler & Gottesman, 2002). Again, the termination efficiency *in vivo* can be stimulated by additional factors, NusG (Nehrke et al, 1993; Sullivan & Gottesman, 1992) and NusA (Zheng & Friedman, 1994).

An additional factor that can terminate prokaryotic RNAP and couple it to the DNA excision repair machinery is the ATP-dependent DNA translocase MFD, that interacts with the polymerase and upstream DNA. Disruption of the elongation complex is achieved by a forward translocation of the polymerase over a DNA lesion (Park et al, 2002; Roberts & Park, 2004).

Eukaryotic cells make use of different mechanisms to terminate the transcription of their polymerases. Ribosomal DNA (rDNA), that is transcribed by Pol I, is arrayed in ~150 copies of the rRNA gene coding for a 35S RNA precursor in *S. cerevisiae.* For each repeat, termination is controlled by two major sites. About 90% of all transcripts are terminated at the Reb1-dependent terminator, which is located ~93 nucleotides downstream of the rRNA-coding sequence (Lang & Reeder, 1993; Reeder & Lang, 1997). There, the Reb1-protein binds 3' of a T-rich sequence element that is coding for the 3'-terminal 10-12 nt of the transcript. This "roadblock" pauses polymerase. By a mechanism that probably involves transcript slippage facilitated by the weak, T-rich hybrid termination occurs (Reeder & Lang, 1997). *In vivo* and *in vitro* data suggests that an additional factor is required for the release (Jansa & Grummt, 1999; Tschochner & Milkereit, 1997). Furthermore, the Pol I subunit A12.2, which is a homolog of Pol II subunit Rpb9, seems to be involved in termination at this site, since a deletion of the gene leads to increased read-through at the Reb1-dependent site (Prescott et al, 2004). Recently it was shown, that A12.2 is required for the intrinsic 3'-RNA cleavage activity of Pol I (Kuhn et al, 2007). Polymerases that fail to terminate at the Reb1-dependent site can be stopped at the "fail-safe" terminator that is located ~250 nt downstream of the rRNA coding sequence (Reeder et al, 1999).

The least regulated form of termination in eukaryotic transcription is the mechanism that is employed by Pol III, as it seems to be largely factor-independent (Geiduschek & Kassavetis, 2001). Short runs of T-residues seem to be enough to elicit termination, the effciency of termination is influenced by flanking sequence and increases with the length of the T-run (Cozzarelli et al, 1983). This involves transcriptional pausing at these sites. It was anticipated that the intrinsic cleavage activity of Pol III, mediated by the subunit C11, is needed for termination (Chedin et al, 1998). However, recent studies show, that rather the subunits C37 and C53 are involved in the recognition of terminator elements, but the cleavage activity is not involved (Landrieux et al, 2006).

The termination mechanism that is still least understood, is the one used by Pol II. A special feature of termination of protein coding genes that are transcribed into mRNA by RNA polymerase II (Pol II), is that this enzyme does not stop transcription at a specific position at the end of a gene like Pol I and Pol III. Rather, the site for termination seems to be random, sometimes up to 1kb downstream of the poly(A) site where the nascent transcript is cleaved and uncoupled from the transcription machinery by factors that are recruited to the Ser_2-phosphorylated CTD of polymerase II. However, it is well established that termination is dependent on the presence of a functional poly(A)-signal and coupled to RNA-processing events (Buratowski, 2005; Proudfoot, 1989). Furthermore, it is clear that Pol II employs different mechanisms for termination of mRNA-transcription and the termination at genes coding for snRNAs (small nuclear RNAs), snoRNAs (small nucleolar RNAs) and CUTs (cryptic unstable transcripts), which are not polyadenylated. Surprisingly, Rat1 and polyadenylation factors localize to these genes, but mutations that disrupt poly(A) site

cleavage or Rat1 activity do not impair termination, which seems to be mediated by Nrd1, Sen1 (a DNA-RNA-helicase) and Ssu72 (Kim et al, 2006; Lykke-Andersen & Jensen, 2007). Two models try to explain termination of Pol II transcribing protein-coding genes, the "torpedo-model" and the "allosteric model". Both are introduced in detail in sections 4.1.1 and 4.1.2.

2 Materials and common methods

2.1 Materials

2.1.1 Bacterial and yeast strains

Table 2: *E. coli* strains

Strain	Description	Source
XL-1 Blue	rec1A; endA1; gyrA96; thi-1; hsdR17; supE44; relA1; lac[F' proAB lacIqZΔM15Tn10(Tet^r)]	Stratagene
BL21-CodonPlus (DE3)RIL	B; F-; ompT; hsdS(rB, mB); dcm+; Tetr; gal λ(DE3); endA; Hte [argU, ileY, leuW, Cam^r]	Stratagene
Rosetta B834	*E.coli* (DE3) hsd metB	Novagen
E. coli GM2163	F- dam-13::Tn9 (Cam^r) dcm-6 hsdR2 (rk-mk+) leuB6 hisG4 thi-1 araC14 lacY1 galK2 galT22 xylA5 mtl-1 rpsL136 (Str^r) fhuA31tsx-78 glnV44 mcrA mcrB1	Fermentas

Table 3: *S. cerevisiae* strains

Strain	Description	Source
S288C	MATα; SUC2; gal2; mal; mel; flo1; flo8-1; hap1; ho; bio1; bio6	Euroscarf
FY119 Spt6ΔC (isogenic to S288C)	MATα; his4-912; lys2-128; leu2-1; ura3-52; trp1-63	Youdell et al., (2008)
S288C Spt6-TAP	MATα; SUC2; gal2; mal; mel; flo1; flo8-1; hap1; ho; bio1; bio6; URA3	This study
FY119 Spt6ΔC TAP (isogenic to S288C)	MATα; his4-912; lys2-128; leu2-1; ura3-52; trp1-63	This study

2.1.2 Plasmids, oligonucleotides and peptides

Table 4: Plasmids

Plasmid	Description	Source
pET21b(+)	T7; T7-Tag, His-Tag 3' of MCS; lacI; pBR 322 origin; f1 origin; bla coding sequence, Ap^r	Novagen
pET24b(+)	T7; T7-Tag; His-Tag 3' of MCS; lacI; pBR 322 origin; f1 origin; Kan^r	Novagen
pET28b(+)	T7; T7-Tag; His-Tag 5' and 3' of MCS; lacI; pBR322 origin; f1 origin; Kan^r	Novagen
pBS1539	K.I. URA3, Amp^r, C-terminal, TEV cleavage site	Euroscarf

Table 5: Oligonucleotides

Name	Sequence	Source
Sc115	GGAGGAGGACATATGAATCATCCTTACTATTTCC	ThermoFisher
Sc116	GGAGGAGGAGCGGCCGCACCGCTTTTGAATTTTTCAC	ThermoFisher
Sc117	GGAGGAGGAGCGGCCGCTCACAAGAGCCTTACCTTGTTTTG	ThermoFisher
Sc118	GGAGGAGGACATATGCATCGTGTTATCAATCATC	ThermoFisher
Sc119	GGAGGAGGAGCGGCCGCGCCTTGGAAGTACAGGTTCTCTTTTTCACTAGATGTCATTTC	ThermoFisher
Sc128	GGAGGAGGAGCGGCCGCTCATTTGAATTTTTCACTAGATG	ThermoFisher
Sc129	GGAGGAGGAGCGGCCGCTCATGTCATTTCATTCAAGAGCC	ThermoFisher
Sc132	GGAGGAGGACATATGGTTATCAATCATCCTTAC	ThermoFisher
Sc133	GGAGGAGGACATATGCATCCTTACTATTTCCCTTTC	ThermoFisher
Sc134	GGAGGAGGAGCGGCCGCTCATTCATTCAAGAGCCTTACCTTG	ThermoFisher
Sc135	GTCTTGATTGTCGATTTAGACCAG	ThermoFisher
Sc136	CTGGTCTAAATCGACAATCAAGAC	ThermoFisher
Sc137	GTCTTGATTGTCTTAGACCAGATC	ThermoFisher
Sc138	GATCTGGTCTAAGACAATCAAGAC	ThermoFisher
Sp118	GGAGGAGGACATATGGCTCGCGTAATTAAGCACCCG	ThermoFisher
Sp129	GGAGGAGGAGCGGCCGCTCATTCATCTATTTTTTTAGCAATGGC	ThermoFisher
Mm118	GGAGGAGGACATATGAAGCGAGTGATTGCACACCCG	ThermoFisher
Mm129	GGAGGAGGAGCGGCCGCTCAGTCTCGAGCAAAGGAAGCCATG	ThermoFisher
Hs118	GGAGGAGGACATATGAAGAGAGTGATCGCACACCC	ThermoFisher
Hs129	GGAGGAGGAGCGGCCGCTCAGTCCCGGGCAAAGGATGC	ThermoFisherC
Cg141	GGATTATATGAGAAGTAAAG	ThermoFisher
Cg142	CTTTACTTCTCATATAATCC	ThermoFisher
Cg143	CATTTGGCTATGACTTGG	ThermoFisher
Cg144	CCAAGTCATAGCCAAATG	ThermoFisher
Cg145	CAAGAAATGGAAAAGGAAAATCC	ThermoFisher
Cg146	GGATTTTCCTTTTCCATTTCTTG	ThermoFisher
Cg147	CTTGCCATGGGTAAAGTCTTGG	ThermoFisher
Cg148	CCAAGACTTTACCCATGGCAAG	ThermoFisher
Cg149	GGAGGAGGAGCGGCCGCCTCGTTCAGTAGTCTAATCATGTTTTG	ThermoFisher
CgAfw	GGATTATATGAGAAGTAAAG	ThermoFisher
CgArv	CTTTACTTCTCATATAATCC	ThermoFisher
CgBfw	CATTTGGCTATGACTTGG	ThermoFisher
CgBrv	CCAAGTCATAGCCAAATG	ThermoFisher
CgCfw	CAAGAAATGGAAAAGGAAAATCC	ThermoFisher
CgCrv	GGATTTTCCTTTTCCATTTCTTG	ThermoFisher
CgDfw	CTTGCCATGGGTAAAGTCTTGG	ThermoFisher
CgDrv	CCAAGACTTTACCCATGGCAAG	ThermoFisher
CgErv	GGAGGAGGAGCGGCCGCCTCGTTCAGTAGTCTAATCATGTTTTG	ThermoFisher
SH2RLfw	TTCGTGATCCTACAGTCTAGCCG	ThermoFisher
SH2RLrv	CGGCTAGACTGTAGGATCACGAA	ThermoFisher
SH2RKfw	TTCGTGATCAAACAGTCTAGCCG	ThermoFisher
SH2RKrv	CGGCTAGACTGTTTGATCACGAA	ThermoFisher
SH2kofw	AGAGAACAATGACATCTAGTGAAAAATTCA	ThermoFisher

Table 5 (continued)

SH2korv	ATGTCATTGTTCTCTTTGCACGGGCTTCT	ThermoFisher
Ctermkofw	GGTAAGGCTCTTGAATGAATAGATGCGTATGTAGTGTCCATTG	ThermoFisher
Ctermkorv	GGACACTACATACGCATCTATTCATTCAAGAGCCTTACCTTG	ThermoFisher
Spt6flTAPfw	CTTCTAAAATCTAACAGTAGTAAGAATAGAATGAACAACTACCGT TCCATGGAAAAGAGAAG	ThermoFisher
Spt6flTAPrv	TAATAATAAAATTAATAATAACAATGGACACTACATACGCATCTA TACGACTCACTATAGGG	ThermoFisher
Spt6dSH2TAPfw	GAGGAGAGGAAATTGATGATGGCAGAAGCCCGTGCAAAGAGA ACATCCATGGAAAAGAGAAG	ThermoFisher
Rat1flfw	GGAGGAGGAGCTAGCATGGGTGTTCCGTCATTTTTCAGATGGC	ThermoFisher
Rat1flrv	GGAGGAGGAGCGGCCGCACGCCTATTTGCTCTTGAATTGTCATA CCG	ThermoFisher
Rai1flfw	GGAGGAGGACATATGGGTGTTAGTGCAAATTTG	ThermoFisher
Rai1flrv	GGAGGAGGAGCGGCCGCTTTCAAAGATTTTCTCCAC	ThermoFisher
Rtt103flfw	GGAGGAGGACATATGCCTTTCTCTTCTGAGCAATTC	ThermoFisher
Rtt103flrv	GGAGGAGGAGCGGCCGCATTTGCAAGCTTACTTAACAAG	ThermoFisher
Rtt103ΔCIDfw	GGAGGAGGACATATGGAGAGCTCACCAGTGGAAGC	ThermoFisher
torpedoCBNT	**B**GGCTACCGACGCTAGGTCAAGGCAGTACTAGTAATGACCAGG CTCAAGTACTTGAGCTTGGAGTCAGTCGACGATGACTGG	Biomers
torpedoCBT	CCAGTCATCGTCGACTGACTCCAAGCTCAAGTACTTGAGCCTGG TCATTACTAGTACTGCCTTGACCTAGCGTCGG	Biomers
torpedoRNA	**P**UAAUCCCAUAUAUAUGCAUAAAGACCAGGC	Biomers
activityRNA	UCCCAUAUAUAUGCAUAAAGACCAGGC	Biomers
RNaseHDNA1	GCCUGGUCUUUAUGCAUAUAUAUGGGA	Metabion
RNaseHDNA2	GCCUGGUCUUUAUGCAU	Metabion
poly(A)1CBNT	**B**CGACGCTAGGTCAAGGCAGTACTAGTAATGACCAGGCTCAACT ACTCAATAAACCCTACACTCCACCATGGGTAGAGTG	Metabion
poly(A)1CBT	CACTCTACCCATGGTGGAGTGTAGGGTTTATTGAGTAGTTGAGC CTGGTCATTACTAGTACTGCCTTGACCTAG	Metabion
poly(A)cCBNT	**B**CGACGCTAGGTCAAGGCAGTACTAGTAATGACCAGGCTCAACT ACTCCTACCACCCTACACTCCACCATGGGTAGAGTG	Metabion
poly(A)cCBT	CACTCTACCCATGGTGGAGTGTAGGGTGGTAGGAGTAGTTGAGC CTGGTCATTACTAGTACTGCCTTGACCTAG	Metabion
poly(A)2CBNT	**B**CGACGCTAGGTCAAGGCAGTACTAGTAATGACCAGGCTCAACTA CTCAATAAACCCTACACTCCACCATGGGTAGAGTG	Metabion
poly(A)2CBT	CACTCTACCCATGGTGGAGTGTAGGGTTTATTGAGTAGTTGAGCC TGGTCATTACTAGTACTGCCTTGACCTAG	Metabion
poly(A)RNA1	**P**UCCCAUAUAUAUGCAUAAAUCAAUAAA	Biomers
poly(A)RNA2	**P**UACAGCGAGUCUAUGAGCAUCAAUAAA	Biomers
poly(A)xtalNT	CAGCTACTTGAGCT	Biomers
poly(A)xtalT	AGCTCAAGTAGCTGCTTTA**BrU**TGCATT	Metabion
poly(A)xtalRNA	UGCAUUUCGCAAUAAA	Biomers

All oligonucleotides are shown in the direction 5' to 3'; **B** = Biotin; **P** = Phosphate; **BrU** = Bromo-dU; in magenta: RNA-oligonucleotides;

Table 6: Synthetic peptides

Name	Sequence	Source
P1	SPSYpSPTS	Anaspec
P2	YpSPTSPSYpSPTSPS	Coring
P3	Fluo-(Linker)-SYpSPTSPSYpSPTSPS	Coring

pS = Phosphoserine; Fluo = Fluorescein; (Linker) = α-aminocaproic acid

2.1.3 Media and supplements

Table 7:Growth media

Media	Description
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)
ZY	1% (w/v) tryptone; 0.5% (w/v) yeast extract
YPD	2% (w/v) peptone; 2% (w/v) glucose; 1.5% (w/v) yeast extract (+1.8% (w/v) agar for selective media plates
SOB	2% (w/v) tryptone; 0.5% (w/v) yeast extract; 8.55 mM NaCl; 2.5 mM KCl; 10 mM $MgCl_2$;
SOC	see SOB + 20 mM glucose (before use)
Minimal medium	7.5 mM $(NH_4)_2SO_4$; 8.5 mM NaCl; 55 mM KH_2PO_4; 100 mM K_2HPO_4; 1mM $MgSO_4$; 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

Table 8: Supplements

Supplement	Description	Applied concentration
Ampicillin	Antibiotic	100 µg/ml for E.coli culture
Kanamycin	Antibiotic	30 µg/ml for E.coli culture
Chloramphenicol	Antibiotic	50 µg/ml for E.coli culture
IPTG	Isopropyl-β-D-thiogalactopyranosid	0.5 mM

2.1.4 Buffers and solutions

Table 9: General buffers, dyes and solutions

Name	Description	Method
1x Bradford dye	1:5 dilution of Bradford concentrate (BioRad)	Protein concentration
4x stacking gel buffer	0.5M Tris; 0.4% (w/v) SDS; pH 6.8 at 25°C	SDS-PAGE
4x separation gel buffer	3 M Tris; 0.4% (w/v) SDS; pH 8.9 at 25°C	SDS-PAGE
electrophoresis buffer	25 mM Tris; 0.1% (w/v) SDS; 250 mM glycine	SDS-PAGE

Table 9 (continued)

5x SDS sample buffer	250 mM Tris/HCl pH 7.0 at 25°C; 50% (v/v) glycerol; 0.5% (w/v) bromophenol blue; 7.5% (w/v) SDS; 12.5% (w/v) β-mercaptoethanol	SDS-PAGE
Gel staining solution	50% (v/v) ethanol; 7% (v/v) acetic acid; 0.125% (w/v) Coomassie Brilliant Blue R-250	Coomassie staining
Gel destaining solution	5% (v/v) ethanol; 7.5% (v/v) acetic acid;	Coomassie staining
TBE	8.9 mM Tris; 8.9 mM boric acid; 2 mM EDTA (pH 8.0, 25°C)	Agarose gel electrophoresis
TE	10 mM Tris pH 7.4; 1 mM EDTA	nucleic acids
10 x TBS	500 mM Tris/HCl pH 7.5; 1.5 M NaCl	ChIP
1 x PBS	2 mM KH_2PO_4; 4 mM Na_2HPO_4; 140 mM NaCl; 3 mM KCl, pH 7.4 @ 25°C	diverse
6x Loading dye (Fermentas)	1.5 g/l bromophenol blue; 1.5 g/l xylene cyanol; 50% (v/v) glycerol	Agarose gel electrophoresis
2x Urea loadig dye	20% (v/v) 10x TBE; 8 M urea; 0.03% (w/v) bromophenol blue; 0.03% (w/v) xylene cyanol FF	denaturing RNA-PAGE
Blotting buffer	10% (v/v) methanol in ddH2O	Edman sequencing
Swelling buffer	200 mM Tris/HCl pH 8.5 at 25°C; 2% (w/v) SDS	Edman sequencing
TFB-1	30 mM KOAc; 50 mM $MnCl_2$; 100 mM RbCl; 10 mM $CaCl_2$; 15% (v/v) glycerol; pH 5.8 at 25°C	chemically competent cells
TFB-2	10 mM MOPS pH 7.0 at 25°C; 10 mM RbCl; 75 mM $CaCl_2$; 15% (v/v) glycerol	chemically competent cells
20x NPS	0.5 M $(NH_4)_2SO_4$; 1 M KH_2PO_4; 1 M Na_2HPO_4	autoinducing protein expression
50x 5052	25% (w/v) glycerol; 140 mM glucose; 300 mM α-lactose	autoinducing protein expression
100x PI	0.028 mg/ml Leupeptin; 0.137 mg/ml Pepstatin A; 0,017 mg/ml PMSF; 0.33 mg/ml benzamidine; in 100% EtOH p.a.	protease inhibitor mix
beads blocking buffer	50 mM Tris/HCl pH 8.0 at 25°C; 150 mM NaCl; 2 mM EDTA pH8.0; 0.1% (w/v) triton X-100; 5% (w/v) glycerol; 0.5% (w/v) BSA; 200 µg/ml insulin; 0.1 mg/ml heparin; 0.5 mM DTT	bead assays
beads breaking buffer	50 mM Tris/HCl pH 8.0 at 25°C; 150 mM NaCl; 0.1% (w/v) triton X-100; 5% (w/v) glycerol; 0.5 mM DTT	bead assays
TELit	10 mM Tris/HCl pH 8.0; 1 mM EDTA; 155 mM lithium acetate	yeast transformation
LitPEG	40% (w/v) polyethylen glycol 3350 in TELit	yeast transformation
LitSorb	100 mM D-sorbitol in TELit	yeast transformation

Table 10: Spt6 SH2 domain purification buffers

Name	Description
Spt6 cell resuspension buffer	50 mM Tris/HCl pH 8.0 at 4°C; 1M NaCl; 10 mM Imidazole; 10 mM β-mercaptoethanol; 10% (v/v) glycerol
Spt6 cell lysis buffer	50 mM Tris/HCl pH 8.0 at 4°C; 1M NaCl; 10 mM Imidazole; 10 mM β-mercaptoethanol; PI
Spt6 IMAC elution buffer	50 mM Tris/HCl pH 8.0 at 4°C; 200 mM NaCl; 10 mM Imidazole; 10 mM β-mercaptoethanol
Spt6 anion exchange buffer A	50 mM Tris/HCl pH 8.0 at 4°C; 100 mM NaCl; 5 mM DTT

Table 10 (continued)

Spt6 anion exchange buffer B	50 mM Tris/HCl pH 8.0 at 4°C; 2 M NaCl; 5 mM DTT
Spt6 *C. glabrata* size exclusion buffer	50 mM Tris/HCl pH 8.0 at 4°C; 200 mM NaCl; 5 mM DTT
Spt6 *S. cerevisiae/ Homo sapiens* size exclusion buffer	50 mM Tris/HCl pH 8.0 at 4°C; 100 mM NaCl; 5 mM DTT
Spt6 *S. pombe* size exclusion buffer	30 mM bicine pH 9.0 at 4°C; 300 mM NaCl; 5 mM DTT
SH2 low salt buffer	50 mM Tris pH 8.0 at 20°C, 20 mM NaCl, 5 mM DTT

Table 11: Rat1, Rai1, Rtt103 purification and reaction buffers

Name	Description
Rat1 cell resuspension buffer	50 mM Tris/HCl pH 8.0 at 25°C; 500 mM NaCl; 10 mM imidazole; 10 mM β-mercaptoethanol; 5% (v/v) glycerol;
Rat1 cell lysis buffer	same as Rat1 cell resuspension buffer + PI
Rat1 IMAC buffer	50 mM Tris/HCl pH 8.0 at 4°C; 100 mM NaCl; 10 mM imidazole;10 mM β-mercaptoethanol
Rat1 heparin buffer A	50 mM Tris/HCl pH 8.0 at 4°C; 100 mM NaCl; 5 mM DTT
Rat1 heparin buffer B	50 mM Tris/HCl pH 8.0 at 4°C; 2 M NaCl; 5 mM DTT
2x Rat1 size exclusion buffer (RNAse free)	50 mM Tris/HCl pH 8.0 at 4°C; 200 mM NaCl; 2 mM MgCl$_2$; 2 mM DTT
Rat1 storage buffer (RNAse free)	25 mM Tris/HCl pH 8.0 at 4°C; 100 mM NaCl; 1 mM MgCl$_2$, 1 mM DTT; 10% (v/v) glycerol;
Rat1/Rai1 cell resuspension buffer	50 mM Tris/HCl pH 8.0 at 25°C; 150 mM NaCl; 10 mM imidazole; 10 mM β-mercaptoethanol; 5% (v/v) glycerol
Rat1/Rai1 cell lysis buffer	same as Rat1/Rai1 cell resuspension buffer + PI
Rat1/Rai1 IMAC buffer	50 mM Tris/HCl pH 8.0 at 25°C; 100 mM NaCl; 10 mM imidazole; 10 mM β-mercaptoethanol
2x Rat1/Rai1 size exclusion buffer (RNAse free)	same as 2x Rat1 size exclusion buffer
Rat1/Rai1 storage Buffer (RNAse free)	same as Rat1 storage buffer
Rtt103 cell resuspension buffer	same as Rat1 cell resuspension buffer
Rtt103 cell lysis buffer	same as Rat1 cell lysis buffer
Rtt103 IMAC buffer	50 mM Tris/HCl pH 8.0 at 4°C; 150 mM NaCl; 10 mM imidazole, 10 mM β-mercaptoethanol
2x Rat1/Rai1/Rtt103 size exclusion buffer	same as 2x Rat1 size exclusion buffer
10 xRat1 reaction buffer	200 mM Tris/HCl pH 8.0 at 30°C; 1.5M NaCl; 20 mM MgCl$_2$; 10 mM DTT
Rat 1 wash buffer	20 mM Tris/HCl pH 8.0 at 30°C; 500 mM NaCl; 2 mM MgCl$_2$; 1 mM DTT

Table 12: Polymerase II purification buffers

Name	Description
TEZ0	10% (v/v) 10x TEZ0; 1 mM DTT; PI
10x TEZ0	500 mM Tris/HCl pH 7.5 at 20°C; 10 mM EDTA; 100 µM $ZnCl_2$
3x Pol II freezing buffer	150 mM Tris/HCl pH 7.9 at 4°C; 3 mM EDTA; 30 µM $ZnCl_2$; 30 % (v/v) glycerol; 3 % (v/v) DMSO; 30 mM DTT, PI
HSB150	50 mM Tris/HCl pH 7.9 at 4°C; 150 mM KCl; 1 mM EDTA; 10 µM $ZnCl_2$; 10% (v/v) glycerol; 10 mM DTT, PI
HSB600	50 mM Tris/HCl pH 7.9 at 4°C; 600 mM KCl; 1 mM EDTA; 10 µM $ZnCl_2$; 10% (v/v) glycerol; 10 mM DTT, PI
Pol II buffer	5 mM HEPES pH 7.25 at 20°C; 40 mM $(NH_4)_2SO_4$; 10 µM $ZnCl_2$; 10 mM DTT
TEZ250	10% (v/v) 10x TEZ0,
TEZ500	10% (v/v) 10x TEZ0; 500 mM $(NH_4)_2SO_4$; 1 mM DTT; PI
TEZ500 + glyerol	same as TEZ500 + 50% (v/v) glycerol; PI

Table 13: Crystallization buffers

Name	Description
Spt6 SH2 seeding buffer	50 mM MES pH 6.5 at 20°C; 200-400 mM Mg acetate; 10-20% (w/v) PEG 3350; 5 mM TCEP
Spt6 SH2-1	100 mM HEPES pH 7.0 at 20°C; 1 M Succinic acid; 1% (w/v) PEG 2000 MME; 5 mM TCEP
Spt6 SH2-2	50 mM bicine pH 8.0 at 20°C; 4.3 M NaCl; 5 mM TCEP
Spt6 SH2 desalting buffer	30 mM bicine pH 8,0; 300 mM NaCl; 50% (w/v) PEG 2000; 5 mM TCEP
Pol II + poly(A)	50 mM HEPES pH 7,0 at 20°C; 3,5% (w/v) PEG 6000; 200 mM ammonium acetate, 5mM TCEP

Table 14: Chromatin immunoprecipitation buffers

Name	Description
FA lysis buffer	50 mM HEPES pH 7.5; 150 mM or 500 mM NaCl; 1 mM EDTA; 1% Triton X-100; 0.1% Na-deoxycholate; 0.1% SDS
ChIP wash buffer	10 mM Tris/HCl pH 8.0; 250 mM LiCl; 1 mM EDTA; 0.5% Nonidet P-40; 0.5% Na-deoxycholate
ChIP elution buffer	50 mM Tris/HCl pH 7.5; 10 mM EDTA; 1% SDS

2.2 Common methods

2.2.1 Molecular cloning

Polymerase Chain Reaction (PCR) primers were designed by using an overhang of several nucleotides (usually 5'-aggaggagg-3') at the 5' end, followed by the restriction side and 20 to 25 nt complementary to the sequence of the gene of interest (2.1.1, Table 5). PCR reactions were carried out with Herculase or Herculase II polymerases (both stratagene) in a volume of 50 µl together with the respective buffer, 100 µM of dNTP mix, 0,5 µM of each primer and variable DMSO concentrations, usually 1% (v/v). About 100 ng genomic DNA or cDNA of the target organism was usually used as a template. In cases, where the gene of interest was allready cloned, the same amount of the specific vector was used.

For the introduction of point mutations and loop-deletions, the overlap extension method was used. Here, two overlapping PCR-products are produced with primers carrying the desired mutation. In a second PCR reaction these products were used as a template to produce the gene of interest containing the mutation.

Thermocycling programs were adjusted to the specific needs of the individual reactions in terms of annealing temperature and elongation times and usually contained 30 cycles (Biometra T3000 Thermocycler). PCR products were visualized by 1% agarose gel electrophoresis and staining with ethidiumbromide. Purification of the DNA was carried out with the QIAquick gel extraction protocol (Quiagen).

Enzymatic restriction cleavage. DNA was digested using restriction endonucleases (New England Biolabs and Fermentas) as recommended by the producer. Cleaved PCR products were purified using the QIAquick PCR purification protocol, cleaved plasmids by the QIAquick gel extraction protocol (both Quiagen).

Ligation of digested DNA into linearized vectors was carried out for 1 hour at room temperature in a volume of 20 µl using T4 DNA ligase (Fermentas) and its corresponding buffer. Concentrations of DNA components were varied depending on the different reactions. Usually a 5- to 10-fold excess of insert, relative to linearized vector was used.

Transformation of *E. coli* and isolation of plasmid DNA. Chemically competent *E. coli* XL-1 blue cells (see 2.1.1, Table 2 and 2.2.2) were transformed with DNA by a heat shock protocol.

3-5 µl of the ligation reaction or 1 µl of plasmid DNA were added to a 50 µl aliquot of competent cells and incubated for 5 minutes on ice. Cells were then heated to 42°C for 30 seconds and put back on ice for 2 minutes. The transformed cells were recovered by incubation at 37°C in 700 µl LB for one hour. After sedimentation (30 sec, 14000 rpm in a

microcentrifuge) the cells were resuspended in 100 µl LB medium and plated on LB-Agar plates containing the corresponding antibiotic for selection of transformed cells.

E. coli cells from a 5 ml overnight culture, grown from a single clone were used for the preparation of plasmid DNA using the QIAquick Miniprep Kit protocol (Qiagen). Isolated plasmids were verified first by restriction analysis, second by DNA sequencing.

2.2.2 Preparation of competent cells

Chemically competent cells were prepared by inoculation of 200 ml LB with 5 ml of an overnight culture of the desired strain. Cells were grown at 37°C to an OD_{600} of 0.4-0.55 and incubated on ice for 10 minutes. All following steps were carried out at 4°C. After sedimentation at 1000 g for 10 minutes, the pellet was washed with 50 ml TFB-1 (see 2.1.4, Table 9). After a second centrifugation step, the pellet was resuspended in 4 ml of TFB-2 (see 2.1.4, Table 9), aliquoted and frozen in liquid nitrogen. Competent cells were stored at –80°C.

Competent yeast cells were prepared by inoculation of YPD medium (see 2.1.3, Table 7) with the appropriate strain to an OD_{600} of 0.2. Cells were grown to OD_{600} of 0.5-0.7 at 30°C, then transferred to Falcon tubes. After sedimentation at 5000 g for 5 min at room temperature, cells were washed with 0.5 volumes of sterile water. After a subsequent centrifugation step, cells were washed with 0.1 volumes of LitSorb (see 2.1.4, Table 9), then resuspended in 360 µl of LitSorb per 50 ml of the inital culture. 40 µl (per 50 ml culture) of pre-heated (10 minutes at 95°C, then put on ice) salmon sperm DNA was added. Aliquots of 50 µl were put on –80°C without freezing in liquid nitrogen.

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

Proteins in this work were expressed recombinantly in *E. coli* BL21-Codon plus (DE3)RIL cells (see 2.1.1, Table 2), where not stated otherwise. For that, plasmids containing genes for the desired protein variants were used for the transformation of the cells (see 2.2.1). Depending on the protein to be expressed, different protocols were used:

For **IPTG-induced protein expression**, an expression culture of the desired volume containing LB-medium (see 2.1.3, Table 7) and the antibiotic corresponding to the resistance cassette of the vector was inoculated from an overnight-culture of the transformed cells in a 1:100 diluton. Cell were grown to an OD_{600} of 0.6-0.9, then put on ice. Protein expression was induced by the addition of 0.5 mM IPTG and was carried out at 18°C overnight.

Autoinducing protein expression medium was used, when the protein showed low solubility and/or low yield in IPTG-induced protein expression. This was carried out as described (Studier, 2005). For that, the desired volume of ZY-medium was supplemented with NPS, 5052, 1 mM MgSO4 and the appropriate antibiotic for selection of the plasmid (2.1.2, Table 4). The expression culture was inoculated from an overnight culture of the transformed cells at a dilution of 1:100. Cells were grown to an OD_{600} of 0.6, then the temperature was shifted to 18°C. Protein expression was carried out overnight.

Cells were harvested by centrifugation at 4400 g (SLC-6000 rotor) for 30 minutes at 4°C, resuspended in the corresponding resuspension buffer (see 2.1.4, Tables 10 and 11) and frozen in liquid nitrogen. Cell pellets were stored at –80°C.

For **selenomethionine incorporation**, the desired expression plasmid was transformed into the methionine auxotroph *E. coli* strain B834 (DE3) (see 2.1.1, Table 2). Cells were grown in LB-medium (see 2.1.3, Table 7) supplemented with the appropriate antibiotic at 37°C to an OD_{600} of 0.5, centrifuged and resuspended in the same amount of minimal medium (see 2.1.3, Table 7) supplemented with selenomethionine and antibiotics. Cells were grown at 37°C until the OD_{600} increased by 0.2, then cultures were shifted to 18°C and protein expression was induced by the addition of 0.5 mM IPTG. Protein was expressed overnight.

2.2.4 Measurement of protein concentration

Protein concentrations were usually determined by the **Bradford protein assay** (Bradford, 1976). The assay was performed according to the instructions of the manufacturer of the Dye reagent (Biorad) . A calibration curve was generated for each new batch of dye reagent using bovine serum albumin (Fraktion V, Roth).

2.2.5 Protein purification

Purification of all different recombinantly expressed protein variants included the same basic steps with slight variations, due to the fact that all variants contained a hexahistidine-tag. The variations in the different protocols for the specific proteins are described in the respective chapters (see 3.2.2 and 4.2.2). Steps that were carried out for all of the purifications the same way, were:

Cell lysis. Cell pellets were resuspended in 50 ml of the respective cell lysis buffer (see 2.1.4, Tables 10 and 11) and sonicated for 20 minutes. The resulting cell extract was cleared by centrifugation (2x 20 minutes at 24000 g in a SA-300 rotor). The pellet was resuspended

in an equal amount of 6 M urea relative to the cell extract and analyzed by SDS-PAGE (see 2.2.7) for the content of expressed protein in the insoluble fraction.

Affinity chromatography. Cleared extracts were applied twice to 1-2 ml of Nickel-NTA-Agarose (Qiagen) in a column. The amount of resin used differed according to the expression level of the different protein variants. Nickel-NTA-Agarose was washed with 10 column volumes of ddH$_2$O and equilibrated with 10 column volumes of the corresponding cell lysis buffer (see 2.1.4, Tables 10 and 11) prior to usage. After binding of the protein, the resin was washed with 10 column volumes of the corresponding cell lysis buffer.

2.2.6 Limited proteolysis

To delineate flexible regions in proteins that might interfere with crystallization, protein variants were probed with proteolytic enzymes: for trypsin and chymotrypsin treatment of purified protein samples, 100 µl of a protein solution with the concentration of 1 mg/ml was mixed with 1 µg of the corresponding protease. The reaction was carried out in the gel filtration buffers of the respective proteins (2.1.4, Tables 10 and 11), supplemented with 1 mM CaCl$_2$. The reaction mixture was incubated at 37°C. Aliquots of 10 µl were taken at different timepoints (usually after 1, 3, 5, 10, 30 and 60 minutes) and the reaction was stopped immediately by the addition of 5 x SDS sample buffer (see 2.1.4, Table 9) and incubation at 95°C for 5 minutes.

For Subtilisin and Proteinase K treatment 1 µl of a dilution of the respective protease (1 µg, 100 ng, 10 ng, 1 ng) was added to 50 µl of a protein solution with the concentration of 1 mg/ml. Samples were incubated for 1 hour on ice, then the reactions were stopped as described for trypsin and chymotrysin.

All samples were analyzed by SDS-PAGE (see 2.2.7). Bands of interest were cut out of the gel and prepared for edman sequencing (see 2.2.8)

2.2.7 Electrophoresis

Electrophoretic separation of DNA was carried out in horizontal 1x TBE (2.1.4, Table 9) agarose gels containing ethidium bromide (0.7 µg/ml). Agarose concentrations varied between 1% to 2%, depending on the size of the DNA-molecules to seperate. Separation was carried out in PerfectBlue Gelsystem electrophoresis chambers from Peqlab. Samples were mixed with 6x loading dye (see 2.1.4, Table 9) and DNA was visualized and documented using a ultraviolet transiluminator from INTAS Science Imaging Instruments (λ=366 nm).

Electrophoretic separation of RNA was conducted by the use of denaturing polyacrylamide gels containing 8% acrylamide, 0.42% bisacrylamide and 6 M urea. TBE was used as buffer for the gel, as well as for running buffer. Depending on the need for separation, electrophoresis was carried out in 1.0 mm cassettes for the XCELL Shure Lock™ gel-system (Invitrogen) for small gels, or Sequigen® GT Sequencing Cell (BioRAD) for large gels. Detection of radioactivly labeled RNA molecules was conducted by exposure of the gels to mounted storage phosphor screens (GE Healthcare) of the appropriate size. Exposure time was usually overnight at 4°C, but could be longer or shorter for individual experiments. Read-out of the phosphor screens was done by the use of Typhoon™ or Storm™ Scanners (GE Healthcare). Quantitation of signals from radioactive RNAs was carried out with the ImageQuant Software (GE Healthcare).

Electrophoretic separation of protein was conducted by **SDS-PAGE** with 15%-17% acrylamide gels (with acrylamide:bisacrylamide = 37.5:1; (Laemmli, 1970) in BioRad gel systems. For buffers see section 2.1.4, Table 9. Gels were stained with Coomassie gel staining solution for 20 minutes and destained overnight in gel destaining solution.

2.2.8 Edman sequencing

Proteins analyzed by SDS-PAGE (2.2.7) and stained with Coomassie staining solution (see 2.4.1, Table 9) were carefully excised from the gel and dried in a speed vac. The dried gel piece was rehydrated in 50 µl of swelling buffer (see 2.1.4, Table 9). Afterwards 200 µl of ddH$_2$O was addet to set up a concentration gradient. For transfer, a small piece of polyvinylidene difluoride (PVDF) membrane (Schleicher & Schuell), which was pre-wet in methanol, was added to the tube. Once the solution turned blue, 24 µl of methanol (final concentration of 8-10%) was added as a catalyst. After 1-2 days of incubation at room temperature, when the membrane turned blue, it was washed 5 times with 10% methanol. After drying, the membrane was used for N-terminal sequencing of the protein in a PROCISE 491 sequencer (Applied Biosystems).

3 Structure and requirement of the Spt6 SH2 domain

3.1 Introduction

3.1.1 Structure and function of SH2 domains

Separation of function in proteins is often achieved by folding these molecules into separate domains. In this way, a single polypeptide chain can be seen as an array of modules that assign different functions to the protein. These functions can involve protein-protein interactions, DNA/RNA-binding, binding to phospholipids, catalysis of chemical reactions and many more.

Such a modular nature is characteristic for proteins involved in cellular signal transduction pathways. In receptor tyrosine kinase (RTK) mediated signaling, ligands bind to their specific transmembrane receptors which dimerize upon this interaction. This leads to the activation of the cytoplasmic kinase domain which then either phosphorylates specific tyrosine-residues within the cytoplasmic part of the receptor (autophosphorylation) or on cytoplasmic proteins. This ligand-induced tyrosine-phosphorylation creates high-affinity docking sites for cytoplasmic signaling proteins. Upon binding, these proteins can be phosphorylated and activated, and transduce the signal to other proteins that are part of their specific signaling cascade (Fig. 3 A). The phosphorylated tyrosine residues can be recognized by SH2 domains or PTB-domains (Blaikie et al, 1994) that are thus transmitting the signal of the activated receptor to a cytoplasmic signaling protein, which can contain a variety of other domains with different specificities (Fig. 3 A): PH-domains as well as FYVE-domains interact mainly with phospho-inositides and can target the protein to the cell membrane. SH3 domains and WW domains bind proteins with proline-rich target motifs, wheras PDZ-domains bind specifically to hydrophobic residues in the C-termini of target molecules (reviewed in Schlessinger, 2000). The signaling proteins can be mere adaptors that transmit the signal from the activated receptor to the next protein in the signaling cascade (e.g. Grb2, Nck), or they can possess intrinsic enzymatic activity to modify downstream targets (e.g. Src, PLCγ), e.g. by another phosphorylation reaction.

With the use of this arsenal of domain functions, a cascade of protein interactions can be constructed, that transmits the signal from the ligand-receptor interaction into the interior of the cell.
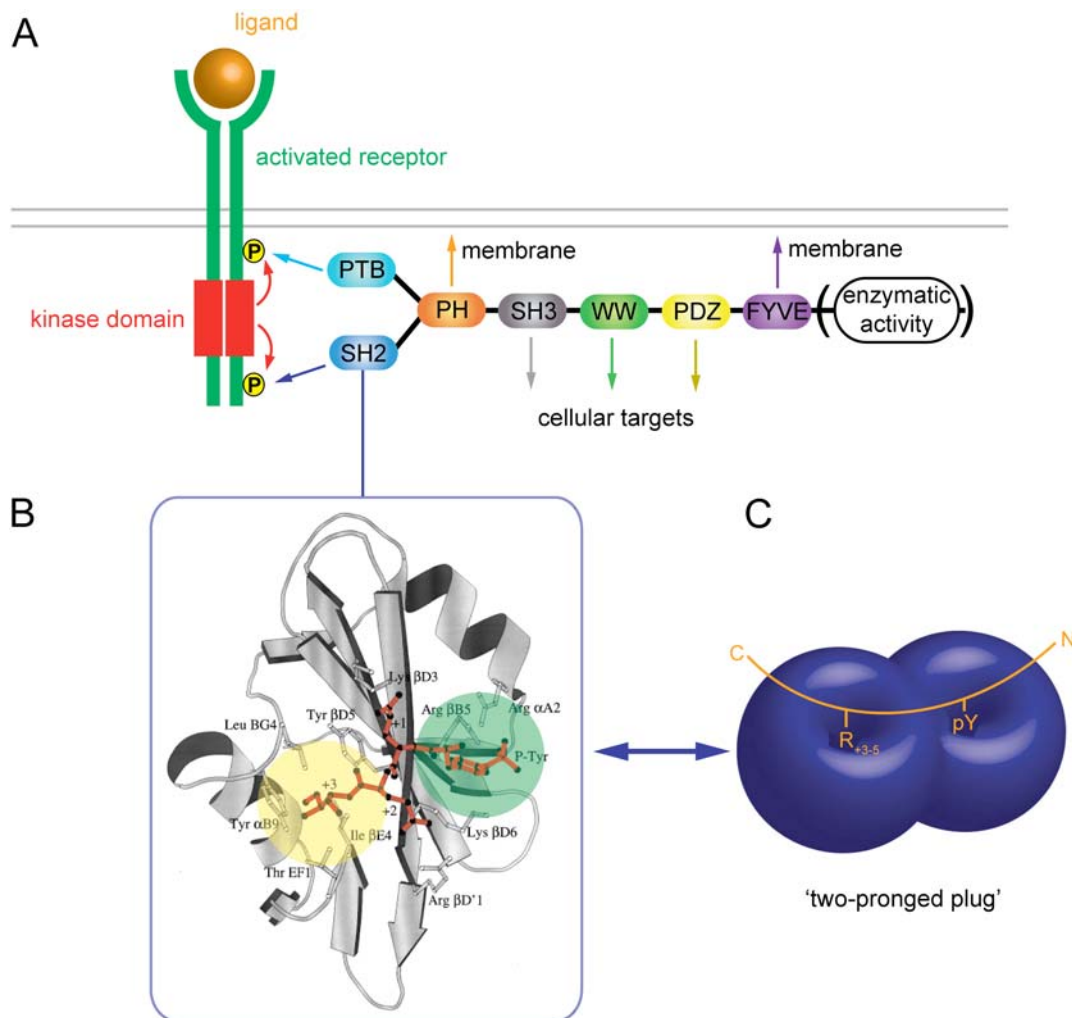
Figure 3: SH2 domains are adaptors in receptor-tyrosine-kinase pathways that bind to phosphorylated tyrosines

(**A**) Protein domains that play a role in cellular signaling pathways. Upon specific binding of a ligand a cell surface receptor is dimerized and activated. The cytoplasmic kinase domain can then autophosphorylate specific tyrosine residues. These phosphates serve as docking sites for cytosolic signaling proteins, that bind via SH2 or PTB domains. The Figure is adapted and modified from Schlessinger (2000).

(**B**) Structure of the Src SH2 domain in complex with its high affinity target peptide pYEEI (Waksman et al, 1993). The phosphotyrosine-binding site is marked green, the pY+3-5 binding-site is marked yellow. The Figure is taken and modified from (Kuriyan & Cowburn, 1997)

(**C**) Schematic representation of the "two-pronged plug" mode of peptide recognition by SH2 domains (reviewed in (Kuriyan & Cowburn, 1997). The SH2 domain is shown in blue, the target peptide in orange. The phosphotyrosine residue is binding to a specific pocket on the surface of the SH2 domain. Specificity of binding is accomplished by interaction of the domain with peptide residues C-terminal of the phospho-tyrosine. A residue at position +3 to +5 C-terminally from the phospho-tyrosine, which inserts into a second pocket on the domain surface, is especially important for this interaction.

SH2 domains are widely used as adaptors in higher eukaryotes. Human cells contain 120 different SH2 domains distributed over 110 proteins (Liu et al, 2006) (see also Fig. 11 A). These domains were originally described in 1986. Insertion mutants in the transforming

protein p130$^{gag-fps}$ of Fujinami sarcoma virus, a cytoplasmic tyrosine-kinase, showed impaired kinase activity *in vivo*. However, when the mutant proteins were expressed in *E. coli*, they showed the same kinase activity as the wild-type. The mutations were located in a highly conserved region of the cytoplasmic tyrosine kinases, immediately N-terminal of the kinase domain. This region was named „Src-homology 2" (SH2) after the cytoplasmic kinase Src, that contains this conserved region (Sadowski et al, 1986). A few years later it was discovered, that SH2 domains bind to phospho-tyrosines on activated receptors, which made the SH2 domain the first targeting module in signaling to be described (Anderson et al, 1990; Kazlauskas et al, 1990; Moran et al, 1990). But these modules are not only used for targeting: the SH2 domain of Src can bind intramolecularly to a phospho-tyrosine in the C-terminal region of the same protein to inhibit its kinase activity (Roussel et al, 1991). Thus, SH2 domains also have a function in the regulation of enzyme activity. In subsequent work it was shown, that the flanking regions of the phospho-tyrosine are important for the specificity of binding (Escobedo et al, 1991; Fantl et al, 1992; Kazlauskas et al, 1992; Ronnstrand et al, 1992; Yoakim et al, 1992) and extensive peptide-library screening could assign specific sequence motifs as targets to various SH2 domains (Songyang et al, 1993).

In 1992, the first structures of SH2 domains emerged: the crystal structure of the Src-domain phosphopeptide complex (Waksman et al, 1992) and the NMR solution structures of the uncomplexed SH2 domains of Abl (Overduin et al, 1992) and of phosphatidylinositol-3-OH kinase (Booker et al, 1992). The structure of the Src-phosphopeptide complex revealed the interaction with the phosphate-group, but not how the domain could identify its specific target peptide (reviewed in Waksman & Kuriyan, 2004). Only one year later, the X-ray structure of the Src SH2 domain in complex with its high affinity target peptide was published (Waksman et al, 1993) revealing the „two-pronged plug" mode of peptide recognition (Fig. 3 C) . This was followed by the elucidation of a large number of SH2 domain structures of numerous signaling factors (reviewed in Kurian & Cowburn, 1997).

The common structure of a SH2 domain contains a central β-sheet, consisting of strands βB-βC-βD (Fig. 3 B). One side of this sheet is flanked by α-helix αA (N-terminal helix), the other side by helix αB (C-terminal helix). Often additional short β-strands add up to the central sheet (βA N-terminal of helix αA and βG C-terminal of helix αB).

In most of the known SH2 ligand structures, the peptide binds in an extended conformation orthogonal to the central β-sheet (Fig. 3 B, reviewed in Kurian & Cowburn, 1997; Yaffe, 2002). The phospho-tyrosine inserts into a pocket with an invariable arginine at its base, that makes contact to the phosphate oxygens together with residues that are part of αA, βB and loop BC (the loop between β-strands βB and βC). This pocked is marked green in Fig. 3 B. The phenyl-ring of the phospho-tyrosine is stabilized by an amino-aromatic interaction with an arginine that is jutting out from helix αA into the binding pocket (Arg αA2, Fig. 3 B; see also Fig. 12 D and E), and in addition by a lysine residue in strand βD (Lys βD6) (Eck et al, 1993; Waksman et al, 1992; Waksman et al, 1993; Yaffe, 2002).

Specificity of binding is accomplished by residues in the loops EF and BG, which are highly variable in SH2 domains. A single mutation in this region can switch the specificity of the Src SH2 domain to that of GRB2 (Marengere et al, 1994). These residues usually form a second pocket, where a peptide residue that is 3-5 positions C-terminal of the phospho-tyrosine is recognized (yellow in Fig. 3 B, reviewed in Kuryan & Cowburn, 1997).

There are exeptions from this binding mode. In GRB2, the +3 binding site is filled with a tryptophane residue that closes the site for interactions with the peptide. In the GRB2-ligand structure, the peptide binds in a bent conformation to the domain (Rahuel et al, 1996).

Binding the phosphate provides about half of the free energy that is needed for the interaction (Bradshaw et al, 1999), the remaining energy is provided by the sum of rather weak interactions with the residues C-terminal of the pTyr described above (Yaffe, 2002). The large contribution of the phosphate to the binding energy provides that the SH2 domain can discriminate between the phosphorylated and unphosphorylated state of the target.

3.1.2 Transcription elongation factor Spt6

The gene encoding Spt6 was originally identified in a genetic screen in yeast as a suppressor of transposon insertion in the promoter region of a reporter gene (Winston et al, 1984). It was later described as an essential, nuclear protein (Clark-Adams & Winston, 1987; Swanson et al, 1990) that is involved in the elongation phase of transcription, like Spt4 and Spt5. Spt4/Spt5 (DSIF in human cells) were identified in the same genetic screen as Spt6. The complex interacts physically with Pol II via Spt5 (Hartzog et al, 1998). Spt4 was characterized as a positive elongation factor (Rondon et al, 2003) but in conjunction with NELF (negative elongation factor), DSIF is responsible for promoter proximal pausing of Pol II (Yamaguchi et al, 1999). Spt6 interacts genetically with Spt4 and Spt5. Spt6 and Spt5 also interact physically, albeit weakly (Swanson & Winston, 1992). Together with Spt5, Spt6 colocalizes with Pol II on actively transcribed genes in yeast (Krogan et al, 2002), in human cells (Endoh et al, 2004) and on *Drosophila* polytene chromosomes (Andrulis et al, 2000; Kaplan et al, 2000).

Mechanistically, Spt6 was found to be important for the maintenance of chromatin structure. It is interacting with histone H3 and has the ability to deposit nucleosomes onto "naked" DNA in a supercoiling assay (Bortvin & Winston, 1996). Furthermore, mutations in Spt6 lead to an altered chromatin structure *in vivo* and can supress a null mutation of the Swi/Snf complex that remodels nucleosome positions at promoters (Bortvin & Winston, 1996). Consistently, recent findings show that the impairment of Spt6-mediated re-assembly of nucleosomes in promoter regions leads to transcription initiation without the need for transcriptional activators (Adkins & Tyler, 2006).

The finding that mutations in Spt6 lead to transcription initiation from cryptic start sites within the coding regions of genes led to the suggestion that Spt6 is involved in the maintenance of proper chromatin structure during elongation (Kaplan et al, 2003). As a consequence, nucleosomes cannot be deposited back onto DNA after the passage of Pol II in a cell where Spt6 is not functional, which leads to higher accessibility of the coding regions of genes.

Another factor that acts together with Spt6 as a histone chaperone is the heterodimeric FACT complex (Spt16/Pob3 in yeast). Whereas intact nucleosomes pose a block to transcription by Pol II, the polymerase can transcribe nucleosomes deficient in one H2A/H2B dimer (Kireeva et al, 2002). Independent studies support the model that FACT removes the H2A/H2B dimer from the nucleosome, thus creating a substrate that can be transcribed by Pol II (Belotserkovskaya et al, 2003; Orphanides et al, 1999). Additionally, FACT also seems to play a role in the maintenance of correct chromatin structure and the suppression of cryptic transcription (Mason & Struhl, 2003). Taken together, Spt6 interacts physically with Pol II, FACT and Spt5 *in vitro* and *in vivo* (Endoh et al, 2004; Krogan et al, 2002; Swanson & Winston, 1992). Besides, an interaction of the homologs of Spt6 and FACT in *Drosophila* with the Paf-complex was reported. There, depletion of Paf1 leads to a significantly decreased recruitment of Spt6 and FACT to the Hsp70 heat shock gene (Adelman et al, 2006) (Fig. 4 A)

Another protein that interacts with Spt6 is Spn1. The essential gene coding for Spn1 was identified in a genetic screen as a suppressor of a mutation in the TATA-binding protein (TBP) that shows a defect in the activation of transcription after binding to the TATA box (Fischbeck et al, 2002). Spn1 was shown to interact directly with Pol II and Spt6 at the promotor of the CYC1 gene, which is regulated post-recruitment of Pol II . Spn1 is a negative regulator of transcription of CYC1 by inhibiting recruitment of the chromatin remodeling complex Swi/Snf. This inhibition is abolished by the interaction of Spn1 with Spt6, which appears at the promotor shortly after induction of CYC1 expression (Zhang et al, 2008). However, the functions of Spn1 and Spt6 do not seem to be generally linked, since an Spn1 mutant that fails to be recruited to the promotor of CYC1 due to a failure of the interaction with Pol II does at the same time not show cryptic transcription at the FLO8 gene as is seen in a Spt6 mutant strain (Kaplan et al, 2003; Zhang et al, 2008). Spn1 is also known as Iws1 and it was further shown that it interacts with the RNA-export factor REF1/Aly (Yra1 in yeast) (Yoh et al, 2007). In yeast, Yra1 interacts with Sub2 and the THO-complex to form the TREX-complex to direct export of mRNP-particles to the cytoplasm (Lei et al, 2001; Strasser & Hurt, 2001). Depletion of Iws1 consequently showed retention of RNA in the nucleus in HeLa cells as well as RNA processing defects. Intriguingly, displacement of the C-terminal SH2 domain of Spt6 from its target - the Pol II C-terminal domain (CTD) phosphorylated at Ser$_2$ residues - by competitive *in vivo* overexpression of the domain alone, showed the same phenotype. This suggests that this interaction is important for proper processing of RNA, at least for the genes tested, probably via the interaction of Spt6 with Iws1 and REF1/Aly (Yra1)

(Yoh et al, 2007). Very recent results from the same lab show another interaction of Iws1 with the HYPB/Setd2 histone methyltransferase, thus directing H3K36 trimethylation (Yoh et al, 2008). Setd2 is interacting with the CTD of Pol II, but in contrast to Spt6 it binds the doubly phosphorylated form (Ser2-Ser5) (Kizer et al, 2005; Li et al, 2005; Vojnic et al, 2006) (Fig. 4 B).
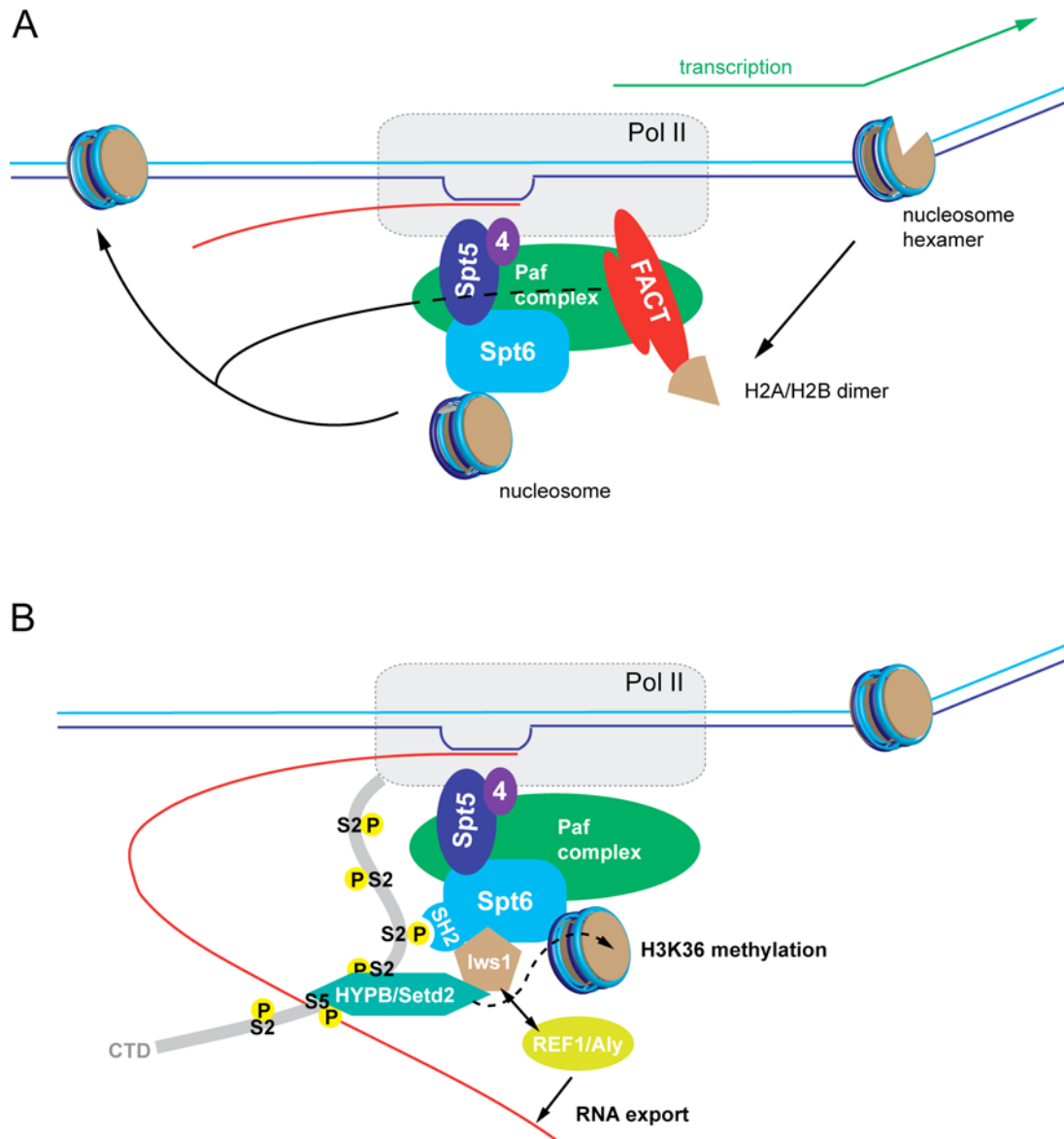
Figure 4: Spt6 is involved in multiple processes during transcription elongation

(**A**) Spt6 is involved in the maintenance of chromatin structure. Histone chaperone FACT removes one H2A/H2B dimer from the nucleosome octamer, thus creating a substrate that can be transcribed by Pol II. Together with Spt6 it re-establishes proper chromatin structure after the passage of Pol II.

(**B**) Spt6 is involved in RNA processing and histone modification. Spt6 binds to Ser_2-phosphorylated CTD and Iws1, which interacts with REF1/Aly and Set2d. REF1/Aly is involved in RNA-export from the nucleus. Set2d is a histone-methylase that interacts with Ser_2-/Ser_5-phosphorylated CTD. Template DNA, non-template DNA and RNA are shown in blue, cyan and red, respectively. Factors in this figure are not drawn to scale. See text for details.

Thus, the interaction between Spt6 and the CTD of Pol II, mediated by a SH2 domain, seems to connect important events like RNA processing and histone modification to transcription elongation.

3.1.3 The SH2 domain of Spt6

Recruitment of Spt6 to the transcription machinery and its interactions with it are complex (see above). Given the importance of Spt6 for transcription elongation as well as for post-transcriptional processes, structural information is highly anticipated to understand this factor on a molecular basis. The C-terminal part of Spt6, that contains an SH2 domain, was shown to interact specifically with the CTD of polymerase II that is phosphorylated at Ser2 residues (Yoh et al, 2007). This makes this particular SH2 domain highly interesting for two reasons:

First, the Spt6 SH2 domain shows a binding specificity that is unusal for SH2 domains. It binds to a Ser_2-phosphorylated target, whereas SH2 domains of signaling factors usually bind to phospho-tyrosines (see section 3.1.1). Although other domains with altered specificity have been characterized (Muller et al, 1992; Pendergast et al, 1991), to our knowledge there is no structural information on such an interaction.

Second, the Spt6 SH2 domain is the only SH2 domain encoded in the yeast genome (Maclennan & Shaw, 1993). Whereas these modules are widely encoded in the genomes of higher eukaryotes, the SH2 domain of Spt6 somehow represents the "minimal equipment" of a simple eukaryotic cell regarding SH2 domains. This domain seems to be an ancestor of the modern SH2 domains, because Spt6 evolved prior to the divergence of the eukaryotic taxa (see also Fig. 11 A). The structure of the Spt6 SH2 domain will consequently give an exciting insight into the evolution of these important domains.

3.1.4 Aim of this work

The C-terminal domain of Pol II cannot be observed in Pol II crystal structures, due to its high flexibility. As a consequence, high resolution structural information can be obtained only in complex with molecules that bind the CTD, as exemplified by the structures of the CTD interacting WW-domain (Verdecia et al, 2000), the CID-domain (CTD-interacting domain) of Pcf11 and Nrd1 (Meinhart & Cramer, 2004; Vasiljeva et al, 2008) and the Set2 SRI-domain (Li et al, 2005; Vojnic et al, 2006). As soon as the interaction of the Spt6 SH2 domain with the CTD was published (Yoh et al, 2007), we started to initiate work on the structure solution of this domain (described in 3.3.1 – 3.3.4) and the analysis of its interaction with the CTD (described in 3.3.7 – 3.3.10). The structure opened way to a deeper analysis of evolutionary relationships between SH2 domains, due to its outstanding position in the development of these domains (described in 3.3.5 and 3.3.6). We were also interested in the influence of the

SH2 domain on overall gene expression in yeast. Since Spt6 is an important regulator of transcription elongation, the disturbance of its direct interaction with Pol II should have a significant effect. This question was tackled in cooperation with Andreas Mayer (3.3.12). In addition, a high resolution occupancy profile of Spt6 on the yeast genome was established in a ChIP-on-chip experiment. From the comparison of this data with the occupancy profile of Rpb3 (Pol II) (described in 3.3.13), we attempted to look for answers on the recruitment of Spt6 to transcribed genes and its overall presence on the genome.

3.2 Specific procedures

3.2.1 Vectors

The following vectors have been created with the methods described in 2.2.1:

Table 15: Vectors containing Spt6 genes

Vector	Source Plasmid (see 2.1.2, Table 4)
Cg118+129	pET28b(+)
Hs118+129	pET28b(+)
Sc118+129	pET28b(+)
Sp118+129	pET28b(+)
SH2RL	pET28b(+)
SH2RK	pET28b(+)
SH2ko	pET28b(+)
Ctermko	pET28b(+)
CgSeMetCD	pET28b(+)
CgSeMetCDE	pET28b(+)

3.2.2 Purification of the Spt6-SH2 domain

Variants of the Spt6 SH2 domain from the organisms *Schizosaccharomyces pombe*, *Candida glabrata* and *Homo sapiens* were purified each from 1 l of IPTG induced expression culture (2.2.3). *S. cerevisiae* variants were purified from 1 l of autoinducing expression culture (2.2.3). After binding to Ni-NTA-Agarose (2.2.5), proteins were eluted by a step gradient of increasing imidazole concentration. For this, 10 ml of Spt6 IMAC elution buffer (see 2.1.4, Table 10) + 20 mM, + 30 mM, + 50 mM, + 100 mM and + 300 mM imidazole were applied to the column. The fractions were analyzed for recombinant protein by SDS-PAGE (see 2.2.7).

Fractions that contained the recombinant protein and showed a sufficient purity for the next step of the purification were pooled. The protein concentration was determined by the Bradford assay (2.2.4) and 1 unit of Thrombin protease (from bovine serum, Sigma) per µg of protein was added to remove the N-terminal Hexahistidine-tag. The sample was then dialyzed against Spt6 IMAC elution buffer (see 2.1.4, Table 10) at 4°C overnight to reduce the imidazole concentration. Subsequently, the sample was applied again to the Ni-NTA-column. The flowthrough, which contained recombinant protein now lacking the His-tag, was collected. The flowthrough from the Ni-NTA-column was applied to a MonoQ 10/100 GL anion exchange column (GE Healthcare). The column was equilibrated with Spt6 anion exchange buffer A (see 2.4.1, Table 10), bound proteins were eluted with a linear gradient of 20 column volumes from 100 mM to 1 M NaCl. Peak fractions and flowthrough of this chromatography step were checked by SDS-PAGE (2.2.7) for the presence of recombinant protein.

Fractions containing the corresponding Spt6 SH2 variant were pooled and concentrated (Amicon Ultra centrifugal filter devices, cutoff 10k, Millipore). Afterwards samples were applied to a Superose 12 10/300 GL column (GE Healthcare), pre-equilibrated with the respective Spt6 size exclusion buffer (see 2.1.4, Table 10). Peak fractions were pooled and concentrated to 10-20 mg/ml for crystallization.

The purification protocol was similar for *C. glabrata* Spt6 SH2 variants containing point mutations for selenomethionine labeling.

3.2.3 Design of selenomethionine mutants of the Spt6 SH2 domain of *C. glabrata*

Selection of positions for methionine mutants in the Spt6 SH2 domain was based on two criteria. First, the position should contain a conserved hydrophobic amino acid residue, preferably a leucine because of its comparable size to a methionine. Second, the probability of such a residue to be part of the hydrophobic core of the domain should be high. Taken together, these criteria should increase the probability to find sites for mutations, which do not disturb the overall structure of the domain and hence its ability to crystallize. A multiple sequence alignment of Spt6 SH2 domains from diverse organisms revealed the desired conserved residues. These were mapped onto the structure of the SH2 domain of Grb2, which was initially used in a structure based alignment with the *S. cerevisiae* Spt6 SH2 domain (see Fig. 5) to check the possible position of these residues in the Spt6 SH2 domain fold. Figure 5 summarizes the selected mutations. Mutation E could not be mapped onto the Grb2 structure because of low sequence conservation in this area of the domain (see Fig. 5 structure based alignment), but was chosen because of a highly conserved methionine residue in the Spt6 molecules of the other organisms. Fig. 5 B shows that mutations A, B and D have a high probability to be part of the hydrophobic core, whereas the probability of mutation C is lower. Nevertheless, mutation C was used because of its high conservation.

3.2.4 Crystallization

Initial screens of crystallization conditions using different protein variants from different organisms were set up with a Hydra II crystallization robot (Matrix). 500 nl drops were set up in Corning 96 well sitting drop crystallization plates with the robot, additionally adding fresh reducing agent (Tris(2-carboxyethyl) phosphine hydrochloride, TCEP) at a concentration of 5 mM. Commercial screens used for initial setups were: Index, Classic screen, Natrix, PEG/Ion (all Hampton), pH-clear, anions suite, cations suite, classic suite (all Qiagen), JB Screen Classic HTS I S and JB Screen Classic HTS II S (both Jena Biosciences). Plates were incubated at 20°C and checked regularly after several days.
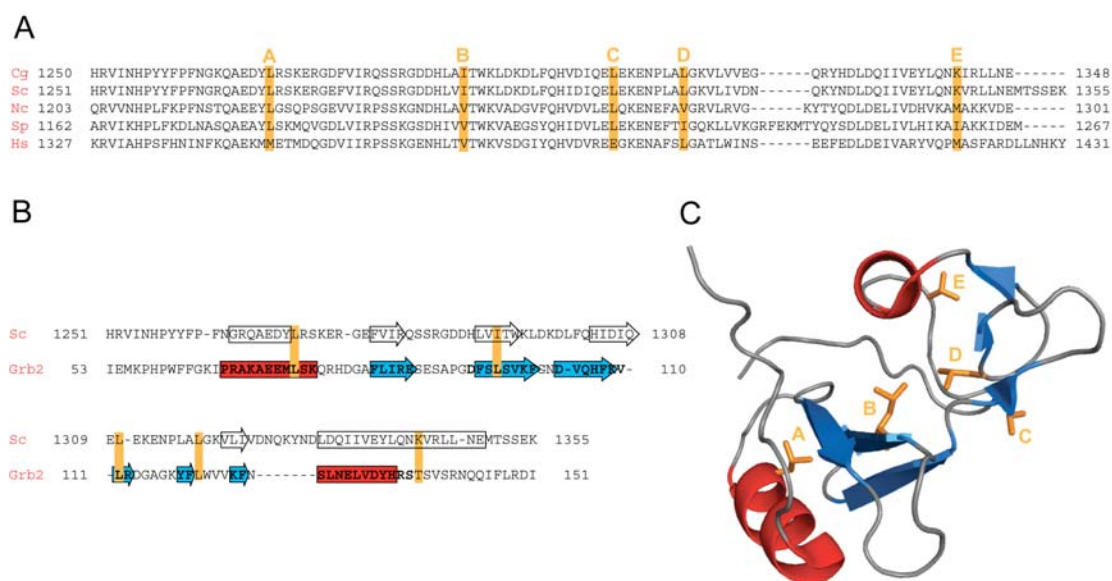
Figure 5: Design of point mutations for seleno-methionine incorporation based on the Grb2-structure

(**A**) Multiple sequence alignment (ClustalW) with Spt6 SH2 domains of *Candida glabrata* (Cg), *Saccharomyces cerevisiae* (Sc), *Gibberella zeae* (Gz), *Neurospora crassa* (Nc), *Neosartorya fischeri* (Nf); *Schizosaccharomyces pombe* (Sp), *Homo sapiens* (Hs); conserved hydrophobic residues chosen for mutation are marked orange.

(**B**) Structure-based alignment of the yeast Spt6 SH2 domain sequence and the human Grb2 SH2 domain sequence. α-helices are shown as red boxes, β-strands as blue arrows. Predicted secondary structure elements of Spt6 are shown as colourless boxes and arrows.

(**C**) structure of the Grb2 SH2 domain (PDB accession Nr.: 1JYR). α-helices and β-strands are coloured as in (B). Residues of Grb2, that align to the highly conserved residues identified in (A) are highlighted in orange.

Promising initial crystals were refined in 24 well hanging drop plates (Easy Xtal Tool, Qiagen) by varying the concentrations of constituents of initial conditions. Diffraction quality crystals of the wild type (wt) Spt6 SH2 domain of *Candida glabrata* were grown by mixing 1 µl of purified protein in the respective size exclusion buffer (see 2.1.4, Table 10) with a concentration of 15 mg/ml with 1 µl of Spt6 SH2 seeding buffer (see 2.1.4, Table 13). Plates were incubated for 2 hours at 20°C. To trigger growth of single crystals, streak seeding (Bergfors, 2003) was performed. Seeds were produced by crushing initial crystals by dilution with 10 µl of the respective reservoir solution and pipeting up and down several times. Seeds were transferred to the pre-equilibrated drops using a cat whisker. Usually, seeds that were taken up were used for seeding sequentially 6 different drops, creating dilutions which increased the probability for growth of suitably sized single crystals. Before freezing, crystals were transferred to Spt6 SH2 seeding buffer containing additionally 10%-22% (w/v) PEG 400 or 5%-17% (w/v) glycerol as cryo-protectants for several minutes. Crystals were then flash-frozen by plunging into liquid nitrogen. This procedure was resulting in diffraction quality crystals as summarized in Fig. 7 on page 42.

Selenomethionine double mutant CD (see 3.2.4 and 3.3.2) was crystallized using the hanging drop method by mixing 1 μl of purified protein in the respective size exclusion buffer (2.1.4, Table 10) with 1 μl of buffer Spt6 SH2-1 or -2 (see 2.1.4, Table 13), resulting in crystal form 1 or 2, respectively (see Fig. 7 B). Before freezing, crystals were transferred to buffer Spt6 SH2-1 containing 13% (w/v) PEG 400 or 15% (w/v) glycerol as cryo-protectant. In case of buffer Spt6 SH2-2, crystals were frozen directly from the drops.

3.2.5 Data collection and structure solution

Synchrotron radiation data of selenomethionine double mutant CD crystal forms A and B were collected at Berliner Elektronen-speicherring - Gesellschaft für Synchrotronstrahlung m.b.H. (BESSY) BL 14.1. Multiwavelength anomalous diffraction (MAD) data were collected from a selenomethionine-labeled crystal form A to a resolution of 1.9 Å. For a selenomethionine-labeled crystal of form B, single wavelength anomalous diffraction (SAD) data was collected to a resolution of 2.4 Å. Data were processed with HKL (Otwinowski, 1996). Selenium sites in crystal form A were located with HKL2MAP (Pape & Schneider, 2004) and MAD phases were calculated with SHARP (Terwilliger, 2002). A model of the SH2 domain was built with Coot (Emsley & Cowtan, 2004) and refined with CNS (Brunger et al, 1998). For crystal form B, phases were calculated by molecular replacement using PHASER (McCoy, 2007) with one molecule of the assymetric unit of crystal A as a search model. Model building was done as for crystal A, Refmac5 was used for refinement (Murshudov et al, 1997). X-ray and refinement statistics are shown in Table 16. In case of the desalted crystals used for peptide soaks (3.2.7 and 3.3.9), XDS was used for Data processing (Kabsch, 1993). Structure was solved as for crystal form B, but not refined to the end (Table 18).

3.2.6 Co-crystallization of the Spt6 SH2 domain with synthetic CTD-peptides

The Spt6 SH2 domain from *C. glabrata* was purified as described in 3.2.2 and dialyzed to a lower salt concentration in a buffer with 50 mM Tris/HCl pH 8.0 at 25°C and 20 mM NaCl. The sample was concentrated to 10 mg/ml and incubated with 3x the molar amount of peptide 1 (P1, see 2.1.2, Table 6) and 2x the molar amount of peptide 2 (P2, see 2.1.2, Table 6) at 20°C for 2 hours. After this time, the samples were used to set up commercial crystallization screens at the crystallization facility of the Max Planck Institute for Biochemistry in Martinsried. The following screens were used: Complex screen I (Qiagen), Magic screen I and II, Index screen (Hampton) and PEGs (Qiagen). Crystallization plates were incubated at 20°C and documented by an imaging system.

3.2.7 Desalting of Spt6 SH2 domain crystals for peptide soaks

Different kinds of PEG (1000 – 8000) were used in different concentrations (10%-50% w/v) to stabilize the crystal while replacing the salt. In these initial tests, crystals were transferred to 10 µl of buffers containing different amounts of PEG at lower salt concentrations (300 mM NaCl) and examined visually by light microscopy. Crystals that were transferred to Spt6 SH2 desalting buffer (see 2.1.4, Table 13), cracked immediately, but "healed" after ~30 seconds (see Fig. 14). These crystals were tested for diffraction quality on an X-ray generator (RINT2000 Series CE Marking, Horizontal Type Rotor Flex).

Since these crystals did not diffract anymore, the desalting protocol was changed to a step protocol: Buffer Spt6 SH2 B and Spt6 SH2 desalting buffer were mixed in the following ratios: 0.9/0.1, 0.8/0.2, 0.6/0.4, 0.4/0.6, 0.2/0.8, 0.15/0.85, 0.1/0.9, 0.05/0.95. Crystals were successively transferred to 100 µl of these solutions and incubated for 1 hour. For each step, one crystal was tested for diffraction on the generator. The last step, where crystals still diffracted, was used for overnight incubation. Subsequently, crystals were again tested for diffraction. In parallel, crystals were transferred imediately into different step conditions without incubation in preceding steps and tested for diffraction.

3.2.8 Soaking of Spt6 SH2 crystals with synthetic peptides

Buffer Spt6 SH2 B and Spt6 SH2 desalting buffer (see 2.1.4, Table 13) were mixed in a ratio of 0.15/0.85. Synthetic peptides were added at a concentration of 1,1 mM for P1 and 0.8 mM for P2 (see 2.1.4, Table 6). Crystals that were grown in buffer Spt6 SH2 B were transferred to 20 µl of these solutions and incubated for 1 hour at 20°C. After this, the crystals were directly frozen in liquid nitrogen. Datasets were collected at the SLS Villigen and the structure was solved as described in 3.2.4.

3.2.9 Fluorescence anisotropy (FA)

FA measurements were carried out on a FluoroMax-P Spectrofluorimeter (Horiba Jobin Yvon). 800 µl or 1 ml of Peptide P3 (see 2.1.2, Table 6) in SH2 low salt buffer (2.1.4, Table 10) were used as a reaction solution. A fluorescence emission spectrum was recorded to evaluate the strength of the fluorescence signal. The Spt6 SH2 domain of different organisms was titrated to the solution and the change in anisotropy of the fluorescence signal was recorded via direct measurement of the polarization of emission light. Binding of the *S. cerevisiae* SH2 domain was measured in SH2 low salt buffer in a titration ranging from 82 nM to 9 µM. The binding of the *C. glabrata* SH2 domain was tested in SH2 low salt buffer and in SH2 low salt buffer +150 mM NaCl in a titration ranging from 250 nM to 6.8 µM. The

human Spt6 SH2 domain was titrated in a concentration range from 36 nM to 2.6 µM in SH2 low salt buffer.

3.2.10 Surface plasmon resonance (SPR)

SPR experiments were carried out on a BIAcore X system. The Spt6-domains of *S. cerevisiae* and *H. sapiens* were bound to a Sensor-Chip NTA (GE Healthcare) via the N-terminal His-Tag (purification analog to 2.2.5 and 3.2.2, omitting the thrombin-cleavage step), using SH2 low salt buffer (2.1.4, Table 10) as running buffer. Binding upon injection of the respective domain (60 ng/µl) was monitored by an increase in the resonance signal (RU). The change in RU upon the domain binding to the chip was ~ 2000. After washing with buffer, petide P1 or P2 (2.1.2, Table 6) were injected at a concentration of 20 µM. The interaction between the respective domain and peptide was monitored by the change in RU.

3.2.11 TAP-tagging of yeast proteins

A C-terminal TAP-tag was added to *S. cerevisiae* Spt6 or Spt6ΔC (deletion of residues 1250-1451) (Youdell et al, 2008) in yeast strain S288C and the isogenic strain FY119 (see 2.1.1, Table 3) as described (Puig et al, 2001). The oligonucleotides „Spt6flTAPfw", „Spt6flTAPrv" and „Spt6dSH2TAPfw" (2.1.2, Table 5) were used to amplify the C-terminal TAP-tagging cassettes from plasmid pBS1539 (2.1.2, Table 4). The PCR-products were purified from a 1% agarose gel usind the QIAquick Gel Purification Kit (Qiagen). The strains were transformed with the respective purified DNA using Lithium acetate as described (Knop et al, 1999). Cells that integrated the foreign DNA into their genome by homologous recombination were identified on selection plates. Single colonies were cultivated and checked by Western blot (Fig. 16). For this, total protein of *S. cerevisiae* strains (S288C, isogenic to strains used in microarray-analysis except for TAP- tags at the C-terminus of the Spt6/Spt6ΔC protein) was resolved by a 8% SDS-PAG and blotted on a PVDF-membrane. The membrane was probed with antibodies directed against the TAP-tag (PAP, Sigma) and tubulin (3H3087, Santa Cruz Biotechnology) as a loading control. Bound antibodies were detected by chemiluminiscence (ECL Plus Western Blotting detection system, GE Healthcare). For tubulin, Peroxidase-conjugated AffiniPure Rabbit Anti-Rat IgG (H+L) (Jackson ImmunoResearch) was used for detection.

3.2.12 Chromatin immunoprecipitation (ChIP)

Chromatin immunoprecipitation experiments were carried out according to an established protocol (Aparicio et al, 2005; Jasiak et al, 2008), together with Kristin Leike.

YPD medium (2.1.3, Table 7) was inoculated to an OD_{600} of 0.2 from a starter culture of the respective strain. The culture was grown at 30°C and 180 rpm until it reached the log-phase (OD_{600} 0.7-0.8). Protein cross-linking was achieved by the addition of 37% formaldehyde to the culture to a final concentration of 1% and slowly shaking at room temperature for 15 minutes. Addition of 3M Glycine to a final concentration of 2.5% and incubation for another 30 minutes stopped the cross-linking reaction. Cells were harvested by centrifugation (4400 g, 5 min, 4°C) and washed three times with cold 1x TBS (2.1.4, Table 9) and once with FA lysis buffer (2.1.4, Table 14) + 2mM PMSF. Cells were frozen in liquid nitrogen and stored at -80°C for further use.

For immunoprecipitation of TAP-tagged proteins, IgG Sepharose™ 6 Fast Flow (GE Healthcare) was used (50 µl of a 50% slurry per IP reaction). Beads were washed with cold 1x TBS buffer an FA-lysis buffer before use. Chromatin was sheared using a BioruptorTM UCD-200 (Diagenode) (25 x 30 seconds with 30 sec breaks at an output of 200 W). 20 µl of the resulting chromatin solution was kept as the input sample. Precipitation was performed for 3 h at room temperature by incubation of the chromatin solution with 50 µl of the washed beads. After precipitation, beads were washed with FA lysis buffer, FA lysis buffer + 500 mM NaCl, ChIP wash buffer (see 2.1.4, Table 14) and TE buffer. Elution of precipitated proteins was carried out in ChIP elution buffer (2.1.4, Table 14) at 65°C for 20 min. Input and IP samples were incubated with Pronase at 42°C for 3h. Reversal of cross-linking was accomplished by a 9 h incubation at 65°C. The nucleic acids were purified using QIAquick PCR Purification Kit (Qiagen) (instead of the standard DNA binding buffer, PB buffer from Qiagen (Cat. No. 19066) was used). After purification RNA was removed by incubation with RNase A at 37°C for 30 min. DNA was purified once more and the sample volume was reduced to the desired concentration using a Speed Vac.

Amplification and Re-amplification of the immunoprecipitated DNA was carried out with GenomePlex® Complete Whole Genome Amplification (WGA) Kit and GenomePlex® WGA Reamplification Kit (Sigma). The quality of the resulting samples was checked on a 1% Agarose Gel, Ethidium bromide staining and visualization in UV light. A correctly amplified sample showed DNA-fragments of various length, with a peak around 300 nt.

3.2.13 ChIP-on-chip

ChIP samples were sent to imaGenes GmbH for labeling, hybridization, array scanning, data extraction and a preliminary data analysis using a *S. cerevisiae* whole genome tiling array

(Cat. No. C4214-00-01, http://www.imagenes-bio.de/services/nimblegen/chip). Two biologically independent samples of each input and IP DNA were sent. The analysis included „dye-swapping" with the fluorophores Cy3 and Cy5, which improves the signal to noise ratio. The bioinformatic analysis of the data was performed by Matthias Siebert and Johannes Soeding from the Gene Center Munich, as part of a collaboration.The data quality was high and the analysis was carried out as described in Jasiak et al. (2008). Briefly, the logarithm of the fluorescence signal of the Chip DNA was divided by the signal from the genomic background and this was used in further analysis. A standard background correction was performed on all such signals by subtracting their genome-wide average. ChIP-chip measurements were repeated with exchanged dyes Cy5 and Cy3 and averaged over measurements to subtract out the strong, systematic, dye-related technical noise, which effectively eliminates intensity-dependent saturation effects. The experiment was carried out as a biologically independent duplicate, resulting in Pearson correlation coefficients of 0.84 for the Cy5 labeled biological duplicates and 0.82 for the Cy3 labeled biological duplicates.

3.3 Results and discussion

3.3.1 Delineation of the Spt6 SH2 domain of *Saccharomyces cerevisiae* for crystallization

Information about sequence conservation (Fig. 9 A) as well as the secondary structure prediction (see Fig. 5 B) was used to design variants of the *S. cerevisiae* Spt6 SH2 domain that differed in length at their C- and N-termini. 10 of these variants were cloned, expressed and purified (chapters 2.2.1, 2.2.3, 2.2.5 and 3.2.2). The relative solubility was estimated by comparing the amount of protein in the extract (soluble fraction) to the amount in the pellet (insoluble fraction) after centrifugation of the cell lysate (2.2.5). The results are summarized in Fig. 6.
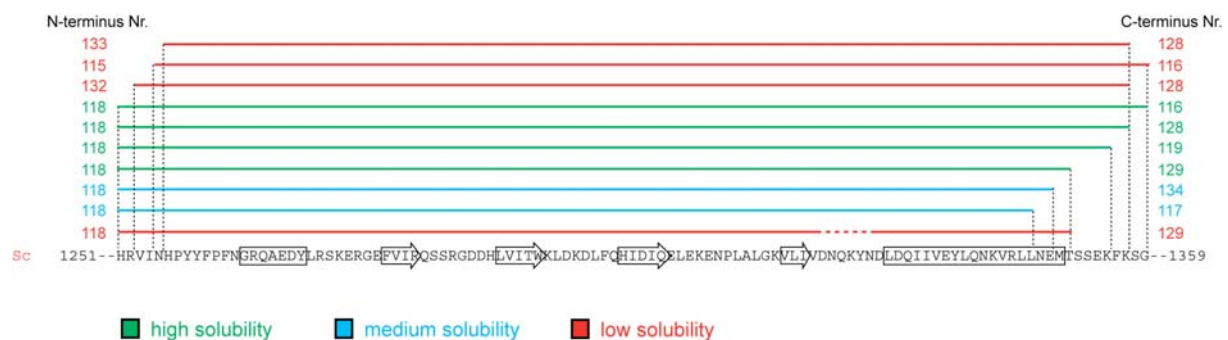
Figure 6: Solubility of *S. cerevisiae* Spt6 SH2 domain variants

The structure based alignment of Fig. 5 B for the Spt6 SH2 domain and the lengths of the different protein variants that were expressed and purified is shown. The solubility is indicated by colours. Numbers refer to the primers (Sc) that were initially used to clone the respective variants (see section 2.1.2, Table 5)

The solubility of the protein variants gives a good estimation about the boundaries of the domain. Unsoluble proteins are considered to be defective in folding because important parts of the domain-fold are removed. The length of the N-terminus could be shortened up to amino acid 1251 (N-terminus Nr. 118), the C-terminus up to amino acid 1351 (C-terminus Nr.129). An internal loop-deletion (dashed line in Fig. 6) rendered this variant insoluble. Thus, the domain boundaries were defined by the protein variant 118+129. This variant was purified to high purity and homogeneity, and was used for extensive screening of crystallization conditions (section 3.2.4). However, the Spt6 SH2 domain from *S. cerevisiae* did not form crystals.

3.3.2 Crystallization and structure solution of the Spt6 SH2 domain of *Candida glabrata*

The variant Sc118+129 of the SH2 domain showed an optimal behavior in terms of solubility and in limited proteolysis experiments (2.2.6, data not shown) but did not form crystals (3.3.1). Due to the high conservation of the domain in various species, the domain borders that were delineated in *S. cerevisiae* could be transferred to the proteins of *Candida glabrata, Schizosaccharomyces pombe* and *Homo sapiens* (Fig. 9 A). In the new variants two additional residues were removed from the C-terminus, based on sequence alignment and secondary structure prediction. The coding sequences for these domain variants were cloned, expressed and purified (chapters 2.2.1, 2.2.3, 2.2.5 and 3.2.2). These variants showed a similar behaviour in the purification compared to the *S. cerevisiae* protein and were purified to high homogeneity. The quality of the protein sample is exemplified for the *C. glabrata* variant in Fig 7 A, but similar results were obtained for the domains from the other organisms.

Initial screenings for formation of crystals gave positive results only for the *C. glabrata* protein. Refinement of these conditions, including streak seeding, lead to an overall improvement of crystals and to a native dataset to 2.8 Å resolution (see 3.2.4 and Fig 7 A). However, despite several attempts to solve the structure by molecular replacement with different search models (not shown), no solution could be obtained.

Thus we attempted to get *de novo* phases from anomalous diffraction of Selenium atoms, by the incorporation of seleno-methionine (2.2.3). For this we had to insert several point mutations into the sequence of the *C. glabrata* SH2 domain, since the native protein did not contain any methionine residue. 5 point mutations were designed (see 3.2.3 and Fig. 5). Four of those could be cloned (mutations A, C,D and E) and the resulting proteins were soluble. Mutations C, D and E were chosen for further subcloning, resulting in double mutant CD and triple mutant CDE, which again showed high solubility. A repeated screening for inital crystallization conditions with the two mutant proteins revealed two novel crystallization conditions for mutant CD (buffers Spt6 SH2-1 and -2, chapter 2.1.4, Table 13). Mutant CDE was crystallized in similar conditions as the native protein, but a SAD dataset to 3.1 Å resolution did not result in solution of the structure because of a weak anomalous signal (data not shown). However, crystals from mutant CD were suitable to solve the structure of the SH2 domain in multiple wavelength anomalous diffraction experiment (MAD, 3.2.5 , Table 16). Subsequently, also the second crystal form of the CD mutant could be solved by molecular replacement, using the structure from crystal form 1 as a search model (Table 16). All diffraction quality crystals that were obtained are summarized in Fig. 7 B.

A

B

	native	SeMet mutant CDE	SeMet mutant CD (crystal form 1)	SeMet mutant CD (crystal form 2)
crystallization buffer	Spt6 SH2 seeding buffer 50 mM MES pH 6.5 350 mM MgAcetate 20% PEG 3350 5 mM TCEP	Spt6 SH2 seeding buffer 50 mM MES pH 6.5 300 mM MgAcetate 10% PEG 3350 5 mM TCEP	Spt6 SH2-1 100 mM Hepes pH 7.0 1 M Succinic acid 1% PEG 2000 MME 5 mM TCEP	Spt6 SH2-2 100 mM Hepes pH 7.5 4.3 M NaCl 5 mM TCEP
diffraction	2,8 Å	3,1 Å	1,9 Å	2,4 Å
experiment	molecular replacement	SAD	MAD	molecular replacement
result	no solution	weak anomalous signal	structure solution	structure solution

Figure 7: Quality of the *C. glabrata* SH2 domain protein sample and the resulting diffraction quality crystals

(**A**) Chromatogram of a Superose12 size exclusion chromatography of the *C. glabrata* Spt6 SH2 domain. Absorption units at 280 nm and 260 nm are shown in blue and red, respectively. In addition, an overloaded, Coomassie stained SDS-PAGE of the pooled peak fractions is shown, to demonstrate the purity of the protein sample.

(**B**) In columns, the different crystals of the *C. glabrata* Spt6 SH2 domain, the respective crystallization buffers (see also 2.1.4, Table 13), diffraction images and resolution, experiment and results are shown.

Table 16: X-ray diffraction and refinement statistics for *C. glabrata* Spt6 selenomethionine double mutant CD crystals

	crystal form 1			crystal form 2
Data collection				
Space group	P65			P32
Cell dimensions				
a, b, c (Å)	54.5, 54.5, 253.4			71.6, 71.6, 87.6
α, β, γ (°)	90, 90, 120			90, 90, 120
	Peak	**Remote**	**Inflection**	
Wavelength (Å)	0.97973	0.90810	0.97987	0.97971
Resolution (Å)	20-1.9	20-1.9	20-1.9	20-2.4
Rsym (%)	5.2 (12.9)	5.3 (22.3)	3.9 (12.6)	6.2 (19.0)
I / σI	42.9 (7.3)	34.35 (7.5)	31.24 (6.0)	45.14 (7.7)
Completeness (%)	99.4 (96.6)	99.9 (100)	99.3 (95.7)	99.5 (95.4)
Redundancy	4.2 (2.8)	7.7 (6.7)	4.1 (2.8)	3.9 (3.7)
Refinement				
Resolution (Å)	1.9			2.4
No. reflections	33162			18513
Rwork / Rfree (%)	19.6 / 24.1			25.27 / 28.50
No. atoms				
Protein	3292			3160
Ligand/ion	20			-
Water	428			76
B-factors				
Protein	26.6			29.8
Ligand/ions	23.5			-
Water	35.4			27.8
R.m.s deviations				
Bond lengths (Å)	0.005			0.007
Bond angles (°)	1.2			1.4

3.3.3 Crystallization of SH2 domains from various species

Interestingly, the SH2 domains of *S. cerevisiae* and *C. glabrata* only differ in 10 amino acid positions, as is shown in the alignment left in Fig. 8. Although most of these positions are highly conserved, they make the difference between crystal formation or no crystals. When these residues are mapped on the four molecules of the asymmetric unit of the CD mutant crystal form 1, all residues except for two (A1291 and V1304) lie on the surface of the individual molecules and thus in between the molecules that build up the crystals (Fig. 8, right). These „evolutionary point mutations" render the *C. glabrata* protein variant suitable for crystallization. Thus, making use of naturally occurring variances in proteins by extending crystallization trials to different source organisms is an appropriate remedy in the crystallization of difficult proteins.
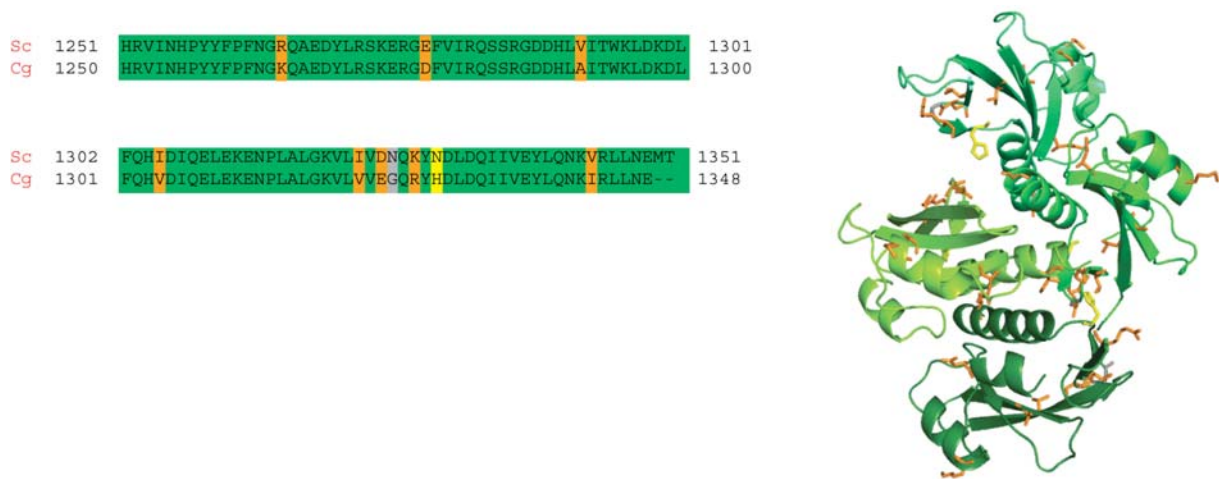
Figure 8: The crystallizable *C. glabrata* protein variant differs in only 10 amino acid positions from the *S. cerevisiae* protein

Alignment of the *S. cerevisiae* and *C.glabrata* SH2 domain sequences (left). Invariant residues are coloured in green, conserved residues are indicated in orange (high) and yellow (low), unconserved residues in gray. On the right, the 4 molecules of the asymmetric unit of crystal „SeMet mutant CD" is shown. Invariant residues from the alignment are shown as a ribbon model (different greens for every molecule in the asymmetric unit). Conserved and unconserved residues are shown as stick models with the same colours as in the alignment.

3.3.4 The structure of the Spt6 SH2 domain reveals a typical SH2 fold with unique features

The structure derived from "SeMet mutant CD crystal form 1" in a MAD experiment (Fig. 7) reveals the classical core fold of SH2 domains (Kuriyan & Cowburn, 1997) with a central three-stranded antiparallel β-sheet (βB-βD) sandwiched between two α-helices (αA and αB, Fig. 9 B). In addition to this core fold, the structure contains an α-helix N-terminal of αA (called αL here), a small anti-parallel β-sheet inserted between βB and βD (βE-βF), and an extended C-terminal α-helix (Fig. 9 B, Fig. 12 D and Table 17). Two surface loops, βB-βC and βD-βB, adopt alternative conformations in two different crystal packings and are thus mobile (Fig. 9 C). Interestingly, the seleomethionine residues in mutant CD were not part of the hydophobic core as predicted (3.2.3). They are positioned at the beginning and the end of loop DE. However they were ordered and did not disturb the overall structure and were thus suited for phasing.

3.3.5 The Spt6 SH2 domain structure contains features of both sub-families of SH2 domains

To understand the unique features of the structure, we reviewed available structures from the PDB database (Table 17) for the corresponding characteristics. It is possible to group SH2-domains into two classes, according to specific features of their primary sequences and their 3D-structures. One group, the "STAT-type" contains the SH2 domains of STAT transcription factors, the second group "Src-type" contains most of the other domains that are part of proteins involved in cellular signaling pathways (Gao et al, 2004): First, we looked for the SH2 core motif sequence, which has the consensus sequence GXF/YBBR (X for any, B for hydrophobic amino acids). This motif resides in the βB-strand (Fig 9 A) and contains the highly conserved arginine residue that is binding the phosphate. A specific feature of all STAT-type domains is a highly conserved phenylalanine immediately following the arginine
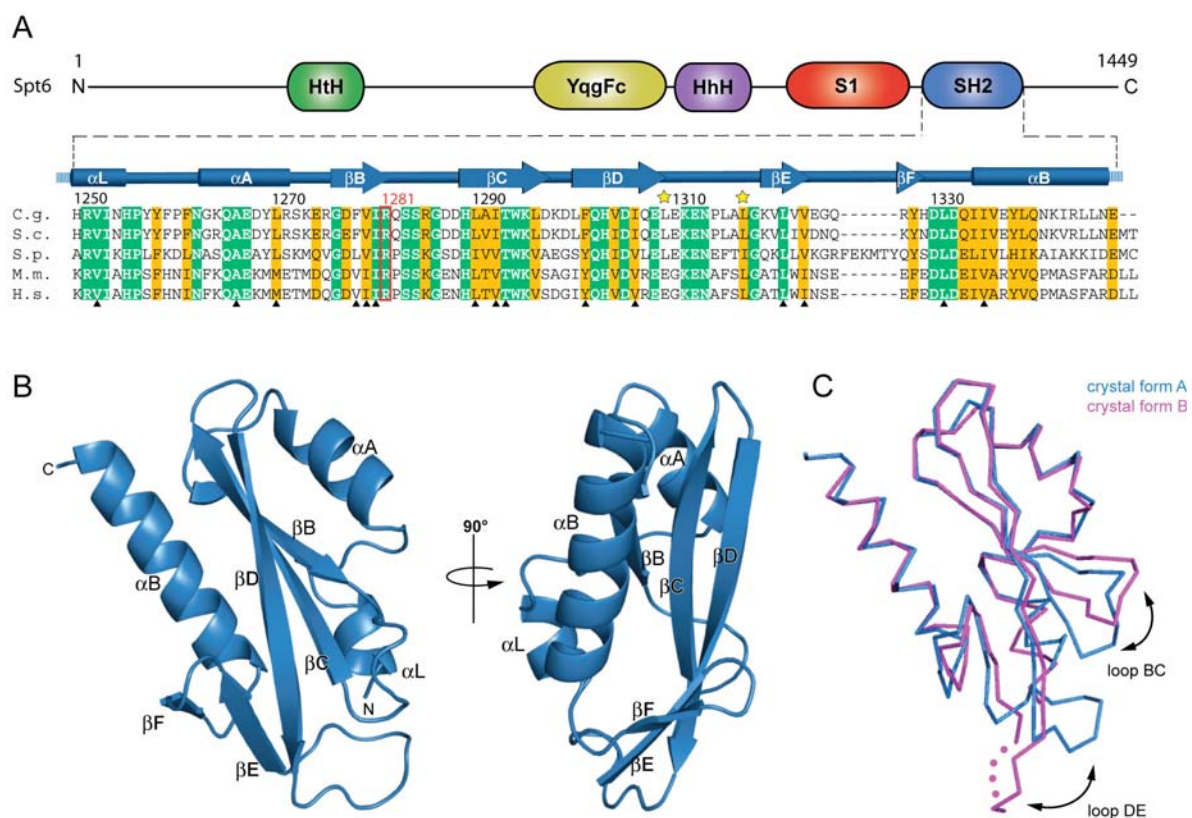

Figure 9: Structure of the Spt6 SH2 domain

(**A**) Domain architecture of Spt6 (Johnson et al, 2008) and alignment of amino acid sequences of the Spt6 SH2 domains of *C. glabrata* (C.g.), *S. cerevisiae* (S.c.), *S. pombe* (S.p.), *M. musculus* (M.m.), and *H. sapiens* (H.s.). Secondary structure elements are indicated above the alignment (cylinders for α-helices, arrows for β-strands).

Figure 9 (continued)

Invariant and conserved residues are highlighted in green and yellow, respectively. The red box indicates the invariant arginine R1281 that binds the phospho group (see also 3.1.1). Stars indicate positions where methionine was introduced for seleno-methionine phasing (sections 3.2.3 and 3.3.2). Residues forming the hydophobic core are marked with a triangle.

(**B**) Two views of a ribbon model of the Spt6 SH2 domain. Secondary structure elements are labeled as in (A).

(**C**) Comparison of the Cα-trace of the SH2 domain structure in two different crystal forms (blue, crystal form 1, magenta, crystal form 2 of SeMet mutant CD, see Fig. 7 B). Structures were aligned by DALILite (Labarga et al, 2007), resulting in a RMSD of 1.2 Å for 91 residues. Unstructured parts in crystal form B are indicated by dots.

C-terminally (Gao et al, 2004), whereas in Src-type domains this residue is different (Fig. 10, Table 17). Spt6 shows no phenylalanine in this position and thus resembles the Src-type in this respect. Interestingly, the consensus sequence of the Spt6 SH2 domain core motif as seen in Fig 9 A is GXF/BBR, thus varying in the third position from the consensus of all other SH2 domains (with the exception of CHK, see table 17). The landmark feature of Src-type domains is the small antiparallel β-sheet βE-βF, that is inserted between strand βD and helix αB (Gao et al, 2004). Residues from this sheet are involved in interactions with residues of the target peptide other than the phospho-aminoacid, and thus important for peptide specificity as outlined in section 3.1.1 (Kimber et al, 2000; Waksman et al, 1993). STAT-type domains lack this structural feature (Fig. 10). Our review in Table 17 confirmed this with the exception of the SH2 domains of Cbl and APS, where a βE-βF sheet could not be detected. The Spt6 SH2 domain has a clearly defined βE-βF sheet and resembles the Src-type in this respect.

The Spt6 domain has an unusually long C-terminal helix αB (at least 26 Å long). Only the SH2 domain of APS, which is a substrate of the insulin receptor, showed this feature in our structure review (Table 17). APS is dimerizing on its substrate and the long C-terminal helix is interfering with the canonical peptide binding path (Hu et al, 2003), which is usually in an extended conformation, perpendicular to the central β-sheet (Kuriyan & Cowburn, 1997). STAT transcription factors also self-dimerize upon the phosphorylation of a specific residue before they can translocate to the nucleus and induce the transcription of specific genes (Becker et al, 1998; Chen et al, 1998; Darnell, 1997). The absence of a βE-βF sheet is a unique feature of the STAT-subfamily of SH2 domains, as is the presence of an extended α-helical structure, termed αB'-αB (Gao et al, 2004) (Fig. 10). The extended C-terminal α-helices αB in Spt6 and APS might resemble a primitive or degenerated fused form of this αB'-αB structural motif. This suggestion is supported by the fact that APS dimerizes on its substrate as STAT transcription factors do. For Spt6, no data on multimerization on target binding is currently available.

The structure of the Spt6 SH2 domain contains a short α-helix (αL) N-terminally of αA (Fig. 9 B). This helix is not a part of the canonical SH2 fold, but is seen in all STAT transcription

factors - where it is part of the STAT linker domain that is preceding the SH2 domain N-terminally – and in a number of Src-type domains (Table 17 and Fig. 10). The absence of such a helix as outlined in Table 17 does not necessarily mean its absence in the protein, it may just not be a part of the crystal structure. However, the position of αL in the Spt6 structure is not artificial, as it is similar in both crystal forms (Fig. 9 C). Its interaction with the rest of the domain is mediated by Arg1251, which contacts Asp1330 of loop FB. Additionally it packs onto a hydrophobic surface made up by residues of the β-sheets βB and βC. The helix αL adopts the same relative position to the rest of the protein as the C-teriminal helix of the STAT linker domain (exemplified by α11 of STATa of *Dictyostelium,* yellow in Fig. 10). Furthermore, secondary structure prediction of Spt6 further C-terminal of the SH2 domain reveals a structural signature consisting of 5 β-strands. This might resemble the S1-RNA-binding-domain found in STAT proteins N-terminal of the linker domain (Becker et al, 1998; Bycroft et al, 1997; Chen et al, 1998). Taken together, helix αL as well as the extended helical structure αB are likely to be evolutionary related to STAT transcription factors.

Table 17: Review of SH2 domain structural elements

Protein, species [a], (PDB accesion no.)	core motif [b]	βEβF	αB [Å] [c]	αL [d]	R155 [e]	K203 [e]	Z-score (Dali server) [f]
Src-type							
Spt6, Cg, (-)	GDFVIRQ	+	26.0	+	G	D	-
Abl1, Mm, (1OPK)	GSFLVRE	+	14.3	-	R	R	-
Aps, Rn (1RPY)	GLFVIRQ	-	29.0	+	R	R	8.6
Blk, Mm, (1BLJ)	GSFLIRE	+	14.8	-	R	K	8.0
PTK6/Brk, Hs, (1RJA)	GAFLIRV	+	16.7	-	R	L	8.1
Cbl, Hs, (2CBL)	GSYIFRL	-	14.2	+	Y	A	6.5
Crk, Hs, (1JU5)	GVFLVRD	+	16.5	-	R	I	8.4
CHK, Hs, (1JWO)	GLVLVRE	+	17.3	-	G	R	9.8
Fes/Fps, Hs, (1WQU)	GDFLVRE	+	17.9	-	R	I	7.2
Fyn, Hs, (1G83)	GTFLIRE	+	17.4	+	R	K	9.8
GADS, Mm, (1R1Q)	GFFIIRA	+	12.9	+	R	K	9.3
Grb10, Hs, (1NRV)	GLFLLRD	+	14.7	-	R	N	-
Grb14, Hs, (2AUG)	GVFLVRD	+	14.7	+	R	N	9.8
Grb2, Hs, (1GRI)	GAFLIRE	+	15.2	-	R	K	8.6
Grb7, Hs, (1MW4)	GLFLVRE	+	10.1	-	R	L	4.5
Hck, Hs, (3HCK)	GSFMIRD	+	15.0	-	R	K	8.5
Itk/Tsk, Mm, (1LUK)	GAFMVRD	+	14.7	-	R	K	7.7
P56-Lck, Hs, (1BHH)	GSFLIRE	+	12.2	-	R	K	9.2
Nck1, Hs, (2CI9)	GDFLIRD	+	16.9	-	R	K	9.6
Nck2, Hs, (2CIA)	GDFLIRD	+	16.5	-	R	K	10.1
P85-N, Hs, (2IUG)	GTFLVRD	+	15.8	+	R	K	9.3
PLCγ1, Bt, (2PLD)	GAFLVRK	+	14.5	-	R	R	6.7
Syp, Mm, (1AYA)	GSFLARP	+	17.0	-	G	K	9.0
SHP-1, Hs, (1X6C)	WTFLVRE	+	17.0	-	G	K	8.1
Sap, Hs, (1D1Z)	GSYLLRD	+	12.9	-	R	R	7.7
Eat2, Mm, (1I3Z)	GNFLIRD	+	11.1	-	K	L	9.6
SH3BP2, Hs, (2CR4)	GLYCIRN	+	14.8	-	S	R	6.9
Shc1, Hs, (1MIL)	GDFLVRE	+	15.3	-	R	L	-
Socs3, Mm, (2BBU)	GTFLIRD	+	15.0	+	G	R	4.2
Syk, Hs, (1CSZ)	GKFLIRA	+	13.6	-	R	R	8.0
Vav1, Hs, (2CRH)	GTFLVRQ	+	14.2	+	R	K	7.2
ZAP70-1, Hs, (1M61)	GLFLLRQ	+	16.4	-	R	P/I	10.4
ZAP70-2, Hs, (1M61)	GKFLLRP	+	14.4	+	R	L	10.4
STAT-type							
STAT1, Hs, (1BF5)	GTFLLRF	-	5.7 + 8.8	+	P/E	K	4.2
STAT3, Hs, (1BG1)	GTFLLRF	-	10.8 + 9.8	+	P/E	K	4.1
STAT5, Mm, (1Y1U)	GTFLLRF	-	8.7 + 9.9	+	K	K	5.5
STAT, Dd, (1UUR)	GTFIIRF	-	7.3 + 10.5	+	R	L	5.7

Structural data was obtained from the Protein Data Bank (PDB) and inspected for the indicated structural elements.

[1] Cg=*Candida glabrata*; Hs=*Homo sapiens*; Mm=*Mus musculus*; Rn=*Rattus norvegicus*; Bt=*Bos taurus*; Dd=*Dictyostelium discoideum*

[2] highly conserved SH2 core motif of consensus sequence GXF/YBBR (X for any, B for hydrophobic amino acids) containing the arginine residue binding the phosphate (red)

[3] length of helix αB in Angstroem

[4] presence of an α helix resembling the αL helix in Spt6 (Fig. 9 B) in position relative to the SH2 fold

[5] residue found at the same position as R155 and K203 in the structure of the Src SH2 domain (see also section 3.1.1)

[6] Z-score derived from a DALI-server (Holm & Sander, 1993) search for related protein structures using the Spt6 SH2 domain as a query. Higher scores indicate higher similarities (similarities with a Z-score < 2 are insignificant)
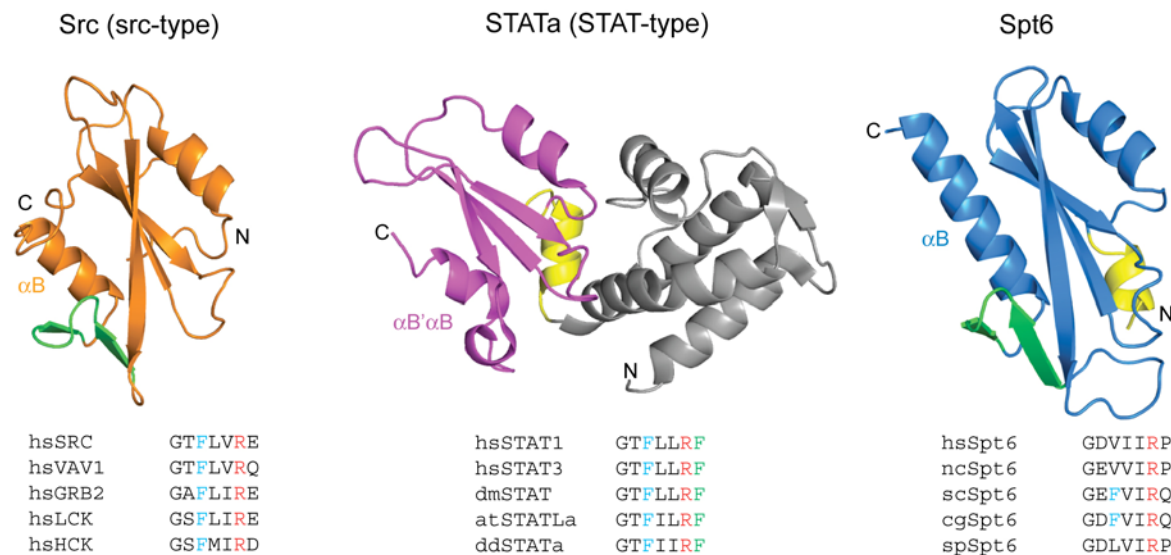
Src (src-type)		STATa (STAT-type)		Spt6	
hsSRC	GTFLVRE	hsSTAT1	GTFLLRF	hsSpt6	GDVIIRP
hsVAV1	GTFLVRQ	hsSTAT3	GTFLLRF	ncSpt6	GEVVIRP
hsGRB2	GAFLIRE	dmSTAT	GTFLLRF	scSpt6	GEFVIRQ
hsLCK	GSFLIRE	atSTATLa	GTFILRF	cgSpt6	GDFVIRQ
hsHCK	GSFMIRD	ddSTATa	GTFIIRF	spSpt6	GDLVIRP

Figure 10: Unique fold of the Spt6 SH2 domain

Comparison of the overall structures of the SH2 domains of *H. sapiens* Src (Waksman et al, 1993), of *D. discoideum* STATa (Soler-Lopez et al, 2004) and of *C. glabrata* Spt6 (blue). The βEβF-sheet is coloured in green, the Spt6 helix αL, as well as STATa linker domain helix α11 are coloured in yellow. The remainder of the STAT linker domain is in grey. Below each structure, alignments of the SH2 βB core motif of five representative sequences are shown (hs = *H. sapiens*; dm = *D. melanogaster*; at = *A. thaliana*; dd = *D. discoideum*; nc = *N. crassa*; sc = *S. cerevisiae*; cg = *C. glabrata*; sp = *S. pombe*). The highly conserved arginine residue that binds the phospho group is highlighted in red, a conserved phenylalanine residue which is a feature of STAT-type domains is highlighted in green. A conserved phenylalanine that is invariant in all SH2 domains except for the Spt6 domain is in cyan.

3.3.6 The SH2 domain of Spt6 is an ancestor of the mammalian SH2 domains involved in signal transduction

The preceding analysis of the structure of the Spt6 SH2 domain identified features, that relates it to both structural SH2 domain subfamilies, the Src-type and the STAT-type (Gao et al, 2004). Both families occur to different extents in the eukaryotic taxa: Human cells contain 120 SH2 domains –the large majority of them belonging to the Src-type- distributed over 110 proteins, that are generally involved in phospho-tyrosine recognition during cell signaling events (Liu et al, 2006) (Fig. 3 A). This high abundance of SH2-containing proteins in animal cells likely reflects the need for fine-tuned intercellular signaling in these complex organisms. *Dictyostelium discoideum*, an organism that can switch between single-cell and multicellular lifestyles, contains 12 different SH2 domain-containing polypeptides, 4 of those as part of STAT transcription factors (Eichinger et al, 2005; Williams et al, 2005). Two bioinformatic studies have found two SH2 domain sequences in plant genomes, which seem to be related

to the STAT-type subfamily (Gao et al, 2004; Williams & Zvelebil, 2004). In contrast to that, yeast cells only contain one SH2 domain in the genome: that of Spt6 (Maclennan & Shaw, 1993).

Spt6, together with its SH2 domain, is found in all eukaryotic taxa. Thus, the protein existed prior to the split of plant and animal lines, as well as the STAT type domains (Fig. 11 A). Because the Spt6 structure bears features from both subfamilies it is highly likely that the Spt6 SH2 domain is the ancestor of the modern SH2 domains. The hallmark feature of the Src-type subfamily – the βEβF-sheet – clearly was already invented with Spt6 (section 3.3.3, Fig. 9 B and Fig. 10), so it is currently the first in SH2 domain development showing this trait. The relationship between the Spt6 SH2 domain and members of the Src-type subfamily can also be shown by a phylogenetic analysis of their protein sequences (Fig. 11 B). There, Spt6 clusters together with other domains of the Src-type, whereas STAT proteins form their own cluster.
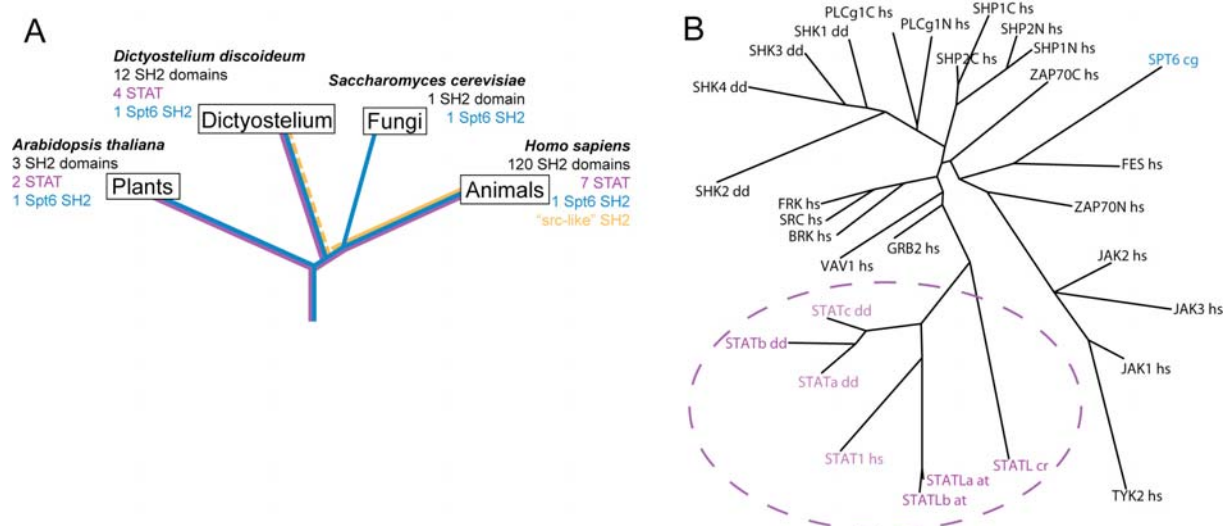
Figure 11: Evolution of SH2 domains

(**A**) Unscaled phylogenetic tree of eukaryotic organisms (Eichinger et al, 2005). For every taxon, a representative model organism is indicated together with the number of SH2 domain-containing proteins and STAT proteins. Colours of the branches of the tree indicate the distribution of SH2 subfamilies and Spt6 in the different taxa (Spt6, blue, STAT-type, magenta, Src-type, orange). The dashed, orange line in the Dictyostelium branch reflects the fact that it is unclear for some of the twelve SH2 domains to which subfamily they belong.

(**B**) The Spt6 SH2 domain (blue) clusters with Src-type domains in a phylogenetic analysis. The cluster of STAT sequences is shown in purple. Organisms are indicated as in Fig. 10. The unrooted phylogenetic tree was calculated with Proml (Protein maximum likelihood) and drawn by Drawtree, which represent tools of the PHYLIP software package version 3.67 (Felsenstein, 1989), figure was provided by Andreas Mayer.

The evolutionary relationship to STAT-proteins is less clear. The αL-helix, as well as the extended αB helix might be a primitive or a degenerated relative of the STAT-linker domain and the αB'αB-motif, respectively. However, the Spt6 SH2 domain is the „minimal equipment" for a eukaryotic cell in terms of SH2 domains, given that the yeast genome encodes only one. This strongly suggests a similar situation in a eukaryotic precursor cell at the root of the eukaryotic tree of life.

Spt6 is not found in prokaryotic cells. However, in pathogenic bacteria, a gene-product was identified which is involved in the expression of toxin genes (Fuchs et al, 1996). Primary sequence analysis of this protein, named Tex (for *t*oxin *ex*pression), identified a similar domain architecture for Tex and parts of Spt6 (Ponting, 2002): YqgF, HhH and S1 RNA-binding domains were identified in both proteins (see also Fig. 9 A). Spt6 contains an additional N-terminal nucleosome binding domain (Johnson et al, 2008) as well as the C-terminal SH2 domain. Thus, the core structure of Spt6 might have evolved from the Tex protein (Johnson et al, 2008) and the additional domains might have evolved as an adaption to eukaryotic processes, like interactions with nucleosomes and with the CTD of Pol II.

3.3.7 The conserved phospho-binding pocket can explain the unusual phospho-serine specificity

The domain surface of the Spt6 SH2 domain shows a conserved patch that includes a small pocket (pocket 1, Fig. 12 A) that contains an invariant arginine residue (R1281). This arginine is present in all known SH2 domains and interacts directly with the phosphate group of the target phosphopeptide in the SH2 domain-phosphopeptide complex (Waksman et al, 1992) (Fig. 12 D). In the "SeMet mutant CD" crystal form 1, R1281 binds a succinate ion that was present in the crystallization buffer. R1281 side chains from two neighboring domains in the crystal each bind a carboxylate of the succinate ion (Fig. 12 B). This observation is consistent with a high affinity of pocket 1 for negatively charged chemical groups, and the conserved phospho-binding function of R1281. Indeed, mutation of R1281 decreased the interaction of the SH2 domain with the phosphorylated CTD *in vitro* (Yoh et al, 2007).

We next modeled the possible CTD interaction with the use of the known Src SH2 domain-phosphopeptide complex structure (Waksman et al, 1993) (Fig. 12 D). The two structures were superimposed with their conserved residues in the central β-sheet of the core domain fold. The phospho-binding pockets are highly similar in structure. The positions of Cα-atoms of the residues that form pocket 1 are essentially identical (Fig. 12 D). The Spt6 residue R1281 perfectly aligns with the phospho-tyrosine binding residue R175 of Src and is thus in a position to make contacts with the phosphate by specific hydrogen-bonding interactions between the two terminal nitrogens and two phospate-oxygens (Fig. 12 E). Although the

position of the residues of the phosphate-binding-loop βB-βC deviates slightly from those of Src (dashed blue line in Fig. 12 D), additional contacts of the phosphate group are likely conserved, including hydrogen bonds to a backbone amide (E178 in Src, S1284 in Spt6) and to the side-chain hydroxyl group of at least one of two residues (S177 and T179 of Src, apparently corresponding to S1283 and S1284, respectively, of Spt6).

Whereas the contacts to the phosphate group are likely conserved, modelling suggests that contacts to an aromatic ring of a phospho-tyrosine side chain are apparently not possible in the Spt6 domain, consistent with binding to a phospho-serine peptide. The specific recognition of the phospho-tyrosine aromatic ring by Src is achieved by amino-aromatic interactions, where the π-electrons of the aromatic ring interact with the amino groups of residues R155 and K203 in Src (Fig. 12 E). The aromatic ring of the phospho-tyrosine is sandwiched between these two residues (Fig. 12 D). Whereas the two residues corresponding to Src residues R155 and K203 are conserved in the majority of SH2 domains, some SH2 domains lack one of them (Table 17). However, the SH2 domains of Spt6 and Cbl are the only SH2 domains within our structural comparisons, in which both residues are not conserved. Src residues R155 and K203 are replaced by residues G1263 and D1305 in Spt6. The negatively charged residue D1305 was only observed in Spt6 and is predicted to repel aromatic π-electrons, consistent with the selection of serine over tyrosine. The alignment also shows that pocket 1 is more shallow in Spt6 than in Src (Fig. 12 F). Thus, the Spt6 pocket 1 is suited for phosphate binding and R1281 is better accessible from the solvent than in Src, making it possible to interact with a short phosphorylated serine side chain, instead of tyrosine.
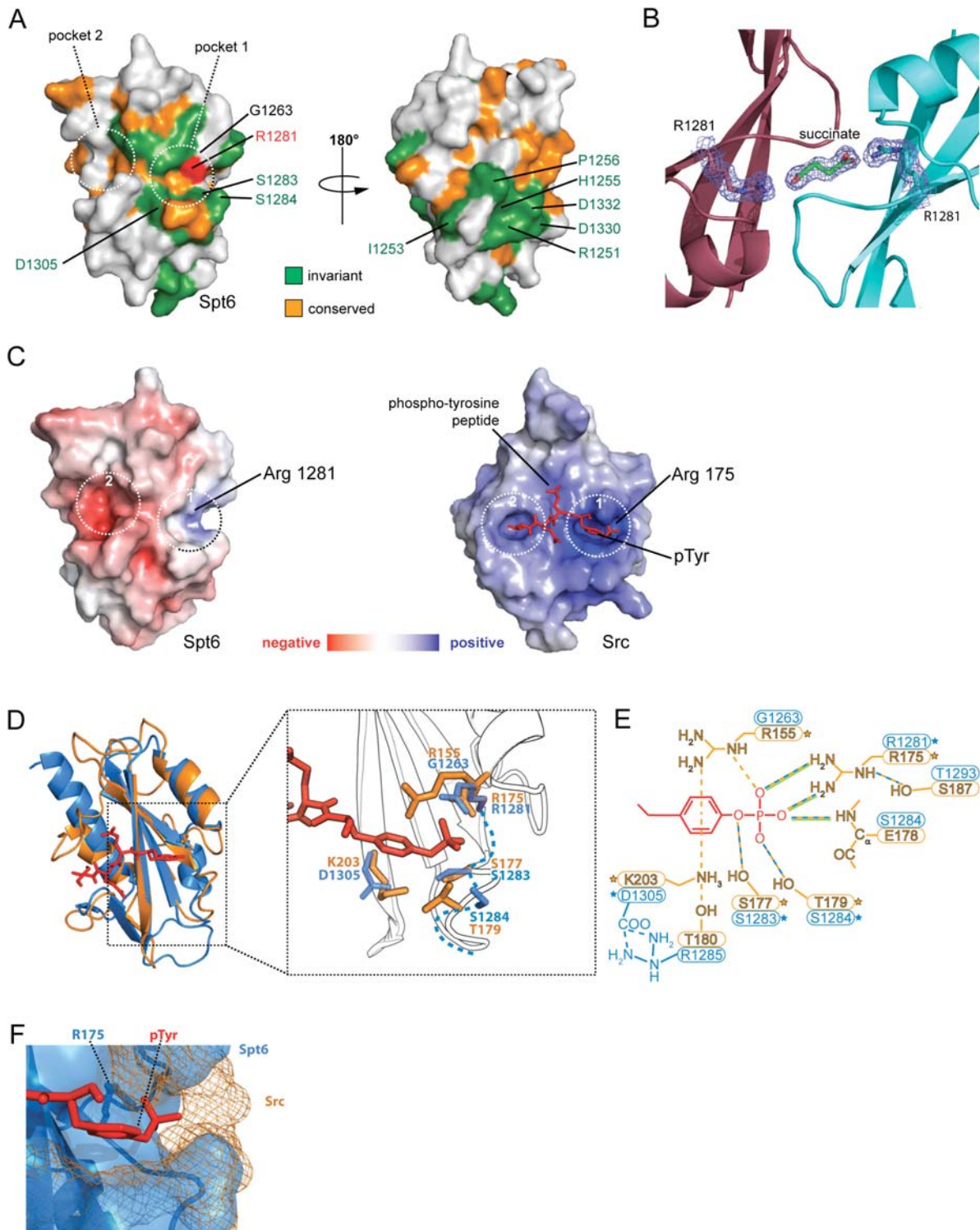
Figure 12: Surface properties of the Spt6 SH2 domain

(**A**) Conservation of surface residues. Residues are coloured according to the alignment in Fig. 9 A, with green for invariant and orange for conserved residues. The putative peptide-binding pockets 1 and 2 and surface residues implicated in peptide binding are indicated.

(**B**) 2Fo-Fc electron density map, contoured at 1σ, for the arginine residue R1281 of neighboring domains in the crystal and a bridging succinate ion.

Figure 12 (continued)

(**C**) Surface charge distribution of the Spt6 SH2 domain (left) and the Src SH2 domain (Waksman et al, 1993). The phosphopeptide bound to the Src domain is shown as a stick model in red, only 4 residues of the original structure are shown. The charge distribution is calculated with APBS (Baker et al, 2001).

(**D**) Superposition of the Src SH2 domain-phosphopeptide complex structure (Waksman et al, 1993) and the Spt6 SH2 domain (blue). Structures were aligned using the most conserved residues (G1276-R1281, L1290-K1295, Q1302-E1308) of the central β-sheet strands. The β-sheets aligned very well, resulting in a RMSD of 0.4 Å for 19 Cα atoms. On the right, a close-up view of the phospho-binding site is shown. Src residues that interact with the phosphotyrosine are in orange, and corresponding residues in Spt6 are in blue. The dashed blue line indicates the position of loop βB-βC of Spt6.

(**E**) Conservation of the phospho-binding pocket. The modeling of (D) was used to delineate the Spt6 residues predicted to be involved in phospho-group binding based on the Src SH2 domain-peptide complex structure. Interactions of Src residues with the phospho-tyrosine (Waksman et al, 1992) are in orange, and potential corresponding interations of the Spt6 residues with the phospho group are in blue. Phospho-mimetic interactions of residues with succinate are in green. Stars indicate the residues depicted in (D).

(**F**) Different shape of the phospho-binding pocket. The molecular surfaces of the phospho-binding pockets in Spt6 and Src are in blue and orange, respectively, and the phospho-tyrosine residue of the peptide bound to Src is in red.

3.3.8 A model for CTD binding

We created a model for possible interactions with the CTD (Fig. 13). The model is based on the interactions of the human Src-protein with its high affinity target peptide (Waksman et al, 1993). In this structure, the peptide runs across the binding surface from pocket 1 to pocket 2 (Fig. 12 C, right), the polaritiy is dictated by the interaction of the phosphotyrosine with pocket 1 and an isoleucine C-terminal of the pY with pocket 2. This binding mode is used by most SH2 domains ("two-pronged plug", see Fig. 3 C) and likely by the SH2 domain of Spt6, since a stretch of conserved surface residues extends from pocket 1 to a second pocket (pocket 2, Fig. 3 A).
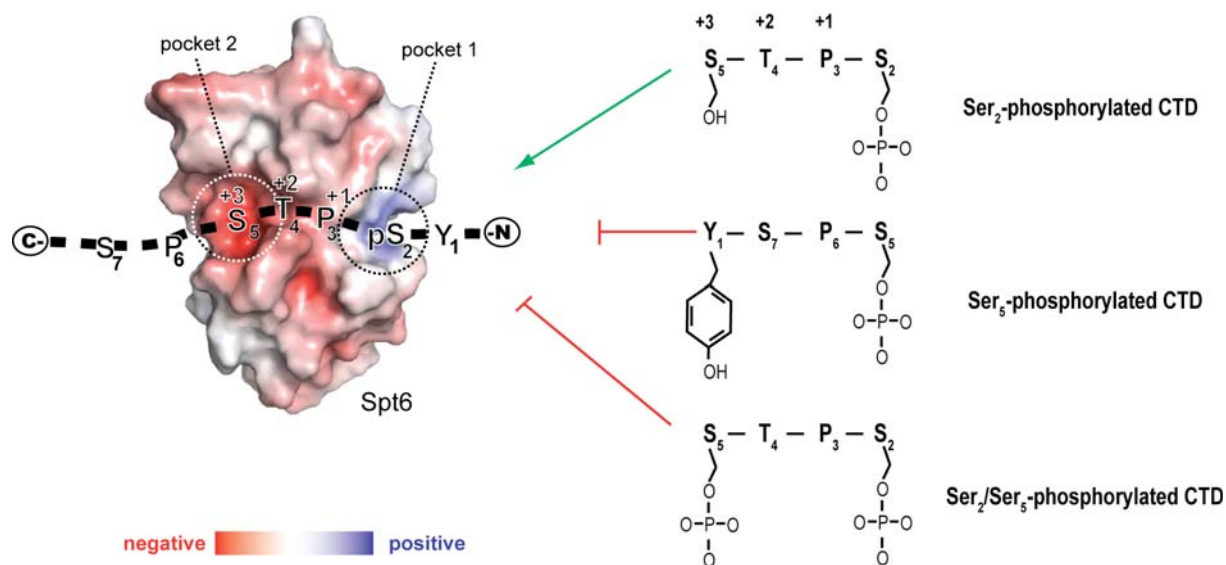
Figure 13: Model for binding of the Spt6 SH2 domain to the Pol II CTD

Suggested path for CTD phosphopeptide binding to the SH2 domain (dashed black line) and implications for specific binding to the Ser2-phosphorylated CTD. The Spt6 SH2 domain is shown with its surface charge distribution as in Fig. 12 C.

In Src, mainly pocket 2 is important for substrate specificity (Kimber et al, 2000; Waksman et al, 1993) and accommodates the residue three residues C-terminal of the phospho-tyrosine (register +3, Fig. 13). Pocket 2 is negatively charged in Spt6, whereas it is slightly positively charged in Src (Fig. 12 C) and is hydrophobic or positively charged in most other SH2 domains. Amongst ten Src-type and four STAT-type SH2 domains that we picked randomly (from Table 17), pocket 2 was negatively charged in only two domains and otherwise generally hydrophobic or positively charged (Appendix Fig. A2). The unusual negatively charged pocket 2 may be important for specific phosphopeptide binding. Assuming that a S_2-phosphorylated CTD peptide runs along the same path as in the Src-peptide complex, the CTD residue Ser_5 at register +3 from the phosphoserine could bind into pocket 2. However, the presence of 2 phospho-aminoacids (S_2 and S_5) that could potentially bind into pocket 1, makes the situation more complex in terms of selectivity of binding. S_5 can also bind into pocket 1, which would bring Y_1 of the CTD heptad repeats into position (register +3) to interact with pocket 2. Selectivity could be accomplished, by sterically rejecting the large, aromatic side chain of tyrosine from the shallow, negatively charged pocket (in comparison to the small, polar side chain of serine in case of S_2-phosphorylated CTD). In case of doubly phosphorylated CTD at S_2 and S_5, the S_5-phosphate would be in register +3. The negative charge of phospho-Ser_5 at the selective position would be repelled by the negatively charged pocket 2. (Fig. 13) This model can explain why the Spt6 SH2 domain exclusively binds to Ser2-phosphorylated CTD and not to CTD phosphorylated at Ser_5 (Yoh et al, 2007).

3.3.9 CTD peptide soaks using existing Spt6 SH2 crystals

To test our model of Spt6-CTD interactions we attempted to solve the complex structure. An efficient method to solve the structure of protein-ligand complexes is to „soak" a small binding molecule into the crystals of the protein. This can be accomplished because of the large solvent content of protein crystals and a relatively good accessibility of protein surfaces that are not involved in crystal contacts.

The „SeMet mutant CD crystal form 1" (Fig. 7 B) was not suited for this task, since succinate molecules were occupying the phosphate-binding pocket (pocket 1), bridging between two molecules of the SH2 domain by binding to the highly conserved R1281, thus making crystal contacts (Fig. 12 B). Indeed, incubation of these crystals with CTD peptides in a control experiment led to the solvatation of the crystals. Likewise, soaking trials with native and „SeMet mutant CDE" crystals (Fig. 7 B) were not successful because incubation with the peptides at different concentrations for different times, led to a loss of diffraction. The crystals of „SeMet mutant CD crystal form 2" (Fig. 7 B) showed a good behaviour, stable diffraction, and their structure was already solved by molecular replacement, which showed that peptide binding pocket 1 of 2 molecules in the asymmetric unit of the crystal was pointed towards solvent and free to bind the target peptide. Unfortunately, these crystals were grown in high ionic strength (4,3 M NaCl, Table 13 and Fig. 7 B). Since the binding of the SH2 domain with the CTD-peptides is mainly based on ionic interactions, these conditions were classified as not suitable for soaking experiments. Thus, we tried to prepare these crystals for soaking, by gradually exchanging the high salt crystallization buffer to a low salt soaking buffer. This screening procedure is described in section 3.2.7. A condition was found were crystals cracked immediately upon transfer, but healed after short time (Fig. 14). This condition was chosen as a starting point for further screening, since the observed healing suggested that the crystal lattice could recover in these conditions and seemed to be intact.
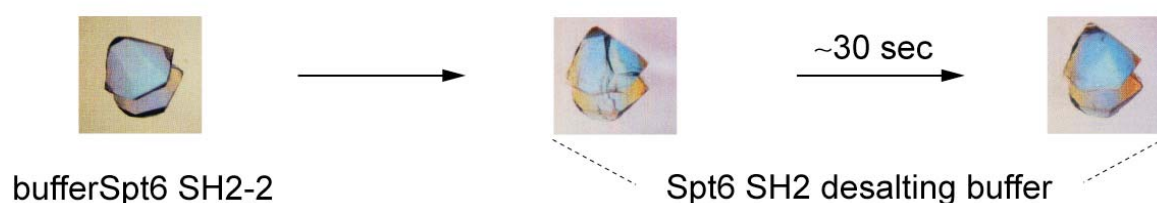

Figure 14: Crystal healing after transfer to low salt buffer
A crystal that was grown in buffer Spt6 SH2-2 (left, see 2.1.4, Table 13) which was transferred directly into Spt6 SH2 desalting buffer (middle) is shown. The crystal showed large cracks upon transfer, which healed about 30 seconds afterwards (right).

Finally, the crystals could be directly transferred to a buffer containing ~900 mM NaCl, stabilized by ~42.5% PEG 2000 containing phosphorylated CTD peptides for one hour without extensively decreasing diffraction quality (3.2.8). Two datasets were obtained with this method, one of a crystal soaked with peptide P1 and one soaked with peptide P2 (2.1.2, Table 6).

The structure was solved by molecular replacement as described in 3.2.5 for crystal form 2, with the exception that the structure was not refined to the end (see Table 18 for statistics). After the first round of refinement, the electron density map was inspected for positive difference electron density in the binding-pocket of the SH2 domain, which would indicate binding of the soaked peptide. The high R_{free} at this point of refinement can be explained by the large rearrangements in loop BC and DE (Fig. 9 C).

Since no density was observed in both datasets that could be assigned to peptide, the refinement procedure was cancelled. The failure of the experiment can be explained by the still high salt concentration in the soaking condition (~900 mM) which might disrupt phosphate binding. Crystals in conditions of lower ionic strength could not be obtained due to a loss of diffraction. Another explanation is that the SH2 domain variant in the crystals as well as the synthetic peptides used might not be sufficient to confer binding. This is adressed in section 3.3.10.

Table 18: : X-ray diffraction and refinement statistics for desalted crystal form B, soaked with synthetic peptides P1 and P2

	soak P1	soak P2
Data collection		
Space group	P32	P32
Cell dimensions		
a, b, c (Å)	70.4, 70.4, 88.8	70.2, 70.2, 89.1
α, β, γ (°)	90, 90, 120	90, 90, 120
Wavelength (Å)	1.00841	1.00841
Resolution (Å)	40-2.7	40-3.0
Rsym (%)	4.4 (94.7)	4.8 (76.9)
I / σI	35.0 (2.2)	34.5 (2.6)
Completeness (%)	99.0 (87.1)	99.1 (88.1)
Redundancy	11.3 (7.2)	11.4 (7.6)
Refinement	Molecular replacement as described in 3.2.5 then restrained refinement using Refmac5	
R_{work} / Rfree (%)	24.8 / 34.9	23.7 / 39.0

3.3.10 The SH2 domain is insufficient for binding short CTD phospho-peptides

We tested whether the Spt6 SH2 domain binds to short, syntehic CTD peptides that were phosphorylated at Ser_2 residues. We used the SH2 domains of Spt6 from *S. cerevisiae* (residues 1251-1351) and human (residues 1327-1427), together with two different peptides. The first one comprising a single CTD heptad repeat and a single phosporylated Ser_2 (peptide P1) and a tandem repeat with two phosphorylated serines (SPSYpSPTS and YpSPTSPSYpSPTSPS, respectively. See 2.1.2, Table 6). For fluorescence anisotropy experiments, peptide P2 was modified by a fluorescein molecule attached to the N-terminus of the peptide by a α-aminocaproic acid linker. An additional serine residue was added to the N-terminus in comparison to peptide P2, to assure enough space for the protein-peptide-interaction (P3, see 2.1.2, Table 6). We assayed binding by fluorescence anisotropy (3.2.9) and by surface plasmon resonance (Biacore, see 3.2.10), but could not detect significant binding. Since the published Spt6-CTD interaction (Yoh et al, 2007) used a C-terminal fragment of murine Spt6 that did not only include the SH2 domain, but also the adjacent C-terminal region (residues 1295-1496, whereas the SH2 domain spans residues 1327-1427), the SH2 domain alone and/or short CTD fragments are apparently insufficient for an interaction between Spt6 and Pol II. Since mutation of the invariant arginine residue abolished CTD binding, the SH2 domain is involved in the interaction but apparently requires its flanking regions and/or multiple phospho-CTD repeats. Very recent results confirm our findings. Yoh and colleagues (2008) show, that a C-terminal fragment of Spt6 containing the SH2 domain does not bind syntethic Ser_2-, and Ser_2-Ser_5-phosphorylated tandem CTD-repeats. In addition, they report that Spt6 binds selectively to the N-terminal and not the C-terminal half of the CTD *in vitro*, dependent on phosphorylation by P-TEFb. This region of the CTD contains 15 consensus tandem repeats of the CTD motif. This data indicates that a higher ordered CTD structure is required for binding of Spt6, consistent with the finding that the CTD can adopt different conformations induced by its binding partner (Meinhart & Cramer, 2004).

3.3.11 Functional architecture of Spt6

The central region of Spt6 shows homology to the bacterial Tex protein, which is involved in toxin gene expression (Fuchs et al, 1996; Johnson et al, 2008). The recent X-ray structure of Tex revealed the folds and relative position of domains HtH, YqgF, HhH, and S1 in the conserved central region (Johnson et al, 2008). Spt6 additionally contains an acidic N-terminal region, which might interact with nucleosomes, the C-terminal SH2 domain, and short regions flanking the SH2 domain. The previously published model of Spt6 (Johnson et

al, 2008) can now be extended with our structure of the SH2 domain (Fig. 15). To create an updated model, we aligned Spt6 and Tex protein sequences (Appendix Fig. A1) and mapped the 13 resulting sequence insertions in Spt6 onto surface regions within the Tex structure (Fig. 15), and added the structure of the SH2 domain. The model illustrates the relative sizes and locations of the various regions of the modular Spt6 protein, and is consistent with the idea that one side of Spt6 is involved in recruiting Spt6 to the transcribing Pol II via interactions with the nascent RNA and the phosphorylated CTD, whereas the other side is involved in nucleosome reassembly.
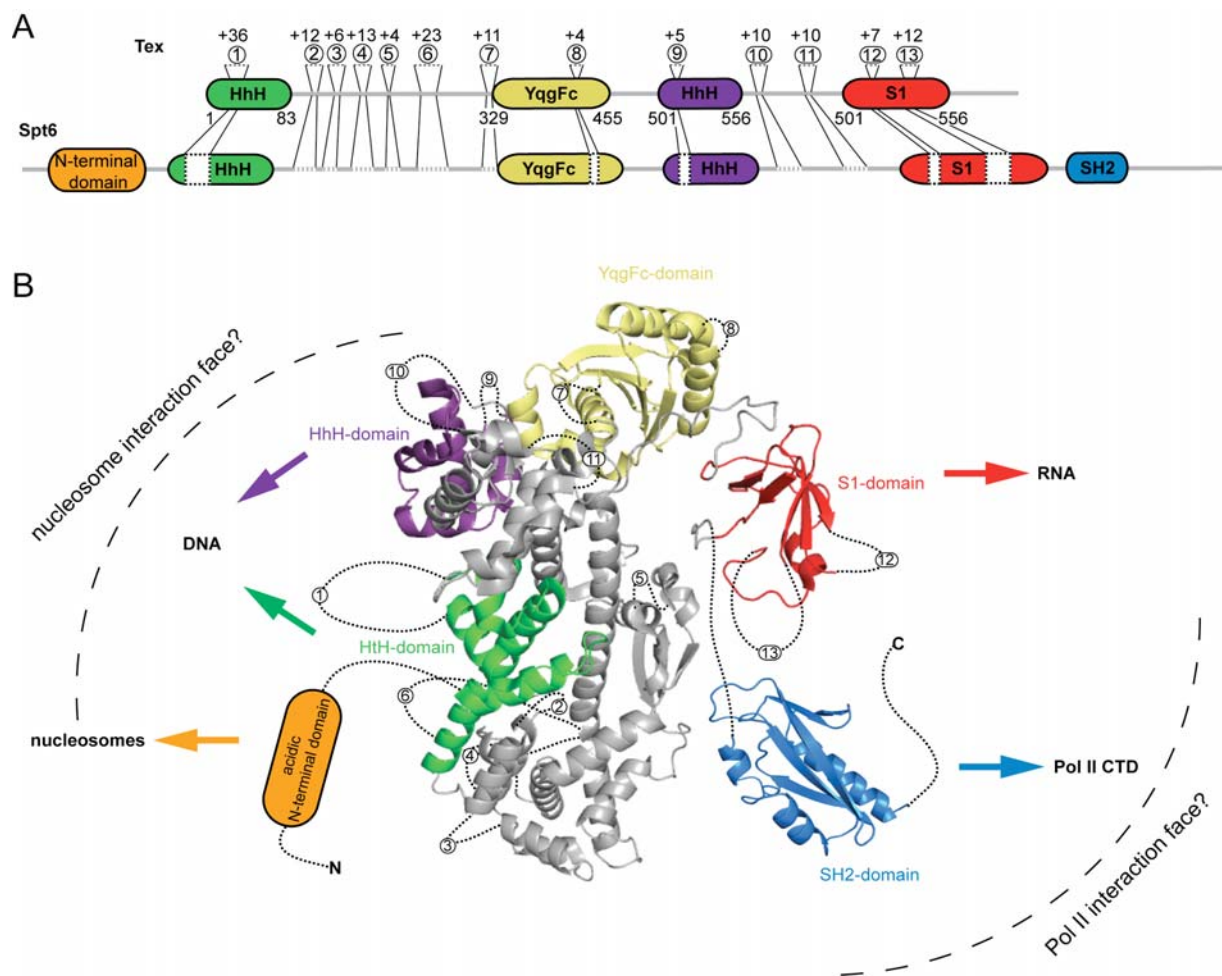
Figure 15: Updated structural model of Spt6

(**A**) Schematic representation of the Tex domain organization. The relative positions of insertions in eukaryotic Spt6 with respect to the bacterial Tex protein were identified by aligning the *C.glabrata* Spt6 and the *P. aeruginosa* Tex sequences (see Appendix Fig. A1) and are indicated with encircled numbers. The number of additional amino acid residues in Spt6 is also indicated. Boundaries of the Tex domains are indicated with numbers below the schematic representation.

(**B**) ribbon representation of the model for Spt6. Insertions in Spt6 are indicated by dashed lines. The relative position of the SH2 domain is arbitrary. Arrows point to potential interaction partners of the different domains.

3.3.12 The Spt6 SH2 domain has a widespread function in vivo

Results in this section are the work of Andreas Mayer and are presented here for discussion. To investigate the importance of the Spt6 SH2 domain *in vivo*, we carried out Affymetrix gene expression profiling (Affymetrix GeneChip Yeast Genome 2.0) with a yeast strain lacking the C-terminal region of Spt6 that includes the SH2 domain (Youdell et al, 2008) (strain spt6ΔC, 2.1.1, Table 3). Compared to a wild type strain, 790 out of 5665 genes that were present on the array showed significantly altered mRNA levels using a fold-change cut-off value of greater +2.0 or smaller than -2.0. Thus the Spt6 SH2 domain is necessary for the regulation of a subset of genes (14%) in *Saccharomyces cerevisiae*. The extent of deregulation of gene expression is comparable to strains carrying deletions of other Pol II elongation factor genes, including genes encoding subunits of the Paf1 complex (13% for paf1 and 15% for ctr9) (Penheiter et al, 2005). Of the mRNAs with significantly altered levels, 465 were up-regulated and 325 were down-regulated (Fig. 17 A), suggesting a repressive function of the SH2 domain at a majority of genes. Western blotting revealed that yeast cells adapt to the deletion by increasing the Spt6 protein levels (Fig. 16), which could compensate a failure of the mutant protein to localize to the transcription machinery. However, this does not restore the wild type phenotype in terms of growth, since the mutant strain shows a slow growth phenotype. Elevated Spt6 levels in the mutant can also root in a less efficient degradation of the TAP-tagged mutant protein. We were not able to resolve this problem because no antibody against the yeast Spt6 protein is available.
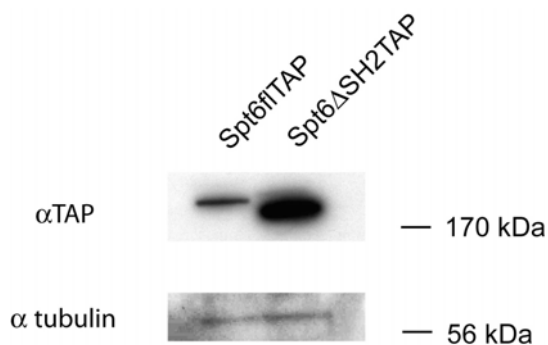
Figure 16: The cellular concentration of Spt6 increases when SH2 containing C-terminus is deleted
Western Blot of Spt6 proteins with a C-terminal TAP-tag, resolved in a 8% SDS-PAG. The mutant variant of Spt6 shows elevated expression levels relative to wild-type Spt6 and tubulin.

We next analyzed biological processes that were significantly affected by deletion of the SH2 domain with the Gene Ontology Enrichment Analysis Software Toolkit (GOEAST) and the web-based Gene Ontology (GO) tools (Ashburner et al, 2000; Zheng & Wang, 2008). This analysis showed that very diverse biological processes were over-represented, including genes involved in the response to toxins, in copper ion transport, and in thiamin metabolic processes.

A deconvolution of gene expression microarray data is generally difficult as it represents the result of primary and secondary effects during gene expression. We nevertheless aimed at detecing a possible global, chromatin-related function of the SH2 domain in the transcriptome data, with use of correlation analysis. We first investigated whether the deregulated genes correlate to genes that were previously described to show cryptic transcription initiation in a spt6 mutant that carried an internal deletion of amino acids 931-949 (corresponding to 930-993 in *C. glabrata* and comprising the HhH-domain marked purple in Fig. 15). From the 960 ORFs that showed cryptic transcription (Cheung et al, 2008), only 147 were included in our spt6ΔC differential gene expression profile (Appendix Fig. A3). We also investigated whether our set of differentially expressed genes shows any correlation with gene length or an unusual number of associated nucleosomes (Lee et al, 2007). However, we could not find any significant correlations. These results are consistent with the view that Spt6 has multiple functions and is not only required for nucleosome assembly but also for mRNA splicing and

export (Yoh et al, 2007), and that it contains different functional surfaces (Fig. 15) that are perturbed in the different mutants.

To address the problem of secondary effects that influence the microarray data, we compared the differentially expressed genes in the spt6ΔC mutant to a list of transcription factors (Hu et al, 2007). 33 genes of transcription factors are contained in our list of genes, which is more than 4% of all affected genes in the spt6ΔC mutant (Appendix Fig. A4). This suggests that a number of alternatively expressed genes can probably be related to secondary effects induced by the Spt6 mutation. With the tools available, it is currently not possible to reveal details of this defective regulation networks.

To analyze whether the expression of similar genes is affected by deletion or mutation of different Pol II elongation factor genes, we compared our gene expression data to available data for the yeast strains *dst1Δ* (*DST1* is the gene encoding TFIIS, (Koschubs et al, 2009), *spt4Δ*, and *rtf1Δ*, (Hu et al, 2007). We expected similarity between these data sets since Spt6 interacts genetically with TFIIS (Hartzog et al, 1998) and with the Rtf1-containing Paf1 complex (Costa & Arndt, 2000; Mueller & Jaehning, 2002), and since Spt6 binds the Spt4-Spt5 complex (Krogan et al, 2002). An unsupervised hierarchical cluster analysis showed that the differential expression data from the *spt4Δ* and *rtf1Δ* strains form a distinct cluster within a dendrogram, indicating similarity of their gene expression profiles (Fig. 17 B, lanes 2 and 3, Methods). However, the *spt6ΔC* mutant exhibits a very different expression profile (Fig. 17 B, lane 1), suggesting that the function of the Spt6 SH2 domain is clearly distinct from the functions of Spt4-Spt5 and the Paf1 complex *in vivo*. This analysis additionally revealed that *dst1Δ* showed the most distinct expression profile (Fig. 17 B, lane 4), maybe because TFIIS is not only required during elongation, but also during initiation (Guglielmi et al, 2007; Kim et al, 2007). Additional correlation studies confirmed these results.
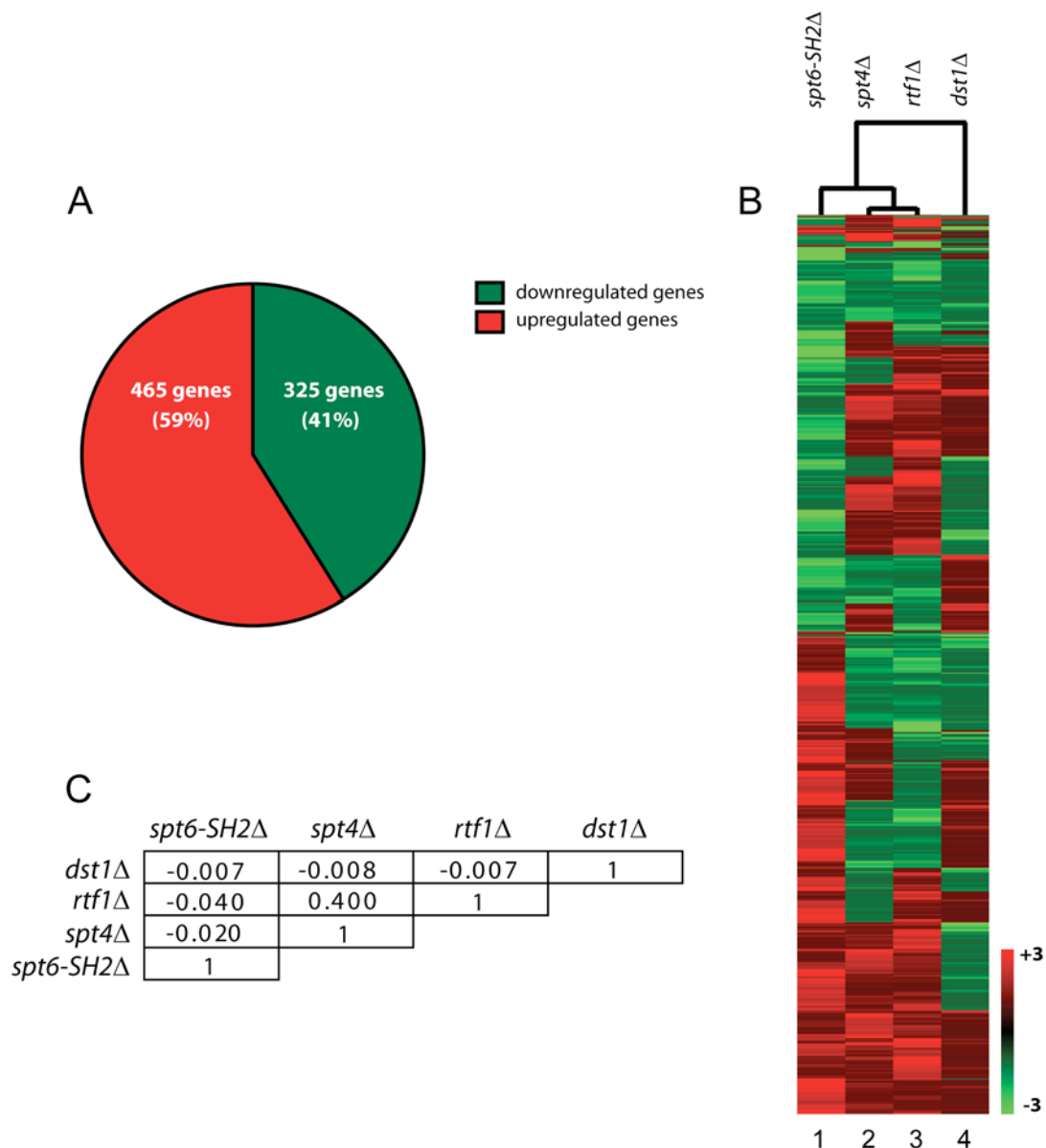
Figure 17: The Spt6 SH2 domain is required for normal gene expression *in vivo*

(**A**) Differentially expressed genes as detected by microarray analysis of the *spt6ΔC* strain. The fraction of up- and down-regulated genes is shown in red and green, respectively. In total, 790 genes showed significantly altered mRNA levels between Spt6 wild-type and the *spt6ΔC* strain. All gene expression analysis was performed with biological duplicates.

(**B**) Cluster analysis of differential gene expression profiles of different Pol II elongation factor deletion strains. The cluster diagram was calculated for 1350 yeast genes of *spt6ΔC*, *spt4Δ*, *rtf1Δ* and *dst1Δ* mutant strains, depicted in lanes 1, 2, 3, and 4, respectively. Each row corresponds to a particular gene and each column corresponds to a particular elongation factor mutant. Changes in mRNA levels compared with the isogenic wild-type strain are depicted in red (increase), green (decrease) or black (no change; see intensity bar). Both rows and columns were clustered using a hierarchical cluster algorithm (Saeed et al, 2003). The dendrogram for column clustering is shown.

(**C**) Pearson's correlation matrix for gene expression profiles of yeast strains *spt6ΔC*, *spt4Δ*, *rtf1Δ* and *dst1Δ*. The corresponding correlation coefficients are given.

The figure was kindly provided by Andreas Mayer.

A weak correlation was detected between expression profiles of *spt4Δ* and *rtf1Δ* strains, but no significant correlations were seen between the profiles of the remaining mutant strains (Fig. 17 C). Taken together, deletion or mutation of various elongation factor genes results in different changes in the transcriptome, despite the observed genetic and physical interations between these factors.

Taken together, this data shows that the deletion of the Spt6 SH2 domain from the genome has a dramatic effect on gene expression. It is currently not resolved if this observation is mainly due to the absence of the domain and a resulting defect of Spt6 to recognize the Ser_2-phosphorylated CTD of Pol II, or to raised Spt6-levels in the mutant strain. However, this data shows the general importance of Spt6 and especially of its SH2 domain for correct transcription of the yeast genome.

3.3.13 Spt6 co-localizes with Pol II on the yeast genome

The occupancy profile of C-terminally TAP-tagged Spt6 on the yeast genome showed a high correlation to the profiles of Pol II subunits Rpb3 (Pearson correlation coefficient 0.86) and Rpb7 (Pearson correlation coefficient 0.87) (Jasiak et al, 2008). Thus, the pearson coefficient is only slightly lower than the correlation coefficient of Rpb3 and Rpb7 (0.91), which are both part of the same multisubuent complex. This means that Spt6 localizes together with Pol II on

Figure 18: Genome-wide occupancy profiling of transcription elongation factor Spt6

A representative 230 kilobase pair sample on chromosome one (genomic positions 0-230.000) of the profiles for Rpb3 and Spt6 are depicted. Each green dot represents the signal for a single oligonucleotide probe on the tiling array, which has one probe every 32 bp. An example for a region in the genome with different Pol II/Spt6 occupancy is marked red.

(**A**) Difference signal between averaged Rpb3 (B) and Spt6 (C) occupancy profiles.

(**B**) Average over three biological replicate traces for Rpb3, one of which with interchanged fluorescent dyes (Jasiak et al., 2008)

(**C**) Average over two biological replicate traces for Spt6, both with interchanged fluorescent dyes

(**D**) Genomic features based on the Saccharomyces genome database annotations

The figure was kindly provided by Matthias Siebert

a genome-wide scale. The significance of these profiles is described in Jasiak et al (2008): housekeeping genes as well snoRNA genes show a high occupancy, which is expected expected, because these genes are highly expressed. tRNA-genes, that are expressed by RNA polymerase III show no increased occupancy. Fig. 18 shows a representative 230 kilobase sample of the profiles for Spt6 and Rpb3. The profiles are virtually similar, but show slight differences in some positions (Fig. 18, marked red).

We asked, if our data shows recruitment of Spt6 relative to Pol II that parallels Ser_2-phosphorylation of the CTD. We would expect to observe an increase of Spt6 occu-pancy towards the 3' end of the transcribed regions, since it was shown that the SH2 domain binds exclusively to the CTD when phosphorylated at Ser_2 residues (Yoh et al, 2007), which predominate only in 3' regions of a gene (Komarnitsky et al, 2000). We could not observe this general behavior of Spt6 in the occupancy profile.

To further analyze this question, we asked if the effect of an interaction between the CTD and the SH2 domain can be seen at genes of similar length, when the ChIP signals of these genes are averaged. For this, based on the Rpb3-data, we filtered out genes with an average ChIP signal < 0.2 to eliminate genes that are not transcribed. In addition, overlapping genes (both on the same strand and on the opposite strand) were eliminated to get clean signals from single genes. For the same reason we excluded neighboring genes that were too close to assign clean signals because of a "smearing" of ChIP signals in 5' and 3' regions. This resulted in 638 genes that showed a distribution in length as shown in Fig. 19 A. We used those classes for our analysis, that contained a sufficiently high number of genes (green in Fig. 19 A).

The results of our analysis are shown in Fig. 19 B. The averaged Rpb3 ChIP-signal is high around the Start-codon (green, dashed line) and has its maximum shortly after that. This behavior can be seen in all classes of gene length and can be explained by polymerases paused at promoter proximal sites. In contrast to that, the averaged Spt6 signal shows a delay in comparison to the Pol II occupancy, with increasing density toward the 3' end of transcribed regions. This effect decreases with increasing gene length. Thus, Spt6 shows an occupancy profile that roughly parallels Ser_2-phosphorylation of the CTD, but only when the signals of many genes of the same length are averaged. Since it is crosslinking efficiency that is directly measured in a ChIP-experiment this effect could be explained by a change in crosslinking efficiency of Spt6 relative to Rpb3. Since it can only be seen in the averaged data, the effect seems to be very subtle and can rather be explained by a rearrangement between Spt6 and Rpb3 (or Pol II) than by a recruitment of Spt6 to the transcription machinery by Ser_2-phosphorylated CTD.

Unexpectedly, the Spt6 occupancy has a peak around the Stop codon and shows a prolonged signal beyond the end of the coding region in comparison to Pol II (Fig. 19 B). To test if this is significant or an artifact of our data processing, we averaged the signals of all genes, independent of their length, around the Start codon (+/- 500 bp) and the Stop codon
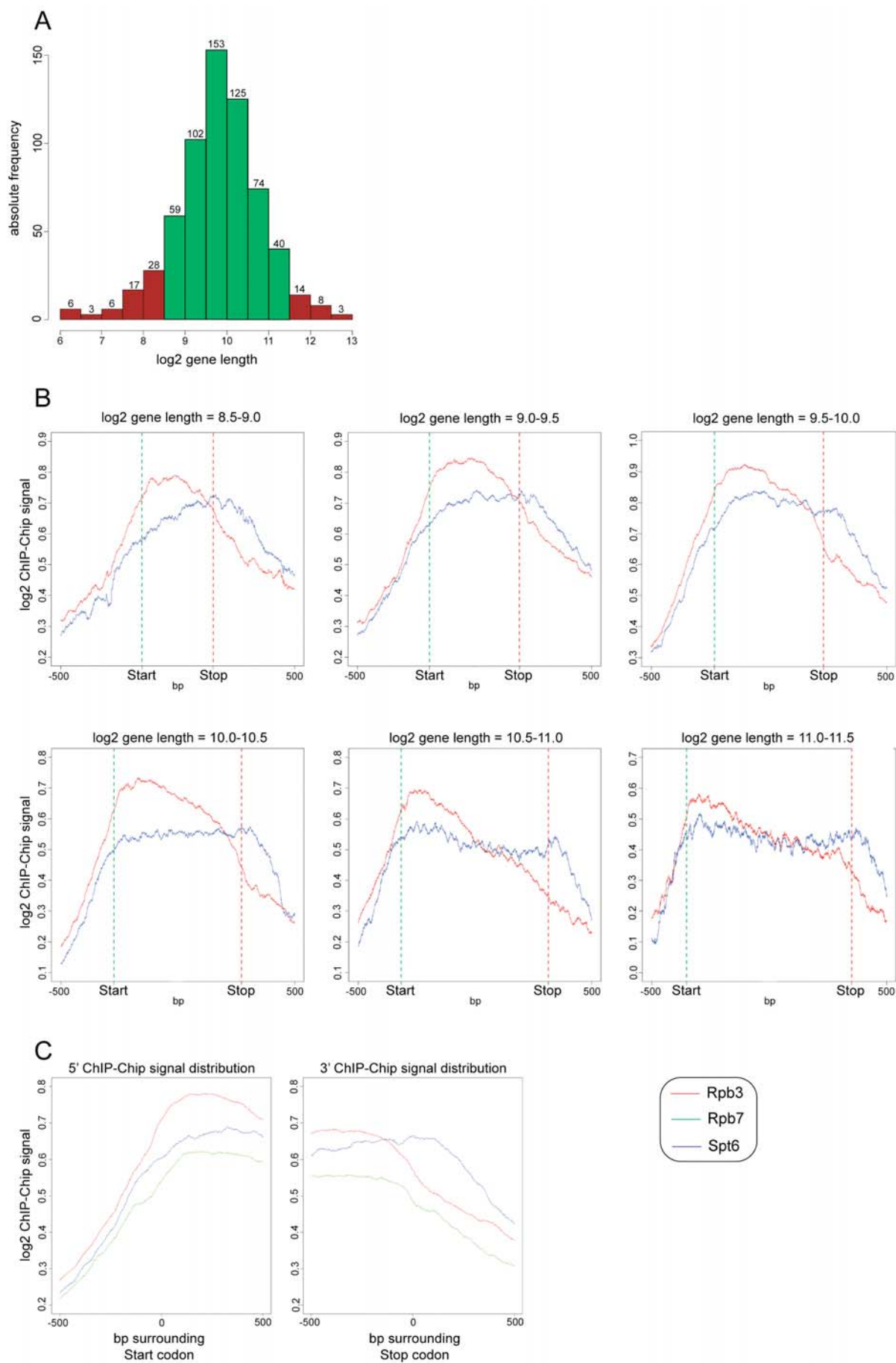
Figure 19 (preceding page): Differences in averaged occupancy profiles of Pol II and Spt6

(**A**) Distribution of genes into classes of different length after filtering out untranscribed, overlapping and closely spaced genes (see text for details). Length classes that were used for further analysis are marked green.

(**B**) Plots of the averaged ChIP-on-chip signals of the different length classes from (A). Classes are indicated above the plots. Averaged Rpb3- and Spt6-ChIP signals are shown in red and blue, respectively. The absolute values of different datasets cannot be related to each other.

(**C**) Plots of averaged ChIP-on-chip signals of the Rpb3-, Rpb7- and Spt6-datasets around regions of the Start and the Stop codon. The ChIP-on-chip signal is averaged from all genes. Rpb3-, Rpb7- and Spt6-signals are shown in red, green and blue, respectively. Rpb3- and Rpb7-ChIP-on-chip data is from Jasiak et al (2008).

The plots were kindly provided by Matthias Siebert.

(+/- 500 bp). For this we used the Rpb3 and the Spt6 datasets and in addition the data of the Rpb7 ChIP-on-chip experiment (Jasiak et al, 2008). The ChIP-data from Rpb7 is independent from the Rpb3-data, but both proteins are Pol II subunits. Indeed, Rpb3 and Rpb7 show similar profiles proximal to the Start and the Stop codon, whereas Spt6 again shows a maximum around the Stop codon and a longer retention in the 3' regions of the genes relative to Pol II, which is strengthening the significance of our data analysis.

3.4 Conclusions and future perspective

In this work, the structure of the SH2 domain of transcription elongation factor Spt6, including a deep analysis of the biological implications of the structure is presented. The results are highly informative for different reasons: First, Spt6 is an essential regulator of transcription elongation through chromatin and the structure of its SH2 domain is the first structural information available for this factor. Second, the Spt6 SH2 domain is the only SH2 domain present in the yeast genome, in contrast to the great variety of these modules we have in human cells (>100 domains). Thus, the Spt6 domain represents the „minimal equipment" of a lower eukaryotic organism in terms of SH2 domains. By analysis of our structure, we were able to put the Spt6 domain in a evolutionary relationship to known structures. This revealed that the domain exhibits features of „modern" SH2 domains that are involved in cell signaling pathways. This makes the Spt6 domain an ancestor of this important modules. Thus, our data gives an exciting insight into the molecular evolution of SH2 domains.

Furthermore, the Spt6 SH2 domain is of special interest because it connects Spt6 directly to the transcription machinery and thus is of great importance for its function. It shows a different specificity in that it binds to Ser_2-phosphorylated CTD, wheres all other SH2 domains on which structural information is available bind to tyrosine-phosphates. Unfortunately, we were not able to elucidate functional information of the interaction of the SH2 domain and the CTD due to the limitations described. A recent publication showed that

the interactions between the Spt6 SH2 domain and the CTD might be more complex compared to the interactions of other SH2 domains and their targets, as higher order CTD-structures seem to be involved (Yoh et al, 2008). Nevertheless, models emerged from our analysis that are experimentally testable as soon as we have obtained the right combination of SH2 domain variant and CTD-peptide to establish a binding assay. Interestingly, we have found a second $\alpha\beta\beta\beta\alpha$-motif directly C-terminal to the SH2 domain. Thus we suspect the presence of a tandem SH2 domain in the C-terminal region of Spt6. We have already initiated work to tackle this question, which led to the crystallization of the whole *C. glabrata* Spt6 C-terminus. Consequently, we will soon learn more about the molecular interactions between the transcription machinery and elongation factor Spt6.

We showed the importance of the Spt6-SH2 domain for the correct transcription of the yeast genome. It is at this point still difficult to decide if the differential gene expression we see stemms from the absence of the domain or from the resulting change in cellular Spt6 concentrations, but we initiated work to decipher this problem. Furthermore, a consistent interpretation of such results is still difficult as we show by our analysis of the transcription factors that are affected in their expression levels in the spt6ΔC mutant (Appendix Fig. A4). This means that a considerable number of the altered expression levels in the dataset are likely produced by secondary effects rather than by the mutation itself. In future, computational tools are needed, that integrate knowledge about cellular regulation networks with the genome-wide expression data to unravel these complex interdependency. It is however fascinating to see the dramatic changes in gene expression upon the deletion of a 100 amino acid module from an elongation factor. This in principle shows the tight and minute regulation of transcription at the level of elongation.

We created a map of Spt6 occupancy on the *S. cerevisiae* genome and compared it to the localization-map of Pol II that was previously published from our lab (Jasiak et al, 2008). The result was unexpected, as we saw Spt6 localizing basically to the same positions as Pol II throughout the genome. In light of the general importance of Spt6 for various processes like chromatin maintenance (Bortvin & Winston, 1996), histone methylation (Yoh et al, 2008) and mRNA processing and export (Yoh et al, 2007) it seems reasonable that Pol II keeps important factors like Spt6 in close proximity for the efficient transcription of genes. Because Pol II- and Spt6-localization was so similar, we could initially not detect recruitment of Spt6 paralleling the Ser_2-phosphorylation of the CTD of Pol II as might be expected from the SH2 domain - CTD interaction. Only by averaging the signals of many genes a slight effect could be detected that seems not to be an artifact of data processing. This effect could be explained by a subtle change in the crosslinking behaviour between Pol II and Spt6 due to a rearrangement within a higher order structure. However, the analysis of complex genome-wide data is still developing and our data analysis suggests a way to extract information that is hidden in the mere occupancy profiles of proteins. The significance of this procedure still has to be evaluated, for example by comparison to the localization profiles of other

elongation factors and to gene specific factors. It will be interesting to see, if proteins that interact with Ser_5-phosphorylated CTD show a similar, but inverted behavior like Spt6 in our data. In addition, such data could be useful to define factors that generally localize together with Pol II on the genome in a higher ordered structure like a „transcription factory".

In summary, a combined, interdisciplinary approach as outlined in this part of the thesis will be increasingly powerful to understand proteins and their interactions in a broad context. However, the development of more powerful computational methods to analyze the complex data derived from genome-wide investigations will be needed to uncover hidden layers of information that will add up on the knowledge that will be gained on the interactions of cellular components.

4 An *in vitro* system to test the "torpedo model" of transcription termination

4.1 Introduction

4.1.1 The 5'-3' exoribonuclease-complex Rat1/Rai and the "torpedo model" of transcription termination

The "torpedo model" of transcription termination was first mentioned in 1988 (Connelly & Manley, 1988). In their work, Connelly and Manley found that in their transient expression assays in human cells, that an intact polyadenylation signal is required for transcription termination by Pol II. They proposed a model, where after poly(A)-site-induced cleavage of the nascent transcript, downstream RNA is degraded by a 5'-3' exonuclease which, upon reaching the still elongating polymerase gives the signal to terminate. Although they did not use the term "torpedo model", it is still used to describe the events that could terminate transcription. The concept was further strengthened when it was discovered, that not only the poly(A)-signal is mandatory for termination (Connelly & Manley, 1988; Logan et al, 1987), but also the factors that accomplish cleavage (or at least some of them) are required for effective termination at the CYC1 gene, whereas mutations in polyadenylation-factors did not show severe effects (Birse et al, 1998). The "torpedo model" gained strong support, when a candidate nuclease was found. Affinity chromatography with Ser_2-phosphorylated CTD led to the purification of Rtt103. This protein contains a CTD-interaction (CID) domain that is related to that of Pcf11 (Meinhart & Cramer, 2004). In addition, the Rat1/Rai1 nuclease complex was found to interact with Pol II/Rtt103 (Kim et al, 2004). Rat1/Rai and Rtt103 crosslink together at the polyadenylation-site and crosslinking of Rtt103 is dependent on CTD-phosphorylation by Ctk1. This suggested that these proteins play a role in 3' end processing.

Rat1 is an essential, nuclear 5' exoribonuclease of 116 kDa, involved in RNA turnover. It has similar enzymatic activity as its cytoplasmic counterpart Xrn1, processively degrading RNA from the 5' end and releasing nucleoside 5'-monophosphates (Stevens & Poole, 1995). Xrn1 was the first 5'-3' exoribonuclease characterized in yeast (Larimer et al, 1992; Larimer & Stevens, 1990; Stevens, 1980; Stevens & Maupin, 1987). Rat1 is an essential protein, whereas cells with a disrupted XRN1 gene are viable (Kenna et al, 1993; Larimer & Stevens, 1990). Both exonucleases prefer RNAs with a 5' monophosphate as a substrate, show weak activity towards RNA with a 5'-hydroxyl group and single stranded DNA (ssDNA) and no

activity for RNAs with a 5' cap or tri-phosphate (Poole & Stevens, 1995; Stevens & Poole, 1995). Interestingly, Xrn1 and Rat1 are functionally interchangeable. When the proteins are directed to opposite compartments by the deletion or addition of a nuclear localization signal (NLS), they can complement defects of their counterpart (Johnson, 1997). In the first purifications of endogenous Rat1, an interacting protein of 45 kDa was seen that co-purified together with the Rat1-exonuclease activity (Stevens & Poole, 1995). This protein was later identified as Rai1 and it was shown that it stabilizes the exonuclease activity of Rat1 (Xue et al, 2000).

Cells with an intact cleavage/polyadenylation machinery, but with a mutated and non-functional Rat1 (Amberg et al, 1992) or a deletion of its co-factor Rai1, showed dramatic termination defects (Kim et al, 2004). In these cells, 3' transcripts after poly(A)-site directed cleavage were greatly stabilized. Similar results were obtained in independent studies on the human β-globin gene . In this gene, the poly(A)-site is associated with an autocatalytic RNA-structure that undergoes rapid self-cleavage after being transcribed (CoTC for co-transcriptional cleavage) (Teixeira et al, 2004). This cleavage was shown to create a free RNA 5' end that is a substrate for Xrn2 (the human homolog of Rat1), and that the consequent degradation of the downstream RNA induces termination (Teixeira et al, 2004; West et al, 2004). Only recently it was shown, that the Rat1/Rai1 activity is also required for efficient termination of Pol I transcription (El Hage et al, 2008). Here, the RNase III-like endonuclease Rnt1 cleaves the nascent rRNA-transcript to create an entry site for Rat1 (Kufel et al, 1999). A knockout of the Rnt1-gene showed an increased read-through at the „Reb-dependent" Pol I terminator (see section 1.3.2), showing that co-transcriptional cleavage is important for efficient termination at this site (Prescott et al, 2004; Reeder et al, 1999)

These results strongly support the "torpedo model" that is depicted in Fig. 20. This model somehow resembles Rho-dependent termination in prokaryotes (Nudler & Gottesman, 2002). There, Rho an ATP-hydrolyzing RNA-DNA-helicase follows bacterial RNAP along the RNA upon a specific signal (without hydrolyzing the RNA) and terminates transcription upon arrival at the elongation complex (see also 1.3.2).

4.1.2 "Torpedo model" vs. "allosteric model"

The "torpedo model" is not the only explanation of the events at the 3' end of genes. First the data was explained with the "allosteric model". Here, transcription of the poly(A) sequence is predicted to trigger a change in the factors that are associated with polymerase II (Logan et al, 1987). This could include the binding of a termination factor or the displacement of an anti-termination factor or both, for example by the cleavage/polyadenylation factors. In fact, there is data that supports this model over the torpedo model. First, poly(A) site-induced cleavage of the transcript, which is necessary for the explanation of the "torpedo model",

does not seem to be a prerequisite for termination in general: disruption of poly(A) site cleavage by mutants of Pcf11 and Ssu72 did not affect termination (Dichtl et al, 2002; He et al, 2003; Sadowski et al, 2003). In an EM study of plasmids injected into *Xenopus* oocytes and in a second one on *Drosophila* chromosomes it was seen that termination occurs before cleavage of the transcript, or that it can even happen without prior cleavage (Osheim et al, 1999; Osheim et al, 2002). Another study presented data using the RNA IP technique, showing that Rat1 and Xrn2 are indeed responsible for co-transcriptional degradation of RNA 3' of the poly(A) site, but are not the factors that cause termination (Luo et al, 2006). It was rather shown that Rat1 is a part of the recruitment cascade of 3' processing factors and by that involved in termination.
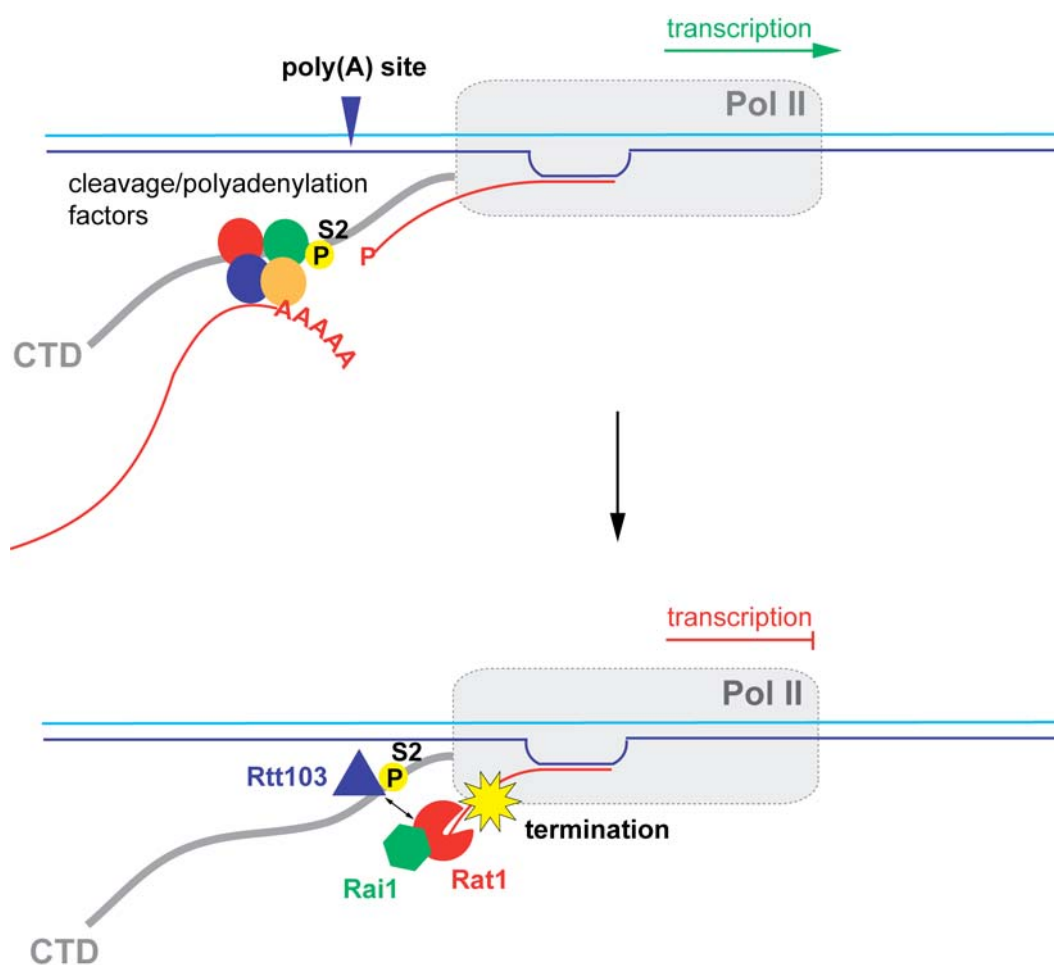
Figure 20: The "torpedo model" of transcription termination

After transcription of the poly(A) site, the nascent RNA is cleaved by cleavage/polyadenylation factors that are assembled on Ser$_2$-phosphorylated CTD. The cleavage creates a new, uncapped RNA 5' end, which is a substrate for the exoribonuclease-complex Rat1/Rai1 that interacts with the Rtt103-Pol II complex and is also assembled based on the Ser$_2$-phosphorylated CTD. Rat1/Rai1 is predicted to degrade the downstream RNA processively. Upon contact with the elongating polymerase, transcription termination is induced. Template DNA, non-template DNA and RNA are shown in blue, cyan and red, respectively. Factors in this figure are not drawn to scale. See text for details.

Consequently, the cytoplasmic counterpart of Rat1, Xrn1 (see section 4.1.2) can rescue the nuclease-activity associated defect of a Rat1 mutation when directed to the nucleus, but does not rescue the termination defect (Luo et al, 2006)

4.1.3 Aim of this work

Despite extensive research from different labs, the mechansim of transcription termination of protein coding genes is still one of the least understood events during the transcription cycle. However, over the years two models were established that try to explain the findings. As summarized in sections 4.1.1 and 4.2.2, existing data both support the "torpedo model" and the "allosteric model". This suggests that the truth about the events that lead to termination lies somewhere in between the two models. To approach this question, it is necessary to characterize the factors that are involved in isolation and to gain knowledge about their intrinsic properties. The "torpedo model" as such was never tested in a highly pure *in vitro* system, to see if Rat1/Rai1 actually possess the capability to disrupt elongation complexes. It is not clear how a nuclease can exert the force on polymerase to disintegrate its interaction with DNA and RNA, especially as it was shown that degradation of the RNA alone is not sufficient to end elongation (Gu et al, 1996). We were engaging this problem by setting up a highly defined *in vitro* elongation assay (described in 4.3.3). In addition a protocol for the production and purification of active, recombinant Rat1 exonuclease complexes was established (described in 4.3.1) and their nuclease activity was characterized (described in 4.3.2). With this in hand we tested transcription termiation as predicted by the the „torpedo model" (described in 4.3.4). The advatage of our assay is, that all the components are easy to exchange and to modify, so we were able to test the model using different types of DNA and RNA sequences that are involved in the processing of mRNA in 3' regions of genes (described in 4.3.5 – 4.3.7).

4.2 Specific procedures

4.2.1 Vectors

The following vectors have been created with the methods described in 2.2.1:

Table 19: Vectors containing Rat1, Rai1, Rtt103 genes

Vector	Source Plasmid (see 2.1.2, Table 4)
Rat1fl	pET21b(+)
Rai1fl	pET24b(+)
Rtt103fl	pET21b(+)
Rtt103ΔCID	pET21b(+)

4.2.2 Purification of Rat1, Rat1/Rai1 and Rtt103

Rat1 was purified from a 1 l autoinducing expression culture (see 2.2.3). After binding of Rat1, the Nickel-NTA-resin was washed with 20 ml Rat1 IMAC buffer (see 2.1.4, Table 11) + 30 mM imidazole to remove proteins binding unspecifically to the matrix. Elution of Rat1 was carried out with 30 ml of Rat1 IMAC buffer + 300 mM imidazole. The eluate from the Nickel-column was applied to a HiTrap Heparin 5ml column (GE Healthcare), pre-equilibrated with Rat1 heparin buffer A (see 2.1.4, Table 11). Bound protein was eluted by increasing the salt concentration in a linear gradient from 100 mM to 1 M NaCl over 12 column volumes. Peak fractions from the heparin affinity chromatography were pooled and concentrated to 500 µl, then applied to a Superose 6 10/300 GL column (GE Healthcare), pre-equilibrated with 2x Rat1 size exclusion buffer (see 2.1.4, Table 11). 2x Rat1 size exlusion buffer was RNAse free, whenever the enzyme was used for nuclease assays. 10% (v/v) RNAse-free glycerol was added to the sample before freezing in liquid nitrogen and storage at –80°C.

The Rat1/Rai1 complex was co-expressed in *E. coli* BL21 (DE3) RIL cells by transforming the two plasmids with different antibiotic-resistance cassettes at the same time. In the complex, only Rat1 has a C-terminal His$_6$-tag. The complex was purified from a 2 l autoinducing medium expression culture (2.2.3).

After binding of Rat1/Rai1 to the Nickel resin and after washing with 10 column volumes of the Rat1/Rai1 cell lysis buffer (see 2.1.4, Table 11), the protein was eluted with 40 ml Rat1/Rai1 IMAC buffer (see 2.1.4, Table 11) + 300 mM imidazole. The eluate from the IMAC was applied to a HiTrap Heparin column (5 ml, GE Healthcare) pre-equilibrated with Rat1/Rai1 IMAC elution buffer (see 2.1.4, Table 11). Bound protein was eluted in a linear gradient from 0.1 M to 1 M NaCl, buffered by 50 mM Tris/HCl pH 8.0 (Rat1 heparin buffer A

and Rat1 heparin buffer B, see 2.1.4, Table 11) over 12 column volumes. Peak fractions from the heparin chromatography were pooled, concentrated to 500 µl and applied to a Superose 6 10/300 GL column (GE Healthcare), pre-eqlibrated with 2x Rat1/Rai1 size exclusion buffer (2.1.4, Table 11). When Rat1/Rai1 was prepared for use in nuclease assays, this buffer was made RNAse free. 10% (v/v) RNAse-free glycerol was added to the sample before freezing in liquid nitrogen and storage at –80°C.

Rtt103 was purified from 1 l of auto-inducing medium expression culture (see 2.2.3). After binding to the Nickel-resin, recombinant Rtt103 was eluted with 30 ml of Rtt103 IMAC buffer (2.1.4, Table 11) + 300 mM imidazole. The purification protocol is summarized in Figure 21.

4.2.3 Assembly of the Rat1/Rai1/Rtt103 trimeric complex

For the assembly of the trimeric exonuclease complex, the Rat1/Rai1 eluate from the heparin column (see 4.2.2) and the Rtt103 eluate from the NiNTA-column (see 4.2.2) were used. Protein concentrations of these samples were determined by the Bradford assay (see 2.2.4). The molarity of the Rat1/Rai1 sample was calculated (M_w of Rat1/Rai1: 160342). For Rtt103 the sample-volume that contained twice the molar amount compared to Rat1/Rai1 was calculated (M_w of Rtt103: 46459). The calculated volume was multiplied by 3, to compensate for the fact that 2/3 of the Rtt103 sample consisted of contaminating proteins (estimated by SDS-PAGE, see 2.2.7). The calculated volumes were mixed gently and incubated for 30 minutes at 20°C. The sample was then concentrated to 500 µl and applied to a Superose 6 10/300 GL column (GE Healthcare) pre-equilibrated with 2x Rat1/Rai1/Rtt103 size exclusion buffer (2.1.4, Table 11). The assembly protocol is summarized in Figure 21.

4.2.4 RNase and RNase H activity assays

For RNase assays, a 27 nt RNA („activityRNA", 2.1.2, Table 5) was labeled at the 3'-end using T4 RNA ligase (Fermentas) and [^{32}P]Cp at a concentration of 3 µCi/pmol of RNA in ligase buffer provided by the manufacturer. Excess radioactive nucleotides were removed with MicroSpin™ G-25 Columns (GE Healthcare) and 3'-end-labeled RNA was used as a substrate for 5'-end digestion. 3 pmol of labeled RNA were mixed with an equimolar amount of pure Rat1/Rai1 or Rat1/Rai1/Rtt103 in 1x Rat1 reaction buffer (2.1.4, Table 11) in a total volume of 30 µl, and incubated at 30°C. After 30, 60 and 180 minutes, 9 µl of each sample were removed and the reaction was stopped by mixing with an equal volume of 2x urea loading dye (2.1.4, Table 9) and incubation at 95°C for 5 minutes. The samples were analyzed by denaturing PAGE in 8 M urea (2.2.7). Remaining RNA was detected by autoradiography and the relative amounts were quantified with ImageQuant software. Radioactive gels were exposed to a storage phosphor screen (Molecular Dynamics) for several hours and scanning of storage screens was carried out with a STORM 860 imaging

system (Molecular Dynamics). All experiments were carried out in triplicates. For RNAseH assays, labeled 27 nt RNA was mixed with a four-fold molar excess of fully complementary DNA ("RNaseHDNA1", 2.1.2, Table 5) or DNA complementary only to 17 nucleotides in the 3' part of the RNA ("RNaseHDNA2", 2.1.2, Table 5) in TE buffer. The nucleic acids were annealed by heating to 95°C and slowly cooling to room temperature. The reaction was carried out and stopped after one hour as above, and the samples were analyzed as above. Control reactions using RNase I (Fermentas) and RnaseH (New England Biolabs) were carried out in the buffers provided by the manufacturer.

4.2.5 Purification of RNA polymerase II core enzyme

RNA Pol II core enzyme from *Saccharomyces cerevisiae* was purified from the deletion strain CB010ΔRpb4 essentially as described (Edwards et al, 1991). Cells, frozen in 3x Pol II freezing buffer (see 2.1.4, Table 12), were thawed in a water bath, then kept at 4°C. After addition of protease inhibitor mix (2.1.4, Table 9) the suspension was transfered to the metal chamber of a BeadBeater containing 200 ml soda lime glass beads ($\varnothing$ = 0.5 mm). After removal of air bubbles by stirring with a glass rod, the chamber was filled with HSB150 (2.1.4, Table 12) The BeadBeater was running 60-70 minutes in cycles of 30 seconds beating followed by a 90 second pause. The chamber with the suspension was kept cold by a salt-ice mix which was regularly renewed. The lysate was filtered through a mesh funnel which was then washed with HSB150 (2.1.4, Table 12), not exceeding a total volume of 1000 ml. The lysate was centrifuged twice for 45 min at 13700 g and 4°C. Afterwards it was filtererd through cheesecloth and paperfilter, to remove the lipid phase.

250 ml Heparin Sepharose 6 FF (GE Healthcare) affinity resin was equilibrated with 3 column volumes of HSB150. After loading of the lysate, the column was washed with 3 column volumes of HSB150, bound protein was eluted with 2 column volumes HSB600 (2.1.4, Table 12). Only milliliters 200 to 400 were used for the subsequent purification steps. Proteins in the eluate were precipitated by addition of 291 g/l fine-ground ammonium sulfate (~ 50% saturation) and overnight stirring at 4°C. The sample was then centrifuged for 45 minutes at 21860 g and 4°C, the precipitate was dissolved in 40 ml 1x TEZ0 (2.1.4, Table 12). Additional buffer was added to set conductivity below the conductivity of TEZ containing 400 mM ammonium sulfate (TEZ400). The sample was centrifuged again for 15 minutes at 34000 g and 4°C, to remove undissolved particles.

In an immunoaffinity step, Pol II was bound by immobilized antibody 8WG16 (NeoClone, Madison, USA) which binds selectively to the unphosphorylated C-terminal domain of Rpb1. Two columns with each 5 ml of the respective resin were equilibrated with 3 column volumes of TEZ250 (2.1.4, Table 12) lacking DTT and PI, followed by 1 column volume of TEZ250. After loading of the sample by gravity flow, the flowthrough of the first column was immediately loaded onto the second column. The columns were equilibrated to room-

temperature for 10 minutes. After washing with 5 column volumes of TEZ500 (2.1.4, Table 12), bound proteins were eluted with TEZ500 + glycerol (2.1.4, Table 12). At least 15 fractions with a volume of 1 ml each were collected and transferred to ice. The protein content of these fractions was qualitatively checked by adding 10 µl of sample to 200 µl 1x Bradford dye (2.1.4, Table 9). Peak fractions were pooled and 10 mM DTT was added before storage overnight at 4°C.

The pooled sample was diluted to 50 ml with Pol II buffer (see 2.1.4, Table 12) and concentrated with an Amicon Ultra-4 Centrifugal Filter Device (100k cutoff, 1300 x g, 4°C). After concentrating to ~5 ml, 10 ml of Pol II buffer were added and concentrated further. More sample was added after this step and this cycle was continued until all of the sample was added. By doing this, the buffer of the sample could be exchanged to the Pol II buffer, which was checked by measuring the conductivity of the flowthrough and comparing it to the conductivity of the Pol II buffer. When this value was constant, the sample was concentrated to ~ 2 mg/ml. The sample was divided into aliquots of 50 to 500 µg of protein. After adding 1.13 volumes of saturated ammonium sulfate solution compared to the aliquot volume, samples were incubated for 1 hour at 4°C on a rotating wheel and then centrifuged for 30 minutes at 16000 x g and 4°C to obtain a pellet. Supernatants were partly removed and samples were frozen in liquid nitrogen and stored at –80°C

4.2.6 Reconstitution of the 12 subunit RNA polymerase II complex

The recombinantly expressed and purified Pol II subcomplex Rpb 4/7 was provided by Elisabeth Lehmann. It was purified essentially as described (Armache et al, 2005b), and added in 5 fold molar excess respective to Pol II core enzyme and incubated for 30 minutes at 20°C. In cases where the sample was used for transcription bead assays, excess Rpb 4/7 was not removed by size exclusion chromatography to prevent unnecessary protein loss. Separation of 12 su Pol II from unbound Rpb 4/7 was carried out by binding fully assembled elongation complexes to magnetic beads (see 4.2.7).

4.2.7 Assembly of RNA polymerase II-nucleic acid complexes

Polymerase II was prepared as described in 4.2.4 and 4.2.5. Nucleic acids were synthetic and are listed in chapter 2.1.2, Table 5.

Artificial transcription scaffolds for crystallization of elongation complexes contained the template strand poly(A)xtalT, the non-template strand poly(A)xtalNT and the poly(A)xtalRNA (2.1.2, Table 5). Nucleic acids were annealed by mixing equimolar amounts of synthetic template DNA, nontemplate DNA, and RNA in TE buffer (see 2.1.4, Table 9) at a final

concentration of 100 µM. The mixture was heated to 95°C for 2 min, and cooled down to room temperature very slowly. Two molar equivalents of this scaffold and five molar equivalents of recombinant Rpb4/7 were incubated with one molar equivalent of Pol II in Pol II buffer (2.1.4, Table 12) for 20 minutes at 20°C. Assembled complexes were purified by Superose 6 size-exclusion chromatography and concentrated to 3.5 mg/ml.

Fully complementary transcription scaffolds for *in vitro* bead-based assays were assembled as described (Komissarova et al, 2003). For this, equimolar amounts of template DNA and RNA were mixed in RNAse-free TE buffer. The oligonucleotides were annealed by heating to 95°C for 2 minutes and slowly cooling to room temperature. Pol II (either core or 12 su Pol II, see 4.2.4 and 4.2.5 respectively) was incubated with a two-fold molar excess of the annealed hybrid for 15 minutes at 20°C while gently shaking. A four-fold molar excess of non-template DNA, containing Biotin at the 5' end, was added and the mixture was incubated for 20 minutes at 25°C while shaking. Control-elongation complexes lacking the biotinylated non-template strand were assembled identically, omitting the last step

4.2.8 Bead-based termination assays

Magnetic, streptavidin-coated beads (Dynabeads® MyOne™ Streptavidin T1, Dynal Biotech, distributed by Invitrogen) were prepared by washing twice with beads breaking buffer (2.1.4, Table 9) followed by incubation in 500 µl of beads blocking buffer (2.1.4, Table 9) overnight at 4°C. After washing again twice with breaking buffer, beads were resuspended in the original volume of breaking buffer. 1-3 pmol fully assembled elongation complexes (4.2.7) were added per reaction (= 10 µl blocked and washed beads) followed by an incubation for 30 minutes at 25°C, gently shaking. Unbound complexes were removed by washing with 50 µl beads breaking buffer, Rat1 wash buffer, and Rat1 reaction buffer (2.1.4, Table 11). Beads were resuspended in 19 µl Rat1 reaction buffer. RNA in the elongation complex was labeled at the 3' end with the use of Pol II activity, by adding 1 µl of $\alpha[^{32}P]UTP$ [10 mCi/ml] (GE Healthcare), followed by an incubation at 28°C for 20 minutes, slowly shaking. Unincorporated nucleotides were washed away by applying twice 30 µl of Rat1 reaction buffer. Beads were resuspended in Rat1 reaction buffer, and nuclease complexes were added at a two-fold molar excess respective to the elongation complex concentration used for binding to the beads, followed by incubation for 1 hour at 30°C. As a single strand-specific control endonuclease, RNase I (100 units, New England Biolabs, #M0243S) was used. After the reaction, nucleases were washed away with Rat1 washing buffer and Rat1 reaction buffer. Beads were resuspended in 20 µl of Rat1 reaction buffer. To test the ability of ECs to elongate the transcript after nuclease digestion, beads were resuspended in 19 µl, and 1 µl of NTP mix was added at a final concentration of 1 mM and incubated at 28°C for

30 minutes. The reactions were stopped by adding 20 µl of 2x urea loading dye (2.1.4, Table 9) and incubating for 5 minutes at 95°C. Samples were analyzed by 6 M urea PAGE and radioactively labeled RNA was detected by exposition of the gels to a storage phosphor screen (Molecular Dynamics) overnight at 4°C. Scanning of storage screens was carried out with a STORM 860 imaging system (Molecular Dynamics).

4.2.9 Crystallization of 12 subunit Pol II elongation complexes and a poly(A) site containing nucleic acid scaffold

Pol II-nucleic acid complexes were purified as described in 4.2.5 and 4.2.6 and assembled as described in 4.2.7 An additional amount of the nucleic acid scaffold was added prior to crystallization to a final concentration of 2 µM. Crystals were grown at 22 °C with the hanging drop vapor diffusion method by mixing 2 µl of sample solution with 1 µl of reservoir solution (2.1.4, Table 13). Crystals were harvested after 10-20 days of growth, when they had reached their maximum size and were transferred stepwise to mother solution containing additionally 0-20 % glycerol over 5 h. After the last step, crystals were slowly cooled to 8°C, and flash-frozen by plunging into liquid nitrogen.

4.2.10 Data collection and structure solution

Complete diffraction data to 4.0 Å resolution were collected at the Swiss light source (SLS) beamline PX1 (Table 20). The structure was solved by molecular replacement with the model of the complete 12-subunit Pol II EC without nucleic acids as a search model (PDB 1Y1W, (Kettenberger et al, 2004b). The molecular replacement solution was subjected to rigid body refinement with CNS (Brunger et al, 1998). The final structure includes 25 nucleotides of DNA and 8 nucleotides of RNA, shows good stereochemistry, and has a free R-factor of 23.5% (Table 20). The nucleic acids were built into the initial F_o-F_c electron density map using the program Coot (Emsley & Cowtan, 2004). A thymine residue in the template strand was replaced for 5-bromouracil. Diffraction data were recorded at the wavelength of the bromine K absorption edge, and the resulting anomalous difference Fourier maps revealed single peaks demarking the positions of the bromine atom. Thus, the register of the nucleic acids was unambiguously defined by the identification of the position of bromine in the template strand (Fig. **24** E). Atomic positions and B-factors were refined with CNS, the progress was monitored with the free R-factor which showed a value of 23.5% at the end of refinement. Statisitics for Data-collection and refinement are listed in Table 20 in chapter 4.3.6.

4.3 Results and discussion

4.3.1 Preparation of recombinant Rat1, Rat1/Rai1 and Rat1/Rai1/Rtt103

To investigate the function of the Rat1 nuclease, we established a protocol to obtain the protein in recombinant form after overexpression in *E. coli* (Fig. 21 A and B, 4.2.2). After optimization of the procedure, about 0.5 mg of the 116 kDa, 1006-residue Rat1 could be obtained from 1 l of bacterial cell culture. We further established a protocol to obtain a stoichiometric pure complex of Rat1 with its cellular partner, Rai1 (Fig. 21 A and B, 4.2.2). We co-expressed Rat1 and Rai1 from individual plasmids with different antibiotic resistance in *E. coli* cells with the use of autoinducing medium (Studier, 2005) . Affinity purification of Rat1, which carried a C-terminal hexahistidine tag, led to co-purification of Rai1 (Fig. 21 A). The Rat1/Rai1 complex was very stable, as it could be purified over several columns. To purify a recombinant trimeric Rat1/Rai1/Rtt103 complex, we incubated the pure Rat1/Rai1 complex with partially purified recombinant full-length Rtt103, and subjected the mixture to size exclusion chromatography (4.2.3). We obtained a symmetric peak for the trimeric complex at a shorter retention than the peak for the dimeric Rat1/Rai1 complex (Fig. 21 B). The identities of the three proteins were confirmed by mass spectrometry (not shown). For the first time this work provided Rat1, and the complexes Rat1/Rai1 and Rat1/Rai1/Rtt103 in pure, bacterially expressed, recombinant form. The results also confirmed that the protein-protein interactions that were previously inferred from affinity co-purifications of endogenous proteins (Kim et al, 2004; Stevens & Poole, 1995) and co-expression of Rat1 and Rai1 in yeast (Johnson, 2001; Xue et al, 2000). In addition, this is the first time a direct interaction between Rat1/Rai1 and Rtt103 is seen.

4.3.2 Stable Rat1 activity requires Rai1, but not Rtt103

Nuclease activity of the Rat1 homolog Xrn1 was extensively analyzed (Stevens, 1978; Stevens, 1980; Stevens & Maupin, 1987) showing its processive 5'-3' mode of degradation and its preference of RNA with a 5'-monophosphate. The characterization of Rat1 basically showed comparable results in this respect (Stevens & Poole, 1995). In early work on Xrn1, which has comparable properties to Rat1, it was shown that RNA with a 5'-OH is used at 1/3 to 1/5 compared to RNA with a 5' phosphate (Stevens & Maupin, 1987).

To investigate the ribonucleolytic activity of recombinant Rat1, and the influence of Rai1 and Rtt103 on this activity, we labeled the 3'-end of a RNA 27-mer of random sequence („activityRNA", 2.1.2, Table 5) with radioactive phosphate. Although the preferred substrate for Rat1 is RNA with a 5'-monophosphate, we used RNA with a 5'-hydroxyl-group. The rea-
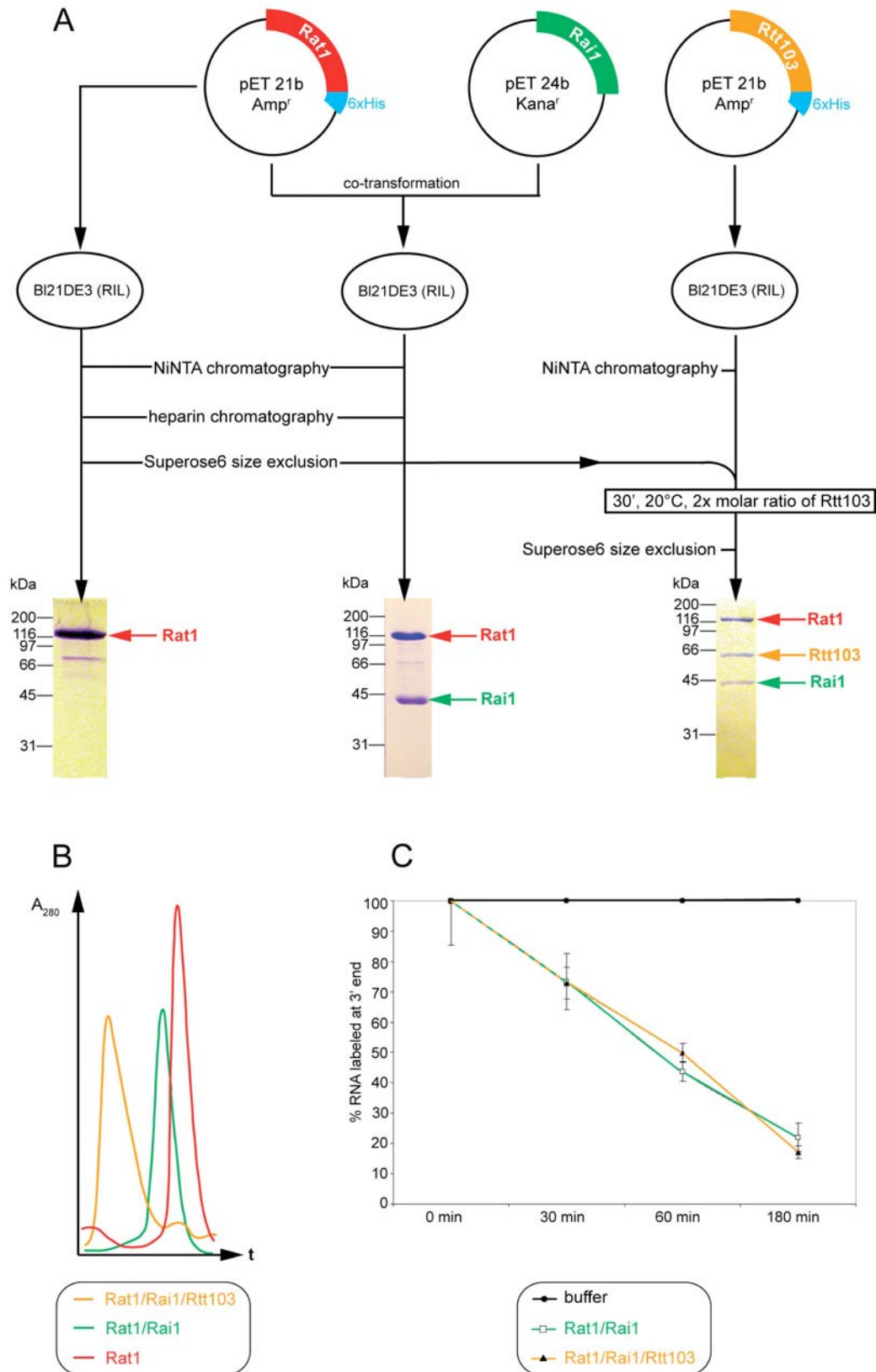
Figure 21: Preparation of recombinant Rat1, Rat1/Rai1 and Rat1/Rai1/Rtt103 and ribonucleolytic activity

(**A**) Expression and purification scheme for Rat1, Rat1/Rai1 and Rat1/Rai1/Rtt103. At the bottom, Coomassie-stained SDS-PAGEs show the resulting proteins.

Figure 21 (continued)

(**B**) Size-exclusion chromatography profiles of the nuclease complexes from (A). A plot of the absorption at 280 nm against the retention time is shown. Profiles are from 3 separate chromatographies.

(**C**) Relative ribonucleolytic activity of Rat1/Rai1 and Rat1/Rai1/Rtt103 using RNA with a 5'-hydoxyl group as a template. The intensity of signals corresponding to the 27mer RNA were quantified with the ImageQuant software resulting in a mean value from three replicates of each experiment. The intensity of the input RNA was set to 100%, the decrease due to degradation by Rat1 was calculated relative to that value and plotted against the duration of the reaction.

son for this is that T4 RNA ligase used in the labeling reaction (4.2.4) produced multimers of the RNA molecules, resulting in a heterogeneous substrate that could not be quantified by the method we used. Nevertheless, ribonucleolytic activity – albeit lower than the expected activity with the natural substrate – could be detected and put into relation for the different complexes (Fig. 21 C). Degradation of the labeled RNA after incubation with enzyme preparations for different times was monitored by separation of the RNA products in denaturing PAGE and autoradiography (Fig. 21 C, 2.2.7). In these assays, recombinant Rat1 either showed no activity or weak, badly reproducible activity. In contrast, the recombinant Rat1/Rai1 complex showed reproducible and stable nuclease activity. This is consistent with the published stabilizing effect of Rai1 on Rat1 activity (Xue et al, 2000), where a dramatic loss of Rat1 nuclease activity was observed during the first minutes of incubation at 30°C when Rai1 was not present. The complex could be stored at –80°C for several months without a significant decrease of its activity. The nuclease activity was due to Rat1 since buffer from the last purification step did not contain contaminating nuclease activities (Fig. 21 C). Thus, recombinant Rat1 shows unstable activity that is stabilized by recombinant Rai1, consistent with similar observations for the endogenous proteins (Xue et al, 2000).

We compared the robust nuclease activity of the Rat1/Rai1 complex to that of the trimeric Rat1/Rai1/Rtt103. The dimeric and trimeric complexes showed indistinguishable activities (Fig. 21 C). This suggests that Rtt103 is not a regulator of Rat1 nuclease activity, consistent with the results that Rtt103 is involved in the recruitment of Rat1/Rai1 to the transcription machinery via its CTD-interacting domain (CID) (Kim et al, 2004).

4.3.3 An improved *in vitro* elongation assay

To establish a defined biochemical system for testing the torpedo model in vitro, we assembled elongation complexes containing Pol II, DNA with a fully complementary transcription bubble, and RNA as described (Komissarova et al, 2003) (Fig. 22 A, top).

Figure 22: The Rat1/Rai1 complex does not terminate RNA polymerase II *in vitro*
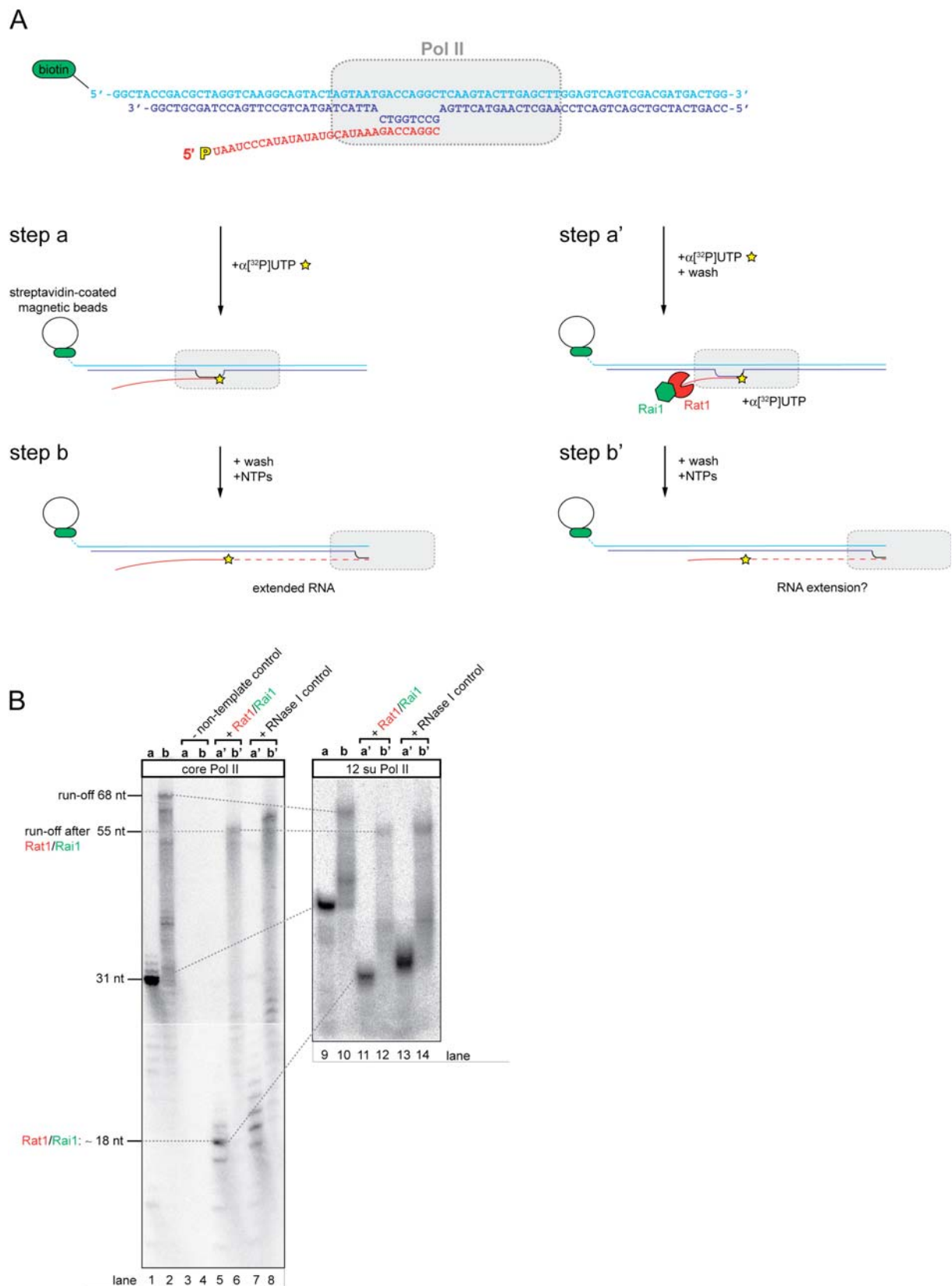
(**A**) Schematic representation of the bead-based elongation/RNA degradation assay. On top the nucleic acid scaffold that was assembled with Pol II into a bead-coupled elongation complex (EC) and used as a substrate for Rat1. Template, non-template and RNA strands are coloured in blue, cyan and red, respectively. Below the diff-

erent steps in the protocol (4.2.8). Steps a and b represent the control reaction without added nuclease, a' and b' represent the reaction containing the nuclease.

(**B**) Autoradiograph of RNA from bead-based termination assay. As substrates, core Pol II ECs (left panel) and 12 subunit Pol II ECs (right panel) were used. Samples that were treated as depicted in (A) were separated by a 6 M urea PAGE and exposed for several hours to a storage phosphor screen. The steps represented in (A) are indicated above the gel, lane numbers below the gel. See text for details.

The presence of a fully complementary transcription bubble is important for a termination assay since transcription termination in prokaryotes requires rewinding of the upstream bubble (Park & Roberts, 2006). ECs were bound to streptavidin-coated magnetic beads with the use of biotin coupled to the 5'-end of the non-template DNA strand (Fig. 22 A). Since incorporation of the non-template strand is the last step in the EC assembly protocol, only fully assembled ECs were bound to the beads. The RNA engaged in active ECs was labeled at the 3'-end by incubation with $\alpha[^{32}P]$-UTP, which leads to Pol II-catalyzed incorporation of radioactive uridine (Fig. 22 A, step a; Fig. 22 B, lane 1). After washing away unincorporated $\alpha[^{32}P]$UTP, and addition of all four NTPs (Fig. 22 A, step b), these ECs were able to elongate the labeled RNA (Fig. 22 B, lane 2). RNA transcripts of various lengths were observed, up to the expected 68 nt run-off product. The occurence of transcripts that are shorter than the run-off RNA was different depending on the scaffolds we used and can probably be related to paused and arrested, or unstable ECs. To prevent unspecific binding of ECs to beads, beads were blocked prior to use (4.2.8). ECs that were assembled without the biotin-labeled non-template DNA strand did not bind to these beads, providing a negative control (Fig. 22 B, lanes 3 and 4). Thus, in this elongation assay, only complete, bead-coupled elongation-competent ECs produced signals for RNA products. Compared to a previously described bead-based elongation assay (Komissarova et al, 2003), our assay uses magnetic beads instead of agarose beads and a bead-blocking protocol. This results in very clear and defined RNA signals.

4.3.4 Functional Rat1 complexes do not terminate Pol II *in vitro*

The bead-based elongation assay (4.3.3), together with the availablity of pure recombinant Rat1 complexes (4.3.1), allowed us to test the "torpedo model" *in vitro*. When Rat1/Rai1 was added to our assay (Fig. 22 A, step a'), the RNA was degraded to a length of approximately 18 nt as judged from a sequencing gel (Fig. 22 B, lane 5). This shows that Rat1/Rai1 could use the pure, bead-coupled ECs as substrates, but also that it was unable to totally degrade the RNA. The length of the obtained products is readily explained based on the known EC structure (Andrecka et al, 2008; Kettenberger et al, 2004b; Vassylyev et al, 2007a). Around 15 nucleotides of RNA are protected within the Pol II hybrid-binding site and the RNA exit

tunnel, and a few additional nucleotides are probably protected due to the probe radius of Rat1. Thus, the obtained results are consistent with the model that Rat1/Rai1 degrades RNA from the 5'-end until it reaches the Pol II surface, and that it does not degrade the 3'-region of the RNA that is protected by Pol II. To examine whether the ECs remained active after partial RNA degradation, Rat1/Rai1 was washed away and NTPs were added (Fig. 22 A, step b'). The truncated transcripts were readily elongated, and reached a length of up to 55 nt, corresponding to the expected run-off product (Fig. 22 B, lane 6). Since the amount of RNA in the Rat1-treated samples was lower (lanes 5 and 6) than in the untreated control sample (lanes 1 and 2), we used a control nuclease to determine if this observation is due to partial termination. However, the same result was obtained when the experiment was repeated with the use of the non-specific endonuclease RNase I instead of Rat1/Rai1 (Fig. 22 B, lanes 7-8), arguing against partial termination. As it was recently shown that transcribing Pol II includes the Rpb4/7 subcomplex *in vivo* (Jasiak et al, 2008; Verma-Gaur et al, 2008), we repeated our assay with the complete, 12-subunit enzyme that comprises Rpb4/7. The same results were obtained (Fig. 22 B, lanes 9-14). Thus, Rat1/Rai1 degrades RNA that is accessible on the EC surface, but does not dissociate Pol II, and leaves the EC intact and transcriptionally competent, such that the RNA 3'-end remains in the active site and can be re-extended. Since transcription termination is defined as a discontinuation of the ability to extend RNA, these results show that Rat1/Rai1 is insufficient to terminate Pol II in a defined *in vitro* system.

4.3.5 A poly(A) site does not trigger termination

At the end of an open reading frame the transcript is cleaved 10-30 nt downstream of a specific nucleotide signal, the poly(A) site. This nucleic acid sequence has the consensus 5'-AAUAAA-3' in the transcript (Proudfoot & Brownlee, 1976), which is highly consereved in almost all polyadenylated mRNAs in higher eukaryotes (Wickens & Stephenson, 1984). Interestingly, the poly(A) site sequence represents sequence-motifs that trigger termination in the other eukaryotic polymerases evolutionary related to Pol II. In case of Pol I, a T-rich sequence in the template strand is mandatory (Lang et al, 1994; Lang & Reeder, 1993) whereas for Pol III a short run of A elicit transcription (Geiduschek & Kassavetis, 2001) together with other factors. Even in intrinsic terminators in prokaryotes, a stretch of U-residues is involved (see also section 1.3.2). All of these sequences are composed of T-A and A-U basepairs between template strand and transcript, respectively, and are referred to as „weak hybrids" here. Although it is known that the poly(A) site does not induce termination in Pol II transcription complexes, we asked the question if this sequence may trigger re-arrangements of nucleic acids within the Pol II cleft and active site, thereby triggering a "termination competent" form of polymerase II. This might be a prerequisite for induced

termination by Rat1/Rai1. We first asked whether Pol II could terminate in our *in vitro* system if the poly(A) site sequence was transcribed. However, the situation in yeast cells is somewhat different, as the consensus AAUAAA is not as highly conserved as in higher eukaryotic systems. Only in ~50% of all yeast genes this sequence can be found and mutations in this sequence do not have a dramatic effect on the 3' end formation of yeast mRNA (Hyman et al, 1991). Yeast poly(A) signals are basically composed of three elements that are important for proper cleavage/polyadenylation in yeast, although they are less conserved than the mammalian poly(A)-site directing sequence. However, these sequences are AT-rich and resemble the "weak hybrid" concept (Graber et al, 1999; Guo & Sherman, 1996; Osborne & Guarente, 1989; Russo et al, 1993). This, in combination with the fact that the process of transcription is conserved, led us to the decision to use the highly conserved AAUAAA consensus sequence in our assays rather than the loosely defined yeast site, to investigate its influence on the elongation complex. We prepared a nucleic acid scaffold containing a poly(A) site 10 nt downstream of the incorporated $\alpha[^{32}P]UTP$ (Fig. 23 A, poly(A) transcription scaffold 1), assembled an EC, and added ATP, CTP, and UTP, but withheld GTP from the reaction mixture. This led to transcription of the poly(A)-site plus 16 additional base pairs, and EC stalling at a defined position (CCC in the template strand, Fig. 23 A, stalling point). RNAs with a length of 41-46 nt were not observed (Fig. 23 B), which means that the poly(A)-sequence was transcribed like a control random sequence in our assay (see Fig. 23 A, „poly(A) transcription scaffold 1 control").

We next tested the model that passage of a poly(A) site would alter EC stability and render Pol II prone to Rat1-induced termination. We assembled ECs containing the poly(A) site sequence within the DNA-RNA hybrid at positions −1 to −6 (Fig. 23 A, poly(A) transcription scaffold 2). RNA was labeled by incubation with CTP and $\alpha[^{32}P]UTP$, which led to incorporation of three cytosines and a radioactive uridine (Fig. 23 C, lane 1). Incorporation of three additional cytosines ensured that the resulting RNA length was as in the other experiments (Figs. 22 , 23 B, 23 D), and mimicked partial polymerase passage of the poly(A) site sequence. Addition of NTPs produced the run-off RNA of 53 nt (Fig. 23 C, lane 2). As before, RNA was trimmed down by Rat1/Rai1 to around 18 nt, and could be re-extended after NTP addition (lanes 3 and 4).

Very similar results were obtained when RNase I was used instead of Rat1/Rai1 (lanes 5 and 6). RNAs longer than the run-off transcript were also detected, but were excluded from Fig. 23 C because they were present in all the samples of the experiment and likely result from transcript slippage as a side reaction. Taken together, Rat1/Rai1 and a transcribed poly(A)-site sequence alone are insufficient to trigger termination of Pol II in these assays.
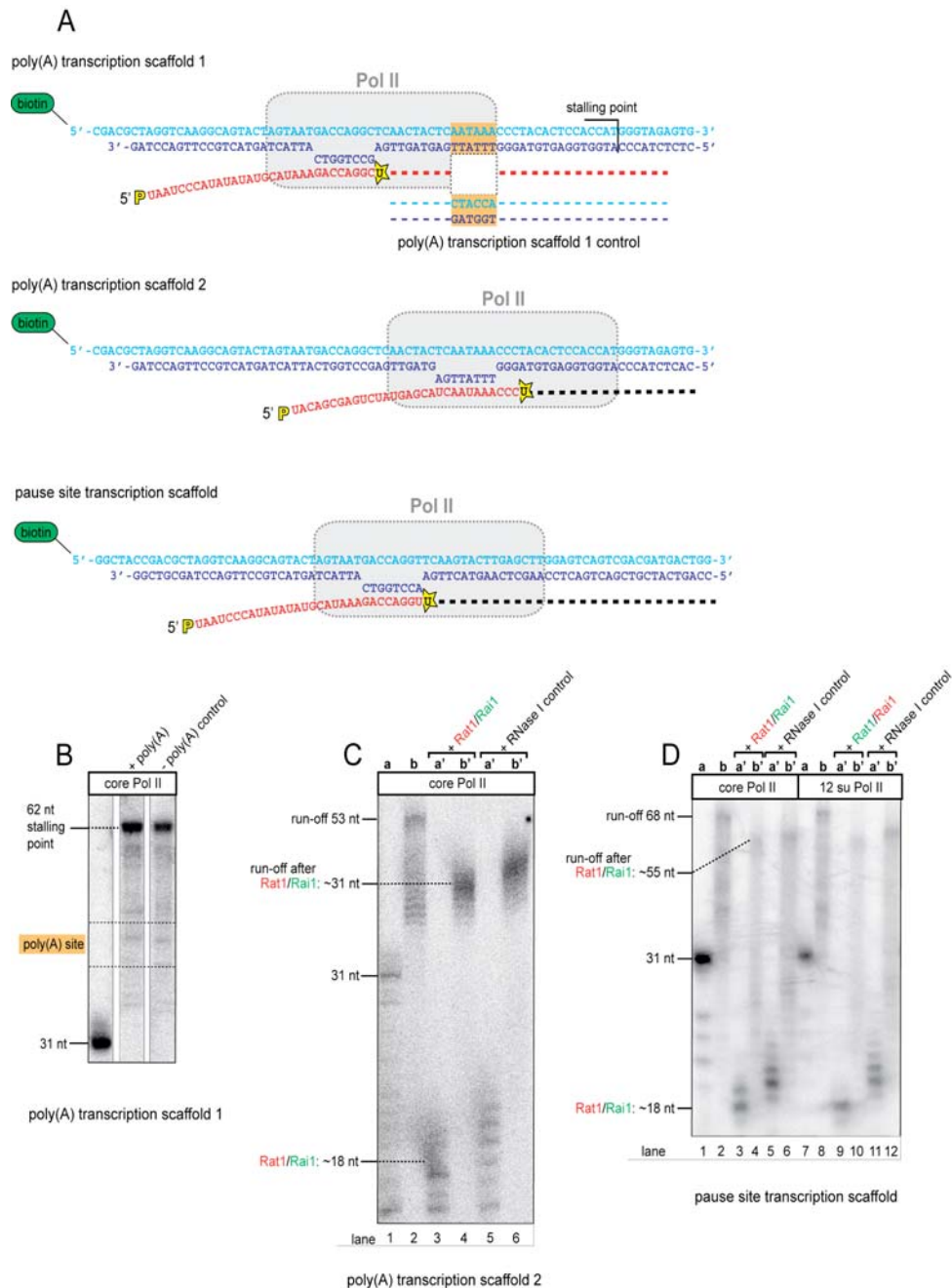
Figure 23: Neither a poly(A) site nor a pause sequence trigger termination

(**A**) Nucleic acid scaffolds that were assembled with Pol II for the indicated experiments.

(**B**) Autoradiograph of RNA extension with a bead-coupled EC including poly(A) transcription scaffold 1. The region where signals of RNAs would be expected if elongation had stopped at the poly(A) site sequence is indicated.

(**C**) Autoradiograph of bead-based assay with an EC containing poly(A)-transcription scaffold 2. Steps a, b, a' and b' correspond to steps in Fig. 22 B.

(**D**) Autoradiograph of bead-based assay with EC containing the pause site transcriptionscaffold. For a detailed description of these experiments refer to text.

4.3.6 A hybrid with poly(A) site sequence does not change EC structure

The results described above suggested that an EC that contains the poly(A) site within the DNA-RNA hybrid does not differ significantly in structure from an EC that contains a random sequence within the hybrid. To examine this, we solved the X-ray structure of *S. cerevisiae* Pol II EC containing a hybrid that harbors the mammalian poly(A)-site sequence in the active center cleft of the enzyme (Fig. **24** and Table 20). The structure could be determined at 4.0 Å resolution with the use of established protocols (Brueckner & Cramer, 2008). The unbiased difference electron density in the hybrid site did not reveal any significant rearrangement of the nucleic acids within the Pol II cleft (Fig. **24** B and C). However, we observed that the hybrid was backstepped compared to the designed scaffold, and that the adenine at the RNA 3'-end was not paired with the DNA template at position +1, but rather disordered (Fig. **24** C). Such fraying of the 3'-terminal RNA nucleotide is typical for a paused state of the EC (Toulokhonov et al, 2007), and has been directly observed in other Pol II structures that we will describe elsewhere (Sydow et al., in preparation). Superposition of our structure with the previous complete Pol II EC structure showed a similar position and conformation of the DNA-RNA hybrid and downstream DNA (Fig. **24** D). These results are consistent with the idea that the poly(A) site sequence is prone to pausing, but do not provide evidence for models that postulate that the EC adopts an alternative structure upon poly(A) site passage.
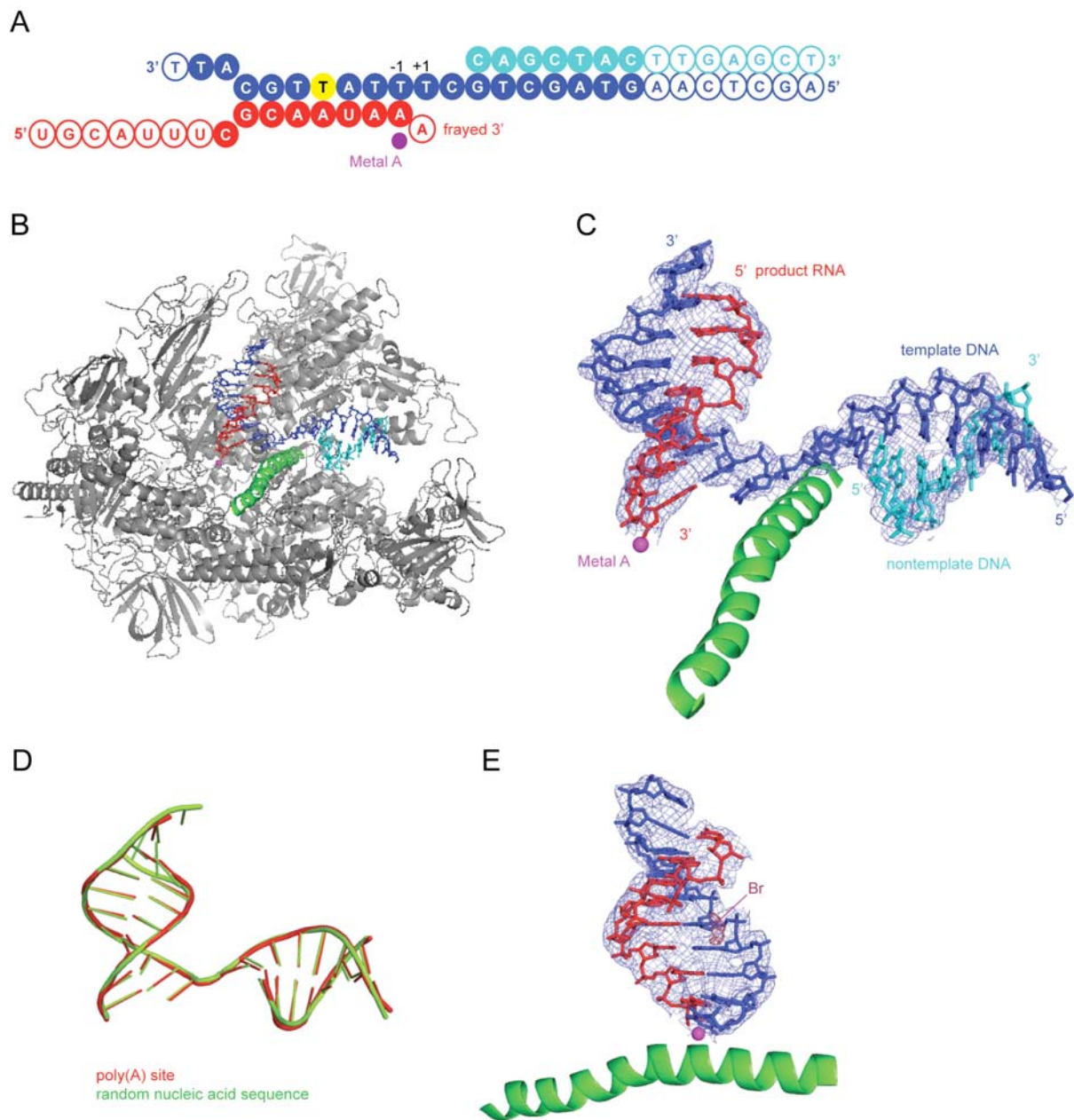
Figure 24: Canonical binding of a poly(A) site-containing hybrid within the Pol II EC

(A) The used nucleic acid scaffold is shown. Filled circles indicate nucleic acids that have interpretable electron density, empty circles indicate nucleic acids that were not ordered. The 5-Bromouracil is shown in yellow. The template strand is shown in blue, the non-template strand in cyan and the RNA in red throughout this figure.

(B) Overview of the Pol II EC structure containing the poly(A) site sequence. Pol II is shown as a ribbon model in grey. Rpb2 residues 1-828 are omitted for clarity. The Pol II bridge helix (residues 811-843 of Rpb1) is shown in green. Nucleic acids are shown in a stick representation, coloured as in (A).

(C) 2Fo-Fc electron density map contoured at 1σ for the nucleic acids in the poly(A) site-containing DNA-RNA hybrid.

Figure 24 **(continued)**

(D) Superposition of the poly(A) site nucleic acids (red) and nucleic acids of random sequence (green, PDB accession code: 1Y1W, Kettenberger et al, 2004b)).

(E) 2Fo-Fc electron density at 1σ of the upstream nucleic acids from (C). In raspberry an anomalous difference Fourier map at 3.5σ, showing the position of the bromine atom. The structure is rotated app. 90° clockwise respective to (C).

Table 20: X-ray diffraction and refinement statistics

	Complete Pol II EC containing poly(A)site sequence
Data collection	
Space group	C222(1)
Cell dimensions	
a, b, c (Å)	222.5, 391.6, 284.1
α, β, γ (°)	90, 90, 90
Wavelength	0,918905
Resolution (Å)	50-4.0
Rsym (%)	10.6 (51.6)
I / σI	9.6 (2.7)
Completeness (%)	99.9 (100)
Redundancy	4.3 (4.3)
Refinement	
Resolution (Å)	4.0
Unique reflections (anomalous pairs unmerged)	202,368 (33,630)
R_{work} / R_{free} (%)	20,4 / 23,5
No. atoms	
Protein	31102
Ions	9
Nucleic acids	671
B-factors	
Protein	137.1
Ions	137.8
Nucleic acids	196.8
R.m.s deviations	
Bond lengths (Å)	0.008
Bond angles (°)	1.5

4.3.7 A paused EC is not terminated by Rat1/Rai1

There is evidence that the EC goes through a paused state before termination (Aranda & Proudfoot, 1999; Birse et al, 1997; Proudfoot, 1989). This makes sense in light of the "torpedo model" as a paused polymerase would create a time window for Rat1/Rai1 to catch up. Thus, we also tested whether a paused EC can be terminated in our assay. We assembled ECs with a scaffold containing a 3'-uridine in the active site (Fig. 23 A, pause site transcription scaffold). Introduction of an additional, radioactively labeled uridine to the RNA results in a pause sequence with a UU RNA 3'-end at the Pol II active center (Fig. 23 A). In ECs that contain this pause sequence, the RNA 3'-nucleotide adopts a frayed position in the pore below the active site (Toulokhonov et al, 2007). We incubated the obtained EC with the recombinant Rat1 complexes and checked if they can still elongate the RNA after it had been partially degraded (Fig. 23 D). The experiments showed that neither core Pol II (lanes 3 and 4), nor the complete Pol II (lanes 9 and 10) were terminated. In both cases the incubation with RNase I gave similar results (lanes 5, 6, 11 and 12). Thus, a paused EC conformation is not sufficient to allow Rat1 to terminate Pol II in our system.

4.3.8 Rat1 contains a putative RNase H-like domain

Rat1 and its cytoplasmatic counterpart Xrn1 contain a eukaryote-specific N-terminal region of sequence conservation that was ascribed to the 5PX superfamily of exoribonucleases (Zuo & Deutscher, 2001) (Fig. 25 A). This region is required for Xrn1 exonuclease activity, since point mutations in D86, E176, E178, D206 and D208 (corresponding to Rat1 residues D102, E203, E205, D233, and D235, respectively, Fig. 25 A) and additional residues (dots in Fig. 25 A) impair activity (Page et al, 1998; Solinger et al, 1999). Consistently, mutation of Rat1 residue D235 abolished exonuclease activity, and resulted in the failure to complement the termination defect seen in rat1 mutant cells (Kim et al, 2004). Thus, the N-terminal regions of the Xrn1 and Rat1 proteins contain the exonuclease domain. In this region, the members of the 5PX superfamiliy were described to show homology in critical catalytic residues with several Mg^{2+}-dependent 5'-3'-exonucleases from different organisms, including also the phage T4 RNase H (Solinger et al, 1999), a 5'-3'-exonuclease that degrades RNA in DNA-RNA hybrids. Thus, we re-examined the Rat1 sequence and found a remote similarity of the conserved N-terminal region to the catalytic domain of RNase H enzymes (Fig. 25 B). The active site of RNase H enzymes contains four conserved acidic residues, three aspartates and a glutamate separated by short sequence stretches and coordinating two
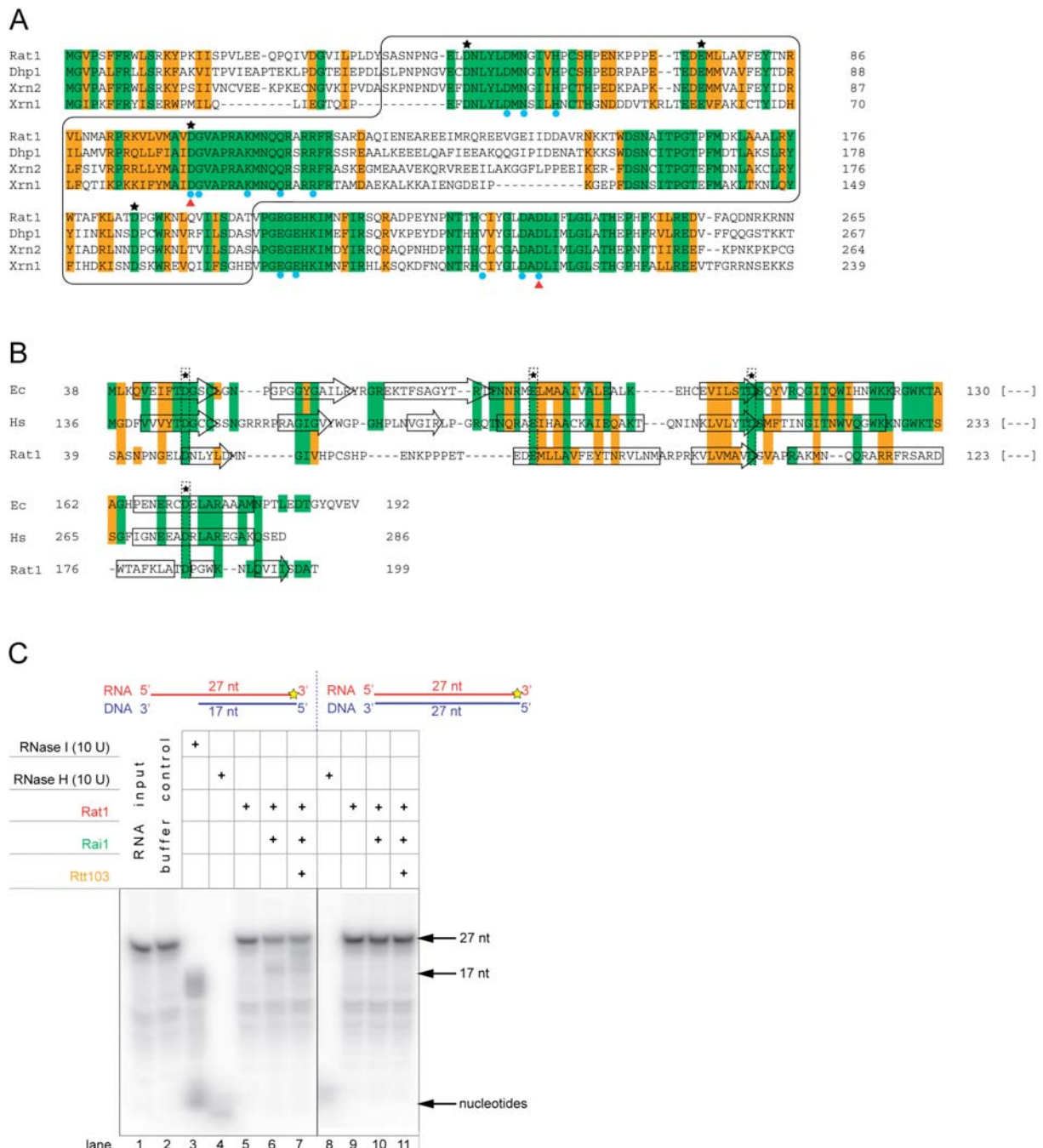
Figure 25: Putative RNase H homology in the Rat1 sequence

(**A**) Alignment of the N-terminal region of Rat1 and its homologs Dhp1 (*S. pombe*), Xrn2 (*H. sapiens*) and Xrn1 (*S. cerevisiae*, cytoplasmic). Invariant and conserved residues are indicated in green and orange, respectively. Blue circles indicate positions of point mutations that impair exonuclease-activity in Xrn1 (Page et al, 1998; Solinger et al, 1999). Red triangles show Rat1 mutations that impair activity as reported in (Kim et al, 2004) and in this work (4.3.9). Stars mark potential RNase H active site residues (see also (B)).

(**B**) Alignment of RNase H sequences from *E. coli* (Ec) and *H. sapiens* (Hs) with *S. cerevisiae* Rat1. The alignment of Ec and Hs sequences is taken from (Nowotny et al, 2007). Secondary structure elements observed in the structures of human RNase H1 Nowotny et al, (2007) and E.coli RNase H (Katayanagi et al, 1990; Yang et al, 1990) are indicated as rectangles for α-helices and arrows for β-strands. Conservation is indicated as in (A).

Figure 25 (continued):

Rat1 secondary structure elements were predicted by PSIPRED (Jones, 1999). Invariant acidic residues that constitute the RNase H active site are indicated by stars.

(**C**) Autoradiograph of RNA from RNase H assay. The templates used are shown schematically on top. The proteins used in the individual reactions are indicated by a (**+**). See text for details.

catalytic magnesium ions (Katayanagi et al, 1990; Nowotny et al, 2005; Yang et al, 1990). We could manually align the sequences of *E. coli* and human RNase H to the Rat1 N-terminal region such that the four acidic residues in the RNase H active site correspond to conserved Rat1 residues (Fig. 25 B). In these alignments, known and predicted secondary structure elements also generally lined up. Consistently, RNase H activity was reported for Xrn1 (Stevens & Maupin, 1987). This cytoplasmic counterpart of Rat1 contains the highly conserved acidic residues we identified in the Rat1 exonuclease-domain (stars in Fig. 25 A and B).

Since Xrn1 can supply the essential functions of Rat1 when directed to the nucleus by a strong localization signal (Johnson, 1997) (although it does not complement a termination defect (Luo et al, 2006)), we wondered whether Rat1 complexes also exhibit RNase H activity. To test this, we used two different DNA-RNA heteroduplex substrates, a radioactively 3'-labeled 27-mer RNA hybridized to a complementary 27-mer DNA strand, or the same RNA hybridized to a complementary 17-mer DNA strand, which leaves a 10-nucleotide 5'-overhang of RNA. The latter substrate resembles the nucleic acids in an EC, where Rat1 would degrade single-stranded RNA from the 5'-end and then encounter the DNA-RNA hybrid (Fig. 25 C). As a control we used RNase I, which specifically degrades single-stranded RNA, and RNase H, which specifically degrades RNA within a DNA-RNA hybrid. As expected, RNase I degraded the single-stranded 10 nt overhang (lane 3), and RNase H degraded RNA within the hybrid region (lanes 4 and 8). However, neither Rat1, nor its complexes Rat1/Rai1 or Rat1/Rai1/Rtt103 led to RNA degradation (lanes 9, 10, 11). It has to be considered, that Rat1 shows a low activity concerning single stranded DNA (Stevens & Poole, 1995). Since our reaction mixture contains an excess of DNA over RNA molecules, the ssDNA can also serve as a substrate and reduce the detectable ribolytic activity. However, the 5'-overhang was partially degraded by Rat1/Rai1 and Rat1/Rai1/Rtt103 (lanes 6 and 7), indicating that the enzyme complexes were active under the assay conditions, and providing the positive control. Thus, Rat1 contains a putative RNase H domain in its N-terminal region, but we did not observe RNase H activity with recombinant Rat1 complexes in vitro.

Nevertheless, the identified RNase H activity of Xrn1, that shows a high sequence conservation with Rat1 (Fig. 25 A), together with the similarity of the conserved N-terminal region of Rat1 to the catalytic domain of RNase H enzymes we identified (Fig. 25 B) leads us to the assumption that Rat1 might exhibit RNase activity *in vivo* we cannot detect in our

highly pure *in vitro* assay. In such a scenario, Rat1 degrades RNA until it reaches elongating Pol II and its associated factors, which may activate a cryptic RNAse H activitiy in Rat1, which then may degrade RNA within the DNA-RNA hybrid, destabilizing the EC and leading to termination. We refer to this model as the „triggered torpedo model" since it implies that the Rat1 torpedo requires a RNAse H activity that must be activated (triggered). The possible RNAse H activity of Rat1 is consistent with a recent study of the role of Rat1 in degrading telomeric repeat-containing RNA (Luke et al, 2008) . In this study, overexpression of RNAse H rescued telomere elongation defects in cells that lacked functional Rat1. The defects stem from an increased concentration of short RNAs that bind to telomeric DNA. These RNAs are kept at a low level by functional Rat1. Interestingly, a deletion of Rnase H genes does not alter the defective Rat1 phenotype *in vivo*. The authors suspect the missing activity to be a helicase, like the ATP-dependent RNA-helicase Upf1. Another explanation could be that Rat1 itself contains RNAse H activity in its functional context *in vivo*.

However, to provide evidence for the triggered torpedo model, activation of a possible RNAse H activity of Rat1 must be demonstrated.

4.3.9 The Rat1 nuclease active site degrades RNA within the EC

To investigate whether the Rat1 N-terminal region was responsible for the observed RNA degradation activity in the context of ECs, we mutated conserved acidic residues and tested for RNA degradation. We mutated the highly conserved residue D102, proposed to be part of a putative RNase H active site (Fig. 25 A and B). In addition we mutated the conserved residue D235 that was shown to be required for nuclease activity *in vitro* (Kim et al, 2004). The Rat1 D102A mutant was slightly impaired in its activity to degrade RNA within the EC (Fig. 26, lanes 9, 6, 3).
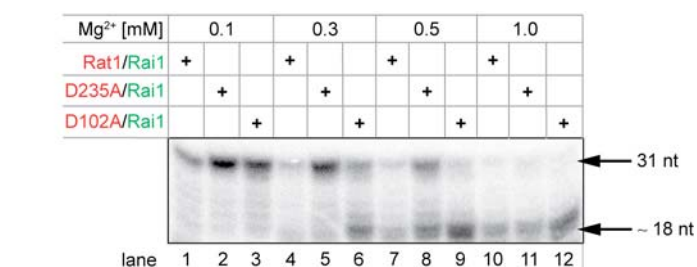
Figure 26: Effect of Rat1 point mutations on ribonucleolytic activity

Autoradiograph of RNAs assembled within ECs, that were used as substrates for mutant Rat1 proteins at different Mg^{2+} concentrations. Samples were treated as depicted in Fig. 22 A (steps a' and b'), separated by 6 M urea PAGE and exposed for several hours to a storage phosphor screen.

The mutation D235A strongly impaired activity at lower magnesium ion concentrations (Fig. 26, lanes 8, 5, 2). Higher magnesium ion concentrations could rescue the mutant defect (Fig. 26, lane 11), apparently by compensating for the decreased magnesium affinity of the active site in the mutant. These results are consistent with mutational studies of Xrn1, where mutations in corresponding conserved residues also affect exonuclease activity to different extents (Page et al, 1998).

Unexpectedly, degradation of RNA at lower magnesium concentrations sometimes did not result in the 18 nt degradation product (Fig 26, lanes 4 and 1). We account this to an overall instability of pure elongation complexes at these conditions, that results in degradation of RNA that is otherwise protected inside of Pol II.

4.4 Conclusions and future perspective

In this work, the production and purification of the Rat1 exoribonuclease in complex with the interacting proteins Rai1 and Rtt103 is shown. The previously reported stabilization of Rat1 nuclease activity by Rai1 was confirmed and it was shown that Rtt103 is not a regulator of Rat1 activity. In addition, the direct interaction of Rat1/Rai1 with the CTD-binding protein Rtt103 was shown for the first time. The results from the *in vitro* termination assay that was established in the course of this work show that Rat1/Rai1 is not a dedicated termination factor as envisaged by the "torpedo model". Nevertheless, existing data strongly suggests the involvement of Rat1 in the termination process and this work strengthens the concept of a "combined allosteric-torpedo model" that was suggested before (Luo et al, 2006). The missing contribution to termination could be a *cis*-element like a DNA sequence, that could change the properties of elongating Pol II. This seems to be reasonable, because the concept of a "weak hybrid" seems to be conserved in the evolution of transcription termination mechanisms (see section 1.3.2). In this light, we tested such a „weak hybrid" that occurs in the 3' regions of protein coding genes, the poly(A) signal. It is known that it does not induce termination, but still it might trigger a "termination competent" form of Pol II that could be terminated by Rat1. This was tested in the *in vitro* assay, but elongation complexes remained stable and were not terminated upon Rat1/Rai1 treatment. Also a crystal structure of the elongation complex bearing the poly(A) site "weak hybrid" within its active site showed no rearrangements of the nucleic acids within Pol II, which might be expected in a "termination competent" Pol II. Neither did essential pause sites influence the outcome of the torpedo reaction. Thus, we conclude that another *trans*-acting factor might be needed for termination. A good candidate for such a factor would be Rtt103. Through binding to the Ser$_2$-phosphorylated CTD of Pol II it could transmit a termination signal to the polymerase, induced by its interaction with Rat1/Rai1. This could be related to a termination mechanism that was reported by Zhang et al (2005). They have shown that the RNA 3'-end processing

factor Pcf11 can "dismantle" elongation complexes. They assume that this happens by Pcf11 forming a bridge between the Pol II CTD and the RNA, that are both bound by this factor. This interaction is thought to exert a force onto the elongation complex that disrupts DNA-RNA hybrid interactions. Unfortunately, we were not able to test this hypothesis in our *in vitro* assay, since we are currently not able to specifically phosphorylate/dephosphorylate residues in the CTD of our Pol II preparations. However, since we can produce the highly pure and active trimeric complex of Rat1/Rai1/Rtt103, this hypothesis should be tested in the future as soon as the experimental limitations have been overcome. Another attractive model for Rat1/Rai1-related termination events is what we call the "triggered-torpedo model". We found homology of the Rat1 N-terminal region to RNase H domains, but could not see RNase H activity (hydrolysis of RNA within DNA-RNA duplexes) in our nuclease preparation (see 4.3.8). Basically, an additional factor could modify the activity of Rat1/Rai1 upon encountering the elongation complex, "triggering" its inherent RNase H activity. Consequently, Rat1 could then attack the RNA within the hybrid of the transcription bubble and disrupt elongation compex interactions. As soon as candidate proteins for such a "trigger" are defined, this model can be tested by our setup. This immediately shows the advantage of our *in vitro* system: it is easily and quickly adjustable to any desired experimental design. First, the coupling of the "substrate" (elongation complexes) to the magnetic beads allows a quick exchange of buffer conditions. Second, the use of sythetic oligonucleotides allows the implementation of any nucleic acid sequence that might modulate the activity of Pol II and/or Rat1/Rai1. Third, availability of fully recombinant exonuclease complexes makes it easy to modify Rat1/Rai1 by point mutations or larger deletions. Fourth, an unlimited number of soluble factors can be added to the assay, to establish a minimal system for termination of Pol II transcription. A good suggestion for such a future experiment would be the addition of the ATP-dependent DNA-RNA-helicase Sen1 when it is available. Sen1 is reported to be involved in termination of sn-/snoRNA transcription and also in termination at some mRNA-coding genes (Lykke-Andersen & Jensen, 2007; Steinmetz et al, 2006). Also the addition of nuclear extract, that is fractionated by biochemical methdos, could help to identify termination activity in one of the fractions. The assay could then be used to monitor this activity and to purify and identify the missing termination factor(s). This procedure would be limited by endogenous nucleases that are active in the nuclear extract. Those need to be specifically inhibited or removed from the extract in order to carry out such an experiment. In addition, the *in vitro* reaction should be tested with endogenous Rat1/Rai1, purified directly from yeast extracts. We cannot exclude that posttranslational modifications, that are not present in our preparation, have an influence on the result of the reaction.

In summary, the *in vitro* assay described in this work, in combination with the fully recombinant and active torpedo-nuclease complexes can be used in future experiments as a tool to elucidate the transcription mechanism for protein coding genes. In addition, the assay was and will be used further in the research on elongation-related events *in vitro*.

5 Appendix

```
cgSpt6    MSDQESASLREEKNEIINGMGSGAPSDEEEGEDVFDSSEEEEEDDEDEARKISEGFIVND  60
paTex     ------------------------------------------------------------


cgSpt6    EDEDEDEDDENVADKKKKRRHKRRAREEEDDRLSEDDLDLLMENAGVKRPTTEGSSGGKL 120
paTex     ------------------------------------------------------------


cgSpt6    KRLKRVGDEVESGAADREQESSQEKESSRLNDFFSDDEDEEPYDEENRERSRGSREYDRG 180
paTex     ------------------------------------------------------------


cgSpt6    ESNHQGDNRKSGMLDELDDFIEEDEFSDEDEETRQQRLLERKMMREQRMKAPTKITGLSS 240
paTex     ------------------------------------------------------------


cgSpt6    DKIDEMYDVFGDGHDYDWALEIENEELEGGDINEQSEEFDEETGMTKSSKKKISLQDIYD 300
paTex     ------------------------------------------------------------


cgSpt6    LQDLKKNLMTDEDMLIRKTDIPERYQELRSGLVNHGNLSDNDQELEKNWISEKIAVDKNF 360
paTex     ------------MDSINTRIAEEISALPSGRVQPQQVAAAVALLSE-------------- 34
                      *    :* *.*.  . * ** *:  :::       *::
                      HtH                                          ①

cgSpt6    SADYDLTEFKEAVGNAIKFITKENLEVPFIYAYRRNYISSRERDGFLLTEDDLWEIVHLD 420
paTex     --------------------GSTVFFIARYRFEVTGSLGDTQLAMLESNLRVLRELE 71
                              .    ****  **::  .*  :   : : *: *   : .*:
              ①

cgSpt6    IEFQSIINKRDYVKRFYSELGISDPLVDEYFKNQSTGSVAELNSLQDIYNYLEFKYAQEI 480
paTex     RREGAILASISIQGKLTPELARDIKLADTKTR------------LEDLY--LPYKQKRRT 117
             .  :*:  . :    :: .**.  .  *.*  :            *:*:*   *  :*  :.
                                                    ②

cgSpt6    NDNLQKESDKSGKKHLKNSNYEKFKSSALYKVIEAVGVSADQIGNNISSQHQIHIPKDHE 540
paTex     KGQIALEAG------LGALADALFDDPTLVPESEAARFVDAEKG-------------FAD 158
          :.::  *:.       *       *...:*     **. .   : *              :
               ③                    *                      ④

cgSpt6    ALKPLELIELVLNENAGDLQVFLSNIKLAMDTIQKYYSWEISKNTKVREKVRADFYRYYL 600
paTex     VKAVLEGAKYILMERFAEDATLLDKLRVFMKNEATLTARVVPG----KEQEGAKFSDYFE 214
          .    **   : :* *.  .:   .:*.:::: *..    :    :.    :*:   *.* *:
                                                       ⑤

cgSpt6    VDVVLTTKGKREIQRGSLYEDIKYAINRTPLHFRREPEIFLKMLEAESLNLMTLKLHMSS 660
paTex     HDEPLKS--APSHRALAIFRGRNEGVLSASLKVGEEAPGTLHPCEVMIAERFGLSNQGRA 272
          *    *.:        . :   :::..  :  .:   :.*:.  .*.    *:   *.   : : *. :

cgSpt6    QKQYVDHLFQIALETTNTSDLAIEWNNFRKAAFTQAIEKIFNDIAQEIKDDLEKTCQKLV 720
paTex     ADKWLAEVVRWTWKVK----------------------LYTHLETDLFGELRDGAEDEA 309
          .:::  .:.:  : :..                       ::..:  :: .:*.. .:. .
                                  ⑥

cgSpt6    CKVVRHKFMTKLDQAPYVPDLKDPKLPKILTLTCGQGRFGSDAIIAAYVNRKGEFVRDFK 780
paTex     ISVFARNLHDLLLAAPAGP-----------RATLGLDPGLRTGVKVAVVDATGKLLDTAT 358
          .*. :: :::    *   ** *            * *   .: .* *:  .*:::
                           ⑦          YqgFC

cgSpt6    ITENPFDRSNPDKFEEVFEDIVQTCQITAIGINGPNPKTQKLFKKLIEVIHKKNLVDSKG 840
paTex     VYPHAP-KNQWDQTLAVLAALCAKHQVELIAIG--NGTASRETDKLAGELIKK----YPG 411
          :  :.  :.: *:   *:  .  . *:  *.*.  *  .::  .**      :  **      *
                                                              ⑧
```

Figure A1: Alignment of Spt6 and Tex protein sequences

Alignment of the *C. glabrata* Spt6 (cgSpt6) and *P. aeruginosa* Tex (paTex) protein sequences by ClustalW. Secondary structure elements are indicated in red and green, for α-helices and β-strands, respectively (information about secondary structure comes from the X-ray structure in the case of Tex and from secondary structure prediction (PROFsec) in the case of Spt6. Only the residues with an expected average accuracy > 82% were highlighted in the alignment). Tex-domains and the Spt6 SH2 domain are colored as in Figs. 9 A and 15. Insertions in Spt6 are numbered as in Fig. 15.

Figure A2: Surface charge distribution of SH2 domains

Surface charge was calculated with APBS (Baker et al, 2001) using the same parameters for all the domains. Binding pockets are indicated by dashed circles (see also Figs. 3, 12 and 13).
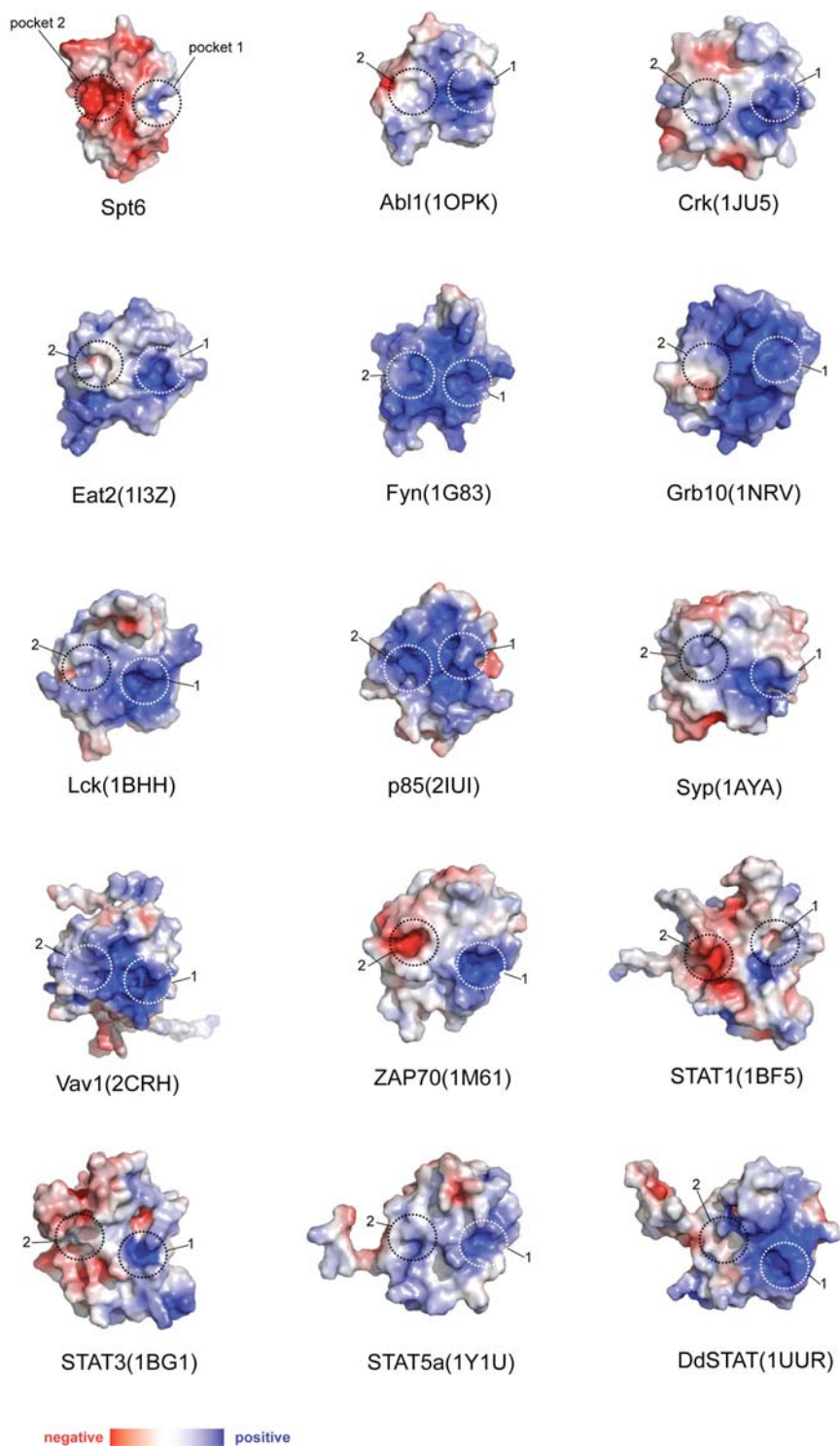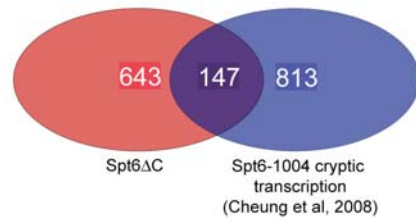
ORF Name	Gene Symbol	ORF Name	Gene Symbol	ORF Name	Gene Symbol
YAL021C	CCR4	YGR217W	CCH1	YNL165W	---
YAL024C	LTE1	YGR233C	PHO81	YNL278W	CAF120
YAL026C	DRS2	YHR106W	TRR2	YNR045W	PET494
YAL063C	FLO1 /// FLO5 /.	YHR131C	---	YNR065C	---
YAR020C	PAU7	YHR165C	PRP8	YNR070W	---
YAR023C	---	YIL005W	EPS1	YOL083W	---
YAR075W	---	YIL073C	SPO22	YOL152W	FRE7
YBL034C	STU1	YJL010C	NOP9	YOL162W	---
YBL060W	YEL1	YJL023C	PET130	YOR005C	DNL4
YBL085W	BOI1	YJL035C	TAD2	YOR011W	AUS1
YBR008C	FLR1	YJL042W	MHP1	YOR049C	RSB1
YBR033W	EDS1	YJL071W	ARG2	YOR071C	THI71
YBR043C	QDR3	YJL078C	PRY3	YOR092W	ECM3
YBR170C	NPL4	YJL109C	UTP10	YOR116C	RPO31
YBR184W	---	YJL197W	UBP12	YOR177C	MPC54
YBR202W	MCM7	YJR092W	BUD4	YOR291W	---
YBR233C	PBP2	YJR094C	IME1	YOR349W	CIN1
YBR270C	---	YJR115W	---	YOR363C	PIP2
YBR280C	SAF1	YKL068W	NUP100	YPL007C	TFC8
YBR284W	---	YKL163W	PIR3	YPL039W	---
YBR294W	SUL1	YKL188C	PXA2	YPL045W	VPS16
YBR295W	PCA1	YKL203C	TOR2	YPL096W	PNG1
YBR296C	PHO89	YKL209C	STE6	YPL120W	VPS30
YCL014W	BUD3	YKL222C	---	YPL164C	MLH3
YCL016C	DCC1	YKR054C	DYN1	YPL167C	REV3
YCL042W	---	YLL039C	UBI4	YPL175W	SPT14
YCR093W	CDC39	YLL053C	---	YPL192C	PRM3
YDL093W	PMT5	YLL060C	GTT2	YPL194W	DDC1
YDR038C	ENA5	YLL061W	MMP1	YPL202C	AFT2
YDR085C	AFR1	YLR106C	MDN1	YPL208W	RKM1
YDR180W	SCC2	YLR137W	---	YPL212C	PUS1
YDR213W	UPC2	YLR139C	SLS1	YPL244C	HUT1
YDR356W	SPC110	YLR214W	FRE1	YPL249C	GYP5
YDR369C	XRS2	YLR256W	HAP1	YPL258C	THI21
YDR386W	MUS81	YLR266C	PDR8	YPR007C	REC8
YDR420W	HKR1	YLR362W	STE11	YPR018W	RLF2
YDR421W	ARO80	YLR430W	SEN1	YPR031W	NTO1
YEL030W	ECM10	YML059C	NTE1	YPR034W	ARP7
YER069W	ARG5,6	YML100W-A	---	YPR105C	COG4
YER113C	---	YML116W	ATR1	YPR111W	DBF20
YER132C	PMD1	YMR006C	PLB2	YPR120C	CLB5
YER184C	---	YMR067C	UBX4	YPR121W	THI22
YGL041C-B	---	YMR109W	MYO5	YPR135W	CTF4
YGL089C	MF(ALPHA)2	YMR207C	HFA1	YPR155C	NCA2
YGL104C	VPS73	YMR266W	RSN1	YPR161C	SGV1
YGL124C	MON1	YMR274C	RCE1	YPR178W	PRP4
YGL141W	HUL5	YMR303C	ADH2	YPR184W	GDB1
YGL143C	MRF1	YMR317W	---	YPR194C	OPT2
YGL156W	AMS1	YNL012W	SPO1		
		YNL040W	---		

Figure A3: Venn diagram depicting the overlap of Spt6ΔC dependent genes with those ORFs showing cryptic transcription in the spt6-1004 mutant.

Of the 960 ORFs known to exhibit cryptic transcription in the Spt6-1004 mutant, the transcript levels of 147 genes are significantly altered in the spt6ΔC mutant. All overlapping genes are listed in the table below. The diagram was drawn with the Partek Genomics Suite 6.3 software. The list of the 960 ORFs is from Cheung et al. The figure was kindly provided by Andreas Mayer.
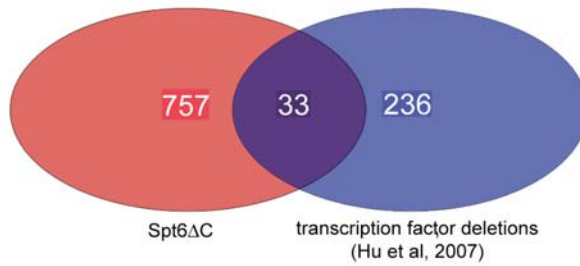
systematic ORF name	gene name
YBR033W	
YBR150C	TBS1
YCL055W	KAR4
YDL170W	UGA3
YDR043C	NRG1
YDR213W	UPC2
YDR216W	ADR1
YDR421W	ARO80
YDR448W	ADA2
YER184C	
YGR249W	MGA1
YIL119C	RPI1
YJL089W	SIP4
YJR094C	IME1
YJR147W	HMS2
YKL043W	PHD1
YKL222C	
YLR136C	TIS11
YLR266C	PDR8
YML007W	YAP1
YML051W	GAL80
YMR273C	ZDS1
YOR162C	YRR1
YOR298C-A	MBF1
YOR363C	PIP2
YPL133C	RDS2
YPL177C	CUP9
YPL202C	AFT2
YPR018W	RLF2
YPR065W	ROX1
YPR193C	HPA2
YPR196W	
YPR199C	ARR1

Figure A4: Venn diagram indicating the overlap of known yeast Pol II transcription factors and spt6ΔC dependent genes.

33 genes coding for various transcription factors are among the significantly changed genes in the spt6ΔC mutant. The diagram was drawn with the Partek Genomics Suite 6.3 software. The list of the 269 yeast transcription factors is taken from Hu et al (2007). The figure was kindly provided by Andreas Mayer.

6 References

Adelman K, Wei W, Ardehali MB, Werner J, Zhu B, Reinberg D, Lis JT (2006) Drosophila Paf1 modulates chromatin structure at actively transcribed genes. *Molecular and cellular biology* **26**(1)**:** 250-260

Adkins MW, Tyler JK (2006) Transcriptional activators are dispensable for transcription in the absence of Spt6-mediated chromatin reassembly of promoter regions. *Molecular cell* **21**(3)**:** 405-416

Ahn SH, Kim M, Buratowski S (2004) Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Molecular cell* **13**(1)**:** 67-76

Amberg DC, Goldstein AL, Cole CN (1992) Isolation and characterization of RAT1: an essential gene of Saccharomyces cerevisiae required for the efficient nucleocytoplasmic trafficking of mRNA. *Genes & development* **6**(7)**:** 1173-1189

Amrani N, Minet M, Wyers F, Dufour ME, Aggerbeck LP, Lacroute F (1997) PCF11 encodes a third protein component of yeast cleavage and polyadenylation factor I. *Molecular and cellular biology* **17**(3)**:** 1102-1109

Anderson D, Koch CA, Grey L, Ellis C, Moran MF, Pawson T (1990) Binding of SH2 domains of phospholipase C gamma 1, GAP, and Src to activated growth factor receptors. *Science (New York, NY* **250**(4983)**:** 979-982

Andrecka J, Lewis R, Bruckner F, Lehmann E, Cramer P, Michaelis J (2008) Single-molecule tracking of mRNA exiting from RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America* **105**(1)**:** 135-140

Andrulis ED, Guzman E, Doring P, Werner J, Lis JT (2000) High-resolution localization of Drosophila Spt5 and Spt6 at heat shock genes in vivo: roles in promoter proximal pausing and transcription elongation. *Genes & development* **14**(20)**:** 2635-2649

Aparicio O, Geisberg JV, Sekinger E, Yang A, Moqtaderi Z, Struhl K (2005) Chromatin immunoprecipitation for determining the association of proteins with specific genomic sequences in vivo. *Curr Protoc Mol Biol* **Chapter 21:** Unit 21 23

Aranda A, Proudfoot NJ (1999) Definition of transcriptional pause elements in fission yeast. *Molecular and cellular biology* **19**(2)**:** 1251-1261

Armache K-J, Mitterweger S, Meinhart A, Cramer P (2005a) Structures of complete RNA polymerase II and its subcomplex Rpb4/7. *J Biol Chem* **280:** 7131-7134

Armache KJ, Kettenberger H, Cramer P (2003) Architecture of initiation-competent 12-subunit RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America* **100**(12)**:** 6964-6968

Armache KJ, Mitterweger S, Meinhart A, Cramer P (2005b) Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *The Journal of biological chemistry* **280**(8)**:** 7131-7134

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics* **25**(1)**:** 25-29

Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America* **98**(18)**:** 10037-10041

Bateman E, Paule MR (1988) Promoter occlusion during ribosomal RNA transcription. *Cell* **54**(7)**:** 985-992

Becker S, Groner B, Muller CW (1998) Three-dimensional structure of the Stat3beta homodimer bound to DNA. *Nature* **394**(6689)**:** 145-151

Belotserkovskaya R, Oh S, Bondarenko VA, Orphanides G, Studitsky VM, Reinberg D (2003) FACT facilitates transcription-dependent nucleosome alteration. *Science (New York, NY* **301**(5636)**:** 1090-1093

Bergfors T (2003) Seeds to crystals. *Journal of structural biology* **142**(1)**:** 66-76

Birse CE, Lee BA, Hansen K, Proudfoot NJ (1997) Transcriptional termination signals for RNA polymerase II in fission yeast. *The EMBO journal* **16**(12)**:** 3633-3643

Birse CE, Minvielle-Sebastia L, Lee BA, Keller W, Proudfoot NJ (1998) Coupling termination of transcription to messenger RNA maturation in yeast. *Science (New York, NY* **280**(5361)**:** 298-301

Blaikie P, Immanuel D, Wu J, Li N, Yajnik V, Margolis B (1994) A region in Shc distinct from the SH2 domain can bind tyrosine-phosphorylated growth factor receptors. *The Journal of biological chemistry* **269**(51)**:** 32031-32034

Booker GW, Breeze AL, Downing AK, Panayotou G, Gout I, Waterfield MD, Campbell ID (1992) Structure of an SH2 domain of the p85 alpha subunit of phosphatidylinositol-3-OH kinase. *Nature* **358**(6388)**:** 684-687

Bortvin A, Winston F (1996) Evidence that Spt6p controls chromatin structure by a direct interaction with histones. *Science (New York, NY* **272**(5267)**:** 1473-1476

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry* **72:** 248-254

Bradshaw JM, Mitaxov V, Waksman G (1999) Investigation of phosphotyrosine recognition by the SH2 domain of the Src kinase. *Journal of molecular biology* **293**(4)**:** 971-985

Brueckner F, Cramer P (2008) Structural basis of transcription inhibition by alpha-amanitin and implications for RNA polymerase II translocation. *Nature structural & molecular biology*

Brueckner F, Hennecke U, Carell T, Cramer P (2007) CPD damage recognition by transcribing RNA polymerase II. *Science (New York, NY* **315**(5813)**:** 859-862

Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta crystallographica* **54**(Pt 5)**:** 905-921

Buratowski S (2003) The CTD code. *Nature structural biology* **10**(9)**:** 679-680

Buratowski S (2005) Connections between mRNA 3' end processing and transcription termination. *Current opinion in cell biology* **17**(3)**:** 257-261

Bushnell DA, Kornberg RD (2003) Complete, 12-subunit RNA polymerase II at 4.1-A resolution: implications for the initiation of transcription. *Proceedings of the National Academy of Sciences of the United States of America* **100**(12)**:** 6969-6973

Bycroft M, Hubbard TJ, Proctor M, Freund SM, Murzin AG (1997) The solution structure of the S1 RNA binding domain: a member of an ancient nucleic acid-binding fold. *Cell* **88**(2)**:** 235-242

Ceruzzi MA, Bektesh SL, Richardson JP (1985) Interaction of rho factor with bacteriophage lambda cro gene transcripts. *The Journal of biological chemistry* **260**(16)**:** 9412-9418

Chapman RD, Heidemann M, Albert TK, Mailhammer R, Flatley A, Meistererrnst M, Kremmer E, Eick D (2007) Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science (New York, NY* **318**(5857)**:** 1780-1782

Chapman RD, Heidemann M, Hintermair C, Eick D (2008) Molecular evolution of the RNA polymerase II CTD. *Trends Genet* **24**(6)**:** 289-296

Chedin S, Riva M, Schultz P, Sentenac A, Carles C (1998) The RNA cleavage activity of RNA polymerase III is mediated by an essential TFIIS-like subunit and is important for transcription termination. *Genes & development* **12**(24)**:** 3857-3871

Chen X, Vinkemeier U, Zhao Y, Jeruzalmi D, Darnell JE, Jr., Kuriyan J (1998) Crystal structure of a tyrosine phosphorylated STAT-1 dimer bound to DNA. *Cell* **93**(5)**:** 827-839

Cheung V, Chua G, Batada NN, Landry CR, Michnick SW, Hughes TR, Winston F (2008) Chromatin- and Transcription-Related Factors Repress Transcription from within Coding Regions throughout the Saccharomyces cerevisiae Genome. *PLoS Biol* **6**(11)**:** e277

Cho EJ, Kobor MS, Kim M, Greenblatt J, Buratowski S (2001) Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. *Genes & development* **15**(24)**:** 3319-3329

Cho EJ, Takagi T, Moore CR, Buratowski S (1997) mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes & development* **11**(24)**:** 3319-3326

Clark-Adams CD, Winston F (1987) The SPT6 gene is essential for growth and is required for delta-mediated transcription in Saccharomyces cerevisiae. *Molecular and cellular biology* **7**(2)**:** 679-686

Connelly S, Manley JL (1988) A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes & development* **2**(4)**:** 440-452

Corden JL, Cadena DL, Ahearn JM, Jr., Dahmus ME (1985) A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America* **82**(23)**:** 7934-7938

Costa PJ, Arndt KM (2000) Synthetic lethal interactions suggest a role for the Saccharomyces cerevisiae Rtf1 protein in transcription elongation. *Genetics* **156**(2)**:** 535-547

Cozzarelli NR, Gerrard SP, Schlissel M, Brown DD, Bogenhagen DF (1983) Purified RNA polymerase III accurately and efficiently terminates transcription of 5S RNA genes. *Cell* **34**(3)**:** 829-835

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

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

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

d'Aubenton Carafa Y, Brody E, Thermes C (1990) Prediction of rho-independent Escherichia coli transcription terminators. A statistical analysis of their RNA stem-loop structures. *Journal of molecular biology* **216**(4)**:** 835-858

Dahmus ME (1996) Reversible phosphorylation of the C-terminal domain of RNA polymerase II. *The Journal of biological chemistry* **271**(32)**:** 19009-19012

Damsma GE, Alt A, Brueckner F, Carell T, Cramer P (2007) Mechanism of transcriptional stalling at cisplatin-damaged DNA. *Nature structural & molecular biology* **14**(12)**:** 1127-1133

Darnell JE, Jr. (1997) STATs and gene regulation. *Science (New York, NY* **277**(5332)**:** 1630-1635

Dichtl B, Blank D, Ohnacker M, Friedlein A, Roeder D, Langen H, Keller W (2002) A role for SSU72 in balancing RNA polymerase II transcription elongation and termination. *Molecular cell* **10**(5)**:** 1139-1150

Eck MJ, Shoelson SE, Harrison SC (1993) Recognition of a high-affinity phosphotyrosyl peptide by the Src homology-2 domain of p56lck. *Nature* **362**(6415)**:** 87-91

Edwards AM, Kane CM, Young RA, Kornberg RD (1991) Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. *The Journal of biological chemistry* **266**(1)**:** 71-75

Egloff S, O'Reilly D, Chapman RD, Taylor A, Tanzhaus K, Pitts L, Eick D, Murphy S (2007) Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science (New York, NY* **318**(5857)**:** 1777-1779

Eichinger L, Pachebat JA, Glockner G, Rajandream MA, Sucgang R, Berriman M, Song J, Olsen R, Szafranski K, Xu Q, Tunggal B, Kummerfeld S, Madera M, Konfortov BA, Rivero F, Bankier AT, Lehmann R, Hamlin N, Davies R, Gaudet P, Fey P, Pilcher K, Chen G, Saunders D, Sodergren E, Davis P, Kerhornou A, Nie X, Hall N, Anjard C, Hemphill L, Bason N, Farbrother P, Desany B, Just E, Morio T, Rost R, Churcher C, Cooper J, Haydock S, van Driessche N, Cronin A, Goodhead I, Muzny D, Mourier T, Pain A, Lu M, Harper D, Lindsay R, Hauser H, James K, Quiles M, Madan Babu M, Saito T, Buchrieser C, Wardroper A, Felder M, Thangavelu M, Johnson D, Knights A, Loulseged H, Mungall K, Oliver K, Price C, Quail MA, Urushihara H, Hernandez J, Rabbinowitsch E, Steffen D, Sanders M, Ma J, Kohara Y, Sharp S, Simmonds M, Spiegler S, Tivey A, Sugano S, White B, Walker D, Woodward J, Winckler T, Tanaka Y, Shaulsky G, Schleicher M, Weinstock G, Rosenthal A, Cox EC, Chisholm RL, Gibbs R, Loomis WF, Platzer M, Kay RR, Williams J, Dear PH, Noegel AA, Barrell B, Kuspa A (2005) The genome of the social amoeba Dictyostelium discoideum. *Nature* **435**(7038)**:** 43-57

El Hage A, Koper M, Kufel J, Tollervey D (2008) Efficient termination of transcription by RNA polymerase I requires the 5' exonuclease Rat1 in yeast. *Genes & development* **22**(8)**:** 1069-1081

Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta crystallographica* **60**(Pt 12 Pt 1)**:** 2126-2132

Endoh M, Zhu W, Hasegawa J, Watanabe H, Kim DK, Aida M, Inukai N, Narita T, Yamada T, Furuya A, Sato H, Yamaguchi Y, Mandal SS, Reinberg D, Wada T, Handa H (2004) Human

Spt6 stimulates transcription elongation by RNA polymerase II in vitro. *Molecular and cellular biology* **24**(8)**:** 3324-3336

Escobedo JA, Kaplan DR, Kavanaugh WM, Turck CW, Williams LT (1991) A phosphatidylinositol-3 kinase binds to platelet-derived growth factor receptors through a specific receptor sequence containing phosphotyrosine. *Molecular and cellular biology* **11**(2)**:** 1125-1132

Fabrega C, Shen V, Shuman S, Lima CD (2003) Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Molecular cell* **11**(6)**:** 1549-1561

Fantl WJ, Escobedo JA, Martin GA, Turck CW, del Rosario M, McCormick F, Williams LT (1992) Distinct phosphotyrosines on a growth factor receptor bind to specific molecules that mediate different signaling pathways. *Cell* **69**(3)**:** 413-423

Farnham PJ, Greenblatt J, Platt T (1982) Effects of NusA protein on transcription termination in the tryptophan operon of Escherichia coli. *Cell* **29**(3)**:** 945-951

Felsenstein J (1989) PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* **5:** 164-166

Fischbeck JA, Kraemer SM, Stargell LA (2002) SPN1, a conserved gene identified by suppression of a postrecruitment-defective yeast TATA-binding protein mutant. *Genetics* **162**(4)**:** 1605-1616

Fish RN, Kane CM (2002) Promoting elongation with transcript cleavage stimulatory factors. *Biochimica et biophysica acta* **1577**(2)**:** 287-307

Fuchs TM, Deppisch H, Scarlato V, Gross R (1996) A new gene locus of Bordetella pertussis defines a novel family of prokaryotic transcriptional accessory proteins. *Journal of bacteriology* **178**(15)**:** 4445-4452

Gao Q, Hua J, Kimura R, Headd JJ, Fu XY, Chin YE (2004) Identification of the linker-SH2 domain of STAT as the origin of the SH2 domain using two-dimensional structural alignment. *Mol Cell Proteomics* **3**(7)**:** 704-714

Geiduschek EP, Kassavetis GA (2001) The RNA polymerase III transcription apparatus. *Journal of molecular biology* **310**(1)**:** 1-26

Gnatt AL, Cramer P, Fu J, Bushnell DA, Kornberg RD (2001) Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. *Science (New York, NY* **292**(5523)**:** 1876-1882

Graber JH, Cantor CR, Mohr SC, Smith TF (1999) Genomic detection of new yeast pre-mRNA 3'-end-processing signals. *Nucleic Acids Res* **27**(3)**:** 888-894

Gu W, Wind M, Reines D (1996) Increased accommodation of nascent RNA in a product site on RNA polymerase II during arrest. *Proceedings of the National Academy of Sciences of the United States of America* **93**(14)**:** 6935-6940

Guglielmi B, Soutourina J, Esnault C, Werner M (2007) TFIIS elongation factor and Mediator act in conjunction during transcription initiation in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **104**(41)**:** 16062-16067

Guo Z, Sherman F (1996) 3'-end-forming signals of yeast mRNA. *Trends in biochemical sciences* **21**(12)**:** 477-481

Gusarov I, Nudler E (1999) The mechanism of intrinsic transcription termination. *Molecular cell* **3**(4)**:** 495-504

Hahn S (2004) Structure and mechanism of the RNA polymerase II transcription machinery. *Nature structural & molecular biology* **11**(5)**:** 394-403

Hani J, Schelbert B, Bernhardt A, Domdey H, Fischer G, Wiebauer K, Rahfeld JU (1999) Mutations in a peptidylprolyl-cis/trans-isomerase gene lead to a defect in 3'-end formation of a pre-mRNA in Saccharomyces cerevisiae. *The Journal of biological chemistry* **274**(1)**:** 108-116

Hartzog GA, Speer JL, Lindstrom DL (2002) Transcript elongation on a nucleoprotein template. *Biochimica et biophysica acta* **1577**(2)**:** 276-286

Hartzog GA, Wada T, Handa H, Winston F (1998) Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in Saccharomyces cerevisiae. *Genes & development* **12**(3)**:** 357-369

He X, Khan AU, Cheng H, Pappas DL, Jr., Hampsey M, Moore CL (2003) Functional interactions between the transcription and mRNA 3' end processing machineries mediated by Ssu72 and Sub1. *Genes & development* **17**(8)**:** 1030-1042

Henderson SL, Ryan K, Sollner-Webb B (1989) The promoter-proximal rDNA terminator augments initiation by preventing disruption of the stable transcription complex caused by polymerase read-in. *Genes & development* **3**(2)**:** 212-223

Holm L, Sander C (1993) Protein structure comparison by alignment of distance matrices. *Journal of molecular biology* **233**(1)**:** 123-138

Hu J, Liu J, Ghirlando R, Saltiel AR, Hubbard SR (2003) Structural basis for recruitment of the adaptor protein APS to the activated insulin receptor. *Molecular cell* **12**(6)**:** 1379-1389

Hu Z, Killion PJ, Iyer VR (2007) Genetic reconstruction of a functional transcriptional regulatory network. *Nature genetics* **39**(5)**:** 683-687

Hyman LE, Seiler SH, Whoriskey J, Moore CL (1991) Point mutations upstream of the yeast ADH2 poly(A) site significantly reduce the efficiency of 3'-end formation. *Molecular and cellular biology* **11**(4)**:** 2004-2012

Izban MG, Luse DS (1992) The RNA polymerase II ternary complex cleaves the nascent transcript in a 3'----5' direction in the presence of elongation factor SII. *Genes & development* **6**(7)**:** 1342-1356

Jansa P, Grummt I (1999) Mechanism of transcription termination: PTRF interacts with the largest subunit of RNA polymerase I and dissociates paused transcription complexes from yeast and mouse. *Mol Gen Genet* **262**(3)**:** 508-514

Jasiak AJ, Hartmann H, Karakasili E, Kalocsay M, Flatley A, Kremmer E, Strasser K, Martin DE, Soding J, Cramer P (2008) Genome-associated RNA polymerase II includes the dissociable Rpb4/7 subcomplex. *The Journal of biological chemistry* **283**(39)**:** 26423-26427

Jenuwein T, Allis CD (2001) Translating the histone code. *Science (New York, NY* **293**(5532)**:** 1074-1080

Johnson AW (1997) Rat1p and Xrn1p are functionally interchangeable exoribonucleases that are restricted to and required in the nucleus and cytoplasm, respectively. *Molecular and cellular biology* **17**(10)**:** 6122-6130

Johnson AW (2001) Rat1p nuclease. *Methods Enzymol* **342:** 260-268

Johnson SJ, Close D, Robinson H, Vallet-Gely I, Dove SL, Hill CP (2008) Crystal Structure and RNA Binding of the Tex Protein from Pseudomonas aeruginosa. *Journal of molecular biology*

Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. *Journal of molecular biology* **292**(2)**:** 195-202

Kabsch W (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Crystallogr* **26:** 795-800

Kaplan CD, Laprade L, Winston F (2003) Transcription elongation factors repress transcription initiation from cryptic sites. *Science (New York, NY* **301**(5636)**:** 1096-1099

Kaplan CD, Morris JR, Wu C, Winston F (2000) Spt5 and spt6 are associated with active transcription and have characteristics of general elongation factors in D. melanogaster. *Genes & development* **14**(20)**:** 2623-2634

Katayanagi K, Miyagawa M, Matsushima M, Ishikawa M, Kanaya S, Ikehara M, Matsuzaki T, Morikawa K (1990) Three-dimensional structure of ribonuclease H from E. coli. *Nature* **347**(6290)**:** 306-309

Kazlauskas A, Ellis C, Pawson T, Cooper JA (1990) Binding of GAP to activated PDGF receptors. *Science (New York, NY* **247**(4950)**:** 1578-1581

Kazlauskas A, Kashishian A, Cooper JA, Valius M (1992) GTPase-activating protein and phosphatidylinositol 3-kinase bind to distinct regions of the platelet-derived growth factor receptor beta subunit. *Molecular and cellular biology* **12**(6)**:** 2534-2544

Kenna M, Stevens A, McCammon M, Douglas MG (1993) An essential yeast gene with homology to the exonuclease-encoding XRN1/KEM1 gene also encodes a protein with exoribonuclease activity. *Molecular and cellular biology* **13**(1)**:** 341-350

Kettenberger H, Armache K-J, Cramer P (2004a) Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol Cell* **16:** 955-965

Kettenberger H, Armache KJ, Cramer P (2003) Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell* **114**(3)**:** 347-357

Kettenberger H, Armache KJ, Cramer P (2004b) Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Molecular cell* **16**(6)**:** 955-965

Kim B, Nesvizhskii AI, Rani PG, Hahn S, Aebersold R, Ranish JA (2007) The transcription elongation factor TFIIS is a component of RNA polymerase II preinitiation complexes. *Proceedings of the National Academy of Sciences of the United States of America* **104**(41)**:** 16068-16073

Kim M, Krogan NJ, Vasiljeva L, Rando OJ, Nedea E, Greenblatt JF, Buratowski S (2004) The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* **432**(7016)**:** 517-522

Kim M, Vasiljeva L, Rando OJ, Zhelkovsky A, Moore C, Buratowski S (2006) Distinct pathways for snoRNA and mRNA termination. *Molecular cell* **24**(5)**:** 723-734

Kimber MS, Nachman J, Cunningham AM, Gish GD, Pawson T, Pai EF (2000) Structural basis for specificity switching of the Src SH2 domain. *Molecular cell* **5**(6)**:** 1043-1049

Kireeva ML, Komissarova N, Waugh DS, Kashlev M (2000) The 8-nucleotide-long RNA:DNA hybrid is a primary stability determinant of the RNA polymerase II elongation complex. *The Journal of biological chemistry* **275**(9)**:** 6530-6536

Kireeva ML, Walter W, Tchernajenko V, Bondarenko V, Kashlev M, Studitsky VM (2002) Nucleosome remodeling induced by RNA polymerase II: loss of the H2A/H2B dimer during transcription. *Molecular cell* **9**(3)**:** 541-552

Kizer KO, Phatnani HP, Shibata Y, Hall H, Greenleaf AL, Strahl BD (2005) A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. *Molecular and cellular biology* **25**(8)**:** 3305-3316

Knop M, Siegers K, Pereira G, Zachariae W, Winsor B, Nasmyth K, Schiebel E (1999) Epitope tagging of yeast genes using a PCR-based strategy: more tags and improved practical routines. *Yeast* **15**(10B)**:** 963-972

Komarnitsky P, Cho EJ, Buratowski S (2000) Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes & development* **14**(19)**:** 2452-2460

Komissarova N, Kireeva ML, Becker J, Sidorenkov I, Kashlev M (2003) Engineering of elongation complexes of bacterial and yeast RNA polymerases. *Methods Enzymol* **371:** 233-251

Korzheva N, Mustaev A, Nudler E, Nikiforov V, Goldfarb A (1998) Mechanistic model of the elongation complex of Escherichia coli RNA polymerase. *Cold Spring Harb Symp Quant Biol* **63:** 337-345

Koschubs T, Seizl M, Lariviere L, Kurth F, Baumli S, Martin DE, Cramer P (2009) Identification, structure, and functional requirement of the Mediator submodule Med7N/31. *The EMBO journal* **28**(1)**:** 69-80

Krogan NJ, Kim M, Ahn SH, Zhong G, Kobor MS, Cagney G, Emili A, Shilatifard A, Buratowski S, Greenblatt JF (2002) RNA polymerase II elongation factors of Saccharomyces cerevisiae: a targeted proteomics approach. *Molecular and cellular biology* **22**(20)**:** 6979-6992

Kufel J, Dichtl B, Tollervey D (1999) Yeast Rnt1p is required for cleavage of the pre-ribosomal RNA in the 3' ETS but not the 5' ETS. *RNA (New York, NY* **5**(7)**:** 909-917

Kuhn CD, Geiger SR, Baumli S, Gartmann M, Gerber J, Jennebach S, Mielke T, Tschochner H, Beckmann R, Cramer P (2007) Functional architecture of RNA polymerase I. *Cell* **131**(7)**:** 1260-1272

Kuriyan J, Cowburn D (1997) Modular peptide recognition domains in eukaryotic signaling. *Annual review of biophysics and biomolecular structure* **26:** 259-288

Labarga A, Valentin F, Anderson M, Lopez R (2007) Web services at the European bioinformatics institute. *Nucleic Acids Res* **35**(Web Server issue)**:** W6-11

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**(5259)**:** 680-685

Landrieux E, Alic N, Ducrot C, Acker J, Riva M, Carles C (2006) A subcomplex of RNA polymerase III subunits involved in transcription termination and reinitiation. *The EMBO journal* **25**(1)**:** 118-128

Lang WH, Morrow BE, Ju Q, Warner JR, Reeder RH (1994) A model for transcription termination by RNA polymerase I. *Cell* **79**(3)**:** 527-534

Lang WH, Reeder RH (1993) The REB1 site is an essential component of a terminator for RNA polymerase I in Saccharomyces cerevisiae. *Molecular and cellular biology* **13**(1)**:** 649-658

Larimer FW, Hsu CL, Maupin MK, Stevens A (1992) Characterization of the XRN1 gene encoding a 5'-->3' exoribonuclease: sequence data and analysis of disparate protein and mRNA levels of gene-disrupted yeast cells. *Gene* **120**(1)**:** 51-57

Larimer FW, Stevens A (1990) Disruption of the gene XRN1, coding for a 5'----3' exoribonuclease, restricts yeast cell growth. *Gene* **95**(1)**:** 85-90

Lee TI, Young RA (2000) Transcription of eukaryotic protein-coding genes. *Annual review of genetics* **34:** 77-137

Lee W, Tillo D, Bray N, Morse RH, Davis RW, Hughes TR, Nislow C (2007) A high-resolution atlas of nucleosome occupancy in yeast. *Nature genetics* **39**(10)**:** 1235-1244

Lei EP, Krebber H, Silver PA (2001) Messenger RNAs are recruited for nuclear export during transcription. *Genes & development* **15**(14)**:** 1771-1782

Li M, Phatnani HP, Guan Z, Sage H, Greenleaf AL, Zhou P (2005) Solution structure of the Set2-Rpb1 interacting domain of human Set2 and its interaction with the hyperphosphorylated C-terminal domain of Rpb1. *Proceedings of the National Academy of Sciences of the United States of America* **102**(49)**:** 17636-17641

Liu BA, Jablonowski K, Raina M, Arce M, Pawson T, Nash PD (2006) The human and mouse complement of SH2 domain proteins-establishing the boundaries of phosphotyrosine signaling. *Molecular cell* **22**(6)**:** 851-868

Logan J, Falck-Pedersen E, Darnell JE, Jr., Shenk T (1987) A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. *Proceedings of the National Academy of Sciences of the United States of America* **84**(23)**:** 8306-8310

Lu PJ, Zhou XZ, Shen M, Lu KP (1999) Function of WW domains as phosphoserine- or phosphothreonine-binding modules. *Science (New York, NY* **283**(5406)**:** 1325-1328

Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**(6648)**:** 251-260

Luke B, Panza A, Redon S, Iglesias N, Li Z, Lingner J (2008) The Rat1p 5' to 3' Exonuclease Degrades Telomeric Repeat-Containing RNA and Promotes Telomere Elongation in Saccharomyces cerevisiae. *Molecular cell* **32**(4)**:** 465-477

Luo W, Johnson AW, Bentley DL (2006) The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model. *Genes & development* **20**(8)**:** 954-965

Lux C, Albiez H, Chapman RD, Heidinger M, Meininghaus M, Brack-Werner R, Lang A, Ziegler M, Cremer T, Eick D (2005) Transition from initiation to promoter proximal pausing requires the CTD of RNA polymerase II. *Nucleic Acids Res* **33**(16)**:** 5139-5144

Lykke-Andersen S, Jensen TH (2007) Overlapping pathways dictate termination of RNA polymerase II transcription. *Biochimie* **89**(10)**:** 1177-1182

Maclennan AJ, Shaw G (1993) A yeast SH2 domain. *Trends in biochemical sciences* **18**(12)**:** 464-465

Marengere LE, Songyang Z, Gish GD, Schaller MD, Parsons JT, Stern MJ, Cantley LC, Pawson T (1994) SH2 domain specificity and activity modified by a single residue. *Nature* **369**(6480)**:** 502-505

Mason PB, Struhl K (2003) The FACT complex travels with elongating RNA polymerase II and is important for the fidelity of transcriptional initiation in vivo. *Molecular and cellular biology* **23**(22)**:** 8323-8333

McCoy AJ (2007) Solving structures of protein complexes by molecular replacement with Phaser. *Acta crystallographica* **63**(Pt 1)**:** 32-41

McCracken S, Fong N, Yankulov K, Ballantyne S, Pan G, Greenblatt J, Patterson SD, Wickens M, Bentley DL (1997) The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**(6614)**:** 357-361

Meinhart A, Cramer P (2004) Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature* **430**(6996)**:** 223-226

Meinhart A, Kamenski T, Hoeppner S, Baumli S, Cramer P (2005) A structural perspective of CTD function. *Genes & development* **19**(12)**:** 1401-1415

Moran MF, Koch CA, Anderson D, Ellis C, England L, Martin GS, Pawson T (1990) Src homology region 2 domains direct protein-protein interactions in signal transduction. *Proceedings of the National Academy of Sciences of the United States of America* **87**(21)**:** 8622-8626

Mueller CL, Jaehning JA (2002) Ctr9, Rtf1, and Leo1 are components of the Paf1/RNA polymerase II complex. *Molecular and cellular biology* **22**(7)**:** 1971-1980

Muller AJ, Pendergast AM, Havlik MH, Puil L, Pawson T, Witte ON (1992) A limited set of SH2 domains binds BCR through a high-affinity phosphotyrosine-independent interaction. *Molecular and cellular biology* **12**(11)**:** 5087-5093

Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta crystallographica* **53**(Pt 3)**:** 240-255

Nehrke KW, Zalatan F, Platt T (1993) NusG alters rho-dependent termination of transcription in vitro independent of kinetic coupling. *Gene Expr* **3**(2)**:** 119-133

Nonet M, Sweetser D, Young RA (1987) Functional redundancy and structural polymorphism in the large subunit of RNA polymerase II. *Cell* **50**(6)**:** 909-915

Nowotny M, Gaidamakov SA, Crouch RJ, Yang W (2005) Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell* **121**(7)**:** 1005-1016

Nowotny M, Gaidamakov SA, Ghirlando R, Cerritelli SM, Crouch RJ, Yang W (2007) Structure of human RNase H1 complexed with an RNA/DNA hybrid: insight into HIV reverse transcription. *Molecular cell* **28**(2)**:** 264-276

Nudler E, Gottesman ME (2002) Transcription termination and anti-termination in E. coli. *Genes Cells* **7**(8)**:** 755-768

Nudler E, Mustaev A, Lukhtanov E, Goldfarb A (1997) The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell* **89**(1)**:** 33-41

Orphanides G, Reinberg D (2002) A unified theory of gene expression. *Cell* **108**(4)**:** 439-451

Orphanides G, Wu WH, Lane WS, Hampsey M, Reinberg D (1999) The chromatin-specific transcription elongation factor FACT comprises human SPT16 and SSRP1 proteins. *Nature* **400**(6741)**:** 284-288

Osborne BI, Guarente L (1989) Mutational analysis of a yeast transcriptional terminator. *Proceedings of the National Academy of Sciences of the United States of America* **86**(11)**:** 4097-4101

Osheim YN, Proudfoot NJ, Beyer AL (1999) EM visualization of transcription by RNA polymerase II: downstream termination requires a poly(A) signal but not transcript cleavage. *Molecular cell* **3**(3)**:** 379-387

Osheim YN, Sikes ML, Beyer AL (2002) EM visualization of Pol II genes in Drosophila: most genes terminate without prior 3' end cleavage of nascent transcripts. *Chromosoma* **111**(1)**:** 1-12

Otwinowski Z, Minor W. (1996) Processing of X-ray diffraction data collected in oscillation mode. *Meth Enzym* **276:** 307-326

Overduin M, Rios CB, Mayer BJ, Baltimore D, Cowburn D (1992) Three-dimensional solution structure of the src homology 2 domain of c-abl. *Cell* **70**(4)**:** 697-704

Page AM, Davis K, Molineux C, Kolodner RD, Johnson AW (1998) Mutational analysis of exoribonuclease I from Saccharomyces cerevisiae. *Nucleic Acids Res* **26**(16)**:** 3707-3716

Pal M, Luse DS (2002) Strong natural pausing by RNA polymerase II within 10 bases of transcription start may result in repeated slippage and reextension of the nascent RNA. *Molecular and cellular biology* **22**(1)**:** 30-40

Pal M, Luse DS (2003) The initiation-elongation transition: lateral mobility of RNA in RNA polymerase II complexes is greatly reduced at +8/+9 and absent by +23. *Proceedings of the National Academy of Sciences of the United States of America* **100**(10)**:** 5700-5705

Pape T, Schneider TR (2004) HKL2MAP: a graphical user interface for phasing with SHELX programs. *Acta Crystallogr D Biol Crystallogr* **37:** 843-844

Park JS, Marr MT, Roberts JW (2002) E. coli Transcription repair coupling factor (Mfd protein) rescues arrested complexes by promoting forward translocation. *Cell* **109**(6)**:** 757-767

Park JS, Roberts JW (2006) Role of DNA bubble rewinding in enzymatic transcription termination. *Proceedings of the National Academy of Sciences of the United States of America* **103**(13)**:** 4870-4875

Pei Y, Shuman S (2002) Interactions between fission yeast mRNA capping enzymes and elongation factor Spt5. *The Journal of biological chemistry* **277**(22)**:** 19639-19648

Pendergast AM, Muller AJ, Havlik MH, Maru Y, Witte ON (1991) BCR sequences essential for transformation by the BCR-ABL oncogene bind to the ABL SH2 regulatory domain in a non-phosphotyrosine-dependent manner. *Cell* **66**(1)**:** 161-171

Penheiter KL, Washburn TM, Porter SE, Hoffman MG, Jaehning JA (2005) A posttranscriptional role for the yeast Paf1-RNA polymerase II complex is revealed by identification of primary targets. *Molecular cell* **20**(2)**:** 213-223

Poglitsch CL, Meredith GD, Gnatt AL, Jensen GJ, Chang WH, Fu J, Kornberg RD (1999) Electron crystal structure of an RNA polymerase II transcription elongation complex. *Cell* **98**(6)**:** 791-798

Ponting CP (2002) Novel domains and orthologues of eukaryotic transcription elongation factors. *Nucleic Acids Res* **30**(17)**:** 3643-3652

Poole TL, Stevens A (1995) Comparison of features of the RNase activity of 5'-exonuclease-1 and 5'-exonuclease-2 of Saccharomyces cerevisiae. *Nucleic acids symposium series*(33)**:** 79-81

Prescott EM, Osheim YN, Jones HS, Alen CM, Roan JG, Reeder RH, Beyer AL, Proudfoot NJ (2004) Transcriptional termination by RNA polymerase I requires the small subunit Rpa12p. *Proceedings of the National Academy of Sciences of the United States of America* **101**(16)**:** 6068-6073

Proudfoot NJ (1989) How RNA polymerase II terminates transcription in higher eukaryotes. *Trends in biochemical sciences* **14**(3)**:** 105-110

Proudfoot NJ, Brownlee GG (1976) 3' non-coding region sequences in eukaryotic messenger RNA. *Nature* **263**(5574)**:** 211-214

Proudfoot NJ, Furger A, Dye MJ (2002) Integrating mRNA processing with transcription. *Cell* **108**(4)**:** 501-512

Puig O, Caspary F, Rigaut G, Rutz B, Bouveret E, Bragado-Nilsson E, Wilm M, Seraphin B (2001) The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* **24**(3)**:** 218-229

Rahuel J, Gay B, Erdmann D, Strauss A, Garcia-Echeverria C, Furet P, Caravatti G, Fretz H, Schoepfer J, Grutter MG (1996) Structural basis for specificity of Grb2-SH2 revealed by a novel ligand binding mode. *Nature structural biology* **3**(7)**:** 586-589

Reeder RH, Guevara P, Roan JG (1999) Saccharomyces cerevisiae RNA polymerase I terminates transcription at the Reb1 terminator in vivo. *Molecular and cellular biology* **19**(11)**:** 7369-7376

Reeder RH, Lang WH (1997) Terminating transcription in eukaryotes: lessons learned from RNA polymerase I. *Trends in biochemical sciences* **22**(12)**:** 473-477

Reines D (1992) Elongation factor-dependent transcript shortening by template-engaged RNA polymerase II. *The Journal of biological chemistry* **267**(6)**:** 3795-3800

Roberts J, Park JS (2004) Mfd, the bacterial transcription repair coupling factor: translocation, repair and termination. *Curr Opin Microbiol* **7**(2)**:** 120-125

Roberts JW (1969) Termination factor for RNA synthesis. *Nature* **224**(5225)**:** 1168-1174

Rondon AG, Garcia-Rubio M, Gonzalez-Barrera S, Aguilera A (2003) Molecular evidence for a positive role of Spt4 in transcription elongation. *The EMBO journal* **22**(3)**:** 612-620

Ronnstrand L, Mori S, Arridsson AK, Eriksson A, Wernstedt C, Hellman U, Claesson-Welsh L, Heldin CH (1992) Identification of two C-terminal autophosphorylation sites in the PDGF beta-receptor: involvement in the interaction with phospholipase C-gamma. *The EMBO journal* **11**(11)**:** 3911-3919

Roussel RR, Brodeur SR, Shalloway D, Laudano AP (1991) Selective binding of activated pp60c-src by an immobilized synthetic phosphopeptide modeled on the carboxyl terminus of pp60c-src. *Proceedings of the National Academy of Sciences of the United States of America* **88**(23)**:** 10696-10700

Russo P, Li WZ, Guo Z, Sherman F (1993) Signals that produce 3' termini in CYC1 mRNA of the yeast Saccharomyces cerevisiae. *Molecular and cellular biology* **13**(12)**:** 7836-7849

Sadowski I, Stone JC, Pawson T (1986) A noncatalytic domain conserved among cytoplasmic protein-tyrosine kinases modifies the kinase function and transforming activity of Fujinami sarcoma virus P130gag-fps. *Molecular and cellular biology* **6**(12)**:** 4396-4408

Sadowski M, Dichtl B, Hubner W, Keller W (2003) Independent functions of yeast Pcf11p in pre-mRNA 3' end processing and in transcription termination. *The EMBO journal* **22**(9)**:** 2167-2177

Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J (2003) TM4: a free, open-source system for microarray data management and analysis. *BioTechniques* **34**(2)**:** 374-378

Schmidt MC, Chamberlin MJ (1987) nusA protein of Escherichia coli is an efficient transcription termination factor for certain terminator sites. *Journal of molecular biology* **195**(4)**:** 809-818

Schroeder SC, Schwer B, Shuman S, Bentley D (2000) Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes & development* **14**(19)**:** 2435-2440

Sims RJ, 3rd, Belotserkovskaya R, Reinberg D (2004) Elongation by RNA polymerase II: the short and long of it. *Genes & development* **18**(20)**:** 2437-2468

Soler-Lopez M, Petosa C, Fukuzawa M, Ravelli R, Williams JG, Muller CW (2004) Structure of an activated Dictyostelium STAT in its DNA-unbound form. *Molecular cell* **13**(6)**:** 791-804

Solinger JA, Pascolini D, Heyer WD (1999) Active-site mutations in the Xrn1p exoribonuclease of Saccharomyces cerevisiae reveal a specific role in meiosis. *Molecular and cellular biology* **19**(9)**:** 5930-5942

Songyang Z, Shoelson SE, Chaudhuri M, Gish G, Pawson T, Haser WG, King F, Roberts T, Ratnofsky S, Lechleider RJ, et al. (1993) SH2 domains recognize specific phosphopeptide sequences. *Cell* **72**(5)**:** 767-778

Steinmetz EJ, Warren CL, Kuehner JN, Panbehi B, Ansari AZ, Brow DA (2006) Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. *Molecular cell* **24**(5)**:** 735-746

Steitz TA (1998) A mechanism for all polymerases. *Nature* **391**(6664)**:** 231-232

Stevens A (1978) An exoribonuclease from Saccharomyces cerevisiae: effect of modifications of 5' end groups on the hydrolysis of substrates to 5' mononucleotides. *Biochemical and biophysical research communications* **81**(2)**:** 656-661

Stevens A (1980) Purification and characterization of a Saccharomyces cerevisiae exoribonuclease which yields 5'-mononucleotides by a 5' leads to 3' mode of hydrolysis. *The Journal of biological chemistry* **255**(7)**:** 3080-3085

Stevens A, Maupin MK (1987) A 5'----3' exoribonuclease of Saccharomyces cerevisiae: size and novel substrate specificity. *Archives of biochemistry and biophysics* **252**(2)**:** 339-347

Stevens A, Poole TL (1995) 5'-exonuclease-2 of Saccharomyces cerevisiae. Purification and features of ribonuclease activity with comparison to 5'-exonuclease-1. *The Journal of biological chemistry* **270**(27)**:** 16063-16069

Strahl BD, Allis CD (2000) The language of covalent histone modifications. *Nature* **403**(6765)**:** 41-45

Strasser K, Hurt E (2001) Splicing factor Sub2p is required for nuclear mRNA export through its interaction with Yra1p. *Nature* **413**(6856)**:** 648-652

Studier FW (2005) Protein production by auto-induction in high density shaking cultures. *Protein expression and purification* **41**(1)**:** 207-234

Sullivan SL, Gottesman ME (1992) Requirement for E. coli NusG protein in factor-dependent transcription termination. *Cell* **68**(5)**:** 989-994

Swanson MS, Carlson M, Winston F (1990) SPT6, an essential gene that affects transcription in Saccharomyces cerevisiae, encodes a nuclear protein with an extremely acidic amino terminus. *Molecular and cellular biology* **10**(9)**:** 4935-4941

Swanson MS, Winston F (1992) SPT4, SPT5 and SPT6 interactions: effects on transcription and viability in Saccharomyces cerevisiae. *Genetics* **132**(2)**:** 325-336

Teixeira A, Tahiri-Alaoui A, West S, Thomas B, Ramadass A, Martianov I, Dye M, James W, Proudfoot NJ, Akoulitchev A (2004) Autocatalytic RNA cleavage in the human beta-globin pre-mRNA promotes transcription termination. *Nature* **432**(7016)**:** 526-530

Terwilliger TC (2002) Automated structure solution, density modification and model building. *Acta crystallographica* **58**(Pt 11)**:** 1937-1940

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

Tschochne H, Milkereit P (1997) RNA polymerase I from S. cerevisiae depends on an additional factor to release terminated transcripts from the template. *FEBS letters* **410**(2-3)**:** 461-466

Tuske S, Sarafianos SG, Wang X, Hudson B, Sineva E, Mukhopadhyay J, Birktoft JJ, Leroy O, Ismail S, Clark AD, Jr., Dharia C, Napoli A, Laptenko O, Lee J, Borukhov S, Ebright RH, Arnold E (2005) Inhibition of bacterial RNA polymerase by streptolydigin: stabilization of a straight-bridge-helix active-center conformation. *Cell* **122**(4)**:** 541-552

Vasiljeva L, Kim M, Mutschler H, Buratowski S, Meinhart A (2008) The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nature structural & molecular biology* **15**(8)**:** 795-804

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

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

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

Verdecia MA, Bowman ME, Lu KP, Hunter T, Noel JP (2000) Structural basis for phosphoserine-proline recognition by group IV WW domains. *Nature structural biology* **7**(8)**:** 639-643

Verma-Gaur J, Rao SN, Taya T, Sadhale P (2008) Genomewide recruitment analysis of Rpb4, a subunit of polymerase II in Saccharomyces cerevisiae, reveals its involvement in transcription elongation. *Eukaryot Cell* **7**(6)**:** 1009-1018

Vojnic E, Simon B, Strahl BD, Sattler M, Cramer P (2006) Structure and carboxyl-terminal domain (CTD) binding of the Set2 SRI domain that couples histone H3 Lys36 methylation to transcription. *The Journal of biological chemistry* **281**(1)**:** 13-15

Wada T, Takagi T, Yamaguchi Y, Ferdous A, Imai T, Hirose S, Sugimoto S, Yano K, Hartzog GA, Winston F, Buratowski S, Handa H (1998) DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes & development* **12**(3)**:** 343-356

Waksman G, Kominos D, Robertson SC, Pant N, Baltimore D, Birge RB, Cowburn D, Hanafusa H, Mayer BJ, Overduin M, et al. (1992) Crystal structure of the phosphotyrosine recognition domain SH2 of v-src complexed with tyrosine-phosphorylated peptides. *Nature* **358**(6388)**:** 646-653

Waksman G, Kuriyan J (2004) Structure and specificity of the SH2 domain. *Cell* **116**(2 Suppl)**:** S45-48, 43 p following S48

Waksman G, Shoelson SE, Pant N, Cowburn D, Kuriyan J (1993) Binding of a high affinity phosphotyrosyl peptide to the Src SH2 domain: crystal structures of the complexed and peptide-free forms. *Cell* **72**(5)**:** 779-790

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

Wang D, Hawley DK (1993) Identification of a 3'-->5' exonuclease activity associated with human RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America* **90**(3)**:** 843-847

Ward DF, Gottesman ME (1981) The nus mutations affect transcription termination in Escherichia coli. *Nature* **292**(5820)**:** 212-215

West ML, Corden JL (1995) Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. *Genetics* **140**(4)**:** 1223-1233

West S, Gromak N, Proudfoot NJ (2004) Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* **432**(7016)**:** 522-525

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

Wickens M, Stephenson P (1984) Role of the conserved AAUAAA sequence: four AAUAAA point mutants prevent messenger RNA 3' end formation. *Science (New York, NY* **226**(4678)**:** 1045-1051

Williams JG, Noegel AA, Eichinger L (2005) Manifestations of multicellularity: Dictyostelium reports in. *Trends Genet* **21**(7)**:** 392-398

Williams JG, Zvelebil M (2004) SH2 domains in plants imply new signalling scenarios. *Trends in plant science* **9**(4)**:** 161-163

Winston F, Chaleff DT, Valent B, Fink GR (1984) Mutations affecting Ty-mediated expression of the HIS4 gene of Saccharomyces cerevisiae. *Genetics* **107**(2)**:** 179-197

Xue Y, Bai X, Lee I, Kallstrom G, Ho J, Brown J, Stevens A, Johnson AW (2000) Saccharomyces cerevisiae RAI1 (YGL246c) is homologous to human DOM3Z and encodes a protein that binds the nuclear exoribonuclease Rat1p. *Molecular and cellular biology* **20**(11)**:** 4006-4015

Yaffe MB (2002) Phosphotyrosine-binding domains in signal transduction. *Nature reviews* **3**(3)**:** 177-186

Yamada T, Yamaguchi Y, Inukai N, Okamoto S, Mura T, Handa H (2006) P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. *Molecular cell* **21**(2)**:** 227-237

Yamaguchi Y, Takagi T, Wada T, Yano K, Furuya A, Sugimoto S, Hasegawa J, Handa H (1999) NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* **97**(1)**:** 41-51

Yang W, Hendrickson WA, Crouch RJ, Satow Y (1990) Structure of ribonuclease H phased at 2 A resolution by MAD analysis of the selenomethionyl protein. *Science (New York, NY* **249**(4975)**:** 1398-1405

Yarnell WS, Roberts JW (1999) Mechanism of intrinsic transcription termination and antitermination. *Science (New York, NY* **284**(5414)**:** 611-615

Yoakim M, Hou W, Liu Y, Carpenter CL, Kapeller R, Schaffhausen BS (1992) Interactions of polyomavirus middle T with the SH2 domains of the pp85 subunit of phosphatidylinositol-3-kinase. *J Virol* **66**(9)**:** 5485-5491

Yoh SM, Cho H, Pickle L, Evans RM, Jones KA (2007) The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. *Genes & development* **21**(2)**:** 160-174

Yoh SM, Lucas JS, Jones KA (2008) The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. *Genes & development* **22**(24)**:** 3422-3434

Youdell ML, Kizer KO, Kisseleva-Romanova E, Fuchs SM, Duro E, Strahl BD, Mellor J (2008) Roles for Ctk1 and Spt6 in regulating the different methylation states of Histone H3 lysine 36. *Molecular and cellular biology*

Zhang G, Campbell EA, Minakhin L, Richter C, Severinov K, Darst SA (1999) Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution. *Cell* **98**(6)**:** 811-824

Zhang J, Corden JL (1991) Phosphorylation causes a conformational change in the carboxyl-terminal domain of the mouse RNA polymerase II largest subunit. *The Journal of biological chemistry* **266**(4)**:** 2297-2302

Zhang L, Fletcher AG, Cheung V, Winston F, Stargell LA (2008) Spn1 regulates the recruitment of Spt6 and the Swi/Snf complex during transcriptional activation by RNA polymerase II. *Molecular and cellular biology* **28**(4)**:** 1393-1403

Zhang Z, Fu J, Gilmour DS (2005) CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11. *Genes & development* **19**(13)**:** 1572-1580

Zheng C, Friedman DI (1994) Reduced Rho-dependent transcription termination permits NusA-independent growth of Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **91**(16)**:** 7543-7547

Zheng Q, Wang XJ (2008) GOEAST: a web-based software toolkit for Gene Ontology enrichment analysis. *Nucleic Acids Res* **36**(Web Server issue)**:** W358-363

Zuo Y, Deutscher MP (2001) Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. *Nucleic Acids Res* **29**(5)**:** 1017-1026

Curriculum vitae

Stefan Dengl

Date of birth: December 10, 1977

Place of birth: Munich, Germany

Nationality: German

Family status: unmarried

Education

1984 - 1988	Grund- und Hauptschule Kirchheim b. München
1988 - 1997	Franz-Marc-Gymnasium Markt Schwaben
1997 - 1998	Civilian service at Bezirkskrankenhaus Haar b. München
1998-2004	Studies in biology at the Ludwig-Maximilians-Universität, München
2003-2004	Diploma thesis at the Department Biologie I, Bereich Mikrobiologie, group of Prof. Dr. Volker Müller
	Title of the thesis: "Biochemische und molekularbiologische Analyse der Hydrogenasen in *Acetobacterium woodii".*
June 2004	Diploma degree in biology
since July 2004	PhD thesis at the Gene Center of the Ludwig-Maximilians-Universität, München in the research group of Prof. Dr. Patrick Cramer
	Title of the thesis: "Structure and requirement of the Spt6 SH2 domain and an *in vitro* system to test the "torpedo model" of transcription termination"