NMR solution structure of the Set2 SRI domain and preparation of RNA polymerase II complexes with the elongation factor Spt4-Spt5

Erika Vojnić

aus Augsburg, Deutschland

2006

Erklärung

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

Ehrenwörtliche Versicherung

Diese Dissertation wurde selbständig und ohne unerlaubte Hilfe erarbeitet.

München, den 06. November 2006

_____

Erika Vojnić

Dissertation eingereicht am 07. November 2006

1. Gutachter: Prof. Dr. Patrick Cramer

2. Gutachter: Prof. Dr. Karl-Peter Hopfner

Mündliche Prüfung am 07. Dezember 2006

Acknowledgments

First of all, I would like to thank my supervisor Prof. Dr. Patrick Cramer for giving me the opportunity to work on a challenging project in an excellent environment both in Munich and Heidelberg.

I am much obliged to the fruitful collaboration with Drs. Michael Sattler and Bernd Simon, who instructed me in NMR spectroscopy and convinced me that not only X-ray crystallography is a prolific technique to solve biological problems.

My thanks go to all present and former members of the Cramer lab for the scientific support and the enjoyable lab atmosphere. I am grateful for the support I got from Hubert, Karim and Florian – Unix and Denzo are not wrapped in mystery any more. Eli – thank you for reading the manuscript. In particular, I would like to thank Kristin, Ania, Sonja and Karim for their friendship.

My special appreciation goes to Gunther Stier who impressed me with his vast knowledge about cloning – there are more than NdeI/NotI restriction sites.

Mojim roditeljima: Zahvaljujem se za ukazanu ljubav, tradiciju i podržavanje sopstvene dece na životnom putu po običaju Vašeg rodnog kraja.

I would also like to thank my friends for their constant support and interest in my work.

Ić – Vielen Dank für deinen Humor!

Danke Simone.

Summary

RNA polymerase II (RNAP II) transiently associates with many different proteins and multiprotein complexes during the mRNA transcription cycle, which includes three phases, initiation, elongation, and termination. This thesis describes structural studies of two factors that facilitate transcription through chromatin. The heterodimeric *Saccharomyces cerevisiae* elongation factor Spt4-Spt5 (human DSIF) has been identified by biochemical and genetic approaches to help RNAP II transcribe through chromatin. It is assumed that Spt4-Spt5 pauses RNAP II to open a time window for capping enzyme recruitment and addition of a cap to the 5'-end of the nascent RNA. The preparation of milligram quantities of soluble Spt4-Spt5 variants that are suited for structural studies has been achieved. Several strategies to resolve the structure of the RNAP II–Spt4-Spt5 complex were unsuccessful, possibly indicating an intrinsic flexibility of the complex. In addition, there is now evidence for direct links between chromatin modification and transcription elongation. A major player in this process is the histone lysine methyltransferase Set2 which has a modular structure. The catalytic activity of Set2 is mediated by the SET [Su(var)3-9, Enhancer of Zeste, Trithorax] domain. During mRNA elongation, the SRI (Set2 Rpb1-interacting) domain of Set2 binds to the phosphorylated CTD (carboxyl-terminal domain) of RNAP II. The NMR solution structure of yeast Set2 SRI domain has been determined. The structure reveals a novel CTD-binding fold consisting of a left-handed three-helix bundle. Unexpectedly, the SRI domain fold resembles the structure of an RNA polymerase-interacting domain in σ factors that mediate transcription initiation in bacteria (domain σ_2 in σ^{70}). NMR titration experiments show that the SRI domain binds a Ser2/Ser5-phosphorylated CTD peptide comprising two heptapeptide repeats and three flanking NH2-terminal residues. Amino acid residues that show strong chemical shift perturbations upon CTD binding cluster in two regions on the SRI surface. The results will enable a detailed analysis of the specific CTD interactions underlying the coupling of transcription and chromatin modification by Set2.

Part of this work has been published

Vojnic E., Simon B., Strahl B. D., Sattler M. and 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, Epub 2005 Nov 14.

Chapter I: Introduction

1 THE TRANSCRIPTION CYCLE

The generation of mature mRNA molecules by eukaryotic RNA polymerase II (RNAP II = 12-subunit RNAP II), is a multistage process consisting of three major steps: initiation, elongation, and termination. Typically the transcription cycle starts with the binding of transcription factors and RNAP II to promoter sites and the onset of RNA synthesis by promoter clearance. During productive transcript elongation RNAP II moves along a protein-coding gene and synthesizes an mRNA copy. Finally, guided by DNA sequences at the end of a gene, RNAP II terminates transcription and releases newly synthesized mRNA. Mature mRNA is capped, cleaved, polyadenylated, and transported to the cytoplasm, where it is finally translated by ribosomes. Concomitantly, RNAP II is recycled and transcription reinitiation can occur [Figure 1; (reviewed by Sims 3rd *et al.*, 2004; reviewed by Hahn, 2004)]. Although much has been learned about processes involved in initiation, much less is known regarding mechanisms that regulate later stages in the transcription process, particularly chain elongation. Therefore, great strides have been made to unveil mechanisms underlying transcription elongation, as it is one of the most important steps in the control of cell growth and differentiation. It is known, however, that the transition from transcriptional initiation to elongation is accompanied by a partial disassembly of the initiation complex and the recruitment of an elongation complex that consists of a different set of factors (Pokholok *et al.*, 2002).
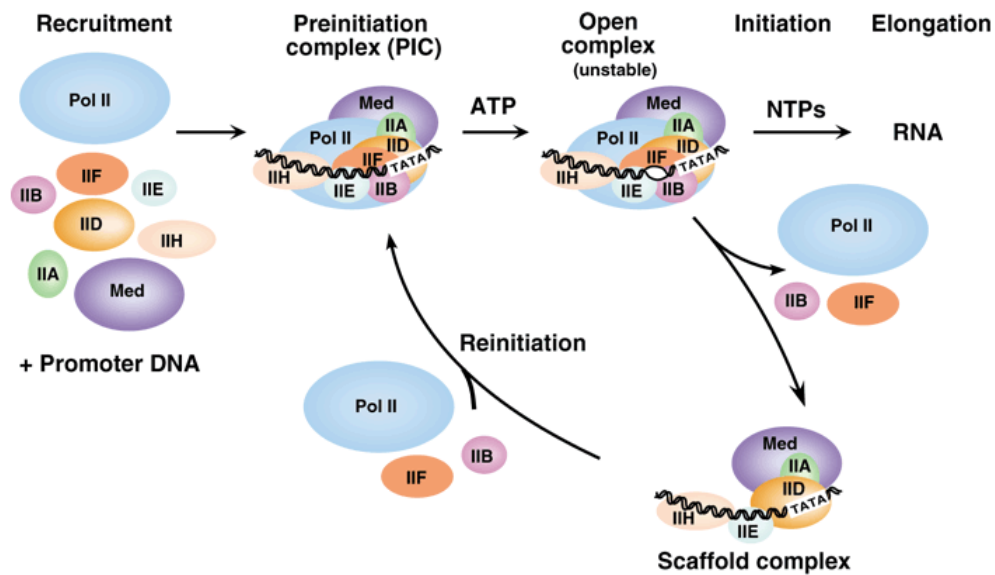
Figure 1. **The RNAP transcription cycle.** The figure is taken from Hahn (2004).

Transcription elongation at the level of chromatin is targeted by a multitude of transcription factors. Several lines of evidence led to the view that, in contrast to a simple linear assembly line, a complex and extensively coupled network has evolved to coordinate the activities of the gene expression machines (reviewed by Maniatis and Reed, 2002). Thus, the ultimate goal of understanding the regulation of gene expression in a physiological context will be achieved by taking into account the coordinated integration of basal transcriptional components, co-factors and chromatin. There are at least two key regulatory mechanisms that must occur during elongation. One is the maintenance of an 'open chromatin' structure so that RNAP II can traverse nucleosomes; the other is the organization of the capping, splicing, and polyadenylation reactions. Factors regulating transcript elongation on chromatin templates are represented by different classes of proteins: histone chaperones, ATP-dependent chromatin remodelers, histone-modifying enzymes, and topoisomerases (Reinberg and Sims 3rd, 2006). On the other hand, vast studies aimed at elucidating the cadre of elongation factors that either modulate the catalytic activity of RNAP II or serve to alleviate impediments to transcript elongation, including drug-induced or sequence dependent transcriptional pause and arrest.

2   STRUCTURE OF CHROMATIN IN TRANSCRIBED GENES

2.1   The Histone Code - Nucleosomes as regulatory units

In the nucleus of cells, chromatin represents the physiological state of DNA, where it is associated with histone and non-histone proteins. The basic building block of chromatin is the nucleosome, a structure consisting of an octamer of four core histone proteins around which 147 base pairs of DNA are wrapped. Histone proteins H2A, H2B, H3 and H4 are each composed of a structured globular domain and an unstructured tail domain (Luger *et al.,* 1997). The amino termini of histones are highly accessible and experience numerous specific post-translational modifications. Especially residues of the amino termini of histones H3 and H4 and the amino and carboxyl termini of histones H2A, H2B and H1, are susceptible to an assortment of covalent alterations, including acetylation of lysines, methylation of lysines and arginines, phosphorylation of serines and threonines, ubiquitination of lysines, sumoylation of lysines, and the ADP-ribosylation of glutamic acids (Figure 2). The first three types of modifications have been studied extensively (Grunstein, 1997; reviewed by Nowak and Corces, 2004; reviewed by Martin and Zhang, 2005; Pokholok *et al.*, 2005).

These chromatin 'marks' are suggested to behave in a combinatorial manner defining a second code, the 'histone code', devoted to epigenetic regulation (Strahl and Allis, 2000; Jenuwein and Allis, 2001), thereby extending the information content of the genome past the genetic code. A systematic nomenclature for modified histones, termed the Brno nomenclature, tries to present the complex encoded information in a consistent and coherent manner (Turner, 2005). For instance, a triple methylation mark on lysine 4 of histone H3 is designated H3K4me3.
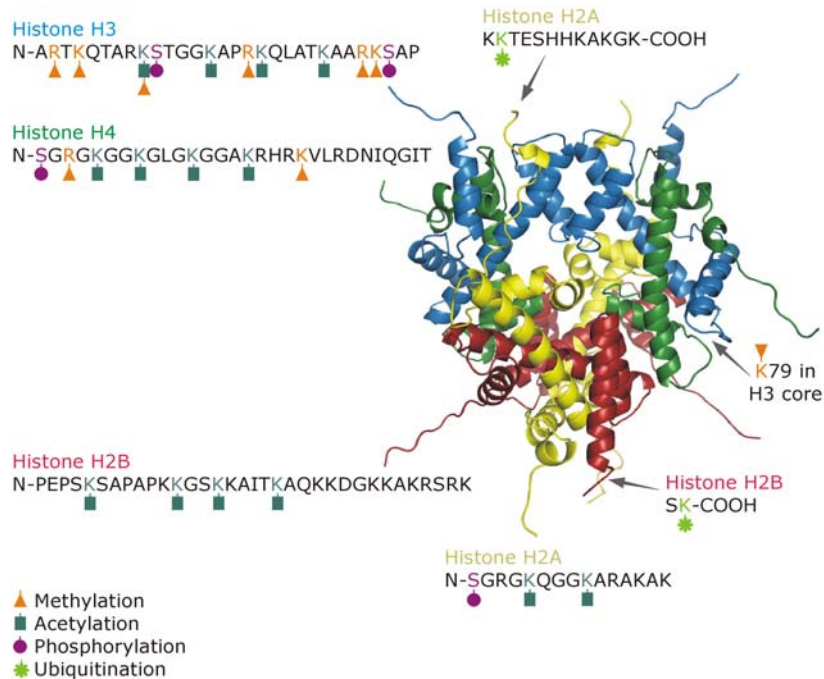
Figure 2. **Posttranslational modifications of histones.** The figure is created according Khorasanizadeh (2004). The 1AOI PDB entry (Luger *et al.*, 1997) without DNA served as model and the color code is adapted from Rhodes (1997).

However, the impact of certain modifications for specific nucleosome regulations has only emerged in the last decade or awaits discovery. Until recently, the methylation of histones appeared to be 'permanent' as nucleosomes were supposed to transmit epigenetic information from one cell generation to the next (reviewed by Bannister *et al.*, 2002). In the absence of evidence for histone lysine demethylases, histone methylation has demonstrated the power of modifications over DNA-based functions, regulating fundamental processes such as gene transcription, DNA repair, and epigenetic inheritance (reviewed by Martin and Zhang, 2005; Huyen *et al.*, 2004; Trojer and Reinberg, 2006; Vire *et al.*, 2006). The recent identification of histone lysine demethylases raises now the question if the reversibility of histone lysine methyl marks jeopardizes their epigenetic status (Trojer and Reinberg, 2006). LSD1 (lysine-specific demethylase 1) displays stringent substrate specificity and demethylates mono- or di-methylated lysine K4 within histone H3 using an amine oxidase reaction (Shi *et al.*, 2004). LSD1 shows extensive sequence homology to metabolic FAD-dependent amine oxidases, but

additionally includes a nuclear localization signal (NLS) and a SWIRM (Swi3p, Rsc8p and Moira) domain that is often found in chromatin-associated proteins. The mechanism is not compatible with any tri-methylated lysine substrate, as especially a protonated nitrogen and FAD (flavin adenine dinucleotide) as cofactor are required. Recently, a crystal structure of human LSD1 revealed a centrally located, spacious active site cavity, which, however, does not confer the ability to determine the degree of methylation of the histone tail. The study rather approves that the chemical mechanism forms the basis for the enzyme's selectivity for mono- and di-methylated substrates (Stavropoulos *et al.*, 2006). This limitation and the caveat that a LSD1 homolog appears to be absent in *Saccharomyces cerevisiae* proposed that additional demethylases with an alternative reaction mechanism to reverse lysine methylation exist. This assumption was confirmed by the discovery of JmjC (jumonji domain C) domain containing proteins that comprise a new subfamily of the large oxygenase family. JHDM1 (JmjC domain containing histone demethylase 1) is the first protein of this family that was shown to specifically act on di-methylated lysine K36 within histone H3 (Tsukada *et al.*, 2006). The *Saccharomyces cerevisiae* homolog scJHDM1 is a genuine H3K36me2 demethylase whereas recombinant Epe1, the *Schizosaccharomyces pombe* homologue, lacks histone demethylase activity (Tsukada *et al.*, 2006; Trewick *et al.*, 2005). Another approach showed that JMJD2A is a lysine tri-methyl-specific histone demethylase. This enzyme belongs to the JMJD2 subfamily of JmjC domain containing proteins and is evolutionarily conserved from *Caenorhabditis elegans* to human (Whetstine *et al.*, 2006). The demethylase signature motif, the JmjC domain, reverses methylation in a radical-based oxidative manner (hydroxylation). This mechanism is chemically compatible with demethylation of trimethylated substrates and uses Fe(II) and α-ketoglutarate as cofactors. JMJD2A activity on both H3K9me3 and H3K36me3 results in di-methylated, but no mono- or unmethylated products.

Despite the different mechanisms used to antagonize histone methylation, these results firmly establish that these marks are not static – rather as dynamic and versatile as the other histone modifications and that histone lysine demethylases offer a new spectrum how methylation influences chromatin, given the plethora of different methylation sites (reviewed by Mellor, 2006).

2.2   **Histone methylation in transcriptional control**

Efficient elongation on chromatin necessitates the displacement of the physical barrier imposed by the nucleosome on the transcribing RNAP II (reviewed by Workmann, 2006). Histone methylation occurs on arginine and lysine residues and is catalyzed by enzymes belonging to three distinct families of proteins – the PRMT (protein arginine methyltransferase) family, the SET-domain-containing [Su(var)3-9, Enhancer of Zeste, Trithorax] protein family, and the non-SET-domain proteins DOT1/DOT1L (disruptor of telomeric silencing-1). However, a methyl group is relatively small and its addition to lysine or arginine residues does not neutralize their charge, so it is unlikely that methylation alone will significantly affect chromatin structure. It is more likely that it serves as an anchorage point for the recruitment of regulatory proteins that contain specialized binding domains. Among the binding modules that have a high specificity for methylated histone residues are chromodomains, Tudor domains and WD40-repeats (Figure 3). For instance, the structure of the hybrid-tudor domains of JMJD2A bound to a tri-methylated H3K4 peptide was reported recently (Huang *et al.*, 2006). Moreover, chromatin immunoprecipitation and biochemical experiments show that the chromodomain of Eaf3 recruits the repressive histone deacetylase complex Rpd3C(S) to nucleosomes methylated by Set2 on H3K36. By generating a localized deacetylated template the nucleosomal organization is restored and aberrant transcription initiation is prevented (Carrozza *et al.*, 2005; Keogh *et al.*, 2005). Repressive proteins, such as HP1 (heterochromatin protein 1) or the Drosophila PC (Polycomb) protein, associate with methylated chromatin in a unique manner governed by their respective chromodomain, which is discriminatory for binding to the tri-methylated K9 and K27 within histone H3 (reviewed by Margueron *et al.*, 2005). By contrast, the CHD1 (chromodomain helicase DNA-binding protein 1) activator protein from human uses both chromodomains to recognize the activating tri-methylated K4 within histone H3 (Sims 3rd *et al.*, 2005; Flanagan, 2005), whereas the same function for *Saccharomyces cerevisiae* is contested, albeit supporting results are on hand (Pray-Grant *et al.*, 2005; Sims 3rd *et al.*, 2005). According to these examples, the ultimate function of the methyl epitope is a reflection of the type of protein it has evolved to recruit – either a repressor or an activator of transcription (Figure 3).
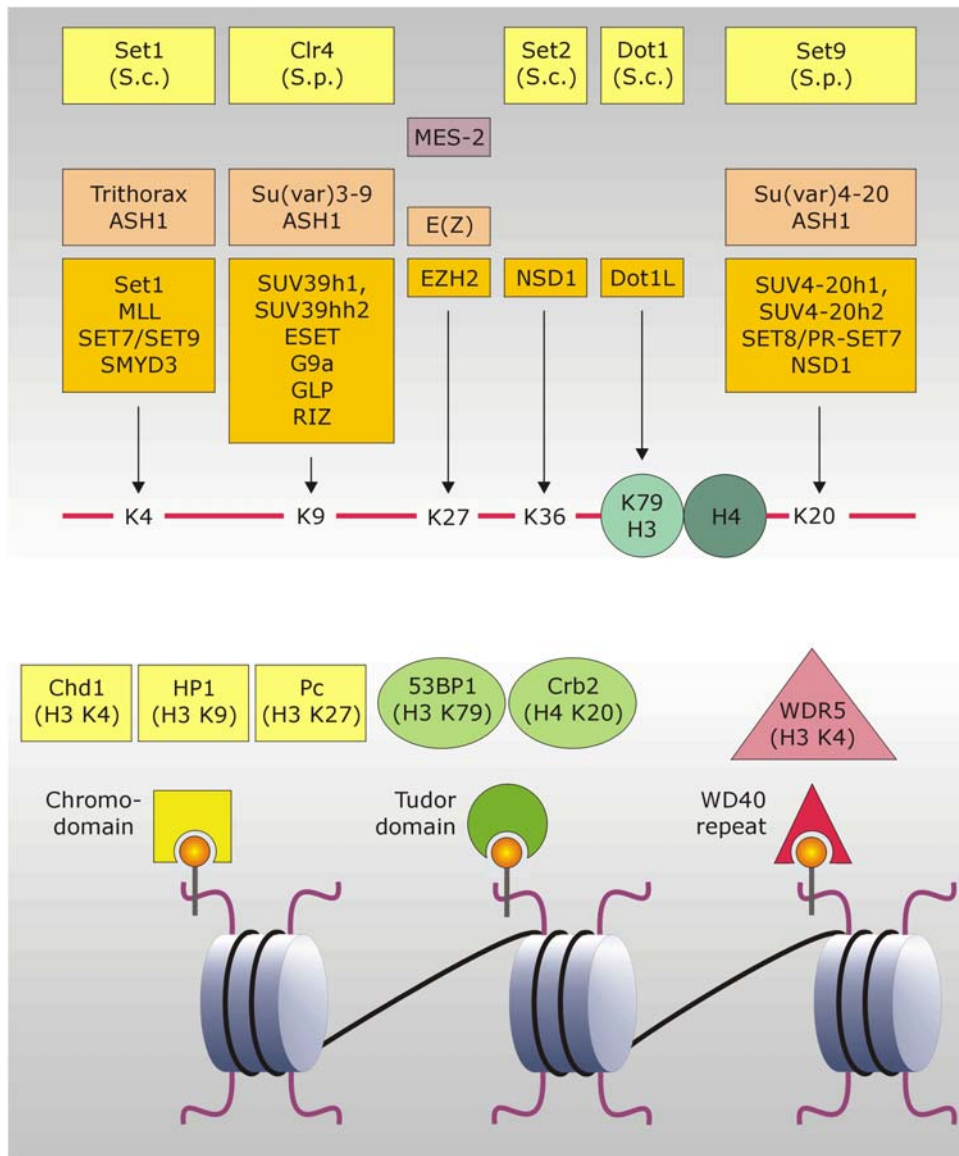
Figure 3. **Histone lysine methyltransferases, their target sites and methyl-lysine binding domains.** The figure is adapted from Martin and Zhang (2005).

11

Regarding lysine methylation, studies show that a number of lysines (K4, K9, K27, K36, and K79 of H3 and K20 of H4) are the major identified sites of methylation, although species-specific differences exist (reviewed by Sims 3rd *et al.*, 2003). Methylation marks that are linked to open chromatin and transcriptional activation (preferentially H3K4-H3K36 and H3K79) are present in all eukaryotes. On the other hand, H3K9, H3K27 and H4K20 methylation are hallmarks of a condensed chromatin state, albeit not found in *Saccharomyces cerevisiae* (reviewed by Sims 3rd *et al.*, 2003). HKMTs (Histone lysine methyltransferases) can add up to three methyl groups to a single lysine, thereby extending the indexing potential of this particular modification. Generally, in lower eukaryotes all three degrees of labeling of a particular histone site are regulated by the same enzyme, whereas in higher eukaryotes histone lysine methyltransferases were identified that specifically transfer only one methyl epitope. To dissect the possible functional reason of different degrees of methylation within chromatin context, the identification of diverse histone methyltransferases was followed in quick succession by a number of crystal structures (Xiao *et al.*, 2003; Zhang *et al.*, 2003; Xiao *et al.*, 2005). Unlike human Set1, Set7/9 acts as a H3K4 mono-methyltransferase. This intrinsic methylation product specificity is apparent from its crystal structure and the structural comparison with DIM-5, an H3K9me3 methyltransferase from *Neurospora crassa*. The lysine-access channel, a common feature of SET enzymes, connects the substrate- and cofactor-binding sites. In the process of subsequent methylation, the steric hindrance resulting from the increasing bulk of the lysine ε-amino group accounts for target and product specificity. In the case of Set7/9, it is speculated that the arrangement of a protein side chain may be the determinant for limited addition of methyl marks (Xiao *et al.*, 2003). By comparison, the cavity at the active site of DIM-5 seems spacious enough to accommodate even a methyl moiety on the substrate (Zhang *et al.*, 2003). Besides the given number of redundant HKMTs, further studies are needed to clarify even contradictory roles in transcription of H3K4 and H3K79. For example, H3K4 tri-methylation was found to be associated with the promoter and 5'-coding regions of exclusively active genes in yeast and higher eukaryotes, whereas H3K4 di-methylation appeared on active and inactive euchromatic genes in yeast (Santos-Rosa *et al.*, 2002). Nevertheless, it is envisaged that these modification states evoke specialized downstream responses.

2.3   Dynamic nucleosomes and gene transcription

Compacted chromatin must be rendered accessible for efficient RNAP II mediated gene expression. This process is amongst others, accomplished by a set of enzymes known as modifying complexes. Recent work gained insight into the impact of methylation profiles onto transcription processes, thus providing evidence for crosstalk between chromatin function and gene expression. The histone methyltransferases Set1 and Set2 (Figure 4 shows a summary of Set2), which modify the histone H3 lysines K4 and K36, respectively, are associated with RNAP II during elongation (reviewed by Gerber and Shilatifard, 2003; reviewed by Hampsey and Reinberg, 2003). Histone methylation apparently controls newly initiated RNAP II, and two phases of histone H3 methylation can be distinguished after transcription initiation (Morillon *et al.*, 2005). Set1 is part of a larger complex termed COMPASS (Complex of proteins associated with Set1) and associates with newly initiated RNAP II when Ser5 of the CTD is phosphorylated whereas this interaction is mediated by the Paf1 complex (Ng *et al.*, 2003). K4 and K79 methylation by Set1 and Dot1, respectively, are coregulated by a preexisting mark on a different histone tail. Ubiquitination of K123 within histone H2B by Rad6 and Bre1 is the necessary prerequisite (Sun and Allis, 2002; Ng *et al.*, 2002; Wood *et al.*, 2003). Latest results suggest that H2B ubiquitylation is in fact dispensable for monomethylation of the histone tail but necessary to stimulate subsequent rounds of methylation (Shahbazian *et al.*, 2005). By contrast, Set2 directly interacts with the phosphorylated CTD of RNAP II and is observed throughout the coding region of genes (Krogan *et al.*, 2003; Strahl *et al.*, 2002; Xiao *et al.*, 2003). Set2 recruitment to RNAP II relies on preceding phosphorylation of serine 2 of the CTD heptapeptide motif. Extensive genetic and interaction studies corroborate the functional link between H3K36 methylation by Set2 and the known serine 2 kinases CTDK-I and Bur1 (Krogan *et al.*, 2003; Xiao *et al.*, 2003; Li *et al.*, 2003; Chu *et al.*, 2006). It is speculated that the kinases might stimulate different degrees of methylation (Chu *et al.*, 2006). In addition, the *cis-trans* interconversion of proline 38 on histone H3 by the proline isomerase Fpr4 is supposed to interact with H3K36 methylation by Set2. This novel, noncovalent histone modification affects in an antagonistic manner the transduction of the signature motif (tri-methylation) and thus the transition from basal to active transcription (Nelson *et al.*, 2006).

The interaction between Set2 and the RNAP II CTD is mediated by a novel SRI (Set2 Rpb1-interacting) domain (Kizer *et al.*, 2005; Phatnani *et al.*, 2004). The SRI domain of *Saccharomyces cerevisiae* comprises the COOH-terminal residues 619 – 718 of Set2 (Kizer *et al.*, 2005). *In vitro*, the yeast Set2 SRI domain binds specifically and with high affinity to the CTD doubly phosphorylated at Ser2 and Ser5 (Kizer *et al.*, 2005). *In vivo*, deletion of the Set2 SRI domain abolishes H3K36 methylation and impairs transcription elongation, suggesting that the SRI domain is responsible for coupling transcription to histone methylation by Set2 (Kizer *et al.*, 2005).
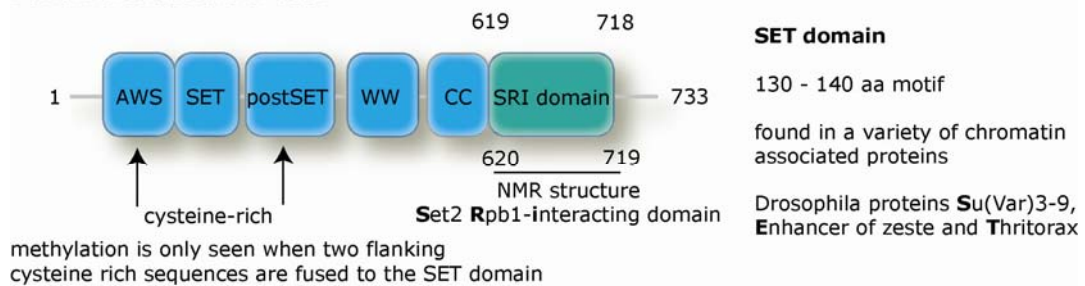
Figure 4. **Summary of the capacity of Set2.**

3 THE mRNA ASSEMBLY LINE

3.1 The CTD code – Heptad repeats as assembly platform

RNAP II shares many similarities at the sequence and structural levels with other multi-subunit RNA polymerases, but the carboxyl-terminal domain (CTD) of its largest subunit Rpb1 is an outstanding feature. It is a unique feature of RNAP II as it is not present in RNAP I or RNAP III or in bacterial or viral RNA polymerases (West and Corden, 1995 and reference therein) and this extraordinary CTD contributes through protein-protein interactions to the functional organization of the nucleus by mediating the association of differential transcription factors with RNAP II. The CTD contains a number of tandemly repeated heptapeptides (52 in mammals and 26 in yeast) with the consensus sequence Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7 (YSPTSPS). These repeats provide a landing pad for macromolecular assemblies involved in transcription or interdependent mRNA processing events (reviewed by Hirose and Manley, 2000; Fong and Bentley, 2001). It is speculated that the CTD interacts dynamically with transcription factors at the appropriate time, respectively, rather than carrying all these components throughout the transcription cycle. The stage of transcription and thereby CTD recognition by specific processing factors is dependent on the phosphorylation pattern of the repeats, which varies during the transcription cycle. Although there are five potential phoshorylation sites in a heptad repeat (Tyr1, Ser2, Thr4, Ser5, and Ser7), mainly the CTD positions Ser2 and Ser5 (Zhang and Corden, 1991) undergo waves of phorsphorylation and dephosphorylation (Dahmus, 1996 and reference therein; Komarnitsky *et al.*, 2000). By virtue of the concerted action of CTD kinases and phosphatases the level and pattern of phosphorylation is determined (reviewed by Meinhart *et al.*, 2005). Both modifications within the recurring motif are independently essential for viability (West and Corden, 1995; Yuryev and Corden, 1996) and play discrete roles in transcription. ChIP data revealed that Ser5 becomes phosphorylated at promoter proximal regions during transcription initiation/early elongation (Komarnitsky *et al.*, 2000) and recruits capping enzymes (McCracken *et al.*, 1997b; Cho *et al.*, 1997; Schroeder *et al.*, 2000). Phosphorylation of Ser2 increases towards the 3'-end of the gene, where the level of Ser5 phosphorylation decreases (Schroeder *et al.*, 2000; Cho *et al.*, 2001), thereby recruiting the mRNA processing, polyadenylation and termination factors to

elongating RNAP II (McCracken *et al.*, 1997a; reviewed by Proudfoot *et al.*, 2002; Ahn *et al.*, 2004). Additionally, the wide diversity of binding sites can be augmented when the *cis-trans* interconversion of peptide bonds N-terminal of prolines is considered. Pin1, a peptidyl-prolyl isomerase, and its homolog in *Saccharomyces cerevisia*e Ess1p specifically act at prolines that are preceded by phosphoserine and phosphothreonine (Hani *et al.*, 1999; Lu *et al.*, 1999). Both are implicated in the regulation of the CTD via their WW domain (Morris *et al.*, 1999; Xu *et al.*, 2003). This series of different phosphorylation patterns and conformation changes generates a 'CTD code' that determines fine-tuned configurations specific for binding of particular factors in the transcription cycle [Figure 5; (Buratowski, 2003)].
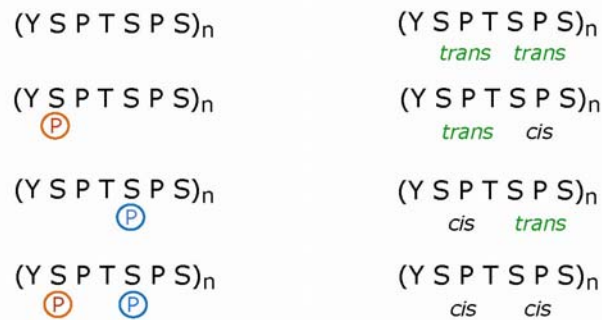
Figure 5. **The CTD code – 16 possible configurations.** The figure is adapted from Buratowski (2003). Possible phosphorylation sites are denoted by circled P (phosphor serine 2 red; phosphor serine 5 blue). Two prolines can adopt either the *cis* or *trans* configuration.

3.2   Induced fit – CTD recognition pattern

The apparent simplicity of the RNAP II CTD heptad repeats is deceptive as the structure of the CTD has been proven to be difficult to establish. The CTD appears as a tail-like extension protruding from the catalytic core of RNAP II, and is flexibly connected by a linker to a region near the RNA exit channel of the enzyme. Because of their largely disordered nature, neither the heptad repeats nor the linker are visible in the crystal structures of yeast RNAP II (Cramer *et al.*, 2001; Armache

et al., 2003). Solution studies of unbound CTD peptides indicate that the free CTD is structurally plastic, although it shows some residual structure and a propensity to form β-turns (reviewed by Meinhart *et al.*, 2005). In essence, an 'induced fit' mechanism is proposed, allowing the CTD to interact with multiple structurally dissimilar partners. Significant advancements in understanding CTD information content have been made by three crystal structures of CTD binding domains in conjunction with synthetic peptides. The structure of a doubly phosphorylated peptide, representing a single heptad repeat with phosphate moieties on Ser2 and Ser5 that is bound to the WW domain of Pin1 has been solved. This analysis revealed that the CTD peptide binds as an extended coil, projecting every third residue onto a unique face of the coil and with both phosphoserine-proline peptide bonds strictly in the *trans* configuration. Additionally, the phosphorylated residues form hydrogen bonds with Pin1 (Verdecia *et al.*, 2000). Another approach demonstrated how four heptad repeats adopt a different conformation when bound to the guanylyltransferase Cgt1. In this case all Ser-5 where phosphorylated (Ser5-P) and although only 17 amino acids were visible in the crystal structure, it was clearly shown how the CTD molds itself to this binding partner. The CTD repeats are bound to an extended surface of Cgt1, anchored at both ends by electrostatic interactions with Ser5-P and with extensive van der Waals contacts between Cgt1 and CTD residues (Fabrega *et al.*, 2003). The fact that RNAP I and RNAP III transcripts are not capped has been attributed to the lack of the CTD in those polymerases (Neugebauer and Roth; 1997; McCracken *et al.*, 1997b). In the case of RNAP II, CTD phosphorylation is required for recruitment of capping enzymes to sites of transcription *in vivo* (Komarnitsky *et al.*, 2000; Schroeder *et al.*, 2000) and the Cgt1-CTD peptide complex reveals some of the interactions necessary for selective and efficient mRNA biogenesis. A third study showed that proteins are capable of recognizing CTD phosphorylation patterns indirectly. Phosphorylation at serine 2 is a hallmark for transcription elongation and triggers the recruitment of factors involved in subsequent processing of the 3'-end of mRNA (Ahn *et al.*, 2004; Bird *et al.*, 2004; Meinhart and Cramer, 2004). Several RNA-processing factors interact with RNAP II through a conserved CID domain (CTD-interacting domain) (Yuryev *et al.*, 1996; Barilla *et al.*, 2000). The structure of CID of yeast Pcf11, an essential protein involved in pre-mRNA 3'-end processing and transcription termination (Amrani *et al.*, 1997), bound to Ser2-phosphorylated CTD peptide was

solved by X-ray crystallography. The central CTD motif Ser2-Pro3-Thr4-Ser5 forms a β-turn, whereas the flanking residues are in an extended conformation. The Ser2 phospho-group points away from the CID surface and may contribute to the β-turn with an additional hydrogen bond. This stabilization seems to be important for the indirect recognition of the phosphorylated peptide (Meinhart and Cramer, 2004). The instructive findings of theses studies are that the phosphorylation array encodes information about the state of the transcriptional apparatus that can be conveyed to factors specialized in recognizing entirely different positional cues in the CTD primary and secondary structure.

In addition, BRCT domains and FF domains are also known to interact via the C-terminus of Rpb1 with RNAP II. Until now there is no structural information of these domains in combination with the CTD.

3.3   FACT and Spt elongation factors

3.3.1   FACT

Intensive work has demonstrated that DNA accessibility is highly restrictive when assembled into chromatin and that eukaryotes have evolved elaborate mechanisms to both utilize and overcome this barrier to regulate and facilitate gene transcription. *In vivo*, RNAP II elongates through nucleosomes at a rate of 25 nucleotides per second (Izban and Luse, 1992). *In vitro,* a similar transcriptional competence can only be achieved on naked DNA templates. This discrepancy pinpointed to the assumption that under physiological conditions cellular factors are assuring efficient transcription within the repressive context of nucleosomes. A biochemical complementation assay was established and succeeded in the identification of an evolutionarily conserved heterodimer that stimulates the progression of RNAP II through nucleosomal templates. This activity from HeLa nuclear extracts was named FACT (Facilitates Chromatin Transcription) (Orphanides *et al.*, 1998). Further biochemical studies have shown that FACT specifically removes one H2A/H2B dimer from the nucleosome, thus disrupting its structure to allow RNAP II passage (Orphanides *et al.*, 1998; Belotserkovskaya *et al.*, 2003). In addition, FACT was found to possess intrinsic histone chaperone activity (Belotserkovskaya *et al.*, 2003). Additionally, Chd1, a chromodomain-containing ATPase, physically associates with FACT (Krogan *et al.*, 2002). Collectively, these

features of FACT maintain chromatin integrity in the wake of transcribing RNAP II, as disrupted chromatin structure becomes re-established during transcription.

Besides its role in chromatin modulation, a capacity of FACT in transcript elongation was simultaneously established. Spt16 genetically interacts with the known elongation factors TFIIS, Spt4/5, Spt6, and the PAF complex (Orphanides *et al.*, 1999; Squazzo *et al.*, 2002; Lindstrom *et al.*, 2003). FACT functions in conjunction with P-TEFb to ameliorate DSIF/NELF-mediated inhibition of transcriptional elongation on naked DNA templates (Wada *et al.*, 2000). FACT travels with elongating RNAP II at transcriptionally active genes *in vivo* and prevents transcriptional initiation from cryptic promoters (Mason and Struhl, 2003). A similar observation is made in Drosophila, where FACT and the RNAP II elongation complex colocalize along *hsp70* (Saunders *et al.*, 2003). FACT is intimately involved in the maintenance of chromatin structure and plays a prominent role in transcription elongation (reviewed by Reinberg and Sims 3rd, 2006).

In human cells, hSPT16 and SSRP1 comprise this chromatin specific complex (Orphanides *et al.*, 1999). The yeast counterparts, Spt16/Cdc68 and Pob3, were previously characterized as members of a CP complex that interacts specifically with DNA polymerase alpha (Wittmeyer *et al.*, 1997; Brewster *et al.*, 1998). However, the complex does not possess nucleosome binding capability as Pob3 lacks a DNA-binding motif (Wittmeyer *et al.*, 1997). This HMG (high-mobility group) box domain is found in the C-terminal part of SSRP1. The chromatin binding activity is provided by another protein named Nhp6. Therefore, Nhp6 and Pob3, structurally analogous to different domains of the vertebrate SSRP1 protein, function as a bipartite yeast analog of SSRP1 (Brewster *et al.*, 2001; Formosa *et al.*, 2001). The gene encoding the larger subunit of FACT, Spt16 belongs to the histone group of *SPT* genes.

3.3.2 *SPT* genes

SPT (Suppressor of Ty) genes were identified in *Saccaromyces cerevisiae* by selection for genetic suppressors of certain promoter insertion mutations (Winston *et al.*, 1984). Under these conditions, transcription initiates from the inserted promoter, while the adjacent proper gene promoter is inhibited. This elegant genetic screen uncovered transcription factor mutations that offset the detrimental effects of inserting a foreign piece of DNA (a TY transposon element or its long-

terminal-repeat δ) in the promoter of a reporter gene. Hence, all *SPT* gene mutations examined, showed reduced transcription initiation from Ty or δ, and conversely, normal transcription of the adjacent gene was restored. These results led to the view that they define factors with fundamental roles in transcription. As mentioned above, *SPT16* belongs to a certain class of *SPT* genes, the histone class, which comprises additionally *SPT4*, *SPT5* and *SPT6* as well as *SPT11* and *SPT12*, which encode histones H2A and H2B. Aside from the functions discussed, recent data from various groups suggest that Spt proteins may also function in other processes. For example Spt4 might play a role in transcription-coupled DNA repair (Jansen *et al.*, 2000). However, I will focus on their influence upon transcription elongation.

3.3.3   Spt6

A confluence of biochemical and genetic approaches has identified Spt6, a conserved protein implicated in both transcription elongation and chromatin structure, to help RNAP II to transcribe through chromatin. Spt6 is essential and it was shown that it has an ATP-independent histone chaperone activity to promote nucleosome assembly by binding to histones H3 and H4 *in vitro* and *in vivo* (Bortvin and Winston, 1996). Recent work characterized that a *spt6* mutation impairs the integrity of chromatin in active genes and permits aberrant transcription initiation from within the coding region of the constitutively active *FLO8* gene (Kaplan *et al.*, 2003). Additionally, Spt6 appears to function as a factor that mediates nucleosome reassembly onto the *PHO5*, *PHO8*, *ADH2*, *ADY2* and *SUC2* promoters and that this unique function in rebuilding nucleosomes at promoters is essential for transcriptional repression (Adkins and Tyler, 2006). These recent results implicate Spt6 as a maintenance factor for chromatin structure in the wake of RNAP II transcript elongation and reveal some of its mechanistic properties.

Another established supposition is that Spt6 acts as a modulator of RNAP II activity during elongation. Spt6 colocalizes with elongating RNAP II and Spt5 on actively transcribed genes in yeast (Krogan *et al.*, 2002) and human (Endoh *et al.*, 2004), as well as it is recruited throughout the transcription unit of heat shock genes in flies (Andrulis *et al.*, 2000; Kaplan *et al.*, 2000), albeit the physical interaction between Spt6 and Spt5 is weak. Classical elongation assays on naked DNA templates revealed that Spt6 functions as an elongation factor that enhances the

rate of transcription elongation of RNAP II both autonomously and together with DSIF (human Spt4-Spt5) in HeLa cells (Endoh *et al.*, 2004). Consistent with the idea that Spt6 may be required for transcription elongation, it genetically interacts with TFIIS and shares one salient mutant phenotype. Mutants of both proteins are sensitive to the drug 6-azauracil (6-AU), which is a common indication for factors involved in this highly regulated multi-enzymatic process (Hartzog *et al.*, 1998).

3.3.4 Spt4/Spt5

Several lines of evidence suggest that the Spt4-Spt5 complex exhibits several attributes implicated in chromatin modulation and transcript elongation (reviewed by Lindstrom *et al.*, 2002 and Sims 3rd *et al.*, 2004). First, DSIF [DRB (5,6-dichloro-1-β-D-ribofuranosylbenzimidazole)-sensitivity-inducing factor] was isolated from HeLa cell nuclear extracts based on its ability to confer DRB sensitivity to a reconstituted *in vitro* transcription system. This evolutionarily conserved heterodimeric complex is composed of p160 (hSpt5) and p14 (hSpt4), which are the human homologs of *Saccharomyces cerevisiae* Spt5 and Spt4 (Wada *et al.*, 1998a; Hartzog *et al.*, 1998). Spt5/DSIF p160 interacts preferentially with RNAP IIa via repeated domains in Spt5, termed KOW-motifs (Figure 6) that are homologous to the *Escherichia coli* transcription elongation factor NusG (Wada *et al.*, 1998a; Wada *et al.*, 1998b; Hartzog *et al.*, 1998; Yamaguchi *et al.*, 1999). Actually, additional experiments led to the conclusion that the phosphorylation pattern of the CTD determines RNAP II-DSIF interaction and that CTD phosphorylation releases p160 from RNAP II (Wada, 1998b; Lindstrom and Hartzog, 2001). Spt5 is essential for cell viability, whereas Spt4 is not. Genetic studies in yeast, *in vitro* transcription assays and ChIP experiments implicate Spt4 as a positive elongation factor with the capacity to antagonize the negative effects of RNAP II pausing imposed by the chromatin-remodeling yeast factor Isw1p (Morillon *et al.*, 2003; Rondon *et al.*, 2004). Moreover, a bioinformatical approach identified Spt4 as the orthologue of archaeal (*Methanococcus jannaschii*) DNA-directed RNA polymerase subunit E" (Ponting 2002). The same study gauged the role of the N-terminal protein segment of Spt5, thereby revealing an additional significant similarity with NusG. Besides the known KOW motifs, Spt5 shares a novel NGN (NusG N-terminal) domain with NusG homologues in archaea, bacteria and eukarya. Both the NGN-domain and a protein segment comprising more than the first KOW-motif are mandatory for Spt4 binding

(Ivanov *et al.*, 2000). A first insight in these evolutionary conserved domains was accomplished by the crystal structure of NusG from *Aquifex aelicus*. In addition, a model for the NGN- and KOW-domain of *Escherichia coli* NusG was the outcome of this study (Steiner *et al.*, 2002). Another bioinformatic elaboration predicts that Spt5 adopts an SH3 domain-like fold and assumes that the binding to RNAP II occurs via the OB-fold of its subunit Rpb7 (Aloy *et al.*, 2004). Spt5 plays in conjunction with Tat a distinct role in HIV-1 transcription by promoting the stability of transcription complexes at terminator sequences and minimizing the amount of polymerase pausing at arrest sites (Bourgeois *et al.*, 2002). Spt4 and Spt5 display an extensive set of genetic and physical interactions with TFIIF, TFIIS, and Rad26 as well as factors affecting chromatin structure like Spt6, FACT, Chd1, and the Paf1 complex (Orphanides *et al.*, 1999; Costa and Arndt, 2000; Jansen *et al.*, 2000; Krogan *et al.*, 2002; Squazzo *et al.*, 2002; Lindstrom *et al.*, 2003; Simic *et al.*, 2003; Endoh *et al.*, 2004). Extensive studies involving *Drosophila melanogaster* depicted the distribution of Spt5 and Spt6 on polytene chromosomes. The fruit fly homologs colocalize with the hyperphosphorylated, elongating form of RNAP II at sites of active transcription (Andrulis *et al.*, 2000, Kaplan *et al.*, 2000). Spt5 may experience methylation at arginine residues by PRMT1 and PRMT5 *in vitro*. It is proposed that methylation of Spt5 together with P-TEFb-mediated phosphorylation of Spt5 and the CTD of RNAP II generates modifications which function in modulating their transcriptional elongation properties (Kwak *et al.*, 2003). In addition, Spt5 interacts with factors associated with mRNA maturation and surveillance (Wen and Shatkin, 1999; Andrulis *et al.*, 2002; Pei and Shuman 2002; Lindstrom *et al.*, 2003). Latest results implicate Spt4-Spt5 in transcription elongation by RNAP I and rRNA processing (Schneider *et al.*, 2006).
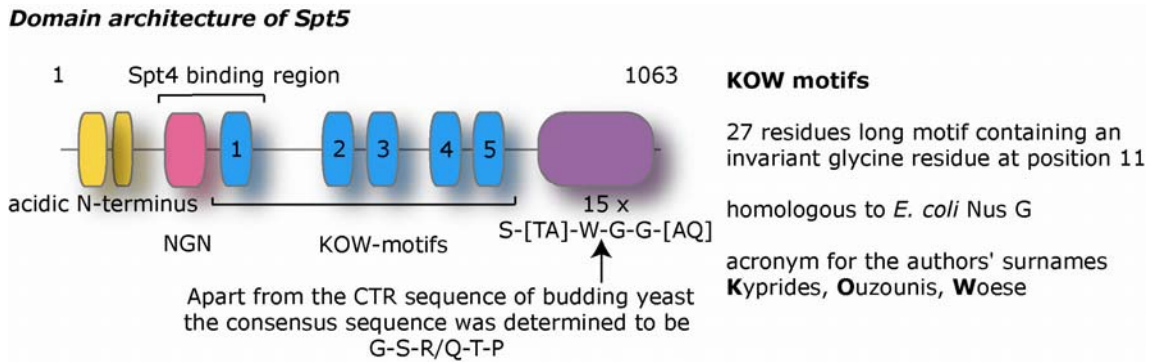
Figure 6. **Domain architecture of Spt5.**

Conclusively, increasing lines of evidence showed that Spt4-Spt5 function depends on the CTD and CTD modifying enzymes like diverse CTD kinases and Fcp1, the major phosphatase (Lindstrom and Hartzog, 2001). In addition, it appeared that the partially purified factors used in the DRB sensitivity assay to identify DSIF contained an additional factor, termed negative elongation factor (NELF), which acts in cooperation with DSIF to inhibit elongation (Yamaguchi *et al.*, 1999b). Notwithstanding the fact that NELF is still not identified in yeast and that the two kinase complexes Bur1/Bur2 and CTDK-I (Ctk1, 2 and 3) appear to functionally reconstitute the activity of the human cyclin-Cdk complex P-TEFb in yeast, a model evolved how DSIF acts as transcription elongation factor. In general, many eukaryotic elongation factors exert their role by either preventing or overcoming RNAP II transcriptional pausing. In the case of DSIF, the heterodimeric factor plays a pivotal role in the temporal coordination of capping and transcriptional elongation.

The model states that DSIF binds to RNAP II shortly after initiation or during formation of the transcription complex at the promoter (Wada *et al.*, 1998a; 1998b) and subsequently recruits NELF, trapping the transcription machinery at promoter proximal sites [Figure 7; (Yamaguchi *et al.*, 1999b)]. This pausing is elicited by NELF only in the presence of DSIF. Between initiation and arrest, the Cdk7 subunit of the general transcription factor TFIIH phosphorylates the CTD of RNAP II on serine 5 of its heptapeptide repeat. The paused RNAP II is then joined by the capping enzymes through stimulated interactions with the serine 5-modified CTD and DSIF (Pei and Shuman, 2002). DSIF/NELF mediated pausing allows a time frame for the capping enzyme recruitment and the faithful addition of a cap to the

5'-end of the nascent RNA ensures that RNAP II commits to productive elongation only after the transcript has been capped. Following the placement of the 7-methylguanosine cap on pre-mRNA, P-TEFb binds to RNAP II and its activation triggers phosphorylation of serine 2 and the CTR (C-terminal region) of DSIF. This abrogates the repressive action of NELF/DSIF and subsequent reactivates RNAP II to synthesize functional mRNA precursors. Thus, P-TEFb and DSIF act antagonistically (Wada *et al.*, 1998b). Latest results suggest that the threonines in the CTR of Spt5, a putative mini-CTD, are phosphorylated by P-TEFb (Yamada *et al.*, 2006). This event serves as a regulatory switch, converting the unphosphorylated 'repressive' Spt5 into a phosphorylated 'stimulating' elongation factor (Yamada *et al.*, 2006). DSIF's attributive opposing role in transcription elongation is thereby reflected (Wada *et al.*, 1998a). These regulatory interactions that pause early transcription appear to have evolved to allow the assembly of mRNA maturation factors on the RNAP II CTD under conditions in which the polymerase is engaged in a productive transcription complex. This would be comparable with the 'checkpoints' that operate during the cell cycle to ensure that each phase of the cycle is complete before the next begins (reviewed by Orphanides and Reinberg, 2002).
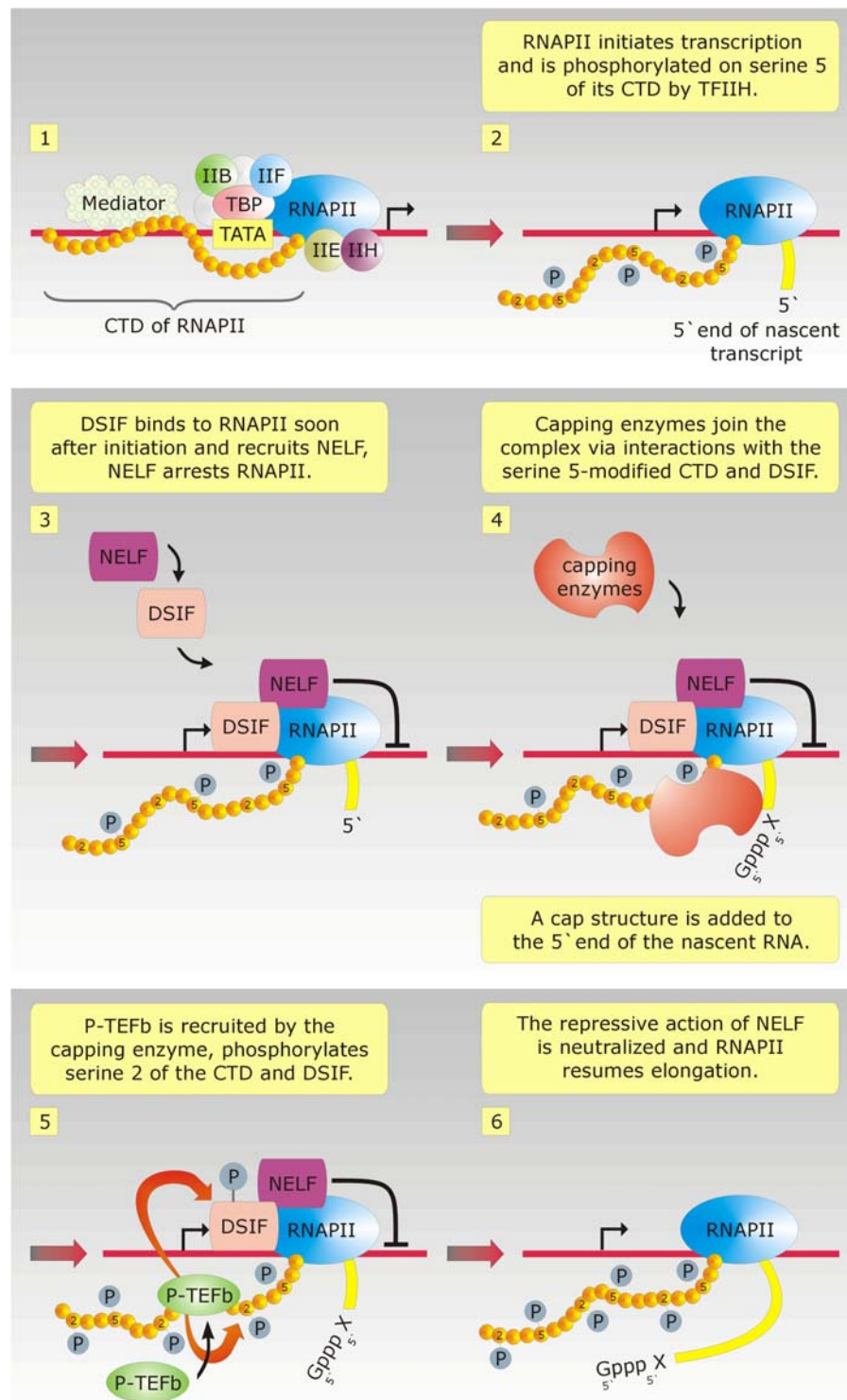
Figure 7. **The elongation checkpoint model.** The figure is adapted from Orphanides and Reinberg (2002).

4   THIS STUDY

With nearly 60 subunits and a mass in excess of 3 million Daltons, the RNAP II transcription machinery poses a formidable challenge for structural analysis. The successful structure determination of the polymerase was the major breakthrough because it revealed the platform upon which all macromolecular components are assembled. The full biological significance of the RNAP II structure lies in the implications for the higher complexes formed with a plethora of transcription factors (Boeger *et al.*, 2005). Against this backdrop, the central aim of this study was to elucidate the atomic structure of components of the transcription elongation machinery to highlight important missing details. A structural perspective on the system should allow us to move towards a molecular picture of this central nuclear process. A heterodimeric complex comprising Spt4 and Spt5 plays a pivotal role in transcription elongation. A part of this study is concerned with the interaction between this elongation factor and RNAP II. I describe here the design and synthesis of Spt4-Spt5 variants. Investigations of the binding to the RNAP II surface could be undertaken by constructing complexes from their component parts. The quest for structural analysis of this interaction has involved X-ray crystallographic methods. Recently, it emerged that transcription is also coupled to the alteration of chromatin structure and the biochemical well-documented histone lysine methyltransferase Set2 has the potential to influence both processes. I tried to capture specific molecular details of the interaction between RNAP II and the SRI (Set2 Rpb1-interacting) domain of Set2. A move in this direction has been attempted by solving the structure of this domain by NMR spectroscopy. Three α-helices are arranged in a left-handed bundle and adopt a new fold which is capable to bind to RNAP II. In order to illuminate the understanding of the interplay among the CTD (C-terminal domain) and the SRI domain, my study involved both attempts of co-crystallizing CTD peptides with SRI variants and NMR titration experiments.

Chapter II: Results and Discussion

5 RECOMBINANT SPT4-SPT5 PROTEINS AND ASSEMBLY OF THE ELONGATION CHECKPOINT COMPLEX

The successful addition of a cap to the 5'-end of the nascent RNA shortly after transcription initiation is reflected by the fact that RNAP II commits to productive elongation. To foster an expanded view of the cellular function of the heterodimeric elongation factor Spt4-Spt5, several endeavors have been made to predict and model the structure of interacting proteins, which combined create the so called 'checkpoint elongation complex' (reviewed by Orphanides and Reinberg, 2002). The resulting structure should identify the interacting surfaces and nature of the protein-protein interaction.

Before I started my project, great strides have been made to obtain endogenous RNAP II of high quality as the purification protocol of ten-subunit RNAP II had already proofed itself. The material used in this study was obtained as explicitly described elsewhere (see PhD-thesis Armache, 2005). In addition, I could adopt the experimental setup of synthetic DNA/RNA scaffolds used by H. Kettenberger in the laboratory (see PhD-thesis Kettenberger, 2005). During this thesis, the establishment of the 200 litres large scale fermentation to meet the growing demand for RNAP II, particularly with regard to illuminate the role of Spt4-Spt5 as transcription elongation factor, posed a challenge.

I describe here the design and production of heterologous Spt4-Spt5 protein variants, and investigations of their binding to ten- or 12-subunit RNAP II. To unravel these interactions I availed myself of protein crystallography. In Table 1 the crystallization conditions used for this approach are summarized. They are slightly adjusted to the demands of the 'checkpoint elongation complex' and therefore differ from previously used reservoir solutions (see PhD-thesis Kettenberger, 2005). Finally, the cryo-protectant solutions contained the same components as the reservoir solutions plus 22 % of glycerol. Screening with increasing amounts of

cryo-protectants like MPD, PEG-400, PEG-600 or ethylenglycol could not rectify the established freezing procedure.

Table 1: Summary of RNAP II crystallization conditions (reservoir solution)

Citrate/PEG	11 – 14 % PEG-6000; 170 mM Tri-Sodium Citrate; 100 mM Hepes pH 7.5; 5 mM DTT.
Natrix #38 modified	4 – 5.5 % PEG-6000; 200 mM Ammonium Acetate; 150 mM Magnesium acetate; 50 mM Hepes pH 7.5; 5 mM DTT.
(NH$_4$)NaTartrate	750 – 825 mM Ammonium Sodium Tartrate; 50 mM Hepes pH 7.5; 5 mM DTT.
(NH$_4$)NaTartrate/ PEG/KSCN	12 – 14 % PEG-6000; 300 mM Ammonium Sodium Tartrate; 100 mM Hepes pH 7.5; 100 mM Potassium Thiocyanate; 5 mM DTT.
MES	18 – 22 % Ammonium Sulfate; 50 mM MES pH 5.7; 100 µm Zinc Chloride; 5 mM DTT.
PEG/PO$_4$	12 – 18 % PEG-6000; 390 Ammonium Sodium Phosphate pH 6.0; 50 mM Dioxane; 5 mM DTT.

5.1 Spt4-Spt5 bicistronic expression

The production of constructs of Spt4 and Spt5 when expressed both separately and in combination, should allow analysis of their interaction with RNAP II. The rational behind the finally selected bicistronic expression strategy was to study a Spt5 variant comprising the homologous NGN- and KOW1-domain of which the *Aquifex aelicus* crystal structure (Steiner *et al.*, 2002) was identified and four additional KOW domains. Production of short Spt5 variants was feasible, but concomitantly accomplished experiments gave the hint that extended Spt5 variants could not be expressed. Only moderate expression levels were obtained. In addition, degradation products were already visible during the first purification step with Ni-NTA affinity column. Insufficient results were obtained as soon as the Spt5 variants comprised the whole Spt4 binding region. Expression of these recombinant proteins in *Escherichia coli* seemed to be impaired, apparently due to a loss of structural integrity of Spt5 when expressed without Spt4. As this approach was unsatisfactory, the full length binding partner Spt4 was expected to stabilize the heterodimeric complex. Bicistronic expression offers the advantage of both placing two protein-coding sequences under the control of the same transcriptional regulation and facilitating co-translational folding of protein pairs.

5.1.1 Purification of Spt4-Spt5 variant3

Initially, a bicistronic expression vector was constructed containing an N-terminal GST-tag linked to Spt4 and a six-histidine-tag (His$_6$-tag) introduced at the C-terminus of a designed Spt5 variant. This variant3 encompasses all modular domains of wild type Spt5 disregarding the predicted unstructured segments found at the far end of the protein, exactly amino acid residues 283 to 849. Equivalent transcript levels of both proteins should be ensured by a purification protocol which makes successive use of both the GST- and the histidine affinity tag. But it turned out that the purification of the protein pair was not straightforward and for improved handling the GST-tag was removed from Spt4. The stoichiometry should be assured with downstream purification procedures. A two-step strategy was established which allowed the purification of a Spt4-Spt5 complex with an approximately 1:1 stoichiometry. The presence of Spt4 in the complex was verified by Edman-Sequencing. The last step was also performed by a Superose6 gel

filtration column. In doing so, the complex showed an aberrant retention volume, which could only be explained by inconvenient interactions with the column matrix. Before adjusting the size exclusion buffer to 100 mM ammonium sulphate, extensive screening of conditions was performed.

5.1.2 Assembly of RNAP II with Spt4-Spt5 variant3

The assembly reaction between Spt4-Spt5 and RNAP II was performed at 4 °C and the complex subsequently subjected to gel filtration. Finally, 35 µl of the peak fractions were mixed with sample buffer and the result of the size exclusion column monitored on a Coomassie stained SDS-polyacrylamid gel (Figure 8). Spt4 is not visible on the gel according to its small size of 11.2 kDa. But in consideration of the fact that independently expressed Spt5 variant3 is heavily degraded and that the Spt4-Spt5 variant3 complex was used for the experiment, I could show the interaction between the recombinant expressed proteins and RNAP II for the fist time by size exclusion chromatography.
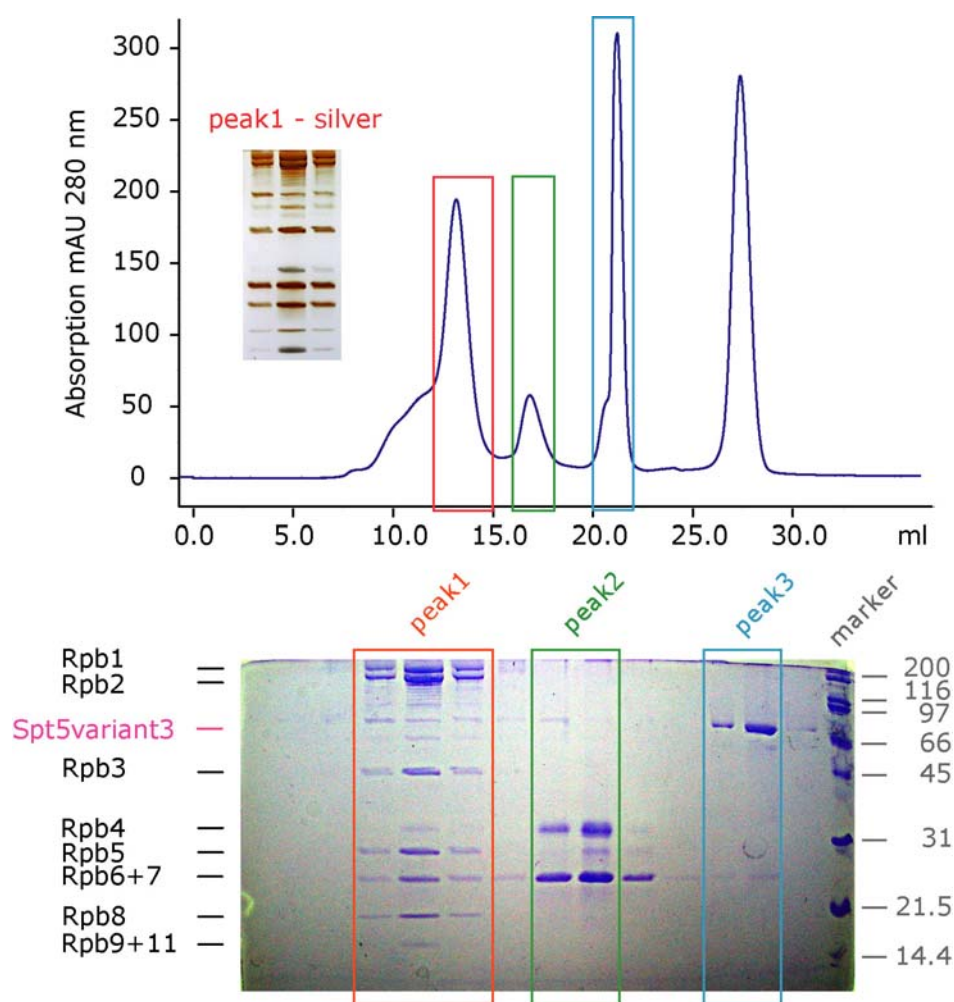
Figure 8. **Assembly of RNAP II with Spt4-Spt5.** The chromatogram depicts the gel filtration profile of the assembly. The elution volume of approximately 13 ml corresponds to RNAP II. The excess of recombinant proteins elutes at approximately 16 ml and 21 ml, respectively. The Coomassie stained SDS-PAGE analysis demonstrates the purity of the final sample as obtained from the gel filtration. In addition, the inset shows the silver-staining of peak1 to highlight Rpb4 which is faintly visible with Coomassie. Neither the RNAP II subunits Rpb10 and Rpb12 nor Spt4 is visible due to their small size.

Nevertheless, the production of sufficient quantities of the Spt4-Spt5 variant3 complex was the bottleneck of this approach. Albeit the expression level of Spt5 was improved by the bicistronic expression strategy, I had to cope with substoichiometric quantities upon assembly. In addition, the speculation of a prevailing propensity to improper folding is reflected by the aberrant retention volume of Spt5. Moreover, evidence that aggregation occurs and possibly impairs

complex formation, comes from the shoulder in the elution profile. The genuine RNAP II peak of Superose6 is of Gaussian shape. Coincident with performing the assembly, the stability of RNAP II alters.

5.2   Rpb7/Spt4 fusion protein

Tethering the components of multiprotein complexes creates new possibilities for exploring the nature and detail of protein–protein interactions, and can also be used to create novel functional combinations. The original impetus for tethering full-length Spt4 to the C-terminus of Rpb7 stems from the finding that in *Sulfolobus acidocaldarius* the DNA-directed RNA polymerase subunit E, which comprises the eukaryotic Rpb7-like N-terminal domain, is linked to another polypeptide E", which is on the other hand homologous to eukaryotic Spt4 (Ponting, 2002).

Here we envisaged that the conjunction of yeast Rpb7 to Spt4 alone, or in combination with a Spt5 variant, may recruit the Spt4-Spt5 complex to the RNAP II surface. The physical attachment of Spt4 to RNAP II should counteract potential transient interactions occurring during crystal growth. In addition, the creation of a fusion protein by sequence-informed tethering of Spt4 to Rbp7 was expected to facilitate the identification of Spt4-Spt5 on the RNAP II surface by defining the amino acid residues either important for interactions between DSIF or necessary for establishing the 'checkpoint complex'. Furthermore, the first attempt to create a complex consisting of RNAP II and Spt4-Spt5 indicated that the conditions of this experiment were still suboptimal. Therefore a new proceeding for creating a composite of both the RNAP II and an elongation factor was chosen.

5.2.1   Purification of an artificial Rpb7/Spt4 and Rpb4 complex

In order to make the fusion protein amenable to random interactions with RNAP II surface, a flexible, glycine rich linker was introduced between the two distinct proteins. According to the cloning procedure a linker containing 15 glycine residues was intended. But additionally, a linker comprising only 13 glycines was obtained and as no frameshift was introduced into the gene sequence this variant was also expressed and purified. The exact protocols for purification of the different linker variants of the Rpb7/Spt4 and Rpb4 complex are given in chapter 10.3.2. As

examined by Coomassie staining, the complex containing the 13 glycine linker showed great promise for the rate of yield and was therefore chosen for all ongoing experiments. After releasing the complex from the Ni-NTA affinity column, numerous different buffer conditions were tested for the following purification steps to achieve homogenous material for crystallization. The final protocol consists of a MonoS- and Superose6 column (Figure 9).
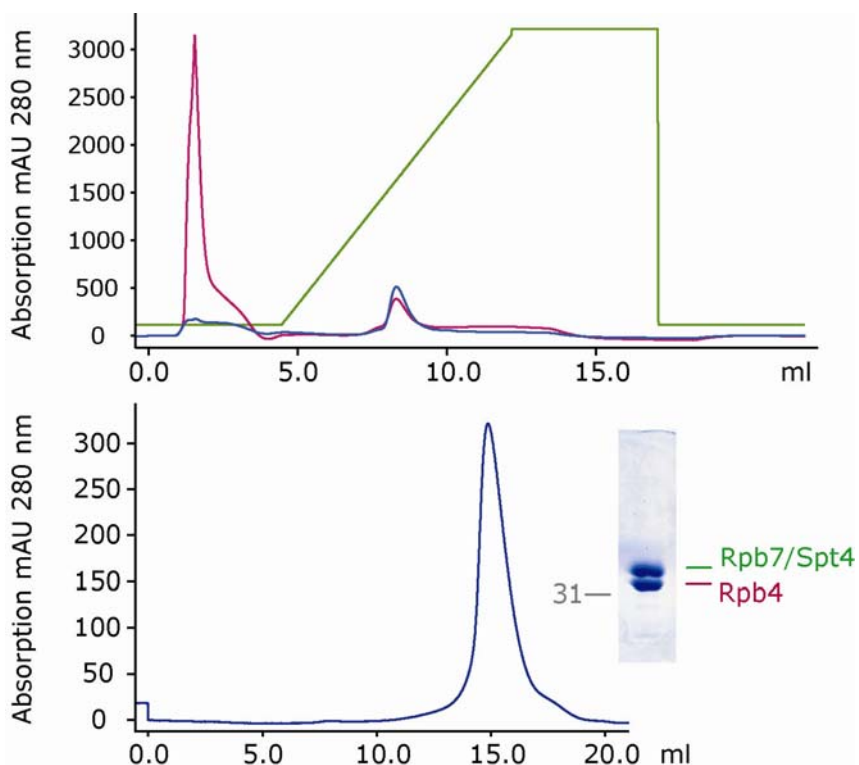
Figure 9. **Purification procedure of an artificial Rpb7/Spt4 and Rpb4 complex.** The chromatograms depict the typical purification procedure of the protein pair. The MonoS profile highlights the high DNA content in the sample. The polishing step comprises a gel filtration. The inset shows 3 µg of concentrated and highly homogenous Rpb7/Spt4-Rpb4 and the expected marker band at 31 kDa.

5.2.2 Assembly of complex12

Another possible benefit of the fusion protein lies in its potential for structure determination by molecular replacement. As the 12-subunit comprising RNAP II was already known (Armache *et al.*, 2003) it was assumed that the attachment of 122 additional amino acids would not influence the overall structure but rather provide a

template where Spt5 might join RNAP II. For this reason complex12 was assembled and the stability monitored by Superose6 gel filtration. The binding was demonstrated by a peak shift of the Rpb7/Spt4-Rpb4 proteins and a Coomassie-stained SDS-polyacrylamid gel (Figure 10). At the first sight, the behaviour of this complex12 was comparable to the genuine Rpb4-Rpb7 subunits of RNAP II (Armache *et al.*, 2003).
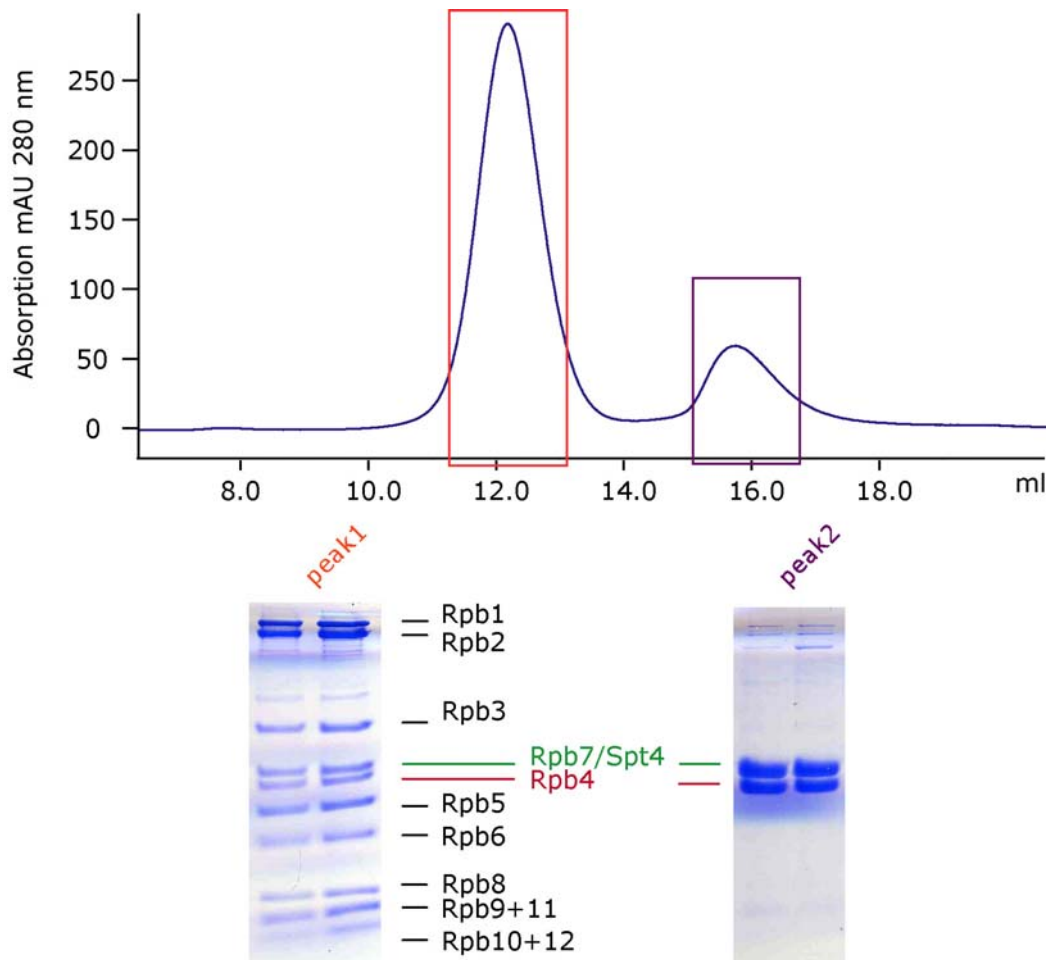
Figure 10. **Assembly of fusion protein Rpb7/Spt4 and Rpb4 with core RNAP II.** The chromatogram depicts a typical gel filtration profile. Like in the case of genuine RNAP II, the elution volume of approximately 12 ml corresponds to the reconstituted enzyme, and the excess of recombinant proteins elutes at approximately 16 ml. The Coomassie stained SDS-PAGE demonstrates the purity of the modified RNAP II. Both the chromatogram and the SDS-PAGE analysis are derived from two different assembly reactions. Nevertheless, they are typical representatives of this highly reproducible experiment.

5.2.3 Crystallization of complex12

Preparations of reconstituted RNAP II were homogeneous according to size exclusion chromatography. The collected peak fractions were concentrated and the complex subjected to different crystallization conditions (Table 1). These were successfully used in former times for crystallizing the genuine RNAP II by the hanging drop vapor diffusion method. A first indication for the influence of the hybrid protein onto the crystal formation was given by the new morphology of the crystals grown in (NH_4)NaTartrat/Hepes conditions (Figure 11) and the downstream freezing procedure. Resulting from the known, high solvent content of RNAP II crystals, the established stepwise freezing protocol was applied (see PhD-thesis Armache, 2005). According to this, glycerol was added up to 22 % to the mother solution and the crystals were allowed to cool down slowly over night to 8 °C before flash frozen in liquid nitrogen. In my case the crystals began to dissolve when event treated with low concentrations of glycerol and this morphology was not reproducible in this condition. In addition, crystals were obtained with other solutions (Table 1). Thus, synchrotron analysis of the crystals only revealed the known RNAP II structure. Despite the size of the crystals of about 0.17 x 0.17 x 0.1 mm, diffraction extended only to 4.0 Å in favourable cases. Albeit I could obtain crystals with a hexagonal morphology (0.17 x 0.17 x 0.07 mm) in the NH_4NaTartrate/PEG/KSCN condition, they did not crystallize in another point group than the genuine RNAP II.

(NH₄)NaTartrate   **NH₄NaTartrate/ PEG/KSCN**   **Na₃Citrate/PEG**
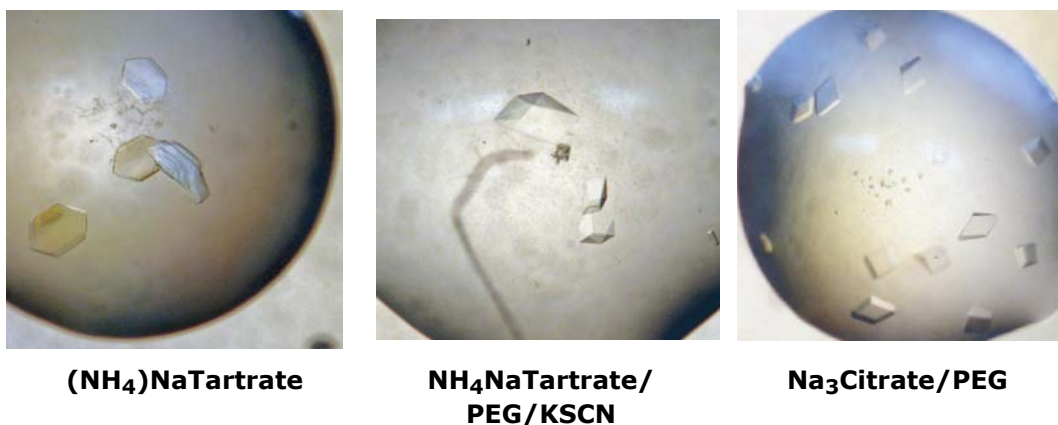
Figure 11. **Single crystals of complex12.** These typical representatives were obtained by the hanging drop vapor diffusion method. The well-established conditions are summarized in Table 1.

5.3 Rpb7/Spt4 and Rpb4 complex plus Spt5 variants

Two kinds of elaboration determined the amino acids which are necessary for Spt5 to interact with RNAP II. One study showed that the part of human Spt5 spanning amino acid residues 313 to 755 is capable to bind to RNAP II (Ivanov *et al.*, 2000). An earlier study even cut the binding region down to amino acid residues 313 to 420, a region which lays in between KOW domain 1 and KOW domain 2 (Yamaguchi *et al.*, 1999). In order to combine the structural insights gained from the *Aquifex aelicus* crystal structure (Steiner *et al.*, 2002) and expression studies with single Spt5 variants, new Spt4-Spt5 protein complexes were designed.

5.3.1 Purification of Rpb7/Spt4 and Rpb4 complex plus Spt5 variant1

On the basis of these results, and guided by sequence conservation and secondary structure analysis I designed three N- and C-terminal deletion constructs. The created Spt5 variants feature the region which is in accordance with the NGN-domain and additionally either cover three or five KOW domains. The shortest variant contains three and the longest all five domains and additional residues with apparent low sequence homology. Already the fist purification step via the C-terminal His$_6$-tag revealed that the fusion protein complex (Rpb7/Spt4 and Rpb4) coexpressed with Spt5 variant1 (amino acid 283 – 620) was to a considerable degree more encouraging than the other variants (Figure 12). The C-terminal extended Spt5 variants seemed to be either impaired in expression (variant2; amino acid 283 – 849) or prone to degradation (variant3; amino acid 283 – 874).
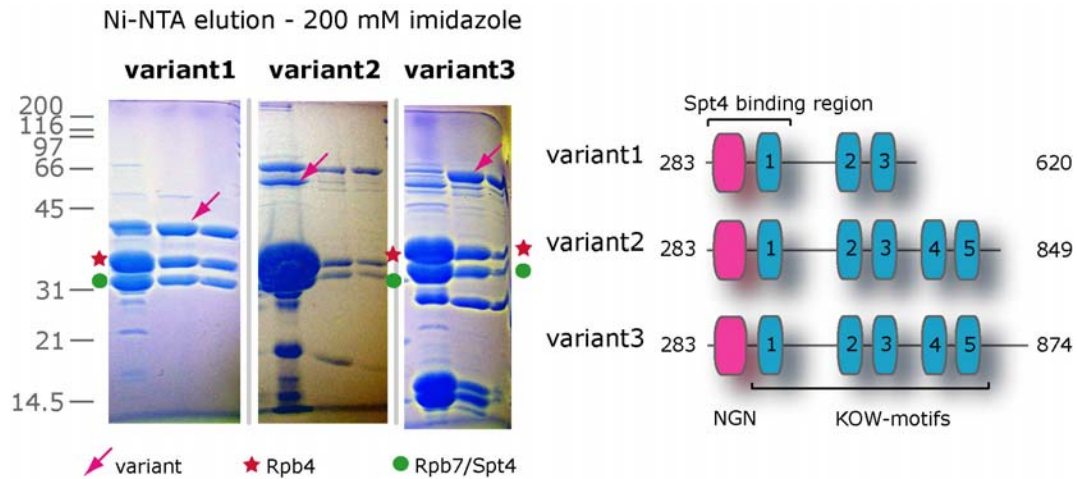
Figure 12. **Coexpression of the fusion protein complex with different Spt5 variants.** The Coomassie stained SDS-acrylamid gel displays the 200 mM Ni-NTA elution profiles of Spt5 variant1 to variant3, which were coexpressed with the fusion protein complex. In addition, the domain architecture of the variants is shown, respectively.

Extensive buffer screening was accomplished to maintain stoichiometry of this artificial complex. The purification protocol involved a MonoQ column, followed by a Superose6 column. This element was introduced as preliminary experiments with PEI (polyethylenimine) showed that DNA sticks to the desired protein and triggers precipitation (data not shown). In the first step the complex retained on the column and after elution two subsequent peaks were the outcome. At best, the peak fraction of the anion exchange column was considered which contained all three components in equal amounts (Figure 13). Already at this stage the purity of the protein sample was high. The latter step was the most crucial as the previously achieved stoichiometry was reversed by diverse buffer conditions. The Spt5 variant1 showed the propensity to either interact with the column material or to precipitate in improper buffer conditions (data not shown). Therefore I abandoned the last polishing step and continued the experiments with concentrated MonoQ samples, keeping in mind that the salt concentration was approximately 450 mM. Nevertheless, aliquots of ammonium sulphate pellets containing 510 µg of protein could be flash frozen in liquid nitrogen and stored at -80 °C for later use. One liter of a standard expression culture yielded a purified, recombinant protein sample of 7.5 mg/ml.
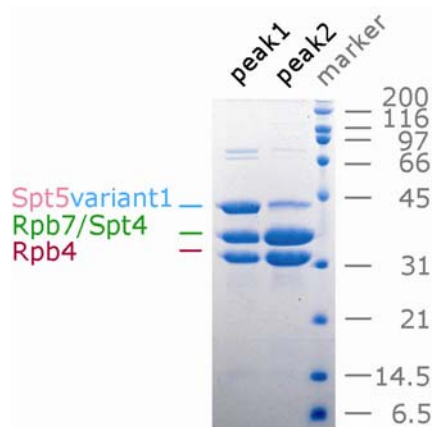
Figure 13. **Purification of fusion protein complex plus variant1.** The Coomassie SDS-PAGE analysis depicts a typical MonoQ elution of two successive peaks and demonstrates the purity/stoichiometry of the sample. The additional subunits are colored according to Figure 12 and chapter 10.3.1.

5.3.2   Assembly of complex13

The competence of Rpb7/Spt4 and Rpb4 plus Spt5 variant1 for complex formation with the core RNAP II was examined by a MembraSpin centrifugal concentrator. It is a common strategy to perform a binding assay by means of a concentrator instead of using a gel filtration column (personal communication I. Artsimovitch). On this account, the 100 kDa cut off was chosen to establish a stoichiometry complex and concomitantly get rid of protein excess. The components of the assembly were provided in their optimal buffer conditions and subsequently transferred into the crystallization buffer.

Essentially, the linkage of Spt4 to Rpb7 aimed at increasing the effective local concentration of Spt5 on the RNAP II surface. Moreover, by providing a variant containing a segment involved in the Spt5 – RNAP II interaction the establishment of the 'elongation checkpoint complex' should be tackled a lot easier. Both assembly sample and flow through were monitored by SDS-PAGE (data not shown). The molecular weight of Rpb3 (45 kDa) is similar to that of Spt5 variant1 (39.7 kDa) and both proteins run at the height near the 45 kDa marker band (Figure 14).
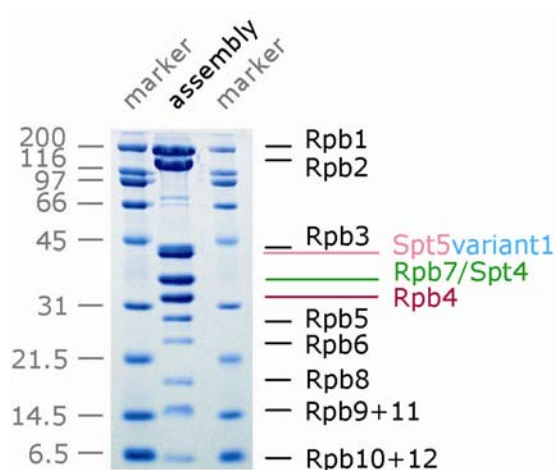
Figure 14. **Assembly of RNAP II with Rpb7/Spt4 and Rpb4 plus Spt5 variant1.** The Coomassie stained SDS-polyacrylamid gel displays the components and the quality of this artificial multiprotein complex. The constituents of the core RNAP II are shown in black. The additional subunits are colored according to Figure 12 and chapter 10.3.1.

5.3.3 Crystallization of complex13

The concentrated sample was used to obtain crystals by the hanging drop vapor diffusion method. Within days three dimensional crystals grew mainly in two conditions [Figure 15; ($NH_4NaTartrat/PEG/KSCN$ and $NH_4NaTartrat/Hepes$)]. They were subjected to synchrotron radiation and diffracted until 4.5 Å.

$NH_4NaTartrate/PEG/KSCN$

Figure 15. **Single crystals of complex13.** These typical representatives were preferentially obtained in the $NH_4NaTartrat/PEG/KSCN$ condition (Table 1).

5.3.4   Assembly of 'elongation checkpoint complex'

Much of the processing of eukaryotic pre-mRNA into mature mRNA occurs cotranscriptionally and the recruitment of the capping machinery to the transcription complex marks the beginning of this multi step procedure (reviewed by Shatkin and Manley, 2000). All RNAP II transcripts are marked at their 5' ends by the addition of a methylated guanosine cap, when nascent RNA is about 20-25 bases long. In this process Spt4-Spt5 plays a pivotal role. On the one hand, the manner of intimate recognition of this transcript by Spt4-Spt5 is still elusive, although the structural similarity between the NGN-domain and the RNP (Ribonucleoprotein) motif led to the speculation of a tuneable RNA binding site (Steiner *et al.*, 2002). On the other hand there are compelling evidences that this bipartite elongation factor brings RNAP II to a halt at promoter proximal sites, thereby ensuring a time window for successful capping (reviewed by Orphanides and Reinberg, 2002). Consistent with the idea that Spt5 latches onto RNAP II in the presence of RNA, I constructed a DNA/RNA hybrid based on the solved structure of the nucleic acids in the elongation complex (Kettenberger *et al.*, 2004). Additionally, a 22 nucleotide long overhang of the RNA was considered (Figure 16). Concomitantly, the 12-subunit containing RNAP II was assembled together with the Spt5 variant1 (chapter 10.4.4). This should constitute a situation whereby all participants of the proposed 'elongation checkpoint model' are on hand (Figure 17).

```
                5' –  GACCAGAAUAUAUAUGCAUAAA
RNA                                          GACCAGGC                  – 3'
TEMPLATE        3' –      CCGTCATGATCATTACTGGTCCGCATTCATGAACTCGAACC    – 5'
NONTEMPLATE     5' –      CCGGCAGTACTAGTAAACTAGTATTGAAAGTACTTGAGCTT    – 3'
```

Figure 16. **RNA/DNA hybrid.** For the reconstitution of an 'elongation checkpoint complex' a 41-mer DNA duplex with an 11 nucleotide mismatched bubble region (blue), and an RNA 30-mer with eight 3'-terminal nucleotides complementary to the DNA template strand in the bubble, was used.

sample 1 = 10 µl fusion protein complex / Spt5 variant1
sample 2 =   2 µl 10-subunit RNAP II
sample 3 =   8 µl assembly reaction before buffer exchange
sample 4 =   2 µl assembly reaction after buffer exchange
sample 5 = 30 µl flowthrough (first 400 µl)
sample 6 =   6 µl 1:5 dilution of sample 4
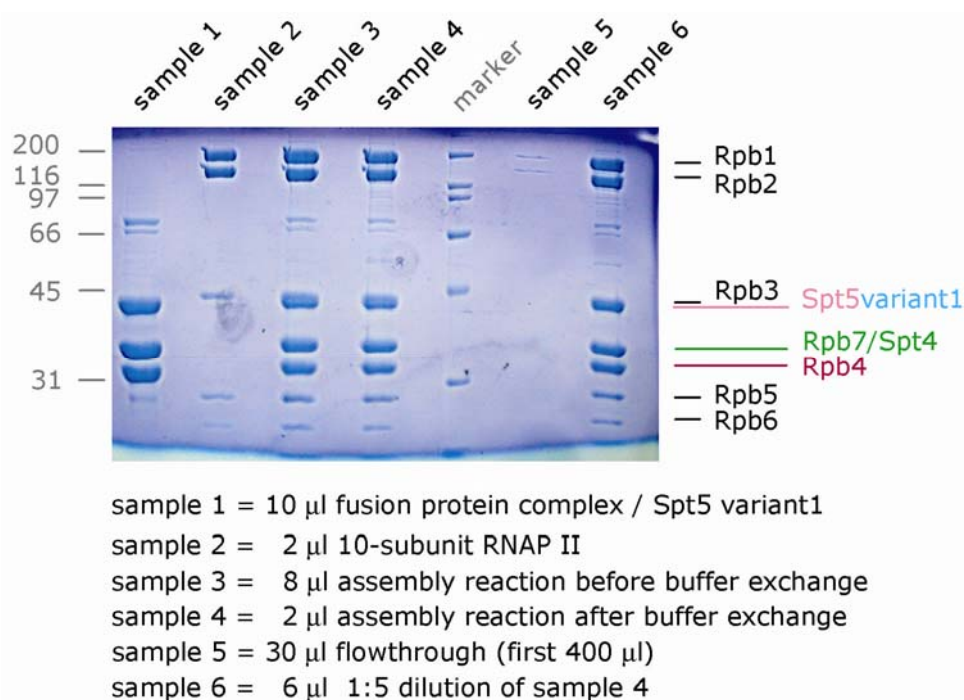
Figure 17. **Assembly of 'elongation checkpoint complex'.** The Coomassie stained SDS-acrylamid gel displays the quality of the artificial 'elongation checkpoint complex'. Subunits of the core RNAP II are shown in black and additional subunits are colored according to Figure 12 and chapter 10.3.1. The two bands found around 66 kDa are impurities.

5.3.5   Crystallization of 'elongation checkpoint complex'

After assembly, the reconstituted RNAP II/Spt5 variant1 and nucleic acid complex was subjected to the usual RNAP II screen (Table 1). Extra DNA/RNA hybrid in the last cryo-solution should increase the occupancy in the crystals. Crystals grew to an ideal size of about 0.3 x 0.15 x 0.1 mm within 1 – 2 weeks at 20 °C. Figure 18 shows typical representatives of crystals obtained under $NH_4NaTartrat/PEG/KSCN$ conditions described in Table 1. In addition, some set-ups were covered with 200 µl of Al's oil (paraffin oil : silicon oil / 1 : 1; Hampton Research). In my case, this treatment did not improve the average crystal size.

with oil layer   **NH₄NaTartrat/PEG
/KSCN**
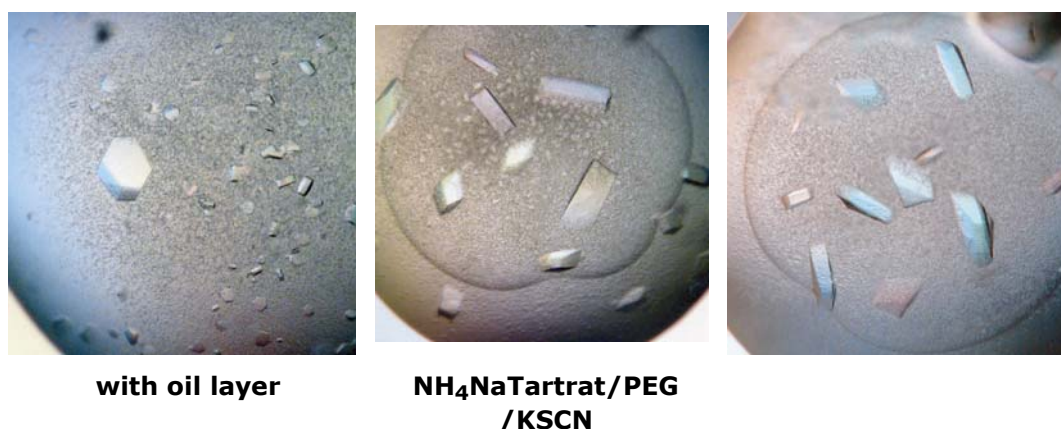
Figure 18. **Single crystals of 'elongation checkpoint complex'.** Crystals were obtained in condition NH₄NaTartrat/PEG/KSCN (Table 1). This condition tends to phase separation and crystals have to be harvested quickly.

5.4   Structural analysis and data survey

The endogenous core RNAP II of yeast was prepared as described elsewhere (PhD thesis Armache, 2005). Crystals were harvested and treated with the appropriate cryo-protectant. The protocol comprises six distinct steps at which the amount of glycerin in the mother solution is increased. Subsequently the crystals were annealed from 20 to 8 °C and frozen in liquid nitrogen. Diffraction data from single cryo-cooled crystals were collected at beamline X06SA at the Swiss Light Source/Switzerland. An overview of the data collection statistics is given in Table 2. The particular name of the crystal was derived from its composition. Based on the core RNAP II extra subunits were appended. Addition of Rpb7/Spt4-Rpb4 generated complex12 (chapter 5.2.3), whereas supplementary added Rpb7/Spt4-Rpb4 and Spt5 variant1 created complex13 (chapter 5.3.3). The complex termed 'elongation checkpoint complex' refers to 12-subunit RNAP II (with artificial Rpb7/Spt4), Spt5variant1 and DNA/RNA hybrid (5.3.5).

Table 2: Statistics of representative crystals

Crystal name	complex12[1]	complex12	complex13	'elongation checkpoint complex'
Space goup	C222$_1$	C222$_1$	C222$_1$	C222$_1$
Unit cell axes (Å)	222.1 393.4 281.4	221.3 393.8 280.9	224.745 399.141 290.7	221.4 393.8 282.5
Wavelength (Å)	0.920039	0.920039	0.97937	0.91929
Resolution range (Å)	50.0 − 4.0 (4.14 − 4.0)[2]	50.0 − 4.31 (4.46 − 4.31)	50.0 − 4.5 (4.66 − 4.50)	50 − 4.2 (4.35 − 4.2)
Unique reflections	101254 9995	150571 15119	68639 5641	88470 8280
Completeness (%)	99.9 (99.8)	94.6 (95.1)	90.9 (75.3)	98.5 (92.9)
Redundancy	4.2 (4.1)	3.8 (3.8)	4.4 (2.6)	5.3 (3.9)
Mosaicity (°)	0.48	0.57	0.70	0.60
R$_{sym}$ (%)	8.5 (29.7)	9.8 (30.6)	7.4 (13.6)	16.6 (36.2)
I/σ(I)	16.0	13.2	10.2	9.0
R$_{cryst}$ after rigid body refinement	30.2	31.8	38.9	33.8

[1] The corresponding electron density will be depicted in Figure 20.

[2] Values in parentheses correspond to the highest resolution shell.

5.4.1 The 12-subunit RNAP II model

As followed from the crystallographic analysis with DENZO, the distinct complexes crystallized in the orthorhombic space group $C222_1$ with quite similar unit cell dimensions. Data reduction was accomplished with DENZO and SCALEPACK. For molecular replacement the program PHASER was chosen. Based on the 1Y1W PDB entry the 12-subunit RNAP II model is shown in Figure 19. The clamp and bridge helix are kept in green. Rpb4 is presented in red and Rpb7 in light blue. In order to highlight the C-terminus of Rbp7 it is depicted in dark blue. The remainder of the enzyme is colored in light pink. The electron density is shown in white, respectively. The model is shown in a back view and tilted downward to emphasize Rpb4-Rpb7.
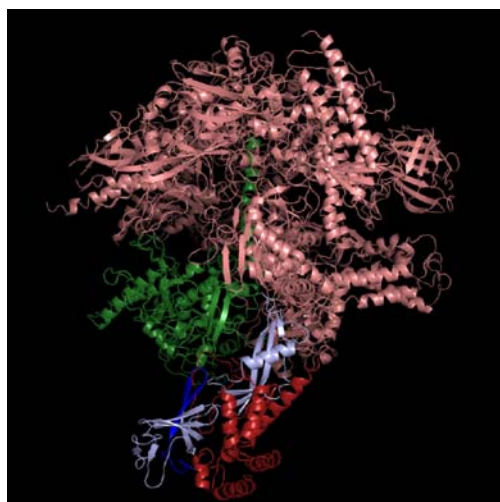
Figure 19. **12-subunit RNAP II model.** The model is derived from the 1Y1W PDB entry.

5.4.2 Electron density map of complex12

At best the crystals of RNAP II comprising the artificial Rpb7-Spt4 subunit were diffracting until 4.0 Å. In Figure 20 A the superposition of the electron density and the clamp/bridge helix are shown to get an idea of the quality of the calculated 2fofc-map at a contour level of 1.7σ. In Figure 20 B the electron density is highlighted which represents Rpb7 and Rpb4 and in which Rpb4 nicely fits. In the case of Rpb7 the situation is different. The N-terminal and middle part of the protein is well defined but for the corresponding extreme C-terminus I could not find an interpretable electron density. The protein segment from leucine 168 to

isoleucine 171 is lacking. But actually this is the most interesting part. The C-terminal region and the glycine linker should make a contribution to position Spt4 in proximity to RNAP II. Unfortunately, no subtle hint of the location of Spt4 could be achieved with this experimental approach. The linker probably allowed Spt4 to be flexible attached to the RNAP II surface and did not enforce a distinct position. The lack of density may suggest the possibility that Spt4 populates alternative conformations. So far, Spt5 has been implicated to recruit the bipartite elongation factor to the transcription machinery. The assembly of RNAP II comprising an artificial subunit could not enforce the formation of a physical interaction between Spt4 and the enzyme additive to the glycine linker. Under these conditions, the experiment suggests that Spt5 is the prerequisite for the aspired multiprotein complex.

A) complex12
electron density of the clamp/bridge helix
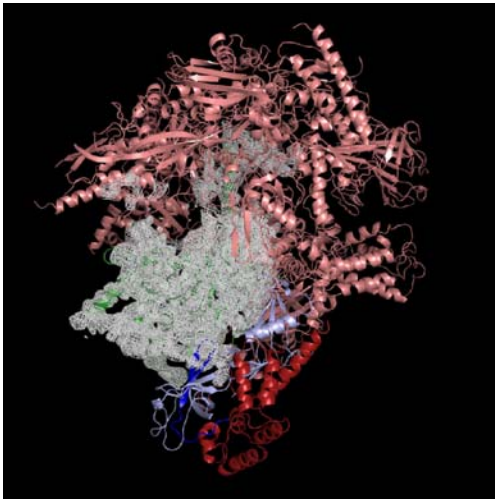
B) complex12
electron density of Rpb7-Rpb4
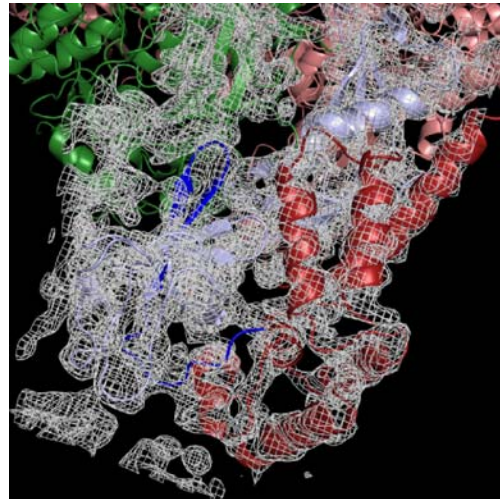
Figure 20. **Electron density map of complex12.** The contour level is 1.7σ. The remaining electron density map belongs to the symmetry mate.

5.4.3   Electron density map of complex13

The crystals comprising core RNAP II, heterologously expressed artificial Rpb7/Spt4-Rpb4 and Spt5 variant1 diffracted until 4.5 Å. In Figure 21 A, the electron density map (contour level 1.7σ of the 2fofc map) covered the residues of the clamp and the bridge helix. Figure 21 B is highlighting the superposition with

Rpb7-Rpb4 at a contour level of 1.7σ. The underlying experimental setup should be a watershed and Spt4 was expected to assist both the recruitment of Spt5 variant1 and the positioning on the RNAP II surface. Strikingly, despite of the crystal size and careful data analysis, it was not possible to localize Spt4-Spt5 on the RNAP II surface by assigning any of the additional density (albeit weak) to this elongation factor. The density of the protein segment of Rpb4 which emanates from the RNAP II surface seems to be slightly shifted. Apparently, the helices three, five and six are identifiable but do not match the predicted and expected position. The early idea that tethering of Spt4 with Rpb7 should create a formidable precursor for macromolecular complex formation did not prove true. But rather the converse was observed. The addition of Spt4-Spt5 to a previously well behaving system seems to influence an established tertiary structure. These findings argue for an alternative assembly of Spt4 and Spt5 variant1 with RNAP II. The brief residency of Spt5 observed here is difficult to reconcile with the opinion that DSIF is a powerful transcription elongation factor and tends to bind as strongly to RNAP II as TFIIS.

A) complex13
electron density of the clamp/bridge helix
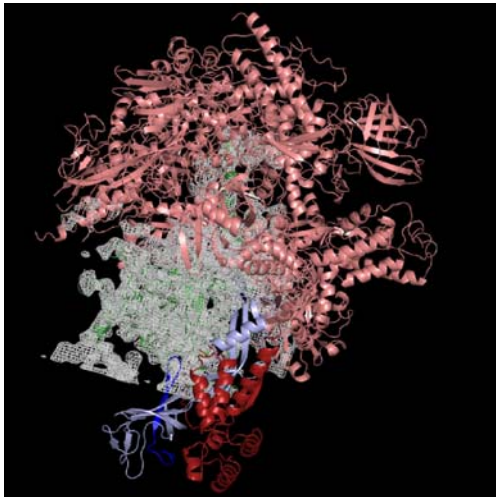
B) complex13
electron density of Rpb7-Rpb4
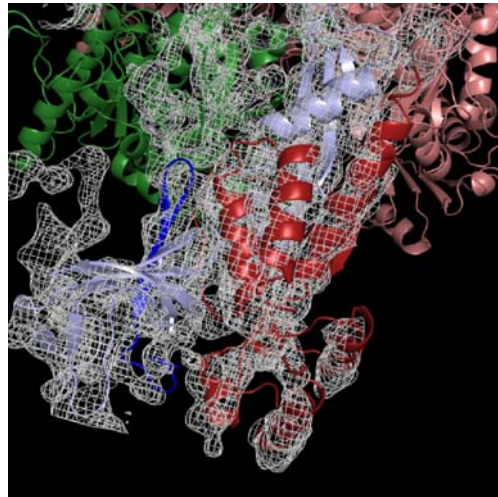
Figure 21. **Electron density map of complex13.** The contour level of both 2fofc maps is 1.7σ.

5.4.4 Electron density map of 'elongation checkpoint complex'

Bringing together different components in a reaction tube and skipping of further polishing but lossy steps should raise the local concentration of the interaction partners and increase the probability of complex formation. The additives were expected to bind to the 'rigid body' of the enzyme. At best the crystals of RNAP II comprising the artificial Rpb7-Spt4 subunit and a DNA/RNA hybrid were diffracting until 4.2 Å. But nevertheless, it does not engender a situation whereby one can reasonably posit a role or location for Spt4-Spt5 in the assumed macromolecular assembly. Rpb4 and Rpb7 fit well into the 2fofc map (contour level 1.7σ) but do not reveal the destiny of the glycine linker and Spt4 as the display detail in Figure 22 A shows. Figure 22 B depicts the model in a front view. In Figure 22 C and 22 D the wall (magenta) and the hybrid binding site (cyan) of Rpb2 are highlighted, respectively. In figure 22 D the perspective is chosen to have a direct view onto the hybrid binding site by omitting interfering protein segments.

Concerning the wall and the hybrid binding site, both regions are defined in the electron density map but confusingly the contribution of the hybrid is missing. It was expected that during the assembly procedure the 12-subunit RNAP II will locate the DNA/RNA hybrid in position allowing the process to be recapitulated. After all, the assembly was performed in the same order as it was done for the complete RNAP II elongation complex structure (Kettenberger *et al.*, 2004). Over and above, evidence that addition of DNA/RNA can improve the affinity of Spt4-Spt5 variant1 for the enzyme could not be perceived. Seemingly, the nucleic acids are not present in the complex and the additional proteins can not be ascertained. The generation of a higher complex containing RNAP II and the heterodimeric protein pair Spt4-Spt5 seems to be hampered. Except for the transcription elongation factor TFIIS, which easily bound to the enzyme, a major lacuna in the search for RNAP II-transcription complexes remains as the full biological significance of the RNAP II structure lies in the implications for the higher complexes formed with different transcription factors. Consequently, this experimental approach to create an 'elongation checkpoint' by adding recombinant components to endogenous core RNAP II does not provide insights into the interacting components and the atomic details of this biochemically well-studied association.

A) 'elongation checkpoint
**   complex'**
electron density of Rpb7-Rpb4
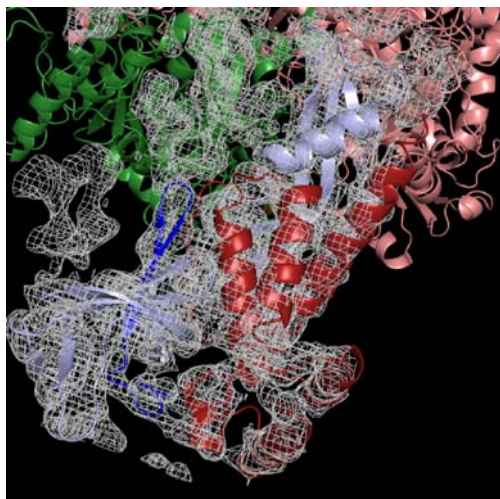
B) 12-subunit model
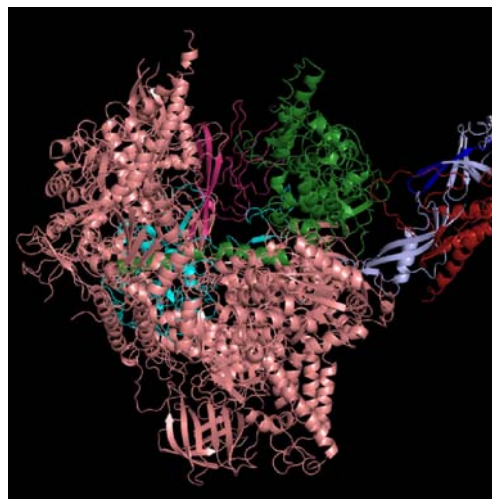
C) wall
electron density of the wall and the clamp
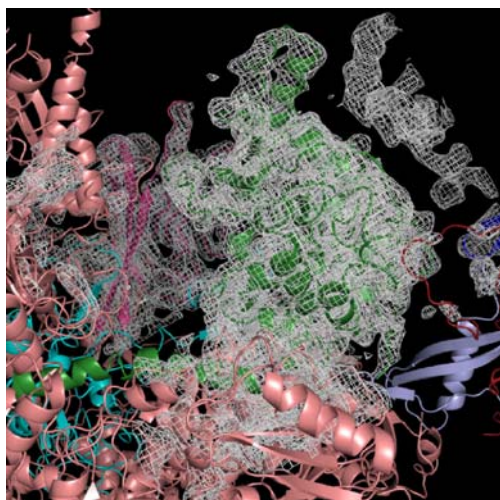
D) hybrid
electron density of hybrid binding region
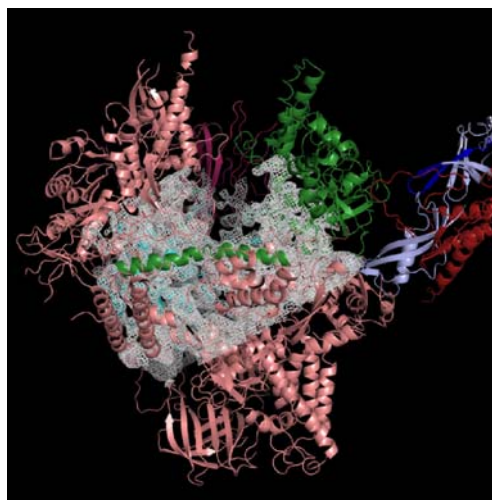
Figure 22. **Electron density map of the bubble-complex.** The contour level of the 2fofc maps is 1.7σ.

5.5   Data interpretation

Although the here presented studies advocate a physical interaction between RNAP II and Spt4-Spt5, the atomic model of this complex as positive proof is still missing. Assembling the components together like pieces of a puzzle does not meet demands of an 'elongation checkpoint complex', allowing the process to be recapitulated using purified constituent parts. But how can such complex formation be accomplished and structurally exploited?

The design of the investigated Spt5 variants was based upon the domain organization of this elongation factor, focussing on the possible molding of the KOW-domains in the presence of RNAP II as its binding partner. I assumed that the proteins will either interact by domain-domain or protein-motif contacts. Separately purified heterologous components assembled efficiently with endogenous core RNAP II but did not form an unequivocally and stable unit which could be captured by protein crystallography. On the first sight, I could conclude from these crystallographic analyses that my recombinant Spt4-Spt5 proteins lack properties of their wild type counterparts, thereby anticipating the complex to be reconstructed *in vitro*. Concerning the fusion protein approach, either an apparently disordered nature of Spt4 hampers a specific interaction with the RNAP II surface or Spt4 lacks an intrinsic binding capability. What's more, it does not confer interaction specificity for the enzyme upon the Spt5 variant1 which could be trapped in a rigid crystal lattice. A phenomenon I could observe was that the contribution of approximately 100 amino acid residues at the N-terminus of Spt5 highly influenced the nature of the protein sample. The segment from amino acid residue 283 to 379 always caused problems. I run the gamut from impaired expression levels to exceptional behaviour on a size exclusion column. As soon as I changed the segment of Spt5 the fidelity changed. This is rather astonishing as the mentioned region has a predicted function (part of the Spt4 binding region) and the *Aquifex aelicus* crystal structure (Steiner *et al.*, 2002) bodes well for revealing the structure of this evolutionary conserved NGN-domain. It may be such that the plasticity of this domain makes Spt5 amenable for specific conformations which I could not trap with the rational design of my Spt5 variants.

Collectively, the model of a Spt4-Spt5/RNAP II complex awaits further proof and I reason that potential antagonistic processes underlying my experimental-setup may also result from either absent posttranslational modifications or an obstructed 'induced fit mechanism'. Spt5 has been demonstrated as a direct methylation target (Kwak *et al.*, 2003). PRMT5 and PRMT1 likely play an important role in altering the protein-protein interaction of DSIF and RNAP II. Methylation of Spt5 is speculated to diminish its association propensity with RNAP II. Regarding my experiments, the Spt5 variant1 is lacking this domain, whereas Spt5 variant3 comprises this segment and both variants bind to RNAP II. Actually, there are no counterparts in yeast for the predicted methylation sites in human Spt5. Guided by the GRG sequence motif stimulating methylation by PRMT1, there could be a single, adequate site lying in between the predicted ones. Nevertheless, it has to be considered that insensible to the existence of this region in the designed proteins, *Escherichia coli* is not capable of accomplishing posttranslational modifications. Notwithstanding these facts, I speculate that the acidic N-terminus and/or the repetitive C-terminus could orchestrate the binding onto RNAP II in a yet not established manner. It has been shown that the KOW-domains potentiate the binding but by virtue of crystallography I neither could confirm this simple rule nor present a mechanism based on the domain organization. The C-terminus of Spt5 is termed 'mini-CTD' which is phosphorylated in the process of transcription elongation (Yamada *et al.*, 2003). Additional posttranslational modifications on both termini or on a distinct terminus could influence the direct and well-defined establishment of the 'elongation checkpoint complex'. Another possibility could be that the generated Spt4-Spt5 conjugates are missing a not previously detected but important protein-protein binding region that undergoes a disorder-to-order transition as Spt5 binds to RNAP II. It is becoming increasingly clear that many functionally important protein segments occur outside of globular domains (Linding *et al.*, 2003) and that a couple of interactions involve unstructured parts of a protein that becomes ordered only on binding to its partner (reviewed by Aloy and Russel, 2006; Radhakrishnan *et al.*, 1997). In my case, upon complex formation either the predicted unstructured N-terminus or the C-terminus could become ordered and extend the interaction interface. As mentioned above, the C-terminus of Spt5 is alike the CTD of RNAP II. Latest results propose an evolutionary conserved reccuring pentapeptide motif in the CTR (C-terminal region) of hSpt5. In addition,

this consensus sequence G-S-R/Q-T-P is the target site of P-TEFb, which also phosphorylates the CTD, thereby controlling the elongation phase of transcription (Yamada *et al.*, 2006; reviewed by Peterlin and Price, 2006). Nevertheless significant differences exist. For instance, the CTR repeats are not contiguous and are not located at the extreme C-terminus. Strikingly, in the case of *Saccharomyces cerevisiae*, it was not possible to pinpoint a consensus motif. Regarding this thesis, the difficult generation and the varying life span of the Spt4-Spt5/RNAP II complex could also be a reflection of the distinctive feature of the C-terminus of budding yeast.

Last but not least, the role of Spt4 in the heterodimeric transcription elongation factor still remains elusive. Spt4 has been exclusively assigned to exert a positive role in transcription elongation. It may regulate the switch to processive transcription as it overcomes the Isw1p-dependent pausing of RNAP II at the onset of elongation (Morillon *et al.*, 2003). In addition, spt4Δ cells are impaired in transcription of *lacZ* and other long and GC-rich DNA sequences driven from the *GAL1* promoter (Rondon *et al.*, 2003). The mechanism how Spt4 interacts with Spt5 has not yet been investigated. This would allow a functional dissection of their complementary but perhaps also different impact on the process of transient pausing of RNAP II. This aspect merits further examination. Latest results suggest that the phosphorylation pattern of Spt5 accounts for its dual function (Yamada *et al.*, 2006). DSIF is clearly capable of both repressing and activating transcription under different conditions (Yamada *et al.*, 2006 and reference therein). According to my experiments, I can deduce that Spt4 has a profound effect on the stability of Spt5. I could hardly get any material from a two liter culture when Spt5 variant1 was expressed in isolation. Even in the case where Spt4 was fused to Rpb7, the amount of coexpressed Spt5 variant1 proved satisfactory and the quality of the sample was suitable for crystallization trials. The basis for specific heterodimerization is unknown, and must be investigated.

6 STRUCTURE AND CTD-BINDING OF THE SET2 SRI DOMAIN THAT COUPLES HISTONE H3 LYSINE 36 METHYLATION TO TRANSCRIPTION

The appreciation that maturation of nascent mRNA occurs by interdependent and cotranscriptional processes that are physically and functionally connected by the CTD has changed the view of gene transcription. Currently, each stage is considered as a subdivision of a continuous process (reviewed by Orphanides and Reinberg, 2002). In addition, the role of chromatin has been adjusted from a packaging device to a highly dynamic unit whose modifications influence gene transcription (reviewed by Mellor, 2006). Until recently, both activities were regarded on their part as autonomous. This narrow perspective is changing as accumulating evidence suggests that histone methylation plays an important role in the process of transcription elongation. While transcribing coding regions, the CTD of RNAP II mediates methylation of H3K36 by Set2. This distinctive histone lysine methyltransferase is associated with RNAP II in a manner that is dependent both on the CTD and the Ctk1/Bur1 kinases that phosphorylate the CTD on serine 2. The precise region in Set2 which is required for crosstalk with RNAP II was determined lately and termed SRI-domain (Kizer *et al.*, 2005). These findings were so innovative to catch my attention and structural biology methods like NMR spectroscopy and X-ray crystallography are prolific techniques to illuminate the relationship between Set2 and RNAP II.

6.1 Domain mapping and crystallization of SRI domain variants

I have used a combination of available bioinformatic methods for predicting protein secondary structure and published interaction studies (Kizer *et al.*, 2005) to design several SRI domain variants. First crystallization attempts with different variants comprising additional segments N- and C-terminal of this domain failed. It was not possible to narrow down the crystallizable protein segment experimentally by repeating experiments with smaller constructs. Extensive screening with various protein samples at discrete concentrations or usage of diverse commercial screens did not lead to crystals. Furthermore, a change in temperature to influence the crystallization process was not fruitful.

In a new approach as it is depicted in Figure 23, I did not confine the SRI domain to the proposed region but rather deleted several additional amino acids to increase the likelihood of crystal growth. An expression plasmid based on pET24d and containing an N-terminal hexahistidine tag was used to produce several N- and C-terminal truncations of the SRI domain.

Figure 23. **Set2 SRI sequence.** The investigated SRI domain comprises amino acids 619 to 718 in yeast. The helical regions predicted by PredictProtein (http://cubic/bioc/columbia/edu/predictprotein.hmtl), a secondary structure prediction surfer, are marked in green. In addition, primer start sites are indicated.

I was expecting that the newly generated SRI variants make it possible to harness the technology of X-ray crystallography to look for the atomic details of the SRI domain alone or in conjunction with a CTD peptide (Figure 24). For this reason two synthetic equivalents were ordered at the companies Jerini (http://www.jerini.com/) and Anaspec (http://www.anaspec.com/). As there was no indication neither for the crystallization condition nor for the sample concentration, I did several varying set-ups as summarized in Table 3. The molar ratio SRI domain : peptide was 1 : 1.25. Unfortunately, cocrystallization of the putative interacting partners did not yield crystals.

Figure 24. **The Sequence of CTD-peptides.** The two CTD-peptides are depicted which were used for the crystallization set-ups.

Concomitantly accomplished crystallization set-ups of the SRI domain alone gave crystals in two cases but only after 3 months (Figure 25). The aberrant crystallizing degradation products could not be identified. The variants comprising amino acids 632 – 707 or 632 – 709 are further truncated at the N-terminus but in contrast did not crystallize. The applicability of this approach was limited as the very thin and tiny crystals could not be reproduced. Albeit the proteins were not crystallizing it

was obvious that they were pure and highly soluble. With this knowledge I pursued a new strategy to shed light into the structure of this extremely interesting domain. I decided to tackle the problem by NMR spectroscopy.

Table 3: Crystallization setups with different SRI variants/peptides

SRI variant	[c]	Crystallization screen	peptide
aa 624 – 705	n. d.		
aa 624 – 707	6 mg/ml	Classic HTS[3]	
	6 mg/ml	Classic HTS[3]	jPT
aa 624 – 709	40.0 mg/ml	Natrix/ PEG-Ion/ MPD suite[1]	
	18.5 mg/ml	Classic HTS[3]	Anaspec
aa 629 – 705	10.5 mg/ml	Classic HTS	Anaspec
aa 629 – 707	31.0 mg/ml	Index[1]	
	31.0 mg/ml	Classic[2]/ Cations[2]	
	16.3 mg/ml	Classic[2]	Anaspec/ jPT
	16.3 mg/ml	Classic HTS[3]	Anaspec/ jPT
aa 629 – 709	12.8 mg/ml	Classic[2]	Anaspec
aa 632 – 705	n. d.		
aa 632 – 707	13.3 mg/ml	Classic[2]	Anaspec
aa 632 – 709	11.5 mg/ml	Classic[2]	Anaspec

[1] Hampton research (www.hamptonresearch.com/)

[2] Nextal (http://www.nextalbiotech.com/)

[3] Jena bioscience (http://www.jenabioscience.com/)

SRI variant aa 624 – 709
CdSO$_4$ HYDRATE

SRI variant aa 629 – 707
CdSO$_4$ HYDRATE
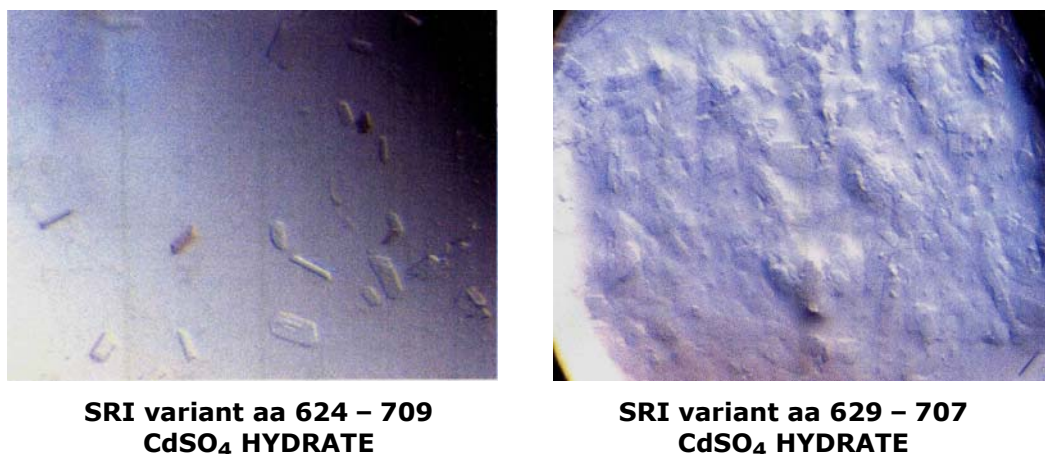
Figure 25. **Preliminary crystals of two SRI domain variants.** Crystals were obtained in conditions of the commercial screen Nextal[2] Classic Suite #39 containing 0.05 mM CdSO4 hydrate/0.1 M Hepes pH 7.5/0.5 M sodium acetate anhydrous.

Before I could start to elucidate the structure by NMR spectroscopy I had to reclone my protein for two reasons. Firstly, a substantially feature of the SRI domain is its vast content of lysines, exactly 20 %. Therefore I had to reduce any additional positive charge by removing the hexahistidine tag. Secondly, for the expression of isotope labelled protein the amount of culture should not extend two litres. To fulfil both requirements I subcloned the coding sequence of the SRI domain into a modified pET9d vector, which was kindly provided by G. Stier from EMBL/Heidelberg. The containing z-tag was thought 'to boost' the expression and by means of the TEV-cleavage side I could remove the tag. This also gave me the opportunity to go back to the 'full length' SRI domain to delineate the whole protein segment promoting the interplay with RNAP II.

6.2   The Set2 SRI domain forms a conserved three-helix bundle

The solution structure of the yeast Set2 SRI domain was determined by multidimensional NMR (Table 4). The structure revealed three α-helices arranged in a left-handed bundle (Figure 26 B). The N-terminal helix $\alpha1$ is slightly kinked at residues F639 and V640, and the linker between helices $\alpha1$ and $\alpha2$ includes a short 3_{10}-helical turn at residues S650-Q652. A hydrophobic core is formed by numerous residues located at the interface between the three helices, including four residues in the two regions linking the helices (Figure 26 C). Consistently, the heteronuclear {1H}-15N NOE measurements demonstrate that the polypeptide backbone in all three helices and the connecting linker regions is rigid (Figure 26 C). The hydrophobic core residues are generally conserved across species (Figure 26 C), demonstrating that our structure is a good model for SRI domains in Set2 of other species.
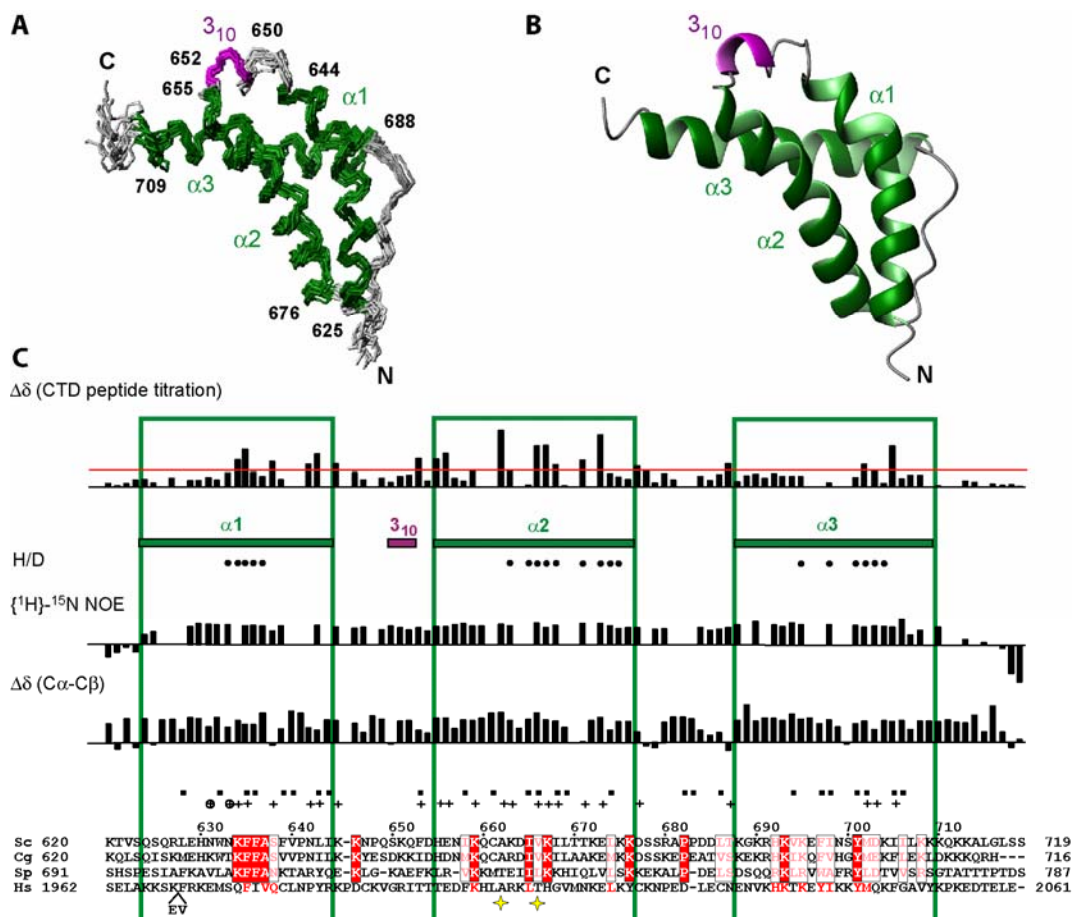
Figure 26. **Structure and CTD binding of the yeast Set2 SRI domain.**

A) ensemble of final NMR structures. The three α-helices are shown in green, and a short 3_{10}-helix is shown in pink.

B) ribbon diagram of the lowest energy structure in A.

C) alignment of SRI domain sequences and NMR structure determination and CTD binding data. The secondary structure is shown above the sequence. Solvent-protected amide protons that show slow H/D exchange are indicated by filled circles. Secondary chemical shifts $\Delta\delta(C\alpha-C\beta)$ are indicated by black bars. Residues that experience large chemical shift perturbations upon addition of the CTD two-repeat phosphopeptide SPS-YpSPTpSPS-YpSPTpSPS (pS = phosphoserine) are indicated above the alignment with crosses and circled crosses for backbone and side chain amides, respectively. Yellow stars indicate residues Ala662 and Val666 that are implicated in binding of a CTD tyrosine side chain. Residues that are identical and conserved in fungal Set2 homologues are on red background and in red, respectively. Hydrophobic core residues are marked with a black square.

Table 4: NMR structural statistics for the yeast Set2 SRI domain

	$<SA>$[1]	$<SA>_{water-ref}$
Number of NOE derived distance restraints		
All (unambiguous/ambiguous)	1958/198	
Long range |i-j|> 4 (unambiguous/ambiguous)	433/26	
R.m.s. deviation (Å) from experimental distance restraints[2]		
R.m.s.d. (NOEs)	0.0147±0.0005	0.025±0.006
Hydrogen bonds (2*20)	0.026 ± 0.004	0.043 ± 0.006
R.m.s. deviation (°) from experimental torsion restraints[3]		
R.m.s.d. (83 Φ/Ψ)	0.68 ± 0.07	0.88 ± 0.12
Coordinate Precision (Å) residues 10-94[4]		
N, Cα, C'	0.38 ± 0.06	0.48 ± 0.08
All heavy atoms	1.01 ± 0.05	1.05 ± 0.06
Structural quality[5]		
Bad contacts	1.8 ± 0.8	0.0 ± 0.0
Ramachandran plot		
% in most favored region	90.2 ± 1.4	93.6 ± 1.6
% in additionally allowed region	9.4 ± 1.4	6.3 ± 1.6

[1] $<SA>$ is an ensemble of ten lowest-energy solution structures (out of 100 calculated) of the Set2 SRI domain before water-refinement. The CNS E_{repel} function was used to simulate van der Waals interactions with an energy constant of 25.0 kcal mol^{-1} Å^{-4} using "PROLSQ" van der Waals radii; r.m.s. deviations for bond lengths, bond angles and improper dihedral angles are 0.0020 ± 0.0001 Å, 0.382 ± 0.008° and 0.31 ± 0.01°. 1 kcal = 4.18 kJ.

[2] Distance restraints were employed with a soft square-well potential using an energy constant of 50 kcal mol^{-1}Å^{2}. For hydrogen bonds, distance restraints with bounds of 1.8-2.3 Å (H-O), and 2.8-3.3 Å (N-O) were derived for slow exchanging amide protons. No distance restraints were violated by more than 0.3 Å in the $<SA>$ structures.

[3] Dihedral angle restraints derived from TALOS (Cornilescu *et al.*, 1999) were applied to backbone angles using energy constants of 200 kcal mol^{-1} rad^{-2}. No dihedral angle restraint was violated by more than 5°.

[4] Coordinate precision is given as the Cartesian coordinate r.m.s. deviation of the 10 lowest-energy structures in the NMR ensemble with respect to their mean structure.

[5] Structural quality of the NMR ensemble was analyzed using PROCHECK (Laskowski *et al.*, 1996).

6.3 The SRI domain defines a novel CTD-binding fold

Comparison with the five known structures of CTD-binding domains reveals that the SRI domain defines a novel CTD-binding fold. Other CTD-binding domains include FF domains, CTD-interacting domains (CIDs), WW domains, BRCT domains, and a domain in the Cgt1 subunit of the 5'-capping enzyme (reviewed by Meinhart *et al.*, 2005). Of these, FF and CID domains also form helical bundles (Allen *et al.*, 2002; Meinhart and Cramer, 2004) but, in contrast to the SRI domain, the superhelical arrangement in these two domains is right-handed (Figure 27). Thus the six CTD-binding domains that have been structurally characterized use different folds for specific CTD recognition.
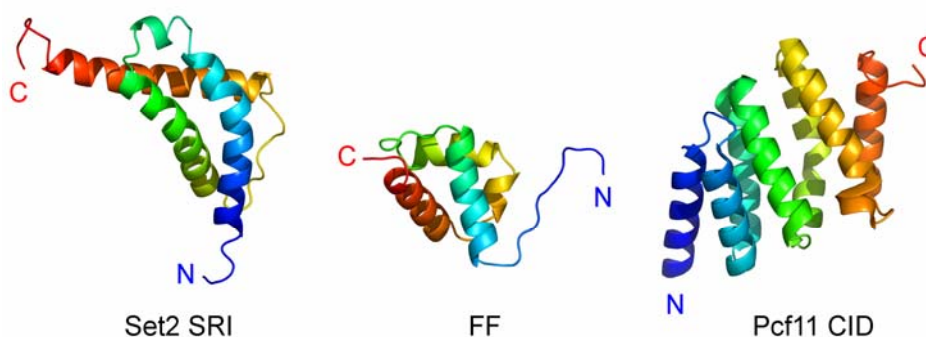
Figure 27. **Comparison of the Set2 SRI domain with known helical CTD-binding folds.** The structures are shown as ribbon models colored from blue to red from the N- to the C-terminus, respectively. Shown are from left to right: the Set2 SRI domain (this study), an FF domain, and the Pcf11 CID domain. For orientation of the structures, the N-terminal helices were superimposed. Note that the SRI domain shows a left-handed superhelical arrangement, whereas the two other domains adopt a right-handed arrangement.

6.4 The SRI domain binds a two-repeat CTD phosphopeptide

To characterize the CTD-binding determinants of the SRI domain, we performed NMR titration experiments with S2/S5-phosphorylated CTD peptides (Figure 28 C). A phosphopeptide consisting of a single CTD repeat (Figure 28 A, YpSPTpSPS, pS=phosphoserine) did not perturb chemical shifts in a 2D ^{1}H, ^{15}N HSQC spectrum, indicating that there is no significant binding (not shown). However, titration with a peptide that comprised two CTD repeats and three flanking N-terminal residues (SPS-YpSPTpSPS-YpSPTpSPS) resulted in many strong chemical shift perturbations

(Figure 26 C, Figure 28). From the titration data the dissociation constant is estimated to be in the low µM range, comparable to the reported approximate affinity of 6 µM for a CTD phosphopeptide comprising three repeats (Kizer et al., 2005).
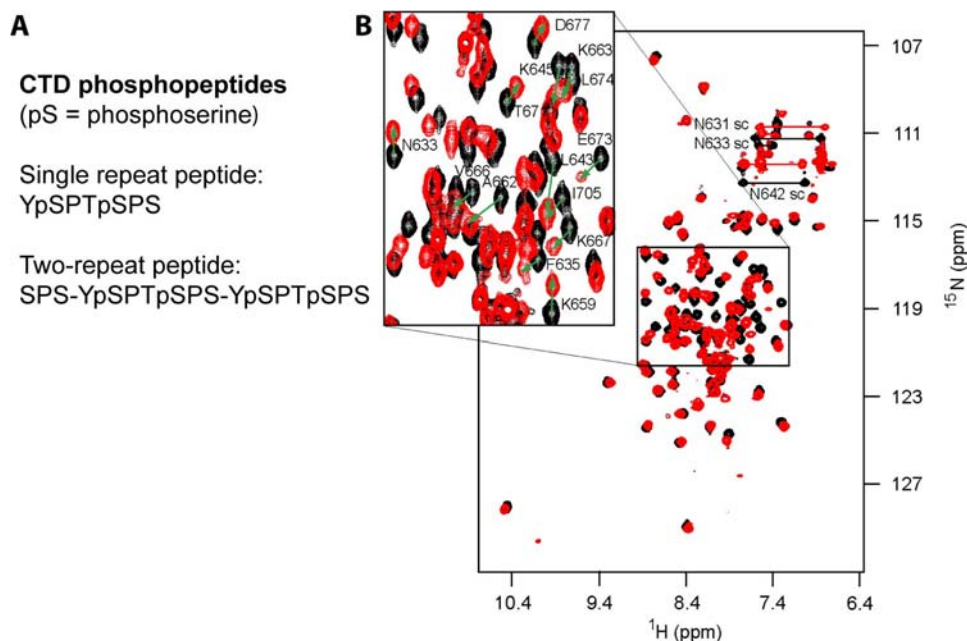
Figure 28**. NMR-monitored CTD binding of the Set2 SRI domain.**

A) Sequence of the two CTD peptides used for binding studies. The upper peptide did not bind the SRI domain, the lower peptide bound with an apparent dissociation constant in the low µM range.

B) NMR titration experiment. Shown are 2D ^{1}H, ^{15}N-HSQC spectra before (black) and after (red) addition of a 1.25-molar excess of the two-repeat.

6.5 Regions in the SRI domain that interact with the CTD

Residues that show strong chemical shift perturbations of their backbone NH groups cluster in two regions on the SRI domain structure (Figure 29 A). The first region includes residues K634, F635 in α1, and A662, V666, K667, T670, T671, and E673 in α2, whereas the second region includes residues F653, H655, E656 in the α1-α2 linker, and residue I705 in α3 (Figure 26 C, Figure 29 A, Figure 28). With the exception of I705, the strongest perturbations upon peptide binding were observed in region 1 (F635, A662, V666, K667, and E673). In this region, the side chain NH2

groups of residues N631 and N633 also show significant chemical shift perturbations (Figure 28 B). Both regions are conserved among fungal Set2 homologues (Figure 29 B), befitting the conserved function of the *Schizosaccharomyces pombe* and *Neurospora crassa* Set2 homologs (Morris *et al.*, 2005; Adhvaryu *et al.*, 2005). The observation of two putative CTD-binding regions, and the finding that two CTD repeats are required for SRI domain binding, indicate that the phospho-CTD extends over a long distance along helices $\alpha1$ and $\alpha2$, and the connecting linker.
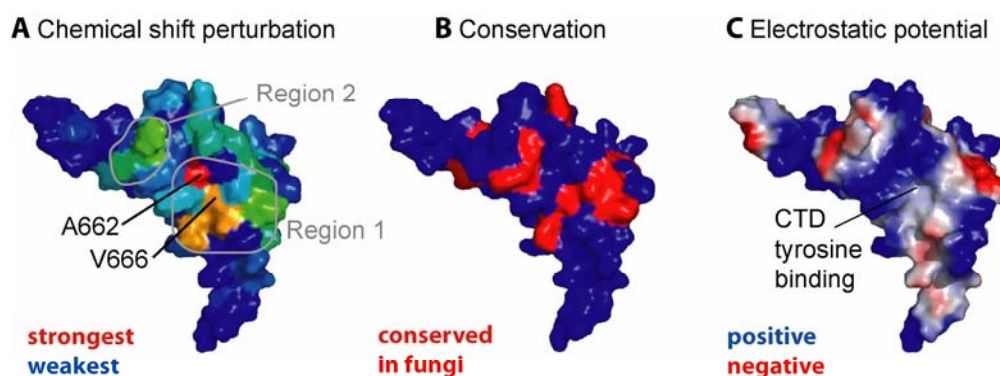
Figure 29. **Surface analysis of the Set2 SRI domain.** Surface representation of the SRI domain color-coded according to chemical shift perturbation of backbone NH and side chain NH_2 groups upon binding of the CTD peptide (Figure 28 A), colored from red to blue for strong to weak perturbations, respectively (A); amino acid conservation among fungal homologues in the alignment of Figure 26 C (B); and electrostatic surface potential (blue and red for positive and negative charges, respectively) (C).

6.6 CTD tyrosine side chains contribute to SRI domain binding

The peptide titration experiments also revealed that the two-repeat CTD peptide (Figure 28 A) binds to the SRI domain via its tyrosine residues. Intermolecular NOEs between both CTD tyrosine side chains and the SRI domain were detected (not shown). Preliminary assignments indicate that one of the tyrosine side chains is in proximity of residues A662 and V666 in region 1 (Figure 26 C, Figure 29 B). These two residues are part of a hydrophobic patch between helices $\alpha1$ and $\alpha2$, and flanked by positively charged surfaces (Figure 29 C), as expected for interaction with the negatively charged phospho-CTD. Interestingly, the tyrosine-proximal residue A662 is identical in human Set2, as are F635, E656, and E673 in the

putative CTD-binding regions (Figure 26 C). In the three known CTD-protein complex structures, the Y1 side chain is also involved in hydrophobic contacts (Meinhart and Cramer, 2004; Fabrega *et al.*, 2003; Verdecia *et al.*, 2000), suggesting that Y1 binding is a general feature of CTD recognition. Previous studies revealed that the CTD can adopt different conformations (reviewed by Meinhart *et al.*, 2005), and this structurally versatile nature of the CTD discourages any detailed model building.

6.7 The SRI domain resembles a polymerase-interacting domain in bacterial sigma factors

Comparison of our structure with known folds in the database DALI (Holm and Sander, 1995) strikingly shows that the SRI domain resembles a region in bacterial σ factors (Figure 30). The four highest hits were the sigma factors σ^{28} (PDB-code 1rp3), σ^E (PDB-code 1or7), σ^R (PDB-code 1h3l), and σ^{70} (PDB-code 1sig), which show DALI scores of 5.6, 5.4, 5.1, and 4.9, respectively, and RMS deviations between 3.3 and 3.7 Å. The region in σ^{70} that is structurally related to the SRI domain is domain 2 (σ_2), which interacts with the clamp region of the core RNAP II upon formation of the holoenzyme (Murakami *et al.*, 2002). The σ_2 domain is involved in binding the −10 element of promoter DNA and contributes to DNA melting during initiation (reviewed by Gross *et al.*, 1998). In the eukaryotic initiation complex, promoter DNA around position -10 lies near the N-terminal domain of the initiation factor TFIIEα (Forget *et al.*, 2004), which shows weak sequence homology (Okhuma *et al.*, 1991) and structural similarity (Meinhart *et al.*, 2003) to the bacterial σ_2 domain. We speculate that the eukaryotic TFIIEα N-terminal domain, which may contact promoter DNA, and the Set2 SRI domain, which binds the negatively charged phospho-CTD, both evolved from the bacterial σ_2 domain.
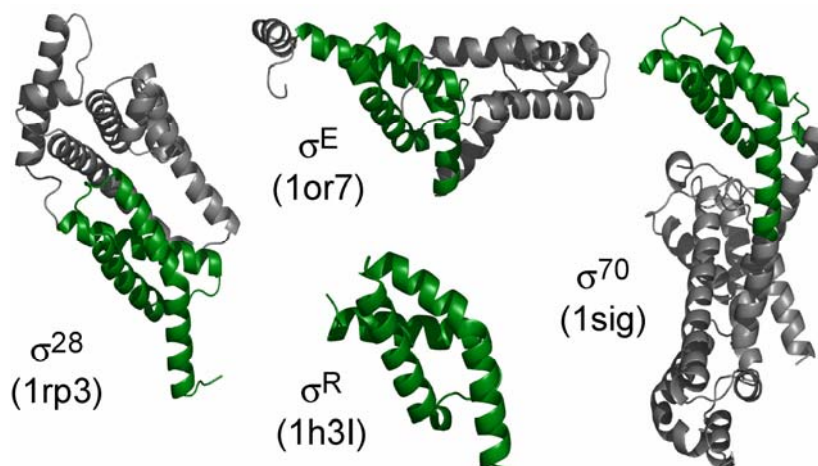
Figure 30. **The Set2 SRI domain resembles a domain in bacterial σ factors.** The domain in the σ factor that resembles the SRI fold is highlighted in green. The PDB codes of the structures are given in parentheses.

6.8 Structural studies of CTD phosphopeptide and SRI domain-peptide complex

6.8.1 Free CTD phosphopeptide

The structural plasticity and divergent phosphorylation pattern of the CTD engenders a versatile landing pad for a consortium of different transcription factors, which recognize this site-specific modifications. In order to examine a putative secondary structure of the synthetic equivalent of the CTD used in this study, I performed several NMR experiments. These were in accordance to previous studies which led to the proposal that CTD-derived peptides do not form stable secondary structure elements (Noble *et al.*, 2005; reviewed by Meinhart *et al.*, 2005).

A set of TOCSY and ROESY experiments recorded on SPS-YpSPTpSPS-YpSPTpSPS revealed several intraresidue correlations and a few crosspeaks consistent with H^α (serine)-H^δ (proline) correlations (Figure 31). On the first sight, the three H^α (serine)-H^δ (proline) correlations could point at the three Pro-Ser amino acid residue pairs in the sequence. Unfortunately, the C^α-H–NH region of the superimposed spectra did not clarify, if the observed NOEs are indeed sequential correlations or may arise from nonsequential residues due to secondary structure or

even belong to one of the Ser-Pro pairs in the peptide. Comparison of *cis* and *trans* Xxx-Pro protein fragments (where Xxx is any amino acid) shows that the distances between NH_i, αH_i, $dCH_{2(i+1)}$, and $\alpha H_{(i+1)}$ are particularly strongly affected by the *cis-trans* isomerization. Thus, the *cis* form allows much closer contacts between αH_i and $\alpha H_{(i+1)}$ and between NH_i and $\alpha H_{(i+1)}$, whereas the *trans* form favors short distances between αH_i and $\delta CH_{2(i+1)}$, and between NH_i and $\delta CH_{2(i+1)}$ (Wüthrich, 1986). Concerning the peptide sequence, I would assume to observe not only three but five of these highly specific interresidue NOEs, if the observed NOEs belong to Ser-Pro pairs. Albeit the $C\alpha$-H–NH region of the ROESY spectra provided additional evidence for two possible pSer-Tyr pairs (data not shown), a sequence specific assignment for this segment could, however, not be established. Conclusively, due to the repetitive character of the peptide it was not possible to infer a sequential connectivity. Hence, for this peptide, no sequence specific assignment could be made and no further structural studies were pursued.
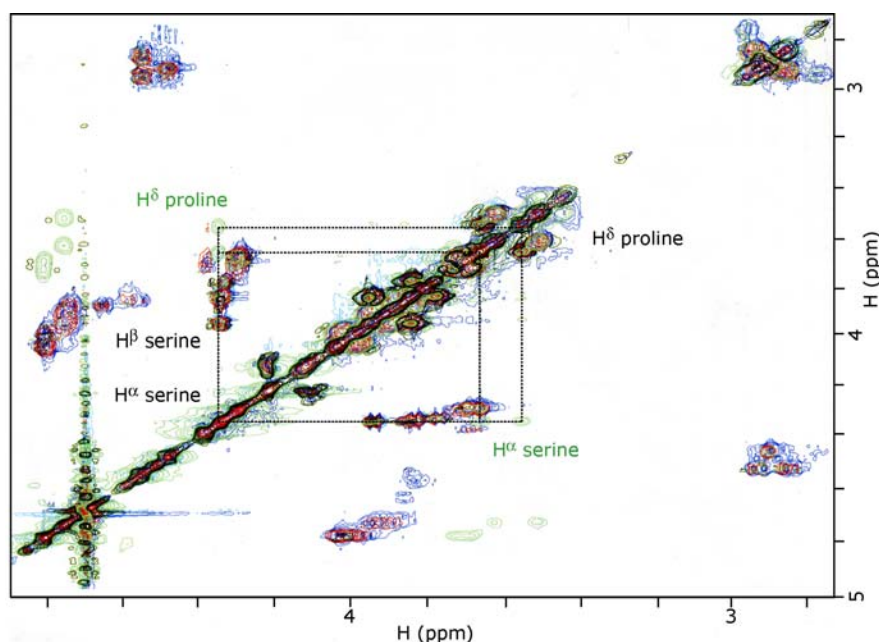
Figure 31. **TOCSY and ROESY spectra of 'free' peptide.** Both TOCSY spectra are superimposed and shown in red (mixing time 60 ms) and blue (mixing time 30 ms). The ROESY spectra is colored in green. One H^α (serine)-H^δ (proline) correlation is explicitly depicted. The spectrum was recorded on a 0.4 mM peptide sample in 100 % D_2O.

6.8.2 Complex structure evaluation

Distance restraints of the putative SRI domain-CTD phosphopeptide complex were derived from 2D NOESY or ^{13}C-resolved 3D NOESY. Essentially, the procedure was the same as for the free SRI-domain. In contrast, the final ensemble of NMR structures was not refined, albeit NOE violations derived from ARIA were considered. The change in conformation of the protein with addition of peptide was weak. From the derived structure I can not deduce why two repeats of the CTD are a necessity for successful binding. The residues which are involved in binding cluster in two regions on the SRI domain structure (Figure 29 A) but the structure of the SRI domain in presence of the CTD phosphopeptide does not pinpoint how the small SRI domain, which comprises only 100 amino acid residues, binds this relatively large peptide (Figure 32).
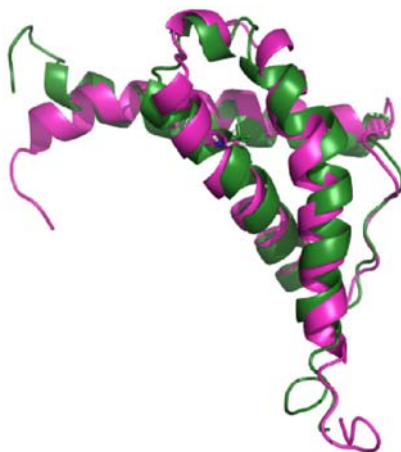
Figure 32. **The preliminary SRI domain-peptide complex structure.** The published free SRI domain is depicted in green and the preliminary complex structure is shown in pink. Ala662 is shown as stick.

Almost concomitantly with the here presented work, the NMR solution structure of the human SRI domain was solved which displays 23/37 % identity/similarity to its yeast counterpart (Li *et al.*, 2005). The SRI domain of HYPB (huntingtin yeast partner B) protein comprises the amino acid residues 1956 to 2056 of the C-terminus and also adopts a left-turned three-helix bundle. After investigating the fact that a CTD peptide containing four contiguous Ser-2/Ser-5 phosphates is sufficient for maximal binding, mutagenesis studies were performed. By mapping

NMR perturbations, the binding site of the peptide was roughly assessed and the nearby side chains actually involved in the interaction should be defined more precisely. The outcome was a model which states that the positively charged residues from helix 2 (His62, R58, K54) contribute to the interface by providing docking sites for the negatively charged phosphoepitopes of Ser5 or Ser2. The principle of electrostatic anchors resembles therein the interface of the Cgt1-CTD peptide complex (Fabrega *et al.*, 2003). It is speculated, that the spacing of the phosphoserines determines the conformation of the CTD which is additionally stabilized by hydrophobic interactions. Tyrosine and proline of the heptapeptide motif accommodate the CTD on the target surface. This binding mode would disregard an extended conformation of the CTD-peptide. The WW-domain of Pin1 binds an entire signature motif of the CTD as an extended coil, with both phosphoserine/proline peptide bonds in the *trans* configuration (Verdecia *et al.*, 2000).

In contrast with this published report, I was unable to obtain an appropriate 'pattern' of amino acid residues involved in binding or deduce the *cis-trans* isomerization state of the CTD peptide. In my case, the model of electrostatic anchors would be tantalizing and indeed explain partially, why at least two repeats are necessary to observe binding between the SRI-domain and a CTD peptide and concomitantly only slight conformational changes are observed in the SRI domain-phosphopeptide structure. As already stated in chapter 6.6, preliminary assignments indicate that one of the tyrosine side chains is in proximity of residues A662 and V666 in region 1 (Figure 26 C, Figure 29 B). Further examination of isotope-filtered experiments and a 3D ^{13}C-edited NOESY implies an interaction between the same tyrosine and an unknown lysine residue. Owing to the high content of lysines, the resulting cross peaks were not well dispersed and a complete assignment of these residues was impaired. In addition, NOEs between a peptide proline and an isoleucine of the SRI domain could be observed but not assigned. The obstacle that the peptide was not labeled, prevented detailed characterization of the SRI domain/peptide interface in my hands. But the present work indicates a new fold of the SRI domain which contributes to the exploitation of CTD recognition modes.

Chapter III: Materials and Methods

7   GENERAL METHODS

7.1   Bacterial strains

Strain	Description	Source or reference
DH5α	F⁻φ80d*lac*ZΔM15 Δ (*lac*ZYA⁻*arg*F) U169 *rec*A1 *end*A1 *hsd*R17(r_k^-, m_k^+) *pho*A *sup*E44 λ⁻ *thi*⁻1 *gyr*A96 *rel*A1	Woodcock et al., 1989
XL-1 blue	*recA1 endA1 gyrA96 thi⁻1 hsdR17 supE44 relA1 lac* [F'*proAB lacI*q*ZΔM15 Tn10* (Tet^r)]	Stratagene
BL21 (DE3)-RIL	*E. coli* B F⁻ *ompT hsdS*(r_B^- m_B^-) *dcm+* Tet^r *gal* λ (DE3) *endA* Hte *[argU ileY leuW* Cam^r]	Stratagene
BL21 (DE3) pLysS	F⁻ *ompT hsdS*$_B$($r_B^-m_B^-$) *gal dcm* (DE3) pLysS (Cam^R)	Stratagene

7.2   Plasmids

#	Name	Protein segment	Vector	Restriction sites	Affinity tag	Expression
1	variant1	Spt5 283 – 620	pET21b[1]	NdeI/NotI	His$_6$	low and degradation
2	variant2	Spt5 380 – 620	pET21b[1]	NdeI/NotI	His$_6$	8 mg/ml
3	variant3	Spt5 283 – 849	pET21b[1]	NdeI/NotI	His$_6$	low and degradation
4	variant4	Spt5 380 – 849	pET21b[1]	NdeI/NotI	His$_6$	7.5 mg/ml
5	variant5	Spt5 283 – 874	pET21b[1]	NdeI/NotI	His$_6$	degradation
6	variant6	Spt5 380 – 874	pET21b[1]	NdeI/NotI	His$_6$	12 mg/ml
7	variant7	Spt5 443 – 849	pET21b[1]	NdeI/NotI	His$_6$	low
8	variant8	Spt5 529 – 849	pET21b[1]	NdeI/NotI	His$_6$	low

9	variant9	Spt5 283 – 543	pET21b[1]	NdeI/NotI	His$_6$	not tested
10	Spt4	Spt4 1 – 102	pET21b[1]	NdeI/NotI	His$_6$	good n. d.[2]
11	complex	Spt5 283 – 849 Spt4 1 – 102	pET24d[3]	NdeI/NotI NheI/EcoR I	His$_6$ GST	good n. d.[2]
12	bicistron	Spt5 283 – 849 Spt4 1 – 102	pET21b[1]	NheI/NotI	His$_6$	1 mg/ml
13	Rpb4	Rpb4 1 – 221	pET21d[3]	BamHI/NcoI	no tag	not tested
14	fusion15	Spt4 1 – 102 Rpb7 1 – 107	pET21b[1]	NdeI/NotI	His$_6$	Good n. d.[2]
15	fusion13	Spt4 1 – 102 Rpb7 1 – 107	pET21b[1]	NdeI/NotI	His$_6$	10 mg/ml
16	SRI_X1	SRI 543 – 733	pET21b[1]	NdeI/NotI	His$_6$	43 mg/ml
17	SRI_X2	SRI 581 – 733	pET21b[1]	NdeI/NotI	His$_6$	not tested
18	SRI_X3	SRI 618 – 733	pET21b[1]	NdeI/NotI	His$_6$	71 mg/ml
19	SRI_X4	SRI 543 – 719	pET21b[1]	NdeI/NotI	His$_6$	not tested
20	SRI_X5	SRI 581 – 719	pET21b[1]	NdeI/NotI	His$_6$	24 mg/ml
21	SRI	SRI 619 – 720	pET9d[4]	NcoI/Acc65I	His$_6$	3 – 7 mg/ml
22	SRI M	SRI 624 – 705	pET24d[3]	Acc65I/NotI	His$_6$	not tested
23	SRI N	aa 624 – 707	pET24d[3]	Acc65I/NotI	His$_6$	6 mg/ml
24	SRI O	aa 624 – 709	pET24d[3]	Acc65I/NotI	His$_6$	40 mg/ml
25	SRI A	aa 629 – 705	pET24d[3]	Acc65I/NotI	His$_6$	10.5 mg/ml
26	SRI B	aa 629 – 707	pET24d[3]	Acc65I/NotI	His$_6$	31 mg/ml
27	SRI C	aa 629 – 709	pET24d[3]	Acc65I/NotI	His$_6$	12.8 mg/ml
28	SRI D	aa 632 – 705	pET24d[3]	Acc65I/NotI	His$_6$	not tested
29	SRI E	aa 632 – 707	pET24d[3]	Acc65I/NotI	His$_6$	13.3 mg/ml
30	SRI F	aa 632 – 709	pET24d[3]	Acc65I/NotI	His$_6$	11.5 mg/ml

[1] pET21b: ampicillin resistance

[2] not determined by Bradford

[3] pET24d and pET21d: kanamycin resistance

[4] pET9d: kanamycin resistance; His$_6$ – ztag – TEVsite – protein sequence

7.3 Media[1]

Luria Bertani medium	
1 %	tryptone
0.5 %	yeast extract
0.5 %	NaCl

Medium A (per liter)[1]	
100 ml	M9 medium (10x)
10 ml	trace elements solution (100x)
20 ml	20 % (w/v) glucose **or** 10 % (w/v) C[13] glucose
1 ml	1 M $MgSO_4$
0.3 ml	1 M $CaCl_2$
1 ml	biotin (1mg/ml)
1 ml	thiamin (1mg/ml)
	appropriate antibiotic(s)

M9 medium (10x) (per liter)[1]	
60 g	Na_2HPO_4
30 g	KH_2PO_4
5 g	NaCl
5 g	[15]NH_4Cl

Trace elements solution (100x) (per liter)[1]	
5 g	EDTA
0.83 g	$FeCl_3$ x 6 H2O
84 mg	$ZnCl_2$
13 mg	$CuCl_2$ x 2 H_2O

10 mg	$CoCl_2$ x 6 H_2O
10 mg	H_3BO_3
1.6 mg	$MnCl_2$ x 6 H_2O

Stock solutions[1]	
1 M	$CaCl_2$ (autoclaved)
1 M	$MgSO_4$ (autoclaved)
20 % (w/v)	Glucose (sterilized)
1 mg/ml	biotin (filter sterilized)
1 mg/ml	thiamin (filter sterilized)

Supplements/Antibiotics	
100 mg/ml in H_2O	ampicillin
50 mg/ml in H_2O	kanamycin
50 mg/ml in H_2O	chloramphenicol
1 M in H_2O	IPTG

100 X proteaseinhibitor mix/ethanol	
3 mg/l	leupeptin
14 mg/l	pepstatin A
1.7 g/l	PMSF
3.3 g/l	benzamidine

Edman buffer	
200 mM	Tris-HCl, pH 8.5
2 %	SDS

8 MICROBIOLOGICAL TECHNIQUES FOR EXPRESSION AND ANALYSIS OF RECOMBINANT PROTEINS

8.1 Transformation

A 50 µl aliquot of competent cells and plasmid DNA were thawed on ice. An amount of either 1 µl of plasmid DNA or 2 µl of PCR product were added to XL-1 blue or DH5α cells, respectively, and then incubated on ice for 20 minutes. Adjacent a 'heat shock' was performed at 42 °C for 60 seconds and the cells were again incubated on ice for 10 minutes. The cells were finally plated on LB agar plates containing either ampicillin (100 mg/ml) or kanamycin (50 mg/ml). The plates were incubated overnight at 37 °C.

In the case of coexpression of two separate plasmids harbouring particular genes this protocol was also executed and the appropriate antibiotics considered. For isotopic labelling BL21 (DE3) pLysS cells were used for transformation and plated on LB agar plates containing chloramphenicol (50 mg/ml) and kanamycin (50 mg/ml).

8.2 Gene expression in LB medium

A fresh colony was picked within 24 h of transformation, and a starter culture was grown to late-log phase in 50 ml LB medium containing ampicillin or kanamycin. Two liter of LB medium supplemented with the appropriate antibiotics were inoculated and shaken (180 rpm) at 37 °C until the culture reached log phase (OD 0.6 - 0.8). After cooling the *E. coli* suspension on ice IPTG at a 1 : 2000 ratio was added and the cells were grown over night at 18 °C. Cells were collected by centrifugation (5000 rpm, SLS6000 rotor) at 4 °C, subsequently suspended in lysis buffer (300 mM NaCl, 50 mM Tris pH 8.0, 5 % glycerol, 10 mM β-mercaptoethanol and 1 : 100 proteaseinhibitor mix) and frozen in liquid nitrogen before storage at -80 °C.

8.3 Gene expression in minimal medium

A fresh colony was picked within 24 h of transformation, and a starter culture was grown overnight in 20 ml medium A containing kanamycin (ratio 1 : 800) and

chloramphenicol (ratio 1 : 3333). This culture served as inoculum for two liter of medium A containing only kanamycin (1 : 1000) as antibiotic. To ensure better aeration only half a liter was filled in a 5 liter flask. The culture was shaken (200 rpm) at 37 °C until the OD 600 was 0.6. After cooling on ice IPTG at a 1 : 2000 ratio was added to induce over expression of the target protein. The cells were grown for 16 hours at 18 °C. Cells were collected by centrifugation (5000 rpm, SLS6000 rotor) at 4 °C, subsequently suspended in lysis buffer (300 mM NaCl, 50 mM Tris pH 8.0, 5 % glycerol, 10 mM β-mercaptoethanol, 1 : 100 proteaseinhibitor mix and frozen in liquid nitrogen before storage at -80 °C.

8.4 Preparation of cleared *E. coli* lysates

The first two steps in all protein isolation processes were cell lysis followed by clarification of the lysate to remove cell debris and nucleic acids. To release proteins from *E. coli* cells the frozen pellets were thawed in a 20 °C water bath and transferred to a glass beaker so that it is about half full. Samples were lysed on ice by sonication for 15 minutes with a Sonifier 450 rom Branson Utrasonics (Danbury, CT) using a duty cycle of 40 % and an output control setting of 4. Lysate clarification was achieved by centrifugation at 15 000 rpm for 20 minutes and the supernatant was collected for downstream steps of protein purification.

8.5 Affinity chromatography

In order to isolate a target protein from complex mixtures several chromatography techniques were applied. Practical aspects of performing a separation of each protein will be covered in the particular chapter, but an affinity chromatography step conducted each purification. The lysate, after clarification by centrifugation, was loaded on a self assembled Ni-NTA column (Quiagen) to retain complexes/proteins containing the His_6-tag.

8.6 Ion exchange chromatography

The separation power of ion exchange chromatography is based on the fact that the relationship between net surface charge and pH is unique for a specific protein. According the results from ProtParam the appropriate Mono S HR 5/5 or Mono Q HR

5/5 (Amersham Biosciences) was chosen to separate molecules on the basis of differences in their net surface charge. The applied salt gradient ranged from 50 mM to 1 M NaCl or $(NH_4)_2SO_4$.

8.7   Gel filtration

Gel filtration separates molecules corresponding to differences in size as they pass through a gel filtration medium packed in a column. For purification purposes Superose6, Superose12 and Superdex75 columns (Amersham Biosciences) were utilised according to the instruction manuals.

8.8   Limited proteolysis experiments

A limited proteolysis approach and Edman sequencing were applied in order to determine the surface accessibility of conserved domains. Limited proteolysis of natively folded proteins occurs at flexible sites and therefore proteolytic probes can be used to pinpoint the sites of local unfolding in a protein chain. For trypsin treatment 1 µg of the protease was added to 20 µg to 50 µg of purified protein. Digests were done in the buffers used for gel filtration and supplemented with $CaCl_2$ to a final concentration of 4 µM. The mixture was incubated at 37 °C and aliquots were removed at 1, 3, 10, 30 and 60 minutes. The reactions were stopped by the addition of SDS sample buffer and were heated immediately to 95 °C for 5 min. All samples were analyzed by SDS-PAGE.

8.9   Protein separation by SDS-PAGE

Denaturing gel electrophoresis was adopted to separate complex protein mixtures into distinct bands on a gel. According to the discontinuous Laemmli system (Sambrook and Russel, 2001), ten gels were cast at once. The percentage of the gels (12 % -17 %) was defined by the size of the monitored proteins. The proteins were totally unfolded by adding β-mercaptoethanol to the SDS loading dye. Gels were then stained with Coomassie (SIGMA) solution and if required subjected to blotting procedures.

8.10   Blotting and Edman Sequencing

For N-terminal sequencing, proteins in a gel cannot be sequenced but must be transferred out of the gel. After electrophoresis, the desired band of interest was excised and dried in a Speed Vac. Following, the gel piece was reswollen in 50 µl of Edman buffer. After setting up a gradient by addition of 200 µl of distilled water, a small piece of pre wet PVDF membrane was added to adsorb the protein. Once the solution began to turn blue from dye, methanol was added to a final concentration of 10 % and the transfer was complete when the membrane was blue. The procedure required two days incubation at room temperature. Finally the membrane was washed with 10 % methanol, dried and the strip was loaded into a PROCISE 491 sequencer (Applied Biosystems).

8.11   Standard techniques

Standard techniques in molecular biology such as isolation of DNA, restriction analysis, polymerase chain reaction (PCR), cloning of DNA, agarose gel electrophoresis etc. were performed essentially as described in Sambrook and Russel, 2001.

9 BIOINFORMATIC TOOLS AND SOFTWARE

Alignments Multiple homologues sequence alignments were initially constructed using CLUSTAL-W (www.ebi.ac.uk/clustalw). The program ESPript (Easy Sequencing in PostScript - espript.ibcp.fr) was applied for the rapid visualization, via PostScript output, of the sequences aligned with CLUSTAL-W.

Secondary structure prediction. Secondary structure prediction was done using the PredictProtein (http://cubic/bioc/columbia/edu/predictprotein.hmtl) secondary structure prediction surfer.

Calculation of molecular weight, absorption coefficient and PI. Calculation of properties of the proteins which are important for the design of the purification strategy as the PI were determined using ProtParam. The absorption coefficients and molecular weights used for quantification were obtained form the same server (www.expasy.org/tools/protparam.html).

Comparing protein structures. The Dali server (http://www.ebi.ac.uk/dali/) is a network service for comparing protein structures in 3D. By submitting the coordinates of a query protein structure to the server, Dali compares them against those in the Protein Data Bank.

Figures. Most oft the molecular figures of RNAP II and the SRI domain were created using the program pymol (http://pymol.sourceforge.net/). Figure 26 is designed with MOLMOL (http://www.mol.biol.ethz.ch/groups/wuthrich_group/software).

10 RECOMBINANT SPT4-SPT5 PROTEINS AND ASSEMBLY OF THE ELONGATION CHECKPOINT COMPLEX – VARIOUS APPROACHES

10.1 Single Spt5 variants

10.1.1 Design and expression of different Spt5 variants

For subcloning the nucleotide sequence of the different Spt5 protein variants into a bacterial expression vector, the forward primer were designed to create an NdeI site and to add the methionine codon before the first residue and the reverse primer were designed to create a NotI site without a stop codon. The desired nucleotide fragment was amplified by PCR using the appropriate oligonucleotides and the genomic *S. cerevisiae* DNA as a template, digested with NdeI and NotI, and subcloned into the NdeI and NotI restriction sites of pET21b (Stratagene) to construct the C-terminal six histidine (His_6)-tagged expression plasmid. Corresponding to the NotI restriction site an AAALE linker between the Spt5 protein and the His_6-tag was introduced. Transformation and expression of the different variants were carried out as mentioned. The cells were lysed by sonication, centrifuged, and the supernatant was loaded onto a Ni-NTA column (Quiagen), equilibrated with lysis buffer. The sample was eluted by a stepwise gradient from 10, 20, 50 to 200 mM imidazole. The polypeptide compositions of the column fractions were monitored by SDS-PAGE. The recombinant His_6-tagged polypeptides were recovered predominantly in the 50 mM or 200 mM imidazole fractions. Only a few of the expressed Spt5 variants could be further purified and yielded a reasonable amount of protein. Either they were prone to degradation or the expression level was low. Spt5 variant2, variant4 or variant6 were promising (see Diploma-thesis Vojnic, 2002).

At best the 50 or 200 mM fractions were pooled and applied onto a MonoS column (HR 5/5) aquilibrated 50 mM $(NH_4)_2SO_4$ (Table 5). The peak fractions were concentrated with an Amicon Ultracentrifuge device (cut off 10 kDa) and passed over a Superose6 gel filtration column. The final fractions were used for crystallization set-ups.

Table 5: Purification protocol of single Spt5 variants

Step	Buffer composition
Lysis buffer	300 mM NaCl, 50 mM Tris-HCl pH 8, 5 % glycerin, 10 mM β-mercaptoethanol, 1/100 proteaseinhibitor mix
Wash solution	2 M NaCl
Wash buffer	150 mM NaCl, 50 mM Tris-HCl pH 8, 5 % glycerin, 50 mM imidazole, 10 mM β-mercaptoethanol
Elution buffer	150 mM NaCl, 50 mM Tris-HCl pH 8, 5 % glycerin, 200 mM imidazole, 10 mM β-mercaptoethanol
MonoS	50 mM Hepes pH 7, 50 mM to 1 M $(NH_4)_2SO_4$
Superose 6	50 mM Hepes pH 7, 150 mM $(NH_4)_2SO_4$

10.2   Bicistronic Spt4-Spt5 variant3 pair

10.2.1 Design and expression of bicistronic Spt4-Spt5 variant3

The bicistronic Spt4-Spt5 variant3 pair was expressed from a pET21b vector. In this case, Spt4 was expressed as untagged full length protein and the adjacent Spt5 variant was expressed as C-terminal His_6-tagged fusion protein.

To create this vector the gene sequences for the proteins were initially subcloned in another vector, kindly provided by S. Baumli. Spt4 was cloned via the restriction sites NheI and EcoRI into a pET24b derived vector containing an N-terminal GST-TEV coding sequence in front of the multiple cloning site. A second ribosomal-binding and multiple cloning site was introduced as described elsewhere (PhD-thesis Baumli, 2005). The PCR Fragment of variant3 was inserted into the second ORF via its NdeI and NotI restriction sites and allowed the expression of a His_6 tagged variant of Spt5. This protein pair was unfortunately not straightforward to purify and hence for convenient handling the cassette containing the Spt4-Spt5 variant3 pair was cut with NheI and NotI restriction enzymes and subcloned into an ampicillin resistant pET21b vector. The basic principles pertaining to the cloning procedure are outlined in Figure 31. The microbiological techniques for expression of recombinant Spt4-Spt5 variant3 proteins in *E.coli* were performed as described in chapter 8.
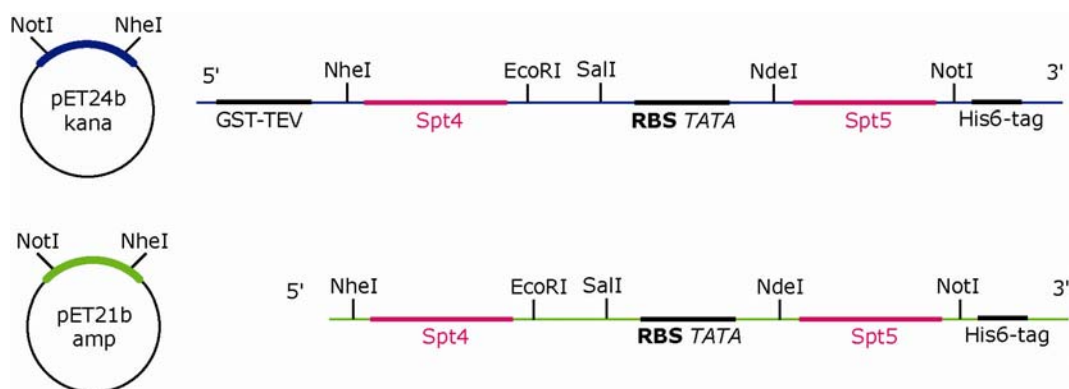
Figure 31. **Flowchart of cloning procedure.**

10.2.2 Purification of Spt4–Spt5 variant3

Thawed cells were lysed by sonication and centrifuged, and the supernatant was loaded onto a Ni-NTA column (Quiagen), which was equilibrated with lysis buffer. The sample was eluted stepwise from 20 to 50 mM imidazole and monitored by SDS-PAGE. The presence of Spt4 in the heterodimeric complex was verified by Edman Sequencing. Due to the aberrant retention volume of the complex on a size exclusion column and the associated loss of material, I performed the buffer exchange via a PD10 column. The column was equilibrated with the appropriate buffer containing 5 mM Hepes pH 7.5, 100 mM $NH_4(SO_4)_2$, 10 µm $ZnCl_2$, 5 mM DTT (Table 6). The protein amount was detected with Bradford Reagenz (biorad).

Table 6: Purification protocol of Spt4-Spt5 variant3 complex

Step	Buffer composition
Lysis buffer	300 mM NaCl, 50 mM Tris-HCl pH 8, 5 % glycerin, 10 mM β-mercaptoethanol, 1/100 proteaseinhibitor mix
Wash buffer	300 mM NaCl, 50 mM Tris-HCl pH 8, 5 % glycerin, 10 mM β-mercaptoethanol, 20 mM imidazole
Wash solution	2 M NaCl
Elution buffer	300 mM NaCl, 50 mM Tris-HCl pH 8, 5 % glycerin, 10 mM β-mercaptoethanol, 50 mM imidazole
PD10 buffer	5 mM Hepes pH 7.5, 100 mM $NH_4(SO_4)_2$, 10 µm $ZnCl_2$, 5 mM DTT

10.2.3 Assembly of RNAP II with Spt4-Spt5 variant3

For assembly of an 'elongation checkpoint complex' consisting of the complete RNAP II and Spt4/Spt5 variant3 the procedure was as follows. Reconstitution of 12-subunit RNAP II was initiated by thawing the ammonium sulphate pellet of the core polymerase at 4 °C and resuspending it in 150 μl PD10 buffer. Rpb4/Rpb7 were obtained frozen in RNAP II buffer (5 mM Hepes pH 7.25; 40 mM $(NH_4)_2SO_4$; 100 μm $ZnCl_2$, 10 mM DTT), thawed, added to the core polymerase and incubated on a rotating wheel at 20 °C for 20 minutes. Subsequent purified Spt4-Spt5 variant3 was combined with RNAP II and incubated at 4 °C for 1 hour. Both 10 fold excess of Rpb4/Rpb7 and Spt4-Spt5 variant3 were used. Meanwhile Superose6 was equilibrated with PD10 buffer and the peak fractions of the reconstitution reaction were monitored by 15 % SDS-PAGE.

10.3 Rpb7/Spt4 fusion protein

10.3.1 Design and expression of an artificial Rpb7/Spt4-Rpb4 complex

According to the proposed homology between archeal subunit E" and Rpb7 from RNAP II and Spt4, a chimeric DNA comprising both sequences connected by a multiple glycine linker was introduced into a pET21b vector. This vector simultaneously permitted the bicistronic expression of Rpb4.

Plasmids for expression of (full length) Rpb7-glycinelinker-Spt4 (full length) and Rpb4 (designated fusion protein complex) were constructed as follows. The Rpb7 ORF was PCR amplified from genomic *S. cerevisiae* DNA as a template with an N-terminal primer that added codons for an NdeI site (designated f1) and one C-terminal primer that comprised the C-terminal region of Rpb7, codons for a 15G-linker, and the N-terminal region of Spt4. Simultaneously, in a second PCR Spt4 was amplified. An N-terminal primer was used, that added codons for the C-terminus of Rpb7 to the 15G-linker and the Spt4 N-terminus (matching the DNA sequence added to Rpb7, above). The C-terminal primer appended codons for a NotI site (designated r2). The fusion was constructed by combining Rpb7-G15-Spt4n with Rpb7c-G15-Spt4 PCR products for amplification with the N-terminal Rpb7 f1 primer and C-terminal Spt4 primer r2 followed by insertion as an NdeI-NotI

fragment into pET21b (adding a C-terminal hexahistidine tag to Spt4). The fusion generates a G15 linker between Rpb7 and Spt4, but due to unknown mechanisms another plasmid with a 13 glycine linker was also generated by accident. In parallel, the Rpb4 ORF was PCR amplified from genomic *S. cerevisiae* DNA as a template with an N-terminal primer that adds an N-terminal M codon and an NcoI site and a C-terminal primer that generates a BamHI site followed by insertion as an NcoI-BamHI fragment into pET21d. This cassette was cut by BamHI-SphI restriction enzymes and the resulting BamHI-SphI fragment was inserted into the BglII-SphI site of pET21b (Rpb7-G15-Spt4) to generate pET21b (Rpb7-G15-Spt4/Rpb4). The heterologous protein expression occurred via two separate T7 promoters. For the sake of clarity, the primer sequences are shown in Table 7. The restriction sites are displayed in bold and the codons for the glycine linker are underlined. The basic principles pertaining to the cloning procedure are outlined in Figure 32.

Table 7: Primer sequences

Primer	Sequence
f1_NdeI	GGGGGGGGGG**CATATG**TTTTTTATTAAAGACCTTTCG
r1	CATACAGGCTCTTTCACTAGACATACCGCCACCGCC<u>ACCGCCACCGCCAC</u> <u>CGCCACCGCCACCGCCACC</u>AATAGCACCCAAATA ATCTTC
f2	GAAGATTATTTGGGTGCTATT<u>GGTGGCGGTGGCGGTGGCGGTGGCGGT</u> <u>GGCGGTGGCGGTGGCGGT</u>ATGTCTAGTGAAAGAGCCTGTATG
r2_NotI	GGGGGGGGGG**GCGGCCGC**CTCAACTTGACTGCCATCCCTCGG
f3_NcoI	AGGTG**CCATGG**CGAATGTTTCTACATCAACC
r3_BamHI	CGC**GGATCC**CTAATAGAGTGTTTCTAGGTTTGAC
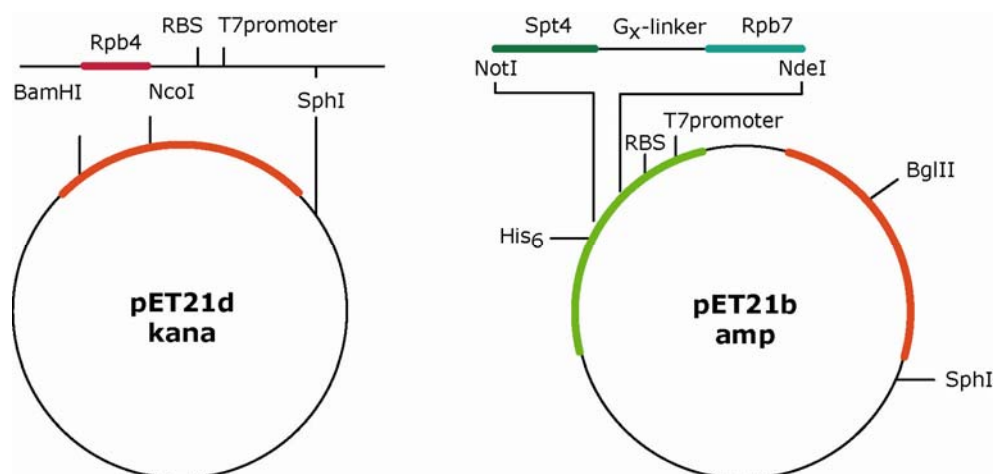
Figure 32. **Cloning procedure (x = linker containing either 13 or 15 glycines).**

10.3.2 Purification of an artificial Rpb7/Spt4-Rpb4 complex

The plasmids containing either fusion protein with a 13- or a 15-glycine linker were transformed in BL21(DE3) RIL (Stratagene) cells and expressed as described in chapter 8. Cells were disrupted by sonication and the lysate, after clarification by centrifugation at 15000 rpm, was loaded onto a one millilitre self assembled Ni-NTA agarose column (Quiagen). The His_6-tagged Rpb7/Spt4 artificial complex should be retained onto the column and constitute a stochiometric subcomplex with Rpb4. In order to get rid of sticky DNA from the expression strain, the resin was washed with 5 column volumes of 2 M NaCl. A second wash step included lysis buffer supplemented with 50 mM imidazol. The proteins were eluted with 5 ml lysis buffer containing 200 mM imidazol. Conditions for optimal purification steps were determined empirically. The eluted proteins were further purified by Mono S chromatography (5/5 Amersham). The column was pre-equilibrated with Mono S buffer containing 50 mM NaCl. A linear gradient was applied and the eluted proteins collected and concentrated by Amicon Ultra centrifugal devices (10 KDa molecular weight cut off). As a last purification step gel filtration was performed to exchange the protein into the running buffer and to separate last impurities from the desired proteins (Table 8). As judged by Coomassie staining, the stoichiometry of both recombinant proteins was given.

Table 8: Purification protocol of an artifical
Rpb7/Spt4 and Rpb4 complex

Step	Buffer composition
Lysis buffer	150 mM NaCl, 50 mM Tris-HCl pH 7.5, 10 µm $ZnCl_2$, 5 % glycerin, 10 mM β-mercaptoethanol, 1/100 proteaseinhibitor mix
Wash solution	2 M NaCl
Wash buffer	150 mM NaCl, 50 mM Tris-HCl pH 7.5, 5 % glycerin, 10 mM β-mercaptoethanol, 50 mM imidazole
Elution buffer	150 mM NaCl, 50 mM Tris-HCl pH 7.5, 5 % glycerin, 10 mM β-mercaptoethanol, 200 mM imidazole
MonoS	50 mM Hepes pH 7.5, 50 mM to 1 M NaCl, 5 mM DTT
Superose6	50 mM Hepes pH 7.5, 150 mM NaCl, 5 mM DTT

10.3.3 Assembly of complex12

Reconstitution of twelve-subunit RNAP II, which contained an artificial Rpb7/Spt4 subunit homolog to archeal E/E", was performed at 4 °C. One ammonium sulphate pellet of endogenous ten-subunit RNAP II was thawed, the supernatant discarded, and the remaining pellet resuspended in 450 µl of Superose6 buffer. Addition of RNAP II to 5-fold excess of fusion protein occurred stepwise. Adjusting the volume to 500 µl was done prior to Superos6 gel filtration. The collected peak fractions were monitored by 15 % SDS-PAGE.

10.4   Rpb7/Spt4 and Rpb4 complex plus Spt5 variant1

10.4.1 Design and expression of complexes containing different Spt5 variants

The expression of all heterologues proteins was achieved by co-transformation of the plasmids harboring the components for establishing the elongation checkpoint complex. Initially, variant1, variant3 and variant5 were produced in *E. coli* as C-terminal His_6-tagged proteins simultaneously with the artificial fusion protein complex. Therefore the existing DNA sequences were cut out from the pET21b

vector by means of their NdeI/NotI restriction sites and ligated into a modified pET24d kanamycin resistant vector. The NcoI site was previously replaced by an NdeI site. The expression procedure was performed essentially as described earlier.

10.4.2 Purification of Rpb7/Spt4 and Rpb4 complex plus Spt5 variant1

Co-expression of Spt4/Rpb7/Rpb4 together with Spt5 variant1 seemed most promising concerning the stoichiometry of all components. The proteins were step-eluted with 200 mM imidazole in lysis buffer (Table 9). The pooled fractions were applied onto a Mono Q column and eluted by a linear gradient, starting with 150 mM NaCl. The peak fractions were visualised by Coomassie-staining and the stochiometric samples concentrated by Amicon Ultra centrifugal devices (100 KDa molecular weight cut off). For further experiments concentrated protein was stored as ammonium sulphate pellet. Hence, purified protein was precipitated after Mono Q with saturated ammonium sulphate, frozen in liquid nitrogen and stored at -80 °C.

**Table 9: Purification protocol of a complex comprising
Rpb7/Spt4 and Rpb4 plus Spt5 variant1**

Step	Buffer composition
Lysis buffer	150 mM NaCl, 50 mM Tris-HCl pH 7.5, 10 μm $ZnCl_2$, 5 % glycerin, 10 mM β-mercaptoethanol, 1/100 proteaseinhibitor mix
Wash solution	2 M NaCl
Wash buffer	150 mM NaCl, 50 mM Tris-HCl pH 7.5, 5 % glycerin, 10 mM β-mercaptoethanol, 50 mM imidazole
Elution buffer	150 mM NaCl, 50 mM Tris-HCl pH 7.5, 5 % glycerin, 10 mM β-mercaptoethanol, 200 mM imidazole
Mono Q	50 mM Bicine pH 8.5 @ 10 °C, 150 to 1 M NaCl, 5 mM DTT

10.4.3 Assembly of complex13

The RNAP II-Spt4/Spt5 complex was assembled by incubating core RNAP II (resuspended in 50 μl of 150 mM ammonium sulphate, 50 mM Hepes pH 7.5, 10

mM DTT) for 10 minutes at 4 °C with a 1.5 molar excess of Spt4/Rpb7/Rpb4-Spt5 variant1, on hand in Mono Q buffer. This complex was transferred to a MembraSpin centrifugal concentrator (100 kDa cutt off) and the buffer adjusted to 150 mM ammonium sulfphate, 50 mM Hepes pH 7.5 and 10 mM DTT. The buffer exchange was achieved by diluting the concentrated sample several times with altogether two milliliter of the desired buffer and monitoring the conductivity.

10.4.4 Assembly of 'elongation checkpoint complex'

A DNA/RNA hybrid was annealed by mixing equimolar amounts of synthetic template DNA, non-template DNA and RNA oligonucleotides (biomers.net) in RNAse free TE buffer (Fluka) at a final concentration of 100 μM. The mixture was heated to 90 °C in a thermoblock (preheated) and slowly cooled to 20 °C over night. All relevant steps for the assembly were done at 4 °C. Core RNAP II was resuspended in 50 μl of 150 mM ammonium sulphate, 50 mM Hepes pH 7.5, 10 mM DTT buffer and incubated with 100 pmol DNA/RNA hybrid for 15 minutes. An ammonium sulphate pellet of Rpb7/Spt4 and Rpb4 complex plus Spt5 variant1 was thawed, dissolved in assembly buffer (200 μl of 150 mM ammonium sulphate, 50 mM Hepes pH 7.5, 10 mM DTT) and followed by the incremental addition of RNAP II-DNA/RNA complex. After ten minutes of incubation the assembly reaction was transferred to a MembraSpin centrifugal concentrator (100 kDa cut off) and the buffer adjusted to 150 mM ammonium sulfphate, 50 mM Hepes pH 7.5 and 10 mM DTT. The procedure was like the assembly of RNAP II-Spt4/Spt5 complex. Prior to the last cryo step, an additional amount of DNA/RNA hybrid was added to a final concentration of 2 μM in order to ensure full occupancy.

11 RECOMBINANT SRI DOMAIN VARIANTS AND PHOSPHO-PEPTIDE STUDIES

11.1 Design and cloning of different SRI domain variants

For following crystallization set-ups DNA fragments comprising varying parts of the SRI domain were amplified by PCR and cloned into the NcoI/NotI sites of a pET24d (G. Stier/EMBL) vector containing an N-terminal His$_6$-tag (as indicated in chapter 7.2). For NMR spectroscopy the DNA fragment encoding for residues 620 – 719 was amplified by PCR and cloned via its NcoI/Acc65I sites into a modified pET9d vector (G. Stier/EMBL) that contained a TEV protease site directly after the N-terminal z-tag and His$_6$-tag.

11.2 Purification of SRI domain variants

11.2.1 Purification of pET24d expressed SRI domain variants for crystallization

The cell lysate was applied to a His-select nickel column (Quiagen) and each protein eluted in 200 mM imidazole. Fractions containing the desired protein were pooled, diluted in the same volume of MonoS buffer without salt and applied onto a MonoS (Amersham) column. The final purification step comprised a Superdex75 gel filtration and the peak fractions were concentrated for crystallization (Table 10).

Table 10: Purification protocol for crystallization set-ups

Step	Buffer composition
Lysis buffer	300 mM NaCl, 50 mM Tris-HCl pH 7.5, 5 % glycerin, 10 mM β-mercaptoethanol, 1/100 proteaseinhibitor mix
Wash solution	2 M NaCl
Wash buffer	300 mM NaCl, 50 mM Tris-HCl pH 7.5, 5 % glycerin, 10 mM β-mercaptoethanol, 50 mM imidazole
Elution buffer	300 mM NaCl, 50 mM Tris-HCl pH 7.5, 5 % glycerin, 10 mM β-mercaptoethanol, 200 mM imidazole
Mono S	30 mM Mes pH 6.5, 50 mM to 1 M NaCl
Superdex75	20 mM Mes pH 6.5, 175 NaCl, 2 mM DTT

11.2.2 Purification of pET9d expressed SRI domain for NMR spectroscopy

Cell lysates were subjected to affinity chromatography on a first Ni-NTA column (Quiagen), followed by cleavage of the hexahistidine tag with TEV protease and dialysis overnight. The tag and the His_6-tagged protease were removed on a second Ni-NTA column. The pooled fractions were diluted carefully with MonoS buffer (without salt) until the conductivity was around 150 µS/cm. DNA was removed by cation exchange chromatography (MonoS, Amersham). After gel filtration the sample was dissolved in buffer containing 20 mM sodium phosphate pH 6.5, 200 mM NaCl and 2 mM DTT (Table 11). Edman sequencing of the protein confirmed the presence of four additional residues (GAMG) at the N-terminus, which resulted from the cloning strategy.

Table 11: Purification protocol for NMR spectroscopy

Step	Buffer composition
Lysis buffer I	300 mM NaCl, 50 mM Tris-HCl pH 8, 5 % glycerin, 10 mM β-mercaptoethanol, 1/100 proteaseinhibitor mix
Wash solution I	2 M NaCl
Wash buffer I	300 mM NaCl, 50 mM Tris-HCl pH 8, 5 % glycerin, 10 mM β-mercaptoethanol, 20 mM imidazole
Elution buffer I	300 mM NaCl, 50 mM Tris-HCl pH 8, 5 % glycerin, 10 mM β-mercaptoethanol, 200 mM imidazole
Dialysis buffer	500 mM NaCl, 50 mM Tris-HCl pH 8, 2 mM DTT
Lysis buffer II	500 mM NaCl, 50 mM Tris-HCl pH 8, 10 mM β-mercaptoethanol,
Elution buffer II	500 mM NaCl, 50 mM Tris-HCl pH 8, 30 mM imidazole, 10 mM β-mercaptoethanol,
Mono S	30 mM Mes pH 6.5, 50 mM to 1 M NaCl, 2 mM DTT
Superdex75	20 mM sodium phosphate pH 6.5, 200 mM NaCl, 2 mM DTT

11.3   Phosphopeptide interaction studies

11.3.1 Crystallization set-ups

The phospho-CTD peptides used for binding experiments were chemically synthesized [three-repeat peptide, SPS-YEPTEPS-YEPTEPS-YEPTEPS, E= glutamate mimics phosophoserine (http://www.jerini.com/); two-repeat peptide, SPS-YSPTpSPS-YpSPTpSPS, pS=phosphoserine (http://www.anaspec.com/)]. For crystallization set-ups 1.25-fold molar excess of peptide was co-crystallized with the appropriate SRI domain protein.

11.3.2 NMR-titration

The phospho-CTD peptides used for binding experiments were chemically synthesized [one-repeat peptide, YpSPTpSPS (G. J. Arnold; gene center); two-repeat peptide, SPS-YpSPTpSPS-YpSPTpSPS, pS=phosphoserine (http://www.anaspec.com/)]. For NMR titration, increasing amounts of the CTD peptide were added to a 0.4 mM solution of $^{15}N,^{13}C$-labeled SRI domain up to a 1.25-fold molar excess. Chemical shifts were monitored in 2D $^1H,^{15}N$ HSQC experiments. The lyophylized, synthetic equivalents of the CTD were dissolved 20 mM sodium phosphate pH 6.5, 200 mM NaCl, 2 mM DTT.

12   STRUCTURE DETERMINATION BY SOLUTION NMR

12.1   NMR data acquisition

NMR spectra were acquired at 292 K on Bruker DRX500, DRX600, or DRX900 spectrometers with cryogenic triple resonance probes. Spectra were processed with NMRPipe (Delaglio *et al.*, 1995) and analyzed using NMRVIEW (Johnson and Blevins, 1994).

12.2   Backbone assignment of chemical shifts

A combined set of heteronuclear multidimensional NMR experiments were recorded for the assignment of the ^{1}H, ^{13}C, and ^{15}N chemical shifts of uniformly labelled protein. The resonance assignment strategy involved the concerted use of four 3D triple-resonance experiments [HN(CO)CA, HNCA, CBCA(CO)NH and ^{13}C-resolved three-dimensional NOESY]. The central feature of this policy was the concurrent assignment of both backbone and side-chain aliphatic atoms, which was critical for overcoming ambiguities in the assignment process. The combination of the 3D experiments HNCA and HN(CO)CA was used to establish backbone sequential connectivities by connecting the resonance frequencies of spins with those of preceding residues. The CBCA(CO)NH experiment was used to extend the connectivities from the backbone to Cβ as the chemical shifts of the side-chain carbons are characteristic for the amino acid type. This information can be utilized to position sequentially connected fragments within the amino acid sequence (Sattler *et al.*, 1999).

12.3   Structure calculation and determination

Distance restraints were derived from 2D NOESY (Nuclear Overhauser effect spectroscopy) and ^{15}N- or ^{13}C-resolved 3D NOESY. Restraints for the backbone dihedral angles Φ and Ψ were derived from TALOS (Cornilescu *et al.*, 1999). The protection of amide protons against chemical exchange was identified from ^{1}H/^{15}N correlation experiments after dissolving of lyophilized protein in 100 % D$_2$O. Analyzing ^{15}N relaxation parameters is a powerful means of characterizing protein backbone dynamics and therefore three sets of relaxation measurements (NOE, T1

and T2) were performed. ^{15}N relaxation (T1, T2) and heteronuclear ^{1}H-^{15}N NOE was measured on a ^{15}N-labelled protein sample at 292 K (Farrow *et al.*, 1994). The experimentally determined distance and dihedral restraints (Table 4, Figure 26 C) were applied in a simulated-annealing protocol using ARIA (Linge *et al.*, 2001) and CNS (Bruenger *et al.*, 1998). NOEs were manually assigned and distance calibrations were performed by ARIA. The final ensemble of NMR structures was refined in a shell of water molecules (Linge *et al.*, 2003). Structural quality was analyzed with PROCHECK (Laskowski *et al.*, 1996).

12.4   NMR titration experiment – ^{1}H-^{15}N HSQC

The ^{1}H-^{15}N HSQC (heteronuclear single quantum coherence) NMR experiment is a two dimensional experiment in which each amino acid residue (except for proline) in a given protein is described by one peak in the spectrum. The chemical shift of a peak at (ω1, ω2), where ω1 and ω2 are the amide ^{15}N and ^{1}H shifts, respectively, depends from the chemical surrounding of the amino acid amide group in the protein. In addition, the side chain amides of glutamine and asparagine are represented in the spectrum. A folded protein structure will generally produce a ^{1}H-^{15}N HSQC spectrum with a broad distribution of well separated signals. Changes in the environment of a spin due to binding of a ligand give rise to chemical shift changes in the NMR spectrum. These changes are expected to be largest near the binding site and the interface of a protein with a ligand can be easily mapped.

12.5   Isotope filtering experiments

To understand the function of biological macromolecules, it is important to illuminate the molecular crosstalk between these molecules. For studying the structures of molecular complexes by NMR spectroscopy, it is essential to distinguish between intra- and intermolecular NOEs. This task can be achieved by heteronuclear filtered NOE experiments performed on a sample of a complex consisting of differentially labeled molecules. In the binary complex consisting of ^{13}C, ^{15}N-labeled SRI domain (A) and unlabeled peptide (B), four kinds of NOE cross peaks can be observed (Table 12).

Table 12: NOE cross peaks observed in complexes

intramolecular:	intermolecular:
A – **A:** **^{1}H-(^{13}C,^{15}N)** – **^{1}H-(^{13}C,^{15}N)**	A – B: **^{1}H-(^{13}C,^{15}N)** – ^{1}H-(^{12}C,^{14}N)
B – B: ^{1}H-(^{12}C,^{14}N) – ^{1}H-(^{12}C,^{14}N)	B – A: ^{1}H-(^{12}C,^{14}N) – **^{1}H-(^{13}C,^{15}N)**

The connectivity of protons to the heteronuclei ^{13}C and ^{15}N can therefore be used to separate intra- and intermolecular NOEs and the intramolecular NOEs of A and B (Sattler *et al.*, 1999). For that purpose, a 2D ω_2-^{13}C,^{15}N-filtered NOESY was acquired to examine especially the intermolecular NOEs between peptide and SRI domain.

12.6   TOCSY experiments

In general, the HCCH-TOCSY (Total correlation spectroscopy) experiment correlates all aliphatic ^{1}H and ^{13}C spins within residues, and is used to assign aliphatic ^{1}H and ^{13}C resonances and connect the side-chain chemicals shifts with the backbone assignment (Teng, 2005). Here, the scalar couplings observed in different TOCSY experiments (mixing time 30 or 60 ms) were used to correlate the spins within the spin system of the 'free' and unlabeled peptide. The measurements were either performed in the same buffer conditions as for the NMR-titration experiment (20 mM sodium phosphate pH 6.5, 200 mM NaCl, 2 mM DTT/2 mM peptide concentration) or in 100 % D_2O (0.4 mM peptide concentration).

12.7   ROESY experiments

ROSY (Rotating-frame Overhauser effect spectroscopy) is an experiment in which homonuclear NOE effects are measured under spin-locked conditions. Both ROESY and NOESY experiments utilize the dipolar interaction in the form of cross relaxation to correlate spins that are close in distance. In the case of small and medium size molecules, the NOESY experiment is limited as the NOE enhancement is close to zero. The ROESY experiment has been developed to overcome this problem (Bax and Davis, 1985). The pulse sequence in this experiment is similar to

TOCSY although the cross-peaks of a ROESY spectrum have an opposite phase to those in the TOCSY spectrum (Teng, 2005).

13   PROTEIN CRYSTALLOGRAPHIC METHODS

13.1   Crystallization and crystal freezing

Crystals of RNAP II in conjunction with varying Spt4-Spt5 proteins were obtained at 20 °C with the hanging drop vapor diffusion method using 24 well plates. The applied buffer solutions varied in the concentration of one component (see Table 1) and thereby narrowed the variation space of the crystallization parameters involved. The crystallization droplets contained the appropriate complex solution mixed with the well solution either in a 1 : 1 or 2 : 1 ratio. Crystals were obtained mostly after one week. For cryo-protection, crystals were stabilized in six steps by replacing the crystallization condition against mother solution containing an increasing amount of glycerin until the final concentration of 22 % was reached. The crystals were stored in the last cryo-solution in a styropor box over night at 8 °C, and then flash frozen in liquid nitrogen.

For crystallization set-ups of the SRI domain variants, I embarked on the established sparse matrix strategy, (Jancarik and Kim, 1991) using the commercially available screens mentioned in chapter 6.1 – Table 3. Sitting drops were performed by the crystallization robot Hydra I.

13.2   Data collection and structure determination

All diffraction data were collected with an increment of 0.5 degree per image at the beamline X06SA at the Swiss Light Source, Villigen, Switzerland and processed with DENZO and SCALEPACK (Otwinowski and Minor, 1996). Crystal structures were solved at 4.0 – 4.5 Å resolutions by molecular replacement with the program PHASER (Storoni *et al.*, 2004), using the complete RNAP II elongation complex (Protein Data Bank ID: 1Y1W) as a search model. The model was improved with iterative cycles of refinement with CNS (Bruenger *et al.*, 1998).

Chapter IV: Literature

Adhvaryu K. K., Morris S. A., Strahl B. D. and Selker E. U. (2005)
Methylation of histone H3 lysine 36 is required for normal development in *Neurospora crassa*. Eukaryot Cell, 4 (8), 1455 – 1464.

Ahn S. H., Kim M. and 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.

Allen M., Friedler A., Schon O. and Bycroft M. (2002)
The structure of an FF domain from human HYPA/FBP11. Journal of Molecular Biology, 323 (3), 411 – 416.

Allison L. A., Moyle M., Shales M. and Ingles C. J. (1985)
Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. Cell, 42 (2), 599 – 610.

Allison L. A., Wong J. K., Fitzpatrick V. D., Moyle M. and Ingles C. J. (1988)
The C-terminal domain of the largest subunit of RNA polymerase II of *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and mammals: a conserved structure with an essential function. Molecular and Cellular Biology, 8 (1), 321 – 329.

Aloy P., Bottcher B., Ceulemans H., Leutwein C., Mellwig C., Fischer S., Gavin A. C., Bork P., Superti-Furga G., Serrano L. and Russell R. B. (2004)
Structure-based assembly of protein complexes in yeast. Science, 303 (5666), 2026 – 2029.

Aloy P. and Russell R. B. (2006)
Structural systems biology: modelling protein interactions. Nature Reviews Molecular Cell Biology, 7 (3), 188 – 197, Review.

Amrani N., Minet M., Wyers F., Dufour M. E., Aggerbeck L. P. and Lacroute F. (1997)
PCF11 encodes a third protein component of yeast cleavage and polyadenylation factor I. Molecular and Cellular Biology, 17 (3), 1102 – 1109.

Andrulis E. D., Werner J., Nazarian A., Erdjument-Bromage H., Tempst P. and Lis J.T. (2002)
The RNA processing exosome is linked to elongating RNA polymerase II in Drosophila. Nature, 420 (6917), 837 – 841.

Armache K. J., Kettenberger H. and Cramer P. (2003)
Architecture of initiation-competent 12-subunit RNA polymerase II. Proceedings of the National Academy of Science, 100 (12), 6964 – 6968.

Bannister A. J. and Kouzarides T. (2005)

Reversing histone methylation. Nature, 436, 1103 – 1106.

Bannister A. J., Zegerman P., Partridge J. F., Miska E. A., Thomas J. O., Allshire R. C. and Kouzarides T. (2001)

Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature, 410 (6824), 120 – 124.

Bannister A. J., Schneider R. and Kouzarides T. (2002)

Histone methylation: dynamic or static? Cell, 109 (7), 801 – 806, Review.

Barilla D., Lee B. A. and Proudfoot N. J. (2001)

Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in Saccharomyces cerevisiae. Proceedings of the National Academy of Science, 98 (2), 445 – 450.

Becker P. B. (2006)

Gene regulation: A finger on the mark. Nature, 442 (7098), 31 – 32.

Belotserkovskaya R., Oh S., Bondarenko V. A., Orphanides G., Studitsky V. M. and Reinberg D. (2003)

FACT facilitates transcription-dependent nucleosome alteration. Science, 301 (5636), 1090 – 1093.

Belotserkovskaya R. and Reinberg D. (2004)

Facts about FACT and transcript elongation through chromatin. Current Opinion in Genetics & Development, 14, 139 – 146.

Belotserkovskaya R., Saunders A., Lis J. T. and Reinberg D. (2004)

Transcription through chromatin: understanding a complex FACT. Biochimica et Biophysica Acta, 677 (1-3), 87 – 99.

Bentley D. (2002)

The mRNA assembly line: transcription and processing machines in the same factory. Current Opinion in Cell Biology, 14 (3), 336 – 342.

Bird G., Zorio D. A. and Bentley D. L. (2004)

RNA polymerase II carboxy-terminal domain phosphorylation is required for cotranscriptional pre-mRNA splicing and 3'-end formation. Molecular and Cellular Biology, 24 (20), 8963 – 8969.

Boeger H., Bushnell D. A., Davis R., Griesenbeck J., Lorch Y., Strattan J. S., Westover K. D. and Kornberg R. D. (2005)

Structural basis of eukaryotic gene transcription. FEBS Letters, 579 (4), 899 – 903, Review.

Brewster N. K., Johnston G. C. and Singer R. A. (1998)
Characterization of the CP complex, an abundant dimer of Cdc68 and Pob3 proteins that regulates yeast transcriptional activation and chromatin repression. Journal of Biological Chemistry, 273 (34), 21972 – 21979.

Brewster N. K., Johnston G. C. and Singer R. A. (2001)
A bipartite yeast SSRP1 analog comprised of Pob3 and Nhp6 proteins modulates transcription. Molecular and Cellular Biology, 21 (10), 3491 – 3502.

Brunger A. T., Adams P. D., Clore G. M., DeLano W. L., Gros P., Grosse-Kunstleve R. W., Jiang J. S., Kuszewski J., Nilges M., Pannu N. S., Read R. J., Rice L. M., Simonson T. and Warren G. L. (1998)
Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallographica. Section D. Biological Crystallography, 54 (5), 905 – 921.

Buratowski S. (2003)
The CTD code. Nature Structural & Molecular Biology, 10, 679 – 680.

Carrozza M. J., Li B., Florens L., Suganuma T., Swanson S. K., Lee K. K., Shia W. J., Anderson S., Yates J., Washburn M. P. and Workman J. L. (2005)
Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell, 123 (4), 581 – 592.

Cho E. J., Kobor M. S., Kim M., Greenblatt J. and Buratowski S. (2001)
Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser2 of the RNA polymerase II C-terminal domain. Genes & Development, 15 (24), 3319 – 3329.

Cho E. J., Takagi T., Moore C. R., and 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.

Cornilescu G., Delaglio F. and Bax A. (1999)
Protein backbone angle restraints from searching a database for chemical shift and sequence homology. Journal of biomolecular NMR, 13 (3), 289 – 302.

Costa P. J. and Arndt K. M. (2000)
Synthetic lethal interactions suggest a role for the *Saccharomyces cerevisiae* Rtf1 protein in transcription elongation. Genetics, 156 (2), 535 – 547.

Cramer P., Bushnell D. A. and Kornberg R. D. (2001)
Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. Science, 292 (5523), 1863 – 1876.

Dahmus M. E. (1996)
Reversible phosphorylation of the C-terminal domain of RNA polymerase II. Journal of Biological Chemistry, 271 (32), 19009 – 19012.

Delaglio F., Grzesiek S., Vuister G. W., Zhu G., Pfeifer J. and Bax A. (1995)

NMRPipe: a multidimensional spectral processing system based on UNIX pipes. Journal of biomolecular NMR, 6, 277 – 293.

Eissenberg J. C. and Elgin S. C. (2000)

The HP1 protein family: getting a grip on chromatin. Current opinion in genetics & development, 10 (2), 204 – 210, Review.

Eissenberg J. C and Elgin S. C. (2005)

Molecular biology: antagonizing the neighbours. Nature, 438 (7071), 1090 – 1091.

Eissenberg J. C. and Shilatifard A. (2006)

Leaving a mark: the many footprints of the elongating RNA polymerase II. Current Opinion in Genetics & Development, 16, 184 – 190.

Fabrega C., Shen V., Shuman S. and Lima C. D. (2003)

Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. Molecular Cell, 11 (6), 1549 – 1561.

Farrow N. A., Muhandiram R., Singer A. U., Pascal S. M., Kay C. M., Gish G., Shoelson S. E., Pawson T., Forman-Kay J. D. and Kay L. E. (1994)

Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. Biochemistry. 33 (19), 5984 – 6003.

Fong N. and Bentley D. L. (2001)

Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. Genes & Development, 15 (14), 1783 – 1795.

Forget D., Langelier M. F., Therien C., Trinh V. and Coulombe B. (2004)

Photo-cross-linking of a purified preinitiation complex reveals central roles for the RNA polymerase II mobile clamp and TFIIE in initiation mechanisms. Molecular and Cellular Biology, 24 (3), 1122 – 1131.

Formosa T., Eriksson P., Wittmeyer J., Ginn J., Yu Y. and Stillman D. J. (2001)

Spt16-Pob3 and the HMG protein Nhp6 combine to form the nucleosome-binding factor SPN. EMBO Journal, 20 (13), 3506 – 3517.

Flanagan J. F., Mi L. Z., Chruszcz M., Cymborowski M., Clines K. L., Kim Y., Minor W., Rastinejad F. and Khorasanizadeh S. (2005)

Double chromodomains cooperate to recognize the methylated histone H3 tail. Nature, 438, 1181 – 1185.

Grunstein M. (1997)

Histone acetylation in chromatin structure and transcription. Nature, 389 (6649), 349 – 352.

Hahn S. (2004)

Structure and mechanism of the RNA polymerase II transcription machinery. Nature Structural & Molecular Biology, 11 (5), 394 – 403, Review.

Hausmann S. and Shuman S. (2002)

Characterization of the CTD phosphatase Fcp1 from fission yeast. Preferential dephosphorylation of serine 2 versus serine 5. Journal of Biological Chemistry, 277 (24), 21213 – 21220.

Hartzog G. A., Speer J. L. and Lindstrom D. L. (2002)

Transcript elongation on a nucleoprotein template. Biochimica et Biophysica Acta, 1577, 276 – 286, Review.

Hartzog G. A., Wada T., Handa H. and 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.

Hani J., Schelbert B., Bernhardt A., Domdey H., Fischer G., Wiebauer K. and Rahfeld J. U. (1999)

Mutations in a peptidylprolyl-cis/trans-isomerase gene lead to a defect in 3'-end formation of a pre-mRNA in *Saccharomyces cerevisiae*. Journal of Biological Chemistry, 274 (1), 108 – 116.

Hirose Y. and Manley J. L. (2000)

RNA polymerase II and the integration of nuclear events. Genes & Development, 14 (12), 1415 – 1429.

Ho C. K. and Shuman S. (1999)

Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. Molecular and Cellular Biology, 3 (3), 405 – 411.

Holm L. and Sander C. (1995)

Dali: a network tool for protein structure comparison. Trends in Biochemical Sciences, 20 (11), 478 – 480.

Huyen Y., Zgheib O., Ditullio R. A. Jr., Gorgoulis V. G., Zacharatos P., Petty T. J., Sheston E. A., Mellert H. S., Stavridi E. S. and Halazonetis T. D. (2004)

Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. Nature, 432 (7015), 406 – 411.

Ivanov D., Kwak Y. T., Guo J. and Gaynor R. B. (2000)

Domains in the SPT5 protein that modulate its transcriptional regulatory properties. Molecular and Cellular Biology, 20 (9), 2970 – 2983.

Izban M. G. and Luse D. S. (1992)

Factor-stimulated RNA polymerase II transcribes at physiological elongation rates on naked DNA but very poorly on chromatin templates. Journal of Biological Chemistry, 267 (19), 13647 – 13655.

Jansen L. E., den Dulk H., Brouns R. M., de Ruijter M., Brandsma J. A. and Brouwer J. (2000)

Spt4 modulates Rad26 requirement in transcription-coupled nucleotide excision repair. EMBO Journal, 19 (23), 6498 – 6507.

Jasiak A. J. , Armache K. J., Martens B., Jansen R. P. and Cramer P. (2005)

Structural Biology of RNA Polymerase III: Subcomplex C17/25 X-Ray Structure and 11 Subunit Enzyme Model. Molecular Cell, 23 (1), 71 – 81.

Jenuwein T. and Allis C. D. (2001)

Translating the histone code. Science, 293 (5532), 1074 – 1080, Review.

Johnson B. A. and Blevins R.A. (1994)

NMR View: A computer program for the visualization and analysis of NMR data. Journal of Biomolecular NMR, 4, 603 – 614.

Jones J. C., Phatnani H. P., Haystead T. A., MacDonald J. A., Alam S. M. and Greenleaf A. L. (2004)

C-terminal repeat domain kinase I phosphorylates Ser2 and Ser5 of RNA polymerase II C-terminal domain repeats. Journal of Biological Chemistry, 279 (24), 24957 – 24964.

Kamenski T., Heilmeier S., Meinhart A. and Cramer P. (2004)

Structure and mechanism of RNA polymerase II CTD phosphatases. Molecular Cell, 15 (3), 399 – 407.

Kaplan C. D., Laprade L. and Winston F. (2003)

Transcription elongation factors repress transcription initiation from cryptic sites. Science, 301 (5636), 1096 – 1099.

Kaplan C. D., Morris J. R., Wu C. and Winston F. (2000)

Spt5 and spt6 are associated with active transcription and have characteristics of general elongation factors in D. melanogaster. Genes, 14 (20), 2623 – 2634.

Jancarik J. and Kim S. H. (1991)

Sparse matrix sampling: a screening method for crystallization of proteins. Journal of applied crystallography. 24, 409 – 411.

Keogh M. C., Kurdistani S. K., Morris S. A., Ahn S. H., Podolny V., Collins S. R., Schuldiner M., Chin K., Punna T., Thompson N. J., Boone C., Emili A., Weissman J. S., Hughes T. R., Strahl B. D., Grunstein M., Greenblatt J. F., Buratowski S. and Krogan N. J. (2005)

Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. Cell, 123 (4), 593 – 605.

Kettenberger H., Armache K. J. and Cramer P. (2004)

Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. Molecular Cell, 16 (6), 955 – 965.

Khorasanizadeh S. (2004)

The Nucleosome: From genomic organization to genomic regulation. Cell, 116, 259 – 272.

Kim J. B. and Sharp P. A. (2001)

Positive transcription elongation factor B phosphorylates hSPT5 and RNA polymerase II carboxyl-terminal domain independently of cyclin-dependent kinase-activating kinase. Journal of Biological Chemistry, 276 (15), 12317 – 12323.

Kizer K. O., Phatnani H. P., Shibata Y., Hall H., Greenleaf A. L. and Strahl B. D. (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.

Knezetic J. A. and Luse D. S. (1986)

The presence of nucleosomes on a DNA template prevents initiation by RNA polymerase II in vitro. Cell, 45 (1), 95 – 104.

Kobor M. S. and Greenblatt J. (2002)

Regulation of transcription elongation by phosphorylation. Biochimica et Biophysica Acta, 1577 (2), 261 – 275, Review.

Komarnitsky P., Cho E. J. and Buratowski S. (2000)

Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes & Development, 14 (19), 2452 – 2460.

Koradi R., Billeter M. and Wuthrich K. (1996)

MOLMOL: a program for display and analysis of macromolecular structures. J. Mol. Graph, 14 (1), 51 – 55, 29 – 32.

Kobor M. S. and Greenblatt J. (2002)

Regulation of transcription elongation by phosphorylation. Biochimica et Biophysica Acta, 1577 (2), 261 – 275.

Krogan N. J., Kim M., Ahn S. H., Zhong G., Kobor M. S., Cagney G., Emili A., Shilatifard A., Buratowski S. and Greenblatt J. F. (2002)

RNA polymerase II elongation factors of *Saccharomyces cerevisiae*: a targeted proteomics approach. Molecular and Cellular Biology, 22 (20), 6979 – 6992.

Krogan N. J., Kim M., Tong A., Golshani A., Cagney G., Canadien V., Richards D. P., Beattie B. K., Emili A., Boone C., Shilatifard A., Buratowski S. and Greenblatt J. (2003)

Methylation of histone H3 by Set2 in *Saccharomyces cerevisiae* is linked to transcriptional elongation by RNA polymerase II. Molecular and Cellular Biology, 12, 4207 – 4218.

Kyrpides N. C., Woese C. R. and Ouzounis C. A. (1996)

KOW: a novel motif linking a bacterial transcription factor with ribosomal proteins. Trends in Biochemical Sciences, 21 (11), 425 – 426.

Lachner S. and Jenuwein T. (2002)

The many faces of histone lysine methylation. Current Opinion in Cell Biology, 14, 286 – 298.

Lachner M., O'Carroll D., Rea S., Mechtler K. and Jenuwein T. (2001)

Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature, 410 (6824), 116 – 120.

Laskowski R. A., Rullmannn J. A., MacArthur M. W., Kaptein R. and Thornton J. M. (1996)

AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. Journal of biomolecular NMR, 8 (4), 477 – 486.

Li, H., Ilin S., Wang W., Duncan E. M., Wysocka J., Allis C. D. and Patel D. J. (2006)

Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. Nature, 442, 91 – 95.

Li M., Phatnani H. P., Guan Z., Sage H., Greenleaf A. L. and 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 Science, 102 (49), 17636 – 17641.

Lima C. D. (2005)

Inducing interactions with the CTD. Nature Structural & Molecular Biology, 12 (2), 102 – 103.

Linding R., Jensen L. J., Diella F., Bork P., Gibson T. J. and Russel R. B. (2003)

Protein disorder prediction: implications for structural proteomics. Structure, 11 (11), 1453 – 1459.

Lindstrom D. L. and Hartzog G. A. (2001)

Genetic interactions of Spt4-Spt5 and TFIIS with the RNA polymerase II CTD and CTD modifying enzymes in *Saccharomyces cerevisiae*. Genetics, 159 (2), 487 – 497.

Lindstrom D. L., Squazzo S. L., Muster N., Burckin T. A., Wachter K. C., Emigh C. A., McCleery J. A., Yates J. R. III and Hartzog G. A. (2003)

Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins. Molecular and Cellular Biology, 23 (4), 1368 – 1378.

Linge J. P., O'Donoghue S. I. and Nilges M. (2001)

Automated assignment of ambiguous nuclear overhauser effects with ARIA. Methods Enzymology, 339, 71 – 90.

Linge J. P., Williams M. A., Spronk C. A., Bonvin A. M. and Nilges M. (2003)

Refinement of protein structures in explicit solvent. Proteins, 50 (3), 496 – 506.

Lu P. J., Zhou X. Z., Shen M. and Lu K. P. (1999)

Function of WW domains as phosphoserine- or phosphothreonine-binding modules. Science, 283 (5406), 1325 – 1328.

Luger K., Mader A. W., Richmond R. K., Sargent D. F. and Richmond T. J. (1997)

Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature, 389 (6648), 251 – 260.

Neugebauer K. M. and Roth M. B. (1997)

Transcription units as RNA processing units. Genes & Development, 11 (24), 3279 – 3285.

Maniatis T. and Reed R. (2002)

An extensive network of coupling among gene expression machines. Nature, 416 (6880), 499 – 506.

Margueron R., Trojer P. and Reinberg D. (2005)

The key to development: interpreting the histone code? Current Opinion in Genetics & Development, 15, 163 – 176.

Martin C. and Zhang Y. (2005)

The diverse functions of histone lysine methylation. Nature Reviews Molecular cell biology, 6 (11), 838 – 849, Review.

McCracken S., Fong N., Rosonina E., Yankulov K., Brothers G., Siderovski D., Hessel A., Foster S., Shuman S. and Bentley D. L. (1997a)

5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. Genes & Development, 11 (24), 3306 – 3318.

McCracken S., Fong N., Yankulov K., Ballantyne S., Pan G., Greenblatt J., Patterson S. D., Wickens M. and Bentley D. L. (1997b)

The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature, 385 (6614), 357 – 361.

Mason P. B. and 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.

Meinhart A., Blobel J. and Cramer P. (2003)

An extended winged helix domain in general transcription factor E/IIE alpha. The Journal of Biological Chemistry, 278 (48), 48267 – 48274.

Meinhart A. and Cramer P. (2004)

Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. Nature, 430, 223 – 226.

Meinhart A., Kamenski T., Hoeppner S., Baumli S. and Cramer P. (2005)

A structural perspective of CTD function. Genes & Development, 19, 1401 – 1415.

Mellor J. (2006)

It takes a PHD to read the histone code. Cell, 126 (1), 22 **–** 24.

Mellor J. (2006)

Dynamic nucleosomes and gene transcription. TRENDS in Genetics, 22 (6), 320 – 329, Review.

Morris D. P., Phatnani H. P. and Greenleaf A. L. (1999)

Phospho-carboxyl-terminal domain binding and the role of a prolyl isomerase in pre-mRNA 3'-end formation. Journal of Biological Chemistry, 274 (44), 31583 – 31587.

Morris S. A., Shibata Y., Noma K., Tsukamoto Y., Warren E., Temple B., Grewal S. I. and Strahl B. D. (2005)

Histone H3 K36 methylation is associated with transcription elongation in *Schizosaccharomyces pombe*. Eukaryotic Cell, 4 (8), 1446 – 1454.

Murakami K. S., Masuda S. and Darst S. A. (2002)

Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. Science, 296 (5571), 1280 – 1284.

Nelson C. J., Santos-Rosa H. and Kouzarides T. (2006)

Proline isomerization of histone h3 regulates lysine methylation and gene expression. Cell, 126 (5), 905 – 916.

Ng H. H., Xu R. M., Zhang Y. and Struhl K. (2002)

Ubiquitination of histone H2B by Rad6 is required for efficient Dot1-mediated methylation of histone H3 lysine 79. Journal of Biological Chemistry, 277 (38), 34655 – 34657.

Noble C. G., Hollingworth D., Martin S. R., Ennis-Adeniran V., Smerdon S. J., Kelly G., Taylor I. A. and Ramos A. (2005)

Key features of the interaction between Pcf11 CID and RNA polymerase II CTD. Nature Structural & Molecular Biology, 12 (2), 144 – 151.

Nowak S. J. and Corces V. G. (2004)

Phosphorylation of histone H3: a balancing act between chromosome condensation and transcriptional activation. TRENDS in Genetics, 20 (4), 214 – 220.

Ohkuma Y., Sumimoto H., Hoffmann A., Shimasaki S., Horikoshi M. and Roeder R.G. (1991)

Structural motifs and potential sigma homologies in the large subunit of human general transcription factor TFIIE. Nature, 354 (6352), 398 – 401.

Orphanides G., LeRoy G., Chang C. H. , Luse D. S. and Reinberg D. (1998)

FACT, a factor that facilitates transcript elongation through nucleosomes. Cell, 92, 105 – 116.

Orphanides G. and Reinberg D. (2000)

RNA polymerase II elongation through chromatin. Nature, 407, 471 – 475.

Orphanides G. and Reinberg D. (2002)

A unified theory of gene expression. Cell, 108 (4), 439 – 451, Review.

Orphanides G., Wu W. H., Lane W. S., Hampsey M. and Reinberg D. (1999)

The chromatin-specific transcription elongation factor FACT comprises human SPT16 and SSRP1 proteins. Nature, 400 (6741), 284 – 288.

Otwinowski Z. and Minor W. (1996)

Processing of X-ray diffraction data collected in oscillation mode. Methodes in Enzymology, 276, 307 – 326.

Pei Y., Hausmann S., Ho C. K., Schwer B. and Shuman S. (2001)

The length, phosphorylation state, and primary structure of the RNA polymerase II carboxyl-terminal domain dictate interactions with mRNA capping enzymes. Journal of Biological Chemistry, 276 (30), 28075 – 28082.

Pei Y. and Shuman S. (2002)

Interactions between fission yeast mRNA capping enzymes and elongation factor Spt5. The journal of biological chemistry, 277 (22), 19639 – 19648.

Pei Y., Schwer B. and Shuman S. (2003)

Interactions between fission yeast Cdk9, its cyclin partner Pch1, and mRNA capping enzyme Pct1 suggest an elongation checkpoint for mRNA quality control. Journal of Biological Chemistry, 28, 278 (9), 7180 – 7188.

Peterlin B. M. and Price D. H. (2006)

Controlling the elongation phase of transcription with P-TEFb. Molecular Cell, 23 (3), 297 – 305, Review.

Pokholok D. K., Hannett N. M. and Young R. A. (2002)

Exchange of RNA polymerase II initiation and elongation factors during gene expression in vivo. Molecular Cell, 9 (4), 799 – 809.

Pokholok D. K., Harbison C. T., Levine S., Cole M., Hannett N. M., Lee T. I., Bell G. W., Walker K., Rolfe P. A., Herbolsheimer E., Zeitlinger J., Lewitter F., Gifford D. K. and Young R. A. (2005)
Genome-wide map of nucleosome acetylation and methylation in yeast. Cell, 122 (4), 517 – 527.

Ponting C. P. (2002)
Novel domains and orthologues of eukaryotic transcription elongation factors. Nucleic Acids Research, 30 (17), 3643 – 3652.

Pray-Grant M. G., Daniel J. A., Schieltz D., Yates J. R. 3rd and Grant P. A. (2005)
Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. Nature, 433 (7024), 434 – 438.

Proudfoot N. J., Furger A. and Dye M. J. (2002)
Integrating mRNA processing with transcription. Cell, 108 (4), 501 – 512.

Radhakrishnan I., Perez-Alvarado G. C., Parker D., Dyson H. J., Montminy M. R. and Wright P. E. (1997)
Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator: coactivator interactions. Cell, 91 (6), 741 – 752.

Reay P., Yamasaki K., Terada T., Kuramitsu S., Shirouzu M. and Yokoyama S. (2004)
Structural and sequence comparisons arising from the solution structure of the transcription elongation factor NusG from Thermus thermophilus. Proteins, 56 (1), 40 – 51.

Reinberg D. and Sims III R. J. (2006)
De FACTo nucleosome dynamics. Journal of Biological Chemistry - Epub ahead of print.

Rhodes D. (1997)
The nucleosome core all wrapped up. Nature, 389 (6648), 231 – 233.

Russell R. B., Alber F., Aloy P., Davis F. P., Korkin D., Pichaud M., Topf M. and Sali A. (2004)
A structural perspective on protein-protein interactions. Current opinion in Structural Biology, 14 (3), 313 – 324.

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

Sambrook J. and Russell D. W. (2001)
Molecular cloning – A laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Santos-Rosa H., Schneider R., Bannister A. J., Sherriff J., Bernstein B. E., Emre N. C., Schreiber S. L., Mellor J. and Kouzarides T. (2002)

Active genes are tri-methylated at K4 of histone H3. Nature, 419 (6905), 407 – 411.

Sattler M., Schleucher J. and Griesinger C. (1999)

Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Progress in nuclear magnetic resonance spectroscopy, 34, 93 – 158.

Saunders A., Werner J., Andrulis E. D., Nakayama T., Hirose S., Reinberg D. and Lis J. T. (2003)

Tracking FACT and the RNA polymerase II elongation complex through chromatin in vivo. Science, 301 (5636), 1094 – 1096.

Schneider D. A., French S. L., Osheim Y. N., Bailey A. O., Vu L., Dodd J., Yates J. R., Beyer A. L. and Nomura M. (2006)

RNA polymerase II elongation factors Spt4p and Spt5p play roles in transcription elongation by RNA polymerase I and rRNA processing. Proceedings of the National Academy of Science, 103 (34), 12707 – 12712.

Schroeder S. C., Schwer B., Shuman S. and Bentley D. (2000)

Dynamic association of capping enzymes with transcribing RNA polymerase II. Genes & Development, 14 (19), 2435 – 2440.

Shahbazian M. D., Zhang K. and Grunstein M. (2005)

Histone H2B ubiquitylation controls processive methylation but not monomethylation by Dot1 and Set1. Molecular Cell, 19 (2), 271 – 277.

Shatkin A. J. and Manley J. L. (2000)

The ends of the affair: capping and polyadenylation. Nature Structural & Molecular Biology, 7 (10), 838 – 842.

Shi X., Hong T., Walter K. L., Ewalt M., Michishita E., Hung T., Carney D., Pena P., Lan F., Kaadige M. R., Lacoste N., Cayrou C., Davrazou F., Saha A., Cairns B. R., Ayer D. E., Kutateladze T. G., Shi Y., Cote J., Chua K. F. and Gozani O. (2006)

ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. Nature, 442, 96 – 99.

Shilatifard A. (2004)

Transcriptional elongation control by RNA polymerase II: a new frontier. Biochimica et Biophysica Acta, 15, 1677 (1–3), 79 – 86, Review.

Simic R., Lindstrom D. L., Tran H. G., Roinick K. L., Costa P. J., Johnson A. D., Hartzog G. A. and Arndt K. M. (2003)

Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes. EMBO Journal, 22 (8), 1846 – 1856.

Sims R. J. 3rd, Belotserkovskaya R. and Reinberg D. (2004)

Elongation by RNA polymerase II: the short and long of it. Genes & Development, 18 (20), 2437 – 2468.

Sims R. J. 3rd, Chen C. F., Santos-Rosa H., Kouzarides T., Patel S. S. and Reinberg D. (2005)

Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains, Journal of Biological Chemistry, 280 (51), 41789 – 41792.

Sims 3rd R. J., Nishioka K. and Reinberg D. (2003)

Histone lysine methylation: a signature for chromatin function. TRENDS in Genetics, 19 (9), 629 – 639.

Squazzo S. L., Costa P. J., Lindstrom D. L., Kumer K. E., Simic R., Jennings J. L., Link A. J., Arndt K. M. and Hartzog G. A. (2002)

The Paf1 complex physically and functionally associates with transcription elongation factors in vivo. EMBO Journal, 21 (7), 1764 – 1774.

Stavropoulos P., Blobel G. and Hoelz A. (2006)

Crystal structure and mechanism of human lysine-specific demethylase-1. Nature Structural & Molecular Biology, 13 (7), 626 – 632.

Steiner T., Kaiser J. T., Marinkovic S., Huber R. and Wahl M. C. (2002)

Crystal structures of transcription factor NusG in light of its nucleic acid- and protein-binding activities. EMBO Journal, 21 (17), 4641 – 4653.

Stiller J. W. and Cook M. S. (2004)

Functional unit of the RNA polymerase II C-terminal domain lies within heptapeptide pairs. Eukaryotic Cell, 3 (3), 735 – 740.

Storoni L. C., McCoy A. J. and Read R. J. (2004)

Likelihood-enhanced fast rotation functions. Acta Crystallographica. Section D. Biological Crystallography, 60 (3), 432 – 438.

Strahl B. D. and Allis C. D. (2000)

The language of covalent histone modifications. Nature, 403 (6765), 41 – 45.

Strahl B. D., Grant P. A., Briggs S. D., Sun Z. W., Bone J. R., Caldwell J. A., Mollah S., Cook R. G., Shabanowitz J., Hunt D. F. and Allis C. D. (2002)

Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. Molecular and Cellular Biology, 22 (5), 1298 – 1306.

Sun Z. W. and Allis C. D. (2002)

Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. Nature, 418 (6893), 104 – 108.

Svejstrup J. Q. (2002)

Chromatin elongation factors. Current opinion in genetics & development, 12 (2), 156 – 161, Review.

Svejstrup J. Q. (2003)

Histones face the FACT. Science, 301 (5636), 1053 – 1055.

Swanson M. S., Carlson M. and 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 M. S., Malone E. A. and Winston F. (1991)

SPT5, an essential gene important for normal transcription in *Saccharomyces cerevisiae*, encodes an acidic nuclear protein with a carboxy-terminal repeat. Molecular and Cellular Biology, 11 (6), 3009 – 3019.

Swanson M. S. and Winston F. (1992)

SPT4, SPT5 and SPT6 interactions: effects on transcription and viability in *Saccharomyces cerevisiae*. Genetics, 132 (2), 325 – 336.

Talbert P. B. and Henikoff S. (2006)

Spreading of silent chromatin: inaction at a distance. Nature Reviews. Genetics, 7 (10), 793 – 803.

Teng Q. (2005)

Structural Biology: Practical NMR applications. Springer Science + Business Media Inc., New York.

Thompson J. D., Higgins D. G. and Gibson T. J. (1994)

CLUSTAL-W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research, 22, 4673 – 4680.

Trewick S. C., McLaughlin P. J. and Allshire R. C. (2005)

Methylation: lost in hydroxylation? EMBO Reports, 6 (4), 315 – 320.

Trojer P. and Reinberg D. (2006)

Histone lysine demethylases and their impact on epigenetics. Cell, 125, 213 – 217.

Trigon S., Serizawa H., Conaway J. W., Conaway R. C., Jackson S. P. and Morange M. (1998)

Characterization of the residues phosphorylated in vitro by different C-terminal domain kinases. Journal of Biological Chemistry, 273 (12), 6769 – 6775.

Tsukada Y., Fang J., Erdjument-Bromage H., Warren M. E., Borchers C. H., Tempst P. and Zhang Y. (2005)

Histone demethylation by a family of JmjC domain-containing proteins. Nature, 16, 439 (7078), 811 – 816.

Turner B. M. (2005)

Reading signals on the nucleosome with a new nomenclature for modified histones. Nature Structural & Molecular Biology, 12, 110 – 112.

Verdecia M. A., Bowman M. E., Lu K. P., Hunter T. and Noel J. P. (2000)

Structural basis for phosphoserine-proline recognition by group IV WW domains. Nature Structural & Molecular Biology, 7 (8), 639 – 643.

Vire E., Brenner C., Deplus R., Blanchon L., Fraga M., Didelot C., Morey L., Van Eynde A., Bernard D., Vanderwinden J. M., Bollen M., Esteller M., Di Croce L., de Launoit Y. and Fuks F. (2005)

The Polycomb group protein EZH2 directly controls DNA methylation. Nature, 439 (7078), 871 – 874.

Wada T., Orphanides G., Hasegawa J., Kim D. K., Shima D., Yamaguchi Y., Fukuda A., Hisatake K., Oh S., Reinberg D. and Handa H. (2000)

FACT relieves DSIF/NELF-mediated inhibition of transcriptional elongation and reveals functional differences between P-TEFb and TFIIH. Molecular Cell, 5 (6), 1067 – 1072.

Wada T., Takagi T., Yamaguchi Y., Ferdous A., Imai T., Hirose S., Sugimoto S., Yano K., Hartzog G. A., Winston F., Buratowski S. and Handa H. (1998a)

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.

Wada T., Takagi T., Yamaguchi Y., Watanabe D. and Handa H. (1998b)

Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription in vitro. EMBO Journal, 17 (24), 7395 – 7403.

West M. L. and Corden J. L. (1995)

Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. Genetics, 140 (4), 1223 – 1233.

Whetstine J. R., Nottke A., Lan F., Huarte M., Smolikov S., Chen Z., Spooner E., Li E., Zhang G., Colaiacovo M. and Shi Y. (2006)

Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. Cell, 125 (3), 467 – 481.

Winston F., Chaleff D. T., Valent B. and Fink G. R. (1984)

Mutations affecting Ty-mediated expression of the HIS4 gene of *Saccharomyces cerevisiae*. Genetics, 107 (2), 179 – 197.

Wittmeyer J. and Formosa T. (1997)

The *Saccharomyces cerevisiae* DNA polymerase alpha catalytic subunit interacts with Cdc68/Spt16 and with Pob3, a protein similar to an HMG1-like protein. Molecular and Cellular Biology, 17 (7), 4178 – 4190.

Wood A., Krogan N. J., Dover J., Schneider J., Heidt J., Boateng M. A., Dean K., Golshani A., Zhang Y., Greenblatt J. F., Johnston M. and Shilatifard A. (2003)
Bre1, an E3 ubiquitin ligase required for recruitment and substrate selection of Rad6 at a promoter. Molecular Cell, 11 (1), 267 – 274.

Workman J. L. (2006)
Nucleosome displacement in transcription. Genes & Development, 20 (15), 2009 – 2017, Review.

Woychik N. A. and Young R. A. (1989)
RNA polymerase II subunit RPB4 is essential for high- and low-temperature yeast cell growth. Molecular and Cellular Biology, 9 (7), 2854 – 2859.

Wüthrich (1986)
NMR of proteins and nucleic acids. John Wiley & Sons Inc., 1986.

Wysocka J., Swigut T., Xiao H., Milne T. A., Kwon S. Y., Landry J., Kauer M., Tackett A. J., Chait B. T., Badenhorst P., Wu C. and Allis C. D. (2006)
A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature, 442 (7098), 86 – 90.

Xiao B., Jing C., Kelly G., Walker P. A., Muskett F. W., Frenkiel T. A., Martin S. R., Sarma K., Reinberg D., Gamblin S. J. and Wilson J. R. (2005)
Specificity and mechanism of the histone methyltransferase Pr-Set7. Genes & Development, 19 (12), 1444 – 1454.

Xiao B., Jing C., Wilson J. R., Walker P. A., Vasisht N., Kelly G., Howell S., Taylor I. A., Blackburn G. M. and Gamblin S. J. (2003)
Structure and catalytic mechanism of the human histone methyltransferase SET7/9. Nature, 421 (6923), 652 – 656.

Xu Y. X., Hirose Y., Zhou X. Z., Lu K. P. and Manley J. L. (2003)
Pin1 modulates the structure and function of human RNA polymerase II. Genes & Development, 17 (22), 2765 – 2776.

Yamada T., Yamaguchi Y., Inukai N., Okamoto S., Mura T. and 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. and Handa H. (1999b)
NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. Cell, 97 (1), 41 – 51.

Yamaguchi Y., Wada T., Watanabe D., Takagi T., Hasegawa J. and Handa H. (1999a)

Structure and function of the human transcription elongation factor DSIF. Journal of Biological Chemistry, 274 (12), 8085 – 8092.

Yuryev A. and Corden J. L. (1996)

Suppression analysis reveals a functional difference between the serines in positions two and five in the consensus sequence of the C-terminal domain of yeast RNA polymerase II. Genetics, 143 (2), 661 – 671.

Yuryev A., Patturajan M., Litingtung Y., Joshi R. V., Gentile C., Gebara M. and Corden J. L. (1996)

The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. Proceedings of the National Academy of Science, 93 (14), 6975 – 6980.

Zhang J. and Corden J. L. (1991)

Identification of phosphorylation sites in the repetitive carboxyl-terminal domain of the mouse RNA polymerase II largest subunit. Journal of Biological Chemistry, 266, 2290 – 2296.

Zhang X., Yang Z., Khan S. I., Horton J. R., Tamaru H., Selker E. U. and Cheng X. (2003)

Structural basis for the product specificity of histone lysine methyltransferases. Molecular Cell, 12 (1), 177 – 185.

Zhao J., Hyman L. and Moore C. (1999)

Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. Microbiology and molecular biology Reviews, 63 (2), 405 – 445, Review.

Curriculum Vitae

E R I K A   V O J N I Ć

Personal data

Date of Birth:	June 18, 1975
Place of Birth:	Augsburg/Germany
Martial status:	unmarried
Citizenship:	Germany

Education

1982 – 1986	Birkenau-Volksschule, Augsburg – Grundschule
1986 – 1988	Birkenau-Volksschule, Augsburg – Hauptschule
1988 – 1992	Maria-Ward-Realschule, Augsburg
1992 – 1996	Holbein-Gymnasium, Augsburg

Academic Training

1996 – 2002	Chemistry studies at the Ludwig-Maximilians-Universität, München Diploma degree in Chemistry
2002 – 2006	December/December
	Dissertation in *NMR solution structure of the Set2 SRI domain and preparation of RNA polymerase II complexes with the elongation factor Spt4-Spt5*
	Prof. Dr. Patrick Cramer Ludwig-Maximilians-Universität, Gene Center, München
2005 2006	May – July/September – October March – April Stay abroad at the European Molecular Biology Laboratory Heidelberg, Germany Dr. Michael Sattler