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

Structure and function of RNA Polymerase II-Bye1 complexes

Kerstin Kinkelin
aus
Tuttlingen, Deutschland

2013

Dissertation zur Erlangung des Doktorgrades

der Fakultät für Chemie und Pharmazie

der Ludwig-Maximilians-Universität München

Structure and function of RNA Polymerase II-Bye1 complexes

Kerstin Kinkelin

aus

Tuttlingen, Deutschland

2013

Erklärung

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

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den 19.08.2013

Kerstin Kinkelin

Dissertation eingereicht am 19.08.2013

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

Mündliche Prüfung am 14.10.2013

Acknowledgements

Thanks to Patrick for not only being my PhD supervisor, but for actually shaping me as a scientist.Your approach of giving people as much freedom as possible, and at the same time keeping track of what they are doing, has taught me to confidently asses my abilities, to not only follow advice, but also come up with my own ideas and solutions, and finally to become a scientist that can work independently, but is never shy to ask for help or take advice. Your skill of creating a multifaceted team of like-minded people, and not allowing lone warriors, brings together what is not seen too often - successful research and happy people. I'm extremely grateful for having worked with you, for your support and for experiencing that research is not mostly about working hard, but about working passionate.

Thanks to all the members of the Cramer group for being an outstanding team that - in my eyes - is hard to find in research these days. The warm-hearted and cheerful working atmosphere experienced in this group is invaluable, and the great team spirit kept many people motivated during difficult times, including me. I can not remember a single occasion where help was refused by anyone, and similarly there were countless occasions where people took their time to figure out a solution for yet another Bye1-related problem. Every person in the lab has contributed to this truly memorable time, and I especially want to thank Amelie, Carrie, Clemens, Kerstin, Laurent, Merle, Nicole, Rebecca, Rike and Steffi for supporting me scientifically and personally, and for always making me believe that Bye1 will be defeated one day. All the fun and success wouldn't have been possible without two very special bench neighbours, Fuen and Sarah. I want to thank both of you not only for all your scientific advice, but mostly for endless hours of fun inside and outside the lab, for being an incredible team at the SLS (we beat the robot!), for hiking and skiing, Tequila and Tamales, Ebenen and parties, and for always being there for me.

Thanks to all the members of my thesis committee: Dr. Dietmar Martin, Prof. Dr. Andreas Ladurner, Prof. Dr. Peter Becker, Dr. Katja Sträßer and Prof. Dr. Mario Halic, for their support and time.

Thanks to the members of my thesis advisory committee: Dr. Esben Lorenzen and Prof. Dr. Jens Michaelis, for their support, discussions and advice.

Thanks to Hans-Jörg, Ingrid and Maxi, for doing an awesome job in keeping an outstanding IMPRS PhD program running and for providing a platform for meeting many new friends. Thanks to the Boehringer Ingelheim Fonds for financial and personal support, and for providing an excellent opportunity of networking.

Danke an Adriane, Leonie, Regina und Vroni für all die Unterstützung, fürs Zuhören, fürs

Motivieren, fürs gemeinsame Lachen und all die anderen schönen Dinge, die wir gemeinsam erlebt haben.

Thanks to Alan, first of all for your patience in handling my craziness and germaness, and for accepting and loving me the way I am. But also and especially for your constant motivation, coddling up and believing in me and my capabilities. Not only at work, but mostly in everyday life.

Ein ganz großes Dankeschön an meine Familie. Das Wissen, dass ihr immer an mich geglaubt habt, und dass ihr immer hinter mir und meinen Ideen standet, hat mich zu dem Menschen gemacht, der ich heute bin. Ohne eure Unterstützung und euren Zuspruch würde ich diese Zeilen heute nicht schreiben. Vielen Dank.

Summary

During transcription of eukaryotic protein-coding genes, RNA Polymerase (Pol) II associates transiently with dozens of transcription factors. Different Pol II-associated factors are required for transcription initiation, RNA chain elongation through chromatin, pre-mRNA 5'-capping, splicing, 3'-RNA processing of the nascent transcript, and transcription termination. Bye1, a transcription factor found in yeast, has been linked to Pol II and transcription elongation in genetic studies.

In this work, we showed that Bye1 binds directly to Pol II, and obtained four crystal structures of different Pol II functional complexes bound by Bye1. The structures revealed that Bye1 contacts Pol II at a similar position as the general transcription factor TFIIS, but the details of the interaction differ. Also, Bye1 does not change Pol II conformation, in contrast to other transcription factors directly contacting Pol II. Functional and structural analysis revealed that basic Pol II mechanisms like nucleotide incorporation, transcription on chromatin-free templates, and arrest and backtracking are not affected by Bye1.

Furthermore, we could show that Bye1 is recruited to the 5' region of actively transcribed genes, and that it interacts in general with histone marks of active chromatin, and specifically binds to trimethylated K4 on H3. In addition, we identified two genes involved in transcription elongation and histone modification that functionally interact with Bye1, namely Paf1 and Tho2. Based on a detailed bioinformatic search, we determined two putative human homologues of Bye1, namely PHF3 and DIDO. Finally, we generated a structural model of Bye1-bound Pol II approaching a +2 nucelosome by combining data obtained during this and other studies. The model shows that it is structurally possible for Bye1 to interact simultaneously with Pol II and histone tails emerging from the +2 nucleosome.

Based on our data, we propose that Bye1 may be the founding member of a new family of transcription factors that link early transcribing Pol II to histones in yeast and human cells.

Publications

Part of this work has been published:

K. Kinkelin, G. G. Wozniak, S. B. Rothbart, M. Lidschreiber, B. D. Strahl, and P. Cramer. Structures of RNA polymerase II complexes with Bye1, a chromatin-binding PHF3/DIDO homologue. *Proc Natl Acad Sci USA* 110, no 38 (2013):15277-82

Author contributions: K. Kinkelin carried out structural work, surface plasmon resonance experiments, SL screens and transcription and RNA extension assays. G.G. Wozniak carried out chromatin assays. S.B. Rothbart carried out peptide arrays. M. Lidschreiber carried out ChIP-chip experiments. B.D. Strahl and P. Cramer designed and supervised research. K. Kinkelin and P. Cramer prepared the manuscript, with help from all authors.

Contents

Part I

Introduction

1 Transcription by DNA-dependent RNA Polymerases

1.1 Transcription in eukaryotes

In all living organisms, genetic information is stored in form of DNA, and DNA-dependent RNA polymerases are required for DNA transcription to synthesize RNA. Although the active site of different RNA polymerases is structurally discrete and evolutionarily unrelated, all RNA polymerases have a common structural framework and operate by closely related molecular mechanisms (Werner, 2007).

Whereas archaea and bacteria use a single RNA polymerase to transcribe all genes, eukaryotes use several classes of polymerases that transcribe distinct subsets of genes (Werner, 2007). RNA Polymerase (Pol) I, II, and III synthesize the 25S rRNA precursor, messenger RNAs (mRNAs), and short untranslated RNAs such as transfer RNAs (tRNAs) and 5S ribosomal RNA (rRNA), respectively. In plants, Pol IV and V additionally transcribe non-coding RNAs required for RNA interference (Matzke *et al.*, 2009). In addition, separate RNA polymerases, which are similar to phage RNA polymerases, are found in chloroplasts and mitochondria, where they specifically transcribe the DNA of those organelles.

Although the different polymerases transcribe distinct subsets of genes, they share several common features (Table 1, Vannini & Cramer, 2012). Pol I, II and III contain 14, 12, and 17 subunits, respectively, and the two largest subunits are related to those of the bacterial polymerase. In addition, five of the core subunits are conserved among Pol I, II and III. Consistent with these structural similarities, the different eukaryotic polymerases share several functional properties, including the need to interact with other proteins to appropriately initiate transcription. As functional differences are required for transcribing distinct subsets of genes, the core of the enzyme is decorated with additional polymerase-specific factors.

Table 1: RNA polymerase subunits in yeast. Adapted from Vannini & Cramer, 2012.

Pol I	Pol II	Pol III
Polymerase core		
A190	Rpb1	C160
A135	Rpb2	C128
AC40	Rpb3	AC40
Rpb5	Rpb5	Rpb5
Rpb6	Rpb6	Rpb6
Rpb8	Rpb8	Rpb8
A12.2 N ribbon	Rpb9	C11 N ribbon
Rpb10	Rpb10	Rpb10
AC19	Rpb11	AC19
Rpb12	Rpb12	Rpb12
Polymerase Stalk		
A14	Rpb4	C17
A43	Rpb7	C25
General transcription factors and their counterparts		
A49 N term	Tfg1 (TFIIFα)	C37
A34.5	Tfg2 (TFIIF β)	C53
	Tfa1 (TFIIEα)	C82
A49 C term	Tfa2 (TFIIEβ)	C34
		C31
TBP	TBP	TBP
TAFs		
Rrn7	TFIIB	Brf1
Specific factors		
		B
Rrn6		
Rrn11		
	SAGA	
	TFIIH	
Rrn3	Mediator	
UAF		
		TFIIIC
		SNAPc

1.2 Regulation of RNA polymerase II transcription

1.2.1 The RNA polymerase II transcription cycle

The transcription cycle of Pol II can be divided into four steps (Fig. 1): pre-initiation, initiation, elongation and termination. Pol II is recruited to the promoter during pre-initiation, where melting of the DNA and pre-initiation complex formation occurs. During initiation, incorporation of the first nucleotides takes place. This process is marked by physical and functional instability of the transcription complex, and results in formation of abortive transcripts. However, when nascent RNA reaches a length of 14-15 nucleotides, transition to the elongation stage of transcription occurs (Dvir, 2002), where processive catalysis generates the bulk of the RNA. Finally, transcription is terminated when Pol II reaches the 3' end of the transcribed gene, and Pol II can then be recycled for another round of transcription.

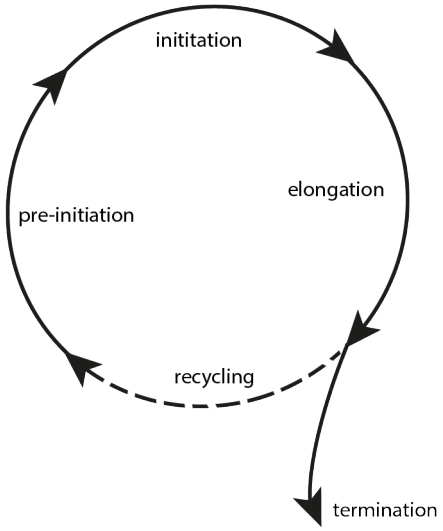

Figure 1: RNA Polymerase II transcription cycle.

In 1980, Robert Roeder and his colleagues discovered that transcription in the eukaryotic system requires distinct transcription factors during all stages of the transcription cycle (Matsui *et al.*, 1980). These so-called general transcription factors are involved in transcription from all Pol II promoters and therefore constitute part of the basic transcription machinery.

1.2.2 General transcription factors in yeast

Five general transcription factors are required for initiation of transcription by Pol II (Table 2, Sikorski & Buratowski, 2009). The first step in formation of a transcription complex is

the binding of the general transcription factor TFIID to the TATA box via its TATA-binding protein (TBP). TBP binding induces a 90° kink in the DNA (Kim *et al.*, 1993a,b), which then allows binding of a second general transcription factor, TFIIB. TFIIB in turn serves as a bridge between the TBP-DNA complex and Pol II (Bushnell *et al.*, 2004; Kostrewa *et al.*, 2009; Liu *et al.*, 2010; Sainsbury *et al.*, 2013), which is recruited to the promoter in association with a third general transcription factor, TFIIF. Binding of two additional factors, TFIIE and TFIIH, is required for initiation of transcription. TFIIE enhances recruitment of TFIIH to promoters, where TFIIH assists in promoter opening and clearance by unwinding DNA and phosphorylating the C-terminal domain (CTD) of Rpb1, the largest Pol II subunit (Lee & Young, 2000).

Table 2: General transcription factors in yeast. Adapted from Sikorski & Buratowski, 2009

Protein complex	Subunits	Function
TFIIB	1	stabilizes TFIID-promoter binding; aids in recruitment of TFIIF/Pol II to the promoter; directs accurate start site selection
TFIID	14	nucleates PIC assembly either through TBP binding to TATA sequences or TAF binding to other promoter sequences; coactivator activity through direct interaction of TAFs and gene specific activators
TFIIE	2	helps recruit TFIIH to promoters; stimulates helicase and kinase activities of TFIIH; binds ssDNA and is essential for promoter melting
TFIIF	3	tightly associates with Pol II; enhances affinity of Pol II for TBP-TFIIB-promoter complex; necessary for recruitment of TFIIE/TFIIH to the PIC; aids in start site selection and promoter escape; enhances elongation efficiency
TFIIH	10	ATPase/helicase necessary for promoter opening and promoter clearance; helicase activity for transcription coupled DNA repair; kinase activity required for phosphorylation of Pol II CTD; facilitates transition from initiation to elongation

1.2.3 General transcription factor TFIIS

The above described general transcription factors are required during initiation of transcription. The general transcription factor TFIIS, however, is required during transcription elongation. In contrast to other elongation factors that stimulate nucleotide incorporation, TFIIS re-activates Pol II during arrest and backtracking (Fig. 2, Sigurdsson *et al.*, 2010; Cheung & Cramer,

2011). Pol II arrest and reactivation occur during promoter-proximal gene regulation and transcription through nucleosomes. Certain DNA sequences cause Pol II to arrest, which is followed by backtracking, displacement of the RNA to a so-called backtrack site, and inhibition of elongation (Adelman *et al.*, 2005; Kireeva *et al.*, 2005), due to misalignment of the catalytic 3' OH group of the RNA (Wang *et al.*, 2009; Cheung & Cramer, 2011).

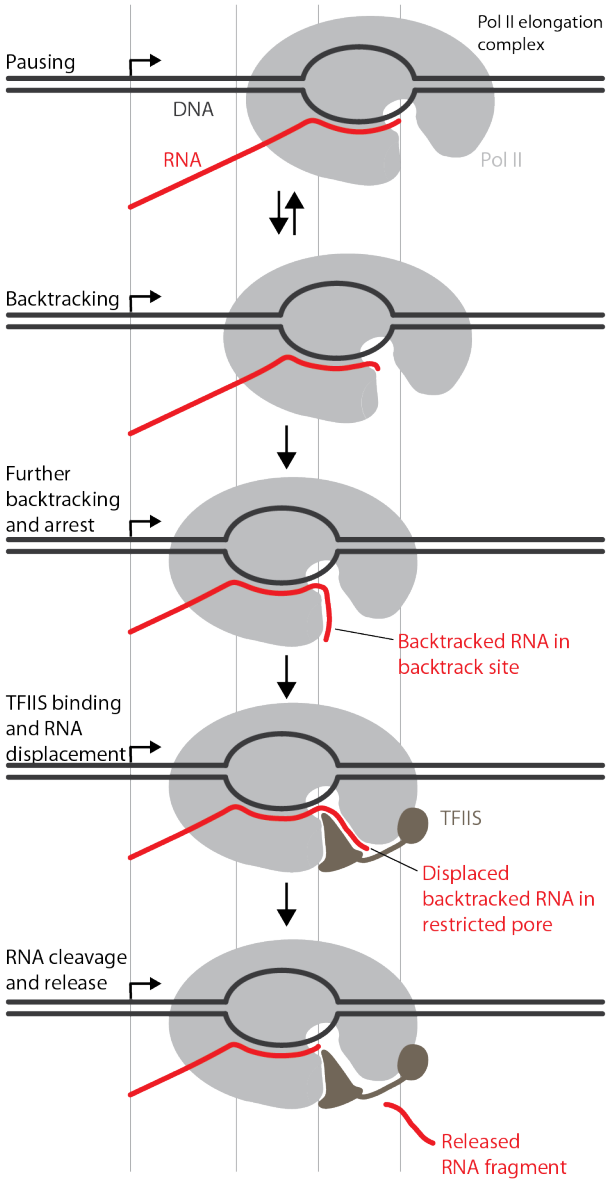

Figure 2: Mechanism of Pol II backtracking, arrest and reactivation. Adapted from Cheung & Cramer, 2011.

TFIIS contains three domains, a mobile N-terminal domain, a central domain that binds directly to the Pol II jaw and funnel domains, and a C-terminal zinc ribbon domain that

inserts into the polymerase pore and reaches the Pol II active site (Kettenberger *et al.*, 2003). During backtracking, the C-terminal zinc ribbon domain stimulates cleavage of backtracked RNA (Wind & Reines, 2000) and thereby re-activates Pol II transcription by restoring the RNA 3' OH in the active site (Wang *et al.*, 2009; Cheung & Cramer, 2011).

2 *S.cerevisiae* RNA Polymerase II

2.1 Architecture and subunit composition

Pol II consists of a 10-subunit core enzyme and a peripheral Rpb4/7 subcomplex (Table 1), and comprises 513 kDa.

The first crystal structure of the ten subunit Pol II core was solved in 2001 (Cramer *et al.*, 2001) at 3 Å. The backbone model revealed that Rpb1 and Rpb2 constitute the bulk of the enzyme and form the active centre, whereas the smaller subunits are distributed around the periphery (Fig. 3a,b). Crystallographic backbone models of the complete, 12-subunit Pol II, revealed the position and orientation of the peripheral Rpb4/7 subcomplex in 2003 (Fig. 3c, and Armache *et al.*, 2003; Bushnell & Kornberg, 2003). A crystal structure of free Rpb4/7 and improved resolution of the complete Pol II crystals finally enabled refinement of a complete atomic model of Pol II (Armache *et al.*, 2005).

In addition to the domains described in the crystal structures, Pol II contains a tail-like domain, the C-terminal domain (CTD) of Rpb1. The CTD is flexibly linked to the core enzyme and consists of heptapeptide repeats of the consensus sequence YSPTSPS, which are differentially phosphorylated during the transcription cycle (Heidemann *et al.*, 2013). CTD-binding proteins recognize a specific CTD phosphorylation pattern and are involved in mRNA biogenesis (Hsin & Manley, 2012).

Pol II is highly conserved across species, with yeast and human Pol II sequences exhibiting 53% overall identity (Cramer *et al.*, 2001), and ten subunits encoded by human genes can replace their counterparts in yeast Pol II without loss of cell viability (Woychik, 1998).

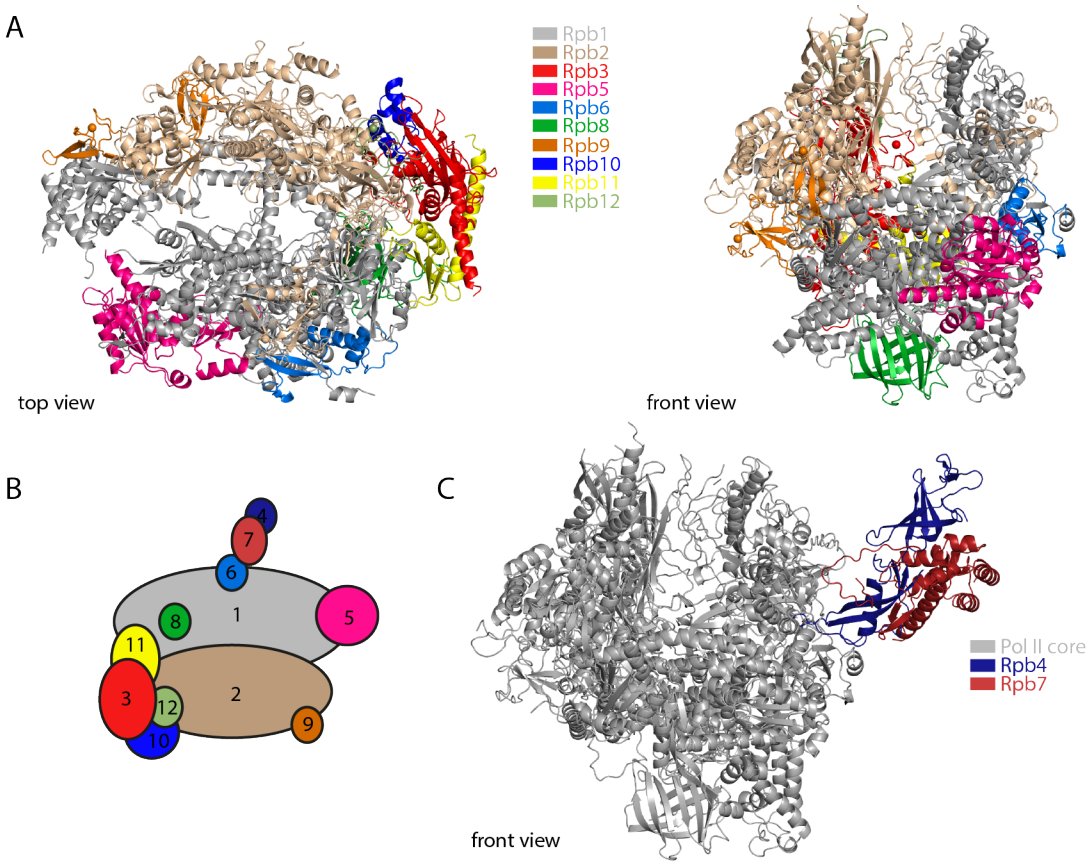

Figure 3: Architecture and subunit composition of *S.cerevisiae* Pol II.
(A) Ribbon model of 10-subunit Pol II (Cramer *et al.*, 2001). (B) Scheme of 12-subunit Pol II showing localisation of all subunits. (C) Ribbon model of 12-subunit Pol II (Armache *et al.*, 2003). The views correspond to the top and front views of Pol II (Cramer *et al.*, 2001).

2.2 Structures of Pol II elongation complexes

The first crystal structure of Pol II in complex with nucleic acids revealed the location of downstream DNA in the cleft, and resolved a 9-base pair DNA–RNA hybrid duplex emanating at right angles to the incoming DNA from the active site at the floor of the cleft (Gnatt *et al.*, 2001). Subsequent higher-resolution structures of core Pol II (Westover *et al.*, 2004a,b) and complete Pol II (Kettenberger *et al.*, 2004) with synthetic nucleic acid scaffolds provided a comprehensive view of the elongation complex, including nucleotides upstream of the hybrid and interactions of nucleic acids with polymerase domains involved in elongation complex maintenance (Fig. 4, and Martinez-Rucobo & Cramer, 2013).

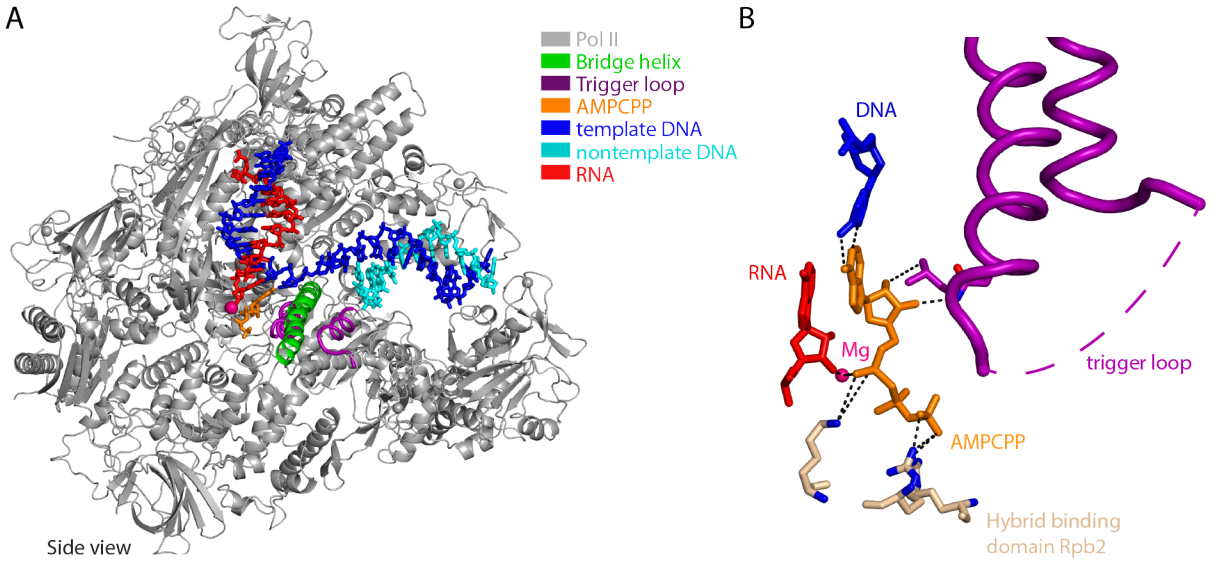

Figure 4: Structures of Pol II elongation complexes.
(A) Ribbon model of the Pol II elongation complex with bound NTP (this work). The bridge helix (green), trigger loop (purple), and metal A (pink) are highlighted. DNA template, DNA non-template, RNA, and NTP are shown in blue, cyan, red, and orange, respectively. (B) Contacts of the NTP substrate with Pol II residues, DNA template, RNA and metal A.

2.3 Structures of Pol II in complex with transcription factors

Thus far, X-ray crystallographic structural information on the complete Pol II in complex with additional factors is limited to two transcription factors, the initiation factor TFIIB (Bushnell *et al.*, 2004; Kostrewa *et al.*, 2009; Liu *et al.*, 2010; Sainsbury *et al.*, 2013), and the elongation factor TFIIS (Kettenberger *et al.*, 2003, 2004; Wang *et al.*, 2009; Cheung & Cramer, 2011). Structures of Pol II in complex with TFIIB and nucleic acids revealed how TFIIB assists in positioning and opening of the DNA, recognition of the initiator sequence and stimulation of initial RNA, followed by directing RNA to its exit tunnel (Fig. 5A). Structures of Pol II in complex with TFIIS and nucleic acids revealed how TFIIS contacts the Pol II jaw and funnel domains via its central domain (domain II), and how the C-terminal zinc ribbon domain (domain III) inserts into the Pol II pore and reaches the Pol II active site to stimulate cleavage of backtracked RNA during transcriptional proofreading and escape from arrest (Fig. 5B).

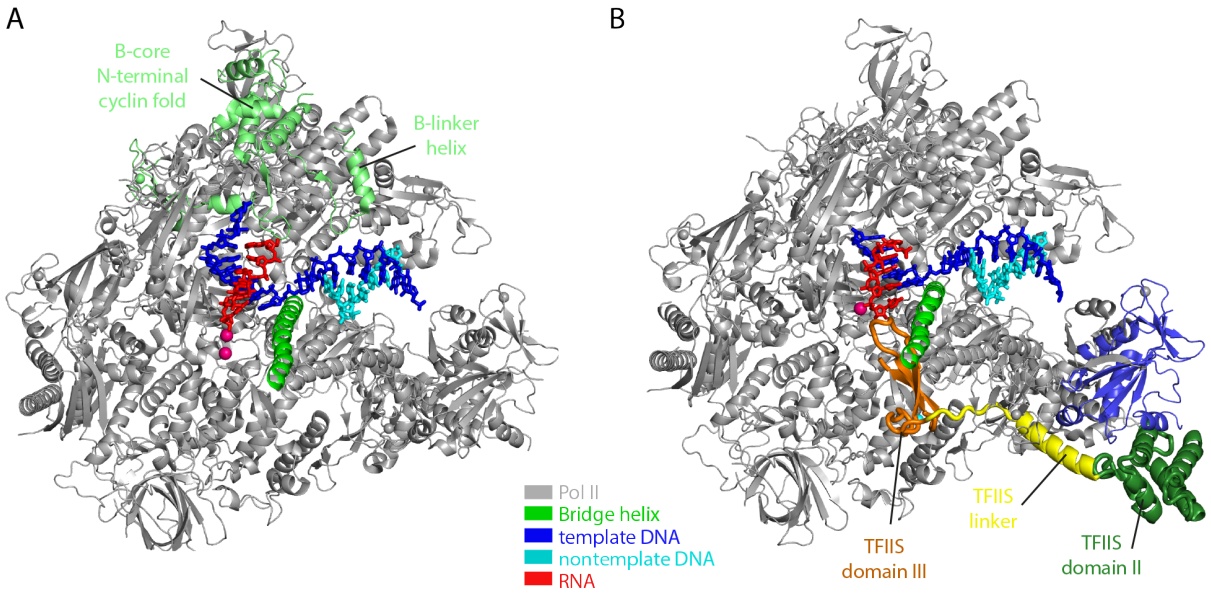

Figure 5: Structures of Pol II-transcription factor complexes.
(A) Ribbon model of the Pol II-nucleic acid-TFIIB complex (Sainsbury *et al.*, 2013). (B) Ribbon model of the Pol II-nucleic acid-TFIIS complex (Cheung & Cramer, 2011).

3 Transcription factor Bye1

In *S.cerevisiae*, there is only a single protein that contains a domain that is distantly homologous to the central, Pol II-associated domain of TFIIS. This protein, Bye1, has been identified as a multi-copy suppressor of Ess1 (Wu *et al.*, 2000), a peptidyl-prolyl cis-trans isomerase involved in proline isomerization of the C-terminal domain (CTD) of Pol II (Morris *et al.*, 1999; Hani *et al.*, 1995). In Bye1, the central TFIIS-like domain (TLD, residues 232-365) is flanked by an N-terminal PHD domain (residues 74-134) and a C-terminal SPOC domain (residues 447-547, Fig.6).

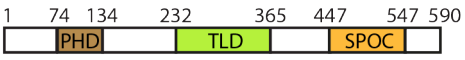

Figure 6: Domain organization of *S.cerevisiae* Bye1.
PHD: Plant Homeodomain, TLD: TFIIS-like domain, SPOC: Spen paralogue and orthologue C-terminal domain. Bordering residue numbers are indicated.

PHD domains are mostly found in proteins involved in chromatin-mediated gene regulation (Aasland *et al.*, 1995). Consistent with this, the Bye1 PHD domain binds to a histone H3 tail peptide containing trimethylated lysine 4 (H3K4me3) (Shi *et al.*, 2007). The function of SPOC domains in yeast is unclear, but in higher eukaryotes SPOC domains are implicated in

developmental signalling (Ariyoshi & Schwabe, 2003). Bye1 localizes to the nucleus (Kumar *et al.*, 2000), consistent with harbouring putative nuclear localization signals in the N-terminal protein region. Based on yeast genetics, it was suggested that Bye1 plays an inhibitory role during transcription elongation (Wu *et al.*, 2003). It is unknown whether Bye1 binds to Pol II directly, and what the consequences of such binding are for polymerase structure and function.

4 Chromatin and transcription in yeast

Packaging of eukaryotic DNA in chromatin plays an essential role in the control of transcription. Chromatin structure has important consequences in terms of availability of DNA as template for transcription, and changes in transcription are therefore correlated with changes in chromatin structure. In general, actively transcribed genes are found in decondensed chromatin, where nucleosome occupancy is reduced and an increased width and depth of nucleosome-free regions (NFR) can be observed (Jiang & Pugh, 2009; Yen *et al.*, 2012). Nucleosomes represent a major barrier for Pol II transcription *in vivo* and *in vitro*, and additional factors are required to increase accessibility for the transcription machinery (Clark & Felsenfeld, 1992; Studitsky *et al.*, 1997). So-called chromatin-modulating factors are recruited by transcription factors or by Pol II, but it is also known that Pol II passage itself can affect chromatin structure (Kulaeva *et al.*, 2009; Kulaeva & Studitsky, 2010; Yen *et al.*, 2012).

4.1 Nucleosome architecture

The nucleosome is the basic repeating unit of chromatin. It comprises an octameric assembly composed of two copies of each of the four canonical histone proteins (H3, H4, H2A, and H2B), around which 147 bp of DNA are wrapped (Fig.7, Luger *et al.*, 1997). Every 10.4 bp, histones contact the phosphate backbone of the DNA via positively charged residues. These histone-DNA contacts are weak by themselves, but provide positional stability when added together. Histone variant proteins can be incorporated into nucleosomes and provide specific functions to distinct regions of chromatin.

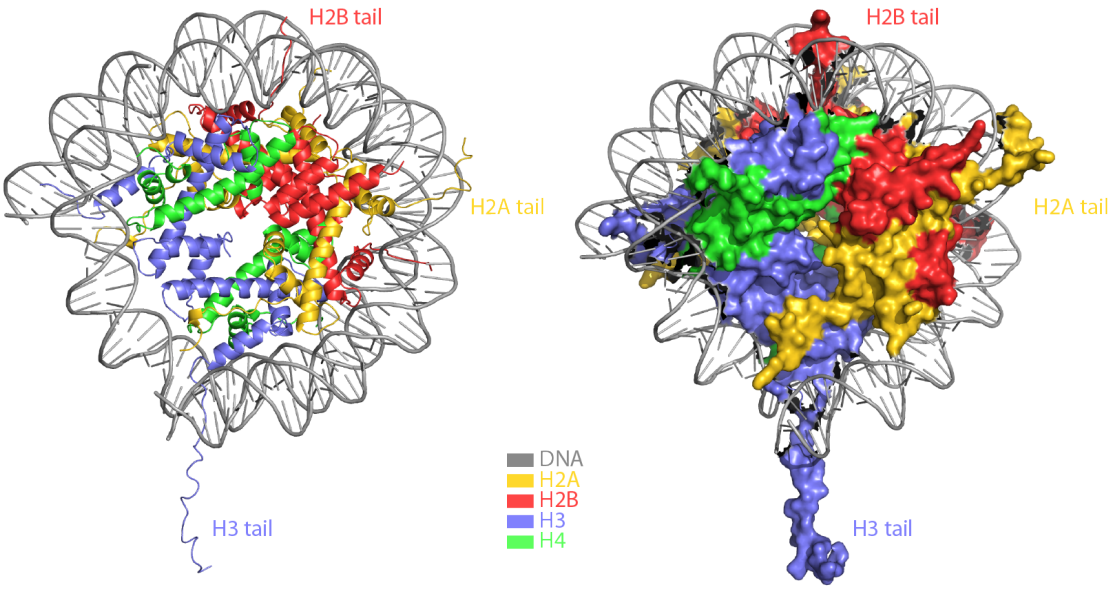

Figure 7: Ribbon (left) and surface (right) model of a core nucleosome (Luger *et al.*, 1997).

4.2 Chromatin remodelling factors

Chromatin-remodelling factors are multi-protein complexes that use the energy of ATP hydrolysis to alter nucleosome position, stability, and composition. Eukaryotic cells contain four families of chromatin-remodelling complexes (Swi/Snf, ISWI, Chd, and Ino80), which share common mechanisms of disrupting histone-DNA interactions (Clapier & Cairns, 2009). Chromatin remodellers often act in concert with transcription factors, and regulate transcription at different stages. For example, Swi/Snf-dependent remodelling has been shown to positively regulate transcription at 5' regulatory regions and during elongation, whereas ISWI-dependent remodelling results in repression of transcription by occluding promoter regions (Rando & Winston, 2012).

4.3 Histone modifications

In addition to chromatin remodelling, post-translational modifications (PTMs) of histone tails play regulatory roles during transcription. The tails of the core histones are unstructured and protrude outward from the nucleosome core and are subject to a range of reversible PTMs, including acetylation, methylation, phosphorylation, ubiquitylation and sumoylation (for an overview of histone modifications in *S.cerevisiae*, see Table 3, Smolle & Workman, 2013). While some histone modifications can directly influence chromatin structure, many of them

can be bound and recognized by proteins and protein complexes that act on chromatin. It is still strongly debated whether single histone modifications or rather the combination of different histone modifications determine the downstream effect on cellular processes such as transcription, replication and DNA repair ('histone code hypothesis', Rando, 2012).

The best studied modifications are lysine acetylation and methylation. Whereas acetylation is associated with active transcription, methylation is associated either with active or repressed chromatin, depending on the context and extent (mono-, di-, trimethylation).

Methylation of H3K4 is exclusively linked to transcriptional activation, and is carried out by the COMPASS methylase complex (Shilatifard, 2008). It consists of Set1/KMT2 and seven other polypeptides (named Cps60-Cps15), and mono-, di- and trimethylates H3K4 (Miller *et al.*, 2001; Roguev *et al.*, 2001). H3K4 methylation is localized near the transcription start site for actively transcribed genes (Krogan *et al.*, 2003; Ng *et al.*, 2003), and this localization pattern was also observed in higher eukaryotes (Bernstein *et al.*, 2005).

Several factors directly associating with di- and trimethylated H3K4 have been identified, including the CHD1 protein and proteins containing a plant homeodomain (PHD), such as the TAF3 subunit of the general transcription factor TFIID (Shi *et al.*, 2007; Pray-Grant *et al.*, 2005). In humans, H3K4me3 has been suggested to be involved in pre-mRNA maturation by bridging spliceosomal components and CHD1 to actively transcribed genes (Sims *et al.*, 2007).

Table 3: *S.cerevisiae* core histone modifications. Adapted from Smolle & Workman, 2013.
ac: acetylation, me: methylation, p: phosphorylation, sumo: sumoylation, ub: ubiquitylation

Histone	Residue	Modification	Enzyme	Recognized by	Function
H2A	K4	ac	Esa1		Transcription activation
	K7	ac	Esa1		Transcription activation
	K126	sumo	Ubc9, Siz1, Siz2		Transcription repression
H2B	K6	sumo	Ubc9, Siz1, Siz2		Transcription repression
	K7				Transcription repression
	S10	p	Ste20		Apoptosis
	K11	ac	Gcn5		Transcription activation
	K16	ac	Gcn5, Esa1		Transcription activation
	K16	sumo	Ubc9, Siz1, Siz2		Transcription repression
	K17				Transcription repression
	K123	ub		Cps35	Transcription activation
H3	R2	me		WD40	Transcription repression/activation

	K4	me	Set1	PHD, Chromo, WD40, ADD, Tudor, MBT, Zf-CW	Transcription activation
	R8	me		WD40	Transcription repression
	K9	ac	Gcn5, Rtt109	PHD, Chromo, Tudor, WD40, Ankyrin	Transcription activation
		me			Silencing
	S10	p	Snf1, Ipl1	Gcn5, 14-3-3	Transcription activation, mitosis, meiosis
	K14	ac	Gcn5, Hpa2, Esa1, Elp3, Sas2, Sas3	Bromo, PHD	Transcription activation
	R17	me		Tudor	Transcription activation
	K18	ac	Gcn5		Transcription activation
	K23	ac	Gcn5, Sas3		Transcription activation
		me		Chromo	
	R26	me			
	K27	ac	Gcn5, Rtt109		Transcription activation
		me		Chromo, WD40	Silencing
	K36	ac	Gcn5		Transcription activation
		me	Set2	Chromo, PHD, PWWP	Transcription elongation
	K56	ac	Rtt109	PHD	Transcription activation
	K79	me	Dot1	Tudor	
H4	R3	me	Rmt1	Tudor, ADD	Transcription activation
	K5	ac	Esa1	Bromo	Transcription activation
		sumo	Ubc9, Siz1, Siz2		Transcription repression
	K8	ac	Esa1, Elp3, Gcn5	Bromo	Transcription activation
		sumo	Ubc9, Siz1, Siz2		Transcription repression
	K12	ac	Hat1, Esa1, Hpa2		Transcription activation
		sumo	Ubc9, Siz1, Siz2		Transcription repression
	K16	ac	Sas2, Esa1	Bromo	Transcription activation
		sumo	Ubc9, Siz1, Siz2		Transcription repression
	K20	me		Tudor, MBT, PWWP, WD40	Transcription activation/repression
		sumo	Ubc9, Siz1, Siz2		Transcription repression

5 Aims and scope

So far, Bye1 has been linked to Pol II in genetic studies, and an association with chromatin has been proposed. However, it is unknown whether Bye1 binds Pol II directly, and what the consequences of such binding are for Pol II structure. Furthermore, it is not known if and how Bye1 affects basic Pol II mechanisms, and whether it influences Pol II transcription on a global level, or rather regulates transcription of a specific subset of genes. In addition, the functions of the individual domains have not been elucidated, and it is not clear whether they contribute to Bye1 function cooperatively, or whether they act in a timely sequence.

In order to answer these questions, we aimed to elucidate the structure of the Pol II-Bye1 complex, and thereby describe the effect of Bye1 on Pol II structure and function. Furthermore, we aimed to elucidate Bye1 function by using genetic, biochemical and biophysical assays as well as systems biology.

Part II

Materials and Methods

6 Materials

6.1 Bacterial strains

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

Strain	Base strain	Genotype	Source
E.coli XL1 Blue	K12	*rec1A; endA1; gyrA96; thi-1; hsdR17; supE44; relA1; lac*[*F' proAB lacI* qZD*M15* Tn*10* (Tetr)]	Stratagene
E.coli BL21 Gold RIL	DE3	B; F-; *ompT*; *hsdS*(rB- mB-);*dcm*+; Tetr; *gal*_(DE3); *endA*; Hte [*argU, ileY, leuW,* Camr]	Stratagene

6.2 Yeast strains

Table 5: *S.cerevisiae* strains used or generated in this study

Strain	Genotype	Source
WT	BY4741; MATa; his3D1; leu2D0; met15D0; ura3D0	Euroscarf
*bye1*Δ	BY4741; MATa; his3D1; leu2D0; met15D0; ura3D0; YKL005C::KanMX	Euroscarf
*paf1*Δ	BY4741; MATa; his3D1; leu2D0; met15D0; ura3D0; YBR279W::clonNAT	this work
*tho2*Δ	BY4741; MATa; his3D1; leu2D0; met15D0; ura3D0; YNL139C::clonNAT	this work
*bye1*Δ*paf1*Δ	BY4741; MATa; his3D1; leu2D0; met15D0; ura3D0; YKL005C::KanMX; YBR279W::clonNAT	this work
*bye1*Δ*tho2*Δ	BY4741; MATa; his3D1; leu2D0; met15D0; ura3D0; YKL005C::KanMX; YNL139C::clonNAT	this work
TAP-Bye1	BY4741; MATa; his3D1; leu2D0; met15D0; ura3D0; YKL005C::YKL005C-TAP-KanMX	Open Biosystems
*bye1*Δ (SL-screen)	BY5563; MATalpha; can1D::MFA1pr-HIS3; lyp1D; ura3D0; leu2D0; YKL005c::ClonNAT	this work

*fyv6*Δ	BY4741; MATa; his3D1; leu2D0; met15D0; ura3D0; YNL133C::clonNAT	this work
*bye1*Δ*fyv6*Δ	BY4741; MATa; his3D1; leu2D0; met15D0; ura3D0; YKL005C::KanMX; YNL133C::clonNAT	this work
pGAL1-3xHA-Bye1	BY4741; MATa; his3D1; leu2D0; met15D0; ura3D0; YKL005C::KanMX-pGAL1-HA-YKL005c	this work
TAP-Bye1 *dst1*Δ*	BY4741; MATa; his3D1; leu2D0; met15D0; ura3D0; YGL043W::clonNAT; YKL005C::YKL005C-TAP-KanMX	this work
TAP-TFIIS	BY4741; MATa; his3D1; leu2D0; met15D0; ura3D0; YGL043W::KanMX-TAP-YGL043W	Andreas Mayer
TAP-Dst1 *bye1*Δ	BY4741; MATa; his3D1; leu2D0; met15D0; ura3D0; YKL005C::clonNAT; YGL043W::KanMX-TAP-YGL043W	this work
*dst1*Δ	BY4741; MATa; his3D1; leu2D0; met15D0; ura3D0; YGL043W::KanMX	Euroscarf
*dst1*Δ*paf1*Δ	BY4741; MATa; his3D1; leu2D0; met15D0; ura3D0; YGL043W::clonNAT; YBR279W::KanMX	this work
*dst1*Δ*tho2*Δ	BY4741; MATa; his3D1; leu2D0; met15D0; ura3D0; YGL043W::KanMX; YNL139C::clonNAT	this work
*dst1*Δ*fyv6*Δ	BY4741; MATa; his3D1; leu2D0; met15D0; ura3D0; YGL043W::clonNAT; YNL133C::KanMX	this work
TAP-Rpb3	BY4741; MATa; his3D1; leu2D0; met15D0; ura3D0; YIL021W::YIL021W-TAP-KanMX	Open Biosystems
TAP-Rpb3 *bye1*Δ	BY4741; MATa; his3D1; leu2D0; met15D0; ura3D0; YKL005C::clonNAT; YIL021W::YIL021W-TAP-KanMX	this work
TAP-Paf1	BY4741; MATa; his3D1; leu2D0; met15D0; ura3D0; YBR279W::YBR279W-TAP-KanMX	Open Biosystems
TAP-Paf1 *bye1*Δ	BY4741; MATa; his3D1; leu2D0; met15D0; ura3D0; YKL005C::clonNAT;YBR279W::YBR279W-TAP-KanMX	this work
TAP-Tho2	BY4741; MATa; his3D1; leu2D0; met15D0; ura3D0; YNL139C::YNL139C-TAP-KanMX	Open Biosystems
TAP-Tho2 *bye1*Δ	BY4741; MATa; his3D1; leu2D0; met15D0; ura3D0; YKL005C::clonNAT; YNL139C::YNL139C-TAP-KanMX	this work
TAP-Fyv6	BY4741; MATa; his3D1; leu2D0; met15D0; ura3D0; YNL133C::YNL133C-TAP-KanMX	Open Biosystems
WT, RNH1 over-expression	BY4741; MATa; his3D1; leu2D0; met15D0; ura3D0; pRS 315 pGAL1-3HA RNase H1	Euroscarf
*bye1*Δ, RNH1 over-expression	BY4741; MATa; his3D1; leu2D0; met15D0; ura3D0; YKL005C::KanMX; pRS 315 pGAL1-3HA RNase H1	Euroscarf

*paf1*Δ, RNH1 over-expression	BY4741; MATa; his3D1; leu2D0; met15D0; ura3D0; YBR279W::clonNAT; pRS 315 pGAL1-3HA RNase H1	this work
*tho2*Δ, RNH1 over-expression	BY4741; MATa; his3D1; leu2D0; met15D0; ura3D0; YNL139C::clonNAT; pRS 315 pGAL1-3HA RNase H1	this work
*fyv6*Δ, RNH1 over-expression	BY4741; MATa; his3D1; leu2D0; met15D0; ura3D0; YNL133C::clonNAT; pRS 315 pGAL1-3HA RNase H1	this work
*bye1*Δ*paf1*Δ, RNH1 over-expression	BY4741; MATa; his3D1; leu2D0; met15D0; ura3D0; YKL005C::KanMX; YBR279W::clonNAT; pRS 315 pGAL1-3HA RNase H1	this work
*bye1*Δ*tho2*Δ, RNH1 over-expression	BY4741; MATa; his3D1; leu2D0; met15D0; ura3D0; YKL005C::KanMX; YNL139C::clonNAT; pRS 315 pGAL1-3HA RNase H1	this work
*bye1*Δ*fyv6*Δ, RNH1 over-expression	BY4741; MATa; his3D1; leu2D0; met15D0; ura3D0; YKL005C::KanMX; YNL133C::clonNAT; pRS 315 pGAL1-3HA RNase H1	this work

** dst1 encodes the general transcription factor TFIIS*

6.3 Plasmids and oligonucleotides

Table 6: Plasmids used in this study

Vector ID	Insert/Description	type	tag	primer (fwd+rev)
KK17	Sc Bye1 1-596 (=full-length)	pOPINF	N-His	KK41+KK44
KK51	Sc Bye1 69-370, 3C to purify TLD	pOPINI	N-His	KK121/133+KK122/134 (inserted 3C site by overlapping PCR)
	Sc Gal4-VP16	pET21b	N-His	(cloned by M. Seizl)
	native HIS4 promoter (-428 to +24 respective to the A in the start codon)	pBlue-script II KS+		(cloned by L. Lariviere)
	Sc TFIIS domain II+III (aa 131-309)	pET 28a	N-His	(cloned by H. Kettenberger)
KK101	pGAL1-3HA RNase H1	pRS 315	N-HA	KK286-289

Table 7: Oligonucleotides used for molecular cloning

Primer ID	Primer name	Sequence (5'-3')
KK41	Bye1 FL fwd	AAGTTCTGTTTCAGGGCCCGATGTCTGTCCGTAC-TTCTTCAAGA
KK44	Bye1 FL rev	ATGGTCTAGAAAGCTTTAATGGATAAATTCCTGTTTCACGA

KK71	*bye1*Δ confirmation fwd	ATATGAACACTGATGGGCACTTTAT
KK74	*bye1*Δ confirmation rev	GAAAGCCTCTGTTATTCACACATCT
KK121	Bye1 PHD 3C TLD fwd	TGCCTCTTGAAGTCCTCTTTCAGGGACCCACTAG-GAAAGATTTTGAAAGCGAAA
KK122	Bye1 PHD 3C TLD rev	CTAGTGGGTCCCTGAAAGAGGACTTCAAGAGGCA-TTTCGTTTTCGGCAGTATCA
KK129	*bye1*Δ kanMX fwd	TTATCCTAGGTTCTCTTTTACTTTTAGCTTTTATTT-AGTTTAAATTTATTCGGATCCCCGGGTTAATTAAGGCGC
KK130	*bye1*Δ kanMX rev	TTTGTTATCATATTACTATAATATTGTTACAACTCT-TAGGAATCTTATATGAATTCGAGCTCGTTTAAACTGGAT
KK133	Bye1 69 fwd	ACCATCACAGCAGCGGCATGGATGAGGGCTATG-TTAGATGTCTG
KK134	Bye1 370 rev	ATGGTCTAGAAAGCTTTATTGTCAGGTACCTCCACGATAAAA
KK165	*paf1*Δ confirmation fwd	ATATCGAAGTTGACCATACGAAAAG
KK173	*tho2*Δ clonNAT fwd	AACGGTTTCAGTTGATACATATTCGCACCAGTATA-CATTTTCAGGACTTTTAACCTTATGTATCATACACA-TACGATTTAGGTGAC
KK174	*tho2*Δ clonNAT rev	CTATCAAAGTACACGTTAAAATTCAGCTCGGGTAT-GTTAAGTACTAGTAATTAGGGGCAGGGCATGCTCATGTAG
KK201	*paf1*Δ clonNAT fwd	AATGTATTCAGTACAATAGAACAGTGCTCATAATA-GTATAAAGGGTCACATAACCTTATGTATCATACAC-ATACGATTTAGGTGAC
KK202	*paf1*Δ clonNAT rev	GTAAAAAGAACTACAGGTTTAAAATCAATCTCCCT-TCACTTCTCAATATTTTAGGGGCAGGGCATGCTCATGTAG
KK212	*tho2*Δ confirmation fwd	ATTTACATGTTCTGAATGAGAAGGC
KK290	*tho2*Δ confirmation rev	CGATAAAAGAAAGAACGGTTTGTTA
KK291	*paf1*Δ confirmation rev	TATTTGTGTCAGTGTAACTGCTGGT
KK123	Bye1 ko ClonNAT fwd	TTATCCTAGGTTCTCTTTTACTTTTAGCTTTTATTT-AGTTTAAATTTATTTAACCTTATGTATCATACACAT-ACGATTTAGGTGAC
KK124	Bye1 ko ClonNAT rev	TTTGTTATCATATTACTATAATATTGTTACAACTCT-TAGGAATCTTATATTTAGGGGCAGGGCATGCTCATGTAG

KK125	Dst1 ko ClonNAT fwd	CAAGCAGCAGAACATTCACAGTGTAGTCAGTCCG-CATAAGAGCATTCATCTAACCTTATGTATCATACAC-ATACGATTTAGGTGAC
KK126	Dst1 ko ClonNAT rev	TCCTATTGTTTCCTTTGCTCTGTCCGTGGCTACTC-TCATCTTTTTCTATTTTAGGGGCAGGGCATGCTCATGTAG
KK138	Dst1 ko confirmation fwd	GGTTCTCTTAGGAAATTCTAAACGC
KK141	Dst1 ko confirmation rev	ACGAACAAAAGATTTTACGTGAGAC
KK164	Fyv6 ko confirmation fwd	GATATACCTGGGAAGTTACAATGGA
KK165	Fyv6 ko confirmation rev	ATATCGAAGTTGACCATACGAAAAG
KK244	Bye1 KanMX6-PGAL1-3HA fwd	ATCAGAAAACCACTTTGTAACAAATATGAACACTG-ATGGGCACTTTATTATTCTTTGACAGAAAGAAATA-AAAAAGCGATGAATTCGAGCTCGTTTAAACTGGATG
KK245	Bye1 KanMX6-PGAL1-3HA rev	GCTTCAGTCTCCTCCTGTAACAAATACTCAATATAT-TTATTCTGACCCTTATTAGATCTTGAAGAAGTAC-GGACAGAAGCGTAATCTGGAACGTCATATGG
KK286	pGAL1 3HA RNH1 fwd	agctccaccgcggtggcggccgcACGGATTAGAAGCCGCCG
KK287	pGAL1 3HA RNH1 rev	ttcctgcagcccgggggatccTTATCGTCTAGATGCTCCTTTC
KK288	pGAL1 3HA RNH1 overlap fwd	GTTCCAGATTACGCTATGGCAAGGCAAGGGAAC
KK289	pGAL1 3HA RNH1 overlap rev	GTTCCCTTGCCTTGCCATAGCGTAATCTGGAAC

Table 8: Oligonucleotides used for crystallization and assays

type	sequence 3'-5'	application
template DNA	TTA CTG GBC CTT ATT CAT GAA CTC GA	tailed template
non-template DNA	TC GAG TTC ATG AAA	
RNA	AG GAC CAG CUU	
template DNA	CAA TCT CGT BGT GGT CCA TTC GAT CGA	arrest template
non-template DNA	TCG ATC GAA TGG AG	
pSH515-5'Cy5	CTC GAA CCG ACA ACG GGC AGA GTG ACC ACT T-Cy5	transcription assay
template DNA	TTG CTG GTC CGG ATT CAT GAA CTC GA	mismatch template
non-template DNA	TC GAG TTC ATG AAT	
RNA	AG GAC CAG CUU UAC UU	

6.4 Reagents and consumables

Chemicals were obtained from Merck (Darmstadt, Germany), Roth (Karlsruhe, Germany) or Sigma-Aldrich (Seelze, Germany), unless stated otherwise. Enzymes and reagents for cloning were obtained from Fermentas (St. Leonrot, Germany), New England Biolabs (Frankfurt am Main, Germany) and Clontech (St Germain-en-Leye, France). For DNA preparation, commercial kits from Qiagen (Hilden, Germany) were used. DNA and RNA oligonucleotides were ordered at ThermoScientific (Ulm, Germany) and Metabion (Planegg, Germany), respectively. Crystallization reagents and tools were ordered at Hampton Research (Aliso Viejo, CA, USA) and Qiagen (Hilden, Germany). Reagents and consumables were ordered at standard laboratory suppliers.

6.5 Media and additives

Table 9: Media for *E.coli* and *S.cerevisiae*

Media	Application	Description
LB	*E.coli* culture	1% (w/v) tryptone; 0.5% (w/v) yeast extract; 0.5% (w/v) NaCl
SOB	*E.coli* transformation	2% (w/v) tryptone; 0.5% (w/v) yeast extract; 8.55 mM NaCl; 2.5 mM KCl; 10 mM $MgCl_2$
SeMet	*E.coli* culture	22 g/l SeMet base; 5 g/l nutrient mix; 40 µg/ml SeMet (all components Molecular Dimensions)
YPD	Yeast culture	2% (w/v) peptone; 2% (w/v) glucose; 1% (w/v) yeast extract

YPDS	Yeast culture	2% (w/v) peptone; 2.2% (w/v) glucose; 1% (w/v) yeast extract; 18.25% (w/v) Sorbitol
YPGal	Yeast culture	2% (w/v) peptone; 2% (w/v) galactose; 1% (w/v) yeast extract
SC	Yeast culture	0.69% (w/v) nitrogen base; 0.6% (w/v) CSM amino acid drop-out mix; 2% (w/v) glucose
GNA	SL screen	3% (w/v) Difco nutrient broth; 1% (w/v) yeast extract; 2.5% (w/v) agar; 5% (w/v) glucose; 20 µg/ml tetracycline
sporulation media	SL screen	0.1% (w/v) yeast extract; 2.5% (w/v) agar; 2% (w/v) glucose; 1% (w/v) K Acetate; 30 mg/L histidine; 30 mg/L uracil; 30 mg/L leucin; 20 µg/ml tetracycline
SD-H	SL screen	0.17% (w/v) yeast nitrogen base w/o aa w/o $(NH_4)_2SO_4$; 2.5% (w/v) agar; 2% (w/v) glucose; 0.5% (w/v) $(NH_4)_2SO_4$; 0.2% (w/v) drop-out powder w/o histidine, arginine, leucine; 50 µg/ml Canavanine; 50 µg/ml S-(2-NH_2ethyl)-L-Cys; 20 µg/ml tetracycline

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

Additive	**Description**	**Stock solution**	**applied concentration**
IPTG	*E.coli* induction	1 M in H_2O	0.25-0.5 mM
Ampicillin	Antibiotic	100 mg/ml in H_2O	100 µg/ml
Chloramphenicol	Antibiotic	50 mg/ml in EtOH	50 µg/ml
6-AU	reduction of ribonucleotide levels	100 mg/ml in DMSO	50-200 µg/ml
MPA	reduction of ribonucleotide levels	powder	15-45 µg/ml
HU	replication fork stalling	powder	100-200 mM
Phleomycin	DNA intercalation	20 mg/ml	1.0-1.5 µg/ml
Etoposide	Topoisomerase II inhibition	powder	0.1-0.25 mM
MMS	replication fork stalling	100%	0.01-0.02% (v/v)

6.6 Buffers and solutions

Table 11: General buffers, dyes and solutions

Name	Description/Source	Application
Electrophoresis buffer	NuPAGE buffer, life technologies	SDS-PAGE
5X SDS sample buffer	250 mM Tris, pH 6.8 at 20°C; 50% (v/v) glycerol; 0.03% (w/v) bromophenol blue; 7.5% (w/v) SDS; 100 mM DTT	SDS-PAGE
Gel staining solution	Instantblue, expedion	Coomassie staining
Silver nitrate solution	0.2% $AgNO_3$, 0.05% formalin (35% formaldehyde)	Silver staining
Developing solution	6% Na_2CO_3; 0.05% formalin (35% formaldehyde)	Silver staining
Transfer buffer	NuPAGE TransferBuffer, life technologies	Western blotting
PBS buffer	137 mM NaCl; 2.7 mM KCl; 10 mM Na_2HPO_4; 1.76 mM KH_2PO_4; pH 7.4	Western blotting
100X PI	0.028 mg/ml leupeptin; 0.137 mg/ml pepstatin A; 0.017 mg/ml PMSF; 0.33 mg/ml benzamidine; in 100% EtOH p.a.	Protease inhibitor mix
10X TAE	50 mM EDTA, pH 8.0; 2.5 M Tris-acetate	Agarose gels
TE	1 mM Tris pH 8.0 at 20°C; 0.1 mM EDTA	dissolving DNA
2X transcription buffer	40 mM HEPES, pH 7.6; 120 mM $(NH_4)_2SO_4$; 16 mM $MgSO_4$; 20 µM $ZnCl_2$; 20 % (v/v) glycerol; 20 mM DTT	RNA extension assay
10X TBE	900 mM Tris; 900 mM boric acid; 20 mM EDTA, pH 8.0	RNA extension assay
2X urea loading buffer	20 % (v/v) 10 × TBE; 8 M RNase free urea	Urea PAGE
2X urea loading dye	20 % (v/v) 10 × TBE; 8 M RNase free urea; 0.03 % (w/v) bromophenol blue; 0.03 % (w/v) xylene cyanol FF	Urea PAGE
5X Acetate transcription buffer	500 mM KAc, pH 7.6; 100 mM HEPES, pH 7.6; 5 mM EDTA; 25 mM MgOAc	Transcription assay
5X primer annealing buffer	25 mM Tris, pH 8.3 at 20°C; 375 mM KCl; 5 mM EDTA, pH 8	Transcription assay
5X synthesis buffer	250 mM Tris, pH 8.3 at 20°C; 375 mM KCl; 22.5 mM $MgCl_2$; 75 mM DTT	Transcription assay
TFB-I	30 mM KaAcetate; 50 mM Mn_2Cl; 100 mM RbCl; 10 mM $CaCl_2$; 15 % (v/v) glycerol	Chemically competent *E.coli*
TFB-II	10 mM MOPS, pH 7.0; 75 mM $CaCl_2$; 10 mM RbCl; 15% (v/v) glycerol	Chemically competent *E.coli*
TeLit, pH 8.0	155 mM LiOAc; 10 mM Tris, pH 8.0 at 20°C; 1 mM EDTA, pH 8.0	chemically competent yeast cells

LitSorb	18.2% (w/v) Sorbitol in TeLit, pH 8.0	chemically competent yeast cells
LitPEG	40% (w/v) PEG 3350 in TeLit, pH 8.0	chemically competent yeast cells
One-step buffer	0.2 M LiOAc; 40% (w/v) PEG 3350; 100 mM DTT	chemically competent yeast cells

Table 12: Pol II purification buffers

Name	Description
3X freezing buffer	150 mM Tris, pH 7.9 at 4°C; 3 mM EDTA; 30 µM $ZnCl_2$; 30 % (v/v) glycerol; 3 % (v/v) DMSO; 30 mM DTT; 3X PI
HSB150	50 mM Tris, pH 7.9 at 4°C; 150 mM KCl; 1 mM EDTA; 10 µM $ZnCl_2$; 10 % (v/v) glycerol; 10 mM DTT; 1X PI
HSB1000/7	50 mM Tris, pH 7.9 at 4°C; 1000 mM KCl; 7 mM imidazole; 1 mM EDTA; 10 µM $ZnCl_2$; 10 % (v/v) glycerol; 2.5 mM DTT; 1X PI
Ni buffer	20 mM Tris, pH 7.9 at 4°C; 150 mM KCl; 7/50/100 mM imidazole
MonoQ 150	20 mM Tris-acetate, pH 7.9 at 4°C; 150 mM KOAc; 10 % (v/v) glycerol; 0.5 mM EDTA, pH 7.9; 10 µM $ZnCl_2$; 10 mM DTT
MonoQ 2000	20 mM Tris-acetate, pH 7.9 at 4°C; 2000 mM KOAc; 10 % (v/v) glycerol; 0.5 mM EDTA, pH 7.9; 10 µM $ZnCl_2$; 10 mM DTT
Pol II buffer	5 mM HEPES, pH 7.25; 40 mM $(NH_4)_2SO_4$; 10 µM $ZnCl_2$; 10 mM DTT

Table 13: Nuclear extract preparation buffers

Name	Description
Resuspension buffer	50 mM Tris, pH 7.5 at 4°C; 20 mM EDTA; 30 mM DTT
Cell lysis buffer	18% Polysucrose 400; 10 mM Tris, pH 7.5 at 4°C; 20 mM K Acetate; 5 mM Mg Acetate; 1 mM EDTA pH 8.0; 0.5 mM Spermidine; 0.15 mM Spermine; 3 mM DTT; 1X PI
Centrifugation buffer	100 mM Tris, pH 7.9 at 4°C; 50 mM K Acetate; 10 mM $MgSO_4$; 20% Glycerol; 2 mM EDTA
Dialysis buffer	20 mM HEPES, pH 7.6; 10 mM $MgSO_4$; 1 mM EGTA, pH 8.0; 20% Glycerol

Table 14: Peptide array buffers

Name	Description
printing buffer	1X protein printing buffer (Arrayit), 1% (w/v) BSA, 5 mg/mL fluorescein-labeled biotin
hybridization buffer	PBS, pH 7.6; 5% (w/v) BSA, 0.1% Tween-20

Table 15: Chromatin immunoprecipitation buffers

Name	Description
TBS	20 mM Tris, pH 7.5 at 4°C; 150 mM NaCl
FA-lysis buffer	50 mM HEPES, pH 7.5; 150 mM NaCl; 1 mM EDTA; 1% (v/v) Triton X-100; 0.1% (v/v) Na deoxycholate; 0.1% (v/v) SDS
ChIP wash buffer	10 mM Tris, pH 8 at 4°C; 0.25 M LiCl; 1 mM EDTA; 0.5% (v/v) NP-40; 0.5% (v/v) Na deoxycholate
ChIP elution buffer	50 mM Tris, pH 7.5 at 25°C; 10 mM EDTA; 1% (v/v) SDS

7 Methods

7.1 General methods

7.1.1 Preparation and transformation of competent cells

Chemically competent *E.coli* cells for cloning Chemically competent *E.coli* cells were prepared from LB pre-cultures grown at 37°C ON. 400 ml LB medium (supplemented with antibiotics if required) was inoculated with the pre-culture and grown at 37°C until an OD_{600} of 0.5 was reached. Cultures were then cooled down on ice and centrifuged for 10 min at 4,000 rpm at 4°C. Pellets were resuspended in 100 ml TFB-I on ice and centrifuged as before. Pellets were then resupended in 8 ml TFB-II on ice. 50 µl aliquots were prepared and immediately frozen in liquid nitrogen.

Cells were transformed by heat-shock using 100 ng vector, or 2.5 µl InFusion product. Cells were incubated on ice for 5 min prior to heat-shock at 42°C for 90 s. Cells were then cooled on ice for 5 min and grown in 1 ml LB medium at 37°C for 1 hour shaking in a thermomixer (Qiagen) for recovery. Finally, cells were plated on selective plates and incubated ON at 37°C.

Electro-competent *E.coli* cells for protein expression Electro-competent *E.coli* cells were prepared LB pre-cultures grown at 37°C ON. 1 L SOB medium (supplemented with antibiotics if required) was inoculated with the pre-culture and grown at 37°C until an OD_{600}

of 0.5 was reached. Cultures were then cooled down on ice and centrifuged for 10 min at 4,000 rpm at 4°C. Pellets were resuspended in 100 ml ice-cold and sterile H_2O and centrifuged as before. Pellets were then resusupended in 8 ml ice-cold 10% glycerol (v/v). 50 µl aliquots were prepared and immediately frozen in liquid nitrogen.

Cells were transformed by electroporation using 100 ng vector. Cells were incubated in a cuvette on ice for 5 min prior to electroporation. Electroporation was carried out at 2.5 kV in a MicroPulser Electroporator (Biorad). Cells were then cooled on ice for 5 min and grown in 1 ml LB at 37°C for 1 hour shaking in a thermomixer (Qiagen) for recovery. Finally, cells were plated on selective plates and incubated over night at 37°C, or were used directly for inoculation of pre-cultures.

Chemically competent yeast cells for homologous recombination Chemically competent *S.cerevisiae* cells were prepared from YPD pre-cultures grown at 30°C ON. 50 ml YPD medium was inoculated with the pre-culture at a starting OD_{600} of 0.2 and grown at 30°C until an OD_{600} of 0.5 was reached. Cultures were then centrifuged for 5 min at 4,000 rpm at RT. Pellets were resuspended in 25 ml sterile H_2O, 5 ml LitSorb and 360 µl LitSorb, with centrifuging inbetween. Salmon sperm DNA was boiled for 10 min and immediately cooled down on ice. 40 µl salmon sperm DNA was added to the cell solution and 50 µl aliquots were prepared for immediate use.

Cells were transformed by adding 1 µg PCR product to 50 µl competent cells. 360 µl LitPEG was added and the mixture was incubated for 30 min at RT. 47 µl DMSO was added and the mixture was incubated for 15 min at 42°C, before spinning down for 3 min at 2,000 rpm at RT. Cells were then resuspended in 1 ml YPD and grown at 30°C for 3 h shaking in a thermomixer (Qiagen) for recovery. Finally, cells were plated on selective plates and incubated for 2-3 days at 30°C.

Chemically competent yeast cells for plasmid transformation Chemically competent *S.cerevisiae* cells were prepared from fresh YPD plates grown ON. A small amount of cells was resuspended in 1 ml sterile H_2O and spun down at 13,000 rpm for 1 min. The pellet was then resuspended in 100 µl One-step buffer and vortexed vigorously. Salmon sperm DNA was boiled for 10 min, immediately cooled down on ice and 10 µl were added to the cell solution. After addition of 500 ng plasmid DNA, cells were incubated for 30 min at 45°C. Cells were washed with 1 ml YPD and immediately plated on selective plates and incubated for 2-3 days at 30°C.

7.1.2 Molecular cloning

Polymerase Chain Reaction Primers for InFusion cloning (Clontech) were designed using an overhang of 15-20 nucleotides at the 5' end, followed by 20 nucleotides complementary to the gene of interest. Purification tags and protease cleavage sites were introduced either by in-frame cloning into according vectors or by PCR. Protease cleavage sites within the gene of interest were introduced by overlapping PCR. For this, two overlapping PCR products are synthesized with primers carrying the desired protease cleavage site. In a second PCR reaction, these products were used as template to synthesize the gene of interest containing the protease cleavage site. PCR reactions were carried out using the Phusion High Fidelity PCR Master Mix (NEB), 0.5 µM of each primer, 5% (v/v) DMSO and variable amounts of DNA template, either genomic or plasmid DNA.

Thermocycling programs comprised 30 cycles and were adjusted to the specific needs of the individual reactions in terms of annealing temperature and elongation times (Biometra T3000 Thermocycler). PCR products were visualized on 0.5-1% (w/v) agarose gel electrophoresis and staining with Sybr Safe (Invitrogen). Purification of PCR products was carried out using the QIAquick gel extraction kit (Qiagen).

Enzymatic restriction cleavage Vectors were digested using restriction endonucleases (NEB, Fermentas) as recommended by the manufacturer. Cleaved vectors were purified using the QIAquick PCR purification kit (Qiagen).

Restriction- and Ligation-free InFusion cloning InFusion cloning (Clontech) allows for restriction and ligation-free cloning. 50 ng of digested vector, 3 fold molar amount of insert, and InFusion reaction mixture was incubated as recommended by the manufacturer. 2.5 µl of InFusion reaction were then transformed as described in 7.1.1.

Homologous recombination In order to insert protein tags genomically or delete genes of interest in yeast, homologous recombination *in vivo* was carried out. For deletion of genes, cassettes containing a resistance marker with 25 bp homologous to the up- and downstream region of the gene of interest at it's 5' and 3' end were synthesized by PCR and transformed into yeast as described in 7.1.1. For insertion of protein tags, cassettes were synthesized as above, additionally containing sequences for protein tags.

After transformation into yeast cells, single colonies were re-streaked on selective plates and grown for 2-3 days at 30°C. A small amount of cells was then resuspended in 100 µl 0.2 M

NaOH, and 50 µl Zirconia beads (Biospec) were added. The mixture was incubated at 95°C for 5 min in a thermomixer at maximum speed. The mixture was then centrifuged for 5 min at max speed and 5 µl of supernatant were used as template for a PCR reaction carried out as described in 7.1.2 to confirm correct genomic insertion.

Isolation and verification of plasmid DNA After transformation of plasmids into *E.coli* cells, single colonies from selective plates were picked and used for colony PCRs. For this, colonies were inoculated in 30 µl H_2O and mixed for 10 min at RT in a thermomixer. 6 µl of the mixture were then used as template for a PCR reaction carried out as described in 7.1.2. Colonies resulting in positive PCR reactions were used to inoculate 5 ml ON cultures. Plasmids were isolated from ON cultures using Miniprep purification kits (Qiagen) and were verified by sequencing (GATC).

After transformation of plasmids into yeast cells, single colonies were re-streaked on selective plates and grown for 2-3 days at 30°C to confirm successful transformation.

Electrophoretic separation of DNA Electrophoretic separation of DNA was carried out in horizontal 1X TAE agarose gels containing 0.7 µg/ml Sybr Safe (Invitrogen) and varying concentrations of agarose, depending on the size of the molecules to be separated. Separation was carried out in PerfectBlue Gelsystem electrophoresis chambers (Peqlab). Samples were mixed with 6X loading dye (Fermentas) prior to loading and DNA was visualized and documented using a Safe Imager blue light transilluminator (Invitrogen, λ=470 nm).

7.1.3 Over-expression of proteins in *S.cerevisiae*

For genomic over-expession, a cassette containing the Gal-promoter sequence and a 3X HA tag sequence was introduced upstream of the ORF by homologous recombination (described in 7.1.1). For over-expression from a plasmid, the Gal-promoter sequence, the 3X HA tag sequence and the ORF of the protein of interest were cloned into pRS315 and the plasmid was transformed into *S.cerevisiae* as described in 7.1.1. Expression was induced by growing transformed cells in YPGal medium and confirmed by Westernblotting using an anti-HA antibody (Roche) as described in 7.1.5.

7.1.4 Protein expression in *E.coli*

Recombinant proteins were expressed in *E.coli* BL21-CodonPlus (DE3) RIL. For this, plasmids encoding proteins of interest were transformed as described in 7.1.1. Cells were grown in

LB medium containing required antibiotics at 37°C until an OD_{600} of 0.8-0.9 was reached. Expression was then induced by addition of varying amounts of IPTG and shifting cells to 20°C for 18 h. Cells were harvested by centrifugation at 4,000 rpm for 10 min and pellets were stored at -20°C.

7.1.5 Protein analysis

Determination of protein concentration Total protein concentrations were determined measuring the absorption at 280 nm using a ND-100 (NanoDrop) spectrophotometer. Individual molar absorption coefficients of proteins measured were calculated using the ProtParam software (Gasteiger *et al.*, 2005). If samples contained large amounts of non-protein molecules absorbing at 280 nm, such as nucleic acids and detergents, protein concentrations were determined by Bradford assay (Bradford, 1976). For this, the dye reagent (BioRad) was used according to manufacturer's instructions and samples were measured at 595 nm. Reference curves were generated for each new batch of dye reagent using bovine serum albumin (Fraction V, Roth).

SDS-Polyacrylamide gel electrophoresis Electrophoretic separation of proteins was carried out using Novex NuPAGE gradient SDS-PAGE (life technologies) according to manufacturer's instructions. Gels were stained with Instantblue (expedion) for up to 1 hour before transferring to H_2O. For more sensitive detection of proteins, silver staining was applied. For this, the gel was fixed in 50% Ethanol for 20 min, followed by 5% Ethanol for another 20 min. After incubating for 5 min in 35 µM DTT, the gel was sensitized in silver nitrate solution for 10 min, washed twice with H_2O and developed in developing solution. The reaction was stopped by addition of 5% HOAc.

Western blotting and immunostaining For western blotting and immunostaining, protein samples were separated on a SDS-PAGE as described (7.1.5). Western blotting onto a PVDF Immobilon membrane (Millipore) was carried out using a Trans-Blot transfer cell (Biorad) at 150 V for 90 min at 4°C. The membrane was blocked for 20 min at RT with 2% milk (w/v) in 1X PBS containing 0.02% Tween 20 (PBS-T), followed by incubation with the primary antibody in 2% milk in PBS-T for 60 min. The membrane was then washed 3X for 5 min with 2% milk in PBS-T before incubating with the secondary antibody for 60 min. After washing the membrane as described before, Pierce ECL Western Blotting Substrate (Thermo Scientific) was added and Amersham Hyperfilms ECL (GE Healthcare) were developed until clear signals

appeared.

7.1.6 Scaffold preparation

To anneal DNA-RNA-scaffolds for crystallization and *in vitro* assays, oligonucleotides were dissolved separately in TE at a final concentration of 400 µM. Equimolar amounts of non-template DNA, template DNA and RNA were mixed and TE was added to reach a final concentration of 100 µM. Scaffolds were annealed in a Thermocycler (Biometra T3000 Thermocycler) using the following program:

time (s)	**Temp or ΔTemp**	
preheat lid	99°C	
90	95°C	
90	95°C	
90	-1°C	75 cycles
∞	4°C	

Annealed scaffolds were stored at -20°C.

7.1.7 Crystallization screening

Initial crystallization screening was performed by the Crystallization Facility at the MPI of Biochemistry, Martinsried. Using various commercial as well as in-house produced screens, sitting-drops were set up at 20°C with a drop size of 100 nl. Initial crystals were refined in 48-well sitting drop plates (Molecular Dimensions) manually by varying pH, precipitant, concentration of components, temperature and drop-size.

7.1.8 Bioinformatic tools

Protein and gene sequences were retrieved from the NCBI or *Saccharomyces cerevisiae* genome (SGD) databases. Sequence data was visualized and processed using Ape (Davis & Hammarlund, 2006). Bioinformatic analysis was performed using the Bioinformatic Toolkit (Biegert *et al.*, 2006). Multiple sequence alignments were generated using MUSCLE (Edgar, 2004) and Aline (Bond & Schüttelkopf, 2009). Protein secondary structures were predicted by HHpred (Söding *et al.*, 2005) and psipred (Jones, 1999).

7.2 Specific methods

7.2.1 Protein expression and purification

Expression and purification of recombinant full-length Bye1 Full-length Bye1 was cloned into pOPINF with an N-terminal hexahistidine tag and expressed in *E. coli* BL21 (DE3) (Novagen). Protein expression was carried out as described in 7.2.1 and expression was induced with 0.25 mM IPTG. Protein was purified by nickel affinity, anion exchange and size-exclusion chromatography. Cells were lysed by sonication in buffer A (20 mM Tris, pH 7.5 at 20°C; 100 mM NaCl; 10 µM $ZnCl_2$; 10% (v/v) glycerol; 5 mM DTT), supplemented with 20 mM imidazole, 1 u/µl DNase (Fermentas) and 1X PI. After centrifugation at 16,000xg for 20 min, the cleared lysate was applied to a pre-equilibrated (buffer A) Ni-NTA agarose column (Qiagen). The column was washed with 10 CV of buffer A containing 20 mM imidazole before step-wise elution of the protein with buffer A containing 50/100/200 mM imidazole. Fractions containing Bye1 were pooled and applied to a MonoQ 10/100 GL column (GE healthcare) equilibrated in buffer A. The protein was eluted with a linear gradient from 100 mM to 1 M NaCl (buffer B: 20 mM Tris, pH 7.5 at 20°C; 1 M NaCl; 10 µM $ZnCl_2$; 10% (v/v) glycerol; 5 mM DTT) over 10 CV. To remove any minor contaminants a final size exclusion step using a Superdex 200 10/300 GL column (GE Healthcare) in buffer A was carried out.

Expression and purification of recombinant Bye1 TLD Bye1 TLD (residues 225-370) was expressed as a larger variant (residues 69-370) containing a protease cleavage site at the N-terminal border of the TLD, cloned into pOPINI with an N-terminal hexahistidine tag. The protein was expressed and purified as above except that expression was induced with 0.5 mM IPTG, buffers did not contain glycerol and the protein was eluted from the Ni-NTA column with 200 mM imidazole. After ion exchange purification, 300 µg precision protease was added and cleavage was carried out ON at 4°C. To separate the cleavage products, the protein was applied to a pre-equilibrated (buffer A) Ni-NTA column. Bye1 TLD could be collected in the flow-through fraction and was then applied to size-exclusion chromatography using a Superdex 75 10/300 GL column (GE Healthcare).

Expression and purification of selenomethionine-substituted Bye1 Selenomethionine (SeMet)-substituted Bye1 was expressed in SeMet medium (6.5). 2 l of SeMet medium were inoculated with an ON preculture and grown at 37°C until an OD_{600} of 0.6 was reached. Temperature was reduced to 20°C and after 30 min, 25 µg/ml SeMet, 50 µg/ml Lys/Thr/Phe, 25 µg/ml Leu/Ile/Val and 0.25 mM IPTG were added. Expression was carried out ON and

protein purification was performed as described above.

Purification of endogenous RNA Polymerase II

Fermentation 50 ml YPD preculture were inoculated from a fresh plate and incubated shaking for 10 h at 30°C. 500 ml YPD preculture were inoculated with 25 ml of the previous preculture and incubated shaking for 14 h at 30°C. 15-17 l YPD culture were inoculated with 500 ml of the second preculture and grown in a small fermenter for 10-11 h at 30°C, stirring at 600 rpm and with an air supply of 800 l/h. 200 l YPD culture were then inoculated with a starting OD_{600} of 0.2 and grown at 30°C stirring at 200 rpm with an air supply of 25 l/h in a large fermenter until OD_{600} of 8.5 was reached. Cells were harvested using a flow-through centrifuge. 1 kg of cell pellet was resuspended in 330 ml 3X freezing buffer at 4°C, flash-frozen in liquid nitrogen in 200 ml aliquots and stored at -80°C.

Protein purification Cells were lysed using a bead beater (Hamilton Beach) filling 200 ml of cell suspension, 200 ml of glass beads and 1 ml 100X PI into a metal chamber. Lysis was carried out for 80 min at 4°C with cycles of 30 s on, 90 s off. Cell lysate was centrifuged twice for 30 min at 9,000 rpm at 4°C and ultracentrifuged for 90 min at 24,000 rpm at 4°C. Proteins were then precipitated with 50% (w/v) $(NH_4)_2SO_4$ stirring ON at 4°C, followed by centrifugation twice for 45 min at 15,000 rpm at 4°C. The pellet was resuspended in 140 ml HSB 150 per 100 g pellet and conductivity was set to that of buffer HSB1000/7. The sample was then applied to pre-equilibrated Ni-NTA resin, washed with 5 CV HSB 1000/7 and 3 CV Ni buffer 7, and eluted with 3 CV Ni buffer 50 and 3 CV Ni buffer 100. Elution fractions were pooled and conductivity was adjusted to that of buffer MonoQ150. Sample was loaded to a pre-equilibrated MonoQ 10/100 (GE Healthcare) and eluted with a linear gradient over 12 CV to 75% buffer MonoQ2000. Fractions containing Pol II were pooled, diluted with 3 V Pol II buffer and concentrated to around 1 ml. 4-fold molar excess of Rpb 4/7 (for purification protocol, see 7.2.1) was added and incubated on ice for 45 min. Sample was applied twice to a pre-equilibrated Superose 6 10/300 (GE Healthcare) and fractions containing Pol II were concentrated to 3.0-3.4 mg/ml, flash-frozen in liquid nitrogen in 20 µl aliquots and stored at -80°C.

Expression and purification of Rpb4/7 Protein expression was carried out as described in 7.2.1 and expression was induced with 0.5 mM IPTG. Cells were lysed by sonication in buffer A (50 mM Tris, pH 7.5 at 4°C; 150 mM NaCl; 10 mM ß-mercaptoethanol) containing 1X PI.

After centrifugation at 16,000xg for 20 min, the cleared lysate was applied to a pre-equilibrated (buffer A) Ni-NTA agarose column (Qiagen). The column was washed with 5 CV buffer A, 3 CV buffer A containing 2 M NaCl, 3 CV buffer A containing 10 mM imidazole and 3 CV buffer A containing 20 mM imidazole. The protein was eluted with 3 CV buffer A containing 50 mM imidazole and 6 CV buffer A containing 200 mM imidazole. Fractions containing Rpb4/7 were pooled and applied to a Source15Q 16/10 column (GE healthcare) equilibrated in buffer A containing 100 mM NaCl. The column was washed with 10 CV buffer A containing 100 mM NaCl and the protein was eluted with a linear gradient to 1 M NaCl over 10 CV. To remove any minor contaminants, a final size exclusion step using a HiLoad 26/60 Superdex75 pg column (GE Healthcare) in Pol II buffer was carried out.

Expression and purification of TFIIS domain II+III TFIIS domain II+III (aa 131-309) was cloned into pET 28a with an N-terminal hexahistidine tag and expressed in *E. coli* BL21 (DE3) (Novagen). Protein expression was carried out as described in 7.2.1 and expression was induced with 0.5 mM IPTG. Protein was purified by nickel affinity and anion exchange. Cells were lysed by sonication in buffer A (25 mM HEPES, pH 7.5 at 20°C; 300 mM NaCl; 5 µM $ZnCl_2$; 2.5% (v/v) glycerol; 10 mM DTT), supplemented with 20 mM imidazole, 1 u/µl DNase (Fermentas) and 1X PI. After centrifugation at 15,000xg for 30 min, the cleared lysate was applied to a pre-equilibrated (buffer A + 500 mM NaCl) Ni-NTA agarose column (Qiagen). The column was washed with 10 CV of buffer A containing 20 mM imidazole before elution of the protein with buffer A containing 500 mM NaCl and 500 mM imidazole in 5 CV. Elution fraction was 5X diluted to 100 mM NaCl and applied to a MonoQ 10/100 GL column (GE healthcare) equilibrated in buffer A containing 100 mM NaCl. The protein was eluted with a linear gradient from 100 mM to 500 mM NaCl (buffer A) over 15 CV.

Expression and purification of Gal4-VP16 Protein expression was carried out as described in 7.2.1 and expression was induced with 0.5 mM IPTG. Cells were lysed by sonication in buffer A (10 mM Tris, pH 8.0 at 20°C; 500 mM NaCl; 10 µM $ZnSO_4$; 10% (v/v) glycerol; 10 mM ß-mercaptoethanol), supplemented with 10 mM imidazole, 1 u/µl DNase (Fermentas) and 1X PI. After centrifugation at 16,000xg for 20 min, the cleared lysate was applied to a pre-equilibrated (buffer A) Ni-NTA agarose column (Qiagen). The column was washed with 10 CV of buffer B (10 mM Tris, pH 8.0 at 20°C; 100 mM NaCl; 10 µM $ZnSO_4$; 10% (v/v) glycerol; 10 mM ß-mercaptoethanol) and 10 CV of buffer B containing 20 mM imidazole before elution of the protein with buffer B containing 200 mM imidazole. Fractions containing Gal4-VP16 were pooled and applied to a HiTrap Q HP column (GE healthcare) equilibrated in buffer C

(20 mM HEPES, pH 7.5; 10 µM Zn acetate; 10% (v/v) glycerol; 1 mM DTT). The protein was eluted with a linear gradient from 0 mM to 700 mM NaCl in buffer C over 10 CV. To remove any minor contaminants, a final size exclusion step using a Superose 12 10/300 GL column (GE Healthcare) in buffer D (20 mM HEPES, pH 7.5; 150 mM K acetate; 10 µM Zn acetate; 10% (v/v) glycerol; 1 mM DTT) was carried out.

7.2.2 Pol II-Bye1 complex formation

Complex formation for crystallization or biochemical analysis was carried out by mixing purified Pol II (3.5 mg/ml) with a tenfold molar excess of recombinant Bye1 and incubating the sample ON at 4°C before further analysis.

7.2.3 Crystallization

Crystallization of the Pol II-Bye1 complex For co-crystallization of Pol II and Bye1 full-length, a stable complex was formed as described (7.2.2) before crystallization by vapour diffusion with 750 mM tri-Na-citrate and 100 mM HEPES pH 7.5 as reservoir solution. Crystals were grown for 13 days, cryo-protected in 22% glycerol, followed by one hour incubation before harvesting and flash-freezing in liquid nitrogen.

Crystallization of Pol II-EC-Bye1 TLD complex Purified Pol II (3.5 mg/ml) was mixed with a two-fold molar excess of tailed template (7.1.6,6.3), and incubated for 1 h at 20°C before crystallization by vapour diffusion with 5-7% PEG 6000, 200 mM ammonium acetate, 300 mM sodium acetate, 50 mM HEPES pH 7.0 and 5 mM TCEP as reservoir solution. Crystals were grown for 5–10 days, cryo-protected in mother solution supplemented with 22% glycerol and containing 4 µM tailed template, followed by ON incubation at 8 °C before harvesting and freezing in liquid nitrogen. Bye1 TLD was added to the cryo-protectant at 1 mg/ml and crystals were incubated ON at 8°C.

Crystallization of arrested Pol II-Bye1 TLD complex Purified Pol II (3.5 mg/ml) was mixed with a two-fold molar excess of arrest template (see 6.3 and 7.1.6), 8 mM magnesium chloride and 2 mM CTP, and incubated for 2 h at 20 °C before crystallization by vapour diffusion with 5-7% PEG 6000, 200 mM ammonium acetate, 300 mM sodium acetate, 50 mM HEPES pH 7.0 and 5 mM TCEP as reservoir solution. Crystals were grown for 5–10 days, cryo-protected in mother solution supplemented with 22% glycerol and containing 4 µM arrested template, 2 mM CTP and 8 mM magnesium chloride, followed by ON incubation at 8°C before

harvesting and freezing in liquid nitrogen. Bye1 TLD was added to the cryo-protectant at 1 mg/ml and crystals were incubated ON at 8°C.

Crystallization of Pol II-EC-Bye1 TLD complex containing an additional nucleotide Crystals were prepared as described for Pol II-EC-Bye1 TLD complexes, with all cryoprotectant solutions containing 2 mM AMPCPP.

Crystallization of Pol II-EC-SeMet Bye1 TLD complex Crystals were prepared as described for Pol II-EC-Bye1 TLD complexes, using selenomethionine-substituted Bye1 TLD.

7.2.4 Data collection and X-ray structure determination

Diffraction data were collected at 100 K at beamline X06SA of the Swiss Light Source. Data were collected at 0.91887 Å, the K-absorption peak of bromine and 0.9797 Å, the K-absorption peak of selenium. Structures were solved with molecular replacement using BUSTER (Bricogne *et al.*, 2012) and the structure of 12-subunit Pol II (1WCM) as search model. Refinement was performed using iterative cycles of model building in COOT (Emsley *et al.*, 2010) and restrained refinement in BUSTER.

7.2.5 Surface plasmon resonance

Approximately 2500 resonance units of yeast Pol II were immobilized in immobilization buffer (Na-Acetate, pH 5) on the surface of a biosensor CM5 chip (Biacore) using the amine coupling kit (Biacore) (Johnsson *et al.*, 1991; Löfås & Johnsson, 1990). Recombinant Bye1 full-length was injected for 60 s at 10 µl/min in running buffer (5 mM HEPES (pH 7.25 at 20 °C), 40 mM $(NH_4)_2SO_4$, 10 µM $ZnCl_2$, 5 mM DTT, 0.005% P20) at different concentrations (19 nM to 20 µM). The complex was allowed to dissociate for 5 min between injections. Affinity was measured for three independent dilution series. Raw data were corrected for the bulk signal from buffer and by identical injection through a flow cell in which no Pol II was immobilized. Data were analysed with BIAevaluation software (Biacore).

7.2.6 Nuclear extract preparation

3 l yeast culture was grown in YPDS to an OD_{600} of 3-5 and harvested by centrifugation at 4,000 rpm for 10 min at 18°C. Pellets were resuspended in 35 ml resuspension buffer and incubated for 15 min at 30°C shaking in a water bath. After centrifugation at 4,000 rpm for 10 min at 18°C, pellets were resuspended in 20 ml YPDS. For cell wall digestion and spheroplasting, 3 ml of 2 M

Sorbitol and 3 ml resuspension buffer, containing 18 mg Zymolyase (Siekagaku), 1X PI and 1X PMSF, were added and the mixture was incubated at 30°C with occasional shaking in a water bath. Spheroplasting was monitored by measuring OD_{600} and was stopped when 1/10 of the initial OD_{600} was reached. 100 ml YPDS were added and cells were centrifuged at 4,000 rpm for 10 min at 18°C before resuspending in 250 ml YPDS and incubation at 30°C for 30 min. Cells were centrifuged as before, resuspended in 200 ml ice-cold YPDS, centrifuged, resuspended in 250 ml ice-cold 1 M Sorbitol and centrifuged again. For cell membrane disruption, pellets were resuspended in 100 ml ice-cold cell lysis buffer and passed three times through a pre-chilled Dounce glass homogenizer (Kontes). Homogenized cells were then centrifuged three times at 5,000 rpm for 8 min at 4°C and pellets were discarded. Crude nuclei were centrifuged at 13,000 rpm for 30 min at 4°C and resuspended in 15 ml centrifugation buffer, washed once by centrifugation and resuspended in 15 ml centrifugation buffer. For nuclear membrane disruption, 3 M $(NH_4)_2SO_4$ was added to a final concentration of 0.5 M and incubated on a spinning wheel at 4°C for 30 min before ultracentrifugation at 28,000 rpm for 90 min at 4°C. To precipitate nuclear extract, 0.35 g $(NH_4)_2SO_4$ was added per ml supernatant, incubated on a spinning wheel at 4°C for 30 min and ultracentrifuged twice at 10,000 rpm for 20 min at 4°C, discarding the supernatant. To resuspend nuclear extracts, pellets were resuspended in maximum 1.5 ml dialysis buffer and dialysed against 500 ml dialysis buffer containing 75 mM $(NH_4)_2SO_4$ for 4.5 hours. Finally, the protein concentration was determined by Bradford assay (see 7.1.5), and 100 µl aliquots were flash-frozen in liquid nitrogen and stored at -80°C.

7.2.7 *In vitro* transcription assay

Pol II *in vitro* transcription was done on the HIS4 yeast promoter inserted into pBluescript II KS+ plasmid (6.3). The 25 µl transcription reaction contained 1X acetate transcription buffer, 200 µg yeast nuclear extract, 200 ng template plasmid, 192 µg phosphocreatine (Sigma), 0.2 µg creatine phosphokinase (Sigma), 10 U RiboLock RNase inhibitor (Fermentas), 100 µM NTPs and 150 ng recombinant Gal4-VP16 (7.2.1). The reaction was incubated for 60 min at 18°C, followed by RNA isolation using the RNeasy MinElute kit (Qiagen). RNA was eluted with 14 µl RNase-free H_2O and *in vitro* transcripts were subsequently analysed by primer extension. The 20 µl primer annealing reaction contained 1X annealing buffer, 12 µl RNA and 0.125 pmol fluorescently labelled oligo (6.3). After boiling the sample for 3 min at 95°C, primer annealing was carried out for 45 min at 48°C, followed by addition of 40 µl synthesis mix containing 1X synthesis buffer, 0.15 mM dNTPs, 12.5 U MuLV reverse transcriptase (Roche) and 1 µg actinomycin D, and incubation for 30 min at 37°C. The resulting cDNA was EtOH precipitated

and resuspended in 4 µl RNase A (40 µg/ml, Thermo Scientific). After incubation for 3 min at 18°C, 4 µl formamide sample buffer was added and the sample boiled for 1 min. Transcripts were separated on a urea-polyacrylamide gel in 1X TBE, scanned with a Typhoon 9400 and quantified with the ImageQuant software (GE Healthcare).

7.2.8 RNA extension assay

For RNA extension assays, tailed templates as listed in 6.3 were used, with the difference that the RNA was labelled with FAM at the 5' end, and the bromine in the template DNA was replaced by a thymine. Scaffolds were assembled as described (7.1.6). 2.5 pmol Pol II were incubated with 5 pmol scaffold for 30 min at 20°C. Additional proteins or equivalent volume of transcription buffer was added and the sample was incubated in a final volume of 6 µl ON at 4°C. To start the reaction, 5 mM NTPs were added and the reaction was incubated for 20 min at 28°C. The reaction was stopped by adding one volume of 2X urea loading buffer and incubation at 95°C for 5 min, followed by incubating samples on ice. Transcripts were separated on a urea-polyacrylamide gel in 1X TBE, scanned with a Typhoon 9400 and quantified with the ImageQuant software (GE Healthcare).

7.2.9 RNA cleavage assay

For RNA cleavage assays, arrest templates were used (6.3), with the difference that the RNA was labelled with FAM at the 5' end, and the bromine in the template DNA was replaced by a thymine. Scaffolds were assembled as described in 7.1.6. 2.5 pmol Pol II were incubated with equimolar amounts of scaffold for 30 min at 20°C. Additional proteins or equivalent volume of transcription buffer was added and the sample was incubated in a final volume of 6 µl ON at 4°C. To start the reaction, TFIIS was added and the reaction was incubated for 20 min at 28°C. The reaction was stopped by adding one volume of 2X urea loading buffer and incubation at 95°C for 5 min, followed by incubating samples on ice. Cleavage products were separated on a urea-polyacrylamide gel in 1X TBE, scanned with a Typhoon 9400 and quantified with the ImageQuant software (GE Healthcare).

7.2.10 Chromatin fractionation

Strains used in yeast chromatin fractionation were derived from W303. Plasmids containing HA-tagged, full-length Bye1, Bye1 ΔPHD (Δ1-177) and Bye1 ΔTLD (Δ177-354) (obtained from S.D. Hanes, (Wu *et al.*, 2003)) were transformed into wild-type yeast. Chromatin frac-

tionation was performed using a combination of previously described methods (Donovan *et al.*, 1997; Keogh *et al.*, 2006). Cells were grown in YPD from a starting OD_{600} of 0.25 to mid log phase (OD_{600} ~1.0). Forty OD_{600} units of cells were harvested by centrifugation and resuspended in 10 ml of sterile water. Following another round of centrifugation, cells were resupended in 10 ml SB buffer (1 M Sorbitol, 20 mM Tris pH 7.4 at 4°C), collected by centrifugation and stored at -80 ° C ON. The cell pellets were then thawed on ice, resuspended in 1.5 ml PSB buffer (20 mM Tris pH 7.4 at 4°C, 2 mM EDTA, 100 mM NaCl, 10 mM β-mercaptoethanol) and transferred to a 2 ml microcentrifuge tube. Cells were allowed to mix for 10 minutes at room temperature on a rotating shaker. Cells were pelleted by a flash spin in a microcentrifuge and the buffer was aspirated. Cell pellets were washed briefly in 1.5 ml SB buffer and centrifuged as before. The pellet was resuspended in 1 ml SB buffer, and 125 µl of 10 mg/ml Zymolyase 20T (Seikagaku Biobusiness) in SB buffer was added. The mixture was incubated at room temperature for 30-60 minutes on a rotating shaker. Spheroplasting progress was assessed by addition of 10 µl of cells to 1 ml 1% SDS and vortexing, followed by measuring the OD_{600} of the liquid. Once OD_{600} measurement decreased by more than 80% the starting value, spheroplasting was stopped with ice-cold SB buffer. Spheroplasts were pelleted at 2,000 rpm for 5 min at 4 ° C. The buffer was removed and the pellet was gently resuspended in 1 ml LB buffer (0.4 M Sorbitol, 150 mM potassium acetate, 2 mM magnesium acetate, 20 mM PIPES-KOH pH 6.8, 1 µg/ml leupeptin, 1 µg/ml pepstatin, 1 µg/ml aprotinin, 1 mM PMSF) and pelleted as above. The LB buffer wash step was repeated. To lyse the cells, the pellet was gently resupended in 250 µl LB buffer with 1% Triton X-100, transferred to a 1.5 ml microcentrifuge tube and incubated on ice for 10 minutes with occasional gentle mixing. Following lysis, 125 µl was removed for the whole cell extract (WCE) and the remainder was centrifuged at 5,000xg for 15 minutes at 4 ° C. The supernatant was collected as the 'soluble' fraction. The 'chromatin' pellet was washed once by resuspension in 125 µl of LB buffer with 1% Triton X-100 and spun as in the previous step. The supernatant was discarded and the 'chromatin' pellet was resuspended in 125 µl of LB buffer with 1% Triton X-100. All samples were normalized to total protein content of WCE as determined by Bradford assay (7.1.5). Normalized WCE and volume equivalents of 'soluble' and 'chromatin' fractions were boiled in 1X SDS loading buffer, separated by 15% SDS-PAGE and analysed by immunoblotting with antibodies HA (Covance) 1:1000, H4 (Millipore) 1:1000, G6PDH (Sigma) 1:100,000.

7.2.11 Histone peptide microarrays

In general, all steps were carried out as described in Rothbart *et al.*, 2012.

Peptide synthesis and validation Peptides were synthesized using a PS3 peptide synthesizer (Protein Technologies, Inc.) and a Symphony multiple peptide synthesizer (Protein Technologies, Inc.). Peptide purification was carried out by reversed-phase high-performance liquid chromatography (RP-HPLC) and fractions were checked for the presence of the correct product by matrix-assisted laser desorption/ionization (MALDI) coupled to time-of-flight (TOF) mass spectrometry. Lyophilized peptides were stored at -80°C. Details on all peptides used can be found in Supplemental Table 18.

Microarray fabrication For microarray printing, 100 µM stock solutions of each biotinylated histone-peptide was prepared. 3 µl of peptide solution and 7 µl of printing buffer were added to each well of a 384-well plate. Prior to printing, the plate was spun down at 500xg for 2 min. Each peptide was spotted in triplicate eight times per array. Triplicate spots were averaged and treated as a single value for subsequent statistical analysis.

Purification of effector protein Full-length Bye1 (residues 1-594) and Bye1 PHD (residues 47-134) were expressed as GST-fusions from exponentially growing (OD_{600} ~0.6) BL21 RIL cells by ON induction with 0.4 mM IPTG at 16°C. Cells were lysed by sonication in cold 1X PBS pH 7.6 containing 1 mM (PHD) or 5 mM (full-length) DTT, 1 mM PMSF, 1 mM $ZnSO_4$, and 10% glycerol (full-length only). Proteins were captured on GST-Bind Resin (Novagen) and eluted in buffer containing 50 mM Tris-HCl pH 8.0 at 4°C and 10 mM glutathione. Proteins were dialysed into buffer containing 20 mM Tris-HCl pH 8.0 at 4°C, 150 mM NaCl, and 1 mM DTT prior to microarray hybridization.

Effector protein hybridization and detection Prior to hybridization, microarray slides were incubated with hybridization buffer for 1 h to block non-specific binding. 200 µl of protein solution containing 1 µM of the protein of interest in hybridization buffer was pipetted drop-wise on the microarray slide. After ON incubation at 4°, the microarray slide was washed 3X for 5 min with PBS, pH 7.6. Bound protein was detected by a series of antibody incubation steps, varying depending on protein or protein-tag to be detected. Antibody incubation was followed by washing 3X for 5 min with PBS, pH 7.6. Slides were then incubated in the dark for 30 min with secondary antibody solution (Alexaflour-647) and washed again 3X for 5 min with PBS, pH 7.6. Excess salt was washed off in 0.1X PBS, pH 7.6, and the slide was dried by centrifugation. Detection of microarray signals was performed using the Typhoon TRIO+ imager (GE Healthcare) with a 526 nm (biotinylated fluorescein) and 670 nm (secondary antibody) emission filter.

Data analysis Signal intensities for each spot on a scanned microarray were determined by densitometry using the microarray analysis program ImageQuant TL (GE Healthcare) and auto-correcting for background by subtracting the local fluorescence spot edge average.

Peptides were printed as four independent subarrays, which account for subtle differences in intensity across the slide. The eight spots per subarray for each peptide were averaged, outliers were removed using a Grubbs test (Grubbs, 1969; Stefansky, 1972), and the average was then treated as a single value for each subarray. The four normalized subarray values were then used to calculate a standard deviation and standard error.

7.2.12 Chromatin immunoprecipitation analysis

Chromatin immunoprecipitation (ChIP) All strains used for chromatin immunoprecipitation were TAP-tagged and are listed in Table 6.2. Strains were validated by gene-specific PCR and westernblotting with the anti-TAP (PAP, Sigma) antibody to confirm TAP-tag expression.

For chromatin immunoprecipitation, a 20 ml YPD pre-culture was inoculated with a single yeast colony and grown ON at 30°C shaking. A 50 ml (for ChIP-qPCR)/600 ml (for ChIP-chip) YPD-culture was inoculated with the preculture with an initial OD_{600} of 0.1, and grown at 30°C shaking to an OD_{600} of 0.8. Cultures were incubated with 1000 ml/16.2 ml 37% formaldehyde for crosslinking at 20°C shaking for 20 min. 5 ml/75 ml 3 M glycine were added and cultures incubated for 5-10 min shaking at 20°C, before centrifugation for 5 min at 4,000 rpm at 4°C. Pellets were washed once with 20 ml/100 ml ice-cold TBS, centrifuged for 5 min at 4,000 rpm at 4°C, and resuspended in 2 ml/10 ml ice-cold FA-lysis buffer containing 1X PI. After centrifugation for 5 min at 4,000 rpm at 4°C, cells were resuspended in 1 ml/1.5 ml ice-cold FA-lysis buffer containing 1X PI. After centrifugation for 5 min at 3,000 rpm at 4°C, pellets were flash-frozen in liquid nitrogen and stored at -80°C.

For cell lysis, pellets were thawed and resuspended in 1 ml ice-cold FA-lysis buffer containing 1X PI. 1 ml of cell suspension was mixed with 1 ml Zirconia beads (Biospec) and cells were lysed in a mixer mill MM 400 (Retsch) for 30 min at 4°C. Lysis efficiency was monitored by OD_{600} measurement. For chromatin fragmentation and resolubilisation, beads were removed from the cell lysate by centrifugation, 200 µl ice-cold FA lysis buffer containing 1X PI were added and samples were treated in a Bioruptor (diagenode) for 35 min at 4°C. Samples were spun down at 13,000 rpm for 30 min at 4°C and chromatin concentration was determined using a Nanodrop (7.1.5). 30 µl chromatin solution were retained as Input-control.

For immunoprecipitation, IgG Sepharose beads (Amersham) were washed once with ice-cold TBS and three times with ice-cold FA lysis buffer before incubating 670 µl chromatin solution

with 25 µl beads for 60 min on a spinning wheel at 4°C. IgG-beads were directed against the Protein A content of the TAP-tag. Beads were spun down at 2500 rpm for 1 min at 4°C and washed 3X with FA lysis buffer, twice with FA lysis buffer containing 500 mM NaCl, twice with ChIP wash buffer and once with TE in a centrifuge-tube filter (Millipore) at 2500 rpm for 30 s at 4°C with inbetween incubation on a spinning wheel for 3 min at 4°C. To elute proteins, beads were incubated with 120 µl ChIP elution buffer for 10 min at 65°C shaking, followed by centrifugation twice for 1 min at 13,000 rpm at RT. To reverse the crosslinks, 80 µl TE and 20 µl Proteinase K (20 mg/ml,) were added to the IP sample, and 100 µl ChIP elution buffer, 60 µl TE and 20 µl Proteinase K were added to the Input sample. Both samples were incubated for 2 h at 37°C, followed by ON incubation at 65°C in a thermocycler (Biometra T3000 Thermocycler). DNA was purified using the QIAquick PCR purification kit (Qiagen) and eluted in 50 µl DNase-free H_2O.

Quantitative real-time PCR (qPCR) For ChIP experiments, input and immunoprecipitated samples (see 7.2.12) were assayed by qPCR to determine protein occupancies at different genomic regions.

Primers for qPCR reactions were 18-24 nt long and were designed with the OligoPerfect Designer (Invitrogen). qPCR products were 60-70 nt long and PCR efficiency was validated using different concentrations of DNA template and determining PCR efficiency using the BioRad CFX Manager software. Only primer pairs with an efficiency $\geq$90% were used in ChIP-qPCR experiments.

A 24 µl master mix was prepared, containing 2 µl of 10 µM primer pairs and 12.5 µl iTaq SYBR Green Supermix (BioRad). 1 µl of DNA was added and all reactions were carried out in 96-well plates. qPCR was performed on a BioRad CFX96 Real-Time System using a 3 min denaturing step at 95°C, followed by 40 cycles of 30 s at 95°C, 30 s at 61°C and 15 s at 72°C. Threshold cycle (Ct) values were determined using the BioRad CFX Manager software setting regression. The IP efficiency and fold enrichment of any given region over control regions (ORF-free heterochromatic region on chromosome V) was determined as described in Fan *et al.* (2008).

Tiling-array hybridization DNA samples purified as described in 7.2.12 were amplified with GenomexPlex Complete Whole Genome Amplification 2 (WGA2) Kit using the Farnham Lab WHA Protocol for ChIP-chip (O'Geen *et al.*, 2006). Briefly, 10 µl of concentrated ChIP DNA was used to generate the PCR-amplifiable OmniPlex Library, consisting of ChIP DNA molecules flanked by universal priming sites. Library preparation was performed as described in

the WGA2 Kit (Sigma). The OmniPlex Library was then amplified by PCR and purified using the QIAquick PCR purification kit (Qiagen). DNA was eluted with 50 µl H_2O and quantity and quality was determined with a Nanodrop and agarose gel electrophoresis. Purified DNA was re-amplified in the presence of 0.4 mM dUTP, which is required for enzymatic fragmentation of DNA.

DNA fragmentation, labelling, hybridization and array scanning were performed according to manufacturer's instructions (AffymetriX Chromatin Immunoprecipitation Assay Protocol P/N 702238). At least two independent biological replicates were analysed for each factor.

Bioinformatic analysis All data normalization procedures were performed using R (R Development Core Team, 2008) and Bioconductor (Gentleman *et al.*, 2004). For data import of the AffymetriX CEL files and the conversion into the basic Bioconductor object class for microarray data Expression Set, we used the R package *Starr* (Zacher *et al.*, 2010). Data normalization consisted of three steps: First, quantile normalization between replicate measurements was carried out. Second, the signal for each probe was averaged by calculating the geometric average over the replicate intensities. Third, data from all factors were normalized using a combined mock IP (untagged WT strains) plus input reference normalization. Additional aspects of data normalization were described in (Mayer *et al.*, 2010). Normalized signals were converted to occupancy values between 0-100% by setting the genome-wide 99.8% quantile to 100% occupancy and the 10% quantile to 0% occupancy.

For gene-averaged occupancy profiles, *Saccharomyces cerevisiae* genes with available TSS and pA assignments from RNA-seq experiments (Nagalakshmi *et al.*, 2008) were grouped into four ORF length classes: extremely short (XS) ranging from 256-511 bp (93 genes), short (S) for 512-937 bp(266 genes), medium (M) for 938-1537 bp (339 genes), and long (L) for 1538 to 2895 bp (299 genes). Only genes containing a TSS-ATG distance $\leq$ 200 bp, a minimal ORF and transcript distance to flanking genes of 250 bp and 200 bp, respectively, and genes being grouped into the 50% highest expressed nuclear protein-coding genes (Dengl *et al.*, 2009) were included in the analysis. Profiles within these groups were scaled to median gene length, and gene-averaged profiles were calculated by taking the median over gene factor profiles.

7.2.13 Synthetic lethality screen

Strain BY5563 *bye1* Δ (see 6.2) was generated as described in 7.1.1 and 7.1.2 by replacing the *bye1* ORF with the ClonNAT ORF, which confers resistance to the antibiotic nourseothricin. In addition, the strain contains a MFA1pr-HIS3 reporter, which allows for selective germination for

MATa meiotic progeny after sporulation. The strain was crossed to an ordered array of viable yeast deletion mutants (Giaever *et al.*, 2002), each carrying a gene deletion mutation linked to a kanamycin-resistance marker (kanMX). Resulting heterozygous diploids were selected on GNA medium, containing 100 µg/ml nourseothricin and 200 µg/ml kanamycin. The heterozygous diploids were transferred to sporulation medium to induce the formation of haploid meiotic spore progeny. Spores were then transferred to media lacking histidine (SD-H) to select for MATa meiotic progeny before selecting for double-mutant meiotic progeny in GNA medium containing both 100 µg/ml nourseothricin and 200 µg/ml kanamycin. Growth of double-mutants was then monitored by visual inspection for 48 h. The screen was performed on a Beckman-Coulter Biomek FX.

Strains used to validate candidates from the synthetic lethality screens (6.2) were derived from BY4741 and generated as described in 7.1.1 and 7.1.2.

7.2.14 Non-homologous end-joining (NHEJ) assay

To assay NHEJ efficiency, the pRS314 vector was linearized by restriction digest with *EcoRI* as described in 7.1.2 and transformed into yeast cells as described in 7.1.1. As the *EcoRI* restriction site is located within the ura3 gene, circularization of the linearised vector by NHEJ is required for cells to allow growth on plates lacking uracil. Transformed WT and mutant cells were plated on -ura plates and incubated for 3 days at 30°C. Colonies were counted to monitor NHEJ efficiency. To normalize for transformation efficiency, the circularized pRS314 vector was transformed, and cells treated and colonies counted as described. Three independent experiments were carried out for each yeast strain.

7.2.15 Fluorescence-activated cell sorting (FACS)

Cells were grown in 50 ml YPD at 30°C at 160 rpm from a starting OD_{600} of 0.1 until an OD_{600} of 0.4 was reached. Before synchronization by addition of α-factor (10 µg/ml final concentration), 1 ml was taken as unsynchronized control and stored at 4°C after addition of 2.5 ml 100% EtOH. Cells were synchronized at 30°C at 160 rpm and checked inbetween for shmoo formation by light microscopy. When shmoo formation was completed (about 2 h), cells were spun down for 2 min at 1600 x g at 30°C, washed with 200 ml pre-warmed YPD, spun down again, resuspended in 50 ml pre-warmed YPD and incubated at 30°C at 160 rpm. 1 ml samples for FACS analysis were taken every 10 min for 120 min and stored at 4°C after addition of 2.5 ml 100% EtOH.

Samples were spun down for 3 min, washed once with 1 ml 50 mM Na citrate (pH 7.0),

resuspended in 100 µl 50 mM Na citrate (pH 7.0) containing RNase A (0.1 mg/ml, Fermentas), and incubated ON at 37°C with slight shaking. 1 mg/ml Proteinase K (Thermo Scientific) was added and samples were incubated for 2 h at 50°C with slight shaking. Samples were washed once with 50 mM Na citrate (pH 7.0) and samples were adjusted to approximately 106 cells/ml. Cells were resuspended in 1 ml 50 mM Na citrate (pH 7.0) containing 1 µM Sytox Green (Invitrogen). After separation of doublets in a bioruptor (6x 10 s, with 10 s pause inbetween), 100 µl cells were added to 900 µl Sytox Green mix and transferred to FACS tubes for analysis on FACS Calibur flow cytometer (BD).

Part III

Results and Discussion

8 Structures of RNA polymerase II complexes with Bye1, a chromatin-binding PHF3/DIDO homologue

Data presented in this chapter have been obtained during this thesis and have been published (see page viii).

8.1 Bye1 interacts with Pol II

To test whether Pol II binds directly to Bye1 *in vitro*, we incubated pure yeast Pol II with recombinant Bye1 (see 7.2.2) and subjected the sample to size-exclusion chromatography. A stable and apparently stoichiometric Pol II-Bye1 complex was obtained (Fig. 8A). To characterize the Pol II-Bye1 interaction, we used surface plasmon resonance. We immobilized Pol II on a Biacore sensor chip (Löfås & Johnsson, 1990; Johnsson *et al.*, 1991) and determined Bye1 association and dissociation rates. The ratio of these rates provided a dissociation constant of K_D=3.8 ± 2.2 µM (Fig. 8B,C).

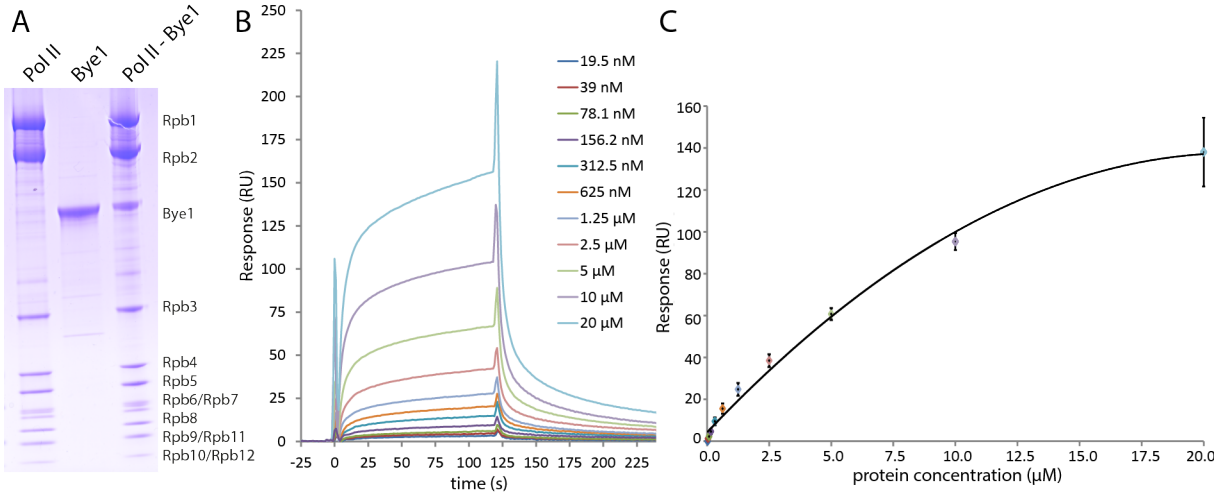

Figure 8: Analysis of the Pol II-Bye1 interaction.
(A) SDS-PAGE analysis (Coomassie staining) of endogenous yeast Pol II (left), recombinant Bye1 (centre), and the Pol II-Bye1 complex after size exclusion chromatography (right) (B) Time-resolved binding of Bye1 dilution series. For reasons of clarity, only one representative curve is shown for each measurement. All measurements have been carried out in triplicates. (C) Corresponding fitted curve (solid line). The curve is reference and blank subtracted.

8.2 Structure of Bye1-bound Pol II elongation complex

Co-crystallisation of Pol II with full-length Bye1 yielded crystals diffracting to 4.8 Å resolution (Table 16). Structure solution by molecular replacement with free Pol II (Armache *et al.*, 2005) revealed positive difference density for the Bye1 TLD on the Rpb1 surface, but no density for the two other Bye1 domains. To obtain better diffraction, the Bye1 TLD was expressed in isolation and soaked into preformed Pol II elongation complex crystals containing a DNA-RNA scaffold. Diffraction data to 3.15 Å resolution were obtained (Table 16). Phasing with the Pol II structure (Armache *et al.*, 2005) revealed positive difference density at the same location observed with full-length Bye1 (Fig. 9A, B).

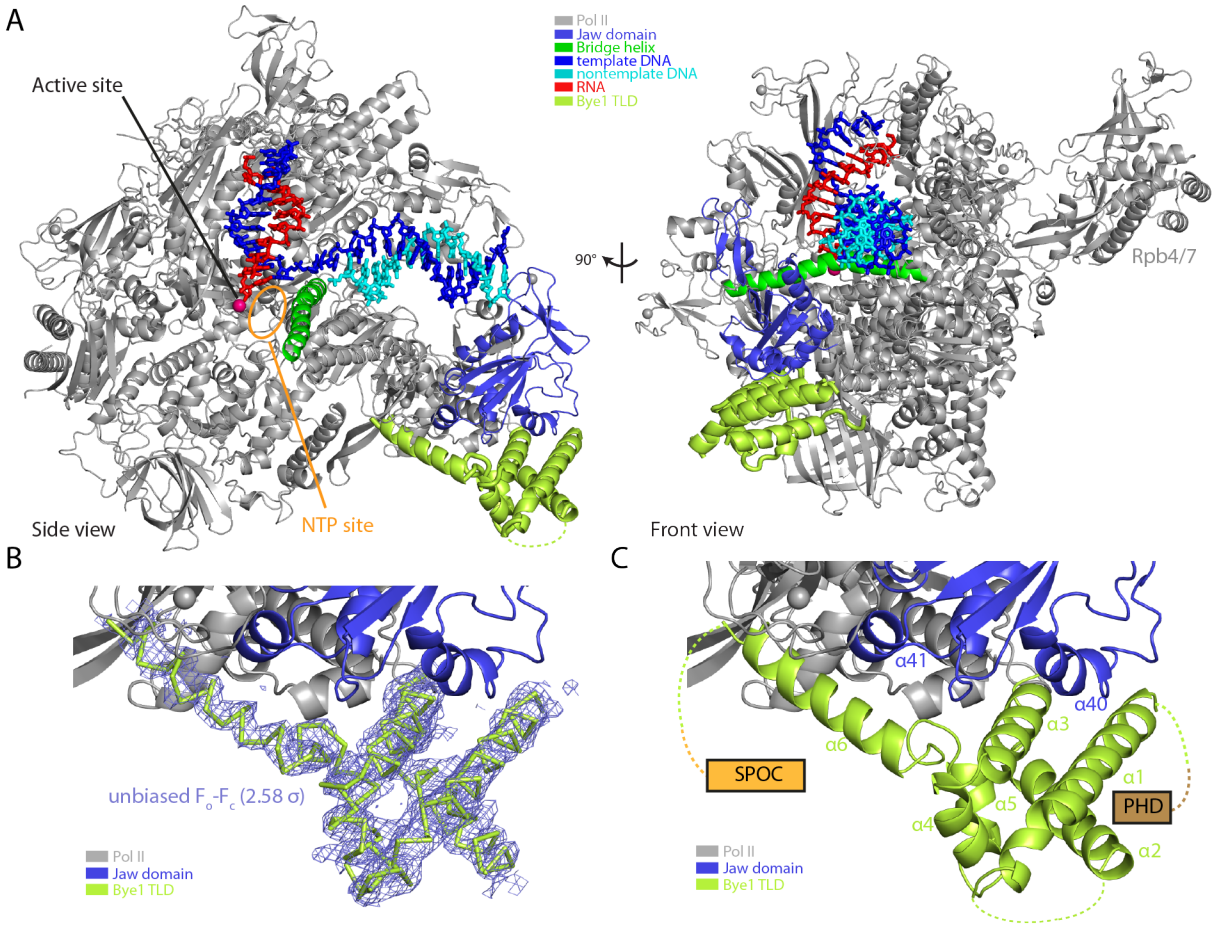

Figure 9: Structure of the Pol II-Bye1 elongation complex.
(A) Ribbon model of the Pol II-Bye1 elongation complex crystal structure. The views correspond to the side and front views of Pol II (Cramer *et al.*, 2001) and are related by a 90° rotation around a vertical axis. (B) Sigma-A weighted difference electron density (blue mesh, contoured at 2.58σ) for the Bye1 TLD (green). (C) Close-up view of the Pol II-Bye1 interaction. Mobile loops are indicated by dashed lines.

The Bye1 TLD structure was built with the aid of sequence markers obtained with seleno-

methionine-labelled protein, and the complex structure was refined to a free R-factor of 21.19% (Table 16).

8.3 Bye1 binds the polymerase jaw

The Bye1 TLD fold comprises an N-terminal three-helix bundle (helices α1-α3) followed by two short helices (α4, α5) that link to an extended C-terminal helix α6 (Fig. 9C). This fold resembles that of TFIIS domain II (helices α1-α5) (Kettenberger *et al.*, 2003), and helix α6 corresponds to the linker between TFIIS domains II and III (Fig. 10A). The Bye1 TLD binds the Rpb1 jaw domain at the location where TFIIS domain II binds the polymerase (Fig. 10B). Despite this overall similarity, the Pol II contacts by the Bye1 TLD and TFIIS differ. The Bye1 helix α3 binds loop ß30-ß31 and helix α40 of the Rpb1 jaw domain and induces ordering of loop α40-ß29. Helix α6 extends from the jaw into the Pol II funnel, contacting the Rpb1 loops α20-α21 and ß29-α41, and strand ß32 of the Rpb1 funnel domain. The Bye1 loop α2-α3 contacts the N-terminus of Rpb5 (Fig. 10C-E).

Table 16: Diffraction data and refinement statistics

	Pol II-Bye1 FL	**Pol II-Bye1 TLD**	**Pol II-Bye1 TLD + AMPCPP**	**Arrested Pol II-Bye1 TLD**
pdb code	4bxz	4by7	4by1	4bxx
Data collection				
Space group	$C222_1$	$C222_1$	$C222_1$	$C222_1$
Unit cell (Å)	220.55	222.50	222.24	222.92
	392.09	390.68	391.58	392.67
	279.80	281.97	281.02	281.04
Unit cell angle (°)	α=β=γ=90°	α=β=γ=90°	α=β=γ=90°	α=β=γ=90°
Resolution range (Å)	49.63-4.80	48.84-3.15	48.95-3.60	49.08-3.28
	(4.92-4.80)	(3.23-3.15)	(3.69-3.60)	(3.37-3.28)
Unique reflections	59394 (4352)	210346 (15471)	141065 (10391)	187168 (13766)
Completeness (%)	99.97 (100)	99.98 (100)	99.98 (100)	99.98 (99.98)
Redundancy	7.50 (7.82)	7.66 (7.74)	7.62 (7.61)	7.66 (7.49)
Rsym (%)	40.9 (173.0)	11.6 (165.2)	21.2 (193.4)	12.9 (185.4)
I/σ(I)	6.05 (1.24)	15.97 (1.60)	9.95 (1.57)	14.66 (1.52)
CC(1/2)	98.5 (60.8)	99.8 (63.2)	99.6 (56.3)	99.8 (67.1)
Refinement				
Non-H atoms	31510	33261	33026	32753
B-factor (mean, Å²)	199.00	115.07	125.08	120.50
Rmsd bonds	0.010	0.010	0.009	0.010
Rmsd angles	1.33	1.22	1.21	1.29
Rcryst (%)	19.06	18.94	17.49	17.98
Rfree (%)	25.27	21.19	20.62	20.77

Values in parenthesis are for the highest resolution shell. All data were collected with a radiation wavelength of 0.9188 Å.

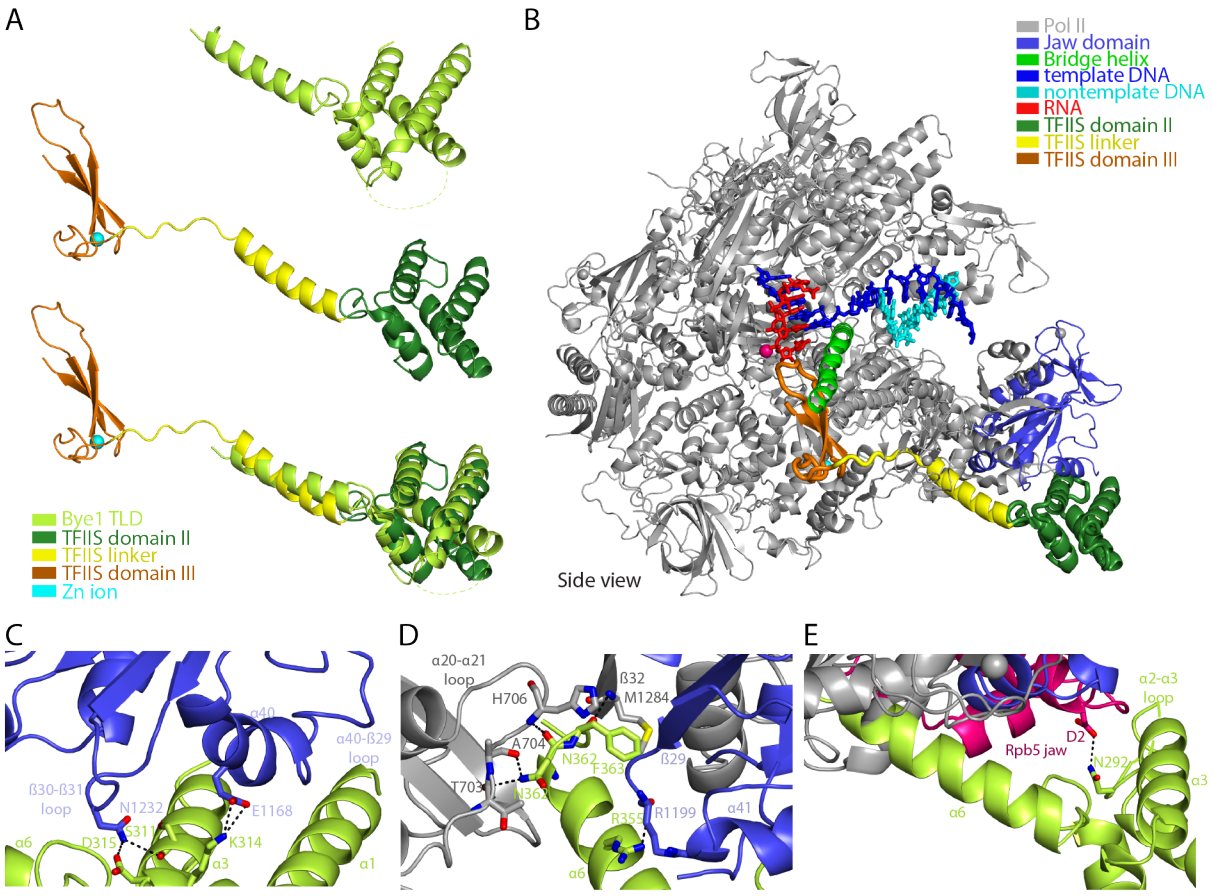

Figure 10: Pol II-Bye1 interaction and comparison with TFIIS.
(A) Side view of the Bye1 TLD bound to Pol II (top), TFIIS in its Pol II-bound state (middle, Cheung & Cramer (2011)) with its central domain II (green), linker helix (yellow) and the C-terminal zinc ribbon domain III (orange), and superposition of the two structures (bottom). (B) Ribbon model of the Pol II-TFIIS complex (Cheung & Cramer, 2011). (C) Details of the interaction of the Bye1 TLD with the Pol II Rpb1 jaw domain (blue) shown in side view. Residues involved in hydrogen bond formation or salt bridges (dashed lines) are shown as sticks. (D) Details of the interaction of the Bye1 TLD with the Pol II Rpb1 funnel and jaw domain shown in side view. (E) Details of the interaction of the Bye1 TLD with the Pol II Rpb5 jaw domain (magenta) shown in side view.

8.4 Bye1 does not change Pol II conformation

TFIIS binding to Pol II induces three major conformational changes in the polymerase elongation complex. It repositions the jaw-lobe module, traps the trigger loop in a locked conformation (Kettenberger *et al.*, 2003), and realigns the RNA in the active site (Kettenberger *et al.*, 2004). Although Bye1 resembles part of TFIIS and binds to a similar position on Pol II, it does not induce conformational changes (Fig. 11A). This lack of conformational changes was observed in structures of Pol II complexes with the Bye1 TLD, but also with full-length Bye1.

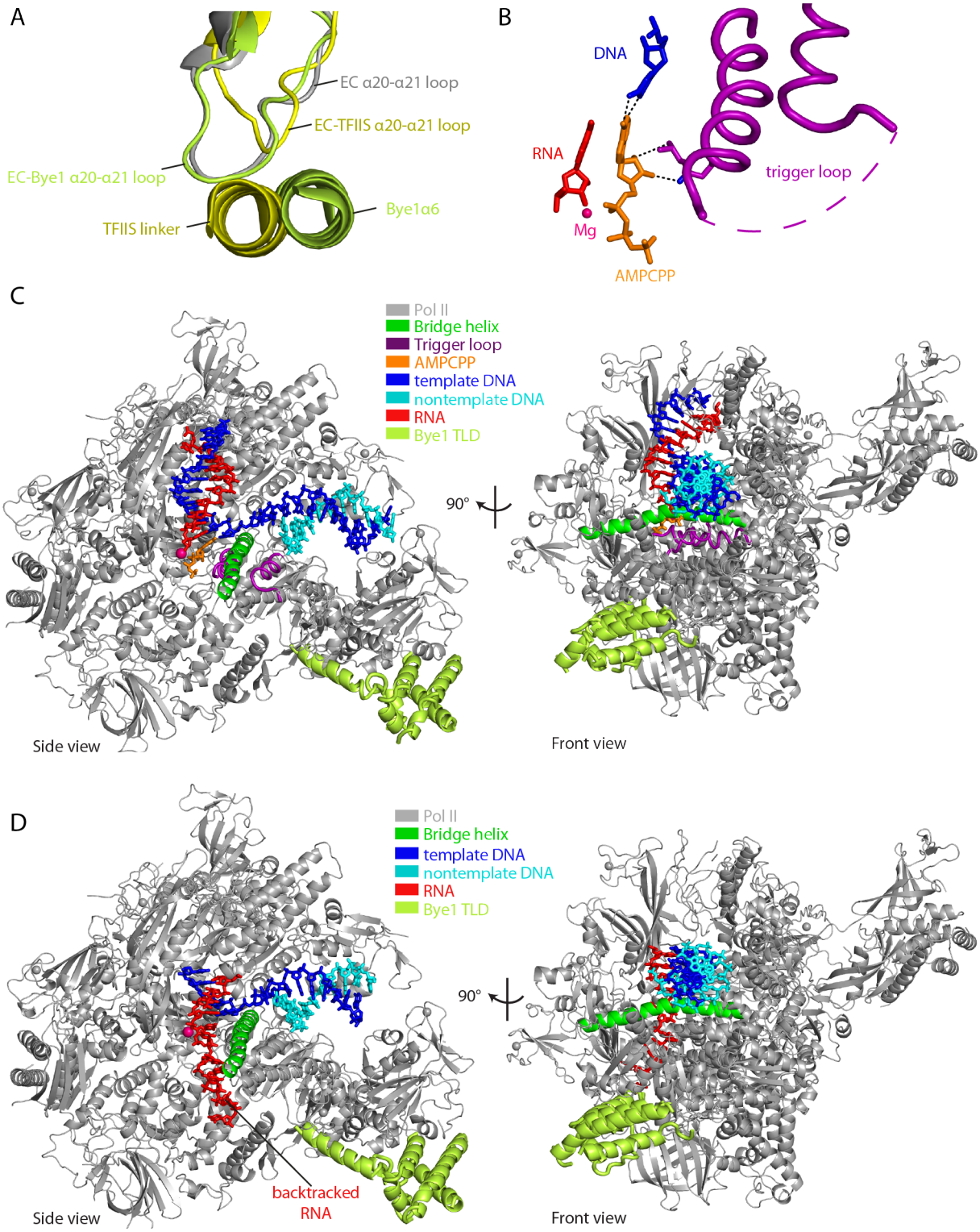

Figure 11: Bye1 does not influence Pol II conformation.
(A) In the Pol II-TFIIS complex structure, conformational changes in Pol II are induced by movements of the Pol II Rpb1 loop α20-α21, which results in opening of a crevice in the polymerase funnel. Loop movements are observed for TFIIS-bound Pol II (yellow, Cheung & Cramer, 2011), but not for Bye1-bound (green, this study) or unbound (silver, Kettenberger *et al.*, 2004) Pol II. (B) Contacts of AMPCPP with the closed trigger loop in the AMPCPP-containing Pol II-Bye1 elongation complex structure. Residues involved in hydrogend bond formation are shown as sticks, hydrogen bonds are indicated by dashed lines. A few residues of the trigger loop that face away from the AMPCPP are mobile (dashed line). (C) Ribbon model of the Pol II-Bye1 complex containing an additional nucleotide. (D) Ribbon model of the arrested Pol II-Bye1 complex.

These observations predicted that Bye1 does not impair nucleoside triphosphate (NTP) binding to Pol II, which requires closure of the trigger loop. Indeed, we were able to show closure of the trigger loop and binding of an NTP substrate analogon in the presence of Bye1 by crystallizing an additional complex of Bye1 TLD bound to the Pol II elongation complex with AMPCPP (Fig. 11B,C and Table 16). Furthermore, Bye1 TLD binding also did not prevent backtracking of RNA into the Pol II pore, as seen in another structure of Bye1 bound to arrested Pol II with backtracked RNA (Fig. 11D and Table 16).

8.5 Bye1 does not influence basic Pol II functions

These observations suggested that Bye1 had no functional influence on basal transcription. Indeed nuclear extracts prepared from yeast cells lacking the gene encoding Bye1 were active in promoter-dependent *in vitro* transcription assays, and addition of purified Bye1 to WT nuclear extracts did not alter their activity (Fig. 12A). In contrast to TFIIS, Bye1 did not induce Pol II backtracking and RNA cleavage on DNA-RNA scaffolds, but allowed for unperturbed elongation activity in RNA extension assays (Fig. 12B,C). Collectively these data indicate that Bye1 neither induces structural changes in Pol II functional complexes nor influences their function *in vitro*.

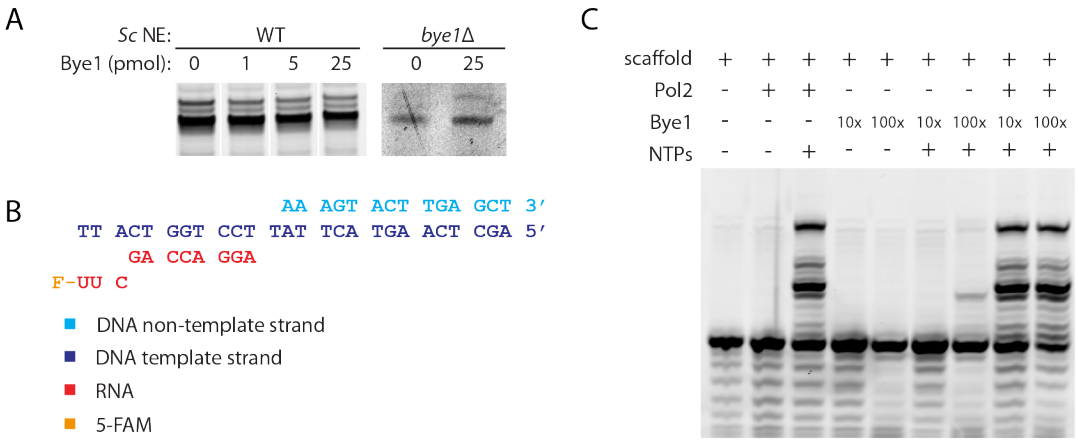

Figure 12: Bye1 does not influence basic Pol II functions.
(A) Transcriptional activity of wildtype (WT) and Bye1-depleted (*bye1*△) nuclear extracts (NE) in an *in vitro* transcription assay using nucleosome-free DNA template. (B) Nucleic acid scaffold for reconstitution of Pol II-EC. (C) Electrophoretic separation of RNA products obtained in RNA extension assays.

8.6 Bye1 associates with chromatin via its TLD domain

The above results suggested that Bye1 functions in a chromatin context. To investigate whether Bye1 associates with chromatin *in vivo* and whether its Pol II-binding TLD is required

for this, we biochemically separated cell extracts into an insoluble fraction, containing chromatin and associated proteins, and a soluble fraction, containing free proteins in the cytoplasm and nucleoplasm. We used strains harbouring plasmids containing HA-tagged full-length Bye1 (WT) or variants lacking either the PHD domain (ΔPHD) or the TLD domain (ΔTLD) (Wu *et al.*, 2003). All variants of Bye1 were present at the same level in unfractionated whole cell extract (Fig. 13, lanes 1-4). Wild-type Bye1 and its ΔPHD variant were present in the chromatin and soluble fractions, but the ΔTLD variant associated with chromatin only very weakly (Fig. 13, lanes 9-12). These results demonstrate that a fraction of the Bye1 protein present in cells associates with chromatin, and that the TLD of Bye1 is important for the association.

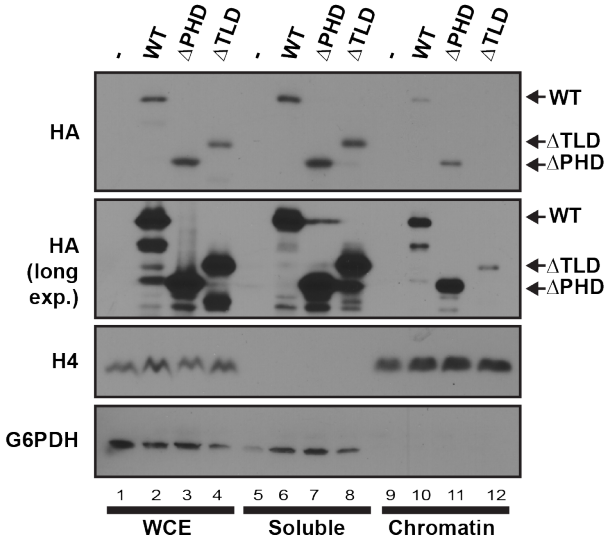

Figure 13: Bye1 associates with chromatin via its TLD.
Immunoblot of whole-cell extract (WCE), soluble and chromatin-bound cell fraction to hemagglutinin tag (HA), histone 4 (H4) and glucose-6-phosphate-1-dehydrogenase (G6PDH). For details compare text.

8.7 Bye1 binds active histone marks via its PHD domain

In addition to the observations described above, it has been reported that the Bye1 PHD domain contributes to chromatin association by binding trimethylated H3K4 peptides (Shi *et al.*, 2007). We therefore investigated binding of the Bye1 PHD domain to roughly 200 differentially modified histone peptides on a microarray (Rothbart *et al.*, 2012). Microarrays were spotted at high density (~4,000 individual features) with histone peptides that encompass known single and combinatorial PTMs on the core and tail domains of the four histone proteins H3, H4, H2A, and H2B, and their variants (see Supplemental Table 18). The Bye1 PHD domain as well as full-length Bye1 bound specifically to H3K4me3 peptides (Fig. 14A). The

high correlation between arrays probed with full length Bye1 and the isolated PHD domain indicated that the histone binding potential of Bye1 is harboured within its PHD domain (Fig. 14B).

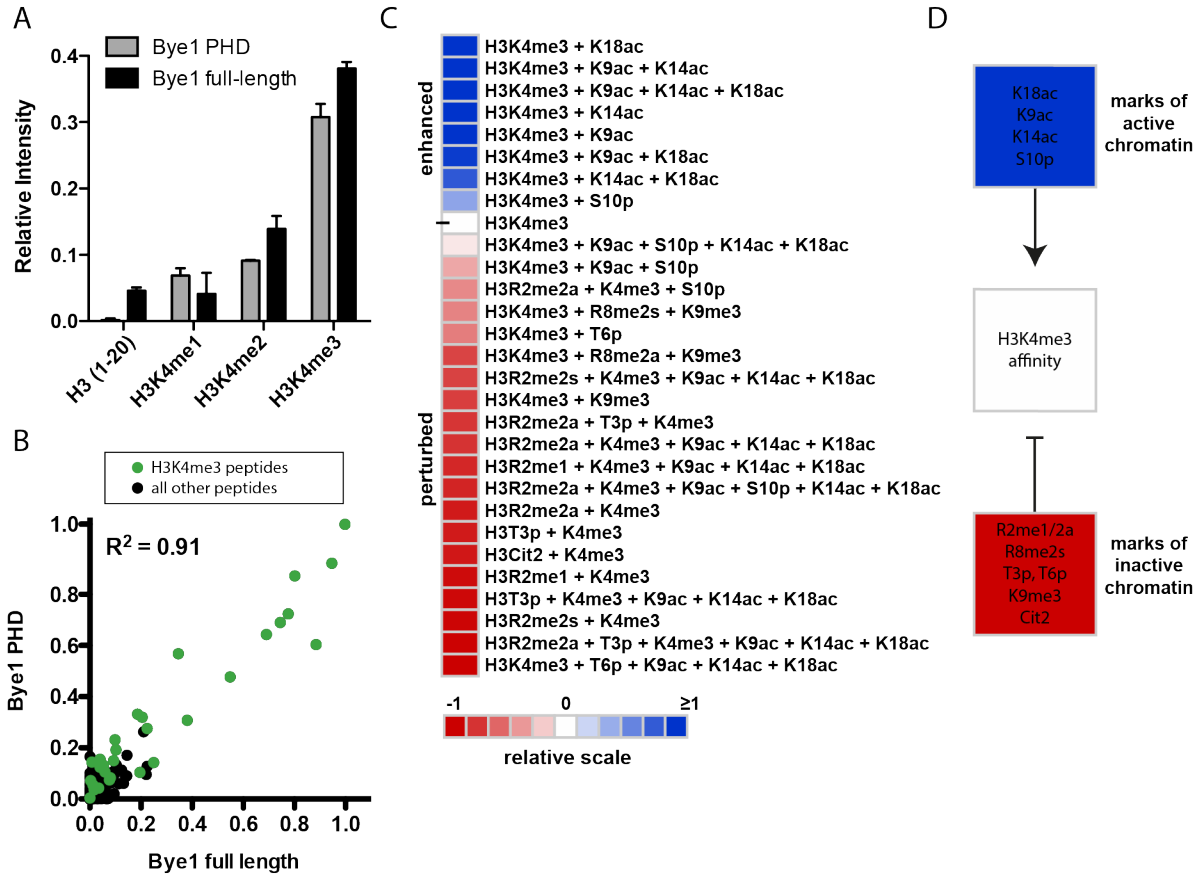

Figure 14: Bye1 preferentially binds histone peptides carrying PTMs of active transcription.
(A) Peptide array binding analysis reveals that Bye1 preferentially associates with H3K4me3 peptides, and its PHD domain is sufficient for this interaction. Results of two independent arrays consisting of 24 individual spots for each peptide are presented as relative mean intensity measurements on a scale from 0 to 1, with 1 being the most significant peptide interaction. (B) Scatter plot correlating relative mean intensity measurements of all peptide interactions from arrays probed with full-length Bye1 and the Bye1 PHD domain. H3K4me3-containing peptides are shown as green dots. All other peptides on the array are shown as black dots. The correlation coefficient was calculated by linear regression analysis using GraphPad Prims v5. (C) Heat map depicting the effects of combinatorial PTMs on the binding of Bye1 to H3K4me3-containing peptides. Average binding intensities are represented relative to the H3K4me3 peptide (0, white). Enhanced (>0, blue) and perturbed (<0, red) interactions are depicted. (D) Summary of modifications enhancing (blue) and perturbing (red) the Bye1 interaction with H3K4me3 peptides.

The interaction of Bye1 with H3K4me3 was strongly influenced by neighbouring PTMs (Fig. 14C). In particular, marks of active transcription (H3K9ac, H3K14ac, H3K18ac, and H3S10p (Oliver & Denu, 2011; Nakanishi *et al.*, 2008; Kurdistani & Grunstein, 2003)) enhanced Bye1 affinity to H3K4me3, whereas marks of transcriptional repression (H3R2 and H3R8 methylation, Cit2, T3 and T6 phosphorylation, and H3K9me3 (Oliver & Denu, 2011; Pal *et al.*, 2004, 2007; Chen *et al.*, 1999)) perturbed the interaction with H3K4me3 (Fig. 14D).

8.8 Bye1 occupies the 5'-region of active genes

To test whether Bye1 is recruited to actively transcribed genes *in vivo*, we carried out genomic occupancy profiling with the use of chromatin immunoprecipitation (ChIP) as described (7.2.12).

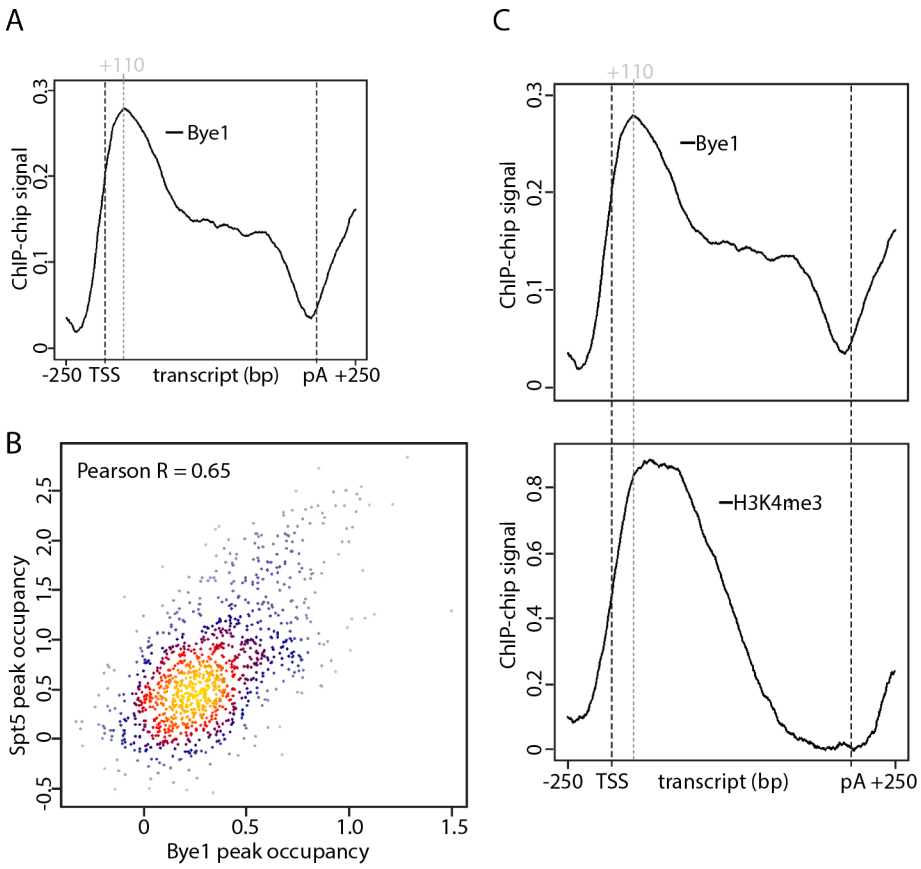

Figure 15: Bye1 occupies the 5' region of active genes.
(A) Gene-averaged Bye1 ChIP occupancy profile for the median length class (1238±300 nt, 339 genes). TSS, transcription start site; pA, polyadenylation site. (B) Correlation of Bye1 and Spt5 occupancies. (C) Comparison of Bye1 and H3K4me3 (Schulze *et al.*, 2011) occupancy profiles.

Metagene analysis by averaging occupancy profiles of genes of similar length revealed a Bye1 occupancy peak 110 nucleotides downstream of the transcription start site (TSS) (Fig. 15A). No significant signals were observed in promoter regions and at the polyadenylation (pA) site. Bye1 was found on all active genes and its occupancy level correlated with those for bona fide Pol II elongation factors such as Spt5 (Fig. 15B). Published ChIP data for H3K4me3 shows a peak at a similar location downstream of the TSS, although the peak is broader (Fig. 15C, Schulze *et al.*, 2011). These results indicated that Bye1 is recruited to the 5'-region of active genes *in vivo*, and suggest that H3K4me3 contributes to Bye1 recruitment. In order to interpret

the ChIP data, we generated a three-dimensional topological model of the Bye1-bound Pol II elongation complex approaching the +2 nucleosome of an active yeast gene (Fig. 16B).

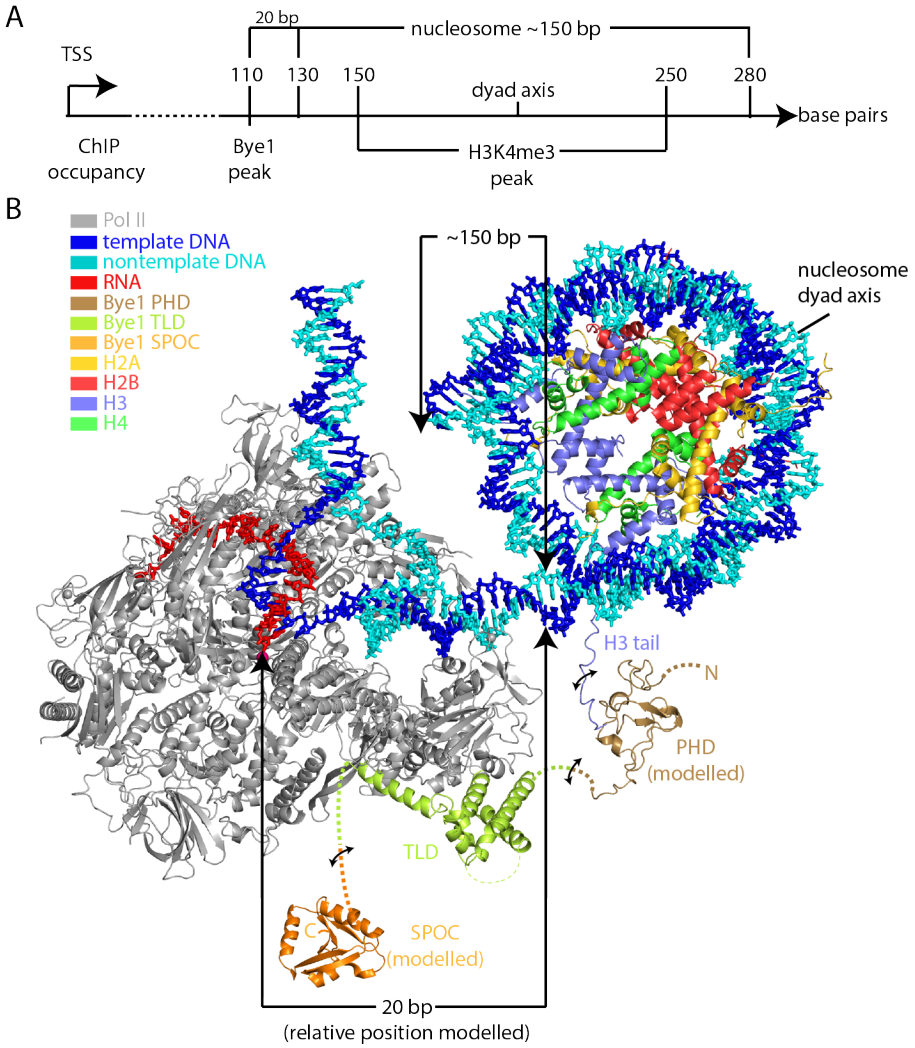

Figure 16: Bye1 can interact simultaneously with Pol II and H3K4me3.
(A) Scheme showing occupancies of Bye1 and H3K4me3 (Schulze *et al.*, 2011) derived from ChIP data and nucleosome position derived from microarray and high-throughput DNA sequencing data (Jiang & Pugh, 2009). (B) Model of a Pol II-nucleosome-Bye1 complex based on crystal structures and ChIP occupancy peak positions. Distances in base pairs (bp) are indicated between the Pol II active centre and the nucleosome as well as for the nucleosomal DNA. The model is based on the structure of the nucleosome core particle (1aoi) by Luger *et al.* (1997). Modelling was performed with Coot (Emsley *et al.*, 2010). Bye1 PHD and SPOC domains were modelled using Modeller (Sánchez & Sali, 1997). The PHD domain model is based on structures 3kqi, 1wem, 1wew, 2lv9 and 1wep, which were identified by HHpred (Söding *et al.*, 2005) to be most similar to Bye1 PHD. Binding of the PHD domain to H3K4me3 was modelled based on structure 2jmj. The SPOC domain model is based on structure 1ow1.

For the modelling we assumed that Bye1 crosslinks to DNA via Pol II in ChIP experiments, and set the Pol II active centre to nucleotide position +110 downstream of the TSS. We positioned the +2 nucleosome based on its experimentally defined average position (Jiang & Pugh, 2009) (Fig. 16A). We also included models of the flexible Bye1 SPOC and PHD

domains, with the latter positioned on the H3 tail emerging from the core nucleosome particle (Fig. 16B). Although the trajectory of the H3 tail is unclear, and although the linkers between the Bye1 domains are flexible, the resulting model explained the position of the ChIP peak with high H3K4me3 occupancy. The model also suggests that it is structurally possible for Bye1 to interact simultaneously with the Pol II core and the trimethylated H3 tail in the 5' region of active genes.

8.9 Bye1 genetically interacts with Paf1 and Tho2

In order to identify genes that interact functionally with the gene encoding Bye1 and thereby further elucidate Bye1 function, we screened a yeast deletion strain collection (Tong *et al.*, 2001, 2004) for synthetic growth defects with *bye1*Δ, which does not show any obvious phenotype (Wu *et al.*, 2003). This screen revealed two candidate genes, *paf1* and *tho2*. Generating *bye1*Δ*paf1*Δ and *bye1*Δ*tho2*Δ double mutants in a different genetic background confirmed the synthetic interaction between these genes (Fig. 17).

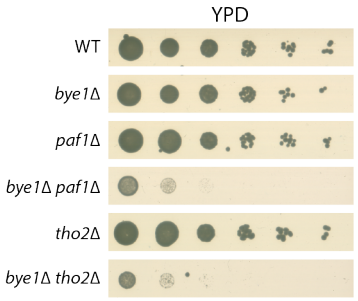

Figure 17: Bye1 genetically interacts with Paf1 and Tho2.
Serial dilutions of strains *bye1*$\triangle$, *paf1*$\triangle$, *bye1*$\triangle$*paf1*$\triangle$, *tho2*$\triangle$, *bye1*$\triangle$*tho2*$\triangle$, and an isogenic wild-type (WT) control strain were placed on YPD plates and incubated at 30°C for 3 days.

The genes *paf1* and *tho2* encode for subunits of two bona fide elongation factor complexes. Paf1 belongs to the 5-subunit Paf complex that recruits the histone methyltransferase Set1 to transcribed genes (Jaehning, 2010; Krogan *et al.*, 2003; Wood *et al.*, 2003; Shilatifard, 2008; Dehé & Géli, 2006). Set1 in turn is responsible for H3K4 trimethylation during transcription (Roguev *et al.*, 2001). The PHD interaction of Bye1 with H3K4me3 is therefore a plausible link between Paf1 and Bye1. Tho2 resides in the 4-subunit THO complex that is required for efficient transcription elongation (Rondón *et al.*, 2003). These results therefore strongly support an involvement of Bye1 in transcription elongation through chromatin.

8.10 PHF3 and DIDO are human homologues of Bye1

No homologues in higher eukaryotes have been reported for Bye1. We performed a bioinformatics search based on the Pfam database (Punta *et al.*, 2012) to identify potential homologues with the same domain organization. We found two human proteins, PHD finger protein 3 (PHF3) and Death-inducer obliterator (DIDO), which show the same domain organization as Bye1 (Fig. 18). Both proteins contain an N-terminal PHD domain, a central TLD domain, and a C-terminal SPOC domain, with linkers of varying lengths in between these domains. Homology for both proteins could not be inferred based on sequence homology (E-value: 2e-04 (PHF3)/ 5e-04 (DIDO)). PHF3 has been associated with glioma development as its expression is significantly reduced or lost in glioblastomas (Fischer *et al.*, 2001). DIDO is a potential tumor suppressor showing abnormal expression patterns in patients with myelodysplastic and myeloproliferative diseases (Fütterer *et al.*, 2005). Specific binding of the DIDO PHD domain to H3K4me3 has been reported recently (Gatchalian *et al.*, 2013).

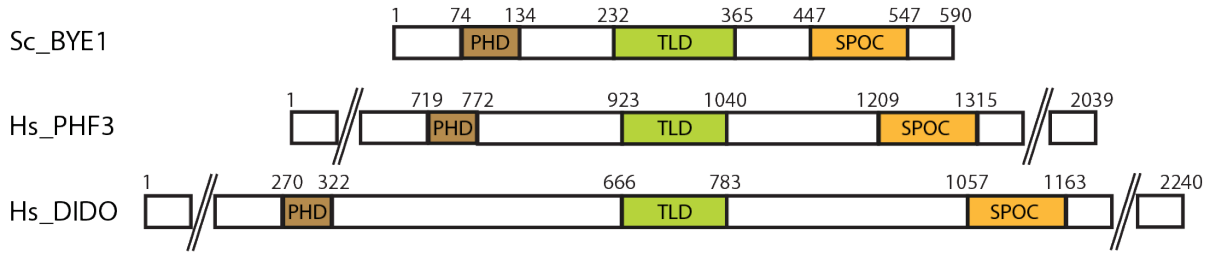

Figure 18: Domain organization of *Saccharomyces cerevisiae* (Sc) Bye1, *Homo sapiens* (Hs) PHF3 and *Homo sapiens* (Hs) DIDO. PHD: Plant Homeodomain, TLD: TFIIS-like domain, SPOC: Spen paralogue and orthologue C-terminal domain. Bordering residue numbers are indicated.

To corroborate the homology of PHF3 and DIDO with Bye1, we analysed the conservation of the Pol II-TLD interface. Both yeast Pol II and Bye1 TLD surfaces forming the interface are well conserved in human Pol II and PHF3/DIDO, respectively (Fig. 19A, B). In particular, a salt bridge between yeast Bye1 residue K314 and E1168 in the largest Pol II subunit Rpb1 is conserved in the predicted human PHF3/DIDO-Pol II complexes. Similarly, many hydrogen bonds observed between the Bye1 TLD and Rpb1 (Bye1 residues N292, S311, D315, R355, N362, F363) are predicted to be conserved in the homologous human complexes. These results indicate that PHF3 and DIDO contain Pol II-binding TLD domains and are human homologues of Bye1, and our structural, biochemical and genetic results provide a starting point for elucidating the function of these proteins.

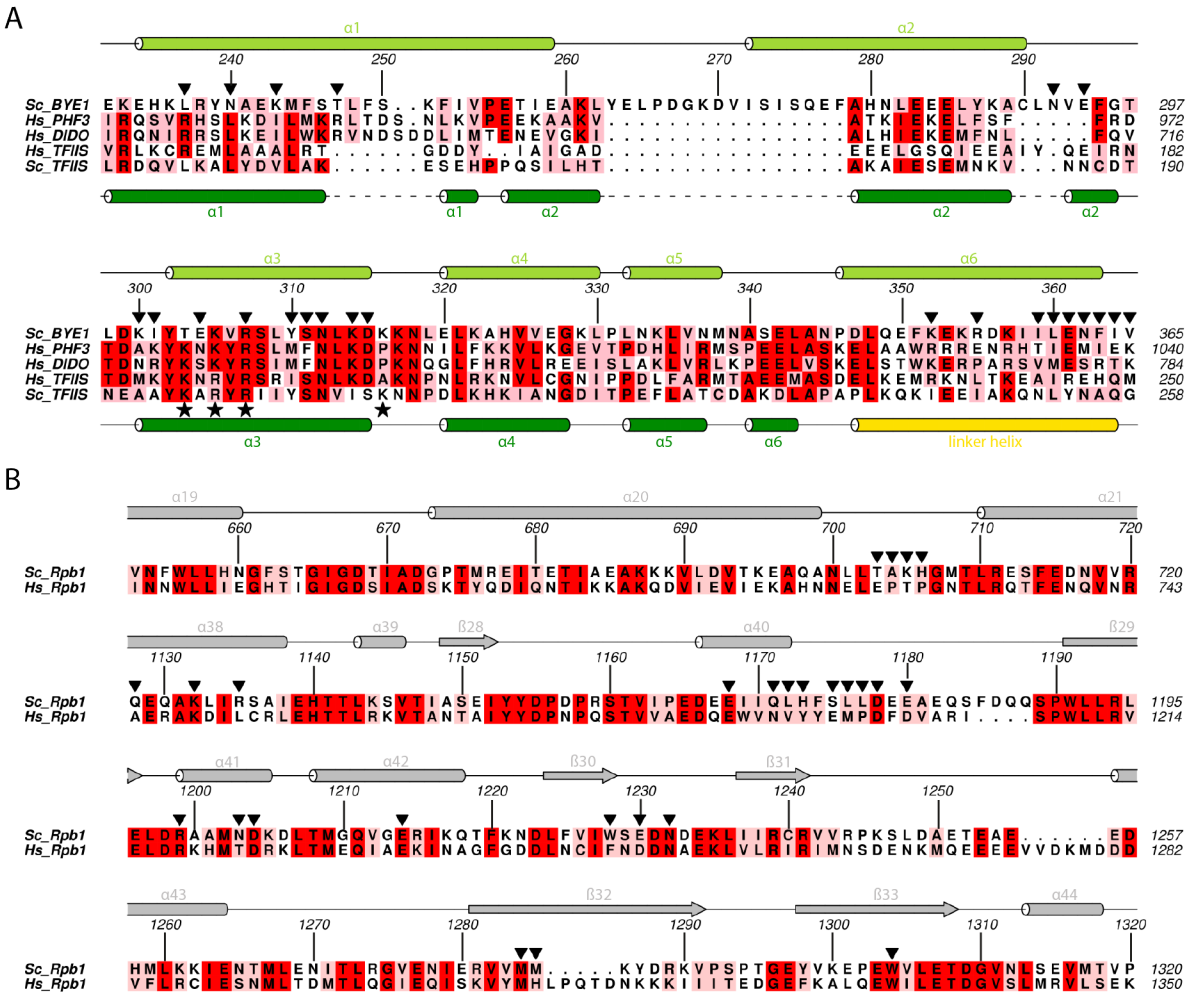

Figure 19: PHF3 and DIDO are human homologues of Bye1.
(A) Amino-acid sequence alignments of *S.cerevisiae* Bye1, *H.sapiens* PHF3, *H.sapiens* DIDO, *H.sapiens* TFIIS and *S.cerevisiae* TFIIS. Secondary structure elements are indicated as arrows (ß strands) or rods (helices). Loops are indicated with solid lines. Residues that are part of the Pol II-Bye1 interface are marked with black triangles. Residues essential for the Pol II-TFIIS interaction (Awrey *et al.*, 1998) are marked with black asterisks. (B) Amino-acid sequence alignments of *S.cerevisiae* Rpb1 and *H.sapiens* Rpb1. Residues that are part of the Pol II-Bye1 interface are marked with black triangles.

9 Further analysis of Bye1 function

Data presented in this chapter have been obtained during this thesis, but have not been published.

9.1 Bye1 genetically interacts with Fyv6

In addition to a genetic interaction of *bye1* with *paf1* and *tho2* (see 8.9), we identified in the same screen a genetic interaction between *bye1* and *fyv6*. Generating *bye1*△*fyv6*△ double mutants in a different genetic background confirmed the synthetic interaction between these genes (Fig. 20A). We also tested for genetic interactions between TFIIS (encoded by the gene *dst1*) and genes interacting with Bye1, but could not observe growth defects. (Fig. 20B)

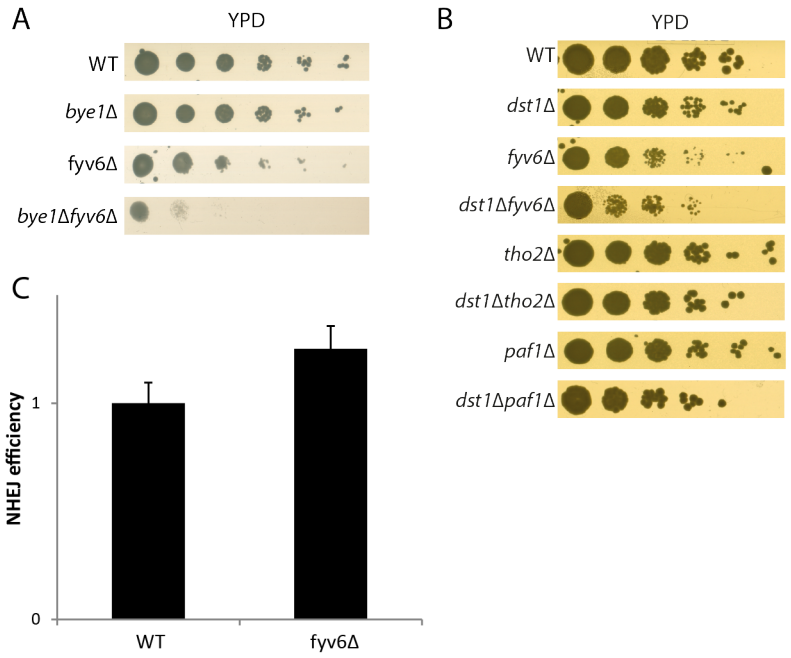

Figure 20: Bye1 genetically interacts with Fyv6.
(A) Serial dilutions of strains *bye1*△, *fyv6*△, *bye1*△*fyv6*△, and an isogenic wild-type (WT) control strain were placed on YPD plates and incubated at 30°C for 3 days. (B) Serial dilutions of strains *dst1*△, *fyv6*△, *dst1*△*fyv6*△, *tho2*△, *dst1*△*tho2*△, *paf1*△, *dst1*△*paf1*△,and an isogenic wild-type (WT) control strain were placed on YPD plates and incubated at 30°C for 3 days. (C) Relative NHEJ efficiency of *fyv6*△ mutants compared to WT cells. Cells were transformed with a linearized vector, and circularization thereof allowed cells to grow on selective plates. Colonies on selective plates were counted to monitor NHEJ efficiency. For details, see 7.2.14.

Based on genome wide screens (Wilson 2002), Fyv6 has been proposed to regulate double-strand break (DSB) repair via non-homologous end-joining (NHEJ). To elucidate this function further, we determined the NHEJ efficiency of *fyv6*△ mutants as described in 7.2.14. As *fyv6*△

mutants did not show any obvious defect in NHEJ (Fig. 20C), the function of Fyv6 and a possible functional link to Bye1 need to be elucidated further.

As Bye1 has been suggested to negatively regulate transcription elongation (Wu *et al.*, 2003), we wanted to investigate this function further in the context of the above described genetic interactions. For this, we screened mutants for their sensitivity to 6-azauracil (6-AU) and mycophenolic acid (MPA). Both 6-AU and MPA lead to a reduction of ribonucleotide levels by inhibiting IMP dehydrogenase (Riles *et al.*, 2004). Therefore, resistance to both 6-AU and MPA depends on fully functional transcription elongation factors for efficient transcription (Archambault *et al.*, 1992). As can be seen in Fig. 21, *fyv6*Δ mutants showed increased sensitivity to 6-AU, and this sensitivity could be partially rescued by *bye1*Δ. No sensitivity to MPA could be observed, and *bye1*Δ*tho2*Δ and *bye1*Δ*paf1*Δ mutants showed no increased sensitivity to either 6-AU or MPA. These findings suggest that Fyv6 function is linked to transcriptional processes, and that Bye1 counteracts this function.

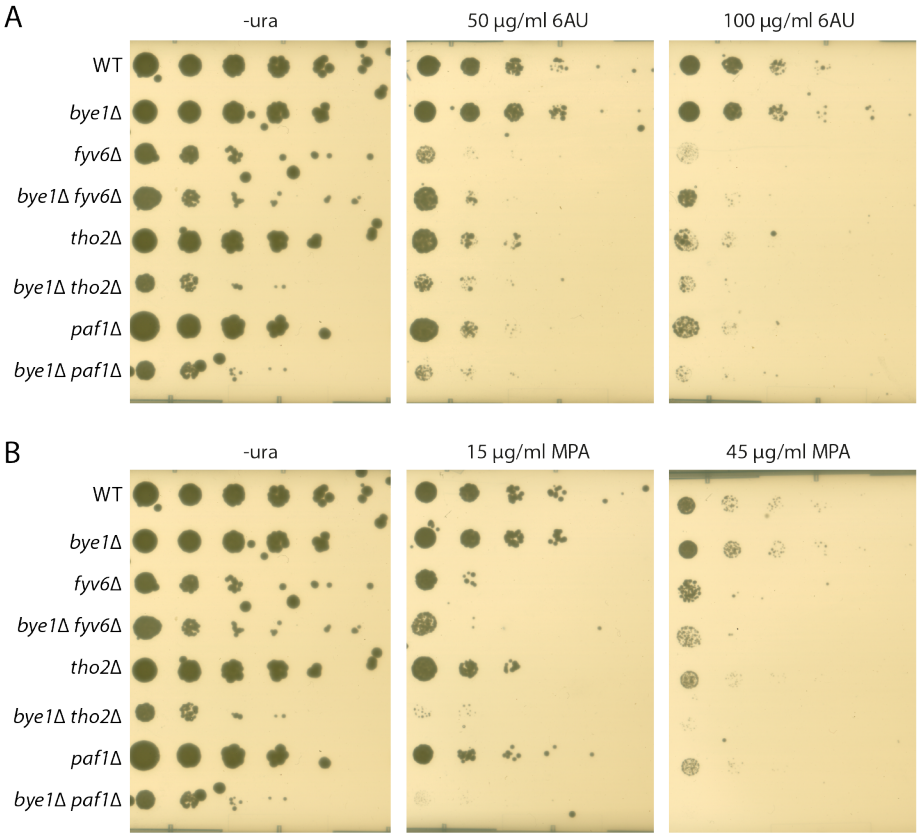

Figure 21: Sensitivity of mutants to 6-AU and MPA.
Serial dilutions of strains *bye1*$\triangle$, *fyv6*$\triangle$, *bye1*$\triangle$*fyv6*$\triangle$, *tho2*$\triangle$, *bye1*$\triangle$*tho2*$\triangle$, *paf1*$\triangle$, *bye1*$\triangle$*paf1*$\triangle$,and an isogenic wild-type (WT) control strain were placed on minimal media plates lacking uracil (-ura) and containing indicated amounts of (A) 6-AU and (B) MPA, and were incubated at 30°C for 3 days.

9.2 Recruitment of Paf1, Tho2, and Pol II is not affected in *bye1*Δ mutants

Based on the observed genetic interactions (8.9 and 9.1), we next asked whether Bye1 is required for recruitment of Paf1, Tho2 and Pol II (Rpb3) to active genes. For this, we performed chromatin immunoprecipitation in WT cells as well as *bye1*Δ mutants and determined occupancy levels on two housekeeping genes, ADH1 and PMA1 (for details, see 7.2.12 and 7.2.12). Both Paf1 and Pol II (Rpb3) showed similar occupancy levels in WT and mutant backgrounds (Fig. 22A, B), indicating that Bye1 is not directly required for their recruitment. Similar results were observed for Tho2 (Fig. 22C), however, occupancy levels were very low and therefore do not allow for detailed interpretation.

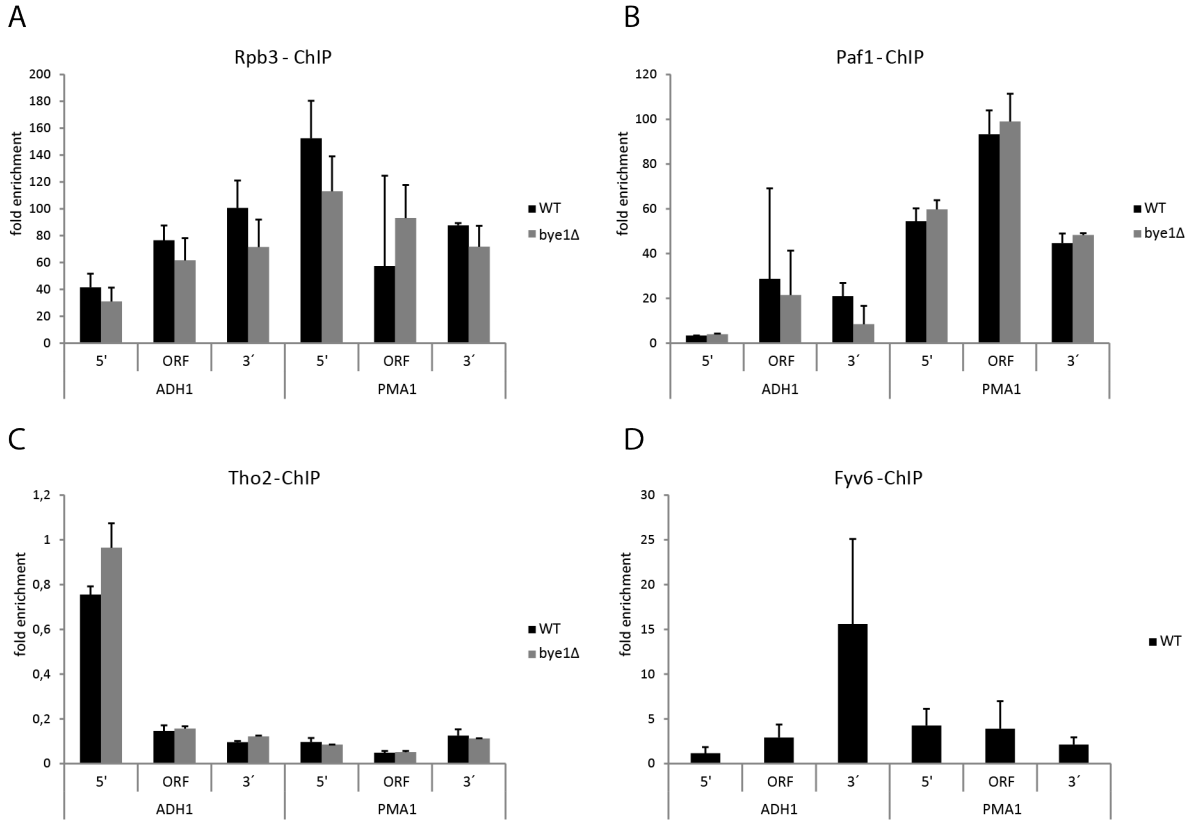

Figure 22: Recruitment of Pol II (Rpb3), Paf1 and Tho2 in WT cells and *bye1*Δmutants.
ChIP-analysis reveals that recruitment of (A) Pol II (Rpb3), (B) Paf1 and (C) Tho2 is not affected in *bye1*Δ mutants (grey) compared to WT cells (black). (D) Occupancy of Fyv6 on housekeeping genes is low and does not show a distinct pattern. The fold enrichments at the ADH1 and PMA1 gene over a non-transcribed region that is located near the centromere of chromosome V are given for the 5' UTR, the ORF region and the 3' UTR. The standard deviations are for two independent ChIP experiments.

As Fyv6 function seemed to be linked to transcriptional processes (see 9.1), we also determined occupancy levels of Fyv6 on housekeeping genes. As shown in Fig. 22D, Fyv6 occupancy

is low on housekeeping genes and does not show a distinct pattern, confirming our previous findings that Fyv6 function is only indirectly linked to transcriptional processes.

9.3 Bye1 is genetically linked to replication fork stalling

To further investigate genetic interactions, above described strains were tested for their sensitivity to DNA damaging agents such as hydroxyurea (HU), Phleomycin, Etoposide and methylmethanesulfonate (MMS) (Fig. 23, and Table 17). HU as well as MMS treatment results in replication fork stalling (Snyder, 1984; Tyrsted, 1982; Slater, 1973; Nickoloff *et al.*, 1998; Tercero & Diffley, 2001), which eventually results in DSBs. Phleomycin treatment results in SSBs as well as DSBs by intercalation into the DNA helix (He *et al.*, 1996). Etoposide in turn inhibits Topoisomerase II, thereby inducing exclusively double strand breaks (Hande, 1998).

Double knockout mutants (*bye1*Δ*fyv6*Δ, *bye1*Δ*tho2*Δ and *bye1*Δ*paf1*Δ) showed increased sensitivity to Phleomycin, whereas no sensitivity could be observed for single knockout mutants (*bye1*Δ, *fyv6*Δ, *tho2*Δ, *paf1*Δ). As none of the mutants showed increased sensitivity to Etoposide, the observed phenotypes suggest that increased sensitivities are a) a direct effect of the Bye1 knockout and b) possibly due to single strand breaks. In addition, increased sensitivity of *fyv6*Δ mutants to HU could be observed, and this phenotype was rescued by *bye1*Δ. Taken together, these findings indicate involvement of Bye1 in HU induced replication fork stalling, or rescue thereof.

One mechanism linking transcription and replication fork stalling is formation of R-loops. R-loops are three-strand nucleic acid structures formed by an RNA:DNA hybrid plus a displaced DNA strand identical to the RNA molecule. The nascent RNA transcript invades the DNA duplex as soon as it exits Pol II during backtracking. Formation of R-loops during transcription has – among other incidents – been observed when replication and transcription machineries collide (Bermejo *et al.*, 2012; Gottipati & Helleday, 2009). To test involvement of Bye1 in R-loop formation, we over-expressed RNase H1 (RNH1) in mutants described above. RNH1 specifically cleaves RNA strands in DNA:RNA duplexes (Crouch *et al.*, 2001), and therefore rescues growth defects resulting from impaired resolving of R-loops.

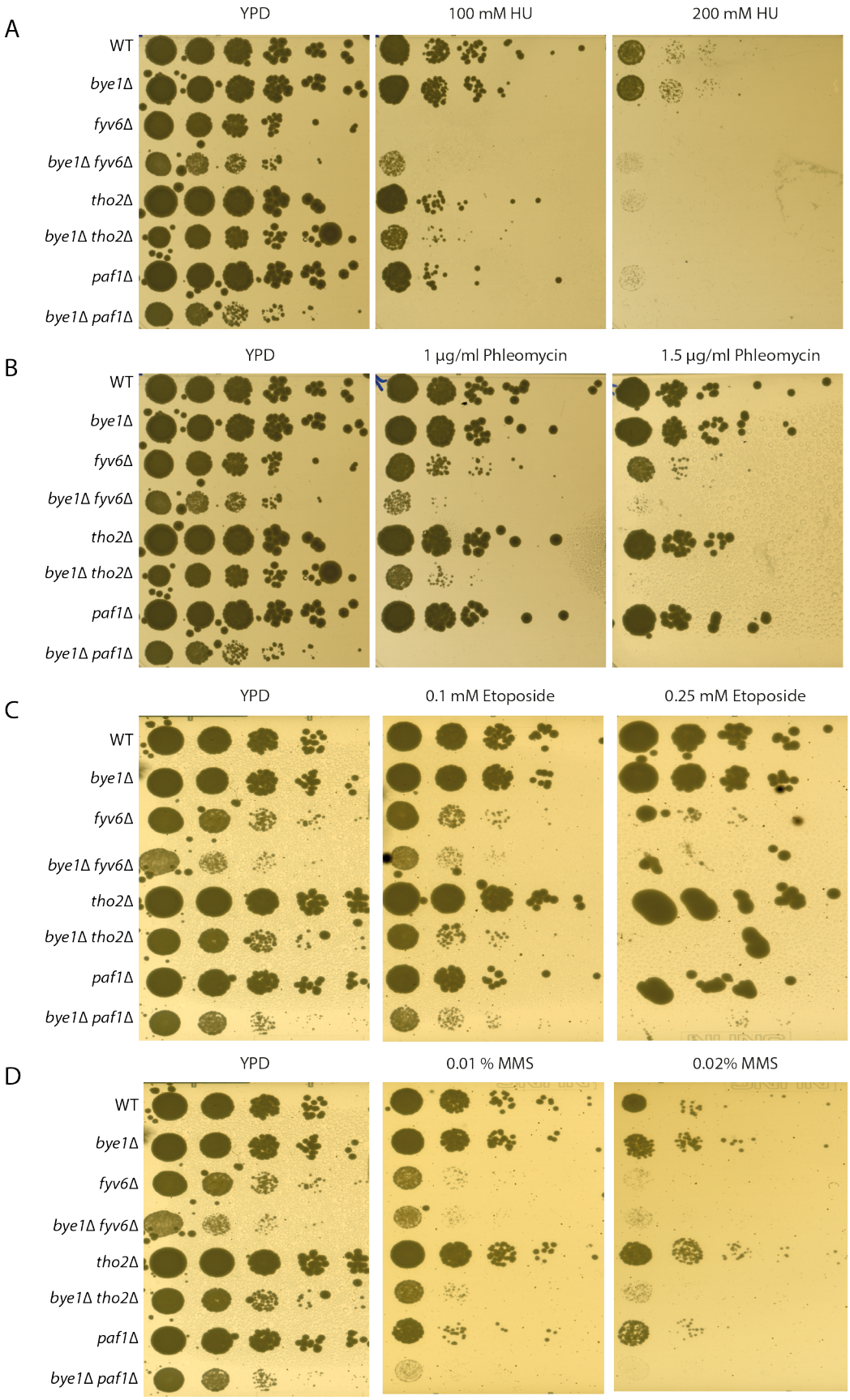

Figure 23: Sensitivity of mutants to DNA damaging agents.
Serial dilutions of strains *bye1Δ*, *fyv6Δ*, *bye1Δ fyv6Δ*, *tho2Δ*, *bye1Δ tho2Δ*, *paf1Δ*, *bye1Δ paf1Δ* , and an isogenic wild-type (WT) control strain were placed on YPD and YPD containing indicated amounts of (A) HU, (B) Phleomycin, (C) Etoposide and (D) MMS, and were incubated at $30\,^{\circ}\mathrm{C}$ for 3 days.

Table 17: Sensitivity of mutants to DNA damaging agents (Summary of Fig. 23).
Growth of WT and mutant cells on media containing different DNA damaging agents is indicated, ranging from normal growth (+++) to no growth/letahlity (-). Parentheses indicate intermediate values.

yeast strain	YPD	HU	Phleomycin	Etoposide	MMS
WT	+++	+(+)	++	++	+(+)
bye1△	+++	+(+)	++	++	+(+)
fyv6△	++	-	+	+	(+)
bye1△*fyv6*△	+	(+)	(+)	+	(+)
tho2△	++(+)	+	++	++	+(+)
bye1△*tho2*△	++	(+)	+	+	(+)
paf1△	++(+)	+	++	++	(+)
bye1△*paf1*△	+	-	-	+	-

As shown in Fig. 24, RNH1 over-expression did not have an effect on growth defects observed before, indicating that tested mutants are not defective in resolving R-loops. To further test whether phenotypes observed in 9.3 result rather from defects in cell cycle progression in general, fluorescence-activated cell sorting (FACS) was carried out as described in 7.2.15. Except for *fyv6*Δ and *bye1*Δ*fyv6*Δ mutants, all mutants showed normal cell cycle progression and DNA content (Fig. 25A). fyv6Δ mutants showed a strong defect in cell cycle progression and aneuploidy (Fig. 25B), and this defect was partially reduced in *bye1*Δ*fyv6*Δ mutants (Fig. 25C).

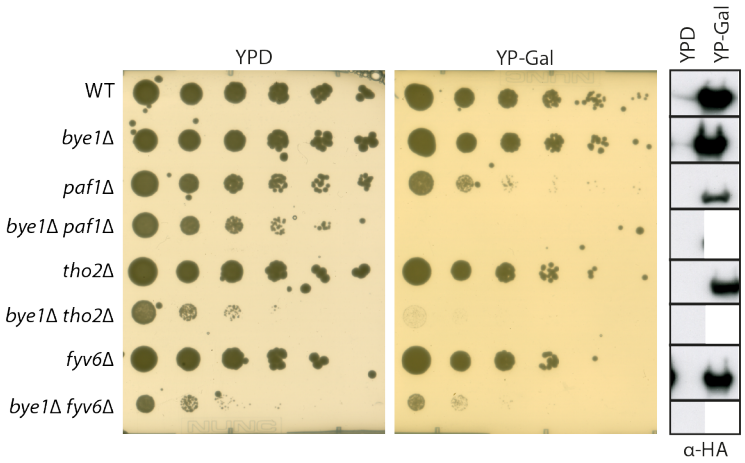

Figure 24: RNase H1 over-expression in mutant yeast cells.
Serial dilutions of strains *bye1*△, *paf1*△, *bye1*△*paf1*△, *tho2*△, *bye1*△*tho2*△, *fyv6*△, *bye*1△*fyv6*△ and an isogenic wild-type (WT) control strain harbouring a plasmid for Gal-induced RNase H1 over-expression were placed on YPD and YP-Gal. Plates were incubated at 30°C for 3 days, and RNase H1 over-expression was confirmed by Westernblotting against HA-tagged RNase H1.

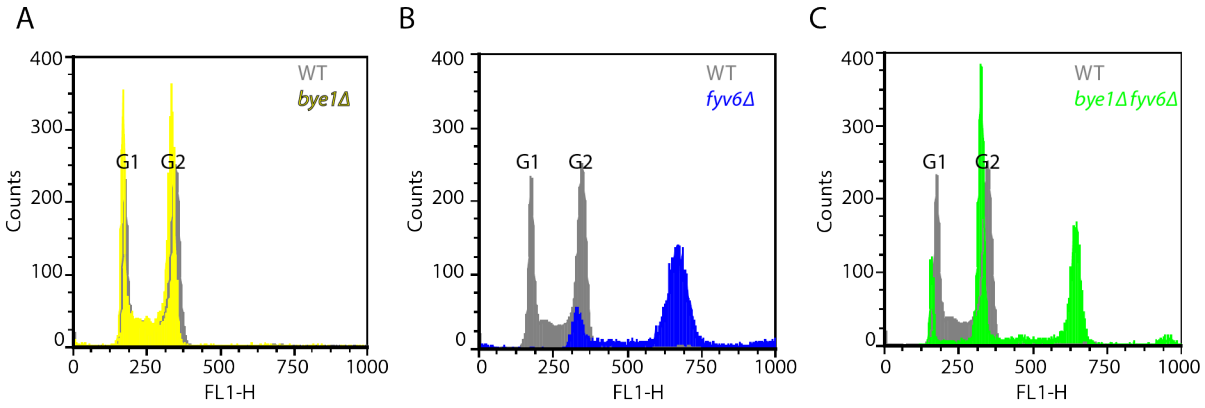

Figure 25: *fyv6*Δ and *bye1*Δ*fyv6*Δ mutants are aneuploid.
FACS analysis of WT and (A) *bye1*Δ, (B) *fyv6*Δ and (C) *bye1*Δ*fyv6*Δ mutants. Cell cycle phases (G1, G2) are indicated.

9.4 Cells over-expressing Bye1 show no growth defects

As depletion of *bye1* in *S.cerevisiae* does not show any obvious phenotypes, we generated a strain over-expressing Bye1 (described in 7.1.3) and checked for growth defects. Over-expression was induced by galactose and confirmed by Westernblotting against HA-tagged Bye1. Strains were placed on YP-Gal plates as well as plates containing 6-AU to check for transcription-related growth defects. As can be seen in Fig. 26, cells over-expressing *bye1* did not show any growth defects or altered sensitivity to 6-AU.

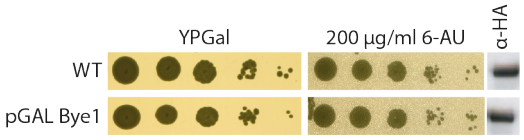

Figure 26: Cells over-expressing *bye1* show no growth defects.
Serial dilutions of wild-type (WT) control strain and a strain harbouring a genomically inserted Gal-promoter upstream of the Bye1 ORF (pGAL Bye1). Cells were placed on YP-Gal and plates containing 200 µg/ml 6-AU and incubated at 30°C for 3 days. Bye1 over-expression was confirmed by WB against HA-tagged Bye1.

9.5 Bye1 and TFIIS compete for Pol II-binding

As the structure of Bye1-bound Pol II revealed binding of Bye1 to Pol II at the same position as TFIIS (see Fig. 10), we wondered whether there is a direct competition for Pol II binding between Bye1 and TFIIS. To test this, we formed a stable Pol II-Bye1 complex (see 7.2.2) containing a mismatch scaffold (see 6.3). After addition of TFIIS, we monitored TFIIS-induced RNA cleavage, which indicates Pol II-TFIIS complex formation and therefore Bye1 displacement from the complex. As can be seen in Fig. 27, TFIIS-induced RNA cleavage could be observed

in the presence of Bye1, even at a 30X molar excess over Pol II. This indicates that TFIIS can displace Bye1 *in vitro*, and that TFIIS affinity to Pol II is higher than that of Bye1. These findings fit with a previously published dissociation constant for the Pol II-TFIIS interaction of 60 nM (Awrey *et al.* 1998), which is more than 60X stronger than that for the Pol II-Bye1 interaction (8.1).

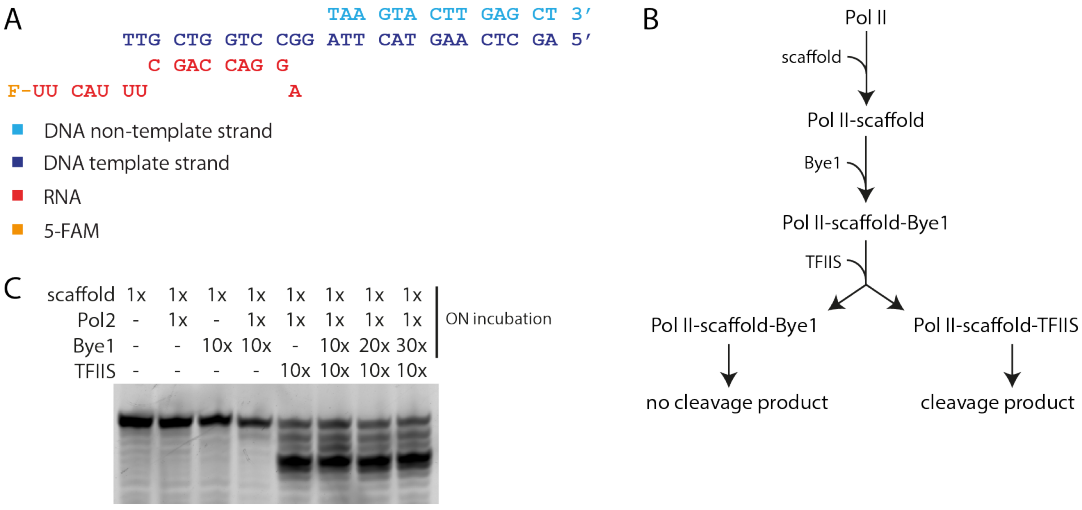

Figure 27: TFIIS has a higher affinity to Pol II than Bye1.
(A) Nucleic acid scaffold for reconstitution of Pol II-nucleic acid complex. (B) Scheme depicting order of addition of components. (C) Electrophoretic separation of RNA cleavage products.

To further investigate a possible competition, we determined occupancy levels of Bye1 and TFIIS at housekeeping genes in WT cells, and *dst1*Δ and *bye1*Δ mutants, respectively. Occupancy levels for Bye1 and TFIIS were similar in WT cells, and were neither affected for Bye1 nor TFIIS in mutant backgrounds (Fig. 28A, B), suggesting that Pol II-Bye1 and Pol II-TFIIS complexes co-exist in cells. As it has been reported that TFIIS sustains Pol II activity in ribosomal protein (RP) genes under transcriptional stress, and is therefore required for the balance between rDNA transcripts and RPs mRNAs (Gómez-Herreros *et al.* 2012), we determined occupancy levels of Bye1 and TFIIS at a RP gene in WT cells, and *dst1*Δ and *bye1*Δ mutants, respectively. Bye1 occupancy levels were unchanged in a *dst1*Δ background compared to WT (Fig. 28C), whereas increased occupancy levels for TFIIS were observed in a *bye1*Δ background, compared to WT (Fig. 28D). These results indicate that Bye1 functions upstream of TFIIS in transcriptional regulation of RP genes, and that the presence of Bye1 limits the number of TFIIS molecules present at these genes.

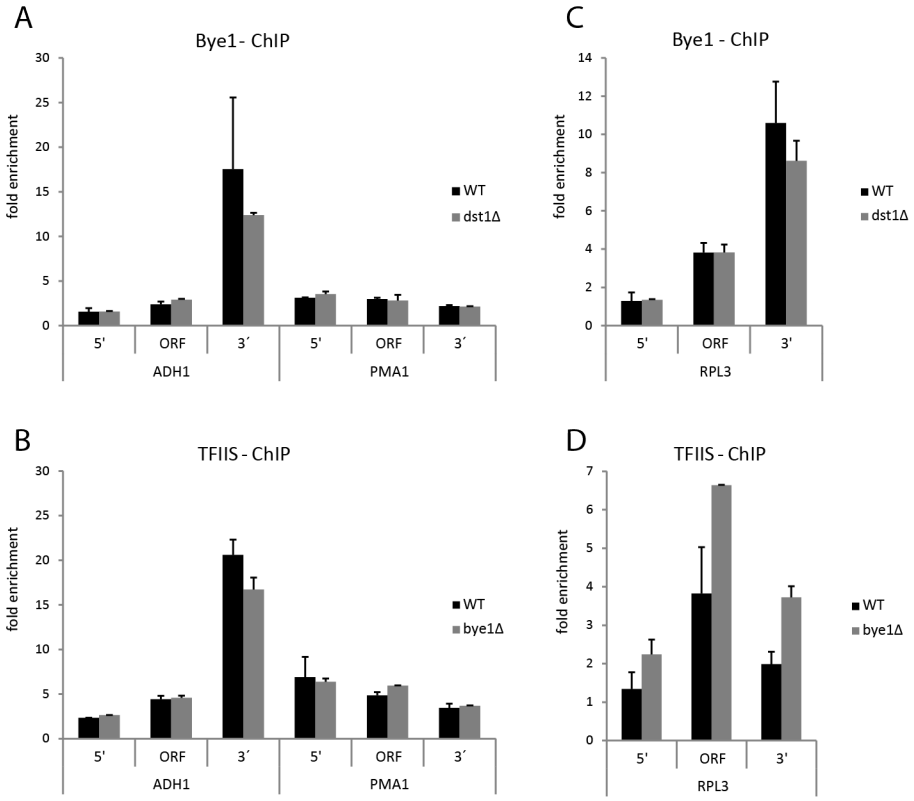

Figure 28: TFIIS recruitment to ribosomal protein genes is altered in *bye1*Δ mutants.
ChIP-analysis reveals that recruitment to housekeeping genes of (A) Bye1 and (B) TFIIS is not affected in *dst1*Δ and *bye1*Δ mutants (grey), respectively, compared to WT cells (black). Similarly, recruitment of Bye1 to the ribosomal protein gene RPL3 is not affected in *dst1*Δ mutants (grey) compared to WT cells (black) (C). In contrast, TFIIS occupancy is increased in *bye1*Δ mutants (grey) compared to WT cells (black) (D). The fold enrichments at the ADH1, PMA1 and RPL3 gene over a non-transcribed region that is located near the centromere of chromosome V are given for the 5' UTR, the ORF region and the 3' UTR. The standard deviations are for two independent ChIP experiments.

Part IV

Conclusions and outlook

From the work described here, we could show that Bye1 binds to Pol II, and report crystal structures of the central TLD domain of Bye1 bound to free Pol II, a Pol II elongation complex with DNA template and RNA transcript, an elongation complex with an NTP analogue, and an arrested elongation complex with backtracked RNA. These studies represent only the third high-resolution structural analysis of a transcription factor complex with the polymerase core. Whereas the previously studied factors TFIIB (Bushnell *et al.*, 2004; Kostrewa *et al.*, 2009; Liu *et al.*, 2010; Sainsbury *et al.*, 2013) and TFIIS (Kettenberger *et al.*, 2003, 2004; Cheung & Cramer, 2011; Wind & Reines, 2000; Wang *et al.*, 2009) alter Pol II function by directly affecting catalytic events, Bye1 does not alter basic Pol II functions *in vitro*. Consistent with this, Bye1 binding to Pol II does not alter Pol II conformation in the structures. Additional functional data *in vitro* and *in vivo* indicate that Bye1 occupies active genes in their 5'-region and can bind to histone H3 tails with post-translational modifications (PTMs) of active transcription using its PHD domain. A chromatin-related function of Bye1 may explain published genetic evidence for a negative role of Bye1 in transcription elongation (Wu *et al.*, 2003).

What could be the function of Bye1 in chromatin transcription? One possibility is the prevention of histone loss during polymerase passage through chromatin. Because the TLD of Bye1 is required for chromatin association of Bye1, it is unlikely that Bye1 first recognizes active chromatin marks, and then recruits Pol II to active chromatin regions. Instead we speculate that Bye1 binds directly to Pol II during early transcription elongation and tethers surrounding histones containing active PTMs. This hypothesis also fits our observations that Bye1 does not alter Pol II conformation or function, but rather uses Pol II as a binding platform. Bye1 may cooperate with other chromatin elongation factors such as Spt6 and FACT (Duina, 2011; Winkler & Luger, 2011) to prevent loss of histones during polymerase passage through chromatin. Therefore, Bye1 might be the founding member of a new family of transcription factors that tether early transcribing Pol II to histones, thereby possibly preventing deregulated transcription that may arise from histone loss. Such a mechanism may account for the apparent function of the human Bye1 homologues PHF3 and DIDO in suppressing cancer development (Fischer *et al.*, 2001; Fütterer *et al.*, 2005). To strengthen this hypothesis, Bye1 function would have to be elucidated further in a chromatin background, possibly by extending established methods of nucleosome localization towards monitoring histone loss in *bye1*Δ mutants. In addition, involvement of the human Bye1 homologues in chromatin transcription should be

analysed to support the hypothesis and to transfer the findings from yeast to humans.

A second possibility is the involvement of Bye1 in cryptic transcription. Unpublished data from the Strahl lab showed that over-expression of Bye1 induces transcription from cryptic promoters in a yeast strain harboring a reporter for cryptic transcription, and this was dependent on the TLD domain. These findings fit with our Bye1 ChIP data, where increased occupancy was observed downstream of the polyA site, indicating divergent transcription. Further indication arises from the genetic link between Bye1 and Ess1, which has been shown to repress the initiation of cryptic unstable transcripts (CUTs) (Ma *et al.*, 2012). Finally, we showed binding of Bye1 to trimethylated H3K4, which has been linked to transcription from cryptic promoters (Pinskaya *et al.*, 2009). To corroborate involvement of Bye1 in cryptic transcription, Bye1 ChIP data has to be analysed in depth in relation to cryptic promoters, and a possible link between Bye1 and the Nrd1-dependent termination of non-coding RNAs (Schulz, Schwalb et al., submitted) should be elucidated, as Nrd1 recruitment is dependent on H3K4me3 and Set1 (Terzi *et al.*, 2011).

Finally, a function of Bye1 in transcription of ribosomal protein (RP) genes could be investigated further. ChIP-occupancies of Bye1 and TFIIS on the RP gene Rpl3 suggested that Bye1 functions upstream and counteracts recruitment of TFIIS, as TFIIS occupancy levels are increasing upon depletion of Bye1. As a role for TFIIS in sustaining Pol II levels on RP genes under transcriptional stress has been described (Gómez-Herreros *et al.*, 2012), occupancy levels of Bye1 and TFIIS on all RP genes, and in mutant backgrounds, should be investigated under transcriptional stress.

In addition, characterization of the SPOC domain should be a major focus to further elucidate Bye1 function. The function of SPOC domains in yeast is unclear, but in higher eukaryotes SPOC domains are implicated in developmental signalling (Ariyoshi & Schwabe, 2003), and are involved in protein-protein interactions (Hiriart *et al.*, 2005; Debeauchamp *et al.*, 2008). Therefore, an interaction screen should be carried out in order to identify possible interacting partners of the SPOC domain. This could be done either by using the SPOC domain as bait in a yeast-two-hybrid screen, or by using a recombinant SPOC domain for pull-down assays with yeast lysate. Crystallization and structure solution of the SPOC domain may allow for further structural and possibly even functional characterization. Unpublished data from our and other labs implied that the human Bye1 homolog, PHF3, interacts with the C-terminal domain of Rpb1, the largest Pol II subunit, and we believe that the SPOC domain could be involved in this interaction. Specifically, PHF3 interacts with Tyrosine1-Serine2-phosphorylated (Y1PS2P) CTD peptides, and this phosphorylation pattern is present during productive transcription elongation (Heidemann *et al.*, 2013). An interaction of Bye1 and

specifically the SPOC domain with differentially phosphorylated CTD peptides should therefore be tested, and if an interaction can be proven, crystallization of a SPOC-CTD complex would be the logical next step.

Supplemental Table

Table 18: List of microarrayed peptides

Peptide	Residues	Sequence	Annotation
P1	H3 1-20	ARTKQTARKSTGGKAPRKQL-K(Biot)-NH2	H3 (1-20)
P2	H3 1-20	ARTKQTARKSTGGK(Ac)APRKQL-K(Biot)-NH2	H3K14ac
P3	H3 1-20	ARTKQTARK(Ac)STGGKAPRKQL-K(Biot)-NH2	H3K9ac
P4	H3 1-20	ARTK(Ac)QTARKSTGGKAPRKQL-K(Biot)-NH2	H3K4ac
P5	H3 1-20	ARTK(Ac)QTARKSTGGK(Ac)APRKQL-K(Biot)-NH2	H3K4ac + K14ac
P6	H3 1-20	ARTKQTARK(Ac)STGGK(Ac)APRKQL-K(Biot)-NH2	H3K9ac + K14ac
P7	H3 1-20	ARTK(Ac)QTARK(Ac)STGGKAPRKQL-K(Biot)-NH2	H3K4ac + K9ac
P8	H3 1-20	ARTK(Ac)QTARK(Ac)STGGK(Ac)APRKQL-K(Biot)-NH2	H3K4ac + K9ac + K14ac
P10	H3 1-20	ARTKQTARKSTGGKAPRK(Ac)QL-K(Biot)-NH2	H3K18ac
P11	H3 1-20	ARTKQTARKSTGGK(Ac)APRK(Ac)QL-K(Biot)-NH2	H3K14ac + K18ac
P12	H3 1-20	ARTKQTARK(Ac)STGGKAPRK(Ac)QL-K(Biot)-NH2	H3K9ac + K18ac
P13	H3 1-20	ARTK(Ac)QTARKSTGGKAPRK(Ac)QL-K(Biot)-NH2	H3K4ac + K18ac
P14	H3 1-20	ARTKQTARK(Ac)STGGK(Ac)APRK(Ac)QL-K(Biot)-NH2	H3K9ac + K14ac + K18ac
P15	H3 1-20	ARTK(Ac)QTARKSTGGK(Ac)APRK(Ac)QL-K(Biot)-NH2	H3K4ac + K14ac + K18ac
P16	H3 1-20	ARTK(Ac)QTARK(Ac)STGGKAPRK(Ac)QL-K(Biot)-NH2	H3K4ac + K9ac + K18ac
P17	H3 1-20	ARTK(Ac)QTARK(Ac)STGGK(Ac)APRK(Ac)QL-K(Biot)-NH2	H3K4ac + K9ac + K14ac + K18ac
P18	H3 1-20	ARTK(Me3)QTARKSTGGKAPRKQL-K(Biot)-NH2	H3K4me3

P19	H3 1-20	ARTK(Me3)QTARK(Ac)STGGKAPRKQL-K(Biot)-NH2	H3K4me3 + K9ac
P20	H3 1-20	ARTK(Me3)QTARKSTGGK(Ac)APRKQL-K(Biot)-NH2	H3K4me3 + K14ac
P21	H3 1-20	ARTK(Me3)QTARKSTGGKAPRK(Ac)QL-K(Biot)-NH2	H3K4me3 + K18ac
P22	H3 1-20	ARTK(Me3)QTARK(Ac)STGGK(Ac)APRKQL-K(Biot)-NH2	H3K4me3 + K9ac + K14ac
P23	H3 1-20	ARTK(Me3)QTARK(Ac)STGGKAPRK(Ac)QL-K(Biot)-NH2	H3K4me3 + K9ac + K18ac
P24	H3 1-20	ARTK(Me3)QTARKSTGGK(Ac)APRK(Ac)QL-K(Biot)-NH2	H3K4me3 + K14ac + K18ac
P25	H3 1-20	ARTK(Me3)QTARK(Ac)STGGK(Ac)APRK(Ac)QL-K(Biot)-NH2	H3K4me3 + K9ac + K14ac + K18ac
P26	H3 1-20	ARpTK(Me3)QTARK(Ac)STGGK(Ac)APRK(Ac)QL-K(Biot)-NH2	H3T3p + K4me3 + K9ac + K14ac + K18ac
P27	H3 1-20	ARpTK(Me3)QTARKSTGGKAPRKQL-K(Biot)-NH2	H3T3p + K4me3
P28	H3 1-20	AR(Me2a)pTK(Me3)QTARK(Ac)STGGK(Ac)APRK(Ac)QL-K(Biot)-NH2	H3R2me2a + T3p + K4me3 + K9ac + K14ac + K18ac
P29	H3 1-20	AR(Me2a)pTK(Me3)QTARKSTGGKAPRKQL-K(Biot)-NH2	H3R2me2a + T3p + K4me3
P30	H3 1-20	AR(Me2a)TK(Me3)QTARKSTGGKAPRKQL-K(Biot)-NH2	H3R2me2a + K4me3
P32	H3 1-20	ARTK(Me2)QTARKSTGGKAPRKQL-K(Biot)-NH2	H3K4me2
P33	H3 1-20	ARTK(Me2)QTARK(Ac)STGGK(Ac)APRK(Ac)QL-K(Biot)-NH2	H3K4me2 + K9ac + K14ac + K18ac
P34	H3 1-20	ARTK(Me)QTARKSTGGKAPRKQL-K(Biot)-NH2	H3K4me1
P35	H3 1-20	ARTK(Me)QTARK(Ac)STGGK(Ac)APRK(Ac)QL-K(Biot)-NH2	H3K4me1 + K9ac + K14ac + K18ac
P36	H3 1-20	ARTKQTARKpSTGGKAPRKQL-K(Biot)-NH2	H3S10p
P37	H3 1-20	ARTK(Ac)QTARK(Ac)pSTGGK(Ac)APRK(Ac)QL-K(Biot)-NH2	H3K4ac + K9ac + S10p + K14ac + K18ac

P38	H3 1-20	ARTK(Me3)QTARKpSTGGKAPRKQL-K(Biot)-NH2	H3K4me3 + S10p
P39	H3 1-20	ARTK(Me3)QTARK(Ac)pSTGGK(Ac)APRK(Ac)QL-K(Biot)-NH2	H3K4me3 + K9ac + S10p + K14ac + K18ac
P40	H3 1-20	AR(Me2a)TK(Me3)QTARKpSTGGKAPRKQL-K(Biot)-NH2	H3R2me2a + K4me3 + S10p
P41	H3 1-20	AR(Me2a)TK(Me3)QTARK(Ac)pSTGGK(Ac)APRK(Ac)QL-K(Biot)-NH2	H3R2me2a + K4me3 + K9ac + S10p + K14ac + K18ac
P42	H3 1-20	ARTKQTARK(Me3)STGGKAPRKQL-K(Biot)-NH2	H3K9me3
P43	H3 1-20	ARTK(Ac)QTARK(Me3)STGGK(Ac)APRK(Ac)QL-K(Biot)-NH2	H3K4ac + K9me3 + K14ac + K18ac
P44	H3 1-20	ARTK(Me2)QTARK(Ac)STGGKAPRK(Ac)QL-K(Biot)-NH2	H3K4me2 + K9ac + K18ac
P45	H3 1-20	ARTK(Me)QTARK(Ac)STGGKAPRK(Ac)QL-K(Biot)-NH2	H3K4me1 + K9ac + K18ac
P47	H3 1-20	AR(Me2a)TKQTARKSTGGKAPRKQL-K(Biot)-NH2	H3R2me2a
P48	H3 1-20	AR(Me2a)TK(Ac)QTARK(Ac)STGGK(Ac)APRK(Ac)QL-K(Biot)-NH2	H3R2me2a + K4ac + K9ac + K14ac + K18ac
P50	H3 1-20	AR(Me2a)TK(Me3)QTARK(Ac)STGGK(Ac)APRK(Ac)QL-K(Biot)-NH2	H3R2me2a + K4me3 + K9ac + K14ac + K18ac
P51	H3 1-20	AR(Me)TK(Me3)QTARKSTGGKAPRKQL-K(Biot)-NH2	H3R2me1 + K4me3
P52	H3 1-20	AR(Me)TK(Me3)QTARK(Ac)STGGK(Ac)APRK(Ac)QL-K(Biot)-NH2	H3R2me1 + K4me3 + K9ac + K14ac + K18ac
P53	H3 1-20	ACitTKQTARKSTGGKAPRKQL-K(Biot)-NH2	H3Cit2
P54	H3 1-20	ACitTK(Me3)QTARKSTGGKAPRKQL-K(Biot)-NH2	H3Cit2 + K4me3
P55	H3 1-20	ACitTK(Me3)QTARK(Ac)STGGK(Ac)APRK(Ac)QL-K(Biot)-NH2	H3Cit2 + K4me3 + K9ac + K14ac + K18ac
P56	H3 1-20	ACitTK(Ac)QTARK(Ac)STGGK(Ac)APRK(Ac)QL-K(Biot)-NH2	H3Cit2 + K4ac + K9ac + K14ac + K18ac
P57	H3 1-20	ARpTKQTARKSTGGKAPRKQL-Peg-K(Biot)-NH2	H3T3p

P58	H4 1-23	Ac-SGRGK5GGKGLGKGGAKRHRKVLR-Peg-Biot	H4 (1-23)
P59	H4 1-23	Ac-SGRGK(Ac)GGK(Ac)GLGK(Ac)GGAK(Ac)RHRKVLR-Peg-Biot	H4K5ac + K8ac + K12ac + K16ac
P60	H3 1-20	AR(Me2a)TK(Me2)QTARKSTGGKAPRKQL-K(Biot)-NH2	H3R2me2a + K4me2
P61	H3 1-20	AR(Me2s)TK(Me2)QTARKSTGGKAPRKQL-K(Biot)-NH2	H3R2me2s + K4me2
P62	H3 1-20	AR(Me)TK(Me2)QTARKSTGGKAPRKQL-K(Biot)-NH2	H3R2me1 + K4me2
P63	H3 1-20	ACitTK(Me2)QTARKSTGGKAPRKQL-K(Biot)-NH2	H3Cit2 + K4me2
P66	H4 1-23	Ac-SGRGK(Ac)GGKGLGKGGAKRHRKVLR-Peg-Biot	H4K5ac
P67	H4 1-23	Ac-SGRGKGGK(Ac)GLGKGGAKRHRKVLR-Peg-Biot	H4K8ac
P68	H4 1-23	Ac-SGRGKGGKGLGK(Ac)GGAKRHRKVLR-Peg-Biot	H4K12ac
P69	H4 1-23	Ac-SGRGKGGKGLGKGGAK(Ac)RHRKVLR-Peg-Biot	H4K16ac
P70	H4 1-23	Ac-SGRGK(Ac)GGKGLGK(Ac)GGAKRHRKVLR-Peg-Biot	H4K5ac + K12ac
P71	H4 1-23	Ac-SGRGKGGK(Ac)GLGKGGAK(Ac)RHRKVLR-Peg-Biot	H4K8ac + K16ac
P72	H4 1-23	Ac-SGRGK(Ac)GGK(Ac)GLGK(Ac)GGAKRHRKVLR-Peg-Biot	H4K5ac + K8ac + K12ac
P73	H4 1-23	Ac-SGR(Me2a)GKGGKGLGKGGAKRHRKVLR-K(Biot)-NH2	H4R3me2a
P74	H4 1-23	Ac-SGR(Me2s)GKGGKGLGKGGAKRHRKVLR-K(Biot)-NH2	H4R3me2s
P75	H4 1-23	Ac-SGR(Me)GKGGKGLGKGGAKRHRKVLR-K(Biot)-NH2	H4R3me1
P76	H4 1-23	Ac-pSGR(Me2a)GKGGKGLGKGGAKRHRKVLR-K(Biot)-NH2	H4S1p + R3me2a
P77	H4 1-23	Ac-pSGR(Me2s)GKGGKGLGKGGAKRHRKVLR-K(Biot)-NH2	H4S1p + R3me2s

P78	H4 1-23	Ac-pSGR(Me)GKGGKGLGKGGAKRHR KVLR-K(Biot)-NH2	H4S1p + R3me1
P79	H4 1-23	Ac-SGR(Me2a)GK(Ac)GGK(Ac)GLGK(Ac) GGAK(Ac)RHRK(Ac)VLR-K(Biot)-NH2	H4R3me2a + K5ac + K8ac + K12ac + K16ac + K20ac
P80	H4 1-23	Ac-SGR(Me2s)GK(Ac)GGK(Ac)GLGK(Ac) GGAK(Ac)RHRK(Ac)VLR-K(Biot)-NH2	H4R3me2s + K5ac + K8ac + K12ac + K16ac + K20ac
P81	H4 1-23	Ac-SGR(Me)GK(Ac)GGK(Ac)GLGK(Ac) GGAK(Ac)RHRK(Ac)VLR-K(Biot)-NH2	H4R3me1 + K5ac + K8ac + K12ac + K16ac + K20ac
P82	H4 11-27	Ac-GKGGAKRHRK(Me3)VLRDNIQ-Peg-Biot	H4K20me3
P83	H4 11-27	Ac-GKGGAKRHRK(Me2)VLRDNIQ-Peg-Biot	H4K20me2
P84	H4 11-27	Ac-GKGGAKRHRK(Me)VLRDNIQ-Peg-Biot	H4K20me1
P85	H4 11-27	Ac-GK(Ac)GGAK(Ac)RHRK(Me3)VLRD NIQ-Peg-Biot	H4K12ac + K16ac + K20me3
P86	H4 11-27	Ac-GK(Ac)GGAK(Ac)RHRK(Me2)VLRD NIQ-Peg-Biot	H4K12ac + K16ac + K20me2
P89	H3 1-20	ARTK(Me3)QTAR(Me2s)K(Me3)STGGK APRKQL-K(Biot)-NH2	H3K4me3 + R8me2s + K9me3
P90	H3 15-43	Ac-APRK18QLATK23AARK27SAPSTG GVK36K37PHRYGGK(Biot)-NH2	H3 (15-41)
P91	H3 15-43	Ac-APRK(Me3)QLATKAARKSAPSTGG VKKPHRY-GG-K(Biot)-NH2	H3K18me3
P93	H3 15-43	Ac-APRKQLATKAARKSAPSTGGV K(Me3)KPHRY-GG-K(Biot)-NH2	H3K36me3
P95	H3 15-43	Ac-APRK(Me3)QLATKAARKSAPSTGG VK(Me3)KPHRY-GG-K(Biot)-NH2	H3K18me3 + K36me3
P96	H3 1-20	ARTK(Me3)QTAR(Me2a)K(Me3)STGGK APRKQL-K(Biot)-NH2	H3K4me3 + R8me2a + K9me3
P99	H4 11-27	Ac-GKGGAKRHRKVLRDNIQ-Peg-Biot	H4 (11-27)
P100	H3 74-84	Ac-IAQDFKTDLRF-Peg-K(Biot)-NH2	H3 (74-84) N-ac
P101	H3 74-84	Ac-IAQDFK(Me3)TDLRF-Peg-K(Biot)-NH2	H3K79me3

P102	H3 74-84	Ac-IAQDFK(Me2)TDLRF-Peg-K(Biot)-NH2	H3K79me2
P103	H3 74-84	Ac-IAQDFK(Me)TDLRF-Peg-K(Biot)-NH2	H3K79me1
P104	H3 74-84	IAQDFKTDLRF-Peg-K(Biot)-NH2	H3 (74-84)
P120	H3 27-45	KSAPSTGGVK(Me3)KPHRYKPGT-G-K(Biot)-NH2	H3K36me3 (27-45)
P121	H3 27-45	KSAPSTGGVK(Me2)KPHRYKPGT-GG-K(Biot)-NH2	H3K36me2 (27-45)
P123	H3 27-45	KSAPSTGGVK(Ac)KPHRYKPGT-GG-K(Biot)-NH2	H3K36ac (27-45)
P124	H3 27-45	KSAPSTGGVKKPHRYKPGT-GG-K(Biot)-NH2	H3 (27-45)
P125	H3 1-20	ARpTKQTARKSTGGKAPRKQL-K(Biot)-NH2	H3T3p
P129	H3 6-30	Ac-TARK(Me2)STGGKAPRKQLATKAARK(Me2)SAP-Peg-K(Biot)-NH2	H3K9me2 + K27me2
P132	H3 1-20	ARTK(Me3)QTARK(Me3)STGGKAPRKQL-K(Biot)-NH2	H3K4me3 + K9me3
P133	H3 1-20	ARTKQTARK(Me2)STGGKAPRKQL-K(Biot)-NH2	H3K9me2
P134	H3 1-20	ARTKQTARK(Me)STGGKAPRKQL-K(Biot)-NH2	H3K9me1
P137	H3 1-20	ARTKQTARKSTGGKAPRK(Me3)QL-K(Biot)-NH2	H3K18me3
P138	H3 1-20	ARTKQTARKSTGGKAPRK(Me2)QL-K(Biot)-NH2	H3K18me2
P139	H3 1-20	ARTKQTARKSTGGKAPRK(Me)QL-K(Biot)-NH2	H3K18me1
P144	H3 1-20	ARTKQTARK(Ac)pSTGGKAPRKQL-K(Biot)-NH2	H3K9ac + S10p
P145	H3 1-20	ARTKQTARK(Me3)pSTGGKAPRKQL-K(Biot)-NH2	H3K9me3 + S10p
P146	H3 1-20	ARTKQTARK(Me2)pSTGGKAPRKQL-K(Biot)-NH2	H3K9me2 + S10p
P147	H3 1-20	ARTKQTARK(Me)pSTGGKAPRKQL-K(Biot)-NH2	H3K9me1 + S10p

P148	H3 1-20	ARTK(Me3)QTARK(Ac)pSTGGKAPRK QL-K(Biot)-NH2	H3K4me3 + K9ac + S10p
P149	H3 1-22	ARTKQTARKSTGGKAPR(Me2a)KQLA T-K(Biot)-NH2	H3R17me2a
P150	H3 1-22	ARTKQTARKSTGGKAPR(Me2s)KQLA T-K(Biot)-NH2	H3R17me2s
P151	H3 1-22	ARTKQTARKSTGGKAPR(Me)KQLAT-K(Biot)-NH2	H3R17me1
P157	H3 1-20	AR(Me2s)TK(Me3)QTARKSTGGKAPR KQL-K(Biot)-NH2	H3R2me2s + K4me3
P162	H3 1-20	ARTKQpTARKSTGGKAPRKQL-K(Biot)-NH2	H3T6p
P163	H3 1-20	ARTK(Me3)QpTARKSTGGKAPRKQL-K(Biot)-NH2	H3K4me3 + T6p
P164	H3 1-20	ARTK(Me2)QpTARKSTGGKAPRKQL-K(Biot)-NH2	H3K4me2 + T6p
P165	H3 1-20	ARTKQpTARK(Ac)STGGK(Ac)APR K(Ac)QL-K(Biot)-NH2	H3T6p + K9ac + K14ac + K18ac
P166	H3 1-20	ARTK(Me3)QpTARK(Ac)STGGK(Ac)AP RK(Ac)QL-K(Biot)-NH2	H3K4me3 + T6p + K9ac + K14ac + K18ac
P167	H3 1-20	ARTK(Me2)QpTARK(Ac)STGGK(Ac)AP RK(Ac)QL-K(Biot)-NH2	H3K4me2 + T6p + K9ac + K14ac + K18ac
P174	H3 1-20	AR(Me2s)TK(Me3)QTARK(Ac)STGG K(Ac)APRK(Ac)QL-K(Biot)-NH2	H3R2me2s + K4me3 + K9ac + K14ac + K18ac
P178	H3 1-20	ARTKQTAR(Me)K(Me3)STGGKAPRK QL-K(Biot)-NH2	H3R8me1 + K9me3
P179	H3 1-20	ARTKQTAR(Me)K(Me2)STGGKAPRK QL-K(Biot)-NH2	H3R8me1 + K9me2
P180	H3 1-20	ARTKQTAR(Me2a)K(Me3)STGGKAPR KQL-K(Biot)-NH2	H3R8me2a + K9me3
P181	H3 1-20	ARTKQTAR(Me2a)K(Me2)STGGKAPR KQL-K(Biot)-NH2	H3R8me2a + K9me2
P182	H3 1-20	ARTKQTAR(Me2a)K(Me)STGGKAPRK QL-K(Biot)-NH2	H3R8me2a + K9me1
P183	H3 1-20	ARTKQTAR(Me2s)K(Me3)STGGKAPRK QL-K(Biot)-NH2	H3R8me2s + K9me3
P184	H3 1-20	ARTKQTAR(Me2s)K(Me2)STGGKAPRK QL-K(Biot)-NH2	H3R8me2s + K9me2

P185	H3 1-20	ARTKQTAR(Me2s)K(Me)STGGKAPRK QL-K(Biot)-NH2	H3R8me2s + K9me1
P186	H3 1-20	ARTK(Ac)QTARK(Me2)STGGK(Ac)APR K(Ac)QL-K(Biot)-NH2	H3K4ac + K9me2 + K14ac + K18ac
P187	H3 1-20	ARTK(Ac)QTARK(Me)STGGK(Ac)APR K(Ac)QL-K(Biot)-NH2	H3K4ac + K9me1 + K14ac + K18ac
P195	H3 15-34	Ac-APRKQLATKAARK(Me3)SAPSTG G-Peg-Biot	H3K27me3
P196	H3 15-34	Ac-APRKQLATKAARK(Me2)SAPSTG G-Peg-Biot	H3K27me2
P197	H3 15-34	Ac-APRKQLATKAARK(Me)SAPSTG G–Peg-Biot	H3K27me1
P198	H3 15-34	Ac-APRKQLATKAAR(Me2a)K(Me3)S APSTGG–Peg-Biot	H3R26me2a + K27me3
P200	H3 15-34	Ac-APRKQLATKAARR(Me2a)K(Me)S APSTGG-Peg-Biot	H3R26me2a + K27me1
P220	H3 1-20	ARTKQpTARK(Me3)STGGKAPRKQL-K(Biot)-NH2	H3T6p + K9me3
P221	H3 1-20	ARTKQpTAR(Me2a)K(Me3)STGGKA PRKQL-K(Biot)-NH2	H3T6p + R8me2a + K9me3
P224	H3 15-34	Ac-APRKQLATKAAR(Me2a)KSAPST GG-Peg-Biot	H3R26me2a
P225	H3 15-34	Ac-APRKQLATKAARK(Me3)pSAPST GG-Peg-Biot	H3K27me3 + S28p
P226	H3 15-34	Ac-APRKQLATKAARK(Me2)pSAPST GG-Peg-Biot	H3S27me2 + S28p
P229	H3 1-20	ARTK(Ac)QTARK(Me3)STGGKAPRK QL-K(Biot)-NH2	H3K4ac + K9me3
P237	H3 1-32	ARTKQTARK(Me2)STGGKAPRKQLA TKAARKSAPAT-Peg-Biot	H3K9me2 (1-32)
P241	H3 15-34	Ac-APRKQLATKAARK(Ac)SAPSTGG-Peg-Biot	H3K27ac
P242	H3 15-34	Ac-APRKQLATKAARK(Ac)pSAPSTGG-Peg-K(Biot)-NH2	H3K27ac + S28p
P243	H3 15-34	Ac-APRKQLATKAARKpSAPSTGG-Peg-K(Biot)-NH2	H3S28p
P253	H3 52-61	Ac-RRYQKSTELL-Peg-Biot	H3 (52-61)

P254	H3 52-61	Ac-RRYQK(Ac)STELL-Peg-Biot	H3K56ac (52-61)
P255	H3 52-61	Ac-RRYQK(Me3)STELL-Peg-Biot	H3K56me3 (52-61)
P259	H3 1-15	ARTK(Me2)QTARK(Me2)STGGKA-Peg-Biot	H3K4me2 + K9me2
P260	H3 1-15	ARTK(Me)QTARK(Me2)STGGKA-Peg-Biot	H3K4me1 + K9me2
P264	H3 1-15	ARTK(Me3)QTARK(Me2)STGGKA-Peg-Biot	H3K4me3 + K9me2
P300	H2A 1-17	Ac-SGRGKQGGKARAKAKTR-Peg-Biot	H2A (1-17)
P301	H2A 1-17	Ac-SGRGK(Ac)QGGK(Ac)ARAK(Ac)AK(Ac)TR-Peg-Biot	H2AK5ac + K9ac + K13ac + K15ac
P302	H2A 1-17	Ac-SGRGK(Ac)QGGKARAKAKTR-Peg-Biot	H2AK5ac
P303	H2A 1-17	Ac-pSGRGK(Ac)QGGKARAKAKTR-Peg-Biot	H2AS1p + K5ac
P304	H2A 1-17	Ac-SGR(Me2a)GK(Ac)QGGKARAKAKTR-Peg-Biot	H2AR3me2a + K5ac
P305	H2A 1-17	Ac-pSGR(Me2a)GK(Ac)QGGKARAKAKTR-Peg-Biot	H2AS1p + R3me2a + K5ac
P306	H2A 1-17	Ac-SGCitGK(Ac)QGGKARAKAKTR-Peg-Biot	H2ACit3 + K5ac
P307	H2A 1-17	Ac-pSGCitGK(Ac)QGGKARAKAKTR-Peg-Biot	H2AS1p + Cit3 + K5ac
P308	H2A 1-17	Ac-pSGRGK(Ac)QGGK(Ac)ARAK(Ac)AK(Ac)TR-Peg-Biot	H2AS1p + K5ac + K9ac + K13ac + K15ac
P309	H2A 1-17	SGRGK(Ac)QGGK(Ac)ARAK(Ac)AK(Ac)TR-Peg-Biot	H2AK5ac + K9ac + K13ac + K15ac (no N-ac)
P310	H2A 1-17	pSGRGK(Ac)QGGK(Ac)ARAK(Ac)AK(Ac)TR-Peg-Biot	H2AS1p + K5ac + K9ac + K13ac + K15ac (no N-ac)
P311	H2A.X	Biot-Peg-G132KKATQAS139QEY142-OH	H2AX (132-142)
P312	H2A.X	Biot-Peg-G132KKATQApS139QEY142-OH	H2AX (S139p)
P350	H4 1-23	Ac-SGR(Me2a)GK(Ac)GGKGLGKGGAKRHRKVLR-K(Biot)-NH2	H4R3me2a + K5ac
P351	H4 1-23	SGRGKGGKGLGKGGAKRHRKVLR-Peg-Biot	H4 (1-23) (no N-ac)

P352	H4 1-23	Ac-SGRGKGGKGLGKGGAKRHRK(Ac)VLRD-Peg-Biot	H4K20ac
P353	H4 1-23	Ac-pSGRGK(Ac)GGK(Ac)GLGK(Ac)GGAK(Ac)RHRKVLR-Peg-Biot	H4S1p + K5ac + K8ac + K12ac + K16ac
P359	H4 1-23	Ac-SGRGK(Ac)GGK(Ac)GLGKGGAKRHRKVLR-Peg-Biot	H4K5ac + K8ac
P360	H4 1-23	Ac-SGRGK(Ac)GGKGLGKGGAK(Ac)RHRKVLR-Peg-Biot	H4K5ac + K16ac
P362	H4 1-23	Ac-SGRGKGGK(Ac)GLGK(Ac)GGAKRHRKVLR-Peg-Biot	H4K8ac + K12ac
P363	H4 1-23	Ac-SGRGKGGK(Ac)GLGKGGAKRHRK(Ac)VLR-Peg-Biot	H4K12ac + K16ac
P366	H4 1-23	Ac-SGRGKGGKGLGKGGAK(Ac)RHRK(Ac)VLR-Peg-Biot	H4K16ac + K20ac
P370	H4 1-23	Ac-SGRGQGGQGLGK(Ac)GGAQRHRQVLR-Peg-Biot	H4K12ac + KQ5,8,16,20
P371	H4 1-23	Ac-SGRGK(Me)GGKGLGKGGAKRHRKVLR-Peg-Biot	H4K5me1
P372	H4 1-23	Ac-SGRGKGGK(Me)GLGKGGAKRHRKVLR-Peg-Biot	H4K8me1
P373	H4 1-23	Ac-SGRGKGGKGLGK(Me)GGAKRHRKVLR-Peg-Biot	H4K12me1
P374	H4 1-23	Ac-SGRGK(Ac)GGK(Me)GLGK(Ac)GGAKRHRKVLR-Peg-Biot	H4K5ac + K8me1 + K12ac
P375	H4 1-23	Ac-SGRGK(Me)GGK(Ac)GLGK(Me)GGAKRHRKVLR-Peg-Biot	H4K5me1 + K8ac + K12me1
P376	H4 1-23	Ac-SGRGK(Me)GGK(Me)GLGK(Me)GGAKRHRKVLR-Peg-Biot	H4K5me1 + K8me1 + K12me1
P381	H4 1-23	Ac-SGRGK(Me)GGK(Ac)GLGK(Ac)GGAK(Ac)RHRKVLR-Peg-Biot	H4K5me1 + K8ac + K12ac + K16ac

P382	H4 1-23	Ac-SGRGK(Ac)GGK(Me)GLGK(Ac)GGA K(Ac)RHRKVLR-Peg-Biot	H4K5ac + K8me1 + K12ac + K16ac
P383	H4 1-23	Ac-SGRGK(Ac)GGK(Ac)GLGK(Me)GGA K(Ac)RHRKVLR-Peg-Biot	H4K5ac + K8ac + K12me1 + K16ac
P400	H2B 1-24	PEPAKSAPAPKKGSKKAVTKAQKK-Peg-Biot	H2B (1-24)
P401	H2B 1-24	PEPAK(Me3)SAPAPKKGSKKAVTKAQ KK-Peg-Biot	H2BK5me3
P402	H2B 1-24	PEPAK(Me2)SAPAPKKGSKKAVTKAQ KK-Peg-Biot	H2BK5me2
P403	H2B 1-24	PEPAK(Me)SAPAPKKGSKKAVTKAQ KK-Peg-Biot	H2BK5me1
P625	H2A.X	Ac-SGRGKTGGKARAKAKSR-Peg-Biotin	H2A.X (1-17)
P626	H2A.X	Ac-SGRGK(Ac)TGGKARAKAKSR-Peg-Biotin	H2A.X K5ac
P789	H3 27-46	KSAPSTGGVK(Me3)KPHRYRPGTV-K(biotin)-NH2	H3K36me3
P790	H3 27-46	KSAPpSTGGVK(Me3)KPHRYRPGTV-K(biotin)-NH2	H3S31p + K36me3
tags	Flag-Tag	Biot-Peg-DYKDDDDK-NH2	Flag-Tag
	HA-Tag	Biot-Peg-YPYDVPDYASL-NH2	HA-Tag
	His-Tag	Biot-Peg-HHHHHH-NH2	His-Tag
	Myc-Tag	Biot-Peg-EQKLISEEDL-NH2	Myc-Tag
	V5-Tag	Biot-Peg-GKPIPNPLLGLDST-NH2	V5-Tag

References

Aasland R., Gibson T. J., and Stewart A. F., Feb 1995. The PHD finger: implications for chromatin-mediated transcriptional regulation. *Trends Biochem Sci*, 20 (2): 56–59.

Adelman Karen, Marr Michael T., Werner Janis, Saunders Abbie, Ni Zhuoyu, Andrulis Erik D., and Lis John T., Jan 2005. Efficient release from promoter-proximal stall sites requires transcript cleavage factor TFIIS. *Mol Cell*, 17 (1): 103–112.

Archambault J., Lacroute F., Ruet A., and Friesen J. D., Sep 1992. Genetic interaction between transcription elongation factor TFIIS and RNA polymerase II. *Mol Cell Biol*, 12 (9): 4142–4152.

Ariyoshi Mariko and Schwabe John W R., Aug 2003. A conserved structural motif reveals the essential transcriptional repression function of Spen proteins and their role in developmental signaling. *Genes Dev*, 17 (15): 1909–1920.

Armache Karim-Jean, Kettenberger Hubert, and Cramer Patrick, Jun 2003. Architecture of initiation-competent 12-subunit RNA polymerase II. *Proc Natl Acad Sci U S A*, 100 (12): 6964–6968.

Armache Karim-Jean, Mitterweger Simone, Meinhart Anton, and Cramer Patrick, Feb 2005. Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J Biol Chem*, 280 (8): 7131–7134.

Awrey D. E., Shimasaki N., Koth C., Weilbaecher R., Olmsted V., Kazanis S., Shan X., Arellano J., Arrowsmith C. H., Kane C. M., and Edwards A. M., Aug 1998. Yeast transcript elongation factor (TFIIS), structure and function. II: RNA polymerase binding, transcript cleavage, and read-through. *J Biol Chem*, 273 (35): 22595–22605.

Bermejo Rodrigo, Lai Mong Sing, and Foiani Marco, Mar 2012. Preventing replication stress to maintain genome stability: resolving conflicts between replication and transcription. *Mol Cell*, 45 (6): 710–718.

Bernstein Bradley E., Kamal Michael, Lindblad-Toh Kerstin, Bekiranov Stefan, Bailey Dione K., Huebert Dana J., McMahon Scott, Karlsson Elinor K., Kulbokas 3rd Edward J, Gingeras Thomas R., Schreiber Stuart L., and Lander Eric S., Jan 2005. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*, 120 (2): 169–181.

Biegert Andreas, Mayer Christian, Remmert Michael, Söding Johannes, and Lupas Andrei N., Jul 2006. The MPI Bioinformatics Toolkit for protein sequence analysis. *Nucleic Acids Res*, 34 (Web Server issue): W335–W339.

Bond Charles Simon and Schüttelkopf Alexander Wolfgang, May 2009. ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta Crystallogr D Biol Crystallogr*, 65 (Pt 5): 510–512.

Bradford M. M., May 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*, 72: 248–254.

Bricogne G., Blanc E., Brandl M., Flensburg C., Keller P., Paciorek P., Roversi P., Sharff A., Smart O., Vonrhein C., Womack T., Matthews B., Ten Eyck L., and Tronrud D., 2012. BUSTER Version 2.11.5.

Bushnell David A. and Kornberg Roger D., Jun 2003. Complete, 12-subunit RNA polymerase II at 4.1-A resolution: implications for the initiation of transcription. *Proc Natl Acad Sci U S A*, 100 (12): 6969–6973.

Bushnell David A., Westover Kenneth D., Davis Ralph E., and Kornberg Roger D., Feb 2004. Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms. *Science*, 303 (5660): 983–988.

Chen D., Ma H., Hong H., Koh S. S., Huang S. M., Schurter B. T., Aswad D. W., and Stallcup M. R., Jun 1999. Regulation of transcription by a protein methyltransferase. *Science*, 284 (5423): 2174–2177.

Cheung Alan C M. and Cramer Patrick, Mar 2011. Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature*, 471 (7337): 249–253.

Clapier Cedric R. and Cairns Bradley R., 2009. The biology of chromatin remodeling complexes. *Annu Rev Biochem*, 78: 273–304.

Clark D. J. and Felsenfeld G., Oct 1992. A nucleosome core is transferred out of the path of a transcribing polymerase. *Cell*, 71 (1): 11–22.

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

Crouch R. J., Arudchandran A., and Cerritelli S. M., 2001. RNase H1 of Saccharomyces cerevisiae: methods and nomenclature. *Methods Enzymol*, 341: 395–413.

Davis M Wayne and Hammarlund Marc, 2006. Single-nucleotide polymorphism mapping. *Methods Mol Biol*, 351: 75–92.

Debeauchamp Jennifer L., Moses Arian, Noffsinger Victoria J P., Ulrich Dagny L., Job Godwin, Kosinski Aaron M., and Partridge Janet F., Apr 2008. Chp1-Tas3 interaction is required to recruit RITS to fission yeast centromeres and for maintenance of centromeric heterochromatin. *Mol Cell Biol*, 28 (7): 2154–2166.

Dehé Pierre-Marie and Géli Vincent, Aug 2006. The multiple faces of Set1. *Biochem Cell Biol*, 84 (4): 536–548.

Dengl Stefan, Mayer Andreas, Sun Mai, and Cramer Patrick, May 2009. Structure and in vivo requirement of the yeast Spt6 SH2 domain. *J Mol Biol*, 389 (1): 211–225.

Donovan S., Harwood J., Drury L. S., and Diffley J. F., May 1997. Cdc6p-dependent loading of Mcm proteins onto pre-replicative chromatin in budding yeast. *Proc Natl Acad Sci U S A*, 94 (11): 5611–5616.

Duina Andrea A., 2011. Histone Chaperones Spt6 and FACT: Similarities and Differences in Modes of Action at Transcribed Genes. *Genet Res Int*, 2011: 625210.

Dvir Arik, Sep 2002. Promoter escape by RNA polymerase II. *Biochim Biophys Acta*, 1577 (2): 208–223.

Edgar Robert C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*, 32 (5): 1792–1797.

Emsley P., Lohkamp B., Scott W. G., and Cowtan K., Apr 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*, 66 (Pt 4): 486–501.

Fan Xiaochun, Lamarre-Vincent Nathan, Wang Qian, and Struhl Kevin, Nov 2008. Extensive chromatin fragmentation improves enrichment of protein binding sites in chromatin immunoprecipitation experiments. *Nucleic Acids Res*, 36 (19): e125.

Fischer U., Struss A. K., Hemmer D., Michel A., Henn W., Steudel W. I., and Meese E., 2001. PHF3 expression is frequently reduced in glioma. *Cytogenet Cell Genet*, 94 (3-4): 131–136.

Fütterer Agnes, Campanero Miguel R., Leonardo Esther, Criado Luis M., Flores Juana M., Hernández Jesús M., San Miguel Jesús F., and Martínez-A Carlos, Sep 2005. Dido gene expression alterations are implicated in the induction of hematological myeloid neoplasms. *J Clin Invest*, 115 (9): 2351–2362.

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., and A. Bairoch, 2005. Protein Identification and Analysis Tools on the ExPASy Server, pages 571–607. Humana Press.

Gatchalian Jovylyn, Fütterer Agnes, Rothbart Scott B., Tong Qiong, Rincon-Arano Hector, Sánchez de Diego Ainhoa, Groudine Mark, Strahl Brian D., Martínez-A Carlos, van Wely Karel H M., and Kutateladze Tatiana G., Jul 2013. Dido3 PHD Modulates Cell Differentiation and Division. *Cell Rep*, 4 (1): 148–158.

Gentleman R., Carey V., Bates D., Bolstad B., Dettling M., Dudoit S., Ellis B., Gautier L., Ge Y., Gentry J., Hornik K., Hothorn T., Huber W., Iacus S., Irizarry R., Leisch F., Li C., Maechler M., Rossini A., Sawitzki G., Smith C., Symth G., Tierney L., Yang J., and Zhang J., 2004. Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biology*, 5: R80.

Giaever Guri, Chu Angela M., Ni Li, Connelly Carla, Riles Linda, Véronneau Steeve, Dow Sally, Lucau-Danila Ankuta, Anderson Keith, André Bruno, Arkin Adam P., Astromoff Anna, El-Bakkoury Mohamed, Bangham Rhonda, Benito Rocio, Brachat Sophie, Campanaro Stefano, Curtiss Matt, Davis Karen, Deutschbauer Adam, Entian Karl-Dieter, Flaherty Patrick, Foury Francoise, Garfinkel David J., Gerstein Mark, Gotte Deanna, Güldener Ulrich, Hegemann Johannes H., Hempel Svenja, Herman Zelek, Jaramillo Daniel F., Kelly Diane E., Kelly Steven L., Kötter Peter, LaBonte Darlene, Lamb David C., Lan Ning, Liang Hong, Liao Hong, Liu Lucy, Luo Chuanyun, Lussier Marc, Mao Rong, Menard Patrice, Ooi Siew Loon, Revuelta Jose L., Roberts Christopher J., Rose Matthias, Ross-Macdonald Petra, Scherens Bart, Schimmack Greg, Shafer Brenda, Shoemaker Daniel D., Sookhai-Mahadeo Sharon, Storms Reginald K., Strathern Jeffrey N., Valle Giorgio, Voet Marleen, Volckaert Guido, Wang Ching-yun, Ward Teresa R., Wilhelmy Julie, Winzeler Elizabeth A., Yang Yonghong, Yen Grace, Youngman Elaine, Yu Kexin, Bussey Howard, Boeke Jef D., Snyder Michael, Philippsen Peter, Davis Ronald W., and Johnston Mark, Jul 2002. Functional profiling of the Saccharomyces cerevisiae genome. *Nature*, 418 (6896): 387–391.

Gómez-Herreros Fernando, de Miguel-Jiménez Lola, Morillo-Huesca Macarena, Delgado-Ramos Lidia, Muñoz-Centeno María C., and Chávez Sebastián, Aug 2012. TFIIS is required for the balanced expression of the genes encoding ribosomal components under transcriptional stress. *Nucleic Acids Res*, 40 (14): 6508–6519.

Gnatt A. L., Cramer P., Fu J., Bushnell D. A., and Kornberg R. D., Jun 2001. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. *Science*, 292 (5523): 1876–1882.

Gottipati Ponnari and Helleday Thomas, May 2009. Transcription-associated recombination in eukaryotes: link between transcription, replication and recombination. *Mutagenesis*, 24 (3): 203–210.

Grubbs Frank E., 1969. Procedures for Detecting Outlying Observations in Samples. *Technometrics*, 11: 1–21.

Hande K. R., Sep 1998. Etoposide: four decades of development of a topoisomerase II inhibitor. *Eur J Cancer*, 34 (10): 1514–1521.

Hani J., Stumpf G., and Domdey H., May 1995. PTF1 encodes an essential protein in Saccharomyces cerevisiae, which shows strong homology with a new putative family of PPIases. *FEBS Lett*, 365 (2-3): 198–202.

He C. H., Masson J. Y., and Ramotar D., Dec 1996. A Saccharomyces cerevisiae phleomycin-sensitive mutant, ph140, is defective in the RAD6 DNA repair gene. *Can J Microbiol*, 42 (12): 1263–1266.

Heidemann Martin, Hintermair Corinna, Voß Kirsten, and Eick Dirk, Jan 2013. Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochim Biophys Acta*, 1829 (1): 55–62.

Hiriart Edwige, Gruffat Henri, Buisson Monique, Mikaelian Ivan, Keppler Selina, Meresse Patrick, Mercher Thomas, Bernard Olivier A., Sergeant Alain, and Manet Evelyne, Nov 2005. Interaction of the Epstein-Barr virus mRNA export factor EB2 with human Spen proteins SHARP, OTT1, and a novel member of the family, OTT3, links Spen proteins with splicing regulation and mRNA export. *J Biol Chem*, 280 (44): 36935–36945.

Hsin Jing-Ping and Manley James L., Oct 2012. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev*, 26 (19): 2119–2137.

Jaehning Judith A., 2010. The Paf1 complex: platform or player in RNA polymerase II transcription? *Biochim Biophys Acta*, 1799 (5-6): 379–388.

Jiang Cizhong and Pugh B Franklin, 2009. A compiled and systematic reference map of nucleosome positions across the Saccharomyces cerevisiae genome. *Genome Biol*, 10 (10): R109.

Johnsson B., Löfås S., and Lindquist G., Nov 1991. Immobilization of proteins to a carboxymethyldextran-modified gold surface for biospecific interaction analysis in surface plasmon resonance sensors. *Anal Biochem*, 198 (2): 268–277.

Jones D. T., Sep 1999. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol*, 292 (2): 195–202.

Keogh Michael-Christopher, Kim Jung-Ae, Downey Michael, Fillingham Jeffrey, Chowdhury Dipanjan, Harrison Jacob C., Onishi Megumi, Datta Nira, Galicia Sarah, Emili Andrew, Lieberman Judy, Shen Xuetong, Buratowski Stephen, Haber James E., Durocher Daniel, Greenblatt Jack F., and Krogan Nevan J., Jan 2006. A phosphatase complex that dephosphorylates gammaH2AX regulates DNA damage checkpoint recovery. *Nature*, 439 (7075): 497–501.

Kettenberger Hubert, Armache Karim-Jean, and Cramer Patrick, Aug 2003. Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell*, 114 (3): 347–357.

Kettenberger Hubert, Armache Karim-Jean, and Cramer Patrick, Dec 2004. Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol Cell*, 16 (6): 955–965.

Kim J. L., Nikolov D. B., and Burley S. K., Oct 1993a. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature*, 365 (6446): 520–527.

Kim Y., Geiger J. H., Hahn S., and Sigler P. B., Oct 1993b. Crystal structure of a yeast TBP/TATA-box complex. *Nature*, 365 (6446): 512–520.

Kireeva Maria L., Hancock Brynne, Cremona Gina H., Walter Wendy, Studitsky Vasily M., and Kashlev Mikhail, Apr 2005. Nature of the nucleosomal barrier to RNA polymerase II. *Mol Cell*, 18 (1): 97–108.

Kostrewa Dirk, Zeller Mirijam E., Armache Karim-Jean, Seizl Martin, Leike Kristin, Thomm Michael, and Cramer Patrick, Nov 2009. RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature*, 462 (7271): 323–330.

Krogan Nevan J., Dover Jim, Wood Adam, Schneider Jessica, Heidt Jonathan, Boateng Marry Ann, Dean Kimberly, Ryan Owen W., Golshani Ashkan, Johnston Mark, Greenblatt Jack F., and Shilatifard Ali, Mar 2003. The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. *Mol Cell*, 11 (3): 721–729.

Kulaeva Olga I., Gaykalova Daria A., Pestov Nikolai A., Golovastov Viktor V., Vassylyev Dmitry G., Artsimovitch Irina, and Studitsky Vasily M., Dec 2009. Mechanism of chromatin remodeling and recovery during passage of RNA polymerase II. *Nat Struct Mol Biol*, 16 (12): 1272–1278.

Kulaeva Olga I. and Studitsky Vasily M., 2010. Mechanism of histone survival during transcription by RNA polymerase II. *Transcription*, 1 (2): 85–88.

Kumar A., Cheung K. H., Ross-Macdonald P., Coelho P. S., Miller P., and Snyder M., Jan 2000. TRIPLES: a database of gene function in Saccharomyces cerevisiae. *Nucleic Acids Res*, 28 (1): 81–84.

Kurdistani Siavash K. and Grunstein Michael, Apr 2003. Histone acetylation and deacetylation in yeast. *Nat Rev Mol Cell Biol*, 4 (4): 276–284.

Lee T. I. and Young R. A., 2000. Transcription of eukaryotic protein-coding genes. *Annu Rev Genet*, 34: 77–137.

Löfås Stefan and Johnsson Bo, 1990. A Novel Hydrogel Matrix on Gold Surfaces in Surface Plasmon Resonance Sensors for Fast and Efficient Covalent Immobilization of Ligands. *Journal of The Chemical Society, Chemical Communications*, no. 21.

Liu Xin, Bushnell David A., Wang Dong, Calero Guillermo, and Kornberg Roger D., Jan 2010. Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism. *Science*, 327 (5962): 206–209.

Luger K., Mäder A. W., Richmond R. K., Sargent D. F., and Richmond T. J., Sep 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, 389 (6648): 251–260.

Ma Zhuo, Atencio David, Barnes Cassandra, DeFiglio Holland, and Hanes Steven D., Sep 2012. Multiple roles for the Ess1 prolyl isomerase in the RNA polymerase II transcription cycle. *Mol Cell Biol*, 32 (17): 3594–3607.

Martinez-Rucobo Fuensanta W. and Cramer Patrick, Jan 2013. Structural basis of transcription elongation. *Biochim Biophys Acta*, 1829 (1): 9–19.

Matsui T., Segall J., Weil P. A., and Roeder R. G., Dec 1980. Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. *J Biol Chem*, 255 (24): 11992–11996.

Matzke Marjori, Kanno Tatsuo, Daxinger Lucia, Huettel Bruno, and Matzke Antonius J M., Jun 2009. RNA-mediated chromatin-based silencing in plants. *Curr Opin Cell Biol*, 21 (3): 367–376.

Mayer Andreas, Lidschreiber Michael, Siebert Matthias, Leike Kristin, Söding Johannes, and Cramer Patrick, Oct 2010. Uniform transitions of the general RNA polymerase II transcription complex. *Nat Struct Mol Biol*, 17 (10): 1272–1278.

Miller T., Krogan N. J., Dover J., Erdjument-Bromage H., Tempst P., Johnston M., Greenblatt J. F., and Shilatifard A., Nov 2001. COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proc Natl Acad Sci U S A*, 98 (23): 12902–12907.

Morris D. P., Phatnani H. P., and Greenleaf A. L., Oct 1999. Phospho-carboxyl-terminal domain binding and the role of a prolyl isomerase in pre-mRNA 3'-End formation. *J Biol Chem*, 274 (44): 31583–31587.

Nagalakshmi Ugrappa, Wang Zhong, Waern Karl, Shou Chong, Raha Debasish, Gerstein Mark, and Snyder Michael, Jun 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*, 320 (5881): 1344–1349.

Nakanishi Shima, Sanderson Brian W., Delventhal Kym M., Bradford William D., Staehling-Hampton Karen, and Shilatifard Ali, Aug 2008. A comprehensive library of histone mutants identifies nucleosomal residues required for H3K4 methylation. *Nat Struct Mol Biol*, 15 (8): 881–888.

Ng Huck Hui, Robert François, Young Richard A., and Struhl Kevin, Mar 2003. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol Cell*, 11 (3): 709–719.

Nickoloff J. A., Spirio L. N., and Reynolds R. J., Oct 1998. A comparison of calcium phosphate coprecipitation and electroporation. Implications for studies on the genetic effects of DNA damage. *Mol Biotechnol*, 10 (2): 93–101.

O'Geen Henriette, Nicolet Charles M., Blahnik Kim, Green Roland, and Farnham Peggy J., Nov 2006. Comparison of sample preparation methods for ChIP-chip assays. *Biotechniques*, 41 (5): 577–580.

Oliver Samuel S. and Denu John M., Jan 2011. Dynamic interplay between histone H3 modifications and protein interpreters: emerging evidence for a "histone language". *Chembiochem*, 12 (2): 299–307.

Pal Sharmistha, Baiocchi Robert A., Byrd John C., Grever Michael R., Jacob Samson T., and Sif Saïd, Aug 2007. Low levels of miR-92b/96 induce PRMT5 translation and H3R8/H4R3 methylation in mantle cell lymphoma. *EMBO J*, 26 (15): 3558–3569.

Pal Sharmistha, Vishwanath Sheethal N., Erdjument-Bromage Hediye, Tempst Paul, and Sif Saïd, Nov 2004. Human SWI/SNF-associated PRMT5 methylates histone H3 arginine 8 and negatively regulates expression of ST7 and NM23 tumor suppressor genes. *Mol Cell Biol*, 24 (21): 9630–9645.

Pinskaya Marina, Gourvennec Stéphanie, and Morillon Antonin, Jun 2009. H3 lysine 4 di- and tri-methylation deposited by cryptic transcription attenuates promoter activation. *EMBO J*, 28 (12): 1697–1707.

Pray-Grant Marilyn G., Daniel Jeremy A., Schieltz David, Yates 3rd John R, and Grant Patrick A., Jan 2005. Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. *Nature*, 433 (7024): 434–438.

Punta Marco, Coggill Penny C., Eberhardt Ruth Y., Mistry Jaina, Tate John, Boursnell Chris, Pang Ningze, Forslund Kristoffer, Ceric Goran, Clements Jody, Heger Andreas, Holm Liisa, Sonnhammer Erik L L., Eddy Sean R., Bateman Alex, and Finn Robert D., Jan 2012. The Pfam protein families database. *Nucleic Acids Res*, 40 (Database issue): D290–D301.

R Development Core Team, 2008. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Rando Oliver J., Apr 2012. Combinatorial complexity in chromatin structure and function: revisiting the histone code. *Curr Opin Genet Dev*, 22 (2): 148–155.

Rando Oliver J. and Winston Fred, Feb 2012. Chromatin and transcription in yeast. *Genetics*, 190 (2): 351–387.

Riles Linda, Shaw Randal J., Johnston Mark, and Reines Daniel, Feb 2004. Large-scale screening of yeast mutants for sensitivity to the IMP dehydrogenase inhibitor 6-azauracil. *Yeast*, 21 (3): 241–248.

Roguev A., Schaft D., Shevchenko A., Pijnappel W. W., Wilm M., Aasland R., and Stewart A. F., Dec 2001. The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *EMBO J*, 20 (24): 7137–7148.

Rondón Ana G., Jimeno Sonia, García-Rubio María, and Aguilera Andrés, Oct 2003. Molecular evidence that the eukaryotic THO/TREX complex is required for efficient transcription elongation. *J Biol Chem*, 278 (40): 39037–39043.

Rothbart Scott B., Krajewski Krzysztof, Strahl Brian D., and Fuchs Stephen M., 2012. Peptide microarrays to interrogate the "histone code". *Methods Enzymol*, 512: 107–135.

Sainsbury Sarah, Niesser Jürgen, and Cramer Patrick, Jan 2013. Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature*, 493 (7432): 437–440.

Schulze Julia M., Hentrich Thomas, Nakanishi Shima, Gupta Arvind, Emberly Eldon, Shilatifard Ali, and Kobor Michael S., Nov 2011. Splitting the task: Ubp8 and Ubp10 deubiquitinate different cellular pools of H2BK123. *Genes Dev*, 25 (21): 2242–2247.

Söding Johannes, Biegert Andreas, and Lupas Andrei N., Jul 2005. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res*, 33 (Web Server issue): W244–W248.

Shi Xiaobing, Kachirskaia Ioulia, Walter Kay L., Kuo Jen-Hao A., Lake Aimee, Davrazou Foteini, Chan Steve M., Martin David G E., Fingerman Ian M., Briggs Scott D., Howe LeAnn, Utz Paul J., Kutateladze Tatiana G., Lugovskoy Alexey A., Bedford Mark T., and Gozani Or, Jan 2007. Proteome-wide analysis in Saccharomyces cerevisiae identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36. *J Biol Chem*, 282 (4): 2450–2455.

Shilatifard Ali, Jun 2008. Molecular implementation and physiological roles for histone H3 lysine 4 (H3K4) methylation. *Curr Opin Cell Biol*, 20 (3): 341–348.

Sigurdsson Stefan, Dirac-Svejstrup A Barbara, and Svejstrup Jesper Q., Apr 2010. Evidence that transcript cleavage is essential for RNA polymerase II transcription and cell viability. *Mol Cell*, 38 (2): 202–210.

Sikorski Timothy W. and Buratowski Stephen, Jun 2009. The basal initiation machinery: beyond the general transcription factors. *Curr Opin Cell Biol*, 21 (3): 344–351.

Sims Robert J, Millhouse Scott, Chen Chi-Fu, Lewis Brian A., Erdjument-Bromage Hediye, Tempst Paul, Manley James L., and Reinberg Danny, Nov 2007. Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. *Mol Cell*, 28 (4): 665–676.

Slater M. L., Jan 1973. Effect of reversible inhibition of deoxyribonucleic acid synthesis on the yeast cell cycle. *J Bacteriol*, 113 (1): 263–270.

Smolle Michaela and Workman Jerry L., Jan 2013. Transcription-associated histone modifications and cryptic transcription. *Biochim Biophys Acta*, 1829 (1): 84–97.

Sánchez R. and Sali A., 1997. Evaluation of comparative protein structure modeling by MODELLER-3. *Proteins*, Suppl 1: 50–58.

Snyder R. D., Oct 1984. Inhibitors of ribonucleotide reductase alter DNA repair in human fibroblasts through specific depletion of purine deoxynucleotide triphosphates. *Cell Biol Toxicol*, 1 (1): 81–94.

Stefansky W., 1972. Rejecting Outliers in Factorial Designs. *Technometrics*, 14: 469–479.

Studitsky V. M., Kassavetis G. A., Geiduschek E. P., and Felsenfeld G., Dec 1997. Mechanism of transcription through the nucleosome by eukaryotic RNA polymerase. *Science*, 278 (5345): 1960–1963.

Tercero J. A. and Diffley J. F., Aug 2001. Regulation of DNA replication fork progression through damaged DNA by the Mec1/Rad53 checkpoint. *Nature*, 412 (6846): 553–557.

Terzi Nihal, Churchman L Stirling, Vasiljeva Lidia, Weissman Jonathan, and Buratowski Stephen, Sep 2011. H3K4 trimethylation by Set1 promotes efficient termination by the Nrd1-Nab3-Sen1 pathway. *Mol Cell Biol*, 31 (17): 3569–3583.

Tong A. H., Evangelista M., Parsons A. B., Xu H., Bader G. D., Pagé N., Robinson M., Raghibizadeh S., Hogue C. W., Bussey H., Andrews B., Tyers M., and Boone C., Dec 2001. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science*, 294 (5550): 2364–2368.

Tong Amy Hin Yan, Lesage Guillaume, Bader Gary D., Ding Huiming, Xu Hong, Xin Xiaofeng, Young James, Berriz Gabriel F., Brost Renee L., Chang Michael, Chen YiQun, Cheng Xin, Chua Gordon, Friesen Helena, Goldberg Debra S., Haynes Jennifer, Humphries Christine, He Grace, Hussein Shamiza, Ke Lizhu, Krogan Nevan, Li Zhijian, Levinson Joshua N., Lu Hong, Ménard Patrice, Munyana Christella, Parsons Ainslie B., Ryan Owen, Tonikian Raffi, Roberts Tania, Sdicu Anne-Marie, Shapiro Jesse, Sheikh Bilal, Suter Bernhard, Wong Sharyl L., Zhang Lan V., Zhu Hongwei, Burd Christopher G., Munro Sean, Sander Chris, Rine Jasper, Greenblatt Jack, Peter Matthias, Bretscher Anthony, Bell Graham, Roth Frederick P., Brown Grant W., Andrews Brenda, Bussey Howard, and Boone Charles, Feb 2004. Global mapping of the yeast genetic interaction network. *Science*, 303 (5659): 808–813.

Tyrsted G., Oct 1982. Effect of hydroxyurea and 5-fluorodeoxyuridine on deoxyribonucleoside triphosphate pools early in phytohemagglutinin-stimulated human lymphocytes. *Biochem Pharmacol*, 31 (19): 3107–3113.

Vannini Alessandro and Cramer Patrick, Feb 2012. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol Cell*, 45 (4): 439–446.

Wang Dong, Bushnell David A., Huang Xuhui, Westover Kenneth D., Levitt Michael, and Kornberg Roger D., May 2009. Structural basis of transcription: backtracked RNA polymerase II at 3.4 angstrom resolution. *Science*, 324 (5931): 1203–1206.

Werner Finn, Sep 2007. Structure and function of archaeal RNA polymerases. *Mol Microbiol*, 65 (6): 1395–1404.

Westover Kenneth D., Bushnell David A., and Kornberg Roger D., Nov 2004a. Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell*, 119 (4): 481–489.

Westover Kenneth D., Bushnell David A., and Kornberg Roger D., Feb 2004b. Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science*, 303 (5660): 1014–1016.

Wilson Thomas E., Oct 2002. A genomics-based screen for yeast mutants with an altered recombination/end-joining repair ratio. *Genetics*, 162 (2): 677–688.

Wind M. and Reines D., Apr 2000. Transcription elongation factor SII. *Bioessays*, 22 (4): 327–336.

Winkler Duane D. and Luger Karolin, May 2011. The histone chaperone FACT: structural insights and mechanisms for nucleosome reorganization. *J Biol Chem*, 286 (21): 18369–18374.

Wood Adam, Schneider Jessica, Dover Jim, Johnston Mark, and Shilatifard Ali, Sep 2003. The Paf1 complex is essential for histone monoubiquitination by the Rad6-Bre1 complex, which signals for histone methylation by COMPASS and Dot1p. *J Biol Chem*, 278 (37): 34739–34742.

Woychik N. A., 1998. Fractions to functions: RNA polymerase II thirty years later. *Cold Spring Harb Symp Quant Biol*, 63: 311–317.

Wu X., Wilcox C. B., Devasahayam G., Hackett R. L., Arévalo-Rodríguez M., Cardenas M. E., Heitman J., and Hanes S. D., Jul 2000. The Ess1 prolyl isomerase is linked to chromatin remodeling complexes and the general transcription machinery. *EMBO J*, 19 (14): 3727–3738.

Wu Xiaoyun, Rossettini Anne, and Hanes Steven D., Dec 2003. The ESS1 prolyl isomerase and its suppressor BYE1 interact with RNA pol II to inhibit transcription elongation in Saccharomyces cerevisiae. *Genetics*, 165 (4): 1687–1702.

Yen Kuangyu, Vinayachandran Vinesh, Batta Kiran, Koerber R Thomas, and Pugh B Franklin, Jun 2012. Genome-wide nucleosome specificity and directionality of chromatin remodelers. *Cell*, 149 (7): 1461–1473.

Zacher Benedikt, Kuan Pei Fen, and Tresch Achim, 2010. Starr: Simple Tiling ARRay analysis of Affymetrix ChIP-chip data. *BMC Bioinformatics*, 11: 194.

List of abbreviations

Δ	genomic knockout
6-AU	6-azauracil
°C	degree celsius
Å	Ångstrom
aa	amino acids
AC	arrested complex
Amp	ampicillin
AMPCPP	α,β-Methyleneadenosine 5´-triphosphate
ATP	adenosine-5'-triphosphate
bp	base pair(s)
BSA	bovine serum albumine
ChIP	Chromatin immunoprecipitation
clonNAT	nourseothricin resistance cassette
CSM	complete supplement mixture
CTD	C-terminal domain
CTP	cytidine 5'-triphosphate
CV	column volumes
Da	Dalton
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	2'-deoxyribonucleoside triphosphate
DSB	double-strand break
dsDNA	double-stranded DNA
DTT	dithiothreitol
EC	elongation complex
ECL	enhanced chemiluminescence
E.coli	*Escherichia coli*
EDTA	ethylene diamine tetraacetate
et al.	and others [Latin: *er alii*]
EtOH	ethanol
FACS	fluorescence-activated cell sorting
FAM	6-carboxyfluorescein
FL	full-length
fwd	forward
g	gram or gravitational acceleration on earth's surface
GST	glutathione-S-transferase
h	hour(s)
H3	histone 3
HA	hemagglutinin
HCl	hydrochloric acid

HEPES	2-[4-(2-hydroxyethyl)-1-piperazinyl]-ethanesulfonic acid
His	Histidine
HOAc	acetic acid
Hs	*Homo sapiens*
H.sapiens	*Homo sapiens*
HU	hydroxyurea
Ile	Isoleucine
IPTG	isopropyl ß-D-1-thiogalactopyranoside
K4	lysine 4
KanMX	kanamycin resistance cassette
ko	knockout
l	liter
LB	lysogeny broth
Leu	Leucine
Lys	Lysine
M	molar (mol/liter)
min	minute(s)
MMS	Methyl methanesulfonate
MOPS	3-(N-morpholino)propanesulfonic acid
MPA	mycophenolic acid
mRNA	messenger RNA
MW	molecular weight
NCBI	National Center for Biotechnology Information
NFR	nucleosome-free region
NHEJ	non-homologous end-joining
Ni-NTA	nickel-nitrilotriacetic acid
nt	nucleotide
NTP	nucleoside triphosphate
OD_{600}	optical density at 600 nm
ON	over night
ORF	open reading frame
pA, polyA	polyadenylation
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	poly(ethylene glycol)
PHD	plant homeo domain
Phe	Phenylalanine
PI	protease inhibitor
PIPES	piperazine-N,N´-bis(2-ethanesulfonic acid)
PMSF	Phenylmethylsulfonylfluorid
Pol	DNA-dependent RNA polymerase
PTM	post-translational modification
qPCR	quantitaive PCR
rev	reverse

RNA	ribonucleic acid
RNase	ribonuclease
RPs	ribosomal proteins
Rpb	RNA polymerase B subunit
Rpb4/7	heterodimer Rpb4 and Rpb 7
rpm	revolutions per minute
rRNA	ribosomal RNA
RT	room temperature
s	second(s)
SC	synthetic complete
Sc, S. cerevisiae	*Saccharomyces cerevisiae*
SeMet	Selenomethionine
SDS	sodium dodecyl sulphate
SL	synthetic lethality
SLS	Swiss Light Source
SOB	super optimal broth
SPOC	Spen paralogue and orthologue C-terminal domain
SSB	single-strand break
ssDNA	single stranded DNA
T	temperature
TAE	Tris-acetate-EDTA
TAP	tandem-affinity purification
Taq	*Thermus aquaticus*
TBE	tris borate EDTA buffer
TBP	TATA-binding protein
TBS	Tris-buffered saline
TCEP	tris(2-carboxyethyl) phosphine hydrochloride
TE	tris-EDTA
TEMED	N,N,N',N'-tetramethylethylenediamine
TF	transcription factor
Thr	Threonine
TLD	TFIIS-like domain
Tris	tris-(hydroxymethyl)-aminomethane
tRNA	transfer RNA
TSS	transcription start site
U	unit
Ura	uracil
V	volt
Val	Valine
v/v	volume per volume
WT	wildtype
w/v	weight per volume
YNB	yeast nitrogen base
YP	yeast extract peptone
YPD	yeast extract peptone dextrose
YPGal	yeast extract peptone galactose

YPDS	yeast extract peptone dextrose sorbitol

Bases in nucleic acids:

A	adenine
C	cytosine
G	guanine
T	thymine
U	uracil