Thesis for doctoral degree (Ph.D.)
2009

# Characterization of RNA Polymerase II Subunit Rpb7 in Silencing and Transcription

Ingela Djupedal

Karolinska Institutet

Thesis for doctoral degree (Ph.D.) 2009

Characterization of RNA Polymerase II Subunit Rpb7 in Silencing and Transcription

Ingela Djupedal

Karolinska Institutet

From DEPARTMENT OF BIOSCIENCES AND NUTRITION
Karolinska Institutet, Stockholm, Sweden

# CHARACTERIZATION OF RNA POLYMERASE II SUBUNIT RPB7 IN SILENCING AND TRANSCRIPTION

Ingela Djupedal


Karolinska Institutet


Stockholm 2009

Published by Karolinska Institutet. Printed by Larserics Digital Print AB.

ISBN 978-91-7409-606-4

Till kärälskligaste Stefan,
Dan & Laura

## ABSTRACT

The DNA in eukaryotes is arranged in fibres of chromatin. The chromatin may be more or less compacted and the degree of condensation of the chromatin affects the accessibility of the DNA. The accessibility of the DNA, in turn, affects transcription and gene regulation. Genes within inaccessible DNA are commonly repressed whereas genes within accessible DNA are active and expressed. This thesis concerns the interplay between chromatin and transcription with focus on the function of the RNA polymerase II (pol II) subunit Rpb7.

We have demonstrated that processing of centromeric transcripts by the ribonuclease III family protein Dcr1 is required for heterochromatin formation at the centromeres of *Schizosaccharomyces pombe.* A point mutation in the pol II subunit Rpb7 caused a specific defect in centromeric heterochromatin formation. We have shown i) that the centromeric transcripts that accumulate in *dcr1Δ* cells are products of pol II, ii) the *rpbG150D* mutation is deficient in recognition and/or initiation of transcription from the centromeric promoter. Transcription by pol II within the centromeres was surprising since insertion of marker genes within these loci normally results in repression of pol II transcription. Here, paradoxically, pol II transcription was required for the construction of the inaccessible heterochromatin structure.

Our analysis of sRNA in *S. pombe* revealed that most centromeric siRNA originate from two clusters, which are repeated several times within the centromeres. This led us to propose a model in which centromeric transcripts fold into double stranded structures that are processed by Dcr1. The resulting siRNAs may contribute with the starting signal for the RNAi feedback loop required for heterochromatin formation at the centromeres.

Finally, we demonstrate that the genome-wide association of Rpb7 is nearly identical to that of the core pol II subunit Rpb2, indicating a general role for Rpb7 in transcription. We further show that the occupancy pattern of Rpb4, a pol II subunit that forms a subcomplex together with Rpb7, differs from those of Rpb2 and Rpb7. Rpb4 may therefore have a less general function in transcription than Rpb7.

Hence, transcription by pol II is required not only for gene expression but also for repression via formation of inaccessible heterochromatin.

## LIST OF PUBLICATIONS

I. Provost P, Silverstein RA, Dishart D, Walfridsson J, **Djupedal I**, Kniola B, Wright A, Samuelsson B, Radmark O, Ekwall K. (2002) Dicer is required for chromosome segregation and gene silencing in fission yeast cells. Proc Natl Acad Sci U S A. 2002 Dec 24;99 (26):16648-53.

II. **Djupedal I***, Portoso M*, Spåhr H, Bonilla C, Gustafsson CM, Allshire RC, Ekwall K. (2005) RNA Pol II subunit Rpb7 promotes centromeric transcription and RNAi-directed chromatin silencing. Genes Dev. Oct 1;19 (19):2301-6.

III. **Djupedal I***, Kos-Braun, CK*, Mosher RA*, Söderholm N, Simmer F, Hardcastle T J, Fender A, Heidrich N, Kagansky A, Bayne E, Wagner EGH, Baulcombe DC, Allshire RA, Ekwall K. Analysis of small RNA in Fission yeast; centromeric siRNAs are potentially generated via a structured RNA. Accepted by EMBO J Nov 5 2009.

IV. **Djupedal I**, Durand-Dubief M, Sinha I, Ekwall K. Differential Genome-wide Occupancies of RNA Polymerase II Subunits Rpb4 and Rpb7 in Fission Yeast. Submitted to JBC.

*These authors contributed equally to this work.

### Related Publications:

**Djupedal I**, Ekwall K. The paradox of silent heterochromatin. Science. 2008 May 2;320 (5876):624-5

**Djupedal I**, Ekwall K. Epigenetics: heterochromatin meets RNAi. Cell Res. 2009 Mar;19 (3):282-95. Review.

# TABLE OF CONTENTS

## LIST OF ABBREVIATIONS

| | |
|---|---|
| 3'-UTR | Untranslated region in mRNA between translation stop codon and poly-A tail |
| 5'-UTR | Untranslated region between TSS and the ATG codon |
| Ago1 | Argonaute (*S. pombe*) |
| ATG | Codon that directs start of protein translation |
| bp | basepairs |
| CTD | C-terminal domain of Rpb1 |
| Cnt | Central core of centromeres |
| ChIP | Chromatin ImmunoPrecipitation |
| Dcr1 | Dicer (*S. pombe*) |
| dsRNA | Double-stranded RNA |
| GTF | General Transcription Factor |
| HAT | Histone Acetyl Transferase |
| HDAC | Histone Deacetylase |
| HP1 | Heterochromatin Protein 1, recognizes and binds H3K9me. |
| HMTase | Histone Methyl Transferase |
| H3K9me | Methylated Lysine 9 of Histone H3 |
| IGR | On Eurogentech microarrays, the 500 bp directly upstream of the ATG codon. |
| Imr | Innermost repeats |
| kDa | kilo Dalton or u, approximately the weight of one Hydrogen atom or $1.66 \times 10^{-21}$ g |
| miRNA | microRNA |
| mRNA | messenger RNA, encodes proteins |
| NELF | Negative Elongation Factor |
| nt | nucleotide |
| OB-fold | Single-stranded nucleic acid binding domain |
| ORF | Open Reading Frame |
| Otr | Outer repeat region of centromeres. |
| Pol II | RNA polymerase II |
| PAZ-domain | PIWI-Argonaute-Zwille domain |
| P-bodies | Processing bodies, also known as GW-bodies, cytoplasmic foci with high concentration of enzymes involved in mRNA degradation |
| piRNA | PIWI-interacting RNA |
| P-TEFb | Positive Transcription Elongation Factor b |

| | |
|---|---|
| Rdp1 | RNA-directed RNA polymerase (*S. pombe*) |
| RISC | RNA-Induced Silencing Complex |
| RITS | RNA-induced Initiation of Silencing Complex |
| RDRC | RNA-directed RNA polymerase complex |
| RNAi | RNA interference |
| RNase III | Specifically binds and cleaves double-stranded RNA |
| Rpb4 | 4:th largest (in *S. cerevisiae*) subunit of RNA polymerase II |
| Rpb7 | 7:th largest subunit of RNA polymerase II |
| rRNA | ribosomal RNA, required for all protein synthesis. |
| RT-PCR | PCR on template produced by reverse transcription, *i. e.*, cDNA |
| S1 domain | Single-stranded nucleic acid binding domain, originally identified in ribosomal protein S1 |
| siRNA | Short interfering RNA |
| snRNA | Small nuclear RNA |
| snoRNA | Small nucleolar RNA |
| sRNA | Small RNA, 20-30 nts |
| Swi6 | HP1 homologue in *S. pombe* |
| TAF | TBP-Associated Factors |
| TBP | TATA-binding protein |
| TBZ | Thiabendazole, interferes with microtubuli formation and is used to assay chromatid segregation defects |
| TSA | Trichostatin A, inhibits class I and II HDACs |
| tRNA | transferRNA, carries amino acids matching their anticodons to ribosomes. |
| Q-PCR | Quantitative Real Time Polymerase Chain Reaction |

# 1 INTRODUCTION

"If any enzyme does the cell's heavy lifting, it's RNA polymerase II. Its job: getting the synthesis of all the proteins in higher cells under way by copying their genes into RNAs, and doing it at just the right time and in just the right amounts. As such, pol II, as the enzyme is called, is the heart of the machinery that controls everything that cells do-from differentiating into all the tissues of a developing embryo to responding to everyday stresses."

Jean Marx, Science, April 20, 2001:411-414.

Regulation of gene expression is a central theme in molecular biology. Transcription of genes by RNA polymerases is one of the main steps in the regulation of which genes are to be expressed and at which levels. Mis-regulation of genes is directly associated with various human diseases, such as cancers and developmental disorders. Therefore, understanding the mechanisms that regulate transcription is key to understanding the processes that govern development and diseases.

This work was initiated with an investigation of the de-repressive effects of a mutant in a subunit of RNA polymerase II (pol II) in *Schizosaccharomyces pombe*. It seemed like a paradox – a mutation in pol II, the enzyme that mediates transcription of all protein coding genes, as well as some other classes of transcripts, would plausibly lead to lower levels of RNA transcripts rather than elevated levels. Furthermore, this de-repression occurred at the centromeres, loci that were supposed to be transcriptionally "silent" due to the repressive conformation of heterochromatin.

In this thesis, I have characterized the phenotype of the *rpb7G150D* mutant in order to understand the function of pol II and Rpb7 in centromeric heterochromatin formation as well as the general functions of Rpb7. Starting with paper I, we investigated the role of the RNA interference (RNAi) component Dcr1 in *S. pombe* where it is involved in cleaving or "dicing" long, centromeric transcripts into siRNA that guide chromatin modifications to homologous loci. In paper II, we investigate the link between pol II, RNAi and heterochromatin formation at the centromere. Paper III is a comprehensive study of sRNAs from the centromeres whereas paper IV provides a comparison of genome-wide maps of protein occupancy of the pol II subunits Rbp2, Rpb4 and Rpb7.

First, I will provide a short introduction to transcription in general as well as the functions of Rpb7, chromatin, epigenetics and RNAi.

## 1.1 TRANSCRIPTION

Transcription is the process in which nucleic acid, DNA or RNA, is used as a template for synthesis of a complementary nucleic acid. Here, the topic is transcription of DNA into RNA. There are five universal types of nucleotides (nt), adenine (A), guanine (G), cytosine (C) and thymine (T), which is specific to DNA and replaced by uracil (U) in RNA. A is complementary to T whereas G is complementary to C. The relative strength of the two types of base pairs (bp) is important; GC forms three hydrogen bonds compared to the two hydrogen bonds between A and T. This affects the chemical properties of the DNA or RNA helices in that GC rich stretches form strong, rigid complexes whereas AT (in RNA AU) rich sequences will disassemble more easily.

Transcription is mediated by RNA polymerases, of which there are three types shared by all eukaryotes: RNA polymerase I, II and III. The RNA polymerases have different substrates. Pol I transcribes rRNA genes, which are present in long arrays and whose products are constituents of the ribosomes, multisubunit complexes that use mRNA as template for protein synthesis. Pol III transcribes small structural RNAs, for example 5S rRNA and tRNAs. Pol II transcribes the most diverse gene set, the protein coding genes, some snoRNAs, snRNAs, miRNA genes and siRNA-precursors. Pol II activity corresponds to 20 % of the total transcription in the cell (Paule and White 2000). Plants have subunits for two additional RNA polymerases, IV and V, which are required for siRNA mediated gene silencing via DNA methylation and heterochromatin formation (Landick 2009).

Transcription is a process of ancient evolutionary origin and all multisubunit RNA polymerases have remarkably similar subunit composition among both eukarytoes and archaeons (Werner 2008). Therefore, understanding the mechanism of transcription in simple eukaryotes, such as yeasts, is useful for understanding transcription in more complex organisms, which are more difficult to study.

### 1.1.1 RNA Polymerase II Transcription

Yeast pol II is a 550 kDa complex, composed of 12 subunits, that transcribes DNA at an average rate of approximately 60 -70 nt/s *in vivo* (Darzacq, Shav-Tal et al. 2007; Singh and Padgett 2009). Pol II subunits are named Rpb1 to Rpb12. Five pol II

subunits, Rpb5, Rpb6, Rpb8, Rpb10 and Rpb12, are shared between all three RNA polymerases. In *S. pombe*, the number of pol II complexes was estimated by immunoblotting to 10.000 per cell (Kimura, Sakurai et al. 2001). The crystal structure of the complete pol II of *S. pombe* was recently solved (FIG. 1) (Spahr, Calero et al. 2009). Pol II is similar to a clamp in which 18 nt of DNA is firmly held and the DNA helix is bent in a 90° angle upon association with pol II. A pair of jaws is formed by Rpb5 together with parts of Rpb1 and Rpb9 (Cramer, Bushnell et al. 2000). Triphosphate ribonucleotides enter the enzyme via a funnel below the catalytic active site, which is composed by the two largest pol II specific subunits Rpb1 and Rpb2. Here, matching ribonucleotides are added to a RNA:DNA hybrid of 8 bp in length. The DNA:RNA hybrid is separated by a wedge and the nascent RNA exits through a positively charged path close to the Rpb4/Rpb7 dimer (Westover, Bushnell et al. 2004; Chen, Chang et al. 2009).

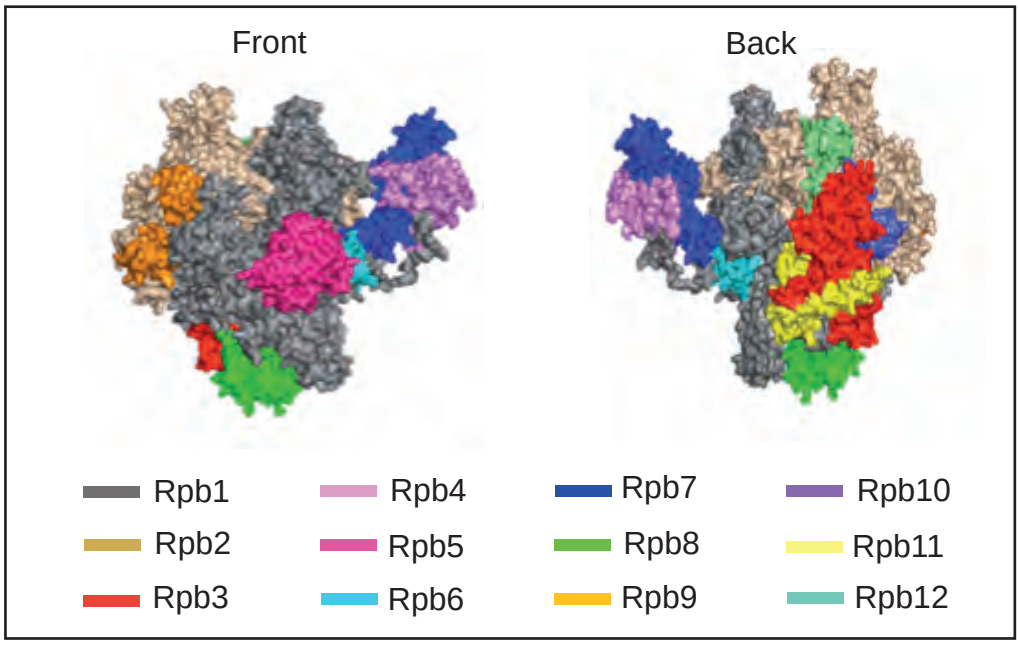


*Figure 1. Front and back view of pol II crystal structure of S. pombe, pol II subunits are colored as indicated. Modified and reprinted with permission from Spåhr et al, PNAS 2009 (Spahr, Calero et al. 2009).*

A feature specific to pol II, which is not included in the crystal structure, is the large C-terminal domain (CTD) of Rpb1 that is composed of heptad repeats of the amino acid sequence YSPTSPS (Tyrosine-Serine-Proline-Threonine-Serine-Proline-Serine). The number of repeats varies between organisms, from 26 in *Saccharomyces cerevisiae* to 56 in *Homo sapiens* (Meinhart, Kamenski et al. 2005). The CTD is an

important platform for assembly for factors that regulate transcription and RNA processing. The CTD is big enough to reach anywhere on the surface of pol II. The binding of specific factors depends on the phosphorylation status of the CTD of which there are four general states, hypophosphorylated, only Ser5-P, Ser-5P together with Ser2-P, and finally Ser2-P alone. In principle, all the serines and threonines residues in the CTD may be phosphorylated or glycosylated and phophorylation of Ser-7 has been reported (Meinhart, Kamenski et al. 2005). Finally, the conformation of the CTD is modified by prolyl isomerases.

#### *1.1.1.1 The Transcription Cycle*

Transcription by pol II has been described as a cycle with different steps, with pol II recruitment, initiation, elongation and termination (FIG. 2). First, the promoter needs to be accessible for pol II binding. Actually, pol II alone does not suffice for promoter recognition and initiation. A number of General Transcription Factors (GTFs) are required for pol II recruitment and initiation (Table 1). In addition, the mediator complex, composed of approximately 25 subunits, transfers environmental signals to activate or repress transcription by pol II. Gene specific activators and repressors further regulate transcription. Such factors may recognize and bind specific DNA sequences or, alternatively, to regions with a specific combination of chromatin modifications. Co-activators and co-repressors regulate transcription by modifications of histones. These activators or repressors may in turn be activated, stabilized or translocated to the nucleus upon addition of post-translational modifications such as phosphorylation. Alternatively, ubiquitylation may induce the degradation of co-activators or co-repressors.

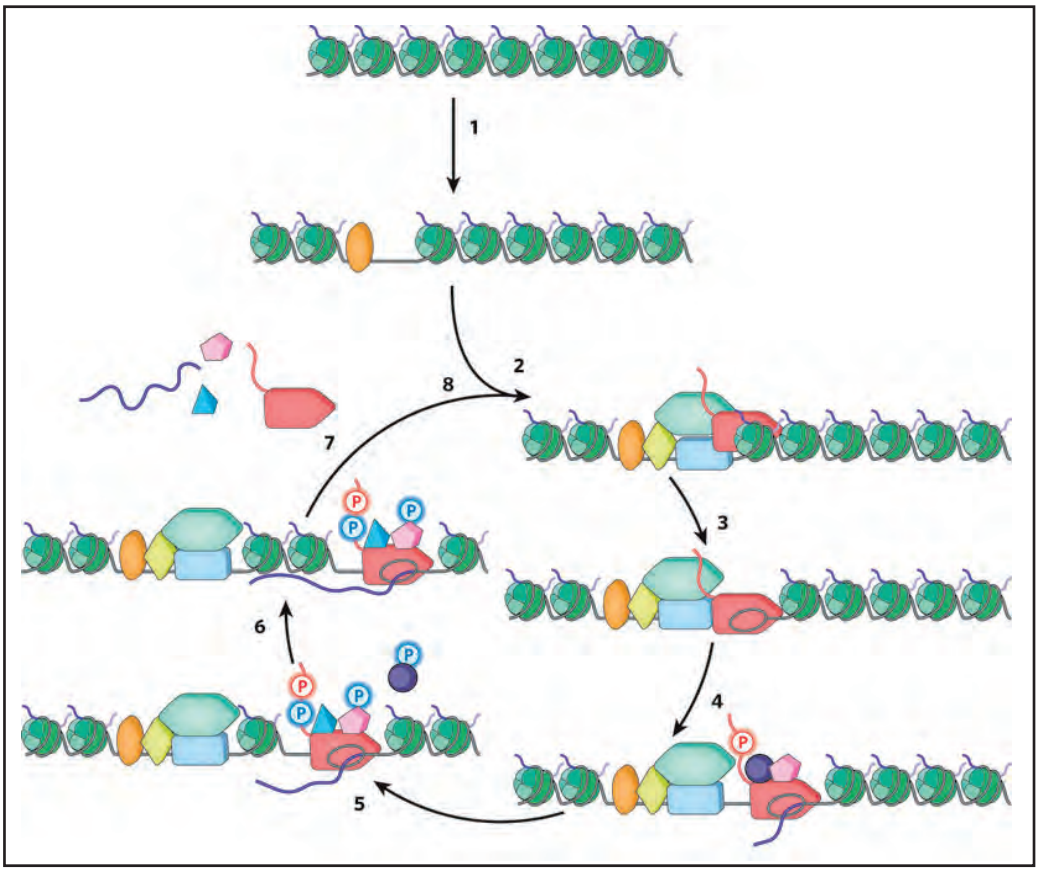


*Figure 2. The pol II transcription cycle. Step 1: Sliding or ejection of nucleosomes (green beads) makes the promoter accessible for pol II binding. Step 2: Formation of the Pre-Inititation Complex includes assembly of activators (orange oval), GTFs (blue rectangle), coactivators (green rhombus) and pol II (red rocket) on the promoter. Step 3: Opening of DNA (blue loop) is followed by initiation of transcription. Step 4: Ser5-phosphorylation of the CTD and promoter clearance, often followed by NELF-mediated (purple circle) pausing after transcription of 20-50 nucleotides. Step 5: Recruitment of the elongation factor P-TEFb that phosphorylates both Ser2 of the CTD and NELF, which then disassociates, resulting in pol II exit of the promoter proximal pause. Step 6: Productive elongation. Step 7: Termination. Step 8: Dephosphorylation of pol II is required prior to initiation of a new transcription cycle. Reprinted by permission from Macmillan Publishers Ltd: Nature, (Fuda et al., copyright 2009).*

Apart from GTFs, the mediator and the co-activators and co-repressors, additional factors that associate with the CTD are required for successful pol II transcription. Among the CTD associating factors are kinases, phosphatases, and factors that associate with the nascent RNA. Capping, splicing and polyadenylation of

mRNA are co-transcriptional processes. Furthermore, progression of pol II along genes creates torsion of the DNA, with positive supercoils ahead of pol II and negative supercoils behind pol II. These supercoils need to be resolved through the action of DNA topoisomerases in order for pol II to progress. Hence, the process of transcription by pol II is highly complex and requires numerous positive and negative regulators *in vivo*.

**Table 1. Factors Associated with Pol II Transcription.**

| Name | Subunits | Function |
|---|---|---|
| Mediator | Approximately 25 | Stimulates basal transcription and transduces positive and negative regulatory signals to pol II |
| NELF | 4 | Stabilizes paused pol II, inhibits TFIIS (Peterlin and Price 2006). |
| pTEF-b | 2 | Cyclin dependent kinase that promotes elongation by phosphorylating the CTD at Ser2 (Peterlin and Price 2006). |
| TFIIA | 3, | Stabilizes binding of TBP to the TATA-box (Hoiby, Zhou et al. 2007). |
| TFIIB | 1 | Binds the BRE and TBP and recruits pol II, assists in melting of DNA and TSS selection (Kostrewa, Zeller et al. 2009) |
| TFIID | TBP + approx 10 TAFs | Promoter recognition and binding. |
| TFIIE | 4 | Recruitment and regulation of TFIIH |
| TFIIF | 4 | Stabilizes the DNA-TBP-TFIIB complex |
| TFIIH | 9, including 2 DNA helicases and a kinase | Required before, during and immediately after initiation, unwinds the promoter DNA and phosphorylates the CTD on Ser5. |
| TFIIS | 1 | Induces mRNA cleavage by pol II to relieve transcriptional pauses or arrests. (Kettenberger, Armache et al. 2003). |

*Compiled from Woychik and Hampsey, (Woychik and Hampsey 2002), unless other reference is cited.*

Analysis of Pol II binding, genome-wide as well as on individual genes, has revealed two major steps where regulation of transcription occurs, i) at the level of the accessibility of the promoter, ii) at the level of promoter proximal pausing, *i. e.* recruitment of elongation factors such as P-TEFb (Fuda, Ardehali et al. 2009).

i) The state of the chromatin at the promoter affects whether transcription factors and pol II may bind to the promoter. If the promoter is situated within densely packed heterochromatin, the promoter is inaccessible and the gene is firmly shut off.

Mapping of the stability of the association of TBP to promoters revealed that genes with high expression levels have stable TBP-binding which allows continuous re-initiation by pol II, whereas TBP-binding to inducible genes was much less stable, indicating that prior to each round of transcription TBP had to be recruited to the promoters (van Werven, van Teeffelen et al. 2009).

ii) Pol II is paused at the beginning of genes for rapid induction of transcription at the correct environmental cue. Only 1 of 90 pol II – promoter interactions were estimated to proceed to elongation (Darzacq, Shav-Tal et al. 2007). Induction of transcription of these genes can increase transcription from one transcript every ten minutes to one every four seconds (Fuda, Ardehali et al. 2009) and high throughput sequencing has demonstrated that these genes are active although at low levels (Core, Waterfall et al. 2008). Hence, these genes are not firmly shut off but produce low levels of transcripts also without induction of transcription.

#### *1.1.1.2 Schematic of a Pol II Gene*

The genes transcribed by pol II are the most diverse of the three eukaryotic RNA polymerases and there may be no properties that are common to all pol II transcribed genes. Classical features of pol II transcripts, however, are the 5'-cap and the poly-A-tail. A crude schematic of a pol II gene is depicted below (FIG. 3). Transcription starts in the vicinity of promoters, DNA elements that recruit TFIID and the rest of the pol II apparatus and direct the transcription start site (TSS). The DNA sequence -250 bp upstream of the TSS is called the proximal promoter and contains binding sites for transcription factors (reviewed in (Juven-Gershon, Hsu et al. 2008; Farnham 2009)). The "classical" pol II promoter motif is the TATA-box, an AT rich sequence approximately 25 bp upstream of the TSS. In fact, only about 20% of pol II gene promoters have a TATA box. There are other characterized motifs that may be present in different combination in different promoters. The TATA-box and the BRE-elements, which are localized immediately upstream and/or downstream of the TATA-box, are the most evolutionarily conserved promoter elements, found in both *Archaea* and *Eukarya*. The TATA-box is bound by TBP and the BRE-elements are bound by TFIIB. The Bre-TFIIB interaction has been reported to guide the direction of transcription (Geiduschek and Ouhammouch 2005). However, promoters of highly expressed genes are associated with bi-directional transcription in both *S. pombe* and mammals, although only transcripts of the sense direction accumulate (Core, Waterfall et al. 2008; Dutrow, Nix et al. 2008; Seila, Calabrese et al. 2008).

Genes with focussed core promoters have one or several TSS in close vicinity whereas dispersed core promoters, which are less well characterized, have a number of TSS separated by up to 100 nt (Juven-Gershon, Hsu et al. 2008). Further upstream is the distal promoter with additional transcription factor binding sites. In addition, DNA elements such as enhancers, insulators or boundary elements, located several kb upstream of the gene may affect gene expression. Genome-wide mapping of transcription factor binding sites of several human transcription factors indicate that the majority of recognizable DNA elements may be located further than 2.5 kb upstream of the TSS (Farnham 2009). Furthermore, it is common for genes to have more than one promoter. In humans 50% of the genes have alternate promoters that may be more than 1 kb apart (Farnham 2009).

*Figure 3. A crude schematic of a Pol II gene. The coding region consists of one or more exons (rectangles). Together these form the Open Reading Frame (ORF), which encodes the protein product. The first exon typically starts with an ATG codon that directs start of translation. Genes may have one or several TSS (arrows) as well as transcription stop sites (crosses).*

Alternative TSS usage leads to variable length of the 5'-untranslated region (5'-UTR). Likewise, there may be several transcription termination signals leading to variable length of the 3'-untranslated regions (3'-UTR) as well. Recently, genome-wide mapping of transcription start and stop sites in *S. pombe* revealed 5'- and 3'-UTRs to be on average 160 bp and 220 bp long, respectively, ranging between 0 and 4000 bp (Dutrow, Nix et al. 2008; Lantermann Submitted manuscript). The lengths of 5'-and 3'-UTRs have been demonstrated to vary in different growth conditions (Wilhelm, Marguerat et al. 2008) and may affect transcript stability.

Only 43% of the genes in *S. pombe* have introns (Dutrow, Nix et al. 2008). Termination of transcription is mediated by 3'-end formation signals and downstream elements that mediate pausing of pol II (Birse, Lee et al. 1997). After termination and release, protein-coding transcripts are transported to the cytoplasm where they are translated. Transcripts that are recognized as aberrant are directly degraded by the

exosome, a complex of 3’ to 5’ exonucleases. The exosome has diverse targets and is involved in RNA surveillance, maturation and sequence dependent- and independent degradation, in the nucleus as well as in the cytoplasm *(Houseley, LaCava et al. 2006)*.

#### *1.1.1.3 Pol II Transcription Factories.*

Transcription by pol II is usually presumed to move along the DNA. However, there is abundant experimental evidence for the opposite, *i. e.* transcribed DNA passes an immobile pol II. It is well established that transcription by pol I is concentrated to the nucleolus. Pol II and pol III may similarly be confined to distinct foci, but the size of these, approximately 50 nm in diameter, makes them difficult to study (Martin and Pombo 2003). The transcription factory model suggests that several pol II enzymes reside in specific foci and that the genes move through these “factories”. One advantage of this model is that the great number of accessory factors required for pol II transcription and mRNA processing would be concentrated in the vicinity, resulting in greater efficiency of the process. Here, the localization of genes adjacent to or far from an active “factory” may affect gene expression. However, simultaneous transcription of adjacent genes from opposites strands of the DNA may be problematic.

#### *1.1.1.4 Non-Canonical Pol II Transcripts*

High throughput sequencing has revealed extensive transcription of both sense and antisense strands of the DNA in mammals as well as in fission yeast (Carninci, Kasukawa et al. 2005; Birney, Stamatoyannopoulos et al. 2007; Dutrow, Nix et al. 2008; Wilhelm, Marguerat et al. 2008). Whereas transcripts corresponding to exons were abundant among sequences derived from the sense strands, sequences derived from the antisense strand corresponded to promoter regions and termination sites (He, Vogelstein et al. 2008). In this study, the distribution of sequences was non-random and differed between cell types. Furthermore, sequences derived from regions upstream of promoter accumulate after knockdown of subunits of the exosome (Preker, Nielsen et al. 2008).

It is well established that non-coding RNA are involved in setting up the inactive X of female mammals and also in parent specific imprinting (Heard and Disteche 2006; Pauler, Koerner et al. 2007; Royo and Cavaille 2008). Likewise, some of the recently discovered classes of RNA may be involved in regulating gene expression. Antisense transcripts may be involved in regulation of the corresponding sense transcripts, as in *S. cerevisiae*, where accumulation of antisense transcripts

resulted in repression of the sense strand mediated by recruitment of chromatin modifying enzymes (Camblong, Iglesias et al. 2007). In humans, antisense RNA corresponding to promoters has been demonstrated to direct changes in chromatin conformation via siRNAs, resulting in either repression or activation of gene expression (Morris, Chan et al. 2004; Kim, Villeneuve et al. 2006; Li, Okino et al. 2006; Morris, Santoso et al. 2008).

In some cases, the process of transcription, rather than the transcript itself, may be important. Transcription upstream of the *fbp1* gene of *S. pombe* was shown to be necessary for gene activation (Hirota, Miyoshi et al. 2008). Transcription disrupts nucleosomal arrays and several chromatin-modifying enzymes bind to the CTD. Thus, pol II transcription may direct post-transcriptional modifications on histones as a mark of recent gene activity.

### 1.1.2 The Rpb7 and Rpb4 Subcomplex.

Pol II is composed of a ten-subunit core and a small, dissociable subcomplex Rpb4/Rpb7. The fact that the Rpb4/Rpb7 subcomplex was demonstrated to be dispensable for promoter-independent initiation and elongation (Edwards, Kane et al. 1991) and that *rpb4Δ* cells are viable in *S. cerevisiae* has led to speculation of the actual roles of these proteins in transcription. There are three possibilities, as proposed by Mordechai Choder, i) Rpb4/Rpb7 stabilize the pol II complex, ii) Rpb4/Rpb7 are only required during some of the steps of the transcription cycle, iii) Rpb4/Rpb7 are only required for transcription of subsets of genes (Choder 2004). There is some experimental support for all three possibilities, as outlined below. Obviously, these possibilities are not mutually exclusive. The possible stabilizing effects of Rpb4/Rpb7 could be particularly important at genes that are more difficult to transcribe, perhaps due to a special chromatin conformation or Rpb4/Rpb7-mediated recruitment of accessory factors, a function that may be especially important at one or more steps of the transcription cycle.

The amino acid sequences and structures of Rpb4 and Rpb7 are well conserved (FIG. 4). For example, the crystal structure of the Rpb4/Rpb7 subcomplex was first solved for the archaeon *Methanococcus jannaschii* (Todone, Brick et al. 2001) and information derived from this structure has been shown to be valid for eukayotes such as yeasts and human as well. A peculiarity in the field is the divergence of the *S. cerevisiae* Rpb4 sequence as compared to most of the Rpb4 orthologues. In *S.*

*cerevisiae*, Rpb4 has a long, non-conserved internal loop, which makes this protein almost twice as large (24 kDa) compared to, for example, *S. pombe* Rpb4 (14 kDa).

#### *1.1.2.1 Rpb4/Rpb7 are required for Stabilization of Pol II Initiation.*

The Rpb4/Rpb7 heterodimer binds to core pol II via the N-terminal tip of Rpb7 in a pocket composed of Rpb1, Rpb2 and Rpb6 (FIG. 5). Rpb4 may stabilize the interaction via contact between its N-terminus and the Rpb1 CTD-linker (Spahr, Calero et al. 2009).

In the archaeon *Pyrococcus furiosus*, the Rpb7 orthologue was shown to stimulate opening of the DNA near the TSS (Naji, Grunberg et al. 2007). This may be achieved by stabilization of a pol II conformation with a closed clamp, as the crystal structure of pol II with Rpb4/Rpb7 was changed towards this conformation. A pol II conformation with a closed clamp resembles that of elongating pol II, whereas the pol II crystal structure lacking Rpb4/Rpb7 took a conformation with an open clamp (Armache, Kettenberger et al. 2003; Bushnell and Kornberg 2003). The Rpb4/Rpb7 heterodimer is situated near TFIIB, a general transcription factor important for transcription activation, promoter recognition and start site selection (Reese 2003). Wild type RNA polymerase II have greater affinity for the TFIIB-TBP complex than *rpb4Δ* RNA polymerase II (Choder 2004). Hence, there is both structural and biological evidence supporting the hypothesis that Rpb4/Rpb7 stabilizes pol II during initiation.

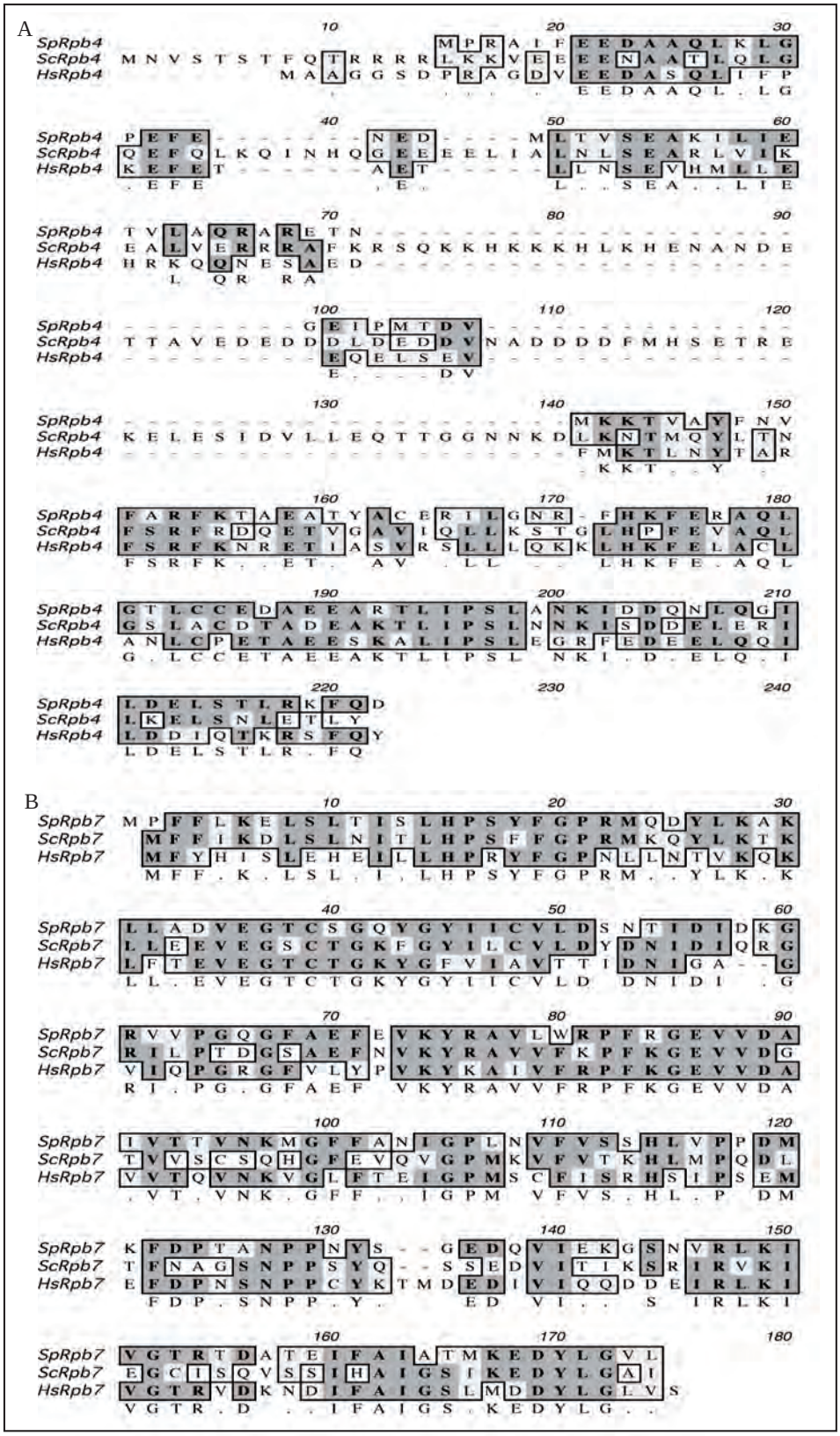


*Figure 4. Alignment of the amino acid sequences of A) Rpb4 and B) Rpb7 from S. pombe, S.cerevisiae and H. sapiens.*

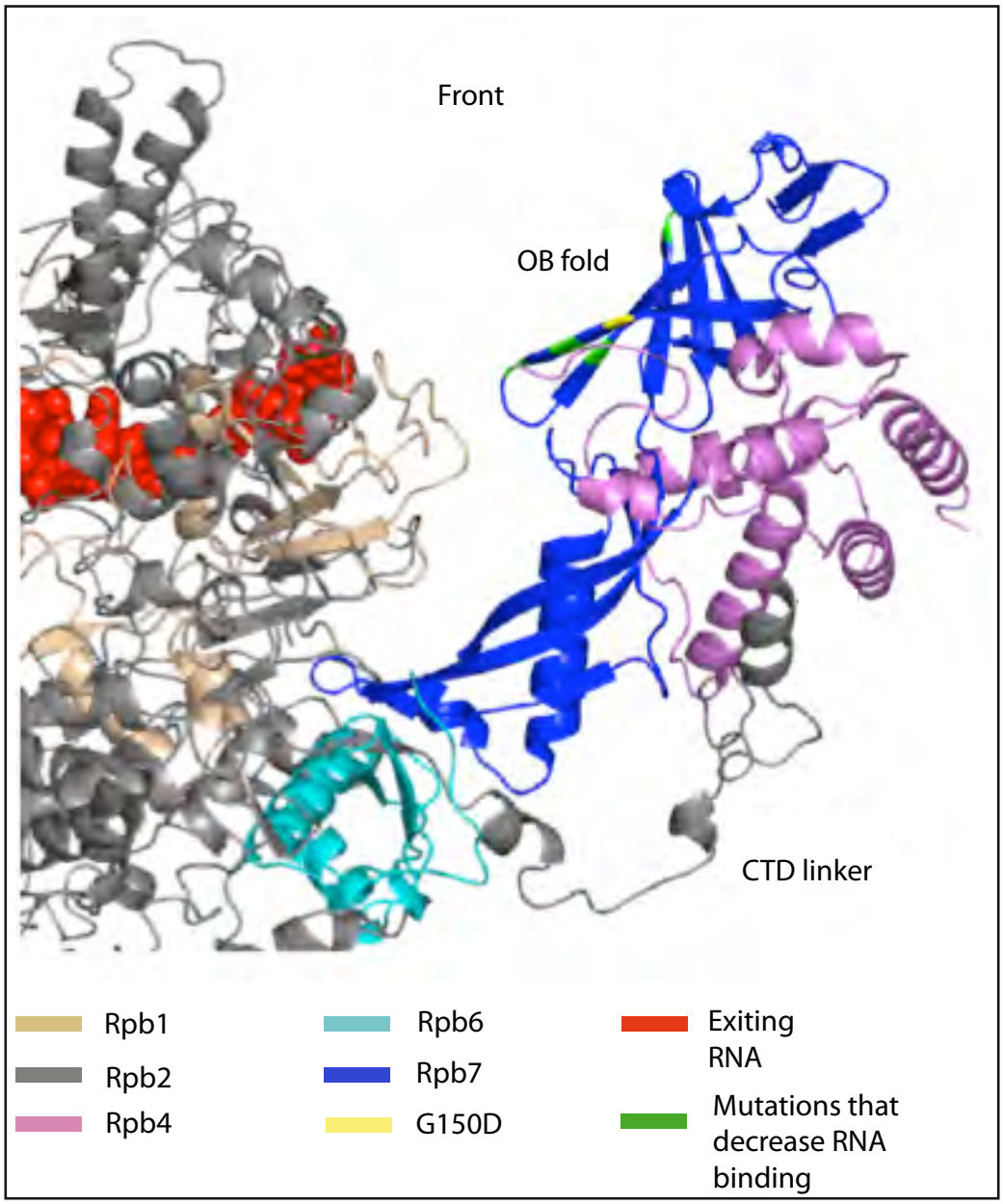


*Figure 5. The position of the Rpb4/Rpb7 subcomplex within pol II. Rpb7 binds to Rpb1, Rpb2 and Rpb6. Binding of Rpb4 to the CTD linker stabilizes the interaction. The rpb7G150D mutation is situated in the OB-fold, close to amino acids (green) that, when mutated, reduced RNA binding (Meka, Werner et al. 2005). Nascent RNA in red. Reprinted with permission from Såhr et al, PNAS 2009 (Spahr, Calero et al. 2009).*

### *1.1.2.2 Rpb4/Rpb7 are required for Certain Steps of the Transcription Cycle*

Rpb7 has two single stranded nucleic acid binding domains and RNA binding activity *in vitro* (Djupedal, Portoso et al. 2005; Meka, Werner et al. 2005), an OB fold and an S1 domain. The exit path of the nascent RNA is close to the Rpb4/Rpb7 subcomplex (Chen, Chang et al. 2009) and Rpb4/Rpb7 has been cross-linked to nascent RNA

(Ujvari and Luse 2006). Hence, Rpb4/Rpb7 is implied in binding the nascent RNA as soon as it exits the core polymerase and thereby stabilizes initiation by pol II (Meka, Werner et al. 2005).

In interacting with both nascent RNA and the CTD, it is easy to imagine that Rpb4/Rpb7 may transfer regulatory information from the surroundings to pol II. Indeed, among the proteins that have been reported to interact with Rpb4 and Rpb7 are the GTF TFIIF, the CTD phosphatase Fcp1, which is required for recycling pol II, and the *S. pombe* homologue of Nrd1, a protein involved in termination of transcription and 3'-end processing (Table 2).

Both the OB-fold and the S1 domain may bind ssDNA as well as ssRNA and Rpb7 has been cross-linked to promoter DNA and in elongating pol II complexes (Wooddell and Burgess 2000; Chen, Mandal et al. 2004). Thus, Rpb4/Rpb7 are implicated in initiation, elongation and recycling of pol II.

**Table 2. Compilation of Proteins Interacting with Rpb4/Rpb7**

| Protein | Interactant | Function | Reference |
|---|---|---|---|
| Actin | SpRpb7 | Actins are highly conserved proteins that are involved in cell motility. | (Mitsuzawa, Kimura et al. 2005) |
| Ess1 | ScRpb7 | Prolyl isomerase that binds the CTD and affects transcription elongation and termination. | (Wu, Wilcox et al. 2000) |
| EWS-Fli | HsRpb7 | Ewing´s sarcoma specific transcription factor. | (Petermann, Mossier et al. 1998) |
| Fcp1 | SpRpb4 | Processively dephosphorylates Ser-2 and Ser-5 of the CTD. | (Kimura, Suzuki et al. 2002) |
| GAPDH | SpRpb7 | Mammalian GAPDH is involved in membrane fusion, microtubule bundling, phosphotransferase activity, nuclear RNA export, prostate cancer progression, programmed neuronal cell death, DNA replication and DNA repair. | (Mitsuzawa, Kimura et al. 2005) |
| Med8 | SpRpb4 | Constituent of the head domain of Mediator. | (Mehta, Miklos et al. 2009) |
| Lsm1/Pat1 | ScRpb4 | Lsm1: involved in degradation of cytoplasmic mRNAs. Pat1: topoisomerase II-associated deadenylation-dependent mRNA-decapping factor. | (Lotan, Bar-On et al. 2005) |
| NOVH | HsRpb7 | Proto-oncogene. | (Perbal 1999) |
| Nsp1 | ScRpb4 | Essential component of the nuclear pore complex, which mediates nuclear import and export. | (Tan, Li et al. 2000) |
| RAR | HsRpb7 | Retinoic acid receptor, positive and | (Shen, Bubulya et |

| | | | |
|---|---|---|---|
| | | negative regulator of transcription. | al. 1999) |
| Rna14, Rna15 | ScRpb4 | 3'-end processing factors. | (Runner, Podolny et al. 2008) |
| Rho1 | ScRpb4 | GTP-binding protein of the rho subfamily of Ras-like proteins, involved in establishment of cell polarity. | (Bourbonnais, Faucher et al. 2001) |
| Seb1 | ScRpb7, SpRpb7 | Homologue of ScNrd1, implicated in 3'-end formation of snRNA and snoRNA, interacts with and stimulates the exosome. | (Mitsuzawa, Kanda et al. 2003) |
| Sro9 | ScRpb4 | Cytoplasmic RNA-binding protein that associates with translating ribosomes; also involved in the organization of actin filaments. | (Bourbonnais, Faucher et al. 2001) |
| Sro77 | ScRpb4 | Sro77 has roles in exocytosis and cation homeostasis; functions in docking and fusion of post-Golgi vesicles with plasma membrane. | (Bourbonnais, Faucher et al. 2001) |
| Tfg1 | ScRpb7 | Largest subunit of TFIIF. | (Chung, Craighead et al. 2003) |
| VHL | HsRpb7 | Tumor suppressor gene. | (Na, Duan et al. 2003) |

*1.1.2.3 The Rpb4/Rpb7 Subcomplex is required for Expression of a Subset of Genes.*

In *S. cerevisiae,* Rpb4 is clearly not required for general transcription under normal conditions, as *rpb4Δ* cells are viable. Estimates from genome wide expression profiling report that approximately 2% or up to 25% of the genome is affected by deletion of Rpb4 (Miyao, Barnett et al. 2001; Pillai, Verma et al. 2003). Under heat stress, however, the number of genes affected by *rpb4Δ* are similar to those of the temperature sensitive mutant *rpb1-1* in which transcription was considered to be totally inactivated (Holstege, Jennings et al. 1998). Hence, Rpb4 seems to be required for general transcription at elevated temperatures. Comparison of the genome-wide expression in *rpb4Δ* cells with wild type cells revealed a requirement for Rpb4 in the transcriptional adaptation to stress. Two recent publications, however, report the genome-wide protein occupancy profiles in *S. cerevisiae* of Rpb4 and Rpb7 to be highly similar to that of the core pol II subunit Rpb3 (Jasiak, Hartmann et al. 2008; Verma-Gaur, Rao et al. 2008). Also in human cells, Rpb7 and Rpb4 were ChIPped both to TSS as well as to downstream regions (Cojocaru, Jeronimo et al. 2008). Cojocaru and colleagues further report increased Rpb7 enrichment at genes after heat shock. In summary, although Rpb4 is not absolutely required for transcription, either *in vivo* or *in vitro*, both Rpb4

and Rpb7 seem to co-localize with pol II along the length of most genes and Rpb4/Rpb7 seem to share a role in transcription at heat stress.

Whereas the subunits of core pol II are mainly localized in the nucleus, Rpb4 and Rpb7 are distributed throughout the nucleus and the cytoplasm. Both proteins were elegantly demonstrated to shuttle from nucleus to cytoplasm in a transcription dependent as well as a transcription independent manner (Selitrennik, Duek et al. 2006; Goler-Baron, Selitrennik et al. 2008). Furthermore, Rpb4 and Rpb7 were demonstrated to have combined as well as independent roles in mRNA degradation. Rpb7 stimulated deadenylation and 3' to 5' degradation of mRNA (Lotan, Goler-Baron et al. 2007) whereas Rpb4 stimulated deadenylation of a specific class of mRNA, mRNAs encoding genes (Lotan, Bar-On et al. 2005). Rpb4/Rpb7 are localized to P-bodies, aggregates of proteins required for de-capping and degradation of mRNA. Here, mRNAs may also be sequestered to avoid translation. P-bodies are the active site of RNAi.

Accumulation of RNA depends on the rate and frequency of transcription as well as the stability and degradation of the transcript. Rpb4/Rpb7 are a unique target for regulatory factors as the subcomplex is involved in both these processes, from promoter recognition and initiation of transcription to RNA degradation. Indeed, Rpb7 has been reported to interact with several proteins involved in cancer, such as NOV, RAR and EWS-FLI (Petermann, Mossier et al. 1998; Perbal 1999; Shen, Bubulya et al. 1999), and could partly mediate these proteins activating or repressive effects on transcription and/or RNA degradation.

#### *1.1.2.4 The rpb7G150D Mutation*

The *rpb7G150D* mutation was found in a screen for mutations that suppressed silencing at the centromeres of *S. pombe* (Ekwall, Cranston et al. 1999). The *rpb7G150D* cells are temperature sensitive and die after exposure to elevated temperatures, presumably due to complete inactivation of pol II (Djupedal, Portoso et al. 2005). At permissive temperatures, the cells grow slowly, with a division time of >3 hours as compared to a wild type cell cycle of two and a half hours. Even at the permissive temperature, a subset of *rpb7G150D* cells display odd morphology such as elongated, banana-shaped or pear-shaped cells (FIG. 6). After prolonged exposure to heat, a handful of revertants usually appear. Since revertants are likely to have a shorter cell cycle than the slow growing *rpb7G150D* cells it is necessary to control the temperature sensitivity in cultures used for RNA preparations and other experiments.

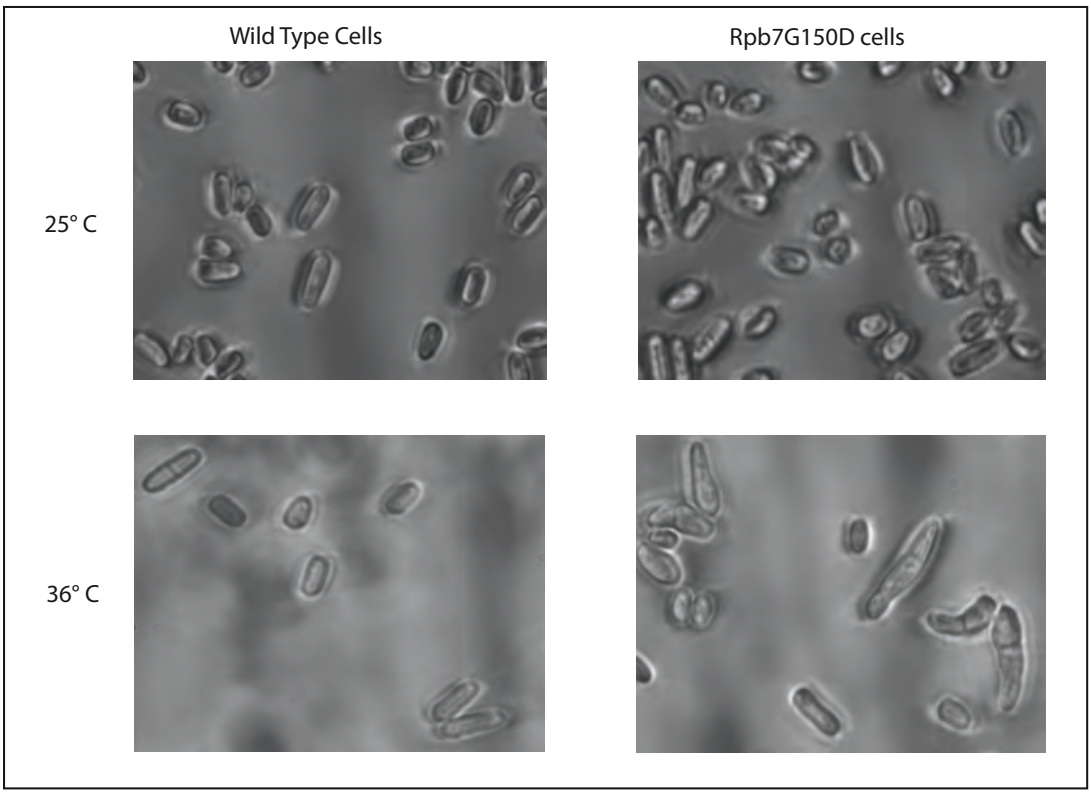


*Figure 6. Wild type and rpb7G150D cells after growth at the permissive temperature of 25° and at the non-permissive temperature of 36° C.*

The *rpb7G150D* mutation is a point mutation within the central β-sheet of the OB-fold of Rpb7, substituting a glycine for an aspartic acid (FIG. 5). We have shown that this mutation has a specific defect in the transcription of centromeric siRNA precursors and rather subtle effects on general transcription. The mutant polymerase has stronger affinity for binding ssRNA (Djupedal, Portoso et al. 2005). This is peculiar since similar mutations, substitutions of charged amino acids to alanine, in the vicinity were reported to cause drastic reductions in ssRNA binding (Meka, Werner et al. 2005). The stronger affinity for ssRNA of *rpb7G150D* may affect the bonds between nascent RNA and Rpb7. This could stabilize the initiating pol II and I speculate that it could have an effect on secondary structure formation, in a process analogous to attenuation in bacteria.

## 1.2 EPIGENETICS AND CHROMATIN

Epigenetics is the study of heritable traits not encoded in the DNA. The best-known mechanisms for epigenetic inheritance are mediated by DNA methylation and the level of chromatin compaction. DNA methylation is a covalent modification that does not alter the DNA sequence but that, usually, represses gene expression. When cells divide,

the DNA methylation is copied to the "new" DNA strand with the "old" DNA strand as template and thereby heritable silencing of gene expression is promoted. The level of chromatin compaction, often in combination with DNA methylation, influences gene expression in a heritable manner. DNA wrapped around histone octamers forms nucleosomes that together with accessory proteins and RNA constitute chromatin. Chromatin may be more or less compact (FIG. 7) and thereby affect gene expression by rendering the DNA accessible or inaccessible to the transcription machinery. Propagation of the different levels of chromatin compaction is long term and may be inherited from mother cell to daughter cell.

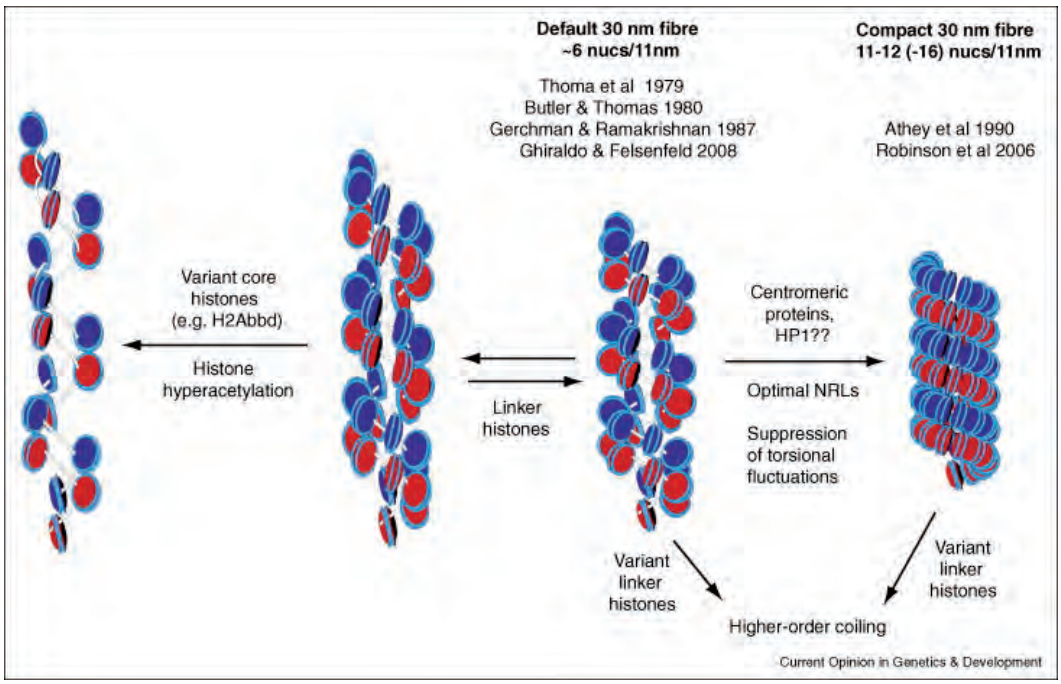


*Figure 7. The different levels of chromatin compaction. The most compact form, to the right, corresponds to heterochromatin. Reprinted from Curr Opin Genet Dev. 2009 Apr; 19(2) Bassett, Cooper, Wu and Travers, The folding and unfolding of eukaryotic chromatin, p 159-65, Copyright 2009, with permission from Elsevier.*

The high degrees of organization and compaction of DNA into chromatin is a prerequisite for the large sizes of the eukaryotic genomes. Chromatin is the natural substrate for all DNA mediated processes, such as transcription, DNA replication, DNA repair and recombination. Hence, chromatin needs to be flexible in order to accommodate these continuous processes. Transcription by pol II must be considered in the context of chromatin rather than naked DNA.

Unfortunately, there is no DNA methylation in *S. pombe*, and this important aspect of epigenetic inheritance will not be described further in this thesis. *S. pombe* is

however, an excellent model organism for understanding chromatin, in that many of the mechanisms that cause different chromatin states, such as histone variants, histone chaperones, histone tail modifications and chromatin remodelers, are conserved from fission yeast to man.

### 1.2.1 Nucleosomes

The classical definition of a nucleosome is 146 bp of DNA wrapped 1.65 laps around a histone octamer consisting of a tetramer of histones H2A and H2B and a tetramer of histones H3 and H4 (Luger and Hansen 2005) There is, however, abundant experimental evidence that nucleosomes are a dynamic family of particles that vary in their composition, stoichiometry and the extent of DNA binding (Zlatanova, Bishop et al. 2009). Apparently, the histone H2A-H2B dimer is easily displaced during transcription leading to, at least temporarily, the creation of nucleosome hexasomes or even hemisomes. Furthermore, binding of DNA to the nucleosomes seems to fluctuate leading to short periods in time when the nucleosome-folded DNA is accessible for other DNA-binding factors (Luger and Hansen 2005; Zlatanova, Bishop et al. 2009).

The DNA in between nucleosomes has been measured to 10-90 bp (Bassett, Cooper et al. 2009). Linker histones, such as histone H1 or H5, stabilize nucleosome arrays by binding and organizing the linker DNA (Luger and Hansen 2005). There are no linker histone homologues in *S. pombe* (Kobori, Yoshino et al. 2003). In addition, accessory proteins bind to nucleosomes and affect the higher order chromatin structure, for example, HP1 and Polycomb. In this way, DNA is organized into chromatin fibres that form higher order structures of which very little is known today.

#### *1.2.1.1 The Importance of Histone Tail Modifications*

The N-terminal tails of histones protrude from the nucleosomes and are subjected to post-transcriptional modifications (FIG. 8). Unmodified histone tails have a general positive charge that balances the negative charge of the DNA, allowing further compaction of the chromatin. The post-transcriptional modifications may alter the positive charge of the tails, as in hyperacetylation, and thereby affect chromatin compaction, or serve as binding sites for chromatin-associated proteins, as in HP1 binding to histone H3K9me. In addition, histone tail modifications may affect nucleosome-nucleosome interactions, as in the acetylation of the N-terminal tail of histone H4 which abrogates the interactions with histone H2A and H2B of neighbouring nucleosomes (Bassett, Cooper et al. 2009). In these ways, histone tail

modifications affect the higher order chromatin structure and make it more or less accessible to pol II.

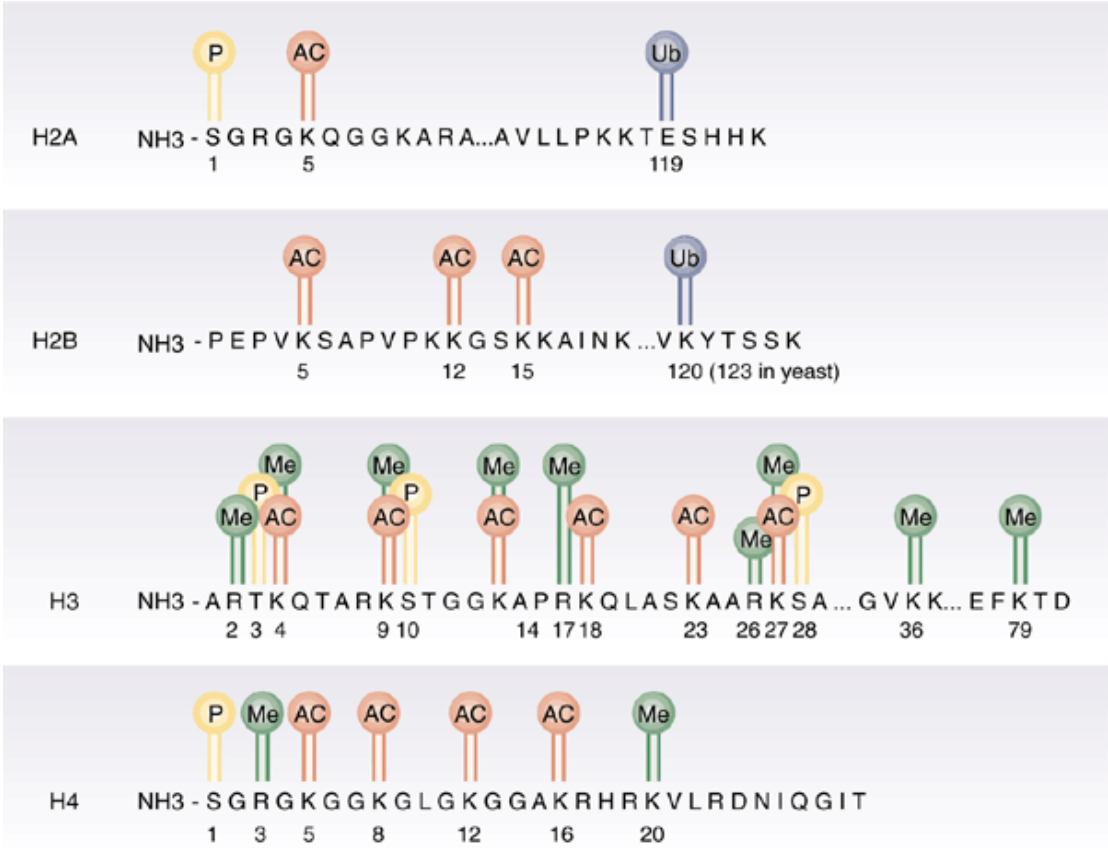


*Figure 8. N-terminal tails of histones, with possible modifications, numbering starting at the N-terminus of the tail. P equals phosphorylation, AC equals acetylation, Me equals Methylation and Ub Ubiquitination. Reprinted by permission from Macmillan Publishers Ltd: [Nature Cell Biology] (Jaskelioff and Peterson), copyright (2003).*

Histone tail modifications are generated by enzymes, such as histone acetyl transferases (HATs), histone deacetylases (HDACs), histone methyl transferases (HMTases) and histone demethylases. Many of these are multsubunit complexes and may be specific to certain histone tail residues. These enzymes may have other substrates in addition to histones. Most of the histone tail modifications are reversible. Methylation was traditionally considered as a more stable mark than acetylation.

Histone modifying enzymes may be recruited to chromatin by other proteins or by DNA motifs. For example, the HMTases Set1 and Set2 in *S. cerevisiae* bind to the pol II CTD that is phosphorylated on Ser2, leading to co-transcriptional generation of their respective histone tail modifications, H3K4me and H3K36me (Eissenberg and Shilatifard 2006). In this way, highly expressed genes carry H3K4me and H3K36me as

a mark of recent transcription activity. Another example is the transcription factor Atf1 that specifically binds to the M26 hotspot, a DNA motif of 6 bp, and mediates chromatin remodelling, which may result in gene activation or repression depending on the specific situation (Davidson, Shandilya et al. 2004; Hirota, Hasemi et al. 2004; Jia, Noma et al. 2004; Kim, Choi et al. 2004).

Histone tail modifications may be positively or negatively correlated to each other. For example, the acetyl group of histone H3K9ac needs to be removed prior to methylation of the same residue, H3K9me. The phosphorylation of a neighbouring residue, H3S10ph prevents HP1 binding of H3K9me. Furthermore, histone H3S10ph is bound by a protein that subsequently recruits a HAT which generates the H4K16ac (Zippo, Serafini et al. 2009). Together, H3S10ph and H4K16ac are recognized by the bromodomain protein Brd4, which recruits P-TEFb resulting in pol II elongation.

#### *1.2.1.2 Histone Variants*

In addition to the canonical histones there are histone variants, for example histones H3.3, H2A.Z, H2AX and CENP-A. The chemical properties of the histone variants may differ from those of the canonical histones, which may cause a different chromatin structure. Histone variants are often incorporated at specific loci. For example, the histone H3 variant CENP-A, or Cnp1 in *S. pombe*, is specifically targeted to centromeres where it is essential for the special centromeric chromatin structure (Silva and Jansen 2009). Another example is histone H2A.Z, which is targeted to promoters (Cairns 2009). The stability of nucleosomes containing histone H2A.Z depends on the type of histone H3 co-occupying the same nucleosome; the combination histone H2A.Z – H3 confers greater stability whereas the combination histone H2A.Z – H3.3 is less stable (Cairns 2009). The stability of the nucleosomes affects the accessibility of the DNA and thereby gene expression. It was recently shown that failure to incorporate histone H2A.Z in combination with deletions of either the HMTase Clr4 or RNAi-component Ago1 lead to an increase in antisense transcripts that would normally be degraded by the exosome (Zofall, Fischer et al. 2009). Hence, the type of histones within the nucleosomes may have effects on gene regulation.

#### *1.2.1.3 Nucleosome Positioning*

The position of nucleosomes on the DNA affects the accessibility of GTF binding sites so that DNA bound in the middle of a nucleosome is inaccessible whereas the DNA between nucleosomes is accessible for recognition and binding by transcription factors.

Therefore, the position and density of nucleosomes at the promoters affect gene expression. The nucleosome positioning and density are determined by a combination of the activities of chromatin remodelers, transcription factor binding, the rate and frequency of transcription and the DNA sequence (Zhang, Moqtaderi et al. 2009).

Active promoters are characterized by long, approximately 150 bp, nucleosome depleted regions in contrast to inactive promoters that are covered by nucleosomes (Cairns 2009). These promoters require the activities of chromatin remodelers in order to become accessible to the transcription machinery. Chromatin remodelers use the energy of ATP to move nucleosomes along the DNA or to disrupt nucleosomes (Radman-Livaja and Rando 2009). The chromatin remodeler SWI/SNF moves nucleosomes along the DNA or ejects nucleosomes, which often results in gene activation. The chromatin remodeler ISWI moves nucleosomes until a fixed length of linker DNA is achieved and often results in repression. In these ways, chromatin remodelers may serve as co-activators or co-repressors in pol II transcription.

### 1.2.2 Heterochromatin

The two extremes of chromatin density are euchromatin and heterochromatin. The less compact euchromatin is characterised by active genes whereas the DNA in heterochromatin is typically non-coding and found at centromeres and telomeres. Different histone tail modifications are associated with the different types of chromatin. Typical histone tail modifications in euchromatin are histone H3K4me$^3$, H3K36me and/or H3K79me in combination with hyperacetylation. In heterochromatin, hypoacetylation is combined with histone H3K9me$^2$ and H3K27me$^3$. However, H3K9me is also found together with H3K4 me at transcribed genes in euchromatin (Eissenberg and Shilatifard 2006). Thus, H3K9me is not an absolute mark of heterochromatin.

Heterochromatin has important structural features that are required for centromere and telomere function. Heterochromatin also represses homologous recombination and thereby suppresses transposition of selfish DNA elements. Insertions of reporter genes within heterochromatic regions are generally associated with silencing of gene expression. Therefore, the heterochromatic structure is regarded as repressive and non-permissive for transcription by pol II. However, there are examples of expressed genes that require heterochromatic surroundings for

transcription (Yasuhara and Wakimoto 2006). Hence, repression of gene expression by heterochromatin is not absolute.

#### *1.2.2.1 Characteristics of the Heterochromatin at the Centromeres of S. pombe.*

In *S. pombe*, centromeres require heterochromatin for correct segregation of sister chromatids at anaphase. The three centromeres of *S. pombe* are large, 35-110 kb, and modular (Wood, Gwilliam et al. 2002). There is a central core (*cnt*) flanked by the inner (*imr*) and outer repeats (*otr*) (see Paper III, Figure 2). The kinetochores connect the centromere to the spindle microtubuli at cell division and anchor at the *cnt* whereas the heterochromatin at *imr* and *otr* maintain sister chromatid cohesion. Loss of centromeric heterochromatin leads to increased sensitivity to the microtubuli-destabilizing drug TBZ. The phenotype also includes loss of silencing of centromeric reporter genes.

The *otr* regions are composed of one or up to seven copies of *dg* and *dh* repeats. These are four to six kb long and composed of non-coding, AT-rich sequence. Analysis of mutants with the characteristic loss of heterochromatin phenotype has lead to a model in which deacetylation of histone H3 and H4 tails by HDACs allows subsequent methylation of histone H3 at Lysine 9 (H3K9me) by the single *SuVar(3-9)* homologue in *S. pombe,* Clr4. Histone H3K9me is the mark of recognition for the HP1 homologue Swi6, which binds H3K9me via a chromodomain (Bannister, Zegerman et al. 2001). Swi6 then self-oligomerizes and heterochromatin forms. Transient inhibition of HDACs by the drug TSA leads to a heritable loss of centromeric heterochromatin, indicating that once established the chromatin structure, whether it is heterochromatin or euchromatin, is propagated epigenetically (Ekwall, Olsson et al. 1997).

The heterochromatin at the centromeres of *S. pombe* has become an extensively investigated model for the mechanism of RNA interference (RNAi) in directing histone tail modifications and chromatin rearrangement. *S. pombe* is, together with the ciliate *Tetrahymena thermophila*, the model organism where RNAi-directed chromatin modifications were first discovered (Mochizuki, Fine et al. 2002; Provost, Silverstein et al. 2002; Taverna, Coyne et al. 2002; Volpe, Kidner et al. 2002).

## 1.3 RNA INTERFERENCE

"Classic" RNAi is a process in which dsRNA or RNA folded into ds structures are "diced" by Dicer into small, 20-30 nt dsRNA (Fire, Xu et al. 1998; Hamilton and Baulcombe 1999). These are incorporated into Argonaute complexes, which select the

active strand. The sRNA subsequently guides the Argonaute-complexes via basepairing to complementary mRNA. The fates of targeted mRNA are either degradation initiated by Argonaute cleavage or sequestration away from the translational machinery in P-bodies. Sequestered mRNA may be released for further translation or may be degraded at a later point in time (Jackson and Standart 2007).

RNAi is a conserved pathway present in animals, plants and simple eukaryotes such as *S. pombe*. Even Budding yeasts were recently demonstrated to have remnants of RNAi; introduction of genes encoding a non-canonical Dicer and Argonaute from the close relative *Saccharomyces castelli* sufficed to induce silencing of an endogenous retrotransposon (Drinnenberg, Weinberg et al. 2009).

There are several examples of evolutionarily conserved microRNA (miRNA) precursor genes, in both animals and plants. Their expression is important for correct development during embryogenesis. A miRNA recognizes the target via basepairing of the first 2-8 nt to mRNAs, either within the coding regions (plants) or in the 3'-UTR (animals). The short stretch of bp infers that one miRNA may have multiple targets. Conversely, one mRNA may be targeted by several miRNA.

The introduction of multicopy genes results in silencing of the endogenous gene copy via the formation of short interfering RNA (siRNA) (van der Krol, Mur et al. 1990). This discovery has had great impact on molecular biology since introduction of exogenous siRNA can be used for gene knockdowns, which is very useful for investigating gene functions. The biological explanation of this phenomenon is that RNAi serves as a defence against viral RNA (in plants) and selfish DNA elements such as transposons. In mammals, introduction of long dsRNA triggers the PKR pathway (Rana 2007).

Endogenous siRNA are usually derived from arrays of non-coding repeats, which are often full of transposons or remnants of transposons. Endogenous siRNA may direct chromatin modifications also known as Transcriptional Gene Silencing (TGS) or, as in *T. thermophila*, the elimination of DNA elements. TGS is achieved by recruitment of DNA methyltransferases, HMTases, chromodomain proteins and/or polycomb to homologous loci, presumably via basepairing of sRNA to complementary RNA or DNA. In *C. elegans*, addition of sRNAs conferred changes in gene expression that were inherited for several generations (Vastenhouw, Brunschwig et al. 2006). The addition of siRNA not only leads to gene silencing but may also result in gene activation (Li, Okino et al. 2006; Janowski, Younger et al. 2007; Vasudevan, Tong et al. 2007).

In addition to miRNA and siRNA, new names of classes of sRNA that regulate gene expression are continually launched. Worth mentioning are the PIWI-associated RNA (piRNA) that are expressed in the germline of animals where they are required for repression of transposable elements (Halic and Moazed 2009) and QDE-2 associated RNA (qiRNA) that are involved in protection against DNA damage (Lee, Chang et al. 2009).

RNAi is a rapidly expanding field of research, mirroring the extensive roles of sRNA in regulation of various cellular processes, such as transcription, mRNA translation and degradation and genome stability via silencing of transposons as well as protection against DNA damage.

### 1.3.1 The RNAi Enzymes

RNAi typically involves enzymes such as Dicer, Argonaute and RNA-directed polymerases (RDP). Most animals and plants have several Dicer and Argonaute isoforms, which may be partly redundant (Table 3). There are also suppressors of RNAi, for example the ribonuclease Eri1, originally identified in *C. elegans*, which specifically degrades siRNA duplexes and thereby prevents accumulation of siRNA (Iida, Kawaguchi et al. 2006), or Small RNA Degrading Nucleases (SDN), recently discovered in *Arabidopsis thaliana*, that degrade miRNA (Ramachandran and Chen 2008).

#### *1.3.1.1 Dicer*

Dicer belongs to the family of RNAse III enzymes with specificity for dsRNA (Robertson, Webster et al. 1967; Bernstein, Caudy et al. 2001). A typical Dicer has two RNAse III domains, an N-terminal helicase domain and a PAZ domain and weighs approximately 200 kDa. The PAZ-domain binds to one end of the dsRNA and has high affinity for the typical 2 nt 3'-overhang that is the typical consequence of Dicer cleavage. The distance between the PAZ-domain and the two RNAse III domains determines the length of the product (FIG. 9) (MacRae, Zhou et al. 2007). The *dcr1* gene in *S. pombe* has no PAZ domain, which infers that there is no inbuilt size determinant in the Dcr1 enzyme.

**Table 3. Factors involved in RNAi-Directed TGS.**

| | *S. pombe* | *C. elegans* | *D. melanogaster* | *H. sapiens* | *A. thaliana* |
|---|---|---|---|---|---|
| DNA methylation | - | - | Low levels | Yes | Yes |
| H3K9me | Yes | Yes | Yes | Yes | Yes |
| HP1 | Swi6, Chp1, Chp2 | HPL1, HPL2 | HP1, HP1b, HP1c | HP1α, HP1β, HP1γ | TFL2, LHP1 |
| sRNA | siRNA | miRNA, siRNA | miRNA, siRNA, piRNA | miRNA, siRNA, piRNA | miRNA, siRNA |
| Dicer | 1 | 1 | 2 | 1 | 4 |
| Argonaute | 1 | 27 | 7 | 8 | 10 |
| RDP | 1 | 4 | 1 | ? | 6 |
| Pol IV, V | - | - | - | - | Yes |
| siRNA ribonuclease | Eri1 | Eri1 | CG6393 | THEX1 | Yes |

*Adapted and reprinted with permission from Macmillan Publishers Ltd [Cell Research](Djupedal and Ekwall 2009), copyright (2009).*

In animals, another RNAse III family member, Drosha, processes primary miRNA into precursor miRNA. Dicer subsequently cleaves precursor miRNA into miRNA. There are no homologues of Drosha outside the animal kingdom and primary miRNA in plants are processed by one of the Dicer isoforms, DCL1 (Meins, Si-Ammour et al. 2005).

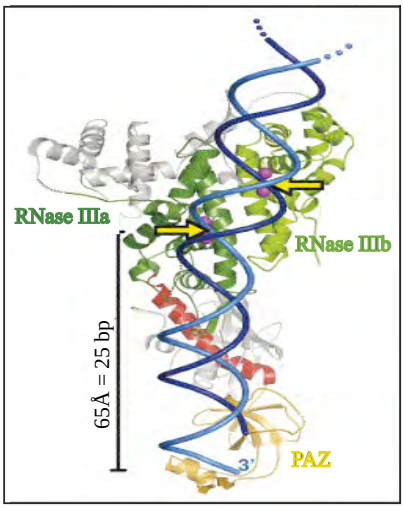


*Figure 9. Crystal structure of the "minimal" Dicer from Giardia intestinalis, composed of only the PAZ domain and two RNAse III domains, with dsRNA substrate (blue). Yellow arrows mark sites of cleavage. Reprinted from Current Opinion in Structural Biology, Vol 17. MacRae, Doudna, Ribonuclease revisited: structural insights into ribonuclease III familyenzymes. Copyright (2007), with permission from Elsevier.*

#### *1.3.1.2 Argonaute*

Argonaute proteins consist of a PAZ-domain, a MID-domain and a PIWI-domain that resembles RNase H. The PAZ- and the MID-domains bind the 3'- and 5'-ends of the siRNA or miRNA. Selection of the active strand is based on the relative thermostability of the bp at the ends of the sRNA duplex (Schwarz, Hutvagner et al. 2003).

In organisms with several Argonaute isoforms, sRNA are sorted into different Argonaute-complexes based on the identity of the 5'-nt and the degree of complementarity of the sRNA duplex – bulges in the duplex are common among miRNA whereas perfect homology leads to entry into the siRNA pathway (Forstemann, Horwich et al. 2007; Mi, Cai et al. 2008). Hence, the MID-domains of different Argonaute isoforms may have evolved specificity for different nucleotides (Jinek and Doudna 2009).

The Argonaute isoforms differ in the ability to cleave or "slice" target RNA. Argonaute proteins are the core components of several complexes such as RISC (RNA-Induced Silencing Complex) and RITS (RNA-Induced Transcriptional Silencing), and mediate TGS and PTGS by miRNA, siRNA, piRNA and qiRNA. In fact, piRNA biogenesis is independent of Dicer and only requires sequential activity of different Argonaute isoforms.

#### *1.3.1.3 RNA-Directed Polymerases*

RDPs are crucial for RNAi in plants and yeast. Recently RDPs involved in RNAi in animals were discovered. In *D. melanogaster* the D-elp1 subunit of the pol II elongator complex was demonstrated to have RDRP activity and to be involved in RNAi (Lipardi and Paterson 2009). D-elp1 has homologues with RDP activity in *S. pombe*, *Caenorhabditis elegans* and humans. It is therefore likely that RDPs participate in the RNAi pathways in animals.

*In vitro* investigation of the properties of an RDP in *Neurospora crassa* demonstrated the usage of ssRNA as template in primer-independent synthesis of full-length or sequential production of 9-21 bp dsRNA (Makeyev and Bamford 2002). In *C. elegans*, such short secondary siRNA constituted the majority of the siRNA population (Aoki, Moriguchi et al. 2007; Pak and Fire 2007; Sijen, Steiner et al. 2007). The contribution of RDP activity to the intensity of the sRNA signal therefore seems to be substantial.

### *1.3.2* Current Model for RNAi at the Centromeres of *S. pombe.*

In *S. pombe* there are single copies of the canonical RNAi-enzymes, Ago1, Dcr1 and Rdp1 (Volpe, Kidner et al. 2002). Ago1 and Dcr1 mediate both TGS and PTGS (Raponi and Arndt 2003; Sigova, Rhind et al. 2004). Whereas TGS at the centromeres has been extensively investigated little is known about RNAi-based PTGS in this organism. RNAi is required for the heterochromatin at the centromeres, telomeres and the silent mating type locus in *S. pombe*, although there are redundant mechanisms for heterochromatin maintenance at the telomeres and the silent mating type locus (Hall, Shankaranarayana et al. 2002).

Several complexes that are required for centromeric heterochromatin formation have been identified, which has led to a proposed model (FIG. 10). Here, pol II transcripts are used as templates for dsRNA synthesis by RDRC whose catalytic activity is derived from Rdp1 (Motamedi, Verdel et al. 2004; Sugiyama, Cam et al. 2005). Dcr1 processes dsRNA into ds-siRNAs, which are incorporated in the Arc complexes, whose major component is Ago1 (Buker, Iida et al. 2007). Ago1 selects the active siRNA strand and the passenger strand is probably degraded rapidly (my own observation of sRNA accumulation from the centromeres). Ago1 loaded with siRNA assemble into a new complex, the RITS (Verdel, Jia et al. 2004). RITS are targeted to centromeres in two ways, via siRNA binding to complementary transcripts and via the

chromodomain of Chp1, a HP1 homologue that recognizes and selectively binds the histone H3K9me mark.

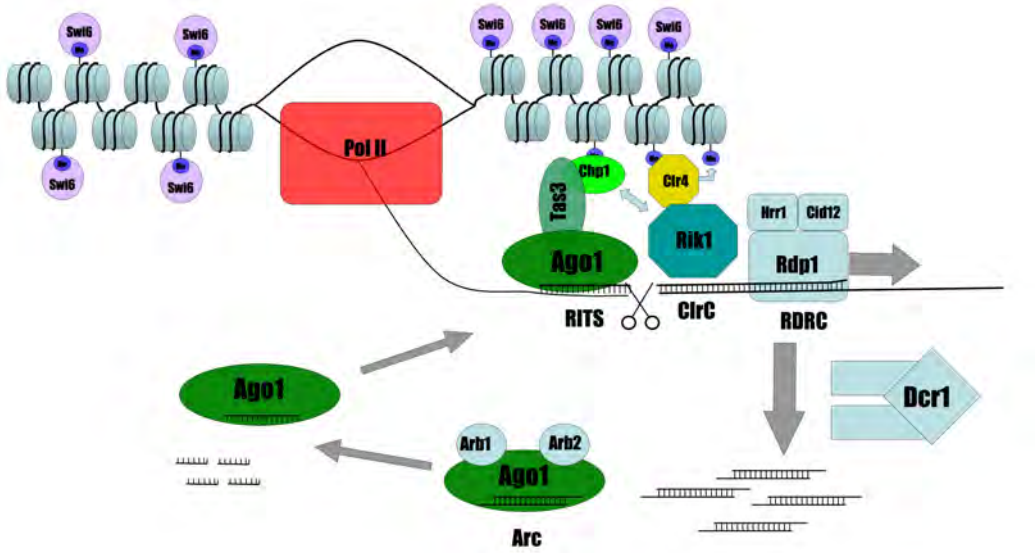


*Figure 10. Simplified model of RNAi-directed heterochromatin formation at the otrs of S. pombe. Ago1-siRNA complexes target centromeres via basepairing to pol II transcripts and via Chp1 binding to H3K9me (blue circles). Histone H3K9me is recognized by the HMTase Clr4 as well as by Swi6. Reprinted by permission from Macmillan Publishers Ltd: [Cell Research] (Djupedal and Ekwall), copyright (2009).*

Subsequently, the ClrC complex, containing the only known HMTase specific for histone H3K9me in *S. pombe*, is recruited. Clr4 not only generates H3K9me but also binds the mark via a shadow-chromodomain (Zhang, Mosch et al. 2008). H3K9me are recognized and bound by Swi6, which like other HP1 homologues directs nucleosome packaging into a dense heterochromatic structure.

In *S. pombe*, deletions of *ago1*, *dcr1*, *rdp1*, *swi6* and *clr4* all result in a characteristic phenotype indicative of centromeric heterochromatin loss. This is evidence of an absolute requirement for all of these enzymes in formation of centromeric heterochromatin. As in other model organisms, introduced RNA hairpins are diced by Dcr1 into siRNAs that may direct silencing in *trans* dependent on the local chromatin structure, amplification by RDRC and the abundance of Swi6 (Iida, Nakayama et al. 2008).

Pol II transcription of the *otr* region has been demonstrated to mainly occur after DNA replication when the chromatin structure at the centromere is less heterochromatic and more permissive for transcription (Chen, Zhang et al. 2008; Kloc, Zaratiegui et al. 2008). If siRNA are only generated at this specific time in the cell cycle, some siRNA need to persist until the next S-phase as there are dual requirements for both siRNA and H3K9me in the current model of RNAi-directed heterochromatin formation.

# 2 COMMENTS ON METHODOLOGY

## 2.1 *S. POMBE* AS A MODEL ORGANISM

Simple model organisms have the advantages of relatively small genomes, few genes and being easy to manipulate genetically. They are also cheap to maintain, rapid to grow and it is easy to obtain plenty of material. Of the simple eukaryotic model organisms, *S. cerevisiae* is the most common. In this study, *S. pombe* was exclusively used to investigate the function of Rpb7 in transcription and silencing. *S. pombe* has most of the advantages of *S. cerevisiae*, although there is less gene annotation. In addition, *S. pombe* bears more resemblance to more complex eukaryotes regarding the structure of the centromeres and number of conserved histone modifications. Unlike *S. cerevisiae*, *S. pombe* has single copies of the genes encoding the RNAi-components Dcr1, Ago1 and Rdp1 (although a recent report demonstrate remnants of RNAi in *S. cerevisiae* (Drinnenberg, Weinberg et al. 2009)). Furthermore, some evidence indicates that *S. pombe* may be a better model for eukaryotic transcription. For example, whereas the distance between TATA-box and promoter is between 40-100 bp in *S. cerevisae*, it is approximately 30 bp in *S. pombe*, which is similar to more complex eukaryotes (Spahr, Calero et al. 2009). In *S. pombe*, Rpb4 is an essential protein and is more similar in sequence and structure to orthologues in plants and animals (Sakurai, Mitsuzawa et al. 1999).

*S. cerevisiae* and *S. pombe* are estimated to have diverged from a common ancestor hundreds of million years ago, and some genes are as diverged between these two yeasts as their metazoan homologues (Wood, Gwilliam et al. 2002). Genes and molecular mechanisms that are conserved in both these yeasts therefore have a good probability of being conserved in other eukaryotes as well. Hence, studies in *S. pombe* may complement studies in *S. cerevisiae* and help us gain insight into the molecular biology of other eukaryotes.

## 2.2 METHODS FOR INVESTIGATION OF RNA LEVELS

During these studies I have used several methods for investigating and quantifying RNA levels; RT-PCR, Northern analysis, 5’ and 3’-RACE, Transcription Run On Analysis (TRO), genome-wide expression profiling (Eurogentech), Real time-RT-PCR, Northern analysis of sRNAs, analysis of 454 pyrosequencing of sRNA and finally,

strand specific transcriptome analysis of total RNA (Affymetrix pombe 1.0 array). In the following section I will discuss advantages and disadvantages of some of these methods.

### 2.2.1 RT-PCR, Northern Analysis and TRO.

With RT-PCR, the presence and levels of specific RNA molecules are probed. Reverse transcriptase and primers such as oligo-dT, random hexamers or gene specific primers are used to produce cDNA. This is followed by PCR with primers for the transcript. By using only one sequence specific primer for reverse transcription, you can potentially obtain strand-specific RT-PCR. However, inherent to reverse transcriptases is the ability to use cDNA as a template, although the affinity is less than for RNA (Perocchi, Xu et al. 2007). Therefore, strand-specific RT-PCR can be very difficult to achieve. If relative transcript levels need to be assessed, the PCR reaction needs to be stopped in the exponential phase of sequence amplification, rather than after entry into stationary phase in which one or several of the components required for sequence synthesis have become limiting.

It is better to choose Real-time-RT-PCR, also known as Q-PCR, where all reactions are run into stationary phase and the amplification rate of all samples are recorded. These recordings are used to calculate differences in the amount of starting template by comparisons of the numbers of PCR-cycles required to reach the exponential phase of amplification between positive and negative controls and the sample of interest. In my experience, cDNA from all negative controls will start to be amplified eventually. My recommendation is to only use reverse transcription in combination with real-time PCR if relative levels of transcripts are investigated. The disadvantages of real-time PCR are that multiplex reactions cannot be run, and the requirement for (at least) triplicates of each sample make it complicated to test even an intermediate number of samples. Handling of too many samples at the same time is often associated with mistakes or errors that occlude the results.

In my opinion, Northern analysis is a more robust method for assessing transcript levels. Size-separated RNA is blotted from gels onto membranes that are directly probed against your favourite transcript. No amplification step is required due to the sensitivity of radioactive labelling. In addition, an estimate of transcript length and splice variants is obtained. The disadvantage is the requirement for radiaoactive probes and all the extra work that the handling of radioactivity demands.

TRO is a method used to distinguish effects on transcription from effects due to changes in RNA stability. Addition of radioactive nucleotides to growing cells is followed by RNA extractions and hybridization to probes of interest. In my experience, TRO is much more complicated than Northern analysis, due to the need to do RNA purification on “hot” material. In addition, high levels of radioactivity are required to get decent signals of moderately expressed transcripts. Furthermore, the safety regulations at Karolinska Institutet limit the number of samples possible to prepare at the same time to two, which is often too few for an optimal experimental design.

### 2.2.2 Genome-wide Analysis of Transcripts

Genome-wide analysis of transcripts is like a reversed northern analysis with thousands of probes on an array to which you hybridize a pre-labelled sample. We used the first generation of commercially available microarrays for *S.pombe*, the Eurogentech cDNA microarray in paper I and II. This microarray contained 500 nt probes complementary to the 500 bp upstream of the translation stop codon of 99% of annotated genes in *S. pombe* (Xue, Haas et al. 2004). The next generation of microarray had, in addition to probes corresponding to the annotated genes, the sequence complementary to the 500 bp upstream of the translation start codon of nearly all annotated genes. This region, called the IGR, would for the average gene cover the 5’-UTR regions, the TSS and the proximal promoter. This microarray was designed for genome-wide ChIP experiments, a method that had recently been developed for *S. cerevisiae* (Robyr, Suka et al. 2002). The method was adjusted to *S. pombe* in our lab (Wiren, Silverstein et al. 2005). Both Eurogentech microarrays were designed for the hybridization of two cDNA samples to the same chip and therefore for relative comparisons between, for example wild type and mutant.

Meanwhile, Affymetrix developed an array, the GeneChip® *S. pombe* Tiling 1.0FR Array, with 25 nt tiles of both strands of the entire *S. pombe* genome. The tiles have a 5 nt overlap at each end. This means that a gene of 1000 bp produces 50 measurements (to be compared with two single measurements from the Eurogentech arrays). Samples are given to the BEA core facility (www.bea.ki.se), which minimizes technical variation due to differences in sample handling and reagents. In my hands, these arrays are highly reproducible in both genome-wide transcription analysis and protein binding analysis. However, although provided and sold by Affymetrix, the company did not support the product *in silico*. Therefore software required for analysis

of the chips was difficult to acquire. This was especially true for transcription analysis, since the chip was developed for ChIP on chip analyses, strand specific files for assigning probes to the position within genome were difficult to acquire. In addition, genome annotation was lagging behind and generation of up-to-date annotation files were generated within our lab. In spite of this, this array is an excellent tool for investigating the effects of various factors on transcription, histone modifications and nucleosome positioning. The tiling of the entire genome allows an unbiased estimation of transcription of both sense and antisense to gene direction.

#### *2.2.2.1 Data Analysis from Genome-Wide Studies*

Thousands or millions of data points generated by one single experiment and the development of genome-wide methods for analysis depend to a great extent on the development of software and knowledge of statistics in order to make any sense. Another aspect is the terminology of DNA elements, which is often confusing or misleading.

Normalization is a treatment of data for removal of technical variation so that biological variation remains. For example, differences in signal intensity are not linear. A common method for normalization of the Eurogentech spotted two-colour array is Lowess. This method assumes that the majority of genes of both datasets are expressed at the same level. For this reason, I did not find this method appropriate for normalization of *rpb7G150D* mutant samples, since pol II transcription in general might be affected by a mutation of an essential subunit of pol II. I therefore used external spiking controls, RNA added in similar amounts to mutant and wild type samples prior to reverse transcription, whose intensity levels were used to adjust the average intensity levels of the samples. Although this method of normalization is not as robust as Lowess, the data obtained after normalization with external spiking controls were in line with results later obtained with *in vitro* transcription with the mutant pol II. Hence, Lowess normalization masks global differences and for studies of any factor that may have general effects on transcript levels it is definitely inappropriate.

For the affymetrix chips, the freeware TAS (Transcription Analysis Software) and IGB (Integrated Genome Browser) were used. Experiments in duplicates or triplicates were normalized with quantile normalization, to adjust the normal distributed curves of measurements, and scaling, which set the average intensity level to the same level. Normalizations of samples together (wild type vs mutant) or single normalization

(wild type and mutants normalized separately) did not result in visible differences, as assessed in IGB.

In addition to TAS and IGB, affymetrix data were also analysed with Partek (www.partek.com), which uses rma normalization as default. In Partek, ANOVA is used for systematic analysis of differences between samples. In the Partek workflow continuous sets of probes that have significantly altered levels of intensity as compared to the control sample are selected. The high number of measurements from each gene or DNA element makes it possible to distinguish regions with subtle changes between samples. Furthermore, annotation of 5´- and 3´-UTR regions as well as other transcribed elements (Dutrow, Nix et al. 2008) allows identification of changes in transcription or protein binding in these regions. This makes the Affymetrix array a powerful tool for genome-wide studies.

### 2.2.3 Detection of sRNA in *S. pombe*

A few years ago, when these studies were initiated, very few methods were designed for sRNA detection. Generally RNA species smaller than a 100 nt were washed out during RNA preparations. As sRNA are estimated to constitute approximately 2% of the total RNA of the cells they are rather abundant. Yet, northern analyses for detection of sRNA in *S. pombe* are difficult, in comparison to other species.

The most successful method for investigating sRNA in *S. pombe* is deep sequencing, such as 454, Illumina or Solexa. Illumina or Solexa can produce tens of millions of reads of about 50 nt per run, whereas 454 produces a few hundred thousand reads of about 500 nt per run (Blencowe, Ahmad et al. 2009). Deep sequencing gives a snapshot of the sRNA in the population of cells from which the RNA was harvested. In addition, deep sequencing may be applied to RNA samples purified from protein complexes. Among the advantages of deep sequencing compared to microarrays are that there is no cross-hybridization and the range is unlimited since the absolute number of sequences are used to estimate relative abundance (Cloonan and Grimmond 2008). However, deep sequencing is expensive and the large number of sequences demands handling by professional bioinformaticians since there is no available software for laymen.

## 2.3 CHIP FOR INVESTIGATION OF PROTEIN BINDING TO DNA

### 2.3.1 ChIP on Chip

Chromatin immunoprecipitation is the most commonly used molecular biology method for analyses of DNA-protein interactions. In summary, log-phase growing cells are fixed by formaldehyde for 30 minutes or up to several hours, followed by sonication of DNA into fragments of approximately 500 bp lengths. Specific antibodies are incubated with the chromatin extract and the antibody-protein-DNA complex is subsequently pulled down by the use of protein A beads. After several washes of the DNA-protein-antibody-protein A complex, the DNA is released from the antibody by reversal of the cross-linking. The enriched DNA is amplified and relative enrichment is assessed by the use of DNA that has been obtained either without addition of antibody or by adding the antibody to a strain without the target protein. Usually the quantity of the DNA sequence of interest is compared to the quantity of a known target, and in addition to ChIPs of proteins that are known to bind/not bind this region.

In ChIP on chip, the amplified DNA from a ChIP is labelled and hybridized to a chip, such as the pombe 1.0 array provided by Affymetrix, in order to obtain genome-wide profiles of protein binding or occupancy.

ChIP is a complex method and there are several important aspects to consider. First, the distance between the protein of interest and the DNA. Second, the abundance of the protein of interest. Best results are obtained with abundant proteins close to the DNA, making histones the optimal target for ChIP on chip. To improve the results, one may increase the time of fixation or add Diadipimidate (DMA) to increase the levels and distance of cross-linking (Kurdistani, Robyr et al. 2002). Third, the specificity and affinity of the antibody used for immunoprecipitation is absolutely crucial for a successful experiment. For this reason, proteins with a tag of a known epitope are often used. Fourth, extensive washing of protein A beads and chromatin is necessary to remove background signal. Last, but not least, it is essential to simultaneously run negative controls with non-immunoprecipitated material. For ChIP on chip, these samples are used in the normalization process to remove signals obtained by background.

In summary, ChIP on chip is a difficult method that for many proteins requires long time to master and many controls. The reward comes from the good, reproducible maps (correlation coefficients above 0.9) of genome-wide protein occupancy.

# 3 RESULTS AND DISCUSSION

## 3.1 AIMS

The aim of this study was to investigate the mechanisms of action of the *rpb7G150D* mutation and pol II in RNAi-dependent heterochromatin formation at the centromeres of *S. pombe*.

In order to gain a better understanding of synthesis of centromeric siRNA precursors, total sRNA content in wild type *S. pombe* cells was assessed. Moreover, the general roles of Rpb7 and its partner Rpb4 in the context of pol II transcription were investigated.

- Functional characterization of *S. pombe* Dcr1 and its role in centromeric heterochromatin formation (paper I).
- Investigation of the general and centromere-specific effects of the *rpb7G150D* mutation (paper II).
- Analysis of sRNA production and accumulation in wild type and *rpb7G150D* cells (paper III).
- Further investigation of the role of Rpb4/Rpb7 in general transcription (paper IV).

## 3.2 PAPER I: THE CENTROMERES OF *S. POMBE* ARE TRANSCRIBED AND PROCESSED BY DICER

Unlike *S. cerevisiae*, a gene homologous to the human Dicer gene was discovered in *S. pombe*. By using homologous recombination a *dcr1Δ* strain and a Dcr1-HA tagged protein were generated (Bahler, Wu et al. 1998). In addition, the *S. pombe dcr1+* gene and the human Dicer gene were cloned in a selectable vector. This material was used for characterization of Dcr1 protein function in *S. pombe*.

We found the *dcr1Δ* to cause slow growth and sensitivity to low temperatures compared to wild type cells. Little or no effect on growth was apparent at elevated temperature. Interestingly, this could be due to the temperature dependence that was recently demonstrated for the RNAi pathway; RNAi was demonstrated to be inactivated at elevated temperatures (Kloc, Zaratiegui et al. 2008).

The *dcr1Δ* made the cells hypersensitive to TBZ, a poison that disrupt spindle-tubule formation and is commonly used to assay defects in sister chromatid segregation. Indeed, *dcr1Δ* cells lead to high numbers of lagging chromosomes during anaphase as visualized with immunofluorescence. Interestingly, *dcr1Δ* TBZ sensitivity was partially rescued by human Dicer, indicating functional conservation of the Dicer proteins of these distantly related species.

TBZ sensitivity and lagging chromosomes are typical for mutations that disrupt centromeric heterochromatin, which is why the expression of a centromeric *ade6+* gene was assessed by RT-PCR (and by colour; expression of *ade6+* results in white yeast colonies whereas repression of the gene produces a nice red colour). The activation of *ade6+* indicates loss of heterochromatin at this locus. RT-PCR with primers for endogenous centromeric RNA revealed accumulation of centromeric transcripts in *dcr1Δ* cells but not in wild type cells.

Since Dcr1 has two RNAse III domains, we purified the HA-tagged Dcr1 and assayed its ability to cleave dsRNA *in vitro*. Dcr1-HA processed the human miRNA precursor *Let-7;* a well characterized Dicer substrate, as well as dsRNA corresponding to the centromeres and the *ura4* gene in *S. pombe* (data not shown) into sRNA of the expected sizes.

Finally, *dcr1Δ* cells were subjected to genome-wide expression profiling analysis in order to estimate the role of Dcr1 in mRNA regulation. At this time, the method was under development and very stringent conditions were used for the data analysis. This may explain why only ten genes were reproducibly detected as induced in *dcr1Δ* cells in this study. Indeed, later reports demonstrate a few hundred genes to be affected by *dcr1Δ* (Hansen, Burns et al. 2005). Of the ten induced genes, two contain an 11-pb DNA motif that was present at 4 loci in total within the entire *S. pombe* genome.

We conclude that Dcr1 of *S. pombe* is functionally conserved to Dicer proteins in other organisms in that it cleaves dsRNA into short dsRNA. Furthermore, we provide support for involvement of Dcr1 in the processing of centromere-derived transcripts, which is necessary for formation of heterochromatin, as described by Volpe et al. earlier that year (Volpe, Kidner et al. 2002). Finally, the possibility of an active RNAi pathway in mRNA regulation was investigated via expression profiling of *dcr1Δ* cells. The effects of *dcr1Δ* on mRNA levels were minor, possibly due to redundancy in gene regulation pathways.

## 3.3 PAPER II: TRANSCRIPTION BY POL II IS REQUIRED FOR HETEROCHROMATIN FORMATION AT THE CENTROMERES

The centromeres of *S. pombe* were considered to be "silent" due to the repressive effects of heterochromatin. Transcripts originating from the centromeres were therefore an unexpected discovery and the polymerase responsible for this non-canonical transcription was not known. However, the *csp3* mutant, which was found in a screen for mutations that caused centromeric silencing defects (Ekwall, Cranston et al. 1999), was found to be identical to *rpb7*, a gene encoding a subunit of pol II. Therefore we suspected that the centromeric siRNA precursors were pol II transcripts.

The *csp3* phenotype is caused by a point mutation that substitutes an aspartic acid for a glycine within one of the two predicted nucleic acid binding motifs of *rpb7*. This amino acid is conserved from yeast to man. First, we wanted to test whether *rpb7G150D* was involved in the same pathway for heterochromatin formation as *dcr1*. Since Dcr1 was expected to cleave all centromeric transcripts in *rpb7G150D* cells, a double mutant was created and tested for accumulation of centromeric transcripts. Indeed, centromeric transcripts accumulated in the double mutant at the permissive temperature but decreased at the non-permissive temperature. Thus, pol II was necessary for transcription of these pre-siRNAs.

Accumulation of centromeric transcripts was assessed by Northern analysis, strand specific RT-PCR and Q-PCR. A peculiar finding was the slight increase of transcripts from the positive DNA strand in *rpb7G150D* cells. Theoretically, Dcr1 would process these into sRNA, as in wild type cells.

To rule out indirect effects of the *rpb7G150D* mutation, expression profiling was done at permissive and non-permissive temperatures. Although fewer than 5 % of annotated transcripts were significantly affected by *rpb7G150D* at the permissive temperature, massive effects on general transcript levels were obvious at the non-permissive temperature. In addition, *in vitro* transcription assays with purified pol II from the *rpb7G150D* strain showed a drastic reduction in transcription at the non-permissive temperature. Furthermore, the pol II subunit composition *in vitro* was similar in the mutant polymerase compared to the wild type, indicating no difference in the stability of the association of Rpb7G150D/Rpb4 to pol II. Therefore, the alleviation of silencing at the centromeres at the permissive temperature seemed to be a specific defect caused by *rpb7G150D* rather than a general problem with transcription.

Next, we wanted to investigate whether it was actual transcription that was affected in *rpb7G150D* cells rather than pre-siRNA stability. For this we used transcription run on analysis. Reduced levels of transcripts from the negative strand and increased levels of transcripts from the positive strand were detected. This supported a hypothesis in which either initiation or elongation of transcription was affected by *rpb7G150D* cells. To further test the step of the transcription cycle in which *rpb7G150D* was deficient, a reporter gene was cloned downstream of the region upstream of the TSS of the centromeric transcripts, as determined by 5'-RACE. These constructs were transformed into wild type and *rpb7G150D* cells and levels of reporter protein activity were measured. Cloning of the region upstream of the TSS of the siRNA-precursors led to high levels of enzyme activity, indicative of a fully functional promoter. In *rpb7G150D* cells, however, the control promoter resulted in levels of enzymatic activity that were comparable to wild type whereas the identified centromeric promoter produced enzymatic activity barely distinguishable from background levels.

In conclusion, pol II synthesizes centromeric siRNA precursors in *S. pombe* and Rpb7 is specifically required for recognition of the pre-siRNA promoter. Lower levels of centromeric siRNA precursors leads to disruption of the repressive heterochromatin at this locus. We speculate that disruption of the heterochromatin causes the increased transcript levels of the opposite strand that were detected in *rpb7G150D* cells.

## 3.4 PAPER III: ANALYSIS OF CENTROMERIC SRNA SYNTHESIS AND ACCUMULATION

Here, we analyse the centromeric sRNA population in *S. pombe*. After filtering of sequences shorter than 16 nt, sequences without a perfect match in the genome or sequences matching structural RNA, some 3500 wild type sequences remained.

Comparisons of our wild type sequences to deep sequencing results of total sRNA from *rpb7G150D* cells (11000 sequences, this study) and deep sequencing of siRNA purified from Ago1-complexes (almost 200,000 sequences, (Buhler, Spies et al. 2008)) revealed interesting differences with implications for the mechanisms of synthesis and accumulation of sRNA. Whereas Ago1-associated siRNA had a pronounced preference for uracil at the 5'-end, were predominantly 22-23 nt long and more than 60% matching to centromeres, sRNA from *rpb7G150D* cells had no preference for uracil at the 5'-end, had a broad size distribution and very few sRNA

matching the centromeres. The wild type sample had characteristics intermediate between these extremes, that is, some preference for uracil, somewhat broader size distribution and more than 40% of the sequences were matching the centromeres.

Mapping of centromeric sRNA revealed two clusters, repeated twice or more at each centromere, from which the majority of sRNA were derived. The DNA sequence of these two clusters had remarkably high degrees of sequence identity. Within the clusters, regions that were devoid of sRNA surrounded hotspots of abundant sRNA. Surprisingly, there was a significant bias for wild type sRNAs corresponding to one of the DNA strands. For the Ago1-associated sRNAs the bias, small but significant, was for the opposite DNA strand. The clusters and a handful of the most abundant sRNA hotspots were validated by Northern analyses. The sRNA were demonstrated to be Dcr1-dependent, have 5-monophosphates and be associated with Ago1. Hence, these sRNA fulfil the criteria for *bona fide* siRNA.

One hypothesis explaining the maintenance of the high degree of sequence identity at the siRNA clusters is that the sequence is important for formation of secondary structures that can be cleaved by Dcr1 into siRNA, in a process analogous to miRNA production. It is well established that miRNA precursors form ds-secondary structures are cleaved into miRNA. The secondary structure of the first 432 nt of the *dg* pre-siRNA, which also include a 300 nt translocation from the *dh* cluster, was experimentally determined *in vitro* and a partially ds structure was revealed. When incubated with human recombinant Dicer, short RNAs of the expected sizes were detected. Finally, centromeric sRNA, some of which match the experimentally determined ds region, were sequenced in *rdp1Δ* cells.

In conclusion, we demonstrate that the species of centromeric sRNA that are synthesized and accumulate in wild type cells differ from those incorporated into Ago1 complexes. Furthermore, not all regions of the siRNA precursors are as likely to end as siRNA. There is little if any “phasing” and low accumulation of siRNA passenger strands. The strong strand bias of wild type sequences compared to the reversed strand bias of Ago1-associated siRNA is puzzling. According to the current model, siRNA are produced from RNA made ds by Rdp1, which would lead to accumulation of siRNA complementary to both strands. In our proposed model, direct cleavage of an ssRNA folded into a hairpin may contribute to the observed excess of siRNA complementary to one strand. RNAi is required for heterochromatin establishment and maintenance at the centromeres, although Swi6 has been reported as the primary determinant for spreading and inheritance of heterochromatin (Hall, Shankaranarayana et al. 2002).

Folding of siRNA precursors into ds structures that are recognized and processed by Dcr1 into siRNA could provide the first spark to initiate the feedback loop of RNAi-dependent heterochromatin formation at the centromeres of *S. pombe*.

## 3.5 PAPER IV: RPB7 AND RPB4 HAVE DIFFERENT PROFILES OF DNA OCCUPANCY

There is extensive literature regarding Rpb4 and Rpb7, yet their roles in transcription remain elusive. On the one hand, in *S. cerevisiae* both Rpb4 and Rpb7 were recently reported to have similar genome-wide occupancy profiles to that of the core pol II subunit Rpb3 (Jasiak, Hartmann et al. 2008; Verma-Gaur, Rao et al. 2008). On the other hand, Rpb4 is not essential for viability in the same organism and neither Rpb4 nor Rpb7 are required for promoter independent initiation and elongation *in vitro* (Edwards, Kane et al. 1991). There are several additional indications that Rpb4/Rpb7 may have a regulatory function and/or that the dimer may transmit regulatory information.

We decided to investigate the genome-wide occupancy profiles of Rpb4, Rpb7 and core pol II subunit Rpb2 with ChIP on chip. We used the Affymetrix 1.0 pombe array that has genome-wide tiles of both strands of 20 bp resolution, resulting in protein occupancy profiles of high reproducibility, specificity and sensitivity. The profiles of all three proteins are similar with Pearson correlation coefficients around 0.9. This indicates that both Rpb4 and Rpb7 are associated to core pol II at most genes. However, although the profiles are very similar, Rpb4 was specifically absent at several loci where Rpb2 and Rpb7 were highly enriched. This was surprising since Rpb4/Rpb7 were considered as a stable subcomplex and therefore likely to have identical enrichment profiles.

Systematic analysis of differences in protein binding was undertaken with variance normalization, which is a common method for assessing differences in highly similar ChIP on chip datasets (Kurdistani, Tavazoie et al. 2004). Due to limitations of the software, the variance normalisation was performed on average enrichment values of only the ORF and IGR regions. Quality clustering was used to generate cladograms of similar gene occupancy profiles. In this way, clusters of genes with similar occupancy profiles were identified.

Clusters of genes in which levels of enrichment of Rpb4 and Rpb7, respectively, were low and that had high degree of correlation with other processes or

factors were analysed further. We show a correlation between the genes that had lower enrichment of Rpb4 relative to Rpb2 and Rpb7, and genes that are induced after exposure to several stress conditions (Chen, Toone et al. 2003). Accordingly, many of these genes were enriched for a DNA motif in the promoter region. This DNA motif is related to the binding site for the transcription factor Atf1, one of the key regulators of stress response in *S. pombe*. In addition, many of these genes were repressed by Rpb4 as well as Rpb7 (Sharma, Marguerat et al. 2006). Rpb4, which has been reported to be necessary for stress adaptation of transcription in *S. cerevisiae* may therefore have a similar function in *S. pombe*.

Although the occupancy profile of Rpb7 was almost identical to that of Rpb2, there were differences at rare loci. Within the clusters with relatively low Rpb7 occupancy there was significant enrichment of genes involved in protein synthesis.

These ChIP data are of high quality and comparisons with our laboratory's extensive database of genome-wide protein occupancies, expression profiles of mutants and histone modification patterns promise interesting correlations with various factors involved in transcription and chromatin formation.

# 4 CONCLUSIONS AND PERSPECTIVE

In paper I we investigate the function of Dcr1 in *S. pombe*. Whereas the role of Dcr1 and RNAi at the centromeres has been extensively investigated, little is known about the function of Dcr1 outside the context of silent heterochromatin. The small effects of *dcr1Δ* on gene expression reported by us and by others (Hansen, Burns et al. 2005) do not necessarily mean that there is nothing relevant going on but rather indicative of the robustness of gene regulatory networks in *S. pombe*. A few publications report RNAi-mediated PTGS in *S. pombe* (Raponi and Arndt 2003; Sigova, Rhind et al. 2004; Gobeil, Plante et al. 2008). Furthermore, Dcr1 and Ago1 but not Rdp1 are required for normal cell cycle progression (Carmichael, Provost et al. 2004). Unlike Rdp1, which is nuclear, the bulk of Dcr1 and Ago1 reside in the cytoplasm in distinct *puncta* that may correspond to P-bodies (Carmichael, Stoica et al. 2006). This is in line with Dcr1 and Ago1 functions in PTGS rather than TGS, which has been demonstrated to occur in *cis*.

Apart from setting up heterochromatin at centromeres, telomeres and the silent mating type locus, RNAi was also demonstrated in TGS at the region between convergent gene pairs, which was necessary for termination of transcription (Gullerova and Proudfoot 2008).

I believe that *S. pombe* has the potential to be a good model organism both for PTGS and for TGS outside the centromere context. A more thorough analysis of the effects of *dcr1Δ* on gene expression is possible with the Affymetrix strand-specific tiling arrays and the new extended annotation that we recently have adapted to the Affymetrix platform.

In paper II we investigated the defect of *rpb7G150D* in the transcription of centromeric siRNA precursors. According to the reporter gene assay and the TRO the mutant has a specific defect in transcription from the centromeric promoter. In this model, lower levels of siRNA precursor transcripts would lead to fewer siRNA, resulting in an overall loss of heterochromatinization. Yet, one may assume that amplification by RDRC would compensate the lower dosage of the primary siRNA signal.

These results implicate Rpb7 in promoter recognition. Rpb7 has been cross-linked to promoter DNA *in vitro* (Chen, Mandal et al. 2004) although the protein is not directly adjacent to the DNA according to the crystal structures.

What makes this promoter different from all other, non-affected promoters? Is it the actual DNA sequence of the promoter or does the special chromatin environment require Rpb7-dependent recruitment of special factors in order to start productive transcription? There are several publications demonstrating that elongating pol II within "silent" loci has non-canonical CTD-phophorylation status, i. e. Ser5-P but not Ser2-P, which interferes with recruitment of factors involved in mRNA maturation and HMTases that normally associate with pol II via binding to Ser2-P (Brookes and Pombo 2009). It would definitely be interesting to investigate the CTD-phosphorylation status of pol II within the centromeres of *S. pombe*, as this could be a key finding in understanding why pol II transcription is required for maintaining silent heterochromatin.

In paper III my aim was to investigate the total sRNA population in *S. pombe* although only the centromeres were coherently analysed in the final manuscript. In hindsight it was unfortunate that we filtered out all sRNA corresponding to structural RNA. In *D. melanogaster*, RNAi has been demonstrated to regulate the nucleolar organization via H3K9me (Peng and Karpen 2007). The large number of sRNA matching rRNA and tRNA sequences seems to be a general feature of sRNA samples from *S. pombe*, indicating potential involvement of RNAi in rDNA chromatin organization in this organism as well. We have seen a sRNA pattern at one tRNA gene, where sRNA unintentionally were not removed by filtering, that was reminiscent of "phasing" with stacks of sequenced sRNA directly adjacent to each other. It would be highly interesting to take a look at sRNA sequences at tRNA and rRNA loci in order to estimate whether these are mere degradation products or whether there is a distinct pattern of siRNA, in analogy to sRNA at the centromeric clusters. I would also enjoy conducting a more thorough analysis of sRNA that match within genes, at promoters or 3'-UTRs. There are some indications that there are gene specific sRNA, as the same sequences were found repeatedly in both wild type and the *rpb7G150D* samples from 454 sequencing. Intriguingly all sRNA matching genes were of the sense orientation. Of course, interesting comparisons with wild type antisense transcription and expression analysis of *dcr1Δ* could be made in order to understand whether these sRNA are involved in gene regulation.

Regarding the clusters and hotspots of siRNA at the centromeres, I would very much like to do a systematic analysis of the differences in characteristics of the DNA sequence of hotspots compared to neighbouring sequence that does not give rise to

siRNA. We have, so far, shown a slight but significant increase in the GC content in the DNA sequences that spawn siRNA. Is there an increased ability to form ds secondary structures in the area surrounding the siRNA hotspots?

In paper IV we report the genome-wide DNA occupancy of Rpb7 compared to Rpb2 and Rpb4. Surprisingly, the profile of Rpb7 was similar to the profile of Rpb2 rather than Rpb4. I therefore believe that especially Rpb4 may influence the expression of groups of genes. Furthermore, there may be a specific requirement for Rpb4 at the termination step of the transcription cycle, as the average enrichment of Rpb4 peaks at the 3'-UTRs, perhaps via interaction of factors involved in 3'-end processing and termination. Rpb4 was recently reported to affect pol II association to 3'-ends of genes and recruitment of mRNA processing factors (Runner, Podolny et al. 2008).

We have seen many interesting correlations with our Rpb4/Rpb7/Rpb2 data to specific histone modifications as well as to expression profiles of various mutants. I am particularly interested in further investigating the role of Rpb4 at the 3'-ends of genes as well as the functions of Rpb4 in stress-induced transcription. Possible experiments to test for a direct role of Rpb4 in stress adaptation is to repeat the ChIP on chip experiments after exposure to different stresses in combination with expression profiling of the same cells, to see whether there is stress-dependent relocation of Rpb4 to induced genes.

In summary, this thesis deals with new aspects of gene regulation. We have shown that transcription does not necessarily result in gene activation but may direct heritable gene repression. The complexity of the pol II transcription apparatus and the plethora of transcripts that it produces is one of the most fascinating fields of research in molecular biology today. Future advances will contribute to our understanding of how it all really works.

# 5 ACKNOWLEDGEMENTS

Tack alla ni som varit med den här tiden! Det har varit bra år!

Särskilt tack till min handledare **Karl** för att jag fått den fantastiska möjligheten att doktorera i din grupp! Du har alltid varit smittande entusiastisk, oerhört generös och har gjort allt detta möjligt. Tack för att jag fått förverkliga mina idéer!

I would like to thank all the past and present members of the KEK- and AWR labs! My co-supervisor **Tony** for scientific discussions and advice!
**Julian**, for my first impression of Södertörn and pombe – for all the protocols, buffers, answering million of questions and scientific discussions, **Henrik**, **Pernilla**, **Rebecca** and **Barbara** for helping me out at the start.

Tack till **Stefan** (Hermann) för all hjälp!

Thanks to **Mickaël** for sharing ideas, protocols and most of all for helping me, on extremely short notice, with manuscript IV!
All the girls **Azadeh**, **Anna** & **Marianna, Yongtao** for all the fun times in and outside the lab! Special thanks to **Anna** for all the practical tips and support when writing this thesis!

Thanks **Shiounan** for all the laughs

Tack till **Coola Monica**! Du är en sann källa till inspiration!

Stort tack till **Fredrik** – vår klippa på Affymetrix – för ditt engagemang och för att du alltid är så trevlig ☺. Tack även till **David & Marika**! Och förstås **Karin Dahlman-Wright**!

**Annelie, Michelle, Andreas, Ulrika, Agata, Jenna, Andreas and Babett** for structuring the lab, planning fika, lab cleaning and so many other practical things! I am sure you all are going to make a positive difference wherever you go. Special thanks to **Jenna** for always critically reading my manuscripts on short notice ☺!

Tack **Peter**, för hjälp med figur 4, manuskript 4! Jag är säker på att du kommer att vara ett strålande tillskott till KEK-gruppen.

Stort tack till **Niklas**, bästa exjobbaren någonsin!

I would like to thank all the people that I have collaborated with throughout the years, especially **Jessica Lindvall**, **Henrik Spåhr, Olga Khorosjutina** and **Claes Gustafsson, Becky Mosher, David Baulcombe, Isabelle Kos-Braun, Gerhart Wagner, Aurélie Fender** and **Nadja Heidrich** for discussions and advice!

Tack för all hjälp till IT-avdelningarna på Södertörn och BioNut, **Sam, Asim, Rickard, Jonas** och **Anders!**

I would also like to thank all the people at Södertörn and NOVUM for pleasant company as well as for advice, especially, **Gayathri**, **Satish**, **Stefan**, **Nassim**, **Magnus**,

**Ida**, **Alex**, **Edmond**, **Robban**, **Richardo**, **Lotta**, **Galina**, **Ivo**, **Gunnar**, **Hussein**, **Annelie**, **Yann, Inger, Inger & Inger**.

Super special extra thanks to my past and present roomies: **Helmi, Karin, Wessam, Indranil & Carolina!** You are all so kind and helpful and I don´t know how I could have managed this without you! I will miss you all so much!
**Helmi**, som outtröttligt hjälper alla med allt och som förmodligen vet allt som är värt att veta. Jag lyssnar alltid främst på dina råd.
Ma chére amie **Karin** – det bästa vore om vi kunde fortsätta vara kollegor så att vi kan ses varje dag, alltid! Jag är så glad och stolt att du är min bästis.
Tack **Wessam**, ängel i människogestalt, för alla bullar, semlor och tårtbitar genom åren! Och för att du är så klok!
Gene spring genious **Indranil** for always being kind, patient and helpful! Especially, this last year, I don´t see how I could have finished manuscript IV without your expertice!
Generösa **Carolina** för alla pratstunder, allt stöd och alla skratt! Jag ser framemot att komma till jobbet för att träffa dig!

Tack till mina vänner för alla trevliga stunder – utan er vet jag inte vad jag skulle göra: Kloka **Evelina**, Änglalika **Maryam**, Trygga **Magdalena**, Glamorösa **Camilla**, Ärliga **Ida**, Fina **Mathias** och förstås och återigen älskade **Karin Karin Karin**.

Mina föräldrar, **Ulla och Lisse**, utan all barnvakt hade det tagit tre år till…
Tack **Erik & Annika, Farmor, Molle** för alla goda minnen!

Min älskade familj, **Dan och Laura**, världens mest underbaraste barn! Lika underbarast som sin far, sötaste, raraste, finaste **Stefan**. Vill alltid vara hos dig!

# 6 REFERENCES


Aoki, K., H. Moriguchi, et al. (2007). "In vitro analyses of the production and activity of secondary small interfering RNAs in C. elegans." EMBO J **26**(24): 5007-5019.

Armache, K. J., H. Kettenberger, et al. (2003). "Architecture of initiation-competent 12-subunit RNA polymerase II." Proc Natl Acad Sci U S A **100**(12): 6964-6968.

Bahler, J., J. Q. Wu, et al. (1998). "Heterologous modules for efficient and versatile PCR-based gene targeting in Schizosaccharomyces pombe." Yeast **14**(10): 943-951.

Bannister, A. J., P. Zegerman, et al. (2001). "Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain." Nature **410**(6824): 120-124.

Bassett, A., S. Cooper, et al. (2009). "The folding and unfolding of eukaryotic chromatin." Curr Opin Genet Dev **19**(2): 159-165.

Bernstein, E., A. A. Caudy, et al. (2001). "Role for a bidentate ribonuclease in the initiation step of RNA interference." Nature **409**(6818): 363-366.

Birney, E., J. A. Stamatoyannopoulos, et al. (2007). "Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project." Nature **447**(7146): 799-816.

Birse, C. E., B. A. Lee, et al. (1997). "Transcriptional termination signals for RNA polymerase II in fission yeast." EMBO J **16**(12): 3633-3643.

Blencowe, B. J., S. Ahmad, et al. (2009). "Current-generation high-throughput sequencing: deepening insights into mammalian transcriptomes." Genes Dev **23**(12): 1379-1386.

Bourbonnais, Y., N. Faucher, et al. (2001). "Multiple cellular processes affected by the absence of the Rpb4 subunit of RNA polymerase II contribute to the deficiency in the stress response of the yeast rpb4(delta) mutant." Mol Gen Genet **264**(6): 763-772.

Brookes, E. and A. Pombo (2009). "Modifications of RNA polymerase II are pivotal in regulating gene expression states." EMBO Rep.

Buchanan, L., Durand-Dubief, M., Roguev, A., Sakalar, C., Wilhelm, B., Strålfors, A., Shevchenko. A., Aasland, R., Shevchenko, A., Ekwall, K., Stewart, F. (2009). "The S. pombe JmjC-protein, Msc1. prevents H2A.Z localization in centromeric and subtelomeric chromatin domains." Plos Biol in press.

Buhler, M., N. Spies, et al. (2008). "TRAMP-mediated RNA surveillance prevents spurious entry of RNAs into the Schizosaccharomyces pombe siRNA pathway." Nat Struct Mol Biol **15**(10): 1015-1023.

Buker, S. M., T. Iida, et al. (2007). "Two different Argonaute complexes are required for siRNA generation and heterochromatin assembly in fission yeast." Nat Struct Mol Biol **14**(3): 200-207.

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

Cairns, B. R. (2009). "The logic of chromatin architecture and remodelling at promoters." Nature **461**(7261): 193-198.

Camblong, J., N. Iglesias, et al. (2007). "Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in S. cerevisiae." Cell **131**(4): 706-717.

Carmichael, J. B., P. Provost, et al. (2004). "ago1 and dcr1, two core components of the RNA interference pathway, functionally diverge from rdp1 in regulating cell cycle events in Schizosaccharomyces pombe." Mol Biol Cell **15**(3): 1425-1435.

Carmichael, J. B., C. Stoica, et al. (2006). "RNA interference effector proteins localize to mobile cytoplasmic puncta in Schizosaccharomyces pombe." Traffic **7**(8): 1032-1044.

Carninci, P., T. Kasukawa, et al. (2005). "The transcriptional landscape of the mammalian genome." Science **309**(5740): 1559-1563.

Chen, B. S., S. S. Mandal, et al. (2004). "High-resolution protein-DNA contacts for the yeast RNA polymerase II general transcription machinery." Biochemistry **43**(40): 12741-12749.
Chen, C. Y., C. C. Chang, et al. (2009). "Mapping RNA exit channel on transcribing RNA polymerase II by FRET analysis." Proc Natl Acad Sci U S A **106**(1): 127-132.
Chen, D., W. M. Toone, et al. (2003). "Global transcriptional responses of fission yeast to environmental stress." Mol Biol Cell **14**(1): 214-229.
Chen, E. S., K. Zhang, et al. (2008). "Cell cycle control of centromeric repeat transcription and heterochromatin assembly." Nature **451**(7179): 734-737.
Choder, M. (2004). "Rpb4 and Rpb7: subunits of RNA polymerase II and beyond." Trends Biochem Sci **29**(12): 674-681.
Chung, W. H., J. L. Craighead, et al. (2003). "RNA polymerase II/TFIIF structure and conserved organization of the initiation complex." Mol Cell **12**(4): 1003-1013.
Cloonan, N. and S. M. Grimmond (2008). "Transcriptome content and dynamics at single-nucleotide resolution." Genome Biol **9**(9): 234.
Cojocaru, M., C. Jeronimo, et al. (2008). "Genomic location of the human RNA polymerase II general machinery: evidence for a role of TFIIF and Rpb7 at both early and late stages of transcription." Biochem J **409**(1): 139-147.
Core, L. J., J. J. Waterfall, et al. (2008). "Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters." Science **322**(5909): 1845-1848.
Cramer, P., D. A. Bushnell, et al. (2000). "Architecture of RNA polymerase II and implications for the transcription mechanism." Science **288**(5466): 640-649.
Darzacq, X., Y. Shav-Tal, et al. (2007). "In vivo dynamics of RNA polymerase II transcription." Nat Struct Mol Biol **14**(9): 796-806.
Davidson, M. K., H. K. Shandilya, et al. (2004). "Atf1-Pcr1-M26 complex links stress-activated MAPK and cAMP-dependent protein kinase pathways via chromatin remodeling of cgs2+." J Biol Chem **279**(49): 50857-50863.
Djupedal, I. and K. Ekwall (2009). "Epigenetics: heterochromatin meets RNAi." Cell Res **19**(3): 282-295.
Djupedal, I., M. Portoso, et al. (2005). "RNA Pol II subunit Rpb7 promotes centromeric transcription and RNAi-directed chromatin silencing." Genes Dev **19**(19): 2301-2306.
Drinnenberg, I. A., D. E. Weinberg, et al. (2009). "RNAi in Budding Yeast." Science.
Dutrow, N., D. A. Nix, et al. (2008). "Dynamic transcriptome of Schizosaccharomyces pombe shown by RNA-DNA hybrid mapping." Nat Genet **40**(8): 977-986.
Edwards, A. M., C. M. Kane, et al. (1991). "Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro." J Biol Chem **266**(1): 71-75.
Eissenberg, J. C. and A. Shilatifard (2006). "Leaving a mark: the many footprints of the elongating RNA polymerase II." Curr Opin Genet Dev **16**(2): 184-190.
Ekwall, K., G. Cranston, et al. (1999). "Fission yeast mutants that alleviate transcriptional silencing in centromeric flanking repeats and disrupt chromosome segregation." Genetics **153**(3): 1153-1169.
Ekwall, K., T. Olsson, et al. (1997). "Transient inhibition of histone deacetylation alters the structural and functional imprint at fission yeast centromeres." Cell **91**(7): 1021-1032.
Farnham, P. J. (2009). "Insights from genomic profiling of transcription factors." Nat Rev Genet **10**(9): 605-616.
Fire, A., S. Xu, et al. (1998). "Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans." Nature **391**(6669): 806-811.
Forstemann, K., M. D. Horwich, et al. (2007). "Drosophila microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1." Cell **130**(2): 287-297.
Fuda, N. J., M. B. Ardehali, et al. (2009). "Defining mechanisms that regulate RNA polymerase II transcription in vivo." Nature **461**(7261): 186-192.
Geiduschek, E. P. and M. Ouhammouch (2005). "Archaeal transcription and its regulators." Mol Microbiol **56**(6): 1397-1407.

Gobeil, L. A., P. Plante, et al. (2008). "Involvement of Dcr1 in post-transcriptional regulation of gene expression in Schizosaccharomyces pombe." Front Biosci **13**: 2203-2215.
Goler-Baron, V., M. Selitrennik, et al. (2008). "Transcription in the nucleus and mRNA decay in the cytoplasm are coupled processes." Genes Dev **22**(15): 2022-2027.
Gullerova, M. and N. J. Proudfoot (2008). "Cohesin complex promotes transcriptional termination between convergent genes in S. pombe." Cell **132**(6): 983-995.
Halic, M. and D. Moazed (2009). "Transposon silencing by piRNAs." Cell **138**(6): 1058-1060.
Hall, I. M., G. D. Shankaranarayana, et al. (2002). "Establishment and maintenance of a heterochromatin domain." Science **297**(5590): 2232-2237.
Hamilton, A. J. and D. C. Baulcombe (1999). "A species of small antisense RNA in posttranscriptional gene silencing in plants." Science **286**(5441): 950-952.
Hansen, K. R., G. Burns, et al. (2005). "Global effects on gene expression in fission yeast by silencing and RNA interference machineries." Mol Cell Biol **25**(2): 590-601.
He, Y., B. Vogelstein, et al. (2008). "The antisense transcriptomes of human cells." Science **322**(5909): 1855-1857.
Heard, E. and C. M. Disteche (2006). "Dosage compensation in mammals: fine-tuning the expression of the X chromosome." Genes Dev **20**(14): 1848-1867.
Hirota, K., T. Hasemi, et al. (2004). "Fission yeast global repressors regulate the specificity of chromatin alteration in response to distinct environmental stresses." Nucleic Acids Res **32**(2): 855-862.
Hirota, K., T. Miyoshi, et al. (2008). "Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs." Nature **456**(7218): 130-134.
Hoiby, T., H. Zhou, et al. (2007). "A facelift for the general transcription factor TFIIA." Biochim Biophys Acta **1769**(7-8): 429-436.
Holstege, F. C., E. G. Jennings, et al. (1998). "Dissecting the regulatory circuitry of a eukaryotic genome." Cell **95**(5): 717-728.
Houseley, J., J. LaCava, et al. (2006). "RNA-quality control by the exosome." Nat Rev Mol Cell Biol **7**(7): 529-539.
Iida, T., R. Kawaguchi, et al. (2006). "Conserved ribonuclease, Eri1, negatively regulates heterochromatin assembly in fission yeast." Curr Biol **16**(14): 1459-1464.
Iida, T., J. Nakayama, et al. (2008). "siRNA-mediated heterochromatin establishment requires HP1 and is associated with antisense transcription." Mol Cell **31**(2): 178-189.
Jackson, R. J. and N. Standart (2007). "How do microRNAs regulate gene expression?" Sci STKE **2007**(367): re1.
Janowski, B. A., S. T. Younger, et al. (2007). "Activating gene expression in mammalian cells with promoter-targeted duplex RNAs." Nat Chem Biol **3**(3): 166-173.
Jasiak, A. J., H. Hartmann, et al. (2008). "Genome-associated RNA polymerase II includes the dissociable Rpb4/7 subcomplex." J Biol Chem **283**(39): 26423-26427.
Jia, S., K. Noma, et al. (2004). "RNAi-independent heterochromatin nucleation by the stress-activated ATF/CREB family proteins." Science **304**(5679): 1971-1976.
Jinek, M. and J. A. Doudna (2009). "A three-dimensional view of the molecular machinery of RNA interference." Nature **457**(7228): 405-412.
Juven-Gershon, T., J. Y. Hsu, et al. (2008). "The RNA polymerase II core promoter - the gateway to transcription." Curr Opin Cell Biol **20**(3): 253-259.
Kettenberger, H., K. J. Armache, et al. (2003). "Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage." Cell **114**(3): 347-357.
Kim, D. H., L. M. Villeneuve, et al. (2006). "Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells." Nat Struct Mol Biol **13**(9): 793-797.
Kim, H. S., E. S. Choi, et al. (2004). "Regulation of Swi6/HP1-dependent heterochromatin assembly by cooperation of components of the mitogen-activated protein kinase pathway and a histone deacetylase Clr6." J Biol Chem **279**(41): 42850-42859.

Kimura, M., H. Sakurai, et al. (2001). "Intracellular contents and assembly states of all 12 subunits of the RNA polymerase II in the fission yeast Schizosaccharomyces pombe." Eur J Biochem **268**(3): 612-619.
Kimura, M., H. Suzuki, et al. (2002). "Formation of a carboxy-terminal domain phosphatase (Fcp1)/TFIIF/RNA polymerase II (pol II) complex in Schizosaccharomyces pombe involves direct interaction between Fcp1 and the Rpb4 subunit of pol II." Mol Cell Biol **22**(5): 1577-1588.
Kloc, A., M. Zaratiegui, et al. (2008). "RNA interference guides histone modification during the S phase of chromosomal replication." Curr Biol **18**(7): 490-495.
Kobori, T., T. Yoshino, et al. (2003). "Hierarchical chromatin structure of Schizosaccharomyces pombe revealed by atomic force microscopy." Curr Microbiol **47**(5): 404-407.
Kostrewa, D., M. E. Zeller, et al. (2009). "RNA polymerase II-TFIIB structure and mechanism of transcription initiation." Nature.
Kurdistani, S. K., D. Robyr, et al. (2002). "Genome-wide binding map of the histone deacetylase Rpd3 in yeast." Nat Genet **31**(3): 248-254.
Kurdistani, S. K., S. Tavazoie, et al. (2004). "Mapping global histone acetylation patterns to gene expression." Cell **117**(6): 721-733.
Landick, R. (2009). "Functional divergence in the growing family of RNA polymerases." Structure **17**(3): 323-325.
Lantermann, A., Straub, T, Strålfors, A, Yuan, GC, Ekwall, K and Korber, P (Nature Structural and Molecular Biology, in press). "Evolutionary Plasticity of Nucleosome Positioning Mechanisms in Yeasts."
Lantermann, A., Straub, T, Strålfors, A, Yuan, GC, Ekwall, K and Korber, P (Submitted manuscript). "Evolutionary Plasticity of Nucleosome Positioning Mechanisms in Yeasts."
Lee, H. C., S. S. Chang, et al. (2009). "qiRNA is a new type of small interfering RNA induced by DNA damage." Nature **459**(7244): 274-277.
Li, L. C., S. T. Okino, et al. (2006). "Small dsRNAs induce transcriptional activation in human cells." Proc Natl Acad Sci U S A **103**(46): 17337-17342.
Lipardi, C. and B. M. Paterson (2009). "Identification of an RNA-dependent RNA polymerase in Drosophila involved in RNAi and transposon suppression." Proc Natl Acad Sci U S A **106**(37): 15645-15650.
Lotan, R., V. G. Bar-On, et al. (2005). "The RNA polymerase II subunit Rpb4p mediates decay of a specific class of mRNAs." Genes Dev **19**(24): 3004-3016.
Lotan, R., V. Goler-Baron, et al. (2007). "The Rpb7p subunit of yeast RNA polymerase II plays roles in the two major cytoplasmic mRNA decay mechanisms." J Cell Biol **178**(7): 1133-1143.
Luger, K. and J. C. Hansen (2005). "Nucleosome and chromatin fiber dynamics." Curr Opin Struct Biol **15**(2): 188-196.
MacRae, I. J., K. Zhou, et al. (2007). "Structural determinants of RNA recognition and cleavage by Dicer." Nat Struct Mol Biol **14**(10): 934-940.
Makeyev, E. V. and D. H. Bamford (2002). "Cellular RNA-dependent RNA polymerase involved in posttranscriptional gene silencing has two distinct activity modes." Mol Cell **10**(6): 1417-1427.
Martin, S. and A. Pombo (2003). "Transcription factories: quantitative studies of nanostructures in the mammalian nucleus." Chromosome Res **11**(5): 461-470.
Mehta, S., I. Miklos, et al. (2009). "The Med8 mediator subunit interacts with the Rpb4 subunit of RNA polymerase II and Ace2 transcriptional activator in Schizosaccharomyces pombe." FEBS Lett **583**(19): 3115-3120.
Meinhart, A., T. Kamenski, et al. (2005). "A structural perspective of CTD function." Genes Dev **19**(12): 1401-1415.
Meins, F., Jr., A. Si-Ammour, et al. (2005). "RNA silencing systems and their relevance to plant development." Annu Rev Cell Dev Biol **21**: 297-318.
Meka, H., F. Werner, et al. (2005). "Crystal structure and RNA binding of the Rpb4/Rpb7 subunits of human RNA polymerase II." Nucleic Acids Res **33**(19): 6435-6444.
Mi, S., T. Cai, et al. (2008). "Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide." Cell **133**(1): 116-127.

Mitsuzawa, H., E. Kanda, et al. (2003). "Rpb7 subunit of RNA polymerase II interacts with an RNA-binding protein involved in processing of transcripts." Nucleic Acids Res **31**(16): 4696-4701.
Mitsuzawa, H., M. Kimura, et al. (2005). "Glyceraldehyde-3-phosphate dehydrogenase and actin associate with RNA polymerase II and interact with its Rpb7 subunit." FEBS Lett **579**(1): 48-52.
Miyao, T., J. D. Barnett, et al. (2001). "Deletion of the RNA polymerase subunit RPB4 acts as a global, not stress-specific, shut-off switch for RNA polymerase II transcription at high temperatures." J Biol Chem **276**(49): 46408-46413.
Mochizuki, K., N. A. Fine, et al. (2002). "Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena." Cell **110**(6): 689-699.
Morris, K. V., S. W. Chan, et al. (2004). "Small interfering RNA-induced transcriptional gene silencing in human cells." Science **305**(5688): 1289-1292.
Morris, K. V., S. Santoso, et al. (2008). "Bidirectional transcription directs both transcriptional gene activation and suppression in human cells." PLoS Genet **4**(11): e1000258.
Motamedi, M. R., A. Verdel, et al. (2004). "Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs." Cell **119**(6): 789-802.
Na, X., H. O. Duan, et al. (2003). "Identification of the RNA polymerase II subunit hsRPB7 as a novel target of the von Hippel-Lindau protein." EMBO J **22**(16): 4249-4259.
Naji, S., S. Grunberg, et al. (2007). "The RPB7 orthologue E' is required for transcriptional activity of a reconstituted archaeal core enzyme at low temperatures and stimulates open complex formation." J Biol Chem **282**(15): 11047-11057.
Pak, J. and A. Fire (2007). "Distinct populations of primary and secondary effectors during RNAi in C. elegans." Science **315**(5809): 241-244.
Paule, M. R. and R. J. White (2000). "Survey and summary: transcription by RNA polymerases I and III." Nucleic Acids Res **28**(6): 1283-1298.
Pauler, F. M., M. V. Koerner, et al. (2007). "Silencing by imprinted noncoding RNAs: is transcription the answer?" Trends Genet **23**(6): 284-292.
Peng, J. C. and G. H. Karpen (2007). "H3K9 methylation and RNA interference regulate nucleolar organization and repeated DNA stability." Nat Cell Biol **9**(1): 25-35.
Perbal, B. (1999). "Nuclear localisation of NOVH protein: a potential role for NOV in the regulation of gene expression." Mol Pathol **52**(2): 84-91.
Perocchi, F., Z. Xu, et al. (2007). "Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D." Nucleic Acids Res **35**(19): e128.
Peterlin, B. M. and D. H. Price (2006). "Controlling the elongation phase of transcription with P-TEFb." Mol Cell **23**(3): 297-305.
Petermann, R., B. M. Mossier, et al. (1998). "Oncogenic EWS-Fli1 interacts with hsRPB7, a subunit of human RNA polymerase II." Oncogene **17**(5): 603-610.
Pillai, B., J. Verma, et al. (2003). "Whole genome expression profiles of yeast RNA polymerase II core subunit, Rpb4, in stress and nonstress conditions." J Biol Chem **278**(5): 3339-3346.
Preker, P., J. Nielsen, et al. (2008). "RNA exosome depletion reveals transcription upstream of active human promoters." Science **322**(5909): 1851-1854.
Provost, P., R. A. Silverstein, et al. (2002). "Dicer is required for chromosome segregation and gene silencing in fission yeast cells." Proc Natl Acad Sci U S A **99**(26): 16648-16653.
Radman-Livaja, M. and O. J. Rando (2009). "Nucleosome positioning: How is it established, and why does it matter?" Dev Biol.
Ramachandran, V. and X. Chen (2008). "Degradation of microRNAs by a family of exoribonucleases in Arabidopsis." Science **321**(5895): 1490-1492.
Rana, T. M. (2007). "Illuminating the silence: understanding the structure and function of small RNAs." Nat Rev Mol Cell Biol **8**(1): 23-36.
Raponi, M. and G. M. Arndt (2003). "Double-stranded RNA-mediated gene silencing in fission yeast." Nucleic Acids Res **31**(15): 4481-4489.
Reese, J. C. (2003). "Basal transcription factors." Curr Opin Genet Dev **13**(2): 114-118.

Robertson, H. D., R. E. Webster, et al. (1967). "A nuclease specific for double-stranded RNA." Virology **32**(4): 718-719.
Robyr, D., Y. Suka, et al. (2002). "Microarray deacetylation maps determine genome-wide functions for yeast histone deacetylases." Cell **109**(4): 437-446.
Royo, H. and J. Cavaille (2008). "Non-coding RNAs in imprinted gene clusters." Biol Cell **100**(3): 149-166.
Runner, V. M., V. Podolny, et al. (2008). "The Rpb4 subunit of RNA polymerase II contributes to cotranscriptional recruitment of 3' processing factors." Mol Cell Biol **28**(6): 1883-1891.
Sakurai, H., H. Mitsuzawa, et al. (1999). "The Rpb4 subunit of fission yeast Schizosaccharomyces pombe RNA polymerase II is essential for cell viability and similar in structure to the corresponding subunits of higher eukaryotes." Mol Cell Biol **19**(11): 7511-7518.
Schwarz, D. S., G. Hutvagner, et al. (2003). "Asymmetry in the assembly of the RNAi enzyme complex." Cell **115**(2): 199-208.
Seila, A. C., J. M. Calabrese, et al. (2008). "Divergent transcription from active promoters." Science **322**(5909): 1849-1851.
Selitrennik, M., L. Duek, et al. (2006). "Nucleocytoplasmic shuttling of the Rpb4p and Rpb7p subunits of Saccharomyces cerevisiae RNA polymerase II by two pathways." Eukaryot Cell **5**(12): 2092-2103.
Sharma, N., S. Marguerat, et al. (2006). "The fission yeast Rpb4 subunit of RNA polymerase II plays a specialized role in cell separation." Mol Genet Genomics **276**(6): 545-554.
Shen, X. Q., A. Bubulya, et al. (1999). "Ligand-free RAR can interact with the RNA polymerase II subunit hsRPB7 and repress transcription." Endocrine **10**(3): 281-289.
Sigova, A., N. Rhind, et al. (2004). "A single Argonaute protein mediates both transcriptional and posttranscriptional silencing in Schizosaccharomyces pombe." Genes Dev **18**(19): 2359-2367.
Sijen, T., F. A. Steiner, et al. (2007). "Secondary siRNAs result from unprimed RNA synthesis and form a distinct class." Science **315**(5809): 244-247.
Silva, M. C. and L. E. Jansen (2009). "At the right place at the right time: novel CENP-A binding proteins shed light on centromere assembly." Chromosoma **118**(5): 567-574.
Singh, J. and R. A. Padgett (2009). "Rates of in situ transcription and splicing in large human genes." Nat Struct Mol Biol.
Spahr, H., G. Calero, et al. (2009). "Schizosacharomyces pombe RNA polymerase II at 3.6-A resolution." Proc Natl Acad Sci U S A **106**(23): 9185-9190.
Sugiyama, T., H. Cam, et al. (2005). "RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production." Proc Natl Acad Sci U S A **102**(1): 152-157.
Tan, Q., X. Li, et al. (2000). "Multiple mechanisms of suppression circumvent transcription defects in an RNA polymerase mutant." Mol Cell Biol **20**(21): 8124-8133.
Taverna, S. D., R. S. Coyne, et al. (2002). "Methylation of histone h3 at lysine 9 targets programmed DNA elimination in tetrahymena." Cell **110**(6): 701-711.
Todone, F., P. Brick, et al. (2001). "Structure of an archaeal homolog of the eukaryotic RNA polymerase II RPB4/RPB7 complex." Mol Cell **8**(5): 1137-1143.
Ujvari, A. and D. S. Luse (2006). "RNA emerging from the active site of RNA polymerase II interacts with the Rpb7 subunit." Nat Struct Mol Biol **13**(1): 49-54.
van der Krol, A. R., L. A. Mur, et al. (1990). "Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression." Plant Cell **2**(4): 291-299.
van Werven, F. J., H. A. van Teeffelen, et al. (2009). "Distinct promoter dynamics of the basal transcription factor TBP across the yeast genome." Nat Struct Mol Biol **16**(10): 1043-1048.
Vastenhouw, N. L., K. Brunschwig, et al. (2006). "Gene expression: long-term gene silencing by RNAi." Nature **442**(7105): 882.

Vasudevan, S., Y. Tong, et al. (2007). "Switching from repression to activation: microRNAs can up-regulate translation." Science **318**(5858): 1931-1934.
Verdel, A., S. Jia, et al. (2004). "RNAi-mediated targeting of heterochromatin by the RITS complex." Science **303**(5658): 672-676.
Verma-Gaur, J., S. N. Rao, et al. (2008). "Genomewide recruitment analysis of Rpb4, a subunit of polymerase II in Saccharomyces cerevisiae, reveals its involvement in transcription elongation." Eukaryot Cell **7**(6): 1009-1018.
Volpe, T. A., C. Kidner, et al. (2002). "Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi." Science **297**(5588): 1833-1837.
Werner, F. (2008). "Structural evolution of multisubunit RNA polymerases." Trends Microbiol **16**(6): 247-250.
Westover, K. D., D. A. Bushnell, et al. (2004). "Structural basis of transcription: separation of RNA from DNA by RNA polymerase II." Science **303**(5660): 1014-1016.
Wilhelm, B. T., S. Marguerat, et al. (2008). "Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution." Nature **453**(7199): 1239-1243.
Wiren, M., R. A. Silverstein, et al. (2005). "Genomewide analysis of nucleosome density histone acetylation and HDAC function in fission yeast." EMBO J **24**(16): 2906-2918.
Wood, V., R. Gwilliam, et al. (2002). "The genome sequence of Schizosaccharomyces pombe." Nature **415**(6874): 871-880.
Wooddell, C. I. and R. R. Burgess (2000). "Topology of yeast RNA polymerase II subunits in transcription elongation complexes studied by photoaffinity cross-linking." Biochemistry **39**(44): 13405-13421.
Woychik, N. A. and M. Hampsey (2002). "The RNA polymerase II machinery: structure illuminates function." Cell **108**(4): 453-463.
Wu, X., C. B. Wilcox, et al. (2000). "The Ess1 prolyl isomerase is linked to chromatin remodeling complexes and the general transcription machinery." Embo J **19**(14): 3727-3738.
Xue, Y., S. A. Haas, et al. (2004). "A DNA microarray for fission yeast: minimal changes in global gene expression after temperature shift." Yeast **21**(1): 25-39.
Yasuhara, J. C. and B. T. Wakimoto (2006). "Oxymoron no more: the expanding world of heterochromatic genes." Trends Genet **22**(6): 330-338.
Zhang, K., K. Mosch, et al. (2008). "Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin." Nat Struct Mol Biol **15**(4): 381-388.
Zhang, Y., Z. Moqtaderi, et al. (2009). "Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo." Nat Struct Mol Biol **16**(8): 847-852.
Zippo, A., R. Serafini, et al. (2009). "Histone crosstalk between H3S10ph and H4K16ac generates a histone code that mediates transcription elongation." Cell **138**(6): 1122-1136.
Zlatanova, J., T. C. Bishop, et al. (2009). "The nucleosome family: dynamic and growing." Structure **17**(2): 160-171.
Zofall, M., T. Fischer, et al. (2009). "Histone H2A.Z cooperates with RNAi and heterochromatin factors to suppress antisense RNAs." Nature **461**(7262): 419-422.