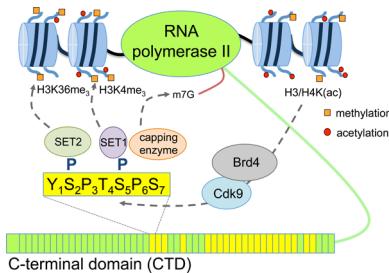


## The RNA Polymerase II Carboxy-Terminal Domain (CTD) Code

Dirk Eick\*,† and Matthias Geyer\*,‡

†Department of Molecular Epigenetics, Helmholtz Center Munich and Center for Integrated Protein Science Munich (CIPSM), Marchioninistrasse 25, 81377 Munich, Germany

‡Center of Advanced European Studies and Research, Group Physical Biochemistry, Ludwig-Erhard-Allee 2, 53175 Bonn, Germany



### CONTENTS

1. Introduction	8456
2. Structure of the CTD	8457
2.1. The CTD Is Composed of Heptad-Repeats	8457
2.2. Functional Units of the CTD	8460
2.3. Properties of the Consensus Repeat	8462
2.4. Post-Translational Modifications of the CTD	8462
2.4.1. CTD Serine Kinases	8463
2.4.2. Other CTD Kinases	8464
2.4.3. CTD Phosphatases	8464
2.4.4. Mechanisms of Template Phosphorylation and Dephosphorylation	8465
2.4.5. Other CTD Modifying Enzymes	8465
2.5. CTD Binding Factors	8466
2.5.1. General Transcription Factors	8466
2.5.2. Histone and Chromatin Modifying Factors	8466
2.5.3. RNA Processing Factors	8466
2.5.4. Retroacting CTD Modifying Factors	8468
2.6. Characteristics of CTD Binding Domains	8468
3. Functional Analysis of CTD Modifications	8470
3.1. Chromatin Immunoprecipitation (ChIP) of RNAPII	8470
3.2. Monoclonal Antibodies for the Analysis of Specific CTD Modifications	8471
3.3. Dynamic Phosphorylation Patterns of the CTD in the Transcription Cycle	8471
4. CTD-Regulated Cellular Processes	8472
4.1. Transcription Initiation	8472
4.2. Promoter Proximal Pausing of RNAPII	8473
4.3. mRNA Capping	8473
4.4. Elongation	8474
4.5. Cotranscriptional Splicing	8474
4.6. 3' Processing of RNA	8475
4.7. Transcription Termination	8476
5. CTD-Chromatin Crosstalk	8477
5.1. Modification of Chromatin and the CTD Is Mutually Controlled	8477
5.2. Transcription through Chromatin	8478
5.3. Epigenetic Regulation of Alternative Pre-mRNA Splicing	8478
6. Acceleration of Evolution in Eukaryotes by the CTD	8479
6.1. Gain and Loss of Introns	8479
6.2. Multifunctional Properties of Introns	8480
7. Conclusions and Future Perspectives	8483
7.1. How Can the CTD Code Be Deciphered?	8483
Author Information	8484
Corresponding Author	8484
Notes	8484
Biographies	8484
Acknowledgments	8484
Abbreviations	8484
References	8484

### 1. INTRODUCTION

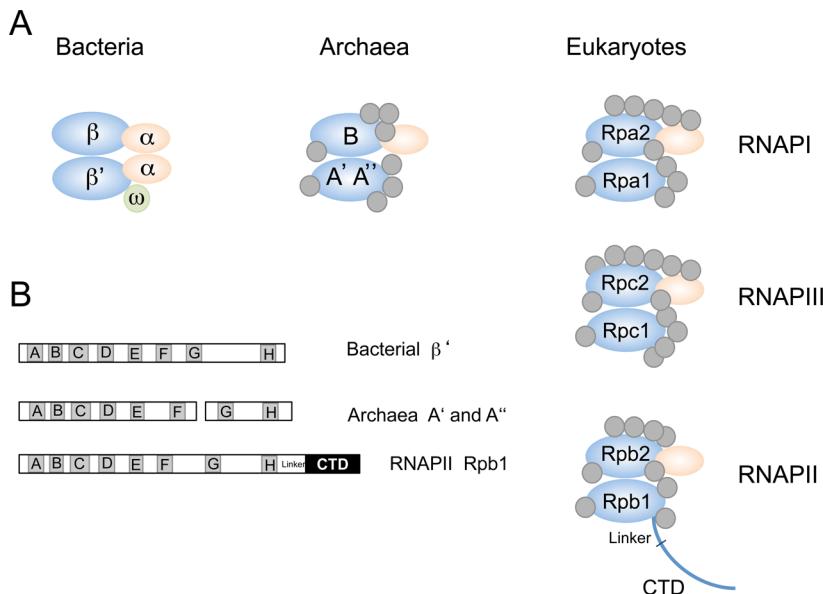
The smallest viable unit of life is the cell. From bacteria to mammals, all cells use the same nucleic acid-based universal code for the maintenance and inheritance of genetic information. All life on earth probably started with a common ancestral cell<sup>1</sup> approximately 4 billion years ago. Today, a huge diversity of organisms live on Earth, many of them still at the evolutionary stage of unicellular organisms, while others are life forms of high complexity with large cell numbers. The development of this diversity was driven by evolution, a process assuring survival and adaption of life to new environmental challenges. While the first evolutionary processes were probably very simple genetic changes, these processes gradually became more sophisticated with the development of higher eukaryotes. In this Review, we will focus on a molecular structure that first evolved approximately 1 billion years ago, but later revolutionized and accelerated evolutionary processes in higher eukaryotes and, thereby, strongly expedited their development. This molecular structure is placed at the carboxy-terminus of the large subunit (Rpb1) of RNA polymerase II (RNAPII) and will be termed the CTD (C-terminal domain) throughout this Review.

RNA polymerases developed early in evolution, and their core structure is conserved from bacteria to mammals (Figure 1). The main task of these enzymes is the transcription of the genetic information of DNA (deoxy-ribonucleic acid) into RNA (ribonucleic acid). In bacteria and many protozoa (unicellular eukaryotes), this is still the main function of RNA polymerases.

**Special Issue:** 2013 Gene Expression

**Received:** February 2, 2013

**Published:** August 16, 2013



**Figure 1.** Multisubunit RNA polymerases in prokaryotes, archaea, and eukaryotes. (A) The two largest subunits,  $\beta$  and  $\beta'$ , of *E. coli* RNA polymerase are highly conserved to the subunits,  $A'A''$  and  $B$ , in archaea and to the two largest subunits in RNAPI, -II, and -III of eukaryotes. Only RNAPII has evolved a CTD. Gray circles and orange ovals represent additional subunits. (B) Regions (A–H) are conserved between the largest subunit in bacteria, archaea, and RNAPII.

However, the evolution of eukaryotes was accompanied by the development of specialized forms of RNA polymerases, RNAPI, RNAPII, and RNAPIII. The activity of RNAPI is restricted to the nucleolus, a large structure in the nucleus of eukaryotic cells accommodating all ribosomal genes. RNAPI transcribes rDNA genes into rRNA and contributes >50% to the transcriptional activity of proliferating cells. The function of RNAPIII is restricted to the transcription of small RNAs, mainly tRNAs. The remaining major part of the genome, including all coding genes, is transcribed by RNAPII. It was first assumed that RNAPII transcription might be limited to genomic loci that give rise to messenger RNA (mRNA) or other stable transcripts, but recent studies have shown a genome wide presence of RNAPII, including at genomic loci with noncoding sequences and at loci that do not produce stable transcripts.<sup>2</sup> This observation posed the question: Is RNAPII involved in additional cellular processes beyond the production of RNAs?

When splicing was discovered for adenoviral mRNAs in 1977,<sup>3</sup> it was not foreseen to be a widespread phenomenon of all cellular mRNAs in higher eukaryotes. Splicing removes so-called intervening sequences (introns) from the primary transcript and leaves the protein coding sequence (exons). Most mRNAs in higher eukaryotes, therefore, represent macromolecules composed of short RNA fragments, which are assembled post- or cotranscriptionally by splicing of a precursor RNA.

In this Review, we will highlight the important features of the CTD. Functions of the CTD are clearly not limited to splicing of mRNA, even though splicing is its most eye-catching function, and this Review will have a specific focus on splicing and the function of introns. The CTD apparently coordinates the entire transcription cycle of RNAPII from initiation, promoter proximal pausing, the elongation phase, choice of exons for splicing, selection of RNA 3' processing sites, and possibly the termination of the transcription process. We will review the function of the CTD in these different processes and discuss the question of whether the extensive modification of

the CTD during the transcription cycle underlies specific rules and displays characteristics of a code. In addition to involvement in the coordination of transcriptional and cotranscriptional RNA maturation processes, the CTD is also likely to be involved in the epigenetic programming of cells and contributes to the writing, reading, and erasing of epigenetic marks in chromatin. Thus, the role of the CTD is multifaceted and extends into many nuclear processes. Yet before we review the CTD's multiple functions, we describe the similarities and differences of CTD structures in various organisms, all of the post translational modifications that have been ascribed to the CTD so far, and discuss how these modifications contribute to the function of the CTD. We would also like to emphasize that many excellent reviews have been published in recent years, giving either a general overview of the CTD,<sup>4</sup> or a focus on CTD structure and modifications,<sup>5</sup> CTD evolution,<sup>6</sup> or CTD function in splicing and mRNA processing.<sup>7</sup>

## 2. STRUCTURE OF THE CTD

Comparison of genes for the largest subunit of RNAPII (Rpb1) and RNAPIII (Rpc1) of the eukaryote *Saccharomyces cerevisiae* with sequences of prokaryotic RNA polymerases revealed that all RNA polymerases have evolved from a common ancestral polymerase.<sup>8</sup> The sequences of Rpb1 and Rpc1 are homologous to the sequence of the largest subunit ( $\beta'$ ) of the bacterial *Escherichia coli* RNA polymerase and display eight conserved domains (A–H) with an identical alignment (Figure 1,B). In addition, Rpb1 carries an unusual heptapeptide sequence tandemly repeated at its carboxy-terminus (the CTD) that is not present in the large subunits of RNAPI and RNAPIII. A repeat-structure similar to that reported for *S. cerevisiae* was described for mammalian Rpb1.<sup>9</sup>

### 2.1. The CTD Is Composed of Heptad-Repeats

In yeast and mammals, a single repeat in the CTD consists of a block of seven amino acids with the consensus sequence tyrosine-serine-proline-threonine-serine-proline-serine. Throughout this Review, we will refer to this sequence as the



**Figure 2.** Comparison of carboxy-terminal domains (CTDs) in higher eukaryotes. The CTDs of mammals (*Homo sapiens*) and zebrafish (*Danio rerio*), both vertebrates, have 52 repeats and differ only marginally at five positions (marked by red arrows). The CTDs of Lancelet (*Branchiostoma floridae*), an invertebrate, and of the plant, *Arabidopsis thaliana*, are shorter and display larger differences. Divergence from the consensus heptad-repeat sequence is highlighted in blue. An insertion of 22 unrelated amino acids in Lancelet is highlighted by red brackets. GenBank accession numbers for protein sequences: *H. sapiens*, NP\_000928.1; *D. rerio*, XP\_682682.1; *B. floridae*, XP\_002593218.1; *A. thaliana*, NP\_195505.2.

consensus repeat and number the position of amino acids from one to seven (Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7). While the sequence of the repeat is conserved in *S. cerevisiae* and mammals, the length of the CTD differs markedly, with 26 and 52 repeats, respectively (Figures 2 and 3). Strikingly, the repeats in *S. cerevisiae* almost all follow the consensus sequence (Figure 3), while the mammalian CTD shows significant deviations from the consensus in the distal part of the CTD (Figure 2). The sequences of more than 100 CTDs from

different organisms have now been determined. We will consider them in detail to discuss and answer questions regarding the fine structure of the CTD. However, it is important to note that yeast (*S. cerevisiae* and *S. pombe*) and mammals are the model organisms in which almost all CTD-related research has been performed. Therefore, some of the conclusions drawn from the structures of other organisms are still speculative and are currently not experimentally substantiated.

S. pombe	S. cerevisiae	B. anathema	A. locustae
YGLTSPS	FSPTSP	YSPASPA	YSPTSPA
YSPSSPG	YSPTSPA	YSPASPA	YSPTSPA
YSTSPA	YSPTSPS	YSPASPA	YSPTSPA
YMPSSPS	YSPTSPS	YSPASPA	YSPTSPA
YSPTSPS	YSPTSPS	YSPASPA	YSPTSPA
YSPTSPS	YSPTSPS	YSPASPA	YSPTSPA
YSPTSPS	YSPTSPS	YSPASPA	YSPTSPA
YSPTSPS	YSPTSPS	YSPASPA	YSPTSPA
YSATSPS	YSPTSPS	YSPASPA	YSPTSPA
YSPTSPS	YSPTSPS	YSPASPA	YSPTSPA
YSPTSPS	YSPTSPS	YSPASPA	YSPTSPA
YSPTSPS	YSPTSPS	YSPASPA	YSPTSPA
YSPTSPS	YSPTSPS	YSPASPA	YSPTSPV
YSPTSPS	YSPTSPS	YSPASPA	YRPSTRAKDQDQ
YSPTSPS	YSPTSPS	YSPASPA	GFNERLKKRRKS
YSPTSPS	YSPTSPA	YSPASPA	
YSPTSPS	YSPTSPS	YSPASPA	
YSPTSPS	YSPTSPS	YSPASPA	
YSPTSPS	YSPTSPS	YSPASPA	
YSPTSPS	YSPTSPN	YGPASPA	
YSPTSPS	YSPTSPS	YSPASPA	
YSPTSPS	YSPTSPG	YSPASPA	
YSPTSPS	YSPGSPA	YSPASPA	
YSPTSPS	YSPKQDEQKHNEENENSR	YDPAEPYDRR	
YSPTSPS			
YSPTSPS			
YSPTSPS			

**Figure 3.** Comparison of CTDs in unicellular eukaryotes. Examples of CTDs rigorously obeying the heptad-repeat rule with exact spacing of SP motifs (yellow). Divergence from the consensus-repeat is highlighted (blue). GenBank accession numbers for protein sequences: *Saccharomyces cerevisiae*, NP\_010141.1; *Schizosaccharomyces pombe*, NP\_595673.1; *Breviata anathema*, AAC62246; *Antonospora locustae*, AAD12605.

T. annulata	C. immitis	A. nidulans
01 FSPTSP	01 FSPAGARSPGG	01 FSPAATSPAG
MSPTSP	YSPTSPFN	YSPSSPFSANPTSPG
MSPTSP	TSPTSPG	YSPTSS
MSPTSAISPV	YSPSSG	YSPTSPGMGI
YSPA	YSPTSPGMSI	TSPRFMTSPG
YSPTSP	TSPTYMTSPG	FSPASPS
MSPTSPDSA	FSPASPT	FAPTSPA
LSPTSPV	FAPTSPTA	YSPTSPAYGQ
YSPA	YSPTSPGYGQA	ASPTSPS
10 YSPTSPDSA	10 QSPTSPS	10 YSPTSPG
MSPTSPV	YSPTSPG	FSPTSPN
YSPA	FSPTSPS	YSPTSPS
YSPTSPNLG	YSPTSPS	FSPTSPA
YAPTSPV	FSPASPA	FSPTSPS
YSPA	FSPTSPS	YSPTSPAIGGARH
YSPTTPNYG	YSPTSPALGGATGRH	LSPTSPTSPK
YSPTSP	LSPTSPTSPK	YTPTSPG
LSPTSPA	YTPTSPG	WSPTSPQT
YSPTDP	WSPTSPAEA	YSPTSPNFA
20 FSPNPFDID	20 YSPTSPNFA	20 GSPTSPGGPTSPG
	GSPTSPGGPTSPG	YSPTSPA
	22 YSPTSPTFNPTSPRQ	22 FSPTSPRQ

**Figure 4.** Unicellular organisms with exact spacing of SP motifs but not obeying the strict heptad-repeat rule. GenBank accession numbers for protein sequences: *Theileria annulata*, XM\_949789; *Coccidioides immitis*, EAS32326; *Aspergillus nidulans*, XM\_653321.

Two simple questions are immediately apparent when the CTDs of *S. cerevisiae* and mammals are compared (Figures 2 and 3). Why does the length of CTDs differ among organisms, and is the length of the CTD critical for its function and for cell viability? The answer to these questions required genetic manipulation of the CTD. Surprisingly, the deletion of a limited number of repeats was tolerated in yeast and mammalian cells, without having a measurable effect on cell viability. Significant

growth defects were observed only when one-half or more than one-half of the repeats were removed from *S. cerevisiae*<sup>10</sup> or mammalian cells.<sup>11</sup> The minimum number of repeats required for cell survival in *S. cerevisiae* was 11,<sup>10</sup> while the minimum number needed for growth of mammalian cells has not been exactly determined.

CTD length differs significantly in various organisms. Short forms with 10 repeats are found in simple organisms, for

**Trypanosoma brucei**

LFDLVLNMAALQO  
AVPQAEA  
VAPGKDVNVYHSLGSTLQQNIQSSIA  
YRPRDHD  
ATPFVNNAASLFLRQGFGGGSS  
SAPVTA  
SAPYNPSTTYHGGRLEASAVHRSQAYS  
TSPALEYGGREASASQMYSVMSASA  
FNPVSTRMSSVAHSYSEYSEASSYHLQHS  
VAPTSMQA  
SLPRTDNNSMTMQGIGSV  
SVPYTPHAMSS  
AAPPSQVYASTEVGRSHSEDSRSQSAL  
YVPTLSPTTHAGYAIRG  
DEPSTHSDSNVMWREAGGGREQDEEDDLSTN  
YMPTAKTPQQ  
VAPPTAAEFGDEEEEQ

**Leishmania major**

LFDLLLDMSKLQH  
VVPLDKATEARTSNVYHTDAS  
VAPGSSTLQGSHG  
ELPPSTVHENSSLAGASS  
VYPRHERAGMYG  
GIPIEASEVQFSS  
ALPTVMRTALACLEHD  
RVPQOPPPLWCFHVLGVVGA  
AERESLDTELSSYHLQS  
VAPSTAYG  
YAPMAAS  
GMPMPGTQLSVGEgg  
SLPYPYEASNGERFGSLGGAASSRA  
SAPYDPT  
QQPQEFSQTEEQEEP

**Figure 5.** Unicellular organisms without CTDs. The sequences correspond to sequences downstream of domain H (see Figure 1B). GenBank accession numbers for protein sequences: *Trypanosoma brucei*, XP\_847569; *Leishmania major*, AF009163.

example, *A. locustae* (Figure 3), while higher eukaryotes have a tendency for longer repeat clusters (worms, insects, fish, mammals) (Figure 2). However, the polyp *Hydra magnipapillata*, a basal metazoan, has a CTD length of >50 repeats; therefore, a general relationship between complexity of an organism and the length of its CTD does not really exist.<sup>12</sup>

The first eye-catching peculiarity of the mammalian CTD is the strong periodicity of the repeats (Figure 2). Most repeats in the mammalian CTD obey the heptad-repeat structure; exceptions are seen only for repeats 49 and 52. This is surprising because many repeats in the distal part of the mammalian CTD diverge from the consensus repeat sequence, despite not breaking the heptad rule. The strong periodicity of the CTD has been discussed as a prerequisite for its function *in vivo*. It was suggested that the CTD might interact with a highly repetitive complementary structure in cells, and that interruption of the periodicity of the CTD would interfere with this interaction and abolish CTD function. The periodic structure of nucleic acids suggested them as possible CTD interactors. Although an interaction of the CTD with DNA or RNA should not be excluded and has even been reported for DNA,<sup>13</sup> we describe evidence (section 2.2) indicating that a strong periodicity of the heptad-repeats is not critical for CTD function.

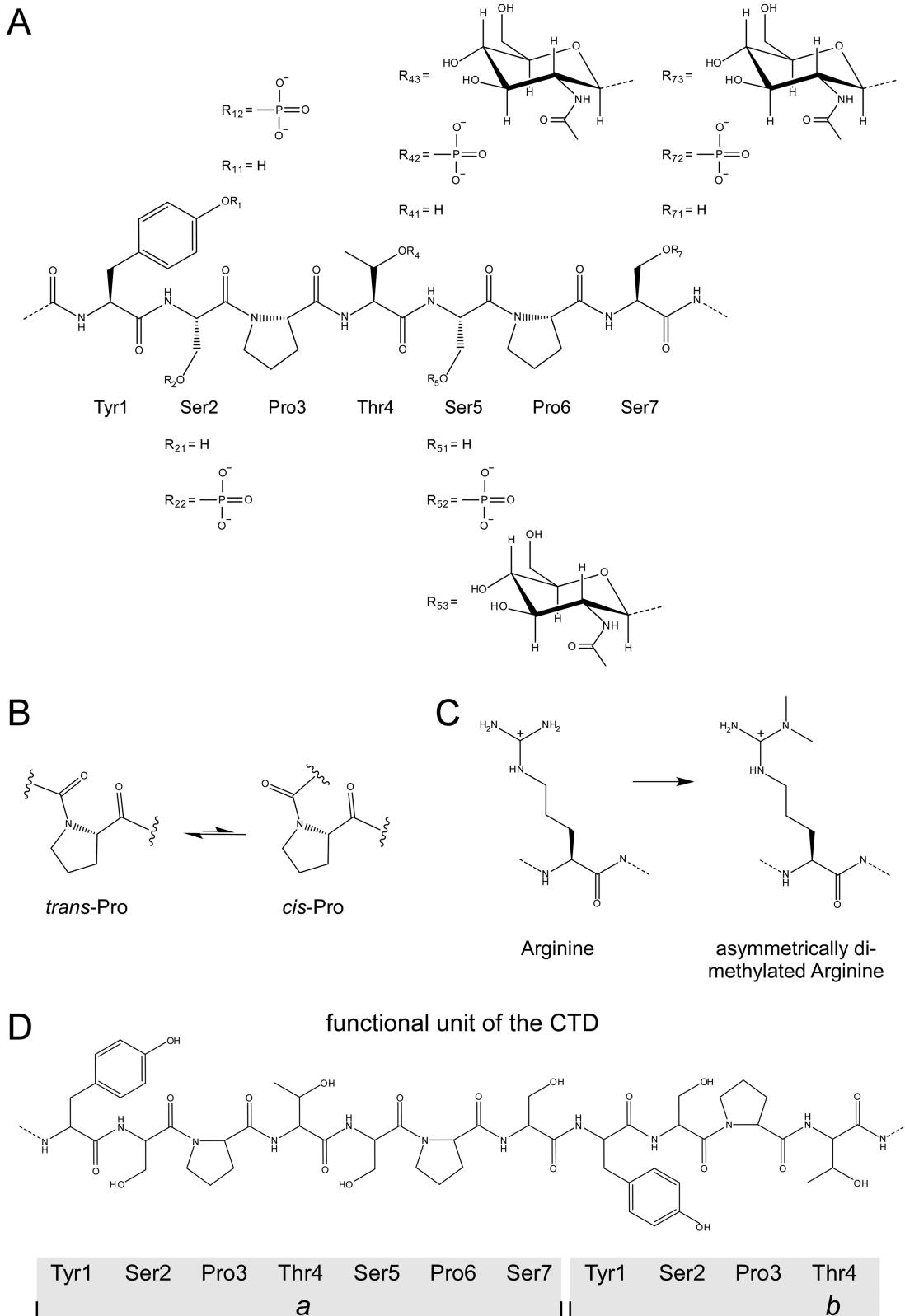
Analysis of different taxa showed that not all eukaryotes developed the typical CTD structure at the carboxy-terminus of Rpb1. Some taxa, including the Apicomplexa, *Plasmodium falciparum*, or the fungi, *Coccidioides immitis* and *Aspergillus nidulans*, clearly show beginnings of a repeat-structure, however, without reaching the strong periodicity for heptads or obeying the consensus sequence (Figure 4). Strikingly, the sequential arrangement of serine-proline (SP) residues at positions 2/3 and 5/6 and of threonine residues between already shows strong conservation and resembles the CTD heptad-structure in mammals. Tyrosine residues at position 1 are frequently replaced by phenylalanines or other amino acids, while the serine at position 7 can be substituted even by short arrays of additional amino acids. In contrast, a CTD or a sequence with similarity to the CTD is essentially absent in the carboxy-terminal region of Rpb1 of very distant taxa, such as *Leishmania*, *Trypanosoma*, *Trichomonas*, and *Giardia* (examples are shown in Figure 5). Comparison of CTDs from different

organisms shows that the spatial arrangement of the two SP motifs in a heptad-repeat is probably the best-conserved element, followed by the tyrosine at position 1. The threonine at position 4 and serine at position 7 show the highest divergence from the consensus in the various taxa. We conclude that different taxa have evolved distinct CTD structures. All vertebrates possess an almost identical CTD and diverge only marginally from each other (Figure 2, mammals and zebrafish). This high conservation is an indication that polymorphisms in the CTD within taxa are hardly tolerated and may cause negative selection rather than being an advantage. The high conservation can also be taken as a strong argument that each of the 52 repeats in the vertebrate CTD and, in particular all nonconsensus repeats, fulfill a specific task. The exact length requirement of the CTD has been tested experimentally in yeast. Remarkably, optimal CTD function and cell growth was achieved at wild-type CTD length rather than number of functional units (see also section 2.2), suggesting that the optimal length of CTD matters.<sup>14</sup> In the next section, we describe genetic experiments to define the minimal CTD functional unit in yeast strains, *S. cerevisiae* and *S. pombe*.

## 2.2. Functional Units of the CTD

Truncation experiments revealed that shortening the CTD within a limited range is compatible with cell viability and growth. Yet the role of each single amino acid within the heptad-repeat for cell growth and viability was unclear. The laboratories of Corden and Stiller studied a series of very informative CTD mutants to answer this question. The substitution of serines at positions two or five in the heptad-repeats of *S. cerevisiae* with alanine or glutamate,<sup>15</sup> or serine at position seven with glutamate<sup>16</sup> is lethal. The change of tyrosine at position one to phenylalanine is lethal in *S. cerevisiae*,<sup>15</sup> but not in *S. pombe*.<sup>17</sup> Thus, replacement of serine 2, 5, or 7 by an acidic mimic is not compatible with life, probably because the CTD cannot be transferred back to its hypo-phosphorylated state by phosphatases.<sup>16</sup> In contrast, the replacement of Thr4 and Ser7 by alanine is nonlethal in *S. cerevisiae*, albeit cells showed higher sensitivity to stress.<sup>18</sup>

The laboratories of Stiller and Shuman also studied the punctuation and syntax of CTD modifications in *S. cerevisiae*<sup>14,18,19</sup> and *S. pombe*,<sup>20</sup> respectively. The authors placed additional amino acids in the CTD and/or downstream



**Figure 6.** Post-translational modifications of the CTD. (A) Tyr1, Ser2, Thr4, Ser5, and Ser7 can be reversibly modified by phosphorylation. Ser5 and Ser7 have been additionally described to be *O*-glycosylated. (B) Pro3 and Pro6 stably exist in *cis*- or *trans*-conformation. (C) The unique Arg1810 is asymmetrically dimethylated at the guanidinium ion. (D) The “functional unit” of the CTD comprises one full repeat including the next four residues of the following repeat.

of each repeat or of every second heptad-repeat. The results were of great surprise. While mutants with an alanine after each repeat were not viable, mutants with additional alanine(s) after

every second repeat grew normally. The authors extended this analysis with more sophisticated mutants and also replaced single amino acids in the first and second repeats with alanine

residues. From these experiments, they were able to define a functional CTD unit. This unit covers an entire first repeat, while from the following repeat only Tyr1-Ser2-Pro3-Thr4 residues were necessary (Figure 6D). Further analysis of the functional unit revealed that only the three SP (serine-proline) motifs and the two tyrosine residues, at their respective positions, were required for a functional unit. The spacing between two consecutive tyrosine residues is important. Additional experiments showed that at least eight of these minimal functional units are required for growth of *S. cerevisiae*.<sup>19</sup> The distance between minimal functional units is not critical as long as flexible spacers are placed between.

This careful analysis answered several important questions with regard to the molecular action of the CTD. First, even in organisms that display a strong periodicity of heptad-repeats, this periodicity is dispensable and not essential for growth. Second, the CTD can be divided into functional units and does not act as an entire structure. Third, the presence and spacing of the SP motifs is the most conserved feature of a minimal functional unit. Finally, the model that the CTD acts as a platform for the recruitment and dissociation of cellular factors is compatible with the presence of functional units. Genetic analyses of the CTD in *S. cerevisiae* and *S. pombe* came to very similar results with respect to sequence requirement and length of a minimal functional unit. However, why *S. pombe*<sup>20</sup> but not *S. cerevisiae*<sup>15</sup> tolerates the replacement of tyrosine by phenylalanine is currently unclear. Possible explanations are that the organisms differ in their requirement of tyrosine for the recruitment of cellular factors to the CTD, or that differences in the construction of the mutants could be relevant. Comparable analyses in mammalian cells have not yet been performed. Because *S. cerevisiae* and *S. pombe*, but not mammalian<sup>21</sup> or chicken cells,<sup>22</sup> tolerate the replacement of Thr4 and Ser7 by alanine, the outcome of such experiments in mammalian cells is open. In the following sections, we describe and discuss the chemical properties and the phosphorylation of the consensus heptad-repeat.

### 2.3. Properties of the Consensus Repeat

The sequence complexity of the CTD consensus repeat is rather low, the heptad-motif being composed of only four different amino acids. The three amino acids with side chain hydroxyl-groups, serine, threonine, and tyrosine, generate a hydrophilic repeating block that is highly amenable to reversible modifications of the hydroxyl group, for example, by esterification (Figure 6A). The proline residues presumably help to stabilize the secondary structure and help prevent proteolytic degradation. Despite its polarity, the consensus sequence is uncharged; the 25 repeats of the *S. cerevisiae* CTD, not counting the last repeat, might indeed be one of the longest naturally occurring polypeptide chains without any charged residue.

The five amino acids Tyr1, Ser2, Thr4, Ser5, and Ser7 can all be reversibly modified by phosphorylation of the side chain hydroxyl group, as research over the past 20 years has revealed.<sup>21–23</sup> In addition, Thr4<sup>24</sup> and the two C-terminal serines, Ser5 and Ser7,<sup>25</sup> were found to be dynamically glycosylated, and the two prolines exist stably in either *cis*- or *trans*-configuration (Figure 6B). Originating from the native, nonmodified state, the possible number of combinations of post-translational modification together with the two proline configurations is  $2 \times 2 \times 2 \times 3 \times 3 \times 2 \times 3 = 432$  possibilities in a single hepta-consensus repeat. In a functional unit, as

described above, these combinations sum to  $2 \times 2 \times 2 \times 3 \times 3 \times 2 \times 3 \times 2 \times 2 \times 3 = 10\,368$  possibilities (Figure 6D). However, currently it is unclear how many of these combinations actually occur *in vivo*, and it is expected that some modifications exclude each other. The sequence variations in the distal part of the mammalian CTD add further to the complexity of the heptad-repeats and to the possible combinations of modification.

The conformation of the CTD remains enigmatic. In the crystal structure of yeast RNAPII, the CTD is not visible because of mobility.<sup>26</sup> Although the CTD shows some residual structure and a tendency to form  $\beta$ -turns, the CTD is thought to be largely flexible.<sup>5b</sup> The two SPXX motifs ( $S_2P_3T_4S_5$  and  $S_5P_6S_7Y_1$ ) were proposed to form  $\beta$ -turn structures stabilized by two hydrogen bonds between the first serine and the backbone amid groups of the third and fourth residue, respectively.<sup>27</sup> Electron micrographs of RNAPII revealed a weak density that was attributed to the CTD and measure only  $\sim 100$  Å,<sup>28</sup> and RNAPII crystals contained a limited space adjacent to the linker that could harbor a compact CTD.<sup>26b</sup> In contrast, in an extended conformation, the length of the yeast CTD would span about 650 Å. Together with the linker, that could make up to 250 Å (Figure 1A); the CTD would thus be 5–6 times longer than the yeast RNAPII is in diameter ( $\sim 150$  Å).<sup>26b</sup> In such a hypothetical conformation, the CTD could reach basically anywhere on the RNAPII core structure but could also locate on the DNA, pre-mRNA, other transcription factors, or histone structures of the transcribing complex. While there is increasing evidence that the CTD is most likely compact, at least in its unphosphorylated state, extensive phosphorylation would lead to a more extended conformation because of charge repulsion.<sup>5b</sup> Consistent with this, electrophoretic, gel filtration, and sucrose gradient analyses revealed that phosphorylation of the CTD results in a far more extended and more protease-sensitive structure.<sup>23d,29</sup> The largely disordered nature of the free CTD in fact allows for many different interactions with target proteins, possibly by an induced fit mechanism upon binding.

### 2.4. Post-Translational Modifications of the CTD

The CTD undergoes a dynamic cycle of post-translational modifications that encompasses all repeats. Similar to the modifications of histones, the enzymes that perform these changes can be described as “writers” and “erasers”. The modification pattern is recognized by interacting proteins, the “readers”, that can specifically bind to these sequence marks. The interplay of the modifying enzymes together with the functional output generated by the recognizing factors gives rise to the idea of a transient “code”. However, these modifications also influence the intrinsic properties of the CTD. For example, phosphorylation introduces negative charges, which lead to repulsion and may induce an extended conformation of the sequence. Likewise, the *cis*–*trans*-proline isomerization alters the conformation of the CTD structure, and O-glycosylation might drastically change the properties of the CTD by increasing its size and mass.

Despite the repetitive nature of the CTD as a substrate, the homogeneity of the post-translational modification pattern is a matter of debate. It is often not known if a specific modification is repeatedly set in several hepta-repeats of the CTD, or if a modification occurs in adjacent repeats, or if a specific pattern is adhered to (e.g., one repeat set, the following left out) and even if modifications predominate in proximal or distal repeats. This

## CTD interacting proteins

writers		readers			erasers	
protein (complex)	function	protein (complex)	function	reader sequence	protein	function
<i>Mammals</i>						
Cdk7 (TFIIEH)	Ser5-P, Ser7-P	TFIIE	preinitiation complex	hypo-CTD	Ssu72	Ser5, Ser7 phosphatase
Cdk8 (Mediator)	Ser5-P	TFIIF	preinitiation complex	hypo-CTD	Scp1	Ser5 phosphatase
Cdk9 (P-TEFb)	Ser2-P, Ser5-P, Ser7-P	TBP (TFIID)	preinitiation complex	hypo-CTD	Rtr1	Ser5 phosphatase
Cdk12	Ser2-P	Mediator	transcription activation	hypo-CTD	Rpap2 *	Ser5 phosphatase
<i>histones and chromatin</i>						
Cdk13	Ser2-P	MLL1, MLL2	histone methylation	hyper-P	Fcp1	Ser2 phosphatase
Plk3	Thr4-P	Set1	histone methylation	hyper-P	Cdc14	Ser2, Ser5 phosphatase
cAbl1, cAbl2	Tyr1-P	Set2	histone methylation	hyper-P		
		Hypb	histone methylation	hyper-P		
CARM1	R1810 methylation	Rpd3S	histone deacetylation	hyper-P		
		Spt6	histone chaperon	hyper-P		
<i>Saccharomyces cerevisiae</i>						
Kin28	Ser5-P, Ser7-P	<i>RNA processing</i>				
Cdk8 (Srb10)	Ser5-P	Mce1	5' capping	hyper-P		
Bur1	Ser5-P, Ser7-P, Ser2-P	Cgt1	5' capping	hyper-P		
Ctk1	Ser2-P	Prp40 (U1 snRNP)	splicing	hyper-P		
		U2AF65/Prp19 (U2 snRNP)	splicing	hyper-P		
<i>Schizosaccharomyces pombe</i>						
Mcs6	Ser5-P, Ser7-P (?)	PSF/p54	multifunctional/splicing	hypo-/hyper-P		
Cdk8	Ser5-P	Sus1	mRNA export	hyper-P		
Cdk9	Ser2-P, Ser5-P	Yra1	mRNA export	hyper-P		
Lsk1	Ser2-P	Nrd1	transcript. termination	hyper-P		
		Sen1	transcript. termination	hyper-P		
		Rtt103	transcript. termination	hyper-P		
		Pcf11	cleavage/poly(A)	hyper-P		
		Ydh1, Yhh1	cleavage/poly(A)	hyper-P		
		Int11 (Integrator)	transcription activation	hyper-P		
		Tdrd3	small RNA processing	hyper-P		
<i>RNAPII CTD modifications</i>						
		Ess1/Pin1	proline isomerase	hyper-P		
		Rpap2 *	pSer5 phosphatase	hyper-P		

**Figure 7.** Overview of CTD interacting proteins. CTD interacting proteins are classified into “writers”, “readers”, and “erasers”, according to their function in modifying the CTD, recognizing a particular CTD modification pattern, or erasing modification patterns from the CTD. Kinases are the writer enzymes and are separately listed for mammals, budding yeast, and fission yeast. Readers are separated into general transcription factors, histone and chromatin modifying factors, RNA processing factors, and interactions retro-acted to the CTD. The CTD recognition pattern is shown for two repeats as this sequence stretch encompasses a functional unit. The asterisk by Rpap2 indicates its dual specificity of reading a modified CTD pattern and acting on the CTD as possible Ser5-P phosphatases. Note that in early studies many interactions were not probed for CTD Ser7-P marks. A possible contribution of this modification to the binding recognition may, therefore, have been missed.

issue is difficult to address by specific antibody recognition, as no information is provided about the localization or the quantitation of the modifications. Among the modifying enzymes, CTD kinases and phosphatases are the best-characterized groups. We will therefore describe the current knowledge on Ser, Thr, and Tyr phosphorylation and the counteracting factors for the reversible removal of the phosphate group.

**2.4.1. CTD Serine Kinases.** Over the past two decades, a series of kinases has been identified to be involved in CTD phosphorylation. As a component of the general transcription factor, TFIIEH, the cyclin-dependent kinase, Cdk7, was initially

discovered in yeast and human to contain CTD kinase activity.<sup>30</sup> Mammalian Cdk7 and its associated subunit, Cyclin H, as well as two orthologous yeast kinases, Kin28 in *S. cerevisiae* and Mcs6 in *S. pombe*, were found to be responsible for most Ser5 phosphorylation (Figure 7). Accordingly, specific inhibition of these kinases caused a drastic reduction of Ser5 phosphorylation *in vivo*.<sup>31</sup> Consistent with their presence in the preinitiation complex, Ser5 phosphorylation occurs early in the transcription cycle. Unexpectedly, Cdk7 and Kin28 were also identified as being crucial for phosphorylating Ser7 in human and yeast cells.<sup>31a,d,32</sup>

Cdk8 is a subunit of the Mediator complex and has been described to have Ser2 and Ser5 phosphorylation activity in vitro.<sup>33</sup> The Mediator complex is an essential transcription regulator in the preinitiation complex that bridges transcription factors with RNAPII. Cdk8 together with Cyclin C are recruited to promoters in a dynamic interaction between the Mediator complex and the Cdk8 module. This interaction is regulated by the ubiquitin ligase, Fbw7, which targets the connecting Mediator subunit, MED13/13L, for degradation.<sup>34</sup> Of note, Cdk8/Cyclin C is regulated differently from other Cdks, as the Cdk8 T-loop appears not to be phosphorylated for activation by upstream kinases.<sup>35</sup> The actual contribution of Cdk8 to CTD phosphorylation in vivo is, however, unclear as a number of recent studies have provided evidence that Cdk8 can play a role as coactivator in the preinitiation complex.<sup>36</sup> For example, the yeast Cdk8 orthologue, Srb10, was shown to target several other transcription factors as well as the yeast transcription activator, Gcn4.<sup>37</sup>

The transition from transcription initiation to productive elongation is connected with the kinase activity of positive transcription elongation factor b (P-TEFb). P-TEFb was first described to overcome promoter-proximal pausing of RNAPII and to stimulate transcription elongation in vitro.<sup>38</sup> Active P-TEFb consists of Cdk9 and Cyclin T, and its kinase activity is potently inhibited by flavopiridol and 5,6-dichloro-1- $\beta$ -D-ribofuranosylbenzimidazole (DRB).<sup>38a,39</sup> The importance of P-TEFb in the transcription cycle emerged when Cyclin T1 was discovered as the cellular target of the HIV-1 Tat/TAR ribonucleoprotein complex.<sup>40</sup> Human retroviruses, HIV and HTLV, have long been known to increase replication by stimulating transcription of viral genes at the level of transcriptional elongation.<sup>41</sup> Likewise, stress-induced genes, such as the heat shock protein family, were known to directly proceed into transcription elongation.<sup>42</sup> P-TEFb achieves these functions by a dual mechanism. First, it phosphorylates negative factors, such as subunit Spt5 of the DRB sensitivity-inducing factor (DSIF) complex and the negative elongation factor (NELF), to release the paused state. After transcription initiation, these two factors associate with the RNAPII and pause the elongating RNAPII approximately 50–150 nucleotides downstream from the transcription start site.<sup>43</sup> In addition, P-TEFb phosphorylates RNAPII CTD at Ser5. From its function in the regulation of transcriptional elongation, it has been concluded that P-TEFb phosphorylates Ser2 if it is associated with the elongating RNAPII. Recent studies, however, showed that Cdk9 preferentially phosphorylates Ser5 of the CTD in vitro, although there seems to be some promiscuity toward Ser7 phosphorylation.<sup>44</sup> Initial studies showed indeed that human Cdk9 and Cdk9 (P-TEFb) in complex with Tat as well as the yeast orthologue CTDK-I (Ctk1 kinase) preferentially phosphorylate Ser5 of the CTD.<sup>45</sup> A recent study of human P-TEFb confirmed the Ser5 specificity of Cdk9 and showed that Cdk9 is unable to phosphorylate a CTD prephosphorylated at Ser2, whereas it readily phosphorylates a CTD prephosphorylated at Ser7 suggesting the simultaneous appearance of both phosphorylation marks in the same repeat.<sup>44</sup> Similar results were obtained for the fission yeast kinase Cdk9 that is primed by Ser7 prephosphorylation.<sup>31e,46</sup> The findings were confirmed by in vivo live imaging analyses that showed the colocalization of Cdk9 with Ser5-P but not Ser2-P marks in the nucleus and the co-occupancy of Cdk9 and RNAPII Ser5 phosphorylation

marks in promoter-proximal regions of transcribed genes by chromatin immunoprecipitation experiments.<sup>47</sup>

Recent studies in *Drosophila* and human cells led to the identification of two additional CTD kinases, Cdk12 and Cdk13.<sup>48</sup> The kinase domains of human Cdk12 and Cdk13 share about 89% sequence identity such that their binding specificity and substrate preference might be considered similar from a biochemical perspective. Cdk12 is suggested to have Ser2 phosphorylation activity, and chromatin immunoprecipitation experiments confirmed that *Drosophila* Cdk12 is present on the transcribed regions of active genes.<sup>48a</sup> Accordingly, its depletion resulted in decreased expression of predominantly long genes with high numbers of exons.<sup>48b</sup> The identification of Cdk12/Cdk13, in addition to Cdk9 (P-TEFb), seemed to resolve a longstanding enigma, as budding yeast and fission yeast each contain two CTD kinases that are associated with the regulation of transcriptional elongation. In *S. cerevisiae*, Bur1 and Ctk1 phosphorylate the CTD in the initiating and elongating state. Likewise, *S. pombe* contains the two kinases, Cdk9 and Lsk1, that can now be assigned, based on mutual sequence alignments, to be orthologues of human Cdk9 and Cdk12/Cdk13, respectively.

Other CTD kinases described to mediate serine phosphorylation were from the MAP/Erk protein kinase family and the atypical kinase Brd4.<sup>49</sup> Erk1 and Erk2 were indeed shown to phosphorylate Ser5 of the CTD in vitro, yet their function in vivo remained elusive. Whereas MAP kinases are known to preferentially phosphorylate Ser-Pro motifs within substrates, Brd4 is suggested to contain a noncanonical N-terminal kinase domain of approximately 700 residues encompassing also the two bromodomains. Brd4 is supposed to phosphorylate Ser2 of the CTD, but the results await further confirmation and a functional proof.

**2.4.2. Other CTD Kinases.** Besides the serine residues, the Tyr1 and Thr4 residues are also phosphorylated in vivo. Thr4-P marks are required for histone mRNA 3' end processing in chicken DT40 cells, probably functioning to facilitate recruitment of 3' processing factors to histone genes.<sup>22</sup> In mammalian cells, Polo-like kinase 3 (Plk3) has been described as a Thr4-specific kinase,<sup>21</sup> while a corresponding kinase in yeast has yet to be identified. Cdk9 inhibitors, DRB and flavopiridol, also lead to diminished Thr4-P levels in chicken and human cells.<sup>21,22</sup> However, as Cdk9 is a known CTD kinase, the inhibitory effect on Thr4 phosphorylation could be indirect and be caused by repression of serine phosphorylation as this modification could be a prerequisite for Plk3 recruitment via its polo-box domain.

Phosphorylation of Tyr1 was described 20 years ago in human cells, but was only recently confirmed in yeast.<sup>23c,50</sup> Indeed, this CTD modification occurs at all active genes and correlates with the distribution of Ser2 phosphorylation. Tyr1 phosphorylation inhibits the binding of the termination factor to RNAPII, because it physically prevents binding of the CTD to the conserved CTD-interaction domain of termination factors such as Nrd1.<sup>23c</sup> Phosphorylation of Tyr1 is mediated by the proto-oncogene kinases, c-Abl1 and c-Abl2, which phosphorylate the CTD to high levels in vivo and in vitro.<sup>50,51</sup> Interestingly, the association with and phosphorylation of the RNAPII CTD depend on an SH2 domain and a CTD interaction domain in c-Abl1.<sup>50</sup>

**2.4.3. CTD Phosphatases.** Phosphorylation of the CTD is a reversible process that allows recycling of RNAPII to the preinitiation complex. Several phosphatases have been

implicated in the removal of phosphates from the CTD, thereby mediating transitions in the transcription cycle, for example, from high Ser5-P levels at the 5' end of genes to high Ser2-P levels at the 3' end. Phosphorylated Ser5-P residues are specifically dephosphorylated by Ssu72, which is conserved from yeast to humans.<sup>52</sup> Ssu72 is enriched at both the promoter and the 3' end of genes, with a greater presence in the 3' end region.<sup>16</sup> Depletion of Ssu72 leads to increased levels of Ser5 phosphorylation toward the 3' end of genes.<sup>53</sup> Ssu72 was also suggested to dephosphorylate Ser7-P, as its depletion also resulted in elevated levels of Ser7-P at the 3' end of genes.<sup>16,53</sup> The ability to associate simultaneously with the promoter and termination site of genes led to the suggestion that Ssu72 mediates gene-loop formation,<sup>54</sup> an effect that enhances transcriptional directionality as mRNA and ncRNA (noncoding RNAs) are often transcribed in opposite directions from bidirectional promoters.

The yeast protein, Rtr1, has been described as a second, atypical Ser5-P phosphatase without a known functional phosphatase motif.<sup>55</sup> Rtr1 localizes to the 5' end of genes, and deletion of the *rtr1* gene consistently causes increased levels of Ser5-P and reduces occupancy of RNAPII on a number of genes. Unlike Ssu72, Rtr1 mediates only Ser5-P but not Ser7-P dephosphorylation.<sup>55</sup> The crystal structure of Rtr1 did not reveal an active site, and extensive experiments to demonstrate its CTD phosphatase activity have been unsuccessful, suggesting that Rtr1 has a noncatalytic role in CTD dephosphorylation.<sup>56</sup> The precise function of Rtr1 in CTD dephosphorylation is therefore currently unclear. The human orthologue of Rtr1, Rpap2, specifically recognizes the Ser7-P mark on the CTD and interacts with Integrator complex subunits.<sup>57</sup> Rpap2 is detected at the 5' end of protein-coding and snRNA genes and is essential for efficient transcription and 3' end processing of snRNA transcripts.

The levels of Ser2 phosphorylation are modulated at the end of the transcription cycle by the Ser2-P phosphatase, Fcp1.<sup>58</sup> Fcp1 efficiently recycles RNAPII O in a processive manner to the RNAPIIA form for PIC assembly. Fcp1 is present at both 5' and 3' ends of active genes,<sup>59</sup> in line with a proposed function of dephosphorylating both Ser5-P and Ser2-P marks and a function in termination.<sup>59,60</sup> Fcp1 dephosphorylation of the RNAPII CTD was also shown to be required for efficient transcription of heat shock genes, and its depletion dramatically increased phosphorylation of the nonchromatin-bound RNA-PII.<sup>61</sup> In addition, a family of three small CTD phosphatases (SCPs) with sequence homology to Fcp1 was identified in higher eukaryotes.<sup>62</sup> These phosphatases preferentially dephosphorylate Ser5-P marks from the CTD. SCP1 is recruited to promoters by the transcription factor REST/NRSF, specifically to regulate expression of neuronal genes in non-neuronal cells.<sup>63</sup> This neuronal gene silencing is achieved by SCP1 mediated Ser5-P dephosphorylation on promoter proximal RNAPII, but the role in general transcription is unclear. SCPs thus appear to play a role in tissue-specific transcriptional regulation. In addition to SCP-1,<sup>64</sup> and CPL-1 and CPL-2 (CTD-phosphatase-like),<sup>65</sup> several plant-specific phosphatases have recently been described for the CTD in *Arabidopsis thaliana*. In addition to transcription cycle associated phosphatases, the mitotic phosphatase, Cdc14, has also recently been reported to remove Ser2 and Ser5 phosphate groups from the RNAPII CTD in yeast and mammals to repress transcription during mitosis.<sup>66</sup>

#### 2.4.4. Mechanisms of Template Phosphorylation and Dephosphorylation.

The repetitive nature of the CTD hepta-repeats allows for the positioning or removal of phosphorylation marks in a repeated manner. The successive modification of such periodic template structure is known as “processive mechanism”, whereas the random modification corresponds to a “distributive mechanism”. However, also the priming of a repetitive substrate template is possible, meaning that once a first modification mark has been set, the following (upstream or downstream) repeat will be preferentially modified due to an increased recognition affinity. Such modification settings implying the directionality action confers to a “consecutive mechanism”.

The current picture of CTD phosphorylation and dephosphorylation appears heterogeneous because of the promiscuity of some kinases. Particular combinations of different modification patterns have not yet been explored, and data resulting from the use of various antibodies seem ambiguous. The periodicity of the CTD substrate appears to be an ideal template for uniform modification patterns. Using analytical methods, it has indeed been shown that Cdk9 (P-TEFb) induces the same number of phosphorylations on a CTD consensus template as the number of hepta-repeats provided, suggesting that every hepta-repeat gets phosphorylated once.<sup>44</sup> The mechanism of phosphorylation, however, was found to be distributive but not processive as might be expected from the uniformly modified sequence. In addition to their function as “writers”, kinases may also be seen as “readers” because they might specifically recognize a CTD modification pattern to set another phosphorylation mark. Similar recognition and reaction mechanisms may hold for phosphatases, and it has to be explored how many repeats are between the recognition site and the site of modification. The advent of analytical methods for the analysis of CTD modifications both in vivo and in vitro might help to explore the combinatorial space of these CTD kinase generated modification patterns (this point is discussed in detail in the section 7.1).

#### 2.4.5. Other CTD Modifying Enzymes.

Besides kinases and phosphatases, several other enzymes have been described that reversibly modify the CTD. Among those, arginine methylation adopts a unique role as there are only two arginine residues present in the human CTD. Arginine R1810 aligns to position 7 of hepta-repeat 31. The guanidinium ion of the R1810 side chain is dimethylated by the coactivator-associated arginine methyltransferase 1, CARM1.<sup>67</sup> CARM1 is a type I protein arginine methyltransferase that catalyzes a methyl-transferase reaction, producing asymmetric dimethylated arginine, and its substrates include histone H3 and the histone acetyltransferase p300. It has been implicated to be involved in coactivation of nuclear receptor-directed transcription as well as mRNA splicing,<sup>68</sup> although the underlying mechanisms are not known. The enzymatic activity of CARM1 toward R1810 is repressed by phosphorylation of Ser5 and Ser2 residues in vitro, but methylated R1810 can be found with hyperphosphorylated CTD in vivo, suggesting that the methylation is placed before early initiation and is present during transcription.<sup>67</sup> On a functional level, the R1810 methyl-mark inhibits general expression of snRNAs and snoRNAs, as its replacement by alanine, as well as genetic knockout of CARM1, leads to the specific upregulation of these RNA species. Interestingly, the repeat 31 with the Arg residues is not essential for viability in mice as a deletion of 13 repeats including repeat 31 viable, if smaller mice.<sup>69</sup> Overall, this and other modifications,

particularly at position 7 within the nonconsensus repeats of the distal part of the mammalian CTD, may serve as a characteristic label to recruit, or keep away, the transcription machinery to or from certain gene loci.

The identification of CTD O-glycosylation two decades ago is a landmark in the analysis of post-translational modifications.<sup>24</sup> Most glycosylation consists of N-linked glycan attached to the nitrogen of asparagine or rarely arginine, whereas O-linked glycosylation is via attachment to the hydroxyl oxygen of serine, threonine, or other hydroxyl group containing residues. N-Glycosylation occurs in the lumen of the endoplasmic reticulum, and modified proteins are then often secreted into the extracellular space, while O-glycosylation is also present in the cytoplasm and nucleus as the O-linked N-acetylglucosamine (O-GlcNAc) modification. Glycopeptides obtained by proteolytic digestion of the CTD show that glycosylation occurs throughout the entire CTD, suggesting that this modification could interfere with and prevent CTD phosphorylation.<sup>24,70</sup> Dynamic glycosylation occurs at Thr4 residues,<sup>24,70</sup> and also at Ser5 and Ser7 of the CTD, and is mediated by O-GlcNAc transferase (OGT).<sup>25</sup> OGT is a component of the preinitiation complex, and its knock down causes a reduction in transcription and RNAPII occupancy at several B-cell promoters. O-GlcNAcylation occurs in higher eukaryotes, but not in yeast; therefore, the cycling of O-GlcNAc could be important for species-specific gene transcription, yet its functional significance is still enigmatic.

The least understood modification of the largest subunit of RNAPII, Rpb1, is ubiquitination. Proteins modified by poly ubiquitination are degraded by the proteasome, so the control of ubiquitination is crucial to avoid unscheduled proteolysis of essential proteins. The marking of a protein with ubiquitin consists of a series of steps, which is often referred to as the E1–E2–E3 cascade. Transcriptional arrest can trigger ubiquitin-mediated proteolysis of RNAPII, and Rpb1 was indeed shown to be poly ubiquitinated, although it is not clear if this modification occurs within the CTD. RNAPII poly ubiquitination induced by DNA-damage after UV irradiation requires cooperation between distinct, sequentially acting ubiquitin ligases, such as Rsp5 in yeast or its human homologue NEDD4, and the Elongin–Cullin complex.<sup>71</sup> The ubiquitin protease, Ubp3, was in turn shown to deubiquitinate RNAPII in yeast.<sup>72</sup> Genetic characterization of Ubp3-deficient cells is consistent with a role in elongation. Ubp3 can be copurified with RNAPII and deubiquitinates both mono- and poly ubiquitinated RNAPII in vitro.<sup>72</sup>

Finally, the two prolines at positions 3 and 6 of the heptad-repeats can undergo conformational changes catalyzed by peptidyl-prolyl *cis/trans* isomerase (PPIases). Two enzymes have been described, Ess1 in yeast and Pin1 in mammals, which specifically recognize the pSer-Pro motif and catalyze transition between the two isomeric states of Pro3 or Pro6 in the CTD, respectively.<sup>73</sup> The proline isomerization state creates binding scaffolds for CTD interacting factors that contribute to the complexity of the CTD code. It has been shown that the cleavage and polyadenylation factor Pcf11 is specifically recruited to CTD repeats carrying Ser2-P marks in combination with the adjacent prolines in a *trans*-conformation.<sup>74</sup> Likewise, the CTD phosphatase, Ssu72, prefers Ser5-P marks in combination with Pro6 in a *cis*-configuration.<sup>52b,75</sup> The isomerization status of the prolines thus influences the recognition of CTD modifying factors and, thereby, contributes to the complexity of the CTD code.

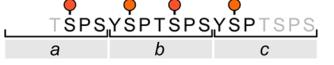
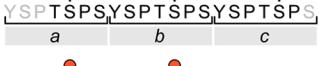
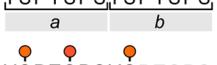
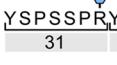
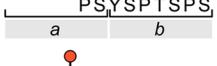
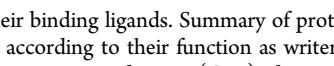
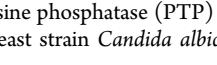
## 2.5. CTD Binding Factors

A number of proteins and multisubunit protein complexes have been reported to interact with the CTD. Here, we describe these reader proteins according to their function in the transcription process. As such, we separate these proteins into the general transcription factors that maintain the transcription cycle, the proteins that crosstalk to the histone and chromatin structures, the multiple factors that are involved in the processing of pre-mRNA and snRNA, and finally reflexive factors that read and change the CTD (Figure 7).

**2.5.1. General Transcription Factors.** The general transcription factors (GTFs) TFIIE, TFIIF, TFIID including the TBP subunit, and Mediator complex all interact with the hypo-phosphorylated CTD in the preinitiation complex.<sup>76</sup> The exact binding of these factors to the CTD is largely unknown and might even be promiscuous as the multisubunit assembly may involve binding to large CTD sections. One example has been reported recently for the Mediator–CTD interaction,<sup>77</sup> which is described in section 2.6.

**2.5.2. Histone and Chromatin Modifying Factors.** Complexes that signal to the histone structures are much better understood and appear to have more specific CTD interaction modules. These factors include the histone methyltransferases, MLL1 and MLL2, Set1 and -2, and Hypb. Mixed-lineage leukemia (MLL) histone methylases, MLL1 and MLL2, control the assembly of transcription preinitiation complexes and RNAPII recruitment. MLL1 and MLL2 are both H3K4 methyl-transferases that specifically associate with a subset of transcriptionally active target genes.<sup>78</sup> Set1 mediates the methylation of histone H3K4<sup>79</sup> and interacts with a CTD that is phosphorylated at Ser5 but not Ser2.<sup>79b</sup> Set2 is a H3K36 methylase that associates with the doubly modified Ser2-P/Ser5-P hyper-phosphorylated form of RNAPII, therefore linking transcriptional elongation to chromatin methylation.<sup>80</sup> Hypb is a human histone H3K36-specific methyltransferase that selectively associates with the hyperphosphorylated form of the RNAPII CTD, but the exact reader sequence is not known.<sup>81</sup> The small form of the histone deacetylase Rpd3, Rpd3S, is specifically recruited by the Ser5 phosphorylated CTD to actively transcribed genes.<sup>82</sup> Spt6 promotes transcription elongation at many genes and functions as a histone H3 chaperone to alter chromatin structure during transcription.<sup>83</sup> Mammalian Spt6 binds the Ser2 phosphorylated CTD through its tandem SH2 domains,<sup>84</sup> which leads to the recruitment of Iws1 and the REF1/Aly mRNA export adaptor to facilitate mRNA export.

**2.5.3. RNA Processing Factors.** The CTD provides the basis for coupling transcription with RNA processing.<sup>85</sup> Multiple factors have been described to interact with different combinations of modified hepta-repeats. The capping complex places the m7G cap on the nascent transcript as it exits the core polymerase, stabilizing the mRNA by preventing its degradation by 5'-3' exonucleases. The interaction of the guanylyltransferases, Mcel1 and Cgt1, with Ser5-P is the first step of mRNA processing.<sup>86</sup> Direct interactions between a splicing factor and the phosphorylated CTD were first shown for the yeast splicing factor, Prp40, which contains multiple binding sites for the CTD.<sup>87</sup> Two other splicing factors, U2AF65 and Prp19, were copurified with hyper-phosphorylated CTD.<sup>88</sup> The splicing and transcription-associated factors, PSF and p54, bind to both hypo- and hyper-phosphorylated CTD matrices with similar selectivity, providing a direct link to pre-mRNA processing components.<sup>89</sup>

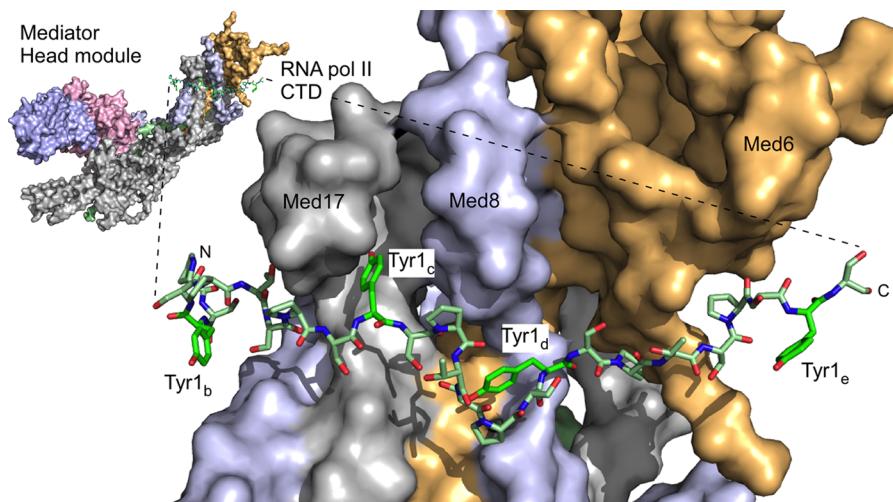
protein	domain fold	organism	CTD peptide sequence	pdb accession code	reference
<b>writers/effectors</b>					
Mediator		(S.c.)		4GWQ	77
<b>readers</b>					
Mce1	NT	(M.m.)		3RTX	102
Cgt1	NT	(C.a.)		1P16	103
Nrd1	CID	(S.c.)		2LO6	97
Scaf8	CID	(H.s.)		3D9K	104
Rtt103	CID	(S.c.)		2L0I	105
Pcf11	CID	(S.c.)		1SZA	74a
Pin1	WW	(H.s.)		1F8A	100
Tdrd3	Tudor	(H.s.)		2LTO	99
<b>erasers</b>					
Scp1	FCPH	(H.s.)		2GHQ	101
Ssu72	PTP	(H.s.)		3O2Q	107
Ssu72	PTP	(D.m.)		3P9Y	75
Ssu72	PTP	(H.s.)		4H3K	108
Ssu72	PTP	(H.s.)		4H3H	108

**Figure 8.** Protein–CTD complex structures and their binding ligands. Summary of protein–CTD complex structures displaying the CTD construct resolved in the structure. Structures were specified according to their function as writers/effectors, readers, and erasers. The domain fold describes the nucleotidyl transferase (NT) domain, the CTD interacting domain (CID), the WW-domain, the Tudor domain, the FCP-homology (FCPH) domain, and the low molecular weight protein tyrosine phosphatase (PTP) domain. Organisms refer to human (H.s.), mouse (M.m.), fruit fly (D.m.), budding yeast (S.c.), fission yeast (S.p.), and the yeast strain *Candida albicans* (C.a.). Accession codes of the protein database are given in the last column.

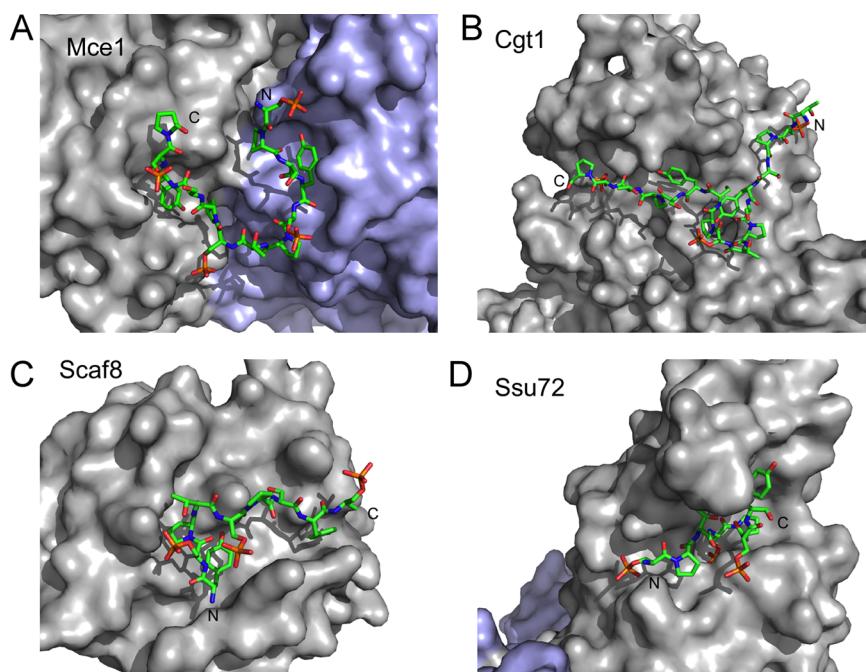
There are two factors involved in mRNA export that interact with the CTD directly. The transcription/export factor, Sus1, plays a key role in coupling transcription activation with mRNA export, and it resides in both the SAGA and the TREX2 complexes.<sup>90</sup> Sus1 is required during transcription elongation and is associated with the elongating form of the RNAPII CTD, which is phosphorylated on Ser5 and Ser2. In addition, Sus1 copurifies with the essential mRNA export factors, Yra1 and Mex67, which bind to the mRNA cotranscriptionally. In yeast cells, Yra1 cotranscriptionally associates with mRNA and delivers it to the nuclear pore complex for export to the cytoplasm. Yra1 directly binds the hyper-phosphorylated form of the CTD characteristic of elongating RNA polymerase II in vitro and contains a phospho-CTD-interacting domain, which also includes an “RNA recognition motif” (RRM).<sup>91</sup> Although the RRM is implicated in both RNA and CTD binding, RRM point mutations can separate these two functions. Both functions are important in vivo as RNA binding-defective or

CTD binding-defective versions of Yra1 result in growth and mRNA export defects.

Termination of transcription requires cleavage and polyadenylation of the mRNA in response to a specific sequence element. Polyadenylation factors are cotranscriptionally recruited by phosphorylation of CTD Ser2. For example, subunit Pcf11 of the yeast cleavage/polyadenylation factor IA directly binds to Ser2 phosphorylated CTD templates.<sup>92</sup> Similarly, Rtt103 is localized at the 3' ends of protein coding genes by the direct recognition of Ser2-P marks.<sup>93</sup> The yeast cleavage/polyadenylation factor (CPF) contains two transcription termination factors, Ydh1 and Yhh1, that bind specifically to the phosphorylated form of the CTD to recruit these complexes to RNAPII.<sup>94</sup> Likewise, the Integrator complex is recruited to phosphorylated CTD for 3' processing of snRNA. Here, the Integrator subunit, Int11, specifically recognizes the Ser7-P/Ser2-P double mark in the CTD.<sup>95</sup> The RNA-binding protein, Nrd1, and the RNA helicase, Sen1, contain CTD



**Figure 9.** Structure of the Mediator–CTD complex. A stretch of 25 nonmodified CTD residues binds in an extended conformation to the Mediator head module contacting the three subunits Med6, Med8, and Med17. The central tyrosine residues appear characteristic for the binding recognition.



**Figure 10.** Structures of CTD binding domains. (A) Recognition of the 5' capping enzyme, Mce1, by Ser2- and Ser5-phosphorylated CTD. (B) Binding of Ctg1 to Ser5-phosphorylated CTD. (C) Interaction of the Scaf8 CID domain with a Ser2-P/Ser5-P CTD matrix. (D) Binding of the protease, Ssu72, to triple serine-phosphorylated CTD.

binding domains that recruit these proteins to the CTD.<sup>96</sup> Whereas Nrd1 interacts with the Ser5 phosphorylated form of the CTD and requires a Ser5P-Pro6 bond in *cis* conformation,<sup>97</sup> Sen1 binds to Ser2-P marks. Both proteins interact with the termination factor, Nab3, suggesting a handoff model in which proteins differentially transfer from the Ser5- to the Ser2-phosphorylated CTD to promote the termination of noncoding transcripts or other cotranscriptional events for protein-coding genes.<sup>98</sup> Finally, Tdrd3 was shown to bind a methylated arginine CTD;<sup>99</sup> its knock down, however, does not affect snRNA and snoRNA processing.<sup>67</sup>

**2.5.4. Retroacting CTD Modifying Factors.** Finally, the peptidyl-prolyl *cis/trans* isomerases Pin1 (human) and Ess1 (yeast) can be considered reader proteins that also act on CTD conformation. They consist of two domains, an N-terminal

WW domain for binding and a C-terminal catalytic peptidyl-prolyl *cis/trans* isomerase (PPIC) domain. While the catalytic PPIC domain is able to bind Ser2-P and/or Ser5-P CTD phosphorylation marks on its own, the binding affinity increases up to 30-fold in the presence of the WW domain.<sup>100</sup>

## 2.6. Characteristics of CTD Binding Domains

The structural basis of protein–CTD binding is manifold due to the flexibility of the CTD polypeptide chain and the multiple combinations of modification marks. A systematic classification of protein domain scaffolds that interact with specific CTD motifs, for example, the binding of the SH2 domain module to phospho-tyrosine motifs, seems impossible as the structures of CTD-binding domains known to date appear too heterogeneous. The specificity of binding instead originates from the recognition of different modification isoforms, modification

patterns, length of bound CTD sequences, and *cis*- or *trans*-conformations of the pSer-Pro peptidyl-prolyl bonds.<sup>5a</sup> The available structures of CTD peptides bound to their “reader” domains illustrate the diversity in the recognition of basically the same peptide chain. The CTD polypeptide backbone adopts distinct types of loop conformations when bound, for example, to the CTD-interaction domain (CID) of Rtt103, the FCPH domain of Scp1, the nucleotidyltransferase (NT) domain of Cgt1, or the WW domain of Pin1 (for reviews, see refs 5a,b). In fact, each reader domain has its own requirement for the functional unit of the recognition sequence, which can be as long as the five repeats seen in the Mediator–CTD complex structure<sup>77</sup> or as short as four residues seen for Scp1.<sup>101</sup>

An overview of the known CTD complex structures, sorted for their function as writers/effectors, readers or erasers, is shown in Figure 8. The recently determined structure of the Mediator head module bound to an unmodified CTD sequence of five repeats is the most comprehensively described CTD structure.<sup>77</sup> Despite a moderate resolution (4.5 Å) of the difference electron density map upon soaking the Mediator head module with the CTD peptide, structural features, such as the sequence register and directionality, were identified (Figure 9). The polypeptide chain associates in an almost entirely extended conformation with three different subunits of the Mediator head module (Med6, Med8, and Med17) spanning a distance of 73 Å, with only the central heptad-repeat twisted in a β-turn structure. Two central tyrosines (Tyr1<sub>c</sub> and Tyr1<sub>d</sub>) are involved in tight interactions with Mediator, while serine residues of the CTD appear to make contacts that would be disrupted by phosphorylation (Figure 9). The structure of the head module–CTD complex thus provides a starting point for the modeling of a complete Mediator–RNAPII complex that would constrain the path and interaction space of the entire CTD.

Capping enzymes, like the guanylyltransferases, add the m7GpppN 5' cap to the nascent transcript as it exits the core polymerase. These enzymes specifically recognize the Ser5 phosphorylated CTD in the transcription initiating complex. There are two structures of guanylyltransferases complexed with CTD peptides known today that are classified into the nucleotidyltransferase (NT) domain family (Figure 10A and B). Although both capping enzymes bind the CTD peptides in an extended β-strand-like conformation, they use different interfaces to read the same modifications. Mouse capping enzyme, Mce1, was crystallized with a doubly phosphorylated Ser2-P/Ser5-P CTD peptide, yet the Ser2-P mark is solvent exposed and not involved in NT domain binding.<sup>102</sup> The residues most prominently involved in specific interaction are Tyr1 of the second repeat and Ser5-P of the first repeat that bind to the interface of the Mce1 NT domain homodimer (Figure 10A). On the other hand, the *Candida albicans* RNA guanylyltransferase, Cgt1, is monomeric, and the Ser5 phosphorylated CTD peptide spans the entire domain.<sup>103</sup> The N- and C-terminal Ser5-P marks of the three heptad-repeats interact with positively charged surface patches of the Cgt1 structure, whereas residues in the central repeat are not recognized but are exposed and can interact with other CTD-binding factors (Figure 10B).

The CTD-interaction domain (CID) is the best-studied scaffold for CTD-binding proteins. It consists of eight α-helices arranged into a right-handed superhelix.<sup>74a</sup> CIDs typically interact with 8–11 residues of the CTD that bind into a groove

formed by three helices of the protein domain. CID–CTD complex structures have been determined for the cleavage factor, Pcf11,<sup>74a</sup> the RNA processing factor, Scaf8,<sup>104</sup> the exonuclease complex, Rtt103,<sup>105</sup> and the transcription termination factor, Nrd1,<sup>97</sup> albeit with different CTD phosphorylation marks (Figure 8). In all structures, CTD residues S<sub>2</sub>P<sub>3</sub>T<sub>4</sub>S<sub>5</sub> adopt a β-turn conformation regardless of the phosphorylation pattern,<sup>104</sup> where Pro3 of the Ser-Pro peptidyl-prolyl bond is in a *trans*-conformation. The side chain hydroxyl group of the Tyr1 preceding the β-turn undergoes specific interactions with a conserved residue of the CID, highlighting its importance for recognition specificity (Figure 10C). The Nrd1 CID interaction surface is longest, with the Ser5-P-Pro6 motif preceding the β-turn structure and being in the *cis*-conformation. These structural differences indicate the versatility of CTD motif recognition even within the same domain family.

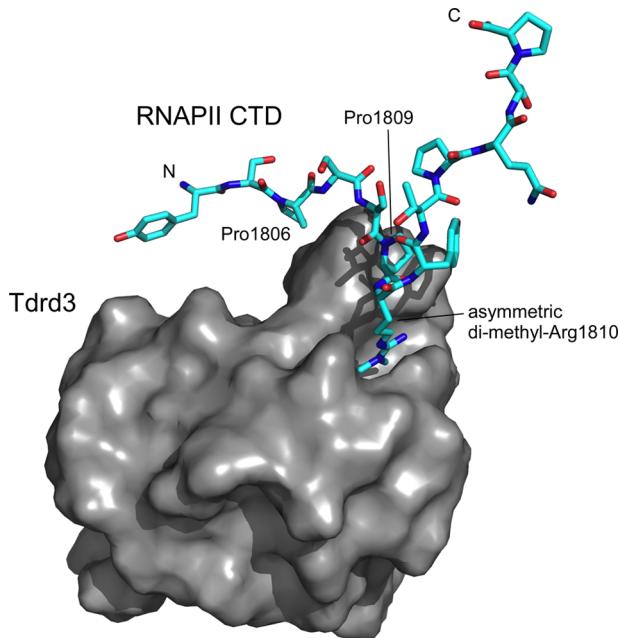
Structures of phosphatases in complex with various CTD peptides have been determined for the small CTD phosphatase, Scp1, and the Ser5/Ser7 phosphatase, Ssu72. Scp1 and Fcp1 belong to a family of Mg<sup>2+</sup>-dependent pSer/pThr-specific phosphatases that share the FCP-homology (FCPH) domain fold. The preference for Ser5-P dephosphorylation was rationalized in the crystal structure of catalytically inactive Scp1 bound to a 9-mer peptide containing Ser2-P and Ser5-P marks.<sup>101</sup> The phosphate group of Ser5-P is positioned to coordinate two catalytically important aspartate residues and the Mg<sup>2+</sup> ion. The preceding sequence stretch, S<sub>2</sub>P<sub>3</sub>T<sub>4</sub>pS<sub>5</sub>, adopts a β-turn-like structure similar to the CID–CTD complex structure in which the Ser2-P group does not contribute to the binding.<sup>101</sup> The 23 kDa encompassing phosphatase, Ssu72, in contrast belongs to the metal-independent phosphohydrolases that act via cysteinyl phosphoenzyme intermediates.<sup>106</sup> As a Ser5-P phosphatase, Ssu72 binds to the CTD by a conformational selection mechanism that allows only for the *cis*-conformation of the pSer5-Pro6 bond.<sup>75,107</sup> The overall domain fold of Ssu72 belongs to the low-molecular-weight protein tyrosine phosphatases (PTPases). Again, the Ser5-P group is tightly bound to the catalytic site, and surrounding residues, Pro3, Thr4, Pro6, and Tyr1, all bind to a narrow groove of the enzyme (Figure 10D). As the *cis* Ser5-P-Pro6 isomer is the minor population in solution, addition of the *cis/trans* isomerase, Ess1, to the reaction facilitates rapid dephosphorylation by Ssu72.<sup>75</sup>

Ssu72 has also been reported to be a Ser7-P phosphatase,<sup>53,73</sup> and the structural basis of this catalytic activity has been described only recently.<sup>108</sup> Surprisingly, a peptide containing Ser7-P instead of Ser5-P bound to the active site of Ssu72 in an opposite direction as compared to binding by the Ser5-P peptide. The Ser7-P phosphatase activity of Ssu72 is about 4000-fold lower toward a CTD peptide substrate as compared to its Ser5-P phosphatase activity *in vitro*. Yet, when all three serine phosphorylations were presented in a single hepta-repeat, the peptide changed its orientation back to that previously observed.

The WW-domain/PPI-domain assembly of human Pin1 bound to one canonical CTD hepta-peptide was the first CTD–protein complex whose crystal structure was determined.<sup>100</sup> WW domains are small units of approximately 40 amino acids that form a compact triple-stranded antiparallel β-sheet structure. The CTD peptide associated to the WW domain, and although both Ser2 and Ser5 were phosphorylated in the substrate peptide, only Ser5-P is directly recognized.

Additional extensive contacts were made only with Pro3 and Pro6 of the CTD. As described above for the Mediator–CTD complex (or as similarly seen for SH3 domain–PxxP motif interactions), the CTD peptide adopts an elongated conformation upon binding to the Pin1 WW-domain.

Finally, the solution structure of the Tudor domain-containing protein 3 (Tdrd3) bound to a 13-mer polypeptide of CTD repeats 31 and 32 modified with the asymmetric dimethylated arginine 1810 has recently been described.<sup>99</sup> The Tudor domain of only 58 residues contains a central aromatic cavity composed of three tyrosines and one phenylalanine that harbors the binding site for the modified arginine guanidinium ion. The binding site is selective for asymmetrically methylated arginines, but interestingly only the two preceding prolines of the heptad-repeat (P1806 and P1809) seem to sustain the interaction (Figure 11). This observation is in line with a very



**Figure 11.** Structure of the Tdrd3 Tudor domain bound to asymmetrically dimethylated arginine. The human Tudor domain interacts with the asymmetrically dimethylated Arg1810 of the human CTD in a specific conformation that includes the two preceding prolines.

weak dissociation constant ( $K_d = 770 \mu\text{M}$ ) determined for the interaction between the CTD peptide and the Tudor domain, which exceeds even the average affinity between acetylated histone motifs and their respective bromodomain recognition factors by 1 order of magnitude. This complex structure is a rare example where the Tyr1 position seems not to contribute to the binding recognition of the CTD.

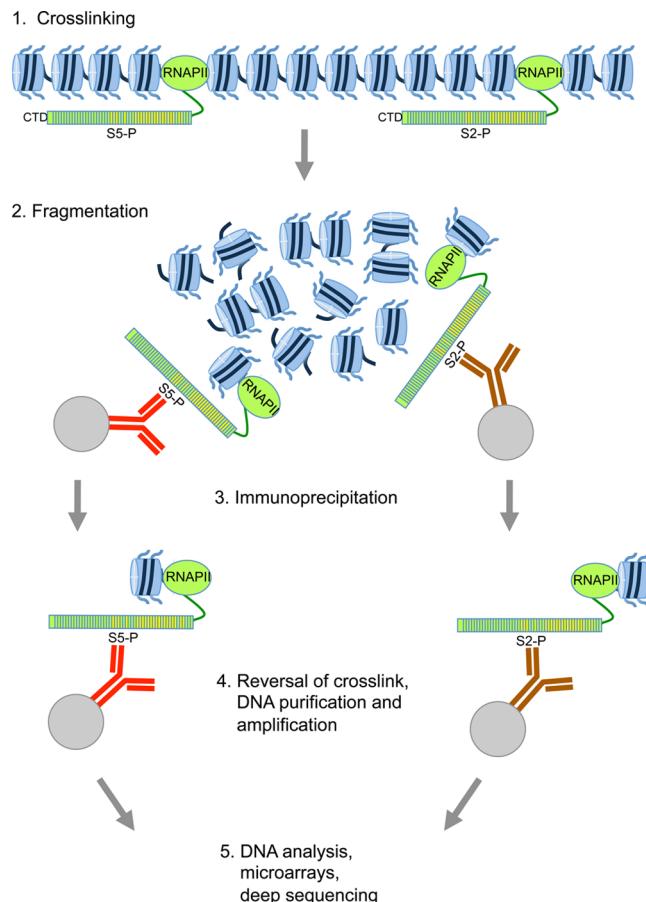
### 3. FUNCTIONAL ANALYSIS OF CTD MODIFICATIONS

The functional analysis of CTD modifications requires a specific technique and tools to study changes in CTD modifications in a gene-specific manner. The tools are monoclonal antibodies (mAb's) that can recognize and discriminate specific CTD modifications. The technique is called chromatin immunoprecipitation (ChIP). This technique is generally applied to study the binding of proteins to genomic DNA loci at high resolution. Before describing changes of CTD phosphorylation during the transcription cycle, we will first

introduce the ChIP technique and describe the potency and limitations of CTD-specific mAb's.

#### 3.1. Chromatin Immunoprecipitation (ChIP) of RNAPII

ChIP analysis is a technique to study the interaction of DNA and proteins at specific genomic loci in cells.<sup>109</sup> The analyzed proteins can be transcription factors, general transcription factors, including RNAPII, but also histones and histone variants, or other chromatin associated factors. The ChIP experiment is divided into several steps (Figure 12). First,



**Figure 12.** Schematic drawing of a chromatin immunoprecipitation (ChIP) experiment. (1) Cells are treated with formaldehyde to cross-link proteins to chromatin. Cells are lysed and (2) chromatin is fragmented by sonification or other methods. (3) Chromatin-bound factors are precipitated with specific bead-coupled antibodies. (4) The precipitated DNA is purified and amplified. (5) The genomic localization of the precipitated DNA is determined by microarray hybridization or deep sequencing.

potential factors are cross-linked to DNA (chromatin) by treatment of living cells with formaldehyde. In a second step, the chromatin is fragmented by sonication or other techniques to generate small DNA fragments ideally with a size of approximately 200 bp. In a third step, DNA fragments are precipitated with the help of specific mAb's recognizing cross-linked proteins, while control mAb's should not lead to the enrichment of this factor in control precipitates. In a fourth step, the precipitated DNA is heated for the reversal of the cross-link, purified, amplified by polymerase chain reaction (PCR), and subsequently analyzed with the help of DNA microarrays or by next generation sequencing. The ChIP technique is suitable and very sensitive for the determination of

genomic binding sites of factors. For the analysis of protein modifications, combinations of two different mAb's are used. The first mAb recognizes and precipitates the protein of interest independently of its modifications and is a measure for total binding of this factor to a specific gene locus. The second mAb recognizes and precipitates the protein in a modification-dependent manner. Control ChIP experiments of the first mAb ensure that changes in the signal obtained for the second, modification-specific mAb are true changes due to the modification state of a protein and not due to changes in the amount of a protein bound to a specific gene locus. Comprehensive ChIP experiments have been applied to study specific CTD modifications of RNAPII during a transcription cycle in yeast and mammalian cells. ChIP experiments do not yield orientation of RNAPII on the template, and thus mapping with this technology should be viewed with caution as potential antisense transcription complexes are precipitated as well. Yet before discussing ChIP data, we will introduce the mAb's that can detect CTD-specific modifications and that have been applied to ChIP analysis.

### 3.2. Monoclonal Antibodies for the Analysis of Specific CTD Modifications

Monoclonal antibodies represent a powerful tool for the investigation of CTD modifications. They recognize either the unphosphorylated CTD or are specific for a single phospho-residue in Tyr1, Ser2, Thr4, Ser5, or Ser7. The mAb's display no pan-activity, meaning that they do not recognize phosphorylated Tyr-, Ser-, or Thr- residues of other proteins. This high specificity is achieved by recognition of additional adjacent amino acids next to the phosphorylated amino acid residue. Unfortunately, adjacent amino acids can also undergo modification and, thereby, hinder the recognition of a specific phospho-epitope by a mAb. Because of this caveat, signals obtained for mAb's in ChIP experiments can be interpreted only in a positive, but not in a negative direction. The lack of a positive signal does not necessarily imply the absence of a modification. Conversely, an increase in the signal for a phospho-specific amino acid in the CTD does not necessarily mean that the phosphorylation level of this residue really was increased. The increase could also mean that adjacent phospho-residues, which hindered the binding of the mAb, were removed by a phosphatase. Another problem constitutes the cross-reactivity of CTD mAb's. The mAb H5 recognizes best, if Ser2 and Ser5 are phosphorylated in the same CTD repeat, but displays also specificity to Ser2-P in the absence of Ser5 phosphorylation.<sup>23a,53</sup> Thus, mAb's are reliable tools for the measurement of changes in the CTD modification pattern, but they fall short of giving a precise interpretation of this change. Finally, because of the highly repetitive structure of the CTD, it also remains unclear which combinations of modifications really exist at the same time in the same CTD heptapeptide *in vivo*. Currently, we have to deal with all of the restrictions described above, but the application of mass spectrometry will help to tackle some of these problems in the near future (see also section 7, Conclusions and Future Perspectives).

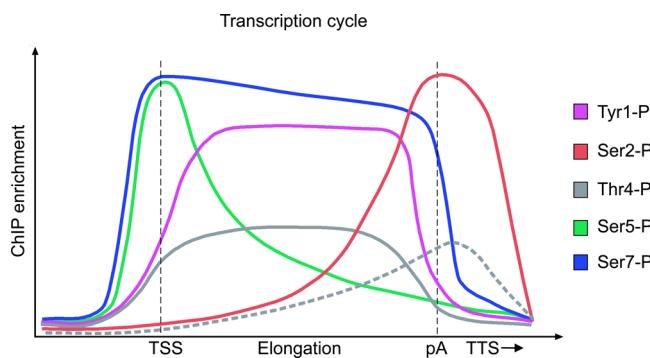
### 3.3. Dynamic Phosphorylation Patterns of the CTD in the Transcription Cycle

The phosphorylation patterns of CTD are subject to dynamic changes in the transcription cycle and undergo significant changes from initiation to termination.

The phosphorylation of CTD Ser2, Ser5, and Ser7 residues has been analyzed genome wide in *S. cerevisiae* by several

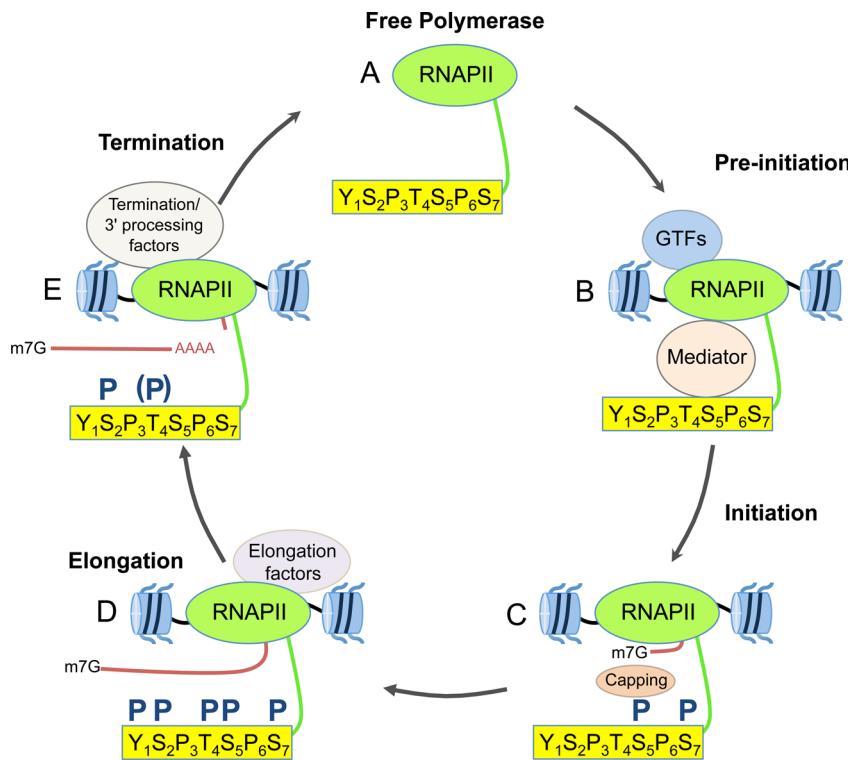
laboratories<sup>4a,53,110</sup> and also in mammalian<sup>2b,111</sup> and plant cells<sup>112</sup> with different sets of CTD phosphorylation specific antibodies. All groups come to very similar results and general conclusions, but also to some different observations and interpretations. The differences can rely on the methods of data mining and evaluation, which were not exactly the same for each analysis. Differences can also be true differences between species, because the demands of long transcription units of mammalian genes and short transcription units yeast genes on CTD functions may differ.

Therefore, the strength of signals obtained for Ser2-P, Ser5-P, and Ser7-P marks in ChIP experiments were plotted either for all genes, or for categories of genes with a large, intermediate, or short transcription unit. The transcription unit is defined as the distance from the transcriptions start site (TSS) to the poly adenylation and transcription termination site (TTS) (Figure 13). The study of Mayer et al.<sup>23c</sup> also



**Figure 13.** Changes of CTD modifications along the transcription cycle visualized by ChIP experiments in *S. cerevisiae*.<sup>4a,53,110</sup> Similar results were obtained for mammalian cells<sup>2b,111a</sup> (dashed line shows Thr4-P marks in mammalian cells<sup>21</sup>). Results are not normalized to RNAPII levels. Increased levels of Ser5-P marks at TSS and Ser2-P at TTS, at least partly, reflect increased levels to RNAPII. TSS, transcription start site; pA, polyadenylation site; TTS, transcription termination site.

analyzed the distribution of signals for Tyr1-P and Thr4-P. These data are included in Figure 13. It is generally accepted from *in vitro* transcription experiments that RNAPII with a hyper-phosphorylated CTD (also termed the RNAPIIO form) is incapable of proper transcription initiation, while the form of RNAPII with a hypo-phosphorylated CTD, RNAPIIA, can bind and initiate at promoters.<sup>113</sup> Whether this statement is applicable for all promoters *in vivo*, however, is currently unknown. Genes in *S. cerevisiae* are generally arranged very close to each other; therefore, those genes with termination and initiation sites in direct proximity were excluded from the evaluation to prevent a mixing of termination and initiation specific CTD phosphorylation signals. ChIP analysis revealed that signals for all CTD phospho-marks in front of the TSS are very low. A strong increase was observed for the Ser5-P and Ser7-P signals exactly at the TSS. The signal for Ser7-P remains high throughout the transcription cycle, while the signal for Ser5-P steadily decreases toward the poly-A (pA) site. The signals for Tyr1-P, Ser2-P, and Thr4-P were low at the TSS, but increase downstream of the TSS. The signals for Ser2-P are highest at and downstream of the poly-A, consistent with the recruitment of 3' RNA processing factors by a Ser2 phosphorylated CTD. High levels of Tyr1-P in the body of genes favor the binding of elongation factors and prevent



**Figure 14.** RNAPII transcription cycle. (A) Free, hypo-phosphorylated RNAPII, not engaged in transcription. (B) General transcription factors (GTFs) assemble the preinitiation complex at the promoter by recruiting RNAPII and the Mediator complex. (C) First wave of Ser5 and Ser7 phosphorylation in the initiation complex and recruitment of the capping enzyme. (D) Dynamic phosphorylation and dephosphorylation of the CTD during the elongation phase probably involving all heptad-repeats and all amino acid residues of the heptad-repeats. (E) Programming the CTD for termination, gradual removal of CTD phosphorylation by phosphatases, and release of RNAPII and transcripts from the template. The CTD is represented by one single consensus repeat in the scheme. Currently, it is unknown whether changes in phosphorylation occur in the proximal, middle, and distal part of the CTD in a uniform manner.

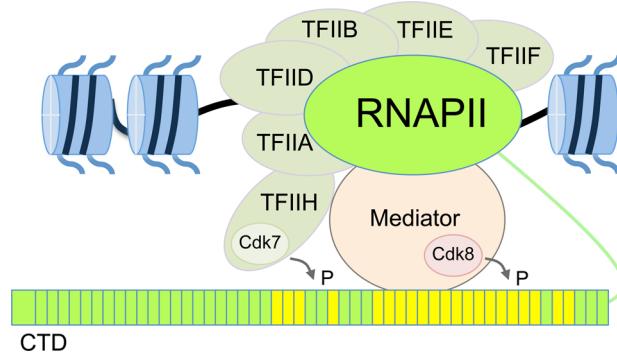
binding of termination factors to the CTD.<sup>23c</sup> Factors binding to Thr4-P have yet to be characterized. During the process of termination, all CTD phospho-marks are removed by phosphatases.

#### 4. CTD-REGULATED CELLULAR PROCESSES

The CTD regulates multiple steps during the transcription process. It starts with the coordination of transcription initiation by interacting with the Mediator complex and other general transcription factors. Subsequently, dynamic CTD phosphorylation changes orchestrate the recruitment of chromatin remodeling and chromatin modifying activities to allow the capping of the pre-mRNA, the release of RNAPII from the promoter and the transition into an elongation mode, the proper splicing and 3' processing of mRNA, and finally termination of transcription and export of the mRNA into the cytoplasm (Figure 14). As pointed out above, the observed sequence of changes in CTD phosphorylation and the suggested uniform transition of CTD phosphorylation stages along genes rely entirely on the application of mAb's in ChIP experiments. In situ changes in CTD phosphorylation are presumably more complex, because the single functional units in the CTD array probably do not all fulfill the same functions at the same stage of the transcription cycle. Thus, changes in mAb reactivity observed for CTD phosphorylation can also be an indication of structural changes in the CTD and may reflect the association to or dissociation of cellular factors from the CTD, or the addition or removal of adjacent CTD modifications, which interfere with mAb reactivity.

#### 4.1. Transcription Initiation

Transcription initiation requires the assembly of a preinitiation complex (PIC) consisting of the general transcription factors, TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIH, RNAPII, and the Mediator complex<sup>114</sup> (Figure 15). In vitro transcription



**Figure 15.** RNAPII initiation complex with general transcription factors TFIIA–TFIIE and the Mediator complex. CTD-specific kinases, Cdk7 and Cdk8, are indicated.

experiments demonstrated that polymerases harboring the hypo-phosphorylated CTD preferentially enter the preinitiation complex and become subsequently phosphorylated.<sup>113,115</sup> Phosphorylation of the CTD does not occur if the Mediator complex is previously immuno-depleted from nuclear extracts.<sup>116</sup> However, neither depletion of the Mediator complex

nor inhibition of CTD phosphorylation by kinase inhibitors prevents PIC formation and basal transcription activity, but it does abolish activated transcription.<sup>115,116</sup> These observations were confirmed in transient transfection experiments with reporter genes. Again, the lack of a CTD abolished activation by enhancers, whereas basal transcription or transcription from promoters driven by SP1, a factor that typically activates housekeeping genes from positions proximal of the initiation site, was not affected in transient reporter gene assays.<sup>117</sup> In contrast, endogenous chromatin-packed genes failed to be transcribed by a CTD deleted RNAPII.<sup>118</sup> Together, these results indicate that the CTD is not required for basal transcription of transiently transfected reporter genes, but combined with the Mediator complex the CTD can activate transcription via distal promoter elements in reporter constructs. Transcription of chromatin-packed genes depends entirely on the presence of the CTD.

The analysis of CTD length mutants in *S. cerevisiae* gave a further clue to the genetic and biochemical interaction of the CTD and the Mediator complex. A large multisubunit complex containing the general transcription factor TBP (TATA-binding protein) together with components of the Mediator complex was described as the SRB-complex.<sup>10</sup> Surprisingly, the lethality of CTD truncation mutants could be rescued on the backgrounds of different SRB mutants. This observation supported the notion that the interaction of the CTD and Mediator complex regulates promoter activity both in a negative and in a positive manner. Meanwhile, the composition and structure of the Mediator complex and its manner of interaction with RNAPII have been studied in detail,<sup>119</sup> and the first structure of the CTD–Mediator interaction has been published.<sup>77</sup> The current initiation model suggests that the hypo-phosphorylated CTD is instrumental for the assembly of the initiation complex, but, thereafter, the CTD acts as a strong repressor and prevents the release of RNAPII into the elongation mode. The negative function of the CTD is abolished by its phosphorylation, which allows dissociation of the CTD and Mediator complex.<sup>120</sup> Because of the tight interaction of several CTD heptad-repeats with the Mediator head domain, phosphorylation and release of the CTD probably leads to major changes in the structure of the Mediator complex. Whether and how the dissociation of the CTD and Mediator complex affects other contacts between RNAPII and the Mediator complex, and whether phosphorylation of the CTD is sufficient for dissociation of RNAPII and the Mediator complex, is currently unknown.

#### 4.2. Promoter Proximal Pausing of RNAPII

Promoter proximal sequences have an extremely high RNAPII density in higher eukaryotes, suggesting that RNAPII pauses at the transcription start site (TSS), or downstream thereof, and requires activating signals for elongation. However, the exact process of pausing has yet to be defined.<sup>121</sup> Polymerases at the pause site reveal melted DNA strands, which are sensitive to oxidation by permanganate and can be visualized in DNA footprint experiments.<sup>122</sup> Yet a footprint does not conclusively indicate whether the polymerase is in a “fixed” stage, or if it constantly transcribes, terminates, and reinitiates at a promoter. Unfortunately, no *in vivo* assay is available to discriminate between the two possibilities.<sup>121</sup> The precise definition of a paused RNAPII at the TSS or downstream thereof also includes knowledge of the phosphorylation status of the CTD. In ChIP experiments, high levels of signals for Ser5-P and Ser7-P are

detectable at promoters, while signals for Ser2-P are essentially absent. At least three kinases, Cdk7, Cdk8, and Cdk9, can target the CTD in the initiation complex. Cdk7 is a subunit of the general transcription factor, TFIIH, and was purified as TFIIH-associated kinase activity.<sup>30b</sup> Cdk7 was first described as a Ser5-specific kinase, but recent studies describe Cdk7 as a dual specificity kinase with activity also for Ser7 phosphorylation.<sup>31a,d,32</sup> Interestingly, small molecule inhibitors that specifically inhibit an engineered Cdk7 yeast homologue, Kin28, significantly reduced Ser5 and Ser7 phosphorylation of RNAPII proximal to promoters, but did not abolish the production of mRNA.<sup>31c</sup>

The role of a second kinase, Cdk8/CyclinC, present in the initiation complex is less well-defined. It was first suggested that phosphorylation of the CTD by Cdk8 is inhibitory for transcription and counteracts the function of Cdk7.<sup>123</sup> *In vitro*, Cdk8 phosphorylates Ser2 and Ser5 residues (Ser7 was not tested);<sup>33</sup> however, because Ser2-P is essentially absent in the CTD at promoters, the specificity of Cdk8 and its contribution to CTD phosphorylation at promoters remains elusive. Because Cdk8 has also been recently reported to act in gene activation, its function could also be gene specific rather than general.<sup>36</sup> The third kinase at the promoter, Cdk9, is not part of the initiation complex but is recruited to the CTD after initiation. Phosphorylation of CTD by Cdk9 abolishes pausing and shifts RNAPII into an elongation mode. Phosphorylation by Cdk9 also abrogates the repressive function of other factors in the initiation complex. Because Cdk9 plays a central role in RNA elongation, its function will be described in detail in section 4.4.

#### 4.3. mRNA Capping

Capping of pre-mRNA larger than 20 nucleotides occurs as soon as its 5' end is extruded from RNAPII.<sup>124</sup> The 5' cap consists of a guanine nucleotide linked to mRNA by a 5' to 5' triphosphate bond. The guanosine of the cap is methylated directly after capping by a methyl transferase and termed 7-methylguanylate or m7G. Capping stabilizes the mRNA and provides protection against 5' exonucleases and facilitates proper translation of the message.<sup>125</sup> Capping is observed for RNAPII-, but not RNAPI- or RNAPIII-specific transcripts. Recruitment of the capping enzymes requires the Ser5 phosphorylated CTD of the CTD.<sup>17,86a,c,102</sup> Interestingly, although Ser5P is essential for the recruitment of capping enzymes in all organisms with a CTD, the recruitment pathway can differ significantly. The recruitment to the Ser5 phosphorylated CTD occurs in metazoa through the RNA guanylyltransferase domain of a bifunctional enzyme with an additional RNA triphosphatase domain.<sup>86b,102</sup> In contrast, the recruitment in *S. pombe* occurs via an additional interface located on the multihelical Foot-domain of Rpb1. Although the Ser5-P interface and the Foot-domain interface individually contribute only weakly to the interaction with the capping enzyme, in combination they form a stable complex with the capping enzyme.<sup>126</sup> The completion of cap formation in *S. pombe* is achieved by recruitment of the 5'-cap methyltransferase, Pcm1, which requires binding to the carboxy-terminal extension of Cdk9 in addition to the CTD.<sup>31e,46</sup> Notably, the recruitment of the capping enzyme is apparently the only essential function of Ser5 in *S. pombe*. Mutants with substitutions of all Ser5 residues to alanine are lethal in *S. pombe* but their viability is rescued after fusion of the capping enzyme to the carboxy-terminus of CTD.<sup>17</sup> The last example

nicely illustrates that the CTD functions as a binding platform for transcription process-coupled factors and that an alternative recruiting route for a binding factor can substitute for the CTD.

#### 4.4. Elongation

The release of RNAPII into an elongation mode requires phosphorylation of CTD Ser2. The recruitment of Ser2-specific kinases is Ser5-P dependent, either in a direct or in an indirect way. In *S. cerevisiae*, two enzymes can phosphorylate Ser2: Bur1,<sup>127</sup> which is recruited directly to RNAPII by Ser5-P,<sup>128</sup> and Ctk1.<sup>59</sup> Equivalent kinases for Ser2 phosphorylation have been identified in *S. pombe* (Figure 7). The Bur1 orthologue, Cdk9, is directed to the CTD by the Ser5-P dependent capping enzyme and Lsk1, which is responsible for the majority of Ser2 phosphorylation, similar to Ctk1 in *S. cerevisiae*.<sup>31c,59,129</sup> A common non-CTD substrate of Bur1 and Cdk9 is the subunit Spt5 of the DRB (5,6-dichloro-1-β-D-ribofuranosylbenzimidazole) sensitivity-inducing factor (DSIF), which contains a carboxy-terminal repeat domain similar to the CTD.<sup>130</sup> DSIF is recruited to RNAPII together with the negative elongation factor, NELF, a complex conserved in higher eukaryotes.<sup>131</sup> Both factors act as inhibitors of transcription elongation.<sup>132</sup> Phosphorylation converts DSIF and NELF into positive factors for elongation.<sup>132a,133</sup> Previously, it was thought that subunit Cdk9 of the positive transcription elongation factor b (P-TEFb) combines the functions of *S. cerevisiae* Bur1/Ctk1 and *S. pombe* Cdk9/Lsk1 in mammals. However, additional CTD Ser2 kinases have been recently identified in higher eukaryotes. Cdk12 and Cdk13 show Ser2 kinase activity in vitro, and Cdk12 was proposed to contribute the majority of Ser2-P on elongating RNAPII.<sup>48a,134</sup> Cdk12 specifically promotes expression of DNA damage response genes,<sup>48b</sup> while the function of Cdk13 during the transcription cycle is still unclear.

Most of our knowledge of Ser2 phosphorylation originates from work with P-TEFb. This factor was first identified as a DRB-sensitive RNA elongation factor. The production of full-length transcripts in vitro and of functional mRNA in vivo is sensitive to the kinase inhibitor, DRB. Price and Marshall isolated a cellular activity from *Drosophila* cell extracts that functioned in a DRB-sensitive manner to allow RNAPII transcription complexes to efficiently synthesize long transcripts.<sup>38b,135</sup> Shortly after its description, P-TEFb achieved fame as a key regulator of human immunodeficiency virus (HIV) propagation. P-TEFb together with viral TAT protein is recruited to a stem-loop structure (TAR) in the nascent transcript derived from the viral LTR promoter and activates elongation by phosphorylation of the CTD.<sup>40,136</sup> Subsequently, promoter-proximal recruitment of P-TEFb for activation of elongation has been described for many cellular promoters.<sup>128,133a,137</sup> However, it is puzzling that Ser2-P signals are generally extremely low at the 5' end of genes in ChIP experiments. Possible explanations could be (i) that RNAPII leaves the promoter immediately after receiving the Ser2-P modification, (ii) that Ser2-P-specific mAb's fail to recognize the Ser2-P mark in promoter proximal regions, either because the high levels of Ser5-P and Ser7-P marks in the CTD simply shield the mAb binding site, or (iii) because the binding of Ser2-P to cellular factors at the promoter prevents mAb binding. A possible candidate for the latter possibility is the FUS (fused in sarcoma) protein. FUS possesses Ser2-P binding activity, and siRNA-mediated knockdown of FUS leads to a strong increase of Ser2-P signals in the promoter region of two-thirds of expressed genes in human HEK293T/17 cells.<sup>138</sup>

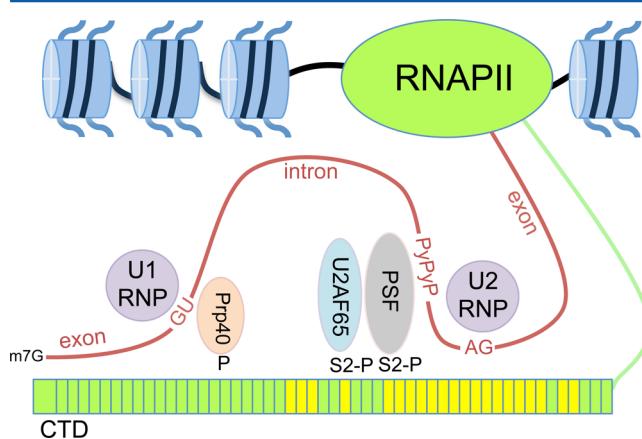
The recruitment and activation of P-TEFb is subject to highly sophisticated control mechanisms. The laboratories of Bensaude and Zhou identified an abundant and evolutionarily conserved small nuclear RNA, 7SK, as a specific P-TEFb associated factor. 7SK snRNA in complex with HEXIM inhibits general and HIV-1 Tat-specific transcriptional activities of P-TEFb in vitro and in vivo by inhibiting the kinase activity of Cdk9 and preventing recruitment of P-TEFb to the HIV-1 promoter.<sup>139</sup> HIV-1 Tat or cellular factors can release P-TEFb from the 7SK/HEXIM complex to allow CTD Ser2 phosphorylation of promoter-proximally paused RNAPII. As for other Cdks, T-loop phosphorylation of Cdk9 is required for activation of the kinase activity of Cdk9. Interestingly, as was similarly observed for cell cycle-specific Cdks, the T-loop of Cdk9 is phosphorylated by Cdk7, a subunit of the general transcription factor, TFIIE.<sup>140</sup> Cdk7 apparently fulfills a dual function in the initiation complex. By phosphorylation of TFIIE, Cdk7 regulates the release of TFIIE and the recruitment of DSIF to RNAPII and, thereby, establishes the pause of RNAPII at the promoter. The phosphorylation of the Cdk9 T-loop by Cdk7 subsequently regulates the release of RNAPII into the elongation mode.<sup>140,141</sup> Thus, the critical regulatory step to allow elongation appears to be the recruitment of P-TEFb to promoters, which is facilitated either by specific transcription factors like c-Myc<sup>137b,142</sup> and NFκB<sup>143</sup> or by factors that bind to specific histone modifications in chromatin, like Brd4. Brd4 is a bromodomain-containing protein that binds to acetylated chromatin with its N-terminal dual bromodomains and interacts with P-TEFb through its C-terminal P-TEFb interaction domain, PID.<sup>144</sup> An increase in Brd4 expression leads to an increased P-TEFb-dependent phosphorylation of the CTD, and the recruitment of P-TEFb to promoters is Brd4-dependent.<sup>137a,139d,144,145</sup> Recently, it was suggested that Brd4 is an atypical protein kinase that phosphorylates CTD Ser2 in the absence of P-TEFb.<sup>49c,146</sup> The individual contribution of P-TEFb and Brd4 kinases to Ser2 phosphorylation however remains unclear if P-TEFb is recruited to promoters via the Brd4 pathway.

#### 4.5. Cotranscriptional Splicing

The regulation and coordination of the mRNA splicing process is probably one of the central functions of the CTD. However, CTD function is not restricted to splicing but also includes the processing of small nuclear RNAs (snRNAs) and other noncoding RNAs (ncRNAs). The first evidence for an involvement of the CTD in splicing came from reporter gene assays with an α-amanitin resistant RNAPII. α-Amanitin binds and specifically inhibits the activity of RNAPII at 2 μg/mL, while RNAPI and RNAPIII are resistant to α-amanitin, or are inhibited at 50-fold higher concentrations, respectively.<sup>147</sup> The availability of α-amanitin resistant mutants of RNAPII with point mutations in Rpb1<sup>148</sup> allowed the construction of Rpb1 expression vectors for the analysis of CTD mutants.<sup>11</sup> After expression of an α-amanitin resistant RNAPII with a manipulated CTD, the endogenous RNAPII was inhibited by addition of α-amanitin to the cell medium. Reporter gene constructs transfected into these cells 24 h later could thus only be transcribed by the recombinant RNAPII.<sup>117</sup> This system provided the first evidence that primary transcripts fail in efficient splicing, 5' capping, and 3' processing, if Rpb1 lacks the CTD.<sup>85,86c</sup> CTD deletion mutants showed that approximately one-half of the length of the CTD is sufficient for proper splicing, 5' capping, and 3' processing of reporter-gene-

derived transcripts.<sup>149</sup> Whether this applies to all cellular genes is currently unknown. Another analysis of deletion mutants revealed that the C-terminal 10 amino acids extension following the last, S2nd repeat is required for CTD stabilization in mammalian cells. The CTD in Rpb1 mutants lacking the S2nd repeat and its extension is sensitive to degradation by unknown proteases, which results in the so-called RNAPIIB form lacking the CTD.<sup>150</sup> The S2nd repeat can recruit the cellular tyrosine kinase, c-Abl, to the CTD for phosphorylation of the Tyr1 residues in vitro.<sup>50,51,151</sup> Whether Tyr1 phosphorylation of the CTD protects it from degradation to the IIB form of RNAPII in vivo is currently unclear. The absence of the S2nd repeat in Rpb1 expression constructs leads to degradation of the CTD in vivo and can therefore affect the interpretation of results for CTD deletion mutants.<sup>152</sup>

Splicing of pre-mRNA is catalyzed by the spliceosome, a large ribonucleoprotein (RNP) complex. The spliceosome is composed of five small nuclear RNP (snRNP) particles and is assembled with a large number of auxiliary factors.<sup>153</sup> In vitro experiments revealed that the hyper-phosphorylated form of RNAPII, but not the hypo-phosphorylated form, supports early steps of spliceosome assembly and strongly enhances the splicing rate,<sup>154</sup> suggesting that CTD phosphorylation is critical for the recruitment and assembly of an active splicing complex. The hypo-phosphorylated form of RNAPII suppresses the basal splicing activity observed in nuclear extracts, possibly by titrating splicing factors.<sup>154</sup> Spliceosome assembly starts with the recruitment of the U1 snRNP to the 5' splice site (Figure 16). In vivo, U1 snRNPs can be recruited to RNAPII in a



**Figure 16.** Splicing of pre-mRNA. Phosphorylation of the CTD coordinates recruitment of splicing factors to the 5'- and 3'-splice sites of introns and formation of the spliceosome. It should be emphasized that it is currently unknown how many and which of the depicted S2 heptad-repeats of the mammalian CTD have to be phosphorylated to support splicing. This statement applies also to Figures 17–23.

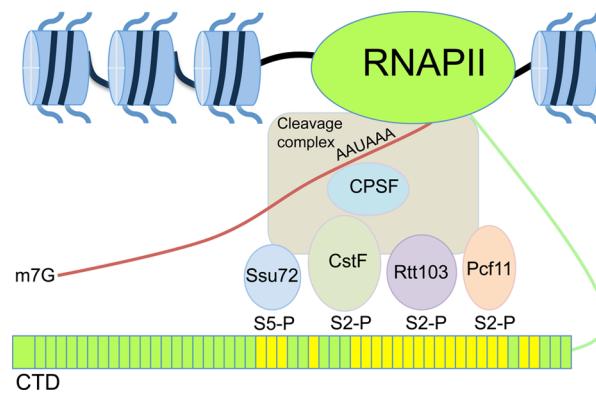
splicing-independent manner.<sup>155</sup> The recruitment occurs via Prp40, the first spliceosomal factor reported to bind to the Ser5 phosphorylated CTD,<sup>87,156</sup> and may also occur with recognition of the 5' splice site.<sup>157</sup> Prp40 also binds to TFIIF<sup>158</sup> and TAF15,<sup>159</sup> but these two factors localize to the promoter and are probably not associated with RNAPII during elongation. Interestingly, while knockout of Prp40 in yeast is lethal, a mutant with deletion of the CTD interaction, WW domains is viable, and shows no defect in splicing and U1 or U2 snRNP recruitment in vivo, or a defect in splicing complex assembly in vitro.<sup>160</sup> However, delays in cotranscriptional U5 snRNA and

Prp19 recruitment suggest that Prp40 WW domains assist in late spliceosome assembly.<sup>160</sup> If other known splicing factors, as CA150, may contribute to 5' splice site recognition via interaction of its FF domain with the phosphorylated CTD has not been analyzed.<sup>161</sup>

Recruitment of U2 snRNP and of the auxiliary U2 snRNP factor, U2AF, occurs in a splicing-dependent manner<sup>88,162</sup> and is not observed in the absence of a functional intron.<sup>163</sup> U2AF binds to the polypyrimidine tract of the 3' splice site and recruits the U2 snRNP to the branch point to form the prespliceosome complex. The recruitment and assembly of the U2 snRNP at the 3' splice site is supported by further auxiliary factors, such as PSF,<sup>89</sup> which bind to the Ser2 phosphorylated CTD (Figure 16). The emerging picture is that the assembly of functional spliceosome complexes is a highly regulated process. In vitro splicing experiments show that many splice sites are weak; therefore, their proper recognition in vivo requires additional factors, which might be recruited by specifically modified forms of the CTD or by other means (see also section 5.3: Epigenetic Regulation of Alternative Splicing). Splicing is a transcription-coupled process, but at the same time the transcription rate appears to be coupled to splicing. Impairment<sup>164</sup> or improvement<sup>165</sup> of splicing reduces or increases, respectively, the transcription rate of a gene.

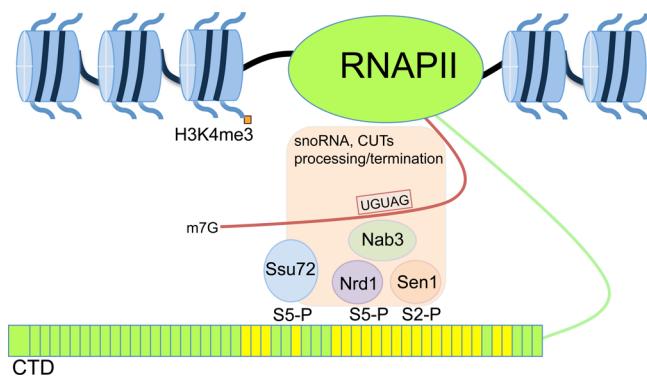
#### 4.6. 3' Processing of RNA

Processing of RNAPII transcripts occurs along distinct, gene class-specific pathways, either via the polyadenylation-dependent pathway (Figure 17), the Nrd1-Sen1-Nab3 pathway (Figure

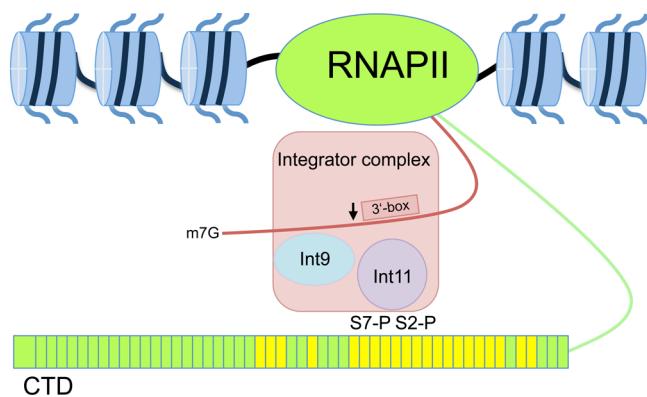


**Figure 17.** Poly(A)-dependent cleavage of mRNA. The sequence, AAUAAA, is recognized by CPSF (cleavage and poly adenylate factor), a major component of the cleavage and poly adenylate complex. Several factors of the complex interact with phosphorylated CTD.

18), or the Integrator-dependent pathway (Figure 19). The first pathway is used for all mRNAs (except for specifically expressed S-phase histone mRNAs), and also for many long noncoding RNAs (ncRNA), while the second pathway is used for many small noncoding RNAs, like snoRNAs, or CUTs (cryptic unstable transcripts).<sup>4g,96b,166</sup> The third pathway is specifically used for snRNAs.<sup>57,95,167</sup> Transcript 3' processing along the polyadenylation pathway involves a specific polyadenylation signal, AAUAAA, in the nascent RNA that is recognized by the polyadenylation complex. The requirement of the CTD for 3' processing and the interaction of the CTD with the polyadenylation complex factors, CPSF and CstF, was first demonstrated in mammalian cells with reporter gene assays<sup>85</sup> and subsequently in *in vitro* splicing assays.<sup>168</sup> Several



**Figure 18.** Nrd1-Sen1-Nab3-dependent pathway for 3' processing of RNA. The sequence, UGUAG,<sup>171a</sup> is recognized by Nrd1-Sen1-Nab3. Nrd1 and Sen1 interact with Ser5-P and Ser2-P, respectively.



**Figure 19.** snRNA processing. The Integrator complex is required specifically for processing snRNAs. Its subunit, Int11, interacts specifically with Ser7-P/Ser2-P phosphorylated CTD.

factors of the polyadenylation-dependent cleavage pathway, in particular PCF11,<sup>74a,92</sup> Rtt103,<sup>93,105</sup> and CstF50,<sup>152a</sup> display binding activity for specific CTD Ser2-P marks, which is consistent with the strong increase of Ser2-P signals at the polyadenylation site seen in ChIP experiments. The phosphatase, Ssu72, removes the Ser5-P marks toward the 3' end of genes and also shows binding specificity for Ser5-P.<sup>52,75,107</sup> Thus, CTD phospho-marks coordinate the binding of several factors involved in 3' processing. This may explain why factor binding platforms of several functional units and a minimal length of CTD are required for the coordination of 3' processing and termination.

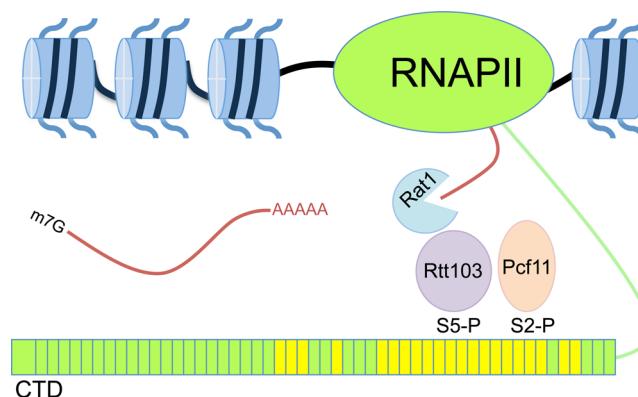
The second, Nrd1-dependent pathway directs poly(A)-independent termination of snoRNAs and CUTs (Figure 18). Nrd1 interacts physically with phosphorylated CTD Ser5, and functionally with trimethylated histone H3 and the PAF complex.<sup>169</sup> This explains its recruitment to promoters and the termination of short cryptic unstable transcripts. The association of Nrd1 with RNAPII is lost downstream of promoters and is associated with increasing Tyr1 phosphorylation of the CTD, probably to prevent premature transcript termination if RNAPII is in the elongation mode.<sup>23c</sup> Nrd1 also regulates the termination and proper processing of snoRNAs.<sup>96b,166a,b,f</sup> The transacting factors involved in Nrd1-dependent and poly(A)-dependent termination differ but also show some overlap. Nrd1-dependent termination specifically requires the RNA-binding Nab3 protein and the putative RNA helicase, Sen1, and RNAPII CTD phosphorylation is involved

in regulating the transcription termination choice.<sup>166e</sup> The processing/termination is further specified by specific sequence motifs in the transcript,<sup>2d,170</sup> which are probably also of relevance, if snoRNAs are processed from intron-embedded sequences independently of transcription termination.<sup>171</sup> Similar to poly(A)-dependent termination, the Nrd1-termination pathway is also dependent on phosphatase, GLC7, a subunit of the cleavage and polyadenylation factor,<sup>172</sup> and the pyrophosphatase, ESS1.<sup>173</sup>

A third, Integrator-dependent processing-termination pathway has been described for snRNA genes. Phosphorylation of CTD Ser7 is required for the recruitment of the Integrator complex (Figure 19). In addition, RNAPII-associated protein 2 (Rrap2)<sup>174</sup> specifically recognizes the phospho-Ser7 mark in the CTD and also interacts with Integrator subunits.<sup>57,167b</sup> Rrap2 is a putative Ser5 phosphatase, and removal of Ser5-P marks is probably a prerequisite for proper binding and function of the Integrator complex. The crystal structure of Rtr1 (regulator of transcription), the yeast homologue of Rrap2, has been determined and suggests that Rrap2 facilitates an interaction with a putative phosphatase, rather than having phosphatase activity itself.<sup>56</sup> Processing snRNAs is also sensitive to Ser2-specific kinase inhibitors, which is consistent with the requirement of Ser2-P marks for the recruitment of the Integrator complex and processing of snRNAs.<sup>4f,g,95,175</sup>

#### 4.7. Transcription Termination

In contrast to our knowledge about mRNA 3' end formation, little is known about signals that control transcription termination of RNAPII in eukaryotes. Transcription continues downstream of the poly(A)-signal and can terminate up to several kb downstream and at multiple sites.<sup>176</sup> Yet, the 3' processing of RNA is a critical determinant for termination. Deletion of the poly(A)-signal in Simian virus 40, a small DNA virus with a circular genome, results in a failure of transcription termination and leads to multiple rounds of transcription.<sup>177</sup> Recent ChIP experiments in mammalian cells revealed that RNAPII occupancy of genes is often much higher downstream of the poly(A)-site as compared to sequences between the TSS and poly(A)-site.<sup>2b,178</sup> The current torpedo model (Figure 20) suggests that the unprotected 5' end of processed RNA is targeted by a 5' → 3' exonuclease that degrades the RNA behind the transcribing RNAPII and that termination is



**Figure 20.** Termination by the torpedo mechanism. After 3' cleavage, the uncapped 5' end of the nascent RNA is recognized and degraded by a 5' → 3' exonuclease, Rat1. Phosphorylation-dependent binding of termination factors, Rtt103 and Pcf11, supports recruitment of Rat1 (Xrn2 in mammals).

impaired in the absence of the exonuclease.<sup>179</sup> However, the torpedo model fails to provide mechanistic explanations of how liberation of RNAPII from the template is achieved and how allosteric changes may contribute to termination. Likewise, the role of the CTD and the removal of all phospho-marks in the CTD by the phosphatases, Fcp1 and Ssu72, as well as the looping of the termination site to the promoter<sup>54a,180</sup> are still controversial in terms of contributing to the liberation of RNAPII from the template. For further reading, we recommend a recent review by Richard and Manley.<sup>166h</sup>

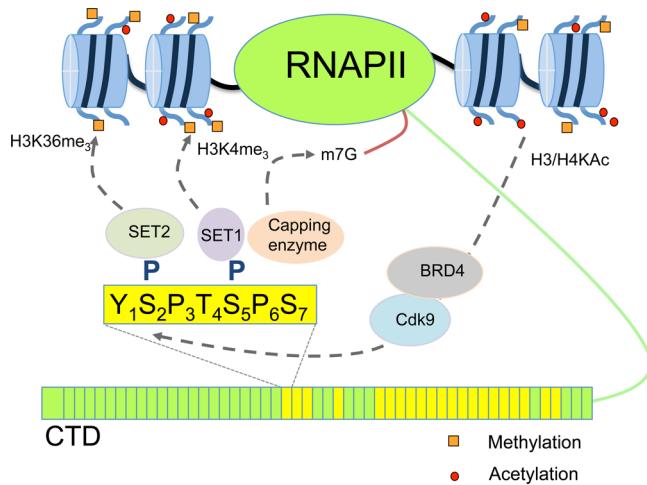
## 5. CTD-CHROMATIN CROSSTALK

Similar to the histone code proposed by Strahl and Allis,<sup>181</sup> the existence of a CTD code has been postulated by Buratowski.<sup>4c</sup> The term code, rather than constituting a general code, is used here to describe the formation (either stable or transient) of structures in histone tails or heptads of the CTD by post-translational modifications to regulate the action of downstream effectors. The combination of modifications (both in histone tails and CTD heptads) specifies and limits the type of action. Because CTD and histone tails coevolved over hundreds of millions of years (as discussed in section 6.1: Gain and Loss of Introns), specific signatures of histone modification may communicate with specific signatures of CTD modification and vice versa. The histone code comprises reversible modifications, including acetylation, phosphorylation, methylation, and ubiquitination, which can remodel the interaction surfaces of chromatin and the binding and dissociation of cellular factors in response to cellular signals. In fact, specific examples show evidence for a linkage of the phosphorylation state of the RNAPII CTD with the modification state of histone tails.

Several newly described modification marks expand the complexity of the CTD, with Ser7 phosphorylation and methylated arginine being the first examples for CTD modifications that are read in a gene-type-specific manner. Ser7-P plays a specific role in recruitment of the Integrator complex, which is required for snRNA 3' processing,<sup>167c</sup> and Arg1880 methylation regulates the steady-state levels of specific small RNAs.<sup>67</sup> Yet how can specific modification patterns be established in the CTD in a gene-specific manner, and where does the information for patterns come from? Does the information, at least in part, arise in a tissue-specific way from epigenetic tags associated with chromatin, and does this histone-encoded information serve to write a CTD code for transcription and maturation of RNA? The CTD and chromatin have an intimate spatial relationship during all stages of transcription. It will be, therefore, a challenging task to get deeper insight into the relationship of these two structures and to uncover if chromatin-modifying processes can be instructed by the CTD and vice versa.

### 5.1. Modification of Chromatin and the CTD Is Mutually Controlled

How are CTD modifying enzymes recruited to chromatin? The dual bromodomain-containing protein, BRD4, binds to acetylated tails of histones H3 and H4<sup>139d,145c,182</sup> (Figure 21). The modification of histone tails depends on the association and activity of locus-specific HATs (histone acetyltransferases) and HDACs (histone deacetylases). HATs and HDACs co-reside in most active and inactive genes,<sup>183</sup> and regulate the pattern of histone acetylation, thereby mediating transcription elongation.<sup>184</sup> As described above, BRD4 recruits



**Figure 21.** Crosstalk between CTD and chromatin. Phosphorylation of Ser2 and Ser5 residues in the CTD facilitates the recruitment of the methyltransferases, SET1 and SET2, respectively, and methylation of specific histone residues. Inversely, acetylated histone tails recruit Cdk9 and BRD4 kinases for phosphorylation of CTD Ser2 residues.

P-TEFb and, thereby, induces CTD phosphorylation. Disruption of the bromodomain-histone acetylation interaction by JQ1, a small-molecule bromodomain inhibitor, results in decreased BRD4 binding, reduced Ser2 phosphorylation, and reduced activity of lineage-specific gene expression in T cell differentiation.<sup>145j</sup> A recent report describes BRD4 itself as an atypical kinase with specificity toward Ser2 phosphorylation.<sup>49c,146</sup> Thus, the pattern of histone acetylation is involved in the recruitment of CTD-specific kinases and, thereby, regulates the activity of genes (Figure 21).

Inverse examples have been described for histone-modifying enzymes, which are recruited by CTD phospho-marks to the transcription machinery. Ser5-P binds and recruits the H3K4 methyltransferase, Set1, to the CTD together with the PAF complex in yeast<sup>79</sup> (Figure 21). Set1 trimethylates H3K4 of promoter proximal histones and is a hallmark of activated genes. Downstream of promoters, Set1 dimethylates H3K4, and the H3K4me2 mark is involved in the recruitment of HDACs to chromatin<sup>185</sup> (see also section 5.2). Similar to yeast, recruitment of H3K4-specific methyltransferases by Ser5-P has been described for mammalian cells.<sup>186</sup> Set1 belongs to the family of mammalian MLL genes, which undergo chromosomal translocation in mixed-lineage leukemia cells. MLL fusion proteins alter the dynamic association of transcriptional regulators with genes critical for leukemia.<sup>187</sup> Further examples for chromatin modifying enzymes, which are recruited by phosphorylated CTD, are the H3K36 methyltransferase, Set2, the histone deacetylase, Rpd3S, and the histone chaperone, Spt6. These factors are critical to the role that chromatin plays in transcription and will be discussed in the next section.

From the examples described above, a picture emerges of an intensive crosstalk between chromatin and the CTD. Both structures are characterized by complex patterns of modifications and association with enzymes that steadily change these patterns in transcribed genes by adding new or removing existing modifications. During transcription, the function of the CTD can be considered as that of an assembly line in a car factory. The RNA (car) is tethered to the CTD (assembly line), which allows the coordinated modification, splicing, and processing (production) of RNA. The production plan for

RNA can vary (depending on the car model that is produced) and is provided by chromatin modifications (working plan). Thus, the CTD is an intelligent platform for RNA maturation. It receives signals and binds factors to fulfill specific operations with RNA, and to send signals after the operations have been executed. As we will see in the next section, the action of the CTD is not restricted to the maturation of RNA; it also supports RNAPII transcription through chromatin.

### 5.2. Transcription through Chromatin

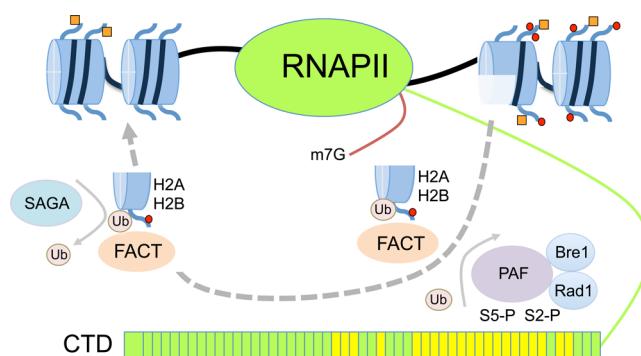
The nucleosomal structure of chromatin provides a natural barrier for RNAPII activity. Chromatin remodeling machines and modification of histone tails by acetylation, methylation, and ubiquitination fulfill critical tasks to allow RNAPII transcription through chromatin. The current model suggests that a nucleosome is not entirely disassembled and removed from DNA to allow RNAPII to pass. Much of our knowledge about transcription of RNAPII through nucleosomes comes from *in vitro* transcription experiments with nucleosomal templates. In *vitro*, up to 95% of nucleosomes survive single-round transcription by purified RNAPII. Nucleosome octamers lose only one H2A/H2B hemidimer, while the resulting hexamer consisting of two H3/H4 dimers and one H2A/H2B hemidimer remains at its position ensuring that the epigenetic information of the nucleosome is not lost. Mechanistic details for the mode of passage of RNAPII through chromatin and in particular through hexameric nucleosomes are described in recent reviews from the laboratories of Studitsky, Kashlev, Bustamante, and Lis.<sup>188</sup>

The CTD supports transcription through chromatin by assisting in the removal of H2A/H2B dimers from nucleosomes and their subsequent replacement. In *vivo*, this process is regulated by the cellular factor FACT (facilitates chromatin transcription), comprising Spt16 and SSRP1 proteins<sup>189</sup> (Figure 22). The action of FACT requires the monoubiquiti-

trimethylation is required for the repression of cryptic transcription initiation, but the mechanisms involved are not completely understood. In yeast, H3K36me3 is involved in the recruitment of the histone deacetylase, Rpd3S, which removes acetyl-residues from histone tails,<sup>191</sup> hindering the passage of RNAPII through chromatin. Acetylation of histone tails decreases the compaction of chromatin but also increases the risk for erroneous initiation. Thus, the erasure of acetyl-residues in histone tails of nucleosomes downstream of RNAPII is important and facilitated by direct binding of HDACs to the phosphorylated CTD.<sup>82b,191b</sup> Interestingly, the inhibition of FACT function and increased cryptic initiation is observed after down-regulation of the methyltransferase, SetD2, in mammalian cells.<sup>192</sup> This suggests that the proper transition of the H2A/H2B dimers and the deacetylation of histone tails are critical steps during transcription, steps that are coordinated by the phosphorylated CTD.

### 5.3. Epigenetic Regulation of Alternative Pre-mRNA Splicing

The mechanisms of alternative splicing are not well understood. Current models suggest that alternative splicing is regulated by splicing enhancers and silencers and by mechanisms controlling the RNA synthesis rate.<sup>7e,193</sup> Given that the majority of mammalian genes undergo alternative splicing and that alternative splicing is an integral part of cell differentiation and gives rise to large number of splice variants in various tissues, the current concepts do not suffice to provide a general explanation of alternative splicing regulation. As described in section 6.2, a single gene can give rise to tens of thousands of splice variants. The decision process by which a cell can produce a single splice variant while thousands of other variants are suppressed is entirely unclear. The underlying mechanism must mark included or excluded exons in a negative or positive way, respectively. A combination of both alternatives is also feasible. As discussed in the section above, many signals contribute to the definition of splice sites: (i) the sequence and structure of the RNA, (ii) factors recruited to the nascent RNA, (iii) modifications in the CTD, (iv) the chromatin of the transcribed template, and (v) the methylation state of DNA (the last point is discussed in section 6.2). All of these variables can contribute to the definition of alternative splice sites. However, the contribution of some of these variables to the definition of alternative splice sites is limited. The sequence of the RNA cannot vary, and changes in its secondary structure or its modification also require factors in trans. Thus, the RNA and its structure for determining the usage of alternative splice sites are questionable. Binding of specific factors to the nascent transcript is another possible option, and examples for this option have been reported. However, this mechanism would require a huge number of different factors to regulate all alternative splicing events in cells. We think that factors specifically binding to RNA are the exception rather than the rule for choosing alternative sites. The contribution of the CTD in making the decision for including or skipping alternative exons is unclear. It could well be that some of the heptad-repeats in the CTD, particularly some of the nonconsensus repeats, become specifically modified at the promoter and that this information is used later for exon selection and the formation of a splice complex. However, no evidence for such a mechanism currently exists. Thus, RNA-binding elements, the control of RNAPII elongation rates, nor CTD modifications

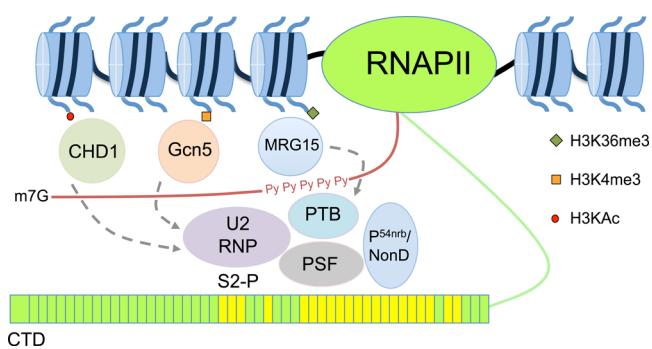


**Figure 22.** Transcription of RNAPII through chromatin. Phosphorylation of CTD Ser2 and Ser5 residues allows the recruitment of the PAF-Bre1-Rad1 complex to facilitate histone H2B ubiquitination. FACT facilitates the dissociation of a H2A/H2B dimer from the nucleosome. The SAGA complex regulates the removal of ubiquitin and acetyl-residues from histone tails and the reincorporation of the H2A/H2B dimer into nucleosomes behind RNAPII.

nation of H2B (H2Bub1) by Bre1/Rad6,<sup>190</sup> which is recruited to the transcription machinery by the PAF complex in a Ser5-P dependent manner. FACT probably not only facilitates the dissociation of H2A/H2B dimers from nucleosomes but may also be involved in the regulation and removal of ubiquitin and reincorporation of the H2A/H2B dimer into the hexameric nucleosome. Previous studies have shown that H3K36me3

appear to be sufficient to explain the regulation of alternative splicing.

The evidence that chromatin structure is involved in the regulation of alternative splicing comes from individual observations for specific genes (for review, see ref 7d), and more recently from the observation that nucleosomes are not randomly distributed genome wide. Nucleosomes are particularly enriched at intron–exon boundaries, thereby marking the exons.<sup>194</sup> The enrichment of nucleosomes at exon–intron boundaries suggests that the modification of histones may contribute to the marking of alternative splice sites and that the information for using or skipping a splice site is encoded in the tails of histones. A hint for a direct role of histone modifications in the regulation of alternative splicing comes from comparative mapping of histone modifications along genes. These studies revealed a strong correlation between specific histone modifications at exon–intron junctions and the outcome of splicing.<sup>7d,195</sup> The authors found that the high prevalence of H3K36me3 marks at exon–intron junctions creates a chromatin platform for the binding factors that regulate alternative splicing. In the case of exons whose alternative splicing depends on PTB (polypyrimidine tract-binding protein), high levels of H3K36me3 modification recruit the chromatin-binding factor, MRG15, which binds PTB and, thereby, facilitates splicing (Figure 23). However, splicing



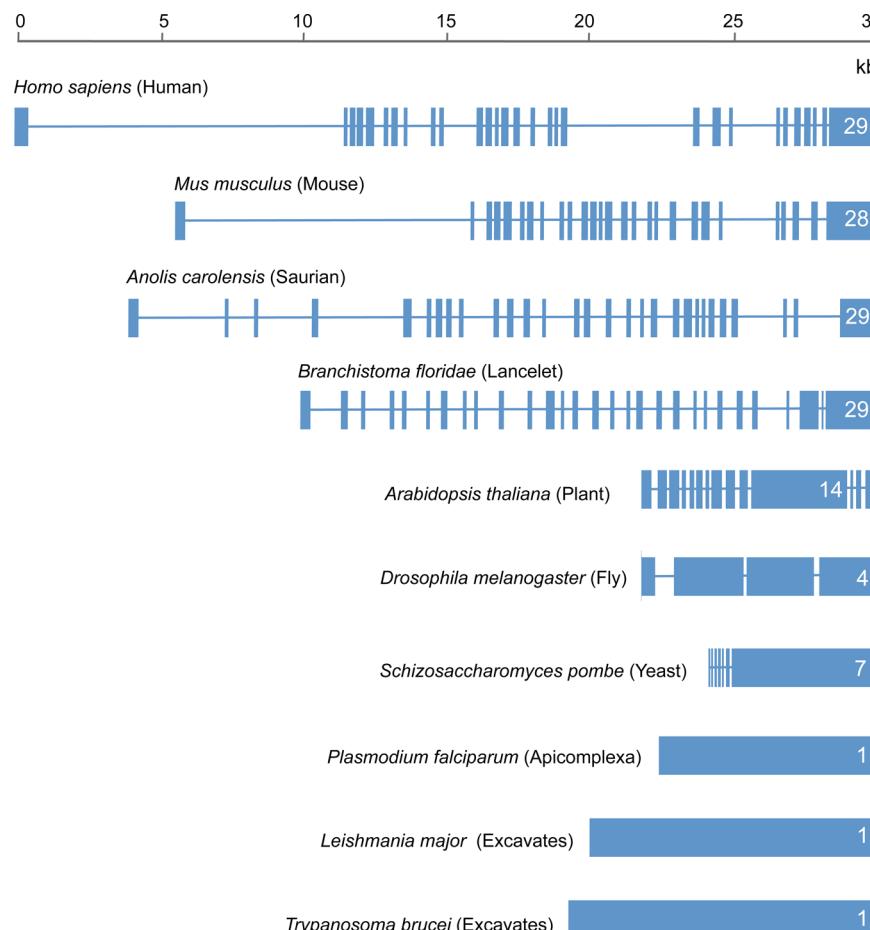
**Figure 23.** Model for chromatin assisted alternative splicing. CHD1, Gcn5, and MRG15 are recruited to chromatin by specific modifications of histone tails and contribute to the recognition and/or specification of the intron's 3' splice site and formation of the spliceosome.

decisions may also be influenced by the methylation state of DNA. Binding of the transcription factor, CTCF, promotes the inclusion of exons with weak splice sites by local pausing of RNAPII. Binding of CTCF and the inclusion of weak exons is inhibited, if the binding site of CTCF is DNA methylated.<sup>196</sup> Retaining alternative splice information by epigenetic mechanisms is clearly superior to other mechanisms. The clear advantage of an epigenetic mechanism is the direct local availability of the information for the splicing machinery, the stability of the information, but also the option of reprogramming this information if required. In addition to MRG15, other histone modification-binding chromatin proteins, such as Gcn5<sup>197</sup> and CHD1,<sup>198</sup> have been reported to interact with splicing factors.<sup>7d</sup> The challenging question will be whether factors as CHD1, Gcn5, and MRG15 can regulate the modification state of CTD and thereby contribute to the formation of stable splice sites.

## 6. ACCELERATION OF EVOLUTION IN EUKARYOTES BY THE CTD

### 6.1. Gain and Loss of Introns

The number of introns in eukaryotes varies from none or only few introns per genome in some unicellular organisms to several hundred thousand introns in mammalian cells.<sup>199</sup> The excision of introns from pre-mRNA requires the action of the spliceosome, a huge complex consisting of five small nuclear RNAs (snRNAs) (U1, U2, U4, U5, and U6), which participate in several RNA–RNA and RNA–protein interactions, and more than 150 proteins.<sup>153b</sup> The vast majority of introns require the spliceosomal machinery for removal, while only a small minority can be excised by self-splicing. The recognition of splicing junctions in introns by the spliceosome is facilitated by several *cis*-elements in intronic sequences of pre-mRNA, including the splice-donor and splice-acceptor sites at the 5' and 3' sites of introns, respectively. Introns can differ considerably in size; the smallest introns can have a size in the range of 30 nucleotides, while 5% of mammalian introns have a size >200.000 nucleotides. There is a long-standing debate about the origin of introns and the question of whether introns existed very early (intron-early theory) or developed late in evolution.<sup>200</sup> Current models suggest that the genome of the last common ancestor of eukaryotic cells was already intron-rich,<sup>201</sup> but spliceosomal splicing was probably not well established at that time point, with splicing of the first introns occurring by more simple mechanisms, including self-splicing. The second, intron-late theory suggests that the large majority of introns evolved after an efficient spliceosomal splicing machinery had evolved, and that the late gain of introns was accompanied with a gain of intron-related functions (discussed in the section below). Yet it is also becoming evident that introns were lost during evolution, and organisms with only a few introns, such as *S. cerevisiae*, have lost their introns. This loss is accompanied by a not yet understood mechanism of genome compaction. We will discuss below the gain and loss of introns for the highly conserved Rpb1 gene, which encodes the largest subunit of RNAPII (Figure 24). At the protein level, Rpb1 is 99.9% conserved in mammals, ~98% in vertebrates, ~70% in metazoa and some unicellular organisms, and bacteria still have eight conserved domains A–H (see Figure 1B). The genomic organization of the human *rpb1* gene differs considerably in eukaryotes (Figure 24). The size of the genomic locus of *POLR2A* is 30 kb, and the coding sequence is interrupted by 28 introns. While mouse, saurian, and lancelet also have 29 or 28 exons, the size of their *rpb1* gene loci and the size of their introns show marked differences. In particular, the organization of the *rpb1* introns in saurian and lancelet has little in common with that in human and mouse. More distant organisms, such as plant, fly, and yeast, display *rpb1* gene loci <10 kb, and the number and size of introns were reduced to 14 in *Arabidopsis thaliana* and 4 in *Drosophila melanogaster*. The *rpb1* genes in *Plasmodium falciparum*, *Leishmania major*, and *Trypanosoma brucei* are characterized by an entire absence of introns. These organisms either have a very rudimentary CTD with only a few nonconsensus repeats, or the CTD structure is entirely absent. The lack of a CTD in *Leishmania major* and *Trypanosoma brucei* is accompanied by strong differences in N-terminal tails of histones H3 and H4 (Figure 25). The structure of histone tails in metazoa and many unicellular eukaryotes, such as *Schizosaccharomyces pombe*, *Aspergillus nidulans*, or *Plasmodium falciparum*, is highly conserved, but shows



**Figure 24.** Number, size, and distribution of introns in the Rpb1 gene of various organisms. The number in the last exon indicates the number of exons. Size of the gene locus is given in kilobases (kb). NCBI GenBank gene ID: *Homo sapiens*, 5430; *Mus musculus*, 20020; *Anolis carolinensis*, 100561860; *Branchiostoma floridae*, 7235237; *Arabidopsis thaliana*, 829734; *Drosophila melanogaster*, 32100; *Schizosaccharomyces pombe*, 2540292; *Plasmodium falciparum*, 814496; *Leishmania major*, 5654193; *Trypanosoma brucei*, 3659748.

significant divergence for the tails of histone H3 and H4 in *Trypanosoma brucei* and *Leishmania major*. While few amino acids at the very N-terminal positions of H3 and H4 still show homology to metazoans, other parts of the tails diverge entirely (Figure 25). It is tempting to speculate that these differences evolved because *Trypanosoma brucei* and *Leishmania major* never developed a CTD structure. Given that the CTD and histone tails in eukaryotes underwent coevolution, *Trypanosoma brucei* and *Leishmania major* evolved different histone tails.

Intronic RNA splicing opened a playground for evolution with unexpected potential. Before introns, elements for regulation of gene activity could only be placed in front or behind the coding sequence of a gene, but never within a gene without disturbing the structure of the encoded protein. With the help of introns, information could now be placed in the gene body and exactly at positions where this information was required. In addition, genes showed almost no limitation for the number and size of new introns. While the total number of introns from human to lancelet remained almost constant, the individual size and sequence of each intron differ substantially. Thus, the evolution of the *rpb1* gene occurred also at the level of its introns. The importance and consequences of this evolution is currently not understood. However, it becomes more evident that the CTD plays a crucial role not only for the removal of intronic RNA from pre-mRNAs, but is involved also in intron-dependent processes beyond splicing. In the following

section, we will give an overview of intron-associated functions with a specific focus on the CTD.

## 6.2. Multifunctional Properties of Introns

Most mammalian genes initiate transcription from multiple promoters.<sup>202</sup> Alternative promoters increase the regulatory options of a gene, can be targeted by different signaling pathways, and allow tissue-specific gene regulation. Additional promoters initiate transcripts with divergent 5' sequences, which may influence the stability, cellular localization, and translation of the mRNA. Thus, the positioning of additional promoters in introns expands the regulatory options for gene transcription but also affects the post-transcriptional regulation of a gene. In addition, initiation of transcription in an antisense orientation from intronic promoters creates a further level of regulation<sup>203</sup> (Figure 26A) and can attenuate the activity of a gene.

Introns may also be integrated into the coding region of genes to provide elements for enhancing, silencing, or insulating gene activity (Figure 26B). Enhancers, silencers, and insulators are DNA elements that can interact with the promoter by DNA looping, but may also contribute to the establishment of a specific form of chromatin to foster or attenuate gene activity.<sup>204</sup> Enhancers and silencers often contain their own promoters, but the function of RNAPII activity in enhancers and silencers and their transcribed RNAs

## Histone 3

		1	10	20	30	40	CTD
Animal	Human ( <i>Homo sapiens</i> )	MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALR					+
	Zebrafish ( <i>Danio rerio</i> )	MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALR					+
	Lancelet ( <i>B. floridae</i> )	MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALR					+
Fungi	<i>S. pombe</i>	MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALR					+
	<i>Aspergillus nidulans</i>	MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALR					+
Alveolates	<i>Plasmodium falciparum</i>	MARTKQTARKSTAGKAPRKQLATKAARKSAPISAGIKKPHRYRPGTVALR					(+)
Excavates	<i>Trypanosoma brucei</i>	MSRTKETARTKKITSKSKKASKGSDAASGVTKAQR---RWRPGTVALR					-
	<i>Leishmania major</i>	MSRTKETARAKRTITSKSKKAPSAGSVKMS-----HRRWRPGTCAIR					-

## Histone 4

		1	10	20	30	40	CTD
Animal	Human ( <i>Homo sapiens</i> )	MSGRGKGKGLGKGGAKRHRKVLRDNIQGITKPAIRRLR					+
	Zebrafish ( <i>Danio rerio</i> )	MSGRGKGKGLGKGGAKRHRKVLRDNIQGITKPAIRRLR					+
	Lancelet ( <i>B. floridae</i> )	MSGRGKGKGLGKGGAKRHRKVLRDNIQGITKPAIRRLR					+
Fungi	<i>S. pombe</i>	MSGRGKGKGLGKGGAKRHRKVLRDNIQGITKPAIRRLR					+
	<i>Aspergillus nidulans</i>	MSGRGKGKGLGKGGAKRHRKVLRDNIQGITKPAIRRLR					+
Alveolates	<i>Plasmodium falciparum</i>	MSGRGKGKGLGKGGAKRHRKVLRDNIQGITKPAIRRLR					(+)
Excavates	<i>Trypanosoma brucei</i>	MA-KGKKSGEAKGSQ-KRQKKVLRENVRGITRGSIRRRLR					-
	<i>Leishmania major</i>	MA-KGKRSTDAGSQ-RRQKKVLRDNIRGITRGCVRMR					-

**Figure 25.** Conservation of N-terminal regions of histone H3 and H4 in evolution. Divergence from the metazoan sequence is highlighted in blue. Organisms with CTDs are indicated. *Plasmodium falciparum* has a rudimentary CTD structure. GenBank accession numbers for protein sequences: *Homo sapiens*, H3, NP\_724345; H4, NP\_003539; *Danio rerio*, H3, NP\_001093643; H4, CAM12247; *Branchiostoma floridae*, H3, EEN69959; H4, EEN69654; *Schizosaccharomyces pombe*, H3, EEN69959; H4, NP\_594682; *Aspergillus nidulans*, H3, CBF88887; H4, CBF88885; *Plasmodium falciparum*, H3, AAO23910; H4, AAP45785; *Trypanosoma brucei*, H3, XP\_001218946; H4, AAX80625; *Leishmania major*, H3, XP\_001682136; H4, XP\_003721607.

are not well understood. Insulators are suggested to mediate inter- and intrachromosomal interactions to regulate transcriptional activity of genes in a spatially defined manner and to contribute to the higher order structure of chromatin and nuclear architecture.

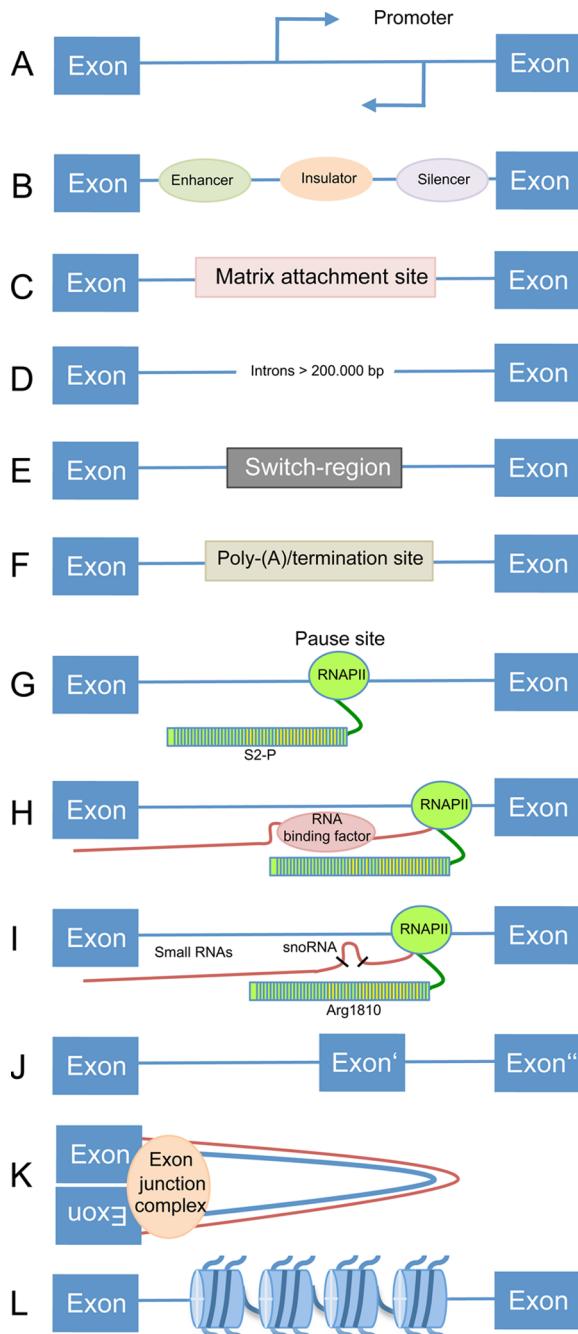
While enhancers, silencers, and insulators may function in a tissue-specific and regulatory manner, matrix attachment regions (MARs) are operationally defined genomic elements, which remain associated with the nuclear matrix after DNase I digestion of chromatin in isolated nuclei.<sup>205</sup> To what extent these regions share functional similarity to enhancers, silencers, and insulators has not been established; however, MARs occur with high incidence, also in introns, and play a crucial role in the spatial organization of chromatin (Figure 26C).

In this context, the varying size of introns may also play an important function. While the vast majority of introns are relatively short, within the range of one or a few kilobases, ~5% of mammalian introns are >200 000 kb (Figure 26D). In particular, the first intron of genes can be extremely large.<sup>206</sup> This extreme size cannot be explained by the positioning of small regulatory elements in genes. It is more likely that the size of long introns itself matters. Possible explanations could be that very long introns can regulate the timing of gene expression and adapt the duration of transcription to other cellular processes, or that long introns act as spacers to allow contact with other regulatory elements in trans, which are spatially separated. Extremely long introns bear the risk that the splice donor can use weak splice acceptor sites further downstream before RNAPII has transcribed the splice acceptor site of the next exon. This undesirable splicing within large introns might be prevented, if large introns become partly degraded before splicing to the downstream exons occurs. Such a mechanism appears possible, because the tethering of exons to the RNAPII transcription machinery has been reported.<sup>207</sup> This observation is consistent with the finding that the

sequences of exon/intron boundaries in nascent transcripts are more stable in pulse-chase labeling experiments using 4-thiouridine than internal sequences of introns.<sup>208</sup> If and how the CTD is involved in the process of exon tethering has not been established.

Introns can also fulfill very specific tasks in DNA recombination during B cell maturation. The gene fragments encoding the B cell receptor and secreted immunoglobulins are realigned during differentiation to plasma cells, and give rise to specific subtypes of immunoglobulins. These rearrangements require specific intronic switch-regions (Figure 26E), which realign immunoglobulin variable and constant regions in a cytokine-regulated manner.<sup>209</sup> Immunoglobulin genes are also a good example for the alternative usage of intronic polyadenylation sites. The decision to express the secreted form of an immunoglobulin and to suppress the cell surface-anchored version of an immunoglobulin is achieved by splicing of the last intron. If splicing is suppressed, the transcript is 3' processed and polyadenylated in the last intron and the secreted immunoglobulin is expressed (Figure 26F); if splicing is permitted, the intronic poly-A/termination site is removed and the last exon encoding the immunoglobulin surface anchor is attached to the mRNA.<sup>210</sup>

Introns can also contain elements for the regulation of RNA elongation. In fact, the control of RNA elongation is a regulatory mechanism directly associated with the phosphorylation of the CTD. The linkage between CTD phosphorylation and RNA elongation was first observed for retroviral HIV gene expression after integration of pro-virus DNA into the cellular genome. RNA elongation requires phosphorylation of the CTD at Ser2 by P-TEFb, a complex of cyclin T and Cdk9, which is recruited to the transcription machinery by a specific stem-loop structure (TAR) in the nascent transcript, and the viral Tat protein.<sup>40,136m</sup> Pausing of RNAPII frequently occurs in introns of cellular genes either to control the level of



**Figure 26.** Gene-specific functions provided by introns. (A) New sense or antisense promoters, (B) gene-specific enhancer, silencer, insulator, (C) matrix attachment site, (D) spacer sequence to increase intron length, (E) immunoglobulin gene switch region, (F) alternative polyadenylation, transcription termination site, (G) RNAPII pause site regulating elongation, (H) binding sites in the nascent intronic transcript for cellular factors, (I) source for small RNAs, for example, snoRNAs, (J) alternative splicing of exons, (K) exon junction complex-mediated quality control mechanism for mRNAs, and (L) storage of epigenetic information in intronic chromatin.

gene expression,<sup>211</sup> or for selection of splice acceptor sites and recruitment of splicing factors.<sup>7a,193b,212</sup> In both cases, the phosphorylation of the CTD at Ser2 residues is critical for the continuation of transcription (Figure 26G). Splice site selection by RNAPII is also affected in UV-irradiated cells, which is

accompanied by pausing and hyperphosphorylation of the CTD.<sup>7e,213</sup>

In addition to the splice donor, splice acceptor, and the pyrimidine tract, intronic RNA can contain further sequence elements for the binding of regulatory factors. Factors binding to the nascent transcript (Figure 26H) can act as bridging elements and can facilitate interaction with the phosphorylated CTD and/or with specific modifications of histone tails. For example, the regulation of glia-specific splicing of Neurexin IV requires the activity of HOW and Cdk12. Binding of HOW to an ACUAA motif in intronic RNA together with phosphorylation of serine 2 residues in the CTD by Cdk12 allows the formation of a HOW-specific spliceosome complex that triggers glia-specific splicing of Neurexin IV exon 3.<sup>214</sup> Another regulatory splicing factor is MRG15. This factor is recruited to chromatin by H3K36me3 marks and interacts with the nascent transcript via the polypyrimidine tract binding protein (PTB).<sup>7d,195b</sup> Thus, several elements in nascent transcripts together with the CTD and chromatin associated factors contribute to the recognition and formation of a functional splice acceptor site. An interesting question is: “How is assembly of the functional spliceosome acceptor complex orchestrated by CTD modifications?”

Introns are also a seemingly inexhaustible source for the generation of small regulatory RNAs of different categories (Figure 26I). These RNAs are transcribed either from their own promoters or from promoters of the corresponding gene. This poses the question of whether specific small RNAs have to be transcribed from particular introns to fulfill their functions. For example, many small nucleolar (sno)RNAs are encoded by introns belonging to genes involved in ribosome biogenesis. These snoRNAs are required for processing rRNA and to guide the methylation and pseudouridylation of rRNA. The evolutionarily stable association of intronic snoRNAs and other microRNAs with their host genes suggests a functional overlap with the host gene.<sup>215</sup> Interestingly, a direct link between snoRNA synthesis and CTD modification has been reported. Methylation of arginine 1810 in heptad 31 of the mammalian CTD is involved in the control of snoRNA synthesis. Mutation of R1810 results in the misexpression of a variety of snoRNAs and other small RNAs.<sup>67</sup> This is the first example showing that the modification of a single amino acid in the CTD can be of regulatory relevance. A function of Ser2 phosphorylation in CTD was recently also suggested for processing of snoRNA U8 in mammalian cells.<sup>216</sup>

The genome projects for many organisms are advanced. Analysis of the corresponding transcriptomes revealed alternative splicing as a widespread phenomenon. Alternative splicing can increase the number of specific protein variants to unbelievable dimensions (Figure 26J). An extreme example is the neural receptor gene, *DSCAM*, in the fruit fly *Drosophila melanogaster*. The mRNA for this gene is composed of 24 exons, but 4 of these exons have the choice of splicing to 12, 48, 33, or 2 alternative exons, respectively. This huge number of options enables the potential production of tens of thousands different *DSCAM* isoforms.<sup>217</sup> Interestingly, neural cells express variant *DSCAM* forms in a mutually exclusive manner. How this exclusion is regulated and whether specific functions or modifications of the CTD are required for this regulation is currently unknown.

Splicing of introns has been adapted to provide a sophisticated quality control mechanism for mRNAs. After splicing and 3' processing, the mature mRNA is transported to

the cytoplasm in a CTD-dependent manner. Truncated forms of the CTD can block the release of mRNA from its gene locus and proper transport,<sup>218</sup> suggesting that CTD-interacting mRNA export factors, such as Yra1,<sup>91</sup> are required for the release of mRNA to the cytoplasm. Whether the function of Yra1 in mRNA export is CTD-dependent has not yet been demonstrated. Under normal conditions, the exported mRNA is decorated with the exon junction complex (EJC), which marks the exon–exon boundary of each splicing event<sup>219</sup> (Figure 26K). The EJC confers instability to the mRNA in the cytoplasm and has to be removed by successful translation of the message. However, in the case of frame shift mutations in the mRNA, the ribosome encounters early stop codons and terminates translation without removing EJCs in the distal part of the mRNA. As a consequence, incompletely translated mRNAs are targeted to the nonsense-mediated decay pathway.<sup>220</sup> Thus, with the evolution of splicing, a very elegant mechanism for the control of mRNA quality could coevolve.

The last example in this survey of intron functions is perhaps the most compelling but also the most unexplored mechanism of intron function. DNA in eukaryotes is compacted in nucleosomes, the fundamental building block of chromatin. Allis and Strahl proposed that modifications of histone tails of one or more nucleosomes act sequentially or in combination to form a “histone code” that is read by other proteins to bring about distinct downstream events.<sup>181</sup> This language of covalent histone modifications is also referred to as an epigenetic code. The intronic chromatin could therefore be used for the gene-specific linkage and storage of epigenetic information in a tissue-specific manner. Chromatin modifying enzymes and chromatin binding factors recruited by the CTD could be involved in writing and reading the “histone code”, while chromatin associated factors could have the function of writing and reading a “CTD code”. The number of nucleosomes in a gene increases with the number and size of its introns. This enables genes to save more epigenetic information.

The survey of intron-specific functions in this section is certainly far from being complete but may provide some insights into how the multifunctional properties of introns contributed to the development of higher eukaryotes. The survey also explains that the function of the CTD was not only a prerequisite for the occurrence and the regulated removal of introns from pre-mRNA, but that the CTD is also involved in many other transcription-coupled processes. There are urgent questions of how CTD modifications can regulate all of its diverse functions, and the most challenging of these is as follows: Does a CTD code exist, and, if yes, how is this code written and read?

## 7. CONCLUSIONS AND FUTURE PERSPECTIVES

### 7.1. How Can the CTD Code Be Deciphered?

Our knowledge of a potential CTD code is based primarily on the application of monoclonal antibodies in genome-wide ChIP experiments. However, constraints for binding of mAb's to epitopes are often not well characterized. Antibodies are selected for their ability to recognize a specific CTD modification in peptides, but recognition of the same epitope in the context of the CTD can be impaired or inhibited if other modifications occur adjacent to the epitope *in vivo*. The restriction of epitope recognition by mAb's therefore has a significant impact not only on the interpretation of data generated by ChIP, but also on data obtained in immunopre-

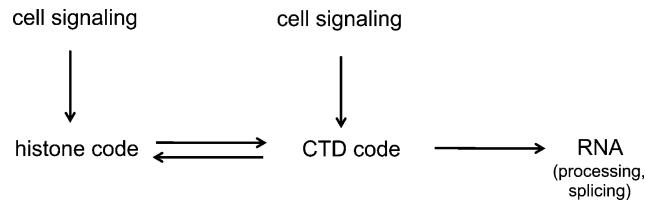
cipitation and immunoblot experiments. For mAb recognition, a single accessible epitope in the CTD is sufficient. Hence, the signal strength depends solely on the number of accessible CTD-marks and not on the number of marks physically present in the CTD. Further, the absence of a CTD-mark can indicate either its physical absence in the CTD or alternatively its masking by other modifications. These are important limitations of mAb's for the analysis of CTD modifications.

mAb's are also not able to discriminate the number of epitopes, if they occur several times within a CTD, nor can they discriminate whether the phosphorylation of specific amino acid residues occurs preferentially in the proximal or distal part of the CTD, or whether the occurrence of a specific phosphorylation mark depends on other, adjacent modifications. All of these limitations and restrictions indicate that the application of mAb's is not sufficient for the deciphering the CTD code. Other techniques like NMR and crystal structure analysis may not be suitable for the analysis of complex CTD modifications. The only current method to decipher a complex array of CTD modification is mass spectrometry. This method can analyze highly complex mixtures of molecules with extreme heterogeneity.

What problems are we facing? A mammalian cell contains approximately 100 000 RNAPII molecules, and the CTD of each of these molecules has 236 potential phosphorylation sites. Alone, the huge number of sites makes it impossible to analyze the modification pattern of the entire CTD as a whole. The current strategy must be to cut the CTD (and thereby the problem) into smaller pieces. This strategy makes sense for several reasons. First, the mammalian CTD in total comprises >350 amino acids. Such a big molecule is not suitable for mass spectrometry analysis and even less for the study of its modification. Second, as described in section 2.2, the CTDs in *S. cerevisiae* and *S. pombe* represent arrays of functional units, each unit with a length of 11 amino acids plus a flexible spacer. Genetically engineered CTDs of >8 functional units are viable in *S. cerevisiae*. If the flexible spacer of each functional unit becomes tagged with a different set of amino acids including a basic lysine or arginine amino acid, each functional unit suddenly becomes an individual unit that can be analyzed by mass spectrometry. Each minimal functional unit (Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7-Tyr1-Ser2-Pro3-Thr4) contains eight phosphorylation sites and the option for several hundreds of different phosphorylation patterns. Currently, it is completely unclear how many of the potential sites can be phosphorylated in a single unit at the same time, nor is anything known about the possible combinations of phosphorylation patterns. It is unclear if a small number of combinations prevails over other possible combinations, and whether distinct combinations occur preferentially in the distal or proximal part of the CTD. The described analysis would also help to answer the question of how specific modification patterns are established by kinases and how these patterns are later changed or removed by phosphatases. It is essential to know the common phosphorylation signatures in the CTD to be able to analyze the mode of CTD operation and its interacting factors.

We have yet not reached the point that enables us to read the language of CTD modifications, and we are even farther away from being able to understand its language. The terms “language” and “code” have often been used to describe the complexity of CTD modifications. From the title of this Review, one can see that we finally decided in favor for the term “code”. In addition to the assembly line in a car factory, we

often compare the function of the CTD with the function of an audio head of a tape recorder. The audio head (CTD) is able to decode information from the tape (chromatin), but is also able to write new information onto or erase old information from the tape. Because of tight crosstalk and mutual dependence of the CTD and chromatin (Figure 27), both structures probably



**Figure 27.** Crosstalk of CTD and chromatin.

coevolved in different taxa. Therefore, we expect that evolutionary close taxa speak the same language, while more distant taxa may have developed significant changes in grammar, syntax, and punctuation<sup>20</sup> for communication of the CTD with chromatin.

## AUTHOR INFORMATION

### Corresponding Author

\*Tel.: +49 (0) 89 7099 512 (D.E.); +49 (0) 228 9656 233 (M.G.). Fax: +49 (0) 89 7099 500 (D.E.); +49 (0) 228 9656 9233 (M.G.). E-mail: eick@helmholtz-muenchen.de (D.E.); Matthias.Geyer@caesar.de (M.G.).

### Notes

The authors declare no competing financial interest.

### Biographies



Dirk Eick graduated in genetics and received his Ph.D. from the University Cologne in 1983, working with Prof. Walter Doerfler. From 1984 until 1988 he was a postdoctoral researcher in the Institute of Virology, University Freiburg, with Prof. Georg Bornkamm. In Freiburg, he discovered the proto-oncogene, *c-myc*, as the first metazoan gene that is regulated at the level of RNA elongation. In 1989 he moved to the Helmholtz Center Munich as group leader. Since 1991 he has been a member of the Faculty of Biology of the Ludwig-Maximilians University (LMU), Munich, and since 2006 a member of the LMU excellence cluster, CIPSM (Center of Integrated Protein Science). Since 2008 he has been Head of the Department for Molecular Epigenetics at the Helmholtz Center Munich. Currently, the work of his laboratory focuses on the connection between ribosome biogenesis and cell cycle control and on the modification

and function of the carboxy-terminal domain (CTD) of RNA polymerase II in the control of gene expression.



Matthias Geyer graduated in biophysics and received his Ph.D. degree from the Ruprecht Karls University in Heidelberg in 1995, where he discovered dynamics in the active center of the proto-oncogene, Ras. After a postdoctoral stay with Prof. B. Matija Peterlin in San Francisco, he became group leader in Roger S. Goody's department at the Max Planck Institute of Molecular Physiology in Dortmund, Germany. In this time, he determined the structure of the retroviral Cyclin T1-Tat-TAR complex and became interested in the regulation of transcriptional elongation. Since 2012 he has led the Physical Biochemistry research group at the Center of Advanced European Studies and Research (caesar) in Bonn. He is interested in understanding the structural basis of protein function and mechanisms of action. His current interests include RNAPII CTD kinases and their mechanisms of regulation.

### ACKNOWLEDGMENTS

The fruitful collaboration and the open exchange of ideas with colleagues working in this area are greatly appreciated. The work of D.E. is supported by the Deutsche Forschungsgemeinschaft (SFB/TR5, SFB684, and SFB1064) and José Carreras Leukämie-Stiftung e.V. M.G. is supported by the Deutsche Forschungsgemeinschaft (GE-976/5,8,9).

### ABBREVIATIONS

Ala	alanine (A)
Arg	arginine (R)
CTD	carboxy-terminal domain of RNA polymerase II
H2A	histone 2A
H2B	histone 2B
H3	histone H3
H4	histone H4
Lys	lysine (K)
Pro	proline (P)
Ser	serine (S)
Thr	threonine (T)
Tyr	tyrosine (Y)

### REFERENCES

- (1) Theobald, D. L. *Nature* **2010**, *465*, 219.
- (2) (a) Heintzman, N. D.; Stuart, R. K.; Hon, G.; Fu, Y.; Ching, C. W.; Hawkins, R. D.; Barrera, L. O.; Van Calcar, S.; Qu, C.; Ching, K. A.; Wang, W.; Weng, Z.; Green, R. D.; Crawford, G. E.; Ren, B. *Nat. Genet.* **2007**, *39*, 311. (b) Koch, F.; Fenouil, R.; Gut, M.; Cauchy, P.; Albert, T. K.; Zacarias-Cabeza, J.; Spicuglia, S.; de la Chapelle, A. L.; Heidemann, M.; Hintermair, C.; Eick, D.; Gut, I.; Ferrier, P.; Andrau, J. C. *Nat. Struct. Mol. Biol.* **2011**, *18*, 956. (c) Koch, F.; Jourquin, F.;

- Ferrier, P.; Andrau, J. C. *Trends Biochem. Sci.* **2008**, *33*, 265.  
 (d) Steinmetz, E. J.; Warren, C. L.; Kuehner, J. N.; Panbehi, B.; Ansari, A. Z.; Brow, D. A. *Mol. Cell* **2006**, *24*, 735.  
 (3) Berk, A. J.; Sharp, P. A. *Cell* **1977**, *12*, 721.  
 (4) (a) Kim, H.; Erickson, B.; Luo, W.; Seward, D.; Gruber, J. H.; Pollock, D. D.; Megee, P. C.; Bentley, D. L. *Nat. Struct. Mol. Biol.* **2010**, *17*, 1279. (b) Heidemann, M.; Hintermair, C.; Voss, K.; Eick, D. *Biochim. Biophys. Acta* **2012**. (c) Buratowski, S. *Nat. Struct. Biol.* **2003**, *10*, 679. (d) Buratowski, S. *Mol. Cell* **2009**, *36*, 541. (e) Drogat, J.; Hermand, D. *Mol. Microbiol.* **2012**, *84*, 995. (f) Egloff, S.; Dienstbier, M.; Murphy, S. *Trends Genet.* **2012**, *28*, 333. (g) Egloff, S.; Murphy, S. *Trends Genet.* **2008**, *24*, 280. (h) Hsin, J. P.; Manley, J. L. *Genes Dev.* **2012**, *26*, 2119. (i) Phatnani, H. P.; Greenleaf, A. L. *Genes Dev.* **2006**, *20*, 2922. (j) Zhang, D. W.; Rodriguez-Molina, J. B.; Tietjen, J. R.; Nemec, C. M.; Ansari, A. Z. *Genet. Res. Int.* **2012**, *2012*, 347214.  
 (5) (a) Jasnovidova, O.; Stefl, R. *Wiley Interdiscip. Rev.: RNA* **2013**, *4*, 1. (b) Meinhart, A.; Kamenski, T.; Hoeppner, S.; Baumli, S.; Cramer, P. *Genes Dev.* **2005**, *19*, 1401. (c) Palancade, B.; Bensaude, O. *Eur. J. Biochem.* **2003**, *270*, 3859.  
 (6) (a) Chapman, R. D.; Heidemann, M.; Hintermair, C.; Eick, D. *Trends Genet.* **2008**, *24*, 289. (b) Stiller, J. W.; Hall, B. D. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6091.  
 (7) (a) Carrillo Oesterreich, F.; Bieberstein, N.; Neugebauer, K. M. *Trends Cell Biol.* **2011**, *21*, 328. (b) Corden, J. L.; Paturajan, M. *Trends Biochem. Sci.* **1997**, *22*, 413. (c) Kornblith, A. R.; de la Mata, M.; Fededa, J. P.; Munoz, M. J.; Nogues, G. *RNA* **2004**, *10*, 1489. (d) Luco, R. F.; Allo, M.; Schor, I. E.; Kornblith, A. R.; Misteli, T. *Cell* **2011**, *144*, 16. (e) Munoz, M. J.; de la Mata, M.; Kornblith, A. R. *Trends Biochem. Sci.* **2010**, *35*, 497. (f) Proudfoot, N. J.; Furger, A.; Dye, M. J. *Cell* **2002**, *108*, 501. (g) Lenasi, T.; Barboric, M. *Wiley Interdiscip. Rev.: RNA* **2012**.  
 (8) Allison, L. A.; Moyle, M.; Shales, M.; Ingles, C. J. *Cell* **1985**, *42*, 599.  
 (9) Corden, J. L.; Cadena, D. L.; Ahearn, J. M., Jr.; Dahmus, M. E. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 7934.  
 (10) Thompson, C. M.; Koleske, A. J.; Chao, D. M.; Young, R. A. *Cell* **1993**, *73*, 1361.  
 (11) Bartolomei, M. S.; Halden, N. F.; Cullen, C. R.; Corden, J. L. *Mol. Cell. Biol.* **1988**, *8*, 330.  
 (12) Chapman, J. A.; Kirkness, E. F.; Simakov, O.; Hampson, S. E.; Mitros, T.; Weinmaier, T.; Rattei, T.; Balasubramanian, P. G.; Borman, J.; Busam, D.; Disbennett, K.; Pfannkoch, C.; Sumin, N.; Sutton, G. G.; Viswanathan, L. D.; Walenz, B.; Goodstein, D. M.; Hellsten, U.; Kawashima, T.; Prochnik, S. E.; Putnam, N. H.; Shu, S.; Blumberg, B.; Dana, C. E.; Gee, L.; Kibler, D. F.; Law, L.; Lindgens, D.; Martinez, D. E.; Peng, J.; Wigge, P. A.; Bertulat, B.; Guder, C.; Nakamura, Y.; Ozbek, S.; Watanabe, H.; Khaliturin, K.; Hemmrich, G.; Franke, A.; Augustin, R.; Fraune, S.; Hayakawa, E.; Hayakawa, S.; Hirose, M.; Hwang, J. S.; Ikeo, K.; Nishimiya-Fujisawa, C.; Ogura, A.; Takahashi, T.; Steinmetz, P. R.; Zhang, X.; Aufschnaiter, R.; Eder, M. K.; Gorny, A. K.; Salvenmoser, W.; Heimberg, A. M.; Wheeler, B. M.; Peterson, K. J.; Bottger, A.; Tischler, P.; Wolf, A.; Gojobori, T.; Remington, K. A.; Strausberg, R. L.; Venter, J. C.; Technau, U.; Hobmayer, B.; Bosch, T. C.; Holstein, T. W.; Fujisawa, T.; Bode, H. R.; David, C. N.; Rokhsar, D. S.; Steele, R. E. *Nature* **2010**, *464*, 592.  
 (13) (a) Lolli, G. *Nucleic Acids Res.* **2009**, *37*, 1260. (b) Suzuki, M. *Nature* **1990**, *344*, 562.  
 (14) Liu, P.; Kenney, J. M.; Stiller, J. W.; Greenleaf, A. L. *Mol. Biol. Evol.* **2010**, *27*, 2628.  
 (15) West, M. L.; Corden, J. L. *Genetics* **1995**, *140*, 1223.  
 (16) Zhang, D. W.; Mosley, A. L.; Ramisetty, S. R.; Rodriguez-Molina, J. B.; Washburn, M. P.; Ansari, A. Z. *J. Biol. Chem.* **2012**, *287*, 8541.  
 (17) Schwer, B.; Shuman, S. *Mol. Cell* **2011**, *43*, 311.  
 (18) Stiller, J. W.; McConaughy, B. L.; Hall, B. D. *Yeast* **2000**, *16*, 57.  
 (19) (a) Liu, P.; Greenleaf, A. L.; Stiller, J. W. *Mol. Biol. Evol.* **2008**, *25*, 719. (b) Stiller, J. W.; Cook, M. S. *Eukaryotic Cell* **2004**, *3*, 735.  
 (20) Schwer, B.; Sanchez, A. M.; Shuman, S. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 18024.  
 (21) Hintermair, C.; Heidemann, M.; Koch, F.; Descotes, N.; Gut, M.; Gut, I.; Fenouil, R.; Ferrier, P.; Flatley, A.; Kremmer, E.; Chapman, R. D.; Andrau, J. C.; Eick, D. *EMBO J.* **2012**, *31*, 2784.  
 (22) Hsin, J. P.; Sheth, A.; Manley, J. L. *Science* **2011**, *334*, 683.  
 (23) (a) Chapman, R. D.; Heidemann, M.; Albert, T. K.; Mailhammer, R.; Flatley, A.; Meisterernst, M.; Kremmer, E.; Eick, D. *Science* **2007**, *318*, 1780. (b) Heidemann, M.; Eick, D. *RNA Biol.* **2012**, *9*, 1144. (c) Mayer, A.; Heidemann, M.; Lidschreiber, M.; Schreieck, A.; Sun, M.; Hintermair, C.; Kremmer, E.; Eick, D.; Cramer, P. *Science* **2012**, *336*, 1723. (d) Zhang, J.; Corden, J. L. *J. Biol. Chem.* **1991**, *266*, 2297. (e) Zhang, J.; Corden, J. L. *J. Biol. Chem.* **1991**, *266*, 2290.  
 (24) Kelly, W. G.; Dahmus, M. E.; Hart, G. W. *J. Biol. Chem.* **1993**, *268*, 10416.  
 (25) Ranuncolo, S. M.; Ghosh, S.; Hanover, J. A.; Hart, G. W.; Lewis, B. A. *J. Biol. Chem.* **2012**, *287*, 23549.  
 (26) (a) Armache, K. J.; Mitterweger, S.; Meinhart, A.; Cramer, P. *J. Biol. Chem.* **2005**, *280*, 7131. (b) Cramer, P.; Bushnell, D. A.; Kornberg, R. D. *Science* **2001**, *292*, 1863.  
 (27) Suzuki, M. *J. Mol. Biol.* **1989**, *207*, 61.  
 (28) Meredith, G. D.; Chang, W. H.; Li, Y.; Bushnell, D. A.; Darst, S. A.; Kornberg, R. D. *J. Mol. Biol.* **1996**, *258*, 413.  
 (29) Laybourn, P. J.; Dahmus, M. E. *J. Biol. Chem.* **1989**, *264*, 6693.  
 (30) (a) Feaver, W. J.; Gileadi, O.; Li, Y.; Kornberg, R. D. *Cell* **1991**, *67*, 1223. (b) Lu, H.; Zawel, L.; Fisher, L.; Egly, J. M.; Reinberg, D. *Nature* **1992**, *358*, 641.  
 (31) (a) Glover-Cutter, K.; Larochelle, S.; Erickson, B.; Zhang, C.; Shokat, K.; Fisher, R. P.; Bentley, D. L. *Mol. Cell. Biol.* **2009**, *29*, 5455. (b) Hong, S. W.; Hong, S. M.; Yoo, J. W.; Lee, Y. C.; Kim, S.; Lis, J. T.; Lee, D. K. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 14276. (c) Kanin, E. I.; Kipp, R. T.; Kung, C.; Slattery, M.; Viale, A.; Hahn, S.; Shokat, K. M.; Ansari, A. Z. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 5812. (d) Kim, M.; Suh, H.; Cho, E. J.; Buratowski, S. J. *Biol. Chem.* **2009**, *284*, 26421. (e) Viladevall, L.; Amour, C. V.; Rosebrock, A.; Schneider, S.; Zhang, C.; Allen, J. J.; Shokat, K. M.; Schwer, B.; Leatherwood, J. K.; Fisher, R. P. *Mol. Cell* **2009**, *33*, 738.  
 (32) Akhtar, M. S.; Heidemann, M.; Tietjen, J. R.; Zhang, D. W.; Chapman, R. D.; Eick, D.; Ansari, A. Z. *Mol. Cell* **2009**, *34*, 387.  
 (33) (a) Liao, S. M.; Zhang, J.; Jeffery, D. A.; Koleske, A. J.; Thompson, C. M.; Chao, D. M.; Viljoen, M.; van Vuuren, H. J.; Young, R. A. *Nature* **1995**, *374*, 193. (b) Sun, X.; Zhang, Y.; Cho, H.; Rickert, P.; Lees, E.; Lane, W.; Reinberg, D. *Mol. Cell* **1998**, *2*, 213.  
 (34) Davis, M. A.; Larimore, E. A.; Fissel, B. M.; Swanger, J.; Taatjes, D. J.; Clurman, B. E. *Genes Dev.* **2013**, *27*, 151.  
 (35) Schneider, E. V.; Bottcher, J.; Blaesse, M.; Neumann, L.; Huber, R.; Maskos, K. *J. Mol. Biol.* **2011**, *412*, 251.  
 (36) Galbraith, M. D.; Donner, A. J.; Espinosa, J. M. *Transcription* **2010**, *1*, 4.  
 (37) (a) Mousley, C. J.; Yuan, P.; Gaur, N. A.; Trettin, K. D.; Nile, A. H.; Deminoff, S. J.; Dewar, B. J.; Wolpert, M.; Macdonald, J. M.; Herman, P. K.; Hinnebusch, A. G.; Bankaitis, V. A. *Cell* **2012**, *148*, 702. (b) Rosonina, E.; Duncan, S. M.; Manley, J. L. *Genes Dev.* **2012**, *26*, 350.  
 (38) (a) Marshall, N. F.; Peng, J.; Xie, Z.; Price, D. H. *J. Biol. Chem.* **1996**, *271*, 27176. (b) Marshall, N. F.; Price, D. H. *J. Biol. Chem.* **1995**, *270*, 12335.  
 (39) Chao, S. H.; Price, D. H. *J. Biol. Chem.* **2001**, *276*, 31793.  
 (40) Wei, P.; Garber, M. E.; Fang, S. M.; Fischer, W. H.; Jones, K. A. *Cell* **1998**, *92*, 451.  
 (41) (a) Kao, S. Y.; Calman, A. F.; Luciw, P. A.; Peterlin, B. M. *Nature* **1987**, *330*, 489. (b) Ruben, S.; Poteat, H.; Tan, T. H.; Kawakami, K.; Roeder, R.; Haseltine, W.; Rosen, C. A. *Science* **1988**, *241*, 89.  
 (42) Rougvie, A. E.; Lis, J. T. *Cell* **1988**, *54*, 795.  
 (43) Fuda, N. J.; Ardehali, M. B.; Lis, J. T. *Nature* **2009**, *461*, 186.  
 (44) Czudnochowski, N.; Bosken, C. A.; Geyer, M. *Nat. Commun.* **2012**, *3*, 842.  
 (45) (a) Jones, J. C.; Phatnani, H. P.; Haystead, T. A.; MacDonald, J. A.; Alam, S. M.; Greenleaf, A. L. *J. Biol. Chem.* **2004**, *279*, 24957. (b) Ramanathan, Y.; Rajpara, S. M.; Reza, S. M.; Lees, E.; Shuman, S.;

- Mathews, M. B.; Pe'ery, T. *J. Biol. Chem.* **2001**, 276, 10913.  
 (c) Ramanathan, Y.; Reza, S. M.; Young, T. M.; Mathews, M. B.; Pe'ery, T. *J. Virol.* **1999**, 73, 5448.
- (46) St. Amour, C. V.; Sanso, M.; Bosken, C. A.; Lee, K. M.; Laroche, S.; Zhang, C.; Shokat, K. M.; Geyer, M.; Fisher, R. P. *Mol. Cell. Biol.* **2012**, 32, 2372.
- (47) Ghamari, A.; van de Corput, M. P.; Thongjuea, S.; van Cappellen, W. A.; van Ijcken, W.; van Haren, J.; Soler, E.; Eick, D.; Lenhard, B.; Grosveld, F. G. *Genes Dev.* **2013**, 27, 767.
- (48) (a) Bartkowiak, B.; Liu, P.; Phatnani, H. P.; Fuda, N. J.; Cooper, J. J.; Price, D. H.; Adelman, K.; Lis, J. T.; Greenleaf, A. L. *Genes Dev.* **2010**, 24, 2303. (b) Blazek, D.; Kohoutek, J.; Bartholomaeusen, K.; Johansen, E.; Hulinkova, P.; Luo, Z.; Cimermancic, P.; Ule, J.; Peterlin, B. M. *Genes Dev.* **2011**, 25, 2158.
- (49) (a) Bellier, S.; Dubois, M. F.; Nishida, E.; Almouzni, G.; Bensaude, O. *Mol. Cell. Biol.* **1997**, 17, 1434. (b) Bonnet, F.; Vigneron, M.; Bensaude, O.; Dubois, M. F. *Nucleic Acids Res.* **1999**, 27, 4399.  
 (c) Devaiah, B. N.; Lewis, B. A.; Cherman, N.; Hewitt, M. C.; Albrecht, B. K.; Robey, P. G.; Ozato, K.; Sims, R. J., III; Singer, D. S. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, 109, 6927.
- (50) Baskaran, R.; Dahmus, M. E.; Wang, J. Y. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, 90, 11167.
- (51) Baskaran, R.; Chiang, G. G.; Wang, J. Y. *Mol. Cell. Biol.* **1996**, 16, 3361.
- (52) (a) Ganem, C.; Devaux, F.; Torchet, C.; Jacq, C.; Quevillon-Cheruel, S.; Labesse, G.; Facca, C.; Faye, G. *EMBO J.* **2003**, 22, 1588. (b) Krishnamurthy, S.; He, X.; Reyes-Reyes, M.; Moore, C.; Hampsey, M. *Mol. Cell* **2004**, 14, 387.
- (53) Bataille, A. R.; Jeronimo, C.; Jacques, P. E.; Laramee, L.; Fortin, M. E.; Forest, A.; Bergeron, M.; Hanes, S. D.; Robert, F. *Mol. Cell* **2012**, 45, 158.
- (54) (a) Ansari, A.; Hampsey, M. *Genes Dev.* **2005**, 19, 2969. (b) Tan-Wong, S. M.; Zaugg, J. B.; Camblong, J.; Xu, Z.; Zhang, D. W.; Mischo, H. E.; Ansari, A. Z.; Luscombe, N. M.; Steinmetz, L. M.; Proudfoot, N. J. *Science* **2012**, 338, 671.
- (55) Mosley, A. L.; Pattenden, S. G.; Carey, M.; Venkatesh, S.; Gilmore, J. M.; Florens, L.; Workman, J. L.; Washburn, M. P. *Mol. Cell* **2009**, 34, 168.
- (56) Xiang, K.; Manley, J. L.; Tong, L. *Nat. Commun.* **2012**, 3, 946.
- (57) Egloff, S.; Zaborowska, J.; Laitem, C.; Kiss, T.; Murphy, S. *Mol. Cell* **2012**, 45, 111.
- (58) Cho, H.; Kim, T. K.; Mancebo, H.; Lane, W. S.; Flores, O.; Reinberg, D. *Genes Dev.* **1999**, 13, 1540.
- (59) Cho, E. J.; Kobor, M. S.; Kim, M.; Greenblatt, J.; Buratowski, S. *Genes Dev.* **2001**, 15, 3319.
- (60) (a) Lin, P. S.; Dubois, M. F.; Dahmus, M. E. *J. Biol. Chem.* **2002**, 277, 45949. (b) Archambault, J.; Chambers, R. S.; Kobor, M. S.; Ho, Y.; Cartier, M.; Bolotin, D.; Andrews, B.; Kane, C. M.; Greenblatt, J. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, 94, 14300. (c) Kobor, M. S.; Archambault, J.; Lester, W.; Holstege, F. C.; Gileadi, O.; Jansma, D. B.; Jennings, E. G.; Kouyoumdjian, F.; Davidson, A. R.; Young, R. A.; Greenblatt, J. *Mol. Cell* **1999**, 4, 55.
- (61) Fuda, N. J.; Buckley, M. S.; Wei, W.; Core, L. J.; Waters, C. T.; Reinberg, D.; Lis, J. T. *Mol. Cell. Biol.* **2012**, 32, 3428.
- (62) Yeo, M.; Lin, P. S.; Dahmus, M. E.; Gill, G. N. *J. Biol. Chem.* **2003**, 278, 26078.
- (63) (a) Yeo, M.; Lee, S. K.; Lee, B.; Ruiz, E. C.; Pfaff, S. L.; Gill, G. N. *Science* **2005**, 307, 596. (b) Yeo, M.; Lin, P. S. *Methods Mol. Biol.* **2007**, 365, 335.
- (64) Feng, Y.; Kang, J. S.; Kim, S.; Yun, D. J.; Lee, S. Y.; Bahk, J. D.; Koiwa, H. *Biochem. Biophys. Res. Commun.* **2010**, 397, 355.
- (65) (a) Koiwa, H.; Hausmann, S.; Bang, W. Y.; Ueda, A.; Kondo, N.; Hiraguri, A.; Fukuhara, T.; Bahk, J. D.; Yun, D. J.; Bressan, R. A.; Hasegawa, P. M.; Shuman, S. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, 101, 14539. (b) Ueda, A.; Li, P.; Feng, Y.; Vikram, M.; Kim, S.; Kang, C. H.; Kang, J. S.; Bahk, J. D.; Lee, S. Y.; Fukuhara, T.; Staswick, P. E.; Pepper, A. E.; Koiwa, H. *Plant Mol. Biol.* **2008**, 67, 683.
- (66) Clemente-Blanco, A.; Sen, N.; Mayan-Santos, M.; Sacristan, M. P.; Graham, B.; Jarmuz, A.; Giess, A.; Webb, E.; Game, L.; Eick, D.; Bueno, A.; Merkenschlager, M.; Aragon, L. *Nat. Cell Biol.* **2011**, 13, 1450.
- (67) Sims, R. J., 3rd; Rojas, L. A.; Beck, D.; Bonasio, R.; Schuller, R.; Drury, W. J., 3rd; Eick, D.; Reinberg, D. *Science* **2011**, 332, 99.
- (68) Cheng, D.; Cote, J.; Shaaban, S.; Bedford, M. T. *Mol. Cell* **2007**, 25, 71.
- (69) Litingtung, Y.; Lawler, A. M.; Sebald, S. M.; Lee, E.; Gearhart, J. D.; Westphal, H.; Corden, J. L. *Mol. Gen. Genet.* **1999**, 261, 100.
- (70) Comer, F. I.; Hart, G. W. *Biochemistry* **2001**, 40, 7845.
- (71) Harreman, M.; Taschner, M.; Sigurdsson, S.; Anindya, R.; Reid, J.; Somesh, B.; Kong, S. E.; Banks, C. A.; Conaway, R. C.; Conaway, J. W.; Svejstrup, J. Q. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, 106, 20705.
- (72) Kvint, K.; Uhler, J. P.; Taschner, M. J.; Sigurdsson, S.; Erdjument-Bromage, H.; Tempst, P.; Svejstrup, J. Q. *Mol. Cell* **2008**, 30, 498.
- (73) Zhang, M.; Wang, X. J.; Chen, X.; Bowman, M. E.; Luo, Y.; Noel, J. P.; Ellington, A. D.; Etkorn, F. A.; Zhang, Y. *ACS Chem. Biol.* **2012**, 7, 1462.
- (74) (a) Meinhart, A.; Cramer, P. *Nature* **2004**, 430, 223. (b) Noble, C. G.; Hollingworth, D.; Martin, S. R.; Ennis-Adeniran, V.; Smerdon, S. J.; Kelly, G.; Taylor, I. A.; Ramos, A. *Nat. Struct. Mol. Biol.* **2005**, 12, 144.
- (75) Werner-Allen, J. W.; Lee, C. J.; Liu, P.; Nicely, N. I.; Wang, S.; Greenleaf, A. L.; Zhou, P. *J. Biol. Chem.* **2011**, 286, 5717.
- (76) (a) Kang, M. E.; Dahmus, M. E. *J. Biol. Chem.* **1995**, 270, 23390. (b) Maxon, M. E.; Goodrich, J. A.; Tjian, R. *Genes Dev.* **1994**, 8, 515. (c) Myers, L. C.; Gustafsson, C. M.; Bushnell, D. A.; Lui, M.; Erdjument-Bromage, H.; Tempst, P.; Kornberg, R. D. *Genes Dev.* **1998**, 12, 45. (d) Svejstrup, J. Q.; Li, Y.; Fellows, J.; Gnatt, A.; Bjorklund, S.; Kornberg, R. D. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, 94, 6075. (e) Usheva, A.; Maldonado, E.; Goldring, A.; Lu, H.; Houbavi, C.; Reinberg, D.; Aloni, Y. *Cell* **1992**, 69, 871.
- (77) Robinson, P. J.; Bushnell, D. A.; Trnka, M. J.; Burlingame, A. L.; Kornberg, R. D. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, 109, 17931.
- (78) (a) Hughes, C. M.; Rozenblatt-Rosen, O.; Milne, T. A.; Copeland, T. D.; Levine, S. S.; Lee, J. C.; Hayes, D. N.; Shanmugam, K. S.; Bhattacharjee, A.; Biondi, C. A.; Kay, G. F.; Hayward, N. K.; Hess, J. L.; Meyerson, M. *Mol. Cell* **2004**, 13, 587. (b) Milne, T. A.; Dou, Y.; Martin, M. E.; Brock, H. W.; Roeder, R. G.; Hess, J. L. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, 102, 14765.
- (79) (a) Krogan, N. J.; Dover, J.; Wood, A.; Schneider, J.; Heidt, J.; Boateng, M. A.; Dean, K.; Ryan, O. W.; Golshani, A.; Johnston, M.; Greenblatt, J. F.; Shilatifard, A. *Mol. Cell* **2003**, 11, 721. (b) Ng, H. H.; Robert, F.; Young, R. A.; Struhl, K. *Mol. Cell* **2003**, 11, 709.
- (80) (a) Kizer, K. O.; Phatnani, H. P.; Shibata, Y.; Hall, H.; Greenleaf, A. L.; Strahl, B. D. *Mol. Cell. Biol.* **2005**, 25, 3305. (b) Li, B.; Howe, L.; Anderson, S.; Yates, J. R., III; Workman, J. L. *J. Biol. Chem.* **2003**, 278, 8897. (c) Xiao, T.; Hall, H.; Kizer, K. O.; Shibata, Y.; Hall, M. C.; Borchers, C. H.; Strahl, B. D. *Genes Dev.* **2003**, 17, 654.
- (81) Sun, X. J.; Wei, J.; Wu, X. Y.; Hu, M.; Wang, L.; Wang, H. H.; Zhang, Q. H.; Chen, S. J.; Huang, Q. H.; Chen, Z. *J. Biol. Chem.* **2005**, 280, 35261.
- (82) (a) Drouin, S.; Laramee, L.; Jacques, P. E.; Forest, A.; Bergeron, M.; Robert, F. *PLoS Genet.* **2010**, 6, e1001173. (b) Govind, C. K.; Qiu, H.; Ginsburg, D. S.; Ruan, C.; Hofmeyer, K.; Hu, C.; Swaminathan, V.; Workman, J. L.; Li, B.; Hinnebusch, A. G. *Mol. Cell* **2010**, 39, 234.
- (83) Yoh, S. M.; Lucas, J. S.; Jones, K. A. *Genes Dev.* **2008**, 22, 3422.
- (84) (a) Sun, M.; Lariviere, L.; Deng, S.; Mayer, A.; Cramer, P. *J. Biol. Chem.* **2010**, 285, 41597. (b) Yoh, S. M.; Cho, H.; Pickle, L.; Evans, R. M.; Jones, K. A. *Genes Dev.* **2007**, 21, 160.
- (85) McCracken, S.; Fong, N.; Yankulov, K.; Ballantyne, S.; Pan, G.; Greenblatt, J.; Patterson, S. D.; Wickens, M.; Bentley, D. L. *Nature* **1997**, 385, 357.
- (86) (a) Cho, E. J.; Takagi, T.; Moore, C. R.; Buratowski, S. *Genes Dev.* **1997**, 11, 3319. (b) Ho, C. K.; Shuman, S. *Mol. Cell* **1999**, 3, 405. (c) McCracken, S.; Fong, N.; Rosonina, E.; Yankulov, K.; Brothers, G.; Siderovski, D.; Hessel, A.; Foster, S.; Shuman, S.; Bentley, D. L. *Genes Dev.* **1997**, 11, 3306.
- (87) Morris, D. P.; Greenleaf, A. L. *J. Biol. Chem.* **2000**, 275, 39935.

- (88) David, C. J.; Boyne, A. R.; Millhouse, S. R.; Manley, J. L. *Genes Dev.* **2011**, *25*, 972.
- (89) (a) Emili, A.; Shales, M.; McCracken, S.; Xie, W.; Tucker, P. W.; Kobayashi, R.; Blencowe, B. J.; Ingles, C. J. *RNA* **2002**, *8*, 1102. (b) Rosonina, E.; Ip, J. Y.; Calarco, J. A.; Bakowski, M. A.; Emili, A.; McCracken, S.; Tucker, P.; Ingles, C. J.; Blencowe, B. J. *Mol. Cell. Biol.* **2005**, *25*, 6734.
- (90) Pascual-Garcia, P.; Govind, C. K.; Queralt, E.; Cuenca-Bono, B.; Llopis, A.; Chavez, S.; Hinnebusch, A. G.; Rodriguez-Navarro, S. *Genes Dev.* **2008**, *22*, 2811.
- (91) MacKellar, A. L.; Greenleaf, A. L. *J. Biol. Chem.* **2011**, *286*, 36385.
- (92) (a) Barilla, D.; Lee, B. A.; Proudfoot, N. J. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 445. (b) Licatalosi, D. D.; Geiger, G.; Minet, M.; Schroeder, S.; Cilli, K.; McNeil, J. B.; Bentley, D. L. *Mol. Cell* **2002**, *9*, 1101.
- (93) Kim, M.; Krogan, N. J.; Vasiljeva, L.; Rando, O. J.; Nedea, E.; Greenblatt, J. F.; Buratowski, S. *Nature* **2004**, *432*, 517.
- (94) (a) Dichtl, B.; Blank, D.; Ohnacker, M.; Friedlein, A.; Roeder, D.; Langen, H.; Keller, W. *Mol. Cell* **2002**, *10*, 1139. (b) Kyburz, A.; Sadowski, M.; Dichtl, B.; Keller, W. *Nucleic Acids Res.* **2003**, *31*, 3936.
- (95) Egloff, S.; Szczepaniak, S. A.; Dienstbier, M.; Taylor, A.; Knight, S.; Murphy, S. *J. Biol. Chem.* **2010**, *285*, 20564.
- (96) (a) Conrad, N. K.; Wilson, S. M.; Steinmetz, E. J.; Paturajan, M.; Brow, D. A.; Swanson, M. S.; Corden, J. L. *Genetics* **2000**, *154*, 557. (b) Steinmetz, E. J.; Conrad, N. K.; Brow, D. A.; Corden, J. L. *Nature* **2001**, *413*, 327.
- (97) Kubicek, K.; Cerna, H.; Holub, P.; Pasulka, J.; Hrossova, D.; Loehr, F.; Hofr, C.; Vanacova, S.; Stefl, R. *Genes Dev.* **2012**, *26*, 1891.
- (98) (a) Chinchilla, K.; Rodriguez-Molina, J. B.; Ursic, D.; Finkel, J. S.; Ansari, A. Z.; Culbertson, M. R. *Eukaryotic Cell* **2012**, *11*, 417. (b) Ursic, D.; Chinchilla, K.; Finkel, J. S.; Culbertson, M. R. *Nucleic Acids Res.* **2004**, *32*, 2441.
- (99) Sikorsky, T.; Hobor, F.; Krizanova, E.; Pasulka, J.; Kubicek, K.; Stefl, R. *Nucleic Acids Res.* **2012**, *40*, 11748.
- (100) Verdecia, M. A.; Bowman, M. E.; Lu, K. P.; Hunter, T.; Noel, J. P. *Nat. Struct. Biol.* **2000**, *7*, 639.
- (101) Zhang, Y.; Kim, Y.; Genoud, N.; Gao, J.; Kelly, J. W.; Pfaff, S. L.; Gill, G. N.; Dixon, J. E.; Noel, J. P. *Mol. Cell* **2006**, *24*, 759.
- (102) Ghosh, A.; Shuman, S.; Lima, C. D. *Mol. Cell* **2011**, *43*, 299.
- (103) Fabrega, C.; Shen, V.; Shuman, S.; Lima, C. D. *Mol. Cell* **2003**, *11*, 1549.
- (104) Becker, R.; Loll, B.; Meinhart, A. *J. Biol. Chem.* **2008**, *283*, 22659.
- (105) Lunde, B. M.; Reichow, S. L.; Kim, M.; Suh, H.; Leeper, T. C.; Yang, F.; Mutschler, H.; Buratowski, S.; Meinhart, A.; Varani, G. *Nat. Struct. Mol. Biol.* **2010**, *17*, 1195.
- (106) Meinhart, A.; Silberzahn, T.; Cramer, P. *J. Biol. Chem.* **2003**, *278*, 15917.
- (107) Xiang, K.; Nagaike, T.; Xiang, S.; Kilic, T.; Beh, M. M.; Manley, J. L.; Tong, L. *Nature* **2010**, *467*, 729.
- (108) Xiang, K.; Manley, J. L.; Tong, L. *Genes Dev.* **2012**, *26*, 2265.
- (109) Orlando, V.; Paro, R. *Cell* **1993**, *75*, 1187.
- (110) (a) Mayer, A.; Lidschreiber, M.; Siebert, M.; Leike, K.; Soding, J.; Cramer, P. *Nat. Struct. Mol. Biol.* **2010**, *17*, 1272. (b) Tietjen, J. R.; Zhang, D. W.; Rodriguez-Molina, J. B.; White, B. E.; Akhtar, M. S.; Heidemann, M.; Li, X.; Chapman, R. D.; Shokat, K.; Keles, S.; Eick, D.; Ansari, A. Z. *Nat. Struct. Mol. Biol.* **2010**, *17*, 1154.
- (111) (a) Brookes, E.; de Santiago, I.; Hebenstreit, D.; Morris, K. J.; Carroll, T.; Xie, S. Q.; Stock, J. K.; Heidemann, M.; Eick, D.; Nozaki, N.; Kimura, H.; Ragoussis, J.; Teichmann, S. A.; Pombo, A. *Cell Stem Cell* **2012**, *10*, 157. (b) Odawara, J.; Harada, A.; Yoshimi, T.; Maehara, K.; Tachibana, T.; Okada, S.; Akashi, K.; Ohkawa, Y. *BMC Genomics* **2011**, *12*, 516.
- (112) Hajheidari, M.; Farrona, S.; Huettel, B.; Koncz, Z.; Koncz, C. *Plant Cell* **2012**, *24*, 1626.
- (113) Lu, H.; Flores, O.; Weinmann, R.; Reinberg, D. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 10004.
- (114) Sikorski, T. W.; Buratowski, S. *Curr. Opin. Cell Biol.* **2009**, *21*, 344.
- (115) Conaway, R. C.; Bradsher, J. N.; Conaway, J. W. *J. Biol. Chem.* **1992**, *267*, 8464.
- (116) Boeing, S.; Rigault, C.; Heidemann, M.; Eick, D.; Meisterernst, M. *J. Biol. Chem.* **2010**, *285*, 188.
- (117) Gerber, H. P.; Hagmann, M.; Seipel, K.; Georgiev, O.; West, M. A.; Littingtung, Y.; Schaffner, W.; Corden, J. L. *Nature* **1995**, *374*, 660.
- (118) (a) Lux, C.; Albize, H.; Chapman, R. D.; Heidinger, M.; Meininghaus, M.; Brack-Werner, R.; Lang, A.; Ziegler, M.; Cremer, T.; Eick, D. *Nucleic Acids Res.* **2005**, *33*, 5139. (b) Meininghaus, M.; Chapman, R. D.; Horndasch, M.; Eick, D. *J. Biol. Chem.* **2000**, *275*, 24375.
- (119) (a) Soutourina, J.; Wydau, S.; Ambroise, Y.; Boschiero, C.; Werner, M. *Science* **2011**, *331*, 1451. (b) Davis, J. A.; Takagi, Y.; Kornberg, R. D.; Asturias, F. A. *Mol. Cell* **2002**, *10*, 409. (c) Lariviere, L.; Plaschka, C.; Seizl, M.; Wenzeck, L.; Kurth, F.; Cramer, P. *Nature* **2012**, *492*, 448. (d) Malik, S.; Roeder, R. G. *Nat. Rev. Genet.* **2010**, *11*, 761.
- (120) Sogaard, T. M.; Sveistrup, J. Q. *J. Biol. Chem.* **2007**, *282*, 14113.
- (121) Core, L. J.; Waterfall, J. J.; Gilchrist, D. A.; Fargo, D. C.; Kwak, H.; Adelman, K.; Lis, J. T. *Cell Rep.* **2012**, *2*, 1025.
- (122) Gilmour, D. S.; Fan, R. *Methods* **2009**, *48*, 368.
- (123) (a) Akoulitchev, S.; Makela, T. P.; Weinberg, R. A.; Reinberg, D. *Nature* **1995**, *377*, 557. (b) Hengartner, C. J.; Myer, V. E.; Liao, S. M.; Wilson, C. J.; Koh, S. S.; Young, R. A. *Mol. Cell* **1998**, *2*, 43.
- (124) (a) Coppola, J. A.; Field, A. S.; Luse, D. S. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 1251. (b) Jove, R.; Manley, J. L. *J. Biol. Chem.* **1984**, *259*, 8513.
- (125) Shuman, S. *Prog. Nucleic Acid Res. Mol. Biol.* **2001**, *66*, 1.
- (126) Suh, M. H.; Meyer, P. A.; Gu, M.; Ye, P.; Zhang, M.; Kaplan, C. D.; Lima, C. D.; Fu, J. *J. Biol. Chem.* **2010**, *285*, 34027.
- (127) Wood, A.; Shilatifard, A. *Cell Cycle* **2006**, *5*, 1066.
- (128) Qiu, H.; Hu, C.; Hinnebusch, A. G. *Mol. Cell* **2009**, *33*, 752.
- (129) Coudreuse, D.; van Bakel, H.; Dewez, M.; Soutourina, J.; Parnell, T.; Vandenhautte, J.; Cairns, B.; Werner, M.; Hermand, D. *Curr. Biol.* **2010**, *20*, 1053.
- (130) Liu, Y.; Warfield, L.; Zhang, C.; Luo, J.; Allen, J.; Lang, W. H.; Ranish, J.; Shokat, K. M.; Hahn, S. *Mol. Cell. Biol.* **2009**, *29*, 4852.
- (131) Narita, T.; Yamaguchi, Y.; Yano, K.; Sugimoto, S.; Chanarat, S.; Wada, T.; Kim, D. K.; Hasegawa, J.; Omori, M.; Inukai, N.; Endoh, M.; Yamada, T.; Handa, H. *Mol. Cell. Biol.* **2003**, *23*, 1863.
- (132) (a) Nechaev, S.; Adelman, K. *Biochim. Biophys. Acta* **2011**, *1809*, 34. (b) Renner, D. B.; Yamaguchi, Y.; Wada, T.; Handa, H.; Price, D. H. *J. Biol. Chem.* **2001**, *276*, 42601.
- (133) (a) Peterlin, B. M.; Price, D. H. *Mol. Cell* **2006**, *23*, 297. (b) Sims, R. J., III; Belotserkovskaya, R.; Reinberg, D. *Genes Dev.* **2004**, *18*, 2437.
- (134) Bartkowiak, B.; Greenleaf, A. L. *Transcription* **2011**, *2*, 115.
- (135) Marshall, N. F.; Price, D. H. *Mol. Cell. Biol.* **1992**, *12*, 2078.
- (136) (a) Chen, D.; Fong, Y.; Zhou, Q. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 2728. (b) Chen, D.; Zhou, Q. *Mol. Cell. Biol.* **1999**, *19*, 2863. (c) Chiu, Y. L.; Coronel, E.; Ho, C. K.; Shuman, S.; Rana, T. M. *J. Biol. Chem.* **2001**, *276*, 12959. (d) Garcia-Martinez, L. F.; Mavankal, G.; Neveu, J. M.; Lane, W. S.; Ivanov, D.; Gaynor, R. B. *EMBO J.* **1997**, *16*, 2836. (e) Isel, C.; Karn, J. *J. Mol. Biol.* **1999**, *290*, 929. (f) Marshall, N. F.; Dahmus, G. K.; Dahmus, M. E. *J. Biol. Chem.* **1998**, *273*, 31726. (g) Okamoto, H.; Sheline, C. T.; Corden, J. L.; Jones, K. A.; Peterlin, B. M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 11575. (h) Parada, C. A.; Roeder, R. G. *Nature* **1996**, *384*, 375. (i) Romano, G.; Kasten, M.; De Falco, G.; Micheli, P.; Khalili, K.; Giordano, A. *J. Cell. Biochem.* **1999**, *75*, 357. (j) Zhou, M.; Halanski, M. A.; Radonovich, M. F.; Kashanchi, F.; Peng, J.; Price, D. H.; Brady, J. N. *Mol. Cell. Biol.* **2000**, *20*, 5077. (k) Zhu, Y.; Pe'ery, T.; Peng, J.; Ramanathan, Y.; Marshall, N.; Marshall, T.; Amendt, B.; Mathews, M. B.; Price, D. H. *Genes Dev.* **1997**, *11*, 2622. (l) Garber, M. E.; Wei, P.; Jones, K. A. *Cold Spring Harbor Symp. Quant. Biol.* **1998**, *63*, 371. (m) Jones, K. A. *Genes Dev.* **1997**, *11*, 2593.

- (137) (a) Bres, V.; Yoh, S. M.; Jones, K. A. *Curr. Opin. Cell Biol.* **2008**, *20*, 334. (b) Eberhardy, S. R.; Farnham, P. J. *J. Biol. Chem.* **2001**, *276*, 48562. (c) Gomes, N. P.; Bjerke, G.; Llorente, B.; Szostek, S. A.; Emerson, B. M.; Espinosa, J. M. *Genes Dev.* **2006**, *20*, 601. (d) Guiguen, A.; Soutourina, J.; Dewez, M.; Tafforeau, L.; Dieu, M.; Raes, M.; Vandenhoute, J.; Werner, M.; Hermand, D. *EMBO J.* **2007**, *26*, 1552. (e) Hanyu-Nakamura, K.; Sonobe-Nojima, H.; Tanigawa, A.; Lasko, P.; Nakamura, A. *Nature* **2008**, *451*, 730. (f) Kumano, G.; Takatori, N.; Negishi, T.; Takada, T.; Nishida, H. *Curr. Biol.* **2011**, *21*, 1308. (g) Lee, D. K.; Duan, H. O.; Chang, C. J. *Biol. Chem.* **2001**, *276*, 9978. (h) Mitra, P.; Pereira, L. A.; Drabsch, Y.; Ramsay, R. G.; Gonda, T. J. *Nucleic Acids Res.* **2012**, *40*, 5988. (i) Murray, S.; Udupa, R.; Yao, S.; Hartzog, G.; Prellich, G. *Mol. Cell. Biol.* **2001**, *21*, 4089. (j) Ni, Z.; Saunders, A.; Fuda, N. J.; Yao, J.; Suarez, J. R.; Webb, W. W.; Lis, J. T. *Mol. Cell. Biol.* **2008**, *28*, 1161. (k) Park, J. H.; Ahn, S. H. *Biochem. Biophys. Res. Commun.* **2010**, *392*, 588.
- (138) Schwartz, J. C.; Ebmeier, C. C.; Podell, E. R.; Heimiller, J.; Taatjes, D. J.; Cech, T. R. *Genes Dev.* **2012**, *26*, 2690.
- (139) (a) Schulte, A.; Czudnochowski, N.; Barboric, M.; Schonichen, A.; Blazek, D.; Peterlin, B. M.; Geyer, M. *J. Biol. Chem.* **2005**, *280*, 24968. (b) Barboric, M.; Yik, J. H.; Czudnochowski, N.; Yang, Z.; Chen, R.; Contreras, X.; Geyer, M.; Matija Peterlin, B.; Zhou, Q. *Nucleic Acids Res.* **2007**, *35*, 2003. (c) He, N.; Jahchan, N. S.; Hong, E.; Li, Q.; Bayfield, M. A.; Maraia, R. J.; Luo, K.; Zhou, Q. *Mol. Cell.* **2008**, *29*, 588. (d) Yang, Z.; Yik, J. H.; Chen, R.; He, N.; Jang, M. K.; Ozato, K.; Zhou, Q. *Mol. Cell.* **2005**, *19*, 535. (e) Yang, Z.; Zhu, Q.; Luo, K.; Zhou, Q. *Nature* **2001**, *414*, 317. (f) Yik, J. H.; Chen, R.; Nishimura, R.; Jennings, J. L.; Link, A. J.; Zhou, Q. *Mol. Cell.* **2003**, *12*, 971. (g) Zhou, Q.; Li, T.; Price, D. H. *Annu. Rev. Biochem.* **2012**, *81*, 119. (h) Barrandon, C.; Bonnet, F.; Nguyen, V. T.; Labas, V.; Bensaude, O. *Mol. Cell. Biol.* **2007**, *27*, 6996. (i) Krueger, B. J.; Jeronimo, C.; Roy, B. B.; Bouchard, A.; Barrandon, C.; Byers, S. A.; Searcey, C. E.; Cooper, J. J.; Bensaude, O.; Cohen, E. A.; Coulombe, B.; Price, D. H. *Nucleic Acids Res.* **2008**, *36*, 2219. (j) Michels, A. A.; Fraldi, A.; Li, Q.; Adamson, T. E.; Bonnet, F.; Nguyen, V. T.; Sedore, S. C.; Price, J. P.; Price, D. H.; Lania, L.; Bensaude, O. *EMBO J.* **2004**, *23*, 2608. (k) Michels, A. A.; Nguyen, V. T.; Fraldi, A.; Labas, V.; Edwards, M.; Bonnet, F.; Lania, L.; Bensaude, O. *Mol. Cell. Biol.* **2003**, *23*, 4859. (l) Nguyen, V. T.; Kiss, T.; Michels, A. A.; Bensaude, O. *Nature* **2001**, *414*, 322.
- (140) Larochelle, S.; Amat, R.; Glover-Cutter, K.; Sanso, M.; Zhang, C.; Allen, J. J.; Shokat, K. M.; Bentley, D. L.; Fisher, R. P. *Nat. Struct. Mol. Biol.* **2012**, *19*, 1108.
- (141) (a) Devaiah, B. N.; Singer, D. S. *J. Biol. Chem.* **2012**, *287*, 38755. (b) Dow, E. C.; Liu, H.; Rice, A. P. *J. Cell. Physiol.* **2010**, *224*, 84.
- (142) (a) Eberhardy, S. R.; Farnham, P. J. *J. Biol. Chem.* **2002**, *277*, 40156. (b) Kanazawa, S.; Soucek, L.; Evan, G.; Okamoto, T.; Peterlin, B. M. *Oncogene* **2003**, *22*, 5707.
- (143) Barboric, M.; Nissen, R. M.; Kanazawa, S.; Jabrane-Ferrat, N.; Peterlin, B. M. *Mol. Cell.* **2001**, *8*, 327.
- (144) Bisgrove, D. A.; Mahmoudi, T.; Henklein, P.; Verdin, E. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 13690.
- (145) (a) Ai, N.; Hu, X.; Ding, F.; Yu, B.; Wang, H.; Lu, X.; Zhang, K.; Li, Y.; Han, A.; Lin, W.; Liu, R.; Chen, R. *Nucleic Acids Res.* **2011**, *39*, 9592. (b) Bartholomeeusen, K.; Xiang, Y.; Fujinaga, K.; Peterlin, B. M. *J. Biol. Chem.* **2012**, *287*, 36609. (c) Jang, M. K.; Mochizuki, K.; Zhou, M.; Jeong, H. S.; Brady, J. N.; Ozato, K. *Mol. Cell.* **2005**, *19*, 523. (d) Palermo, R. D.; Webb, H. M.; West, M. J. *PLoS Pathog.* **2011**, *7*, e1002334. (e) Schroder, S.; Cho, S.; Zeng, L.; Zhang, Q.; Kaehlcke, K.; Mak, L.; Lau, J.; Bisgrove, D.; Schnolzer, M.; Verdin, E.; Zhou, M. M.; Ott, M. *J. Biol. Chem.* **2012**, *287*, 1090. (f) Vollmuth, F.; Blankenfeldt, W.; Geyer, M. *J. Biol. Chem.* **2009**, *284*, 36547. (g) Wang, R.; Li, Q.; Helfer, C. M.; Jiao, J.; You, J. *J. Biol. Chem.* **2012**, *287*, 10738. (h) Wu, S. Y.; Chiang, C. M. *J. Biol. Chem.* **2007**, *282*, 13141. (i) Yang, Z.; He, N.; Zhou, Q. *Mol. Cell. Biol.* **2008**, *28*, 967. (j) Zhang, W.; Prakash, C.; Sum, C.; Gong, Y.; Li, Y.; Kwok, J. J.; Thiessen, N.; Pettersson, S.; Jones, S. J.; Knapp, S.; Yang, H.; Chin, K. C. *J. Biol. Chem.* **2012**, *287*, 43137. (k) Zhou, M.; Huang, K.; Jung, K. J.; Cho, W. K.; Klase, Z.; Kashanchi, F.; Pise-Masison, C. A.; Brady, J. N. *J. Virol.* **2009**, *83*, 1036. (l) Zhou, Q.; Yik, J. H. *Microbiol. Mol. Biol. Rev.* **2006**, *70*, 646. (146) Devaiah, B. N.; Singer, D. S. *Transcription* **2012**, *4*. (147) (a) Kedinger, C.; Gniazdowski, M.; Mandel, J. L., Jr.; Gissinger, F.; Chambon, P. *Biochem. Biophys. Res. Commun.* **1970**, *38*, 165. (b) Lindell, T. J.; Weinberg, F.; Morris, P. W.; Roeder, R. G.; Rutter, W. J. *Science* **1970**, *170*, 447. (148) (a) Bartolomei, M. S.; Corden, J. L. *Mol. Cell. Biol.* **1987**, *7*, 586. (b) Bartolomei, M. S.; Corden, J. L. *Mol. Gen. Genet.* **1995**, *246*, 778. (149) Rosonina, E.; Blencowe, B. J. *RNA* **2004**, *10*, 581. (150) (a) Chapman, R. D.; Conrad, M.; Eick, D. *Mol. Cell. Biol.* **2005**, *25*, 7665. (b) Chapman, R. D.; Palancade, B.; Lang, A.; Bensaude, O.; Eick, D. *Nucleic Acids Res.* **2004**, *32*, 35. (151) (a) Baskaran, R.; Chiang, G. G.; Mysliwiec, T.; Kruh, G. D.; Wang, J. Y. *J. Biol. Chem.* **1997**, *272*, 18905. (b) Baskaran, R.; Escobar, S. R.; Wang, J. Y. *Cell Growth Differ.* **1999**, *10*, 387. (c) Duyster, J.; Baskaran, R.; Wang, J. Y. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 1555. (152) (a) Fong, N.; Bentley, D. L. *Genes Dev.* **2001**, *15*, 1783. (b) Fong, N.; Bird, G.; Vigneron, M.; Bentley, D. L. *EMBO J.* **2003**, *22*, 4274. (153) (a) Rappsilber, J.; Ryder, U.; Lamond, A. I.; Mann, M. *Genome Res.* **2002**, *12*, 1231. (b) Wahl, M. C.; Will, C. L.; Luhrmann, R. *Cell* **2009**, *136*, 701. (154) Hirose, Y.; Tacke, R.; Manley, J. L. *Genes Dev.* **1999**, *13*, 1234. (155) Spiluttini, B.; Gu, B.; Belagal, P.; Smirnova, A. S.; Nguyen, V. T.; Hebert, C.; Schmidt, U.; Bertrand, E.; Darzacq, X.; Bensaude, O. *J. Cell Sci.* **2010**, *123*, 2085. (156) Kang, C. H.; Feng, Y.; Vikram, M.; Jeong, I. S.; Lee, J. R.; Bahk, J. D.; Yun, D. J.; Lee, S. Y.; Koiwa, H. *Arch. Biochem. Biophys.* **2009**, *484*, 30. (157) Phatnani, H. P.; Jones, J. C.; Greenleaf, A. L. *Biochemistry* **2004**, *43*, 15702. (158) Kwek, K. Y.; Murphy, S.; Furger, A.; Thomas, B.; O'Gorman, W.; Kimura, H.; Proudfoot, N. J.; Akoulitchev, A. *Nat. Struct. Biol.* **2002**, *9*, 800. (159) Jobert, L.; Pinzon, N.; Van Herreweghe, E.; Jady, B. E.; Guiialis, A.; Kiss, T.; Tora, L. *EMBO Rep.* **2009**, *10*, 494. (160) Gornemann, J.; Barrandon, C.; Hujer, K.; Rutz, B.; Rigaut, G.; Kotovic, K. M.; Faux, C.; Neugebauer, K. M.; Seraphin, B. *RNA* **2011**, *17*, 2119. (161) (a) Carte, S. M.; Goldstrohm, A. C.; Sune, C.; Garcia-Blanco, M. A.; Greenleaf, A. L. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9015. (b) Goldstrohm, A. C.; Albrecht, T. R.; Sune, C.; Bedford, M. T.; Garcia-Blanco, M. A. *Mol. Cell. Biol.* **2001**, *21*, 7617. (c) Smith, M. J.; Kulkarni, S.; Pawson, T. *Mol. Cell. Biol.* **2004**, *24*, 9274. (162) David, C. J.; Manley, J. L. *Transcription* **2011**, *2*, 221. (163) Gu, B.; Eick, D.; Bensaude, O. *Nucleic Acids Res.* **2012**. (164) Eberle, A. B.; Hesse, V.; Helbig, R.; Dantoft, W.; Gimber, N.; Visa, N. *PLoS One* **2010**, *5*, e11540. (165) (a) Bres, V.; Gomes, N.; Pickle, L.; Jones, K. A. *Genes Dev.* **2005**, *19*, 1211. (b) Damgaard, C. K.; Kahns, S.; Lykke-Andersen, S.; Nielsen, A. L.; Jensen, T. H.; Kjems, J. *Mol. Cell.* **2008**, *29*, 271. (c) Furger, A.; O'Sullivan, J. M.; Binnie, A.; Lee, B. A.; Proudfoot, N. J. *Genes Dev.* **2002**, *16*, 2792. (d) Lin, S.; Coutinho-Mansfield, G.; Wang, D.; Pandit, S.; Fu, X. D. *Nat. Struct. Mol. Biol.* **2008**, *15*, 819. (166) (a) Arigo, J. T.; Carroll, K. L.; Ames, J. M.; Corden, J. L. *Mol. Cell* **2006**, *21*, 641. (b) Arigo, J. T.; Eyler, D. E.; Carroll, K. L.; Corden, J. L. *Mol. Cell* **2006**, *23*, 841. (c) Birse, C. E.; Minvielle-Sebastia, L.; Lee, B. A.; Keller, W.; Proudfoot, N. J. *Science* **1998**, *280*, 298. (d) Buratowski, S. *Curr. Opin. Cell Biol.* **2005**, *17*, 257. (e) Gudipati, R. K.; Villa, T.; Boulay, J.; Libri, D. *Nat. Struct. Mol. Biol.* **2008**, *15*, 786. (f) Kim, M.; Vasiljeva, L.; Rando, O. J.; Zhelkovsky, A.; Moore, C.; Buratowski, S. *Mol. Cell* **2006**, *24*, 723. (g) Lykke-Andersen, S.; Jensen, T. H. *Biochimie* **2007**, *89*, 1177. (h) Richard, P.; Manley, J. L. *Genes Dev.* **2009**, *23*, 1247. (167) (a) Baillat, D.; Hakimi, M. A.; Naar, A. M.; Shilatifard, A.; Cooch, N.; Shiekhattar, R. *Cell* **2005**, *123*, 265. (b) Egloff, S. *RNA Biol.*

- 2012, 9, 1033. (c) Egloff, S.; O'Reilly, D.; Chapman, R. D.; Taylor, A.; Tanzhaus, K.; Pitts, L.; Eick, D.; Murphy, S. *Science* 2007, 318, 1777. (168) Hirose, Y.; Manley, J. L. *Nature* 1998, 395, 93.
- (169) (a) Terzi, N.; Churchman, L. S.; Vasiljeva, L.; Weissman, J.; Buratowski, S. *Mol. Cell. Biol.* 2011, 31, 3569. (b) Tomson, B. N.; Crisucci, E. M.; Heisler, L. E.; Gebbia, M.; Nislow, C.; Arndt, K. M. *Mol. Cell. Biol.* 2013, 33, 170. (c) Vasiljeva, L.; Kim, M.; Mutschler, H.; Buratowski, S.; Meinhart, A. *Nat. Struct. Mol. Biol.* 2008, 15, 795.
- (170) Carroll, K. L.; Pradhan, D. A.; Granek, J. A.; Clarke, N. D.; Corden, J. L. *Mol. Cell. Biol.* 2004, 24, 6241.
- (171) (a) Creamer, T. J.; Darby, M. M.; Jamonnak, N.; Schaughency, P.; Hao, H.; Wheelan, S. J.; Corden, J. L. *PLoS Genet.* 2011, 7, e1002329. (b) Jamonnak, N.; Creamer, T. J.; Darby, M. M.; Schaughency, P.; Wheelan, S. J.; Corden, J. L. *RNA* 2011, 17, 2011.
- (172) Nedea, E.; Nalbant, D.; Xia, D.; Theoharis, N. T.; Suter, B.; Richardson, C. J.; Tatchell, K.; Kislinger, T.; Greenblatt, J. F.; Nagy, P. L. *Mol. Cell* 2008, 29, 577.
- (173) Singh, N.; Ma, Z.; Gemmill, T.; Wu, X.; Defiglio, H.; Rosettini, A.; Rabeler, C.; Beane, O.; Morse, R. H.; Palumbo, M. J.; Hanes, S. D. *Mol. Cell* 2009, 36, 255.
- (174) Ni, Z.; Olsen, J. B.; Guo, X.; Zhong, G.; Ruan, E. D.; Marcon, E.; Young, P.; Guo, H.; Li, J.; Moffat, J.; Emili, A.; Greenblatt, J. F. *Transcription* 2011, 2, 237.
- (175) (a) Medlin, J.; Scurry, A.; Taylor, A.; Zhang, F.; Peterlin, B. M.; Murphy, S. *EMBO J.* 2005, 24, 4154. (b) Medlin, J. E.; Uguen, P.; Taylor, A.; Bentley, D. L.; Murphy, S. *EMBO J.* 2003, 22, 925. (c) Uguen, P.; Murphy, S. *EMBO J.* 2003, 22, 4544.
- (176) Hagenbuchle, O.; Wellauer, P. K.; Cribbs, D. L.; Schibler, U. *Cell* 1984, 38, 737.
- (177) Connolly, S.; Manley, J. L. *Genes Dev.* 1988, 2, 440.
- (178) (a) Anamika, K.; Gyenis, A.; Poidevin, L.; Poch, O.; Tora, L. *PLoS One* 2012, 7, e38769. (b) Anamika, K.; Gyenis, A.; Tora, L. *Transcription* 2012, 4.
- (179) (a) Kawauchi, J.; Mischo, H.; Braglia, P.; Rondon, A.; Proudfoot, N. J. *Genes Dev.* 2008, 22, 1082. (b) West, S.; Gromak, N.; Proudfoot, N. J. *Nature* 2004, 432, 522.
- (180) O'Sullivan, J. M.; Tan-Wong, S. M.; Morillon, A.; Lee, B.; Coles, J.; Mellor, J.; Proudfoot, N. J. *Nat. Genet.* 2004, 36, 1014.
- (181) Strahl, B. D.; Allis, C. D. *Nature* 2000, 403, 41.
- (182) Dey, A.; Chitsaz, F.; Abbasi, A.; Misteli, T.; Ozato, K. *Proc. Natl. Acad. Sci. U.S.A.* 2003, 100, 8758.
- (183) Wang, Z.; Zang, C.; Cui, K.; Schones, D. E.; Barski, A.; Peng, W.; Zhao, K. *Cell* 2009, 138, 1019.
- (184) Zippo, A.; Serafini, R.; Rocchigiani, M.; Pennacchini, S.; Krepelova, A.; Oliviero, S. *Cell* 2009, 138, 1122.
- (185) Kim, T.; Buratowski, S. *Cell* 2009, 137, 259.
- (186) Lee, J. H.; Skalnik, D. G. *Mol. Cell. Biol.* 2008, 28, 609.
- (187) (a) Mueller, D.; Bach, C.; Zeisig, D.; Garcia-Cuellar, M. P.; Monroe, S.; Sreekumar, A.; Zhou, R.; Nesvizhskii, A.; Chinnaiyan, A.; Hess, J. L.; Slany, R. K. *Blood* 2007, 110, 4445. (b) Smith, E. R.; Lin, C.; Garrett, A. S.; Thornton, J.; Mohaghegh, N.; Hu, D.; Jackson, J.; Saraf, A.; Swanson, S. K.; Seidel, C.; Florens, L.; Washburn, M. P.; Eissenberg, J. C.; Shilatifard, A. *Mol. Cell* 2011, 44, 954.
- (188) (a) Bintu, L.; Ishibashi, T.; Dangkulwanich, M.; Wu, Y. Y.; Lubkowska, L.; Kashlev, M.; Bustamante, C. *Cell* 2012, 151, 738. (b) Bintu, L.; Kopaczynska, M.; Hodges, C.; Lubkowska, L.; Kashlev, M.; Bustamante, C. *Nat. Struct. Mol. Biol.* 2011, 18, 1394. (c) Hodges, C.; Bintu, L.; Lubkowska, L.; Kashlev, M.; Bustamante, C. *Science* 2009, 325, 626. (d) Kulaeva, O. I.; Gaykalova, D. A.; Pestov, N. A.; Golovastov, V. V.; Vassilyev, D. G.; Artsimovitch, I.; Studitsky, V. M. *Nat. Struct. Mol. Biol.* 2009, 16, 1272. (e) Kulaeva, O. I.; Hsieh, F. K.; Chang, H. W.; Luse, D. S.; Studitsky, V. M. *Biochim. Biophys. Acta* 2013, 1829, 76. (f) Petesch, S. J.; Lis, J. T. *Trends Genet.* 2012, 28, 285.
- (189) (a) Orphanides, G.; LeRoy, G.; Chang, C. H.; Luse, D. S.; Reinberg, D. *Cell* 1998, 92, 105. (b) Orphanides, G.; Wu, W. H.; Lane, W. S.; Hampsey, M.; Reinberg, D. *Nature* 1999, 400, 284.
- (190) (a) Belotserkovskaya, R.; Oh, S.; Bondarenko, V. A.; Orphanides, G.; Studitsky, V. M.; Reinberg, D. *Science* 2003, 301, 1090. (b) Pavri, R.; Zhu, B.; Li, G.; Trojer, P.; Mandal, S.; Shilatifard, A.; Reinberg, D. *Cell* 2006, 125, 703.
- (191) (a) Carrozza, M. J.; Li, B.; Florens, L.; Suganuma, T.; Swanson, S. K.; Lee, K. K.; Shia, W. J.; Anderson, S.; Yates, J.; Washburn, M. P.; Workman, J. L. *Cell* 2005, 123, 581. (b) Joshi, A. A.; Struhl, K. *Mol. Cell* 2005, 20, 971. (c) Keogh, M. C.; Kurdustani, S. K.; Morris, S. A.; Ahn, S. H.; Podolny, V.; Collins, S. R.; Schuldtner, M.; Chin, K.; Punna, T.; Thompson, N. J.; Boone, C.; Emili, A.; Weissman, J. S.; Hughes, T. R.; Strahl, B. D.; Grunstein, M.; Greenblatt, J. F.; Buratowski, S.; Krogan, N. J. *Cell* 2005, 123, 593.
- (192) Carvalho, S.; Raposo, A. C.; Martins, F. B.; Gross, A. R.; Sridhara, S. C.; Rino, J.; Carmo-Fonseca, M.; de Almeida, S. F. *Nucleic Acids Res.* 2013.
- (193) (a) Chasin, L. A. *Adv. Exp. Med. Biol.* 2007, 623, 85. (b) de la Mata, M.; Alonso, C. R.; Kadener, S.; Fededa, J. P.; Blaustein, M.; Pelisch, F.; Cramer, P.; Bentley, D.; Kornblith, A. R. *Mol. Cell* 2003, 12, 525.
- (194) (a) Andersson, R.; Enroth, S.; Rada-Iglesias, A.; Wadelius, C.; Komorowski, J. *Genome Res.* 2009, 19, 1732. (b) Chodavarapu, R. K.; Feng, S.; Bernatavichute, Y. V.; Chen, P. Y.; Stroud, H.; Yu, Y.; Hetzel, J. A.; Kuo, F.; Kim, J.; Cokus, S. J.; Casero, D.; Bernal, M.; Huijser, P.; Clark, A. T.; Kramer, U.; Merchant, S. S.; Zhang, X.; Jacobsen, S. E.; Pellegrini, M. *Nature* 2010, 466, 388. (c) Dhami, P.; Saffrey, P.; Bruce, A. W.; Dillon, S. C.; Chiang, K.; Bonhoure, N.; Koch, C. M.; Bye, J.; James, K.; Foad, N. S.; Ellis, P.; Watkins, N. A.; Ouwehand, W. H.; Langford, C.; Andrews, R. M.; Dunham, I.; Vetrici, D. *PLoS One* 2010, 5, e12339. (d) Kolasinska-Zwierz, P.; Down, T.; Latorre, I.; Liu, T.; Liu, X. S.; Ahringer, J. *Nat. Genet.* 2009, 41, 376. (e) Nahkuri, S.; Taft, R. J.; Mattick, J. S. *Cell Cycle* 2009, 8, 3420. (f) Schwartz, S.; Meshorer, E.; Ast, G. *Nat. Struct. Mol. Biol.* 2009, 16, 990. (g) Spies, N.; Nielsen, C. B.; Padgett, R. A.; Burge, C. B. *Mol. Cell* 2009, 36, 245. (h) Tilgner, H.; Nikolaou, C.; Althammer, S.; Sammeth, M.; Beato, M.; Valcarcel, J.; Guigo, R. *Nat. Struct. Mol. Biol.* 2009, 16, 996.
- (195) (a) de Almeida, S. F.; Gross, A. R.; Koch, F.; Fenouil, R.; Carvalho, S.; Andrade, J.; Levezinho, H.; Gut, M.; Eick, D.; Gut, I.; Andrau, J. C.; Ferrier, P.; Carmo-Fonseca, M. *Nat. Struct. Mol. Biol.* 2011, 18, 977. (b) Luco, R. F.; Pan, Q.; Tominaga, K.; Blencowe, B. J.; Pereira-Smith, O. M.; Misteli, T. *Science* 2010, 327, 996.
- (196) Shukla, S.; Kavak, E.; Gregory, M.; Imashimizu, M.; Shutinoski, B.; Kashlev, M.; Oberdoerffer, P.; Sandberg, R.; Oberdoerffer, S. *Nature* 2011, 479, 74.
- (197) Gunderson, F. Q.; Johnson, T. L. *PLoS Genet.* 2009, 5, e1000682.
- (198) Sims, R. J., III; Millhouse, S.; Chen, C. F.; Lewis, B. A.; Erdjument-Bromage, H.; Tempst, P.; Manley, J. L.; Reinberg, D. *Mol. Cell* 2007, 28, 665.
- (199) (a) Lane, C. E.; van den Heuvel, K.; Kozera, C.; Curtis, B. A.; Parsons, B. J.; Bowman, S.; Archibald, J. M. *Proc. Natl. Acad. Sci. U.S.A.* 2007, 104, 19908. (b) Mair, G.; Shi, H.; Li, H.; Djikeng, A.; Aviles, H. O.; Bishop, J. R.; Falcone, F. H.; Gavrilescu, C.; Montgomery, J. L.; Santori, M. I.; Stern, L. S.; Wang, Z.; Ullu, E.; Tschudi, C. *C. RNA* 2000, 6, 163. (c) Morrison, H. G.; McArthur, A. G.; Gillin, F. D.; Aley, S. B.; Adam, R. D.; Olsen, G. J.; Best, A. A.; Cande, W. Z.; Chen, F.; Cipriano, M. J.; Davids, B. J.; Dawson, S. C.; Elmendorf, H. G.; Hehl, A. B.; Holder, M. E.; Huse, S. M.; Kim, U. U.; Lasek-Nesselquist, E.; Manning, G.; Nigam, A.; Nixon, J. E.; Palm, D.; Passamanec, N. E.; Prabhu, A.; Reich, C. I.; Reiner, D. S.; Samuelson, J.; Svard, S. G.; Sogin, M. L. *Science* 2007, 317, 1921. (d) Sakharkar, M. K.; Chow, V. T.; Kanguane, P. *In Silico Biol.* 2004, 4, 387.
- (200) Chorev, M.; Carmel, L. *Front. Genet.* 2012, 3, 55.
- (201) (a) Carmel, L.; Wolf, Y. I.; Rogozin, I. B.; Koonin, E. V. *Genome Res.* 2007, 17, 1034. (b) Csuros, M.; Rogozin, I. B.; Koonin, E. V. *PLoS Comput. Biol.* 2011, 7, e1002150.
- (202) Sandelin, A.; Carninci, P.; Lenhard, B.; Ponjavic, J.; Hayashizaki, Y.; Hume, D. A. *Nat. Rev. Genet.* 2007, 8, 424.
- (203) Katayama, S.; Tomaru, Y.; Kasukawa, T.; Waki, K.; Nakanishi, M.; Nakamura, M.; Nishida, H.; Yap, C. C.; Suzuki, M.; Kawai, J.; Suzuki, H.; Carninci, P.; Hayashizaki, Y.; Wells, C.; Frith, M.; Ravasi, T.; Pang, K. C.; Hallinan, J.; Mattick, J.; Hume, D. A.; Lipovich, L.;

- Batalov, S.; Engstrom, P. G.; Mizuno, Y.; Faghihi, M. A.; Sandelin, A.; Chalk, A. M.; Mottagui-Tabar, S.; Liang, Z.; Lenhard, B.; Wahlestedt, C. *Science* **2005**, *309*, 1564.
- (204) Maston, G. A.; Evans, S. K.; Green, M. R. *Annu. Rev. Genomics Hum. Genet.* **2006**, *7*, 29.
- (205) (a) Bode, J.; Benham, C.; Knopp, A.; Mielke, C. *Crit. Rev. Eukaryotic Gene Expression* **2000**, *10*, 73. (b) Chattopadhyay, S.; Pavithra, L. *Subcell. Biochem.* **2007**, *41*, 213. (c) Gaszner, M.; Felsenfeld, G. *Nat. Rev. Genet.* **2006**, *7*, 703.
- (206) Hong, X.; Scofield, D. G.; Lynch, M. *Mol. Biol. Evol.* **2006**, *23*, 2392.
- (207) Dye, M. J.; Gromak, N.; Proudfoot, N. J. *Mol. Cell* **2006**, *21*, 849.
- (208) Windhager, L.; Bonfert, T.; Burger, K.; Ruzsics, Z.; Krebs, S.; Kaufmann, S.; Malterer, G.; L'Hernault, A.; Schilhabel, M.; Schreiber, S.; Rosenstiel, P.; Zimmer, R.; Eick, D.; Friedel, C. C.; Dolken, L. *Genome Res.* **2012**, *22*, 2031.
- (209) Min, I. M.; Selsing, E. *Adv. Immunol.* **2005**, *87*, 297.
- (210) Alt, F. W.; Bothwell, A. L.; Knapp, M.; Siden, E.; Mather, E.; Koshland, M.; Baltimore, D. *Cell* **1980**, *20*, 293.
- (211) Stadhouders, R.; Thongjuea, S.; Andrieu-Soler, C.; Palstra, R. J.; Bryne, J. C.; van den Heuvel, A.; Stevens, M.; de Boer, E.; Kockx, C.; van der Sloot, A.; van den Hout, M.; van IJcken, W.; Eick, D.; Lenhard, B.; Grosveld, F.; Soler, E. *EMBO J.* **2012**, *31*, 986.
- (212) Alexander, R. D.; Innocente, S. A.; Barrass, J. D.; Beggs, J. D. *Mol. Cell* **2010**, *40*, 582.
- (213) Munoz, M. J.; Perez Santangelo, M. S.; Paronetto, M. P.; de la Mata, M.; Pelisch, F.; Boireau, S.; Glover-Cutter, K.; Ben-Dov, C.; Blaustein, M.; Lozano, J. J.; Bird, G.; Bentley, D.; Bertrand, E.; Kornblith, A. R. *Cell* **2009**, *137*, 708.
- (214) Rodrigues, F.; Thuma, L.; Klammt, C. *Development* **2012**, *139*, 1765.
- (215) Hoeppner, M. P.; White, S.; Jeffares, D. C.; Poole, A. M. *Genome Biol. Evol.* **2009**, *1*, 420.
- (216) Burger, K.; Muehl, B.; Rohrmoser, M.; Coordes, B.; Heidemann, M.; Kellner, M.; Gruber-Eber, A.; Heissmeyer, V.; Strasser, K.; Eick, D. *J. Biol. Chem.* **2013**.
- (217) Schmucker, D.; Chen, B. *Genes Dev.* **2009**, *23*, 147.
- (218) Custodio, N.; Vivo, M.; Antoniou, M.; Carmo-Fonseca, M. J. *Cell Biol.* **2007**, *179*, 199.
- (219) Bono, F.; Gehring, N. H. *RNA Biol.* **2011**, *8*, 24.
- (220) Chang, Y. F.; Imam, J. S.; Wilkinson, M. F. *Annu. Rev. Biochem.* **2007**, *76*, 51.