CHEMICAL REVIEWS

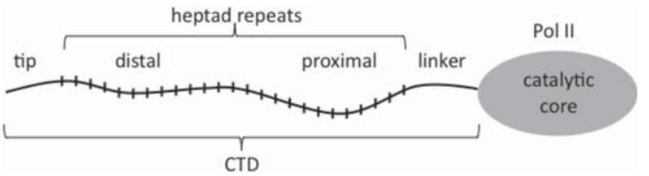
Review

pubs.acs.org/CR

RNA Polymerase II C-Terminal Domain: Tethering Transcription to Transcript and Template

Jeffry L. Corden*

Department of Molecular Biology and Genetics, The Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore Maryland 21205, United States



CONTENTS

1. Introduction 8424
2. Evolution of the CTD 8424
2.1. Origin of the CTD 8425
2.2. Expansion of the CTD 8426
2.3. CTD Clade 8426
2.4. Vertebrate CTDs 8426
2.5. Summary of CTD Evolution 8427
3. CTD Structure 8427
3.1. Modeling 8427
3.2. Solution Structure 8427
3.3. Crystallographic Analysis 8427
3.4. Potential Dimensions in Vivo 8428
4. Genetic Dissection of CTD Function 8428
4.1. CTD Deletions in Yeast 8428
4.2. Deletions in Metazoa 8428
4.3. Mutation of the Consensus Heptad 8429
4.4. Spacing of Heptad Repeats 8429
4.5. Proximal vs Distal Consensus Repeats 8430
4.6. CTD Mutation Summary 8430
5. Functional Analysis of CTD Mutants 8430
5.1. Suppression of CTD Mutations 8430
5.2. Changes in Transcription and Processing Due to CTD Mutation 8430
5.2.1. Activation and Initiation 8430
5.2.2. Elongation 8431
5.2.3. RNA Processing: Capping 8431
5.2.4. RNA Processing: Splicing 8431
5.2.5. RNA Processing: 3′-End Formation and Termination 8431
5.2.6. RNA Transport 8432
5.2.7. Chromatin Modification 8432
5.2.8. DNA Replication and Repair 8432
6. Writing the CTD Code: CTD Kinases 8433
6.1. CTD Phosphorylation 8433
6.1.1. Characterization of CTD Phosphorylation Sites 8433
6.1.2. Chemical Genetic Approach to Kinase Function 8433
6.1.3. Limitations to Characterization of Phosphorylation Sites 8434
6.2. Cyclin-Dependent Kinases Targeting the CTD 8434
6.2.1. CDK7: Structure and Biochemistry 8435
6.2.2. CDK7: Function and Genetics 8435
6.2.3. Cdk8 8436
6.2.4. Cdk9: Structure and Biochemistry 8436
6.2.5. Cdk9: Function and Genetics 8436
6.2.6. Cdk12 8437
6.2.7. Cdk1 8437
6.3. Other CTD Kinases 8437
6.3.1. MAP Kinases 8437
6.3.2. DNA-PK 8437
6.3.3. BRD4 8437
6.3.4. Plk3 8438
6.3.5. Tyrosine Kinases 8438
6.3.6. Casein Kinase II 8438
6.4. Summary of CTD Kinases 8438
6.5. CTD Glycosylation 8438
6.6. Other CTD Modifications 8439
7. CTD Code Erasers: CTD Phosphatases 8439
7.1. Fcp1 CTD Phosphatase Family 8439
7.2. CTD Phosphatase Ssu72 8439
7.3. Rtr1/RPAP2 Plays a Role in CTD Dephosphorylation 8440
8. Rearranging the CTD Code: Proline Isomerases 8440
8.1. Pin1 8440
8.2. Ess1 8440
8.3. Rrd1 8440
9. Reading the CTD Code 8440
9.1. Mediator 8441
9.2. Capping Complex 8441
9.3. WW Domain Interacts with Phosphorylated CTD Repeats 8442
9.3.1. Pin1/Ess1 8442
9.4. CTD-Interacting Domains (CIDs) and Transcription Termination 8442
9.4.1. PCF11 8443
9.4.2. Nrd1 CID 8443
9.4.3. SCAF8/RBM16 CID 8443
9.4.4. Rtt103 CID 8443
9.4.5. CID-CTD Interaction Summary 8443
9.5. SH2 Domain of SPT6 Interacts with the CTD 8444
9.6. Histone Methylase Set2-Rpb1 Interacting (SRI) Domain 8444
9.7. FF Domain Proteins 8444
9.8. FUS and Related Proteins 8444
9.9. Other CTD-Interacting Proteins 8444

Special Issue: 2013 Gene Expression

Received: March 13, 2013
Published: September 16, 2013

dx.doi.org/10.1021/cr400158h | *Chem. Rev.* 2013, 113, 8423−8455

9.10. Summary of CTD Interactions 8444
10. Global Changes in CTD Phosphorylation 8444
10.1. Stress-Induced Changes in CTD Phosphorylation 8444
10.2. Growth Stimulation and CTD Phosphorylation 8445
10.3. CTD Phosphorylation in Early Development 8445
11. CTD Modification and Pol II Subnuclear Localization 8446
12. Concluding Remarks 8446
Author Information 8448
Corresponding Author 8448
Notes 8448
Biography 8448
Acknowledgments 8448
References 8448

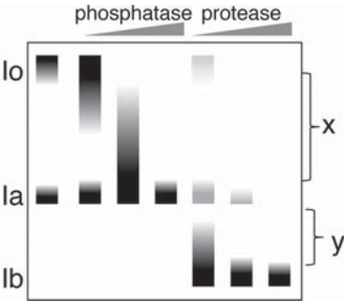


Figure 2. Idealized SDS PAGE separation of the Rpb1 subunit. Increasing phosphatase treatment increases the mobility of the IIo form while proteolysis increases the mobility of both forms. x refers to intermediate phosphorylation states while y refers to intermediates in CTD degradation.

1. INTRODUCTION

The RNA polymerase II (Pol II) C-terminal domain (CTD) is a repetitive disordered domain that extends from the catalytic core of the enzyme. This "tail" domain is heavily modified by phosphorylation, glycosylation, and proline isomerization. In addition to the enzymes that modify the tail a number of RNA processing factors and chromatin modification factors interact with the CTD. Thus, this domain acts as a tether to bring into close proximity the machinery necessary to synthesize and process Pol II transcripts.

By definition the CTD consists of the amino acid sequences extending from the largest subunit, *RPO21* or *RPB1* in yeast and *POLR2A* in human. Specifically the CTD consists of sequences beyond the most C-terminal region that is conserved among the largest subunits of all multisubunit RNA polymerases.[1] This "H" homology region plays a role in positioning the catalytic core and provides surfaces for interactions with Rpb2 and Rpb6.[2] The CTD thus consists of three regions (Figure 1). In order of proximity to the catalytic core these are the linker, the heptad repeats, and a C-terminal nonrepeat or "tip" domain. The linker region is not conserved among different organisms but does contain an enrichment of amino acids found in the CTD.

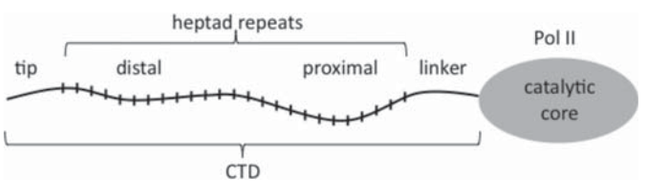


Figure 1. Diagram showing the three regions of the CTD.

The heptad repeat domain consists of multiple tandem repeats of the consensus sequence YSPTSPS. Since this sequences is tandemly repeated the consensus could be any one of seven permutations of this sequence. The selection of Tyr as position one is essentially arbitrary and other permutations have been proposed.[1] Because the Tyr in position one is more universally adopted in the literature we will use this nomenclature in this review.

The discovery of the CTD in the mid-1980s[3] explained the multiple forms of Pol II that were identified by ion exchange chromatography.[4,5] These forms, termed Pol IIO, IIA, and IIB differed by the mobility in SDS gels of the largest subunit termed IIo, IIa, and IIb, respectively (Figure 2). Comparing the amino acid content of these different forms to the coding sequence revealed that the most rapidly migrating form, IIb, is a proteolytic breakdown product lacking the CTD.[3a] Forms IIo and IIa both contain the CTD but differ in that IIo is highly phosphorylated as could be shown by labeling with ^{32}P or by conversion of IIo to IIa by phosphatase treatment.[6] The ability to separate phosphorylated and unphosphorylated CTD species by SDS gel electrophoresis has allowed the characterization of CTD phosphorylation states established both in vitro and in vivo.

The existence of both the IIO and IIA forms of Pol II in vivo led to the discovery that Pol II undergoes reversible phosphorylation with the unphosphorylated IIA form functioning in initiation and the phosphorylated IIO form carrying out the elongation step.[7] Genetic and biochemical studies showed that multiple sites in the CTD heptad repeats could be phosphorylated and these sites were not functionally equivalent.[8] Mapping different CTD phosphorylations to different chromatin sites in coding regions led to the "CTD code" hypothesis in which different phosphorylation states exist in the initiation, elongation and termination phases of the transcription cycle for the purpose of recruiting the appropriate factors to carry out the needed processing reactions.[9]

In this review we will focus on the general properties of the CTD and its modification and interactions with CTD-binding proteins. Gene-specific aspects of CTD function will be discussed in accompanying reviews by Jeronimo et al. and Eick and Geyer. To set the stage for discussing the CTD we will consider the evolution of this domain. We will then discuss structural studies of the CTD. Genetic and gene expression effects of altering the CTD will then be considered. Finally, we will discuss the kinases phosphatases and proline isomerases that establish the code and the CTD-binding proteins that read the code.

2. EVOLUTION OF THE CTD

C-terminal extensions are specific to the largest subunit of RNA polymerase II and related subunits. No similar extension of the largest subunit is seen on Pol I or Pol III nor in any prokaryotic or archaeal largest subunit.[10] This suggests that the CTD emerged as a Pol II specific adaptation after the duplication of subunit genes and the specialization of the three eukaryotic RNA polymerases. In plants, the existence of extensions from the largest subunits of RNA polymerases IV and V is consistent with this view as these plant-specific enzymes are most closely related to Pol II.[11] In this section we will consider the evolutionary origin

dx.doi.org/10.1021/cr400158h | *Chem. Rev.* 2013, 113, 8423−8455

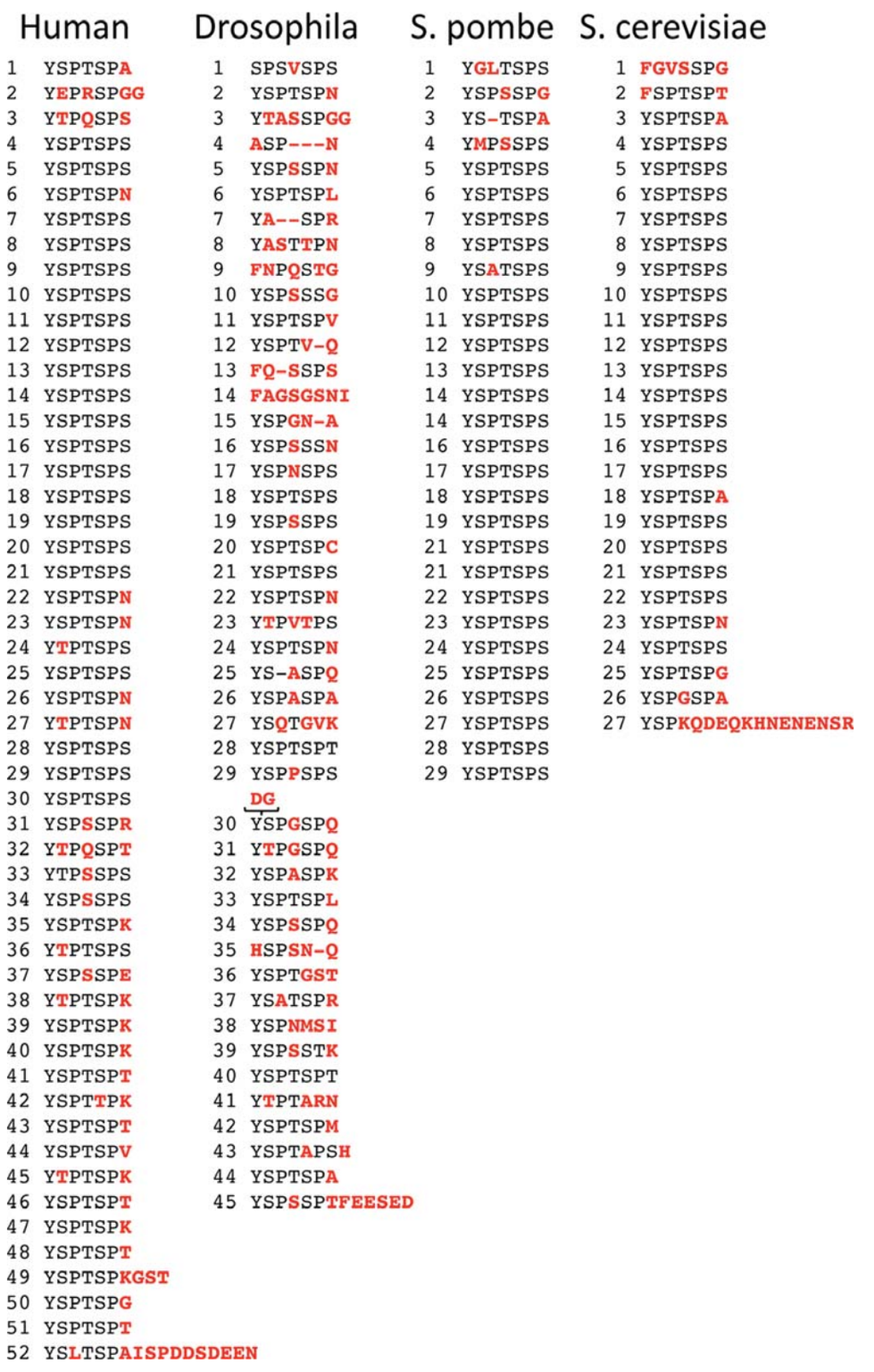


Figure 3. Sequences of CTDs from different organisms. The sequences are aligned to emphasize the heptad repeat. Residues in red deviate from the consensus.

of the Pol II CTD and the conditions that have shaped the evolution of this domain.

2.1. Origin of the CTD

While virtually every Pol II largest subunit (Rpb1) has a sequence extending from the conserved "H" homology region[2] there is a wide difference in the sequence, repetitiveness, consensus, spacing, and length of the CTD. Stiller and Hall derived an evolutionary tree based on the catalytic domains of Rpb1 and used this tree to distinguish between a CTD-clade that consists of organisms in which the CTD is fixed in the YSPTSPS consensus and primordial CTDs like those of *Trypanosoma brucei*, *Giardia lamblia*, and *Trichomonas vaginalis*.[12] In *T. brucei* and *G. lamblia* there are extensions of 291 and 267 amino acids that are rich in Ser and Thr but with no discernible repeat (GenBank

dx.doi.org/10.1021/cr400158h | *Chem. Rev.* 2013, 113, 8423−8455

AAZ13503.1 and EDO76544.1). *Trichomonas* is even further from the CTD-clade, and its extension of 320 amino acids is rich in Leu, Lys, Glu, and Asp (GenBank EAY20967.1). Thus, in the most deeply branched eukaryotes there is little evidence for a repeated sequence. Chapman et al.[1] have proposed that in primordial CTDs the presence of submotifs SPXY and YSPX (where X is any amino acid) coalesced to form a heptad motif SPXYSPX. Amplification of this sequence could have provided selective advantages that enabled the development of more sophisticated gene regulatory mechanisms.

While it seems logical that CTD clade organisms containing the YSPTSPS repeat originated from a common ancestor, there are several organisms that deviate from this consensus suggesting the possibility that repetitive C-terminal repeat domains emerged more than once. *Aspergillus oryzae* has multiple repeats with Phe in position one (**F**SPTSPS)[13] while in *Plasmodium falciparum* the consensus is YSPTSP**K**.[14] *Mastigamoeba invertens* is the most challenging to explain as this brown algae species contains repeats with the consensus YSP**A**SP**A**.[15] It is hard to imagine how two positions in multiple repeats could have been altered so completely through single amino acid substitution starting with the YSPTSPS consensus. One possibility is that changes in one repeat were duplicated and conferred a selective advantage over the YSPTSPS consensus that was lost over time. It remains difficult, however, to rule out the possibility that *Mastigamoeba* and the other nonconsensus repeats evolved independently from a primordial Ser Pro rich C-terminal extension.

2.2. Expansion of the CTD

In organisms that evolved more than one repeat, the stage was set for further expansion. There are two ways that repeated DNA sequences might lead to further duplication. For short repeats (3−4 nt) replication slippage is the predominant cause of insertion or deletion of repeats.[16] For longer sequences like the 21 nt DNA repeat encoding the CTD heptad the mechanism of further duplication is unlikely to be replication slippage but rather out of register recombination between repeat sequences.[17] The results of such unequal crossovers would lead to rapid expansion of the heptad repeat. An example of this process is provided in *S. cerevisiae* where eight consensus repeats are sufficient for viability but result in slow growth and sensitivity to extremes of temperature.[18] Large colonies derived from this strain represent rapidly growing revertants that contain increased number of heptad repeats.[18] Further culturing this strain for about 100 population doublings leads to the appearance of subpopulations with CTDs containing 13−19 repeats and a loss of the 8 repeat parental strain (Creamer and Corden, unpublished). This observation indicates that growth in laboratory culture provides sufficient selective pressure for expansion of the CTD.

Clearly, the reconstruction of the *S. cerevisiae* CTD from identical 21 nt DNA oligos enhances the rate of recombination.[18] Naturally occurring 21 nt sequences have diverged from one another likely stabilizing the CTD from rampant expansion and contraction by recombination. The presence of many Ser residues with 6 codons each has left an evolutionary clue, however, to the expansion process.[1] In humans the distribution of Ser codons indicates that the consensus repeats (Figure 3) proximal to the catalytic core arose by tandem duplications. Thus, recombination among repeat domains can lead to the rapid evolution of longer CTDs.

2.3. CTD Clade

The presence of consensus YSPTSPS repeats in the CTD of animals, plants, and fungi led Stiller and colleagues to propose the CTD clade hypothesis. The crux of their argument is that the CTD provides an essential function that requires tandem repeats of YSPTSPS. Among these organisms there is wide variation in the number of repeats and the presence of nonconsensus amino acids. There is little variation, however in the tandem arrangement of the heptad repeats. Within the CTD-clade the number of repeats varies noticeably from 52 repeats in vertebrates to ∼20 repeats in fungi yielding a rough correlation between the complexity of the organism and the number of repeats.[12,19]

In the *Saccharomyces* family including *cerevisiae*, *bayanus*, *mikatae*, *castellii*, and *kulyveri* there are 27, 25, 26, 24, and 25 repeats, respectively (*Saccharomyces* Genome Database). The positions of the few nonconsensus amino acids also differ in these species. These differences in both number of repeats and the position of nonconsensus repeats argues for considerable instability during the 20 million years since their last common ancestor (www.timetree.org[20]). This instability is further supported by the observation that two commonly used in laboratory strains of *S. cerevisiae* (S288c and A364A) contain 27 and 26 repeats, respectively, and based on DNA sequence comparison, this difference results from two changes; an insertion of two repeats and a deletion of one repeat.[21] Adaptation by CTD repeat expansion is also seen with CTD truncation mutants that result in cold-sensitivity. Selection for spontaneous revertants of this phenotype yielded CTDs which contain more heptad repeats.[22] These observations indicate that the naturally occurring *Saccharomyces cerevisiae* CTD is genetically unstable.

In more complicated multicellular organisms there are generally more repeats and considerable variation in the heptapeptide sequence. In *Drosophila melanogaster* the consensus heptad is YSPTSPS but only 2 of 45 repeats match the consensus (Figure 3). Despite this degree of degeneration the tandem register of repeats is largely maintained indicating that the repetitive nature of the CTD is more important than the actual sequence. Examining different *Drosophila* species (*melanogaster, virilis, mojavensis*, GenBank AAF48057.1, EDW66298.1, and EDW06178.1) indicates that the number of repeats is maintained although these CTDs differ at more than 20 amino acid positions over 40 million years of evolution.[20] Comparing *Drosophila* melanogaster to another dipteran *Aedes agypti* that diverged about 250 million years ago[20] one observes considerable differences in sequence (GenBank: EJY57389.1). In total, the *A. aegypti* CTD amino acid sequence differs at over 100 positions in the CTD. Whereas the *Drosophila* CTD is very degenerate with only two consensus repeats, the *A. aegypti* CTD contains 11 consensus repeats. Despite these differences the *melanogaster* and *aegypti* CTDs contain the same number of repeats.

2.4. Vertebrate CTDs

The CTD of vertebrate animals contains 52 repeats with about half adhering to the consensus and with most of the nonconsensus repeats distal to the catalytic core (Figure 3). Position seven is the most often substituted with Lys, Thr or Asn appearing in multiple repeats. Among mammals, the 52 repeats are identical (including nonconsensus amino acids) in human, marmoset, rat, cow, elephant and opossum and differ by a single amino acid only in mouse (GenBank: NP_000928.1, XP_002724554.1, NP_001193242.1, XP_003416946.1,

dx.doi.org/10.1021/cr400158h | *Chem. Rev.* 2013, 113, 8423−8455

XP_001364837.1, and NP_033115.1, respectively). This striking level of conservation spans about 175 million years of evolution.[20] The number of repeats also appears to have been maintained from zebrafish (*Daneo rerio*, GenBank: XP_682682.1) to humans spanning more than 400 million years of vertebrate evolution. The zebrafish CTD differs from the mammalian CTD at only nine positions, mostly at the least conserved seventh position and most of these changes are conservative and/or create consensus repeats. This striking level of conservation suggests a powerful stabilizing selection among vertebrates.

2.5. Summary of CTD Evolution

The extreme conservation of this domain argues for an essential function in gene regulation. Several possible reasons for conservation are that a specific structure is formed by this sequence. Alternatively the conservation could be due to coevolution of proteins that interact with CTD. These possibilities are discussed in the following sections.

3. CTD STRUCTURE

3.1. Modeling

While predicting the three-dimensional structure of proteins based solely on sequence remains difficult, recent progress has been made in identifying sequences that are likely to be intrinsically disordered.[23] Such disordered regions contain few bulky hydrophobic residues and a high percentage of polar or charged amino acids. The CTD fills this expectation with the possible exception of the high percentage of Tyr. Despite the presence of tyrosine, computer algorithms predict that the CTD is highly disordered.[24]

Although the CTD consensus sequence is predicted to lack order, it has the potential to form secondary structures. Suzuki first proposed that the SPSY and SPTS motifs repeated in the CTD form ß-turns similar to the SPXX motifs found in histones.[25] More elaborate CTD models consisting of helices comprised of ß-turns and proline helices have been proposed, but no evidence supporting extensive formation of such helices in vivo has been obtained.[26]

Another possible CTD structure derives from the recent observation that low complexity (LC) sequences in proteins that contain multiple copies of the motif [G/S]-Y-[G/S] are able to reversibly form amyloid-like fibers.[27] These LC sequences are present in many RNA-binding proteins and McKnight and colleagues have shown that proteins like FUS, RBM3, hnRNP A2, CPEB2, TIA, and hnRNP A1 are able to form both homotypic and heterotypic fibers.[27] The LC sequence motif occurs once in each CTD repeat, and preliminary results from McKnight and colleagues indicate that the CTD can form ameloid-like fibers with other LC sequences (Kwan and Kato et al., submitted). The conformation of proteins in amyloid-lke fibers is a cross-ß structure in which the peptide backbone folds to form a series of ß-sheets that stack upon each other.[28] It is possible that multiple repeats within the CTD are able to form a ß-sheet that can stack with other complementary sheets to form a cross-ß steric zipper which acts as the building block of an amyloid-like fibril. In the case of the CTD this association is reversible, as phosphorylation of the CTD withing these structures releases the CTD (Kwan and Kato et al., submitted).

3.2. Solution Structure

NMR studies of single repeat and multiple repeat peptides have indicated that the CTD is mainly disordered in solution but with a slight propensity to form ß-turns as indicated by a small fraction of folded structures.[26b,29] Addition of the hydrogen bond promoting solvent trifluoroethanol (TFE) increases the population of CTD peptides adopting a ß-turn conformation.[26b] Circular dichroism (CD) of CTD peptides also indicates a predominantly unordered conformation,[26b,29b,30] but careful examination of the CD spectra of CTD peptides in water indicate small but measurable populations of polyproline II helix (P_{II}) and ß-turns.[30] When the CD spectra are measured in TFE the population of ß-turns is greatly increased. Taken together these data argue that the CTD in solution is a dynamic population with fluctuating elements of secondary structure. In the appropriate environment the CTD may contain more or less secondary structure.

In another NMR study focused on a single heptad repeat Dobbins et al.[31] showed that altering the i+2 position (underlined) in the SP<u>X</u>X motif of the heptad to Ala or Gly stabilized turn formation. Thus, the natural occurrence of these nonconsensus amino acids may have a structural role favoring the formation ß-turns. Woody and colleagues further showed using CD that altering Ser2 to Ala in each of eight repeats had little effect on solution structure in water or TFE.[30] In contrast, Ser5 to Ala substitution leads to an increase in P_{II} in water and a loss of ß-turn conformation in TFE.[30] This latter effect of changing Ser5 suggests that some caution must be taken in interpreting the results of mutations that convert Ser to Ala in CTD repeats.

The structural effect of CTD phosphorylation has been addressed using CD. In this case phosphorylation of Ser2 appears to lead to increasing disorder or alteration of the ß-turn conformation.[30] The increased disorder of a phosphorylated CTD is also consistent with an increase in the apparent Stokes radius of the CTD in solution upon phosphorylation.[32] Whether these changes are due to altered backbone conformations or simply a shift in equilibrium between alternative backbone conformations due to charge repulsion is not known.

One limitation to structures that the CTD can adopt is that X-Pro peptide bonds can exist as either cis or trans isomers. In most proteins the trans isomer is favored because there is less steric clash between the Pro amide hydrogen and the preceding Cα atom. NMR spectroscopy has been used to show that for a thirteen residue CTD peptide containing four prolines the trans conformation is the most highly populated at about 70%.[29b] If this is the same for the entire CTD then the mammalian CTD will contain about 30 *cis* Pro bonds at any one time. Since the conversion between cis and trans isomers is slow, this places a limit on the rate of folding of potential CTD structures.[33] Folding can be accelerated by peptidyl prolyl isomerases that will be discussed in a later section.

The length of the CTD peptide also plays a role in its conformational stability. Tyr side chain ordering as determined by CD spectra is different in an eight repeat peptide compared to a two repeat peptides.[30] This suggests that Tyr side chains interact over a distance of more than two repeats. Comparing the eight repeat peptide with the mouse CTD or with synthetic peptides containing ~90 repeats[34] revealed negligible difference suggesting that approximately eight repeats is sufficient to adopt any potential secondary structure.

3.3. Crystallographic Analysis

A second line of evidence for CTD flexibility comes from X-ray and low-resolution electron crystallographic analysis of Pol II. Neither the CTD nor the 80 amino acid linker region are visible in the crystal structure of yeast Pol II although the base of the

dx.doi.org/10.1021/cr400158h | *Chem. Rev.* 2013, 113, 8423−8455

linker is stable and associates with Rpb7 near the RNA exit channel.[2,35] While negative stain is able to fully penetrate the CTD in solution, the 2-dimensional crystallization of Pol II against a lipid layer enabled visualization of a low-density large volume region consistent with a high degree of CTD conformational mobility.

3.4. Potential Dimensions in Vivo

The inherent disorder of the CTD might suggest that this domain occupies a large volume in vivo. This volume may be subject to influences that could have opposing effects on the overall dimensions of the CTD. First, phosphorylation of the CTD is likely to produce a more extended structure as the negatively charged side chains will tend to repel one another. This is seen in vitro in an increased Stokes radius of the phosphorylated CTD[32] and an increase in susceptibility to proteolysis.[36] If fully extended the mammalian CTD could reach over 500 Å, many times the diameter of the catalytic core of Pol II. When proteins are bound to the CTD this could relieve charge repulsion and allow the CTD to form a more compact structure. Indeed, EM images of the CTD bound to Mediator suggest that the CTD adopts a compact structure,[37] although the exact dimensions of this structure were not determined.

4. GENETIC DISSECTION OF CTD FUNCTION

Genetic analysis of the CTD began shortly after its discovery and led to two fundamental insights. First, deletion of the CTD is lethal indicating that it plays an essential role in yeast,[21,38] *Drosophila*[39] and mammalian cells in culture.[40] This essential function is not RNA synthesis as CTD deletions leave the catalytic core intact and biochemical studies of Pol II in which the CTD was removed showed normal catalytic activity.[3c,41] A second important result from deletion studies is that cells are able to grow with less than the natural number of heptad repeats.[21,38,40a] This result indicates that the heptad repeats are functionally redundant.

Before discussing the results of CTD deletions and amino acid substitutions the different genetic systems used to carry out these experiments will be discussed. The bulk of the data have been obtained from the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* and in mammalian cells in tissue culture. In each of these three model systems there are slightly different approaches to testing viability of CTD mutations. In the case of *S. cerevisiae* the effect of CTD mutation can be assessed by plasmid shuffle.[42] Starting with a diploid strain one copy of the *RPB1* gene is deleted and the strain is then transformed with a plasmid containing the WT *RPB1* gene and a selectable *URA3* marker gene. This strain is then sporulated and a haploid strain containing the *RPB1* on a plasmid and harboring a deletion of the endogenous *RPB1* is selected. This strain is then transformed with a plasmid expressing a mutated version of *rpb1* on a plasmid with a *LEU2* marker. Growth in 5-fluoroorotic acid counter-selects against the *URA2* plasmid with the WT *RPB1* thus leaving the cell with only the mutant *rpb1* gene for survival.[18,21] In *S. pombe* there is no plasmid shuffle available and thus mutant CTDs along with a selectable marker gene are recombined into one copy of the *RPB1* gene in a diploid strain. The effect of CTD mutation on viability is then assessed by sporulating the strain and scoring for the presence of the selectable marker linked to the mutant CTD.[43] This same approach was also taken in *S. cerevisiae*.[38] These two approaches, plasmid shuffle and sporulation, while broadly similar may exert slightly different selective pressure. For example, in plasmid shuffle there will usually be multiple copies of the plasmid and this may yield more of the mutant Rpb1 subunit. In sporulation, the colonies must emerge from spores and this may require transcriptional programs not assessed in the plasmid shuffle assay.

To test the role of the CTD in mammalian cells a codominant α-amanitin resistant form of the Rpb1 subunit is expressed from a plasmid. Mutation of the CTD repeats in this α-amanitin resistant gene are assessed by scoring α-amanitin-resistant colonies after transfection[40a] or by coselecting a second marker gene and then assessing the long-term ability to grow in the presence of α-amanitin.[40c,44] Finally, the CTD can be altered by recombination in embryonic stem cells and creation of mice with deletion of CTD repeats.[45] In this approach the effect of mutation is assessed after breeding to obtain mice homozygous for the deletion.

Functional analysis of CTD mutants must take into account sequences in the tip domain located C-terminal to the heptad repeats (Figure 1). This domain in mammals contains a number of acidic residues, is phosphorylated by casein kinase II[46] and binds to Abl1/2 tyrosine kinases.[47] While the sequence of this tip domain is not critical, deletions that remove the tip render the Rpb1 subunit unstable[40c,48] making it difficult to interpret the CTD deletion phenotype. Instability due to deletion of the tip has not been demonstrated in yeast, although deletions that alter the reading frame leading to longer heterologous tip sequences are lethal.[21] Interestingly, the *S. pombe* CTD does not contain a tip domain.

4.1. CTD Deletions in Yeast

Yeast CTD deletion construction has taken several approaches. In the first studies exonuclease digestion was used to remove sequences progressively from the 3′-end of the Rpb1 gene.[21] In this study, maintaining less than 10 of the most proximal repeats rendered cells inviable. CTDs containing the proximal 10 or 11 repeats were viable although grew poorly and CTDs with more than 11 proximal repeats grew normally. Similar results were obtained through a process that removed internal repeats and tested for viability in a sporulation assay.[38] In a separate study the CTD was reconstructed from 21 nt oligonucleotides leaving the most distal and proximal sequences intact.[18] In this case CTDs containing 10 or 11 repeats grew normally while shorter CTDs with 9 or 8 repeats grew slowly and were temperature and cold sensitive.[18]

Similar truncation mutants have been made in *S. pombe*.[43a] This study showed that 16 of 29 proximal heptad repeats are sufficient for normal growth. Deletions that leave 10, 11, 12, or 13 repeats grow increasingly poorly as the number of repeats is reduced. The shortest of these mutations fail to form colonies at high or low temperatures. Reducing the *S. pombe* CTD to the 8 proximal repeats is lethal. Thus, in both *S. cerevisiae* and *S. pombe* about two-thirds of the heptad repeats are dispensable for growth.

4.2. Deletions in Metazoa

Genetic analysis of the murine CTD was made possible by the isolation of an α-amanitin resistance mutation in the mouse *RPB1* gene.[49] This mutation serves as a selectable marker to test the effect of mutations in the CTD.[40a] Growing cells in otherwise lethal concentrations of α-amanitin assessed the function of ectopically expressed amanitin-resistant CTD mutants. Using this approach deletion of the CTD is lethal while partial deletions that contain 29, 31, or 32 of 52 repeats were viable. Removing more than 23 repeats is lethal.[40a,50]

The effect of different CTD deletions in mammalian systems is complicated by the presence of many nonconsensus repeats in the distal part of the CTD (Figure 3). To determine whether the nonconsensus repeats are functionally equivalent to the consensus repeats CTDs containing only consensus or nonconsensus repeats were constructed. These experiments showed that CTDs with only nonconsensus repeats were not viable even if more than the minimum number of consensus repeats were present.[40a,c,44]

Given the difference in function of consensus and nonconsensus repeats it is difficult to assess the minimal number of repeats in mammalian systems as the nonconsensus repeats, while not required for cell growth in culture, may have roles in development. Deletion of 13 repeats containing several nonconsensus repeats in the mouse CTD was tested in this regard by constructing a targeted deletion in the mouse germline.[45] Mice homozygous for this deletion were born at normal rates but exhibited a high rate of neonatal lethality and were significantly smaller than normal littermates. Thus, these nonconsensus repeats likely play a role in normal growth and development.

Truncation of the Drosphila CTD through insertion of a transposon yielded an *RPB1* allele expressing 20 of the proximal repeats and resulting in lethality although this polymerase was still able to support early embryogenesis and transcribe in vitro.[3c,39] To what degree wild-type Rpb1 derived from maternally inherited mRNA contributed to embryonic development was not determined.

4.3. Mutation of the Consensus Heptad

The presence of mainly consensus repeats in budding and fission yeast, together with facile genetics provided an opportunity to examine the role of different amino acids within each heptad repeat. Rather than mutating each repeat individually, new CTDs have been constructed by the concatenation of DNA oligonucleotides such each repeat (or pair of repeats) contains the same substitution. The results of these analyses reveal amino acid substitutions that are allowed or disallowed at each position.

Most of the CTD mutations analyzed to date alter the phosphorylatable residues to nonphosphorylatable Ala or the phosphomimetic Glu. Replacement in *S. cerevisiae* of the phosphorylation sites, Ser2 or Ser5 with either Ala or Glu did not support viability.[18] This result is consistent with essential roles for both the phosphorylated and unphosphorylated form of the CTD. Altering the order of Ser2 and Pro3 also is lethal indicating the need for correct spacing of the heptad Pro residues.[18] Changing the Tyr residue in position one to Phe was also lethal arguing for a possible role for Tyr phosphorylation.[18] Substitution of the *S. cerevisiae* CTD with that of *M. invertans* in which each Thr4 and Ser7 is replaced by Ala supports viability indicating that modification of these residues in not essential for viability.[51] Finally, substitution of Ser7 with Glu is lethal suggesting that at some point this residue must be dephosphorylated.[52] Phenotypes of these mutants are summarized in Figure 4.

More recently, similar mutations have been made in the *S. pombe* CTD.[43b,c] In this organism Ala substitutions at Pro3, Ser5, and Pro6 were found to be lethal, but in contrast to the *S. cerevisiae* CTD, Ser2 to Ala or Tyr1 to Phe substitutions were not lethal (Figure 4). In the case of Ser2 to Ala substitutions in 12 repeats there is a mating defect[43b] and a failure of septation following mitosis.[53] Substituting either Val or Ala at Thr4 was not lethal indicating that phosphorylation of this residues in *S. pombe* is not essential. Substituting Thr for Ser2 or Ser5 yielded different results. In the case of Ser2, Thr substitution is not lethal but a

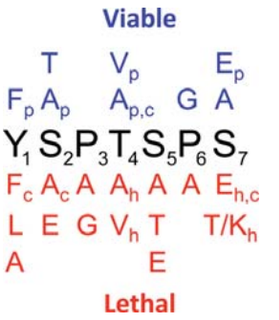


Figure 4. Phenotypes of CTD substitution mutants. Indicated phenotypes are due to substitution in each repeat. Red indicates lethal mutations while blue indicates viable mutations. Some mutations are listed as both viable and inviable depending one the organism as designated in subscripts (c = *S. cerevisiae*, p = *S. pombe*, and h = human).

Ser5 to Thr is lethal indicating that at this position the presence of an extra methyl group interferes with CTD conformation or interaction with modifying enzymes. Finally, Gly substitution at Pro3 is lethal while substitution at Pro6 is not, although these cells grow slowly.[43b]

One difference between the *S. cerevisiae* and *S. pombe* experiments is that the *S. pombe* constructs all contained four degenerate repeats proximal to the catalytic core. It is possible that these repeats supply partial function compensating for the loss of Ser2 or Tyr1. An alternate explanation is that Ser2 is generally a less important residue. Individual deletion or mutation of several Mediator subunit genes allows the growth of *S. cerevisiae* containing Ala in position 2 demonstrating that phosphorylation of this residue is not essential in all genetic backgrounds.[8b]

Mutation of the consensus heptad has not been extensively studied in mammals. Chapman et al. have examined the effect of substituting Ser7 with various amino acids in a reconstructed 48 repeat CTD. Only the substitution with Ala was viable while substitution of each repeat with Glu or Thr/Lys in alternating repeats were not viable. This indicates that phosphorylation of Ser7 may not be essential and is consistent with CTD mutants that replaced the consensus repeats with the distal nonconsensus repeats.[40a,c,44] In separate studies, the CTD of human Rpb1 was replaced with a CTD in which Thr4 was substituted with either Val or Ala. In each case the mutant Rpb1 subunit was not able to support growth.[54] This result is consistent with an essential function for Thr4 perhaps requiring phosphorylation (see later section).

4.4. Spacing of Heptad Repeats

The tandem nature of the CTD suggests that the heptad repeat is the functional unit. However, mutations in which the spacing of repeats is altered has led to the realization that the functional unit comprises more than one repeat. Stiller and colleagues first showed that addition of a single alanine residue in each heptad repeat is lethal in each of the four permutations. In contrast, inserting a single alanine between Ser7-Tyr1 or Tyr1-Ser2 in every other repeat supported viability.[55] Shuman and colleagues extended this result in *S. pombe* showing that insertion of a single alanine in all seven permutations of a diheptad repeat is not lethal.[43c] Taken together, these results provide strong evidence that the functional unit of the CTD lies with two tandem heptad repeats.

Tandem repeats of diheptads in which residues of the conserved consensus repeats are deleted or substituted with

dx.doi.org/10.1021/cr400158h | *Chem. Rev.* 2013, 113, 8423−8455

Ala have further defined the essential unit of CTD function. In *pombe*, substituting positions 5−7 of the distal diheptad $(YSPTSPSYSPTAAA)_7$ had no discernible effect on growth.[43c] Changing one or two more residues to Ala in the distal repeat resulted in impaired growth. Taken together, these results indicate that the Ser5-Pro6 sequence is not required in every repeat. Substitutions of residues in the proximal repeat of a tandem diheptad further refine the functional unit. Replacing the first three residues in *pombe* is lethal as is replacing Tyr1 in every other repeat. However, the Pro3 to Ala change in the proximal repeat was not lethal indicating that Ser2-Pro3 need not be present in each repeat. Taking these results together leads to the minimal functional unit in *pombe* of YSPTSPSYSP.[43c] This result confirmed an earlier result in *S. cerevisiae* in which a tandem repeat of a partial diheptad lacking the last three residues of the distal heptad grows normally. Thus, Stiller and colleagues proposed minimal "252" CTD functional unit consisting of three SP units Ser2-Pro3−Ser5-Pro6−Ser2-Pro3.[56]

The spacing of consecutive CTD functional units is not critical as up to five Ala residues between diheptads is viable although cells grow slowly.[57] Insertion of seven Ala residues between diheptads is lethal but between triheptads is not.[57] The tendency of poly alanine to form α-helices indicates that the helical nature of the insert may alter the conformation of adjacent heptads. This was directly demonstrated by substituting two Pro residues in a seven Ala insert. The helix breaking residues restored viability to this CTD.[57] Helical inserts may alter the functional unit when each heptad is adjacent to a helical domain but in the case of the triheptad the internal repeat would be shielded.

4.5. Proximal vs Distal Consensus Repeats

The *S. cerevisiae* CTD consists almost entirely of consensus repeats that are redundant in their function. It is therefore unexpected that sequence requirements for the distal repeats should differ from those of the proximal repeats. This is the case, however, as Ser2 to Glu substitutions in the proximal repeats are viable while the same substitutions in the distal repeats are lethal.[18] The converse is observed with Ser5 to Glu substitutions; proximal substitutions are lethal while distal substitutions are viable.[18] Together these results indicate that distal and proximal repeats have at least partially nonredundant functions.

4.6. CTD Mutation Summary

Mutations in the CTD of a variety of organisms have shown that the CTD performs an essential function. The genetic malleability of this domain indicates a degree of functional redundancy that is somewhat surprising given the evolutionary conservation, especially in mammals. Insertion of residues between heptad repeats has shown that the functional unit of the CTD extends beyond a single heptad and can be separated by nonconsensus residues. Finally, the effect of substitutions of individual residues in each heptad varies from organism to organism. The wide variety of available CTD mutants has greatly aided the functional dissection of the CTD.

5. FUNCTIONAL ANALYSIS OF CTD MUTANTS

Cells harboring mutant CTDs have been exploited to identify the genetic pathways that are perturbed thereby providing information about CTD functions. Genetic suppression of the conditional growth phenotype of the most severe CTD mutants has been used to identify factors that interact with the CTD. In addition, the effect of CTD mutations on transcription of endogenous or reporter genes has been used to identify the step(s) in transcription that are effected by mutation of the CTD.

5.1. Suppression of CTD Mutations

In *S. cerevisiae* the most severe CTD truncation mutants grow slowly at 30 °C and not at all at low or high temperature.[21] The Young laboratory used the inability of strains with CTD of 10 or 11 repeats to grow at low temperature to isolate spontaneous suppressors[22,58] of this phenotype. This screen yielded nine different SRB (suppressors of RNA polymerase B) genes. Characterization of Srb proteins led eventually to the identification of the mediator complex that contains the Srbs and a number of other transcription regulators. The association between Mediator and the CTD will be discussed in a later section, but the identification of Srbs led to a number of experiments aimed at understanding the role of the CTD in transcription.

Suppressor screens were also used to show that Ser2 and Ser5 have different functions in the CTD.[8b] Substitution of yeast Ser2 with Ala can be suppressed by a number of *srb* mutations, but Ser5 to Ala substitutions cannot.[8b] This implies that the function impaired in CTD truncation is similar to the defect caused by lack of Ser2 and suggests that these Srbs may have a repressive function that counters a positive function provided by Ser2.

5.2. Changes in Transcription and Processing Due to CTD Mutation

The slow growth of CTD truncation mutants suggests an underlying defect in some aspect of Pol II function and examination of transcription in these cells has pointed to the involvement of the CTD at many steps in the transcription cycle. In this section we will discuss data implicating the CTD in various stages in the biogenesis of Pol II transcripts. The historical progress of this endeavor follows the transcription cycle from early work on initiation to the more recent work on Pol II termination.

5.2.1. Activation and Initiation. Yeast cells harboring truncated CTDs show defects in activation of some genes like *GAL10* and *INO1* but not others like *HIS4*.[59] The similar binding of activators to the UASs of the effected genes suggested that the CTD might act through directly contacting the activation domains of transcription factors like Gal4. Deletion of the transcription suppressor *SIN1* suppresses a CTD truncation mutant suggesting that the CTD may also be involved in removing repressors to allow transcription activation.[60] Defects in the ability to respond to activators are also observed in CTD truncation mutants in mammalian systems.[61]

Studies in mammalian transcription extracts gave mixed results concerning the role of the CTD in transcription initiation. Transcription in vitro from the adenovirus-2 major late promoter and the Drosophila HSP70 and actin promoters apparently does not require the CTD.[3c,41,62] In contrast, transcription from the mouse DHFR gene does require the CTD.[63] This latter result is consistent with other studies indicating that monoclonal antibodies against the CTD could inhibit transcription of the adenovirus MPL and DHFR promoters.[64]

One possible role of the CTD in initiation would be to recruit Pol II to the preinitiation complex through interaction with the general transcription factors. In this regard several reports showed interaction between the CTD and the TATA Box binding protein TBP[65] or its larger TFIID complex.[66] While direct interactions with TBP have not been further pursued recent data indicates that the CTD can interact with the Taf15 component of TFIID (Kwan and Kato et al., in preparation).

Interest in interactions between the CTD and GTFs began to subside with the discovery that the Srb proteins form a complex

dx.doi.org/10.1021/cr400158h | *Chem. Rev.* 2013, 113, 8423−8455

that associates with Pol II.[58a] This complex was later named the Mediator after its biochemical function in mediating the activation of Pol II transcription in an activator-dependent reaction Purified Mediator also stimulates the phosphorylation of the CTD by TFIIH kinase.[67] The CTD was shown to interact with the Mediator,[68] but whether this interaction is defective in CTD truncations was not explored. Only recently has the basis for this interaction been determined,[37] and this interaction will be discussed more fully in a later section.

5.2.2. Elongation. The transition from initiation to elongation was shown to correlate with phosphorylation of the CTD.[46,69] Two CTD phosphorylation dependent steps have been identified. First, phosphorylation has been proposed in yeast to release the CTD from essential contacts with the initiation complex, particularly with the Mediator. As will be discussed in a later section, this release is provided by a CTD kinase associated with the general transcription factor TFIIH.

A second step requiring CTD phosphorylation is the transition between the early, nonprocessive elongation complex and the processive elongation complex. In mammalian cells the CTD is required for the transition from promoter proximal pausing to productive elongation.[50,70] In yeast, the positive role of CTD phosphorylation on transcription elongation was first documented by Lee and Greenleaf, who showed that the yeast CTD kinase I (CTDK-I) enhances elongation but not initiation in vitro. Inhibition of CTD phosphorylation has the opposite effect on elongation. Price and colleagues, who discovered P-TEFb, showed that this kinase phosphorylates the CTD in the early elongation complex. Inhibition of P-TEFb or removal of the CTD prohibit the transition from the early elongation complex to processive elongation.[71] The details of P-TEFb phosphorylation are discussed in a later section.

5.2.3. RNA Processing: Capping. The 5′-ends of eukaryotic pre-mRNAs are modified by addition of a nontemplated methylated guanyl cap through a process that takes place in three steps. First, the 5′ γ phosphate is removed by an RNA triphosphatase (RTase). Second, GMP is added to the 5′ β phosphate by a GTP-dependent guanylyltransferase (GTase). Finally, a methyl group is added at position 7 of the 5′ cap guanine by a methyltransferase. Capping is the initial step in processing nascent Pol II transcripts occurring when the nascent transcript has just emerged from the elongating Pol II.[72]

CTD deletion mutants are synthetically lethal with mutations in the capping enzymes suggesting a role for the CTD in this process.[73] The GTase enzyme has been shown to interact directly with the CTD in both yeast and mammalian cells,[73,74] and this interaction requires that the CTD is phosphorylated.[9a,75] The structure of GTase bound to the CTD will be discussed in a later section.

Recruitment of the capping enzymes to the CTD ensures that Pol II transcripts are preferentially modified. Mutations in the CTD that disrupt this interaction are expected to have deleterious effects on capping and this may have further consequences. For example, if CTD mutants like Ser5 to Ala fail to cap normally this would render transcripts unstable and lead to premature termination through the Rat1 pathway.[76] Whether CTD truncation mutations are lethal because of a failure to cap is not clear, however. As few as two heptads are required for recognition of the CTD by the mammalian capping enzyme[77] suggesting that deletions removing only half of the repeats will still be capped normally.

5.2.4. RNA Processing: Splicing. CTD truncation markedly reduces the efficiency of splicing in mammalian cells.[78] Bentley and colleagues used an α-amanitin-resistant Pol II with a CTD truncated to 5 heptad repeats to drive expression of reporter genes containing either SV40 or ß-globin introns.[78a] In this system transcription is reduced in the presence of α-amanitin and the transcripts that are synthesized display a 3−5-fold reduction in the percentage of spliced transcript.[78a] This observation fits nicely with CTD binding studies indicating that the CTD interacts with SR-like proteins and other splicing factors.[78b,79]

Various mechanisms have been proposed to explain the necessity of the CTD for efficient splicing in vivo. One possible role of the CTD is to sequester the upstream exon close to the elongating Pol II so that when the downstream exon is synthesized it will be in close proximity to its partner.[80] The fact that in vitro splicing takes place in the absence of Pol II indicates that the CTD is not required for the actual splicing reactions. However, in vitro splicing is much slower than in vivo splicing and cannot efficiently join exons that are more than a few hundred base pairs apart. In contrast, long introns are efficiently removed in vivo[81] supporting the idea that the CTD tethers the upstream exon to the Pol II elongation complex. Furthermore, the addition of Pol II with a phosphorylated CTD (or the phosphorylated CTD alone) to an in vitro reaction can stimulate splicing[82] although short CTD peptides are inhibitory.[79b]

Several proteins have been proposed as the functional link between the spliceosome and the CTD. In yeast, the U1 snRNP protein Prp40 has been demonstrated to bind the CTD through its WW domain.[83] This observation suggested that the recognition of 5′ splice sites could be facilitated through recruitment of the U1 snRNP. Deletion of the Prp40 WW domains is not lethal, however, and does not results in any observable splicing defect.[84]

More recent work has identified a CTD-dependent splicing activity consisting of a complex containing U2AF65 and Prp19C.[85] U2AF65 binds directly to the CTD phosphorylated on both Ser2 and Ser5.[85a] Transcripts produced from a CTD mutant in which Ser2 is substituted in all repeats with Ala is defective for splicing suggesting that this CTD phosphoisoform, present in the middle and 3′-end of genes is required for interaction with the splicosome.[86]

The CTD has also been shown to play a role in alternative splicing.[87] The inclusion of an alternative exon in the fibronectin gene is inhibited by recruitment of the SRp20 SR protein by the CTD.[88] Transcription by a Pol II lacking most of the CTD results in inclusion of this exon.[88] Although SRp20 coimmunoprecipitates with Pol II[89] the mechanism of SR protein recruitment has not been established. In addition to splicing factor recruitment alternative splicing is also regulated by chromatin modification and Pol II elongation rates, both of which may be regulated by the CTD.[87] A more comprehensive discussion of the role of the CTD in splicing is found in an accompanying review by Eick and Geyer (this issue).

5.2.5. RNA Processing: 3′-End Formation and Termination. CTD truncation mutants are also defective in 3′-end processing and termination.[90] The first evidence connecting the CTD to 3′-end formation came from the same experiments showing the CTD-dependence of splicing.[78a] In these experiments transcripts synthesized by Pol II with a truncated CTD readthrough the 3′ processing signals and fail to cleave and polyadenylate the nascent transcript. In support of a role for the CTD in 3′-end formation a number of cleavage/polyadenylation factors have been shown to interact with the CTD.[78a,91] One of the most important factors in the 3′-end machinery is the protein

dx.doi.org/10.1021/cr400158h | Chem. Rev. 2013, 113, 8423−8455

Pcf11. This protein is part of pre-mRNA cleavage complex II and binds directly to the Ser2 phosphorylated form of the CTD that is prominent at the 3′-end of genes.[92]

The CTD is also required for proper termination of yeast nonpolyadenylated Pol II transcripts through the Nrd1-Nab2-Sen1 pathway.[91e,93] Nrd1 and Nab3 are RNA-binding proteins that act as sensors binding to terminator elements in nascent transcripts.[93,94] Nrd1 binds that Ser5 phosphorylated CTD[95] through it CTD-interacting domain (CID) and this leads to termination of Pol II in a process that directs the 3′ end of the transcript to the nuclear exosome.[96] CTD truncation mutants are defective in this pathway, and many snoRNA and other noncoding RNAs fail to terminate properly.[93,97] In addition, a number of protein coding genes are regulated by attenuation through the Nrd1-Nab3-Sen1 pathway.[98] These genes are overexpressed in CTD truncation mutants. No similar nonpoly-(A) termination pathway has been discovered in metazoa although the mammalian SCAF8/RBM16 gene encodes a protein with similarity to Nrd1,[79b] colocalizes with sites of transcription,[99] and binds Ser2P + Ser5P CTD fusion proteins and peptides.[99,100]

5.2.6. RNA Transport. Several studies have indicated that transport of mRNA from the nucleus to the cytoplasm involves proteins that interact with the CTD. The *S. cerevisiae* proteins Npl3 and Yra1 have been implicated in different stages of mRNA biogenesis; transcription elongation, splicing, mRNA transport, and translation.

Npl3 is similar to mammalian SR proteins containing tandem RRMs and a Ser/Arg-rich domain. Npl3 copurifies with Pol II, interacts with the Pol II CTD, and is recruited to chromatin early in the transcription cycle[101] and mutations of Npl3 that reduce RNA binding lead to a reduced elongation rate.[102] After transcription termination Npl3 remains bound to the mRNA and facilitates its export to the cytoplasm.[103] Once in the cytoplasm Npl3 plays a role in translation through interaction with ribosomal proteins and the mRNA poly(A)-binding protein Pab1.[104] Furthermore, *npl3* mutants display impaired translation suggesting a repressive role for Npl3.[104a] In more recent studies Npl3 has been shown to play a role in translation termination fidelity.[105] Npl3 has also been shown to repress translation through an interaction between its RGG domain and the translation initiation factor eIF4G.[106] The roles of Npl3 thus span the mRNA biogenesis pathway, and this implies that conditions that alter CTD interactions with Npl3 could have consequences for translation in the cytoplasm.

Yra1 is a second yeast transport protein that binds the CTD.[107] This RNA-binding protein is homologous to the mammalian Aly/REF export factor.[108] Yra1 is cotranscriptionally bound to a subset of nascent mRNAs[101a,109] and interacts with the mRNA export factor Mex67 that escorts the mRNP to the nuclear pore.[110] Mutations in *YRA1* lead to nuclear retention of completed transcripts.[111] Greenleaf and colleagues have shown that Yra1 binds to doubly phosphorylated Ser2P + Ser5P CTD present in transcription elongation complexes. What role this binding plays is not clear, but mutations that impair CTD binding but not RNA binding have a slight growth phenotype and a minor transport defect under stress conditions.[107b] Given that Yra1 binds as many as a thousand mRNAs[101a,109] the sum of these minor processing defects could have wide-ranging deleterious effects. Clearly, more work will be needed to establish the role of the CTD in establishing the unique mRNP signatures of pre- and mature mRNAs.

5.2.7. Chromatin Modification. Nucleosomes present a barrier to Pol II elongation and histone modification and nucleosome remodeling play critical roles in transcriptional regulation.[112] While Pol II lacking most of the CTD can transcribe transfected reporter gene shortly after transfection[61a,78a] transcription of endogenous genes is blocked early in the transcription cycle.[40b,50,70] One difference in these classes of genes is that endogenous genes are assembled into chromatin while the transiently transfected DNA is likely not fully assembled into native chromatin. The first indication of a connection between the CTD and chromatin was the observation that histone methyltransferses Set1 and Set2 are recruited to actively transcribed genes.[113]

The Set1 histone methyltransferase is part of the COMPASS complex that methylates histone 3 on lysine 4 (H3K4).[114] This modification requires Ser5P and the PAF complex[113c] and is localized to the 5′ ends of genes.[115] Recruitment of Set1/COMPASS also requires Bur1 which is surprising since this kinase phosphorylates Ser2 (discussed in a later section). One possible explanation is that Bur1 phosphorylates a component of COMPASS to stimulate the H3K4Me3 modification.[116] What role modification by Set1 plays in transcription remains unclear as mutation of SET1 or H3K4 have only minor effects on gene expression.[117] Because Set1 is required for repression of Ty1 and *PHO84* in yeast and both transcripts are regulated by cryptic unstable transcripts, it is possible that H3K4Me plays a role in repression of Pol II transcription by trans-acting RNAs.[97,118]

Set2 methyltransferase methylates histone H3 on K36,[119] and this modification localizes to the middle and 3′-end of genes[115] through interaction of Set2 with the elongating form of Pol II phosphorylated on both Ser2 and Ser5 of the CTD.[113a,b,120] Deletion of the *SET2* gene is not lethal, but acetylated histones accumulate over coding regions leading to the expression from cryptic promoters.[121] This *set2* defect is due to the failure of the histone acetylase complex Rpd3s to deacetylate histone in the wake of elongating Pol II.[121,122] Surprisingly, Rpd3s is not recruited by the H3K36me mark but rather through interaction directly with the Ser2+Ser5 phosphorylated CTD.[123] The H3K36me mark is, however, required for deacetylase activity arguing for an allosteric mechanism of the H3K36me mark on HDAC1.

Spt6 is a factor required for positioning nucleosomes in transcribed regions.[124] Working together with Spt4 and Spt5, Spt6 is also required for transcription elongation.[125] The C-terminal region of Spt6 contains several SH2 domains that facilitate recruitment of Spt6 to the transcription elongation complex.[126] Spt6 in this complex is also involved in H3K36 methylation through its interaction with IWS1/Spn1 that recruits the mammalian Set2 to coding regions.[126a]

Other chromatin proteins have been shown to interact with the CTD including heterochromatin protein 1 (HP1) that acts to recruit FACT to the CTD.[127] The Chromodomain-helicase-DNA-binding protein 8 (CHD8) protein has also been shown to associate with the CTD.[128] Taken together, these results indicate that the CTD plays an important role in modulating the structure of the chromatin template.

5.2.8. DNA Replication and Repair. The first indication of a link between the CTD and DNA replication came from the identification of the replication factor mini chromosome maintenance (MCM) helicase complex in high molecular weight Pol II holoenzyme complexes from *Xenopus* oocytes and HeLa cells.[129] This connection was supported by genetic interactions between a mcm5 ts mutation and a CTD truncation.[130] In

dx.doi.org/10.1021/cr400158h | *Chem. Rev.* 2013, 113, 8423−8455

addition, CTD truncation mutations show increased instability of minichromosomes.[130] More recent experiments have suggested that stalled Pol II elongation complexes recruit components of the origin replication complex (ORC) through interactions with the CTD.[131] This model is consistent with genome-wide mapping studies showing a correlation between Pol II promoters and ORC complexes.[132]

The CTD is also involved in DNA repair. Truncation of the yeast CTD results in a reduced DNA damage response[133] and deletion of the CTD kinase catalytic subunit gene *CTK1* renders cells sensitive to chemical and irradiative damage.[134] Furthermore, a number of proteins that bind to the phosphorylated CTD are involved in the DNA damage response pathway.[134]

Errors in processing of nascent transcripts can lead to R-loop formation and this has been linked to an increased rate of recombination and genome instability.[135] While the Nrd1-Nab3-Sen1 complex that associates with the CTD has been implicated in this process[135e] the detailed mechanisms have not been established. Another factor that could couple transcription and repair is the DNA helicase RecQ5ß. This protein is required for the maintenance of genome integrity[136] and has been shown to interact with the CTD.[137] The presence of DNA repair components in Pol II elongation complexes could enable the enzyme to respond rapidly to encounters with damaged DNA or in the event of collisions with the DNA replication machinery.

6. WRITING THE CTD CODE: CTD KINASES

The strong selective pressures that maintain the CTD through evolution have been proposed to originate through coselection of the proteins that interact with the CTD.[138] Among the most coconserved proteins are the kinases and phosphatases that modify the CTD and the RNA processing factors that bind the CTD and couple transcription and processing have fixed the CTD structure in its present state. In this section we will describe the CTD kinases that modify the CTD. There are a number of different CTD kinases and a number of different possible phosphorylation sites in the CTD heptad. The preferential phosphorylation of different residues in the heptad repeat at different stages of the transcription cycle has been termed the CTD code.[9b,43c,92b,139] While there does not seem to be a strict correlation between the phosphorylation state and the function of the CTD there are definite correlations between different patterns of CTD phosphorylation and the position of Pol II in the transcription cycle. In this section we first discuss the analysis of CTD phosphorylation. We will then discuss the kinases individually and finally address the interplay among these kinases.

6.1. CTD Phosphorylation

6.1.1. Characterization of CTD Phosphorylation Sites.

Consensus CTD heptad peptide substrates phosphorylated by kinases in vitro and subsequently sequenced by Edman degradation established that the human Cdc2 kinase phosphorylates both Ser2 and Ser5.[8a] Mass spec analysis has also been used to identify phosphorylation sites on in vitro phosphorylated model CTD substrates[140] and in vivo phosphorylated CTD as part of a phosphoproteome analysis.[141] These approaches are useful for identifying which of the consensus residues is phosphorylated but identifying which of the many repeats is phosphorylated is limited to the nonconsensus heptads.

The most commonly used method for mapping CTD phosphorylation sites employs antibodies against phosphorylated CTD epitopes. The earliest experiments consisted of polyclonal antibodies raised against a CTD fusion protein phosphorylated in vitro by CTDK-I.[142] This antibody was affinity purified and used to show that Pol II on transcription puffs contained a phosphorylated CTD while Pol II near the promoter contained an unphosphorylated CTD.[143] The problem with this approach is that the epitopes generated by CTDK-I, although unknown at the time, consisted of both Ser2P and Ser5P. Thus, while this approach led to the first in vivo evidence supporting a CTD code,[143] the polyclonal nature of the antibody prep limited its utility in identifying specific phosphorylation sites.

Monoclonal antibodies (mAbs) are able to discriminate between different phosphoepitopes. The first phosphoepitope-specific mAbs were generated against mixtures of proteins and the target of individual mAbs was determined by characterizing the bound protein. In this way two mAbs recognizing different CTD phosphoepitopes were identified.[144] The identity of the phosphoepitope (H5 = Ser2P and H14 = Ser5P) was determined by the ability of each antibody to identify phosphorylated CTD fusion proteins containing either Ser2 or Ser5 to Ala substitutions.[144b] More recently, Eick and colleagues have used CTD antigens consisting of peptides phosphorylated on specific residues to generate mAbs and their specificity has been confirmed using reactivity to a similar bank of phosphorylated CTD peptides and/or CTD fusion proteins containing mutations in the consensus heptad repeat.[44,54b,145] Monoclonal antibodies against various CTD isoforms have standardized the mapping of in vivo phosphorylation patterns although the binding of these antibodies can be blocked by modification of adjacent residues introducing a degree of uncertainty in the resulting maps. These antibodies are described in Table 1.

6.1.2. Chemical Genetic Approach to Kinase Function.

The characterization of CTD phosphorylation patterns in vivo

Table 1. Anti-CTD Monoclonal Antibody Specificities[a]

monoclonal antibody	recognition site(s)	blocked recognition
8WG16	Y1 **S2** P3 T4 S5 P6 S7	Y1 **S*2** P3 T4 S5 P6 S7
H5	Y1 **S*2** P3 T4 S5 P6 S7	**Y*1 S*2** P3 T4 S5 P6 S7
	S*5 P6 S7 Y1 **S*2** P3 T4	
	Y1 **S*2** P3 T4 **S*5** P6 S7	
	S5 P6 **S*7** Y1 **S*2** P3 T4	
H14	T4 **S*5** P6 S7 Y1 **S*2** P3	
3E10	Y1 **S*2** P3 T4 S5 P6 S7	**Y*1 S*2** P3 T4 S5 P6 S7
	T4 **S*5** P6 S7 Y1 **S*2** P3	**S*7** Y*1 **S*2** P3 T4 S5 P6
		Y1 **S*2** P3 T4 **S*5** P6 S7
3E8	T4 **S*5** P6 S7 Y1 S2 P3	Y1 **S*2** P3 T4 **S*5** P6 S7
	T4 **S*5** P6 S7 Y1 **S*2** P3	Y1 S2 P3 **T*4 S*5** P6 S7
	T4 **S*5** P6 **S*7** Y1 S2 P3	
4E12	S5 P6 **S*7** Y1 S2 P3 T4	T4 **S*5** P6 **S*7** Y1 S2 P3
	S5 P6 **S*7** Y1 **S*2** P3 T4	
6D7	Y1 S2 P3 **T*4** S5 P6 S7	Y1 S2 P3 **T*4** S5 P6 S7
	Y*1 S2 P3 **T*4** S5 P6 S7	Y1 **S*2** P3 **T*4** S5 P6 S7
	S2 P3 **T*4** S5 P6 S7 **Y*1**	Y1 S2 P3 **T*4 S*5** P6 S7
	Y1 S2 P3 **T*4** S5 P6 **S*7**	
3D12	S5 P6 S7 **Y*1** S2 P3 T4	S5 P6 S7 **Y*1** S2 P3 **T*4**
	S5 P6 S7 **Y*1 S*2** P3 T4	P6 S7 **Y*1** S2 P3 T4 **S*5**
		T4 S5 P6 **S*7 Y*1** S2 P3
		T4 **S*5** P6 S7 **Y*1** S2 P3
		T*4 S5 P6 S7 **Y*1** S2 P3

[a]The asterisks indicate phosphorylated residues. The data in this table was derived from work published by the Eick laboratory in Munich.

dx.doi.org/10.1021/cr400158h | *Chem. Rev.* 2013, 113, 8423−8455

Table 2. CTD-Associated Cyclin-Dependent Kinases

	human	*S. cerevisiae*	*S. pombe*	AKA	CTD P-sit
CDK7	Cdk7/cyclin H/Mat1	Kin28/Ccl1/Tfb3	Mcs6/Msc2/Pmh1	TFFIIH kinase	**Ser5**, **S**
	CAK = CDK7	CAK = Cak1	CAK= Csk1		
CDK9	Cdk9/cyclin T	Bur1/Bur2	Cdk9/Pch1	P-TEFb	**Ser2**, Ser5
	CAK = CDK7	CAK = Cak1	CAK = Csk1		
CDK12	Cdk12/cyclin K Cdk13/cyclin K	Ctk1/Ctk2/Ctk3	Lsk1/Lsc1	CTDK-I	**Ser2**, Ser5
	CAK = CDK7	CAK = Cak1	CAK = Csk1		
CDK8	Cdk8/cyclin C/MED12/MED13	Ssn3/Ssn8/Srb8/Ssn2	Cdk8/CycC/Med12/Med13	Mediator CDK module (CKM)	Ser2, S
	CAK = none?	CAK = none?	CAK = none?		

has benefitted greatly from a chemical genetic manipulation of CTD kinase subunits.[146] In this approach the ATP binding site of the kinase of interest is genetically altered to allow binding of a bulky ATP analog inhibitor that does not bind to any other kinase. The assumption is that altering the ATP binding pocket does not alter the specificity of the kinase. This has been demonstrated in at least one case.[147] Using cell-permeable inhibitors like 1NM-PP1 or 1NA-PP1 allows selective in vivo inhibition of a single kinase. This selectivity can be assured by demonstrating a lack of effect on the wild-type cell. For in vitro analyses the use of radioactive bulky ATP analogs can be used to identify kinase targets in complex mixtures of potential substrates.[148] These chemical genetic approaches have proven extremely useful in clarifying the roles of different CTD kinases.

6.1.3. Limitations to Characterization of Phosphorylation Sites. Given the positive aspects of the above approaches to mapping phosphorylation sites in the CTD there are certain limitations that must be taken into account in the interpretation of many experiments. For example, there are some problems with antiphospho-CTD antibodies that are not always adequately addressed and thus some caution must be used in the interpretation of CTD phosphorylation patterns. First, the recognition of the epitope may be altered by phosphorylation of adjacent sites. This looks to be the case for H5 that recognizes Ser2P but binds more efficiently if the adjacent Ser5 is also phosphorylated.[149] mAb 3E10 that also recognizes Ser2P, in contrast does not bind if there is a Ser5P in the next heptad.[44] Another problem is that the reactivity on a Western blot may not be linearly dependent on the level of phosphorylation. This is particularly problematic for the pentavalent IgM antibodies H14 and H5. IgM antibodies generally have lower affinity that is compensated for by their increased valency. Thus H14 and H4 may not detect low levels of CTD phosphorylation preferring to bind to the most highly phosphorylated CTDs containing multiple closely spaced epitopes. The valency problem also arises in immunoprecipitation, for example in ChIP experiments. Titrating the amount of antibody can yield different results due to ratio of antibody binding sites and the density of the epitopes.[150]

Another problem with analysis of CTD phosphorylation state using antibodies is in determining whether all modifications occur on the same CTD or whether transcribing Pol IIs are heterogeneous with some CTDs having mainly one phosphoepitope and another Pol II predominantly a different epitope. Immunoprecipitation of Pol II phosphorylated in vivo with the anti-Thr4P mAb 6D7 left substantial Pol IIO in the supernatant.[54b] This Pol IIO contained Ser5P and Ser7P but little Ser2P. Conversely, precipitation of Pol II with a Ser5P-specific mAb left Ser2P and Thr4P containing Pol IIO in the supernatant. These results suggest the presence of three types of Pol IIO containing primarily Ser5P, Ser5P+Ser7P, and Ser2P+Thr4P.[54b] While these classes of Pol IIO may be separable by IP it is far from certain how discrete their phosphorylation patterns are. Some sets of repeats may be differentially modified and this may lead to optimal IP conditions while other repeats in the same CTD may have a different pattern of phosphorylation. In the future it will be important to map the different Pol IIO subforms on the genome to determine their functional role(s).

There are also limitations in the use of analog sensitive kinases. One problem is that the inhibitors may not work at 100% efficiency and thus remaining phosphorylation may not be due to phosphorylation by other kinases but rather by the low level of kinase activity remaining in the presence of inhibitor. Another possibility is that the specificity of the analog sensitive kinase may be different. This has been checked for some kinases on some substrates but has not been thoroughly explored.

Finally, the use of in vitro kinase reactions on peptide or fusion protein substrates may yield specificities that are not evident in vivo. For example, Ctk1 can phosphorylate Ser5 inefficiently on an unphosphorylated substrate[149] but in vivo prefers substrates already phosphorylated on Ser5 by CDK7 or CDK8.

6.2. Cyclin-Dependent Kinases Targeting the CTD

The CTD code is primarily written by members of the cyclin-dependent kinase (CDK) class of enzymes that were first identified as controllers of cell cycle transitions leading to the initiation of DNA synthesis and entry into mitosis.[151] Later work identified transcription-related CDKs as integral components of the Pol II transcription machinery.[152] These two functions in metazoans are carried out by two sets of CTKs. CDK1, CDK2, CDK4, and CDK6 play central roles in cell cycle control while CDK8, CDK9, CDK12, and CDK13 play different roles in transcription through phosphorylation of the CTD.

The CDK catalytic subunits are not active as single polypeptide chains but need to be activated. This takes place through the interaction with regulatory cyclin subunits and through phosphorylation of a Thr residue in the activation segment (T-loop).[153] The cyclin subunit can also provide a degree of substrate specificity. The transcription-related CDKs from human and fission and budding yeast are listed in Table 2 and discussed individually in a following section.

The specificity of cell cycle CDKs was initially established through the use of libraries of potential peptide substrates.[154] Using this approach the preferred motif of CDK1 is S/T-P-X-R/K. With lower specificity CDK1 can recognize and phosphorylate Ser or Thr in the S/T-P diamino acid. Only the minimal CDK recognition motif is present in the CTD at Ser2 and Ser5. For some cell cycle substrates a nearby RXL motif that binds to a hydrophobic patch on the cyclin can enhance phosphorylation,[155] but this motif is not present in the CTD.

Given the caveats concerning the specificity of kinases and the characterization of CTD phosphorylation sites we will now discuss the kinases that have been shown to phosphorylate the CTD. This discussion will focus mainly on the role of these

dx.doi.org/10.1021/cr400158h | *Chem. Rev.* 2013, 113, 8423−8455

kinases in the transcription cycle in human cells and in the yeasts *S. pombe* and *S. cerevisiae*. This restriction is due to the elegant recent work employing analog sensitive kinases that have helped to clarify earlier work. The review by Jeronimo et al. (this issue) provides a more in-depth discussion of the genomic distribution of kinases and the resulting genome-wide CTD phosphorylation pattern in the accompanying review. After covering the CTD kinases individually the interrelationships of these kinases and their CTD targets will be addressed.

6.2.1. CDK7: Structure and Biochemistry. This CTD kinase is comprised of Kin28/Ccl1 in *S. cerevisiae*, Mcs6/Mcs2 in *S. pombe* and CDK7/cyclin H in metazoa. In metazoa CDK7 was initially identified as the CDK-activating kinase CAK that phosphorylates the T-loop in the cell cycle CDKs to render them catalytically active.[151,156] In yeast, this activity is provided by an independent kinase Cak1 that is distantly related to CDKs.[157] Subsequently, Cdk7 was shown to form a heterotrimeric complex with the ring finger protein Mat1 and this complex is part of the general transcription factor TFIIH that phosphorylates the CTD.[158]

CDK7 can apparently execute these dual functions by recognizing different classes of substrate. CDKs are known to prefer Ser-Pro recognition sites and this site is clearly present in the CTD at Ser2 and Ser5. The activation loop sequence in mammalian CDKs is TXXVVTL in which the underlined T is the phospho-acceptor. The first X residue in this sequence in Cdk1, Cdk2, and Cdk6 is not proline. Thus, CDK7 can recognize and phosphorylate a number of nonstandard amino acid sequences.

Two mechanisms have been proposed to explain the diversity of CDK7 phosphorylation sites. First, CDK7 may be promiscuous in recognizing phosphorylation target sequences. This could either be intrinsic to the active site or could depend on conformational arrangement of nonconsensus targets. A second explanation is that CDK7 has more than one way of interacting with substrates but both modes are very specific. Using an analog sensitive version of human CDK7 in in vitro reactions Fisher and colleagues showed that the specificity for CTD required the presence of Mat1 as the dimeric complex of Cdk7/cyclin H did not efficiently phosphorylate the CTD.[159] This is consistent with the observation that Mat1 is required in vivo for stimulation of CTD phosphorylation by CDK7.[160] The difference between CTD-like and T-loop substrates is also seen in the ability of Cdk7 to phosphorylate peptides in solution. CTD peptides are effective substrates while T-loop substrates are not. Presumably additional contacts with the CDK polypeptide are required to achieve T-loop phosphorylation.

In vitro phosphorylation reactions on CTD substrates were initially used to show that human CDK7 prefers to phosphorylate Ser5.[158d,161] Roy et al. first showed that mammalian Cdk7 phosphorylates a peptide with Ala2 substitutions but cannot phosphorylate a similar peptide with Ser5 converted to Ala.[158d] Similar results were obtained with fusion proteins containing multiple repeats in which each Ser2 or Ser5 is converted to Ala.[161c] Peptides with Ser7 converted to Ala are also effective Cdk7 substrates.[161b,d] CDK7 can phosphorylate a peptide already phosphorylated on Ser2 but cannot phosphorylate a similar peptide phosphorylated on Ser5.[161d] The *S. cerevisiae* Kin28/Ccl1/Mat complex also phosphorylates Ser5 in vitro.[162] Taken together, these in vitro results indicate that CDK7 has a strong preference for Ser5.

With the discovery of Ser7 phosphorylation,[44] the hunt began for the kinase that deposits this mark. Surprisingly, CDK7 was shown to be responsible for at least the promoter proximal deposition of this mark.[150,163] The Ser7 consensus diverges markedly from the S/T-P motif common to CDKs raising the question of how this phosphoacceptor site is recognized by the kinase catalytic center. Two possible explanations are that the kinase is promiscuous or that two distinct recognition modes are present in the catalytic center. Evidence for the latter explanation has come from in vivo work of Fisher and colleagues.[159] Whether Ser7 phosphorylation falls into this second category is not known but in vitro the efficiency of this phosphorylation is less than that for Ser5.[150]

Phosphorylation of peptides or recombinant CTDs of different lengths indicate a preference for CTDs longer than eight repeats suggesting that Cdk7 may recognize some type of secondary structure not present in shorter peptides.[161b,c] CTD length is particularly important for phosphorylation of Ser7 as CTDs with fewer than 24 repeats are poorly phosphorylated.[44] Perhaps secondary structure favored in the longer CTD positions the nonstandard Ser7 phosphoacceptor side chain in the Cdk7 active site.

Mammalian CTD fusions containing either the conserved proximal repeats or the distal nonconserved repeats are phosphorylated with similar kinetics. When specific nonconsensus repeat peptides are used as substrates, however, an approximately 4-fold preference is observed for peptides with Lys7 in place of Ser7.[161c] This substitution occurs in nine of the distal mammalian repeats. No similar preference was observed for Asn7 or Thr7 peptides. The basic nature of Lys located +2 residues from the phosphoacceptor is similar to the consensus of the cell cycle CDKs suggesting that a kinase like CDK1 may have an unappreciated role in CTD phosphorylation.

6.2.2. CDK7: Function and Genetics. CDK7 is the first kinase to phosphorylate the CTD within the PIC.[158] This TFIIH kinase activity is stimulated by the Mediator complex that binds to the unphosphorylated form of the CTD and helps recruit Pol II to the preinitiation complex (PIC).[67] Phosphorylation of Ser5 and Ser7 residues in the CTD[150,163] disrupts this interaction[162a,164] permitting promoter escape and entry into the elongation phase of the transcription cycle. The early stages of elongation coincide with assembly of the elongation complex that consists of factors promoting processive elongation, chromatin modification, and RNA processing.[165] This stage of the transcription cycle differs in yeasts and metazoa. In yeast, the Pol II elongation complex is rapidly assembled and the transition to elongation occurs without an apparent delay. In metazoa there is a delay in this transition leading to a paused Pol II early elongation complex about 30 nt downstream of the transcription start site.[166] CTD kinases play an important role in these early elongation complexes, and thus, although these kinases are similar in yeast and metazoa their functions are subtly different.

Despite its proposed central role in releasing Pol II from the PIC and the fact that essential genes encode CDK7, transcription of some genes occurs in the absence of CDK7 activity. This was first observed in vitro where neither basal nor activated transcription was reduced in the absence of TFIIH kinase activity.[62c,167] Later in vivo experiments supported this observation. Inhibition of analog-sensitive Kin28 has only a minimal effect on global Pol II transcription.[168] In the first case, Kanin et al showed no decline in global transcription of Pol II genes.[168b] In the second case Hong et al normalized their data differently to show that steady state levels of 58% of Pol II transcripts were sensitive to Kin28 inhibition.[168a] Despite this decrease in steady state RNA the occupancy of Pol II on chromatin did not appreciably decline. Phosphorylation of Ser5

dx.doi.org/10.1021/cr400158h | *Chem. Rev.* 2013, 113, 8423−8455

did decrease while that of Ser2 remained nearly constant. Taking these results together, a picture emerges that Kin28 phosphorylation provides Ser5P for the purpose of recruiting the capping enzymes. In the absence of this phosphorylation capping is reduced and, although transcription continues the uncapped transcripts are unstable. CDK8/Srb10 and/or Cdc2/Cdc28 likely provide some residual Ser5 phosphorylation.

Global transcription seems to be independent of CDK7 in *S. pombe* and humans cells. Microarray analysis of transcription in a *mcs6* temperature sensitive mutant grown at the nonpermissive temperature for a short period demonstrated only a minor effect on global transcription.[169] Similarly, there is little effect on global transcription in a human cell line expressing an analog sensitive Cdk7 gene.[163c] In this latter case the accumulation of Pol II near the 5′ end of genes is reduced suggesting that promoter proximal pausing is dependent on CDK7.[163c] This effect on pausing is most likely not due to changes in CTD phosphorylation but rather to changes in phosphorylation of TFIIE and/or DSIF.[170]

Although Cdk7 is essential for organismal viability there has been some difficulty in discriminating between its physiological roles in transcription and cell cycle control. In *C. elegans* loss of Cdk7 effects both transcription and mitosis.[171] In *Drosophila*, adult animals homozygous for a *cdk7*^{ts} mutation are viable for over 40 days at the nonpermissive temperature arguing against an essential role in transcription. A role in cell cycle control is suggested by a reduction in the number of rapidly dividing follicle cells in females. Mice homozygous for a Cdk7 deletion die early in embryogenesis[172] but demonstrate little effect on Pol II transcription. The major effect in these cells is in cessation of cell division brought on by the lack of Cak activity. Tissue specific knockouts of Cdk7 in adults cause the loss of proliferating cells but little change in nonproliferating cells. This result argues that TFIIH kinase is not required for transcription in postmitotic cells. Furthermore, inhibition of Rb family proteins with ectopic expression of the SV40 T antigen restored growth to cdk7 null embryonic fibroblasts.[172] Taken together these results indicate that in some cells, Cdk7 is not essential for transcription. The likely explanation is that other CDKs can provide the missing CTD kinase activity.

6.2.3. Cdk8. CDK8 is the catalytic subunit of the CDK8 kinase module (CKM) that associates dynamically with the Mediator complex.[173] The CKM consists of Cdk8, cyclin C (CycC), Med12 and Med13 in human and Srb8/9/10/11 in *S. cerevisiae*.[58c,174] Med12 is required for CTD kinase activity but MED13 is not.[175] About one-third of the CDK module is not associated with the Mediator complex suggesting that some CDK8 functions may be attributable to activities that occur apart from the transcription initiation or elongation complexes.[173a,174f,175]

In vitro phosphorylation of CTD substrates indicates that CDK8 phosphorylates both Ser2 and Ser5 with a slight preference for Ser5.[58c,161c,d,162a] Unlike CDK7, CKM displays no preference for heptad repeats that contain Lys in position seven.[161c] Consistent with this finding CDK8 displays a preference for phosphorylating the more conserved mammalian proximal repeats.[161d]

Early work on the kinase catalytic subunit Srb10 in *S. cerevisiae* suggested a repressive function for this kinase in phosphorylating the CTD before PIC formation thereby preventing a stable PIC to form.[162a] This was supported by transcription profiling of an *srb10* deletion strain that displays up regulation of over 150 genes.[176] Later work employing analog sensitive kinases showed that Srb10 has a positive role in transcription both in vivo and in vitro.[177] Inhibition of *srb10-as* reduced transcription but not to the extent as inhibition of Cdk7. Inhibiting both kinases together yielded even more inhibition suggesting that both of these kinases have similar, partially overlapping positive roles in transcription.[177] These roles are clearly not fully redundant as deletion of Cdk7 is lethal while deletion of Srb10 is not.

6.2.4. Cdk9: Structure and Biochemistry. Cdk9 and its regulatory cyclin subunit form the positive transcription elongation factor b (P-TEFb).[71,178] The closest relative to Cdk9 in *S. cerevisiae* is the Bur1/Bur2 kinase complex.[179] In *S. pombe* the equivalent kinase is comprised of a Cdk9/Pch1 heterodimer that copurifies with components of the capping complex.[53,180] In both yeasts this Cdk9 heterodimer is essential. In yeasts there is a second gene encoding a nonessential Cdk9-related kinase (Ctk1 in *S. cerevisiae* and Lsk1 in *S. pombe*, Table 2) and two additional kinases Cdk12 and Cdk13 in metazoa. These kinases will be discussed in a later section.

Mammalian CDK9 was initially cloned as a cdc2-related kinase and named PITALRE after a unique sequence motif that occurs in this kinase.[181] There are two isoforms of this subunit, Cdk9^{p42} and Cdk9^{p55} that differ in size due to alternative transcription start sites.[182] These catalytic subunits form a complex with one of four cyclin subunits T, T2a or T2b creating a diverse set of heterodimers that add a degree of functional complexity beyond the scope of this review

Whereas Ser2 is thought to be the main in vivo target of P-TEFb in vitro biochemical analysis indicates that this kinase phosphorylates primarily Ser5. CTD peptide substrates prephosphorylated at Ser5 are not phosphorylated by CDK9 while substrates phosphorylated at Ser2 are.[161d] CTD peptides with Ala substitutions at Ser2 are phosphorylated while those substituted as Ser5 are not. Furthermore, Western blot analysis with mAbs against different phosphorylated forms of the CTD generated by P-TEFb revealed primarily Ser5P with a small amount of Ser7P and virtually no Ser2P.[140a] Interestingly, Ser7P substrates are phosphorylated more rapidly than unphosphorylated substrates indicating that Ser7P may prime phosphorylation by P-TEFb.[140a,183]

The *S. cerevisiae* Bur1/Bur2 kinase phosphorylates a GST-CTD fusion protein on mainly on Ser5, although phosphorylation of Ser7 and Ser2 has also been observed.[162b,179b,184] Ser2 is only slightly phosphorylated when probed with mAb 3E10 that recognizes Ser2P but when the same substrate is probed with mAb H5 that recognizes Ser2P along with an adjacent Ser5P there is an increase in reactivity. These observations suggest that Cdk9/Bur1 kinases phosphorylate Ser2 on CTD substrates that are already phosphorylated on Ser5 and Ser7 CDK7.[185] Once Ser5 is phosphorylated Ser2 may become a preferred substrate. This could also be a priming reaction at the substrate level or could be a change in substrate specificity brought about by phosphorylation of the CDK9 T-loop by CDK7.[170] In this connection it should be noted that the in vitro assays often employ immunoprecipitated or tagged kinases. These kinases may not be phosphorylated on the T-loop as this phosphorylation takes place on chromatin.[170]

6.2.5. Cdk9: Function and Genetics. In metazoans Pol II pauses about 30 nt downstream of many promoters.[166] This was first observed at *Drosophila* heat shock genes by Lis and colleagues.[186] More recent genome-wide experiments have shown that many genes in eukaryotes display a peak of Pol II occupancy just downstream of the start site.[187] Work done by the Price and Handa laboratories led to the identification of proteins that establish this paused elongation complex.[178g,188] The DRB-

dx.doi.org/10.1021/cr400158h | *Chem. Rev.* 2013, 113, 8423−8455

sensitivity-inducing-factor (DSIF) consisting of Spt4 and Spt5 along with the negative-elongation-factor (NELF) and the newly discovered protein GDOWN1[189] are involved in blocking the early elongation complex. P-TEFb is recruited to this complex in part through interaction with transcription factors or the bromodomain protein Brd4.[187d,190]

The transition from the paused complex to productive elongation requires P-TEFb kinase activity.[178a] Among the P-TEFb targets are Spt5 which is converted from an elongation inhibitor to a processivity factor and the CTD that is further phosphorylated to create binding sites for elongation, processing and chromatin modification factors.[178g] Inhibition of P-TEFb with flavopiridol blocks this transition but does not inhibit Pol II that has already entered the productive elongation phase.[187d,189,191] This observation argues that P-TEFb is required only near the promoter.

The activity of P-TEFb within the paused Pol II elongation complex is stimulated in several ways. First, previous phosphorylation of the CTD on Ser7 primes the kinase for phosphorylation of Ser2.[140a] In addition, the Cdk9 activation loop is phosphorylated either by Cdk7[170] or Brd4 kinases.[192] The activation of P-TEFb is likely to be an important regulatory step.[166]

Recruitment of CDK9 in *S. pombe* and Bur1/Bur2 in *S. cerevisiae* takes place through a different mechanism. NELF and GDOWN1 are not present in yeast[193] an observation that is consistent with a lack of promoter proximal pausing.[194] Despite the lack of promoter proximal pausing there is evidence for a sequential action of CDKs at yeast promoters. Recruitment of Cdk9 or Bur1 requires the activity of Cdk7.[180b,185] In the case of Bur1 there is a domain that interacts with Ser5P sites produced by Cdk7.[185] In *S. pombe* previous phosphorylation of Ser7 primes Cdk9 activity.[180b] Cdk9 and Bur1 thus provide the initial Ser2 phosphorylation on the early elongation complex. This process may also involve the atypical Brd4 kinase in mammalian cells (see later section). The phosphorylation of Ser2 on downstream elongation complexes involves another family of CDKs discussed in the next section.

6.2.6. Cdk12. This family of nonessential CTD kinases was initially discovered in *S. cerevisiae* as CTDK-I, the first CTD kinase identified.[142,188d,195] For many years it was thought that the functions of P-TEFb were split between Bur1 and Ctk1 in *S. cerevisiae*.[188d,196] With the discovery that metazoan Cdk12 is a CTD kinase this has clarified the situation and the Cdk12-related kinases are now recognized as distinct from Cdk9 providing a CTD kinase activity that functions later in the transcription cycle.[188d] The *S. pombe* CTDK-I kinase is comprised of Lsk1, the catalytic subunit and Lsc1 the cyclin subunit and Lsg1 corresponds to Ctk3.[197] The human CDK12 and CDK13 genes are closely related by sequence and are the closest relative of the *S. cerevisae* Ctk1 and *S. pombe* Lsk1 catalytic subunits. Recent work from the Greenleaf lab has shown that CDK12/13 are CTD kinases both in vitro and in vivo.[198]

In yeast, the CDK12 homologue CTDK-I is comprised of the catalytic subunit Ctk1, the cyclin subunit Ctk2 and Ctk3, a subunit of unknown function.[188d] CTDK-I is the major Ser2 kinase in *S. cerevisiae*. Bur1 and Ctk1 make approximately equal contributions to Ser2P near promoters but deletion of Ctk1 reduces SerP in the downstream region by 90%.[185] Similar inhibition of Bur1 reduces Ser2P by 50%, much greater than the expected 10% remaining after Ctk1 inhibition.[185] This result suggests that Bur1 phosphorylation enhances Ctk1 phosphorylation and loss of Bur1 activity also leads to a reduction in Ctk1 activity.

The role of these kinases in depositing the Ser2P mark is apparently to encourage the recruitment of factors that bind Ser2P or the dually phosphorylated Ser5P-Ser2P heptad repeats. This includes splicing and termination factors as well as factors that alter chromatin. These factors are discussed in a later section.

6.2.7. Cdk1. The cyclin-dependent kinase I (Cdk1) is also known as Cdc2 in metazoa and *S. pombe* or Cdc28 in *S. cerevisiae*. The catalytic subunit assembles with cyclin B to form the active CDK. The primary function of this kinase complex is to regulate cell cycle[151] but several recent reports suggest a role in transcription.[199]

The first evidence that Cdk1 is a CTD kinase came several decades ago when it was identified as the first metazoan CTD kinase.[200] With the identification of other, transcription-specific CDKs it was assumed that Cdk1 phosphorylation of the CTD was an artifact of promiscuous in vitro activity. Recent results, however, have suggested that the yeast Cdc28 kinase plays an important role in CTD phosphorylation and transcription. Enserink and colleagues have used ChIP assays to show that Cdc28 is present on some but not all genes.[199a] Using an analog sensitive Cdc28 they show that inhibition of this kinase results in lower Pol II density on these genes and show that the overall level of Ser5P is reduced. Their work further shows that Cdc28 acts together with Kin28 to provide Ser5P. Inhibition of either kinase produces a modest decrease in Ser5P while inhibition of both kinases causes a major reduction. These experiments suggest that Cdc28 may play a similar role to Kin28, providing a boost in this function as cells enter the cell cycle.

6.3. Other CTD Kinases

6.3.1. MAP Kinases. Mitogen-activated protein kinases (MAPKs) are Ser/Thr kinases most closely related to the CDKs.[201] The MAPKs are activated by mitogens, osmotic stress, and heat shock.[202] Bensaude and colleagues showed that serum stimulation of resting fibroblasts causes a marked increase in the highly phosphorylated IIO form of Rpb1.[203] The enhanced CTD kinase activity in these cells copurified with MAP kinases and purified MAP kinases were shown to phosphorylate the CTD in vitro.[203] The MAPKs ERK1 and ERK2 were later shown to specifically phosphorylate Ser5 in CTD peptides.[161b]

6.3.2. DNA-PK. DNA-dependent protein kinase (DNA-PK) is a DNA-activated protein kinase that has been implicated in DNA repair processes like homologous recombination and nonhomologous end joining.[204] Although initial studies indicated that DNA-PK preferentially targets S/T-Q or Q-S/T phosphoacceptor sites several studies have shown that DNA-PK phosphorylates CTD fusion proteins or peptides.[161b,205] Surprisingly, DNA-PK can phosphorylate all three Ser residues in the CTD.[206] DNA-PK was identified as part of the Pol II holoenzyme[207] and has recently been implicated in transcription of the HIV genome[206] and in transcription-coupled repair.[208] Phosphorylation of the CTD by DNA-PK depends not only on DNA but is also stimulated by transcription activators.[209]

6.3.3. BRD4. BRD4 is a bromodomain protein that has been implicated in a variety of human diseases.[210] Recently Singer and colleagues have shown that BRD4 is an atypical protein kinase that phosphorylates the CTD on Ser2.[211] Whereas P-TEFb phosphorylation of the CTD generates epitopes recognized by both mAbs H5 and 3E10, BRD4 generates only the Ser2P-specific 3E10 epitope. This difference is possibly due to the ability of P-TEFb to phosphorylate both Ser2 and Ser5 whereas

dx.doi.org/10.1021/cr400158h | *Chem. Rev.* 2013, 113, 8423−8455

BRD4 is apparently restricted to Ser2. The BRD4 kinase activity differs from P-TEFb in that it is insensitive to the inhibitor flavopiridol but is inhibited by apigenin.[211] Inhibition of P-TEFb with flavopiridol in an in vitro transcription reaction did not substantially reduce Ser2 phosphorylation while apigenin reduced SerP levels by 90%.[211] Overexpression of BRD4 in vivo leads to a marked increase in Ser2P while drugs that block the acetylated histone binding of the BRD4 bromodomain reduce the in vivo level of Ser2P. These results indicate that BRD4 is likely to play an important role in the transcription cycle potentially acting at the transition between initiation and elongation, before the recruitment of pTEFb.[210]

6.3.4. Plk3. Polo-like kinase 3 is a member of the polo Ser/Thr kinase family.[212] Plk3 is unique to mammalian cells and is widely expressed and localized to the nucleus.[212] Like other members of this family Plk3 contains a polo box domain (PBD) that binds phosphoserine/threonine containing motif.[213] Plks phosphorylate Ser or Thr embedded in acidic sequences, preferring acidic residues in all positions between +4 and −4 from the phosphoacceptor.[214] There are no amino acid sequences in the CTD that correspond to this motif but phosphorylation of multiple serines could enhance the phosphorylation of adjacent sites by Plk3.

Eick and colleagues identified an antibody (6D7) that recognizes Thr4P in the CTD and have shown that in human cells Plk3 deposits this mark.[54b,215] In vitro, phosphorylation occurs on Thr4 and can convert Pol IIA to Pol IIO. Thr4P is found in vivo on a subset of Pol IIO containing Ser2P but lacking Ser5P and Ser7P.[54b] Plk3 is activated by oxidative stress[216] and treating HeLa cells with H_2O_2 results in a marked increase in Thr4P in vivo.[54b] Ser2P peaks before Thr4P along a gene indicating that Ser2P is a prerequisite for Thr4P.[54b] This is consistent with the known preference for acidic phosphoacceptor sites. Phosphorylation of Thr4 has also been observed in budding[54b] and fission[217] yeast but the kinase(s) responsible for this modification have not been identified.

6.3.5. Tyrosine Kinases. While phosphoserine and phosphothreonine are the most abundant phosphoamino acids obtained from Rpb1 immunoprecipitated from in vivo extracts small amounts of phosphotyrosine are also obtained.[47b] In vitro phosphorylation of a GST-CTD fusion protein by c-Abl or the Arg kinase suggested that the origin of the phosphotyrosine was Tyr1 of the CTD.[47]

Recently Eick and colleagues have described a monoclonal antibody that is specific for Tyr1P in the CTD although reactivity is blocked by phosphorylation of adjacent Ser5 residues.[215] This antibody detects tyrosine phosphorylation both in metazoa and yeast. ChIP using this antibody has been used to show that Tyr1P CTD is widely distributed throughout the genome being low at promoters, increasing toward the middle of genes and reduced in abundance at the polyadenylation site.[145] The abundance of Ser5P near the 5′-end may lead to an underestimation of promoter-proximal Y1P. Tyr1P-containing peptides do not bind to the CTD interacting domains (CIDs) of several termination factors. Taken together these results suggest a role for Tyr1P in regulating access of termination factors to the CTD during elongation.[145] Kinase(s) that deposit this mark in yeast have not yet been identified.

6.3.6. Casein Kinase II. Casein kinase II (CK2) is a Ser/Thr kinase that phosphorylates acidic target sites with the consensus S/T-X-X-D/E.[218] This sequence occurs several times in the mammalian CTD and has been reported to be phosphorylated by CK2.[152] Although this kinase cannot phosphorylate the consensus repeats,[152] it is possible that the presence of phosphates deposited by other kinases can prime phosphorylation by CK2.

6.4. Summary of CTD Kinases

The broad outline of the sequential phosphorylation of CTD targets is now becoming clear, and we can see the similarities between the sequential chain of protein kinases that regulate the cell cycle and those that regulate the transcription cycle.[219] . Emerging evidence indicates that many of the CTD kinases interact in ways that enhance or inhibit CTD phosphorylation. These interactions fall into three main categories. First, the kinases can phosphorylate each other to either inhibit of activate CTD phosphorylation. Second, phosphorylation of some CTD sites may prime phosphorylation by other kinases. Third, phosphorylation may facilitate recruitment of the kinase to the transcription complex. All of these possible interactions have been shown to occur and the remaining challenge is to determine how the interplay between kinases functions on individual genes.

While the CTD code offers a solid foundation for understanding the regulation of the transcription cycle and its coupling with processing of the nascent transcript, we still lack details about many steps in this process One problem is that the tools used to characterize the CTD phosphorylation pattern are not fully understood and are not able to distinguish between heptads within a CTD. In addition, the genetic tools used to alter the function of different CTD kinases have limitations. For example, there are likely to be many other substrates of the CTD kinases and some of these may play critical roles in regulating transcription. Future studies will need to address the spectrum of substrates of the CTD kinases.

6.5. CTD Glycosylation

The CTD in mammalian cells is postsynthetically modified by the addition of O-linked *N*-acetylglucosamine by the enzyme O-GlcNAc-transferase (OGT).[220] Removal of O-GlcNAc is catalyzed by an enzyme named *N*-acetyl-β-D-glucosaminidase (OGA).[221] For almost two decades the significance of CTD glycosylation remained elusive but the recent development of inhibitors of OGT[222] and OGA[223] has allowed the characterization of CTD glycosylation demonstrating that this modification plays a significant role in regulation of transcription in higher eukaryotes.

Early experiments showed that the unphosphorylated IIa form of Rpb1 is O-GlcNAcylated at Thr4 and Ser5.[220a] The stoichiometry of the modification was very low, however, leaving open the question of function. More recent work has used an OGA inhibitor to stabilize the O-GlcNAcylated CTD.[224] These studies showed the presence of a highly modified form of the unphosphorylated CTD that was termed Pol IIγ.[224] The retarded mobility suggests that this form is multiply GlcNAcalated but the stoichiometry was not determined. In vitro modification of GST-CTD fusion proteins containing Ser5 or Ser7 to Ala substitution are poor substrates for OGT. Taken together with the earlier identification of Ser5 and Thr4 as the targets site for modification this indicates the Ser5 is the most likely site of CTD GlcNAcalation.

OGT and OGA inhibitors block transcription indicating that O-GlcNAc modification is an essential step in transcription. Since both the addition and removal of the modification are essential this has led to model in which O-GlcNAc cycles on and off the CTD in the PIC.[224] What role this "γ-cycle" may play is unclear and the possibility that the prevalence of hexosaminidases removes the GlcNAc from the CTD rapidly in vitro could

dx.doi.org/10.1021/cr400158h | *Chem. Rev.* 2013, 113, 8423−8455

mean that the unmodified IIa form of Rpb1 is a rare intermediate. In this view Pol II would cycle in vivo between the IIO and IIγ forms. The rationale for modifying phosphorylation sites with O-GlcNAc may be to protect these sites from inadvertent phosphorylation.

6.6. Other CTD Modifications

Several other CTD modifications have been observed in mammalian cells. These modifications generally target non-consensus amino acids like the methylation of the arginine residue in position seven of the 31st repeat. This modification is carried out by the coactivator-associated arginine methyl transferase I (CARM1). The methylated CTD acts to inhibit expression of snRNAs and snoRNAs.[225] Methylation of this CTD arginine is suppressed when the CTD is phosphorylated on nearby Ser2 and Ser5 residues.[225]

7. CTD CODE ERASERS: CTD PHOSPHATASES

CTD phosphorylation is a dynamic process in vivo with different sites phosphorylated and dephosphorylated during the transcription cycle to provide binding sites for proteins involved in processing the nascent transcript. In addition, dephosphorylation is required to recycle Pol II after transcription termination providing an unphosphorylated Pol IIA for assembly of new PICs. In this section we will discuss the phosphatases that act to counter the CTD kinases and thus shape the changing CTD phosphorylation pattern.

7.1. Fcp1 CTD Phosphatase Family

The Fcp1 CTD phosphatase was first identified biochemically as an enzyme activity that converts the phosphorylated IIO form of Pol II to the unphosphorylated IIA form.[226] This phosphatase is essential in yeast[227] and highly conserved in evolution. Fcp1 is a multidomain protein with an N-terminal catalytic domain and a breast cancer protein related C-terminal (BRCT) domain.[228] The catalytic domain contains a DXDX(T/V) motif commonly found in phosphotransferases and hydrolases.[229] In addition to Fcp1 animal cells contain a related family of small CTD phosphatases (SCPs) encoded by separate genes and consisting of an Fcp1-related catalytic domain but lacking the BRCT domain.[230] The human SCPs act to silence neuronal genes in non-neuronal tissues.[231]

The first structure of an FCP1-like phosphatase to be solved was human SCP1.[232] These studies showed that the active site is located in a depression on the surface that normally contains a Mg^{2+} ion. The first Asp in the conserved motif DXDX(T/V) acts as a phosphoacceptor for the phosphatase reaction and mutation of this residue to Ala in the yeast Fcp1 protein results in an inactive enzyme.[228a,232a,233] There is a fundamental difference in the specificities of Scp1 and Fcp1. The small phosphatase Scp1 recognizes Ser5P about 60-fold better than Ser2P.[232b] In contrast, while Fcp1 will dephosphorylate both Ser2P and Ser5P it displays a preference for Ser2P of about 6-fold.[234] In both cases the preferred substrate contains an SP sequence in the trans conformation.[235] The difference in specificity between Scp1 and Fcp1 lies in the particular arrangement of amino acids that lie in the active site groove and by the presence of additional domains in the Fcp1 enzyme.[234a] Scp1 makes specific contacts with Tyr1, Pro3 and Thr4 thus positioning Ser5P in the catalytic center. Fcp1 cannot make the same contacts as many of the contact residues are buried in the Fcp1 structure. Structure–activity experiments show that mutations in the Fcp1 CTD-binding pocket alter specificity for either Ser2P or Ser5P. Most importantly, mutations that reduce Ser5P specificity have no effect on the growth of *S. cerevisiae* strongly supporting the idea that Fcp1 is primarily a Ser2P-specific phosphatase.[236]

Fcp1 interacts with the catalytic core of Pol II,[237] and this docking site includes the Rpb4/7 subunits[227b,232a] located adjacent to the CTD linker domain.[35,238] In addition, Fcp1 interacts with TFIIF and this interaction stimulates the phosphatase activity.[227b,239] These interactions are not essential for phosphatase activity however, as Fcp1 can dephosphorylate CTD peptides[233] and Pol II that is not engaged in transcription.[240]

The CTD is dynamically phosphorylated in vivo and Fcp1 plays an important role in this process. A major question is whether CTD Ser2P dephosphorylation occurs continuously along a gene as it is being transcribed or alternatively is restricted to the 3′ end where it could serve to recycle recently terminated Pol II. Consistent with an ongoing role in elongation *fcp1* mutants show an increase in the level of phosphorylation of Ser2 detected by mAb H5.[184,236] Surprisingly, the level of reactivity of the Ser2P specific antibody 3E10 is not changed arguing that the role of Fcp1 may not be able to remove all Ser2P but rather those adjacent to Ser5P.[184]

Transcription both in vivo and in vitro is stimulated by Fcp1 activity.[228a,c,236,241] One mechanism leading to higher levels of transcription is the removal of CTD phosphates resulting in a higher level of transcription initiation. Recent work by Lis and colleagues have shown that Fcp1 depletion in *Drosophila* does not change the phosphorylation state of chromatin bound Pol II but leads instead to an increase in the amount of nonchromatin bound phosphorylated Pol II.[241] This result suggests a role for Fcp1 in providing the pool of unphosphorylated initiation competent Pol II.[241]

7.2. CTD Phosphatase Ssu72

The *SSU72* gene was initially identified as a modifier of a mutation in the gene encoding the general transcription factor TFIIB that causes a change in Pol II start-site selection.[242] The presence of a CX5R sequence motif in Ssu72 suggested that this protein is related to protein tyrosine phosphatases and this prediction was confirmed through biochemical and genetic analyses.[243]

Ssu72 is present in Pol II complexes throughout the transcription cycle. It plays a role in start site selection through interaction with TFIIB.[244] Mutations in *ssu72* alter the sensitivity of yeast to 6-azauracil arguing for a role in elongation.[244c] Ssu72 also plays a role in 3′end formation as a part of the cleavage and polyadenylation complex.[91e,245] Finally, Ssu72 brings the transcription cycle full circle by promoting interactions between the 5′ and 3′ ends in a process called gene looping.[246]

Biochemical studies of Ssu72 show that this phosphatase preferentially removes Ser5P from the CTD.[234b,247] Rather than acting at the 5′ end of genes where Ser5P predominates, Ssu72 has recently been shown to remove Ser5P that persists into the 3′ end. Mutation of Ssu72 causes the persistence of Ser5P and the termination of Pol II that retains Ser5P.[184] Ser7P mirrors Ser5P at the 3′ end and mutation of *ssu72* leads to a build up of this phosphorylation mark suggesting that Ssu72 is a dual specificity phosphatase or that it indirectly affects removal of phosphate from Ser7P .

The activity of Ssu72 is enhanced by the peptidyl prolyl isomerases Ess1[248] and structural studies of Ssu72-CTD substrates show that the phosphophatase preferentially interacts with the *cis* configuration of the Ser5P-Pro6 bond.[249] Mutation of *ess1* leads to a retention of both Ser5P and Ser7P on Pol II

dx.doi.org/10.1021/cr400158h | *Chem. Rev.* 2013, 113, 8423–8455

mapping to the 3′ ends of genes. Taken together, these results indicate that Ssu72 plays an important role in recycling Pol II.

7.3. Rtr1/RPAP2 Plays a Role in CTD Dephosphorylation

Yeast Rtr1 is related to the human Pol II-associated protein RPAP2.[250] The yeast Rtr1 protein was identified as a protein that interacts with Pol II and in yeast localizes to the coding region of genes in a ChIP assay.[251] Deletion of Rtr1 leads to an increase in Ser5 CTD phosphorylation suggesting that Rtr1 is a CTD phosphatase and two different groups have reported CTD phosphatase activity for purified Rtr1 or Rpap2.[251,252]

Although poorly conserved overall, the present of a conserved Zn-finger motif suggested this might be a phosphatase active site. However, the structure of Rtr1 does not contain an active site cleft with conserved residues at the bottom and bacterially expressed Rtr1 does not display any CTD phosphatase activity suggesting that Rtr1 may play a regulatory role in CTD dephosphorylation.[253] Future studies will need to address the precise role of Rtr1 in CTD dephosphorylation.

8. REARRANGING THE CTD CODE: PROLINE ISOMERASES

Prolyl peptide bonds can adopt either of two conformations designated cis and trans. Isomerization between the cis and trans conformations in proteins has an activation energy barrier of about 20−25 kcal mol^{-1} and thus the isomerization reaction is slow and is often the rate-limiting step in protein folding.[254] In globular proteins the percentage of cis X-P bonds is low and depends on the preceding amino acid and on the structural context.[33] Given these constraints, the majority of cis bonds in proteins occur in turn motifs.[33] In more disordered protein domains like the CTD the Ser-Pro bond is less constrained. If 30% of the S−P bonds in the mammalian CTD were to adopt the cis conformation[29b] this would mean that on average every CTD would contain about 30 cis S−P bonds. The rate of cis−trans isomerization is greatly enhanced by peptidyl-prolyl-cis−trans-isomerases (PPIases).[255] Several PPIases have been shown to interact with the CTD including Pin1, Ess1, and Rrd1.

8.1. Pin1

Pin1 was the first PPIase shown to alter the conformation of the CTD.[256] This metazoan enzyme was initially identified as a protein that interacts with NIMA to regulate cell cycle[257] and together with its yeast homologue Ess1 are members of the parvulin class of PPIases.[258] These PPIases do not bind immunosuppressive drugs and thus are distinct from the immunophilin PPIases.[259] Pin1 is dependent on phosphorylation of its target site, Ser or Thr followed by Pro.[260] While it is impossible to measure the isomerization state of the CTD in vivo, genetic manipulations of PPIase gene function cause changes in the CTD modification state and thus one can infer a role for proline isomerization in this process. Pin1 inhibits dephosphorylation of the CTD by Fcp1.[261] Overexpression of Pin1 causes Pol II to dissociate from the chromatin template and a hyperphosphorylated form designated Pol IIoo accumulates in speckle structures in the nucleus.[262] Depletion of Pin1 enhances transcription and Xu and Manley have shown that Pin1-induced transcription inhibition occurs early in the transcription cycle at the transition between initiation and elongation.[262]

8.2. Ess1

Ess1 is the yeast homologue of Pin1 and is the only essential PPIase gene in *Saccharomyces cerevisiae*.[263] Like Pin1, Ess1 has been shown to interact with the CTD.[264] While Pin1 has been shown to isomerize many target proteins,[265] Ess1 has only been shown to act on the CTD. In vitro Ess1 isomerizes the Ser5-Pro6 bond about 5-times more effectively than the Ser2-Pro3 bond.[266] Ess1 plays several roles in RNA synthesis and processing. Hani et al identified *ESS1* (*PTF1*) in a screen for mutants that failed to terminate transcription at the *ADH1* 3′-end.[258,267] Subsequent work from Hanes and colleagues has expanded this to show that Ess1 plays a role in both mRNA and ncRNA termination and processing.[184,268] One possible mechanism for Ess1 function is to enable the Ser5P-Pro6 bond to adopt the *cis* conformation and thus enable Ssu72 to bind and dephosphorylate this residue.[249,269] Pcf11 prefers to bind Ser2P and the continued presence of Ser5P could reduce Ser2 phosphorylation.[91c] In addition, failing to remove Ser5P would lead to the persistence of Nrd1 binding and thus enhance premature termination of many transcripts without polyadenylation.[91e]

8.3. Rrd1

Rrd1 is an *S. cerevisiae* PPIase that is distinct from the Pin1/Ess1 family in that it is sensitive to the immunosuppressive drug rapamycin. This nonessential proline isomerase was initially identified by its similarity to the human phosphotyrosyl phosphatase activator PTPA.[270] Deletion of Rrd1 confers resistance to rapamycin[270] and subsequent studies demonstrated that Rrd1 is a PPIase.[271] Rdr1 has been shown to act at the transcriptional level.[272] Genes that are normally activated at the diauxic shift are not properly up regulated in an *rrd1* mutant and expression of genes like the ribosomal protein genes that are normally down-regulated are not reduced in an *rrd1* mutant.[272,273]

While the mechanism of Rrd1 has not been elucidated several possible targets have been identified. Rrd1 binds to the yeast PP2A-like phosphatase Sit4.[274] One possibility is that Rrd1 isomerizes and thereby activates this phosphatase and this leads to changes in the phosphorylation state of transcription activators and repressors. A second potential target is the CTD. Recent studies have indicated that Rrd1 binds to chromatin in vivo and to the CTD and isomerizes it both in vitro and in vivo.[273,275] Like the case of Pin1 this leads to removal of Pol II from the chromatin template.[275] In contrast to the case with Ess1 there is no evidence that Rdr1-induced changes in CTD conformation alter the posphorylation state of the CTD.[275]

9. READING THE CTD CODE

The highly conserved tandem arrangement of CTD heptad repeats initially suggested that the CTD interacts with a regularly repeating structure like DNA or RNA and some experiments indicated such interactions can occur.[25b,60,276] However, the functions of these interactions have not been forthcoming and the emphasis is now on proteins that interact with specific sequences or modifications within the CTD to execute CTD functions.

CTD-interacting proteins have been identified both by in vivo genetic approaches[79b] and through in vitro biochemical techniques.[107a,277] This has led to identification of a large number of potential interacting proteins but in vivo verification of these interactions is more limited. In this section we will focus on a subset of these interactions that have been verified as interacting directly with the CTD. Much of our knowledge of the stereospecificity of these interactions has come from three-dimensional structures of CTD-binding proteins bound to CTD peptides. A number of proteins that bind to different phosphorylated forms of the CTD are involved in various stages

dx.doi.org/10.1021/cr400158h | *Chem. Rev.* 2013, 113, 8423−8455

of the transcription cycle and several of these will be discussed in general terms as to their roles. The reader is directed to the reviews by Jeronimo et al. and Eick and Geter in this series for a further discussion of gene-specific functions of CTD-binding proteins.

9.1. Mediator

The CTD undergoes reversible phosphorylation during the transcription cycle.[278] The unphosphorylated Pol IIA form assembles in the preinitiation complex[36,46] and is converted to the IIO form by phosphorylation on Ser5 by the TFIIH kinase.[158a,c,161b] This phosphorylation event coincides with promoter clearance leading to the idea that phosphorylation of the CTD releases Pol II from the PIC allowing the transition to elongation. While elements of the model are correct, the situation is likely more complex.

Early genetic and biochemical approaches suggested that the CTD interacts with the Mediator complex that enables regulation of Pol II transcription.[173b,c,279] The genes encoding Mediator subunits were initially identified as suppressors of CTD truncation mutants[22] and later were shown to be components of the megaDalton Mediator complex.[67] Mediator contains over 30 subunits in four different submodules. The tail module interacts with activators and repressors, while the head and middle modules interact with Pol II. The CKM containing the Ckd8 CTD kinase activity reversibly associates with the middle module. Mediator can bind to the uphosphorylated CTD in vitro[68] and can be displaced from the Pol II holoenzyme complex by mAb 8WG16[67,280] that binds to the unphosphorylated CTD.[144b] Truncation of the CTD weakens the response to activators that function through the Mediator and lengthening the CTD acts to suppress mutations in Mediator genes.[281] Together, these results pointed to a functional interaction between Mediator and the CTD.

Purification of the mediator using an antibody to the CTD strongly suggests that the CTD interacts directly with the CTD but early efforts to identify interacting subunits were unsuccessful. The reason for this failure has recently become apparent as new studies have shown that the CTD makes multiple contacts with the Mediator including both the head[282] and the middle module.[37] The strongest interactions occur between the CTD and the middle module and are thought to represent the initial stages of holoenzyme formation.[37] Although the precise interaction site has not been identified, the CTD binding site on the middle module overlaps with the CKM binding site and the presence of CKM on the Mediator interferes with Pol II binding and holoenzyme formation.[37] This nicely explains the identification of the CKM genes as suppressors of CTD truncation mutations. Presumably truncation of the CTD weakens interaction with Mediator and this weak interaction can be suppressed by mutations that weaken the competitive interaction of CKM with Mediator.[37]

Once the initial complex between the CTD and the middle module forms it has been proposed that the conformation of the preholoenzyme changes, positioning the CTD on the head module. In the head module the CTD contact surface spreads over 73 Å and involves almost four heptad repeats in an extended conformation.[282] Specific contacts are made with highly conserved regions of Med6, Med8, and Med17. The distributed nature of these interactions may explain why single mutations are unlikely to eliminate CTD interactions with the Mediator.

What releases the CTD from Mediator contacts? In vitro experiments first showed that TFIIH is able to disrupt holoenzyme complexes containing Mediator and Pol II assembled on promoters in vitro.[283] The recent finding that in vivo inhibition of analog sensitive TFIIH kinase does not inhibit transcription leaves open the possibility that CTD phosphorylation is not necessary for releasing Pol II from the PIC. The Mediator makes non-CTD contacts with Pol II,[284] and the act of initiation may alter the conformation of the catalytic core in a way that weakens the affinity for Mediator[285] without the need for CTD phosphorylation by TFIIH. An alternative explanation is that other CTD kinases are able to substitute for TFIIH and by phosphorylating the CTD release Pol II from the PIC. In vitro, Mediator can be released by CTDK-I arguing that phosphorylation itself may be sufficient for release and the Ser2P or Ser5P are sufficient to loosen the CTD-Mediator interactions.[283b] Precisely which contacts are disrupted are not clear but it is interesting to note that CTD phosphorylation may induce a conformational change in the Mediator.[286]

9.2. Capping Complex

The m^7GpppN 5′ cap is added to Pol II transcripts early in the transcription process[72b,287] and serves to protect the nascent transcript from premature decay and direct further processing and translation of the mRNA (reviewed in Ghosh and Lima, 2010[288]). Targeting of the capping reaction to Pol II transcripts is generated through the interaction of the capping complex with Pol II phosphorylated on Ser5 in the early elongation complex.[73,74,77,288] How the capping machinery interacts with the CTD has been addressed through crystallographic analysis of complexes formed between the guanyltransferase enzymes and CTD peptides.

Lima, Shuman and colleagues crystallized the nucleotidyl transferase domain of the *Candida albicans* Cgt1 GTase subunit bound to a Ser5 phosphorylated CTD peptide containing four repeats.[289] The GTase domain adopts a mixed α/β fold with an interaction surface consisting of a 40 Å long channel in which 17 noncontiguous amino acids of the CTD peptide lie in an extended ß-like conformation. Within this channel there are two CTD docking sites (CDS1 and CDS2, Figure 5) that make specific contacts with CTD side chains. In each of the docking sites there is a positively charged depression that binds the phosphate on Ser5 of the first and third heptad repeats. Critical interactions are also made with tyrosine residues both through hydrophobic interaction (CDS1) and hydrogen bonding (CDS2). All of these interactions are fully supported by genetic structure−activity experiments.[289]

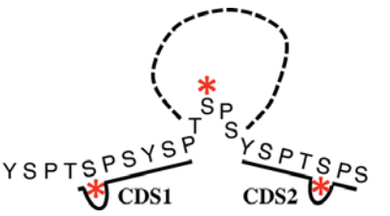


Figure 5. Diagram showing the recognition of adjacent heptad repeats by the *C. albicans* capping enzyme CDS1 and CDS2 binding pockets. The dotted line indicates the potential position of additional repeats if the CDS1 and CDS2 binding pockets recognize heptads that are not adjacent but separated by one or more heptad repeat.

The four residues between CDS1 and CDS2 do not make contact and form a bend that protrudes from the surface of the protein. This feature of the CTD-Cgt1 complex has important implications for the interaction not only of Cgt1 but for other

dx.doi.org/10.1021/cr400158h | *Chem. Rev.* 2013, 113, 8423−8455

CTD-binding proteins. First, the Ser5P in the second repeat makes no contacts with the protein indicating that Ser5 on contiguous heptads need not be phosphorylated to interact with Cgt1. Thus, low levels of Ser5 might be sufficient to recruit the capping complex. Second, the looping out of one or more repeats may result in CTD heptads adopting conformations that are stabilized relative to CTD repeats free in solution. This may enhance further interactions or modifications.

The structure of the human Cgt1 homologue Mce1 has recently been solved in a complex with the same phosphorylated peptide.[290] The CTD in this structure is also in an extended ß-like conformation but there are important differences in the details of the interactions compared to the yeast complex. Mce1 interacts directly with only six amino acids, S*PSYS*P in which Ser2 and Ser5 are phosphorylated (asterisks). In this structure both Pro3 and Pro6 adopt a trans conformation and Tyr1 is packed into a hydrophobic pocket with its hydroxyl is hydrogen-bonded to a Glu residue at the bottom of the pocket. The Ser5P in this structure makes direct contact with two Arg side chains but the Ser2P does not contact the enzyme. Interactions with Tyr and Ser5 are supported by earlier mutational studies showing that mutations that alter these two CTD positions fail to bind the enzyme.[291] The observation that Ser2P makes no direct contacts and points away from the Mce1-CTD interface is inconsistent with previous binding studies showing that Mce1 binds equally to CTD peptides phosphorylated at either Ser2 or Ser5. These binding studies indicate a second Ser2P binding site on Mce1 but the nature of this interaction remains unclear. Given the difference in the manner of interaction with the CTD in Mce1 and Cgt1 it is not surprising that the Mce side chains that interact with the CTD are not conserved in yeast. The one conserved aspect of the interaction with the CTD is the dependence of both structures on interactions with Ser5P and Tyr1. These interactions serve to direct one of the most important CTD functions, coupling transcription to 5′-end-capping.

9.3. WW Domain Interacts with Phosphorylated CTD Repeats

WW domains are approximately 40 amino acid domains characterized by a pair of conserved tryptophan residues.[292] These domains interact specifically with proline-rich peptides[293] and in the case of type IV WW domains with phosphorylated SerP/ThrP-Pro diamino acids. The function of WW domains is typically in providing an interaction module for proteins involved in signal transduction.

9.3.1. Pin1/Ess1. The proline isomerases Pin1 and its yeast homologue Ess1 have been shown to interact with the phosphorylated CTD.[256,264a,267,294] These proteins contain two domains; an N-terminal WW domain and a C-terminal peptidyl prolylisomerase (PPIase) domain of the parvulin family.[295] The PPIase of these enzymes shows a weak preference for binding to phosphorylated Ser/Thr-Pro sequences but the WW domain has a higher affinity and thus is the major determinant of substrate specificity.[296]

The structure of the Pin1 WW domain has been determined by X-ray crystallography both in the absence of a peptide in the binding site[295a] and in a complex with a canonical CTD heptapeptide phosphorylated at both Ser2 and Ser5.[296b] The CTD peptide lies in a cavity between the PPIase and the WW domain making contacts on the surface of the WW domain but not with the PPIase domain. The CTD peptide adopts an extended coil conformation in which the last five residues (PTS*PS) make contacts. Ser5P makes contacts with three residues that form a phosphate binding module. Ser2P, in contrast, makes no contacts. The proline residues are in the trans conformation making both hydrophobic and van der Waals interactions and the backbone of the TS*P sequence is clamped in place by a pair of aromatic residues. Interestingly, the Tyr1 residue makes no contacts and is apparently not involved in binding. Solution binding studies,[296b] however, do indicate that Tyr1 is important. An alternative interpretation is that longer peptides may bind through interactions with multiple phosphates as well as the Tyr residue.

9.4. CTD-Interacting Domains (CIDs) and Transcription Termination

The CTD-interacting domain (CID) was initially identified by its ability to interact with the CTD in a yeast 2-hybrid assay and by direct binding to CTD fusion proteins.[79b,99] The CID is an approximately 140 amino acid domain with similarity to the VHS domain[26c,297] and like this domain is commonly found at the N-terminus of proteins. The first identified CID-containing proteins were mammalian SCAF8/RBM16 and SCAF4/SFRS1, RNA-binding proteins of unknown function.[79b] More recently mammalian proteins RPRD1A and RPRD1B have been shown to contain a CID and to play an unspecified role in regulating the CTD phosphorylation state.[298] Yeast proteins containing this domain (Nrd1,[93,299] PCF11,[91a,300] and Rtt103[76,301]) have been implicated in transcription termination through their interaction with the CTD.[91e] Each of these CIDs has a slightly different specificity for CTD phosphorylation states and the structures of CIDs from several of these proteins have been described and shown to interact with the CTD in a variety of different ways (Figure 6a).

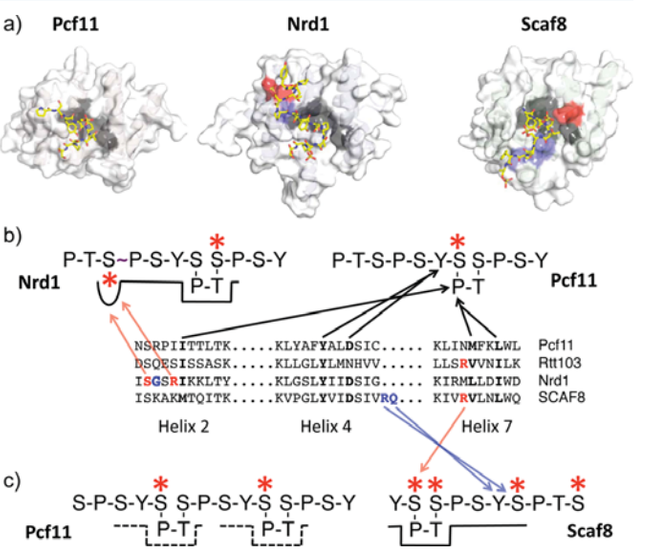


Figure 6. CTD-interacting domain (CID). (a) Structures of the CID of Pcf11, Nrd1, and Scaf8 bound to CTD peptides. (b) Diagram showing the interaction of CID side chains with CTD residues. Black arrows indicate interactions with the ß-turn and preceding Tyr residue. Blue arrows indicate Scaf8 interactions with the downstream Tyr residue. Red arrows indicate interactions with phosphorylated Ser residues by Nrd1 and Scaf8. The position of these interactions is indicated by the same colors in part A. The wavey bond in the CTD peptide bound to Nrd1 indicates the cis conformation of this paptide bond. (c) Dotted lines indicate alternative binding modes for Pcf11 bound to a doubly phosphorylated CTD peptide. The solid line represents the interaction of SCAF8 with a doubly phosphorylated CTD repeat.

dx.doi.org/10.1021/cr400158h | Chem. Rev. 2013, 113, 8423−8455

9.4.1. PCF11. PCF11 is a subunit of the cleavage and polyadenylation machinery that processes the 3′-ends of mRNAs.[91e,302] The amino-terminal CID of PCF11 binds directly to the CTD.[91a,300] Although this interaction favors Ser2 phosphorylated CTD repeats[91c] binding to other phosphoisoforms is nearly as strong.[301] Pcf11 also binds RNA and has been implicated in removal of Pol II from the template during termination.[303] The structure of the Pcf11 CID has been determined both by crystallography and NMR spectroscopy.[26c,29b]

The Pcf11 CID module consists of eight α-helices that form a compact right-handed superhelix. The helices are linked by short well-ordered turns and the module is stabilized by a large hydrophobic core. The CTD peptide in this structure lays in a groove formed by conserved residues in helices 2, 4, and 7 and the CTD adopts a ß-turn structure consisting of Ser2P-Pro3-Thr4-Ser5 while flanking residues are in an extended conformation.[26c] Within this channel seven hydrogen bonds are formed between the CTD backbone and CID side chains (Figure 6b). Pro3 of the CTD ß-turn is buried in a hydrophobic pocket as is the upstream Tyr1 side chain. This Tyr1 residue also makes a hydrogen bond with a conserved Asp residue possibly explaining why Tyr1 to Phe mutations in the CTD are lethal in *S. cerevisiae*.[18] Surprisingly, the Ser2P does not make any contact with the CID.

The interaction between the Pcf11 CID and the CTD has also been studied in solution using chemical shift mapping. This approach finds interactions similar to those in the crystal structure but extending over a broader area (Figure 6C). Using a longer peptide containing three repeats two of the turn structures are able to interact with the CID. Because no evidence was obtained for simultaneous binding of two CIDs this indicates that the CTD peptide can alternate between two distinct binding modes (dotted lines) with adjacent binding sites alternatively occupying the CID binding pocket. This type of fuzzy complex[304] has interesting implications for movement of CTD binding proteins along the CTD.

A major question concerning the binding of the CTD is whether interacting proteins recognize preformed structures or rather the interaction induces a conformation that interacts specifically. Using both calorimetry and NMR Taylor and colleagues measured the binding properties of the Pcf11 CID with this two repeat peptide containing a single Ser2P.[29b] The K_d (equilibrium dissociation constant) for this interaction was about 10^{-4} M^{-1} indicating a short lifetime for the interaction. The entropic term of the interaction is unfavorable (TdS = −4.3 kCal mol^{-1}, 293.14 K) suggesting an induced-fit binding mode rather than docking of a preformed turn.[23a,29b]

9.4.2. Nrd1 CID. Nrd1 is an *S. cerevisiae* RNA-binding protein that forms a complex with another RNA-binding protein Nab3 and the RNA helicase Sen1.[299b,305] This complex is required for termination of noncoding Pol II transcripts like snoRNAs, snRNAs, and cryptic unstable transcripts (CUTs).[93,96−98,306] The N-terminal CID of Nrd1 binds to the CTD phosphorylated on Ser5 rather than Ser2[95] consistent with its termination function near the promoter where Ser5P predominates.[9,165] Ser7 phosphorylation slightly weakens this interaction.[307]

The Nrd1 CID binds the same ß-turn (Ser2-Pro3-Thr4-Ser5P) that is the central feature of other CID-CTD interactions.[307] This turn binds in a hydrophobic pocket created by helices 2, 4, and 7. Where Nrd1 differs from Pcf11 is in binding to a second upstream repeat (Figure 6b). This interaction occurs between residues at the tip of helix 2 and the upstream Ser5P. Specific contacts are made with Nrd1 residues Ser25 and Arg28. Interestingly, the CTD Ser5P-Pro6 bond in this upstream repeat is found in the *cis* conformation and this feature allows the Ser5P to fit into an electropositive pocket formed by a Gly26.[307] This position in the CID is occupied by more bulky side chains in other CIDs making Nrd1 unique in its specificity for Ser5P.

9.4.3. SCAF8/RBM16 CID. SCAF8/RBM16 is a mammalian RNA-binding protein that was identified in a two-hybrid screen for proteins that interact with the CTD.[79b] The sequence organization of SCAF8 and a related mammalian protein SCAF4/SFRS15 are most closely related to Nrd1 in that they contain a single RRM domain. The function of SCAF8/RBM16 has not yet been elucidated but this protein has been shown to colocalize with active transcription sites in nuclei.[99] The CID of this protein binds selectively to CTD fusion proteins phosphorylated on both Ser2 and Ser5.[99] More recent binding assays have supported this preference for Ser2P-Ser5P and have shown that Ser2P is the dominant feature recognized by SCAF8.[100]

Meinhart and colleagues have solved the structure of the SCAF8 CID complexed with various phosphorylated CTD peptides.[100] The structure of the SCAF8 CID-CTD complexes are generally similar to that of other CID-CTD complexes but show several key differences that can explain the altered specificity for doubly phosphorylated heptad repeats. In the complex containing two repeats of the consensus sequence phosphorylated on both Ser2 and Ser5 of both repeats the first repeat forms the Ser2P-Pro3-Thr4-Ser5P ß-turn common to other CID-CTD structures. This structure brings Ser2P and Ser5P into closer juxtaposition than might be expected when considering the negative charge on the phosphates (Figure 6c). The SCAF8 CID deals with this in two ways. First, distinct from other complexes the SCAF8 CID interacts specifically with Ser2P through Arg-118 (Figure 6b). Mutation of this residue reduces the affinity of SCAF8 for the peptide by an order of magnitude to a level exhibited by Pcf11 which does not specifically recognize this residue. Second, the charge on Ser5P is further compensated for by a water molecule (hydronium ion) in the crystal structure.[100] A third distinct feature of this complex is contacts between the C-terminal tip of helix 4 and the downstream heptad repeat. These additional contacts are only seen when Ser2 and Ser5 doubly phosphorylated peptides are bound. Taken together, the SCAF8 CID is adapted to recognize a heavily phosphorylated form of the CTD seen in transcription elongation complexes located in the coding regions of genes.

9.4.4. Rtt103 CID. Rtt103 was initially identified in a screen for regulators of Ty transposition in yeast[308] and later shown to interact with Rat1 and Rai1 and to play a role in transcription termination of Pol II.[76] Like PCF11 and Nrd1, Rtt103 has an N-terminal CID but unlike these other CID-containing proteins there is no evidence for an RNA-binding domain. The Rtt103 CID binds exclusively to heptad repeats phosphorylated on Ser2.[76] The structure of the Rtt103 CID-bound to a CTD peptide containing two repeats both phosphporylated on Ser2 has been determined by NMR spectroscopy and the overall structure of the CTD peptide is similar to that seen in the Pcf11-CTD complex with a ß-turn consisting of Ser2-Pro3-Thr4-Ser5.

9.4.5. CID-CTD Interaction Summary. The CID is a key binding module on several proteins that direct termination and processing of nascent Pol II transcripts. Although these CIDs share a common structural framework there is sufficient diversity that many different phosphoisomers can be selectively bound. Solution binding studies argue that the interaction of the CTD

dx.doi.org/10.1021/cr400158h | *Chem. Rev.* 2013, 113, 8423−8455

with CID domains takes place through an induced fit mechanism and such interactions are characteristically of low affinity.[23a] Several features of this type of interaction allow for an increase in specificity. First, as seen with Pcf11 the CID can rapidly shift registers along a CTD containing multiple binding sites. This not only serves to increase binding affinity but may also play a role in movement of this factor along the CTD. A second feature seen primarily in Rtt103 is the ability of individual CIDs to cooperatively bind to the CTD.[301] This would serve to increase the local concentration of key RNA processing factors.

9.5. SH2 Domain of SPT6 Interacts with the CTD

The SH2 (Src homology 2) domain is a conserved structural module of about 100 amino acids that binds phosphorylated Tyr and plays significant roles in signal transduction pathways.[309] Spt6 has two SH2 domains that are unusual in recognizing phosphoserine rather than phosphotyrosine. While a single Ser2P is sufficient for binding, a second SerP in the peptide enhances binding. The basis of this specificity is thought to lie in the phospho-binding pockets of the two noncanonical SH2 domains. In each case the deep pocket that normally accommodates a phosphotyrosine is shallower and more appropriate for a phosphoserine. The consequences of mutation of key residues in these pockets are consistent with this view but the details of the CTD-SH2 interactions await a crystal structure of the complex with a CTD peptide. Nonetheless, the interaction of Spt6 with Ser2 phosphorylated CTD is consistent with genome-wide ChIP experiments that place Spt6 in the body and especially near the 3′-ends of genes, regions that display the highest levels of Ser2 phosphorylation.[92c]

9.6. Histone Methylase Set2-Rpb1 Interacting (SRI) Domain

The histone methyltransferases Set1 and Set2 are part of the Pol II elongation complex.[310] Set2 interacts directly with Pol II and ChIP experiments place this protein in the coding region of transcribed genes.[113b,119a,120b] Recruitment of Set2 requires the CTD kinase CTDK-I suggesting that this recruitment requires Ser2 phosphorylation.[113b,120b,311] Set2 contains a C-terminal domain of about 100 amino acids that is required for the interaction with Ser2 phosphorylated CTD. This SRI domain consists of the three-helix bundle that resembles the s2 domain of bacterial sigma factors.[312313]The Set2 SRI domain binds a two repeat CTD phosphopeptide with affinities in the low micromolar range.[113a,312] Maximum binding requires two repeats phosphorylated on both Ser2 and Ser5. While the structure of the CTD in this complex was not determined, the interaction covers a broad region and involved interactions with two Tyr residues in the CTD.

9.7. FF Domain Proteins

The FF domain is a 50−60 amino acid phosphopeptide interaction module characterized by conserved phenylalanine residues near the N- and C-termini.[314] The domain is comprised of three α-helices arranged in an orthogonal bundle with a 310 helix connecting the second and third helices.[315] FF domains are usually found in repeated arrays of 4−6 domains separated by linkers of variable length.[314]

Three proteins with FF domains have been shown to bind to the CTD; the human transcription factor TCERG1 (CA150), and the yeast splicing factor Prp40 and mammalian HYPA/FBP11.[83,316] Each of these factors falls into a family of RNA-processing proteins that contain a pair of WW domains followed by four FF domains. TCERG1/CA150 is a factor that regulates transcription elongation and pre-mRNA splicing.[317] This factor interacts with the phosphorylated CTD through multiple FF domains.[316] Interestingly, the interaction of FF domains 4−6 with a variety of phosphorylated CTD peptides shows that this cluster of FF domains requires phosphorylation of Ser2, Ser5, and Ser7, a phosphorylation state that is observed in the middle of genes where cotranscriptional splicing occurs.[318]

9.8. FUS and Related Proteins

A recent report has indicated that the RNA-binding protein FUS (fused in sarcoma) can interact with the CTD both in vivo and in vitro.[319] This protein is of great interest as it is mutated in several neurological diseases.[320] FUS is a member of the FET family of RNA-binding proteins that include EWS (Ewing sarcoma) and the TFIID subunit TAF15.[320a,321] These proteins all contain a central conserved RNA-recognition motif (RRM) and a C-terminal Zn-finger motif.[320a] In addition, each protein contains a N-terminal low-complexity domain rich in Gln, Gly, Ser and Tyr residues. Recent work from McKnight and colleagues has shown that low complexity sequences in the FET proteins contain multiple copies of the motif [G/S]-Y-[G/S] and this sequence allows these domains to form both homotypic and hetertypic ameloid-like cross-ß structures.[27] The CTD also contains this motif and a recent study shows that the CTD can associate with FET proteins through the formation of heterotypic cross-ß structures (Kwan and Kato et al., submitted).

9.9. Other CTD-Interacting Proteins

The proteins we have discussed in this section have really only scratched the surface of the family of CTD interacting proteins. A large number of proteins identified in systematic binding assays have not yet been subjected to further study.[8c,107a,277] In addition a variety of proteins like Bur1,[185] RecQ5,[137] Cdc73,[185] and the MCM[129,130] proteins have been shown to bind the CTD but have not been extensively studied. Future work will illuminate the roles these proteins play in elaborating the CTD code.

9.10. Summary of CTD Interactions

The large network of interactions formed by the CTD is based on two basic principles. First, the CTD is structurally flexible and thus can adopt a wide variety of conformations. Second, CTD-interacting proteins are likely to bind using an induced-fit mode. Such coupled folding and binding yields highly specific interactions but with relatively low affinity.[23a] One advantage of conformational flexibility is that the CTD can bind both to modifying enzymes and to the readers of these modifications using overlapping sites that can adopt different conformations. The adaptability of CTD structure is seen in the variety of conformations the CTD adopts in complexes with CTD-binding proteins. These interactions are weak and thus have a short lifetime but are boosted by the number of different binding motifs present in the CTD.

10. GLOBAL CHANGES IN CTD PHOSPHORYLATION

While the CTD code primarily addresses the regulatory processes occurring on the CTD during the transcription cycle there are other instances where transcription within the cell is altered on a global scale. Several examples of this global alteration involve changes in CTD phosphorylation.

10.1. Stress-Induced Changes in CTD Phosphorylation

Sudden changes in environmental conditions like temperature, pH, osmolarity, oxidative state or the presence of DNA damaging agents can induce rapid and widespread changes in gene expression.[322] In *S. cerevisiae*, changing the temperature from 25 to 37 °C alters expression of about one-third of all genes.[322b] This

dx.doi.org/10.1021/cr400158h | *Chem. Rev.* 2013, 113, 8423−8455

change in gene expression allows the cell to respond to the changing condition by expressing a small set of genes including heat shock proteins (chaparones) that enable cells to deal with the effect of temperature on protein folding and assembly.[322c] In addition, transcription of highly expressed genes like those encoding ribosomal proteins and glycolytic enzymes are rapidly shut off. Since these genes recruit most of the active Pol II in the cell, heat shock is expected to alter the transcriptional state of almost all Pol II.

Bensaude and colleagues first observed that in HeLa cells heat shock results in the accumulation of the phosphorylated IIo form of the largest subunit.[323] Further work showed that this change is brought about, not necessarily directly, by stress-activated kinases,[324] inhibition of cyclin-dependent kinases and inactivation of CTD phosphatase FCP1.[325] The increased reactivity toward the CC3 mAb suggests that the increase in IIo abundance is due to phosphorylation of Ser2.[144b] When *S. cerevisiae* cultures are shifted from 25 to 37 °C there is a rapid but transient increase in the level of Ser2P.[144b,326] The timing of this change correlates with the rapid changes seen in steady state RNAs by microarray.[322b] As with the case in HeLa cells this increase is mainly due to Ser2P as there is no change in the reactivity to mAbs that recognize Ser5P.

Similar changes in CTD phosphorylation are also seen in other stress conditions. Yeast cells in logarithmic growth transcribe genes necessary for glycolysis and protein synthesis at high frequency. As glucose is depleted the Ras and Tor signaling pathways combine to alter expression of a large set of genes in what has been termed the diauxic shift.[327] The end result of this shift is a reduction in expression of glycolytic and protein synthesis genes and activation of genes required for oxidative phosphorylation. During the diauxic shift there is an increase in the phosphorylation of Ser2 as detected by the H5 mAb.[144b]

In *S. pombe* a similar increase in Ser2 is observed as cells are starved for nitrogen.[197b] This change in CTD phosphorylation pattern has been linked to a shift in gene expression leading to sporulation. As in *S. cerevisiae* the kinase that phosphorylates the Ser2 sites is the Cdk1 homologue Lsk1 which is activated through a MAP kinase cascade. Global changes in CTD phosphorylation are also brought about by DNA damaging agents.[328] In this case as well an increase in Ser2P is observed and this increase is mediated by CTDK-I.

A common theme of the stress response is an increase in the apparent level of Ser2 phosphorylation. This change in the global CTD phosphorylation pattern may be accomplished in several ways. First, there could be an increase in Ser2 phosphorylation through activation of Cdk9, Cdk12, or the *S. cerevisiae* homologues Bur1 or Ctk1. In *S. pombe* the Ctk1 homologue Lsk1 is activated through the MAP kinase signaling pathway.[197b] An alternative method for increasing Ser2P is through inhibition of Fcp1.[241] Finally, the inhibition of Ser5 phosphorylation may increase the reactivity of Ser2P to antibodies that are blocked by adjacent Ser5.

What is the rationale for increasing global Ser2P levels? For most genes initiation requires unphosphorylated Pol II that becomes phosphorylated on Ser5 while in the PIC. An increase in the level of Ser2 phosphorylated Pol II would have the effect of depleting the pool of Pol II available for initiation. In the case of stress response this is a desired result as most gene expression must be reduced. Ser2 phosphorylation correlates with promoter escape and the transition to productive elongation but for stress-related genes this paradigm may not apply. Transcription of heat shock genes is less dependent on TFIIH than are other genes.[329] In Drosophila, Pol II on heat shock genes appears to be less phosphorylated than on developmentally induced genes.[143a] In a similar manner the *S. pombe* small heat shock protein hsp9 is up regulated by Ser2 phosphorylation.[197b] Taken together this indicates that a subset of genes involved in responding to stress are transcribed through a CTD code that differs from the bulk of growth promoting genes.

10.2. Growth Stimulation and CTD Phosphorylation

In addition to changes in stress, the phosphorylation state of the CTD also changes in response to the sudden availability of nutrients and/or growth factors. Quiescent fibroblasts contain equal amounts of the IIA and IIO forms of Pol II. When these cells are stimulated to grow by the addition of serum there is a marked increase in the level of Pol IIO that persists for several hours and then subsides.[203] This change in phosphorylation state is not transcription dependent as a CTD fusion protein is similarly phosphorylated in response to serum stimulation. MAP kinases co purify with the serum induced CTD kinase activity and are thought to carry out this global modification.[203] Stimulation of quiescent lymphocytes yields a 5-fold increase in TFIIH kinase resulting an a similar increase in the level of CTD Ser5 phosphorylation.[330] This change in CTD phosphorylation is coincident with an increase in global promoter melting as PICs with unphosphorylated CTDs are converted to open complexes and execute promoter escape.

A recent report has also indicated that in yeast Cdc28 (Cdc2) is able to phosphorylate the CTD of highly transcribed genes.[199] Since Cdc2 activity is increased as cells enter the cell cycle after serum stimulation it is entirely possible that in mammalian cells Cdc2 plays a role in generating the high level of Pol IIo seen after growth stimulation.

10.3. CTD Phosphorylation in Early Development

Oocytes from different organisms contain large amounts of maternally derived Pol II that is held in store for use in the early cell division cycles.[331] The CTD of this stored maternal Pol II is phosphorylated despite the fact that it is not engaged in transcription. In *Xenopus* oocyte development the CTD is largely unphosphorylated despite the fact that these cells are transcriptionally active.[332] At the final stage of maturation the CTD becomes hyperphosphorylated through the action of MAP kinase. Within one hour after fertilization the level of phosphorylation is reduced.[332a] A similar hyperphosphorylated form of the CTD is present in mammalian oocytes.[333] As with *Xenopus*, the hypophosphorylated IIa form of Rbp1 can be detected several hours after fertilization.[333] What role CTD phosphorylation plays in the storage and mobilization of Pol II in oocytes and early zygotes is not known.

In a variety of organisms the specification of germ cell fate is based on transcription repression in a process that is thought to involve inhibition of CTD phosphorylation.[334] This was first observed in *C. elegans* where Seydoux and colleagues showed that the levels of both Ser5P and Ser2P are reduced in germ line nuclei compared to somatic nuclei.[335] They went on to show that PIE-1, a maternal protein that segregates with germ cells is capable of inhibiting Ser2 and Ser5 phosphorylation.[336]

The inhibition of Ser2 phosphorylation takes place through a sequence in PIE-1, YAPMAPT that acts like a CTD heptad repeat binding to the cyclin T subunit of P-TEFb and blocking CTD phosphorylation.[337] PIE-1 deleted for this sequence fails to repress P-TEFb in germ cells leading to elevated levels of Ser2P.[338] In a more recent study it was shown that Cdk12 is responsible for the bulk of Ser2 phosphorylation in germ cells.[339]

dx.doi.org/10.1021/cr400158h | *Chem. Rev.* 2013, 113, 8423−8455

It is possible that PIE-1 also inhibits Cdk12 or alternatively, the small amount of Ser2P remaining after Cdk12 knockdown may represent a small number of genes where Cdk9 provides Ser2 phosphorylation. Further work will be required to delineate that roles of Cdk9 and Cdk12 in the *C. elegans* soma and germline.

Inhibition of P-TEFb in germ cells also occurs in early *Drosophila* development. In this case the polar granule component Pgc is expressed in germ cell precursors that contain low levels of Ser2P.[340] Pgc forms an inactive complex with P-TEFb in these cells and overexpression of P-TEFb creates a similar phenotype as a *pgc* mutant. These results suggest that similar to PIE-I the Pgc protein is able to inhibit Ser2 phosphorylation. Pgc1 is not related to PIE-I indicating convergent evolution of this process.

Germline specification in mammals is different from *C. elegans* and *Drosophila* in that the earliest embryonic cells (epiblasts) are equally likely to contribute to the germline.[341] Only after about one week of mouse development are PGCs identifiable and as these cells migrate to the gonad they lose both Ser2 and Ser5 phosphorylation and RNA synthesis in these cells is markedly reduced.[342] The common theme of these observations is that inhibition of CTD Ser2 and Ser5 phosphorylation in germ cells is important in specifying germ cell fate. Since Ser2P and Ser5P are usually associated with active transcription by Pol II the implication is that many genes are inactive in germ cells. Clearly there are genes that escape this regulation and maintain the viability of germ cells. How these genes escape is not know but may involve an override of the CTD code.

11. CTD MODIFICATION AND POL II SUBNUCLEAR LOCALIZATION

Pol II forms large complexes that contain all of the components necessary to synthesize and process nascent transcripts. Pulse labeling RNA with Br-UTP first showed that RNA synthesis in HeLa cells is localized to 300−500 sites that have been termed transcription factories.[343] Each of these factories contain an average of eight Pol II molecules[344] with an apparent mass of >8 M Da.[343d] In metazoa these factories are bound to a nuclear substructure.[345] Whether such transcription factories exist in yeast has not been shown.

HeLa cell transcription factories only contain about one-quarter of the total Pol II in the cell, the remainder is not localized and apparently not engaged in transcription.[344,345b] Within factories two different CTD phosphorylation states have been observed; those with Ser5P only and those with both Ser2P and Ser5P.[346] The former are proposed contain primarily poised Pol II complexes while the later contain actively elongating complexes.[346] Recent experiments have shown that the Ser5P factories contain Cdk9 and are distinct from the Ser2P factories.[347]

In addition to transcription factories several studies have shown that Pol II is localized to subnuclear substructures that are not associated with transcription including speckles (interchromatin granule clusters),[79a,144a,348] paraspeckles,[348b] and Cajal bodies (CBs).[332b,349] Speckles are irregular shaped punctae found in the nucleoplasm of mammalian cells and consisting of splicing factors and other RNA processing proteins.[350] Although transcription does not occur in speckles, the Pol II found in these structures is phosphorylated preferentially on Ser2.[79a,144a,348b] This observation suggests that CTD phosphorylation may have an alternative function in storage or assembly of Pol II complexes. Pol II has also been identified in Cajal bodies (CBs). These are small spherical structures 0.1−1 μm in diameter located in the nucleoplasm.[351] There are typically only a few (~5) CBs in most cells although more are seen in *Xenopus* germinal vesicles. CBs contain many transcription and processing proteins and have been proposed as sites of assembly of the transcription and pre-mRNA processing machinery.[352] Pol II in CBs is phosphorylated primarily on Ser5 which is surprising since there is no evidence for ongoing transcription in these structures. The CTD may have a role in localization of Pol II to CBs. A GST-CTD fusion protein expressed in *Xenopus* germinal vesicles is rapidly phosphorylated and localized to CBs.[332b] Phosphorylation of this domain is not required, however, as a fusion protein with Ser2 and Ser5 converted to Ala is also localized to CBs. The localization of Pol II within the nucleus plays an as yet poorly defined role in transcription regulation and the localization of Pol II with different phosphorylated CTD isoforms to different subnuclear structures will be an important area of future research.

12. CONCLUDING REMARKS

The last three decades since the discovery of the Pol II CTD have seen a dramatic expansion in our understanding of this enigmatic domain. Research has focused both on the enzymes that establish the wide array of modification patterns and the myriad of proteins that recognize these marks. The overarching theme of these studies is the central role that the CTD plays in integrating Pol II transcription with processing of the nascent transcript and the interaction of Pol II with the chromatin template.

The nature of the CTD structure is well adapted to its role in tethering multiple functional activities to the Pol II catalytic core. The extended structure enables the CTD to interact with other proteins over a distance of up to a thousand angstroms. The flexible nature of the CTD allows different repeats to adopt different conformations in response to interactions with CTD modifiers or RNA processing proteins. While the number of tandem repeats seems to be far in excess of that necessary for viability in a number of different organisms CTD deletions have only been tested in a limited number of conditions and it seems certain that the natural length of the CTD confers a selective advantage under some conditions. Similarly, while the consensus heptad repeats are found in tandem in most organisms, mutations that insert amino acids between every other repeat are viable indicating that the minimal functional unit is contained within two repeats.[43c,56] Why then is the tandem register maintained in evolution? There must be some interaction that the CTD makes that covers more than a pair of repeats and this type of interaction must confer a selective advantage. One possibility is that the CTD makes contacts with other low complexity regions in proteins. Such interactions are often based on cross-beta structures in which polypeptide chains are in a fully extended conformation.[27] If the CTD were to make such interactions either with other proteins or within the CTD itself then this could explain the need to maintain the tandem heptad register.

Dynamic modification of the CTD enables changes in the tethered activities as Pol II proceeds through the transcription cycle. This is the basis of the CTD code that has emerged as a unifying concept in transcription regulation. According to this model particular CTD modifications specify particular outcomes for the transcription machinery, the chromatin template and the nascent transcript. These dynamic changes are inferred from ChIP experiments showing that Pol II at different points along a template carries different CTD modifications. In the simplest form of the code the CTD is unphosphorylated when Pol II enters the PIC allowing the CTD to interact with the Mediator.

Once initiation occurs the CTD is phosphorylated on Ser5 and Ser7 breaking the contacts with the Mediator and allowing promoter escape. Ser5P also serves to recruit the capping machinery ensuring that Pol II transcripts are modified at the 5′-end. Once capping is completed Pol II enters a stable elongation complex and the CTD phosphorylation pattern changes with an increase in Ser2P, Thr4P, and Tyr1P and a decrease in Ser5P. These changes allow the recruitment of splicing factors and enzymes that modify the chromatin template. Finally, as Pol II nears the 3′-end there is a further decrease in Ser5P, Ser7P and Thr4P allowing Ser2P to predominate. This modification recruits the cleavage and polyadenylation machinery and, following 3′-end formation and termination the CTD is further modified by complete dephosphorylation thus regenerating initiation-competent Pol II.

The CTD code model has been very effective in focusing attention on the role of CTD modifications. The prevailing model is that dynamic changes in CTD modification underlie many of the steps in the transcription cycle. Much of this model is based on static data, however, in which the Pol II machinery is frozen in place following formaldehyde treatment. This type of data yields snapshots of Pol II along the gene and some inference is necessary to paint a dynamic picture. This picture is also clouded by the fact that ChIP yields a population average of the CTD modifications at a particular point along a given gene. For example, near the 5′-end most genes have more Ser5P than Ser2P. This could be due to the presence of high levels of Ser5P and a small amount of Ser2P on each CTD or rather could mean that most of the Pol II near the 5′-end has only Ser5P and a small number of Pol II molecules have mainly Ser2P. This issue becomes more critical in the middle of genes where both modifications are present in similar amounts.

Further complicating the CTD code are the uncertain specificities and binding properties of the antibodies used to derive the ChIP data. The specificity of some of these antibodies are masked by adjacent phosphorylations. Other antibodies preferentially bind to multiply phosphorylated repeats. Finally, and perhaps most importantly, these antibodies are multivalent and thus bind to CTDs modified at multiple sites. Part of the affinity is thus not dependent solely on modification of a single repeat but is due to the probability that a nearby repeat is also modified. Until these antibody properties are fully understood the CTD code will only be a low-resolution model. To increase the resolution it will be necessary to carry out sequential immunoprecipitations to characterize the distribution of different CTD phosphoisomers. This has been done to a limited extent using western analysis and this has revealed subpopulations of Pol II with CTDs either phosphorylated on Ser5 and Ser7 or on Ser2 and Thr4.[54b] Net-seq or PAR-CLIP using this type of sequential precipitation will allow a more detailed view of the distribution of CTD phosphoisomers on the genome. This will also address the issue of whether the CTD code is the same for all genes.

Another emerging theme in the CTD code is the degree to which different modifications can interact. One example is the priming of one CTD kinase by phosphorylation by another CTD kinase. This can take place either through modification of the CTD substrate or by the direct phosphorylation of one kinase by another[140a,170,183,185,192] thus establishing an ordered series of phosphorylation reactions during the transcription cycle much like the series of phosphorylations that govern the cell cycle.[219] Another emerging principle is the allosteric effect of CTD phosphorylation marks. Ser5P has been shown to enhance the capping activity.[77,291,353] Isomerization of the Ser-Pro bonds in the CTD also cooperates with different CTD binding and modification steps. Ess1 and/or Pin1 are necessary for full activity of the Ssu72 phosphatase that recognizes Ser5P in the context of a cis peptide bond.[249,269] Finally, the binding of Nrd1 to the CTD requires one of three S−P bonds to be in the cis conformation.[307] All of these interactions work together to provide a set of sequential binding complexes that direct the processing of nascent Pol II transcripts. The picture that this paints is of a series of complexes that are recruited to a moving Pol II elongation complex and fall off once they have executed their required function.

The problem with this picture is that the various CTD −interacting proteins bind weakly to the CTD and are unlikely to remain bound long enough to complete a transcription cycle. The most well studied interactions with the CTD involve the CTD-interacting domain (CID) found in Pcf11, Rtt103, SCAF8 and Nrd1. This domain binds to the same basic structure in the CTD but details of the interaction vary among the proteins leading to specificity for Ser2P, Ser5P or both. The key feature recognized by these proteins is a ß-turn comprising Ser2-Pro3-Thr4-Ser5. This turn is not abundant in the CTD in solution and binding thus requires a disorder to order transition in the CTD. Such coupled folding and binding reactions typically are of low affinity and high specificity and that seems to hold for many CTD interactions. Some CTD interactors may overcome low binding affinity by cooperative interactions.[301] Alternatively, stable interaction may not be required. Because there are many potential binding sites within the CTD it may not be necessary for an "interacting" protein to be stably bound. In this case the CTD could act to increase the local concentration of CTD-binding sites and thus increase the local concentration of any interacting proteins.[354]

In this later model the CTD is not so much acting as a tether but as a fly casting apparatus luring the RNA-processing proteins closer to the site of their function. Appendages like the CTD are not seen on Pol I or Pol III perhaps indicating that their processing need not be tethered or otherwise concentrated. One possible reason for not tethering the processing factors for Pol I is that transcription by this polymerase is sequestered in the nucleolus. Pol II transcription, while not restricted to a subnuclear organelle is, however, localized in transcription factories. Whether the CTD is involved in localizing Pol II in such structures is an open question. Several lines of investigation suggest that the CTD may target Pol II to structures in the nucleus. For example, the interaction with SCAFs (CTD associated SR-like factors) that colocalize to sites of active transcription may direct Pol II to subnuclear structures.[79b,99] Another possibility is that the CTDs from different Pol II molecules could interact with each other or with a common low complexity factor to form a cluster of polymerases that form a transcription factory.

Many challenges remain before the role of the CTD is understood in detail. As is apparent from the studies cited here, the CTD is involved in many steps in the synthesis and processing of Pol II transcripts. Despite key insights into CTD function there are still many details that are not clear. Part of the problem in teasing out detailed mechanisms is the degree to which transcription and RNA processing are coupled. Thus, defects in one step caused by a CTD mutation may yield defects in subsequent or preceding RNA processing steps. Future advances will depend on our ability to biochemically reconstitute these coupled steps in transcription and processing and assess the

dx.doi.org/10.1021/cr400158h | Chem. Rev. 2013, 113, 8423−8455

role of the CTD at each step along the way. The large number of CTD mutations already constructed will be a valuable asset in this endeavour.

AUTHOR INFORMATION

Corresponding Author

*Phone: 410-955-4719. Fax: 410-502-6718. E-mail: jcorden@jhmi.edu.

Notes

The authors declare no competing financial interest.

Biography

Jeffry Corden received his Ph.D. in Biochemistry and Biophysics from Oregon State University for his work on viral chromatin with Drs. George Pearson and Ken van Holde. He then joined the laboratory of Pierre Chambon at the Faculte de Medcine of the University of Strasbourg, France where he worked to define RNA polymerase II promoter elements. Since 1982 he has been on the faculty at Johns Hopkins Medical School where his laboratory investigates the function of the RNA polymerase II CTD.

ACKNOWLEDGMENTS

The author thanks Dr. Anton Meinhart for Figure 6a. This work was funded by NIH Grant GM066108 to J.L.C.

REFERENCES

(1) Chapman, R. D.; Heidemann, M.; Hintermair, C.; Eick, D. *Trends Genet* **2008**, *24*, 289.
(2) Cramer, P.; Bushnell, D. A.; Kornberg, R. D. *Science* **2001**, *292*, 1863.
(3) (a) Corden, J. L.; Cadena, D. L.; Ahearn, J. M., Jr.; Dahmus, M. E. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 7934. (b) Allison, L. A.; Moyle, M.; Shales, M.; Ingles, C. J. *Cell* **1985**, *42*, 599. (c) Zehring, W. A.; Lee, J. M.; Weeks, J. R.; Jokerst, R. S.; Greenleaf, A. L. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 3698.
(4) (a) Roeder, R. G. In *RNA Polymerase*; Losick, R., Chamberlin, M., Eds.; Cold Spring Harbor Laboratory: Cold Spring Harbor, New York, 1976. (b) Dahmus, M. E. *J. Biol. Chem.* **1983**, *258*, 3956.
(5) Greenleaf, A. L.; Haars, R.; Bautz, E. K. *FEBS Lett.* **1976**, *71*, 205.
(6) Cadena, D. L.; Dahmus, M. E. *J. Biol. Chem.* **1987**, *262*, 12468.
(7) Dahmus, M. E. *Biochim. Biophys. Acta* **1995**, *1261*, 171.
(8) (a) Zhang, J.; Corden, J. L. *J. Biol. Chem.* **1991**, *266*, 2290. (b) Yuryev, A.; Corden, J. L. *Genetics* **1996**, *143*, 661. (c) Phatnani, H. P.; Greenleaf, A. L. *Genes Dev.* **2006**, *20*, 2922.
(9) (a) Komarnitsky, P.; Cho, E. J.; Buratowski, S. *Genes Dev.* **2000**, *14*, 2452. (b) Buratowski, S. *Nat. Struct. Biol.* **2003**, *10*, 679.
(10) Werner, F.; Grohmann, D. *Nat. Rev. Microbiol.* **2011**, *9*, 85.
(11) Haag, J. R.; Pikaard, C. S. *Nat. Rev. Mol. Cell. Biol.* **2011**, *12*, 483.
(12) Stiller, J. W.; Hall, B. D. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6091.
(13) Nakajima, K.; Chang, Y. C.; Suzuki, T.; Jigami, Y.; Machida, M. *Biosci. Biotechnol. Biochem.* **2000**, *64*, 641.
(14) Li, W. B.; Bzik, D. J.; Gu, H. M.; Tanaka, M.; Fox, B. A.; Inselburg, J. *Nucleic Acids Res.* **1989**, *17*, 9621.
(15) Stiller, J. W.; Duffield, E. C.; Hall, B. D. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11769.
(16) Mirkin, S. M. *Nature* **2007**, *447*, 932.
(17) (a) Jeffreys, A. J.; Royle, N. J.; Wilson, V.; Wong, Z. *Nature* **1988**, *332*, 278. (b) Jeffreys, A. J.; Wilson, V.; Thein, S. L. *Nature* **1985**, *314*, 67.
(18) West, M. L.; Corden, J. L. *Genetics* **1995**, *140*, 1223.
(19) (a) Corden, J. L. *Trends Biochem. Sci.* **1990**, *15*, 383. (b) Corden, J. L., Ingles, C. J. In *Transcriptional Regulation*; McKnight, S. L., Yamamoto, K. R., Ed.; Cold Spring Harbor Laboratory Press: Plainview, NY, 1992; Vol. *1*.
(20) Hedges, S. B.; Dudley, J.; Kumar, S. *Bioinformatics* **2006**, *22*, 2971.
(21) Nonet, M.; Sweetser, D.; Young, R. A. *Cell* **1987**, *50*, 909.
(22) Nonet, M. L.; Young, R. A. *Genetics* **1989**, *123*, 715.
(23) (a) Dyson, H. J.; Wright, P. E. *Nat. Rev. Mol. Cell. Biol.* **2005**, *6*, 197. (b) Tompa, P. *Trends Biochem. Sci.* **2002**, *27*, 527.
(24) Li, X.; Romero, P.; Rani, M.; Dunker, A. K.; Obradovic, Z. *Genome Inform. Ser. Workshop Genome Inform.* **1999**, *10*, 30.
(25) (a) Suzuki, M. *J. Mol. Biol.* **1989**, *207*, 61. (b) Suzuki, M. *Nature* **1990**, *344*, 562. (c) Suzuki, M.; Yagi, N. *Proc. Biol. Sci.* **1991**, *246*, 231.
(26) (a) Matsushima, N.; Creutz, C. E.; Kretsinger, R. H. *Proteins* **1990**, *7*, 125. (b) Cagas, P. M.; Corden, J. L. *Proteins* **1995**, *21*, 149. (c) Meinhart, A.; Cramer, P. *Nature* **2004**, *430*, 223.
(27) (a) Han, T. W.; Kato, M.; Xie, S.; Wu, L. C.; Mirzaei, H.; Pei, J.; Chen, M.; Xie, Y.; Allen, J.; Xiao, G.; McKnight, S. L. *Cell* **2012**, *149*, 768. (b) Kato, M.; Han, T. W.; Xie, S.; Shi, K.; Du, X.; Wu, L. C.; Mirzaei, H.; Goldsmith, E. J.; Longgood, J.; Pei, J.; Grishin, N. V.; Frantz, D. E.; Schneider, J. W.; Chen, S.; Li, L.; Sawaya, M. R.; Eisenberg, D.; Tycko, R.; McKnight, S. L. *Cell* **2012**, *149*, 753.
(28) (a) Sawaya, M. R.; Sambashivan, S.; Nelson, R.; Ivanova, M. I.; Sievers, S. A.; Apostol, M. I.; Thompson, M. J.; Balbirnie, M.; Wiltzius, J. J.; McFarlane, H. T.; Madsen, A. O.; Riekel, C.; Eisenberg, D. *Nature* **2007**, *447*, 453. (b) Fitzpatrick, A. W.; Debelouchina, G. T.; Bayro, M. J.; Clare, D. K.; Caporini, M. A.; Bajaj, V. S.; Jaroniec, C. P.; Wang, L.; Ladizhansky, V.; Muller, S. A.; MacPhee, C. E.; Waudby, C. A.; Mott, H. R.; De Simone, A.; Knowles, T. P.; Saibil, H. R.; Vendruscolo, M.; Orlova, E. V.; Griffin, R. G.; Dobson, C. M. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 5468.
(29) (a) Harding, M. M. *J. Med. Chem.* **1992**, *35*, 4658. (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.
(30) Bienkiewicz, E. A.; Moon Woody, A.; Woody, R. W. *J. Mol. Biol.* **2000**, *297*, 119.
(31) Dobbins, J. R.; Murali, N.; Long, E. C. *Int. J. Pept. Protein Res.* **1996**, *47*, 260.
(32) Zhang, J.; Corden, J. L. *J. Biol. Chem.* **1991**, *266*, 2297.
(33) Lu, K. P.; Finn, G.; Lee, T. H.; Nicholson, L. K. *Nat. Chem. Biol.* **2007**, *3*, 619.
(34) Nishi, N.; Ohiso, I.; Sakairi, N.; Tokura, S.; Tsunemi, M.; Oka, M. *Biochem. Biophys. Res. Commun.* **1995**, *206*, 981.
(35) Armache, K. J.; Mitterweger, S.; Meinhart, A.; Cramer, P. *J. Biol. Chem.* **2005**, *280*, 7131.
(36) Laybourn, P. J.; Dahmus, M. E. *J. Biol. Chem.* **1989**, *264*, 6693.
(37) Tsai, K. L.; Sato, S.; Tomomori-Sato, C.; Conaway, R. C.; Conaway, J. W.; Asturias, F. J. *Nat. Struct. Mol. Biol.* **2013**, *20*, 611.
(38) Allison, L. A.; Wong, J. K.; Fitzpatrick, V. D.; Moyle, M.; Ingles, C. J. *Mol. Cell. Biol.* **1988**, *8*, 321.
(39) Brickey, W. J.; Greenleaf, A. L. *Genetics* **1995**, *140*, 599.
(40) (a) Bartolomei, M. S.; Halden, N. F.; Cullen, C. R.; Corden, J. L. *Mol. Cell. Biol.* **1988**, *8*, 330. (b) Meininghaus, M.; Eick, D. *FEBS Lett.* **1999**, *446*, 173. (c) Chapman, R. D.; Conrad, M.; Eick, D. *Mol. Cell. Biol.* **2005**, *25*, 7665.
(41) Kim, W. Y.; Dahmus, M. E. *J. Biol. Chem.* **1989**, *264*, 3169.
(42) Boeke, J. D.; Trueheart, J.; Natsoulis, G.; Fink, G. R. *Methods Enzymol* **1987**, *154*, 164.

(43) (a) Schneider, S.; Pei, Y.; Shuman, S.; Schwer, B. *Mol. Cell. Biol.* **2010**, *30*, 2353. (b) Schwer, B.; Shuman, S. *Mol. Cell* **2011**, *43*, 311. (c) Schwer, B.; Sanchez, A. M.; Shuman, S. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 18024.
(44) Chapman, R. D.; Heidemann, M.; Albert, T. K.; Mailhammer, R.; Flatley, A.; Meisterernst, M.; Kremmer, E.; Eick, D. *Science* **2007**, *318*, 1780.
(45) Litingtung, Y.; Lawler, A. M.; Sebald, S. M.; Lee, E.; Gearhart, J. D.; Westphal, H.; Corden, J. L. *Mol. Gen. Genet.* **1999**, *261*, 100.
(46) Payne, J. M.; Laybourn, P. J.; Dahmus, M. E. *J. Biol. Chem.* **1989**, *264*, 19621.
(47) (a) Baskaran, R.; Chiang, G. G.; Mysliwiec, T.; Kruh, G. D.; Wang, J. Y. *J. Biol. Chem.* **1997**, *272*, 18905. (b) Baskaran, R.; Dahmus, M. E.; Wang, J. Y. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 11167.
(48) Chapman, R. D.; Palancade, B.; Lang, A.; Bensaude, O.; Eick, D. *Nucleic Acids Res.* **2004**, *32*, 35.
(49) Bartolomei, M. S.; Corden, J. L. *Mol. Cell. Biol.* **1987**, *7*, 586.
(50) Meininghaus, M.; Chapman, R. D.; Horndasch, M.; Eick, D. *J. Biol. Chem.* **2000**, *275*, 24375.
(51) Stiller, J. W.; McConaughy, B. L.; Hall, B. D. *Yeast* **2000**, *16*, 57.
(52) 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.
(53) Karagiannis, J.; Balasubramanian, M. K. *PLoS One* **2007**, *2*, e433.
(54) (a) Hsin, J. P.; Sheth, A.; Manley, J. L. *Science* **2011**, *334*, 683. (b) Hintermair, C.; Heidemann, M.; Koch, F.; Descostes, 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.
(55) Stiller, J. W.; Cook, M. S. *Eukaryot Cell* **2004**, *3*, 735.
(56) Liu, P.; Greenleaf, A. L.; Stiller, J. W. *Mol. Biol. Evol.* **2008**, *25*, 719.
(57) Liu, P.; Kenney, J. M.; Stiller, J. W.; Greenleaf, A. L. *Mol. Biol. Evol.* **2010**, *27*, 2628.
(58) (a) Thompson, C. M.; Koleske, A. J.; Chao, D. M.; Young, R. A. *Cell* **1993**, *73*, 1361. (b) Hengartner, C. J.; Thompson, C. M.; Zhang, J.; Chao, D. M.; Liao, S. M.; Koleske, A. J.; Okamura, S.; Young, R. A. *Genes Dev.* **1995**, *9*, 897. (c) 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.
(59) (a) Allison, L. A.; Ingles, C. J. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 2794. (b) Scafe, C.; Chao, D.; Lopes, J.; Hirsch, J. P.; Henry, S.; Young, R. A. *Nature* **1990**, *347*, 491.
(60) Peterson, C. L.; Kruger, W.; Herskowitz, I. *Cell* **1991**, *64*, 1135.
(61) (a) Gerber, H. P.; Hagmann, M.; Seipel, K.; Georgiev, O.; West, M. A.; Litingtung, Y.; Schaffner, W.; Corden, J. L. *Nature* **1995**, *374*, 660. (b) Chun, R. F.; Jeang, K. T. *J. Biol. Chem.* **1996**, *271*, 27888.
(62) (a) Buratowski, S.; Sharp, P. A. *Mol. Cell. Biol.* **1990**, *10*, 5562. (b) Zehring, W. A.; Greenleaf, A. L. *J. Biol. Chem.* **1990**, *265*, 8351. (c) Serizawa, H.; Conaway, J. W.; Conaway, R. C. *Nature* **1993**, *363*, 371.
(63) Kang, M. E.; Dahmus, M. E. *J. Biol. Chem.* **1993**, *268*, 25033.
(64) (a) Dahmus, M. E.; Kedinger, C. *J. Biol. Chem.* **1983**, *258*, 2303. (b) Moyle, M.; Lee, J. S.; Anderson, W. F.; Ingles, C. J. *Mol. Cell. Biol.* **1989**, *9*, 5750.
(65) Usheva, A.; Maldonado, E.; Goldring, A.; Lu, H.; Houbavi, C.; Reinberg, D.; Aloni, Y. *Cell* **1992**, *69*, 871.
(66) (a) Koleske, A. J.; Buratowski, S.; Nonet, M.; Young, R. A. *Cell* **1992**, *69*, 883. (b) Conaway, R. C.; Bradsher, J. N.; Conaway, J. W. *J. Biol. Chem.* **1992**, *267*, 8464.
(67) Kim, Y. J.; Bjorklund, S.; Li, Y.; Sayre, M. H.; Kornberg, R. D. *Cell* **1994**, *77*, 599.
(68) Myers, L. C.; Gustafsson, C. M.; Bushnell, D. A.; Lui, M.; Erdjument-Bromage, H.; Tempst, P.; Kornberg, R. D. *Genes Dev.* **1998**, *12*, 45.
(69) (a) Laybourn, P. J.; Dahmus, M. E. *J. Biol. Chem.* **1990**, *265*, 13165. (b) Arias, J. A.; Peterson, S. R.; Dynan, W. S. *J. Biol. Chem.* **1991**, *266*, 8055.
(70) Lux, C.; Albiez, 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.
(71) Marshall, N. F.; Peng, J.; Xie, Z.; Price, D. H. *J. Biol. Chem.* **1996**, *271*, 27176.
(72) (a) Coppola, J. A.; Field, A. S.; Luse, D. S. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 1251. (b) Rasmussen, E. B.; Lis, J. T. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 7923.
(73) Cho, E. J.; Takagi, T.; Moore, C. R.; Buratowski, S. *Genes Dev.* **1997**, *11*, 3319.
(74) (a) 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. (b) Yue, Z.; Maldonado, E.; Pillutla, R.; Cho, H.; Reinberg, D.; Shatkin, A. J. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 12898. (c) Ho, C. K.; Sriskanda, V.; McCracken, S.; Bentley, D.; Schwer, B.; Shuman, S. *J. Biol. Chem.* **1998**, *273*, 9577.
(75) Schroeder, S. C.; Schwer, B.; Shuman, S.; Bentley, D. *Genes Dev.* **2000**, *14*, 2435.
(76) Kim, M.; Krogan, N. J.; Vasiljeva, L.; Rando, O. J.; Nedea, E.; Greenblatt, J. F.; Buratowski, S. *Nature* **2004**, *432*, 517.
(77) Ho, C. K.; Shuman, S. *Mol. Cell* **1999**, *3*, 405.
(78) (a) McCracken, S.; Fong, N.; Yankulov, K.; Ballantyne, S.; Pan, G.; Greenblatt, J.; Patterson, S. D.; Wickens, M.; Bentley, D. L. *Nature* **1997**, *385*, 357. (b) Misteli, T.; Spector, D. L. *Mol. Cell* **1999**, *3*, 697.
(79) (a) Mortillaro, M. J.; Blencowe, B. J.; Wei, X.; Nakayasu, H.; Du, L.; Warren, S. L.; Sharp, P. A.; Berezney, R. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8253. (b) Yuryev, A.; Patturajan, M.; Litingtung, Y.; Joshi, R. V.; Gentile, C.; Gebara, M.; Corden, J. L. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6975. (c) Vincent, M.; Lauriault, P.; Dubois, M. F.; Lavoie, S.; Bensaude, O.; Chabot, B. *Nucleic Acids Res.* **1996**, *24*, 4649. (d) Kim, E.; Du, L.; Bregman, D. B.; Warren, S. L. *J. Cell Biol.* **1997**, *136*, 19.
(80) (a) Greenleaf, A. L. *Trends Biochem. Sci.* **1993**, *18*, 117. (b) Corden, J. L.; Patturajan, M. *Trends Biochem. Sci.* **1997**, *22*, 413. (c) Goldstrohm, A. C.; Greenleaf, A. L.; Garcia-Blanco, M. A. *Gene* **2001**, *277*, 31. (d) Dye, M. J.; Gromak, N.; Proudfoot, N. J. *Mol. Cell* **2006**, *21*, 849.
(81) Singh, J.; Padgett, R. A. *Nat. Struct. Mol. Biol.* **2009**, *16*, 1128.
(82) (a) Hirose, Y.; Tacke, R.; Manley, J. L. *Genes Dev.* **1999**, *13*, 1234. (b) Zeng, C.; Berget, S. M. *Mol. Cell. Biol.* **2000**, *20*, 8290. (c) Millhouse, S.; Manley, J. L. *Mol. Cell. Biol.* **2005**, *25*, 533.
(83) Morris, D. P.; Greenleaf, A. L. *J. Biol. Chem.* **2000**, *275*, 39935.
(84) Gornemann, J.; Barrandon, C.; Hujer, K.; Rutz, B.; Rigaut, G.; Kotovic, K. M.; Faux, C.; Neugebauer, K. M.; Seraphin, B. *RNA* **2011**, *17*, 2119.
(85) (a) David, C. J.; Boyne, A. R.; Millhouse, S. R.; Manley, J. L. *Genes Dev.* **2011**, *25*, 972. (b) David, C. J.; Manley, J. L. *Transcription* **2011**, *2*, 221.
(86) Gu, B.; Eick, D.; Bensaude, O. *Nucleic Acids Res.* **2012**.
(87) Munoz, M. J.; de la Mata, M.; Kornblihtt, A. R. *Trends Biochem. Sci.* **2010**, *35*, 497.
(88) de la Mata, M.; Kornblihtt, A. R. *Nat. Struct. Mol. Biol.* **2006**, *13*, 973.
(89) Das, R.; Yu, J.; Zhang, Z.; Gygi, M. P.; Krainer, A. R.; Gygi, S. P.; Reed, R. *Mol. Cell* **2007**, *26*, 867.
(90) (a) Proudfoot, N. *Curr. Opin. Cell. Biol.* **2004**, *16*, 272. (b) Buratowski, S. *Curr. Opin. Cell. Biol.* **2005**, *17*, 257.
(91) (a) Barilla, D.; Lee, B. A.; Proudfoot, N. J. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 445. (b) Kyburz, A.; Sadowski, M.; Dichtl, B.; Keller, W. *Nucleic Acids Res.* **2003**, *31*, 3936. (c) Licatalosi, D. D.; Geiger, G.; Minet, M.; Schroeder, S.; Cilli, K.; McNeil, J. B.; Bentley, D. L. *Mol. Cell* **2002**, *9*, 1101. (d) Hirose, Y.; Manley, J. L. *Nature* **1998**, *395*, 93. (e) Kuehner, J. N.; Pearson, E. L.; Moore, C. *Nat. Rev. Mol. Cell. Biol.* **2011**, *12*, 283.
(92) (a) Kim, H.; Erickson, B.; Luo, W.; Seward, D.; Graber, J. H.; Pollock, D. D.; Megee, P. C.; Bentley, D. L. *Nat. Struct. Mol. Biol.* **2010**, *17*, 1279. (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. (c) Mayer, A.; Lidschreiber, M.; Siebert, M.; Leike, K.; Soding, J.; Cramer, P. *Nat. Struct. Mol. Biol.* **2010**, *17*, 1272.
(93) Steinmetz, E. J.; Conrad, N. K.; Brow, D. A.; Corden, J. L. *Nature* **2001**, *413*, 327.

(94) (a) Carroll, K. L.; Ghirlando, R.; Ames, J. M.; Corden, J. L. *RNA* **2007**, *13*, 361. (b) Carroll, K. L.; Pradhan, D. A.; Granek, J. A.; Clarke, N. D.; Corden, J. L. *Mol. Cell. Biol.* **2004**, *24*, 6241. (c) Porrua, O.; Hobor, F.; Boulay, J.; Kubicek, K.; D'Aubenton-Carafa, Y.; Gudipati, R. K.; Stefl, R.; Libri, D. *EMBO J.* **2012**, *31*, 3935.
(95) (a) Gudipati, R. K.; Villa, T.; Boulay, J.; Libri, D. *Nat. Struct. Mol. Biol.* **2008**, *15*, 786. (b) Vasiljeva, L.; Kim, M.; Mutschler, H.; Buratowski, S.; Meinhart, A. *Nat. Struct. Mol. Biol.* **2008**, *15*, 795.
(96) (a) Arigo, J. T.; Eyler, D. E.; Carroll, K. L.; Corden, J. L. *Mol. Cell* **2006**, *23*, 841. (b) Thiebaut, M.; Kisseleva-Romanova, E.; Rougemaille, M.; Boulay, J.; Libri, D. *Mol. Cell* **2006**, *23*, 853.
(97) Wyers, F.; Rougemaille, M.; Badis, G.; Rousselle, J. C.; Dufour, M. E.; Boulay, J.; Regnault, B.; Devaux, F.; Namane, A.; Seraphin, B.; Libri, D.; Jacquier, A. *Cell* **2005**, *121*, 725.
(98) (a) Arigo, J. T.; Carroll, K. L.; Ames, J. M.; Corden, J. L. *Mol. Cell* **2006**, *21*, 641. (b) Kuehner, J. N.; Brow, D. A. *Mol. Cell* **2008**, *31*, 201. (c) Jenks, M. H.; O'Rourke, T. W.; Reines, D. *Mol. Cell. Biol.* **2008**, *28*, 3883. (d) Thiebaut, M.; Colin, J.; Neil, H.; Jacquier, A.; Seraphin, B.; Lacroute, F.; Libri, D. *Mol. Cell* **2008**, *31*, 671.
(99) Patturajan, M.; Wei, X.; Berezney, R.; Corden, J. L. *Mol. Cell. Biol.* **1998**, *18*, 2406.
(100) Becker, R.; Loll, B.; Meinhart, A. *J. Biol. Chem.* **2008**, *283*, 22659.
(101) (a) Lei, E. P.; Krebber, H.; Silver, P. A. *Genes Dev.* **2001**, *15*, 1771. (b) Lei, E. P.; Silver, P. A. *Genes Dev.* **2002**, *16*, 2761. (c) Dermody, J. L.; Dreyfuss, J. M.; Villen, J.; Ogundipe, B.; Gygi, S. P.; Park, P. J.; Ponticelli, A. S.; Moore, C. L.; Buratowski, S.; Bucheli, M. E. *PLoS One* **2008**, *3*, e3273.
(102) (a) Bucheli, M. E.; Buratowski, S. *EMBO J.* **2005**, *24*, 2150. (b) Bucheli, M. E.; He, X.; Kaplan, C. D.; Moore, C. L.; Buratowski, S. *RNA* **2007**, *13*, 1756.
(103) Henry, M.; Borland, C. Z.; Bossie, M.; Silver, P. A. *Genetics* **1996**, *142*, 103.
(104) (a) Windgassen, M.; Sturm, D.; Cajigas, I. J.; Gonzalez, C. I.; Seedorf, M.; Bastians, H.; Krebber, H. *Mol. Cell. Biol.* **2004**, *24*, 10479. (b) Sachs, A. B.; Varani, G. *Nat. Struct. Biol.* **2000**, *7*, 356.
(105) Estrella, L. A.; Wilkinson, M. F.; Gonzalez, C. I. *J. Mol. Biol.* **2009**, *394*, 410.
(106) (a) Rajyaguru, P.; Parker, R. *Cell Cycle* **2012**, *11*, 2594-9. (b) Rajyaguru, P.; She, M.; Parker, R. *Mol. Cell* **2012**, *45*, 244.
(107) (a) Phatnani, H. P.; Jones, J. C.; Greenleaf, A. L. *Biochemistry* **2004**, *43*, 15702. (b) Mackellar, A. L.; Greenleaf, A. L. *J. Biol. Chem.* **2011**, *286*, 36385.
(108) Kim, V. N.; Dreyfuss, G. *Mol Cells* **2001**, *12*, 1.
(109) Hieronymus, H.; Silver, P. A. *Nat. Genet.* **2003**, *33*, 155.
(110) Kohler, A.; Hurt, E. *Nat. Rev. Mol. Cell. Biol.* **2007**, *8*, 761.
(111) (a) Strasser, K.; Hurt, E. *EMBO J.* **2000**, *19*, 410. (b) Zenklusen, D.; Vinciguerra, P.; Strahm, Y.; Stutz, F. *Mol. Cell. Biol.* **2001**, *21*, 4219.
(112) (a) Li, B.; Carey, M.; Workman, J. L. *Cell* **2007**, *128*, 707. (b) Rando, O. J. *Curr. Opin. Genet. Dev.* **2012**, *22*, 148. (c) Rando, O. J.; Winston, F. *Genetics* **2012**, *190*, 351.
(113) (a) Kizer, K. O.; Phatnani, H. P.; Shibata, Y.; Hall, H.; Greenleaf, A. L.; Strahl, B. D. *Mol. Cell. Biol.* **2005**, *25*, 3305. (b) Krogan, N. J.; Kim, M.; Tong, A.; Golshani, A.; Cagney, G.; Canadien, V.; Richards, D. P.; Beattie, B. K.; Emili, A.; Boone, C.; Shilatifard, A.; Buratowski, S.; Greenblatt, J. *Mol. Cell. Biol.* **2003**, *23*, 4207. (c) 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. (d) Ng, H. H.; Robert, F.; Young, R. A.; Struhl, K. *Mol. Cell* **2003**, *11*, 709.
(114) (a) Miller, T.; Krogan, N. J.; Dover, J.; Erdjument-Bromage, H.; Tempst, P.; Johnston, M.; Greenblatt, J. F.; Shilatifard, A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 12902. (b) Nagy, P. L.; Griesenbeck, J.; Kornberg, R. D.; Cleary, M. L. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 90. (c) Roguev, A.; Schaft, D.; Shevchenko, A.; Pijnappel, W. W.; Wilm, M.; Aasland, R.; Stewart, A. F. *EMBO J.* **2001**, *20*, 7137.
(115) Pokholok, D. K.; Harbison, C. T.; Levine, S.; Cole, M.; Hannett, N. M.; Lee, T. I.; Bell, G. W.; Walker, K.; Rolfe, P. A.; Herbolsheimer, E.; Zeitlinger, J.; Lewitter, F.; Gifford, D. K.; Young, R. A. *Cell* **2005**, *122*, 517.
(116) Laribee, R. N.; Krogan, N. J.; Xiao, T.; Shibata, Y.; Hughes, T. R.; Greenblatt, J. F.; Strahl, B. D. *Curr. Biol.* **2005**, *15*, 1487.
(117) (a) Venkatasubrahmanyam, S.; Hwang, W. W.; Meneghini, M. D.; Tong, A. H.; Madhani, H. D. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 16609. (b) Jiang, H.; Shukla, A.; Wang, X.; Chen, W. Y.; Bernstein, B. E.; Roeder, R. G. *Cell* **2011**, *144*, 513. (c) Lenstra, T. L.; Benschop, J. J.; Kim, T.; Schulze, J. M.; Brabers, N. A.; Margaritis, T.; van de Pasch, L. A.; van Heesch, S. A.; Brok, M. O.; Groot Koerkamp, M. J.; Ko, C. W.; van Leenen, D.; Sameith, K.; van Hooff, S. R.; Lijnzaad, P.; Kemmeren, P.; Hentrich, T.; Kobor, M. S.; Buratowski, S.; Holstege, F. C. *Mol. Cell* **2011**, *42*, 536.
(118) (a) Camblong, J.; Beyrouthy, N.; Guffanti, E.; Schlaepfer, G.; Steinmetz, L. M.; Stutz, F. *Genes Dev.* **2009**, *23*, 1534. (b) Berretta, J.; Pinskaya, M.; Morillon, A. *Genes Dev.* **2008**, *22*, 615.
(119) (a) Strahl, B. D.; Grant, P. A.; Briggs, S. D.; Sun, Z. W.; Bone, J. R.; Caldwell, J. A.; Mollah, S.; Cook, R. G.; Shabanowitz, J.; Hunt, D. F.; Allis, C. D. *Mol. Cell. Biol.* **2002**, *22*, 1298. (b) Schaft, D.; Roguev, A.; Kotovic, K. M.; Shevchenko, A.; Sarov, M.; Neugebauer, K. M.; Stewart, A. F. *Nucleic Acids Res.* **2003**, *31*, 2475.
(120) (a) Rao, B.; Shibata, Y.; Strahl, B. D.; Lieb, J. D. *Mol. Cell. Biol.* **2005**, *25*, 9447. (b) Xiao, T.; Hall, H.; Kizer, K. O.; Shibata, Y.; Hall, M. C.; Borchers, C. H.; Strahl, B. D. *Genes Dev.* **2003**, *17*, 654.
(121) (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.
(122) Keogh, M. C.; Kurdistani, S. K.; Morris, S. A.; Ahn, S. H.; Podolny, V.; Collins, S. R.; Schuldiner, M.; Chin, K.; Punna, T.; Thompson, N. J.; Boone, C.; Emili, A.; Weissman, J. S.; Hughes, T. R.; Strahl, B. D.; Grunstein, M.; Greenblatt, J. F.; Buratowski, S.; Krogan, N. J. *Cell* **2005**, *123*, 593.
(123) (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.
(124) (a) Kaplan, C. D.; Laprade, L.; Winston, F. *Science* **2003**, *301*, 1096. (b) Adkins, M. W.; Tyler, J. K. *Mol. Cell* **2006**, *21*, 405. (c) Bortvin, A.; Winston, F. *Science* **1996**, *272*, 1473.
(125) (a) Swanson, M. S.; Winston, F. *Genetics* **1992**, *132*, 325. (b) Hartzog, G. A.; Wada, T.; Handa, H.; Winston, F. *Genes Dev.* **1998**, *12*, 357.
(126) (a) Yoh, S. M.; Cho, H.; Pickle, L.; Evans, R. M.; Jones, K. A. *Genes Dev.* **2007**, *21*, 160. (b) Sun, M.; Lariviere, L.; Dengl, S.; Mayer, A.; Cramer, P. *J. Biol. Chem.* **2010**, *285*, 41597. (c) Dengl, S.; Mayer, A.; Sun, M.; Cramer, P. *J. Mol. Biol.* **2009**, *389*, 211.
(127) Kwon, S. H.; Florens, L.; Swanson, S. K.; Washburn, M. P.; Abmayr, S. M.; Workman, J. L. *Genes Dev.* **2010**, *24*, 2133.
(128) Rodriguez-Paredes, M.; Ceballos-Chavez, M.; Esteller, M.; Garcia-Dominguez, M.; Reyes, J. C. *Nucleic Acids Res.* **2009**, *37*, 2449.
(129) Yankulov, K.; Todorov, I.; Romanowski, P.; Licatalosi, D.; Cilli, K.; McCracken, S.; Laskey, R.; Bentley, D. L. *Mol. Cell. Biol.* **1999**, *19*, 6154.
(130) Gauthier, L.; Dziak, R.; Kramer, D. J.; Leishman, D.; Song, X.; Ho, J.; Radovic, M.; Bentley, D.; Yankulov, K. *Genetics* **2002**, *162*, 1117.
(131) Mayan, M. D. *PLoS One* **2013**, *8*, e53405.
(132) Cadoret, J. C.; Meisch, F.; Hassan-Zadeh, V.; Luyten, I.; Guillet, C.; Duret, L.; Quesneville, H.; Prioleau, M. N. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 15837.
(133) Wong, J. M.; Ingles, C. J. *Mol Gen Genet* **2001**, *264*, 842.
(134) Winsor, T. S.; Bartkowiak, B.; Bennett, C. B.; Greenleaf, A. L. *PLoS One* **2013**, *8*, e60909.
(135) (a) Aguilera, A. *EMBO J.* **2002**, *21*, 195. (b) Gonzalez-Aguilera, C.; Tous, C.; Gomez-Gonzalez, B.; Huertas, P.; Luna, R.; Aguilera, A. *Mol. Biol. Cell* **2008**, *19*, 4310. (c) Li, X.; Manley, J. L. *Cell* **2005**, *122*, 365. (d) Stirling, P. C.; Chan, Y. A.; Minaker, S. W.; Aristizabal, M. J.; Barrett, I.; Sipahimalani, P.; Kobor, M. S.; Hieter, P. *Genes Dev.* **2012**, *26*, 163. (e) Mischo, H. E.; Gomez-Gonzalez, B.; Grzechnik, P.; Rondon, A. G.; Wei, W.; Steinmetz, L.; Aguilera, A.; Proudfoot, N. J. *Mol. Cell* **2011**, *41*, 21.

(136) (a) Branzei, D.; Foiani, M. *Genes Dev.* **2007**, *21*, 3019. (b) Hu, Y.; Lu, X.; Barnes, E.; Yan, M.; Lou, H.; Luo, G. *Mol. Cell. Biol.* **2005**, *25*, 3431. (c) Hu, Y.; Raynard, S.; Sehorn, M. G.; Lu, X.; Bussen, W.; Zheng, L.; Stark, J. M.; Barnes, E. L.; Chi, P.; Janscak, P.; Jasin, M.; Vogel, H.; Sung, P.; Luo, G. *Genes Dev.* **2007**, *21*, 3073.
(137) Kanagaraj, R.; Huehn, D.; MacKellar, A.; Menigatti, M.; Zheng, L.; Urban, V.; Shevelev, I.; Greenleaf, A. L.; Janscak, P. *Nucleic Acids Res.* **2010**, *38*, 8131.
(138) Guo, Z.; Stiller, J. W. *BMC Genomics* **2004**, *5*, 69.
(139) (a) Egloff, S.; Dienstbier, M.; Murphy, S. *Trends Genet* **2012**, *28*, 333. (b) Egloff, S.; Murphy, S. *Trends Genet* **2008**, *24*, 280. (c) Heidemann, M.; Hintermair, C.; Voss, K.; Eick, D. *Biochim. Biophys. Acta* **2012**. (d) Jasnovidova, O.; Stefl, R. *Wiley Interdiscip Rev. RNA* **2012**.
(140) (a) Czudnochowski, N.; Bosken, C. A.; Geyer, M. *Nat. Commun.* **2012**, *3*, 842. (b) Morris, D. P.; Stevens, R. D.; Greenleaf, A. L. *Biochem. Biophys. Res. Commun.* **1998**, *245*, 53.
(141) (a) Dinkel, H.; Chica, C.; Via, A.; Gould, C. M.; Jensen, L. J.; Gibson, T. J.; Diella, F. *Nucleic Acids Res.* **2011**, *39*, D261. (b) Gnad, F.; Gunawardena, J.; Mann, M. *Nucleic Acids Res.* **2011**, *39*, D253. (c) Hornbeck, P. V.; Kornhauser, J. M.; Tkachev, S.; Zhang, B.; Skrzypek, E.; Murray, B.; Latham, V.; Sullivan, M. *Nucleic Acids Res.* **2012**, *40*, D261.
(142) Lee, J. M.; Greenleaf, A. L. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 3624.
(143) (a) Weeks, J. R.; Hardin, S. E.; Shen, J.; Lee, J. M.; Greenleaf, A. L. *Genes Dev.* **1993**, *7*, 2329. (b) O'Brien, T.; Hardin, S.; Greenleaf, A.; Lis, J. T. *Nature* **1994**, *370*, 75.
(144) (a) Bregman, D. B.; Du, L.; van der Zee, S.; Warren, S. L. *J. Cell Biol.* **1995**, *129*, 287. (b) Patturajan, M.; Schulte, R. J.; Sefton, B. M.; Berezney, R.; Vincent, M.; Bensaude, O.; Warren, S. L.; Corden, J. L. *J. Biol. Chem.* **1998**, *273*, 4689.
(145) Mayer, A.; Heidemann, M.; Lidschreiber, M.; Schreieck, A.; Sun, M.; Hintermair, C.; Kremmer, E.; Eick, D.; Cramer, P. *Science* **2012**, *336*, 1723.
(146) (a) Bishop, A. C.; Ubersax, J. A.; Petsch, D. T.; Matheos, D. P.; Gray, N. S.; Blethrow, J.; Shimizu, E.; Tsien, J. Z.; Schultz, P. G.; Rose, M. D.; Wood, J. L.; Morgan, D. O.; Shokat, K. M. *Nature* **2000**, *407*, 395. (b) Knight, Z. A.; Shokat, K. M. *Cell* **2007**, *128*, 425.
(147) Witucki, L. A.; Huang, X.; Shah, K.; Liu, Y.; Kyin, S.; Eck, M. J.; Shokat, K. M. *Chem Biol* **2002**, *9*, 25.
(148) (a) Shah, K.; Liu, Y.; Deirmengian, C.; Shokat, K. M. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 3565. (b) Shah, K.; Shokat, K. M. *Chem. Biol.* **2002**, *9*, 35.
(149) Jones, J. C.; Phatnani, H. P.; Haystead, T. A.; MacDonald, J. A.; Alam, S. M.; Greenleaf, A. L. *J. Biol. Chem.* **2004**, *279*, 24957.
(150) Kim, M.; Suh, H.; Cho, E. J.; Buratowski, S. *J. Biol. Chem.* **2009**, *284*, 26421.
(151) Morgan, D. O. *Annu. Rev. Cell. Dev. Biol.* **1997**, *13*, 261.
(152) Bregman, D. B.; Pestell, R. G.; Kidd, V. J. *Front. Biosci.* **2000**, *5*, D244.
(153) Pavletich, N. P. *J. Mol. Biol.* **1999**, *287*, 821.
(154) Songyang, Z.; Blechner, S.; Hoagland, N.; Hoekstra, M. F.; Piwnica-Worms, H.; Cantley, L. C. *Curr. Biol.* **1994**, *4*, 973.
(155) Chen, J.; Saha, P.; Kornbluth, S.; Dynlacht, B. D.; Dutta, A. *Mol. Cell. Biol.* **1996**, *16*, 4673.
(156) (a) Fesquet, D.; Labbe, J. C.; Derancourt, J.; Capony, J. P.; Galas, S.; Girard, F.; Lorca, T.; Shuttleworth, J.; Doree, M.; Cavadore, J. C. *EMBO J.* **1993**, *12*, 3111. (b) Matsuoka, M.; Kato, J. Y.; Fisher, R. P.; Morgan, D. O.; Sherr, C. J. *Mol. Cell. Biol.* **1994**, *14*, 7265. (c) Solomon, M. J.; Lee, T.; Kirschner, M. W. *Mol. Biol. Cell* **1992**, *3*, 13.
(157) (a) Thuret, J. Y.; Valay, J. G.; Faye, G.; Mann, C. *Cell* **1996**, *86*, 565. (b) Kaldis, P.; Sutton, A.; Solomon, M. J. *Cell* **1996**, *86*, 553.
(158) (a) Feaver, W. J.; Gileadi, O.; Li, Y.; Kornberg, R. D. *Cell* **1991**, *67*, 1223. (b) Feaver, W. J.; Svejstrup, J. Q.; Henry, N. L.; Kornberg, R. D. *Cell* **1994**, *79*, 1103. (c) Lu, H.; Zawel, L.; Fisher, L.; Egly, J. M.; Reinberg, D. *Nature* **1992**, *358*, 641. (d) Roy, R.; Adamczewski, J. P.; Seroz, T.; Vermeulen, W.; Tassan, J. P.; Schaeffer, L.; Nigg, E. A.; Hoeijmakers, J. H.; Egly, J. M. *Cell* **1994**, *79*, 1093. (e) Shiekhattar, R.; Mermelstein, F.; Fisher, R. P.; Drapkin, R.; Dynlacht, B.; Wessling, H. C.; Morgan, D. O.; Reinberg, D. *Nature* **1995**, *374*, 283.
(159) Larochelle, S.; Batliner, J.; Gamble, M. J.; Barboza, N. M.; Kraybill, B. C.; Blethrow, J. D.; Shokat, K. M.; Fisher, R. P. *Nat. Struct. Mol. Biol.* **2006**, *13*, 55.
(160) Yankulov, K. Y.; Bentley, D. L. *EMBO J.* **1997**, *16*, 1638.
(161) (a) Gebara, M. M.; Sayre, M. H.; Corden, J. L. *J. Cell. Biochem.* **1997**, *64*, 390. (b) Trigon, S.; Serizawa, H.; Conaway, J. W.; Conaway, R. C.; Jackson, S. P.; Morange, M. *J. Biol. Chem.* **1998**, *273*, 6769. (c) Rickert, P.; Corden, J. L.; Lees, E. *Oncogene* **1999**, *18*, 1093. (d) Ramanathan, Y.; Rajpara, S. M.; Reza, S. M.; Lees, E.; Shuman, S.; Mathews, M. B.; Pe'ery, T. *J. Biol. Chem.* **2001**, *276*, 10913.
(162) (a) Hengartner, C. J.; Myer, V. E.; Liao, S. M.; Wilson, C. J.; Koh, S. S.; Young, R. A. *Mol. Cell* **1998**, *2*, 43. (b) Keogh, M. C.; Podolny, V.; Buratowski, S. *Mol. Cell. Biol.* **2003**, *23*, 7005.
(163) (a) Akhtar, M. S.; Heidemann, M.; Tietjen, J. R.; Zhang, D. W.; Chapman, R. D.; Eick, D.; Ansari, A. Z. *Mol. Cell* **2009**, *34*, 387. (b) Boeing, S.; Rigault, C.; Heidemann, M.; Eick, D.; Meisterernst, M. *J. Biol. Chem.* **2010**, *285*, 188. (c) Glover-Cutter, K.; Larochelle, S.; Erickson, B.; Zhang, C.; Shokat, K.; Fisher, R. P.; Bentley, D. L. *Mol. Cell. Biol.* **2009**, *29*, 5455.
(164) (a) Akoulitchev, S.; Makela, T. P.; Weinberg, R. A.; Reinberg, D. *Nature* **1995**, *377*, 557. (b) Jiang, Y.; Yan, M.; Gralla, J. D. *Mol. Cell. Biol.* **1996**, *16*, 1614.
(165) Buratowski, S. *Mol. Cell* **2009**, *36*, 541.
(166) Adelman, K.; Lis, J. T. *Nat. Rev. Genet.* **2012**, *13*, 720.
(167) (a) Makela, T. P.; Parvin, J. D.; Kim, J.; Huber, L. J.; Sharp, P. A.; Weinberg, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 5174. (b) Tirode, F.; Busso, D.; Coin, F.; Egly, J. M. *Mol. Cell* **1999**, *3*, 87.
(168) (a) 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. (b) 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.
(169) Lee, K. M.; Miklos, I.; Du, H.; Watt, S.; Szilagyi, Z.; Saiz, J. E.; Madabhushi, R.; Penkett, C. J.; Sipiczki, M.; Bahler, J.; Fisher, R. P. *Mol. Biol. Cell* **2005**, *16*, 2734.
(170) 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.
(171) Wallenfang, M. R.; Seydoux, G. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5527.
(172) Ganuza, M.; Saiz-Ladera, C.; Canamero, M.; Gomez, G.; Schneider, R.; Blasco, M. A.; Pisano, D.; Paramio, J. M.; Santamaria, D.; Barbacid, M. *EMBO J.* **2012**, *31*, 2498.
(173) (a) Galbraith, M. D.; Donner, A. J.; Espinosa, J. M. *Transcription* **2010**, *1*, 4. (b) Conaway, R. C.; Conaway, J. W. *Curr. Opin. Genet. Dev.* **2011**, *21*, 225. (c) Malik, S.; Roeder, R. G. *Nat. Rev. Genet.* **2010**, *11*, 761.
(174) (a) Borggrefe, T.; Davis, R.; Erdjument-Bromage, H.; Tempst, P.; Kornberg, R. D. *J. Biol. Chem.* **2002**, *277*, 44202. (b) Wang, G.; Cantin, G. T.; Stevens, J. L.; Berk, A. J. *Mol. Cell. Biol.* **2001**, *21*, 4604. (c) Samuelsen, C. O.; Baraznenok, V.; Khorosjutina, O.; Spahr, H.; Kieselbach, T.; Holmberg, S.; Gustafsson, C. M. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 6422. (d) Leclerc, V.; Tassan, J. P.; O'Farrell, P. H.; Nigg, E. A.; Leopold, P. *Mol. Biol. Cell* **1996**, *7*, 505. (e) Larschan, E.; Winston, F. *Mol. Cell. Biol.* **2005**, *25*, 114. (f) Knuesel, M. T.; Meyer, K. D.; Bernecky, C.; Taatjes, D. J. *Genes Dev.* **2009**, *23*, 439.
(175) Knuesel, M. T.; Meyer, K. D.; Donner, A. J.; Espinosa, J. M.; Taatjes, D. J. *Mol. Cell. Biol.* **2009**, *29*, 650.
(176) Holstege, F. C.; Jennings, E. G.; Wyrick, J. J.; Lee, T. I.; Hengartner, C. J.; Green, M. R.; Golub, T. R.; Lander, E. S.; Young, R. A. *Cell* **1998**, *95*, 717.
(177) Liu, Y.; Kung, C.; Fishburn, J.; Ansari, A. Z.; Shokat, K. M.; Hahn, S. *Mol. Cell. Biol.* **2004**, *24*, 1721.
(178) (a) Marshall, N. F.; Price, D. H. *J. Biol. Chem.* **1995**, *270*, 12335. (b) 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. (c) Peng, J.; Marshall, N. F.; Price, D. H. *J. Biol. Chem.* **1998**, *273*, 13855. (d) Peng, J.; Zhu, Y.; Milton, J. T.; Price, D. H. *Genes Dev.* **1998**, *12*, 755. (e) Wei, P.; Garber, M. E.; Fang, S. M.; Fischer, W. H.; Jones, K. A. *Cell*

1998, *92*, 451. (f) Mancebo, H. S.; Lee, G.; Flygare, J.; Tomassini, J.; Luu, P.; Zhu, Y.; Peng, J.; Blau, C.; Hazuda, D.; Price, D.; Flores, O. *Genes Dev.* **1997**, *11*, 2633. (g) Peterlin, B. M.; Price, D. H. *Mol. Cell* **2006**, *23*, 297.

(179) (a) Prelich, G.; Winston, F. *Genetics* **1993**, *135*, 665. (b) Murray, S.; Udupa, R.; Yao, S.; Hartzog, G.; Prelich, G. *Mol. Cell. Biol.* **2001**, *21*, 4089.

(180) (a) Pei, Y.; Du, H.; Singer, J.; Stamour, C.; Granitto, S.; Shuman, S.; Fisher, R. P. *Mol. Cell. Biol.* **2006**, *26*, 777. (b) Viladevall, L.; St 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.

(181) Grana, X.; De Luca, A.; Sang, N.; Fu, Y.; Claudio, P. P.; Rosenblatt, J.; Morgan, D. O.; Giordano, A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 3834.

(182) (a) Shore, S. M.; Byers, S. A.; Dent, P.; Price, D. H. *Gene* **2005**, *350*, 51. (b) Shore, S. M.; Byers, S. A.; Maury, W.; Price, D. H. *Gene* **2003**, *307*, 175.

(183) St Amour, C. V.; Sanso, M.; Bosken, C. A.; Lee, K. M.; Larochelle, S.; Zhang, C.; Shokat, K. M.; Geyer, M.; Fisher, R. P. *Mol. Cell. Biol.* **2012**, *32*, 2372.

(184) 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.

(185) Qiu, H.; Hu, C.; Hinnebusch, A. G. *Mol. Cell* **2009**, *33*, 752.

(186) Lis, J. *Cold Spring Harb. Symp. Quant. Biol.* **1998**, *63*, 347.

(187) (a) Core, L. J.; Waterfall, J. J.; Lis, J. T. *Science* **2008**, *322*, 1845. (b) Min, I. M.; Waterfall, J. J.; Core, L. J.; Munroe, R. J.; Schimenti, J.; Lis, J. T. *Genes Dev.* **2011**, *25*, 742. (c) Nechaev, S.; Fargo, D. C.; dos Santos, G.; Liu, L.; Gao, Y.; Adelman, K. *Science* **2010**, *327*, 335. (d) Rahl, P. B.; Lin, C. Y.; Seila, A. C.; Flynn, R. A.; McCuine, S.; Burge, C. B.; Sharp, P. A.; Young, R. A. *Cell* **2010**, *141*, 432. (e) Zeitlinger, J.; Stark, A.; Kellis, M.; Hong, J. W.; Nechaev, S.; Adelman, K.; Levine, M.; Young, R. A. *Nat. Genet.* **2007**, *39*, 1512. (f) Guenther, M. G.; Levine, S. S.; Boyer, L. A.; Jaenisch, R.; Young, R. A. *Cell* **2007**, *130*, 77. (g) Kim, T. H.; Barrera, L. O.; Zheng, M.; Qu, C.; Singer, M. A.; Richmond, T. A.; Wu, Y.; Green, R. D.; Ren, B. *Nature* **2005**, *436*, 876.

(188) (a) Kephart, D. D.; Marshall, N. F.; Price, D. H. *Mol. Cell. Biol.* **1992**, *12*, 2067. (b) Wada, T.; Takagi, T.; Yamaguchi, Y.; Ferdous, A.; Imai, T.; Hirose, S.; Sugimoto, S.; Yano, K.; Hartzog, G. A.; Winston, F.; Buratowski, S.; Handa, H. *Genes Dev.* **1998**, *12*, 343. (c) Yamaguchi, Y.; Takagi, T.; Wada, T.; Yano, K.; Furuya, A.; Sugimoto, S.; Hasegawa, J.; Handa, H. *Cell* **1999**, *97*, 41. (d) Bartkowiak, B.; Greenleaf, A. L. *Transcription* **2011**, *2*, 115.

(189) Cheng, B.; Li, T.; Rahl, P. B.; Adamson, T. E.; Loudas, N. B.; Guo, J.; Varzavand, K.; Cooper, J. J.; Hu, X.; Gnatt, A.; Young, R. A.; Price, D. H. *Mol. Cell* **2012**, *45*, 38.

(190) (a) Eberhardy, S. R.; Farnham, P. J. *J. Biol. Chem.* **2002**, *277*, 40156. (b) Wu, S. Y.; Chiang, C. M. *J. Biol. Chem.* **2007**, *282*, 13141.

(191) Ni, Z.; Saunders, A.; Fuda, N. J.; Yao, J.; Suarez, J. R.; Webb, W. W.; Lis, J. T. *Mol. Cell. Biol.* **2008**, *28*, 1161.

(192) Devaiah, B. N.; Singer, D. S. *J. Biol. Chem.* **2012**, *287*, 38755.

(193) (a) Li, T.; Price, D. *Transcription* **2012**, *3*, 177. (b) 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.

(194) Steinmetz, E. J.; Warren, C. L.; Kuehner, J. N.; Panbehi, B.; Ansari, A. Z.; Brow, D. A. *Mol. Cell* **2006**, *24*, 735.

(195) (a) Lee, J. M.; Greenleaf, A. L. *Gene Expr.* **1991**, *1*, 149. (b) Sterner, D. E.; Lee, J. M.; Hardin, S. E.; Greenleaf, A. L. *Mol. Cell. Biol.* **1995**, *15*, 5716.

(196) Wood, A.; Shilatifard, A. *Cell Cycle* **2006**, *5*, 1066.

(197) (a) Liu, J.; Kipreos, E. T. *Mol. Biol. Evol.* **2000**, *17*, 1061. (b) Sukegawa, Y.; Yamashita, A.; Yamamoto, M. *PLoS Genet.* **2011**, *7*, e1002387.

(198) 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.

(199) (a) Chymkowitch, P.; Eldholm, V.; Lorenz, S.; Zimmermann, C.; Lindvall, J. M.; Bjoras, M.; Meza-Zepeda, L. A.; Enserink, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 10450. (b) Chymkowitch, P.; Enserink, J. M. *Transcription* **2012**, *4*, 3.

(200) (a) Cisek, L. J.; Corden, J. L. *Nature* **1989**, *339*, 679. (b) Cisek, L. J.; Corden, J. L. *Methods Enzymol.* **1991**, *200*, 301.

(201) Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. *Science* **2002**, *298*, 1912.

(202) Pearson, G.; Robinson, F.; Beers Gibson, T.; Xu, B. E.; Karandikar, M.; Berman, K.; Cobb, M. H. *Endocr. Rev.* **2001**, *22*, 153.

(203) Dubois, M. F.; Nguyen, V. T.; Dahmus, M. E.; Pages, G.; Pouyssegur, J.; Bensaude, O. *EMBO J.* **1994**, *13*, 4787.

(204) (a) Kong, X.; Shen, Y.; Jiang, N.; Fei, X.; Mi, J. *Cell Signal* **2011**, *23*, 1273. (b) Anderson, C. W.; Carter, T. H. *Curr. Top. Microbiol. Immunol.* **1996**, *217*, 91.

(205) Dvir, A.; Stein, L. Y.; Calore, B. L.; Dynan, W. S. *J. Biol. Chem.* **1993**, *268*, 10440.

(206) Tyagi, S.; Ochem, A.; Tyagi, M. *J. Gen. Virol.* **2011**, *92*, 1710.

(207) Maldonado, E.; Shiekhattar, R.; Sheldon, M.; Cho, H.; Drapkin, R.; Rickert, P.; Lees, E.; Anderson, C. W.; Linn, S.; Reinberg, D. *Nature* **1996**, *381*, 86.

(208) An, J.; Yang, T.; Huang, Y.; Liu, F.; Sun, J.; Wang, Y.; Xu, Q.; Wu, D.; Zhou, P. *BMC Biochem.* **2011**, *12*, 2.

(209) Peterson, S. R.; Jesch, S. A.; Chamberlin, T. N.; Dvir, A.; Rabindran, S. K.; Wu, C.; Dynan, W. S. *J. Biol. Chem.* **1995**, *270*, 1449.

(210) Devaiah, B. N.; Singer, D. S. *Transcription* **2012**, *4*, 13-7.

(211) Devaiah, B. N.; Lewis, B. A.; Cherman, N.; Hewitt, M. C.; Albrecht, B. K.; Robey, P. G.; Ozato, K.; Sims, R. J., 3rd; Singer, D. S. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 6927.

(212) de Carcer, G.; Manning, G.; Malumbres, M. *Cell Cycle* **2011**, *10*, 2255.

(213) Elia, A. E.; Rellos, P.; Haire, L. F.; Chao, J. W.; Ivins, F. J.; Hoepker, K.; Mohammad, D.; Cantley, L. C.; Smerdon, S. J.; Yaffe, M. B. *Cell* **2003**, *115*, 83.

(214) Johnson, E. F.; Stewart, K. D.; Woods, K. W.; Giranda, V. L.; Luo, Y. *Biochemistry* **2007**, *46*, 9551.

(215) Heidemann, M.; Eick, D. *RNA Biol.* **2012**, *9*, 1144-6.

(216) Wang, L.; Gao, J.; Dai, W.; Lu, L. *J. Biol. Chem.* **2008**, *283*, 25928.

(217) Sakurai, H.; Ishihama, A. *Genes Cells* **2002**, *7*, 273.

(218) Kuenzel, E. A.; Mulligan, J. A.; Sommercorn, J.; Krebs, E. G. *J. Biol. Chem.* **1987**, *262*, 9136.

(219) Sanso, M.; Fisher, R. P. *Transcription* **2013**, *4*.

(220) (a) Kelly, W. G.; Dahmus, M. E.; Hart, G. W. *J. Biol. Chem.* **1993**, *268*, 10416. (b) Haltiwanger, R. S.; Blomberg, M. A.; Hart, G. W. *J. Biol. Chem.* **1992**, *267*, 9005.

(221) Gao, Y.; Wells, L.; Comer, F. I.; Parker, G. J.; Hart, G. W. *J. Biol. Chem.* **2001**, *276*, 9838.

(222) Gross, B. J.; Kraybill, B. C.; Walker, S. *J. Am. Chem. Soc.* **2005**, *127*, 14588.

(223) Macauley, M. S.; Whitworth, G. E.; Debowski, A. W.; Chin, D.; Vocadlo, D. J. *J. Biol. Chem.* **2005**, *280*, 25313.

(224) Ranuncolo, S. M.; Ghosh, S.; Hanover, J. A.; Hart, G. W.; Lewis, B. A. *J. Biol. Chem.* **2012**, *287*, 23549.

(225) 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.

(226) Chambers, R. S.; Dahmus, M. E. *J. Biol. Chem.* **1994**, *269*, 26243.

(227) (a) 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. (b) Kimura, M.; Suzuki, H.; Ishihama, A. *Mol. Cell. Biol.* **2002**, *22*, 1577.

(228) (a) 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. (b) Archambault, J.; Pan, G.; Dahmus, G. K.; Cartier, M.; Marshall, N.; Zhang, S.; Dahmus, M. E.; Greenblatt, J. *J. Biol. Chem.* **1998**, *273*, 27593. (c) Cho, H.; Kim, T. K.; Mancebo, H.; Lane, W. S.; Flores, O.; Reinberg, D. *Genes Dev.* **1999**, *13*, 1540.

(229) Collet, J. F.; Stroobant, V.; Pirard, M.; Delpierre, G.; Van Schaftingen, E. *J. Biol. Chem.* **1998**, *273*, 14107.

(230) Yeo, M.; Lin, P. S.; Dahmus, M. E.; Gill, G. N. *J. Biol. Chem.* **2003**, *278*, 26078.
(231) Yeo, M.; Lee, S. K.; Lee, B.; Ruiz, E. C.; Pfaff, S. L.; Gill, G. N. *Science* **2005**, *307*, 596.
(232) (a) Kamenski, T.; Heilmeier, S.; Meinhart, A.; Cramer, P. *Mol. Cell* **2004**, *15*, 399. (b) 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.
(233) Hausmann, S.; Shuman, S. *J. Biol. Chem.* **2002**, *277*, 21213.
(234) (a) Ghosh, A.; Shuman, S.; Lima, C. D. *Mol. Cell* **2008**, *32*, 478. (b) Hausmann, S.; Koiwa, H.; Krishnamurthy, S.; Hampsey, M.; Shuman, S. *J. Biol. Chem.* **2005**, *280*, 37681.
(235) Zhang, M.; Wang, X. J.; Chen, X.; Bowman, M. E.; Luo, Y.; Noel, J. P.; Ellington, A. D.; Etzkorn, F. A.; Zhang, Y. *ACS Chem. Biol.* **2012**, *7*, 1462.
(236) Cho, E. J.; Kobor, M. S.; Kim, M.; Greenblatt, J.; Buratowski, S. *Genes Dev.* **2001**, *15*, 3319.
(237) (a) Chambers, R. S.; Wang, B. Q.; Burton, Z. F.; Dahmus, M. E. *J. Biol. Chem.* **1995**, *270*, 14962. (b) Suh, M. H.; Ye, P.; Zhang, M.; Hausmann, S.; Shuman, S.; Gnatt, A. L.; Fu, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17314.
(238) Bushnell, D. A.; Westover, K. D.; Davis, R. E.; Kornberg, R. D. *Science* **2004**, *303*, 983.
(239) (a) Wostenberg, C.; Kumar, S.; Noid, W. G.; Showalter, S. A. *J. Phys. Chem. B* **2011**, *115*, 13731. (b) Kobor, M. S.; Simon, L. D.; Omichinski, J.; Zhong, G.; Archambault, J.; Greenblatt, J. *Mol. Cell. Biol.* **2000**, *20*, 7438.
(240) Palancade, B.; Dubois, M. F.; Dahmus, M. E.; Bensaude, O. *Mol. Cell. Biol.* **2001**, *21*, 6359.
(241) 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.
(242) Sun, Z. W.; Hampsey, M. *Mol. Cell. Biol.* **1996**, *16*, 1557.
(243) (a) Ganem, C.; Devaux, F.; Torchet, C.; Jacq, C.; Quevillon-Cheruel, S.; Labesse, G.; Facca, C.; Faye, G. *EMBO J.* **2003**, *22*, 1588. (b) Meinhart, A.; Silberzahn, T.; Cramer, P. *J. Biol. Chem.* **2003**, *278*, 15917.
(244) (a) Wu, W. H.; Pinto, I.; Chen, B. S.; Hampsey, M. *Genetics* **1999**, *153*, 643. (b) Pappas, D. L., Jr.; Hampsey, M. *Mol. Cell. Biol.* **2000**, *20*, 8343. (c) Dichtl, B.; Blank, D.; Ohnacker, M.; Friedlein, A.; Roeder, D.; Langen, H.; Keller, W. *Mol. Cell* **2002**, *10*, 1139.
(245) (a) He, X.; Khan, A. U.; Cheng, H.; Pappas, D. L., Jr.; Hampsey, M.; Moore, C. L. *Genes Dev.* **2003**, *17*, 1030. (b) Nedea, E.; He, X.; Kim, M.; Pootoolal, J.; Zhong, G.; Canadien, V.; Hughes, T.; Buratowski, S.; Moore, C. L.; Greenblatt, J. *J. Biol. Chem.* **2003**, *278*, 33000. (c) Steinmetz, E. J.; Brow, D. A. *Mol. Cell. Biol.* **2003**, *23*, 6339.
(246) (a) Ansari, A.; Hampsey, M. *Genes Dev.* **2005**, *19*, 2969. (b) O'Sullivan, J. M.; Tan-Wong, S. M.; Morillon, A.; Lee, B.; Coles, J.; Mellor, J.; Proudfoot, N. J. *Nat. Genet.* **2004**, *36*, 1014. (c) Singh, B. N.; Hampsey, M. *Mol. Cell* **2007**, *27*, 806.
(247) Krishnamurthy, S.; He, X.; Reyes-Reyes, M.; Moore, C.; Hampsey, M. *Mol. Cell* **2004**, *14*, 387.
(248) Krishnamurthy, S.; Ghazy, M. A.; Moore, C.; Hampsey, M. *Mol. Cell. Biol.* **2009**, *29*, 2925.
(249) 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.
(250) (a) Gibney, P. A.; Fries, T.; Bailer, S. M.; Morano, K. A. *Eukaryot Cell* **2008**, *7*, 938. (b) Jeronimo, C.; Forget, D.; Bouchard, A.; Li, Q.; Chua, G.; Poitras, C.; Therien, C.; Bergeron, D.; Bourassa, S.; Greenblatt, J.; Chabot, B.; Poirier, G. G.; Hughes, T. R.; Blanchette, M.; Price, D. H.; Coulombe, B. *Mol. Cell* **2007**, *27*, 262.
(251) 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.
(252) Egloff, S.; Zaborowska, J.; Laitem, C.; Kiss, T.; Murphy, S. *Mol. Cell* **2012**, *45*, 111.
(253) Xiang, K.; Manley, J. L.; Tong, L. *Nat. Commun.* **2012**, *3*, 946.
(254) Schiene, C.; Fischer, G. *Curr. Opin. Struct. Biol.* **2000**, *10*, 40.
(255) Fanghanel, J.; Fischer, G. *Front Biosci.* **2004**, *9*, 3453.
(256) Albert, A.; Lavoie, S.; Vincent, M. *J. Cell Sci.* **1999**, *112* (Pt 15), 2493.
(257) Lu, K. P.; Hanes, S. D.; Hunter, T. *Nature* **1996**, *380*, 544.
(258) Hani, J.; Stumpf, G.; Domdey, H. *FEBS Lett.* **1995**, *365*, 198.
(259) Arevalo-Rodriguez, M.; Wu, X.; Hanes, S. D.; Heitman, J. *Front. Biosci.* **2004**, *9*, 2420.
(260) Yaffe, M. B.; Schutkowski, M.; Shen, M.; Zhou, X. Z.; Stukenberg, P. T.; Rahfeld, J. U.; Xu, J.; Kuang, J.; Kirschner, M. W.; Fischer, G.; Cantley, L. C.; Lu, K. P. *Science* **1997**, *278*, 1957.
(261) (a) Xu, Y. X.; Hirose, Y.; Zhou, X. Z.; Lu, K. P.; Manley, J. L. *Genes Dev.* **2003**, *17*, 2765. (b) Palancade, B.; Marshall, N. F.; Tremeau-Bravard, A.; Bensaude, O.; Dahmus, M. E.; Dubois, M. F. *J. Mol. Biol.* **2004**, *335*, 415.
(262) Xu, Y. X.; Manley, J. L. *Genes Dev.* **2007**, *21*, 2950.
(263) Hanes, S. D.; Shank, P. R.; Bostian, K. A. *Yeast* **1989**, *5*, 55.
(264) (a) Morris, D. P.; Phatnani, H. P.; Greenleaf, A. L. *J. Biol. Chem.* **1999**, *274*, 31583. (b) Wu, X.; Chang, A.; Sudol, M.; Hanes, S. D. *Curr. Genet.* **2001**, *40*, 234. (c) Wu, X.; Wilcox, C. B.; Devasahayam, G.; Hackett, R. L.; Arevalo-Rodriguez, M.; Cardenas, M. E.; Heitman, J.; Hanes, S. D. *EMBO J.* **2000**, *19*, 3727.
(265) (a) Lu, K. P.; Zhou, X. Z. *Nat. Rev. Mol. Cell. Biol.* **2007**, *8*, 904. (b) Yeh, E. S.; Means, A. R. *Nat. Rev. Cancer* **2007**, *7*, 381.
(266) Gemmill, T. R.; Wu, X.; Hanes, S. D. *J. Biol. Chem.* **2005**, *280*, 15510.
(267) Hani, J.; Schelbert, B.; Bernhardt, A.; Domdey, H.; Fischer, G.; Wiebauer, K.; Rahfeld, J. U. *J. Biol. Chem.* **1999**, *274*, 108.
(268) Ma, Z.; Atencio, D.; Barnes, C.; DeFiglio, H.; Hanes, S. D. *Mol. Cell. Biol.* **2012**, *32*, 3594.
(269) (a) Singh, N.; Ma, Z.; Gemmill, T.; Wu, X.; Defiglio, H.; Rossettini, A.; Rabeler, C.; Beane, O.; Morse, R. H.; Palumbo, M. J.; Hanes, S. D. *Mol. Cell* **2009**, *36*, 255. (b) Xiang, K.; Nagaike, T.; Xiang, S.; Kilic, T.; Beh, M. M.; Manley, J. L.; Tong, L. *Nature* **2010**, *467*, 729.
(270) Rempola, B.; Kaniak, A.; Migdalski, A.; Rytka, J.; Slonimski, P. P.; di Rago, J. P. *Mol. Gen. Genet.* **2000**, *262*, 1081.
(271) Jordens, J.; Janssens, V.; Longin, S.; Stevens, I.; Martens, E.; Bultynck, G.; Engelborghs, Y.; Lescrinier, E.; Waelkens, E.; Goris, J.; Van Hoof, C. *J. Biol. Chem.* **2006**, *281*, 6349.
(272) (a) Douville, J.; David, J.; Lemieux, K. M.; Gaudreau, L.; Ramotar, D. *Genetics* **2006**, *172*, 1369. (b) Heitman, J.; Movva, N. R.; Hall, M. N. *Science* **1991**, *253*, 905. (c) Powers, T.; Walter, P. *Mol. Biol. Cell* **1999**, *10*, 987. (d) Rohde, J.; Heitman, J.; Cardenas, M. E. *J. Biol. Chem.* **2001**, *276*, 9583.
(273) Jouvet, N.; Poschmann, J.; Douville, J.; Bulet, L.; Ramotar, D. *BMC Mol. Biol.* **2010**, *11*, 92.
(274) (a) Douville, J.; David, J.; Fortier, P. K.; Ramotar, D. *Curr. Genet.* **2004**, *46*, 72. (b) Mitchell, D. A.; Sprague, G. F., Jr. *Mol. Cell. Biol.* **2001**, *21*, 488. (c) Van Hoof, C.; Martens, E.; Longin, S.; Jordens, J.; Stevens, I.; Janssens, V.; Goris, J. *Biochem. J.* **2005**, *386*, 93.
(275) Poschmann, J.; Drouin, S.; Jacques, P. E.; El Fadili, K.; Newmarch, M.; Robert, F.; Ramotar, D. *PLoS One* **2011**, *6*, e23159.
(276) Kaneko, S.; Manley, J. L. *Mol. Cell* **2005**, *20*, 91.
(277) Phatnani, H. P.; Greenleaf, A. L. *Methods Mol. Biol.* **2004**, *257*, 17.
(278) (a) Dahmus, M. E. *J. Biol. Chem.* **1996**, *271*, 19009. (b) Svejstrup, J. Q. *Biochim. Biophys. Acta* **2004**, *1677*, 64.
(279) (a) Kornberg, R. D. *Trends Biochem. Sci.* **2005**, *30*, 235. (b) Myers, L. C.; Kornberg, R. D. *Annu. Rev. Biochem.* **2000**, *69*, 729.
(280) Svejstrup, J. Q.; Li, Y.; Fellows, J.; Gnatt, A.; Bjorklund, S.; Kornberg, R. D. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 6075.
(281) Scafe, C.; Nonet, M.; Young, R. A. *Mol. Cell. Biol.* **1990**, *10*, 1010.
(282) 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.
(283) (a) Yudkovsky, N.; Ranish, J. A.; Hahn, S. *Nature* **2000**, *408*, 225. (b) Sogaard, T. M.; Svejstrup, J. Q. *J. Biol. Chem.* **2007**, *282*, 14113.
(284) (a) Asturias, F. J.; Jiang, Y. W.; Myers, L. C.; Gustafsson, C. M.; Kornberg, R. D. *Science* **1999**, *283*, 985. (b) Davis, J. A.; Takagi, Y.; Kornberg, R. D.; Asturias, F. A. *Mol. Cell* **2002**, *10*, 409.
(285) Takagi, Y.; Kornberg, R. D. *J. Biol. Chem.* **2006**, *281*, 80.

(286) Naar, A. M.; Taatjes, D. J.; Zhai, W.; Nogales, E.; Tjian, R. *Genes Dev.* **2002**, *16*, 1339.
(287) Hagler, J.; Shuman, S. *Science* **1992**, *255*, 983.
(288) Ghosh, A.; Lima, C. D. *Wiley Interdiscip. Rev. RNA* **2010**, *1*, 152.
(289) Fabrega, C.; Shen, V.; Shuman, S.; Lima, C. D. *Mol. Cell* **2003**, *11*, 1549.
(290) Ghosh, A.; Shuman, S.; Lima, C. D. *Mol. Cell* **2011**, *43*, 299.
(291) Pei, Y.; Hausmann, S.; Ho, C. K.; Schwer, B.; Shuman, S. *J. Biol. Chem.* **2001**, *276*, 28075.
(292) (a) Andre, B.; Springael, J. Y. *Biochem. Biophys. Res. Commun.* **1994**, *205*, 1201. (b) Sudol, M.; Chen, H. I.; Bougeret, C.; Einbond, A.; Bork, P. *FEBS Lett.* **1995**, *369*, 67.
(293) (a) Chen, H. I.; Sudol, M. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 7819. (b) Sudol, M. *Prog. Biophys. Mol. Biol.* **1996**, *65*, 113.
(294) Myers, J. K.; Morris, D. P.; Greenleaf, A. L.; Oas, T. G. *Biochemistry* **2001**, *40*, 8479.
(295) (a) Ranganathan, R.; Lu, K. P.; Hunter, T.; Noel, J. P. *Cell* **1997**, *89*, 875. (b) Yaffe, M. B.; Rittinger, K.; Volinia, S.; Caron, P. R.; Aitken, A.; Leffers, H.; Gamblin, S. J.; Smerdon, S. J.; Cantley, L. C. *Cell* **1997**, *91*, 961.
(296) (a) Lu, P. J.; Zhou, X. Z.; Shen, M.; Lu, K. P. *Science* **1999**, *283*, 1325. (b) Verdecia, M. A.; Bowman, M. E.; Lu, K. P.; Hunter, T.; Noel, J. P. *Nat. Struct. Biol.* **2000**, *7*, 639.
(297) (a) Lohi, O.; Poussu, A.; Mao, Y.; Quiocho, F.; Lehto, V. P. *FEBS Lett.* **2002**, *513*, 19. (b) Misra, S.; Puertollano, R.; Kato, Y.; Bonifacino, J. S.; Hurley, J. H. *Nature* **2002**, *415*, 933.
(298) 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.
(299) (a) Steinmetz, E. J.; Brow, D. A. *Mol. Cell. Biol.* **1996**, *16*, 6993. (b) Conrad, N. K.; Wilson, S. M.; Steinmetz, E. J.; Patturajan, M.; Brow, D. A.; Swanson, M. S.; Corden, J. L. *Genetics* **2000**, *154*, 557.
(300) Sadowski, M.; Dichtl, B.; Hubner, W.; Keller, W. *EMBO J.* **2003**, *22*, 2167.
(301) 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.
(302) Gross, S.; Moore, C. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 6080.
(303) (a) Hollingworth, D.; Noble, C. G.; Taylor, I. A.; Ramos, A. *RNA* **2006**, *12*, 555. (b) Zhang, Z.; Fu, J.; Gilmour, D. S. *Genes Dev.* **2005**, *19*, 1572.
(304) Fuxreiter, M.; Tompa, P. *Adv. Exp. Med. Biol.* **2012**, *725*, 1.
(305) Vasiljeva, L.; Buratowski, S. *Mol. Cell* **2006**, *21*, 239.
(306) Steinmetz, E. J.; Brow, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 6699.
(307) Kubicek, K.; Cerna, H.; Holub, P.; Pasulka, J.; Hrossova, D.; Loehr, F.; Hofr, C.; Vanacova, S.; Stefl, R. *Genes Dev.* **2012**, *26*, 1891.
(308) Scholes, D. T.; Banerjee, M.; Bowen, B.; Curcio, M. J. *Genetics* **2001**, *159*, 1449.
(309) Liu, B. A.; Nash, P. D. *Philos. Trans. R. Soc. London B Biol. Sci.* **2012**, *367*, 2556.
(310) (a) Gerber, M.; Shilatifard, A. *J. Biol. Chem.* **2003**, *278*, 26303. (b) Hampsey, M.; Reinberg, D. *Cell* **2003**, *113*, 429.
(311) Li, B.; Howe, L.; Anderson, S.; Yates, J. R., 3rd; Workman, J. L. *J. Biol. Chem.* **2003**, *278*, 8897.
(312) Vojnic, E.; Simon, B.; Strahl, B. D.; Sattler, M.; Cramer, P. *J. Biol. Chem.* **2006**, *281*, 13.
(313) Gross, C. A.; Chan, C.; Dombroski, A.; Gruber, T.; Sharp, M.; Tupy, J.; Young, B. *Cold Spring Harb. Symp. Quant. Biol.* **1998**, *63*, 141.
(314) Bedford, M. T.; Leder, P. *Trends Biochem. Sci.* **1999**, *24*, 264.
(315) Allen, M.; Friedler, A.; Schon, O.; Bycroft, M. *J. Mol. Biol.* **2002**, *323*, 411.
(316) Carty, S. M.; Goldstrohm, A. C.; Sune, C.; Garcia-Blanco, M. A.; Greenleaf, A. L. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9015.
(317) (a) Sune, C.; Garcia-Blanco, M. A. *Mol. Cell. Biol.* **1999**, *19*, 4719. (b) Sune, C.; Hayashi, T.; Liu, Y.; Lane, W. S.; Young, R. A.; Garcia-Blanco, M. A. *Mol. Cell. Biol.* **1997**, *17*, 6029.
(318) Liu, J.; Fan, S.; Lee, C. J.; Greenleaf, A. L.; Zhou, P. *J. Biol. Chem.* **2013**.
(319) Schwartz, J. C.; Ebmeier, C. C.; Podell, E. R.; Heimiller, J.; Taatjes, D. J.; Cech, T. R. *Genes Dev.* **2012**, *26*, 2690.
(320) (a) Kovar, H. *Sarcoma* **2011**, *2011*, 837474. (b) Mackenzie, I. R.; Neumann, M. *Brain Res.* **2012**, *1462*, 40.
(321) Bertolotti, A.; Lutz, Y.; Heard, D. J.; Chambon, P.; Tora, L. *EMBO J.* **1996**, *15*, 5022.
(322) (a) Chen, D.; Toone, W. M.; Mata, J.; Lyne, R.; Burns, G.; Kivinen, K.; Brazma, A.; Jones, N.; Bahler, J. *Mol. Biol. Cell* **2003**, *14*, 214. (b) Gasch, A. P.; Spellman, P. T.; Kao, C. M.; Carmel-Harel, O.; Eisen, M. B.; Storz, G.; Botstein, D.; Brown, P. O. *Mol. Biol. Cell* **2000**, *11*, 4241. (c) Morimoto, R. I. *Genes Dev.* **2008**, *22*, 1427.
(323) (a) Dubois, M. F.; Bensaude, O.; Morange, M. *CR Acad. Sci. III* **1991**, *313*, 165. (b) Dubois, M. F.; Bellier, S.; Seo, S. J.; Bensaude, O. *J. Cell. Physiol.* **1994**, *158*, 417. (c) Venetianer, A.; Dubois, M. F.; Nguyen, V. T.; Bellier, S.; Seo, S. J.; Bensaude, O. *Eur. J. Biochem.* **1995**, *233*, 83.
(324) Legagneux, V.; Morange, M.; Bensaude, O. *Eur. J. Biochem.* **1990**, *193*, 121.
(325) Dubois, M. F.; Marshall, N. F.; Nguyen, V. T.; Dahmus, G. K.; Bonnet, F.; Dahmus, M. E.; Bensaude, O. *Nucleic Acids Res.* **1999**, *27*, 1338.
(326) Patturajan, M.; Conrad, N. K.; Bregman, D. B.; Corden, J. L. *J. Biol. Chem.* **1999**, *274*, 27823.
(327) (a) Herman, P. K. *Curr. Opin. Microbiol.* **2002**, *5*, 602. (b) DeRisi, J. L.; Iyer, V. R.; Brown, P. O. *Science* **1997**, *278*, 680.
(328) (a) Jeong, S. J.; Kim, H. J.; Yang, Y. J.; Seol, J. H.; Jung, B. Y.; Han, J. W.; Lee, H. W.; Cho, E. J. *J. Microbiol.* **2005**, *43*, 516. (b) Ostapenko, D.; Solomon, M. J. *Eukaryot Cell* **2003**, *2*, 274.
(329) (a) Lee, D.; Lis, J. T. *Nature* **1998**, *393*, 389. (b) McNeil, J. B.; Agah, H.; Bentley, D. *Genes Dev.* **1998**, *12*, 2510.
(330) Kouzine, F.; Wojtowicz, D.; Yamane, A.; Resch, W.; Kieffer-Kwon, K. R.; Bandle, R.; Nelson, S.; Nakahashi, H.; Awasthi, P.; Feigenbaum, L.; Menoni, H.; Hoeijmakers, J.; Vermeulen, W.; Ge, H.; Przytycka, T. M.; Levens, D.; Casellas, R. *Cell* **2013**, *153*, 988.
(331) (a) Latham, K. E.; Solter, D.; Schultz, R. M. *Dev. Biol.* **1992**, *149*, 457. (b) Roeder, R. G. *J. Biol. Chem.* **1974**, *249*, 249. (c) Toyoda, T.; Wolffe, A. P. *Dev. Biol.* **1992**, *153*, 150.
(332) (a) Bellier, S.; Dubois, M. F.; Nishida, E.; Almouzni, G.; Bensaude, O. *Mol. Cell. Biol.* **1997**, *17*, 1434. (b) Doyle, O.; Corden, J. L.; Murphy, C.; Gall, J. G. *J. Struct. Biol.* **2002**, *140*, 154.
(333) Bellier, S.; Chastant, S.; Adenot, P.; Vincent, M.; Renard, J. P.; Bensaude, O. *EMBO J.* **1997**, *16*, 6250.
(334) Nakamura, A.; Seydoux, G. *Development* **2008**, *135*, 3817.
(335) Seydoux, G.; Dunn, M. A. *Development* **1997**, *124*, 2191.
(336) Batchelder, C.; Dunn, M. A.; Choy, B.; Suh, Y.; Cassie, C.; Shim, E. Y.; Shin, T. H.; Mello, C.; Seydoux, G.; Blackwell, T. K. *Genes Dev.* **1999**, *13*, 202.
(337) Zhang, F.; Barboric, M.; Blackwell, T. K.; Peterlin, B. M. *Genes Dev.* **2003**, *17*, 748.
(338) Ghosh, D.; Seydoux, G. *Genetics* **2008**, *178*, 235.
(339) Bowman, E. A.; Bowman, C. R.; Ahn, J. H.; Kelly, W. G. *Development* **2013**, in press.
(340) (a) Hanyu-Nakamura, K.; Sonobe-Nojima, H.; Tanigawa, A.; Lasko, P.; Nakamura, A. *Nature* **2008**, *451*, 730. (b) Deshpande, G.; Calhoun, G.; Schedl, P. *Development* **2004**, *131*, 1247. (c) Martinho, R. G.; Kunwar, P. S.; Casanova, J.; Lehmann, R. *Curr. Biol.* **2004**, *14*, 159.
(341) Tam, P. P.; Zhou, S. X. *Dev. Biol.* **1996**, *178*, 124.
(342) Seki, Y.; Yamaji, M.; Yabuta, Y.; Sano, M.; Shigeta, M.; Matsui, Y.; Saga, Y.; Tachibana, M.; Shinkai, Y.; Saitou, M. *Development* **2007**, *134*, 2627.
(343) (a) Jackson, D. A.; Hassan, A. B.; Errington, R. J.; Cook, P. R. *EMBO J.* **1993**, *12*, 1059. (b) Chakalova, L.; Fraser, P. *Cold Spring Harb. Perspect. Biol.* **2010**, *2*, a000729. (c) Cook, P. R. *J. Mol. Biol.* **2010**, *395*, 1. (d) Melnik, S.; Deng, B.; Papantonis, A.; Baboo, S.; Carr, I. M.; Cook, P. R. *Nat. Methods* **2011**, *8*, 963. (e) Rieder, D.; Trajanoski, Z.; McNally, J. G. *Front. Genet.* **2012**, *3*, 221.
(344) Jackson, D. A.; Iborra, F. J.; Manders, E. M.; Cook, P. R. *Mol. Biol. Cell* **1998**, *9*, 1523.

(345) (a) Jackson, D. A.; Cook, P. R. *EMBO J.* **1985**, *4*, 919. (b) Kimura, H.; Tao, Y.; Roeder, R. G.; Cook, P. R. *Mol. Cell. Biol.* **1999**, *19*, 5383.
(346) Ferrai, C.; Xie, S. Q.; Luraghi, P.; Munari, D.; Ramirez, F.; Branco, M. R.; Pombo, A.; Crippa, M. P. *PLoS Biol.* **2010**, *8*, e1000270.
(347) 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.
(348) (a) Mintz, P. J.; Patterson, S. D.; Neuwald, A. F.; Spahr, C. S.; Spector, D. L. *EMBO J.* **1999**, *18*, 4308. (b) Xie, S. Q.; Martin, S.; Guillot, P. V.; Bentley, D. L.; Pombo, A. *Mol. Biol. Cell* **2006**, *17*, 1723.
(349) Xie, S. Q.; Pombo, A. *Histochem. Cell Biol.* **2006**, *125*, 21.
(350) Spector, D. L.; Lamond, A. I. *Cold Spring Harb Perspect Biol* **2011**, *3*, a000646.
(351) Gall, J. G. *Nat. Rev. Mol. Cell. Biol.* **2003**, *4*, 975.
(352) (a) Gall, J. G. *FEBS Lett.* **2001**, *498*, 164. (b) Kiss, T. *J. Cell Sci.* **2004**, *117*, 5949.
(353) Cho, E. J.; Rodriguez, C. R.; Takagi, T.; Buratowski, S. *Genes Dev.* **1998**, *12*, 3482.
(354) Pontius, B. W. *Trends Biochem. Sci.* **1993**, *18*, 181.