Research review paper

Modifications of RNA polymerase II CTD: Connections to the histone code and cellular function

Rakesh Srivastava, Seong Hoon Ahn *

Division of Molecular and Life Sciences, College of Science and Technology, Hanyang University, Ansan, Republic of Korea

**Abstract**

At the onset of transcription, many protein machineries interpret the cellular signals that regulate gene expression. These complex signals are mostly transmitted to the indispensable primary proteins involved in transcription, RNA polymerase II (RNAPII) and histones. RNAPII and histones are so well coordinated in this cellular function that each cellular signal is precisely allocated to specific machinery depending on the stage of transcription. The carboxy-terminal domain (CTD) of RNAPII in eukaryotes undergoes extensive posttranslational modification, called the ‘CTD code’, that is indispensable for coupling transcription with many cellular processes, including mRNA processing. The posttranslational modification of histones, known as the ‘histone code’, is also critical for gene transcription through the reversible and dynamic remodeling of chromatin structure. Notably, the histone code is closely linked with the CTD code, and their combinatorial effects enable the delicate regulation of gene transcription. This review elucidates recent findings regarding the CTD modifications of RNAPII and their coordination with the histone code, providing integrative pathways for the fine-tuned regulation of gene expression and cellular function.

© 2015 Elsevier Inc. All rights reserved.

**Contents**

1. Introduction......................................................................................................................... 857
2. The RNA polymerase II CTD: structure, modification and signature code ...................... 857
3. Phosphorylation of the CTD ............................................................................................... 858
   3.1. CTD serine phosphorylation ....................................................................................... 858
   3.2. CTD tyrosine phosphorylation .................................................................................... 859
   3.3. CTD threonine phosphorylation .................................................................................. 860
4. Other modifications to the CTD .......................................................................................... 861
   4.1. CTD glycosylation ....................................................................................................... 861
   4.2. CTD proline isomerization ......................................................................................... 861
   4.3. CTD acetylation .......................................................................................................... 861
   4.4. CTD methylation ......................................................................................................... 861
   4.5. CTD ubiquitylation ...................................................................................................... 861
5. Encoders, readers, and decoders of the CTD code signature ............................................... 863
6. Apprehending the synchronized function of the CTD and the histone code during transcription ................................................................. 863
   6.1. RNAPII CTD and histone acetylation ........................................................................ 863
   6.2. RNAPII CTD and histone methylation ...................................................................... 864
   6.3. RNAPII CTD and histone ubiquitylation .................................................................... 864
   6.4. RNAPII CTD and histone phosphorylation ............................................................... 865
7. Application of RNAPII CTD to therapeutic targets for various diseases ............................ 866
8. Perspectives and conclusions for the synchronized code .................................................... 868
Acknowledgements .................................................................................................................. 868
Appendix A. Supplementary data ............................................................................................ 868
References ................................................................................................................................ 868

* Corresponding author at: Division of Molecular and Life Sciences, College of Science and Technology, Hanyang University, 1271 Sa 3-dong, Sangnok-gu, Ansan, Gyeonggi-do 426-791, Republic of Korea.
E-mail address: hoon320@hanyang.ac.kr (S.H. Ahn).

http://dx.doi.org/10.1016/j.biotechadv.2015.07.008
0734-9750/© 2015 Elsevier Inc. All rights reserved.
1. Introduction

Gene expression in eukaryotic cells is dynamically regulated and contains multiple pathways, with each biological process being tightly controlled and organized. This complicated regulation entails the coordinated stimulation of gene promoters and enhancers, as well as co-activators, transcription factors, and many other regulatory proteins. RNA polymerase (RNAP) is one of the most essential multi-subunit holoenzymes for gene transcription and is conserved throughout all eukaryotes and prokaryotes. In most eukaryotes, three RNAPs control all gene transcription. RNAPI transcribes the genes encoding the 5.8S, 18S, and 25S (28S in mammals) ribosomal RNAs (rRNAs) (Richard and Manley, 2009). RNAPII transcribes transfer RNAs (tRNAs), the 5S rRNA, and the 6S small nuclear RNA (snRNA) (Diec et al., 2007). The RNAPII holoenzyme transcribes all protein-coding genes, as well as other various types of RNAs, including snRNA, small nuclear RNA (snRNA), microRNA, long noncoding RNA, cryptic unstable transcripts (PUTs), stable unannotated transcripts (SUTs) and Xrn1-dependent unstable transcripts (XUTs) (Diec et al., 2007; Neil et al., 2009; van Dijk et al., 2011). RNAPI and RNAPII jointly account for 90% of transcriptional throughput, and RNAPII transcription produces the remaining 10% (Rudra and Warner, 2004). Among the RNAPs, the gene-transcribing RNAPII (550 kDa) was first isolated and purified from Saccharomyces cerevisiae (Frederick et al., 1969) and was found to comprise 10–12 subunits in eukaryotes (12 subunits in yeast and humans) (Acker et al., 1997; Edwards et al., 1991). In particular, RNAPII has an unusual repetitive disordered domain at the carboxy-terminal domain (CTD), which extends from the catalytic core of the enzyme (Cramer et al., 2001). Notably, these unusual CTD repeats are found in the largest RNAPII subunit in all eukaryotes.

Accumulating evidence reveals that the RNAPII CTD is dynamically modified throughout the transcription cycle, permitting the association or dissociation of various protein complexes (Gosh et al., 2011; Lima, 2005; Lunde et al., 2010). Although the CTD is unnecessary for RNAPII catalytic activity, it is essential for cell viability and functions as a scaffold to orchestrate the assembly of numerous protein complexes. Indeed, RNAPII CTD modification plays pivotal roles in the progression of the transcription cycle and in mRNA processing, including 5′ capping, splicing, or 3′ cleavage and polyadenylation (poly(A)). Additionally, recent reports have revealed that RNAPII CTD modification is closely linked to histone modification and to other cellular processes such as DNA damage and repair (Corden, 2013; Hsin and Manley, 2012).

Histone posttranslational modification, described as the ‘histone code’, was first proposed by Brian D. Strahl and C. David Allis (Strahl and Allis, 2000). Different chemical modifications occur on various histone amino acids, which primarily include acetylation, ADP ribosylation, butyrylation, β-N-acetylglucosaminylatation, crotonylation, citrullination, formylation, hydroxylation, hydroxyisobutyrylation, methylation, malonylation, proline isomerization, propionylation, phosphorylation, sumoylation, succinylation or ubiquitylation (Huang et al., 2014). The recent identification of 130 different posttranslational modification sites on human histones further expands their understood functions and emphasizes the complexities of chromatin-associated pathways (Tan et al., 2011). Histone modifications most frequently occur within the histone amino- or carboxy-terminal tails, which protrude from the surface of the nucleosome, but are also found in the globular core regions (Cosgrove et al., 2004). Although these dynamic and reversible histone modifications significantly affect chromosome structure and gene regulation, their co-transcriptional connections with the RNAPII CTD are just beginning to be elucidated. Here, we highlight the combinatorial control of gene regulation by RNAPII CTD modifications and histone modifications, along with their synchronized function in the regulation of transcription. This review also provides insight into the cellular functions of CTD modifications under disease or stress conditions.

2. The RNA polymerase II CTD: structure, modification and signature code

The unique CTD structure of the largest subunit of RNAPII, which was first revealed in 1985 from mouse Rpb1 and budding yeast Rpb1, is largely conserved from fungi to humans. The RNAPII CTD in eukaryotes mostly comprises multiple tandemly repeated heptadepptides with the consensus sequence Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7 (Allison et al., 1985; Corden et al., 1985). However, the total number of CTD repeats and the numbers of consensus repeats vary in many organisms; in particular, Ostreococcus lucimarinus and Micromonas sp. RCC299 have only non-consensus CTD sequences (Liu et al., 2010; Stump and Ostrozhynska, 2013). Additionally, the number of CTD repeats varies from organism to organism, and the variation in total CTD repeats is generally correlated with the degree of organismal complexity (Table 1).

The RNAPII CTD contains three parts: the flexible linker region, the region comprising the mostly conserved heptad repeats with consensus or non-consensus sequences and the tip region at the end of the CTD repeats (Fig. 1A and B). The length of the Rpb1 CTD in budding yeast reaches 650 Å in extended conformations, and this CTD is positioned near the mRNA exit tunnel, where interactions occur between the CTD and many proteins associated with the nascent mRNA (Cramer et al., 2001). The CTD repeats are highly disordered, with a dynamic plasticity to form potential secondary structures, mainly due to the presence of a high percentage of hydrophobic and polar amino acids. Nuclear magnetic resonance (NMR) spectroscopy reveals that β-turns at two SPXX motifs (i.e., S5P7T5S6 and S9P17T16) of the CTD are formed and stabilized by hydrogen bonds (Meinhart et al., 2005; Suzuki, 1989, 1990). NMR and circular dichroism analyses further show that the CTD S2-P leads to a less ordered backbone conformation or alterations in the β-turn conformation in the hydroxide-bond-promoting solvent trifluoroethanol (Bienkiewicz et al., 2000). Similarly, glycosylation of CTD T4 leads to the formation of a turn-like structure (Simanek et al., 1998). Indeed, iteration of the CTD peptide structure is thought to result in a compact β-spiral model, but during the mRNA transcription-processing cycle, these compact spiral regions are unraveled and regenerated in a phosphorylation-dependent manner (Meinhart and Cramer, 2004).

Because of the structural flexibility of the CTD and the variety of binding surfaces generated by extensive posttranslational modifications on the heptadepptide repeat sequences, the RNAPII CTD has the ability to interact with and recruit distinct proteins at various stages of transcription (Lunde et al., 2010; Meinhart et al., 2005; Noble et al., 2005). The modifications of the CTD repeats primarily include phosphorylation, isomerization between cis/trans conformations and glycosylation during the transcription cycle and mRNA processing (Egloff et al., 2012a). All serine, tyrosine and threonine residues within the heptad CTD are phosphorylated. CTD glycosylation occurs on threonine and serine residues, whereas the cyclic amino acid proline is subject to isomerization (Fig. 1C). In addition to these conventional modifications of the RNAPII CTD consensus sequence, recent investigations report other modifications of non-consensus sequences, including methylation, acetylation and ubiquitylation (Dauny et al., 2008; Schroder et al., 2013; Sims et al., 2011). Analyses of the frequency and conservation of amino acids within the consensus and non-consensus sequences of the RNAPII CTD further show that most variation occurs at the 4th and 7th positions in yeast and humans, implying a more complicated repertoire of CTD modifications (Fig. 1D). Although the distribution or stoichiometry of individual modifications within the RNAPII CTD repeats is not known, the complex, combinatorial landscape of modification patterns, which are read by CTD-interacting proteins to recruit other co-activators or co-repressors during transcription, leads to the concept of the ‘CTD code’ (Buratowski, 2003). Together, the existence of numerous and complicated CTD modifications of consensus or non-consensus sequences suggests that the CTD code is used for a broad range of transcriptional regulation.

3. Phosphorylation of the CTD

The side chain hydroxyl groups of serine at positions 2 (S2), 5 (S5), and 7 (S7), tyrosine at position 1 (Y1), and threonine at position 4 (T4) of the RNAPII CTD are phosphorylated. CTD phosphorylation is dynamically regulated at various stages of transcription and is closely linked to the recruitment of several transcriptional complexes (hereafter, see Fig. 2 and Supplementary Table S1 for the functions of CTD modifications during transcription).

3.1. CTD serine phosphorylation

During the early events of the transcription cycle, unphosphorylated RNAPII, general transcription factors (GTFs) and a mediator complex are recruited onto the gene promoter (Thomas and Chiang, 2006). The mediator complex stimulates TFIIH-associated kinase activity (CDK7 in mammals and Kin28 in S. cerevisiae). CDK7/Kin28, along with its cyclin, then phosphorylates S5 and S7 of the RNAPII CTD (Egloff et al., 2007; Kim et al., 2009; Rodriguez et al., 2000). Another kinase, CDK8

Table 1

<table>
<thead>
<tr>
<th>Group</th>
<th>Organisms</th>
<th>Species</th>
<th>CTD repeats</th>
<th>Consensus repeats</th>
<th>Non-consensus repeats†</th>
<th>Protein reference sequence*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower eukaryotes</td>
<td>Amoebozoa</td>
<td>Dictyostelium discoideum</td>
<td>23</td>
<td>18</td>
<td>5</td>
<td>XP_641735</td>
</tr>
<tr>
<td>Fungi</td>
<td>Saccharomyces cerevisiae</td>
<td>26</td>
<td>19</td>
<td>7</td>
<td>NP_010141</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Schizosaccharomyces pombe</td>
<td>29</td>
<td>24</td>
<td>5</td>
<td>NP_595673</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Schizosaccharomyces japonicus</td>
<td>29</td>
<td>22</td>
<td>7</td>
<td>XP_002172959</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Candida albicans</td>
<td>25</td>
<td>12</td>
<td>10</td>
<td>XP_719414</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ashbya gossypii</td>
<td>27</td>
<td>23</td>
<td>4</td>
<td>NP_584182</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Physcomitrella patens</td>
<td>29</td>
<td>13</td>
<td>16</td>
<td>XP_001785319</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ostreococcus lucimus</td>
<td>16</td>
<td>0</td>
<td>16</td>
<td>XP_001419944</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Micromonas sp. RCC299</td>
<td>24</td>
<td>0</td>
<td>24</td>
<td>XP_002503529</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Arabidopsis thaliana</td>
<td>34</td>
<td>15</td>
<td>19</td>
<td>NP_195305</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Oroza sativa</td>
<td>29</td>
<td>12</td>
<td>17</td>
<td>NP_001054670</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sorghum bicolor</td>
<td>28</td>
<td>5</td>
<td>23</td>
<td>XP_002450841</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vitis vinifera</td>
<td>34</td>
<td>17</td>
<td>19</td>
<td>XP_002285900</td>
<td></td>
</tr>
<tr>
<td>Higher eukaryotes</td>
<td>Plant</td>
<td>Brachioles florales</td>
<td>43</td>
<td>20</td>
<td>23</td>
<td>XP_002210199</td>
</tr>
<tr>
<td></td>
<td>Caenorhabditis elegans</td>
<td>35</td>
<td>10</td>
<td>25</td>
<td>NP_500523</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Drosophila melanogaster</td>
<td>44</td>
<td>2</td>
<td>42</td>
<td>NP_511124</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anopheles gambiæ</td>
<td>35</td>
<td>9</td>
<td>26</td>
<td>XP_317690</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Danio rerio</td>
<td>52</td>
<td>22</td>
<td>30</td>
<td>XP_682682</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Homo sapiens</td>
<td>52</td>
<td>21</td>
<td>31</td>
<td>NP_009328</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mus musculus</td>
<td>52</td>
<td>21</td>
<td>31</td>
<td>NP_033115</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rattus norvegicus</td>
<td>52</td>
<td>21</td>
<td>31</td>
<td>XP_343923</td>
<td></td>
</tr>
</tbody>
</table>

† Non-consensus sequences: The CTD repeats containing variations in any amino acid residue of the YSPTSPS heptad repeats.
* Consensus and non-consensus repeat sequences were calculated using NCBI GenBank accession numbers for protein sequences (Liu et al., 2010; Stump and Ostrozhynska, 2013).

3. Phosphorylation of the CTD

The side chain hydroxyl groups of serine at positions 2 (S2), 5 (S5), and 7 (S7), tyrosine at position 1 (Y1), and threonine at position 4 (T4) of the RNAPII CTD are phosphorylated. CTD phosphorylation is dynamically regulated at various stages of transcription and is closely linked to the recruitment of several transcriptional complexes (hereafter, see Fig. 2 and Supplementary Table S1 for the functions of CTD modifications during transcription).

3.1. CTD serine phosphorylation

During the early events of the transcription cycle, unphosphorylated RNAPII, general transcription factors (GTFs) and a mediator complex are recruited onto the gene promoter (Thomas and Chiang, 2006). The mediator complex stimulates TFIIH-associated kinase activity (CDK7 in mammals and Kin28 in S. cerevisiae). CDK7/Kin28, along with its cyclin, then phosphorylates S5 and S7 of the RNAPII CTD (Egloff et al., 2007; Kim et al., 2009; Rodriguez et al., 2000). Another kinase, CDK8

Fig. 1. The structure of the RNAPII CTD and its modification signature. a) A systematic representation of RNAPII showing different regions of the CTD structure. The structure of the RNAPII catalytic core was generated using the RaptorX web server (Kallberg et al., 2012). b) Yeast and human RNAPII CTD sequences are shown. All the sequences of the RNAPII CTD were obtained from the NCBI reference sequence database (NP_010141 for S. cerevisiae and NP_000928 for H. sapiens). The RNAPII CTD mostly comprises repetitions of the consensus YSPTSPS sequence (in round brackets) and non-consensus sequences (in red). The number next to the round brackets indicates the repeat number of the CTD heptapeptide consensus sequences. The sequences found at the end of each CTD indicate the TIP or the CTD C-terminal motif, which contributes to CTD stabilization. c) The RNAPII CTD modifications within the consensus and the non-consensus sequences during transcription. The consensus CTD modifications include phosphorylation (P), isomerization between cis- and trans-conformation (I), and glycosylation (G) on the indicated CTD residues. The non-consensus CTD modifications include acetylation (Ac), methylation (Me), and ubiquitination (Ub) at the 7th residue of the CTD. d) Amino acid frequency at each position of the RNAPII CTD heptad repeats in S. cerevisiae and H. sapiens. The relative frequency and conservation of each amino acid is represented by the height of the corresponding amino acid symbol. The fourth and seventh positions of the CTD heptapeptide repeat display more sequence variation than other CTD residues in both yeast and humans. The graphics were generated using the WebLogo web server (Crooks et al., 2004).
(in mammals) or Srb10 (in yeast), a component of the mediator complex, phosphorylates CTD S5 and S2 in vitro (Belakavadi and Fondell, 2010; Hengartner et al., 1998). S5 phosphorylation (S5-P) induces the dissociation of the mediator from the pre-initiation complex (PIC) and stimulates RNAPII for promoter clearance, processes that are required to transition from transcription initiation to elongation (Rodriguez et al., 2000; Sogaard and Svejstrup, 2007; Viladevall et al., 2009). S5-P is also necessary to facilitate the recruitment of pre-mRNA capping enzymes (Ho and Shuman, 1999; Schroeder et al., 2000; Wen and Shatkin, 1999). The levels of S5-P are enriched at the promoter and the 5′-end of the gene during transcription initiation and significantly decrease as transcription proceeds to the 3′-end of the gene (hereafter, see Fig. 3 for the patterns of CTD modifications in yeast and in human) (Bataille et al., 2012; Komarnitsky et al., 2000; Rodriguez et al., 2000).

In budding yeast, two main S2 kinases are known, namely, Bur1 and Ctk1 (Cdk9 and Lsk1, respectively, in Schizosaccharomyces pombe). S2-P levels caused by these kinases increase as transcription elongation progresses toward the 3′ end of genes and decreases beyond the poly(A) sites (Ahn et al., 2004; Bataille et al., 2012; Kim et al., 2004; Viladevall et al., 2009). Yeast Bur1 kinase (Bur1 and its cyclin Bur2) binds to RNAPII S5-P and subsequently phosphorylates S2 during elongation. However, unlike Ctk1, Bur1 primarily acts at the 5′-end of genes and stimulates Ctk1-mediated phosphorylation (Liu et al., 2009; Qiu et al., 2009). Additionally, Bur1 phosphorylates CTD S5 and S7 in vitro (Bowman and Kelly, 2014; Qiu et al., 2009; Tietjen et al., 2010). Yeast Ctk1, a component of the CTD kinase 1 (CTDK1) complex (Ctk1 and cyclin Ctk2/3), is another kinase that phosphorylates RNAPII CTD S2. Ctk1 is required for the release of RNAPII from general transcription factors (GTFs) in the PIC at the 5′-end of genes and the subsequent RNAPII S2-P by Ctk1 also plays an essential role in coupling transcription with the 3′-end processing of nascent mRNA transcripts by recruiting 3′-end processing factors by (Ahn et al., 2004; Qiu et al., 2009). In humans, four CTD S2-P kinases have been reported: CDK9, a catalytic subunit of P-TEFb (positive transcriptional elongation factor b), CDK12, CDK13 and BRD4, a bromodomain protein (Bartkowiak et al., 2010; Bosken et al., 2014; Bowman and Kelly, 2014; Devaiah et al., 2012; Liang et al., 2015). CDK13 phosphorylates S2 of the CTD only in vitro (Bartkowiak et al., 2010; Liang et al., 2015). BRD4 is an atypical protein kinase that phosphorylates S2 of the CTD primarily at the 5′-ends of genes and stimulates P-TEFb (Devaiah and Singer, 2012; Devaiah et al., 2012).

The seventh position of each CTD repeat is degenerate and allows several modifications to serine, arginine, or lysine residues of the CTD (Fig. 1C). Among these modifications, serine 7 phosphorylation (S7-P) was initially observed in snRNA genes and protein-coding genes in mammalian cells (Chapman et al., 2007; Egloff et al., 2007; Glover-Cutter et al., 2009). The S7-P mark is found at gene promoters and remains stable toward the 3′-ends of coding regions in yeast or increases toward the 3′-ends of coding regions in mammals (Bataille et al., 2012; Chapman et al., 2007; Kim et al., 2010). Budding yeast Kin28 or Bur1, fission yeast Mcs6 or Cdki9, and human CDK7 phosphorylate CTD S7 in vivo (Glover-Cutter et al., 2009; Tietjen et al., 2010). Human DNA-PK (DNA-dependent protein kinase) and P-TEFb also phosphorylate S7 in vitro (Glover-Cutter et al., 2009; Tyagi et al., 2011). CTD S7-P recruits RAP22 (RNA polymerase II associated protein 2), an S5-P phosphatase and integrator complex, to snRNA genes, ensuring proper transcription and transcript processing (Albrecht and Wagner, 2012; Egloff et al., 2012b), and S7-P also plays a role in stimulating the CTD S2 kinase CDK9 (Czudnochowski et al., 2012; St Amour et al., 2012). However, the substitution of S7 with alanine in all CTD repeats is not lethal and has little effect on the expression of protein-coding genes; only 1.2% of annotated protein-coding RNAs were significantly dysregulated by at least twofold by this mutation in fission yeast (Chapman et al., 2007; Schwer et al., 2014). Thus, it appears that the role of CTD S7-P in regulating gene transcription is confined to regulating the expression of snRNAs or other specifically induced genes.

3.2. CTD tyrosine phosphorylation

The phosphorylation of CTD tyrosine 1 (Y1-P) was first described in mammals and has recently been reported in yeast (Baskaran et al., 1993; Mayer et al., 2012). CTD Y1-P in mammals is mainly found at promoters and decreases toward the 3′-ends of genes, similar to the pattern found for CTD S5-S7 (Descoste et al., 2014). By contrast, in yeast, the CTD Y1-P is found at all active genes, starting low at promoters and increasing toward the 3′-ends of genes, similar to the S2-P pattern (Mayer et al., 2012). Notably, several functions of Y1-P in regulating gene transcription have been described in yeast and mammals. For example, CTD Y1-P activates the recruitment of the elongation factor Spt6 but significantly decreases as transcription proceeds to the 3′-end of the gene.
impairs the recruitment of the CID (CTD-interacting domain)-containing termination factors Nrd1, Pcf11, and Rtt103 in yeast (Mayer et al., 2012). Another report shows that cytoplasmic and nucleoplasmic Rpb1 is phosphorylated exclusively on CTD Y1 in chicken DT40 cells and that CTD Y1-P prevents CTD degradation by the 20S proteasome in vitro (Hsin et al., 2014a). In addition, CTD Y1-P is associated with antisense promoter transcription and active enhancers in mammalian cells (Descostes et al., 2014). To date, human c-Abl kinase is the only kinase that phosphorylates Y1, and no other such kinases have been identified in other organisms (Baskaran et al., 1993, 1999).

3.3. CTD threonine phosphorylation

The phosphorylation of CTD threonine 4 (T4-P) is found in coding regions in patterns similar to those of CTD S2-P or Y1-P in yeast (Mayer et al., 2012). In humans, Polo-like kinase 3 (PLK3) and CDK9 are known to phosphorylate CTD T4 (Hintermair et al., 2012; Hsin et al., 2014b). ChIP sequencing from B-cells shows that the CTD T4-P level in humans is low at the TSS but strongly increases toward the 3′ ends of genes, and it further shows that this modification increases at approximately 300-bp downstream of CTD S2-P, suggesting that CTD S2-P is a prerequisite for the subsequent T4-P (Hintermair et al., 2012). Additionally, a genome-wide study in yeast shows that CTD T4-P is mainly present in the coding region and is low at the poly(A) site, facilitating the binding of termination factors such as Pcf11 or Rtt103 to the 3′-ends of genes (Mayer et al., 2012). Several other reports show that CTD T4-P is required for the induction of certain specific genes: in budding yeast, T4-P plays a critical role in the expression of certain inducible genes that are associated with phosphate and galactose metabolism (Rosonina et al., 2014). In fission yeast, T4-P is associated with protein expression involved in phosphate homeostasis, specifically, the phosphate transporter SPBC8E4.01c and the acid phosphatase Pho1.

**Fig. 3.** Coordination of RNAPII modification with histone modification during the transcription cycle. The phosphorylation and dephosphorylation at S2 and S5 of the RNAPII CTD are co-transcriptionally coordinated with several histone modifications. The histone-modifying complexes/factors that bind to S2 or S5 of the RNAPII CTD and couple the CTD code with the histone code during gene transcription are shown at the bottom of the figure (see text for details). The color of each complex/factor corresponds to the histone modification found at the top of the figure. The serine residues phosphorylated at each stage of transcription are shown in bold red. The genome-wide distribution of each histone modification pattern is adapted from Li et al. (2007a) and Smolle and Workman (2013). The average patterns of each RNAPII CTD modification level during each stage of transcription are found in yeast and mammals, based on genome-wide studies, and they are shown in the middle of the figure (see text for details). Proteins are shown from representative organisms: yeast (y) and human (h).
(Schwer et al., 2014). In chicken cells, T4-P is required for processing intron-less replication activated histone genes, with no effect on the expression of other protein-coding genes or noncoding RNA genes (Hsin et al., 2011).

4. Other modifications to the CTD

4.1. CTD glycosylation

CTD glycosylation involves the addition of the nucleotide sugar uridine diphosphate-N-acetylglucosamine (UDP-GlcNAc) to the hydroxyl groups of the CTD SS, S7 and T4 side chains (Kelly et al., 1993; Ranunculo et al., 2012). In human cells, UDP-GlcNAc is added to the RNAII CTD by O-GlcNac transferase (OGT) and is removed by N-acetyl-β-glucosaminidase (OGA) during PIC assembly, and transcription in vitro is dependent on both OGT and OGA (Gao et al., 2001; Haltiwanger et al., 1992; Ranunculo et al., 2012). In particular, the glycosylation attaches a rather large sugar moiety to the CTD peptide chain and thus primarily acts as a steric block to prevent aberrant CTD phosphorylation (Comer and Hart, 2001; Kelly et al., 1993). Thus, the question of how CTD phosphorylation occurs on the glycosylated RNAII CTD during transcription remains open. Comer and Hart (2001) suggest that OGT and CTD kinases have distinct CTD repeat requirements and different modes of substrate binding, thereby providing potential mechanisms for the differential regulation of these two enzymes (Comer and Hart, 2001). That study and another hypothesized that the glycosylated CTD is recruited to the promoter during PIC assembly, preventing premature CTD phosphorylation and RNAII degradation, but upon the completion of PIC assembly, CTD glycosylation is removed by OGA, consequently facilitating the phosphorylation of the RNAII CTD for the progression of transcription (Comer and Hart, 2001; Lewis, 2013; Ranunculo et al., 2012).

4.2. CTD proline isomerization

CTD proline cis and trans isomerization is another essential modification that controls the binding of numerous proteins (Zhang et al., 2010a). This non-covalent posttranslational modification is induced in the cyclic amino acid proline at the 3rd and the 6th positions of the CTD sequence (P3 and P6). PIN1 in mammals and Ess1 or Rrd1 in yeast are the peptidyl-prolyl isomerase (PPIase) domains that catalyze cis/trans propyl-peptide bond inter-conversions (Albert et al., 1999; Jouvet et al., 2010; Morris et al., 1999). Both PIN1 and Ess1 contain two domains: the N-terminal WW domain that identifies and interacts with the phosphorylated CTD mark and the catalytic C-terminal PPlase domain (Hanes, 2014; Ranganathan et al., 1997). The presence of two propyl-peptide bonds within CTD repeats results in four possible cis–trans configurations, further expanding the complexity of the CTD code signature by providing a scaffold for the recruitment of various CTD-modifying proteins (Hanes, 2014). For example, the yeast phosphatase Ssu72 and the yeast RNA-binding protein Nrd1 recognize CTD repeats with S5-P and the downstream P6 in the cis-form, whereas both the human CTD S5-P phosphatase SCP1 and the yeast poly(A)/termination factor Pcf11 bind to CTD repeats with P3 and P6 in the trans-configuration (Kubicek et al., 2012; Noble et al., 2005; Werner-Allen et al., 2011; Zhang et al., 2006). Additionally, PIN1 has been reported to inhibit RNAII CTD dephosphorylation by the CTD phosphatase Fcp1 (TFIIF-associated CTD phosphatase 1), and PIN1-dependent RNAII hyperphosphorylation was shown to occur in M-phase cells (Xu et al., 2003). By contrast, Ess1 appears to have the opposite effect on CTD phosphorylation — ess1 mutant analysis suggests that Ess1 promotes CTD dephosphorylation by Fcp1 (Wu et al., 2000).

4.3. CTD acetylation

A recent study identified acetylation as a new modification of the mammalian RNAII CTD, showing that this modification is required to induce growth factor response genes (Schroder et al., 2013). In this report, the 7th lysine of non-consensus CTD distal heptad repeats is subject to acetylation by p300/KAT3B acetyltransferases both in vivo and in vitro. Notably, a genome-wide study showed that CTD acetylation is specifically enriched downstream of the TSSs of RNAII-occupied genes, suggesting that CTD acetylation contributes to the early stages of transcription. However, CTD acetylation has yet to be reported in yeast.

4.4. CTD methylation

Another recent study revealed that the CTD of RNAII is methylated in mammalian cells (Sims et al., 2011). Intriguingly, only one arginine residue (R1810) found at the 31st non-consensus repeat of the hypophosphorylated RNAII CTD is site-specifically methylated by the co-activator-associated arginine methyltransferase 1 (CARMT1) in vitro and in vivo. This modification was found to be required to regulate the early stages of transcription during the expression of snRNAs and snoRNAs.

4.5. CTD ubiquitylation

The 7th lysine of non-consensus CTD distal heptad repeats is subject to ubiquitylation in mammals. Only six of the eight lysines of the non-consensus CTD repeats are ubiquitylated by Wwp2, a HECT-domain ubiquitin E3 ligase, leading to the degradation of Rpb1 by 26S proteasome-mediated proteolysis (Li et al., 2007b). Intriguingly, CTD S2-P and S5-P are increased upon knockdown of the Wwp2 protein in embryonic pluripotent stem cells, suggesting that CTD ubiquitylation is necessary for maintaining normal Rpb1 protein levels in embryonic pluripotent stem cells (Li et al., 2007b). In yeast, two lysine CTD residues, the lysine at the 4th position of the 26th CTD repeat (K1720) and the lysine within the CTD tip region (K1725), are ubiquitylated by a ubiquitin E3 ligase, Ars1 (see Fig. 1B) (Daunly et al., 2008). However, this Ars1-mediated ubiquitylation is a mono-ubiquitylation and plays a role in promoting the ejection of the Rpb4/7 heterodimer from the RNAII decameric core, which is required for the initiation of transcription and the coordination of various gene expression steps by associating with mRNAs (Duan et al., 2013; Harel-Sharvit et al., 2010).

5. Encoders, readers, and decoders of the CTD code signature

Different proteins recognize and interact with specific modification patterns on the RNAII CTD repeats. Although biochemical, molecular and genetic studies have identified several CTD-modifying enzymes, including CTD kinases, phosphatases and isomerases, how these proteins coordinate the combinatorial establishment of precise CTD modifications throughout the transcription cycle remains elusive. The proteins that change the CTD to produce the various modification states of RNAII are known as the ‘encoders’ or ‘writers’ of RNAII CTD modifications (Supplementary Table S1) (Jeronimo et al., 2013). An RNAII with a modified or unmodified CTD is recognized by other groups of proteins called the ‘readers’ of the RNAII CTD (Jeronimo et al., 2013) (Table 2). CTD-binding proteins are categorized based on the distinct protein domains through which they interact with RNAII. Several CTD-binding domains have been reported, including the CTD domain, the WW domain (40 amino acids long; W, tryptophan), the SH2 (Src Homology 2) domain, the SR1 (Set2 Rpb1 Interaction) domain, the FF domain (60 amino acids long; F, phenylalanine), the low-complexity sequence domain, the RRM (RNA recognition motif) domain and the GTase NT (guanylyltransferase nucleotidyltransferase) domain (Corden, 2013; Jeronimo et al., 2013).
Several readers of the RnapII CTD interact with different forms of RNAPII. For example, TBP interacts with unphosphorylated RNAPII CTD by recognizing S2-P, and Rsc protein interacts with RNAPII CTD in vivo and is found at both the 5′ and 3′-ends of actively transcribed genes. In addition, in vivo and in vitro activities of Fcp1 and Sen1 CTD interacting/binding proteins and readers of CTD modification. Several readers of the RNAPII CTD interact with different forms of RNAPII. For example, TBP interacts with unphosphorylated RNAPII CTD by recognizing S2-P, and Rsc protein interacts with RNAPII CTD in vivo and is found at both the 5′ and 3′-ends of actively transcribed genes. In addition, in vivo and in vitro activities of Fcp1 and Sen1 CTD interacting/binding proteins and readers of CTD modification.
against T4-P and SS5-P/S7-P of the CTD, respectively, have also been reported (Cho et al., 2001; Hsin et al., 2014b; Xiang et al., 2012). Yeast Ssu27 was first designated as an SS5 phosphatase, but it also targets the CTD S7 residue (Krishnamurthy et al., 2004; Xiang et al., 2012). Yeast Rtr1 (Regulator of transcription 1) is an atypical SS and Y1 CTD phosphatase that plays an essential role in the transition of the CTD phosphorylation status from SS5-P to S2-P at the 5′-ends of genes (Hsu et al., 2014; Mosley et al., 2009). More recently, RPAP2 was identified as the human homologue of the yeast Rtr1 phosphatase and dephosphorylates CTD SS5-P in vitro and in vivo (Egloff et al., 2012b). Human CDC14A and CDC14B are the orthologs of the yeast M-phase-specific phosphatase Cdc14 and regulate the expression of cell-cycle-specific genes (Guillamot et al., 2011). CDC14A decreases both SS5-P and the S2-P in vitro, whereas CDC14B displays a preference for SS5 both in vitro and in vivo (Guillamot et al., 2011). Mammalian cells also contain many SCPs (small CTD phosphatases) that include the FCPH (FCP-homology) domain but lack the BRCT (BRCA1 C-terminal) domain of the FCP1 phosphatase (Yeo et al., 2003). An FCP1-like phosphatase, SCP1, preferentially catalyzes the dephosphorylation of CTD SS5-P, a modification that is important for transitioning from transcription initiation to elongation (Yeo et al., 2003; Zhang et al., 2010b). Finally, yeast Glc7, a cleavage/poly(A) factor, acts as a Y1 CTD phosphatase at the 3′-ends of protein-coding genes both in vitro and in vivo (Schreieck et al., 2014).

6. Appreciating the synchronized function of the CTD and the histone code during transcription

Chromatin in eukaryotes adopts a dynamic and highly ordered structure. The basic unit of chromatin is the nucleosome, which contains 147 base pairs of DNA tightly wrapped around an octamer of the core histones (two each of H2A, H2B, H3, and H4). This chromatin structure generally impedes RNA polymerase-mediated transcription but can be dynamically regulated by histone posttranslational modifications, chromatin remodeling, or the incorporation or eviction of histone variants (Li et al., 2007a; Zentner and Henikoff, 2013).

Among the many histone modifications, four key modifications highly facilitate chromatin dynamics: lysine acetylation, lysine/arginine methylation, serine/threonine phosphorylation, and lysine ubiquitylation (Lalonde et al., 2014). Histone acetylation neutralizes the positive charges of histone lysine residues and weakens the interactions with negatively charged nucleosomal DNA, leading to an open chromatin structure that increases the accessibility of DNA to the transcriptional machinery (Zentner and Henikoff, 2013). Histone phosphorylation imparts a negative charge to histones and is assumed to have a similar function to histone acetylation in modulating nucleosome dynamics (Banerjee and Chakravarti, 2011). However, unlike histone acetylation or phosphorylation, histone methylation does not alter the histone charge. Adding methyl groups to histone residues produces steric bulk and removes a potential hydrogen bond donor, thereby affecting the interactions between histones and DNA to destabilize the nucleosome (Casadio et al., 2013). Recent work also demonstrated that histone ubiquitylation interferes with chromatin compaction by inhibiting nucleosome array folding and inter-chromatin fiber oligomerization, leading to an open chromatin conformation (Fierz et al., 2011).

The consensus and non-consensus RNAPII CTD repeats are associated with co-transcriptional chromatin modifications or chromatin remodeling (Rosonina et al., 2014; Zhang et al., 2012). Notably, posttranslational modifications of the RNAPII CTD, along with histone modifications, positively or negatively affect gene transcription by adopting different combinations of modification patterns. Accumulating evidence suggests that there are three possible steps that support the coordination between the CTD and histone modifications to reinforce the proper assembly of the transcriptional machinery to regulate transcription. First, CTD modifications by CTD code regulators affect the association of histone code regulators with chromatin (Fig. 4). For example, the HAT complex SAGA (Spt-Ada-Gcn5 acetyltransferase) is recruited to the GAL1 and ARG1 genes, which are dependent on SS5-P of RNAPII CTD in yeast (Govind et al., 2007). Second, changes in histone modifications by histone code regulators affect the association of CTD code regulators with chromatin. For example, Ubp8 deubiquitylase, a subunit of the SAGA and SALSA/SILK complexes, triggers the association of Ctk1 with a subset of SAGA-dependent genes (Wyce et al., 2007). Third, coordinated functions between the CTD and histone codes lead to synchronized effects on the fine-tuning of gene expression at each step. For example, Ubp8, which is recruited to the coding regions of genes through SS5-P of the RNAPII CTD, deubiquitylates H2B, thereby stimulating the association of Ctk1 with actively the transcribed regions of genes (Wyce et al., 2007). The subsequent SS2-P of RNAPII CTD by Ctk1 is important for coupling transcription with the 3′-end processing of nascent mRNA transcripts (Ahn et al., 2004). Although the 3-step coordination between the RNAPII CTD and histone modifications is a good example of the coordinated pathways for the fine-tuned regulation of gene transcription by these two codes, most synchronized effects have thus far been shown from changes in histone modifications following RNAPII CTD modifications (Figs. 3 and 5).

6.1. RNAPII CTD and histone acetylation

Histone acetylation, a reversible modification that is regulated by histone acetyltransferases (HATs) and histone deacetylases (HDACs), facilitates the recruitment of RNAPII to the transcribing unit and is primarily associated with euchromatin. Several studies link histone acetylation to phosphorylated RNAPII CTD during gene transcription, showing that the coordination of these two modifications is required for transcriptional progression and regulation. The lysine 27 acetylation of histone H3 (H3K27ac) is known as an important enhancer mark for distinguishing between active and poised enhancer elements (Creighton et al., 2010). Studies of the impact of histone acetylation on RNAPII activation and downstream transcriptional kinetics in mouse cells indicate that H3K27ac is associated with RNAPII CTD SS2-P during gene transcription, facilitating RNAPII promoter escape at a broad set of endogenous genes (Stasevich et al., 2014a). Histone H3 lysine 9 acetylation (H3K9ac) found near the TSS is tightly related to transcriptional initiation (Kimura, 2013). Fab-based live endogenous modification labeling was used to quantify the concentrations of H3K9ac in single living cells and their progeny and revealed that H3K9ac is correlated with both RNAPII CTD SS5-P and S2-P throughout the cell cycle. In particular, a relatively strong correlation between H3K9ac and CTD SS5-P is observed a few hours before mitosis, which appears to reflect gene bookmarking for efficient re-initiation following mitosis (Stasevich et al., 2014b).

Histone acetylation is found across the coding regions of genes during transcription. As mentioned earlier, the coactivator SAGA occupies the coding regions of the GAL1 and ARG1 genes depending on SS5-P of the RNAPII CTD in yeast (Govind et al., 2007). The SAGA complex is involved in the acetylation of histone H2B on K16 residues and of histone H3 on K4, K9, K14, K18 and K23 (Rando and Winston, 2012). Indeed, the occupancy of the SAGA complex on actively transcribed genes is reduced in yeast ino80 mutant cells at the 5′- and 3′-ends of genes (Govind et al., 2007). Interestingly, Sus1, a component of both the SAGA and TRE2X (transcription/export 2) complexes, undergoes CTD SS5-P-dependent targeting to coding regions but is also associated with the elongating form of RNAPII phosphorylated at CTD S2 (Pascual-Garcia et al., 2008). The HAT complex NuA4 (nucleosomal acetyltransferase of H4) acetylates histones H4, H2A and H2AZ and is co-transcriptionally targeted to genes by the phosphorylation of CTD S5, particularly during Gcn4-targeted gene activation (Ginsburg et al., 2009; Rando and Winston, 2012). The chromatin remodeling protein Rvb1, which is associated with the Ino80 and Tip60/NuA4 complexes, promotes H4/H3 histone acetylation and is required for the recruitment
of RNAPII and its CTD SS-P at interferon (IFN)-α-stimulated gene promoters in humans (Gnatovskiy et al., 2013).

RNAPII and HDAC complexes associate and coordinate with each other during transcription elongation to suppress cryptic initiation or nucleosome eviction (Carrozza et al., 2005; Govind et al., 2010; Rando and Winston, 2012). Rpd3C(S) is a class-I type HDAC complex that is recruited to coding regions based on the ability of both CTD S5-P by Kin28 and CTD S2-P by Ctk1 to suppress cryptic initiation (Drouin et al., 2010; Govind et al., 2010). HDACs such as Hos2, a subunit of the Set3 and Rpd3C(L) complexes, and Hda1, a class-II HDAC, are co-transcriptionally recruited to the ARG1 gene based on the ability of RNAPII CTD S5-P to repress nucleosome eviction (Govind et al., 2007, 2010).

6.2. RNAPII CTD and histone methylation

Reversible histone methylation is regulated by histone methyltransferases (HMTs) and histone demethylases (HDMs). Histone methylation on lysine and arginine residues shows different modification patterns: mono-, di-, or trimethyl forms for lysines and mono- or di-(asymmetric or symmetric) for arginines (Zentner and Henikoff, 2013). There are three lysine HMTs in budding yeast (Set1/KMT2, Set2/KMT3 and Dot1/KMT4) and eight (KMT1 to 8) in mammals (Allis et al., 2007). Among lysine HMTs in yeast, the methylations of histone H3 lysine 4 (H3K4) and those of histone H3 lysine 36 (H3K36) are well characterized and interconnected with the phosphorylation of the RNAPII CTD (Shilatifard, 2012).

Set1 is a component of COMPASS (complex of proteins associated with Set1), the yeast homologue of the mammalian complex MLL, and is recruited to the promoters and 5′-ends of actively transcribed genes (Shilatifard, 2012). Importantly, the phosphorylation of CTD S5 by Kin28 is necessary for the recruitment of COMPASS in vivo (Ng et al., 2003). The recruitment of a transcription elongation factor, the PAF1 complex (polymerase-associated factor 1), to genes is also facilitated by SS-P of the CTD, showing a strong association with the di-phosphorylated form of the CTD at either S2/S5 or S5/S7. This PAF1 complex then stimulates the association between Set1 and promoter regions (Krogan et al., 2003; Ng et al., 2003; Qiu et al., 2012). Furthermore, the different methylation forms of Set1 establish two distinct chromatin zones: the tri-methylation of histone H3K4 (H3K4me3) is found at the promoter region, and the di-methylation of histone H3K4 (H3K4me2) is found just downstream of the promoter. H3K4me3 at promoters is associated with HATs such as SAGA, NuA3 and NuA4, whereas the H3K4me2 serves as a binding site for the Set3 HDAC, a complex that mediates the effects of overlapping noncoding transcription (Kim and Buratowski, 2009; Kim et al., 2009, 2012b). Therefore, the RNAPII CTD SS-P is a critical modification for recruiting the histone encoder Set1, through which H3K4me3 and H3K4me2 mark distinct stages of transcription to recruit other histone modifiers such as HATs or HDACs.

Many studies in yeast and humans show that the recruitment of the histone methyltransferase Set2, which methylates histone H3K36, is associated with CTD phosphorylation (Krogan et al., 2003; Li et al., 2002, 2003). Set2 interacts with both CTD S5 and S2 through its SRI (Set2-Rpb1 Interacting) domain (Kizer et al., 2005). Histone H3K36 methylation by Set2, particularly at the 3′ regions of transcribed genes, acts as a mark to recruit the Rpd3C(S) HDAC complex, leading to histone deacetylation in this region (Keogh et al., 2005; Kim and Buratowski, 2009). DOT1 (disruptor of telomeric silencing 1), which methylates histone H3K79, is a histone methyltransferase without a SET domain (Nguyen and Zhang, 2011). Human DOT1 binds directly
to RNAPII phosphorylated at S5, S2, or both, via its CTD-binding patch (Kim et al., 2012a). Genome-wide analysis further shows a strong occupancy correlation at the TSS between the mono-, di-, and tri-methylation of histone H3 at K79 by DOT1 and RNAPII CTD S5-P, suggesting a coordinated role for H3K79 methylation with phosphorylated RNAPII in active transcription (Kim et al., 2012a).

Two lysine HDM families, HDMs containing an amine oxidase-domain such as LSD1/2 (also called lysine demethylase1/2 or KDM1/2) and jumonji C (JmjC)-domain containing HDMs such as JHDM1/2/3, are also associated with the phosphorylation of the RNAPII CTD. A lysine HDM, JMJD3 (also known as KDM6B), reverts a repressive tri-methylation of histone H3K27 (H3K27me3) and de-represses the gene expression required for neural and epidermal differentiation (Kooistra and Helin, 2012; Stock et al., 2007). A genome-wide analysis further shows that JMJD3 directly interacts with CTD S2-P in transforming growth factor β-stimulated mouse neural stem cells (Estaras et al., 2013). Similarly, the mono- or di-methylation of histone H3K27 is demethylated by an HDM enzyme, JHDM1D (also known as KIAA1718 or KMD7A), and knocking down JHDM1D or JMJD3 reduces the enrichment of RNAPII CTD S2-P across the CXCL3 and ZC3H12c genes in human promyelocytic leukemia cells; this suggests that the demethylase activity of JHDM1D or JMJD3 is necessary to release poised RNAPII into an active state during transcription elongation because of their close association with RNAPII phosphorylated at S2 of the CTD (Chen et al., 2012).

6.3. RNAPII CTD and histone ubiquitylation

Non-proteolytic protein ubiquitylation, or mono-ubiquitylation, occurs on histones H2A, H2B, and H4 to regulate gene transcription (Geng et al., 2012; Kim et al., 2013). The major mono-ubiquitylation site in histone H2B is K123 in budding yeast, K119 in fission yeast and K120 and K34 in mammals (Cao and Yan, 2012; Fuchs and Oren, 2014). Although there are specific histone ubiquitin-conjugating enzymes, E2, Rad6, and a specific histone ubiquitin ligase enzyme E3, Bre1, in budding yeast (Cao and Yan, 2012; Fuchs and Oren, 2014), different types of histone E2s and E3s are known in humans: UBE2E1 (also known as UbcH6), UBE2A/B (also known as RAD6A/B), BAF250B (BRG1-associated factor 250B) and UBE2D3 (UbcH5c) for E2 enzymes and RNF20/40 (Ring finger protein 20/40), ISIL2 (Male-specific lethal 2 homologue), CUL4A (Cullin 4A) and RING1A/B for E3 enzymes. The mono-ubiquitylation of histones is reversible, and ubiquitin can be removed by deubiquitylating enzymes such as Ubp8 or Ubp10 in yeast and by the ubiquitin-specific peptidases USP3, USP22, USP42 or UBP12/46 in humans (Cao and Yan, 2012).

Histone mono-ubiquitylation is associated with elongating RNAPII. Mutating the yeast S5 CTD kinase Kin28 (kin28ts16) or deleting the CTD dramatically abolishes the Rad6-mediated H2B ubiquitylation of K123 (H2BK123ub) (Xiao et al., 2005). Additionally, loss of the yeast CTD S2 kinase Bur1 also results in reduced levels of Rad6-mediated H2BK123ub, suggesting that CTD phosphorylation at either S5 or S2 is
required for histone H2B ubiquitylation (Wood et al., 2005). Notably, histone H2BK123ub regulates histone H3 methylation at K4 or K79 through a pathway known as histone ‘trans-tail regulation’ (Daniel et al., 2004; Henry et al., 2003; Sun and Allis, 2002). In mammals, there are similar connections between histone ubiquitylation and RNAPII CTD phosphorylation. Knockdown of either of the CTD S2 kinases CDK9 or BRD4 or the substitution of CTD S2 with alanine leads to a loss of H2BK123ub; by contrast, knocking down RNF20/40 or mutating UBE2A (S120A) decreases the RNAPII CTD S2-P (Nagarajan et al., 2014; Pirngruber et al., 2009; Schchet et al., 2012; Wu et al., 2014). Similarly, the knockdown of human E3 ubiquitin ligase MSL2, a ubiquitylase for histone H2B on K34, causes a decrease in the chromatin binding of CDK9 (Wu et al., 2011, 2014). CUL4A, a human E3 ubiquitin ligase for histone H4 on K31, is required for linker histone H1.2-dependent transcription elongation associated with phosphorylated RNAPII at CTD S2 (Kim et al., 2013). Additionally, the E3 ubiquitin ligase RING1A/B, which is present in the Polycomb Repressor Complex 1 (PRC1), mediates the mono-ubiquitylation of histone H2A on K119, which enforces the poised RNAPII configuration with phosphorylation at CTD S5 at bivalent genes in embryonic stem cells (Stock et al., 2007).

Histone deubitylation has been found to be as significant as ubiquitylation for activating transcription. Although H2BK123ub is closely associated with CTD S5-P, histone ubiquitylation itself impedes the nucleosomal association of Ctk1 and acts as a barrier for the subsequent S2-P of the RNAPII CTD in yeast (Wyce et al., 2007). This problem is solved by Ubp8 deubitylase, which deubitylates H2BK123ub and triggers Ctk1 association at a subset of SAGA-dependent genes (Wyce et al., 2007). Additionally, the Ik1-associated protein Lge1 stimulates H2BK123ub by assisting in the recruitment of Ik1 and by inhibiting the recruitment of the deubitylase Ubp8 that is associated with RNAPII CTD S5-P at the early stages of transcription, providing an example of a histone modifier that controls the balance of histone ubiquitylation and deubitylation during transcription (Song and Ahn, 2010). In humans, the deubitylating enzyme USP42 co-localizes with RNAPII phosphorylated at S2 of the CTD (Hock et al., 2014). Therefore, these reports from yeast and mammalian cells show that both histone ubiquitylases and deubitylases and their association with RNAPII phosphorylated at either S5 or S2 are critical for regulating transcription.

6.4. RNAPII CTD and histone phosphorylation

Phosphorylation is one of the most common posttranslational modifications of proteins, and histones contain specific phosphorylation sites, both on core histones and on the linker histone H1. Histone phosphorylation is associated with transcription activation in a broad range of organisms from yeast to humans, and there are several histone serine/threonine/tyrosine kinases, including PIM1 (provalional integration site for Moloney murine leukemia virus kinase 1), MSK1/2 (mitogen and stress activated protein kinase 1/2), RSK2 (ribosomal S6 kinase 2), Aurora B kinase and JAK2 (Janus kinase 2) in mammals and JIL1 (chromosomal serine/threonine-protein kinase) in Drosophila (Banerjee and Chakravarti, 2011; Sawicka and Seiser, 2012).

The phosphorylation of histone H3 on S10 and S28 (H3S10P and H3S28P, respectively) is known to be associated with phosphorylated RNAPII during transcriptional activation in humans and Drosophila (Ivaldi et al., 2007; Rossetto et al., 2012). In humans, PIM1 mediates H3S10P during MYC-dependent transcriptional activation and cellular transformation (Zippo et al., 2007), and this modification stimulates the recruitment of the CTD S2 kinases CDK9 or BRD4 during transcription elongation (Zippo et al., 2007, 2009). H3S28P is mediated by human MSK1 and is associated with RNAPII phosphorylation at CTD S5 at the C-FOS and A-GLOBIN promoters (Drobo et al., 2010; Keum et al., 2013; Lau and Cheung, 2011). Human Aurora B kinase phosphorylates H3S10 and H3S28 and plays significant roles during mitosis (Crosio et al., 2002; Frangini et al., 2013). Genome-wide analysis further shows that Aurora B kinase phosphorylates histone H3S28 and is associated with RNAPII phosphorylation at CTD S5 at the active gene promoters of resting human B-cells, preventing H2A ubiquitylation by the polycomb protein RING1B (Frangini et al., 2013). In Drosophila, JIL-1 kinase, a homologue to mammalian MSK1, phosphorylates both H3S10 and H3S28 in interphase cells and regulates transcription elongation by associating with S2-phosphorylated RNAPII at the Hsp70 (heat shock protein 70) gene (Ivaldi et al., 2007; Wang et al., 2013).

7. Application of RNAPII CTD to therapeutic targets for various diseases

Because modifications to the RNAPII CTD are closely associated with the regulation of gene transcription during the normal course of the transcription cycle, it is likely that these modifications are susceptible to the effects of diseases and various extracellular stimuli, including stress. However, alterations in the modification status of the RNAPII CTD lead to changes in gene expression that induce certain notable changes during cellular differentiation in eukaryotes; these changes trigger a broad range of diseases and requires adaptability toward stress conditions.

Viruses affect the function of host RNAPII in different ways, either through the ubiquitylation and proteasome-dependent degradation of RNAPII or by impeding RNAPII CTD phosphorylation (Rodriguez et al., 2007). One such example is the expression of the herpes simplex virus 1-encoded ICP22, which causes specific loss of S2-P or Y1-P of the RNAPII CTD in infected human host cells (Fraser and Rice, 2007). The Bunyamwera virus nonstructural protein NSs also prevents CTD S2-P, leading to the inhibition of type I interferon (IFN)-α gene expression and to impaired host immunity in mammalian cells (Thomas et al., 2004). Another case is the association of the influenza virus RNA-dependent RNA polymerase with host RNAPII phosphorylated at CTD S5, which inhibits RNAPII elongation (Chan et al., 2006). Likewise, several other viruses, such as cytomegalovirus, Epstein–Barr virus, human immunodeficiency virus, or adenovirus, also affect the phosphorylation status of the RNAPII CTD and alter gene transcription in host cells (Bark-Jones et al., 2006; Chen et al., 2014; Tamrakar et al., 2005; Zhou et al., 2000). Thus, viral infection severely affects RNAPII CTD modifications, facilitating the viral life cycle by interrupting gene expression in host cells.

Several reports show close associations between the dysregulation of RNAPII CTD phosphorylation and various types of cancer cells, such as breast cancer, liver cancer, and chronic lymphocytic leukemia cells (Bi et al., 2013; Ji et al., 2014; Pallis et al., 2013; Walsby et al., 2014). Moreover, because most of these cancerous cells exhibit the dysregulation of CTD-modifying enzymes, thereby causing CTD-dependent transcriptional irregularities, several types of inhibitors targeting CTD-modifying enzymes are considered promising candidates for cancer therapy or future cancer diagnosis. For example, LDC3140 and LDC4297 are CDK7 inhibitors and inhibit RNAPII CTD S5-P and S7-P in multiple tumor cell lines; LY2857785 is a CDK9 inhibitor, and the application of this drug to leukemia or solid tumor cell lines significantly reduces RNAPII S2-P, leading to apoptosis (Kelsd et al., 2014; Yin et al., 2014). Likewise, several other drugs inhibit either RNAPII CTD S2-P or S5-P in various cancer cells (Table 3).

The modification of the RNAPII CTD is also associated with neurologic diseases and the inflammatory response. The RNA-binding protein FUS (fused in sarcoma) directly binds to the RNAPII CTD, and the disruption of this interaction leads to neurodegenerative diseases such as amyotrophic lateral sclerosis (ALS) (Schwartz et al., 2012, 2014). The binding of FUS to the RNAPII CTD inhibits the inappropriate hyperphosphorylation of the CTD S2 at thousands of human genes, and the knockdown of FUS in normal fibroblasts causes high occupancy of RNAPII at the TSS, which is correlated with an inappropriate accumulation of RNAPII CTD hyper-S2-P near the TSS in ALS patients’ cells.
Stress is a broad term used to define responses to environmental, physiological or biological stimuli. It is apparent that changes in environmental conditions such as temperature, pH, osmotic stress, nutrient availability, starvation, toxic chemical agents or radiation severely affect and alter the status of RNAPII CTD modifications and cause significant changes in gene expression patterns (Baugh et al., 2005; Liu and Tao, 2013; Miguel et al., 2013). However, it remains unclear how the RNAPII CTD responds and adapts to changes in stressful environments at the cellular and molecular levels. Only a few studies have shown that RNAPII CTD modifications are associated with various stress signals, indicating their potential roles in stress signaling pathways. For example, heat shock from 37 °C to 45 °C decreases the levels of S5-P and S7-P, indicating their potential roles in stress signaling pathways. For example, heat shock from 37 °C to 45 °C decreases the levels of S5-P and S7-P, indicating their potential roles in stress signaling pathways.
8. Perspectives and conclusions for the synchronized code

The combination of the CTD code with the histone code provides a new model for understanding transcriptional regulation as a whole. However, further studies are required to unravel the new types of potential modifications of both the RNAPII CTD and histones. The interaction and coordination of these two codes must also be seamlessly defined to understand the integration of the tangled pathways of each transcription process that is influenced by the two codes. Thus, a promising approach is to trace and characterize the roles of the machineries that induce modifications of RNAPII CTD repeats and of histones, such as the p300 complex or CARM1. Moreover, the variation of non-consensus CTD repeats and their modification deserve more attention because these repeats significantly increase the probability of new modifications and enhance the complexity of combinations within the CTD code. Finally, detailed information regarding when or how CTD and histone modifications occur and are regulated and which specific target genes are influenced by non-consensus CTD modifications would be interesting topics for future research. RNAPII CTD modifications and their coordination with the histone code enable the delicate regulation of gene transcription and cellular function, and there is a growing list of stress factors that are responsible for the manifestation of diseases caused by genetic defects in the transcription machinery. Thus, it appears to be important to unravel the exact defects in the modification patterns of the RNAPII CTD during transcription in various stress pathways to better understand and implement relevant measures for the proper reversal of stress and disease.

Acknowledgements

This work was supported by the National Research Foundation of Korea grant funded by the Korean government (MSIP) (No. 20110030049) and by the research fund of Hanyang University (HY-2014-P).

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.biotechadv.2015.07.008.

References


Chesnut, J.D., Stephens, J.H., Dahmus, M.E., 1992. Interactions of RNAPII CTD repeats and of histones, such as the p300 complex or CARM1. Moreover, the variation of non-consensus CTD repeats and their modification deserve more attention because these repeats significantly increase the probability of new modifications and enhance the complexity of combinations within the CTD code. Finally, detailed information regarding when or how CTD and histone modifications occur and are regulated and which specific target genes are influenced by non-consensus CTD modifications would be interesting topics for future research. RNAPII CTD modifications and their coordination with the histone code enable the delicate regulation of gene transcription and cellular function, and there is a growing list of stress factors that are responsible for the manifestation of diseases caused by genetic defects in the transcription machinery. Thus, it appears to be important to unravel the exact defects in the modification patterns of the RNAPII CTD during transcription in various stress pathways to better understand and implement relevant measures for the proper reversal of stress and disease.

282, 21901–21906.


