



Phosphorylation and functions of the RNA polymerase II CTD

Hemali P. Phatnani and Arno L. Greenleaf

Genes Dev. 2006 20: 2922-2936

Access the most recent version at doi:[10.1101/gad.1477006](https://doi.org/10.1101/gad.1477006)

References

This article cites 139 articles, 72 of which can be accessed free at:
<http://genesdev.cshlp.org/content/20/21/2922.full.html#ref-list-1>

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

Topic Collections

Articles on similar topics can be found in the following collections
[Chromatin and Gene Expression](#) (166 articles)

To subscribe to *Genes & Development* go to:
<http://genesdev.cshlp.org/subscriptions>

REVIEW

Phosphorylation and functions of the RNA polymerase II CTD

Hemali P. Phatnani¹ and Arno L. Greenleaf²

Department of Biochemistry, Duke University Medical Center, Durham, North Carolina 27710, USA

The C-terminal repeat domain (CTD), an unusual extension appended to the C terminus of the largest subunit of RNA polymerase II, serves as a flexible binding scaffold for numerous nuclear factors; which factors bind is determined by the phosphorylation patterns on the CTD repeats. Changes in phosphorylation patterns, as polymerase transcribes a gene, are thought to orchestrate the association of different sets of factors with the transcriptase and strongly influence functional organization of the nucleus. In this review we appraise what is known, and what is not known, about patterns of phosphorylation on the CTD of RNA polymerases II at the beginning, the middle, and the end of genes; the proposal that doubly phosphorylated repeats are present on elongating polymerase is explored. We discuss briefly proteins known to associate with the phosphorylated CTD at the beginning and ends of genes; we explore in more detail proteins that are recruited to the body of genes, the diversity of their functions, and the potential consequences of tethering these functions to elongating RNA polymerase II. We also discuss accumulating structural information on phosphoCTD-binding proteins and how it illustrates the variety of binding domains and interaction modes, emphasizing the structural flexibility of the CTD. We end with a number of open questions that highlight the extent of what remains to be learned about the phosphorylation and functions of the CTD.

The C-terminal repeat domain (CTD) of RNA polymerase II (RNAPII) is an amazing sequence arrangement at the end of the largest RNAPII subunit (apologies to Chow et al. 1977). This "domain" is inherently unstructured yet evolutionarily conserved, and in fungi, plants, and animals it comprises from 25 to 52 tandem copies of the consensus repeat heptad $Y_1S_2P_3T_4S_5P_6S_7$ (Corden 1990). The CTD is essential for life: Cells containing only RNAPII from which two-thirds or more of the repeats have been removed are inviable (Nonet et al. 1987; Zehring et al. 1988; for a review, see Corden 1990). It is

important that the heptads be in tandem: Insertion of an Ala residue between heptads is lethal in yeast, whereas insertion of an Ala between heptad *pairs* can be tolerated (Stiller and Cook 2004). While the CTD is indispensable in vivo, it is frequently not required for general transcription factor (GTF)-mediated initiation and RNA synthesis in vitro (Zehring et al. 1988; Kim and Dahmus 1989; Buratowski and Sharp 1990; Kang and Dahmus 1993; Akoulitchev et al. 1995). Thus, the CTD does not form part of the catalytic essence of RNAPII; rather, it must perform other functions. The nature and variety of these functions are currently being elucidated and are a main topic of this review.

A feature of the CTD that was discovered early, and that clearly carries functional implications, was that it is subject to hyperphosphorylation. RNAPII can exist in a form with a highly phosphorylated CTD (subunit II₀; RNAPII₀) and a form with a nonphosphorylated CTD (subunit II_a; RNAPII_a) (for a review, see Dahmus and Dynan 1992). Phosphorylation occurs principally on Ser2 and Ser5 of the repeats (Dahmus 1995, 1996), although these positions are not equivalent (West and Corden 1995; Yuryev and Corden 1996). A consequence of hyperphosphorylation is that the mobility in SDS gels of the II₀ form of the largest subunit is markedly reduced relative to that of form II_a (e.g., see Greenleaf 1992). Learning that RNAPII could exist in two forms led to efforts to understand functional differences between them. We now know that the phosphorylation state changes as RNAPII progresses through the transcription cycle.

Early results from Dahmus suggested that the initiating RNAPII was form IIA while the elongating enzyme was form IIO (Cadena and Dahmus 1987; Payne et al. 1989). In the meantime, the first CTD kinase (yeast CTDK-I) was purified (Lee and Greenleaf 1989, 1991; Sterner et al. 1995) and used to prepare biochemical amounts of hyperphosphorylated recombinant CTD, which was then employed to generate and affinity purify antiphosphoCTD antibodies (Lee and Greenleaf 1991; Weeks et al. 1993). These antibodies were used in fluorescence microscopy to investigate the in vivo distribution of RNAPII₀ on *Drosophila* polytene chromosomes. Consistent with the results from Dahmus, this approach demonstrated that sites of active transcription contained RNAPII₀, whereas some inactive genes and promoter-

[Keywords: CTD; RNA polymerase II; cotranscriptional; nuclear organization; phosphorylation]

¹Present address: Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138.

²Corresponding author.

E-MAIL: arno@biochem.duke.edu; FAX (919) 684-8885.

Article is online at <http://www.genesdev.org/cgi/doi/10.1101/gad.1477006>.

proximal sites with paused polymerases contained RNAPIIA (Weeks et al. 1993). Using a different cross-linking method and assay, Lis and colleagues (O'Brien et al. 1994) observed the same distribution. These results reinforced the idea that promoter binding and early events are carried out by RNAPIIA, whereas elongation is carried out by RNAPIIO (Dahmus 1994). Almost all subsequent experiments are consistent with this overall notion. However, it should be kept in mind that some genes may differ from this picture (e.g., Lee and Lis 1998). An example of gene class-specific differences in CTD phosphorylation was already found in 1993: By immunofluorescence, elongating RNAPs on developmentally induced loci in *Drosophila* (ecdysone puffs on polytene chromosomes) were recognized exclusively as "II0" enzymes, whereas RNAPs on stress-induced loci (heat-shock puffs) were recognized as both "II0" and "IIA" forms (Weeks et al. 1993).

It is very important to note that the "II0" designation simply indicates hyperphosphorylation of the CTD (as detected originally by mobility shift of the Rpb1 subunit); RNAPIIO, however, is not necessarily a homogeneous population of molecules. While RNAPIIO does consist of RNAPs with hyperphosphorylated CTDs, the *patterns* of phosphorylation on individual CTDs can vary widely. This variation can be due to differential phosphorylation of Ser2 versus Ser5 residues and/or to differential phosphorylation of repeats along the length of the CTD. As expanded on below, modulating these patterns regulates the affinity of the CTD for its binding partners, and consequently different phosphorylation patterns present at different stages of transcription control the timely recruitment to transcribing RNAPII of factors important for RNA maturation and other events. Much of this review deals with what these CTD phosphorylation patterns may be, how they are created, and what their functional significance is. While the recent past has witnessed significant progress toward answering these questions, our hope is that this review will underscore the point that we have a great deal to learn about virtually every aspect of CTD phosphorylation and function.

Patterns and consequences of CTD phosphorylation

The CTD as a binding scaffold: linking nuclear processes to transcription

If the CTD is not required for catalyzing the synthesis of RNA chains, what does it do? While its purpose was nebulous for some time, it is now clear that a major function of the CTD is to serve as a binding scaffold for a variety of nuclear factors. Since the activities of bound factors become physically associated with RNAPII, the processes they represent become linked to this transcriptase. The early proposal that the PCTD (phosphoCTD) physically links pre-mRNA processing to transcription by tethering processing factors to elongating RNAPII (Corden 1990; Greenleaf 1993) has been borne out experimentally over the last decade (for reviews, see Corden and Patturajan 1997; Goldstrohm et al. 2001; Maniatis

and Reed 2002; Proudfoot et al. 2002). We now understand that the PCTD, via its recruitment of PCTD-binding factors, plays a major role in coordinating a number of nuclear processes with RNA chain synthesis and the translocation of RNAPII along a gene.

The role of CTD phosphorylation in facilitating pre-mRNA processing has thus far been best characterized for 5'-end capping and 3'-end cleavage and polyadenylation. The 7-methyl G5'ppp5'N cap is added when the transcript is ~25 bases long, soon after its 5' end emerges from the exit channel of RNAPII (Jove and Manley 1984; Rasmussen and Lis 1993). That acquisition of such a cap is unique to RNAPII transcripts (Shatkin 1976), and transcripts made by a CTD-less RNAPII are very inefficiently capped (McCracken et al. 1997a), suggested that capping enzyme associates with the transcription complex via interactions with the CTD. An exploration of this hypothesis led to the finding that capping enzyme indeed associates physically with the CTD of RNAPII *in vitro* (Cho et al. 1997; McCracken et al. 1997a). Subsequent cross-linking studies (chromatin immunoprecipitation, or CHIP) showed that capping enzyme also associates with transcribed genes *in vivo*, in a manner that requires CTD phosphorylation; consistent with 5' capping being an early event in the life of a nascent transcript, capping enzyme localizes to genes near their 5' ends (Komarnitsky et al. 2000; Schroeder et al. 2000).

Analogous to capping, the formation of 3' ends of messages is also coupled to transcription by RNAPII through interactions between the CTD and the processing machinery (for a review, see Proudfoot 2004). Attempts to uncover the biochemical basis of this functional link led to the finding that cleavage and polyadenylation factors bind to the PCTD *in vitro* (McCracken et al. 1997b; Birse et al. 1998). CHIP experiments reveal that CF IA, a factor involved in 3'-end formation, accumulates toward the 3' ends of genes (Licatalosi et al. 2002), and its cross-linking is dependent on CTD phosphorylation (Licatalosi et al. 2002; Ahn et al. 2004).

CTD phosphorylation patterns along genes

With the advent of CHIP it became feasible to explore the phosphorylation status of the CTD on RNAPs at different positions along a transcription unit. The commercial availability of anti-CTD monoclonal antibodies (mAbs) with phosphorylation pattern-dependent specificities helped spur these studies. A major finding was that phosphorylation of Ser5 residues predominates near the beginning of genes, whereas polymerases near the ends of genes are extensively phosphorylated on Ser2 residues (Komarnitsky et al. 2000; Morris et al. 2005).

In vivo, Ser5 phosphorylation near the 5' ends of genes depends principally on the kinase activity of TFIIF (Kin28 in yeast; CDK7 in metazoans) (Komarnitsky et al. 2000; Schroeder et al. 2000). *In vitro*, this kinase correspondingly adds phosphates to Ser5 positions of CTD repeats (Hengartner et al. 1998; Sun et al. 1998). Subsequent to the action of TFIIF kinase, Ser2 residues are phosphorylated by CTDK-I in yeast (CDK9 kinase in

metazoans) (Marshall et al. 1996; Lee and Greenleaf 1997; Prelich 2002). Fittingly, *in vitro* CTDK-I preferentially adds phosphates to Ser2 residues of repeats already containing Ser5PO₄ (Jones et al. 2004).

In parallel with its location at the 5' end of genes *in vivo*, capping enzyme binds directly to and its activity is modulated by Ser5P CTD repeats *in vitro* (e.g., E.J. Cho et al. 1998; Ho and Shuman 1999). These findings are consistent with the observed interaction between genes encoding capping enzyme and TFIIH kinase (Rodriguez et al. 2000). Analogously, the accumulation of CF IA at the 3' end of genes depends on CTDK-I (Ahn et al. 2004), the gene for the catalytic subunit of CTDK-I shows genetic interactions with 3'-end-forming factors (Lindstrom and Hartzog 2001; Skaar and Greenleaf 2002), and Pcf11, a subunit of CF IA, binds directly to repeats with Ser2 phosphates (Licatalosi et al. 2002).

This developing picture of CTD phosphorylation, which has revealed most about the two ends of a gene, raises a number of important questions: What is the phosphorylation pattern in the middle of a gene? How does Ser5P transition into Ser2P? Can Ser2P and Ser5P residues coexist in the same heptad? Why are contiguous repeats required for viability? Recent results are beginning to answer some of these questions.

Analogous to the studies on 5'- and 3'-end factors, investigations into PCTD-binding proteins that are found specifically in the body of a gene should provide insights into CTD phosphorylation patterns internal to transcription units. A case in point is the histone methyltransferase Set2, a PCTD-binding protein found at internal sites along transcription units (Li et al. 2002, 2003; Krogan et al. 2003; Schaft et al. 2003; Xiao et al. 2003). Recent data indicate that the histone H3 K36 methyl groups added by Set2 recruit a histone deacetylase that dampens the activity potential of just-transcribed chromatin (Carozza et al. 2005; Joshi and Struhl 2005; Keogh et al. 2005; Chu et al. 2006). Set2-mediated methylation of H3 K36 *in vivo* requires the presence of CTDK-I (Krogan et al. 2003; Xiao et al. 2003). Correspondingly, the PCTD-binding domain of Set2 (the SRI [Set2-Rpb1-interacting] domain) is essential for cotranscriptional methylation (Kizer et al. 2005). The binding specificity of this ~100-amino-acid domain was determined using a series of synthetic CTD peptides of varying length and phosphorylation patterns (Kizer et al. 2005; M. Li et al. 2005). The SRI domain displays several novel and notable requirements for optimal binding: It needs repeats phosphorylated on *both* Ser2 *and* Ser5; it needs at least four phosphate groups; and these phosphates need to be on contiguous heptads. Thus the SRI domain binds optimally to a diheptad comprising doubly phosphorylated repeats.

The SRI domain's binding requirements argue for the presence of doubly phosphorylated (Ser2,5P) repeats on the CTD of elongating RNAPII. The presence of such repeats is also supported by the demonstrated proficiency of CTDK-I at generating 2,5P repeats *in vitro* (Jones et al. 2004). Both of these features are consistent with the requirement of Set2 for the presence of CTDK-I *in vivo*.

What we do and do not know: a working model of CTD phosphorylation and function

Much of the above information has been incorporated into a working model shown in Figure 1. This overview contains many features of recently published models (Sims et al. 2004; Zorio and Bentley 2004), but it differs from the others in explicitly including proteins that bind to Ser2,5P repeats as major components of the elongating complex. This view of RNAPII action implies that the CTD exists in at least four major phosphorylation states during the transcription of a gene. RNAPII at a promoter initially carries a largely unphosphorylated CTD, and the enzyme is associated with a set of factors, such as Mediator, that interact with this form of the CTD. Integrator, a newly described factor involved in snRNA 3'-end processing, is another such multiprotein complex (Baillat et al. 2005). In the case of Mediator, there are contacts between the factor and both the CTD and the body of RNAPII (Asturias et al. 1999; Dotson et al. 2000), although the identity of the Mediator subunits that bind directly to the CTD are not yet known.

Early in the transition from preinitiation to elongation, the CTD is phosphorylated on Ser5 residues to create the second CTD phosphorylation state; 5'-end processing factors, such as capping enzyme, now bind. Since phosphorylation patterns on the CTD ultimately depend on the combined action of CTD kinases and phosphatases, several labs have investigated the specificity of these enzymes (for reviews, see Prelich 2002; Meinhart et al. 2005). The best agreement between *in vitro* enzymatic studies and *in vivo* ChIP results exists for TFIIH kinase at the 5' end of genes. In addition to demonstrating that

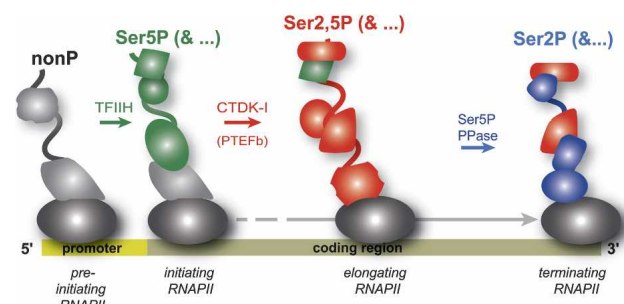


Figure 1. CTD phosphorylation patterns dictate which factors associate with RNAPII. RNAPII (gray oval) is depicted at four positions along a gene, and at each position its CTD (wavy line) is a different color to indicate different phosphorylation states: Gray indicates nonphosphorylated repeats; green indicates Ser5P repeats; red indicates Ser2,5P repeats (doubly phosphorylated); and blue indicates Ser2P repeats. Proteins bound to a type of repeat are indicated in the same color as the repeat. Note that some proteins remain bound as RNAPII changes position on the gene; the repeats to which they are bound (obscured by the protein) are assumed not to change phosphorylation state. Thus, for example, the CTD of “initiating RNAPII” is proposed to comprise two types of repeat: nonP and Ser5P. Analogous reasoning applies to RNAPs at other positions on the gene. As pointed out in the text, the exact number and position of each type of repeat is not known [indicated by “(& . . .)”].

the kinase modifies mainly Ser5 of CTD repeats (Hengartner et al. 1998; Sun et al. 1998), in vitro experiments also indicate that TFIIF acts before the elongation phase-specific kinase (Marshall et al. 1996; Lee and Greenleaf 1997). These results are consistent with ChIP experiments utilizing a mAb (H14) specific for Ser5P, which yield a strong Ser5P signal at the 5' ends of transcription units (Komarnitsky et al. 2000). However, because the CTD is highly repetitive and TFIIF can generate potentially many H14 epitopes, it is currently impossible to know, based solely on ChIP, either the number of Ser5 residues phosphorylated by TFIIF or their positions along the CTD in vivo. That the disposition of Ser5Ps along the CTD is important is suggested by genetic studies of West and Corden (1995). They found that replacing five proximal Ser5 residues in the CTD with Glu residues results in lethality, whereas replacing seven distal Ser5 residues with Glu does not.

After initiation, an elongation-phase kinase (CTDK-I in yeast; P-TEFb in metazoa) (Marshall et al. 1996; Lee and Greenleaf 1997) modifies mainly Ser2 residues to generate elongation-proficient RNAPII; elongation-related factors such as Set2 bind to the CTD in this third state of phosphorylation. (It is not clear whether CTK1 is the actual ortholog of CDK9. A detailed study of evolutionary relationships among CDKs extant in sequence databases of humans, *Drosophila melanogaster*, and several simpler eukaryotes suggests that the yeast protein closest to human and *Drosophila* CDK9 is Bur1 [Liu and Kipreos 2000]. The analysis further suggests that yeast Ctk1 is most related to two uncharacterized human proteins, gi|14110386| and gi|20521690|, and one uncharacterized *Drosophila* protein, gi|24668141|.) Because of the binding specificity of Set2 and because CTDK-I efficiently uses Ser5P repeats to generate Ser2,5P repeats in vitro (Jones et al. 2004), we propose that during elongation the CTD contains repeats phosphorylated at both Ser2 and Ser5 positions. This proposal is not at odds with the observation that the ChIP signal generated by mAb H5 (usually used to detect Ser2P) increases when RNAPII is at sites within transcription units (e.g., Komarnitsky et al. 2000), because mAb H5 actually binds Ser2,5P repeats better than Ser2P repeats (Jones et al. 2004). Also, the H14 (Ser5P-specific mAb) signal persists as RNAPII moves from the 5' end into the gene (Boehm et al. 2003), even though it may decrease (Ahn et al. 2004; Morris et al. 2005); the remaining H14 signal could then indicate the presence of either Ser5P or Ser2,5P repeats (Jones et al. 2004). Furthermore, the initial step in the ChIP procedure, formaldehyde cross-linking, can couple proteins to the CTD (A.L. Greenleaf, unpubl.) and may block an unknown number of epitopes; such epitope masking will reduce the ChIP signal of the cognate antibody by an unknown amount. Overall, extant data do not permit an unequivocal assignment of the number of Ser2,5P repeats or their distribution along the CTD of elongating RNAPII. It will be interesting to see if all proteins that are recruited to the middle of genes via PCTD binding display Set2-like specificity for doubly phosphorylated repeats. This need not be the case, be-

cause there may be a mixture of differently phosphorylated repeats on the CTD at any given time.

Finally, near the 3' end of the gene it is widely believed that CTD phosphorylation is dominated by Ser2P residues; this is consistent with the binding specificity and localization of some 3'-end processing factors (Licatalosi et al. 2002; Ahn et al. 2004). If there are actually fewer Ser5P residues at the 3' end, a Ser5P-specific protein phosphatase must act on the PCTD, as indicated in Figure 1 (for a review, see Meinhart et al. 2005). As with the 5' end and middle of the gene, however, there are caveats to the idea that Ser2 phosphorylation predominates at 3' ends. First, not all genes analyzed by ChIP experiments show a loss of mAb H14 reactivity (Ser5P) at the 3' end (Boehm et al. 2003). Second, the antibody usually used to detect Ser2P repeats (mAb H5) actually reacts better with Ser2,5P repeats, as mentioned (Jones et al. 2004). Third, the repeat nature of the CTD bespeaks a large number of potential phospho-epitopes, and quantifying these is extremely difficult. Fourth, proteins cross-linked to the CTD will block access of the cognate antibodies, altering ChIP signals by an unknown amount. Thus, while the overall idea that Ser2P residues increase in abundance toward the 3' end is likely to be upheld, it will take substantial additional effort to establish the actual number, distribution, and protein occupancy of heptad repeat types on the CTD of RNAPII at the 3' ends of genes.

To stimulate discussion and experiments, we have incorporated some of the facts and caveats discussed above into a set of hypothetical results that relate phosphorylation patterns on different CTDs to antibody signals that might be generated by ChIP analysis of those CTDs (Fig. 2). We emphasize that these results are hypothetical, and are presented to stimulate discussion. CTD #1, for example, comprises eight nonphosphorylated (NP) repeats and nine Ser5P repeats, as might be found on an initiating RNAPII after TFIIF acts on it. There are also CTD-binding proteins cross-linked to CTD #1, obscuring five NP repeats and two 5P repeats. Thus, three NP repeats and seven 5P repeats are available for antibody binding. In a ChIP gedanken experiment we employ the most commonly used anti-CTD mAbs to analyze CTD #1. The resulting signal strengths will be proportional to the number of repeats recognized: three for mAb 8WG16 (NP) and seven for mAb H14 (5P). The mAb H5, which reacts with Ser2P (and Ser2,5P repeats) gives a signal on NP and 5P repeats we will call background (-).

The other example CTDs in Figure 2 are considered analogously, and the resulting signals are presented in the right half of the figure. A noteworthy feature of this exercise is that it illustrates how ChIP signals may be nonconcordant with CTD phosphorylation states (both because signals may not result uniquely from Ser5P or Ser2P residues, and because there may be signal alteration due to epitope masking). In these examples the H14 signal shows a decreasing trend (going from CTD #1 to #4), whereas the H5 signal shows an increasing overall trend, much as in some real experiments. In summary, it seems prudent to keep in mind that the ChIP approach

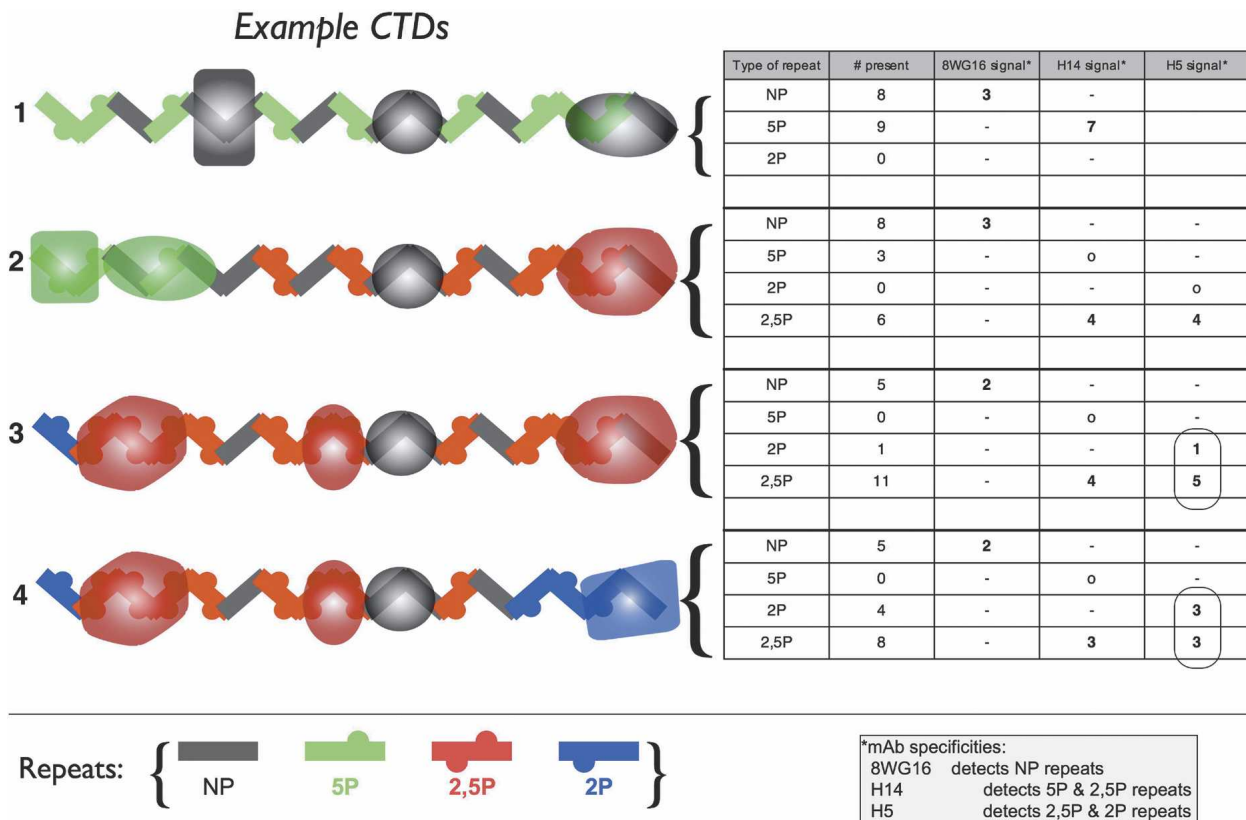


Figure 2. Phosphorylation patterns versus antibody signals for “example” CTDs. Repeats making up hypothetical CTDs (#1–#4) are indicated by colored bars with zero, one, or two knobs (phospho-Ser residues). (Gray) Nonphospho (NP); (green) 5P; (red) 2,5P; (blue) 2P. Partially transparent objects placed over repeats represent proteins that bind to repeats of the same color; this binding is assumed to block antibody access to these repeats. In some cases, binding obscures neighboring repeats of a different type (e.g., right-most gray oval in CTD #1; right-most red protein in CTD #2). The table at the right indicates for each CTD the number of each repeat type present and the ChIP signal that would be generated by that CTD for each of three commonly used mAbs. The box at bottom right, lists the type of repeats detected by each antibody (note that very weak affinities are presumed not to generate a signal (e.g., 8WG16 for Ser5P [Patturajan et al. 1998a; Cho et al. 2001] and H5 for Ser5P [Jones et al. 2004]). (H5 has also been shown to react with phosphorylated SR proteins [Doyle et al. 2002]; in some instances this cross-reactivity could generate spurious ChIP signals.) A notable feature of the results is that the ChIP signals do not necessarily parallel the repeat types present.

alone may not reveal a complete picture of the phospho-modification state of the CTD.

PCTD-associating proteins (PCAPs) and functions of the CTD

The main function of the CTD is to serve as a flexible binding scaffold for a variety of nuclear factors, and binding of a factor links the process it represents to RNAPII. Because factor binding depends on the CTD’s phosphorylation state, which changes as the RNAP moves from one end of a gene to the other, we now understand that CTD phosphorylation plays a major role in coordinating several kinds of nuclear event with RNA chain synthesis and translocation of RNAPII along a gene. This section presents a look at the development of this understanding. Before delving into its binding partners, however, we want to mention briefly some relevant features of the PCTD scaffold itself.

The unique amino acid sequence and restricted amino

acid composition of the CTD engender it with some unusual properties. For example, the CTD is very hydrophilic, and in aqueous solution it has little stable secondary structure (Cagas and Corden 1995; Bienkiewicz et al. 2000); thus, it has the ability to adopt numerous conformations that should enable it to bind cognate factors of different structural types. Its length, if stretched out, is potentially >1200 Å in mammals (1500 Å if the linker region is included); thus there is room for binding of several factors. It is easy to see how the PCTD could tether several discrete functional entities to an elongating RNAPII at the same time. Of evolutionary interest, the CTDs in most animals, plants and fungi contain many identical repeats; it has been proposed that divergence of CTD repeat sequences has been constrained during much of eukaryotic evolution by essential interactions between different CTD-binding factors and canonical CTD repeats (Stiller and Cook 2004; Guo and Stiller 2005).

The first PCAPs and RNA processing

The first systematic effort to identify CTD-binding proteins involved a yeast two-hybrid screen that used part of the mammalian CTD as bait. This screen uncovered two classes of protein (carrying a CTD-interacting domain [CID] either at the N or C terminus), now called SCAFs (SR-like CTD-associated factors) (Yuryev et al. 1996; Conrad et al. 2000). While the functions of the mammalian SCAFs are not yet known, the yeast homolog of one class, Nrd1 protein, in fact, functions in processing of RNAPII transcripts (Steinmetz et al. 2001; Arigo et al. 2006). Following quickly on the heels of finding the SCAFs, several other PCAPs were recognized, largely as a result of studying CTD-truncated RNAPII in mammalian cells. Unexpectedly, RNAPII missing most of its CTD caused more drastic defects in pre-mRNA processing than in transcription itself; this observation led to demonstrations that, for example, cleavage/polyadenylation factors display CTD associations (McCracken et al. 1997b; Barilla et al. 2001; Licatalosi et al. 2002; Maniatis and Reed 2002; Proudfoot et al. 2002; Kyburz et al. 2003). As mentioned earlier, it was found that CF-IA subunit Pcf11 binds directly to the CTD, preferentially to repeats phosphorylated on Ser2 (Licatalosi et al. 2002). Interestingly, the CID of Pcf11 is homologous to that of Nrd1; these and other PCTD-interacting domains (PCIDs) are discussed more in a later section.

RNA processing factors acting at the other end of the gene were also found to bind the PCTD, as several groups demonstrated capping enzyme/PCTD connections (Cho et al. 1997; McCracken et al. 1997a; Yue et al. 1997). Subsequently Shuman and colleagues (Ho and Shuman 1999) showed that whereas mammalian guanylyltransferase (GTase) binds CTD repeats carrying either Ser2P or Ser5P, only the repeats with Ser5P allosterically activate the enzyme. The groups of Buratowski (Komarnitsky et al. 2000) and Bentley (Schroeder et al. 2000) showed that capping enzyme is cross-linkable to chromatin at promoter regions of genes but not at internal or 3'-terminal regions, nicely correlating its localization with its function. Recently the structure of a capping GTase complexed with a PCTD peptide was solved (Fabrega et al. 2003). This structure is a good example of how a flexible CTD can fit into an extended docking site on its binding partner.

Splicing factors also were shown to associate with the phosphorylated form of RNAPII by coprecipitation or colocalization approaches (e.g., Kim et al. 1997; Misteli and Spector 1999); however, these demonstrations could not reveal which factor or subunit contacted the PCTD. Additional splicing factors that bind the PCTD and the transcription-splicing connection are discussed more in a later section.

A few more PCAPs

A number of additional proteins that bind directly to the PCTD were found by diverse approaches. Recently, use of CTD phospho-peptides in an affinity chromatography

approach identified a novel PCTD-binding protein called Rtt103 (YDR289c) that binds specifically to Ser2P repeats, presumably via a domain with homology with the CID of Nrd1 and Pfl1; identifying proteins that interact with Rtt103 led to implicating the exonuclease Rat1 in transcription termination (Kim et al. 2004). Another two-hybrid screen yielded some of the same SCAFs mentioned above, and in addition, revealed a putative prolyl isomerase, SRCyp (Bourquin et al. 1997); as proposed for Ess1 (below) (Morris et al. 1999; Wu et al. 2000), this activity may be involved in modulating the structure of the PCTD and/or its associated factors. Investigations of large human RNAPII complexes revealed that the histone acetyltransferase PCAF interacts with the phosphorylated form of RNAPII (H. Cho et al. 1998); it may be that this association is important for facilitating the movement of RNAPII through chromatin. Along related lines, a multisubunit "Elongator" has been copurified with RNAPII (although direct binding to the CTD has not been demonstrated), and contains among its subunits a HAT activity (Otero et al. 1999; Wittschleben et al. 1999). Interestingly, genetic studies (Jona et al. 2001) revealed interactions between genes encoding Elongator subunits and *CTK1*, the gene encoding the catalytic subunit of CTDK-I. A number of other proteins of disparate function have been shown to bind directly to the PCTD, including the KRAB/Cys2-His2 zinc finger protein ZNF74 (Grondin et al. 1997), and the splicing and transcription-associated proteins PSF and p54^{nrb}/NonO (Emili et al. 2002). Interestingly, it should be noted that a pool of nontranscribing RNAPII carries a PCTD and is associated with certain transcription and processing factors in potential assembly areas called "transcriptosomes" (cf. Gall 2000).

Not just RNA processing anymore: many more PCAPs and functions

Because most of the PCAPs mentioned to this point were not discovered in a systematic way, and because the original and subsequent two-hybrid screens picked up only a very small number of PCAPs (Yuryev et al. 1996; Bourquin et al. 1997; Guo et al. 2004), the existence of additional PCAPs seemed very likely. A biochemical approach that included PCTD direct-binding assays and affinity-matrix purification procedures was applied to yeast extracts and resulted in identification of novel PCAPs. The initial group included a prolyl isomerase (Ess1), a splicing factor (Prp40), and a ubiquitin ligase (Rsp5) (Morris et al. 1999; Morris and Greenleaf 2000). Subsequently, an improved approach, applied on a larger scale, revealed >100 proteins in the yeast proteome that are retained specifically on an affinity matrix carrying a synthetic three-repeat CTD peptide in which both Ser2 and Ser5 residues of each repeat are phosphorylated ("2,5P" peptide column) (Phatnani et al. 2004); recall that this is the pattern preferentially generated by CTDK-I. A striking feature of the proteins identified by this approach is the number of functional classes into which they fall. In addition to pre-mRNA RNA process-

ing factors, proteins recovered from the 2,5P peptide column represent factors with known or proposed roles in transcription, chromatin structure modification, DNA damage/repair, protein degradation, protein synthesis, RNA degradation, snRNA modification, and snoRNP biogenesis.

The idea that PCTD binding by these proteins is functionally meaningful has been investigated to date for several of the proteins; most is currently known about the histone methyltransferase, Set2, whose PCTD-mediated link to elongating RNAPII was described earlier. It is worth reiterating that binding studies with recombinant Set2 constructs demonstrated that its SRI domain binds with high selectivity to Ser2,5P CTD repeats (Kizer et al. 2005). NMR structure determination (M. Li et al. 2005; Vojnic et al. 2006) together with point mutagenesis and phospho-peptide binding studies (M. Li et al. 2005, see a later section) provide a molecular picture of the SRI domain and suggest how it binds to 2,5P CTD repeats, tethering Set2 to elongating RNAPII. Recent ideas about Set2 function (Carozza et al. 2005; Joshi and Struhl 2005; Keogh et al. 2005; Chu et al. 2006) tie in nicely with the notion that it is part of a transcription elongation mega-complex.

A few of the other proteins identified in the biochemical search through the yeast proteome have already been shown to bind directly to the PCTD; these include Ess1, Prp40, Ssd1, and Hrr25 (Morris et al. 1999; Morris and Greenleaf 2000; Phatnani et al. 2004). Interestingly, the binding domains of three of these proteins bind best to Ser2,5P repeats, whereas the binding domain of Ssd1 binds equally well to Ser2,5P and Ser2P repeats (Phatnani et al. 2004); the functional significance of these specificities have yet to be explored *in vivo*. Of this group, the protein with perhaps the most novel implications is Hrr25, a protein kinase involved in response to DNA damage (Ho et al. 1997). The selective binding of Hrr25 to 2,5P repeats suggests a role in DNA damage responses for RNAPII carrying repeats phosphorylated in this pattern. This suggestion is in line with published studies showing that *ctk1Δ* strains are sensitive to certain DNA damaging agents (Ostapenko and Solomon 2003). On the other hand, PCTD-associated Hrr25 may be involved in other processes. For example, recent results implicate Hrr25 and two other 2,5P-binding proteins (Enp1, Tsr1) in ribosome biogenesis (Schafer et al. 2006).

Another kinase isolated as a 2,5P-binding protein is Hog1, a stress-activated protein kinase that plays an essential role in adaptation to conditions of high osmolarity. Interestingly, it has recently been found to interact directly with Rpb1 in a manner that appears to depend both on CTD phosphorylation and osmotic stress. Moreover, in osmotically stressed cells, Hog1 can be cross-linked to the coding regions of osmoregulated genes (Proft et al. 2006). It will be interesting to see whether the recruitment and/or activity of Hog1 is dependent on its direct binding to the PCTD. Along similar lines, preliminary studies on Cbf5, another protein isolated by 2,5P-affinity chromatography, have now revealed that it

is a direct-binding PCAP (R.J. Boruta, H.P. Phatnani, and A.L. Greenleaf, unpubl.). Cbf5 is a component of the H/ACA snoRNP, an RNA/protein particle that converts certain U residues in rRNAs and other RNAs to pseudouridine (Meier 2005). This result fits nicely with the developing understanding of cotranscriptional assembly of the H/ACA snoRNP, and especially with the contemporaneous discovery that the process depends on CTDK-I (Ballarino et al. 2005; Yang et al. 2005). It also provides another example supporting the view that many of the proteins isolated by 2,5P-affinity chromatography will ultimately be found to interact with the PCTD in a functionally meaningful way.

Unexpectedly, a large number of the 2,5P-repeat-binding proteins have known or proposed roles in protein synthesis or degradation. In addition there are several proteins with connections to the proteasome or with potential chaperone-like functions. These associations are consistent with a role for the PCTD in a cotranscriptional process that involves translation (Iborra et al. 2001; Brogna et al. 2002); alternatively, they may bind the PCTD in the course of executing other functions (e.g., some ribosomal proteins regulate splicing of their own message [Dabeva and Warner 1993; Fewell and Woolford 1999; Warner 1999]; others perform different extraribosomal functions [Wool 1996; Jeffery 2003; Zimmermann 2003]). If some kind of nuclear translation does occur (Dahlberg and Lund 2004; Iborra et al. 2004b), it likely participates in nonsense-mediated decay (Iborra et al. 2004a). Consistent with this idea, recent systems biology analyses strongly suggest connections between the transcription/pre-mRNA processing/export machinery and the NMD machinery (Maciag et al. 2006).

PCIDs and binding modes

How do factors actually bind to the CTD? How do factors bind to PCTDs with different phosphorylation states? Is the conformation of the CTD the same when bound to different proteins? Are (P)CIDs all related? These and other outstanding questions are being answered as structural information about PCTD-interacting proteins is generated (e.g., Meinhart et al. 2005).

Pin1, a prolyl isomerase

The first structure of a PCAP, mammalian Pin1 (homologous to yeast Ess1), was reported in 1997 (Ranganathan et al. 1997), and the structure of a fungal homolog was reported more recently (Z. Li et al. 2005). Pin1/Ess1 consists of a prolyl isomerase domain and a small N-terminal WW domain. In the mammalian protein the domains are coupled via a flexible linker, whereas the fungal enzyme has a more rigid linker, leading to a different spatial relationship between the domains in the two structures. The structure of mammalian Pin1 complexed with a CTD one-repeat phospho-peptide shows that the peptide, which assumes an extended coil-like conformation, contacts only the WW domain (Verdecia

et al. 2000). Curiously, while the peptide was phosphorylated on both Ser2 and Ser5, only the phosphate on Ser5 made contact with the protein. One explanation for this may be that with only seven amino acids (YSPTSPS; phosphorylated serines are underlined) the peptide was not long enough to contain the actual *in vivo* binding epitope, which might extend across the canonical repeat boundaries (consideration of repeat “phasing” is discussed in Greenleaf 2003). Later binding studies with yeast Ess1 WW domain, using multirepeat CTD peptides, indicate that it does have a strong preference for Ser2,5P repeats (Phatnani et al. 2004). It may be that because the peptides used in these later tests comprised three canonical repeats they contained binding epitopes that extend across repeat boundaries.

The 5' end: capping enzyme

The next structure of a complex between the PCTD and a cognate binding protein was that of Cgt1 capping GTase bound to a four-repeat peptide carrying phosphates on each Ser5 residue (Fabrega et al. 2003). In this case, both the type of protein bound and the phosphorylation pattern on the CTD peptide were different from the earlier studies. Nevertheless, a number of the questions posed earlier were answered in these experiments, the results of which included many novel findings. For example, unlike Pin1, the region of Cgt1 that binds the PCTD is not a separate domain; rather, it is a part of the GTase domain, but separate from the active site. The PCTD binds in a long groove that extends some 40 Å along the protein surface. At each end there is a “docking site” that binds a Ser5-PO₄ and several adjacent residues of the CTD peptide. Each docking site makes critical contacts with highly conserved CTD repeat residues, such as the almost invariant Tyr, in addition to the Ser5-PO₄. An important feature of docking site 1 is that it binds residues from two consecutive canonical repeats; thus, observing this binding mode depended on using a peptide containing more than one canonical repeat. Very interestingly, the central portion of the three repeats involved in interacting with the Cgt1 protein loops out from the surface of the protein and does not participate directly in binding. The conformation of the CTD segments bound to Cgt1 contrasts with that in the Pin1 complex by not being coiled, but extended and nonhelical. Thus the first two structures solved demonstrated that not all PCIDs are the same and also that the flexible CTD sequence can adopt different conformations in binding to different proteins.

The 3' end: polyA/cleavage factor Pcf11

Yet another mode of binding was observed when the structure of the CID of yeast Pcf11 was determined. The CID is a domain of ~140 amino acids discovered a decade ago (Yuryev et al. 1996). In some proteins, such as yeast Pcf11, the CID shows binding specificity for repeats carrying Ser2P (Licatalosi et al. 2002). In other proteins,

however, it apparently can have a different binding specificity; mammalian SCAF8, for example, binds best to doubly phosphorylated repeats (Patturajan et al. 1998b). Meinhart and Cramer (2004) solved the crystal structure of the Pcf11 CID by itself, and then they soaked in a 12-residue repeat peptide in which the central Ser2 was phosphorylated. Unexpectedly, the Ser2 phosphate group does not make any contacts with the CID. Noble et al. (2005), who also recently solved the CID structure, determined that the K_D for a similar peptide was ~180 μM; intriguingly, Hollingworth et al. (2006) found that RNA also binds weakly to the CID of Pcf11, displaying an apparent competition with CTD phospho-peptides. It will be interesting to see if this competition is functionally significant for 3'-end processing *in vivo*. In addition, experiments utilizing differently phosphorylated peptides will be important in comparing binding modes for CIDs with homologous structures but different binding specificities (e.g., Pcf11 and SCAF8).

The middle: FF domains and SRI domain

The solution structures of two other classes of PCID have been solved by NMR methods, and additional binding motifs and modes have emerged. Certain FF domains, ~50-residue motifs characterized by two conserved Phe residues (Bedford and Leder 1999), were shown to bind to the PCTD (Carty et al. 2000; Morris and Greenleaf 2000). Interestingly, the FF1 domain from the yeast splicing factor Prp40 (FF1_{Prp40}) has a 3D structure extremely similar to that of the FF1 domain from the mammalian splicing-related protein HYPA/FBP11 (FF1_{FBP11}), but its binding specificity is very different (Allen et al. 2002; Gasch et al. 2005). Whereas FF1_{FBP11} binds the PCTD, FF1_{Prp40} instead binds to N-terminal TPR repeats of the multifunctional yeast protein Clf1. Such differences are not too surprising, since the amino acid sequences of different FF domains are poorly conserved. For example, Gasch et al. (2005) carried out a phylogenetic analysis of FFs from splicing-related factors and found that they could be placed in several different subgroups. To a large extent, this grouping placed FF domains with similar pK_as together. As might be expected for a domain that binds the negatively charged PCTD, FF1_{FBP11} has a basic pK_a of 9.6; in contrast the FF1_{Prp40} that does not bind the PCTD has a pK_a of 4.7. The other individual FF domain previously shown to bind the PCTD, FF5 of CA150 (Carty et al. 2000), has a pK_a of 9.1, consistent with this analysis. It will be interesting to see if all basic FF domains, dispersed among different proteins, bind the PCTD. Also, the identification of the binding partners of the other FF domain classes (neutral and acidic) should be very informative.

The solution structure of another small domain, which is found only in one class of chromatin-modifying enzyme, was solved recently. The histone methyltransferase Set2 contains at its C terminus, a 100 residue domain that binds the PCTD, tethering Set2 to elongating RNAPII and coupling methylation of Lys 36 in histone H3 to transcription elongation (Kizer et al. 2005). Human

and yeast SRI domains are structurally quite similar even though the amino acids sequences are only ~20% identical [M. Li et al. 2005; Vojnic et al. 2006]. NMR resonance perturbation experiments suggest that the PCTD-binding sites are similarly positioned on the two domains, which have similar binding characteristics [M. Li et al. 2005; Vojnic et al. 2006].

As additional PCIDs are identified and their structures are solved, it will be instructive to see how many families of PCID there are, how they are distributed among factors of different functions, and how they bind to the PCTD. Elucidating this structural information and combining it with functional studies will be important for filling in gaps in our understanding of the CTD and its functions.

The PCTD as a major organizer of nuclear functions

Summary and overview

We have seen that the CTD of actively transcribing RNAPII is phosphorylated at multiple sites, and that the pattern of phosphorylation changes as polymerase traverses a transcription unit; in turn, different phosphorylation patterns recruit different proteins to the

CTD. Thus, during the process of RNA chain synthesis the PCTD orchestrates formation of a megacomplex that is linked to RNAPII. However, the elongation megacomplex is not static in composition, but changes components and capabilities as RNAPII moves through different regions of a gene. A simplified overview of these events was depicted in the model of Figure 1. A more detailed snapshot of a hypothetical elongation megacomplex in the middle of a gene is shown in Figure 3.

As depicted in Figure 3, a fully phosphorylated CTD is likely to extend multiple diameters out from the globular portion of RNAPII (a stretched-out yeast CTD would extend ~650 Å [Meinhart et al. 2005], and the mammalian CTD is twice as long; the diameter of the globular portion of the enzyme is ~150 Å [Cramer et al. 2001]). There is thus ample room for simultaneous binding of multiple PCAPs. The phosphorylation pattern shown is not homogeneous because the distribution of phosphates along the CTD at any given time is not known. For this and other reasons, the PCAPs in the figure are arbitrarily positioned along the CTD. An additional consequence of CTD length is that factors bound to it can easily reach either the nascent transcript (e.g., to tether exons in the course of splicing [Goldstrohm et al. 2001; Maniatis and Reed 2002; Takahara et al. 2005; Dye et al. 2006]) or the

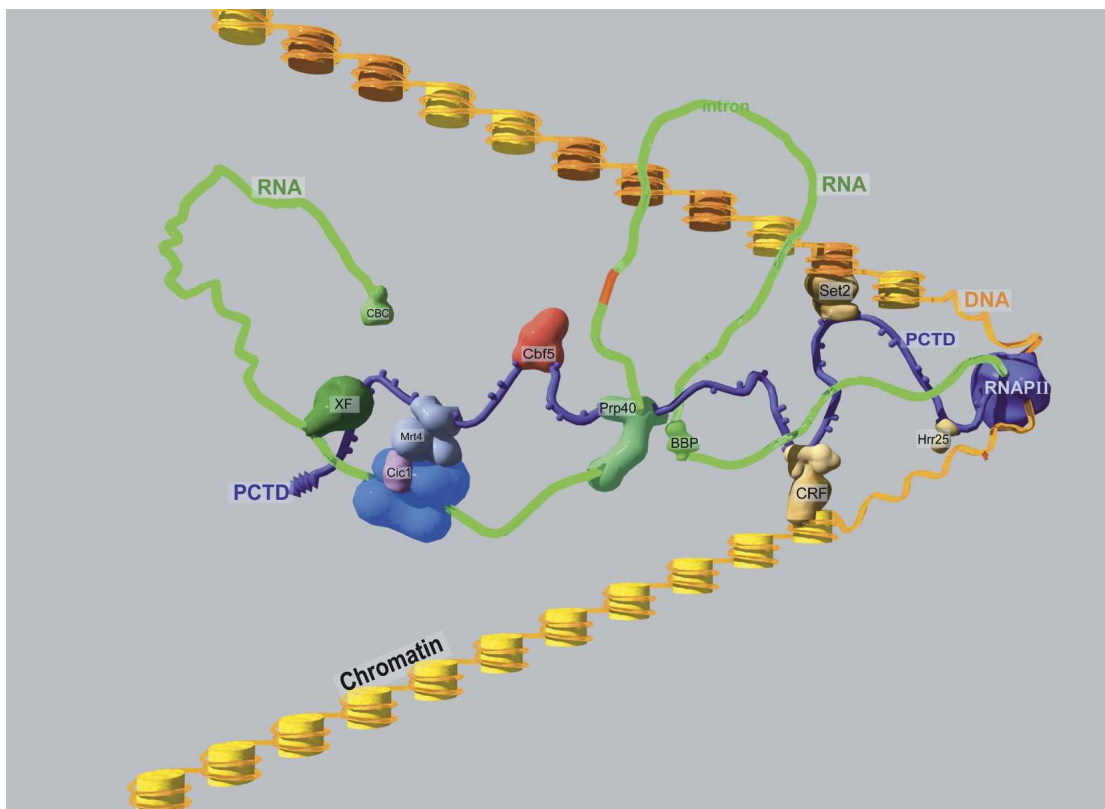


Figure 3. A hypothetical RNAPII elongation megacomplex. RNAPII (including the extended CTD with SerPO₄ knobs) is purple; the globular and CTD portions are drawn approximately to scale for mammalian RNAPII. Orange DNA is wrapped around yellow nucleosomal histones; nucleosomes modified by Set2 are shaded darker. The nascent RNA transcript is green. Yeast names are used for PCAPs [e.g., Phatnani et al. 2004], not all of which are shown. (CBC) cap-binding complex; (CRF) chromatin remodeling factor; (XF) processing/export factor. See text for additional discussion.

DNA/chromatin template (e.g., to modify nearby nucleosomes [Kizer et al. 2005]).

Considering first the DNA/chromatin template, the histone methyltransferase Set2 is depicted as simultaneously contacting the PCTD and a nucleosome near transcribing RNAPII, since strong evidence exists that Set2 binds directly to the PCTD via its SRI domain and cotranscriptionally modifies histone H3 in nucleosomes (discussed above). A chromatin remodeling factor (CRF) is included to represent potentially CTD-bound factors that modify chromatin structure to facilitate transcription by RNAPII. Hrr25, a PCTD-interacting protein implicated in the response to DNA damage, is also shown attached to the PCTD, where it might receive a signal from the polymerase that DNA damage has been encountered (red adduct about to enter RNAPII).

As for the RNA transcript, we show Prp40 binding to the PCTD, positioning its associated U1snRNP to recognize a 5' splice site in the transcript, tethering it to the PCTD until branchpoint-binding protein (BBP) and the associated 3' splice site are encountered. We have also positioned the H/ACA snoRNP component Cbf5 such that it can access a hypothetical intron-encoded small nucleolar RNA (snoRNA) (red recognition sequence indicated in RNA) to initiate cotranscriptional snoRNP assembly (discussed above). Analogously, it may be that U3 snoRNP assembly also begins cotranscriptionally, since we found two of its components, Utp20 and Rrp5, in the collection of proteins bound to doubly phosphorylated CTD repeats (Phatnani et al. 2004). In the hypothetical megacomplex, we have also included putative PCAPs involved in proteasome function (e.g., Cic1) and RNA degradation (e.g., Mrt4). Also present is a representative factor (XF) that links RNA processing to nuclear export (for a review, see Maniatis and Reed 2002). Finally, we point out that the order in which these interactions occur is not known; for instance, PCAPs that also bind RNA (e.g., Prp40, Cbf5) could bind the PCTD either before or after binding their cognate RNA. In addition, PCAP binding to the PCTD may be stabilized by interactions with other components of the megacomplex (e.g., Phatnani et al. 2004, and its Supplementary Tables 1, 2).

The snapshot of the elongation megacomplex illustrated in Figure 3 represents only one of many possible configurations. Because of the number of possible phosphorylation sites in the CTD and the existence of multiple CTD kinases and phosphatases (and one or more prolyl isomerases), the extent and pattern of CTD phosphorylation potentially can be modulated to generate a vast number of different phospho-epitopes (Sudol et al. 2001; Buratowski 2003). The arrays of binding sites thus generated have the potential to recruit many combinations of binding partners. We propose that this feature contributes to remodeling or fine-tuning the functional capabilities of the RNAPII elongation megacomplex in response to different signals. These signals could indicate, for example, position along the transcription unit, presence and nature of introns (Batsche et al. 2006), alterations in cellular physiology, actions of gene-specific

modulators, or presence of DNA damage. Determining the nature of such signals and how they function should form the basis for exciting future experiments.

Open questions

The past decade has seen significant progress in our understanding of the CTD and its interacting factors, but crucial questions about many aspects of CTD phosphorylation and function remain open. Among many important questions about the CTD that remain to be answered are the following baker's dozen:

1. What initiates CTD phosphorylation, and to which repeats does the CTD kinase activity in TFIIFH add phosphates?
2. Which factors bound to (unphosphorylated repeats of) preinitiating RNAPII are displaced due to TFIIFH kinase activity? Which remain bound? When do they dissociate? What causes their dissociation?
3. How is CTDK-I (P-TEFb) recruited to RNAPII, and what Ser residues does it initially phosphorylate? Ultimately, what is the complete set of residues phosphorylated by CTDK-I?
4. Mechanistically, how does phosphorylation of CTD residues by CTDK-I lead to dissociation of some factors and recruitment of others?
5. Does a factor (PCAP) always bind to the same place on the CTD? That is, is there a particular order of factors bound along the length of the CTD? If it exists, what determines such an order?
6. In animal CTDs, what is the significance of non-canonical repeats? For example, the distal half of all mammalian CTDs contain nine repeats with Arg or Lys in position seven (Barron-Casella and Corden 1992); are these repeats less constrained evolutionarily (Guo and Stiller 2005)? Do they recruit specific factors?
7. Is the PCID of a particular PCAP sufficient for in vivo recruitment to the PCTD? Or, does association of a PCAP with elongating RNAPII depend on additional interactions?
8. When do CTD phosphatases act? Which specific phosphates are removed, and under what conditions? What are the direct consequences of each type of dephosphorylation?
9. How dynamic is the phosphorylation pattern on the CTD during elongation? How does the pattern (and its dynamics) on one class of genes compare with that on another class?
10. How are the activities of kinases and phosphatases coordinated to produce changes in CTD phosphorylation patterns? Do signals encountered by elongating RNAPII—for example, DNA damage within transcription units—result in changes in phosphorylation and remodeling of the PCTD-associated proteome?
11. Do RNAPII molecules on different genes have different factors bound to the PCTD?
12. How does the association of a factor with the PCTD influence its function? Does PCTD recruitment sim-

ply serve to increase local concentration of the factor, or, as for capping enzyme, does binding to the PCTD modulate specific activity?

13. How are signals sent and received by proteins bound to the PCTD of an RNAPII molecule? For example, if the globular catalytic portion of RNAPII encounters DNA damage that blocks transcription, how is that information conveyed to other components of the elongation complex?

We have listed these questions because it is useful to realize what we still do *not* know about the CTD. Keeping these unknowns in mind should both guide interpretations of experimental data and help stimulate new experiments. We have made great strides in the last 10 yr, but the amount left to learn suggests that the next decade of CTD investigation will be at least as productive and even more exciting.

Acknowledgments

Special thanks to Janice Jones for advice and assistance. Our sincere thanks to John Stiller, Craig Bennett, and members of the Greenleaf laboratory for helpful discussions and suggestions. We thank Ashley Greenleaf for Figure 3. Work in the Greenleaf laboratory was supported by the NIH. We apologize to those colleagues whose work could not be cited.

References

- Ahn, S.H., Kim, M., and Buratowski, S. 2004. Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Mol. Cell* **13**: 67–76.
- Akoulitchev, S., Makela, T.P., Weinberg, R.A., and Reinberg, D. 1995. Requirement for TFIIF kinase activity in transcription by RNA polymerase II. *Nature* **377**: 557–560.
- Allen, M., Friedler, A., Schon, O., and Bycroft, M. 2002. The structure of an FF domain from human HYPB/FBP11. *J. Mol. Biol.* **323**: 411–416.
- Arigo, J.T., Carroll, K.L., Ames, J.M., and Corden, J.L. 2006. Regulation of yeast NRD1 expression by premature transcription termination. *Mol. Cell* **21**: 641–651.
- Asturias, F.J., Jiang, Y.W., Myers, L.C., Gustafsson, C.M., and Kornberg, R.D. 1999. Conserved structures of mediator and RNA polymerase II holoenzyme. *Science* **283**: 985–987.
- Baillat, D., Hakimi, M.A., Naar, A.M., Shilatifard, A., Cooch, N., and Shiekhattar, R. 2005. Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* **123**: 265–276.
- Ballarino, M., Morlando, M., Pagano, F., Fatica, A., and Bozzoni, I. 2005. The cotranscriptional assembly of snRNPs controls the biosynthesis of H/ACA snoRNAs in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **25**: 5396–5403.
- Barilla, D., Lee, B.A., and Proudfoot, N.J. 2001. Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci.* **98**: 445–450.
- Barron-Casella, E. and Corden, J.L. 1992. Conservation of the mammalian RNA polymerase II largest-subunit C-terminal domain. *J. Mol. Evol.* **35**: 405–410.
- Batsche, E., Yaniv, M., and Muchardt, C. 2006. The human SWI/SNF subunit Brm is a regulator of alternative splicing. *Nat. Struct. Mol. Biol.* **13**: 22–29.
- Bedford, M.T. and Leder, P. 1999. The FF domain: A novel motif that often accompanies WW domains. *Trends Biochem. Sci.* **24**: 264–265.
- Bienkiewicz, E.A., Moon Woody, A., and Woody, R.W. 2000. Conformation of the RNA polymerase II C-terminal domain: Circular dichroism of long and short fragments. *J. Mol. Biol.* **297**: 119–133.
- Birse, C.E., Minvielle-Sebastia, L., Lee, B.A., Keller, W., and Proudfoot, N.J. 1998. Coupling termination of transcription to messenger RNA maturation in yeast. *Science* **280**: 298–301.
- Boehm, A.K., Saunders, A., Werner, J., and Lis, J.T. 2003. Transcription factor and polymerase recruitment, modification, and movement on dhsp70 in vivo in the minutes following heat shock. *Mol. Cell Biol.* **23**: 7628–7637.
- Bourquin, J.P., Stagljar, I., Meier, P., Moosmann, P., Silke, J., Baechli, T., Georgiev, O., and Schaffner, W. 1997. A serine/arginine-rich nuclear matrix cyclophilin interacts with the C-terminal domain of RNA polymerase II. *Nucleic Acids Res.* **25**: 2055–2061.
- Brogna, S., Sato, T.A., and Rosbash, M. 2002. Ribosome components are associated with sites of transcription. *Mol. Cell* **10**: 93–104.
- Buratowski, S. 2003. The CTD code. *Nat. Struct. Mol. Biol.* **10**: 679–680.
- Buratowski, S. and Sharp, P.A. 1990. Transcription initiation complexes and upstream activation with RNA polymerase II lacking the C-terminal domain of the largest subunit. *Mol. Cell Biol.* **10**: 5562–5564.
- Cadena, D.L. and Dahmus, M.E. 1987. Messenger RNA synthesis in mammalian cells is catalyzed by the phosphorylated form of RNA polymerase II. *J. Biol. Chem.* **262**: 12468–12474.
- Cagas, P.M. and Corden, J.L. 1995. Structural studies of a synthetic peptide derived from the carboxyl-terminal domain of RNA polymerase II. *Proteins* **21**: 149–160.
- Carrozza, M.J., Li, B., Florens, L., Sugauma, T., Swanson, S.K., Lee, K.K., Shia, W.J., Anderson, S., Yates, J., Washburn, M.P., et al. 2005. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **123**: 581–592.
- Carty, S.M., Goldstrohm, A., Suñe, C., Garcia-Blanco, M.A., and Greenleaf, A.L. 2000. Protein-interaction modules that organize nuclear function: FF domains of CA150 bind the phospho-CTD of RNA polymerase II. *Proc. Natl. Acad. Sci.* **97**: 9015–9020.
- Cho, E.J., Takagi, T., Moore, C.R., and Buratowski, S. 1997. mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes & Dev.* **11**: 3319–3326.
- Cho, E.J., Rodriguez, C.R., Takagi, T., and Buratowski, S. 1998. Allosteric interactions between capping enzyme subunits and the RNA polymerase II carboxy-terminal domain. *Genes & Dev.* **12**: 3482–3487.
- Cho, H., Orphanides, G., Sun, X., Yang, X.J., Ogryzko, V., Lees, E., Nakatani, Y., and Reinberg, D. 1998. A human RNA polymerase II complex containing factors that modify chromatin structure. *Mol. Cell Biol.* **18**: 5355–5363.
- Cho, E.J., Kobor, M.S., Kim, M., Greenblatt, J., and Buratowski, S. 2001. Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. *Genes & Dev.* **15**: 3319–3329.
- Chow, L.C., Gelinis, R.E., Broker, T.R., and Roberts, R.J. 1977. An amazing sequence arrangement at the 5' ends of adeno-

- virus 2 messenger RNA. *Cell* **12**: 1–8.
- Chu, Y., Sutton, A., Sternglanz, R., and Prelich, G. 2006. The BUR1 cyclin-dependent protein kinase is required for the normal pattern of histone methylation by SET2. *Mol. Cell. Biol.* **26**: 3029–3038.
- Conrad, N.K., Wilson, S.M., Steinmetz, E.J., Patturajan, M., Brow, D.A., Swanson, M.S., and Corden, J.L. 2000. A yeast heterogeneous nuclear ribonucleoprotein complex associated with RNA polymerase II. *Genetics* **154**: 557–571.
- Corden, J.L. 1990. Tails of RNA polymerase II. *Trends Biol. Sci.* **15**: 383–387.
- Corden, J.L. and Patturajan, M. 1997. A CTD function linking transcription to splicing. *Trends Biochem. Sci.* **22**: 413–416.
- Cramer, P., Bushnell, D.A., and Kornberg, R.D. 2001. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**: 1863–1876.
- Dabeva, M.D. and Warner, J.R. 1993. Ribosomal protein L32 of *Saccharomyces cerevisiae* regulates both splicing and translation of its own transcript. *J. Biol. Chem.* **268**: 19669–19674.
- Dahlberg, J.E. and Lund, E. 2004. Does protein synthesis occur in the nucleus? *Curr. Opin. Cell Biol.* **16**: 335–338.
- Dahmus, M.E. 1994. The role of multisite phosphorylation in the regulation of RNA polymerase II activity. *Prog. Nucleic Acid Res. Mol. Biol.* **48**: 143–179.
- Dahmus, M.E. 1995. Phosphorylation of the C-terminal domain of RNA polymerase II. *Biochim. Biophys. Acta* **1261**: 171–182.
- Dahmus, M. 1996. Reversible phosphorylation of the C-terminal domain of RNA polymerase II. *J. Biol. Chem.* **271**: 19009–19012.
- Dahmus, M.E. and Dynan, W.S. 1992. Phosphorylation of RNA polymerase II as a transcriptional regulatory mechanism. In *Transcriptional regulation* (eds. K. Yamamoto and S. McKnight), pp. 109–129. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
- Dotson, M.R., Yuan, C.X., Roeder, R.G., Myers, L.C., Gustafson, C.M., Jiang, Y.W., Li, Y., Kornberg, R.D., and Asturias, F.J. 2000. Structural organization of yeast and mammalian mediator complexes. *Proc. Natl. Acad. Sci.* **97**: 14307–14310.
- Doyle, O., Corden, J.L., Murphy, C., and Gall, J.G. 2002. The distribution of RNA polymerase II largest subunit (RPB1) in the *Xenopus* germinal vesicle. *J. Struct. Biol.* **140**: 154–166.
- Dye, M.J., Gromak, N., and Proudfoot, N.J. 2006. Exon tethering in transcription by RNA polymerase II. *Mol. Cell* **21**: 849–859.
- Emili, A., Shales, M., McCracken, S., Xie, W., Tucker, P.W., Kobayashi, R., Blencowe, B.J., and Ingles, C.J. 2002. Splicing and transcription-associated proteins PSF and p54nrb/nonO bind to the RNA polymerase II CTD. *RNA* **8**: 1102–1111.
- Fabrega, C., Shen, V., Shuman, S., and Lima, C.D. 2003. Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Mol. Cell* **11**: 1549–1561.
- Fewell, S.W. and Woolford Jr., J.L. 1999. Ribosomal protein S14 of *Saccharomyces cerevisiae* regulates its expression by binding to RPS14B pre-mRNA and to 18S rRNA. *Mol. Cell. Biol.* **19**: 826–834.
- Gall, J.G. 2000. Cajal bodies: The first 100 years. *Annu. Rev. Cell Dev. Biol.* **16**: 273–300.
- Gasch, A., Wiesner, S., Martin-Malpartida, P., Ramirez-Espain, X., Ruiz, L., and Macias, M.J. 2005. The structure of PRP40 FF1 domain and its interaction with the CRN-TPR1 motif of CLF1 gives a new insight into the binding mode of FF domains. *J. Biol. Chem.* **281**: 356–364.
- Goldstrohm, A.C., Greenleaf, A.L., and Garcia-blanco, M.A. 2001. Co-transcriptional splicing of Pre-messenger RNAs: Considerations for the mechanism of alternative splicing. *Gene* **277**: 31–47.
- Greenleaf, A.L. 1992. RNA polymerase II subunits, transcription factors, and kinases: Investigations in *Drosophila melanogaster* and *Saccharomyces cerevisiae*. In *Transcriptional regulation* (eds. K. Yamamoto and S. McKnight), pp. 55–80. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
- Greenleaf, A.L. 1993. Positive patches and negative noodles: Linking RNA processing to transcription? *Trends Biochem. Sci.* **18**: 117–119.
- Greenleaf, A. 2003. Getting a grip on the CTD of Pol II. *Structure* **11**: 900–902.
- Grondin, B., Cote, F., Bazinet, M., Vincent, M., and Aubry, M. 1997. Direct interaction of the KRAB/Cys2-His2 zinc finger protein ZNF74 with a hyperphosphorylated form of the RNA polymerase II largest subunit. *J. Biol. Chem.* **272**: 27877–27885.
- Guo, Z. and Stillier, J.W. 2005. Comparative genomics and evolution of proteins associated with RNA polymerase II C-terminal domain. *Mol. Biol. Evol.* **22**: 2166–2178.
- Guo, D., Hazbun, T.R., Xu, X.J., Ng, S.L., Fields, S., and Kuo, M.H. 2004. A tethered catalysis, two-hybrid system to identify protein–protein interactions requiring post-translational modifications. *Nat. Biotechnol.* **22**: 888–892.
- Hengartner, C.J., Myer, V.E., Liao, S.M., Wilson, C.J., Koh, S.S., and Young, R.A. 1998. Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases. *Mol. Cell* **2**: 43–53.
- Ho, C.K. and Shuman, S. 1999. Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. *Mol. Cell* **3**: 405–411.
- Ho, Y., Mason, S., Kobayashi, R., Hoekstra, M., and Andrews, B. 1997. Role of the casein kinase I isoform, Hrr25, and the cell cycle-regulatory transcription factor, SBF, in the transcriptional response to DNA damage in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci.* **94**: 581–586.
- Hollingworth, D., Noble, C.G., Taylor, I.A., and Ramos, A. 2006. RNA polymerase II CTD phosphopeptides compete with RNA for the interaction with Pcf11. *RNA* **12**: 555–560.
- Iborra, F.J., Jackson, D.A., and Cook, P.R. 2001. Coupled transcription and translation within nuclei of mammalian cells. *Science* **293**: 1139–1142.
- Iborra, F.J., Escargueil, A.E., Kwek, K.Y., Akoulitchev, A., and Cook, P.R. 2004a. Molecular cross-talk between the transcription, translation, and nonsense-mediated decay machineries. *J. Cell Sci.* **117**: 899–906.
- Iborra, F.J., Jackson, D.A., and Cook, P.R. 2004b. The case for nuclear translation. *J. Cell Sci.* **117**: 5713–5720.
- Jeffery, C.J. 2003. Moonlighting proteins: Old proteins learning new tricks. *Trends Genet.* **19**: 415–417.
- Jona, G., Wittschieben, B.O., Svejstrup, J.Q., and Gileadi, O. 2001. Involvement of yeast carboxy-terminal domain kinase I (CTDK-I) in transcription elongation in vivo. *Gene* **267**: 31–36.
- Jones, J.C., Phatnani, H.P., Haystead, T.A., MacDonald, J.A., Alam, S.M., and Greenleaf, A.L. 2004. C-terminal repeat domain kinase I phosphorylates Ser2 and Ser5 of RNA polymerase II C-terminal domain repeats. *J. Biol. Chem.* **279**: 24957–24964.
- Joshi, A.A. and Struhl, K. 2005. Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol

Phatnani and Greenleaf

- II elongation. *Mol. Cell* **20**: 971–978.
- Jove, R. and Manley, J.L. 1984. In vitro transcription from the adenovirus 2 major late promoter utilizing templates truncated at promoter-proximal sites. *J. Biol. Chem.* **259**: 8513–8521.
- Kang, M.E. and Dahmus, M.E. 1993. RNA polymerases IIA and IIO have distinct roles during transcription from the TATA-less murine dihydrofolate reductase promoter. *J. Biol. Chem.* **268**: 25033–25040.
- 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., et al. 2005. Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* **123**: 593–605.
- Kim, W.Y. and Dahmus, M.E. 1989. The major late promoter of adenovirus-2 is accurately transcribed by RNA polymerases IIO, IIA, and IIB. *J. Biol. Chem.* **264**: 3169–3176.
- Kim, E., Du, L., Bregman, D.B., and Warren, S.L. 1997. Splicing factors associate with hyperphosphorylated RNA polymerase II in the absence of pre-mRNA. *J. Cell Biol.* **136**: 19–28.
- Kim, M., Krogan, N.J., Vasiljeva, L., Rando, O.J., Nedeá, E., Greenblatt, J.F., and Buratowski, S. 2004. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* **432**: 517–522.
- Kizer, K.O., Phatnani, H.P., Shibata, Y., Hall, H., Greenleaf, A.L., and Strahl, B.D. 2005. A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. *Mol. Cell Biol.* **25**: 3305–3316.
- Komarnitsky, P., Cho, E.J., and Buratowski, S. 2000. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes & Dev.* **14**: 2452–2460.
- Krogan, N.J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D.P., Beattie, B.K., Emili, A., Boone, C., et al. 2003. Methylation of histone H3 by Set2 in *Saccharomyces cerevisiae* is linked to transcriptional elongation by RNA polymerase II. *Mol. Cell Biol.* **23**: 4207–4218.
- Kyburz, A., Sadowski, M., Dichtl, B., and Keller, W. 2003. The role of the yeast cleavage and polyadenylation factor subunit Ydh1p/Ctf2p in pre-mRNA 3'-end formation. *Nucleic Acids Res.* **31**: 3936–3945.
- Lee, J.M. and Greenleaf, A.L. 1989. A protein kinase that phosphorylates the C-terminal repeat domain of the largest subunit of RNA polymerase II. *Proc. Natl. Acad. Sci.* **86**: 3624–3628.
- Lee, J.M. and Greenleaf, A.L. 1991. CTD kinase large subunit is encoded by *CTK1*, a gene required for normal growth of *Saccharomyces cerevisiae*. *Gene Expr.* **1**: 149–167.
- Lee, J.M. and Greenleaf, A.L. 1997. Modulation of RNA polymerase II elongation efficiency by C-terminal heptapeptide repeat domain kinase I. *J. Biol. Chem.* **272**: 10990–10993.
- Lee, D. and Lis, J.T. 1998. Transcriptional activation independent of TFIIF kinase and the RNA polymerase II mediator in vivo. *Nature* **393**: 389–392.
- Li, J., Moazed, D., and Gygi, S.P. 2002. Association of the histone methyltransferase Set2 with RNA polymerase II plays a role in transcription elongation. *J. Biol. Chem.* **277**: 49383–49388.
- Li, B., Howe, L., Anderson, S., Yates III, J.R., and Workman, J.L. 2003. The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J. Biol. Chem.* **278**: 8897–8903.
- Li, M., Phatnani, H.P., Guan, Z., Sage, H., Greenleaf, A.L., and Zhou, P. 2005. Solution structure of the Set2–Rpb1 interacting domain of human Set2 and its interaction with the hyperphosphorylated C-terminal domain of Rpb1. *Proc. Natl. Acad. Sci.* **102**: 17636–17641.
- Li, Z., Li, H., Devasahayam, G., Gemmill, T., Chaturvedi, V., Hanes, S.D., and Van Roey, P. 2005. The structure of the *Candida albicans* Ess1 prolyl isomerase reveals a well-ordered linker that restricts domain mobility. *Biochemistry* **44**: 6180–6189.
- Licaltosi, D.D., Geiger, G., Minet, M., Schroeder, S., Cilli, K., McNeil, J.B., and Bentley, D.L. 2002. Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. *Mol. Cell* **9**: 1101–1111.
- Lindstrom, D.L. and Hartzog, G.A. 2001. Genetic interactions of Spt4–Spt5 and TFIIS with the RNA polymerase II CTD and CTD modifying enzymes in *Saccharomyces cerevisiae*. *Genetics* **159**: 487–497.
- Liu, J. and Kipreos, E.T. 2000. Evolution of cyclin-dependent kinases (CDKs) and CDK-activating kinases (CAKs): Differential conservation of CAKs in yeast and metazoa. *Mol. Biol. Evol.* **17**: 1061–1074.
- Maciag, K., Altschuler, S.J., Slack, M.D., Krogan, N.J., Emili, A., Greenblatt, J.F., Maniatis, T., and Wu, L.F. 2006. Systems-level analyses identify extensive coupling among gene expression machines. *Mol. Syst. Biol.* **2**: 2006.0003.
- Maniatis, T. and Reed, R. 2002. An extensive network of coupling among gene expression machines. *Nature* **416**: 499–506.
- Marshall, N.F., Peng, J., Xie, Z., and Price, D.H. 1996. Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *J. Biol. Chem.* **271**: 27176–27183.
- McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Program, A.E., Shuman, S., et al. 1997a. 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes & Dev.* **11**: 3306–3318.
- McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., and Bentley, D.L. 1997b. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**: 357–361.
- Meier, U.T. 2005. The many facets of H/ACA ribonucleoproteins. *Chromosoma* **114**: 1–14.
- Meinhart, A. and Cramer, P. 2004. Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature* **430**: 223–226.
- Meinhart, A., Kamenski, T., Hoepfner, S., Baumli, S., and Cramer, P. 2005. A structural perspective of CTD function. *Genes & Dev.* **19**: 1401–1415.
- Misteli, T. and Spector, D.L. 1999. RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo. *Mol. Cell* **3**: 697–705.
- Morris, D.P. and Greenleaf, A.L. 2000. The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J. Biol. Chem.* **275**: 39935–39943.
- Morris, D.P., Phatnani, H., and Greenleaf, A.L. 1999. Phospho-CTD binding and the role of a prolyl isomerase in pre-mRNA 3' end formation. *J. Biol. Chem.* **274**: 31583–31587.
- Morris, D.P., Michelotti, G.A., and Schwinn, D.A. 2005. Evidence that phosphorylation of the RNA polymerase II carboxyl-terminal repeats is similar in yeast and humans. *J. Biol. Chem.* **280**: 31368–31377.
- Noble, C.G., Hollingworth, D., Martin, S.R., Ennis-Adeniran, V., Smerdon, S.J., Kelly, G., Taylor, I.A., and Ramos, A. 2005. Key features of the interaction between Pcf11 CID and RNA polymerase II CTD. *Nat. Struct. Mol. Biol.* **12**: 144–151.
- Nonet, M., Sweetser, D., and Young, R.A. 1987. Functional re-

- dundancy and structural polymorphism in the large subunit of RNA polymerase II. *Cell* **50**: 909–915.
- O'Brien, T., Hardin, S., Greenleaf, A.L., and Lis, J.T. 1994. Phosphorylation of RNA polymerase II C-terminal domain and transcriptional elongation. *Nature* **370**: 75–77.
- Ostapenko, D. and Solomon, M.J. 2003. Budding yeast CTDK-I is required for DNA damage-induced transcription. *Eukaryot. Cell* **2**: 274–283.
- Otero, G., Fellows, J., Li, Y., de Bizemont, T., Dirac, A.M., Gustafsson, C.M., Erdjument-Bromage, H., Tempst, P., and Svejstrup, J.Q. 1999. Elongator, a multisubunit component of a novel RNA polymerase II holoenzyme for transcriptional elongation. *Mol. Cell* **3**: 109–118.
- Patturajan, M., Schulte, R.J., Sefton, B.M., Berezney, R., Vincent, M., Bensaude, O., Warren, S.L., and Corden, J.L. 1998a. Growth-related changes in phosphorylation of yeast RNA polymerase II. *J. Biol. Chem.* **273**: 4689–4694.
- Patturajan, M., Wei, X., Berezney, R., and Corden, J.L. 1998b. A nuclear matrix protein interacts with the phosphorylated C-terminal domain of RNA polymerase II. *Mol. Cell. Biol.* **18**: 2406–2415.
- Payne, J.M., Laybourn, P.J., and Dahmus, M.E. 1989. The transition of RNA polymerase II from initiation to elongation is associated with phosphorylation of the carboxyl-terminal domain of subunit IIa. *J. Biol. Chem.* **264**: 19621–19629.
- Phatnani, H.P., Jones, J.C., and Greenleaf, A.L. 2004. Expanding the functional repertoire of CTD kinase I and RNA polymerase II: Novel phosphoCTD-associating proteins in the yeast proteome. *Biochemistry* **43**: 15702–15719.
- Prelich, G. 2002. RNA polymerase II carboxy-terminal domain kinases: Emerging clues to their function. *Eukaryot. Cell* **1**: 153–162.
- Proft, M., Mas, G., de Nadal, E., Vendrell, A., Noriega, N., Struhl, K., and Posas, F. 2006. The stress-activated hog1 kinase is a selective transcriptional elongation factor for genes responding to osmotic stress. *Mol. Cell* **23**: 241–250.
- Proudfoot, N. 2004. New perspectives on connecting messenger RNA 3' end formation to transcription. *Curr. Opin. Cell Biol.* **16**: 272–278.
- Proudfoot, N.J., Furger, A., and Dye, M.J. 2002. Integrating mRNA processing with transcription. *Cell* **108**: 501–512.
- Ranganathan, R., Lu, K.P., Hunter, T., and Noel, J.P. 1997. Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent. *Cell* **89**: 875–886.
- Rasmussen, E.B. and Lis, J.T. 1993. In vivo transcriptional pausing and cap formation on three *Drosophila* heat shock genes. *Proc. Natl. Acad. Sci.* **90**: 7923–7927.
- Rodriguez, C.R., Cho, E.-J., Keogh, M.C., Moore, C.L., Greenleaf, A.L., and Buratowski, S. 2000. Kin28, the TFIIB-associated CTD kinase, facilitates the recruitment of mRNA processing machinery to RNA polymerase II. *Mol. Cell. Biol.* **20**: 104–112.
- Schafer, T., Maco, B., Petfalski, E., Tollervey, D., Bottcher, B., Aebi, U., and Hurt, E. 2006. Hrr25-dependent phosphorylation state regulates organization of the pre-40S subunit. *Nature* **441**: 651–655.
- Schaft, D., Roguev, A., Kotovic, K.M., Shevchenko, A., Sarov, M., Shevchenko, A., Neugebauer, K.M., and Stewart, A.F. 2003. The histone 3 lysine 36 methyltransferase, SET2, is involved in transcriptional elongation. *Nucleic Acids Res.* **31**: 2475–2482.
- Schroeder, S.C., Schwer, B., Shuman, S., and Bentley, D. 2000. Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes & Dev.* **14**: 2435–2440.
- Shatkin, A.J. 1976. Capping of eucaryotic mRNAs. *Cell* **9**: 645–653.
- Sims III, R.J., Belotserkovskaya, R., and Reinberg, D. 2004. Elongation by RNA polymerase II: The short and long of it. *Genes & Dev.* **18**: 2437–2468.
- Skaar, D.A. and Greenleaf, A.L. 2002. The RNA polymerase II CTD kinase CTDK-I affects pre-mRNA 3' cleavage/polyadenylation through the processing component Pti1p. *Mol. Cell* **10**: 1429–1439.
- Steinmetz, E.J., Conrad, N.K., Brow, D.A., and Corden, J.L. 2001. RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature* **413**: 327–331.
- Sterner, D., Lee, J.M., Hardin, S.E., and Greenleaf, A.L. 1995. Yeast carboxyl-terminal repeat domain kinase CTDK-I is a divergent cyclin-cyclin-dependent kinase complex. *Mol. Cell. Biol.* **15**: 5716–5724.
- Stiller, J.W. and Cook, M.S. 2004. Functional unit of the RNA polymerase II C-terminal domain lies within heptapeptide pairs. *Eukaryot. Cell* **3**: 735–740.
- Sudol, M., Sliwa, K., and Russo, T. 2001. Functions of WW domains in the nucleus. *FEBS Lett.* **490**: 190–195.
- Sun, X., Zhang, Y., Cho, H., Rickert, P., Lees, E., Lane, W., and Reinberg, D. 1998. NAT, a human complex containing Srb polypeptides that functions as a negative regulator of activated transcription. *Mol. Cell* **2**: 213–222.
- Takahara, T., Tasic, B., Maniatis, T., Akanuma, H., and Yanagisawa, S. 2005. Delay in synthesis of the 3' splice site promotes *trans*-splicing of the preceding 5' splice site. *Mol. Cell* **18**: 245–251.
- Verdecia, M.A., Bowman, M.E., Lu, K.P., Hunter, T., and Noel, J.P. 2000. Structural basis for phosphoserine-proline recognition by group IV WW domains. *Nat. Struct. Biol.* **7**: 639–643.
- Vojnic, E., Simon, B., Strahl, B.D., Sattler, M., and Cramer, P. 2006. Structure and carboxyl-terminal domain (CTD) binding of the Set2 SRI domain that couples histone H3 Lys36 methylation to transcription. *J. Biol. Chem.* **281**: 13–15.
- Warner, J.R. 1999. The economics of ribosome biosynthesis in yeast. *Trends Biochem. Sci.* **24**: 437–440.
- Weeks, J.R., Hardin, S.E., Shen, J., Lee, J.M., and Greenleaf, A.L. 1993. Locus-specific variation in phosphorylation state of RNA polymerase II in vivo: Correlations with gene activity and transcript processing. *Genes & Dev.* **7**: 2329–2344.
- West, M.L. and Corden, J.L. 1995. Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. *Genetics* **140**: 1223–1233.
- Wittschieben, B.O., Otero, G., de Bizemont, T., Fellows, J., Erdjument-Bromage, H., Ohba, R., Li, Y., Allis, C.D., Tempst, P., and Svejstrup, J.Q. 1999. A novel histone acetyltransferase is an integral subunit of elongating RNA polymerase II holoenzyme. *Mol. Cell* **4**: 123–128.
- Wool, I.G. 1996. Extraribosomal functions of ribosomal proteins. *Trends Biochem. Sci.* **21**: 164–165.
- Wu, X., Wilcox, C.B., Devasahayam, G., Hackett, R.L., Arevalo-Rodriguez, M., Cardenas, M.E., Heitman, J., and Hanes, S.D. 2000. The *ess1* prolyl isomerase is linked to chromatin remodeling complexes and the general transcription machinery. *EMBO J.* **19**: 3727–3738.
- Xiao, T., Hall, H., Kizer, K.O., Shibata, Y., Hall, M.C., Borchers, C.H., and Strahl, B.D. 2003. Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. *Genes & Dev.* **17**: 654–663.
- Yang, P.K., Hoareau, C., Froment, C., Monsarrat, B., Henry, Y., and Chanfreau, G. 2005. Cotranscriptional recruitment of the pseudouridyltransferase Cbf5p and of the RNA binding protein Naf1p during H/ACA snoRNP assembly. *Mol. Cell.*

Phatnani and Greenleaf

- Biol.* **25**: 3295–3304.
- Yue, Z., Maldonado, E., Pillutla, R., Cho, H., Reinberg, D., and Shatkin, A.J. 1997. Mammalian capping enzyme complements mutant *Saccharomyces cerevisiae* lacking mRNA guanylyltransferase and selectively binds the elongating form of RNA polymerase II. *Proc. Natl. Acad. Sci.* **94**: 12898–12903.
- Yuryev, A. and Corden, J.L. 1996. Suppression analysis reveals a functional difference between the serines in positions two and five in the consensus sequence of the C-terminal domain of yeast RNA polymerase II. *Genetics* **143**: 661–671.
- Yuryev, A., Patturajan, M., Litingtung, Y., Joshi, R.V., Gentile, C., Gebara, M., and Corden, J.L. 1996. The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. *Proc. Natl. Acad. Sci.* **93**: 6975–6980.
- Zehring, W.A., Lee, J.M., Weeks, J.R., Jokerst, R.S., and Greenleaf, A.L. 1988. The C-terminal repeat domain of RNA polymerase II largest subunit is essential in vivo but is not required for accurate transcription initiation in vitro. *Proc. Natl. Acad. Sci.* **85**: 3698–3702.
- Zimmermann, R.A. 2003. The double life of ribosomal proteins. *Cell* **115**: 130–132.
- Zorio, D.A. and Bentley, D.L. 2004. The link between mRNA processing and transcription: Communication works both ways. *Exp. Cell Res.* **296**: 91–97.