Published in final edited form as:

Biochim Biophys Acta. 2016 April ; 1864(4): 372–387. doi:10.1016/j.bbapap.2016.01.007.

Dephosphorylating Eukaryotic RNA Polymerase II

Joshua E. Mayfield¹, Nathaniel T. Burkholder¹, and Yan Jessie Zhang^{1,*}

¹Department of Molecular Biosciences and Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, TX 78712, USA.

Abstract

The phosphorylation state of the C-terminal domain of RNA Polymerase II is required for the temporal and spatial recruitment of various factors that mediate transcription and RNA processing throughout the transcriptional cycle. Therefore, changes in CTD phosphorylation by site-specific kinases/phosphatases are critical for the accurate transmission of information during transcription. Unlike kinases, CTD phosphatases have been traditionally neglected as they are thought to act as passive negative regulators that remove all phosphate marks at the conclusion of transcription. This over-simplified view has been disputed in recent years and new data asserts the active and regulatory role phosphatases play in transcription. We now know that CTD phosphatases ensure the proper transition between different stages of transcription, balance the distribution of phosphorylation for accurate termination and re-initiation, and prevent inappropriate expression of certain genes. In this review, we focus on the specific roles of CTD phosphatases in regulating transcription. In particular, we emphasize how specificity and timing of dephosphorylation is achieved for these phosphatases and consider the various regulatory factors that affect these dynamics.

Keywords

transcription; x-ray crystallography; RNA polymerase II; phosphorylation/dephosphorylation; post-translational modification

Introduction

In eukaryotes, RNA polymerase II (RNA Pol II) is a workhorse for cellular transcription, producing all mRNAs as well as small nuclear, small nucleolar, and some micro RNAs [1-4]. Like many important eukaryotic enzymes, the function of RNA Pol II is tightly regulated by post-translational modifications such as phosphorylation [4-13], which has a rich and sprawling history (Box 1). A search of PhosphoSitePlus®, a high throughput proteomics database for post-translational modifications [14], reveals over 100 sites of phosphorylation throughout the subunits of human RNA Pol II identified by proteomic

* Correspondence should be address to: Yan Jessie Zhang, jzhang@cm.utexas.edu; Phone: 512-471-8645..

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

analysis or molecular biology approaches. However, the most clearly physiologically relevant phosphorylation sites with validation are densely localized within one region of RNA Pol II: the C-terminal region of the largest subunit RPB1 now universally referred to as the C-terminal domain (CTD) of RNA polymerase II [15, 16]. The primary sequence of the CTD consists of multiple repeats of the consensus heptad sequence Tyr₁Ser₂Pro₃Thr₄Ser₅Pro₆Ser₇, with the number of repeats ranging from 5 in *Plasmodium yoelii* (GenBank Accession: XM_726075) to 52 in humans (GenBank Accession: NM_000937) [17]. This sequence is highly enriched in potential phosphorylation sites with each of the five non-proline residues in the heptad repeat being recognized as phosphate acceptors *in vivo* [9, 12, 13, 18].

The CTD of RNA Pol II is conserved among eukaryotes, but not present in prokaryotic RNA polymerases or in other eukaryotic DNA-dependent RNA polymerases (RNA Pol I & III) [15, 46]. The CTD acts as a nexus for the flow of information between DNA, RNA, and protein codes [4, 47]. The primary role of CTD phosphorylation during transcription is to recruit proteins that facilitate mRNA production and processing such as initiation factors [48, 49], capping enzymes [7, 8, 50], splicing factors [51, 52], and poly-adenylation/termination complexes [12, 53-56]. Furthermore, different phosphorylated forms of the CTD link transcription to other cellular events like histone modification [57], DNA repair [58], and cell cycle regulation [59-61].

The phosphorylation status of the two best-understood CTD residues, Ser2 and Ser5, is tightly correlated with the progress of transcription [62, 63]. Genome wide analysis has allowed for mapping of not only Ser2 and Ser5 levels during transcription, but also the occurrence of other phosphorylated residues of the heptad repeat in yeast [13, 62, 64-66] (Fig. 1). The CTD must be hypophosphorylated for the polymerase to bind the promoter of transcribed genes [5, 67]. During promoter clearance, the transcribing RNA Pol II leaves the pre-initiation complex and polymerizes the first few nucleotides. This step in transcription is associated with increasing levels of Ser5 phosphorylation that quickly plateau. Within the next 750bp Ser5 is significantly dephosphorylated, although phosphorylated Ser5 seems to be present throughout transcript elongation [62]. Phosphorylated Ser2 levels rise later in elongation and reach their peak when approximately 1kb is transcribed in *S. cerevisiae* [62]. Phosphorylated Ser2 remains present past the site of poly-adenylation and is removed at transcription termination [62]. Less well-understood phosphorylation at Tyr1, Thr4, and Ser7 have been identified and are also accumulated in distinct patterns throughout the transcription cycle. Unlike Ser2 and Ser5, the biological implication of phosphorylation events at this site might differ in different model organisms. In yeast, phosphorylated Tyr1 levels follow a similar trend to phosphorylated Ser2, being initially low at the promoter and increasing throughout transcript elongation in yeast [13]. Notably, vertebrate Tyr1 phosphorylation occurs at the transcription start site, enhancer regions, and to a lesser extent in the coding region of genes suggesting differential functionality in these organisms [68, 69]. Unlike Ser2, Tyr1 phosphorylation does not persist after the polyadenylation site [13, 68]. Thr4 phosphorylation is evident throughout transcript elongation but is low at the poly-adenylation site in yeast [13]. Ser7 phosphorylation rises along side Ser5 phosphorylation, but remains at a more constant level throughout transcription and drops as the polymerase

reaches the poly-adenylation site [62]. All phosphorylation marks are removed before the initiation of another round of transcription [70] (Fig. 1). Ser2 and Ser5 phosphorylation of the CTD help couple transcription to vital co-transcriptional processes like 5'-capping [8, 50] and 3'-end processing [56], and are therefore most likely important to the majority of RNA polymerase II transcripts. Tyr1 also appears to play a role in general transcription by preventing the recruitment of termination factors and promoting elongation [13]. Thr4 and Ser7, on the other hand, may be playing „gene-specific' roles. For example, Ser7 phosphorylation is essential for snRNA gene expression [10] and Thr4 phosphorylation is linked specifically to histone mRNA 3'-processing in mammals [12] likely due to elongation defects inherent to Thr4 mutants [71].

The dynamic pattern of post-translational modification of the RNA Pol II C-terminal tail is termed “the CTD code” [72]. This code is generated through the interplay of molecular writers, modifiers, and erasers; and interpreted by readers [4]. Consistent with this metaphor, enzymes like CTD kinases and phosphatases embody writers and erasers and determine the post-translational modification status of CTD for the readers, or transcription regulators. Together, these factors ensure proper timing of transcriptional processes [4].

Interest in key writers of the CTD code, the CTD kinases, has generated a wealth of information to explain fundamental determinants of gene expression [49, 73-75]. Intriguingly, most of the CTD kinases belong to the cyclin-dependent kinase family (CDK), implying a link between cell-cycle regulation and transcription. Ser5 phosphorylation by the TFIIF complex subunit Cdk7 coincides with the clearance of RNA Pol II from the promoter [76, 77] and disruption of the interaction between RNA Pol II and mediator [78]. The most essential function of phosphorylated Ser5 is the recruitment of 5'-capping factors and the enhancement of guanylyltransferase activity to form the 5'-end of nascent RNA [8, 79]. This point was elegantly demonstrated in *S. pombe*, where bypassing the requirement for Ser5 phosphorylation was accomplished by covalently tethering 5'capping machinery to the CTD [50]. CDK9 as a part of the PTEFb complex further promotes transition from initiation to elongation by phosphorylating Ser2 [75], as well as negative elongation factors (NELF) and DRB sensitivity-inducing factor (DSIF) which inhibit transcript elongation [80-83], all contributing to promote elongation. Ser2 phosphorylation also facilitates the recruitment of 3'-end processing factors [56]. Several other CDKs such as CDK8 [84], CDK12 [85] and CDK13 [86] have been reported to complement the function of CDK7 and 9 by phosphorylating Ser2 and Ser5 of CTD heptad repeats. The function of the CTD kinases has been discussed extensively in review articles of CTD including recent reviews by Eick & Geyer [47]; Jeronimo, Bataille, & Robert [4]; and Corden [87].

In contrast to the actively regulatory role commonly attributed to CTD kinases, CTD phosphatases are often viewed as passive players in the transcription cycle. However, recent discoveries demonstrate that CTD phosphatases are highly regulated in each stage of transcription to promote accuracy. These phosphatases operate in concert with CTD kinases and modifying enzymes, like prolyl isomerases, to produce a combinatorial code that fine-tunes and directs transcriptional outcomes [88]. In this review we focus specifically on the role and regulation of CTD phosphatases, their cross talk to CTD modifiers, and their global impact on transcription.

1. Studying CTD Phosphatases

Compared to the intensive study of CTD kinases spanning over three decades, the CTD phosphatases have been overshadowed due to their perceived passive role of simply erasing CTD phosphorylation at the end of transcription. However, this view has been disputed with overwhelming evidence that phosphatases are dynamic and active participants in transcription regulation. Cells lacking certain CTD phosphatases experience termination defects [89-92], reduced transcription levels due to failure to turnover RNA Pol II [92-94], and even cell death [95, 96].

Another reason for the delayed research on CTD phosphatases roots from the perception that phosphatases lack specificity *in vitro*. Many phosphatases require regulatory factors in order to selectively target their substrates ([97], for review). Because of this, *in vitro* assays conducted using phosphatases and their substrate candidates may not reflect the physiological events *in vivo*. Unlike kinases, which can be divided into tyrosine and serine/threonine kinases based on their substrate selectivity, phosphatases are versatile and recognize phosphate-containing substrates with considerably less distinction. This makes the identification of physiological phosphatase-substrate pairs challenging. This challenge is intensified in the study of CTD phosphatases since the flanking residues around Ser2 and Ser5 on CTD are similar. For instance, a proline residue follows both Ser2 and Ser5 and forms a common recognition motif for kinases and phosphatases. In eukaryotic transcription, the precise removal of phosphate groups from the CTD seems to be just as important as adding them accurately in the first place. Therefore, phosphatases must utilize strategies like localization and complex assembly to gain specificity *in vivo*. This allows for the timely dephosphorylation of CTD and the recruitment of correct regulatory factors.

Due to the general lack of phosphatase specificity *in vitro*, additional credentials are required to establish whether an *in vitro* substrate is a physiological target for a CTD phosphatase. Before we delve into the discussion of individual CTD phosphatases we will clarify our proposed requirements to assign phosphatase specificity. The first requirement is confirmation of phosphatase activity against a phosphorylated substrate *in vitro*. This is commonly tested using colorimetric or fluorescence based kinetic assays or by detecting the post-translational modifications state of a substrate using a phospho-specific antibody. The second credential is that loss or inactivation of the phosphatase should result in the accumulation of the phosphorylated substrate species *in vivo*. Several methods of inactivating phosphatases are available including gene-knockout, gene-knockdown by siRNA or shRNA, or directly blocking phosphatase activity with a small molecule inhibitor. Generally, detection of the phosphorylated substrate can be done using western blot or mass spectrometry. The third and final credential is the requirement of the phosphatase being “in the right place at the right time”. Since the specificity of phosphatases is highly dependent on its spatial and temporal localization, evidence must be provided that the phosphatase and its presumed substrate indeed coexist in close proximity of one another. This can be at least partially fulfilled by testing the direct interaction between the substrate and its cognate phosphatase or phosphatase complex *in vitro*. For detecting *in vivo* interactions, researchers often utilize yeast two-hybrid and pull-down assays with the phosphatase and supposed substrate. Only when all three credentials are met can a phosphatase be firmly established as

a *bona fide* physiologically relevant phosphatase for a specific substrate. It is technically challenging to fulfill all three requirements, which has led to several instances of phosphatases acting on particular substrates being circumstantial

Unlike kinases in which most family members appear to derive from a common ancestor and display similar folding and reaction mechanisms [42, 43], phosphatases can be divided into three families based on catalytic mechanism (Fig. 2). Approximately 150 protein phosphatases have been identified in humans. Most of these (approximately 100) comprise the first category that uses cysteine as a nucleophile [97]. Cysteine-based PTP (protein tyrosine) phosphatases form a phosphoryl-intermediate between the substrate phosphate and cysteine nucleophile, which is eventually hydrolyzed to release inorganic phosphate (Fig. 2A). These phosphatases can dephosphorylate lipids and secondary metabolites as well as tyrosine, serine, and threonine residues making this family highly versatile. Notable family members of the cysteine-based phosphatases include PTEN, almost all tyrosine phosphatases, and many dual specificity phosphatases [98]. The second category, consisting of the PPP (phosphoprotein phosphatases) and PPM (metal-dependent protein phosphatases) families of phosphatases, utilizes a di-metal ion assisted mechanism. Less than 40 members of this category have been identified in humans, but they direct the majority of dephosphorylation reactions in human cells [97, 99]. These phosphatases use two metal ions to coordinate and activate a water molecule that breaks the covalent bond connecting the phosphate to substrate (Fig. 2B). PP1 and PP2 phosphatase subfamilies are prominent members of this group and are major players in many signaling events involving serine/threonine phosphorylation [97]. The third category resembles cysteine-based PTP phosphatases, but instead of cysteine utilize an aspartate nucleophile. In this group of enzymes, the negative charges present on both nucleophile and general acid/base (also an aspartate) are neutralized by a magnesium ion that is required for substrate binding and enzymatic activity [100]. Upon nucleophilic attack a phospho-aspartyl intermediate is formed and eventually resolved by active site water to yield a dephosphorylated serine/threonine and an inorganic phosphate (Fig. 2C). CTD phosphatases are a diverse group of enzymes and contain members from all three of these major phosphatase families (Fig. 2).

2. CTD Phosphatases

Phosphatase activity against the CTD was first detected in a fraction of HeLa cell extract [101]. A factor, or factors, in this extract was capable of converting hyperphosphorylated RNA Pol II (RNAP II_o) to a hypophosphorylated state (RNAP II_a). Such a transition allows RNA Pol II to bind the adenovirus-2 major late promoter [67], providing early evidence that the phosphorylation state of the CTD is critical for promoter recognition. Research has since expanded upon CTD dephosphorylation, leading to the discovery of several distinct phosphatases that play important roles at different stages of the transcription cycle in both general and gene-specific transcription. These CTD phosphatases can be catalogued by both the sites they target within the CTD heptad repeats and by the stage(s) of transcription at which they are recruited. This dual classification is essential because even phosphatases targeting the same sites within the heptad can lead to different transcriptional outcomes depending on the stage of transcription in which they act. The critical question for these enzymes is how their specificity is achieved both spatially and temporally in the context of

eukaryotic transcription. Here we provide an in depth analysis of this heterogeneous group of enzymes and highlight their diverse transcriptional functions, unique regulatory strategies, and interesting biochemical properties.

2.1 Fcp1

Shortly after CTD phosphatase activity was found in HeLa cell extract, a phosphatase with similar properties was identified and isolated in yeast [102] and later named Fcp1 (TFIIF-associated component of the CTD phosphatase) [96]. Sequence alignment of Fcp1 homologues from yeast, humans, and 15 other eukaryotic species revealed three conserved domains among them [96]: (1) An N-terminal Fcp1 homology region (FCPH) containing the catalytically important residues, (2) a BRCT domain, which commonly mediates protein-protein interactions, and (3) a C-terminal region that binds to general transcription factor RAP74 of TFIIF (RAP30/74) [96] (Fig. 3A).

Fcp1 was initially characterized as a PP2C phosphatase that utilized the PPP/PPM mechanism (Fig. 2B), based on its requirement for magnesium and resistance to okadaic acid [101]. However, the mechanism of Fcp1 more closely resembles that of the haloacid dehydrogenase (HAD) protein family, a huge protein family including enzymes that mediate C-P or O-P bond cleavage [100]. The first aspartate of the signature Fcp1 DXDX(T/V) motif acts as a nucleophile to form a phosphoryl intermediate which is later broken down to release an inorganic phosphate (Fig. 2C) [94, 100]. These findings positioned Fcp1 as the founding member of a new aspartate-based serine/threonine phosphatase family in eukaryotes, the FCP/SCP phosphatases [97, 103].

There is substantial evidence that the phosphatase activity of Fcp1 is vital to general transcription mediated by RNA Pol II. Yeast strains lacking the Fcp1 gene are not viable, presumably due to an inability to dephosphorylate and recycle RNA Pol II for subsequent rounds of transcription [94, 96]. Yeast containing Fcp1 variants with attenuated phosphatase activity can survive but exhibit defects like read-through transcription and globally reduced RNA levels [94]. These findings strongly linked Fcp1 phosphatase activity to general transcription *in vivo*.

In addition to Fcp1's essential role in RNA Pol II recycling, *in vitro* transcription reconstruction highlights Fcp1 as an important elongation factor [70, 104]. The addition of Fcp1 to a reconstituted *in vitro* transcription assay resulted in enhanced transcript elongation. Surprisingly, the positive effect on elongation didn't require the phosphatase activity of Fcp1 as it can be achieved with an isoform lacking the N-terminal catalytic region [70]. A reasonable explanation for this enhancement is that a C-terminal region of Fcp1 associates with Rap74, a component of the TFIIF transcription factor known to increase the rate of RNA Pol II elongation [105]. Furthermore, transcript elongation is enhanced three-fold by Fcp1 in the absence of initiation promoting TFIIF and has an additive effect when both elongation factors are present [104]. Therefore, it appears that Fcp1 enhances elongation along at least two pathways *in vitro* independent of its phosphatase activity: through a TFIIF-independent mechanism in which Fcp1 alone contributes to elongation rate and a TFIIF-dependent pathway that is additive with TFIIF stimulated elongation. The observed importance of Fcp1 to transcript elongation was also investigated *in vivo* using

mutant Fcp1 alleles. Mutant alleles of Fcp1 were found to suppress slow-growth phenotypes associated with inhibition of transcript elongation and confer a synthetic lethal phenotype when combined with RNA Pol II mutations known to reduce transcript elongation [104]. Although these experiments do not provide direct evidence of Fcp1 impacting transcript elongation *in vivo* they do inform a model in which Fcp1 regulates elongation and warrant further investigation.

Since Fcp1 was identified as a CTD phosphatase, it has been of great interest to identify which residues of the CTD heptad it targets. Fcp1 can dephosphorylate Ser2 and Ser5 residues in synthetic phosphoryl-CTD peptides *in vitro* [106]. This is not surprising since the residues flanking these two serine residues are quite similar (Y₁S₂P₃ vs. T₄S₅P₆). However, yeast cells with inactive Fcp1 accumulate only phosphorylated Ser2, suggesting that Fcp1 controls the phosphorylation state of Ser2 rather than Ser5 *in vivo* [107]. Fcp1 favored synthetic CTD peptides phosphorylated at Ser2 over Ser5 by 10-fold in *in vitro* kinetic assays with specific activities of 16 nmol vs. 1.6 nmol of inorganic phosphate released per µg of Fcp1 per hour against phosphorylated Ser2 and Ser5 peptides, respectively [106]. The biological implication of this substrate specificity is substantial since the phosphorylation states of specific residues in the heptad repeat of CTD, in particular Ser2 and Ser5, are associated with different stages of transcription (Fig. 1) [8, 62]. Fcp1 has also been implicated as the phosphatase responsible for the newly identified Thr4 dephosphorylation *in vivo* [108, 109] and Fcp1 is active against phosphorylated Ser7 *in vitro*, but the physiological significance of this activity is unclear [62, 110]. These results together suggest that Fcp1 fulfills the specific and vital role of Ser2 CTD dephosphorylation, but maintains secondary activity against Thr4, Ser5, and Ser7 to ensure the complete dephosphorylation of the CTD at the end of transcription.

Fcp1 satisfies all three of our requirements to be a *bona fide* CTD phosphatase: (1) It shows *in vitro* phosphatase activity against phosphorylated CTD [70, 101, 106]. (2) Compromised activity of Fcp1 in cells leads to the accumulation of phosphorylated Ser2 [107]. (3) Both pull-down [111] and ChIP assay experiments [91, 107] demonstrate that Fcp1 associates with RNA Pol II at initiation and into elongation, but becomes most evident near the end of transcription where the drop in Ser2 phosphorylation is most apparent. Altogether, Fcp1 neatly fits our requirements for a CTD phosphatase and represents a model for classifying other suspected CTD phosphatases.

Because of its essential role in transcription there is considerable interest in determining how Fcp1 recognizes CTD. One explanation is that the molecular architecture of the active site facilitates discrimination between distinct phosphoryl species along CTD. To visualize this potential means of selection Fcp1 complex crystal structures with phosphorylated CTD peptide substrates were attempted using a catalytically inactive Fcp1. Unfortunately, even in crystallization conditions containing high concentrations of phosphorylated CTD peptide no substrate could be resolved in the final electron density. Although no complex structure was obtained, a partial crystal structure of the Fcp1 homology and BRCT domains was solved (Fig. 3C). Additionally, these *S. pombe* Fcp1 structures contain chemical inhibitors that mimic the transition state of the Fcp/Scp reaction mechanism (Fig. 2C) and provide at least some insight into substrate positioning [112]. The difficulty in obtaining Fcp1 complex

structures may stem from limitations inherent to packing of the crystal lattice that disfavor ligand incorporation or from Fcp1's weak affinity for substrate, as evidenced by low *in vitro* activity [112].

The weak binding between CTD and Fcp1 suggests that association of the CTD and Fcp1 could involve additional biological factors. This is not surprising since many phosphatases present ubiquitous activity against phosphoryl species and rely on an intricate network of binding partners or additional protein domains to garner substrate specificity [97]. Several proteins have been proposed to play a role in targeting Fcp1 to the CTD. The top candidate is the general transcription factor TFIIF, which has a known binding site at the C-terminus of Fcp1 and whose binding to Fcp1 results in increased phosphatase activity [96, 113]. Complex structures of the Rap74 subunit of TFIIF and a peptide derived from the Fcp1 C-terminal domain have been reported using both crystallography and NMR spectroscopy [114-116]. Disrupting the TFIIF and Fcp1 interaction *in vivo* leads to altered CTD phosphorylation and transcription defects supporting the functional role of this interaction in transcription [103]. Further interactions between Fcp1, RNA Pol II, and TFIIF appear to be mediated by Fcp1's BRCT domain. The BRCT domain binds to phosphorylated RNA Pol II *in vitro* [117]. Introduction of a single point mutation (Fcp1W575A) at the binding interface of the BRCT domain and RAP74 resulted in accumulation of RNA Pol II phosphorylation, decreased mRNA synthesis, and a temperature sensitive phenotype at 37°C in *S. cerevisiae* [103]. Importantly, the TFIIF/Fcp1 interaction may also be regulated by phosphorylation [118], although this requires further investigation. Fcp1 also interacts with the Rpb4 subunit of the RNA polymerase II holoenzyme [111]. Depletion of Rpb4 results in decreased Fcp1 and accumulated CTD phosphorylation in immunoprecipitated RNA Pol II holoenzyme complexes [111]. A more recent report corroborates these earlier findings and demonstrate that Rpb4 mutation results in reduced Fcp1 occupancy at genes and accumulated Ser2 phosphorylation, although similar data is presented for another major CTD phosphatase Ssu72 and Ser5 phosphorylation suggesting Rpb4 is important to additional facets of CTD dephosphorylation [108]. The association of Fcp1 with TFIIF and Rpb4 likely contributes to its association with RNA Pol II throughout elongation, its stimulation of elongation, and helps prime RNA Pol II for timely dephosphorylation by Fcp1 at the end of transcription where Fcp1 levels peak.

2.2 Scp1-3

The characterization of Fcp1 as a Ser2 CTD phosphatase sparked interest in finding other phosphatases acting on the CTD. A bioinformatics approach for proteins containing regions similar to the Fcp1 homology domain revealed three closely related human proteins [119]. The highly similar Small CTD Phosphatases 1-3 (Scp1-3 or CTDSP1/2/L) contain the conserved DXDX(T/V) motif, but lack the BRCT domain and C-terminal TFIIF-interacting domain that are required for Fcp1 recognition of the CTD [119]. The three isoforms share high homology from the DXDX(T/V) motifs to the end of the proteins, but vary at their leading N-terminal sequences that are presumed to be involved in targeting. Importantly, the three isoforms perform identically in all *in vitro* biophysical and kinetic characterizations; and although they locate in different chromosomes, the tissue specific expression profile for the three isoforms do not appear to differ [120].

Even though the overall primary sequences of the Scps and Fcp1 phosphatase domains are not well conserved, the active site residues are identical and the arrangement of the catalytically important residues adopt similar positions [121]. By mutating individual residues participating in particular steps of the phosphoryl-transfer reaction of Scp1 against CTD the mechanistic detail of this reaction was delineated structurally and snapshots of substrate binding, phosphoryl-enzyme intermediate, and phosphate release were obtained (Fig. 2C) [122]. Despite having an identical reaction mechanism to Fcp1, Scp1 has a nearly 69-fold higher preference for phosphorylated Ser5 of the CTD as opposed to Ser2 [119], with specificity constants of $11.4\text{s}^{-1}\text{mM}^{-1}$ against doubly phosphorylated Ser5 di-heptad CTD peptides compared to only $0.166\text{s}^{-1}\text{mM}^{-1}$ for equivalent Ser2 peptides [123]. The structure of catalytically-dead Scp1 (D96N) bound to the phosphorylated Ser5 of this peptide explains its substrate specificity (Fig. 4A) [123]. A hydrophobic pocket formed by the residues Phe106, Val118, Ile120, Val127, and Leu155 recognizes the Pro3 residue of the CTD heptad repeat (Fig. 4A) [123]. Attenuating the hydrophobicity of this pocket through mutation reduces Scp1 specificity. For instance, a single mutation of Phe106 to alanine nearly eliminates substrate discrimination [123]. Therefore, this hydrophobic pocket found in Scps but not in other Fcp/Scp family members appears to play an important role in their specificity.

While initially characterized as a CTD phosphatase, there is strong evidence suggesting that the Scps are not major Ser5 phosphatases in general transcription. Firstly, phosphatases required for general transcription should be conserved throughout evolution. The closest yeast homologues are the DXDX HAD phosphatases Nem1 and Psr1/2, which appear to be involved in membrane organization rather than transcription [124, 125]. Additionally, although RNA Pol II is associated with immunoprecipitated Scp1 as shown by CTD specific antibodies [119], inactivating Scps showed no obvious change in general transcription [120, 126]. These lines of evidence suggest that Scps are not required in the transcription of every gene.

Scps function in transcription seems gene-specific. Loss of functional Scp1 leads to de-repression of several neuronal genes as well as an induction of neuronal phenotypes in neural progenitor cells [120, 126]. Interestingly, the adult expression pattern of Scps mirrors that of the master neuronal regulator RE-1 Silencing Factor (REST) and the two proteins co-immunoprecipitate indicating a direct interaction between them [120]. REST specifically binds signature DNA sequences called RE-1 elements that are found upstream of certain neuronal genes [127]. REST acts as a hub for complexes that are capable of efficiently repressing ~10% of all neuronal genes [128, 129]. As a constituent of the REST complex, Scps contribute to regulating neuronal gene expression [120].

Based on these initial findings [119, 120], a plausible mechanism for Scps in neuronal silencing was that they are recruited to RE-1-controlled neuronal genes via their interaction with REST where they dephosphorylate any active RNA Pol II. In this model, Scps dephosphorylate the CTD at Ser5 and other residues efficiently due to their high phosphatase activity. Scps prevents RNA Pol II from entering active transcription by maintaining RNA Pol II in the hypophosphorylated form, consistent with a role in gene silencing. An additional mode of action has been recently proposed for Scps role in gene

silencing. REST degradation and association with REST-complex cofactors is regulated by phosphorylation at S861 and S864 [130]. Phosphorylation of these residues increases REST affinity to E3 ubiquitin ligase SCF ^{β} -TRCP and reduces its affinity to CoREST [130], a protein that binds the C-terminus of REST and helps recruit additional silencing factors [131]. Scps directly dephosphorylate REST at S861/S864 both *in vitro* and *in vivo*, and can abolish these negative effects on REST function [130]. By preventing ubiquitin-mediated degradation of REST and promoting its assembly with CoREST Scps stabilize the silencing complex and prevent neuronal gene expression [130]. These two models are not mutually exclusive and may coexist. Therefore, Scp1 negatively regulates RE-1 linked neuronal genes through dephosphorylation of REST to promote its stability/silencing function and may interfere with phosphorylation of RNA polymerase II to prevent transcript generation (Fig. 4B).

As a neuronal gene silencer, Scp mRNA is carefully regulated by inhibitory micro RNAs (miRNAs) that bind the 3' UTR and a class of mRNA splicing regulators called polypyrimidine tract binding proteins (PTBPs). The highly conserved miR-124 accounts for a majority (~25-50%) of miRNAs in the mammalian nervous system [132]. REST helps prevent neuronal differentiation by repressing three genomic miR-124 sites via upstream RE-1 elements in nonneuronal cells [133]. In neuronal cells, miR-124 down-regulates the expression of Scps as well as REST and other REST complex components by directing their mRNAs for degradation [134, 135]. Counter to the effect of miR-124, PTBP1 enhances Scp1 expression by directly competing for binding of the 3' UTR of Scp1 mRNA with miR-124 [126]. PTBP1 is also repressed by miR-124 in neuronal cells [136]; further reinforcing the notion that neurogenesis is carefully managed by a series of interacting regulatory loops. The intricate genetic feedback loops among REST complex components, miRNAs, and PTBPs appear to tightly control entry and maintenance of the neuronal phenotype. Another unique mechanism that controls Scps expression is through microRNAs miR-26a/b encoded within Scp introns. Under certain conditions, these introns mature into functional miR26a/b and down regulate their host genes [137, 138]. For example, miR-26b is up regulated during zebrafish embryonic development and down regulates its host gene *scp2* [138]. Loss of Scp in this manner likely promotes increased REST instability and initiation of the miR-124 regulatory program. Although more work is needed to identify conditions that trigger miR26a/b maturation and activity, a model has been proposed for Scp genetic regulation that fits with its suspected neuronal silencing function (Fig. 4C).

Inactivation of Scps specifically promote neuronal gene expression without significantly affecting global gene expression, making them attractive targets for stimulating neuron regeneration. Even though Scps are phosphatases, a protein family generally avoided by drug designers, the unique architecture of their active site is sufficiently different from the cysteine-based or di-metal phosphatases to provide a unique opportunity for targeted inhibitor design. A major consideration is cross-activity of Scp inhibitors towards other family members that share the same reaction mechanism, since inhibiting an essential protein like to Fcp1 would lead to cell death. A region in the Scps named the "insertion domain" is composed of approximately 40 residues that follow the DXDX(T/V) motif and exhibits significant sequence variation within the Fcp/Scp family [121] (Fig. 4A). This

region forms an additional binding pocket proximate to the DXDX(T/V) active site (~7Å removed) which enhances phosphorylated Ser5 specificity through recognition of Pro3 of the CTD [123]. Due to divergence of this insertion domain, compounds that specifically bind this hydrophobic pocket are less likely to cross-react with other Fcp/Scp family members. This notion is supported by the identification of the small molecule rabeprazole that specifically inhibits Scp1 and has no significant activity against Fcp1 or Dullard, an Fcp/Scp phosphatase involved in membrane biogenesis [139] (Fig. 4D). Rabeprazole provides a proof-of-concept for Scp inhibitor design, but compounds with greater potency and bioavailability are needed for the application of neuron regeneration.

In addition to Fcp1 and the Scps, HSPC129 and UBLCP1 are two other members of the Fcp/Scp family that have been reported to have activity against phosphorylated CTD *in vitro* [140, 141]. However, no *in vivo* CTD phosphatase activity has been characterized for either protein. As mentioned earlier, phosphatase specificity is achieved by a combination of factors like selectivity, localization, and timing. Therefore, *in vitro* activity alone is not enough to assign either protein as CTD phosphatase. In fact, despite its sequence similarity and *in vitro* activity UBLCP1 is specifically recruited to the 26S proteasome rather than RNA Pol II in order to regulate proteasomal assembly [142].

2.3 Ssu72

As Scps only regulate a subset of genes they cannot fulfill the requirement of dephosphorylating Ser5 of the CTD in general transcription. In the search for a physiologically relevant Ser5 CTD phosphatase, researchers identified Ssu72 (gene SSU72, *suppressor of sua7*, gene 2) [143]. SSU72 is an essential gene for yeast cell growth and is conserved throughout eukaryotes [95, 144]. SSU72 was originally identified in a genetic screen for genes that suppress the phenotype of mutations in *SUA7*, the *S. cerevisiae* homologue of the gene encoding human TFIIB [95]. Functional interaction between Ssu72, TFIIB, and Rpb2 suggested that this protein played a role in transcript initiation [145]. Further investigation implicated Ssu72 as a key player in 3' end processing and termination [89, 92]. Ssu72 was identified as a constituent of the Cleavage and Polyadenylation Factor (CPF) complex in yeast through proteomic studies and confirmed using immunoprecipitation [92, 146]. *In vivo* and *in vitro* transcription assays showed that Ssu72 mutant variants are defective in transcript elongation and termination [92, 147]. Indeed, Ssu72 activity is essential for the accurate termination of RNA Pol II-mediated transcription since Ssu72 variants with compromised phosphatase activity resulted in read-through transcription of almost half of snRNA and snoRNA as well as some mRNAs [89]. This phenotype can be partially rescued by reducing the elongation speed through mutation or introduction of chemical agents [89, 92]. It seems that Ssu72 balances transcript elongation and termination by decreasing the rate of elongation and allowing for appropriate binding of transcription factors and/or termination machinery [92]. The function of Ssu72 in transcription termination is also conserved in higher eukaryotes where Ssu72 is found to be a constituent of human Cleavage and Polyadenylation Specificity Factor (CPSF) complex (homologue of yeast CPF complex in mammalian systems) [144]. Based on ChIP analysis, the majority of Ssu72 is present at the end of the transcription cycle with a small portion found at the promoter region due to its interaction with TFIIB [91, 95]. This dual

localization might be functionally important as Ssu72 has been implicated in gene looping, which can allow RNA Pol II to re-initiate transcription of a gene instead of diffusing away [148, 149]. These early observations indicated that Ssu72 could be one of the missing players in CTD dephosphorylation.

Ssu72-mediated transcriptional regulation is reliant on its phosphatase activity. However, it was not immediately apparent that Ssu72 was indeed a phosphatase. The phosphatase activity of Ssu72 was supported by two bioinformatic lines of evidence: (1) Ssu72 orthologues contain a conserved CxxxxxRS sequence motif commonly found in protein tyrosine phosphatases [150] and (2) structure alignment revealed Ssu72 has a comparable fold to the low molecular weight tyrosine phosphatase (LMW-PTP) despite sharing little sequence homology [151]. Purified Ssu72 was first shown to be active using p-nitro-phenyl phosphate (pNPP) as a substrate [89], which is inline with its cysteine-based PTP mechanism considering the chemical similarity between pNPP and phosphorylated tyrosine. Interestingly, yeast mutants with decreased Ssu72 activity accumulated Ser5, but not Ser2 CTD phosphorylation [143]. This result is consistent with *in vitro* phosphatase assays against phosphorylated CTD peptide substrate, showing that Ssu72 is specific for Ser5 and not Ser2 or Tyr1 [110, 152].

The complex structures of Ssu72 bound to phosphorylated CTD peptides are highly informative in elucidating its substrate recognition. The most shocking revelation from the structure of Ssu72 in complex with a phosphorylated Ser5 CTD peptide is that Pro6 is in the *cis* conformation (Fig. 5A) [144, 153]. Since *cis*-proline is a minor species and crystallography captures an averaged signal, this configuration strongly suggests that Ssu72 is selective for *cis*-Pro6. Furthermore, substitution of proline with alanine, which does not assume a stable *cis* conformation, in CTD peptides results in profoundly reduced activity [154]. As far as we know, this is the first *cis*-proline specific phosphatase currently reported. Ser-Pro is a signature motif recognized by major kinases like mitogen activated protein kinases and the cyclin-dependent kinases [155], which yield product containing phosphorylated Ser-Pro motifs in the *trans* configuration [156, 157]. The requirement of *cis*-proline for Ssu72 implies that Pro6 has to undergo isomerization after the preceding Ser5 is phosphorylated by CTD kinases, like the cyclin-dependent kinase subunit of TFIIF (Cdk7) [158]. This requirement of Ssu72 ensures the CTD remains in the phosphorylated state without getting dephosphorylated immediately. This mechanism, and how it pertains to other CTD phosphatases, is discussed at length in Section 3 of this review. Additional sequence elements along the CTD also appear to facilitate Ssu72 binding. Mutation of the Tyr1 following phosphorylated Ser5 or the Thr4 preceding phosphorylated Ser5 greatly reduces Ssu72's activity against CTD substrate [154, 159]. These findings can be rationalized in the context of crystal structures where Tyr1 forms hydrophobic interactions with Phe50 of Ssu72 and Thr4 forms an intramolecular hydrogen bond with other CTD peptide residues to stabilize *cis*-Pro6 in the active site [159].

More recently, evidence suggests that Ssu72 is also the physiological Ser7 CTD phosphatase [62, 91]. *In vitro*, Ssu72 is 4000 fold less active against phosphorylated Ser7 than Ser5 peptides [110]. However, this activity seems to be much greater *in vivo* as long as Ser5 is also phosphorylated, suggesting that Ssu72 first targets Ser5 phosphorylated heptads and

then removes phosphates from Ser7 [91]. Coincidentally, Ssu72 is recruited to RNA Pol II just when Ser7 phosphorylation starts to fall [62, 91]. Furthermore, depletion of Ssu72 leads to an increase in Ser7 phosphorylation at the 3' end of genes [62]. The crystal structure of Ssu72 bound to phosphorylated Ser7 CTD peptide reveals an unexpected and distinct binding mode compared to the Ssu72-Ser5 structure. The phosphorylated Ser7 CTD peptide is rotated $\sim 180^\circ$ about the phosphorylated residue compared to the Ser5 substrate [110]. Between these two structures both phosphorylated Ser5 and Ser7 bind the same active site pocket while Pro6 remains in the same position irrespective of the substrate. In order to maintain the hydrophilic interaction between the phosphate group and active site residues as well as the hydrophobic interaction with Pro6, the N and C terminus of the CTD must be reversed (Fig. 5B). The crystal structures of Ssu72 binding to the CTD in these opposite orientations imply that rather than sliding, Ssu72 does the tango along the CTD reversing its orientation between dephosphorylating Ser5 and Ser7. There is also the possibility that different molecules of Ssu72 independently sample and subsequently dephosphorylate Ser5 and Ser7. Both models are intriguing and require further investigation to fully elucidate.

The activity of Ssu72 is regulated by its binding partner in cleavage and poly-adenylation complexes, Pta1 in yeast and Symplekin in vertebrates [55, 160, 161]. *In vivo*, this protein forms a scaffold for the complex and helps recruit additional transcription and termination factors to the poly-adenylation site [55, 160, 161]. Ssu72 is recruited to this complex by binding the concave side of Symplekin, characterized by a double layer of parallel helices containing HEAT repeats (Fig. 5C) [144, 153]. *In vitro*, Symplekin binding enhances the phosphatase activity of Ssu72 towards Ser5 phosphorylated GST-CTD despite the active site being removed by $\sim 20\text{\AA}$ from the binding interface [144]. Interestingly, comparison of *Drosophila* Ssu72 bound to CTD peptide with or without Symplekin association shows no difference in CTD binding mode [153, 159], suggesting that *in vitro* Pta1/Symplekin does not alter active site specificity but may facilitate CTD recruitment and dephosphorylation in a yet to be characterized way. *In vivo*, the main function of Pta1/Symplekin appears to be localization of Ssu72 to cleavage and Polyadenylation complexes [160].

2.4 Rtr1

None of the CTD phosphatases are more controversial than Rtr1. Since Fcp1 and Ssu72 are primarily recruited in the later stages of transcription [91], many have searched for phosphatases that modify the CTD phosphorylation status earlier in transcription cycle [62]. An early drop in Ser5 phosphorylation levels, detected using Ser5 specific antibodies, is coincident with the recruitment of a Pol II-associated protein, Rtr1 [90]. Incubation of *S. cerevisiae* Rtr1 with *in vitro* phosphorylated GST-CTD and synthetic peptides resulted in a reduction of Ser5 phosphorylated species [90]. Furthermore, yeast strains lacking Rtr1 display growth defects, accumulate phosphorylated Ser5, and experience problems in transcription [90, 162]. RPAP2, the human homologue of Rtr1, also associates with RNA Pol II and dephosphorylates Ser5 *in vitro* and *in vivo* [163]. This combination of *in vitro* and *in vivo* evidence supports the notion that Rtr1 and RPAP2 are Ser5 CTD phosphatases.

Surprisingly, the crystal structures of Rtr1 argue against its assignment as a CTD phosphatase. Analysis of the primary sequence of Rtr1 homologues revealed a conserved N-

terminal zinc finger motif and a variable C-terminal region with unknown function [164]. Puzzlingly, the primary sequence of Rtr1 does not reveal any known phosphoryl-transfer motif. Efforts to obtain crystals from *S. cerevisiae* Rtr1 have failed owing to low diffraction, with the highest resolution reported at 4 Å and the data has not been deposited in protein data bank [162]. High-resolution structures are obtained for the Rtr1 homologue from thermophilic yeast *K. lactis* (Fig. 5D) [162, 165]. Both structures highlighted the zinc finger motif but no deep surface pocket to serve as active site was detected in *K. lactis* Rtr1 [162, 165].

The discrepancy between these biological and structural results might stem from the fact that the *K. lactis* Rtr1 has a great degree of flexibility. Other than the four helices coordinating a zinc ion, very little secondary structure is observed. In fact, more than 65% of the *K. lactis* Rtr1 sequence assumes a random coil configuration. Part of the *K. lactis* Rtr1 structure was disordered (residue 88-100), breaking the continuous backbone of the polypeptide (Fig. 5D). Since the surface model of the protein was generated in the absence of these residues, it is possible that the elusive active site pocket can be found in this region. Therefore, a crystal structure of Rtr1 with a continuous backbone is needed to determine if the putative active site pocket is present.

The weak activity of Rtr1 against the CTD *in vitro* likely stems from multiple causes. First, it appears that Rtr1 and its homologue RPAP2 both favor multiple-phosphorylated CTD. For instance, RPAP2 requires Ser7 phosphorylation for recruitment to dephosphorylate Ser5 [163]. Similarly, Ser2 or Tyr1 phosphorylation appear to be required for the recruitment of Rtr1 [162, 166]. These results are consistent with Rtr1/RPAP2 recruitment to RNA pol II during the transition from initiation to elongation when the CTD becomes hyperphosphorylated [90]. This suggests a model in which Rtr1 might be recruited to the CTD through a higher order combination of phosphorylation marks along CTD. This complex combination of marks would be difficult to replicate *in vitro* and may partially account for Rtr1's low apparent activity. Secondly, Rtr1 might be targeted to the CTD through its interaction with other proteins. Two human proteins containing C-terminal binding domains (CID), RPRD1a and RPRD1b, are recently found to interact with RPAP2 (human homologue of Rtr1) and mediate its association with the CTD [167]. The authors propose that the two CID proteins form a clamp to stabilize the interaction between RPAP2 and the CTD. A recent publication provided another explanation for the discrepancy between the lack of a detectable active site in Rtr1 and the biological data showing phosphatase activity *in vivo*. A highly sensitive kinetic assay was employed to show that Rtr1 can indeed mediate phosphoryl-transfer, with wild type full-length *K. lactis* having a k_{cat} of $0.0010 \pm 0.00003 \text{ s}^{-1}$ and a K_M of $587 \pm 70 \text{ }\mu\text{M}$ against pNPP. Interestingly, deletion of C-terminal region of the *K. lactis* Rtr1 marginally increased its activity (k_{cat} of $0.0015 \pm 0.00005 \text{ s}^{-1}$ and K_M of $694 \pm 74 \text{ }\mu\text{M}$), indicating that the C-terminus may be involved in negatively regulating Rtr1 activity *in vivo* [162].

With low intrinsic activity, it seems doubtful Rtr1 is fully responsible for the dramatic reduction of Ser5 phosphorylation levels during the transition from initiation to elongation. This reasoning is supported by the fact yeast Rtr1 knockout confers temperature sensitivity rather than lethality [90, 164]. Other phosphatases such as Ssu72 might also supplement

Rtr1 phosphatase function during the transition from initiation to elongation [168]. Therefore, further investigation of Rtr1's catalytic mechanism, regulatory factors, and unique contribution to transcription must be undertaken to better understand this emerging CTD phosphatase.

2.5 Cdc14

Cdc14 is an essential cell cycle phosphatase that regulates key events during late mitosis. Cdc14 reverses Cdk-dependent phosphorylation and promotes late mitotic exit [169]. Cdc14 has been identified as a component of the silencing complex RENT. RENT is similar to the Scp associated REST complex in that it helps to silence gene expression [170, 171]. Cdc14 is thought to contribute to transcription silencing by dephosphorylating CTD [60, 61]. Recombinant Cdc14 can dephosphorylate purified, hyperphosphorylated CTD substrate *in vitro*. This dephosphorylation occurs both at Ser2 and Ser5, as determined by site-specific antibodies [60, 61]. Phosphorylated Ser5 accumulates in Cdc14-null cell lines and tagged Cdc14 coimmunoprecipitates with Rpb1 [60]. Cdc14-null cells display a shift in cyclin transcript expression, suggesting Cdc14 may be controlling a subset of genes during cell cycle [60], but it is unclear if this can be linked entirely to its CTD phosphatase activity. These experiments suggest an interesting role for Cdc14 dephosphorylation of CTD in cell cycle regulation and expand the critical nature of this regulated process.

2.6 Glc7

Recently identified Tyr1 phosphorylation is most abundant at the end of the transcription cycle and is believed to play an important role in timing termination [13]. Purified yeast cleavage and poly-adenylation factor (CPF) complex can dephosphorylate the CTD at Tyr1, Ser2, and Ser5 [152]. Ssu72 and Glc7 are the only two known phosphatases in this complex [160]. Ssu72 has no significant Tyr1 phosphatase activity *in vitro* or *in vivo*, even though it is structurally similar to established tyrosine phosphatases [110, 152]. This leaves Glc7, a di-metal ion phosphatase that is essential for the termination of snoRNA [172] and mRNA export [173]. Inactivation of Glc7 by broad-spectrum di-metal phosphatase inhibitors EDTA and microcystin abolished CPF phosphatase activity against Tyr1 and Ser2 *in vitro*. Since EDTA and microcystin do not affect the activity of PTP phosphatases like Ssu72, it was reasoned that Glc7 is responsible for the Tyr1 and Ser2 phosphatase activity of CPF *in vitro*. Depletion of Glc7 *in vivo* led to dephosphorylation defects specific to Tyr1 and resulted in increased occupancy of RNA Pol II downstream of genes, suggesting defects in termination [152]. *In vitro* purified Glc7 isolated from CPF has not been tested for phosphatase activity against CTD, owing in part to the low specificity expected of di-metal dependent phosphatases in the absence of auxiliary domains or binding partners [97, 152]. Therefore, the identification and role of Glc7 binding partners in directing Glc7 activity against CTD must be investigated to unambiguously characterize this phosphatase.

3. Proline Isomerization and Regulation of CTD Dephosphorylation

Beyond CTD kinases and phosphatases, other CTD modifiers can impact the phosphorylation state of the CTD. These include factors that alter the conformation of the CTD, which can either enhance or decrease the affinity of CTD binding partners. In turn,

changes in affinity towards the CTD over the course of transcription can regulate the timing of transcriptional processes. Primary examples of such CTD modifiers are the prolyl isomerases Ess1 in yeast and Pin1 in humans.

Prolyl isomerization plays a major role in signal transduction [174]. Proline is the only natural amino acid that stably assumes either a cis or a trans conformation about its prolyl peptide bond (Fig. 6A). The trans form is thermodynamically favored and naturally occurs ~70-90% of the time [175]. This difference in conformation can alter the interaction between proteins if the protein-protein interface contains proline residues. Additionally, the two isoforms interchange according to a thermal equilibrium, but this process is slow in the context of an entire protein and can be rate-limiting [176, 177]. To overcome this slow isomerization, nature has evolved a class of enzymes called prolyl isomerases that increase the rate of isomer conversion [178, 179]. It is important to note that prolyl isomerases do not alter total isomer ratios but instead can rapidly re-balance isomer pools when the equilibrium is broken and maintain isomer ratios at a constant level. Ess1 and Pin1 are essential phospho-specific proline isomerases that act on a variety of substrates [174]. In terms of the CTD, Ess1/Pin1 equilibrate Pro3 and Pro6 residues adjacent to phosphorylated Ser2 and Ser5 residues [179, 180]. Depletion or mutation of Ess1/Pin1 leads to global transcription defects and accumulation of CTD Ser5 phosphorylation, suggesting that proline isomerization of the CTD plays a role in transcriptional signaling [181, 182]. Ess1/Pin1 enhances CTD dephosphorylation and recognizes the same Ser-Pro binding motif as CTD phosphatases. One explanation for the role of Ess1/Pin1 in CTD dephosphorylation is that prolyl isomerases and phosphatases are functionally linked. Specifically, the proline isomer preference of CTD phosphatases may contribute to their apparent activity and be regulated by prolyl isomerases.

Unfortunately, little mechanistic data is available to explain the impact of proline isomerization on transcription at the molecular level. This is primarily due to the subtle conformational differences introduced by proline isomerization that cannot be detected using current technologies like sequencing or mass spectrometry. Until very recently only structural approaches like X-ray crystallography and NMR spectroscopy could decipher the proline isomer preference of proteins. Structural approaches require extensive data analysis and have technique specific artifacts that prohibit clear interpretation. For example, in X-ray crystallography the low abundant cis-proline is easily masked by the more abundant trans isomer. Furthermore, both X-ray crystallography and NMR require extremely high amounts of pure protein and sufficient resolution to distinguish the two isoforms.

A recent chemical biology approach has overcome a variety of these confounding factors and provided a direct, easily interpretable, and quantitative means of deciphering the proline isomer preference of CTD phosphatases [182]. By replacing the native prolyl peptide bond with a carbon-carbon double bond, proline residues can be locked into either the cis or trans conformation (Fig. 6B). When incorporated into synthetic peptides, these “locked-prolines” can be used to determine proline isomer preference of protein binding partners via kinetic or binding assays. Using this method, the proline isomer preference of the CTD phosphatases Scp1, Fcp1, and Ssu72 have been established. Scp1 and Fcp1 recognize phosphorylated CTD substrates containing trans or cis proline, but both prefer the trans proline containing

substrate (Fig. 6C). On the other hand, Ssu72 is strictly cis-proline specific (Fig. 6D). Due to these inherent proline isomer preferences, CTD phosphatases behave differently in the context of thermal isomerization versus the prolyl isomerase supplemented state (Fig. 6E & 6F). Since Scp1 and Fcp1 recognize the major trans-proline species, they are unlikely to experience rate limitation due to proline isomerization with or without prolyl isomerases and have equivalent activity in either condition (Fig. 6E). However, cis-proline specific phosphatases like Ssu72 recognize only the minor species and require prolyl isomerases to generate a constant supply of substrate, boosting apparent activity [182, 183] (Fig. 6F). This selective enhancement of CTD phosphatases translates to cellular systems where upon Pin1 depletion Ssu72 target phosphorylated Ser5 is accumulated and Fcp1 target phosphorylated Ser2 is unaffected [182]. This is in line with the *in vitro* observation that Ssu72 is cis-specific, while Fcp1 is trans-preferred. Proline isomerization directed by Ess1/Pin1, therefore, allows for divergent phosphorylation signaling along CTD and may direct different transcriptional outcomes by altering CTD phosphorylation state.

4. Perspective

The excellent work in the CTD phosphatase field has elucidated many aspects of the specificity, localization, and timing of these enzymes as well as provided several exciting avenues for future research. Undoubtedly, Ser2 and Ser5 are the best characterized phosphorylation sites on the CTD, but novel sites like Tyr1, Thr4, and Ser7 are attracting greater attention in recent years for their unique contributions to transcription [12, 13, 71, 184]. Particularly, substantial focus is being placed on how or if phosphorylation of individual sites can contribute to the status of others. Functional links at the kinase/phosphatase level have already been observed for most of these marks. Ser5 and Ser7 are both phosphorylated by TFIIF [11, 185] and dephosphorylated by Ssu72 [91, 143]. Similarly, Ser2 and Thr4 phosphorylation is catalyzed by the same kinase in yeast, PTEFb [12, 186] and are dephosphorylated at least in part by Fcp1 [108, 109]. Higher order interactions between these phosphorylation marks and the transcriptional machinery also occur in cells. For instance, Ser7 phosphorylation helps recruit RPAP2 (the human homologue to Rtr1) to the CTD, which then dephosphorylates Ser5 on snRNA genes [163] and primes the CTD for recognition by the Ser2 kinase PTEFb [184]. Furthermore, the complete removal of Ser2 marks by Fcp1 requires dephosphorylation of Ser5 by Ssu72, suggesting a level of interplay between these two marks and the phosphatases that recognize them [62]. These connections between the phosphorylation state of the CTD and CTD phosphatases support the existence of a complex network to time modification along CTD and impact transcription.

Another recurring theme for the study of CTD phosphatases is the role of binding partners to their transcriptional function. The importance of these binding partners is evident for all known CTD phosphatases. Scps, Ssu72, Cdc14, and Glc7 all operate within protein complexes that position these CTD phosphatases at the correct point in the transcription cycle [120, 160, 170]. Fcp1 and RPAP2, the human homologue of Rtr1, rely on binding partners for association with the RNA polymerase II holoenzyme [108, 167]. However, gaps in our knowledge of the of binding partners for CTD phosphatases limit our biological interpretations and more transient protein-protein interactions may not even be apparent with

current technologies. Therefore, high sensitivity proteomic analysis will be of key importance in the near future to expand our understanding of CTD phosphatases. Once full repertoires of binding partners are available CTD dephosphorylation can be better rationalized physiologically.

RNA Pol II dephosphorylation is required for cell survival and gene knockouts of CTD phosphatases can have devastating results like cell death. In such scenarios, selective chemical inhibitors have the unique advantage of shutting down CTD phosphatases quickly and reversibly. However, inhibitors that are highly potent and specific against CTD phosphatases are rarely available. Admittedly, phosphatases have historically been a challenging target for inhibitor design. This difficulty stems from three major causes: (1) affinity; the binding pockets of phosphatases tend to be small, and therefore the potential for high affinity binding of inhibitors is limited; (2) specificity; phosphatase specificity is often encoded beyond the local active site configuration, so inhibitors bound at the active site tend to cross inhibit other phosphatase family members in the cell due to the high conservation of active site geometry; and (3) permeability; phosphatases bind negatively charged phosphate groups and inhibitors with similarly charged groups are incapable of crossing cell membranes.

The difficulty in identifying inhibitors is obvious for cysteine-based phosphatases since the reactive thiol group of cysteine contributes to high false positive rates in compound screening. To overcome such limitations, a new strategy has been developed in the field of phosphatase inhibitor design using fragment-based drug discovery. In addition to the active site pocket, the phosphatase is examined for unique proximal pockets not present in other phosphatases to target by specific inhibitors. By targeting both pockets in a combinatorial fashion, potent and specific inhibitors can be obtained. The fragment-based drug discovery approach has successfully overcome these inherent difficulties for several phosphatases to yield potent and selective inhibitors. Zhong-Yin Zhang *et al.* have identified a highly selective inhibitor of the cysteine-based phosphatase TC-PTP with high potency ($K_i=4.3$ nM) that causes an accumulation of phosphorylated target proteins in cells [187]. An inhibitor against the aspartate based phosphatase UBLCP1 has recently been developed which targets the active site and an adjacent binding pocket lending both potency and specificity [188]. Finally, Scp1 is inhibited by the small molecule rabeprazole, which binds a hydrophobic pocket adjacent to the reaction center and does not cross-inhibit the essential phosphatase Fcp1 or other homologues [139]. Utilizing this strategy of targeting sites apart from the conserved reaction centers of phosphatases, potent inhibitors against CTD phosphatases will hopefully be developed. These inhibitors will allow us to directly link CTD phosphatase activity to transcription phenotypes and greatly advance the field.

The greatest hurdle remaining in the CTD field today is the resolution in which we can observe the landscape of phosphorylation marks along the CTD. Owing to current technological limitations, how these marks are patterned in individual heptads and along the sprawling CTD remain fairly nebulous. Phosphorylation specific CTD antibodies provided the leap necessary to identify the presence and abundance of specific phosphorylated residues in the CTD. Unfortunately, these antibodies do not provide the intricate molecular detail required to fully characterize the pattern of post-translational modification at specific

residues in specific locations along the CTD. Mass spectrometry provides greater resolution and the ability to assess various modifications in a discovery based fashion, but this technique has some inherent challenges in terms of investigating the CTD. These challenges include the lack of basic residues along the CTD that limit protease digestion, the repetitive nature of CTD's amino acid sequence that makes it difficult to distinguish modification sites, and CTD's inherently low ionization potential. Overcoming the difficulties of analyzing the CTD with mass spectrometry and determining the pattern of post-translational modifications along the entire CTD will change the landscape of future studies in this field.

5. Conclusion

The substantial advances in CTD phosphatases described here and the evidence that they are essential to the transcription process establishes CTD dephosphorylation as equal to the recognized importance of CTD phosphorylation. They are a diverse group of enzymes capable of enhancing and silencing gene expression. They participate in both general and gene specific transcription and contribute to cell level phenomenon like neuronal differentiation. This overcomes the misconception that phosphatases play a passive role and expands the paradigm of reversible phosphorylation beyond that of a kinase driven one-way road. Instead, CTD phosphatase and kinases behave in a circular and regulated harmony that maintains cellular balance and drives transcriptional events.

References

1. Steinmetz EJ, Warren CL, Kuehner JN, Panbehi B, Ansari AZ, Brow DA. Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. *Molecular cell*. 2006; 24:735–746. [PubMed: 17157256]
2. Koch F, Jourquin F, Ferrier P, Andrau JC. Genome-wide RNA polymerase II: not genes only! *Trends in biochemical sciences*. 2008; 33:265–273. [PubMed: 18467100]
3. Koch F, Fenouil R, Gut M, Cauchy P, Albert TK, Zacarias-Cabeza J, Spicuglia S, de la Chapelle AL, Heidemann M, Hintermair C, Eick D, Gut I, Ferrier P, Andrau JC. Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nature structural & molecular biology*. 2011; 18:956–963.
4. Jeronimo C, Bataille AR, Robert F. The writers, readers, and functions of the RNA polymerase II C-terminal domain code. *Chemical reviews*. 2013; 113:8491–8522. [PubMed: 23837720]
5. Kang ME, Dahmus ME. RNA polymerases IIA and IIO have distinct roles during transcription from the TATA-less murine dihydrofolate reductase promoter. *The Journal of biological chemistry*. 1993; 268:25033–25040. [PubMed: 8227067]
6. Kelly WG, Dahmus ME, Hart GW. RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc. *The Journal of biological chemistry*. 1993; 268:10416–10424. [PubMed: 8486697]
7. Cho EJ, Takagi T, Moore CR, Buratowski S. mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes & development*. 1997; 11:3319–3326. [PubMed: 9407025]
8. Ho CK, Shuman S. Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. *Molecular cell*. 1999; 3:405–411. [PubMed: 10198643]
9. Chapman RD, Heidemann M, Albert TK, Mailhammer R, Flatley A, Meisterernst M, Kremmer E, Eick D. Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science*. 2007; 318:1780–1782. [PubMed: 18079404]

10. Egloff S, O'Reilly D, Chapman RD, Taylor A, Tanzhaus K, Pitts L, Eick D, Murphy S. Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science*. 2007; 318:1777–1779. [PubMed: 18079403]
11. Glover-Cutter K, Larochelle S, Erickson B, Zhang C, Shokat K, Fisher RP, Bentley DL. TFIIF-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II. *Molecular and cellular biology*. 2009; 29:5455–5464. [PubMed: 19667075]
12. Hsin JP, Sheth A, Manley JL. RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing. *Science*. 2011; 334:683–686. [PubMed: 22053051]
13. Mayer A, Heidemann M, Lidschreiber M, Schreieck A, Sun M, Hintermair C, Kremmer E, Eick D, Cramer P. CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science*. 2012; 336:1723–1725. [PubMed: 22745433]
14. Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic acids research*. 2015; 43:D512–520. [PubMed: 25514926]
15. Allison LA, Moyle M, Shales M, Ingles CJ. Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell*. 1985; 42:599–610. [PubMed: 3896517]
16. Corden JL, Cadena DL, Ahearn JM Jr. Dahmus ME. A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America*. 1985; 82:7934–7938. [PubMed: 2999785]
17. Chapman RD, Heidemann M, Hintermair C, Eick D. Molecular evolution of the RNA polymerase II CTD. *Trends in genetics : TIG*. 2008; 24:289–296. [PubMed: 18472177]
18. Patturajan M, Schulte RJ, Sefton BM, Berezney R, Vincent M, Bensaude O, Warren SL, Corden JL. Growth-related changes in phosphorylation of yeast RNA polymerase II. *The Journal of biological chemistry*. 1998; 273:4689–4694. [PubMed: 9468530]
19. Hammarsten O. Zur Frage ob Casein ein einheitlicher Stoff sei., *Hoppe-Seyler's Z. Physiol. Chem.* 1883:227–273.
20. Suzuki, K, UY.; Takaishi, M. The bulletin of the College of Agriculture. Vol. 7. Tokyo Imperial University; 1907. Ueber ein Enzym "Phytase," das "Anhydro-oxy-methylen diphosphorsäure" spaltet.; p. 507-512.
21. McComb, RB.; Bowers, GN.; Posen, S. Alkaline phosphatase. Plenum Press; New York: 1979.
22. Robison R. The Possible Significance of Hexosephosphoric Esters in Ossification. *The Biochemical journal*. 1923; 17:286–293. [PubMed: 16743183]
23. Robison R, Soames KM. The Possible Significance of Hexosephosphoric Esters in Ossification: Part II. The Phosphoric Esterase of Ossifying Cartilage. *The Biochemical journal*. 1924; 18:740–754. [PubMed: 16743453]
24. Levene FALPA. Serinephosphoric acid obtained on hydrolysis of vitellinic acid. *The Journal of biological chemistry*. 1932; 98:109–114.
25. Green GTCAA. Crystalline muscle phosphorylase: II. prosthetic group. *The Journal of biological chemistry*. 1943; 151:31–38.
26. Wosilait WD, Sutherland EW. The relationship of epinephrine and glucagon to liver phosphorylase. II. Enzymatic inactivation of liver phosphorylase. *The Journal of biological chemistry*. 1956; 218:469–481. [PubMed: 13278354]
27. De Verdier CH. Isolation of phosphothreonine from bovine casein. *Nature*. 1952; 170:804–805. [PubMed: 13002453]
28. Burnett G, Kennedy EP. The enzymatic phosphorylation of proteins. *The Journal of biological chemistry*. 1954; 211:969–980. [PubMed: 13221602]
29. Fischer EH, Krebs EG. Conversion of phosphorylase b to phosphorylase a in muscle extracts. *The Journal of biological chemistry*. 1955; 216:121–132. [PubMed: 13252012]
30. Sutherland EW Jr. Wosilait WD. Inactivation and activation of liver phosphorylase. *Nature*. 1955; 175:169–170. [PubMed: 13235837]
31. Krebs EG, Fischer EH. The phosphorylase b to a converting enzyme of rabbit skeletal muscle. *Biochimica et biophysica acta*. 1956; 20:150–157. [PubMed: 13315361]

32. Krebs EG, Kent AB, Fischer EH. The muscle phosphorylase b kinase reaction. *The Journal of biological chemistry*. 1958; 231:73–83. [PubMed: 13538949]
33. Walsh DA, Perkins JP, Krebs EG. An adenosine 3',5'-monophosphate-dependant protein kinase from rabbit skeletal muscle. *The Journal of biological chemistry*. 1968; 243:3763–3765. [PubMed: 4298072]
34. Cohen P, Watson DC, Dixon GH. The hormonal control of activity of skeletal muscle phosphorylase kinase. Amino-acid sequences at the two sites of action of adenosine-3':5'-monophosphate-dependent protein kinase. *European journal of biochemistry / FEBS*. 1975; 51:79–92. [PubMed: 164350]
35. Daile P, Carnegie PR, Young JD. Synthetic substrate for cyclic AMP-dependent protein kinase. *Nature*. 1975; 257:416–418. [PubMed: 170533]
36. Kemp BE, Benjamini E, Krebs EG. Synthetic hexapeptide substrates and inhibitors of 3':5'-cyclic AMP-dependent protein kinase. *Proceedings of the National Academy of Sciences of the United States of America*. 1976; 73:1038–1042. [PubMed: 177970]
37. Zetterqvist O, Ragnarsson U, Humble E, Berglund L, Engstrom L. The minimum substrate of cyclic AMP-stimulated protein kinase, as studied by synthetic peptides representing the phosphorylatable site of pyruvate kinase (type L) of rat liver. *Biochemical and biophysical research communications*. 1976; 70:696–703. [PubMed: 180994]
38. Eckhart W, Hutchinson MA, Hunter T. An activity phosphorylating tyrosine in polyoma T antigen immunoprecipitates. *Cell*. 1979; 18:925–933. [PubMed: 229973]
39. Collett MS, Erikson RL. Protein kinase activity associated with the avian sarcoma virus src gene product. *Proceedings of the National Academy of Sciences of the United States of America*. 1978; 75:2021–2024. [PubMed: 205879]
40. Vogt PK. Retroviral oncogenes: a historical primer. *Nature reviews. Cancer*. 2012; 12:639–648.
41. Tonks NK, Diltz CD, Fischer EH. Purification of the major protein-tyrosinephosphatases of human placenta. *The Journal of biological chemistry*. 1988; 263:6722–6730. [PubMed: 2834386]
42. Hanks SK, Quinn AM, Hunter T. The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science*. 1988; 241:42–52. [PubMed: 3291115]
43. Knighton DR, Zheng JH, Ten Eyck LF, Ashford VA, Xuong NH, Taylor SS, Sowadski JM. Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science*. 1991; 253:407–414. [PubMed: 1862342]
44. Barford D, Flint AJ, Tonks NK. Crystal structure of human protein tyrosine phosphatase 1B. *Science*. 1994; 263:1397–1404. [PubMed: 8128219]
45. Goldberg J, Huang HB, Kwon YG, Greengard P, Nairn AC, Kuriyan J. Three-dimensional structure of the catalytic subunit of protein serine/threonine phosphatase-1. *Nature*. 1995; 376:745–753. [PubMed: 7651533]
46. Engel C, Sainsbury S, Cheung AC, Kostrewa D, Cramer P. RNA polymerase I structure and transcription regulation. *Nature*. 2013; 502:650–655. [PubMed: 24153182]
47. Eick D, Geyer M. The RNA polymerase II carboxy-terminal domain (CTD) code. *Chemical reviews*. 2013; 113:8456–8490. [PubMed: 23952966]
48. Lu H, Flores O, Weinmann R, Reinberg D. The nonphosphorylated form of RNA polymerase II preferentially associates with the preinitiation complex. *Proceedings of the National Academy of Sciences of the United States of America*. 1991; 88:10004–10008. [PubMed: 1946417]
49. Conaway RC, Bradsher JN, Conaway JW. Mechanism of assembly of the RNA polymerase II preinitiation complex. Evidence for a functional interaction between the carboxyl-terminal domain of the largest subunit of RNA polymerase II and a high molecular mass form of the TATA factor. *The Journal of biological chemistry*. 1992; 267:8464–8467. [PubMed: 1569096]
50. Schwer B, Shuman S. Deciphering the RNA polymerase II CTD code in fission yeast. *Molecular cell*. 2011; 43:311–318. [PubMed: 21684186]
51. McCracken S, Fong N, Yankulov K, Ballantyne S, Pan G, Greenblatt J, Patterson SD, Wickens M, Bentley DL. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature*. 1997; 385:357–361. [PubMed: 9002523]

52. Morris DP, Greenleaf AL. The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. *The Journal of biological chemistry*. 2000; 275:39935–39943. [PubMed: 10978320]
53. Hirose Y, Manley JL. RNA polymerase II is an essential mRNA polyadenylation factor. *Nature*. 1998; 395:93–96. [PubMed: 9738505]
54. Richard P, Manley JL. Transcription termination by nuclear RNA polymerases. *Genes & development*. 2009; 23:1247–1269. [PubMed: 19487567]
55. Kuehner JN, Pearson EL, Moore C. Unravelling the means to an end: RNA polymerase II transcription termination. *Nature reviews. Molecular cell biology*. 2011; 12:283–294. [PubMed: 21487437]
56. Ahn SH, Kim M, Buratowski S. Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Molecular cell*. 2004; 13:67–76. [PubMed: 14731395]
57. Ng HH, Robert F, Young RA, Struhl K. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Molecular cell*. 2003; 11:709–719. [PubMed: 12667453]
58. Winsor TS, Bartkowiak B, Bennett CB, Greenleaf AL. A DNA damage response system associated with the phosphoCTD of elongating RNA polymerase II. *PloS one*. 2013; 8:e60909. [PubMed: 23613755]
59. Di Vona C, Bezdan D, Islam AB, Salichs E, Lopez-Bigas N, Ossowski S, de la Luna S. Chromatin-wide profiling of DYRK1A reveals a role as a gene-specific RNA polymerase II CTD kinase. *Molecular cell*. 2015; 57:506–520. [PubMed: 25620562]
60. Guillaumot M, Manchado E, Chiesa M, Gomez-Lopez G, Pisano DG, Sacristan MP, Malumbres M. Cdc14b regulates mammalian RNA polymerase II and represses cell cycle transcription. *Scientific reports*. 2011; 1:189. [PubMed: 22355704]
61. Clemente-Blanco A, Sen N, Mayan-Santos M, Sacristan MP, Graham B, Jarmuz A, Giess A, Webb E, Game L, Eick D, Bueno A, Merkenschlager M, Aragon L. Cdc14 phosphatase promotes segregation of telomeres through repression of RNA polymerase II transcription. *Nature cell biology*. 2011; 13:1450–1456. [PubMed: 22020438]
62. Bataille AR, Jeronimo C, Jacques PE, Laramée L, Fortin ME, Forest A, Bergeron M, Hanes SD, Robert F. A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. *Molecular cell*. 2012; 45:158–170. [PubMed: 22284676]
63. Komarnitsky P, Cho EJ, Buratowski S. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes & development*. 2000; 14:2452–2460. [PubMed: 11018013]
64. Kim H, Erickson B, Luo W, Seward D, Graber JH, Pollock DD, Megee PC, Bentley DL. Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nature structural & molecular biology*. 2010; 17:1279–1286.
65. Mayer A, Lidschreiber M, Siebert M, Leike K, Soding J, Cramer P. Uniform transitions of the general RNA polymerase II transcription complex. *Nature structural & molecular biology*. 2010; 17:1272–1278.
66. Tietjen JR, Zhang DW, Rodriguez-Molina JB, White BE, Akhtar MS, Heidemann M, Li X, Chapman RD, Shokat K, Keles S, Eick D, Ansari AZ. Chemical-genomic dissection of the CTD code. *Nature structural & molecular biology*. 2010; 17:1154–1161.
67. Chesnut JD, Stephens JH, Dahmus ME. The interaction of RNA polymerase II with the adenovirus-2 major late promoter is precluded by phosphorylation of the C-terminal domain of subunit IIa. *The Journal of biological chemistry*. 1992; 267:10500–10506. [PubMed: 1316903]
68. Descostes N, Heidemann M, Spinelli L, Schuller R, Maqbool MA, Fenouil R, Koch F, Innocenti C, Gut M, Gut I, Eick D, Andrau JC. Tyrosine phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells. *eLife*. 2014; 3:e02105. [PubMed: 24842994]
69. Hsin JP, Li W, Hoque M, Tian B, Manley JL. RNAP II CTD tyrosine 1 performs diverse functions in vertebrate cells. *eLife*. 2014; 3:e02112. [PubMed: 24842995]

70. Cho H, Kim TK, Mancebo H, Lane WS, Flores O, Reinberg D. A protein phosphatase functions to recycle RNA polymerase II. *Genes & development*. 1999; 13:1540–1552. [PubMed: 10385623]
71. Hintermair C, Heidemann M, Koch F, Descostes N, Gut M, Gut I, Fenouil R, Ferrier P, Flatley A, Kremmer E, Chapman RD, Andrau JC, Eick D. Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. *The EMBO journal*. 2012; 31:2784–2797. [PubMed: 22549466]
72. Buratowski S. The CTD code. *Nature structural biology*. 2003; 10:679–680. [PubMed: 12942140]
73. Lu H, Zawel L, Fisher L, Egly JM, Reinberg D. Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II. *Nature*. 1992; 358:641–645. [PubMed: 1495560]
74. Trigon S, Serizawa H, Conaway JW, Conaway RC, Jackson SP, Morange M. Characterization of the residues phosphorylated in vitro by different C-terminal domain kinases. *The Journal of biological chemistry*. 1998; 273:6769–6775. [PubMed: 9506978]
75. Ni Z, Saunders A, Fuda NJ, Yao J, Suarez JR, Webb WW, Lis JT. P-TEFb is critical for the maturation of RNA polymerase II into productive elongation in vivo. *Molecular and cellular biology*. 2008; 28:1161–1170. [PubMed: 18070927]
76. Akoulitchev S, Makela TP, Weinberg RA, Reinberg D. Requirement for TFIIF kinase activity in transcription by RNA polymerase II. *Nature*. 1995; 377:557–560. [PubMed: 7566158]
77. Jiang Y, Yan M, Gralla JD. A three-step pathway of transcription initiation leading to promoter clearance at an activation RNA polymerase II promoter. *Molecular and cellular biology*. 1996; 16:1614–1621. [PubMed: 8657136]
78. Jeronimo C, Robert F. Kin28 regulates the transient association of Mediator with core promoters. *Nature structural & molecular biology*. 2014; 21:449–455.
79. Ghosh A, Shuman S, Lima CD. Structural insights to how mammalian capping enzyme reads the CTD code. *Molecular cell*. 2011; 43:299–310. [PubMed: 21683636]
80. Wada T, Takagi T, Yamaguchi Y, Watanabe D, Handa H. Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription in vitro. *The EMBO journal*. 1998; 17:7395–7403. [PubMed: 9857195]
81. Yamaguchi Y, Takagi T, Wada T, Yano K, Furuya A, Sugimoto S, Hasegawa J, Handa H. NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell*. 1999; 97:41–51. [PubMed: 10199401]
82. Ping YH, Rana TM. DSIF and NELF interact with RNA polymerase II elongation complex and HIV-1 Tat stimulates P-TEFb-mediated phosphorylation of RNA polymerase II and DSIF during transcription elongation. *The Journal of biological chemistry*. 2001; 276:12951–12958. [PubMed: 11112772]
83. Fujinaga K, Irwin D, Huang Y, Taube R, Kurosu T, Peterlin BM. Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. *Molecular and cellular biology*. 2004; 24:787–795. [PubMed: 14701750]
84. Donner AJ, Ebmeier CC, Taatjes DJ, Espinosa JM. CDK8 is a positive regulator of transcriptional elongation within the serum response network. *Nature structural & molecular biology*. 2010; 17:194–201.
85. Bartkowiak B, Liu P, Phatnani HP, Fuda NJ, Cooper JJ, Price DH, Adelman K, Lis JT, Greenleaf AL. CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes & development*. 2010; 24:2303–2316. [PubMed: 20952539]
86. Liang K, Gao X, Gilmore JM, Florens L, Washburn MP, Smith E, Shilatifard A. Characterization of human cyclin-dependent kinase 12 (CDK12) and CDK13 complexes in C-terminal domain phosphorylation, gene transcription, and RNA processing. *Molecular and cellular biology*. 2015; 35:928–938. [PubMed: 25561469]
87. Corden JL. RNA polymerase II C-terminal domain: Tethering transcription to transcript and template. *Chemical reviews*. 2013; 113:8423–8455. [PubMed: 24040939]
88. Yogesha SD, Mayfield JE, Zhang Y. Cross-talk of phosphorylation and prolyl isomerization of the C-terminal domain of RNA Polymerase II. *Molecules*. 2014; 19:1481–1511. [PubMed: 24473209]

89. Ganem C, Devaux F, Torchet C, Jacq C, Quevillon-Cheruel S, Labesse G, Facca C, Faye G. Ssu72 is a phosphatase essential for transcription termination of snoRNAs and specific mRNAs in yeast. *The EMBO journal*. 2003; 22:1588–1598. [PubMed: 12660165]
90. Mosley AL, Pattenden SG, Carey M, Venkatesh S, Gilmore JM, Florens L, Workman JL, Washburn MP. Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation. *Molecular cell*. 2009; 34:168–178. [PubMed: 19394294]
91. Zhang DW, Mosley AL, Ramisetty SR, Rodriguez-Molina JB, Washburn MP, Ansari AZ. Ssu72 phosphatase-dependent erasure of phospho-Ser7 marks on the RNA polymerase II C-terminal domain is essential for viability and transcription termination. *The Journal of biological chemistry*. 2012; 287:8541–8551. [PubMed: 22235117]
92. Dichtl B, Blank D, Ohnacker M, Friedlein A, Roeder D, Langen H, Keller W. A role for SSU72 in balancing RNA polymerase II transcription elongation and termination. *Molecular cell*. 2002; 10:1139–1150. [PubMed: 12453421]
93. Fuda NJ, Buckley MS, Wei W, Core LJ, Waters CT, Reinberg D, Lis JT. Fcp1 dephosphorylation of the RNA polymerase II C-terminal domain is required for efficient transcription of heat shock genes. *Molecular and cellular biology*. 2012; 32:3428–3437. [PubMed: 22733996]
94. Kobor MS, Archambault J, Lester W, Holstege FC, Gileadi O, Jansma DB, Jennings EG, Kouyoumdjian F, Davidson AR, Young RA, Greenblatt J. An unusual eukaryotic protein phosphatase required for transcription by RNA polymerase II and CTD dephosphorylation in *S. cerevisiae*. *Molecular cell*. 1999; 4:55–62. [PubMed: 10445027]
95. Sun ZW, Hampsey M. Synthetic enhancement of a TFIIB defect by a mutation in SSU72, an essential yeast gene encoding a novel protein that affects transcription start site selection in vivo. *Molecular and cellular biology*. 1996; 16:1557–1566. [PubMed: 8657130]
96. Archambault J, Chambers RS, Kobor MS, Ho Y, Cartier M, Bolotin D, Andrews B, Kane CM, Greenblatt J. An essential component of a C-terminal domain phosphatase that interacts with transcription factor IIF in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America*. 1997; 94:14300–14305. [PubMed: 9405607]
97. Shi Y. Serine/threonine phosphatases: mechanism through structure. *Cell*. 2009; 139:468–484. [PubMed: 19879837]
98. Tonks NK. Protein tyrosine phosphatases--from housekeeping enzymes to master regulators of signal transduction. *The FEBS journal*. 2013; 280:346–378. [PubMed: 23176256]
99. Barford D, Das AK, Egloff MP. The structure and mechanism of protein phosphatases: insights into catalysis and regulation. *Annual review of biophysics and biomolecular structure*. 1998; 27:133–164.
100. Allen KN, Dunaway-Mariano D. Phosphoryl group transfer: evolution of a catalytic scaffold. *Trends in biochemical sciences*. 2004; 29:495–503. [PubMed: 15337123]
101. Chambers RS, Dahmus ME. Purification and characterization of a phosphatase from HeLa cells which dephosphorylates the C-terminal domain of RNA polymerase II. *The Journal of biological chemistry*. 1994; 269:26243–26248. [PubMed: 7929341]
102. Chambers RS, Kane CM. Purification and characterization of an RNA polymerase II phosphatase from yeast. *The Journal of biological chemistry*. 1996; 271:24498–24504. [PubMed: 8798710]
103. Kobor MS, Simon LD, Omichinski J, Zhong G, Archambault J, Greenblatt J. A motif shared by TFIIF and TFIIB mediates their interaction with the RNA polymerase II carboxy-terminal domain phosphatase Fcp1p in *Saccharomyces cerevisiae*. *Molecular and cellular biology*. 2000; 20:7438–7449. [PubMed: 11003641]
104. Mandal SS, Cho H, Kim S, Cabane K, Reinberg D. FCP1, a phosphatase specific for the heptapeptide repeat of the largest subunit of RNA polymerase II, stimulates transcription elongation. *Molecular and cellular biology*. 2002; 22:7543–7552. [PubMed: 12370301]
105. Archambault J, Pan G, Dahmus GK, Cartier M, Marshall N, Zhang S, Dahmus ME, Greenblatt J. FCP1, the RAP74-interacting subunit of a human protein phosphatase that dephosphorylates the carboxyl-terminal domain of RNA polymerase II. *The Journal of biological chemistry*. 1998; 273:27593–27601. [PubMed: 9765293]

106. Hausmann S, Shuman S. Characterization of the CTD phosphatase Fcp1 from fission yeast. Preferential dephosphorylation of serine 2 versus serine 5. *The Journal of biological chemistry*. 2002; 277:21213–21220. [PubMed: 11934898]
107. Cho EJ, Kobor MS, Kim M, Greenblatt J, Buratowski S. Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. *Genes & development*. 2001; 15:3319–3329. [PubMed: 11751637]
108. Allepuz-Fuster P, Martinez-Fernandez V, Garrido-Godino AI, Alonso-Aguado S, Hanes SD, Navarro F, Calvo O. Rpb4/7 facilitates RNA polymerase II CTD dephosphorylation. *Nucleic acids research*. 2014; 42:13674–13688. [PubMed: 25416796]
109. Hsin JP, Xiang K, Manley JL. Function and control of RNA polymerase II C-terminal domain phosphorylation in vertebrate transcription and RNA processing. *Molecular and cellular biology*. 2014; 34:2488–2498. [PubMed: 24752900]
110. Xiang K, Manley JL, Tong L. An unexpected binding mode for a Pol II CTD peptide phosphorylated at Ser7 in the active site of the CTD phosphatase Ssu72. *Genes & development*. 2012; 26:2265–2270. [PubMed: 23070812]
111. Kimura M, Suzuki H, Ishihama A. Formation of a carboxy-terminal domain phosphatase (Fcp1)/TFIIF/RNA polymerase II (pol II) complex in *Schizosaccharomyces pombe* involves direct interaction between Fcp1 and the Rpb4 subunit of pol II. *Molecular and cellular biology*. 2002; 22:1577–1588. [PubMed: 11839823]
112. Ghosh A, Shuman S, Lima CD. The structure of Fcp1, an essential RNA polymerase II CTD phosphatase. *Molecular cell*. 2008; 32:478–490. [PubMed: 19026779]
113. Chambers RS, Wang BQ, Burton ZF, Dahmus ME. The activity of COOH-terminal domain phosphatase is regulated by a docking site on RNA polymerase II and by the general transcription factors IIF and IIB. *The Journal of biological chemistry*. 1995; 270:14962–14969. [PubMed: 7797476]
114. Kamada K, Roeder RG, Burley SK. Molecular mechanism of recruitment of TFIIF- associating RNA polymerase C-terminal domain phosphatase (FCP1) by transcription factor IIF. *Proceedings of the National Academy of Sciences of the United States of America*. 2003; 100:2296–2299. [PubMed: 12591941]
115. Nguyen BD, Chen HT, Kobor MS, Greenblatt J, Legault P, Omichinski JG. Solution structure of the carboxyl-terminal domain of RAP74 and NMR characterization of the FCP1-binding sites of RAP74 and human TFIIB. *Biochemistry*. 2003; 42:1460–1469. [PubMed: 12578358]
116. Yang A, Abbott KL, Desjardins A, Di Lello P, Omichinski JG, Legault P. NMR structure of a complex formed by the carboxyl-terminal domain of human RAP74 and a phosphorylated peptide from the central domain of the FCP1 phosphatase. *Biochemistry*. 2009; 48:1964–1974. [PubMed: 19215094]
117. Yu X, Chini CC, He M, Mer G, Chen J. The BRCT domain is a phospho-protein binding domain. *Science*. 2003; 302:639–642. [PubMed: 14576433]
118. Friedl EM, Lane WS, Erdjument-Bromage H, Tempst P, Reinberg D. The C-terminal domain phosphatase and transcription elongation activities of FCP1 are regulated by phosphorylation. *Proceedings of the National Academy of Sciences of the United States of America*. 2003; 100:2328–2333. [PubMed: 12591939]
119. Yeo M, Lin PS, Dahmus ME, Gill GN. A novel RNA polymerase II C-terminal domain phosphatase that preferentially dephosphorylates serine 5. *The Journal of biological chemistry*. 2003; 278:26078–26085. [PubMed: 12721286]
120. Yeo M, Lee SK, Lee B, Ruiz EC, Pfaff SL, Gill GN. Small CTD phosphatases function in silencing neuronal gene expression. *Science*. 2005; 307:596–600. [PubMed: 15681389]
121. Kamenski T, Heilmeyer S, Meinhart A, Cramer P. Structure and mechanism of RNA polymerase II CTD phosphatases. *Molecular cell*. 2004; 15:399–407. [PubMed: 15304220]
122. Zhang M, Liu J, Kim Y, Dixon JE, Pfaff SL, Gill GN, Noel JP, Zhang Y. Structural and functional analysis of the phosphoryl transfer reaction mediated by the human small C-terminal domain phosphatase, Scp1. *Protein science : a publication of the Protein Society*. 2010; 19:974–986. [PubMed: 20222012]

123. Zhang Y, Kim Y, Genoud N, Gao J, Kelly JW, Pfaff SL, Gill GN, Dixon JE, Noel JP. Determinants for dephosphorylation of the RNA polymerase II C-terminal domain by Scp1. *Molecular cell*. 2006; 24:759–770. [PubMed: 17157258]
124. Siniossoglou S, Santos-Rosa H, Rappsilber J, Mann M, Hurt E. A novel complex of membrane proteins required for formation of a spherical nucleus. *The EMBO journal*. 1998; 17:6449–6464. [PubMed: 9822591]
125. Siniossoglou S, Hurt EC, Pelham HR. Psr1p/Psr2p, two plasma membrane phosphatases with an essential DXDX(T/V) motif required for sodium stress response in yeast. *The Journal of biological chemistry*. 2000; 275:19352–19360. [PubMed: 10777497]
126. Xue Y, Ouyang K, Huang J, Zhou Y, Ouyang H, Li H, Wang G, Wu Q, Wei C, Bi Y, Jiang L, Cai Z, Sun H, Zhang K, Zhang Y, Chen J, Fu XD. Direct conversion of fibroblasts to neurons by reprogramming PTB-regulated microRNA circuits. *Cell*. 2013; 152:82–96. [PubMed: 23313552]
127. Chong JA, Tapia-Ramirez J, Kim S, Toledo-Aral JJ, Zheng Y, Boutros MC, Altshuler YM, Frohman MA, Kraner SD, Mandel G. REST: a mammalian silencer protein that restricts sodium channel gene expression to neurons. *Cell*. 1995; 80:949–957. [PubMed: 7697725]
128. Naruse Y, Aoki T, Kojima T, Mori N. Neural restrictive silencer factor recruits mSin3 and histone deacetylase complex to repress neuron-specific target genes. *Proceedings of the National Academy of Sciences of the United States of America*. 1999; 96:13691–13696. [PubMed: 10570134]
129. Jepsen K, Hermanson O, Onami TM, Gleiberman AS, Lunyak V, McEvilly RJ, Kurokawa R, Kumar V, Liu F, Seto E, Hedrick SM, Mandel G, Glass CK, Rose DW, Rosenfeld MG. Combinatorial roles of the nuclear receptor corepressor in transcription and development. *Cell*. 2000; 102:753–763. [PubMed: 11030619]
130. Nesti E, Corson GM, McCleskey M, Oyer JA, Mandel G. C-terminal domain small phosphatase 1 and MAP kinase reciprocally control REST stability and neuronal differentiation. *Proceedings of the National Academy of Sciences of the United States of America*. 2014; 111:E3929–3936. [PubMed: 25197063]
131. Andres ME, Burger C, Peral-Rubio MJ, Battaglioli E, Anderson ME, Grimes J, Dallman J, Ballas N, Mandel G. CoREST: a functional corepressor required for regulation of neural-specific gene expression. *Proceedings of the National Academy of Sciences of the United States of America*. 1999; 96:9873–9878. [PubMed: 10449787]
132. Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W, Tuschl T. Identification of tissue-specific microRNAs from mouse. *Current biology : CB*. 2002; 12:735–739. [PubMed: 12007417]
133. Conaco C, Otto S, Han JJ, Mandel G. Reciprocal actions of REST and a microRNA promote neuronal identity. *Proceedings of the National Academy of Sciences of the United States of America*. 2006; 103:2422–2427. [PubMed: 16461918]
134. Wu J, Xie X. Comparative sequence analysis reveals an intricate network among REST, CREB and miRNA in mediating neuronal gene expression. *Genome biology*. 2006; 7:R85. [PubMed: 17002790]
135. Visvanathan J, Lee S, Lee B, Lee JW, Lee SK. The microRNA miR-124 antagonizes the anti-neural REST/SCP1 pathway during embryonic CNS development. *Genes & development*. 2007; 21:744–749. [PubMed: 17403776]
136. Makeyev EV, Zhang J, Carrasco MA, Maniatis T. The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. *Molecular cell*. 2007; 27:435–448. [PubMed: 17679093]
137. Sowa N, Horie T, Kuwabara Y, Baba O, Watanabe S, Nishi H, Kinoshita M, Takanabe-Mori R, Wada H, Shimatsu A, Hasegawa K, Kimura T, Ono K. MicroRNA 26b encoded by the intron of small CTD phosphatase (SCP) 1 has an antagonistic effect on its host gene. *Journal of cellular biochemistry*. 2012; 113:3455–3465. [PubMed: 22678827]
138. Dill H, Linder B, Fehr A, Fischer U. Intronic miR-26b controls neuronal differentiation by repressing its host transcript, ctdsp2. *Genes & development*. 2012; 26:25–30. [PubMed: 22215807]

139. Zhang M, Cho EJ, Burstein G, Siegel D, Zhang Y. Selective inactivation of a human neuronal silencing phosphatase by a small molecule inhibitor. *ACS chemical biology*. 2011; 6:511–519. [PubMed: 21348431]
140. Qian H, Ji C, Zhao S, Chen J, Jiang M, Zhang Y, Yan M, Zheng D, Sun Y, Xie Y, Mao Y. Expression and characterization of HSPC129, a RNA polymerase II C-terminal domain phosphatase. *Molecular and cellular biochemistry*. 2007; 303:183–188. [PubMed: 17487459]
141. Zheng H, Ji C, Gu S, Shi B, Wang J, Xie Y, Mao Y. Cloning and characterization of a novel RNA polymerase II C-terminal domain phosphatase. *Biochemical and biophysical research communications*. 2005; 331:1401–1407. [PubMed: 15883030]
142. Guo X, Engel JL, Xiao J, Tagliabracchi VS, Wang X, Huang L, Dixon JE. UBLCP1 is a 26S proteasome phosphatase that regulates nuclear proteasome activity. *Proceedings of the National Academy of Sciences of the United States of America*. 2011; 108:18649–18654. [PubMed: 21949367]
143. Krishnamurthy S, He X, Reyes-Reyes M, Moore C, Hampsey M. Ssu72 Is an RNA polymerase II CTD phosphatase. *Molecular cell*. 2004; 14:387–394. [PubMed: 15125841]
144. Xiang K, Nagaike T, Xiang S, Kilic T, Beh MM, Manley JL, Tong L. Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex. *Nature*. 2010; 467:729–733. [PubMed: 20861839]
145. Pappas DL Jr. Hampsey M. Functional interaction between Ssu72 and the Rpb2 subunit of RNA polymerase II in *Saccharomyces cerevisiae*. *Molecular and cellular biology*. 2000; 20:8343–8351. [PubMed: 11046131]
146. Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM, Remor M, Hofert C, Schelder M, Brajenovic M, Ruffner H, Merino A, Klein K, Hudak M, Dickson D, Rudi T, Gnau V, Bauch A, Bastuck S, Huhse B, Leutwein C, Heurtier MA, Copley RR, Edelman A, Querfurth E, Rybin V, Drewes G, Raida M, Bouwmeester T, Bork P, Seraphin B, Kuster B, Neubauer G, Superti-Furga G. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*. 2002; 415:141–147. [PubMed: 11805826]
147. Steinmetz EJ, Brow DA. Ssu72 protein mediates both poly(A)-coupled and poly(A)-independent termination of RNA polymerase II transcription. *Molecular and cellular biology*. 2003; 23:6339–6349. [PubMed: 12944462]
148. Hampsey M, Singh BN, Ansari A, Laine JP, Krishnamurthy S. Control of eukaryotic gene expression: gene loops and transcriptional memory. *Advances in enzyme regulation*. 2011; 51:118–125. [PubMed: 21036187]
149. Tan-Wong SM, Zaugg JB, Camblong J, Xu Z, Zhang DW, Mischo HE, Ansari AZ, Luscombe NM, Steinmetz LM, Proudfoot NJ. Gene loops enhance transcriptional directionality. *Science*. 2012; 338:671–675. [PubMed: 23019609]
150. Meinhart A, Silberzahn T, Cramer P. The mRNA transcription/processing factor Ssu72 is a potential tyrosine phosphatase. *The Journal of biological chemistry*. 2003; 278:15917–15921. [PubMed: 12606538]
151. Zhang Y, Zhang M, Zhang Y. Crystal structure of Ssu72, an essential eukaryotic phosphatase specific for the C-terminal domain of RNA polymerase II, in complex with a transition state analogue. *The Biochemical journal*. 2011; 434:435–444. [PubMed: 21204787]
152. Schreieck A, Easter AD, Etzold S, Wiederhold K, Lidschreiber M, Cramer P, Passmore LA. RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7. *Nature structural & molecular biology*. 2014; 21:175–179.
153. Werner-Allen JW, Lee CJ, Liu P, Nicely NI, Wang S, Greenleaf AL, Zhou P. cis-Proline-mediated Ser(P)5 dephosphorylation by the RNA polymerase II C-terminal domain phosphatase Ssu72. *The Journal of biological chemistry*. 2011; 286:5717–5726. [PubMed: 21159777]
154. Hausmann S, Koiwa H, Krishnamurthy S, Hampsey M, Shuman S. Different strategies for carboxyl-terminal domain (CTD) recognition by serine 5-specific CTD phosphatases. *The Journal of biological chemistry*. 2005; 280:37681–37688. [PubMed: 16148005]
155. Pelech SL. Networking with proline-directed protein kinases implicated in tau phosphorylation. *Neurobiology of aging*. 1995; 16:247–256. discussion 257–261. [PubMed: 7566335]

156. Weiwad M, Kullertz G, Schutkowski M, Fischer G. Evidence that the substrate backbone conformation is critical to phosphorylation by p42 MAP kinase. *FEBS letters*. 2000; 478:39–42. [PubMed: 10922466]
157. Etzkorn FA, Zhao S. Stereospecific phosphorylation by the central mitotic kinase Cdk1-cyclin B. *ACS chemical biology*. 2015; 10:952–956. [PubMed: 25603287]
158. Sanso M, Fisher RP. Pause, play, repeat: CDKs push RNAP II's buttons. *Transcription*. 2013; 4:146–152. [PubMed: 23756342]
159. Luo Y, Yogesha SD, Cannon JR, Yan W, Ellington AD, Brodbelt JS, Zhang Y. novel modifications on C-terminal domain of RNA polymerase II can fine-tune the phosphatase activity of Ssu72. *ACS chemical biology*. 2013; 8:2042–2052. [PubMed: 23844594]
160. Nedeia E, He X, Kim M, Pootoolal J, Zhong G, Canadien V, Hughes T, Buratowski S, Moore CL, Greenblatt J. Organization and function of APT, a subcomplex of the yeast cleavage and polyadenylation factor involved in the formation of mRNA and small nucleolar RNA 3'-ends. *The Journal of biological chemistry*. 2003; 278:33000–33010. [PubMed: 12819204]
161. Ghazy MA, He X, Singh BN, Hampsey M, Moore C. The essential N terminus of the Pta1 scaffold protein is required for snoRNA transcription termination and Ssu72 function but is dispensable for pre-mRNA 3'-end processing. *Molecular and cellular biology*. 2009; 29:2296–2307. [PubMed: 19188448]
162. Hsu PL, Yang F, Smith-Kinnaman W, Yang W, Song JE, Mosley AL, Varani G. Rtr1 is a dual specificity phosphatase that dephosphorylates Tyr1 and Ser5 on the RNA polymerase II CTD. *Journal of molecular biology*. 2014; 426:2970–2981. [PubMed: 24951832]
163. Egloff S, Zaborowska J, Laitem C, Kiss T, Murphy S. Ser7 phosphorylation of the CTD recruits the RPA2 Ser5 phosphatase to snRNA genes. *Molecular cell*. 2012; 45:111–122. [PubMed: 22137580]
164. Gibney PA, Fries T, Bailer SM, Morano KA. Rtr1 is the *Saccharomyces cerevisiae* homolog of a novel family of RNA polymerase II-binding proteins. *Eukaryotic cell*. 2008; 7:938–948. [PubMed: 18408053]
165. Xiang K, Manley JL, Tong L. The yeast regulator of transcription protein Rtr1 lacks an active site and phosphatase activity. *Nature communications*. 2012; 3:946.
166. Smith-Kinnaman WR, Berna MJ, Hunter GO, True JD, Hsu P, Cabello GI, Fox MJ, Varani G, Mosley AL. The interactome of the atypical phosphatase Rtr1 in *Saccharomyces cerevisiae*. *Molecular bioSystems*. 2014; 10:1730–1741. [PubMed: 24671508]
167. Ni Z, Xu C, Guo X, Hunter GO, Kuznetsova OV, Tempel W, Marcon E, Zhong G, Guo H, Kuo WH, Li J, Young P, Olsen JB, Wan C, Loppnau P, El Bakkouri M, Senisterra GA, He H, Huang H, Sidhu SS, Emili A, Murphy S, Mosley AL, Arrowsmith CH, Min J, Greenblatt JF. RPRD1A and RPRD1B are human RNA polymerase II C-terminal domain scaffolds for Ser5 dephosphorylation. *Nature structural & molecular biology*. 2014; 21:686–695.
168. Rosado-Lugo JD, Hampsey M. The Ssu72 phosphatase mediates the RNA polymerase II initiation-elongation transition. *The Journal of biological chemistry*. 2014; 289:33916–33926. [PubMed: 25339178]
169. Visintin R, Craig K, Hwang ES, Prinz S, Tyers M, Amon A. The phosphatase Cdc14 triggers mitotic exit by reversal of Cdk-dependent phosphorylation. *Molecular cell*. 1998; 2:709–718. [PubMed: 9885559]
170. Shou W, Seol JH, Shevchenko A, Baskerville C, Moazed D, Chen ZW, Jang J, Shevchenko A, Charbonneau H, Deshaies RJ. Exit from mitosis is triggered by Tem1-dependent release of the protein phosphatase Cdc14 from nucleolar RENT complex. *Cell*. 1999; 97:233–244. [PubMed: 10219244]
171. Visintin R, Hwang ES, Amon A. Cfi1 prevents premature exit from mitosis by anchoring Cdc14 phosphatase in the nucleolus. *Nature*. 1999; 398:818–823. [PubMed: 10235265]
172. Nedeia E, Nalbant D, Xia D, Theoharis NT, Suter B, Richardson CJ, Tatchell K, Kislinger T, Greenblatt JF, Nagy PL. The Glc7 phosphatase subunit of the cleavage and polyadenylation factor is essential for transcription termination on snoRNA genes. *Molecular cell*. 2008; 29:577–587. [PubMed: 18342605]

173. Gilbert W, Guthrie C. The Glc7p nuclear phosphatase promotes mRNA export by facilitating association of Mex67p with mRNA. *Molecular cell*. 2004; 13:201–212. [PubMed: 14759366]
174. Liou YC, Zhou XZ, Lu KP. Prolyl isomerase Pin1 as a molecular switch to determine the fate of phosphoproteins. *Trends in biochemical sciences*. 2011; 36:501–514. [PubMed: 21852138]
175. Brandts JF, Halvorson HR, Brennan M. Consideration of the Possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues. *Biochemistry*. 1975; 14:4953–4963. [PubMed: 241393]
176. Brandl CJ, Deber CM. Hypothesis about the function of membrane-buried proline residues in transport proteins. *Proceedings of the National Academy of Sciences of the United States of America*. 1986; 83:917–921. [PubMed: 3456574]
177. Wedemeyer WJ, Welker E, Scheraga HA. Proline cis-trans isomerization and protein folding. *Biochemistry*. 2002; 41:14637–14644. [PubMed: 12475212]
178. Fischer G, Bang H, Mech C. [Determination of enzymatic catalysis for the cis-trans isomerization of peptide binding in proline-containing peptides]. *Biomedica biochimica acta*. 1984; 43:1101–1111. [PubMed: 6395866]
179. Hanes SD. The Ess1 prolyl isomerase: traffic cop of the RNA polymerase II transcription cycle. *Biochimica et biophysica acta*. 2014; 1839:316–333. [PubMed: 24530645]
180. Lu KP, Finn G, Lee TH, Nicholson LK. Prolyl cis-trans isomerization as a molecular timer. *Nature chemical biology*. 2007; 3:619–629. [PubMed: 17876319]
181. Singh N, Ma Z, Gemmill T, Wu X, Defiglio H, Rossetini A, Rabeler C, Beane O, Morse RH, Palumbo MJ, Hanes SD. The Ess1 prolyl isomerase is required for transcription termination of small noncoding RNAs via the Nrd1 pathway. *Molecular cell*. 2009; 36:255–266. [PubMed: 19854134]
182. Mayfield JE, Fan S, Wei S, Zhang M, Li B, Ellington AD, Etzkorn FA, Zhang YJ. Chemical Tools To Decipher Regulation of Phosphatases by Proline Isomerization on Eukaryotic RNA Polymerase II. *ACS chemical biology*. 2015
183. Zhang M, Wang XJ, Chen X, Bowman ME, Luo Y, Noel JP, Ellington AD, Etzkorn FA, Zhang Y. Structural and kinetic analysis of prolyl-isomerization/phosphorylation crosstalk in the CTD code. *ACS chemical biology*. 2012; 7:1462–1470. [PubMed: 22670809]
184. Czudnochowski N, Bosken CA, Geyer M. Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. *Nature communications*. 2012; 3:842.
185. Akhtar MS, Heidemann M, Tietjen JR, Zhang DW, Chapman RD, Eick D, Ansari AZ. TFIIF kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Molecular cell*. 2009; 34:387–393. [PubMed: 19450536]
186. Ni Z, Schwartz BE, Werner J, Suarez JR, Lis JT. Coordination of transcription, RNA processing, and surveillance by P-TEFb kinase on heat shock genes. *Molecular cell*. 2004; 13:55–65. [PubMed: 14731394]
187. Zhang S, Chen L, Luo Y, Gunawan A, Lawrence DS, Zhang ZY. Acquisition of a potent and selective TC-PTP inhibitor via a stepwise fluorophore-tagged combinatorial synthesis and screening strategy. *J Am Chem Soc*. 2009; 131:13072–13079. [PubMed: 19737019]
188. He Y, Guo X, Yu ZH, Wu L, Gunawan AM, Zhang Y, Dixon JE, Zhang ZY. A potent and selective inhibitor for the UBLCP1 proteasome phosphatase. *Bioorganic & medicinal chemistry*. 2015; 23:2798–2809. [PubMed: 25907364]

Box 1. Major events in reversible eukaryotic protein phosphorylation

Year	Event
1883	Casein is identified as the first phosphoprotein [19].
1907	Suzuki characterizes phytase, an enzyme that converts phytic acid to inositol and inorganic phosphate. This protein is generally accepted as the founding member of modern phosphatases [20, 21].
1923	Robinson discovers the earliest member of modern alkaline phosphatases. This enzyme has a pH optimum at the alkaline side of neutral. These phosphatases are now known to dephosphorylate a broad range of substrates including proteins, nucleic acids, and small molecules [21-23].
1932	Serinephosphoric acid is obtained from vitellin, providing a physical mechanism for phosphate's association with proteins. [24].
1943	Prosthetic group removing (PR) enzyme is found to convert phosphorylase A to phosphorylase B. Although unrecognized at the time, this was the first example of protein regulation by reversible phosphorylation and the PR enzyme was the first protein serine/threonine phosphatase characterized [25, 26].
1952	Phosphothreonine is obtained from casein [27].
1954	Burnett & Kennedy detect protein kinase activity for the first time against casein [28].
1955	Fischer & Krebs show that phosphorylase B can be converted to phosphorylase A in cell free extracts. This conversion was dependent on certain divalent metal ions and, in some conditions, ATP [29].
1955	Sutherland & Wosilait show that enzymatic inactivation of phosphorylase A is accompanied by phosphate release from phosphorylase A into solution [30].
1956	Krebs & Fischer purify "converting enzyme" from rabbit muscle extract and use it in an isolated system to convert phosphorylase B into phosphorylase A in the presence of radioactive ATP. The generated phosphorylase A is firmly bound to isotopic phosphate. The "converting enzyme" is later identified as Phosphorylase Kinase. [31].
1958	Krebs, Kent, & Fischer describe the stoichiometry of the kinase reaction converting phosphorylase B to phosphorylase A [32].
1968	Walsh, Perkin's, & Krebs identify the first protein kinase cascade. cAMP-dependent protein kinase (PKA) activates kinase by phosphorylation [33].
1975	Cohen, Watson, & Dixon demonstrate PKA phosphorylates phosphorylase kinase at specific amino acid sequences. Only 2 of 200 serines present in phosphorylase kinase are subject to phosphorylation by PKA, demonstrating a level of intrinsic specificity [34].
1975	Daile, Carnegie, & Young phosphorylate a synthetic peptide based on the basic protein of human myelin. This was the first demonstration of kinase activity against a synthetic peptide [35]. This study in combination with two subsequent papers[36, 37] using synthetic peptides to probe kinase specificity represent a major technical advance in the study of protein phosphorylation.
1979	Tyrosine phosphorylation is identified. v-Src is a protein tyrosine kinase and the first retroviral oncogene [38-40].
1988	Protein tyrosine phosphatases are purified to homogeneity for the first time and characterized [41].
1988	Sequence analysis of known protein kinases reveals a conserved catalytic domain [42].
1991	The first protein kinase crystal structure is determined [43].
1994	The first protein tyrosine phosphatase crystal structure is determined [44].
1995	The first protein serine/threonine phosphatase crystal structure is determined [45].

Figure 1.

Phosphorylation of CTD residues through transcription in *S. cerevisiae*. The y-axis indicates approximate ChIP-enrichments found in genomic tiling arrays in *S. cerevisiae*. The x-axis indicates approximate stage of transcription. (TSS = transcription start site. TTS = transcription termination site) [13, 62].

Figure 2.

CTD phosphatase mechanisms. A) PTP family phosphatase mechanism. PTP phosphatases utilize a cysteine nucleophile to generate a cysteinyl-phosphate intermediate. This intermediate is then resolved utilizing active site water. This mechanism is utilized by CTD phosphatases Ssu72 (Ssu72 in yellow, Symplekin in blue, *D. melanogaster*, PDB ID: 4IMJ) and Cdc14b (cyan, human, PDB ID: 1OHC). B) PPP/PPM family phosphatase mechanism. PPP/PPM phosphatases utilize a di-metal mechanism to coordinate and activate a water molecule that acts as nucleophile and liberates phosphate from substrate. This mechanism is

utilized by Glc7 (salmon, human PP1 (86% identity), PDB ID: 4MOV). C) Fcp/Scp family phosphatase mechanism. Fcp/Scp phosphatases (Fcp1 orange and blue, *S. pombe*, PDB ID: 3EF0; Scp1 in purple, human, PDB ID: 2GHT) utilize a conserved DXDX(T/V) motif and magnesium ion to coordinate phosphorylated serine. Aspartic acid acts as a nucleophile and forms an aspartyl-phosphate intermediate that is subsequently resolved by active site water.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 3.

Structure and organization of Fcp1 from *Saccharomyces pombe*. A) Domain diagram of Fcp1 with FCPH domain (orange), BRCT domain (blue), and TFIIF interacting region (green). The Fcp/Scp signature motif is labeled with the nucleophile Asp170 highlighted (red). B) Structural view of Fcp1 with domains colored as in panel A and active site labeled (PDB ID: 3EFO). The TFIIF interacting domain was not present in the solution of Fcp1. C) View of the Fcp1 active site with binding pocket residues, catalytic magnesium (pink), and AlF₃ mimicking the target phosphate (PDB ID: 3EFO).

Figure 4.

Structure of human Scp1 and models of transcriptional regulation of the Scp family. A) Structure of Scp1 (purple) bound to a phosphorylated Ser5 (P.Ser5) CTD peptide (white) with active site residues (blue) and unique hydrophobic pocket (yellow) highlighted (PDB ID: 2GHT). The hydrophobic pocket interacts with Thr4 and Pro3 of the CTD peptide. B) Two predicted models of Scp mediated regulation of transcription: Scps are recruited to RE-1 elements by REST (orange) to block transcription initiation through dephosphorylation of Ser5 of the RNA Pol II CTD (green) or promote REST stability through dephosphorylation of residues S861/864. Neuronal genes (blue), phosphates (P, yellow), ubiquitin (Ub, grey), and the co-repressor of REST (CoREST, pink) have been highlighted. C) Predicted model of PTB and miRNA mediated regulation of SCP transcription (not drawn to scale). In non-neuronal cells, PTBP1 (green) binds the SCP 3' UTR promoting expression of Scp. During neuronal differentiation, mature miR-124 and SCP intron encoded miR-26a/b (orange) are produced and bind the 3' UTR to repress SCP expression. The red lines represent competitive binding of PTBP1 and miRNAs. D) Structure of the inhibitor rabeprazole (green) bound at the hydrophobic pocket consisting of residues Phe106, Val118, Ile120, Val127, Leu155, and Tyr158 (PDB ID: 3PGL).

Figure 5.

Structures of CTD phosphatases Ssu72 and Rtr1. A) Structure of Ssu72 (pink) with phosphorylated Ser5 peptide (green) bound in the active site with catalytic residues Cys12 and Arg18 highlighted (blue) (PDB ID: 3O2Q). The Pro6 bond (green) was found in the less abundant cis-conformation. B) Structure of Ssu72 with phosphorylated Ser7 peptide bound in the active site. The orientation of the bound CTD peptide flips 180° depending on whether phosphoryl-Ser7 or Ser5 are bound at the active site (PDB ID: 4H3H). C) Structure of Ssu72 in complex with Symplekin (blue) (PDB ID: 3O2Q). D) Structure of Rtr1 from *Kluyveromyces lactis*. Chain A of the Rtr1 structure (green) lacking the disordered region between residues Asp91 and Tyr104 (PDB ID: 4M3O).

Figure 6.

Specific enhancement of Ssu72 activity by proline isomerization. A) The heptad repeat of the CTD is highly enriched with phospho-serine/proline motifs. The proline of these motifs is unique among natural amino acids in its ability to stably assume both a trans (left) and cis (right) conformation about its prolyl peptide bond. B) A chemical biology approach was used to generate phospho-Ser5 CTD peptides containing trans- (left) and cis-locked (right) proline. The prolyl-peptide bond is replaced with a carbon-carbon double bond that does not undergo thermal isomerization. These chemical tools allow for the investigation of proline

isomer preference of proteins known to recognize phospho-Ser5 in the CTD, like CTD phosphatases Ssu72 and Scp1. C) The human Scp1 (Figure C, purple and white, PDB ID: 2GHT) and D) *D. melanogaster* Ssu72 (Figure D, yellow and white, PDB ID: 4IMJ) structures provide clues to proline isomer preference. Pro6 was found in the trans conformation in the Scp1/pSer5 peptide complex structures suggesting either trans proline preference or no inherent preference. Pro6 was resolved in the cis conformation in Ssu72/pSer5 peptide complex structures suggesting a cis-proline preference. Isomerized bonds are highlighted with a dashed red line. E) Trans-specific or preferred phosphatases like Scp1 and Fcp1 (Scp1 in purple, human, PDB ID: 2GHT; Fcp1 orange and blue, *S. pombe*, PDB ID: 3EF0) bypass regulation by proline isomerization state due to their preference for the major proline species and have equivalent apparent activity with or without prolyl isomerases. F) *In vitro* and cellular experiments demonstrate Ssu72 (Ssu72 in yellow, Symplekin in blue, *D. melanogaster*, PDB ID: 4IMJ) activity is enhanced by the presence of phospho-specific prolyl isomerase Pin1 (green, human, PDB ID: 2Q5A) due to its structural requirement for cis-proline. The activity of prolyl isomerases boosts the apparent dephosphorylation of CTD by Ssu72, while thermal isomerization of prolines in CTD becomes rate limiting and leads to an accumulation of phospho-Ser5 CTD in the cell.