



Published in final edited form as:

ACS Chem Biol. 2015 October 16; 10(10): 2405–2414. doi:10.1021/acscchembio.5b00296.

## Chemical Tools To Decipher Regulation of Phosphatases by Proline Isomerization on Eukaryotic RNA Polymerase II

Joshua E. Mayfield<sup>†</sup>, Shuang Fan<sup>‡</sup>, Shuo Wei<sup>§,||</sup>, Mengmeng Zhang<sup>†,⊥</sup>, Bing Li<sup>#</sup>, Andrew D. Ellington<sup>†</sup>, Felicia A. Etzkorn<sup>‡</sup>, and Yan Jessie Zhang<sup>\*,†</sup>

<sup>†</sup>Department of Molecular Biosciences and Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, Texas 78712, United States

<sup>‡</sup>Department of Chemistry, Virginia Tech, Blacksburg, Virginia 24061, United States

<sup>§</sup>1 Cancer Research Institute, Beth Israel Deaconess Cancer Center, Harvard Medical School, Boston, Massachusetts 02215, United States

<sup>||</sup>Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts 02215, United States

<sup>#</sup>Department of Molecular Biology, UT Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas 75390, United States

### Abstract

Proline isomerization greatly impacts biological signaling but is subtle and difficult to detect in proteins. We characterize this poorly understood regulatory mechanism for RNA polymerase II carboxyl terminal domain (CTD) phosphorylation state using novel, direct, and quantitative chemical tools. We determine the proline isomeric preference of three CTD phosphatases: Ssu72 as *cis*-proline specific, Scp1 and Fcp1 as strongly *trans*-preferred. Due to this inherent characteristic, these phosphatases respond differently to enzymes that catalyze the isomerization of proline, like Ess1/Pin1. We demonstrate that this selective regulation of RNA polymerase II phosphorylation state exists within human cells, consistent with *in vitro* assays. These results support a model in which, instead of a global enhancement of downstream enzymatic activities, proline isomerases selectively boost the activity of a subset of CTD regulatory factors specific for *cis*-proline. This leads to diversified phosphorylation states of CTD *in vitro* and in cells. We

\*Corresponding Author: Yan Jessie Zhang. jzhang@cm.utexas.edu. Phone: 512-471-8645.

⊥Present Address

Wyss Institute for Biologically Inspired Engineering at Harvard University, 3 Blackfan Circle, Boston, Massachusetts 02115, United States.

### Supporting Information

This material is available free of charge via the Internet. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acscchembio.5b00296.

Synthetic peptide sequences, reaction conditions for CTD phosphatases, complete methods, NMR spectra, HRMS ESI, and HPLC results for 1–6 (PDF)

### Accession Codes

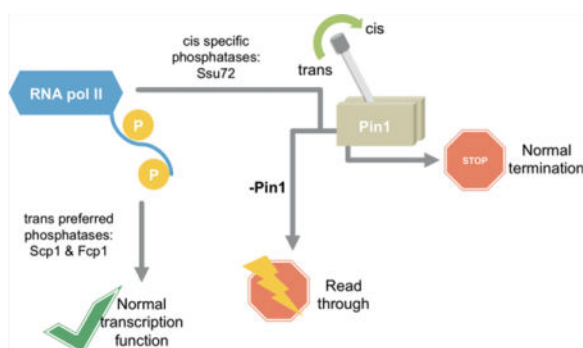
Structure factors and coordinates reported in this paper are deposited in Protein Data Bank. Crystal Structure of *D. melanogaster* Ssu72 + Symplekin bound to *cis* peptidomimetic CTD phospho-Ser5 peptide is assigned accession code 4YGX. Complex structures of Scp1 with *cis* and *trans* peptidomimetic CTD peptides are assigned as 4YH1 and 4YGY, respectively.

### Notes

The authors declare no competing financial interest.

provide the chemical tools to investigate proline isomerization and its ability to selectively enhance signaling in transcription and other biological contexts.

## Graphical abstract



The carboxyl terminal domain (CTD) of RNA polymerase II coordinates the initiation, elongation, and termination of transcription, as well as coprocessing of nascent mRNA to maturation.<sup>1</sup> This diverse set of functions is coordinated through multiple post-translational modifications (PTM) of the CTD.<sup>2-4</sup> The CTD is composed of the consensus heptad YSPTSPS, which repeats 26–52 times depending on the organism.<sup>2</sup> Various PTMs of this heptad recruit different regulatory factors to active RNA polymerase II allowing for effective transcription. The post-translational modification states of CTD fluctuate throughout transcription as the result of interplay between modifying enzymes like kinases, phosphatases, and prolyl isomerases.<sup>2,3,5</sup>

The most prominent modifications of the CTD are phosphorylation of Ser2 and Ser5 during each round of transcription.<sup>1</sup> These phosphorylation states are temporally regulated and are correlated to different stages of the transcription cycle.<sup>3</sup> Ser5 phosphorylation is dominant at the early stages of transcription but gradually decreases as RNA polymerase II proceeds into elongation.<sup>1</sup> Ser2 phosphorylation becomes dominant during elongation and termination.<sup>1</sup> All phosphorylations are removed at the termination of transcription, allowing RNA polymerase II to recycle and bind promoters for subsequent rounds of transcription.<sup>1</sup>

Proline residues, Pro3 and Pro6, flank these major phosphorylation sites. When Ser2 and Ser5 are phosphorylated, the isomerization states of Pro3 and Pro6 are equilibrated by phospho-specific prolyl isomerases Ess1 in yeast and Pin1 in humans.<sup>6,7</sup> Ess1/Pin1 has been identified to play an important role in transcription regulation.<sup>8-11</sup> Ess1 mutations are synthetic lethal with truncated CTD alleles, linking it to CTD-mediated transcription.<sup>10</sup> In human cells, Pin1 isomerase activity can impact RNA polymerase II phosphorylation state and alter RNA polymerase II localization.<sup>11</sup> Furthermore, defective transcription termination phenotypes are associated with compromised prolyl isomerase activity in yeast.<sup>8</sup>

While these data suggest that Ess1/Pin1 impacts transcription by altering the proline isomerization state of the CTD and in turn its phosphorylation, this is difficult to prove mechanistically. One hypothesis is that proline isomerization state impacts the relative activities of modification enzymes. Three CTD phosphatases have been well characterized:

Ssu72, Scp1, and Fcp1.<sup>12–15</sup> Both Ssu72 and Scp1 dephosphorylate Ser5 of CTD, but they lead to different transcriptional outcomes. Ssu72 plays a pivotal role in general transcription elongation, 3'-end processing, and termination.<sup>13,16,17</sup> Scp1 is a component of the RE1-silencing transcription factor (REST) complex and is found only in higher eukaryotes. REST complex prevents the transcription of a subset of neuronal genes.<sup>14</sup> Fcp1, the only Ser2 phosphatase characterized to date, is essential for the recycling of RNA polymerase II.<sup>18,19</sup>

Little is known about the use of proline isomerization as a regulatory mechanism during transcription.<sup>7</sup> Unique among the natural amino acids, the peptide bond of proline can stably assume a *cis* or *trans* isomer conformation, and isomer interconversion occurs naturally. The *trans* form is preferred and occurs 70–90% of the time.<sup>20</sup> Prolines located in fully folded proteins can assume the *cis* or *trans* form exclusively,<sup>21,22</sup> but context-specific equilibration and the activity of prolyl isomerases, like Ess1/Pin1, can greatly increase the conversion rate.<sup>6</sup> Proline isomerization is not evident in sequence or molecular weight analysis, and conversion between the two isomers is challenging to monitor in cells.

To surmount these limitations, we developed “locked-proline” analogues that mimic proline but cannot undergo isomer conversion.<sup>23–27</sup> By incorporating locked-proline analogues in place of proline residues, we can differentiate their subtle regulatory effect on protein modification.<sup>23</sup> In this study, we designed peptidomimetic compounds to characterize the prolyl isomeric requirement for substrates of CTD phosphatases and demonstrate that Pin1 up-regulates only *cis*-specific phosphatases. Using yeast GST-CTD as substrate, we show that Pin1 isomerase activity promotes dephosphorylation by Ssu72 in the context of full length CTD. We translate our *in vitro* observations to a cellular system by investigating the accumulation of CTD phosphorylation marks with and without Pin1 activity in HeLa cells. On the basis of these results, we propose a model of divergent Pin1 regulation on CTD phosphatases and identify Pin1 as a kinetic switch that helps determine effective and accurate transcription.

## RESULTS AND DISCUSSION

### Synthetic CTD Peptidomimetic Analogues Incorporating *cis*- and *trans*-Locked Isosteres

*In vitro* proteomic analysis revealed that more than 100 yeast proteins are found associated directly or indirectly with CTD,<sup>28</sup> and the number is suspected to be higher in mammals. Although it is suggested that many of these proteins are recruited to CTD by binding phospho-Ser-Pro motifs, the configuration of proline is established for very few.<sup>27,29–32</sup> Since proline isomerization results in subtle changes in conformation, not in sequence or molecular weight, detection of isomer-specific binding is difficult.

We use a chemical biology approach to determine whether a protein is selective toward a given proline isomeric state in of the proline is substituted with a carbon atom and the prolyl peptide bond is replaced with a carbon-carbon double bond to prevent thermal isomerization. These locked isosteres are good mimics for proline residues in different isomeric states, and Pin1 recognizes both forms.<sup>26,27</sup> Furthermore, recent studies using peptides with alkene isosteres in place of SP motifs show that the *trans* conformation is phosphorylated by Cdk1-cyclin B kinase.<sup>23</sup>

The crystal structures of Scp1 and Ssu72 in complex with phosphorylated-Ser5 CTD peptides are solved.<sup>15,30,31,33</sup> Although these phosphatases both recognize phosphorylated Ser5 in CTD, the Pro6 isomerization states captured at their active sites differ. Scp1 binds to CTD peptides when Pro6 is in the *trans* conformation,<sup>15</sup> whereas Pro6 is in the *cis* conformation when bound to Ssu72.<sup>30,31,33</sup> Since X-ray crystallography only captures the species that is favorable for crystallization, we want to establish selectivity for proline isomers by CTD phosphatases in solution using peptides incorporating *cis*- or *trans*-locked proline analogues. Two native peptides and four different CTD peptidomimetic compounds were synthesized as 11-mer or 12-mer repeats with Pro6 or Pro3 following phosphorylated Ser5 or Ser2 replaced by *cis*- or *trans*-locked proline analogues (Supplementary Table 1).

### Ssu72 Is a *cis*-Specific CTD Ser5 Phosphatase

Ssu72 is a conserved eukaryotic Ser5 phosphatase that is important to transcription termination and mRNA coprocessing.<sup>13,16,17,34,35</sup> The phosphatase activity of *Drosophila* Ssu72–Symplekin was tested against synthetic CTD peptides with Pro6 replaced by a *cis*- or *trans*-locked moiety. Ssu72–Symplekin shows robust phosphatase activity against *cis*-locked peptide ( $k_{\text{cat}}/K_m$  of  $5.24 \pm 0.08 \text{ mM}^{-1} \text{ s}^{-1}$ ; Figure 2A), but no phosphatase activity was detected against the *trans*-locked peptide (Figure 2A). As a control experiment, we used a synthetic peptide with the same sequence incorporating native pSer–Pro. The activity of Ssu72 against the natural peptide is substantially lower than that against *cis*-locked peptidomimetic (Figure 2A). This is because the effective *cis*-substrate for Ssu72 constitutes only a small portion of the natural peptide pool, estimated at 10–30%.<sup>20</sup> These quantitative measurements establish Ssu72 as a *cis*-specific phosphatase for Ser5–Pro6 motifs of CTD.

The crystal structure of Ssu72–Symplekin bound to the *cis*-locked peptide provides a structural explanation for proline isomer specificity. The complex structure was solved to 2.95 Å using a catalytically inactive variant (C13D/D144N) to capture the mode of isostere peptide binding (Table 1). Strong electron density was observed at the active site of Ssu72 with 6 of 11 substrate residues observed (Figure 2B). Importantly, the *cis*-locked proline moiety is bound into a small hydrophobic pocket formed by Met85, Leu45, Pro46, and the side chain of Met17. This pocket is approximately 5 Å deep, allowing it to snugly accommodate the proline side chain (Figure 2C). The *cis*-proline isostere binds tightly to this deep and narrow pocket, whereas the alternative *trans*-proline configuration clashes sterically. Alignment with native peptide complex structure<sup>33</sup> reveals conserved geometry and nearly identical substrate positioning (Figure 2D), with the *cis*-locked peptide's carbon–carbon double bond about the same length as a peptide bond (both around 1.33 Å). Since the peptide bond between Ser5 and Pro6 does not form hydrogen bonds with Ssu72 active site residues, replacement of the amide bond by an alkene minimally disturbs hydrogen bonding and allows the *cis*-locked peptide to act as optimal Ssu72 substrate. These observations coupled with our kinetic analysis suggest that Ssu72 has a strict selectivity for *cis*-Pro6 of CTD.

### Scps Strongly Favor *trans*-Proline as Substrate

As a component of master silencing complex REST, Scp phosphatases (Scp1–3) are CTD Ser5 phosphatases whose activity is implicated in repression of neuronal genes.<sup>14</sup> Since the

three isoforms of Scps have identical catalytic activity and all catalytic residues conserved, we used the best-characterized protein of the Scps, Scp1, to study their prolyl isomeric selectivity.<sup>15</sup> Both *cis*- and *trans*-locked CTD peptidomimetics were used as substrates in a phosphatase assay for human Scp1. Different from Ssu72, Scp1 shows robust phosphatase activity against both locked peptides (Figure 3A). However, Scp1 presents a significant preference toward *trans*-locked and native CTD peptides ( $k_{cat}/K_m = 323 \pm 22$  and  $386 \pm 21 \text{ mM}^{-1} \text{ s}^{-1}$ , respectively). Scp1 can also recognize the *cis*-locked peptide as substrate, although there is a 8.5-fold reduction in activity ( $k_{cat}/K_m = 38.0 \pm 1.3 \text{ mM}^{-1} \text{ s}^{-1}$ ; Figure 3A). This explains why only the *trans* form of proline is found in X-ray crystal structures of Scp1 bound to native CTD peptide. Because of the averaging effect of X-ray crystallography experiments, they present an averaged structure in which the *cis*-proline signal would contribute very little to the final electron density due to its low population and weaker affinity. For the first time, our chemical tools allow for discrete and quantitative partitioning of proline isomer substrate preference.

We obtained crystal structures of Scp1 bound with each locked isostere compound using an inactive Scp1 variant (D96N) in which the nucleophile aspartate was mutated to asparagine to prevent product turnover. The Scp1 + *trans*-locked peptide structure was solved to 2.36 Å, and the Scp1 + *cis*-locked peptide structure was solved to 2.20 Å (Table 1). In both structures, five of the 11 synthetic peptide residues can be visualized at the active site (Figure 3B). The Scp1 structure containing the *cis*-locked peptide reveals nearly identical conformation for most residues with the only significant difference being the change of configuration of proline analogues (Figure 3C). The locked isosteres, which are flipped 180° in the two structures, locate at the edge of the active site binding pocket and provide a structural explanation for Scp1's less stringent proline isomer requirement. Unlike Ssu72, in which Pro6 extends into a deep and narrow pocket, Scp1 binds solvent exposed Pro6 at the rim of the active site pocket (Figure 3D,E). The openness of the Scp1 active site for Pro6 binding explains the more promiscuous nature of Scp1 prolyl selectivity since both isomers can be accommodated. These complex structures are consistent with our kinetic results showing that both the *cis* and *trans* configurations at Pro6 serve as substrates for Scp1. Our chemical tools not only provide insight into the kinetic impact of proline isomerization state but also help explain this functional data in terms of protein structure. By visualizing proteins of interest bound to each proline isoform, we can better understand the substrate promiscuity inherent to some protein active sites and develop a physical explanation for this variability.

### Prolyl Isomeric Selectivity of CTD Ser2 Phosphatase Fcp1

Fcp1 is the only Ser2 phosphatase reported and its activity is essential for the recycling of RNA polymerase II.<sup>19,36</sup> The effect of proline isomerization of CTD on Fcp1 has been debated since Fcp1 phosphatase activity was first characterized.<sup>11,37</sup> Unfortunately, even with high concentrations of CTD peptides included in crystallization conditions, no peptide was resolved in the active site of Fcp1.<sup>12</sup> To identify the prolyl selectivity of Fcp1, we designed Fcp1 substrates with our locked-proline isosteres at the Ser2–Pro3 position. We synthesized 12-mer synthetic CTD peptides with Pro3 replaced by *cis*- or *trans*-locked isosteres. These two peptidomimetic compounds were used as substrates for Fcp1 in

phosphatase assays. Fcp1 shows activity against the *trans*-locked isostere compound at levels comparable to the activity observed against the native peptide (Figure 4A). Lower activity was observed against *cis*-locked peptide (Figure 4A). Therefore, our results indicate that Fcp1 prefers *trans*-proline next to the serine subject to dephosphorylation but can also accept *cis*-proline containing substrate. This proline isomeric preference is identical to Scp1, though their mode of CTD binding is expected to be different.<sup>12</sup>

### Impact of Pin1 Isomerase Activity on Fcp1 Mediated Dephosphorylation

Since CTD phosphatases have different preferences for proline configuration within their recognition phospho-Ser-Pro motifs, enzymes that catalyze proline isomerization could affect downstream phosphatase function. Indeed, it has been shown that Ess1/Pin1 isomerase activity can alter the transcription profile in yeast<sup>8,9,38</sup> and humans,<sup>11</sup> but downstream effectors for such regulation on transcription are not well established.<sup>6</sup> Pin1 shows strong affinity to the two SP islands in the CTD consensus sequence with a  $K_d$  of 30  $\mu$ M for phospho-Ser5-Pro6 and a  $K_d$  of 61  $\mu$ M for phospho-Ser2-Pro3 in peptides of about a single repeat in length.<sup>39</sup> Since Pin1 recognizes the same motif as CTD phosphatases, we hypothesize that the enzymatic activity of Pin1 can alter downstream phosphatase activity leading to changes in transcription pattern. This hypothesis is consistent with our previous observation that there are dramatic differences between Ssu72 and Scp1 activities when Pin1 is present: Ssu72 showed a 3–4-fold increase in phosphatase activity upon Pin1 supplementation, while Scp1 activity was unaffected.<sup>27</sup> This enhancement is specific to Pin1 isomerase activity and binding CTD, since mutation to prevent recognition of CTD by Pin1 eliminates this effect.<sup>27</sup>

To see how Pin1 activity affects dephosphorylation of Ser2 in CTD by Fcp1, we measured the phosphatase activity of Fcp1 in the presence and absence of Pin1 (Figure 4B). Previous data suggest contradictory roles of Pin1 on Fcp1 phosphatase activity, showing both stimulatory and inhibitory effects.<sup>11,37,40</sup> We believe confusion stems from several factors: (1) the identity and concentration of the *de facto* substrate was not determined, (2) nonphysiological kinases were utilized to generate CTD substrate as a mixture of phosphorylated species, (3) Pin1 concentrations utilized in the *in vitro* assays were sometimes quite high and may have competed with Fcp1 for binding CTD substrate, and finally (4) overexpression of Pin1 *in vivo* can affect many other human Pin1 substrates. To overcome these limitations, we monitored the phosphatase activity of Fcp1 with or without Pin1 against saturating amounts of synthetic CTD peptide phosphorylated at Ser2. Additionally, by performing a control reaction with the truncated and mutated PPIase K77/82Q domain of Pin1, which is incapable of binding and isomerizing the CTD, we can determine whether the observed effects on phosphatase activity are the result of Pin1's prolyl isomerase activity against CTD. Our quantitative assay reveals Fcp1 phosphatase activity against the phosphorylated 12-mer CTD peptide is unaffected by Pin1 isomerase activity (Figure 4B). This observation is consistent with our determination that Fcp1 is a *trans* preferred Ser2 phosphatase, reminiscent of the Scp1/Pin1 profile in which Pin1 isomerase activity does not affect phosphate release by Scp1.<sup>27</sup>



The observation that phosphatases like Fcp1/Scp1 that show strong preference to *trans*-proline are not affected by Pin1 activity is surprising. It has been generally assumed that since Pin1 can convert the *cis* and *trans* isomers whenever either population is below equilibrium, Pin1-mediated proline isomerization could promote the activity of any downstream proteins by replenishing substrate pools. However, the rate of *cis/trans* thermal conversion is highly dependent on the local protein configuration and can range from minutes to hundreds of hours.<sup>21,22</sup> The isomerase activity of Pin1/Ess1 would only boost the downstream enzyme activity when this thermal conversion is rate limiting and too slow to sustain the supply of substrate. Since *trans*-proline is the major natural species, we hypothesize that Ess1/Pin1 isomerase activity will not show any obvious effect until most of the substrate is depleted. Therefore, instead of a global effect of up-regulation for any downstream phosphatases recognizing SP motifs, Ess1/Pin1 only has a significant regulatory effect on proteins with a strong preference for the minor *cis*-proline species. CTD phosphatases with specificity or preference for the *trans* conformation proline can bypass Pin1 regulation with no alteration of the signaling pathway even though Pin1 performs the isomerization reaction.

### ***In Vitro* Reconstruction of Pin1 Mediated Ssu72 Activity Enhancement**

Quantitative measurement of Ssu72 activity against phosphorylated CTD peptides established that Ssu72 phosphatase activity is enhanced 3–4-fold upon Pin1 supplementation.<sup>27</sup> Since the *cis/trans* proline conversion rate is highly dependent on context, we investigated whether proline isomerization can be rate limiting in the context of full-length CTD and whether the enhancement of Ssu72 by Pin1 is still evident. To test this, we reconstructed a minimalist system *in vitro* using GST-CTD phosphorylated by physiologically relevant kinase TFIIF to enrich the substrate for phospho-Ser5 marks; Ssu72 then dephosphorylated this substrate with or without Pin1. The level of Ser5 phosphorylation was determined by Western blot using phosphorylated Ser5 specific antibodies (Figure 5A). Reactions containing Pin1 displayed a higher degree of dephosphorylation relative to the zero time point than the reactions containing Ssu72 alone at all time points (Figure 5B). These results are consistent with our kinetic results using short CTD peptides.<sup>27</sup> These data imply that Pin1 increases apparent Ssu72 activity against both short peptides and full length CTD.

### **Prolyl Isomerase Activity Regulates *cis*-Specific Phosphatases in the Cell**

Our *in vitro* observations suggest that the CTD phosphorylation state is differentially regulated by phosphatases based on their proline isomeric selectivity. To determine whether this extends to cells, we monitored the effect of proline isomerization on the phosphorylation states of RNA polymerase II in HeLa cells. In this experiment, Pin1 expression was knocked down by more than 90%,<sup>41</sup> and the phosphorylation levels of Ser2 and Ser5 were monitored relative to vector control cells. During general transcription, Fcp1 is the main phosphatase for Ser2 dephosphorylation,<sup>18</sup> whereas Ssu72 is the workhorse for Ser5 dephosphorylation.<sup>32</sup> Western blot for RNA polymerase II using a phospho-specific Ser5 antibody consistently showed the accumulation of phospho-Ser5 CTD by 30–60% in the Pin1 knockdown cells compared with an empty vector control (Figure 6A,B). Due to the lack of Pin1 activity, CTD repeats containing phospho-Ser5-*cis*-Pro6 are subject to

depletion, and Ssu72 must wait for the slow *trans* to *cis* thermal conversion to dephosphorylate the remaining substrate. This apparent reduction of Ssu72 activity due to the loss of Pin1 causes an accumulation of phospho-Ser5 CTD in the cell. Importantly, the phosphorylation level of Ser2 is not impacted by Pin1 knockdown (Figure 6A,B). This is consistent with our *in vitro* data showing that Ssu72 activity is increased by Pin1,<sup>27</sup> whereas Fcp1 is unaffected.

### Differential Regulation of CTD Phosphatases by Pin1

The isomerase activity of Ess1/Pin1 has been found to affect the outcome of eukaryotic transcription, yet its biological mechanism is not understood.<sup>6</sup> In this study, we utilized locked-proline analogues to address how subtle conformational variation in proline isomerization alters signal transduction and results in differentiated regulation. Specifically, we determined the prolyl isomeric preference of three CTD phosphatases, Ssu72, Scp1, and Fcp1. Based on the different abundances of *cis*- and *trans*-proline, we reason that the impact of prolyl isomerase activity on CTD phosphatases varies. For CTD in the absence of prolyl isomerases, *cis*-proline containing motifs become depleted. This reduced availability of substrate acts as a “kinetic trap” to hinder the dephosphorylation of RNA polymerase II by *cis*-specific CTD phosphatases. The resultant accumulation of phospho-Ser5 would lead to global transcription termination defects such as read-through.<sup>17</sup> However, Pin1-mediated *cis*–*trans* conversion overcomes this kinetic trap and provides sufficient *cis*-proline for Ssu72 consumption. Using GST-CTD, we show that the enhanced activity of Ssu72 by Pin1 extends to full-length substrate. This is consistent with the change in CTD phosphorylation states observed in cells upon Pin1 knockdown. Our kinetic, structural, and cellular results support a model in which Ess1/Pin1 alters the conformation of CTD and provides a kinetic switch that leads to differentiated phosphorylation states and results in different transcriptional outcomes (Figure 7).

Conversely, CTD phosphatases Scp1 and Fcp1 strongly favor the proline residue next to the phospho-serine to be in *trans* conformation. Our results challenge a model where Ess1/Pin1 globally enhances the activity of proteins with isomeric preferences. Instead, we show Pin1 has no effect on Scp1 and Fcp1 activity *in vitro*. We reason that these enzymes, utilizing the major *trans*-proline species, have abundant substrate pools available, and *cis* to *trans* conversion is less likely to become rate limiting until almost all substrate is depleted, which is unlikely *in vivo*. Therefore, Pin1 isomerase activity has little effect on *trans*-proline specific or preferred phosphatases. This mechanism is confirmed via Western blot against cells lacking Pin1 in which Ser2 phosphorylation, regulated by Fcp1, is not affected but Ser5 phosphorylation, regulated by Ssu72, is impacted. Therefore, even though Ess1/Pin1 has been reported to have widespread effects on transcription, the direct regulation is restricted to proteins with *cis*-specific proline selectivity (Figure 7).

Proteomic studies have reported that at least 100 proteins associate with CTD *in vitro*, mostly when it is phosphorylated.<sup>28</sup> The proline residues situated next to the phosphorylated serine are usually recognized by CTD binding proteins. With the configuration difference introduced by the isomeric states of proline, the interacting domains of CTD binding proteins could accommodate one isomer better due to steric restrictions. Proline



isomerization state can alter the suitability of CTD as substrate for binding partners or downstream CTD modification enzymes. As shown here, such selectivity plays a pivotal role in regulatory pathways controlled by Pin1 and leads to differentiated outcomes. Furthermore, the isomeric states of proline could play a crucial role for the recruitment of protein factors and the assembly of transcription complexes. Nrd1, for example, is a protein factor for small noncoding RNA termination that forms the Nrd1-Nab3-Sen1 (NNS) complex and is the only other known *cis*-specific CTD binding protein.<sup>29</sup> Therefore, prolyl isomeric selectivity might direct the assembly of different complexes to process nascent RNA polymerase II products. Since prolyl isomeric selectivity is hard to detect, the locked isostere compounds described here are useful as chemical tools to directly establish the preference of proteins toward proline isomers. As shown here for CTD biology, locked isosteres can be used to elucidate regulatory pathways involving proline-containing motifs such as MAP kinase signaling, cyclin-dependent kinase signaling, and the GSK3 $\beta$  pathway.<sup>23,42</sup>

## METHODS

Methods, peptide sequences, reaction conditions, and antibodies are provided in Supporting Information.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

### Funding

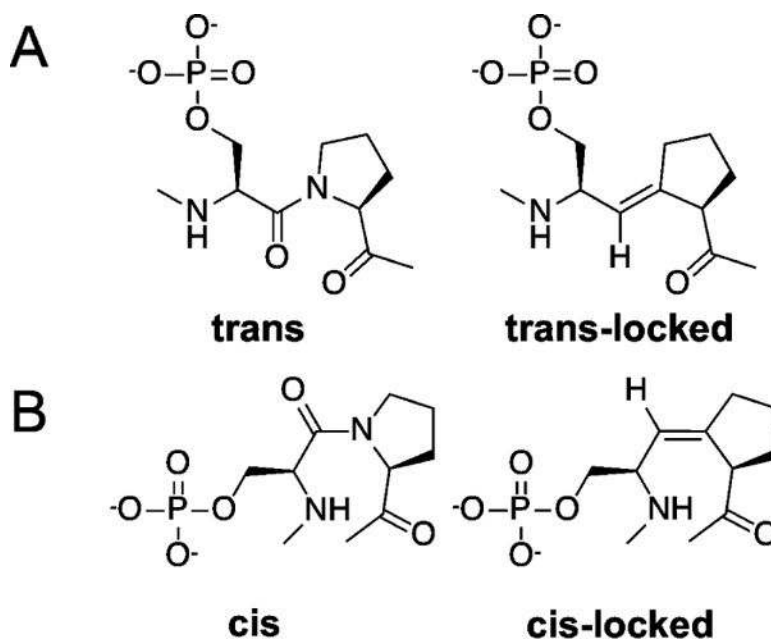
This work is supported in part by grants from the National Institutes of Health (R01 GM104896 to Y.Z., R01 GM090077 to B.L., and R01 CA110940 to F.A.E.) and Welch Foundation (F-1778 to Y.Z., I-1713 to B.L.). B.L. is a W.A. “Tex” Moncrief, Jr. Scholar in Medical Research. Crystallographic data collections were conducted at advanced light sources (Beamline 5.0.3) and advanced photon sources (BL23-ID-B), Department of Energy (DOE) national user facility.

## References

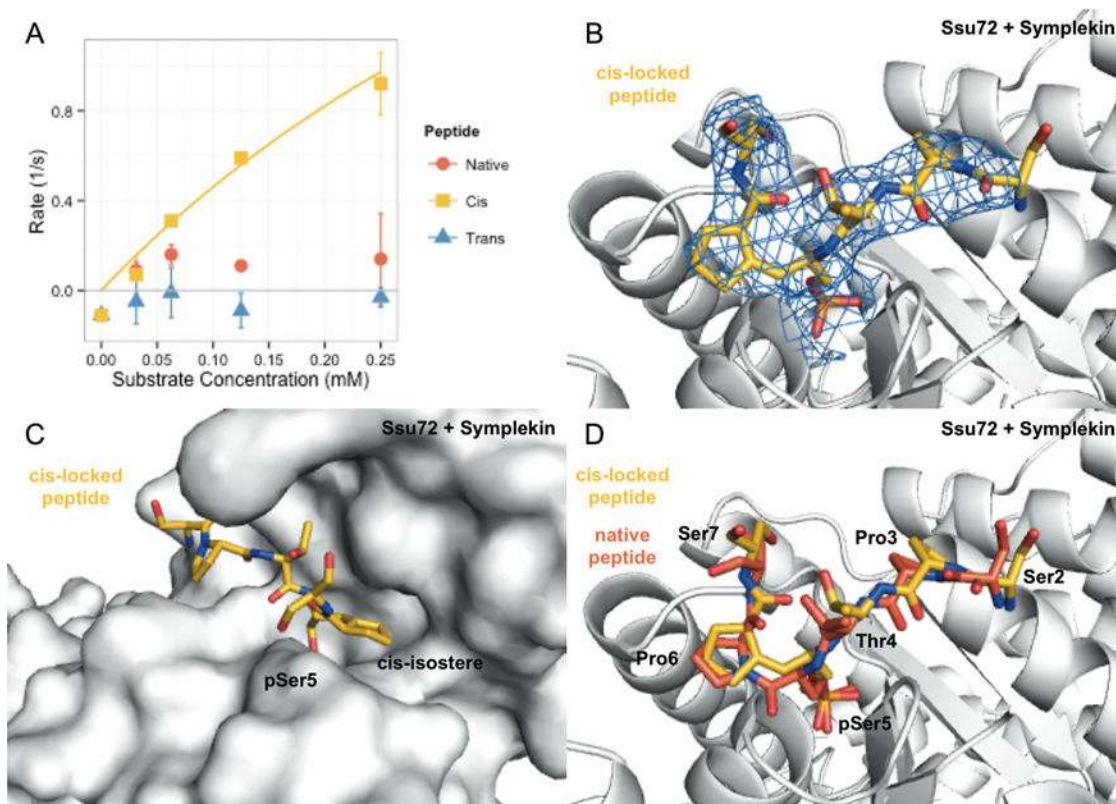
1. Eick D, Geyer M. The RNA polymerase II carboxyterminal domain (CTD) code. *Chem Rev* (Washington, DC, US). 2013; 113:8456–8490.
2. Bataille AR, Jeronimo C, Jacques PE, Laramée L, Fortin ME, Forest A, Bergeron M, Hanes SD, Robert F. A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. *Mol Cell*. 2012; 45:158–170. [PubMed: 22284676]
3. Buratowski S. Progression through the RNA polymerase II CTD cycle. *Mol Cell*. 2009; 36:541–546. [PubMed: 19941815]
4. Schroder S, Herker E, Itzen F, He D, Thomas S, Gilchrist DA, Kaehlcke K, Cho S, Pollard KS, Capra JA, Schnolzer M, Cole PA, Geyer M, Bruneau BG, Adelman K, Ott M. Acetylation of RNA polymerase II regulates growth-factor-induced gene transcription in mammalian cells. *Mol Cell*. 2013; 52:314–324. [PubMed: 24207025]
5. Conaway JW, Shilatfard A, Dvir A, Conaway RC. Control of elongation by RNA polymerase II. *Trends Biochem Sci*. 2000; 25:375–380. [PubMed: 10916156]

6. Hanes SD. The Ess1 prolyl isomerase: traffic cop of the RNA polymerase II transcription cycle. *Biochim Biophys Acta, Gene Regul Mech.* 2014; 1839:316–333.
7. Lu KP, Finn G, Lee TH, Nicholson LK. Prolyl cis-trans isomerization as a molecular timer. *Nat Chem Biol.* 2007; 3:619–629. [PubMed: 17876319]
8. 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. *Mol Cell.* 2009; 36:255–266. [PubMed: 19854134]
9. Wu X, Rossetini A, Hanes SD. The ESS1 prolyl isomerase and its suppressor BYE1 interact with RNA pol II to inhibit transcription elongation in *Saccharomyces cerevisiae*. *Genetics.* 2003; 165:1687–1702. [PubMed: 14704159]
10. Wu X, Wilcox CB, Devasahayam G, Hackett RL, Arevalo-Rodriguez M, Cardenas ME, Heitman J, Hanes SD. The Ess1 prolyl isomerase is linked to chromatin remodeling complexes and the general transcription machinery. *EMBO J.* 2000; 19:3727–3738. [PubMed: 10899126]
11. Xu YX, Hirose Y, Zhou XZ, Lu KP, Manley JL. Pin1 modulates the structure and function of human RNA polymerase II. *Genes Dev.* 2003; 17:2765–2776. [PubMed: 14600023]
12. Ghosh A, Shuman S, Lima CD. The structure of Fcp1, an essential RNA polymerase II CTD phosphatase. *Mol Cell.* 2008; 32:478–490. [PubMed: 19026779]
13. Krishnamurthy S, He X, Reyes-Reyes M, Moore C, Hampsey M. Ssu72 Is an RNA polymerase II CTD phosphatase. *Mol Cell.* 2004; 14:387–394. [PubMed: 15125841]
14. 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]
15. 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. *Mol Cell.* 2006; 24:759–770. [PubMed: 17157258]
16. 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. *Mol Cell.* 2002; 10:1139–1150. [PubMed: 12453421]
17. 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. *EMBO J.* 2003; 22:1588–1598. [PubMed: 12660165]
18. 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 Dev.* 2001; 15:3319–3329. [PubMed: 11751637]
19. Cho H, Kim TK, Mancebo H, Lane WS, Flores O, Reinberg D. A protein phosphatase functions to recycle RNA polymerase II. *Genes Dev.* 1999; 13:1540–1552. [PubMed: 10385623]
20. 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]
21. Brandl CJ, Deber CM. Hypothesis about the function of membrane-buried proline residues in transport proteins. *Proc Natl Acad Sci U S A.* 1986; 83:917–921. [PubMed: 3456574]
22. Wedemeyer WJ, Welker E, Scheraga HA. Proline cis-trans isomerization and protein folding. *Biochemistry.* 2002; 41:14637–14644. [PubMed: 12475212]
23. Etkorn FA, Zhao S. Stereospecific phosphorylation by the central mitotic kinase Cdk1-cyclin B. *ACS Chem Biol.* 2015; 10:952–956. [PubMed: 25603287]
24. Namanja AT, Wang XJ, Xu B, Mercedes-Camacho AY, Wilson BD, Wilson KA, Etkorn FA, Peng JW. Toward flexibility-activity relationships by NMR spectroscopy: dynamics of Pin1 ligands. *J Am Chem Soc.* 2010; 132:5607–5609. [PubMed: 20356313]
25. Wang XJ, Hart SA, Xu B, Mason MD, Goodell JR, Etkorn FA. Serine-cis-proline and serine-trans-proline isosteres: stereoselective synthesis of (Z)- and (E)-alkene mimics by Still-Wittig and Ireland-Claisen rearrangements. *J Org Chem.* 2003; 68:2343–2349. [PubMed: 12636401]
26. Wang XJ, Xu B, Mullins AB, Neiler FK, Etkorn FA. Conformationally locked isostere of phosphoSer-cis-Pro inhibits Pin1 23-fold better than phosphoSer-trans-Pro isostere. *J Am Chem Soc.* 2004; 126:15533–15542. [PubMed: 15563182]

27. Zhang M, Wang XJ, Chen X, Bowman ME, Luo Y, Noel JP, Ellington AD, Etkorn FA, Zhang Y. Structural and kinetic analysis of prolyl-isomerization/phosphorylation cross-talk in the CTD code. *ACS Chem Biol.* 2012; 7:1462–1470. [PubMed: 22670809]
28. Phatnani HP, Greenleaf AL. Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev.* 2006; 20:2922–2936. [PubMed: 17079683]
29. Kubicek K, Cerna H, Holub P, Pasulka J, Hrossova D, Loehr F, Hofr C, Vanacova S, Stefl R. Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. *Genes Dev.* 2012; 26:1891–1896. [PubMed: 22892239]
30. 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. *J Biol Chem.* 2011; 286:5717–5726. [PubMed: 21159777]
31. Xiang K, Nagaïke 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]
32. 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. *Biochem J.* 2011; 434:435–444. [PubMed: 21204787]
33. 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 Chem Biol.* 2013; 8:2042–2052. [PubMed: 23844594]
34. He X, Khan AU, Cheng H, Pappas DL Jr, Hampsey M, Moore CL. Functional interactions between the transcription and mRNA 3' end processing machineries mediated by Ssu72 and Sub1. *Genes Dev.* 2003; 17:1030–1042. [PubMed: 12704082]
35. Reyes-Reyes M, Hampsey M. Role for the Ssu72 C-terminal domain phosphatase in RNA polymerase II transcription elongation. *Mol Cell Biol.* 2007; 27:926–936. [PubMed: 17101794]
36. 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*. *Mol Cell.* 1999; 4:55–62. [PubMed: 10445027]
37. Kops O, Zhou XZ, Lu KP. Pin1 modulates the dephosphorylation of the RNA polymerase II C-terminal domain by yeast Fcp1. *FEBS Lett.* 2002; 513:305–311. [PubMed: 11904169]
38. Hani J, Schelbert B, Bernhardt A, Domdey H, Fischer G, Wiebauer K, Rahfeld JU. Mutations in a peptidylprolyl-cis/trans-isomerase gene lead to a defect in 3'-end formation of a pre-mRNA in *Saccharomyces cerevisiae*. *J Biol Chem.* 1999; 274:108–116. [PubMed: 9867817]
39. Verdecia MA, Bowman ME, Lu KP, Hunter T, Noel JP. Structural basis for phosphoserine-proline recognition by group IV WW domains. *Nat Struct Biol.* 2000; 7:639–643. [PubMed: 10932246]
40. Palancade B, Marshall NF, Tremeau-Bravard A, Bensaude O, Dahmus ME, Dubois MF. Dephosphorylation of RNA polymerase II by CTD-phosphatase FCP1 is inhibited by phospho-CTD associating proteins. *J Mol Biol.* 2004; 335:415–424. [PubMed: 14672652]
41. Min SH, Lau AW, Lee TH, Inuzuka H, Wei S, Huang P, Shaik S, Lee DY, Finn G, Balastik M, Chen CH, Luo M, Tron AE, Decaprio JA, Zhou XZ, Wei W, Lu KP. Negative regulation of the stability and tumor suppressor function of Fbw7 by the Pin1 prolyl isomerase. *Mol Cell.* 2012; 46:771–783. [PubMed: 22608923]
42. Lu KP, Liou YC, Zhou XZ. Pinning down proline-directed phosphorylation signaling. *Trends Cell Biol.* 2002; 12:164–172. [PubMed: 11978535]



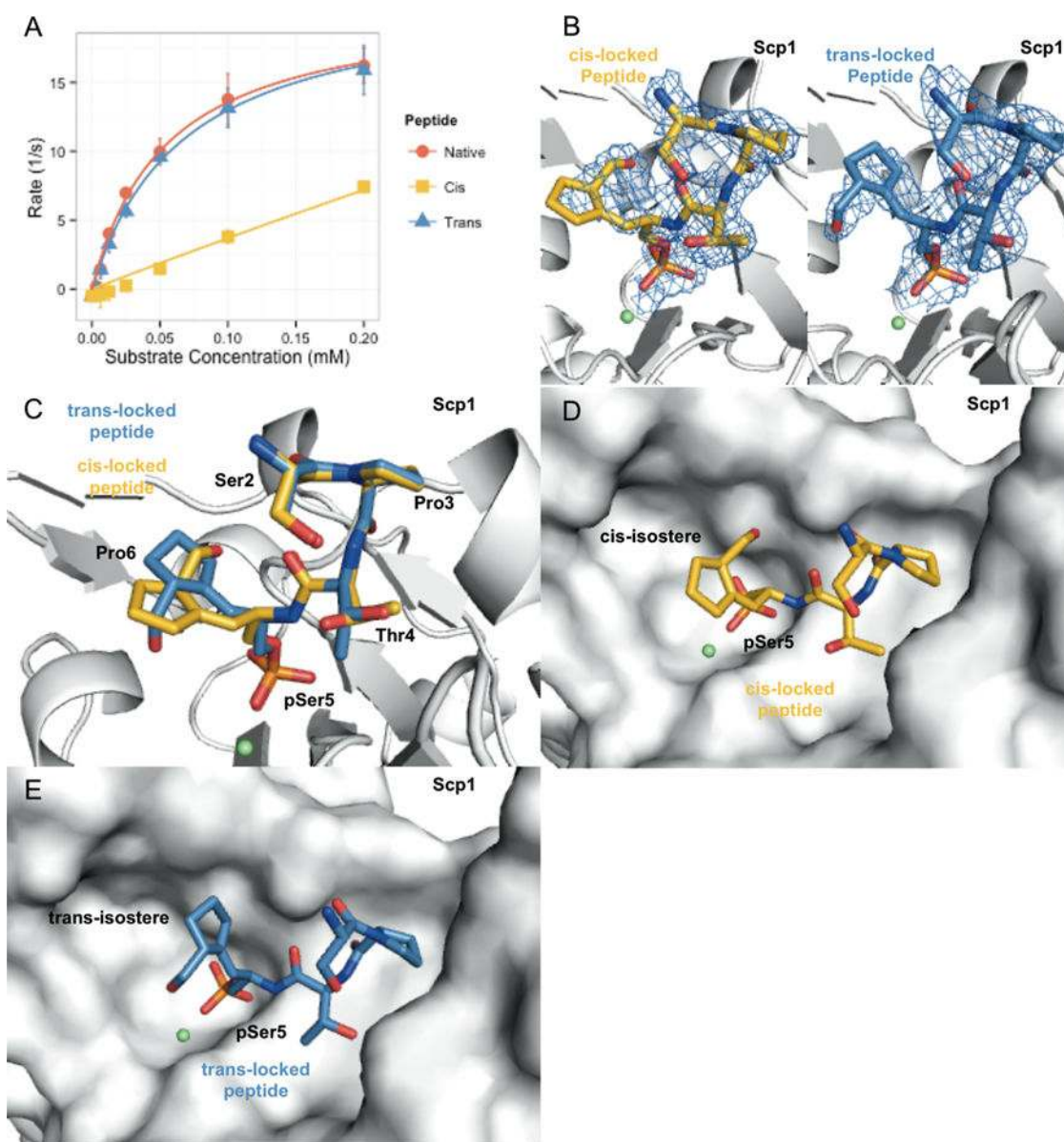
**Figure 1.** Locked proline isosteres. (A) Native proline containing motif in the *trans*-proline configuration (left) and *trans*-locked proline isostere (pSer-Ψ[(*E*)C=CH]-Pro) (right). (B) Native proline containing motif in the *cis*-proline configuration (left) and *cis*-locked proline isostere (pSer-Ψ[(*Z*)C=CH]-Pro) (right). Sequences of synthesized peptides incorporating isosteres are provided in Supplementary Table 1.



**Figure 2.**

*Drosophila melanogaster* Ssu72 + Symplekin analysis using locked proline peptides. (A) Kinetic analysis of Ssu72 + Symplekin against native, *cis*-locked, and *trans*-locked Pro6 containing phospho-Ser5 peptides. Ssu72 + Symplekin shows considerable activity against the *cis*-locked (goldenrod) compound with significantly lower activity against the native peptide (tomato). Activity against the *trans*-locked peptide was not detected above background (blue). Data are from three experimental replicates, error bars indicate standard deviation ( $n = 3$ ). (B)  $2F_o - F_c$  map about *cis*-locked peptide (goldenrod) contoured to  $1\sigma$ . Density accounts for residues analogous to Ser2 through Ser7 of a consensus CTD heptad repeat. (C) Surface depiction of Ssu72 + Symplekin. The image has been rotated with respect to panels B and D about a vertical axis along the Ser5 position by  $\sim 90^\circ$  counterclockwise and tilted toward the viewer by an additional  $\sim 90^\circ$ . The locked-proline isostere fits into a restrictive hydrophobic pocket. (D) Alignment of complex crystal structures of Ssu72 + Symplekin containing *cis*-locked (goldenrod, PDB ID 4ygx) and native (tomato, PDB ID 4imj) peptides. Residues numbered to indicate position in consensus CTD heptad repeat.





**Figure 3.**

Human Scp1 analysis using locked proline peptides. (A) Kinetic analysis of Scp1 against native, *cis*-locked, and *trans*-locked Pro6 containing phospho-Ser5 peptides. Scp1 shows comparable and high activity against native (tomato) and *trans*-locked (blue) peptides. Activity against *cis*-locked peptide (goldenrod) is observed but nearly 10-fold smaller than that observed for native and *trans*-locked substrates. Data are from three experimental replicates; error bars indicate standard deviation ( $n = 3$ ). (B)  $2F_o - F_c$  map about *cis*-locked peptide (goldenrod, left) and *trans*-locked (blue, right) contoured to  $1\sigma$ . Density accounts for residues analogous to Ser2 through Pro6 of a consensus CTD heptad repeat. (C) Alignment of complex crystal structures of Scp1 containing *cis*-locked (goldenrod, PDB ID 4yh1) and *trans*-locked (blue, PDB ID 4ygy) peptides. The structures align well except at the Pro6 location, where they are flipped  $180^\circ$  relative to one another. Residues numbered to indicate



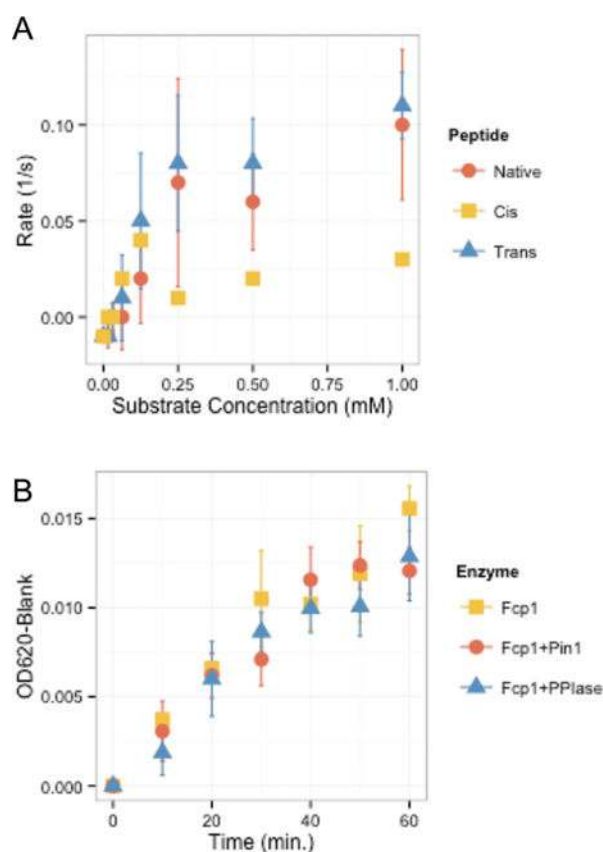
position in consensus CTD heptad repeat. (D) Surface depiction of Scp1 and *cis*-locked peptide (goldenrod). (E) Surface depiction of Scp1 and *trans*-locked peptide (blue).

Author Manuscript

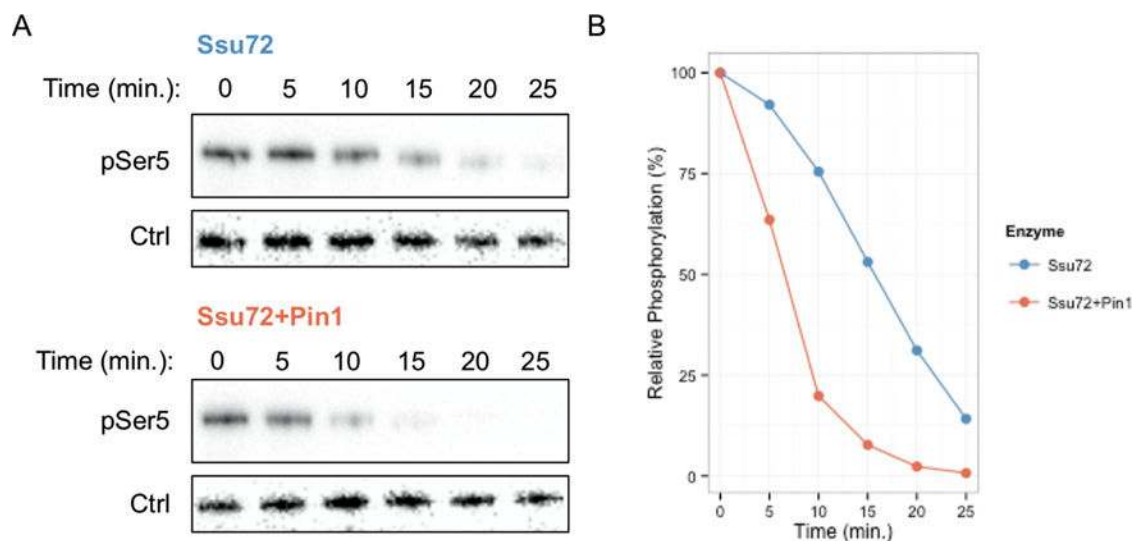
Author Manuscript

Author Manuscript

Author Manuscript

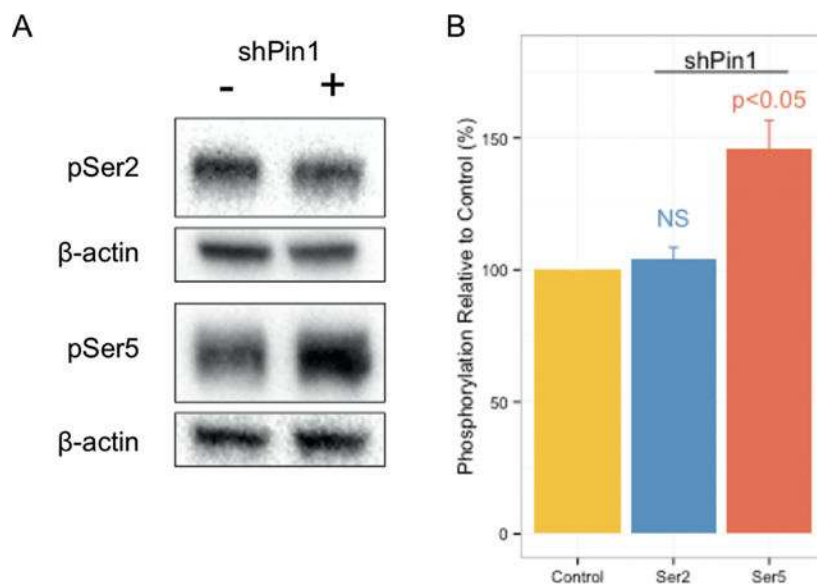


**Figure 4.** Proline isomer specificity of Fcp1. (A) Kinetic analysis of Fcp1 against native, *cis*-locked, and *trans*-locked Pro3 containing phospho-Ser2 peptides. Fcp1 shows comparable and higher activity against native (tomato) and *trans*-locked (blue) peptides. Activity against *cis*-locked peptide (goldenrod) is observed but lower than that observed for native and *trans*-locked substrates. Native and *trans*-locked data are from three experimental replicates; error bars indicate standard deviation ( $n = 3$ ). *cis*-Locked data is from one experimental replicate ( $n = 1$ ). (B) Fcp1/Pin1 coupled assay. All trials show comparable activity with or without Pin1 supplementation. Error bars indicate standard deviation ( $n = 3$ ).

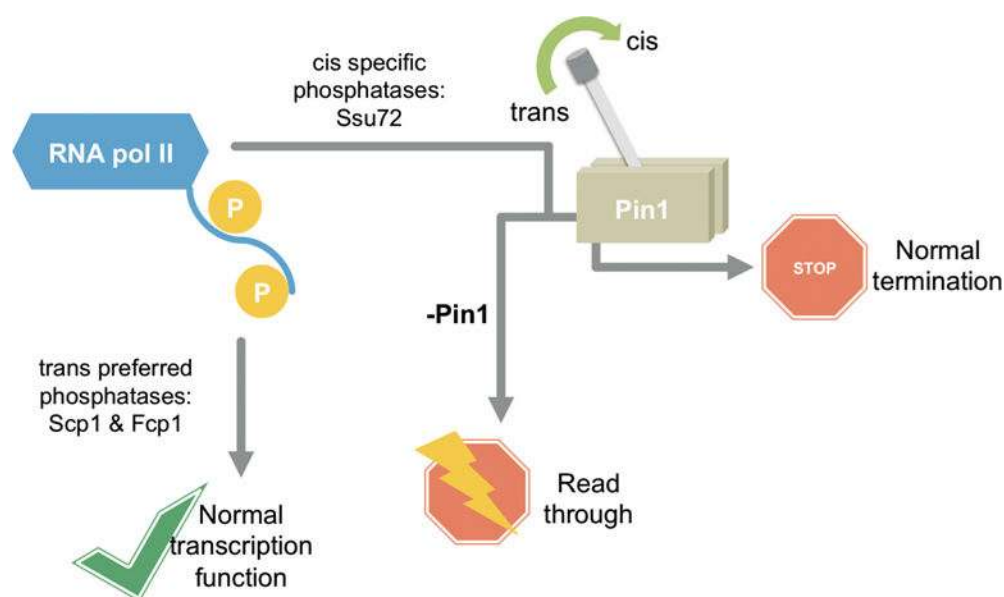


**Figure 5.**

*In vitro* reconstruction Pin1 mediated Ssu72 enhancement. (A) Western blot against TFIIH phosphorylated GST-CTD dephosphorylated by Ssu72 with or without Pin1. Phospho-Ser5 (pSer5) was monitored for reactions containing Ssu72 alone (top) or Ssu72 + Pin1 (bottom) over the indicated time course. Mouse IgG heavy chain (Ctrl) was introduced during reaction quenching to provide a loading control. (B) Quantification of Western blot. Phospho-Ser5 bands were first normalized to loading control and then relative to the respective zero time point for each condition. Blot and quantification represent one experimental replicate of three independent experimental replicates, all displaying increased dephosphorylation upon Pin1 supplementation.



**Figure 6.** Impact of Pin1 knockdown on CTD phosphorylation states in HeLa cell lines. (A) Western blot analysis of phosphorylated Ser2 (top) and phosphorylated Ser5 (bottom) in HeLa cell lines transformed with either empty vector (left) or shPin1 containing vector (right). Blots were performed using three biological replicates of both the empty vector control and shPin1 containing vector cells. Control and knockdown protein sample pairs were prepared in parallel. The three paired sample sets were analyzed on three separate blots. (B) Quantification of Western blot. Phosphorylated Ser5 levels increase 30–60% upon Pin1 knockdown. Quantification of Western blot was performed by first normalizing control and shPin1 samples to the endogenous loading control ( $\beta$ -actin). The shPin1 samples were then normalized to the paired vector control samples. Significance was assessed using Welch's *t* test ( $n = 3$ ).



**Figure 7.** Model of differentiated regulation mediated by proline isomerization of CTD in RNA polymerase II dephosphorylation. Trans-preferred or specific phosphatases, Scp1 and Fcp1, have substrate consistently available due to the thermodynamic preference for *trans*-proline in the CTD. Therefore, these enzymes bypass regulation by prolyl isomerases like Pin1. However, *cis*-specific phosphatases like Ssu72 rely on a minor substrate pool containing the *cis*-proline isomer and quickly deplete their available substrate. Prolyl isomerases, like Pin1, can restore the equilibrium between *cis* and *trans* isomers and replenish substrate pools. This regulatory switch provides for proper RNA polymerase II CTD phosphorylation levels and normal transcription termination. Upon Pin1 disruption or knockdown, global transcription defects, like read-through, may occur.

Table 1

## X-ray Crystallography Data Collection and Refinement Statistics

|  | Ssu72 + <i>cis</i> -locked peptide | Scp1 + <i>trans</i> -locked peptide | Scp1 + <i>cis</i> -locked peptide |
|--|------------------------------------|-------------------------------------|-----------------------------------|
| PDB code   | 4YGX                               | 4YGY                                | 4YH1                              |
|  | Data Collection                    |                                     |                                   |
| wavelength (Å)   | 1.0332                             | 1.03334                             | 0.97648                           |
| space group  | <i>P</i> 4                         | <i>C</i> 2                          | <i>C</i> 2                        |
| cell dimensions:   |                                    |                                     |                                   |
| <i>a</i> , <i>b</i> , <i>c</i> (Å)                                   | 127.9, 127.9, 105.9                | 125.3, 78.3, 63.0                   | 125.1, 78.8, 62.9                 |
| $\alpha$ , $\beta$ , $\gamma$ (deg)                                  | 90.0, 90.0, 90.0                   | 90.0, 112.6, 90.0                   | 90.0, 112.54, 90.0                |
| resolution <sup>a</sup> (Å)  | 127.88–2.95 (3.00–2.95)            | 64.86–2.36 (2.40–2.36)              | 65.11–2.20 (2.24–2.20)            |
| no. of unique reflections  | 35810                              | 22534                               | 27838                             |
| <i>I</i> / $\sigma$ ( <i>I</i> )                                     | 20.5(1.4)                          | 29.3(3.3)                           | 11.3(1.4)                         |
| Completeness (%)   | 99.2(96.2)                         | 96.9(85.2)                          | 97.0(83.0)                        |
| redundancy   | 6.9(4.5)                           | 3.7(3.1)                            | 3.7(2.8)                          |
| <i>R</i> <sub>sym</sub> (%)  | 10.4(89.9)                         | 5.0(31.0)                           | 11(49.5)                          |
|  | Refinement                         |                                     |                                   |
| resolution (Å)   | 50.00–2.95                         | 50.00–2.36                          | 50.00–2.20                        |
| no. of reflns (test set)   | 33965 (1844)                       | 21427 (1105)                        | 26369 (1469)                      |
| <i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub> <sup>b</sup> (%) | 20.6/25.4                          | 18.3/24.6                           | 19.3/24.6                         |
| no. of atoms   |                                    |                                     |                                   |
| protein  | 8113                               | 2908                                | 2908                              |
| Mg <sup>2+</sup>   | NA                                 | 2                                   | 2                                 |
| ligand   | 42                                 | 72                                  | 72                                |
| water  | 10                                 | 62                                  | 141                               |
| <i>B</i> -factors (Å <sup>2</sup> )                                  |                                    |                                     |                                   |
| protein  | 88.7                               | 46.8                                | 32.7                              |
| Mg <sup>2+</sup>   | NA                                 | 53.5                                | 53.6                              |
| ligand   | 109.6                              | 69.3                                | 67.8                              |
| water  | 62.6                               | 50.3                                | 38.0                              |
| RMS deviations   |                                    |                                     |                                   |
| bond lengths (Å)   | 0.013                              | 0.016                               | 0.017                             |
| bond angles (deg.)   | 1.74                               | 1.86                                | 2                                 |
| Ramachandran plot <sup>c</sup> (%)                                   |                                    |                                     |                                   |
| favored  | 96.9                               | 94.7                                | 95.5                              |
| allowed  | 3.1                                | 5.3                                 | 4.5                               |
| outlier  | 0                                  | 0                                   | 0                                 |

<sup>a</sup> Highest resolution shell is shown in parentheses.

<sup>b</sup> *R*<sub>free</sub> is calculated with 5% of the data randomly omitted from refinement.

<sup>c</sup> Ramachandran statistics generated in MolProbity.