Published in final edited form as:

Methods Enzymol. 2018 ; 607: 269–297. doi:10.1016/bs.mie.2018.04.020.

Chemical Tools for Studying the Impact of *cis/trans* Prolyl Isomerization on Signaling: A Case Study on RNA Polymerase II Phosphatase Activity and Specificity

Nathaniel Tate Burkholder^{*}, Brenda Medellin^{*}, Seema Irani[†], Wendy Matthews^{*}, Scott A. Showalter^{‡,§}, Yan Jessie Zhang^{*,¶,1}

^{*}Department of Molecular Biosciences, The University of Texas at Austin, Austin, TX, United States

[†]Department of Chemical Engineering, The University of Texas at Austin, Austin, TX, United States

[‡]Department of Chemistry, Pennsylvania State University, University Park, PA, United States

[§]Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA, United States

[¶]Institute for Cellular and Molecular Biology, The University of Texas at Austin, Austin, TX, United States

Abstract

Proline isomerization is ubiquitous in proteins and is important for regulating important processes such as folding, recognition, and enzymatic activity. In humans, peptidyl-prolyl isomerase *cis-trans* isomerase NIMA interacting 1 (Pin1) is responsible for mediating fast conversion between *cis*- and *trans*-conformations of serine/threonine–proline (S/T–P) motifs in a large number of cellular pathways, many of which are involved in normal development as well as progression of several cancers and diseases. One of the major processes that Pin1 regulates is phosphatase activity against the RNA polymerase II C-terminal domain (RNAPII CTD). However, molecular tools capable of distinguishing the effects of proline conformation on phosphatase function have been lacking. A key tool that allows us to understand isomeric specificity of proteins toward their substrates is the usage of proline mimicking isosteres that are locked to prevent *cis/trans*-proline conversion. These locked isosteres can be incorporated into standard peptide synthesis and then used in replacement of native substrates in various experimental techniques such as kinetic and thermodynamic assays as well as X-ray crystallography. We will describe the application of these chemical tools in detail using CTD phosphatases as an example. We will also discuss alternative methods for analyzing the effect of proline conformation such as ¹³C NMR and the biological implications of proline isomeric specificity of proteins. The chemical and analytical tools presented in this chapter are widely applicable and should help elucidate many questions on the role of proline isomerization in biology.

¹Corresponding author: jzhang@cm.utexas.edu.

1. INTRODUCTION

Among the 20 natural amino acids, proline is the only residue with a significant fraction (5%–20% depending on context) found in isomeric *cis*-conformations (Fig. 1A). The capacity for interconversion between *cis*- and *trans*-conformations at peptidyl-prolyl bonds can promote energetically favorable folding patterns in proteins or even act as regulatory switches for protein recognition (Lu, Liou, & Zhou, 2002). However, the presence of *cis*-proline is generally low and isomerization of peptidyl-prolyl bonds is inherently slow (Wedemeyer, Welker, & Scheraga, 2002). In order to regulate *cis/trans*-proline isomerization, organisms have evolved a large family of peptidyl-prolyl *cis/trans* isomerases (PPIases) that increase the rate of interconversion between *cis*- and *trans*-proline residues to mediate the equilibration of the population of *cis/trans*-prolines (Shaw, 2002). An important PPIase in humans is called PPIase NIMA-interacting 1 (Pin1), which targets phosphorylated Ser/Thr-Pro (S/T-P) motifs that are enriched in phosphorylation-mediated signaling pathways (Yaffe et al., 1997). Regulation of S/T-P motifs by Pin1 can alter the enzymatic activity, protein stability, and cellular localization of its substrates (Liou, Zhou, & Lu, 2011; Lu, Finn, Lee, & Nicholson, 2007; Wulf, Finn, Suizu, & Lu, 2005; Zhou & Lu, 2016). Studying the roles and mechanisms of *cis/trans*-proline-mediated regulation is critical for understanding key signaling pathways as well as elucidating the underlying mechanisms of certain cancers and diseases (Liou, Zhou, & Lu, 2011; Lu & Zhou, 2007). However, it is methodologically difficult to distinguish *cis*- and *trans*-isomers due to their identical chemical composition and molecular weight. Thus, innovative strategies are needed to detect proline conformation in various biological contexts.

One of the essential processes that proline isomerization regulates is transcription mediated by RNA polymerase II (RNAPII), the main enzyme responsible for generation of messenger RNAs as well as small nuclear RNAs in eukaryotic cells (Hanes, 2014). In order to efficiently coordinate transcription of vastly complex genomes, RNAPII has evolved a highly repetitive C-terminal domain (CTD) consisting of consensus heptad repeats $Y_1S_2P_3T_4S_5P_6S_7$ that undergo a variety of posttranslational modifications (PTMs). Various PTMs on the CTD of RNAPII specifically and temporally recruit different transcriptional proteins to execute their appropriate functions at different stages of transcription (Eick & Geyer, 2013). The combination of writing, reading, and erasing of posttranslational marks on the RNAPII CTD is collectively referred to as the “CTD code,” which is essential for proper cycling through each stage of general transcription (Buratowski, 2003; Corden, 2013; Eick & Geyer, 2013; Jeronimo, Bataille, & Robert, 2013; Mayfield, Burkholder, & Zhang, 2016). Of the five residues within the heptad repeats that can be phosphorylated, the two SP motifs (YS₂P₃TSPS₆) are the major sites of dynamic phosphorylation/dephosphorylation required in each round of transcription (Eick & Geyer, 2013; Harlen & Churchman, 2017; Mayfield et al., 2016; Schüller et al., 2014). The presence of SP motifs in the CTD raises the possibility that *cis/trans* prolyl isomerization plays a role in CTD recognition. Indeed, temperature sensitive mutants of Ess1 (Pin1 homolog in *Saccharomyces cerevisiae*) have a slow growth or temperature sensitive phenotype along with transcriptional defects in elongation and termination (Wu et al., 2000). The termination defects may be in part related to an important regulator of termination called Nrd1 which prefers pS₅-*cis*-P₆ as its substrate

(Kubicek et al., 2012). Studying how *cis/trans* prolyl isomerization impacts the specificity of CTD-binding proteins, as well as how altered recognition impacts transcriptional processing, is crucial for deciphering the CTD code.

In addition to altering protein–protein recognition, another mechanism for *cis/trans*-proline isomerization affecting the CTD code is through altering the enzymatic activity of CTD modification enzymes. A number of phosphatases have been found to dephosphorylate RNAPII CTD, some of which play critical roles in transcription. Ssu72 is one such CTD phosphatase which has a high specificity for phosphorylated S₅ of CTD heptad repeats (Krishnamurthy, He, Reyes-Reyes, Moore, & Hampsey, 2004) and has been implicated in both 3′ end processing and transcriptional termination (Steinmetz & Brow, 2003). Interestingly, the crystal structure of the complex between Ssu72 and phosphorylated S₅ (pS₅) CTD substrate revealed that the pS₅P₆ is selectively captured in the *cis*-conformation (Luo et al., 2013; Werner-Allen et al., 2011; Xiang et al., 2010; Fig. 2A) compared to another CTD phosphatase that captured P₆ in the *trans*-conformation (Zhang et al., 2006; Fig. 2B). For crystal structures showing substrate bound in the *trans*-conformation, this does not preclude the possibility of bound substrate being in the *cis*-conformation in some of the crystal units. However, since the *trans*-form substrate is more abundant in these structures the signal from bound *cis*-form substrates will be averaged out. The structural evidence for Ssu72 recognition of pS₅-*cis*-P₆ is especially convincing then, as the averaging effect resulted in the proline being observed only in the *cis*-conformation (Xiang et al., 2010). Furthermore, the availability of *cis* proline is critical for efficient dephosphorylation of pS₅ substrate by Ssu72 as shown through addition of Pin1 (Mayfield et al., 2015; Zhang et al., 2012). However, to fully elucidate the dynamics of Ssu72 recognition of *cis* and *trans* proline conformations, we and others have developed specialized chemical tools and methods to decipher the effects of one conformation vs the other.

While *cis/trans*-proline isomerization plays important biological roles in the function of many proteins, this process is difficult to study because tools such as DNA sequencing or mass spectrometry are not able to detect the differences in *cis*- and *trans*-proline within proteins. On the other hand, structural biology methods such as X-ray crystallography and nuclear magnetic resonance (NMR) can provide detailed information on the interaction between the target protein's isomeric specificity and the ligand containing the proline residue. However, both of these techniques are highly dependent on the applicability of the methodology for each target protein and substrate. If the protein does not crystallize or does not incorporate the peptide ligand easily, then X-ray crystallography will be not suitable. NMR is a powerful method for detecting atomic detailed structure that can distinguish *cis*- and *trans*-prolines, but this technique is incompatible with larger proteins and complexes. In this chapter, we introduce the design of peptidomimetic isosteres that are locked in specific isomeric conformations that can be incorporated into any ligand peptide sequence for testing isomeric specificity of a target protein (Fig. 1B). As an example, we show how we generated peptidomimetic compounds with the sequence of the heptad repeats of RNAPII CTD and used them to understand how prolyl isomerization affects the activity and specificity of CTD phosphatases. In this chapter, we describe in detail how the locked proline peptidomimetics were designed and how these compounds were used as tools to determine the isomeric preference of CTD phosphatases. In essence, we used these locked peptidomimetics in place

of native substrate in kinetic assays to determine the rates of dephosphorylation as well as in X-ray crystallography to show the binding modes of substrates with different isomeric conformations. We also compared the crystal structures of CTD phosphatases bound to locked peptidomimetics with structures bound to native peptides, showing that the locked isosteres accurately mimic natural substrates. Complementary studies using ^{13}C NMR to analyze isomeric states of prolines in CTD are also discussed briefly. Finally, how Pin1-mediated regulation of CTD phosphatase substrate availability can lead to differential transcription resulting in alternative cellular outcomes will be considered. While we will focus attention toward the effect of *cis/trans*-proline isomerization on CTD phosphatase activity and specificity, analogous strategies can be adapted for studying any protein whose function is regulated by the proline conformation.

2. USING *cis/trans*-LOCKED SERINE-PROLINE ISOSTERES TO STUDY ACTIVITY AND SPECIFICITY OF CTD PHOSPHATASES

Enzymatic reactions against protein substrates can often be biochemically characterized by producing a piece of the substrate protein sequence and using it in *in vitro* kinetic assays. However, native proline-containing polypeptides cannot be used to assess the effect of *cis/trans*-proline isomerization because a mixture of *cis*- and *trans*-isomers of prolines will coexist and autoconvert during the measurement. To prevent interconversion and obtain a homogenous population of specific isomers, we designed peptidomimetic compounds using unnatural proline homologues (Wang et al., 2003). In such peptidomimetics, the proline residues are replaced by locked isostere moieties in which the peptide bond connecting proline to the preceding residue is replaced by a carbon double bond either in the *cis*- or *trans*-configuration. Unlike peptidyl-prolyl bonds, a carbon double bond cannot be easily rotated preventing interconversion between *cis*- and *trans*-prolines. Combining the aspect of isomeric locking with efficient synthesis and isolation strategies for one or the other isomer essentially guarantees that the final product will consist of a stable and homogenous species of a specific isomer configuration. Thus, these compounds are invaluable tools for studying the effect of *cis/trans*-proline isomerization on protein function.

2.1. Design and Synthesis of *cis/trans*-Locked Serine-Proline Isosteres

When designing *cis*- and *trans*-locked proline peptidomimetics, we aim to alter the chemical nature of the peptidyl-prolyl bond to prevent free rotation but minimize structural alterations to the rest of the peptide. Therefore, we kept the primary sequence of the substrate peptides the same as the natural protein sequence while replacing the peptidyl-prolyl bond with a carbon-carbon double bond. In order to accommodate a double bond between the proline and its preceding amino acid, the prolyl nitrogen was substituted with a carbon and the carbonyl oxygen of the preceding residue was removed (Wang et al., 2003). Using the carbon-carbon double bond to lock the isostere has several advantages: (1) the geometry of a carbon-carbon double bond is similar to a peptide bond, but the p-orbitals of the pi-bond between double bonded carbons must remain parallel to each other and therefore cannot be rotated (resulting in a planar locked configuration); (2) carbon-carbon double bonds are more stable than carbon-nitrogen double bonds in aqueous solution, increasing their versatility in experimental design; and (3) carbon-carbon double bonded peptidomimetics

are less likely to be degraded by proteases that target peptidyl-prolyl bonds. Loss of the carboxyl group and adoption of a planar conformation in the dipeptide isosteres are unlikely to significantly change conformation of the entire peptide. Overall, modifying peptidyl-prolyl dipeptides with carbon-carbon double bonds poses minimal structural disturbances to the peptide chains while securing proline conformation.

Locked XP (X for any amino acid) isosteres can be chemically synthesized and then incorporated during solid-phase peptide synthesis (Hart, Sabat, & Etzkorn, 1998). Locked proline isosteres are not yet commercially available, but there are relatively efficient synthesis strategies for their production. The key aspect in the chemical synthesis of these isosteres is obtaining sufficient quantities of product with the correct stereochemistry in the double-bond formation step. Two chemical synthesis strategies have been successfully applied to generate stereospecific XP-locked isosteres. The first scheme uses careful application of organocopper-mediated anti-S_N2' reactions (Otaka et al., 2002; Sasaki et al., 2006). This strategy is particularly effective with the synthesis of the *trans*-isostere, in which an SP-like oxazolidinone (**4a**) is modified with a silylmethyl group in an organocopper-mediated anti-S_N2' attack to predominantly form the *trans*-alkene precursor **5a** (Fig. 3; Otaka et al., 2002). However, syntheses based on organocopper chemistry have very low yield in the production of L-form X-*cis*-P isosteres which mimic naturally occurring amino acids (Otaka et al., 2002; Sasaki et al., 2006, Table 1). Instead, rearrangement chemistry can be used to position a double bond in the desired peptidyl-prolyl position to efficiently synthesize SP analogs in either *cis*- or *trans*-configuration (Wang et al., 2003). Synthesis of either isostere begins by making a dipeptide-like precursor through condensation of a Weinreb amide (**2b** and **1c**) with cyclopentyllithium derived from 1-iodocyclopentene (**3b**) (Figs. 4 and 5). The key stereoselective step for the S-*cis*-P is a Still-Wittig rearrangement of the ether functional group and the cyclopentene to form ~53% of the *cis*-isomer (Fig. 4, **6b**→**7b-cis**). Yield for the *trans*-isomer in this synthesis is poor (~28%), so instead an Ireland-Claisen rearrangement between an ester functional group and the double bond of a cyclopentene can be performed to acquire ~52% of the *trans*-isomer (Fig. 5, **4c**→**5c**; Wang et al., 2003). Synthesis of the S-*trans*-P isosteres using either method takes eight steps with overall yields of about 17% using organocopper chemistry vs 13% for Ireland-Claisen rearrangement (Table 1). On the other hand, S-*cis*-P can be synthesized with Still-Wittig rearrangement with many fewer steps and higher overall yield (10 steps, 20% yield) than with organocopper chemistry (Table 1).

The synthesis strategies described in this section can be generally performed for any Xaa-Pro sequence and produce abundant amounts of *cis*- or *trans*-isomer with high purity. Although these isosteres are not yet commercially available, increased usage will hopefully encourage industrial chemical providers to generate locked dipeptide mimetic compounds for scientific use. Once the dipeptide mimetic precursors are synthesized and purified, they can be incorporated into routine solid-phase peptide synthesis as a unit as if they were a single amino acid. Phosphate groups can be added at this stage to the target serine/threonine dephosphorylation site. The polypeptides (incorporating X-*E* or X-C = C-P isosteres) are then purified through chromatography and used as substrates or ligands in characterizing the isomeric specificity of proteins of interest. For our studies, we used peptides consisting of

RNAPII CTD sequences with phosphorylated *S-cis/trans*-P-locked moieties to determine the effect of SP motif conformation on CTD phosphatase function (Table 2).

2.2. Malachite Green Kinetic Analysis of CTD Phosphatase Activity Against *cis/trans*-Locked Isosteres

In order to study RNAPII CTD recognition, we generated two sets of peptide mimetics with either SP motif of the CTD replaced by locked, dipeptide isosteres (phosphoryl-S-(*E*, *Z*)C=CH)-P] (Table 2). These peptidomimetic compounds containing locked pSP isosteres can thus be used in place of natural substrates in kinetic or thermodynamic measurements to investigate whether CTD-binding proteins prefer *cis*- or *trans*-SP-motifs. Here, we used CTD peptidomimetics containing proline-locked isosteres to investigate proline isomeric specificity of different CTD phosphatases including: (1) Fcp1, which dephosphorylates pS₂ during general transcription (Ghosh, Shuman, & Lima, 2008; Hausmann & Shuman, 2002; Kobor et al., 1999); (2) Ssu72, the phosphatase responsible for pS₅ dephosphorylation during general transcription (Krishnamurthy et al., 2004; Xiang et al., 2010); and (3) Scp1, which inhibits transcription of neuronal genes and dephosphorylates pS₅ (Yeo et al., 2005; Yeo, Lin, Dahmus, & Gill, 2003; Zhang et al., 2006).

The specific activity of phosphatases is often measured by the malachite green phosphate detection method (MGPD) using short phosphorylated peptides as substrates. MGPD is a colorimetric assay based on a complex formation between malachite green molybdate (BIOMOL® Green, Enzo Life Sciences) and free orthophosphate, the absorbance of which can be easily detected at wavelengths of 620–650 nm. The overall procedure for MGPD is listed below:

Step 1: Phosphate Standard curve determination (see note 1)

1. Allow an aliquot of Malachite Green Reagent to warm up to room temperature (see note 2).
2. Perform triplicate serial dilutions of a phosphate standard (see note 3) (usually Malachite Green Reagents will be sold along with a phosphate standard). For a 96-well clear microplate scale assay, perform eight serial dilutions in which the last tube should be solvent only acting as a blank.
3. Using a multichannel pipette, take 20 µL of each dilution and add to a clean, clear microplate.
4. Using a multichannel pipette, add 40 µL of room temperature Malachite Green Reagent (as an extra precaution, filtered tips may be used to prevent phosphate contamination).
5. Cap microplate or seal with tape. Incubate at room temperature for 15–20 min to allow green color development.
6. Measure absorbance at 620 nm (precipitation should not be observed as it indicates oversaturation of the solution and will yield inaccurate readings).

7. Average triplicate absorbance measurements and subtract the blank from each reading. Plot absorbance at 620 nm on the y -axis and [phosphate] (M) on the x -axis. Fit to a linear equation and evaluate goodness of fit.

Notes

1. Ideally, the standard curve will be performed using the protein buffer as solvent as this will provide a more comparable standard to the enzymatic reaction. For different buffers and/or old reagent, it is recommended to repeat the standard curve.
2. Malachite Green Reagent is light sensitive so using dark opaque tubes or wrapping tubes in foil is recommended.
3. Malachite Green Reagent should be highly sensitive to phosphate in the nanomolar range; however, carefully inspecting the linearity of standard curve is advised to ensure that the lower limit of phosphate concentration can be reasonably detected.

Step 2: Enzyme activity assay: Phosphatase activity against peptidomimetic substrate

1. Allow an aliquot of Malachite Green Reagent to warm up to room temperature.
2. Prepare $10 \times$ concentrated working stocks of buffer, substrate, and enzyme.
3. Perform a serial dilution of the $10 \times$ substrate at a concentration that when diluted 10-fold results in a maximal rate of phosphate release for the target phosphatase%. Leave the last dilution without enzyme as a control for background phosphate signal.
4. Mix buffer, water, and enzyme in a total volume of 36 μL each (excluding substrate) in an eight-strip tube ($3 \times$ reactions and $1 \times$ control without enzyme for background phosphate from each substrate dilution).
5. Prepare a 96-well clear microplate with 40 μL of Malachite Green Reagent for quenching each reaction. Seal the plate with plate to prevent contamination.
6. Start reactions by multichannel pipetting 4 μL of $10 \times$ substrate serial dilutions to each reaction and control strip tube. Incubate reactions at an appropriate temperature and time for the target enzyme for as long as the reactions remain in the steady state (see note). Cap strip tubes to prevent evaporation.
7. Quench reactions by adding 20 μL from each strip tube to 40 μL of Malachite Green Reagent in the microplate.
8. Reseal the microplate with tape to avoid contamination and incubate at room temperature for 10–20 min for green color development. During this time the P_i + MG complex will be formed.
9. Read the absorbance in the plate reader at 620 nm. Using the standard curve, convert absorbance to concentration of phosphate in order to calculate the rate of phosphate release for each phosphatase reaction.

Notes: Determine the linear range of the steady state and the concentration of substrate to result in maximal rate of dephosphorylation for the target enzyme by performing a time course with constant concentrations of substrate.

We used the highly sensitive malachite green assay in combination with proline-locked CTD peptides to determine the rates of CTD phosphatases against *cis*- and *trans*-specific substrates (Mayfield et al., 2015). We determined the specificity constants for each enzyme and substrate by plotting the substrate concentration against the initial rate and fitting to the classical steady-state Michaelis–Menten kinetic equation. Using MGPD, we found that Scp1 has an almost 10-fold preference for the pS₅-*trans*-P₆ substrate with a k_{cat}/K_m of $323 \pm 22 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$ compared to the *cis*-proline substrate with a k_{cat}/K_m of only $38.0 \pm 1.3 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$ (Mayfield et al., 2015). Similarly, Fcp1 can dephosphorylate pS₂-P₃ of the CTD in either conformation, but it has a twofold preference for the *trans*-locked peptidomimic compared to the *cis*-proline substrate (Mayfield et al., 2015). In contrast, Ssu72 had virtually no activity against the pS₅-*trans*-P₆ peptide and very low activity against the native peptide. Ssu72 did have robust activity against the *cis*-peptide with a k_{cat}/K_m of $5.24 \pm 0.08 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$ (Mayfield et al., 2015), which supports the crystallographic data suggesting that Ssu72 is CTD phosphatase is specific for *cis*-proline in substrates. Combining the malachite green assay with our proline-locked peptides yielded accurate and quantitative results revealing the *cis/trans*-proline specificity of several CTD phosphatases as well as their activity toward each distinct isomeric substrate.

While the malachite green assay is highly sensitive, there are issues that must be addressed for successful application of this assay on studying phosphatase activity. Like all kinetic assays, the enzymatic conditions (e.g., enzyme concentration, temperature, etc.) need to be optimized in order to obtain the best signal/noise and sensitivity. The reaction time is a particular important factor to consider for this assay because the color development buffer tends to precipitate when it is saturated with the ortho-phosphate complex, leaving a narrow window for detection. Additionally, one must take into account that certain additives may interfere with the assay by either showing a strong absorbance at 620 nm or interfering with molybdate–phosphate complex formation. For example, commonly used detergents such as Triton X-100, SDS, Tween-20, and NP-40 can adversely affect signal readings at 620 nm (Carter & Karl, 1985). Divalent cations such as Zn²⁺ and Ca²⁺ can also affect signal by forming salts with phosphates which frequently precipitate due to their low solubility. Finally, the high sensitivity of the malachite green assay makes it prone to high noise interference from any source of free inorganic phosphate. While it is difficult to completely eliminate phosphate contamination, negative controls lacking enzyme and/or phosphorylated substrate should be used to subtract unwanted noise from activity curves.

2.3. X-Ray Crystallography of CTD Phosphatases Bound to *cis/trans*-Locked Peptides

After kinetically quantifying the proline conformation preference of several CTD phosphatases, we obtained the binding mode of locked peptidomimic substrates using X-ray crystallography. X-ray crystallography has unexpectedly been an invaluable tool for studying the substrate specificity of CTD phosphatases (Xiang et al., 2010; Zhang et al., 2006). It can directly visualize the bound substrate conformation and identify protein residues involved in

substrate recognition. However, since X-ray crystallography captures the overall signal from many repeating units of protein molecules, signal from any minor conformations that exist in the crystal may not be great enough to be observed in the occupancy-weighted average. Structure determination of CTD phosphatases bound to locked peptidomimetics on the other hand allowed us to directly visualize binding modes of less favorable but still biologically relevant substrate proline-conformations.

Since locked peptidomimetics are structurally analogous to native peptides, the procedure for determining complex structures by X-ray crystallography is identical to structural determination of proteins in complex with natural peptide ligands. If the target protein turns over the substrate through enzymatic activity, a mutant variant of the target protein must be used in order to capture the binding of substrate in the active site without releasing the product. Protein crystallization is generally an empirical process often requiring much trial and error for each different protein. Müller's recent review provides a good general procedure of the protein–ligand complex crystallization process if you are interested in a more detailed overview (Müller, 2017). To obtain protein–ligand complexes within target protein crystals, soaking is often a good strategy to start with. We begin by testing a range of ligand concentrations and soaking times with good quality crystals while watching for signs of crystal deterioration upon complex formation. Stably soaked crystals can then be cryocooled and data can be collected to obtain the structures of protein–ligand complex. Ligand binding may cause substantial conformational changes and disrupt crystal packing, so alternatively cocrystallization of the ligand and protein can be attempted. However, there is no guarantee that cocrystallization will work in the unliganded crystallization conditions thus requiring rescreening with ligand included. Incorporating ligand into protein crystals depends on both the protein and substrate, but obtaining complex structures with locked peptides should be no different than for native peptides due to the highly similarity in structure between peptidomimetic and natural peptides.

For soaking experiments with Ssu72, we began by crystallizing inactive variants of *Drosophila* Ssu72 (C13D/D144N)–symplekin complexed to substrate peptides. Symplekin is a scaffold protein bound to Ssu72 during transcription, which has been shown to stabilize Ssu72 (Luo et al., 2013; Xiang et al., 2010). We have shown that the association between Ssu72 and symplekin binding does not compromise the enzymatic activity but significantly improves the diffraction limits of crystals (Luo et al., 2013). Preformed crystals of Ssu72–symplekin were transferred to crystal trays containing a serial dilution of locked peptidomimetic compounds (highest concentration at 2 mM) in mother liquor (Mayfield et al., 2015). During soaking, Ssu72–symplekin crystals appeared to be stable and showed no sign of visible deterioration. The crystals were harvested and cryoprotected after incubation at room temperature overnight. In cases where protein crystals lose diffraction after compound soaking, several strategies should be attempted to obtain optimal diffraction of the complex crystals: (1) reducing the ligand concentration; (2) reducing soak time (even soaking less than a minute may successfully incorporate the ligand); and (3) including additives that may help improve the stability of a protein crystal upon ligand complexation (Hassell et al., 2007). If the data resolution is good but the ligand density is weak, increasing the ligand concentration and soaking times may improve complex formation. Obtaining

optimal protein–ligand complex structures via crystal soaking requires a delicate balance between ligand binding and crystal maintenance.

Guided by these strategies, we obtained the complex structure of Ssu72–symplekin bound to the *cis*-proline isostere compound at high resolution (2.95 Å; Fig. 6A) but not of the *trans*-proline peptidomimic (Mayfield et al., 2015). Comparison of the structures of Ssu72–symplekin bound to the pS₅-*cis*-P₆ isostere or native pS₅P₆ revealed near identical binding modes (Fig. 6B and C), demonstrating that the *cis*-locked isostere is a great mimic for *cis*-proline. Careful inspection of the substrate-bound structure of Ssu72 provides a plausible reason for its specificity toward *cis*-proline. Effective substrate recognition by Ssu72 requires phosphoserine occupancy at its phosphate-binding site as well as the flanking proline residue in an adjacent hydrophobic pocket (Fig. 6D). P₆ of the CTD fits tightly in the hydrophobic pocket consisting of residues L₄₅, P₄₆, L₈₂, and M₈₅ of Ssu72, but only the *cis*-isomer seems to be compatible with the cavity shape. When we tried to model a *trans*-proline into the substrate, unfavorable steric clashes occurred with active site residues L₈₂ and M₈₅. Overall, the structure of Ssu72–symplekin bound to the *cis*-proline peptidomimic replicates the native substrate-binding mode very well, solidifying the hypothesis that Ssu72 is a *cis*-proline-dependent phosphatase.

Unlike Ssu72, our kinetic characterization Scp1 indicated that while it heavily favors the *trans*-form of CTD substrates it can still efficiently dephosphorylate either isomer (Mayfield et al., 2015). With significantly lower activity against *cis*-proline substrates that occur much less frequently in nature, it is not surprising that the structure of Scp1 in complex with native CTD peptides only shows the *trans*-proline configuration (Zhang et al., 2006). To compare the binding modes of Scp1 with CTD substrates of different isomeric configurations, we obtained the complex structures of Scp1 bound to *cis*- or *trans*-locked isostere peptidomimetic compounds (Mayfield et al., 2015). We were able to obtain diffracting crystals of Scp1 D96N (inactive variant) after soaking with 1 mM of either peptidomimic for 2 h at room temperature (crystals diffract to 2.20 Å for the complex with *cis*- and 2.36 Å for the complex with *trans*-locked compounds). The structure of Scp1 in complex with *trans*-locked peptidomimetic compounds echoed the conformation of Scp1 bound to the natural CTD peptide, confirming again that locked isosteres accurately mimic native peptidyl-prolyl conformations (Fig. 7A and B). The comparison between the structures containing the two locked isosteres shows that there are conserved molecular interactions that are essential for Scp1 to recognize its substrates (Fig. 7C and D). One of the major sites of substrate recognition for Scp1 is its hydrophobic pocket, which appears to tightly interact with residues T₄ and P₆ of both CTD phosphomimetics (Fig. 7E). This conserved interaction demonstrates how Scp1 can still efficiently dephosphorylate substrates with P₆ in either isomeric conformation. However, unlike in Ssu72 where P₆ strongly contributes to substrate recognition by tightly binding in a deep pocket, Pro6 likely contributes little to Scp1 substrate recognition as it is situated at the rim of the active site with very few discernible interactions. In particular, locked *cis*-P₆ is much more solvent exposed than locked *trans*-Pro6 (Fig. 7D), which likely explains why Scp1 prefers *trans*-over *cis*-proline substrates. The availability of peptidomimetics with *cis*-locked isosteres allows us to directly visualize the binding mode of a less favored substrate conformation in our target protein that X-ray crystallography would normally not be able to detect.

3. DETERMINATION OF PHOSPHATASE PROLINE CONFORMATION SPECIFICITY USING ^{13}C NMR

A powerful method to detect subtle structural changes in biological molecules is protein NMR, which is embraced by scientists who try to study changes in protein conformations due to its ability to accommodate and even quantify those structural dynamics. However, an obstacle to observing the proline isomeric configuration in a biological sample is the lack of an amide proton on these residues, which often impairs necessary chemical shift assignments near prolines through proton-detected NMR. In any case, the absence of an amide proton prevents the direct detection of signal from the proline peptide bond for the purpose of assigning the isomerization state, although readout of the proline C β and C γ chemical shifts via the adjacent residue is possible if the amide chemical shifts of that residue have been assigned. Using these methods, structural information can also be derived on the bound state of ligand prolines, as was reported for Nrd1 binding of *cis*-P $_6$ in native pS $_5$ CTD peptides (Kubicek et al., 2012). Recently, advances in ^{13}C NMR methodology and instrumentation have provided an opportunity to gain straightforward information on the backbone conformation of proteins, especially in the study of intrinsically disorder proteins (IDPs; Brutscher et al., 2015). One key aspect of ^{13}C NMR is that it is not limited by pH as in proton NMR. Further limiting the application of traditional proton NMR in studying IDPs, the lack of unique chemical environments and rapid conformational averaging cause the amide proton chemical shift dispersion for these systems to collapse to a roughly 1 ppm window, thus foiling higher resolutions typically achievable with modern spectrometers. Simultaneously, the mostly exposed backbone protons of these proteins often rapidly exchange with solvent which causes loss in signal. ^{13}C NMR improves resonance dispersion of IDPs and provides structural information of prolines which lack amide protons (Gibbs & Showalter, 2015), thus leading to many novel discoveries concerning IDPs (Gibbs et al., 2017; Lawrence & Showalter, 2012). ^{13}C NMR can also be used in combination with the other forms of NMR to obtain more comprehensive and clear assignment of resonances of a protein–ligand complex. A particularly applicable review covers strategies for overcoming the technical challenges of employing ^{13}C NMR for studying IDPs (Bastidas, Gibbs, Sahu, & Showalter, 2015).

For the investigation of RNAPII CTD, the application of ^{13}C NMR provides a very useful complement to the peptidomimic approaches described in Sections 2.1–2.3. In each heptad of the CTD, there are generally two critical SP motifs where reversible phosphorylation occurs during transcription, but the resonances associated with S $_2$ P $_3$ and S $_5$ P $_6$ are very difficult to distinguish using traditional protein NMR methodologies. By including ^{13}C NMR in the study of the CTD, information on the overall conformation as well as the relative population of *cis*- and *trans*- conformations for these prolines can be determined (Gibbs et al., 2017). The yeast and human CTD sequences are highly repetitive making it difficult to sequence-specifically assign residues, even when using ^{13}C NMR, but the *Drosophila* CTD has more divergent heptad sequences making it an ideal subject for NMR (Fig. 8). Intriguingly, developmentally complex organisms including all mammals and *Drosophila* have evolved multiple CTD heptads in which the S $_7$ position has been substituted with an asparagine residue. In the context of the *Drosophila* CTD sequence,

pS₅P₆N₇ heptads were found to possess significantly increased *cis*-P₆ populations in solution using ¹³C NMR (Gibbs et al., 2017). Formation of an additional intramolecular hydrogen bond between N₇ and T₄ in solution likely helps stabilize the *cis*-conformation (Fig. 9), which should make these heptads favorable substrates for Ssu72. Indeed, kinetic data using pS₅ CTD peptides with N at the seventh position showed that they are better substrates for Ssu72 dephosphorylation compared to consensus CTD peptides (Gibbs et al., 2017). Collectively, peptidomimic chemistry allows for unambiguous demonstration that Ssu72 requires *cis*-proline in its substrates, while ¹³C NMR clearly demonstrates that certain conserved sequence variations of the CTD heptad dramatically enhance the equilibrium population of pS₅-*cis*-P₆ available in solution. Altogether, the advances in ¹³C direct-detection NMR and its application on CTD structure and function support the role of this technique as an invaluable tool for studying IDPs.

4. SELECTIVE REGULATION OF PHOSPHATASE ACTIVITY BY PIN1 PROLYL ISOMERIZATION

The critical question regarding protein specificity for proline isomeric states of its substrate is how this special property can rewire down-stream signaling pathways. Differences in proline isomeric specificity of regulatory proteins can provide unique opportunities for differential regulation of signaling pathways. For proteins that can easily accommodate either proline conformation, the availability of *cis*- or *trans*-proline substrate should not significantly affect the whole pathway. On the other hand, if the protein only recognizes one isomeric proline substrate, it is likely that equilibrium interconversion between *cis*- and *trans*-proline is too slow to replenish the consumed isomeric form of substrate thus becoming the rate-limiting step of the biological pathway. This effect should be more prominent in proteins specific for *cis*-proline, since this conformation is the minority species in nature thus making it easier to be depleted. Under such a scenario, mechanisms that accelerate reequilibration of the *cis/trans* isomer ratio will increase the amount of apparent enzymatic activity. Pin1 in eukaryotes (or Ess1 in *S. cerevisiae*) catalyzes the conversion between *cis*- and *trans*-proline which can help more quickly equilibrate the population (Hanes, 2014; Lu et al., 2007). One example of how Pin1 can affect enzymatic function is when coincubated with Ssu72 there is a three-to fourfold increase in pS₅ CTD dephosphorylation compared to control reactions when the phosphate turnover was measured using the malachite green assay (Mayfield et al., 2015; Xiang et al., 2010). This increase in apparent Ssu72 activity is dependent on accelerated CTD proline isomerization since coincubation with an activity-deficient variant of Pin1 does not produce the same effect (Werner-Allen et al., 2011). In contrast, the activity of Scp1 and Fcp1 is not affected by Pin1 in vitro (Mayfield et al., 2015), consistent with their general activity against either *cis*- or *trans*-SP motifs.

To show whether the specific enhancement of Ssu72 activity by Pin1 is physiologically relevant, we tested if the phosphatase activity of Ssu72 is linked to Pin1 activity in cells (Mayfield et al., 2015). We performed lipofectamine (Thermo Fisher Scientific) transfections of HEK293T cells with a pLKO.1 plasmid containing shPin1 along with packaging and envelope plasmids to produce viral particles containing the shRNA coding

MGPD	malachite green phosphate detection method
NMR	nuclear magnetic resonance
pS₅	phosphorylated serine 5
Pin1	peptidyl-prolyl isomerase <i>cis-trans</i> isomerase NIMA interacting 1
PPIase	peptidyl-prolyl <i>cis/trans</i> isomerase
PTM	posttranslational modification
RNAPII CTD	RNA polymerase II C-terminal domain
SP motif	serine–proline motif

REFERENCES

- Bastidas M, Gibbs EB, Sahu D, & Showalter SA (2015). A primer for carbon-detected NMR applications to intrinsically disordered proteins in solution. *Concepts in Magnetic Resonance Part A*, 44(1), 54–66.
- Brutscher B, Felli IC, Gil-Caballero S, Hošek T, Kümmerle R, Piai A, et al. (2015). NMR methods for the study of intrinsically disordered proteins structure, dynamics, and interactions: General overview and practical guidelines. *Advances in Experimental Medicine & Biology*, 870, 49–122. [PubMed: 26387100]
- Buratowski S (2003). The CTD code. *Nature Structural Biology*, 10(9), 679–680. [PubMed: 12942140]
- Carter SG, & Karl DW (1985). Inorganic phosphate assay with malachite green: An improvement and evaluation. *Journal of Biochemical and Biophysical Methods*, 7(1), 7–13.
- Corden JL (2013). RNA polymerase II C-terminal domain: Tethering transcription to transcript and template. *Chemical Reviews*, 113(11), 8423–8455. [PubMed: 24040939]
- Dichtl B, Blank D, Ohnacker M, Friedlein A, Roeder D, Langen H, et al. (2002). A role for Ssu72 in balancing RNA polymerase II transcription elongation and termination. *Molecular Cell*, 10(5), 1139–1150. [PubMed: 12453421]
- Eick D, & Geyer M (2013). The RNA polymerase II carboxy-terminal domain (CTD) code. *Chemical Reviews*, 113(11), 8456–8490. [PubMed: 23952966]
- Ganem C, Devaux F, Torchet C, Jacq C, Quevillon-Cheruel S, Labesse G, et al. (2003). Ssu72 is a phosphatase essential for transcription termination of snoRNAs and specific mRNAs in yeast. *The EMBO Journal*, 22(7), 1588–1598. [PubMed: 12660165]
- Ghosh A, Shuman S, & Lima CD (2008). The structure of Fcp1, an essential RNA polymerase II CTD phosphatase. *Molecular Cell*, 32(4), 478–490. [PubMed: 19026779]
- Gibbs EB, Lu F, Portz B, Fisher MJ, Medellin BP, Laremore TN, et al. (2017). Phosphorylation induces sequence-specific conformational switches in the RNA polymerase II C-terminal domain. *Nature Communications*, 8(15233), 1–11.
- Gibbs EB, & Showalter SA (2015). Quantitative biophysical characterization of intrinsically disordered proteins. *Biochemistry*, 54(6), 1314–1326. [PubMed: 25631161]
- Hanes SD (2014). The Ess1 prolyl isomerase: Traffic cop of the RNA polymerase II transcription cycle. *Biochimica et Biophysica Acta*, 1839(4), 316–333. [PubMed: 24530645]
- Harlen KM, & Churchman LS (2017). The code and beyond: Transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nature Reviews. Molecular Cell Biology*, 18(4), 263–273. [PubMed: 28248323]
- Hart SA, Sabat M, & Etzkorn FA (1998). Enantio- and regioselective synthesis of a (Z)-alkene cis-proline mimic. *The Journal of Organic Chemistry*, 63(22), 7580–7581.

- Hassell AM, An G, Bledsoe RK, Bynum JM, Carter HL 3rd, Deng SJ, et al. (2007). Crystallization of protein-ligand complexes. *Acta Crystallographica. Section D, Biological Crystallography*, 63(Pt. 1), 72–79. [PubMed: 17164529]
- Hausmann S, & Shuman S (2002). Characterization of the CTD phosphatase Fcp1 from fission yeast preferential dephosphorylation of serine 2 versus serine 5. *The Journal of Biological Chemistry*, 277(24), 21213–21220. [PubMed: 11934898]
- Jeronimo C, Bataille AR, & Robert F (2013). The writers, readers, and functions of the RNA polymerase II C-terminal domain code. *Chemical Reviews*, 113(11), 8491–8522. [PubMed: 23837720]
- Kobor MS, Archambault J, Lester W, Holstege FC, Gileadi O, Jansma DB, et al. (1999). An unusual eukaryotic protein phosphatase required for transcription by RNA polymerase II and CTD dephosphorylation in *S. cerevisiae*. *Molecular Cell*, 4(1), 55–62. [PubMed: 10445027]
- Krishnamurthy S, He X, Reyes-Reyes M, Moore C, & Hampsey M (2004). Ssu72 is an RNA polymerase II CTD phosphatase. *Molecular Cell*, 14(3), 387–394. [PubMed: 15125841]
- Kubicek K, Cerna H, Holub P, Pasulka J, Hrossova D, Loehr F, et al. (2012). Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. *Genes & Development*, 26(17), 1891–1896. [PubMed: 22892239]
- Lawrence CW, & Showalter SA (2012). Carbon-detected 15N NMR spin relaxation of an intrinsically disordered protein: FCP1 dynamics unbound and in complex with RAP74. *Journal of Physical Chemistry Letters*, 3(10), 1409–1413. [PubMed: 26286791]
- Liou YC, Zhou XZ, & Lu KP (2011). Prolyl isomerase Pin1 as a molecular switch to determine the fate of phosphoproteins. *Trends in Biochemical Sciences*, 36(10), 501–514. [PubMed: 21852138]
- Lu KP, Finn G, Lee TH, & Nicholson LK (2007). Prolyl cis-trans isomerization as a molecular timer. *Nature Chemical Biology*, 3(10), 619–629. [PubMed: 17876319]
- Lu KP, Liou YC, & Zhou XZ (2002). Pinning down proline-directed phosphorylation signaling. *Trends in Cell Biology*, 12(4), 164–172. [PubMed: 11978535]
- Lu KP, & Zhou XZ (2007). Pinning down proline-directed phosphorylation signaling. *Trends in Cell Biology*, 12(4), 164–172.
- Luo Y, Yogesha SD, Cannon JR, Yan W, Ellington AD, Brodbelt JS, et al. (2013). Novel modifications on C-terminal domain of RNA polymerase II can fine-tune the phosphatase activity of Ssu72. *ACS Chemical Biology*, 8(9), 2042–2052. [PubMed: 23844594]
- Mayfield JE, Burkholder NT, & Zhang YJ (2016). Dephosphorylating eukaryotic RNA polymerase II. *Biochimica et Biophysica Acta*, 1864(4), 372–387. [PubMed: 26779935]
- Mayfield JE, Fan S, Wei S, Zhang M, Li B, Ellington AD, et al. (2015). Chemical tools to decipher regulation of phosphatases by proline isomerization on eukaryotic RNA polymerase II. *ACS Chemical Biology*, 10(10), 2405–2414. [PubMed: 26332362]
- Müller I (2017). Guidelines for the successful generation of protein-ligand complex crystals. *Acta crystallographica. Section D, Structural biology*, 73(Pt. 2), 79–92. [PubMed: 28177304]
- Otaka A, Katagiri F, Kinoshita T, Odagaki Y, Oishi S, Tamamura H, et al. (2002). Regio- and stereoselective synthesis of (E)-alkene trans-Xaa-pro dipeptide mimetics utilizing organocopper-mediated Anti-SN2' reactions. *The Journal of Organic Chemistry*, 67(17), 6152–6161. [PubMed: 12182656]
- Sasaki Y, Niida A, Tsuji T, Shigenaga A, Fujii N, & Otaka A (2006). Stereoselective synthesis of (Z)-alkene-containing proline dipeptide mimetics. *The Journal of Organic Chemistry*, 71(13), 4969–4979. [PubMed: 16776529]
- Schüller R, Forné I, Straub T, Schreieck A, Texier Y, Shah N, et al. (2014). Heptad-specific phosphorylation of RNA polymerase II CTD. *Molecular Cell*, 61(2), 305–314.
- Shaw PE (2002). Peptidyl-prolyl isomerases: A new twist to transcription. *EMBO Reports*, 3(6), 521–526. [PubMed: 12052773]
- Steinmetz EJ, & Brow DA (2003). Ssu72 protein mediates both poly(A)-coupled and poly(A)-independent termination of RNA polymerase II transcription. *Molecular and Cellular Biology*, 23(18), 6339–6349. [PubMed: 12944462]
- Wang XJ, Hart SA, Xu B, Mason MD, Goodell JR, & Etzkorn F (2003). Serine- cis-proline and serine-trans-proline isosteres: Stereoselective synthesis of (Z)- and (E)-alkene mimics by still-Wittig and

- Ireland-Claisen rearrangements. *The Journal of Organic Chemistry*, 68(6), 2343–2349. [PubMed: 12636401]
- Wedemeyer WJ, Welker E, & Scheraga HA (2002). Proline cis-trans isomerization and protein folding. *Biochemistry*, 41(50), 14637–14644. [PubMed: 12475212]
- Werner-Allen JW, Lee CJ, Liu P, Nicely NI, Wang S, Greenleaf AL, et al. (2011). cis-Proline-mediated Ser(P)5 dephosphorylation by the RNA polymerase II C-terminal domain phosphatase Ssu72. *The Journal of Biological Chemistry*, 286(7), 5717–5726. [PubMed: 21159777]
- Wu X, Wilcox CB, Devasahayam G, Hackett RL, Arevalo-Rodriguez M, Cardenas ME, et al. (2000). The Ess1 prolyl isomerase is linked to chromatin remodeling complexes and the general transcription machinery. *The EMBO Journal*, 19(14), 3727–3738. [PubMed: 10899126]
- Wulf G, Finn G, Suizu F, & Lu KP (2005). Phosphorylation-specific prolyl isomerization: Is there an underlying theme? *Nature Cell Biology*, 7(5), 435–441. [PubMed: 15867923]
- Xiang K, Nagaike T, Xiang S, Kilic T, Beh MM, Manley JL, et al. (2010). Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex. *Nature*, 467(7316), 729–733. [PubMed: 20861839]
- Yaffe MB, Schutkowski M, Shen M, Zhou XZ, Stukenberg PT, Rahfeld JU, et al. (1997). Sequence-specific and phosphorylation-dependent proline isomerization: A potential mitotic regulatory mechanism. *Science*, 278(5345), 1957–1960. [PubMed: 9395400]
- Yeo M, Lee SK, Lee B, Ruiz EC, Pfaff SL, & Gill GN (2005). Small CTD phosphatases function in silencing neuronal gene expression. *Science*, 307(5709), 596–600. [PubMed: 15681389]
- Yeo M, Lin PS, Dahmus ME, & Gill GN (2003). A novel RNA polymerase II C-terminal domain phosphatase that preferentially dephosphorylates serine 5. *The Journal of Biological Chemistry*, 278(28), 26078–26085. [PubMed: 12721286]
- Zhang Y, Kim Y, Genoud N, Gao J, Kelly JW, Pfaff SL, et al. (2006). Determinants for dephosphorylation of the RNA polymerase II C-terminal domain by Scp1. *Molecular Cell*, 24(5), 759–770. [PubMed: 17157258]
- Zhang M, Wang XJ, Chen X, Bowman ME, Luo Y, Noel JP, et al. (2012). Structural and kinetic analysis of prolyl-isomerization/phosphorylation cross-talk in the CTD code. *ACS Chemical Biology*, 7(8), 1462–1470. [PubMed: 22670809]
- Zhou XZ, & Lu KP (2016). The isomerase PIN1 controls numerous cancer-driving pathways and is a unique drug target. *Nature Reviews. Cancer*, 16(7), 463–478.

FURTHER READING

- Mayfield JE, Robinson MR, Cotham VC, Irani S, Matthews WL, Ram A, et al. (2017). Mapping the phosphorylation pattern of *Drosophila melanogaster* RNA polymerase II carboxyl-terminal domain using ultraviolet photodissociation mass spectrometry. *Biochimica et Biophysica Acta*, 1839(4), 316–333.

Fig. 1.

Isomers of dipeptide motifs. (A) General structures of *cis*- and *trans*-proline peptides. Isomerization occurs around the peptidyl-prolyl bond between proline and the preceding residue. The *trans*-proline conformation is more abundant than the *cis*-conformation (~5%–20%) in nature. (B) Core structures of the phosphorylated SP *cis*- and *trans*-isosteres. Installing a double bond between the serine and proline residues locks the isostere in the *cis*- or *trans*-conformation.

Fig. 2.

Crystal structures of Ssu72 and Scp1 bound to native pS₅-P₆ CTD peptides. (A) Ssu72 (*green*) bound to the native pS₅ CTD peptide (*gray*) with P₆ in the *cis*-conformation (PDB: 3O2Q). (B) Scp1 (*goldenrod*) bound to native pS₅ CTD peptide with P₆ in the *trans*-conformation (PDB: 2GHT). The Van der Waals exterior of each protein is represented as a transparent *gray* surface. The 2Fo-Fc maps representing electron density are contoured to 1 σ and illustrated as *blue meshes* around the CTD substrates.

Fig. 3.

Synthesis of the *S-trans*-P isostere via organocopper-mediated chemistry. A protected serine derivative (**1a**) is first reacted with cyclopentenyl-lithium (**2a**) to generate an SP-like precursor. Cyclization and protection of **3a** produces 54% of the *trans*-oxazolidinone derivative (**4a**), which can then be treated with an silyl-organocopper complex ($(i\text{-PrO})\text{Me}_2\text{SiCH}_2\text{Cu}(\text{CN})\text{MgCl} \cdot 2\text{LiCl}$) to form 98% of the anti- $\text{S}_{\text{N}}2'$ product (**5a**). Conversion of the silyl group to a carboxyl group results in the final *S-trans*-P isostere precursor (**7a**), which can then have the *tert*-butyl blocking group removed to form serine.

Fig. 4.

Synthesis of the *S-cis-P* isostere via Still-Wittig rearrangement. A protected serine derivative (**2b**) is reacted with 1-iodo-cyclopentene (**3b**) in the second synthesis step to generate an SP-like precursor. The Still-Wittig rearrangement of **6b** results in two major isomeric products: ~28% in the *trans*-conformation (**7b-trans**) and ~53% in the *cis*-conformation (**7b-cis**) which is used in the subsequent reaction. Product **9b** is the acid form of **8b** which is converted to the final *S-cis-P* isostere (**10b**).

Fig. 5. Synthesis of the *S-trans*-P isostere via Ireland-Claisen rearrangement. 1-Iodo-cyclopentene (**3b**) is used as a reactant in the first synthesis step to generate a proline-like precursor. In the third step of synthesis (**3c** $\rightarrow$ **4c**), *tert*-butyldimethylsilyloxyacetyl chloride is used to generate the Ireland-Claisen ester precursor. The Ireland-Claisen rearrangement of **4c** is carried out in two steps with activation of TMSCl followed by removal of the TBS group with *tert*-butylammonium fluoride (Bu_4NF) in order to generate ~52% of the *trans*-like precursor (**5c**). Subsequent oxidation and deblocking steps result in the final *S-trans*-P isostere (**7c**).

Fig. 6. Crystal structures of Ssu72-symplekin bound to pS₅-cis-P₆ peptidomimetic or native pS₅ peptide. (A) Overall structure of Ssu72 (*light blue*) complexed to symplekin (*purple*) and pS₅-cis-P₆ peptidomimetic (sand; PDB: 4YGX). (B) Structure of *cis*-locked peptidomimetic bound at the Ssu72 active site. The Van der Waals exterior of Ssu72 is represented as a transparent *gray* surface. The 2Fo-Fc map representing electron density is contoured to 1 σ and illustrated as a *blue mesh* around the CTD peptidomimetic. (C) Overlay of Ssu72 bound to pS₅-cis-P₆ peptidomimetic and Ssu72 bound to native pS₅ peptide (*green* and *gray*, respectively; PDB: 3O2Q). (D) Hydrophobic pocket of Ssu72 consisting of residues L₄₅, P₄₆, L₈₂, and M₈₅ (*orange*) that contribute specificity for *cis*-P₆ substrates.

Fig. 7. Crystal structures of Scp1 bound to *cis/trans*-locked pS₅ CTD peptidomimetics or native pS₅ peptide. (A) Structure of Scp1 D96N (*cerulean*) bound to pS₅-*trans*-P₆ peptidomimetic (*dark green*, PDB: 4YGY). (B) Overlay of Scp1 bound to pS₅-*trans*-P₆ peptidomimetic and Scp1 bound to native pS₅ peptide (*yellow* and *gray*, respectively; PDB: 2GHT). (C) Structure of Scp1 D96N (*violet*) bound to pS₅-*cis*-P₆ peptidomimetic (*sand*, PDB: 4YH1). The Van der Waals exterior of Scp1 is represented as a transparent *gray* surface. The 2Fo-Fc maps representing electron density are contoured to 1 σ and illustrated as *blue meshes* around the CTD peptidomimetics. (D) Overlay of the Scp1 structures bound to either *trans*- or *cis*-locked peptidomimetics. (E) Scp1-binding pocket interactions stabilizing complexation formation with peptidomimetic substrates. Residues F₁₀₆, V₁₁₈, I₁₂₀, and V₁₂₇ form a

hydrophobic pocket around CTD-P₃. Y₁₅₈ interacts with the side chain of CTD-T₄ while R₁₇₈ can form hydrogen bonds with the backbone carboxyl groups of CTD-S₂/T₄.

<i>H. sapiens</i>	<i>D. melanogaster</i>	<i>S. cerevisiae</i>
1 YSPTSPA	1 SPSVSPS	1 FGVSSPG
2 YEPSPGG	2 YSPTSPN	2 FSPTSPT
3 YTPQSPS	3 YTASSPGG	3 YSPTSPA
4 YSPTSPS	4 ASP---N	4 YSPTSPS
5 YSPTSPS	5 YSPSSPN	5 YSPTSPS
6 YSPTSPN	6 YSPTSPL	6 YSPTSPS
7 YSPTSPS	7 YA--SPR	7 YSPTSPS
8 YSPTSPS	8 YASTTPN	8 YSPTSPS
9 YSPTSPS	9 FNPQSTG	9 YSPTSPS
10 YSPTSPS	10 YSPSSSG	10 YSPTSPS
11 YSPTSPS	11 YSPTSPV	11 YSPTSPS
12 YSPTSPS	12 YSPTV-Q	12 YSPTSPS
13 YSPTSPS	13 FQ-SSPS	13 YSPTSPS
14 YSPTSPS	14 FAGSGSNI	14 YSPTSPS
15 YSPTSPS	15 YSPGN-A	15 YSPTSPS
16 YSPTSPS	16 YSPSSPN	16 YSPTSPS
17 YSPTSPS	17 YSPNSPS	17 YSPTSPS
18 YSPTSPA	18 YSPTSPS	18 YSPTSPA
19 YSPTSPS	19 YSPSSPS	19 YSPTSPS
20 YSPTSPS	20 YSPTSPC	20 YSPTSPS
21 YSPTSPS	21 YSPTSPS	21 YSPTSPS
22 YSPTSPN	22 YSPTSPN	22 YSPTSPS
23 YSPTSPN	23 YTPVTPS	23 YSPTSPN
24 YTPSPS	24 YSPTSPN	24 YSPTSPS
25 YSPTSPS	25 YS-ASPQ	25 YSPTSPG
26 YSPTSPN	26 YSPASPA	26 YSPGSPA
27 YTPSPN	27 YSQTGVK	27 YSPKQDEQKHNEENSR
28 YSPTSPS	28 YSPTSPT	
29 YSPTSPS	29 YSPSPS	
30 YSPTSPS	30 YDGSPPGSPQ	
31 YSPSSPR	31 YTPGSPQ	
32 YTPQSPT	32 YSPASPK	
33 YSPSSPS	33 YSPTSPL	
34 YSPSSPS	34 YSPSSPQ	
35 YSPTSPK	35 HSPSN-Q	
36 YTPSPS	36 YSPTGST	
37 YSPSSPE	37 YSATSPR	
38 YTPSPK	38 YSPNMSI	
39 YSPTSPK	39 YSPSSPK	
40 YSPTSPK	40 YSPTSPS	
41 YSPTSPT	41 YTPTARN	
42 YSPTTPK	42 YSPTSPM	
43 YSPTSPT	43 YSPTAPSH	
44 YSPTSPV	44 YSPTSPA	
45 YTPSPK	45 YSPSSPTFEESED	
46 YSPTSPT		
47 YSPTSPK		
48 YSPTSPT		
49 YSPTSPKGST		
50 YSPTSPG		
51 YSPTSPT		
52 YSLTSPAISPDSDDEEN		

Fig. 8.

Comparison of complete RNAPII CTD sequences from *Homo sapiens*, *Drosophila melanogaster*, and *Saccharomyces cerevisiae*. Each heptad within the CTDs has been numbered and aligned to highlight residue variations from the consensus Y₁S₂P₃T₄S₅P₆S₇ sequence. Residue variations were colored as follows: *orange* for aromatic residues (F, Y, and W), *green* for hydrophobic residues (e.g., A and V), *light blue* for polar uncharged residues (e.g., S and T), *purple* for basic residues (H, K, and R), and *red* for acidic residues (D and E).

Fig. 9.

Structural comparison of Ssu72 (*light green*) bound to native CTD peptide and model of S7N peptide. (A) The *cis*-conformation of P₆ in the native peptide (*white*) is stabilized by three intramolecular hydrogen bonds between T₄, S₇, and Y₁' (*apostrophe* denotes residue from the next heptad; PDB: 3O2Q). (B) Modeling N in place of S₇ in the native peptide (*yellow*) forms an additional hydrogen bond between the N₇ side chain and T₄ carboxyl group that may stabilize the *cis*-conformation of P₆. The Van der Waals exterior of Ssu72 is represented as a transparent *gray surface*.

Fig. 10.

Model of Pin1 effect on Ssu72-dependent dephosphorylation of RNAPII CTD. (1) Rate of isomerization of phosphorylated SP motifs with or without Pin1. The *thin arrows* without Pin1 represent the slow rate of normal proline isomerization, while the *bold arrows* represent the accelerated rate of isomerization with Pin1. (2A) In cells with low Pin1 activity, Ssu72 activity will become limited as it consumes its preferred pS₅-*cis*-P₆ substrate and the relatively slow rate of isomerization limits further dephosphorylation at this site. (2B) Therefore, disruption in Pin1 and subsequently Ssu72 causes transcriptionally active

RNAPII to remain hyperphosphorylated and read through termination sites. (3A) With normally functioning Pin1 catalyzing the *cis/trans* isomerization, Ssu72 rate will no longer be limited by a lack of pS₅-*cis*-P₆ substrate as *trans*-substrate can be converted to the *cis*-conformation in the appropriate time frame. (3B) Proper functioning of Pin1 and Ssu72 results in accurate transcriptional termination and thus better coordination of transcriptional cycling.

Table 1Strategies for Synthesizing *cis/trans* Proline Dipeptide Isosteres

Synthesis Strategy	Product	Number of Steps	Overall Yield (%)
Organocopper ^a	S- <i>trans</i> -P	8	17
Organocopper ^b	A- <i>cis</i> -P	>20	<11
Ireland–Claisen ^c	S- <i>trans</i> -P	8	13
Still–Wittig ^c	S- <i>cis</i> -P	10	20

^aOtaka et al., 2002.^bSasaki et al., 2006.^cWang et al., 2003.

Table 2

Synthetic Peptide Sequences Used in Malachite Green Kinetics and X-Ray Crystallography

Peptide	Sequence
Native pS ₅	Ac-S-P-Y-S-P-T-S(PO ₃ H ₂)-P-S-Y-S-NH ₂
<i>trans</i> -locked pS ₅	Ac-S-P-Y-S-P-T-S(PO ₃ H ₂) $\Psi[(E)C = CH]$ -P-S-Y-S-NH ₂
<i>cis</i> -locked pS ₅	Ac-S-P-Y-S-P-T-S(PO ₃ H ₂) $\Psi[(Z)C = CH]$ -P-S-Y-S-NH ₂
Native pS ₂	Ac-S-P-S-Y-S(PO ₃ H ₂)-P-T-S-P-S-Y-S-NH ₂
<i>trans</i> -locked pS ₂	Ac-S-P-S-Y-S(PO ₃ H ₂) $\Psi[(E)C = CH]$ -P-T-S-P-S-Y-S-NH ₂
<i>cis</i> -locked pS ₂	Ac-S-P-S-Y-S(PO ₃ H ₂) $\Psi[(Z)C = CH]$ -P-T-S-P-S-Y-S-NH ₂

The symbol *E* represents *trans* and the symbol *Z* represents *cis*-conformations. The symbol Ψ represents the psi bond of the SP-like isostere.