Published as: *Mol Cell*. 2006 December 8; 24(5): 759–770.

Determinants for Dephosphorylation of the RNA Polymerase II C-Terminal Domain by Scp1

Yan Zhang[1], **Youngjun Kim**[3,4,6], **Nicolas Genoud**[2], **Jianmin Gao**[7], **Jeffery W. Kelly**[7], **Samuel L. Pfaff**[2], **Gordon N. Gill**[5,6], **Jack E. Dixon**[3,4,6], and **Joseph P. Noel**[1,*]

[1]Howard Hughes Medical Institute Jack H. Skirball Center for Chemical Biology and Proteomics

[2]Gene Expression Laboratory The Salk Institute for Biological Studies La Jolla, California 92037

[3]Department of Pharmacology University of California, San Diego La Jolla, California 92093

[4]Department of Chemistry and Biochemistry University of California, San Diego La Jolla, California 92093

[5]Department of Medicine University of California, San Diego La Jolla, California 92093

[6]Department of Cellular and Molecular Medicine University of California, San Diego La Jolla, California 92093

[7]Department of Chemistry The Skaggs Institute for Chemical Biology The Scripps Research Institute La Jolla, California 92037

Summary

Phosphorylation and dephosphorylation of the C-terminal domain (CTD) of RNA polymerase II (Pol II) represent a critical regulatory checkpoint for transcription. Transcription initiation requires Fcp1/Scp1-mediated dephosphorylation of phospho-CTD. Fcp1 and Scp1 belong to a family of Mg^{2+}-dependent phosphoserine (P.Ser)/phosphothreonine (P.Thr)-specific phosphatases. We recently showed that Scp1 is an evolutionarily conserved regulator of neuronal gene silencing. Here, we present the X-ray crystal structures of a dominant-negative form of human Scp1 (D96N mutant) bound to mono- and diphosphorylated peptides encompassing the CTD heptad repeat ($Y_1S_2P_3T_4S_5P_6S_7$). Moreover, kinetic and thermodynamic analyses of Scp1-phospho-CTD peptide complexes support the structures determined. This combined structure-function analysis discloses the residues in Scp1 involved in CTD binding and its preferential dephosphorylation of $P.Ser_5$ of the CTD heptad repeat. Moreover, these results provide a template for the design of specific inhibitors of Scp1 for the study of neuronal stem cell development.

Introduction

Pol II is the key component in the transcription apparatus used for nascent mRNA synthesis and also mediates posttranscriptional regulation and modification of mRNA (Proudfoot et al., 2002). Through the CTD of the largest subunit of Pol II, a repertoire of processing and regulatory factors assemble on nascent mRNA (Dahmus, 1996; Meinhart et al., 2005).

*Correspondence: noel@salk.edu.

Supplemental Data Supplemental Data include Supplemental References, four figures, and one table and can be found with this article online at http://www.molecule.org/cgi/content/full/24/5/759/DC1/.

Accession Numbers Coordinates for human Scp1 in complex with mono- or doubly phosphorylated 9-mer CTD peptides have been deposited in the Protein Data Bank as 2ghq and 2ght, respectively.

Coordination of the temporal, structural, and mechanistic features of this ensemble depend on the timing of the transcription and cell cycles (Bregman et al., 2000).

Notably, the CTD of Pol II is unique as it is not found in RNA Pol I or III (Stiller and Hall, 2002). In contrast to the elaborate spatial and temporal regulatory functions governed by the CTD in cells, its primary composition is surprisingly simple. The CTD, found only in eukaryotes, consists of tandem repeated heptapeptide segments generally consisting of a consensus sequence, $Y_1S_2P_3T_4S_5P_6S_7$. Repeat number varies from 26 in budding yeast to 52 in humans (Corden, 1990). Although the CTD is simple in primary sequence, the presence of Pro residues capable of adopting *cis* or *trans* peptide-bond configurations (Morris et al., 1999) and the inclusion of Tyr, Thr, and Ser residues as possible kinase targets (Corden et al., 1985) encode enormous potential for physiochemical diversity of Pol II. Indeed, CTD phosphorylation is a major mechanism by which cells regulate gene expression (Palancade and Bensaude, 2003). The extent of Tyr, Thr, and Ser phosphorylation varies temporally throughout the life cycle of cells. For instance, Ser_2 and Ser_5 are major targets for proline-directed kinases (Zhang and Corden, 1991), whereas Tyr_1 and Thr_4 undergo phosphorylation to minimal extents (Baskaran et al., 1993). Pro_3 and Pro_6 can modulate the structure and accessibility of the CTD to kinases and phosphatases given their capacity to reside in *cis* and *trans* peptide-bond configurations. Significantly, the phosphorylation-specific peptidyl-prolyl *cis-trans* isomerase (PPIase) Pin1 regulates transcription through Pol II by altering the *cis-trans* ratio of Pro-containing peptide bonds in the CTD (Verdecia et al., 2000; Xu and Manley, 2004). Given the large number of possible phosphorylation and conformational states accessible to the CTD, a “CTD code” can act as a transcriptional rheostat to modulate mRNA production and processing (Buratowski, 2003).

Coordinately regulated phosphorylation and dephosphorylation of the CTD is not only essential for the recruitment and assembly of transcription complexes but also temporally controls transcription and mRNA processing during the cell cycle. Cell-cycle effects are largely mediated through the participation of proline-directed kinases such as cyclin-dependent kinases. Conversely, few CTD-specific phosphatases have been identified (Lin et al., 2002; Majello and Napolitano, 2001). Fcp1 is a conserved phosphatase found in all eukaryotes and is essential for cell survival in budding and fission yeasts (Archambault et al., 1997; Chambers and Dahmus, 1994). Specifically, the phosphatase activity of Fcp1 is essential for recycling Pol II (Cho et al., 1999).

Recently, another CTD-specific phosphatase belonging to the family of Fcp1-like enzymes known as Scp1 was characterized (Yeo et al., 2003) and shown to act as an evolutionarily conserved transcriptional corepressor for inhibiting neuronal gene transcription in nonneuronal cells (Yeo et al., 2005). Scp1 as well as closely related family members Scp2 and Scp3 is present in neuronal progenitor cells and nonneuronal cells and targets neuronal genes by interacting with the REST/NRSF complex (Yeo et al., 2005). Notably, blocking Scp1 function in P19 stem cells with a dominant-negative form of Scp1 (D96ED98N) capable of binding phospho-CTD, but not catalyzing dephosphorylation, derepressed neuronal gene expression and induced neuronal differentiation of P19 stem cells (Yeo et al., 2005). Given Scp1’s role in limiting inappropriate expression of neuronal-specific genes in pluripotent cells and its marked down-regulation as these cells initiate neuronal differentiation, Scp1 serves as a target for small molecule inhibitors.

One notable distinction between Fcp1 and Scp1 is their respective preferences for either $P.Ser_2$ or $P.Ser_5$ dephosphorylation. Fcp1 is reported to be 10-fold more active against $P.Ser_2$ than $P.Ser_5$ (Hausmann and Shuman, 2002), whereas Scp1 prefers $P.Ser_5$ as a substrate (Yeo et al., 2003). This catalytic selectivity likely plays a major role in regulating the timing of transcription and mRNA processing, as Ser_2 and Ser_5 phosphorylation are

temporally controlled by selective kinases acting at specific stages of transcription and cell-cycle progression (Ho and Shuman, 1999). For instance, Ser_5 phosphorylation plays a critical role in recruitment and regulation of mRNA capping factors, and the appearance of $P.Ser_5$ is considered a hallmark for transcript initiation (Cho et al., 1997; McCracken et al., 1997). During mRNA elongation, $P.Ser_5$ undergoes dephosphorylation coincident with the release of a 5′ processing factor, whereas phosphorylation of Ser_2 leads to recruitment of 3′ mRNA processing factors (Cho et al., 2001; Komarnitsky et al., 2000). Dephosphorylation of $P.Ser_2$ is also critical for the regeneration of the hypophosphorylated CTD necessary for recycling Pol II to initiate the next round of transcription (Cho et al., 1999). Although previous efforts have provided circumstantial evidence for site selectivity in $P.Ser_2$-$P.Ser_5$ dephosphorylation (Yeo et al., 2003), the determination of the specificity of dephosphorylation based on kinetic analyses and confirmatory structural characterization is necessary.

We report X-ray crystal structures of complexes of a substrate trapping human Scp1 mutant (D96N) bound to singly and doubly phosphorylated peptides spanning both a single and double CTD repeat. These structures complement in vitro kinetic characterization of Scp-mediated dephosphorylation of phospho-CTD peptides. Together, this structure-function analysis illuminates the specificity determinants of Scp1 for $P.Ser_5$-directed dephosphorylation, delineates a more complete picture of the dephosphorylation reaction catalyzed by the Fcp1-Scp1 family of CTD phosphatases, and provides a structural and kinetic starting point to design selective inhibitors of Scp1 and Fcp1.

Results and Discussion

Fcp1/Scp1 CTD Phosphatase Family

Fcp1 and Scp1 share the same catalytic domain architecture, with identical Mg^{2+} binding DXDX(T/V) motifs. This divalent metal-ion coordination site corresponds to the signature sequence found in a family of structurally related metal-dependent phosphohydrolases and haloacid dehalogenases (Allen and Dunaway-Mariano, 2004). To date, eight putative CTD phosphatases sharing the Fcp1/Scp1 DXDX(T/V) signature sequence have been identified in the human genome (Figure 1A). Fcp1 is the largest of the Fcp1/Scp1 family members, with a conserved catalytic domain that includes a breast cancer protein-related carboxy-terminal (BRCT) domain and a C -terminal region that binds regulatory TFIIF (Figure 1B). Other putative CTD phosphatases, except the ubiquitin-like domain containing CTD phosphatase (UBLCP, MGC10067) (Zheng et al., 2005), contain only the conserved catalytic domain (Figure 1B). Antisensemediated reduction of Dullard, another Scp-like phosphatase, in frogs, leads to an abnormal developmental phenotype (Satow et al., 2002). TIMM 50 plays an important role in the transport of proteins into the mitochondria (Yamamoto et al., 2002).

Crystallization and Structure Determination of Human Scp1 Bound to Phospho-CTD Peptides

A truncated version of Scp1, used in crystallization experiments, improves the solubility and limits the flexibility of the protein (Kamenski et al., 2004). In order to trap a complex of Scp1 with phospho-CTD peptide substrates, the catalytically essential Asp96, used as an attacking nucleophile for phosphoryl transfer, was mutated to Asn (D96N) to create a dominant-negative Scp1. The D96N Scp1 crystals used for these studies belong to a different space group (monoclinic C2) than the citrate-bound Scp1 crystals (trigonal $P3_121$) previously reported (Kamenski et al., 2004).

Phosphorylated CTD peptides were soaked into D96N Scp1 crystals (Figure 1C). We initially used 9-mer phosphopeptides encompassing a single heptad repeat of the CTD and

phosphorylated on Ser_2, Ser_5, or both (Figure 1C). The structure of D96N Scp1 in complex with a monophosphorylated $P.Ser_2$ peptide shows the presence of a low-occupancy phosphate group in the Mg^{2+}-bound active site, whereas other parts of the CTD peptide are not visible in electron-density maps. This observation is consistent with previous biochemical studies that demonstrated that Scp1 prefers to dephosphorylate $P.Ser_5$ but will slowly dephosphorylate $P.Ser_2$ (Yeo et al., 2003).

In contrast, two additional phosphopeptides containing $P.Ser_5$ or $P.Ser_2/P.Ser_5$ and soaked into D96N Scp1 crystals exhibit clear electron density, permitting modeling of all residues of the respective phosphopeptides (Figures 2A and 2B). The structures were refined to 2.05Å for the complex with the doubly phosphorylated CTD 9-mer peptide and to 1.8Å for the singly phosphorylated $P.Ser_5$ CTD peptide (Table 1). As will be described in more detail later, the two phosphopeptides possessing $P.Ser_5$ bind to Scp1 in an identical manner. However, the phosphate group on $P.Ser_2$ participates in no obvious intermolecular interactions with Scp1 and instead extends outward from the peptide binding groove (Figure 2A).

Longer peptides encompassing two CTD heptad repeats with different arrangements of $P.Ser_2/P.Ser_5$ sites were used to kinetically and structurally probe possible secondary phosphate binding sites residing outside of the phosphopeptide binding groove shown in Figure 2. Five different 14-mer doubly phosphorylated CTD peptides were used that span the $P.Ser_2$ and $P.Ser_5$ sites in various combinations (Figure 1C). Although these phosphopeptides are relatively long for crystal soaking, crystal structures (2.1–2.6Å) were obtained with clear electron density for the peptides (Table 1 and Table S1 in the Supplemental Data available with this article online). The doubly phosphorylated ($P.Ser_5$-$P.Ser_5$) CTD 14-mer peptide (Figure 1C, 14-mer peptide 4) binds with high occupancy to D96N Scp1 (Figure 2D) (Table 1). Notably, the resultant intermolecular arrangement mirrors the core Scp1 interactions first observed in the 9-mer phosphopeptide complexes possessing $P.Ser_5$.

The statistics for the remaining complexes with the other doubly phosphorylated 14-mers (Figure 1C) are summarized in Table S1. Again, for 14-mers possessing at least one $P.Ser_5$ residue, the initial electron-density maps recapitulate the maps calculated for the 9-mer $P.Ser_5$ complex (Figure 2B) and the 14-mer $P.Ser_5$-$P.Ser_5$ complex (Figure 2D). In contrast, the doubly phosphorylated $P.Ser_2$ 14-mer (Figure 1C, 14-mer peptide 8) exhibited weak electron density first observed for the $P.Ser_2$ 9-mer phosphopeptide. For all the well-defined phosphopeptide complexes solved, which contain at least one $P.Ser_5$ group (Figure 1C, 14-mer peptides 4–7, Table 1, and Table S1), the same binding mode encompassing Ser_2, Pro_3, Thr_4, and $P.Ser_5$ localizes in an ordered Scp1 binding groove, whereas the N- and C-terminal residues extend outward from the Scp1 active site and exhibit considerable disorder (Figure 2D). These results suggest that no flanking P.Ser interacting sites exist at least within the vicinity of the CTD binding groove and extending to the insertion domain.

Dephosphorylation Reaction

Although the D96N Scp1-phospho-CTD complexes described here were obtained by using strikingly different crystallization conditions from the previously reported X-ray crystal structures of Scp1 (Kamenski et al., 2004), the overall architecture of Scp1 is identical to the previously reported structures. The global architecture of Scp1 resembles several DXDX family proteins with limited sequence homology (Figures 3A and 3B). For example, β-phosphoglucomutase shares 9% sequence identity with Scp1, but the active-site residues involved in metal coordination are conserved both in primary sequence and in tertiary structures (Figure 3C). Structural conservation extends to the solvent molecules occupying coordination sites around the catalytically essential Mg^{2+} ion (Figure 3C). Notably, the CTD

peptides observed here adopt a unique binding conformation that encompasses a β turn (Figure 2A), distinct from the extended binding for CTD peptides observed in another doubly phosphorylated CTD peptide bound to human Pin1 (Verdecia et al., 2000)(Figure S1A).

Fcp1 and Scp1 share 21% sequence identity in their catalytic domains. The active sites, including the conserved DXDX(T/V) motifs used for Mg^{2+} coordination and the catalytic residues involved in phosphoryl-transfer, are almost identical (Figure 2C). The catalytic pocket of Scp1 exhibits a negatively charged surface owing to the cluster of conserved Asp residues used for Mg^{2+} coordination and catalysis (Figure 3D). As shown previously for structures of Scp1 complexed with BeF_3^- (Kamenski et al., 2004), residues coordinating Mg^{2+} in Scp1 are conserved with Fcp1 and are essential for the dephosphorylation reaction. Hausmann et al. conducted a mutagenesis screen in order to determine which residues contribute to dephosphorylation and identified 11 residues essential for phosphatase activity (Hausmann and Shuman, 2003). Seven of the residues include the Mg^{2+} coordination sphere or interact with residues contacting Mg^{2+} (Figure 3D).

Dephosphorylation catalyzed by Scp1 shares mechanistic features with other DXDX family members, including β-phosphoglucomutase (Lahiri et al., 2003) (Figures 3A and 3C) and phosphoserine phosphatase (Wang et al., 2002). The reaction encompasses a two-step mechanism comprising phosphoryl transfer to Asp96 followed by water-mediated hydrolysis of the covalent enzyme intermediate (Figure 4A). The identity of the general base and nucleophilic water leading to hydrolysis of the phosphoryl-enzyme intermediate is unclear.

Two ordered waters present in both copies of Scp1 found in the crystallographic asymmetric unit coordinate Mg^{2+} (Figure 3D). Using the D96N Scp1 dominant-negative mutant, we trapped the P.Ser_5 phosphate coordinated to Mg^{2+}. In turn, the mutant Asn96 side chain forms a hydrogen bond with the P.Ser_5 phosphate in a manner resembling the approach to the transition state for the phosphoryl transfer. This picture of the catalytic complex implicates the water labeled H_2O-1 in Figure 3D as the nucleophile necessary for hydrolysis of the phosphoryl-enzyme intermediate. Moreover, this arrangement posits that Asp206 will serve as the general base to activate H_2O-1 (Figure 4A). In support, replacement of the residue equivalent to Asp206 (Asp297 in *S. pombe* Fcp1) by Ala abolished Fcp1 phosphatase activity (Hausmann and Shuman, 2003), as did mutation of Asp206 in Scp1 (M. Yeo and G.N.G., unpublished data).

Moreover, Lys190 (Lys280 in *S. pombe* Fcp1) was also identified in the original mutagenesis screen of Fcp1 as important for phosphatase activity. Notably, the side chain amino group of Lys190 forms a salt bridge with the side chain carboxyl moiety of Asp206 to position the carboxyl group next to the putative nucleophilic water molecule (Kamenski et al., 2004). Conversely, Wang et al. showed that the residue equivalent to Asp98 may function as the general base for this final catalytic step in phosphoserine phosphatase (Wang et al., 2002). The second coordinating water molecule mentioned above (H_2O-2 in Figure 3D) could be responsible for the final water-mediated hydrolysis of the phosphoryl-enzyme intermediate. In this case, the most suitable candidate to act as a general base for water activation is Glu 99 in human Scp1, whose carboxylate resides 2.6Å from H_2O-2. However, Glu 99 is not conserved across the Fcp1-Scp1family, i.e., Gln in yeast and human Fcp1. As shown in Figures 3D and 4A, the side-chain carboxylate of Asp98 or Asp206 more likely serves to activate the putative nucleophilic water necessary for hydrolysis of the phosphoryl-enzyme intermediate (Asp96 phosphate).

Phospho-CTD Recognition by Scp1

The Scp1 active site forms a groove into which a portion (Ser_2-Pro_3-Thr_4-Ser_5) of the CTD heptad repeat binds (Figure 2C). The primary contacts between the phospho-CTD peptides and Scp1 involve Mg^{2+} coordination of the $P.Ser_5$ phosphate and sequestration of the Pro_3 ring in a hydrophobic pocket formed by Phe106, Val118, Ile120, Val127, and Leu155 (Figure 4B). This arrangement provides a satisfying explanation for the observed $P.Ser_5$ specificity of Scp1 previously reported (Yeo et al., 2003). Moreover, the limited contact surface of the phospho-CTD peptide with Scp1 combined with the specificity of the interactions suggests that this recognition surface will serve as a starting point for structure-based design of Scp1-specific inhibitors. This limited recognition surface contrasts sharply with the extended interface observed between Cgt1 and a CTD peptide (Fabrega et al., 2003) (Figure S1B). The arrangement of Pro_3 in the hydrophobic pocket of Scp1 is similar to the interface used by another phospho-CTD binding protein, namely Pin1, in which Pro_6 inserts into an aromatic-rich hydrophobic pocket (Verdecia et al., 2000)(Figure S1A). However, unlike Pin1, which sequesters the Pro residue immediately following the P.Ser site, Scp1 recognizes the Pro residue two residues N-terminal to $P.Ser_5$. Conversely, Pro_6, next to $P.Ser_5$, serves no role in Scp1-phospho-CTD recognition.

Thr_4 of the Scp1-bound phospho-CTD peptides shows no direct contact with Scp1, but its backbone is involved in van der Waals interactions with the Scp1-phospho-CTD binding groove (Figure 4C). This interaction explains the role of Tyr158 (Tyr242 in human Fcp1 and Tyr249 in *S. pombe* Fcp1) in Fcp1 phosphatase activity (Hausmann and Shuman, 2003). However, it was not clear from the apo Scp1 structure previously reported why this Tyr might be important catalytically, as Tyr158 resided outside of the immediate active site defined centrally by the Mg^{2+} ion (Kamenski et al., 2004). Inspection of the current phospho-CTD peptide complexes makes it clear that the side chain of Tyr158 forms an intimate van der Waals interaction surface with the backbone atoms surrounding Thr_4 of the CTD heptad repeat. Moreover, the side chain of Tyr158 is anchored by a hydrogen bond with the active site side chain of Asp98 (Figure 4C). Another residue, Arg178 (Arg271 in *S. pombe* Fcp1), is conserved in the Fcp1-Scp1 family, whereas previous replacement of Arg178 by Ala unexpectedly had no effect on the intrinsic phosphatase activity against nonphysiological para-nitrophenyl phosphate (pNPP) but alters activity against physiological targets (Hausmann and Shuman, 2003). Although Arg178 provides no direct contact to the seryl-phosphate, the side chain guanido forms a hydrogen-bond network with the backbone carbonyls of Ser_2 and Thr_4, possibly playing a critical role in binding phospho-CTD (Figure 4C).

In the doubly phosphorylated CTD peptide-Scp1 complex, $P.Ser_2$ faces out of the active site, making no direct contact with Scp1, and exhibits weak electron density. Indeed, the two molecules in the asymmetric unit exhibit different conformations for the $P.Ser_2$ phosphate. Of the two Scp1 polypeptide chains in each asymmetric unit, only one, chain B, exhibits traceable electron density for Tyr_1 of the phospho-CTD peptides. The location of Tyr_1 in this Scp1 complex most likely does not reflect a physiologically relevant conformation, as the neighboring Scp1 chain in the crystal lattice appears to limit the rotational freedom of the Tyr_1 residue. In contrast, Tyr_1 has been implicated to play an essential role in phospho-CTD recognition by Fcp1, with a 6-fold reduction in binding affinity for CTD peptides when mutated to Ala (Hausmann et al., 2004). This result can be explained by the preferential $P.Ser_2$ dephosphorylation activity exhibited by Fcp1. Given the proximity of Tyr_1 to $P.Ser_2$ in an Fcp1-phospho-CTD complex, Tyr_1 would be predicted to reside at the entrance to, or wholly within, the CTD binding groove in Fcp1 overlapping with the Pro_6 or Thr_4 binding site observed in Scp1.

Xaa-Pro peptide bonds can exist in energetically favorable *cis* or *trans* configurations with a thermodynamic preference for the *trans* regioisomer (Brandts and Lin, 1986). The P.Ser-Pro peptide bonds for both $P.Ser_2$ and $P.Ser_5$ observed in these Scp1-phospho-CTD complexes exist exclusively in *trans* (omega bond = 180°) configurations. The central residues of the phospho-CTD peptides used in these current studies (Pro_3-Thr_4-$P.Ser_5$) form an interaction surface with Scp1 (Figure 2C). Displacement of these residues due to rotation around either P.Ser-Pro peptide bond would likely disrupt this interaction surface, leading to lowered binding energies for the *cis*-containing forms of the phospho-CTD. Moreover, if the positions of these core binding residues are maintained while either prolyl peptide bond at $P.Ser_2$-Pro_3 or $P.Ser_5$-Pro_6 adopts a cis configuration, steric clashes with Scp1 or with the phospho-CTD itself would result (Figures S2A and S2B).

This model for preferential recognition of all *trans*-containing phospho-CTD heptad repeats by Scp1 hints at additional mechanisms underlying the regulation of CTD phosphorylation. Notably, another phospho-CTD substrate, Pin1, catalyzes isomerization of the *cis-trans* prolyl-peptide bond in a phosphorylation-dependent manner (Ranganathan et al., 1997; Yaffe et al., 1997). As suggested previously (Verdecia et al., 2000), Pin1 can modulate the structure and function of Pol II, possibly by changing the balance of *cis-trans* peptide bonds and, by logical extension, availability of substrate for *trans*-specific CTD phosphatases.

Mono- and Diphosphorylated CTD Peptides

The crystallographically observed modes of phospho-CTD peptide recognition by Scp1 superimpose in nearly identical fashions, suggesting that the phosphorylation state of Ser_2 is not important for Scp1 recognition. However, one unique intramolecular interaction observed in the CTD peptide itself lies between the side chain hydroxyl of Thr_4 and the side chain hydroxyl of Ser_2 (Figures 2B and 4C) in the complex with $P.Ser_5$-monophosphorylated CTD peptide. This intramolecular hydrogen bond stabilizes the β turn conformation of the CTD peptide. Notably, this conformation may form in the absence of Scp1 or without the need for Ser_5 phosphorylation, as analysis of the CTD peptide conformation by NMR found that a turn encompassing Ser_2-Pro_3-Thr_4-Ser_5 is more stable than one spanning Ser_5-Pro_6-Ser_7-Tyr_1 (Kumaki et al., 2001). In the doubly phosphorylated CTD 9-mer peptide complex, the hydroxyl of Thr_4 is also within hydrogen bonding distance of the $P.Ser_2$ side chain. However, the unreliability of the seryl-phosphate position in this latter complex makes it difficult to ascertain whether phosphorylation of Ser_2 can modify the preferred conformation of the CTD in solution. Certainly, electrostatic repulsion of the $P.Ser_2$ and $P.Ser_5$ phosphates is likely to lead to a more extended CTD conformation that may weaken its interaction with Scp1 through breakage of this side chain-side chain turn.

Similar intramolecular hydrogen bonds of CTD peptides are found in complexes of CTD peptides with the CID domain of the transcription capping enzyme Pcf11 (Figure S1C) (Meinhart and Cramer, 2004; Noble et al., 2005). However, in this case, the hydrogen bond forms between the $P.Ser_2$ and Thr_4 side chains (Meinhart and Cramer, 2004). The authors speculate that the CTD domain forms a β spiral stabilized by a series of side-chain hydrogen bonds linking $P.Ser_2$ and Thr_4. Our current structure of the CTD phosphatase Scp1 bound to phospho-CTD peptides supports the notion that intramolecular hydrogen bonds may serve as a conformational regulatory mechanism in modulating CTD function. Preformed hydrogen bonds between either phosphorylated or nonphosphorylated Ser_2 may regulate the accessibility of Ser_5 and $P.Ser_5$ to kinases and phosphatases, respectively, during transcription. Moreover, the intramolecular hydrogen bond can only form when the Ser_2-Pro_3 peptide bond resides in the *trans* configuration, possibly linking PPIases like Pin1 to additional regulatory roles.

Kinetic Basis for the Preferential Dephosphorylation of P.Ser$_5$

The preferences of Fcp1 and Scp1 for P.Ser$_2$ or P.Ser$_5$, respectively, may play a pivotal role in the temporal and conformational regulation of the CTD. The residues lining the bottom part of the CTD backbone binding groove in Scp1 are conserved in Fcp1, suggesting conservation of the Fcp1/Scp1 recognition of the CTD polypeptide backbone. The specificity of the interaction and thus the register of the CTD segment bound to either Fcp1 or Scp1 are most likely achieved by recognition of key CTD residues directed at the cavities accessible in the CTD binding groove (Figure 2C).

In order to define the preferred P.Ser site turned over by Scp1, we used synthetic CTD phospho-peptides for the steady-state kinetic analysis of Scp1 using in vitro phosphatase assays. First, we used 9-mer CTD peptides singly phosphorylated at Ser$_2$ or Ser$_5$. Scp1 catalyzed phosphate loss on these peptides in a pH-dependent manner, with maximum activity at pH 5.5. Kinetic analyses defined a K_M value of 0.287 mM and an apparent turnover number, k_{cat}, of 2.88 s^{-1} for the P.Ser$_5$ 9-mer. In contrast, Scp1 activity for P.Ser$_2$ was beyond the limits of quantitative determination under our in vitro assay conditions (K_M > 5 mM). These results agree with affinity measurements of the substrate-trapping (D96N) dominant-negative Scp1 mutant. Using 9-mer peptides labeled at the N terminus with the tetramethylrhodamine fluorophore, measurements of the change in fluorescence anisotropy (Verdecia et al., 2000) of the respective peptides resulted in dissociation constants, K_Ds, of ~0.3 mM for the singly or doubly phosphorylated P.Ser$_5$ or P.Ser$_2$/P.Ser$_5$ 9-mers, respectively (Figures S3A and S3B). In contrast, the K_D for the singly phosphorylated P.Ser$_2$ 9-mer was beyond the limit of detection used in these in vitro binding assays. This suggests the binding energy and register of the phospho-CTD segment recognized by Scp1 are set by interactions flanking the core P.Ser recognition element coordinating the catalytically essential Mg^{2+} ion.

Second, we synthesized two phosphorylated peptides, each of which contains a diheptad repeat and a monophosphorylated serine residue, namely $Y_{1a}S_{2a}P_{3a}T_{4a}S_{5a}P_{6a}S_{7a}Y_{1b}S_{2b}P_{3b}T_{4b}(P.S_{5b})P_{6b}S_{7b}$ (P.Ser$_5$-14-mer) and $Y_{1a}S_{2a}P_{3a}T_{4a}S_{5a}P_{6a}S_{7a}Y_{1b}(P.S_{2b})P_{3b}T_{4b}S_{5b}$-$P_{6b}S_{7b}$ (P.Ser$_2$-14-mer). We then measured the steady-state kinetics of Scp1 with these peptides. Scp1 exhibited a K_M of 0.21 mM and an apparent k_{cat} of 2.44 s^{-1} for the P.Ser$_5$-14-mer and a K_M of 5.6 mM and a k_{cat} of 0.935 s^{-1} for the P.Ser$_2$-14-mer (Figure S4). From these constants (Table 2), we calculated the specificity constant (k_{cat}/K_M) resulting in values of 11.4 $s^{-1}mM^{-1}$ for the P.Ser$_5$-14-mer and 0.166 $s^{-1}mM^{-1}$ for the P.Ser$_2$-14-mer (Figure S4). Thus, Scp1 exhibits significantly higher kinetic specificity (67-fold) for dephosphorylation of P.Ser$_5$ than for P.Ser$_2$,. Because k_{cat} for the P.Ser$_5$-14-mer is only 2.6-fold higher than for the P.Ser$_2$-14-mer, the kinetic specificity of Scp1 for dephophorylation of P.Ser$_5$ is largely determined by the specific binding affinity of Scp1 for the polypeptide segment of phospho-CTD encompassing P.Ser$_5$.

Third, we measured steady-state kinetics by using doubly phosphorylated 14-mer CTD peptides (Figure 1C). The data using combinations of doubly phosphorylated serine sites demonstrate that the determinant for the kinetic efficiency of dephosphorylation is the number of P.Ser$_5$ sites present (Table 2). These results are consistent with our crystallographic structures showing that the Ser$_2$-Pro$_3$-Thr$_4$-P.Ser$_5$ segment is key to the Scp1 interaction, whereas additional residues present on the N- and C- termini provide little to Scp1 binding. Phosphopeptides containing multiple CTD repeats will likely exhibit even greater catalytic efficiency due to higher local concentrations of P.Ser$_5$ and possible catalytic processivity.

The X-ray crystal structure of Scp1 in complex with a phosphorylated CTD peptide enabled identification of several Scp1 residues involved in phospho-CTD recognition. Among these

residues, Phe106 appears to provide a significant amount of binding energy for site selectivity, as its phenyl ring stacks against the prolyl ring of Pro_3 (Figure 2C). Given this favorable van der Waals interaction of Pro_3 with Phe106 in the Scp1 CTD binding groove, we examined a F106A Scp1 mutant and measured the kinetic constants for F106A Scp1 turnover (Table 2). The K_M and k_{cat} values using pNPP as a substrate were comparable to previous published values (Kamenski et al., 2004), suggesting that this mutation does not alter the structure and intrinsic catalytic potential of Scp1.

Using the F106A mutant, we measured the kinetic parameters (Table 2) for the P.Ser_5-14-mer and P.Ser_2-14-mer peptides. Interestingly, F106A exhibited poorer K_M values (6-fold poorer than wild-type Scp1) for the P.Ser_5-14-mer but more favorable K_M values (5-fold better than wild-type Scp1) for the P.Ser_2-14-mer. In both cases, F106A Scp1 exhibited similar turnover numbers. The preference of F106A for the P.Ser_5-14-mer versus the P.Ser_2-14-mer is 4.4 compared to a value of 67 for wild-type Scp1, suggesting that Phe106 is an important determinant for the specific selection of P.Ser_5 for Scp1-mediated dephosphorylation.

Fcp1, which shares a high percentage of sequence identity/conservation in the active site and along the CTD binding groove involved in backbone interactions, possesses a divergent and polar cavity off the main CTD binding groove involved in Pro_3 sequestration in human Scp1 (Figure 4C). Notably, Phe106 exhibits variability in Fcp1 sequences (Val in *S. pombe* Fcp1 and Glu in human Fcp1). Ile 120 is conservatively replaced by a Leu, and Leu155 is replaced by a smaller polar Thr residue (Figure 4B), with all substitution circumstantially supporting the regiospecificity of dephosphorylation in the Fcp1/Scp1 family dictated by the physiochemical nature of the Pro_3 binding cavity.

Comparison of CTD-Binding Modules

A large fraction of CTD-binding proteins exploit a CTD-binding module minimally recognizing a portion of a CTD heptad repeat to sequester Pol II. These binding modules include the WW domain of Pin1 (Verdecia et al., 2000), the CID domain of Pcf11 (Meinhart and Cramer, 2004), and most recently, the SRI domain from Set2 (Li et al., 2005; Vojnic et al., 2006). At first glance, Scp1 does not appear to contain a stably folded domain that specifically binds to the CTD. However, when the apo Scp1 structure was determined, the authors noticed an “insertion domain” that is found only in a subset of the larger family of Fcp1/Scp1 phosphatases (Figure 1B; βID1-3 in Figure 2A). As we were investigating the Scp1 structure, we noticed the striking similarity between this insertion domain of human Scp1 and the WW domain of Pin1 (rmsd of 1.05Å for structure-based alignment of the main chain atoms), even though they share no notable sequence identity (Figure 5A).

The structural similarity between the Pin1 WW domain and the Scp1 insertion domain together with a citrate molecule present in the previously reported apo Scp1 structure (Kamenski et al., 2004) prompted us to consider the possibility that the position of citrate observed in a previous Scp1 crystal form represents a flanking and noncatalytic second phosphate binding site. However, use of longer CTD peptides containing multiple phosphate groups (Figure 1C) showed no such secondary binding site. Moreover, through superimpositioning of our Scp1-CTD complex structures reported here and the apo Scp1 structure previously reported (Kamenski et al., 2004), we found that the location of the citrate molecule overlaps with the polypeptide backbone interactions spanning Pro_3-Thr_4 (Figure 5B). In fact, there is a paucity of interactions between the citrate group and Scp1 in the previously reported apo structure (Kamenski et al., 2004). Finally, we obtained apo Scp1 crystals by using different conditions from Kamenski et al. (2004), i.e., ammonium sulfate (2.2Å) and polyethylene glycol (PEG) 8000 (1.8Å) (Table S1) as precipitants. None of the structures obtained in the absence of phosphopeptides show a negatively charged group in

the citrate-binding site even when multivalent anions such as sulfate are present at molar concentrations. Therefore, we believe the citrate molecule observed previously may not be representative of a physiologically relevant interaction.

Experimental Procedures

Purification and Mutagenesis

An *E. coli* plasmid containing the human Scp1 gene spanning residues 77–256 was subcloned into the *E. coli* expression vector pHIS8 (Jez et al., 2000). The protein was expressed and purified by using a protocol previously described (Kamenski et al., 2004), resulting in 35 mg of Scp1 (>98% purity based upon SDS-PAGE analysis and Coomassie staining) per liter of *E. coli* culture. The D96N and F106A mutants were prepared with the QuikChange Site-Directed Mutagenesis Kit (Stratagene, CA), expressed, and purified as described above.

Crystallization and Data Collection

The CTD phosphopeptides were obtained commercially (Anaspec, San Jose CA). Scp1 D96N was concentrated to 20 mg/ml in 100 mM NaCl, 10 mM HEPES-Na^+ (pH 7.5), and 1 mM DTT. Crystals were grown by using several different conditions, including (1) 0.5 M ammonium sulfate containing 0.2 MLi_2SO_4, 100 mM HEPES-Na^+ (pH 7.5) and (2) 1 M sodium tartrate, 100 mM Tris-HCl (pH 8.5). We reasoned that a high-salt environment may negatively affect the binding of phospho-peptides soaked into crystals; therefore, crystals were transferred to a mother liquor consisting of 30% (w/v) PEG 4000, 0.2 M $MgCl_2$, 35 mM CTD peptide, and 50 mM Tris-HCl (pH 8.0). The crystals were then transferred into 2 μl of a cryoprotection solution consisting of mother liquor containing 30% (v/v) glycerol after 48 hr of equilibration in the CTD peptide binding mother liquor. After freezing in liquid nitrogen, data were collected at a wavelength of 0.9537Å at 100 K on beam line 8.2.2 of the Advanced Light Source (ALS). Data were processed with HKL2000 (Otwinowski and Minor, 1997) and are summarized in Table 1.

Structure Solution and Refinement

The structures of human Scp1 in complex with CTD peptides were determined by molecular replacement (MR) using the apo Scp1 structure as a search model (PDB code 1ta0) with the program AmoRe (Navaza, 1994) available in the CCP4 software package (CCP4, 1994). MR solutions were refined first by using CNS (Brunger et al., 1998), reserving 5% of the measured and reduced structure factor amplitudes as an unbiased test set for cross validation (R_{free}) (Brunger, 1992). SigmaA-weighted electron-density maps ($2F_o - F_c$ and $F_o - F_c$) were calculated after each cycle of refinement for model building by using O (Jones et al., 1991). The locations of the peptides were unequivocal in $F_o - F_c$ maps even after the first round of refinement. The peptide models were built into the electron density by using O (Jones et al., 1991). The final models were evaluated by PROCHECK (Laskowski et al., 1993). For the Scp1 complex with the doubly phosphorylated CTD peptide, 87.2% of the residues reside in the most favored regions of the Ramachandran plot and the remaining 12.2% are found in additionally allowed areas with 0.6% in generally allowed areas. For the Scp1 complex with the monophosphorylated P.Ser_5 CTD peptide, 88.1% of the residues are found in the most favored regions, with 11.3% in additionally allowed regions and 0.6% in generally allowed regions with no residues found in the disallowed regions. Statistics are summarized in Table 1. Figures were prepared with PyMol (DeLano, 2002).

Phosphatase Assays

pNPP mixtures (20 ml) containing 50 mM Tris-acetate (pH 5.5), 10 mM $MgCl_2$, 0.5–50 mM of pNPP, and 50 ng of recombinant Scp1 were incubated at 37°C for 10 min. The reactions were quenched by adding 80 μl of 0.25 N NaOH. Release of p-NP was determined by measuring the absorbance at 410 nm. The data were analyzed as reported previously (Taylor et al., 1997). For phosphopeptides, reaction mixtures (20 μl) containing 50 mM Tris-acetate (pH 5.5), 10 mM $MgCl_2$, 5 μM–2 mM phosphopeptide, and 5 ng of wild-type or mutant Scp1 were incubated for 10 min at 37°C. The reactions were quenched by adding 40 μl of malachite green reagent. Samples were allowed to sit for 40 min for color development before measuring the absorbance at 620 nm. Inorganic phosphate release was quantified by comparison to a standard curve constructed by using a range of KH_2PO_4 concentrations diluted in distilled H_2O. The malachite green reagent was prepared as reported previously (Taylor and Dixon, 2003). To derive K_M and k_{cat} values, the data were fit by nonlinear regression to the Michaelis-Menten equation.

Binding Studies

CTD peptides were labeled with tetramethylrhodamine isothiocyanate mixed isomers (Invitrogen, Eugene, OR) by using a protocol previously reported (Verdecia et al., 2000). Fluorescence anisotropy was measured on an AVIV model ATF-105 spectrofluorimeter (Aviv Associates, Lakewood, NJ). Dissociation equilibrium constants were calculated with a range of D96N Scp1 mutant concentrations titrated into a binding buffer consisting of 25 mM HEPES-Na^+ (pH 7.5), 100 mM NaCl, and 3 mM DTT containing a fixed concentration (2 μM–1.5 mM) of labeled CTD peptide and fit to a binding isotherm by using the procedure previously reported (Verdecia et al., 2000).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank C. Worby for making the first version of Figure 1A and the staff of the Advanced Light Source, Berkeley, California for assistance during data collection at beam line 8.2.2. This work was supported by the Howard Hughes Medical Institute (J.P.N.), a National Institutes of Health grant GM18024 to (J.E.D.), funds from the Walther Cancer Institute (J.E.D.) and project A.L.S (S.L.P). N.G. was supported by the Swiss National Science Foundation. J.P.N. is an investigator of the Howard Hughes Medical Institute.

References

Allen KN, Dunaway-Mariano D. Phosphoryl group transfer: evolution of a catalytic scaffold. Trends Biochem. Sci. 2004; 29:495–503. [PubMed: 15337123]

Archambault J, Chambers RS, Kobor MS, Ho Y, Cartier M, Bolotin D, Andrews B, Kane CM, Greenblatt J. An essential component of a C-terminal domain phosphatase that interacts with transcription factor IIF in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA. 1997; 94:14300–14305. [PubMed: 9405607]

Baskaran R, Dahmus ME, Wang JY. Tyrosine phosphorylation of mammalian RNA polymerase II carboxyl-terminal domain. Proc. Natl. Acad. Sci. USA. 1993; 90:11167–11171. [PubMed: 7504297]

Brandts JF, Lin LN. Proline isomerization studied with proteolytic enzymes. Methods Enzymol. 1986; 131:107–126. [PubMed: 3534521]

Bregman DB, Pestell RG, Kidd VJ. Cell cycle regulation and RNA polymerase II. Front. Biosci. 2000; 5:D244–D257. [PubMed: 10704151]

Brunger AT. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature. 1992; 355:472–475. [PubMed: 18481394]

Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, et al. Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. 1998; 54:905–921. [PubMed: 9757107]

Buratowski S. The CTD code. Nat. Struct. Biol. 2003; 10:679–680. [PubMed: 12942140]

CCP4 (Collaborative Computational Project, Number 4). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. 1994; 50:760–763. [PubMed: 15299374]

Chambers RS, Dahmus ME. Purification and characterization of a phosphatase from HeLa cells which dephosphorylates the C-terminal domain of RNA polymerase II. J. Biol. Chem. 1994; 269:26243–26248. [PubMed: 7929341]

Cho EJ, Takagi T, Moore CR, Buratowski S. mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. Genes Dev. 1997; 11:3319–3326. [PubMed: 9407025]

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]

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]

Corden JL. Tails of RNA polymerase II. Trends Biochem. Sci. 1990; 15:383–387. [PubMed: 2251729]

Corden JL, Cadena DL, Ahearn JM Jr. Dahmus ME. A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. Proc. Natl. Acad. Sci. USA. 1985; 82:7934–7938. [PubMed: 2999785]

Dahmus ME. Reversible phosphorylation of the C-terminal domain of RNA polymerase II. J. Biol. Chem. 1996; 271:19009–19012. [PubMed: 8759772]

DeLano, WL. The PyMol Molecular Graphics System. DeLano Scientific; San Carlos, CA: 2002.

Fabrega C, Shen V, Shuman S, Lima CD. Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. Mol. Cell. 2003; 11:1549–1561. [PubMed: 12820968]

Hausmann S, Shuman S. Characterization of the CTD phosphatase Fcp1 from fission yeast. Preferential dephosphorylation of serine 2 versus serine 5. J. Biol. Chem. 2002; 277:21213–21220. [PubMed: 11934898]

Hausmann S, Shuman S. Defining the active site of Schizosaccharomyces pombe C-terminal domain phosphatase Fcp1. J. Biol. Chem. 2003; 278:13627–13632. [PubMed: 12556522]

Hausmann S, Erdjument-Bromage H, Shuman S. Schizosaccharomyces pombe carboxyl-terminal domain (CTD) phosphatase Fcp1: distributive mechanism, minimal CTD substrate, and active site mapping. J. Biol. Chem. 2004; 279:10892–10900. [PubMed: 14701811]

Ho CK, Shuman S. Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. Mol. Cell. 1999; 3:405–411. [PubMed: 10198643]

Jez JM, Ferrer JL, Bowman ME, Dixon RA, Noel JP. Dissection of malonyl-coenzyme A decarboxylation from polyketide formation in the reaction mechanism of a plant polyketide synthase. Biochemistry. 2000; 39:890–902. [PubMed: 10653632]

Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for building models in electron density maps and the location of errors in these models. Acta Crystallogr. 1991; A47:110–119.

Kamenski T, Heilmeier S, Meinhart A, Cramer P. Structure and mechanism of RNA polymerase II CTD phosphatases. Mol. Cell. 2004; 15:399–407. [PubMed: 15304220]

Komarnitsky P, Cho EJ, Buratowski S. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes Dev. 2000; 14:2452–2460. [PubMed: 11018013]

Kumaki Y, Matsushima N, Yoshida H, Nitta K, Hikichi K. Structure of the YSPTSPS repeat containing two SPXX motifs in the CTD of RNA polymerase II: NMR studies of cyclic model peptides reveal that the SPTS turn is more stable than SPSY in water. Biochim. Biophys. Acta. 2001; 1548:81–93. [PubMed: 11451441]

Lahiri SD, Zhang G, Dunaway-Mariano D, Allen KN. The pentacovalent phosphorus intermediate of a phosphoryl transfer reaction. Science. 2003; 299:2067–2071. [PubMed: 12637673]

Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. 1993; 26:283–291.

Li M, Phatnani HP, Guan Z, Sage H, Greenleaf AL, Zhou P. Solution structure of the Set2-Rpb1 interacting domain of human Set2 and its interaction with the hyperphosphorylated C-terminal domain of Rpb1. Proc. Natl. Acad. Sci. USA. 2005; 102:17636–17641. [PubMed: 16314571]

Lin PS, Marshall NF, Dahmus ME. CTD phosphatase: role in RNA polymerase II cycling and the regulation of transcript elongation. Prog. Nucleic Acid Res. Mol. Biol. 2002; 72:333–365. [PubMed: 12206456]

Majello B, Napolitano G. Control of RNA polymerase II activity by dedicated CTD kinases and phosphatases. Front. Biosci. 2001; 6:D1358–D1368. [PubMed: 11578967]

McCracken S, Fong N, Rosonina E, Yankulov K, Brothers G, Siderovski D, Hessel A, Foster S, Shuman S, Bentley DL. 5′-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. Genes Dev. 1997; 11:3306–3318. [PubMed: 9407024]

Meinhart A, Cramer P. Recognition of RNA polymerase II carboxy-terminal domain by 3′-RNA-processing factors. Nature. 2004; 430:223–226. [PubMed: 15241417]

Meinhart A, Kamenski T, Hoeppner S, Baumli S, Cramer P. A structural perspective of CTD function. Genes Dev. 2005; 19:1401–1415. [PubMed: 15964991]

Morris DP, Phatnani HP, Greenleaf AL. Phosphocarboxyl-terminal domain binding and the role of a prolyl isomerase in pre-mRNA 3′-End formation. J. Biol. Chem. 1999; 274:31583–31587. [PubMed: 10531363]

Navaza J. AMoRe: an automated package for molecular replacement. Acta Crystallogr. 1994; A50:157–163.

Noble CG, Hollingworth D, Martin SR, Ennis-Adeniran V, Smerdon SJ, Kelly G, Taylor IA, Ramos A. Key features of the interaction between Pcf11 CID and RNA polymerase II CTD. Nat. Struct. Mol. Biol. 2005; 12:144–151. [PubMed: 15665873]

Otwinowski Z, Minor W. HKL: processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 1997; 276:307–326.

Palancade B, Bensaude O. Investigating RNA polymerase II carboxyl-terminal domain (CTD) phosphorylation. Eur. J. Biochem. 2003; 270:3859–3870. [PubMed: 14511368]

Proudfoot NJ, Furger A, Dye MJ. Integrating mRNA processing with transcription. Cell. 2002; 108:501–512. [PubMed: 11909521]

Ranganathan R, Lu KP, Hunter T, Noel JP. Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent. Cell. 1997; 89:875–886. [PubMed: 9200606]

Satow R, Chan TC, Asashima M. Molecular cloning and characterization of dullard: a novel gene required for neural development. Biochem. Biophys. Res. Commun. 2002; 295:85–91. [PubMed: 12083771]

Stiller JW, Hall BD. Evolution of the RNA polymerase II C-terminal domain. Proc. Natl. Acad. Sci. USA. 2002; 99:6091–6096. [PubMed: 11972039]

Taylor GS, Dixon JE. PTEN and myotubularins: families of phosphoinositide phosphatases. Methods Enzymol. 2003; 366:43–56. [PubMed: 14674238]

Taylor GS, Liu Y, Baskerville C, Charbonneau H. The activity of Cdc14p, an oligomeric dual specificity protein phosphatase from Saccharomyces cerevisiae, is required for cell cycle progression. J. Biol. Chem. 1997; 272:24054–24063. [PubMed: 9295359]

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]

Vojnic E, Simon B, Strahl BD, Sattler M, Cramer P. Structure and carboxyl-terminal domain (CTD) binding of the Set2 SRI domain that couples histone H3 Lys36 methylation to transcription. J. Biol. Chem. 2006; 281:13–15. [PubMed: 16286474]

Wang W, Cho HS, Kim R, Jancarik J, Yokota H, Nguyen HH, Grigoriev IV, Wemmer DE, Kim SH. Structural characterization of the reaction pathway in phosphoserine phosphatase: crystallographic "snapshots" of intermediate states. J. Mol. Biol. 2002; 319:421–431. [PubMed: 12051918]

Xu YX, Manley JL. Pinning down transcription: regulation of RNA polymerase II activity during the cell cycle. Cell Cycle. 2004; 3:432–435. [PubMed: 14739776]

Yaffe MB, Schutkowski M, Shen M, Zhou XZ, Stukenberg PT, Rahfeld JU, Xu J, Kuang J, Kirschner MW, Fischer G, et al. Sequence-specific and phosphorylation-dependent proline isomerization: a potential mitotic regulatory mechanism. Science. 1997; 278:1957–1960. [PubMed: 9395400]

Yamamoto H, Esaki M, Kanamori T, Tamura Y, Nishikawa S, Endo T. Tim50 is a subunit of the TIM23 complex that links protein translocation across the outer and inner mitochondrial membranes. Cell. 2002; 111:519–528. [PubMed: 12437925]

Yeo M, Lin PS, Dahmus ME, Gill GN. A novel RNA polymerase II C-terminal domain phosphatase that preferentially dephosphorylates serine 5. J. Biol. Chem. 2003; 278:26078–26085. [PubMed: 12721286]

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]

Zhang J, Corden JL. Identification of phosphorylation sites in the repetitive carboxyl-terminal domain of the mouse RNA polymerase II largest subunit. J. Biol. Chem. 1991; 266:2290–2296. [PubMed: 1899239]

Zheng H, Ji C, Gu S, Shi B, Wang J, Xie Y, Mao Y. Cloning and characterization of a novel RNA polymerase II C-terminal domain phosphatase. Biochem. Biophys. Res. Commun. 2005; 331:1401–1407. [PubMed: 15883030]

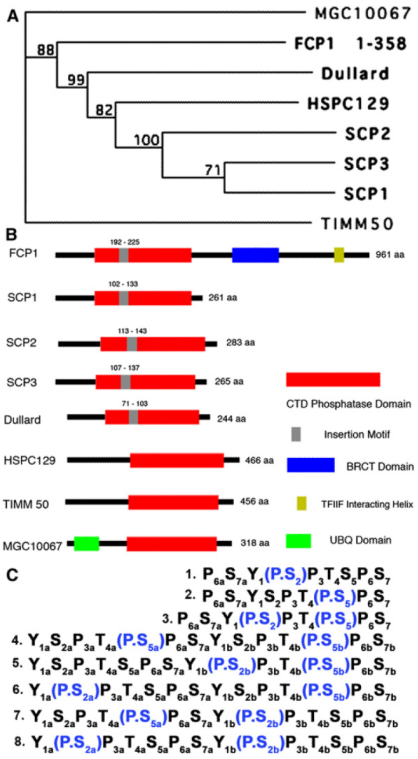

Figure 1. Family of Human CTD Phosphatases

(A) Dendrogram of human CTD phosphatases. The bootstrap values within the dendrogram are shown at nodes on the dendrogram. The NCBI accession numbers for the proteins are Fcp1 (AAC64549), Scp1 (AAH12977), Scp2 (AAH65920), Scp3 (NP005799), HSPC 129 (AAF29093), Dullard (AAH09295), TIMM50 (NP001001563), and MGC10067 (AAH13425).

(B) Domain structure of CTD phosphatases. Each domain is represented by a colored block. The catalytic domain is red. An insertion domain is partially conserved and is gray.

(C) Phosphopeptides used for the experiments. Phosphoresidues (blue) are highlighted in parentheses. Residue numbers are shown as subscripts.

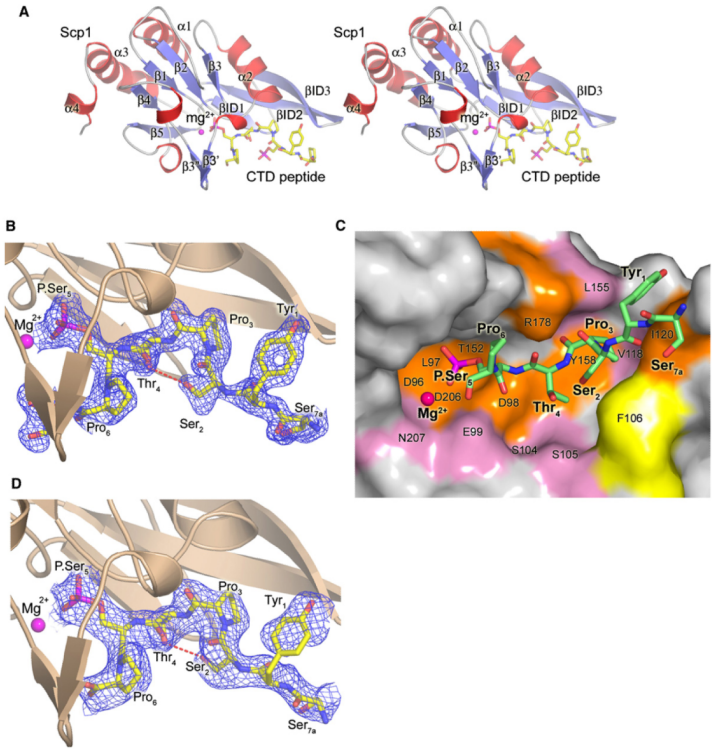

Figure 2. Structures of Human Scp1 Complexed to CTD Phosphopeptides

(A) Stereo ribbon diagram of human Scp1 bound to a CTD phosphopeptide with helices as red coils and b strands as blue arrows. The three-stranded insert is labeled βID1-3. The CTD peptide is a stick diagram with color-coded bonds. Yellow is carbon, red is oxygen, blue is nitrogen, and magenta is phosphorus. The Mg^{2+} ion is shown as a magenta van der Waals sphere.

(B) Model of the monophosphorylated CTD peptide complex ($P.Ser_5$) as a half-colored bond diagram with the blue SIGMAA weighted $2F_o - F_c$ electron-density map contoured at 1σ. An intramolecular hydrogen bond between the Ser_2 and Thr_4 is shown as red cylinders.

(C) Accessible surface of Scp1 bound to monophosphorylated $P.Ser_5$ CTD peptide. Surfaces conserved between human Fcp1 and human Scp1 are orange, and chemically similar residues are pink. Phe106 is yellow. The peptide is shown as half-colored bonds with carbon atoms light green.

(D) Model of the doubly phosphorylated 14-mer CTD peptide complex ($P.Ser_5$-$P.Ser_5$) as a half-colored bond diagram with the blue SIGMAA-weighted $2F_o - F_c$ electron-density map contoured at 1σ.

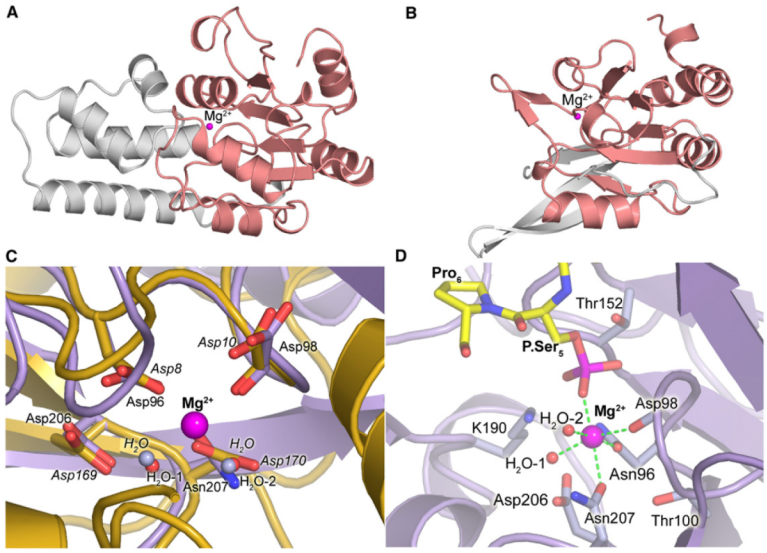

Figure 3. DXDX Family of Mg^{2+}-Dependent Enzymes

(A) β-phosphoglucomutase, a member of the DXDX family, shares its three-dimensional architecture with Scp1 (PDB code 1o08). The conserved core is merlot, and the elements that do not align structurally with Scp1 are light gray. The Mg^{2+} is depicted as a van der Waals sphere.

(B) Three-dimensional architecture of Scp1 shown in the same orientation.

(C) Active sites of Scp1 and β-phosphoglucomutase. The β-phosphoglucomutase ribbons are gold, and the Scp1 ribbons are lavender. Labels are in italic for β-phosphoglucomutase.

(D) Mg^{2+} coordination of the phospho-CTD peptide with bonds as green cylinders. The peptide is shown as half-colored bonds. Residues essential for the dephosphorylation reaction (Hausmann and Shuman, 2003) are highlighted as color-coded half bonds.

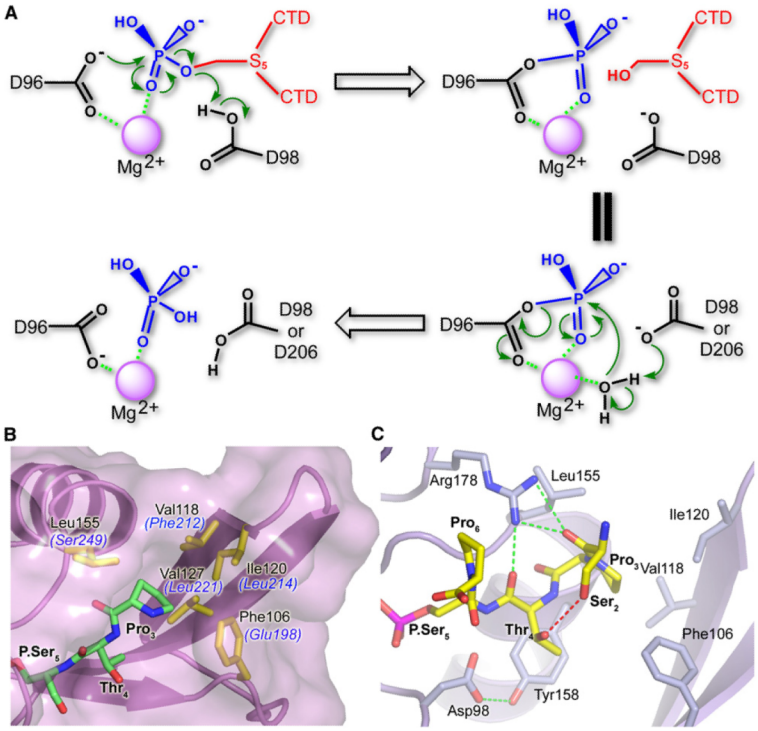

Figure 4. Reaction and Binding Mechanisms of Scp1

(A) Reaction mechanism of Scp1. The active site geometry, water coordination, and position of Asp206 supports its participation as a general base for activating a water for breakdown of the mixed anhydride intermediate formed after nucleophilic attack of Asp96 on $P.Ser_5$. Alternatively, the residue equivalent to Asp98 may function as the general base for this final catalytic step (Wang et al., 2002).

(B) Pro_3 binding pocket of Scp1 and comparison with Fcp1. The lavender ribbon underlies a transparent surface used to illustrate the steric volume surrounding the Pro_3 moiety. Peptide and side chains are shown as half-colored bonds with green highlighting carbon for the peptide and yellow highlighting carbon for Scp1. The equivalent residues in human Fcp1 are labeled blue in italic.

(C) Scp1 residues involved in the binding of the phospho-CTD. Intermolecular hydrogen bonds are shown as rendered green cylinders, and an intramolecular hydrogen bond in the CTD peptide is shown as rendered red cylinders.

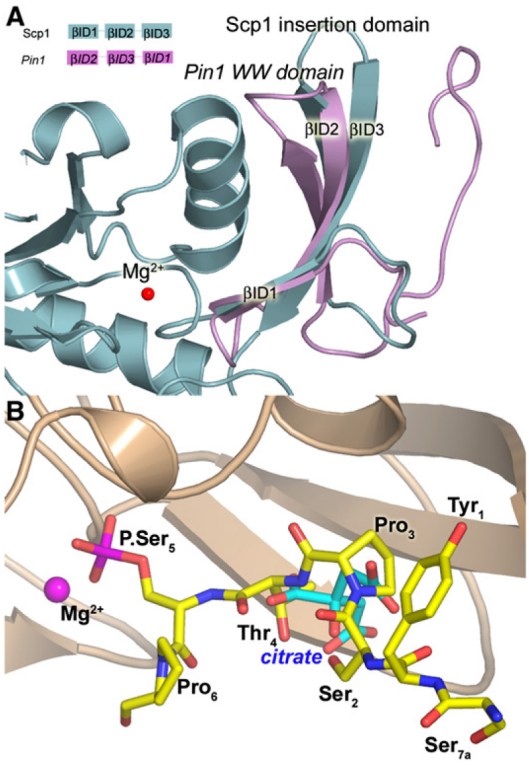

Figure 5. CTD-Binding Proteins

(A) Superimposition of the Pin1 WW domain and the Scp1 insertion domain. The connectivity pattern of β strands for both Pin1 (lavender) and Scp1 (cyan) is shown schematically as an inset.

(B) Comparison of the bound citrate molecule previously observed in apo Scp1 (Kamenski et al., 2004) and the bound phospho-CTD peptides reported here. The citrate molecule and the phospho-CTD peptide are shown as rendered bonds (CTD peptide is yellow and the citrate molecule is cyan). Citrate superimposes with the CTD peptide backbone and the Pro_3 and Thr_4 side chains.

Table 1

Data Collection and Refinement Statistics

	Scp1 Bound to Doubly Phosphorylated ($P.Ser_2P.Ser_5$) CTD 9-mer Peptide	**Scp1 Bound to Singly Phosphorylated ($P.Ser_5$) CTD 9-mer Peptide**	**Scp1 Bound to Doubly Phosphorylated ($P.Ser_5$) CTD 14-mer Peptide**
Data Collection			
Space group	C2	C2	C2
Cell dimensions			
a, b, c (AÅ)	126.0, 78.7, 62.8	125.7, 78.5, 62.8	125.7, 78.5, 62.8
α,g=b, γ (°)	90.0, 112.3, 90.0	90.0, 112.1, 90.0	90.0, 112.2, 90.0
Resolution (AÅ)	49.2–2.05 (2.15–2.05)[a]	39.2–1.80 (1.90–1.80)[a]	32.2–2.35 (2.40–2.35)[a]
R_{sym} or R_{merge} (%)	8.1 (45.4)[a]	4.1 (31.2)[a]	7.1 (29.6)[a]
I/σI	12.1 (2.8)[a]	14.2 (2.9)[a]	15.7 (1.6)[a]
Completeness (%)	95.7 (83.5)[a]	90.7 (78.5)[a]	94.1 (91.9)[a]
Refinement			
Resolution (AÅ)	49.2–2.05	39.2–1.80	32.2–2.35
No.reflections	35,936	52,830	24,261
R_{work}/R_{free} (%)	21.8 / 24.8[b]	21.8 / 23.3[b]	24.8 / 26.8[b]
Number of atoms			
Protein	2928	2928	2928
peptide/Mg^{2+}	116	100	97
Water	103	246	97
B factors (AÅ^2)			
Protein	37.4	31.1	37.4
Peptide/Mg^{2+}	52.5	44.2	52.5
Water	39.8	38.2	39.1
Rmsds			
Bond lengths (AÅ)	0.016	0.011	0.011
Bond angles (°)	1.82	1.39	1.60

[a]Highest-resolution shell is shown in parentheses.

[b] R_{free} is calculated with 5% of the data randomly omitted from refinement.

Table 2

Kinetic Characterization of Wild-Type Scp1 and F106A Scp1

Enzyme	Peptide Substrate	K_M (mM)	k_{cat} (s^{-1}	k_{cat} / K_M ($s^{-1}mM^{-1}$)
Wild-type	pNPP	5.80 ± 0.571	4.20 ± 0.212	0.724 ± 0.061
	PSY(P.S)PTSPS	ND[a]	ND	ND
	PSYSPT(P.S)PS	0.287 ± 0.074	2.88 ± 0.192	10.0 ± 0.026
	YSPTSPSY(P.S)PTSPS	5.61 ± 0.269	0.935 ± 0.129	0.166 ± 0.038
	YSPTSPSYSPT(P.S)PS	0.214 ± 0.051	2.44 ± 0.041	11.4 ± 0.019
	Y(P.S)PTSPSY(P.S)PTSPS	3.59 ± 0.606	0.670 ± 0.126	0.187 ± 0.028
	Y(P.S)PTSPSYSPT(P.S)PS	0.185 ± 0.051	3.22 ± 0.254	17.4 ± 2.81
	YSPT(P.S)PSYSPT(P.S)PS	0.103 ± 0.015	4.09 ± 0.055	39.7 ± 4.36
	YSPTSPSY(P.S)PT(P.S)PS	0.223 ± 0.039	1.49 ± 0.094	6.68 ± 0.968
	YSPT(P.S)PSY(P.S)PTSPS	0.273 ± 0.050	2.89 ± 0.085	10.6 ± 0.086
F106A	pNPP	6.32 ± 0.631	4.00 ± 0.328	0.632 ± 0.059
	YSPTSPSY(P.S)PTSPS	1.11 ± 0.083	0.851 ± 0.237	0.766 ± 0.180
	YSPTSPSYSPT(P.S)PS	1.31 ± 0.051	4.40 ± 0.698	3.36 ± 0.431

[a]ND, not determined.Data are averages of three measurements with associated 6 standard deviations.