Molecular Cell, Vol. 15, 399–407, August 13, 2004, Copyright ©2004 by Cell Press

Structure and Mechanism of RNA Polymerase II CTD Phosphatases

Tomislav Kamenski, Susanna Heilmeier, Anton Meinhart, and Patrick Cramer*
Department of Chemistry and Biochemistry
Gene Center
University of Munich
Feodor-Lynen-Str. 25
81377 Munich
Germany

Summary

Recycling of RNA polymerase II (Pol II) after transcription requires dephosphorylation of the polymerase C-terminal domain (CTD) by the phosphatase Fcp1. We report the X-ray structure of the small CTD phosphatase Scp1, which is homologous to the Fcp1 catalytic domain. The structure shows a core fold and an active center similar to those of phosphotransferases and phosphohydrolases that solely share a DXDX(V/T) signature motif with Fcp1/Scp1. We demonstrate that the first aspartate in the signature motif undergoes metal-assisted phosphorylation during catalysis, resulting in a phosphoaspartate intermediate that was structurally mimicked with the inhibitor beryllofluoride. Specificity may result from CTD binding to a conserved hydrophobic pocket between the active site and an insertion domain that is unique to Fcp1/Scp1. Fcp1 specificity may additionally arise from phosphatase recruitment near the CTD via the Pol II subcomplex Rpb4/7, which is shown to be required for binding of Fcp1 to the polymerase in vitro.

Introduction

RNA polymerase II (Pol II) consists of a folded region that is responsible for mRNA synthesis, and the mobile CTD (Cramer et al., 2001). The CTD is a unique tail-like feature of the largest Pol II subunit that consists of up to 52 heptapeptide repeats of the consensus sequence Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7. The CTD is required for mRNA processing in vivo and binds to processing factors in vitro (Hirose and Manley, 2000; Proudfoot et al., 2002). During the transcription cycle, the CTD undergoes dynamic phosphorylation and dephosphorylation (Orphanides and Reinberg, 2002). Transcription initiation requires an unphosphorylated CTD, whereas transcription elongation is carried out by Pol II with a hyperphosphorylated CTD that binds mRNA processing factors for transcription-coupled mRNA maturation. For recycling of Pol II and reinitiation of transcription, the CTD must be dephosphorylated.

Dephosphorylation of CTD serine residues is catalyzed by the CTD phosphatase Fcp1 (Archambault et al., 1997; Archambault et al., 1998; Chambers and Dahmus, 1994; Chambers and Kane, 1996; Cho et al., 1999; Kobor et al., 1999; Lin et al., 2002b). Fcp1 is the founding member of a new phosphatase family with no sequence similarity to other phosphatases, except for the signature motif DXDX(T/V) that is shared within a superfamily of phosphotransferases and phosphohydrolases (Collet et al., 1998). The Fcp1 sequence comprises two conserved regions. The N-terminal Fcp1 homology (FCPH) region includes the DXDX(T/V) signature motif with residues important for catalysis. The C-terminal BRCT (breast cancer protein-related carboxy-terminal) domain binds to the phosphorylated CTD (Yu et al., 2003).

Fcp1 is conserved throughout eukaryotes and is essential for cell viability. Genome-wide expression analysis of a temperature-sensitive Fcp1 mutant shows a nearly global defect in Pol II transcription (Kobor et al., 1999). Dephosphorylation of the Pol II CTD by Fcp1 can facilitate recycling of the hyperphosphorylated form of the polymerase for a new round of transcription (Cho et al., 1999). Fcp1 dephosphorylates the CTD in solution (Chambers et al., 1995; Cho et al., 1999) and when associated with a transcription elongation complex (Cho et al., 1999; Lehmann and Dahmus, 2000). The phosphatase activity of Fcp1 is stimulated by the general transcription factor TFIIF, and the general factor TFIIB inhibits this stimulation (Chambers et al., 1995). Taken together, these results indicate that Fcp1 is a central regulator of the mRNA transcription cycle, but its structure and mechanism remained unclear.

Recently, a second CTD phosphatase in higher eukaryotes, the small CTD phosphatase Scp1, was described (Yeo et al., 2003). Scp1 shows homology to the catalytic FCPH domain of Fcp1 but lacks the BRCT domain. Here, we present the high-resolution crystal structure of human Scp1. Biochemical and further structural studies of recombinant Scp1 and Fcp1 reveal the catalytic mechanism of these enzymes. We also present data that suggest a role of the Pol II subcomplex Rpb4/7 in Fcp1 recruitment to the CTD.

Results and Discussion

Domain Organization and Structure Determination

For structure determination of Fcp1 from the yeast *Saccharomyces cerevisiae*, a soluble core variant comprising the FCPH and BRCT regions could be expressed, purified, and crystallized (Fcp1c, residues 168–606; Figure 1A). However, Fcp1c crystals were not suited for structural studies. Limited proteolysis of Fcp1c suggested that the FCPH and BRCT regions form structured domains connected by a partially exposed linker (Figure 1A). Since the catalytic FCPH domain was generally protected from proteolytic cleavage, we prepared Fcp1 variants comprising only the FCPH domain for crystallization, but these were not soluble under the conditions tested.

We therefore turned to human Scp1, which corresponds to the FCPH domain with an N-terminal extension (Figure 1A). The first 63 residues of the extension were cleaved from the full-length protein during expression in *E. coli* (Figure 1A), prompting us to prepare sev-

*Correspondence: cramer@LMB.uni-muenchen.de

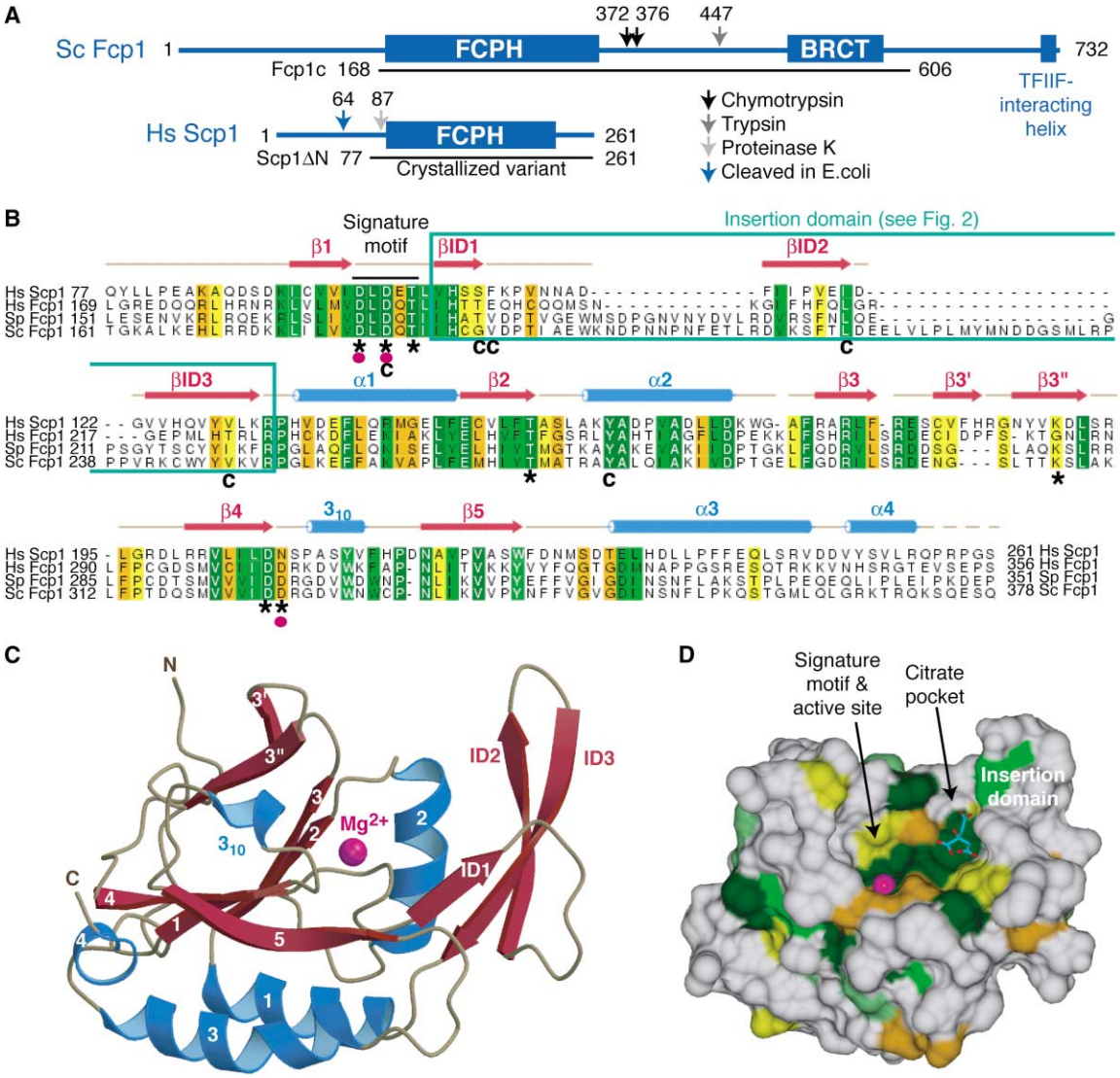

Figure 1. Structure of CTD Phosphatase Scp1

(A) Domain organization of *Saccharomyces cerevisiae* Fcp1 (Sc Fcp1), and human Scp1 (Hs Scp1). FCPH and BRCT denote the Fcp1 homology regions and the breast cancer protein-related carboxy-terminal domains, respectively. The amino acids C-terminal of protease cleavage sites are depicted with arrows. Protein variants used in this study are indicated below the diagrams.

(B) Alignment of human Scp1 (Hs Scp1) with the catalytic FCPH domains of Fcp1 from human (Hs), *Schizosaccharomyces pombe* (Sp), and *Saccharomyces cerevisiae* (Sc). α helices and β strands are indicated as cylinders and arrows, respectively. Active site residues are marked with an asterisk below the alignment. Residues involved in metal ion and citrate binding are marked below the alignment with a pink dot and with a "c," respectively. Conserved residues are highlighted with the degree of conservation decreasing from dark green to yellow.

(C) Ribbon model of Scp1. Secondary structure elements are colored according to (B). The catalytic metal ion is shown as a pink sphere.

(D) Surface conservation and putative specificity pocket. The molecular surface of Scp1 is colored according to amino acid residue conservation as in (B). A citrate ion that binds to the largely conserved putative specificity pocket is shown as a stick model.

eral Scp1 variants with N-terminal truncations. A variant lacking 76 N-terminal residues (Scp1ΔN) was highly soluble and crystallized with a plate-like morphology. Although these crystals were only about 5 μm thick, they diffracted synchrotron radiation to better than 2 Å resolution. The X-ray structure was determined by multiwavelength anomalous diffraction with the selenomethionine-substituted double methionine mutant L165M/L205M of Scp1ΔN (Table 1). The structure was refined at 2.3 Å resolution, has very good stereochemistry, and reveals chemical details.

Overall Structure

Scp1 forms a central, 5-stranded, parallel β sheet with the strand order 3-2-1-4-5, flanked by two α helices on one side and a two-stranded β sheet and a short 3_{10} helix on the other side (Figure 1). The signature motif is located at the end of strand β1 and, together with other conserved residues, lines a central depression that forms the active site of the enzyme. Although there is no sequence similarity outside the signature motif, a DALI search (Holm and Sander, 1995) revealed that the core fold of Scp1 largely corresponds to that of phos-

Table 1. Structure Determination and Refinement

Crystal	Mutant (L165M/L205M) SeMet MAD			Wild-Type + BeF_3^-
Data collection				
Space group	$P2_12_12$			$P2_12_12$
Unit cell dimensions (Å)	117.6 × 47.2 × 40.0			117.8 × 47.2 × 40.1
Wavelength (Å)	0.9795 peak	0.9797 inflection	0.9686 remote	0.9919
Resolution (Å)	20–2.3 (2.38–2.3)	20–2.3 (2.38–2.3)	20–2.3 (2.38–2.3)	20–2.1 (2.2–2.1)
Completeness (%)	88.7 (78.7)	90.8 (81.9)	90.9 (80.1)	97.7 (87.5)
Unique reflections	9,386 (815)	9,639 (857)	9,635 (840)	13,356 (1,520)
Redundancy	17.4	17.2	17.0	3.8
R_{sym} (%)	6.3 (9.8)	5.2 (10.2)	5.2 (8.5)	5.7 (6.2)
Mean I/σ I	21.8 (15.5)	21.1 (12.7)	22.7 (14.0)	19.0 (16.8)
f'	−7.0	−10.0	−2.3	
f''	5.1	2.5	3.5	
Refinement				
Residues			181 (M76-Q255)	181 (M76-Q255)
Water molecules			131	148
Magnesium ion			1	1
Citrate ion			1	1
Rmsd bonds (Å)			0.006	0.006
Rmsd angles (°)			1.29	1.25
R_{cryst} (%)			18.7	20.7
R_{free} (%)			23.4	22.7

phoserine phosphatase (PSP) (Wang et al., 2001) and β-phosphoglucomutase (β-PGM) (Lahiri et al., 2003), both members of the DXDX(T/V) superfamily of phosphohydrolases and phosphotransferases, and to 1, 2-L-haloacid dehalogenase (HAD) (Ridder et al., 1999) (Figure 2). The major structural difference between these enzymes lies in a large insertion after strand β1, which may determine substrate specificity ("insertion domain," Figures 1 and 2; see below). Within the core domain, Scp1 lacks a sixth β strand that is present in the central β sheet of the other enzymes (Figure 2).

Homology modeling of the FCPH domain of human Fcp1 showed that residues forming the hydrophobic core are conserved, although the overall sequence identity in the FCPH domains of human Scp1 and Fcp1 is only 21%. This indicates that our structure is a valid model for the Fcp1 FCPH domain, and that mutational data on Fcp1 can be interpreted based on the Scp1 structure.

Phosphatase Activity

To investigate whether the structural resemblance of Fcp1 and Scp1 is reflected in comparable enzymatic properties, we subjected highly purified recombinant

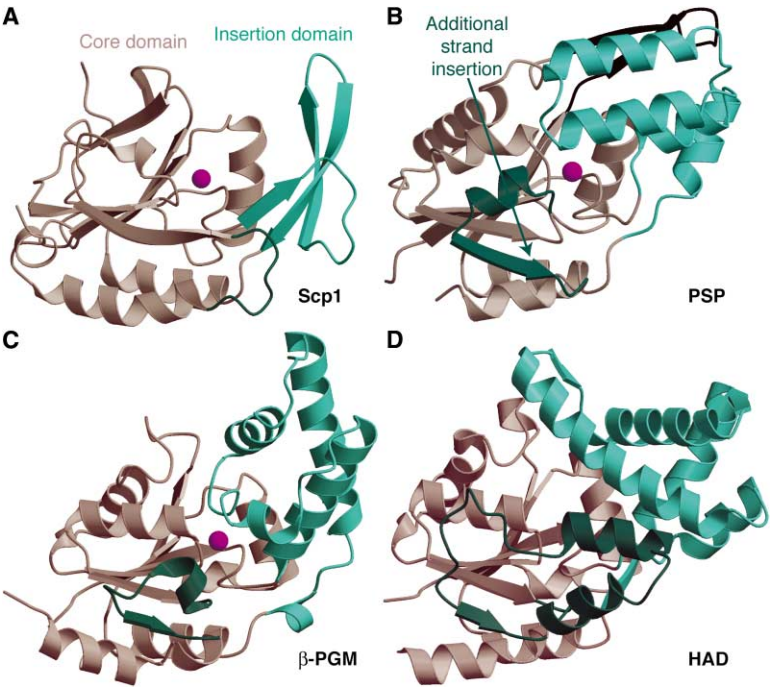

Figure 2. Comparison of DXDX(T/V) Superfamily Enzyme Structures

(A–D) Ribbon models of the DXDX(T/V) superfamily phosphotransferases (A) Scp1 (this study), (B) phosphoserine phosphatase PSP (PDB code 1F5S [Wang et al., 2001]), (C) β-phosphoglucomutase β-PGM (PDB code 1O03 [Lahiri et al., 2003]), and (D) 1, 2-haloacid dehalogenase HAD (PDB code 1QQ5 [Ridder et al., 1999]). Structural similarities were found with the DALI server (Holm and Sander, 1995), which detects fold similarities by superposing a query structure onto representative structures in the database and calculating the RMS deviation of carbon-α atom positions. The catalytic core domain of the DXDX(T/V) enzymes, which includes a central parallel β sheet and its flanking regions, is shown in brown. The insertion domain of Scp1 and the corresponding regions in the other enzymes are shown in turquoise. The catalytic metal ion is depicted as a pink sphere.

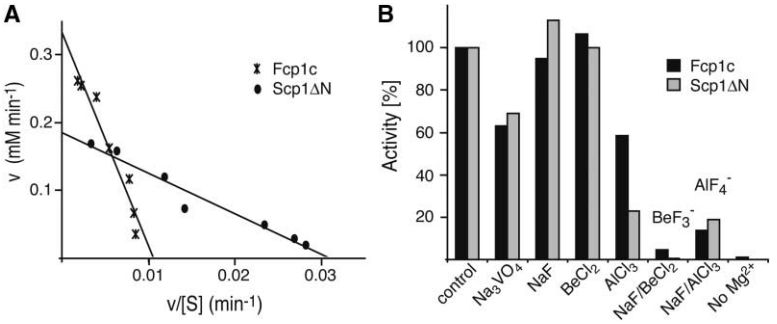

Figure 3. Catalytic Activity and Specific Inhibition of CTD Phosphatases

(A) Determination of catalytic parameters. Catalytic activity of Scp1ΔN and Fcp1c was determined as described in the Experimental Procedures. To derive K_M and k_{cat}, the initial velocities (v) at seven different substrate concentrations were plotted according to Eadie-Hofstee against the velocities divided by the substrate concentration. The plot was fitted to the equation $v = V_{max} - K_M(v/[S])$.

(B) Specific inhibition. The inhibitors were added to reaction mixtures (see the Experimental Procedures) in the following concentrations: 1 mM sodium vanadate (Na_3VO_4), 1 mM sodium fluoride (NaF), 100 μM beryllium chloride ($BeCl_2$), 100 μM aluminum chloride ($AlCl_3$), a mixture of 1 mM sodium fluoride and 100 μM beryllium chloride (to produce the beryllofluoride anion BeF_3^- in situ), and a mixture of 1 mM sodium fluoride and 100 μM aluminum chloride (to produce AlF_4^- in situ).

human Scp1ΔN and yeast Fcp1c to a spectrophotometric assay based on cleavage of the nonspecific substrate *para*-nitrophenylphosphate (pNPP), and we determined the catalytic parameters (see the Experimental Procedures). Both proteins readily cleaved pNPP in a pH-dependent manner, with maximum activity at pH 5.5. Michaelis-Menten kinetic analysis revealed K_M values of 6 mM and 32 mM for human Scp1ΔN and yeast Fcp1c, respectively (Figure 3A). These K_M values are comparable to published values for endogenous *S. cerevisiae* Fcp1 (60 mM) (Kobor et al., 1999) and full-length recombinant *Schizosaccharomyces pombe* Fcp1 (19 mM) (Hausmann and Shuman, 2002). The turnover k_{cat} numbers were 17 s^{-1} and 33 s^{-1} for human Scp1ΔN and yeast Fcp1c, respectively. Thus, the recombinant proteins used in our structural studies are active, and the structural similarity of the FCPH domains of Scp1 and Fcp1 is reflected in similar enzymatic properties, although Fcp1 has slightly higher catalytic activity.

Specific Inhibition

To investigate the catalytic mechanism, we tested the effect of various inhibitors of phosphoryl-transfer reactions on the activity of Scp1ΔN and Fcp1c (Figure 3B). Phosphatase activity of both enzymes was essentially abolished by the beryllofluoride anion BeF_3^-, produced in situ from $BeCl_2$ and NaF. In contrast, $BeCl_2$ or NaF alone did not have a strong effect on activity. The inhibitory effect of AlF_4^- was less severe, and sodium vanadate only inhibited activity slightly. Magnesium ions were essential for activity. The data indicate the presence of a catalytic aspartate residue, as BeF_3^- can form a stable tetrahedral adduct with catalytic aspartate side chains, mimicking a labile phosphoaspartate intermediate (Cho et al., 2001; Yan et al., 1999), wheras AlF_4^- tends to mimic a phosphate (Chabre, 1990). The response of Fcp1 and Scp1 to the inhibitors was very similar, indicating that these enzymes share a common catalytic mechanism.

Active Center

The enzymatic and inhibitory data are reflected in the active site structure. Mutational analysis of Fcp1 from *S. pombe* has defined 11 amino acid residues in the FCPH domain that are important for catalytic activity (Hausmann et al., 2004; Hausmann and Shuman, 2003). The Scp1 structure shows that 7 of these residues, including 3 residues in the signature motif, cluster in the central depression and form the active site (Table 2, Figure 4). The remaining 4 residues are not found in the active center depression and are involved in salt bridges or hydrophobic core interactions; based on these findings, we predict that their mutation disrupts the domain structure (*S. pombe* Fcp1 residues R223, Y237, Y249, and D258, corresponding to Scp1 residues R132, F164, Y158, and D167, respectively; Figure 1B). The 7 active site residues generally superimpose well with corresponding residues in structures of other enzymes of the DXDX(T/V) superfamily (Figure 4C). In the case of K190, the counterpart residues in PSP and PGM protrude from different locations of the protein backbones, but functional head groups occupy the same position. This is

Table 2. Active Site Residues in DXDX(T/V) Phosphotransferases and Phosphohydrolases

H.s. Scp1[a]	S.c. Fcp1[a]	S.p. Fcp1	PSP[a]	β-PGM[a]	Functional or Structural Role
D96	D180	D170	D11	D8	Phosphoryl acceptor, metal coordination
D98	D182	D172	D13	D10	General acid/base, metal coordination
T100	T184	T174	T15	V12	Positions side chains of D96 and N207
T152	T270	T243	S99	S114	Transition state stabilization
K190	K307	K280	K144[b]	K145[b]	Transition state charge stabilization
D206	D324	D297	D171[b]	E169	Salt bridge with K190 in Scp1
N207	D325	D298	D167	D170	Metal coordination

[a] Scp1, small CTD phosphatase 1; Fcp1, TFIIF-dependent CTD phosphatase 1; PSP, phosphoserine phosphatase (PDB codes 1F5S, 1J97); β-PGM, β-phosphoglucomutase (PDB codes 1O08, 1O03).

[b] These residues protrude from different sites of the protein backbones, but their functional head groups are at positions equivalent to those in the Scp1 structure (see Figure 4C).

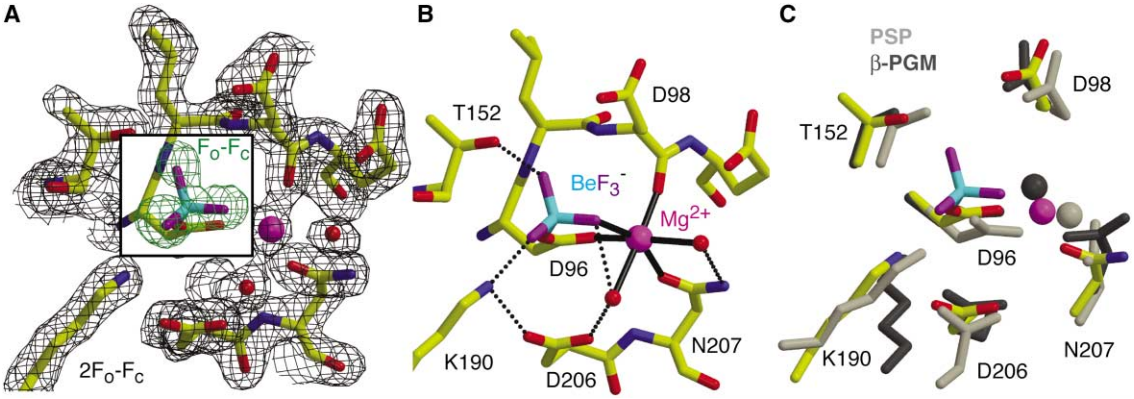

Figure 4. Active Site and Mimicry of the Phosphoaspartate Intermediate

(A) Electron density. The final $2F_o–F_c$ density is shown for residues in the active site (black, contoured at 1.4σ). For the beryllofluoride anion, which mimics the phosphoasparate-96 intermediate, the $F_o–F_c$ density calculated from the model lacking the beryllofluoride is shown (green, contoured at 3.0σ), and the $2F_o - F_c$ density is omitted.

(B) Active site interactions. Hydrogen bonds are shown as dotted lines, and metal ion-ligand interactions are shown as solid black lines. The metal ion is coordinated by the D96 side chain oxygen (2.3 Å), by the D98 carbonyl oxygen (2.1 Å), by the N207 side chain oxygen (2.3 Å), by a fluorine atom in the beryllofluoride (2.2 Å), and by two water molecules (2.4 and 2.5 Å).

(C) Superposition of active site residues in Scp1 (yellow), phosphoserine phosphatase (light gray, PDB code 1F5S; Wang et al., 2001), and β-phosphoglucomutase (dark gray, PDB code 1O08 [Lahiri et al., 2003]). Six out of seven active site residues are shown.

also true for the counterpart of D206 in PSP. Consistent with the requirement for metal ions in catalysis, the active site comprises a metal ion that is coordinated by residues D96, D98, N207, and two water molecules. The metal ion shows the highest peak in a difference Fourier map phased with the final model lacking any water molecules or ions (9.8σ).

Catalytic Mechanism

The structural and functional data strongly suggest that the catalytic mechanism of Scp1 and Fcp1 involves the metal-dependent formation of a phosphoaspartate intermediate. This mechanism was suggested for *S. pombe* Fcp1 based on biochemical data (Hausmann and Shuman, 2003) and was shown for other DXDX(T/V) superfamily enzymes, which use the N-terminal aspartate in the signature motif as the phosphoryl acceptor (Cho et al., 2001; Lahiri et al., 2003).

To directly show that the first aspartate in the signature motif acts as the phosphoryl acceptor in Scp1, we soaked crystals with trifluoroberyllate (beryllofluoride). A difference Fourier map revealed a beryllofluoride anion bound to one oxygen atom of the D96 side chain carboxylate (Figure 4A). The metal ion is bound to the other D96 side chain oxygen atom and shows an approximate octahedral coordination with metal-ligand distances of 2.0–2.3 Å (Figure 4B). A metal ion is found at an equivalent position in structures of PSP (Wang et al., 2001) and β-PGM (Lahiri et al., 2003). Thus, beryllofluoride mimics the labile phosphoaspartate intermediate formed at residue D96 in the active center of Scp1. Consistently, mutation of this aspartate in Scp1ΔN or Fcp1c to alanine abolished activity (Katharina Gmelch, T.K., S.H. and P.C., unpublished data). The same mutation abolishes activity of full-length *S. cerevisiae* Fcp1 (Kobor et al., 1999) and *S. pombe* Fcp1 (Hausmann and Shuman, 2002). These data establish D96 as the phosphoryl acceptor in Scp1. The conservative Scp1 mutation D96E retains most of the activity (Yeo et al., 2003), apparently because the carboxylate group can still act as the phosphoryl acceptor.

The functional role of the other active site residues can be inferred from their relative location and from comparison with other enzymes of the DXDX(T/V) superfamily. In addition to metal ion binding, D98 may act as a general acid/base, donating a proton to the leaving group (the CTD serine side chain). D98 may also position and activate a water molecule for the second step, dephosphorylation of the phosphoasparate and regeneration of the free D96 side chain. The metal ion and the

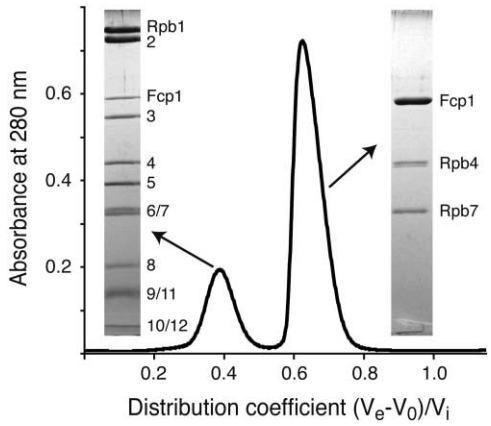

Figure 5. Reconstitution of the Pol II-Fcp1 Complex

Purified endogenous yeast Pol II core enzyme was incubated with an excess of recombinant Rpb4/7 subcomplex and Fcp1c, and the resulting mixture was separated on a Superose-6 gel filtration column. The elution profile was normalized by the distribution coefficient (V_e = elution volume, V_0 = void volume, V_i = inner volume, as determined by retention of acetone). The two peak fractions were analyzed by SDS-PAGE. The first peak corresponds to a reconstituted Pol II-Fcp1 complex.

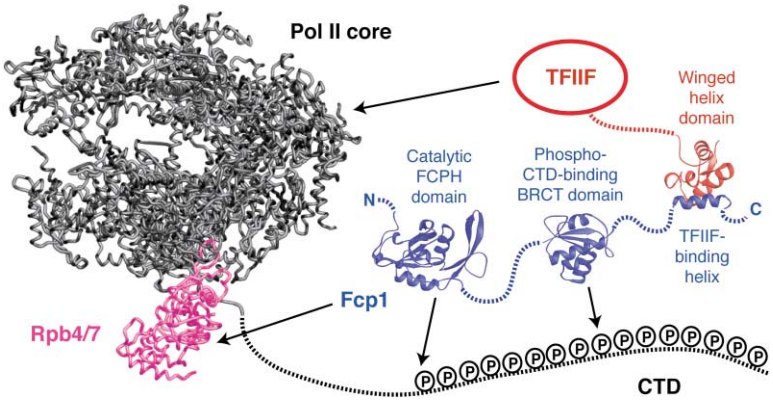

Figure 6. Protein-Protein Interaction Network in a Pol II-TFIIF-Fcp1 Complex

The Pol II core enzyme and the Rpb4/7 subcomplex are shown in black and pink, respectively (Armache et al., 2003). Fcp1 is in blue, with the FCPH domain modeled with the Scp1 structure lacking the two C-terminal helices, and the BRCT domain modeled as a canonical BRCT domain (PDB code 1JNX [Williams et al., 2001]). The C-terminal Fcp1 helix is shown in complex with the TFIIF large-subunit winged helix domain (red, PDB code 1J2X [Kamada et al., 2003]). Protein-protein interactions are indicated by arrows.

head groups of the side chains of T152 and K190 all lie in a plane slightly above the position of the beryllium atom, and thus within a presumed equatorial plane of a pentavalent trigonal bipyramidal transition state. The metal ion and residues T152 and K190 are therefore expected to stabilize the geometry of the transition state and to partially neutralize its charge.

CTD Specificity

Specificity for the CTD may result from recognition of CTD residues neighboring the phosphorylated target side chain. There is evidence that CTD recognition involves the specific insertion domain of Scp1 and Fcp1, which directly follows the signature motif (Figures 1B and 2A). The insertion domain and the core fold line a deep pocket, which binds a citrate ion that was present in the crystallization solution (Figure 1D). The citrate is only 8 Å away from the beryllofluoride. The pocket is largely hydrophobic, with many of the lining residues conserved between Scp1 and Fcp1, including several residues that contact the citrate (Figure 1B). The insertion domain of Scp1 forms a 3-stranded β sheet, and sequence conservation indicates that it must be similar in Fcp1, except for the two loops of the sheet that are longer in yeast Fcp1 (Figure 1B). The insertion domain of PSP, β-PGM, and HAD is involved in substrate binding, but it has a totally different structure (Figure 2), as expected for different substrate specificities. Together these findings suggest that the Fcp1/Scp1 pocket between the insertion domain and the active site binds the CTD and confers substrate specificity. Mutational analysis of the presumed substrate pocket could confirm this proposal.

Consistent with CTD binding to the specificity pocket, *S. pombe* Fcp1 activity requires at least four N-terminal and two C-terminal CTD residues flanking phosphoserine 2, and single alanine mutations of the flanking Tyr1 and Pro3 decrease activity 6-fold (Hausmann et al., 2004). However, Fcp1 and Scp1 can dephosphorylate both Ser5 and Ser2, with some preference for either of these serines (Hausmann and Shuman, 2002; Lin et al., 2002a; Yeo et al., 2003). Since both serines are flanked on the C-terminal side by a proline, we speculate that the CTD phosphatases bind the adjacent prolines Pro3 or Pro6, and preferential dephosphorylation of Ser2 and Ser5 is achieved by binding to other nearby residues. Indeed, the Pro3 side chain binds to a hydrophobic pocket in the known CTD peptide complex structures (Fabrega et al., 2003; Verdecia et al., 2000), and specific recognition of a flanking proline is consistent with Fcp1 inhibition by the prolyl isomerase Pin1 (Xu et al., 2003). A better understanding of CTD specificity, however, requires structure determination of Fcp1 or Scp1 in complex with a phosphorylated CTD peptide.

The Pol II Subcomplex Rpb4/7 Recruits Fcp1

Specificity of Fcp1 for the CTD may not only result from direct recognition of the CTD residues, but also from binding of Fcp1 to a docking site on Pol II that is distinct from the CTD (Chambers et al., 1995). To investigate if the docking site is on the 10-subunit core of Pol II or on the heterodimeric polymerase subcomplex Rpb4/7, we tried to reconstitute complexes of Fcp1 with the Pol II core and complete Pol II (the Pol II core plus recombinant Rpb4/7) (Armache et al., 2003). To this end, we incubated pure endogenous yeast Pol II core with an excess of recombinant Fcp1c, either in the presence or the absence of an additional excess of recombinant Rpb4/7 complex, and subjected the samples to size exclusion chromatography. In both cases, two separated peaks were obtained. In the absence of Rpb4/7, the first peak corresponded to the Pol II core alone, and the second peak contained Fcp1c. In the presence of Rpb4/7, however, the first peak contained all 12 subunits of Pol II and Fcp1c, representing a reconstituted Pol II-Fcp1c complex (Figure 5). Since the formation of a Pol II-Fcp1c complex relied on the presence of Rpb4/7, we conclude that Fcp1 interacts with Pol II mainly via the Rpb4/7 subcomplex. The direct interaction of Fcp1 with Pol II via the Rpb4/7 complex may result in functional differences between Fcp1 and Scp1, but further studies are required to analyze these.

Rpb4/7 is located directly adjacent to the last ordered residues of the largest Pol II subunit, which form the beginning of a linker to the disordered CTD, and may therefore recruit Fcp1 to the phosphorylated CTD (Figure 6) (Armache et al., 2003; Bushnell and Kornberg, 2003). Stable association of Fcp1 with Pol II in vivo may, however, require additional interactions (Figure 6). In particular, Fcp1 binds the phosphorylated CTD (Yu et al., 2003) and the polymerase-associated general transcription factor TFIIF, which also stimulates Fcp1 activity (Chambers et al., 1995). The C-terminal end of Fcp1 includes a short helix that interacts with the large subunit

of TFIIF (Archambault et al., 1998; Kamada et al., 2003; Nguyen et al., 2003). These multiple interactions may play a role in regulating Fcp1 activity during the transcription cycle and may also allow for dephosphorylation of the repetitive CTD in a pseudo-processive manner, although isolated Fcp1 is a distributive enzyme, apparently associating with the CTD and dissociating from it for each catalytic cycle (Hausmann et al., 2004).

Conclusions

The structure of the CTD phosphatase Scp1 reveals a core fold similar to that of other enzymes of the DXDX(T/V) superfamily that are essentially unrelated in sequence, and it is a good model for the catalytic FCPH domain of all Fcp1 and Scp1 enzymes. Functional and crystallographic data show that catalysis involves the metal-assisted phosphorylation of the first aspartate in the DXDX(T/V) signature motif. A conserved hydrophobic pocket formed between the active site and the Fcp1/Scp1-specific insertion domain is likely involved in CTD recognition. Specificity of Fcp1 for the Pol II CTD may additionally arise from multiple interactions with the Pol II machinery, including docking of Fcp1 to the Rpb4/7 subcomplex, which is located adjacent to a protein linker connecting to the CTD. Open questions on CTD phosphatase and Pol II recycling include details of the interactions between the phosphatase, Pol II, and the CTD, and the timing of these transient protein-protein interactions during the transcription cycle.

Experimental Procedures

Cloning and Protein Purification

The gene for human Scp1 (or CTDSP1) was amplified from HUVEC cDNA and subcloned into pET21b vector (Stratagene). Variants of human Scp1 were expressed for 15 hr at 20°C in *E. coli* BL21 (DE3) RIL (Stratagene). Cells were harvested by centrifugation and resuspended in buffer A (300 mM NaCl, 50 mM Tris [pH 8.0], 5% glycerol, 10 mM β-mercaptoethanol). Cells were lysed by sonication. After centrifugation, the supernatant was loaded onto a 2 ml Ni-NTA column (Qiagen) equilibrated with buffer A. After washing with 10 ml of buffer A, bound protein was eluted with buffer A containing 200 mM imidazole and, in the case of full-length Scp1, was further purified by anion exchange chromatography (MonoQ, Amersham). The column was equilibrated with buffer B (50 mM NaCl, 50 mM Tris [pH 7.5], 5% glycerol, 3 mM DTT), and the protein was eluted with a gradient of 10 column volumes from 50 mM to 1 M NaCl. After concentration, the sample was applied to a Superose-12 HR gel filtration column (Amersham) equilibrated with buffer C (150 mM NaCl, 10 mM HEPES [pH 7.5], 3 mM DTT). Pooled peak fractions were concentrated for crystallization to 20 mg ml^{-1}. Variants of *Saccharomyces cerevisiae* Fcp1 were obtained by PCR amplification of the corresponding regions in the gene from yeast genomic DNA and were subcloned into pET21b (Stratagene). Expression and purification of variant Fcp1c (residues 168–606) were performed essentially as for Scp1. The variant Scp1ΔN, comprising residues 77–261, with a C-terminal HisTag was subcloned into pET21b (Stratagene), expressed in *E. coli*, and purified as described above, but without the anion exchange step. For MAD phasing, the Scp1ΔN mutant L165M/L205M, which contains two additional methionines at positions of conserved hydrophobic residues, was constructed with the PCR overlap extension method, and selenomethionine was incorporated as described (Meinhart et al., 2003).

Crystallization and Structure Determination

Samples were crystallized at 20°C with the hanging drop method, and 20%–30% PEG 3300 or PEG 6000 or PEG 8000, 200 mM NH_4OAc, 100 mM citrate buffer (pH 5.6), 5 mM DTT was used as reservoir solution. The thin plate-like crystals grew to a maximum size of 400 × 150 × 5 μm. To the drops containing crystals, 100 μl cryo solution (25% PEG 3300 or PEG 6000 or PEG 8000, 15% PEG 400, 200 mM NH_4OAc, 100 mM citrate buffer [pH 5.6], 5 mM DTT) was added, and, after 10 min, crystals were flash-cooled by plunging them into liquid nitrogen. A MAD experiment was performed on crystals from the selenomethionine-labeled L165M/L205M double mutant at the Swiss Light Source (Table 1). Data were processed with DENZO and SCALEPACK (Mutant L165M/L205M SeMet MAD data; Otwinowski and Minor, 1996) or with XDS (wild-type beryllofluoride data; Kabsch, 1993). The crystals belong to space group $P2_12_12$. SOLVE (Terwilliger, 2002) was used for detection of five selenium sites (M76, M142, M165, M205, and M229) and for MAD phasing (Z score = 43.3, FOM = 0.62). Phases were further improved by SHARP (La Fortelle and Bricogne, 1997). The resulting electron density map allowed building of an atomic model, which was refined with CNS (Brunger et al., 1998) against the remote wavelength dataset to 2.3 Å resolution (Table 1). To trap the beryllofluoride inhibitor in the active site, preformed wild-type crystals were soaked with 5 mM $BeCl_2$/24 mM NaF for 10 min before flash-cooling and data collection at the Swiss Light Source. After phasing with the refined model, clear difference density at residue D96 indicated the presence of beryllofluoride. Except for minor changes in the water structure in the active site, the structure is essentially identical and was refined to 2.1 Å resolution (Table 1). In both refined structures, 99% of the residues fall in allowed and additionally allowed regions of the Ramachandran plot, and none of the residues are in disallowed regions.

Phosphatase Kinetics and Inhibitor Studies

A discontinuous, indirect spectrophotometric assay for *p*-nitrophenol (pNP) was used to determine enzymatic parameters. Reactions were carried out at 37°C in a total volume of 100 μl and contained 50 mM Tris-acetate [pH 5.5], 10 mM $MgCl_2$, 0.19 μM Scp1 (or 0.1 μM Fcp1), and varying concentrations of *p*-nitrophenylphosphate (pNPP). Reactions were initiated by the addition of the enzyme and were quenched subsequently in 30-s or 60-s increments by the addition of 900 μl 1 M sodium carbonate. Prior to enzyme addition, reaction mixtures were incubated for 5 min to allow for temperature equilibration. Rates of pNP release were determined by monitoring absorbance at 405 nm and interpolating the value to a pNP standard curve. A minimum of five colinear data points for each substrate concentration were used to carry out linear fits to obtain the initial reaction velocity (not shown). Velocities were plotted against substrate concentration to obtain standard Michaelis-Menten curves (not shown). To derive K_M and k_{cat}, velocities were plotted against velocity divided by substrate concentration according to Eadie-Hofstee (see Figure 3A). To obtain the molar extinction coefficient ε, a standard curve was determined by measuring the absorbance at 405 nm of known pNP concentrations (Sigma) in the assay mixture. This plot was linear between 5 and 150 mM pNP and yielded $\epsilon = 18{,}300\ (M \times cm)^{-1}$. To study the effect of various inhibitors on phosphatase reactions, 100 μl reaction mixtures containing 10 mM pNPP and additional different phosphatase inhibitors were incubated for 30 min at 37°C and quenched with 900 μl 1 M sodium carbonate, and the amount of released pNP was determined by measuring the absorbance at 405 nm with the use of the pNP standard curve (compare with Figure 3B).

Reconstitution of a Pol II-Fcp1 Complex

For reconstitution of a Pol II-Fcp1 complex, a total of 0.5 mg of 10-subunit core Pol II in 40 mM ammonium sulfate, 10 μM zinc chloride, 5 mM HEPES [pH 7.3], and 5 mM DTT was incubated with a 5-fold excess of recombinant Rpb4/7 subcomplex at 20°C for 45 min, followed by a 5-fold excess of Fcp1 with respect to Rpb4/7 and an additional incubation of 2 hr. The Pol II-Fcp1 complex was separated from free Rpb4/7 and Fcp1 by gel filtration on a Superose 6 column (Amersham). Pol II and Rpb4/7 were prepared as described (Armache et al., 2003).

Acknowledgments

Part of this work was performed at the Swiss Light Source, Paul Scherrer Institut, Villigen, Switzerland. We thank C. Schulze-Briese

and the staff of beamline X06SA of the Swiss Light Source for help. We thank G. A. Vlastos and A. K. Hatzopoulos for providing human cDNA. We thank K. Armache and H. Kettenberger for help with data collection. P.C. is supported by the Deutsche Forschungsgemeinschaft, the EMBO Young Investigator Programme, and the Fonds der chemischen Industrie. T.K. is supported by a Boehringer Ingelheim Fonds Ph.D. fellowship. A.M. is supported by EMBO postdoctoral fellowship ALTF-399-2003.

Received: April 2, 2004
Revised: May 21, 2004
Accepted: May 26, 2004
Published: August 12, 2004

References

Archambault, J., Chambers, R.S., Kobor, M.S., Ho, Y., Cartier, M., Bolotin, D., Andrews, B., Kane, C.M., and Greenblatt, J. (1997). An essential component of a C-terminal domain phosphatase that interacts with transcription factor IIF in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA *94*, 14300–14305.

Archambault, J., Pan, G., Dahmus, G.K., Cartier, M., Marshall, N., Zhang, S., Dahmus, M.E., and Greenblatt, J. (1998). FCP1, the RAP74-interacting subunit of a human protein phosphatase that dephosphorylates the carboxyl-terminal domain of RNA polymerase IIO. J. Biol. Chem. *273*, 27593–27601.

Armache, K.-J., Kettenberger, H., and Cramer, P. (2003). Architecture of the initiation-competent 12-subunit RNA polymerase II. Proc. Natl. Acad. Sci. USA *100*, 6964–6968.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.

Bushnell, D.A., and Kornberg, R.D. (2003). Complete RNA polymerase II at 4.1 Å resolution: implications for the initiation of transcription. Proc. Natl. Acad. Sci. USA *100*, 6969–6972.

Chabre, M. (1990). Aluminofluoride and beryllofluoride complexes: a new phosphate analogs in enzymology. Trends Biochem. Sci. *15*, 6–10.

Chambers, R.S., and Dahmus, M.E. (1994). Purification and characterization of a phosphatase from HeLa cells which dephosphorylates the C-terminal domain of RNA polymerase II. J. Biol. Chem. *269*, 26243–26248.

Chambers, R.S., and Kane, C.M. (1996). Purification and characterization of an RNA polymerase II phosphatase from yeast. J. Biol. Chem. *271*, 24498–24504.

Chambers, R.S., Wang, B.Q., Burton, Z.F., and Dahmus, M.E. (1995). The activity of COOH-terminal domain phosphatase is regulated by a docking site on RNA polymerase II and the by the general transcription factors IIF and IIB. J. Biol. Chem. *270*, 14962–14969.

Cho, H., Kim, T.K., Mancebo, H., Lane, W.S., Flores, O., and Reinberg, D. (1999). A protein phosphatase functions to recycle RNA polymerase II. Genes Dev. *13*, 1540–1552.

Cho, H., Wang, W., Kim, R., Yokota, H., Damo, S., Kim, S.H., Wemmer, D., Kustu, S., and Yan, D. (2001). BeF(3)(−) acts as a phosphate analog in proteins phosphorylated on aspartate: structure of a BeF(3)(−) complex with phosphoserine phosphatase. Proc. Natl. Acad. Sci. USA *98*, 8525–8530.

Collet, J.F., Stroobant, V., Pirard, M., Delpierre, G., and Van Schaftingen, E. (1998). A new class of phosphotransferases phosphorylated on an aspartate residue in an amino-terminal DXDX(T/V) motif. J. Biol. Chem. *273*, 14107–14112.

Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. Science *292*, 1863–1876.

Fabrega, C., Shen, V., Shuman, S., and Lima, C.D. (2003). Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. Mol. Cell *11*, 1549–1561.

Hausmann, S., and Shuman, S. (2002). Characterization of the CTD phosphatase Fcp1 from fission yeast. Preferential dephosphorylation of serine 2 versus serine 5. J. Biol. Chem. *277*, 21213–21220.

Hausmann, S., and Shuman, S. (2003). Defining the active site of *Schizosaccharomyces pombe* C-terminal domain phosphatase Fcp1. J. Biol. Chem. *278*, 13627–13632.

Hausmann, S., Erdjument-Bromage, H., and Shuman, S. (2004). *Schizosaccharomyces pombe* CTD phosphatase Fcp1: distributive mechanism, minimal CTD substrate, and active site mapping. J. Biol. Chem. *279*, 10892–10900.

Hirose, Y., and Manley, J.L. (2000). RNA polymerase II and the integration of nuclear events. Genes Dev. *14*, 1415–1429.

Holm, L., and Sander, C. (1995). Dali: a network tool for protein structure comparison. Trends Biochem. Sci. *20*, 478–480.

Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J. Appl. Crystallogr. *26*, 795–800.

Kamada, K., Roeder, R.G., and Burley, S.K. (2003). Molecular mechanism of recruitment of TFIIF-associating RNA polymerase C-terminal domain phosphatase (FCP1) by transcription factor IIF. Proc. Natl. Acad. Sci. USA *100*, 2296–2299.

Kimura, M., Suzuki, H., and Ishihama, A. (2002). Formation of a carboxy-terminal domain phosphatase (Fcp1)/TFIIF/RNA polymerase II (pol II) complex in *Schizosaccharomyces pombe* involves direct interaction between Fcp1 and the Rpb4 subunit of pol II. Mol. Cell. Biol. *22*, 1577–1588.

Kobor, M.S., Archambault, J., Lester, W., Holstege, F.C., Gileadi, O., Jansma, D.B., Jennings, E.G., Kouyoumdjian, F., Davidson, A.R., Young, R.A., et al. (1999). An unusual eukaryotic protein phosphatase required for transcription by RNA polymerase II and CTD dephosphorylation in *S. cerevisiae*. Mol. Cell *4*, 55–62.

La Fortelle, E.D., and Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. Methods Enzymol. *276*, 472–494.

Lahiri, S.D., Zhang, G., Dunaway-Mariano, D., and Allen, K.N. (2003). The pentacovalent phosphorus intermediate of a phosphoryl transfer reaction. Science *299*, 2067–2071.

Lehmann, A.L., and Dahmus, M.E. (2000). The sensitivity of RNA polymerase II in elongation complexes to C-terminal domain phosphatase. J. Biol. Chem. *275*, 14923–14932.

Lin, P.S., Dubois, M.F., and Dahmus, M.E. (2002a). TFIIF-associating carboxyl-terminal domain phosphatase dephosphorylates phosphoserines 2 and 5 of RNA polymerase II. J. Biol. Chem. *277*, 45949–45956.

Lin, P.S., Marshall, N.F., and Dahmus, M.E. (2002b). CTD phosphatase: role in RNA polymerase II cycling and the regulation of transcript elongation. Prog. Nucleic Acid Res. Mol. Biol. *72*, 333–365.

Meinhart, A., Blobel, J., and Cramer, P. (2003). An extended winged helix domain in general transcription factor E/IIEalpha. J. Biol. Chem. *278*, 48267–48274.

Nguyen, B.D., Abbott, K.L., Potempa, K., Kobor, M.S., Archambault, J., Greenblatt, J., Legault, P., and Omichinski, J.G. (2003). NMR structure of a complex containing the TFIIF subunit RAP74 and the RNA polymerase II carboxyl-terminal domain phosphatase FCP1. Proc. Natl. Acad. Sci. USA *100*, 5688–5693.

Orphanides, G., and Reinberg, D. (2002). A unified theory of gene expression. Cell *108*, 439–451.

Otwinowski, Z., and Minor, W. (1996). Processing of X-ray diffraction data collected in oscillation mode. Meth. Enzym. *276*, 307–326.

Proudfoot, N.J., Furger, A., and Dye, M.J. (2002). Integrating mRNA processing with transcription. Cell *108*, 501–512.

Ridder, I.S., Rozeboom, H.J., Kalk, K.H., and Dijkstra, B.W. (1999). Crystal structures of intermediates in the dehalogenation of haloalkanoates by L-2-haloacid dehalogenase. J. Biol. Chem. *274*, 30672–30678.

Terwilliger, T.C. (2002). Automated structure solution, density modification and model building. Acta Crystallogr. D Biol. Crystallogr. *58*, 1937–1940.

Verdecia, M.A., Bowman, M.E., Lu, K.P., Hunter, T., and Noel, J.P.

(2000). Structural basis for phosphoserine-proline recognition by group IV WW domains. Nat. Struct. Biol. *7*, 639–643.

Wang, W., Kim, R., Jancarik, J., Yokota, H., and Kim, S.H. (2001). Crystal structure of phosphoserine phosphatase from *Methanococcus jannaschii*, a hyperthermophile, at 1.8 Å resolution. Structure *9*, 65–71.

Williams, R.S., Green, R., and Glover, J.N. (2001). Crystal structure of the BRCT repeat region from the breast cancer-associated protein BRCA1. Nat. Struct. Biol. *8*, 838–842.

Xu, Y.X., Hirose, Y., Zhou, X.Z., Lu, K.P., and Manley, J.L. (2003). Pin1 modulates the structure and function of human RNA polymerase II. Genes Dev. *17*, 2765–2776.

Yan, D., Cho, H.S., Hastings, C.A., Igo, M.M., Lee, S.Y., Pelton, J.G., Stewart, V., Wemmer, D.E., and Kustu, S. (1999). Beryllofluoride mimics phosphorylation of NtrC and other bacterial response regulators. Proc. Natl. Acad. Sci. USA *96*, 14789–14794.

Yeo, M., Lin, P.S., Dahmus, M.E., and Gill, G.N. (2003). A novel RNA polymerase II C-terminal domain phosphatase that preferentially dephosphorylates serine 5. J. Biol. Chem. *278*, 26078–26085.

Yu, X., Chini, C.C., He, M., Mer, G., and Chen, J. (2003). The BRCT domain is a phospho-protein binding domain. Science *302*, 639–642.

Accession Numbers

Atomic coordinates and structure factors for the Scp1 and Scp1-BeF_3^- structures have been deposited in the Protein Data Bank with accession numbers 1T9Z and 1TA0, respectively.