Different Strategies for Carboxyl-terminal Domain (CTD) Recognition by Serine 5-specific CTD Phosphatases*  

Stéphane Hausmann, Hisashi Koïwa, Shankarling Krishnamurthy, Michael Hampsey, and Stewart Shuman  

The phosphorylated carboxyl-terminal domain (CTD) of RNA polymerase II, consisting of [(YSPTSPS)ₙ]ₙ heptad repeats, encodes information about the state of the transcriptional apparatus that can be conveyed to factors that regulate mRNA synthesis and processing. Here we describe how the CTD code is read by two classes of protein phosphatases, plant CPLs and yeast Ssu72, that specifically dephosphorylate Ser₅ in vitro. The CPLs and Ssu72 recognize entirely different positional cues in the CTD primary structure. Whereas the CPLs rely on Tyr¹ and Pro⁶ located on the upstream side of the Ser⁵-PO₄ target site, Ssu72 recognizes Thr⁴ and Pro⁶ flanking the target Ser⁵-PO₄ plus the downstream Tyr¹ residue of the adjacent heptad. We surmise that the reading of the CTD code does not obey uniform rules with respect to the location and phasing of specificity determinants. Thus, CTD code, like the CTD structure, is plastic.

The carboxyl-terminal domain (CTD)² of the largest subunit of RNA polymerase II is a landing pad for proteins and multiprotein complexes that regulate transcription and catalyze mRNA processing. The CTD is composed of a tandemly repeated heptapeptide of consensus sequence ¹YSPTSPS². CTD positions Ser² and Ser⁵ undergo waves of phosphorylation and dephosphorylation during the transcription cycle. The potential complexity of the CTD serine phosphorylation array comprises 4ⁿ different structures, where n is the number of heptad repeats (n varies from 15 in microsporidia to 52 in mammals). Thus, the CTD is a vast repository of information about the state of the transcriptional apparatus (a so-called “CTD code”)¹ that can be read by trans-acting factors. Studies of the interaction of mRNA capping enzymes with the phosphorylated CTD have illuminated its atomic structure and how CTD length, amino acid sequence, and the phosphorylation array influence CTD-PO₄ effector functions. ² Three important points about the CTD code have emerged: (i) the phosphorylated CTD is structurally plastic and can assume markedly different conformations depending on its binding partner (4 – 7); (ii) CTD primary structure is recognized independently of CTD phosphorylation state (2, 3, 8, 9); and (iii) encoded CTD-PO₄ information can be assembled from multiple noncontiguous repeats (4).

The information content of the CTD reflects the instantaneous balance between the activities of CTD kinases and CTD phosphatases at either some or all of the serine phosphorylation sites. A subject of intense interest (and debate) is whether and how CTD kinases and phosphatases are themselves responsive to specificity determinants within the CTD. Among the CTD phosphatases, some enzymes fail to discriminate between Ser²-PO₄ and Ser⁵-PO₄ substrates (9, 10), whereas others display a preference (2- to 10-fold) for either Ser²-PO₄ or Ser⁵-PO₄ (11, 12). Recently, we described two types of CTD phosphatases that act exclusively on Ser⁵-PO₄: (i) the paralogous plant enzymes CPL1 and CPL2, which were identified genetically as regulators of osmotic stress and abscisic acid-responsive transcription in Arabidopsis thaliana (13, 14), and (ii) yeast Ssu72, which interacts with TFIIB and with proteins involved in RNA 3’ -end formation (15–20).

CPL1/2 and Ssu72 belong to different enzyme families. The plant CPL1/2 proteins, like the prototypical CTD phosphatase Fcp1, are members of the DDX1 superfamily of metal-dependent phosphotransferases that act via an aspartyl-phosphoenzyme intermediate (8, 11, 13, 21–23). Ssu72 belongs to the CXXXXR superfamily of metal-independent phosphohydrolases that act via a cysteiny1 phosphoenzyme intermediate (15, 24, 25). This situation, whereby nature has selected two entirely different structural solutions to perform the same site-specific chemical transformation of the CTD, provides an opportunity to address key questions about how the CTD code is read. Here we demonstrate that CPL1/2 and Ssu72 rely on different positional cues in the CTD primary structure to recognize and hydrolyze Ser⁵-PO₄.

**EXPERIMENTAL PROCEDURES**

Recombinant Proteins—The CPL1-(1–646) and CPL2-(1–649) coding sequences were amplified by PCR using primers designed to introduce BglII sites at the start codon and immediately 3′ of the stop codon. The BglII-digested PCR products were insert into pET28-His₁₀Smt3 that had been linearized with BamHI. The resulting expression plasmids pET-His₁₀Smt3-CPL1-(1–646) and pET-His₁₀Smt3-CPL2-(1–649) encode the respective CPL proteins fused in-frame to an amino-terminal His₁₀Smt3 domain consisting of a His₁₀ leader (MGHHHHHHHHHHHHHHSSGHIEGRH) followed by the 98-amino acid Saccharomyces cerevisiae Smt3 protein and a single serine. (Smt3 is the yeast ortholog of the small ubiquitin-like modifier SUMO.) pET-His₁₀Smt3 plasmids encoding mutated versions CPL1-(1–646)-D161A and CPL2-(1–649)-D144A were generated by two-stage overlap extension PCR and cloning of the BglII-digested mutated PCR products into pET28-His₁₀Smt3.

The pET-His₁₀Smt3-CPL plasmids were transformed into Escherichia coli BL21-CodonPlus(DE3). Cultures (500 ml) of derived from single transformants were grown at 37 °C in Luria Bertani medium containing 50 µg/ml kanamycin and 50 µg/ml chloramphenicol until the A₆₀₀ reached 0.6. The cultures were adjusted to 0.2 mM isopropyl-1-thio-β-D-galactopyranoside and incubation was continued for 20 h at 17 °C. Cells were harvested by centrifugation and stored at

---

* This work was supported by National Institutes of Health Grants GM52470 (to S. S.) and GM68887 (to M. H.) and National Science Foundation Grant MCB0421889 (to H. K.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

1. An American Cancer Society Research Professor. To whom correspondence should be addressed. E-mail: s-shuman@ski.mskcc.org.

2. The abbreviations used are: CTD, carboxyl-terminal domain; CDS, CTD-docking site; BSA, bovine serum albumin.
Serine 5-specific CTD Phosphatases

−80 °C. All subsequent procedures were performed at 4 °C. Thawed bacteria were resuspended in 25 ml of buffer A (50 mM Tris–HCl, pH 8.0, 200 mM NaCl, 10% glycerol). Phenylmethylsulfonyl fluoride and lysosome were added to final concentrations of 500 μM and 100 μg/ml, respectively. After incubation on ice for 30 min, Triton X-100 was added to a final concentration of 0.1%, and the lysate was sonicated to reduce viscosity. Insoluble material was removed by centrifugation. The soluble extracts were mixed for 30 min with 1 ml of Ni²⁺-nitritotriacetate acid-agarose (Qiagen) that had been equilibrated with buffer A containing 0.1% Triton X-100. The resin was recovered by centrifugation, resuspended in buffer A, and poured into columns. The columns were washed with 10 ml of 20 mM imidazole in buffer A and then eluted stepwise with 1.5 ml of buffer A containing 50, 100, 250, and 500 mM imidazole. The polypeptide compositions of the column fractions were monitored by SDS-PAGE. The recombinant His₆-Smt3–CPL polypeptides were recovered predominantly in the 250 mM imidazole fractions. The 250 mM imidazole eluates were dialyzed against buffer A containing 2 mM dithiothreitol and 0.01% Triton X-100 and then stored at −80 °C. The CPL1 and CPL2 concentrations were determined by SDS-PAGE analysis of serial dilutions of the CPL preparations in parallel with serial dilutions of a BSA standard. The gels were stained with Coomassie Blue, and the staining intensities of the His₆-Smt3–CPL and BSA polypeptides were quantified using a Fujifilm FLA-5000 digital imaging and analysis system. CPL1 and CPL2 concentrations were calculated by interpolation to the BSA standard curve.

Wild-type Ssu72 and the C15S mutant were produced in E. coli as glutathione S-transferase fusions and purified from soluble bacterial extracts by glutathione-Sepharose affinity chromatography as described (15). Protein concentration was measured with the Bio-Rad dye reagent.

Velocity Sedimentation—Aliquots (40 μg) of the nickel–agarose preparations of CPL1 and CPL2 were mixed with catalase (45 μg), BSA (45 μg), and cytochrome c (45 μg), and the mixtures were applied to a 15–30% glycerol gradient containing 50 mM Tris–HCl (pH 8.0), 0.2 M NaCl, 1 mM EDTA, 2 mM dithiothreitol, 0.01% Triton X-100. The gradients were centrifuged in a SW50 rotor at 50,000 rpm for 17 h at 4 °C. Fractions (~0.26 ml) were collected from the bottoms of the tubes.

CTD Phosphopeptides—CTD Ser-PO₄ peptides were synthesized and purified by the Sloan–Kettering Microchemistry Core Laboratory as described previously (2, 8). The peptides were dissolved in 10 mM Tris–HCl (pH 7.4), 1 mM EDTA and stored at 4 °C. The molar concentrations of the phosphopeptides were initially estimated from the absorbance at 274 nm using an extinction coefficient of 1.4 × 10³ M⁻¹ cm⁻¹ for tyrosine. The content of Ser-PO₄ was then determined for each peptide, measuring the release of inorganic phosphate after digestion with calf intestinal phosphatase as described (8).

CTD Phosphatase Assay—Reaction mixtures (25 μl) containing 50 mM Tris acetate (pH 5.5), 10 mM MgCl₂, CTD phosphopeptide, and CPL1 or CPL2 were incubated for 60 min at 37 °C. Reaction mixtures (25 μl) containing 50 mM Tris acetate (pH 7.0), 5 mM dithiothreitol, CTD phosphopeptide, and Ssu72 as specified were incubated for 120 min at 30 °C. The reactions were quenched by adding 0.5 ml of mala-chite green reagent (BIOMOL Research Laboratories, Plymouth Meeting, PA). Release of phosphate was determined by measuring A₆₃₀ and interpolating the value to a phosphate standard curve. The amounts of CTD substrate in the CTD phosphatase reactions are expressed as input phosphoserine, as determined by CIP digestion performed in parallel with each assay.

RESULTS AND DISCUSSION

Physical and Biochemical Characterization of CPL1 and CPL2—CPL1 and CPL2 were shown previously to dephosphorylate in vitro Ser⁵ of the Arabidopsis CTD (consisting of 34 heptad repeats), but not Ser² (13). Exclusive hydrolysis of Ser³-PO₄, but not Ser²-PO₄, was also demonstrated using defined synthetic CTD phosphopeptides. Physical and biochemical characterization of the plant phosphatases was hampered initially by the poor solubility of the recombinant proteins in E. coli and their susceptibility to proteolysis in vivo. Deletion analysis showed that the carboxyl-terminal segments of CPL1 and CPL2 could be removed without diminishing phosphatase activity or affecting the Ser⁵ specificity (13). Here, we produced the catalytically active amino-terminal domains as His₆-Smt3–CPL fusions; this maneuver improved their yield and solubility compared with previous expression strategies. The His₆-Smt3-tagged proteins CPL1-(1–646) and CPL2-(1–649) were isolated from soluble bacterial extracts by adsorption to nickel-agarose and elution with imidazole. SDS-PAGE revealed the presence of polypeptides corresponding to the intact fusion proteins (Fig. 1A, arrows). The apparent sizes of the Smt3-CPL fusion proteins by SDS-PAGE (97–98 kDa) were larger than their calculated molecular masses of 87 kDa; this is because the Smt3 domain migrates aberrantly during SDS-PAGE, appearing ~10 kDa larger than its predicted size. The CPL1 and CPL2 preparation contained several polypeptides in the 60–90-kDa range and a cluster of smaller polypeptides migrating at 20–27 kDa (the latter correspond to His₆-Smt3- and His₆-Smt3-peptide fusions arising via proteolysis) (Fig. 1A and data not shown).

The quaternary structure of recombinant CPL1 and CPL2 was investigated by zonal velocity sedimentation through a 15–30% glycerol gradient (Fig. 1, B and C). Marker protein catalases (native size 248 kDa), BSA (66 kDa), and cytochrome c (12 kDa) were included as internal standards in the gradient. The tagged CPL1 and CPL2 proteins sedimented as discrete peaks coincident with BSA. CPL1 and CPL2 were clearly separated from the cluster of low molecular weight polypeptides, which cosedimented with cytochrome c. The CTD Ser⁵ phosphatase activity profile (measured by the release of Pi from a synthetic phosphopeptide) paralleled the abundance of the intact CPL1 and CPL2 polypeptides. These results are consistent with a monomeric quaternary structure for the catalytic domains of CPL1 and CPL2.

CPL1- and CPL2-catalyzed release of Pi from a 28-amino acid tetraheptad CTD Ser³-PO₄ substrate was proportional to enzyme concentration (expressed as the amount of full-length CPL fusion protein); 95% of the input Ser³-PO₄ residues were hydrolyzed at saturating CPL levels (Fig. 2A). From the slope of the titration curves we estimated that CPL1 and CPL2 hydrolyzed a 6400- and 9300-fold molar excess of Ser³-PO₄/ enzyme, respectively, during the 60-min reaction. Neither CPL1 nor CPL2 catalyzed Pi release from a tetraheptad CTD Ser²-PO₄ substrate (Fig. 2A). A kinetic analysis is shown in Fig. 2B of the release of CPL2 with the tetraheptad CTD-Ser³-PO₄ substrate at two different levels of input enzyme. Product accumulated steadily with a pseudo-first order profile. The initial rate was proportional to enzyme concentration; the turnover number was 13 s⁻¹.

Reaction of CPL1 and CPL2 with a 14-amino acid diheptad CTD Ser³-PO₄ substrate resulted in preferential conversion of the input phosphoserine in vivo (Fig. 2C). 85% of the input phosphoserine was hydrolyzed at saturating enzyme levels (Fig. 2C and data not shown). CPL1 and CPL2 released an ~11,000-fold molar excess of Pi/enzyme during the 60-min reaction. In contrast, CPL1 and CPL2 were inert in hydrolyzing a diheptad CTD Ser³-PO₄ substrate YSPTSPYSPTSPS (Fig. 2C). This experiment showed that two heptad repeats suffice for Ser³-specific CTD phosphatase activity.

S. Hausmann, unpublished information.
CPL1 and CPL2 are putative members of the DXDXT phosphotransferase family. The signature feature of these enzymes is the formation of an intermediate in which phosphate is attached covalently to the first Asp residue of the DXDXT motif (23). We mutated the presumptive Asp nucleophiles of CPL1 (Asp161) and CPL2 (Asp144) to alanine, produced the His<sub>10</sub>Smt3-CPL1-(1–646) and His<sub>10</sub>Smt3-CPL2-(1–649) polypeptides (arrows) were analyzed by SDS-PAGE. The gel was stained with Coomassie Blue dye. The positions and sizes (kDa) of marker polypeptides are indicated on the left. B and C, sedimentation analysis of CPL1 (B) and CPL2 (C) was performed as described under “Experimental Procedures.” Aliquots (20 μl) of odd-numbered glycerol gradient fractions were analyzed by SDS-PAGE. The Coomassie Blue-stained gels are shown (top panels). The phosphatase activity profiles are shown in the bottom panels. Phosphatase reaction mixtures containing 2 μl of a 1/8 dilution of the gradient fractions and 2.9 nmol CTD phosphopeptide SPSYSPTSPS were incubated for 60 min at 37 °C.

CPL1 and CPL2 are putative members of the DXDXT phosphotransferase family. The signature feature of these enzymes is the formation of an intermediate in which phosphate is attached covalently to the first Asp residue of the DXDXT motif (23). We mutated the presumptive Asp nucleophiles of CPL1 (Asp<sup>161</sup>) and CPL2 (Asp<sup>144</sup>) to alanine, produced the His<sub>10</sub>Smt3-CPL1-(1–646)-D161A and His<sub>10</sub>Smt3-CPL2-(1–649)-D144A proteins in bacteria, and isolated them from soluble extracts by affinity chromatography. The polypeptide compositions of the mutant proteins were virtually identical to those of the wild-type CPL1 and CPL2 (not shown). The CPL1-D161A and CPL2-D144A preparations displayed no detectable CTD-Ser<sup>5</sup> phosphatase activity at levels of input protein that were saturating for the respective wild-type enzymes (not shown). These results suggest that the plant CPLs are bona fide acyl-phosphatases and, in conjunction with the sedimentation data, prove that the observed CTD-Ser<sup>5</sup>-specific CTD phosphatase activity is intrinsic to the recombinant CPLs.

CTD Ser<sup>5</sup> Phosphatase Activity of Ssu72—The specificity of Ssu72 was characterized using a recombinant glutathione S-transferase-Ssu72 protein. The enzyme was expressed in bacteria and purified by affinity chromatography. The specific activity of Ssu72 was measured using a CTD phosphopeptide (SPSYSPTSPS) as substrate. The enzyme displayed no activity when the Ser residue was replaced by Thr (not shown). These results suggest that Ssu72 is a specific CTD-Ser<sup>5</sup> phosphatase.
Serine 5-specific CTD Phosphatases

CPLs and Ssu72 Recognize Different Specificity Determinants in the CTD—To probe the role of CTD primary structure in Ser5 phosphatase activity, we tested mutated versions of the tetraheptad CTD Ser5-PO4 peptide (YSPTSPS)4, wherein every Tyr1, Ser2, Pro3, Thr4, or Pro6 was replaced by alanine. The titration profiles of the reaction of CPL1 with the S2A, T4A and P6A substrates were similar to that of the wild-type CTD (Fig. 4A). However, the CPL1 activity profiles with the P3A and Y1A substrates displayed a shift to the right (Fig. 4A). The specific activities of CPL1 with the P3A and Y1A substrates were 16 and 4%, respectively, of the activity with the wild-type (YSPTSPS)4 substrate. CPL2 hydrolyzed the S2A and T4A CTDs as well as the wild-type CTD. Activity was reduced modestly by the P6A change (to 27% of wild-type) and drastically by the P3A (3%) and Y1A (1%) CTD mutations (Fig. 4B). These experiments demonstrate that Tyr1 and Pro6 are the critical determinants of CPL phosphatase activity at position Ser5.

Ssu72 responded quite differently to changes in CTD primary structure (Fig. 4C). The P3A change that was deleterious to the CPLs had virtually no ill effect on Ssu72, whereas the T4A mutation, which was well tolerated by the CPLs, was inimical to Ssu72, reducing its specific activity to 3% of the wild-type CTD level (Fig. 4C). Moreover, the P6A lesion abolished Ser5 phosphatase activity of Ssu72, though it had no effect on CPL1 and only a modest effect on CPL2. Ssu72 activity was reduced severely (to 4% of wild-type) by the Y1A change and was unaffected by S2A. We conclude that (i) Tyr1, Thr4, and Pro6 are the critical determinants of Ssu72 phosphatase, and (ii) different CTD phosphatases with the same exquisite specificity for Ser5-PO4 achieve their specificities by recognizing different structural cues in the CTD.

By titrating the wild-type (YSPTSPS)4 substrate against fixed amounts of CPL1, CPL2, and Ssu72, we determined apparent $K_m$ values of 60, 140, and 280 $\mu M$, respectively (not shown). These values are higher than the substrate concentrations (30–300 $\mu M$) used in the specific activity determinations in Fig. 4. Thus, we retested the activity of CPL1 and CPL2 with the wild-type, Y1A, P3A, and P6A tetraheptad Ser5-PO4 peptides at substrate concentrations (240–260 $\mu M$) in excess of the $K_m$ for wild-type peptide (Fig. 5). The hierarchy of mutational effects at higher substrate was similar to what was observed at limiting substrate. The extents of phosphate release by CPL1 with the P6A, P3A, and Y1A substrates were 96, 27, and 13%, respectively, of the activity with the wild-type (YSPTSPS)4 substrate (Fig. 5, left panel). Phosphate release by CPL2 with the P6A, P3A, and Y1A substrates was 64, 7, and 4% of activity with the wild-type peptide (middle panel). Thus, raising the CTD concentration above $K_m$ only slightly mitigated the effects of the Y1A or P3A changes on CPL1 activity, or the effects of Y1A, P3A, and P6A on CPL2 function, vis-à-vis their activity with the wild-type tetraheptad substrate. A parallel experiment with Ssu72 at the higher substrate concentrations showed that phosphate release with the Y1A, T4A, and P6A substrates was 8, 12, and 3% of its activity with the wild-type peptide (Fig. 5, right panel).

Effects of a Lysine Substitution at Position 7 of the CTD Heptad—The identity of the amino acid at position 7 of the CTD is either highly conserved (e.g. all serines in the Schizosaccharomyces pombe CTD (GenBank™ accession code CAB57941) and Encephalitozoon cuniculi CTD (GenBank™ accession code CAD26175)) or subject to variation from one heptad to another within the same CTD. In the case of human RNA polymerase

---

**FIGURE 3.** Ser5 specificity of Ssu72. **A**, reaction mixtures containing 29 $\mu M$ tetraheptad Ser5-PO4 substrate (2.9 nmol Ser-PO4) or 30 $\mu M$ tetraheptad Ser2-PO4 substrate (3.0 nmol Ser-PO4) and Ssu72 as specified were incubated for 2 h at 30°C. **B**, reaction mixtures containing (per 25 $\mu l$) 24 $\mu M$ tetraheptad Ser2-PO4 substrate (2.4 nmol Ser-PO4) or 32 $\mu M$ tetraheptad Ser5-PO4 (3.2 nmol Ser-PO4) substrate and 1.25 $\mu g$ of Ssu72 were incubated at 30°C. Aliquots (25 $\mu l$) were withdrawn at the times specified and quenched immediately. **C**, reaction mixtures containing the diheptad Ser5-PO4 substrate (5.1 nmol Ser-PO4) or diheptad Ser2-PO4 substrate (4.2 nmol Ser-PO4) and Ssu72 as specified were incubated for 2 h at 30°C. Aliquots (5 $\mu l$) of the glutathione S-transferase-tagged wild-type Ssu72 and C15S mutant proteins were analyzed by SDS-PAGE. The polypeptides were visualized by staining the gel with Coomasie Blue dye. The positions and sizes (kDa) of marker polypeptides are indicated on the left.
II (GenBank™ P24928), one-half (26/52) of the heptads have a consensus Ser7, and the others are substituted, most commonly with Lys7 (7 heptads) or Asn7 (5 heptads). The CTD of the malaria parasite Plasmodium falciparum (GenBank™ Z98551) is dominated by Lys7-containing heptads (11/15 repeats).

To gauge the role, if any, of the position 7 side chain as a CTD phosphatase specificity determinant, we tested a tetraheptad CTD Ser5-PO4 peptide (YSPTSPK)4 in which every Ser7 was replaced by lysine. The titration profiles of the reaction of CPL1 and CPL2 with the S7K substrates were similar to that of the wild-type CTD (Fig. 6, A and B). Ssu72 was ~30% as active with the S7K CTD substrate as it was with the wild-type CTD (Fig. 6C). Thus, neither the CPLs nor Ssu72 depend on Ser7 for their Ser5-specific phosphatase activity.

**Different Phasing of Specificity Determinants for CPLs and Ssu72—** Further insight into the location of the specificity determinants emerged from an analysis of the reaction of CPLs and Ssu72 with diheptad monophosphorylated CTDs, which contained Ser5-PO4 in either the amino-terminal heptad (YSPTSPKYSPTSPK) or the carboxyl-terminal heptad (YSPTSPKYSPTSPK). Ssu72 displayed a 10-fold preference for hydrolysis of Ser5-PO4 within the amino-terminal heptad (Fig. 7C). This result implies that the recognition of Tyr1 by Ssu72 occurs on the carboxyl side of the Ser5-PO4 moiety that is being hydrolyzed, i.e. Ssu72
has higher activity on the proximal repeat because Ssu72 recognizes the Thr⁴ and Pro⁶ flanking the Ser⁵-PO⁴ and the Tyr¹ in the distal repeat (see Fig. 9). Ssu72 is only weakly active on the distal repeat because there is no Tyr¹ available on the carboxyl side of the phosphoserine. From a substrate titration experiment at a fixed level of Ssu72, we determined an apparent $K_m$ of 850 μM for the YSPTSPSYSPTSPS peptide (data not shown); thus, the affinity of the enzyme for a tetraheptad CTD with four phosphoserines is ~3-fold higher than for a diheptad with a single phosphoserine.

CPL1 and CPL2 display the opposite preference; they are 2- to 3-fold more active in hydrolyzing Ser⁵-PO⁴ when it is located in the carboxyterminal heptad (Fig. 7, A and B). These results suggest that: (i) the CPLs recognize the Tyr¹ residue upstream of the phosphorylated Ser⁵ that is being hydrolyzed, and (ii) 2 amino acids downstream of Ser⁵-PO⁴ suffice. From substrate titration experiments, we determined apparent $K_m$ values of CPL1 and CPL2 for the preferred YSPTSPSYSPTSPS peptide of 100 and 570 μM, respectively (data not shown). The disfavored CPL substrate does contain a Tyr¹ upstream of Ser⁵-PO⁴ in the first repeat; perhaps the free amine of the terminal Tyr¹ (which adds a positive charge not present normally) might account for the lower activity of this substrate. Alternatively, 1 or more amino acids on the upstream side of Tyr¹ might contribute to substrate recognition, either directly or by limiting the flexibility of the Tyr¹ residue. To address these issues, we tested a series of incrementally truncated 12-mer (PTSPSYSPTSPS), 10-mer (SPSYSPTSPS), and 8-mer (SYSPTSPS) peptides containing a complete Ser⁵-PO⁴ heptad at their carboxyl termini and either 5, 3, or 1 amino acids from an upstream heptad (Fig. 8). CPL1 and CPL2 readily hydrolyzed the 12- and 10-mer substrates, whereas their activities with the 8-mer were reduced by factors of 2 and 3, respectively (Fig. 8). We surmise that CPL1/2 activity is optimal with as few as 3 amino acids on the amino-terminal side of Tyr¹. Apparent $K_m$ values of CPL1 and CPL2 for the 12-mer PTSPSYSPTSPS were 180 and 950 μM, respectively (not shown).

**Properties of the CTD Code**—Here we have begun to illuminate how the CTD code is read by two classes of protein phosphatases that specifically dephosphorylate Ser⁵. The instructive findings are that the
plant CPLs and yeast Ssu72 recognize entirely different constellations of specificity determinants in the CTD. Whereas the CPLs rely on Tyr1 and Pro3 located on the upstream side of the Ser5-PO4 target site, Ssu72 recognizes Thr4 and Pro6 flanking the target Ser5-PO4 plus the downstream Tyr1 residues of the adjacent heptad (Fig. 9). The two classes of phosphatases not only see different amino acids, they see them in entirely different registers with respect to Ser5-PO4.

The simple interpretation of our results is that: (i) the functionally relevant CTD structural elements comprise the respective binding sites for the CPLs and Ssu72, and (ii) those CTD elements are likely to be disposed on the surface of the CTD Ser5-PO4 substrate to which the phosphatase docks. The most plausible scenario is that the proper CTD conformation for catalysis is templated by interaction with the phosphatase via an induced fit mechanism, as shown for other CTD-binding proteins. Although we do not yet have an atomic structure of the CTD bound to either of the phosphatases studied here, we can make an educated guess as to what conformations the phosphatases might select for, based on principles emerging from the structure of the mRNA capping enzyme bound to the Ser5-PO4 CTD (4).

It is remarkable that the CTD segments and the CTD conformations at CDS1 and CDS2 of the capping enzyme satisfy reasonably well the specificity parameters elucidated here for Ssu72 and the CPLs, respectively. To wit, the Tyr1, Pro3, and Ser-PO4 moieties that comprise the recognition code for CPL1/2 project onto a common surface within CDS2, whereas Ser2 and Thr4, which are not needed for CPL1/2 activity, project away from the docking site (Fig. 9). Pro6 also projects onto the protein side of the interface in CDS2; this residue is not critical for CPL1 phosphatase activity but does play a modest role in CPL2 function. We envision that the CTD adopts a CDS2-like conformation in its interaction with the CPL1 and CPL2 phosphatases.

The Ser5-PO4, Pro6, and Tyr1 moieties that Ssu72 recognizes are located on the protein side of the interface in CDS1, whereas Ser2 and Pro3, to which Ssu72 is indifferent, project away from the protein surface (Fig. 9). We therefore speculate that the CTD adopts a CDS1-like conformation in its interaction with Ssu72. Although Thr4 does not contribute to the CTD-protein interface at CDS1 of the capping enzyme, the Thr O\(\gamma\) is in position to donate a hydrogen bond to the vicinal phosphate group at Ser5 (4).

The finding of at least two CTD coding elements that signal specific hydrolysis of Ser5-PO4 has implications for understanding CTD information content. First, the functional “footprints” of CPL1/2 and Ssu72 at Ser5-PO4 positions highlighted in shaded boxes. Reaction mixtures containing 3.1–3.3 nmol CTD peptide and CPL1 (A) or CPL2 (B) as specified were incubated for 60 min at 37 °C.

**FIGURE 8. Minimizing the substrate for plant CPLs.** The sequences of the CTD phosphopeptides are shown with Ser5-PO4 positions highlighted in shaded boxes. Reaction mixtures containing 3.1–3.3 nmol CTD peptide and CPL1 (A) or CPL2 (B) as specified were incubated for 60 min at 37 °C.

**FIGURE 9. CPLs and Ssu72 read distinct overlapping CTD codes to hydrolyze Ser5-PO4.** The CTD structural determinants of CPL and Ssu72 phosphatase activity are indicated by arrows above and below the CTD primary structure. The phosphorylated Ser5 position is colored red. It is presumed that the specificity determinants are likely to residue on a common face of the CTD with which the phosphatase interacts. The CTD requirements for CPL activity are consistent with the β-like conformation of the corresponding CTD segment bound to CTD-docking site 2 (CDS2) of yeast mRNA capping enzyme, which is illustrated at top right. The requirements for Ssu72 activity suggest a β-like conformation of the same CTD segment bound to CDS1 of the capping enzyme, which is shown at bottom left.

The Ser5-PO4, Pro6, and Tyr1 moieties that Ssu72 recognizes are located on the protein side of the interface in CDS1, whereas Ser2 and Pro3, to which Ssu72 is indifferent, project away from the protein surface (Fig. 9). We therefore speculate that the CTD adopts a CDS1-like conformation in its interaction with Ssu72. Although Thr4 does not contribute to the CTD-protein interface at CDS1 of the capping enzyme, the Thr O\(\gamma\) is in position to donate a hydrogen bond to the vicinal phosphate group at Ser5 (4).

The finding of at least two CTD coding elements that signal specific hydrolysis of Ser5-PO4, with Ser5-PO4, Pro6, and Tyr1 moieties that Ssu72 recognizes are located on the protein side of the interface in CDS1, whereas Ser2 and Pro3, to which Ssu72 is indifferent, project away from the protein surface (Fig. 9). We therefore speculate that the CTD adopts a CDS1-like conformation in its interaction with Ssu72. Although Thr4 does not contribute to the CTD-protein interface at CDS1 of the capping enzyme, the Thr O\(\gamma\) is in position to donate a hydrogen bond to the vicinal phosphate group at Ser5 (4).

The finding of at least two CTD coding elements that signal specific hydrolysis of Ser5-PO4, with SYSP, whereas CDS2 interacts with a downstream segment SYSP located two heptads away. Thus, the capping enzyme samples two distinct CTD structures that are phased in different registers with respect to Ser5-PO4. The CTD bound at sites CDS1 and CDS2 adopts a β-like extended conformation, whereby every other residue is oriented toward or away from the protein surface (Fig. 9).

Serine 5-specific CTD Phosphatases  NOVEMBER 11, 2005 • VOLUME 280 • NUMBER 45 • JOURNAL OF BIOLOGICAL CHEMISTRY 37687
Serine 5-specific CTD Phosphatases

Ssu72 is identified here as a stringently Pro6-directed phosphatase, whereas CPL1 is Pro6-independent and CPL2 is only modestly affected by elimination of Pro6. Third, even though the CTD is structurally plastic, there are likely to exist a set of CTD conformational states that are utilized reiteratively by different CTD-binding proteins, even where the proteins are themselves not structurally related (e.g. the CTD phosphatases and the mRNA capping enzyme). Fourth, the information content of the CTD can be amplified by blending multiple distinct CTD coding elements (such as those seen here for the CPLs and Ssu72) to form a bipartite, or even higher order, recognition site for CTD-associated factors (such as the capping enzyme).

REFERENCES
1. Buratowski, S. (2003) *Nature Struct. Biol.* 10, 679–680
2. Ho, C.K., and Shuman, S. (1999) *Mol. Cell* 3, 405–411
3. Pei, Y., Hausmann, S., Ho, C. K., Schwer, B., and Shuman, S. (2001) *J. Biol. Chem.* 276, 28075–28082
4. Fabrega, C., Shen, V., Shuman, S., and Lima, C. D. (2003) *Mol. Cell* 11, 1549–1561
5. Noble, C. G., Hollingsworth, D., Martin, S. R., Ennis-Adeniran, V., Smerdon, S. J., Kelly, G., Taylor, A. A., and Ramos, A. (2005) *Nature Struct. Mol. Biol.* 12, 144–151
6. Meinhart, A., and Cramer, P. (2004) *Nature* 430, 223–226
7. Verdecia, M. A., Bowman, M. E., Lu, K. P., Hunter, T., and Noel J. P. (2000) *Nature Struct. Biol.* 7, 639–643
8. Hausmann, S., Erdjument-Bromage, H., and Shuman, S. (2004) *J. Biol. Chem.* 279, 10892–10900
9. Hausmann, S., Schwer, B., and Shuman, S. (2004) *Biochemistry* 43, 7111–7120
10. Lin, P. S., Dubois, M. F., and Dahmus, M. E. (2002) *J. Biol. Chem.* 277, 45949–45956
11. Hausmann, S., and Shuman, S. (2002) *J. Biol. Chem.* 277, 21213–21220
12. Yeo, M., Lin, P. S., Dahmus, M. E., and Gill, G. N. (2003) *J. Biol. Chem.* 278, 26078–26085
13. Koiva, H., Hausmann, S., Bang, W. Y., Ueda, A., Kondo, N., Hiraguri, A., Fukuhara, T., Bahk, J. D., Yun, D. J., Bressan, R. A., Hasegawa, P. M., and Shuman, S. (2004) *Proc. Natl. Acad. Sci. U. S. A.* 101, 14539–14544
14. Koiva, H., Barb, A. W., Xiong, L., Li, F., McCully, M. G., Lee, B., Sokolchik, I., Zhu, J., Gong, Z., Reddy, M., Shakhshuu, A., Manabe, Y., Yokoi, S., Zhu, J. K., Bressan, R. A., and Hasegawa, P. M. (2002) *Proc. Natl. Acad. Sci. U. S. A.* 99, 10893–10898
15. Krishnamurthy, S., He, X., Reyes-Reyes, M., Moore, C., and Hampsey, M. (2004) *Mol. Cell* 14, 387–394
16. Sun, Z., and Hampsey, M. (1996) *Mol. Cell. Biol.* 16, 1557–1566
17. He, X., Khan, A. U., Cheng, H., Pappas, D. L., Hampsey, M., and Moore, C. L. (2003) *Genes Dev.* 17, 1030–1042
18. Dichtl, B., Blank, D., Ohnacker, M., Friedlin, A., Roeder, D., Langen, H., and Keller, W. (2002) *Mol. Cell* 10, 1139–1150
19. Nedea, E., He, X., Kim, M., Pootoolal, J. H., Zhong, G., Canadien, V., Hughes, T., Buratowski, S., Moore, C. L., and Greenblatt, J. (2003) *J. Biol. Chem.* 278, 33000–33010
20. Steinmetz, E. J., and Brow, D. A. (2003) *Mol. Cell. Biol.* 23, 6339–6349
21. Hausmann, S., and Shuman, S. (2003) *J. Biol. Chem.* 278, 13627–13632
22. Kamenski, T., Heilmeier, S., Meinhart, A., and Cramer, P. (2004) *Mol. Cell* 15, 399–407
23. Lahiri, S., Zhang, G., Dunaway-Mariano, D., and Allen, K. N. (2003) *Science* 299, 12067–2071
24. Ganem, C., Devaux, F., Torchet, C., Jacq, C., Quevillon-Cheruel, S., Labesse, G., Facca, C., and Faye, G. (2003) *EMBO J.* 22, 1588–1598
25. Meinhart, A., Silberzahn, T., and Cramer, P. (2003) *J. Biol. Chem.* 278, 15917–15921
26. Stillier, J. W., and Cook, M. S. (2004) *Eukaryotic Cell* 3, 735–740