Specific Interaction of the Transcription Elongation Regulator TCERG1 with RNA Polymerase II Requires Simultaneous Phosphorylation at Ser2, Ser5, and Ser7 within the Carboxyl-terminal Domain Repeat

Received for publication, February 7, 2013, and in revised form, February 15, 2013 Published, JBC Papers in Press, February 22, 2013, DOI 10.1074/jbc.M113.460238

Jiangxin Liu1, Shilong Fan2, Chul-Jin Lee3, Arno L. Greenleaf4, and Pei Zhou5
From the 4Department of Biochemistry, Duke University Medical Center, Durham, North Carolina 27710 and 5Laboratory of Molecular Biology, Medical Research Council (MRC) Centre, Cambridge CB2 0QH, United Kingdom

Background: TCERG1 interacts with hyperphosphorylated RNAII CTD through FF domains.

Results: We determined the structure of TCERG1 FF4–6 domain and its specific binding requirement of the CTD phosphoepitope.

Conclusion: FF4–6 forms a rigid structure of tandem FF repeats and requires simultaneous Ser2, Ser5, and Ser7 phosphorylation of the CTD for high affinity binding.

Significance: This study provides molecular insights into Ser7P-mediated co-transcriptional splicing events.

The human transcription elongation regulator TCERG1 physically couples transcription elongation and splicing events by interacting with splicing factors through its N-terminal WW domains and the hyperphosphorylated C-terminal domain (CTD) of RNA polymerase II through its C-terminal FF domains. Here, we report biochemical and structural characterization of the C-terminal three FF domains (FF4–6) of TCERG1, revealing a rigid integral domain structure of the tandem FF repeat that interacts with the hyperphosphorylated CTD (PCTD). Although FF4 and FF5 adopt a classical FF domain fold containing three orthogonally packed a helices and a 3_10 helix, FF6 contains an additional insertion helix between a1 and a2. The formation of the integral tandem FF4–6 repeat is achieved by merging the last helix of the preceding FF domain and the first helix of the following FF domain and by direct interactions between neighboring FF domains. Using peptide column binding assays and NMR titrations, we show that binding of the FF4–6 tandem repeat to the PCTD requires simultaneous phosphorylation at Ser2, Ser5, and Ser7 positions within two consecutive Y1S2P3T4S5P6S7 heptad repeats. Such a sequence-specific PCTD recognition is achieved through FF6-docking sites on FF4 and FF5 of TCERG1 but not FF6. Our study presents the first example of a nuclear factor requiring all three phospho-Ser marks within the heptad repeat of the CTD for high affinity binding and provides a molecular interpretation for the biochemical connection between the Ser7 phosphorylation enrichment in the CTD of the transcribing RNA polymerase II over introns and co-transcriptional splicing events.

RNA polymerase II (RNAPII)2 carries an intrinsically unstructured, flexible domain at the C terminus of its largest subunit, Rpb1 (1, 2). This C-terminal domain (CTD) consists of multiple repeats of a consensus heptamer, Y1S2P3T4S5P6S7 (3, 4). During each transcription cycle, the CTD undergoes waves of phosphorylation and dephosphorylation events at the Ser positions (Ser2, Ser5, and Ser7) within the heptad repeats, producing a large number of phosphorylation states (5, 6). Ser5 of the CTD is strongly phosphorylated upon formation of the pre-initiation complex followed by an increase of Ser2 phosphorylation during the elongation phase (7–9). Although Ser2 phosphorylation was first implicated in snRNA processing (10), recent studies have revealed high levels of Ser7 phosphorylation in the CTD throughout protein-coding genes, hinting at a broader function of Ser7 phosphorylation beyond snRNA processing (11–13). These serine phosphorylation states, together with phosphorylation of Tyr1 (14) and Thr4 (15) and glycosylation (16), and the distinct configurations of the Pro (Pro3 and Pro4) residues (17–19) form a “CTD code” that is recognized by a myriad of RNA processing factors and other nuclear proteins participating in co-transcriptional events (20, 21).

The human transcription elongation regulator TCERG1 (CA150) is one of the first few identified nuclear proteins that specifically bind to the hyperphosphorylated CTD (PCTD). TCERG1 is involved in trans-activator protein (Tat)-mediated transcriptional regulation of human immunodeficiency virus type 1 gene expression (22). Early lines of experimentation also implicated TCERG1 in transcription elongation via association with elongation factors, such as Tat-SF1 and positive transcription elongation factor b (23, 24). TCERG1 has been detected in highly purified native spliceosomes, suggesting that it partici-

---

* This work was supported, in whole or in part, by National Institutes of Health Grants GM079376 (to P. Z.) and GM040505 (to A. L. G.). The atomic coordinates and structure factors (code 4FQG) have been deposited in the Protein Data Bank (http://wwpdb.org/).

1 To whom correspondence should be addressed. Tel.: 919-668-6409; Fax: 919-684-8885; E-mail: peizhou@biochem.duke.edu.

2 The abbreviations used are: RNAPII, RNA polymerase II; CTD, C-terminal domain; Ser2P, Ser2 phosphorylation; Ser5P, Ser5 phosphorylation; Ser7P, Ser7 phosphorylation; PCTD, hyperphosphorylated CTD; Tat, trans-activator protein; HSQC, heteronuclear single quantum correlation; RDC, residual dipolar coupling; HYPAR, huntingtin yeast partner A; SRI, Set2-Rpb1 interacting.
pates in mRNA splicing (25–27). Consistent with this notion, in vivo splicing assays have revealed a critical role for TCERG1 in activating pre-mRNA splicing (28), and RNAi-mediated knockdown of TCERG1 has identified transcripts whose splicing decisions are dependent on TCERG1 in microarray analysis (29). Taken together, these observations establish TCERG1 as an important adaptor protein that physically couples active transcription with splicing.

TCERG1 contains three WW domains in the N-terminal half and six FF domains in the C-terminal half. The N-terminal WW domains of TCERG1 are required for binding spliceosome components, such as pre-mRNA splicing factors SF1 (26) and U2AF65 (23, 26, 28), whereas the C-terminal FF domains in TCERG1 are essential for its localization in splicing factor-rich nuclear speckles and its interaction with the PCTD (23, 30). The FF domain is a compact protein-protein interaction module of 50–60 residues that is characterized by two highly conserved phenylalanine residues at the N and C termini (31). FF domains are primarily found in two protein families, p190 family of Rho GTPase-activating proteins and N-terminal WW domains are primarily found in two protein families, p190 family of Rho GTPase-activating proteins and N-terminal WW domains-containing proteins, such as yeast pre-mRNA processing factor (Prp40) and human TCERG1 (31). Although proteins containing isolated FF domains have been identified (32), most FF domains are found as a tandem array of two to six FF repeats connected by linkers of variable lengths (31), suggesting that the biological functions of these proteins may require the cooperative interaction of multiple FF domains.

The structures of isolated FF domains have been determined, revealing a highly conserved fold of three orthogonal helices and a short $3_{10}$ helix (33–36). Despite the structural similarity of the FF domains, their ligand-binding surfaces share little similarity (37). For example, the binding of the splicing factor Prp40 FF1 domain to the crooked necklike factor 1 has been mapped to a surface encompassing $\alpha2$, the following loop, the $3_{10}$ helix, and the N-terminal half of $\alpha3$ (36). In contrast, NMR titration of the FF1 domain of formin-binding protein (FBP11/HYPA) with a Ser$^2$/Ser$^3$ doubly phosphorylated CTD peptide has implicated FBP11 residues at the N-terminal parts of $\alpha1$ and $\alpha3$ in the PCTD interaction (33).

In addition to the FF domain in FBP11/HYPA, FF domains in TCERG1 have also been implicated in the PCTD recognition (30), although such an interaction has not been characterized in detail. Among the six identified FF domains in TCERG1, the first three FF domains do not show appreciable binding to the PCTD (35, 38), consistent with the early report that C-terminal FF domains are the major contributors to PCTD binding (30). Given the well established role of TCERG1 in coupling transcription elongation and splicing, to gain insight into the interactions between C-terminal FF domains and the PCTD, we determined the crystal structure of FF4–6 and probed its binding specificity with PCTD using NMR spectroscopy and peptide column binding assays. Our combined structural and biochemical studies have revealed an integral tandem FF4–6 repeat and a previously unobserved CTD phosphoepitope required for high affinity interaction.

**Experimental Procedures**

**Molecular Cloning**—The DNA fragments encoding FF2 (residues 728–784), FF5–6 (residues 952–1081), and FF4–6 (residues 895–1081) of human TCERG1 were PCR-amplified and cloned into a pET15b vector between the Ndel and BamHI sites (EMD Biosciences). A series of TCERG1 FF4–6 point mutants (W918F, R922A, R922E, R923A, R926A, W931F, K942A, L953M, W977F, K981A, K982A, K985A, K1000A, and K1000E) were generated according to the QuikChange mutagenesis protocol (Stratagene) using the FF4–6-containing pET15b plasmid as the template. The presence of the correct mutation was verified by DNA sequencing.

**Protein Expression and Purification**—The FF2, FF5–6, and FF4–6 tandem repeat with an N-terminal His$_{10}$ tag were over-expressed in *Escherichia coli* BL21(DE3)STAR cells. Cultures were grown in LB medium in the presence of 100 µg/ml ampicillin at 37 °C until $A_{600}$ reached 0.6. Cells were induced with 0.25 mM isopropyl $\beta$-d-thiogalactopyranoside at 20 °C for 20 h. After harvest by centrifugation, cells were resuspended in the lysis buffer containing 50 mM NaH$_2$PO$_4$ (pH 8), 300 mM NaCl, and 0.1% β-mercaptoethanol (v/v) and lysed by passing through a French pressure cell at 20,000 p.s.i. Cellular debris was pelleted by centrifugation at 66,800 × g for 1 h, and the supernatant was loaded onto a Ni$_2$-nitrilotriacetic acid column. The column was extensively washed with the lysis buffer, and then the protein was eluted with a buffer containing 50 mM NaH$_2$PO$_4$ (pH 8), 300 mM NaCl, and 0.1% β-mercaptoethanol (v/v) and lyzed with passing through a French pressure cell at 20,000 p.s.i. Cellular debris was pelleted by centrifugation at 66,800 × g for 1 h, and the supernatant was loaded onto a Ni$_2$-nitrilotriacetic acid column. The column was extensively washed with the lysis buffer, and then the protein was eluted with a buffer containing 50 mM NaH$_2$PO$_4$ (pH 8), 300 mM NaCl, and 0.1% β-mercaptoethanol (v/v) and directly with tobacco etch virus protease at room temperature overnight. The digested sample was exchanged into the lysis buffer and passed through a second Ni$_2$-nitrilotriacetic acid column to remove the cleaved His$_{10}$ tag. The final purification was achieved using size exclusion chromatography (Superdex 75, GE Healthcare). Fractions containing purified protein were pooled and exchanged into crystallization buffer containing 25 mM HEPES (pH 7), 100 mM KCl, and 0.1% β-mercaptoethanol (v/v) and digested with tobacco etch virus protease at room temperature overnight. The digested sample was exchanged into the lysis buffer and passed through a second Ni$_2$-nitrilotriacetic acid column to remove the cleaved His$_{10}$ tag. The final purification was achieved using size exclusion chromatography (Superdex 75, GE Healthcare). Fractions containing purified protein were pooled and exchanged into crystallization buffer containing 25 mM HEPES (pH 7), 100 mM KCl, and 0.1% β-mercaptoethanol. The purified protein contained an N-terminal overhang of three additional residues (SHM) as a result of tobacco etch virus cleavage and primer design.

**Yeast Whole-cell Extracts**—Yeast strains with 14 repeats of consensus sequence (YSPTPS)$_1$ or all-S7A mutant CTD (YSPTPS)$_1$ fused to residue Gly$^{1541}$ in *Saccharomyces cerevisiae* Rpb1 were generously provided by Prof. Beate Schwer (39). The growth of these two strains is identical to that of strains with a full-length *S. cerevisiae* CTD, and they are referred to as WT CTD$_{14}$ and S7A CTD$_{14}$, respectively, in this study. WT CTD$_{14}$ and S7A CTD$_{14}$ strains were grown in dropout medium minus histidine at 30 °C until $A_{600}$ reached 1.0. The cells were harvested at 4 °C. The cell pellet was washed twice with the PBS buffer and then transferred to a syringe with a spatula and scoop. The cells were slowly extruded into liquid nitrogen and frozen into small pieces. The frozen yeast cells were ground using a Retsch Mixer Mill MM 400 in liquid nitrogen to a fine powder and then stored at −80 °C. Aliquots of cell powder were suspended in a buffer containing 25 mM HEPES (pH 7.0), 100
Structure of TCERG1 FF4–6 and Its PCTD Binding Specificity

mm NaCl, 1 mm PMSF, and a Complete Mini protease inhibitor mixture tablet (Roche Applied Science) and centrifuged at 15,000 rpm for 30 min to remove the debris.

Whole-cell Extract Pulldown Assay and Western Blot—Purified His-tagged FF4–6 was applied repeatedly onto a column containing 200 μl of TALON cobalt resin (Clontech), and unbound FF4–6 was removed by extensive wash. About 1 mg of WT or S7A CTD cell extract was loaded onto the column followed by extensive wash with a buffer containing 25 mM HEPES (pH 7.0), 0.15 mM NaCl, and 12 mM imidazole to eliminate nonspecific binding. A high salt buffer containing 25 mM HEPES (pH 7) and 1 mM NaCl was then used to disrupt the interaction between the PCTD and FF4–6 and elute the PCTD. WT CTD cell extract loaded onto a blank cobalt resin column was loaded onto the peptide column. The flow-through was extensively washed with PBS buffer containing 0.2% Tween 20, 5 mM NaF, 0.1% PMSF, and 5 mM DTT. GST, GST-tagged yeast CTD containing 26 heptad repeats (GST-CTD26), and GST-tagged CTD containing three heptad repeats (GST-CTD3) were hyperphosphorylated with yeast CTD kinase I for 6 h in vitro (40). The nitrocellulose membrane was extensively washed with PBS buffer containing 0.2% Tween 20 and then probed with hyperphosphorylated GST-CTD fusion protein for 2 h at 4 °C. The probe was washed four times and detected with rabbit anti-GST antibody (Sigma) and then with the IRDye 680 donkey anti-rabbit IgG (heavy + light) (LI-COR Biosciences) antibody. The blots were scanned with an Odyssey scanner (LI-COR Biosciences).

Immobilized CTD Peptide Binding Assay—The PCTD peptides (Table 1) were dissolved in a buffer containing 25 mM HEPES (pH 7) and 100 mM KCl and loaded repetitively onto a blank column containing 200 μl of TetraLink tetrameric avidin resins (Invitrogen) to generate the PCTD peptide column. Unbound PCTD peptides were removed by an extensive buffer wash (15 ml). 50–100 μg of TCERG1 in a buffer containing 25 mM HEPES (pH 7), 100 mM KCl, and 0.1% β-mercaptoethanol was loaded onto the peptide column. The flow-through was collected, and the column was washed twice with 200 μl of the loading buffer followed by 15-ml buffer wash. Proteins bound to the PCTD peptide column were eluted in three fractions of 200 μl each with an elution buffer containing 25 mM HEPES (pH 7.0), 8% glycerol, and 0.3 M NaCl. All of the fractions were analyzed by SDS-PAGE.

X-ray Crystallography—Crystallization was performed using the hanging drop vapor diffusion method at 4 °C. The crystallization buffer contained 0.016 M NiCl2, 0.1 M Tris-HCl (pH 9), 16% polyethylene glycol monomethyl ether 2000, and 0.13 M glycine. Selenomethionine-labeled FF4–6 containing an L953M mutation was expressed by incorporation of selenomethionine into the SelenoMet Medium Base containing SelenoMet Nutrient Mix (Molecular Dimensions Ltd., UK) and purified as described above. The extent of selenomethionine incorporation was determined by mass spectrometry. Harvested native and selenomethionine-labeled protein crystals were cryoprotected with a reservoir solution containing 30% (v/v) of ethylene glycol and with perfluoropolyether (PF-OX175/08) oil, respectively, and flash frozen with liquid nitrogen.

Diffraction data of native and selenomethionine-labeled crystals were collected at the Southeast Regional Collaborative Access Team (SER-CAT) 22-BM beamline at the Advanced Photon Source, Argonne National Laboratory. Diffraction data were processed with HKL2000 (41). The experimental phases were determined by the multiwavelength anomalous dispersion method using data sets collected on a selenomethionine-labeled crystal of the FF4–6 L953M mutant. Programs SOLVE and RESOLVE were used to locate the selenomethionine sites, calculate the initial multiwavelength anomalous dispersion phases, and modify the density map (42–44). Initial automated model building gave excellent electron density for regions containing FF4 and FF5 (residues 895–1010) but poor density for the region containing FF6 (residues 1011–1081). Successive rounds of model building using Coot (45) and refinement using PHENIX (42) were used to build the complete model, which was validated with MolProbity (46). Data collection, phasing, and refinement statistics are summarized in Table 2.

NMR Spectroscopy—Isotopically enriched proteins were overexpressed in M9 minimal medium using [15N]NH4Cl and [13C]glucose as the sole nitrogen and carbon sources (Cambridge Isotope Laboratories). Protein deuteriation was achieved by growing cells in 100% D2O M9 minimal medium. NMR spectra were acquired with Agilent INOVA 600- or 800-MHz spectrometers at 25 °C. Backbone resonance assignments were obtained using standard triple resonance experiments (47). Spectra were processed by NMRPipe (48) and analyzed using Sparky (49). T1 and T2 experiments were conducted to measure the rotational correlation time, τc. A series of 1H,15N HSQC’s on the FF4–6 sample was collected. The delays used for data collection were 10, 200, 400, 600, 800, 1000, and 1400 ms for longitudinal relaxation (T1) and 10, 20, 30, 40, 50, 60, and 70 ms for transverse relaxation (T2). T1 and T2 values were determined by fitting peak intensities to an exponential delay function using the rate analysis tool in NMRView (50), and rotational correlation time was calculated as described previously (51).

An HNCO-based experiment for measurement of residual dipolar coupling (RDC) was performed on a sample of 0.8 mM deuterated and uniformly 13C/15N-labeled FF4–6 (52). 2D 1H-13C NMR data were obtained by taking the difference in 1H NH couplings in aligned 9 mg/ml Pf1 phage medium (ASLA Biotech Ltd.) and isotropic (water) medium. Errors of RDC measurements were estimated on the basis of duplicate experiments. RDC values from residues within secondary structural regions were analyzed by the MODULE program (53) using the crystal structure of FF4–6 as the input coordinate.
NMR Titration—NMR samples contained 0.2 mM 15N-labeled TCERG1 FF4–6 domain in a buffer containing 25 mM HEPES (pH 7), 100 mM KCl, and 10 mM DTT. Synthetic three-heptad-repeat 2,5,2,5,2,5-Ser(P) and 2,5,7,2,5,7-Ser(P) CTD peptides were dissolved in the NMR buffer and titrated into the 15N-labeled FF4–6 sample. HSQC spectra were analyzed, and the chemical shift perturbation was calculated by
\[
\delta = \frac{\delta_{\text{max}}(L+K_g+P) - \sqrt{(L+K_g+P)^2 - 4PL}}{2E_T}
\]
where \(\delta_{\text{max}}\) refers to the maximum chemical shift change between bound and free states, \(P\) refers to the protein concentration, and \(L\) refers to the ligand concentration, respectively.

RESULTS

TCERG1 FF4–6 Tandem Repeat Forms a Rigid Integral Domain Structure—Previous biochemical studies on TCERG1-PCTD interaction have revealed a high affinity interaction between the hyperphosphorylated CTD and FF1–6 and have mapped such an interaction to the C-terminal FF domains (30). It is important to note that although the structures of individual FF domains of TCERG1 have been determined by solution NMR (34, 35) (also see Protein Data Bank codes 2DOD, 2DOE, 2DOF, and 2E71 deposited by the RIKEN Structural Genomics/Proteomics Initiative), a recent crystallography study has shown that the three N-terminal FF domains (FF1–3) fold into a rigid integral structure with neighboring FF domains connected by a long helix (38). A close examination of the FF domain sequences of TCERG1 reveals a long disordered linker between FF3 and FF4, whereas FF4, FF5, and FF6 are only separated by a single residue (Ala\(^{552}\) between FF4 and FF5 and Asp\(^{1010}\) between FF5 and FF6). Thus, there is a strong likelihood for FF4, FF5, and FF6 to form an integral tandem domain structure. To examine this possibility, we expressed and purified 15N-labeled FF2, FF5–6, and FF4–6 domains of TCERG1. All three proteins display high quality 1H–15N HSQC spectra (data not shown). Measurements of rotational correlation times using \(T_1\) and \(T_2\) experiments for FF2, FF5–6, and FF4–6 revealed distinct values of 4.9, 10.5, and 14.1 ns for FF2, FF5–6, and FF4–6, respectively. These highly different values of the NMR-determined rotational correlation times suggest that FF4, FF5, and FF6 do not behave as isolated FF domains and that they fold into a rigid integral domain. Thus, the FF4–6 tandem repeat domain, but not shorter peptides, is the minimal functional unit for interacting with the hyperphosphorylated CTD in solution.

FF4–6 Tandem Repeat Domain Binds Hyperphosphorylated CTD Containing Three Heptad Repeats—Because the previously mapped minimal PCTD-binding module of FF5 (30) is inconsistent with the notion that FF4–6 represents the minimal functional unit of C-terminal FF domains of TCERG1 in solution, we evaluated whether FF4–6 can similarly bind to the PCTD using far-Western blotting assays. Briefly, purified FF4–6 and FF1–6 (as a positive control) were used as input for SDS-PAGE, transferred to a nitrocellulose membrane, and probed with CTD kinase I-treated, hyperphosphorylated GST-yeast CTD containing 26 heptad repeats (GST-PCTD\(_{26}\); Fig. 1). The retention of GST-yeast CTD by the TCERG1 FF repeats after extensive buffer wash was probed by primary and secondary antibodies. In our assays, both FF4–6 and FF1–6 bound to the hyperphosphorylated GST-yeast CTD strongly, whereas none of the FF repeats bound to GST alone, suggesting a specific interaction between FF4–6 and the hyperphosphorylated CTD. We next probed the minimal functional unit of the CTD required for recognition. Constructs with different numbers of CTD heptad repeats were made and investigated. Our far-Western analysis shows that hyperphosphorylated CTD with as few as three heptad repeats (GST-PCTD\(_3\)) was sufficient for high affinity interaction with both FF4–6 and FF1–6 (Fig. 1B).

Specific CTD Recognition by the FF4–6 Tandem Repeat Requires Simultaneous Phosphorylation at Ser\(^2\), Ser\(^3\), and Ser\(^7\)—After elucidating that a three-repeat hyperphosphorylated CTD peptide is sufficient for high affinity interaction with the FF4–6 tandem repeat, we investigated its specific CTD phosphopeptide requirement using PCTD peptide column binding assays and NMR titration experiments. Because CTD kinase I has been shown to phosphorylate Ser\(^2\) within the heptad repeat of the CTD already containing Ser\(^5\) phosphorylation and generate Ser\(^2\)/Ser\(^5\) doubly phosphorylated heptads starting from unphosphorylated heptad repeats in vitro (54, 55), we anticipated that the TCERG1 FF4–6 tandem repeat binds to a three-heptad-repeat CTD hydrophosphorylated at Ser\(^2\) and Ser\(^5\) positions. To test this hypothesis, we synthesized a three-heptad-repeat CTD peptide containing Ser\(^2\)P and Ser\(^5\)P (Y\(^1\)p\(^2\)Sp\(^4\)p\(^5\)p\(^7\)Y\(^1\)p\(^2\)Sp\(^4\)p\(^5\)p\(^7\)Y\(^1\)p\(^2\)Sp\(^4\)p\(^5\)p\(^7\); referred to as 2,5,2,5,2,5-Ser(P); Table 1) and evaluated its interaction with the FF4–6. Surprisingly, no specific binding was detected between FF4–6 and the 2,5,2,5,2,5-Ser(P) peptide in the column binding assay after an extensive wash with buffer containing 100 mM NaCl (Fig. 2A, left panel). In contrast, consistent with previous observations (56), the SRI domain of the histone methyltransferase SET2 displayed specific and high affinity binding to the 2,5,2,5,2,5-Ser(P) CTD peptide under identical conditions (Fig. 2A, right panel), verifying that the 2,5,2,5,2,5-Ser(P) CTD peptide was functionally active. We also tested two other CTD peptides containing three heptad repeats and Ser phosphorylation at the

FIGURE 1. Interaction between the tandem FF4–6 repeat of TCERG1 and hyperphosphorylated CTD. A, purified TCERG1 FF1–6 and FF4–6 samples stained with Coomassie Blue on the SDS-PAGE gel. B, far-Western blots probing the interactions of TCERG1 FF1–6 and FF4–6 with GST, hyperphosphorylated GST-yeast CTD containing 26 heptad repeats (GST-PCTD\(_{26}\)), and hyperphosphorylated GST-yeast CTD containing three heptad repeats (GST-PCTD\(_3\)). Interactions were detected with rabbit anti-GST antibody and IRDye 680 donkey anti-rabbit IgG (heavy + light).
Structure of TCERG1 FF4–6 and Its PCTD Binding Specificity

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>7,7-Ser(P)</td>
<td>Biotin-GGGYTPSSpYTPYTPSSpYTPYS</td>
</tr>
<tr>
<td>2,5,2,5-Ser(P)</td>
<td>Biotin-GGGGSSpYPSSpYPSSpYPSSpY</td>
</tr>
<tr>
<td>5,7,5,7-Ser(P)</td>
<td>Ac-YpTPSpYpSPYpSPYpSPyYpYpSPy</td>
</tr>
<tr>
<td>7,2,7,2-Ser(P)</td>
<td>Biotin-GGGYTPSSpYTPYTPSSpYTPYS</td>
</tr>
<tr>
<td>2,5,2,5,2,5-Ser(P)</td>
<td>Ac-YpTPSpYpSPYpSPYpSPyYpYPYpYpSPy</td>
</tr>
<tr>
<td>2,5,7,2,5,7-Ser(P)</td>
<td>Ac-SCPYySPSpYPsPYpSPYpY</td>
</tr>
</tbody>
</table>

5,2,5,2 positions (Y₁S²P³T⁴S⁵P⁶Y⁷pS⁷pT⁸pS⁹pT⁴S¹pS²pT³S⁴; referred to as 5,2,5,2-Ser(P)) or 2,5,2,5 positions (S⁵pS⁶pT⁷pS⁸pY⁹pS¹pT²pS³pY⁴pS²pT³; referred to as 2,5,2,5-Ser(P)). Again, no specific interaction was observed after extensively washing the column with buffer containing 100 mM salt (data not shown). Therefore, the Ser²/Ser⁵ double phosphorylation of CTD repeating units is insufficient for high affinity interaction with the TCERG1 FF4–6 tandem repeat.

Because our far-Western blotting assay showed that FF4–6 binds to the hyperphosphorylated CTD containing only three heptad repeats, we reasoned that FF4–6 might recognize a previously unobserved CTD phosphoepitope. In particular, given the recent discovery of prevalent Ser⁷ phosphorylation in CTD during transcription (11–13) and the observation that bacterially overexpressed GST-CTD contains a low level of Ser⁷ phosphorylation (57), we wondered whether the CTD recognition by TCERG1 FF4–6 tandem repeat requires Ser⁷ phosphorylation. To test this idea, we evaluated the binding of the FF4–6 tandem repeat to PCTD peptides containing Ser phosphorylation at 2,5,2,5,2,5; 5,7,5,7; or 7,2,7,2 positions using peptide column binding assays. The 2,5,7,2,5,7-Ser(P) peptide exhibited strong binding to FF4–6 (Fig. 2B, left panel), requiring 0.3 M NaCl to elute the PCTD-bound FF4–6. In contrast, neither the 5,7,5,7-Ser(P) nor the 7,2,7,2-Ser(P) PCTD peptides showed specific interactions with FF4–6, and FF4–6 can be washed off in the presence of 0.1 M NaCl. Taken together, these results suggest that the specific CTD recognition by the FF4–6 tandem repeat requires all six serine residues within the two heptad repeats to be phosphorylated (Fig. 2B). It is important to note that such a high affinity interaction is not due to nonspecific charge-charge interactions as FF4–6 did not bind the 2,5,2,5,2,5-Ser(P) CTD peptide containing the same number of phosphate groups in the peptide column binding assay (Fig. 2A, left panel), highlighting the high degree of specificity of this interaction.

To obtain a more quantitative measurement of the interaction between TCERG1 FF4–6 and the 2,5,7,2,5,7-Ser(P) CTD peptide, we determined their binding affinity using NMR titration. A series of ¹H HSQC spectra of TCERG1 FF4–6 tandem repeat was collected in the presence of increasing molar ratios of the 2,5,7,2,5,7-Ser(P) CTD peptide (from 0:1 to 3:1). A number of resonances were notably perturbed during titration. Resonances with obvious chemical shift perturbations and no signal overlaps were selected for extraction of the binding affinity. Fitting of the titration curve yielded a dissociation constant ($K_d$) of $13 \pm 5 \mu M$ to the 2,5,7,2,5,7-Ser(P) peptide (Fig. 3A). In comparison, FF4–6 binds with a much weaker binding affinity to the 2,5,2,5,2,5-Ser(P) CTD peptide with an extracted $K_d$ value of $102 \pm 33 \mu M$ (Fig. 3B). Compared with the 2,5,2,5,2,5-Ser(P) peptide, the much higher affinity of FF4–6 for the 2,5,7,2,5,7-Ser(P) peptide strongly argues for specific binding to the triply phosphorylated repeats.

In Vivo Phosphorylation on Ser⁷ of PCTD Is Required for TCERG1 FF4–6 Interaction—We next evaluated whether TCERG1 FF4–6 is able to bind in vivo modified RNAPII CTD in the absence of Ser⁷ phosphorylation. Pulldown assays were carried out using lysates from yeast cells expressing 14 repeats of consensus CTD sequence (YSPTSPS)₁₄ (WT CTD₁₄) or S7A mutant sequence (YSPTSPA)₁₄ (S7A CTD₁₄) (39). Western blotting of cell lysates revealed Ser2P, Ser5P, and Ser7P marks of the CTD in the WT CTD₁₄ cell lysates, whereas only Ser2P and Ser5P marks can be detected in the S7A CTD₁₄ cell lysates (data not shown), confirming that all of the Ser² residues in the CTD have been replaced with Ala. Importantly, when His₁₀-tagged TCERG1 FF4–6 was immobilized on a cobalt column, it was able to selectively pull down the hyperphosphorylated CTD from the WT CTD₁₄ whole-cell lysate but not from the S7A CTD₁₄ whole-cell lysate (Fig. 4), suggesting that Ser⁷ phosphorylation is required for the specific interaction of TCERG1 FF4–6 with hyperphosphorylated CTD.

Structure of the Tandem FF4–6 Repeat—Having characterized the specific CTD phosphoepitope requirement of FF4–6, we went on to determine its structure, which is composed of residues 895–1081 of TCERG1, to further characterize the molecular basis of the FF4–6 tandem repeat-PCTD interaction. The structure of FF4–6 was solved by x-ray crystallography and refined to 2.0 Å. Two molecules, protomer A and protomer B, are observed in one asymmetric unit. Except for the C-terminal four residues of protomer A, clear electron density can be observed for all the residues, including the entirety of protomer B. Structural superimposition shows excellent agreement between the two protomers with an all-atom root mean square deviation of 0.6 Å, indicating a high degree of structural consistency. Because of its completeness in electron density, protomer B was selected as the representative monomer structure of FF4–6 in the following discussion. Final statistics are reported in Table 2.

The overall architecture of FF4–6 adopts an inverted V shape with FF5 centering at the vertex and FF4 and FF6 occupying the ends (Fig. 5A). Among the three FF domains, FF4 and FF5 adopt a canonical FF domain fold (33). Each domain consists of three α helices arranged in an orthogonal bundle with a 3₁₀ helix in the loop connecting the α₂ and α₃ helices. The N-terminal helix (α₁) and C-terminal helix (α₃) are pointing in opposite directions with a cross-angle of $\approx 120^\circ$. FF6 deviates from such a canonical FF domain fold. It contains an 18-residue linker between helices α₁ and α₂ in comparison with a linker of six to nine residues in other FF domains (Fig. 5B). These extra residues in the linker region adopt a helical conformation (α₁') in FF6, whereas the shorter linker in the canonical FF domain fold adopts a loop conformation. The formation of such an insertion α₁' helix in FF6 does not disrupt the relative orienta-
tion of other helices as all canonical helices, including $\alpha_1$, $\alpha_2$, $3_{10}$ and $\alpha_3$, superimpose very well in FF4, FF5, and FF6 (Fig. 5B).

The integral domain structure of the FF4–6 tandem repeat is forged by merging the C-terminal helix of the preceding FF domain with the N-terminal helix of the following FF domain into a single, continuous $\alpha$ helix that sequentially connects FF4 and FF5 and FF5 and FF6, respectively. The connecting helices do not show elevated B-factors compared with individual FF domain residues (data not shown), consistent with the notion that FF4–6 forms a rigid domain structure. The rigidity of the tandem FF4–6 repeat is buttressed by hydrogen bonds and van der Waals interactions between neighboring FF domains. In particular, side chains of Ser914 and Asp915 in the $\alpha_1$-$\alpha_2$ loop of FF4 form three hydrogen bonds with the backbone of Phe993 and Ser995 and the side chain of Ser994 from the loop connecting the $3_{10}$ helix and the $\alpha_3$ helix in FF5 (Fig. 5C). This hydrogen bond network is strengthened by the formation of an additional interdomain hydrogen bond between the side chain of Lys957 of FF5 and the backbone carbonyl of Lys992 within FF5. In contrast, hydrophobic interactions dominate the FF5-FF6 interface. Residues Thr972 and Thr974 located in the $\alpha_1$-$\alpha_2$ loop of FF5 and residues Leu1060, 

FIGURE 2. Interaction between the TCERG1 FF4–6 domain and PCTD requires simultaneous phosphorylation of Ser2, Ser5, and Ser7 within the heptad repeat of the CTD. A, column binding assays of FF4–6 (left) and human SRI (right; positive control) to the 2,5,2,5,2,5-Ser(P) CTD peptide. B, column binding assays of FF4–6 to PCTD peptides containing different phosphoepitopes. PCTD peptides harboring different phosphoepitopes were immobilized onto each streptavidin column. The same amount of FF4–6 protein was loaded onto the column as input. The flow-through (FT) fraction was collected, and the column was washed twice with 200 $\mu$l of buffer containing 25 mM HEPES (pH 7.0) and 0.1 M NaCl followed by 15-mI buffer wash (Wash, fractions 1–3). Protein bound to the PCTD peptide column was eluted with three fractions of 200 $\mu$l buffers containing 25 mM HEPES (pH 7), 0.3 M NaCl (Elution, fractions 1–3). All fractions were analyzed by SDS-PAGE and stained with Coomassie Blue. M, molecular mass markers.

FIGURE 3. Binding affinities of TCERG1 FF4–6 to CTD peptides containing 2,5,7,2,5,7-Ser(P) (A) or 2,5,2,5,2,5-Ser(P) (B) phosphoepitopes determined by NMR titration experiments.

FIGURE 4. TCERG1 FF4–6 selectively pulls down hyperphosphorylated WT RNAPII CTD from yeast lysate but not S7A mutant RNAPII CTD. Purified His$_{10}$-tagged TCERG1 FF4–6 was immobilized on a cobalt column, and empty cobalt resin was used as a negative control. Whole-cell lysates from yeast strains containing WT RNAPII CTD$_{14}$ or S7A RNAPII CTD$_{14}$ were loaded onto the column followed by an extensive buffer wash and high salt elution. The whole-cell lysate (input) and the elution fraction were loaded on an SDS-PAGE gel, which was Western blotted using the Ser5P-specific CTD antibody 3E8.
**Structure of TCERG1 FF4–6 and Its PCTD Binding Specificity**

### TABLE 2

Crystallographic data, phasing, and refinement statistics

<table>
<thead>
<tr>
<th>Data collection and phasing</th>
<th>Native FF4–6</th>
<th>MAD data collection, Se-Met FF4–6 L953M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (Å)</td>
<td>1.0001</td>
<td>0.9719</td>
</tr>
<tr>
<td>Space group</td>
<td>P 1 21 1</td>
<td>0.9794</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
<td>0.9796</td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>27.9, 77.1, 95.2</td>
<td>83,492</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
<td>107,683</td>
</tr>
<tr>
<td>Total reflections</td>
<td>10896</td>
<td>26,788</td>
</tr>
<tr>
<td>Completeness</td>
<td>98.3 (99.4)</td>
<td>98.1 (90.2)</td>
</tr>
<tr>
<td>Rmerge (%)</td>
<td>33.2 (11.7)</td>
<td>98.9 (97.2)</td>
</tr>
<tr>
<td>I/σ (all)</td>
<td>0.05 ppm</td>
<td>7.3 (32)</td>
</tr>
<tr>
<td>Ramachandran plot</td>
<td></td>
<td>8.8 (35.8)</td>
</tr>
<tr>
<td>Refinement</td>
<td></td>
<td>8.3 (32)</td>
</tr>
<tr>
<td>Bond length (Å)</td>
<td>0.006</td>
<td>13.8 (2.8)</td>
</tr>
<tr>
<td>Bond angle (°)</td>
<td>0.917</td>
<td>15.6 (3.0)</td>
</tr>
<tr>
<td>Ramachandran plot</td>
<td></td>
<td>11.7 (2.8)</td>
</tr>
<tr>
<td>r.m.s. deviations</td>
<td>0.189/0.240</td>
<td>0.32</td>
</tr>
<tr>
<td>B-factor (Å²)</td>
<td>23.45</td>
<td></td>
</tr>
<tr>
<td>Rwork/Rfree</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>22.74</td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>27.64</td>
<td></td>
</tr>
<tr>
<td>MoltProbity score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All-atomic clashscore</td>
<td>8.66</td>
<td></td>
</tr>
<tr>
<td>Clashscore percentile</td>
<td>87th</td>
<td></td>
</tr>
</tbody>
</table>

Note: Values in parentheses are for the highest resolution shells.

Refinement statistics were generated using MolProbity (46).

---

Cys¹⁰⁶², and Val¹⁰⁶³ from the loop connecting the 3₁₀ helix and the α₃ helix of FF₆ form extensive interdomain van der Waals contacts (Fig. 5D). This hydrophobic interface is additionally supported by interdomain contacts between Tyr¹⁰¹² of FF₆ and Leu⁹⁷₃ of FF₅ and intradomain interactions involving Tyr¹⁰¹² and Leu¹⁰⁶⁰ of FF₆.

To evaluate whether the rigid tandem domain structure of FF₄–₆ is also preserved in solution, we assigned the backbone resonances of FF₄–₆ using transverse relaxation optimized spectroscopy-based triple resonance experiments and a²H/¹²C/¹⁵N-labeled protein sample (47). TALOS+ analysis of the backbone resonances predicted a nearly uniform distribution of order parameters derived from the random coil index (RCI-²5) (58, 59), including the two linker helices connecting FF₄–FF₅ and connecting FF₅–FF₆ (Fig. 5E), suggesting that TCERG₁ FF₄–₆ also adopts a rigid structure in solution. Furthermore, the experimentally measured ¹³DHiN residual dipolar couplings showed good correlation with calculated values from the crystal structure (Fig. 6) with an RDC quality factor (Q-factor; Ref. 60) of 0.29, suggesting that the solution state conformation of the FF₄–₆ is consistent with that observed in the crystal structure.

### Tandem FF₄–₆ Repeat Domain Binds PCTD through FF₄ and FF₅—To determine the PCTD-binding surface of the TCERG₁ FF₄–₆ tandem repeat, we analyzed the chemical shift perturbation of the backbone resonances and Trp side chain resonances based on known assignments. Our analysis showed that titration of the 2,5,7,2,5,7-Ser(P) CTD peptide resulted in noticeable chemical shift perturbations (δₑ > 0.05 ppm) for the resonances of the following residues: Ser⁹¹⁹, Arg⁹²₃, Arg⁹₂₆, Trp⁹₃₁, Gly⁹₃₄, Thr⁹₇₂, Thr⁹₇₆, Lys⁹₈₁, Lys⁹₈₂, Lys⁹₈₅, and Gly⁹₈₆ and side chains of Trp⁹₁₈, Trp⁹₃₁, and Trp⁹₇₇ (Fig. 7A). The side chain HeNe resonance of the Trp⁹₃₁ in particular undergoes a large scale chemical shift perturbation (δₑ > 0.5 ppm; Fig. 3A), indicating that the Trp⁹₃₁ side chain is likely involved directly in PCTD interaction. These perturbed residues are located within FF₄ and FF₅ but not in FF₆ of TCERG₁ (Fig. 7A), arguing that FF₄ and FF₅ are the main PCTD-interacting modules. Furthermore, these residues cluster in the middle of α₂ helices of FF₄ (Ser⁹¹⁹, Arg⁹₂₃, and Arg⁹₂₆) and FF₅ (Thr⁹₇₂, Thr⁹₇₆, Lys⁹₈₁, and Lys⁹₈₂) and the following 3₁₀ helices of FF₄ (Thr⁹₇₁ and Gly⁹₃₄), and they define two neighboring CTD-docking sites enriched with basic residues that are ideally suited for interacting with hyperphosphorylated CTD peptides (Fig. 7B).

To verify that the CTD-docking sites defined by the NMR titration experiment are the bona fide binding interface of TCERG₁ to the 2,5,7,2,5,7-Ser(P) CTD peptide, we selectively mutated positively charged Arg and Lys residues within these two sites that are most likely to be directly involved in the binding of the phospho-Ser of the CTD and evaluated their effects on the FF₄–₆-PCTD interaction. ¹⁵N HSQC spectra of mutated proteins were collected to verify the structural integrity of the FF₄–₆ point mutations (data not shown). Well folded mutant proteins were probed for their ability to interact with the 2,5,7,2,5,7-Ser(P) peptide using peptide column binding assays. Under the same washing condition used for the FF₄–₆ tandem repeat domain of the wild-type TCERG₁ protein (Fig. 2B, left panel), point mutations R₉₂₂E, R₉₂₃A, and
R926A of FF4 and K942A of FF5 within FF4 completely eliminated the TCERG1 interaction with the 2,5,7,2,5,7-Ser(P) CTD peptide, and the R922A mutation weakened the interaction. Similarly, point mutations K981A, K982A, and K985A of FF4 and K1000A or K1000E of FF5 in FF5 either completely eliminated or severely diminished the TCERG1 interaction with the 2,5,7,2,5,7-Ser(P) CTD peptide (Fig. 7C). Taken together, these data define the FF4, the following 310 helix, and the N terminus of FF5 as the primary CTD-docking sites of TERG1 FF4–6. Interestingly, these CTD-docking sites of the TERG1 FF4–6 tandem repeat are distinct from that of the PCTD-binding FF domain of HYPA/FBP11 that shows perturbation on residues at the N terminus of FF1 and FF3 (33), but they are similar to the binding surface of Prp40 FF1 that interacts with crooked necklike factor 1, a peptide unrelated to the PCTD.

DISSCUSSION

PCTD Binding Specificity of FF4–6—Modern structural biology is based on a reductionist approach in that the minimal functional unit of a target protein is isolated and probed at the atomic level. In the case of TERG1, the structures of individual FF domains have been studied in detail (34, 35) (also see Protein Data Bank codes 2DOD, 2DOE, 2DOF, and 2E71

FIGURE 5. Structure of the TERG1 FF4–6 tandem repeat. A, ribbon diagram of the FF4–6 tandem repeat revealing a rigid integral domain. Individual FF domains are color-coded with FF4 in cyan, FF5 in orange, and FF6 in pink. B, superimposition of FF4, FF5, and FF6. C and D depict residues involved in interdomain interactions between FF4 and FF5 and between FF5 and FF6, respectively. E, sequence-specific order parameters derived from the random coil index (RCI-S2) (58). Secondary structures are labeled on the top, and regions of individual FF domains are color-coded as in A.
deposited by the RIKEN Structural Genomics/Proteomics Initiative). In contrast to the notion of individual FF domains as functional units, our NMR study of FF4–6 has revealed an integral tandem repeat domain in solution with a rotational correlation time far exceeding that of isolated FF domains, and our crystallographic study has further revealed a rigid FF4–6 tandem repeat fold. The FF4–6 tandem repeat is topologically similar to the previously reported tandem repeat structure of FF1–3 (38), but it is much less flexible than FF1–3 as the assembly of the FF4–6 is held together not only by an undisrupted helix connecting neighboring FF domains but also by direct domain-domain interactions between FF4 and FF5 and between FF5 and FF6 (Fig. 8A). Taken together, these observations argue that the minimal function units of TCERG1 are not individual FF domains but rather tandem FF repeats of FF1–3 and FF4–6, suggesting that functional studies utilizing individual FF domains or double FF domains may need to be re-evaluated.

Except for FF6, which contains an insert helix (α1'), all of the remaining FF domains of TCERG1 adopt a canonical FF domain fold consisting of three orthogonal helices and a short 3_10 helix. Among the six FF domains of TCERG1, FF1, FF2, FF5, and FF6 are highly basic with pI values exceeding 9.0, whereas FF3 and FF4 have pI values slightly less than 7.0. Gasch et al. (36) argue that the pI values dictate whether individual FF domains are involved in PCTD binding. Contrary to this
structure of TCERG1 FF4–6 and Its PCTD Binding Specificity

propose, several groups reported that TCERG1 FF1–3 domain shows a very weak and barely detectable interaction with the PCTD (35, 38). Our result presented here further invalidates this notion as the slightly acidic FF4 and basic FF5 are involved in binding to the 2,5,7,2,5,7-Ser(P) CTD peptide rather than the highly basic FF6. Furthermore, our mutagenesis studies revealed two CTD-docking sites enriched with basic residues, including Arg922, Arg923, and Arg926 of α2 and Lys942 of α3 within FF4 and Lys981, Lys982, and Lys985 of α2 and Lys1000 of α3 within FF5, that are required for high affinity interaction between TCERG1 and PCTD. Because a significant portion of these basic residues is either not conserved or replaced with oppositely charged acidic residues in other FF domains of TCERG1 (Fig. 8B), those FF domains, despite their overall highly basic pl values, do not interact with the 2,5,7,2,5,7-hyperphosphorylated CTD. Therefore, the pl value alone is insufficient for prediction of the PCTD binding property of an FF domain.

It is important to note that although PCTD-associating protein binding to singly phosphorylated CTD at Ser2, Ser5, or Ser7 positions or doubly phosphorylated CTD at Ser2 and Ser5 positions have been reported previously (61), no other protein has been observed to require phosphorylation of all three Ser residues, including Ser2, Ser5, and Ser7, within the heptad repeat of the CTD for high affinity binding. In contrast, our peptide column assays showed that the high affinity interaction of TCERG1 FF4–6 with PCTD peptides requires the simultaneous phosphorylation at Ser2, Ser5, and Ser7 positions; additionally, our in vivo pulldown assay showed that TCERG1 FF4–6 specifically interacts with hyperphosphorylated CTD only in the presence of Ser7P but not when all of the Ser7 residues in the heptad CTD repeats are replaced with Ala. The ~8-fold affinity difference of TCERG1 FF4–6 binding to the 2,5,7,2,5,7-Ser(P) CTD peptide over the same CTD peptide with a less optimal phosphorylation pattern (2,5,2,5,2,5-Ser(P)) is comparable with the affinity variations reported for other well characterized PCTD-associating domains, such as the Nrd1 CTD-interacting domain and the SRI domain, for specific PCTD recognition (56, 62). Taken together, these results suggest that TCERG1 FF4–6 is the first example of a PCTD-associating protein specifically recognizing Ser2P, Ser5P, and Ser7P of the heptad repeat for high affinity CTD binding.

Implication of the Distinct 2,5,7-Ser(P) CTD-binding Epitope of TCERG1—CTD has been implicated in a wide range of transcription-associated functions. Different forms of CTD predominate at each stage of the transcription cycle and act as recognition sites for recruiting various mRNA processing factors, therefore coupling transcription with mRNA processing (7, 63). The most extensively studied aspect of CTD modification has been the phosphorylation of Ser2 and Ser5 within the consensus heptad repeat. For example, Ser5 phosphorylation is primarily detected at the 5'-end of the genes, and its recognition by mRNA-capping enzymes enhances the activity of capping enzymes (63, 64). In contrast, Ser2 phosphorylation is enriched at the 3'-end of the genes, recruiting transcription termination factors, such as Rtt103 and Pcf11, and coordinating the 3'-end processing (65, 66).

Besides Ser2 and Ser5 phosphorylation, Ser7P has recently been discovered in both mammalian and yeast cells (10, 57, 67). Although Ser7P has initially been implicated in snRNA gene expression (10), recent high resolution genome-wide occupancy profiling has revealed widespread marks of Ser7P in the RNAPII CTD for coding genes, indicating that the function of Ser7P goes beyond snRNA processing (11–13). The profiles of Ser2P, Ser5P, and Ser7P overlap in coding genes, hinting at the possibility of simultaneous phosphorylation at Ser2, Ser5, and Ser7 positions. Importantly, Ser7P is specifically enriched over introns (12), suggesting a role for Ser7P in the regulation of co-transcriptional splicing events. How Ser7P mediates the assembly of the splicing complex remains a mystery.

Our results presented here provide a structural interpretation for the connection between the Ser7P enrichment at intron and co-transcriptional splicing events. We show that TCERG1, a transcription elongation regulator that interacts with the splicing factors and the transcribing RNAPII, specifically recognizes the hyperphosphorylated CTD containing Ser2P,
Ser5P, and Ser7P marks. Therefore, Ser7P enrichment at introns may likely serve as a signaling post for recruiting adaptor proteins, such as TCERG1, to couple transcribing RNA Pol II with spliceosomes to regulate co-transcriptional splicing events.


REFERENCES


interaction with the crn-TPR1 motif of Clf1 gives a new insight into the binding mode of FF domains. J. Biol. Chem. 281, 356–364