

Gene Regulation:
**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**

Jiangxin Liu, Shilong Fan, Chul-Jin Lee, Arno
L. Greenleaf and Pei Zhou

J. Biol. Chem. 2013, 288:10890-10901.

doi: 10.1074/jbc.M113.460238 originally published online February 22, 2013



Access the most updated version of this article at doi: [10.1074/jbc.M113.460238](https://doi.org/10.1074/jbc.M113.460238)

Find articles, minireviews, Reflections and Classics on similar topics on the [JBC Affinity Sites](http://www.jbc.org/).

Alerts:

- [When this article is cited](#)
- [When a correction for this article is posted](#)

[Click here](#) to choose from all of JBC's e-mail alerts

This article cites 65 references, 26 of which can be accessed free at
<http://www.jbc.org/content/288/15/10890.full.html#ref-list-1>

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 Liu[‡], Shilong Fan[§], Chul-Jin Lee[‡], Arno L. Greenleaf[‡], and Pei Zhou^{‡#1}

From the [‡]Department of Biochemistry, Duke University Medical Center, Durham, North Carolina 27710 and [§]Laboratory of Molecular Biology, Medical Research Council (MRC) Centre, Cambridge CB2 0QH, United Kingdom

Background: TCERG1 interacts with hyperphosphorylated RNAPII 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 α helices and a 3_{10} helix, FF6 contains an additional insertion helix between $\alpha 1$ and $\alpha 2$. 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 CTD-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 Ser⁷ 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 Pro⁶) 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://www.pdb.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; HYP A, 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 knock-down 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 domain-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 $\alpha 2$, the following loop, the 3_{10} helix, and the N-terminal half of $\alpha 3$ (36). In contrast, NMR titration of the FF1 domain of formin-binding protein (FBP11/HYPA) with a Ser²/Ser⁵ doubly phosphorylated CTD peptide has implicated FBP11 residues at the N-terminal parts of $\alpha 1$ and $\alpha 3$ 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 NdeI and BamHI sites (EMD Biosciences). A series of TCERG1 FF4–6 point mutants (W918E, R922A, R922E, R923A, R926A, W931F, K942A, L953M, W977E, 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 His₁₀ tag were over-expressed in *Escherichia coli* BL21(DE3)STAR cells. Cultures were grown in LB medium in the presence of 100 μ g/ml ampicillin at 37 °C until A_{600} reached 0.6. Cells were induced with 0.25 mM isopropyl β -D-thiogalactopyranoside at 20 °C for 20 h. After harvest by centrifugation, cells were resuspended in the lysis buffer containing 50 mM NaH₂PO₄ (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 $\times g$ for 1 h, and the supernatant was loaded onto a Ni²⁺-nitrilotriacetic acid column. The column was extensively washed with the lysis buffer, and then the protein was eluted with a buffer containing 50 mM NaH₂PO₄ (pH 8), 300 mM NaCl, 250 mM imidazole, and 0.1% β -mercaptoethanol (v/v). The eluted protein was exchanged into the FPLC 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 Ni²⁺-nitrilotriacetic acid column to remove the cleaved His₁₀ 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 (YSPTSPS)₁₄ or all-S7A mutant CTD (YSPTSPA)₁₄ fused to residue Gly¹⁵⁴¹ 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 CTD₁₄ and S7A CTD₁₄, respectively, in this study. WT CTD₁₄ and S7A CTD₁₄ 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 μ l 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 M NaCl, and 12 mM imidazole to eliminate nonspecific binding. A high salt buffer containing 25 mM HEPES (pH 7) and 1 M 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 served as a negative control. The input and elution fractions were analyzed by SDS-PAGE followed by Western blotting with Ser5P-specific CTD antibody 3E8. Peroxidase-conjugated anti-rat IgG (heavy + light) antibody was used as the secondary antibody and was visualized using an enhanced chemiluminescence system (PerkinElmer Life Sciences).

Far-Western Blot Analysis—Duplicate protein samples were loaded into two precast SDS gels (4–20%; Bio-Rad). One was stained with Coomassie Blue, whereas the other one was transferred to a nitrocellulose membrane at 0.75 A for 2 h at 4 °C. The membrane was incubated overnight at 4 °C in the blocking/renaturation buffer containing 1 \times PBS (10 mM Na₂HPO₄/NaH₂PO₄ (pH 7.2), and 150 mM NaCl), 3% nonfat dry milk, 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 μ l 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 μ g 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 μ l of the loading buffer followed by 15-ml buffer wash. Proteins bound to the PCTD peptide column were eluted in three fractions of 200 μ l 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 (PFO-X175/08) 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 Advance 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 deuteration 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). *T*₁ and *T*₂ experiments were conducted to measure the rotational correlation time, τ_c . A series of ¹H-¹⁵N HSQCs 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 (*T*₁) and 10, 20, 30, 40, 50, 60, and 70 ms for transverse relaxation (*T*₂). *T*₁ and *T*₂ values were determined by fitting peak intensities to an exponential decay function using the rate analysis tool in NMRView (50), and rotational correlational time was calculated as described previously (51).

An HNC0-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). ¹D_{NH} RDC data were obtained by taking the difference in ¹J_{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 ^{15}N -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 ^{15}N -labeled FF4–6 sample. HSQC spectra were analyzed, and the chemical shift perturbation was calculated by $\delta = \sqrt{\delta_H^2 + (0.17 \times \delta_N)^2}$, where δ_H and δ_N are chemical shift changes in the ^1H and ^{15}N dimensions, respectively. The dissociation constant K_d was deduced from the Morrison equation, $\delta = \frac{\delta_{\max}}{2L}((L + K_d + P) - \sqrt{(L + K_d + P)^2 - 4PL})$ where δ_{\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 (Ala⁹⁵² between FF4 and FF5 and Asp¹⁰¹⁰ 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 ^{15}N -labeled FF2, FF5–6, and FF4–6 domains of TCERG1. All three proteins display high quality ^1H - ^{15}N 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 correlational 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-

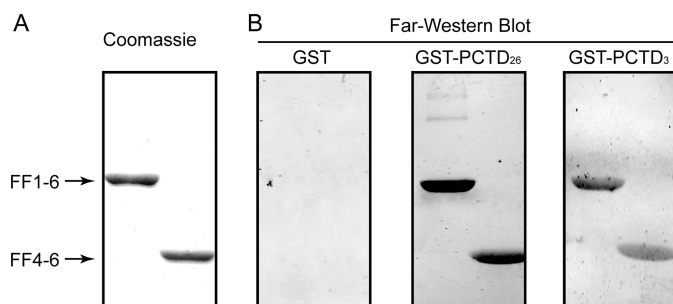


FIGURE 1. Interaction between the tandem FF4–6 repeat of TCERG1 and hyperphosphorylated CTD. A, purified TCERG1 FF1–6 and FF4–6 samples stained with Coomassie Blue on the SDS-PAGE gel. B, far-Western blots probing the interactions of TCERG1 FF1–6 and FF4–6 with GST, hyperphosphorylated GST-yeast CTD containing 26 heptad repeats (GST-PCTD₂₆), and hyperphosphorylated GST-CTD containing three heptad repeats (GST-PCTD₃). Interactions were detected with rabbit anti-GST antibody and IRDye 680 donkey anti-rabbit IgG (heavy + light).

yeast CTD containing 26 heptad repeats (GST-PCTD₂₆; 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-PCTD₃) 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 Ser², Ser⁵, and Ser⁷—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 phosphoepitope requirement using PCTD peptide column binding assays and NMR titration experiments.

Because CTD kinase I has been shown to phosphorylate Ser² within the heptad repeat of the CTD already containing Ser⁵ phosphorylation and generate Ser²/Ser⁵ 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 hyperphosphorylated at Ser² and Ser⁵ positions. To test this hypothesis, we synthesized a three-heptad-repeat CTD peptide containing Ser2P and Ser5P (Y¹pS²pS³T⁴pS⁵pS⁶S⁷Y¹pS²pS³T⁴pS⁵pS⁶S⁷-Y¹pS²pS³T⁴pS⁵pS⁶S⁷; 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

Structure of TCERG1 FF4–6 and Its PCTD Binding Specificity

TABLE 1
Phospho-CTD peptides
pS refers to phospho-Ser.

Peptide	Sequence
7,7-Ser(P)	Biotin-GGGGYSPTSPpSYSPpSYSPpSYSP
2,5,2,5-Ser(P)	Biotin-GGGGSPSYpSPTpSPSYpSPTpSPSYSP
5,2,5,2-Ser(P)	Ac-YSPpTSPSYpSPTpSPSYpSPTSPS
5,7,5,7-Ser(P)	Biotin-GGGGYSPTpSPpSYSPpSPpSYSPpSPTSPS
7,2,7,2-Ser(P)	Biotin-GGGGYSPTSPpSYpSPTSPpSYpSPTSPS
2,5,2,5,2,5-Ser(P)	Biotin-YpSPTpSPSYpSPTpSPSYpSPTpSPS Ac-YpSPTpSPSYpSPTpSPSYpSPTpSPS
2,5,7,2,5,7-Ser(P)	Biotin-GGGSPSYpSPTpSPpSYpSPTpSPpSYSP Ac-SPSYpSPTpSPpSYpSPTpSPpSYSP

5,2,5,2 positions ($Y^1S^2P^3T^4pS^5P^6S^7Y^1pS^2P^3T^4pS^5P^6S^7Y^1pS^2P^3T^4S^5P^6S^7$; referred to as 5,2,5,2-Ser(P)) or 2,5,2,5 positions ($S^5P^6S^7Y^1pS^2P^3T^4pS^5P^6S^7Y^1pS^2P^3T^4$; 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 2,5,7,2,5,7; 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 ¹⁵N 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 $13 \pm 5 \mu\text{M}$ 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 $102 \pm 33 \mu\text{M}$ (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. Pull-down 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 Ser⁷ 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 ~120°. 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-

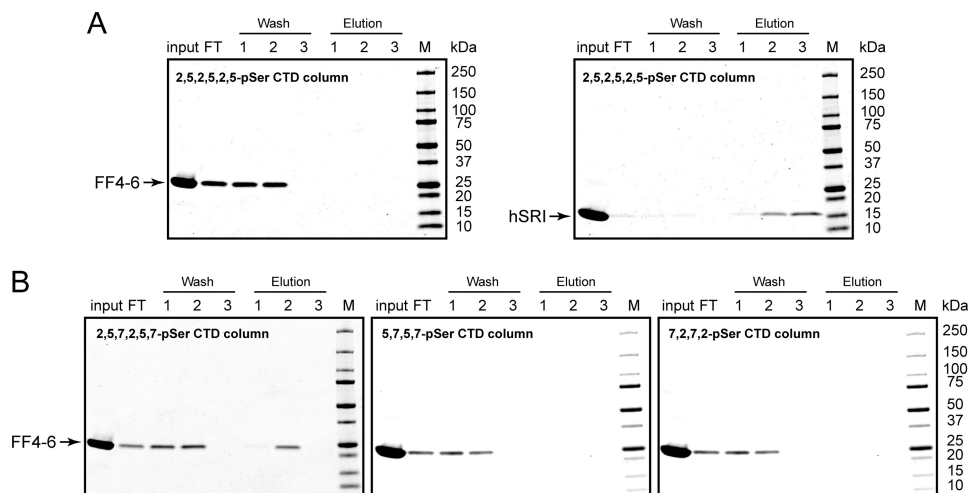


FIGURE 2. Interaction between the TCERG1 FF4–6 domain and PCTD requires simultaneous phosphorylation of Ser², Ser⁵, and Ser⁷ within the heptad repeat of the CTD. *A*, column binding assays of FF4–6 (left) and human SRI (right; positive control) to the 2,5,2,5,2,5-Ser(P) CTD peptide. *B*, column binding 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 μ l of buffer containing 25 mM HEPES (pH 7.0) and 0.1 M NaCl followed by 15-ml buffer wash (Wash, fractions 1–3). Protein bound to the PCTD peptide column was eluted with three fractions of 200 μ 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.

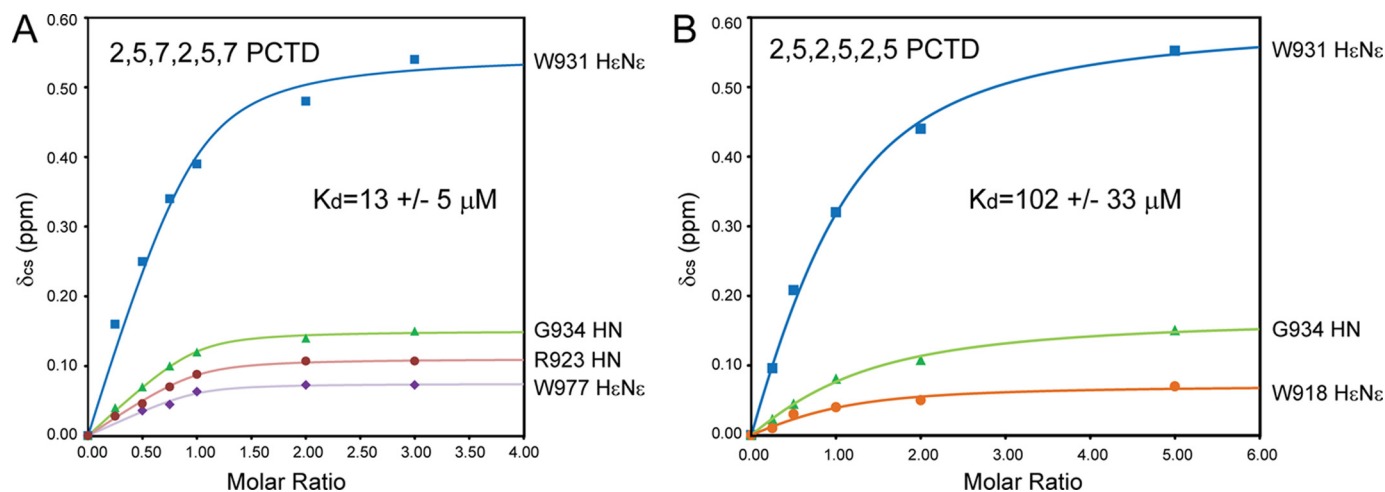


FIGURE 3. Binding affinities of TCERG1 FF4–6 to CTD peptides containing 2,5,7,2,5,7-Ser(P) (A) or 2,5,2,5,2,5-Ser(P) (B) phosphoepitopes determined by NMR titration experiments.

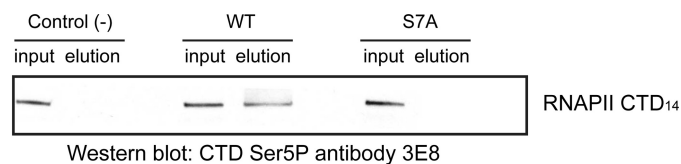


FIGURE 4. TCERG1 FF4–6 selectively pulls down hyperphosphorylated WT RNAPII CTD from yeast lysate but not S7A mutant RNAPII CTD. Purified His₁₀-tagged TCERG1 FF4–6 was immobilized on a cobalt column, and empty cobalt resin was used as a negative control. Whole-cell lysates from yeast strains containing WT RNAPII CTD₁₄ or S7A RNAPII CTD₁₄ 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.

tion of other helices as all canonical helices, including α 1, α 2, 3_{10} and α 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 α 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 α 1– α 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 α 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 of Phe⁹¹² of FF4 as well as an intradomain hydrogen bond involving the side chain of Lys⁹⁵⁷ and backbone carbonyl of Lys⁹⁹² within FF5. In contrast, hydrophobic interactions dominate the FF5–FF6 interface. Residues Thr⁹⁷² and Thr⁹⁷⁴ located in the α 1– α 2 loop of FF5 and residues Leu¹⁰⁶⁰,

Structure of TCERG1 FF4–6 and Its PCTD Binding Specificity

TABLE 2

Crystallographic data, phasing, and refinement statistics

MAD, multiwavelength anomalous dispersion; r.m.s., root mean square.

	Native FF4–6	MAD data collection, Se-Met FF4–6 L953M		
		$\lambda 1$ (remote)	$\lambda 2$ (peak)	$\lambda 3$ (edge)
Data collection and phasing				
Wavelength (Å)	1.0001	0.9719	0.9794	0.9796
Space group	P 1 21 1	P 1 21 1		
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	27.9, 77.1, 95.2	28.1, 78.8, 92.6		
α , β , γ (°)	90, 96, 90	90, 95.6, 90		
Resolution (Å)	50–2.0 (2.03–2.00) ^a	50–2.2 (2.24–2.20)	50–2.3 (2.32–2.28)	50–2.5 (2.59–2.53)
Total reflections	107,683	83,492	107,680	55,548
Unique reflections	26,788	20,151	20,196	16,857
Completeness	98.3 (99.4)	98.1 (90.2)	98.9 (97.2)	93 (91.7)
<i>R</i> _{merge} (%)	6.3 (15.0)	7.3 (32)	8.8 (35.8)	8.3 (32)
<i>I</i> / σ <i>I</i>	33.2 (11.7)	13.8 (2.8)	15.6 (3.0)	11.7 (2.8)
Figure of merit for MAD phasing		0.32		
Refinement				
Resolution range (Å)	24.7–2.0			
Unique reflections	26,272			
<i>R</i> _{work} / <i>R</i> _{free}	0.189/0.240			
r.m.s. deviations				
Bond length (Å)	0.006			
Bond angle (°)	0.917			
B-factor (Å ²)	23.45			
Protein	22.74			
Water	27.64			
Ramachandran plot ^b				
Favored (%)	98.9			
Allowed (%)	1.1			
Disallowed (%)	0			
MolProbity score				
All-atom clashscore	8.66			
Clashscore percentile	87th			

^a Values in parentheses are for the highest resolution shells.

^b Ramachandran plot statistics were generated using MolProbity (46).

Cys¹⁰⁶², and Val¹⁰⁶³ from the loop connecting the 3₁₀ helix and the $\alpha 3$ 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 a ²H/¹³C/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-S²) (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 ¹D_{HN} 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 ($\delta_{cs} > 0.05$ ppm) for the

resonances of the following residues: Ser⁹¹⁹, Arg⁹²³, Arg⁹²⁶, Trp⁹³¹, Gly⁹³⁴, Thr⁹⁷², Thr⁹⁷⁶, Lys⁹⁸¹, Lys⁹⁸², Lys⁹⁸⁵, and Glu⁹⁸⁶ and side chains of Trp⁹¹⁸, Trp⁹³¹, and Trp⁹⁷⁷ (Fig. 7A). The side chain H_N resonance of the Trp⁹³¹ in particular undergoes a large scale chemical shift perturbation ($\delta_{cs} > 0.5$ 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 $\alpha 2$ helices of FF4 (Ser⁹¹⁹, Arg⁹²³, and Arg⁹²⁶) and FF5 (Thr⁹⁷⁶, Trp⁹⁷⁷, Lys⁹⁸¹, and Lys⁹⁸²) and the following 3₁₀ helices of FF4 (Trp⁹³¹ 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

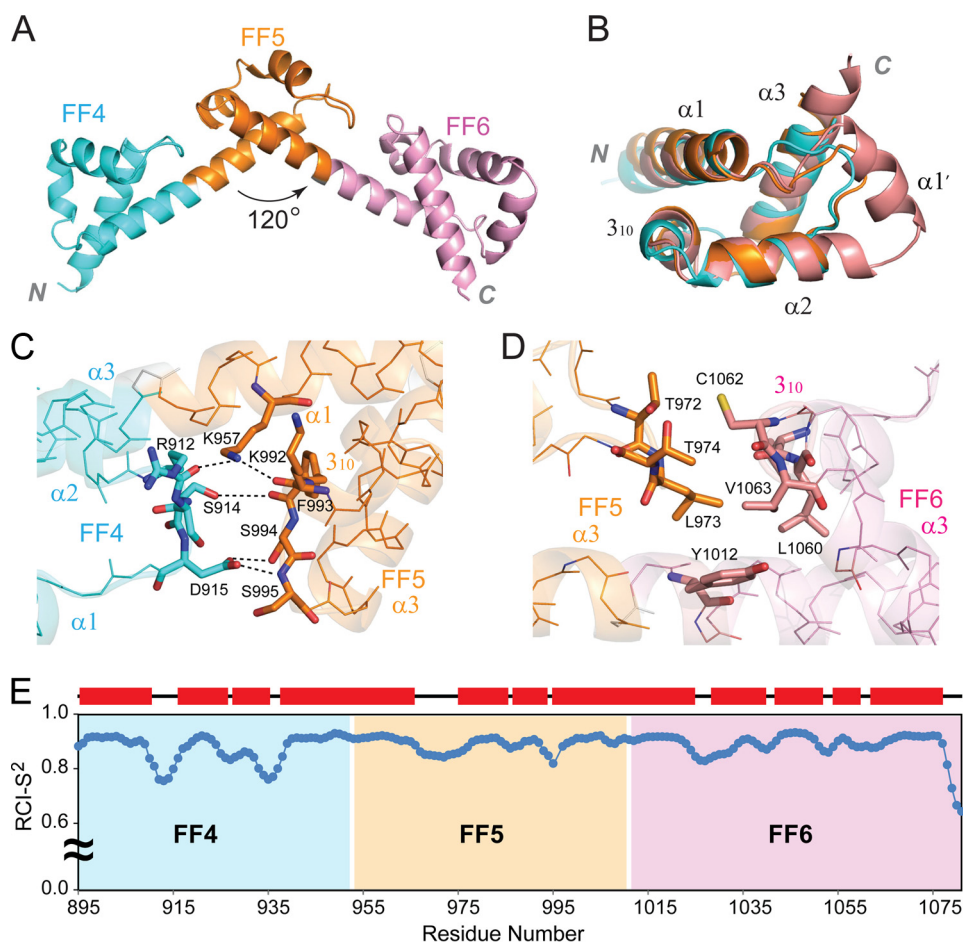


FIGURE 5. Structure of the TCERG1 FF4–6 tandem repeat. *A*, ribbon diagram of the FF4–6 tandem repeat revealing a rigid integral domain. Individual FF domains are color-coded with FF4 in cyan, FF5 in orange, and FF6 in pink. *B*, superimposition of FF4, FF5, and FF6. *C* and *D* depict residues involved in interdomain interactions between FF4 and FF5 and between FF5 and FF6, respectively. *E*, sequence-specific order parameters derived from the random coil index ($RCI-S^2$) (58). Secondary structures are labeled on the top, and regions of individual FF domains are color-coded as in *A*.

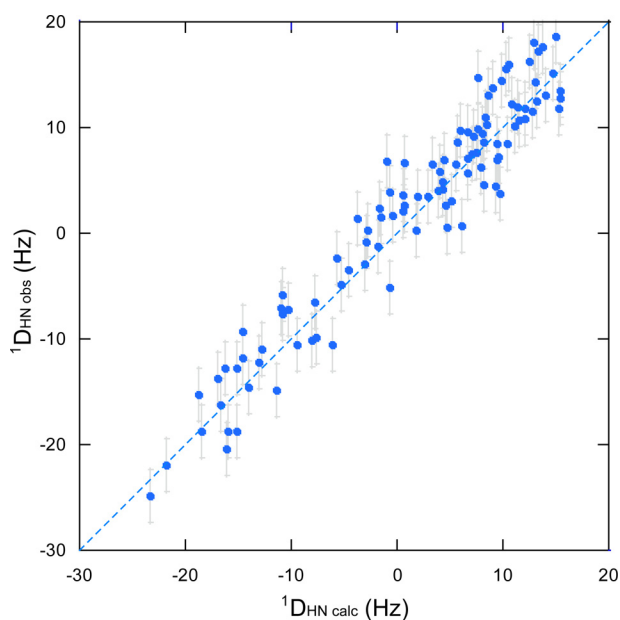


FIGURE 6. Correlation between observed (*obs*) and calculated (*calc*) $^1D_{HN}$ from the crystal structure of TCERG1 FF4–6. Error bars indicate uncertainties of the RDC measurements.

R926A of $\alpha 2$ and K942A of $\alpha 3$ 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 $\alpha 2$ and K1000A or K1000E of $\alpha 3$ 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 $\alpha 2$, the following 3_{10} helix, and the N terminus of $\alpha 3$ of FF4 and FF5 as the primary CTD-docking sites of TCERG1 FF4–6. Interestingly, these CTD-docking sites of the TCERG1 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 $\alpha 1$ and N terminus of $\alpha 3$ (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.

DISCUSSION

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 TCERG1, the structures of individual FF domains have been studied in detail (34, 35) (also see Protein Data Bank codes 2DOD, 2DOE, 2DOF, and 2E71

Structure of TCERG1 FF4–6 and Its PCTD Binding Specificity

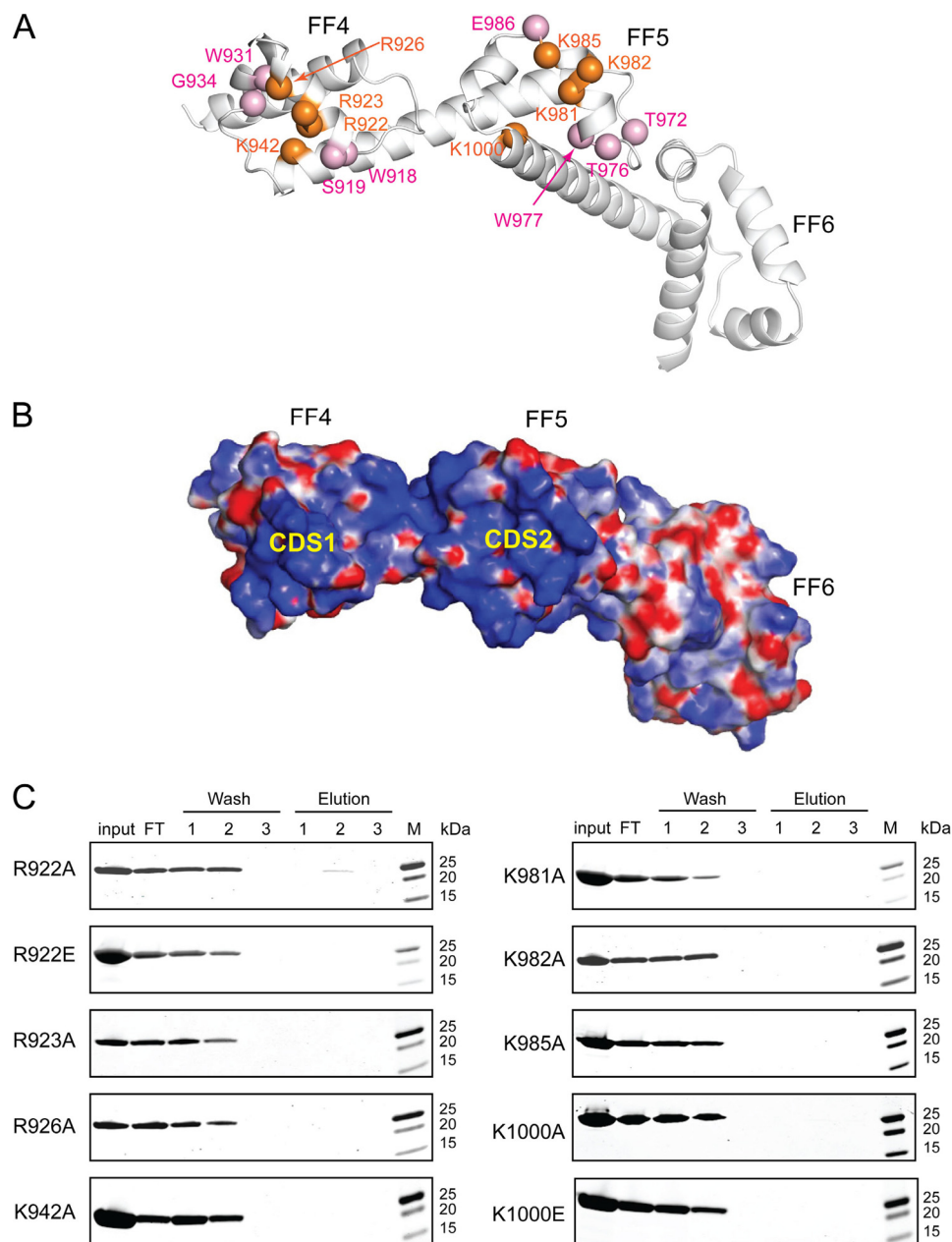


FIGURE 7. PCTD-docking sites of TCERG1 FF4–6. *A*, PCTD recognition by FF4–6 is mediated by residues within FF4 and FF5. TCERG1 residues experiencing resonance perturbations during NMR titration of the 2,5,7,2,5,7-Ser(P) CTD peptide are mapped on the ribbon diagram of FF4–6 with $C\alpha$ atoms colored in pink. Residues important for PCTD interaction as revealed by point mutagenesis studies are also mapped onto the ribbon diagram with $C\alpha$ atoms colored in orange. *B*, electrostatic surface of FF4–6 highlighting the enrichment of basic residues in the two CTD-docking sites (CDS1 and CDS2). *C*, point mutations of basic residues in the CTD-docking sites of FF4–6 disrupt or reduce its interaction with the PCTD in peptide column binding assays. *M*, molecular mass markers.

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 observa-

tions 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 ($\alpha 1'$), 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. Contradictory to this

Structure of TCERG1 FF4–6 and Its PCTD Binding Specificity

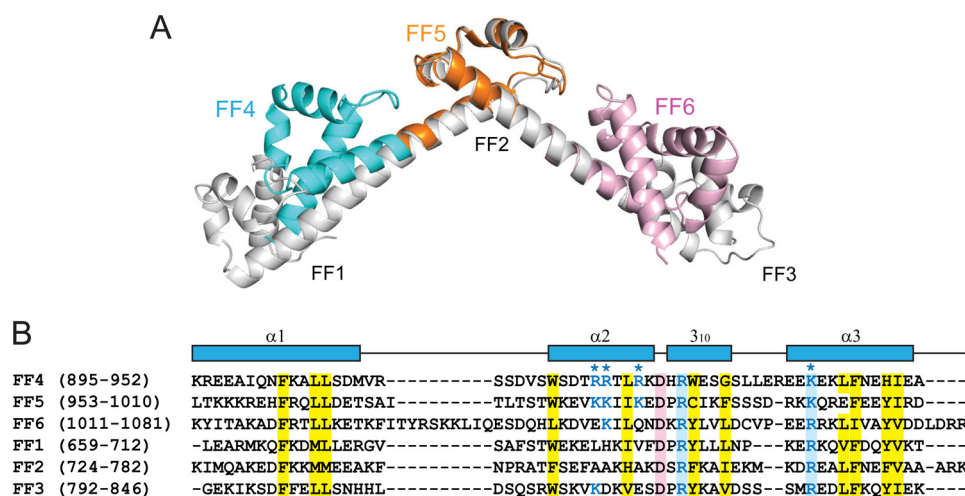


FIGURE 8. **Structural and sequence comparison of tandem FF repeats of TCERG1.** *A*, comparison of the FF1–3 and FF4–6 tandem domains superimposed with FF2 and FF5. *B*, sequence alignment of individual FF domains of TCERG1. Secondary structures are displayed on top of the sequences. Conserved hydrophobic residues are colored in yellow, basic residues are in blue, and acidic residues are in pink. FF4 and FF5 residues important for PCTD interaction are indicated by asterisks.

proposal, 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 $\alpha 2$ and Lys⁹⁴² within FF4 and Lys⁹⁸¹, Lys⁹⁸², and Lys⁹⁸⁵ of $\alpha 2$ and Lys¹⁰⁰⁰ of $\alpha 3$ 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 pI values, do not interact with the 2,5,7,2,5,7-hyperphosphorylated CTD. Therefore, the pI 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,

Structure of TCERG1 FF4–6 and Its PCTD Binding Specificity

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 RNAPII with spliceosomes to regulate co-transcriptional splicing events.

Acknowledgments—Crystal screening was performed at the Duke University Medical Center X-ray Crystallography Shared Resource. X-ray diffraction data were collected at the Southeast Regional Collaborative Access Team (SER-CAT) 22-BM beamline at the Advanced Photon Source, Argonne National Laboratory. Yeast strains with 14 repeats of the consensus sequence (YSPTSPS)₁₄ or all-S7A mutant CTD (YSPTSPA)₁₄ fused to residue Gly¹⁵⁴¹ in *S. cerevisiae* Rpb1 were generously provided by Prof. Beate Schwer (Weill Cornell Medical College). We thank Jeffrey Boyles for assistance with NMR analysis. We thank Dr. Charles W. Pemble IV, Dr. Jon W. Werner-Allen, and Dr. Stuart Endo-Streeter for inspiring discussions and helpful advice; Dr. Nathan I Nicely for assistance during the x-ray data collection; and Bart Bartkowiak, Dr. April MacKellar, and Dr. Pengda Liu for assistance with the far-Western blot assay.

REFERENCES

1. Cramer, P., Bushnell, D. A., and Kornberg, R. D. (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**, 1863–1876
2. Meinhart, A., Kamenski, T., Hoepfner, S., Baumli, S., and Cramer, P. (2005) A structural perspective of CTD function. *Genes Dev.* **19**, 1401–1415
3. Allison, L. A., Wong, J. K., Fitzpatrick, V. D., Moyle, M., and Ingles, C. J. (1988) The C-terminal domain of the largest subunit of RNA polymerase II of *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and mammals: a conserved structure with an essential function. *Mol. Cell. Biol.* **8**, 321–329
4. Corden, J. L. (1990) Tails of RNA polymerase II. *Trends Biochem. Sci.* **15**, 383–387
5. Phatnani, H. P., and Greenleaf, A. L. (2006) Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev.* **20**, 2922–2936
6. Egloff, S., Dienstbier, M., and Murphy, S. (2012) Updating the RNA polymerase CTD code: adding gene-specific layers. *Trends Genet.* **28**, 333–341
7. Komarnitsky, P., Cho, E. J., and Buratowski, S. (2000) Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev.* **14**, 2452–2460
8. Cho, E. J., Kobor, M. S., Kim, M., Greenblatt, J., and Buratowski, S. (2001) Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. *Genes Dev.* **15**, 3319–3329
9. Bartkowiak, B., MacKellar, A. L., and Greenleaf, A. L. (2011) Updating the CTD story: From tail to epic. *Genet. Res. Int.* **2011**, 623718
10. Egloff, S., O'Reilly, D., Chapman, R. D., Taylor, A., Tanzhaus, K., Pitts, L., Eick, D., and Murphy, S. (2007) Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science* **318**, 1777–1779
11. Tietjen, J. R., Zhang, D. W., Rodríguez-Molina, J. B., White, B. E., Akhtar, M. S., Heidemann, M., Li, X., Chapman, R. D., Shokat, K., Keles, S., Eick, D., and Ansari, A. Z. (2010) Chemical-genomic dissection of the CTD code. *Nat. Struct. Mol. Biol.* **17**, 1154–1161
12. Kim, H., Erickson, B., Luo, W., Seward, D., Graber, J. H., Pollock, D. D., Megee, P. C., and Bentley, D. L. (2010) Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nat. Struct. Mol. Biol.* **17**, 1279–1286
13. Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Söding, J., and Cramer, P. (2010) Uniform transitions of the general RNA polymerase II transcription complex. *Nat. Struct. Mol. Biol.* **17**, 1272–1278
14. Baskaran, R., Dahmus, M. E., and Wang, J. Y. (1993) Tyrosine phosphorylation of mammalian RNA polymerase II carboxyl-terminal domain. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 11167–11171
15. Zhang, J., and Corden, J. L. (1991) Identification of phosphorylation sites in the repetitive carboxyl-terminal domain of the mouse RNA polymerase II largest subunit. *J. Biol. Chem.* **266**, 2290–2296
16. Kelly, W. G., Dahmus, M. E., and Hart, G. W. (1993) RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc. *J. Biol. Chem.* **268**, 10416–10424
17. Shaw, P. E. (2007) Peptidyl-prolyl cis/trans isomerases and transcription: is there a twist in the tail? *EMBO Rep.* **8**, 40–45
18. Werner-Allen, J. W., Lee, C. J., Liu, P., Nicely, N. I., Wang, S., Greenleaf, A. L., and Zhou, P. (2011) cis-Proline-mediated Ser(P)5 dephosphorylation by the RNA polymerase II C-terminal domain phosphatase Ssu72. *J. Biol. Chem.* **286**, 5717–5726
19. Xiang, K., Nagaike, T., Xiang, S., Kilic, T., Beh, M. M., Manley, J. L., and Tong, L. (2010) Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex. *Nature* **467**, 729–733
20. Buratowski, S. (2003) The CTD code. *Nat. Struct. Mol. Biol.* **10**, 679–680
21. Egloff, S., and Murphy, S. (2008) Cracking the RNA polymerase II CTD code. *Trends Genet.* **24**, 280–288
22. Suñé, C., Hayashi, T., Liu, Y., Lane, W. S., Young, R. A., and Garcia-Blanco, M. A. (1997) CA150, a nuclear protein associated with the RNA polymerase II holoenzyme, is involved in Tat-activated human immunodeficiency virus type 1 transcription. *Mol. Cell. Biol.* **17**, 6029–6039
23. Sánchez-Alvarez, M., Goldstrohm, A. C., Garcia-Blanco, M. A., and Suñé, C. (2006) Human transcription elongation factor CA150 localizes to splicing factor-rich nuclear speckles and assembles transcription and splicing components into complexes through its amino and carboxyl regions. *Mol. Cell. Biol.* **26**, 4998–5014
24. Smith, M. J., Kulkarni, S., and Pawson, T. (2004) FF domains of CA150 bind transcription and splicing factors through multiple weak interactions. *Mol. Cell. Biol.* **24**, 9274–9285
25. Jurica, M. S., Licklider, L. J., Gygi, S. R., Grigorieff, N., and Moore, M. J. (2002) Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *RNA* **8**, 426–439
26. Goldstrohm, A. C., Albrecht, T. R., Suñé, C., Bedford, M. T., and Garcia-Blanco, M. A. (2001) The transcription elongation factor CA150 interacts with RNA polymerase II and the pre-mRNA splicing factor SF1. *Mol. Cell. Biol.* **21**, 7617–7628
27. Zhou, Z., Licklider, L. J., Gygi, S. P., and Reed, R. (2002) Comprehensive proteomic analysis of the human spliceosome. *Nature* **419**, 182–185
28. Lin, K. T., Lu, R. M., and Tarn, W. Y. (2004) The WW domain-containing proteins interact with the early spliceosome and participate in pre-mRNA splicing *in vivo*. *Mol. Cell. Biol.* **24**, 9176–9185
29. Pearson, J. L., Robinson, T. J., Muñoz, M. J., Kornblihtt, A. R., and Garcia-Blanco, M. A. (2008) Identification of the cellular targets of the transcription factor TCERG1 reveals a prevalent role in mRNA processing. *J. Biol. Chem.* **283**, 7949–7961
30. Carty, S. M., Goldstrohm, A. C., Suñé, C., Garcia-Blanco, M. A., and Greenleaf, A. L. (2000) Protein-interaction modules that organize nuclear function: FF domains of CA150 bind the phosphoCTD of RNA polymerase II. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9015–9020
31. Bedford, M. T., and Leder, P. (1999) The FF domain: a novel motif that often accompanies WW domains. *Trends Biochem. Sci.* **24**, 264–265
32. Bonet, R., Ramirez-Espain, X., and Macias, M. J. (2008) Solution structure of the yeast URN1 splicing factor FF domain: comparative analysis of charge distributions in FF domain structures-FFs and SURPs, two domains with a similar fold. *Proteins* **73**, 1001–1009
33. Allen, M., Friedler, A., Schon, O., and Bycroft, M. (2002) The structure of an FF domain from human HYPA/FBP11. *J. Mol. Biol.* **323**, 411–416
34. Zeng, J., Boyles, J., Tripathy, C., Wang, L., Yan, A., Zhou, P., and Donald, B. R. (2009) High-resolution protein structure determination starting with a global fold calculated from exact solutions to the RDC equations. *J. Biomol. NMR* **45**, 265–281
35. Murphy, J. M., Hansen, D. F., Wiesner, S., Muhandiram, D. R., Borg, M., Smith, M. J., Sicheri, F., Kay, L. E., Forman-Kay, J. D., and Pawson, T. (2009) Structural studies of FF domains of the transcription factor CA150 provide insights into the organization of FF domain tandem arrays. *J. Mol. Biol.* **393**, 409–424
36. Gasch, A., Wiesner, S., Martin-Malpartida, P., Ramirez-Espain, X., Ruiz, L., and Macias, M. J. (2006) The structure of Prp40 FF1 domain and its

- 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
37. Ester, C., and Uetz, P. (2008) The FF domains of yeast U1 snRNP protein Prp40 mediate interactions with Luc7 and Snu71. *BMC Biochem.* **9**, 29
 38. Lu, M., Yang, J., Ren, Z., Sabui, S., Espejo, A., Bedford, M. T., Jacobson, R. H., Jeruzalmi, D., McMurray, J. S., and Chen, X. (2009) Crystal structure of the three tandem FF domains of the transcription elongation regulator CA150. *J. Mol. Biol.* **393**, 397–408
 39. Schwer, B., and Shuman, S. (2011) Deciphering the RNA polymerase II CTD code in fission yeast. *Mol. Cell* **43**, 311–318
 40. Phatnani, H. P., and Greenleaf, A. L. (2004) Identifying phosphoCTD-associating proteins. *Methods Mol. Biol.* **257**, 17–28
 41. Otwinowski, Z., and Minor, W. (1997) Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326
 42. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1948–1954
 43. Terwilliger, T. C., and Berendzen, J. (1999) Automated MAD and MIR structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **55**, 849–861
 44. Terwilliger, T. C. (2000) Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr.* **56**, 965–972
 45. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132
 46. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21
 47. Cavanagh, J., Fairbrother, W. J., Palmer, A. G. I., Skelton, J. N., and Rance, M. (2007) *Protein NMR Spectroscopy: Principles and Practice*, 2nd Ed., Elsevier Academic Press, Burlington, MA
 48. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293
 49. Goddard, T. D., and Kneller, D. G. (2008) *Sparky 3*, University of California, San Francisco
 50. Johnson, B. A., and Blevins, R. A. (1994) NMRView: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4**, 603–614
 51. Kay, L. E., Torchia, D. A., and Bax, A. (1989) Backbone dynamics of proteins as studied by ¹⁵N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* **28**, 8972–8979
 52. Yang, D., Venters, R. A., Mueller, G. A., Choy, W. Y., and Kay, L. E. (1999) TROSY-based HNC0 pulse sequences for the measurement of ¹HN-¹⁵N, ¹⁵N-¹³CO, ¹HN-¹³CO, ¹³CO-¹³C^α and ¹HN-¹³C^α dipolar couplings in ¹⁵N, ¹³C, ²H-labeled proteins. *J. Biomol. NMR* **14**, 333–343
 53. Dosset, P., Hus, J. C., Marion, D., and Blackledge, M. (2001) A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings. *J. Biomol. NMR* **20**, 223–231
 54. Jones, J. C., Phatnani, H. P., Haystead, T. A., MacDonald, J. A., Alam, S. M., and Greenleaf, A. L. (2004) C-terminal repeat domain kinase I phosphorylates Ser2 and Ser5 of RNA polymerase II C-terminal domain repeats. *J. Biol. Chem.* **279**, 24957–24964
 55. Phatnani, H. P., Jones, J. C., and Greenleaf, A. L. (2004) Expanding the functional repertoire of CTD kinase I and RNA polymerase II: novel phosphoCTD-associating proteins in the yeast proteome. *Biochemistry* **43**, 15702–15719
 56. Li, M., Phatnani, H. P., Guan, Z., Sage, H., Greenleaf, A. L., and Zhou, P. (2005) Solution structure of the Set2-Rpb1 interacting domain of human Set2 and its interaction with the hyperphosphorylated C-terminal domain of Rpb1. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 17636–17641
 57. Kim, M., Suh, H., Cho, E. J., and Buratowski, S. (2009) Phosphorylation of the yeast Rpb1 C-terminal domain at serines 2, 5, and 7. *J. Biol. Chem.* **284**, 26421–26426
 58. Berjanskii, M. V., and Wishart, D. S. (2005) A simple method to predict protein flexibility using secondary chemical shifts. *J. Am. Chem. Soc.* **127**, 14970–14971
 59. Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44**, 213–223
 60. Lipsitz, R. S., and Tjandra, N. (2004) Residual dipolar couplings in NMR structure analysis. *Annu. Rev. Biophys. Biomol. Struct.* **33**, 387–413
 61. Zhang, D. W., Rodríguez-Molina, J. B., Tietjen, J. R., Nemeč, C. M., and Ansari, A. Z. (2012) Emerging views on the CTD code. *Genet. Res. Int.* **2012**, 347214
 62. Kubicek, K., Cerna, H., Holub, P., Pasulka, J., Hrossova, D., Loehr, F., Hofr, C., Vanacova, S., and Stefl, R. (2012) Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. *Genes Dev.* **26**, 1891–1896
 63. Schroeder, S. C., Schwer, B., Shuman, S., and Bentley, D. (2000) Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes Dev.* **14**, 2435–2440
 64. Ghosh, A., Shuman, S., and Lima, C. D. (2011) Structural insights to how mammalian capping enzyme reads the CTD code. *Mol. Cell* **43**, 299–310
 65. McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S. D., Wickens, M., and Bentley, D. L. (1997) The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**, 357–361
 66. Lunde, B. M., Reichow, S. L., Kim, M., Suh, H., Leeper, T. C., Yang, F., Mutschler, H., Buratowski, S., Meinhart, A., and Varani, G. (2010) Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nat. Struct. Mol. Biol.* **17**, 1195–1201
 67. Chapman, R. D., Heidemann, M., Albert, T. K., Mailhammer, R., Flatley, A., Meisterernst, M., Kremmer, E., and Eick, D. (2007) Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science* **318**, 1780–1782