Supplementary Information

Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain

Bradley M. Lunde[1,2], Steve L. Reichow[3], Minkyu Kim[4,5], Hyunsuk Suh[4], Thomas C. Leeper[3], Fan Yang[3], Hannes Mutschler[6], Stephen Buratowski[4], Anton Meinhart[6], and Gabriele Varani[1,3,*]

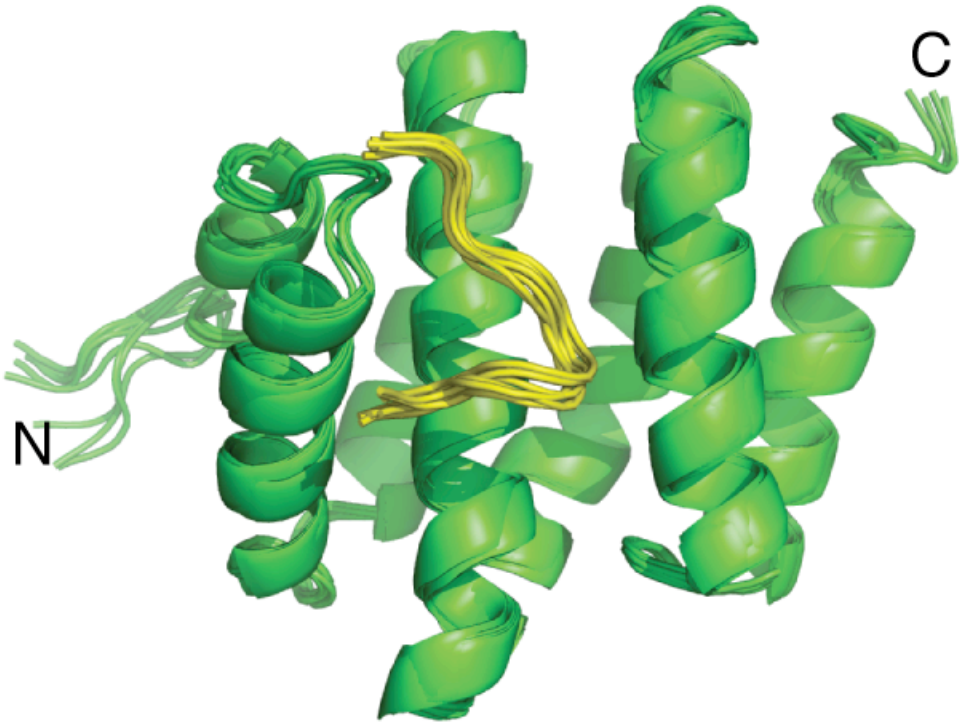
[1]Department of Biochemistry, [2]Biomolecular Structure and Design Program, [3]Department of Chemistry, University of Washington, Seattle, Washington 98195, USA

[4]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115

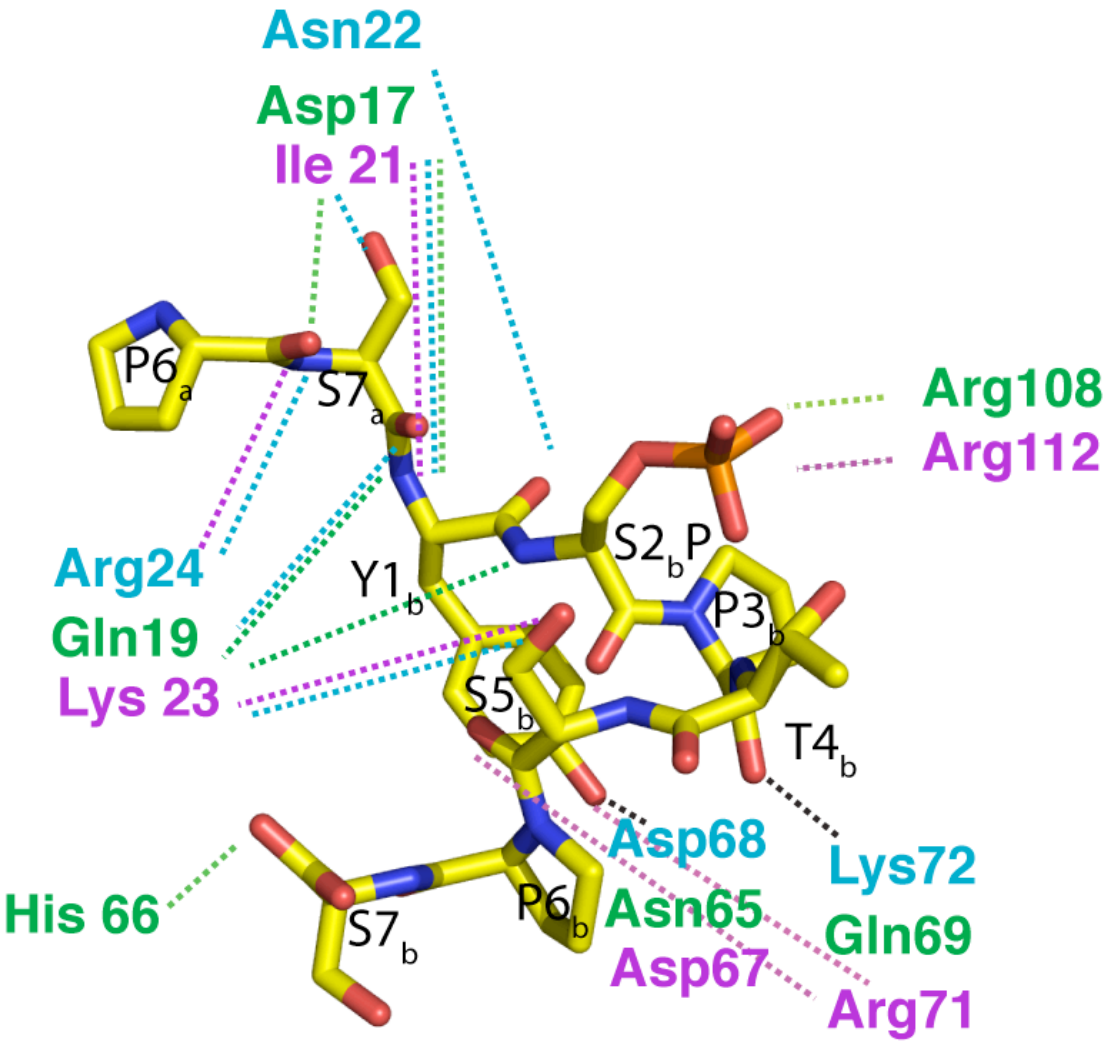
[5] Department of Biophysics and Chemical Biology, Seoul National University, Seoul 151-742, Korea

[6]Department of Biomolecular Mechanisms, Max-Planck-Institute for Medical Research, Jahnstr 29, 69120 Heidelberg, Germany

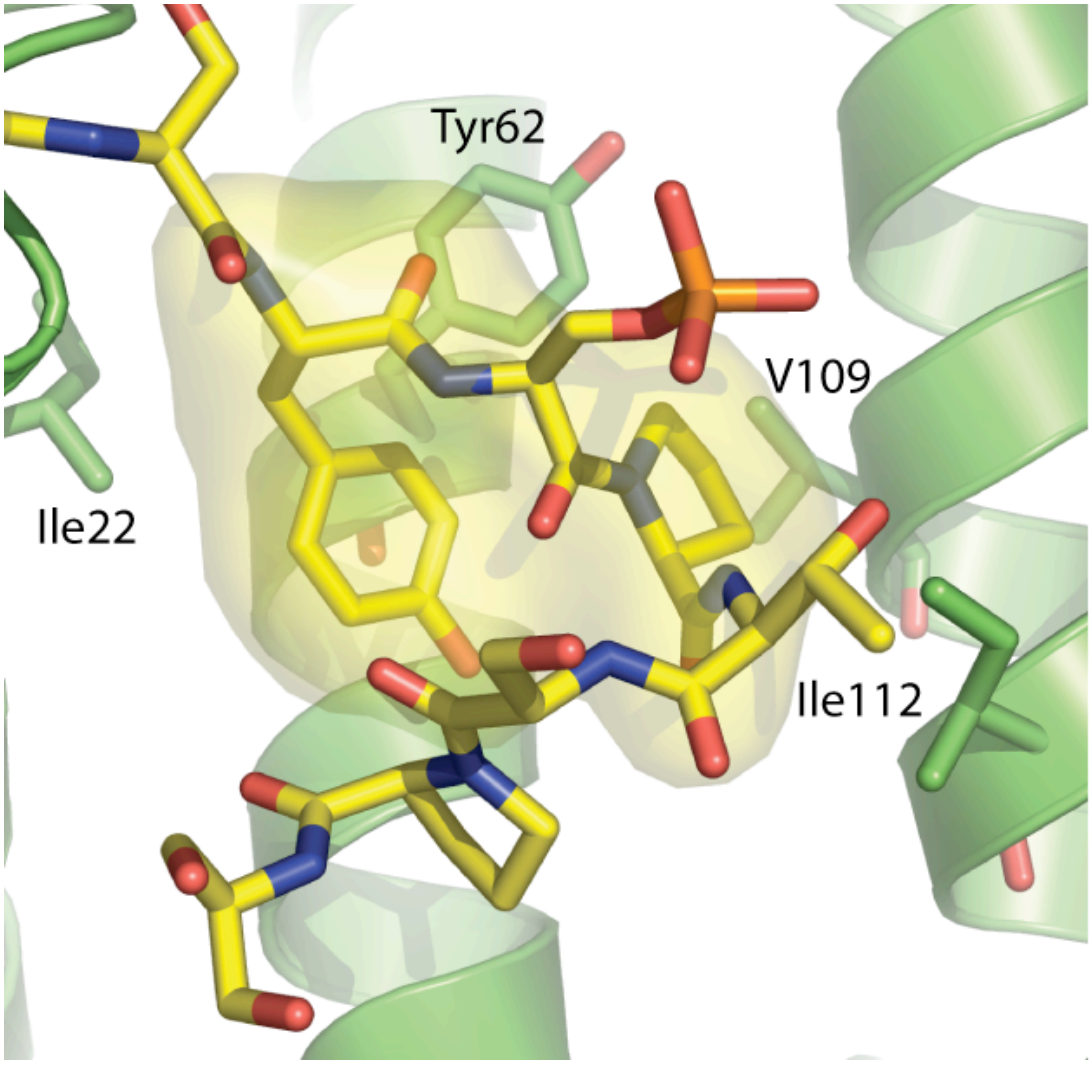
Supplementary Figures



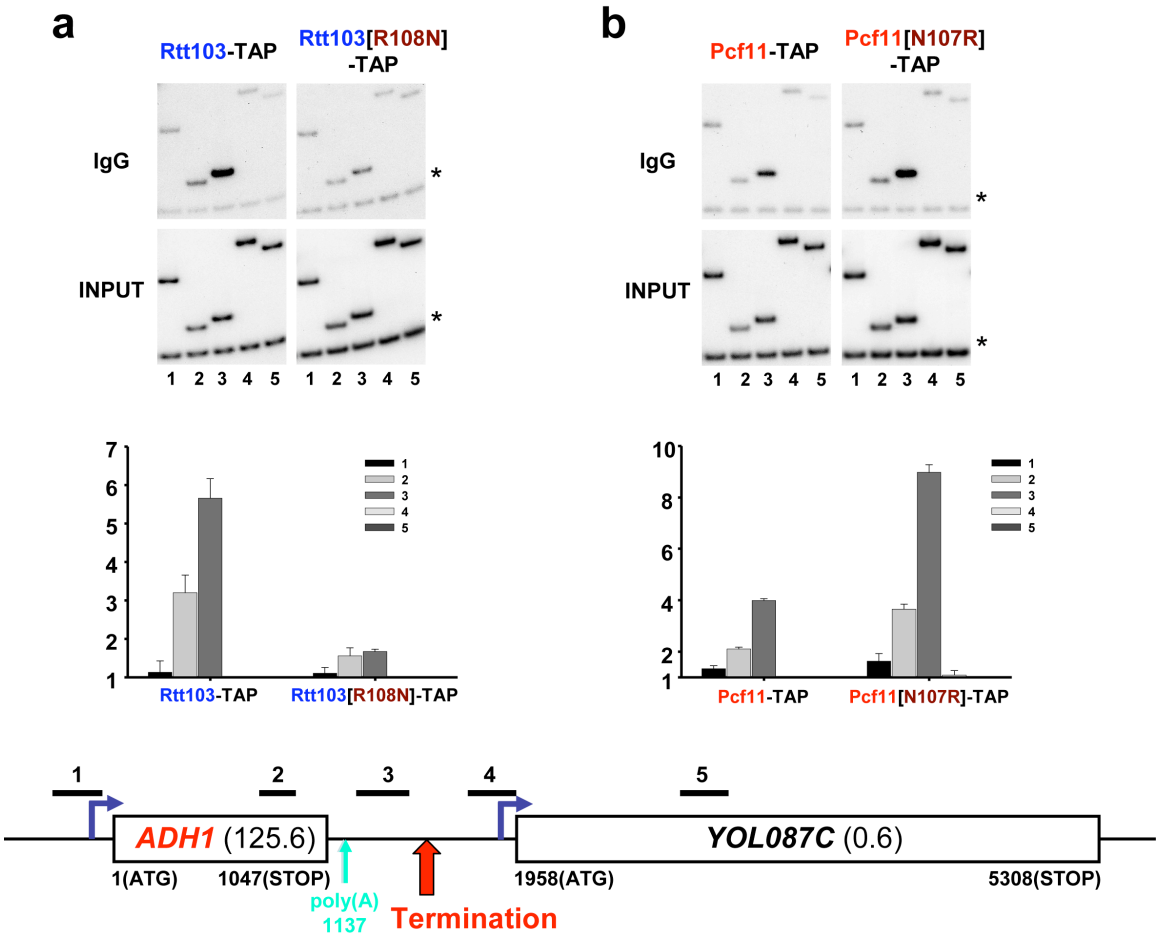
Supplementary Figure 1: Superposition of 10 lowest energy structures from HADDOCK of the Rtt103-CID bound to the Ser2P CTD. The average Cα RMSD for the complex is 0.94 ± 0.08 Å. Rtt103 and the Ser2P CTD peptide are shown as ribbons.



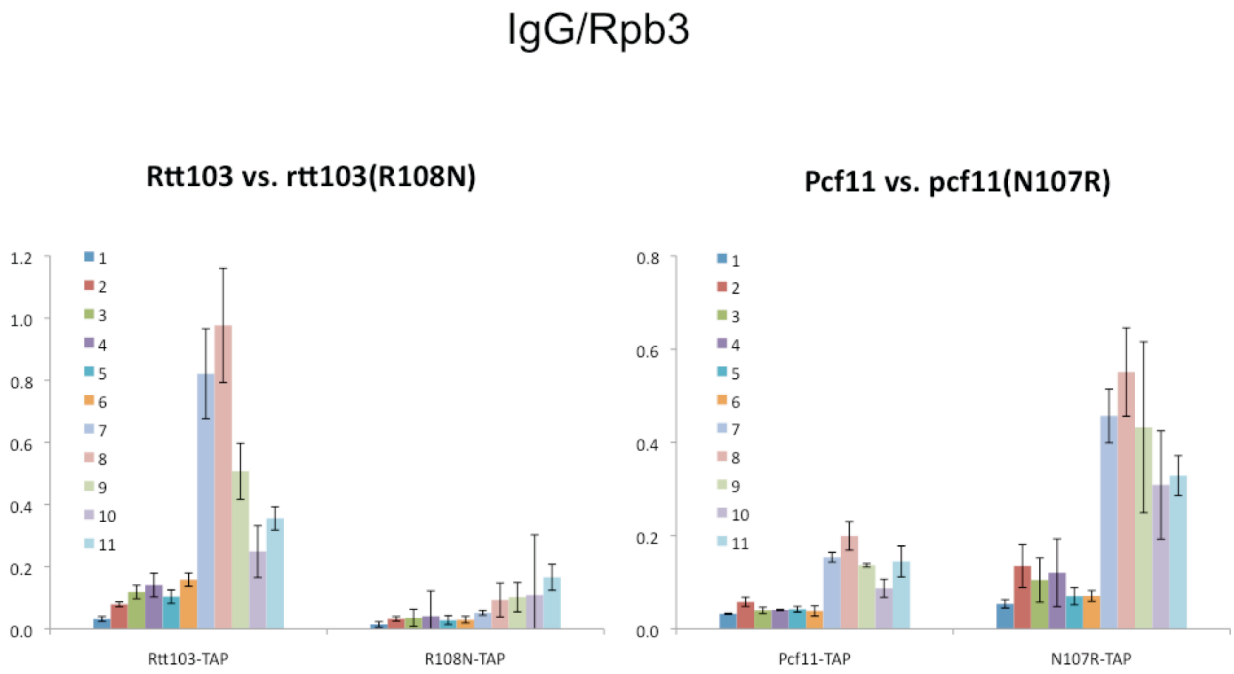
Supplementary Figure 2: Comparison of CID residues that make specific contacts to the CTD for Rtt103 (green), Pcf11 (cyan), and SCAF8 (magenta). The CTD is shown colored in yellow. Dashed lines indicate the contacts between each protein residue and the CTD.



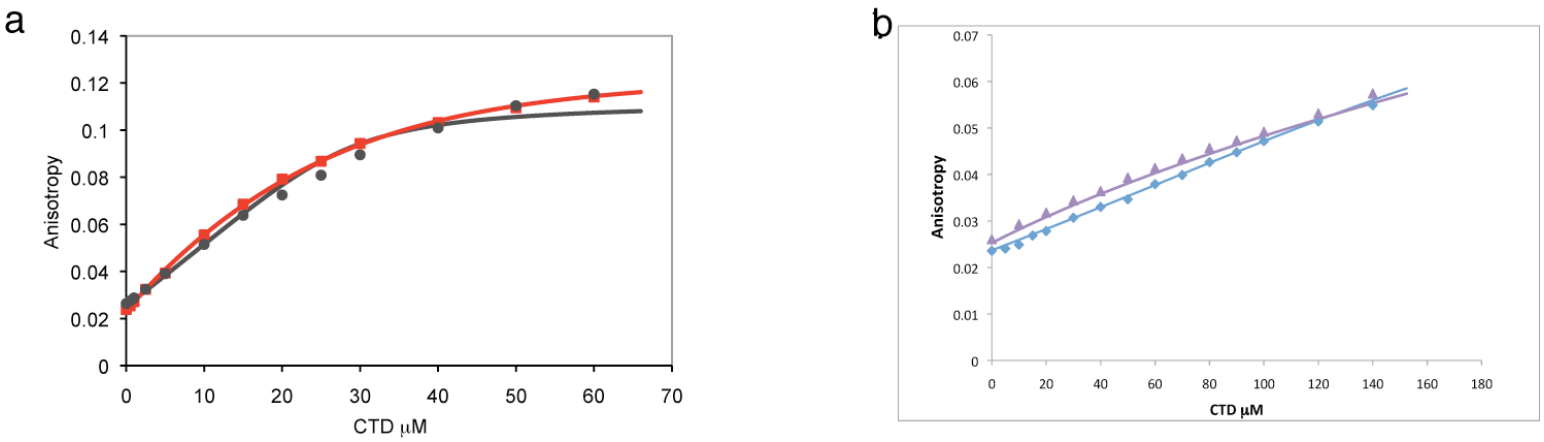
Supplementary Figure 3: Close-up view of the CTD binding pocket for Rtt103 showing the binding site of the CTD Tyr1 and Pro3 shown in yellow. CID residues that contact the CTD Tyr1 and Pro3 are labeled and shown in the stick representation.



Supplementary Figure 4: Recruitment of CID-mutated Rtt103 and Pcf11 on the *ADH1* gene. (**a**) Recruitment of wild type Rtt103 and Rtt103 (R108N) was monitored by ChIP assay using C-terminal TAP tag. Immunoprecipitated DNA was amplified with *ADH1* primers as depicted at the bottom. The top band is the *ADH1*-specific band, while the common lower band (marked by an asterisk) is an internal background control from a non-transcribed region on chromosome VI. The middle panel shows quantification of ChIP data, expressed as fold enrichment over the background. (**b**) Recruitment of wild type Pcf11 and Pcf11 (N107R) by ChIP assay.



Supplementary Figure 5: Normalized data for Pcf11(N107R) and Rtt103 (R108N) mutants. Chromatins were prepared from Rtt103-TAP, rtt103(R108N)-TAP, Pcf11-TAP, and pcf11(N107R)-TAP strains and subsequently immunoprecipitated (ChIPed) with IgG sepharose or anti-Rpb3 antibody. Shown is the Fold enrichment of each TAP-tagged protein along the *PMA1* gene which is normalized against Rpb3 level.



Supplementary Figure 6: Fluorescence Anisotropy data for the 4-repeat CTD peptides. (**a**) wtRtt103 (red squares) and Rtt103 E115R (grey circles) titrated into 2 mM FAM-labeled Ser2P-CTD diheptad repeat peptide which competes against 25 mM Ser2P-CTD 4-repeat peptide for binding. (**b**) Titration of wtPcf11 except in this case the competing peptide is either 150 mM Ser2P-CTD 4-repeat peptide (purple triangles) or 150 mM $Ser2P_{x2}noP_{x2}$ (third peptide Fig. 5a).

Supplementary Table 1

NMR distance and dihedral constraints	Free	Bound
Intraresidue (i-j = 0)	718	716
Sequential (|i-j| = 1)	613	734
Medium range (2 ≤ |i-j| ≤ 4)	815	968
Long range (|i-j| ≥ 5)	612	814
Intermolecular	-	10
AIRs	-	
Hydrogen bonds	96	66
Total	2662	3308
Dihedral angle restraints from TALOS		
Phi (ϕ), Psi(ψ)	232	220
Structure Statistics		
Average pairwise RMSD		
Backbone heavy atoms	0.56 ± 0.10	0.46 ± 0.09
All heavy atoms	1.27 ± 0.14	1.02 ± 0.07
Violations		
Distances >0.5Å and dihedral angles >5°	0	0
Ramachandran plot analysis (% residues)		
Most favorable region	93.7	94.3
Additional allowed regions	6.3	5.7
Generously allowed regions	0	0
Diallowed regions	0	0

Supplementary Table 1: NMR structural statistics for the 20 lowest energy conformers of Rtt103-CID free and bound to the Ser2 phosphorylated CTD calculated using CYANA 2.1. Hydrogen bonding restraints were determined from protected amides in D_2O. Dihedral angle restraints determined using TALOS. Ramachandran statistics were calculated using PROCHECK-NMR.

Supplementary Methods

Plasmids: The plasmid used to produce Pcf11-CID has been previously described[1]. The DNA encoding residues 1-131 of yeast Rtt103-CID was amplified by PCR and inserted into a pET28b expression vector (Novagen) yielding an expression construct with a C-terminal hexa-histidine tag connected by a 5-residue linker. The construct has a single Pro->Ala point mutation at residue 2 due to the use of the NcoI site at the N-terminus; Rtt103-CID (R108N) and Pcf11-CID (N107R) variants were obtained by site-directed mutagenesis using Stratagene's Quikchange site-directed mutagenesis kit. All constructs were verified by DNA sequencing.

Protein Expression and purification: The Pcf11-CID was expressed and purified as previously described[1]. The Rtt103-CID was expressed into *E. coli* BL21(DE3) RIL cells (Stratagene). Cells harboring the plasmid were grown at 310 K in either 2XYT or M9 minimal media, as required, until the OD_{600} reached 0.6. Protein expression was induced by adding IPTG to a final concentration of 0.5 mM, and expression carried out at 293 K for 18-20 hours. Cells were harvested by centrifugation (6,000 x g), resuspended in Buffer A (50 mM potassium phosphate, pH 6.8, 500 mM KCl, 10 mM β-mercaptoethanol) supplemented with a complete protease inhibitor cocktail tablet (Roche) and stored at -70°C. ^{15}N- and $^{15}N/^{13}C$-labeled Rtt103 and Pcf11 proteins were prepared in the same way using M9 minimal media containing 0.5 g/L $^{15}NH_4Cl$

and/or 2 g/L $^{13}C_6$-D-glucose as the sole nitrogen and/or carbon source, respectively.

The cell pellets were thawed on ice and soluble protein was extracted by sonication and clarified using centrifugation (30,000 x g). The protein was purified using a 5 ml HisTrap FF Ni-Sepharose (GE Healthcare) column followed by gel filtration on a 16/60 Superdex 75 column (GE Healthcare). Purified samples were concentrated using an Amicon centrifugal filter device (MWCO 3,000 Da) and either used directly for NMR or flash-frozen and stored at -70°C for later use. All samples were purified to homogeneity and were free of impurities as determined by SDS-PAGE and Coomassie blue staining.

Methods

Protein and peptide preparation: All peptides were purchased from AnaSpec (San Jose, CA) and desalted using a G-10 sephadex resin prior to use.

NMR Spectroscopy: NMR experiments on free Rtt103-CID were recorded on a $^{15}N/^{13}C$-labeled sample at ~1 mM protein concentration, prepared in 35 mM potassium phosphate, pH 6.8, 100 mM KCl in either 92 % H_2O/8 % D_2O or 100 % D_2O. Data were collected at 25°C on Bruker DMX 750, 500 and 600 MHz spectrometers and a Varian 800 MHz spectrometer at the Environmental Molecular Sciences Laboratory (EMSL) at Pacifc Northwest National Laboratory

(PNNL). Spectra were processed using NMRPipe[2] and analyzed using Sparky (http://www.cgl.ucsf.edu/home/sparky/).

Backbone and side-chain resonance assignments were obtained from standard heteronuclear NMR experiments[3]. NOE distance restraints for structure determination were obtained from 3D ^{15}N-edited NOESY, ^{13}C-edited NOESY and 2D-NOESY collected in 100 % D_2O (for aromatic resonances) with a mixing time of 100 ms. T_1 and T_2 relaxation experiments were collected for the free protein, as well as for Rtt103 and Pcf11 bound to both 2- and 4-repeat Ser2 phosphorylated CTD peptides[4].

The structure of Rtt103-CID bound to the CTD was determined using multiple samples. For most experiments, the sample consisted of a 1:1.5 molar ratio of Rtt103 to the CTD Ser2P diheptad repeat in a 10 mM deuterated Tris-HCl buffer, pH 7.4, 120 mM NaCl. Assignments of Rtt103 bound to CTD peptides were obtained by following each resonance in both ^{1}H-^{15}N HSQC and ^{1}H-^{13}C-HSQC titrations until a fully bound complex was observed. These assignments were confirmed with HCCH-TOCSY and 3D ^{15}N-edited and ^{13}C-edited NOESYs. Assignments for the bound peptide were obtained from 2D ^{13}C/^{15}N-filtered TOCSYs and NOESYs (100 and 300 ms mixing time) recorded in both water and D_2O[5]. Intermolecular NOEs between Rtt103 and the CTD peptide were observed with 2D-filtered-edited NOESY experiments[6].

NMR structure calculation and analysis: NOE assignments and structure calculation for both free Rtt103-CID and for Rtt103-CID bound to the CTD were performed using CYANA2.1 (ref. 7). Backbone dihedral angles restraints (ϕ/ψ) were included from the chemical-shift based TALOS database[8]. Preliminary structural analysis was combined with amide D_2O protection data to identify hydrogen-bonding restraints (1.8-2.0 Å) for final refinement. Structures of the CTD peptide were calculated using CNS based on NOE distances from the filtered NOESYs. Structural statistics are given in Supplementary Table 1. PDB coordinates were visualized and prepared for illustration using PyMOL[9].

Structure of the Rtt103-CTD complex: The 10 best scoring structures of the CTD from CNS were docked against the 20 top scoring structures from CYANA2.1 using HADDOCK 2.0. Active ambiguous interaction restraints (AIRs) for Rtt103 were defined as residues that showed an average chemical shift perturbation of at least 0.1 p.p.m. and greater than 50 % solvent accessibility as calculated by NACCESS; passive AIRs were defined as residues neighboring active restraints with greater than 50 % solvent accessibility (AIRs are listed in Supplemental Methods). For the CTD peptide, residues involved in the formation of the β-turn and those C-terminal to the turn were defined as active restraints, while the rest of the peptide was considered passive. Additionally, docking was performed using 10 intermolecular NOEs between the CTD $Tyr1_b$, $Pro3_b$, $Thr4_b$, and the protein. Docking was performed using HADDOCK in the default

configuration[10] and the lowest energy cluster of structures was used for further analysis.

NMR titration: The strength of the interaction between Pcf11-CID and Rtt103-CID with various CTD phosphoisoforms was measured using NMR titration performed on 0.1 mM Pcf11-CID or 0.3 mM Rtt103-CID samples in 10 mM Tris-HCl, pH 7.4, 120 mM NaCl (the Pcf11 sample also contained 2 mM DTT). The CTD peptides were added in a stepwise manner to give peptide:protein ratios ranging from 0:1 to 14:1. At each titration point, we recorded a ^{1}H-^{15}N HSQC and the chemical shift of each amide resonance in each protein was measured. Combined average chemical shift perturbations were calculated as $\Delta\delta_{avg}=([0.2\Delta\delta_N]^2+[\Delta\delta_H]^2)^{1/2}$, where $\Delta\delta_H$ and $\Delta\delta_N$ are the chemical shift changes in the ^{1}H and ^{15}N dimensions, respectively. Apparent binding constants for well-resolved resonances in fast-exchange were determined using the software XCRVFIT by plotting the chemical shift change vs [L]/[P] where [L] and [P] are the total ligand and protein concentrations[11]:

$$\Delta = \Delta_0 * \frac{(K_d+[L]+[P]) - \sqrt{(K_d+[L]+[P])^2 - 4[P][L]}}{2[P]}$$

(Equation 1)

Here Δ is the observed change in chemical shift, Δ_0 is the total change in chemical shift at saturation; [L] and [P] are the total ligand and protein concentrations, respectively. The residues used for determining the binding constants were resonances that showed greater than average chemical shift

perturbations for all peptides tested. These are (Rtt103) L15, E16, D17, S18, E20, S21, I22, S24, L59, Y62, L63, N65, V67, V68, K72, Q74, R108, V109, N111, and I112. For Pcf11, they are D11, T20, R24, I26, T28, T31, Y65, A66, L67, D68, S69, M108, and F109.

Fluorescence Anisotropy: N-terminally fluorescently-tagged CTD peptides were ordered from AnaSpec (San Jose, CA) and desalted prior to use. Fluorescence anisotropy titration experiments using di-peptides were performed as competitive experiments in a buffer containing 25 mM HEPES pH 8.0, 100 mM NaCl, 1 mM EDTA, 1 mM DTE, and 2 μM N-terminal labelled 5,6-carboxyfluoresceine (FAM) Ser2P-CTD peptides, essentially as described elsewhere[12]. Concentrations used in the individual experiments are given in the figure legend. For the 4-repeat peptides, FA competition experiments were performed in a buffer containing 10 mM Tris-HCl pH 7.4, 120 mM NaCl (plus 2 mM DTT for Pcf11) as described above. The data were fit to both 2:1 and 1:1 binding models using non-linear least squares algorithm as implemented in the program DynaFit[13]. Model selection was done based on the Akaike Information Criterion.

***Yeast strains*:** YF884/Rtt103-TAP [MATa, ura3Δ0, leu2Δ0, his3Δ1, met15Δ0, RTT103-TAP::HIS3]; YSB2391 [MATa, ura3Δ0, leu2Δ0, his3Δ1, met15Δ0, rtt103(R108N)-TAP::HIS3]; YF406 [MATa, ura3-1, leu2-3,112, trp1-0, his3-

11,15, can1-100, ade2-1, PCF11-TAP::TRP1(KL), pFL38-PCF11]; YSB2389 [MATa, ura3-1, leu2-3,112, trp1-0, his3-11,15, can1-100, ade2-1, pcf11(N107R)-TAP::TRP1(KL), pFL38-PCF11(URA3)]; YSB2535 [MATa, ura3-1, leu2-3,112, trp1Δ, his3-11,15, ade2-1, pcf11Δ::TRP1, pRS315-Pcf11-3xHA(LEU2)]; YSB2536 [MATa, ura3-1, leu2-3,112, trp1Δ, his3-11,15, ade2-1, pcf11Δ::TRP1, pRS315-pcf11(D117A)-3xHA(LEU2)]; YSB2537 [MATa, ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2Δ202, rtt103Δ::HIS3, pRS315-RTT103-3xHA(LEU2)]; YSB2538 [MATa, ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2Δ202, rtt103Δ::HIS3, pRS315-rtt103(E115R)-3xHA(LEU2)].

***Plasmid construction*:** The genomic region spanning from -458 to +2566 of wild-type *PCF11* was subcloned into pBSII-KS(+) between the SalI and NotI sites to generate pBS-Pcf11. Similarly, *RTT103* from -402 to +1339 was inserted into the BamHI site of pBSII-KS(+) to generate pBS-Rtt103. Using these wild-type plasmids as template, inverse PCR-mediated site-directed mutagenesis was carried out to generate pBS-pcf11 (D117A) or pBS-rtt103 (E115R). For Pcf11, the region from -458 to +1878 (just upstream of the stop codon) was amplified by PCR from pBS-Pcf11 or pBS-pcf11 (D117A), and subcloned into pRS315-3xHA between NotI and BamHI sites. For Rtt103, the region from -402 to +1227 (just upstream of the stop codon) was amplified by PCR from pBS-Rtt103 or pBS-rtt103 (E115R), and subcloned into pRS315-3xHA between NotI and BamHI sites. The sequences of oligonucleotides are shown below.

Name	Sequence
Pcf11(D117A)-F	5'-GGTTGAATCCTAATGCCACCGGCCTGCCTCT-3'
Pcf11(D117A)-R	5'-AGAGGCAGGCCGGTGGCATTAGGATTCAACC-3'
Not1-Pcf11(-458)-F	5'-GACTGCGGCCGCTATACAGAACTTTCTCAGTC-3'
BamH1-Pcf11(no stop)-R	5'-GACTGGATCCTTTTGTGACCAATTTCTTTAAGTC-3'
RTT103(-402)-F	5'-CCTCGAGATCTCAAAGCAGTTTCTATAGGTC-3'
RTT103(+1339)-R	5'-CTTCGAGATCTATACGAAATAAACTATGGG-3'
Rtt103(E115R)-F	5'-GTGAATATACTAAAAAGAAGAAATATATTTTCC-3'
Rtt103(E115R)-R	5'-GGAAAATATATTTCTTCTTTTTAGTATATTCAC-3'
Not1-Rtt103(-402)-F	5'-GACTGCGGCCGCTCAAAGCAGTTTCTATAGGT-3'
Bgl2-Rtt103(no stop)-R	5'-GACTAGATCTATTTGCAAGCTTACTTAACAAGTCTTG-3'

Chromatin immunoprecipitations*:* ChIP assays of TAP-tagged Rtt103 and Pcf11 were performed using IgG sepharose (GE Healthcare) as previously described[14]. *PMA1* and *ADH1* primers used in PCR are as listed in Kim *et al*[15]. ChIP assays for 3xHA tagged Pcf11 or Rtt103 were carried out as previously described[14]. For Rpb3 ChIP assay, immunoprecipitation was carried out with 1Y26 (NeoClone) preincubated with protein G-sepharose (Amersham).

Supplementary References

1. Meinhart, A. & Cramer, P. Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature* 430, 223-6 (2004).
2. Delaglio, F. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6, 277-93 (1995).
3. Sattler, M., Schleucher, J. & Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in Nuclear Magnetic Resonance Spectroscopy* 34, 93-158 (1999).
4. Farrow, N.A. et al. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry* 33, 5984-6003 (1994).
5. Breeze, A.L. Isotope-filtered NMR methods for the study of biomolecular structure and interactions. *Progress in Nuclear Magnetic Resonance Spectroscopy* 36, 323-372 (2000).
6. Zwahlen, C. et al. Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: Application to a bacteriophage lambda N-peptide/boxB RNA complex. *Journal of the American Chemical Society* 119, 6711-6721 (1997).
7. Guntert, P. Automated NMR structure calculation with CYANA. *Methods Mol Biol* 278, 353-78 (2004).
8. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* 13, 289-302 (1999).
9. DeLano, W.L. The PyMOL Molecular Graphics System. in *DeLano Scientific LLC, San Carlos, CA, USA.*
10. Dominguez, C., Boelens, R. & Bonvin, A.M. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* 125, 1731-7 (2003).
11. Boyko, R., Sykes, B. xcrvfit: a graphical X-windows program for binding curve studies and NMR spectroscopic analysis. (University of Alberta, 1994).
12. Becker, R., Loll, B. & Meinhart, A. Snapshots of the RNA processing factor SCAF8 bound to different phosphorylated forms of the carboxyl-terminal domain of RNA polymerase II. *J Biol Chem* 283, 22659-69 (2008).
13. Kuzmic, P. Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. *Anal Biochem* 237, 260-73 (1996).
14. Kim, M., Ahn, S.H., Krogan, N.J., Greenblatt, J.F. & Buratowski, S. Transitions in RNA polymerase II elongation complexes at the 3' ends of genes. *EMBO J* 23, 354-64 (2004).
15. Kim, M. et al. Distinct pathways for snoRNA and mRNA termination. *Mol Cell* 24, 723-34 (2006).