



Structure and dynamics of the RNAPII CTDosome with Rtt103

Olga Jasnovidova^a, Tomas Klumpler^a, Karel Kubicek^a, Sergei Kalynych^a, Pavel Plevka^a, and Richard Steff^{a,1}

^aCEITEC—Central European Institute of Technology, Masaryk University, CZ-62500 Brno, Czech Republic

Edited by Juli Feigon, University of California, Los Angeles, CA, and approved September 12, 2017 (received for review July 13, 2017)

RNA polymerase II contains a long C-terminal domain (CTD) that regulates interactions at the site of transcription. The CTD architecture remains poorly understood due to its low sequence complexity, dynamic phosphorylation patterns, and structural variability. We used integrative structural biology to visualize the architecture of the CTD in complex with Rtt103, a 3'-end RNA-processing and transcription termination factor. Rtt103 forms homodimers via its long coiled-coil domain and associates densely on the repetitive sequence of the phosphorylated CTD via its N-terminal CTD-interacting domain. The CTD–Rtt103 association opens the compact random coil structure of the CTD, leading to a beads-on-a-string topology in which the long rod-shaped Rtt103 dimers define the topological and mobility restraints of the entire assembly. These findings underpin the importance of the structural plasticity of the CTD, which is templated by a particular set of CTD-binding proteins.

RNA polymerase II | CTD | structural biology | transcription | Rtt103

The C-terminal domain (CTD) of the largest subunit of RNA polymerase II (RNAPII) consists of multiple tandem repeats (26 in yeast, 52 in humans) of the heptapeptide consensus Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7, which is highly conserved from yeast to human (1–3). The CTD serves as a binding platform for many RNA/protein-binding factors involved in the regulation of the transcription cycle (1, 3). Yeast are inviable if the CTD is trimmed to less than 11 repeats of the heptapeptide consensus (4) or if the periodicity of two repeats is perturbed (5), suggesting the importance of both the CTD length and its repetitiveness.

The CTD interaction network is regulated by posttranslational modifications of the CTD, which yield specific phosphorylation and subsequent factor-binding patterns in coordination with the transcription cycle (the “CTD code”) (1, 6–11). Phosphorylations at Y₁, S₂, T₄, S₅, and S₇ are the most common and well-studied posttranslational modifications of the CTD (12). Mass spectrometry studies of the CTD showed that the CTD heptads are homogeneously phosphorylated along the entire length of the domain in proliferating yeast and human cells (13, 14). Major phosphorylation sites are S₂ and S₅, whereas Y₁, T₄, and S₇ are minor phosphorylation sites (13, 14), but all sites are important for transcription regulation and proper functioning of the cell. On average, each CTD heptad is phosphorylated once and the occurrence of two phosphorylations per repeat is a rare event (13, 14). The coimmunoprecipitation of specific CTD phosphoisoforms revealed distinct functional sets of factors (CTD-interactome) related to each CTD phosphoisoform (15).

The CTD has no well-defined 3D structure and, therefore, is not observed in the crystal structures of RNAPII (16–19) and forms fuzzy densities on electron microscopy images (20, 21). Nevertheless, the first structural information of the unbound CTD has recently been reported in the fruit fly (22, 23), where it was shown that the CTD forms a compact random coil and that its phosphorylation induces a modest extension and stiffening of the CTD (22, 23).

Current structural knowledge of interactions between the CTD and its recognition factors is based on short peptides

mimicking the CTD bound to given CTD binding factors (1, 19, 24). However, the atomic-level structural architecture of the full-length CTD modulated by associated factors remains unknown. Several studies attempted to propose a structural model for the full-length CTD. For example, in the complex of the CTD peptide with the CTD-interacting domain (CID) of Pcf11, a subunit of cleavage factor IA (25), the CTD heptad was found to adopt a β -turn conformation (26). Therefore, a compact, left-handed, β -spiral model of the CTD was proposed (26). A β -spiral conformation would allow the CTD chain with a length of 100 Å to fold into a compact structure, which corresponds to the observed densities in low-resolution electron microscopy images of RNAPII (20). The heterodimer composed of the human proteins RPRD1A and RPRD1B was found to bind the CTD, thereby stimulating the recruitment and phosphatase activity of RPAP2 (pS₅-CTD-phosphatase) (27). These findings led to the proposal of a model in which the CTD and accessory molecules form a high-order arrangement dubbed the “CTDosome” (27).

To probe the CTDosome architecture experimentally, we set out to apply integrative structural biology methods and investigate how the termination factor Rtt103 decorates the sequence of the CTD. First, we independently solved high-resolution structures of stable subunits by solution NMR spectroscopy (NMR) and X-ray crystallography. Then, we corroborated the obtained structural information with small-angle X-ray scattering (SAXS) data to reconstruct the overall architecture of the Rtt103–CTD complex. We show that Rtt103 contains a coiled-coil domain that mediates Rtt103 dimerization and uses its

Significance

RNA polymerase II (RNAPII) not only transcribes protein coding genes and many noncoding RNA, but also coordinates transcription and RNA processing. This coordination is mediated by a long C-terminal domain (CTD) of the largest RNAPII subunit, which serves as a binding platform for many RNA/protein-binding factors involved in transcription regulation. In this work, we used a hybrid approach to visualize the architecture of the full-length CTD in complex with the transcription termination factor Rtt103. Specifically, we first solved the structures of the isolated subcomplexes at high resolution and then arranged them into the overall envelopes determined at low resolution by small-angle X-ray scattering. The reconstructed overall architecture of the Rtt103–CTD complex reveals how Rtt103 decorates the CTD platform.

Author contributions: O.J. and R.S. designed research; O.J. and K.K. performed research; O.J., T.K., K.K., S.K., P.P., and R.S. analyzed data; and O.J. and R.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This is an open access article distributed under the [PNAS license](#).

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.wwpdb.org (PDB ID codes 5M48 and 5M9D); SAXS data are deposited in Small Angle Scattering Biological Data Bank (SASBDB ID code SASDCZ2).

¹To whom correspondence should be addressed. Email: richard.steff@ceitec.muni.cz.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1712450114/-DCSupplemental.

N-terminal CID to read adjacent repetitive phosphorylation marks on the CTD independently of one other. Our reconstruction demonstrates how Rtt103 explores the repetitiveness and length of the CTD sequence while keeping the entire arrangement partially flexible.

Results

Limited Proteolysis of Rtt103 Reveals a Coiled-Coil Domain That Mediates Dimerization. In our divide-and-conquer approach, we first identified the overall domain organization of Rtt103. Trypsin digestion of the full-length Rtt103 coupled with mass spectrometry

revealed that the protein fragment harboring amino acid residues 1–246 (Rtt103_{1–246}) is protected from proteolytic cleavage (Fig. 1*A* and Fig. S1). The remaining C-terminal part of Rtt103 (amino acid residues 247–409) was efficiently digested by trypsin, suggesting the absence of additional structured domains (Fig. 1*A* and Fig. S1). Subsequent biochemical characterization of the identified stable constructs revealed that Rtt103_{141–246} and Rtt103_{1–246} form homodimers (Fig. S2*A* and *B*). Subsequent crystallization screens of the Rtt103_{141–246} and Rtt103_{1–246} constructs showed that only the Rtt103_{141–246} construct formed well-diffracting crystals. The structure of Rtt103_{141–246} was determined to a resolution of 2.6 Å

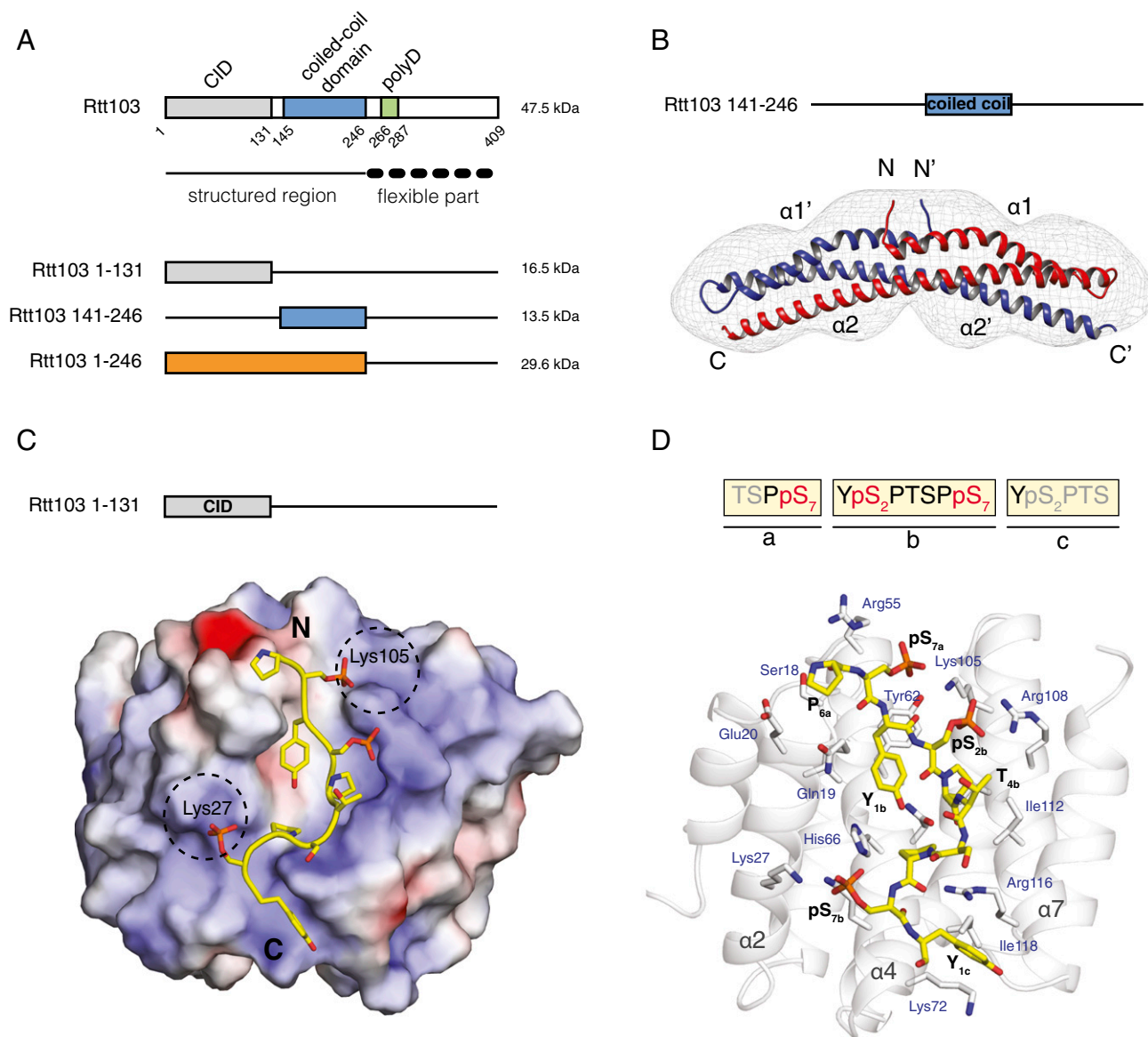


Fig. 1. Dimerization and RNAPII CTD recognition by Rtt103. (A) Scheme of Rtt103 domain organization (Upper). The numbers below the scheme represent borders of the amino acid segments. Structured and flexible regions were determined based on the limited proteolysis study (Fig. S1). The recombinant protein constructs used in the study, along with their respective molecular masses, are shown (Lower). CID, CTD-interacting domain; polyD, polyaspartate stretch. (B) Crystal structure of the Rtt103_{141–246} coiled-coil domain shown superimposed with an ab initio model (gray mesh) derived using DAMMIN (40) from SAXS scattering data. The two different polypeptide chains of the coiled-coil dimer are shown in red and blue; their respective N- and C-termini, as well as α -helices, are indicated. (C) Electrostatic surface representation of the Rtt103 CID (electropositive in blue, electronegative in red, neutral in white) in complex with the pS₂pS₇-CTD peptide (yellow sticks; PDB ID code: 5M9D). The N- and C-termini of the peptide are indicated. Dashed black circles indicate electro-positive areas that accommodate pS₇ residues. (D) Detailed view of the Rtt103 CID (gray cartoons) bound to the pS₂pS₇-CTD peptide (yellow sticks). Highlighted Rtt103 CID residues (gray sticks, blue labels) form hydrophobic contacts and putative hydrogen bonds with the pS₂pS₇-CTD peptide (yellow sticks, black labels). The sequence of the peptide used for structure determination is indicated above; residues that showed interaction with the Rtt103 CID are shown in black and red.

(Tables S1 and S2). We found that each Rtt103_{141–246} subunit consists of two α -helices, namely the α 1-helix (Gln146–Glu177) with a small bend in the middle and a long α 2-helix of \sim 105 Å in length (Val184–Asp246) (Fig. 1B). In the crystal, two protein chains form a dimer where the α 2-helices are arranged in an antiparallel fashion (Fig. 1B). This architecture of the dimer is in agreement with findings from gel filtration experiments (Fig. S24) and SAXS data (Fig. S2B). Importantly, the central region of the α 2-helix (Lys200–Ile238) contains a coiled-coil signature, which is arranged in trans in the antiparallel dimer assembly of the two Rtt103 molecules. The coiled-coil domain contains a characteristic knobs-into-holes packing with mixed “a” and “d” layers (Fig. S2C), with an average pitch of 172 Å (defined by CCCP; ref. 28). The dimer structure is also stabilized by multiple intermolecular (Asp149–Lys152, Asp153–Lys216, Lys168–Asp172, Asp223–Arg226) and intramolecular (Lys200–Glu239, Glu231–Arg210–Glu224) salt bridges. Altogether, the key findings regarding Rtt103 architecture are as follows: (i) Rtt103 contains a coiled-coil domain that follows the CID and (ii) the C-terminal half of the Rtt103 is disordered.

The Rtt103 CID Binds the Extended pS₂-pS₇-CTD Peptide. The structure of Rtt103_{1–131} (or CTD-interacting domain; CID) bound to a short Ser2-phosphorylated CTD moiety has previously been reported (29). Here, we used NMR to determine the structure of Rtt103_{1–131} bound to a longer CTD substrate (Ser2/7-phosphorylated; Tables S3 and S4), which revealed that the recognition interface of the CID is, in fact, larger than previously reported (29). Our NMR structure of Rtt103_{1–131} bound to the extended TSPpS₇ YpS₂PTSPpS₇ YpS₂PTS peptide (termed pS₂pS₇-CTD) confirmed the previously reported observation regarding the recognition of the upstream region of pS₂pS₇-CTD, and further revealed information regarding the recognition of the downstream region of pS₂pS₇-CTD (Fig. 1C and D). The structure of Rtt103 CID is formed by eight α -helices in a right-handed superhelical arrangement. The NMR data show that the pS₂pS₇-CTD peptide binds at the conserved surface formed by helices α 2, α 4, and α 7 of the Rtt103 CID (Fig. 1D and Fig. S3). NMR revealed intermolecular contacts between Rtt103_{1–131} and the residues P_{6a}, pS_{7a}, Y_{1b}, pS_{2b}, P_{3b}, T_{4b}, S_{5b}, and Y_{1c} of the pS₂pS₇-CTD peptide. Specifically, P_{6a} lies in the proximity of the hydrophobic area formed by the N-terminal tip of the α 2-helix, being involved in multiple intermolecular contacts with Ser18, Gln19, and Glu20. Residue Y_{1b} is also docked into a hydrophobic pocket (Ile22, Tyr62) and stabilized by a hydrogen bond between its hydroxyl group and the side-chain amide group of Asn65. Residue P_{3b} forms hydrophobic interactions with Val109 and Ile112. Residues pS_{2b}P_{3b}T_{4b}S_{5b} form a β -turn stabilized by hydrogen bonds between the pS_{2b} carbonyl and the S_{5b} amide, between the pS_{2b} γ -oxygen and T_{4b} amide, and between the pS_{2b} phosphate and T_{4b} hydroxyl. Perturbation of the above-described hydrophobic pocket (not affecting the structural integrity; Fig. S4C and refs. 27 and 30) caused a drop of 30- to 50-fold in the affinity between pS₂pS₇-CTD and Rtt103_{1–131} ($K_D = 33 \pm 1.2 \mu\text{M}$ for Ile112Ala, $K_D = 80 \pm 11 \mu\text{M}$ for Ile112Gly) (Fig. S4). In agreement with previous structural observations (29), we noted that the phosphorylation of S_{2b} is recognized by the side chain of Arg108. Interestingly, we observed multiple close contacts between Y_{1c} and the C-terminal parts of helices α 4 and α 7. The positioning of Y_{1c} near the tip of helices α 4 and α 7 induces a second sharp turn in the pS₂pS₇-CTD peptide. The side chain of Y_{1c} forms a broad range of hydrophobic contacts with Lys72, Gly73, and Ile118, whereas the guanidinium group of Arg116 coordinates the backbone of pS₂pS₇-CTD. We found that charge-swapping mutations at the interacting sites of Rtt103 (not affecting the structural integrity; Fig. S4C) resulted in pronounced affinity decrease between Rtt103 and pS₂pS₇-CTD ($K_D = 9.7 \pm 0.7$, 51 ± 2.2 , and $65 \pm 8.9 \mu\text{M}$ for Lys72Glu, Arg116Glu, and Lys72Glu/Arg116Glu, respectively), highlighting the importance

of the CTD backbone interactions with Arg116. A large area of the Rtt103 CID surface is positively charged and enriched in residues that could stabilize interaction with negatively charged sites of the phosphorylated CTD peptide (Fig. 1C). Although our data did not indicate the presence of intermolecular contacts for the pS₇ residues, the positions of these residues are indirectly defined by the nuclear Overhauser effects from the neighboring residues. Therefore, residues pS₇ are likely involved in charge–charge interactions with Lys27 and Lys105 (which is part of the poly-Lys tract Lys103–Lys104–Lys105) (Fig. 1C). We found that the Lys27Glu mutant (perturbation of one of the pS₇ binding pockets; Fig. 1C) showed lower binding only for the pS₇-containing peptide (K_D of $13.2 \pm 0.3 \mu\text{M}$ and $28.5 \pm 1 \mu\text{M}$ for wild type and Lys27Glu, respectively) but not for the pS₂-containing peptide (K_D of $1.6 \pm 0.07 \mu\text{M}$ and $2 \pm 0.8 \mu\text{M}$ for wild type and Lys27Glu, respectively). Altogether, the key finding is that the Rtt103 CID interacts with pS₂pS₇-CTD via a larger area than previously reported (29), specifically recognizing the downstream region of the CTD peptide. Our structure reveals that P_{6a}pS_{7a}Y_{1b}pS_{2b}P_{3b}T_{4b}S_{5b}P_{6b}pS_{7b}Y_{1c} is the minimal CTD-binding moiety recognized by Rtt103.

Two CIDs Tethered by a Coiled-Coil Domain Tumble Independently. As a result of the antiparallel arrangement of the coiled coils, the Rtt103 CIDs are attached by a linker of 15 amino acids to the middle region of the coiled-coil domain. NMR investigations of the Rtt103_{1–246} and CID constructs showed that the CID structure is not influenced by the presence of the coiled-coil domain and that the CIDs are likely to tumble independently (Fig. S5). To visualize the arrangement of the Rtt103 CIDs relative to the coiled-coil domains, we analyzed SAXS scattering data of purified Rtt103_{1–246} using available atomic structures (PDB ID codes: 2KM4, 5M48) by ensemble-optimization method (EOM 2.0) (31). This approach enables deconvolving the conformational averaging into the contribution of individual conformers. The obtained models suggest that the coiled-coil domain defines the length of the protein (\sim 105 Å) (Fig. 2), and the linker allows the CIDs to reach all across the 105-Å-long coiled-coil domain. Thus, the CIDs could be positioned relatively close to each other but are also able to sample a large surrounding space to recognize the substrate (Movie S1). Next, we tested whether the coiled-coil-mediated dimerization of Rtt103 affects the binding to the CTD using fluorescence anisotropy (FA). We measured the binding affinity for the minimal CTD-binding moiety (SPS YpSPTSPpS YS) and a long CTD substrate (harboring two minimal CTD-binding moieties connected with a spacer; SPS YpSPTSPpS YSPTSPS YpSPTSPpS YS) with Rtt103_{1–246} and with the CID. We found that Rtt103_{1–246} binds to the minimal CTD-binding moiety with a K_D of $3.3 \pm 0.06 \mu\text{M}$, and to the long CTD substrate with a K_D of $0.3 \pm 0.01 \mu\text{M}$. The isolated CID binds to the minimal CTD-binding moiety with a K_D of $1.65 \pm 0.13 \mu\text{M}$, and to the long CTD substrate with a K_D of $0.5 \pm 0.01 \mu\text{M}$. This suggests that the dimerization increases the local concentration of CIDs that are available for the CTD binding. Altogether, the key finding is that dimerization of the Rtt103 coiled-coil domains does not promote formation of a rigid structure between the Rtt103 CIDs, but in fact helps the CIDs sample multiple conformations restricted only by their tethering to the flexible linker.

The Rtt103 Coiled-Coil Domain Restricts the Variability of the CTD–CIDs Assembly. Next, we asked whether the coiled-coil-mediated dimerization of Rtt103 affects the overall fashion in which the repetitive CTD sequence is decorated with Rtt103 CIDs. The complex between Rtt103_{1–246} and pS₂E-CTD {pSer₂-CTD mimic [SPEFTCEPTSPS-(YEPTSPS)₁₃-YEPAAADYKDDDDK]; Fig. S6} was prepared by mixing individual proteins with molar excess of Rtt103_{1–246}, followed by size-exclusion chromatography and SAXS data collection. The estimation of the molecular mass

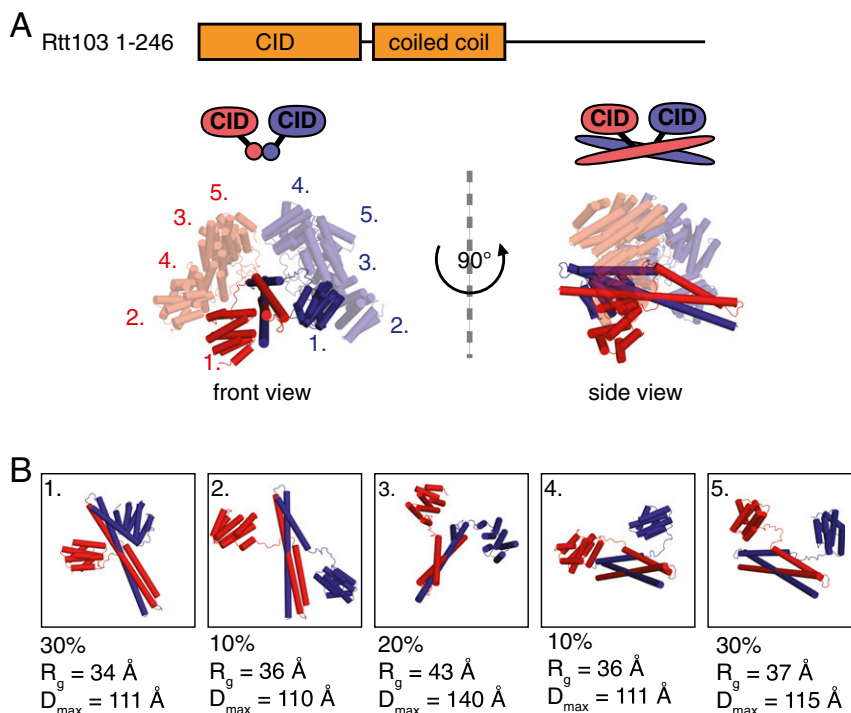


Fig. 2. The two Rtt103 CIDs are tethered by a coiled-coil domain but tumble independently. (A) Overlay of individual conformations from the ensemble of free Rtt103₁₋₂₄₆ structures derived using the ensemble-optimization method (EOM 2.0) (31) ($\chi^2 = 1.001$); front and side views are provided. Conformations are superimposed based on structure of the coiled-coil domain. Conformations 2–5 are shown with 60% opacity. The two different polypeptide chains are shown in red and blue. (B) Individual conformations from the Rtt103₁₋₂₄₆ ensemble derived by EOM 2.0; the fraction (%), radius of gyration (R_g), and maximum intraparticle distance (D_{max}) are indicated for each conformation.

(MM) of the complex was done using DAMMIF (32), which yielded a MM of 200 ± 5 kDa, that is close to the theoretical MM of the Rtt103₁₋₂₄₆:CTD complex with a ratio of 6:1 (190 kDa). In terms of molecular architecture, it suggests that three Rtt103 dimers bind to a 13-repeat-long CTD upon saturation. The interpretation of the scattering data was performed using the CORAL software (33). The structures of the Rtt103 CID (PDB ID code: 5M9D) and coiled-coil domain (PDB ID code: 5M48) were combined together with the distance constraints between the CTD and Rtt103 CID and fitted against the experimental SAXS data. The calculation was repeated 10 times for each interaction scenario with a ratio of 6:1 for Rtt103₁₋₂₄₆:CTD, which provided the best fit to the experimental data (Fig. 3 and Fig. S7). All reconstituted complexes displayed a similar elongated architecture characterized by a D_{max} value of 180–250 Å (Fig. S7B). One Rtt103 dimer is accommodated on four CTD repeats (**PS YEPTSPS YEPTSPS YEPTSPS YEPTS**; CID recognition sites are shown in bold). In this architecture, the coiled-coil domains surround the individual Rtt103 CIDs accommodated on the CTD (Fig. 3). The shielding provided by the coiled coils restricts the flexibility of CIDs on the CTD to some extent, but promotes the stretching of the compact random coil structure of the CTD (22, 23). Interestingly, we obtained a similar quality fit to the experimental data without including the constraint that two CIDs of the dimer must bind neighboring CTD epitopes (Fig. S7). The obtained models contain CIDs accommodated in different areas of the CTD, supporting the hypothesis of residual flexibility in the core of the CTDsome shielded from the outside by coiled coils. Altogether, the key finding is that the CTDsome architecture is dynamic and allows for optimal recognition of available phosphorylation signals in the CTD. This variability is essential as the CTD contains some poorly conserved heptads (2) whose recognition is promoted by dimerization in which the

coiled-coil domains prime the sampling of the CTD epitopes (Fig. 3 and Fig. S7).

Discussion

Assembly and reassembly of the CTDsome during transcription by RNAPII is important for regulation of transcription and RNA processing. However, due to the structural complexity and dynamics of the CTDsome, the mechanistic aspects of this process remain poorly understood. Here, we report an experimentally based structural model of the CTDsome, which has been derived using a combination of X-ray, NMR, and SAXS data (Fig. 3).

First, we determined that Rtt103 is capable of dimerizing in free form via a coiled-coil domain. Several CID-containing proteins are known to have multimerization regions such as the Nrd1-Nab3 heterodimerization region (34), coiled-coil region in Pcf11 (35), and coiled-coil regions in RPRD1A, RPRD1B, and RPRD2 (27, 36). The RPRD1A-RPRD1B heterodimer binds to multiple pS₂-CTD repeats and exposes the pS₅ sites on the CTD, which stimulates the activity of RPAP2 pS₅-CTD phosphatase. It is likely that the Rtt103 scaffold is used to recruit other factors or enzymes (e.g., Rat1-Rai1) that act on the CTD. In contrast to RPRD1A and RPRD1B, which include only the CID and dimerization domain, Rtt103 has a long unstructured C-terminal part that occupies half of the protein and, therefore, could greatly impact multisubunit architectures and interactions with other RNA/protein-binding factors.

Here, we determined that the Rtt103 CID binds across three CTD heptads and that the minimal CTD binding moiety consists of the P_{6a}S_{7a}Y_{1b}P_{2b}P_{3b}T_{4b}S_{5b}P_{6b}S_{7b}Y_{1c} sequence. These findings indicate that the Rtt103 CID binds a longer CTD stretch than previously reported (29). Accommodation of the core P_{6a}S_{7a}Y_{1b}P_{2b}P_{3b}T_{4b}S_{5b} stretch of the CTD in the CID binding pocket is highly conserved among CID-CTD peptide complexes (26, 27, 37, 38). In contrast, the conformation of the upstream

SAXS Core Facility, and the Proteomics Core Facility of CEITEC, supported by the project CZ.1.05/1.1.00/02.0068, financed from the European Regional Development Fund. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme Grant Agreement 649030 (to R.S.). This work was supported by the Czech Science Foundation Grants 13-18344S (to R.S.)

and 15-24117S (to K.K.). The research leading to these results received funding from ERC under the European Union's Seventh Framework Program Grant FP/2007-2013/ERC Grant Agreement 355855' and from EMBO installation Grant 3041 (to P.P.). This publication reflects only the author's view and the Research Executive Agency is not responsible for any use that may be made of the information it contains.

- Eick D, Geyer M (2013) The RNA polymerase II carboxy-terminal domain (CTD) code. *Chem Rev* 113:8456–8490.
- Chapman RD, Heidemann M, Hintermair C, Eick D (2008) Molecular evolution of the RNA polymerase II CTD. *Trends Genet* 24:289–296.
- Harlen KM, Churchman LS (2017) The code and beyond: Transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nat Rev Mol Cell Biol* 18:263–273.
- West ML, Corden JL (1995) Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. *Genetics* 140:1223–1233.
- Liu P, Kenney JM, Stillier JW, Greenleaf AL (2010) Genetic organization, length conservation, and evolution of RNA polymerase II carboxyl-terminal domain. *Mol Biol Evol* 27:2628–2641.
- Buratowski S (2003) The CTD code. *Nat Struct Biol* 10:679–680.
- Mayer A, et al. (2010) Uniform transitions of the general RNA polymerase II transcription complex. *Nat Struct Mol Biol* 17:1272–1278.
- Mayer A, et al. (2012) CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science* 336:1723–1725.
- Bataille AR, et al. (2012) A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. *Mol Cell* 45:158–170.
- Kim H, et al. (2010) Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nat Struct Mol Biol* 17:1279–1286.
- Tietjen JR, et al. (2010) Chemical-genomic dissection of the CTD code. *Nat Struct Mol Biol* 17:1154–1161.
- Heidemann M, Hintermair C, Voß K, Eick D (2013) Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochim Biophys Acta* 1829:55–62.
- Suh H, et al. (2016) Direct analysis of phosphorylation sites on the Rpb1 C-terminal domain of RNA polymerase II. *Mol Cell* 61:297–304.
- Schüller R, et al. (2016) Heptad-specific phosphorylation of RNA polymerase II CTD. *Mol Cell* 61:305–314.
- Harlen KM, et al. (2016) Comprehensive RNA polymerase II interactomes reveal distinct and varied roles for each phospho-CTD residue. *Cell Rep* 15:2147–2158.
- Cramer P, et al. (2000) Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* 288:640–649.
- Cramer P, Bushnell DA, Kornberg RD (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* 292:1863–1876.
- Spähr H, Calero G, Bushnell DA, Kornberg RD (2009) Schizosaccharomyces pombe RNA polymerase II at 3.6-Å resolution. *Proc Natl Acad Sci USA* 106:9185–9190.
- Meinhart A, Kamenski T, Hoepfner S, Baumli S, Cramer P (2005) A structural perspective of CTD function. *Genes Dev* 19:1401–1415.
- Meredith GD, et al. (1996) The C-terminal domain revealed in the structure of RNA polymerase II. *J Mol Biol* 258:413–419.
- Tsai K-L, et al. (2013) A conserved Mediator-CDK8 kinase module association regulates Mediator-RNA polymerase II interaction. *Nat Struct Mol Biol* 20:611–619.
- Portz B, et al. (2017) Structural heterogeneity in the intrinsically disordered RNA polymerase II C-terminal domain. *Nat Commun* 8:15231.
- Gibbs EB, et al. (2017) Phosphorylation induces sequence-specific conformational switches in the RNA polymerase II C-terminal domain. *Nat Commun* 8:15233.
- Jasnovidova O, Stefl R (2013) The CTD code of RNA polymerase II: A structural view. *Wiley Interdiscip Rev RNA* 4:1–16.
- Barillà D, Lee BA, Proudfoot NJ (2001) Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in Saccharomyces cerevisiae. *Proc Natl Acad Sci USA* 98:445–450.
- Meinhart A, Cramer P (2004) Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature* 430:223–226.
- Ni Z, et al. (2014) RPRD1A and RPRD1B are human RNA polymerase II C-terminal domain scaffolds for Ser5 dephosphorylation. *Nat Struct Mol Biol* 21:686–695.
- Grigoryan G, Degrado WF (2011) Probing designability via a generalized model of helical bundle geometry. *J Mol Biol* 405:1079–1100.
- Lunde BM, et al. (2010) Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nat Struct Mol Biol* 17:1195–1201.
- Jasnovidova O, Krejčíková M, Kubicek K, Stefl R (2017) Structural insight into recognition of phosphorylated threonine-4 of RNA polymerase II C-terminal domain by Rtt103p. *EMBO Rep* 18:906–913.
- Tria G, Mertens HDT, Kachala M, Svergun DI (2015) Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCr* 2:207–217.
- Franke D, Svergun DI (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J Appl Crystallogr* 42:342–346.
- Petoukhov MV, et al. (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J Appl Crystallogr* 45:342–350.
- Vasiljeva L, Kim M, Mutschler H, Buratowski S, Meinhart A (2008) The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nat Struct Mol Biol* 15:795–804.
- Xu X, Pérébaskine N, Minvielle-Sébastien L, Fribourg S, Mackereth CD (2015) Chemical shift assignments of a new folded domain from yeast Pcf11. *Biomol NMR Assign* 9:421–425.
- Mei K, et al. (2014) Structural basis for the recognition of RNA polymerase II C-terminal domain by CREPT and p15RS. *Sci China Life Sci* 57:97–106.
- Becker R, Loll B, Meinhart A (2008) Snapshots of the RNA processing factor SCAF8 bound to different phosphorylated forms of the carboxyl-terminal domain of RNA polymerase II. *J Biol Chem* 283:22659–22669.
- Kubicek K, et al. (2012) Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. *Genes Dev* 26:1891–1896.
- Noble CG, et al. (2005) Key features of the interaction between Pcf11 CID and RNA polymerase II CTD. *Nat Struct Mol Biol* 12:144–151.
- Svergun DI (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* 76:2879–2886.
- Mossessova E, Lima CD (2000) Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol Cell* 5:865–876.
- Svergun D, Barberato C, Koch MHJ (1995) CRYSOLOG-A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr* 28:768–773.
- Kozin MB, Svergun DI (2001) Automated matching of high- and low-resolution structural models. *J Appl Crystallogr* 34:33–41.
- Kuzmic P (2009) DynaFit-A software package for enzymology. *Methods Enzymol* 467:247–280.
- Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* 66:125–132.
- Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* 62:72–82.
- Winn MD, et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67:235–242.
- Adams PD, et al. (2011) The Phenix software for automated determination of macromolecular structures. *Methods* 55:94–106.
- McCoy AJ, Storoni LC, Read RJ (2004) Simple algorithm for a maximum-likelihood SAD function. *Acta Crystallogr D Biol Crystallogr* 60:1220–1228.
- Grosse-Kunstleve RW, Adams PD (2003) Substructure search procedures for macromolecular structures. *Acta Crystallogr D Biol Crystallogr* 59:1966–1973.
- Terwilliger T (2004) SOLVE and RESOLVE: Automated structure solution, density modification and model building. *J Synchrotron Radiat* 11:49–52.
- Terwilliger TC, et al. (2009) Decision-making in structure solution using Bayesian estimates of map quality: The PHENIX AutoSol wizard. *Acta Crystallogr D Biol Crystallogr* 65:582–601.
- Cowtan K (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* 62:1002–1011.
- Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126–2132.
- McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Crystallogr* 40:658–674.
- Afonine PV, et al. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* 68:352–367.
- Pettersen EF, et al. (2004) UCSF Chimera-A visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612.
- Walshaw J, Woolfson DN (2001) Socket: A program for identifying and analysing coiled-coil motifs within protein structures. *J Mol Biol* 307:1427–1450.
- Larkin MA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948.

Supporting Information

Jasnovidova et al. 10.1073/pnas.1712450114

SI Methods

Cloning and Purification of Phospho-Serine Mimetic. Template for phospho-serine mimetic was synthesized by GeneArt, Thermo Fisher Scientific. Sequence was amplified and cloned into pET28b-SMT3 (41) using BamHI and HindIII restriction sites with the addition of C-terminal FLAG-tag [protein sequence after 6xHIS-SUMO cleavage: SPEFTCEPTSPS-(YEPTSPS)₁₃-YEPAAADYKDDDDK]. Resulted construct was verified by DNA sequencing and transformed into *E. coli* BL21-Codon Plus (DE3)-RIPL cells (Stratagene). For protein expression, bacteria was grown in M9 medium with 50 mg/L of kanamycin at 37 °C until OD₆₀₀ ~ 0.6, cooled down to 30 °C, induced with 0.1 mM IPTG and protein was overexpressed at 30 °C overnight. Cells were harvested by centrifugation and resuspended in the denaturing buffer (8 M urea, 50 mM Tris, 500 mM NaCl, 10 mM BME, 5 mM imidazole, 0.1% Nonidet P-40, pH_{20 °C} = 8). The lysate was rigorously stirred at 4 °C for 1 h, cleared then by centrifugation (50,000 × *g* for 1 h). Soluble fraction was loaded on Ni-NTA beads (Qiagen), the bound protein was refolded on the column with the refolding buffer (50 mM Tris, 500 mM NaCl, 10 mM BME, 5 mM imidazole, 0.1% Nonidet P-40, pH_{4 °C} = 8), and eluted with refolding buffer containing 300 mM imidazole. 6xHIS-SUMO-tag was cleaved overnight by Ulp1 protease during the dialysis into the refolding buffer. Cleaved fraction of 6xHIS-SUMO-tag was bound to Ni-NTA beads (Qiagen), flow-through containing was concentrated using Vivaspin 20 (Sartorius) concentrator with 3,000 Da cutoff, loaded on Superdex 75 10/600 (GE Healthcare), and eluted with 50 mM Tris, 500 mM NaCl, 10 mM BME, pH_{20 °C} = 8 (room temperature).

Cloning and Purification of Rtt103. The pET28b-Rtt103-CID S18A, Q19A, I112A, and I112G plasmids were gifts from T. Kabzinski, CEITEC, Brno, Czech Republic. Rtt103-CID-6xHIS point mutants (K72E, K72E/R116E, R108E, R116E) were obtained by QuikChange site-directed mutagenesis kit (Stratagene). Resulting constructs were verified by DNA sequencing, and then transformed into *E. coli* BL21-Codon Plus (DE3)-RIPL cells (Stratagene). Rtt103-CID-6xHIS (gift from A. Meinhart, Max Planck Institute for Medical Research, Heidelberg) and point mutants were expressed and purified as previously described (29).

pET21b-Rtt103-1-246 was amplified from pET21b-Rtt103 (gift from C. L. Moore, Tufts University School of Medicine, Boston) by removing residues 247-409. Rtt103₁₄₁₋₂₄₆ was amplified from pET21b-Rtt103 and cloned into pET22b using NdeI and XhoI restriction sites. Resulting constructs containing C-terminal 6xHIS-tag were verified by DNA sequencing and then transformed into *E. coli* BL21-Codon Plus (DE3)-RIPL cells (Stratagene).

For the expression of Rtt103₁₋₂₄₆, bacterial culture was grown in M9 medium with 50 mg/L of ampicillin at 37 °C until OD₆₀₀ ~ 0.3, cooled down to 20 °C, induced with 0.5 mM IPTG, and protein was overexpressed at 20 °C overnight. Cells were harvested by centrifugation and resuspended in lysis buffer [50 mM KH₂PO₄, 500 mM KCl, 0.5 mM EDTA, 10 mM BME, 40 mM imidazole, 10% glycerol, pH_{20 °C} = 7 supplemented with protease inhibitor mixture (cOmplete; Roche)]. After disruption of cells, the lysate was cleared by centrifugation (50,000 × *g* for 1 h) and soluble fraction was loaded on HisTrap FF Crude (GE Healthcare). The column was washed by lysis buffer containing 75 mM imidazole, protein eluted by lysis buffer containing 500 mM imidazole. Buffer was exchanged to the ion-exchange buffer (25 mM KH₂PO₄, 50 mM KCl, 1 mM BME pH_{20 °C} = 6.5) using HiPrep26/10 desalting (GE Healthcare). Protein was loaded on HiTrap SP HP column (GE Healthcare) equilibrated with the ion-exchange buffer

and eluted with 50 mM to 1 M KCl gradient. Eluted protein was further purified on HiLoad Superdex 200 (GE Healthcare) in 25 mM KH₂PO₄, 300 mM KCl, 10 mM BME pH_{20 °C} = 6.5 buffer. Protein sample was concentrated using Vivaspin 20 (Sartorius) concentrator with 10,000 Da cutoff.

For expression of Rtt103₁₄₁₋₂₄₆, bacterial culture was grown in M9 medium with 50 mg/L ampicillin at 37 °C until OD₆₀₀ ~ 0.3, cooled down to 25 °C, induced with 0.5 mM IPTG, and protein was overexpressed at 25 °C overnight. Cells were harvested by centrifugation and resuspended in the lysis buffer [50 mM Tris, 500 mM NaCl, 10 mM BME, 40 mM imidazole, 10% glycerol, pH_{20 °C} = 8 supplemented with protease inhibitor mixture (cOmplete; Roche)]. After disruption of cells, the lysate was cleared by centrifugation (50,000 × *g* for 1 h) and soluble fraction was loaded on Ni-NTA column (Qiagen), washed with lysis buffer containing 1 M NaCl and 75 mM imidazole, protein was eluted using lysis buffer containing 500 mM imidazole. Eluted protein was loaded on HiLoad 16/600 Superdex 75 (GE Healthcare) and eluted with 50 mM Tris, 500 mM NaCl, 5% glycerol, 10 mM BME, pH_{20 °C} = 8. Protein was dialyzed to 25 mM Tris, 200 mM NaCl, 1 mM BME, pH_{4 °C} = 8.0 and concentrated using Vivaspin 20 (Sartorius) concentrator with 3,000 Da cutoff to 6 mg/mL for crystallization. Protein was crystallized in 3.75 M sodium formate at 20 °C. For crystallization, the protein was labeled with selenomethionine by feedback inhibition of the methionine biosynthesis pathway in M9 media.

pET21b-Rtt103 was transformed into *E. coli* BL21-Codon Plus (DE3)-RIPL cells (Stratagene). For the expression of Rtt103, bacterial culture was grown in LB medium with 50 mg/L ampicillin at 37 °C until OD₆₀₀ ~ 0.3, cooled down to 16 °C, induced with 0.5 mM IPTG, and protein was overexpressed at 16 °C overnight. Cells were harvested by centrifugation and resuspended in the lysis buffer [50 mM Tris, 500 mM NaCl, 10 mM BME, 20 mM imidazole, pH_{20 °C} = 8 supplemented with protease inhibitor mixture (cOmplete; Roche)]. After disruption of cells, the lysate was cleared by centrifugation (50,000 × *g* for 1 h) and soluble fraction was loaded on Ni-NTA column (Qiagen), washed with lysis buffer containing 1 M NaCl, protein was eluted using lysis buffer containing 500 mM imidazole. Eluted protein was loaded on HiLoad 16/600 Superdex 200 (GE Healthcare) and eluted with 50 mM Tris, 500 mM NaCl, 10% glycerol, 10 mM BME, pH_{20 °C} = 8.

Size-Exclusion Chromatography Analysis. Molecular weight estimation was performed using Superdex 200 10/300 GL column in buffer containing 50 mM Tris, 250 mM NaCl, 1 mM BME, pH_{20 °C} = 8.0 at room temperature. Column was calibrated using Gel Filtration Molecular Weight Markers Kit for Molecular Weights 12,000-200,000 Da (Sigma Aldrich).

Sample Preparation and SAXS Analysis. Rtt103₁₄₁₋₂₄₆ was measured in 50 mM Tris, 300 mM NaCl, 10 mM BME, pH = 8.0 (4 °C) with 1.9, 4.9, and 7 mg/mL concentration at 4 °C. Data from low and high concentration measurements was merged. Rtt103₁₋₂₄₆ was measured in 25 mM KH₂PO₄, 300 mM KCl, 10 mM BME, pH_{20 °C} = 6.5 with 5.2 mg/mL concentration at 4 °C. Rtt103₁₋₂₄₆-phosphomimetic complex was prepared by mixing phospho-mimetic with 10-fold molar excess of Rtt103₁₋₂₄₆. The formed complex was purified using Superdex 200 Increase 10/300 (GE Healthcare) in 25 mM KH₂PO₄, 300 mM KCl, 10 mM BME, pH_{4 °C} = 6.5. The peak fractions were pooled and concentrated on

Amicon Ultra 0.5 mL Centrifugal filters (Millipore) with 3,000 Da cutoff. Dilution series were measured at 4 °C.

The SAXS data were collected on the BioSAXS-1000, Rigaku at CEITEC. Data were collected at X-ray beam wavelength $\lambda = 1.54 \text{ \AA}$. Sample to detector (PILATUS 100K; Dectris Ltd.) distance was 0.485 m, covering a scattering vector range from 0.008 to 0.65 \AA^{-1} . For buffer and sample one 2D image was collected with an exposure time of 60 min per image. Data are plotted using GnuPlot 4.6. Evaluation of the solution scattering of the atomic models and the fitting to experimental data were performed by CRY SOL (42). Ab initio modeling was performed by DAMMIN (40). Superposition of ab initio and atomic models was performed by SUPCOMB (43).

CORAL Modeling. Rigid body modeling was performed by CORAL (33) using Rtt103_{141–246} crystal structure (PDB ID code: 5M48), Rtt103-CID—pS₂pS₇-CTD complex structure (PDB ID code: 5M9D), protein sequences of phospho-serine mimetic and sequence of Rtt103_{1–246} as a constrain.

In CORAL modeling configuration setup each dimer of Rtt103_{1–246} was composed of two Rtt103-CIDs connected by 12-aa flexible linker to the Rtt103_{141–246} coiled-coil dimer. The interaction between Rtt103-CID and the CTD stretch was restrained based on the Rtt103-CID—pS₂pS₇-CTD complex structure, the nonbound regions of CTD were kept flexible. Different arrangements of CIDs were tested the calculation for each interaction scenario was repeated 10 times.

EOM Modeling. Rtt103_{1–246} was modeled into SAXS experimental data by EOM 2.0 (31) applying standard setup using Rtt103-CID (PDB ID code: 2KM4), Rtt103_{141–246} (PDB ID code: 5M48) and keeping unstructured elements flexible.

Fluorescence Anisotropy. The equilibrium binding of Rtt103 CID and its mutants to different CTD-peptides was analyzed by FA. The CTD peptides were N-terminally labeled with the 5,6-carboxyfluorescein (FAM); for the list of peptides, see Table S5. The measurements were conducted on a FluoroLog-3 spectrofluorometer (Horiba Jobin-Yvon Edison). The instrument was equipped with a thermostated cell holder with a Neslab RTE7 water bath (Thermo Scientific). Samples were excited with vertically polarized light at 467 nm, and both vertical and horizontal emissions were recorded at 516 nm. All measurements were conducted at 10 °C in 35 mM KH₂PO₄, 100 mM KCl (pH 6.8). Each data point is an average of three measurements. The experimental binding isotherms were analyzed by DynaFit (44).

Crystal Structure Determination. Selenomethionine-containing crystals were subjected to multiple anomalous dispersion diffraction experiments. Specifically, the data were collected at the Se absorption edge, at the inflection point, and at high and low energy remote wavelengths. Data were processed and integrated using XDS package (45). Unmerged XDS.ASCII file containing all of the reflections was processed by the program Pointless of CCP4 software suite (46, 47), which resulted in the unambiguous F4132 space group assignment. Reflections were scaled using the program Aimless (47) of the CCP4 software suite. The same procedure was applied to all four datasets. Scaled data were analyzed by the Xtriage module of Phenix software suite (48), which indicated reliable anomalous signal extended to 3.8 Å. The data were then subjected to heavy atom search by the Hyss module of Phenix software suite while keeping the resolution at 3.8 Å (49, 50). Four possible solutions were found by Hyss (each containing six Se atoms) with a Hyss CC of 0.461. The correct solution was determined by calculating the phase and carrying out density modification by programs SOLVE and RESOLVE, respectively [both parts of Phenix software suite (48, 51)]. Correct solution statistics as described in ref. 52: Estimated MAP CC $\times 100 = 65.13$, SKEW of 0.37, FOM (figure of merit) = 0.551, R factor (of density modified map) of 0.331. The phases for

the protein model were calculated by the program SOLVE, and the resulting density was further modified by the program RESOLVE (51). Phasing was carried out at the resolution of 3 Å. The resulting electron density was used for the automated model building using the program Buccaneer of the CCP4 Software Suite (47, 53). This resulted in a model, which was deemed to be 90% complete. Manual modifications included several Ramachandran plot and rotamer outliers corrections using the program Coot (54). The model was then used as a molecular replacement search model to phase a dataset processed to 2.593 Å (low energy remote) using Phaser (55). The resulting solution was subjected to Phenix.Refine (56) module of Phenix crystallographic suite of programs (48). Individual XYZ coordinates and individual B factors were refined and included TLS groups as determined by TLS motion determination server, which resulted in the overall crystallographic *R* and *R*_{free} factors of 24.55% and 26.40%, respectively. Ramachandran-plot statistics: favored, 99%; allowed, 1% as defined by wwPDB validation report.

NMR Measurements and Structure Determination. All NMR spectra for the backbone and side-chain assignments were recorded on Bruker AVANCE III HD 950, 850, and 700 MHz spectrometers equipped with cryoprobes at a sample temperature of 20 °C using 1 mM uniformly ¹⁵N, ¹³C-labeled Rtt103-CID in 35 mM KH₂PO₄, 100 mM KCl, pH 6.8 (20 °C) (90% H₂O/10% D₂O), and 2.5 mM pS₂pS₇-CTD peptide. Initial nuclei assignment was transferred from Biological Magnetic Resonance Bank entries 17044 and 16411 and confirmed by CBCACONH and HCCH-TOCSY spectra. The structure of the complex was determined as previously described (30). Ramachandran-plot statistics: favored, 94%, allowed, 4%; outliers, 2% as defined by wwPDB validation report. Determination of chemical shift perturbation (CSP) value was performed as previously described (30).

Trypsin Digestion and Mass-Spectrometry Analysis. For the limited proteolysis study, 0.8 mg/mL Rtt103 was mixed with 50 ng/μL trypsin (1:100 wt/wt; Promega). Samples were incubated at 20 °C for 5, 10, 15, 30, 45, 60, 120 min and overnight. The cleavage reaction was stopped by mixing with 4xSDS-gel loading dye and 5 min boiling at 95 °C. The samples were analyzed on 18% SDS/PAGE. One-dimensional gel areas to be analyzed were excised from the gel and after destaining and washing procedures, each gel plug was incubated with trypsin (125 ng of trypsin, 2 h, 40 °C). MALDI-MS and MS/MS analyses of tryptic digests were performed on an Ultraflextreme mass spectrometer (Bruker Daltonik) operated by FlexControl 3.3 software (Bruker Daltonik). Peptide maps were acquired in reflectron positive mode (25 kV acceleration voltage) with 800 laser shots. Peaks with minimum S/N = 10 were picked out for MS/MS analysis employing LIFT arrangement with 600 laser shots for each peptide. α -Cyano-4-hydroxycinnamic acid was used as the matrix in combination with AnchorChip target. External mass calibration procedure was employed, using a mixture of seven peptide standards (Bruker Daltonik) covering the mass range of 700–3,100 Da. The FlexAnalysis 3.3 and MS BioTools 3.2 (Bruker Daltonik) software were used for data processing. MASCOT 2.2 (MatrixScience) search engine was used for processing the MS and MS/MS data. For MALDI MS data, mass tolerance of 30 ppm was allowed for peptide mapping and 0.5 Da for MS/MS ion searches. Oxidation of methionine and propionamidylation of cysteine as optional modifications and one enzyme miscleavage were set for all searches. Peptides with statistically significant peptide score ($P < 0.05$) were considered. Manual MS/MS spectra assignment validation was done.

Visualization of Structures and Preparation of Movies. Structures were visualized and movies prepared using PyMOL (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC) or UCSF Chimera (57).

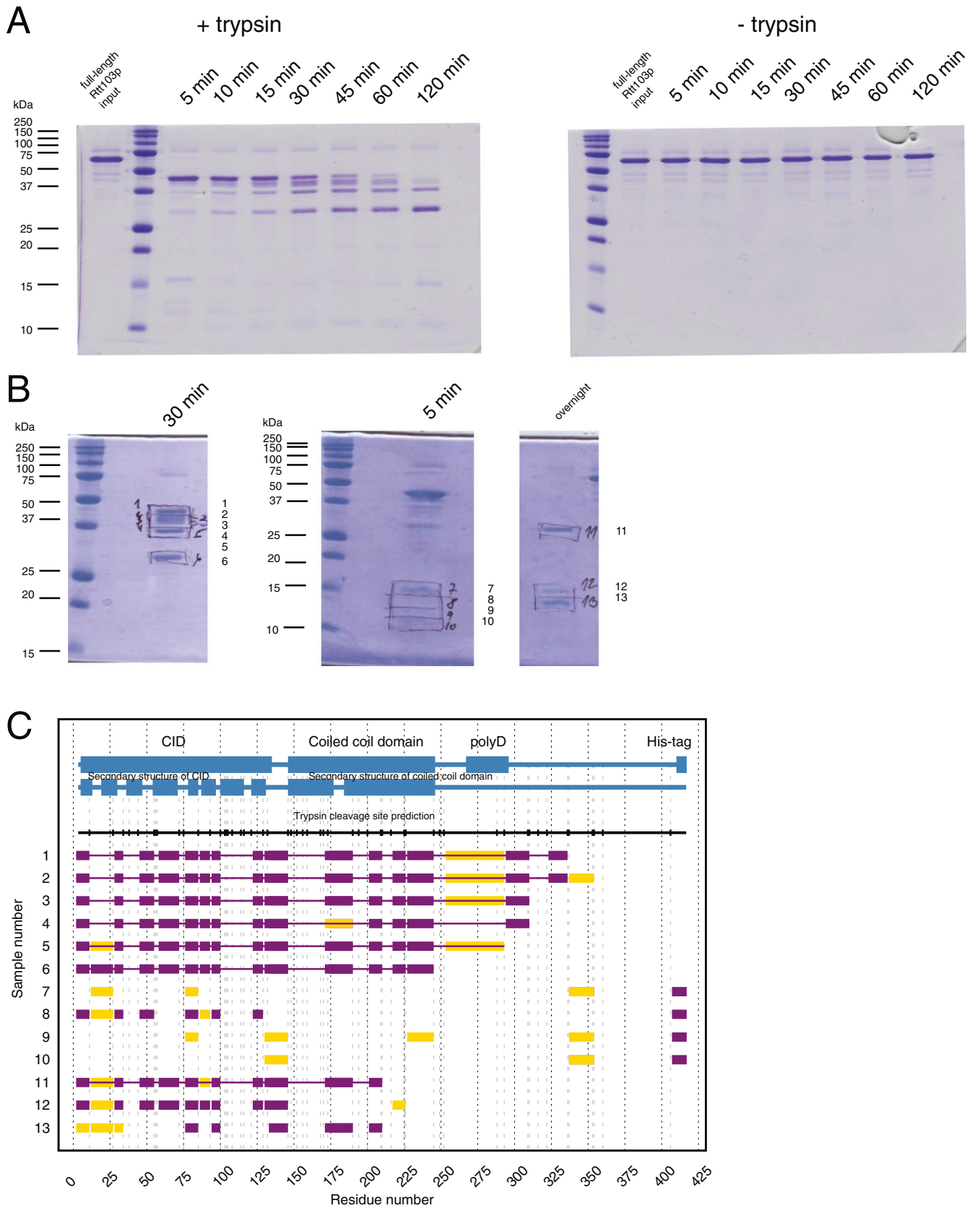


Fig. S1. Evaluation of Rtt103 structure by limited proteolysis. Trypsin digestion of Rtt103, which was recombinantly expressed in *E. coli* and purified to homogeneity, revealed a stable protein fragment of ~30 kDa. The remaining C-terminal part (amino acids 247–409) was efficiently digested. (A) Limited proteolysis study of the Rtt103 visualized on 18% SDS/PAGE. Sample of Rtt103 with addition of trypsin protease (+trypsin, *Right*) and the negative control (–trypsin, *Left*); 5, 10, 15, 30, 45, 60, 120 min time points; molecular mass standards from Precision Plus Protein, Bio-Rad. (B) Image of 18% SDS/PAGE with samples from A subjected to the subsequent MALDI-MS/MS analysis. The number of the sample is indicated on the *Right*. (C) Summary of the results of the MALDI-MS/MS analysis. The sample number indicated on the *Right* corresponds to the number given on B. Blue boxes indicate the domain organization of Rtt103 and known α -helical regions. Black vertical serif continued as dashed vertical gray lines indicate predicted trypsin cleavage sites. Purple boxes indicate peptides detected by trypsin digestion. Yellow boxes indicate peptides detected with low signal. Horizontal purple line indicates the presumed span of Rtt103 fragment present in the gel bands.

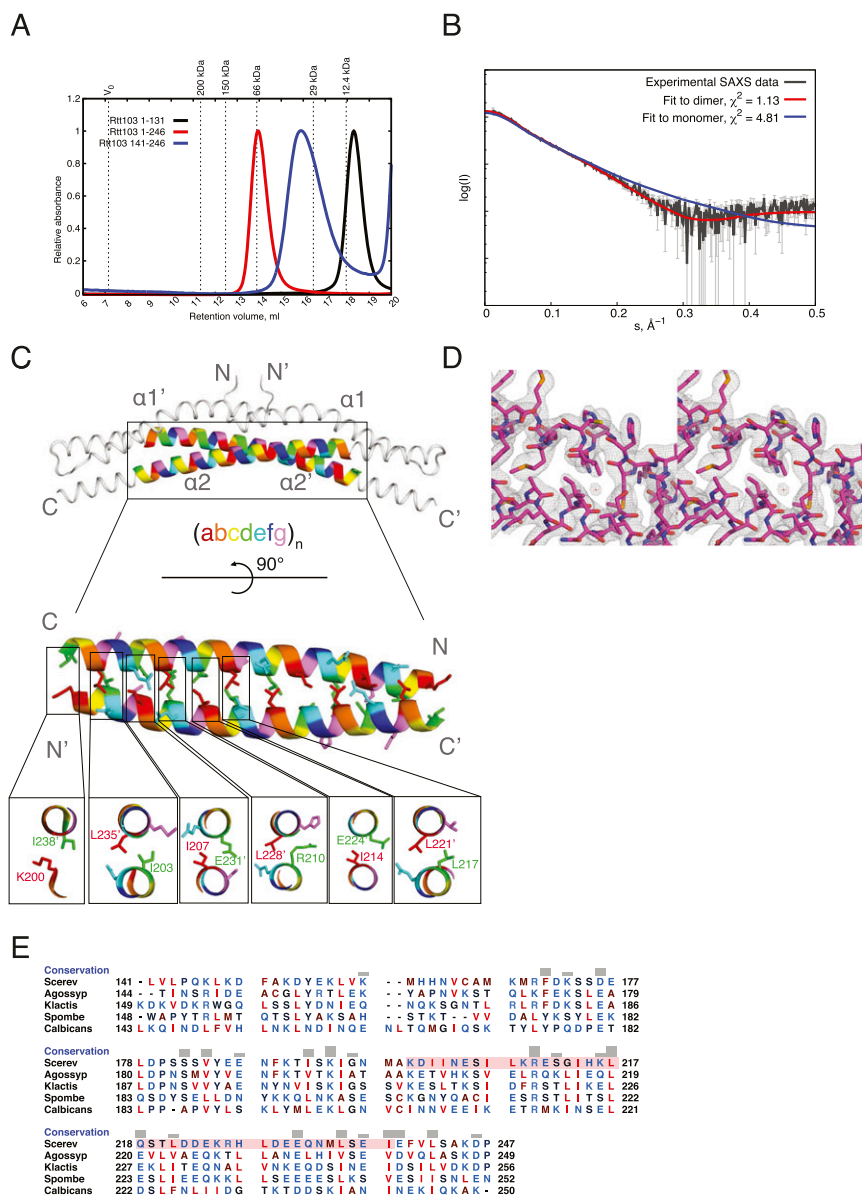


Fig. S2. Study of the Rtt103 coiled-coil domain. (A) Size-exclusion chromatography analysis of recombinant Rtt103 truncation constructs. A Superdex 200/300 GL column was used. Void volume (V_0), retention volumes, and molecular masses of the protein standards are indicated at the top (gray dashed lines). (B) Evaluation of Rtt103_{141–246} experimental SAXS scattering data (black) against theoretical scattering data pertaining to the monomer (blue) and dimer (red) structures derived by CRYSOLOG (42). (C) Crystal structure of the Rtt103_{141–246} coiled-coil domain with the characteristic coiled-coil region (defined by SOCKET software; ref. 58). The residues that form the “knobs-into-holes” packing are highlighted according to the position in the classical coiled-coil heptad (abcdefg). The zoom of the cross-sections from selected residues is shown at Lower. Respective N- and C-termini, as well as α -helices, of two polypeptide chains are indicated. (D) Stereoview of the coiled-coil domain electron density map ($2F_0 - F_c$ density map contoured at 1.0σ ; gray mesh). (E) Multiple sequence alignment of the dimerization domain region in Rtt103 homologs (*S. cerevisiae*, Q05543; *A. gossypii*, AAL081Cp; *K. lactis*, KLLA0C10758p; *S. pombe*, CAA21273.1; *C. albicans*, CaO19.7662). Residues are colored according to the hydrophobicity index. Alignment prepared in ClustalX (59), visualized in UCSF Chimera (57). The coiled-coil region of *S. cerevisiae* Rtt103 is indicated in pink.

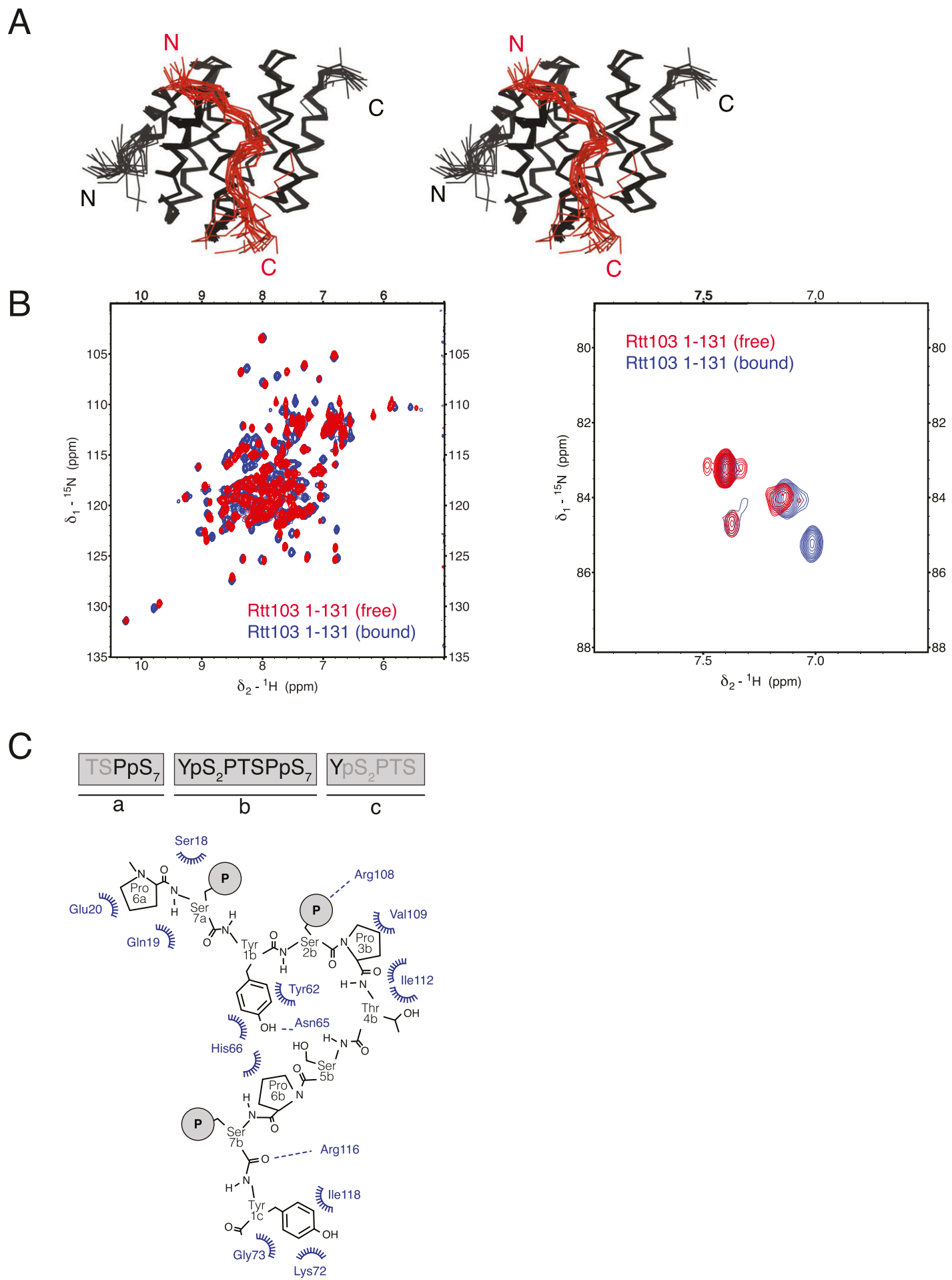


Fig. S3. Solution NMR structure of the Rtt103 CID with the pS₂pS₇-CTD peptide. (A) Overlay of the 20 lowest-energy structures of the Rtt103 CID backbone (black ribbon) in complex with pS₂pS₇-CTD (red ribbon), shown in stereoview. N- and C-termini are indicated. (B) Overlay of [¹H/¹⁵N]-HSQC spectra, with Rtt103 CID in free form (red) and bound to the pS₂pS₇-CTD peptide (blue). Region corresponding to the amides of protein (*Left*); region corresponding to the Arg side chains (*Right*). (C) Schematic diagram of Rtt103 CID (blue labels) and pS₂pS₇-CTD (black) interactions. Hydrophobic contacts are indicated by spoked arcs, whereas hydrogen bonds are indicated by dashed lines. The pS₂pS₇-CTD sequence used for structure determination is indicated at *Upper*, where residues that showed interaction with the Rtt103 CID are shown in black.

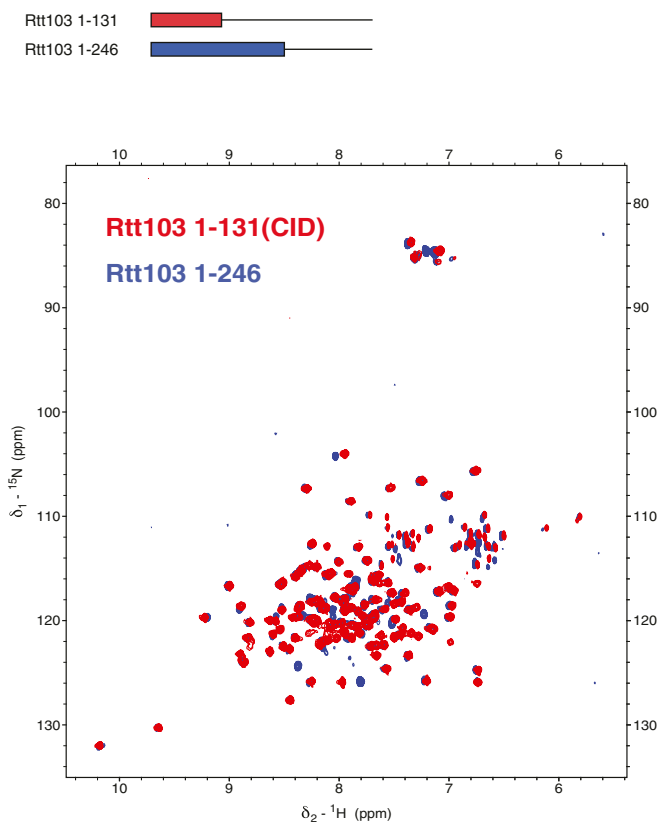


Fig. S5. Presence of the coiled-coil domain does not influence the structure of the Rtt103 CID. Overlay of $[^{15}\text{N}, ^1\text{H}]$ -TROSY spectra of Rtt103_{1–246} (blue) and Rtt103_{1–131} (red), measured on a 950-MHz spectrometer. CID, CTD interacting domain.

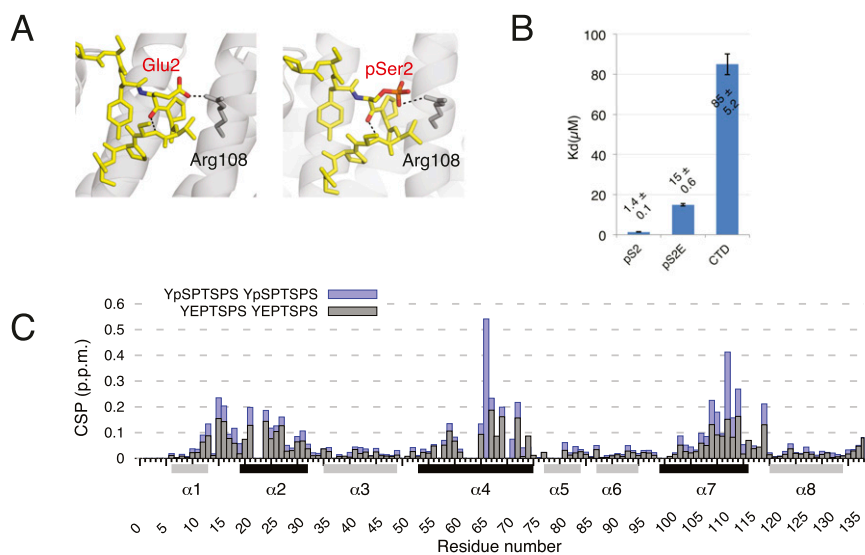


Fig. S6. Glutamate substitution mimics CTD phosphorylation. To overcome production of high yields of homogeneously phosphorylated CTD, we prepared a phospho-mimicking mutant of CTD where pS₂ is substituted by a glutamate residue (Glu2, E). $[^{15}\text{N}, ^1\text{H}]$ -HSQC titration data and binding studies confirmed that the Rtt103 CID interacts similarly with the pS₂-CTD peptide and the corresponding mutant pS₂E-CTD. (A) Scheme showing a putative interaction (black dashed line) between the Arg108 residue of Rtt103 and the E residue of pS₂E-CTD (Left) or pS₂ of pS₂-CTD (Right). CTD peptides are shown as yellow sticks, whereas the Rtt103 CID is shown as gray cartoons. The conformation of the glutamate residue was modeled using PyMOL, based on the structure with PDB ID code 2L0I. (B) Binding affinities of the Rtt103 CID toward 5,6-carboxyfluorescein-labeled pS₂, pS₂E, and unphosphorylated CTD peptides, measured by fluorescence anisotropy. (C) Chemical shift perturbations (CSP) of the Rtt103 CID upon interaction with pS₂-CTD (blue) or pS₂E-CTD (gray), plotted against residue numbers corresponding to the Rtt103 CID sequence. Secondary structure elements are shown below the x axis; helices involved in the interaction with the phosphopeptides are colored in black.

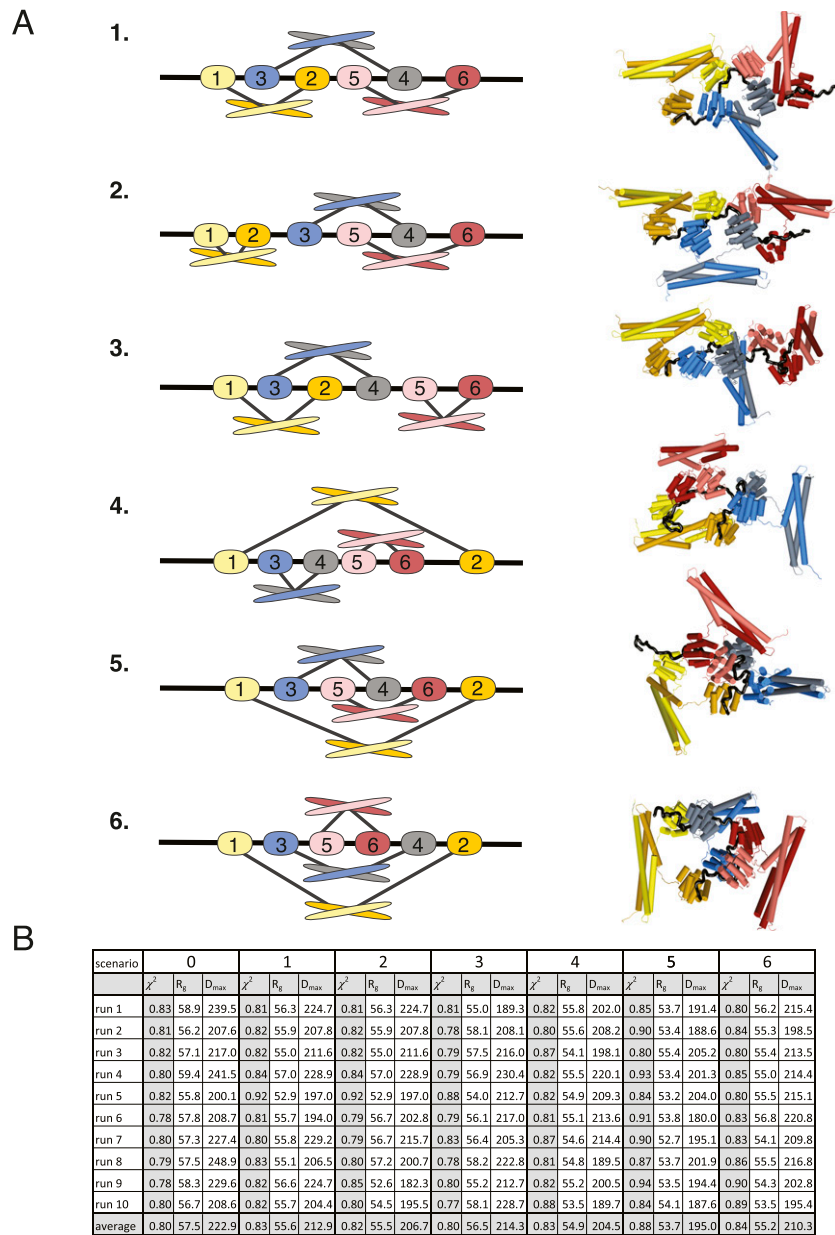


Fig. S7. Multisubunit arrangement of Rtt103_{1–246} across half the length of the CTD. (A) Scheme of six different interaction scenarios (*Left*) used for CORAL modeling. The order in which the CIDs bind the CTD (black ribbon) repeats is indicated by numbers. Ten CORAL runs were performed for each scenario, and a representative model is shown for each (*Right*). (B) Table with values of goodness of the fit (χ^2), radius of gyration (R_g), and maximum intraparticle distance (D_{max}) for each CORAL run of each interaction scenario. Scenarios 1–6 correspond to those in A, whereas scenario 0 corresponds to that shown in Fig. 3. The values of R_g and D_{max} are determined using CRYSOLOG (42) for each CORAL model.

Table S1. Data collection and phasing for the structure of the construct Rtt103_{141–246}

Data collection	Peak	Inflection	High-energy remote	Low-energy remote
Space group	F4 ₁ 23	F4 ₁ 23	F4 ₁ 23	F4 ₁ 23
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> , Å	217.14, 217.14, 217.14	217.14, 217.14, 217.14	217.14, 217.14, 217.14	217.14, 217.14, 217.14
α , β , γ , °	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Wavelength	0.979	0.9792	0.9713	0.9919
Resolution, Å	63.18–2.65 (2.74–2.65)	63.18–2.65 (2.74–2.65)	63.09–2.60 (2.69–2.60)	54.54–2.59 (2.70–2.59)
<i>R</i> _{sym} , %	15.8 (192)	16.1 (187)	35 (485)	21.6 (306)
<i>I</i> / σ <i>I</i>	23.2 (2.0)	25.9 (2.4)	21.6 (1.3)	21.9 (1.7)
Completeness, %	99.9 (99.3)	99.9 (99.2)	100 (99.9)	100 (99.9)
Redundancy	27.8	37.0	46.8	29.7

Values in parentheses are for highest-resolution shell.

Table S2. Refinement statistics for the structure of the construct Rtt103_{141–246}

Data collection	Refinement statistics
Resolution, Å	48.55–2.59
No. reflections	14,127
<i>R</i> _{work} / <i>R</i> _{free}	0.246/0.260
No. atoms	978
Protein	935
Water	43
<i>B</i> factors	
Protein	59.81
Water	46.12
rmsds	
Bond lengths, Å	0.021
Bond angles, °	0.463

Table S3. NMR distance and dihedral constraints for the complex between the Rtt103 CID and the pS₂pS₇-CTD peptide

NMR distance and dihedral constraints	Value
Distance restraints	
Total NOE	3,564
Intraresidue	842
Interresidue	2,722
Sequential ($ i - j = 1$)	1,667
Nonsequential ($ i - j > 1$)	1,897
Hydrogen bonds	96
Intermolecular distance restraints	49
Total dihedral angle restraints*	227
Protein	
ϕ	101
ψ	99

* α -helical dihedral angle restraints imposed for the backbone based on the CSI.

Table S4. Structure statistics for the complex between the Rtt103 CID and the pS₂pS₇-CTD peptide

Structure statistics	Value
Violations (mean and SD)	
Distance constraints (>0.4 Å)	0.1 ± 0.31
Dihedral angle constraints (>15°)	15.05 ± 1.6
Max. dihedral angle violation, °	52 ± 11
Max. distance constraint violation, Å	0.31 ± 0.05
Deviations from idealized geometry	
Bond lengths, Å	
Bond angles, °	1.5 ± 0.022
Average pairwise rmsd,* Å	
Rtt103 CID (7-12; 19-31; 54-73; 77-94; 100-116; 121-133)	
Heavy atoms	1.13 ± 0.16
Backbone atoms	0.41 ± 0.16
CTD (141-150)	
Heavy atoms	2.38 ± 0.57
Backbone atoms	1.17 ± 0.33
Complex	
All heavy atoms	2.23 ± 0.43
All backbone atoms	1.77 ± 0.46

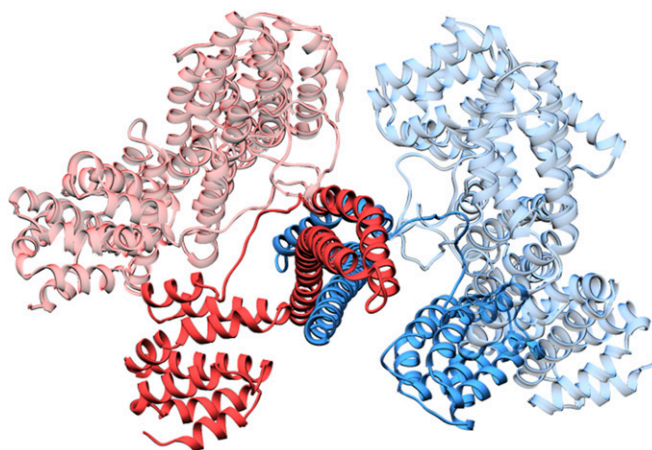
*Pairwise rmsd was calculated among 20 refined structures.

Table S5. List of peptides used in the study

Name	(Label)-sequence (N to C terminus)	Synthesis company	Description/use
pS ₂ pS ₇ -CTD	TSP(pS) Y(pS)PTSP(pS) Y(pS)PTS	Clonestar	Solution NMR structure determination
FAM-extrapS ₂ -CTD	FAM-TSPS Y(pS)PTSPS Y(pS)PTSPS	Clonestar	Fluorescence anisotropy measurements; NMR titration
FAM-pS ₂ -CTD	FAM- Y(pS)PTSPS Y(pS)PTSPS	Clonestar	Fluorescence anisotropy measurements; NMR titration
FAM-CTD	FAM- YSPTSPS YSPTSPS	Clonestar	Fluorescence anisotropy measurements
FAM-short pS ₂ pS ₇ -CTD	FAM- SPS Y(pS)PTSP(pS) YS	Clonestar	Fluorescence anisotropy measurements
FAM-long pS ₂ pS ₇ -CTD	FAM- SPS Y(pS)PTSP(pS) YSPTSPS Y(pS)PTSP(pS) YS	Clonestar	Fluorescence anisotropy measurements
FAM-pS ₇	FAM- YSPTSP(pS) YSPTSP(pS)	Clonestar	Fluorescence anisotropy measurements
FAM-pS ₂ E-CTD	FAM- YEPTSPS YEPTSPS	Clonestar	Fluorescence anisotropy measurements
pS ₂ E -CTD	YEPTSPS YEPTSPS	Clonestar	NMR titration
pS ₂ -CTD	Y(pS)PTSPS Y(pS)PTSPS	Clonestar	NMR titration

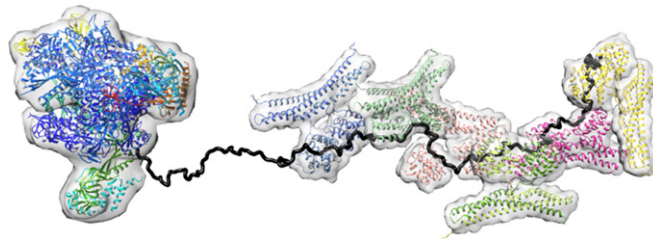
Table S6. Oligonucleotides used in the study

Name	Sequence (5' to 3')	Description/use
Rtt103_141_NdeI_forw	CAGCATATGTTGGTGTACCCAG	Forward primer for cloning of Rtt103 ₁₄₁₋₂₄₆ into pET22b
Rtt103_246_XhoI_rev	CACCTCGAGGTCTTTAGCAGATAAAAC	Reverse primer for cloning of Rtt103 ₁₄₁₋₂₄₆ into pET22b
Rtt103_246_forw	GTTTTATCTGCTAAAGACCTCGAGCACCACCACCACC	Forward primer for cloning of Rtt103 ₁₋₂₄₆
Rtt103_246_rev	GGTGGTGGTGGTGGTGCCTCGAGGTCTTTAGCAGATAAAA C	Reverse primer for cloning of Rtt103 ₁₋₂₄₆
Rtt103_R108E_forw	GGGACCTAAAAAAGAAGTTGTCAGAAGTTGTGAATATAC	Forward primer for Rtt103-CID R108E mutant
Rtt103_R108E_rev	GTATATTCACAACTTCTGACAACTTCTTTTTTAGGTCCC	Reverse primer for Rtt103-CID R108E mutant
K72E_forw	CATGTTGTTCAACAGGCTGAAGGTCAAAAAATTATTC	Forward primer for Rtt103-CID K72E and K72E/R116E mutant
K72E_rev	GAATAATTTTTTGACCTTCAGCCTGTGAACAACATG	Reverse primer for Rtt103-CID K72E and K72E/R116E mutant
Rtt103_R116E_forw	GTGAATATACTAAAGAAGAGAATATATTTTTCCAAGCAGG	Forward primer for Rtt103-CID R116E and K72E/R116E mutant
Rtt103_R116E_rev	CCTGCTTGAAAAATATATTCTCTCTTTTAGTATATTAC	Reverse primer for Rtt103-CID R116E and K72E/R116E mutant
sCTD universal Bam HI forw	CGTGGATCCCCGGAATTCACCTG	Forward primer for cloning of phophomimicking CTD into pET28b-SMT3
sE2 13 FLAG HindIII rev	GGAAGCTTACTTATCGTCGTCATCCTTGTAACTGCTGCTGCTGTTTCATA	Reverse primer for cloning of phophomimicking CTD into pET28b-SMT3



Movie S1. The two CIDs tethered to the coiled-coil domain tumble independently. The SAXS data were combined with available atomic-resolution structures (PDB ID codes: 2KM4, 5M48) using the ensemble-optimization method (EOM 2.0) to yield the conformational distribution of CID- and coiled-coil-domain arrangements. The movie starts with the overlay of individual conformers (Fig. 3) obtained from the ensemble of Rtt103₁₋₂₄₆ structures derived by EOM 2.0. These conformers are then displayed in sequence one after another. Then the morph that shows conversion from conformer 1 to conformer 5 is shown and rotated. The movie ends with the overlay of all conformers, where conformers 2 through 5 are shown with lower opacity. The two different polypeptide chains forming the dimer are shown in red and blue. The conformers are superimposed relative to the coiled-coil domain. In the resulting model, the coiled-coil domain defines the length of the protein. Connection via the 15-amino acid linker allows the CIDs to be positioned relatively close to each other but also to reach all across the coiled-coil domain and sample a large surrounding space to recognize the substrate. The ability of the CIDs to sample multiple conformations is restricted only by their tethering to the flexible linker. The movie was prepared using UCSF Chimera (57).

[Movie S1](#)



Movie S2. Model of the Rtt103-CTDome assembly involving the full-length RNAPII CTD. The model of RNAPII (Fig. 4) with the full-length CTD (26 repeats, budding yeast) is decorated with six Rtt103₁₋₂₄₆ dimers. The structure of RNAPII (PDB ID: 5F12) is combined with two CORAL models of the Rtt103₁₋₂₄₆-CTD complex, where the CTD-interacting domains (CIDs) are arranged in a consecutive manner. The movie was prepared using UCSF Chimera (57).

[Movie S2](#)