Structure and dynamics of the RNAPII CTDsome with Rtt103

Olga Jasnovidova[a], Tomas Klumpler[a], Karel Kubicek[a], Sergei Kalynych[a], Pavel Plevka[a], and Richard Stefl[a,1]

[a]CEITEC–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 Y_1, S_2, T_4, S_5, and S_7 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 S_2 and S_5, whereas Y_1, T_4, and S_7 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 (pS_5-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 "CTDsome" (27).

To probe the CTDsome 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.

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).

[1]To whom correspondence should be addressed. Email: richard.stefl@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

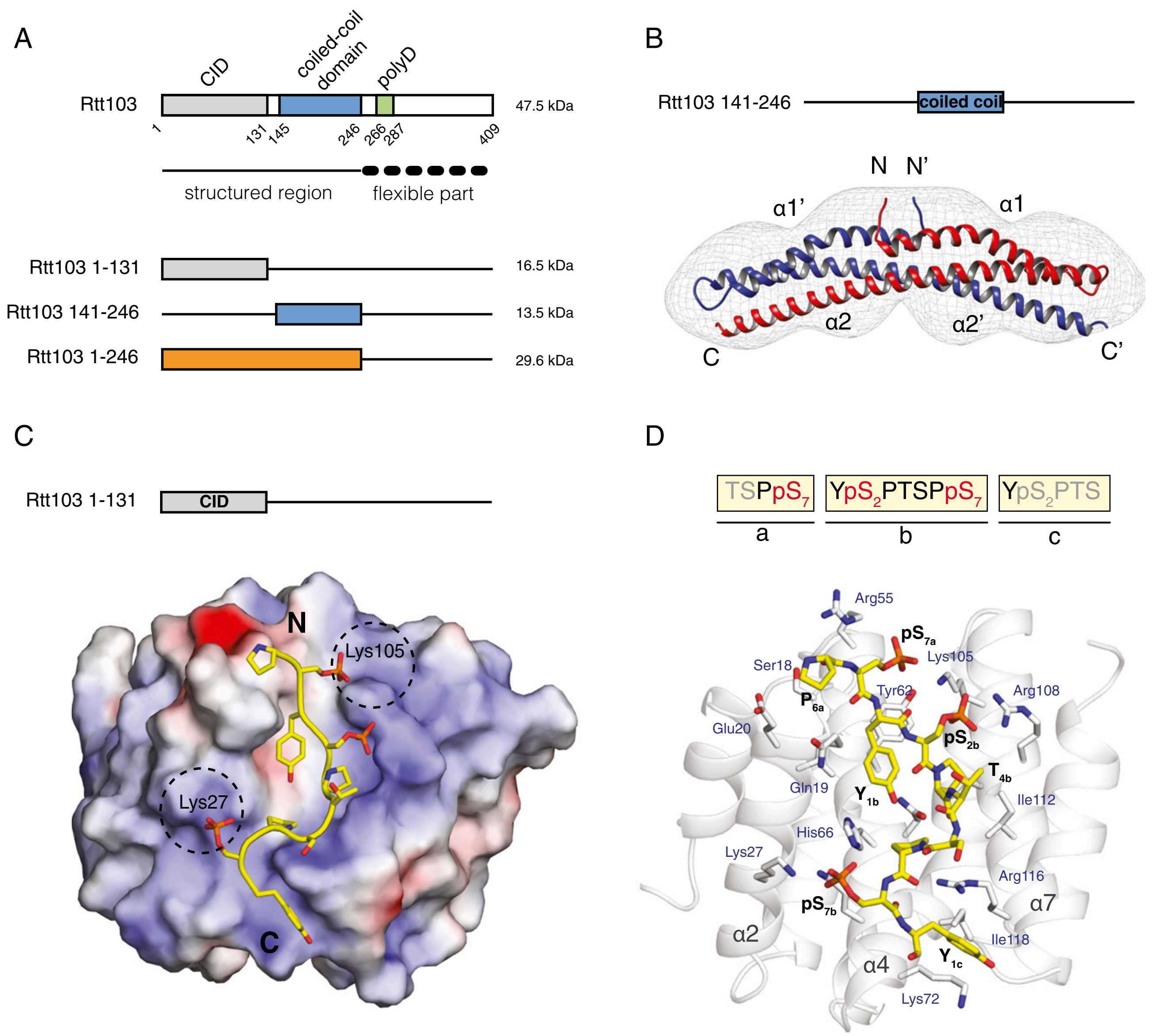

Fig. 1. Dimerization and RNAPII CTD recognition by Rtt103. (*A*) Scheme of Rtt103 domain organization (Upper). The numbers below the scheme represent borders of the amino acid segments. Structured and flexible regions were determined based on the limited proteolysis study (Fig. S1). The recombinant protein 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 pS_2pS_7-CTD peptide (yellow sticks; PDB ID code: 5M9D). The N- and C-termini of the peptide are indicated. Dashed black circles indicate electropositive areas that accommodate pS_7 residues. (*D*) Detailed view of the Rtt103 CID (gray cartoons) bound to the pS_2pS_7-CTD peptide (yellow sticks). Highlighted Rtt103 CID residues (gray sticks, blue labels) form hydrophobic contacts and putative hydrogen bonds with the pS_2pS_7-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 α1-helix (Gln146-Glu177) with a small bend in the middle and a long α2-helix of ∼105 Å in length (Val184-Asp246) (Fig. 1*B*). In the crystal, two protein chains form a dimer where the α2-helices are arranged in an antiparallel fashion (Fig. 1*B*). This architecture of the dimer is in agreement with findings from gel filtration experiments (Fig. S2*A*) and SAXS data (Fig. S2*B*). Importantly, the central region of the α2-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. S2*C*), 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 pS_2-pS_7-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 $TSPpS_7$ $YpS_2PTSPpS_7$ YpS_2PTS peptide (termed pS_2pS_7-CTD) confirmed the previously reported observation regarding the recognition of the upstream region of pS_2pS_7-CTD, and further revealed information regarding the recognition of the downstream region of pS_2pS_7-CTD (Fig. 1 *C* 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 pS_2pS_7-CTD peptide binds at the conserved surface formed by helices α2, α4, and α7 of the Rtt103 CID (Fig. 1*D* 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 pS_2pS_7-CTD peptide. Specifically, P_{6a} lies in the proximity of the hydrophobic area formed by the N-terminal tip of the α2-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. S4*C* and refs. 27 and 30) caused a drop of 30- to 50-fold in the affinity between pS_2pS_7-CTD and $Rtt103_{1–131}$ (K_D = 33 ± 1.2 μM for Ile112Ala, K_D = 80 ± 11 μM 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 α4 and α7. The positioning of Y_{1c} near the tip of helices α4 and α7 induces a second sharp turn in the pS_2pS_7-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 pS_2pS_7-CTD. We found that charge-swapping mutations at the interacting sites of Rtt103 (not affecting the structural integrity; Fig. S4*C*) resulted in pronounced affinity decrease between Rtt103 and pS_2pS_7-CTD (K_D = 9.7 ± 0.7, 51 ± 2.2, and 65 ± 8.9 μM 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. 1*C*). Although our data did not indicate the presence of intermolecular contacts for the pS_7 residues, the positions of these residues are indirectly defined by the nuclear Overhauser effects from the neighboring residues. Therefore, residues pS_7 are likely involved in charge–charge interactions with Lys27 and Lys105 (which is part of the poly-Lys tract Lys103-Lys104-Lys105) (Fig. 1*C*). We found that the Lys27Glu mutant (perturbation of one of the pS_7 binding pockets; Fig. 1*C*) showed lower binding only for the pS_7 containing peptide (K_D of 13.2 ± 0.3 μM and 28.5 ± 1 μM for wild type and Lys27Glu, respectively) but not for the pS_2-containing peptide (K_D of 1.6 ± 0.07 μM and 2 ± 0.8 μM for wild type and Lys27Glu, respectively). Altogether, the key finding is that the Rtt103 CID interacts with pS_2pS_7-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 (∼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 ± 0.06 μM, and to the long CTD substrate with a K_D of 0.3 ± 0.01 μM. The isolated CID binds to the minimal CTD-binding moiety with a K_D of 1.65 ± 0.13 μM, and to the long CTD substrate with a K_D of 0.5 ± 0.01 μM. 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 pS_2E-CTD {$pSer_2$-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

BIOCHEMISTRY

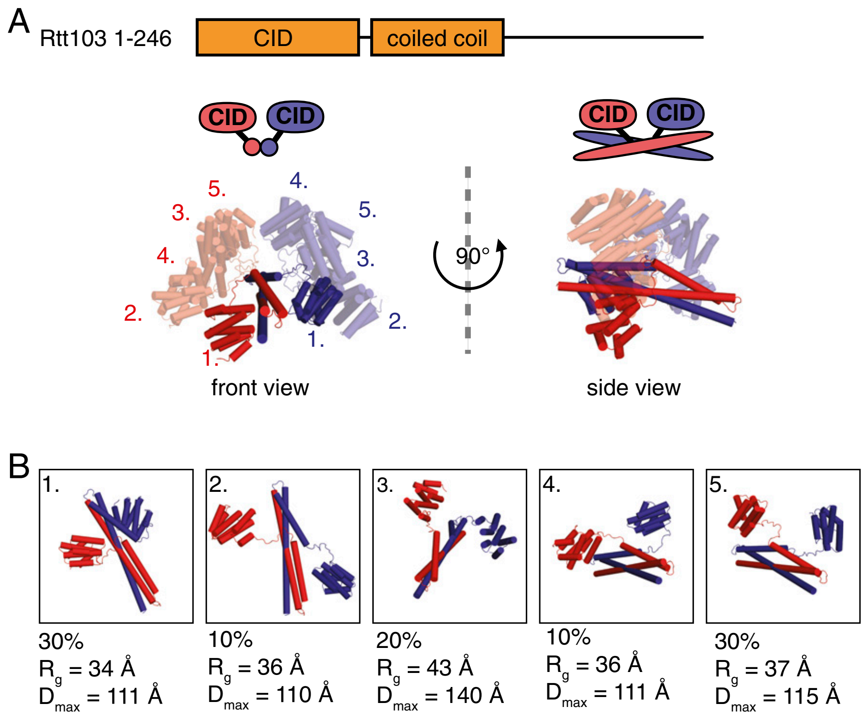

Fig. 2. The two Rtt103 CIDs are tethered by a coiled-coil domain but tumble independently. (*A*) Overlay of individual conformations from the ensemble of free Rtt103$_{1–246}$ structures derived using the ensemble-optimization method (EOM 2.0) (31) (χ^2 = 1.001); front and side views are provided. Conformations 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 Rtt103$_{1–246}$ 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 Rtt103$_{1–246}$: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 Rtt103$_{1–246}$: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. S7*B*). One Rtt103 dimer is accommodated on four CTD repeats (**PS YEPTSPS** YEPT**SPS YEPTSPS Y**EPTS; 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 pS$_2$-CTD repeats and exposes the pS$_5$ sites on the CTD, which stimulates the activity of RPAP2 pS$_5$-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 $P_{6a}pS_{7a}Y_{1b}pS_{2b}P_{3b}T_{4b}S_{5b}P_{6b}pS_{7b}Y_{1c}$ sequence. These findings indicate that the Rtt103 CID binds a longer CTD stretch than previously reported (29). Accommodation of the core $P_{6a}S_{7a}Y_{1b}S_{2b}P_{3b}T_{4b}S_{5b}$ 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 and especially downstream region of the CTD peptide on the CID surface varies among CID–CTD peptide complexes (26, 27, 37, 38). In our solution NMR structure, the pS_2pS_7-CTD peptide makes multiple contacts with a conserved Arg116 residue and exits the binding pocket between helices α4 and α7, thereby occupying almost the whole conserved surface of the CID. Similarly, the solution structure of pT_4-CTD–Rtt103-CID and crystal structure of RPRD1B/RPRD1A also exhibit an elongated conformation of the CTD (27, 30). In these structures, the binding of additional residues at the C-terminal of the β-turn stretch of the CTD significantly changes the conformation of bound CTD, which could be important for the higher order arrangement of CIDs and exposure of the nonbound CTD residues to other factors (27). Additionally, the extended interaction surface of the CIDs in Rtt103 and RPRD1B/RPRD1A may partially

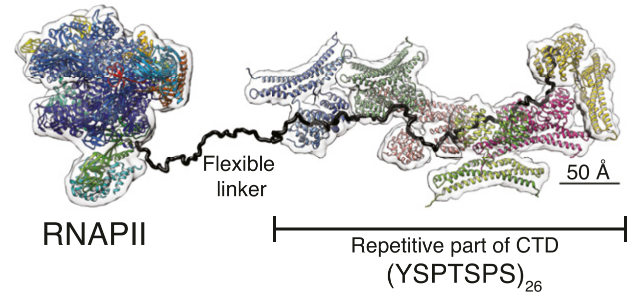

Fig. 4. Model of the Rtt103-CTDsome assembly involving the full-length RNAPII CTD. The model of the RNAPII with the full-length CTD is decorated with six dimers of $Rtt103_{1–246}$ (Movie S2). The structure of RNAPII (PDB ID code: 5F12) is combined with two CORAL models of the $Rtt103_{1–246}$–CTD complex, where C-interacting domains (CIDs) are arranged in a consecutive manner.

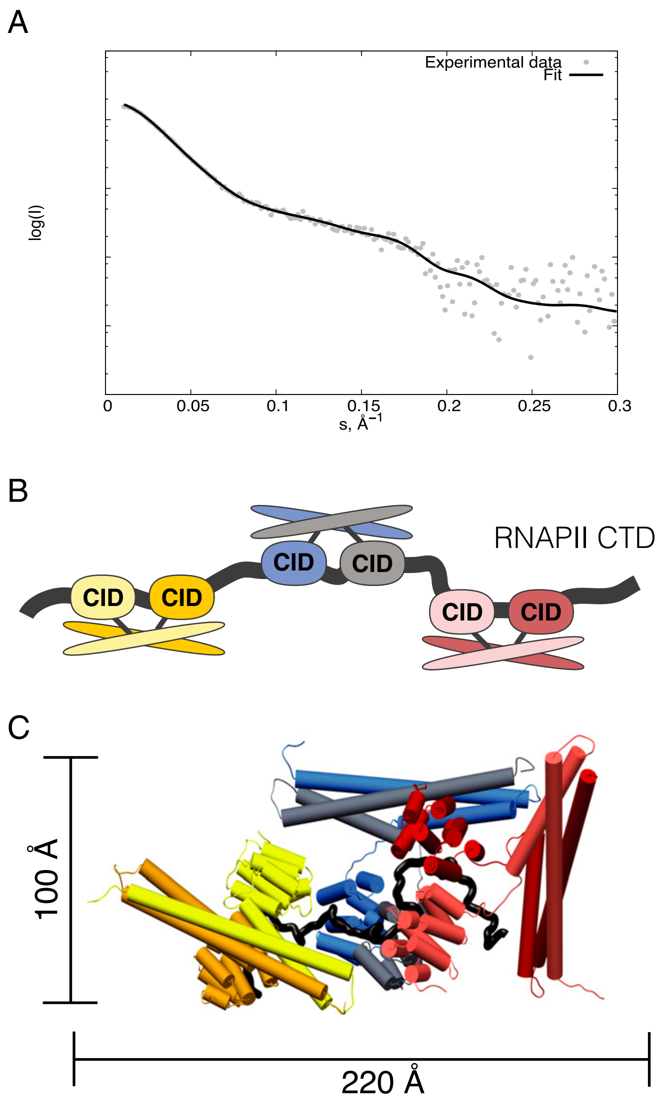

Fig. 3. Multisubunit arrangement of $Rtt103_{1–246}$ across half the length of the CTD. (*A*) Comparison of the theoretical scattering (black) derived using CORAL based on the $Rtt103_{1–246}$-CTD model against the experimental scattering data (gray). (*B*) Scheme showing the arrangement of the individual $Rtt103_{1–246}$ molecules across half the length of the CTD (13 heptad repeats), which is the CTD construct employed during modeling using CORAL (consecutive interaction, "scenario 0," see Fig. S7). (*C*) Representative model obtained from the CORAL calculation for the consecutive interaction scenario. Color coding is according to the scheme in *B*.

explain their higher affinity toward the CTD compared with the affinity of the CIDs in Nrd1 and Pcf11 (38, 39).

The regulation of transcription requires a complex interplay involving fast and dynamic exchange of multiple RNA/protein-binding factors. This network is largely maintained and balanced by means of a structurally adaptable CTD which increases the local concentration of factors and allosterically regulates their activity of transcription and RNA processing factors near the emerging nascent transcript. Our reconstruction of the CTD in complex with Rtt103 shows that Rtt103 can fully explore the repetitiveness and length of the CTD sequence by occupying CTD in a repetitive manner ("beads-on-a-string") while keeping the entire arrangement flexible and dynamic. Rtt103 dimerization creates topological and mobility restraints, which, in turn, tune the protein's affinity toward the CTD by increasing the local concentration of CIDs, and further governs the exposure of the CTD sequence to other protein-binding factors. We suggest that CTD code readers, such as Rtt103, and other CTD effector molecules form a high-order structure that is essential for the conception and interpretation of the CTD code (Fig. 4 and Movie S2). The tail-like architecture allows for quick exchange of binding factors and coordinates the regulatory networks necessary for efficient gene regulation. Interestingly, the structure of the CTD tail decorated with Rtt103 dimers appears to be fully extended and protrudes away from the invariant core of the RNAPII (Fig. 4 and Movie S2). The structural model of the Rtt103–CTDsome demonstrates how the CTD allows forming diverse and tuneable protein assemblies around the invariant core of the RNAPII, supporting the complex networks necessary for efficient gene regulation.

Methods

A full description of the methods for protein expression, purification, and fluorescence anisotropy measurements as well as NMR, X-ray, and SAXS data collection and analysis is provided in *SI Methods* and Tables S5 and S6. SAXS data are deposited in Small Angle Scattering Biological Data Bank (SASBDB ID: SASDCZ2). The model and the diffraction data containing phase information was deposited to Protein Data Bank, PDB ID code: 5M48. The atomic coordinates and restraints for the NMR ensemble of the Rtt103-CID—pS_2pS_7–CTD complex have been deposited in the Protein Data Bank, PDB ID code: 5M9D.

ACKNOWLEDGMENTS. We thank P. Kuzmic and C. Hofr for helpful advice on binding data analysis; J. Houser and J. Kosourova for help with crystallization screens; O. Sedo and D. Fridrichova for mass-spectrometry analysis; L. Mukhamedova for collecting crystal diffraction data; and C. Jeffries and A. Panjkovich for helpful advice on SAXS data analysis. The results of this research have been acquired within the CEITEC 2020 (LQ1601) project with financial contribution made by the Ministry of Education, Youths and Sports of the Czech Republic (MEYS CR) and special support paid from the National Programme for Sustainability II funds. The scientific data were obtained with the support of Josef Dadok National NMR Centre of CEITEC, supported by the Czech Infrastructure for Integrative Structural Biology research infrastructure (LM2015043 funded by MEYS CR), the X-ray Diffraction and Bio-

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.

1. Eick D, Geyer M (2013) The RNA polymerase II carboxy-terminal domain (CTD) code. *Chem Rev* 113:8456–8490.
2. Chapman RD, Heidemann M, Hintermair C, Eick D (2008) Molecular evolution of the RNA polymerase II CTD. *Trends Genet* 24:289–296.
3. 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.
4. West ML, Corden JL (1995) Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. *Genetics* 140:1223–1233.
5. Liu P, Kenney JM, Stiller JW, Greenleaf AL (2010) Genetic organization, length conservation, and evolution of RNA polymerase II carboxyl-terminal domain. *Mol Biol Evol* 27:2628–2641.
6. Buratowski S (2003) The CTD code. *Nat Struct Biol* 10:679–680.
7. Mayer A, et al. (2010) Uniform transitions of the general RNA polymerase II transcription complex. *Nat Struct Mol Biol* 17:1272–1278.
8. Mayer A, et al. (2012) CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science* 336:1723–1725.
9. 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.
10. Kim H, et al. (2010) Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nat Struct Mol Biol* 17:1279–1286.
11. Tietjen JR, et al. (2010) Chemical-genomic dissection of the CTD code. *Nat Struct Mol Biol* 17:1154–1161.
12. 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.
13. 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.
14. Schüller R, et al. (2016) Heptad-specific phosphorylation of RNA polymerase II CTD. *Mol Cell* 61:305–314.
15. 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.
16. Cramer P, et al. (2000) Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* 288:640–649.
17. Cramer P, Bushnell DA, Kornberg RD (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* 292:1863–1876.
18. Spåhr H, Calero G, Bushnell DA, Kornberg RD (2009) Schizosacharomyces pombe RNA polymerase II at 3.6-A resolution. *Proc Natl Acad Sci USA* 106:9185–9190.
19. Meinhart A, Kamenski T, Hoeppner S, Baumli S, Cramer P (2005) A structural perspective of CTD function. *Genes Dev* 19:1401–1415.
20. Meredith GD, et al. (1996) The C-terminal domain revealed in the structure of RNA polymerase II. *J Mol Biol* 258:413–419.
21. 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.
22. Portz B, et al. (2017) Structural heterogeneity in the intrinsically disordered RNA polymerase II C-terminal domain. *Nat Commun* 8:15231.
23. Gibbs EB, et al. (2017) Phosphorylation induces sequence-specific conformational switches in the RNA polymerase II C-terminal domain. *Nat Commun* 8:15233.
24. Jasnovidova O, Stefl R (2013) The CTD code of RNA polymerase II: A structural view. *Wiley Interdiscip Rev RNA* 4:1–16.
25. 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.
26. Meinhart A, Cramer P (2004) Recognition of RNA polymerase II carboxy-terminal domain by 3′-RNA-processing factors. *Nature* 430:223–226.
27. 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.
28. Grigoryan G, Degrado WF (2011) Probing designability via a generalized model of helical bundle geometry. *J Mol Biol* 405:1079–1100.
29. 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.
30. Jasnovidova O, Krejcikova 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.
31. Tria G, Mertens HDT, Kachala M, Svergun DI (2015) Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ* 2:207–217.
32. Franke D, Svergun DI (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J Appl Crystallogr* 42:342–346.
33. Petoukhov MV, et al. (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J Appl Crystallogr* 45:342–350.
34. 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.
35. Xu X, Pérébaskine N, Minvielle-Sébastia L, Fribourg S, Mackereth CD (2015) Chemical shift assignments of a new folded domain from yeast Pcf11. *Biomol NMR Assign* 9: 421–425.
36. 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.
37. 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.
38. Kubicek K, et al. (2012) Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. *Genes Dev* 26:1891–1896.
39. 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.
40. Svergun DI (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* 76:2879–2886.
41. 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.
42. Svergun D, Barberato C, Koch MHJ (1995) CRYSOL–A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr* 28:768–773.
43. Kozin MB, Svergun DI (2001) Automated matching of high- and low-resolution structural models. *J Appl Crystallogr* 34:33–41.
44. Kuzmic P (2009) DynaFit–A software package for enzymology. *Methods Enzymol* 467: 247–280.
45. Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* 66:125–132.
46. Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* 62:72–82.
47. Winn MD, et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67:235–242.
48. Adams PD, et al. (2011) The Phenix software for automated determination of macromolecular structures. *Methods* 55:94–106.
49. McCoy AJ, Storoni LC, Read RJ (2004) Simple algorithm for a maximum-likelihood SAD function. *Acta Crystallogr D Biol Crystallogr* 60:1220–1228.
50. Grosse-Kunstleve RW, Adams PD (2003) Substructure search procedures for macromolecular structures. *Acta Crystallogr D Biol Crystallogr* 59:1966–1973.
51. Terwilliger T (2004) SOLVE and RESOLVE: Automated structure solution, density modification and model building. *J Synchrotron Radiat* 11:49–52.
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.
53. Cowtan K (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* 62:1002–1011.
54. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126–2132.
55. McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Crystallogr* 40: 658–674.
56. Afonine PV, et al. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* 68:352–367.
57. Pettersen EF, et al. (2004) UCSF Chimera–A visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612.
58. Walshaw J, Woolfson DN (2001) Socket: A program for identifying and analysing coiled-coil motifs within protein structures. *J Mol Biol* 307:1427–1450.
59. Larkin MA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948.