Structure of Ctk3, a subunit of the RNA polymerase II CTD kinase complex, reveals a noncanonical CTD-interacting domain fold

Wolfgang Mühlbacher,[1] Andreas Mayer,[2] Mai Sun,[2] Michael Remmert,[2] Alan C. M. Cheung,[2] Jürgen Niesser,[1] Johannes Soeding,[1] and Patrick Cramer[1]*

[1] Max Planck Institute for Biophysical Chemistry, Department of Molecular Biology, Am Fassberg 11, Göttingen, 37077, Germany

[2] Gene Center Munich and Department of Biochemistry, Ludwig-Maximilians-Universität München, Feodor-Lynen-Str. 25, Munich, 81377, Germany

ABSTRACT

CTDK-I is a yeast kinase complex that phosphorylates the C-terminal repeat domain (CTD) of RNA polymerase II (Pol II) to promote transcription elongation. CTDK-I contains the cyclin-dependent kinase Ctk1 (homologous to human CDK9/CDK12), the cyclin Ctk2 (human cyclin K), and the yeast-specific subunit Ctk3, which is required for CTDK-I stability and activity. Here we predict that Ctk3 consists of a N-terminal CTD-interacting domain (CID) and a C-terminal three-helix bundle domain. We determine the X-ray crystal structure of the N-terminal domain of the Ctk3 homologue Lsg1 from the fission yeast *Schizosaccharomyces pombe* at 2.0 Å resolution. The structure reveals eight helices arranged into a right-handed super-helical fold that resembles the CID domain present in transcription termination factors Pcf11, Nrd1, and Rtt103. Ctk3 however shows different surface properties and no binding to CTD peptides. Together with the known structure of Ctk1 and Ctk2 homologues, our results lead to a molecular framework for analyzing the structure and function of the CTDK-I complex.

INTRODUCTION

The C-terminal repeat domain (CTD) of the largest subunit of RNA polymerase II (Pol II) from the yeast *Saccharomyces cerevisiae* consists of 26 heptapeptide repeats with the consensus sequence Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7.[1,2] The CTD serves as a binding platform for various factors during transcription, including pre-mRNA processing factors. During the transcription cycle, the CTD changes its phosphorylation pattern and this alters its binding affinity to factors.[3–6] Phosphorylation at position Ser5 is associated with pre-mRNA capping in early transcription elongation complexes.[7] Ser2 phosphorylation has been implicated in both elongation and termination events. The CTD residues Tyr1, Thr4, and Ser7 can also be phosphorylated.[8–11]

CTD phosphorylation is accomplished by four different cyclin-dependent kinases (CDKs) in yeast, namely Kin28, Srb10, Bur1, and Ctk1.[5] The Kin28 kinase and its human counterpart CDK7 are subunits of the initiation factor TFIIH and phosphorylate the CTD at position Ser5.[12–14] The Srb10 kinase associates with cyclin Srb11 and resides within the Mediator coactivator complex. The Srb10-Srb11 pair phosphorylates both Ser2 and Ser5 residues and is related to the mammalian pair CDK8-cyclin C.[15,16] The CTD kinases Bur1 and Ctk1 are Ser2 kinases and both share homology with mammalian CDK9, a subunit of positive transcription elongation factor b (P-TEFb), which induces productive

Figure 1

CTDK-I domains and conservation of Ctk3 N-terminal domain. **A**: Domain architecture of the three *S. pombe* CTDK-I subunits Ctk1 (Lsk1), Ctk2 (Lsc1), Ctk3 (Lsg1). **B**: Structural alignment of *S. pombe* Ctk3 (Lsg1) N-terminal domain (1–140) to known CID domains. Conserved residues are colored yellow, invariant residues green. Sequences were ordered from highest (top) to lowest conservation, relative to *S. pombe* (*Sp, Schizosaccharomyces pombe; Sc, Saccharomyces. cerevisiae; Hs, Homo sapiens*). α-Helices are indicated above the alignment. Residues forming the hydrophobic core are depicted with black triangles. **C**: Alignment of Ctk3 N-terminal domains in different yeast species (*Pb, Paracoccidioides brasiliensis; Kp, Komagataella pastoris; Ca, Candida albicans*). Alignments were performed by ClustalW2 and the results were visualized by ESPript 3.0, using default values and "%Equivalent" as similarities depiction parameters.[55,56] [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

elongation.[17–19] Ctk1 is the main Ser2 kinase in yeast, whereas Bur1 phosphorylates both Ser2 and the elongation factor Spt4/5.[20–22] Bur1 also plays a role in histone modification.[23] The transition from transcription initiation to elongation requires, in addition to Ser2 phosphorylation, dephosphorylation of Ser5 residues by Rtr1 and Ssu72.[24,25] Bur1 activity is controlled by cyclin Bur2.[19,26] Yeast Ctk1 and Bur1 kinases appear to be orthologues to metazoan CDK12 and CDK9, respectively.[27]
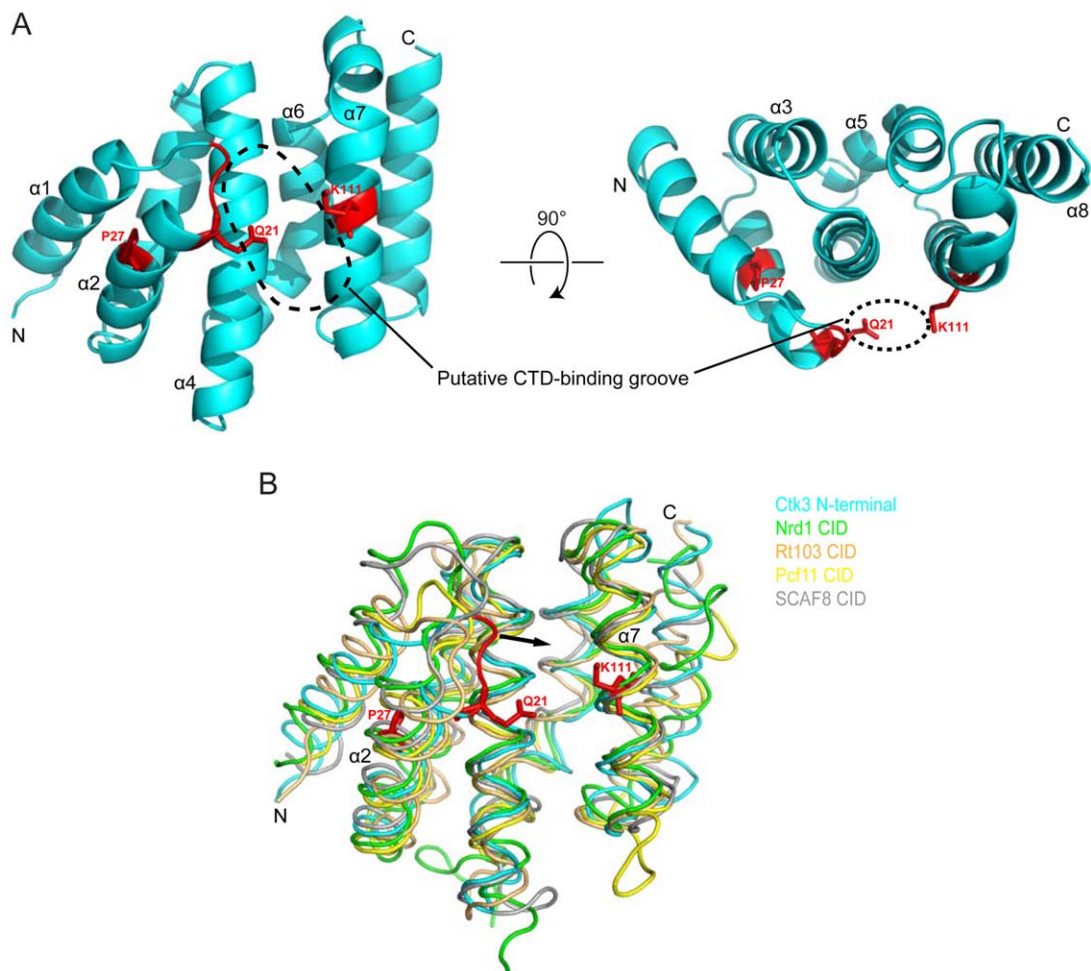
Figure 2

Crystal structure of the Ctk3 N-terminal domain at 2.0 Å resolution. **A:** Two views of a ribbon representation of the Ctk3 N-terminal domain, related by a 90° rotation around a horizontal axis. Residues that partially occupy the space of the surface region corresponding to the CTD-binding groove in canonical CID domains are depicted in red. The proline on position 27, which bends helix α2, is also colored red. The CTD-binding groove is indicated with black dashed lines and is corresponding to the CID in *Sc* Pcf11.[38] **B:** Superposition of known CID structures onto the Ctk3 N-terminal domain reveals its similarity to the CID fold. The key structural difference between the Ctk3 CID-like domain and the canonical CIDs is shown in red (part of helix α2 in Ctk3). The N- and C-termini of the protein domains are indicated.

Ctk1 (also known as Lsk1 in *S. pombe*) associates with its cyclin partner Ctk2 (*S. pombe* Lsc1) and a third subunit, Ctk3 (*S. pombe* Lsg1), to form the CTD kinase I (CTDK-I) complex.[28–32] This trimeric structure is unique amongst CDK complexes.[29,33] *S. cerevisiae* Ctk3 and *S. pombe* Lsg1 share 24% amino acid sequence identity, and associate with Ctk1/Ctk2 and *S. pombe* Lsk1/Lsc1, respectively.[30–32,34,35] It was shown that Lsg1 resembles the function of the Ctk3 subunit of CTDK-I in *S. cerevisiae* despite the low sequence identity.[32] Throughout this work, we refer to *S. pombe* Lsg1 as Ctk3, to be consistent with most literature. In *S. cerevisiae*, the activity of Ctk1 and Ctk2 are strongly dependent on the binding to Ctk3.[28,36] The Ctk3 C-terminal region is involved in the stabilization of the Ctk2-Ctk3 heterodimer and CTDK-I function.[28] Recruitment of Ctk1

in vivo relies to some extent on the completion of the pre-mRNA 5'-cap structure,[37] but there are additional, unknown mechanisms of CTDK-I recruitment. Ctk3 may function in CTDK-I recruitment, although Ctk3 does not have counterparts in metazoan.[29,33]

Here we investigate the structure of Ctk3. We predict a possible CTD-binding domain (CID) in the N-terminal region of Ctk3. The CID fold consists of eight α-helices in a right-handed superhelical arrangement and was thus far discovered in the three yeast proteins Pcf11, Nrd1, and Rtt103.[38–40] We use X-ray crystallography to show that Ctk3 indeed contains a CID fold. However, unlike other CID domains, this domain does not bind the Pol II CTD. We therefore refer to this domain as "CID-like domain." In addition, we predict a three-helix bundle in the C-terminal region of Ctk3.

Table I
Crystallographic Diffraction Data and Refinement Statistics

Data processing			
Space group		P4$_3$2$_1$2	
Unit cell dimensions (Å)		$a = b = 51.3$, $c = 119.1$	
	Peak	Inflection	Remote
Wavelength (Å)	0.97964	0.98012	0.97197
Resolution (Å)	47.1–2.00 (2.05–2.00)[a]	47.2–2.50 (2.57–2.50)	47.2–2.15 (2.21–2.15)
No. of observed reflections	157,934	80,837	126,934
No. of unique reflections	20,524	10,523	16,590
Completeness (%)	100 (99.9)	100 (100)	100 (100)
$I/\sigma(I)$	20.9 (4.9)	17.1 (5.1)	22.1 (5.0)
R_{merge} (%)	8.2 (32.8)	10.4 (32.9)	7.8 (32.4)
R_{meas} (%)	8.3 (48.8)	11.0 (46.7)	7.8 (47.4)
CC$_{(1/2)}$[d] (%)	99.9 (93.8)	99.9 (93.3)	99.9 (92.9)
Wilson B value (Å^2)	27.8	33.3	31.4
Refinement			
Resolution range (Å)		47.12–2.0 (2.05–2.0)[a]	
No. of protein atoms		1194	
Amino acid residues		140	
Water molecules/ionic molecules (EPE ligands[b])		67/2	
R-factor/free R-factor (%)		19.2 (R-work + test data set)/24.10	
B-factor (averaged)/protein/ligands/solvent (Å^2)		23.25/64.50/32.70	
Ramachandran plot: Favored/allowed/disallowed[c] (%)		99.3/0.7/0.0	
RMSD bond lengths/bond angles		0.008 Å/1.047°	

[a]Values in parentheses refer to the highest resolution shell.
[b]4-(2-hydroxyethyl)−1-piperazineethanesulfonic acid.
[c]As calculated using MolProbity.[58]
[d]CC1/2 = percentage of correlation between intensities from random half-datasets.[59]

MATERIALS AND METHODS

Sample preparation

DNA constructs of *S. pombe* full-length Ctk3 (residues 1–218) and the Ctk3 N-terminal domain (residues 1-140) were synthesized (Mr. Gene GmbH) and cloned into pET28b+ expression vector (Novagen) resulting in C-terminal hexahistidine tags. Both, Ctk3 (1-218) and Ctk3 (1-140) protein variants were expressed overnight at 18°C in *Escherichia coli* (*E. coli*) BL21 (DE3) RIL cells (Stratagene). *E. coli* strain B834 (DE3) pLsyS (Novagen) was used for selenomethionine (SeMet) labeling. Cells were harvested and resuspended in 50 m*M* Tris-HCl pH 8.0, 300 m*M* NaCl and 2 m*M* DTT, followed by sonication. The resulting slurry was cleared by centrifugation. The cleared lysate was subjected to affinity chromatography on a Ni-NTA column (Qiagen) before dialyzing against 50 m*M* Tris-HCl pH 8.0, 50 m*M* NaCl, and 1 m*M* DTT to reduce the high imidazole concentration. The hexahistidine tag was removed by thrombin cleavage at 4°C overnight. The protein variants were further purified by anion exchange chromatography (MonoQ column, GE healthcare life science). After size exclusion chromatography (Superpose-12 column, GE healthcare life science) in gelfiltration buffer (50 m*M* HEPES pH 8.0, 50 m*M* NaCl, 1 mM DTT), the pure protein was concentrated to 14.5 mg mL^{-1}. *S. cerevisiae* Ctk3 could not be expressed in *E. coli* in soluble form.

Crystal structure determination

Crystals for the Ctk3 N-terminal domain variant (residues 1–140) were grown at 4°C using hanging-drop vapor diffusion. The reservoir solution contained 26% PEG 6000, 100 m*M* citric acid pH 4.0, 0.8 *M* lithium chloride and 5 m*M* Tris(2-carboxyethyl)phosphin (TCEP). Grown crystals were transferred to reservoir buffer containing 10% PEG400. Diffraction data were collected at the Swiss Light Source (SLS) in Villigen, Switzerland. Data were processed by XDS and scaled using XSCALE.[41] The crystal structure was solved by multiwavelength anomalous diffraction (MAD) from SeMet-labeled crystals using SOLVE.[42] Density modification was carried out with RESOLVE.[42] An initial model was automatically built with ARP/wARP.[43] Manual model building was carried out in Coot.[44] The peak dataset was used for model refinement by PHENIX.[45] Individual isotropic B-factors and bulk solvent correction were used to generate a final model with a free *R*-factor of 24.1% (see updated table and PDB entry [5CE7]) at 2.0 Å resolution.

Peptide interaction analysis

We measured the protein−peptide interactions by fluorescence anisotropy. The synthetic CTD peptides were labelled by N-terminal aminocaproic-linked fluorescein. Changes in fluorescence anisotropy of the peptide solution were measured by titration of Ctk3 N-terminal domain or full-length Ctk3 (FluoroMaxP, HORIBA). All

peptides were dissolved in 20 mM HEPES pH 8.0, 10 mM NaCl, and 5 mM DDT, and adjusted to a concentration of 0.4 mM. Ctk3 N-terminal domain and full-length Ctk3 proteins were dissolved in gelfiltration buffer. The FluoroMaxP analyzer was calibrated at 20°C with gelfiltration buffer and 1 μL peptide in a quartz cuvette (0.4 mM). Before analysis, solutions were mixed by magnetic stirring for 1 min and incubated for an additional minute. The protein solution was titrated to the peptide solution in steps of 20 μL, and three measurements were recorded to enable calculation of an arithmetic average. The excitation wavelength was set to 495 nm (slit width = 2) and the emission wavelength to 520 nm (slit width = 1).

RESULTS

Prediction of a CID domain in Ctk3

We tried to predict a possible structure and function of Ctk3 by using HHblits,[46] an iterative sequence search tool that represents both query and database sequences by profile hidden Markov models (HMMs). Such alignment methods are the most sensitive class of sequence search methods and the best choice for structure prediction and 3D homology modelling. We started with the protein sequence of Ctk3 and generated a profile HMM by performing two iterations of HHblits against the UniProt20 database, a clustered profile HMM database with 20% maximum pairwise sequence identity based on the UniProt sequence database (www.uniprot.org). The resulting profile HMM was then used for a final search against the PDB70 database, a clustered version of the protein data bank (PDB) filtered to 70% maximum pairwise sequence identity.

In this search, the best matches to Ctk3 were the three known CIDs present in the yeast genome. These reside in the transcription termination factors Rtt103, Pcf11, and Nrd1, and gave rise to predicted probabilities of 98%, 94%, and 93%, respectively. The matched alignment covered the entire CID, except for a weakly conserved C-terminal helix where the confidence values were lower. Based on these three alignments we generated a structural model of Ctk3 with the use of the MODELLER software.[47] The Ctk3 model showed a high conservation in the N-terminal five helices of the CID with an insertion between the first and second helix of approximately 20 amino acid residues in length. The detected similarities strongly suggested that Ctk3 contains a CID, thus apparently representing a forth CID-containing protein in the yeast genome.

Crystal structure analysis of Ctk3 N-terminal domain

To clarify whether Ctk3 indeed contains the predicted CID fold, we prepared the putative CID-containing region of Ctk3 from *S. pombe* (Ctk3 (1-140), Fig. 1) in
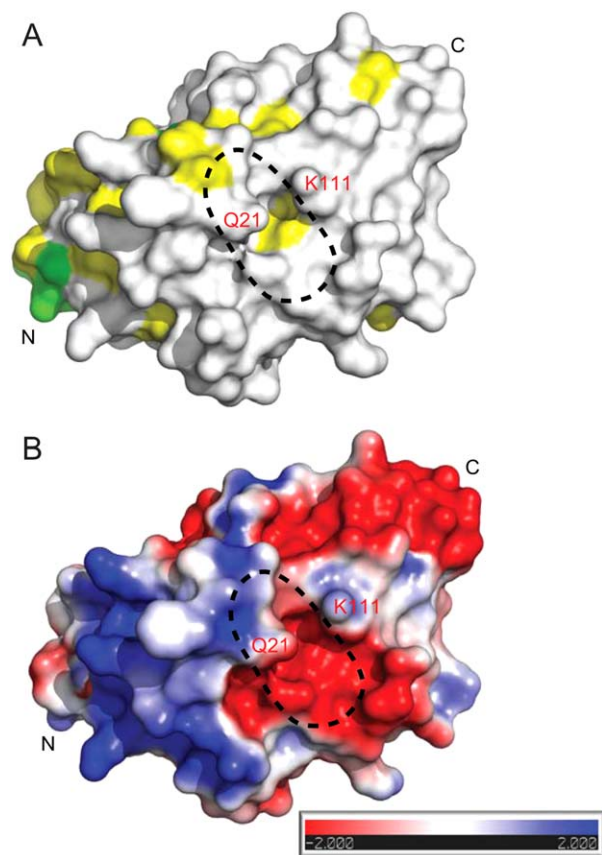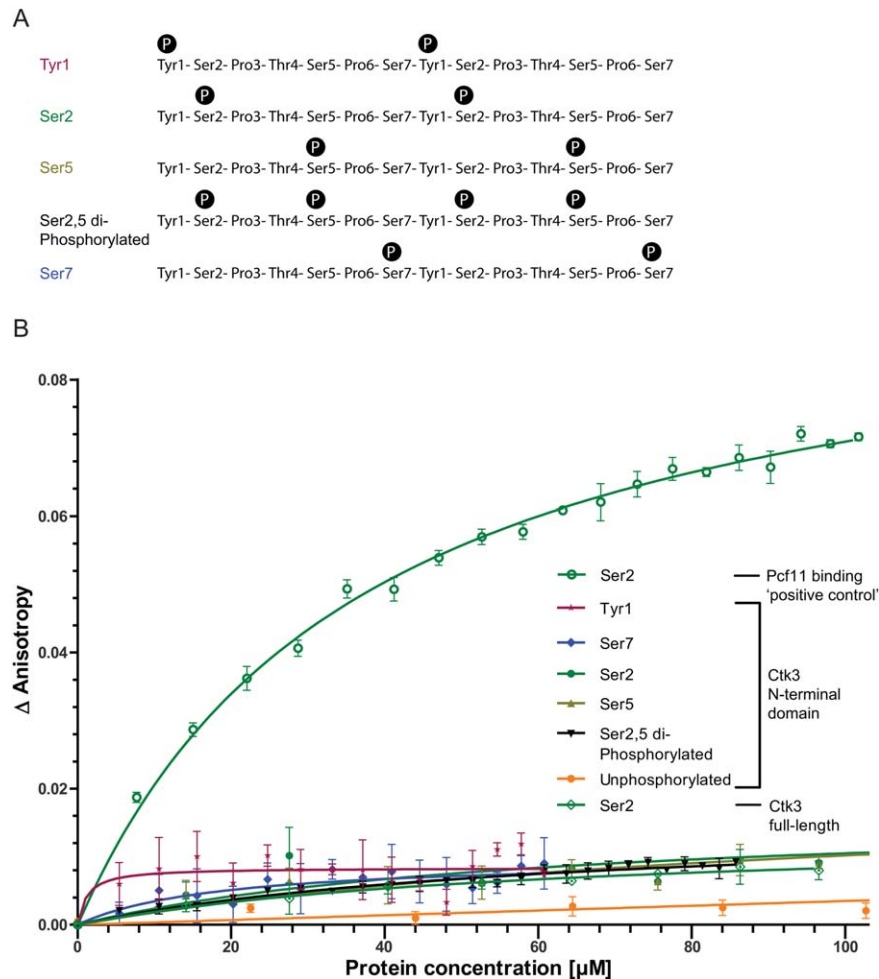

Figure 3

Surface properties of the Ctk3 N-terminal domain structure. **A**: Lack of a conserved CTD-binding groove in the Ctk3 N-terminal domain. The area corresponding to the CTD-binding groove in the CID domain of Pcf11 is indicated by a black dashed line.[38] Surface model generated by ESPript 3.0[56] and colored according to conservation as in Figure 1(C). The alignment in Figure 1(C) was used as an input for ESPript. Residues that partially occupy the space of the surface region corresponding to the CTD-binding groove in CID domains are labeled in red. **B**: Electrostatic surface potentials of Ctk3 N-terminal domain. Positive and negative charges are in blue and red, respectively. Surface potentials were calculated with APBS.[48] [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

recombinant form after overexpression in *E. coli* (Materials and Methods). The Ctk3 N-terminal domain was purified and crystallized by vapor diffusion (Materials and Methods). X-ray diffraction data were collected at the Swiss Light Source to a resolution of 2.0 Å (Table I). The structure was determined by selenomethionine (SeMet) incorporation and multiwavelength anomalous diffraction (MAD) phasing, and the resulting model was refined to a free R-factor of 24.1% (see updated table and PDB entry [5CE7]) and showed very good stereochemistry (Materials and Methods). The overall fold of the Ctk3 N-terminal domain consists of eight α-helices in a right-handed superhelical arrangement [Fig. 2(A)]. As predicted, the fold closely resembles the known CID domains in the *S. cerevisiae* proteins Rtt103, Pcf11, and

A

Tyr1   Tyr1- Ser2- Pro3- Thr4- Ser5- Pro6- Ser7- Tyr1- Ser2- Pro3- Thr4- Ser5- Pro6- Ser7

Ser2   Tyr1- Ser2- Pro3- Thr4- Ser5- Pro6- Ser7- Tyr1- Ser2- Pro3- Thr4- Ser5- Pro6- Ser7

Ser5   Tyr1- Ser2- Pro3- Thr4- Ser5- Pro6- Ser7- Tyr1- Ser2- Pro3- Thr4- Ser5- Pro6- Ser7

Ser2,5 di-Phosphorylated   Tyr1- Ser2- Pro3- Thr4- Ser5- Pro6- Ser7- Tyr1- Ser2- Pro3- Thr4- Ser5- Pro6- Ser7

Ser7   Tyr1- Ser2- Pro3- Thr4- Ser5- Pro6- Ser7- Tyr1- Ser2- Pro3- Thr4- Ser5- Pro6- Ser7

B

Figure 4

Ctk3 N-terminal domain does not bind CTD phosphopeptides. **A**: CTD-derived diheptad repeat peptides with phosphorylation positions indicated. From top to bottom, sequences of Tyr1-, Ser2-, Ser5-, Ser2,5-, and Ser7-phosphorylated peptides are given. **B**: Fluorescence anisotropy titration curves, using purified Ctk3 N-terminal domain (residues 1–140) and fluorescently labeled peptides shown in panel A. All measurements based on technical triplicate data. Plotted are the mean and standard deviation. As a positive control, we monitored binding of purified Pcf11-CID to Ser2-phosphorylated CTD peptide (green[57]). Ser2-phosphorylated binding was additionally measured with full-length Ctk3 (residues 1–218). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Nrd1, and in the human protein SCAF8 [Fig. 2(B)]. Superposition of the known CID structures revealed a difference in helix α2 of Ctk3 (residue 19–21), which is bent at its residue Pro27 toward helix α4 that flanks the proposed CTD-binding region of Ctk3 [Fig. 2(B)].

The Ctk3 N-terminal domain has a noncanonical surface

To analyze whether the CID fold in Ctk3 also shares surface properties with canonical CID domains, we analyzed the conservation and electrostatic properties of the domain surface. We first colored the molecular surface of our structure according to conservation of amino acid residues over species [Fig. 3(A), alignment Fig. 1(C)]. Comparison with the surface of CID domains in Rtt103,

Pcf11, and Nrd1 revealed that the Ctk3 N-terminal domain has a distinct surface conservation. Whereas canonical CID domains show a very high conservation of the CTD-binding groove between helices α2, α4, and α7, Ctk3 only shows weak and partial conservation in this region [Fig. 3(A)]. Moreover, Ctk3 residues Q21 and K111 partially obstruct the region corresponding to the CTD-binding groove in CID domains [Fig. 2(A)].

The Ctk3 N-terminal domain also differs from canonical CID domains with respect to its surface charge distribution. To determine the charge distribution on the surface of the structure we used the APBS Tool via PyMOL [Fig. 3(B)].[48] For the canonical CID structures in Rtt103, Pcf11, and Nrd1, the conserved CTD-binding groove is positively charged, as required for binding to the negatively charged, phopshorylated CTD. In contrast,
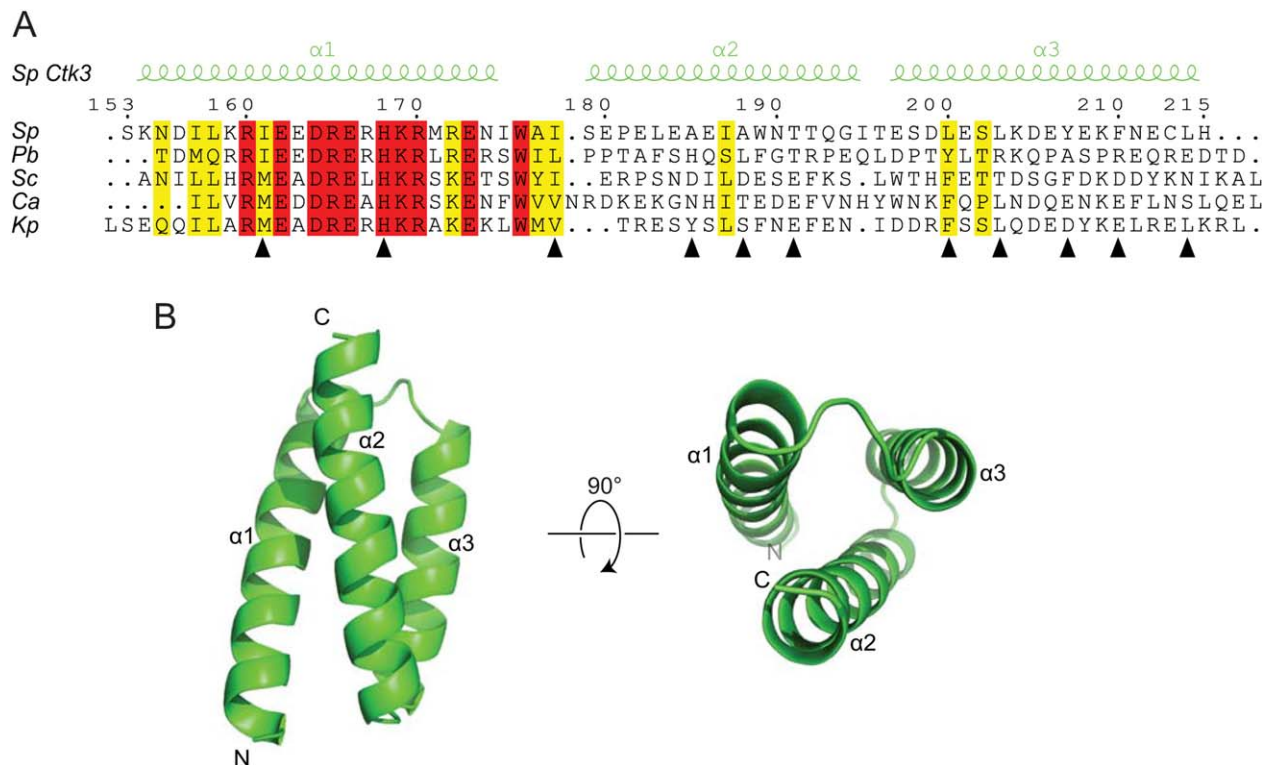
Figure 5

The Ctk3 C-terminal region contains a predicted helical bundle domain. **A**: Sequence alignment and secondary structure prediction of Ctk3 C-terminal region comparing different yeast species. Conserved and invariant residues colored yellow and red, respectively. Sequences ordered from highest conservation (on top) to lowest, relative to *S. pombe* Ctk3 C-terminal domain (153–215). (*Sp, Schizosaccharomyces pombe*; *Pb, Paracoccidioides brasiliensis*; *Sc, Saccharomyces cerevisiae*; *Ca, Candida albicans*; *Kp, Komagataella pastoris*). α-Helices are indicated above the alignment. Residues forming the hydrophobic core are depicted with black triangles. **B**: Two views of the predicted Ctk3 C-terminal domain model obtained with I-TASSER.[52] The N- and C-termini of the protein domain are indicated. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

the corresponding region of the Ctk3 N-terminal domain is mainly positively charged. Taken together, analysis of the surface properties of the Ctk3 domain revealed that the putative CTD-binding groove was not conserved and differently charged, arguing against a CTD-binding function.

The Ctk3 N-terminal domain does not bind CTD-derived peptides

Structural analysis of Ctk3 revealed the fold of a CID domain, but also surface features that clearly differ from canonical CID domains and therefore question whether Ctk3 binds directly the CTD. To test CTD binding, we used a fluorescence-based phosphopeptide interaction assay in solution. We monitored changes in fluorescence anisotropy of fluorescently labeled CTD peptides upon addition of increasing amounts of purified Ctk3 N-terminal domain. The peptides were based on a di-heptad sequence [Fig. 4(A)], which represents the functional unit of the Pol II CTD[49] and comprise different phosphorylation sites, mimicking various phosphorylation states of the CTD that are adopted during the transcription cycle.

We tested binding of CTD peptides phosphorylated at positions Tyr1, Ser2, Ser5, or Ser7, or phosphorylated at both Ser2 and Ser5, and also used a non-phosphorylated di-heptad peptide as a control [Fig. 4(A)]. We did not observe any binding of the Ctk3 N-terminal domain or full-length Ctk3 to any of the CTD-based peptides under the conditions tested [Fig. 4(B)]. As a positive control we reproduced binding of purified *S. pombe* Pcf11-CID to one of the Ser2-phosphorylated peptide, which demonstrated that the assay was working very well under our conditions. The data obtained from the positive control titration could be fit to the Hill equation as expected [Fig. 4(B)]. Taken together, we were unable to detect any binding of the Ctk3 N-terminal domain to any of the CTD-based peptides, strongly indicating that the domain does not directly bind the CTD *in vitro*.

Ctk3 contains a highly conserved C-terminal bundle domain

In order to gain structural insights also into the C-terminal region of Ctk3, we returned to bioinformatic
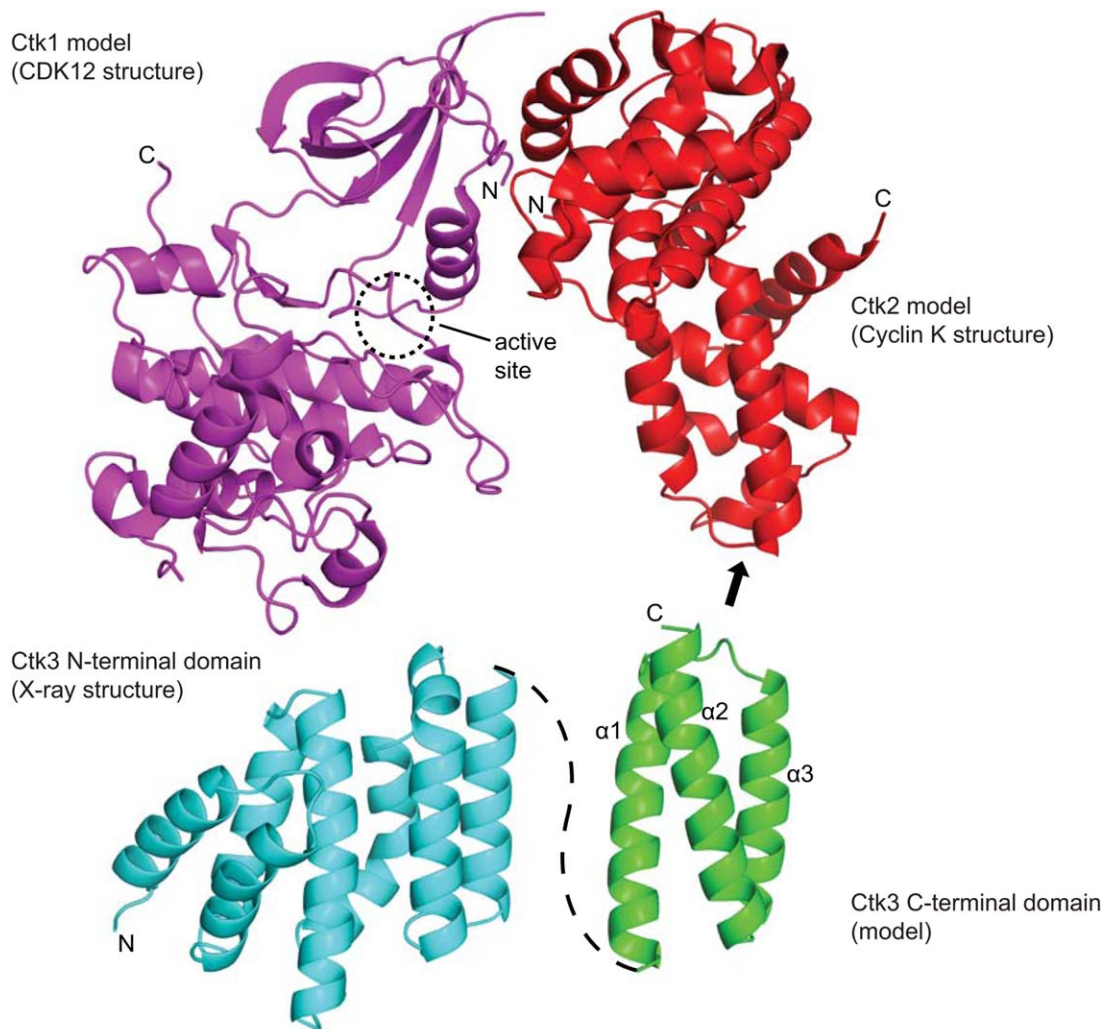
Figure 6

Overview of CTDK-I architecture. Relative size of the structure of mammalian Ctk1-Ctk2 counterpart CDK12–Cyclin K (PDB-code 4UN0) and *S. pombe* Ctk3 N-terminal domain structure (this work) and predicted C-terminal domain (this work). Arrow indicate putative interaction of the C-terminal domains with Ctk2 (compare text). The active center of the CDK12 kinase structure is indicated with a black dashed circle. The N- and C-termini of the proteins are labeled. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

analysis. *S. pombe* Ctk3 residues 153-215 were annotated as a possible conserved domain in the PFAM database (http://pfam.xfam.org/) . This region shows even higher sequence conservation than the N-terminal domain of Ctk3 [Fig. 5(A)]. We used the prediction algorithm PSIPRED[50] to assign potential secondary structure to this region. This revealed three helical stretches in this region that we named putative helices α1, α2, and α3 [Fig. 5(A)]. Throughout different yeast species, helices α2 and α3 were only weakly conserved, whereas helix α1 was highly conserved, showing 163 yeast species comprising helix α1 (Ctk3 residues 153–183) in a BLAST search (conservative expect threshold of 1e-6).[51] We generated a three-dimensional model for the Ctk3 C-terminal region with the I-TASSER software,[52] which uses

structural templates from the Protein Data Bank. The model with the highest score was a three-helix bundle [Fig. 5(B)]. Analysis of the model revealed a conserved putative hydrophobic core, supporting the accuracy of the prediction, and a conserved surface patch on helix α1 that may be involved in interactions with other regions of CTDK-I.

DISCUSSION

Here we predicted that Ctk3 consists of an N-terminal CTD-interacting domain (CID) and a C-terminal helical bundle domain. X-ray crystallography confirmed the N-terminal CID fold despite the low sequence conservation,

consistent with the general observation that the three-dimensional structure of proteins is more conserved than their sequence. The Ctk3 N-terminal domain however differs from canonical CID domains in its surface conservation and charge distribution. In particular, the putative CTD-binding groove between helices α2, α4, and α7 is not conserved and lacks positively charged residues. Consistent with this, the N-terminal domain of Ctk3 failed to bind CTD phosphopeptides in solution. The CID fold resembles the fold of VHS (Vsp27p/Hrs/STAM) domains,[53] as detected by DALI.[54] Thus the N-terminal domain in Ctk3 could have arisen from a CID domain that lost its CTD-binding surface groove, or from a VHS domain of unknown function.

The C-terminal domain of Ctk3 is more conserved than its N-terminal domain, especially at its putative α1 helix. It remains unclear whether the C-terminal domain of Ctk3 associates stably with the N-terminal domain or whether it forms a distinct domain connected with a short linker. Consistent with a two-domain, flexible structure, our attempts to crystallize full-length Ctk3 failed. The predicted C-terminal domain is essential for cell growth in *S. cerevisiae,* and a truncation that removes most of the C-terminal domain destabilizes the interaction of Ctk3 with Ctk2 *in vitro*, whereas interaction with Ctk1 was maintained.[28] From these results and considerations, a topological model of CTDK-I emerges with Ctk1 and Ctk2 forming a canonical kinase-cyclin pair, as observed for CDK12–Cyclin K[27] and Ctk3 being anchored to Ctk2 via its C-terminal helical bundle domain (Fig. 6). In this model, the newly found N-terminal domain in Ctk3 remains available for other, currently unknown functions.

DATA DEPOSITION

The refined structure and structure factors were deposited in the Protein Data Bank under accession code 5CE7.

ACKNOWLEDGMENTS

REFERENCES

1. Eick D, Geyer M. The RNA polymerase II carboxy-terminal domain (CTD) code. Chem Rev 2013;113:8456–8490.
2. Napolitano G, Lania L, Majello B. RNA polymerase II CTD modifications: how many tales from a single tail. J Cell Physiol 2014;229:538–544.
3. Corden JL. Transcription. Seven ups the code. Science 2007;318:1735–1736.
4. Buratowski S. Progression through the RNA polymerase II CTD cycle. Mol Cell 2009;36:541–546.
5. Heidemann M, Hintermair C, Voss K, Eick D. Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. Biochim Biophys Acta 2013;1829:55–62.
6. Egloff S, Dienstbier M, Murphy S. Updating the RNA polymerase CTD code: adding gene-specific layers. Trends Genet 2012; 28:333–341.
7. Schroeder SC, Schwer B, Shuman S, Bentley D. Dynamic association of capping enzymes with transcribing RNA polymerase II. Genes Dev 2000;14:2435–2440.
8. Cho EJ, Kobor MS, Kim M, Greenblatt J, Buratowski S. Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. Genes Dev 2001;15:3319–3329.
9. Jona G, Wittschieben BO, Svejstrup JQ, Gileadi O. Involvement of yeast carboxy-terminal domain kinase I (CTDK-I) in transcription elongation in vivo. Gene 2001;267:31–36.
10. Lee JM, Greenleaf AL. Modulation of RNA polymerase II elongation efficiency by C-terminal heptapeptide repeat domain kinase I. J Biol Chem 1997;272:10990–10993.
11. Ho CK, Shuman S. Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. Mol Cell 1999;3:405–411.
12. Cismowski MJ, Laff GM, Solomon MJ, Reed SI. KIN28 encodes a C-terminal domain kinase that controls mRNA transcription in Saccharomyces cerevisiae but lacks cyclin-dependent kinase-activating kinase (CAK) activity. Mol Cell Biol 1995;15:2983–2992.
13. Feaver WJ, Svejstrup JQ, Henry NL, Kornberg RD. Relationship of CDK-activating kinase and RNA polymerase II CTD kinase TFIIH/TFIIK. Cell 1994;79:1103–1109.
14. Keogh MC, Cho EJ, Podolny V, Buratowski S. Kin28 is found within TFIIH and a Kin28-Ccl1-Tfb3 trimer complex with differential sensitivities to T-loop phosphorylation. Mol Cell Biol 2002;22:1288–1297.
15. Borggrefe T, Davis R, Erdjument-Bromage H, Tempst P, Kornberg RD. A complex of the Srb8, −9, −10, and −11 transcriptional regulatory proteins from yeast. J Biol Chem 2002;277:44202–44207.
16. Hengartner CJ, Myer VE, Liao SM, Wilson CJ, Koh SS, Young RA. Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases. Mol Cell 1998;2:43–53.
17. Marshall NF, Price DH. Purification of P-TEFb, a transcription factor required for the transition into productive elongation. J Biol Chem 1995;270:12335–12338.
18. Cheng B, Price DH. Properties of RNA polymerase II elongation complexes before and after the P-TEFb-mediated transition into productive elongation. J Biol Chem 2007;282:21901–21912.
19. Keogh MC, Podolny V, Buratowski S. Bur1 kinase is required for efficient transcription elongation by RNA polymerase II. Mol Cell Biol 2003;23:7005–7018.
20. Zhou K, Kuo WH, Fillingham J, Greenblatt JF. Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5. Proc Natl Acad Sci USA 2009;106:6956–6961.
21. Liu Y, Warfield L, Zhang C, Luo J, Allen J, Lang WH, Ranish J, Shokat KM, Hahn S. Phosphorylation of the transcription elongation factor Spt5 by yeast Bur1 kinase stimulates recruitment of the PAF complex. Mol Cell Biol 2009;29:4852–4863.
22. Qiu H, Hu C, Hinnebusch AG. Phosphorylation of the Pol II CTD by KIN28 enhances BUR1/BUR2 recruitment and Ser2 CTD phosphorylation near promoters. Mol Cell 2009;33:752–762.
23. Wood A, Schneider J, Dover J, Johnston M, Shilatifard A. The Bur1/Bur2 complex is required for histone H2B monoubiquitination

by Rad6/Bre1 and histone methylation by COMPASS. Mol Cell 2005;20:589–599.

24. Mosley AL, Pattenden SG, Carey M, Venkatesh S, Gilmore JM, Florens L, Workman JL, Washburn MP. Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation. Mol Cell 2009;34:168–178.

25. Krishnamurthy S, He X, Reyes-Reyes M, Moore C, Hampsey M. Ssu72 Is an RNA polymerase II CTD phosphatase. Mol Cell 2004; 14:387–394.

26. Yao S, Neiman A, Prelich G. BUR1 and BUR2 encode a divergent cyclin-dependent kinase-cyclin complex important for transcription in vivo. Mol Cell Biol 2000;20:7080–7087.

27. Bartkowiak B, Liu P, Phatnani HP, Fuda NJ, Cooper JJ, Price DH, Adelman K, Lis JT, Greenleaf AL. CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. Genes Dev 2010;24:2303–2316.

28. Hautbergue G, Goguel V. Activation of the cyclin-dependent kinase CTDK-I requires the heterodimerization of two unstable subunits. J Biol Chem 2001;276:8005–8013.

29. Sterner DE, Lee JM, Hardin SE, Greenleaf AL. The yeast carboxyl-terminal repeat domain kinase CTDK-I is a divergent cyclin-cyclin-dependent kinase complex. Mol Cell Biol 1995; 15:5716–5724.

30. Karagiannis J, Bimbo A, Rajagopalan S, Liu J, Balasubramanian MK. The nuclear kinase Lsk1p positively regulates the septation initiation network and promotes the successful completion of cytokinesis in response to perturbation of the actomyosin ring in Schizosaccharomyces pombe. Mol Biol Cell 2005;16:358–371.

31. Karagiannis J, Balasubramanian MK. A cyclin-dependent kinase that promotes cytokinesis through modulating phosphorylation of the carboxy terminal domain of the RNA Pol II Rpb1p sub-unit. PloS One 2007;2:e433

32. Sukegawa Y, Yamashita A, Yamamoto M. The fission yeast stress-responsive MAPK pathway promotes meiosis via the phosphorylation of Pol II CTD in response to environmental and feedback cues. PLoS Genet 2011;7:e1002387

33. Lee JM, Greenleaf AL. CTD kinase large subunit is encoded by CTK1, a gene required for normal growth of Saccharomyces cerevisiae. Gene Expr 1991;1:149–167.

34. Viladevall L, St Amour CV, Rosebrock A, Schneider S, Zhang C, Allen JJ, Shokat KM, Schwer B, Leatherwood JK, Fisher RP. TFIIH and P-TEFb coordinate transcription with capping enzyme recruitment at specific genes in fission yeast. Mol Cell 2009;33:738–751.

35. Coudreuse D, van Bakel H, Dewez M, Soutourina J, Parnell T, Vandenhaute J, Cairns B, Werner M, Hermand D. A gene-specific requirement of RNA polymerase II CTD phosphorylation for sexual differentiation in S. pombe. Curr Biol 2010;20:1053–1064.

36. Hautbergue G, Goguel V. The yeast C-type cyclin Ctk2p is phosphorylated and rapidly degraded by the ubiquitin-proteasome pathway. Mol Cell Biol 1999;19:2527–2534.

37. Lidschreiber M, Leike K, Cramer P. Cap completion and C-terminal repeat domain kinase recruitment underlie the initiation-elongation transition of RNA polymerase II. Mol Cell Biol 2013;33:3805–3816.

38. Meinhart A, Cramer P. Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. Nature 2004; 430: 223–226.

39. Vasiljeva L, Kim M, Mutschler H, Buratowski S, Meinhart A. The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. Nat Struct Mol Biol 2008;15:795–804.

40. Lunde BM, Reichow SL, Kim M, Suh H, Leeper TC, Yang F, Mutschler H, Buratowski S, Meinhart A, Varani G. Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. Nat Struct Mol Biol 2010;17:1195–1201.

41. Kabsch W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J Appl Crystallography 1993;26:795–800.

42. Terwilliger T. SOLVE and RESOLVE: automated structure solution, density modification and model building. J Synchrotron Radiat 2004;11:49–52.

43. Langer G, Cohen SX, Lamzin VS, Perrakis A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat Protocols 2008;3:1171–1179.

44. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallographica Section D, Biol Crystallography 2004; 60(Part 12 Part 1):2126–2132.

45. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica Section D, Biol Crystallography 2010;66(Part 2):213–221.

46. Remmert M, Biegert A, Hauser A, Soding J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat Methods 2012;9:173–175.

47. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993;234:779–815.

48. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci USA 2001;98:10037–10041.

49. Stiller JW, Cook MS. Functional unit of the RNA polymerase II C-terminal domain lies within heptapeptide pairs. Eukaryot Cell 2004; 3:735–740.

50. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol 1999;292:195–202.

51. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403–410.

52. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. Nat Methods 2014; 12:7–8.

53. Lohi O, Poussu A, Mao Y, Quiocho F, Lehto VP. VHS domain—a longshoreman of vesicle lines. FEBS Lett 2002;513:19–23.

54. Holm L, Rosenstrom P. Dali server: conservation mapping in 3D. Nucleic Acids Res. 2010/05/12 ed., 2010;38:W545–W549.

55. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal W, Clustal X version 2.0. Bioinformatics 2007;23:2947–2948.

56. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res 2014;42(Web Server issue):W320–W324.

57. Mayer A, Heidemann M, Lidschreiber M, Schreieck A, Sun M, Hintermair C, Kremmer E, Eick D, Cramer P. CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. Science 2012;336:1723–1725.

58. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, 3rd Snoeyink J, Richardson JS Richardson DC. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 2007;35(Web Server issue):W375–W383.

59. Karplus PA, Diederichs K. Linking crystallographic model and data quality. Science 2012;336:1030–1033.