LETTER

doi:10.1038/nature09391

Crystal structure of the human symplekin–Ssu72–CTD phosphopeptide complex

Kehui Xiang[1], Takashi Nagaike[1]*, Song Xiang[1]*†, Turgay Kilic[1], Maia M. Beh[1], James L. Manley[1] & Liang Tong[1]

Symplekin (Pta1 in yeast) is a scaffold in the large protein complex that is required for 3′-end cleavage and polyadenylation of eukaryotic messenger RNA precursors (pre-mRNAs)[1–4]; it also participates in transcription initiation and termination by RNA polymerase II (Pol II)[5,6]. Symplekin mediates interactions between many different proteins in this machinery[1,2,7–9], although the molecular basis for its function is not known. Here we report the crystal structure at 2.4 Å resolution of the amino-terminal domain (residues 30–340) of human symplekin in a ternary complex with the Pol II carboxy-terminal domain (CTD) Ser 5 phosphatase Ssu72 (refs 7, 10–17) and a CTD Ser 5 phosphopeptide. The N-terminal domain of symplekin has the ARM or HEAT fold, with seven pairs of anti-parallel α-helices arranged in the shape of an arc. The structure of Ssu72 has some similarity to that of low-molecular-mass phosphotyrosine protein phosphatase[18,19], although Ssu72 has a unique active-site landscape as well as extra structural features at the C terminus that are important for interaction with symplekin. Ssu72 is bound to the concave face of symplekin, and engineered mutations in this interface can abolish interactions between the two proteins. The CTD peptide is bound in the active site of Ssu72, with the pSer 5-Pro 6 peptide bond in the *cis* configuration, which contrasts with all other known CTD peptide conformations[20,21]. Although the active site of Ssu72 is about 25 Å from the interface with symplekin, we found that the symplekin N-terminal domain stimulates Ssu72 CTD phosphatase activity *in vitro*. Furthermore, the N-terminal domain of symplekin inhibits polyadenylation *in vitro*, but only when coupled to transcription. Because catalytically active Ssu72 overcomes this inhibition, our results show a role for mammalian Ssu72 in transcription-coupled pre-mRNA 3′-end processing.

Human symplekin contains 1,274 amino-acid residues (Fig. 1a) and its sequence is well conserved in higher eukaryotes (Supplementary Fig. 1). In comparison, symplekin shares only weak sequence similarity with yeast Pta1 (ref. 1) (Supplementary Fig. 2), and Pta1 lacks the C-terminal 500 residues of symplekin (Fig. 1a). Symplekin and Pta1 do not have any recognizable homology with other proteins. Predictions of secondary structure suggest the presence of an all-helical segment in the N-terminal region of symplekin and Pta1 (Fig. 1a and Supplementary Figs 1 and 2). Recent studies in yeast suggested that the N-terminal segment of Pta1 is important for interaction with Ssu72 (ref. 9). Ssu72 is required for pre-mRNA 3′-end cleavage in yeast[7], although its phosphatase activity is not necessary for this function[13]. The catalytic activity of Ssu72 may instead be important for Pol II transcription and termination and for gene looping[17]. Ssu72 is highly conserved in the eukaryotes (Supplementary Fig. 3), but so far no evidence exists implicating mammalian Ssu72 in 3′-end processing.

To determine the structure of a symplekin–Ssu72–CTD phosphopeptide ternary complex, residues 30–360 of human symplekin and full-length human Ssu72 were overexpressed and purified separately. The two proteins were mixed, with Ssu72 in slight molar excess, and the symplekin–Ssu72 complex was purified by gel-filtration chromatography. This procedure also demonstrated strong interactions between the two human proteins, consistent with observations on their yeast counterparts[9]. The decamer CTD phosphopeptide used in this study, Ser-Tyr 1-Ser 2-Pro 3-Thr 4-pSer 5-Pro 6-Ser 7-Tyr-Ser, where Ser 5 is phosphorylated, contained an entire CTD heptad repeat as well as a serine residue from the previous repeat and Tyr-Ser from the following repeat. To prevent hydrolysis, the active-site nucleophile Cys 12 of Ssu72 was mutated to Ser in the ternary complex. We have also determined the crystal structure of the symplekin–Ssu72(wild-type) binary complex and the structures of the symplekin N-terminal domain alone (for residues 30–395 or 1–395). All the structures are in excellent agreement with the crystallographic data and the expected geometric parameters (Supplementary Table 1).

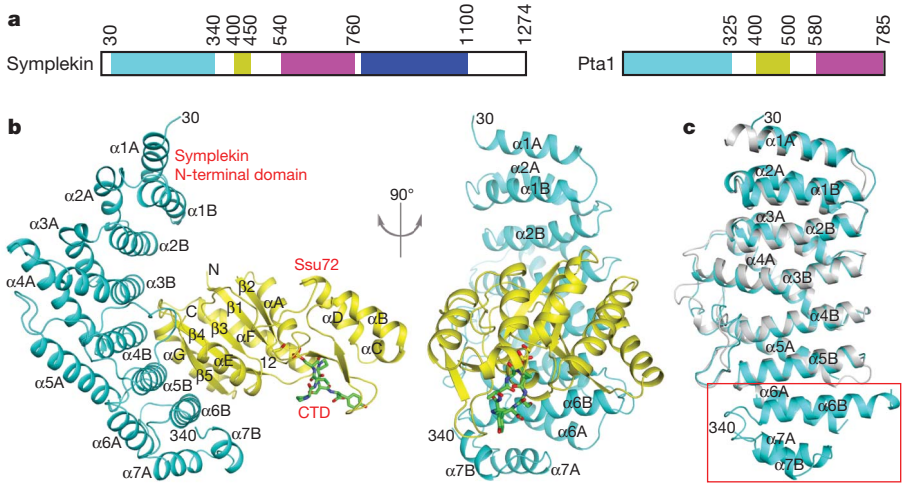

Figure 1 | Structure of the human symplekin–Ssu72–CTD phosphopeptide ternary complex. **a**, Domain organization of human symplekin and *Saccharomyces cerevisiae* Pta1. The domains are suggested by secondary-structure predictions, and the middle region of Pta1 is suggested by functional studies[9]. **b**, Schematic drawing of the structure of human symplekin–Ssu72–CTD phosphopeptide ternary complex, in two views. The N-terminal domain of symplekin is shown in cyan, and Ssu72 in yellow. The CTD phosphopeptide is shown as a stick model (in green for carbon atoms). **c**, Overlay of the structures of the N-terminal domain of human (in cyan) and *Drosophila* (in grey) symplekin[22]. *Drosophila* symplekin lacks the last two pairs of helical repeats (boxed). All structure figures were produced with PyMOL (http://www.pymol.org).

[1]Department of Biological Sciences, Columbia University, New York, New York 10027, USA. †Present address: Key Laboratory of Nutrition and Metabolism, Institute for Nutritional Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China.
*These authors contributed equally to this work.

The structures show that residues 30–340 of symplekin (Symp-N) form seven pairs of antiparallel α-helices, whereas residues 1–29 and 341–395 are disordered (Supplementary Fig. 1). The pairs of helices are arranged in an arc, with the first helix in each pair, the αA helix, being located on the convex face of the arc, and the αB helix on the concave face (Fig. 1b). Most of the loops connecting the helices are short, except for that linking helices α4B and α5A, with 31 residues (Fig. 1b). The overall fold of Symp-N is found in many other proteins, including those with the ARM or HEAT repeats. These structures are often involved in protein–protein interactions, which is consistent with the proposed scaffold function of symplekin.

The structure of the N-terminal domain of *Drosophila* symplekin (residues 22–270) was reported recently[22]. Its overall conformation is similar to that of human Symp-N, with a root mean squared distance of 1.0 Å among their equivalent Cα atoms, although the *Drosophila* symplekin structure lacks the two pairs of helices (α6 and α7) at the C terminus (Fig. 1c). Noting the good sequence conservation of residues in this region (Supplementary Fig. 1), it is likely that these helices are also present in the N-terminal domain of *Drosophila* symplekin. Our studies show that helix α6B is important for interactions with Ssu72 (Fig. 1b; see below).

The structure of Ssu72 contains a central five-stranded β-sheet (β1–β5) that is surrounded by helices on both sides (Fig. 2a and Supplementary Fig. 4). The closest structural homologue of Ssu72 is the low-molecular-mass phosphotyrosine protein phosphatase (Fig. 2b)[18,19], as suggested previously[11,12,14], even though the two proteins share only 16% sequence identity. However, our studies show that Ssu72 possesses three unique structural features compared with this other phosphatase (Fig. 2b), which are formed by highly conserved residues (Supplementary Fig. 3) and have important functions. A small, two-stranded antiparallel β-sheet (β2A and β2B) is located near the active site (Fig. 2a). The αD helix is in a different conformation in Ssu72 and also contributes to phosphopeptide binding. Finally, Ssu72 contains an extra helix (αG) and a β-strand (β5) at the C terminus, which are essential for interactions with symplekin (Fig. 1b).

Our structure of the ternary complex showed that the CTD phosphopeptide, with good electron density for residues Thr 4 to Ser 7 (Fig. 2c), is bound with the peptide bond between pSer 5 and Pro 6 in the *cis* configuration (Fig. 2d). This is in sharp contrast to the conformations of the CTD phosphopeptides observed in other structures, which all have the Pro residue(s) in the *trans* configuration (Supplementary Fig. 5)[20,21]. With the *cis* configuration, the backbone of the phosphopeptide makes a 180° turn at the pSer-Pro residues, whereas the peptide in the *trans* configuration (Supplementary Fig. 5) would clash with Ssu72. Therefore, Ssu72 can only bind and dephosphorylate CTD substrates

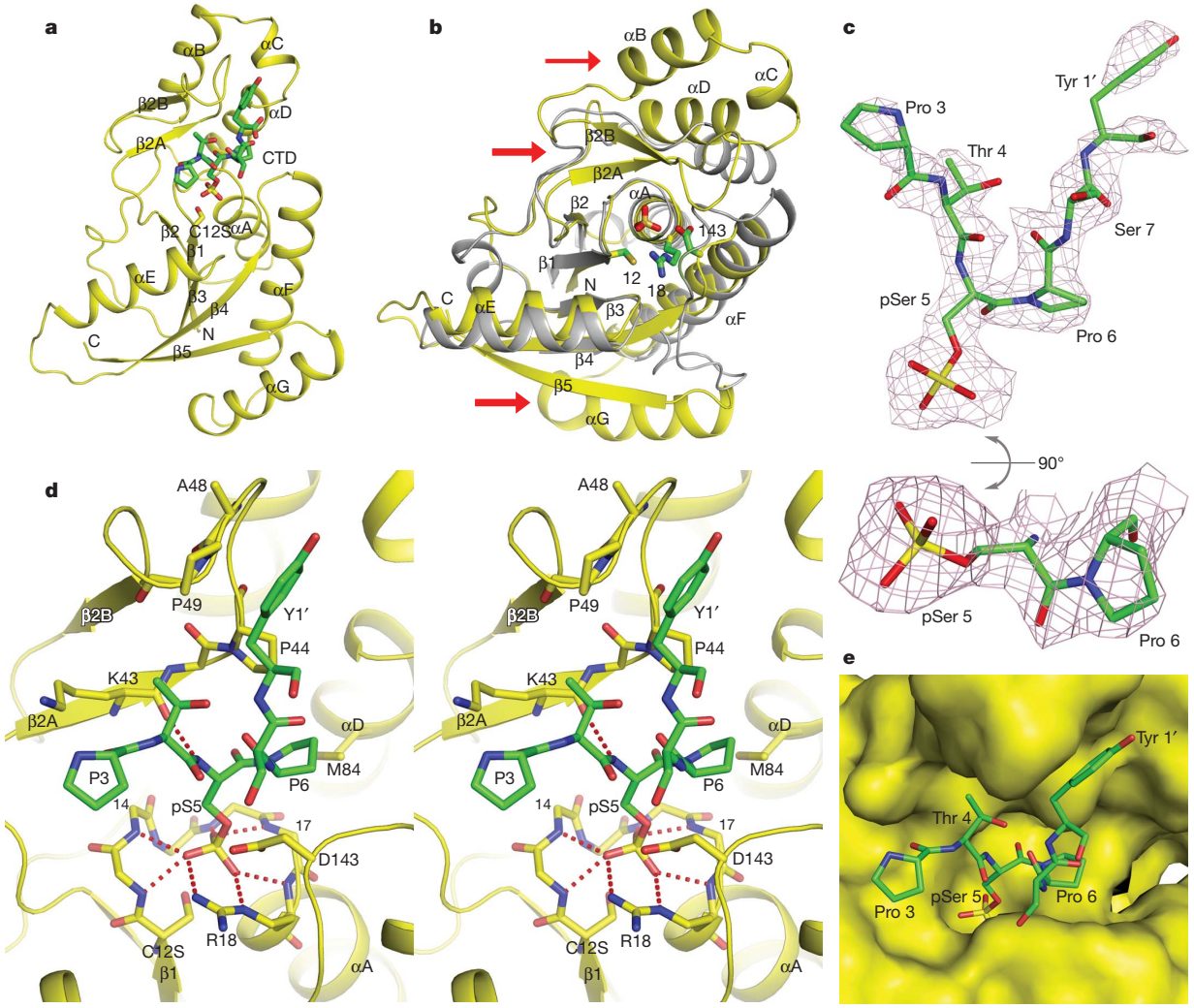

Figure 2 | Recognition of the CTD phosphopeptide by human Ssu72. **a**, Schematic drawing of the structure of human Ssu72–CTD phosphopeptide complex. **b**, Overlay of the structures of Ssu72 (yellow) and low-molecular-mass phosphotyrosine protein phosphatase (grey)[18,19]. Arrows indicate unique structural features in Ssu72. Stereo versions of **a** and **b** are given in Supplementary Fig. 4. **c**, Two views of the omit $F_o - F_c$ electron density at 2.4 Å resolution for the CTD phosphopeptide, contoured at 3σ. **d**, Stereo pair showing detailed interactions between the CTD phosphopeptide and the active site of Ssu72. Ion-pair and hydrogen-bonding interactions are indicated by red dashed lines. **e**, Molecular surface of the active-site region of Ssu72. The CTD phosphopeptide is shown as a stick model.

with the pSer-Pro peptide bond in the *cis* configuration, in contrast to all other known CTD phosphatases (Supplementary Information).

Our observation of a *cis* configuration for the CTD also provides a different interpretation for the role of the peptidyl-prolyl isomerase Pin1 (Ess1 in yeast) in regulating Pol II transcription[23–25]. It has been proposed that Pin1/Ess1 promotes the *trans* configuration of the CTD for dephosphorylation by Ssu72 (refs 24, 25), whereas our structure indicates that the opposite must be true. Our *in vitro* phosphatase assays demonstrate that Pin1 strongly stimulates the phosphatase activity of Ssu72 (Supplementary Information and Supplementary Fig. 6), consistent with its specificity for the *cis* configuration.

The active site of Ssu72 is located at the bottom of a narrow groove (Fig. 2e), one wall of which is formed by the small β-sheet (β2A and β2B) and the loop linking the two strands (Fig. 2d). This severely limits the possible conformation of the CTD, ensuring that only the *cis* configuration of the pSer-Pro peptide bond can be accommodated in the active site. In fact, the Thr 4-pSer 5 peptide bond is π-stacked with the Pro 6-Ser 7 peptide bond (Fig. 2d), suggesting a highly restrained conformation for the CTD phosphopeptide in this region. Residues Thr 4, pSer 5 and Pro 6 of the same repeat as well as Tyr 1 of the following repeat have interactions with the enzyme (Fig. 2d and Supplementary Information), explaining the preference for pSer 5 by Ssu72 and consistent with results from earlier biochemical studies on yeast Ssu72 (ref. 14).

The phosphate group of the peptide is bound deepest in the structure, having extensive ion-pair and hydrogen-bonding interactions with the enzyme (Fig. 2d). In addition, the main-chain amide group of pSer 5 is hydrogen-bonded to the main-chain carbonyl of Lys 43 (in β2A). The catalytic nucleophile of Ssu72, Cys 12, is located directly below the phosphate group and can be in the correct position for the inline nucleophilic attack on the phosphorus atom to initiate the reaction (Supplementary Fig. 7). The side chain of Asp 143 is located 3.5 Å from the Oγ atom of Ser 5, consistent with its role as the general acid to protonate the leaving group. There are some conformational changes in the active-site region of Ssu72, especially for the β2A–β2B loop, on binding of the CTD phosphopeptide (Supplementary Fig. 7), although this loop seems to be flexible and can assume different conformations in the various structures.

In the structures of the binary and ternary complexes, Ssu72 is bound to the concave face of Symp-N (Fig. 1b). About 950 $Å^2$ of the surface area of each protein is buried in the interface of this complex, which involves helices α3B–α6B of Symp-N (Fig. 3a and Supplementary Fig. 1), and helix αE, the following αE–β4 loop, helix αG and strand β5 of Ssu72 (Supplementary Fig. 3). In addition, residue Arg 206, at the tip of the long loop connecting helices α4B and α5A of Symp-N, is also located in the interface (Fig. 3a). Ion-pair, hydrogen-bonding and hydrophobic interactions make contributions to the formation of this complex (Supplementary Information). In particular, the side chains of Val 191 and Phe 193 of Ssu72 (in strand β5) establish hydrophobic interactions with those of Lys 185 (α4B) and Ile 251 (α5B) of symplekin in the centre of this interface (Fig. 3a). In addition, the side-chain hydroxyl group of Thr 190 (β5) of Ssu72 is hydrogen-bonded to the side chain of Asn 300 (α6B) of symplekin. The relative positions of Symp-N and Ssu72 seem to be somewhat variable among

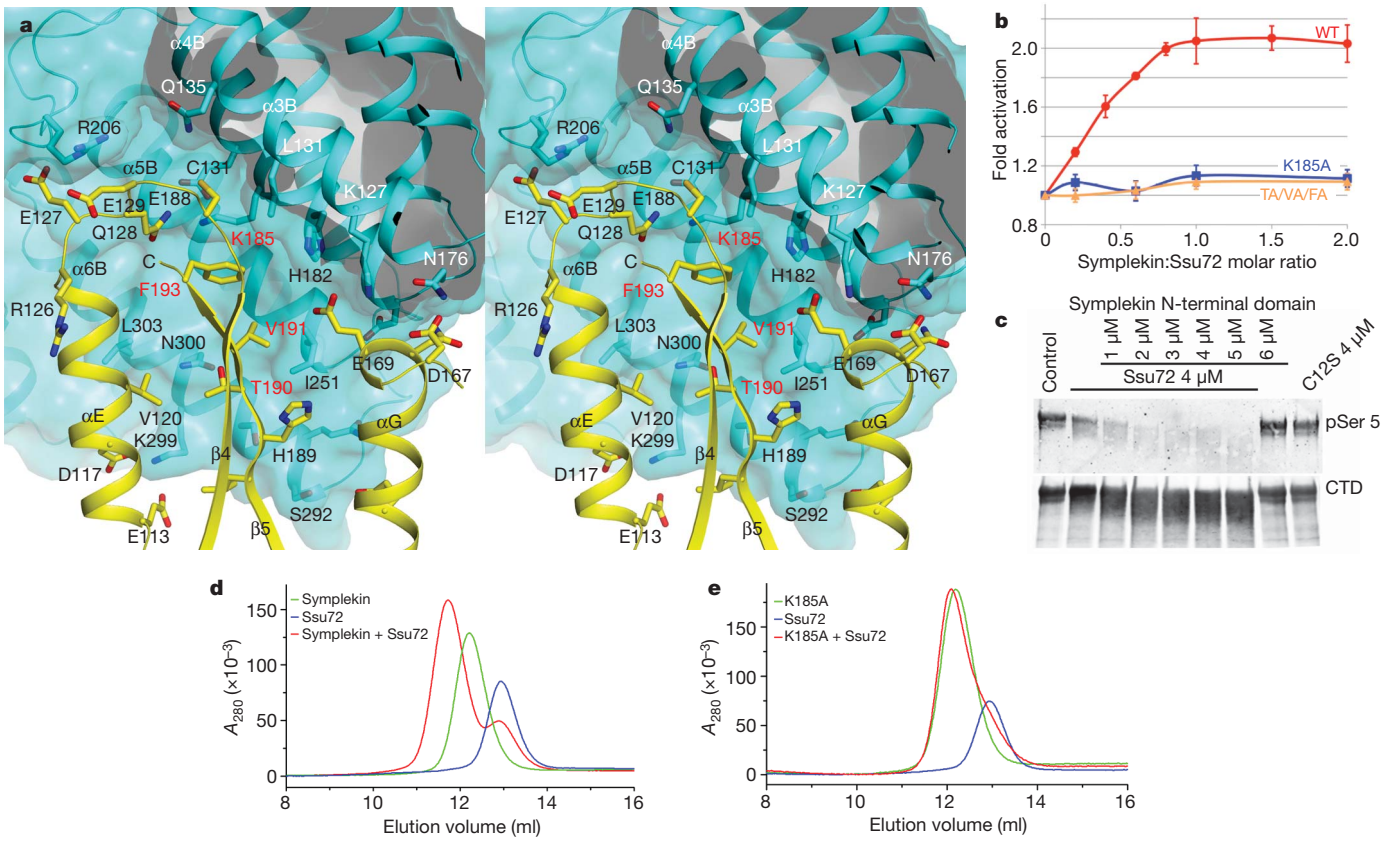

Figure 3 | Structural and biochemical characterizations of the symplekin–Ssu72 interface. **a**, Stereo pair showing detailed interactions between symplekin (in cyan) and Ssu72 (in yellow) in the interface. The molecular surface of symplekin is also shown (in cyan). Side chains making important contributions to the interface are shown as stick models. Residues labelled in red were selected for mutagenesis. **b**, The activity of Ssu72, measured by the hydrolysis of pNPP, as a function of the molar ratio of symplekin. TA/VA/FA, T190A/V191A/F193A mutant of Ssu72 (stimulation by wild-type (WT) symplekin). Results are shown as means ± s.d. for three independent experiments. **c**, Stimulation of the CTD Ser 5 phosphatase activity of Ssu72 by symplekin. The levels of pSer 5 and total CTD were determined using the H14 and 8WG16 antibodies, respectively. **d**, Gel-filtration profiles for wild-type symplekin N-terminal domain alone, wild-type human Ssu72 (full-length) alone, and a mixture of the two (with Ssu72 present in roughly twofold molar excess). **e**, Gel-filtration profiles for wild-type Ssu72 alone, K185A mutant of symplekin alone, and a mixture of the two.

the binary and ternary complexes (Supplementary Information and Supplementary Fig. 8).

The symplekin–Ssu72 interface is located about 25 Å from the active site of Ssu72 (Fig. 1b). However, phosphatase assays measuring the hydrolysis of a *p*-nitrophenyl phosphate (pNPP) model substrate[11,12] showed that Symp-N stimulated Ssu72 activity (Fig. 3b), and maximal activation was achieved when the two proteins were at a 1:1 molar ratio. To assess whether this stimulation also occurs with a natural substrate, we first used the decamer CTD phosphopeptide in the assay, monitoring the release of inorganic phosphate, and observed a similar stimulation (Supplementary Fig. 6). We next prepared a glutathione *S*-transferase (GST)–CTD fusion protein that had been phosphorylated on Ser 2 and Ser 5 with HeLa nuclear extract[26]. As demonstrated by western blotting with a pSer 5-specific antibody, Ssu72 dephosphorylated this protein on Ser 5, in a manner that was also stimulated by Symp-N (Fig. 3c). Ssu72 was specific for dephosphorylating pSer 5, because Ser 2 phosphorylation, as monitored by a pSer 2-specific antibody, was not affected (data not shown).

Our data indicate that the symplekin–Ssu72 interaction activated Ssu72 phosphatase activity, probably through stabilization of the Ssu72 structure and/or an allosteric mechanism. This is consistent with previous studies on the R129A mutant (*ssu72-2*) of yeast Ssu72, equivalent to Arg 126 in human Ssu72 (Supplementary Fig. 3). This mutant shows a twofold decrease in catalytic activity compared with wild-type Ssu72 and produces a severe growth defect at the non-permissive temperature[16]. Arg 126 is far from the active site and is in fact near the interface with symplekin (Fig. 3a). However, it does not contribute directly to interactions with symplekin, and the R126A mutation did not disrupt interaction with Symp-N (data not shown).

To assess the importance of individual residues for the stability of the symplekin–Ssu72 complex, we introduced mutations in the interface and characterized their effects on the complex by using gel-filtration chromatography and phosphatase assays. The presence of wild-type Ssu72 gave rise to a clear shift in the peak for Symp-N from a gel-filtration column (Fig. 3d), corresponding to the formation of the symplekin–Ssu72 complex. Ssu72 was present in twofold molar excess in this experiment, and only half of this protein was incorporated into the complex (Fig. 3d), demonstrating a 1:1 stoichiometry for the complex. Mutation of a symplekin residue in the interface, K185A (Fig. 3a), essentially abolished the interaction with wild-type Ssu72 (Fig. 3e), and mutation of three Ssu72 residues in the interface, T190A/V191A/F193A, abolished the interaction with wild-type symplekin. The chromatographic behaviour of the mutants alone was similar to that of the wild-type protein (Fig. 3e), suggesting that the mutations did not disrupt the structure of the proteins. This was also confirmed by the crystal structure of the K185A mutant (data not shown). Consistent with the gel-filtration data, the symplekin K185A mutant failed to stimulate Ssu72 phosphatase activity, and the T190A/V191A/F193A mutant of Ssu72 could not be stimulated by wild-type Symp-N (Fig. 3b).

We next wished to assess the functional importance of the symplekin–Ssu72 interaction with respect to 3′-end formation. Given the roles of their yeast counterparts in both transcription and polyadenylation, we used a transcription-coupled 3′-end processing assay[27]. HeLa nuclear extract was preincubated with increasing concentrations of Symp-N, which led to a pronounced inhibition of polyadenylation (Fig. 4a), similar to an effect observed earlier with the yeast Pta1 N-terminal domain in a transcription-independent assay[9]. Transcription, as measured by the accumulation of unprocessed pre-mRNA, was not affected (Fig. 4b). RNase protection assays showed that 3′-end cleavage was also not affected (Supplementary Fig. 9), indicating that Symp-N affects only the polyadenylation step of 3′-end formation. Inclusion of purified Ssu72 during the preincubation with Symp-N blocked the inhibition, whereas Ssu72 alone had no effect (Fig. 4c). The K185A mutation in Symp-N abolished this inhibitory effect, whereas the T190A/V191A/F193A mutant of Ssu72 failed to overcome the inhibition by wild-type Symp-N (Fig. 4c). These results provide strong evidence that the

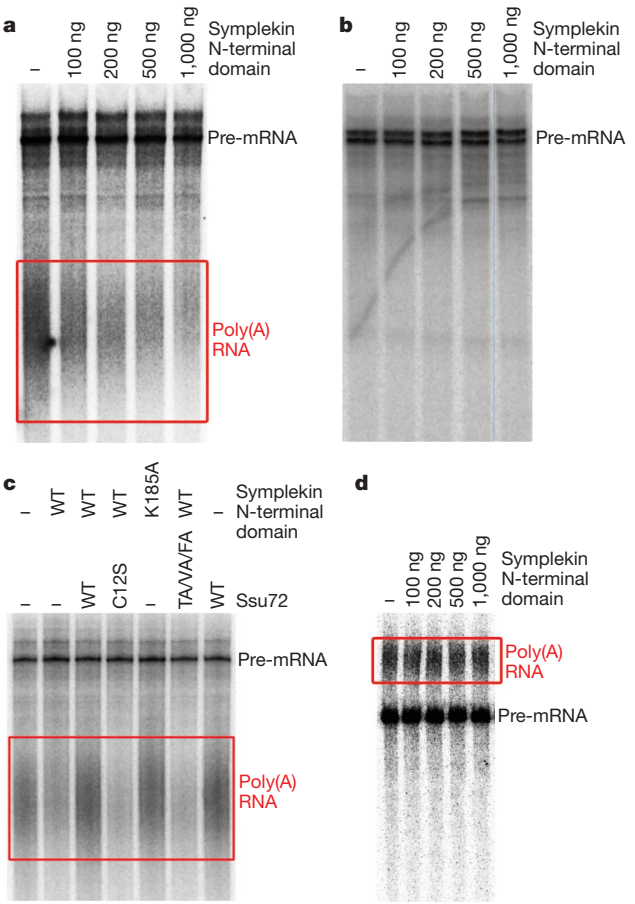

Figure 4 | Functional characterization of the symplekin–Ssu72 interaction. **a**, Transcription-coupled polyadenylation is inhibited in a dose-dependent fashion by symplekin N-terminal domain. Transcription-processing was performed in HeLa nuclear extract; RNAs were purified and separated into poly(A)$^-$ and poly(A)$^+$ fractions, and resolved by denaturing PAGE. Positions of unprocessed 'run-off' RNA (pre-mRNA) and cleaved and polyadenylated RNA (poly(A) RNA) are indicated. **b**, Symplekin N-terminal domain does not inhibit transcription. Poly(A)$^-$ RNAs (2%) from the transcription-coupled polyadenylation assays in the presence of increasing concentrations of symplekin N-terminal domain are shown. **c**, Ssu72 overcomes the inhibition of transcription-coupled polyadenylation by symplekin. The effects of the K185A mutant of symplekin and the C12S and T190A/V191A/F193A (TA/VA/FA) mutants of Ssu72 on polyadenylation are also shown. **d**, Polyadenylation of SV40 late pre-mRNA, uncoupled to transcription, is not inhibited by symplekin N-terminal domain.

inhibitory effect of Symp-N reflects its interaction with Ssu72, and thus implicates Ssu72 in mammalian 3′-end processing. In contrast with studies in yeast[7,9], the catalytically inactive C12S mutant of Ssu72 failed to overcome this inhibition (Fig. 4c), and Symp-N had no detectable effect on transcription-independent polyadenylation (Fig. 4d, and data not shown). Together, these results indicate that Ssu72 phosphatase activity is required for polyadenylation of pre-mRNAs, but only when processing is coupled to transcription.

Our finding that a CTD phosphopeptide is bound to Ssu72 with the pSer-Pro peptide bond in the *cis* configuration indicates the existence of a novel CTD conformation. Although Ssu72 has been well studied in yeast and has functions in transcription and 3′-end processing, essentially nothing was known about its mammalian counterpart. In fact, whereas the yeast enzyme is a stable component of the polyadenylation machinery and is required for processing, mammalian Ssu72 has not been found associated with polyadenylation factors and was not detected in a recent proteomic analysis of the assembled polyadenylation complex[28]. Consistent with this, our results provide evidence that in mammals Ssu72 is only necessary for polyadenylation when processing is coupled to

transcription. A parsimonious model is that symplekin recruits Ssu72 to the transcription complex and activates its phosphatase activity, which promotes polyadenylation. Conceivably, this occurs by facilitating the recruitment of poly(A) polymerase, known for many years to be only weakly associated with other 3′-end processing factors[28,29], to the complex. Given that the CTD is necessary for efficient 3′-end formation in mammalian cells[26,30], and that CTD pSer 5 is the only known target of Ssu72, CTD pSer 5 dephosphorylation may well be important in facilitating polyadenylation during transcription.

METHODS SUMMARY

Crystallography. The N-terminal domain of human symplekin and full-length human Ssu72 were overexpressed separately in *Escherichia coli* and purified. The symplekin–Ssu72 complex was purified by gel filtration of a mixture of the two proteins. Crystals were obtained by the sitting-drop vapour-diffusion method, and the structures were determined by the selenomethionyl single-wavelength anomalous diffraction method and the molecular replacement method.

***In vitro* transcription-coupled polyadenylation assay.** Transcription-coupled polyadenylation was performed with a DNA construct containing GAL4-binding sites upstream of the adenovirus E4 core promoter and SV40 late poly(A) site downstream. Recombinant symplekin and Ssu72 proteins were preincubated with HeLa nuclear extract before transcription was started by adding the DNA template and purified GAL4–VP16. RNA products were purified, separated into nonpolyadenylated and polyadenylated fractions and analysed on 5% denaturing gel. Radioactivity was detected with a PhosphorImager. Assays were repeated multiple times with consistent results.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 March; accepted 3 August 2010.
Published online 22 September 2010.

1. Takagaki, Y. & Manley, J. L. Complex protein interactions within the human polyadenylation machinery identify a novel component. *Mol. Cell. Biol.* **20,** 1515–1525 (2000).
2. Zhao, J., Kessler, M. M., Helmling, S., O'Connor, J. P. & Moore, C. L. Pta1, a component of yeast CFII, is required for both cleavage and poly(A) addition of mRNA precursor. *Mol. Cell. Biol.* **19,** 7733–7740 (1999).
3. Zhao, J., Hyman, L. & Moore, C. L. Formation of mRNA 3′ ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol. Mol. Biol. Rev.* **63,** 405–445 (1999).
4. Mandel, C. R., Bai, Y. & Tong, L. Protein factors in pre-mRNA 3′-end processing. *Cell. Mol. Life Sci.* **65,** 1099–1122 (2008).
5. Calvo, O. & Manley, J. L. Strange bedfellows: polyadenylation factors at the promoter. *Genes Dev.* **17,** 1321–1327 (2003).
6. Moore, M. J. & Proudfoot, N. J. Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* **136,** 688–700 (2009).
7. He, X. *et al.* Functional interactions between the transcription and mRNA 3′ end processing machineries mediated by Ssu72 and Sub1. *Genes Dev.* **17,** 1030–1042 (2003).
8. Nedea, E. *et al.* Organization and function of APT, a subcomplex of the yeast cleavage and polyadenylation factor involved in the formation of mRNA and small nucleolar RNA 3′-ends. *J. Biol. Chem.* **278,** 33000–33010 (2003).
9. Ghazy, M. A., He, X., Singh, B. N., Hampsey, M. & Moore, C. The essential N terminus of the Pta1 scaffold protein is required for snoRNA transcription termination and Ssu72 function but is dispensible for pre-mRNA 3′-end processing. *Mol. Cell. Biol.* **29,** 2296–2307 (2009).
10. Dichtl, B. *et al.* A role for SSU72 in balancing RNA polymerase II transcription elongation and termination. *Mol. Cell* **10,** 1139–1150 (2002).
11. Meinhart, A., Silberzahn, T. & Cramer, P. The mRNA transcription/processing factor Ssu72 is a potential tyrosine phosphatase. *J. Biol. Chem.* **278,** 15917–15921 (2003).
12. Ganem, C. *et al.* Ssu72 is a phosphatase essential for transcription termination of snoRNAs and specific mRNAs in yeast. *EMBO J.* **22,** 1588–1598 (2003).
13. Krishnamurthy, S., He, X., Reyes-Reyes, M., Moore, C. L. & Hampsey, M. Ssu72 is an RNA polymerase II CTD phosphatase. *Mol. Cell* **14,** 387–394 (2004).
14. Hausmann, S., Koiwa, H., Krishnamurthy, S., Hampsey, M. & Shuman, S. Different strategies for carboxyl-terminal domain (CTD) recognition by serine 5-specific CTD phosphatases. *J. Biol. Chem.* **280,** 37681–37688 (2005).
15. St-Pierre, B. *et al.* Conserved and specific functions of mammalian ssu72. *Nucleic Acids Res.* **33,** 464–477 (2005).
16. Reyes-Reyes, M. & Hampsey, M. Role for the Ssu72 C-terminal domain phosphatase in RNA polymerase II transcription elongation. *Mol. Cell. Biol.* **27,** 926–936 (2007).
17. Ansari, A. & Hampsey, M. A role for the CPF 3′-end processing machinery in RNAP II-dependent gene looping. *Genes Dev.* **19,** 2969–2978 (2005).
18. Su, X.-D., Taddei, N., Stefani, M., Ramponi, G. & Nordlund, P. The crystal structure of a low-molecular-weight phosphotyrosine protein phosphatase. *Nature* **370,** 575–578 (1994).
19. Zhang, M., Van Etten, R. L. & Stauffacher, C. V. Crystal structure of bovine heart phosphotyrosyl phosphatase at 2.2-Å resolution. *Biochemistry* **33,** 11097–11105 (1994).
20. Meinhart, A., Kamenski, T., Hoeppner, S., Baumli, S. & Cramer, P. A structural perspective of CTD function. *Genes Dev.* **19,** 1401–1415 (2005).
21. Zhang, Y. *et al.* Determinants for dephosphorylation of the RNA polymerase II C-terminal domain by Scp1. *Mol. Cell* **24,** 759–770 (2006).
22. Kennedy, S. A. *et al.* Crystal structure of the HEAT domain from the pre-mRNA processing factor symplekin. *J. Mol. Biol.* **392,** 115–128 (2009).
23. Xu, Y.-X., Hirose, Y., Zhou, X. Z., Lu, K. P. & Manley, J. L. Pin1 modulates the structure and function of human RNA polymerase II. *Genes Dev.* **17,** 2765–2776 (2003).
24. Krishnamurthy, S., Ghazy, M. A., Moore, C. & Hampsey, M. Functional interaction of the Ess1 prolyl isomerase with components of the RNA polymerase II initiation and termination machineries. *Mol. Cell. Biol.* **29,** 2925–2934 (2009).
25. Singh, N. *et al.* The Ess1 prolyl isomerase is required for transcription termination of small noncoding RNAs via the Nrd1 pathway. *Mol. Cell* **36,** 255–266 (2009).
26. Hirose, Y. & Manley, J. L. RNA polymerase II is an essential mRNA polyadenylation factor. *Nature* **395,** 93–96 (1998).
27. Rozenblatt-Rosen, O. *et al.* The tumor suppressor Cdc73 functionally associates with CPSF and CstF 3′ mRNA processing factors. *Proc. Natl Acad. Sci. USA* **106,** 755–760 (2009).
28. Shi, Y. *et al.* Molecular architecture of the human pre-mRNA 3′ processing complex. *Mol. Cell* **33,** 365–376 (2009).
29. Takagaki, Y., Ryner, L. C. & Manley, J. L. Separation and characterization of a poly(A) polymerase and a cleavage/specificity factor required for pre-mRNA polyadenylation. *Cell* **52,** 731–742 (1988).
30. McCracken, S. *et al.* The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385,** 357–361 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank F. Forouhar and J. Seetharaman for help with data collection; T. Auperin for initial experiments with symplekin; R. Abramowitz and J. Schwanof for setting up the X4C beamline; S. Myers for setting up the X29A beamline at the National Synchrotron Light Source; M. Heidemann and D. Eick for providing the 3E10 antibody; N. Rao and C. Logan for HeLa nuclear extract and help with *in vitro* assays; and J. Decatur for characterizing the phosphopeptide by NMR. This research is supported in part by grants from the National Institutes of Health to L.T. (GM077175) and J.L.M. (GM028983).

Author Contributions K.X., S.X., T.K. and M.M.B. performed protein expression, purification and crystallization experiments. K.X., S.X. and L.T. conducted crystallographic data collection, structure determination and refinement. T.N. and K.X. performed polyadenylation experiments. K.X. performed Ssu72 phosphatase assays. All authors commented on the manuscript. L.T. and J.L.M. designed the experiments, analysed the data and wrote the paper.

Author Information Atomic coordinates have been deposited at the Protein Data Bank (accession codes 3O2Q, 3O2S, 3O2T, 3ODR and 3ODS). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to L.T. (ltong@columbia.edu).

METHODS

Protein expression and purification. Residues 30–395 of human symplekin were subcloned into the pET28a vector (Novagen). The recombinant protein carries a hexahistidine tag at the N terminus. The plasmids were transformed into *E. coli* BL21(DE3) Star cells. After induction with 0.5 mM isopropyl β-D-thiogalactoside, the cells were allowed to grow at 20 °C for 14–16 h, collected by centrifugation, and lysed by sonication. Soluble symplekin was purified by Ni^{2+}-nitrilotriacetate (Qiagen) and gel-filtration (Sephacryl S-300; GE Healthcare) chromatography. Purified protein was concentrated to 15 mg ml^{-1} in a buffer containing 20 mM Tris-HCl pH 8.5, 200 mM NaCl, 10 mM dithiothreitol (DTT) and 5% (v/v) glycerol, flash-frozen in liquid nitrogen and stored at −80 °C. The N-terminal His-tag was not removed for crystallization.

The selenomethionyl protein sample was produced in *E. coli* B834(DE3) cells, and the bacteria were grown in defined LeMaster medium supplemented with selenomethionine[31]. The purification procedure was the same as for the native protein.

To prepare the symplekin–Ssu72 complex, residues 30–360 of human symplekin and full-length human Ssu72 were cloned separately into the pET28a vector. Both proteins carried a hexahistidine tag at the N terminus and were purified separately by Ni^{2+}-nitrilotriacetate (Qiagen) and gel-filtration (Sephacryl S-300; GE Healthcare) chromatography. The purified proteins were then mixed, with Ssu72 present in slight molar excess (1:1.2 molar ratio), and the symplekin–Ssu72 complex was purified by gel-filtration chromatography. Purified protein was concentrated to 10 mg ml^{-1} in a buffer containing 20 mM Tris-HCl pH 8.5, 200 mM NaCl, 10 mM DTT and 5% (v/v) glycerol, flash-frozen in liquid nitrogen and stored at −80 °C.

Protein crystallization. Crystals of symplekin (residues 30–395) were obtained with the sitting-drop vapour-diffusion method at 20 °C. The reservoir solution contained 50 mM Bis-Tris pH 6.5, 40 mM ammonium sulphate and 40% (v/v) pentaerythritol ethoxylate (15/4 EO/OH). The crystals belong to space group $P2_1$, with unit cell parameters of $a = 41.6$ Å, $b = 63.0$ Å, $c = 62.3$ Å and $\beta = 90.6°$. There is one symplekin molecule in the asymmetric unit.

Crystals of the symplekin–Ssu72 complex were obtained with the sitting-drop vapour-diffusion method at 20 °C. The reservoir solution contained 0.32 M ammonium sulphate and 26% (w/v) PEG 3350. The crystals belong to space group $P2_12_12$, with unit cell parameters of $a = 99.4$ Å, $b = 113.6$ Å and $c = 59.1$ Å. There is one symplekin–Ssu72 complex in the asymmetric unit.

Crystals of the symplekin–Ssu72–CTD phosphopeptide ternary complex were obtained with the sitting-drop vapour-diffusion method at 20 °C. The reservoir solution contained 1.6 M ammonium chloride and 27% (w/v) PEG 3350. Sodium potassium tartrate (10 mM) was included as an additive in the drop. These crystals of the symplekin–Ssu72 binary complex were soaked overnight in a solution containing 30 mM CTD phosphopeptide, 25% (w/v) PEG 3000 and 100 mM Tris-HCl pH 8.5 at 20 °C. The crystal belongs to space group $P2_1$, with unit cell parameters of $a = 67.0$ Å, $b = 97.6$ Å, $c = 105.0$ Å and $\beta = 98.7°$. There are two copies of the symplekin–Ssu72 complex in the asymmetric unit. The CTD phosphopeptide was observed in only one of the Ssu72 molecules, and a phosphate is bound in the active site of the other Ssu72 molecule. We found that commercial PEG 3350 contains some phosphate (about 1 mM) as a contaminant.

The crystals were cryoprotected by the reservoir solution, supplemented with 25% (v/v) ethylene glycol if necessary, and flash-frozen in liquid nitrogen for data collection at 100 K.

Data collection and structure determination. The structure of symplekin alone was determined by the selenomethionyl anomalous diffraction method[32]. A single-wavelength anomalous diffraction (SAD) data set to 2.0 Å resolution was collected on a SeMet-substituted crystal at beamline 9-2 of the Stanford Synchrotron Radiation Laboratory. A native reflection data set to 1.4 Å resolution was collected at the same beamline. The diffraction data were processed and scaled with the HKL package[33]. The data processing statistics are summarized in Supplementary Table 1.

The Se atoms were located with the program BnP[34], and the reflections were phased with the program Solve[35]. Most of the residues were built automatically by the program Resolve, and the model was completed by manual building with the programs O[36] and Coot[37]. The structure refinement was performed with the programs CNS[38] and Refmac[39].

The structure of the symplekin–Ssu72 complex was determined by the molecular replacement method, with the program COMO[40]. A native reflection data set to 2.5 Å resolution was collected at the X4C beamline of the National Synchrotron Light Source. The structures of symplekin and *Drosophila* Ssu72 (PDB ID 3FDF) were used as the search models.

A native reflection data set to 2.4 Å resolution was collected at the X29 beamline on the symplekin–Ssu72–CTD phosphopeptide ternary complex.

Symplekin–Ssu72 interactions. Symplekin and Ssu72 mutants were made with the QuikChange kit (Stratagene) and verified by sequencing. The mutant proteins were expressed and purified by following the same protocol as that for the wild-type protein.

Analytical gel-filtration experiments were carried out on a Superose-12 10/30 column (GE Healthcare), with a buffer containing 20 mM Tris-HCl pH 7.5 and 200 mM NaCl. Symplekin (210 μg) and Ssu72 (380 μg) were mixed and diluted to a final volume of 500 μl with the gel-filtration buffer. The mixture was incubated on ice for 1 h before being loaded on the column. The proteins were also run separately on the column to determine their migration behaviour alone.

Ssu72 CTD peptide phosphatase assays. Reaction mixtures (50 μl) containing 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM DTT, the indicated amount of CTD phosphopeptide, Ssu72 and symplekin were incubated at 37 °C. Time-point samples were taken and quenched by the addition of 0.5 ml of malachite green reagent (BIOMOL Research Laboratories). Phosphate release was determined by measuring A_{620} and comparing it with a phosphate standard curve.

To study the effect of Pin1 on Ssu72, reaction mixtures (50 μl) containing 50 mM Tris-HCl pH 7.5, 50 mM NaCl, 5 mM DTT, 500 μM CTD phosphopeptide and the indicated amount of Ssu72 and GST–Pin1 were incubated at 10 °C. Phosphate release was determined with the malachite green reagent.

Ssu72 GST–CTD phosphatase assays. GST–CTD fusion protein was expressed, purified and phosphorylated *in vitro* as described previously[41]. GST–CTD phosphatase assays were performed in a total volume of 10 μl containing 50 mM Tris-HCl pH 7.9, 100 mM KCl, 12.5 mM $MgCl_2$, 0.5 mM EDTA, 5% (v/v) glycerol and the indicated amount of phosphorylated GST–CTD as a substrate, Ssu72, symplekin or their mutants. Reactions were incubated for various durations at 37 °C, stopped by the addition of SDS loading buffer and resolved by 8% SDS–PAGE. pSer 2 and pSer 5 levels were detected by western blot with 3E10 (ref. 42) and H14 (Covance) antibodies, respectively.

***In vitro* transcription-coupled polyadenylation assay.** The DNA construct used for the transcription-coupled polyadenylation assay contained GAL4-binding sites upstream of an E4 core promoter and an SV40 late poly(A) site downstream. Transcription-coupled polyadenylation was performed at 30 °C for 1 h in 20-μl reaction mixtures containing 10 μl of nuclear extract, 100 ng of Gal4–VP16, recombinant proteins (symplekin and Ssu72), 12 mM HEPES pH 7.9, 500 ng of the DNA templates, 0.5 mM each of ATP, GTP and CTP, 15 μM unlabelled UTP, 10 μCi of [α-^{32}P]UTP, 4 mM $MgCl_2$, 20 mM creatine phosphate (di-tris), 2.4% PEG 8000, 12% glycerol, 60 mM KCl, 0.12 mM EDTA, 0.12 mM DTT and 0.3 mM phenylmethylsulphonyl fluoride. Recombinant symplekin and Ssu72 proteins were preincubated with nuclear extract (30 min at 23 °C) before transcription was started by adding the DNA templates. The reaction was stopped by adding proteinase K. RNA products were separated into non-polyadenylated and polyadenylated fractions by oligo(dT) selection; thereafter 2% of non-polyadenylated and 100% of polyadenylated fractions were analysed on 5% denaturing gel. Radioactivity was detected with a PhosphorImager.

RNase protection assay. To make the probe, pG3SVL-A was linearized with *Sal*I and transcribed with T7 RNA polymerase (Promega) for 2 h at 37 °C, uniformly labelled with [α-^{32}P]UTP. The RNA was gel purified and resuspended in hybridization buffer containing 40 mM PIPES pH 6.4, 400 mM NaCl, 1 mM EDTA and 80% (v/v) formamide.

Transcription-coupled polyadenylation was performed as mentioned above except that 0.5 mM each of ATP, GTP, UTP and CTP, but no [α-^{32}P]UTP, were used in the reaction mixture. The RNA/DNA mixture was resuspended in 20 μl of Turbo DNase buffer (Ambion) with 1 U of Turbo DNase (Ambion) and incubated for 1 h at 37 °C to remove the DNA template. The remaining RNA products were hybridized overnight with the probe (5×10^5 c.p.m.) at 42 °C. Each reaction was then incubated for 45 min with 14 μg of RNase A and 0.7 μg of RNase T1 at 30 °C. The final RNA products were resolved on a 10% denaturing gel. Radioactivity was detected with a PhosphorImager.

31. Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* **9,** 1665–1672 (1990).
32. Hendrickson, W. A. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254,** 51–58 (1991).
33. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
34. Weeks, C. M. & Miller, R. The design and implementation of SnB v2.0. *J. Appl. Cryst.* **32,** 120–124 (1999).
35. Terwilliger, T. C. SOLVE and RESOLVE: automated structure solution and density modification. *Methods Enzymol.* **374,** 22–37 (2003).
36. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110–119 (1991).
37. Emsley, P. & Cowtan, K. D. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2126–2132 (2004).

38. Brunger, A. T. *et al.* Crystallography & NMR System: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54,** 905–921 (1998).
39. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53,** 240–255 (1997).
40. Jogl, G., Tao, X., Xu, Y. & Tong, L. COMO: A program for combined molecular replacement. *Acta Crystallogr. D Biol. Crystallogr.* **57,** 1127–1134 (2001).
41. Hirose, Y. & Manley, J. L. RNA polymerase II is an essential mRNA polyadenylation factor. *Nature* **395,** 93–96 (1998).
42. Chapman, R. D. *et al.* Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science* **318,** 1780–1782 (2007).