LETTERS

Polyadenylation factor CPSF-73 is the pre-mRNA 3′-end-processing endonuclease

Corey R. Mandel[1], Syuzo Kaneko[1]*, Hailong Zhang[1]*†, Damara Gebauer[1]†, Vasupradha Vethantham[1], James L. Manley[1] & Liang Tong[1]

Most eukaryotic messenger RNA precursors (pre-mRNAs) undergo extensive maturational processing, including cleavage and polyadenylation at the 3′-end[1–8]. Despite the characterization of many proteins that are required for the cleavage reaction, the identity of the endonuclease is not known[4,9,10]. Recent analyses indicated that the 73-kDa subunit of cleavage and polyadenylation specificity factor (CPSF-73) might be the endonuclease for this and related reactions[10–15], although no direct data confirmed this. Here we report the crystal structures of human CPSF-73 at 2.1 Å resolution, complexed with zinc ions and a sulphate that might mimic the phosphate group of the substrate, and the related yeast protein CPSF-100 (Ydh1) at 2.5 Å resolution. Both CPSF-73 and CPSF-100 contain two domains, a metallo-β-lactamase domain and a novel β-CASP (named for metallo-β-lactamase, CPSF, Artemis, Snm1, Pso2) domain[12]. The active site of CPSF-73, with two zinc ions, is located at the interface of the two domains. Purified recombinant CPSF-73 possesses RNA endonuclease activity, and mutations that disrupt zinc binding in the active site abolish this activity. Our studies provide the first direct experimental evidence that CPSF-73 is the pre-mRNA 3′-end-processing endonuclease.

CPSF-73 belongs to the metallo-β-lactamase superfamily of zinc-dependent hydrolases[11,12]. Canonical metallo-β-lactamases contain five signature sequence motifs—Asp (motif 1), His-X-His-X-Asp-His (motif 2), His (motif 3), Asp (motif 4) and His (motif 5), most of which are ligands to the two zinc ions in their active site. Sequence conservation between CPSF-73 and the canonical metallo-β-lactamases is limited to these signature motifs. Whereas the first four motifs have been identified in the N-terminal segment of CPSF-73 (see Supplementary Fig. 1a, Supplementary Table 1), the fifth motif was uncertain, with three candidates, A (Asp or Glu), B (His) and C (His) (see Supplementary Fig. 1a), in the so-called β-CASP motif[12]. Motif B was proposed to be equivalent to motif 5 in the canonical metallo-β-lactamases. Another subunit of CPSF, CPSF-100, shares sequence conservation (see Supplementary Fig. 1b) and a similar domain architecture (see Supplementary Fig. 1a) with CPSF-73 but lacks the putative zinc-binding residues.

To understand the roles of CPSF-73 and CPSF-100 in pre-mRNA 3′-end processing, we determined the structures of human CPSF-73 (residues 1–460) and yeast CPSF-100 (residues 1–720) (the crystallographic data are summarized in Supplementary Table 2). The two structures obtained for CPSF-73 were crystallized in the absence or presence of 0.5 mM zinc (although both structures contained zinc atoms; see below). We discovered serendipitously that *in situ* proteolysis by a fungal protease is crucial for the crystallization of yeast CPSF-100 (ref. 16).

The structure of CPSF-73 can be divided into two domains (Fig. 1a). The amino-terminal residues (amino acids 1–208) form a

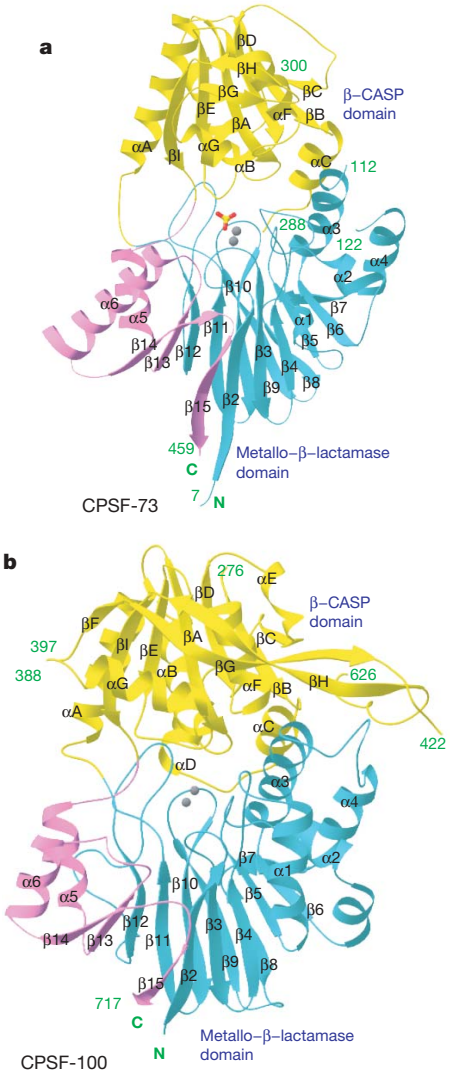

Figure 1 | Structures of human CPSF-73 and yeast CPSF-100 (Ydh1). **a**, Schematic representation of the structure of human CPSF-73. The β-strands and α-helices are labelled, and the two zinc atoms in the active site are shown as grey spheres. The sulphate ion is shown as a stick model. **b**, Schematic representation of the structure of yeast CPSF-100. The zinc atoms in the CPSF-73 structure are shown for reference. See Supplementary Fig. 2 For Schematic drawings of the metallo-β-lactamase domains of the two proteins.

[1]Department of Biological Sciences, Columbia University, New York, New York 10027, USA. †Present addresses: Array BioPharma Inc., Boulder, Colorado 80301, USA (H.Z.); Johnson & Johnson Pharmaceutical Research & Development LLC, San Diego, California 92121, USA (D.G.).
*These authors contributed equally to this work.

domain similar to the structure of canonical metallo-β-lactamases, with a four-layered αβ/βα architecture (see Supplementary Fig. 2a). A close structural homologue of this domain is the L1 metallo-β-lactamase from *Stenotrophomonas maltophilia* (see Supplementary Fig. 2c)[17], with a root mean squared (r.m.s.) distance of 3.3 Å for 167 equivalent Cα atoms, calculated using the program Dali[18]. However, the sequence identity for these structurally equivalent residues is only 17%.

Despite the overall similarity to canonical metallo-β-lactamases, the structure of this domain of CPSF-73 contains several distinctive features. Most remarkably, residues 395–460, at the carboxy-terminal end of the expression construct (see Supplementary Fig. 1a), are also part of the metallo-β-lactamase domain (Fig. 1a). These residues add three strands (β13 to β15) to the central β-sandwich, and the last residue of the structure is located next to the first residue (see Supplementary Fig. 2a). These additional β-strands in the metallo-β-lactamase fold are also seen in the structure of CPSF-100 (Fig. 1b, Supplementary Fig. 2b; see below) and RNase Z (see Supplementary Fig. 2d)[19,20], indicating that they might be unique to RNA/DNA-processing nucleases. This feature of the metallo-β-lactamase fold is crucial for the activity of these proteins, as the loop after strand β13 provides one of the ligands (motif C) to the zinc ions in the active site (see Supplementary Fig. 2a and see below). The r.m.s. distance for 210 equivalent Cα atoms between CPSF-73 and RNase Z is 2.6 Å, but the sequence identity is only 18%. RNase Z is also distinct from CPSF-73 in that it lacks the second domain (Fig. 1a, Supplementary Fig. 2d).

The second domain of CPSF-73 covers residues 209–394 (Supplementary Fig. 1a), and contains a central, fully parallel, six-stranded β-sheet that is surrounded by five α-helices on both faces (Fig. 2a). In addition, a small β-hairpin structure is on the surface of the domain. This second domain can be considered as a cassette inserted into the metallo-β-lactamase domain, between strand β12 and helix α5 (Fig. 1a). As it is related to the β-CASP motif identified earlier on the basis of sequence analyses[12], we have named it the β-CASP domain (see Supplementary Fig. 1a).

The closest structural homologue of the β-CASP domain is the nucleotide-binding fold (NBF, Supplementary Fig. 3), with a *Z* score of 6 from Dali[18]. However, the β-CASP domain lacks the Walker A motif for binding nucleotide phosphate groups[21] (see Supplementary Information and Supplementary Fig. 3). Therefore, the β-CASP domain seems to be a new example of the NBF superfold but is unlikely to bind nucleotides. This domain is highly conserved among CPSF-73 proteins, and probably has important functions in regulating their activity (see below).

The overall structure of yeast CPSF-100 (Fig. 1b) is remarkably similar to that of human CPSF-73, despite the low degree of sequence conservation between them (see Supplementary Fig. 1b). The r.m.s. distance for 235 equivalent Cα atoms in the metallo-β-lactamase domain (see Supplementary Fig. 2b) of the two structures is 2.2 Å (with 21% sequence identity), and that for 154 equivalent Cα atoms in the β-CASP domain is 2.3 Å (with 17% identity).

There are, however, significant differences in the structures of CPSF-100 and CPSF-73 (see Supplementary Fig. 4). Most importantly, motifs for zinc binding in the metallo-β-lactamase domain are missing in CPSF-100 (see Supplementary Fig. 2b), and therefore this protein cannot bind zinc and is unlikely to possess nuclease activity. In addition, the β-CASP domain of CPSF-100 (Fig. 2b) is much larger than that of CPSF-73, but contains a highly flexible segment (see Supplementary Information and Supplementary Fig. 5).

Clear electron density for two zinc ions (confirmed on the basis of anomalous scattering data) was observed in the CPSF-73 active site (see Supplementary Fig. 6a), regardless of whether zinc was present in the crystallization solution. Even in the absence of added zinc, the zinc ions in the structure seem to have full occupancy as their temperature factor values are comparable to those of their protein ligands. This indicates that the active site of CPSF-73 might possess extremely high affinity for zinc ions, making it impossible to remove the zinc from the protein, consistent with the known resistance of 3′ cleavage to relatively high concentrations of EDTA[22]. Two additional zinc ions were observed on the surface of CPSF-73 in the crystal grown in the presence of 0.5 mM zinc (see Supplementary Information and Supplementary Fig. 7).

The two zinc atoms in the active site are each bound in an octahedral environment, with a hydroxide ion as one of the bridging ligands (Fig. 3a, Supplementary Fig. 6b). The other bridging ligand is the side-chain carboxylate oxygen atom of Asp 179 from motif 4 (Fig. 3a). His 71, His 73 (motif 2) and His 158 (motif 3) provide three more ligands to the first zinc atom (Zn1), and Asp 75, His 76 (motif 2) and His 418 (motif C, see below) are the ligands to Zn2 (Fig. 3a). Finally, the two zinc ions are coordinated by oxygen atoms from a sulphate ion (Fig. 3a), which was present at 300 mM in the crystallization solution and has clear electron density (Supplementary Fig. 6a).

The structure of its active site strongly implies that CPSF-73 possesses hydrolase/nuclease activity. In canonical metallo-β-lactamases, the hydroxide that bridges the two zinc ions is the nucleophile for the hydrolysis reaction, and the substrate is directly liganded to the zinc atoms[23]. In our structure, the hydroxide is directly below the sulphate group (Fig. 3a), which is probably a good mimic for the phosphate group of the pre-mRNA substrate at the cleavage site. Therefore, the hydroxide is located at the perfect position for an inline nucleophilic attack on the phosphate group to initiate the nuclease reaction (Fig. 3a).

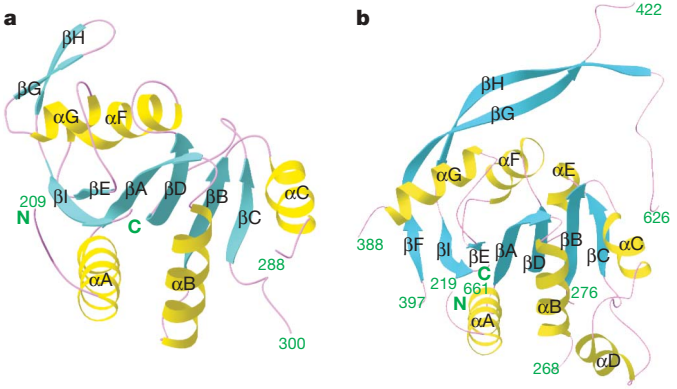

Figure 2 | The β-CASP domain of CPSF-73 and CPSF-100. Schematic drawings of the β-CASP domains of human CPSF-73 (**a**) and yeast CPSF-100 (**b**).

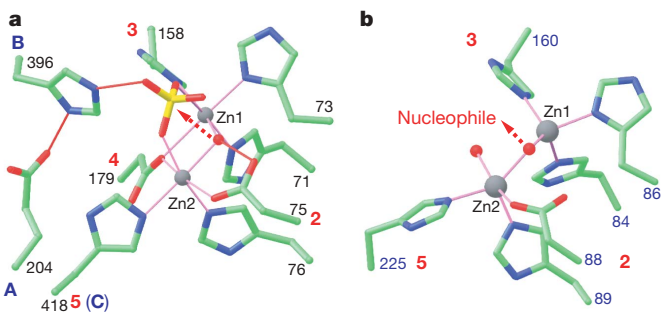

Figure 3 | The active site of CPSF-73. **a**, The zinc binding site for human CPSF-73. The motifs are labelled, and the bridging hydroxide ion is shown as a red sphere. Liganding interactions are indicated by thin magenta lines, and hydrogen-bonding interactions by thin red lines. The arrow indicates the nucleophilic attack from the hydroxide ion. **b**, The zinc binding site in L1 metallo-β-lactamase[17].

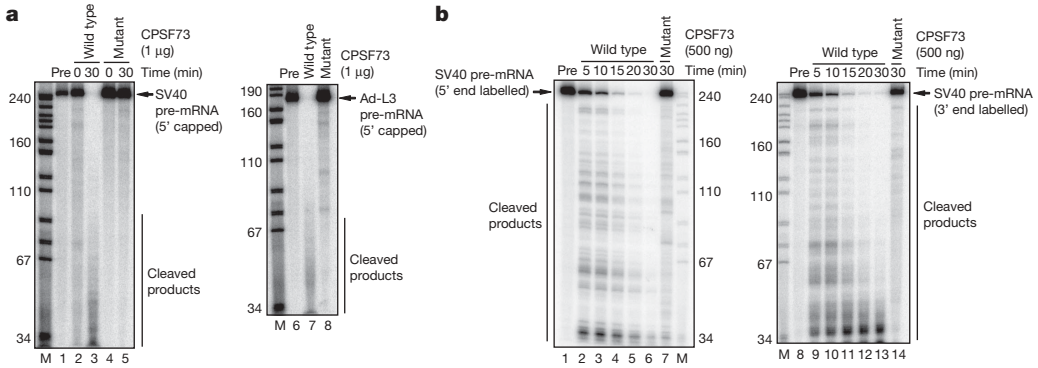

Figure 4 | CPSF-73 possesses endoribonuclease activity. a, Cleavage of 5′-capped SV40 late pre-mRNA (SVL) was performed with wild-type (lane 3) or mutant (lane 5) CPSF-73 at 30 °C for 30 min. Incubations for 0 min are shown in lanes 2 and 4. Cleavage of 5′-capped adenovirus L3 pre-mRNA (Ad-L3) is shown in lanes 7 and 8. Precursor RNAs (Pre) are shown in lanes 1 and 6. M, *Msp*I-digested pBR322DNA markers. **b**, Time-course analysis of RNA cleavage by CPSF-73. Cleavage of 5′ (lanes 2–6) and 3′ (lanes 9–13) end-labelled SVL pre-mRNAs was performed with wild-type CPSF-73 (500 ng); time points were 5, 10, 15, 20 and 30 min. Mutant CPSF-73 (500 ng) did not exhibit RNase activity (lanes 7 and 14).

Our analysis indicates that His 396 (motif B, in the linker between the β-CASP domain and the metallo-β-lactamase domain; see Supplementary Fig. 2a) is the general acid for catalysis. This residue is hydrogen-bonded to the side chain of Glu 204 (motif A, last residue of strand β12 in the metallo-β-lactamase domain, see Supplementary Fig. 2a) as well as an oxygen atom of the sulphate (Fig. 3a). This oxygen is located opposite to the direction of attack of the hydroxide (Fig. 3a), and therefore probably represents the leaving group in the hydrolysis reaction. After hydrolysis, the His 396 side chain can function as the general acid to protonate the oxyanion in the reaction product. This pair of Glu-His residues is also found in the structure of RNase Z, hydrogen-bonded to a phosphate group in the active site[19]. In yeast, mutation of either of these residues is lethal[13].

Although an earlier sequence analysis indicated that motif B might be the best candidate for the zinc ligand in CPSF-73 (ref. 12), our structure shows that the ligand is the His residue from motif C. This residue is conserved among CPSF-73 proteins and its close homologues (RC-68 or Int11)[24,25], but not in the DNA nuclease Artemis[12,26] (see Supplementary Fig. 1b). Therefore, the Zn2 atom might have a different coordination sphere in Artemis. The structure-based sequence alignment also indicates a change in the assignment of residues equivalent to motifs B and C in CPSF-100 (ref. 12; see Supplementary Fig. 1b), although neither residue is His in CPSF-100.

The binding modes of the zinc ions in CPSF-73 (Fig. 3a) are similar to that in RNase Z but distinct from those in canonical metallo-β-lactamases (Fig. 3b; see Supplementary Information). Another important difference between CPSF-73 and other metallo-β-lactamases is the presence of the β-CASP domain. Although most metallo-β-lactamases (Supplementary Fig. 2c), and RNase Z (Supplementary Fig. 2d), have an open zinc-binding site, the active site in CPSF-73 is located deep in the interface between the metallo-β-lactamase and the β-CASP domains (Fig. 1a). In fact, the β-CASP domain severely restricts access to the active site (Supplementary Fig. 8), and the scissile phosphate group probably cannot reach the zinc ions in the current structure. This indicates that a conformational change in the enzyme might be required for pre-mRNA binding (see Supplementary Information).

To show that CPSF-73 has endonuclease activity *in vitro* and to provide direct evidence for its role in 3′-end processing, we assessed the nuclease activity of the bacterially expressed CPSF-73, together with a derivative containing a double mutation (D75K/H76A) that destroys two of the zinc ligands in motif 2. We used a 5′-capped, uniformly labelled RNA containing the SV40 late polyadenylation site (SVL) as the substrate[13,27]. Using standard conditions for 3′ processing reactions, there was evidence for weak cleavage with the wild-type but not the mutant derivative (Supplementary Fig. 10a); a similarly prepared CPSF-100 derivative was also inactive (results not shown). This induced us to attempt to optimize conditions to obtain more efficient cleavage.

Among the possibilities tested to increase nuclease activity was preincubation with divalent cations. Our structural studies showed that there are binding sites for additional Zn^{2+} on the surface (see Supplementary Fig. 7), and it is possible that (non-specific) binding of divalent cations to the protein surface could affect its activity. Preincubation with Zn^{2+} and all other divalent cations tested, except Ca^{2+}, resulted in precipitation of the enzyme. Strikingly, after preincubation with Ca^{2+}, the wild-type enzyme degraded the SVL substrate almost entirely to oligonucleotides, whereas the mutant was still inactive (Fig. 4a). A similar-sized RNA from adenovirus was also entirely degraded by wild-type but not mutant CPSF-73 (Fig. 4a). Preincubation with Ca^{2+}, although required for enzyme activation (results not shown), was unnecessary for catalysis. It was diluted to 50 μM in the reactions and nuclease activity was not sensitive to the addition of 5 mM EGTA (results not shown). Furthermore, there is no evidence that Ca^{2+} is required in authentic 3′ processing, indicating that its activating function is probably performed by other components of the polyadenylation machinery. At lower CPSF-73 enzyme levels, partially degraded, higher-molecular-weight intermediates were seen (see Supplementary Fig. 10b). Significant cleavage required relatively high concentrations (10 ng μl^{-1} or ~200 nM), which we attribute to only a small fraction of the enzyme being activated by the preincubation.

Although our data are consistent with purified CPSF-73 possessing non-specific endonuclease activity, they do not conclusively rule out that it might have exonuclease activity, as previously suggested[14]. Indeed, there is evidence that Artemis has exonuclease activity[28]. Although the presence of a 5′ cap on the RNAs analysed above largely excludes 5′→3′ exonuclease activity, we nonetheless compared the ability of CPSF-73 to degrade 5′ and 3′ end-labelled RNA (Fig. 4b). The two RNAs were degraded with comparable kinetics and generated similar patterns of intermediates, consistent only with endonuclease activity.

These results, together with previous data[13,14], establish CPSF-73 as a strong, sequence-independent endonuclease that functions with specificity in 3′-end processing of both polyadenylated and histone pre-mRNAs by interacting with other factors. In addition, our data indicate that the close sequence homologue of CPSF-73, RC-68 or Int11, is probably the endonuclease in the 3′-end processing of small nuclear RNAs[24].

METHODS

Protein expression, purification and crystallization. Residues 1–460 of human CPSF-73 were sub-cloned into the pET28a vector (Novagen) and overexpressed in *Escherichia coli* at 20 °C. The soluble protein was purified by nickel–agarose

affinity chromatography and gel-filtration chromatography. Yeast CPSF-100 (Ydh1, residues 1–720) was expressed and purified by nickel–agarose affinity chromatography, anion exchange and gel-filtration chromatography.

Crystals of human CPSF-73 were obtained at room temperature by the sitting-drop vapour diffusion method. In an attempt to increase the occupancy of zinc ions, some crystals were grown from a reservoir solution that also contained 0.5 mM $ZnCl_2$. Crystals of yeast CPSF-100 were obtained at 4 °C by the same protocol. A solution that had been infected by a fungus was crucial for the crystallization of this protein[16]. The crystals were frozen in liquid propane for data collection at 100 K.

Data collection, structure determination and refinement. X-ray diffraction data were collected on an ADSC CCD at the X4A beamline of Brookhaven National Laboratory. The diffraction data were collected at the zinc absorption peak for crystals of CPSF-73, and at the gold absorption peak for a $KAu(CN)_2$ derivative of CPSF-100. The structure of CPSF-73 was determined by the single-wavelength anomalous diffraction method[29], using the anomalous signal of zinc. The structure of CPSF-100 was determined by the single-isomorphous replacement method, supplemented with anomalous diffraction. The crystallographic information are summarized in Supplemental Table 2. Figs 1–3 were produced with Ribbons[30].

Pre-mRNA 3′-end cleavage assay. CPSF-73 was pre-incubated with 5 mM $CaCl_2$ at 37 °C for 30 min. Cleavage assays were carried out in reaction mixture (10 μl) containing ~1 ng labelled RNA substrates, 10 mM HEPES (pH 7.9), 10% (v/v) glycerol, 50 mM KCl, 0.25 mM DTT, 0.25 mM PMSF, 0.125 mM EDTA, 50 μM $CaCl_2$, RNase inhibitor (2 units), 500 ng BSA and indicated amounts of recombinant CPSF-73. Cleaved RNAs were isolated and fractionated on 6% urea PAGE. The data were analysed by Phospho-Imager.

Received 18 July; accepted 20 October 2006.
Published online 26 November 2006.

1. Colgan, D. F. & Manley, J. L. Mechanism and regulation of mRNA polyadenylation. *Genes Dev.* **11,** 2755–2766 (1997).
2. Wahle, E. & Ruegsegger, U. 3′-end processing of pre-mRNA in eukaryotes. *FEMS Microbiol. Rev.* **23,** 277–295 (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. Proudfoot, N. J. New perspectives on connecting messenger RNA 3′ end formation to transcription. *Curr. Opin. Cell Biol.* **16,** 272–278 (2004).
5. Zorio, D. A. R. & Bentley, D. The link between mRNA processing and transcription: communication works both ways. *Exp. Cell Res.* **296,** 91–97 (2004).
6. Wilusz, C. J., Wormington, M. & Peltz, S. W. The cap-to-tail guide to mRNA turnover. *Nature Rev. Mol. Cell Biol.* **2,** 237–246 (2001).
7. Calvo, O. & Manley, J. L. Strange bedfellows: polyadenylation factors at the promoter. *Genes Dev.* **17,** 1321–1327 (2003).
8. Vinciguerra, P. & Stutz, F. mRNA export: an assembly line from genes to nuclear pores. *Curr. Opin. Cell Biol.* **16,** 285–292 (2004).
9. Shatkin, A. J. & Manley, J. L. The ends of the affair: capping and polyadenylation. *Nature Struct. Biol.* **7,** 838–842 (2000).
10. Wickens, M. & Gonzalez, T. N. Knives, accomplices, and RNA. *Science* **306,** 1299–1300 (2004).
11. Daiyasu, H., Osaka, K., Ishino, Y. & Toh, H. Expansion of the zinc metallo-hydrolase family of the β-lactamase fold. *FEBS Lett.* **503,** 1–6 (2001).
12. Callebaut, I., Moshous, D., Mornon, J.-P. & de Villartay, J.-P. Metallo-β-lactamase fold within nucleic acids processing enzymes: the β-CASP family. *Nucleic Acids Res.* **30,** 3592–3601 (2002).
13. Ryan, K., Calvo, O. & Manley, J. L. Evidence that polyadenylation factor CPSF-73 is the mRNA 3′ processing endonuclease. *RNA* **10,** 565–573 (2004).
14. Dominski, Z., Yang, X.-C. & Marzluff, W. F. The polyadenylation factor CPSF-73 is involved in histone-pre-mRNA processing. *Cell* **123,** 37–48 (2005).
15. Kolev, N. G. & Steitz, J. A. Symplekin and multiple other polyadenylation factors participate in 3′-end maturation of histone mRNAs. *Genes Dev.* **19,** 2583–2592 (2005).
16. Mandel, C. R., Gebauer, D., Zhang, H. & Tong, L. A serendipitous discovery that *in situ* proteolysis is required for the crystallization of yeast CPSF-100 (Ydh1p). *Acta Crystallogr.* **F62,** 1041–1045 (2006).
17. Ullah, J. H. *et al.* The crystal structure of the L1 metallo-β-lactamase from *Stenotrophomonas maltophilia* at 1.7Å resolution. *J. Mol. Biol.* **284,** 125–136 (1998).
18. Holm, L. & Sander, C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233,** 123–138 (1993).
19. de la Sierra-Gallay, I. L., Pellegrini, O. & Condon, C. Structural basis for substrate binding, cleavage and allostery in the tRNA maturase RNase Z. *Nature* **433,** 657–661 (2005).
20. Ishii, R. *et al.* Crystal structure of the tRNA 3′ processing endoribonuclease tRNase Z from *Thermotoga maritima*. *J. Biol. Chem.* **280,** 14138–14144 (2005).
21. Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **1,** 945–951 (1982).
22. Hirose, Y. & Manley, J. L. Creatine phosphate, not ATP, is required for 3′ end cleavage of mammalian pre-mRNA *in vitro*. *J. Biol. Chem.* **272,** 29636–29642 (1997).
23. Garrity, J. D., Bennett, B. & Crowder, M. W. Direct evidence that the reaction intermediate of metallo-β-lactamase L1 is metal bound. *Biochemistry* **44,** 1078–1087 (2005).
24. Baillat, D. *et al.* Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* **123,** 265–276 (2005).
25. Dominski, Z., Yang, X.-C., Purdy, M., Wagner, E. J. & Marzluff, W. F. A. CPSF-73 homologue is required for cell cycle progression but not cell growth and interacts with a protein having features of CPSF-100. *Mol. Cell. Biol.* **25,** 1489–1500 (2005).
26. Moshous, D. *et al.* Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency. *Cell* **105,** 177–186 (2001).
27. 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).
28. Ma, Y., Pannicke, U., Schwarz, K. & Lieber, M. R. Hairpin opening and overhand processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell* **108,** 781–794 (2002).
29. Hendrickson, W. A. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254,** 51–58 (1991).
30. Carson, M. Ribbon models of macromolecules. *J. Mol. Graph.* **5,** 103–106 (1987).

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

Acknowledgements We thank K. Ryan for discussions; B. Tweel for characterizing the fungus; R. Abramowitz, J. Schwanof and X. Yang for setting up the X4A beamline at the NSLS; and J. Khan and Y. Shen for help with data collection at the synchrotron. This research is supported in part by grants from the National Institutes of Health.

Author Information The atomic coordinates have been deposited at the Protein Data Bank (accession numbers 2I7T, 2I7V and 2I7X). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.T. (ltong@columbia.edu).