Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallographic analysis of the conserved C-terminal domain of transcription factor Cdc73 from *Saccharomyces cerevisiae* reveals a GTPase-like fold

Hongkai Chen,[a,b] Nuo Shi,[a,b] Yongxiang Gao,[a,b] Xu Li,[a,b] Maikun Teng[a,b]* and Liwen Niu[a,b]*

[a]Hefei National Laboratory for Physical Sciences at the Microscale and School of Life Sciences, University of Science and Technology of China, Hefei, Anhui 230026, People's Republic of China, and [b]Key Laboratory of Structural Biology, Chinese Academy of Sciences, Hefei, Anhui 230026, People's Republic of China

Correspondence e-mail: mkteng@ustc.edu.cn, lwniu@ustc.edu.cn

Received 7 February 2012
Accepted 18 April 2012

PDB Reference: conserved domain of yeast Cdc73, 4dm4.

The yeast Paf1 complex (Paf1C), which is composed of the proteins Paf1, Cdc73, Ctr9, Leo1 and Rtf1, accompanies RNA polymerase II from the promoter to the 3′-end formation site of mRNA- and snoRNA-encoding genes. As one of the first identified subunits of Paf1C, yeast Cdc73 (yCdc73) takes part in many transcription-related processes, including binding to RNA polymerase II, recruitment and activation of histone-modification factors and communication with other transcriptional activators. The human homologue of yCdc73, parafibromin, has been identified as a tumour suppressor linked to breast, renal and gastric cancers. However, the functional mechanism of yCdc73 has until recently been unclear. Here, a 2.2 Å resolution crystal structure of the highly conserved C-terminal region of yCdc73 is reported. It revealed that yCdc73 appears to have a GTPase-like fold. However, no GTPase activity was observed. The crystal structure of yCdc73 will shed new light on the modes of function of Cdc73 and Paf1C.

1. Introduction

Eukaryotic transcription is a highly regulated process mediated by the RNA polymerase II holoenzyme, which is composed of RNA polymerase II and many other accessory proteins, including TBP, TFIIB, TFIIE, TFIIF and TFIIH (Orphanides *et al.*, 1996; Buratowski, 1994). In addition to their physical and genetic interactions, the subunits of the RNA polymerase II holoenzyme also transmit information between themselves. The holoenzyme exists in multiple functional forms *in vivo* (Shi *et al.*, 1997), one of which contains Paf1 and Cdc73. However, only this form of the holoenzyme is involved in full expression of a subset of cell-cycle-regulated genes (Porter *et al.*, 2002).

The Paf1 complex (Paf1C), which is composed of the proteins Paf1, Ctr9, Cdc73, Rtf1 and Leo1, accompanies RNA polymerase II from the promoter to the 3′-end formation site of mRNA- and snoRNA-encoding genes; it is also found associated with RNA polymerase I on rDNA (Zhang *et al.*, 2009; Pokholok *et al.*, 2002; Sheldon *et al.*, 2005). Paf1 and Cdc73 were first found as RNA polymerase II-associated proteins (RAPs; Wade *et al.*, 1996). Ctr9, Rtf1 and Leo1 were subsequently found to be the other members of Paf1C (Koch *et al.*, 1999; Squazzo *et al.*, 2002). Paf1C also has genetic and physical links with elongation factors, including Spt4/5 (DSIF complex), Spt16/Pob3 (FACT complex) and TFIIS (Squazzo *et al.*, 2002). Genetic and physical links between Paf1C and COMPASS, which is required for methylation of histone H3

on lysines 4 and 79, have subsequently established that Paf1C participates in histone methylation (Krogan *et al.*, 2003). Recent research has found that Paf1C genes can also affect telomere length (Mozdy *et al.*, 2008).

Cdc73 was first described as a cell-division cycle protein related to the yeast mating-pheromone signalling pathway (Reed *et al.*, 1988). As an important part of Paf1C, it participates in the functional activities of Paf1C; for example, it plays a direct role in H3 Lys36 methylation (Chu *et al.*, 2007). It can also affect telomere length by decreasing Tlc1 (telomerase component 1) levels. When Cdc73 is lost from the complex, Paf1C dissociates from RNA polymerase II (Mueller *et al.*, 2004). The homologue of yCdc73 in human Paf1C is parafibromin, which has been identified as a tumour suppressor linked to breast, renal and gastric cancers (Newey *et al.*, 2009). Several other functions have been described in eukaryotes. In flies, dPaf1C directly contacts the transcriptional activator β-catenin (the key protein in the Wnt pathway) through interactions with dCdc73 (Hyrax; Mosimann *et al.*, 2006). In mice, Cdc73 directly regulates genes involved in cell growth (Wang *et al.*, 2008).

Although diverse functions of Paf1C have been reported, little is known about its functional mechanism owing to a lack of structural information. In this study, we targeted yCdc73 for structural studies. Primary-sequence analyses revealed that yCdc73 does not contain an easily identified domain in the N-terminal region, but shows a C-terminal region that is conserved across species and shares 27% identity and 45% similarity to that of parafibromin. Here, we report the crystal structure of the C-terminal domain of yCdc73 (yCdc73_C). Our study revealed that yCdc73_C has a typical small GTPase-like fold, although no hydrolytic activity was observed in a GTPase-activity assay. Structural comparison of yCdc73_C with several typical small GTPase-family proteins revealed that yCdc73_C does not contain the conserved motifs that are responsible for GTP hydrolysis in the typical small GTPase-family proteins in its corresponding regions, which may account for the lack of any detectable GTPase activity.

2. Materials and methods

2.1. Cloning, expression and purification

The full-length *Saccharomyces cerevisiae* Cdc73 gene and a C-terminal fragment (residues 235–393; yCdc73_C) were amplified by PCR from the yeast genome. Both gene fragments were subcloned into the *Nde*I–*Xho*I sites of the pET-22b vector (Novagen). The pET-22b vector adds a C-terminal 6×His tag to the recombinantly expressed protein. The yCdc73 and yCdc73_C constructs were transformed into *Escherichia coli* BL21 (DE3) cells, which were grown in LB medium containing ampicillin (100 µg ml^{-1}) overnight at 310 K. When the absorbance at 600 nm reached 0.6–0.8, protein expression was induced by adding 0.25 m*M* isopropyl β-D-1-thiogalactopyranoside at 289 K. After 20 h induction, the cells were harvested by centrifugation at 6080*g* for 6 min. The cell pellets were resuspended in ice-cold 20 m*M* Tris–HCl pH 7.5 containing 500 m*M* NaCl and the resuspended cells were lysed by sonication. Cell supernatants were obtained by centrifugation at 20 000*g* for 30 min at 277 K and the soluble lysate was purified using a nickel-chelating column (GE Healthcare) and a HiLoad 16/60 Superdex 200 size-exclusion column (GE Healthcare). The peak fractions were concentrated to 15 mg ml^{-1} and stored at 193 K.

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

	Native	SAD
Data collection		
Wavelength (Å)	1.5417	1.5000
Space group	*I*222	*I*222
Unit-cell parameters (Å)	a = 90.94, b = 96.89, c = 114.03	a = 91.85, b = 93.55, c = 113.27
Resolution range (Å)	50–2.20 (2.24–2.20)	50–2.50 (2.54–2.50)
Unique reflections	26095 (1281)	17202 (813)
Completeness (%)	100 (99.8)	99.5 (94.2)
Multiplicity	6.4 (5.6)	7.4 (5.2)
Average $I/\sigma(I)$	36.6 (5.7)	43.3 (4.2)
$R_{\rm merge}$†	0.06 (0.39)	0.06 (0.34)
Refinement statistics		
Resolution range (Å)	50.0–2.20 (2.28–2.20)	
No. of reflections	24765	
$R_{\rm work}$‡ (%)	22.0	
$R_{\rm free}$§ (%)	26.1	
R.m.s.d. bond lengths (Å)	0.009	
R.m.s.d. bond angles (°)	1.074	
Mean B factor (Å^2)	20.3	
Ramachandran plot¶, residues in		
Most favoured regions (%)	98.6	
Allowed regions (%)	1.4	

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$. ‡ $R = \sum_{hkl}\big||F_{\rm obs}| - |F_{\rm calc}|\big|/\sum_{hkl}|F_{\rm obs}|$. § $R_{\rm free}$ is the R factor calculated using 5% of the data that were excluded from refinement. ¶ Categories were defined by *MolProbity*.

2.2. Crystallization and data collection

Crystals of yCdc73_C were grown using the hanging-drop vapour-diffusion method. 1 µl protein sample at a concentration of ∼15 mg ml^{-1} in 20 m*M* Tris–HCl pH 7.5, 500 m*M* NaCl was mixed with 1 µl reservoir solution (2.7 *M* sodium formate pH 7.0). Crystals appeared after 4 d at 287 K.

A set of native data was collected to 2.2 Å resolution using in-house Cu $K\alpha$ X-rays generated by a Rigaku MicroMax-007 rotating-anode X-ray source; the diffraction images were collected using a MAR 345 dtb imaging-plate detector (MAR Research).

Crystals were soaked in 0.5 *M* KI solution for several seconds and were then quickly transferred to crystallization buffer with 20% glycerol. A SAD data set was collected on beamline 3W1A of Beijing Synchrotron Radiation Facility (BSRF) at the Institute of High Energy Physics, Chinese Academy of Sciences. The data were collected at a wavelength of 1.5 Å. The SAD data and the native data were processed using *HKL*-2000 (Otwinowski & Minor, 1997). The native and derivative crystals both belonged to space group *I*222, with similar unit-cell parameters. Data-collection statistics are presented in Table 1.

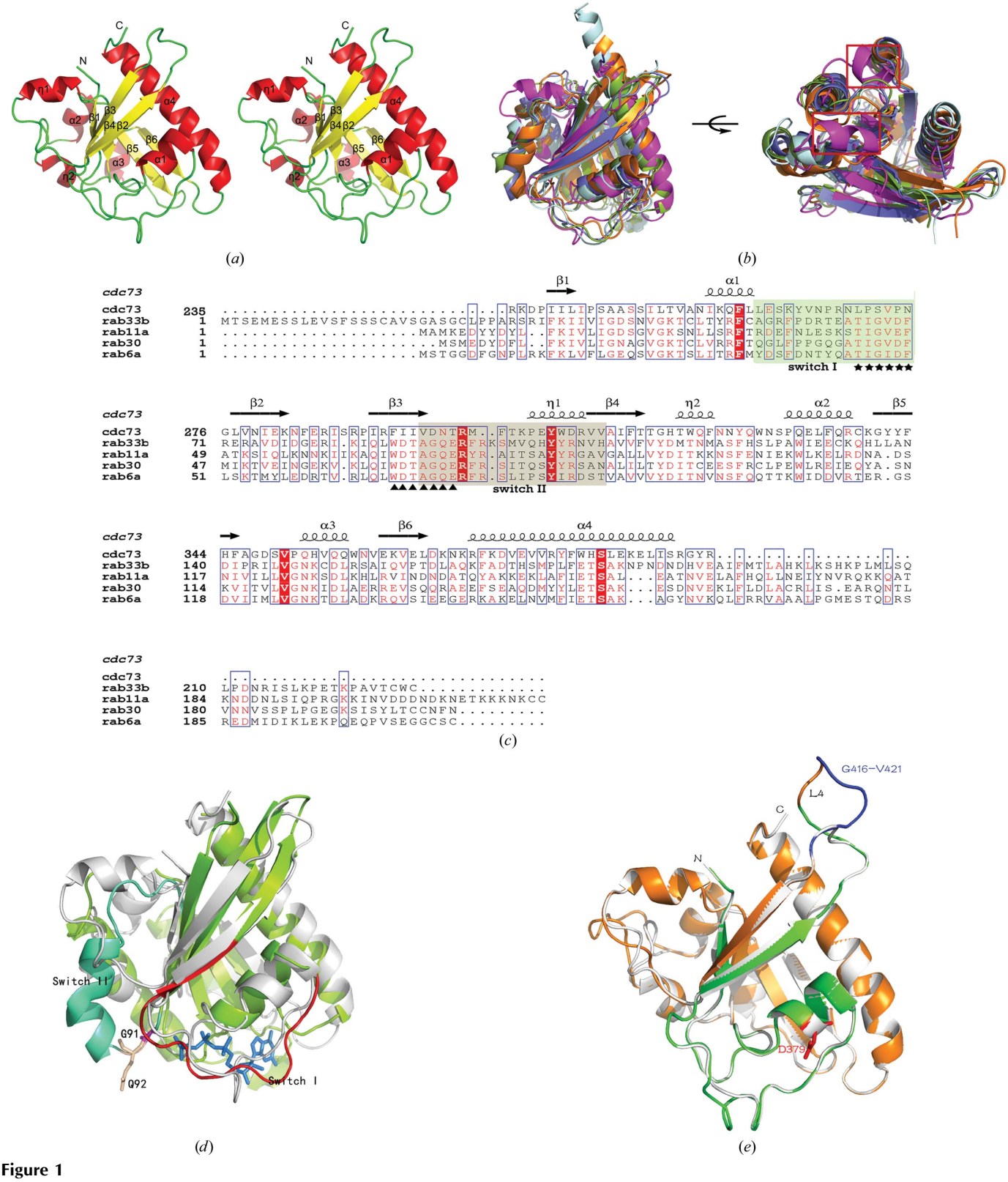

Figure 1

(*a*) Ribbon diagram of yCdc73_C. (*b*) Comparison of yCdc73_C (magenta) and several Rab-family proteins (Rab33B, Rab30, Rab11a and Rab6a, coloured green, orange, pale cyan and blue, respectively). (*c*) Sequence alignment between yCdc73_C and some members of the Rab family aligned using the *ClustalW*2 tool and coloured using *ESPript*. The green and grey backgrounds indicate the switch I and switch II regions, respectively. The WDTAGQE and TiGxdF motifs are marked with stars and triangles, respectively. (*d*) Superposition of yCdc73_C (grey) and Rab33 (green). The switch I and switch II regions of Rab33B are shown in red and cyan, respectively. The residues Gly91 and Gln92 are shown in light pink. GDP and AlF_4^- are shown in sky blue. (*e*) Alignment of the cartoon skeletons of yCdc73_C and the hCdc73_C model derived from *SWISS-MODEL*. The yCdc73_C structure and the model of hCdc73_C are coloured grey and orange, respectively. The residues Gly416–Val421 are shown in blue and residue Asp379 is shown in red. The green regions (358–413) of the hCdc73_C model represent the binding zone that interacts with other subunits of the PAF1 complex.

2.3. Structure determination and refinement

Heavy-atom location and primary phasing were carried out using *SOLVE* (Terwilliger & Berendzen, 1999). Phase modification by solvent flattening was carried out using *RESOLVE* (Terwilliger, 2003). The initial model was automatically built and refined using *Buccaneer* from the *CCP*4 package (Cowtan, 2006; Winn *et al.*, 2011). Refinement was continued using 2.2 Å resolution data from the native crystal. The complete model was built using *Coot* (Emsley & Cowtan, 2004) and refinement was carried out using *REFMAC*5 (Murshudov *et al.*, 2011) until the crystallographic *R* factor and free *R* factor converged to 21.9% and 26.1%, respectively. The stereochemistry of the structure was checked using *PROCHECK* (Laskowski *et al.*, 1993). Figures were prepared using *PyMOL* (DeLano, 2002).

2.4. Structural model of the C-terminal domain of hCdc73

A homology model of parafibromin (hCdc73_C) was built by *SWISS-MODEL* (http://swissmodel.expasy.org/?pid=smh01) using the structure of yCdc73_C as a template (Arnold *et al.*, 2006). The sequence identity between hCdc73_C and yCdc73_C is 27% and the QMEAN4 global score, which is a composite score consisting of a linear combination of four statistical potential terms (giving an estimated model reliability of between 0 and 1), is 0.639 (Benkert *et al.*, 2011).

3. Results and discussion

3.1. Structure determination

Full-length yCdc73 that had been purified to homogeneity was subjected to crystal screening; however, no crystals could be obtained. Based on its sequence conservation, we targeted the C-terminal domain (yCdc73_C) for structure determination and a complete data set was collected and indexed in space group *I*222. In the absence of any known homologous model, we prepared an iodine derivative by soaking the native crystal in KI. Phases were calculated using SAD data collected from the iodine derivative. The initial model was then refined against the native data at 2.2 Å resolution. There were two molecules in the asymmetric unit. The majority of the residues were reliably modelled in the electron-density map, except for the C-terminal His_6 tag in both subunits. 98.6% of the residues in the final refined model had main-chain torsion angles in the most favoured regions; no residues were in disallowed regions. The data-collection and refinement statistics are summarized in Table 1.

3.2. Overall structure

yCdc73_C is an approximately spherical molecule with dimensions of 42 × 39 × 47 Å. The structure adopts an α/β-fold consisting of a six-stranded β-sheet, with five parallel strands and one antiparallel strand, flanked by six α-helices (Fig. 1*a*). The overall structure of yCdc73_C closely resembles those of Rab-family proteins as revealed by the *DALI* server (Holm & Rosenström, 2010). As shown in Fig. 1(*b*), structural superposition of yCdc73_C onto Rab33B, Rab6a, Rab11a and Rab30 shows that yCdc73_C has overall structural similarity to the Rab-family proteins, although there are some differences in two α-helices, as highlighted by the red rectangular box. Rab33B, which is the most similar, superimposes onto yCdc73_C with an r.m.s.d. of 3.1 Å.

3.3. Structural comparison of yCdc73_C with Rab33B

Structural studies of a number of GTP-binding proteins have shown that a conserved architecture exists for efficient GTP binding and hydrolytic activity. Rab GTPases, which belong to the Ras superfamily of small GTPases, play central roles in regulating vesicle budding, motility and fusion. Like other regulatory GTPases, the Rab proteins switch between two distinct conformations: GTP-bound and GDP-bound. The main differences between the two conformations occur in the regions denoted switch I and switch II (Schlichting *et al.*, 1990; Pan *et al.*, 2006). The *Mus musculus* Rab33B protein is a member of the Rab GTPase family. As shown in Fig. 1(*d*), although structural superposition of yCdc73_C on Rab33B shows an overall structural similarity, the Rab33B switch I and switch II regions show great differences when compared with the corresponding regions of yCdc73_C. The WDTAGQE sequence motif encompassing the C-terminal end of β3 and the N-terminal portion of switch II (Lee *et al.*, 2008) is >90% conserved among Rab GTPases (Fig. 1*c*). This motif, which is located close to the catalytic centre, is responsible for contacting the γ-phosphate of the nucleotide and undergoes conformational changes between the active state and the inactive state. In the structure of yCdc73_C, however, the corresponding region of switch II is distant from the supposed catalytic centre. The conserved motif TiG*x*dF (in which uppercase and lowercase letters indicate >90% and >65% sequence conservation) in the switch I region also contacts the γ-phosphate of the nucleotide (Lee *et al.*, 2008). The Gln92 and Gly91 residues of Rab33B, which are highly conserved in the Rab GTPase family and play key roles in GTP hydrolysis, do not exist in the corresponding region of yCdc73_C.

3.4. Is yCdc73 a GTPase?

The structural resemblance between yCdc73_C and Rab GTPase prompted us to speculate that yCdc73 possesses GTPase activity. We thus addressed this question by using yCdc73_C and full-length yCdc73 (yCdc73_full) in functional assays. However, no obvious enzyme activity was observed in a series of enzyme assays (Supplementary Fig. 2[1]) and no interaction was detected between GTP and yCdc73_full/yCdc73_C using the ITC (isothermal titration calorimetry) method (data not shown).

In small GTPase family proteins, the GTP-bound conformation is the form that interacts with downstream effector proteins and is generally regarded as 'active', while the GDP-bound conformation is often regarded as 'inactive'. The conformational changes between these two forms are confined

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: BE5203). Services for accessing this material are described at the back of the journal.

to the switch I and switch II regions. The high affinity of GTPases for effectors is a consequence of the conformation of these two switch motifs. However, the corresponding regions of yCdc73_C differ from those of typical small GTPases. For example, the surface-charge distribution of Rab33B shows a channel with positive potential at the GTP-bound site, but yCdc73_C does not have such a channel in the corresponding region (Fig. 2*a*), which is consistent with the lack of unassigned density at the assumed nucleotide-binding site even after several cycles of density modification. Furthermore, the low *B* factors for the residues corresponding to switch I and switch II are comparable to those of the residues located in the close vicinity. The different conformations and low *B* factors may explain why yCdc73 makes no interaction with GTP and suggests that the 'switch I and switch II' motifs in yCdc73 may be degenerate GTPase hydrolysis motifs that mediate protein–protein interactions.

Based on the structural analysis and biochemical data, we suggest the following reasons as to why yCdc73 has no detectable GTPase activity. Firstly, although there is an overall structural similarity between yCdc73_C and Rab33b, they have low sequence identity (11%); moreover, yCdc73 does not possess the features of the canonical switch I and switch II motifs in either primary sequence or tertiary structure. Secondly, since GTPases usually require GTPase-activating proteins (GAPs) to stimulate activity, yCdc73 may possess low intrinsic activity and may require other proteins or cofactors to perform GTPase activity.

3.5. Structural implications

As mentioned above, the N-terminal region of yCdc73 is not conserved and has been reported not to be essential for yCdc73 function under normal growth conditions. However, the conservation presented by the yCdc73_C domain suggests that it fulfils a special function, which is exemplified by human parafibromin mutations causing hyperparathyroidism–jaw tumour syndrome (HPT-JT), an autosomal dominant disease characterized by the occurrence of parathyroid tumours and ossifying fibromas of the maxilla/mandible. More than 100 mutations in the *hCDC73* gene have been identified, of which seven occur in the conserved C-terminal domain (Carpten *et al.*, 2002; Teh *et al.*1998; Szabó *et al.*, 1995, Jackson *et al.*, 1990). Over 80% of the mutations are predicted to cause premature protein truncation and include five mutations (four frameshift mutations and one nonsense mutation) in the C-terminal domain, reflecting the importance of the C-terminal domain of hCdc73 (Supplementary Table 1) (Newey *et al.*, 2009, 2010; Siu *et al.*, 2011). Our crystal structure establishes the structure of this conserved domain for the first time.

A three-dimensional model of hCdc73_C was built using the yCdc73_C structure as the template. As shown in Fig. 2(*c*), there are four frameshift mutations (Asn376LeufsX10, Gln411ArgfsX17, Gln413HisfsX15 and Leu478GlufsX3) and one nonsense mutation (Trp492X) that cause truncations in hCdc73_C. Based on the model of hCdc73_C, the missense mutation (Asp379Asn) is predicted to be located within a

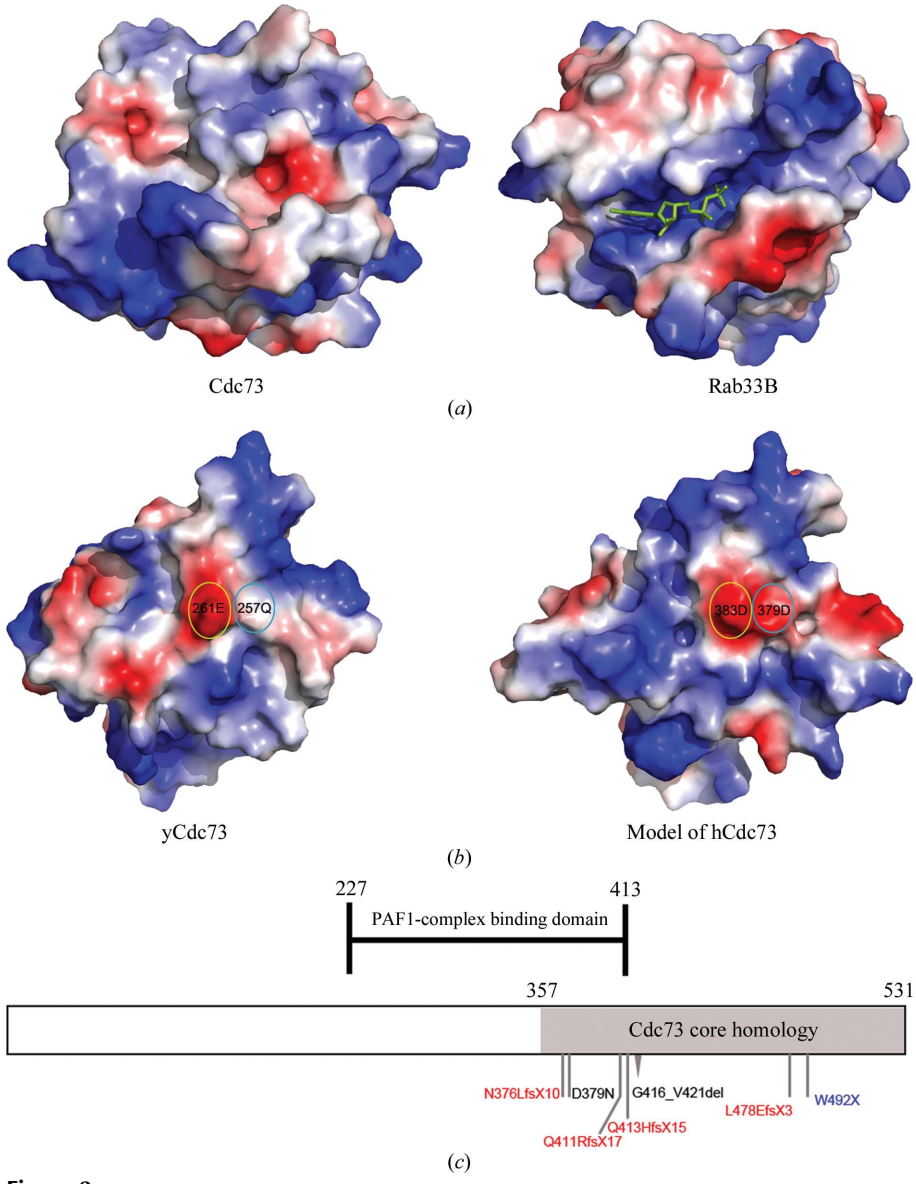

Figure 2
(*a*) Surface-charge diagrams of yCdc73_C and Rab33B. GDP and AlF_4^- were located in the negatively charged channel of Rab33B. The surface-charge diagrams were drawn in the same orientation after superposition. (*b*) Surface-charge diagrams of yCdc73 and the model of hCdc73. The cyan-coloured circles indicate the surfaces of Gln257 and Asp379. The yellow-coloured circles indicate the surfaces of Glu261 and Asp381. (*c*) Schematic representation of hCdc73.

region that interacts with another Paf1C subunit. Although Gln257 in yeast Cdc73 is not an acidic residue as is the human Asp379, the existence of Glu261 yields a negatively charged surface area (Fig. 2*b*, Supplementary Fig. 1). Similarly, hCdc73 also presents a negatively charged surface at a similar region contributed by Asp379 and Asp383. Thus, the Asp379Asn mutation may significantly alter the distribution of surface charge. Given the conservation of electrostatic surface potential at this site, we speculate that this region is an important protein–protein interaction site and thus that the Asp379Asn mutation would cause reduced affinity between human Cdc73 and its binding partners. The deletion mutation Gly416_Val421del causes a truncation of Cdc73 loop4, which may also reside in a protein–protein interface of the Paf1 complex [as shown in Fig. 1(*e*), the binding regions (residues 358–413) form an interaction surface, with the exception of the first β-sheet]. The disruption of protein–protein interactions caused by the above mutations may explain their pathogenicity in humans.

4. Conclusion

Cdc73, which is one subunit of Paf1C, is important for the transcription of a subset of genes. The C-terminal domain of yCdc73 is conserved and may have an important function that has yet to be described. Here, we determined the crystal structure of the C-terminal domain of Cdc73 from *S. cerevisiae* at 2.2 Å resolution and compared it with the structure of Rab33B. Although the sequence identity between the C-terminal domain of yCdc73 and Rab33B is only 11.4%, their three-dimensional structures are similar. We did not find any GTPase activity for yCdc73 using a series of enzyme assays. A structural comparison of yCdc73_C with small GTP-binding proteins also indicated some differences in the conserved catalytic regions. It remains a puzzle whether yCdc73 has GTPase activity. It may be that the ancestor of Cdc73 was indeed a GTPase but lost its GTPase activity during evolution. We built a three-dimensional model of the C-terminal domain of human Cdc73 (hCdc73_C) and mapped several mutations onto hCdc73_C, which might provide some clues to explain why some mutations of hCdc73_C are observed in tumours. We hope that our results will shed light on further functional studies of Cdc73 and Paf1C.

We appreciate the assistance from the Beijing Synchrotron Radiation Facility. Financial support for this project was provided by research grants from the Chinese Ministry of Science and Technology (grant Nos. 2012CB917200 and 2009CB825500), the Chinese National Natural Science Foundation (grant Nos. 31130018 and 31170726) and the Natural Science Foundation of Anhui Province (grant No. 090413085).

References

Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. (2006). *Bioinformatics*, **22**, 195–201.
Benkert, P., Biasini, M. & Schwede, T. (2011). *Bioinformatics*, **27**, 343–350.
Buratowski, S. (1994). *Cell*, **77**, 1–3.
Carpten, J. D. *et al.* (2002). *Nature Genet.* **32**, 676–680.
Chu, Y., Simic, R., Warner, M. H., Arndt, K. M. & Prelich, G. (2007). *EMBO J.* **26**, 4646–4656.
Cowtan, K. (2006). *Acta Cryst.* D**62**, 1002–1011.
DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
Jackson, C. E., Norum, R. A., Boyd, S. B., Talpos, G. B., Wilson, S. D., Taggart, R. T. & Mallette, L. E. (1990). *Surgery*, **108**, 1006–1012.
Koch, C., Wollmann, P., Dahl, M. & Lottspeich, F. (1999). *Nucleic Acids Res.* **27**, 2126–2134.
Krogan, N. J., Dover, J., Wood, A., Schneider, J., Heidt, J., Boateng, M. A., Dean, K., Ryan, O. W., Golshani, A., Johnston, M., Greenblatt, J. F. & Shilatifard, A. (2003). *Mol. Cell*, **11**, 721–729.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lee, S. H., Baek, K. & Dominguez, R. (2008). *FEBS Lett.* **582**, 4107–4111.
Mosimann, C., Hausmann, G. & Basler, K. (2006). *Cell*, **125**, 327–341.
Mozdy, A. D., Podell, E. R. & Cech, T. R. (2008). *Mol. Cell. Biol.* **28**, 4152–4161.
Mueller, C. L., Porter, S. E., Hoffman, M. G. & Jaehning, J. A. (2004). *Mol. Cell*, **14**, 447–456.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Newey, P. J., Bowl, M. R., Cranston, T. & Thakker, R. V. (2010). *Hum. Mutat.* **31**, 295–307.
Newey, P. J., Bowl, M. R. & Thakker, R. V. (2009). *J. Intern. Med.* **266**, 84–98.
Orphanides, G., Lagrange, T. & Reinberg, D. (1996). *Genes Dev.* **10**, 2657–2683.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pan, X., Eathiraj, S., Munson, M. & Lambright, D. G. (2006). *Nature (London)*, **442**, 303–306.
Pokholok, D. K., Hannett, N. M. & Young, R. A. (2002). *Mol. Cell*, **9**, 799–809.
Porter, S. E., Washburn, T. M., Chang, M. & Jaehning, J. A. (2002). *Eukaryot. Cell*, **1**, 830–842.
Reed, S. I., Ferguson, J. & Jahng, K. Y. (1988). *Cold Spring Harb. Symp. Quant. Biol.* **53**, 621–627.
Schlichting, I., Almo, S. C., Rapp, G., Wilson, K., Petratos, K., Lentfer, A., Wittinghofer, A., Kabsch, W., Pai, E. F., Petsko, G. A. & Goody, R. S. (1990). *Nature (London)*, **345**, 309–315.
Sheldon, K. E., Mauger, D. M. & Arndt, K. M. (2005). *Mol. Cell*, **20**, 225–236.
Shi, X., Chang, M., Wolf, A. J., Chang, C.-H., Frazer-Abel, A. A., Wade, P. A., Burton, Z. F. & Jaehning, J. A. (1997). *Mol. Cell. Biol.* **17**, 1160–1169.
Siu, W. K., Law, C. Y., Lam, C. W., Mak, C. M., Wong, G. W. K., Ho, A. Y. Y., Ho, K. Y., Loo, K. T., Chiu, S. C., Chow, L. T. C., Tong, S. F. & Chan, A. Y. (2011). *Fam. Cancer*, **10**, 695–699.
Squazzo, S. L., Costa, P. J., Lindstrom, D. L., Kumer, K. E., Simic, R., Jennings, J. L., Link, A. J., Arndt, K. M. & Hartzog, G. A. (2002). *EMBO J.* **21**, 1764–1774.
Szabó, J., Heath, B., Hill, V. M., Jackson, C. E., Zarbo, R. J., Mallette, L. E., Chew, S. L., Besser, G. M., Thakker, R. V., Huff, V., Leppert, M. F. & Heath, H. (1995). *Am. J. Hum. Genet.* **56**, 944–950.
Teh, B. T., Farnebo, F., Twigg, S., Höög, A., Kytölä, S., Korpi-Hyövälti, E., Wong, F. K., Nordenström, J., Grimelius, L., Sandelin, K., Robinson, B., Farnebo, L.-O. & Larsson, C. (1998). *J. Clin. Endocrinol. Metab.* **83**, 2114–2120.
Terwilliger, T. C. (2003). *Acta Cryst.* D**59**, 38–44.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.

Wade, P. A., Werel, W., Fentzke, R. C., Thompson, N. E., Leykam, J. F., Burgess, R. R., Jaehning, J. A. & Burton, Z. F. (1996). *Protein Expr. Purif.* **8**, 85–90.
Wang, P., Bowl, M. R., Bender, S., Peng, J., Farber, L., Chen, J., Ali, A., Zhang, Z., Alberts, A. S., Thakker, R. V., Shilatifard, A., Williams, B. O. & Teh, B. T. (2008). *Mol. Cell. Biol.* **28**, 2930–2940.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Zhang, Y., Sikes, M. L., Beyer, A. L. & Schneider, D. A. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 2153–2158.