Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Expression, crystallization and preliminary X-ray crystallographic analyses of two N-terminal acetyltransferase-related proteins from *Thermoplasma acidophilum*

Sang Hee Han,[a]‡ Jun Yong Ha,[a]‡ Kyoung Hoon Kim,[a]‡ Sung Jin Oh,[a] Do Jin Kim,[a] Ji Yong Kang,[a] Hye Jin Yoon,[a] Se-Hee Kim,[b] Ji Hae Seo,[b] Kyu-Won Kim[b] and Se Won Suh[a]*

[a]Department of Chemistry, College of Natural Sciences, Seoul National University, Seoul 151-742, South Korea, and [b]NeuroVascular Coordination Research Center, Research Institute of Pharmaceutical Sciences, College of Pharmacy, Seoul National University, Seoul 151-742, South Korea

‡ These authors contributed equally to this work.

Correspondence e-mail: sewonsuh@snu.ac.kr

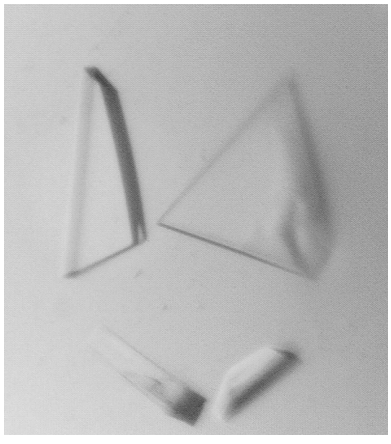
Received 15 September 2006
Accepted 30 September 2006

N-terminal acetylation is one of the most common protein modifications in eukaryotes, occurring in approximately 80–90% of cytosolic mammalian proteins and about 50% of yeast proteins. ARD1 (arrest-defective protein 1), together with NAT1 (*N*-acetyltransferase protein 1) and possibly NAT5, is responsible for the NatA activity in *Saccharomyces cerevisiae*. In mammals, ARD1 is involved in cell proliferation, neuronal development and cancer. Interestingly, it has been reported that mouse ARD1 ($mARD1^{225}$) mediates ε-acetylation of hypoxia-inducible factor 1α (HIF-1α) and thereby enhances HIF-1α ubiquitination and degradation. Here, the preliminary X-ray crystallographic analyses of two N-terminal acetyltransferase-related proteins encoded by the *Ta0058* and *Ta1140* genes of *Thermoplasma acidophilum* are reported. The Ta0058 protein is related to an N-terminal acetyltransferase complex ARD1 subunit, while Ta1140 is a putative N-terminal acetyltransferase-related protein. Ta0058 shows 26% amino-acid sequence identity to both $mARD1^{225}$ and human $ARD1^{235}$.The sequence identity between Ta0058 and Ta1140 is 28%. Ta0058 and Ta1140 were overexpressed in *Escherichia coli* fused with an N-terminal purification tag. Ta0058 was crystallized at 297 K using a reservoir solution consisting of 0.1 *M* sodium acetate pH 4.6, 8%(*w*/*v*) polyethylene glycol 4000 and 35%(*v*/*v*) glycerol. X-ray diffraction data were collected to 2.17 Å. The Ta0058 crystals belong to space group $P4_1$ (or $P4_3$), with unit-cell parameters $a = b = 49.334$, $c = 70.384$ Å, $\alpha = \beta = \gamma = 90^\circ$. The asymmetric unit contains a monomer, giving a calculated crystal volume per protein weight (V_M) of 2.13 $Å^3$ Da^{-1} and a solvent content of 42.1%. Ta1140 was also crystallized at 297 K using a reservoir solution consisting of 0.1 *M* trisodium citrate pH 5.6, 20%(*v*/*v*) 2-propanol, 20%(*w*/*v*) polyethylene glycol 4000 and 0.2 *M* sodium chloride. X-ray diffraction data were collected to 2.40 Å. The Ta1140 crystals belong to space group *R*3, with hexagonal unit-cell parameters $a = b = 75.174$, $c = 179.607$ Å, $\alpha = \beta = 90$, $\gamma = 120^\circ$. Two monomers are likely to be present in the asymmetric unit, with a V_M of 2.51 $Å^3$ Da^{-1} and a solvent content of 51.0%.

1. Introduction

N-terminal acetylation is one of the most common protein modifications in eukaryotes, occurring in approximately 80–90% of cytosolic mammalian proteins and about 50% of yeast proteins (Polevoda & Sherman, 2003). Studies with *Saccharomyces cerevisiae* revealed three major N-terminal acetyltransferases, NatA, NatB and NatC, which exhibit different substrate specificities (Polevoda & Sherman, 2003). ARD1 (arrest-defective protein 1), together with NAT1 (*N*-acetyltransferase protein 1), is required for NatA activity in yeast (Lee *et al.*, 1989; Mullen *et al.*, 1989; Park & Szostak, 1992). The yeast N-terminal acetyltransferase complex NatA was quantitatively bound to ribosomes *via* NAT1 and contained a third subunit, NAT5 (Gautschi *et al.*, 2003). Human homologues of ARD1 and NAT1 also form an evolutionarily conserved N-terminal acetyltransferase complex (Arnesen, Anderson *et al.*, 2005) and a human homologue of NAT5 has also been identified (Arnesen, Anderson *et al.*, 2006). In mammals, the N-terminal acetyltransferase complex plays an important role in cell proliferation, neuronal development and cancer (Sugiura *et al.*, 2003; Fisher *et al.*, 2005; Arnesen, Gromyko *et al.*, 2005).

ARD1 and NAT5 proteins have been assigned to the ARD1 subfamily of acetyltransferases. They possess an acetyl-CoA-binding



motif (Q/R)*XX*G*X*(G/A) (Neuwald & Landsman, 1997; Arg82–Ala87 in ARD1 and Arg84–Ala89 in NAT5 for both human and mouse). A recent study has identified three mouse (mARD1^{198}, mARD1^{225} and mARD1^{235}) and two human (hARD1^{131} and hARD1^{235}) variants of ARD1 (Kim *et al.*, 2006). Of the ARD1 isoforms, mARD1^{225} differs completely in its C-terminal region (residues 158–225) from mouse and human ARD1^{235} (Kim *et al.*, 2006). Interestingly, a mouse ARD1 variant (mARD1^{225}) was reported to mediate ε-acetylation of Lys532 within the oxygen-dependent degradation (ODD) domain of hypoxia-inducible factor 1α (HIF-1α), thereby enhancing HIF-1α ubiquitination and degradation (Jeong *et al.*, 2002; Kim *et al.*, 2006). However, other investigators could not demonstrate such an activity for hARD1^{235} (Fisher *et al.*, 2005; Bilton *et al.*, 2005). Overexpression or silencing of hARD1^{235} had no impact on HIF-1α stability (Arnesen, Kong *et al.*, 2005) and the recombinant hARD1^{235} did not catalyze the acetylation of HIF-1α ODD domain (Murray-Rust *et al.*, 2006). In constrast, Chang and coworkers recently provided strong evidence for the role of hARD1 in HIF-1α stability in human cells (Chang *et al.*, 2006). They suggested that connective tissue-growth factor (CTGF) mediated HIF-1α degradation was attributed to hARD1-dependent acetylation of HIF-1α at Lys532 in human lung cancer cells (Chang *et al.*, 2006). Structurally, hARD1^{235} consists of a compact globular region (residues 1–178) and a flexible unstructured C-terminus (Sanchez-Puig & Fersht, 2006). The globular region encompasses the predicted acetyltransferase domain of hARD1^{235} (residues 44–129; Arnesen, Betts *et al.*, 2006). It was also reported that hARD1^{235} forms protofilament aggregates under physiological conditions of pH and temperature (Sanchez-Puig & Fersht, 2006). In particular, the region encompassing the acetyltransferase domain is responsible for fibre formation (Sanchez-Puig & Fersht, 2006), which in turn may abrogate the catalytic activity of hARD1^{235}.

Despite the important functional role of protein N-terminal acetyltransferases, their structural studies are in an early stage. To provide a structural basis to obtain a better understanding of protein N^{α}-acetylation by N-terminal acetyltransferases and to facilitate structural comparisons with N^{ε}-acetyltransferases, we have initiated three-dimensional structure determination of the *Ta0058* gene product from *Thermoplasma acidophilum*. Its 154-residue protein is related to the N-terminal acetyltransferase complex ARD1 subunit. It shows amino-acid sequence identity of 26% with mARD1^{225}, mARD1^{235} and hARD1^{235}. We have also performed crystallization studies of the *Ta1140* gene product from *T. acidophilum*, a putative N-terminal acetyltransferase-related protein. Its 149-residue polypeptide chain shows sequence identity of 28% with Ta0058, but bears no detectable sequence homology with either mouse or human ARD1 variants. As the first step toward structure determination of the Ta0058 and Ta1140 proteins from *T. acidophilum*, we overexpressed them in *Escherichia coli* and crystallized them. Here, we report the crystallization conditions and preliminary X-ray crystallographic data. Both of these crystals are suitable for structure determination at sufficiently high resolution. It is hoped that the ultimate structural information on these proteins will provide a framework for obtaining a better insight into the molecular functions of N-terminal acetyltransferase-related proteins.

2. Experimental

2.1. Protein expression and purification

The *Ta0058* and *Ta1140* genes were amplified by polymerase chain reaction (PCR) using the genomic DNA of *T. acidophilum* as the template and the following oligonucleotide primers, which were designed using the published genome sequence (Ruepp *et al.*, 2000). The forward and reverse primers were 5′-G GAA TTC **CAT ATG** GCT ATA AAT GCT GTG GCC G-3′ and 5′-CCG CCG **CTC GAG** TTA CTA AAC TAT GCG CCA CAT AGT GT A-3′ for Ta0058 and 5′-G GAA TTC **CAT ATG** ATC CTG AGA AGA TAC AGA AGT ACG G-3′ and 5′-CCG **CTC GAG** TTA TCT GCA CAT CAA CCT CAT TCT ATA-3′ for Ta1140. The bases in bold represent the *Nde*I and *Xho*I restriction-enzyme cleavage sites. The PCR product was then digested with *Nde*I and *Xho*I and inserted into the *Nde*I/*Xho*I-digested expression vector pET-28b(+) (Novagen). This vector construction added a 20-residue tag (MGSSHHHHHHSSGLVPR-GSH) to the N-terminus of the gene product in order to facilitate protein purification.

The Ta0058 protein was expressed in *E. coli* Rosetta II(DE3)pLysS cells. The cells were grown at 310 K to an OD_{600} of ∼0.5 in Terrific Broth medium containing 30 µg ml^{-1} kanamycin and protein expression was induced by the auto-induction method (Studier, 2005). Cell growth continued at 303 K for 18 h after auto-induction and cells were harvested by centrifugation at 4200*g* (6000 rev min^{-1}; Sorvall GSA rotor) for 10 min at 277 K. The cell pellet was resuspended in ice-cold lysis buffer [5 m*M* imidazole, 500 m*M* sodium chloride, 20 m*M* Tris–HCl pH 7.9, 5%(*v*/*v*) glycerol, 1 m*M* phenylmethylsulfonyl fluoride, 0.3%(*v*/*v*) Tween-20] and was then homogenized with an ultrasonic processor. The crude cell extract was centrifuged at 36 000*g* (18 000 rev min^{-1}; Hanil Supra 21K rotor) for 60 min at 277 K and the recombinant protein in the supernatant fraction was purified in two chromatographic steps. The first step utilized the N-terminal hexahistidine tag by metal-chelate chromatography on Ni–NTA resin (Qiagen). Gel filtration was then performed on a HiLoad XK 16 Superdex 200 prep-grade column (Amersham-Pharmacia) previously equilibrated with buffer *A* (20 m*M* Tris–HCl pH 7.9) containing 200 m*M* sodium chloride. The homogeneity of the purified protein was assessed by SDS–PAGE. The protein solution was concentrated using a YM10 ultrafiltration membrane (Millipore-Amicon). The protein concentration was estimated by measuring the absorbance at 280 nm.

The Ta1140 protein was expressed in *E. coli* BL21(DE3)pLysS cells. The cells were grown at 310 K to an OD_{600} of ∼0.6 in Terrific Broth medium containing 30 µg ml^{-1} kanamycin and protein expression was induced by 0.5 m*M* isopropyl β-D-thiogalactopyranoside (IPTG). Cell growth continued at 293 K for 26 h after IPTG induction and cells were harvested by centrifugation at 4200*g* (6000 rev min^{-1}; Sorvall GSA rotor) for 10 min at 277 K. The cell pellet was resuspended in ice-cold lysis buffer [5 m*M* imidazole, 500 m*M* sodium chloride, 20 m*M* Tris–HCl pH 7.9, 5%(*v*/*v*) glycerol, 1 m*M* phenylmethylsulfonyl fluoride] and was then homogenized with an ultrasonic processor. The crude cell extract was centrifuged at 36 000*g* (18 000 rev min^{-1}; Hanil Supra 21K rotor) for 60 min at 277 K and the recombinant protein in the supernatant fraction was purified in two chromatographic steps, essentially as above, except that the gel-filtration buffer contained 2 m*M* β-mercaptoethanol and its pH was 7.0.

2.2. Crystallization

Crystallization trials were carried out at 297 K using the hanging-drop vapour-diffusion method with 24-well VDX plates (Hampton Research). Each hanging drop was prepared by mixing equal volumes (2 µl each) of the Ta0058 protein solution (at 7.1 mg ml^{-1} concentration in a buffer consisting of 20 m*M* Tris–HCl pH 7.9, 200 m*M* sodium chloride) and the reservoir solution. Each hanging drop was

Table 1
Data-collection statistics for Ta0058.

Values in parentheses are for the highest resolution shell.

X-ray source	Photon Factory beamline BL-5A
X-ray wavelength (Å)	1.0000
Temperature (K)	100
Space group	$P4_1$ (or $P4_3$)
Unit-cell parameters (Å, °)	a = 49.334, b = 49.334, c = 70.384, α = 90, β = 90, γ = 90
Resolution range (Å)	50–2.17 (2.25–2.17)
Total/unique reflections	61931/8885
$R_{\rm merge}$† (%)	0.070 (0.182)
Data completeness (%)	99.0 (92.2)
Average $I/\sigma(I)$	28.5 (5.56)

† $R_{\rm merge} = \sum_h \sum_i |I(h)_i - \langle I(h)\rangle| / \sum_h \sum_i I(h)_i$, where $I(h)_i$ is the intensity of the ith measurement of reflection h and $\langle I(h)\rangle$ is the mean value of $I(h)$ for all i measurements.

Table 2
Data-collection statistics for Ta1140.

Values in parentheses are for the highest resolution shell.

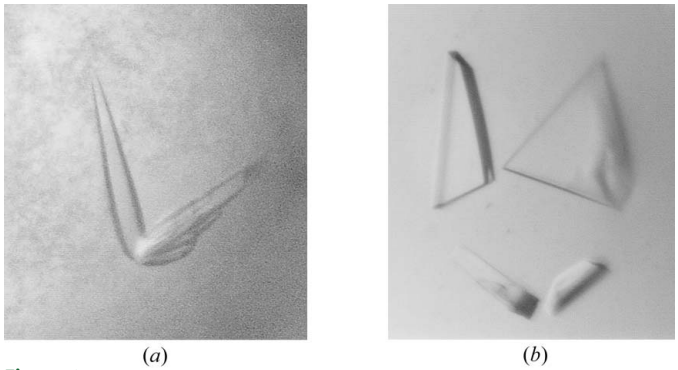
X-ray source	Pohang Light Source beamline BL-4A
X-ray wavelength (Å)	0.9919
Temperature (K)	100
Space group	$R3$
Unit-cell parameters (hexagonal setting) (Å, °)	a = 75.174, b = 75.174, c = 179.607, α = 90, β = 90, γ = 120
Resolution range (Å)	50–2.40 (2.49–2.40)
Total/unique reflections	74875/15318
$R_{\rm merge}$† (%)	0.053 (0.478)
Data completeness (%)	99.9 (99.7)
Average $I/\sigma(I)$	31.8 (3.26)

† $R_{\rm merge}$ as in Table 1.

prepared by mixing equal volumes (2 µl each) of the Ta1140 protein solution (at 6 mg ml^{-1} concentration in a buffer consisting of 20 m*M* Tris–HCl pH 7.0, 2 m*M* β-mercaptoethanol and 200 m*M* sodium chloride) and the reservoir solution. It was placed over 0.45 ml reservoir solution. Screening of crystallization conditions was carried out using commercial kits from Hampton Research (Crystal Screens I and II, MembFac, Index I and II and SaltRX I and II) and from Emerald Biostructures Inc. (Wizard I and II).

2.3. X-ray diffraction experiments

The reservoir solution used to crystallize Ta0058, which consisted of 0.1 *M* sodium acetate pH 4.6, 8%(w/*v*) polyethylene glycol 4000 and 35%(*v*/*v*) glycerol, served as a cryoprotectant solution. Therefore, the crystals could be flash-frozen by dipping into liquid nitrogen. Ta1140 crystals were cryoprotected using a solution consisting of 95 m*M* trisodium citrate pH 5.6, 19%(*v*/*v*) 2-propanol, 19%(*w*/*v*) polyethylene glycol 4000 and 5%(*v*/*v*) glycerol. The crystal was briefly (∼10 s) soaked in the cryoprotectant solution and was flash-frozen by dipping into liquid nitrogen. For Ta0058, we collected X-ray diffraction data at 100 K using a Quantum 315 CCD detector (Area Detector System Corporation, Poway, CA, USA) at beamline BL-5A of Photon Factory, Japan. The crystal was rotated through a total of 180° with a 1.0° oscillation range per frame. For Ta1140, X-ray data were collected at 100 K using a Quantum 210 CCD detector (Area Detector System Corporation, Poway, CA, USA) at beamline BL-4A of Pohang Light Source, Korea. The crystal was rotated through a total of 152°. The raw data for both Ta0058 and Ta1140 were processed and scaled using the *HKL*-2000 program package (Otwinowski & Minor, 1997).



(*a*) (*b*)

Figure 1
Crystals of N-terminal acetyltransferase-related proteins from *T. acidophilum*. (*a*) Crystals of Ta0058. The large crystals have approximate dimensions of 0.4 × 0.05 × 0.05 mm. (*b*) Crystals of Ta1140. The large crystals have approximate dimensions of 0.2 × 0.2 × 0.02 mm.

3. Results

When we expressed Ta0058 in *E. coli* strain BL21(DE3) as an intact form, it was not expressed. When we expressed it as a fusion with the N-terminal 20-residue tag (MGSSHHHHHHSSGLVPRGSH), it was highly overexpressed but the protein was insoluble. When the N-terminally tagged Ta0058 was expressed in Rosetta II(DE3)pLysS cells, it was highly overexpressed and was mostly soluble. The yield was ∼11 mg purified enzyme per litre of culture. When Ta1140 was expressed as an intact form in the *E. coli* BL21(DE3)pLysS strain, the protein was expressed at a very low level with poor solubility (below 10%). However, when we expressed it as a fusion with the N-terminal 20-residue tag in BL21(DE3)pLysS cells, the expression level was much higher and ∼80% of the expressed protein was in the soluble fraction. The yield was ∼8 mg of purified enzyme per litre of culture. Despite the presence of the N-terminal tag, we could obtain well diffracting crystals of both recombinant proteins.

The best crystallization conditions of Ta0058 were obtained after optimizing the initial crystallization condition No. 37 of Crystal Screen I (Hampton Research), with the refined reservoir solution comprising 0.1 *M* sodium acetate pH 4.6, 8%(*w*/*v*) polyethylene glycol 4000 and 35%(*v*/*v*) glycerol. The crystals grew to approximate dimensions of 0.4 × 0.05 × 0.05 mm within a day (Fig. 1). The crystal used for data collection diffracted strongly to 2.17 Å, but only very weakly beyond this resolution. An X-ray diffraction data set was collected to 2.17 Å and the crystals belonged to space group $P4_1$ (or $P4_3$), with unit-cell parameters $a = b = 49.334$, $c = 70.384$ Å, $\alpha = \beta = \gamma = 90°$. One monomer is present in the crystallographic asymmetric unit, with a calculated crystal volume per protein weight ($V_{\rm M}$) of 2.13 Å^3 Da^{-1} and a solvent content of 42.1%. Table 1 summarizes the X-ray data-collection statistics for Ta0058.

We grew the best crystals of Ta1140 after optimizing the initial crystallization condition No. 40 of Crystal Screen I (Hampton Research), with the refined reservoir solution comprising 0.1 *M* trisodium citrate pH 5.6, 20%(*v*/*v*) 2-propanol, 20%(*w*/*v*) polyethylene glycol 4000 and 0.2 *M* sodium chloride. The crystals grew to approximate dimensions of 0.2 × 0.2 × 0.02 mm within 2 d (Fig. 1). The crystals of Ta1140 diffracted more poorly than those of Ta0058. Diffraction data were collected to 2.40 Å and the space group was determined to be *R*3 on the basis of systematic absences and diffraction intensity symmetry. The unit-cell parameters are $a = b = 75.174$, $c = 179.607$ Å, $\alpha = \beta = 90$, $\gamma = 120°$ in the hexagonal setting. Two monomers are likely to be present in the asymmetric

unit, giving a V_M of 2.51 Å^3 Da^{-1} and a solvent content of 51.0%. Table 2 summarizes the X-ray data-collection statistics for Ta1140.

We thank the staff of beamlines BL-5A, BL-6A and NW-12A of Photon Factory, Tsukuba, Japan and beamlines BL-4A and BL-6B of Pohang Light Source, Korea for assistance during X-ray experiments. This work was supported by a grant from the Center for Functional Analysis of Human Genome to SWS and by the Creative Research Initiatives (NeuroVascular Coordination Research Center) of MOST/KOSEF, Korea to KWK. SHH, JYH, KHK, OSJ, DJK and JYK are recipients of BK21 fellowships.

References

Arnesen, T., Anderson, D., Baldersheim, C., Lanotte, M., Varhaug, J. E. & Lillehaug, J. R. (2005). *Biochem. J.* **386**, 433–443.
Arnesen, T., Anderson, D., Torsvik, J., Halseth, H. B., Varhaug, J. E. & Lillehaug, J. R. (2006). *Gene*, **371**, 291–295.
Arnesen, T., Betts, M. J., Pendino, F., Liberles, D. A., Anderson, D., Caro, J., Kong, X., Varhaug, J. E. & Lillehaug, J. R. (2006). *BMC Biochem.* **7**, 13.
Arnesen, T., Gromyko, D., Horvli, O., Fluge, O., Lillehaug, J. & Varhaug, J. E. (2005). *Thyroid*, **15**, 1131–1136.
Arnesen, T., Kong, X., Evjenth, R., Gromyko, D., Varhaug, J. E., Lin, Z., Sang, N., Caro, J. & Lillehaug, J. R. (2005). *FEBS Lett.* **579**, 6428–6432.
Bilton, R., Mazure, N., Trottier, E., Hattab, M., Déry, M. A., Richard, D. E., Pouysségur, J. & Brahimi-Horn, M. C. (2005). *J. Biol. Chem.* **280**, 31132–31140.
Chang, C. C., Lin, M. T., Lin, B. R., Jeng, Y. M., Chen, S. T., Chu, C. Y., Chen, R. J., Chang, K. J., Yang, P. C. & Kuo, M. L. (2006). *J. Natl Cancer Inst.* **98**, 984–995.
Fisher, T. S., Etages, S. D., Hayes, L., Crimin, K. & Li, B. (2005). *J. Biol. Chem.* **280**, 17749–17757.
Gautschi, M., Just, S., Mun, A., Ross, S., Rucknagel, P., Dubaquie, Y., Ehrenhofer-Murray, A. & Rospert, S. (2003). *Mol. Cell Biol.* **23**, 7403–7414.
Jeong, J. W., Bae, M. K., Ahn, M. Y., Kim, S. H., Sohn, T. K., Bae, M. H., Yoo, M. A., Song, E. J., Lee, K. J. & Kim, K. W. (2002). *Cell*, **111**, 709–720.
Kim, S. H., Park, J. A., Kim, J. H., Lee, J. W., Seo, J. H., Jung, B. K., Chun, K. H., Jeong, J. W., Bae, M. K. & Kim, K. W. (2006). *Biochem. Biophys. Res. Commun.* **340**, 422–427.
Lee, F. J., Lin, L. W. & Smith, J. A. (1989). *J. Bacteriol.* **171**, 5795–5802.
Mullen, J. R., Kayne, P. S., Moerschell, R. P., Tsunasawa, S., Gribskov, M., Colavito-Shepanski, M., Grunstein, M., Sherman, F. & Sternglanz, R. (1989). *EMBO J.* **8**, 2067–2075.
Murray-Rust, T. A., Oldham, N. J., Hewitson, K. S. & Schofield, C. J. (2006). *FEBS Lett.* **580**, 1911–1918.
Neuwald, A. F. & Landsman, D. (1997). *Trends Biochem. Sci.* **22**, 154–155.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Park, E. C. & Szostak, J. W. (1992). *EMBO J.* **11**, 2087–2093.
Polevoda, B. & Sherman, F. (2003). *Biochem. Biophys. Res. Commun.* **308**, 1–11.
Ruepp, A., Graml, W., Santos-Martinez, M. L., Koretke, K. K., Volker, C., Mewes, H. W., Frishman, D., Stocker, S., Lupas, A. N. & Baumeister, W. (2000). *Nature (London)*, **407**, 508–513.
Sanchez-Puig, N. & Fersht, A. R. (2006). *Protein Sci.* **15**, 1968–1976.
Studier, F. W. (2005). *Protein Expr. Purif.* **41**, 207–234.
Sugiura, N., Adams, S. M. & Corriveau, R. A. (2003). *J. Biol. Chem.* **278**, 40113–40120.