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

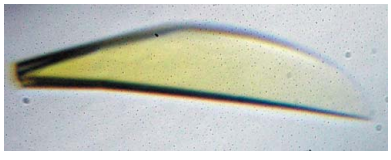
Purification and crystallization of yeast glycosylphosphatidylinositol transamidase subunit PIG-S (PIG-$S_{71–467}$)

Neelagandan Kamariah,[a] Frank Eisenhaber,[a,b,c] Sharmila Adhikari,[a] Birgit Eisenhaber[a] and Gerhard Grüber[a,c]*

[a]Bioinformatics Institute, Agency for Science, Technology and Research (A*STAR), 30 Biopolis Street, #07-01 Matrix, Singapore 138671, Singapore, [b]School of Computer Engineering, Nanyang Technological University, 50 Nanyang Drive, Singapore 637553, Singapore, and [c]School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551, Singapore

Correspondence e-mail: ggrueber@ntu.edu.sg

Received 26 May 2011
Accepted 20 June 2011



The transfer of glycosylphosphatidylinositol (GPI) anchors onto eukaryotic proteins is catalyzed by the transamidase complex, which is composed of at least five subunits (PIG-K, PIG-S, PIG-T, PIG-U and GPAA1). Here, the recombinant protein PIG-$S_{71–467}$ from *Saccharomyces cerevisiae*, including residues 71–467 of the entire 534-residue protein, was cloned, expressed and purified to homogeneity. The monodisperse protein was crystallized by the vapour-diffusion method. A diffraction data set was collected to 3.2 Å resolution with 91.6% completeness. The crystals belonged to space group *C*2, with unit-cell parameters $a = 106.72$, $b = 59.33$, $c = 124.3$ Å, $\beta = 114.19^\circ$, and contained two molecules in the asymmetric unit.

1. Introduction

Many eukaryotic proteins are attached to the extracellular side of the plasmalemma by glycosylphosphatidylinositol (GPI) lipid anchoring (Eisenhaber *et al.*, 2003; Paulick & Bertozzi, 2008). Examples of GPI-anchored proteins include cell-surface receptors such as urokinase receptor, CD14 and CD16, enzymes such as alkaline phosphatase and 5′-nucleotidase, and carbonic anhydrase, and adhesion and antigen proteins such as Thy-1, DAF and MACIF (Haziot *et al.*, 1988; Scallon *et al.*, 1989; Eisenhaber *et al.*, 2003). Defective GPI-anchor biosynthesis often has biological and medical implications, as in the case of paroxysmal nocturnal haemoglobinuria (Eisenhaber *et al.*, 2003; Orlean & Menon, 2007). Disruption of GPI lipid-anchor biosynthesis in knockout experiments leads to embryonic lethality (Nozaki *et al.*, 1999). The GPI lipid anchor is biosynthesized stepwise by a series of enzymes located in the endoplasmic reticulum (ER) membrane (Eisenhaber *et al.*, 2003). Typically, the nascent substrate proprotein is exported into the ER after ribosomal synthesis *via* the signal peptide pathway. The so-called transamidase complex in the ER lumen recognizes a specific GPI-anchor attachment signal located at the C-terminus of protein substrates (Eisenhaber *et al.*, 1998, 2004; Eisenhaber & Eisenhaber, 2010). Removal of the C-terminal propeptide and its replacement by a pre-synthesized GPI moiety are catalyzed by the GPI transamidase complex (Eisenhaber *et al.*, 2003; Orlean & Menon, 2007). The GPI lipid anchor is a complex organic structure made up of a lipid-modified phosphatidylinositol, a tetrasaccharide with variable elaborations and phosphoethanolamine subunits (Paulick & Bertozzi, 2008). Although the general outline of the GPI lipid-anchor biosynthesis pathway is common among eukaryotes, there are distinctive taxon-specific differences that might be useful for the design of species-specific inhibitors, as described for the suppression of unicellular parasites such as trypanosomes (Izquierdo *et al.*, 2009; Urbaniak *et al.*, 2008).

The human and yeast transamidase complexes consist of at least five different proteins named PIG-K, PIG-T, PIG-U, PIG-S and GPAA1 in humans and Gpi8p, GpI16p, CDC91/GAB1, Gpi17p and GAA1 in yeast (Eisenhaber *et al.*, 2003; Orlean & Menon, 2007). All five proteins are essential for GPI transamidase function. PIG-K shares sequence similarity with the cysteine protease of the C13 family and is responsible for C-terminal proteolytic processing of the substrate proteins (Maxwell *et al.*, 1995; Ramalingam *et al.*, 1996). It has been revealed that PIG-T is disulfide-linked to PIG-K (Ohishi *et*

al., 2003) and probably 'regulates the protein substrate's access to the enzyme's catalytic centre' (Eisenhaber *et al.*, 2003). The GPAA1 and PIG-U subunits have been suggested to be involved in GPI recognition (Vainauskas & Menon, 2004; Hong *et al.*, 2003). PIG-S has been speculated to play a structural role in transamidase substrate-protein selectivity and/or may fix the hydrophobic tail of the substrate protein in the transition state (Eisenhaber *et al.*, 2003).

Subunits of the GPI transamidase complex for which structures have been solved to date include low-resolution solution structures of PIG-S (yPIG-$S_{38–467}$) and PIG-K (yPIG-$K_{24–337}$) from *Saccharomyces cerevisiae* (Toh *et al.*, 2011). yPIG-$K_{24–337}$ consists of two structural domains, an egg-like portion and a small globular segment, which are linked by a 1.9 nm stalk. yPIG-$S_{38–467}$ forms an elongated molecule with a larger domain of 10.1 nm in length and 9.1 nm in diameter and a smaller domain of 6.7 nm in length and 3.4 nm in width. The two domains of yPIG-$S_{38–467}$ are tilted relative to each other. Overall, the shape of yPIG-$S_{38–467}$ appears like an open-hand domain attached to a wrist made up by the smaller domain (Toh *et al.*, 2011).

Sequence analysis of yeast PIG-S with the *ANNIE* suite (Ooi *et al.*, 2009) revealed that the N-terminus (residues 9–29) and the C-terminus (residues 473–493) contain transmembrane segments, whereby a major region between residues 30 and 472 is predicted to be exposed to the lumenal part of the endoplasmic reticulum. Based on sequence-analytic considerations including primary- and secondary-structure prediction, we generated a new construct, yPIG-$S_{71–467}$, with the first residue presumably starting from the second predicted helix after the transmembrane helix. In comparison to yPIG-$S_{38–467}$, the new recombinant protein yPIG-$S_{71–467}$ showed a higher degree of solubility and enabled us to grow three-dimensional crystals of yPIG-$S_{71–467}$ which diffracted to 3.2 Å resolution.

2. Materials and methods

2.1. Cloning and overexpression

The coding region for yPIG-S (amino acids 71–467; UniProt Q04080; EMBL U33007.1) was amplified by PCR using *S. cerevisiae* genomic DNA (strain AH104) as a template. The forward primer 5′-CATGCCATGGCATGTGTTCATGACGCTATCCAA-3′ with an *Nco*I restriction site and the reverse primer 5′-CGAGCTCGTTA-GAAGAAATTTTGTTGAACCAT-3′ with a *Sac*I restriction site were used for amplification. The amplified products were ligated into pET9-d1-His3 vector (Grüber *et al.*, 2002), which was then transformed into *Escherichia coli* cells (strain RG-2). To express the respective protein, consisting of the N-terminal amino-acid sequence HHH followed by the yPIG-$S_{71–467}$ protein sequence, liquid cultures were shaken in LB medium containing chloramphenicol (34 µg ml^{-1}), tetracycline (12.5 µg ml^{-1}) and kanamycin (30 µg ml^{-1}) for about 6 h at 310 K until an optical density OD_{600} of 0.6–0.7 was reached. To induce the production of proteins, cultures were supplemented with isopropyl β-D-1-thiogalactopyranoside to a final concentration of 1 m*M* followed by overnight incubation at 293 K.

2.2. Protein purification

E. coli cells containing recombinant yPIG-$S_{71–467}$ were harvested from 2 l cultures by centrifugation at 8000*g* for 10 min at 279 K. The cells were lysed on ice in buffer *A* (50 m*M* Tris–HCl pH 8, 500 m*M* NaCl, 2 m*M* PMSF, 1 m*M* Pefabloc, 0.8 m*M* DTT) by sonication using an ultrasonic homogenizer (Bandelin, KE76 tip) for 3 × 1 min. After sonication, the cell lysate was centrifuged at 10 000*g* for 35 min at 277 K. The resulting supernatant was passed through a filter (0.45 mm pore size) and supplemented with Ni–NTA resin pre-equilibrated in buffer *A*. The His-tagged protein was allowed to bind to the matrix for 2 h at 277 K by mixing using a sample rotator (Neolab) and was eluted with an imidazole gradient (0–100 m*M*) in buffer *A*. Fractions containing the required protein were identified by SDS–PAGE (Laemmli, 1970), pooled and concentrated using a Millipore spin concentrator with a molecular-mass cutoff of 10 kDa. The sample was applied onto a Resource Q anion-exchange column (6 ml; GE Healthcare) and eluted at room temperature. Prior to injection, the sample was diluted ten times with buffer composed of 50 m*M* Tris–HCl pH 8 to a final NaCl concentration of 50 m*M*. Elution was performed with a linear gradient of buffer *B* (50 m*M* Tris–HCl pH 8 and 50 m*M* NaCl) and buffer *C* (50 m*M* Tris–HCl pH 8 and 1 *M* NaCl). Recombinant yPIG-$S_{71–467}$ eluted at 21–24% buffer *C*. Respective fractions were pooled, concentrated in a Millipore spin concentrator and simultaneously exchanged with buffer *D* (50 m*M* Tris–HCl pH 8 and 230 m*M* NaCl). The purity of the protein sample was analyzed by SDS–PAGE (Laemmli, 1970) stained with Coomassie Brilliant Blue R250. Protein concentration was determined by the bicinchoninic acid assay (BCA; Pierce, Rockford, Illinois, USA).

2.3. Circular-dichroism (CD) spectroscopy

Steady-state CD spectra were measured in the far UV (190–260 nm) using a Chirascan spectropolarimeter (Applied Photophysics). Spectra were collected using a 60 µl quartz cell (Hellma) with a path length of 0.1 mm at 293 K and a step resolution of 1 nm. Readings were obtained for an average of 2 s at each wavelength and the recorded ellipticity values were the average of three determinations for each sample. CD spectroscopy of recombinant yPIG-$S_{71–467}$ (2.0 mg ml^{-1}) was performed in a buffer consisting of 50 m*M* Tris–HCl pH 8 and 200 m*M* NaCl. The buffer spectrum was subtracted from the protein spectrum. CD values were converted to mean residue ellipticity in units of deg cm^2 dmol^{-1} using the software *Chirascan* v.1.2 (Applied Photophysics). This baseline-corrected spectrum was used as an input for computer methods to obtain predictions of secondary structure. In order to analyze the CD spectrum, the following algorithms were used: *Varselec* (Manavalan & Johnson, 1987), *Selcon* (Sreerama & Woody, 1993), *CONTIN* (Provencher, 1982), *K2D* (Andrade *et al.*, 1993) (all as incorporated into the program *DICROPROT*; Deléage & Geourjon, 1993) and *CDNN* (Böhm *et al.*, 1992). The normalized root-mean-square deviation (NRMSD) was calculated according to Whitmore & Wallace (2004).

2.4. Crystallization conditions

The purified protein was concentrated to 10 mg ml^{-1} in 50 m*M* Tris pH 8, 230 m*M* NaCl using a 10 kDa cutoff concentrator and stored at 193 K prior to crystallization. Preliminary screening for initial crystallization conditions was performed by the hanging-drop vapour-diffusion method using screens from Hampton Research at 296 K by mixing 1 µl droplets of concentrated protein solution with an equal volume of reservoir solution and equilibrating against 500 µl reservoir solution using 24-well Cryschem plates (Hampton Research, USA). Further screenings to find optimal crystallization conditions were performed by varying the protein concentrations, using different additives (Additive Screen from Hampton Research) and seeding method. Microseeding was performed by picking needle crystals into a seed bead (Hampton Research) containing 100 µl of well solution from the same drop, which was mixed for 90 s. Three different seed stocks were prepared from the seed solution by serial dilution (10^{-1}, 10^{-2} and 10^{-3}) in well solution and 0.2 µl seed stock was added to the protein drop. Finally, a good diffraction-quality

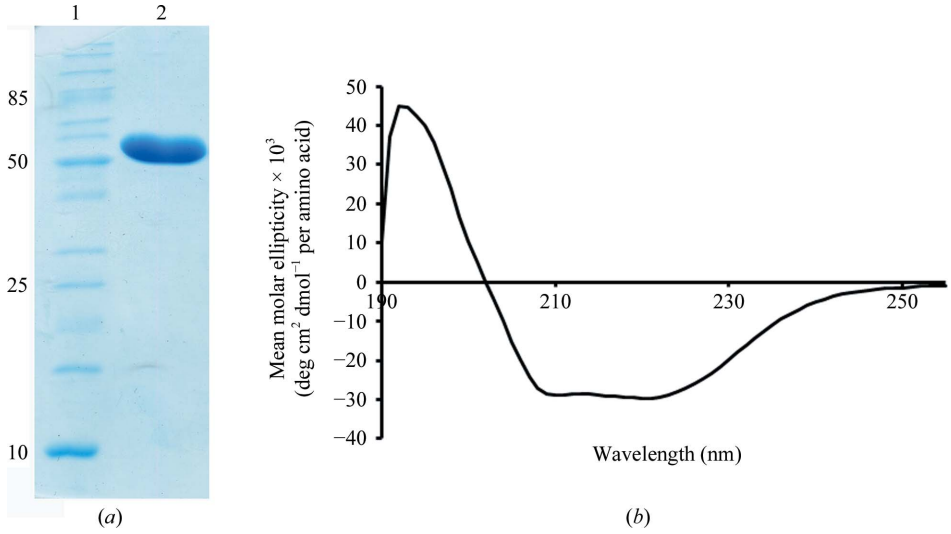


Figure 1
(*a*) SDS–PAGE (17% total acrylamide and 0.4% cross-linked acrylamide) of the purified recombinant yPIG-$S_{71–467}$; lane 1 contains protein marker (labelled in kDa). (*b*) Far-UV CD spectrum of yPIG-$S_{71–467}$ (2 mg ml^{-1}).

plate-shaped native crystal was grown at 3 mg ml^{-1} protein concentration in 0.1 *M* Tris–HCl pH 8.5, 2 *M* ammonium sulfate, 10 m*M* cadmium chloride.

2.5. X-ray data collection and analysis

The crystal was quick-soaked in a cryoprotectant solution consisting of 20% glycerol in mother liquor and flash-cooled in liquid nitrogen at 100 K. Single-wavelength data sets were collected from yPIG-$S_{71–467}$ crystals at 140 K on beamline 13B1 at the National Synchrotron Radiation Research Centre (NSRRC, Hsinchu, Taiwan) using an ADSC Quantum 315 CCD detector. The diffraction data were indexed, integrated and scaled using the *HKL*-2000 suite of programs (Otwinowski & Minor, 1997). Data-collection statistics are given in Table 1.

3. Results and discussion

3.1. Protein characterization

SDS–PAGE of recombinant yPIG-$S_{71–467}$ revealed a prominent band of about 50 kDa found entirely within the soluble fraction, which was purified by affinity and ion-exchange chromatography (Fig. 1*a*). The higher migration of the recombinant protein (sequence-based predicted mass of 45.2 kDa, including the His tag) in gel electrophoresis may be attributed to high hydration of the protein in solution, which is characteristic for extended molecules with large specific accessibility to solvent, and supports the recently determined elongated shape of yeast PIG-$S_{38–467}$ of 16.8 nm in length (Toh *et al.*, 2011). Based on secondary-structural predictions from the amino-acid sequence of yPIG-$S_{71–467}$, the secondary-structure content was 49% α-helical, 17% β-sheet and 34% random coil. The secondary structure of this protein was determined from the circular-dichroism spectrum (Fig. 1*b*), resulting in a secondary-structure composition of 52% α-helix, 18% β-sheet and 30% random coil.

3.2. Crystallization and preliminary X-ray analysis

In the initial screening, small needle-shaped crystals were obtained after several days in condition No. 4 of Crystal Screen from Hampton Research (0.1 *M* Tris pH 8.5, 2.0 *M* ammonium sulfate) at 293 K. Attempts to optimize this condition by further screening identified an additive (cadmium chloride hydrate) that produced a cluster of

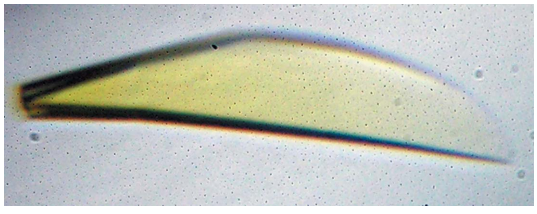


Figure 2
Crystals of yPIG-$S_{71–467}$. The crystals are approximately 0.08 × 0.22 × 0.02 mm in size.

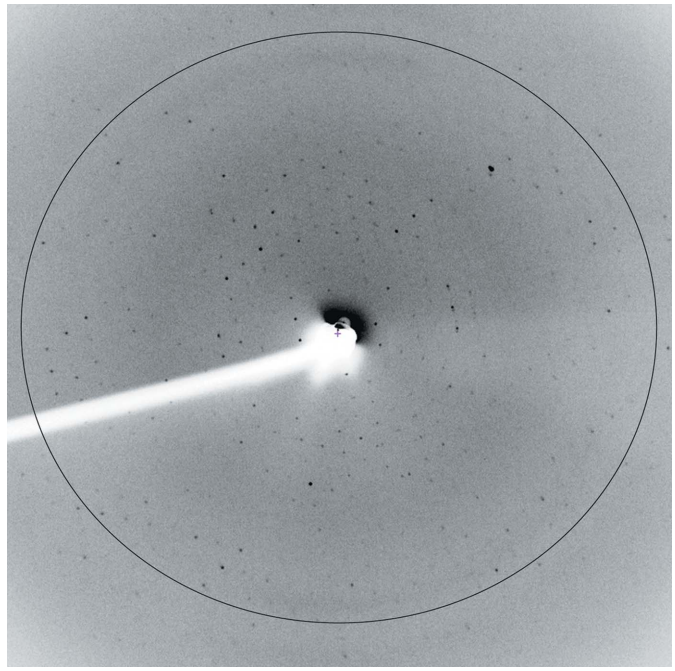


Figure 3
Diffraction pattern of yPIG-$S_{71–467}$. The ring indicates 3 Å resolution.

Table 1
Statistics of crystallographic data collection and processing for yPIG-$S_{71–467}$.

Values in parentheses are for the highest resolution shell.

Wavelength (Å)	1.5417
Space group	*C*2
Unit-cell parameters (Å, °)	a = 106.72, b = 59.33, c = 124.3, $\alpha = \gamma$ = 90.0, β = 114.19
Resolution range (Å)	50–3.2 (3.31–3.20)
No. of unique reflections	10981 (1066)
Total No. of reflections	19587
Completeness (%)	91.6 (92.1)
R_{merge}† (%)	13.2 (40.8)
Multiplicity	2.8 (2.6)
Mean $I/\sigma(I)$	6.9 (2.4)
Mosaicity (°)	0.93
Wilson B factor (Å^2)	50.1

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl) \times 100\%$, where $\langle I(hkl)\rangle$ is the average intensity of reflection *hkl* and $I_i(hkl)$ is the intensity of the *i*th measurement of reflection *hkl*.

needles. Microseeding with these needle crystals using a seed bead (Hampton Research) yielded good diffraction-quality plate-shaped single crystals. Finally, optimized native crystals were grown at 3 mg ml^{-1} protein concentration in 0.1 *M* Tris–HCl pH 8.5, 2 *M* ammonium sulfate, 10 m*M* cadmium chloride. A typical crystal is shown in Fig. 2 with dimensions of 0.08 × 0.22 × 0.02 mm. These crystals diffracted to 3.2 Å resolution (Fig. 3) and belonged to space group *C*2, with unit-cell parameters a = 106.72, b = 59.33, c = 124.3 Å, β = 114.19°. Assuming the presence of two molecules in the asymmetric unit, the solvent content was 39% and V_M was 2.1 Å^3 Da^{-1} (Matthews, 1968). Research is in progress to determine the three-dimensional structure of yPIG-$S_{71–467}$.

We thank the NSRRC (National Synchrotron Radiation Research Center, a national user facility supported by the National Science Council of Taiwan, ROC; the Synchrotron Radiation Protein Crystallography Facility is supported by the National Research Program for Genomic Medicine) staff at beamline 13B1 and Dr M. S. S. Manimekalai and Dr A. Balakrishna (SBS, NTU) for expert help with data collection.

References

Andrade, M. A., Chacón, P., Merelo, J. J. & Morán, F. (1993). *Protein Eng.* **6**, 383–390.
Böhm, G., Muhr, R. & Jaenicke, R. (1992). *Protein Eng.* **5**, 191–195.
Deléage, G. & Geourjon, C. (1993). *Comput. Appl. Biosci.* **9**, 197–199.
Eisenhaber, B., Bork, P. & Eisenhaber, F. (1998). *Protein Eng.* **11**, 1155–1161.
Eisenhaber, B. & Eisenhaber, F. (2010). *Methods Mol. Biol.* **609**, 365–384.
Eisenhaber, B., Maurer-Stroh, S., Novatchkova, M., Schneider, G. & Eisenhaber, F. (2003). *Bioessays*, **25**, 367–385.
Eisenhaber, B., Schneider, G., Wildpaner, M. & Eisenhaber, F. (2004). *J. Mol. Biol.* **337**, 243–253.
Grüber, G., Godovac-Zimmermann, J., Link, T. A., Coskun, Ü., Rizzo, V. F., Betz, C. & Bailer, S. M. (2002). *Biochem. Biophys. Res. Commun.* **298**, 383–391.
Haziot, A., Chen, S., Ferrero, E., Low, M. G., Silber, R. & Goyert, S. M. (1988). *J. Immunol.* **141**, 547–552.
Hong, Y., Ohishi, K., Kang, J. Y., Tanaka, S., Inoue, N., Nishimura, J., Maeda, Y. & Kinoshita, T. (2003). *Mol. Biol. Cell*, **14**, 1780–1789.
Izquierdo, L., Nakanishi, M., Mehlert, A., Machray, G., Barton, G. J. & Ferguson, M. A. (2009). *Mol. Microbiol.* **71**, 478–491.
Laemmli, U. K. (1970). *Nature (London)*, **227**, 680–685.
Manavalan, P. & Johnson, W. C. (1987). *Anal. Biochem.* **167**, 76–85.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Maxwell, S. E., Ramalingam, S., Gerber, L. D. & Udenfriend, S. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 1550–1554.
Nozaki, M., Ohishi, K., Yamada, N., Kinoshita, T., Nagy, A. & Takeda, J. (1999). *Lab. Invest.* **79**, 293–299.
Ohishi, K., Nagamune, K., Maeda, Y. & Kinoshita, T. (2003). *J. Biol. Chem.* **278**, 13959–13967.
Ooi, H. S., Kwo, C. Y., Wildpaner, M., Sirota, F. L., Eisenhaber, B., Maurer-Stroh, S., Wong, W. C., Schleiffer, A., Eisenhaber, F. & Schneider, G. (2009). *Nucleic Acids Res.* **37**, W435–W440.
Orlean, P. & Menon, A. K. (2007). *J. Lipid Res.* **48**, 993–1011.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Paulick, M. G. & Bertozzi, C. R. (2008). *Biochemistry*, **47**, 6991–7000.
Provencher, S. W. (1982). *Comput. Phys. Commun.* **27**, 213–227.
Ramalingam, S., Maxwell, S. E., Medof, M. E., Chen, R., Gerber, L. D. & Udenfriend, S. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 7528–7533.
Scallon, B. J., Scigliano, E., Freedman, V. H., Miedel, M. C., Pan, Y.-C., Unkeless, J. C. & Kochan, J. P. (1989). *Proc. Natl Acad. Sci. USA*, **86**, 5079–5083.
Sreerama, N. & Woody, R. W. (1993). *Anal. Biochem.* **209**, 32–44.
Toh, Y. K., Kamariah, N., Stroh, S. M., Roessle, M., Eisenhaber, F., Adhikari, S., Eisenhaber, B. & Grüber, G. (2011). *J. Struct. Biol.* **173**, 271–281
Urbaniak, M. D., Yashunsky, D. V., Crossman, A., Nikolaev, A. V. & Ferguson, M. A. (2008). *ACS Chem. Biol.* **3**, 625–634.
Vainauskas, S. & Menon, A. K. (2004). *J. Biol. Chem.* **279**, 6540–6545.
Whitmore, L. & Wallace, B. A. (2004). *Nucleic Acids Res.* **32**, W668–W673.