Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Novel β-structure of YLR301w from *Saccharomyces cerevisiae*

Kook-Han Kim, Hyung Jun Ahn, Won-Kyu Lee, Cheolju Lee, Myeong-Hee Yu and Eunice EunKyeong Kim*

Biomedical Research Institute, Korea Institute of Science and Technology, 39-1 Hawolkok-dong, Sungbuk-gu, Seoul 136-791, Republic of Korea

Correspondence e-mail: eunice@kist.re.kr

Received 26 September 2011
Accepted 4 February 2012

PDB Reference: YLR301w, 3rby.

When the Z-type variant of human α_1-antitrypsin was overexpressed in *Saccharomyces cerevisiae*, proteomics analysis identified YLR301w as one of the up-regulated proteins. YLR301w is a 27.5 kDa protein with no sequence homology to any known protein and has been reported to interact with Sec72 and Hrr25. The crystal structure of *S. cerevisiae* YLR301w has been determined at 2.3 Å resolution, revealing a novel β-structure. It consists of an N-terminal ten-stranded β-barrel with two short α-helices connected by a 23-residue linker to a seven-stranded half-barrel with two short helices at the C-terminus. The N-terminal barrel has a highly conserved hydrophobic channel that can bind hydrophobic molecules such as PEG. It forms a homodimer both in the crystal and in solution. YLR301w binds Sec72 with a K_d of 6.2 μM, but the biological significance of this binding requires further investigation.

1. Introduction

Proteins are the essential operators of cellular functions and at least 30 000 different proteins are expected to play various physiological roles in humans. However, the accumulation of misfolded or unfolded proteins in the endoplasmic reticulum (ER) causes ER stress, and chronic or unresolved ER stress can lead to apoptosis. This stress-related apoptosis contributes to pathophysiological processes involved in a number of prevalent diseases such as diabetes, atherosclerosis, neurodegenerative diseases and renal diseases (Marciniak & Ron, 2006; Herczenik & Gebbink, 2008; Tabas & Ron, 2011).

One of the better characterized protein-folding diseases is human α_1-antitrypsin deficiency, which is an autosomal co-dominant genetic disorder caused by the defective production of α_1-antitrypsin. α_1-Antitrypsin inhibits a wide variety of proteases and protects tissues from the enzymes of inflammatory cells; *e.g.* in the absence of α_1-antitrypsin neutrophil elastase is free to break down elastin, resulting in respiratory complications such as emphysema or chronic obstructive pulmonary disease in adults (Crystal, 1990; Stoller & Aboussouan, 2005; Gooptu & Lomas, 2009). α_1-Antitrypsin is also associated with cirrhosis in adults and children. The forms and degree of deficiency depend on whether the patient has one or two copies of the affected genes. To date, over 100 different variants of α_1-antitrypsin have been described, including the Z-type with a mutation of Glu342 to Lys (Lomas *et al.*, 1992). In the case of Z-type α_1-antitrypsin, protein folding is extremely slow compared with the wild type owing to loop–sheet polymerization and this ultimately results in protein aggregation (Yu *et al.*, 1995).

In order to better understand the cellular responses at the protein level when such protein aggregation occurs, we overexpressed Z-type α_1-antitrypsin in *Saccharomyces cerevisiae*, as yeast has been widely used as a model system to understand cellular processes (Gasch *et al.*, 2000). In order to observe the effect, we carried out proteomics analysis. Two-dimensional gel electrophoresis and mass analysis of differentially expressed proteins resulted in chaperones, antioxidant proteins, mitochondrial proteins for ATP generation and some others (data to be published elsewhere). Interestingly, YLR301w, which has no sequence homology to any proteins that have been characterized, is among the up-regulated proteins. The *YLR301w* (UniProt code Q05905) gene is located in chromosome 12 of *S. cerevisiae* and encodes a protein of 244 amino acids (M_r = 27 501) which does not appear to have an apparent signal sequence or transmembrane domain.

Genomic neighbourhood analysis searches suggest that YLR301w is a protein of yeast origin of unknown function. Searches in the Pfam database (Bateman *et al.*, 2004) as well as the Protein Data Bank (Berman *et al.*, 2000) showed no shared domains or motifs for YLR301w. However, YLR301w has previously been identified as an interacting partner of Sec72, which is associated with Sec62, Sec63 and Sec71, three integral membrane proteins in the ER. The Sec63 complex is required for translocation of pre-secretory proteins into the ER of *S. cerevisiae* (Willer *et al.*, 2003; Fang & Green, 1994; Rapoport, 2007; Zimmermann *et al.*, 2006; Jermy *et al.*, 2006). The Sec72 null mutant accumulated a subset of secretory precursors *in vivo*, but Sec72 is not essential for cell growth. In addition, YLR301w was reported to increase by 1.7-fold at the mRNA level when yeast was treated with rapamycin (Bandhakavi *et al.*, 2008) and YLR301w expression was down-regulated in the presence of 300 m*M* citric acid pH 3.5 (Lawrence *et al.*, 2004). In summary, YLR301w is a novel protein that is located in the cytosol with no sequence homology to any known proteins and that may play a role in various environmental stresses. Protein structure is closely linked to protein function and several recent structural genomics efforts have resulted in the identification of function based on structure (Chandonia & Brenner, 2006; Hrmova & Fincher, 2009; Kim *et al.*, 2010; Shin *et al.*, 2007; Chin *et al.*, 2006). Therefore, we investigated the structure of YLR301w in order to gain some insight into its function. Here, we report a novel crystal structure of *S. cerevisiae* YLR301w at 2.3 Å resolution.

2. Materials and methods

2.1. Protein expression and purification

The YLR301w (Swiss-Prot entry Q05905) and Sec72 (Swiss-Prot entry P39742) genes were amplified by polymerase chain reaction (PCR) from *S. cerevisiae* genomic DNA using *Pfu*Turbo polymerase (Stratagene). The PCR product of YLR301w was cloned into plasmid pET28a, which encodes a purification tag consisting of the amino acids Leu-Glu-His_6 at the C-terminus of the protein. For Sec72–YLR301w interaction, YLR301w was cloned into a modified pGEX-2T vector with an N-terminal GST tag and residues 39–193 of Sec72 were cloned into pET32a with a thioredoxin-His_6 tag added at the N-terminus. All three constructs were expressed in *Escherichia coli* BL21 (DE3) cells.

YLR301w was prepared and expressed from *E. coli* BL21 (DE3) as described previously (Ahn *et al.*, 2007). Cells overexpressing YLR301w were harvested by centrifugation and resuspended in buffer consisting of 50 m*M* Tris–HCl pH 7.4, 100 m*M* NaCl, 2 m*M* β-mercaptoethanol, 1 m*M* phenylmethylsulfonyl fluoride. The cells were disrupted by sonication; the crude lysate was centrifuged at 16 000*g* (Sorvall GSA rotor) for 30 min at 277 K and the cell debris was discarded. The supernatant was loaded onto a nickel-chelated HiTrap chelating column (GE Healthcare) and eluted with a linear gradient of 20–500 m*M* imidazole in 50 m*M* Tris–HCl pH 7.4, 100 m*M* NaCl, 2 m*M* β-mercaptoethanol. Fractions containing YLR301w were pooled based on SDS–PAGE analysis. YLR301w was further purified by gel filtration using a HiLoad 26/60 Superdex 75 prep-grade column (GE Healthcare) and was concentrated to 21 mg ml^{-1} in 50 m*M* Tris–HCl pH 7.4, 100 m*M* NaCl, 2 m*M* β-mercaptoethanol using a concentrator. Since YLR301w only has a starting methionine, an additional methionine was introduced at position Leu35 or Leu128 for phasing. The Leu128Met mutant gave a clear single band on SDS–PAGE, while the Leu35Met mutant did not. The Leu128Met mutant was then cultured in *E. coli* BL21 B834 (DE3) cells in M9 culture medium which contained extra amino acids including SeMet (Guerro *et al.*, 2001). Purification of the SeMet protein was carried out in the same manner as for the native protein.

GST-YLR301w and Sec72 were overexpressed as described above. YLR301w was purified using a GST affinity column (GE Healthcare) followed by a HiLoad 26/60 Superdex 75 prep-grade column (GE Healthcare) and was concentrated to 6 mg ml^{-1} in 50 m*M* Tris–HCl pH 7.4, 100 m*M* NaCl, 1 m*M* DTT using a concentrator. Sec72 was purified utilizing a nickel-chelating HiTrap column (GE Healthcare) and the thioredoxin tag was removed by cleavage using thrombin. The cleaved Sec72 was further purified by gel filtration using a HiLoad 26/60 Superdex 75 prep-grade column (GE Healthcare) which was pre-equilibrated in 50 m*M* Tris–HCl pH 7.4, 300 m*M* NaCl, 1 m*M* DTT and concentrated to 10 mg ml^{-1}. Purified GST-YLR301w and Sec72 were subjected to gel filtration using a Superdex 200 HR 10/30 column (GE Healthcare) in PBS buffer (50 m*M* NaH_2PO_4, 150 m*M* NaCl pH 7.4).

2.2. Crystallization and data collection

The crystallization condition for YLR301w protein was initially obtained from sitting-drop vapour-diffusion experiments using 96-well Intelli-Plates (Hampton Research) and a Hydra II Plus One robot (Matrix Technology) at 295 K and was optimized using hanging-drop vapour diffusion in 24-well plates with a 1:1 mixture of protein solution and reservoir

solution. Diffraction-quality crystals of YLR301w were obtained by mixing a protein solution consisting of 21 mg ml^{-1} protein in 50 m*M* Tris–HCl pH 7.4, 100 m*M* NaCl, 2 m*M* β-mercaptoethanol with a reservoir solution consisting of 33% polyethylene glycol monomethyl ether 550 (PEG MME 550), 0.1 *M* HEPES pH 7.3. The crystals grew as hexagonal bipyramids in 1–2 d. SeMet protein crystals were obtained using an identical condition to that used for the native crystals. The crystals were flash-cooled in liquid nitrogen and diffraction data were collected on beamline 4A of Pohang Light Source, Pohang, Republic of Korea using an ADSC Quantum 210 CCD detector at 100 K. The native crystal diffracted to 2.3 Å resolution, while the SeMet crystal diffracted to 3.0 Å resolution. Data were processed and scaled using the programs *DENZO* and *SCALEPACK* from the *HKL*-2000 program suite (Otwinowski & Minor, 1997). The Matthews coefficient for YLR301w was 3.29 Å^3 Da^{-1} and the estimated solvent content was 62.6%; there were two YLR301w molecules in an asymmetric unit. Data statistics are given in Table 1.

2.3. Structure solution and refinement

Refinement of heavy-atom parameters and phase calculation was carried out using *SOLVE* (Terwilliger, 2004), which found all four potential Se sites for the $P6_522$ space group. The identification of four Se sites by single-wavelength anomalous dispersion (SAD) phasing led to a mean FOM of 0.75 to a resolution of 3.5 Å. The electron-density map was generated through Fourier transformation. Maximum-likelihood density modification and automated model building was carried out using *RESOLVE* (Terwilliger, 2004). The resulting electron-density map with a partial model revealed clear main-chain density with substantial side-chain details. Manual building was performed using *Coot* (Emsley & Cowtan, 2004) and refinement was carried out using *CNS* (Brünger *et al.*, 1998) and *PHENIX* (Zwart *et al.*, 2008). The final model was validated using *PROCHECK* (Laskowski *et al.*, 1993). The *DALI* server (Holm & Rosenström, 2010) was used to search for proteins with a similar fold. Solvent-accessible and interaction areas were calculated by *PISA* (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html). Figures were generated using *PyMOL* (DeLano, 2002).

2.4. Size-exclusion chromatography

In order to determine the quaternary structure of YLR301w in solution, the hydrodynamic volume was determined using a Superdex 75 10/300 GL column (GE Healthcare) installed on an FPLC (GE Healthcare) protein-purification system. YLR301w and standard molecular-weight markers (Sigma) were prepared separately in buffers consisting of 50 m*M* Tris–HCl pH 7.4, 100 m*M* NaCl, 1 m*M* DTT. The proteins were passed through the Superdex 75 10/300 GL column and the elution profile was compared with the standard molecular-weight marker profile to estimate the size of the proteins in solution.

Table 1
Crystallographic data-collection and refinement statistics for YLR301w.

Values in parentheses are for the highest resolution shell.

Data set	Se peak	Native
Data-collection statistics		
Beamline	BL-4A	BL-4A
X-ray wavelength (Å)	0.97939	1.00000
Resolution range (Å)	50.0–3.0 (3.11–3.00)	50.0–2.3 (2.36–2.30)
Space group	$P6_522$	$P6_522$
Unit-cell parameters (Å)	a = 121.44, b = 121.44, c = 171.45, $\alpha = \beta$ = 90, γ = 120	a = 122.32, b = 122.32, c = 174.00, $\alpha = \beta$ = 90, γ = 120
Total/unique reflections	1723672/28778	840218/34825
Completeness (%)	99.9 (98.2)	97.9 (97.4)
Multiplicity	14.7 (8.2)	8.7 (6.3)
Mean $I/\sigma(I)$	30.3 (3.9)	27.8 (3.3)
$R_{\rm merge}$† (%)	9.8 (28.9)	8.5 (31.4)
Refinement statistics		
Resolution range (Å)		50.0–2.3
$R/R_{\rm free}$‡ (%)		20.4/24.7
No. of protein atoms		3949
No. of water molecules		146
No. of ligand molecules		7
Average B factor (Å^2)		48.3
R.m.s.d. bond length (Å)		0.009
R.m.s.d. bond angles (°)		1.199
Ramachandran analysis (%)		
Most favoured		96.1
Additionally allowed		3.9
Outliers		0

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the ith measurement of reflection hkl and $\langle I(hkl)\rangle$ is the mean value of $I(hkl)$ for all i measurements. ‡ $R_{\rm free}$ was calculated from a randomly selected 5% set of reflections not included in the calculation of the R value.

2.5. Circular-dichroism spectroscopy and dynamic light scattering

The secondary structure of YLR301w was characterized using a Jasco J-715 spectropolarimeter equipped with a temperature-control unit. Standard far-UV circular-dichroism (CD) spectra were collected at room temperature at a scan speed of 20 nm min^{-1} with a 0.1 nm step resolution using a 0.1 cm path-length cell with a 1 nm bandwidth and a 4 s response time. Ten accumulations taken from 260 to 190 nm were added and averaged, followed by subtraction of the solvent CD signal. Dynamic light-scattering (DLS) measurements were performed on a DynaPro molecular-sizing instrument equipped with a microsampler (Protein Solutions). The data were analyzed using the *Dynamics* v.4.0 and *DynaLS* software as described previously (Moradian-Oldak *et al.*, 1994). *Dynamics* v.4.0 was used to calculate the $R_{\rm h}$.

2.6. Enzyme-linked immunosorbent assay (ELISA)

An ELISA was carried out in order to estimate the interaction between YLR301w and Sec72. Each well of a nickel-coated 96-well plate (Pierce) was incubated with 1 µg His-Sec72 protein in 100 µl PBS (50 m*M* NaH_2PO_4, 150 m*M* NaCl pH 7.4) for 1 h at room temperature. The wells were blocked with 5%(*w*/*v*) skimmed milk in PBST (0.1% Tween-20 in PBS) for 1 h at room temperature and different concentrations of GST-YLR301w protein were then added to the wells and incubated at 1 h at room temperature. The wells

were incubated with HRP-labelled GST-tag antibodies (1/1000 dilution in PBST) for 1 h at room temperature; the colour-developing reaction was initiated by adding OPD solution (Pierce) and the reaction was terminated by adding 2.5 *N* H_2SO_4. The absorbance of each well was measured at 492 nm using a TRIAD microplate reader (Dynex Technologies).

2.7. PDB code

The atomic coordinates and structure factors for YLR301w from *S. cerevisiae* have been deposited in the RCSB Protein Data Bank (PDB) with code 3rby.

3. Results and discussion

3.1. Structure of YLR301w

YLR301w from *S. cerevisiae* was overexpressed in *E. coli* and purified. The CD spectra indicated significant β-structure content, *i.e.* 47% of the structure was β-structure, as shown in Fig. 1(*a*). Since YLR301w has only one methionine (that at the N-terminus), we introduced an additional methionine for phasing by mutating leucine residues randomly. Replacement of Leu128 by Met gave soluble protein and the SAD method gave a structure solution at 2.3 Å resolution. All atoms of YLR301w were well defined in the electron-density maps. The final model included two protein molecules (each consisting of residues 1–244 plus one N-terminal His residue), six PEG molecules, one sulfate molecule and 146 water molecules. The Ramachandran plot produced by *PROCHECK* (Laskowski *et al.*, 1993) showed that 96% of the residues were in the most favoured regions and 4% were in additional allowed regions. Statistics of the data collection and refinement are summarized in Table 1.

The topology of YLR301w is shown in Fig. 1(*b*) and a ribbon drawing of the molecule is shown in Fig. 1(*c*). As shown in the figure, the structure is almost exclusively comprised of β-strands and loops, with a few short helices. The structure is composed of three parts: the N-terminal 131 or so residues form a full antiparallel β-barrel consisting of ten strands (β1–β10) and two short helices, the C-terminal 80 residues (residues 155–244) form a half-barrel consisting of seven antiparallel strands (β11–β17) with two short helices and the central region (residues 132–154) forms a long linker loop between the two (Fig. 1*b*). The total β-structural content amounts to 45.7% and this is similar to the value estimated based on the CD data. There are two YLR301w molecules in the asymmetric unit and the two molecules are related by a noncrystallographic twofold axis. The two molecules are almost identical to each other, with an overall root-mean-square deviation (r.m.s.d.) of 0.51 Å over 245 C^{α} atoms.

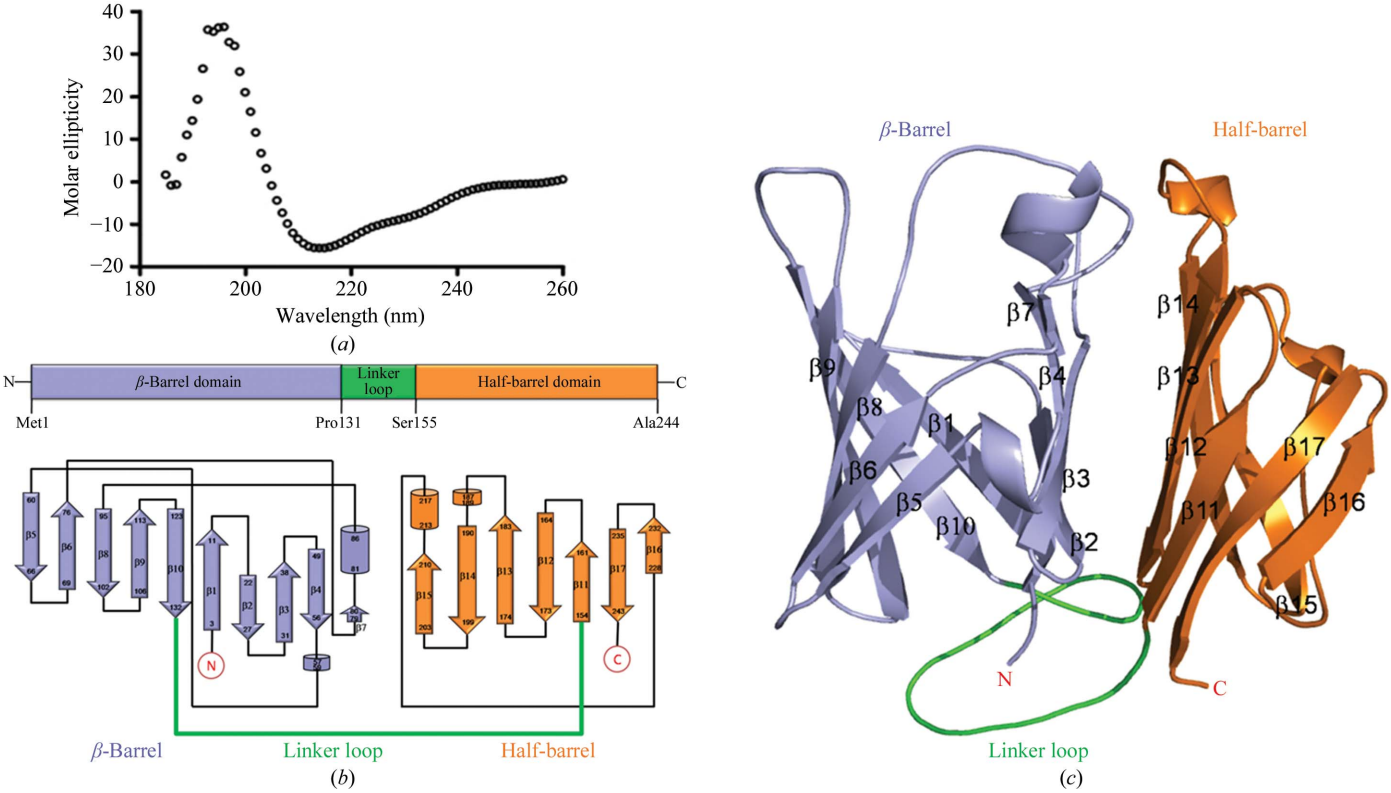

Figure 1
Structure of YLR301w from *S. cerevisiae*. (*a*) Far-UV CD spectra of YLR301w. The spectra indicate about 47% β-structure. (*b*) Topological diagram of YLR301w. YLR301w is comprised of an N-terminal domain (residues 1–131), a linker (residues 132–154) and a C-terminal domain (residues 155–244), which are coloured blue, green and orange, respectively. The N-terminal domain has a ten-stranded β-structure with two short helices and the C-terminal domain consists of a seven-stranded β-structure with two short helices. Helices and strands are shown as cylinders and arrows, respectively. (*c*) Ribbon diagram of YLR301w. The ten-stranded N-terminal domain forms a β-barrel, while the seven-stranded C-terminal domain forms a half-barrel. The domains and linker are coloured blue, orange and green as in (*b*).

Although YLR301w is annotated as an uncharacterized protein with unknown function, a *BLAST* search using the UniProt database (Wu *et al.*, 2006) gave a number of hits: putative uncharacterized proteins from *Candida glabrata* (Q6FR71), *Vanderwaltozyma polyspora* (A7TL22), *Zygosaccharomyces rouxii* (C5DWI2), *Kluyveromyces lactis* (Q6CUE9) and *Lachancea thermotolerans* (C5E2Q7) were identified with sequence identities of 54, 52, 54, 51 and 47%, respectively (see Fig. 2). Proteins with sequence identities in the range 40–60% are generally considered to be homologous, with potentially similar functions. Sequence alignment of YLR301w with these proteins shows that the conserved residues are distributed in both the N-terminal and C-terminal domains. As shown in Fig. 3, the conserved residues in the N-terminal domain mostly face the interior of the barrel. The residues in both the N-terminal and the C-terminal domain that face each other, *i.e.* the residues at the interface, are also relatively conserved. The residues lining the entrance of the N-terminal barrel, *i.e.* Phe10, Trp52, Phe54, Phe75, Phe125 and Trp129, are conserved among the YLR301w homologues. Together, these observations suggest that these proteins most likely to all have the same structural arrangement as YLR301w.

3.2. Interactions between the N-terminal and C-terminal domains

The N-terminal domain and the C-terminal domain are connected by a 23-residue linker, as shown in Fig. 1. However, the two domains make extensive interactions with each other, as indicated by the buried surface area of 1208 Å^2. The interactions involve both hydrophobic and polar residues. In addition, residues from the linker loop make interactions with the two barrels, with the interfaces being 420 and 480 Å^2, respectively. The corresponding values are 1217, 421 and 472 Å^2 for molecule *B*. The conserved residues at the

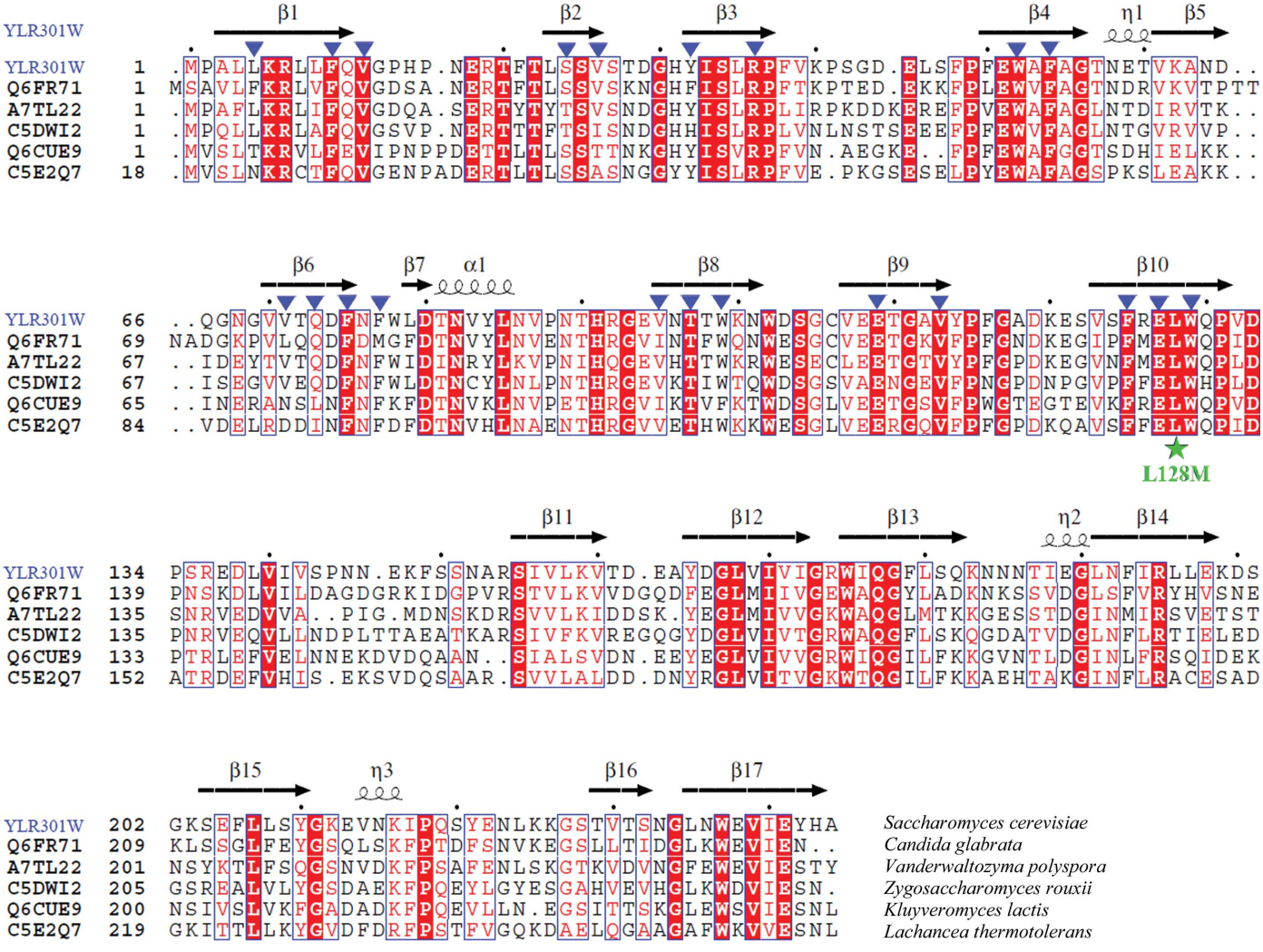

Figure 2
Amino-acid sequence alignment. A *BLAST* search shows uncharacterized proteins from *C. glabrata* (Q6FR71), *V. polyspora* (A7TL22), *Z. rouxii* (C5DWI2), *K. lactis* (Q6CUE9) and *L. thermotolerans* (C5E2Q7) as homologues of YLR301w. The residues in the red boxes are strictly conserved. Secondary structures of YLR301w are depicted above the sequence; α-helices and β-strands are indicated by coils and arrows, respectively. Residues within 3.5 Å of the two PEG MME 550 molecules bound inside the barrel are highlighted by inverted blue triangles. The sequence alignment was performed using *ClustalW* and the image was produced using *ESPript* 2.2.

N-terminal and C-terminal domain interface include both hydrophobic and hydrophilic amino acids, *i.e.* Ser25, Ser27, Leu35, Pro37, Pro49, Leu85 and Pro131 from the N-terminal domain, and Ser155, Leu168, Gly173, Trp175, Gln177, Glu241 and Arg195 from the C-terminal domain (see Fig. 2). The backbone atoms of Phe21, Leu23, Thr28, Asp29, Ser47 and Gln182 make hydrogen bonds to the side-chain atoms of Arg195, Gln177, Ser155, Glu241, Tyr210 and Tyr84, respectively, while the side-chain atoms of Ser25, Ser27, Asp29 and His31 make hydrogen bonds to the side-chain atoms of Gln177, Glu241, His243 and Glu241, respectively.

3.3. PEG binding to YLR301w

After all the protein atoms had been placed, extra electron density still remained at the centre of the barrel in both molecules *A* and *B*, as shown in Fig. 4(*a*). The electron density was clearly separated from the protein atoms; it was surrounded by the side-chain atoms of mostly hydrophobic residues, *i.e.* Leu5, Phe10, Trp52, Phe54, Val71, Phe75, Phe77, Val95, Trp99, Val107, Val113 and Phe125, and is therefore highly hydrophobic in nature (see Fig. 4*a*). Since the crystallization buffer included PEG MME 550 we modelled this molecule, and two molecules of PEG MME 550 accounted for the observed extra density as shown in Fig. 4(*a*). These molecules make hydrophobic interactions with the side-chain atoms of Phe10, Tyr32, Trp52, Phe54, Phe75, Phe77, Trp99 and Phe125. In addition, hydrogen bonds are formed to residues Ser24, Arg36, Glu109 and Glu127 through three water molecules. An additional PEG molecule is located between β7 and the β3–β4 loop in the domain interface. In this case the PEG molecule makes interactions with Phe38, Pro49, Phe50 and Leu208. The same interactions are observed in both molecules of YLR301w.

It is interesting to note that the aforementioned hydrophobic residues lining the pocket of the full barrel are

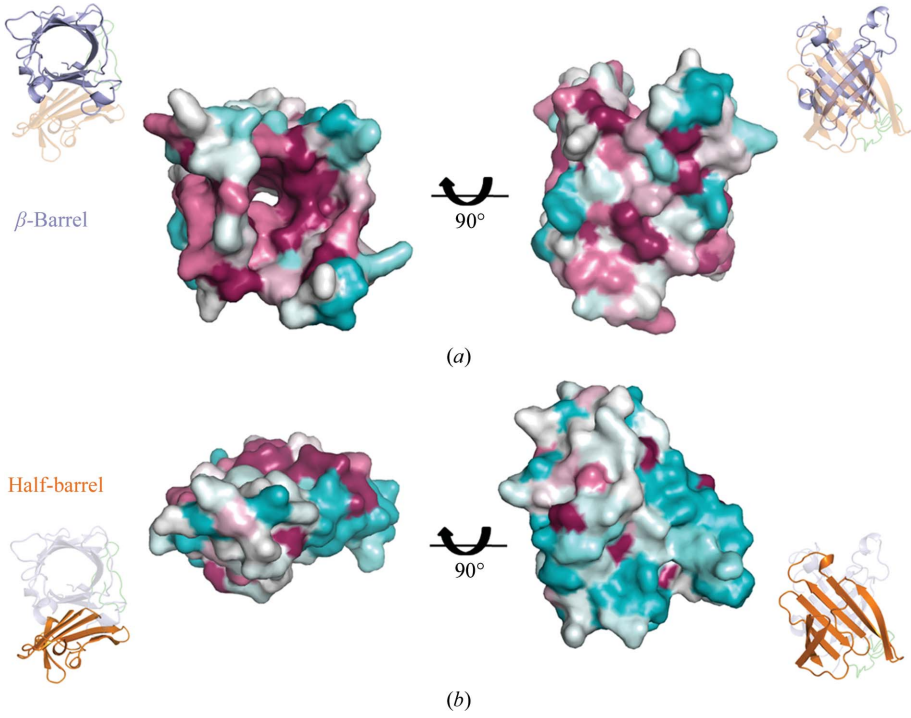

Figure 3
Conservation of YLR301w. Residues are colour-coded based on sequence conservation (see Fig. 2) going from maroon to blue to white as the degree of conservation decreases. The conservation of (*a*) the N-terminal β-barrel domain and (*b*) the C-terminal half-barrel domain are shown as surface models. The two domains are shown separately for clarity and the ribbon shows the relative orientation of the molecule.

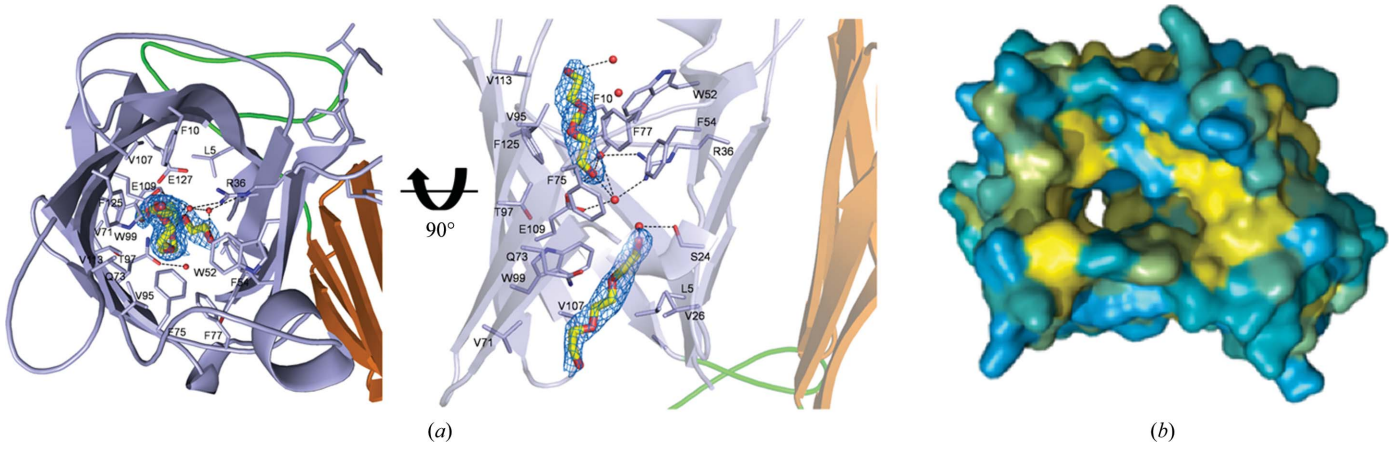

Figure 4
Internal cavity of YLR301w. (*a*) Bound PEG MME 550 at the centre of the β-barrel. There are two PEG MME 550 molecules bound inside the β-barrel. The electron-density map ($2F_o - F_c$) for the PEG molecules is contoured at the 1σ level and the residues involved in PEG binding are included. The right-hand panel shows a side view. (*b*) Hydrophobicity of YLR301w. The molecular surface is coloured according to the hydrophobicity of the side chain, with yellow representing hydrophobic residues and blue representing residues with polar side chains.

highly conserved, as indicated by blue triangles in Fig. 2. Both sides of the N-terminal barrel are open to the solvent, *i.e.* a narrow channel runs through the barrel. It is somewhat narrower in the middle of the barrel and widens at both ends. The entry to the hydrophobic cavity is defined by four loops: Val12–Thr22, Phe38–Pro49, Asn76–Glu94 and Tyr114–Ser122. This can be seen even more clearly in Fig. 4(*b*). The volume of the channel calculated using the *Pocket-Finder* algorithm (Burgoyne & Jackson, 2006) is about 910 Å^3; this is about 7% of that of the entire protein. The change in the accessible surface area owing to PEG MME 550 binding is about 435 Å^2. The same was observed for the second YLR301w molecule in the asymmetric unit; the corresponding values are 670 and 430 Å^2, respectively. Based on this, it is tempting to suggest that YLR301w may bind other hydrophobic compounds that are similar to PEG MME 550 and that YLR301w and its homologues may function as carriers of some sort.

3.4. Comparison with other structures

In order to determine whether there are any structurally homologous proteins despite the lack of sequence homology, we searched the PDB using the *DALI* server (Holm & Rosenström, 2010). Initial attempts using either the monomer or the dimer of YLR301w all failed, showing no hits. However, when the N-terminal β-barrel alone was used as a search model hits were obtained. The hits with the highest Z scores were the C-terminal haem-binding domain of human THAP domain-containing protein 4 (cTHAP4; PDB entry 3ia8; Bianchetti *et al.*, 2011), cellular retinoic acid-binding protein 2 (PDB entry 3d95; Vaezeslami *et al.*, 2008), hypothetical protein MTUBF_01000852 (PDB entry 2fwv; Shepard *et al.*, 2007) and human brain fatty-acid-binding protein (B-FABP; PDB entry 1fdq; Balendiran *et al.*, 2000). There were several additional hits as possible structural homologues, but they were either hypothetical proteins or functionally annotated as fatty acid-binding proteins. They all had a Z score of about 8, with an r.m.s.d. ranging from 3.2 to 3.4 Å. However, a search using only the C-terminal half-barrel gave no hits. Therefore, the N-terminal domain of YLR301w has a β-barrel fold that is similar to those of fatty acid-binding proteins and the half-barrel structure in YLR301w is rather unique, suggesting that the overall structure of YLR301w as a whole is quite unique and novel.

Similar to YLR301w, these fatty acid-binding proteins bind ligands or chromophores inside the barrel. However, unlike YLR301w they have a cavity rather than a channel, as shown in Fig. 5(*a*). In the case of B-FABP the internal cavity is estimated to be about 960 Å^3 in volume. The structure of the complex of B-FABP with oleic acid showed that the hydrocarbon tail of oleic acid assumes a 'U-shaped' conformation, while the hydrocarbon tail in the complex with docosahexaenoic acid adopts a helical conformation. The binding of these ligands has been reported to have a K_d in the range 28–53 n*M* (Balendiran *et al.*, 2000).

THAP4 belongs to a novel protein family termed thanatos-associated proteins, which have a conserved N-terminal C2CH zinc-finger DNA-binding motif. They are found in the proteomes of all species except plants, yeast and bacteria; 12 human THAP domain-containing proteins have been identified. Although the exact function of the THAP proteins is not well defined, they have been reported to play a role in many cellular functions and have been implicated in a number of human disease states, including heart disease and neurological defects as well as cancers (Bianchetti *et al.*, 2011). It is interesting to note that THAP has also been found to be up-regulated in response to heat shock (Gau *et al.*, 2008). The C-terminal domain of THAP4, which is evolutionarily conserved, has a haem moiety which is located in the hydrophobic cavity; therefore, the structure is similar to that of nitrobindin (r.m.s.d. of 0.68 Å, Z score of 32.8; Bianchetti *et al.*, 2010), which is involved in nitric oxide transport. In this case, the internal cavity volume excluding the haem is estimated to be ∼830 Å^3.

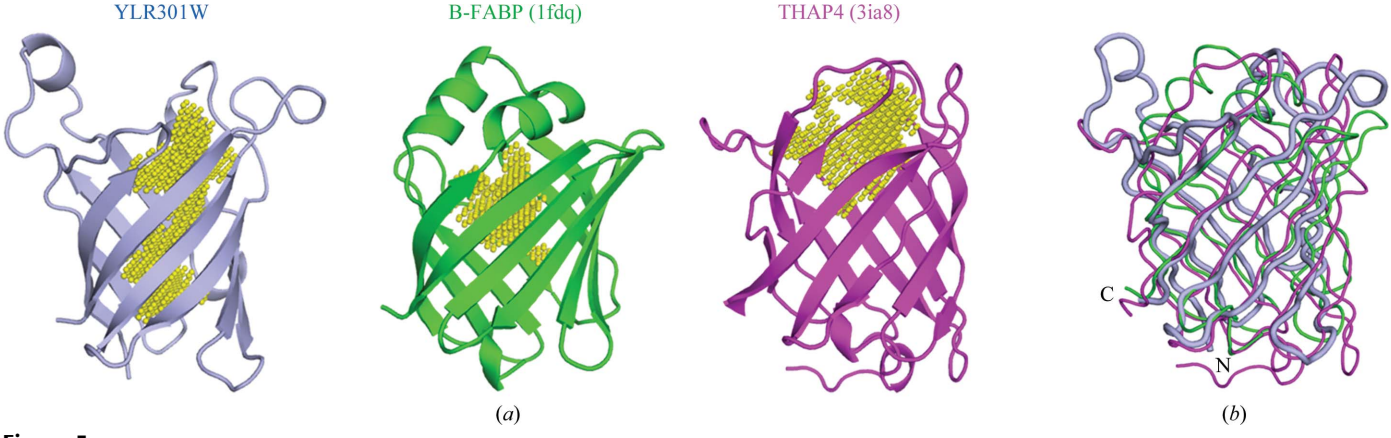

Figure 5
Structural comparison of the N-terminal β-barrel. (*a*) The N-terminal β-barrel of YLR301w from *S. cerevisiae*, fatty acid-binding protein from human brain (PDB entry 1fdq; Balendiran *et al.*, 2000) and the C-terminal domain of human THAP domain-containing protein 4 (PDB entry 3ia8; Bianchetti *et al.*, 2011) are shown in blue, green and magenta, respectively. The yellow sphere represents the channel or the cavity of the barrel calculated using the *Pocket-Finder* algorithm. (*b*) Representation of structural comparisons of YLR301w, B-FABP and cTHAP4. The three β-barrel structures are shown superimposed and coloured as in (*a*). B-FABP and cTHAP show r.m.s.d.s of 2.3 Å over 82 C$^\alpha$ atoms and 2.8 Å over 105 C$^\alpha$ atoms, respectively.

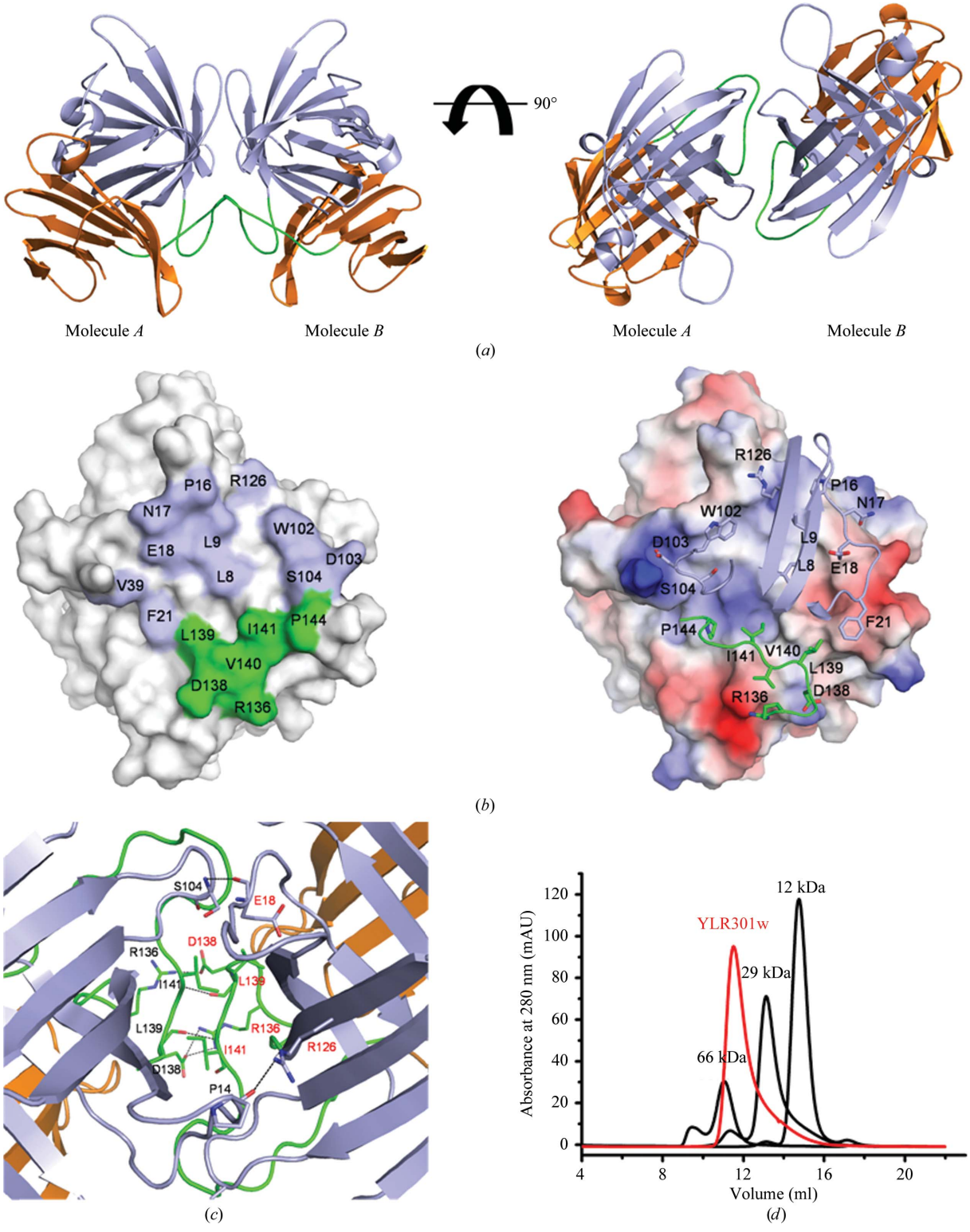

Figure 6
Molecular arrangement of the YLR301w dimer. (*a*) Dimeric arrangement. The two molecules in the asymmetric unit are related by a noncrystallographic twofold axis; the overall arrangement of the dimer is a butterfly shape with dimensions of 47 × 57 × 81 Å. (*b*) Dimer-interface residues of YLR301w. Surface representation of YLR301w; residues involved in the dimer interaction in molecule *A* are coloured and labelled. The β-barrel domain is coloured blue and the linker-loop domain is coloured green. The right panel shows the electrostatic potential of the surface area of YLR301w (blue, positively charged area; red, negatively charged area). The residues involved in the dimer interaction in molecule *B* are shown as stick models and the electrostatic surface model was calculated by the program *PyMOL*. (*c*) A close-up view showing the residues involved in hydrogen bonds in the dimer interface. These residues are shown as stick representations. (*d*) Analytical size-exclusion chromatography of YLR301w. Gel filtration was performed as described in §2. Chromatographic absorbance traces at 280 nm and elution positions are shown for molecular standards and YLR301w in black and red, respectively. The standard proteins used were bovine serum albumin (66 kDa), carbonic anhydrase (29 kDa) and cytochrome *c* (12 kDa).

Although all three proteins form a compact ten-stranded β-barrel, there are several differences. Firstly, both ends of the β-barrel are accessible in the case of YLR301w, while only one end is accessible in the other two, as shown in Fig. 5(*b*). Secondly, in the case of B-FABP the bound oleic acid is located towards the rim of the β-barrel, just beneath the α-helical lid, and the haem moiety in cTHAP4 is also bound towards the rim, while the two PEG molecules bound in YLR301w are buried deeper in the β-barrel. Thirdly, the amino-acid composition of the β-barrel is significantly more hydrophobic in the case of YLR301w. There are seven Trp, 14 Phe and seven Tyr residues in the case of YLR301w, while there are two Trp, eight Phe and two Tyr residues in B-FABP and two Trp, eight Phe and three Tyr residues in cTHAP4. Despite this, both ends of the β-barrel in YLR301w display electronegative surfaces.

3.5. Dimeric arrangement of YLR301w

As mentioned earlier, there are two YLR301w molecules in the asymmetric unit and the two molecules are related by a rotation of $\sim$100° about the noncrystallographic axis. The overall arrangement of the dimer is as a butterfly shape with dimensions of 47 × 57 × 81 Å, as shown in Fig. 6(*a*). The interface is relatively flat, but the interaction between the two molecules at the interface is quite extensive, as indicated by the change in accessible surface area, ΔASA, of about 1340 Å^2 (for molecule *A*; 1355 Å^2 for molecule *B*). These values correspond to about 11% of the surface area of a monomer. The interactions of the residues from the two N-terminal domain barrels and the linker loops are shown in Fig. 6(*b*). They are coloured blue and green, respectively. As shown in the middle panel, the interactions involve both hydrophobic and hydrophilic residues, with the hydrophobic residues at the centre and the polar residues on the peripheral surface. In addition to the two salt bridges Lys40–Asp103 and Arg136–Asp138, the backbone atoms of Pro14, Ser104, Leu139 and Ile141 make hydrogen bonds to Arg126, Glu18, Ile141 and Leu139 (see Fig. 6*c*). Further contacts are made by residues Leu8, Leu9, Pro16, Trp102, Leu139 and Pro144 from each monomer. Ser104, Glu18 and Trp102 are conserved (Fig. 2). The methionine that was introduced at position 128 for phasing is located at the dimer interface. The side chain is harboured between Leu8 and Trp102. It is somewhat accessible, in contrast to Leu35, which is located at the interface and tightly packed by hydrophobic residues.

It is worth noting that B-FABP and cTHAP4 are reported to form a dimer in their crystal structures. The orientation of the dimer is also somewhat similar to that of YLR301w; the two barrels in B-FABP and cTHAP4 are related by 150° and 170°, respectively. However, the dimeric interfaces or ΔASA in B-FABP and cTHAP4 are 470 and 1310 Å^2, respectively, corresponding to about 7 and 15% of the surface area of each monomer. Interestingly, cTHAP is found as a dimer both in the crystal and in the solution state (Bianchetti *et al.*, 2011), while B-FABP is only found as a dimer in the crystal (Balendiran *et al.*, 2000); this can be understood from the ΔASA values of the two dimers. The ΔASA of 1340 Å^2 for YLR301w is about the same as that of cTHAP. However, since this value includes the interaction provided by the residues from the linker loop (460 Å^2), it was of interest to determine whether YLR301w is a dimer in solution. In order to check this, we carried out DLS and gel-filtration analysis. As shown in Fig. 6(*d*), both DLS measurements (data not shown) and analytical gel-filtration chromatography suggested a molecular weight of 56–60 kDa, suggesting that YLR301w exists as a dimer in solution. Therefore, the dimeric YLR301w observed in the crystal structure may be biologically relevant and required for proper functioning of the YLR301w protein.

3.6. Interaction with Sec72

Sec72 has been reported to contribute to the selective recognition of signal peptides by the secretory polypeptide translocation complex (Fang & Green, 1994; Rapoport, 2007). Recently reported cryo-electron microscopic analysis of the Sec62–Sec63 complex showed that Sec72 and Sec71 form a stable complex with Sec63, although the exact mechanism requires further characterization (Harada *et al.*, 2011). Yeast two-hybrid library screening using Sec72 as bait previously gave YLR301w as an interaction partner (Willer *et al.*, 2003). To confirm this, we analyzed the interaction between the two by measuring the amount of YLR301w bound to immobilized Sec72 and measuring the binding kinetics of GST-YLR301w to His-Sec72 protein. Application of GST-YLR301w to surface-immobilized His-Sec72 yielded a saturated binding curve, as shown in Fig. 7. The dissociation constant (K_d) between the two proteins is estimated to be 6.2 μ*M*. It is worth mentioning that YLR301w only appears to be stable, and perhaps functional, in its full form, since all attempts to obtain soluble forms of the individual domains separately failed (data not shown). Nonetheless, the biological significance of the interaction between YLR301w and Sec72 requires further investigation. Interestingly, recent investigation of yeast protein

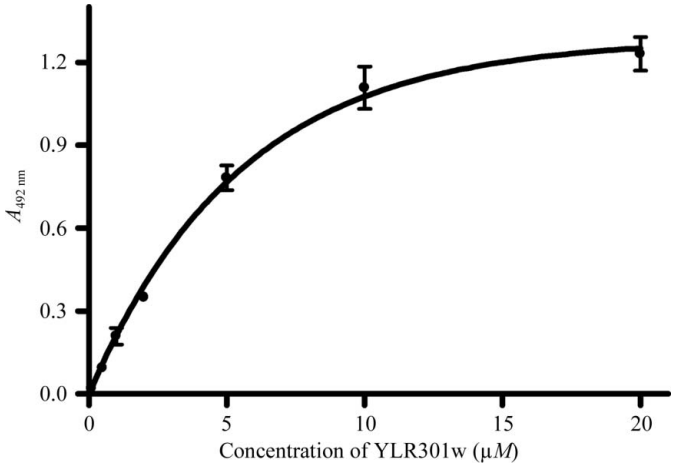

Figure 7
Interaction between GST-YLR301w and His-Sec72. The amount of GST-YLR301w bound to a His-Sec72-coated plate was measured using a modified ELISA method as described in §2.

kinases using protein microarrays revealed that YLR301w was one of the interacting proteins for Hrr25 along with Crz1, Pcl10 and Tmt1; based on this, YLR301w was proposed to be named Hri1 (Hrr25-interacting protein 1). Hrr25 is a member of the casein kinases that is required for normal cellular growth, nuclear segregation, DNA repair and meiosis (Fasolo *et al.*, 2011).

4. Conclusions

The crystal structure of *S. cerevisiae* YLR301w has been determined. The structure of YLR301w shows a unique β-structure composed of a ten-stranded β-barrel with two short α-helices at the N-terminus connected by a 23-residue linker to a seven-stranded β-sandwich with two short helices at the C-terminus. The N-terminal barrel with two PEG molecules bound has hydrophobic residues lining the channel and highly conserved residues leading to the channel. It appears that YLR301w represents a new class of proteins that have yet to be functionally characterized.

We thank the beamline staff for assistance in data collection. This work was supported financially by the Functional Proteomics Center, the 21C Frontier Research and Development Program and the Global Research Laboratory Program of the Korea Ministry of Science and Technology and an institutional grant from Korea Institute of Science and Technology.

References

Ahn, W.-S., Ahn, J.-Y., Jung, C.-H., Hwang, K. Y., Kim, E. E., Kim, J., Im, H., Kim, J.-O., Yu, M.-H. & Lee, C. (2007). *J. Microbiol. Biotechnol.* **17**, 1868–1874.

Balendiran, G. K., Schnutgen, F., Scapin, G., Borchers, T., Xhong, N., Lim, K., Godbout, R., Spener, F. & Sacchettini, J. C. (2000). *J. Biol. Chem.* **275**, 27045–27054.

Bandhakavi, S., Xie, H., O'Callaghan, B., Sakurai, H., Kim, D.-H. & Griffin, T. J. (2008). *PLoS One*, **3**, e1598.

Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., Studholme, D. J., Yeats, C. & Eddy, S. R. (2004). *Nucleic Acids Res.* **32**, D138–D141.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.

Bianchetti, C. M., Bingman, C. A. & Phillips, G. N. (2011). *Proteins*, **79**, 1337–1341.

Bianchetti, C. M., Blouin, G. C., Bitto, E., Olson, J. S. & Phillips, G. N. (2010). *Proteins*, **78**, 917–931.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Burgoyne, N. J. & Jackson, R. M. (2006). *Bioinformatics*, **22**, 1335–1342.

Chandonia, J. M. & Brenner, S. E. (2006). *Science*, **311**, 347–351.

Chin, K.-H., Chou, C.-C., Wang, A. H.-J. & Chou, S.-H. (2006). *Proteins*, **65**, 1046–1050.

Crystal, R. G. (1990). *J. Clin. Invest.* **85**, 1343–1352.

DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Fang, H. & Green, N. (1994). *Mol. Biol. Cell*, **5**, 933–942.

Fasolo, J., Sboner, A., Sun, M. G., Yu, H., Chen, R., Sharon, D., Kim, P. M., Gerstein, M. & Snyder, M. (2011). *Genes Dev.* **25**, 767–778.

Gasch, A. P., Spellman, P. T., Kao, C. M., Carmel-Harel, O., Eisen, M. B., Storz, G., Botstein, D. & Brown, P. O. (2000). *Mol. Biol. Cell*, **11**, 4241–4257.

Gau, B.-H., Chu, I.-M., Huang, M.-C., Yang, K.-T., Chiou, S.-H., Fan, Y.-H., Chen, M.-Y., Lin, J.-H., Chuang, C.-K., Huang, S.-Y. & Lee, W.-C. (2008). *Theriogenology*, **69**, 758–766.

Gooptu, B. & Lomas, D. A. (2009). *Annu. Rev. Biochem.* **78**, 147–176.

Guerrero, S. A., Hecht, H.-J., Hofmann, B., Biebl, H. & Singh, M. (2001). *Appl. Microbiol. Biotechnol.* **56**, 718–723.

Harada, Y., Li, H., Wall, J. S., Li, H. & Lennarz, W. J. (2011). *J. Biol. Chem.* **286**, 2956–2965.

Herczenik, E. & Gebbink, M. F. (2008). *FASEB J.* **22**, 2115–2133.

Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.

Hrmova, M. & Fincher, G. B. (2009). *Methods Mol. Biol.* **513**, 199–227.

Jermy, A. J., Willer, M., Davis, E., Wilkinson, B. M. & Stirling, C. J. (2006). *J. Biol. Chem.* **281**, 7899–7906.

Kim, D. J. *et al.* (2010). *Proc. Natl Acad. Sci. USA*, **107**, 21418–21423.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Lawrence, C. L., Botting, C. H., Antrobus, R. & Coote, P. J. (2004). *Mol. Cell. Biol.* **24**, 3307–3323.

Lomas, D. A., Evans, D. L., Finch, J. T. & Carrell, R. W. (1992). *Nature (London)*, **357**, 605–607.

Marciniak, S. J. & Ron, D. (2006). *Physiol. Rev.* **86**, 1133–1149.

Moradian-Oldak, J., Simmer, J. P., Lau, E. C., Sarte, P. E., Slavkin, H. C. & Fincham, A. G. (1994). *Biopolymers*, **34**, 1339–1347.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Rapoport, T. A. (2007). *Nature (London)*, **450**, 663–669.

Shepard, W., Haouz, A., Graña, M., Buschiazzo, A., Betton, J. M., Cole, S. T. & Alzari, P. M. (2007). *J. Bacteriol.* **189**, 1899–1904.

Shin, D. H., Hou, J., Chandonia, J.-M., Das, D., Choi, I.-G., Kim, R. & Kim, S.-H. (2007). *J. Struct. Funct. Genomics*, **8**, 99–105.

Stoller, J. K. & Aboussouan, L. S. (2005). *Lancet*, **365**, 2225–2236.

Tabas, I. & Ron, D. (2011). *Nature Cell Biol.* **13**, 184–190.

Terwilliger, T. (2004). *J. Synchrotron Rad.* **11**, 49–52.

Vaezeslami, S., Jia, X., Vasileiou, C., Borhan, B. & Geiger, J. H. (2008). *Acta Cryst.* D**64**, 1228–1239.

Willer, M., Jermy, A. J., Young, B. P. & Stirling, C. J. (2003). *Yeast*, **20**, 133–148.

Wu, C. H. *et al.* (2006). *Nucleic Acids Res.* **34**, D187–D191.

Yu, M.-H., Lee, K. N. & Kim, J. (1995). *Nature Struct. Biol.* **2**, 363–367.

Zimmermann, R., Müller, L. & Wullich, B. (2006). *Trends Mol. Med.* **12**, 567–573.

Zwart, P. H., Afonine, P. V., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., McKee, E., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., Storoni, L. C., Terwilliger, T. C. & Adams, P. D. (2008). *Methods Mol. Biol.* **426**, 419–435.