Structure of the human Tim44 C-terminal domain in complex with pentaethylene glycol: ligand-bound form

Noriko Handa,[a] Seiichiro Kishishita,[a] Satoshi Morita,[a] Ryogo Akasaka,[a] Zhongmin Jin,[b] John Chrzas,[b] Lirong Chen,[b] Zhi-Jie Liu,[b]‡ Bi-Cheng Wang,[b] Sumio Sugano,[c] Akiko Tanaka,[a] Takaho Terada,[a] Mikako Shirouzu[a] and Shigeyuki Yokoyama[a,d]*

[a]Protein Research Group, Genomic Sciences Center, Yokohama Institute, RIKEN, 1-7-22 Suehiro, Tsurumi, Yokohama 230-0045, Japan, [b]Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30603, USA, [c]Department of Medical Genome Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Japan, and [d]Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan

‡ Present address: National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Chaoyang District, Beijing 100101, People's Republic of China.

Correspondence e-mail: yokoyama@biochem.s.u-tokyo.ac.jp

Familial oncocytic thyroid carcinoma is associated with a missense mutation, P308Q, in the C-terminal domain of Tim44. Tim44 is the mitochondrial inner-membrane translocase subunit and it functions as a membrane anchor for the mitochondrial heat-shock protein 70 (mtHsp70). Here, the crystal structure of the human Tim44 C-terminal domain complexed with pentaethylene glycol has been determined at 1.9 Å resolution. The overall structure resembles that of the nuclear transport factor 2-like domain. In the crystal structure, pentaethylene glycol molecules are associated at two potential membrane-binding sites: the large hydrophobic cavity and the highly conserved loop between the $\alpha 1$ and $\alpha 2$ helices near Pro308. A comparison with the yeast homolog revealed that lipid binding induces conformational changes around the $\alpha 1$–$\alpha 2$ loop, leading to slippage of the $\alpha 1$ helix along the large β-sheet. These changes may play important roles in the translocation of polypeptides across the mitochondrial inner membrane.

1. Introduction

Oncocytic or Hürthle cell carcinoma of the thyroid gland accounts for 3–10% of all differentiated thyroid cancers. The cytological features are hypercellularity, with a predominance (more than 75%) of oncocytic or Hürthle cells. These cells are characterized by the abnormal proliferation of mitochondria, which appears as an eosinophilic granular cytoplasm. Oncocytic carcinoma is thought to be more aggressive and to have a higher incidence of metastasis than other thyroid cancers. In a cell model of oncocytic tumors, the activity of mitochondrial complexes I and III is decreased, leading to extensive production of reactive oxygen species (ROS; Bonora, Porcelli et al., 2006). Recently, the P308Q point mutation in the Tim44 C-terminal domain was found to be linked to an autosomal dominant familial oncocytic thyroid carcinoma (Bonora, Evangelisti et al., 2006). On the other hand, Tim44 is up-regulated in streptozotocin (STZ) induced diabetic mouse kidneys (Wada & Kanwar, 1998). The gene delivery of Tim44 into diabetic model rats or mice facilitates the mitochondrial import of antioxidative enzymes, such as superoxide dismutase and glutathione peroxidase, and reduces ROS production and renal cell proliferation (Matsuoka et al., 2005; Zhang et al., 2006).

Most of the proteins present in mitochondria are encoded by nuclear DNA and are synthesized in the cytosol as precursor proteins (Pfanner & Geissler, 2001). Preproteins

doi:10.1107/S0907444907051463 **1225**

destined for the mitochondrial matrix carry cleavable N-terminal targeting signals that direct the preproteins across both the inner and outer mitochondrial membranes. This process is mediated by two proteinaceous channels formed by the translocase of the outer membrane (TOM) complex in the outer membrane and the translocase of the inner membrane (TIM) complex in the inner membrane (Pfanner & Geissler, 2001; Neupert & Brunner, 2002; van der Laan *et al.*, 2006). Tim44 is the peripheral inner mitochondrial membrane protein and it tethers the mitochondrial heat-shock protein 70 (mtHsp70) to the TIM23 complex. Tim44 recruits soluble mtHsp70 to the TIM23 complex and mtHsp70 binds to the translocating peptide. Two different translocation mechanisms, the Brownian-ratchet model and the power-stroke model, have been proposed (Neupert & Brunner, 2002; Glick, 1995; Pfanner & Geissler, 2001; Matouschek *et al.*, 2000).

Human Tim44 shares 25.8% identity with the yeast homolog, Tim44p, with greater conservation in the C-terminal halves (Bauer *et al.*, 1999). Limited proteolysis revealed that the C-terminal domain of yeast Tim44p folds stably (Weiss *et al.*, 1999). While yeast Tim44p is tightly associated with the matrix side of the inner membrane (Blom *et al.*, 1993), human and rat Tim44 are loosely associated with the inner membrane (Bauer *et al.*, 1999; Ishihara & Mihara, 1998). The crystal structure of the yeast homolog at 3.2 Å resolution has recently been reported (Josyula *et al.*, 2006). The structure showed that the C-terminal domain forms a monomer and contains six α-helices and four antiparallel β-strands (Josyula *et al.*, 2006). The large hydrophobic cavity on the domain surface was proposed to interact with the acyl chains in the mitochondrial inner membrane (Josyula *et al.*, 2006).

In this study, we report the crystal structure of the human Tim44 C-terminal domain complexed with pentaethylene glycol molecules at 1.9 Å resolution, which probably represents the membrane-binding state. The domain forms a monomer and its topology is very similar to that of the yeast homolog. However, the binding of pentaethylene glycol to the domain surface induces conformational changes in the loop between the α1 and α2 helices, resulting in the slippage of the α1 helix along the large β-sheet. The Pro308 residue lies within the highly conserved α2–α3 loop near the α1–α2 loop. These conformational changes may trigger structural changes in the Tim44–mtHsp70 complex that play an important role in transporting an incoming polypeptide chain.

2. Materials and methods

2.1. Protein expression and purification

The C-terminal domain (residues 266–452) of human Tim44 was produced as a 227-amino-acid protein with an N-terminal histidine-affinity tag and a tobacco etch virus (TEV) protease cleavage site. The selenomethionine (SeMet) substituted protein was synthesized by the *Escherichia coli* cell-free system using the dialysis method (Kigawa *et al.*, 2004; Kigawa, Yabuki & Yokoyama, 1999; Kigawa, Yabuki, Yoshida *et al.*, 1999; Wada *et al.*, 2003). The internal solution (27 ml) was

dialyzed in three dialysis tubes (Spectra/Por 7, MWCO 15 000, Spectrum) against the external solution (270 ml) at 303 K for 4 h with shaking. After the dialysis, the internal solution was centrifuged at 16 000g at 277 K for 20 min. The supernatant was loaded onto a HiTrap Chelating (GE Healthcare Biosciences) column (5 ml) previously equilibrated with 20 m*M* Tris–HCl buffer pH 8.0 containing 1 *M* NaCl and 15 m*M* imidazole and was eluted with 20 m*M* Tris–HCl buffer pH 8.0 containing 500 m*M* NaCl and 500 m*M* imidazole. The sample buffer was exchanged to 20 m*M* Tris–HCl buffer pH 8.0 containing 1 *M* NaCl and 15 m*M* imidazole using a HiPrep 26/10 Desalting column. The histidine-affinity tag was cleaved by 100 µl TEV protease (4 mg ml^{-1}) at 303 K for 1 h. The reaction solution was loaded onto a HiTrap Chelating column (5 ml) and was eluted as described previously. The protein sample was desalted on a HiPrep 26/10 Desalting column and was eluted with 20 m*M* Tris–HCl buffer pH 8.5 containing 10 m*M* NaCl and 5 m*M* β-mercaptoethanol. Next, the protein sample was loaded onto a HiTrap Q (5 ml; GE Healthcare Biosciences) column previously equilibrated with 20 m*M* Tris–HCl buffer pH 8.5 containing 10 m*M* NaCl and 5 m*M* β-mercaptoethanol and was eluted with a linear gradient of 10 m*M*–1.0 *M* NaCl in 20 m*M* Tris–HCl buffer pH 8.5 with 5 m*M* β-mercaptoethanol. Finally, the protein sample was loaded onto a HiLoad 16/60 Superdex 75 (GE Healthcare Biosciences) column previously equilibrated with 20 m*M* Tris–HCl buffer pH 8.0 containing 150 m*M* NaCl and 2 m*M* DTT and was eluted using this buffer. The native protein prepared for the analytical ultracentrifugation experiments was synthesized and purified in the same manner as the SeMet-substituted protein.

2.2. Protein crystallization and data collection

Crystals of the SeMet-substituted protein were grown at 293 K by the sitting-drop vapor-diffusion method (protein at 19.2 mg ml^{-1}) against a reservoir solution containing 85 m*M* Na HEPES pH 7.5, 1.7 *M* ammonium sulfate, 1.7% PEG 400 and 15% glycerol. The crystals belong to space group $P6_522$, with unit-cell parameters $a = b = 108.12$, $c = 114.06$ Å. There is one monomer in the asymmetric unit. X-ray diffraction data for the SAD method were collected at the Southeast Regional Collaborative Access Team (SER-CAT) 22-ID beamline at the Advanced Photon Source (APS), Argonne National Laboratory. Supporting institutions may be found at http://www.ser.anl.gov/new/members.html. All data were processed using the *HKL*-2000 and *SCALEPACK* programs (Otwinowski & Minor, 1997). The data-processing statistics are summarized in Table 1.

2.3. Structure determination and refinement

The positions of the Se atoms and the initial SAD phases were determined using the program *SOLVE* (Terwilliger & Berendzen, 1999) and the SAD phases were improved with *RESOLVE* (Terwilliger, 2004), which automatically located about 66% and 7% of the residues with and without side chains, respectively. The resulting electron-density map was

clear. Two pentaethylene glycol molecules were traced unambiguously. The model was built with the program *TURBO-FRODO* and multiple cycles of model building and refinement were performed. The model was refined using *CNS* v.1.1 (Brünger *et al.*, 1998). The final model has good geometry as examined by *PROCHECK* (Laskowski *et al.*, 1996): 93.3% of the residues have φ/ψ angles in the most favored region of the Ramachandran plot and 100% are in allowed regions. The refinement statistics are summarized in Table 1.

2.4. Analytical ultracentrifugation

All analytical ultracentrifugation experiments were carried out with a Beckman Optima XL-I analytical ultracentrifuge. The sample buffer was 20 m*M* Tris–HCl pH 8.0, 150 m*M* NaCl and 10 m*M* β-mercaptoethanol and all experiments were performed at 277 K. The solvent density ($\rho = 1.0045$ g ml^{-1}) and the protein partial specific volume ($\overline{v} = 0.740$) were estimated with *SEDNTERP* (Laue *et al.*, 1992). Sedimentation-velocity data were obtained at 40 000 rev min^{-1} using an Epon two-channel centerpiece, with loading concentrations of 0.45 mg ml^{-1} native protein. The data were analyzed with the program *SEDFIT* (Schuck, 1998). Sedimentation-equilibrium experiments were carried out with six-channel centerpieces, with loading concentrations of 0.70, 0.35 and 0.18 mg ml^{-1}. Data were obtained at 11, 13 and 15 krev min^{-1}. A total equilibration time of 16 h was used for each speed, with scans taken at 12 and 14 h to ensure that equilibrium had been reached. The absorbance wavelength was 280 nm and the optical baseline was determined by overspeeding at 40 krev min^{-1} at the end of data collection. The equilibrium data were fitted using the manufacturer's software.

3. Results and discussion

3.1. Overall structure of the human Tim44 C-terminal domain

The crystal structure of the human Tim44 C-terminal domain (residues 266–452) has been determined at 1.9 Å resolution by the single-wavelength anomalous dispersion (SAD) method (Table 1). In the electron-density map, the N-terminal artificial linker, and four and one residue(s) at the N- and C-termini, respectively, were disordered. The final model consists of 182 residues, two pentaethylene glycol molecules (PEG1 and PEG2) and 151 water molecules. PEG2, one of the two pentaethylene glycol molecules, also interacts with a symmetry-related protein molecule (Fig. 1*a*). Thus, there are two pentaethylene glycol-binding surfaces in the Tim44 C-terminal domain (details will be discussed later). The Tim44 C-terminal domain folds as a monomer in the crystal. The molecular weight of the native protein was measured by analytical ultracentrifugation. Sedimentation equilibrium using UV absorption yielded a molecular-weight value of 22 880 Da, which is close to the calculated value of the monomer (21 633 Da). The sedimentation-velocity data showed that the Tim44 C-terminal domain sedimented as a single species, with a sedimentation coefficient of 1.3 S. This

Table 1
Data-collection, phasing and refinement statistics.

|	
Data collection and processing	
Wavelength (Å)	0.9792
Resolution range (Å)	50–1.9
Unique reflections	31350
Measured reflections	256545
Redundancy	8.18
Completeness (%)	99.1 (95.7)
R_{sym}† (%)	6.5 (31.3)
$I/\sigma(I)$	25.6 (3.9)
Phasing statistics	
Resolution range (Å)	15–2.1
Se sites per monomer	6
FOM$_{MAD}$‡	0.40
Model refinement	
Resolution range (Å)	20–1.9
No. of reflections	31278
No. of protein atoms	1441
No. of water molecules	151
No. of pentaethylene glycol atoms	32
R_{work}/R_{free}§ (%)	21.8/22.1
B factor estimated from Wilson plot (Å^2)	20.5
Mean B factor for all atoms (Å^2)	34.7
Mean B factor for protein atoms (Å^2)	33.4
Mean B factor for waters (Å^2)	43.2
Mean B factor for pentaethylene glycol atoms (Å^2)	49.2
Mean B factor for PEG1 atoms (Å^2)	55.0
Mean B factor for PEG2 atoms (Å^2)	43.5
Stereochemistry	
R.m.s.d. for bond lengths (Å)	0.005
R.m.s.d. for bond angles (°)	1.10
Residues in the Ramachandran plot	
Most favored region (%)	93.3
Additional allowed regions (%)	6.7
Generously allowed regions (%)	0
Disallowed regions (%)	0

† $R_{sym} = (\sum_h \sum_i |I_{hi} - \langle I_h \rangle|/\sum_h \sum_i |I_{hi}|)$, where h indicates unique reflection indices and i indicates symmetry-equivalent indices. ‡ Figure of merit after *SOLVE* phasing. § $R_{work} = \sum |F_{obs} - F_{calc}|/\sum F_{obs}$ for all reflections; R_{free} was calculated using randomly selected reflections (5%).

low value indicates that the domain has an extended structure. The gel-filtration assay of full-length yeast Tim44p revealed that Tim44p forms dimers, suggesting that Tim44p recruits two molecules of mtHsp70 to an import site (Moro *et al.*, 1999). Our result shows that the Tim44 homodimeric interaction does not occur *via* the C-terminal domain. It was proposed that the sequential hand-over-hand interaction mode of mtHsp70 with a peptide promotes protein import and that the highly conserved coiled-coil domains in the N-terminal portion of Tim44 are involved in the homodimeric conformation (Moro *et al.*, 1999).

The C-terminal domain consists of four long β-strands (β1b, β2a, β3b and β4), three short β-strands (β1a, β2b and β3a), four long α-helices (α1, α2, α3 and α5) and one short α-helix (α4) (Fig. 1*a*). The overall structure has a flat pyramid-like appearance. The base is a triangle formed by the α1 and α2 helices and the N-terminal end of the α3 helix. The angle between the α1 and α2 helices is almost a right angle and the α3 helix is almost perpendicular to the plane formed by the α1 and α2 helices. The top is formed by the loop between β2b and β3a. The α3, α4 and α5 helices and the four long β-strands (β1b, β2a, β3b and β4) form an $\alpha + \beta$ barrel, which is similar to that of the nuclear transport factor 2-like (NTF2-like) fold (Bayliss *et al.*, 2002; Lundqvist *et al.*, 1994). In the middle of the

α3 helix, there is a kink induced by Pro325. The four long β-strands form a large antiparallel β-sheet, which curves extensively around one side of the α3 helix before the kink (α3a). The β-sheet is extended by the additional small anti-parallel β-sheet formed by the three short β-strands. The loop between β3b and β4 contains a 3_{10}-helix.
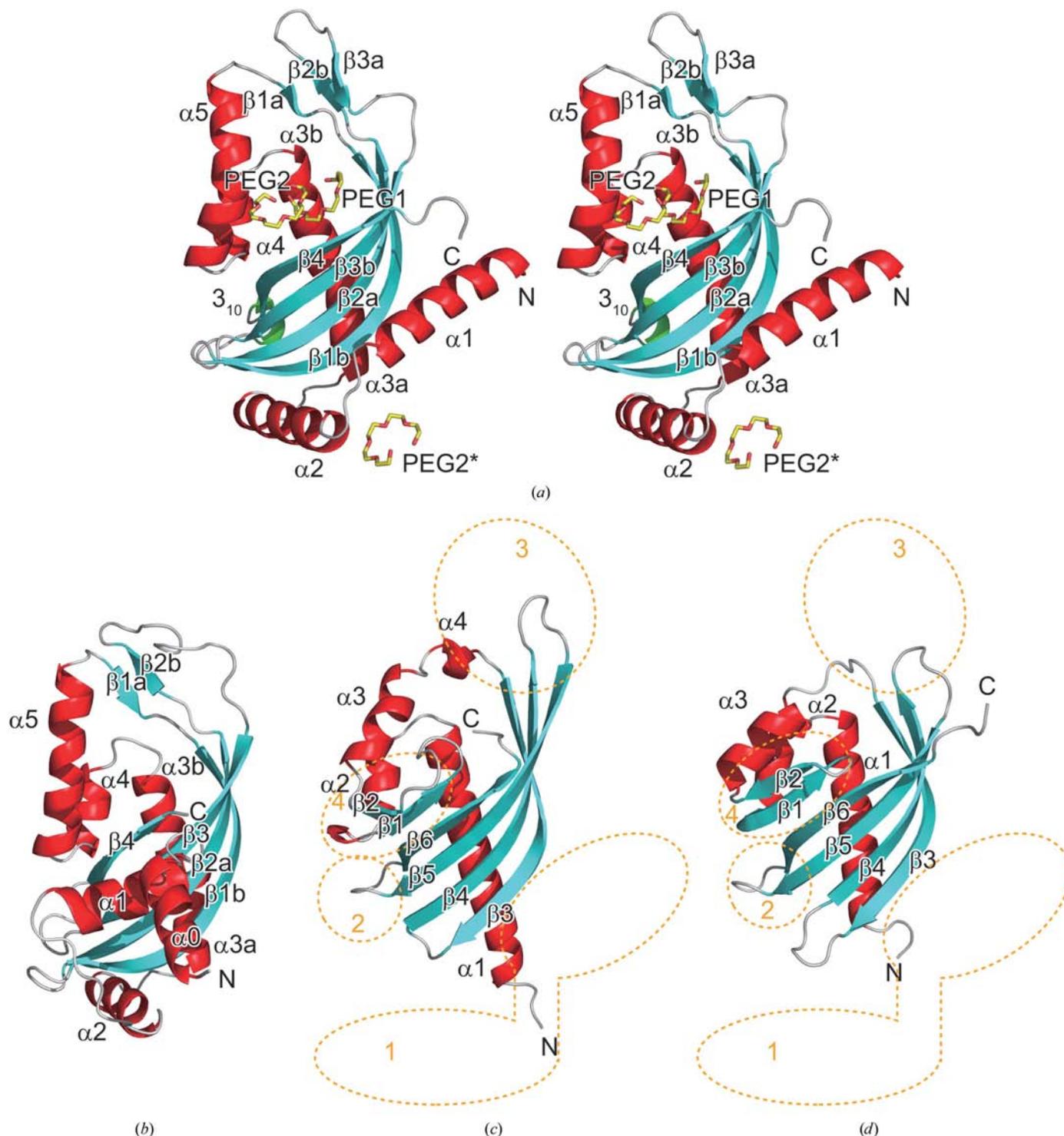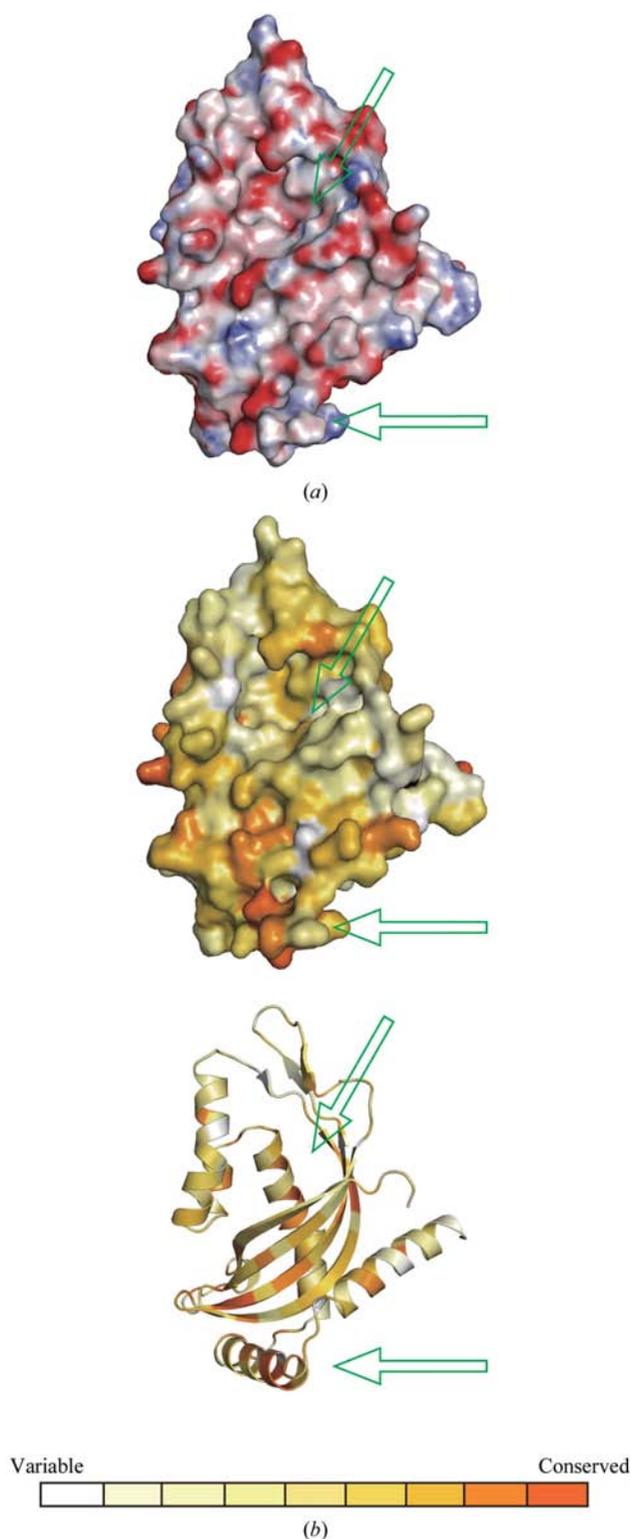

Figure 1
Structures of Tim44 and its structure homologs. (a) Ribbon diagram of the human Tim44 C-terminal domain complexed with pentaethylene glycol molecules (stereoview). The β-strands are cyan, the α-helices are red, the 3_{10}-helix is green and the random coils are gray. The bound pentaethylene glycol molecules are represented by yellow stick models with the O atoms shown in red. PEG2 and PEG2* are symmetry-related. This figure was drawn using *PyMOL* (http://www.pymol.org). (b), (c), (d) Ribbon diagrams of the yeast Tim44p C-terminal domain (b), the association domain of Ca^{2+}/ calmodulin-dependent kinase II (CaMKII) (c) and the nuclear transport factor 2 (NTF2) (d). The four major differences between Tim44 and CaMKII or between Tim44 and NTF2 are enclosed by orange dotted lines.
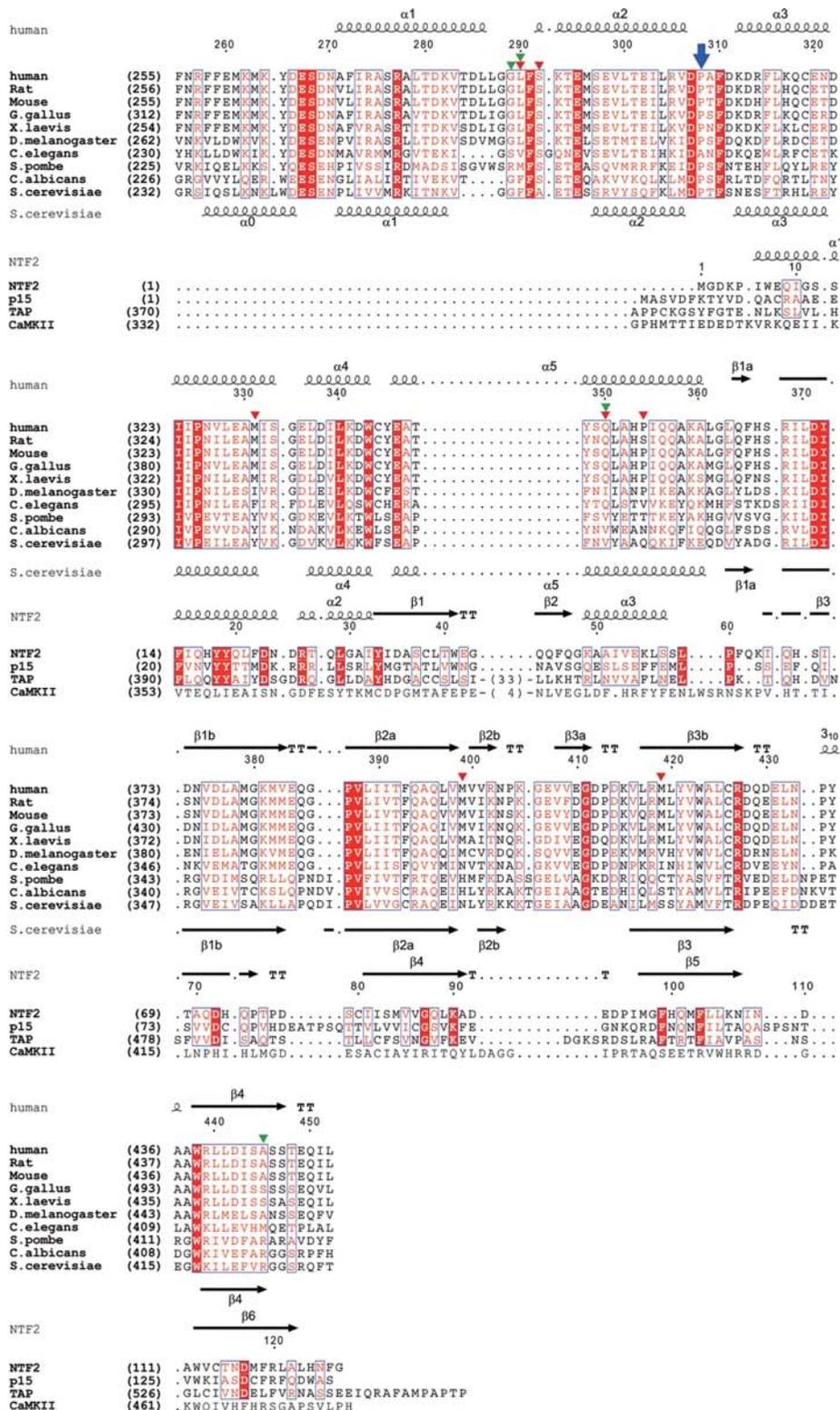
Figure 2
Sequence alignments of Tim44 homologs. The upper lines are the sequence alignments of human Tim44 with the sequence homologs from *Rattus norvegicus*, *Mus musculus*, *Gallus gallus*, *Xenopus laevis*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Schizosaccharomyces pombe*, *Candida albicans* and *Saccharomyces cerevisiae*. The lower lines are the structure-based sequence alignments of human Tim44 with the structural homologs NTF2, p15, TAP and CaMKII, which share high levels of structural similarity [PDB codes 1oun, 1jn5 (chain *A*), 1jn5 (chain *B*) and 1hkx, respectively]. The secondary structures of the human and yeast Tim44 C-terminal domains and NTF2 are shown. Strictly conserved and similar residues are represented by white characters on a red background and by red characters, respectively. Residues involved in contacting pentaethylene glycol molecules in the human Tim44 crystal structure with their main chains and side chains are indicated by green and red triangles, respectively. Pro308 in human Tim44 is shown by a blue arrow. The alignment was produced using *ClustalX* (Thompson *et al.*, 1997) and was manually modified. The figure was generated using *ESPript* (Gouet *et al.*, 1999).

3.2. Comparison to known homologous structures

Human Tim44 and yeast Tim44p share 27% sequence identity in the full-length protein and 30% sequence identity in the C-terminal domain region (residues 266–452 in Tim44 and 244–431 in Tim44p; Fig. 2). Recently, the crystal structure of the yeast Tim44p C-terminal domain was reported at 3.2 Å resolution (Josyula *et al.*, 2006). The protein used for crystallization of the yeast structure is 34 amino acids (residues 210–243) longer at the N-terminus. This portion contains an additional α-helix, which protrudes from the domain and may be flexible (α0 in Figs. 1*b* and 2; Josyula *et al.*, 2006). The overall structure and topology of the human Tim44 C-terminal domain are similar to those of yeast Tim44p (Figs. 1*a*, 1*b* and 2). A pairwise comparison with the program *DALI* generated a *Z* score of 14.7 and the root-mean-square (r.m.s.) deviation is 3.5 Å for 147 aligned residues. Despite the similarities, there are significant differences between the two structures (details will be discussed later).

A search of the Protein Data Bank using the program *DALI* revealed that the Tim44 C-terminal domain exhibits the highest structure similarity with the association domain of Ca^{2+}/calmodulin-dependent kinase II (CaMKII; Hoelz *et al.*, 2003), the hypothetical protein PA1314 from *Pseudomonas aeruginosa*, the transport protein p15 (Fribourg *et al.*, 2001), the NTF2-like domain of yeast Bre5 (Li *et al.*, 2005), the periplasmic domain from *Agrobacterium tumefaciens* VirB8 (Bailey *et al.*, 2006), NTF2 (Bullock *et al.*, 1996) and scytalone dehydratase, with *Z* scores ranging between 10.0 and 9.2, and r.m.s. deviations ranging between 2.6 and 3.5 Å. However, their amino-acid sequences share only weak similarities with that of Tim44 (identities ranging between 4 and 15%; Fig. 2). For comparison, the structures of Tim44, CaMKII and NTF2 are placed side by side in Figs. 1(*a*), 1(*c*) and 1(*d*), respectively. The Tim44 C-terminal domain fold differs from the other NTF2-like fold proteins at the following points: (i) it has two additional long α-helices (α1 and α2; Figs. 1*c* and 1*d*, orange circle 1) at the N-terminal region, (ii) its α3 helix has a kink in the middle, (ii) it has a short 3_{10}-helix between β3b and β4 (Figs. 1*c* and 1*d*, orange circle 2), (iv) it has an additional short β-strand (β1a) between α5 and β1a and two additional short β-strands (β3b and β3a) between β2a and β3b, forming a small antiparallel β-sheet (Figs. 1*c* and 1*d*, orange circle 3), and (v) it lacks two short β-strands between α2 and α3 (Figs. 1*c* and 1*d*, orange circle 4). Features (iii) and (v) are also seen in the structures of the pVirB8 proteins from *Brucella suis* and *A. tumefaciens*, although the additional helix is an α-helix rather than a 3_{10}-helix (Bailey *et al.*, 2006; Terradot *et al.*, 2005).

3.3. Large hydrophobic cavity bound to polyethylene glycol

As also found in many NTF2-like proteins, Tim44 has a hydrophobic cavity formed between three helices, α3, α4 and α5, on one side and two β-sheets on the other (Figs. 1*a*, 3*a* and 3*b*). In the crystal, two molecules of pentaethylene glycol, PEG1 and PEG2, were unambiguously traced in this cavity (Figs. 4*a* and 4*b*). In NTF2, the cavity formed at the corresponding region is also hydrophobic and is utilized to bind

Figure 3
Surface representation and ribbon diagram of the human Tim44 C-terminal domain. (*a*) Blue and red surfaces represent positive and negative potentials, respectively. The figure is in the same orientation as that in Fig. 1(*a*). Two pentaethylene glycol binding sites are indicated by arrows. This figure was drawn using *APBS* (http://apbs.sourceforge.net/) and *PyMOL* (http://www.pymol.org). (*b*) The residues are colored according to sequence conservation, ranging from white (variable residues) to orange (conserved residues). This figure was drawn using *ClustalX* (Thompson *et al.*, 1997), *CONSURF* (Glaser *et al.*, 2003) and *PyMOL* (http://www.pymol.org).

Phe72 in the switch II loop of Ran (Stewart *et al.*, 1998; Figs. 1*d* and 4*d*). The hydrophobic cavity of Tim44 is much larger than that of NTF2, mainly because Tim44 lacks the two short β-strands between α2 and α3 found in many NTF2-like domains (Figs. 1*c* and 1*d*, orange circle 4). In Tim44, the hydrophobic residues Leu328, Met331, Ile332, Leu340, Leu351, Pro354, Ile355, Phe365, Leu397, Met399, Leu417, Met419, Tyr421, Ile443 and Ala445 constitute the hydrophobic cavity, whereas hydrophilic residues surround the cavity entrance (Fig. 4*a*). Despite the similar location and hydrophobicity of the cavity, Tim44 does not share significant sequence similarity with NTF2 or other NTF2-like fold domains, including CaMKII (Fig. 2). In the Tim44 crystal structure, the bound pentaethylene glycol molecules contact the side chains of Met331, Gln350, Pro354, Met399 and Met419, and the main chains of Gln350 and Ala445 (Fig. 4*b*).

A study of yeast Tim44p using model membranes suggested that Tim44p associates with the phospholipids of the mitochondrial inner membrane *via* its C-terminal domain (Weiss *et al.*, 1999). Therefore, the hydrophobic cavity of Tim44 may be a lipid-binding site in the mitochondrial inner membrane and the residues that associate with the pentaethylene glycol molecules in the crystal may play an important role in the direct protein–membrane association.

The color-coded sequence-conservation score of the Tim44 C-terminal domain surface indicates that the hydrophobic cavity has relatively weak conservation, although its hydrophobicity is highly conserved (Figs. 2 and 3*b*). This finding suggests that the Tim44 cavity differs in size and shape between species and thus that the lipid-binding affinity may also differ between species. Actually, the size and the shape of the cavity differ between the human and yeast structures (Figs.
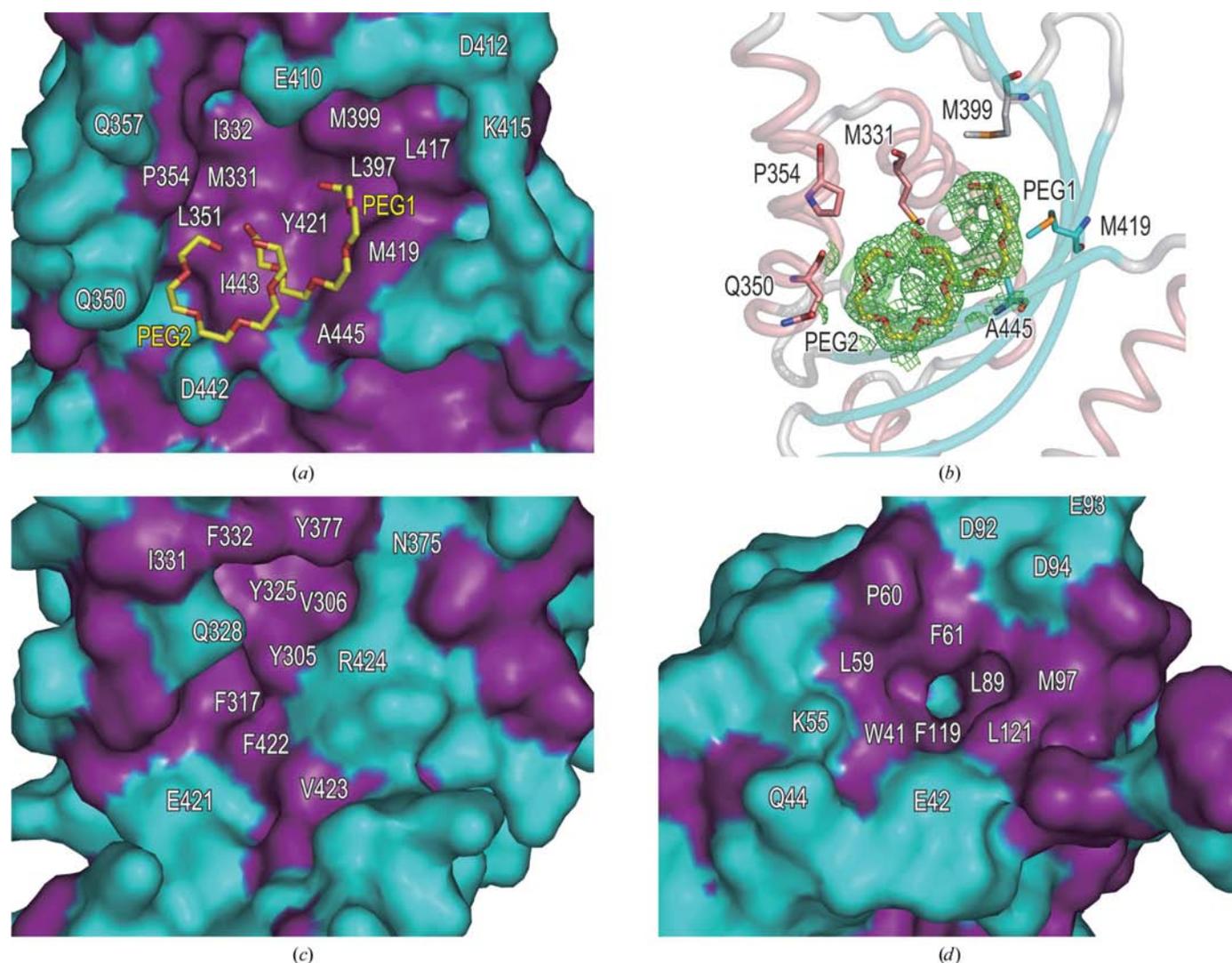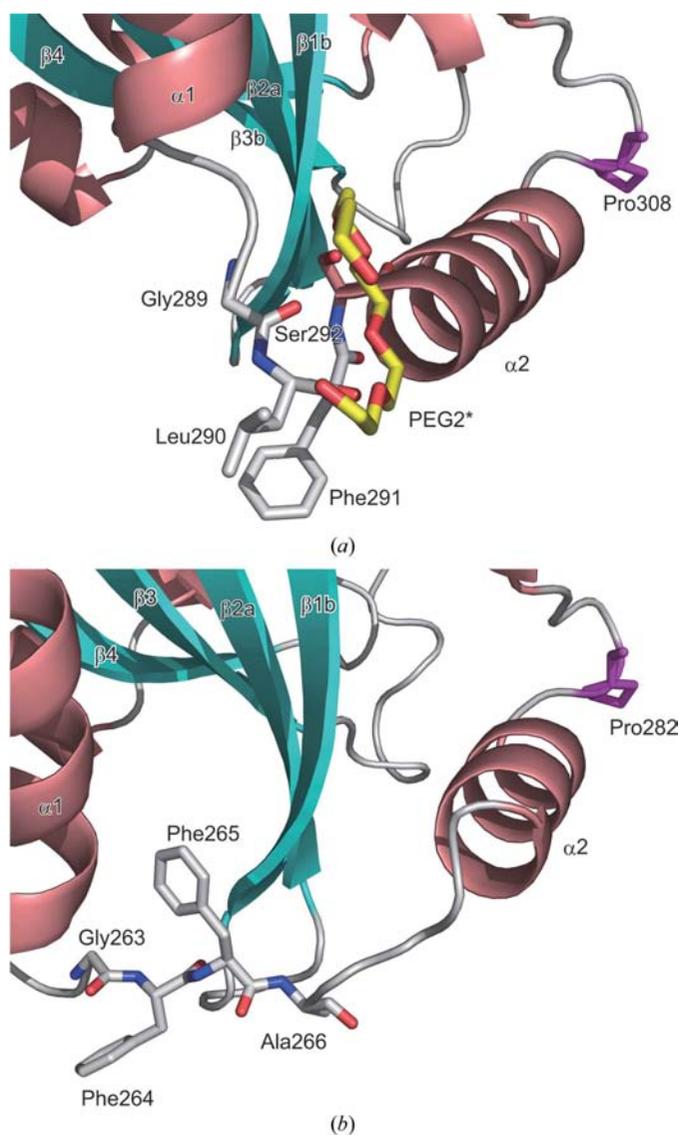

Figure 4
Hydrophobic pocket. (*a*) Surface representation of the hydrophobic pocket of human Tim44. The figure represents a 30° rotation of Fig. 1(*a*) around the *x* axis. The bound pentaethylene glycol molecules are represented as stick models. The hydrophobic and hydrophilic residues are purple and cyan, respectively. (*b*) Cartoon representation of the interaction with pentaethylene glycol molecules. A simulated-annealing OMIT $\alpha_{calc}(|F_o| - |F_c|)$ map was calculated without the pentaethylene glycol molecule atoms to 1.9 Å resolution and was contoured at 3.0σ. The orientation is the same as in (*a*). The interacting residues are shown as stick models, with O, N and S atoms shown in red, blue and orange, respectively. The β-strands are shown in cyan, the α-helices in salmon, the 3$_{10}$-helix in green and the random coils in gray. (*c*), (*d*) Surface representation of the hydrophobic pocket of yeast Tim44 (*c*) and NTF2 (*d*), viewed in a similar orientation as in Fig. 4(*a*).

4*a* and 4*c*; Josyula *et al.*, 2006). This may explain the fact that mammalian Tim44 is found in both the soluble matrix fraction and the mitochondrial membranes (Bauer *et al.*, 1999; Ishihara & Mihara, 1998), whereas yeast Tim44p is tightly bound to the mitochondrial inner membrane (Blom *et al.*, 1993).

3.4. Polyethylene glycol-induced structural changes

As mentioned in the previous section, two molecules of pentaethylene glycol, PEG1 and PEG2, are bound in the large hydrophobic cavity (Figs. 4*a* and 4*b*). In the human Tim44 C-terminal domain, the region consisting of the loop between α1 and α2 interacts with the PEG2* molecule, which is generated by a crystallographic symmetry operation on PEG2

(a)

(b)
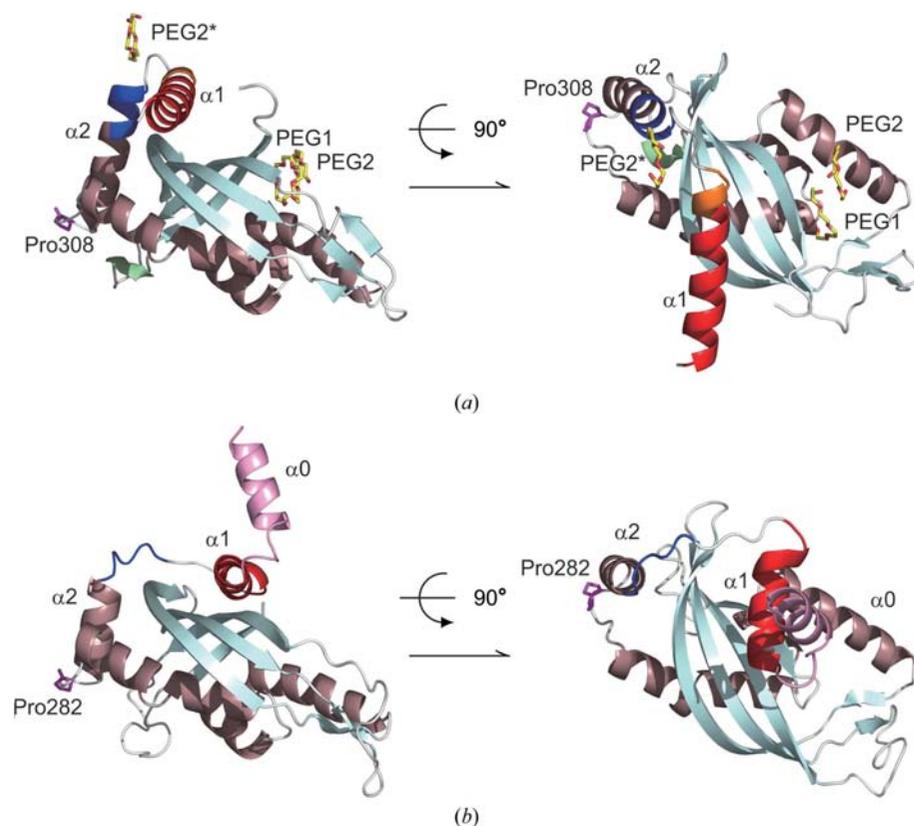
Figure 5
Comparison of the loop between the α1 and α2 helices of human and yeast Tim44. Close-up view of human (*a*) and yeast (*b*) Tim44 shown in similar orientations after optimal superposition. The Pro308 residue in human Tim44 and the corresponding residue, Pro282, in yeast Tim44 are highlighted by purple stick models. The other coloring for the cartoon diagram is the same as in Fig. 4(*b*). The highly conserved G-(L/F)-F-(S/A) sequence is highlighted as a stick model.

(Fig. 1*a*). This binding site may also be the mitochondrial inner membrane-binding site. The residues around this binding site are highly conserved, whereas those situated on the surface of the large hydrophobic cavity show relatively weak conservation (Figs. 2 and 3*b*). The consensus sequence around the pentaethylene glycol-binding site is 289-G-(L/F)-F-(S/A)-(K/E)-T-E-295. The side chain and main chain of Leu290, the main chain of Gly289 and the side chain of Ser292 contact the pentaethylene glycol molecule in the human Tim44 crystal structure (Fig. 5*a*). However, the three-dimensional structure of this consensus sequence is completely different in the human and yeast structures (Figs. 5*a* and 5*b*; Josyula *et al.*, 2006). The α2 helix of human Tim44 is five residues longer at the N-terminus than that of yeast Tim44 because the corresponding residues in yeast form an elongated loop structure rather than a helical structure (Figs. 2, 5*a* and 5*b*). The B factors around this region are relatively low in the human structure (21.1–30.0 Å^2; Thr284–Met296), while those in the corresponding region in the yeast Tim44 crystal structure are quite high (greater than 90 Å^2; Val261–Ser270), suggesting that the binding of the pentaethylene glycol molecule stabilizes the three-dimensional structure of this region. This ligand-induced conformational change results in the slippage of the α1 helix along the large β-sheet surface (Fig. 6). In the human Tim44 crystal structure, the α1 helix associates with the β1b and β2a strands *via* direct or water-mediated hydrogen bonds (Fig. 7). In the yeast structure, the α1 helix associates with the β3 and β4 strands *via* van der Waals or hydrophobic interactions (Josyula *et al.*, 2006).

We speculated that the lipid-induced conformational changes in Tim44 might play an important role in importing mitochondrial matrix-destined proteins across the inner membrane. Based on the pentaethylene glycol binding sites, the sliding α1 helix may be on the membrane side of the large β-sheet surface of the Tim44 C-terminal domain (Fig. 6*a*). Thus, the lipid-induced sliding of α1 may contribute to the stabilization of Tim44 and the membrane interaction to some extent.

The deletion of an 18-residue segment (residues 185–202) of yeast Tim44p weakens the interaction of Tim44p with mtHsp70 (Merlin *et al.*, 1999). This 18-residue segment is close to the C-terminal domain and therefore the lipid-induced conformational change of Tim44 in the mtHsp70 complex directly alters the position of mtHsp70 relative to the membrane. Results obtained with purified yeast mtHsp70 and Tim44p showed that these proteins form a stable complex in the presence of ATP or ADP (Liu *et al.*, 2003). The mtHsp70–ATP–Tim44p complex is released without ATP hydrolysis upon binding a model peptide, consistent with the Brownian-ratchet model of translation (Liu *et al.*, 2003). It is possible that the lipid-induced conformational change of Tim44 facilitates the dissociation of the mtHsp70–ATP–Tim44 complex. The large hydrophobic cavity of Tim44 might be stably bound to the inner membrane and the α1–α2 loop might associate with the membrane when mtHsp70 binds the incoming peptide chain. It is also possible that this conformational change of Tim44 produces a mechanical power stroke to pull on the

Figure 6
α1 Helix slippage along the large β-sheet of Tim44. Ribbon diagrams of human (*a*) and yeast (*b*) Tim44 shown in similar orientations after optimal superposition. The right panel represents a 90° rotation of the left panel around the *x* axis. The α0 helix in yeast Tim44 is colored pink. The α1 helix of yeast Tim44 and the corresponding region in the α1 helix of human Tim44 are colored red. Four residues, Thr284, Asp285, Leu286 and Leu287, of human Tim44 are colored orange. The yeast Tim44 lacks the corresponding amino acids. The five residues Ser292, Lys293, Thr294, Glu295 and Met296 in the α2 helix of human Tim44 and the corresponding five residues Ala266, Glu267, Thr268, Glu269 and Ser270 in an elongated loop structure of yeast Tim44 are colored blue. The other coloring for the cartoon diagram is the same as in Fig. 5.
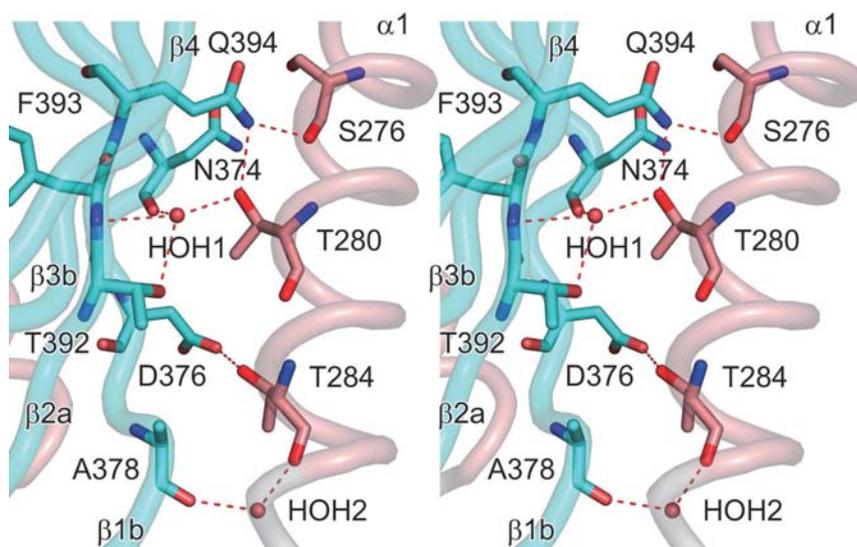
Figure 7
Cartoon representation of the interaction of the α1 helix and the β-sheet of human Tim44 (stereoview). The residues that interact by direct or water-mediated hydrogen bonds are shown as stick models. The interacting water molecules are shown as red dots. Hydrogen bonds are indicated by red dashed lines.

incoming polypeptide chain, although recent studies have suggested that the translocation of polypeptides can be accelerated by unfolding and pulling in the absence of a mechanical power stroke (De Los Rios *et al.*, 2006; Neupert & Brunner, 2002; Okamoto *et al.*, 2002).

3.5. The familial oncocytic thyroid carcinoma mutation site in Tim44

The missense mutation P308Q in human Tim44 causes familial oncocytic thyroid carcinoma (Bonora, Evangelisti *et al.*, 2006). Pro308 is located in the loop between α2 and α3 within a highly conserved motif, 307-D-P-A/T/S/N-F-310 (Fig. 2). An *in vitro* study revealed that the interaction between Tim44 and mtHsp70 is not abolished by the P308Q mutation (Bonora, Evangelisti *et al.*, 2006). Our crystal structure suggests that Pro308 stabilizes the structure of the α2–α3 loop because the proline residue restricts the main chain (Fig. 5). The *B* factors of this highly conserved motif, D-P-A/T/S/N-F, are relatively low in both the free and lipid-bound forms of Tim44 (the backbone *B* factors of the motif are 67.1–88.6 and 25.2–26.3 Å^2, respectively, whereas the average backbone *B* factors of the entire structure are 87.6 and 31.1 Å^2, respectively). We attempted to crystallize the P308 mutant of the human Tim44 C-terminal domain, but no crystals were obtained using either the same crystallization conditions employed for the wild type or several commercially available crystallization screening kits. The P308Q mutation would reduce the stability of the α2–α3 loop, thus shifting the α2 helix away from the correct position. This may cause the structural instability of the C-terminal domain of Tim44. Proline mutations often lead to conformational changes and thus are a factor in several diseases (Kanekura *et al.*, 2006).

Structural instability of the C-terminal domain, especially the α2–α3 loop and the α2 helix, would affect the slippage of the α1 helix along the β-sheet surface. This would impair the protein import across the mitochondrial inner membrane. The import of superoxide

research papers

dismutase and glutathione peroxidase is carried out by Tim44 and delivery of the Tim44 gene into STZ-induced diabetic model rats or mice reduces ROS production and renal cell proliferation (Matsuoka *et al.*, 2005; Zhang *et al.*, 2006). The oncocytic carcinoma cell model is associated with enhanced ROS production (Bonora, Porcelli *et al.*, 2006). These findings suggest that the P308Q mutation in Tim44 impairs the anti-oxidative enzyme import and increases ROS production, thus triggering carcinogenesis. Further studies based on this structure will clarify the mechanisms of the mitochondrial protein import and the carcinogenesis of the oncocytic tumors.

References

Bailey, S., Ward, D., Middleton, R., Grossmann, J. G. & Zambryski, P. C. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 2582–2587.

Bauer, M. F., Gempel, K., Reichert, A. S., Rappold, G. A., Lichtner, P., Gerbitz, K. D., Neupert, W., Brunner, M. & Hofmann, S. (1999). *J. Mol. Biol.* **289**, 69–82.

Bayliss, R., Leung, S. W., Baker, R. P., Quimby, B. B., Corbett, A. H. & Stewart, M. (2002). *EMBO J.* **21**, 2843–2853.

Blom, J., Kubrich, M., Rassow, J., Voos, W., Dekker, P. J., Maarse, A. C., Meijer, M. & Pfanner, N. (1993). *Mol. Cell. Biol.* **13**, 7364–7371.

Bonora, E., Evangelisti, C., Bonichon, F., Tallini, G. & Romeo, G. (2006). *Br. J. Cancer*, **95**, 1529–1536.

Bonora, E., Porcelli, A. M., Gasparre, G., Biondi, A., Ghelli, A., Carelli, V., Baracca, A., Tallini, G., Martinuzzi, A., Lenaz, G., Rugolo, M. & Romeo, G. (2006). *Cancer Res.* **66**, 6087–6096.

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.

Bullock, T. L., Clarkson, W. D., Kent, H. M. & Stewart, M. (1996). *J. Mol. Biol.* **260**, 422–431.

De Los Rios, P., Ben-Zvi, A., Slutsky, O., Azem, A. & Goloubinoff, P. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 6166–6171.

Fribourg, S., Braun, I. C., Izaurralde, E. & Conti, E. (2001). *Mol. Cell*, **8**, 645–656.

Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E. & Ben-Tal, N. (2003). *Bioinformatics*, **19**, 163–164.

Glick, B. S. (1995). *Cell*, **80**, 11–14.

Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). *Bioinformatics*, **15**, 305–308.

Hoelz, A., Nairn, A. C. & Kuriyan, J. (2003). *Mol. Cell*, **11**, 1241–1251.

Ishihara, N. & Mihara, K. (1998). *J. Biochem. (Tokyo)*, **123**, 722–732.

Josyula, R., Jin, Z., Fu, Z. & Sha, B. (2006). *J. Mol. Biol.* **359**, 798–804.

Kanekura, K., Nishimoto, I., Aiso, S. & Matsuoka, M. (2006). *J. Biol. Chem.* **281**, 30223–30233.

Kigawa, T., Yabuki, T., Matsuda, N., Matsuda, T., Nakajima, R., Tanaka, A. & Yokoyama, S. (2004). *J. Struct. Funct. Genomics*, **5**, 63–68.

Kigawa, T., Yabuki, T. & Yokoyama, S. (1999). *Tanpakushitsu Kakusan Koso*, **44**, 598–605.

Kigawa, T., Yabuki, T., Yoshida, Y., Tsutsui, M., Ito, Y., Shibata, T. & Yokoyama, S. (1999). *FEBS Lett.* **442**, 15–19.

Laan, M. van der, Rissler, M. & Rehling, P. (2006). *FEMS Yeast Res.* **6**, 849–861.

Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. (1996). *J. Biomol. NMR*, **8**, 477–486.

Laue, T. M., Shah, B., Ridgeway, T. M. & Pelletier, S. L. (1992). In *Analytical Ultracentrifugation in Biochemistry and Polymer Science*, edited by S. E. Harding, A. J. Rowe & J. Horton. Cambridge: Royal Society of Chemistry.

Li, K., Zhao, K., Ossareh-Nazari, B., Da, G., Dargemont, C. & Marmorstein, R. (2005). *J. Biol. Chem.* **280**, 29176–29185.

Liu, Q., D'Silva, P., Walter, W., Marszalek, J. & Craig, E. A. (2003). *Science*, **300**, 139–141.

Lundqvist, T., Rice, J., Hodge, C. N., Basarab, G. S., Pierce, J. & Lindqvist, Y. (1994). *Structure*, **2**, 937–944.

Matouschek, A., Pfanner, N. & Voos, W. (2000). *EMBO Rep.* **1**, 404–410.

Matsuoka, T., Wada, J., Hashimoto, I., Zhang, Y., Eguchi, J., Ogawa, N., Shikata, K., Kanwar, Y. S. & Makino, H. (2005). *Diabetes*, **54**, 2882–2890.

Merlin, A., Voos, W., Maarse, A. C., Meijer, M., Pfanner, N. & Rassow, J. (1999). *J. Cell Biol.* **145**, 961–972.

Moro, F., Sirrenberg, C., Schneider, H. C., Neupert, W. & Brunner, M. (1999). *EMBO J.* **18**, 3667–3675.

Neupert, W. & Brunner, M. (2002). *Nature Rev. Mol. Cell Biol.* **3**, 555–565.

Okamoto, K., Brinker, A., Paschen, S. A., Moarefi, I., Hayer-Hartl, M., Neupert, W. & Brunner, M. (2002). *EMBO J.* **21**, 3659–3671.

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

Pfanner, N. & Geissler, A. (2001). *Nature Rev. Mol. Cell Biol.* **2**, 339–349.

Schuck, P. (1998). *Biophys. J.* **75**, 1503–1512.

Stewart, M., Kent, H. M. & McCoy, A. J. (1998). *J. Mol. Biol.* **277**, 635–646.

Terradot, L., Bayliss, R., Oomen, C., Leonard, G. A., Baron, C. & Waksman, G. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 4596–4601.

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

Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). *Nucleic Acids Res.* **25**, 4876–4882.

Wada, J. & Kanwar, Y. S. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 144–149.

Wada, T., Shirouzu, M., Terada, T., Ishizuka, Y., Matsuda, T., Kigawa, T., Kuramitsu, S., Park, S.-Y., Tame, J. R. H. & Yokoyama, S. (2003). *Acta Cryst.* D**59**, 1213–1218.

Weiss, C., Oppliger, W., Vergeres, G., Demel, R., Jeno, P., Horst, M., de Kruijff, B., Schatz, G. & Azem, A. (1999). *Proc. Natl Acad. Sci. USA*, **96**, 8890–8894.

Zhang, Y., Wada, J., Hashimoto, I., Eguchi, J., Yasuhara, A., Kanwar, Y. S., Shikata, K. & Makino, H. (2006). *J. Am. Soc. Nephrol.* **17**, 1090–1101.