
Structure and ligand binding of the soluble domain of a *Thermotoga maritima* membrane protein of unknown function TM1634

CLARE J. MCCLEVERTY,^{1,3} LINDA COLUMBUS,²⁻⁴ ANDREAS KREUSCH,¹
AND SCOTT A. LESLEY^{1,2}

¹The Genomics Institute of the Novartis Research Foundation, San Diego, California 92121, USA

²The Joint Center for Structural Genomics and The Scripps Research Institute, Department of Molecular Biology, La Jolla, California 92037, USA

(RECEIVED January 3, 2008; FINAL REVISION February 12, 2008; ACCEPTED February 12, 2008)

Abstract

As a part of the Joint Center for Structural Genomics (JCSG) biological targets, the structures of soluble domains of membrane proteins from *Thermotoga maritima* were pursued. Here, we report the crystal structure of the soluble domain of TM1634, a putative membrane protein of 128 residues (15.1 kDa) and unknown function. The soluble domain of TM1634 is an α -helical dimer that contains a single tetratricopeptide repeat (TPR) motif in each monomer where each motif is similar to that found in Tom20. The overall fold, however, is unique and a DALI search does not identify similar folds beyond the 38-residue TPR motif. Two different putative ligand binding sites, in which PEG200 and Co^{2+} were located, were identified using crystallography and NMR, respectively.

Keywords: TPR motif; *Thermotoga maritima*; Co^{2+} ligand; PEG binding site; membrane protein; protein of unknown function

One of the aims of the Joint Center for Structural Genomics (JCSG, www.jcsg.org) is the complete protein structural coverage of a simple model organism *Thermotoga maritima* (≈ 1858 proteins) (Lesley et al. 2002; DiDonato et al. 2004). At present, 272 *T. maritima* protein structures (nonredundant) have been deposited in the Protein Data Bank with 166 of these structures determined by the JCSG. Approximately, 24% of the *T. maritima* proteome is classified as membrane proteins, with 169 putative single transmembrane α -helical proteins comprising 9% of the entire *T. maritima* proteome. This distribution is similar to other bacterial membrane

proteomes; however, included in this count are proteins with signal peptides for secretion, localization to the periplasm, and lipid modification. Of the 169 putative single transmembrane α -helix proteins, approximately 130 (7% of the proteome) are predicted to localize to the inner membrane.

Forty-nine of these proteins were recloned without the N-terminal hydrophobic sequences and expressed and purified as soluble proteins. Seven of these proteins now have structures deposited in the Protein Data Bank. TM1410 (pdb id 2aam) and TM1622 (pdb id 1vr8) (Xu et al. 2006) are putative lipoproteins (Madan Babu and Sankaran 2002). TM1223 (pdb id 1vr5) is a periplasmic protein (Nanavati et al. 2006), and TM0189 (pdb id 2etv), TM0957 (pdb id 2f4i), TM0961 (pdb id 2etd), and TM1553 (pdb id 1vrm) (Han et al. 2006) are proteins that contain a putative, single, N-terminal, transmembrane α -helix.

Here, we report the crystal structure of the soluble domain of TM1634, a putative membrane protein of 128 residues (15.1 kDa) and unknown function. The topology

³These authors contributed equally to this work.

⁴Present address: Department of Chemistry, University of Virginia, Charlottesville, VA 22904-4319, USA.

Reprint requests to: Linda Columbus, Department of Chemistry, University of Virginia, Charlottesville, VA 22904-4319, USA; e-mail: columbus@virginia.edu; fax: (434) 924-3567.

Article published online ahead of print. Article and publication date are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.083432208>.

of TM1634 is also unknown; however, the positive-inside rule suggests that the acidic soluble domain (calculated $pI \approx 5.6$) is located in the periplasm (von Heijne 1986, 1989). Sequence comparison identifies only one other homolog that is found in *Thermotoga petrophila*. A sequence alignment shows they are 98% identical (Fig. 1). A single tetratricopeptide repeat (TPR) motif was predicted for residues 75–108 by PROSITE (Hulo et al. 2006) and is colored red in Figure 1. The consensus sequence of a TPR motif is $W_4G_8Y_{11}G_{15}Y_{17}A_{20}Y_{24}A_{27}P_{32}$; however, no position is invariant. The TPR structural motif is found throughout prokaryotic and eukaryotic proteomes. TPR motifs mediate protein–protein interactions and the assembly of multi-protein complexes. TPR containing proteins are involved in biological processes, such as cell cycle regulation, transcriptional control, mitochondrial and peroxisomal protein transport, neurogenesis, and protein folding (D’Andrea and Regan 2003). The TPR motif adopts a helix (A)–turn–helix (B) arrangement and consists of three to 16 tandem repeats. The repeats pack in a parallel arrangement where helix A interacts with helix B and the preceding A helix (A’). As a result, a right-handed superhelix is formed with an inner concave surface predominantly formed by the A helices and a convex surface formed by residues from both the A and B helices (D’Andrea and Regan 2003). To date, the only available protein structure with just a single TPR motif is the soluble domain of the mitochondrial protein Tom20 (Abe et al. 2000).

The crystal structure of the soluble domain of TM1634 reveals a dimer that contains a single TPR motif in each monomer where each motif is similar to that found in Tom20. The overall fold is unique and a DALI search does not identify similar folds beyond the 38-residue TPR motif (less than 1/3 of the entire fold) (Holm and Sander

1995). Two different putative ligand binding sites, PEG200 and Co^{2+} , were identified using crystallography and NMR, respectively.

Results

Crystal structure of $\Delta 26TM1634$

The structure of the soluble domain of TM1634 ($\Delta 26TM1634$) was determined in three crystal forms (Table 1). $\Delta 26TM1634$ is a dimer in all crystal forms (Fig. 2). Each monomer is composed of five α -helices that fold into two domains. The N-terminal domain (residues 27–61) contains the first two α -helices and the C-terminal domain is a bundle of three α -helices, which contains the TPR motif ($\alpha 3$ and $\alpha 4$; residues 70–104) (Figs. 1, 2A). The dimer interface is highly hydrophobic (Fig. 2B) and primarily formed by $\alpha 1$ with 2340 \AA^2 of buried surface area. The $\alpha 1$ sequence contains a leucine heptad and forms a coiled-coil structure.

The crystallographic dimer is asymmetric: The orientation of the C-terminal domain with respect to the N-terminal domain is different for each monomer. The superposition of the N-terminal domain of the two conformers and the resulting position of the C-terminal domains are shown in Figure 3. The two distances shown in Figure 3B,C emphasize the difference between conformers *a* and *b* in the asymmetric dimer. A salt bridge is present in conformer *b* between Lys100 and Asp49 (2.8 \AA), as well as a hydrogen bond between the hydroxyl of Tyr45 and the backbone amide of Lys74 (3.1 \AA) that are absent in conformer *a* (9.5 \AA and 6.1 \AA , respectively). A domain motion analysis using the program DynDom (Hayward and Berendsen 1998) identifies residues Leu70, Leu71, Lys72, Glu73, as the hinge region (Fig. 3A). The observed difference between the open and closed conformers (*a* and *b*, respectively) results from a 64° rotation of the C-terminal domain around the hinge. The asymmetric dimer is observed in both native crystal forms with a 3.3 \AA pairwise RMSD for the $C\alpha$ of 103 equivalent residues.

Oligomerization of $\Delta 26TM1634$ in solution

Although the gel-filtration chromatogram (Fig. 4A) suggests the protein is trimeric in solution, the analytic ultracentrifugation and chemical cross-linking data are consistent with a dimer in solution (Fig. 4B and C, respectively). Trace amounts of trimer are observed in the analytical ultracentrifugation and cross-linking (120 min time point) experiments. The aberrant behavior during gel-filtration chromatography can be explained by the nonglobular shape of the protein dimer as observed in the crystal. The soluble domain dimer is also consistent with

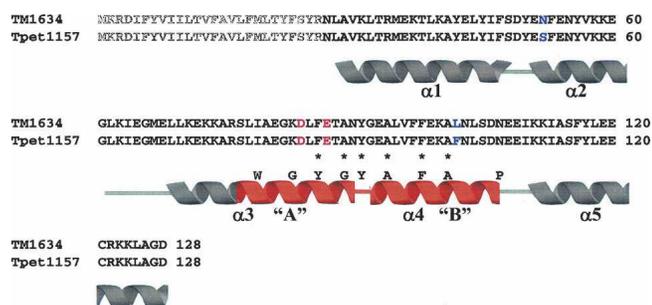


Figure 1. Sequence alignment of TM1634 and Tpet1157. The sequences of TM1634 and Tpet1157 of *T. petrophila* were aligned using BLAST. The hydrophobic N-terminal sequence is colored gray, the secondary structure is schematically drawn for α -helices, and the TPR consensus sequence is indicated underneath the protein sequences with the conserved residues with TM1634 and Tpet1157 indicated with stars. The two residues that are different between the two sequences are colored blue. Residues implicated in Co^{2+} binding are colored magenta.

Table 1. Data collection and refinement statistics

	TM1634 (Se) SAD	TM1634 (I)	TM1634 (II)
Crystallization conditions			
Buffer	30% PEG 400, 0.2 M Li ₂ SO ₄ , 0.1 M sodium cacodylate, pH 6.5	30% PEG 400, 0.2 M Li ₂ SO ₄ , 0.1 M sodium cacodylate, pH 6.5	50% PEG 200, 0.1 M sodium citrate, pH 5.5
Temperature (°C)	20	4	20
Crystal parameters			
Space group	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2	<i>P</i> 2 ₁
Cell dimensions (Å)	<i>a</i> = 46.1 <i>b</i> = 86.5 <i>c</i> = 110.5	<i>a</i> = 68.2 <i>b</i> = 76.2 <i>c</i> = 46.2	<i>a</i> = 50.9 <i>b</i> = 82.0 <i>c</i> = 68.3 β = 101.4°
Number of molecules/AU	4	2	4
Data collection			
Wavelength (Å)	0.97887	1.0000	1.0000
Resolution (Å)	50–2.7	50–1.65	50–1.8
Total reflections	79,862	104,892	163,714
Unique reflections	12,321	28,757	51,063
<i>R</i> _{merge} ^{a,b}	0.071 (0.248)	0.061 (0.409)	0.079 (0.568)
<i>I</i> / σ ^a	8.9	23.8	17.9
Completeness (%) ^a	96.6 (75.4)	96.6 (88.3)	99.2 (99.6)
Redundancy	6.5 (3.9)	3.6 (2.5)	3.2 (3.2)
Refinement			
Resolution (Å)		50–1.65	50–1.8
Reflections used		27,259	48,429
<i>R</i> _{work} / <i>R</i> _{free} ^c		0.20/0.23	0.24/0.28
No. of protein atoms		1726	3384
No. of water molecules		215	213
Average B value (Å ²) ^d		27.7	34.2 (39.6)
Ramachandran Statistics ^c			
Molecule A		96.0/4.0	96.9/3.1
Molecule B		96.0/4.0	95.9/4.1
Molecule C			96.9/3.1
Molecule D			97.9/2.1
RMSD bond lengths (Å)		0.011	0.011
RMSD bond angles (°)		1.4	1.3

^aNumbers in parentheses are for outer shell.

^b $R_{\text{merge}} = \frac{\sum_h \sum_j |I_{hj} - I_h|}{\sum_h \sum_j I_{hj}}$, where I_h is the weighted mean intensity of the symmetry-related reflections I_{hj} .

^c R_{work} and $R_{\text{free}} = \frac{\sum_{hkl} |F_{\text{obs}} - F_{\text{calc}}|}{\sum_{hkl} F_{\text{obs}}}$ for the working set and test set (5%) of reflections, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively.

^dNumber in parenthesis is for ligand.

^ePercentage of residues in most favored/additionally allowed regions according to PROCHECK (Laskowski et al. 1993).

the oligomeric state of the full-length protein in several detergents (Columbus et al. 2006).

PEG binding site

Crystal form II showed additional density that was not part of the protein. Based on the components in the crystal condition and the shape of the electron density, PEG200 was modeled into the density (Figs. 2A, 5). PEG200 only binds to molecule B of the dimer between the N- and C-terminal domains. Residues from $\alpha 1$ to $\alpha 3$ contribute to the binding site. Specifically, Lys74 forms a bifurcated hydrogen bond to two oxygen atoms of the PEG200 (2.9 and 3.2 Å) and Phe97, Tyr45, and Leu78 are within van der Waals distance to the PEG200 carbon atoms (Fig. 5). The PEG molecule does not seem to be responsible for the observed asymmetry in the dimers

since crystal form I also contains an asymmetric dimer but lacks the additional ligand electron density. It is possible that the precipitant in the crystallization conditions for crystal form I, PEG400, is too large to occupy the binding site observed in crystal form II.

Ligand binding in solution

In order to investigate the binding of ligands in solution, the backbone atoms of $\Delta 26$ TM1634 were assigned by NMR. Three potential ligands were investigated, PEG200, glycyl-glycyl-glycine, and CoCl₂. PEG200 was chosen based on the electron density in the crystal structure. Glycyl-glycyl-glycine was chosen to investigate if the observed PEG200 molecule was occupying a potential peptide binding site. CoCl₂ was chosen because, when the thrombin-treated protein was loaded onto a Co²⁺-chelating

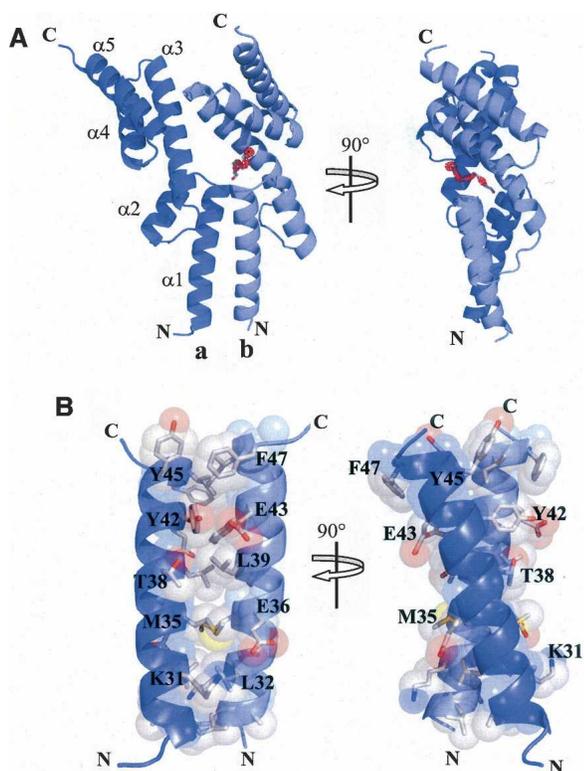


Figure 2. The crystal structure of the soluble domain of TM1634. (A) The dimer is shown as a ribbon diagram with one monomer colored slate and the other colored light blue. The FoFc electron density (red) corresponds to the PEG200 molecule in one of the monomers and is contoured at 3σ . The five α -helices are labeled $\alpha 1$ – $\alpha 5$. (B) The $\alpha 1$ dimer interface is shown with the backbone rendered as a cartoon and the interfacial residues rendered as sticks with transparent spheres. The side chains are colored by atom (C, gray; N, blue; and O, red).

column to remove the cleaved His-tag and uncleaved protein, the flow-through containing cleaved $\Delta 26$ TM1634 was an intense pink color that could not be removed by dialysis. Ligand binding was assessed with native gel protein shifts and NMR chemical shift perturbation mapping. PEG200 and glycyl-glycyl-glycine binding could not be detected with either of these methods. A protein shift in a native gel was only observed for Co^{2+} (Fig. 6A) and localized chemical shift perturbations were observed in the ^{15}N , ^1H -HSQC spectra (Fig. 6B).

The chemical shift perturbation is localized to a single helical turn (Lys83–Glu87) in the C-terminal TPR domain. In addition to changes in chemical shift, the paramagnetic Co^{2+} enhances the relaxation of the amide proton and, therefore, broadens the line widths (in some cases beyond detection). Histidine, aspartate, glutamate, and cysteine are residues that are known to coordinate cobalt in either an octahedral or tetrahedral geometry. The two residues within this sequence that could coordinate Co^{2+} are Asp84 and Glu87. These residues are a

single helical-turn away from each other and near the dimer interface (Fig. 6C), which suggests that the Co^{2+} could be bound by these four residues; however, the asymmetric dimer in the crystal is not in a configuration that would facilitate such coordination.

It should be noted that the line widths are broad for a protein with a molecular weight of 25 kDa; however, there is no evidence for the asymmetric dimer in solution, i.e., two sets of resonances are not observed for a subset of cross-peaks. In addition, the ^{15}N , ^1H -HSQC spectrum is different from that reported for the full-length protein in LDAO (Columbus et al. 2006). The latter spectrum can be

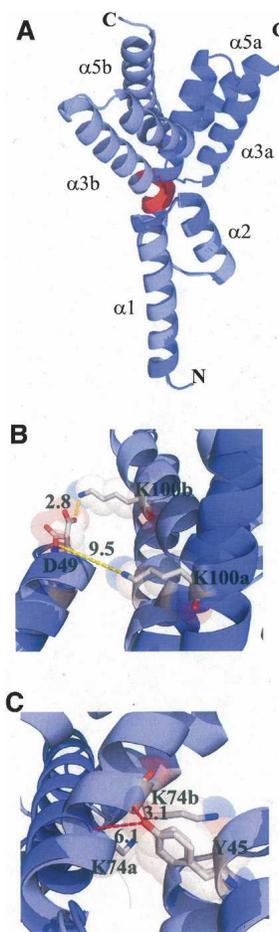


Figure 3. The $\Delta 26$ TM1634 asymmetric dimer. (A) A superposition of α -helices 1 and 2 from monomer *a* onto the equivalent residues from monomer *b* results in an RMSD of 0.55 \AA and highlights the different conformations of the C-terminal domains ($\alpha 3$, $\alpha 4$, and $\alpha 5$) in each monomer. The hinge region (residues 70–73) is colored red. (B) A salt bridge between K100 and D49 is shown in conformer *b* (2.8 \AA) that is not present in conformer *a* (9.5 \AA). (C) A hydrogen bond between the hydroxyl of Y46 and the backbone amide of K74 (3.1 \AA) is observed in conformer *b* but is not present in conformer *a* (6.1 \AA). Residues in panels *B* and *C* are labeled and distances (\AA) are shown as dashed lines.

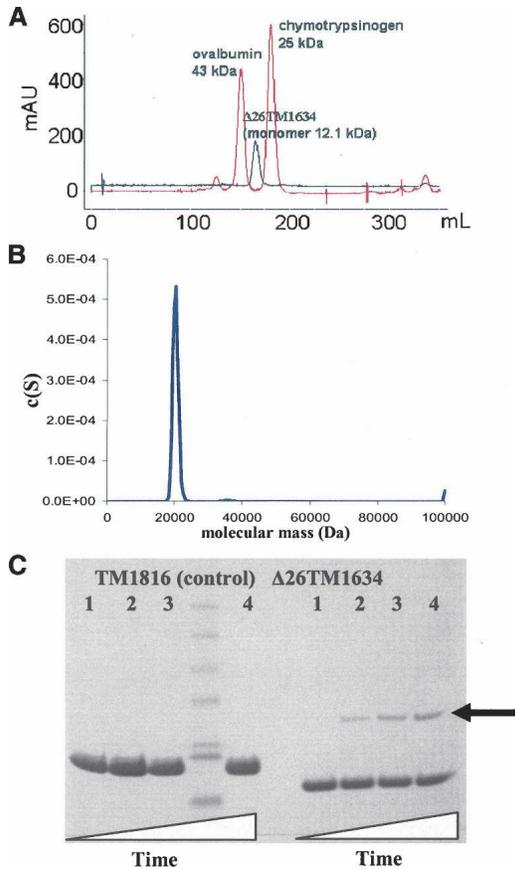


Figure 4. $\Delta 26\text{TM1634}$ is a dimer in solution. (A) Gel-filtration chromatogram of $\Delta 26\text{TM1634}$ and molecular weight standards ovalbumin and chymotrypsinogen. (B) Sedimentation velocity $c(s)$ distribution of $\Delta 26\text{TM1634}$ displaying a single monodisperse peak corresponding to a dimer. (C) Denaturing SDS-PAGE gel of $\Delta 26\text{TM1634}$ after incubation with EDC (0, 40, 60, and 120 min; lanes 1–4, respectively). TM1816 (a monomer in solution) was used as a control for collisional complex formation (Columbus et al. 2005). The arrow indicates dimeric $\Delta 26\text{TM1634}$. The unmarked lane between TM1816 lanes 3 and 4 contains the molecular weight markers (from top to bottom: 62, 49, 38, 28, 17, 14, and 6 kDa).

obtained by adding LDAO to the soluble domain (data not shown) implying that the detergent alters the conformation of the soluble region in the protein–detergent complex.

Discussion

Ligand binding to $\Delta 26\text{TM1634}$

The crystal structure implies that there is a “closed” state of $\Delta 26\text{TM1634}$ to which a PEG200 molecule can bind. PEG molecules have mimicked ligand binding in several crystal structures, such as a putative odorant binding site of the odorant binding protein AgamOBP1 (Wogulis et al. 2006) and the cephalosporin binding site of the putative methyltransferase CmcI (Oster et al. 2006). However, the

binding site in $\Delta 26\text{TM1634}$ could not be confirmed in solution using PEG200 or glycyl-glycyl-glycine.

Co-crystallization experiments with $\Delta 26\text{TM1634}$ and CoCl_2 were not successful. For some crystal forms, analysis of X-ray data sets from native crystals soaked with CoCl_2 did not reveal a defined cobalt binding site. In other instances, the crystals dissolved with the addition of CoCl_2 . These results suggest that the protein in the crystal is not in a conformation suitable for cobalt binding without disruption of the lattice. A binding site could be mapped with solution NMR and specificity demonstrated with native gel protein shifts (Fig. 6). Several *T. maritima* proteins have been shown to bind Co^{2+} : alkaline phosphatase (TM0156) (Wojciechowski et al. 2002), cytosolic α -mannosidase (TM1851), and the N-terminal regulatory domain of CorA (TM0561) (Eshaghi et al. 2006). It is also likely that a class II xylose isomerase (TM1667) (Epting et al. 2005) and a methionine aminopeptidase (TM1478) (Spraggon et al. 2004) bind Co^{2+} , although the metal was not observed in the crystal structures.

T. maritima needs to import Co^{2+} from the environment, but only a few components for cobalt transport have been annotated (TM1663 and TM1868); however, a periplasmic binding domain for the transporter has still not been identified. In addition, uptake of Co^{2+} needs to be tightly regulated to avoid toxic effects. Thus, the possible role for a Co^{2+} binding membrane protein, such as TM1634, awaits further functional characterization.

Oligomeric state of $\Delta 26\text{TM1634}$

$\Delta 26\text{TM1634}$ is a dimer in solution as well as in the crystal. However, the asymmetrical dimer observed in the crystal is not in agreement with the solution NMR data. A single set of resonances is observed in the $^{15}\text{N}, ^1\text{H}$ -HSQC

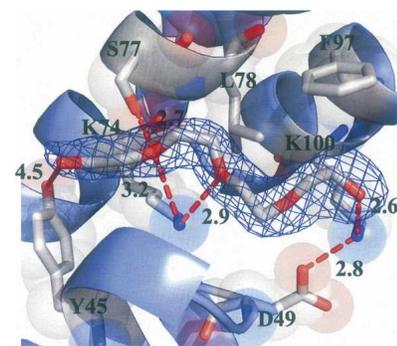


Figure 5. PEG200 binding site in the $\Delta 26\text{TM1634}$ crystal structure. D49, Y45, K74, S77, L78, F97, and K100 form the PEG200 binding site in the crystal. The 2FoFc electron density for the PEG200 is contoured at 1σ and the PEG200 molecule rendered as sticks. The protein backbone is rendered as a cartoon. The side chains are rendered as sticks with transparent spheres, labeled, and colored by atom (C, gray; N, blue; and O, red). Distances (Å) between atoms are indicated by dashed lines and labeled.

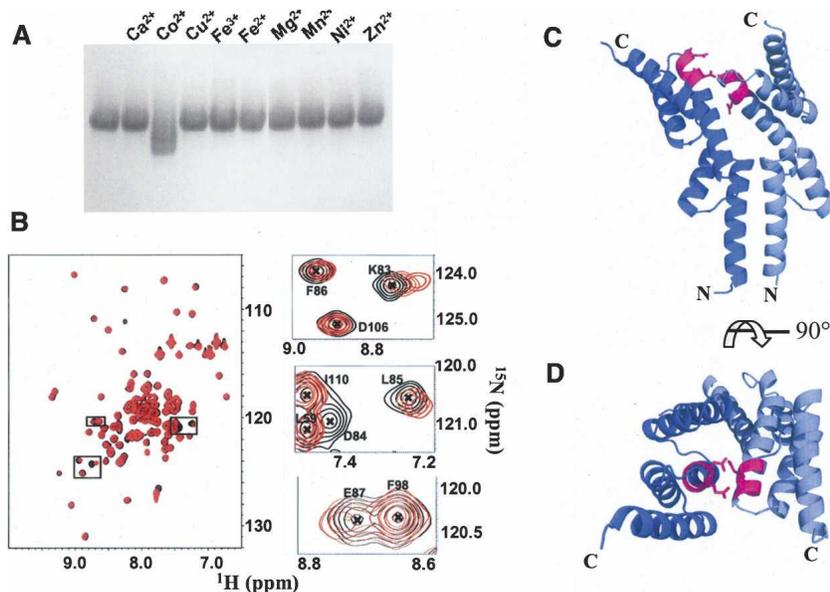


Figure 6. $\Delta 26\text{TM1634}$ binds Co^{2+} in solution. (A) A native gel of $\Delta 26\text{TM1634}$ incubated with different ions. (B) Two-dimensional ^{15}N , ^1H -HSQC spectra of ^{15}N -labeled $\Delta 26\text{TM1634}$. The red spectrum is without Co^{2+} and the black spectrum is with $600 \mu\text{M}$ CoCl_2 added. The panels on the *right* correspond to the boxes drawn on the entire spectrum and the peaks are labeled with the residue assignment, and an additional spectrum recorded with $300 \mu\text{M}$ CoCl_2 is shown in orange. (C) The $\Delta 26\text{TM1634}$ crystal structure with the region in which chemical shifts were observed in the presence of Co^{2+} . E87 and D84 are rendered as sticks. (D) A 90° rotation of the protein in panel C.

spectrum, which is consistent with a symmetrical dimer in solution. In addition, the Co^{2+} chemical shift mapping and structural analysis implies the solution dimer is symmetrical. Therefore, the asymmetry observed in the crystal structure may be due to crystal packing forces.

The TPR motif in the structure of $\Delta 26\text{TM1634}$

TM1634 contains a single TPR motif in each monomer and shares considerable structural similarity with the second TPR motif of the peroxin PEX5 (Kumar et al. 2001). Superposition of the structures is shown in Figure 7A. The structures share one and a half TPR motifs (55 residues) with a pairwise RMSD of 2.7 \AA . It is interesting that this motif was found to chelate Mg^{2+} , not at a dimer interface, but with a symmetry-related molecule (Kumar et al. 2001). Only the $\Delta 26\text{TM1634}$ monomers individually share the structural homology. The TPR motif tandem repeat requires the B helices to interact as well as the A helices (Fig. 7A). The $\Delta 26\text{TM1634}$ dimer has the B helices on the opposite sides of the A helices; therefore, the B helix of a TPR motif of one monomer does not interact with the B helix of the other.

The TPR motif in TM1634 is most similar to the only other structure with a single TPR motif, Tom20 (Abe et al. 2000; Saitoh et al. 2007). Tom20 is also a membrane protein containing a single transmembrane α -helix and a structural alignment of the $\Delta 26\text{TM1634}$ C-terminal domain results in a pairwise RMSD of 3.0 \AA for 57 residues

(Fig. 7B). Tom20 is a mitochondrial protein that recognizes the signal peptide for proteins being imported into the mitochondria. Using the structural similarity of the proteins, the signal peptide binding site of Tom20 was mapped onto TM1634 (Fig. 7C,D; Abe et al. 2000) where the structurally equivalent region is at the dimer interface and likely does not bind a signal peptide.

Model of the full-length TM1634 membrane protein

TM1634 is an integral membrane protein with each monomer containing a single transmembrane α -helix. A model of the full-length dimeric protein is depicted in Figure 8. The N-terminal helices were extended by including the 26 residues which were removed to obtain the structure. It is likely that only the three charged residues (Lys2, Arg3, and Asp4) at the N terminus extend from the membrane opposite of the soluble domain, and topology experiments in the bacterial membrane need to be performed to determine if the soluble domain is cytoplasmic or periplasmic.

Materials and Methods

Preparation of $\Delta 26\text{TM1634}$

Cloning and expression of $\Delta 26\text{TM1634}$

Primers were designed to amplify the soluble domain (residues 27–128) of TM1634 from the clone of the full-length gene

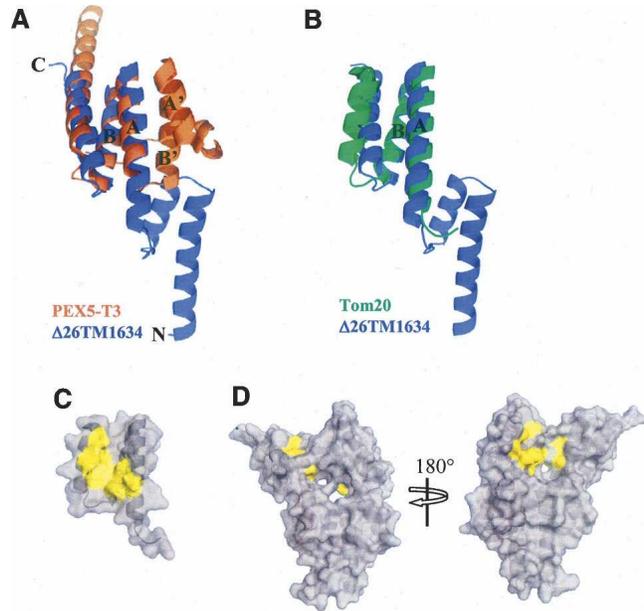


Figure 7. Structural similarity to the TPR motif. (A) The structural alignment of $\Delta 26\text{TM1634}$ (slate blue) and the third TPR motif of the peroxin pex5 from *Trypanosoma brucei* (pdb id: 1HXI, orange); the A and B helices of the aligned TPR motif are labeled, and the helices of the tandem repeat are labeled A' and B'. (B) Structural alignment of $\Delta 26\text{TM1634}$ and the soluble domain of the mitochondrial import receptor subunit Tom20 from *Rattus norvegicus* (pdb id: 2V1T); $\Delta 26\text{TM1634}$ is colored slate blue and Tom20 is colored green. In both panels, the TPR helices are labeled A and B and correspond to those in Figure 1. (C) Tom20 is rendered as a cartoon with a semitransparent surface. The signal peptide-binding region is colored yellow. (D) $\Delta 26\text{TM1634}$ is rendered as a cartoon with a semitransparent surface and the corresponding region to the Tom20 peptide-binding region is colored yellow.

(Lesley et al. 2002). Using standard molecular biology techniques, the amplified soluble domain was ligated into the pET28b vector between the NdeI and BamHI restriction sites. This vector encodes a Thrombin cleavable N-terminal His₆-tag (MGSSHHHHHHSSGLVPRGSHM). The final protein construct after Thrombin digestion has four additional N-terminal residues (GSHM) before N27 of TM1634. For expression, the $\Delta 26\text{TM1634}$ -containing plasmid was transformed into a BL21(DE3) *Escherichia coli* strain. Cell cultures were grown at 37°C to an OD₆₀₀ \approx 0.8 and protein expression was induced with 1 mM isopropyl- β -thio-D-galactoside. Unlabeled samples were grown in Luria-Bertani medium. For uniformly labeled $\Delta 26\text{TM1634}$, cells were grown in M9 minimal media supplemented with ¹⁵NH₄Cl (1 g/L) and/or [¹³C]-D-glucose (4 g/L) for obtaining ¹⁵N- or ¹⁵N/¹³C-labeled $\Delta 26\text{TM1634}$.

Purification of $\Delta 26\text{TM1634}$

The cells were harvested 3 h after induction by centrifugation at 5000g for 15 min. Bacteria were resuspended in lysis buffer (50 mM Tris, pH 8.0, 100 mM NaCl, 1 Complete protease inhibitor pellet [Roche]) and lysed using a microfluidizer. Cell debris was removed by centrifugation at 15,000g for 30 min. The supernatant was applied to a column containing cobalt chelating resin (GE Healthcare) previously equilibrated with

lysis buffer. The resin was washed with 10 column volumes of wash buffer (25 mM sodium phosphate, pH 7.8, 100 mM NaCl, 25 mM imidazole) and the protein was eluted with five column volumes of elution buffer (wash buffer with 600 mM imidazole). The eluate was concentrated (MWCO = 10 kDa) to 10 mL and dialyzed against 2 \times 4 L of Thrombin cleavage buffer (50 mM Tris, pH 8.0, 100 mM NaCl, and 10 mM CaCl₂). Thrombin (10–20 μ g) was added and the cleavage reaction was allowed to proceed at room temperature for 2 d. The extent of cleavage was monitored using SDS-PAGE denaturing gels. The Thrombin was then removed using a p-aminobenzamide-agarose resin (Sigma-Aldrich). Initially, a second Co²⁺ chelating column was used to remove the cleaved His-tag and any uncleaved protein; however, the cleaved protein in the flow-through was an intense pink color that could not be removed, even after extensive dialysis. Therefore, for crystallization, NMR, and biochemical studies, the second cobalt chelating column was eliminated.

X-ray crystallography

Crystallization and data collection

$\Delta 26\text{TM1634}$ was dialyzed into crystallization buffer (20 mM Tris, pH 7.8, and 150 mM NaCl) and concentrated to 33 mg/mL. Crystals were grown by nanodrop vapor diffusion using the precipitants shown in Table 1. Crystals were flash-frozen to 100 K and native diffraction data were collected at Beamline 5.0.1, Advanced Light Source, Berkeley, CA. Single wavelength anomalous dispersion (SAD) data from a single selenomethionine-substituted crystal were collected at the selenium K-edge (peak) at Beamline 5.0.2, Advanced Light Source, Berkeley, CA. All data were processed with the HKL2000 package (Otwinowski and Minor 1997).

Phasing, model building, and refinement

The scaled intensity SAD data were input into SOLVE (Terwilliger and Berendzen 1999) and the positions of 12 selenium atoms in the asymmetric unit were located. Phases were calculated to 3.1 Å resolution by SOLVE with a figure of merit of 0.33 and the phases refined with RESOLVE (Terwilliger and Berendzen 1999). The C α atom trace for the four molecules in the asymmetric unit was manually built with COOT (Emsley



Figure 8. Structural model of the full-length TM1634. The lipid bilayer is indicated by the gray box and marks the region of the protein modeled by extending the N-terminal helices in the crystal structure.

and Cowtan 2004). A molecular replacement solution for the 1.65 Å native data set I was subsequently found with the program Phaser (Storoni et al. 2004) using a monomer as the search model. The two monomers in the asymmetric unit were refined to 1.65 Å resolution using Refmac5 within the CCP4 program suite (Collaborative Computational Project, Number 4 1994). An additional native data set II was collected, solved by molecular replacement, and refined as described above for the native data set I. The final model contains residues 27–128 in each molecule with 96% of the main chain torsion angles of the non-glycine residues in the most favored regions of the Ramachandran plot and the remaining 4% in the additionally allowed regions as evaluated by PROCHECK (Laskowski et al. 1993). All data collection and refinement statistics are shown in Table 1. Protein graphics were prepared using PyMOL (DeLano Scientific). The atomic coordinates and structure factors (codes 2VKJ and 2VKO, respectively) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

Characterizing the oligomerization state in solution

For solution studies, $\Delta 26\text{TM}1634$ was dialyzed into phosphate buffer (20 mM phosphate, pH 6.2, and 150 mM NaCl) and concentrated to 2 mM.

AUC

Analytical ultracentrifugation was performed on a Beckman XL-I equipped with an AnTi50 eight-hole rotor. Interference and absorbance optics were used with cells constructed from sapphire windows and 1.2 mm charcoal-filled Epon centerpieces. Samples for velocity experiments were equilibrated at 20°C for at least 1 h prior to initiation of runs at 50,000 rpm for 4 h. Scans were fit using the $c(s)$ analysis method with the programs SedFit or SedPhat (Schuck 2000).

Gel filtration

A HiLoad Superdex 75 FPLC column (GE Healthcare) was equilibrated with 20 mM phosphate buffer (pH 6.2) and 150 mM NaCl. A 0.5-mL sample of 1.5 mM $\Delta 26\text{TM}1634$ was injected onto the column and the elution profile was recorded. The procedure was repeated with a set of molecular weight standards (BioRad).

Chemical cross-linking

$\Delta 26\text{TM}1634$ was reacted with the hydrophilic cross-linker Ethyl-3-[3-dimethylaminopropyl]carbodiimide Hydrochloride (EDC). Protocols from Pierce were followed using the following reaction conditions: 50 mM HEPES (pH 8.0), 150 mM NaCl, 2.5 mM EDC, and ~ 10 μM protein. A 20- μL sample was removed from the reaction at 0, 40, 60, and 120 min and mixed with SDS loading buffer to stop the cross-linking reaction. The reaction products were evaluated with SDS-PAGE denaturing gels and the oligomeric state was estimated based on the migration of the reacted protein compared to unreacted protein. TM1816, a monomeric protein, was used as a control.

Ligand binding

Co^{2+} , PEG, and glycyl-glycyl-glycine NMR titrations

All NMR data were recorded at 303 K on a Bruker Avance600 spectrometer. For the sequence-specific resonance assignments

of the polypeptide backbone atoms, the following experiments were recorded: 2D ^{15}N , ^1H -HSQC, 3D HNCACB, 3D HNCA, and 3D CBCA(CO)NH (Bax and Grzesiek 1993). All assignments were done interactively using the program XEASY (Bartels et al. 1995). The backbone resonance assignments were obtained for all residues except Ala42, Glu100, and Lys101. A 2D ^{15}N , ^1H -HSQC spectrum was recorded for 300 μM ^{15}N -labeled $\Delta 26\text{TM}1634$ with 50 μM , 100 μM , 300 μM , 600 μM , 3.4 mM, and 15 mM CoCl_2 , 50 μM , 100 μM , 300 μM , 600 μM , 3.4 mM, and 15 mM PEG200, and 50 μM , 100 μM , 300 μM , and 600 μM glycyl-glycyl-glycine using a BACS120 autosampler.

Native gel assay of metal binding and selectivity

$\Delta 26\text{TM}1634$ was incubated with 10-fold molar excess of divalent salts for 10 min at room temperature. The complexes were monitored on a Coomassie-stained native 4%–20% gradient gel, which was run for 3 h. Gel migration was compared to the protein without divalent salt.

Acknowledgments

We thank Dr. Kurt Wüthrich for support and Drs. Gerard Kroon, Bernhard Geierstanger, and David Jones for helpful discussions and technical NMR assistance; Drs. Ted Foss and Jeffery Kelly for analytical ultracentrifugation support; Dr. Lukasz Jaroszewski for bioinformatics assistance; Heath Klock for clones; Eric Hampton for technical differential scanning calorimetry assistance; and the staff at the Advanced Light Source, BL501/502/503 (Berkeley, CA), for beamline assistance. Support for this research was provided by the National Institutes of Health, Protein Structure Initiative grant numbers: P50 GM62411 and U54 GM074898 and NIH grant 1F32GM068286 (L.C.).

References

- Abe, Y., Shodai, T., Muto, T., Mihara, K., Torii, H., Nishikawa, S., Endo, T., and Kohda, D. 2000. Structural basis of presequence recognition by the mitochondrial protein import receptor Tom20. *Cell* **100**: 551–560.
- Bartels, C., Xia, T.H., Billeter, M., Guntert, P., and Wuthrich, K. 1995. The program XEASY for computer-supported NMR spectral-analysis of biological macromolecules. *J. Biomol. NMR* **6**: 1–10.
- Bax, A. and Grzesiek, S. 1993. Methodological advances in protein NMR. *Acc. Chem. Res.* **26**: 131–138.
- Collaborative Computational Project, Number 4. 1994. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **D50**: 760–763.
- Columbus, L., Peti, W., Etezady-Esfarjani, T., Herrmann, T., and Wuthrich, K. 2005. NMR structure determination of the conserved hypothetical protein TM1816 from *Thermotoga maritima*. *Proteins* **60**: 552–557.
- Columbus, L., Lipfert, J., Klock, H., Millett, I., Doniach, S., and Lesley, S.A. 2006. Expression, purification, and characterization of *Thermotoga maritima* membrane proteins for structure determination. *Protein Sci.* **15**: 961–975.
- D'Andrea, L.D. and Regan, L. 2003. TPR proteins: The versatile helix. *Trends Biochem. Sci.* **28**: 655–662.
- DiDonato, M., Deacon, A.M., Klock, H.E., McMullan, D., and Lesley, S.A. 2004. A scalable and integrated crystallization pipeline applied to mining the *Thermotoga maritima* proteome. *J. Struct. Funct. Genomics* **5**: 133–146.
- Emsley, P. and Cowtan, K. 2004. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**: 2126–2132.
- Epting, K.L., Vieille, C., Zeikus, J.G., and Kelly, R.M. 2005. Influence of divalent cations on the structural thermostability and thermal inactivation kinetics of class II xylose isomerases. *FEBS J.* **272**: 1454–1464.
- Eshghi, S., Niegowski, D., Kohl, A., Martinez Molina, D., Lesley, S.A., and Nordlund, P. 2006. Crystal structure of a divalent metal ion transporter CorA at 2.9 angstrom resolution. *Science* **313**: 354–357.

- Han, G.W., Sri Krishna, S., Schwarzenbacher, R., McMullan, D., Ginalski, K., Elslinger, M.A., Brittain, S.M., Abdubek, P., Agarwalla, S., Ambing, E., et al. 2006. Crystal structure of the ApbE protein (TM1553) from *Thermotoga maritima* at 1.58 Å resolution. *Proteins* **64**: 1083–1090.
- Hayward, S. and Berendsen, H.J. 1998. Systematic analysis of domain motions in proteins from conformational change: New results on citrate synthase and T4 lysozyme. *Proteins* **30**: 144–154.
- Holm, L. and Sander, C. 1995. DALI: A network tool for protein structure comparison. *Trends Biochem. Sci.* **20**: 478–480.
- Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., De Castro, E., Langendijk-Genevaux, P.S., Pagni, M., and Sigrist, C.J. 2006. The PROSITE database. *Nucleic Acids Res.* **34**: D227–D230. doi: 10.1093/nar/gkj06.
- Kumar, A., Roach, C., Hirsh, I.S., Turley, S., deWalque, S., Michels, P.A., and Hol, W.G. 2001. An unexpected extended conformation for the third TPR motif of the peroxin PEX5 from *Trypanosoma brucei*. *J. Mol. Biol.* **307**: 271–282.
- Laskowski, R.J., MacArthur, N.W., Moss, D.S., and Thornton, J.M. 1993. PROCHECK: A program to check stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**: 283–290.
- Lesley, S.A., Kuhn, P., Godzik, A., Deacon, A.M., Mathews, I., Kreusch, A., Spraggon, G., Klock, H.E., McMullan, D., Shin, T., et al. 2002. Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. *Proc. Natl. Acad. Sci.* **99**: 11664–11669.
- Madan Babu, M. and Sankaran, K. 2002. DOLOP—database of bacterial lipoproteins. *Bioinformatics* **18**: 641–643.
- Nanavati, D.M., Thirangoon, K., and Noll, K.M. 2006. Several archaeal homologs of putative oligopeptide-binding proteins encoded by *Thermotoga maritima* bind sugars. *Appl. Environ. Microbiol.* **72**: 1336–1345.
- Oster, L.M., Lester, D.R., Terwisscha van Scheltinga, A., Svenda, M., van Lun, M., Genereux, C., and Andersson, I. 2006. Insights into cephamycin biosynthesis: The crystal structure of CmcI from *Streptomyces clavuligerus*. *J. Mol. Biol.* **358**: 546–558.
- Otwinowski, Z. and Minor, W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**: 307–326.
- Saitoh, T., Igura, M., Obita, T., Ose, T., Kojima, R., Maenaka, K., Endo, T., and Kohda, D. 2007. Tom20 recognizes mitochondrial presequences through dynamic equilibrium among multiple bound states. *EMBO J.* **26**: 4777–4787.
- Schuck, P. 2000. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* **78**: 1606–1619.
- Spraggon, G., Schwarzenbacher, R., Kreusch, A., McMullan, D., Brinen, L.S., Canaves, J.M., Dai, X., Deacon, A.M., Elslinger, M.A., Eshagi, S., et al. 2004. Crystal structure of a methionine aminopeptidase (TM1478) from *Thermotoga maritima* at 1.9 Å resolution. *Proteins* **56**: 396–400.
- Storoni, L.C., McCoy, A.J., and Read, R.J. 2004. Likelihood-enhanced fast rotation functions. *Acta Crystallogr. D Biol. Crystallogr.* **60**: 432–438.
- Terwilliger, T.C. and Berendzen, J. 1999. Automated MAD and MIR structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **55**: 849–861.
- von Heijne, G. 1986. Net N-C charge imbalance may be important for signal sequence function in bacteria. *J. Mol. Biol.* **192**: 287–290.
- von Heijne, G. 1989. Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. *Nature* **341**: 456–458.
- Wogulis, M., Morgan, T., Ishida, Y., Leal, W.S., and Wilson, D.K. 2006. The crystal structure of an odorant binding protein from *Anopheles gambiae*: Evidence for a common ligand release mechanism. *Biochem. Biophys. Res. Commun.* **339**: 157–164.
- Wojciechowski, C.L., Cardia, J.P., and Kantrowitz, E.R. 2002. Alkaline phosphatase from the hyperthermophilic bacterium *T. maritima* requires cobalt for activity. *Protein Sci.* **11**: 903–911.
- Xu, Q., Krishna, S.S., McMullan, D., Schwarzenbacher, R., Miller, M.D., Abdubek, P., Agarwalla, S., Ambing, E., Astakhova, T., Axelrod, H.L., et al. 2006. Crystal structure of an ORFan protein (TM1622) from *Thermotoga maritima* at 1.75 Å resolution reveals a fold similar to the Ran-binding protein Mog1p. *Proteins* **65**: 777–782.