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

Structure of the twin-arginine signal-binding protein DmsD from *Escherichia coli*

Suresh Kumar Ramasamy and William M. Clemons Jr*

Division of Chemistry and Chemical Engineering, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA

Correspondence e-mail: clemons@caltech.edu

Received 28 May 2009
Accepted 21 June 2009

PDB Reference: DmsD, 3cw0, r3cw0sf.

The translocation of folded proteins *via* the twin-arginine translocation (Tat) pathway is regulated to prevent the futile export of inactive substrate. DmsD is part of a class of cytoplasmic chaperones that play a role in preventing certain redox proteins from premature transport. DmsD from *Escherichia coli* has been crystallized in space group $P4_12_12$, with unit-cell parameters $a = b = 97.45$, $c = 210.04$ Å, in the presence of a small peptide. The structure has been solved by molecular replacement to a resolution of 2.4 Å and refined to an R factor of 19.4%. There are four molecules in the asymmetric unit that may mimic a higher order structure *in vivo*. There appears to be density for the peptide in a predicted binding pocket, which lends support to its role as the signal-recognition surface for this class of proteins.

1. Introduction

The twin-arginine translocation (Tat) pathway of bacteria is responsible for translocation across the cytoplasmic membrane of folded proteins. Typical Tat substrates are complexes and proteins that contain cofactors that are assembled in the cytoplasm. The minimal Tat system is composed of the integral membrane proteins TatA, TatB and TatC. The current model for translocation is that TatB and TatC form a receptor complex and bind the Tat signal motif, subsequently targeting the preprotein for translocation across the membrane through a TatA complex (reviewed in Berks *et al.*, 2000; Müller & Klösgen, 2005; Lee *et al.*, 2006). During the process, it is essential to prevent the futile export of immature protein.

The dimethylsulfoxide reductase complex (DmsABC) is targeted to the bacterial periplasm *via* the Tat pathway and is involved in anaerobic respiration using dimethylsulfoxide (DMSO) as an electron acceptor. It is a member of a group of Tat substrates that have a targeting chaperone, which is usually found in their operon. These chaperones, termed redox-enzyme maturation proteins (REMPs; Turner *et al.*, 2004), bind to the Tat signal sequence and presumably regulate folding, substrate addition and complex assembly prior to targeting to the translocation machinery (Berks *et al.*, 2000; Sargent, 2007).

DmsD is the REMP for DmsABC. It is a cytoplasmic protein and its size ranges between 20 and 30 kDa across species. DmsD null mutants fail to grow under anaerobic conditions, implying an essential role in anaerobic respiration (Oresnik *et al.*, 2001). The DmsA preprotein contains the Tat motif (S/TRR*x*LVK) in its N-terminal signal sequence and this region has been shown to bind directly to DmsD (Oresnik *et al.*, 2001). The remainder of DmsA is a large globular domain with a catalytic molybdopterin (MoPt) cofactor. DmsA is synthesized in the cytoplasm and loaded with cofactor and the accessory protein DmsB prior to translocation.

The best studied REMP, TorD, for the trimethylamine *N*-oxide (TMAO) reductase TorA, performs a parallel function to the DmsA/DmsD system (Ilbert *et al.*, 2004), and TorD and DmsD fall into a class of REMPs known by the first characterized structure TorD. The REMP motifs (Y/F/W)*xx*LF and E(P*x* or *x*P)D(H/Y) are conserved among all members of this family and mutations in these regions

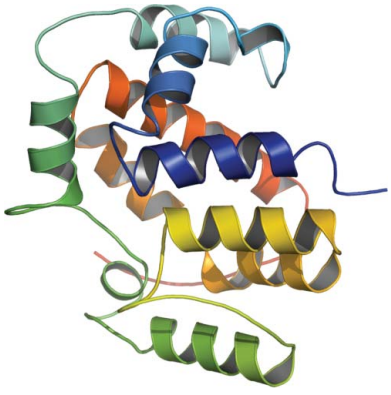

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

Data-collection statistics	
Space group	$P4_12_12$
Wavelength used (Å)	0.98
Unit-cell parameters (Å)	$a = b = 97.45$, $c = 210.04$
Resolution range (Å)	30.0–2.4 (2.49–2.40)
Total no. of observations	75194 (7411)
Total no. of unique observations	40654
Redundancy	3.8 (3.5)
Mean $I/\sigma(I)$	10.6 (2.2)
Completeness (%)	99.2 (97.3)
$R_{\rm merge}$ (%)	13.3 (58.8)
Wilson B factor (Å^2)	33.9
Refinement statistics	
Asymmetric unit content	4 molecules
R factor (%)	19.4
$R_{\rm free}$ factor (%)	24.0
No. of reflections in working set	40574
No. of reflections in test set	2028
Protein atoms	6588
Water atoms	509
Mean temperature factor (Å^2)	13.8
Matthews coefficient (Å^3 Da^{-1})	2.8
R.m.s.d. bond lengths (Å)	0.014
R.m.s.d. bond angles (°)	1.41
Ramachandran plot, residues in	
Most favored region (%)	97.3
Additionally allowed region (%)	2.7

lower the binding affinity to substrates (Chan *et al.*, 2008). Although physical interaction data between REMPs and partners is limited, several studies have looked directly at the problem. DmsD binds to the twin-arginine signal peptide from DmsA *via* a hydrophobic interaction with micromolar affinity in an equimolar ratio (Winstone *et al.*, 2006). DmsD can form a complex with the N-terminus of DmsA independently of the globular domain of the protein (Oresnik *et al.*, 2001) and has been shown *in vivo* to bind to the TatBC complex under anaerobic conditions, implying a direct role in targeting (Papish *et al.*, 2003). *In vitro*, TorD is able to form a complex with the unfolded mature portions of TorA (Pommier *et al.*, 1998). DmsD/TorD REMPs are not essential for substrate translocation; instead, they appear to be important for transport of the correctly folded complex (Ray *et al.*, 2003).

Many mechanistic questions about REMPs remain. The coordination between cofactor loading and the subsequent release of the protein after folding is unknown. There is no direct evidence for interaction between the Tat machinery, the Tat substrate and the corresponding REMP protein. The Tat translocase itself has been shown to accept or reject protein for transport through a quality-control mechanism (Matos *et al.*, 2008), so the exact role of REMPs is unclear. This paper reports an additional crystal structure of DmsD from *Escherichia coli* (*Ec*DmsD) and provides further evidence for functional interactions.

2. Methods

2.1. Cloning, overexpression and protein purification

The gene encoding *Ec*DmsD was amplified from *E. coli* genomic DNA using the primers 5′-GGGCGG**GTCGAC**ATGACCCATTT-TTCACAG-3′ and 5′-GGGGCGG**AGATCT**CTATCGAAACAGC-GG-3′, which incorporate the restriction-endonuclease sites *Sal*I and *Bgl*II (highlighted in bold), respectively. Digested PCR product was ligated into a modified T7 promotor-based pET33b expression vector (Novagen) that contained an N-terminal 6×His tag followed by a thrombin cleavage site and a modified multiple cloning site and was then transformed into *E. coli* BL21 (DE3) host strain. The plated cells were directly inoculated into 1 l LB medium containing 35 µg ml^{-1} kanamycin. The culture was grown at 310 K to an OD_{600} of 0.6, followed by induction with 300 µ*M* isopropyl *β*-D-1-thiogalacto-pyranoside (IPTG from Anatrace) for 3 h. The cells were harvested by centrifugation and the cell pellet was resuspended in 100 ml buffer

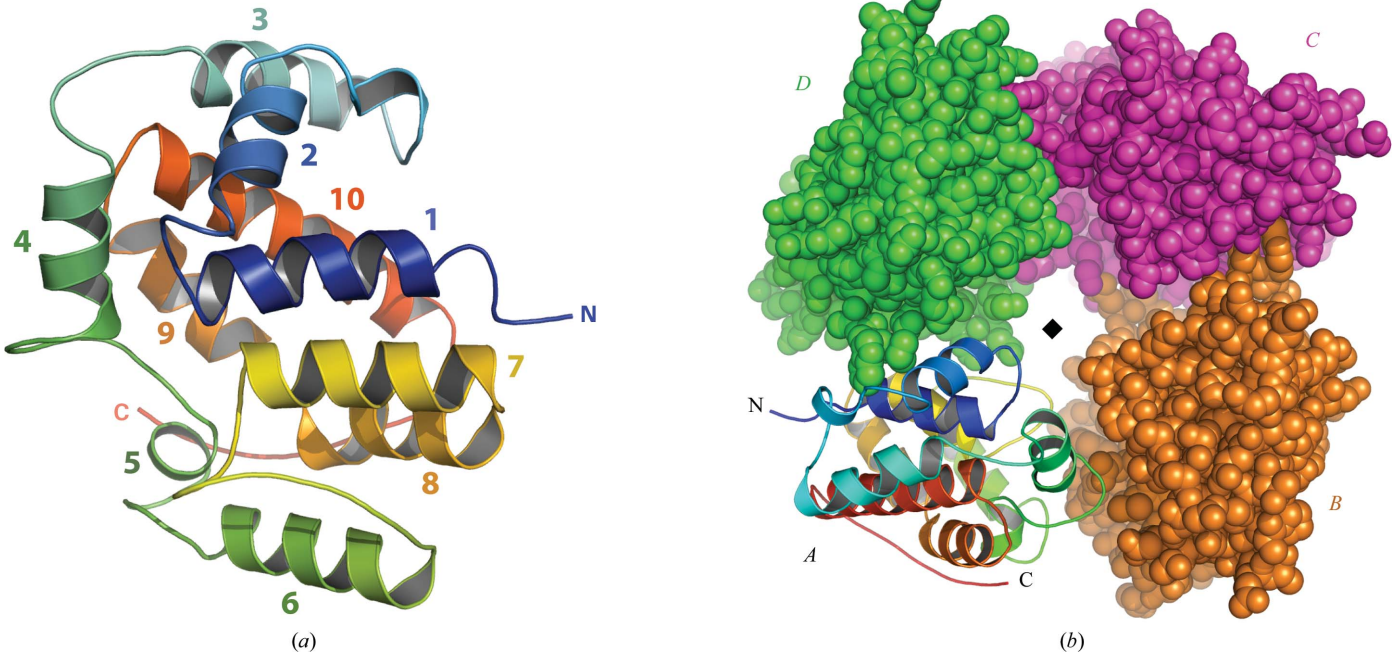

Figure 1
(*a*) Ribbon diagram of a monomer of *Ec*DmsD color-ramped blue to red from the N-terminus to the C-terminus. (*b*) The tetramer in the asymmetric unit of *Ec*DmsD, with each monomer colored differently. Monomer *A* ribbons are colored in a similar way to those in (*a*) and monomers *B*–*D* are colored by chain and shown as spheres. Except where indicated otherwise, structural figures were produced using *PyMOL* (DeLano, 2002).

A (100 m*M* NaCl, 10 m*M* β-mercaptoethanol, 30 m*M* imidazole and 50 m*M* Tris–HCl pH 7.5) and then lysed by passing it three times though a microfluidizer (Microfluidics). Cell debris was removed by centrifugation at 18 000*g* for 20 min at 277 K. The supernatant was loaded onto a 5 ml Ni–NTA affinity resin (Qiagen) pre-equilibrated and washed with 30 ml buffer *A* and then eluted with buffer *A* containing 300 m*M* imidazole. The protein was diluted into PBS buffer and the 6×His tag was cleaved with 20 units of thrombin incubated overnight at 295 K, resulting in an additional four residues at the native N-terminus (GSVD). The remaining uncleaved protein was removed by passage over a second Ni–NTA affinity resin. The flowthrough was concentrated and run on a Superdex 200 10/300 GL column (GE Healthcare) in buffer *A* without imidazole. The purified protein was concentrated to 20 mg ml^{-1} using a centrifugal concentration device (Amicon Ultra, Millipore) before storage at 193 K. The 98.7% pure peptide (SRRDFLK) was ordered from Genescript, USA.

2.2. Multi-angle light-scattering analysis

Purified *Ec*DmsD (20 mg ml^{-1}) was loaded onto a Shodex Protein KW-803 size-exclusion column equilibrated with 50 m*M* Tris–HCl, 150 m*M* NaCl, 10 m*M* β-mercaptoethanol and connected in-line with a Dawn 18-angle light-scattering detector coupled to an Optilab interferometric refractometer and a WyattQELS Quasi-Elastic Light-Scattering instrument (Wyatt Technologies). Data analysis was performed with the *ASTRA* v5.3.4.14 software (Wyatt Technologies) and molecular weights were calculated using the Zimm fit method.

2.3. Crystallization, data collection and processing

The purified DmsD was incubated for 2 h with a 1:1 molar ratio of peptide to protein before crystallization. A Mosquito pipetting robot was used to set up 200 nl drops, which were equilibrated against 50 µl reservoir solution using sitting-drop vapor diffusion in a 96-well plate format using standard crystal screening kits (Hampton Research and Qiagen). Large single crystals appeared in drops made up from 1 µl 1.5 *M* ammonium tartrate, 0.1 *M* Tris pH 7.5 and 1 µl protein solution and equilibrated against 500 µl reservoir solution using a 24-well sitting-drop plate (Hampton Research). A single crystal was briefly transferred to mother liquor supplemented with 30%(*v*/*v*) glycerol, mounted in a cryoloop and flash-cooled in liquid nitrogen.

Diffraction data were collected on beamline BL9-2 at the Stanford Synchrotron Radiation Laboratory (SSRL). The crystal was maintained at a constant 100 K throughout the experiment. A total of 180 images were collected with an oscillation range of 0.5° per image, an exposure time of 8.1 s and a crystal-to-detector distance of 275 mm. Images were collected on a MAR Mosaic 325 CCD detector and indexed using *HKL*-2000 (Otwinowski & Minor, 1997). The crystals belonged to space group $P4_12_12$, with unit-cell parameters $a = b = 97.45$, $c = 210.04$ Å and a solvent content of 56% (Matthews, 1968) with four molecules in the asymmetric unit. Scaling of the data was performed using *SCALA* (Evans, 2006). Data statistics are presented in Table 1.

2.4. Structure determination and refinement

The structure of *Ec*DmsD was solved by molecular replacement using the program *Phaser* (McCoy, 2007) with a single chain of the DmsD structure from *Salmonella typhimurium* (68% sequence identity; PDB code 1s9u) as a search model. The initial model was built using a combination of automatic building and refinement with *ARP*/*wARP* (Langer *et al.*, 2008). Rigid-body and TLS refinement were performed using *REFMAC* (Vagin *et al.*, 2004). Structure refinement continued with successive rounds of model building in *Coot* (Emsley & Cowtan, 2004) followed by *REFMAC* refinement. At the end of the analysis, waters were added to the model using as criteria a 3σ peak height in the difference maps and hydrogen-bonding distances to protein atoms of 2.0–3.5 Å. The structure was validated using *MolProbity* (Lovell *et al.*, 2003). The final coordinates and structure factors have been deposited in the Protein Data Bank (http://www.rcsb.org/pdb) with PDB code 3cw0.

3. Results

The preparation of recombinant protein resulted in a reliable supply of pure protein for crystallization, with a final yield of 15–18 mg pure protein per litre of cell culture. The purity of the protein was confirmed using SDS–PAGE. Similar to previous results (Sarfo *et al.*, 2004), native gels showed many bands, possibly indicating multiple oligomeric forms of the protein; however, gel-filtration data, along with multi-angle light scattering, only showed the presence of a monomer in solution with or without peptide (data not shown).

There are four molecules of *Ec*DmsD in the asymmetric unit and we were able to build a continuous chain trace for residues 2–204 that only lacked the residues left from the thrombin cleavage site (Fig. 1*a*). The most likely biological unit is a monomer, as evidenced by gel-filtration chromatography. A total of 6588 protein atoms and 509 water molecules were included in the structural model. The structure was successfully refined to an *R* factor of 19.6% and a free *R* factor of 23.9% at 2.4 Å resolution. Further refinement statistics are presented in Table 1.

DmsD has an all α-helical architecture composed of ten helices with dimensions of approximately 52 × 43 × 33 Å (Fig. 1*a*). The root-mean-square deviation (r.m.s.d.) between the C^{α} atoms in this structure and the search model 1s9u is 0.9 Å for all residues. The average solvent-accessible surface area per monomer is 9979.6 Å^2 and the four monomers in the asymmetric unit superimpose with an average r.m.s.d. of 0.29 Å, burying an average of 624 Å^2 of surface at each interface (values calculated using *PISA*; Krissinel & Henrick, 2007; Fig. 1*b*).

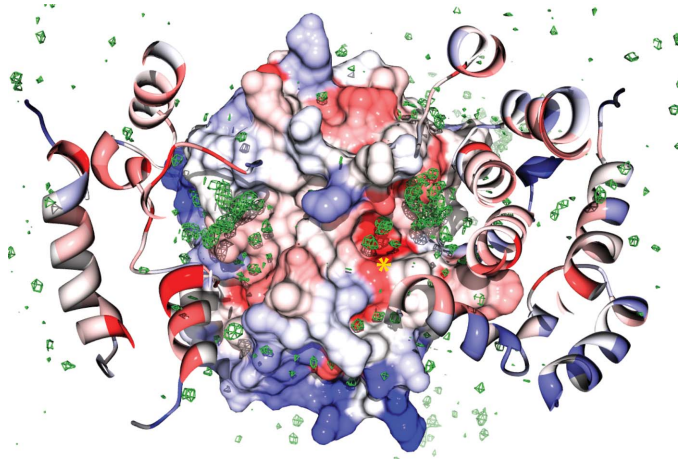

Figure 2
A view from inside the tetramer looking at monomer *A*. Monomer *A* is shown as an accessible surface and monomers *B* and *D* are shown as ribbons. Subunits are colored based on surface conservation calculated by the *ConSurf* server (http://consurf.tau.ac.il/) with ramping from unconserved (blue) to 25% conserved (grey) and then to 100% conserved (red). The pocket formed by the REMP motif is shown by a yellow asterisk. Density not accounted for by protein is shown in green contoured at 2σ. This figure was produced using *Chimera* (Pettersen *et al.*, 2004).

The crystallization conditions included a small peptide, SRRD-FLK, in an attempt to generate a complex between a minimal DmsA Tat signal and DmsD. Addition of the peptide led to a marked improvement in crystal size and quality (data not shown); however, there was no density in the final maps that could conclusively be built as peptide. We did observe several blobs of density that could not be accounted for as waters or components of the mother liquor that may be consistent with partially ordered peptide (Fig. 2).

4. Discussion

4.1. Structural comparison with other REMPs

Several REMP structures of the DmsD/TorD class have previously been solved, including those of TorD from *Shewanella massilia* (*Sm*TorD; PDB code 1n1c), DmsD from *Salmonella typhimurium* (*St*DmsD; PDB code 1s9u) and a putative REMP from *Archaeglobus fulgidus* (*Af*REMP; PDB code 2o9x) (Qiu *et al.*, 2008; Tranier *et al.*, 2003; Kirillova *et al.*, 2007). The overall tertiary structure of each of these REMPs is similar. When superimposing relevant regions of monomers on *Ec*DmsD, the r.m.s.d. of each is 3.0, 0.9 and 2.9 Å, respectively (Fig. 3).

Many of the salient differences in the various REMP structures have been mentioned previously, so we will only highlight those that pertain to *Ec*DmsD (Qiu *et al.*, 2008; Tranier *et al.*, 2003; Kirillova *et al.*, 2007). In *St*DmsD the loop between helix 6 and helix 7 (117–122) is missing owing to disorder and we were able to clearly build this in *Ec*DmsD (Fig. 3*a*). The structure of *Sm*TorD is a dimer containing monomers formed by swapped domains, with the last α-helix (residues 202–210) extending out of the globular domain (Fig. 3*c*). In both DmsD structures the C-terminus forms a long extended region that packs tightly with the globular domain (Figs. 3*a* and 3*b*). The orientation of helix 7 of TorD, located before the hinge region (121–135), is considerably different in both DmsD structures (Qiu *et al.*, 2008). The N-termini of both *Sm*TorD and *St*DmsD are solvent-exposed, while in *Ec*DmsD this hydrophobic stretch packs against the rest of the protein in a hydrophobic pocket.

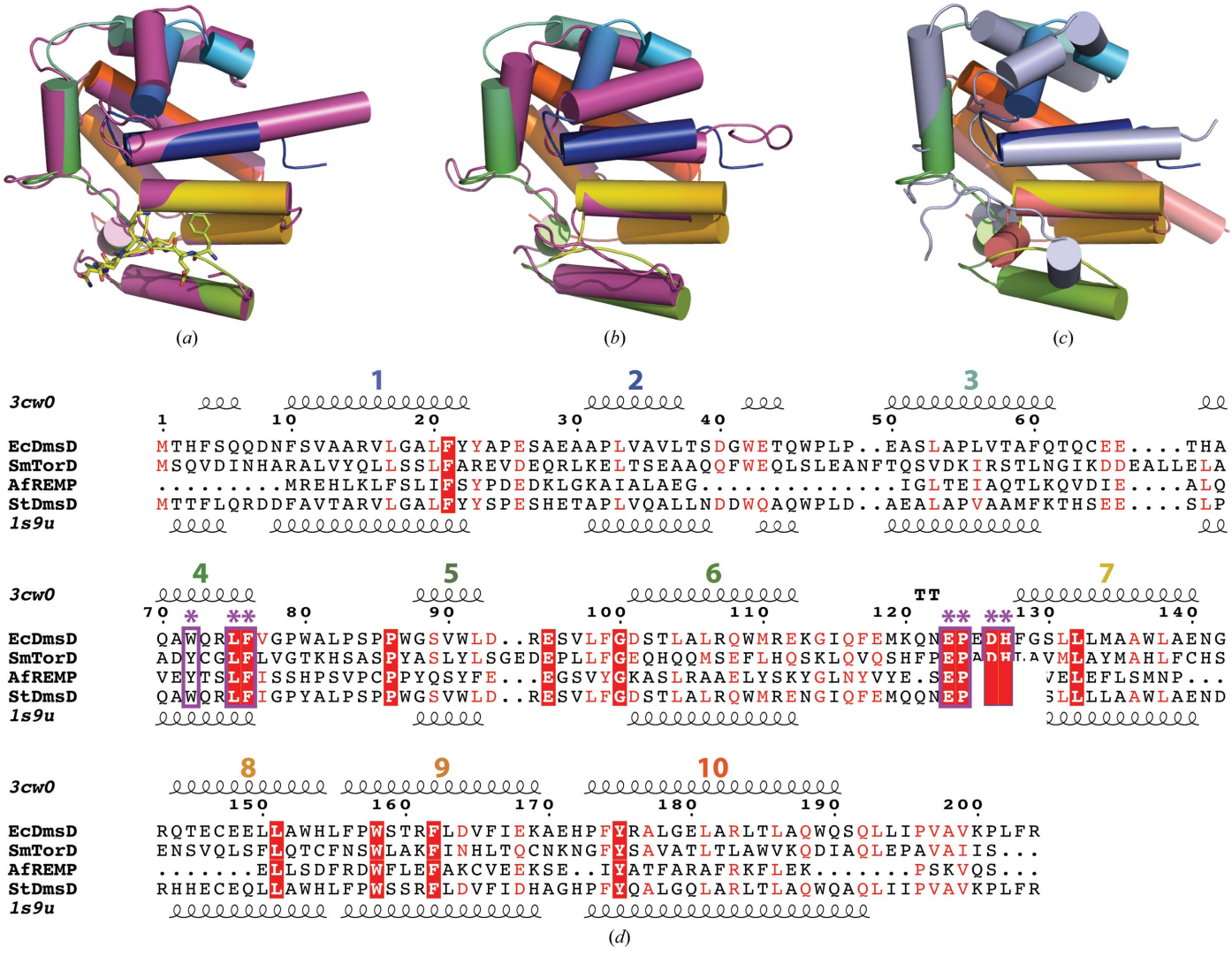

Figure 3
(*a*) Superposition of *Ec*DmsD (rainbow) with other REMPs (magenta): (*a*) *St*DmsD (magenta), with the missing loop visible in *Ec*DmsD shown with side-chain sticks, (*b*) *Af*REMP (magenta) and (*c*) *Sm*TorD, with one monomer formed by swapped domains from the crystal dimer colored with the N-terminus in light blue and the C-terminus in light red. All are oriented relative to Fig. 1(*a*). (*d*) A sequence alignment between the REMP structures discussed. Secondary-structure elements are displayed for *Ec*DmsD (top) and for *St*DmsD (bottom) with helices numbered as in Fig. 1(*a*). Semi-conserved residues are shown in red and completely conserved residues are shown in white with red boxes. The REMP motif is highlighted by magenta asterisks and boxes.

The tetrameric packing in the asymmetric unit is a pseudo-fourfold presenting a repeating interface between monomers (Fig. 1*b*). The interface contains three hydrogen-bond interactions; however, a hole formed at this interface is solvent-filled and contained density which we were unable to fit (Fig. 2). Although the calculated interface implies that the packing would not constitute a tight interface, many of the conserved DmsD/TorD residues cluster at this interface. The packing in the *St*DmsD and *Af*REMP crystals clearly suggests that they exist as monomers.

In this REMP structural family the conserved sequence motifs contain two residues that have been shown to be crucial for function: Asp126 and His127 in *Ec*DmsD (Fig. 3*d*; Chan *et al.*, 2008). The motifs occur in helix 4 and a long loop (*Ec*DmsD 112–129) forming a mainly hydrophilic groove. These regions in *St*DmsD and *Sm*TorD have been predicted to be involved in signal peptide binding (Qiu *et al.*, 2008). In *Af*REMP, this groove opens to form a more open funnel-shaped cavity connected to a slightly hydrophobic surface that is predicted to be a region that recognizes partially folded substrate (Kirillova *et al.*, 2007). This groove in *Ec*DmsD contains the extra density that is unaccounted for in our crystal structure (Fig. 2). We could not fit the peptide we cocrystallized owing to poor density and it is unclear whether this is representative of low occupancy. Further investigations are necessary to confirm the peptide-binding region.

During the writing of this manuscript the coordinates for another structure of *Ec*DmsD in a different space group ($P3_121$) were released (PDB code 3efp; Stevens *et al.*, 2009). Despite the different space groups and crystallization conditions, the two *E. coli* monomers are virtually identical, with an r.m.s.d. of 0.5 Å. Interestingly, the 3efp crystal form contains a dimer in the asymmetric unit; however, in the crystal packing the same tetramer is formed as seen in our crystal form. In the subsequent manuscript, the authors also conclude that the *Ec*DmsD structure is representative of the TorD REMP family and the biochemical evidence points toward a specific binding pocket for signal peptide (Stevens *et al.*, 2009). The authors note that several small molecules from their crystallization conditions occupy the putative binding pocket, similar to our unfitted density. We had considered this possibility for our crystals but were unable to fit any of the compounds from our crystallization conditions into our density (Fig. 2).

5. Conclusion

Despite the many crystal structures that are now available, it is still difficult to interpret the functional interactions between the REMP proteins and their substrates. Clearly, the consensus points toward a binding groove for the Tat recognition motif and this report lends support to this model. Curiously, although DmsD/TorD REMPs appear to function as monomers, they behave anomalously in solution and in native gels, with the formation of dimers and multimers. The crystallographic packing of *Ec*DmsD perhaps sheds light on this and may indicate a possible functional interface.

We thank A. Müller and D. Rees for discussion and comments on the manuscript. We thank Gordon and Betty Moore for support of the Molecular Observatory at Caltech. All data collection was performed on beamline 9-2 at SSRL. Operations at SSRL are supported by the US DOE and NIH. WMC is supported by awards from the Searle Scholar program and the Burroughs–Wellcome fund career award for biological sciences.

References

Berks, B. C., Sargent, F. & Palmer, T. (2000). *Mol. Microbiol.* **35**, 260–274.
Chan, C. S., Winstone, T. M., Chang, L., Stevens, C. M., Workentine, M. L., Li, H., Wei, Y., Ondrechen, M. J., Paetzel, M. & Turner, R. J. (2008). *Biochemistry*, **47**, 2749–2759.
Delano, W. L. (2002). *The PyMOL Molecular Graphics System.* http://www.pymol.org.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Ilbert, M., Méjean, V. & Iobbi-Nivol, C. (2004). *Microbiology*, **150**, 935–943.
Kirillova, O., Chruszcz, M., Shumilin, I. A., Skarina, T., Gorodichtchenskaia, E., Cymborowski, M., Savchenko, A., Edwards, A. & Minor, W. (2007). *Acta Cryst.* D**63**, 348–354.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nature Protoc.* **3**, 1171–1179.
Lee, P. A., Tullman-Ercek, D. & Georgiou, G. (2006). *Annu. Rev. Microbiol.* **60**, 373–395.
Lovell, S. C., Davis, I. W., Arendall, W. B. III, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003). *Proteins*, **50**, 437–450.
Matos, C. F., Robinson, C. & Di Cola, A. (2008). *EMBO J.* **27**, 2055–2063.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
McCoy, A. J. (2007). *Acta Cryst.* D**63**, 32–41.
Müller, M. & Klösgen, R. B. (2005). *Mol. Membr. Biol.* **22**, 113–121.
Oresnik, I. J., Ladner, C. L. & Turner, R. J. (2001). *Mol. Microbiol.* **40**, 323–331.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Papish, A. L., Ladner, C. L. & Turner, R. J. (2003). *J. Biol. Chem.* **278**, 32501–32506.
Pettersen, E., Goddard, T., Huang, C., Couch, G., Greenblatt, D., Meng, E. & Ferrin, T. (2004). *J. Comput. Chem.* **25**, 1605–1612.
Pommier, J., Méjean, V., Giordano, G. & Iobbi-Nivol, C. (1998). *J. Biol. Chem.* **273**, 16615–16620.
Qiu, Y., Zhang, R., Binkowski, T. A., Tereshko, V., Joachimiak, A. & Kossiakoff, A. (2008). *Proteins*, **71**, 525–533.
Ray, N., Oates, J., Turner, R. J. & Robinson, C. (2003). *FEBS Lett.* **534**, 156–160.
Sarfo, K. J., Winstone, T. L., Papish, A. L., Howell, J. M., Kadir, H., Vogel, H. J. & Turner, R. J. (2004). *Biochem. Biophys. Res. Commun.* **315**, 397–403.
Sargent, F. (2007). *Biochem. Soc. Trans.* **35**, 835–847.
Stevens, C. M., Winstone, T. M., Turner, R. J. & Paetzel, M. (2009). *J. Mol. Biol.* **389**, 124–133.
Tranier, S., Iobbi-Nivol, C., Birck, C., Ilbert, M., Mortier-Barrière, I., Méjean, V. & Samama, J.-P. (2003). *Structure*, **11**, 165–174.
Turner, R. J., Papish, A. L. & Sargent, F. (2004). *Can. J. Microbiol.* **50**, 225–238.
Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F. & Murshudov, G. N. (2004). *Acta Cryst.* D**60**, 2184–2195.
Winstone, T. L., Workentine, M. L., Sarfo, K. J., Binding, A. J., Haslam, B. D. & Turner, R. J. (2006). *Arch. Biochem. Biophys.* **455**, 89–97.