doi:10.1016/j.jmb.2009.03.069 J. Mol. Biol. (2009) 389, 124–133

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Structural Analysis of a Monomeric Form of the Twin-Arginine Leader Peptide Binding Chaperone *Escherichia coli* DmsD

Charles M. Stevens[1], Tara M. L. Winstone[2], Raymond J. Turner[2] and Mark Paetzel[1]*

[1]*Department of Molecular Biology and Biochemistry, Simon Fraser University, South Science Building, 8888 University Drive, Burnaby, British Columbia, Canada V5A 1S6*

[2]*Department of Biological Sciences, University of Calgary, 2500 University Drive NW, Calgary, Alberta, Canada T2N 1N4*

Received 3 February 2009; received in revised form 27 March 2009; accepted 31 March 2009
Available online 8 April 2009

Edited by R. Huber

The redox enzyme maturation proteins play an essential role in the proofreading and membrane targeting of protein substrates to the twin-arginine translocase. Functionally, the most thoroughly characterized redox enzyme maturation protein to date is *Escherichia coli* DmsD (EcDmsD). Herein, we present the X-ray crystal structure of the monomeric form of the EcDmsD refined to 2.0 Å resolution, with clear electron density present for each of its 204 amino acid residues. The structural data presented here complement the biochemical data previously generated regarding the function of these twin-arginine translocase leader peptide binding chaperone proteins. Docking and molecular dynamics simulation experiments were used to provide a proposed model for how this chaperone is able to recognize the leader peptide of its substrate DmsA. The interactions observed in the model are in agreement with previous biochemical data and suggest intimate interactions between the conserved twin-arginine motif of the leader peptide of *E. coli* DmsA and the most conserved regions on the surface of EcDmsD.

Keywords: DmsD; twin-arginine translocase (Tat); redox enzyme maturation protein (REMP); chaperone; signal peptide

Introduction

The process by which proteins are translocated across the plasma membrane is of fundamental importance in biology. One system employed by bacteria to accomplish this task is the twin-arginine translocase (Tat) apparatus.[1] This translocon is named for the characteristic arginine motif (S/T–R–R–x–F–L–K) present near the amino-terminus of Tat substrates and is capable of translocating fully folded, multiprotein complexes with bound cofactors in place.[2] The translocation machinery is composed of three proteins: TatA, which is thought to polymerize to form the pore; TatB, a regulatory component; and TatC, which binds the substrate proteins.[3,4]

The folding and assembly of Tat substrates with their cofactors must be complete prior to their targeting and translocation. A family of chaperone proteins has been identified in the targeting of Tat substrates to the membrane. These redox enzyme maturation proteins (REMPs)[5] bind to the leader sequence of the Tat preprotein and maintain the substrate in a state competent for cofactor insertion, ensure correct folding and assembly, and, finally, target the substrate to the membrane.

Mutagenesis studies have suggested, using the model REMP–substrate interaction between *Escherichia coli* DmsD (EcDmsD) and the *E. coli* DmsA (EcDmsA) leader peptide, that a number of conserved REMP residues are responsible for the interaction with the leader peptide.[6] The REMPs and their corresponding Tat leader peptides appear to form tight associations.[7,8]

A number of structures for the REMP proteins are available in the Protein Data Bank (PDB). These

*Corresponding author. E-mail address: mpaetzel@sfu.ca.

Abbreviations used: EcDmsD, *Escherichia coli* DmsD; Tat, twin-arginine translocase; REMP, redox enzyme maturation protein; EcDmsA, *Escherichia coli* DmsA; SmTorD, *Shewanella massilia* TorD; StDmsD, *Salmonella typhimurium* DmsD; TBS, Tris-buffered saline.

include *Shewanella massilia* TorD (SmTorD; PDB ID 1N1C),[9] *Salmonella typhimurium* DmsD (StDmsD; PDB ID 1S9U),[10] *Archaeoglobus fulgidus* AF0173 (PDB ID 2O9X),[11] *A. fulgidus* AF0160 (PDB ID 2IDG), and, finally, the NMR structure of the *E. coli* NapD protein (PDB ID 2JSX).[12] The crystal structure of SmTorD describes a domain-swapped dimer in which a single lobe is made up of residues 1–142 from chain A and residues 130–211 from chain B. SmTorD has been shown to be functional both as a monomer and a dimer, which is also the case with EcDmsD.[13,14] The primary amino acid sequence of the StDmsD protein is 78% identical to the EcDmsD protein; however, the crystal structure of StDmsD does not contain electron density for residues 116–122, a section of a conserved loop that contributes to leader peptide binding based on the biochemical analysis performed by Chan *et al*.[6] NapD is structurally distinct from the majority of REMPs as it bears a ferredoxin-like fold rather than the familiar TorD-like fold characteristic of the other REMPs that have been solved to date.[12] The putative *A. fulgidus* REMPs AF0173 (NarJ homolog) and AF0160 have little sequence similarity to DmsDs and have been solved with comparatively low resolution diffraction data at 3.4 and 2.7 Å, respectively.

In this study, we report the 2.0 Å resolution crystal structure of EcDmsD, a chaperone that recognizes and binds to the twin-arginine leader peptide of its substrate EcDmsA. The structure is unique among the REMP structures solved to date in that it provides clear electron density for all 204 residues of the protein. It is also the first X-ray crystal structure of a Tat chaperone from the model organism *E. coli*, with EcDmsD being one of the most thoroughly biochemically characterized Tat chaperones to date. The crystal structure, along with docking experiments, molecular dynamics simulation experiments, and use of previous biochemical data,[6] was used to generate a proposed model for how this chaperone is able to recognize and bind to the leader peptide of its substrate.

Results and Discussion

Structure solution

The EcDmsD crystal structure was refined to 2.0 Å resolution. Clear electron density was observed for all 204 amino acid residues for both molecules in the asymmetric unit. The final refined model has an *R*-factor of 17.8% with an R_{free} of 21.2%. The average *B*-factor of the structure is 25.8 (Table 1). The two molecules in the asymmetric unit superimposed with an RMSD of 1.11 Å for all atoms, or with that of 0.68 Å when the superimposition was restricted to the polypeptide backbone.

Table 1. Crystallographic statistics

Crystal parameters	
Space group	$P3_121$ (152)
a, *b*, *c* (Å)	128.0, 128.0, 78.7
Data collection statistics	
Wavelength (Å)	1.5418
Resolution (Å)	30.7–2.0 (2.08–2.01)
Total reflections	278,055
Unique reflections	48,639
R_{merge}[a]	0.089 (0.362)
Mean $(I)/\sigma(I)$	11.5 (4.5)
Completeness (%)	97.7 (95.6)
Redundancy	5.7 (5.6)
Refinement statistics	
Protein molecules (chains)	2
Residues	412
Water molecules	411
Ligands	11
Total no. of atoms	3860
R_{cryst}/R_{free} (%)[b,c]	17.8/21.2
Average *B*-factor (Å^2)	
Proteins	25.8
Water molecules	36.9
Ligands	57.0
RMSD on angles (°)	1.1
RMSD on bonds (Å)	0.009
Residues in most favored regions (%)[d]	92
Residues in additionally allowed regions (%)[d]	8

The data in brackets are the values for the highest-resolution shell.
[a] $R_{merge} = (\sum_h \sum_i | I_i - \langle I \rangle |)/(\sum_h \sum_i I_i)$.
[b] $R_{cryst} = (\sum | |F_{obs}| - |F_{calc}| |)/(\sum |F_{obs}|)$.
[c] R_{free} is calculated as R_{cryst}, but it uses a test set of 5% of the total reflections to generate F_{obs}.
[d] Ramachandran plot.

Protein fold

The EcDmsD structure is mainly α-helical (Fig. 1). There are 11 α-helices and 1 3_{10}-helix arranged in a single globular domain with approximate dimensions of 49 Å × 37 Å × 34 Å. The EcDmsD monomer has a surface area of approximately 8894 Å^2 and a volume of approximately 26,120 Å^3. The EcDmsD fold can be classified within the TorD-like family of proteins according to the SCOP database,[15] but there are no domain swapping interactions in this EcDmsD structure as were observed in the SmTorD crystal structure.[9]

Structural comparison with other REMPs

Alignment of the EcDmsD amino acid sequence with sequences for DmsD molecules from other

Fig. 1. The protein fold of EcDmsD. (a) Cartoon rendering colored spectrally from the N-terminus (blue) to the C-terminus (red) with the amino- and carboxy-termini labeled. The 11 α-helices are numerically labeled. (b) Schematic diagram of the EcDmsD topology and cartoon representation of EcDmsD. Helices represented as cylinders are labeled, as are the amino and carboxy-termini. Shaded and white helices correspond to the domain-swapped dimer conformation observed in SmTorD. The conserved loops that make up the putative leader peptide binding site are shown in blue (residues 77–88), red (residues 93–100), and green (residues 113–128). (c) A stereo view of EcDmsD. Every 20th residue is labeled with a sphere.

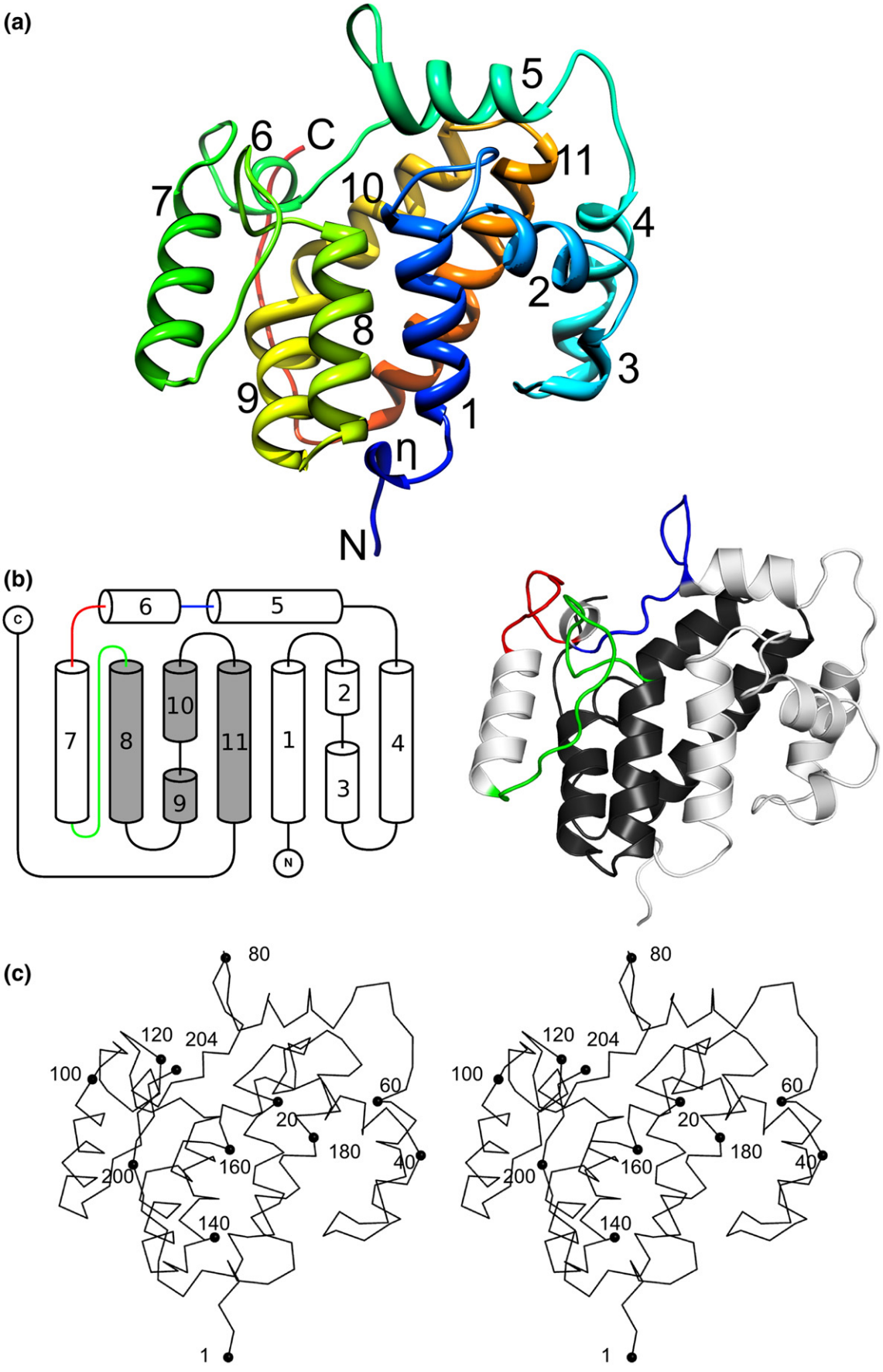

Fig. 1 *(legend on previous page)*

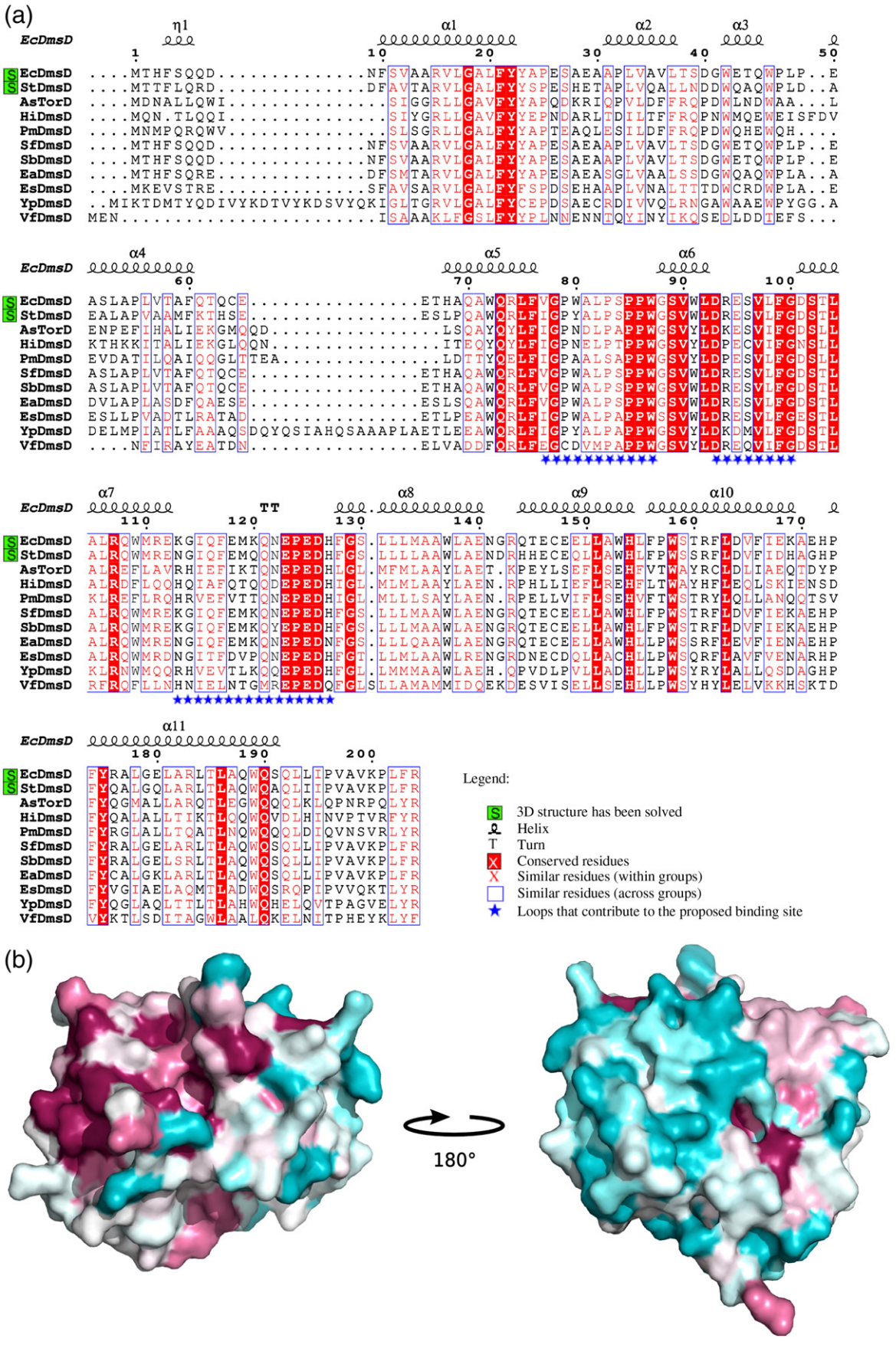

Fig. 2 *(legend on next page)*

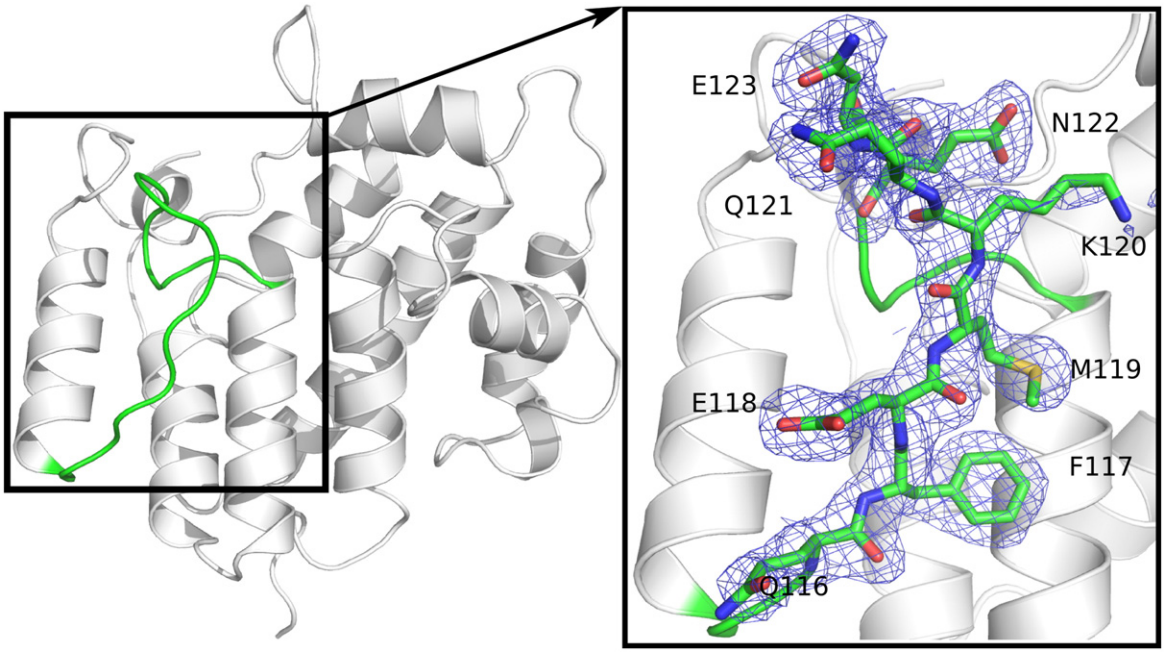

Fig. 3. All the loops that make up the putative leader peptide binding site of EcDmsD are visible in the electron density. A sample of the $2F_{obs}-F_{calc}$ electron density map contoured at 1.0σ is shown here for the loop between residues 116 and 121.

species shows a moderate level of sequence identity (15.7%), yet the alignment also reveals two highly conserved regions that map to three loops (residues 77–88, 93–100, and 113–128) on the surface of the protein (Figs. 1b and 2a). When the conservation is mapped onto the molecular surface of EcDmsD, it is clear that the region of high conservation corresponds with many of the residues that were previously determined by mutagenesis to be important for Tat leader peptide binding[6] (Fig. 2b). This conserved region on the surface of EcDmsD also corresponds to the location of the most significant depression on the EcDmsD molecular surface.

The superimposition of EcDmsD on StDmsD yields a C^{α} RMSD of 0.71 Å. The EcDmsD structure shows the location of a presumably flexible solvent-exposed loop that was unresolved in StDmsD (residues 116–122). Another notable difference between the two structures is the presence of an N-terminal 3_{10}-helix in the EcDmsD structure, while the corresponding residues in the StDmsD structure are part of the neighboring α-helix (helix 1).

Electron density was clearly observed for the region of EcDmsD between residues 116 and 123 (Fig. 3). This region was not resolved in the structure of StDmsD.[10] These residues lie on one of the three conserved surface loops (Fig. 1b), part of which forms the putative leader peptide binding site.[6] In the SmTorD structure, a homologous loop is involved in bridging the two domains that are swapped to form the dimer.[9] The EcDmsD structure presented here is the first structure of a DmsD to have experimental electron density for the complete protein and therefore the complete refined model.

When comparing EcDmsD with the AF0173 protein from *A. fulgidus*, the two structures superimpose with a C^{α} RMSD of 2.4 Å despite having only 17% sequence identity and AF0173 being 45 amino acid residues shorter than EcDmsD. Interestingly, the structure of AF0173 was solved such that the AF0173 protein was bound to the TEV protease recognition sequence of a symmetry-related molecule. This is despite the absence of any sequence similarity between the TEV protease cleavage sequence, E–N–L–Y–F–Q–S, and

Fig. 2. (a) Sequence alignment of EcDmsD with homologous proteins. The secondary structure is shown above the alignment. The sequences were acquired from the SwissProt/TrEMBL database. The accession numbers for each sequence are as follows: EcDmsD, P69853; StDmsD, Q8ZPK0; *Actinobacillus succinogenes* TorD family protein, A6VPI1; *Haemophilus influenzae* DmsD, A5UD55; *Pasteurella multocida* DmsD, Q9CK76; *Shigella flexneri* DmsD, P69855; *Shigella boydii* DmsD, Q320U8; *Escherichia albertii* DmsD, B1EN93; *Enterobacter* sp. 638 DmsD, A4WA71; *Yersinia pseudotuberculosis* TorD family protein, B1JJB8; and *Vibrio fischeri* DmsD, Q5E1E3. Absolutely conserved residues are shown in white with red fill, similar residues within groups are shown in red, and similar residues across groups are surrounded by a blue box. Sequences for which three-dimensional coordinates are available are highlighted in green. (b) A view of DmsD conservation mapped onto the EcDmsD surface generated using the above alignment. Individual amino acid residues are colored according to the degree to which they are conserved: absolutely conserved residues are shown in maroon, while highly variable residues are shown in green.

the twin-arginine motif from the preAF0174 sequence, S–R–R–D–F–I–K.[11] In the AF0173 structure, the TEV protease cleavage sequence binds into a region of AF0173 that is on the opposite pole of the structure from the putative binding site described by Chan *et al.*[6]

When EcDmsD is compared with SmTorD, there are a number of significant differences that can be observed. The superimposition of EcDmsD with a single domain-swapped monomer of SmTorD, containing residues 1–129 from one chain and residues 130–211 from the complementary chain, yields an RMSD of 2.6 Å. The major differences arise primarily in the bridge point between the two domains (Fig. 1b) and at the C-termini. Notably, both SmTorD and EcDmsD have been reported to exist in monomeric and dimeric forms.[13,14]

The putative Tat leader peptide binding pocket on EcDmsD

Previous mutagenesis work by Chan *et al.*[6] identified a number of residues on the surface of EcDmsD that are important for EcDmsA leader peptide binding. Most of these residues map to a pocket on the surface of EcDmsD. This putative leader peptide binding site is composed of sections of three conserved loops (Fig. 2). The first loop is made up of residues 77–88 and is contained between helices α5 and α6. The next loop lies between helices α6 and α7 and encompasses residues 93–100. The third loop lies between helices α7 and α8 and is made up of residues 113–128 (Fig. 4). These loops form a curved trench along the surface of the protein approximately 17.1 Å in length (Arg204 NH1 to Leu82 $C^{\delta 1}$ to Trp72 CH2) and approximately 8.5 Å in width at the narrowest point (Val77 O′ to Glu123 C^{δ}). The pocket is predominantly hydrophobic with small regions of positive charge (Arg204 and Lys120) and a region of negative charge centered at residue Glu29 (Fig. 4). Interestingly, in our crystal structure of EcDmsD, we found strong electron density for five small molecules [three glycerol molecules and two tris(hydroxymethyl)aminomethane molecules] within the proposed leader peptide binding pocket of EcDmsD (Fig. 4a). Electron density for a polyethylene glycol molecule was found in a similar region in the structure of StDmsD.[10]

Docking and molecular dynamics simulation

The mutagenesis studies by Chan *et al.*[6] identified several residues that, if mutated, disrupt the

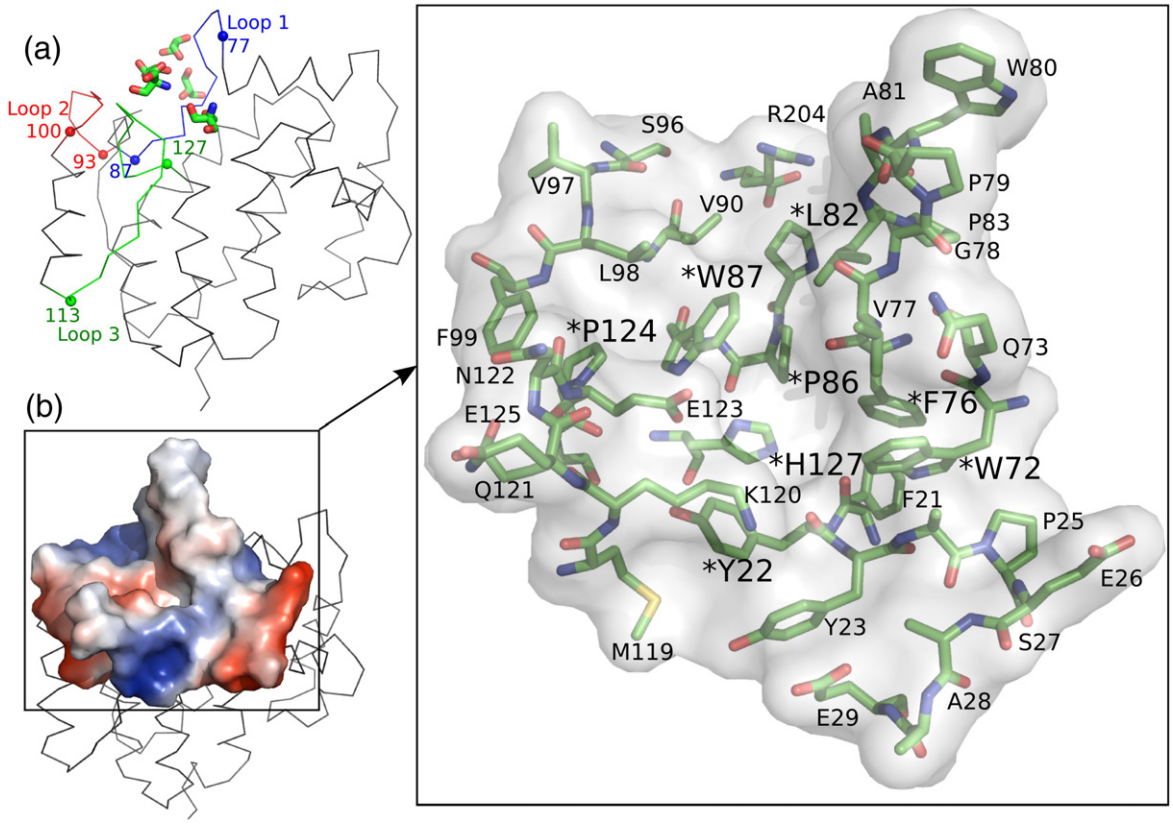

Fig. 4. The putative leader peptide binding pocket on EcDmsD. (a) Ribbon diagram of EcDmsD with stick representation of the glycerol and tris(hydroxymethyl)aminomethane molecules bound in the putative leader peptide binding site. The conserved loops that make up this pocket are shown in red, green, and blue. (b) Surface representation of the putative leader peptide binding site on a ribbon rendering of EcDmsD. The electrostatic potential is mapped onto the surface. The residues that make up the proposed pocket are shown within the semitransparent surface and labeled. Those residues previously shown by mutagenesis to be important for leader peptide binding are labeled with a larger font and an asterisk.

interaction between EcDmsD and the EcDmsA leader peptide. We have mapped these residues onto the surface of the EcDmsD crystal structure (Fig. 5a). We used the docking server 3D-Garden[16] to dock an EcDmsA leader peptide, modeled in an extended conformation, onto the crystal structure of EcDmsD. Previous work with Tat leader peptides has shown them to be unstructured in aqueous solution.[17] The docking procedure repeatedly placed the leader peptide across the EcDmsD surface, with the twin-arginine motif placed between the three conserved surface loops. This is roughly the position taken by the glycerol and tris molecules that were modeled into the electron density in the putative leader peptide binding pocket of the EcDmsD crystal structure (Fig. 4a). The position of the docked peptide is consistent with several studies that have highlighted the importance of the twin-arginine motif of the leader peptide in REMP–substrate interaction.[18,19,20]

The docked peptide was used as a starting point for a molecular dynamics simulation. After 63.5 ns of simulation, the EcDmsD–EcDmsA leader peptide complex became considerably more integrated (Fig. 5b), with a series of hydrogen bonds and van der Waals interactions formed between the EcDmsA leader peptide and EcDmsD (Fig. 5c). The dynamics simulation experiment resulted in small adjustments in the EcDmsD structure at helices 2, 3, 4, and 7, the N-terminal coil, and various side chain rotamers on the EcDmsD molecular surface. Interestingly, the regions that moved the most in the simulation corresponded to the sites of intermolecular contact within the crystalline lattice. Some of the residues that were previously shown by mutagenesis to be important for leader peptide binding were not observed on the molecular surface of the crystal structure (Fig. 5a). After the simulation, more of these proposed functionally important residues can be observed interacting with the bound peptide (Fig. 5b). Table 2 lists the putative molecular interactions between EcDmsD and the leader peptide of EcDmsA.

Conclusions

In this article, we report the first DmsD crystal structure with observed electron density for all residues in the protein. This 2.0 Å resolution crystal structure provides insight into the molecular details of the REMP that has been the most thoroughly characterized biochemically, that of the EcDmsD.[6] Additionally, we have presented a molecular dynamics

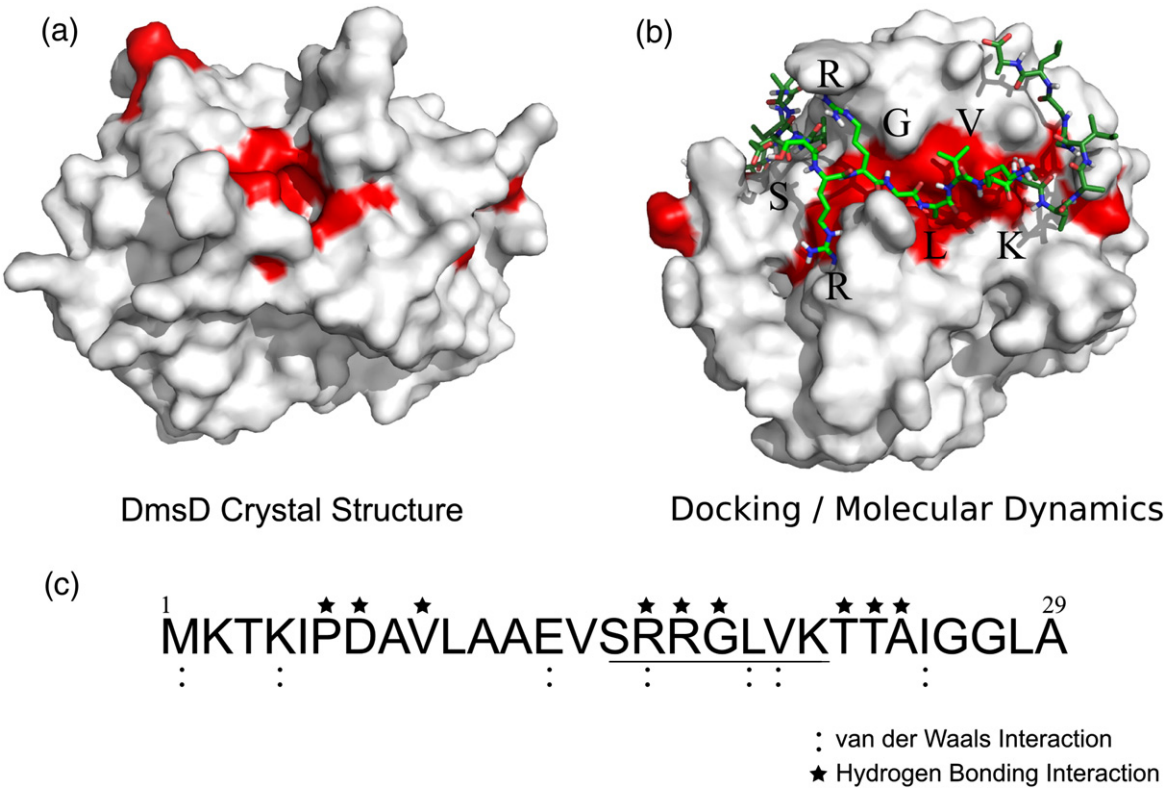

Fig. 5. Docking and molecular dynamics simulation experiments to predict the EcDmsD–EcDmsA leader peptide interactions. The molecular surface of EcDmsD is shown in white with red highlights for those residues previously shown by mutagenesis to be important for EcDmsA leader peptide binding. (a) Surface representation of the X-ray crystal structure of EcDmsD (chain A). (b) Molecular dynamics simulation of EcDmsD with the region of EcDmsA leader peptide that harbors the twin-arginine motif. Residues that are part of the twin-arginine consensus motif are labeled. (c) The sequence of the EcDmsA leader peptide. The twin-arginine consensus motif is underlined.

Table 2. Proposed interactions between EcDmsD and the leader peptide of EcDmsA suggested from docking and molecular dynamics studies

EcDmsA leader peptide	EcDmsD	Type of interaction
Met1	Pro201 Arg204	vdW
Lys4	Leu202	vdW
Pro6	Phe203 Arg204	HB
Asp7	Asp93	HB
Val9	Pro83 Trp91	HB
Glu13	Arg94	vdW
Arg16	Gln121 Asn122 Glu123 Glu125 Glu123 Glu125	HB/vdW
Arg17	Trp80 Asn122	HB
Gly18	Phe76	HB/vdW
Leu19	Phe76 His127 Tyr22	vdW
Val20	Trp72	vdW
Thr22	His68 Ala69	HB
Thr23	Glu65	HB
Ala24	Glu65	HB
Ile25	Glu65	vdW

vdW indicates van der Waals interactions; HB, hydrogen bond.

simulation-based prediction of an EcDmsD–EcDmsA leader peptide complex that is consistent with previous biochemical analysis of this interaction.

Materials and Methods

Expression and purification of EcDmsD

The EcDmsD protein was expressed using the plasmid pTDMS67 generated previously by Winstone *et al.*[7] in the host strain C41(DE3).[21] Overnight cultures were diluted (1%) into LB broth containing ampicillin (50 μg/mL), grown at 37 °C for 3 h, and induced with a final concentration of 1 mM IPTG for a further 3 h. Cells were harvested by centrifugation and lysed with an Avestin Emulsiflex-3C cell homogenizer. The lysate was clarified by centrifugation (30,000*g*) for 30 min. The supernatant was applied to a Ni^{++}–NTA column (5-mL column volume, Qiagen) equilibrated with Tris-buffered saline (TBS) (100 mM NaCl and 20 mM Tris–HCl, pH 8.0). The column was then washed with 10 column volumes of TBS, followed by 2 column volumes of TBS with 50 mM imidazole, and elution was carried out with a stepwise gradient of imidazole to 500 mM in 100 mM increments. EcDmsD was eluted from the column between 100 and 400 mM imidazole, and fractions containing EcDmsD were analyzed by SDS-PAGE, pooled, and concentrated using an Amicon ultracentrifuge filter (Millipore). The concentrated protein was then applied to a Sephacryl S-100 HiPrep 26/60 size-exclusion chromatography column on an AKTA Prime system (Pharmacia Biotech) at 1 mL/min using TBS as the buffer. Fractions containing EcDmsD were analyzed by SDS-PAGE, pooled, and concentrated to 36 mg/mL. The tag was removed by digestion with 1 U of recombinant enterokinase (Novagen) per 500 μg of protein for 24 h at 25 °C according to the manufacturer's instructions. The free hexahistidine affinity tag and uncleaved protein were removed by application of the protein mixture to Ni^{++}–NTA resin (5-mL column volume, Qiagen) equilibrated with TBS. The protein was further purified on a Sephacryl S-100 HiPrep 26/60 size-exclusion chromatography column for a second time. Fractions containing purified EcDmsD were analyzed by SDS-PAGE, pooled, and concentrated to 14.5 mg/mL using an Amicon ultracentrifuge filter (Millipore). The final 208-amino-acid protein construct consists of the full 204 residues of EcDmsD and a 4-residue N-terminal extension (RWGS), a byproduct of the proteolytic removal of the hexahistidine tag. This sequence has a calculated molecular mass of 23,831 Da and a theoretical isoelectric point of 5.0.

Crystallization

Crystals of EcDmsD were produced through hanging-drop vapor diffusion. Drops were prepared by combining 1 μL of protein solution (14.5 mg/mL) with 1 μL of reservoir solution. The optimized reservoir conditions were as follows: 100 mM Bis-tris (Bis[2-hydroxyethyl] amino- tris [hydroxymethyl]-methane), pH 6.5, 12% glycerol, and 1.25 M $(NH_4)_2SO_4$. The crystals were grown at 18 °C. This condition was derived from an initial hit in the Hampton Research sparse matrix crystal screen #2. Optimized crystals appeared after 72 h. The crystals belong to space group $P3_121$ with unit cell dimensions of 128.0 Å×128.0 Å×78.7 Å, with two molecules in the asymmetric unit and a Matthews coefficient of 3.9 $Å^3$ Da^{-1} (68.6% solvent).

Data collection

Crystals were incubated for 5 min in a cryoprotectant solution consisting of the mother liquor in which 20% of the water was replaced with glycerol. Diffraction data were collected at the Simon Fraser University Macromolecular X-Ray Diffraction Data Collection Facility using a MicroMax-007 rotating-anode microfocus generator operating at 40 KeV and 20 mA, VariMax Cu HF optics, an X-stream 2000 cryosystem, and an R-AXIS IV^{++} imaging-plate area detector (MSC-Rigaku). The data were collected and processed using the CrystalClear software package.[22] Reflections were collected beyond 2.0 Å for 180° of rotation using 0.5° oscillations. See Table 1 for data collection statistics.

Structure determination and refinement

The structure was solved using the molecular replacement program Phaser.[23] The search model was provided by the structure 1S9U, the EcDmsD homolog from *S. typhimurium*.[10] The model was adjusted manually using the program Coot,[24] and refinement was carried out using refmac 5.[25] The final round of restrained refinement with TLS restraints used TLS models generated by the TLS motion determination server.[26] The final refined structure was evaluated by PROCHECK.[27]

Structural analysis

Superimpositions were carried out using SSM superimposition in the program Coot.[24] Volume and surface area calculations were performed with UCSF Chimera.[28] Intramolecular interaction and fold analysis was performed with PROMOTIF 3.0.[29] The surface electrostatics analysis was performed with the adaptive Boltzmann–Poisson solver plug-in[30] and displayed using PyMOL.[31] *B*-factor analysis was performed by the program Baverage within the CCP4 suite of programs.[32]

Docking and molecular dynamics

Docking was performed on the 3D-Garden Web server.[16] Chain A of the EcDmsD asymmetric unit was used for the docking experiment. A *de novo*-generated polypeptide corresponding to the sequence of the preEcDmsA leader peptide (MKTKIPDAVLAAEVSRRGLVKTTAIGGLAMAS-SALTLPFSRIAHA) in an extended conformation was submitted to 3D-Garden using the default set of parameters. After docking, the leader peptide was truncated at residue 29. The package GROMACS version 3.3.3[33] was used to perform the simulations. The docked EcDmsA leader peptide–EcDmsD complex was processed using the GROMOS96 G43a2 force field, and simulations were run in an environment that keeps the number of atoms, pressure, and temperature constant. The complex was energy minimized *in vacuo* using the steepest descents algorithm such that the maximum of force on any atom (F_{max}) was less than 250.0 kJ mol^{-1} nm^{-1}. The complex was then embedded in a cubic box with a 9-Å space between the edge of the protein and the edge of the box and solvated using the spc216 simple point charge water model. The net charge of the system was made zero by replacing solvent molecules with sodium or chloride ions. The solvated system was energy minimized using the steepest descents algorithm to an F_{max} <1000.0 kJ mol^{-1} nm^{-1} and equilibrated for 1 ns with a time step of 0.002 ps, with position restraints placed on all atoms of the protein and peptide. Interactions were calculated using a twin-range pair list with long- and short-range cutoffs at 10 and 0.8 Å, respectively. Berendsen coupling was applied for temperature and pressure coupling at 300 K using a τ_T value of 0.1 and a τ_P value of 1.0. The simulation cube was periodic in all dimensions. After equilibration, the position restraints on the protein atoms were replaced with LINCS bond length constraints and bond angle restraints. The simulation was run on the WestGrid computing cluster "matrix" at variable intervals for a total of 63,500 ps, followed by steepest descents energy minimization to an F_{max} of 250.0 kJ mol^{-1} nm^{-1}. Analyses of the simulations were carried out using Visual Molecular Dynamics[34] and the GROMACS suite of programs.[33]

Figure preparation

Figures were prepared using PyMOL.[31] The alignment figure was prepared using the programs CLUSTALW[35] and ESPript[36] based on a multiple-sequence alignment generated by PSI-BLAST with five iterations.[37]

PDB accession code

Coordinates and structure factor amplitudes have been deposited with accession code 3EFP.

Acknowledgements

This work was supported in part by the Canadian Institute of Health Research (M.P., R.J.T.), the National Science and Engineering Research Council of Canada (M.P.), the Michael Smith Foundation for Health Research (M.P.), and the Canadian Foundation of Innovation (M.P.). The WestGrid computing resources used in this research are funded in part by the Canada Foundation for Innovation, Alberta Innovation and Science, BC Advanced Education, and the participating research institutions. WestGrid equipment is provided by IBM, Hewlett Packard, and SGI.

References

1. Weiner, J. H., Bilous, P. T., Shaw, G. M., Lubitz, S. P., Frost, L., Thomas, G. H. *et al.* (1998). A novel and ubiquitous system for membrane targeting and secretion of cofactor-containing proteins. *Cell*, **93**, 93–101.
2. Berks, B. C. (1996). A common export pathway for proteins binding complex redox cofactors? *Mol. Microbiol.* **22**, 393–404.
3. Buchanan, G., de Leeuw, E., Stanley, N. R., Wexler, M., Berks, B. C., Sargent, F. & Palmer, T. (2002). Functional complexity of the twin-arginine translocase TatC component revealed by site-directed mutagenesis. *Mol. Microbiol.* **43**, 1457–1470.
4. Papish, A. L., Ladner, C. L. & Turner, R. J. (2003). The twin-arginine leader-binding protein, DmsD, interacts with the TatB and TatC subunits of the *Escherichia coli* twin-arginine translocase. *J. Biol. Chem.* **278**, 32501–32506.
5. Turner, R. J., Papish, A. L. & Sargent, F. (2004). Sequence analysis of bacterial redox enzyme maturation proteins (REMPs). *Can. J. Microbiol.* **50**, 225–238.
6. Chan, C. S., Winstone, T. M., Chang, L., Stevens, C. M., Workentine, M. L., Li, H. *et al.* (2008). Identification of residues in DmsD for twin-arginine leader peptide binding, defined through random and bioinformatics-directed mutagenesis. *Biochemistry*, **47**, 2749–2759.
7. Winstone, T. L., Workentine, M. L., Sarfo, K. J., Binding, A. J., Haslam, B. D. & Turner, R. J. (2006). Physical nature of signal peptide binding to DmsD. *Arch. Biochem. Biophys.* **455**, 89–97.
8. Hatzixanthis, K., Clarke, T. A., Oubrie, A., Richardson, D. J., Turner, R. J. & Sargent, F. (2005). Signal peptide–chaperone interactions on the twin-arginine protein transport pathway. *Proc. Natl Acad. Sci. USA*, **102**, 8460–8465.
9. Tranier, S., Iobbi-Nivol, C., Birck, C., Ilbert, M., Mortier-Barriere, I., Mejean, V. & Samama, J. P. (2003). A novel protein fold and extreme domain swapping in the dimeric TorD chaperone from *Shewanella massilia*. *Structure*, **11**, 165–174.
10. Qiu, Y., Zhang, R., Binkowski, T. A., Tereshko, V., Joachimiak, A. & Kossiakoff, A. (2008). The 1.38 Å crystal structure of DmsD protein from *Salmonella typhimurium*, a proofreading chaperone on the Tat pathway. *Proteins*, **71**, 525–533.
11. Kirillova, O., Chruszcz, M., Shumilin, I. A., Skarina, T., Gorodichtchenskaia, E., Cymborowski, M. *et al.* (2007). An extremely SAD case: structure of a putative redox-enzyme maturation protein from *Archaeoglobus fulgidus* at 3.4 Å resolution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **63**, 348–354.
12. Maillard, J., Spronk, C. A., Buchanan, G., Lyall, V., Richardson, D. J., Palmer, T. *et al.* (2007). Structural diversity in twin-arginine signal peptide-binding proteins. *Proc. Natl Acad. Sci. USA*, **104**, 15641–15646.
13. Sarfo, K. J., Winstone, T. L., Papish, A. L., Howell, J. M., Kadir, H., Vogel, H. J. & Turner, R. J. (2004). Folding forms of *Escherichia coli* DmsD, a twin-arginine leader binding protein. *Biochem. Biophys. Res. Commun.* **315**, 397–403.

14. Tranier, S., Mortier-Barriere, I., Ilbert, M., Birck, C., Iobbi-Nivol, C., Mejean, V. & Samama, J. P. (2002). Characterization and multiple molecular forms of TorD from *Shewanella massilia*, the putative chaperone of the molybdoenzyme TorA. *Protein Sci.* **11**, 2148–2157.
15. Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536–540.
16. Lesk, V. I. & Sternberg, M. J. (2008). 3D-Garden: a system for modelling protein–protein complexes based on conformational refinement of ensembles generated with the marching cubes algorithm. *Bioinformatics*, **24**, 1137–1144.
17. San Miguel, M., Marrington, R., Rodger, P. M., Rodger, A. & Robinson, C. (2003). An *Escherichia coli* twin-arginine signal peptide switches between helical and unstructured conformations depending on the hydrophobicity of the environment. *Eur. J. Biochem.* **270**, 3345–3352.
18. Gross, R., Simon, J. & Kroger, A. (1999). The role of the twin-arginine motif in the signal peptide encoded by the hydA gene of the hydrogenase from *Wolinella succinogenes*. *Arch. Microbiol.* **172**, 227–232.
19. Stanley, N. R., Sargent, F., Buchanan, G., Shi, J., Stewart, V., Palmer, T. & Berks, B. C. (2002). Behaviour of topological marker proteins targeted to the Tat protein transport pathway. *Mol. Microbiol.* **43**, 1005–1021.
20. Li, H., Faury, D. & Morosoli, R. (2006). Impact of amino acid changes in the signal peptide on the secretion of the Tat-dependent xylanase C from *Streptomyces lividans*. *FEMS Microbiol. Lett.* **255**, 268–274.
21. Miroux, B. & Walker, J. E. (1996). Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260**, 289–298.
22. Pflugrath, J. W. (1999). The finer things in X-ray diffraction data collection. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **55**, 1718–1725.
23. McCoy, A. J. (2007). Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **63**, 32–41.
24. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2126–2132.
25. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **53**, 240–255.
26. Painter, J. & Merritt, E. A. (2006). Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **62**, 439–450.
27. Morris, A. L., MacArthur, M. W., Hutchinson, E. G. & Thornton, J. M. (1992). Stereochemical quality of protein structure coordinates. *Proteins*, **12**, 345–364.
28. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612.
29. Hutchinson, E. G. & Thornton, J. M. (1996). PROMOTIF—a program to identify and analyze structural motifs in proteins. *Protein Sci.* **5**, 212–220.
30. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.
31. DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* DeLano Scientific, San Carlos, CA.
32. Collaborative Computational Project No. 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **50**, 760–763.
33. Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E. & Berendsen, H. J. (2005). GROMACS: fast, flexible, and free. *J. Comput. Chem.* **26**, 1701–1718.
34. Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graphics*, **14**, 27–28.
35. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.
36. Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). ESPript: multiple sequence alignments in PostScript. *Bioinformatics*, **15**, 305–308.
37. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.