Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallographic analysis of *Bacillus subtilis* CsaA

Yuliya A. Shapova and Mark Paetzel*

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

Correspondence e-mail: mpaetzel@sfu.ca

Received 8 December 2006
Accepted 31 January 2007

PDB References: CsaA, $P3_221$, 2nzo, r2nzosf; $P42_12$, 2nzh, r2nzhsf.

Bacillus subtilis CsaA (BsCsaA) has been proposed to act as a protein-secretion chaperone in the Sec-dependent translocation pathway, possibly compensating for the lack of SecB in the Gram-positive eubacterium *Bacillus subtilis*. This paper presents the cloning, purification, crystallization and structures of BsCsaA in two space groups ($P42_12$ and $P3_221$) solved and refined to resolutions of 1.9 and 2.0 Å, respectively. These structures complement the previously available crystal structure of CsaA from the Gram-negative eubacterium *Thermus thermophilus* (TtCsaA) and provide a direct structural basis for the interpretation of previously available biochemical data on BsCsaA. The sequence and structure of the proposed substrate-binding pocket are analyzed and discussed. A comparison with the TtCsaA structure reveals a different pattern of electrostatic potential in the vicinity of the binding site, which overlaps with a region of high sequence variability. In addition, the dimerization interface of this homodimeric protein is analyzed and discussed.

1. Introduction

The Sec-dependent protein-targeting and translocation pathway is universally conserved across all three domains of life (Pohlschroder *et al.*, 2005). In eubacteria, secreted proteins are synthesized in the cytosol as precursors carrying an amino-terminal signal peptide (Driessen *et al.*, 2001). These precursors are targeted to the translocation machinery at the cytosolic membrane. In *Escherichia coli*, the translocation machinery (translocase) involves a translocation channel composed of three integral membrane proteins SecYEG, SecA, an ATPase that provides energy for the translocation process, and several accessory proteins, such as SecD, SecF, YajC (de Keyzer *et al.*, 2003; Driessen *et al.*, 2001; Stephenson, 2005) and YidC (Yi & Dalbey, 2005). In bacteria, most of the protein secretion is carried out post-translationally (Pohlschroder *et al.*, 2005), with the homotetrameric SecB functioning as a targeting factor that binds to the core regions of the newly synthesized proteins and targets them to the SecA subunit of the translocase while maintaining them in an unfolded translocation-competent state (Driessen *et al.*, 2001). Interestingly, certain species of eubacteria, and Gram-positive bacteria in particular, lack SecB. The Gram-positive eubacterial species that has been investigated the most from a protein-secretion perspective is *Bacillus subtilis* (Kunst *et al.*, 1997).

The *B. subtilis csaA* gene was identified as being capable of suppressing growth and secretion defects in *E. coli secA* and *secB* mutants (Muller *et al.*, 1992). The *B. subtilis* CsaA (BsCsaA) protein restored the function of thermally inactivated firefly luciferase in chaperone mutant strains of *E. coli*; in addition, it prevented the aggregation of thermally inactivated luciferase *in vitro* (Muller, Bron *et al.*, 2000). Furthermore, it has been demonstrated that CsaA interacts with the SecA subunit of the Sec translocase, as well as with a number of secreted precursors, including *B. subtilis* prePhoB and preYvaY (Muller, Ozegowski *et al.*, 2000; Linde *et al.*, 2003). BsCsaA induced the translocation of prePhoB in *E. coli* membrane vesicles containing translocation machinery from *B. subtilis* (Muller, Ozegowski *et al.*, 2000). More recently, it has been demonstrated that the levels of expression of the *csaA* gene were upregulated 3.5-fold in response to severe secretion stress in *B. subtilis* (Hyyrylainen *et al.*, 2005). The above evidence suggests that *B. subtilis* CsaA acts as a protein-secretion chaperone of the Sec-dependent protein-targeting and translocation pathway, possibly compensating for the lack of SecB.

There are several Gram-negative species of eubacteria that have both CsaA and SecB. One of these species is *Thermus thermophilus*. The crystal structure of CsaA from *T. thermophilus* has been solved to 2.0 Å resolution (Kawaguchi *et al.*, 2001). The functional unit of CsaA is a homodimer, in which each monomer is composed of a β-barrel domain which resembles an oligonucleotide/oligosaccharide-binding (OB) fold, as well as short N- and C-terminal extensions from the β-barrel core which form the dimer interface (Kawaguchi *et al.*, 2001). The OB-fold proteins are incredibly diverse and have been shown to bind a variety of substrates, such as RNA, ssDNA, oligosaccharides and proteins (Arcus, 2002). There is no sequence or structural similarity between CsaA and SecB, which functions as a dimer of dimers with each monomer composed of two α-helices and four β-strands (Xu *et al.*, 2000; Dekker *et al.*, 2003).

CsaA shares sequence and structural homology with TRBP111, a tRNA-binding protein, as well as with the C-terminal domain of methionyl-tRNA synthetase (MetRS). TRBP111 binds to the outside corner of the L-shaped tRNA, possibly *via* electrostatic interactions with the tRNA phosphate backbone, and thus may act as a structure-specific tRNA chaperone (Swairjo *et al.*, 2000). The C-terminal domain of MetRS also possesses general tRNA-binding ability and serves as a dimerization domain in several species (Crepin *et al.*, 2002). Based on the structural similarity of these two proteins to CsaA, it has been proposed that CsaA may possess a second tRNA-binding ability (Kawaguchi *et al.*, 2001).

This paper presents two crystal structures of CsaA from *B. subtilis*, a Gram-positive eubacterium, solved at 2.0 and 1.9 Å resolution. These structures provide a basis for the interpretation of previous biochemical characterization of BsCsaA, as well as a comparison with the available TtCsaA structure from the Gram-negative organism *T. thermophilus*. In addition, these structures may provide further clues to the mode of binding of CsaA to its proposed substrates.

2. Materials and methods

2.1. Cloning, overexpression and purification

Genomic DNA from *B. subtilis* was purified using the standard phenol–chloroform method (Sambrook *et al.*, 1989). A region of *B. subtilis* genomic DNA corresponding to the *csaA* gene was amplified by PCR using the forward primer 5′-g agc tga ata CAT ATG gca gtt att gat gac ttt gag aaa ttg gat atc, incorporating the *Nde*I restriction site, and the reverse primer 5′-g aat gct cat GTC GAC tta tta tcc gat ttt tgt gcc gtt tgg gac agg ctg, incorporating the *Sal*I restriction site. PCR was carried out using HotStar Taq polymerase (Qiagen) and the PCR products were cloned into the overexpression vector pET-28a(+) (Novagen), incorporating a hexahistidine tag at the N-terminus of the protein. Sequencing was performed to confirm the BsCsaA sequence. *E. coli* BL21(DE3) cells transformed with the recombinant vector were grown in LB media containing kanamycin at 310 K to an OD of 0.6 and induced with 0.5 m*M* IPTG at 298 K for 16 h. Harvested cells were resuspended in 50 m*M* Tris–HCl pH 8.0, 0.1 *M* NaCl, 10 m*M* imidazole and subjected to sonication followed by lysis in a cell homogenizer (Avestin EmulsiFlex-C3). The cell lysate was centrifuged at 31 000*g* for 30 min and the supernatant containing the overexpressed His-tagged CsaA protein was passed over nickel–nitriloacetic acid beads (Qiagen) and eluted in a stepwise manner with 50 m*M* Tris pH 8.0, 300 m*M* NaCl buffer containing 100–400 m*M* imidazole. Fractions that contained CsaA were pooled together and applied onto a HiTrap Sephadex G-25 desalting column (Amersham Biosciences) in order to remove imidazole. The His tag was removed by adding 20 units of thrombin per milligram of CsaA and incubating at room temperature for 16 h. The cleaved CsaA was again passed over Ni–NTA resin and the flowthrough fraction was concentrated and applied onto a HiPrep 16/60 Sephacryl S-100 HR size-exclusion column (Amersham Biosciences) pre-equilibrated with 20 m*M* Tris–HCl pH 8.0, 100 m*M* NaCl. Fractions containing CsaA were pooled together and concentrated to 10 mg ml^{-1} using an Amicon ultracentrifugal filter (Millipore). The protein concentration was determined using the bicinchoninic acid (BCA) protein assay (Pierce).

2.2. Crystallization and data collection

The initial crystallization conditions were obtained by the hanging-drop vapor-diffusion method using sparse-matrix screens from Hampton Research. The crystallization conditions were optimized using grid screens based on the initial hits. All drops contained 1 µl protein and 1 µl reservoir solution and were equilibrated over 1 ml reservoir solution. The reservoir condition that produced the $P3_221$ crystals contained 0.1 *M* potassium dihydrogen phosphate, 12% PEG 4000. The protein solution contained 12.5 mg ml^{-1} BsCsaA in 20 m*M* Tris–HCl pH 8.0, 100 m*M* NaCl. Crystals appeared after 2 d incubation at room temperature. The reservoir condition that produced the $P42_12$ crystals contained 0.2 *M* ammonium sulfate and 30% PEG 8000. The protein solution contained 9 mg ml^{-1} BsCsaA in 20 m*M* Tris pH 8.0, 1 *M* NaCl. The

crystals were transferred into a cryoprotectant solution that consisted of mother liquor in which 25% of the water had been replaced by glycerol. The 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 mV and 20 mA, VariMax Cu HF optics, an X-stream 2000 cryosystem and an R-AXIS IV^{++} imaging-plate area detector (MSC-Rigaku). All data were collected and processed using the *CrystalClear* software package (Pflugrath, 1999). The trigonal crystals ($P3_221$) diffracted to beyond 2.0 Å resolution. The tetragonal crystals ($P42_12$) diffracted to beyond 1.9 Å resolution. Complete data sets were collected for each crystal form. See Table 1 for data-collection statistics.

2.3. Structure determination and refinement

The structures of *B. subtilis* CsaA were solved by molecular replacement using the program *Phaser* (McCoy *et al.*, 2005). The coordinates of *T. thermophilus* CsaA chain *A* (1gd7A) were used as a search model. Several rounds of restrained refinement with *REFMAC*5 (Murshudov *et al.*, 1997) and manual adjustment and manipulation using *Coot* (Emsley & Cowtan, 2004) were used to build the BsCsaA models. *CNS* (Brünger *et al.*, 1998) was utilized as an additional tool to carry out the combined simulated annealing, energy minimization and *B*-factor refinement. The final models were obtained by running restrained refinement in *REFMAC*5 with TLS restraints obtained from the TLS motion-determination server (Painter & Merritt, 2006). The quality of the final models was assessed with the program *PROCHECK* (Morris *et al.*, 1992).

2.4. Structural analysis

Superimpositions were carried out using the program *SuperPose* (Maiti *et al.*, 2004). The surface/binding-pocket analysis was carried out using *CASTp* (Binkowski *et al.*, 2003) with a 1.4 Å probe radius. The mapping of the sequence conservation onto the three-dimensional structure was performed using *CONSURF* (Glaser *et al.*, 2003). The figures were produced using *PyMOL* (DeLano, 2002). The sequence-alignment analysis (Fig. 1) was prepared using *ClustalW* (Thompson *et al.*, 1994) and *ESPript* v.2.2 (Gouet *et al.*, 2003). The protein–protein interaction server was used to analyze the dimer interface (Jones & Thornton, 1995). The surface electrostatic analysis was performed using the vacuum electrostatics utility in the program *PyMOL*.

3. Results and discussion

3.1. Sequence-alignment analysis

There are three residues that appear to be universally conserved among the CsaA, TRBP111 and MetRS proteins: Gly38, Asn69 and Ser80 (Fig. 1). Asn69 and Ser80 have been identified as residues that are crucial for tRNA binding in TRBP111 (Swairjo *et al.*, 2000). The following residues appear to be conserved in most CsaA proteins and are different in TRBP111 and MetRS: 26, 29–30, 42, 46–51, 70, 83 and 86.

Table 1
Crystallographic data-collection and refinement statistics for *B. subtilis* CsaA.

Values in parentheses are for the highest resolution shell.

Data collection		
Space group	$P3_221$	$P42_12$
Unit-cell parameters (Å)	$a = b = 148.4$, $c = 54.1$	$a = b = 109.2$, $c = 37.4$
Resolution range (Å)	28.05–2.00 (2.07–2.00)	54.57–1.90 (1.97–1.90)
Total No. of reflections	209396	190906
No. of unique reflections	45675	17190
Average redundancy	4.58 (4.25)	11.11 (9.99)
Completeness (%)	98.6 (97.0)	93.3 (87.0)
$R_{\rm merge}$†	0.044 (0.310)	0.051 (0.286)
$\langle I/\sigma(I)\rangle$	20.4 (4.6)	31.9 (7.2)
Refinement		
Molecules in ASU	4	2
No. of protein atoms	3306	1654
No. of solvent atoms	320	123
Water	284	117
Glycerol	36	6
$R_{\rm work}$‡	0.192	0.202
$R_{\rm free}$§	0.230	0.245
R.m.s.d. bond lengths (Å)	0.019	0.017
R.m.s.d. bond angles (°)	1.78	1.61
Average *B* factor, protein (Å^2)	17.4	17.2
Average *B* factor, solvent (Å^2)	32.0	32.5
Ramachandran analysis (%)		
Favored	91.5	90.3
Allowed	8.5	9.1
Generously allowed	0.0	0.6
Disallowed	0.0	0.0
Residues missing from the models owing to a lack of electron density	Chain *C*, 24–28	Chain *B*, 1–2
Residues modeled as alanines owing to a lack of side-chain density	Chain *C*, 32, 40, 91; chain *D*, 8, 22, 40	Chain *A*, 1; chain *B*, 25, 27

† $R_{\rm merge} = \sum|I - \langle I\rangle|/\sum\langle I\rangle$, where I is the observed intensity obtained from multiple observations of symmetry-related reflections after rejections. ‡ $R_{\rm work} = \sum\big||F_{\rm o}| - |F_{\rm c}|\big|/\sum|F_{\rm o}|$. § $R_{\rm free}$ is calculated in the same way as the R factor for data omitted from refinement (5% of reflections for all data sets).

Based on the phylogenetic tree analysis of 18 CsaA and 18 TRBP111 and MetRS (C-terminal domain only) proteins (data not shown), CsaA proteins are distinct from the other group and form their own subfamily.

It is notable that the *csaA* gene is found in many species of Gram-positive and Gram-negative eubacteria, as well as archaea, yet in the Gram-positive eubacteria CsaA seems to be present only in species of the genera *Bacilli* and *Clostridia*.

3.2. Structural overview

CsaA from *B. subtilis* (BsCsaA) is a homodimeric molecule in which each monomer is 110 amino acids in length and has a molecular weight of 12 kDa. The core structure of each monomer displays a well described oligonucleotide/oligosaccharide-binding (OB) fold: a five-stranded β-barrel with a short capping α-helix (Murzin, 1993). In the case of BsCsaA, the β-barrel is formed by strands s1, s2, s3, s4 and s7 and an α-helix h3 located between s3 and s4 (Fig. 2*a*). An additional short helix h2 is found in the loop region between s2 and s3. Two short β-strands, s5 and s6, hydrogen bond to each other and are connected by a type II β-turn. In addition to the

β-barrel, an α-helix h1 is found at the N-terminus of the protein and the C-terminus contains strands s8 and s9. These elements participate in the formation of the dimeric structure of CsaA.

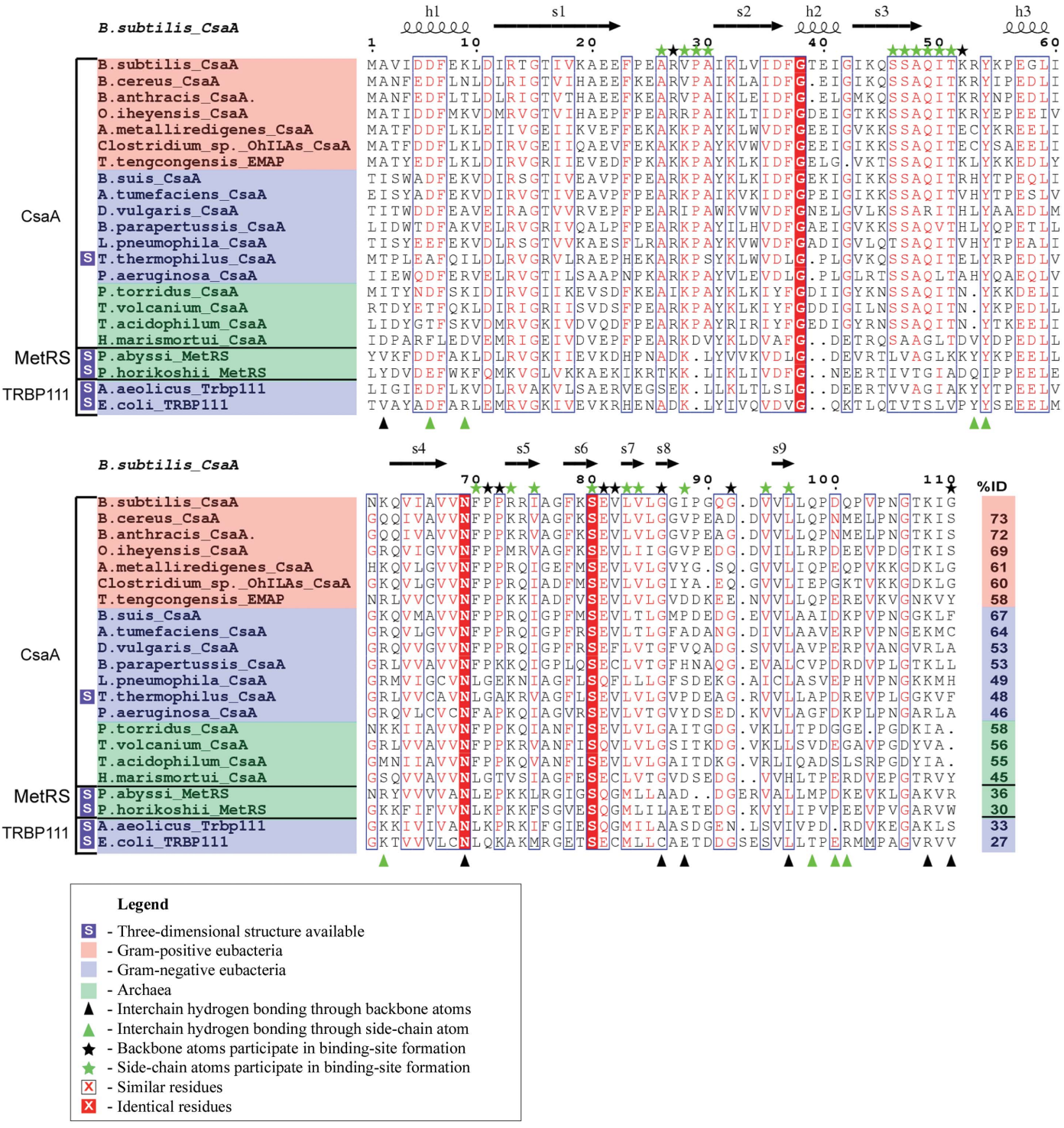

Figure 1
Sequence alignment of BsCsaA and other CsaA proteins and the tRNA-binding proteins MetRS and TRBP111. The percentage identity of each protein with BsCsaA is listed at the bottom right of each sequence. The secondary structure of BsCsaA, as determined by *DSSP* (Kabsch & Sander, 1983), is shown above the alignment. The sequence numbering is that of BsCsaA. The SWISS-PROT accession Nos. for the sequences used in the alignment are as follows: *Bacillus subtilis*, P37584; *B. cereus*, Q739K5; *B. anthracis*, Q81RI2; *Oceanobacillus iheyensis*, Q8ELP7; *Alkaliphilus metalliredigenes*, Q3C798; *Clostridium* sp. OhILAs, Q1ETM8; *Thermoanaerobacter tengcongensis*, Q8RDK8; *Brucella suis*, Q8FZF9; *Agrobacterium tumefaciens*, Q8UDB9; *Desulfovibrio vulgaris*, Q72G33; *Bordetella parapertussis*, Q7W9Y8; *Legionella pneumophila*, Q5ZSH0; *Thermus thermophilus*, Q9AQH8; *Pseudomonas aeruginosa*, Q9HZ19; *Picrophilus torridus*, Q6L208; *Thermoplasma volcanium*, Q97BA7; *T. acidophilum*, Q9HJC4; *Haloarcula marismortui*, Q5V1H5; *Pyrococcus abyssi*, Q9V011; *P. horikoshii*, O58023; *Aquifex aeolicus*, O66738; *Escherichia coli*, P42589.

The two structures of BsCsaA differ somewhat, particularly in the positions of the atoms in residues 23–32 and 73–79 (Fig. 2*b*). These regions contain residues that contribute to the formation of the putative substrate-binding site (Fig. 1). The four chains in the asymmetric unit of the structure from the trigonal crystals ($P3_221$) superimpose over the backbone atoms with a root-mean-square deviation (r.m.s.d.) of 0.3 Å. The two chains in the asymmetric unit of the structure with space group $P42_12$ superimpose with an r.m.s.d. of 0.7 Å. The r.m.s.d. of superposition of all six chains over backbone atoms is 0.4 Å.

The structural neighbours of *B. subtilis* CsaA were found by performing a VAST search of the medium-redundancy PDB database (Gibrat *et al.*, 1996). The closest structural neighbor is the CsaA protein from *T. thermophilus*, with an r.m.s.d. of superposition over the backbone atoms of 1.6 Å (Fig. 2*d*). Other structural neighbors include the tRNA-binding protein TRBP111 (PDB code 1pyb and 1pxf), the C-terminal domain of methionyl-tRNA synthetase (1mkh), a MetRS-related protein (2cwp) and the EMAPII domain of the p43 protein from the human aminoacyl-tRNA synthetase complex (1e7z). The r.m.s.d. of superposition of these proteins and BsCsaA ranges from 2.1 to 3.8 Å, while the sequence identity ranges from 27 to 48%.

3.3. The dimerization interface

The two monomers of BsCsaA are held together by 21 hydrogen bonds (Table 2). Since the dimer has a local twofold axis of symmetry, the same residues form the hydrogen-bonding interactions in both chains. The majority of the inter-chain hydrogen-bonding network is localized in the C-terminal portion of the protein and occurs through main-chain atoms. On average, 1550 Å^2 (about 22.5%) of the total accessible surface area is buried in the interface, which is mostly comprised of nonpolar atoms. There are two notable hydrogen-bonding interactions that occur in the residues near the N-terminus. Firstly, there is a hydrogen bond between the side chains of Lys9 and Asp6′ (and Lys9′ and Asp6, respectively). In the $P42_12$ structure, this hydrogen bond is formed directly from Lys9 NZ to Asp6 OD1. However, in the $P3_221$ structure this hydrogen bond is indirect and occurs through a water molecule (W151). All four residues (Lys9, Asp6′, Lys9′ and Asp6) make hydrogen bonds to water W151. The $P3_221$ structure has two hydrogen bonds that occur at the N-terminus between the backbone atoms of Ala2 and Asn69′ (and *vice versa*). This interaction is not observed in the $P42_12$ structure. Among the residues that make hydrogen-bonding interactions through their side chains, Tyr54 is notable because this residue is highly conserved among the CsaA proteins, TRBP111 and the C-terminal portion of MetRS, all of which are dimers. The Tyr54 side-chain OH forms hydrogen-bonding interactions with the side-chain OD1 and OD2 of Asp100 (Fig. 2*c*). Although Asp100 is conserved to a lesser degree than Tyr54, most related proteins

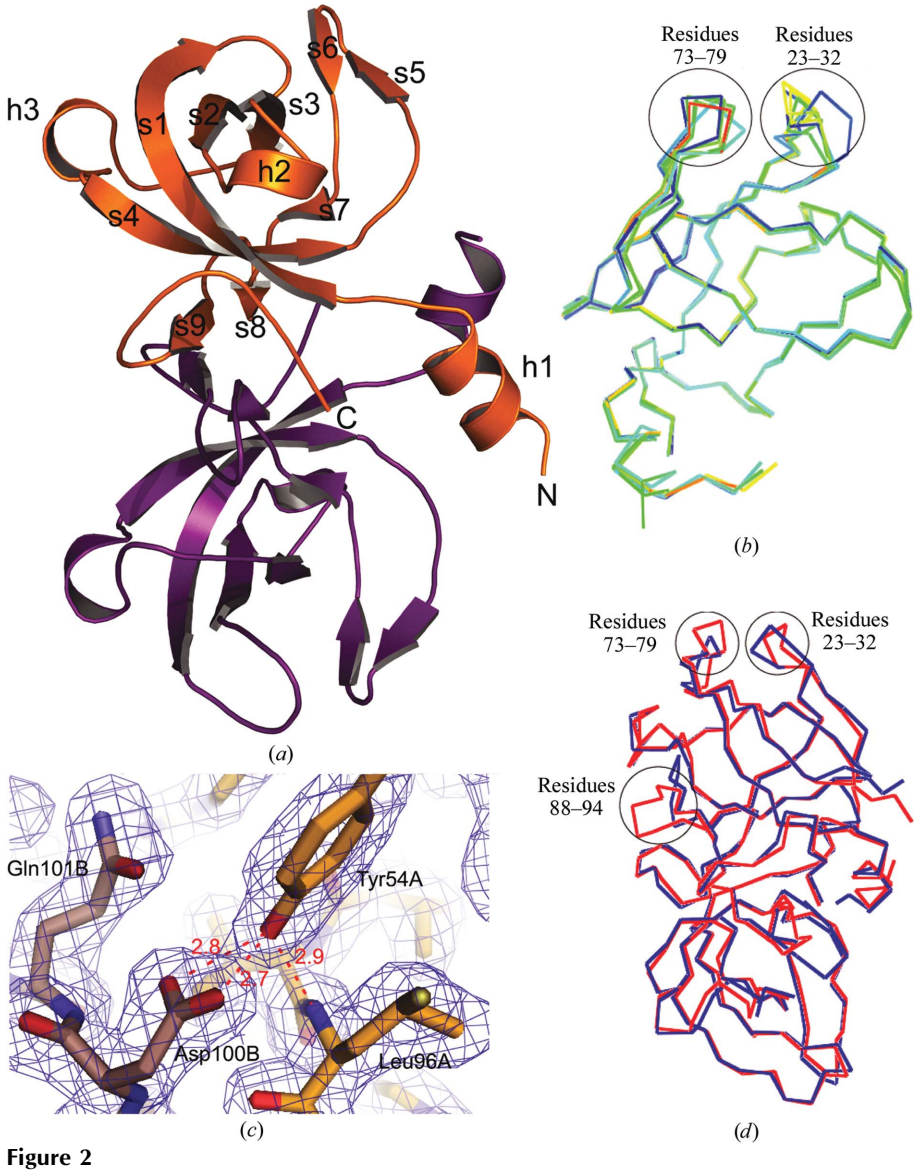

Figure 2
(*a*) A cartoon diagram of BsCsaA. Chains *A* and *B* are shown in orange and purple, respectively. (*b*) A Cα trace of the six superimposed chains from the two structures of BsCsaA, colored by *B* factor. Areas with the lowest *B* factor are colored blue and areas with the highest *B* factor are colored red. (*c*) A fragment of the $2F_o - F_c$ electron density at 1.0σ, demonstrating dimerization interactions *via* the side-chain OH of Tyr54 hydrogen bonding to the side chain of Asp100′. (*d*) A Cα trace of the dimeric structures of BsCsaA (blue) and TtCsaA (red). Regions that show different conformations in different chains are labeled.

contain residues at this position that are capable of providing a hydrogen-bond acceptor for Tyr54 OH. It is therefore possible that Tyr54 may be important for CsaA dimerization.

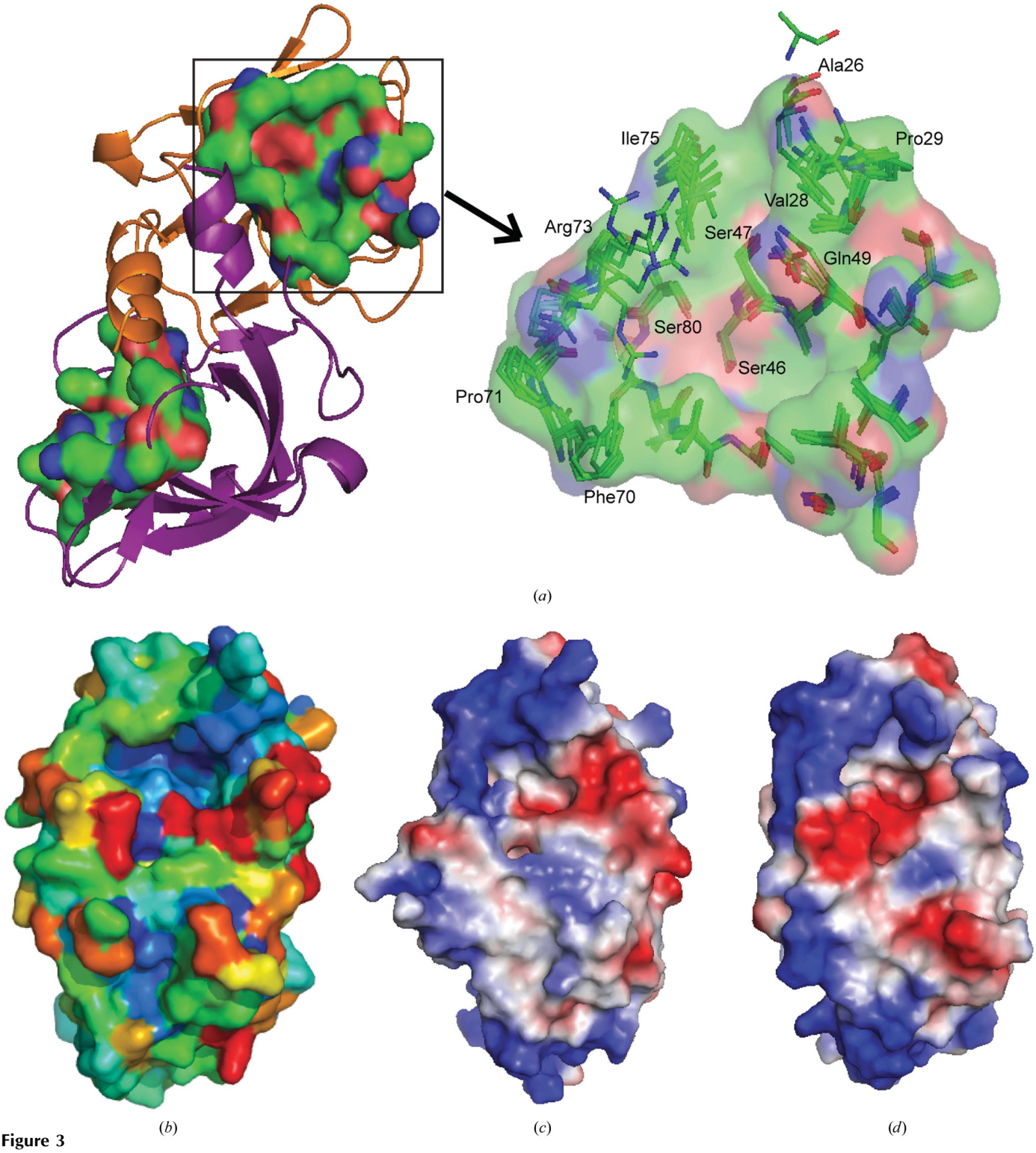

Figure 3
(*a*) Left: a cartoon representation of the BsCsaA dimer with the substrate-binding cavities shown as a surface. C atoms are shown in green, O atoms in red and N atoms in blue. Right: the superimposition of the residues forming the binding site from the six chains of the two structures of BsCsaA. The surface corresponds to the putative substrate-binding cavity of the BsCsaA structure in space group $P42_12$, chain *A*. (*b*) The surface representation of BsCsaA, colored by the conservation score, with the most conserved residues colored dark blue and the least conserved residues colored red. The figure was produced using *ConSurf* (Glaser *et al.*, 2003). The conservation scores were obtained from sequence alignment of 36 CsaA, TRBP and MetRS (C-terminal domains only) sequences. (*c*) The protein surface electrostatics of *T. thermophilus* CsaA. (*d*) The protein surface electrostatics of *B. subtilis* CsaA. Areas colored white, red and blue correspond to neutral, negative and positive surface electrostatic potentials, respectively.

Table 2
The inter-chain hydrogen bonds between the two monomers of BsCsaA, as determined from the optimal hydrogen-bonding network (Hooft *et al.*, 1996).

These hydrogen bonds occur in both the $P3_221$ and $P42_12$ crystal forms. Owing to the presence of a local twofold axis of symmetry in the dimer, the same residues form hydrogen bonds in both chains of the dimer (except as noted).

Donors		Acceptors	
Residue	Atom	Residue	Atom
Ala2†	N	Asn69′	O
Lys9‡	NZ	Asp6′	OD1
Arg53	NH1	Gln101′	OE1
Tyr54	OH	Asp100′	OD2
Tyr54	OH	Asp100′	OD1
Lys62	NZ	Asp100′	OD2
Gly86	N	Gly110′	O
Ile88	N	Lys108′	O
Gln98	N	Gln98′	O
Gln98§	NE2	Gln98′	OE1
Asp100	N	Leu96′	O
Gly110	N	Gly86′	O
Gly110	OXT	Gly86′	O

† Occurs only in $P3_221$ structure. ‡ Occurs only in $P42_12$ structure. § Occurs only once in each dimer.

3.4. The potential substrate-binding site

CastP (Binkowski *et al.*, 2003) analysis of the solvent-accessible surface revealed that BsCsaA contains two large T-shaped cavities, one in each monomer. The two binding sites are separated by a rotation of approximately 90° about the axis of the dimer (Fig. 3*a*). These cavities are located on one side of each β-barrel and are predominantly formed by loops. The following residues contribute to the binding-cavity formation: 26–30, 46–52, 70–73, 75, 80–84, 86, 88, 92, 94, 96 and 110′. The cavity dimensions are approximately 15 × 15 Å, with a depth ranging from ∼3 to 6 Å. The residues forming the cavity are mostly hydrophobic and come from the same monomer, except for the C-terminal Gly110. The residues that line the floor of the cavity are Ser46, Ser47, Ala48, Ile50, Ser80, Glu81, Val82, Val84 and Leu96. The three serines (Ser46, Ser47 and Ser80) form a hydrophilic patch in the center of the cavity floor. The cavity walls are formed by Ala26, Val28, Gln49, Phe70, Pro71, Pro72, Arg73, Ile75, Leu83, Gly86, Ile88, Val94 and Gly110′. Most of the residues forming the binding site do not show large variations in the position of their atoms among the six chains seen in the two BsCsaA structures; however, the following residues show significant variations in position: Ala26, Val28, Pro29, Phe70, Pro71, Pro72, Arg73 and Ile75. These residues are predominantly located in loops that together form one wall of the binding cavity (Fig. 3*a*). These regions demonstrated weaker electron density and higher *B* factors, which are consistent with greater mobility in this region. These residues superimpose with an r.m.s.d. of 1.4 Å (over all atoms), whereas the r.m.s.d. for all residues of the binding site is 0.9 Å and that for all atoms in the six models is 0.8 Å (Fig. 3*a*). It has previously been demonstrated that CsaA has an affinity for binding multiple peptides (Linde *et al.*, 2003). It is possible that the flexibility of this wall is important to accommodate the binding of a variety of peptide substrates, which is consistent with the general chaperone function.

It has previously been shown that BsCsaA has a higher affinity for denatured peptides, thus indicating preferred binding to unfolded proteins (Linde *et al.*, 2003). The hydrophobic nature of the binding cavity is consistent with the chaperone activity of BsCsaA, allowing it to bind the exposed hydrophobic residues in an unfolded protein substrate. The hydrophilic patch in the floor of the binding cavity, formed by the three serines (Ser46, Ser47 and Ser80), provides the possibility of hydrogen-bonding interactions between the residues of the cavity and the backbone atoms of the protein substrate in an extended conformation. It is possible that the protein substrate wraps around the surface of CsaA, much like the substrate–chaperone interactions recently described for SecB (Crane *et al.*, 2006).

Overall, the residues of the binding site are well conserved among the sequences of CsaA proteins (Fig. 3*b*). However, some residues that are well conserved among the CsaA proteins are different in TRBP111 and MetRS, such as Pro29, Ser46, Gln49 and Thr51. Swairjo and coworkers have identified TRBP111 residues important for tRNA binding (Swairjo *et al.*, 2000). It is worth noting that most of these residues are conserved between TRBP and CsaA, such as Ser80 (Ser82 in TRBP111), Arg73 and Asn69. However, two of these residues are not conserved in CsaA proteins: Met85 (Leu83 in BsCsaA) and Glu45 (Ile43 in BsCsaA). Mutating these residues in *E. coli* TRBP111 reduced the binding affinity of $tRNA^{Met}$ 66-fold and eightfold, respectively (Swairjo *et al.*, 2000). It has been proposed that CsaA may bind dual substrates: preproteins and tRNA (Kawaguchi *et al.*, 2001). While it is possible that CsaA is capable of binding tRNA owing to its structural and sequence similarities to other tRNA-binding proteins, the tRNA-binding ability of CsaA has not yet been demonstrated.

3.5. The electrostatics and conservation analysis

Electrostatics analysis of the protein surface revealed two prominent regions of electrostatic surface potential in the vicinity of the binding cavity (Fig. 3*d*). These two areas of positive and negative electrostatic surface potential flank the opposite sides of the binding cavity. The negative surface potential occurs near the entrance to the binding cavity and is formed by Asp5, Asp6, Glu8, Asp11 and the C-terminal carboxylate (Gly110). This negative surface potential is consistent with the proposed preference of CsaA to bind positively charged peptides (Linde *et al.*, 2003). An area of positive electrostatic surface potential arises owing to a cluster of basic residues at the ridge surrounding the binding site: Arg27, Arg73, Arg74, Lys32, Lys44 and Lys79.

The electrostatics in the vicinity of the binding site differs somewhat in BsCsaA and TtCsaA (Figs. 3*c* and 3*d*). The area of negative electrostatic potential is weaker in TtCsaA than in BsCsaA and its location is shifted. This is a consequence of the replacement of Asp6 and Glu8 with Ala and Gln, respectively, in TtCsaA. BsCsaA, on the other hand, contains a lysine at

position 52 and a glycine at position 90 instead of the glutamic acids in TtCsaA. These replacements are responsible for the different positions of negative surface potentials in the vicinity of the binding sites of BsCsaA and TtCsaA.

The analysis of the BsCsaA surface colored by the conservation score (Fig. 3*b*) reveals that the following residues are highly variable: Ile4′, Asp5′, Glu8′, Lys52, Ile88, Gly90, Gln91, Asp93 and Gly110′. It is interesting to note that these variable residues occur in a region that overlaps the area of negative electrostatic surface potential at the entrance to the binding site in BsCsaA. This region has a different pattern of electrostatic potential in TtCsaA. Based on the high sequence variability of this region, it is possible that each CsaA protein has its own unique pattern of electrostatic surface potential in the vicinity of the binding-site entrance.

4. Conclusion

CsaA is a small dimeric protein that is present in some species of Gram-negative and Gram-positive eubacteria and archaea. The available biochemical data indicate that CsaA may act as a chaperone in the Sec-dependent protein-secretion system. The dimeric structure of *B. subtilis* CsaA is held together by 21 hydrogen bonds that are mostly localized in the C-terminus. 19 of the hydrogen bonds are the same in both the $P3_221$ and $P42_12$ structures, while two hydrogen bonds are unique to each structure. Analysis of the proposed substrate-binding site reveals that it is mostly hydrophobic, with several residues forming a hydrophilic patch that may allow binding of unfolded peptides in an extended conformation. One wall of the proposed binding cavity appears to be flexible, which may allow CsaA to bind a broad spectrum of unfolded pre-protein substrates. The presence of an area of negative surface potential near the entrance to the binding site is correlated with the preference of CsaA to bind positively charged peptides. A region of negative electrostatic surface potential at the entrance to the binding site in BsCsaA contains residues that are highly variable among the sequences of CsaA, TRBP111 and the C-terminal regions of MetRS.

This work was supported in part by a Canadian Institute of Health Research operating grant (to MP), a National Science and Engineering Research Council of Canada operating grant (to MP), a Michael Smith Foundation for Health Research Scholar award (to MP) and a Canadian Foundation of Innovation grant (to MP).

References

Arcus, V. (2002). *Curr. Opin. Struct. Biol.* **12**, 794–801.
Binkowski, T. A., Naghibzadeh, S. & Liang, J. (2003). *Nucleic Acids Res.* **31**, 3352–3355.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Crane, J. M., Suo, Y., Lilly, A. A., Mao, C., Hubbell, W. L. & Randall, L. L. (2006). *J. Mol. Biol.* **363**, 63–74.
Crepin, T., Schmitt, E., Blanquet, S. & Mechulam, Y. (2002). *Biochemistry*, **41**, 13003–13011.
Dekker, C., de Kruijff, B. & Gros, P. (2003). *J. Struct. Biol.* **144**, 313–319.
DeLano, W. L. (2002). *The PyMOL Molecular Visualization System.* DeLano Scientific, San Carlos, CA, USA. http://www.pymol.org.
Driessen, A. J., Manting, E. H. & van der Does, C. (2001). *Nature Struct. Biol.* **8**, 492–498.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Gibrat, J. F., Madej, T. & Bryant, S. H. (1996). *Curr. Opin. Struct. Biol.* **6**, 377–385.
Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E. & Ben-Tal, N. (2003). *Bioinformatics*, **19**, 163–164.
Gouet, P., Robert, X. & Courcelle, E. (2003). *Nucleic Acids Res.* **31**, 3320–3323.
Hooft, R. W., Sander, C. & Vriend, G. (1996). *Proteins*, **26**, 363–376.
Hyyrylainen, H. L., Sarvas, M. & Kontinen, V. P. (2005). *Appl. Microbiol. Biotechnol.* **67**, 389–396.
Jones, S. & Thornton, J. M. (1995). *Prog. Biophys. Mol. Biol.* **63**, 31–65.
Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.
Kawaguchi, S., Muller, J., Linde, D., Kuramitsu, S., Shibata, T., Inoue, Y., Vassylyev, D. G. & Yokoyama, S. (2001). *EMBO J.* **20**, 562–569.
Keyzer, J. de, van der Does, C. & Driessen, A. J. (2003). *Cell Mol. Life Sci.* **60**, 2034–2052.
Kunst, F. *et al.* (1997). *Nature (London)*, **390**, 249–256.
Linde, D., Volkmer-Engert, R., Schreiber, S. & Muller, J. P. (2003). *FEMS Microbiol. Lett.* **226**, 93–100.
McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). *Acta Cryst.* D**61**, 458–464.
Maiti, R., Van Domselaar, G. H., Zhang, H. & Wishart, D. S. (2004). *Nucleic Acids Res.* **32**, W590–W594.
Morris, A. L., MacArthur, M. W., Hutchinson, E. G. & Thornton, J. M. (1992). *Proteins*, **12**, 345–364.
Muller, J. P., Bron, S., Venema, G. & van Dijl, J. M. (2000). *Microbiology*, **146**, 77–88.
Muller, J. P., Ozegowski, J., Vettermann, S., Swaving, J., Van Wely, K. H. & Driessen, A. J. (2000). *Biochem. J.* **348**, 367–373.
Muller, J., Walter, F., van Dijl, J. M. & Behnke, D. (1992). *Mol. Gen. Genet.* **235**, 89–96.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Murzin, A. G. (1993). *EMBO J.* **12**, 861–867.
Painter, J. & Merritt, E. A. (2006). *Acta Cryst.* D**62**, 439–450.
Pflugrath, J. W. (1999). *Acta Cryst.* D**55**, 1718–1725.
Pohlschroder, M., Hartmann, E., Hand, N. J., Dilks, K. & Haddad, A. (2005). *Annu. Rev. Microbiol.* **59**, 91–111.
Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor, NY, USA: Cold Spring Harbor Laboratory Press.
Stephenson, K. (2005). *Mol. Membr. Biol.* **22**, 17–28.
Swairjo, M. A., Morales, A. J., Wang, C. C., Ortiz, A. R. & Schimmel, P. (2000). *EMBO J.* **19**, 6287–6298.
Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.
Xu, Z., Knafels, J. D. & Yoshino, K. (2000). *Nature Struct. Biol.* **7**, 1172–1177.
Yi, L. & Dalbey, R. E. (2005). *Mol. Membr. Biol.* **22**, 101–111.