0907-4449

Bacillus subtilis

*

8 2006
31 2007

B 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 *B* . This paper presents the cloning, purification, crystallization and structures of BsCsaA in two space groups ($4_2 1 2$ and $3_2 21$) 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 (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.

, $P3_2 21$, ; $P42_1 2$,

1.

The Sec-dependent protein-targeting and translocation pathway is universally conserved across all three domains of life (Pohlschroder ., 2005). In eubacteria, secreted proteins are synthesized in the cytosol as precursors carrying an amino-terminal signal peptide (Driessen ., 2001). These precursors are targeted to the translocation machinery at the cytosolic membrane. In , 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 ., 2003; Driessen ., 2001; Stephenson, 2005) and YidC (Yi & Dalbey, 2005). In bacteria, most of the protein secretion is carried out post-translationally (Pohlschroder ., 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 ., 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 *B* (Kunst ., 1997).

© 2007

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 *C* *C* software package (Pflugrath, 1999). The trigonal crystals (3_221) diffracted to beyond 2.0 Å resolution. The tetragonal crystals (42_12) diffracted to beyond 1.9 Å resolution. Complete data sets were collected for each crystal form. See Table 1 for data-collection statistics.

. .' t t

The structures of *B.* CsaA were solved by molecular replacement using the program (McCoy ., 2005). The coordinates of . CsaA chain *A* (1gd7A) were used as a search model. Several rounds of restrained refinement with */AC*5 (Murshudov ., 1997) and manual adjustment and manipulation using *C* (Emsley & Cowtan, 2004) were used to build the BsCsaA models. *C* (Brünger ., 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 */AC*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 *C* (Morris ., 1992).

. .' t t

Superimpositions were carried out using the program (Maiti ., 2004). The surface/binding-pocket analysis was carried out using *CA* (Binkowski ., 2003) with a 1.4 Å probe radius. The mapping of the sequence conservation onto the three-dimensional structure was performed using *C* (Glaser ., 2003). The figures were produced using (DeLano, 2002). The sequence-alignment analysis (Fig. 1) was prepared using *C* (Thompson ., 1994) and v.2.2 (Gouet ., 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 .

. t

.1.' t

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 ., 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. 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 *A* 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 *B* and *C* .

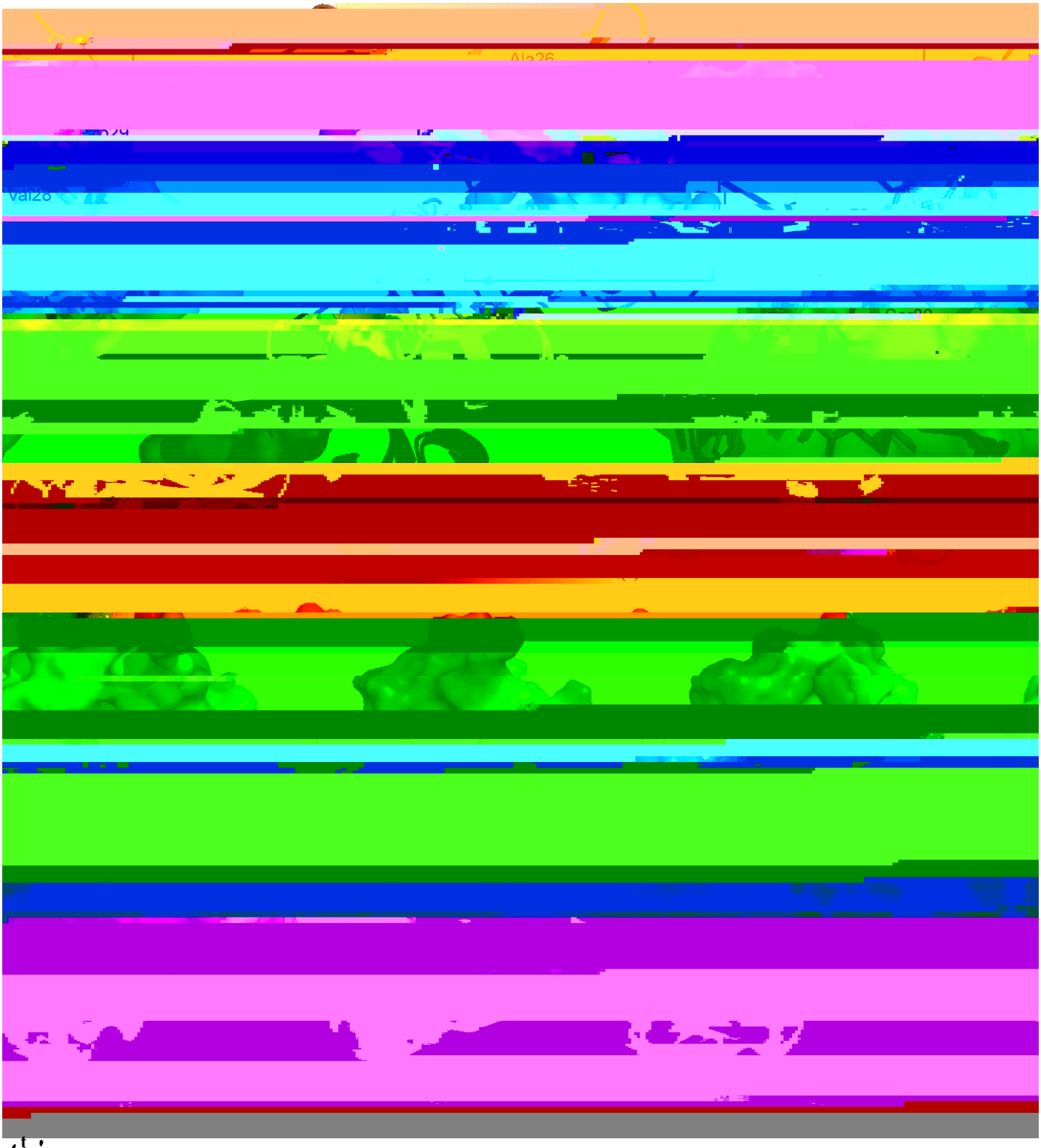
. .' t t

CsaA from *B.* (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

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.

The two structures of BsCsaA differ somewhat, particularly in the positions of the atoms in residues 23–32 and 73–79 (Fig. 2). 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 (3_221) superimpose over the backbone

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.



( ) 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 42_12, chain *A*. ( ) 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 *C* (Glaser ., 2003). The conservation scores were obtained from sequence alignment of 36 CsaA, TRBP and MetRS (C-terminal domains only) sequences. ( ) The protein surface electrostatics of CsaA. ( ) The protein surface electrostatics of *B.* CsaA. Areas colored white, red and blue correspond to neutral, negative and positive surface electrostatic potentials, respectively.

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) 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.

.. .t

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.* 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 3_221 and 42_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).

, . ; , .,

Arcus, V. (2002). *C . . . B .* **12**, 794–801.
Binkowski, T. A., Naghibzadeh, S. & Liang, J. (2003). *A . .* **1**, 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). *A C . D* , 905–921.
Crane, J. M., Suo, Y., Lilly, A. A., Mao, C., Hubbell, W. L. & Randall, L. L. (2006). *. . B .* , 63–74.
Crepin, T., Schmitt, E., Blanquet, S. & Mechulam, Y. (2002). *B* , **1**, 13003–13011.
Dekker, C., de Kruijff, B. & Gros, P. (2003). *. . B .* **1** , 313–319.
DeLano, W. L. (2002). . DeLano Scientific, San Carlos, CA, USA. http://www.pymol.org.
Driessen, A. J., Manting, E. H. & van der Does, C. (2001). *. . B .* , 492–498.
Emsley, P. & Cowtan, K. (2004). *A C . D* **0**, 2126–2132.
Gibrat, J. F., Madej, T. & Bryant, S. H. (1996). *C . . . B .* , 377–385.
Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E. & Ben-Tal, N. (2003). *B* , **1** , 163–164.
Gouet, P., Robert, X. & Courcelle, E. (2003). *A* . **1**, 3320–3323.
Hooft, R. W., Sander, C. & Vriend, G. (1996). , **2** , 363–376.
Hyyrylainen, H. L., Sarvas, M. & Kontinen, V. P. (2005). *A . . B* , 389–396.
Jones, S. & Thornton, J. M. (1995). *. B . . B .* , 31–65.
Kabsch, W. & Sander, C. (1983). *B* , **22**, 2577–2637.
Kawaguchi, S., Muller, J., Linde, D., Kuramitsu, S., Shibata, T., Inoue, ,