STRUCTURAL BIOLOGY COMMUNICATIONS

ISSN 2053-230X

Structural insights into the substrate specificity of SP_0149, the substrate-binding protein of a methionine ABC transporter from *Streptococcus pneumoniae*

Bhavya Jha,[a]‡ Rajan Vyas,[a]‡§ Jaya Bhushan,[b]¶ Devinder Sehgal[b] and Bichitra Kumar Biswal[a]*

[a]Structural and Functional Biology Laboratory, National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi 110 067, India, and [b]Molecular Immunology Laboratory, National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi 110 067, India. *Correspondence e-mail: bbiswal@nii.ac.in

Received 24 April 2019
Accepted 24 June 2019

Edited by I. Tanaka, Hokkaido University, Japan

‡ These authors contributed equally to this work.
§ Present address: Department of Biochemistry and Microbiology, Central University of Punjab, City Campus, Mansa Road, Bathinda 151 001, India.
¶ Present address: Department of Molecular Microbiology, Washington University School of Medicine in St Louis, 660 South Euclid Avenue, St Louis, Missouri, USA.

Keywords: *Streptococcus pneumoniae*; ABC transporter; substrate-binding protein; methionine; X-ray crystallography.

PDB reference: crystal structure of the substrate binding protein of a methionine transporter from *Streptococcus pneumoniae*, 6jf1

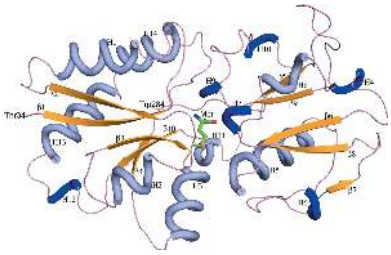
Supporting information: this article has supporting information at journals.iucr.org/f

Successful pathogenesis is a cumulative effect of the virulence factors of a pathogen and its capability to efficiently utilize the available nutrients from the host. *Streptococcus pneumoniae*, a Gram-positive opportunistic pathogen, may either reside asymptomatically as a nasopharyngeal commensal inside the human host or cause lethal diseases, including pneumonia, meningitis and sepsis. *S. pneumoniae* is known to acquire methionine (Met) from its host through a Met importer. Here, the crystal structure of the substrate-binding protein (SBP; SP_0149) of an ABC importer with Met bound is reported at a resolution of 1.95 Å. The three-dimensional structure of SBP shows that it is composed of two distinct domains, each consisting of a mixed β-sheet flanked by helices. The substrate, Met, is bound in the central part of the interface between the two domains. The overall structure of SP_0149 resembles those of SBPs from other reported bacterial Met and Gly-Met dipeptide transporters. However, a detailed analysis of these structures shows notable variations in the amino-acid composition of the substrate-binding pockets of the SP_0149–Met and GmpC–Gly-Met structures. In particular, SP_0149 harbors Thr212 and Tyr114, whereas the corresponding residues in GmpC are Gly and Val. This difference is likely to be the underlying basis for their differential substrate specificity. In summary, the structure of the SP_0149–Met complex provides insights into the transport function of SP_0149 and its interactions with methionine. It opens up avenues for the rational design of inhibitors of SP_0149 through a structure-mediated approach.

1. Introduction

Streptococcus pneumoniae (also called pneumococcus) is the etiological agent of pneumonia, bacteremia, meningitis and middle-ear infection. Pneumococcus is the leading cause of morbidity and mortality worldwide. Children <5 years of age, the elderly (>65 years) and immunocompromised individuals are the most likely to develop pneumococcal infections (Henriques-Normark & Tuomanen, 2013). It has been estimated that pneumococcus was responsible for the deaths of 294 000 (uncertainty interval 192 000–366 000) children of <5 years of age who were not infected with HIV in 2015. The majority of these deaths occurred in developing countries such as India and Nigeria (Wahl *et al.*, 2018).

The current commercially licensed pneumococcal vaccines (*i.e.* 23-valent polysaccharide vaccine and polyvalent glyco-conjugate vaccines) are based on the capsular polysaccharide. The introduction of pneumococcal vaccines has led to a decrease in the prevalence of vaccine serotypes (Croucher *et*



al., 2018). The replacement of serotypes that are a part of current pneumococcal vaccines with nonvaccine serotypes has been reported (Ciruela *et al.*, 2018; Sakata *et al.*, 2018; Usuf *et al.*, 2018). The treatment of pneumococcal infections has become a global public health problem owing to the emergence of multidrug-resistant strains and the replacement of vaccine serotypes with nonvaccine serotypes. There is an urgent need to make a global effort to discover and develop new drugs and approaches against multidrug-resistant respiratory bacterial pathogens such as *S. pneumoniae* (Farha & Brown, 2013; Guitor & Wright, 2018).

The pneumococcal cell surface is decorated with various protein and polysaccharide components that have been shown to be involved in virulence and pathogenesis (Kadioglu *et al.*, 2008). The ATP-binding cassette (ABC) transporter superfamily is one of the largest classes of transporters. ABC transporters have diverse molecular architectures. They are typically composed of an external substrate-binding protein (SBP), a transmembrane domain containing a permease, and a cytoplasmic nucleotide-binding domain (ATPase). ABC transporters utilize ATP to translocate a variety of substrates into (importers) or out of (exporters) the cell. Importers are present in prokaryotes but are absent in eukaryotes. Exporters are present in both prokaryotes and eukaryotes. The SBPs of Gram-negative bacteria are present in the periplasm, whereas they are covalently anchored to the outer cell membrane via a lipid moiety in the case of Gram-positive bacteria (Rees *et al.*, 2009). The size of SBPs varies from 25 to 70 kDa and they exhibit little similarity at the primary-sequence level (Berntsson *et al.*, 2010). SBPs of importers are thought to be responsible for the specificity and directionality of substrate translocation as they are absent in exporters. The SBPs of ABC importers take up essential substrates and have been demonstrated to be involved in the pathogenesis, virulence and fitness of bacterial pathogens (Kohler *et al.*, 2016). The surface location of SBPs and the fact that they are present only in prokaryotes and are absent in the host makes SBPs good candidates for the development of new drugs against bacterial pathogens (Garmory & Titball, 2004). An understanding of the molecular architecture and mechanism of action would help in the rational design of novel post-infection therapeutics and strategies (Lewinson & Livnat-Levanon, 2017). Knowledge regarding the substrate specificities of SBPs can potentially be exploited either to deliver a lethal substrate mimic or to inhibit the binding of the substrate to the SBP (Tanaka *et al.*, 2018).

We have previously reported the crystallization of SP_0149, the SBP of an ABC transporter from *S. pneumoniae* (Bhushan *et al.*, 2011). SP_0149 is highly conserved in all publically available pneumococcal genomes (>97% identity at the protein-sequence level; Basavanna *et al.*, 2013). Biochemical studies indicated that SP_0149 has a high specificity for L-methionine (Met) and is required for the uptake of Met. The virulence of a pneumococcal mutant that lacked *sp_0149* was compromised in a mouse sepsis model (Basavanna *et al.*, 2013). Here, we report the crystal structure of SP_0149 at 1.95 Å resolution. The three-dimensional structure reveals the presence of Met bound in the substrate-binding site of SP_0149. The detailed architecture of the molecule, including the Met-binding pocket, and a possible structural explanation of its substrate specificity form the theme of this study.

Table 1
Data-collection and processing statistics.

Values in parentheses are for the highest resolution shell.

Diffraction source	Rigaku FR-E+ SuperBright microfocus rotating copper-anode generator
Wavelength (Å)	1.54178
Space group	$P2_12_12_1$
Unit-cell parameters (Å, °)	a = 54.7, b = 75.8, c = 75.6, $\alpha = \beta = \gamma$ = 90
Matthews coefficient (Å^3 Da^{-1})	2.58
Subunits per asymmetric unit	1
Solvent content (%)	52.26
Temperature (K)	100
Detector	Rigaku R-AXIS IV^{++}
Diffraction limit (Å)	35.00–1.95 (2.02–1.95)
$R_{\rm merge}$ (%)†	12.9 (80.3)
Mean $I/\sigma(I)$	10.43 (2.02)
No. of unique reflections	22518 (1643)
Average multiplicity	3.6 (2.3)
Completeness (%)	96.0 (71.7)

† $R_{\rm merge}(I) = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$ for i observations of a given reflection hkl. $\langle I(hkl)\rangle$ is the average intensity of the i observations.

2. Materials and methods

2.1. Cloning, expression, purification and crystallization

The details of the cloning, expression, purification and crystallization of SP_0149 have been reported elsewhere (Bhushan *et al.*, 2011). Briefly, the *sp_0149* gene was PCR-amplified using specific primers to obtain amplicons coding for amino acids 30–284 and was ligated into *Escherichia coli* expression vector pQE-30 Xa. Thus, the recombinant construct was N-terminally truncated by 29 residues. It was overexpressed in *E. coli* K-12 expression strain SG13009. The protein was purified using Ni–NTA affinity and gel-filtration chromatography. The recombinant protein was concentrated in a buffer consisting of 10 m*M* Tris pH 8.0, 50 m*M* NaCl and used for crystallization.

Diffraction-quality crystals of SP_0149 were grown using drops composed of the recombinant protein at 20 mg ml^{-1} and a precipitant solution consisting of 5% PEG 8000, 100 m*M* MES sodium salt pH 6.5, 200 m*M* zinc acetate in a 1:2.5 protein:precipitant solution ratio; the drops were equilibrated against 1 ml reservoir solution. Complete intensity data sets were collected at 100 K with Cu $K\alpha$ radiation using the in-house X-ray diffraction facility: a Rigaku FR-E+ SuperBright microfocus rotating-anode dual-wavelength X-ray generator mounted with an R-AXIS IV^{++} detector. The best data set was indexed, integrated and scaled using *HKL*-2000 (Otwinowski & Minor, 1997). The data-collection and processing statistics are presented in Table 1.

2.2. Structure solution and refinement

The structure of recombinant SP_0149 was solved by molecular replacement using the structure of the Gly-Met substrate-binding protein of an ABC transporter from

Table 2
Refinement statistics.

R_{work} (%)	18.3
R_{free} (%)	23.1
No. of reflections in test set	1193
B factors (Å^2)	
All atoms	30.35
Proteins	28.74
Ligands	39.71
Water	44.26
Wilson *B* factor (Å^2)	29.53
R.m.s.d.	
Bond angles (°)	1.5638
Bond lengths (Å)	0.0119
Ramachandran statistics (%)	
Allowed regions	96.85
Moderately allowed regions	2.76
Outliers	0.39
PDB code	6jf1

Staphylococcus aureus (Williams *et al.*, 2004; PDB entry 1p99) in *Phaser* (McCoy *et al.*, 2007) from the *CCP*4 suite (Winn *et al.*, 2011). Initially, the structure was subjected to 50 cycles of rigid-body and simulated-annealing refinement using the structure-refinement program *REFMAC* (Murshudov *et al.*, 2011) employing the maximum-likelihood method. At this stage, electron density was computed using the observed structure factors and calculated phases for the search model. Amino acids specific to SP_0149 were incorporated/replaced in the polypeptide chain based on electron-density maps ($2|F_o| - |F_c|$ at 1σ and $|F_o| - |F_c|$ at 3σ contour levels) using the model-building program *Coot* (Emsley & Cowtan, 2004). The structure was refined and rebuilt iteratively. Positional and individual *B*-factor refinements were carried out after every cycle of model building. Many parts of the structure showed significant structural differences from the search model; therefore, extensive model building was performed. Bulk-solvent correction and anisotropic *B*-factor scaling were incorporated during the refinement. Water molecules, substrate, zinc, acetate, chloride and glycerol were modelled based on the electron density ($2|F_o| - |F_c|$ at 1σ and $|F_o| - |F_c|$ at 3σ contour levels). Residues included in the hexahistidine affinity tag that came from the vector as part of the recombinant construct were also modelled in the structure. The stereochemistry of each and every residue of the structure was checked using the *PROCHECK* (Laskowski *et al.*, 1993). The R_{work} and R_{free} values of the final refined model were 18.3% and 23.1%, respectively. The refinement statistics are presented in Table 2. The secondary-structural elements were assigned using *DSSP* (Kabsch & Sander, 1983). Figs. 1–3 and 5 were prepared using *PyMOL* (v1.8; Schrödinger). An alignment file for preparing Fig. 4 was obtained using *Clustal Omega* (Sievers & Higgins, 2014), while the image was prepared in *ESPript*3 (Robert & Gouet, 2014).

3. Results and discussion

3.1. The three-dimensional structure of SP_0149

The recombinant SP_0149, with 29 residues truncated from the N-terminus, consists of amino-acid residues 30–284. The refined crystal structure of SP_0149 shows that its tertiary structure has an ellipsoidal shape with a surface area of ∼11 608 Å^2 as calculated using the *PISA* server (Krissinel & Henrick, 2007). The protein model consists of residues 34–284; residues 30–33 could not be traced in the electron-density map [Fig. 1(*a*)]. It possesses 14 helices and ten β-strands, and 22 connecting loops. Of the 14 helices, eight are α-type and six are 3_{10}-type [Fig. 1(*a*)]. The entire tertiary structure consists of two distinct domains, A and B, which are represented in different colors in Fig. 1(*b*). Domain A containing two amino-acid stretches (34–115 and 234–284) folds into a central five-stranded ($\beta\uparrow\beta\uparrow\beta\uparrow\beta\downarrow\beta\uparrow$) mixed twisted β-sheet flanked by three short helices (H2, H3 and H11) on one side and three relatively longer helices (H1, H13 and H14) on the opposite side [Fig. 1(*b*)]. These two groups of helices are connected by a short 3_{10}-helix (H12). On the other hand, domain B consists of one continuous stretch of amino acids (116–233) and folds into a globular shape comprising a five-stranded mixed

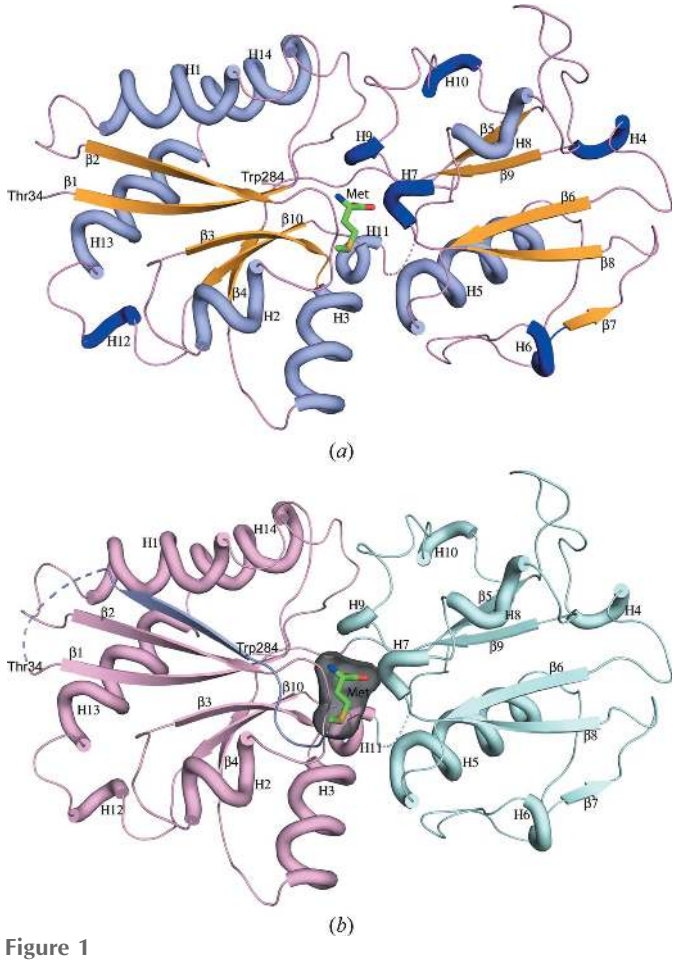


Figure 1
(*a*) Cartoon representation of the three-dimensional structure of SP_0149. α-Helices are depicted in light blue, while 3_{10}-helices are in dark blue. β-Sheets and loops are presented in bright orange and pink, respectively. Helices and β-sheets are numbered. The first (Thr34) and the last (Trp284) amino-acid residues of SP_0149 observed in the electron density are marked. (*b*) The monomer consists of two domains, A and B, which are represented in light pink and pale cyan, respectively. The initial stretch of residues from the vector are shown in purple. The bound Met is shown in stick representation in a surface cavity.

($\beta\uparrow\beta\downarrow\beta\downarrow\beta\downarrow\beta\uparrow$) β-sheet flanked by one short helix (H8) and four 3_{10}-helices (H4, H7, H9 and H10) on one side, and one long helix (H5) and one 3_{10}-helix (H6) on the other side [Fig. 1(*b*)]. A large number of loops, each connecting two secondary-structural elements, account for about 40% of the total amino acids. These loops are exposed to the solvent and are likely to provide the necessary structural flexibility of the molecule that is required for its function.

3.2. Mapping the Met-binding site

No exogenous Met was added during the purification and crystallization; however, unambiguous electron density for bound Met was observed in the cleft between the domains [Fig. 2(*a*)]. This finding clearly indicates that the physiological substrate of SP_0149 is Met. The electron density clearly shows that the amino and carboxylic groups of Met are positioned towards the hinge region, whereas its side chain is oriented towards helix H3. The tip of the β3 strand provides a platform for positioning the side chain of the Met substrate [Fig. 2(*a*)]. Computation of the electrostatic potential surface shows that charge complementarity assists the favorable binding of Met to SP_0149 [Fig. 2(*b*)]. The residues that line the substrate-binding pocket within a distance of 4 Å from Met are Tyr73, Phe90, Gln91, His92, Phe95, Tyr114 and Asn239 of domain A, and Asn151, Arg154, Asn210, Asn211 and Thr212 of domain B [Fig. 3(*a*)]. Notably, with the exceptions of Gln91 and Asn239, all of the residues that protrude from domain A are aromatic. The substrate makes noncovalent interactions such as hydrogen bonds and van der Waals contacts with SP_0149 (Table 3). The hydrogen-bonding interactions are depicted in Fig. 3(*b*).

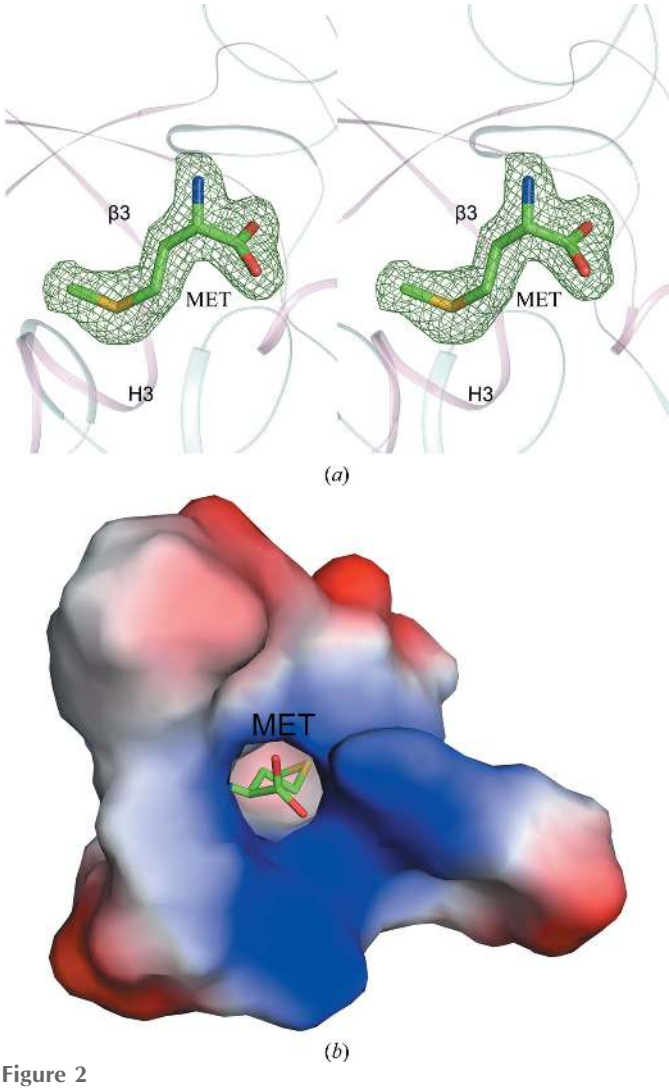


Figure 2
(*a*) An unbiased OMIT $|F_o| - |F_c|$ map at 3σ indicated the presence of Met, which was eventually satisfied by adding Met to the model. The final refined Met is superimposed on the electron-density map; it is shown in stick representation and represented in stereo mode. (*b*) The electrostatic potential surface of the Met-binding pocket, with Met shown in stick representation, shows that charge complementarity exists in this region.

3.3. Comparison of the structure of SP_0149 with those of other SBPs

3.3.1. Overall structure. To compare the similarities and the differences between reported structures of SBPs from various bacterial Met ABC transporters, a search for proteins that are

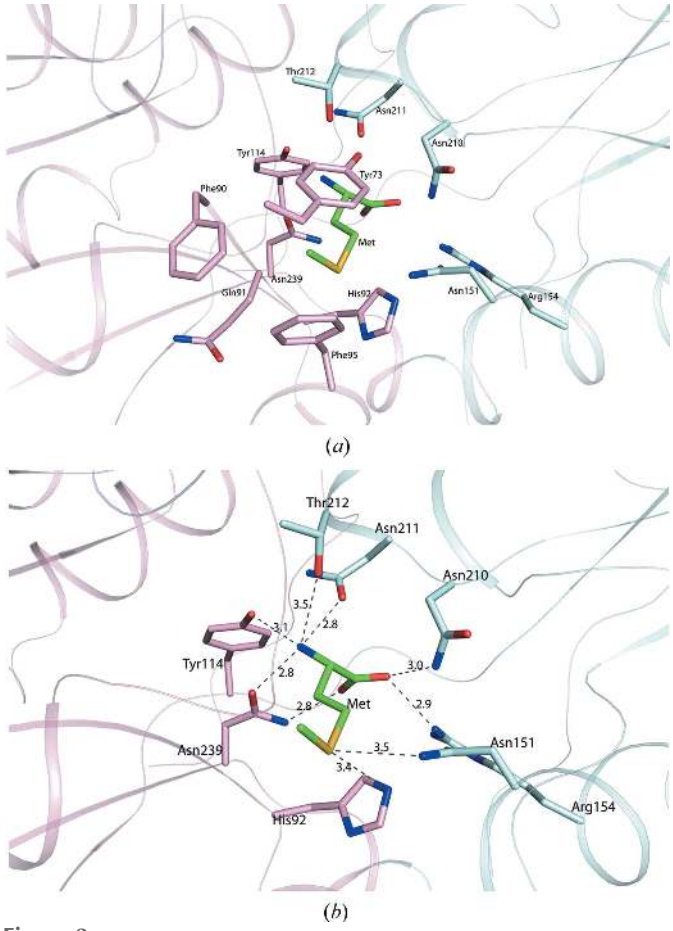


Figure 3
(*a*) The substrate-binding pocket of SP_0149 is shown. The residues lining the pocket are represented as sticks. Both domain A (light pink) and domain B (light cyan) contribute to forming the pocket. (*b*) Hydrogen-bonding interactions of SP_0149 with its natural substrate (Met) are depicted. The color code of the monomers is the same as in (*a*).

Table 3
Interactions (distance cutoff of 4 Å) of binding-pocket residues with the substrate Met.

vdW, van der Waals contact.

Source atom	Target atom	Distance (Å)	Type of interaction
Met C	Asn210 ND2	3.92	vdW
	Asn211 OD1	3.30	vdW
	Arg154 NH1	3.76	vdW
	His92 CD2	3.94	vdW
	Asn239 ND2	3.87	vdW
Met OXT	Asn211 OD1	3.68	vdW
	Asn239 OD1	3.57	vdW
	Asn239 ND2	2.81	Hydrogen bond
	Asn239 CG	3.61	vdW
	His92 CD2	3.98	vdW
Met N	Tyr114 OH	3.07	Hydrogen bond
	Asn211 CG	3.73	vdW
	Asn211 OD1	2.77	Hydrogen bond
	Asn211 ND2	3.96	vdW
	Thr212 OG1	3.52	Hydrogen bond
	Phe90 O	3.91	vdW
	Asn239 OD1	2.83	Hydrogen bond
	Asn239 ND2	3.92	vdW
	Asn239 CG	3.74	vdW
Met O	Asn210 CG	3.95	vdW
	Asn210 ND2	2.97	Hydrogen bond
	Asn211 OD1	3.64	vdW
	Arg154 CZ	3.95	vdW
	Arg154 NH1	2.88	Hydrogen bond
Met CA	Tyr73 OH	3.59	vdW
	Asn211 OD1	3.22	vdW
	Thr212 OG1	3.71	vdW
	Asn239 OD1	3.70	vdW
Met CB	Tyr73 OH	3.90	vdW
	Phe90 O	3.38	vdW
	Gln91 CA	3.98	vdW
	His92 CD2	3.78	vdW
	Asn239 OD1	3.63	vdW
Met CG	Asn151 OD1	3.96	vdW
	Asn210 ND2	3.71	vdW
	Asn151 ND2	3.75	vdW
	Tyr73 CZ	3.69	vdW
	Tyr73 CE2	3.82	vdW
	Tyr73 OH	3.62	vdW
	His92 CD2	3.44	vdW
	His92 NE2	3.54	vdW
Met SD	Phe95 CG	3.63	vdW
	Phe95 CD2	3.72	vdW
	Asn151 ND2	3.49	Hydrogen bond
	Gln91 CA	3.75	vdW
	His92 CE1	3.71	vdW
	Phe95 CB	3.50	vdW
	His92 CD2	3.55	vdW
	His92 CG	3.92	vdW
	His92 ND1	3.98	vdW
	His92 NE2	3.39	Hydrogen bond
	His92 N	3.68	vdW
Met CE	Tyr73 CD1	3.50	vdW
	Tyr73 CE1	3.66	vdW
	Tyr73 CG	3.51	vdW
	Phe95 CG	3.67	vdW
	Phe95 CD2	3.71	vdW
	Tyr73 CZ	3.76	vdW
	Tyr73 CE2	3.72	vdW
	Tyr73 CD2	3.59	vdW
	Phe90 CE2	3.81	vdW
	Phe95 CD1	3.98	vdW
	Phe90 CD2	3.84	vdW
	Phe90 O	3.63	vdW
	Gln91 CB	3.66	vdW
	Gln91 CA	3.61	vdW

structurally related to SP_0149 (>30% sequence identity) was performed using the protein structure comparison service *PDBeFold* at the European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/ssm; Krissinel & Henrick, 2004). Four proteins with a *Z*-score of >15 were identified (Table 4). This search employing *PDBeFold* provided us with *Z*-score, r.m.s.d., $N_{\rm align}$ and $\%_{\rm seq}$ values, details and definitions of which are given in Table 4. Importantly, the overall r.m.s.d. between the structures of SP_0149 and of the model of GmpC from *S. aureus* used for molecular replacement (PDB entry 1p99) is 1.29 Å (233 C^{α} atoms). All of the proteins mentioned in Table 4 bind Met or selenomethionine (SeMet), except for GmpC which binds the Gly-Met dipeptide, as reported in the

Table 4
Structural homologs of SP_0149.

Z-score†	R.m.s.d.‡ (Å)	$N_{\rm align}$§	$\%_{\rm seq}$¶	Bound substrate	PDB code	Protein name and description	Source
17.8	1.09	248	63	SeMet	4q5t	AtmB (a putative membrane lipoprotein)	*Streptococcus mutans*
16.7	1.29	233	39	Gly-Met	1p99	Membrane-associated lipoprotein-9 GmpC that binds the dipeptide Gly-Met	*Staphylococcus aureus* subsp. *aureus*
15.2	1.36	223	41	Met	4qhq	A nutrient-binding protein (bound to methionine in the reported structure)	*Burkholderia cenocepacia*
15.0	1.59	224	36	Met	3k2d	Immunogenic lipoprotein A	*Vibrio vulnificus*

† The *Z*-score measures the statistical significance of a match in terms of Gaussian statistics. If the quality of matching is considered a random value for randomly picking structures from a sufficiently large database it should obey the Gaussian distribution, with the probability that the same or a better quality match may be found expressed as $P = \mathrm{erfc}(Z/2)$. If two structures are matched uniquely with each other, then *P* equals the *P*-value. The higher the *Z*-score, the higher the statistical significance of the match. ‡ R.m.s.d. stands for the root-mean-square deviation calculated between C^{α} atoms of matched residues at the best three-dimensional superposition of the query and target structures. In simple words, the r.m.s.d. gives an idea of how separated, at the best three-dimensional superposition, a 'typical' pair of matched C^{α} atoms is. Generally, the larger the r.m.s.d., the more distant the matched structures are. § The length of alignment $N_{\rm align}$ or number of matched residues is calculated at the best three-dimensional superposition of the query and target structures. The residues are aligned in three dimensions on the basis of their spatial closeness. The orientation of the structures is optimized such as to minimize the r.m.s.d. and to maximize the number of aligned residues. ¶ The sequence identity $\%_{\rm seq}$ is a quality characteristics of C^{α} alignment. It is calculated as the fraction of pairs of identical residues among all aligned: $\%_{\rm seq} = N_{\rm ident}/N_{\rm align}$. The program *SSM* reports sequence identity as a percentage. Thus, if two identical chains are aligned such that only identical residues are mapped onto each other then $\%_{\rm seq} = 100$. Sequence identity in *SSM* is calculated from structure (three-dimensional), rather than sequence, alignment. Therefore, two almost identical sequences may be estimated to have low sequence identity if they fold into slightly different structures.

respective crystal structures. A comparison of the structural homologs of SP_0149 is given in Table 5. The resulting r.m.s.d. values suggest that the backbone structures of all SBPs have essentially the same conformation (Table 5).

A sequence alignment was performed of the recombinant SP_0149 protein and its homologs mentioned in Tables 4 and 5, the sequences of which were obtained from the corresponding PDB files. The alignment file obtained using *Clustal Omega* was used as an input to *ESPript*3 and an alignment image was obtained (Robert & Gouet, 2014; Fig. 4). It was observed that the residues belonging to the substrate-binding pocket of SP_0149 (73, 90, 91, 92, 95, 114, 151, 154, 210, 211, 212 and 239) were largely conserved in its homologs (Fig. 4). In particular, residues 73, 90, 91, 92, 151, 154, 210 and 239 were completely conserved. Superimposition of the structures of SP_0149, GmpC (the MR model) and AtmB (the closest structurally characterized homolog of SP_0149; PDB entry 4q5t; Joint Center for Structural Genomics, unpublished work) showed that the overall three-dimensional structures of these binding proteins are mostly similar; however, some differences are observed in the loop regions, mainly owing to their flexibility [Fig. 5(*a*)].

3.3.2. Substrate-binding pocket. A detailed analysis of the corresponding substrate-binding pockets of SP_0149, GmpC

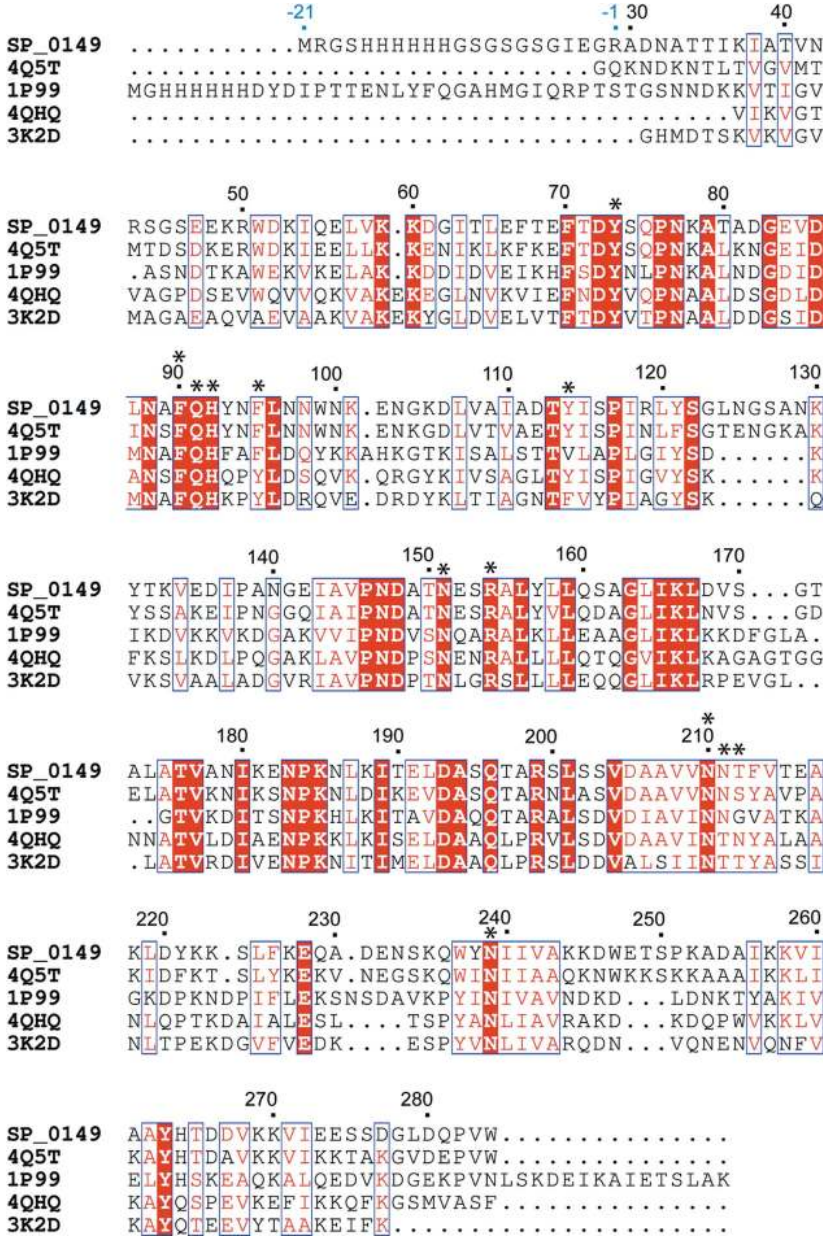


Figure 4
An *ESPript* representation of sequence alignment of SP_0149 with several of its homologs, showing the conservation of key residues, is depicted. The recombinant SP_0149 sequence was used in the process. The residues that come from the vector are numbered −21 to −1. Since the recombinant protein is truncated by 29 residues, the numbering of amino-acid residues belonging to the protein starts at 30 and they are numbered in intervals of ten. The residues that form the substrate-binding pocket in SP_0149 [73, 90, 91, 92, 95, 114, 151, 154, 210, 211, 212 and 239; marked with an asterisk (*) in the figure] are largely conserved in its homologs.

and AtmB led to some interesting findings. The position of Met in SP_0149 not only matches that of its counterpart SeMet in AtmB but also that of the Met of the Gly-Met dipeptide bound to GmpC [Figs. 5(*a*) and 5(*b*)].

The structures of SP_0149 and AtmB are largely similar [Fig. 5(*a*)], with Met and SeMet, respectively, bound to them. The major residues that interact with the substrate in the binding pocket are identical, with the exception of Ser208 in AtmB in place of Thr212 in SP_0149 [Figs. 5(*b*) and 5(*d*)]. The side chain of Asn207 in AtmB is positioned away from the substrate, in contrast to SP_0149, in which the corresponding Asn211 faces towards the substrate [Figs. 5(*b*) and 5(*d*)]. Overall, the interacting residues are closer to the substrate in

Table 5
Comparison of the structural homologs of SP_0149.

R.m.s.d. values (in Å) are presented in the upper diagonal boxes of the matrix, while the numbers of C^{α} atoms aligned are given in the lower diagonal boxes.

PDB code	6jf1 (SP_0149)	4q5t	1p99	4qhq	3k2d
6jf1 (SP_0149)	—	1.09	1.29	1.36	1.59
4q5t	248	—	1.29	1.39	1.56
1p99	233	231	—	1.18	1.30
4qhq	223	218	230	—	0.96
3k2d	224	221	228	227	—

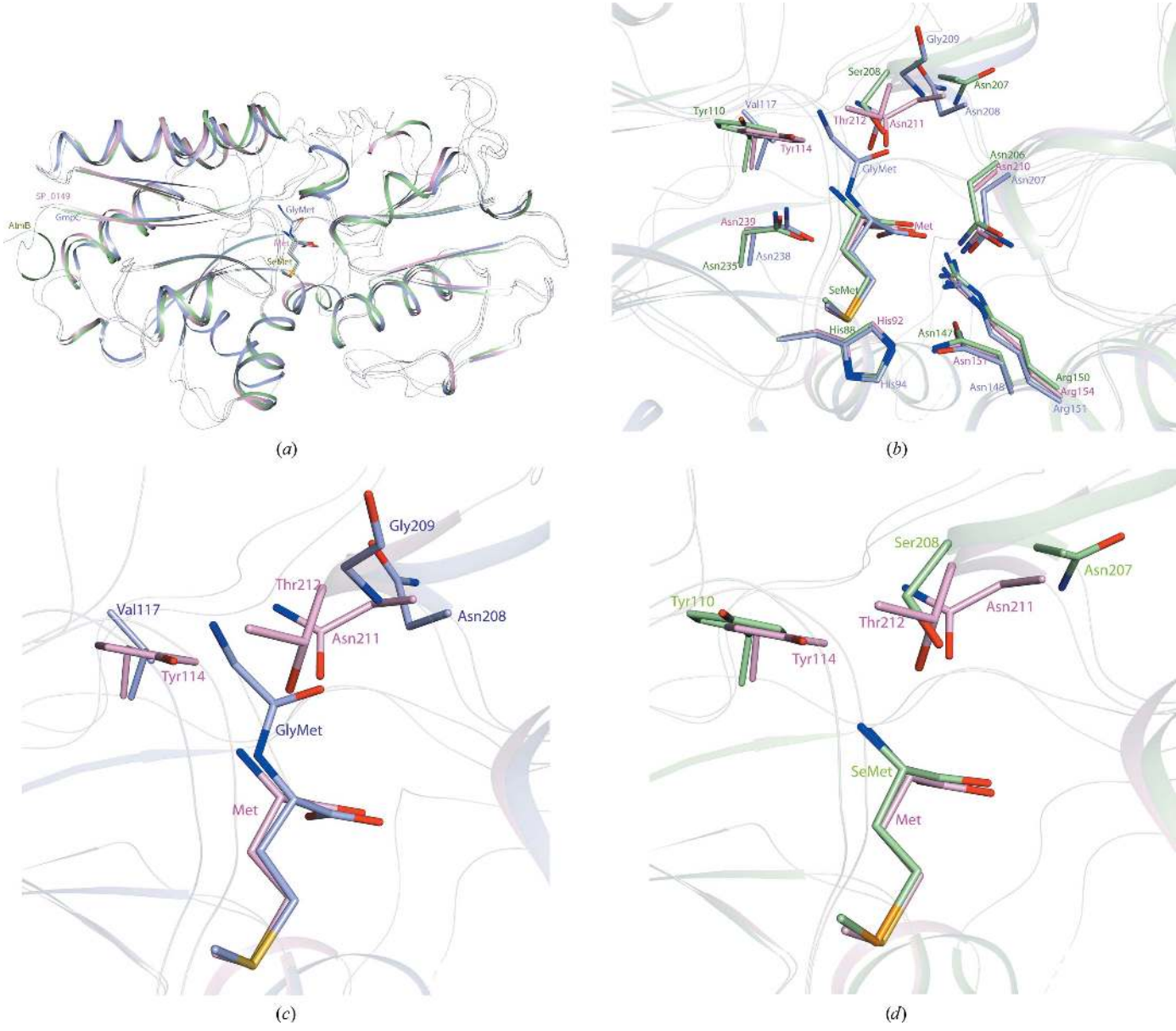


Figure 5
(*a*) Superimposition of the structures of SP_0149 from *S. pneumoniae*, GmpC from *S. aureus* and AtmB from *S. mutans*. SP_0149, GmpC and AtmB are represented in light pink, light purple and light green, respectively. The corresponding bound substrates are shown in stick representation. (*b*) Residues interacting with the corresponding bound substrates in SP_0149, GmpC and AtmB are depicted in stick representation. Entities belonging to SP_0149, GmpC and AtmB are represented in light pink, light purple and light green, respectively. Residues are labeled in the corresponding colors. (*c*, *d*) Interacting residues which differ in SP_0149 and GmpC are represented in (*c*), while those which have a different disposition in SP_0149 and AtmB are shown in (*d*). The color code is the same as in (*b*).

SP_0149 (Tyr114 and Asn211) than in AtmB (Tyr110 and Asn207), implying that Met binding might be more efficient in SP_0149 compared with SeMet binding in AtmB [Figs. 5(*b*) and 5(*d*)].

A few key differences were observed in the SP_0149 and GmpC structures. The Gly of the Gly-Met dipeptide in GmpC is positioned towards a helix corresponding to H9 of SP_0149 [Fig. 5(*a*)]. Importantly, the side chain of Asn208 in GmpC is shifted by an angle of around 90° from the corresponding residue Asn211 in SP_0149 [Figs. 5(*b*) and 5(*c*)]. The subsequent residue being a glycine (it is a threonine in the case of SP_0149) provides a high degree of conformational flexibility to this region and, along with the specific conformation of the side chain of Asn208, accounts for a spatially favorable environment for the dipeptide to bind [Figs. 5(*b*) and 5(*c*)]. The next residue is Val210 (Phe213 in SP_0149), which is most likely to make the helix even more flexible (not shown in the figure). Therefore, the conformational flexibility of this region of the molecule is perhaps the major determinant for GmpC to accommodate a dipeptide substrate. Notably, another residue in the substrate-binding pocket is altered. In this case, Tyr114 of SP_0149 protruding from the loop (113–118) that connects the two domains is replaced by Val117 in GmpC [Figs. 5(*b*) and 5(*c*)]. The structure superimposition clearly indicates that a dipeptide substrate in SP_0149 would have a steric clash with this residue. It appears reasonable to conclude that owing to its substrate-binding pocket being rigid and smaller in size, SP_0149 acquires substrate specificity for a single amino acid such as Met.

4. Conclusions

ABC transporters have a major impact on pneumococcal physiology, and their disruption can have an adverse effect on virulence (Kohler *et al.*, 2016). The involvement of some ABC transporters in survival and/or virulence, their location on the pneumococcal surface and the absence of these importers in humans make them good targets for the development of novel anti-pneumococcal drugs and immunotherapy (Garmory & Titball, 2004). SP_0149, the SBP of the Met ABC transporter, is required for the full virulence of *S. pneumoniae* (Basavanna *et al.*, 2013). The present study reports the crystal structure of SP_0149 with inherently bound Met and elucidates the interactions between the substrate and the SBP. The indigenous binding of Met to SP_0149 clearly indicates that Met is likely to be its physiological substrate. Although the overall fold is similar to those of other SBPs, the different architecture of the substrate-binding region as a consequence of the different amino-acid composition of the binding pocket may contribute to its specificity for Met. Importantly, the three-dimensional structure of SP_0149 in complex with Met opens up new avenues for the rational design of inhibitors of this transporter which may successfully curb the virulence of *S. pneumoniae*.

Acknowledgements

The X-ray data were collected using the in-house X-ray facility, which was established with financial support from the Department of Biotechnology (DBT), Government of India. We thank Ravikant Pal for his help during data collection.

Funding information

Both BKB and DS acknowledge the core funding received from the National Institute of Immunology, New Delhi, India for carrying out this study.

References

Basavanna, S., Chimalapati, S., Maqbool, A., Rubbo, B., Yuste, J., Wilson, R. J., Hosie, A., Ogunniyi, A. D., Paton, J. C., Thomas, G. & Brown, J. S. (2013). *PLoS One*, **8**, e49638.
Berntsson, R. P.-A., Smits, S. H. J., Schmitt, L., Slotboom, D.-J. & Poolman, B. (2010). *FEBS Lett.* **584**, 2606–2617.
Bhushan, J., Vyas, R., Sharma, T., Sehgal, D. & Biswal, B. K. (2011). *Acta Cryst.* F**67**, 797–799.
Ciruela, P., Izquierdo, C., Broner, S., Muñoz-Almagro, C., Hernández, S., Ardanuy, C., Pallarés, R., Domínguez, A., Jané, M., Esteva, C., de Sevilla, M. F., Henares, D., Grau, I., Marco, F., Llaberia, J., González-Cuevas, A., Díaz, A., Martin, M. T., Sierra, M., Curriu, M., Gallés, C., Hernández, P., Gassiot, P., Martínez-Zurita, M., Martí, C., Morta, M., Sauca, G., Gassós, A., Sanfeliu, E., Ballester, F., Pujol, I., Olsina, M., Raga, X., Gómez-Bertomeu, F., Pérez-Moreno, M. O., Vilamala, A., Navarro, M., Ribelles, M., Padilla, E., Prim, N., Fontanals, D., Benítez, M. A., Jou, E., Sanjosé, C., Giménez, M., Quesada, M. D., de la Fuente, J. C., Calderon, A., Ayala, P. J., Vega, L., Pérez-Jové, J., Balado, C. & Valle, I. (2018). *Vaccine*, **36**, 7744–7752.
Croucher, N. J., Løchen, A. & Bentley, S. D. (2018). *Annu. Rev. Microbiol.* **72**, 521–549.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Farha, M. A. & Brown, E. D. (2013). *Nature Biotechnol.* **31**, 120–122.
Garmory, H. S. & Titball, R. W. (2004). *Infect. Immun.* **72**, 6757–6763.
Guitor, A. K. & Wright, G. D. (2018). *Chest*, **154**, 1202–1212.
Henriques-Normark, B. & Tuomanen, E. I. (2013). *Cold Spring Harb. Perspect. Med.* **3**, a010215.
Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.
Kadioglu, A., Weiser, J. N., Paton, J. C. & Andrew, P. W. (2008). *Nature Rev. Microbiol.* **6**, 288–301.
Kohler, S., Voss, F., Gómez Mejia, A., Brown, J. S. & Hammerschmidt, S. (2016). *FEBS Lett.* **590**, 3820–3839.
Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lewinson, O. & Livnat-Levanon, N. (2017). *J. Mol. Biol.* **429**, 606–619.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Rees, D. C., Johnson, E. & Lewinson, O. (2009). *Nature Rev. Mol. Cell Biol.* **10**, 218–227.
Robert, X. & Gouet, P. (2014). *Nucleic Acids Res.* **42**, W320–W324.
Sakata, H., Sato, Y., Toyonaga, Y., Hanaki, H. & Drug-Resistant Pathogen Surveillance Group in Pediatric Infectious Disease (2018). *Pediatr. Int.* **60**, 52–56.
Sievers, F. & Higgins, D. G. (2014). *Methods Mol. Biol.* **1079**, 105–116.
Tanaka, K. J., Song, S., Mason, K. & Pinkett, H. W. (2018). *Biochim. Biophys. Acta*, **1860**, 868–877.
Usuf, E., Bottomley, C., Bojang, E., Cox, I., Bojang, A., Gladstone, R., Kampmann, B., Hill, P. C. & Roca, A. (2018). *Clin. Infect. Dis.* **67**, 1191–1197.

Wahl, B., O'Brien, K. L., Greenbaum, A., Majumder, A., Liu, L., Chu, Y., Luksic, I., Nair, H., McAllister, D. A., Campbell, H., Rudan, I., Black, R. & Knoll, M. D. (2018). *Lancet Glob. Health*, **6**, e744–e757.
Williams, W. A., Zhang, R. G., Zhou, M., Joachimiak, G., Gornicki, P., Missiakas, D. & Joachimiak, A. (2004). *Biochemistry*, **43**, 16193–16202.
Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta Cryst.* D**67**, 235–242.