PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

STRUCTURE NOTE

A novel "open-form" structure of sortaseC from *Streptococcus suis*

Guangwen Lu,[1,2] Jianxun Qi,[1] Feng Gao,[3,4] Jinghua Yan,[1] Jiaqi Tang,[5] and George F. Gao[1,2,6,7]*

[1] CAS Key Laboratory of Pathogenic Microbiology and Immunology (CASPMI), Institute of Microbiology, Chinese Academy of Sciences (CAS), Beijing 100101, China

[2] Graduate University, Chinese Academy of Sciences, Beijing 100049, China

[3] National Laboratory of Macromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China

[4] School of Life Sciences, Sichuan University, Chengdu 610064, Sichuan, China

[5] Department of Epidemiology, Research Institute for Medicine of Nanjing Command, Nanjing 210002, China

[6] China-Japan Joint Laboratory of Molecular Immunology and Molecular Microbiology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China

[7] Beijing Institutes of Life Science, Chinese Academy of Sciences, Beijing 100101, China

Key words: protein structure; class C sortase; sortaseC-specific lid; helical conformation; substrate binding site.

INTRODUCTION

Surface proteins of Gram-positive bacteria play critical roles in host–pathogen interactions. The vast majority of these surface targeting proteins are covalently linked to the cell wall peptidoglycan via the action of membrane-bound transpeptidases called sortases.[1,2] According to Shaynoor Dramsi, sortases are suggested to be sub-grouped into four classes denoted sortaseA–D.[3] Among these, sortaseA is responsible for the localization of a majority of surface targeting proteins[1,2]; and sortaseC plays more specialized roles in pili assembly and is therefore also conceived as pilus-associated sortases.[4–6]

To date, the atomic structures of several sortases have been reported.[7–12] All these structures contain a multi-stranded anti-parallel β-barrel core. This conserved fold is further surrounded by surface helices whose number and steric arrangement may vary dramatically among sortases of different classes but are relatively conserved for members of the same subfamily. Distinct from sortaseA or sortaseB, all thus far reported sortaseC structures exhibit a long loop (usually more than 25 amino acids in length) covering the external surface of the substrate binding groove, locking the site in a close state like a "lid."[7,8,11] Nevertheless, this lid is not necessarily required for sortaseC-catalyzed reactions[11,13] and is extremely flexible. These observations indicate that this mobile lid of class C sortase is movable; thereby enabling a free binding of the substrate to the enzyme. Yet the mechanism of lid-opening to allow for substrate access remains elusive.

Streptococcus suis serotype 2 (*S. suis* 2) is a Gram-positive zoonotic pathogen that can cause many serious diseases. Recent outbreaks of *S. suis* 2 infections in China in 1998 and 2005 with human cases of severe streptococcal toxic shock syndrome (STSS) have posed worldwide public health concerns.[14] The whole genome sequencing of one virulent isolate (05ZYH33) from the 2005 outbreak reveals the presence of several sortaseC coding genes.[15] Among these, a gene with the annotation number of SSU05_2098 (accession number: YP_001199464) encodes a sortase product of 295 amino acids in length, which shares over 50% sequence identity with sortaseC1s from *Streptococcus pneumoniae* and Group B Streptococcus. A topological prediction of this enzyme reveals a typical

Grant sponsor: National Science Foundation of China (NSFC); Grant number: 81021003

*Correspondence to: George Fu Gao, CAS Key Laboratory of Pathogenic Microbiology and Immunology (CASPMI), Institute of Microbiology, Chinese Academy of Sciences (CAS), NO. 1 West Beichen Road, Chaoyang District, Beijing 100101, China. E-mail: gaof@im.ac.cn

Received 4 April 2011; Revised 6 May 2011; Accepted 9 May 2011
Published online 20 May 2011 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.23093

character of class C sortase with two hydrophobic domains at both the N- and C-termini. In addition, this gene locates within a cluster whose organization resembles that of the *rlrA* pilus islet of *S. pneumoniae*.[16] Therefore, despite that the functional activity of this protein has not been elucidated yet, the enzyme should be a classical sortaseC. We denote it sortaseC1_suis here in this study. A construct encoding residues 67–244 of sortaseC1_suis was designed which retains the sortase core structure but excludes the hydrophobic N- and C-termini and the predicted disordered terminal regions.

Here, we reported the crystal structure of sortaseC1_suis. The overall structure of this enzyme resembles the previously reported sortaseC structures. Unexpectedly, rather than covering the external surface of the substrate binding groove as in other class C sortases, the sortaseC-specific lid of this *S. suis* enzyme was observed to form a long helix in the structure, leaving the whole substrate binding site wide-open. This is, to our knowledge, the first "open-form" sortaseC structure with a novel helical conformation for the lid region of pilus-associated sortases.

MATERIALS AND METHODS

Cloning and expression of sortaseC1_suis (67–244)

The coding sequence of sortaseC1_suis (covering amino acids 67–244) was PCR amplified from the genomic DNA extract of *Streptococcus suis* 05ZYH33 (serotype2) using primers 5′-AACCCATATGCGAATA GATTTGGCTCAAG-3′ and 5′-CCGCTCGAGATGCTCT TCTGCAACATAAG-3′. The subsequent fragment was inserted into pET-21a vector via *Nde1*/*Xho1* sites. The resultant recombinant plasmid, designated as pET-21a-suis-sortase, was verified by double enzyme digestion and direct DNA sequencing.

For protein expression, 1 μL plasmid was transformed into *Escherichia coli* BL-21 (DE3) competent cells and spread over a LB plate containing 50 mg/L ampicillin. The resultant single colony was inoculated into liquid LB medium supplemented with 50 mg/L ampicillin and incubated at 37°C overnight. The overnight culture was then transferred into 2-L fresh LB medium at a ratio of 1:100 and incubated until OD600 = 0.8, and induced by the addition of IPTG at a final concentration of 0.2 m*M* for 10 h at 16°C. Selenomethionine-labeled enzyme was produced in BL21 (DE3) *E. coli* grown in SeMet minimal medium (0.65% YNB, 5% glucose, 1 m*M* MgSO4 in M9 medium) supplemented with L-selenomethionine at 60 mg/L, lysine, threonine, and phenylalanine at 100 mg/L and leucine, isoleucine, and valine at 50 mg/L.

Protein purification and crystallization

For protein purification, the induced cells were harvested and resuspended in cold lysis buffer (20 m*M* Tris-HCl pH 8.0 and 50 m*M* NaCl) and homogenized by sonication. After removal of the cell debris by centrifugation at 16,000 rpm for 30 min, the resultant supernatant was added to Ni-NTA resin (GE) and gently mixed at 4°C for 1 h. The resin was then collected and washed with 20 column volumes of lysis buffer to remove the unspecific bound proteins. SortaseC1_suis was eluted using 200 m*M* imidazole in lysis buffer and further purified by gel filtration using a Superdex 75 Hiload 16/60 column (GE). The protein fractions were pooled and concentrated to about 18 mg/mL. The purification procedure for the selenomethionine-labeled protein follows that for the native protein.

The initial crystallization screening of the recombinant enzyme was performed using the Hampton Kits, and those conditions support the growth of protein crystals were optimized. The best crystals for both the native and the selenomethionine-labeled sortaseC1_suis enzymes were obtained by vapor diffusion in hanging drops consisting of 1 μL of reservoir solution (0.1*M* Tris-HCl pH 8.5, 0.54*M* sodium citrate, 8% V/V iso-propanol, 8% V/V *tert*-butanol) and 1 μL of concentrated protein solution (18 mg/mL in 20 m*M* Tris-HCl pH 8.0, 50 m*M* NaCl), following by incubation at 18°C for 5 days.

Data collection, phasing, and refinement

Crystals were flash-frozen in liquid nitrogen in cryoprotectant solution containing 20% glycerol and 50% reservoir solution. Multiwavelength X-ray diffraction data were collected from crystals of SeMet-labeled protein on beamline 3W1B of the Beijing Synchrotron Radiation Facility and processed using HKL2000.[17] Two selenium sites were successfully located with the data sets collected at peak, edge, and remote wavelengths using SOLVE.[18] An initial main-chain trace was obtained using RESOLVE[19] followed by model refinement using REFMAC5 (CCP4 suite).[20]

The diffraction data of the native protein was collected on an in-house Rigaku MicroMax007 rotating-anode X-ray generator equipped with an image plate detector. The data set was phased using MR procedure in MOLREP as implemented in CCP4,[20] with the initially determined structure as the search model. A series of iterative cycles of manual rebuilding were performed in COOT[21] and refined with Refmac5. The R_{free} and R_{work} values of the final structure are 26.1 and 20.6, respectively. During the course of model building and refinement, the stereochemistry of the structure was monitored by PROCHECK.[22] Data collection, phasing, and refinement statistics are summarized in Table I. Structural figures were generated with PyMOL (http://www.pymol.org).

Protein structure accession number

The coordinates and associated structure factors have been deposited into the Protein Data Bank with PDB code of 3RE9.

Table I
Data Collection, Phasing, and Refinement Statistics

Data collection				
	Native	SeMet MAD		
Space group	$P2_12_12$	$P2_12_12$		
Cell dimensions				
A, *b*, *c* (Å)	50.3, 101.0, 36.9	50.1, 100.8, 36.6		
		Peak	Edge	Remote
Wavelength (Å)	1.5418	0.97893	0.98050	0.90000
Resolution (Å)[a]	36.9–2.40 (2.49–2.40)	50.0–2.30 (2.38–2.30)	50.0–2.70 (2.80–2.70)	50.0–2.70 (2.80–2.70)
R_{sys} or R_{merge}[a,b]	0.093 (0.221)	0.145 (0.435)	0.116 (0.340)	0.129 (0.412)
$I/\sigma I$[a]	11.8 (6.6)	14.2 (2.85)	20.8 (7.03)	11.6 (3.67)
Completeness (%)[a]	99.7 (100.0)	98.8 (95.0)	100.0 (100.0)	99.9 (100.0)
Redundancy[a]	6.7 (6.8)	12.5 (7.7)	14.0 (14.4)	7.0 (7.2)
No. total reflections	52,427	108,065	77,504	38,543
No. unique reflections	7831	8683	5532	5535
Refinement				
Resolution (Å)	36.9–2.40			
R_{work}/R_{free}[c]	0.206/0.261			
No. reflections	7469			
No. protein atoms	1422			
No. water molecules	93			
Average *B*-factors				
Protein (Å^2)	19.0			
Water (Å^2)	20.6			
R.m.s deviations				
Bonds lengths (Å)	0.014			
Angles angles (°)	1.540			
Ramachandran plot[d]				
favored region	96.0%			
allowed region	4.0%			
outlier region	0			

[a]Values for the outmost resolution shell are given in parentheses.
[b]$R_{merge} = \Sigma_i\Sigma_{hkl}|I_i/\langle I\rangle|/\Sigma_i\Sigma_{hkl}I_i$, where I_i is the observed intensity and $\langle I\rangle$ is the average intensity from multiple measurements.
[c]$R_{work} = \Sigma||F_o|-|F_c||/\Sigma|F_o|$, where F_o and F_c are the structure-factor amplitudes from the data and the model, respectively. R_{free} is the *R* factor for a subset (5%) of reflections that was selected before refinement calculations and was not included in the refinement.
[d]Ramachandran plots were generated by using the program PROCHECK.

RESULTS AND DISCUSSION

Structure determination and the quality of the refined structure

So far, the structures of five class C sortases (soraseC1-C3 from *S. pneumoniae*, sortaseC1 from Group B Streptococcus and sortaseC1 from *Actinomyces oris*) have been reported.[7,8,11,23] Among them, the sortaseC1s from *S. pneumoniae* and GBS share the highest sequence identities of 54.3% and 55.3% with our enzyme. Nevertheless, we failed to solve the structure of our protein by molecular replacement using either of the two aforementioned structures as the search models. Therefore, a seleno-methionine-based multiwavelength anomalous dispersion (MAD) method was used to solve the phase problem.[18] The successful localization of two crystallographically independent selenium atoms within the recombinant protein enable us to build a structural model with good stereochemistry (For details see "Materials and Methods").

The final model contains 179 residues including R67 to H244 of sortaseC1_suis and the initiation amino acid methionine. As shown in Table I, the refined structure solved to 2.4 Å resolution, has R_{work} = 20.6 and R_{free} = 26.1, respectively. A total of 96.0% of the residues lie in the favored region of the Ramachandran plot and 4.0% lie in the allowed region based on the calculation using PROCHECK.[22] The excellent electron density map for the recombinant enzyme and an overall temperature factor of 19.0 Å^2 enable us to model the positions of the main chain, side chain atoms, as well as the solvent molecules unambiguously.

The overall structure of sortaseC1_suis

Like other sortases with known structures, sortaseC1 from *S. suis* is also a single-domained enzyme. The core of the protein is composed of 8 β-strands (β1–β8) arranged in an anti-parallel manner, forming a compact barrel. This β-barrel core is further surrounded by several short loops, one 3_{10} helix and one α-helix (α3), together forming the typical fold that could be observed in all class A and C sortases [Fig. 1(A)]. At the N-terminus of the protein, two α-helices (α1–α2), whose orientations are almost perpendicular to each other, are suspended on

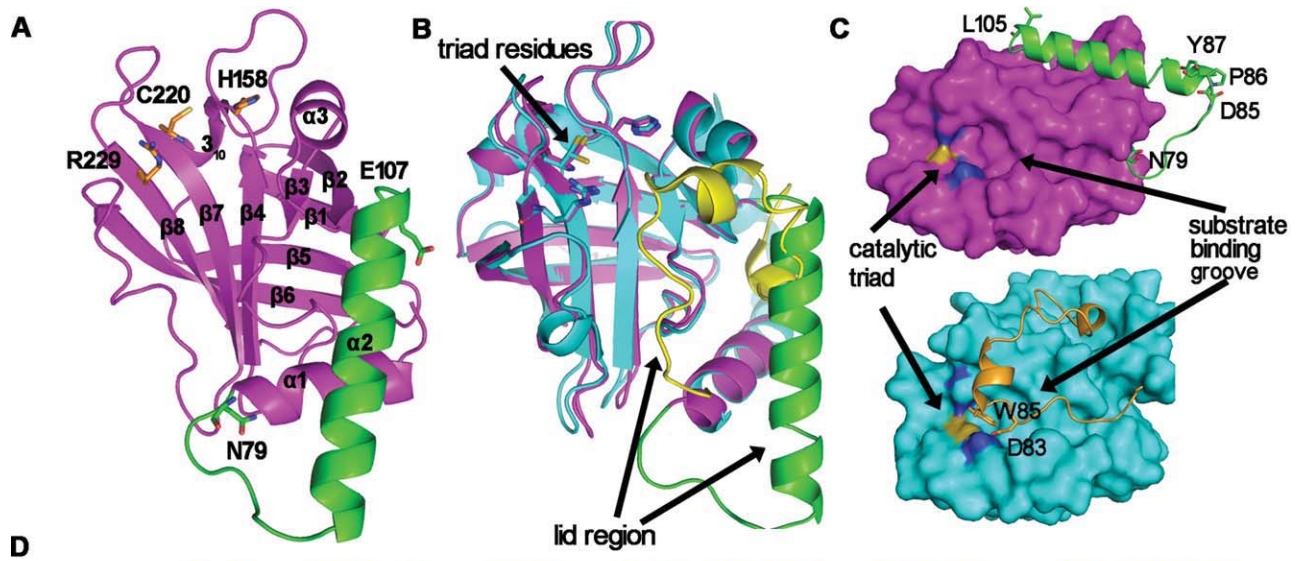

Figure 1

SortaseC1_suis exhibits a novel helical conformation for the sortaseC-specific lid. **A**: The overall structure of sortaseC1_suis. α-helices and β-strands are marked α1–α3 and β1–β8, respectively, according to their occurrence along the primary structure. The 3_{10} helix is indicated. The catalytic triad, which is composed of H158, C220, and R229, is shown as orange sticks. The region that covers the sortaseC-specific lid is highlighted in green. Residues N79 and E107 which marks the starting and ending points of the lid region are indicated. **B**: Superimposition of sortaseC1_suis (magenta) and sortaseC1_pneu (cyan). The well-aligned catalytic triads are highlighted. The lid regions in the suis and pneumococcal enzymes, which exhibits great structural discrepancies, are colored green and yellow, respectively. **C**: Surface representation of the structures of sortaseC1_suis (top) and sortaseC1_pneu (bottom) to highlight the sortaseC-specific lid in respective enzymes and its steric position relative to the substrate binding groove. In sortaseC1_pneu, the lid region covers the external surface of the substrate binding groove. Nevertheless, the corresponding region was shown to form a helix for residues P86-L105, leaving the substrate binding site fully solvent exposed in the *S. suis* enzyme. The residues referred to in the text are indicated. **D**: Sequence alignment of the sortaseC1 enzymes from *S. suis* and *S. pneumoniae*. The lid region in respective enzymes is highlighted by enclosure with a black box.

one side of the enzyme molecule that is opposite to the nucleophilic residue C220. Two basic amino acids including H158 and R229 were observed to locate in the proximity to C220 [Fig. 1(A)]. These three residues most likely constitute the catalytic triad of sortaseC1_suis. Recent studies on other class C sortases have demonstrated that enzymes of this family also use a three-residue triad for catalysis, substituting any of which by Ala could totally abolish their activities.[11,13]

Structural comparison with other class C sortases reveals a novel helical conformation for the sortaseC-specific lid

All thus far reported sortaseC structures could be described as a combination of a barrel core plus peripheral helices and a flexible loop acting as a "lid", that are sterically arranged in a similar mode.[7,8,11] Of these sortases with known structures, the transpeptidases sortaseC1 from *S. pneumoniae* (sortaseC1_pneu) and sortaseC1 of Group B streptococcus exhibit the highest similarities with our enzyme when searching through the Protein Data Bank (PDB) database. Since soraseC1_pneu is the only one whose backbone for the whole lid region is traceable, this pneumococcal enzyme was selected as the representative protein whose structure was superimposed onto that of sortaseC1_suis. As expected, striking structural similarity was observed for the two enzymes with a root mean square deviation (R.M.S.D) of about 1.29 Å for 146 analogous Cα atoms. The β-barrel cores of respective proteins are well structurally conserved

(with an R.M.S.D of only 0.48 Å for the β-barrel alone), as are the steric positions of the catalytic triad residues [Fig. 1(B)].

Despite the great similarities between the two sortaseC1 proteins, our enzyme exhibits a large structural incongruence with its pneumococcal counterpart for the region extending from N79 to E107 [Fig. 1(B)]. The corresponding residues have been shown to form a unique lid structure covering the external surface of the substrate binding groove in other class C sortases.[7,8,11] As represented by sortaseC1_pneu, this lid region is mainly a structure of two 3_{10} helices connected and flanked by several short loops, with residues D83 and W85 locating near the catalytic centre interacting with the triad residues [Fig. 1(C) bottom]. Accordingly, the nucleophilic Cys in the structure of sortaseC1_pneu was shown to adopt two conformations, which are postulated to be the result of a sulfur-aromatic interaction conferred by the tryptophan side-chain.[7] Nevertheless, in sortaseC1_suis, residues P86-L105 were observed to form a long helix whose axis is almost parallel to the orientation of the substrate binding groove. Amino acids N79 to D85 remain in a loop conformation. However, this loop is rotated around the base-positioned N79 residue over 90° relative to the corresponding loop in the pneumococcal enzyme, thereby leaving the whole substrate binding site of sortaseC1_suis fully solvent exposed [Fig. 1(C) top]. With D85 and Y87 being far away from the catalytic triad, C220 displays a consistent conformation with its side-chain pointing toward H158 [Fig. 1(A)].

It is noteworthy that the sequence similarity for the lid region between sortaseC1_suis and sortaseC1_pneu (47.4%) is relatively comparable to that of the overall similarity throughout the whole protein (54.3%) [Fig. 1(D)]. Therefore, it is totally unexpected that this region in the two sortases was structurally quite different, in terms of both its composition of secondary elements and its steric position relative to the β-barrel core. The observance of a lid structure covering the substrate binding groove in all thus far reported sortaseC structures raises the intrinsic question about the possible way of lid-opening upon substrate binding to the enzyme. Yet the fact that we were able to capture such a sortaseC structure with its lid region forming a unique helix and its substrate binding groove consequently being in a wide-open state, might reveal some clues about the mechanism of lid-opening for class C sortases which is necessary to allow substrate access.

ACKNOWLEDGMENTS

GFG is a leading principal investigator of the Innovative Research Group of National Natural Science Foundation of China (NSFC). The authors thank Qiangmin Zhang, Youjun Feng, and Hao Peng for their help with the project. They also thank Christopher J. Vavricka for his discussion of the manuscript.

REFERENCES

1. Paterson GK, Mitchell TJ. The biology of Gram-positive sortase enzymes. Trends Microbiol 2004;12:89–95.
2. Marraffini LA, Dedent AC, Schneewind O. Sortases and the art of anchoring proteins to the envelopes of Gram-positive bacteria. Microbiol Mol Biol Rev 2006;70:192–221.
3. Dramsi S, Trieu-Cuot P, Bierne H. Sorting sortases: a nomenclature proposal for the various sortases of Gram-positive bacteria. Res Microbiol 2005;156:289–297.
4. Ton-That H, Schneewind O. Assembly of pili on the surface of *Corynebacterium diphtheriae*. Mol Microbiol 2003;50:1429–1438.
5. Falker S, Nelson AL, Morfeldt E, Jonas K, Hultenby K, Ries J, Melefors O, Normark S, Henriques-Normark B. Sortase-mediated assembly and surface topology of adhesive pneumococcal pili. Mol Microbiol 2008;70:595–607.
6. Fittipaldi N, Takamatsu D, de la Cruz Dominguez-Punaro M, Lecours MP, Montpetit D, Osaki M, Sekizaki T, Gottschalk M. Mutations in the gene encoding the ancillary pilin subunit of the *Streptococcus suis* srtF cluster result in pili formed by the major subunit only. PLoS One 2010;5:e8426.
7. Manzano C, Contreras-Martel C, El Mortaji L, Izore T, Fenel D, Vernet T, Schoehn G, Di Guilmi AM, Dessen A. Sortase-mediated pilus fiber biogenesis in *Streptococcus pneumoniae*. Structure 2008; 16:1838–1848.
8. Neiers F, Madhurantakam C, Falker S, Manzano C, Dessen A, Normark S, Henriques-Normark B, Achour A. Two crystal structures of pneumococcal pilus sortase C provide novel insights into catalysis and substrate specificity. J Mol Biol 2009;393:704–716.
9. Zong Y, Bice TW, Ton-That H, Schneewind O, Narayana SV. Crystal structures of *Staphylococcus aureus* sortase A and its substrate complex. J Biol Chem 2004;279:31383–31389.
10. Zhang R, Wu R, Joachimiak G, Mazmanian SK, Missiakas DM, Gornicki P, Schneewind O, Joachimiak A. Structures of sortase B from *Staphylococcus aureus* and *Bacillus anthracis* reveal catalytic amino acid triad in the active site. Structure 2004;12:1147–1156.
11. Cozzi R, Malito E, Nuccitelli A, D'Onofrio M, Martinelli M, Ferlenghi I, Grandi G, Telford JL, Maione D, Rinaudo CD. Structure analysis and site-directed mutagenesis of defined key residues and motives for pilus-related sortase C1 in Group B Streptococcus. FASEB J 2011;25:1874–1886.
12. Race PR, Bentley ML, Melvin JA, Crow A, Hughes RK, Smith WD, Sessions RB, Kehoe MA, McCafferty DG, Banfield MJ. Crystal structure of *Streptococcus pyogenes* sortase A: implications for sortase mechanism. J Biol Chem 2009;284:6924–6933.
13. Manzano C, Izore T, Job V, Di Guilmi AM, Dessen A. Sortase activity is controlled by a flexible lid in the pilus biogenesis mechanism of Gram-positive pathogens. Biochemistry 2009;48:10549–10557.
14. Tang J, Wang C, Feng Y, Yang W, Song H, Chen Z, Yu H, Pan X, Zhou X, Wang H, Wu B, Zhao H, Lin Y, Yue J, Wu Z, He X, Gao F, Khan AH, Wang J, Zhao GP, Wang Y, Wang X, Gao GF. Streptococcal toxic shock syndrome caused by *Streptococcus suis* serotype 2. PLoS Med 2006;3:e151.
15. Chen C, Tang J, Dong W, Wang C, Feng Y, Wang J, Zheng F, Pan X, Liu D, Li M, Song Y, Zhu X, Sun H, Feng T, Guo Z, Ju A, Ge J, Dong Y, Sun W, Jiang Y, Yan J, Yang H, Wang X, Gao GF, Yang R, Yu J. A glimpse of streptococcal toxic shock syndrome from comparative genomics of *S. suis* 2 Chinese isolates. PLoS One 2007;2:e315.
16. Barocchi MA, Ries J, Zogaj X, Hemsley C, Albiger B, Kanth A, Dahlberg S, Fernebro J, Moschioni M, Masignani V, Hultenby K, Taddei AR, Beiter K, Wartha F, von Euler A, Covacci A, Holden DW, Normark S, Rappuoli R, Henriques-Normark B. A pneumo-

coccal pilus influences virulence and host inflammatory responses. Proc Natl Acad Sci USA 2006;103:2857–2862.
17. Otwinowski Z, Minor W. Processing of X-ray Diffraction Data Collected in Oscillation Mode, Methods in Enzymology, Volume 276: Macromolecular Crystallography, part A, C.W. Carter, Jr. & R. M. Sweet, Eds., New York: Academic Press 1997;276:307–326.
18. Terwilliger TC, Berendzen J. Automated MAD and MIR structure solution. Acta Crystallogr D Biol Crystallogr 1999;55 (Part 4):849–861.
19. Terwilliger TC. Maximum-likelihood density modification. Acta Crystallogr D Biol Crystallogr 2000;56 (Part 8):965–972.
20. Collaborative Computing Project Number 4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 1994;50 (Part 5):760–763.
21. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 2004;60 (Part 12 Part 1): 2126–2132.
22. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–291.
23. Persson K. Structure of the sortase AcSrtC-1 from Actinomyces oris. Acta Crystallogr D Biol Crystallogr 2011;67(Pt 3):212–217.