Interrogation of 3D-swapped structure and functional attributes of quintessential Sortase A from *Streptococcus pneumoniae*

Tora Biswas[1¶], Anurag Misra[2¶], Sreetama Das[2¶], Prity Yadav[1], Suryanarayanarao Ramakumar[2‡] and Rajendra P. Roy[1‡]

[1]National Institute of Immunology, Delhi 110067, India.

[2]Department of Physics, Indian Institute of Science, Bangalore 560012, India.

[‡]Correspondence: Rajendra P. Roy, National Institute of Immunology, Delhi 110067, India Email: rproy@nii.ac.in and Suryanarayanarao Ramakumar, Department of Physics, Indian Institute of Science, Bangalore 560012, India, Email: ramak@iisc.ac.in

[¶]These authors contributed equally to the work.

Keywords: sortase, transpeptidase, 3D domain swapping, crystal structure, peptide substrate

Abstract

The anchoring of the surface proteins to the cell wall in gram-positive bacteria involves a peptide ligation reaction catalyzed by transpeptidase sortase. Most bacterial genomes encode multiple sortases with dedicated functions. *Streptococcus pneumoniae* (Sp) carries four sortases; a housekeeping sortase (SrtA), and three pilin specific sortases (SrtC1, C2, C3) dedicated to the biosynthesis of covalent pilus. Interestingly, SrtA, meant for performing housekeeping roles, is also implicated in pilus assembly of Sp. The allegiance of SpSrtA to the pathogenic pilus assembly makes it an ideal target for clinical inhibitor development. In this paper, we describe biochemical characterization, crystal structure and peptide substrate preference of SpSrtA. Transpeptidation reaction with a variety of substrates revealed that the enzyme preferred elongated LPXTG sequences and transferred them equally well to both Ala- and Gly-terminated peptides. Curiously, crystal structure of both wild type and an active site (Cys to Ala) mutant of SpSrtA displayed inter-twined 3D-swapped dimers in which each protomer generated a classic eight stranded beta-barrel “sortase fold”. Size-exclusion chromatography and sedimentation equilibrium measurements revealed predominant presence of a dimer in equilibrium with its monomer. The crystal structure-based Cys-Cys distance mapping with defined chemical cross-linkers established the existence of 3D-swapped structure in solution. The swapping in SpSrtA, unprecedented for sortase family, may be physiologically relevant and meant to perform regulatory functions.

Introduction

Streptococcus pneumoniae (S. pneumoniae) is a gram-positive bacterium and a commensal of human nasopharyngeal cavity. Under favourable conditions, *S. pneumoniae* can infect other host tissues and cause pneumonia, meningitis, endocarditis, and cellulites [1]. The surface proteins of *S. pneumoniae* play critical roles in bacterial pathogenesis by facilitating attachment of the bacterium to the host and promoting degradation of host-tissue components [2–6]

Several surface proteins of *S. pneumoniae*, and other gram-positive bacteria, contain a LPXTG type of pentapeptide sequence in their C-terminal region that serves as a recognition motif for a class of cysteine transpeptidases called sortase [7].The bacterial genomes generally encode multiple sortases including a ubiquitous housekeeping sortase A (SrtA) which covalently anchors the surface proteins to the cell wall peptidoglycan [8]. Other sortases are meant for dedicated functions, such as, assembly of large protein appendages built by covalent linking of one or more pilin subunits [9]. The sortase-mediated covalent anchoring of proteins proceeds through a thioacyl-enzyme intermediate generated by nucleophilic attack of the thiol group of the catalytic Cys residue on the scissile T-G peptide bond of the LPXTG motif [10]. Subsequent reaction of the thioacyl-enzyme intermediate with the terminal amine group of the pentaglycine/dialanine branch of the peptidoglycan results in the formation of a peptide bond leading to the covalent attachment of the surface protein to the peptidoglycan. Sortases involved in pilin assembly use epsilon amine of a specific Lys residue present in the pilin subunit to resolve the thioacyl-enzyme intermediate and link the pilin chains by an isopeptide bond [9].

S. pneumoniae expresses a housekeeping SrtA and three other sortases (SrtC1, SrtC2, and SrtC3) responsible for pilus assembly from constituent RrgA, RrgB and RrgC pilin subunits [11,12]. The mechanism of pilus assembly *vis a vis* specificity of sortases for pilin subunits has been a subject of intense investigation because of the role of pilus in

pathogenicity of *S. pneumoniae*. Cumulative biochemical evidence gathered in the past years have defined the structure and specificity of sortases for pilin subunits. SrtC2 is specific for RrgA, while both SrtC1 and SrtC3 are selective for RrgB [11]. Interestingly, RrgC is recognized by housekeeping SrtA and not by pilin sortases. The ability of SrtA of *S. pneumoniae* to perform dual function of housekeeping and pilus synthesis may be critical for virulence and pathogenesis [12]. Notably, ablation of SrtA gene in *S. pneumoniae* affects display of LPXTG- proteins on the bacterial cell surface, and SrtA-deficient bacteria do not cause disease in animal models [13–15]. Thus, SrtA is considered an excellent target for rational inhibitor development.

The crystal structures of pilin sortases, SrtC1 (PDB id: 2w1j), SrtC2 (PDB id: 3g66, 3g69), and SrtC3 (PDB id: 2w1k), have been elucidated and provide some clues about substrate selectivity and regulatory mechanism in pilus biosynthesis [16]. We herein report functional attributes, LPXTG peptide substrate preference, and crystal structure of the catalytic domain encompassing residues 82-247 of housekeeping SrtA of *S. pneumoniae* (SpSrtA). Curiously, Δ81SpSrtA forms intertwined 3D-swapped dimers in the crystals in which each protomer generates a conserved eight-stranded beta-barrel “sortase fold” by exchanging the C-terminal beta-strands. We have probed the fidelity of the swapped structure in solution using site-directed mutagenesis, size-exclusion chromatography, analytical ultracentrifugation, and crystal structure-based distance mapping with specific chemical cross-linkers of defined length. The results suggest the existence of a swapped dimeric species in solution as well.

Materials and Methods

Cloning, expression, and purification of SpSrtA

SpSrtA sequence (Uniport ID: Q8DPM3) corresponding to residues, 60-247 and 82-247, respectively, was expressed in *E.coli* BL-21 cells as described previously [17]. Briefly, PCR was carried out to amplify the DNA encoding residues 60-247 of *S. pneumoniae* SrtA

(Δ59SpSrtA) from the genomic DNA of *S. pneumoniae* Strain R6 using specific primers (Table S1). The PCR product was digested with enzymes NdeI and HindIII and subsequently ligated into NdeI/HindIII digested pET28c vector. The Δ81SpSrtA (residue 82-247) was subcloned from pET28c-Δ59SpSrtA construct, using primers listed in Table S1, following the same protocol as above. DNA sequencing was done to verify the identity of individual clones.

The respective constructs were transformed into *E. coli* BL-21 cells for expression. The transformed cells were grown in LB media, containing kanamycin (50 μg/ml) till mid-log phase (OD_{600}~0.5) at 37 °C. Protein expression was induced by addition of 0.5 mM IPTG and grown at 30 °C for 5 hrs. The cells were harvested, re-suspended in a buffer composed of 10 mM Tris-HCl (pH 7.5), 40 mM NaCl, 2 mM β-mercaptoethanol and lysed by sonication. The supernatant was collected after sonication and purified using Ni-NTA based affinity chromatography as described previously [17].

Site-directed mutagenesis

Site-directed mutagenesis was carried out according to the QuickChange (Stratagene) protocol. Briefly, mutation was introduced by PCR using pET28c-Δ81SpSrtA as the template DNA and mutagenic primers to introduce desired mutation in the template DNA. The PCR product was digested with *Dpn1* for 1 hr at 37 °C and the reaction mixture was transformed into *E. coli* DH5α cells. Plasmid DNA was isolated from *E. coli* DH5α cells and mutation was confirmed by DNA sequencing. The plasmid DNA was transformed into *E. coli* BL-21 cells for expression. The mutagenic primers are listed in Table S2.

Synthesis and purification of peptide substrates

Peptides were synthesized using standard solid phase method using Fmoc chemistry on a peptide synthesizer (Applied Biosystems, ABI 433A or Advanced Chemtech, Model ACT90). Wang resin pre-loaded with desired amino acid was used as the starting material for synthesis. 20% piperidine in NMP was used for deprotection of the Fmoc group. Activation and coupling were achieved using HBTU/HoBt/DIEA reagents. The peptide was cleaved

from the resin with 95% aqueous trifluoroacetic acid, filtered, precipitated in cold diethyl ether, and extracted in water. The peptide was purified by RP-HPLC using a preparative C18 column (Phenomenex, 100 Å, 10 µ, 30 X 250 mm, using a gradient of 8-72% acetonitrile in 130 min, flow rate: 30 ml/min). The purity of the peptide was checked by analytical RP-HPLC (Phenomenex, 100 Å, 5 µ, 4.6 X 250 mm, gradient 4-72% acetonitrile in 130 min, flow rate: 1ml/min), and MALDI-TOF analysis (Table S3).

Crystallization, diffraction data collection and processing

The Ni-NTA purified protein was desalted on a PD-10 column, concentrated using a Millipore Amicon filters (10 kDa cut-off) and used for crystallization. Conditions supporting the growth of wild type Δ81SpSrtA crystals were initially ascertained using the hanging-drop vapor-diffusion and micro-batch under-oil methods at 20 °C with commercially available crystallization screens as reported previously [17]. The best diffracting crystal was grown using the under-oil method in a solution containing 25 mg/mL of protein (10 mM Tris buffer, pH 7.5), 0.2 M tri-ammonium citrate and 20% (*w/v*) PEG 3350, pH 7.0 along with 40% *v/v* (±)-1,3-butanediol as an additive agent. Diffraction quality crystals of the Δ81SpSrtA(C207A) mutant protein were grown in a condition containing 0.2 M tri-ammonium citrate and 20% (*w/v*) PEG 3350, pH 7.0, with 1.0 M guanidine hydrochloride as an additive agent. Crystals grew to their maximum dimensions in 25-30 days for the wild-type and 5-7 days for the mutant.

X-ray diffraction data from the wild-type protein crystal were collected at home source (λ = 1.5418 Å) without any cryoprotectant, whereas diffraction data for the mutant protein (C207A) crystal were collected at the synchrotron (λ = 0.9772 Å, BM-14, ESRF) using 20% glycerol as cryoprotectant. Crystals of the wild type and mutant protein diffracted to 2.70 Å and 2.48 Å, respectively. The diffraction data for both the crystals were processed using MOSFLM [18] and scaled using SCALA from the CCP4 program suite [19]. Data-collection and processing statistics for the crystals are presented in (Tables1 and 2). The quality of

data sets was assessed using SFCHECK [20] which suggested no twinning in the data.

Structure solution, refinement and analysis

Structure of Δ81SpSrtA was solved by molecular replacement using Phaser [21] with the homologous *S. pyogenes* SrtA (78% sequence identity, PDB id: 3fn5, [22]) as the search model. The molecular-replacement (MR) solution yielded four molecules in the crystal asymmetric unit, consistent with expectation from Matthews coefficient and solvent content. The MR solution was subjected to 10 cycles of rigid body refinement followed by several cycles of restrained refinement using the program REFMAC5 [23] from the CCP4 package [19], with alternate rounds of inspection and manual model building in COOT [24]. A few rounds of simulated annealing were also performed using phenix.refine from the PHENIX suite [25] followed by TLS refinement in REFMAC5. 5% of the measured reflections were kept aside for calculating R_{free} [26]. The electron density map showed breaks in the backbone between P187 and R189 in all four chains. The structure was finally modeled as a swapped dimer, which led to a convergence of R_{work}/R_{free} (18.10% / 23.37%) (Figure S1).

The Δ81SpSrtA(C207A) model was built using molecular replacement with *S. pyogenes* SrtA as model, and not the wild type enzyme, since 3D-swapping was not assured. Subsequently, the model was refined using the same protocol employed for Δ81SpSrtA. 3D-swapping was observed in the 2Fo-Fc and Fo-Fc maps of the mutant protein as well. The final model was refined with R_{work} = 18.23% and R_{free} = 22.48%.

The stereochemical acceptability of the structures was analysed using MOLPROBITY [27] and validated using PROCHECK [28] Structural superposition was performed using SUPERPOSE [29]. Domain motions and dimer interface statistics were evaluated using the DYNDOM server [30] and PISA server [31] respectively. Coordinates and structure factors for Δ81SpSrtA and Δ81SpSrtA(C207A) have been deposited in the Protein Data Bank with accession codes 4o8l and 4o8t, respectively.

The hinge region was estimated using the method proposed by Shingate and

Sowdhamini [32]. First, the swapped monomer structure of Δ81SpSrtA was superposed with the unswapped monomer from the closest homolog *S. pyogenes*; the swapped domain remained unaligned. The unaligned C-terminal stretch was then superposed separately with the equivalent region from *S. pyogenes* SrtA. Stretches that were unaligned in both instances were likely to form the hinge region (Table S4). Further, in the swapped dimer, the hinge loop is 'extended', whereas in the unswapped homolog it forms a turn structure; hence the difference in backbone torsion angles was also evaluated using the method proposed by Bennett and co-workers [33]. We calculated delta-torsion as $\sqrt{(\Delta\varphi^2 + \Delta\psi^2)}$, where φ, ψ are the backbone torsion angles. The stretch of non-superposable residues with delta-torsion > 30° (Table S5) was considered as the hinge region.

Transpeptidation assay

Transpeptidation reaction between LPXTG (donor peptide substrate) and AAKY/GGGKY (acceptor nucleophilic peptide) was carried out at pH 7.5 (50 mM Tris-HCl, 150 mM NaCl**,** 2 mM β-mercaptoethanol) in the presence of desired concentration of Δ81SpSrtA. Reactions were set up in a volume of 50 µl and incubated at 37 °C for indicated time. Each reaction was quenched by addition of 10-fold excess of 0.1% TFA and analysed on analytical RP-HPLC (C18, 4.6x250 mm, gradient 4-72% acetonitrile in 0.1% TFA over 130 min, flow rate-1 ml/min). The relevant peaks were collected, lyophilised and characterized by mass spectrometry (Table S6). For, assessing the effect of Ca^{2+}on Δ81SpSrtA activity, transpeptidase reaction was carried out as above but in the presence of 5 mM $CaCl_2$.

Transpeptidation activity of ∆81SpSrtA was also assayed in the crystallization buffer (0.2 M tri-ammonium citrate and 20% (*w/v*) PEG 3350, pH 7.0 along with 40% *v/v* (±)-1,3-butanediol) using the above protocol.

Size exclusion chromatography

Size-exclusion chromatography was carried out by FPLC on a Superdex200 (10/300 GL) column. The column was pre-equilibrated with degassed buffer composed of 50 mM Tris

(pH 7.5) containing 150 mM NaCl at a flow rate of 1 ml/min. Protein sample was loaded on the column using a 100 µl injection loop and elution was monitored at 280 nm. The fractions corresponding to individual peaks were pooled and concentrated using protein concentrator (Amicon, Millipore, 10 kDa MW cut-off).

Analytical ultracentrifugation

Sedimentation equilibrium experiments were performed using an Optima XL-A analytical ultracentrifuge equipped with an An60Ti rotor (Beckman Inc). Sedimentation equilibrium studies were carried out at 13000, 17000, 23000, and 27000 rpm, at 20 °C using six-channel charcoal-filled centerpieces. Data were collected by scanning samples at 280 nm with the resolution of 0.003 cm and average of 5-7 scans per step. The partial specific volume and solvent density were calculated using SEDNTERP. Equilibrium data were edited with WinREEDIT and analyzed by nonlinear least squares using WINNONLIN and subsequent simulations were performed using SCIENTIST (Micromath, Salt Lake City, UT, USA). Absorbance profiles were analyzed using a sum of two exponential model which describes the monomer-dimer equilibrium as given below (eq 1-4).

$$2M \xleftrightarrow{K_A} M_2 \quad (1)$$

$$A_T = \sum \exp([\ln A_{ref}] + \ln K_{A,abs} + n\sigma\xi) + d \quad (2)$$

Where M and M_2 represent monomer and dimer respectively; K_A and $K_{A,abs}$, are association equilibrium constant and association constant defined in absorbance units; A_T, total absorbance at given radial position, A_{ref}, absorbance at reference radial position; $\xi = (r^2 - r_{ref}^2)/2$, d, base line offset, and σ, is the reduced molecular mass of monomer and is described as

$$\sigma = M(1 - \upsilon\rho)\omega^2/RT \quad (3)$$

Where M, molecular weight of the monomer, υ is the partial specific volume, ρ solvent

density, ω, is the rotor angular speed; R and T are gas constant and temperature respectively. The overall equilibrium association constant K_A can be calculated from the association constant estimated in absorbance units, $K_{A,obs}$ as given by

$$K_A = K_{A,obs} (\varepsilon l/n) \qquad (4)$$

where ε, is the extinction coefficient of monomer, and l is the path length of the sedimentation cell (1.2 cm).

Chemical cross-linking

Cross-linking reaction was carried out in 100 µL of Tris-HCl buffer (50 mM Tris-HCl containing 150 mM NaCl, pH 7) at 37 °C. Sortase (100 µM) was incubated with 50 µM individual cross-linker (Pierce Chemical Co, USA) for 15 min. The reaction was quenched by addition of 50 mM DTT and analyzed by SDS-PAGE. Cross-linking experiments employing above conditions were also performed in the presence of 8 M urea.

Results

N-terminal truncated versions (Δ59SpSrtA and Δ81SpSrtA) of SpSrtA are active

The amino acid sequences of SrtA from the pathogenic *S. pneumoniae* **(**TIGR4 strain) and non-pathogenic R6 strain are identical except for a single amino acid (N233D) substitution. SrtA sequence corresponding to residues, 60-247 and 82-247, respectively, was expressed in *E coli* as described previously [17]. Both the constructs contained a hexa-His tag and were purified using Ni-NTA affinity chromatography (Figure S2). The ES-MS mass of purified Δ59SpSrtA was found to be 23493 Da which was in accord with the calculated mass of the protein (23495 Da) without a Met residue (Figure S3). Likewise, Δ81SpSrtA yielded a mass of 20949 Da that fit the computed mass of the des-Met protein. Thus, N-terminus Met residue was processed by *E. coli* aminopeptidases in both constructs of SpSrtA.

We first assessed the transpeptidation activity of Δ59SpSrtA to see if the truncated version of sortase was active. Previous studies revealed that native LPXTG pentapeptide substrate was not recognized by *S. aureus* SrtA [34] However, LPXTG turned into an effective substrate when its termini were derivatized by acetylation and amidation or extended by an amino acid. Thus, it was important to see if Δ59SpSrtA displays a similar trend. We used LPXTG based donor substrates against both Ala and Gly- based nucleophilic acceptor peptides considering the presence of Ala-Ala branch in the peptidoglycan of *S. pneumoniae*. Accordingly, sortase-mediated transpeptidation reaction was examined with native and appropriately modified LPXTG pentapeptides using AAKY and GGGKY as nucleophile acceptors (Figure 1A, Figure S4).

RPHPLC analyses of the transpeptidation reaction mixture carried out in the presence of Δ59SpSrtA with native LPNTG (free termini) and AAKY or GGGKY as substrates did not yield any product. The results were similar and no product was seen when either end of LPNTG pentapeptide (acetylation of the amino terminus or amidation of the carboxyl end) was modified. However, blocking of both the ends of the pentapeptide motif (Ac-LPNTG-NH_2) facilitated the transpeptidation reaction with both AAKY and GGGKY in a similar fashion leading to a product yield of about 8%.

Next, the effect of Ala residue at the LPNTG termini was evaluated using ALPNTGA. Under identical conditions as above, heptapeptide ALPNTGA produced transpeptidation yields of 10-12% which was marginally higher than the yield for Ac-LPNTG-NH_2. Further, we placed Gln at N-terminus of LPNTG because Gln residue precedes the LPNTG sorting motif in β-galactosidase which is annotated as a substrate of SrtA in *S. pneumoniae*. The resulting QLPNTGA peptide behaved in a similar way as ALPNTGA and produced about 13% yield of the transpeptidation product indicating that blocking of the termini of LPXTG recognition motif by chemical modification or placement of an amino acid exerts similar effect on Δ59SpSrtA-mediated transpeptidation reaction.

Subsequently an Ala residue was added at the N-terminus of QLPNTGA to see if extension of the motif has any bearing on the transpeptidation reaction. Interestingly, AQLPNTGA peptide produced about 25% yield that was almost two-fold or more than that obtained when termini modified LPNTG or ALPNTGA or QLPNGA peptides were used as donor substrates. Finally, we extended the AQLPNTGA sequence by placing a Tyr residue at the N-terminus to further evaluate the effect of peptide length on transpeptidation reaction as also to use Tyr residue for spectroscopic estimation of peptide concertation. The transpeptidation yield of about 28-30% associated with YAQLPNTGA was marginally higher than that obtained with AQLPNTGA. Taken together, the above results indicate that Δ59SpSrtA preferred relatively longer peptide substrates. This feature of *S. pneumoniae* SrtA was in sharp contrast with *S. aureus* SrtA which was shown to process Ac-LPNTG-NH_2 as effectively as the octapeptide YALPNTGK [34].

Next, we assessed the transpeptidase activity of Δ81SpSrtA to see if further truncation (22 residues) was detrimental to enzyme activity. The transpeptidation assay was carried out using YAQLPNTGA as the substrate peptide and AAKY or GGGKY as the acceptor peptide (Figure 1B). Analysis of transpeptidation reaction with both the acceptor peptides proceeded in a similar fashion, and the transpeptidation reaction was found to be independent of calcium ion (Figure 1C). Thus, both versions of the enzyme namely, Δ81SpSrtA and Δ59SpSrtA, were active.

SpSrtA crystallizes as a 3D-swapped dimer

Both truncated forms of sortase, namely, Δ59SpSrtA and Δ81SpSrtA, could be crystallized. However, Δ59SpSrtA crystals were recalcitrant to diffraction. The preliminary account of crystallization and data collection at 2.9 Å resolution for the wild type Δ81SpSrtA was reported earlier [17]. In the present work, diffraction data of Δ81SpSrtA crystals was collected at 2.7 Å (data collection and processing statistics in Table 1, 2). The crystal asymmetric unit contains two independent homodimers which are not related by

crystallographic symmetry (Figure 2A). The modeled structure of Δ81SpSrtA comprises residues 82-247 for the three chains (A, C and D), 92-247 for the fourth chain (B) and 54 water molecules (residues 75-81 in the structure are derived from the expression vector). Interpretable electron density is present for the backbone of all residues within the chains; however, clear electron density is absent for some of the atoms in the side chains. While the overall topology of the four chains is similar (H1-β1-H2-β2-H3-β3-β4-H4-β5-β6-H5-β7-β8-H6)(Figure S1B), chains A, C and D have an additional 3_{10} helix each between β4 and β5 (residues 149-151), and between β6 and β7 strands (residues 191-193); on the other hand, chain B displays an additional 3_{10} helix beyond β8 at the C-terminal of the chain (residues 228-230). The N-terminal helix in chain D of Δ81SpSrtA does not align with the helix in the other monomers. Residues 86-91 form a short helix, while the portion before it assumes beta strand structure that is aligned with a small beta strand formed at the C-terminal end of chain D. This feature is absent in chains A and C (interpretable electron density is not observed for residues preceding Leu92 in chain B), indicating flexibility at the N- and C-termini.

Unlike the structures of other SrtA homologs (from *Streptococcus pyogenes*, PDB id: 3fn5; *Streptococcus agalactiae*, PDB id: 3rcc [35] and *Staphylococcus aureus*, PDB id: 1t2p [36] (X-ray structure), 2kid [37] (NMR structure)), the Δ81SpSrtA monomer alone does not form the conserved eight-stranded beta-barrel sortase-fold, which is unique to the sortase superfamily. One monomer exchanges the C-terminal end, consisting of the β7 and β8 strands and a helix, with another monomer, thereby forming a 3D-swapped dimer. There are two such dimers in the asymmetric unit, comprising of chains A and B, and chains C and D respectively. Strands 1-6 from one monomer and the swapped strands 7-8 from the other monomer constitute the 8-stranded beta-barrel fold. Residues Pro187, Asp188, Arg189 and Val190 in the β6/β7 loop are estimated to form the hinge region based on the differences in structural superposition with its closest unswapped homolog, *S. pyogenes* SrtA (Table S4),

and also from the differences in torsion angles of the equivalent residues between the two structures (Table S5). It may be noted that Pro residues have been found to be present in the hinge loop in many 3D-swapped proteins observed so far.

Structural superposition of the four Δ81SpSrtA monomers show that the stretch from the N-terminal up to the hinge region has RMSD ~ 0.4 Å, while the swapped stretch alone has higher RMSD (>1 Å), indicating that the swapped stretch is dynamic and hence oriented differently in the different chains. This RMSD is comparable to that for the superposition of monomers of SrtA homologs – *S. pyogenes* (PDB id: 3fn5, chain B) and *S. agalactiae* (PDB id: 3rcc, chain N) (RMSD = 1.773 Å for an alignment length of 129). Analysis of domain motion revealed that there is a rotation of the swapped β7-β8 strands about the larger domain (β1-β6 strands) by ~29° (Figure 2B). B-factors indicate flexibility in the loops, with the highest B-factors occurring in the hinge region (Figure S5) and in the stretch beyond β8 strand. Similar flexibility has been observed in *S. agalactiae* SrtC1 where there are subtle differences among the different chains in the asymmetric unit.

A strong dimer interface is formed between chains A and B, and between chains C and D, whereas weaker interfaces (crystallographic interfaces) are formed between A and C, A and D, and B and C. This is evident from the large interface area (> 10000 $Å^2$) and large number of inter-chain hydrogen bonds and van der Waals contacts across the AB and CD interfaces (interface statistics in Table S7). As a result of the dimerization via 3D-swapping, a secondary interface is formed by residues 113-120 (in β2/β3 loop), 125 (in β3/β4 loop), 143-146 (part of β4/H4 loop after active site His141), 187-190 (hinge region in β6/β7 loop), residues 205-212 (β7/β8 loop containing active site Cys207) and active site Arg215 (Figure 2C). Interestingly, the equivalent regions (β2/β3, β4/H1 and β7/β8 loops) in *Bacillus anthracis* SrtC (PDB id: 2ln7) have been implicated in the formation of a transient dimer [38].

It is pertinent to mention here that the structure of *S. pneumoniae* D39 Sortase A (PDB id: 5dv0) which shares 98% sequence identity with our SpSrtA (4o8l), and used 4o8l

as the starting model for molecular replacement, displays one molecule per asymmetric unit. The data for 5dv0 has been reported at a poor resolution of 3.3 Å and the description of the structure has not been published thus far. However, PISA analysis shows that the assembly of 5dv0 represents a thermodynamically stable homodimer with a buried surface area of 10710 $Å^2$ (ΔG of -61.0 kcal/mol) which is similar to the present (4o8l) structure (Table S7). The structure of 5dv0 when viewed as biological assembly with the cyclic symmetry C2 in PyMol reveals a 3D-swapped structure arising from exchange of the β7-β8 strands of the monomer (Figure S6).

Structural features of the active site

The active site of sortases is comprised of a catalytic triad of highly conserved His, Cys and Arg residues. In prototypic *S. aureus* SrtA, the active site is composed of Cys184, His120 and Arg197.The equivalent residues in SpSrtA are His141 (at C-terminal of β4), Cys207 (at C-terminal of β7) and Arg215 (at N-terminal of β8). We carried out mutation of each putative active site residue and generated three single mutants (H141A, C207A and R215A respectively) of Δ81SpSrtA. The mutants were found to be inactive in standard transpeptidation assays, establishing the critical role of individual residues in catalysis.

The Cys207 residue is found to be oxidized to its sulphenic acid form in chain B (Figure S7); a similar feature has been observed for one of the structures of *S. pyogenes* SrtA (PDB id: 3fn6). 3D-swapping leads to two sets of active site residues at the open interface (Figure 3A). His141 from one monomer, and Cys207 and Arg215 from the other monomer of a swapped dimer constitute a pair of active site residues. In the present structure, the Arg215 residue is partially exposed, whereas the Cys207 and His141 residues are buried, unlike the SrtA from *S. pyogenes* and *S. agalactiae*. The superposition of the active site residues from the four chains in the crystallographic asymmetric unit depicts some flexibility in the side chains of Cys207 and Arg215 (Figure 3B) but that of His141 is rigid. Of the two catalytic Arg residues (Arg215) at the dimeric interface, one is found to interact with

Asp209 in the other protomer of the swapped dimer (Figure 3C). The other Arg215 in each dimer is not involved in the salt-bridge interaction. Notably, structural superposition of SpSrtA dimers (4o8l) onto *S. pneumoniae* D39 Sortase A (5dv0) symmetrical dimer also shows a similar disposition of the catalytic residues (Figure S6).

Several co-crystallization trials and crystal soaking experiments with a variety of peptide substrates based on LPNTG pentapeptide motif was attempted to glean structural insights into substrate recognition and catalysis. However, none of the datasets showed the presence of any electron density for the peptide substrate. To eliminate the possibility of hydrolysis of the enzyme-substrate intermediate in the crystallization condition, crystallization trials and crystal soaking experiments were carried out with the active site mutant Δ81SrtA(C207A). The asymmetric unit in this case was found to contain three independent swapped homodimers (Figure S8) and a glycerol molecule, but no density for the peptide substrate. The final model comprises residues 75-247 for chain A, residues 91-247 in chain B, residues 73-247 in chain C, and residues 74-247 in chains D, E and F, and 243 water molecules (residues 73-81 in the structure are derived from the expression vector). There is no missing electron density within the backbone of the protein chains; however, some of the atoms in the side chains do not have clear electron density. The different orientation of the N-terminus observed in chain D of the wild type protein is absent in the C207A mutant structure.

Comparison with other known SrtA structures

The individual domains in the swapped dimer are similar to the sortase fold observed in homologous structures. SpSrtA has the highest sequence similarity with SrtA from *S. pyogenes* (78% sequence similarity, 67% identity; PDB id: 3fn5), followed by that from *S. agalactiae* (77% sequence similarity, 58% identity; PDB id: 3rcc). Figure 4 shows the structure-based sequence alignment of SpSrtA with the other sortases of known structure. The RMSD for superposition of chain B of 3fn5 on the larger domain of chain A (up to the

hinge region) is 0.717 Å when residues 89-187 (*S. pneumoniae* numbering) are used, whereas, the RMSD is 1.104 Å (over an alignment length 152) when the 3fn5 monomer is superposed onto a complete sortase domain in *S. pneumoniae* formed by portions of 2 different chains. The C-terminus of β6 strand from chain A (start of β6/β7 loop), the C-terminus of β7 from chain B and the N-terminus of β8 from chain B in SpSrtA do not overlap with the equivalent regions in the *S. pyogenes* SrtA (PDB id: 3fn5, chain B). The β7/β8 loop region from the four monomers in the swapped SpSrtA dimer also does not superpose either among themselves or with the *S. pyogenes* SrtA, pointing to its conformational flexibility. Notably, the β7/β8 loop occupies the active site in the swapped dimer and LAATE sequence stretch of this loop mimics the LPXTG peptide substrate. This is reminiscent of the *S. agalactiae* SrtC structure (PDB id: 3tb7) in which the loop of one molecule occupies the active site of a neighbouring molecule.

Absence of a Ca^{2+} binding site

The prototype sortase, *S. aureus* SrtA has a flexible β6/β7 loop. A 3_{10} helix is formed in β6/β7 loop upon substrate-binding, which immobilizes the loop and leads to the formation of hydrophobic contacts around the Leu in LPXTG sorting motif. This is further stabilized by Ca^{2+}-binding to the C-terminus region of the loop [39]. This has led to the idea of an 'induced fit' mechanism for its activity. In contrast, *S. pyogenes* SrtA and *S. agalactiae* SrtA have a preformed substrate-binding site that doesn't require large scale conformational changes for substrate-binding [40].

Comparison of *S. aureus* SrtA and Δ81SpSrtA structure reveal that calcium binding residues in *S. aureus* SrtA (PDB id: 2kid), namely, Glu105, Glu108, Asp112 and Glu171, are substituted in Δ81SpSrtA by Lys126, Gln129 and Glu133 from one monomer, and Asp195 from the swapped arm of the other monomer in the dimer (Figure S9). In *S. aureus* SrtA, the Ca^{2+} ion balances electrostatic repulsion among the Glu105, Glu108, Glu171 and Asp112 at the site. In Δ81SpSrtA, NZ atom of Lys126 occupies the place of Ca^{2+}. Lys126 interacts via

hydrogen bonds with Gln129 and Asp195, and with Asn135 instead of Glu133, which has moved away from the equivalent position occupied by Asp112 in *S. aureus* SrtA. The above description is consistent with the calcium- independent transpeptidation activity of Δ81SpSrtA shown in Figure 1C. The *S. pyogenes* SrtA behaves in much the same way as Δ81SpSrtA and does not require Ca^{2+} for its activity. The equivalent residues in *S. pyogenes* SrtA are Lys126, Gln129, Gly133 and Asp196.

Δ81SpSrtA exists predominantly as a dimer in solution

We performed size-exclusion chromatography to assess the oligomerization status of Δ81SpSrtA in solution (Figure 5A). The purified preparation of Δ81SpSrtA, when subjected to chromatography on a Superdex200 column, eluted as two peaks (elution volumes of ~ 16.0 ml and 17.3 ml respectively). Based on the mass of a standard calibration protein mixture, mass of the proteins in the two peaks were calculated as 41 kDa and 24 kDa, respectively. These masses fit well with the expected mass of a monomer (~21 kDa) and a dimer (42 kDa) of the enzyme.

To further investigate the oligomeric status of Δ81SpSrtA, individual peaks were re-chromatographed on the same column (Figure 5A, inset). For this, concentrated samples prepared from each peak, comprising protein fractions pooled from rising portion of the dimeric peak and falling portion of the monomeric peak, respectively, were analysed to avoid cross-contamination. Under identical conditions, the monomeric protein sample resolved into two peaks corresponding to the monomeric and a dimeric form indicating the existence of a monomer-dimer equilibrium. In contrast, the dimeric fraction eluted in a single peak as a dimer suggesting that dimerization results in a gain of stability.

The strength of dimerization of Δ81SpSrtA was further probed by analytical equilibrium sedimentation ultracentrifugation (Figure 5B). Experiments were performed using two concentrations of the protein (20 μM and 40 μM) and at four rotor speeds (13000,

17000, 23000 and 27000 rpm) at 20 °C. Scans were taken at 280 nm at different time-points and absorbance data obtained at the end of the experiment was analysed individually by fitting to monomer-dimer equilibrium model (equation 1-4 in methods). Figure 5B represents the equilibrium state data fitted into a monomer-dimer model. The overall dissociation constant (K_d) of 3.4± 0.1 μM was obtained.

Mapping of distance between catalytic Cys207 residues in the 3D-swapped dimer

The oligomeric status of Δ81SpSrtA as seen through size-exclusion chromatography and sedimentation equilibrium measurements established that the enzyme predominantly forms dimer in solution. It was therefore, pertinent to investigate if the swapped dimer observed in the crystal also existed in solution. Based on the crystal structure, the distance between the thiol groups of Cys207 in chains A & B of the swapped dimer was calculated to be 13.0 Å and those between chains C & D was 8.8 Å and 11.4 Å (Figure 6). We employed maleimide thiol-cross-linkers (Figure 6B) of varying spacer length (8 Å, 10.2 Å, 13 Å, 14.7 Å and 17.8 Å) to map the distance between thiol group of catalytic Cys207 residues in the protomers.

Chemical-crosslinking reaction of Δ81SpSrtA was carried out at 37 °C for 15 min with sortase to crosslinker molar ratio of 2:1. The reaction was quenched by addition of 50 mM DTT and analysed by SDS-PAGE (Figure 6C). Interestingly, a predominant intense band (Figure 7C, lane 3) corresponding to a dimer of Δ81SpSrtA, was observed with the sample treated with Bismaleimidohexane (BMH, 13 Å spacer) in accord with the crystal structure based Cys207-Cys207 distance (13.04 Å) in A and B protomers of the swapped dimer. Intensity of the dimeric band in Δ81SpSrtA sample treated with other cross-linkers (lanes 2, 4 and 5), corresponding to 10.2 Å, 14.7 Å, and 17.8 Å respectively, was marginal *albeit* relatively higher with 8 Å (lane1) presumably due to its favourable access to the Cys residues in Chains C and D (8.8 Å) of the swapped dimer. The above cross-linking data appear

consistent with the crystal structure data (Figure 6A).

The cross-linking reaction was also carried out in the presence of 8M urea to ensure that the facile cross-linking of Δ81SpSrtA with BMH (13 Å) was a consequence of 3-dimensional structure of the protein. Interestingly, SDS-PAGE of Δ81SpSrtA incubated with various cross-linkers in the presence of 8 M urea showed the presence of a predominant dimeric species in all samples (Figure 6D) suggesting non-specific, random conjugation.

Further, reaction of Δ81SpSrtA(C207A) mutant with the above cross-linking molecules was also studied to evaluate the thiol specificity of the reagents. Δ81SpSrtA(C207A) mutant was expected to be inert to cross-linking due to the absence of Cys residue. Cross-linking experiments of Δ81SpSrtA(C207A) mutant were performed in much the same way as the wild type enzyme. SDS-PAGE analysis of the reaction mixture (Figure 6E) revealed the presence of only a single protein band (monomeric) in case of both untreated mutant enzyme as well as the cross-linker treated samples. Absence of a dimeric species or other high-molecular weight bands indicated that the chemical cross-linking reaction was highly specific to Cys207 residue. Taken together, the cross-linking data provide compelling evidence for the existence of a 3D-swapped dimeric species in solution as well.

DISCUSSION

Sortase enzymes are implicated in bacterial pathogenesis, and also serve as effective protein labelling tool [8]. Therefore, structure and specificity elucidation as a prelude to developing clinical inhibitors or improved sortases for protein engineering applications remains a topical subject in sortase research. Here we report the peptide substrate specificity and crystal structure of housekeeping SrtA from *S. pneumoniae.* The housekeeping SrtA encoded, along with pilus subunits (RrgA, B, and C) and pilin sortases (SrtC1, SrtC2, SrtC3), in the *rlrA* pathogenicity islet is apparently the most crucial

transpeptidase among *S. pneumoniae* sortases because of its additional role in the assembly of pathogenic pilus [11].

We generated two N-terminal truncated versions of SrtA (Δ59SpSrtA and Δ81SpSrtA) to explore the limit of truncation for expression of a soluble protein endowed with catalytic activity. We examined the donor substrate specificity of the above constructs with the classic LPNTG motif using AAKY or GGGKY as an acceptor peptide. Curiously, end-capped LPNTG pentapeptide (acetylated/amidated) or Ala extended heptapeptide (ALPNTGA), which was earlier shown to be a robust substrate for archetypal housekeeping SrtA of *S. aureus* [34], turned out to be poor substrates for Δ59SpSrtA. In contrast, transpeptidation yield with an octapeptide (AQLPNTGA) or a nonapeptide (YAQLPNTGA) was found to be almost two-fold higher (13% vs 28%) than that of the heptapeptide. Interestingly, Δ81SpSrtA also displayed a donor LPXTG substrate preference similar to Δ59SpSrtA and produced comparable transpeptidation yields against both, AAKY and GGGKY, acceptors. Notably, only Gly-based amine-acceptors, such as GGGKY, are effective substrates for quintessential housekeeping SaSrtA. Besides, unlike SaSrtA, transpeptidation reaction of SpSrtA was independent of calcium ion. The cumulative results indicate that substrate recognition propensity of SpSrtA is quite different from quintessential housekeeping SrtA of *S. aureus* [34].

Although both constructs of SpSrtA were endowed with catalytic activity, only Δ81SpSrtA produced good quality crystals for structure analysis. Interestingly, Δ81SpSrtA displayed a 3D-swapped structure comprising of two independent homodimers in the asymmetric unit wherein each protomer generated a characteristic 'sortase fold' by exchanging the β7-β8 strands. Size exclusion chromatography indicated the predominant presence of a dimeric species in solution in equilibrium with a monomer, and sedimentation equilibrium measurements yielded a K_d in low micromolar range (3.4 ± 0.1 μM) reflecting enhanced stability of the dimeric form. The distance between thiol groups of Cys207 in protomers A/B, and that of the protomer C/D of the 3D-swapped dimer (as seen in the

asymmetric unit of the crystal structure) mapped with cross-linking spacers of 13 Å and 8.8 Å, respectively, indicating the existence of the swapped dimer in solution as well.

Is the 3D-swapped dimer of Δ81SpSrtA catalytically competent? 3D-swapping generates two sets of active site residues at the dimer interface in which His141 from one chain with Cys207 and Arg215 of the other constitute a pair of active site residues. However, His141 and Cys207 are buried, and Arg215 of one chain forms a salt-bridge with Asp209 of the other chain. PISA analyses show that chain A forms a strong dimer with chain B (total buried surface area of 10300 $Å^2$) with a solvation free energy gain of -59.8 kcal/mol, and chain C forms another strong dimer with chain D (buried surface area of 11020 $Å^2$) with a solvation free energy gain of -60.2 kcal/mol. Apparently, the dimer interface is quite strong and stable, and the active site may not be accessible to the substrate. This perhaps could also be one of the reasons for our failure to crystallize SpSrtA with peptide substrates. Taken together, these considerations lead us to conclude that the monomer of SpSrtA is the catalytically active species, and perhaps, the sole active species. This conclusion appears consistent with the recent work of Nikghalb et al [41] who reported enhanced transpeptidase activity in a sample preparation of Δ80SpSrtA that was enriched in its monomeric form.

The occurrence of a 3D-swapped dimer as observed in SpSrtA is not known among the sortase family of transpeptidases but dimer-monomer equilibrium has been reported in some cases [42-44]. Sortase A of *S. aureus* (SaSrtA) has been shown to exist in a monomer-dimer equilibrium in solution [42]. However, functional evaluation of individual species of SaSrtA produced curious results [45]; while *in vitro* studies showed increased enzyme activity in the dimeric species of SaSrtA, dimer disruptive mutants of SaSrtA were found to elicit enhanced surface display of associated substrate proteins indicating monomer as the relevant catalytic form of SaSrtA *in vivo*. Notwithstanding the identity of the active species, monomer-dimer equilibrium may be envisioned as a mechanism for regulation of *in vivo* enzyme activity. The low micromolar K_d of SpSrtA suggest that a large fraction of enzyme in

micromolar concentration range (Figure S10) would exist as dimer (>90% dimer at 100 μM). However, it is probable that the equilibrium bias toward the dimer could be a consequence of truncation. Furthermore, shifting of the equilibrium to a monomeric active form by preferential binding of a hitherto unknown cell surface effector molecule to the dimer is also conceivable. In this connection, it would be interesting to explore if the membrane anchorage propensity of SpSrtA remains unaffected by truncation and whether the full-length protein can form a 3D-swapped dimer.

In summary, our results demonstrate that SpSrtA prefers relatively longer LPXTG peptide substrates as compared to the archetypal SrtA of *S. aureus* and other housekeeping sortases. Results of biochemical experiments indeed demonstrate the predominant presence of a dimer, and existence of a 3D-swapped dimer as observed in the crystal structure. However, active site residues appear tightly packed at the interface, and are buried, rendering them inaccessible to a peptide substrate. This is also borne out by our unsuccessful attempts to crystallize a peptide substrate with wild type or active site Cys207Ala mutant of SpSrtA. However, 3D-swapping observed in Cys207Ala mutant emphasizes the natural tendency of SpSrtA to assume this type of structure. The causation of 3D-swapping of SpSrtA, unprecedented among sortases, may have regulatory consequences given that the enzyme is meant to perform housekeeping functions as well as a critical role in the pilus assembly of *S. pneumoniae*.

Competing interest

The authors declare no conflict of interest.

Funding

This work was supported by core grants to the National Institute of Immunology from the Department of Biotechnology (DBT), Government of India (GOI) and J. C. Bose National Fellowship grant (SB/S2/JCB-016/2016) to RPR from the Department of Science and Technology (DST, GOI). X-ray data were collected at the X-ray Facility for Structural Biology

at the Molecular Biophysics Unit, Indian Institute of Science, Bangalore (supported by DST and DBT, GOI), and at the BM14 beamline, European Synchrotron Radiation Facility (ESRF) Grenoble, France.

Acknowledgements

The helpful discussions with Dr S. Kumaran in measurement and analysis of analytical ultracentrifugation data is gratefully acknowledged. TB and SD received research fellowship from CSIR, India. AM received research fellowships from UGC and DBT, GOI.

Abbreviations

Sp: *Streptococcus pneumoniae*, SrtA: Sortase A, SpSrtA: *Streptococcus pneumoniae* sortase A, Δ59SpSrtA: SpSrtA construct with truncation of 59 residues from the N-terminal, Δ81SpSrtA: SpSrtA construct with truncation of 81 residues from the N-terminal, RP-HPLC: reverse-phase high performance liquid chromatography, Fmoc: Fluoromethoxycarbonyl, NMP: N-methyl-2-pyrrolidone. RrgA (Rlr-regulated gene A), RrgB (Rlr-regulated gene B) and RrgC (Rlr-regulated gene C) are pilus constituents of *S. pneumoniae*.

References

1 Cartwright, K. (2002) Pneumococcal disease in western Europe: Burden of disease, antibiotic resistance and management. *Eur. J. Pediatr.* **161**, 188-195.

2 Kausmally, L., Johnsborg, O., Lunde, M., Knutsen, E. and Håvarstein, L. S. (2005) Choline-binding protein D (CbpD) in Streptococcus pneumoniae is essential for competence-induced cell lysis. *J. Bacteriol.***187**, 4338–4345.

3 Gosink, K. K., Mann, E. R., Guglielmo, C., Tuomanen, E. I. and Masure, H. R. (2000) Role of novel choline binding proteins in virulence of Streptococcus pneumoniae. *Infect. Immun*. **68**, 5690–5695.

4 Shaper, M., Hollingshead, S. K., Benjamin, W. H. and Briles, D. E. (2004) PspA protects Streptococcus pneumoniae from killing by apolactoferrin, and antibody to PspA enhances killing of pneumococci by apolactoferrin. *Infect. Immun.* **72**, 5031–5040.

5 Bierne, H., Mazmanian, S. K., Trost, M., Pucciarelli, M. G., Liu, G., Dehoux, P., Jänsch, L., Garcia-del Portillo, F., Schneewind, O. and Cossart, P. (2002) Inactivation of the srtA gene in Listeria monocytogenes inhibits anchoring of surface proteins and affects virulence. *Mol. Microbiol.* **43**, 869–881.

6 Rosenow, C., Ryan, P., Weiser, J. N., Johnson, S., Fontan, P., Ortqvist, A. and Masure, H. R. (1997) Contribution of novel choline-binding proteins to adherence, colonization and immunogenicity of Streptococcus pneumoniae. *Mol. Microbiol.* **25**, 819–829.

7 Navarre, W. W. and Schneewind, O. (1999) Surface poteins of gram-positive bacteria

and mechanisms of their targeting to the cell wall envelope. *Microbiol. Mol. Biol. Rev.* **63**, 174–229.
8 Spirig, T., Weiner, E. M. and Clubb, R. T. (2011) Sortase enzymes in gram-positive bacteria. *Mol. Microbiol*. **82**, 1044–1059.
9 Hendrickx, A. P. A., Budzik, J. M., Oh, S.-Y. and Schneewind, O. (2011) Architects at the bacterial surface - sortases and the assembly of pili with isopeptide bonds. *Nat. Rev. Microbiol.* **9**, 166–176.
10 Marraffini, L. A., Dedent, A. C. and Schneewind, O. (2006) Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria. *Microbiol. Mol. Biol. Rev.* **70**, 192–221.
11 Fälker, S., Nelson, A. L., Morfeldt, E., Jonas, K., Hultenby, K., Ries, J., Melefors, Ö., Normark, S. and Henriques-Normark, B. (2008) Sortase-mediated assembly and surface topology of adhesive pneumococcal pili. *Mol. Microbiol.* **70**, 595–607.
12 LeMieux, J., Woody, S. and Camilli, A. (2008) Roles of the sortases of Streptococcus pneumoniae in assembly of the RlrA pilus. *J. Bacteriol.* **190**, 6002–6013.
13 Kharat, A. S. and Tomasz, A. (2003) Inactivation of the srtA gene affects localization of surface proteins and decreases adhesion of Streptococcus pneumoniae to human pharyngeal cells in vitro. *Infect. Immun.* **71**, 2758–2765.
14 Paterson, G. K. and Mitchell, T. J. (2006) The role of Streptococcus pneumoniae sortase A in colonisation and pathogenesis. *Microbes Infect*. **8**, 145–153.
15 Chen, S., Paterson, G. K., Tong, H. H., Mitchell, T. J. and DeMaria, T. F. (2005) Sortase A contributes to pneumococcal nasopharyngeal colonization in the chinchilla model. *FEMS Microbiol. Lett*. **253**, 151–154.
16 Neiers, F., Madhurantakam, C., Fälker, S., Manzano, C., Dessen, A., Normark, S., Henriques-Normark, B. and Achour, A. (2009) Two Crystal Structures of Pneumococcal Pilus Sortase C Provide Novel Insights into Catalysis and Substrate Specificity. *J. Mol. Biol.* **393**, 704–716.
17 Misra, A., Biswas, T., Das, S., Marathe, U., Sehgal, D., Roy, R. P. and Suryanarayanarao, R. (2011) Crystallization and preliminary X-ray diffraction studies of sortase A from Streptococcus pneumoniae. *Acta Crystallogr. Sect. F*. **67**, 1195–1198.
18 Leslie, A. G. W. and Powell, H. R. (2007) In: *Read R.J., Sussman J.L. (eds) Evolving Methods for Macromolecular Crystallography*. NATO Science Series, vol **245**. Processing diffraction data with mosflm, pp 41–51, Springer, Dordrecht.
19 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., et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr*. **67**, 235-242.
20 Vaguine, A. A., Richelle, J. and Wodak, S. J. (1999) SFCHECK: A unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr. Sect. D Biol. Crystallogr*. **55**, 191–205.
21 McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674.
22 Race, P. R., Bentley, M. L., Melvin, J. A., Crow, A., Hughes, R. K., Smith, W. D., Sessions, R. B., Kehoe, M. A., McCafferty, D. G. and Banfield, M. J. (2009) Crystal structure of Streptococcus pyogenes sortase A: implications for sortase mechanism. *J. Biol. Chem.* **284**, 6924–6933.
23 Murshudov, G. N., Vagin, A. A. and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **53**, 240-255.

24 Emsley, P. and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132.
25 Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Echols, N., Headd, J. J., Hung, L. W., Jain, S., Kapral, G. J., Grosse Kunstleve, R. W., et al. (2011) The Phenix software for automated determination of macromolecular structures. *Methods*, **55**, 94-106.
26 Brünger, A. T. (1992) Free R value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–475.
27 Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. and Richardson, D. C. (2010) MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 12–21.
28 Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr*. **26**, 283–291.
29 Krissinel, E. and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2256–2268.
30 Hayward, S., Kitao A. and Berendsen, H. J. C. (1997) Model-free methods of analyzing domain motions in proteins from simulation: a comparison of normal mode analysis and molecular dynamics simulation of lysozyme. *Proteins*, **27,** 425-437.
31 Krissinel, E. and Henrick, K (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774--797.
32 Shingate, P. and Sowdhamini, R. (2012) Analysis of domain-swapped oligomers reveals local sequence preferences and structural imprints at the linker regions and swapped interfaces. *PLoS One*, **7,** e39305.
33 Bennett, M. J., Schlunegger, M. P. and Eisenberg, D. (1995) 3D domain swapping: A mechanism for oligomer assembly. *Protein Sci*. **4**, 2455-2468.
34 Biswas, T., Pawale, V. S., Choudhury, D. and Roy, R. P. (2014) Sorting of LPXTG Peptides by Archetypal Sortase A: Role of invariant substrate residues in modulating the enzyme dynamics and conformational signature of a productive substrate. *Biochemistry* **53**, 2515–2524.
35 Khare, B., Krishnan, V., Rajashankar, K. R., I-Hsiu, H., Xin, M., Ton-That, H. and Narayana, S. V. (2011) Structural differences between the Streptococcus agalactiae housekeeping and pilus-specific sortases: SrtA and SrtC1. *PLoS One*. **6**, e22995.
36 Zong, Y., Bice, T. W., Ton-That, H., Schneewind, O. and Narayana, S. V. L. (2004) Crystal structures of Staphylococcus aureus sortase A and its substrate complex. *J. Biol. Chem.* **279**, 31383–31389.
37 Suree, N., Liew, C. K., Villareal, V. A., Thieu, W., Fadeev, E. A., Clemens, J. J., Jung, M. E. and Clubb, R. T. (2009) The Structure of the *Staphylococcus aureus* sortase-substrate complex reveals how the universally conserved LP*X*TG sorting signal is recognized. *J. Biol. Chem.* **284**, 24465–24477.
38 Robson, S. A., Jacobitz, A. W., Phillips, M. L. and Clubb, R. T. (2012) Solution structure of the sortase required for efficient production of infectious *Bacillus anthracis* Spores. *Biochemistry* **51**, 7953–7963.
39 Naik, M. T., Suree, N., Ilangovan, U., Liew, C. K., Thieu, W., Campbell, D. O., Clemens, J. J., Jung, M. E. and Clubb, R. T. (2006) Staphylococcus aureus Sortase A transpeptidase. Calcium promotes sorting signal binding by altering the mobility and structure of an active site loop. *J. Biol. Chem.* **281**, 1817–1826.
40 Weiner, E. M., Robson, S., Marohn, M. and Clubb, R. T. (2010) The Sortase A enzyme that attaches proteins to the cell wall of *Bacillus anthracis* contains an unusual active site architecture. *J. Biol. Chem.* **285**, 23433–23443.

41 Nikghalb, K. D., Horvath, N. M., Prelesnik, J. L., Banks, O. G. B., Filipov, P. A., Row, R. D., Roark, T. J. and Antos, J. M. (2018) Expanding the scope of sortase-mediated ligations by using sortase homologues. *ChemBioChem* **19**, 185–195.
42 Wallock-Richards, D. J., Marles-Wright, J., Clarke, D. J., Maitra, A., Dodds, M., Hanley, B. and Campopiano, D. J. (2015) Molecular basis of Streptococcus mutans sortase A inhibition by the flavonoid natural product trans-chalcone. *Chem. Commun.* **51**, 10483-85.
43 Lu, C., Zhu, J., Wang, Y., Umeda, A., Cowmeadow, R. B., Lai, E., Moreno, G. N., Person, M. D. and Zhang, Z. (2007) Staphylococcus aureus sortase A exists as a dimeric protein in vitro. *Biochemistry*. **46**, 9346-54.
44 Zhu, J., Lu, C., Standland, M., Lai, E., Moreno, G. N., Umeda, A., Jia, X. and Zhang, Z. (2008) Single mutation on the surface of Staphylococcus aureus Sortase A can disrupt its dimerization. *Biochemistry*. **47**, 1667-74.
45 Zhu, J., Xiang, L., Jiang, F. and Zhang, Z. J. (2016) Equilibrium of sortase A dimerization on Staphylococcus aureus cell surface mediates its cell wall sorting activity. *Exp. Biol. Med.* **241**, 90-100.
46 Gouet, P., Robert, X. and Courcelle, E. (2003) ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res*. **31**, 3320–3323.

Legends to Figures

Figure 1. Substrate preference of ∆59SpSrtA and ∆81SpSrtA. Transpeptidation reactions were carried out using 1 mM LPXTG donor peptide and 2 mM AAKY or GGGKY in the presence of 100 μM sortase at 37 °C for indicated time. The reaction was quenched by addition of 10-fold excess 0.1% TFA and analysed by RP-HPLC. A) HPLC-trace for ∆59SpSrtA -catalyzed transpeptidation reaction of Ac-LPNTG-NH_2, ALPNTGA, QLPNTGA, AQLPNTGA or YAQLPNTGA and AAKY. B) HPLC profile of the ∆81SpSrtA-catalyzed transpeptidation reaction of YAQLPNTGA and AAKY or GGGKY. C) Time course of ∆81SpSrtA-catalyzed transpeptidation reaction of YAQLPNTGA and AAKY or GGGKY in the presence and absence of Ca^{2+}.

Figure 2. *S. pneumoniae* SrtA structure. A) Asymmetric unit containing two dimers AB and CD (chains A, B, C & D are shown in green, blue, yellow and red colours, respectively). Each dimer contains two characteristic 8-stranded beta barrel “sortase fold”, but each fold in the dimer is constituted by the combination of beta-strands from both the chains, that is, a complete fold is made through 3D-swapping. Folded catalytic domains are labeled as protomer-A, protomer-B, protomer-C and protomer-D depending upon the chain identifier contributing first six beta strands. B) Flexibility of the monomers in the 3D-swapped dimer. Superposition of the larger domain (β1-β6 strands) in the swapped monomers chain A (Green) and chain C (Blue) through DynDom server (Hayward ***et al***., 1997) showed a rotation of ~29º around the rotation axis C) Primary and Secondary interfaces in the swapped dimer. The primary interface is formed by sections of two monomers in the 3D-swapped structure, whereas it would be part of the same monomer in an unswapped structure. The secondary interface is the new interface formed due to the proximity of the two protomers in the dimer.

Figure 3. A) Active site residues (two sets) in AB dimer at the secondary interface. B) Structural comparison of active site residues within the four sortase protomers. Side chains of Cys207 and Arg215 are flexible but His141 is relatively rigid (SpSrtA: Green A-chain, Blue B-chain, Yellow C-chain and Pink D-chain). C) Salt-bridge interactions between Arg-215 and Asp-209 from opposite chains at the secondary interface in the (a) 3D-swapped AB and (b) CD dimers. In both the dimers only one Arg-Asp residue pair involving catalytic Arg-215 shows salt bridge interaction, and presumably, enhances the stability of the 3D-swapped dimeric structure.

Figure 4. Multiple sequence alignment of SpSrtA where sequence numbering starts with 82 which is the first protein residue in the construct. The secondary structures of SrtA (A-chain) are shown on top of the sequence alignment, where coils and arrows represent helices and strands, respectively. α, β, η and TT correspond to α-helix, β-strand, 3_{10}-helix and β-turn respectively. The residues shown with blue background correspond to the active site residues. Sequence alignment was done with ClustalW (http://www.ebi.ac.uk/clustalw) and coloured using ESPRIPT [46].

Figure 5. Oligomerization status of SpSrtA. A) Size-exclusion chromatography (SEC) of ∆81SpSrtA was performed on a Superdex200 (10/300 GL) column pre-equilibrated with 50 mM Tris-HCl (pH 7.5) containing 150 mM NaCl at a flow rate of 1 ml/min. ∆81SpSrtA eluted as two peaks with an elution volume of 16.0 ml and 17.3 ml respectively. Inset shows re-chromatography of the SEC eluted peaks. a) The monomer fractions were pooled from multiple runs, concentrated, and reloaded onto the pre-equilibrated Superdex200 (10/300 GL) column. b) The dimer fraction pooled from multiple runs was concentrated and reloaded. B) Sedimentation equilibrium measurements. Data was collected at four rotor speeds (13000, 17000, 23000, and 27,000 rpm) at 20 μM and 40 μM respectively. The solid line represents the best fit of absorbance data to a monomer-dimer model yielding an overall dissociation constant of about 3.4 μM. The residuals for both concentrations and four speeds are shown in the lower panel.

Figure 6. Cys-Cys distance mapping in 3D-swapped dimer of ∆81SpSrtA. A) Based on the crystal structure, a distance of 13.04 Å was observed between thiol group of Cys207 of the dimers composed of chains A (marked as C207a) & B (marked as C207b). The corresponding Cys-Cys distance in Chain C (C207c) and Chain D (Cys207d) is 8.86 Å and 11.48 Å. B) Chemical structure and spacer length description of the cross-linking molecules. (C-E) SDS-PAGE analyses of cross-linking reaction. C) ∆81SpSrtA in neat buffer, D) ∆81SpSrtA pre-treated with 8M urea, and E) mutant ∆81SpSrtA(C207A). The lane labelled NT indicates “not treated” (control sample), other lanes (labelled 1-5) are sortase samples treated with cross-linkers as defined in panel B.

Table 1 Summary of the crystallographic data-collection and processing statistics for the Δ81SpSrtA and its mutant Δ81SpSrtA(C207A) crystals.

	Δ81SpSrtA	Δ81SpSrtA(C207A)
Wavelength (Å)	1.5418	0.9772 (Synchrotron)
Resolution range (Å)	32.08-2.70 (2.85-2.70)*	47.15-2.48 (2.61-2.48)*
Crystal system	Monoclinic	Monoclinic
Space group	$P2_1$	$C2$
Unit cell parameters		
a (Å)	66.94	155.57
b (Å)	103.45	113.33
c (Å)	74.87	81.34
β (°)	115.65	90.80
Volume ($Å^3$)	467377.9	1433947.1
Temperature (K)	100	100
Mosaicity (°)	1.52	0.93
Total number of reflections	187659 (26715)	207901 (29473)
Unique reflections	26168 (3621)	50010 (7245)
Multiplicity	7.5 (7.4)	4.2 (4.1)
Mean $\langle I/\sigma(I)\rangle$	14.3 (3.8)	10.8 (2.5)
Completeness (%)	99.2 (97.3)	99.9 (99.7)
R_{merge}† (%)	9.4 (52.9)	9.2 (54.7)
Overall *B* factor from Wilson plot ($Å^2$)	66.5	53.9
No. of molecules in the asymmetric unit	4 (Two dimers)	6 (Three dimers)
V_M ($Å^3\,Da^{-1}$)	2.79	2.86
Solvent content (%)	56	57

† $R_{merge} = \Sigma_{hkl}\Sigma_i|I_i(hkl) - \langle I(hkl)\rangle|/ \Sigma_{hkl}\Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the i^{th} observation of reflection *hkl* and $\langle I(hkl)\rangle$ is the average intensity over all observations.

*Values in parentheses are for the last resolution shell.

Table 2: Refinement statistics and the model parameters.

Refinement statistics	Δ81SpSrtA	Δ81SpSrtA(C207A)
Number of protein residues	666	1027
Number of protein atoms	5221	8012
Number of solvent water	54	234
Number of ligands	-	1 (Glycerol)
R_{work}%	18.098	18.245
R_{free} %	23.386	22.388
RMS deviations		
Bond distance (Å)	0.012	0.010
Bond angles (°)	1.162	1.226
Luzzati coordinate error (Å)		
Working set	0.363	0.302
Ramachandran plot statistics (%)		
Residues in allowed regions	90.6	90.7
Residues in additionally allowed regions	8.2	7.6
Residues in generously allowed regions	1.2	1.4
Residues in disallowed regions	0.0	0.2
Mean B-factors (Overall, Å^2)	60.53	43.67

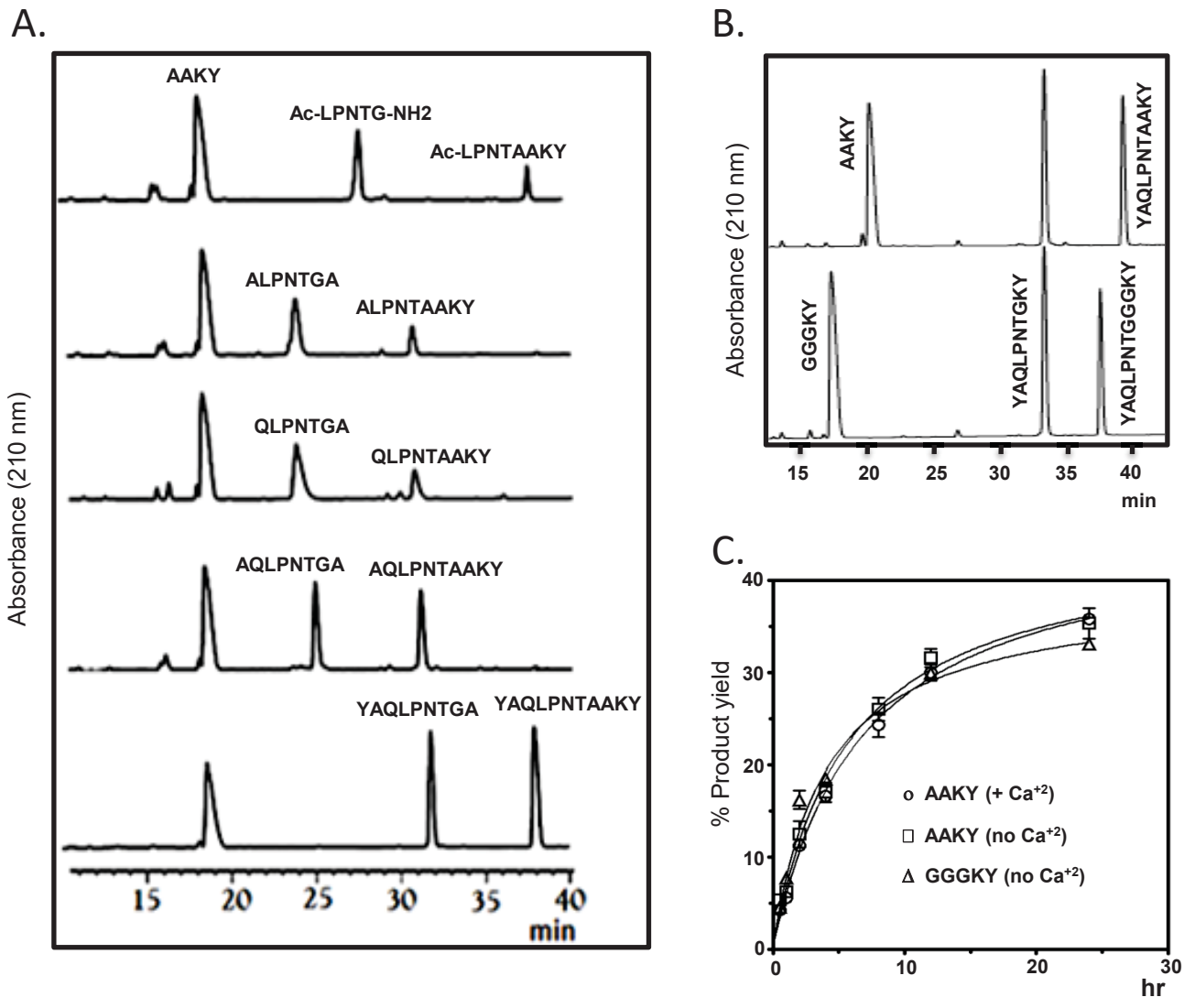

Figure 1

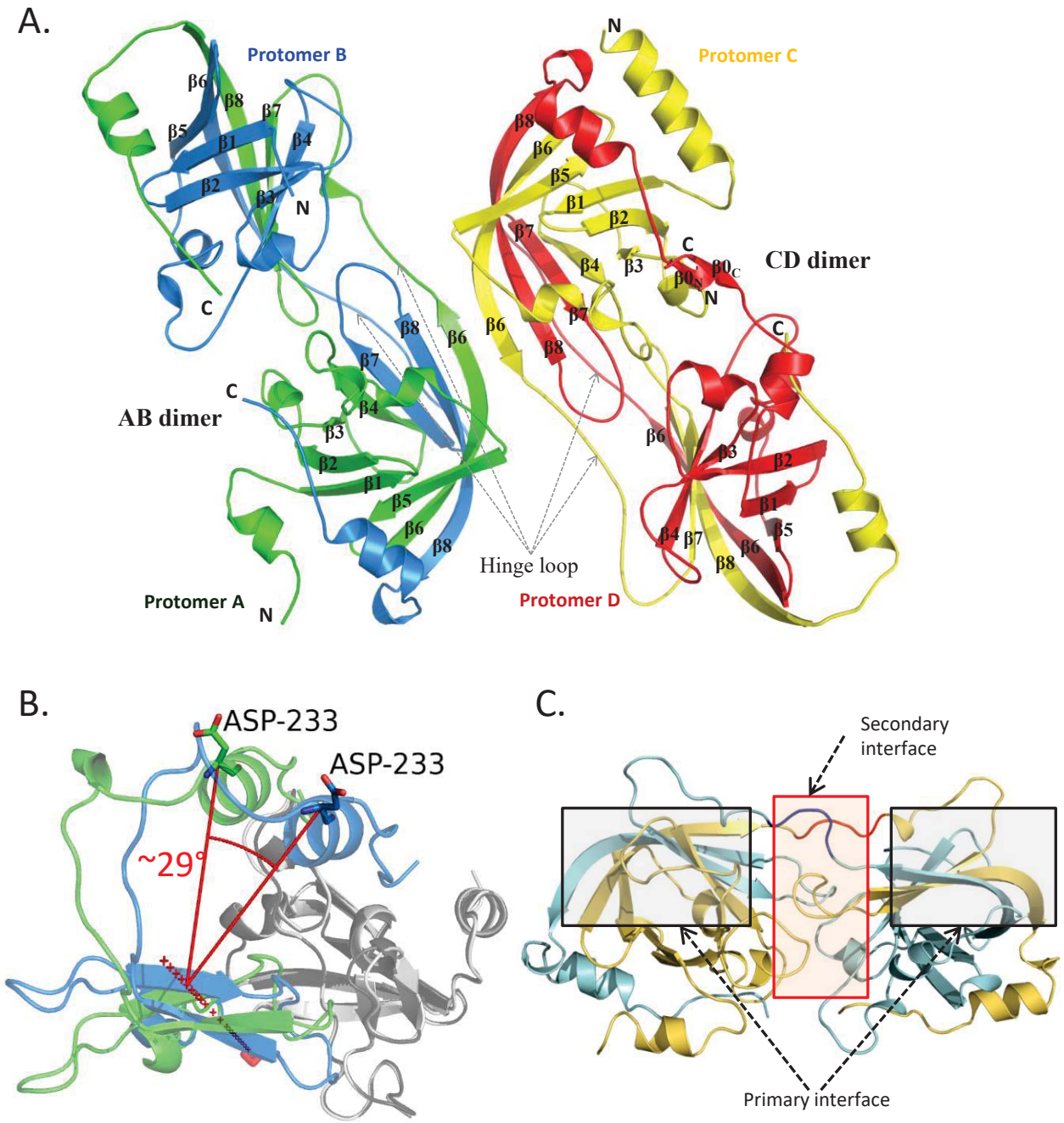

Figure 2

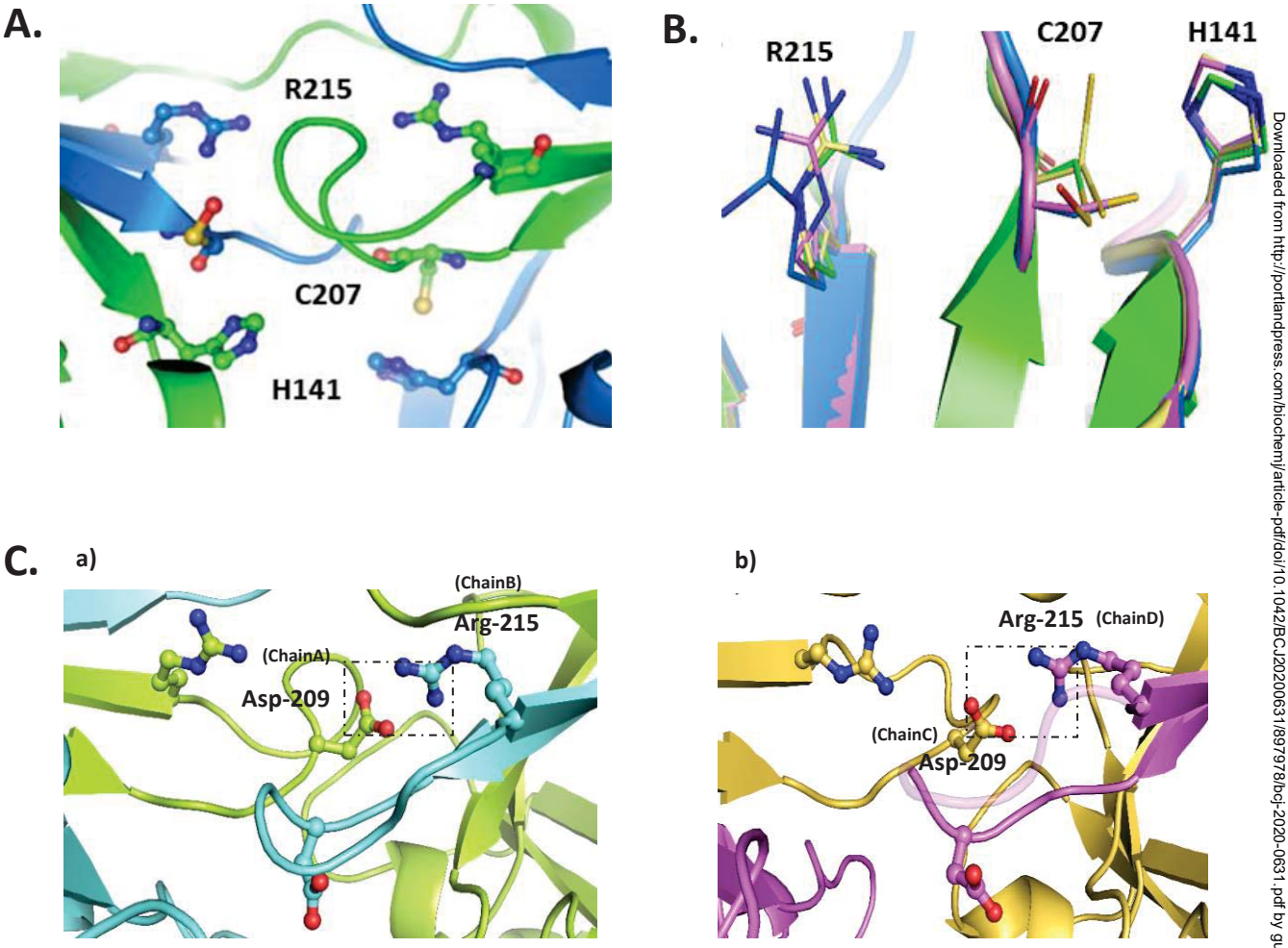

Figure 3

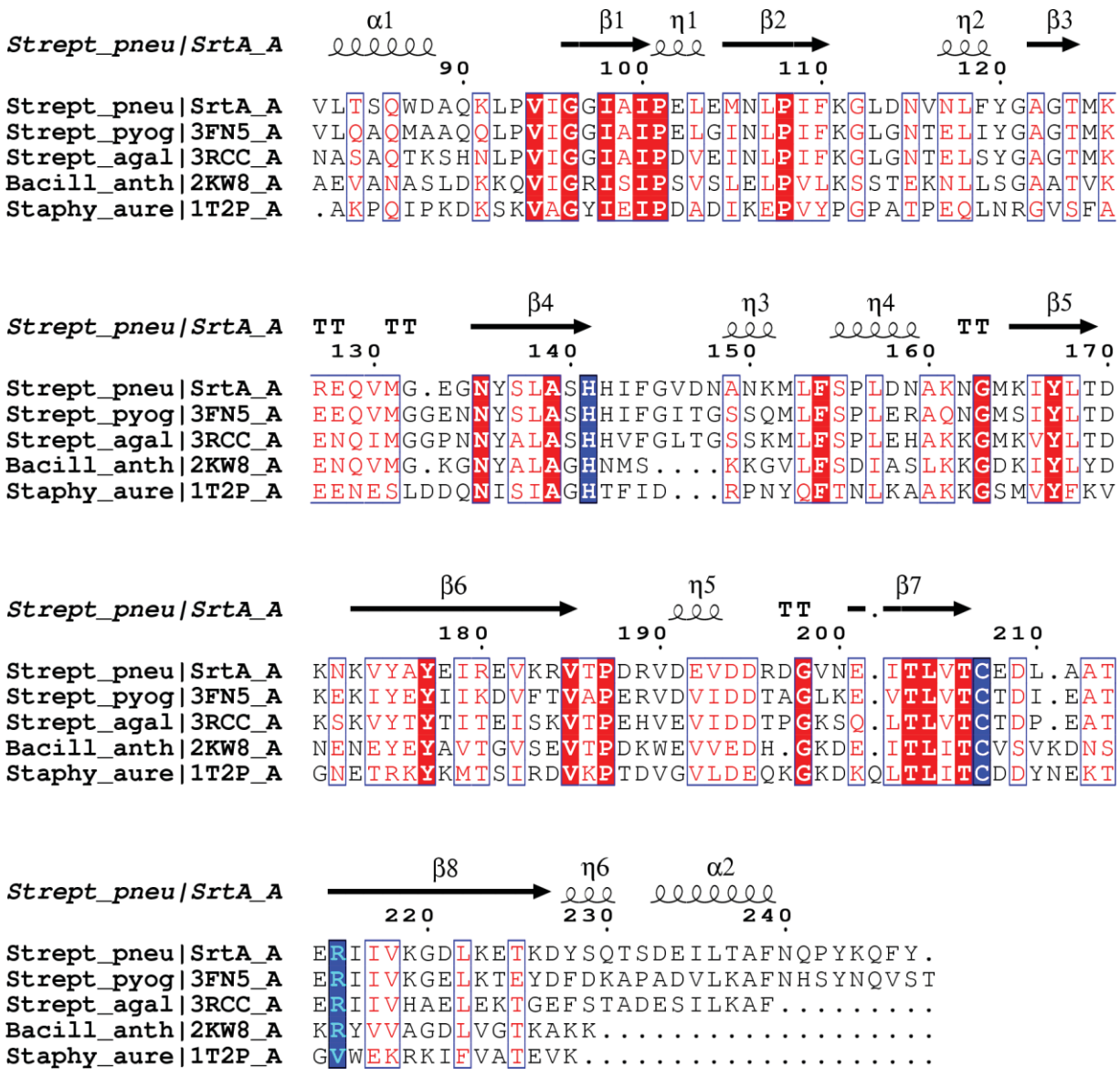

Figure 4

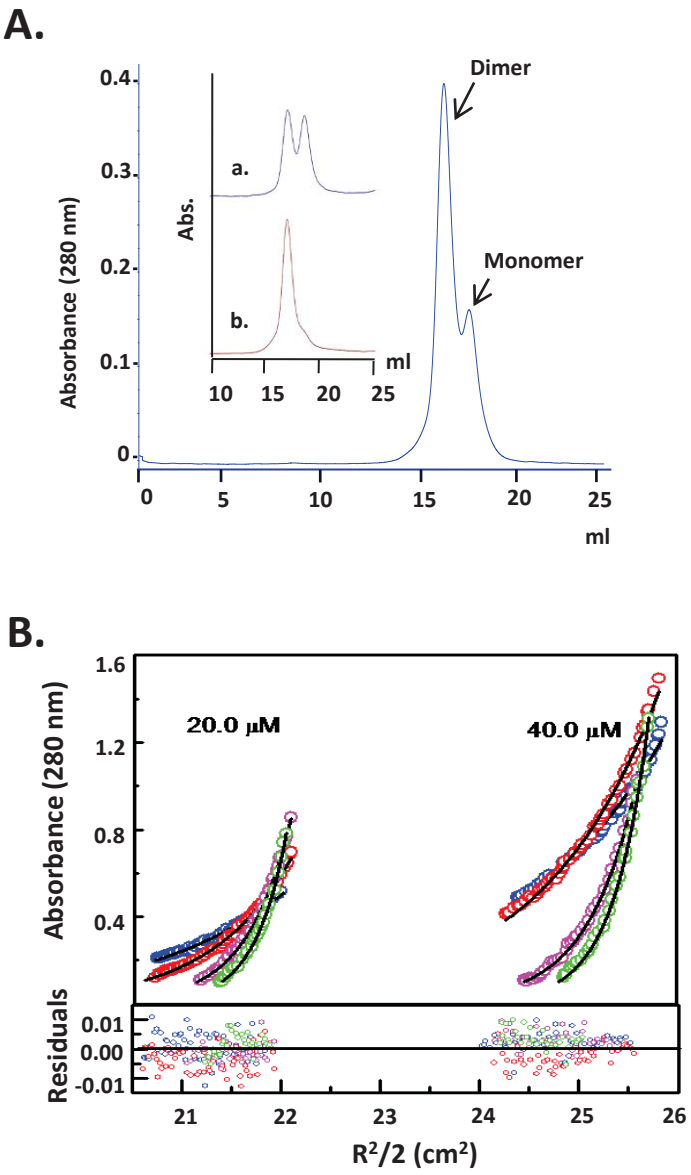

Figure 5

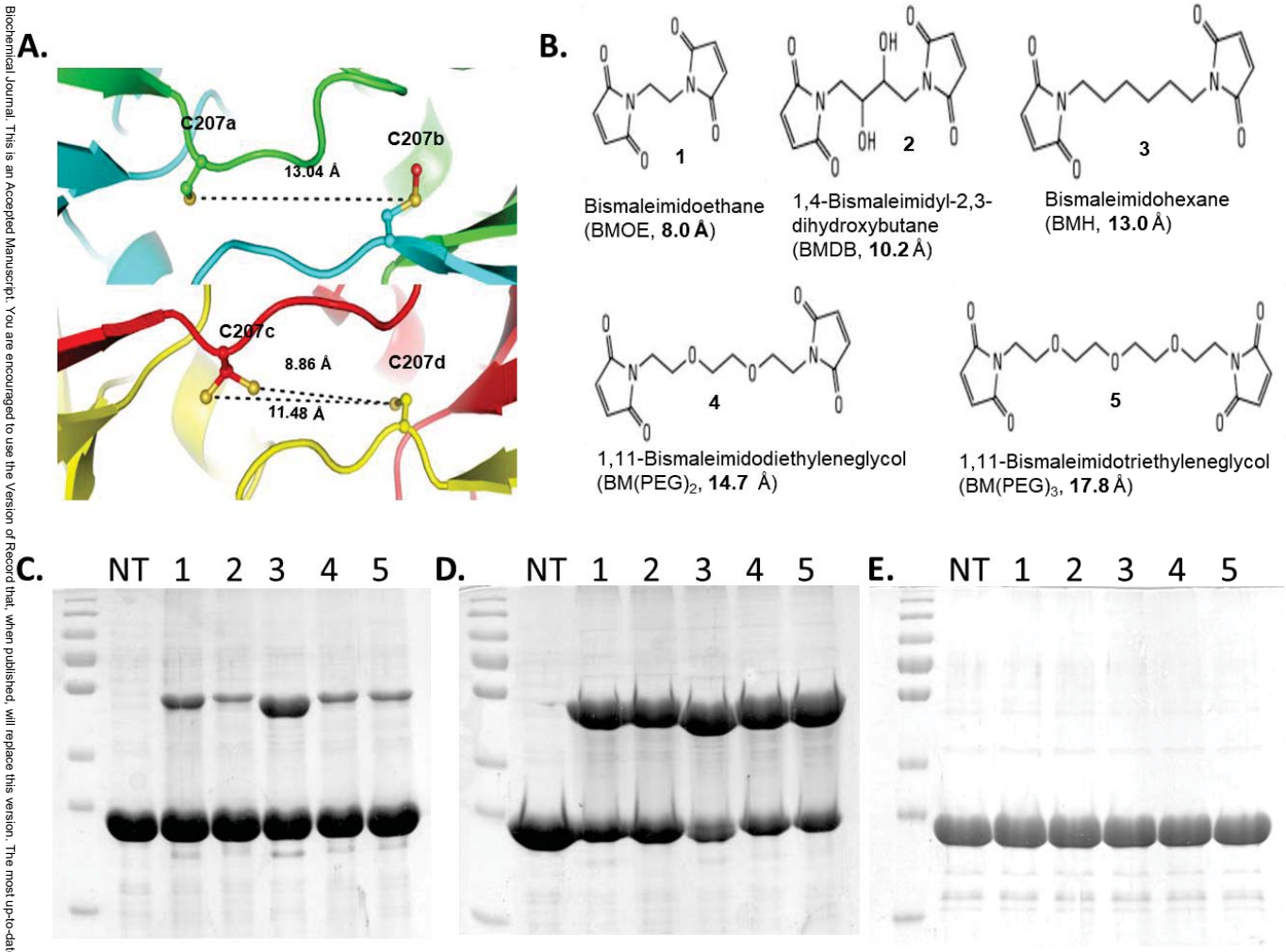

Figure 6