



Solution structure of the Big domain from *Streptococcus pneumoniae* reveals a novel Ca²⁺-binding module

Tao Wang¹, Jiahai Zhang¹, Xuecheng Zhang², Chao Xu³ & Xiaoming Tu¹

¹Hefei National Laboratory for Physical Sciences at Microscale, School of Life Sciences, University of Science and Technology of China, Hefei, Anhui 230026, P.R. China, ²School of Life Sciences, Anhui University, Hefei, Anhui 230039, P.R. China, ³Structural Genomics Consortium, University of Toronto, 101 College St., Toronto, Ontario, M5G 1L7, Canada.

SUBJECT AREAS:
SOLUTION-STATE NMR
PROTEINS
PATHOGENS
CALCIUM

Received
28 August 2012

Accepted
3 December 2012

Published
16 January 2013

Correspondence and
requests for materials
should be addressed to
X.M.T. (xmtu@ustc.
edu.cn)

Streptococcus pneumoniae is a pathogen causing acute respiratory infection, otitis media and some other severe diseases in human. In this study, the solution structure of a bacterial immunoglobulin-like (Big) domain from a putative *S. pneumoniae* surface protein SP0498 was determined by NMR spectroscopy. SP0498 Big domain adopts an eight- β -strand barrel-like fold, which is different in some aspects from the two-sheet sandwich-like fold of the canonical Ig-like domains. Intriguingly, we identified that the SP0498 Big domain was a Ca²⁺ binding domain. The structure of the Big domain is different from those of the well known Ca²⁺ binding domains, therefore revealing a novel Ca²⁺-binding module. Furthermore, we identified the critical residues responsible for the binding to Ca²⁺. We are the first to report the interactions between the Big domain and Ca²⁺ in terms of structure, suggesting an important role of the Big domain in many essential calcium-dependent cellular processes such as pathogenesis.

S *treptococcus pneumoniae* is a gram-positive pathogen that can cause acute respiratory infection, otitis media and some other severe diseases in human¹. The completed sequencing of the *S. pneumoniae* genome provides valuable information which contributes a lot to the research on the diseases caused by this pathogen². Surface proteins are supposed to play key roles in the pathogenesis of *S. pneumoniae*^{3,4}. Bacterial surface proteins consist of diverse groups which are thought to be involved in a variety of important processes when a pathogen invades the host body, such as adhesion to and invasion into the host cells, interaction with the immune system and so on. The surface proteins facilitate pathogens' invading the host by interacting with the surface of host cells^{5,6}. Therefore, the characterization of their structures may advance the understanding of some details of the pathogenesis.

The genome of a virulent isolate of *S. pneumoniae*, TIGR4, was analyzed, and several groups of proteins were identified as potential surface proteins according to the typical domains or motifs present on surface proteins². SP0498 is a putative endo-beta-N-acetylgluco-saminidase possessing an LPxTG motif at the C-terminus, which is the symbol of surface proteins discovered in many other bacteria. The LPxTG motif is recognized by the sortase which is responsible for the covalent attachment of specific proteins to the cell wall of gram-positive bacteria^{6,7}. SP0498 is a ~182 KD protein composed of multiple domains (Figure 1A), one of which is a bacterial immunoglobulin-like (Big) domain (group 3).

Immunoglobulin super family is composed of a number of classical proteins involved in immune system (such as immunoglobulins, Fc fragment, T-cell receptors, and so on) and proteins participating in other important processes (such as cell recognition, adhesion, inter-cellular interaction and so on)^{8,9}. Immunoglobulin-like (Ig-like) domain proteins are also found in bacteria and these bacterial immunoglobulin-like (Big) domains disperse among proteins with diverse functions¹⁰. Although the functional and structural studies on the eukaryotic Ig-like domains are extensively carried out, information of the Big domains in prokaryotes is relatively lagging behind.

In this study, we determined the solution structure of SP0498 Big domain by NMR spectroscopy and identified it as a novel Ca²⁺-binding module. Furthermore, we explored its potential binding sites for calcium by chemical shift perturbation, ITC and Stains-all assays. Mutation of residues H68, Y69, G71, H72 and E73 to alanines abolished the Ca²⁺-binding ability of SP0498 Big domain, indicating these residues are involved in and critical for the Ca²⁺ binding.

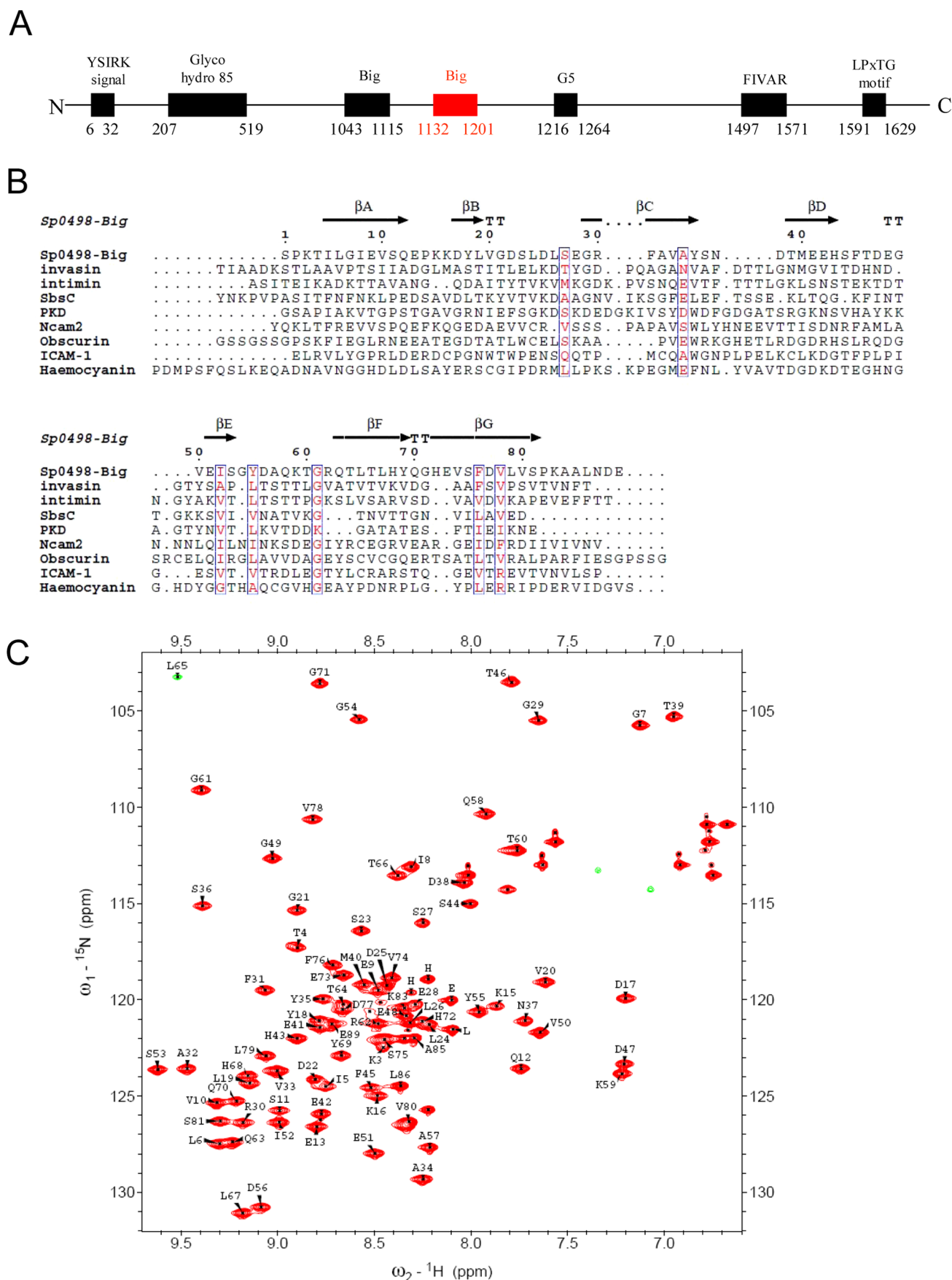


Figure 1 | (A) Diagram of the complete domain organization of SP0498. (B) Sequence alignment of SP0498 Big domain with other Ig-like domains by ClustalW and decorated using ESPrict^{31,32}. (C) ¹H-¹⁵N HSQC spectrum of SP0498 Big domain with the assigned residues.



Results

SP0498 Big domain shares low sequence similarity with other Ig-like domains. Ig-like domains are widespread in diverse proteins from the immunoglobulin to other divergent bacterial proteins. They share low sequence homology as a result of the diversity of evolution. SP0498 Big domain also shows very low homology compared with other Ig-like domains, with the sequence identity less than 20% and sequence similarity less than 30% (Figure 1B). Although Ig-like domains share low sequence similarity, they have similar secondary structure pattern, which is composed of 7–10 β -strands.

Solution structure of SP0498 Big domain. SP0498 Big domain was recombinantly expressed and purified. ^1H - ^{15}N HSQC spectrum of the recombinant SP0498 Big domain is shown in Figure 1C. Almost all the residues of SP0498 Big domain were identified: 99% of ^{15}N , ^{13}C and ^1H in the backbone, and 93% of ^{13}C and 99% of ^1H in the side chain were assigned. The solution structure of SP0498 Big domain was determined by a set of standard NMR spectra. The chemical shifts of atoms from SP0498 Big domain have been deposited into the Biological Magnetic Resonance Data Bank (accession number 17381), and the atomic coordinates of the calculated structures were deposited in Protein Data Bank with the PDB ID code of 2L7Y. The assembly of 20 lowest-energy structures and a best representative structure are shown in Figure 2A. The statistics of the calculated structures are listed in Table 1. The residues in the

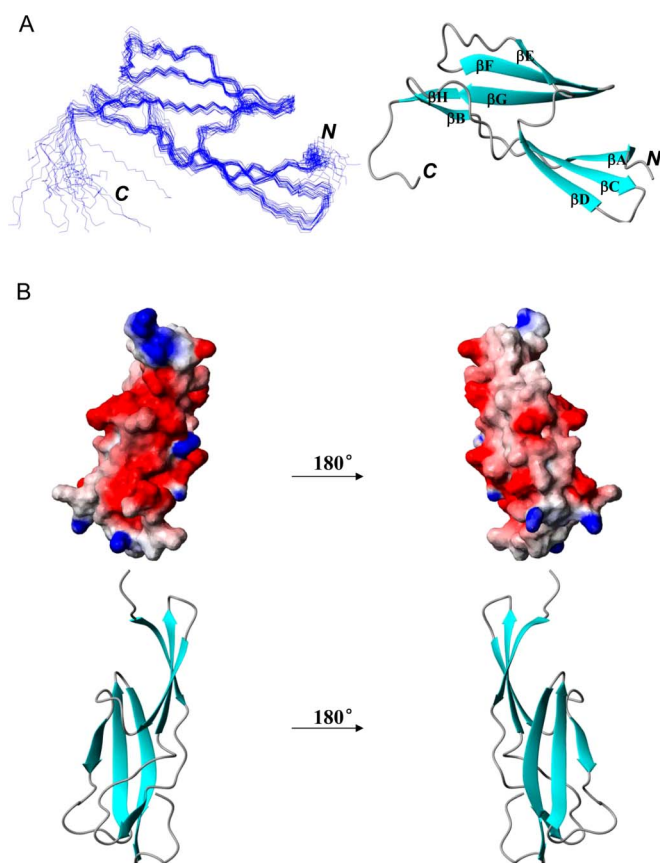


Figure 2 | NMR structure of SP0498 Big domain. (A) Ribbon representation of the minimized averaged structure of SP0498 Big domain with the secondary structure elements highlighted (left) and backbone superposition of 20 selected conformers with the lowest energy from the final CYANA calculation (right). This figure was produced with MOLMOL. (B) Electrostatic surface diagram of SP0498 Big domain. The surface color represents the magnitude of the electrostatic potential: red, negative; blue, positive; white, neutral.

Table 1 | NMR structural statistics

NMR restraints in the structure calculation	
Intraresidue	225
Sequential ($ i - j = 1$)	405
Medium-range ($ i - j < 5$)	106
Long-range ($ i - j \geq 5$)	366
Hydrogen bonds	38
Total distance restraints	1140
Dihedral angle restraints	90
Residual violations	
CYANA target functions, \AA	1.08 ± 0.21
NOE upper distance constrain violation	0.22 ± 0.09
Maximum, \AA	1 ± 1
Number $> 0.2 \text{\AA}$	
Dihedral angle constrain violations	
Maximum, $^\circ$	3.05 ± 0.71
Number $> 5^\circ$	0 ± 0
Vander Waals violations	
Maximum, \AA	0.24 ± 0.04
Number $> 0.2 \text{\AA}$	1 ± 1
Average structural rmsd to the mean coordinates, \AA	
Secondary structure backbone ^a	0.46
Secondary structure heavy atoms ^a	1.33
All backbone atoms ^b	0.63
All heavy atoms ^b	1.45
Ramachandran statistics, % of all residues	
Most favored regions	76.6
Additionally allowed regions	19.5
Generously allowed regions	2.6
Disallowed regions	1.3

^aIncludes residues in secondary structure.

^bObtained for residues L6–S81.

secondary structure are well defined with the RMSD of 0.46 \AA for the backbone of these regions. However, the loops were not well restrained because long-range restraints identified were relatively few.

SP0498 Big domain adopts a barrel-like conformation composed of eight β -strands which form three relatively separated regions: βA (residues 7–9), βC (residues 31–34) and βD (residues 41–43) constitute a short β -sheet composed of all anti-parallel strands at one side of the structure; C-terminal βH (residues 79–81) and N-terminal βB (residues 18–20) pack together in a parallel pattern at the other side of the structure; βE (residues 51–53), βF (residues 63–69) and βG (residues 72–78) form another anti-parallel β -sheet. Meanwhile, electrostatic surface analysis of SP0498 Big domain structure was carried out for better understanding its structural characteristics (Figure 2B). The surface of the SP0498 Big domain is composed of residues with negative charge dominantly. Positive charge is rarely present on the surface of SP0498 Big domain except the N-terminus and certain individual sites.

Structural comparison of SP0498 Big domain with classical Ig-like domains.

Structural comparison reveals that the solution structure of SP0498 Big domain is different from those of classical Ig-like domain family members, illustrated by the comparison between SP0498 Big domain and one Ig-like domain of the human Fc fragment (Figure 3A–C). The classical Ig-like domains adopt a Greek Key pattern which is conserved among the members of this super family^{11,12}. They are all composed of β -strands, number of which varies from 7 to 10. The β -strands are arranged into two separate sheets connected by a conserved disulfide bond. The strands of A, B, C, E, F and G are generally conserved in classical Ig-like domains, and the strands of B, C, E and F constitute a hydrophobic core, among which a disulfide bond is formed by cysteines in the strands of B and F. Distinctive from the classical Ig-like domain, SP0498 Big domain displays a barrel-like shape instead of typical two-sheet sandwich-like one. The two

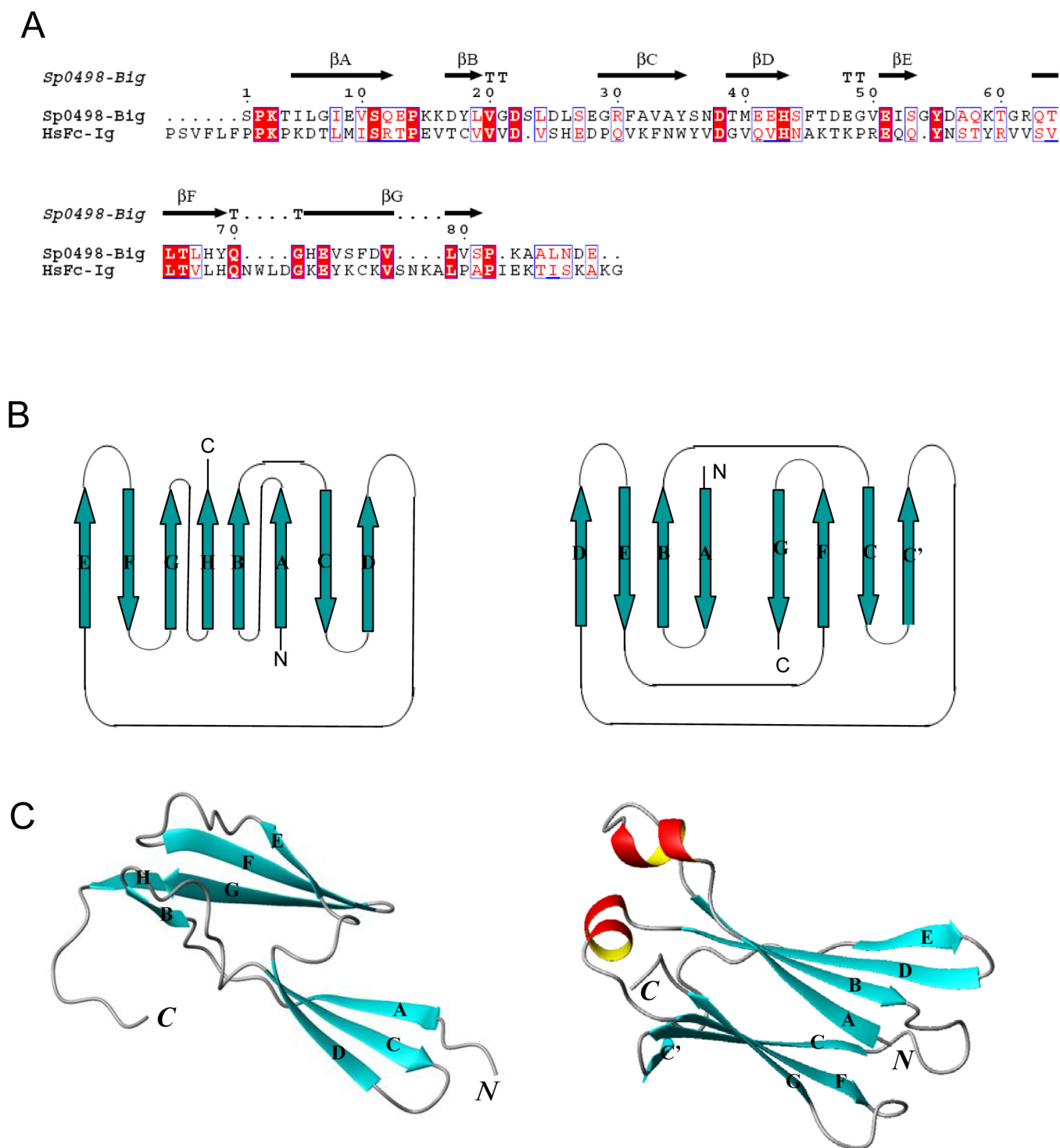


Figure 3 | Structural comparison between SP0498 Big domain and a classical Ig-like domain (human Fc fragment, HsFc). (A) Sequence alignment of SP0498 Big domain and HsFc fragment; (B) Topology diagram of Ig-like structures from SP0498 Big domain (left) and HsFc fragment (right); (C) Ribbon structures of SP0498 Big domain (left, PDB code of 2L7Y) and HsFc fragment (right, PDB code of 1FC1).

anti-parallel β -sheets, with few connections between each other, are not positioned in a face-to-face pattern adopted by classical Ig-like domains. Besides, an additional pair of parallel strands present in SP0498 Big domain is not identified in the classical Ig-like domains (Figure 3B & C). Furthermore, the conserved disulfide bond is not identified in SP0498 Big domain either. Finally, although both SP0498 Big domain structure and classical Ig-like fold are composed of β -strands, the strands in SP0498 Big domain are relatively shorter^{13,14}.

Identification of Big domain as a novel Ca^{2+} -binding domain. The interactions between SP0498 Big domain and Ca^{2+} were identified by Stains-all assay¹⁵. After incubated with SP0498 Big domain, the CD spectrum of the Stains-all showed a J-band around 650 nm which indicated an obvious Ca^{2+} -binding ability of SP0498 Big domain (Figure 4A). The binding was further confirmed by ITC assay and the value of the dissociation constant (K_d) was determined to be $0.24 \pm 0.02 \mu M$, which indicated the strong interactions between SP0498 Big domain and Ca^{2+} (Figure 4B). Additionally, we cloned and

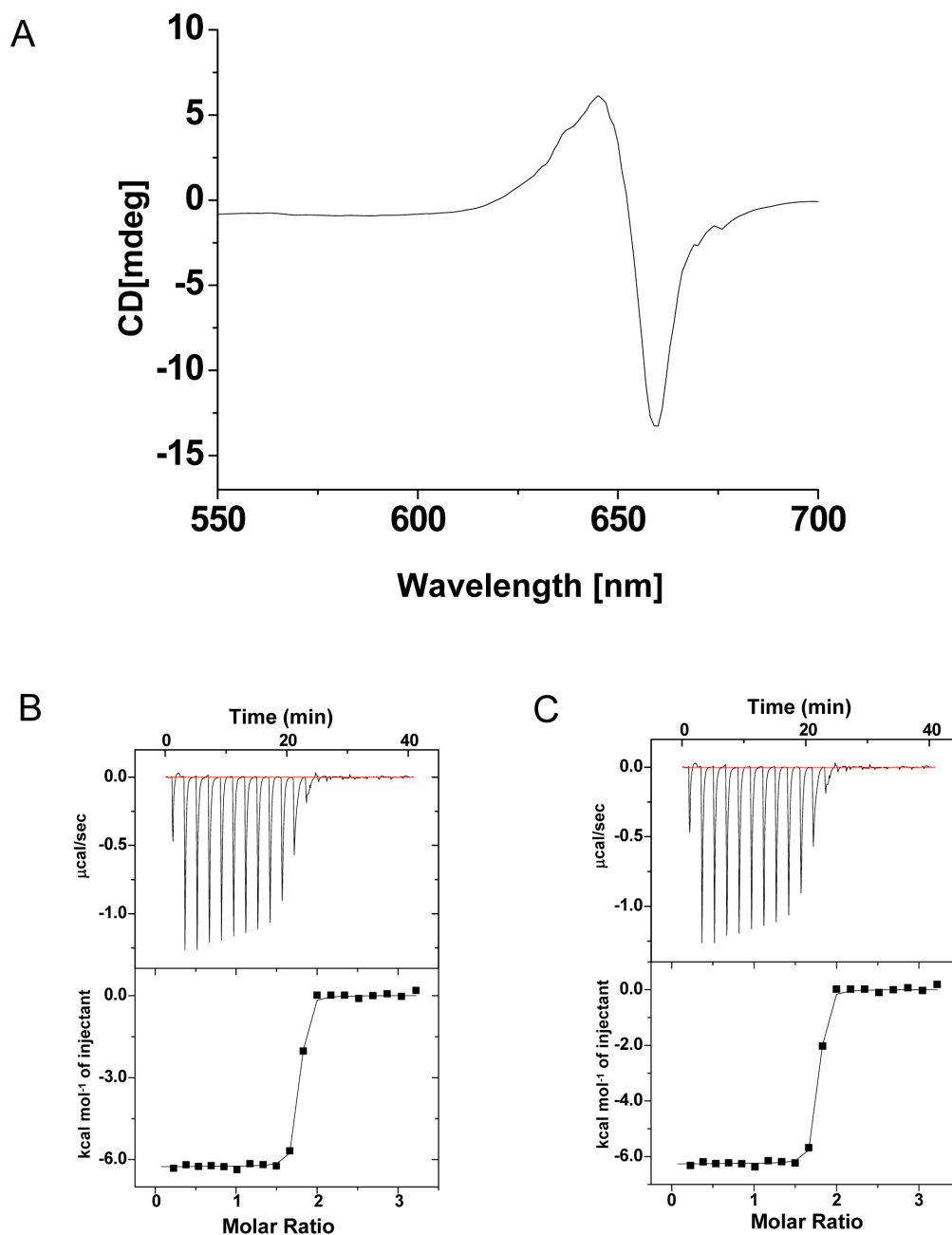


Figure 4 | The interactions between Big domains and Ca^{2+} . (A) CD spectrum of SP0498 Big domain mixed with Stains-all dye at molar ratio of 10 : 1. (B) Titration of SP0498 Big domain with Ca^{2+} , measured by ITC in 20 mM HEPES buffer (containing 100 mM NaCl) at pH 6.5 and 20°C. The upper thermogram panel shows the observed heats for each injection of CaCl_2 at 120 s intervals after baseline correction, whereas the lower panel depicts the binding enthalpies vs. Ca^{2+} /protein molar ratio. (C) Titration of Lig A3 (the third Big domain of Lig A with the accession number of FJ030917) with Ca^{2+} , measured by ITC in 20 mM HEPES buffer (containing 100 mM NaCl) at pH 6.5 and 20°C. The upper thermogram panel shows the observed heats for each injection of CaCl_2 at 120 s intervals after baseline correction whereas the lower panel depicts the binding enthalpies vs. Ca^{2+} /protein molar ratio.

expressed another bacterial immunoglobulin-like domain (the Big domain of Lig A in *Leptospira interrogans*)¹⁶. By ITC assay, we also confirmed the interactions between another Big domain, the Big domain of Lig A, and Ca^{2+} (Figure 4C). Altogether, two different bacterial immunoglobulin-like (Big) domains were independently proved to be able to bind Ca^{2+} , confirming that Big domain is a Ca^{2+} binding module.

As the structure of SP0498 Big domain is different from other typical Ca^{2+} -binding domain such as EF-hand, crystallin, C2 domain and so on, it may represent a novel Ca^{2+} -binding module (Figure 5).

We also tested the ability of SP0498 Big domain to bind Mg^{2+} . It showed that SP0498 Big domain didn't bind Mg^{2+} according to ITC

assay (supplementary materials Figure S1). Therefore, SP0498 Big domain might bind to Ca^{2+} specifically.

Identification of potential Ca^{2+} -binding sites in SP0498 Big domain. ^1H - ^{15}N HSQC spectra were recorded for ^{15}N -labeled SP0498 Big domains (His-tag was removed) before and after addition of increasing amounts of Ca^{2+} to identify the potential Ca^{2+} binding sites in SP0498 Big domain. The spectral changes that occurred after Ca^{2+} addition were characterized by the chemical shift variation of certain residues. When the concentration of Ca^{2+} was increased to 45 mM or higher, there was an obvious perturbation of the HSQC spectrum of SP0498 Big domain,

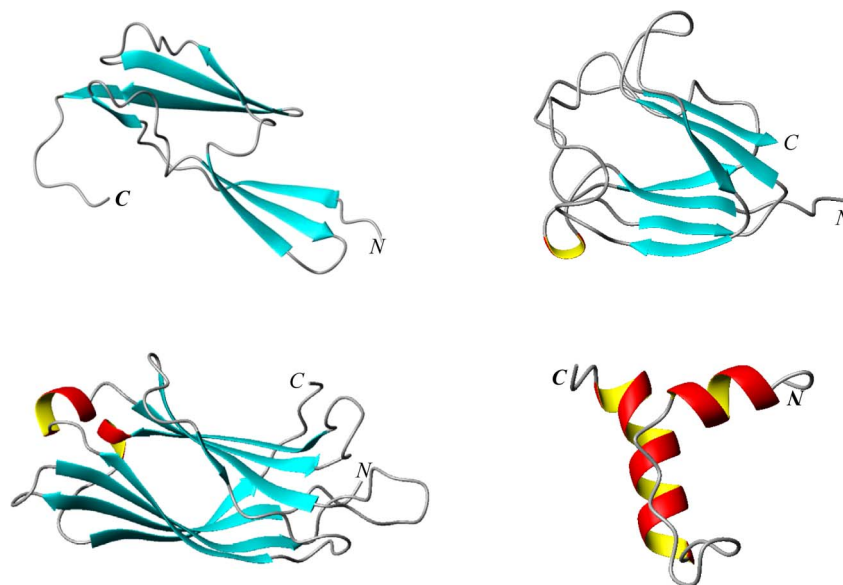


Figure 5 | Structural comparison between SP0498 Big domain and the typical Ca^{2+} -binding modules. SP0498 Big domain, upper left; crystallin domain of *Methanosarcina acetivorans* M-crystallin (PDB code 2K1W), upper right; C2 domain of rat Synaptogamin I (PDB code 1BYN), lower left; and EF-hand domain of human cardiac sodium channel NaV1.5 (PDB code 2KBI), lower right.

indicating the interactions between SP0498 Big domain and Ca^{2+} (Figure 6A). The residues with obvious chemical shift perturbation include I8, E9, E28, G29, R30, G49, I52, H68, G71, H72, and E73 (Figure 6B). Interestingly, these residues are located at the N-terminal half of the barrel-like structure and form a potential cavity which might be responsible for accommodating and binding to Ca^{2+} (Figure 6C). Furthermore, mutants corresponding to these residues were constructed to determine whether they are involved in the calcium binding. These mutants were tested by Stains-all and ITC assays to detect their Ca^{2+} -binding ability. Altogether 8 mutants were constructed and tested: T4A, I8A/E9A, S11A/Q12A, D17A, E28A/G29A/R30A, Y35A/S36A, S44A/E48A/G49A/I52A and H68A/Y69A/G71A/H72A/E73A mutants. Except for E28A/G29A/R30A mutant which was unstable, all the other 7 mutants were tested by Stains-all and ITC assays for the ability to bind calcium (supplementary materials Figure S2 A–F & Figure S3 A–F). Except for H68A/Y69A/G71A/H72A/E73A mutant, all the other tested mutants retained the Ca^{2+} -binding ability (Table 2). When H68, Y69, G71, H72, and E73 were mutated simultaneously to alanines, the mutated SP0498 Big domain lost the Ca^{2+} -binding ability completely (Figure 7A & B). We further constructed H68A/Y69A and G71A/H72A/E73A mutants separately in order to explore more details about this binding site. However, we found that both of mutants retained the Ca^{2+} binding ability with almost the same K_d value (Table 2). To exclude that the abolishment of Ca^{2+} binding ability was caused by the collapse of the global structure of H68A/Y69A/G71A/H72A/E73A mutant, we compared the CD and HSQC spectra between this mutant and the wild-type SP0498 Big domain. There was no obvious change between their CD and HSQC spectra (supplementary materials Figure S4A & B), which indicated that the mutation did not affect the global structure of SP0498 Big domain. All these confirmed that the residues (H68A/Y69A/G71A/H72A/E73A) are involved in and critical for the Ca^{2+} binding.

Discussion

Bacterial Ig-like (Big) domains belong to the immunoglobulin-like domain superfamily. They are identified in a variety of potential surface proteins in bacteria. *Escherichia coli* intimin and *Yersinia pseudotuberculosis* invasins, which also contain Big domain, were proved to play important roles in invading host cells^{17,18}.

Additionally, Big domains in Lig proteins in *Leptospira interrogans* were indicated to regulate the adhesion of pathogenic leptospires to host cells¹⁶. Thus, it is possible that SP0498 Big domain might be also involved in regulating adhesion to host cell or pathogenic invasion.

In this study, we determined the solution structure of SP0498 Big domain from *S. pneumoniae* by NMR. The solution structure of SP0498 Big domain is the first reported structure of bacterial Ig-like domain family subgroup 3 (Big 3). Instead of a typical Greek Key pattern, which is present in classic Ig-like domains, SP0498 Big domain adopts a different fold. Although it is also composed of β -strands, it displays a barrel-like conformation with eight β -strands forming three relatively separate regions. Compared with the classical Ig-like domains, SP0498 Big domain shows divergent structural characteristics, which might imply unique functions of the Big domain. Intriguingly, this domain was identified as a novel Ca^{2+} -binding module by Stains-all and ITC assays. The determined dissociation constant ($0.24 \pm 0.02 \mu\text{M}$) indicates the strong interactions between this domain and Ca^{2+} . Moreover, chemical shift perturbation and site-directed mutagenesis implied that residues H68, Y69, G71, H72, and E73 are involved in the calcium-binding process. These residues are located on a loop linking strands of βF (residues 63–69) and βG (residues 72–78) in SP0498 Big domain. This is similar to other calcium binding motifs such as EF-hands and C2 domains, in which the residues responsible for the calcium binding are also located in certain loops. Furthermore, the surface of SP0498 Big domain is dominantly composed of negative charge, which may facilitate the access of Ca^{2+} . As SP0498 is a potential surface protein of *S. pneumoniae*, the identification of the interactions between SP0498 Big domain and Ca^{2+} implies a potential role of the Big domain and calcium in the regulation of *S. pneumoniae* surface proteins' interactions with host cells.

Ca^{2+} is critical for a variety of biological processes in organisms from human to bacteria. In bacteria, Ca^{2+} is involved in many cellular processes including cell cycle, pathogenesis and chemotaxis^{19–21}. Additionally, the activities or stability of some enzymes, many of which are extracellular, are dependent on Ca^{2+} ^{22,23}. The recognition of Ca^{2+} requires Ca^{2+} -binding modules which are present in a number of proteins. The well-known EF-hand is among the earliest identified Ca^{2+} -binding modules. This domain is composed of 2 helices spaced by loops. Other modules like the crystallin domain and the C2

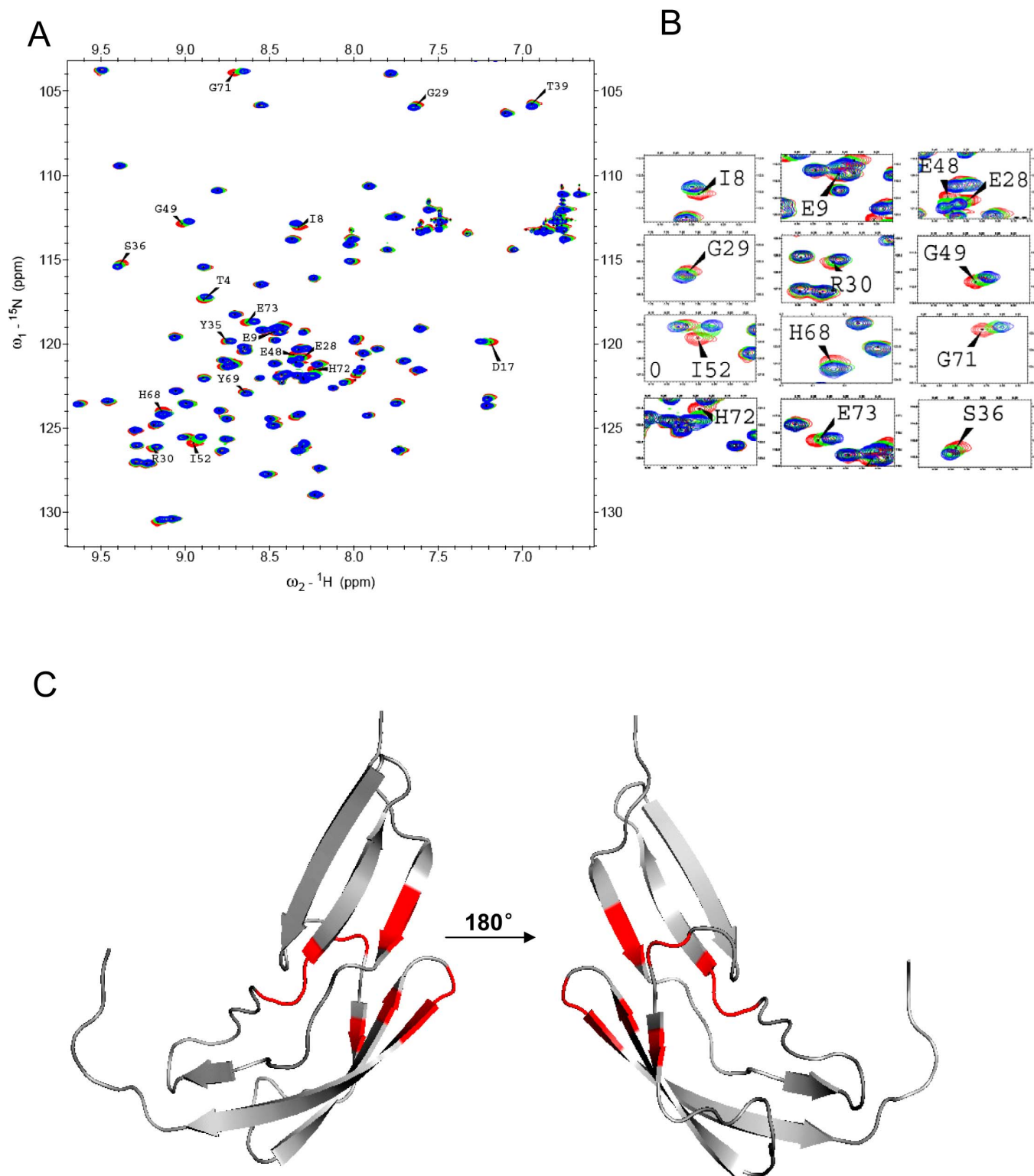


Figure 6 | Potential residues involved in the interactions between SP0498 Big domain and Ca^{2+} . (A) The ^1H - ^{15}N HSQC spectra of SP0498 Big domain titrated with different concentrations of Ca^{2+} . (B) The representative residues of SP0498 Big domain with chemical shift significantly changed. Four colours represent different concentrations of Ca^{2+} : red, 0 mM; green, 15 mM; and blue, 45 mM. (C) The ribbon structure of SP0498 Big domain produced by PyMOL, showing the residues (red color) interacting with Ca^{2+} .

domain are both composed of β -strands^{13,14}. The former adopts a Greek key-like fold while the latter displays a wedge-like shape. The structural comparison between SP0498 Big domain and these typical Ca^{2+} -binding modules shows that the structure of the Big domain is different from that of any of these well-known Ca^{2+} -binding

domains, indicating it may represent a novel Ca^{2+} -binding module subfamily.

In summary, we solved the solution structure of SP0498 Big domain and revealed the Big domain as a novel Ca^{2+} -binding module. As some surface proteins are involved in mediating the pathogens'


Table 2 | Binding constants of SP0498 Big domains and Ca²⁺ determined by ITC

Mutated Site	Dissociation constants K _d (μM)
Wild type	0.24 ± 0.02
T4A	0.32 ± 0.02
I8A/E9A	0.24 ± 0.00
S11A/Q12A	0.26 ± 0.02
D17A	0.19 ± 0.02
E28A/G29A/R30A	non-available ^a
Y35A/S36A	0.23 ± 0.01
S44A/E48A/G49A/I52A	0.31 ± 0.04
H68A/Y69A/G71A/H72A/E73A	no binding was tested
H68A/Y69A	0.19 ± 0.01
G71A/H72A/E73A	0.18 ± 0.01

^aThis mutation causes the protein unstable.

interaction with hosts more effectively in the presence of Ca²⁺^{13,24,25}, our identification of the interactions between Big domain and Ca²⁺ suggests a potential role of this domain in the host cell adhesion and invasion of the pathogens.

Methods

Cloning, expression and protein purification. The fragment of the gene of the Big domain (residues 1146–1228) of SP0498 from *Streptococcus pneumoniae* was amplified by PCR and cloned into pET-22 b (+) (Novagen). Recombinant protein with a C-terminal 6-histidine-tag was expressed in *Escherichia coli* strain BL21 (DE3). Cells expressing the recombinant protein were grown in Luria-Bertani (LB) medium containing 100 μg/mL of ampicillin at 37° and induced with 1.0 mM isopropyl-β-D-thiogalacto-pyranoside (IPTG) for 4 hours. The cells were harvested and sonicated in buffer containing 20 mM Tris and 500 mM NaCl at pH 7.8. The cell lysate was then centrifuged to remove precipitate and purified by affinity chromatography using Ni²⁺-NTA resin. The eluted protein was further purified by size exclusion chromatography using a Superdex 75 column on an AKTA purification system. ¹⁵N, ¹³C-labeled protein was purified in the same way except for that LB medium was replaced by M9 medium containing 0.5 g/L 99% ¹⁵N-labeled ammonium chloride and 2.5 g/L ¹³C-labeled glucose as the sole nitrogen and carbon source, respectively. The final NMR samples contained 0.7 mM SP0498 Big domain, 25 mM phosphate (pH 6.5), 100 mM sodium chloride and 2 mM EDTA in either 90% H₂O/10% D₂O or 100% D₂O.

For the SP0498 Big domain with His-tag cleaved, an additional fragment, which can be recognized and cleaved by TEV protease, was cloned between the SP0498 big domain fragment and the C-terminal His-tag. The target protein was then expressed and purified as described above. Then the protein was incubated with TEV protease in 20 mM Tris with 150 mM NaCl and 2 mM β-mercaptoethanol at pH 7.7 and 4°C overnight. Target protein with His-tag cleaved was separated from the cleaved His-tag and the TEV protease (containing His-tag) by affinity chromatography using Ni²⁺-NTA resin.

NMR experiments and structure calculation. All the NMR experiments were carried out at 298 K on a Bruker DMX500 spectrometer using a cryoprobe. ¹H-¹⁵N HSQC, HNCACB, CACB(CO)NH, HNHA, HCC(CO)NH, 3D ¹⁵N-edited NOESY, and ¹³C-edited NOESY spectra were recorded. The acquired NMR data were processed with the soft wares of NMRPipe and NMRDraw and analyzed with Sparky^{326,27}. Distance restraints were derived from ¹⁵N-edited NOESY and ¹³C-edited NOESY spectra and dihedral angle restraints were obtained by TALOS program for structural calculation²⁸. Structures were calculated using CYANA 3.0²⁹. The finally calculated 20 structures with lowest energy were analyzed with MOLMOL³⁰. The Ramachandran plot was analyzed with PROCHECK online (http://nihserver.mbi.ucla.edu/SAVES_3/).

Circular dichroism spectroscopy (CD). CD experiments for Stains-all assay were performed on a Jasco-810 spectrophotometer over the wavelength range of 550–700 nm at room temperature. SP0498 Big domain was purified as described above and incubated with 2 mM EDTA immediately for 1 hour immediately after the purification. Then EDTA was removed by dialysis against 20 mM HEPES buffer (containing 100 mM NaCl, pH 7.0). 0.01 mM SP0498 Big domain was mixed with 0.1 mM Stains-all dye solution which was prepared in 2 mM MOPS buffer, pH 7.2 and contained 30% ethylene glycol. The mixture was incubated in dark room for 10 min before measurement. Measurements were taken in a 1 mm path-length quartz cuvette at a rate of 50 nm/min and a data pitch of 1 nm. CD spectrum of the buffer was measured as control and subtracted. Three successive scans were recorded and averaged. CD spectra for monitoring the secondary structure of the proteins were measured over the wavelength range of 210–250 nm at room temperature.

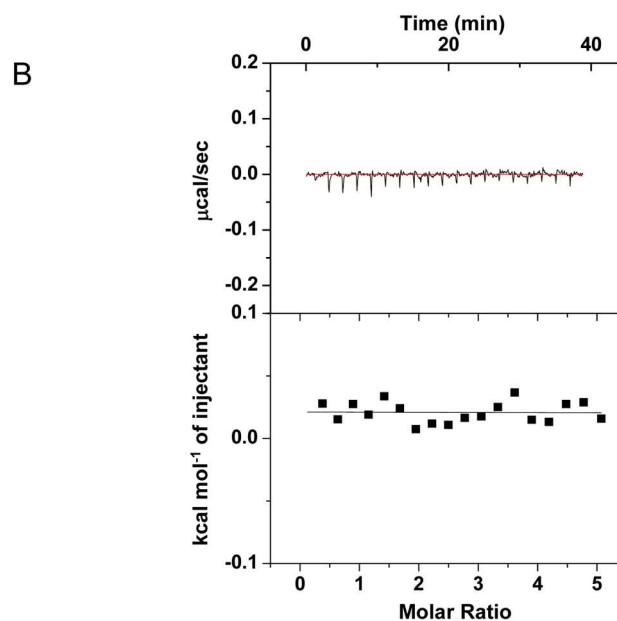
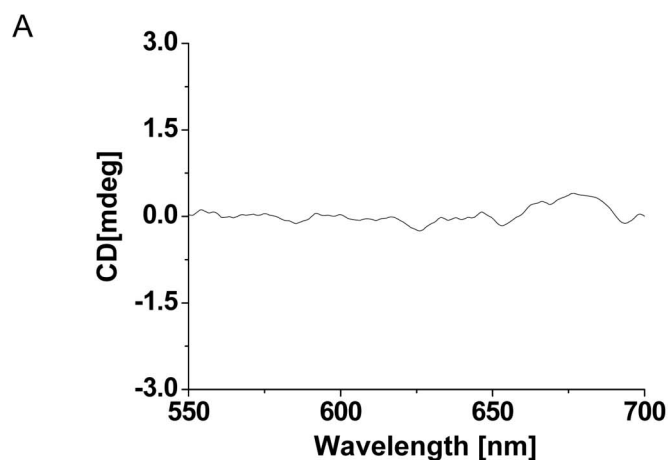


Figure 7 | Interactions between H68A/Y69A/G71A/H72A/E73A mutant of SP0498 Big domain and Ca²⁺ measured by Stains-all and ITC assays. (A) CD spectrum of this mutated SP0498 Big domain mixed with Stains-all dye at molar ratios of 10 : 1. (B) Titration of this mutated SP0498 Big domain with Ca²⁺ was performed in 20 mM HEPES buffer pH 6.5 (containing 100 mM NaCl) measured by ITC at 20°C. The upper thermogram panel shows the observed heats for each injection of CaCl₂ at 120 s intervals after baseline correction whereas the lower panel depicts the binding enthalpies vs. Ca²⁺/protein molar ratio.

Isothermal titration calorimetry (ITC). The interactions between the SP0498 Big domain and Ca²⁺ were tested by ITC (ITC₂₀₀, GE Company) and the data were analyzed by MicroCal LLC ITC software (MicroCal). The SP0498 Big domain was purified and decalcified as described above. 0.20 mM SP0498 Big domain was loaded in the cell and 2.75 mM CaCl₂ was loaded in syringe. 20 injections of 2 μL of CaCl₂ (2.75 mM) were added into the calorimeter cell containing 200 μL of 200 μM SP0498 Big domain at 120 seconds intervals. The interactions between the mutated SP0498 Big domains and Ca²⁺ were tested by the same procedure.

Chemical shift perturbation. Chemical shift perturbation was carried out to investigate the interactions between SP0498 Big domain and Ca²⁺. 0.15 mM ¹⁵N-labeled SP0498 Big domain (decalcified as described above) was titrated with increasing amount of Ca²⁺. ¹⁵H-¹⁵N HSQC spectra were recorded at each point for analysis.

Site-directed mutagenesis. A series of point mutations were introduced into the recombinant pET22b(+)-SP0498 Big domain vector. The plasmid pET22b(+)-



SP0498 Big domain was used as a template, amplified by PCR using PrimeSTARTTM HS DNA polymerase (TaKaRa, Dalian, China) and two complementary (partially overlapping) primers containing the desired mutation. The 50 μ L PCR reaction was carried out with 50–100 ng templates, 10 μ M primer pair, 200 μ M dNTPs and 2 U of DNA polymerase, and started at 95°C for 3 min, followed by 20 cycles of 95°C for 45 s, 55°C for 45 s, and a final extension at 72°C for 8 min. After PCR reaction, the PCR product was digested with DpnI (TaKaRa, Dalian, China) overnight to remove methylated parental non-mutated plasmid, and then transformed into *Escherichia coli* BL21 (DE3). The mutated proteins were overexpressed and purified as described above.

- Greenwood, B. The epidemiology of pneumococcal infection in children in the developing world. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **354**, 777–785 (1999).
- Tettelin, H. *et al.* Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science* **293**, 498–506 (2001).
- Bergmann, S. & Hammerschmidt, S. Versatility of pneumococcal surface proteins. *Microbiology* **152**, 295–303 (2006).
- Jedrzejas, M. Unveiling molecular mechanisms of bacterial surface proteins: *Streptococcus pneumoniae* as a model organism for structural studies. *Cellular and molecular life sciences* **64**, 2799–2822 (2007).
- Tuomanen, E. Molecular and cellular biology of pneumococcal infection. *Current opinion in microbiology* **2**, 35–39 (1999).
- Bierne, H. & Cossart, P. *Listeria monocytogenes* surface proteins: from genome predictions to function. *Microbiology and molecular biology reviews* **71**, 377–397 (2007).
- Paterson, G. K. & Mitchell, T. J. The biology of Gram-positive sortase enzymes. *Trends in microbiology* **12**, 89–95 (2004).
- Williams, A. The immunoglobulin superfamily in cell surface recognition. *Comptes rendus de l'Académie des sciences. Série 3, Sciences de la vie* **314**, 27–29 (1992).
- Barclay, A. N. Membrane proteins with immunoglobulin-like domains—a master superfamily of interaction molecule. *Semin Immunol* **15** (4), 215–223 (2003).
- Bateman, A., Chothia, C. & Eddy, S. R. Members of the immunoglobulin superfamily in bacteria. *Protein science* **5**, 1939–1941 (1996).
- Bork, P., Holm, L. & Sander, C. The immunoglobulin fold. *J. Mol. Biol.* **242**, 309–320 (1994).
- Halaby, D., Poupon, A. & Mornon, J. P. The immunoglobulin fold family: sequence analysis and 3D structure comparisons. *Protein engineering* **12**, 563–571 (1999).
- Aravind, P. *et al.* The $\beta\gamma$ -Crystallin Superfamily Contains a Universal Motif for Binding Calcium. *Biochemistry-U S A* **48**, 12180–12190 (2009).
- Rizo, J. & Südhof, T. C. C2-domains, structure and function of a universal Ca^{2+} -binding domain. *J Biol Chem* **273**, 15879–15882 (1998).
- Caday, C. & Steiner, R. The interaction of calmodulin with the carbocyanine dye (Stains-all). *J Biol Chem* **260**, 5985–5990 (1985).
- Raman, R. *et al.* Big Domains Are Novel Ca^{2+} -Binding Modules: Evidences from Big Domains of *Leptospira* Immunoglobulin-Like (Lig) Proteins. *PLoS one* **5**, e14377 (2010).
- Luo, Y. *et al.* Crystal structure of enteropathogenic *Escherichia coli* intimin–receptor complex. *Nature* **405**, 1073–1077 (2000).
- Hamburger, Z. A., Brown, M. S., Isberg, R. R. & Bjorkman, P. J. Crystal structure of invasins: a bacterial integrin-binding protein. *Science* **286**, 291–295 (1999).
- Yu, X. C. & Margolin, W. Ca^{2+} -mediated GTP-dependent dynamic assembly of bacterial cell division protein FtsZ into asters and polymer networks in vitro. *The EMBO journal* **16**, 5455–5463 (1997).
- Straley, S. C., Plano, G. V., Skrzypek, E., Haddix, P. L. & Fields, K. A. Regulation by Ca^{2+} in the *Yersinia* low- Ca^{2+} response. *Molecular microbiology* **8**, 1005–1010 (1993).
- Tisa, L. S. & Adler, J. Calcium ions are involved in *Escherichia coli* chemotaxis. *Proceedings of the National Academy of Sciences* **89**, 11804 (1992).
- Zakharov, S. D., Li, X., Red'ko, T. P. & Dilley, R. A. Calcium binding to the subunit c of *E. coli* ATP-synthase and possible functional implications in energy coupling. *Journal of bioenergetics and biomembranes* **28**, 483–494 (1996).
- Smith, C. A., Toogood, H. S., Baker, H. M., Daniel, R. M. & Baker, E. N. Calcium-mediated thermostability in the subtilisin superfamily: the crystal structure of *Bacillus* Ak. 1 protease at 1.8 Å resolution. *Journal of molecular biology* **294**, 1027–1040 (1999).
- Dardanelli, M., Angelini, J. & Fabra, A. A calcium-dependent bacterial surface protein is involved in the attachment of rhizobia to peanut roots. *Canadian journal of microbiology* **49**, 399–405 (2003).
- Josefsen, E., O'Connell, D., Foster, T. J., Durussel, I. & Cox, J. A. The binding of calcium to the B-repeat segment of SdrD, a cell surface protein of *Staphylococcus aureus*. *J Biol Chem* **273**, 31145–31152 (1998).
- Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of biomolecular NMR* **6**, 277–293 (1995).
- Goddard, T. D. & K. D. G. SPARKY 3. *University of California, San Francisco*.
- Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of biomolecular NMR* **44**, 213–223 (2009).
- Herrmann, T., Güntert, P. & Wüthrich, K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *Journal of molecular biology* **319**, 209–227 (2002).
- Koradi, R., Billeter, M. & Wüthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *Journal of molecular graphics* **14**, 51 (1996).
- Larkin, M. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
- Gouet, P., Courcelle, E. & Stuart, D. I. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305–308 (1999).

Acknowledgments

We thank F. Delaglio and A. Bax for providing NMRPipe and NMRDraw, T. D. Goddard and D. Kneller for Sparky, T. Herrmann and K. Wüthrich for CYANA, R. Koradi and K. Wüthrich for MOLMOL. This work was supported by the National Basic Research Program of China (973 Program), grant number 2009CB918804, the Knowledge Innovation Program of the Chinese Academy of Science, grant number KSCX2-EW-Q-4, and the Chinese National Natural Science Foundation, grant number 31270780.

Author contributions

Tao Wang and Xiaoming Tu designed the research. Tao Wang and Jiahai Zhang performed the experiments and data analysis, and Tao Wang, Xiaoming Tu, Xuecheng Zhang and Chao Xu wrote the paper.

Additional information

Supplementary information accompanies this paper at <http://www.nature.com/scientificreports>

Competing financial interests: The authors declare no competing financial interests.

License: This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>

How to cite this article: Wang, T., Zhang, J., Zhang, X., Xu, C. & Tu, X. Solution structure of the Big domain from *Streptococcus pneumoniae* reveals a novel Ca^{2+} -binding module. *Sci. Rep.* **3**, 1079; DOI:10.1038/srep01079 (2013).