Structural Insights into SraP-Mediated Staphylococcus aureus Adhesion to Host Cells

Yi-Hu Yang1*, Yong-Liang Jiang1*, Juan Zhang1, Lei Wang1, Xiao-Hui Bai1, Shi-Jie Zhang1, Yan-Min Ren1, Na Li2, Yong-Hui Zhang1, Zhiyong Zhang1, Qingguo Gong1, Yide Mei1, Ting Xue1, Jing-Ren Zhang3, Yuxing Chen1*, Cong-Zhao Zhou1*

1Hefei National Laboratory for Physical Sciences at the Microscale and School of Life Sciences, University of Science and Technology of China, Hefei Anhui, People’s Republic of China, 2Shanghai Institute of Biochemistry and Cell Biology, Shanghai, China, 3Center for Infectious Disease Research, School of Medicine, Tsinghua University, Beijing, People’s Republic of China

Abstract

Staphylococcus aureus, a Gram-positive bacterium causes a number of devastating human diseases, such as infective endocarditis, osteomyelitis, septic arthritis and sepsis. S. aureus SraP, a surface-exposed serine-rich repeat glycoprotein (SRRP), is required for the pathogenesis of human infective endocarditis via its ligand-binding region (BR) adhering to human platelets. It remains unclear how SraP interacts with human host. Here we report the 2.05 Å crystal structure of the BR of SraP, revealing an extended rod-like architecture of four discrete modules. The N-terminal legume lectin-like module specifically binds to N-acetylmuramuramic acid. The second module adopts a β-grasp fold similar to Ig-binding proteins, whereas the last two tandem repetitive modules resemble eukaryotic cadherins but differ in calcium coordination pattern. Under the conditions tested, small-angle X-ray scattering and molecular dynamic simulation indicated that the three C-terminal modules function as a relatively rigid stem to extend the N-terminal lectin module outwards. Structure-guided mutagenesis analyses, in addition to a recently identified trisaccharide ligand of SraP, enabled us to elucidate that SraP binding to sialylated receptors promotes S. aureus adhesion to and invasion into host epithelial cells. Our findings have thus provided novel structural and functional insights into the SraP-mediated host-pathogen interaction of S. aureus.

Introduction

The serine-rich repeat glycoproteins (SRRPs) are a family of adhesins encoded by Gram-positive bacteria that mediate attachment to a variety of host cells or bacteria themselves [1]. SRRPs typically consist of a signal peptide at the N-terminus, a short SRR (SRR1, ~50–170 residues), a ligand-binding region (BR, ~250–500 residues) followed by a much longer SRR (SRR2, ~400–4000 residues), and a C-terminal LPXTG motif anchoring to the cell wall [1]. The BRs of SRRPs from different pathogenic bacteria have varying primary sequences and bind to diverse targets from carbohydrates to proteins [1].

In addition to having highly variable sequences, the BRs from different bacteria are composed of distinct modules. The diversity of BR modules and combinations contributes to the multiple functions of SRRPs. The only four known BR structures to date have identified five distinct modules [2–4]. The BR of Streptococcus pneumoniae Fap1 contains two modules: an N-terminal helical module and a C-terminal CnaA module [2], whereas Streptococcus gordoni GspB has a BR of three modules: CnaA, Siglec and a unique module of unknown function [3]. In addition, the recently reported BR structures of the two SRRP paralogs (Srr1 and Srr2) from Streptococcus agalactiae defined two immunoglobulin-fold modules, which specifically bind to the host fibrinogen [4]. However, the module composition and corresponding molecular functions of most BRs remain unknown, which largely impedes the understanding of the pathogenesis mechanism of SRRPs.

S. aureus is a human pathogen that causes a wide range of debilitating and life-threatening infections [5]. S. aureus encodes a 2,271-residue SRRP termed serine-rich adhesin for binding to platelets (SraP), that is involved in the pathogenesis of infective endocarditis [6]. Moreover, the BR (residues Phe245–Asn751) of SraP, termed SraPBR, mediates intraspecies interaction and promotes bacterial aggregation [7]. We determine the 2.05 Å crystal structure of SraPBR, revealing a rod-like tandem organization of four discrete modules: a legume lectin-like module, a module with a β-grasp fold, and two tandem cadherin-like modules that create the rigid stem of SraPBR. Further structural and biochemical analyses reveal that the legume lectin-like module specifically binds to N-acetylmuramuramic acid (Neu5Ac), which may mediate adhesion to host sialylated receptors. These findings increase our knowledge of the diverse BR modules of SRRPs, and provide structural insights into a novel surface protein that mediates interaction of S. aureus with host epithelial cells.

Citation: Yang Y-H, Jiang Y-L, Zhang J, Wang L, Bai X-H, et al. (2014) Structural Insights into SraP-Mediated Staphylococcus aureus Adhesion to Host Cells. PLoS Pathog 10(6): e1004169. doi:10.1371/journal.ppat.1004169

Editor: Paul M. Sullam, University of California, San Francisco, United States of America

Received December 4, 2013; Accepted April 22, 2014; Published June 5, 2014

Copyright: © 2014 Yang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the Ministry of Science and Technology of China (http://www.most.gov.cn, Grant No. 2014CB910100 and 2013CB835300), the National Natural Science Foundation of China (http://www.nsfc.gov.cn/Portal0/default152.htm, Grant No. 31270781), and the Fundamental Research Funds for the Central Universities. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: cxying@ustc.edu.cn (YC); zcz@ustc.edu.cn (CZZ)

These authors contributed equally to this work.
Author Summary

Staphylococcus aureus is an important pathogen that causes a range of human diseases, such as infective endocarditis, osteomyelitis, septic arthritis and sepsis. The increasing resistance of S. aureus to most of the current antibiotics emphasizes the need to develop new approaches to control staphylococcal infections. As a surface-exposed serine-rich repeat glycoprotein (SRRP), S. aureus SraP is involved in the pathogenesis of infective endocarditis via its ligand-binding region (BR) adhering to human platelets. However, little is known about how SraP interacts with its host receptor(s). Through structural and functional analyses of the BR domain, we have discovered a specific binding of SraP to N-acetylneuraminic acid (Neu5Ac), in agreement with a recent report of the trisaccharide ligand of SraP. Further mutagenesis analysis showed that SraP binding to Neu5Ac and the trisaccharide promotes S. aureus adhesion to and invasion into host epithelial cells. These findings increase our knowledge of surface protein mediated interaction of S. aureus with host epithelial cells.

Results

Overall structure of SraP<sub>BR</sub>

Each asymmetric unit of the final model at 2.05 Å resolution contains a single SraP<sub>BR</sub> molecule of residues Thr251–Asn751. The N-terminal residues Phe245–Thr250 are not visible due to their poor electron density. SraP<sub>BR</sub> folds into a slightly bent, rod-like structure of 160 Å in length that has four discrete modules: a head-like N-terminal module followed a stem of three all-β modules (Fig. 1A). All modules have a dominant β-strand secondary structure. During the model building and refinement process, three peaks of electron density at the 24 σ level were observed in the |Fo|−|Fc| Fourier difference map, indicating the presence of three metal ions. Atomic absorption spectroscopy assigned these metal ions to Ca<sup>2+</sup>; we thus termed them Ca-1, Ca-2 and Ca-3 accordingly. The structure also contains a sucrose and a 2-(N-morpholino)-ethanesulfonic acid (MES) molecule, which were introduced from the cryoprotectant and crystallization buffer, respectively.

The N-terminal module adopts a jelly-roll fold with a β-sandwich (β1–β17) core architecture of two antiparallel β-sheets packed against each other (Fig. 1B). Beyond the core structure, two α-helices (α1 and α2) pack on either side of the lateral of the β-sandwich and partially seal the hydrophobic lateral openings between the two β-sheets. A molecule of sucrose binds to the protruding loops at the distal end of the N-terminal module (Fig. 1B). A structural similarity search using the DALI server revealed that this module is most similar to legume lectins, despite sharing a sequence identity of ≤20%. The top hits include lectins from legume plants such as Pisum sativum (PDB 2BQP) [9] and Robinia pseudoacacia (PDB: 1FNY) [10], with a Z-score of 22–23 and root mean square deviation (RMSD) of 2.3–2.5 Å over ~200 Ca atoms. Thus, we termed the N-terminal module L-lectin.

The second module possesses a ubiquitin-like β-grasp fold (β-GF) in the immunoglobulin (Ig)-binding protein superfamily [11]. The module contains a twisted four-stranded mixed β-sheet (β19–β18–B23–B21) packed against a two-stranded antiparallel β-sheet (β20–β22) and an α-helix α3 (Fig. 1B).

The C-terminal tandem repeat of two modules share a sequence identity of 56% and a similar overall structure with an RMSD of 0.81 Å over 81 Ca atoms. The two modules are linked in a head-to-tail fashion, indicating duplication of the coding region during evolution. Each module consists of a β-sandwich of three β-sheets (Fig. 1B & 1C). Structural analysis revealed that the two modules resemble eukaryotic cadherins of known-structure (PDB 3UBF, Z-score 8.6, RMSD 2.3 Å over 85 Ca atoms and PDB 4APX, Z-score 9.5, RMSD 2.3 Å over 87 Ca atoms) [12,13]. Thus the tandem cadherin-like (CDHL) modules were termed CDHL-1 and CDHL-2, respectively.

Specific binding of the L-lectin module to N-acetylneuraminic acid

The L-lectin module is structurally similar to legume lectins (Fig. 2A), a large family of carbohydrate-binding proteins with diverse activities [14]. Legume lectins commonly coordinate a Ca<sup>2+</sup> in addition to a transition metal ion, usually Mn<sup>2+</sup> [15]. In contrast, the L-lectin module has only a Ca<sup>2+</sup>-binding site. At the apex of the L-lectin module, the Ca<sup>2+</sup> (named Ca-l) was embedded in a 7-coordinate geometry (Fig. 2B). The seven coordinates are from the side-chain oxygen atoms of Asp365 (bidentate coordination from Oδ1 and Oδ2), Asp382 (Oδ2), Asn369 (Oδ1), the main-chain oxygen atom of Tyr367 and two water molecules (Wat1 and Wat2). The two water molecules were further stabilized by the main-chain oxygen atoms of Asp330, Asp331, Asn353, and Asp354.
Ala349, Oδ1 of Asp382 and another water molecule (Wat3) (Fig. 2B).

A molecule of sucrose is fixed by a cluster of loops protruding from the apex of the L-lectin module. The sucrose molecule has a "bent-back" conformation with the glucose and fructose moieties perpendicular to each other (Fig. 2B). The glucose moiety is inserted in the pocket, and adopts a conformation nearly parallel to the two β-sheet layers. The sugar ring makes hydrophobic interactions with Tyr367 and the morpholine ring of MES, whereas the hydroxyl groups are stabilized by hydrogen bonds with Ala478-Nα, Asp330-Oδ1, and three water molecules (Wat3–5), which are further fixed by residues Asp330, Asn347 and Ala349 (Fig. 2B). In contrast, the solvent-exposed fructose moiety is bent through interactions with MES and two water molecules (Wat6 and Wat7) (Fig. 2B). The sucrose binding residues, especially the stacking residue Tyr367, are structurally conserved in the legume lectins (Fig. 2C).

To identify the favored saccharide of SraP, we first detected the binding affinity of the L-lectin module towards eight common monosaccharides using the surface plasmon resonance assays. Among these monosaccharides, only Neu5Ac bound to the L-lectin module (Fig. 2D). In consequence, we determined that the L-lectin module has an equilibrium dissociation constant (Kd) of 0.54 mM towards Neu5Ac (Fig. 2E), comparable to previously reported values of legume lectins [16].

Afterwards, we attempted to obtain the Neu5Ac-complexed structure without success. Therefore, we docked Neu5Ac to the structure of the L-lectin module with the sucrose-binding pocket as the search grid. Neu5Ac was docked at a position overlapping the glucose moiety of sucrose, with a shift of ~1.5 Å towards Ca-1. In the model, Neu5Ac is stacked against Tyr367 via hydrophobic interactions, and makes direct polar interactions with the side chains of Ser293, Asn347, Tyr367 and Asn369 and the Nα atoms of Gly477 and Ala478 (Fig. 2F).

Structural similarity of the β-GF module to Ig-binding proteins

The β-GF module adopts a ubiquitin-like β-grasp fold in the Ig-binding superfamily [11]. DALI search [8] suggested the module resembles the B1 domain of mucus-binding protein type 2 repeat Mub-R5 from Lactobacillus reuteri [17] (PDB code 3I57, Z-score 9.1, RMSD 1.6 Å, over 68 Cα atoms), and protein L (PpL) from Peptostreptococcus magnus [18] (PDB code 1HEZ, Z-score 4.5, RMSD 1.6 Å, over 68 Cα atoms).
2.5, over 55 Cα atoms). The B1 domain belongs to a family of Ig-binding proteins [19] that have a core structure of an α-helix packed against a four-stranded β-sheet [17,20]. The major differences are from the helix α3 and the two lateral β-strands (β19 and β21). In addition, the β-GF module of SraP BR contains two extra β-strands, β20 and β22, which are substituted by loops or α-helix extensions in the B1 domain (Fig. 3). Mub-R5 interacts in vitro with a large repertoire of mammalian Ig proteins including secretory IgA, whereas PpL binds to the V_L domain of Ig κ chain [19,21]. Complex structures indicated that formation of a β-zipper is necessary for the binding of PpL to the V_L domain of Ig κ chain [18] and IgG or IgM [22]. However, the corresponding β-strands β19 and β21 are much shorter in the β-GF module, which might not be capable of forming a β-zipper.

The tandem cadherin-like modules create a relatively rigid SraPBR stem

The tandem CDHL modules resemble eukaryotic cadherins in a superfamily of calcium-dependent adhesions (Fig. 4A & 4B). Unlike the eukaryotic cadherins, each of which coordinates three Ca^{2+} ions (PDB: 1L3W) [23], CDHL-1 and CDHL-2 binds to Ca-2 and Ca-3 with a 7-coordinate geometry, respectively. Ca-2 is fixed by Asp573 and Asp601 (bidentate coordination from Oδ1 and Oδ2), Asn602 (Oδ1), Asp645 (Oδ2), and the main-chain oxygen atom of Lys575 (Fig. 4C), whereas Ca-3 is coordinated to Asp661 and Asp690 (bidentate coordination from Oδ1 and Oδ2), Asn691 (Oδ1), Asp734 (Oδ2), and the main-chain oxygen atom of Thr663 (Fig. 4D). The coordinate bonds between Ca^{2+} and corresponding residues have a length from 2.2 to 2.6 Å.

The small-angle scattering of X-rays (SAXS) is usually applied to address the flexibility and conformational states of biological macromolecules in solution [24]. To explore the role of Ca^{2+}, we used SAXS to compare the overall structure of SraPBR in the presence or absence of Ca^{2+}. A difference in wide-angle scattering curves of SAXS indicated that SraPBR adopts different conformations with or without Ca^{2+}. Compared to the Ca^{2+}-free form, the envelope of Ca^{2+}-bound SraPBR correlates much better with the SraPBR crystal structure (Fig. 5A). The larger discrepancies of the Ca^{2+}-free SraPBR are mainly resulted from the tandem CDHL modules, indicating that binding of Ca^{2+} makes the tandem CDHL modules more rigid, and thereby facilitates the extended conformation of SraPBR in solution.

We also performed molecular dynamics simulations of SraPBR in the presence or absence of Ca^{2+}. Ca^{2+}-bound SraPBR remains extended and shows slight conformational changes over the simulation time, whereas removal of Ca^{2+} caused the curling of SraPBR (Fig. 5B). Geometric analysis of the Ca^{2+}-coordinating residues revealed larger fluctuations at the two Ca^{2+}-binding junctions of Ca^{2+}-free SraPBR, suggesting the importance of Ca^{2+} for the structural integrity of SraPBR. Moreover, the increased RMSD values indicated the two junctions in Ca^{2+}-free SraPBR undergo dramatic conformational changes. Given the CDHL-2 modules superimposed, the other three modules adopt a more curved conformation and are projected to the opposite side upon the loss of Ca^{2+} (Fig. 5B).

Figure 3. Superpositions of the β-GF module against A) the Ig-binding proteins of B1 domain of mucus-binding protein type 2 repeat Mub-R5 from L. reuteri (PDB 3I57), and B) that of Protein L (PpL) from P. magnus (PDB 1HEZ). The β-GF module, the B1 domain and PpL are colored in red, grey and light pink, respectively.

doi:10.1371/journal.ppat.1004169.g003

Figure 4. Structural comparison and the calcium coordination patterns of the CDHL modules. Structural comparison of A) CDHL-1 and B) CDHL-2 to the classic cadherin (PDB 3MVS). The CDHL-1 module, CDHL-2 module and the classic cadherin and corresponding binding calcium ions are shown in yellow, green and red, respectively. The coordination patterns of C) Ca-2 and D) Ca-3. The coordinate bonds are presented in green and yellow sticks, respectively. The coordinate bonds are indicated as blue dotted lines.

doi:10.1371/journal.ppat.1004169.g004

PLOS Pathogens | www.plospathogens.org
The rod-like, four-module structure of SraP BR has three junctions with a buried interface of 700, 400 and 220 Å², respectively (Fig. 1B). The interface between L-lectin and β-GF is maintained by several hydrogen bonds including Asn257-Glu493, Arg303-Asn519, Arg303-Ser494, Leu341-Asn519, and Asn466-Asn562 (Fig. 5C). The second interface is formed by residues from β-GF and CDHL-1 via hydrogen bonds of Lys536-Asn602, Gly537-Asn602 and Asp503-Thr604 (Fig. 5D). At the third junction, Glu587 and Val588 of CDHL-1 form two hydrogen bonds with Asn691 of CDHL-2 (Fig. 5E). The second
and third interfaces are relatively small; however, they are more stable due to the contribution from coordinate bonds of Ca-2 to Asn602 and Ca-3 to Asn691.

To further investigate the plasticity of these junctions, we determined three crystal structures for each of two consecutive modules: L-lectin&β-GF (Phe245-Lys575), β-GF&CDHL-1 (Ser579-Thr663), and CDHL-1&2 (Ala576-Asn751). We superimposed each of these three structures over the corresponding two modules of the full-length SraPR structure, always with the C-terminal modules aligned. The results revealed slight twisting and/or translation with intermodule angle changes of 9.4°, 11.7° and 14.6° for the three junctions, respectively (Fig. 5C–E). In detail, residues Val496-Glu498 in L-lectin&β-GF, Phe571&Thr572 in β-GF&CDHL-1, and Asp661-Thr663 in CDHL-1&2 undergo slight conformational changes. Except for a short disordered segment at the N-terminus of CDHL-1&2 (Fig. 5E) due to the deletion of two Ca-2 coordinate residues Asp573 and Lys575, we did not find secondary-structure change in any module pairs. Moreover, merging these three structures via sequential alignment of the same module resulted in a total twist of 5.4° along the axis of SraPR structure (Fig. S1), suggesting that the slight interdomain twists of the three junctions are randomly occurred and could be canceled out. Together, the results indicated that in the presence of calcium, SraPR adopts a relatively rigid rod-like structure under all conditions tested.

SraP promotes S. aureus adhesion and invasion to A549 cells through sialylated receptors

Bacterial attachment and colonization at the surface of host cells have been thought to be mediated by specific binding of BRs to glycoconjugates [1]. It has recently been reported that SraP binds to the salivary agglutinin gp340 via the Neu5Ac moiety of the trisaccharide Neu5Ac(2–3)Galβ1–4GlcNAc [25]. In fact, gp340 and homologs are also expressed in lung epithelial cells [26]. To determine if the L-lectin module mediates SraPR adhesion, we incubated a monolayer of human lung epithelial A549 cells with green fluorescent protein (GFP)-fused SraPR and individual modules. The results indicated that full-length SraPRR and the L-lectin module, but not the CDHL-1&2 modules or the β-GF module, specifically adhered to the A549 cells (Fig. 6). Try367 is a key residue to make hydrophobic interactions with sucrose and the docked Neu5Ac, we thus constructed the Y367G mutant proteins docking the Neu5Ac, we thus constructed the Y367G mutant proteins (Fig. 5E) due to the deletion of two Ca-2 coordinate residues Asp573 and Lys575, we did not find secondary-structure change in any module pairs. Moreover, merging these three structures via sequential alignment of the same module resulted in a total twist of 5.4° along the axis of SraPR structure (Fig. S1), suggesting that the slight interdomain twists of the three junctions are randomly occurred and could be canceled out. Together, the results indicated that in the presence of calcium, SraPR adopts a relatively rigid rod-like structure under all conditions tested.

SraP promotes S. aureus adhesion and invasion to A549 cells through sialylated receptors

Bacterial attachment and colonization at the surface of host cells have been thought to be mediated by specific binding of BRs to glycoconjugates [1]. It has recently been reported that SraP binds to the salivary agglutinin gp340 via the Neu5Ac moiety of the trisaccharide Neu5Ac(2–3)Galβ1–4GlcNAc [25]. In fact, gp340 and homologs are also expressed in lung epithelial cells [26]. To determine if the L-lectin module mediates SraPR adhesion, we incubated a monolayer of human lung epithelial A549 cells with green fluorescent protein (GFP)-fused SraPR and individual modules. The results indicated that full-length SraPRR and the L-lectin module, but not the CDHL-1&2 modules or the β-GF module, specifically adhered to the A549 cells (Fig. 6). Try367 is a key residue to make hydrophobic interactions with sucrose and the docked Neu5Ac, we thus constructed the Y367G mutant proteins docking the Neu5Ac, we thus constructed the Y367G mutant proteins (Fig. 5E) due to the deletion of two Ca-2 coordinate residues Asp573 and Lys575, we did not find secondary-structure change in any module pairs. Moreover, merging these three structures via sequential alignment of the same module resulted in a total twist of 5.4° along the axis of SraPR structure (Fig. S1), suggesting that the slight interdomain twists of the three junctions are randomly occurred and could be canceled out. Together, the results indicated that in the presence of calcium, SraPR adopts a relatively rigid rod-like structure under all conditions tested.

SraP promotes S. aureus adhesion and invasion to A549 cells through sialylated receptors

Bacterial attachment and colonization at the surface of host cells have been thought to be mediated by specific binding of BRs to glycoconjugates [1]. It has recently been reported that SraP binds to the salivary agglutinin gp340 via the Neu5Ac moiety of the trisaccharide Neu5Ac(2–3)Galβ1–4GlcNAc [25]. In fact, gp340 and homologs are also expressed in lung epithelial cells [26]. To determine if the L-lectin module mediates SraPR adhesion, we incubated a monolayer of human lung epithelial A549 cells with green fluorescent protein (GFP)-fused SraPR and individual modules. The results indicated that full-length SraPRR and the L-lectin module, but not the CDHL-1&2 modules or the β-GF module, specifically adhered to the A549 cells (Fig. 6). Try367 is a key residue to make hydrophobic interactions with sucrose and the docked Neu5Ac, we thus constructed the Y367G mutant proteins docking the Neu5Ac, we thus constructed the Y367G mutant proteins (Fig. 5E) due to the deletion of two Ca-2 coordinate residues Asp573 and Lys575, we did not find secondary-structure change in any module pairs. Moreover, merging these three structures via sequential alignment of the same module resulted in a total twist of 5.4° along the axis of SraPR structure (Fig. S1), suggesting that the slight interdomain twists of the three junctions are randomly occurred and could be canceled out. Together, the results indicated that in the presence of calcium, SraPR adopts a relatively rigid rod-like structure under all conditions tested.

Discussion

The specific recognition of the L-lectin module to sialylated receptors may be a universal mechanism for staphyloccocal adhesion to host cells

Structural analyses combined with epithelial adhesion experiments demonstrate that SraP plays an important role in mediating bacterial adhesion to host cells by recognizing the L-lectin module. Legume lectins are a large family of proteins primarily found in the seeds of legume plants that have a similar fold but distinct carbohydrate-binding specificities [14]. They typically adopt a quaternary structure of dimers or tetramers that enhances sugar binding specificity or affinity [27]. The architecture of the legume lectin fold has also been found in the animal calcium-dependent lectin ERGIC-53/MR60, a mannose-binding protein involved in the export of soluble glycoproteins from the endoplasmic reticulum [28]. We report here the first legume lectin fold-containing protein in bacteria. This prokaryotic monomeric L-lectin module mediates adhesion to host cells by recognizing Neu5Ac, usually the non-reducing terminal residue of glycoconjugates of extracellular receptors. Bioinformatics analysis suggested that the L-lectin module might also exist in other proteins in Staphylococci and Streptococci (Fig. S4A). Moreover, the residues involved in Neu5Ac binding are relatively conserved among these putative L-lectins (Fig. S4B). Therefore, we hypothesize that this adhesion mechanism mediated by the L-lectin module operates in other staphyloccocal and streptococcal species.

The L-lectin module is projected outwards mainly by two cadherin-like modules

To scan the host receptors, the L-lectin module should be projected outwards from the bacterial surface. In addition to a long region of SRR2 that crosses the bacterial cell wall, SraP has a relatively rigid stem of three modules: a β-GF and two CDHL
modules. The classic cadherins in vertebrates bridge the intermembrane space between neighboring cells by forming trans-adhesive homodimers through membrane-distal extracellular domains [29]. The extracellular domain of most cadherins often contains three conserved Ca\(^{2+}\)-coordination sites at the interdomain junction [30]. In contrast, each CDHL module in SraP BR binds to only a single Ca\(^{2+}\) that does not superimpose on any of the three Ca\(^{2+}\) ions in eukaryotic cadherins. Nevertheless, multiple-sequence alignments suggested strong conservation among Gram-positive bacteria of SraP BR Ca\(^{2+}\)-binding residues (Fig. 8A).

Unlike the monomeric form of SraP BR, the recombinant CDHL1&2 exists as a dimer in solution as confirmed by size-exclusion chromatography and chemical cross-linking assays (Fig. S5). The structure revealed that the homodimer of CDHL1&2 buries an interface area of \(600 \, \text{Å}^2/\text{per subunit}\), as calculated by PISA (http://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver) server [31]. The interface is mainly stabilized by polar interactions, in contrast to the hydrophobic dimeric interfaces of eukaryotic cadherins [29]. CDHL-2 from one subunit packs against the junction between CDHL-1 and CDHL-2 from the symmetric subunit, and vice versa (Fig. 8B). This dimerization pattern may explain the result that SraP mediates intraspecies interaction and promotes aggregation of \(S. \, \text{aureus}\) ISP479C [7]. However, we did not observe significant decrease of aggregation or biofilm formation upon the deletion of \(sraP\) in \(S. \, \text{aureus}\) NCTC 8325.

The different results might be due to the variation of \(S. \, \text{aureus}\) strains. Sequence homology search indicated that some surface proteins in Gram-positive bacteria contain two or more CDHL modules, suggesting that those proteins with multiple CDHL modules have a high potential to mediate intraspecies aggregation.

**SraP\(_{\text{BR}}\) possesses a unique functional modularization pattern**

SRKPs have been identified in a variety of Gram-positive bacteria and function as virulence factors in a wide spectrum of infections [1]. The diversity of these infections (e.g. endocarditis, meningitis and pneumonia) correlates with the variability of BRs. Furthermore, the few BR ligands that have been identified to date range from carbohydrates (such as sialyl-T antigen) [32] to proteins (such as keratins) [33,34]. In addition to the five distinct modules defined in the four previously reported BR structures [2–4], our SraP\(_{\text{BR}}\) structure identified three types of unique modules. Notably, although the \(\beta\)-GF module resembles the Ig-binding proteins, such as B1 domain of Mub-R5 and PpL [19,21], the binding of the \(\beta\)-GF module towards human IgG, IgA or IgM could not be detected, in agreement with our structural analysis. We thus propose that the \(\beta\)-GF module, in addition to the two CDHL modules, functions as a relatively rigid stem to project the L-lectin module outwards.

The two structure-known SRRPs, Fap1 and GspB, undergo significant inter-module angle changes, which are regulated by pH and ligand binding, respectively [2,3]. In contrast, SraP\(_{\text{BR}}\) appears to adopt a relatively rigid, rod-like conformation when colonizing a host, for the concentration of free Ca\(^{2+}\) in the extracellular space or blood is stringently maintained at 1.1–1.3 mM [35]. The rigid architecture of SraP\(_{\text{BR}}\) enables the globular L-lectin module to extend outwards from the bacterial surface for scanning host receptors. This strategy to expose the functional modules has been observed for other bacterial adhesins such as the fibrillar antigen I/II from \(S. \, \text{mutans}\) [36], and the rod-like surface protein SasG from \(S. \, \text{aureus}\) [37].

**Materials and Methods**

**Cloning, expression, and purification of SraP\(_{\text{BR}}\)**

The genomic DNA from \(Staphylococcus \, \text{aureus}\) NCTC 8325 was prepared for gene cloning. The DNA sequences (GeneBank, the accession number of YP_501439.1 encoding SraP\(_{\text{BR}}\) (Phe245–Asn751) and other SraP truncates were cloned into pET28a with an N-terminal His\(_6\)-tag or pET28a with a C-terminal GFP-tag, respectively. The constructs were overexpressed in \(E. \, \text{coli}\) strain BL21 (Novagen) using LB culture medium (10 g NaCl, 10 g Bacto-Tryptone, and 5 g yeast extract per liter). The cells were
grown at 37°C to an OD$_{600}$ of 0.6. Expression of the recombinant protein was induced with 0.2 mM isopropyl b-D-1-thiogalactopyranoside (IPTG) at 16°C for another 20 hr before harvesting. Bacteria were collected by centrifugation at 8,000 g for 10 min and resuspended in 30 ml lysis buffer (20 mM Tris-Cl, pH 8.8, 100 mM NaCl). After sonication for 2.5 min followed by centrifugation at 12,000 g for 25 min, the supernatant containing the His-tagged protein was collected and loaded onto a Ni-NTA column (GE Healthcare) equilibrated with the binding buffer (20 mM Tris-Cl, pH 8.0, 100 mM NaCl). The target protein was eluted with 300 mM imidazole, and loaded onto a Superdex 200 column or Superdex 75 column (GE Healthcare; 20 mM Tris-Cl, pH 8.0, 100 mM NaCl). The purity of protein was assessed by electrophoresis and the protein sample was stored at −20°C.

The selenium-Met (SeMet) labeled L-lectin&b-GF protein was expressed in E. coli strain B834 (DE3) (Novagen). Transformed cells were grown at 37°C in SeMet medium (M9 medium with 25 μg/ml SeMet and the other essential amino acids at 50 μg/ml) containing 30 μg/ml kanamycin until the OD$_{600}$ reached 0.6, and were then induced with 0.2 mM IPTG at 16°C for 20 hr. SeMet substituted protein was purified with the same procedure as the native protein.

The oligomerization state of CDHL-1&2 was analyzed by Superdex 75 column. The protein markers are bovine serum albumin, ovalbumin, chymotrypsinogen A, myoglobin and ribonuclease A, which have a molecular weight of 67, 44, 25, 17 and 13.7 kDa, respectively (GE Healthcare).

Crystallization, data collection and processing

All crystals were grown using the hanging drop vapor diffusion method, with a drop of 1 μl protein solution mixed with 1 μl of reservoir solution equilibrated against 500 μl of the reservoir solution. The proteins for crystallization were concentrated by ultrafiltration (Millipore Amicon) to 30, 38, 20 and 20 mg/ml for the full-length SraP$_{BR}$, L-lectin&b-GF, b-GF&CDHL-1 and CDHL-1&2, respectively. The SeMet substituted L-lectin&b-GF protein for crystallization was concentrated to 38 mg/ml. Crystals of SraP$_{BR}$ were grown at 28°C, whereas others were grown at 16°C. Crystals were obtained from 0.8 M (NH$_4$)$_2$SO$_4$ and 0.1 M MES, pH 6.0 for SraP$_{BR}$; 2.5 M sodium formate, 0.1 M sodium acetate pH 4.6 for the native L-lectin&b-GF; 10% PEG 6000, 0.1 M MES, pH 6.0, and 1.0 M lithium chloride for the SeMet substituted L-lectin&b-GF; 2.0 M (NH$_4$)$_2$PO$_4$, 0.1 M Tris-Cl,
pH 8.5 for b-GF&CDHL-1 and 1.8 M (NH₄)₂SO₄, 0.1 M HEPES, pH 7.5 for CDHL-1&2. The crystals of SraP BR, b-GF&CDHL-1 and CDHL1&2 were transferred to the cryoprotectant with the reservoir solution supplemented with 50% sucrose. The cryoprotectant for the crystals of L-lectin&b-GF consists of the reservoir solution supplemented with 30% glycerol. All crystals in the cryoprotectant were flash-cooled with liquid nitrogen prior to X-ray diffraction. Data for a single crystal were collected at 100 K in a liquid nitrogen stream using beamline 17U with a Q315r CCD (ADSC, MARresearch, Germany) at the Shanghai Synchrotron Radiation Facility (SSRF). All diffraction data were integrated and scaled with the program HKL2000 [38].

Structure determination and refinement

The crystal structure of L-lectin&b-GF was determined using single-wavelength anomalous dispersion (SAD) phasing [39] method from a single crystal of SeMet-substituted protein to a maximum resolution of 2.10 Å. The AutoSol program implemented in PHENIX [40] was used to locate the selenium atoms and calculate the phase, which was further improved with the program Buccaneer [41]. Automatic model building was carried out using Autobuild in PHENIX. The initial model was refined in REFMAC5 [42] and Phenix.refine and rebuilt interactively using the program COOT [43]. The model was used as the search model against 2.05 Å SraP BR data by molecular replacement using Molrep program as part of CCP4i [44] program suite. Electron density maps showed clear features of secondary structural elements for automatically building the C-terminal tandem cadherin-like modules using IPCAS [45]. The structures of b-GF&CDHL-1 and CDHL1&2 were determined by molecular replacement using the corresponding modules in the full-length SraP BR structure as the search model. The initial models were refined by simulated annealing using Phenix.refine to reduce the phase bias. Then the models were refined interactively using COOT and REFMAC5 until the R-factor and R-free values converged. All final models were evaluated with the programs MOLPROBITY [46] and PROCHECK [47]. Crystallographic parameters were listed in Table 1. The |Fo| - |Fc| omit map of the sucrose molecule contoured at 3 σ was calculated by FFT implemented in CCP4i. All structure figures were prepared with PyMOL (http://www.pymol.org/).

Determination of metals binding to the protein

The purified SraP BR, CDHL-1&2, BRV367G and L-lectinV367G in 20 mM Tris-Cl, pH 8.0, 100 mM NaCl were concentrated to 30, 38, 30 and 21 mg/ml, respectively, and applied to the analyses. Briefly, 500 μl of protein sample was subjected to digestion by the aqueous method using the HNO₃ and HClO₄.
Afterwards, the digested samples were diluted with deionized water and analyzed by atomic absorption spectroscopy (Atomscan Advantage, Thermo Ash Jarell Corporation, USA).

Surface plasmon resonance (SPR) assays

The binding affinities of the L-lectin module towards varying monosaccharides were determined by SPR. SPR experiments were performed at 25°C using a Biacore 3000 instrument using HBS (10 mM HEPES, pH 7.5, 150 mM NaCl) containing 0.005% (v/v) Tween 20 and a flow rate of 5 μl/min. The L-lectin module was covalently immobilized on the carboxymethyldextran surface of the CM5 chip. The chip was activated with EDC (N-ethyl-N-[3-dimethylaminopropyl] carbodi-imide)/NHS (N-hydroxysuccinimide) solution, and the L-lectin module in 10 mM acetate buffer (pH 5.5) was injected into the flow channel. At the end, the sensor surface was blocked with 1 M ethanolamine. The blank channel was treated in the same way without protein injected. Each monosaccharide in the running buffer was incubated for 1 min in the flow-cells using the kinject mode. Both injection and dissociation steps last for 5 min. The sensor surface was regenerated with 50 mM NaOH. All analyses were performed with the BIAnval software. The equilibrium responses were plotted versus monosaccharide concentrations and fitted to a 1:1 Langmuir binding model using the Origin 8.0 software (OriginLab Corp.).

Chemical cross-linking

Chemical cross-linking of purified CDHL-1&2 was performed using formaldehyde and bis(sulfosuccinimidyl) suberate (BS³), which is a homobifunctional sulfo-N-hydroxysuccinimide ester (4:1, v/v) method. Afterwards, the digested samples were diluted with deionized water and analyzed by atomic absorption spectroscopy (Atomscan Advantage, Thermo Ash Jarell Corporation, USA).

### Table 1. Crystal parameters, data collection and structure refinement.

|                      | BR | SeMet- L-lectin&β-GF | β-GF&CDHL-1 | CDHL-1&2 |
|----------------------|----|----------------------|-------------|----------|
| **Data collection**  |    |                      |             |          |
| Space group          | P3_21 | C2                  | P2_1        | P3_21    |
| Unit cell            |      | a, b, c (Å)         |             |          |
| α, β, γ (°)          | 90.00 | 90.00, 120.00       | 90.00, 89.77, 90.00 | 90.00, 90.000, 120.00 | 90.00, 90.000, 120.00 |
| Resolution range (Å) | 50.00–2.05 (2.15–2.05) | 50.00–2.10 (2.18–2.10) | 50.00–1.60 (1.64–1.60) | 50.00–2.24 (2.32–2.24) |
| Unique reflections   | 69,226 (6,805) | 99,475 (9,904) | 18,581 (1,236) | 8,890 (866) |
| Completeness (%)     | 100 (100) | 100 (100) | 99.6 (99.1) | 96.6 (96.7) |
| I/σ(I)               | 13.8 (5.8) | 14.3 (5.6) | 10.0 (3.7) | 16.7 (8.8) |
| Rmerge (%)           | 13.7 (38.5) | 15.8 (35.7) | 10.3 (32.3) | 13.1 (38.9) |
| Average redundancy   | 6.9 (6.8) | 6.9 (6.9) | 3.0 (2.6) | 8.6 (5.7) |
| **Structure refinement** | | | | |
| Resolution range (Å) | 50.00–2.05 | 44.91–2.10 | 46.44–1.60 | 50.00–2.24 |
| R-factor/ R-free (%) | 18.2/20.5 | 15.8/19.5 | 16.4/20.2 | 27.9/32.6 |
| Number of protein atoms | 3,675 | 9,566 | 1,251 | 1,181 |
| Number of water atoms | 618 | 1,448 | 267 | 8 |
| RMSD bond lengths (Å) | 0.009 | 0.011 | 0.008 | 0.012 |
| RMSD bond angles (°) | 1.285 | 1.337 | 1.220 | 1.569 |
| Mean B factors (Å²) | 29.5 | 21.4 | 21.0 | 51.9 |
| Ramachandran plot (%) | 98.2 | 98.8 | 98.2 | 97.5 |
| Most favored (%)     | 98.2 | 98.8 | 98.2 | 97.5 |
| Additional allowed (%) | 1.8 | 1.2 | 1.8 | 2.5 |
| Outliers (%)         | 0 | 0 | 0 | 0 |
| PDB entry            | 4M00 | 4M01 | 4M02 | 4M03 |

*The values in parentheses refer to statistics in the highest bin.

[1] Rmerge = \(\sum_{i<j} \sum_{h} \sum_{k} ((I_{hkij} - \langle I_{hkij} \rangle)/\langle I_{hkij} \rangle)\), where \(I_{hkij}\) is the intensity of an observation and \(\langle I_{hkij} \rangle\) is the mean value for its unique reflection; Summations over all reflections.

[2] R-factor = \(\sum_{i<j} |F_i| - |F_c|)/\sum_{i<j} |F_i|\), where \(|F_i|\) and \(|F_c|\) are the observed and calculated structure-factor amplitudes, respectively.

[3] R-free was calculated with 5% of the data excluded from the refinement.

[4] Root-mean square-deviation from ideal values.

[5] Categories were defined by Molprobity.

[6] doi:10.1371/journal.ppat.1004169.t001

Crystal Structure of the Binding Region of SraP

PLOS Pathogens | www.plospathogens.org 10 June 2014 | Volume 10 | Issue 6 | e1004169
analogue with a spacer arm length of 1.14 nm (Pierce). Briefly, for formaldehyde cross-linking assays of CDHL-1&2, 20 µl of the recombinant protein (2 mg/ml) was mixed with 20 µl PBS containing 2% formaldehyde, and the samples were incubated at 25°C for 30 min and 1 hr, respectively. For BS³ cross-linking assay of CDHL-1&2, 100 µl recombinant protein (2 mg/ml) was incubated with 5 mM BS³ at 25°C for 30 min and 1 hr, respectively. The reaction was quenched by the addition of 20 mM Tris·HCl, pH 8.0. Then the samples were separated by 10% SDS-PAGE, and were stained by Coomassie Brilliant Blue G250. Molecular mass markers for SDS-PAGE were purchased from Thermo Scientific (Wilmington, DE); beta-galactosidase, bovine serum albumin, ovalbumin, lactate dehydrogenase, R Fuse Bsp98I, beta-lactoglobulin, lysozyme, which have a molecular weight of 116, 66.2, 45, 35, 25, 18.4 and 14.4 kDa, respectively.

**Bacterial strains, media and growth conditions**

*Staphylococcus aureus* NCTC 8325 and its derivative strains were grown in LB medium, and when necessary, erythromycin (2.5 mg/ml) and chloramphenicol (15 mg/ml) were added. To generate the insertion-deletion mutagenesis, approximately 500 bp of upstream and downstream fragments of the BR region of *sraP* gene was amplified by polymerase chain reaction (PCR), digested with *Pst*I and *Sal*I and *Bam*HI/Xhol, respectively, and then ligated to either end of a double-digested (*Bam*HI/SalI) erythromycin-resistance gene (*ermβ*), which was amplified from plasmid pEC1. The three fragments were ligated with the erythromycin-resistance gene in the middle, and then cloned in the temperature-sensitive shuttle vector pBT2. The resulting plasmid was transformed by electroporation into *S. aureus* strain RN4220 for propagation, and then transformed into *S. aureus* NCTC 8325 for allelic exchange. Mutants were screened and further checked by PCR and sequencing.

**Cell adhesion and invasion assays**

Adhesion and invasion experiments were performed as described previously [50]. A549 human respiratory epithelial cells were grown in Dulbecco's modified Eagle's medium (DMEM, Gibco) supplemented with 10% fetal bovine serum (FBS), 5 mM glutamine, penicillin (5 µg/ml) and streptomycin (100 µg/ml). Approximately 5×10⁵ cells were seeded into 24-well tissue culture plates, and allowed to grow in 5% CO₂ at 37°C. Before use, the monolayers were washed twice with PBS. For adhesion assays, *S. aureus* grown to OD₆00nm of 0.5 in LB medium were resuspended in cell culture medium without serum and antibiotics, and incubated on ice for 2 hr with the monolayer of A549 cells, The cells were incubated with DMEM media (Gibco) supplemented with 10% fetal bovine serum, 5 mM glutamine, penicillin (5 µg/ml) and streptomycin (100 µg/ml). The monolayers were washed three times with sterile PBS and lysed with 0.05% Triton X-100. The internalized bacteria were counted by plating serial dilutions of the recovered bacterial suspensions onto LB agar. Experiments were performed in triplicate. Data corresponding to adhesion and invasion were compared using the Mann-Whitney tests. Statistical differences were determined with the t-test.

**Immunofluorescence assays**

Immunofluorescence assays were performed as described previously [51]. Briefly, GFP and GFP-fused proteins at 10 µM were suspended in the media in the absence of serum and antibiotics, and incubated on ice for 2 hr with the monolayer of A549 cells grown on the 12-mm glass coverslips. For inhibition assays, Neu3Ac at 5 and 10 mM was pre-incubated with GFP-fused SraPBR protein, respectively. After incubation, cells were sequentially washed three times with cold PBS, fixed with 4% paraformaldehyde for 10 min, permeabilized with 0.1% Triton X-100 for 2 min, and incubated with GFP-tag mouse antibody and a fluorescein isothiocyanate (FITC)-conjugated goat anti-mouse IgG antibody. The nuclei were stained with 4',6-diamidino-2-phenylindole (DAPI) reagent. Slides were examined with a Zeiss LSM710 confocal scanning fluorescence microscope (Carl Zeiss, Jena, Germany) with a Plan-Apocromat 20×/0.8NA objective. Confocal parameters set for immunofluorescence detection were taken as standard settings. The excitation wavelength is 405 nm and the emission wavelength is 410–492 nm. The confocal images were collected and processed with the software ZEN 2009. Experiments were performed in triplicate, with three or more replicate wells tested for each experimental condition.

**Molecular dynamics simulations**

Molecular dynamics simulations were performed for SraPBR in the Ca²⁺-bound and the Ca²⁺-free states, respectively. For each simulation, the system was placed in a TIP3P [52] water box with a distance of at least 12 Å to the box boundaries. Ions were added to neutralize the system and to result in a concentration of 0.15 M NaCl. The solvated protein was subjected to energy minimization employed the steepest descent algorithm and conjugate gradient, respectively. Simulations were performed with a parallel implementation of the GROMACS (version 4.5.5) package [53] using the AMBER03 force field [54]. MD productions were run for 20 ns using a time step of 2 fs and the NPT ensemble [55]. Covalent bonds were constrained using the LINCS algorithm [56], while the cutoff distances for the Coulomb and van der Waals interactions were set to 0.9 and 1.4 nm, respectively. The long-range electrostatic interactions were treated by the PME algorithm [57] with a tolerance of 10⁻³ and an interpolation order of 4. Structure visualization was performed with VMD [58].

**Small-angle X-ray scattering (SAXS) experiments**

SAXS was used to investigate the overall conformations of SraPBR in the presence or absence of calcium. The full-length SraPBR at 1.0 and 5.0 mg/ml was analyzed, either in 10 mM calcium chloride or in 50 mM EDTA. SAXS data were collected on the 12ID beamline of Advanced Photon Sources (APS) at the Argonne National Laboratory using the Plutus 2M detector (DECTRIS, Switzerland). The scattering patterns were measured with a 1–2 second exposure time for each collected frame, and twenty frames were taken for each sample to optimize the signal-to-noise ratio. To reduce the radiation damage, a flow cell made of a cylindrical quartz capillary with a diameter of 1.5 mm and a wall of 10 µm was used during the data collection process. No concentration effect was observed. All SAXS curves were measured at the room temperature over the range of momentum transfer 0.006<s<0.82 Å⁻¹ (where s = 4π sin(θ)/λ, 2θ is the scattering angle, and the X-ray wavelength λ = 1.03 Å). The data corresponding to adhesion and invasion were compared using the Mann-Whitney tests. Statistical differences were determined with the t-test.
were processed using the PRIMUS [59] program package and standard procedures. The forward scattering (I(0)) and the radius of gyration (Rg) were evaluated using the Guinier approximation assuming that at very small angles (s<1.3/Rg) the intensity is represented as I(0) - s^2R_g^2.

The program of GNOM [60] was used to calculate D_max and the interatomic distance distribution function p(r). The particle shape of each measured sample was reconstructed ab initio using the programs DAMMIN [61] and GASBOR [62]. The scattering patterns of the atomic crystal structures for SraP were calculated using the program CRYSOL. For the ab initio analyses and modeling, multiple runs were performed to verify the stability of the solution.

Supporting Information

**Figure S1** The intermodule twist along the axis of SraP BR. The merged structure of SraP BR (in orange) was generated by sequentially superimposing the same module against each other from the three structures (L-lectin&β-GF, β-GF&CDHL-1 and CDHL-1&2). The L-lectin, β-GF, CDHL-1 and CDHL-2 modules of SraP BR are shown in cyan, red, yellow and green, respectively. The merged SraP BR was superimposed against the crystal structure of SraP BR with the two CDHL-2 modules aligned. (TIF)

**Figure S2** The CD spectra of A) SraP BR and B) the L-lectin module. The results demonstrated that mutation of Y367G did not introduce significant changes to the protein structures. (TIF)

**Figure S3** Quantitation of the GFP fluorescent signals of SraP BR mutants. The fluorescent signals for each protein were quantified by calculating the mean gray values of three replicates using the ImageJ software (http://imagej.nih.gov/ij/). The average mean gray value and standard error of the mean (SEM) derived from triplicate treatments are indicated as bar graph. (TIF)

References

1. Lizardo A, Sanchez CJ, Orihuela CJ (2012) A role for glycosylated serine-rich repeat proteins in Gram-positive bacterial pathogenesis. Mol Oral Microbiol 27: 237–269.

2. Ramboarina S, Garnett JA, Zhou M, Li Y, Peng Z, et al. (2010) Structural insights into serine-rich fimbrine from Gram-positive bacteria. J Biol Chem 285: 32446–32457.

3. Pyburn TM, Bensing BA, Xiong YQ, Melancon BJ, Tomasiak TM, et al. (2011) A structural model for binding of the serine-rich repeat adhesin GspH to host cellular receptors. PLoS Pathog 7: e1002112.

4. Seo HS, Misnov G, Serepansad R, Doran KS, Dubrovskia I, et al. (2013) Characterization of fibronectin binding by glycoproteins Sr1 and Sr2 of *Streptococcus agalactiae*. J Biol Chem 288:35962–96.

5. Lowy FD (1998) Staphylococcus aureus infections. N Engl J Med 339: 520–532.

6. Shibor RI, Chambers HF, Sullam PM (2005) Role of SraP, a serine-rich surface protein of *Staphylococcus aureus*, in binding to human platelets. Infect Immun 73: 2273–2280.

7. Sanchez CJ, Shivshankar P, Stol K, Trakhtenbrot S, Sullam PM, et al. (2010) The pneumococcal serine-rich repeat protein is an intra-species bacterial carbohydrate receptor. PLoS Pathog 6: e100044.

8. Holm L, Rosenstrom P (2010) Dali server: conservation mapping in 3D. Nucleic Acids Res 38: W345–349.

9. Pietrocz VZ, Raul JC, AS, Nt, Tsyanik NN, Mikhailova IV, Duax W, et al. (1997) The structure of pea lectin-D-glucopyranose complex at a 1.9 A resolution. RUSs. JBioorganic Chem 23: 469–478.

10. Rahibis A, Verboven C, Rouge P, Barre A, Van Damme EJ, et al. (2001) Structure of a leucine lectin from the bark of Robinia pseudoacacia and its complex with N-acetylgalactosamine. Proteins 44: 470–478.

11. Murzin AG, Brenner SE, Hubbard T, Chothia C (1995) SCOP: a structural classification of proteins database for investigation of sequences and structures. J Mol Biol 246: 536–540.

12. Jin X, Walker MA, Felsowal V, Vendome J, Bahna F, et al. (2012) Crystal structures of *Daphnia* N-aceladherin ectodomain regions reveal a widely used class of Ca2+-free interdomain linkers. Proc Natl Acad Sci U S A 109: E127–134.

13. Somotayor M, Weihofen WA, Gaudet R, Corey DP (2012) Structure of a force-conveying cadherin bond essential for inner-ear mechanotransduction. Nature 492: 128–132.

14. Sharon N, Li H (1990) Legume lectins—a large family of homologous proteins. FASEB J 4: 3198–3202.

15. Loris R, Hameltyck T, Bourcart J, Wyns L (1998) Legume lectin structure. Biochim Biophys Acta 1393: 9–36.

16. Dowerer E, Fison N, Roche AG, Monsigny M (2003) Carbohydrate-lectin interactions assessed by surface plasmon resonance. Biochimie 85: 167–179.

17. MacKenzie DA, Tailford LE, Hamings AM, Juge N (2009) Crystal structure of a mucus-binding protein repeat reveals an unexpected functional immunoglobulin binding activity. J Biol Chem 284: 32444–32453.

18. Graille M, Stura EA, Beckingham JA, Bottomley SP, et al. (2001) Complex between *P. magnus* protein L and a human antibody reveals structural convergence in the interaction modes of Fab binding proteins. Structure 9: 679–687.

19. Bjorck L (1986) Protein L: A novel bacterial cell wall protein with affinity for Ig L chains. Immunol 140: 1194–1197.

20. Wikstrom M, Drakenberg T, Forsen S, Sjobling U, Bjorck L (1994) Three-dimensional solution structure of an immunoglobulin light chain-binding domain of protein L. Comparison with the IgG-binding domains of protein G. Biochemistry 33: 14011–14017.

21. Beckingham JA, Bottomley SP, Hintz R, Sutton BJ, Geoe MG (1999) Interactions between a single immunoglobulin-binding domain of protein L from *P. magnus* and a human k light chain. Biochem J 340 (Pt 1): 193–199.

22. Graille M, Harrison S, Crump MP, Finlow SB, Housden NG, et al. (2002) Evidence for plasticity and structural mimicry at the immunoglobulin light chain-protein L interface. J Biol Chem 277: 47500–47506.

23. Berman H, Heurick K, Nakamura H, Markley JL (2007) The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. Nucleic Acids Res 35: D301–303.

24. Hura GL, Menon AL, Hannel M, Rambo RP, Plooe FL, 2nd, et al. (2009) Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). Nat Methods 6: 606–612.

**Figure S4** The L-lectin module is conserved in *Staphylococci* and some species of *Streptococci*. A) Schematic of *Staphylococci* and *Streptococci* proteins containing an L-lectin module. The first line represents the SraP protein from *S. aureus*. Similar modularization is observed in *S. epidermidis* and *S. warneri* (2nd and 3rd line). The 4th line represents uncharacterized protein from *S. australis*. SP: signal peptide, CWA: cell wall anchor motif, Rib: a repeat motif in Rib protein of group B *Streptococcus*, Ig/albumin-bd: immunoglobulin and albumin-binding domain. B) Multiple-sequence alignment of the L-lectin module and homologs. The Neu5Ac binding residues (red triangle) are conserved. (TIF)

**Figure S5** CDHL1&2 exists as a dimer in solution. A) The size-exclusion chromatogram profile of CDHL-1&2. The standard curve was inserted as an inlet. Chemical cross-linking of the purified CDHL-1&2 proteins using B) formaldehyde (FA) and C) bis(sulfosuccinimidy) suberate (BS). The protein samples were separated by 10% SDS-PAGE. The bands corresponding to the monomer and the dimer of CDHL1&2 are labeled. (TIF)

Acknowledgments

We thank the staff at the Shanghai Synchrotron Radiation Facility for technical assistance.

Author Contributions

Conceived and designed the experiments: YHY YLJ YC CZZ. Performed the experiments: YHY YLJ JZ LW XHB SJZ YMR YHZ. Analyzed the data: YHY YLJ. Contributed reagents/materials/analysis tools: NL ZZ QG YM TX. Wrote the paper: YHY YLJ JRZ YC CZZ.
25. Kukita K, Kawada-Matsuo M, Oto H, Nagatomo M, Oosugi Y, et al. (2013) *Streptococcus agalactiae* SaaxA is responsible for binding to the salivary agglutinin gp340, derived from human saliva. Infect Immun 81: 1870–1879.

26. Holmåker U, Lawson P, Teisner B, Torneé I, Willis AC, et al. (1997) Isolation and characterization of a new member of the scavenger receptor superfamily, glycoprotein-340 (gp-340), as a lung surfactant protein-D binding molecule. *J Biol Chem* 272: 13743–13749.

27. Srinivas VR, Reddy GB, Ahmad N, Swaminathan CP, Mitra N, et al. (2001) Legume lectin family, the 'natural mutants of the quaternary state', provide insights into the relationship between protein stability and oligomerization. *Biochim Biophys Acta* 1527: 102–111.

28. Vellonos LM, Svensson K, Schauier G, Petersson RF, Lindqvist Y (2002) Crystal structure of the carbohydrate recognition domain of p58/ERGIC-53, a protein involved in glycoprotein export from the endoplasmic reticulum. *J Biol Chem* 277: 13997–13984.

29. Brasch J, Harrison OJ, Honig B, Shapiro L (2012) Thinking outside the cell: how cadherins drive adhesion. *Trends Cell Biol* 22: 299–310.

30. Boggon TJ, Murray J, Chappuis-Flament S, Wong E, Gumbiner BM, et al. (2002) C-cadherin ectodomain structure and implications for cell adhesion mechanisms. *Science* 296: 1308–1313.

31. Kristján E, Henrik K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372: 774–797.

32. Birnbaum BA, Lopez JA, Sullam PM (2004) The *Streptococcus gordonii* surface proteins GspB and Hsa mediate binding to sialylated carbohydrate epitopes on the platelet membrane glycoprotein IX. *Infect Immun* 72: 6526–6537.

33. Samen U, Eikmanns BJ, Reinscheid DJ, Borges F (2007) The surface protein Ssr1 of *Streptococcus agalactiae* binds human Keratin 4 and promotes adherence to epithelial HEP-2 Cells. *Infect Immun* 75: 5405–5414.

34. Shivshankar P, Sanchez C, Rose LF, Orihuela CJ (2009) The *Streptococcus pneumoniae* adhesin PspA binds to Keratin 10 on lung cells. *Mol Microbiol* 73: 663–679.

35. Maurer P, Hohenester E, Engel J (1996) Extracellular calcium-binding proteins. *Curr Opin Cell Biol* 8: 609–617.

36. Larson MR, Rajashankar KR, Patel MH, Robinette RA, Crowley PJ, et al. (2010) Elongated fibrillar structure of a *Staphylococcal* adhesin assembled by the high-affinity association of α- and PPII-helices. *Proc Natl Acad Sci U S A* 107: 5983–5988.

37. Gruszka DT, Wojdyla JA, Bingham RJ, Turkenburg JP, Manfield IW, et al. (2002) *C-cadherin* ectodomain structure and implications for cell adhesion. *Acta Crystallogr D Biol Crystallogr* 62: 1002–1011.

38. Otwonowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol*, Pt A 276: 307–326.

39. Brodersen DE, de La Fortelle E, Vonrhein C, Bricogne G, Nyborg J, et al. (2000) PHASE: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 56: 431–441.

40. Adams PD, Afonine PV, Bankoczi G, Chen VB, Davis IW, et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66: 215–221.

41. Cowtan K (2006) The Buccaneer software for automated model building. I. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* 62: 1002–1011.

42. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53: 240–255.

43. Emaley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60: 2126–2127.

44. Collaborative Computational Project N (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50: 760–765.

45. Duan Y, Gu YX, Zheng CD, Fan HF (2000) OASIS: a computer program for breaking phase ambiguity in one-wavelength anomalous scattering or single isomorphous substitution (replacement) data. *J Appl Crystallogr* 33: 980–981.

46. Davis I, Leaver-Fay A, Chen V, Block J, Kapoor G, et al. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35: W375–383.

47. Laskowski R, Macarthur M, Moss D, Thornton J (1993) Procheck - A program to Check the Stereochemical Quality of Protein Structures. *J Appl Crystallogr* 26: 283–291.

48. Trout O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31: 455–461.

49. Samer MF (1999) Python: a programming language for software integration and development. *J Mol Graph Model* 17: 57–61.

50. Valle J, Latasa G, Gil C, Toledo-Arana A, Solano C, et al. (2012) Bap, a biofilm matrix protein of *Streptobacillus moniliformis* prevents cellular internalization through binding to GP96 host receptor. *PLoS Pathog* 8: e1002843.

51. Nelson AL, Ries J, Bagnoli F, Dahlberg S, Falkert S, et al. (2007) BrgA is a pilus-associated adhesin in *Streptococcus pneumoniae*. *Mol Microbiol* 66: 329–340.

52. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79: 926–935.

53. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* 4: 143–147.

54. Hao Q, Gu YX, Zhou J, Fan HF (2000) OASIS: a computer program for breaking phase ambiguity in one-wavelength anomalous scattering or single isomorphous substitution (replacement) data. *J Appl Crystallogr* 33: 980–981.

55. Konarev PV, Volkov VV, Sokolova AV, Koch MJH, Svergun DI (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J Appl Crystallogr* 36: 1277–1282.

56. Svergun DI (1992) Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. *J Appl Crystallogr* 25: 495–505.

57. Svergun DI (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing (vol 76, pg 2079, 1999). *Biophys J* 77: 2086–2086.

58. Svergun DI, Petoukhov MV, Koch MJH (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* 80: 2946–2953.