

# Structural and Functional Analysis of the N-terminal Domain of the *Streptococcus gordonii* Adhesin Sgo0707

Åsa Nylander<sup>1</sup>, Gunnel Svensäter<sup>2</sup>, Dilani B. Senadheera<sup>3</sup>, Dennis G. Cvitkovitch<sup>3</sup>, Julia R. Davies<sup>2</sup>, Karina Persson<sup>4\*</sup>

**1** Department of Odontology, Division of Oral Microbiology, Umeå University, Umeå, Sweden, **2** Department of Oral Biology, Faculty of Odontology, Malmö University, Malmö, Sweden, **3** Dental Research Institute, University of Toronto, Toronto, Canada, **4** Department of Chemistry, Umeå University, Umeå, Sweden

## Abstract

The commensal *Streptococcus gordonii* expresses numerous surface adhesins with which it interacts with other microorganisms, host cells and salivary proteins to initiate dental plaque formation. However, this Gram-positive bacterium can also spread to non-oral sites such as the heart valves and cause infective endocarditis. One of its surface adhesins, Sgo0707, is a large protein composed of a non-repetitive N-terminal region followed by several C-terminal repeat domains and a cell wall sorting motif. Here we present the crystal structure of the Sgo0707 N-terminal domains, refined to 2.1 Å resolution. The model consists of two domains, N1 and N2. The largest domain, N1, comprises a putative binding cleft with a single cysteine located in its centre and exhibits an unexpected structural similarity to the variable domains of the streptococcal Antigen I/II adhesins. The N2-domain has an IgG-like fold commonly found among Gram-positive surface adhesins. Binding studies performed on *S. gordonii* wild-type and a Sgo0707 deficient mutant show that the Sgo0707 adhesin is involved in binding to type-1 collagen and to oral keratinocytes.

**Citation:** Nylander Å, Svensäter G, Senadheera DB, Cvitkovitch DG, Davies JR, et al. (2013) Structural and Functional Analysis of the N-terminal Domain of the *Streptococcus gordonii* Adhesin Sgo0707. PLoS ONE 8(5): e63768. doi:10.1371/journal.pone.0063768

**Editor:** Indranil Biswas, University of Kansas Medical Center, United States of America

**Received:** January 11, 2013; **Accepted:** April 5, 2013; **Published:** May 17, 2013

**Copyright:** © 2013 Nylander 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 study was supported by Umeå Center for Microbial Research and the Swedish Research Council grant 2011-4186 to KP, NIH grant R01DE013230-03 and CIHR-MT15431 to DGC, the Swedish Dental Society and the Knowledge Foundation to JRD and GS. The Aberdeen Proteome Facility is funded by the SHEFC and BBSRC. ÅN is supported by The Swedish National Graduate School in Odontological Science. Case No 01944647. 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: karina.persson@chem.umu.se

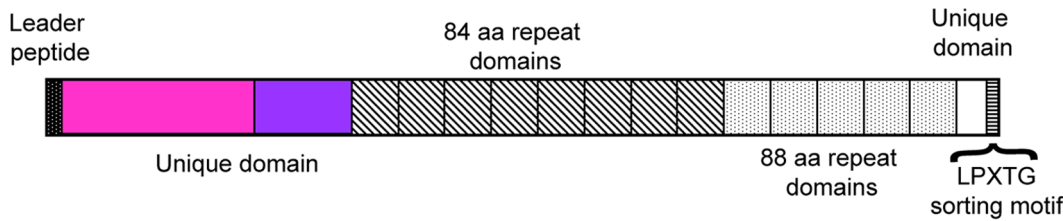
## Introduction

Oral streptococci are the most abundant bacteria in the oral cavity and approximately 70% of early colonizers belong to the streptococcal family [1]. Colonization occurs through bacterial adherence to molecules in the saliva-derived pellicle which covers all surfaces in the oral cavity. The salivary pellicle contains proteins, peptides and other molecules and so far 130 proteins have been identified [2]. The bacteria utilize a variety of cell surface proteins in order to adhere to a surface and thereby avoid clearance through swallowing. After adhesion, the colonizing bacteria themselves present new surfaces for adhesion by secondary colonizers. The commensal strain *Streptococcus gordonii* expresses an array of surface adhesins, for example the Antigen I/II (AgI/II) proteins SspA and SspB, and CshA and CshB that mediate interactions with salivary agglutinin (gp340) [3,4]. Other examples are the two serine-rich cell surface glycoproteins GspB and Hsa that have a large number of binding partners, such as salivary MUC7 [5], secretory IgA [6], gp340 [7] and the platelet glycoprotein Ib $\alpha$  [8]. Although *S. gordonii* is mainly beneficial for oral health, the bacteria can become pathogenic if they spread to non-oral sites such as the heart valves with infective endocarditis as the result [9].

Surface adhesins on Gram-positive bacteria can adopt very different structures and depend on different forms of bioassembly. Many bacteria express pili which are long polymers of covalently linked pilins with an adhesin presented at the tip [10]. In contrast,

the AgI/II adhesins expressed by oral streptococci adopt a unique monomeric structure where a central variable domain is presented as the tip on a stalk formed by intertwining flanking regions [11]. A third form of surface adhesins is built up from an N-terminal adhesion-mediating domain presented on a fibrillar stalk formed by a number of C-terminal repeat units as described for *e.g.* the *S. gordonii* adhesins CshA and GspB [12,13], the *Streptococcus parasanguinis* adhesin Fap1 [14] and the *Streptococcus pyogenes* adhesin Epf [15]. A common feature among all these proteins is the presence of a C-terminal LPXTG-like sorting motif which is recognized by the enzyme sortase A (SrtA). This enzyme covalently links the protein to the cell wall. In a study on a *S. gordonii* SrtA deletion mutant, four previously unidentified LPXTG-containing surface proteins were detected at high levels by the examination of excreted proteins [16].

One of the proteins identified, Sgo0707, was present in high amounts and was predicted to be a fibrillar adhesin. The protein precursor for Sgo0707 is 1643 amino acids long and can be divided into several regions by examination of the sequence. A leader peptide is followed by an N-terminal domain of 419 amino acids. Next follows an 84-amino acid sequence that is repeated eight times followed by an 88-amino acid sequence, repeated five times. Prior to the LPXTG sorting motif and transmembrane helix is a small unique domain (Fig. 1). We hypothesize that the N-terminal domain functions as the adhesin and that the repeat



**Figure 1. Domain architecture of Sgo0707.** Schematic drawing of the domain architecture of Sgo0707. The 1643-amino acid long protein consists of several regions: a leader peptide, a unique N-terminal domain, a number of repeat sequences, a short unique C-terminal domain and an LPXTG sorting motif.

doi:10.1371/journal.pone.0063768.g001

domains build up the stalk of the protein, similar to that described previously for streptococcal adhesins [12,13,14,15].

To add to our knowledge regarding the structure and function of adhesins from Gram-positive bacteria in general and oral streptococci in particular, we have solved the crystal structure of the N-terminal domain from *S. gordonii* Sgo0707 to 2.1 Å resolution. We also present the result of binding studies of Sgo0707 to keratinocytes, serum, saliva, type-1 collagen and a set of glycan structures.

## Materials and Methods

### Bacterial Strains and Culture Conditions

Wild-type *S. gordonii* DL-1 (Challis), and the  $\Delta$ Sgo0707 strain, stored at  $-70^{\circ}\text{C}$  in skim milk (Oxoid), were grown on blood agar in 5%  $\text{CO}_2$  in air at  $37^{\circ}\text{C}$  for 24 h. For biofilm experiments, inocula were prepared by transferring colonies into modified *Actinomyces* defined medium (m-ADM) and incubating in an atmosphere of 5%  $\text{CO}_2$  in air at  $37^{\circ}\text{C}$  overnight. Erythromycin was added to the medium at a concentration of 50  $\mu\text{g}/\text{ml}$  for  $\Delta$ Sgo0707. Aliquots (500  $\mu\text{l}$ ) were transferred to 4.5 ml fresh m-ADM and incubated in 5%  $\text{CO}_2$  in air at  $37^{\circ}\text{C}$ ; the  $\text{OD}_{600}$  was monitored until the mid-exponential growth phase ( $\text{OD}_{600} = 0.6 \pm 0.1$ ) was reached. Cells were harvested by centrifugation (3000  $g$ , 5 min at  $4^{\circ}\text{C}$ ), washed in 10 mM PBS (pH 7.5) and re-suspended in m-ADM to a cell concentration of  $1 \times 10^7/\text{ml}$ .

### Cloning of the Sgo0707 N-terminal Domain

Residues encoding 36–458 from *sgo0707* (UniProt A8AW49) were amplified by PCR using forward primer 5'-aaaaaacatggcgttagaagagattaaaaat-3' and reverse primer 5'-tttttggtacctactttcataaagaagagc-3'. The PCR product was digested with *NcoI* and *Acc65I* (restriction sites in italics) and ligated into the corresponding sites of a pET-M11 expression vector containing a His-tag at the N-terminus. The final construct, Sgo0707N, encodes MKHHHHHHHPMSDYDIPTTENLYFQGAM-Sgo0707<sub>36–458</sub>. The plasmid was transformed into *Escherichia coli* DH5 $\alpha$  and subsequently selected on kanamycin plates. The positive clones were verified by DNA sequencing.

### Overexpression and Purification of Sgo0707N

The Sgo0707N construct was overexpressed in *E. coli* BL21 and grown in Luria Broth supplemented with 50  $\mu\text{g}/\text{ml}$  kanamycin. The culture was grown at  $37^{\circ}\text{C}$  until it reached an  $\text{OD}_{600}$  of 0.6 when the temperature was lowered to  $28^{\circ}\text{C}$  and protein expression was induced with 0.5 mM IPTG. The culture was allowed to grow for an additional 4 h. Cells were harvested by centrifugation at 5300  $g$  for 20 min and the pellets were stored at  $-80^{\circ}\text{C}$ . Pellets were resuspended in lysis buffer (20 mM Tris pH 7.5, 150 mM NaCl and 10 mM imidazole pH 8.0) supplemented with EDTA-

free protease inhibitor cocktail (Roche). The cells were lysed on ice by sonication and cellular debris was removed by centrifugation at 39000  $g$  for 35 min. The supernatant was loaded onto a Ni-NTA agarose column (Qiagen). The column was washed with lysis buffer containing 20 mM imidazole and eluted in the same buffer containing 300 mM imidazole. The protein was then concentrated and the buffer exchanged to 20 mM Tris pH 7.5 and 0.5 mM EDTA. The protein was further purified on an UNO Q6 anion exchange column (Bio-Rad) equilibrated with 20 mM Tris pH 7.5 and eluted with a gradient of the same buffer containing 1 M NaCl. Finally the protein was purified using size exclusion chromatography on a Superdex200 16/60 PG column (Amersham Biosciences) equilibrated with 20 mM Tris pH 7.5, 200 mM NaCl and 1 mM EDTA. The protein was concentrated to 84  $\text{mg}/\text{ml}$  in 20 mM Tris pH 7.5.

Selenomethionine (SeMet)-substituted Sgo0707N was expressed in *E. coli* BL21 grown in M9 medium supplemented with glucose at  $37^{\circ}\text{C}$ . At an optical density of  $\sim 0.6$  at 600 nm, lysine, threonine, phenylalanine at 100  $\text{mg}/\text{l}$ , leucine, isoleucine, valine, proline and SeMet at 50  $\text{mg}/\text{l}$  were added to down-regulate the methionine synthesis [17]. The temperature was lowered to  $20^{\circ}\text{C}$  and the expression was induced with 0.5 mM IPTG. The culture was grown overnight. The SeMet labelled protein was purified as described above with the exception that 0.5 mM Tris (2-carboxyethyl)-phosphine hydrochloride (TCEP) was present throughout all steps. The SeMet protein was concentrated to 92  $\text{mg}/\text{ml}$ .

### Construction of the *sgo0707* Deletion Mutant

A knockout mutant of *sgo0707* (strain  $\Delta$ Sgo0707) was constructed using a PCR-ligation mutagenesis strategy, as previously described [18]. Briefly, genomic DNA from *S. gordonii* DL-1 was used to amplify the upstream flanking region of *sgo0707* using P1: 5'-tcagccataccaccgtcaac-3' and P2: 5'-ggcgcgccacaaagccgaagc-caaac-3' primers, whereas the downstream flanking region was amplified using primers P3: 5'-ggcggccaccatcagccactcaatg-3' and P4: 5'-tgctgtaagacggtgg-3' (italicized bases represent *AscI* and *FseI* cut sites, respectively). After digesting with respective restriction enzymes, both amplicons were ligated with an erythromycin resistant cassette with exposed *AscI* and *FseI* cut sites. The ligated product was amplified using P1 and P4 primers, transformed into wild-type cultures, and selected for resistance to erythromycin. Successful mutagenesis was validated using PCR and nucleotide sequence analyses.

### Preparation of Cell Wall Proteins and 2DE

Mid-exponential growth phase cells were washed with PBS and centrifuged at 2000  $g$  for 10 min at  $4^{\circ}\text{C}$ . The pellet was then resuspended in 0.2% sulfobetaine (3–10) and shaken for 1 h at 100 rpm,  $28^{\circ}\text{C}$  before being centrifuged at 6000  $g$  for 10 min at

4°C. After washing three times in ultrapure water, the cells were re-suspended in spheroplasting buffer (20 mM Tris-HCl, pH 6.8, 10 mM MgCl<sub>2</sub>, 26% w/v raffinose), 100 U/ml mutanolysin (Sigma) added, and the sample incubated for 75 min at 37°C by 20 min at 60°C. Samples were placed on ice before being centrifuged at 12000 g for 20 min at 4°C. The supernatant was then dialysed against ultrapure water and freeze-dried. The resulting material was dissolved in two-dimensional polyacrylamide gel electrophoresis (2DE) rehydration buffer (8 M urea, 2% CHAPS, 10 mM DTT, 2% immobilized pharmlayte gradient buffer (GE Healthcare Life Sciences)) and stored at -20°C until subjected to 2DE and identification with liquid chromatography-tandem mass spectroscopy (LC-MS/MS) as described previously [16].

### Adherence of *S. gordonii* DL1 Wild-type or ΔSgo0707 Strain to Protein-coated Surfaces

Binding of the wild-type or mutant strain to different surface coatings was investigated in an ibiTreat μ-slide VI flow-cell model. Stimulated whole saliva, pooled from six individuals was prepared according to the method described by Palmer and co-workers [19]. 5% human serum was prepared by diluting human serum (Lonza Group Ltd) with distilled water. Human type-1 collagen (Coating matrix kit, GIBCO) was used according to the manufacturer's instructions. Flow-cells were coated overnight at room temperature with 100 μl of each protein preparation (saliva, serum, or collagen) to create conditioning films and gently rinsed with 3×100 μl m-ADM before use. Mid-exponential growth phase cells of the wild-type strain and the ΔSgo0707 strain were passed over the flow-cell surfaces for 2 h at a rate of 3.6 ml/h. The flow-cells were subsequently rinsed with m-ADM at the same flow rate for 30 min to remove non-attached cells. The LIVE/DEAD® BacLight™ Bacterial Viability Kit was used to stain the adhered cells which were then visualized with confocal laser scanning microscopy (CLSM) (Nikon Eclipse TE2000-E confocal microscope). Twenty images were collected in each experiment and experiments were repeated three times using independent bacterial cultures. Image analysis was performed using the bioImage\_L software package [20]. The mean surface coverage for the wild-type and ΔSgo0707 strains was compared using student's t-test.

To investigate the inhibitory effect of exogenous Sgo0707N recombinant protein on binding of wild-type *S. gordonii* DL1 to type-1 collagen coated surfaces, surfaces were incubated overnight at 37°C with recombinant Sgo0707N protein at three concentrations (0.01, 0.05 and 0.1 mg/ml). Wild-type cells of *S. gordonii* DL1 in M-ADM containing the same concentration of recombinant protein were then flowed over the surfaces and the adherence assessed as described above. The mean surface coverage in the presence of the three concentrations of the recombinant protein were compared with the control using ANOVA (n = 3).

### Adherence of *S. gordonii* DL1 Wild-type and ΔSgo0707 Strains to Oral Keratinocytes

Immortalized normal human keratinocytes OKF6/TERT-2 [21] were seeded into Ibidi μ-Slide VI ibiTreat flow-cell chambers (Ibidi GmbH, Germany) and allowed to grow until 30% confluence was reached as described previously [16]. The keratinocyte layer was washed with 3×100 μl m-ADM before loading 100 μl of bacterial cultures in mid-exponential growth phase into each lane. After 1 h at 37°C in 5% CO<sub>2</sub> in air, the lanes were washed with 3×100 μl PBS, pH 7.2 and stained with LIVE BacLight™ Bacterial Gram Stain Kit (Invitrogen) which stains the

Gram-positive bacteria red and the keratinocytes green. Adhered bacteria were visualized as above and twenty images were collected in each experiment. Experiments were repeated three times using independent bacterial cultures. Image analysis was performed manually by counting the number of keratinocytes and adhered bacteria in each image and calculating a ratio of bacteria/keratinocyte. For each of the three experiments, 1500–2000 keratinocytes were analysed for the wild-type or ΔSgo0707 strains respectively. The mean of adhered bacteria per thousand keratinocytes for the wild-type and ΔSgo0707 strains was compared using student's t-test.

### Auto-aggregation Assay

The wild-type and mutant strains were grown to mid-exponential growth phase in Todd Hewitt broth and the degree of auto-aggregation assessed as described previously [22]. Briefly, cultures were centrifuged (3000 g, 5 min at 4°C), washed twice with PBS and then dispersed in co-aggregation buffer (1 mM Tris pH 8, 0.15 M NaCl, 0.1 mM CaCl<sub>2</sub>, 0.1 mM MgCl<sub>2</sub> and 0.02% NaN<sub>3</sub>). The suspensions were vortexed for 10 s, adjusted to OD<sub>660</sub> = 0.5 and 0.7 ml of bacterial suspension transferred to a sterile cuvette. The optical density was recorded. The cuvettes were allowed to stand at room temperature for 1 h before the optical density was recorded again. The degree of auto-aggregation was calculated as the percentage decrease in optical density over this time.

### Glycan Array

Recombinant Sgo0707N in 25 mM Tris pH 7.5 was concentrated and lyophilized (total amount 2.1 mg). The protein was dissolved in 20 mM Tris pH 7.4, 150 mM NaCl, 2 mM CaCl<sub>2</sub>, 2 mM MgCl<sub>2</sub>, 0.05% Tween 20 and 1% BSA to a concentration of 200 μg/ml. A total of 611 glycans in replicates of six were screened and binding was detected with an anti-his antibody. The array was performed by the Consortium for Functional Glycomics, Core H (<https://www.functionalglycomics.org>).

### Thermal Shift Assay

Sgo0707N was screened for stabilizing metal-buffer conditions using the thermofluor method [23]. In short, 25 μl solutions containing 3×Sypro Orange (Molecular Probes), 100 mM Tris pH 7.5, 6 mM metal and 7.5 mg/ml protein were added to 0.1 ml PCR tubes and heated in a qPCR detection system (Rotor-Gene 6000, Corbett Life Science) from 28°C to 95°C in increments of 0.2°C. Changes in fluorescence were monitored and the melting temperature (T<sub>m</sub>) was determined by calculating the derivative of the midpoint of the protein unfolding transition. The metals that were screened were NaCl, LiCl, CaCl<sub>2</sub>, MnCl<sub>2</sub>, MgCl<sub>2</sub>, and Zn(OAc)<sub>2</sub>. The addition of water to an identical protein mix was used as a control. The experiment was performed in triplicate.

### Crystallization and Data Collection of Sgo0707N

Initial crystallization trials were performed by a Mosquito pipetting robot (TTP Labtech) using sitting-drop vapour diffusion and standard crystal screening kits (Hampton Research and Molecular Dimensions) in 96-well plate format. The initial hits were optimized using the sitting-drop method by mixing 2 μl of purified protein at a concentration of 15 mg/ml and 2 μl of reservoir solution. Crystals were initially obtained in condition A3 and G6 of the Structure Screen I/II (Molecular Dimensions). The final crystallization condition was optimized to 0.1 M sodium acetate pH 5.0, 0.2 M ammonium sulphate and 20% (w/v)

polyethylene glycol 4000. Crystals grew within one week in space group *I*222 with cell dimensions  $a=152.1$ ,  $b=158.1$  and  $c=164.3$  Å. Crystals of the SeMet labelled protein were obtained from the same conditions. Prior to data collection the crystals were soaked in crystallization solution supplemented with 20% polyethylene glycol 400, mounted on loops, and vitrified in liquid nitrogen. Diffraction data were collected at 100 K. Data from both native and SeMet containing crystals were collected at beamline ID-23 at the European Synchrotron Radiation Facility, Grenoble, France to 2.1 and 2.4 Å resolution respectively. Data were processed with XDS [24] and scaled with SCALA from the CCP4 program suit [25].

### Structure Solution and Refinement

The SeMet-containing structure was solved with SAD-phasing using AutoRickshaw [26]. Density modification and automatic model building were performed using AutoRickshaw and ArpWarp [27] and resulted in a readily interpretable map. The model was further built using rounds of manual building in COOT [28] and refined using phenix.refine [29]. The native structure was solved by molecular replacement using the program Phaser [30], with the SeMet containing structure as the search model. The native structure was similarly refined with phenix.refine. The first refinement step included rigid body refinement and simulated annealing starting at 5000 K. Manual inspection, rebuilding and addition of water molecules were performed with COOT [31]. For refinement 5% of the reflections were removed for the calculation of  $R_{\text{free}}$ . The quality of the model was analyzed with WHATCHECK [32] and Ramachandran Statistics were obtained from COOT. The model was subjected to four-fold NCS restraints throughout the refinement. Refinement statistics are given in Table 1. Figures were drawn with CCP4MG [33].

Structure factors and coordinates have been deposited in the Protein Data Bank (PDB; <http://www.rcsb.org/pdb>) under accession code 4IGB.

### Docking of type-1 Collagen

Three web-based docking programs (GRAMMx [34], Firedock [35] and Hex Server [36]) were used to dock an 11 residue long type-1 collagen fragment (derived from PDB 2F6A [37]) to the Sgo0707 crystal structure.

## Results

### $\Delta$ Sgo0707 Shows Reduced Binding to Type-1 Collagen but Not to Saliva or Serum-Coated Surfaces

To investigate the role of Sgo0707 as an adhesin, a mutant deficient in the protein was constructed in *S. gordonii* DL1 ( $\Delta$ Sgo0707). In order to confirm that Sgo0707, previously identified from the culture fluid of a *S. gordonii* DL1 sortase deficient mutant, was expressed in the cell wall of the wild-type strain, cell wall proteins were prepared and subjected to 2DE (Fig. 2). This revealed a prominent cluster of spots with  $M_r$  130–180 kDa. Using mass spectroscopy, this cluster was identified as corresponding to Sgo0707. This cluster was absent from cell wall preparations from the  $\Delta$ Sgo0707 strain, thus confirming that the mutant did not express this protein.

The wild-type and  $\Delta$ Sgo0707 strains were then compared for their binding to surface-associated proteins; type-1 collagen, saliva and serum (Fig. 3a). The wild-type strain bound well to all the coatings tested although adherence was slightly better to saliva than to type-1 collagen or serum. The mutant strain showed a 40% lower level of binding to collagen than its wild-type counterpart ( $p<0.05$ ), whereas for saliva and serum, no such

**Table 1.** Data collection, refinement and model quality statistics for Sgo0707.

	Native Sgo0707	SeMet Sgo0707
<b>Data collection</b>		
Space group	<i>I</i> 222	<i>I</i> 222
Cell dimensions a, b, c (Å)	152.13, 158.12, 164.26	152.12, 157.40, 165.26
Wavelength (Å)	0.979300	0.979300
Resolution (Å)	48.45–2.09	48.47–2.39
Highest resolution shell (Å)	2.20–2.09	2.52–2.39
Total reflections <sup>a</sup>	674,645 (57,281)	885,346 (76,170)
Unique reflections <sup>a</sup>	114,899 (15,141)	77,177 (10,121)
$I/\sigma(I)$ <sup>a</sup>	13.9 (3.9)	22.8 (5.6)
$R_{\text{sym}}(\%)$ <sup>a, b</sup>	8.0 (32.4)	8.6 (34.7)
Completeness (%) <sup>a</sup>	98.4 (89.8)	98.5 (89.8)
Overall redundancy	5.9 (3.8)	11.5 (7.5)
<b>Refinement</b>		
No. of reflections	112615	
$R_{\text{work}}/R_{\text{free}}(\%)$ <sup>c</sup>	17.9/22.5	
Average B-factors (Å <sup>2</sup> )		
Wilson plot	29.6	
Protein (Chain A/B/C/D)	27.6/28.6/30.0/31.6	
Water	35.8	
Metal ions	37.9	
Sulfate ions	48.5	
No. protein atoms	13424	
No. metal ions	4	
No. sulfate ions	8	
No. water molecules	1282	
RMSD from ideal		
Bond lengths (Å)	0.007	
Bond angles (°)	1.103	
Ramachandran plot		
Preferred, allowed, outliers (%)	93.88/5.36/0.77	

<sup>a</sup>Values in parentheses indicate statistics for the highest resolution shell.

<sup>b</sup> $R_{\text{sym}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$ , where  $I_i(hkl)$  is the intensity of the  $i$ th observation of reflection  $hkl$  and  $\langle I(hkl) \rangle$  is the average over of all observations of reflection  $hkl$ . <sup>c</sup> $R_{\text{work}} = \frac{\sum |F_{\text{obs}}| - |F_{\text{calc}}|}{\sum |F_{\text{obs}}|}$ , where  $F_{\text{obs}}$  and  $F_{\text{calc}}$  are the observed and calculated structure factor amplitudes, respectively.  $R_{\text{free}}$  is  $R_{\text{work}}$  calculated using 5% of the data, randomly omitted from refinement.

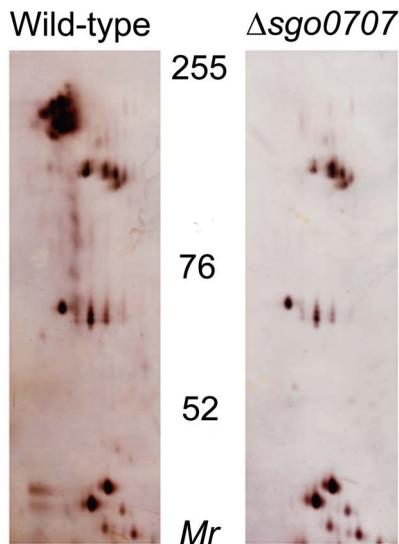
doi:10.1371/journal.pone.0063768.t001

decrease was seen (Fig. 3b) This suggests that Sgo0707 is involved in binding of *S. gordonii* to type-1 collagen but not to saliva or serum-coated surfaces.

Wild-type *S. gordonii* also bound well to oral keratinocytes (mean number of bacteria per 1000 cells =  $3602 \pm 439$ ) while binding of  $\Delta$ Sgo0707 was 30% lower (mean number of bacteria per 100 cells =  $2820 \pm 52$ ). This difference was significant ( $p<0.01$ ) (Fig. 3).

To determine whether Sgo0707 plays a role in cell-cell interactions in *S. gordonii*, the auto-aggregation of the wild-type and mutant strains was also investigated. The percentage auto-aggregation seen for the wild-type strain over 1 h was  $9 \pm 0.6\%$  while the corresponding value for the  $\Delta$ Sgo0707 strain was





**Figure 2. Silver-stained 2DE gels of cell wall proteins from wild-type and  $\Delta$ Sgo0707 strains of *S. gordonii*.** Proteins prepared as described in materials and methods were subjected to isoelectric focussing at pH 4–7 followed by SDS-PAGE in 7% gels. The migration positions of molecular mass markers are shown. doi:10.1371/journal.pone.0063768.g002

11 ± 0.9%. The difference between these values is not significant at the 5% level suggesting that Sgo0707 does not play a role in auto-aggregation.

### Binding of Sgo0707 to type-1 Collagen is Mediated via the N-terminal Domain

To test whether the Sgo0707 N-terminal domain mediates binding to type-1 collagen, recombinant Sgo0707N was overexpressed and the ability of the protein to inhibit binding was tested (Fig. 4). This showed that the protein competitively inhibited binding of wild-type *S. gordonii* in a concentration-dependent manner. However, inhibition of binding was not complete even at high concentrations of the recombinant protein.

### Carbohydrate Binding

In order to investigate whether or not Sgo0707N binds to carbohydrates, a glycan array screening was performed at the Consortium for Functional Glycomics, Core H. A total of 611 glycans (of mammalian and pathogen origin) were screened but no binding was detected for any of the sugars.

### Metal Stabilization

A thermally induced melting analysis was performed on Sgo0707N in combination with a set of mono- and divalent metals using the thermofluor method [23]. Results showed that the protein was most stable when supplemented with  $\text{Ca}^{2+}$  ions ( $T_m = 62^\circ\text{C}$ ). Other divalent ions ( $\text{Zn}^{2+}$ ,  $\text{Mg}^{2+}$  and  $\text{Mn}^{2+}$ ) resulted in melting temperatures of 55–57°C whereas monovalent ions ( $\text{Na}^+$  and  $\text{Li}^+$ ), gave the same melting temperature as the non-metal supplemented protein ( $T_m = 54^\circ\text{C}$ ). The melting temperatures are presented in Table 2.

### Overall Structure of the N-terminal Domain of Sgo0707

A construct of Sgo0707N representing residues 36–458 was expressed, purified and crystallized. The structure of *S. gordonii* Sgo0707N was solved to a resolution of 2.4 Å by single anomalous

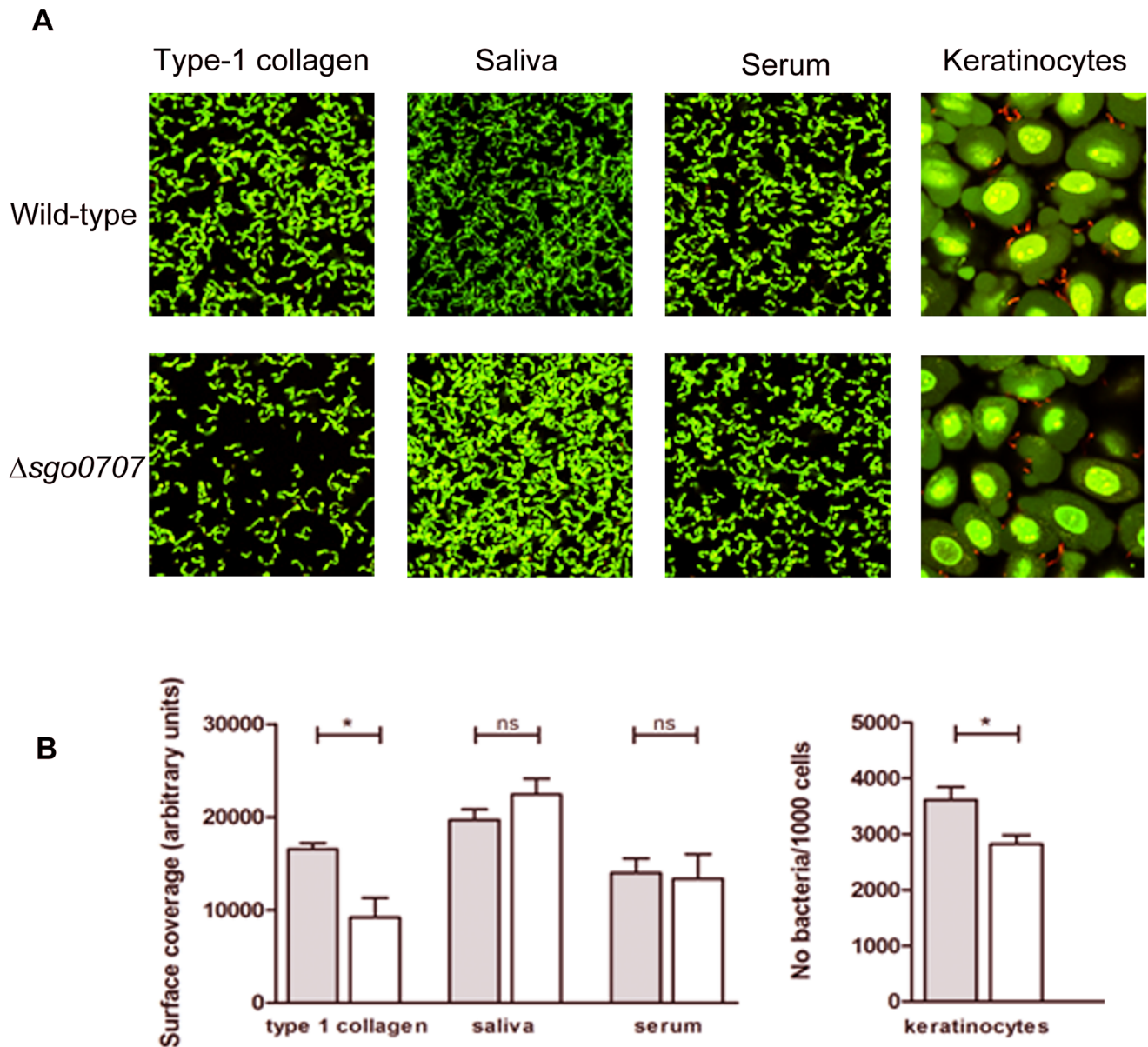
dispersion using one crystal of SeMet substituted protein belonging to the  $I222$  space group with four molecules in the asymmetric unit. The initial model was subsequently refined against native data to 2.1 Å resolution. The final model has an average B-factor of 29.5 Å<sup>2</sup>, calculated on all protein atoms, and a crystallographic R-factor of 17.9% ( $R_{\text{free}} = 22.5\%$ ). The model consists of amino acids 36–455 as well as 6–12 residues from the preceding linker. No, or weak, electron density was observed for the loop residues 427–430. The structure can be divided into two domains, N1 and N2, which both adopt  $\beta$ -sandwich folds with antiparallel  $\beta$ -sheets (Fig. 5). The N1-domain, (residues 36–311), comprises a  $\beta$ -sandwich built up from two sheets consisting of nine (S1) and eight (S2) strands respectively. The first strand of S1,  $\beta_1$ , is connected to  $\beta_3$  via a long segment that runs across the domain over to the S2 sheet. The S1 sheet is broken with four strands on one side (S1a) and five strands on the other (S1b). A small subdomain (SDA), mostly consisting of the  $\alpha$ -helix ( $\alpha_C$ ) that connects  $\beta_{22}$  with  $\beta_{23}$  and the first part of the N-terminal segment that runs across the  $\beta$ -sheet, is located over the S1a sheet. SDA also comprises a loop containing a small  $\beta$ -hairpin  $\beta_{17}/\beta_{18}$ . On top of S1b another small subdomain (SDB) is located, consisting of a loop segment connecting  $\beta_4$  and  $\beta_7$ , a coiled segment connecting  $\beta_{13}$  and  $\beta_{15}$  and a helix,  $\alpha_A$ , connecting  $\beta_{11}$  and  $\beta_{12}$ . In addition three short  $\beta$ -strands,  $\beta_5$ ,  $\beta_6$  and  $\beta_{14}$ , form a small  $\beta$ -sheet. The N2-domain (residues 312–456) consists of two  $\beta$ -sheets of five  $\beta$ -strands each. The domain also comprises a small three-stranded  $\beta$ -sheet at the corner of the  $\beta$ -sandwich. The loops connecting the  $\beta$ -strands on the side facing the N1-domain are long and one contains a short  $\alpha$ -helix ( $\alpha_D$ ). In three of the four molecules in the asymmetric unit the linker preceding the cloned construct is visible, six residues in molecule A, and 11–12 residues in molecule B and C. In B and C the linker folds as a  $\beta$ -strand ( $\beta_0$ ) that becomes the first of the S1 sheet. The extra strand,  $\beta_0$ , participates in crystal packing by linking the S1a sheets of two molecules, A with C and B with D.

### A Putative Binding Cleft in the N1-domain

An open cleft is formed between the subdomains SDA and SDB and the split between  $\beta_{16}$  and  $\beta_{25}$  in the S1-sheet (Fig. 6a). The underlying S2 sheet constitutes the floor of the cleft. Furthermore, the N-terminal segment that runs across the domain and a small anti-parallel  $\beta$ -sheet consisting of strands  $\beta_2$  and  $\beta_{26}$  form the rim of the cleft. As calculated by the CASTp server [38] the cleft has a volume of 333 Å<sup>3</sup>.

The pocket exhibits a predominantly negative surface potential (Fig. 6b) formed by Glu81, Glu169, Glu174, Glu232, Asp221, Asp223, Glu277, Glu300 together with the polar residues Thr220, Asn278 Ser274, Ser303 and Cys302. The cleft is also built up from several aromatic residues, Tyr80, Tyr82, Tyr88, Trp168, Tyr180, Tyr296, Phe306, Tyr309, mainly originating from the loops of SDB.

In all four chains positive electron density is found in the cleft but this is more pronounced in chains A and D. Although the protein was crystallized without the addition of metal ions positive electron density, indicating a partly occupied metal ion was observed in association with Asp221 (OD1), Asp223 (OD2) and Thr271 (O) of the A-chain. The electron density in the corresponding sites in chains B–D was less resolved. The putative metal is further coordinated by two water molecules and what was modeled as a sulfate ion. An additional stretch of positive electron density is observed in the binding cleft, stacked between the side chains of Tyr80, Tyr88 and Trp168. In chain A, it is unambiguous that this density originates from the final C-terminal residues 456–458 (YEK) from a symmetry related molecule. The density is however not of sufficient quality for the side chains to be modeled



**Figure 3. Role of Sgo0707 in *S. gordonii* binding to type-1 collagen, saliva, serum and oral keratinocytes.** A: Fluorescence micrographs showing adherence of wild-type and  $\Delta$ Sgo0707 strains to protein-coated surfaces and to oral keratinocytes. In panels 1–3 bacteria were flowed over surfaces coated with type-1 collagen, 25% saliva or 10% serum for 2 h, as described, and adhered bacteria stained green using the LIVE/DEAD<sup>®</sup> BacLight<sup>™</sup> Bacterial Viability stain. Images were obtained using CLSM and the surface coverage determined by image analysis using the *bioimage\_L* software package. In panel 4, bacteria were flowed over keratinocyte layers for 1 h, stained using the LIVE BacLight<sup>™</sup> Bacterial Gram Stain Kit and viewed with CLSM. The mean number of bacteria (stained red)/1000 keratinocytes (stained green) was determined by manual image analysis. B: Graphs showing the adherence of the wild-type (grey bars) and  $\Delta$ Sgo0707 (white bars) strains to different surface coatings. For type-1 collagen, saliva and serum, adherence is expressed as surface coverage (arbitrary units) whereas for the keratinocytes, binding is expressed as number of bacteria/1000 keratinocytes. The mean and standard error of three independent experiments is shown (\* indicates a significant difference  $p < 0.05$ ). doi:10.1371/journal.pone.0063768.g003

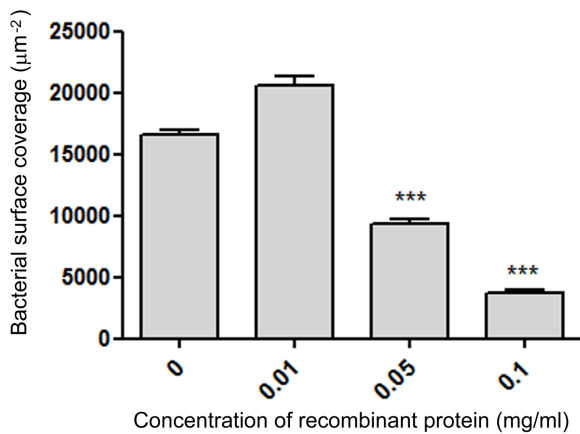
with certainty. Residual electron density is also observed in the cleft of chain C although no symmetry related protein chains are present in the close vicinity.

#### Structural Relatives of Sgo0707N

A Dali server search [39] was carried out on the N1- and N2-domains separately. Surprisingly the closest structural relatives of the N1-domain were identified as the variable domains of the AgI/II proteins SspB from *S. gordonii* (PDB code 2WD6 [40]) and SpaP from *Streptococcus mutans* (PDB code 1JMM [11,41]). Both share

approximately 10% sequence identity with Sgo0707N1 with a root mean square deviation (r.m.s.d.) of 3.7 Å and a Z-score of 12.2 and 11.0 respectively. A comparison between the Sgo0707 N1-domain and the variable domain of SspB is shown in Fig. 7.

The structural similarities include the size of the putative binding cleft, 333 and 418 Å<sup>3</sup> respectively for Sgo0707N1 and SspB, whereas the cleft in SpaP is larger, 2280 Å<sup>3</sup>. However, whereas SspB and SpaP have a tightly bound metal ion in the cleft, Sgo0707N1 does not. In addition, the cleft of Sgo0707N1 has a more predominant negative charge than the two AgI/II proteins.



**Figure 4. Inhibition of binding of *S. gordonii* DL1 to type-1 collagen coated surfaces using recombinant Sgo0707<sub>36–455</sub> protein.** Flow-cell surfaces were coated with type-1 collagen followed by different concentrations of recombinant Sgo0707<sub>36–455</sub> protein. Bacteria were flowed over the surfaces for 2 h in the presence of the same concentrations of recombinant Sgo0707<sub>36–455</sub> and adhered bacteria were visualized using the BacLight Live/Dead stain and analysed with *bioImage\_L* software package. The mean and standard error of three independent experiments is shown (\*\*\*) indicates a significant difference  $p < 0.001$  from control. doi:10.1371/journal.pone.0063768.g004

The N2-domain consists mainly of two  $\beta$ -sheets of five  $\beta$ -strands each forming an IgG-like domain of CnaA fold. The Dali server search of the N2-domain recognized the structure of the middle domain of the *Bacillus cereus* BcpA pilin (PDB code 3KPT [42]) as the closest structural relative with a Z-score of 16.6, 15% sequence identity of 131 aligned C $\alpha$  atoms and an r.m.s.d. of 2.3 Å. In general all hits in the Dali search represent IgG-like modules that build up surface proteins on Gram-positive bacteria.

Interestingly, many structures in this group are stabilized by intra-molecular isopeptide bonds [10] between the side chains of a lysine and an asparagine or an aspartic acid. Such bond formation is facilitated by a close aspartic or glutamic acid and is believed to protect the protein from mechanical and chemical stress. No isopeptide bonds are however observed in the Sgo0707 N2-domain and the residues in the equivalent lysine-asparagine positions are Leu319 and Asn434. The position where a catalytic acid would be expected is occupied by Gln360. Thus the N2-domain lacks the prerequisites for isopeptide bond formation.

### Collagen Docking

Three web-based docking programs (GRAMMx, HexServer and Firedock) were used to dock a type-1 collagen fragment to the Sgo0707N crystal structure. The programs were consistent in predicting two areas of the protein as putative interaction sites. Firstly, the groove in the N1-domain, discussed above as a putative binding cleft, and secondly the concave surface formed by the S4 sheet of the N2-domain and loops from the N1-domain (Fig. 8). By superimposing the collagen binding proteins ACE19 of *Enterococcus faecalis* [43] and CNA of *Staphylococcus aureus* [37] onto Sgo0707N it could be shown that the collagen triple helix was docked onto the opposite side of the  $\beta$ -sandwich in Sgo0707N compared to that observed for the collagen binding proteins.

### Discussion

In a previous study, using a  $\Delta$ SrtA mutant, we identified a novel LPXTG-linked protein, Sgo0707, from the growth medium [16].

**Table 2. Protein melting temperatures when supplemented with metal ion solutions.**

Additive (6 mM)	T <sub>m</sub> (°C)
No additive (H <sub>2</sub> O)	54.4
Ca <sup>2+</sup>	62.4
Zn <sup>2+</sup>	57.3
Mn <sup>2+</sup>	56.6
Mg <sup>2+</sup>	55.8
Na <sup>+</sup>	54.0
Li <sup>+</sup>	53.7

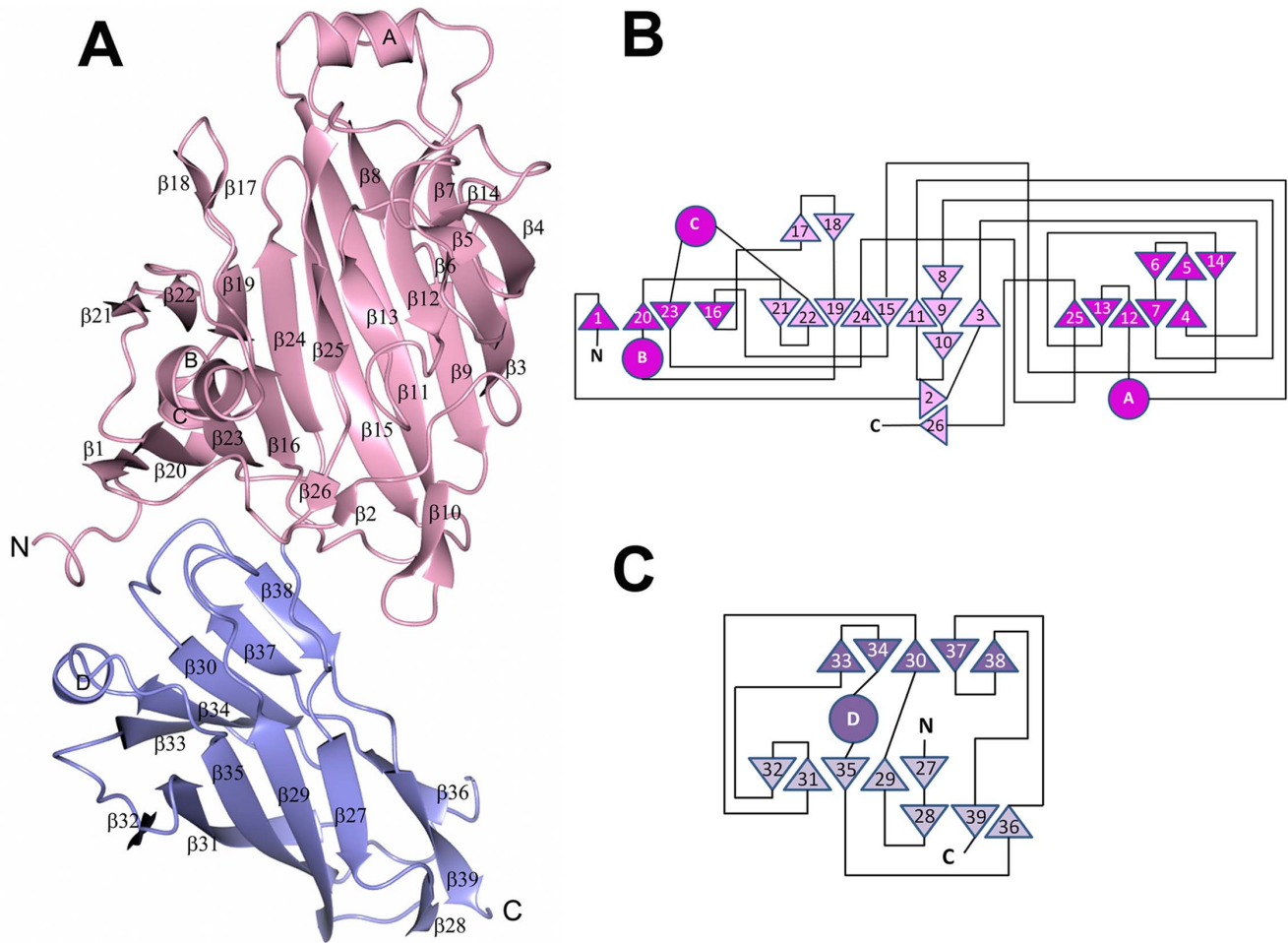
doi:10.1371/journal.pone.0063768.t002

The Sgo0707 protein consists of 1643 amino acids including a leader peptide, an N-terminal region, eight 84-amino acid repeat domains, five 88-amino acid repeat domains and a C-terminal domain containing the LPXTG sorting motif. The domain organization is similar to that of the fibrillar protein CshA from *S. gordonii* [13]. In a study by McNab and coworkers, CshA was shown to be involved in bacterial binding to fibronectin [44], an interaction which could be blocked using antibodies recognizing the N-terminal domain (amino acids 42–886) of the protein. Based on their related domain organization, our hypothesis was that the N-terminal region of Sgo0707 also represents an adhesive domain. Therefore we chose to investigate the structural and functional characteristics of the N-terminus of Sgo0707.

*S. gordonii* is known to express a number of multi-ligand adhesins, covalently bound to the cell wall. In this study, through selective extraction of cell wall proteins, Sgo0707 was revealed as an abundant component of the cell wall proteome. Therefore we investigated the binding of Sgo0707 to a range of proteins found in the oral cavity. For saliva- or serum-coated surfaces, no significant difference was seen between the wild-type strain and an isogenic  $\Delta$ Sgo0707 mutant. Similarly, no differences in auto-aggregation were seen. Interestingly, binding to type-1 collagen was significantly reduced in the mutant strain. These data suggest that Sgo0707 plays a role in initial adherence of *S. gordonii* to type-1 collagen. Interactions between the wild-type strain and collagen-coated surfaces could be competitively inhibited using a recombinant polypeptide encompassing residues 36–458 of Sgo0707, suggesting that binding is mediated via the N-terminal region. However, binding was not completely abolished even at high concentrations of the recombinant protein confirming that other cell surface proteins of *S. gordonii* also bind collagen. This is in agreement with previous studies showing that SspA and SspB, members of the AgI/II family [45], bind type-1 collagen [46]. Thus, as proposed previously, *S. gordonii* shows redundancy with respect to cell surface adhesins.

The crystal structure of amino acids 36–455 from the N-terminal of Sgo0707 revealed that it was divided into two domains, N1 and N2, that mainly consist of  $\beta$ -strands. The N1-domain is the largest domain (residues 36–311) harbouring a putative binding cleft which strengthens the original hypothesis of the N-terminal, non-repetitive domain, functioning as an adhesin. To our surprise we found that the N1-domain, despite very low sequence similarity, exhibits a striking structural similarity to the variable domains of the AgI/II adhesins, also expressed by oral streptococci. Despite having fundamentally different domain organizations; the AgI/II variable domain is located in the centre of the protein but presented at the tip due to intertwining flanking regions and the formation of a fibrillar structure [11], and



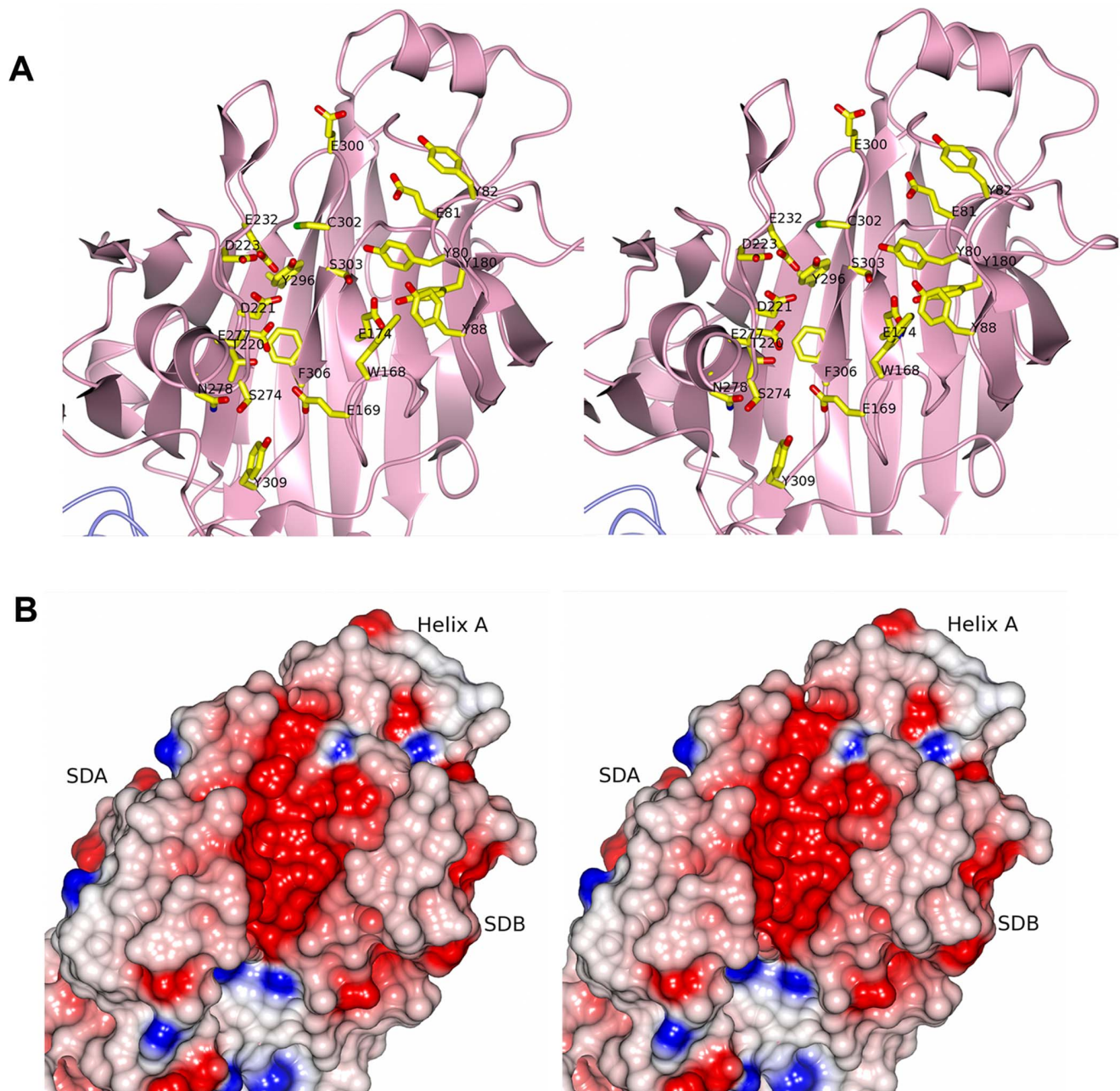


**Figure 5. The overall structure of Sgo0707N.** A: Ribbon representation of Sgo0707N. The N1-domain is coloured pink and the N2-domain purple. B: Topology diagram of Sgo0707N1. C: Topology diagram of Sgo0707N2.  $\beta$ -strands are depicted as triangles and  $\alpha$ -helices as circles. The N1- and N2-domains are shown in pink and purple respectively. doi:10.1371/journal.pone.0063768.g005

Sgo0707N1 being expressed as the N-terminal domain followed by a repetition of C-terminal domains, they both obtain similar folds, including putative binding clefts. This points to an interesting divergent evolution, possibly due to gene duplication of a domain hitherto only found in oral streptococci. The AgI/II adhesins are the best studied adhesins expressed by oral bacteria and several interaction partners are known, such as gp340, type-1 collagen, integrin and the Mfa1 fimbria from *Porphyromonas gingivalis* [47], however a ligand specific for the variable domain has never been described. Despite some similarities the binding pockets of the AgI/II proteins (SspB and SpaP) and Sgo0707 also show significant differences. Whereas the Sgo0707 cleft is predominantly negatively charged the AgI/II clefts are less charged and have a tightly bound metal ion. Thermal melting analysis showed that Sgo0707 is indeed stabilized by  $\text{Ca}^{2+}$ -ions, as is SspB [40] and electron density also indicates the presence of metal ions also in the Sgo0707 cleft. However, the electron density is not of such quality that it can be concluded that a true metal binding site is present. The AgI/II variable domains have been proposed to be carbohydrate binding [41] but in a previous study we performed a glycan array screen for the variable domain of the AgI/II adhesin SspB which showed that the domain most likely does not interact with carbohydrates [40]. A similar screen with Sgo0707N

gave the same negative result. However, these screens do not rule out that the adhesins may interact with carbohydrates not included in the array or with glycoproteins. In addition, the two forms of adhesins are related with regards to collagen binding. AgI/II shows affinity for type-1 collagen, but the interaction is not described in molecular detail. Similarly, Sgo0707 interacts with type-1 collagen but the interaction surface is not obvious. A docking of collagen to Sgo0707N yielded two different suggestions. The first alternative presented the collagen triple helix bound along the putative binding cleft in the N1-domain and as a second alternative collagen was placed at the concave side of the S4-sheet of the N2-domain, at the interface between the N1- and N2-domains. By looking at the electrostatic surfaces of the two sites it is obvious that the characteristics of the two suggested sites are not consistent. Whereas the cleft in the N1-domain is predominantly negatively charged, the concave side of the S4-sheet in the N2-domain is mostly non-polar and shows more resemblance to the binding surfaces of the collagen-binding proteins CNA and ACE. However in those proteins, the collagen triple-helix is bound on the opposite side of the  $\beta$ -sandwich (corresponding to the S3 sheet of Sgo0707) and is embraced by a linker that connects two domains. In Sgo0707, collagen is modelled to bind to the S4 sheet and interacts with loops and turns from the N1-domain. This





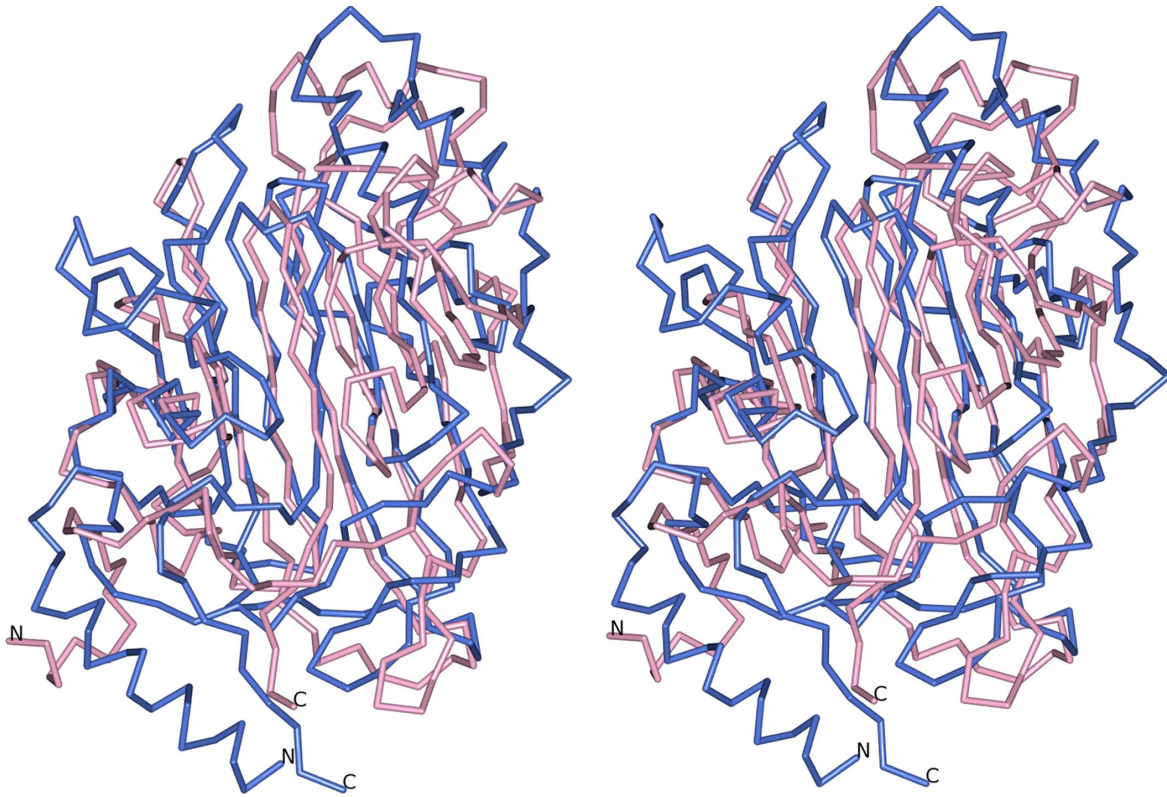
**Figure 6. The putative binding cleft in the N1-domain.** A: Stereo representation of the cleft in the N1-domain. Amino acids in the pocket are represented as cylinders and are labeled. B: The N1-domain is presented as an electrostatic surface, colored in red and blue according to negative and positive electrostatic potential, respectively. The cleft has considerable negative electrostatic potential. The figure is shown in stereo. doi:10.1371/journal.pone.0063768.g006

putative interaction cannot be as tight as the “collagen hug” model in the true collagen-binding proteins. Another mechanism for collagen adhesion, observed both in eukaryotes [48] and the *Streptococcus pneumoniae* pilus adhesin RrgA [49], involves the interaction between collagen and a MIDAS motif coordinated metal ion. However, neither a high occupancy metal ion, nor a MIDAS motif is found in Sgo0707.

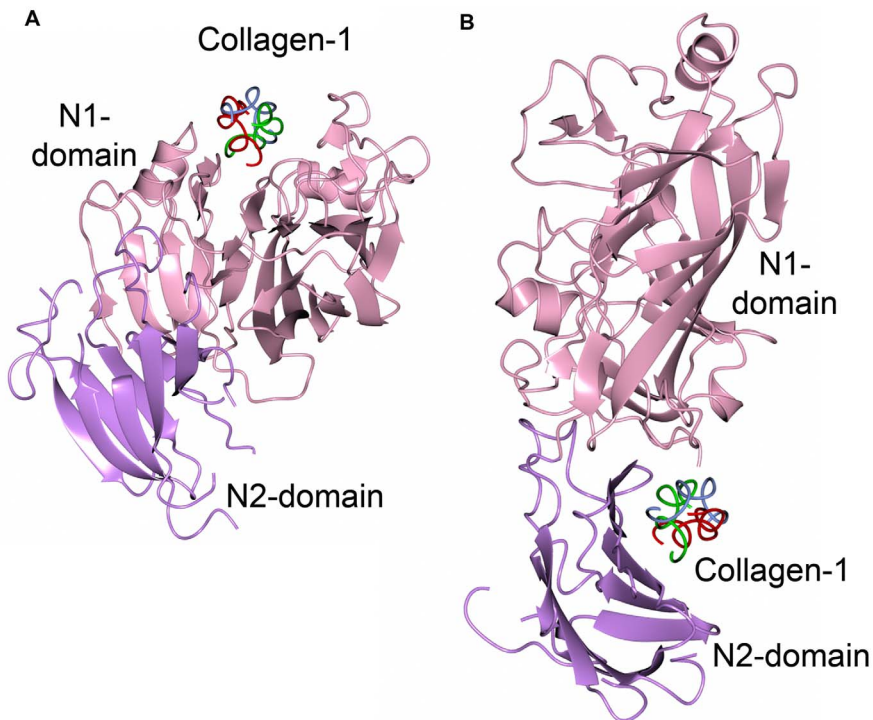
Binding to collagen is generally not an interaction that is necessary for normal colonization of the oral cavity, however it may be an important invasion mechanism at non-oral sites. The oral streptococcal adhesin most studied with regards to pathogenicity is GspB. Its binding to the platelet membrane glycoprotein

GPIIb $\alpha$  is implied to be an important factor for causing infective endocarditis [6]. Similarly the collagen binding properties of oral streptococci are considered important virulence factors for infective endocarditis [50]. Interestingly the collagen-binding protein Cnm, expressed by certain serotypes of the major pathogen of dental caries, *S. mutans*, was recently identified as an important factor in the onset of hemorrhagic stroke [51]. Although collagen may not be the main ligand for Sgo0707, its ability to interact (non-specifically) with exposed collagen tissue may be of importance when attaching to non-oral surfaces.

We also show that Sgo0707 makes significant contributions in binding to oral keratinocytes, the deletion mutant showed a 30%



**Figure 7. Superposition of Sgo0707N1 and the V-domain of SspB.** The N1-domain is superimposed on the variable domain of SspB from *S. gordonii*. Sgo0707 is depicted in pink and SspB in blue. The figure is shown in stereo.  
doi:10.1371/journal.pone.0063768.g007



**Figure 8. Sgo0707 collagen docking.** A model of collagen docked to the Sgo0707 crystal structure using FireDock. A: The collagen triple helix is docked into the binding pocket of the N1-domain. B: Collagen is docked in the groove formed by the interface between the N1- and N2-domains.  
doi:10.1371/journal.pone.0063768.g008

decrease in binding compared to wild-type. The particular molecule that is recognized on the keratinocyte cell surface cannot be determined at this point, however these cells are known to express a wide variety of surface molecules, for example integrins, cadherins, selectins, and members from the immunoglobulin superfamily [52].

Interestingly, the 1643 amino acid long protein contains only one cysteine, Cys302, and that single residue is located in the floor of the putative binding cleft in the N1-domain. In the *S. pyogenes* pilus adhesin Cpa [53] there is also a single cysteine in a putative binding cleft. In the crystal structure of Cpa, the cysteine forms an intra thioester bond with a glutamine in the same cleft, priming for a reaction with a host cell ligand and a subsequent covalent linkage. The authors show that the involved residues indeed are important for the attachment to host cells. The position of the cysteine in Sgo0707N1 similarly implies a crucial function where the residue may be involved in interacting with host cell ligands or to surface molecules of other bacteria in the oral biofilm. However, no internal thioester is formed in Sgo0707N1 and further studies are needed to examine if the cysteine has a role in ligand recognition.

In conclusion, the oral biofilm is known to harbor hundreds of different bacterial strains and oral streptococci, such as *S. gordonii*, are among the most abundant. Due to the complex environment oral streptococci can be expected to have multiple co-aggregation partners and hosts. The possibility of Sgo0707 having a palette of ligands and multiple binding activities is therefore to be expected.

## References

- Nyvad B, Kilian M (1990) Comparison of the initial Streptococcal microflora on dental enamel in caries-active and in caries-inactive individuals. *Caries Res* 24: 267–272.
- Siqueira WL, Zhang W, Helmerhorst EJ, Gygi SP, Oppenheim FG (2007) Identification of protein components in *in vivo* human acquired enamel pellicle using LC-ESI-MS/MS. *J Proteome Res* 6: 2152–2160.
- Demuth DR, Duan Y, Brooks W, Holmes AR, McNab R, et al. (1996) Tandem genes encode cell-surface polypeptides SspA and SspB which mediate adhesion of the oral bacterium *Streptococcus gordonii* to human and bacterial receptors. *Molecular Microbiology* 20: 403–413.
- Jakubovics NS, Stromberg N, van Dolleweerd CJ, Kelly CG, Jenkinson HF (2005) Differential binding specificities of oral streptococcal antigen I/II family adhesins for human or bacterial ligands. *Molecular Microbiology* 55: 1591–1605.
- Kesimer M, Kilic N, Mehrotra R, Thornton DJ, Sheehan JK (2009) Identification of salivary mucin MUC7 binding proteins from *Streptococcus gordonii*. *BMC Microbiol* 9: 163.
- Takamatsu D, Bensing BA, Cheng H, Jarvis GA, Siboo IR, et al. (2005) Binding of the *Streptococcus gordonii* surface glycoproteins GspB and Hsa to specific carbohydrate structures on platelet membrane glycoprotein Ibalph. *Molecular Microbiology* 58: 380–392.
- Jakubovics NS, Kerrigan SW, Nobbs AH, Stromberg N, van Dolleweerd CJ, et al. (2005) Functions of cell surface-anchored antigen I/II family and Hsa polypeptides in interactions of *Streptococcus gordonii* with host receptors. *Infection & Immunity* 73: 6629–6638.
- Bensing 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 Ibalph. *Infect Immun* 72: 6528–6537.
- Takahashi Y, Takashima E, Shimazu K, Yagishita H, Aoba T, et al. (2006) Contribution of sialic acid-binding adhesin to pathogenesis of experimental endocarditis caused by *Streptococcus gordonii* DL1. *Infect Immun* 74: 740–743.
- Kang HJ, Baker EN (2012) Structure and assembly of Gram-positive bacterial pili: unique covalent polymers. *Curr Opin Struct Biol* 22: 200–207.
- Larson MR, Rajashankar KR, Patel MH, Robinette RA, Crowley PJ, et al. (2010) Elongated fibrillar structure of a streptococcal adhesin assembled by the high-affinity association of alpha- and PPII-helices. *Proc Natl Acad Sci U S A* 107: 5983–5988.
- Pyburn TM, Bensing BA, Xiong YQ, Melancon BJ, Tomasiak TM, et al. (2011) A structural model for binding of the serine-rich repeat adhesin GspB to host carbohydrate receptors. *PLoS Pathog* 7: e1002112.
- McNab R, Forbes H, Handley PS, Loach DM, Tannock GW, et al. (1999) Cell wall-anchored CshA polypeptide (259 kilodaltons) in *Streptococcus gordonii* forms surface fibrils that confer hydrophobic and adhesive properties. *J Bacteriol* 181: 3087–3095.
- Ramboarina S, Garnett JA, Zhou M, Li Y, Peng Z, et al. (2010) Structural insights into serine-rich fimbriae from Gram-positive bacteria. *J Biol Chem* 285: 32446–32457.
- Linke C, Siemens N, Oehmcke S, Radjainia M, Law RH, et al. (2012) The extracellular protein factor Epf from streptococcus pyogenes is a cell-surface adhesin that binds to cells through an N-terminal domain containing a carbohydrate-binding module. *J Biol Chem*.
- Davies JR, Svensater G, Herzberg MC (2009) Identification of novel LPXTG-linked surface proteins from *Streptococcus gordonii*. *Microbiology* 155: 1977–1988.
- Van Duyn GD, Standaert RF, Karplus PA, Schreiber SL, Clardy J (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *Journal of Molecular Biology* 229: 105–124.
- Lau PC, Sung CK, Lee JH, Morrison DA, Cvitkovich DG (2002) PCR ligation mutagenesis in transformable streptococci: application and efficiency. *J Microbiol Methods* 49: 193–205.
- Palmer RJ, Jr., Kazmerzak K, Hansen MC, Kolenbrander PE (2001) Mutualism versus independence: strategies of mixed-species oral biofilms *in vitro* using saliva as the sole nutrient source. *Infect Immun* 69: 5794–5804.
- Chavez de Paz LE (2009) Image analysis software based on color segmentation for characterization of viability and physiological activity of biofilms. *Appl Environ Microbiol* 75: 1734–1739.
- Dickson MA, Hahn WC, Ino Y, Ronfard V, Wu JY, et al. (2000) Human keratinocytes that express hTERT and also bypass a p16(INK4a)-enforced mechanism that limits life span become immortal yet retain normal growth and differentiation characteristics. *Mol Cell Biol* 20: 1436–1447.
- Shen S, Samaranyake LP, Yip HK (2005) Coaggregation profiles of the microflora from root surface caries lesions. *Arch Oral Biol* 50: 23–32.
- Ericsson UB, Hallberg BM, Detitta GT, Dekker N, Nordlund P (2006) Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Analytical Biochemistry* 357: 289–298.
- Kabsch W (2010) Xds. *Acta Crystallogr D Biol Crystallogr* 66: 125–132.
- Collaborative Computational Project N (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50: 760–763.
- Panjikar S, Parthasarathy V, Lamzin VS, Weiss MS, Tucker PA (2005) Auto-Rickshaw: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Crystallographica Section D-Biological Crystallography* 61: 449–457.
- Langer G, Cohen SX, Lamzin VS, Perrakis A (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nature Protocols* 3: 1171–1179.
- Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66: 486–501.

We suggest that Sgo0707 may have a specific set of binding partners that is recognized via the binding pocket in the N1-domain but that it will be a considerable challenge to identify these ligands. We also suggest that the protein interacts non-specifically with other ligands such as the interaction with type-1 collagen that is described here. The ability to bind to collagen may however be very important for the bacteria when located at non-oral sites such as the heart valves. In addition the Sgo0707N crystal structure that is presented here constitutes only a fraction of the total protein. The C-terminal repeats that are believed to build up the stalk of the protein are not yet characterized but may be equally important for the function of the protein.

## Acknowledgments

We are grateful for access to the beamlines ID23-1, ESRF, Grenoble and for the support of the beamline staff. We acknowledge The Aberdeen Proteome Facility and the University of Aberdeen for help with protein identification using LC-MS/MS and we also thank Dr James Reinwald (Brigham and Women's Hospital, Boston, USA) and Gunter Stier (EMBL, Germany) for the kind gifts of OKF6/TERT-2 cells and the expression vector, respectively.

## Author Contributions

Conceived and designed the experiments: GS DBS DGC JRD KP. Performed the experiments: AN KP DBS JRD. Analyzed the data: AN KP DBS JRD GS DGC. Contributed reagents/materials/analysis tools: KP JRD DGC. Wrote the paper: KP JRD GS DBS AN.



29. Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, et al. (2012) Towards automated crystallographic structure refinement with phenix-refine. *Acta Crystallogr D Biol Crystallogr* 68: 352–367.
30. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, et al. (2007) Phaser crystallographic software. *J Appl Crystallogr* 40: 658–674.
31. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60: 2126–2132.
32. Hoofstede RW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. *Nature* 381: 272.
33. Potterton L, McNicholas S, Krissinel E, Gruber J, Cowtan K, et al. (2004) Developments in the CCP4 molecular-graphics project. *Acta Crystallogr D Biol Crystallogr* 60: 2288–2294.
34. Tovchigrechko A, Vakser IA (2006) GRAMM-X public web server for protein-protein docking. *Nucleic Acids Research* 34: W310–314.
35. Mashich E, Schneidman-Duhovny D, Andrusier N, Nussinov R, Wolfson HJ (2008) FireDock: a web server for fast interaction refinement in molecular docking. *Nucleic Acids Research* 36: W229–232.
36. Macindoe G, Mavridis L, Venkatraman V, Devignes MD, Ritchie DW (2010) HexServer: an FFT-based protein docking server powered by graphics processors. *Nucleic Acids Research* 38: W445–449.
37. Zong Y, Xu Y, Liang X, Keene DR, Hook A, et al. (2005) A ‘Collagen Hug’ model for *Staphylococcus aureus* CNA binding to collagen. *EMBO J* 24: 4224–4236.
38. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, et al. (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res* 34: W116–118.
39. Holm L, Kaariainen S, Wilton C, Plewczynski D (2006) Using Dali for structural comparison of proteins. *Current Protocols in Bioinformatics* Chapter 5: Unit 5.5.
40. Forsgren N, Lamont RJ, Persson K (2009) Crystal structure of the variable domain of the *Streptococcus gordonii* surface protein SspB. *Protein Sci*.
41. Troffer-Charlier N, Ogier J, Moras D, Cavarelli J (2002) Crystal structure of the V-region of *Streptococcus mutans* antigen I/II at 2.4 Å resolution suggests a sugar preformed binding site. *Journal of Molecular Biology* 318: 179–188.
42. Budzik JM, Poor CB, Faull KF, Whitelegge JP, He C, et al. (2009) Intramolecular amide bonds stabilize pili on the surface of bacilli. *Proc Natl Acad Sci U S A* 106: 19992–19997.
43. Ponnuraj K, Narayana SV (2007) Crystal structure of ACE19, the collagen binding subdomain of *Enterococcus faecalis* surface protein ACE. *Proteins* 69: 199–203.
44. McNab R, Holmes AR, Clarke JM, Tannock GW, Jenkinson HF (1996) Cell surface polypeptide CshA mediates binding of *Streptococcus gordonii* to other oral bacteria and to immobilized fibronectin. *Infection & Immunity* 64: 4204–4210.
45. Jenkinson HF, Demuth DR (1997) Structure, function and immunogenicity of streptococcal antigen I/II polypeptides. *Molecular Microbiology* 23: 183–190.
46. Heddle C, Nobbs AH, Jakubovics NS, Gal M, Mansell JP, et al. (2003) Host collagen signal induces antigen I/II adhesin and invasin gene expression in oral *Streptococcus gordonii*. *Molecular Microbiology* 50: 597–607.
47. Brady LJ, Maddocks SE, Larson MR, Forsgren N, Persson K, et al. (2010) The changing faces of *Streptococcus* antigen I/II polypeptide family adhesins. *Mol Microbiol* 77: 276–286.
48. Emsley J, Knight CG, Farndale RW, Barnes MJ, Liddington RC (2000) Structural basis of collagen recognition by integrin  $\alpha 2\beta 1$ . *Cell* 101: 47–56.
49. Izore T, Contreras-Martel C, El Mortaji L, Manzano C, Terrasse R, et al. (2010) Structural basis of host cell recognition by the pilus adhesin from *Streptococcus pneumoniae*. *Structure* 18: 106–115.
50. Moreillon P, Que YA (2004) Infective endocarditis. *Lancet* 363: 139–149.
51. Nakano K, Hokamura K, Taniguchi N, Wada K, Kudo C, et al. (2011) The collagen-binding protein of *Streptococcus mutans* is involved in haemorrhagic stroke. *Nat Commun* 2: 485.
52. Thomas GJ, Speight PM (2001) Cell adhesion molecules and oral cancer. *Crit Rev Oral Biol Med* 12: 479–498.
53. Pointon JA, Smith WD, Saalbach G, Crow A, Kehoe MA, et al. (2010) A highly unusual thioester bond in a pilus adhesin is required for efficient host cell interaction. *J Biol Chem* 285: 33858–33866.