Staphylococcal biofilm-forming protein has a contiguous rod-like structure

Dominika T. Gruszka[a], Justyna A. Wojdyla[a], Richard J. Bingham[b], Johan P. Turkenburg[c], Iain W. Manfield[d], Annette Steward[e], Andrew P. Leech[a], Joan A. Geoghegan[f], Timothy J. Foster[f], Jane Clarke[e], and Jennifer R. Potts[a,c,1]

[a]Department of Biology, University of York, York YO10 5DD, United Kingdom; [b]Department of Chemical and Biological Sciences, University of Huddersfield, Huddersfield HD1 3DH, United Kingdom; [c]Department of Chemistry, University of York, York YO10 5DD, United Kingdom; [d]Astbury Centre for Structural Molecular Biology, Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, United Kingdom; [e]Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom; and [f]Microbiology Department, Moyne Institute of Preventive Medicine, Trinity College, Dublin 2, Ireland

Edited by Roberto Kolter, Harvard Medical School, Boston, MA, and accepted by the Editorial Board March 1, 2012 (received for review November 25, 2011)

Staphylococcus aureus **and** ***Staphylococcus epidermidis*** **form communities (called biofilms) on inserted medical devices, leading to infections that affect many millions of patients worldwide and cause substantial morbidity and mortality. As biofilms are resistant to antibiotics, device removal is often required to resolve the infection. Thus, there is a need for new therapeutic strategies and molecular data that might assist their development. Surface proteins *S. aureus* surface protein G (SasG) and accumulation-associated protein (*S. epidermidis*) promote biofilm formation through their "B" regions. B regions contain tandemly arrayed G5 domains interspersed with approximately 50 residue sequences (herein called E) and have been proposed to mediate intercellular accumulation through Zn^{2+}-mediated homodimerization. Although E regions are predicted to be unstructured, SasG and accumulation-associated protein form extended fibrils on the bacterial surface. Here we report structures of E–G5 and G5–E–G5 from SasG and biophysical characteristics of single and multidomain fragments. E sequences fold cooperatively and form interlocking interfaces with G5 domains in a head-to-tail fashion, resulting in a contiguous, elongated, monomeric structure. E and G5 domains lack a compact hydrophobic core, and yet G5 domain and multidomain constructs have thermodynamic stabilities only slightly lower than globular proteins of similar size. Zn^{2+} does not cause SasG domains to form dimers. The work reveals a paradigm for formation of fibrils on the 100-nm scale and suggests that biofilm accumulation occurs through a mechanism distinct from the "zinc zipper." Finally, formation of two domains by each repeat (as in SasG) might reduce misfolding in proteins when the tandem arrangement of highly similar sequences is advantageous.**

device infection | protein biophysics | X-ray crystallography | protein domains

Health care-associated infections affect many millions of patients every year worldwide, causing substantial morbidity and mortality and high costs to health services (1). Health care-associated infections are a particular problem in adult (2) and neonatal intensive care units (3), and frequently arise as a result of formation of biofilms on the surfaces of indwelling medical devices (4). *Staphylococcus aureus* and *Staphylococcus epidermidis* are a common cause of such infections (2). Device-related staphylococcal infections are difficult to eradicate and treat clinically (5) because bacteria in the biofilm are protected from antimicrobial agents and the host immune system (6). Prolonged antibiotic therapy and device removal can be required to resolve the infection (5).

A biofilm is a functional multilayered community of microorganisms, adhering to a surface and organized within a self-produced exopolymeric matrix (4). Initiation of biofilm formation is a two-step process involving attachment (in which bacteria adhere to a surface) and subsequent maturation (when a 3D structure evolves). The maturation phase requires intercellular aggregation, during which bacteria divide and accumulate (6). Staphylococci can mediate cell-to-cell adhesion using two types of exopolymers: the polysaccharide intercellular adhesin and proteins (7). Polysaccharide intercellular adhesin is also known as poly-*N*-acetyl-glucosamine and is synthesized by enzymes encoded on the *icaADBC* operon (7). Accumulation can occur independently of *ica*, instead relying on the expression of surface proteins such as accumulation-associated protein (Aap) (8) in *S. epidermidis* and *S. aureus* surface protein G (SasG) (9) in *S. aureus*.

Aap and SasG have a domain organization that is typical of the LPXTG protein family, comprising an N-terminal secretion signal and a C-terminal sorting peptide, which is essential for covalent linkage to wall peptidoglycan (10) (Fig. 1*A*). They contain an N-terminal A domain, followed by a stretch of tandemly arrayed "B" repeats. The A domains of SasG and Aap have been implicated in adhesion of bacteria to desquamated epithelial cells (10) and B repeat regions are responsible for cell-to-cell accumulation during biofilm formation (11, 12). Although annotations in the literature differ (13, 14), the Pfam database (15) recognizes a B-repeat sequence as containing a G5 domain (~80 residues) (14) followed by an approximately 50-residue sequence (herein called E; Fig. 1*A*). The formation of Aap-mediated biofilm requires proteolytic cleavage of the N-terminal A domain (11). A similar processing pathway was proposed for SasG (9). However, recent findings indicated that the full-length protein exposed on the bacterial surface undergoes limited processing within B repeats (12). The current model of protein-mediated intercellular adhesion promoted by Aap and SasG in staphylococci is known as a zinc zipper, in which Zn^{2+}-mediated self-association events occur between stretches of B repeats on opposing Aap or SasG molecules (12, 13).

SasG and Aap have sequence identities of approximately 34% and approximately 50% for G5 and E repeats, respectively, and contain a variable number of repeats [three to 10 in SasG (16) and four to 17 in Aap (17)], dependent on the strain. High DNA sequence identity results in pairwise protein sequence similarity between B repeats within SasG (Fig. S1*A*) and within Aap of 90% to 100% and 82% to 91%, respectively. It has been shown recently that tandemly arrayed domains with high sequence identity are prone to misfolding events (18). This is likely to be a particular problem for long-lived proteins, or those which un-

Author contributions: D.T.G., R.J.B., J.A.G., T.J.F., J.C., and J.R.P. designed research; D.T.G., J.A.W., I.W.M., A.S., and A.P.L. performed research; D.T.G., J.A.W., R.J.B., J.P.T., A.P.L., and J.R.P. analyzed data; and D.T.G., J.C., and J.R.P. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. R.K. is a guest editor invited by the Editorial Board.

Freely available online through the PNAS open access option.

Data deposition: The atomic coordinates and structure factors reported in this paper have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 3TIP and 3TIQ).

[1]To whom correspondence should be addressed. E-mail: jennifer.potts@york.ac.uk.

See Author Summary on page 6370 (volume 109, number 17).

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1119456109/-/DCSupplemental.

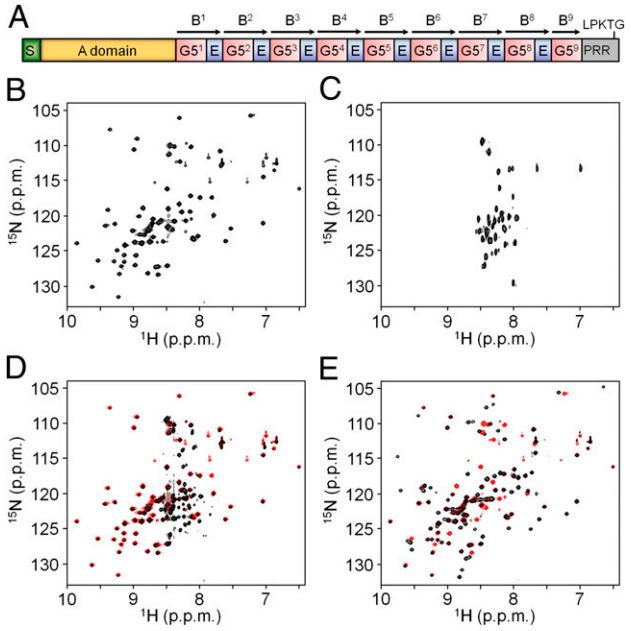

Fig. 1. SasG from *S. aureus* NCTC 8325 (UniProt no. Q2G2B2). (*A*) Schematic representation of the domain arrangement showing nine 78-residue G5 domains and eight 50-residue E segments. The location of the signal sequence (S), A domain, proline-rich region (PRR), and sorting peptide (LPKTG) are also presented. B repeats [according to the previous annotation (16)] are indicated with arrows. HSQC spectra (^{1}H-^{15}N) of G5^2 (*B*) and E (*C*). (*D*) Superimposed ^{1}H-^{15}N HSQC spectra of G5^2-E (black) and G5^2 (red). (*E*) Superimposed ^{1}H-^{15}N HSQC spectra of E-G5^2 (black) and G5^2 (red).

dergo shear stresses, and might explain the apparent evolutionary pressure for sequences of adjacent domains to have less than 40% sequence identity (19). Thus, the SasG and Aap biofilm-forming region also raises the question of how misfolding is avoided when sequence identity between repeats is otherwise advantageous.

The minimum number of SasG B repeats required for biofilm formation is five. SasG variants with more than four repeats blocked the binding of *S. aureus* surface adhesins to their ligands; for example, the clumping factor B binding to cytokeratin and fibrinogen (9). It is therefore possible that a minimum number of repeats is required for projection of the biofilm forming domains beyond other surface proteins. EM shows that SasG and Aap form highly elongated fibrils on the bacterial surface. *S. aureus* strains expressing SasG with eight full-length B repeats form peritrichous fibrils of varying density and a mean length of 53 ± 3 nm (9). Aap-expressing strains produce localized tufts of fibrillar appendages, usually in a lateral position in relation to the septum. The mean estimated length of the fibrils formed by an Aap variant with 12 full-length B repeats was 122 ± 11 nm for *S. epidermidis* NCTC 11047 and 159 ± 35 nm for RP62A (20). Although it was proposed that fibrils correspond to a single molecule of Aap (20), individual fibrils could not be distinguished as a result of very close packing.

Here we show that the biofilm-forming region of SasG has an elongated, contiguous structure formed by folded E and G5 domains connected by mutually stabilizing interfaces in head-to-tail fashion. The high resolution crystal structures of E-G5 (1.7 Å) and G5-E-G5 (1.85 Å) reveal that the domain structures are composed of flat, single-layer β-sheets and thus lack a compact hydrophobic core. Interdomain interfaces form interlocking connections between G5s and Es, leading to extended rod-like structures, which explain the appearance of the SasG fibrils on the bacterial surface.

Results

Structural Annotation of SasG Repeat Region. The SasG B repeat region of *S. aureus* strain NCTC 8325 contains nine 78-residue G5 domains and eight 50-residue segments (herein called E; Fig. 1*A* and Fig. S1*A*). Although there is one G5 domain structure (21) in the Protein Data Bank (PDB; with low ~15% sequence identity to SasG G5 domains), the E segments are predicted to be disordered (with high probability) by algorithms such as PONDR (22) and IUPred (23). The ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) NMR spectrum of G5^2 shows that most peaks have a wide distribution in the ^{1}H dimension, indicating that, as expected, it has a stable fold (Fig. 1*B*). In addition, as predicted by the sequence analysis tools, the shorter E segment appears disordered in isolation (Fig. 1*C*), with peaks having a narrow range of ^{1}H chemical shifts. The spectrum of G5^2-E (a B repeat in the previous annotation; Fig. 1*D*) shows a subpopulation of intense and poorly dispersed peaks whereas the remaining peaks are less intense but widely distributed and overlay almost exactly with the spectrum of G5 in isolation. Thus, the G5 domain retains its fold in the G5-E (B-repeat) context, E is disordered, and there is no evidence of a significant interface between G5 and E. Surprisingly, in the spectrum of E-G5^2 (Fig. 1*E*), most peaks show wide ^{1}H chemical shift dispersion, and an overlay of the G5^2 and E-G5^2 spectra shows that, although the majority of peaks associated with G5 in isolation remain in the same position, several are shifted in the E-G5^2 context. This implies that E is folded and that, although the overall structure of the G5 domain is similar, there is a significant E–G5 interface. Backbone dynamics of the E segment in the context of G5^2-E and E-G5^2 were estimated by using a ^{1}H-^{15}N heteronuclear nuclear Overhauser effect (NOE) experiment (Fig. 2*A*). In the case of G5^2-E, the average NOE values calculated for E residues were significantly lower than those for the G5^2 residues, indicating the high flexibility of the E region. The NOE values obtained for E-G5^2 were similar for most residues, demonstrating similar backbone dynamics for both subdomains on the subnanosecond timescale, and thus that E is folded.

To test the thermodynamic significance of the interdomain interfaces, we monitored the unfolding of a G5 domain in isolation and linked with E. Differential scanning calorimetry (DSC) thermograms (Fig. 2*B* and Table S1) show a significant increase in the melting temperature (T_m) of E-G5^2 (54 °C) compared with an isolated G5^2 domain (47 °C), implying a stabilizing effect of the N-terminal E. Consistent with the NMR studies, the T_m of the G5^2 domain was unaffected by the presence of a C-terminal E segment. Furthermore, the urea-induced equilibrium unfolding transitions (Fig. 2*C*) reveal that E-G5^2 is significantly more stable (m-value = 1.4 kcal·mol^{-1}·M^{-1}) than an individual G5^2 domain (m-value = 0.9 kcal·mol^{-1}·M^{-1}) and that both unfold as single cooperative units.

Crystal Structures of E-G5^2 and G5^1-E-G5^2 Explain the Elongated Nature of SasG Fibrils. Having established the domain boundaries of stably folded segments of SasG, we sought to characterize them structurally. Crystal structures of E-G5^2 and G5^1-E-G5^2 were determined at a resolution of 1.70 Å and 1.87 Å, respectively (statistics are provided in Tables S2 and S3). Both structures reveal a highly extended topology, which can be depicted as a cylinder with a diameter of approximately 20 Å and lengths of 115 Å for E-G5^2 and 170 Å for G5^1-E-G5^2 (Fig. 3 *A* and *B*). Each structure is formed from consecutive single-layer triple-stranded β-sheets, and their unusual elongation arises from the head-to-tail arrangement of the β-sheets.

The crystal structures, together with the NMR spectroscopy analysis, reveal that G5 domains and E segments are the building blocks of the B repeat region. G5 and E share 24% sequence identity (Fig. S1*B*) and show similar overall topology (Figs. 3 and 4*B* and *C*). They are each composed of triple-stranded single-

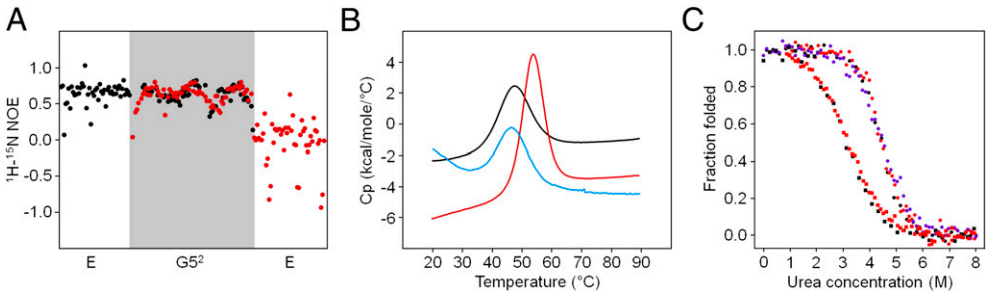

Fig. 2. Dynamics and stability of SasG domains. (*A*) Heteronuclear NOE experiment (^{1}H-^{15}N) measuring the dynamics of backbone ^{15}N nuclei of E-G5^2 (black) and G5^2-E (red). The G5^2 residues were assigned in both E-G5^2 and G5^2-E, whereas the E segment values were measured but not assigned to individual residues. (*B*) DSC thermograms for G5^2 (black), G5^2-E (blue), and E-G5^2 (red; Table S1). (*C*) Urea-induced equilibrium denaturation curves for G5^2 (■) and E-G5^2 (●). Folding was followed by monitoring changes in intrinsic tyrosine fluorescence starting with folded (black) or unfolded (red) protein. The purple points show unfolding of E-G5^2 monitored by changes in the circular dichroism at 235 nm.

layer β-sheets connected by an intertwined motif, which leads to a strand switch. Such a supersecondary structural motif was denoted by Ruggiero et al. as a β–triple helix–β and was first described for the G5 domain from RpfB from *Mycobacterium tuberculosis* (21). The SasG E segment is approximately 45 Å in length and is composed of two β-sheets: an N-terminal antiparallel β-sheet and a C-terminal β-sheet with a mixed parallel/antiparallel arrangement of β-strands (Fig. 3). The G5 domain extends to approximately 70 Å and is assembled from two (G5^1) or three (G5^2) triple-stranded β-sheets. The N-terminal antiparallel β-sheet of G5^1 is longer than the C-terminal sheet and corresponds to the two most N-terminal β-sheets in the G5^2 structure, which are also antiparallel. The C-terminal β-sheets of both G5 domains are nearly identical and show a mixed parallel/antiparallel arrangement of β-strands (Fig. 3). SasG G5^1 and G5^2 are nearly identical (rmsd = 1.34 Å; Fig. 4*A*) and show significant structural homology to the previously reported structure of the G5 domain. However, the RpfB G5 is composed of two β-sheets that are bent with respect to each other, providing the molecule with an overall arch-like structure, whereas SasG G5 domains are planar (Fig. 4 *D* and *E*).

The SasG β-sheets are exposed to the solvent on both faces, yet the exclusion of apolar side chains from aqueous solvent is considered to be a key stabilizing force in folded proteins. As expected from the appearance of the structure, the total relative accessible surface area (RSA) and the RSA for nonpolar side chains are larger for SasG domains than for globular proteins with a similar number of residues (Table S4). The ratio of the RSA calculated for apolar side chains and for all atoms is, however, comparable between the globular proteins and the extended SasG domains, suggesting a sufficient burial of hydrophobic residues. Moreover, the DSC thermograms (Fig. 2*B* and Table S1) clearly demonstrate a positive change in heat capacity upon thermal unfolding (and that the melting temperatures are not unusually low; Table S1). Thus,

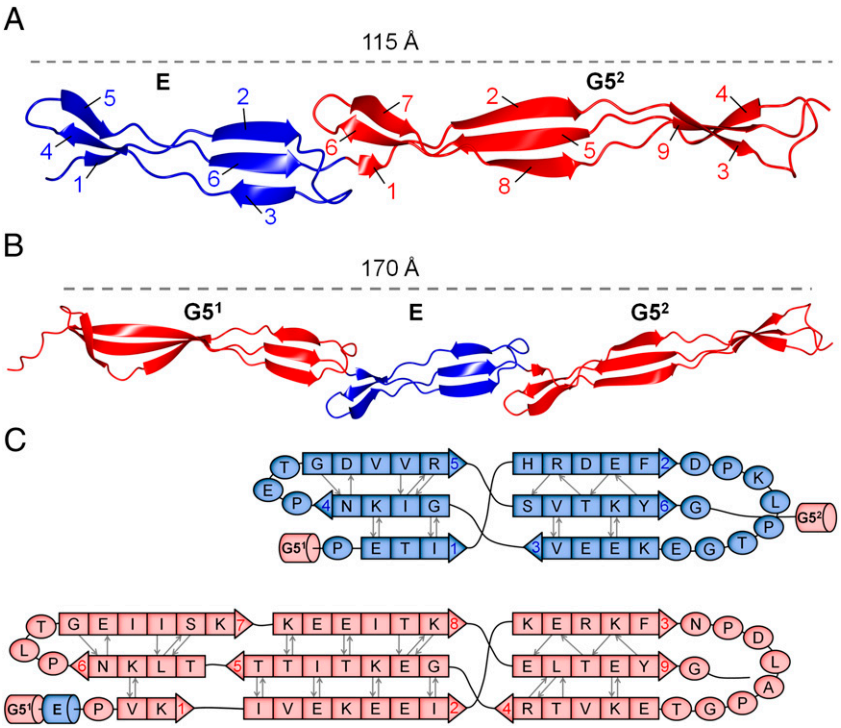

Fig. 3. Crystal structures of SasG domains; E and G5 domains are shown in blue and red, respectively. (*A*) Structure of E-G5^2. The β-strands are numbered for E and G5 domains. (*B*) Structure of G5^1-E-G5^2. (*C*) Schematic of the secondary structure of E (*Upper*) and G5^2 (*Lower*), as defined by the method of Kabsch and Sanders (64).

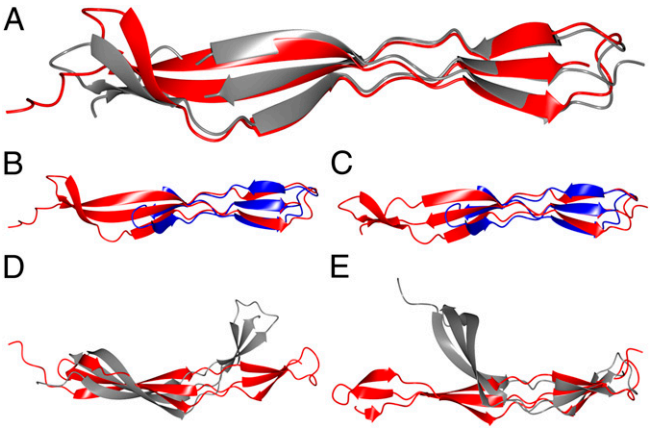

Fig. 4. Structural superposition of G5 and E domains. Automatic secondary structure matching in CCP4mg was implemented (65): (*A*) $G5^1$ (red) and $G5^2$ (gray), (*B*) $G5^1$ (red) and E (blue), (*C*) $G5^2$ (red) and E (blue), (*D*) $G5^1$ (red) and G5 from RpfB (gray), and (*E*) $G5^2$ (red) and G5 from RpfB (gray).

although they lack a typical compact hydrophobic core, the burial of nonpolar groups makes an important contribution to the stabilization of SasG domains. This appears to be achieved by strategic distribution of nonpolar side chains and aromatic residues throughout the SasG sequence. Bulky aromatics (tyrosines and phenylalanines) and longer hydrophobic side chains (isoleucines, leucines) are located at the interdomain interfaces, where they contribute to the formation of pseudohydrophobic cores. Smaller nonpolar residues are distributed along the single-layer β-sheets and pack against the hydrophobic moieties of long charged side chains, such as glutamates and lysines.

Interdomain Interfaces. The structural similarity in the N- and C-terminal β-sheets of SasG G5 and E domains results in highly homologous interdomain interfaces (G5–E and E–G5; Fig. 5*A*). The interfaces are formed by equivalent residues from G5 and E domains, which are involved in similar networks of van der Waals interactions. For example, Phe441 from $G5^1$ interacts with Pro499, Pro531, and Val537 from E, which corresponds to Phe510 from E interacting with Pro549, Pro599, and Ile605 from $G5^2$ (Fig. 5 *A* and *B*). In comparison with G5–E, the E–G5 interface is additionally stabilized by the presence of Leu600 in $G5^2$, which interacts with Phe510 and Tyr547 from E (Fig. 5*C*). In the context of the G5–E interface, this position is occupied by Glu532, which appears to distort the packing of nonpolar side chains. These structural differences partially rationalize the NMR and DSC data.

The residues involved in the E–G5 and G5–E interfaces are distributed across three strands of both domains and interdigitate (Fig. 6). Thus, the interdomain flexibility of G5s and Es is likely to be restricted, providing rigidity to multidomain SasG constructs. To test this hypothesis, we carried out sedimentation velocity analytical ultracentrifugation to estimate the overall shape of SasG constructs containing different numbers of domains. The calculated frictional ratios from the known molecular weights and sedimentation coefficients, as well as the prolate axial ratios (*a*/*b*), confirmed that all tested proteins are highly elongated monomers (Table 1). These results demonstrate that E-$G5^2$ and $G5^1$-E-$G5^2$ (Fig. 3 *A* and *B*) also adopt an extended conformation in solution. Furthermore, the length of constructs appears to be directly proportional to the number of G5 and E domains; that is, the long axis dimensions for modules of varying length are approximately additive. For example, the experimental *a*/*b* value for $G5^1$-E-$G5^2$ is 10.4, whereas the estimated axial ratio would be 10.7 if based on *a*/*b* values obtained for $G5^1$ (4.8) and E-$G5^2$ (5.9; Table 1). This additive nature of the axial ratios confirms the head-to-tail arrangement of domains and the presence of substantial rigidity in the interdomain connections.

SasG Domains Are Elongated Monomers in Solution. Previously, we and others showed that SasG and Aap B repeats dimerize in the presence of Zn^{2+}; however, the structural basis of dimerization was not clear (12, 13). The crystal structure of a SasG $G5^2$ dimer (Fig. S2) reveals that Zn^{2+} is coordinated by an N-terminal non-

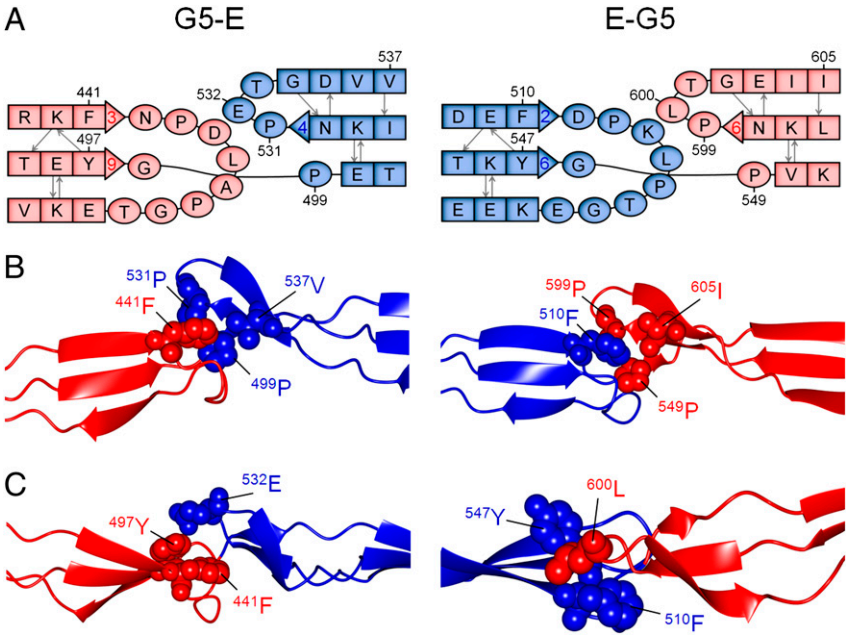

Fig. 5. G5–E and E–G5 interfaces in SasG. G5 and E domains are shown in red and blue, respectively. (*A*) Schematic representation of the interfaces indicating the similarity between the N- and C-terminal sheets of each domain. (*B*) Structural representation of the interfaces showing the equivalent residues involved in interdomain interactions. (*C*) Structural representation of the interfaces highlighting the major difference between G5–E and E–G5 (stereo images showing a portion of electron density are shown in Fig. S4).

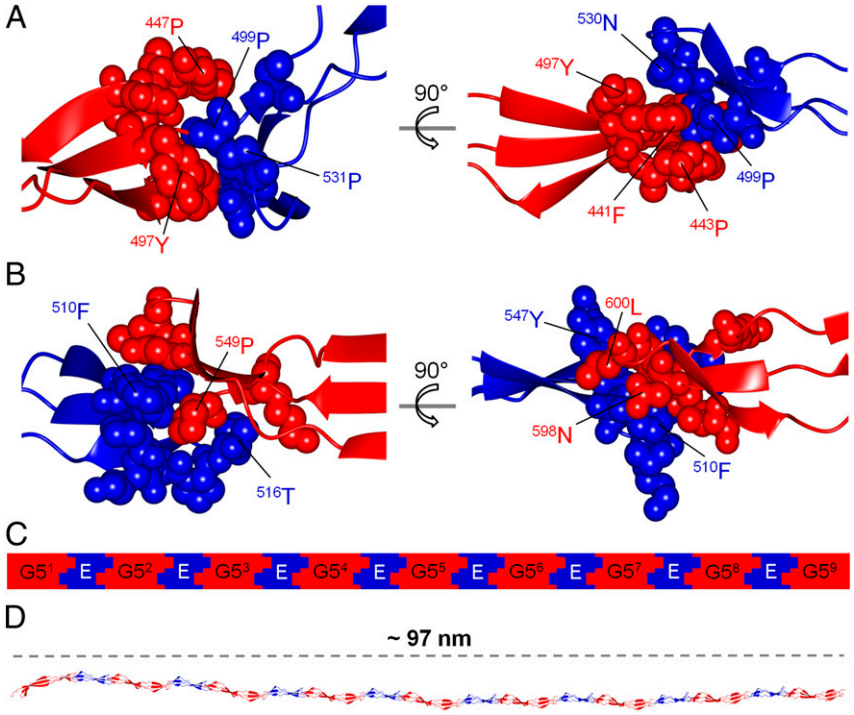

Fig. 6. Interlocking domain interfaces in SasG. G5 and E domains are shown in red and blue, respectively. (*A*) The $G5^1$–E interface highlighting the interdigitated residues. (*B*) The E–$G5^2$ interface highlighting the interdigitated residues. (*C*) Schematic representation and (*D*) structural model of interlocking G5 and E domains within SasG with nine G5 domains.

native histidine and two glutamate residues from each monomer. Mutation of the histidine abolished the Zn^{2+}-mediated dimerization of $G5^2$ and of a previously reported B repeat ($G5^2$-E; Fig. S3). Using the new SasG domain boundaries (defined based on the crystal structure of $G5^1$-E-$G5^2$), size-exclusion chromatography with multiangle laser light scattering (SEC-MALLS) confirmed that all stably folded modules ($G5^1$, $G5^2$, E-$G5^2$, $G5^1$-E-$G5^2$, and E-$G5^2$-E-$G5^3$) and partially or fully disordered SasG modules (E, $G5^1$-E, $G5^2$-E and $G5^1$-E-$G5^2$-E) are monomeric in solution at neutral pH in the absence and presence of 5 mM Zn^{2+} (Fig. 7, Fig. S3, and Table S5).

Discussion

Bacterial biofilm matrices are heterogeneous composed mainly of polysaccharides, proteins, nucleic acids, and lipids (24). Proteins detected in biofilms are either enzymes, responsible for degradation and modification of the extracellular biopolymers, or structural proteins that are attached to the bacterial surface and are involved in the formation and stabilization of the matrix. Some nonenzymatic surface proteins, such as LecA (25) and LecB (26) of *Pseudomonas aeruginosa*, bind to extracellular saccharides and link the bacterial surface with the matrix. Other proteinaceous components of the biofilm matrix are functional amyloids. Examples include amyloid fibers of Gram-positive *Bacillus subtilis* composed of TasA (27), as well as curli (28) and Tafi (29) amyloid fibrils of *Escherichia coli* and *Salmonella* spp., composed of the highly homologous subunit proteins CsgA and AgfA, respectively. Staphylococcal biofilm-associated surface proteins, including Aap and SasG, represent another group of matrix proteins. They also form elongated fibrillar structures at the bacterial cell surface, but via a mechanism distinct from amyloid assembly.

Elongated filamentous structures on bacterial surfaces are most commonly assembled from multiple polypeptide chains. For example, the M protein from *Streptococcus pyogenes* forms hair-like fimbriae that extend approximately 50 nm from the cell surface (30) and are composed of dimeric parallel α-helical coiled-coil structures (31). Pili in Gram-negative bacteria are typically formed by noncovalent homopolymerization of major pilus subunit proteins (pilins) (32). Recently discovered pili in Gram-positive bacteria are formed by covalent polymerization of pilin subunits in a process that requires a specific sortase enzyme (33). The additive long axis dimensions for SasG modules of varying length (Table 1) suggest that the interdomain interfaces provide rigidity. This, taken together with the head-to-tail arrangement of the domains, implies that in vivo SasG structures (with as many as 10 G5 domains) could form the highly extended structures observed on the bacterial surface. As a single-chain structure composed of single-layer β-sheets, SasG would form an unusually thin filament on the cell surface, providing a highly efficient solution to the formation of an extended structure on this scale (50–100 nm). Thus, SasG represents a new paradigm in the production of thin, rod-like protein structures.

A striking feature of the SasG structures (Fig. 3 *A* and *B*) is the apparent lack of a compact hydrophobic core, although the T_m, free energies of unfolding, and *m*-values for urea denaturation

Table 1. Summary of analytical ultracentrifugation data

Protein	Theoretical MW (Da)	*s** (S)	$s_{20,w}$ (S)	f/f_0*	*a/b**	f/f_0	*a/b*
$G5^1$	9,699.1	1.007	1.047	1.53	5.5	1.47	4.8
$G5^2$	9,654.8	1.066	1.107	1.56	5.8	1.45	4.4
E-$G5^2$	14,471.2	1.291	1.341	1.72	8.0	1.57	5.9
$G5^1$- E-$G5^2$	23,729.8	1.472	1.530	1.89	10.7	1.87	10.4
E-$G5^1$- E-$G5^2$	28,512.1	1.594	1.655	2.06	13.5	1.98	12.2

a/b, prolate axial ratio calculated from f/f_0; *a/b**, axial ratio for a prolate ellipsoid; f/f_0, frictional ratio calculated from the known molecular weight and sedimentation coefficient; f/f_0*, experimental frictional ratio; MW, molecular weight; *s** (S), experimental sedimentation coefficient; $s_{20,w}$ (S), sedimentation coefficient corrected to water.

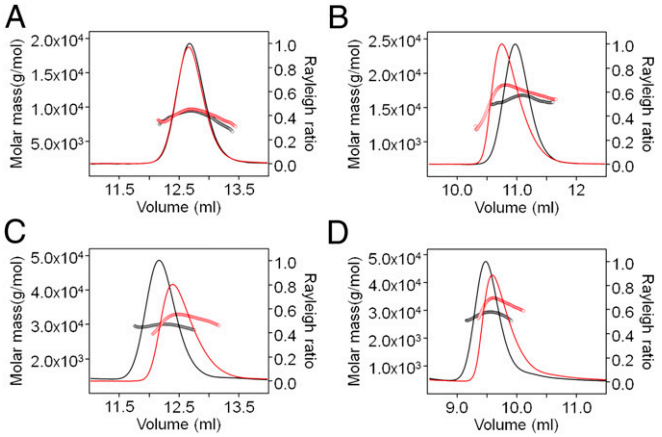

Fig. 7. SEC-MALLS analysis of the oligomeric state of SasG domains in the presence (red) and absence (black) of Zn^{2+} (5 mM): (*A*) $G5^2$, (*B*) E-$G5^2$, and (*C*) $G5^1$-E-$G5^2$, i.e., native sequence with four additional N-terminal amino acids (i.e., GPHM); and (*D*) E-$G5^1$-E-$G5^2$, i.e., native sequence with four additional N-terminal amino acids (i.e., GPHM). The exact molar masses and sequences are listed in Table S5.

of the G5 and E–G5 constructs suggest they have stabilities only slightly lower than globular domains of similar size. Typically, β-sheets, whether in extracellular or cytoplasmic proteins, are amphipathic, and their hydrophobic faces form a compact hydrophobic core with other secondary structure elements. In addition, extracellular β-sheet–containing domains are often stabilized through disulfide bond formation. Despite the increased solvent exposure of SasG domains, the burial of apolar residues appears to be sufficient to stabilize their fold. Moreover, more than 95% refolding efficiencies (even after multiple unfolding cycles) were observed in DSC experiments for structured SasG constructs, at protein concentrations at which misfolding and aggregation are typically observed.

Individual (rather than tandemly arrayed) single-layer β-sheet domains have been reported previously; for example, OspA, a surface protein of *Borrelia burgdorferi*, contains a three-stranded β-sheet between two globular domains (34), and WRKY4 transcription factors (35) contain a four-stranded single-layer β-sheet that is stabilized by zinc. In OspA and WRKY4, the thermodynamic stability of the sheets is suggested to depend on interactions of nonpolar side chains with the hydrophobic parts of long hydrophilic amino acids. The sequences of G5 and E domains show a high content of long charged residues (Glu, ∼15%; and Lys, ∼15%), and analysis of the E-$G5^2$ and $G5^1$-E-$G5^2$ crystal structures reveal that short nonpolar side chains (Ala, Val, Ile, Leu) form small hydrophobic clusters surrounded by large hydrophilic amino acids. In addition, the long, charged side chains form cross-strand arrays of alternating charges observed previously in antiparallel β-sheets (36).

Despite the difference in length, the SasG G5 and E domains show significant structural similarity (Figs. 3 and 4 *B* and *C*). As in the G5 domain from RpfB (21), the central "switch" region of the SasG G5 and E domains resembles the collagen triple-helical structural motif. The pseudotriple helix of SasG domains has a PPII-like conformation and is rich in proline and glycine residues. The packing of intertwined SasG strands within the switch region is tight, despite their mixed parallel/antiparallel arrangement. This feature of the β–triple helix–β fold is stabilized by two (21) of the five conserved glycines after which the G5 domain was named (14). The SasG E segment contains structurally equivalent glycine residues within its pseudotriple helix. The distribution of hydrophobic, charged, and proline residues within the E segment also resembles that of the G5 domain (Fig. 3). This suggests that E–G5 has evolved through processes of G5 duplication, mutation, fusion, and partial deletion. Despite the structural similarity between the two SasG domains, the E segment is significantly less stable than G5 (Figs. 1 *B–E* and 2 *A* and *B*) and dependent on the E–G5 interface for folding.

The sequences of SasG (and Aap) repeats are very similar at the DNA level (16) and vary in number dependent on the bacterial strain (16, 17). However, although sequence identity at the DNA level might be advantageous in facilitating recombination events, the resulting protein sequence identity is a potential problem from a protein folding perspective. The immediate juxtaposition of domains with identical sequence can promote misfolding events (18). This might explain the apparent evolutionary pressure to maintain sequence identity at less than 40% between adjacent domains that was revealed by an analysis of proteins containing strings of Ig and fibronectin-type III domains (19). The formation of two domains by each sequence repeat, as in SasG, could provide an elegantly simple solution to this problem. As there are very few structures of native tandem repeats of domain size in the PDB (37), this observation will assist future structural studies of such proteins, several of which [e.g., Rib and alphaC (38)] are expressed on the surface of pathogens.

The structure of a Zn^{2+}-mediated $G5^2$ dimer (Fig. S3) and SEC-MALLS analysis of mutants (Fig. S2) show that our previously reported dimerization of a SasG B repeat at millimolar Zn^{2+} concentrations (12) was dependent on two glutamate residues and a single nonnative histidine in each monomer. This phenomenon is highly context-dependent, as the presence of this histidine did not result in Zn^{2+}-dependent dimerization of other SasG constructs (Fig. 7). Notwithstanding, in vivo studies demonstrated that zinc is essential for protein-mediated biofilm formation in *S. epidermidis* (13) and *S. aureus* (12). Aap G5 domains, which contain native C-terminal histidine residues (not present in SasG), dimerize in a zinc-dependent manner (13). However, there is no experimental evidence that directly links this in vitro dimerization to the Zn^{2+}-dependence of protein-mediated biofilm formation in *S. epidermidis.* In addition, free Zn^{2+} is normally present at much lower concentrations in vivo [estimated as low as femtomolar (39) in bacterial cells and subnanomolar in mammalian plasma (40)] than was required for Aap dimerization. Thus, our study implies that, rather than homodimerization, zinc might mediate binding to another polymeric component of the staphylococcal biofilm in *ica*-null strains, such as extracellular teichoic acid or DNA (41). Alternatively, the rod-like structure might play a role in maintaining bacterial separation in the biofilm. Furthermore, a recent study on Aap-mediated biofilm formation by *S. epidermidis* showed that anti-G5 monoclonal antibodies enhanced bacterial accumulation, whereas those with an epitope in the E segment inhibited biofilm accumulation to 60% of the maximum (42). This suggests that G5 and E domains might play different roles during bacterial accumulation, despite their structural similarity. Further studies will be required to reveal the full molecular basis of protein-mediated staphylococcal biofilm formation. The domain arrangement, high resolution structures, and lack of Zn^{2+}-dependent dimerization of SasG domains reported here provide a very significant step toward this goal and the overall aim of informing the development of new therapeutic strategies for the treatment or prevention of staphylococcal biofilm infections.

Materials and Methods

Cloning. DNA sequences of B^1, B^2, and E^1G5^2, codon-optimized for *Escherichia coli*, were synthesized (GenScript) and subcloned into the pSKB2 expression vector providing an N-terminal hexahistidine tag. Other SasG modules were generated based on these three templates by using standard cloning and mutagenesis techniques.

Protein Production. Unlabeled, ^{15}N-labeled, and ^{13}C-^{15}N-labeled proteins were produced in *E. coli* BL21(DE3) by using Luria–Bertani, ^{15}N-M9, and ^{13}C-^{15}N-M9 medium, respectively. Selenomethionylated E-G5^2-L17M-L103M was produced in *E. coli* B834(DE3) (Novagen) by using SeMet-supplemented minimal medium. Standard procedures were used. After induction with isopropyl β-D-1-thiogalactopyranoside at an OD_{600} of 0.6, *E. coli* cultures were grown at 20 °C for 24 h. SasG domains were purified by nickel-affinity purification using a HisTrap HP column (GE Healthcare). The His-tag was cleaved by using HRV 3C protease (Novagen), and removed with a HisTrap HP column. In case of insufficient purity, size exclusion chromatography was applied with a HiLoad 16/60 Superdex 75 column (GE Healthcare).

NMR Spectroscopy. NMR spectra were acquired at 25 °C on a Bruker AVANCE II 700 MHz spectrometer equipped with a triple-resonance probe. Samples for HSQC experiments contained ^{15}N-labeled protein (0.5 mM) in 20 mM Tris-HCl, pH 7.0, 100 mM NaCl, and 10% ^2H_2O. Steady-state ^{1}H-^{15}N NOE values were determined by recording HSQC spectra in the presence (i.e., NOE) and absence (i.e., NONOE) of ^{1}H saturation. NOE and NONOE experiments were deconvoluted and the NOE value was calculated from the intensity (i.e., volume) of the cross peaks by using the following formula:

$$\mathrm{NOE} = (I_{\mathrm{NOE}})/(I_{\mathrm{NONOE}}) \quad [1]$$

For the assignment of the G5^2 HSQC spectrum, a series of 3D experiments was recorded (HNCO, HNCACO, CBCACONH, and CBCANH) for a sample containing ^{13}C-^{15}N-G5^2 (1 mM) in 20 mM Tris-HCl, pH 7.0, 100 mM NaCl, and 10% ^2H_2O. Spectra were processed by using NMRPipe (43) and viewed in NMRView (44). NMR backbone resonances were assigned using CcpNmr Analysis 2.1 (45, 46).

DSC. DSC scans were acquired on a MicroCal VP-DSC calorimeter for protein samples at a concentration of 1 mg/mL in 20 mM Tris-HCl, pH 7.4, 150 mM NaCl. Scans of degassed buffer and proteins were recorded for temperatures ranging from 10 to 90 °C at a scan rate of 90 °C/h. Data analysis was performed using MicroCal Origin 7.0. A progress baseline was used for the sample baseline correction before area integration or fitting of the unfolding endotherm. Data points between 20 and 90 °C were used for fitting. The reversibility of thermal unfolding was verified by repetitive scans on the same sample.

Folding Studies. Protein stability was determined by using urea-induced equilibrium denaturation. Fluorescence measurements were performed on a Cary Eclipse fluorescence spectrophotometer. CD measurements were performed on an Applied Photophysics Chirascan CD spectrometer. The experiments were carried out by using 5 μM protein in PBS solution (pH 7.4) at 25 °C after equilibration overnight. The fluorescence excitation wavelength was 274 nm, and emission was followed at 302 nm. For CD, the ellipticity at 235 nm was followed. The data were fitted to a two-state equation (48).

Crystallography. Crystallization trials were performed by using the sitting drop vapor diffusion method. E-G5^2 crystallized in 0.2 M NH_4Cl and 20% (wt/vol) PEG3350, using the protein solution at a concentration of 2 mM (30 mg/mL) in 20 mM Tris, pH 7.0, 100 NaCl. Crystals of selenomethionylated E-G5^2-L17M-L103M were obtained in 0.05 M Tris, 0.05 M Bicine, pH 8.5, 12.5% (vol/vol) 2-methyl-2, 4-pentanediol, 12.5% (wt/vol) PEG 1000, 12.5% (wt/vol) PEG 3350, 0.02 M sodium formate, 0.02 M sodium citrate, and 0.02 M sodium oxamate, for the protein solution at a concentration of 2 mM (30 mg/mL) in 20 mM Tris, pH 7.0, 100 NaCl. G5^1-E-G5^2 crystallized in 0.2 M $MgCl_2$, 0.1 M Bis-Tris, pH 5.5, and 25% (wt/vol) PEG 3350, using the protein solution at a concentration of 1.2 mM (27.6 mg/mL) in 20 mM Tris, pH 7.0, 100 NaCl.

Diffraction data were collected at 100 K at the Diamond Light Source on beamline I04 (native E-G5^2 and G5^1-E-G5^2) and I02 (SeMet-E-G5^2-L17M-L103M). Images recorded for E-G5^2 and G5^1-E-G5^2 were processed with XDS (48), and the data were scaled by using SCALA (49) from the CCP4 program suite (50). Multiple-wavelength anomalous data collected for SeMet-E-G5^2-L17M-L103M were processed by using the HKL2000 suite (51).

The structure of SeMet-E-G5^2-L17M-L103M was solved by the multi-wavelength anomalous dispersion method by using SHELXC/D/E (52). Four heavy atom sites were identified, and a partial model with two molecules in the asymmetric unit was produced by Buccaneer (53). One of the molecules was used as a search model for molecular replacement for E-G5^2 data. The solution, produced by Phaser (54), consisted of one molecule in the asymmetric unit. A complete model of the E-G5^2 structure was generated in ARP/wARP (55). The structure of G5^1-E-G5^2 was also solved by molecular replacement. First, the E-G5^2 structure was used as a search model in Phaser, revealing two E segments and consequently two G5^1-E-G5^2 molecules in the asymmetric unit. Subsequent molecular replacement runs with MOLREP (56), followed by model building in ARP/wARP, localized the remaining four G5 domains. MOLREP was implemented because of its shorter run time. Initial refinement was carried out with REFMAC5 (57) using the "jelly body" option. Final refinement runs were performed in Phenix 1.7.1 (58) using TLS restraints [generated with TLSMD (59)] and simulated annealing. The structures were visualized and manually rebuilt in COOT (60). The stereochemistry of the final model was evaluated with MolProbity (61). Data collection and refinement statistics are shown in Tables S2 and S3. Atomic coordinates and structure factor amplitudes have been deposited to PDB under codes 3TIP and 3TIQ.

Analytical Ultracentrifugation. Sedimentation velocity experiments were conducted on a Beckman Optima XL/I analytical ultracentrifuge using an An-60 Ti rotor at 20 °C. Protein concentrations of 2 mg/mL in 20 mM Tris, 150 mM NaCl buffer, pH 7.4, were centrifuged at 58,000 and 52,000 rpm (G5^1, G5^2, and E-G5^2) or 42,000 rpm (G5^1-E-G5^2, E-G5^2-E-G5^3) collecting absorbance data at 280 nm and interference data. The program SEDFIT (62) was used to determine sedimentation coefficients and frictional ratios and to convert these to axial ratios. Buffer density and viscosity and protein partial specific volumes were calculated by using SEDNTERP (63).

SEC-MALLS. SEC-MALLS experiments were performed by using a Superdex 75 HR10/30 column (GE Healthcare) and a Shimadzu HPLC System. Protein samples (100 μL) at a concentration of 1.5 mg/mL were loaded onto a gel filtration column and eluted with one column volume (24 mL) of an appropriate running buffer at a flow rate of 0.5 mL/min. The eluting fractions were monitored using a DAWN HELEOS-II 18-angle light scattering detector (Wyatt Technologies), a SPD20A UV/Vis detector (Shimadzu), and an Optilab rEX refractive index monitor (Wyatt Technologies). Data were analyzed by using Astra (Wyatt Technologies).

ACKNOWLEDGMENTS. We thank Garib Murshudov for assistance with refinement of the E-G5^2 and G5^1-E-G5^2 structures and Alexey Murzin for helpful discussions. This work was carried out with the support of the Diamond Light Source. This work was supported by British Heart Foundation nonclinical doctoral Studentship FS/08/025/24765 (D.T.G.); British Heart Foundation Senior Basic Science Research Fellowship FS/07/034 (to J.R.P.); Wellcome Trust Grant 064417 (to J.C.); Biotechnology and Biological Sciences Research Council LoLa Grant BB/G020671/1 (J.A.W.); and Science Foundation Ireland Programme Investigator Grant 08/IN.1/B1854 (to T.J.F.). J.C. is a Wellcome Trust Senior Research Fellow.

1. World Health Organization (2011) *Report on the Burden of Endemic Health Care-Associated Infection Worldwide* (World Health Organization, Geneva).
2. Rosenthal VD, et al.; International Nosocomial Infection Control Consortium Members (2008) International Nosocomial Infection Control Consortium report, data summary for 2002-2007, issued January 2008. *Am J Infect Control* 36:627–637.
3. Brady MT (2005) Health care-associated infections in the neonatal intensive care unit. *Am J Infect Control* 33:268–275.
4. Costerton JW, Stewart PS, Greenberg EP (1999) Bacterial biofilms: A common cause of persistent infections. *Science* 284:1318–1322.
5. Harris LG, Richards RG (2006) Staphylococci and implant surfaces: A review. *Injury* 37(Suppl 2):S3–S14.
6. Dunne WM, Jr. (2002) Bacterial adhesion: Seen any good biofilms lately? *Clin Microbiol Rev* 15:155–166.
7. O'Gara JP (2007) ica and beyond: Biofilm mechanisms and regulation in *Staphylococcus epidermidis* and *Staphylococcus aureus*. *FEMS Microbiol Lett* 270:179–188.
8. Hussain M, Herrmann M, von Eiff C, Perdreau-Remington F, Peters G (1997) A 140-kilodalton extracellular protein is essential for the accumulation of *Staphylococcus epidermidis* strains on surfaces. *Infect Immun* 65:519–524.
9. Corrigan RM, Rigby D, Handley P, Foster TJ (2007) The role of *Staphylococcus aureus* surface protein SasG in adherence and biofilm formation. *Microbiology* 153:2435–2446.
10. Roche FM, Meehan M, Foster TJ (2003) The *Staphylococcus aureus* surface protein SasG and its homologues promote bacterial adherence to human desquamated nasal epithelial cells. *Microbiology* 149:2759–2767.
11. Rohde H, et al. (2005) Induction of *Staphylococcus epidermidis* biofilm formation via proteolytic processing of the accumulation-associated protein by staphylococcal and host proteases. *Mol Microbiol* 55:1883–1895.
12. Geoghegan JA, et al. (2010) Role of surface protein SasG in biofilm formation by *Staphylococcus aureus*. *J Bacteriol* 192:5663–5673.
13. Conrady DG, et al. (2008) A zinc-dependent adhesion module is responsible for intercellular adhesion in staphylococcal biofilms. *Proc Natl Acad Sci USA* 105:19456–19461.

14. Bateman A, Holden MTG, Yeats C (2005) The G5 domain: A potential N-acetylglucosamine recognition domain involved in biofilm formation. *Bioinformatics* 21:1301–1303.
15. Bateman A, et al. (2004) The Pfam protein families database. *Nucleic Acids Res* 32(database issue):D138–D141.
16. Roche FM, et al. (2003) Characterization of novel LPXTG-containing proteins of *Staphylococcus aureus* identified from genome sequences. *Microbiology* 149:643–654.
17. Rohde H, et al. (2007) Polysaccharide intercellular adhesin or protein factors in biofilm accumulation of *Staphylococcus epidermidis* and *Staphylococcus aureus* isolated from prosthetic hip and knee joint infections. *Biomaterials* 28:1711–1720.
18. Borgia MB, et al. (2011) Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins. *Nature* 474:662–665.
19. Wright CF, Teichmann SA, Clarke J, Dobson CM (2005) The importance of sequence diversity in the aggregation and evolution of proteins. *Nature* 438:878–881.
20. Banner MA, et al. (2007) Localized tufts of fibrils on *Staphylococcus epidermidis* NCTC 11047 are comprised of the accumulation-associated protein. *J Bacteriol* 189: 2793–2804.
21. Ruggiero A, et al. (2009) Crystal structure of the resuscitation-promoting factor (DeltaDUF)RpfB from *M. tuberculosis*. *J Mol Biol* 385:153–162.
22. Romero P, et al. (2001) Sequence complexity of disordered protein. *Proteins* 42:38–48.
23. Dosztányi Z, Csizmok V, Tompa P, Simon I (2005) IUPred: Web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* 21:3433–3434.
24. Flemming HC, Wingender J (2010) The biofilm matrix. *Nat Rev Microbiol* 8:623–633.
25. Diggle SP, et al. (2006) The galactophilic lectin, LecA, contributes to biofilm development in *Pseudomonas aeruginosa*. *Environ Microbiol* 8:1095–1104.
26. Tielker D, et al. (2005) *Pseudomonas aeruginosa* lectin LecB is located in the outer membrane and is involved in biofilm formation. *Microbiology* 151:1313–1323.
27. Romero D, Aguilar C, Losick R, Kolter R (2010) Amyloid fibers provide structural integrity to *Bacillus subtilis* biofilms. *Proc Natl Acad Sci USA* 107:2230–2234.
28. Vidal O, et al. (1998) Isolation of an *Escherichia coli* K-12 mutant strain able to form biofilms on inert surfaces: involvement of a new ompR allele that increases curli expression. *J Bacteriol* 180:2442–2449.
29. Austin JW, Sanders G, Kay WW, Collinson SK (1998) Thin aggregative fimbriae enhance *Salmonella enteritidis* biofilm formation. *FEMS Microbiol Lett* 162:295–301.
30. Fischetti VA (1989) Streptococcal M protein: Molecular design and biological behavior. *Clin Microbiol Rev* 2:285–314.
31. McNamara C, et al. (2008) Coiled-coil irregularities and instabilities in group A *Streptococcus* M1 are required for virulence. *Science* 319:1405–1408.
32. Geibel S, Waksman G (2011) Crystallography and electron microscopy of chaperone/usher pilus systems. *Adv Exp Med Biol* 715:159–174.
33. Proft T, Baker EN (2009) Pili in Gram-negative and Gram-positive bacteria - structure, assembly and their role in disease. *Cell Mol Life Sci* 66:613–635.
34. Li H, Dunn JJ, Luft BJ, Lawson CL (1997) Crystal structure of Lyme disease antigen outer surface protein A complexed with an Fab. *Proc Natl Acad Sci USA* 94:3584–3589.
35. Yamasaki K, et al. (2005) Solution structure of an Arabidopsis WRKY DNA binding domain. *Plant Cell* 17:944–956.
36. Wouters MA, Curmi PMG (1995) An analysis of side chain interactions and pair correlations within antiparallel beta-sheets: The differences between backbone hydrogen-bonded and non-hydrogen-bonded residue pairs. *Proteins* 22:119–131.
37. Jorda J, Xue B, Uversky VN, Kajava AV (2010) Protein tandem repeats - the more perfect, the less structured. *FEBS J* 277:2673–2682.
38. Lachenauer CS, Creti R, Michel JL, Madoff LC (2000) Mosaicism in the alpha-like protein genes of group B streptococci. *Proc Natl Acad Sci USA* 97:9630–9635.
39. Outten CE, O'Halloran TV (2001) Femtomolar sensitivity of metalloregulatory proteins controlling zinc homeostasis. *Science* 292:2488–2492.
40. Magneson GR, Puvathingal JM, Ray WJ, Jr. (1987) The concentrations of free Mg^{2+} and free Zn^{2+} in equine blood plasma. *J Biol Chem* 262:11140–11148.
41. Otto M (2008) Staphylococcal biofilms. *Curr Top Microbiol Immunol* 322:207–228.
42. Hu J, et al. (2011) Monoclonal antibodies against accumulation-associated protein affect EPS biosynthesis and enhance bacterial accumulation of *Staphylococcus epidermidis*. *PLoS ONE* 6:e20918.
43. Delaglio F, et al. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6:277–293.
44. Johnson BA, Blevins RA (1994) NMR View - a computer-program for the visualization and analysis of NMR data. *J Biomol NMR* 4:603–614.
45. Fogh RH, et al. (2005) A framework for scientific data modeling and automated software development. *Bioinformatics* 21:1678–1684.
46. Vranken WF, et al. (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins* 59:687–696.
47. Pace CN (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol* 131:266–280.
48. Kabsch W (1988) Evaluation of single-crystal X-ray-diffraction data from a position-sensitive detector. *J Appl Cryst* 21:916–924.
49. Evans PR (1993) Data reduction. *Proceedings of CCP4 Study Weekend, 1993, on Data Collection and Processing* (Daresbury Lab, Warrington, UK), pp 114–122.
50. Collaborative Computational Project, Number 4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50:760–763.
51. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326.
52. Sheldrick GM (2010) Experimental phasing with SHELXC/D/E: Combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr* 66:479–485.
53. Cowtan K (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* 62:1002–1011.
54. McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Cryst* 40:658–674.
55. Perrakis A, Morris R, Lamzin VS (1999) Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* 6:458–463.
56. Vagin A, Teplyakov A (1997) MOLREP: An automated program for molecular replacement. *J Appl Cryst* 30:1022–1025.
57. Vagin AA, et al. (2004) REFMAC5 dictionary: Organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr* 60:2184–2195.
58. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66:213–221.
59. Painter J, Merritt EA (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr* 62: 439–450.
60. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126–2132.
61. Davis IW, et al. (2007) MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35(Web server issue):W375–W383.
62. Schuck P (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J* 78:1606–1619.
63. Laue TM, Shah BD, Ridgeway TM, Pelletier SL (1992) Computer-aided interpretation of analytical sedimentation data for proteins. *Analytical Ultracentrifugation in Biochemistry and Polymer Science*, eds Harding SE, et al. (Royal Society of Chemistry, Cambridge, UK), pp 90–125.
64. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22:2577–2637.
65. McNicholas S, Potterton E, Wilson KS, Noble ME (2011) Presenting your structures: The CCP4mg molecular-graphics software. *Acta Crystallogr D Biol Crystallogr* 67:386–394.