Article:
Bruce, AJE, Paci, E and Brockwell, DJ orcid.org/0000-0002-0802-5937 (2023) Collagen-like motifs of SasG: a novel fold for protein mechanical strength. Journal of Molecular Biology, 435 (6). 167980. ISSN 0022-2836

https://doi.org/10.1016/j.jmb.2023.167980

Collagen-like motifs of SasG: a novel fold for protein mechanical strength.

*Alexander J. E. Bruce, Emanuele Paci[‡] and David J. Brockwell**

Astbury Centre for Structural and Molecular Biology and School of Molecular and Cellular Biology, Faculty of Biological Sciences, University of Leeds, LS2 9JT, UK

Single-molecule force spectroscopy (SMFS), protein unfolding, SasG, mechanobiology, collagen-related disease and Osteogenesis Imperfecta (OI).

AUTHOR INFORMATION

Corresponding Author

*David J. Brockwell

Tel: +44 (0) 113 343 7821

Email: D.J.Brockwell@leeds.ac.uk

[‡]E.P. present address: University of Bologna, Department of Physics, Bologna, Italy

Abstract

The *Staphylococcus aureus* surface protein G (SasG) is associated with host colonisation and biofilm formation. As colonisation occurs at the liquid-substrate interface bacteria are subject to a myriad of external forces and, presumably as a consequence, SasG displays extreme mechanical strength. This mechanical phenotype arises from the B-domain; a repetitive region composed of alternating E and G5 subdomains. These subdomains have an unusual structure comprising collagen-like regions capped by triple-stranded β-sheets. To identify the determinants of SasG mechanical strength, we characterised the mechanical phenotype and thermodynamic stability of 18 single substitution variants of a pseudo-wildtype protein. Visualising the mechanically-induced transition state at a residue-level by ϕ-value analysis reveals that the main force-bearing regions are the N- and C-terminal 'Mechanical Clamps' and their side-chain interactions. This is tailored by contacts at the pseudo-hydrophobic core interface. We also describe a novel mechanical motif - the collagen-like region and show that glycine to alanine substitutions, analogous to those found in *Osteogenesis Imperfecta* (brittle bone disease), result in a significantly reduced mechanical strength.

Keywords: Single-molecule force spectroscopy (SMFS), protein unfolding, SasG, mechanobiology, collagen-related disease and Osteogenesis Imperfecta (OI).

ABBREVIATIONS

Aap, accumulation-associated protein; AFM, Atomic Force microscopy; CLM, collagen-like motif; CWA, cell-wall anchored; *E. coli*, *Escherichia coli*; F_U, unfolding force; FWHM, full width at half maximum; LC-MS, liquid chromatography-mass spectroscopy; OI, Osteogenesis Imperfecta; PBS, phosphate buffered salineMD, molecular dynamic; SasG, *S. aureus* surface protein G; *S. aureus*, *Staphylococcus aureus*; SMFS, single-molecule force spectroscopy; TS, transition state; WLC, worm-like chain; WT, wild-type;

Introduction

Bacteria typically populate liquid-substrate interfaces, rather than bulk liquid, by the formation of biofilms - a bacterial community encapsulated in an extracellular polysaccharide substance [1,2]. The attachment and growth of bacteria to the substrate during biofilm formation is usually mediated by cell-surface structures such as cell-wall anchored (CWA) proteins through binding of their A (adhesin) domains to cognate host ligands [3–5]. These interactions are typically strong and long lasting, such as the interaction between the *Staphylococcal* surface protein SpsD and fibrinogen, which boasts a bond strength of 1.8 nN measured by Atomic Force Microscopy (AFM) at a retraction velocity of 1000 nms^{-1} [5]. As these adhesion:ligand interactions rupture at forces similar to covalent bonds [6], the remaining structure may either be expected to extend under tensile force (to reduce the load on the non-covalent adhesin-ligand bond to extend lifetime) or to exhibit high mechanical strength. For example, type 1 pili are helical rods composed of 1000 globular subunits utilised by *E. coli* to attach to the host epithelium [7]. Under external force, pili extend reversibly by unwinding of the helical quaternary structure. In contrast, the E and G5 subdomains of the *Staphylococcus epidermidis* accumulation-associated protein (Aap) (a 'periscope' protein [8]) sequentially unfold at extreme forces of ~312 and ~475 pN, respectively, prior to adhesion:ligand bond dissociation (~3 nN at 1000 nms^{-1}) [9]. In a similar fashion, the E and G5 subdomains from the analogous CWA 'periscope' protein, SasG, also exhibit remarkable mechanostability, unfolding at ~ 230 and ~ 380 pN at 1500 nms^{-1} using the AFM [10].

SasG is a multi-domain protein thought to promote host colonisation and biofilm formation [11–14] and comprises a host-ligand binding A domain [11] and a B domain of 3-10 tandemly arrayed B-repeats [15] (Figure 1A). Each B-repeat is formed from one copy each of structurally related E and G5 subdomains [16]. These subdomains have an unusual flat elongated structure, comprising

three interlaced strands forming a collagen-like region, which is capped by a triple stranded β-sheet with a mixed parallel/anti-parallel arrangement [16] (Figure 1B). The interfaces between the E and G5 subdomains are compact and the presence of hydrophobic side chains forms a pseudo-hydrophobic core [16]. The collagen-like motif (CLM) comprises a mixed parallel/antiparallel PPII-like chain arrangement leading to a staggered distorted triple helix resembling the structure of collagen but involving a single polypeptide chain only (Figure 1C) [17]. The triple helical structure of collagen (a right-handed triple helix comprising left-handed helical polypeptides) is known to perform a mechanical function in vertebrates, but its complex quaternary structure has hindered previous investigations of collagen mechanical properties at the single-molecule level [18]. Consequently, characterisation of the mechanical effects of amino-acid substitutions associated with Osteogenesis Imperfecta (OI) a disease of the connective tissues characterised by extremely fragile bones has not been investigated at the single-molecule level.

Which of the unusual structural features of SasG endow the observed extreme mechanical strength of this protein? Single-molecule force spectroscopy (SMFS) is a powerful tool for delineating the effect of force at the molecular level. SMFS permits the measurement of force response to changes in the end-to-end length of a protein. In general, all α-helical proteins or local helical regions are the most mechanically labile (unfolding at a low force or in the noise of the instrument), followed by α/β structures, with all β-proteins exhibiting the highest mechanostability [19,20]. The most mechanically stable proteins share a common shear topology whereupon two directly hydrogen-bonded β-strands are extended in opposite directions parallel to the β-strand long-axis. This requires the simultaneous rupture of hydrogen bonds between the β-strands creating a ‘mechanical clamp’ motif [21–23]. Several of the most mechanically robust proteins characterised to date are derived from bacterial extracellular repetitive Immunoglobulin

(Ig)-like fold domains including the scaffoldin c7A domain [24] and the SdrG B1 domain [25]. The mechanostability of the latter is enhanced by the coordination of Ca^{2+} ions across the hydrogen bonded 'mechanical clamp', as also observed for the extender domains of *Mp*AFP and *Mh*Lap adhesion proteins [26]. Instead of mechanical stability enhancement by non-covalent bridging across β-strands, the formation of post-translational isopeptide bonds across the terminal strands of Spy0128 [27] and the Cna B domain [28] renders them mechanically inextensible. In addition to topology, hydrophobic contacts across the shearing mechanical interface have also been shown to tailor mechanical stability in protein L [29] and protein GB1 [30].

The B-domains of SasG lack many of the canonical features of 'mechanically strong' proteins: directly hydrogen bonded β-strands are not sheared, there is no typical hydrophobic core, known metal ion coordination or intramolecular covalent crosslinks, suggesting the presence of novel mechanical motif(s). Previous molecular dynamic (MD) forced unfolding simulations suggested, however, that the mechanical strength originates from the previously described mechanical clamps motifs and the associated side-chain packing interactions between the β-strands [10].

Here, we created a pentameric pseudo-wildtype polyprotein of the second B-repeat of SasG (pWT $(E\text{-}G5^2)_5$) that enabled the rapid generation of 18 variants containing single residue substitutions introduced throughout the secondary structural motifs of SasG. AFM force spectroscopy combined with measurement of thermodynamic stability allowed the identification of residues key to the extreme mechanostability of SasG and visualisation of the mechanical unfolding transition state (TS) by ϕ-value analysis. SasG mechanical stability is multi-faceted as

substitutions in the E-G5^2 interface and previously identified 'mechanical clamp' regions significantly destabilised the G5^2 and/or E subdomain. Additionally, we identify a novel mechanical motif – the collagen-like region and show that the same glycine to alanine substitutions that are found in OI also result in a drastically reduced mechanical strength of SasG, directly linking genotypic changes to the mechanical phenotype of a collagen-like structure.

Results

Construction, over-expression and biophysical characterisation of a pentameric E-G5^2 homo-polyprotein

The determinants of protein mechanical strength are usually investigated by comparing changes in unfolding force upon single amino acid substitutions introduced into different sequence or structural motifs [27,29–31]. For SasG, however, the high DNA sequence similarity across the tandem B-repeats within the wild-type SasG sequence from *S. aureus* 8325-4 (~ 97 %) and that of a synonymous sequence for expression in *E. coli* (~ 80 %) precluded this approach.

Previous studies on the wild-type (WT) SasG sequence (G5^1-G5^7, Figure 2A) showed that, in contrast to chemical denaturant induced cooperative (un)folding of B-repeats, the mechanical unfolding of each G5 subdomain was uncoupled from the corresponding E subdomain [10]. Despite this, each E subdomain requires a C-terminal G5 subdomain to form a stable, folded structure [10,32]. Accordingly, a recombinant pentameric homo-polyprotein was designed as a pseudo-WT protein, termed pWT (E-G5^2)$_5$ (Figure 2B). Each repeat within the pentamer comprised the second G5 domain of SasG and its N-terminal E domain (E-G5^2, residues Glu500 to Tyr625 using the numbering system of 3TIP [10]) with each B-repeat (Figure 2C) separated by unstructured linkers [33] (Figure 2B). Following gene construction and protein over-expression (Methods), pWT (E-G5^2)$_5$ was purified using a standard three-step purification procedure (Methods), yielding 2-10 mg of soluble protein per L of culture medium. Full details of the DNA and protein sequences can be found in the Supporting Information. Biophysical analysis indicated pWT (E-G5^2)$_5$ was folded with a secondary structure content similar to SasG (Figure S1). Furthermore, when pWT (E-G5^2)$_5$ was subjected to equilibrium denaturation analysis, a single transition was observed (Figure S2) with identical parameters to monomeric E-G5^2 (Table S1)

[10,32]. This demonstrates that in equilibrium chemical denaturation experiments, the E and $G5^2$ subdomains within each repeat unfold cooperatively, but each E-$G5^2$ repeat unfolds independently (acts as monomeric units).

pWT (E-$G5^2$)$_5$ as a model to investigate SasG mechanical strength

To validate pWT (E-$G5^2$) as a suitable substitute for SasG in SMFS studies, force-extension profiles (obtained as shown schematically in Figure S3) were measured for both pWT (E-$G5^2$)$_5$ (Figure 2D) and native SasG (Figure S4, gene construction, protein over-expression and purification outlined in Supporting Information) at a retraction velocity of 1500 nms^{-1} using the AFM in PBS, pH 7.4 in triplicate for both constructs (data summarised in Tables S2-S5). The 'sawtoothed' force-extension profile for pWT (E-$G5^2$)$_5$ displayed five smaller peaks then five larger peaks, followed by a detachment peak (Figure 2D). The numbers of each subdomain unfolding event reflect the composition of pWT (E-$G5^2$)$_5$ (Figure 2B). Each force-extension profile was described well by the worm-like chain (WLC) model [34,35] allowing the change in contour length (ΔL_C, corresponding to the total length of amino acids 'released' upon unfolding) for each unfolding event to be determined. This gave average values of 147.7 ± 2.9 and 214.7 ± 1.2 Å for the E and $G5^2$ subdomains, respectively, in close agreement to those measured for the E (150.5 ± 3.7 Å) and G5 (218.1 ± 4.3 Å) subdomains of native SasG. Comparison of the unfolding forces for G5 ($F_{U,G5}$) and E ($F_{U,E}$) domains reveal a small but significant increase in $F_{U,G5}$ for pWT (E-$G5^2$)$_5$ relative to SasG (431.5 and 408.4 pN, respectively), yet a small, but significant decrease in $F_{U,E}$ (238.5 and 258.4 pN for pWT (E-$G5^2$)$_5$ and SasG, respectively).

F_U is strongly affected by the polyprotein in which the protein domains under study reside [36]. Firstly, the fewer G5 subdomains in pWT $(E\text{-}G5^2)_5$ relative to SasG (five versus seven) reduces the number of thermally assisted unfolding attempts per unit of time resulting in a greater observed unfolding force. Secondly, the reduced number of domains (either folded or unfolded) reduces the compliance of the polyprotein increasing the effective loading rate, resulting in an increase in F_U. These effects would also be expected to increase $F_{U,E}$ for pWT $(E\text{-}G5^2)_5$ but to a lesser extent (five versus six E subdomains in pWT $(E\text{-}G5^2)_5$ and SasG, respectively). To investigate the origins of the decrease in $F_{U,E}$ for pWT $(E\text{-}G5^2)_5$ relative to SasG, forced unfolding was further performed at retraction velocities of 200, 800, 3000 and 5000 nms^{-1} (Tables S2-S5). The difference in $F_{U,E}$ values for pWT $(E\text{-}G5^2)_5$ and SasG was found to increase upon increasing retraction velocity (Figure S5). The decrease in the speed dependence of $F_{U,E}$ for pWT $(E\text{-}G5^2)_5$ suggests a change in the underlying energy landscape (a small increase in the x_U value or the distance from the native state well to the mechanical TS barrier). This is sometimes interpreted as a more mechanically 'malleable' protein[19]. This effect has been observed previously for constructs of different domain number, for example a recombinant I27 octamer ($I27_8$) displayed a larger x_U value than its recombinant pentameric ($I27_5$) counterpart [37]. Alternatively, the origin of this difference may be due to the loss of G5-E interfaces between each mechanical E-G5 unit. As the mechanical strength of SasG will be delineated by quantifying the effect of amino-acid substitutions in the same pWT $(E\text{-}G5^2)_5$ background, these effects are obviated.

Selection and the structural and mechanical analysis of variants

Guided by previously obtained thermodynamic and kinetic folding data [32], eighteen $E\text{-}G5^2$ variants were designed and assembled into pentamers using Golden Gate cloning (Methods) and

subjected to SMFS to identify the mechanically-stabilising motifs in SasG (Figure 3A). These substitutions were introduced into the E-G5^2 interface; the collagen-like region and the putative 'mechanical clamps' (Figure 2C). After construction, each variant was expressed and purified as described for pWT (E-G5^2)$_5$ and their identity verified by liquid chromatography-mass spectroscopy (LC-MS) (Table S6). The secondary and tertiary structure of each variant was shown to be similar to pWT (E-G5^2)$_5$ using far-UV circular dichroism and fluorescence emission spectroscopy, respectively (Figures S6-S11), with the exception of three of the variants within the 'mechanical clamp' motif of the E (V522P and I502P) and G5^2 subdomain (V550P) (discussed below). Finally, the effects of each substitution on thermodynamic stability was measured by equilibrium denaturation in urea and the resulting fluorescence emission intensity data fit to a two-state model to extract the stability (ΔG_{UN}) and m_{UN} value (related to the solvent accessible surface area upon unfolding [38]) for pWT (E-G5^2)$_5$ and variants for which spectroscopy suggested significantly altered/unfolded structures. The m_{UN} value was fixed at 6 kJ mol^{-1} M^{-1} (the value extracted for pWT (E-G5^2)$_5$) for all other variants (Figure S12 and Table S1).

Forced unfolding experiments were then carried out in triplicate at five retraction velocities to quantify the speed dependence of the F_U values for each variant. This is necessary to identify substitutions which induce significant changes to x_U, characterised by changes in the gradient of a plot of F_U versus log retraction velocity. For each replicate, at least 20 unfolding events for both E and G5^2 subdomains were accumulated (note the average number of unfolding events for E and G5^2 subdomains was 101 ± 45 and 75 ± 32, respectively, Table S13-S46). Example force-extension profiles and F_U versus log retraction velocity plots for a variant from each category are shown in Figure 3B and all variants in Figures S13-19 (these figures also show the location of each variant in SasG). All variants, with the exception of V522P (E-G5^2)$_5$ and V550P (E-G5^2)$_5$ displayed

unfolding spectra comprising five unfolding events for both E and $G5^2$ subdomains and a detachment peak. These two variants displayed a featureless force-extension for ~90 nm (Figure S17), then five unfolding peaks with ΔL_c values identical to that for $G5^2$ subdomain (216.9, 215.4 and 214.7 ± 1.8 Å, for V522P, V550P and pWT $(E\text{-}G5^2)_5$ at 5000 nms^{-1}, respectively). These data, together with changes in their CD spectra and a significant reduction in the thermodynamic parameter m_{UN} to values (4.0 ± 0.1 and 4.1 ± 0.3 kJ mol^{-1} M^{-1}, Table S1) close to that for $G5^2$ in isolation (4.2 ± 0.2 kJ mol^{-1} M^{-1}) [32] suggests that substitution for proline at positions 522 and 550 results in unfolding of the E subdomain. These two variants were not studied further. For all other variants, ΔL_C values (Table S7) were within error to that for pWT $(E\text{-}G5^2)_5$, indicating the pathway to unfolding is coarsely the same to pWT $(E\text{-}G5^2)_5$ as a similar amount of structure is being 'released' after unfolding.

Similar substitutions result in diverse mechanical phenotypes

The aim of this study is to identify the residues and structural motifs that endow SasG mechanical strength. To do this, the average values of $F_{U,E}$ and $F_{U,G5}$ and the speed dependences of variants containing substitutions introduced into each structural motif were calculated. ANCOVA analysis was used to identify variants with a significantly different mechanical phenotype to pWT $(E\text{-}G5^2)_5$ ($p < 0.05$). For those variants that displayed a significant mechanical difference yet a pWT-like speed dependence (and therefore TS placement, characterised by F_U versus log retraction velocity plots parallel to those for pWT $(E\text{-}G5^2)_5$), a threshold difference in F_U of at least 12.0 and 21.5 pN for E and $G5^2$ subdomains (5 % error to take into account the intrinsic cantilever calibration error and number of replicates) was used to identify those residues important for mechanical strength.

E-G5^2 interface variants

The E-G5 interface is important for both the thermodynamic stability and co-operativity between each subdomain [10,32]. To investigate its role in mechanical unfolding where each subdomain is uncoupled, we designed homo-E-G5^2 pentamers containing single amino-acid substitutions in the E subdomain (G517A), the short linker (P549A) and in G5^2 (N598A and T601A) (Figure 3A). Despite the large change in thermodynamic stability ($\Delta\Delta G_{UN}$ = 13.7 $\pm$ 0.8 kJ mol^{-1}) measured for T601A (E-G5^2)$_5$, its mechanical phenotype was identical to pWT (E-G5^2)$_5$ (Figures S12 and S13). Both G517A (E-G5^2)$_5$ and P549A (E-G5^2)$_5$ displayed a significant decrease in the mechanical strength of the E subdomain across all retraction velocities (e.g. ΔF_U = 22.0 $\pm$ 11.2 and 29.0 $\pm$ 13.0 pN at a retraction velocity of 1500 nms^{-1}, respectively) (Figures S13 and S14). P549 is located in the short GP linker connecting E to G5^2 and is part of the pseudo-hydrophobic core interacting with P510 of E and P599 and I605 of G5^2 (Figure S20), also forming a main chain hydrogen bond with G517. Substitutions at these positions may thus affect mechanical strength by changing the packing of the pseudo-hydrophobic core. Finally, a significant increase in the speed dependence of the unfolding force for the E subdomain was observed for N598A (Figure S14). N598 is located in the N-terminal region of the G5^2 subdomain (Figure 3A) yet changes were only observed for the E subdomain. This observation, together with the effect of P549A, suggests that despite their mechanical independence, changes to the E-G5^2 interface affects the mechanical strength of the E subdomain, irrespective of its subdomain location with E-G5^2.

Collagen-like region variants

Next, given the known mechanical function of collagen, we sought to identify the role of the collagen-like region in the mechanical strength of SasG. Substitutions of glycine or proline with

alanine were introduced into either the E (G524A, G527A and P540A) or $G5^2$ subdomains (G584A, G587A and P562A) to affect the packing (Gly to Ala) or to disrupt the PPII-like twists of this region (Pro to Ala). Irrespective of their location, all Gly to Ala variants displayed large decreases in mechanical strength but localised to the subdomain where the substitution was introduced (Figures S15 and S16). For example, a substitution in the E subdomain (G524A) resulted in ΔF_U values of 31.3 ± 10.1 and 3.4 ± 8.2 pN at 1500 nms^{-1} for E and $G5^2$ unfolding events, respectively, while G587A ($G5^2$ subdomain) resulted in $\Delta F_{U,E}$ and $\Delta F_{U,G5}$ values of 1.9 ± 11.2 and 99.7 ± 14.0 pN, respectively. In contrast to these results, the consequences on the mechanical behaviour of proline to alanine substitutions within each subdomain (P540A (E subdomain) and P562A ($G5^2$ subdomain)) were found to be distinct, even though changes in their thermodynamic stability were both small ($\Delta\Delta G_{UN}$ = 0.1 and 2.3 kJ mol^{-1}, respectively). More specifically, the force-extension profiles (Figure S15) and speed dependence of F_U for both E and $G5^2$ subdomains for P540A $(E\text{-}G5^2)_5$ was similar to pWT $(E\text{-}G5^2)_5$ (Figure S16). By contrast, substituting alanine for proline at residue 562 in $G5^2$ resulted in a ~40 pN decrease in $F_{U,G5}$ (with no significant change in $F_{U,E}$) at all retraction velocities (Figures S15 and S16). P540 and P562 are located on distinct strands of the structurally similar E and $G5^2$ subdomains (Figure 3A), suggesting differences in force propagation through each strand of the collagen-like region (see ϕ-value analysis below).

'Mechanical clamp' variants

To assess the role of the mixed β-sheets in SasG mechanical strength, proline substitution and charge reversal mutants were introduced into each subdomain to disrupt the main-chain hydrogen bonding networks and side-chain charge interactions of the 'mechanical clamps' in the E and G5

subdomains previously identified using MD [10] (Figure 3A). The five hydrophobic residue to proline substitutions characterised were I502P and V522P (E subdomain), and V550P, V556P and V580P ($G5^2$). As described above, the E subdomain for V522P and V550P was unfolded at equilibrium (even though residue 550 is in $G5^2$) as assessed by spectroscopic and thermodynamic analysis (Figures S10-S12). Accordingly, only peaks corresponding to $G5^2$ were observed in force-extension profiles for these variants (Figure S17). The ability of V550P to prevent E subdomain folding can be rationalised by the observation that the E subdomain will only fold after the interface between E and $G5^2$ has been formed [32], suggesting that V550P interferes with interface formation. The mechanical phenotype of V556P (Figures S17 and S19) was found not to be statistically significant from pWT $(E\text{-}G5^2)_5$, despite the large change in thermodynamic stability ($\Delta\Delta G_{UN}$ = 14.1 $\pm$ 0.8 kJ mol^{-}1, Table 1).

The remaining two proline substitution variants were found to display different mechanical phenotypes. For I502P (substitution in the E subdomain), $\Delta F_{U,E} \sim 50$ pN at all retraction velocities, with the $G5^2$ subdomain unaffected (Figures S17 and S19). As the spectroscopic evidence (Figure S11) suggests partial unfolding of the E subdomain (with a high probability of this being the 'unlatching' of a β-strand), this large decrease in mechanical strength suggests a role of the 'mechanical clamp' motif in the mechanical strength of the E subdomain. A valine to proline substitution at residue 580 in $G5^2$ resulted in a change to the speed dependence of $F_{U,G5}$ while $F_{U,E}$ was unaffected. (Figures S17 and S19). These data suggest that introduction of proline into the β-strands of the mechanical clamp can result in a perturbed unfolding pathway as observed previously [39]. The increase in the gradient of speed dependence of $F_{U,G5}$ indicates a decrease in x_U contrasting with the effects of valine to proline mutations in β-sheets of other proteins that display an increase in x_U [39].

The mechanical clamp region of $G5^2$ was also probed by charge-swap substitutions E588K, K589E and E624K. As expected for this type of substitution, all three resulted in large reductions in thermodynamic stability (~ 6-18 kJ mol^{-1}, Table 1 and Supplementary Table 1) yet the mechanical unfolding behaviour of E588K was indifferent from pWT $(E\text{-}G5^2)_5$ (Figure S18 and S19). K589E resulted in a large decrease ($\Delta F_{U,G5}$ ~ 70 pN) in mechanical strength across all retraction velocities. As biophysical analysis suggests no difference in the native secondary or tertiary structure (Figures S10 and S11), we can speculate that the side chain interactions of K589 are part of the electrostatic packing of the predicted N-terminal 'Mechanical Clamp' of $G5^2$.

Similarly for the majority of other variants in this study, while large differences were observed for the subdomain hosting the substitution ($G5^2$ in this case, see below), no differences in unfolding behaviour relative to pWT $(E\text{-}G5^2)_5$ was observed for the other subdomain (E in this case), evidencing the robustness of our data accumulation and analysis (Table 1). The E624K $G5^2$ subdomain displayed an intriguing mechanical phenotype. In addition to only the $G5^2$ subdomain displaying a large decrease in mechanical strength and a substantial change in gradient (Figure 3BIII), $F_{U,G5}$ values for each unfolding event were observed to vary significantly both within and between approach-retract cycles (Figure S21). As a consequence of this heterogeneity, the width of $F_{U,G5}$ force-frequency distribution is significantly wider than for pWT $(E\text{-}G5^2)_5$ (Figure S22) (i.e. the average FWHM values are ~116 and ~57 pN for E624K and pWT $(E\text{-}G5^2)_5$, respectively). Additionally, the average ΔL_C distribution was significantly wider for the E624K $G5^2$ subdomain compared to pWT $(E\text{-}G5^2)_5$ (~ 28 vs ~ 15 Å) (Figure S23), indicating there is a larger variety in amount of structure released prior to the detected unfolding peak. These observations offer strong evidence for alternative unfolding pathway(s) for the E624K $G5^2$ [40–43], despite a similar CD spectrum to pWT $(E\text{-}G5^2)_5$ (Figure S10), presumably due to destabilisation of the C-terminal

363 'mechanical clamp' (Figure 2C) by changes to the charge network. Interestingly, Gruszka and
364 colleagues observed that the Y625W variant of monomeric E-G5[2] folded via a distinct pathway
365 [32]. Together with our SMFS data, this suggests that interactions in this region are pivotal for
366 both folding and forced unfolding of G5[2].
367
368 **Revealing the mechanical TS structure through ϕ-value analysis**
369 The data above shows that introduction of amino-acid substitutions at specific locations within
370 all three regions can have profound effects on the mechanical strength of SasG (Figure 4A).
371 Additionally, while some of these exert their effect by destabilising the ground state, as
372 exemplified by G584A ($\Delta F_{U,G5}$ = ~ 40 pN, $\Delta\Delta G_{UN}$ = 17.7 ± 0.9 kJ mol^{-1}), others show similar
373 changes in unfolding force, with small changes in ΔG_{UN} (e.g. P562A $\Delta F_{U,G5}$ = ~ 35 pN, $\Delta\Delta G_{UN}$ =
374 2.3 ± 0.8 kJ mol^{-1}). The latter variant exerts its effect on the TS. By comparing the relative effects
375 of mutation on the ground state and TS, mechanical ϕ-value analysis [40] has allowed visualisation
376 of the mechanically-induced TS of proteins including I27 [44], TNfn3 [45], Protein L [29] and
377 Protein GB1 [30]. As mechanical unfolding experiments probe the local kinetic stability between
378 two defined points, such transition states are usually distinct from those probed by chemical
379 denaturation experiments [44]. In mechanical ϕ-value analysis, ΔF_U values can be used to assess
380 the change in kinetic stability as the change in unfolding rate constant is proportional to the change
381 in F_U relative to the WT if both show similar speed dependences of their unfolding force (i.e.
382 similar placement of the transition state along the unfolding pathway) as other parameters that
383 control loading rate are identical [40,46]. This is discussed further in the Supplemental methods.
384 The ϕ-value is defined as the ratio of change in the height of the TS barrier of mechanical protein
385 unfolding ($\Delta\Delta_{TS-N}^{WT-MUT}$), which is obtained from the measured unfolding force (Methods) to the

386 change in free energy difference between the folded and unfolded state of the protein ($\Delta\Delta G_{UN}$)
387 (measured by equilibrium denaturation) as defined:
388
389 $$\phi = 1 - \left(\frac{\Delta\Delta_{TS-N}^{WT-MUT}}{\Delta\Delta G_{UN}}\right)$$
390
391 These mechanical ϕ-values describe the structural preservation of the residue's local
392 environment at the mechanical TS upon the application of force, with a value of 0 and 1 indicating
393 the interactions to be fully unformed or fully formed, respectively, with partial ϕ-values indicating
394 partial structuring/contacts. To maintain native-like structure, ϕ-value analysis usually measures
395 the effects of conservative hydrophobic volume reduction mutants [47] but as the E and $G5^2$
396 subdomains lack a hydrophobic core and conservative substitution of solvent exposed residues
397 rarely sufficiently destabilises proteins for reliable estimation of ϕ-values, more radical
398 substitutions have been included in this study. This approach has successfully been used to
399 elucidate the "innate" ensemble-derived folding transition state for SasG. This previous study by
400 the Clarke group showed that no intermediate ϕ-values were obtained using similar substitutions
401 to the ones used here, suggesting no major perturbation of the native or unfolded states [32].
402 Additionally, large effects on the folded state of E-$G5^2$ of each substitution was obviated by CD
403 and fluorescence emission spectra that were similar to pWT $(E\text{-}G5^2)_5$ (Figure S6-11).
404 The calculated ϕ-values for all variants with a similar speed dependence of unfolding force to
405 pWT $(E\text{-}G5^2)_5$ are listed in Table 1 and shown schematically in Figure 4B. For each variant, ϕ-
406 values for the subdomain in the E-$G5^2$ repeat without the substitution of ~ 1 demonstrates the
407 quality of these data and the robustness of our approach. Mechanical ϕ-value analysis is inherently
408 low throughput due to both the need to produce polyproteins and obtain statistically robust force

409 unfolding data at a range of retraction velocities. Examination of ensemble-derived ϕ-values shows
410 that these change gradually throughout the transition-state structure [44], suggesting that our
411 sampling (which is the highest reported to date for mechanical analyses) is sufficient to elucidate
412 the coarse TS structure as substitutions have been introduced throughout the structures of both
413 subdomains. The data in Table 1 and shown schematically in Figure 4, reveal a largely native-like
414 TS.

415 Figure 4B suggests that both the E-$G5^2$ interface (G517A and P549A) and the collagen-like
416 regions of both E (G524A and G527A) and $G5^2$ (G584A and G587A) subdomains have undergone
417 structural remodeling at the mechanical TS, yielding ϕ-values of 0.0-0.8 (Table 1). P562A ($G5^2$
418 subdomain) has a ϕ-value of -0.1, suggesting that stabilising non-native contacts are formed at the
419 mechanical TS [49]. This unusual value may, however, arise due to error associated with small
420 $\Delta\Delta G_{UN}$ values [44,48]. In contrast to these regions, several residues in the 'mechanical clamp'
421 regions have ϕ-values ~1, suggesting that they retain native-like contacts at the TS and that they
422 are the main force-bearing regions for SasG.

423

424 **Discussion**

425 Concatenating a single domain of SasG (E-$G5^2$) and its variants into a pentameric homo-
426 polyprotein (pWT $(E\text{-}G5^2)_5$) has allowed the determinants of SasG mechanical strength to be
427 elucidated. While an increase in the mechanical strength of the $G5^2$ subdomain of pWT $(E\text{-}G5^2)_5$
428 relative to SasG could be ascribed to domain number and compliance effects [36], the change in
429 the malleability of the E subdomain was unexpected and is rationalised by the lack of G5-E
430 interface in pWT $(E\text{-}G5^2)_5$. This hypothesis was confirmed by an E-$G5^2$ interfacial variant

(N598A), which not only displayed a change in the malleability of the E subdomain, but also restored the x_U to that observed by native SasG.

Our ϕ-value analysis (Figure 4B) confirms that the 'mechanical clamp' region is native at the TS, consistent with the MD predictions by Gruszka and colleagues [10]. Disrupting the long stretches of hydrogen bonding and sidechain packing interactions in this region resulted in a change in the mechanical phenotype, by decreasing the unfolding force or altering the TS position (or both). The 'mechanical clamp' motif is typically observed in proteins requiring mechanical strength to perform their function such as I27 [39] and C7a [24]. The origin of this resistance to extension is thought to be due to the requirement to simultaneously break all of the non-covalent interactions between the pulling positions. As the majority of proteins are extended via their termini, such shearing generally occurs between directly hydrogen bonded parallel β-strands. As proteins with these structures have complex topology, it has been observed that increasing mechanical strength correlates with increasing relative contact order (RCO) [19,50]. As shown in Supplementary Figure S24, while this correlation generally holds for both subdomains of SasG, the higher mechanical strength of the $G5^2$ subdomain relative to E (~170 pN difference at a retraction speed of 600 nms^{-1}), despite an identical RCO suggests that sequence in addition to topology plays an important role.

The mechanical properties of SasG are further tailored by contacts in the pseudo-hydrophobic core, similar to that observed for the effects of hydrophobic core substitutions in both Protein L [29] and Protein GB1 [30].

Comparing $\Delta\Delta G_{UN}$ and ΔF_U (Figure 4A) shows clearly that force resistance does not always correlate with ΔG_{UN} (exemplified by G527A: $\Delta F_{U,E}$ = ~ 40 pN, $\Delta\Delta G_{UN}$ = 4.9 ± 0.8 kJ mol^{-1} and V556P: $\Delta F_{U,G5}$ = ~ 0 pN, $\Delta\Delta G_{UN}$ = 14.1 ± 0.8 kJ mol^{-1}). The lack of correlation between changes

in these parameters (Figure S25) is unsurprising [29,30,44] as ΔG_{UN} measures the global equilibrium stability and F_U measures the kinetic stability of a localised structure between two defined extension points. This comparison identifies CLM as an important factor in the mechanical strength of SasG as it has a large effect on force, despite a generally moderate effect on ΔG_{UN}. As the ϕ-value analysis suggests gross rearrangement of the CLM at the TS, we utilised MD to investigate in more detail (Methods). These analyses displayed the elongation and untwisting of the collagen-like regions prior to global unfolding (Methods, Figure S26, Movie S1), which may bestow subtle flexibility of the collagen-like regions. The collagen-like regions may act as an intrinsic molecular spring when there are several in tandem, permitting dissipation of force and regulating the ability of the bacterial cell to sustain attachment under shear force by preventing the global unfolding of the E and $G5^2$ subdomains, in accord with the low (0) and partial ϕ-values observed for residues 524 and 527 (in E) and 584 and 587 (in $G5^2$).

G524/G527 (E subdomain) and G584/G587 ($G5^2$ subdomain) pairs are structurally equivalent glycine residues located in one of the three left-handed polyproline PPII-like helices following the canonical collagen sequence of $X_{aa}Y_{aa}$Gly with glycine required to accommodate the packing of the sidechains at the central axis [51]. Here we have shown that substitution of these glycines with residues with bulkier side-chains results in a large decrease in F_U for SasG. This is probably due to the structural perturbation that arises from over-packing (Figure S27). This type of substitution is also found in OI - a rare genetic disease (brittle bone disease) affecting the connective tissues characterised by extremely fragile bones that break or fracture easily. Structural and MD studies on collagen-like peptides show that accommodation of a methyl side-chain results in local twist relaxation and bulging due to loss of hydrogen bonding and electrostatic repulsions (Figure S27) [52,53]. In line with this observation, spectroscopic analysis of our variants displays a subtle

deviation in secondary structure (Figure S8) indicating similar structural perturbation of our collagen-like regions. In summary, OI-like substitutions significantly decrease the mechanical strength of the E and $G5^2$ subdomains, identifying the collagen-like region as a novel mechanical motif. We speculate that the changes in mechanical phenotype uncovered for SasG, also affects the mechanical properties of collagen fibres found in OI patients.

Materials and Methods

Homo-polyprotein construction by Golden Gate assembly

Polymerase chain reactions (PCR) were performed with Q5® DNA polymerase (NEB, Hertfordshire, UK). All mutagenic, cassette amplification and destination vector linearisation primers are found in Tables S8-S11. Briefly, a gene encoding a monomer of E-$G5^2$ (using a *E. coli* optimised native SasG sequence as template, provided by Professor Jennifer Potts, University of Sydney, Australia) was inserted into a linearised pET14b vector using Gibson Assembly® [54]. This WT E-$G5^2$ monomer sequence was then used as template to make pentameric polyproteins of the wild-type and variant sequences by Golden Gate assembly. Firstly, two BsaI restriction sites were removed from the destination pET14b vector by PCR to create pET14b$\Delta bsaI$. For each construct, five DNA cassettes were generated by PCR that encoded one E-$G5^2$ repeat and part of the unstructured linkers separating each repeat together with BsaI sites (at the DNA level) for assembly. The purified cassettes and PCR-linearised pET14b$\Delta bsaI$ were then assembled using the NEB Golden Gate assembly kit (BsaI-HF®v2) following the manufacturers protocol, using 75 ng of linearised destination vector and a two-fold molar excess of cassettes. The reaction mixture was thermo-cycled (T100™ thermal cycler, BioRad, California, USA) at 37 ˚C for one minute and then

16 °C for one minute for a total of 30 cycles. Subsequently, the mixture was then held at 60 °C for five minutes. 2 µl of the product was transformed into SURE2 competent *E. coli* cells (Agilent, California, USA). The sequence of each assembled variant was confirmed by Sanger Sequencing (Twist Bioscience, Cambridge, UK). The final construct (and variants thereof) comprises five identical domains of E-G5^2 separated by unstructured linkers and two C-terminal cysteine residues for protein immobilisation to gold substrate for SMFS experiments:

MSSHHHHHHSS(E-G5^2)LSVGATI(E-G5^2)TVIGLAS(E-G5^2)ALSGTIV(E-G5^2)VITGSLA(E-G5^2)CC

The full protein sequence of pWT (E-G5^2)$_5$ and variants thereof are found in Supplementary Information).

Protein over-expression and purification

Plasmids encoding the homo-polyprotein constructs were transformed into *E. coli* BL21 (DE3) pLysS (Agilent, California, USA). 10 × 0.5 L of autoinduction medium was prepared containing 1 % (w/v) yeast extract, 2 % (w/v) bactotryptone, 0.01 % (w/v) glycerol, 0.001 % (w/v) D-glucose anhydrous, 0.004 % (w/v) α-Lactose monohydrate, 50 mM ammonium chloride, 5 mM sodium sulphate, 25 mM potassium phosphate monobasic, 25 mM sodium phosphate dibasic, 100 µg/ml ampicillin and 25 µg/ml chloramphenicol) was inoculated with overnight starter culture (LB medium containing 100 µg/ml ampicillin and 25 µg/ml chloramphenicol) and protein expression carried out at 28 °C, 200 rpm as described [55]. After 24 hours, the cells were harvested and the cell pellet resuspended in lysis buffer (20 mM Na_2HPO_4/ NaH_2PO_4, 8 M urea, 500 mM NaCl, 10

mM imidazole, 2 mM DTT, 1mM PMSF, 2 mM benzamidine and a rice grain of lysozyme and DNaseI). After cell disruption by sonication the cleared lysate was applied to a lab-packed 200 ml pre-charged Ni Sepharose™ Fast Flow column (Cytiva, Massachusetts, USA). Wash/refolding buffer (20 mM Na_2HPO_4-NaH_2PO_4, 500 mM NaCl, 10 mM imidazole, 2 mM DTT, 1mM PMSF, 2 mM benzamidine) was applied until baseline A_{280} was achieved and the protein eluted using wash buffer supplemented with 150 mM imidazole. The protein of interest was dialysed into 20 mM Tris-HCl, 50 mM NaCl, 1 mM EDTA and 2 mM DTT for 16 hours. This was loaded onto 2× 5 ml HiTrap SP HP columns (Cytiva) stacked atop of 4× 5 ml HiTrap Q HP columns (Cytiva). After removal of the HiTrap SP HP columns, the protein was eluted with a linear 50-185 mM NaCl gradient of 7.5 column volumes. Subsequent purification was performed with gel filtration using a 320 ml Hiload Superdex™ 26/600 75 column (Cytiva) in 1× PBS, pH 7.4. Successful isolation was confirmed by LC-MS (Table S6). The recombinant protein was snap frozen in liquid nitrogen and stored at -80 °C.

Circular Dichroism spectroscopy

200 µl of 0.2 mg/ml protein solution (5 mM NaH_2PO_4, pH 7.4) was transferred to a 1 mm path-length cuvette (Hellma, Müllheim, Germany). A far-UV spectrum (180-260 nm) was obtained using a 2 nm bandwidth, 1 s time step (25 °C) with an average of three scans per sample on a Chirascan CD spectrometer (Applied Photophysics®, Leatherhead, UK).

Single-Molecule Force spectroscopy

Single-molecule force spectroscopy experiments were carried out for the native SasG construct ($G5^1$-$G5^7$), pWT $(E\text{-}G5^2)_5$ and variants as described previously[10]. Briefly, an MFP-3D™ Stand

Alone AFM (Asylum Research, Buckinghamshire, UK) was mounted with a silicon nitride cantilever (MLCT, Bruker, California, USA). The deflection sensitivity and natural resonance of the cantilever were found, and the spring constant calculated using the equi-partition theorem [56]. 100 μl 250 μg/ml of protein in PBS, pH 7.4 was deposited on a freshly cleaved gold substrate [10], resulting in covalent attachment between the C-terminal cysteine residues and the gold atoms of the surface. After incubation for ~ 15 mins, the surface was washed with PBS, pH 7.4 and allowed to equilibrate for another 15 minutes prior to accumulating force-extension data. Mechanical unfolding experiments were carried out at retraction velocities of 200, 800, 1500, 3000 and 5000 nms^{-1} at room temperature over a distance of 600 nm, with the exception of V522P $(E\text{-}G5^2)_5$ and V550P $(E\text{-}G5^2)_5$, which were carried out at 200 and 5000 nms^{-1}. Three separate cantilevers were used for each construct, with the exception of V522P $(E\text{-}G5^2)_5$ and V550P $(E\text{-}G5^2)_5$ (n=1), to decrease systematic errors from cantilever spring constant calculation.

Analysis of Single-Molecule Force spectroscopy data

Data were filtered to include only traces that display the unfolding of four or more subdomains and a detachment peak. All mechanical unfolding data were processed using IGOR pro (version 6.37, Wavemetrics, Oregon, USA) with an Asylum Research extension (MFP3DXop v30). After filtering, spectra were imported into Fodis [57]. The apex of each peak was taken as the observed unfolding force and L_C values obtained by fitting a worm-like chain (WLC) model for polymer elasticity [35]:

$$F(x) = \frac{k_B T}{p}\left(0.25\left(1-\frac{x}{L_C}\right)^{-2} - 0.25 + \frac{x}{L_C}\right)$$

Where L_C is the contour length (maximum predicted extension of the polypeptide chain under external force), $F(x)$ is the force as a function of extension, x. k_B is the Boltzmann's constant, T is temperature and a persistence length (p) of 3.8 Å. Both the unfolding forces and ΔL_C were binned and plotted in histograms, with bin sizes of 10 pN and 5 Å, respectively. Subsequent Gaussian fits to these histograms were used to obtain median values from each cantilever repeat, which were averaged to yield the mean unfolding force and ΔL_C value.

Equilibrium denaturation

Fluorescence spectroscopy was performed on a Photon Technology International spectrofluorometer (PTI, New Jersey, USA) at 25 ˚C using a 1 ml quartz Hellma cuvette (Hellma, Müllheim, Germany). For equilibrium denaturation of pWT $(E\text{-}G5^2)_5$ and variants thereof, PBS pH 7.4 and PBS, 9 M urea pH 7.4 were prepared alongside 1 mg/ml of the protein being analysed in PBS pH 7.4. Five stocks of 0.2 mg/ml containing 0, 2, 4, 6 and 8 M urea were prepared and mixed to create protein samples increasing in urea concentration in 0.2 M increments from 0 to 8 M. After equilibration for 16 hours at 25 °C in a circulating water bath (NesLab, Massachusetts, USA), the samples were transferred to the fluorimeter and after five minutes of equilibration, the sample was excited at 276 nm and a time drive scan was taken at 305 nm for 30 seconds (1 second per data point) at 25 °C. The average signal was calculated for each concentration and plot as a function of urea concentration. A standard two-state transition chemical denaturant model [58] was fit to the data using Igor Pro (version 7.02, Wavemetrics, Oregon, USA) to obtain thermodynamic fitting values. For all variants (with the exception of pWT $(E\text{-}G5^2)$, I502P, V522P, and V550P), the m_{UN} value was fixed to 6 kJ mol^{-1} M^{-1} (the value extracted for pWT $(E\text{-}G5^2)_5$).

590 The raw data was normalised to a fraction of natively folded protein to allow qualitative
591 comparison between pWT (E-G5^2)$_5$ and variants thereof.
592
593 **Mechanical ϕ-value analysis**
594 $\Delta\Delta G_{UN}$ values are the difference of the thermodynamic stability (ΔG_{UN}) between each variant
595 and pWT (E-G5^2)$_5$. The $\Delta\Delta G_{TS-N}^{WT-MUT}$ values of variants were calculated directly from the speed
596 dependence as described previously [40]:
597
598 $$\Delta\Delta G_{TS-N}^{WT-MUT} = RT(f_{WT} - f_{MUT})/m$$
599
600 Where f_{WT} and f_{MUT} are the average unfolding forces of pWT (E-G5^2)$_5$ and the variant being
601 measured, respectively. R is the gas constant, T is temperature and m is the average gradient of
602 the weighted best-fits to the speed dependence of unfolding force for those variants and
603 subdomains within error (of the gradient value) of pWT (E-G5^2)$_5$ at a retraction velocity of 1500
604 nms^{-1}. Errors were propagated using an error of 5 % (lower confidence limit due to inherent
605 cantilever calibration systematic error) and errors on $\Delta\Delta G_{UN}$ were the propagated errors of the fits.
606 ϕ-value can then be calculated:
607
608 $$\phi = 1 - \frac{\Delta\Delta G_{TS-N}^{WT-MUT}}{\Delta\Delta G_{UN}}$$
609
610 **E-G5^2 forced unfolding simulations**
611 Simulations were carried out using an all atom force field (CHARMM22) and fast analytical
612 continuum treatment of solvation (FACTS) implicit solvent model. Forced unfolding was

simulated by attaching an ideal spring to the N and C atoms of the two termini and retracting them at a constant velocity (in the range 10^5-10^7 nms^{-1}) at 300 K and 0 K (local minimum energy pathway). The main features of the unfolding pathways does not depend on the pulling speed or the temperature.

Supplementary Information

All analysed spectroscopic, equilibrium denaturation and SMFS data (raw data tables also included). pWT $(E\text{-}G5^2)_5$ DNA sequence and protein sequence (and variants thereof) included.

ACKNOWLEDGMENTS

The authors would like to thank Professor Jennifer Potts for a gift of the *E. coli* codon-optimised native SasG DNA and members of the Radford, Brockwell, and Calabrese groups for helpful discussions and advice throughout. This work was supported by a White Rose Research Studentship (White Rose University Consortium) to A.J.E.B. We thank the Wellcome Trust (094232) and University of Leeds for the purchase of the Chiroscan CD spectrometer.

REFERENCES

[1] J.W. Costerton, Z. Lewandowski, D.E. Caldwell, D.R. Korber, H.M. Lappin-Scott, Microbial Biofilms., Annu. Rev. Microbiol. 49 (1995) 711–745.

[2] M. Otto, Physical stress and bacterial colonisation, FEMS Microbiol. Rev. 38 (2014) 1250–1270.

[3] P. Herman, S. El-Kirat-Chatel, A. Beaussart, J.A. Geoghegan, T.J. Foster, Y.F. Dufrêne, The binding force of the staphylococcal adhesin SdrG is remarkably strong, Mol. Microbiol. 93 (2014) 356–368.

[4] P. Herman-Bausier, C. Labate, A.M. Towell, S. Derclaye, J.A. Geoghegan, Y.F. Dufrêne, Staphylococcus aureus clumping factor A is a force-sensitive molecular switch that activates bacterial adhesion, Proc. Natl. Acad. Sci. U.S.A. 115 (2018) 5564–5569.

[5] M. Mathelié-Guinlet, F. Viela, G. Pietrocola, P. Speziale, D. Alsteens, Y.F. Dufrêne, Force-clamp spectroscopy identifies a catch bond mechanism in a Gram-positive pathogen, Nat. Commun. 11 (2020) 5431.

[6] M. Grandbois, M. Beyer, M. Rief, H. Clausen-Schaumann, H.E. Gaub, How strong is a covalent bond, Science. 283 (1999) 1727–1730.

[7] E. Miller, T. Garcia, S. Hultgren, A.F. Oberhauser, The mechanical properties of *E. coli* type 1 pili measured by Atomic Force Microscopy techniques, Biophys. J. 91 (2006) 3848–3856.

[8] F. Whelan, A. Lafita, J. Gilburt, C. Dégut, S.C. Griffiths, H.T. Jenkins, A.N. St John, E. Paci, J.W.B. Moir, M.J. Plevin, C.G. Baumann, A. Bateman, J.R. Potts, Periscope Proteins are variable-length regulators of bacterial cell surface interactions, Proc. Natl. Acad. Sci. U.S.A. 118 (2021) e2101349118.

[9] C. Chantraine, M. Mathelié-Guinlet, G. Pietrocola, P. Speziale, Y.F. Dufrêne, AFM identifies a protein complex involved in pathogen adhesion which ruptures at three nanonewtons, Nano Lett. 21 (2021) 7595–7601.

[10] D.T. Gruszka, F. Whelan, O.E. Farrance, H.K.H. Fung, E. Paci, C.M. Jeffries, D.I. Svergun, C. Baldock, C.G. Baumann, D.J. Brockwell, J.R. Potts, J. Clarke, Cooperative folding of intrinsically disordered domains drives assembly of a strong elongated protein, Nat. Commun. 6 (2015) 7271.

[11] R.M. Corrigan, D. Rigby, P. Handley, T.J. Foster, The role of *Staphylococcus aureus* surface protein SasG in adherence and biofilm formation, Microbiology (Reading). 153 (2007) 2435–2446.

[12] F.M. Roche, M. Meehan, T.J. Foster, The *Staphylococcus aureus* surface protein SasG and its homologues promote bacterial adherence to human desquamated nasal epithelial cells, Microbiology (Reading). 149 (2003) 2759–2767.

[13] C. Formosa-Dague, P. Speziale, T.J. Foster, J.A. Geoghegan, Y.F. Dufrêne, Zinc-dependent mechanical properties of Staphylococcus aureus biofilm-forming surface protein SasG, Proc. Natl. Acad. Sci. U.S.A. 113 (2016) 410–415.

[14] D.G. Conrady, C.C. Brescia, K. Horii, A.A. Weiss, D.J. Hassett, A.B. Herr, A zinc-dependent adhesion module is responsible for intercellular adhesion in staphylococcal biofilms, Proc. Natl. Acad. Sci. U.S.A. 105 (2008) 19456–19461.

673 [15] F.M. Roche, R. Massey, S.J. Peacock, N.P.J. Day, L. Visai, P. Speziale, A. Lam, M. Pallen,
674 T.J. Foster, Characterization of novel LPXTG-containing proteins of *Staphylococcus*
675 *aureus* identified from genome sequences, Microbiology (Reading). 149 (2003) 643-654.

676 [16] D.T. Gruszka, J.A. Wojdyla, R.J. Bingham, J.P. Turkenburg, I.W. Manfield, A. Steward,
677 A.P. Leech, J.A. Geoghegan, T.J. Foster, J. Clarke, J.R. Potts, Staphylococcal biofilm-
678 forming protein has a contiguous rod-like structure, Proc. Natl. Acad. Sci. U.S.A. 109
679 (2012).

680 [17] R.Z. Kramer, J. Bella, P. Mayville, B. Brodsky, H.M. Berman, Sequence dependent
681 conformational variations of collagen triple-helical structure, Nat. Struct. Biol. 6 (1999)
682 454–457.

683 [18] L. Bozec, M. Horton, Topography and mechanical properties of single molecules of type I
684 collagen using Atomic Force Microscopy, Biophys. J. 88 (2005) 4223–4231.

685 [19] T. Hoffmann, K.M. Tych, M.L. Hughes, D.J. Brockwell, L. Dougan, Towards design
686 principles for determining the mechanical stability of proteins, Phys. Chem. Chem. Phys.
687 15 (2013) 15767–15780.

688 [20] N. Crampton, D.J. Brockwell, Unravelling the design principles for single protein
689 mechanical strength., Curr. Opin. Struct. Biol. 20 (2010) 508–517.

690 [21] M.L. Hughes, L. Dougan, The physics of pulling polyproteins: A review of single molecule
691 force spectroscopy using the AFM to study protein unfolding, Rep. Prog. Phys. 79 (2016)
692 076601.

693 [22] M. Sikora, J.I. Sułkowska, B.S. Witkowski, M. Cieplak, BSDB: the biomolecule stretching
694 database, Nucleic Acids Res. 39 (2011) D443–D450.

695 [23] M. Sikora, J.I. Sułkowska, M. Cieplak, Mechanical strength of 17,134 model proteins and
696 cysteine slipknots, PLOS Comput. Biol. 5 (2009) e1000547.

697 [24] A. Valbuena, J. Oroz, R. Hervás, A.M. Vera, D. Rodríguez, M. Menéndez, J.I. Sulkowska,
698 M. Cieplak, M. Carrión-Vázquez, On the remarkable mechanostability of scaffoldins and
699 the mechanical clamp motif, Proc. Natl. Acad. Sci. U.S.A. 106 (2009) 13791–13796.

700 [25] L.F. Milles, E.M. Unterauer, T. Nicolaus, H.E. Gaub, Calcium stabilizes the strongest
701 protein fold, Nat. Commun. 9 (2018) 4764.

702 [26] A.S. Oude Vrielink, T.D.R. Vance, A.M. de Jong, P.L. Davies, I.K. Voets, Unusually high
703 mechanical stability of bacterial adhesin extender domains having calcium clamps, PLoS
704 One. 12 (2017) e0174682.

705 [27] B. Wang, S. Xiao, S.A. Edwards, F. Gräter, Isopeptide bonds mechanically stabilize
706 spy0128 in bacterial pili, Biophys. J. 104 (2013) 2051–2057.

707 [28] P. Herman-Bausier, C. Valotteau, G. Pietrocola, S. Rindi, D. Alsteens, T.J. Foster, P.
708 Speziale, Y.F. Dufrêne, Mechanical strength and inhibition of the *Staphylococcus aureus*
709 collagen-binding protein Cna, MBio. 7 (2016) e01529.

[29] D.P. Sadler, E. Petrik, Y. Taniguchi, J.R. Pullen, M. Kawakami, S.E. Radford, D.J. Brockwell, Identification of a mechanical rheostat in the hydrophobic core of protein L, J. Mol. Biol. 393 (2009) 237–248.

[30] T. Bu, H.C.E. Wang, H. Li, Single molecule force spectroscopy reveals critical roles of hydrophobic core packing in determining the mechanical stability of protein GB1, Langmuir. 28 (2012) 12319–12325.

[31] M. Carrión-Vázquez, A.F. Oberhauser, T.E. Fisher, P.E. Marszalek, H. Li, J.M. Fernandez, Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering, Prog. Biophys. Mol. Biol. 74 (2000) 63–91.

[32] D.T. Gruszka, C.A.T.F. Mendonça, E. Paci, F. Whelan, J. Hawkhead, J.R. Potts, J. Clarke, Disorder drives cooperative folding in a multidomain protein, Proc. Natl. Acad. Sci. U.S.A. 113 (2016) 11841–11846.

[33] T. Hoffmann, K.M. Tych, T. Crosskey, B. Schiffrin, D.J. Brockwell, L. Dougan, Rapid and robust polyprotein production facilitates single-molecule mechanical characterization of β-Barrel Assembly Machinery Polypeptide Transport Associated domains, ACS Nano. 9 (2015) 8811–8821.

[34] C. Bustamante, J.F. Marko, E.D. Siggia, S. Smith, Entropic elasticity of λ-phage DNA, Science. 265 (1994) 1599–1600.

[35] J.F. Marko, E.D. Siggia, Stretching DNA, Macromolecules. 28 (1995) 8759–8770.

[36] R.C. Zinober, D.J. Brockwell, G.S. Beddard, A.W. Blake, P.D. Olmsted, S.E. Radford, D.A. Smith, Mechanically unfolding proteins: The effect of unfolding history and the supramolecular scaffold, Protein Sci. 11 (2002) 2759–2765.

[37] D.J. Brockwell, G.S. Beddard, J. Clarkson, R.C. Zinober, A.W. Blake, J. Trinick, P.D. Olmsted, D.A. Smith, S.E. Radford, The effect of core destabilization on the mechanical resistance of I27, Biophys. J. 83 (2002) 458–472.

[38] J.K. Myers, C.N. Pace, J.M. Scholtz, Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding, Protein Sci. 4 (1995) 2138–2148.

[39] H. Li, M. Carrión-Vázquez, A.F. Oberhauser, P.E. Marszalek, J.M. Fernandez, Point mutations alter the mechanical stability of immunoglobulin modules, Nat. Struct. Biol. 7 (2000) 1117–1120.

[40] R.B. Best, S.B. Fowler, J.L. Toca-Herrera, J. Clarke, A simple method for probing the mechanical unfolding pathway of proteins in detail, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 12143–12148.

[41] E. Evans, Probing the relation between force--lifetime--and chemistry in single molecular bonds., Annu. Rev. Biophys. Biomol. Struct. 30 (2001) 105–128.

[42] K.L. Anderson, S.E. Radford, D.A. Smith, D.J. Brockwell, The dynamical response of proteins under force, in: Handbook of Molecular Force Spectroscopy, 2008: pp. 202–246.

[43] M. Schlierf, M. Rief, Temperature softening of a protein in single-molecule experiments, J. Mol. Biol. 354 (2005) 497–503.

[44] R.B. Best, S.B. Fowler, J.L.T. Herrera, A. Steward, E. Paci, J. Clarke, Mechanical unfolding of a titin Ig domain: structure of transition state revealed by combining atomic force microscopy, protein engineering and molecular dynamics simulations., J. Mol. Biol. 330 (2003) 867–877.

[45] S.P. Ng, R.W.S. Rounsevell, A. Steward, C.D. Geierhaas, P.M. Williams, E. Paci, J. Clarke, Mechanical unfolding of TNfn3: The unfolding pathway of a FnIII domain probed by protein engineering, AFM and MD simulation, J. Mol. Biol. 350 (2005) 776–789.

[46] R. Rounsevell, J.R. Forman, J. Clarke, Atomic force microscopy: Mechanical unfolding of proteins, Methods. 34 (2004) 100–111.

[47] A.R. Fersht, A. Matouschek, L. Serrano, The folding of an enzyme: I. Theory of protein engineering analysis of stability and pathway of protein folding, J. Mol. Biol. 224 (1992) 771–782.

[48] I.E. Sánchez, T. Kiefhaber, Origin of unusual φ-values in protein folding: evidence against specific nucleation sites, J. Mol. Biol. 334 (2003) 1077–1085.

[49] A. Zarrine-Afsar, S. Wallin, A.M. Neculai, P. Neudecker, P.L. Howell, A.R. Davidson, H.S. Chan, Theoretical and experimental demonstration of the importance of specific nonnative interactions in protein folding, Proc. Natl. Acad. Sci. U.S.A. 105 (2008) 9999–10004.

[50] G.H.M. Huysmans, S.A. Baldwin, D.J. Brockwell, S.E. Radford, The transition state for folding of an outer membrane protein, Proc. Natl. Acad. Sci. U.S.A. 107 (2010) 4099–4104.

[51] M.D. Shoulders, R.T. Raines, Collagen structure and stability, Annu. Rev. Biochem. 78 (2009) 929–958.

[52] J. Bella, M. Eaton, B. Brodsky, H.M. Berman, Crystal and molecular structure, of a collagen-like peptide at 1.9 Å resolution, Science. 266 (1994) 75–81.

[53] V. Punitha, S.S. Raman, R. Parthasarathi, V. Subramanian, J.R. Rao, B.U. Nair, T. Ramasami, Molecular dynamics investigations on the effect of D amino acid substitution in a triple-helix structure and the stability of collagen, J. Phys. Chem. B. 113 (2009) 8983–8992.

[54] D.G. Gibson, L. Young, R.-Y. Chuang, J.C. Venter, C.A. Hutchison, H.O. Smith, Enzymatic assembly of DNA molecules up to several hundred kilobases, Nat. Methods. 6 (2009) 343–345.

[55] F.W. Studier, Protein production by auto-induction in high density shaking cultures, Protein. Expr. Purif. 41 (2005) 207–234.

[56] J.L. Hutter, J. Bechhoefer, Calibration of atomic-force microscope tips, Rev. Sci. Instrum. 64 (1993) 1868.

[57] N. Galvanetto, A. Perissinotto, A. Pedroni, V. Torre, Fodis: software for protein unfolding analysis, Biophys. J. 114 (2018) 1264–1266.

[58] M.M. Santoro, D.W. Bolen, Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl α-chymotrypsin using different denaturants, Biochemistry. 27 (1988) 8063–8068.

Tables

Group	Variant	Subdomain Location of mutation	**ΔΔ**G_{UN} (kJ mol^{-1})[a]	**Δ**$F_{U,E}$ [b]	**Δ**$F_{U,G5}$ [b]	ϕ_E	ϕ_{G5}
I	G517A	E	5.6 ± 0.8	**22.0 ± 11.2**	5.5 ± 17.2	**0.5 ± 0.1**	0.9 ± 0.2
	P549A	GP Linker	3.8 ± 0.8	29 ± 13.0	-11.8 ± 16.3	0.1 ± 0.1	1.2 ± 0.3
	N598A	$G5^2$	16.7 ± 0.9	22.5 ± 6.1	**8.1 ± 11.9**	ND	**1.0 ± 0.1**
	T601A	$G5^2$	13.7 ± 0.8	-6.1 ± 9.2	**11.6 ± 14.0**	1.1 ± 0.2	**0.9 ± 0.1**
CLM	G524A	E	6.4 ± 0.8	**31.3 ± 10.1**	3.4 ± 8.2	**0.4 ± 0.1**	1.0 ± 0.1
	G527A	E	4.9 ± 0.8	**41.6 ± 11.5**	-3.5 ± 17.3	**0.0 ± 0.0**	1.1 ± 0.2
	P540A[c]	E	0.1 ± 0.8	**-6.1 ± 10.0**	-7.5 ± 8.9	ND	ND
	P562A[d]	$G5^2$	2.3 ± 0.8	-1.9 ± 5.2	**34.7 ± 6.4**	1.1 ± 0.4	**-0.1 ± 0.0**
	G584A	$G5^2$	17.7 ± 0.9	-4.1 ± 7.6	**42.6 ± 9.7**	1.0 ± 0.1	**0.8 ± 0.1**
	G587A	$G5^2$	20.7 ± 0.9	1.9 ± 9.1	**99.7 ± 14.0**	1.0 ± 0.1	**0.7 ± 0.1**
MC	I502P	E	11.2 ± 1.1	**50.3 ± 7.0**	3.7 ± 7.8	**0.5 ± 0.1**	1.0 ± 0.1
	V522P[e]	E	15.6 ± 0.9	-	-	ND	ND
	V550P[e]	$G5^2$	17.7 ± 0.9	-	-	ND	**1.0**[f]
	V556P	$G5^2$	14.1 ± 0.8	8 ± 14.6	**0.6 ± 9.3**	0.9 ± 0.1	**1.0 ± 0.1**
	V580P	$G5^2$	12.7 ± 0.8	-5.3 ± 11.6	**20.8 ± 13.2**	1.0 ± 0.2	ND

	E588K	$G5^2$	5.7 ± 0.8	-4.1 ± 10.8	**-0.3 ± 8.5**	1.1 ± 0.2	**1.0 ± 0.2**
	K589E	$G5^2$	16.8 ± 0.9	-3.8 ± 6.7	**73.8 ± 15.8**	1.0 ± 0.1	**0.7 ± 0.1**
	E624K	$G5^2$	17.7 ± 0.9	-0.8 ± 6.5	**99.2 ± 11.4**	1.0 ± 0.1	ND

Table 1. Thermodynamic and mechanical stability of variants and their mechanical ϕ-values. Mechanical ϕ-values describing the TS of the $(E\text{-}G5^2)_5$ variants in this study with a pWT-like speed dependence. [a]$\Delta\Delta G_{UN} = \Delta G_{UN,pWT} - \Delta G_{UN,Mut}$ (± propagated error of fit). [b]$\Delta F_U = F_{U,pWT} - F_{U,Mut}$ at 1500 nms^{-1} (± propagated standard deviation). [c]ϕ-value was not calculated due to small change in $\Delta\Delta G_{UN}$. [d]P562A $(E\text{-}G5^2)_5$ has a small $\Delta\Delta G_{UN}$ value, which can cause artefactual ϕ-values [44,48], so this should be analysed with caution. [e]V522P and V550P were only unfolded at retraction velocities of 200 and 5000 nms^{-1}, so the accurate unfolding force at 1500 nms^{-1} is unknown. [f]As V550P $(E\text{-}G5^2)_5$ displayed mechanical unfolding forces of the $G5^2$ subdomain indifferent from pWT $(E\text{-}G5^2)_5$ at unfolding forces of 200 and 5000 nms^{-1} (Table S13), we can assume the distance to TS is the same and as there is no change in the activation energy to unfolding (unperturbed mechanical strength), we can estimate the ϕ-value is ~ 1. Bolded values are those calculated for the subdomain in which the substitution is located. Values in red are partial ϕ-values and indicate structural perturbation/loss of contacts in the mechanical unfolding TS. As the E subdomain of N598A and $G5^2$ subdomain of V580P and E624K displayed a change in their speed dependence gradient, the ϕ-value was not calculated. I: interface, CLM: collagen-like motif, MC: 'mechanical clamp', and ND: not done. All errors are propagated.

Figure Legends

Figure 1. Domain arrangement of full length SasG and structural details of a B-repeat domain. **A)** Schematic of full-length SasG. S: Signal sequence, A: A domain, W: Wall-spanning region and SS: Sorting signal. A single B-repeat is highlighted. **B)** $E\text{-}G5^2$ crystal structure (PDB: 3TIP) displaying the structural similarity between the subdomains E and $G5^2$. E and $G5^2$ shown in blue and grey, respectively. The collagen-like motifs (CLM) of the E and $G5^2$ subdomains are highlighted with a red circle. **C)** E collagen-like region (from PDB: 3TIP, coloured blue) aligned with collagen (PDB: 1BKV, coloured purple) to display the collagen-like homology and the staggering of 1/3 strands.

Figure 2. The mechanical properties of pWT $(E\text{-}G5^2)_5$. **A)** Modelled structure of WT SasG $G5^1$-$G5^7$ **(**taken from reference [10]) with $E\text{-}G5^2$ subdomains used in this study highlighted. **B)**

820 Schematic of pWT $(E\text{-}G5^2)_5$ with the E and $G5^2$ subdomains highlighted in blue and grey,
821 respectively. Linkers (purple), His tag (H_6, orange) and the C-terminal double cysteine (CC,
822 yellow) are displayed as blocks. **C)** Topology diagram of $E\text{-}G5^2$ highlighting boundary residues
823 for E and $G5^2$, with GP linker residues highlighted in green. Putative 'mechanical clamp' regions
824 are displayed as blue and grey boxes. **D)** Three example pWT $(E\text{-}G5^2)_5$ 'sawtooth' force-extension
825 profiles (black) displaying five unfolding events of both E and $G5^2$ subdomains at a retraction
826 velocity of 1500 nms^{-1}. Solid lines: WLC model fitted to E and $G5^2$ subdomains (blue and grey,
827 respectively), and the detachment peak (purple). The average unfolding forces for pWT $(E\text{-}G5^2)_5$
828 E and $G5^2$ at a retraction velocity of 1500 nms^{-1} are shown as blue and grey dashed lines,
829 respectively. **E)** Scatterplot with associated histograms reveals a bimodal distribution in both
830 rupture force and ΔL_C values. E and $G5^2$ Gaussian fits in black solid and dashed lines, respectively.
831 Black scatterplot crosshairs are the mode and the FWHM (full width at half maximum) from the
832 corresponding histogram Gaussian model fits, and red scatterplot crosshairs are the mode and the
833 FWHM from the corresponding histogram Gaussian model fits (not shown here) from an example
834 replicate of SasG at the equivalent retraction velocity.

835

836 **Figure 3.** Variant location and example SMFS data. **A)** Topology diagram of $E\text{-}G5^2$ showing
837 location of each substitution labelled according to the type of mechanical motif: green (interface
838 residue), blue (collagen-like region) or purple ('mechanical clamp'). The putative 'Mechanical
839 clamp' regions are shaded in blue or grey. **B)** Example data for interface (P549A), collagen-like
840 motif (G587A) and 'Mechanical Clamp' (E624K) variants. **I)** Force extension profiles where the
841 dashed blue and grey lines correspond to the average unfolding force of the E and $G5^2$ subdomains
842 of pWT $(E\text{-}G5^2)_5$, respectively. The fits to a WLC model are displayed as solid black lines. **II)** F_U-

843 ΔL_c scatterplots at a retraction velocity of 1500 nms^{-1}, where red crosshairs are the mode and
844 FWHM from Gaussian fits to corresponding frequency histograms (not shown). Crosshairs from
845 one repeat at the same retraction velocity are shown for pWT $(E\text{-}G5^2)_5$ for reference (black). **III)**
846 Speed dependence of measured unfolding force, where the points (triangles (lower forces) and
847 circles (higher forces) represent E and $G5^2$, respectively) and errors are the mean and standard
848 deviation of triplicate datasets, respectively. The linear fit (broken lines in their respective colour)
849 is weighted with the inverse of the standard deviation error of the triplicate data sets, and pWT (E-
850 $G5^2)_2$ data is displayed in black for reference. Full data sets are found in Supplementary
851 Information.

852

853 **Figure 4.** Comparison of ΔF_U, mechanical ϕ-value and $\Delta\Delta G_{UN}$ of E-$G5^2$ variants. Topology
854 diagrams displaying the **A)** decrease in mechanical strength relative pWT $(E\text{-}G5^2)_5$ at a retraction
855 velocity of 1500 nms^{-1} vs $\Delta\Delta G_{UN}$ values and the **B)** mechanical ϕ-values for the variants of E-$G5^2$.
856 Comparison of these schematics shows that the majority of the substitutions are exerting their
857 effects on the ground state, whereas P562A is acting on the TS. *I502P is displayed as both 0.5
858 and ~1 as spectroscopic evidence suggests partial unfolding of the structure in the native state
859 (Figures S10B and S11B), which implies a value of 0.5 may be an artefact and could actually be
860 closer to ~ 1 at this residue [47]. V550P is assumed to have a <5 % decrease in the mechanical
861 strength of $G5^2$ as the unfolding force at 200 and 5000 nms^{-1} remain unchanged.

862

863

864

865