Folding, stability and mechanical strength of a biofilm-forming protein from *Staphylococcus aureus*

Gruszka D.T. [1*], **Farrance O.E.**[2], **Whelan F.**[3], **Brockwell D.J.**[2], **Paci E.**[2], **Potts J.R.**[3] **and Clarke J.**[1]

[1] Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, UK
[2] School of Molecular and Cellular Biology, University of Leeds, Leeds, UK
[3] Department of Biology, University of York, Heslington, York, UK
[*] E-mail: dg418@cam.ac.uk

UNIVERSITY OF CAMBRIDGE

Background & Objectives

- SasG is a protein covalently attached to the surface of *Staphylococcus aureus*[1] that forms elongated fibrils and promotes cell-to-cell accumulation during biofilm formation[2] (Fig. 1).
- SasG consists of an N-terminal A domain followed by a B region, composed of tandemly arrayed and highly homologous 128-residue repeats[2]. Each sequence repeat is comprised of a G5 domain[3] of 78 amino acids and a smaller subdomain of 50 amino acids, annotated as E and predicted to be disordered[4]. The G5 domain is stable in isolation and multi-domain constructs, whereas E folds only in the context of a C-terminal G5. The crystal structure of $G5^1EG5^2$ revealed a highly extended topology, in which G5 and E domains share the same fold (β-triple helix-β), and are arranged in a head-to-tail manner[4]. The domains are formed from two single-layer triple-stranded β-sheets connected by a collagen-like triple helical region[4] (Fig. 1 and 2).
- Towards the broader aim of understanding the role of SasG in biofilm formation, the specific objectives of this research were to determine the stability and (un)folding pathways of SasG domains using protein engineering, biophysical techniques, AFM and MD simulations.

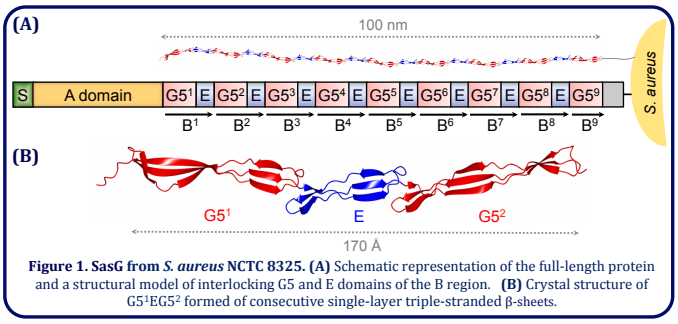

Figure 1. SasG from *S. aureus* NCTC 8325. (A) Schematic representation of the full-length protein and a structural model of interlocking G5 and E domains of the B region. **(B)** Crystal structure of $G5^1EG5^2$ formed of consecutive single-layer triple-stranded β-sheets.

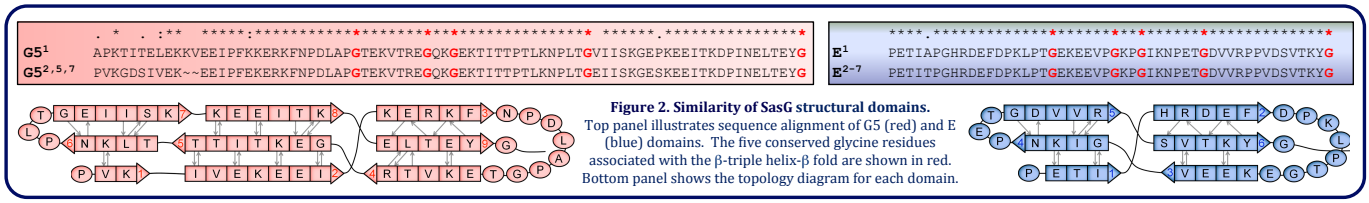

Figure 2. Similarity of SasG structural domains. Top panel illustrates sequence alignment of G5 (red) and E (blue) domains. The five conserved glycine residues associated with the β-triple helix-β fold are shown in red. Bottom panel shows the topology diagram for each domain.

Methods & Results

1. Urea-induced unfolding studies reveal that $G5^2$ and $EG5^2$ fold in a reversible, apparent two-state manner with a folding rate constant of ~12 s^{-1} (Fig. 3).

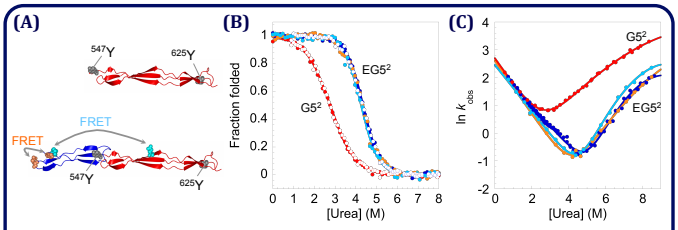

Figure 3. Urea-induced unfolding studies of $G5^2$ (red) and $EG5^2$ (blue). (A) Location of fluorophore probes. **(B)** Equilibrium unfolding (closed circles) and refolding (open circles) curves. **(C)** Chevron plots.

2. The Φ-value analysis of $G5^2$ and $EG5^2$ supported by Go model simulations indicate that the C-terminal loop/β-sheet region of G5 domain forms early in the folding process (Fig. 4).

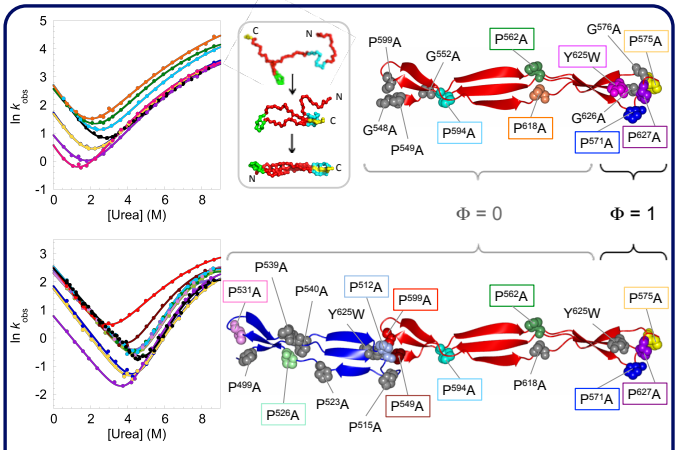

Figure 4. Folding kinetics of $G5^2$ (top), $EG5^2$ (bottom) and their selected mutants. Wild-type proteins are shown in black. Mutants are colour-coded and their location in the structure is indicated. Grey-framed panel illustrates a refolding mechanism for $G5^2$ domain based on Go model simulations.

3. SasG domains are mechanically resistant, as determined by atomic force microscopy (AFM) pulling experiments. G5 domain unfolds at a higher force than E (Fig. 5).

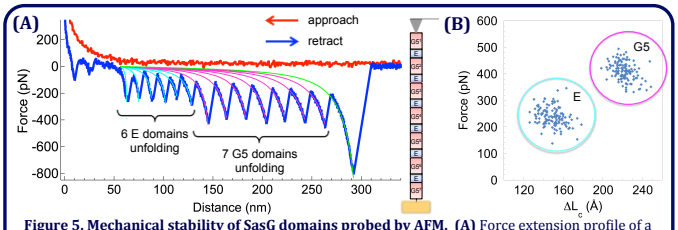

Figure 5. Mechanical stability of SasG domains probed by AFM. (A) Force extension profile of a SasG construct containing 7 G5 and 6 E domains at a tip retraction rate of 800 nm/s. Schematic of the experimental setup is presented. **(B)** Scatter plot of force against the difference in contour length.

4. Molecular dynamics (MD) simulations confirm that G5 is mechanically stronger than E and reveal forced unfolding mechanisms for both domains (Fig. 6).

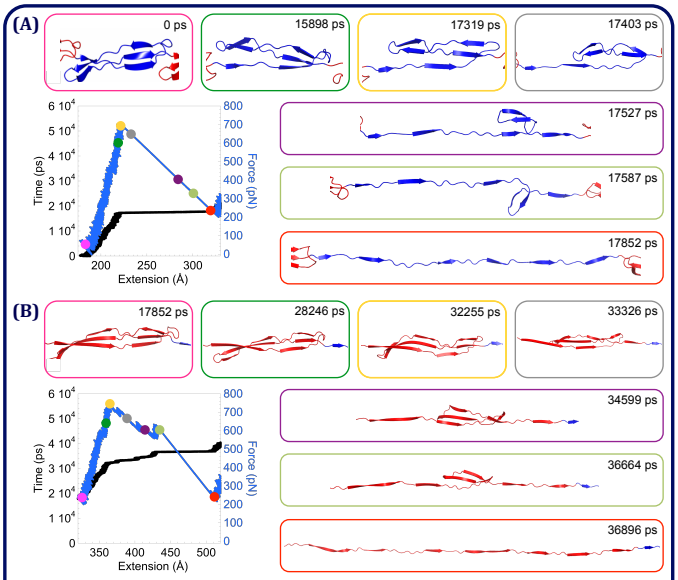

Figure 6. Constant velocity MD simulations of SasG domains. In this figure, simulations were carried out at 300K, 5 pN/nm and 0.01 nm/ps. **(A)** Forced unfolding mechanism of E domain in $G5^1EG5^2$ construct. **(B)** Forced unfolding mechanism of $G5^1$ domain in $G5^1EG5^2$.

Summary & Conclusions

- The B region of SasG is composed of two related structural domains, 78-residue G5 and a smaller E subdomain, which is unstable in isolation, but folds cooperatively with the C-terminal G5. Thus, from the folding perspective, SasG is formed from the N-terminal G5 domain, followed by a series of EG5 'bi-domains'. Both G5 and EG5 fold fast, given their high relative contact order (30% and 20%, respectively), with the C-terminal loop/β-strand region of G5 formed early in that process.
- SasG G5 and E domains are mechanically strong, compared to other strong all-β proteins, such as immunoglobulin domains. Simulations show that the mechanism of forced unfolding is similar for both domains, however G5 unfolds from a native-like metastable intermediate.

References

1. Banner M.A. *et al.* (2007) *J Bacteriol* **189**, 2793-2804
2. Corrigan R.M. *et al.* (2007) *Microbiology* **153**, 2435-2446
3. Bateman A. *et al.* (2005) *Bioinformatics* **21**, 1301-1303
4. Gruszka D.T. *et al.* (2012) *PNAS USA* **109** E1011-1018

Acknowledgements

- BBSRC
- Nathaniel Wand
- Judith Hawkhead
- Alexandra Travis

BBSRC bioscience for the future