

Structure, Volume 22

Supplemental Information

**The Statistical Conformation
of a Highly Flexible Protein: Small-Angle X-Ray
Scattering of *S. aureus* Protein A**

Jo A. Capp, Andrew Hagarman, David C. Richardson, and Terrence G. Oas

Supplemental Information

Supplemental Figures

Figure S1 – Related to Figure 2. The dimensionless Kratky plot indicates that (2-5)-BdpA and SpA-N may not be globular.

When the dimensionless Kratky plot fails to return to 0 at high $q \cdot R_g$ values, it is likely that the biopolymer exhibits significant flexibility (Durand et al., 2010; Rambo and Tainer, 2011). 1-BdpA converges to 0 at $q \cdot R_g = 4$, indicating that it is a globular protein. The Kratky plots of (2-5)-BdpA and SpA-N do not return to zero, indicating that these molecules are flexible or spherically asymmetric. Shown: A dimensionless Kratky plot of N-BdpA. Blue: 1-BdpA, Red: 2-BdpA, Purple: 3-BdpA, Green: 4-BdpA, Cyan: 5-BdpA, Black: SpA-N.

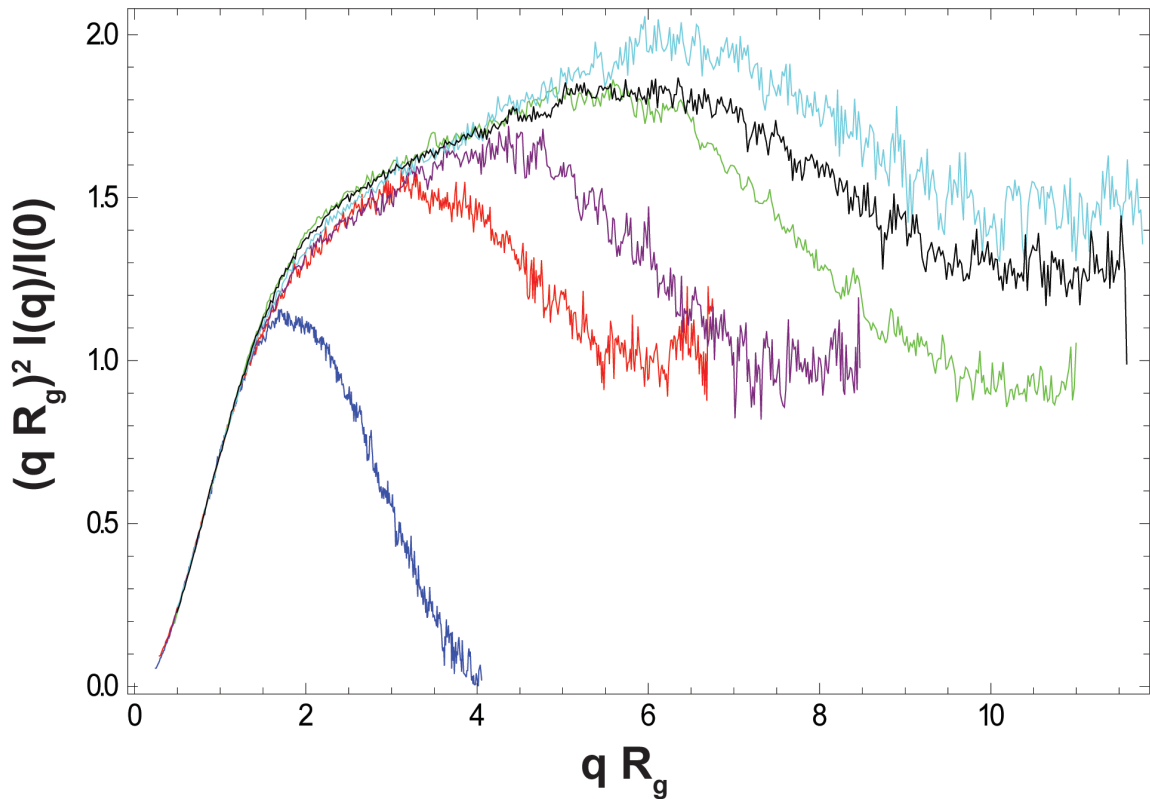


Figure S2 – Related to Figure 5. Ensemble analysis of (2-5)-BdpA

A) CRY SOL (Svergun et al., 1995) was used to calculate the theoretical scattering curve for each structure in the unconstrained parent ensemble. The aggregate scattering curve was calculated by summing the scattering curves for each structure in the parent ensemble. The aggregate scattering curve was normalized by $I(q)/I(0)$. Black: data. Red: calculated aggregate scattering curve. The χ^2 statistics are as follows: 2-BdpA: 1.04, 3-BdpA: 1.12, 4-BdpA: 0.94, 5-BdpA: 1.04. **B,C)** A converged unconstrained ensemble of structures were generated for (2-5)-BdpA using RanCH (Bernado et al., 2007) in random coil mode. GAJOE was used to generate 3 “minimal ensembles” using the same unconstrained parent ensemble. The experimental scattering curves were used to constrain the minimal ensembles. B) the R_g distributions for each minimal ensemble (red, green, blue) and the parent ensemble (black). C) D_{max} distributions for the minimal ensembles (red, green, blue) and the parent ensemble (black). The χ^2 statistics for each minimal ensemble are as follows: 2-BdpA ensembles (red, green, blue): 1.03, 3-BdpA ensembles (red, green, blue): 0.94, 4-BdpA ensembles (red, green, blue): 1.02, 5-BdpA ensembles (red, green, blue): 1.05. For each dataset, all the χ^2 statistics were within 0.004 of each other.

Figure S2

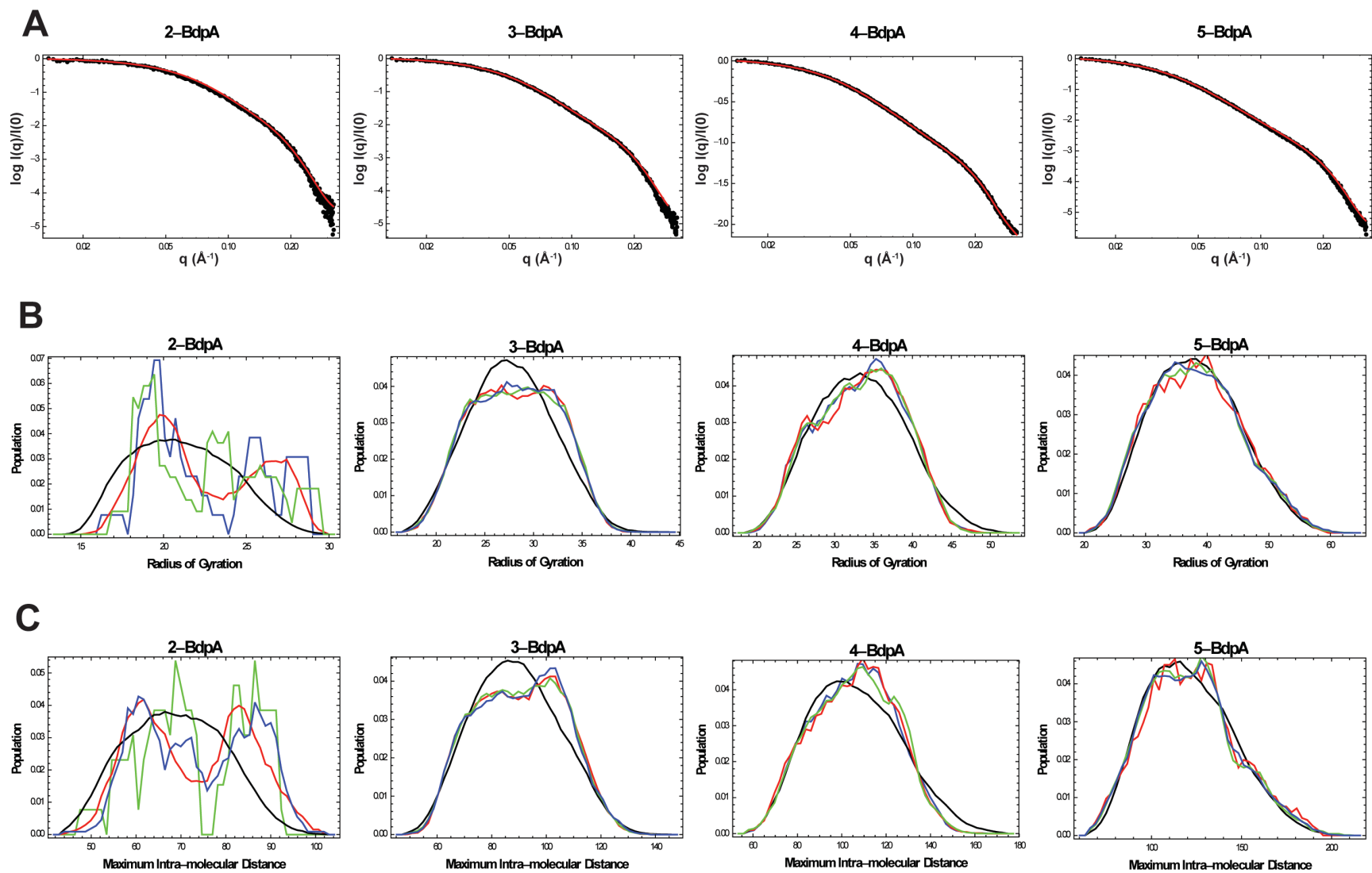


Figure S3 – Related to Figure 7 – SAXS Analysis of LAR12P

Application of Figure 7 protocol to a semi-flexible system for which ensemble modeling is appropriate

To demonstrate the utility of the protocol outlined in Fig. 7 to detect residual inter-domain structure, we applied the method to another minimally flexible protein whose published SAXS data supported a finer-grain description modeled through an ensemble of atomistic model structures. The SAXS data for this counter example, the immunoglobulin-like domains 1-2 of LAR3 (Biersmith et al., 2011) (BIOISIS ID: LAR12P), has been deposited in the BIOISIS database (<http://Biolsis.net>). This is a 2-domain protein with a flexible N- and C-terminal tags. The Kratky plot of the scattering data does not return to zero (Figure S4A) and there is no plateau in the Porod-Debye plot (step 1 and Figure S4B), providing strong evidence that the protein may not be completely globular. Although the flexible cylinder ellipse polymer model, fitted to the data in SASView, fit the LAR12P scattering data well (Figure S4C), a more detailed model may be justified. To determine if this is the case for LAR12P, a converged ensemble of structures was generated with RanCH, with residues 1 - 8 (N-terminus), 126 - 136 (inter-domain linker), and 226 - 237 (C-terminus) dihedrals randomized. Since this protein is minimally flexible, and the two domain orientations appear to be fixed in the protein, then an aggregate scattering curve where the domain orientations are not fixed should fit the data poorly. However, a minimal ensemble, where these two domains are in contact with one another, should fit the data reasonably well. In contrast to our results for (3-5)-BdpA and SpA-N, scattering curve of the unconstrained RanCH ensemble for LAR12P did not completely fit the experimental scattering data (step 3 and Figure S4D), justifying the use of GAJOE to select a more restricted ensemble of structures. The resultant scattering curve for the minimal ensemble fit the experimental data well (Figure S4E). The resultant R_g and D_{max} distributions from the minimal ensemble were smaller than the R_g and D_{max} distribution from the parent ensemble (step 4 and Figure S4F,G). The best fit ensemble was a two member ensemble where the domains were in contact with

each other and in an anti-parallel conformation. Based on this analysis following the protocol outlined in Fig. 7, we conclude that the LAR12P SAXS data supports limited ensemble modelling. In this case we can conclude that there are more constraints on the statistical conformation than steric constraints alone. Thus, not all flexible proteins are devoid of inter-domain interactions. These interactions can lead to SAXS data with more information than can be fit with simple polymer physics models and can be used to infer higher resolution details about the statistical conformation. However, it is important to use the protocol we have outlined to assure that such details are justified by the SAXS data. In the case of (3-5)-BdpA and SpA-N they are not.

Figure S3 Legend. **(A)** The Kratky plot of LAR12P (Durand et al., 2010) does not return to zero, indicating flexibility or spherical asymmetry in the statistical conformation. **(B)** There is no plateau in the Porod-Debye plot, indicating that the protein may be non-globular (Rambo and Tainer, 2011). **(C)** Fit of the flexible-cylinder-ellipse model (red) to the LAR12P SAXS data (black). Residuals (red) are shown below the SAXS curves. Note that the flexible-cylinder-ellipse model fits the data well at $q < 0.20$, indicating that the global structure of this protein may be described by this model. **(D)** Ensemble analysis - Fit of the aggregate scattering curve calculated from the unconstrained parent ensemble (red) to the experimental data (black). Residuals (red) are shown below the SAXS curves. The aggregate scattering curve does not fit the SAXS data well. **(E)** Fit of the calculated scattering curve for the minimal ensembles (red, green, blue) to the LAR12P data (black). **(F)** R_g distributions of the unconstrained parent ensemble (black) and the minimal ensembles (red, green, blue). The R_g distributions of the minimal ensembles are considerably narrower and smaller than the R_g distribution of the parent ensemble. **(G)** R_g distributions of the unconstrained parent ensemble (black) and the minimal ensembles (red, green, blue). The D_{max} distributions of the minimal ensembles are considerably narrower and smaller than the D_{max} distribution of the parent ensemble.

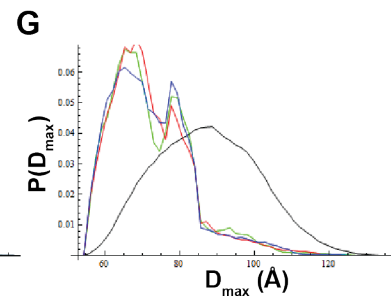
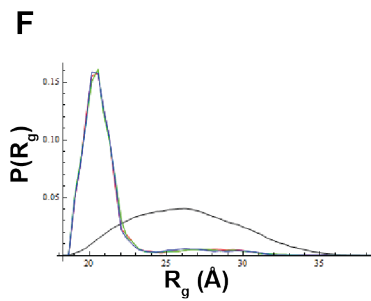
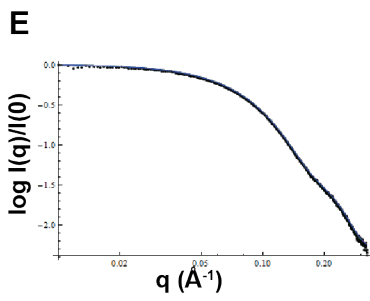
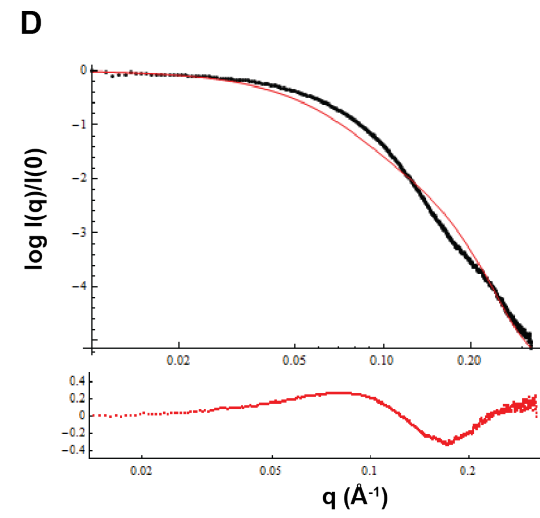
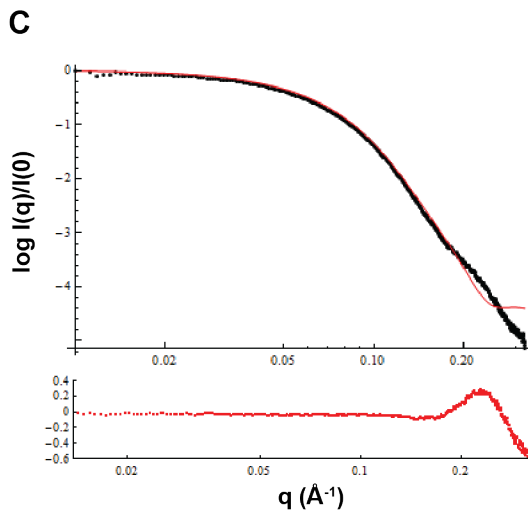
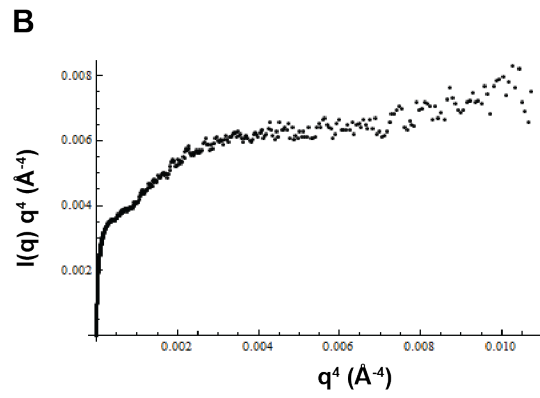
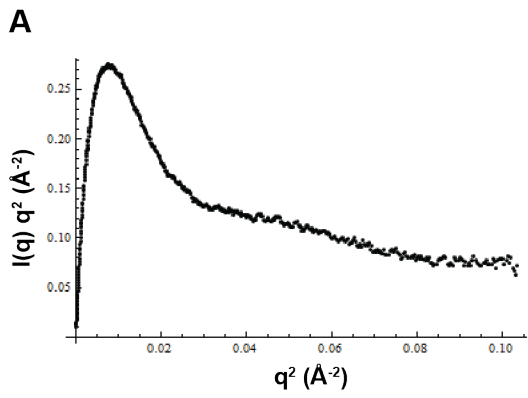


Table S1 – Related to Figure 4. A summary of the scattering functions for the various polymer models used to fit the 5-BdpA SAXS data.

Polymer Model	Domain Form Factor	Domain Structure Factor	Linker Form Factor	Linker Structure Factor	Cross-term
Swollen Gaussian Coil	Swollen Gaussian Coil (domains are dimensionless)	None	None	None	None
Linear Pearl Necklace	Sphere	Infinitely thin rod	None	None	None
Random-Flight Pearl Necklace	Sphere	Random-flight polymer	Rod	Random-flight polymer	Included
Excluded Volume Pearl Necklace	Sphere	Swollen Gaussian coil	Swollen Gaussian coil	None	Included

Table S2 – Related to Figure 5. Ensemble analysis of (2-5)-BdpA – Sizes of the parent and minimal ensembles for each protein.

n-BdpA protein	Parent ensemble	Minimal ensemble #1	Minimal ensemble #2	Minimal ensemble #3
2-BdpA	100,014	44	26	22
3-BdpA	10,000	34	28	38
4-BdpA	10,000	48	32	32
5-BdpA	84,669	40	28	38

Supplemental Experimental Procedures: Theory

Small-angle x-ray scattering is a biophysical technique that probes the global conformational space of a biopolymer in solution. It is assumed that the biopolymers are randomly orientated, so the observed intensity, $I(q)$, as a function of momentum transfer, q , is related to the shape and spatial arrangement of monomers by the Debye scattering equation (Koch et al., 2003):

$$I(q) = \sum_i \sum_j F_i(q) F_j^*(q) \left\langle \frac{\sin(qr_{ij})}{qr_{ij}} \right\rangle \quad (\text{S1a})$$

where the sum is over all monomers within a single biopolymer molecule; F_i is the scattering-amplitude-weighted form factor of monomer i , F_j^* is the complex conjugate of the form factor of monomer j ; r_{ij} is the distance between monomers i and j ; and $\langle \rangle$ denotes an orientational average. Under non-anomalous conditions, $F^*(q) = F(q)$. For simple models the spatial arrangement of the monomers, $\left\langle \frac{\sin(qr_{ij})}{qr_{ij}} \right\rangle$, can be analytically described, and is referred to as the structure factor, $S(q)$. Thus,

$$I(q) = \sum_{i=1}^n \sum_{j=1}^n F_i(q) F_j(q) S_{ij}(q) \quad (\text{S1b})$$

This basic scattering function can be adapted to complex systems using the form factor for a specific shape of monomer and a specific structure factor corresponding to the arrangement of these monomers in solution (linear, random flight, excluded volume, sticky hard sphere, etc.). The structure factor describes interference between the monomers, whose shape (sphere, ellipsoid, cylinder, ring, etc.) is described by their form factor.

The Barbell Model (BM): Introduction

2-BdpA can be modeled as a polymer composed of two phases: homogeneous domains and flexible linkers. The two domains are described as spheres with a form factor, F_{sphere} . The structure factor (S_{lin}), which describes the interference due to the two spheres separated by a fixed distance. S_{lin} can be fit to an experimental SAXS profile to determine the average distance

between the two domains. The flexible regions of the protein (N-terminus and inter-domain linker) are modeled as excluded volume Gaussian coils. This form factor is denoted F_{ev} . The analytical derivations of F_{sphere} , F_{ev} , and S_{lin} are described in detail below (equations S4, S6, and S9). Substituting these form and structure factors into Eq S1b, gives the following result for the sums over all components:

$$I_{BM}(q) = 2 F_{sphere}^2(q) S_{lin}(q) + 4 F_{sphere}(q) F_{ev}(q) S_{lin}(q) + 2 F_{ev}^2(q) \quad (S2)$$

In $I_{BM}(q)$ the length of the linkers and the distance between two domains is uncorrelated. The first term of Eq. S2 accounts for scattering of the spheres and the interference between them; the third term accounts for coil scattering, and the middle term accounts for coil-sphere interference. Eq. S2 ignores interference between coils because this term represents a very small fraction of the total scattering of 2-BdpA.

The Excluded Volume Pearl-Necklace Model (EV-PNM): Introduction

N-BdpA can be similarly modeled using an excluded volume pearl-necklace model (EV-PNM) for $n > 2$. The two phases of this model are made up of n homogeneous domains and n homogenous flexible regions of the protein (N-terminus and inter-domain linkers). The domains are modeled as spheres with a form factor, F_{sphere} . The spatial arrangement of domain spheres is represented by the structure factor for an excluded volume Gaussian coil, S_{ev} . S_{ev} includes parameters that model the flexibility of the chain of spheres: the persistence length (l_p) and the Flory coefficient (ν). The model for the flexible regions is identical to the description in the BM. Substituting these form and structure factors into equation 4b gives the following results for the sums over all components:

$$I_{EVPNM}(q) = N F_{sphere}^2(q) S_{ev}(q) + 2 N F_{sphere}(q) F_{ev}(q) S_{ev}(q) + N F_{ev}^2(q) \quad (S3)$$

N is the number of BdpA domains in the protein. For SpA-N, $N=5$. In equation S3 the length of the linkers and the distance between two domains is uncorrelated. Equation S3 also ignores interference between coils because this term represents a very small fraction of the total

scattering of N-BdpA.

Scattering function: Barbell model

$I_{BM}(q)$ can be decomposed into three components: scattering and interference from spheres ($F_{sphere}^2(q) S_{ev}(q)$), scattering from coils ($F_{ev}^2(q)$), and a cross term ($2 F_{sphere}(q) F_{ev}(q) S_{ev}(q)$), that accounts for correlations between interconnecting spheres and coils. The so-called “coils” of the model are equivalent to the inter-domain linker and the 6 residues at the termini.

Scattering from domain spheres: $F_{sphere}^2(q) S_{BM}(q)$.

The normalized scattering amplitude of the domains is assumed to be the form factor of a sphere (Glatter and Kratky, 1982):

$$F_{spheres}(q) = 3 \frac{\sin(qR) - (qR)\cos(qR)}{(qR)^3} W_s \quad (S4)$$

R is the radius of the spheres, and W_s is the ratio of the scattering mass of the domains to the scattering mass of the coil. For 2-BdpA, this ratio has been determined by NMR dynamics data. The coils consist of 12 (6 residues each) residues, and the spheres consist of 52 residues each. This gives a scattering mass ratio of $11647.2:1425.6 = 8.17$.

The structure factor describing spatial arrangement of the spheres in the biopolymer is (Schweins and Huber, 2004):

$$S_{BM}(q) = \frac{2}{1 - \frac{\sin(qb)}{qb}} - 1 - \frac{1 - \left(\frac{\sin(qb)}{qb}\right)^2}{\left(1 - \frac{\sin(qb)}{qb}\right)^2} * \frac{\sin(qb)}{qb} \quad (S5)$$

where b is the distance between the domain spheres, the only fitted parameter in this term.

Scattering from the coils: $F_{ev}^2(q)$

As Hammouda notes (Hammouda, 1993), the square of the form factor for an excluded volume Gaussian coil is not simply $F_{ev}^2(q)$, but is instead:

$$F_{ev}^2(q) = S_{ev}(q) = \frac{1}{vX^{\frac{1}{2}}} * \Gamma\left(\frac{1}{2v}, X\right) - \frac{1}{X^{\frac{1}{2v}}} * \Gamma\left(\frac{1}{v}, X\right) \quad (S6)$$

where $\Gamma(a, X)$ and X are:

$$\Gamma(a, X) = \int_0^X dt \exp(-t)t^{a-1}, \text{ and} \quad (\text{S7})$$

$$X = \frac{q^2 \left(\frac{l_p}{2}\right) N^{2\nu}}{6} \quad (\text{S8})$$

In the case of the 2-BdpA linker, $\nu = 0.588$ (the Flory coefficient for unfolded peptides) (Baldwin, 2002), $l_p = 6.6$ (the persistence length of unfolded proteins) (Lairez et al., 2003), and $N = 6$ (the number of residues in the linker). There are no adjustable parameters in this term.

Cross-terms: $2 F_{sphere}(q) F_{ev}(q) S_{BM}(q)$

The components $F_{sphere}(q)$ and $S_{BM}(q)$ have been defined above. In this case, $S_{ev}(q)$ refers to the spatial arrangement of the spheres. $F_{ev}(q)$ is the form factor for the coils and is (Hammouda, 1993):

$$F_{ev}(q) = \frac{1}{\nu X^{2\nu}} * \Gamma\left(\frac{1}{2\nu}, X\right) - \frac{1}{\nu X^{\frac{1}{\nu}}} * \Gamma\left(\frac{1}{\nu}, X\right) \quad (\text{S9})$$

The fitted parameters in this term are R (the radius of the domain spheres) and b (the distance between spheres).

Scattering function - excluded volume pearl necklace model

$I_{EVPNM}(q)$ can be decomposed into three components: scattering and interference from spheres ($F_{sphere}^2(q) S_{ev}(q)$), scattering from coils ($F_{ev}^2(q)$), and a cross term ($2 F_{sphere}(q) F_{ev}(q) S_{ev}(q)$), accounting for correlations between interconnecting spheres and coils.

Scattering from domain spheres: $F_{sphere}^2(q) S_{ev}(q)$

The normalized scattering amplitude of the domains is assumed to be sphere, identical to that in the Barbell model:

$$F_{spheres}(q) = 3 \frac{\sin(qR) - (qR)\cos(qR)}{(qR)^3} W_s \quad (\text{S10})$$

R is the radius of the spheres, and W_s is the ratio of the scattering mass of the domains to the scattering mass of the coil. For SpA-N and N-BdpA, this ratio is 8.17/1.

The structure factor describing the interference due to the spatial arrangement of spheres within

the biopolymer is (Hammouda, 1993):

$$S_{ev}(q) = \frac{1}{vX^{\frac{1}{2}}} * \Gamma\left(\frac{1}{2v}, X\right) - \frac{1}{X^{\frac{1}{2v}}} * \Gamma\left(\frac{1}{v}, X\right) \quad (\text{S11})$$

where $\Gamma(a, X)$ is the incomplete gamma function:

$$\Gamma(a, X) = \int_0^X dt \exp(-t) t^{a-1}, \quad (\text{S12})$$

and X is:

$$X = \frac{q^2 \left(\frac{l_p}{2}\right) N^{2v}}{6}. \quad (\text{S13})$$

v is the Flory coefficient, l_p is the persistence length, and N is the number of domains. The adjustable parameters in $F_{sphere}^2(q) S_{ev}(q)$ are R , v , and l_p .

Scattering from the coils: $F_{ev}^2(q)$

This term is identical to that of the Barbell model, Eq.S6.

Cross-terms: $2 F_{sphere}(q) F_{ev}(q) S_{ev}(q)$

The components $F_{sphere}(q)$ and $S_{ev}(q)$ has been defined above. In this case, $S_{ev}(q)$ refers to the spatial arrangement of the spheres. $F_{ev}(q)$ is the form factor for the coils and is (Hammouda, 1993):

$$F_{ev}(q) = \frac{1}{vX^{\frac{1}{2v}}} * \Gamma\left(\frac{1}{2v}, X\right) - \frac{1}{vX^{\frac{1}{v}}} * \Gamma\left(\frac{1}{v}, X\right) \quad (\text{S14})$$

where $\Gamma(a, X)$ and X are defined above. The fitted for parameters in this function are R (the radius of the domain spheres), v (the Flory coefficient for the sphere spatial relationship), and l_p (the persistence length of the entire chain).

Generation of the Graphical Abstract

The graphical abstract models the distribution of SpA on a *S. aureus* cell wall. The red spheres represent the root-mean-squared radius of gyration of SpA-N (Figure 2), the attached red lines represent the C-terminal half of SpA anchored on the cell wall (blue cuboid), and the blue lines depict the peptidoglycan molecules in the cell wall. On average, we calculated that

there are approximately 117 molecules of SpA per 10^6 \AA^2 of cell surface.

The diameter of a spherical *S. aureus* cell is 0.5 – 1.5 μm (Harris et al., 2002). It has been reported that SpA makes up 1.7% of the total mass of the cell (Sjoquist et al., 1972). Chang, 2013 (Chang et al., 2013), reported that 1 OD_{600} contains 1.5×10^8 *S. aureus* cells/ml. Using the dry weight of *S. aureus* calculated by Downer, 2002 (Downer et al., 2002), and the molecular weight of SpA (57,320 Da), we calculated that there are ~398000 molecules of SpA per cell. If we assume that the diameter of a single cell is 1.5 μm , and that SpA is uniformly distributed in the cell, a $1000 \text{ \AA} \times 1000 \text{ \AA}$ section of the cell surface will contain 117 molecules of SpA.

Supplemental References

Baldwin, R.L. (2002). A new perspective on unfolded proteins. *Advances in protein chemistry* 62, 361-367.

Bernado, P., Mylonas, E., Petoukhov, M.V., Blackledge, M., and Svergun, D.I. (2007). Structural characterization of flexible proteins using small-angle X-ray scattering. *Journal of the American Chemical Society* 129, 5656-5664.

Chang, Y.C., Yang, C.Y., Sun, R.L., Cheng, Y.F., Kao, W.C., and Yang, P.C. (2013). Rapid single cell detection of *Staphylococcus aureus* by aptamer-conjugated gold nanoparticles. *Scientific reports* 3, 1863.

Downer, R., Roche, F., Park, P.W., Mecham, R.P., and Foster, T.J. (2002). The elastin-binding protein of *Staphylococcus aureus* (EbpS) is expressed at the cell surface as an integral membrane protein and not as a cell wall-associated protein. *The Journal of biological chemistry* 277, 243-250.

Durand, D., Vives, C., Cannella, D., Perez, J., Pebay-Peyroula, E., Vachette, P., and Fieschi, F. (2010). NADPH oxidase activator p67(phox) behaves in solution as a multidomain protein with semi-flexible linkers. *Journal of structural biology* 169, 45-53.

Franke, D., and Svergun, D.I. (2009). DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *Journal of Applied Crystallography* 42, 342-346.

Glatter, O., and Kratky, O. (1982). *Small angle x-ray scattering* (London ; New York: Academic Press).

Hammouda, B. (1993). SANS from homogenous polymer mixtures: A unified overview. *Advances in Polymer Science* 106, 87-133.

Harris, L.G., Foster, S.J., and Richards, R.G. (2002). An introduction to *Staphylococcus aureus*, and techniques for identifying and quantifying *S. aureus* adhesins in relation to adhesion to biomaterials: review. *European cells & materials* 4, 39-60.

Koch, M.H., Vachette, P., and Svergun, D.I. (2003). Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Quarterly reviews of biophysics* 36, 147-227.

Lairez, D., Pauthe, E., and Pelta, J. (2003). Refolding of a high molecular weight protein: salt effect on collapse. *Biophysical journal* 84, 3904-3916.

Rambo, R.P., and Tainer, J.A. (2011). Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers* 95, 559-571.

Schweins, R., and Huber, K. (2004). Particle scattering factor of pearl necklace chains. *Macromolecular Symposia* 211, 25-42.

Sjoquist, J., Meloun, B., and Hjelm, H. (1972). Protein A isolated from *Staphylococcus aureus* after digestion with lysostaphin. *European journal of biochemistry / FEBS* 29, 572-578.

Svergun, D., Barberato, C., and Koch, M.H.J. (1995). CRY SOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of Applied Crystallography* 28, 768-773.

Svergun, D.I. (1992). Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. *Journal of Applied Crystallography* 25, 495-503.