SUPPORTING INFORMATION (SI) APPENDIX

The Ag43 structure reveals a molecular Velcro-like mechanism of autotransporter-mediated bacterial clumping

Begoña Heras[a,1,2], Makrina Totsika[b,2], Kate M. Peters[b], Jason J. Paxman[a], Christine Gee[c], Russell Jarrott[d], Matthew A. Perugini[a], Andrew E. Whitten[d] and Mark A. Schembri[b,1]

[a]Department of Biochemistry, La Trobe University, Melbourne VIC 3086, Australia. [b]Australian Infectious Diseases Research Centre, School of Chemistry and Molecular Biosciences, The University of Queensland, St Lucia, Brisbane QLD 4072, Australia. [c]Australian Synchrotron, 800 Blackburn Rd, Clayton, Melbourne VIC 3168, Australia. [d]Institute for Molecular Bioscience The University of Queensland, St Lucia, Brisbane QLD 4072, Australia

[1]Corresponding authors:

[2]Both authors contributed equally to this work and hence are joint first authors.

TABLE OF CONTENTS

SI MATERIAL AND METHODS

SAXS data collection and analysis

Data were collected on the SAXS-WAXS beamline at the Australian Synchrotron with a sample to detector distance set at 1567 mm and an X-ray wavelength of λ = 1.033 Å, allowing access to a q-range spanning ~0.01-0.55 Å^{-1}. The A_{280} of the stock solution was 17.0 mg/mL (ε_{280} = 26470 cm^{-1}mol^{-1}, molecular mass =50.1 kDa). The stock solution was diluted to ~3.4 mg/mL. Immediately prior to loading, all samples were centrifuged at 10,000 *g* to remove large particles from the solution. To minimise the effects of radiation damage, samples were flowed (100 µl) past the beam in 1.5 mm quartz capillaries (Hampton Research) at 283 K Data measuring 5 frames (each with a 2 sec exposure time). Data reduction was carried out using SAXS15ID software (1), averaging all 5 measured frames and correcting for solvent scattering, sample transmission, detector sensitivity and background radiation. Data were placed on an absolute scale, by normalization against a water standard.

Data quality was assessed by inspection of the linearity of the Guinier region of the data, and estimated molecular mass of the protein complex. To determine the estimated molecular mass the contrast and partial specific volume were estimated from the protein sequence (2). The estimated mass was ~20% too low, thus, to establish whether this discrepancy was due to poor concentration estimates or aggregation, the mass of the particle was estimated using the concentration independent approach (3). This mass estimate was very close to the expected mass, suggesting that the concentration of the protein was over-estimated.

The pair-distance distribution function ($p(r)$) was generated from the experimental data using *GNOM* (4), from which $I(0)$, R_g and D_{max} were determined. The predicted scattering profile of the crystal structure was calculted using the program *CRYSOL*. The program *DAMMIN* (5) was used to generate 10 molecular envelopes, which were averaged using the program *DAMAVER* (6).

Sedimentation Velocity-Analytical Ultracentrifugation (SV-AUC)

SV-AUC experiments were conducted using a Beckman model XL-A analytical ultracentrifuge at a temperature of 20 °C. Samples (380 μl) of α^{43a} (peak 1 and peak 2 obtained from size exclusion chromatography) and mutant (α^{43a} ΔH1H2) solubilized in 25 mM HEPES, 150mM NaCl, pH 7.1; and reference solutions (400 μl) were loaded into a conventional double sector quartz cell and mounted in a Beckman 8-hole An50 Ti rotor. Absorbance data were collected in continuous mode at a single wavelength between (230 or 280 nm) using a rotor speed of 40,000 rpm and a step-size of 0.003 cm without averaging. Frictional ratios (f/f_0) were calculated employing SEDNTERP (7) using the measured sedimentation coefficient of the given species, molecular mass calculated from amino acid sequence, a solvent density of 1.005 g/ml, viscosity of 1.023 cp and a partial specific volume of 0.7194 ml/g (Ag43 native) and 0.7187 ml/g (Ag43 ΔH1H2 mutant). The resulting sedimentation velocity absorbance data at multiple time points were fitted to a continuous size-distribution model using the program SEDFIT employing a resolution of 200 species ranging from 0 S to 10 S and a P-value of 0.95 (8-10).

SI REFERENCES

1. Cookson D (2006) Saxs15ID—Software for acquiring, processing and viewing SAXS/WAXS image data at ChemMatCARS. . http://millenia.cars.aps.anl.gov/chemmat/documents/saxswaxs/saxs15id_050807.doc.
2. Whitten AE, Cai S, & Trewhella J (2008) MULCh: modules for the analysis of small-angle neutron contrast variation data from biomolecular assemblies *J Appl Crystallogr* 41(1):222-226.
3. Fischer H, de Oliveira Neto M, Napolitano HB, Polikarpov I, & Craievich AF (2010) Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale *J Appl Crystallogr* 43:101-109.
4. Svergun DI (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Crystallogr* 25:495-503
5. Svergun DI (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* 76(6):2879-2886.
6. Volkov VV & Svergun DI (2003) Uniqueness of ab initio shape determination in small-angle scattering. *J Appl Crystallogr* 36:860-864.
7. Laue TM, Shah BD, & Pelletier SL (1992) Computer-aided interpretation of analytical sedimentation data for proteins. In Analytical Ultracentrifugation in Biochemistry and Polymer Science. *The Royal Society of Chemistry Cambridge*:90-125.
8. Schuck P (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J* 78(3):1606-1619.
9. Schuck P, Perugini MA, Gonzales NR, Howlett GJ, & Schubert D (2002) Size-distribution analysis of proteins by analytical ultracentrifugation: strategies and application to model systems. *Biophys J* 82(2):1096-1111.
10. Perugini MA, Schuck P, & Howlett GJ (2000) Self-association of human apolipoprotein E3 and E4 in the presence and absence of phospholipid. *J Biol Chem* 275(47):36758-36765.
11. Svergun D, Barberato C, & Koch MHJ (1995) CRYSOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J Appl Crystallogr* 28:768-773
12. Baker NA, Sept D, Joseph S, Holst MJ, & McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* 98(18):10037-10041.
13. Poirot O, O'Toole E, & Notredame C (2003) Tcoffee@igs: A web server for computing, evaluating and combining multiple sequence alignments. *Nucleic Acids Res* 31(13):3503-3506.
14. Emsley P, Charles IG, Fairweather NF, & Isaacs NW (1996) Structure of Bordetella pertussis virulence factor P.69 pertactin. *Nature* 381(6577):90-92.
15. Johnson TA, Qiu J, Plaut AG, & Holyoak T (2009) Active-site gating regulates substrate selectivity in a chymotrypsin-like serine protease the structure of haemophilus influenzae immunoglobulin A1 protease. *J Mol Biol* 389(3):559-574.
16. Meng G, Spahich N, Kenjale R, Waksman G, & St Geme JW, 3rd (2011) Crystal structure of the Haemophilus influenzae Hap adhesin reveals an intercellular oligomerization mechanism for bacterial aggregation. *EMBO J* 30(18):3864-3874.
17. Gangwer KA, *et al.* (2007) Crystal structure of the Helicobacter pylori vacuolating toxin p55 domain. *Proc Natl Acad Sci U S A* 104(41):16293-16298.
18. Khan S, Mian HS, Sandercock LE, Chirgadze NY, & Pai EF (2011) Crystal structure of the passenger domain of the Escherichia coli autotransporter EspP. *J Mol Biol* 413(5):985-1000.
19. Otto BR, *et al.* (2005) Crystal structure of hemoglobin protease, a heme binding autotransporter protein from pathogenic Escherichia coli. *J Biol Chem* 280(17):17339-17345.

20. Krissinel E & Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2256-2268.
21. Emsley P, Lohkamp B, Scott WG, & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486-501.

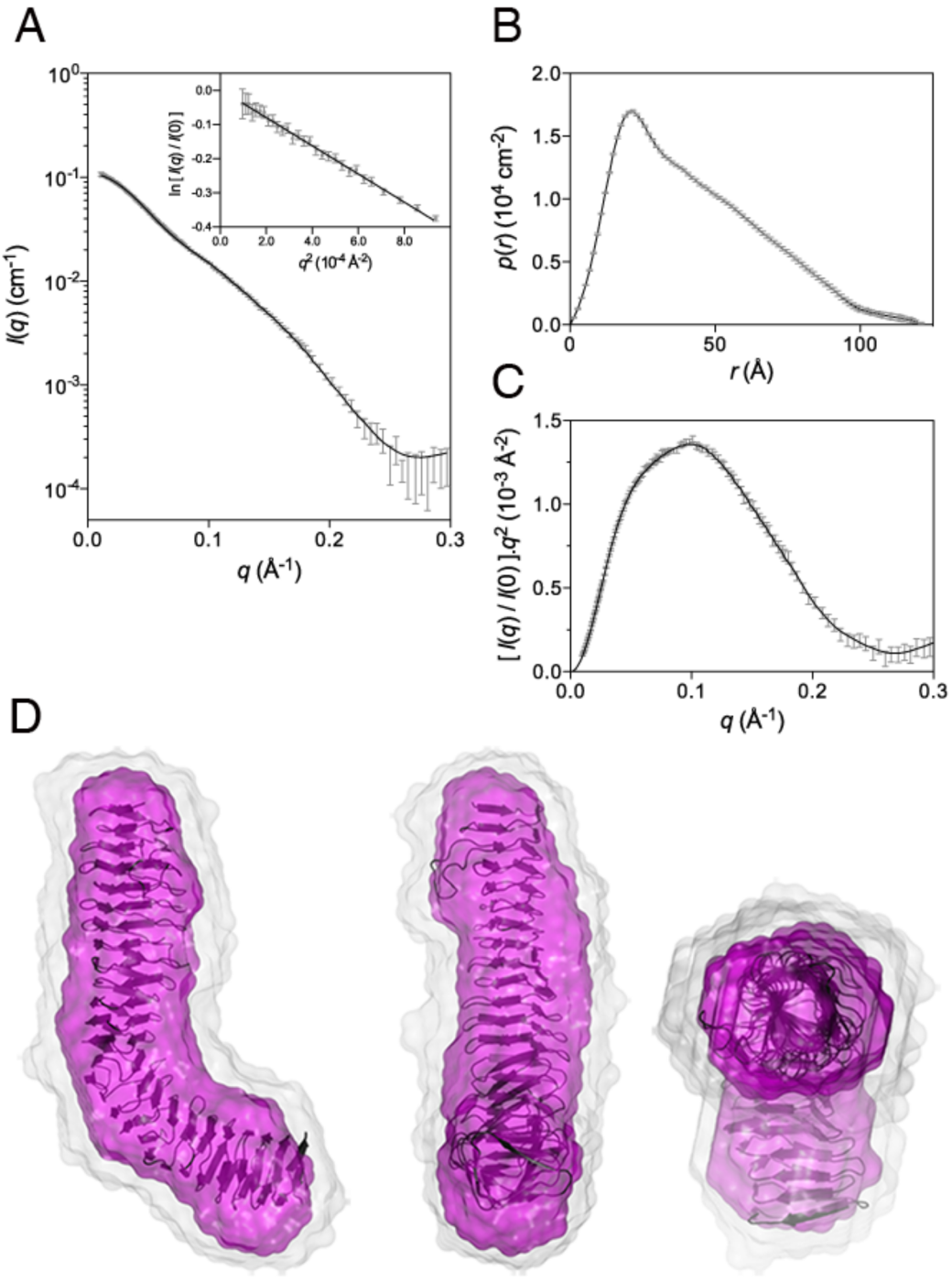


Figure S1. Small-angle X-ray scattering from α^{43a}. (*A*) Predicted scattering profile of the α^{43a} crystal structure calculated using CRYSOL (11) (solid black line) overlaid on the measured scattering data (grey data points, $\chi^2 = 0.74$). Inset: Guinier plots of the low-angle portion of the scattering data are linear consistent with monodisperse solutions. (*B*) Pair-distance distribution function derived from the scattering data using GNOM (4); (*C*) Porod plot of the scattering data, consistent with a well folded protein. (*D*) Comparison between the crystal structure (black cartoon) and the model of the protein structure obtained from *ab intio* modelling against the scattering data. The grey envelope represents the total volume encompassed by the 10 aligned models obtained from the modelling program DAMMIN (5) (χ^2=0.37 ± 0.01 for all models). The probable shape of the protein obtained by the averaging and filtering processed employed by DAMAVER (6) is shown in magenta. The normalised spatial discrepancy of the averaging procedure is 0.570 ± 0.012, where none of the 10 models were rejected.

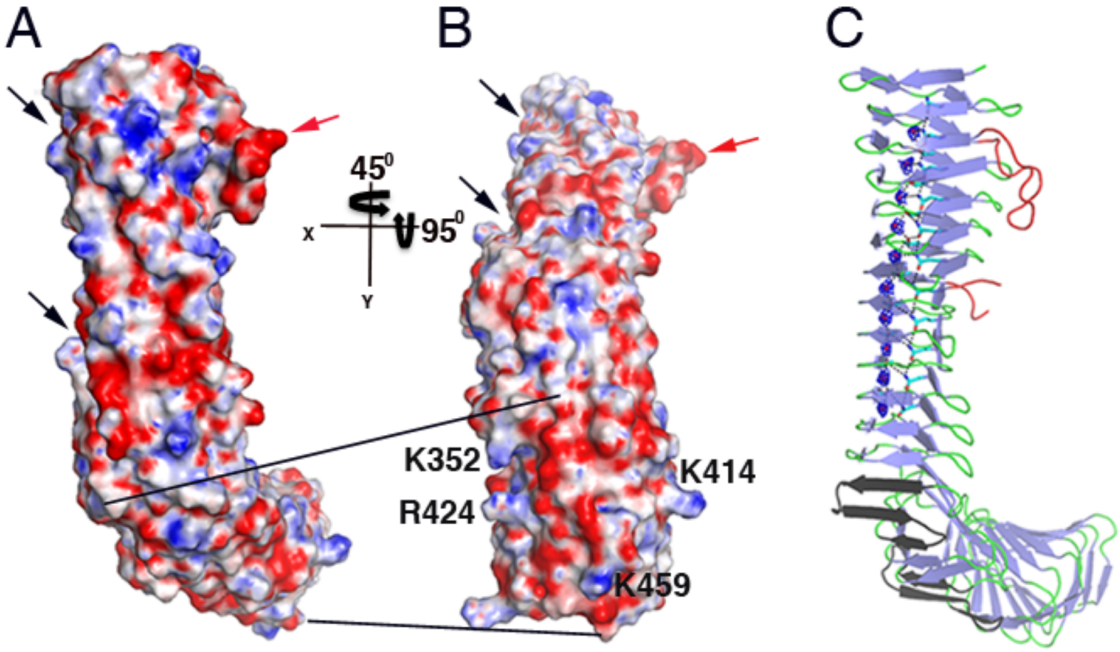


Figure S2. Structural analysis of α^{43a}. (*A*) Electrostatic surface of α^{43a}. Positive and negative electrostatic potentials are shown in blue and red, respectively (saturation at 15 kT/e) for each of the proteins. The orientation of the panel corresponds to that in Figure 1B. Black arrows indicate the long grooves running diagonally along the β-helix. Red arrow indicates the acidic patch generated by the negatively charged loop 1. (*B*) Charged surface of the foot of the L-shaped α^{43a}. The orientation in this panel corresponds to the protein in panel a rotated by 45° along the axis of the β-helix (Y axis) and 95° along the axis perpendicular to the β-helix (X axis). Positively charged residues protruding from the β-helix are labelled. (*C*) Ribbon diagram of α^{43a} showing an intricate network of hydrogen bonds in the interior of the protein binding stacks of polar residues and a single-file water chain running down the passenger domain of the Ag43 adhesin**.** Figures of the electrostatic potential were generated using APBS (12).

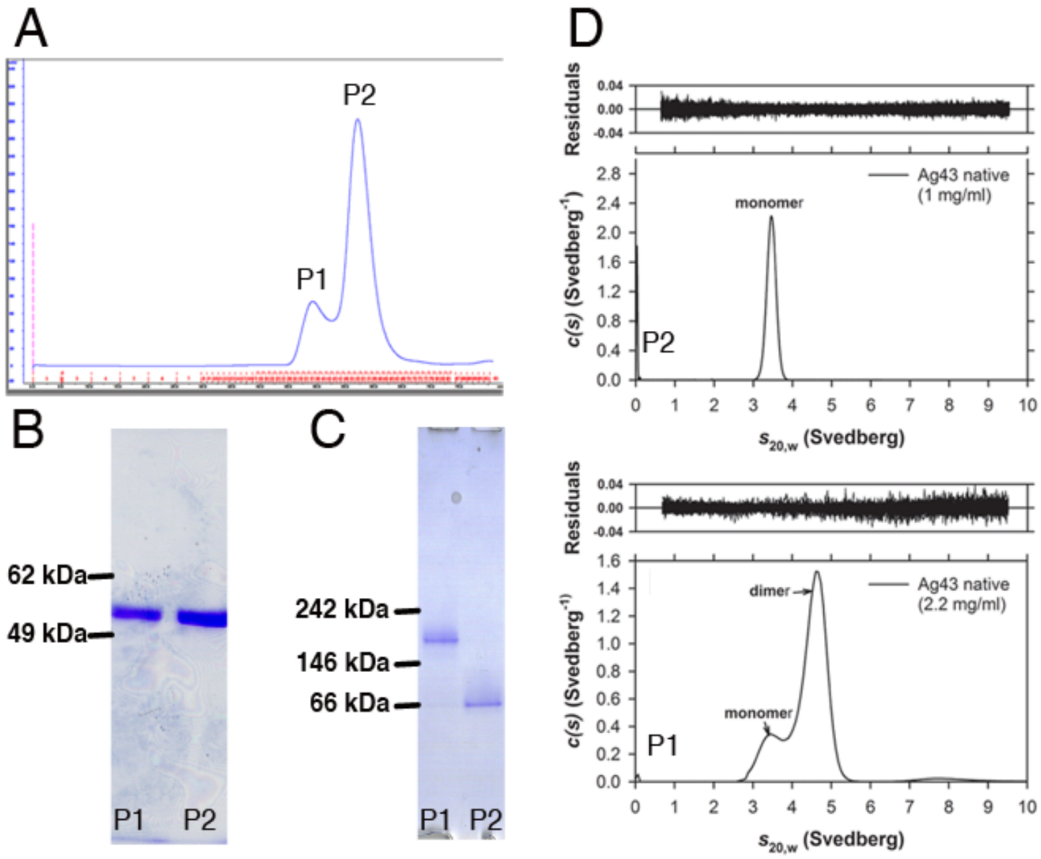


Figure S3. Oligomerization state of α^{43a} *in vitro* (*A*) Size-exclusion chromatography (SEC) profile of α^{43a} concentrated to 18 mg/ml on the Superdex-75 column (GE Healthcare). The concentrated protein eluted in two peaks with an apparent molecular weight of 150 and 60 kDa respectively. (*B*) SDS Polyacrylamide Gel Electrophoresis (SDS-PAGE) analysis of α^{43a} SEC samples. The fractions corresponding to peak 1 and 2 after SEC, were analyzed by SDS-PAGE followed by Coomassie brilliant blue staining. Both peaks consisted exclusively of α^{43a}, which migrates with an apparent molecular weight of 50 kDa. (*C*) Blue Native Polyacrylamide Gel Electrophoresis (BN-PAGE) analysis of α^{43a} SEC samples. Pre-cast NativePAGE Novex 4-16% Bis-Tris gels were used to analyze 2 micrograms of SEC peak 1 and peak 2 samples. Peak 1 showed the presence of two α^{43a} species corresponding to monomer and dimer α^{43a}, whereas peak 2 only consisted of the monomeric form. (*D*) Sedimentation Velocity-Analytical Ultracentrifugation (SV-AUC) analyses of α^{43a} SEC samples. Continuous standardized sedimentation coefficient [*c(s)*] distribution analyses of α^{43a} in peak 2 (top panel) and peak 1 (bottom panel) at an initial protein concentration of 1 and 2.2 mg/ml respectively. The resulting distribution shows that α^{43a} exist as monomer (3.4 S) in peak 2 and in peak 1 the protein self-associates to form dimers (4.6 S) and a small proportion of higher order oligomer (~8 S) consistent with a tetrameric species. Top of each panel show the residuals resulting from the *c(s)* distribution best fits plotted as a function of radial position from the axis of rotation. The *c(s)* distribution analyses were performed using the program SEDFIT with a sedimentation coefficient range of 0-10 S at a resolution of 200 and a P-value of 0.95. The r.m.s.d. and runs test Z values for all fits were <0.0048 and 10. The distribution has been modified to filter a non-equilibrium 1.8 S degradation product.

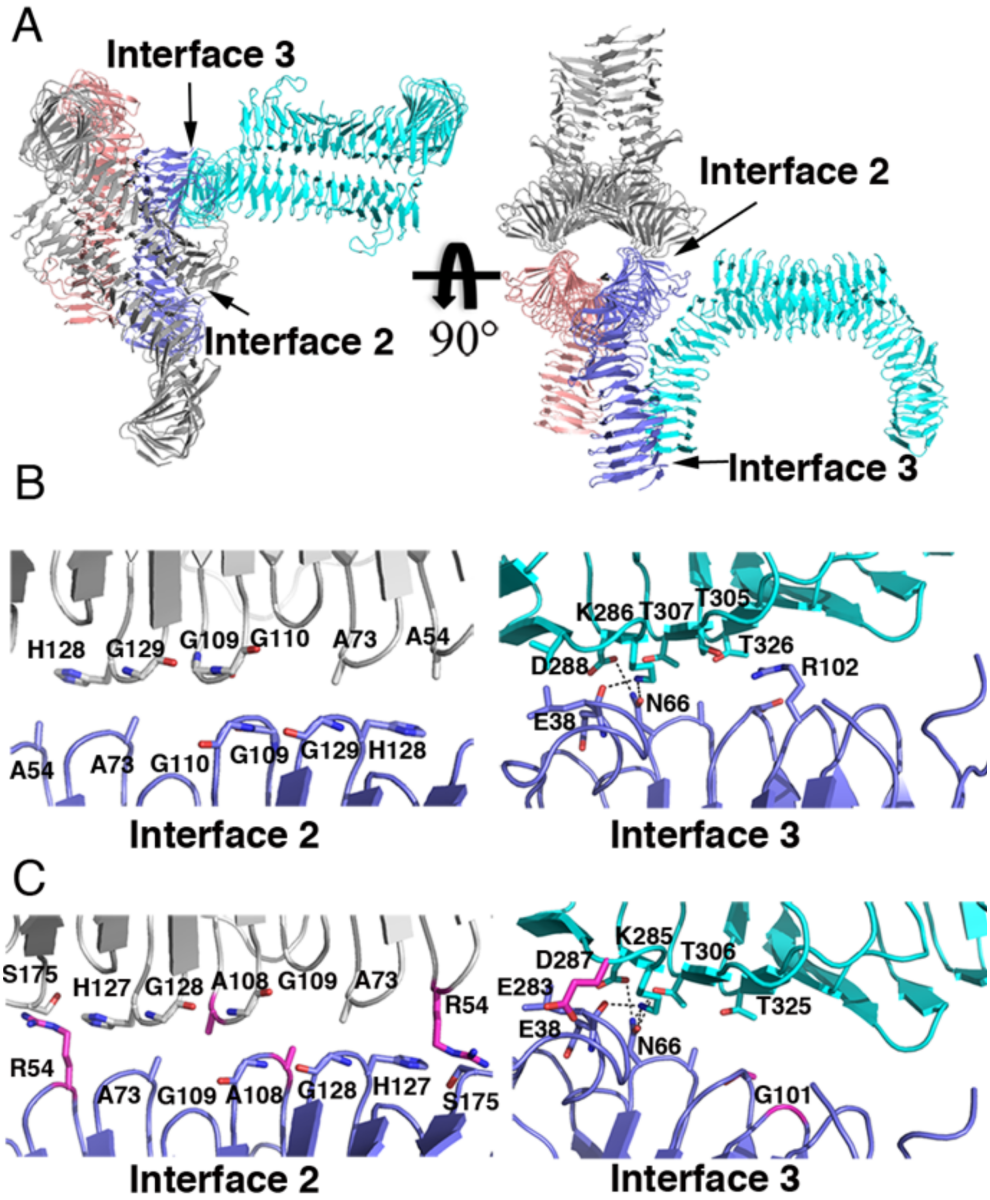


Figure S4. (*A*) Crystal packing of α^{43a} showing interface 2 and interface 3. (*B*) Close up view of α^{43a} interface 2 (left panel) and interface 3 (right panel). At Interface 2 the two α^{43a} molecules are *stabilized* solely by *van der Waals forces*. Interface 3 contains 3 hydrogen bonds (E38 (O) – K286 (NZ), N66 (ND2) – D288 (OD1), N66 (OD1) – K286 (NZ)) as well as *van der Waal* contacts. (*C*) Close up view of α^{43b} interface 2 (left panel) and interface 3 (right panel). The α^{43a} interface 2 and 3 residues that are not conserved in α^{43b} are colored in magenta. R54 in interface 2 and E283 in interface 3 could hinder the association α^{43b} molecules.

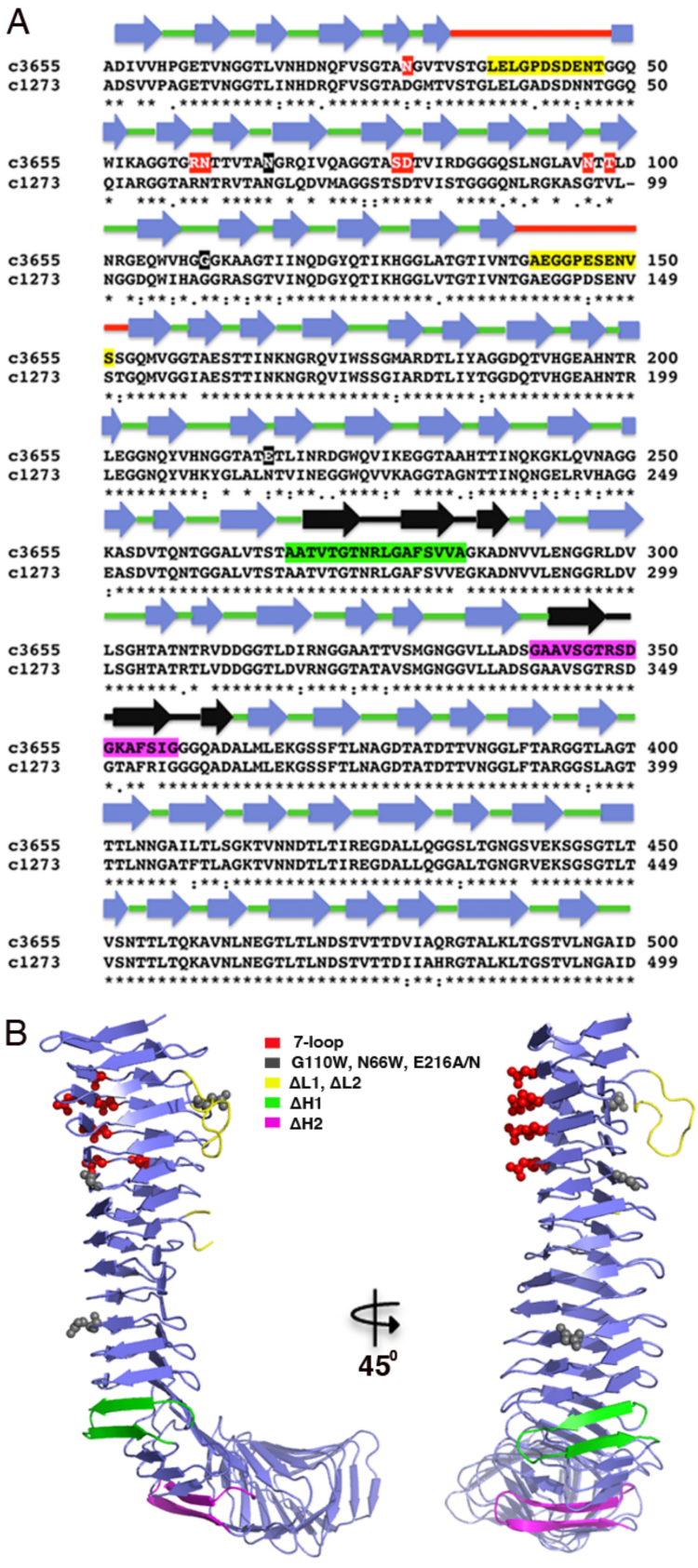


Figure S5. (*A*) Sequence alignment of the passenger domain of Ag43a (c3655) and Ag43b (c1273) using Tcoffee (13). Secondary structural elements are based on the α^{43a} structure. Identical and conserved residues are indicated with asterisks and points respectively. Residues modified in the different mutants are highlighted: red (7 loop mutant), yellow (ΔL1, ΔL2 and ΔL1,L2 deletion mutants), green (ΔH1 mutant), magenta (ΔH2 mutant) and gray (E216A, E216N, G110W and N66W mutants). (*B*) Ribbon representation of α^{43a} displaying all the residues truncated in this work.

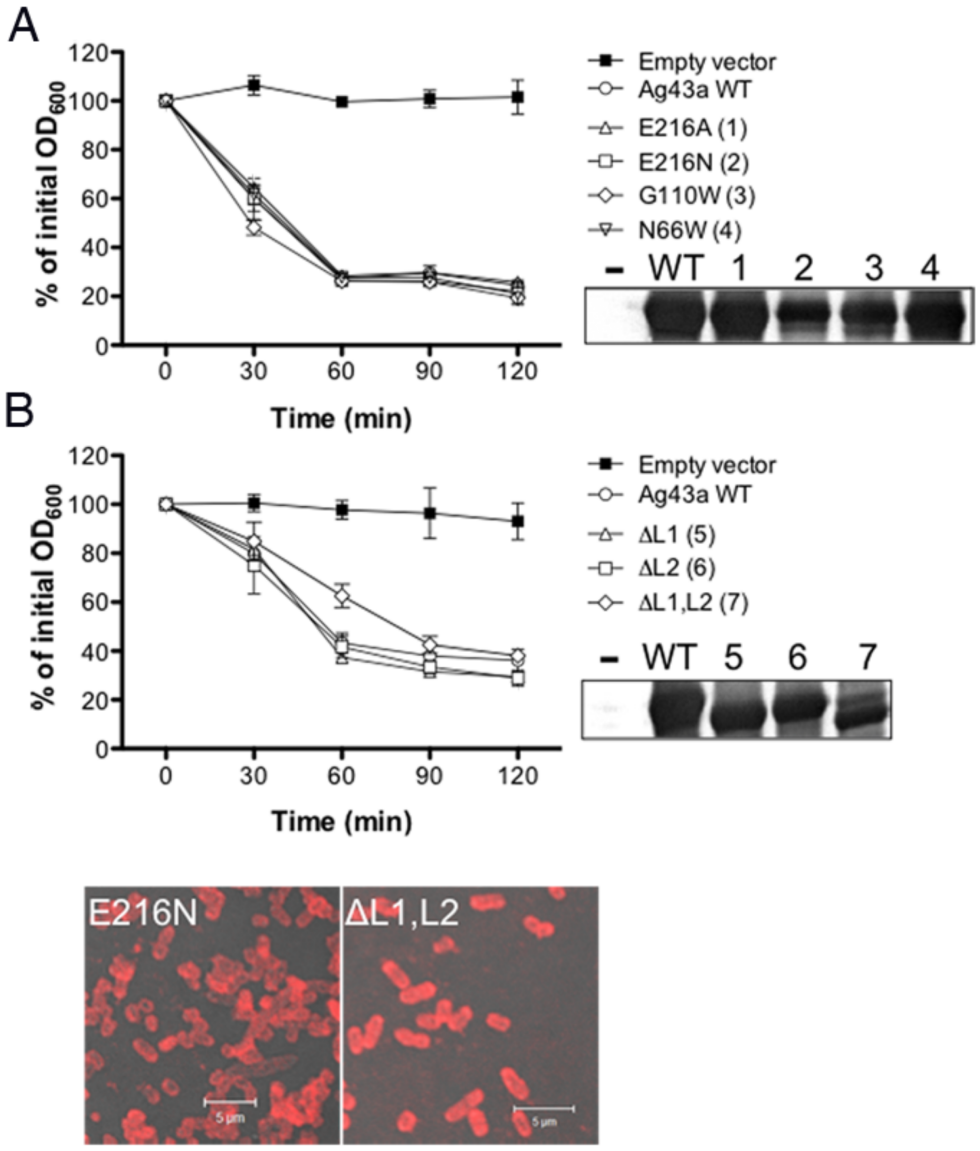
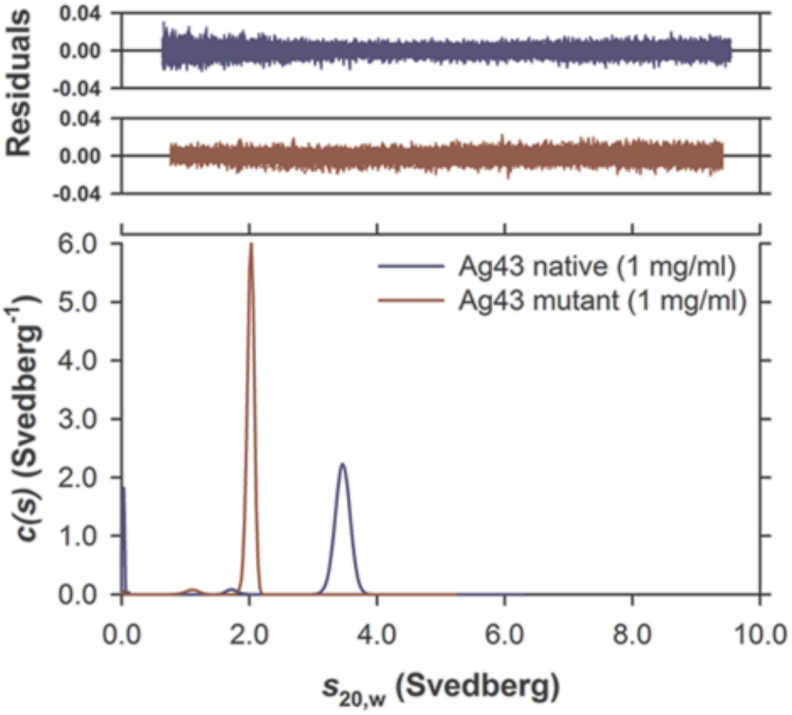


Figure S6. Functional characterization of additional Ag43a mutants. Cell-cell aggregation profiles of *E. coli* strains overexpressing (*A*) *ag43a* substitution mutants E216A, E216N, G110W and N66W and (*B*) *ag43a* loop deletion mutants ΔL1, ΔL2 and ΔL1,L2. Each panel of mutants was assessed together with isogenic control strains containing *ag43a* WT (positive control, shown in open circles) or empty vector (negative control, filled squares). Ag43a protein production at the bacterial cell surface was evaluated in Ag43a overexpressing cultures using an anti-Ag43a specific polyclonal serum by (*A, B insets*) heat-release of the α^{43a} from the cell surface and western blot analysis or (*C*) immunofluorescence labelling and confocal laser microscopy. α^{43a} was detected in all strains containing *ag43a*-encoding plasmids (WT and mutants 1-7) but not the empty vector control (-).

A



B

Sample	Oligomeric Species	$s_{20,w}$ (Svedberg)[1]	f/f_0[2]
α^{43a}	Monomer	3.4	1.5
$\alpha^{43a}_{\Delta H1, \Delta H2}$	Monomer	2.0	2.4

[1]Standardized sedimentation coefficient taken from the ordinate maximum of the c(s) distribution
[2]Frictional ratio

Figure S7. SV-AUC analyses of double β-hairpin deletion mutant α^{43a}ΔH1,H2. (A) Continuous standardized sedimentation coefficient [*c(s)*] distribution analyses of native α^{43a} (blue line) and α^{43a}ΔH1,H2 (red line) at an initial protein concentration of 1 mg/ml. The distributions show that both proteins exist as monomers but with significantly different $s_{20,w}$ values corresponding to 3.4 S (native) 2.0 S (mutant). Top panel shows residuals resulting from the *c(s)* distribution best fits plotted as a function of radial position from the axis of rotation. The *c(s)* distribution analyses were performed using the program SEDFIT with a sedimentation coefficient range of 0-10 S at a resolution of 200 and a P-value of 0.95. The r.m.s.d. and runs test Z values for both fits were 0.0045 and 9, respectively. (B) Table showing the hydrodynamic properties derived from SV-AUC of α^{43a} versus α^{43a}ΔH1,H2. Frictional ratio analyses show that the differences in $s_{20,w}$ values are due to shape changes, with the mutant protein forming a significantly more elongated monomer.

Table S1. Data collection and refinement statistics (SAD phasing)

	SeMeth-α^{43a}	**Native α^{43a}**
Data collection		
Temperature (K)	100	100
Wavelength (Å)	0.953663	0.954383
Space group	*P* 42_12	*P* 42_12
Distance (Å)	200	150
Cell dimensions		
a, *b*, *c* (Å)	106.9,106.9, 148.8	106.0,106.0, 147.5
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	50-2.81 (2.86-2.81)*	19.7-2.5 (2.64-2.50)*
R_{merge}	12.5 (47.6)	12.4 (57.5)
I / sI	20.39 (6.66)	17.3 (4.6)
Completeness (%)	99.6 (99.0)	99.8 (99.8)
Redundancy	24.7 (23.4)	14.2 (13.9)
Refinement		
Resolution (Å)		19.7-2.5
No. reflections		29742
R_{work} / R_{free}		16.76/21.56
No. atoms		
Protein		3493
malonate		49
Water		306
B-factors		
Protein/malonate		29.59
Water		37.08
R.m.s. deviations		
Bond lengths (Å)		0.007
Bond angles (°)		1.004
Ramachandran plot		
Residues in most favored/additionally allowed regions (%)		96.75/ 3.25

*The structure of α^{43a} was determined from one crystal of the selenomethionine derivative and one native crystal. *Values in parentheses are for highest-resolution shell

Table S2 Summary of Ag43a structural homologues present in the PDB

Organism	*Protein name*	*PDB code*	*Resolution (Å)*	*Domains*	*# of turns in β-helix*	*Ref.*	*Seq. id. (%)*[1]	*Rmsd (Å), # Cα*[1]
E. coli CT073	Ag43a α	4KH3	2.5	SAAT[2]	23		100	
Bordetella pertussis	Pertactin	1DAB	2.5	SAAT[2]	18	(14)	11.5	3.5, 260
Haemophilus influenzae	IgA1 protease (IgAP)	3H09	1.75	SAAT[2] Serine protease Domain 2[3]	24	(15)	10.8	4.6, 250
Haemophilus influenzae	Hap adhesin	3SYJ	2.2	SAAT[2] Serine protease	7	(16)	10.4	5.9, 287
Helicobacter pylori	Vacuolating Toxin p55 (VacA)	2QV3	2.4	SAAT[2]		(17)	7.8	4.2, 206
E. coli O157:H7	EspP	2SZE	2.5	SAAT[2] Serine protease	23	(18)	10.5	8.9, 228
E. coli strain EB1	Hemoglobin protease (Hbp)	1WXR	1.9	SAAT[2] Serine protease Domain 2[3]	24	(19)	8.8	9.3, 226

[1]Compared with Ag43a α domain (PDB code 4KH3) and calculated using SSM (20) within Coot (21).
[2]SAAT Self-associating autotransporter domain consisting of a β-helix
[3]Chitinase β-like domain with unknown function

Table S3 SAXS data collection and analysis details

Data Collection Parameters	
Instrument	SAXS-WAXS (Australian Synchrotron)
Beam geometry	Point
Wavelength (Å)	1.033
q-range (Å^{-1})	0.01-0.55
Exposure time (s)	10 (5 × 2 sec. exposures)
Protein concentration (mg/mL)	3.4
Temperature (K)	283
Standard	Water
Structural parameters	
$I(0)$ (cm^{-1}) [from $p(r)$]	0.1088 ± 0.0008
R_g (Å) [from $p(r)$]	34.5 ± 0.4
$I(0)$ (cm^{-1}) [from Guinier]	0.1101 ± 0.0009
R_g (Å) [from Guinier]	34.1 ± 0.4
D_{max} (Å)	120 ± 5
Porod volume (Å^3)	62000 ± 1500
R_g (Å) [crystal structure]	33.9
D_{max} (Å) [crystal structure]	115
Dry volume (Å^3) [crystal structure]	60700
Molecular mass determination	
Partial specific volume (cm^{3}g^{-1})	0.726
Contrast, $\Delta\rho$ (10^{10} cm^{-2})	3.065
Molecular mass M_r [from $I(0)$]	40000 ± 2000
Molecular mass M_r [from Porod volume]*	49300 ± 1000
Molecular mass M_r [crystal structure]	50100
Software employed	
Primary data reduction	*SAXS15ID*
Data processing	*PRIMUS and GNOM*
Ab initio modelling	*DAMMIN*
Validation and averaging	*DAMAVER*
Computation of model intensities	*CRYSOL*
Three-dimensional graphics representations	*PyMOL*

*Molecular mass determined using the approach outlined by Fischer et al. (3)

Table S4 Primers used in this study

Primer name	Primer sequence (5'- 3')
Agn43 LIC F	tacttccaatccaatgcggctgacatcgttg
Agn43 LIC R	ttatccacttccaatgttagtcaatggcaccg
Agn43 SacI F	ttagcggatcctacctgacg
Agn43 KpnI R	agcgtgaatgaactgccttt
Agn43 StuI R	tggcaaagcaaaagttcaaa
Agn43 L1 R	cattgcccgcccccggtactgacagtcacg
Agn43 L1 F	cagtaccgggggcgggcaatggataaaa
Agn43 L2 R	tctgaccgctaccggtgttgacgatggtt
Agn43 L2 F	caacaccggtagcggtcagatggtcgg
Agn43 H1 R	cagccttacctgtactggtaaccagtgcacc
Agn43 H1 F	taccagtacaggtaaggctgataatgtcgtactg
Agn43 H2 R	gcctgaccgccggaatcggccagcagtaca
Agn43 H2 F	ggccgattccggcggtcaggcggatgc
Agn43 Seq #1 F	acggataccacggtaaatgg
Agn43 Seq #2 F	atggcacagaacaatgctga