Supplemental information

Ligand-binding properties and conformational dynamics of autolysin repeat domains in staphylococcal cell wall recognition

Running title: Structure and function of major Atl repeats

Sebastian Zoll[1#], Martin Schlag[2#], Alexander V. Shkumatov[3#], Maren Rautenberg,[4] Dmitri I. Svergun[3], Friedrich Götz[2], Thilo Stehle[1,5*]

[#] Contributed equally to this publication

*Address correspondence to Thilo Stehle, thilo.stehle@uni-tuebingen.de

[1]Interfaculty Institute of Biochemistry, University of Tuebingen, Tuebingen, Germany, [2]Department of Microbial Genetics, Faculty of Biology, University of Tuebingen, Tuebingen, Germany, [3]European Molecular Biology Laboratory, Hamburg Outstation, EMBL c/o DESY, Hamburg, Germany, [4]Department of Medical Microbiology and Hygiene, Faculty of Biology, University of Tuebingen, Tuebingen, Germany, [5]Department of Pediatrics, Vanderbilt University School of Medicine, Nashhville, TN 37232, USA

SAXS modeling

Low-resolution shapes were reconstructed *ab initio* from the scattering data using DAMMIN. This program represents the particle by an assembly of densely packed spheres and employs simulated annealing to construct a compact interconnected model fitting the experimental data to minimize the discrepancy:

$$\chi = \sqrt{\frac{1}{(N-1)}\sum_{j}\left[\frac{\left(I_{\exp}(S_j) - cI_{calc}(S_j)\right)}{\left(?(S_j)\right)}\right]^2} \quad (1)$$

where N is the number of experimental points, c is a scaling factor, $I_{exp}(s)$, $I_{calc}(s)$ and $\sigma(s_j)$ are the experimental intensity, the calculated intensity and experimental error at the momentum transfer s_j, respectively. Higher-resolution *ab initio* models were constructed using GASBOR, which models the particle in solution as a protein-like assembly of dummy residues and represents the internal structure more accurately than DAMMIN. Multiple DAMMIN and GASBOR calculations were performed to assess the stability of resulting solutions. Ten to fifteen independent reconstructions were performed and the models were averaged with the program DAMAVER (17), which provided a value of Normalized Spatial Discrepancy (NSD). NSD values close to one indicate that the two models are similar. DAMAVER generates an average model of the set of superposed structures and also outputs the most typical one (i.e. that having the lowest average NSD with all other models in the set). The average models were further refined using DAMMIN.

High resolution crystal structures of two domains (AM, R2ab) as well as a homology model (R1ab) were used for a combined *ab initio*/rigid-body modeling with BUNCH (13). The rigid-body models were generated for AM-R1ab-R2ab, AM-R1ab and R1ab-R2ab. Starting from a random domain arrangement, BUNCH uses simulated annealing to guide the translations and rotations of domains to minimize the discrepancy, χ between the experimental and calculated data (Eq. 1) while maintaining chain connectivity without steric clashes. The missing linkers between the individual subunits are modeled using dummy residues, starting from a random initial configuration generated by PRE_BUNCH. For AM-R1ab-R2ab, either a single scattering curve was fitted or multiple curves (AM-R1ab-R2ab and R1ab-R2ab) were fitted simultaneously, as the SAXS-based modeling is most informative when one concurrently uses the scattering patterns from several deletion mutants, provided they retain its overall structure as a part of the larger construct (13). Ten BUNCH runs were performed, yielding stable and consistent rigid body models for each sample. These models were clustered using DAMCLUST and DAMAVER (9).

The scattering patterns from AM-R1ab-R2ab, AM-R1ab and R1ab-R2ab were analyzed using the Ensemble Optimization Method (EOM), which takes the

flexibility into account by allowing for the coexistence of multiple conformations in solution in order to fit the experimental SAXS data (1). EOM selects appropriate ensembles of configurations from large pools of random models of the protein. In the first step, RANCH (1) generates a pool of models (typically 10.000) with random arrangement of high resolution structures of individual domains connected by modeled linkers. The theoretical scattering curve is then calculated for each model by CRYSOL (16). In the second step, a genetic algorithm (GAJOE) (1) selects subsets of scattering curves (and hence models) such that the averages over the ensembles fits the experimental data. The goodness of fit, for each individual experimental curve, is characterized by the discrepancy (Eq. 1). This goodness of fit guides the genetic algorithm by allowing only the ensembles with the best scores to proliferate to the next generations of the genetic algorithm.

With the assumption that the R1ab-R2ab retains its overall structure as a part of AM-R1ab-R2ab, simultaneous fitting of two experimental data sets was also performed. Synchronous fitting of multiple scattering patterns from deletion mutants, if available, provides yet more detailed local information about the structure (1).

Multiple runs of GAJOE with default parameters either with single or multiple curves were performed and the obtained subsets analyzed to yield the R_g distributions in the optimal ensembles. In order to check the minimum number of structures in the ensemble required to fit the experimental data, GAJOE was run with default parameters, except for number of curves per ensemble, which was set to 3.

SUPPLEMENTAL TABLE 1. Strains and plasmids.

Strains and plasmids	Description	Source or reference
E. coli		
DH5α	F- φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (rk-, mk+) *gal- pho*A *sup*E44 λ- *thi*-1 *gyr*A96 *rel*A1	Invitrogen
BL21 (DE3)	*huA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*	New England Biolabs
S. epidermidis		
SE O-47	Clinical Isolate	(6)
SE mut1	*atl* transposon mutant of SE O-47	(5)
S. aureus		
SA113	Derivative of NCTC 8325	(7)
RN4220	Derivative of NCTC 8325	(10)
RN4220 (4S5)	RN4220 *spa*/*ltaS* suppresor	(3)
Plasmids		
pGEX-AM	Overexpression plasmid for GST-AM fusion protein in *E. coli*; Amp^R	(11)
pGEX-R1ab-R2ab	Overexpression plasmid for GST-R1ab-R2ab fusion protein in *E. coli*; Amp^R	(15)
pET52b-AM-R1ab	Overexpression plasmid for Strep-AM-R1ab fusion protein in *E. coli*; Amp^R	This work
pGEX-AM-R1ab-R2ab	Overexpression plasmid for GST-AM-R1ab-R2ab fusion protein in *E. coli*; Amp^R	This work
pGEX-mutR1a/R2a	Overexpression plasmid for expression of mutant protein mutR1a/R2a in *E. coli*; Amp^R	This work
pGEX-mutR1b/R2b	Overexpression plasmid for expression of mutant protein mutR1b/R2b in *E. coli*; Amp^R	This work

pRC20	Shuttle plasmid for expression of atlE-amidase; Amp^R in *E. coli* Cm^R in *Staphylococcus* sp.	(4)
pRC-mutR1a/R2a	Shuttle plasmid for expression of mutant protein mutR1a/R2a; Amp^R in *E. coli* Cm^R in *Staphylococcus* sp.	This work
pRC-mutR1b/R2b	Shuttle plasmid for expression of mutant protein mutR1b/R2b; Amp^R in *E. coli* Cm^R in *Staphylococcus* sp.	This work

SUPPLEMENTAL TABLE 2. Data collection and structure refinements statistics. Values in parentheses correspond to the highest resolution bin.

	Native	**$HgCl_2$**	**K_2PtCl_4**
Data collection			
Space group	$P6_122$	$P6_122$	$P6_122$
Cell dimensions (Å)	95.36, 95.36, 233.67	94.07, 94.07, 233.21	95.01, 95.01, 231.87
Wavelength (Å)	0.934	1.5418	1.5418
Resolution (Å)	30.0-2.90 (2.98-2.90)	30.0-3.70 (3.79-3.70)	30.0-4.1 (4.21-4.10)
R_{sym}* (%)	6.5 (38.8)	15.9 (49.3)	25.2 (45.1)
I/σI	15.3 (2.6)	15.3 (5.6)	12.7 (6.3)
Completeness (%)	98.9 (99.8)	98.9 (98.8)	99.1 (100.0)
Redundancy	3.6 (3.7)	13.1 (12.8)	11.2 (10.2)
Total refelections	51923	160036	102264
Unique reflections	14514	12231	9113
Refinement			
Resolution (Å)	25.7-2.9		
R_{work}/R_{free}** (%)	26.3/29.8		
N(mol) ASU	1		
Solvent content (%)	70.0		
Vm ($Å^3$/Da)	4.1		
Number of atoms			
Protein	1217		
Water	18		
B-factors ($Å^2$)			
Protein	55.6		
Water	43.9		
r.m.s deviations			
Bond lengths(Å)	0.01		
Bond angles (°)	1.23		
Ramachandran plot (%)			
Most favored region	98.6		
Additionally allowed region	1.4		
Clashscore, all atoms^	0.83		
	#100th percentile (N=97, 2.90Å ± 0.25Å)		
MolProbity score^^	1.24		
	#100th percentile (N=3760, 2.90Å ± 0.25Å)		

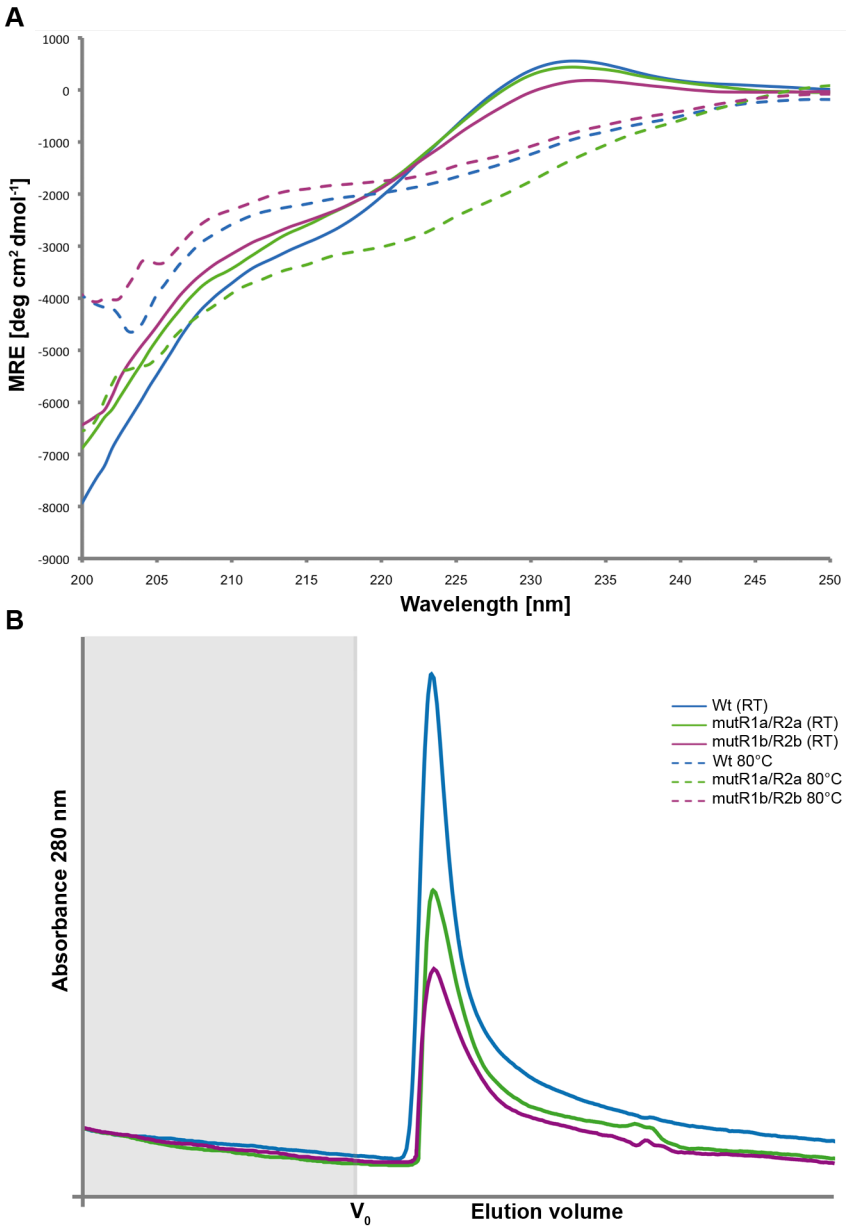
* $R_{sym} = \Sigma \mid I - \langle I \rangle \mid / \Sigma I$

** R-factor = $\Sigma| |F_{obs}(hkl)| - |F_{calc}(hkl)| | / \Sigma |F_{obs}(hkl)$

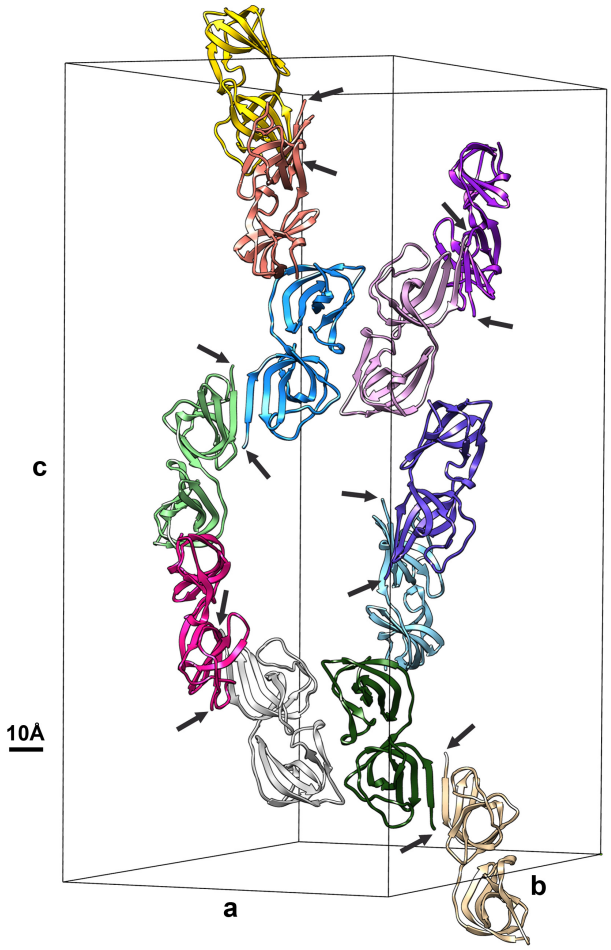
100th percentile is the best among structures of comparable resolution; 0th percentile is the worst.

^ Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms

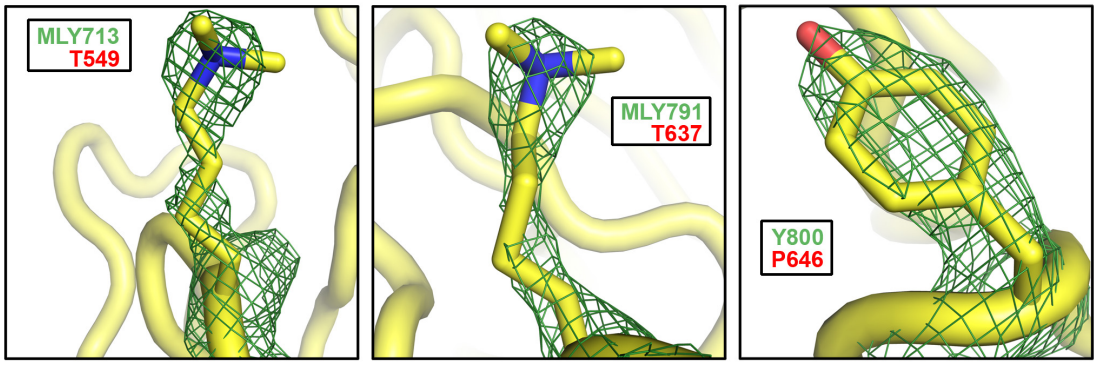
^^ MolProbity score is defined as the following: 0.42574*log(1+clashscore) + 0.32996*log(1+max(0,pctRotOut-1)) + 0.24979*log(1+max(0,100-pctRama Favored-2)) + 0.5 (2)



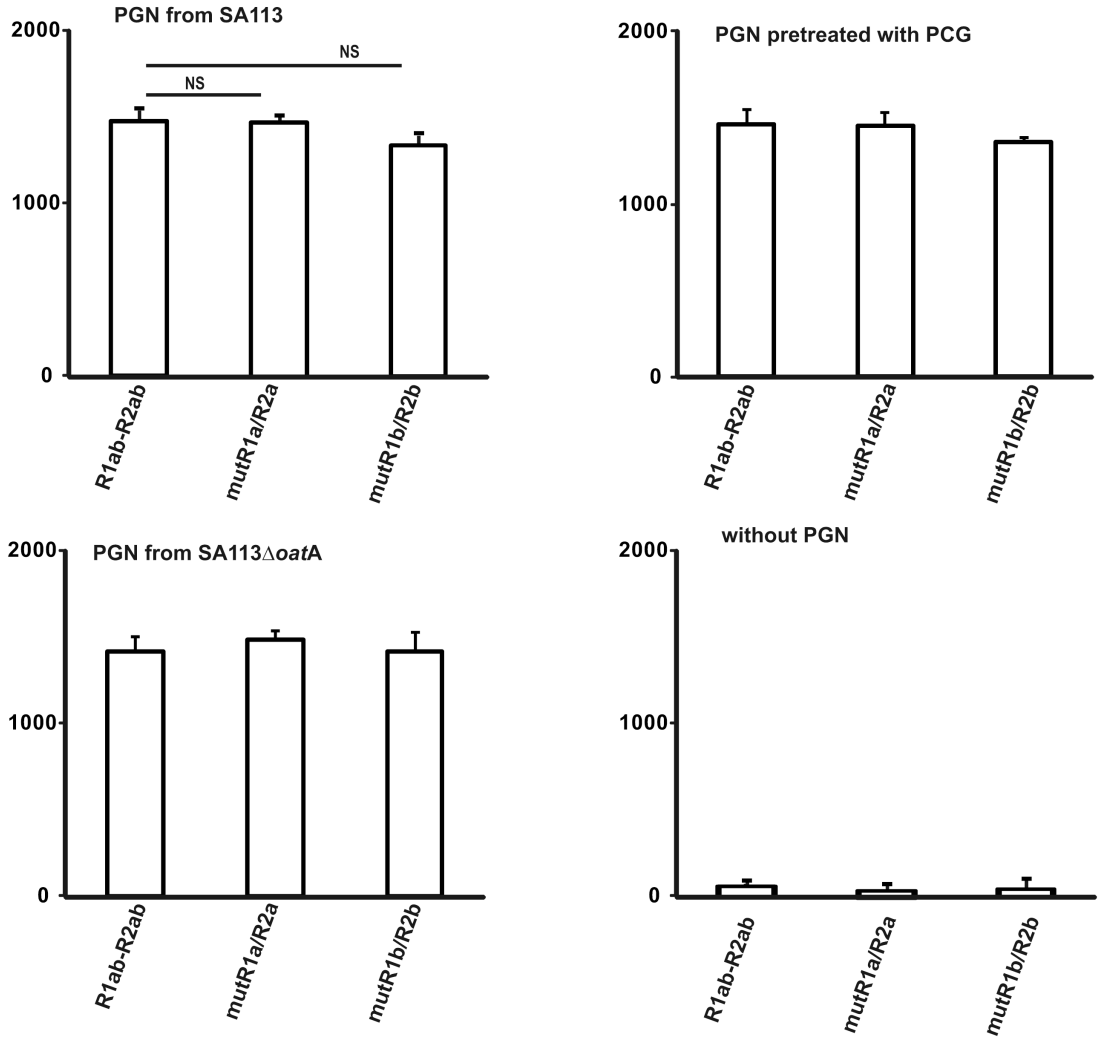
Supplemental Figure S1. Surface mutations do not alter the structure of R1abR2ab. (A) Far-UV circular dicroism (CD) spectra of mutR1a/R2a and mutR1b/R2b closely resemble the Wt spectrum at RT. Mutants and the Wt all have a maximum at 232 nm that is characteristic for β-strands (8, 12, 14). Increasing the temperature to 80°C results for all proteins in a loss of the maximum which indicates unfolding. Proteins were measured at a concentration of 0.2 mg/ml in 6 mM phosphate buffer, 70 mM NaCl, pH 7.5 on a J-720 spectropolarimeter (Jasco) equipped with a 10 mm pathlength cell. For each spectrum 10 runs were accumulated at an acquisition speed of 50 nm/min. MRE= mean residue ellipticity (B) Repeat mutants and Wt elute from a gel filtration column at the same volume. 25 µl of purified mutR1a/R2a, mutR1b/R2b and Wt were applied to a Superdex 75 (3.2/30) analytical gel filtration column installed on an Ettan chromatography system (GE Healthcare). The running buffer was 50 mM Tris, 150 mM NaCl, pH 8.0.



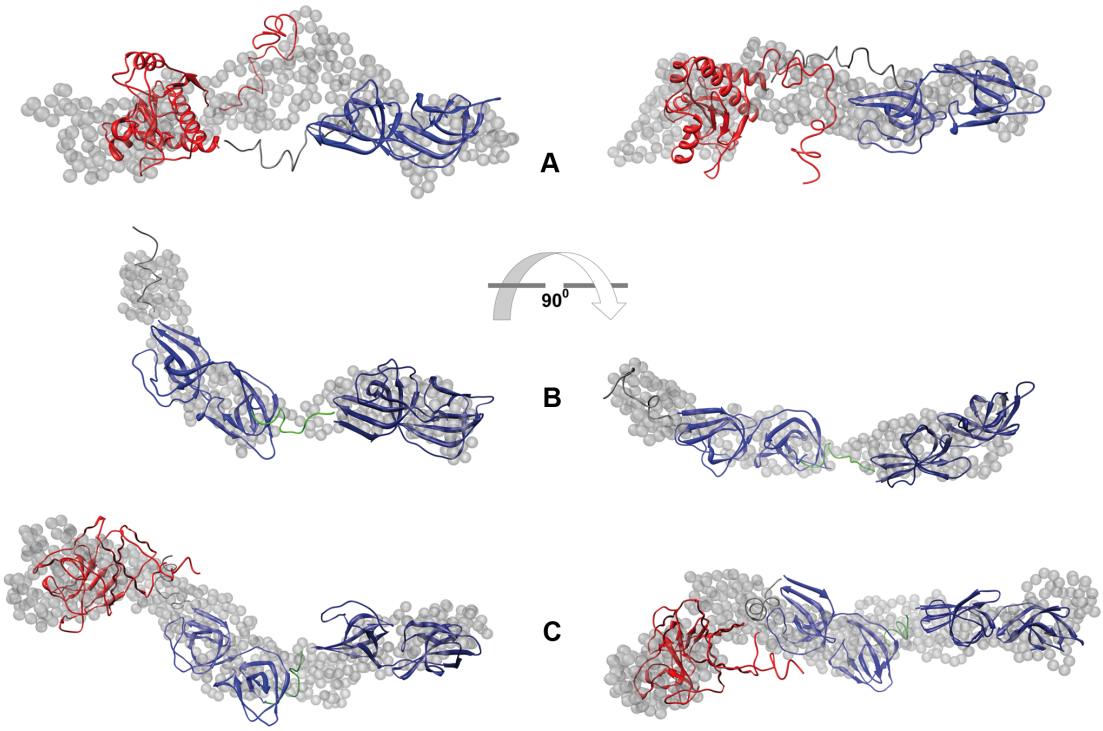
Supplemental Figure S2. R2ab is sufficient to explain the crystal lattice in space-group $P6_122$. Arrows indicate the position of linker L2. L2 and R1ab protrude into a solvent channel and are not visible in the electron density map.



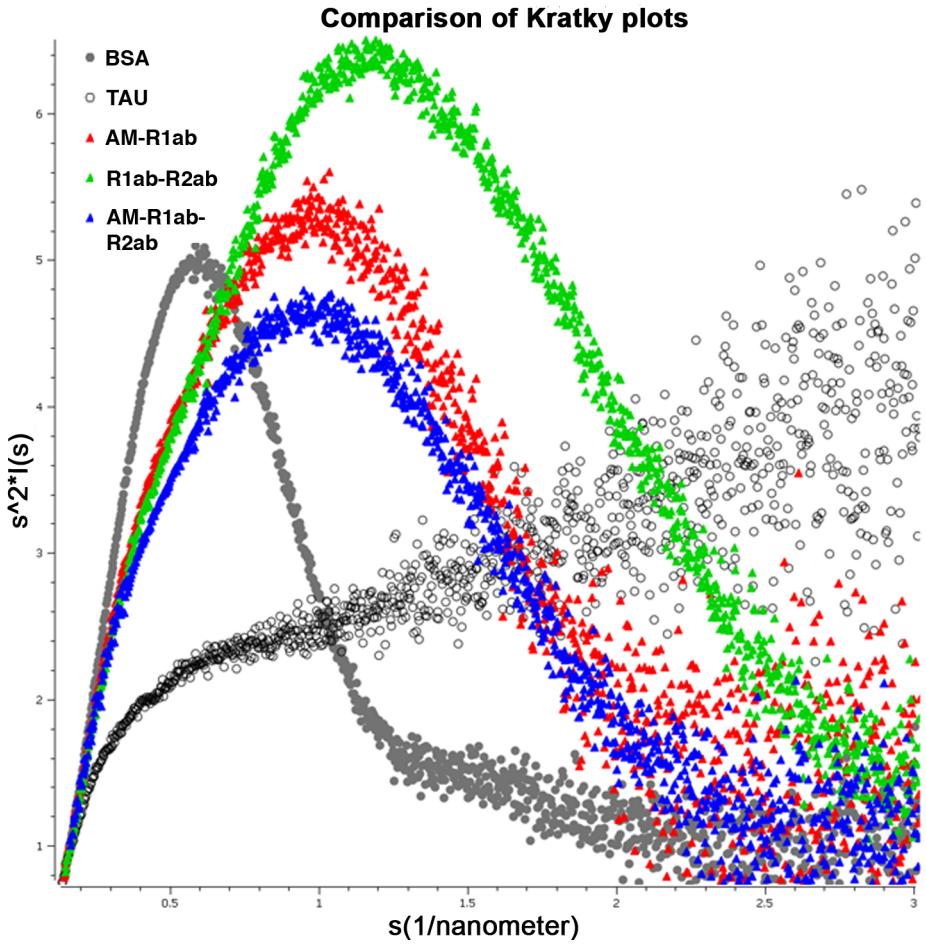
Supplemental Figure S3. Electron densities of particularly well resolved amino acids in R2ab.The characteristic side chains of amino acids MLY713, MLY791 and Y800 (green labels) show well defined electron density and were, amongst others, used to unambiguously confirm the identity of R2ab in the final electron density map. Equivalent positions in R1ab would be occupied by amino acids with significantly shorter or no side chains, respectively (red labels). Simulated annealing difference density omit map contoured at 3.0 σ for 1.6 Å around selected side chains. MLY= di-methylated lysine.



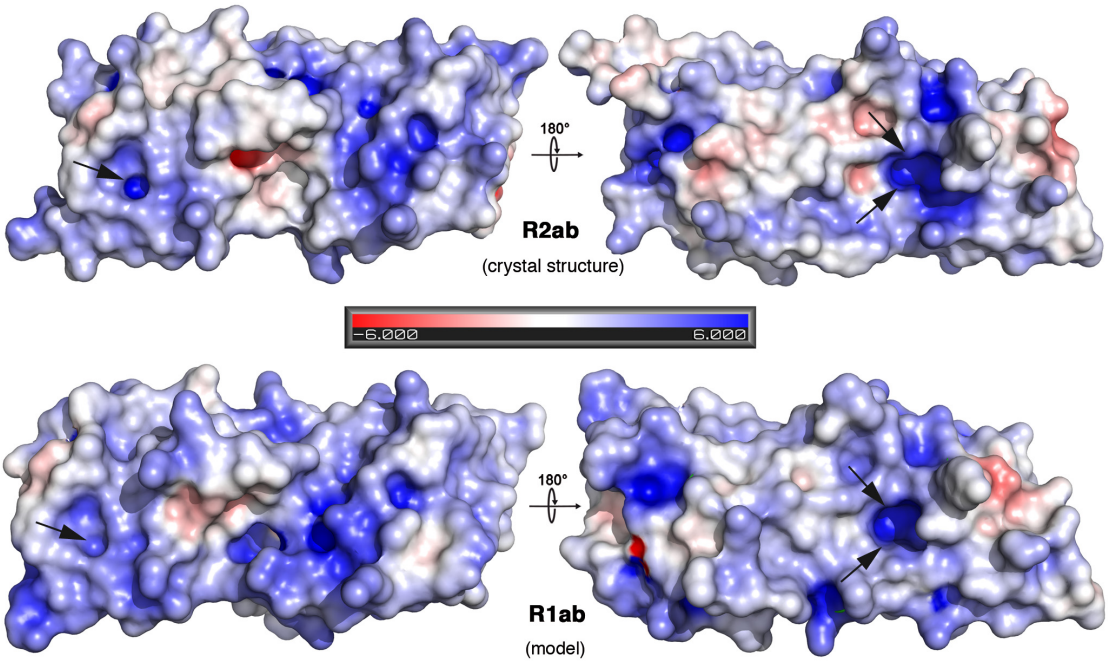
Supplemental Figure S4. Binding of repeat mutants to PGN. Wt repeats and mutants were incubated with purified, untreated PGN (upper left), PGN purified from PCG-treated cells (upper right) and PGN from the *S. aureus* SA113 Δ*oat*A mutant that lacks O-acetylation in the PGN sugar backbone (lower left). To rule out precipitation of protein or unspecific binding, protein without PGN was used as negative control (lower right). Bound proteins were pulled down and separated by SDS-PAGE. The amount of bound protein was determined from the band intensity. No significant differences (NS) could be observed for repeats that carry mutations in the putative binding grooves.



Supplemental Figure S5. Rigid body and *ab initio* models of Atl amidase variants. **(**A) AM-R1ab (B) R1ab-R2ab and (C) AM-R1ab-R2ab. *Ab initio* models are depicted using transparent grey beads. Folded domains of AM-R1ab-R2ab and modeled linkers are shown as cartoons (AM – red, R1ab – blue, R2ab – dark blue, linker between AM and R1ab – grey, linker between R1ab and R2ab – green).



Supplemental Figure S6. Kratky plots. Comparison of Kratky plots of AM-R1ab (*red triangle*), R1ab-R2ab (*green triangle*), and AM-R1ab-R2ab (*blue triangle*), with bovine serum albumin (*filled grey circle*), and protein tau (*open black circle*).



Supplemental Figure S7. Electrostatic surface potential of R2ab and R1ab in two different views. Arrows mark sites where point mutations were introduced. The conserved regions in R2ab and R1ab are also areas with a positive electrostatic potential. Rab and R1ab have the orientation as in figure 4.

REFERENCES

1. **Bernado, P., E. Mylonas, M. V. Petoukhov, M. Blackledge, and D. I. Svergun.** 2007. Structural characterization of flexible proteins using small-angle X-ray scattering. J Am Chem Soc **129:**5656-64.
2. **Chen, V. B., W. B. Arendall, 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, and D. C. Richardson.** MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr **66:**12-21.
3. **Corrigan, R. M., J. C. Abbott, H. Burhenne, V. Kaever, and A. Gründling.** 2011. c-di-AMP is a new second messenger in *Staphylococcus aureus* with a role in controlling cell size and envelope stress. PLoS Pathog **7:**e1002217.
4. **Heilmann, C., C. Gerke, F. Perdreau-Remington, and F. Gotz.** 1996. Characterization of Tn917 insertion mutants of Staphylococcus epidermidis affected in biofilm formation. Infect Immun **64:**277-82.
5. **Heilmann, C., C. Gerke, F. Perdreau-Remington, and F. Götz.** 1996. Characterization of Tn917 insertion mutants of *Staphylococcus epidermidis* affected in biofilm formation. Infect Immun **64:**277-82.
6. **Heilmann, C., J. Hartleib, M. S. Hussain, and G. Peters.** 2005. The multifunctional Staphylococcus aureus autolysin aaa mediates adherence to immobilized fibrinogen and fibronectin. Infect Immun **73:**4793-802.
7. **Iordanescu, S., and M. Surdeanu.** 1976. Two restriction and modification systems in Staphylococcus aureus NCTC8325. J Gen Microbiol **96:**277-81.
8. **Kidric, M., H. Fabian, J. Brzin, T. Popovic, and R. H. Pain.** 2002. Folding, stability, and secondary structure of a new dimeric cysteine proteinase inhibitor. Biochem Biophys Res Commun **297:**962-7.
9. **Konarev, P. V., M. V. Petoukhov, V. V. Volkov, and D. I. Svergun.** 2006. ATSAS 2.1, a program package for small-angle scattering data analysis. J. Appl. Crystallogr. **39:**277-286.
10. **Kreiswirth, B. N., S. Lofdahl, M. J. Betley, M. O'Reilly, P. M. Schlievert, M. S. Bergdoll, and R. P. Novick.** 1983. The toxic shock syndrome exotoxin structural gene is not detectably transmitted by a prophage. Nature **305:**709-12.
11. **Lützner, N., B. Patzold, S. Zoll, T. Stehle, and H. Kalbacher.** 2009. Development of a novel fluorescent substrate for Autolysin E, a bacterial type II amidase. Biochem Biophys Res Commun **380:**554-8.
12. **Neu, U., M. S. Maginnis, A. S. Palma, L. J. Stroh, C. D. Nelson, T. Feizi, W. J. Atwood, and T. Stehle.** Structure-function analysis of the human JC polyomavirus establishes the LSTc pentasaccharide as a functional receptor motif. Cell Host Microbe **8:**309-19.
13. **Petoukhov, M. V., and D. I. Svergun.** 2005. Global rigid body modelling of macromolecular complexes against small-angle scattering data. Biophys J **89:**1237-1250.
14. **Pohleven, J., M. Renko, S. Magister, D. F. Smith, M. Kunzler, B. Strukelj, D. Turk, J. Kos, and J. Sabotic.** 2012. Bivalent Carbohydrate Binding Is Required for Biological Activity of Clitocybe

nebularis Lectin (CNL), the N,N'-Diacetyllactosediamine (GalNAcbeta1-4GlcNAc, LacdiNAc)-specific Lectin from Basidiomycete C. nebularis. J Biol Chem **287:**10602-12.

15. **Schlag, M., R. Biswas, B. Krismer, T. Kohler, S. Zoll, W. Yu, H. Schwarz, A. Peschel, and F. Gotz.** 2010. Role of staphylococcal wall teichoic acid in targeting the major autolysin Atl. Mol Microbiol **75:**864-73.
16. **Svergun, D. I., C. Barberato, and M. H. J. Koch.** 1995. CRYSOL - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. J. Appl. Crystallogr. **28:**768-773.
17. **Volkov, V. V., and D. I. Svergun.** 2003. Uniqueness of ab initio shape determination in small-angle scattering. J. Appl. Cryst. **36:**860-864