Supplementary Material for

Structural basis of lipoprotein signal peptidase II action and inhibition by the antibiotic globomycin

Lutz Vogeley, Toufic El Arnaout, Jonathan Bailey, Phillip J. Stansfeld, Coilin Boland, Martin Caffrey*

*Corresponding author. E-mail: martin.caffrey@tcd.ie

This PDF file includes:

Materials and Methods
Supplementary Text
Figs. S1 to S9
Full Reference List

Other Supplementary Material for this manuscript includes the following:
(available at www.sciencemag.org/content/351/6275/876/suppl/DC)

Movies S1 to S5
Database S1as a separate Excel file

Materials and Methods

<u>LspA expression and purification</u>

The wild-type *lspA* gene from *P. aeruginosa,* strain PAO1, was cloned into the expression vector pET28a using restriction sites *Nde*I and *Xho*I (*6*). The recombinant construct was the wild-type *P. aeruginosa* PAO1 LspA modified N-terminally with an MGSS sequence followed by a 6 x His-tag and the thrombin cleavage sequence SSGLVPRGSH. Sequencing (MWG Biotech) confirmed the identity of the recombinant gene. The pET28a-*lspa* recombinant plasmid was transformed into chemically competent *E. coli* C41 (DE3) (NEB) cells. Cells were grown in LB broth (Sigma) supplemented with 50 mg/L kanamycin (Melford) at 37 ^{o}C to an OD_{600} of 0.5-0.6 and induced with 1 mM isopropyl ß-D-1-thiogalactopyranoside (IPTG) (Sigma) at 30 ^{o}C and 180 rpm for 18 h. Cells were harvested by centrifugation at 5,000*g* for 10 min at 4 ^{o}C. Cells were either used immediately or stored at -80 ^{o}C for a maximum of 2 months.

For SeMet labelling of LspA the methionine auxotroph *E. coli* strain B834 (DE3) (Novagen) was transformed with the pET28a-*lspa* plasmid. SeMet medium was made using the SelenoMet™ kit (Molecular Dimensions) according to the manufacturer's instructions. The medium was supplemented with 50 mg/L SeMet (Sigma) and 50 mg/L kanamycin. Because expression levels were low production of biomass was carried out on a scale of 20 L. Biomass production and protein purification of the SeMet form of LspA were as described for unlabelled enzyme.

LspA cell mass was re-suspended at 1 g/mL Buffer A (50 mM MES pH 6.15, 150 mM NaCl, 10 %(v/v) glycerol) at room temperature (RT, 20-22 ^{o}C) and a single Complete Protease Inhibitor Cocktail tablet (Roche) was added per 50 mL of resuspended material. Cells were lysed by three passages through a high pressure homogenizer (Emulsiflex-C5; Avestin). Cell debris was removed by centrifugation at 24,000*g* for 12 min at 4 ^{o}C. A membrane fraction was separated from the supernatant by ultra-centrifugation at 150,000*g* for 45 min at 4 ^{o}C. The membrane pellet was resuspended in 100 mL Buffer A containing 1 %(w/v) fos choline-12 (FC12) (Anatrace) and solubilisation was carried out for 1 h at 4 ^{o}C with agitation using a magnetic stirrer. Unsolubilized material was separated by ultra-centrifugation at 150,000*g* for 30 min at 4 ^{o}C. The supernatant to which 20 mM imidazole was added was incubated with 5 mL Ni-NTA Superflow resin (Qiagen) for 1 h at 4 °C on a rotor circulating at 20 rpm. The resin was added to a BioRad Econo Column (4 cm diameter) and washed with 150 mL Buffer B (50 mM MES pH 6.15, 150 mM NaCl, 10 %(v/v) glycerol, 0.14 %(w/v) FC12) containing 50 mM imidazole. Bound protein was eluted with 15 mL Buffer B containing 300 mM imidazole. 1 mL fractions were collected and protein concentration was determined in each by absorbance at 280 nm (A_{280}) measured with a Nanodrop 1000 (Wilmington, DE, USA). Fractions containing ≥ 1 mg/mL LspA were combined and concentrated to a volume of 1.5 mL using a 15 mL Millipore centrifuge filter with a molecular weight cut off of 50 kDa. The concentrated sample was loaded onto a HiLoad 16/600 column packed with Superdex 200 size-exclusion resin pre-equilibrated with 180 mL Buffer B at 4 °C using a GE Healthcare ÄKTA Purifier system. Size-exclusion chromatography was performed at 4 °C at a flow rate of 0.5 mL/min. LspA eluted as a single, symmetrical A_{280} peak at 76 mL. Peak fractions with

A_{280} values larger than half the peak maximum were combined and concentrated to 10 - 30 mg/mL using a 4 mL Millipore centrifuge filter with a molecular weight cut off of 50 kDa. The pure protein was aliquoted into 20 μL fractions, flash frozen in liquid nitrogen and stored at -80 ^{o}C. The identity of the protein was confirmed as LspA by mass-spectrometry.

LspA mutant genes were obtained either by site-directed mutagenesis of the wild-type *lspA* construct in pET28a (D115A, R116A, D143A, D143N) or as synthetic genes cloned into pET28a (D115N, D124N) (GenScript). That the correct mutations had been made was confirmed by sequencing (MWG Biotech). Expression and purification of all LspA mutants was performed according to the wild-type protocol.

Native and SeMet-labelled LspA was also produced using an *E. coli*-based cell-free expression system, mainly to facilitate the production of the SeMet derivative. The pET28a-*lspA* recombinant plasmid was purified from *E. coli* DH5α cells (NEB) using a Plasmid Midiprep kit (Qiagen) according to the manufacturer's instructions. Optimisation of magnesium concentration for protein production was carried out in a continuous exchange, precipitate-generating cell-free (P-CF) expression system on a 50-μL scale, as described (*16*). The optimal magnesium concentration was subsequently used for preparative-scale (3 mL reaction mix, 42 mL feeder mix) cell-free expression, as described (*16*). Expression was carried out in an incubator at 30 °C with continuous shaking at 150 rpm over a 16-h period. For SeMet labelling, the cell-free amino acid mix was supplemented with SeMet (Sigma) in place of Met. Purification of cell-free LspA was carried out as described above for *in vivo* expressed LspA starting with the solubilisation step.

Crystallisation

Aliquots of purified LspA at 12 mg/mL in Buffer B were thawed on ice and incubated for 5 min on ice with a 10-fold molar excess of the antibiotic globomycin from *Streptomyces hagronensis* (Sigma). This solution was then used to prepare the cubic phase with monoolein as host lipid following well-established procedures (*7, 17*).

In meso crystallisation trials were set up by transferring 50 nL of the protein-laden mesophase onto a silicanised 96-well glass crystallisation plate which was then covered with 800 nL of precipitant solution using *in meso* robots (*18*). Plates were sealed with a glass cover plate and incubated in and imaged at 20 ^{o}C with a Formulatrix Rock Imager. Best quality crystals were obtained using precipitant solutions containing 100 mM MES pH 5.6-6.0, 35-43 %(v/v) PEG400 and 60-100 mM ammonium phosphate monobasic. Plate-shaped crystals appeared after 1 - 3 days and continued to grow reaching dimensions of 3 x 30 x 80 μm^{3} in 2 weeks. Crystals were harvested by opening the wells with a tungsten carbide glass cutting tool, removing crystals with a minimum of adhering mesophase using MiTeGen cryoloops followed by snap-cooling in liquid nitrogen without added cryoprotectant (*17*).

Data collection and processing

X-ray diffraction data were collected on the 23-ID-B beamline of the General Medicine and Cancer Institute Collaborative Access Team (GM/CA-CAT) at the Advanced Photon Source (APS), Argonne, Illinois, USA, the PXII beamline at the Swiss Light Source (SLS), Villigen, Switzerland and the I24 beamline at the Diamond Light Source (DLS), Didcot, Oxford, UK. Data were collected at 100 K using a 10 x 10 μm^{2} collimated microbeam at GM/CA-CAT, while 10 x 18 μm^{2} and 10 x 10 μm^{2} microfocus beams were used at SLS and DLS, respectively. Software-implemented grid-scan protocols were used to locate and center crystals obscured by the opaque frozen lipid cubic phase.

Data were processed with xia2 using *XDS* (*19*) for indexing, integration and scaling, *XSCALE* (*19*) for relative scaling of multiple datasets and *AIMLESS* (*20*) for merging of reflections. An 8-fold redundant SeMet derivative data set to 3.3 Å was obtained by merging 8 data sets from 6 crystals collected at the K-edge of selenium, determined by means of a fluorescence scan to be 0.9793 Å. A 2.8 Å native data set was obtained by merging 16 data sets collected from 9 crystals of native protein collected at 0.9824 Å.

Structure solution and refinement

Due to a weak anomalous signal extending only to 6 Å solving the structure was not straightforward. The selenium substructure search was performed using *SHELXD* (*21*) via the *HKL2MAP* GUI (*22*) with a large variety of input parameters. Since no clear solution was evident and *SHELXE* (*21*) failed to deliver acceptable maps or models all *SHELXD* selenium substructure solutions were screened for solutions with slightly favourable statistics and a reasonable substructure occupancy distribution. Solutions selected in this way were submitted to *PHENIX* Autosol (*23*) for substructure refinement and phasing. A favourable solution emerged with 12 out of 13 selenium sites of the refined SHELXD output located at anomalous peaks while also being reasonably positioned in relation to the protein envelopes discernible after solvent flattening. This solution was submitted to the *PHENIX* AutoBuild routine which built 397 out of 676 residues representing four molecules in the asymmetric unit with R / R_{free} factors of 0.36 / 0.44. R / R_{free} were improved to 0.219 / 0.245 by using the 2.8 Å native data and rounds of iterative model building and refinement in Coot (*24*) and *PHENIX*. Refinement was performed with automatic optimization of X-ray / stereochemistry and X-ray / ADP weights and torsion-angle NCS restraints. In the MolProbity (*25*) Ramachandran plot 97.3 % of residues were in the favoured region, 2.5 % were in the allowed region. There was one Ramachandran outlier fit into poor density. The clash score was 6.69 and the MolProbity score was 1.50. There were no rotamer outliers and the average B-factor of 68.2 $Å^2$ (66.0 $Å^2$ for the peptide chain) corresponded well with the Wilson B-factor of 67.7 $Å^2$.

For Molecule A in the asymmetric unit (AU) all residues except the N-terminal methionine and 11 disordered residues at the C-terminus could be placed in electron density. The models for the other 3 molecules in the AU were less complete with 20 residues not modelled for each. Thus, 604 of 676 possible LspA wild-type residues in the 4 protein chains in the AU could be placed. Clear doughnut-shaped density for the globomycin ligand was observed after the *PHENIX* autobuild step, allowing the ligand to be placed unambiguously. Twenty four monoolein (9.9 MAG) molecules were added to the model based mainly on density of the glycerol headgroups and the first few carbon atoms in the acyl chain. The monoolein molecule with the strongest density in all four of the asymmetric units was bound in a groove adjacent to the globomycin binding site (**fig. S5**).

Proteins for LspA assay

The LspA substrate used in this work is the dagylated form of the PAO1 pre-prolipoprotein, ICP (pre-proICP). In order to generate the substrate (proICP) for use in *in vitro* assays, it was necessary to produce purified Lgt, the dagylating enzyme, and ICP.

The *lgt* wild-type gene from *P. aeurginosa* PAO1 was synthesized (GenScript) and cloned into pET28a using restriction sites *NdeI* and *HindIII*. The recombinant Lgt construct contained an N-terminal 6 x His-tag and the TEV cleavage site sequence ENLYFQG. The pET28a-*lgt* recombinant plasmid was transformed into *E. coli* CD43 (NEB) cells for expression. Cells were grown in 1 L TB broth (Sigma) in 6 x 3 L flasks at 37 ^{o}C to an OD_{600}

of 0.5-0.6. Protein expression was induced with 0.25 mM IPTG at 20 ^{o}C with shaking at 180 rpm for 18 h. Cells was harvested by centrifugation at 5,000g for 10 min at 4 ^{o}C. The cell mass was either used immediately or stored at -80 ^{o}C for a maximum of 2 months.

The protocol for Lgt purification was similar to that used with LspA with the following modifications. The buffer used throughout the purification was 50 mM Tris pH 7.5, 500 mM NaCl, and 1 mM *tris*(2-carboxyethyl)phosphine (TCEP). Lgt was solubilised using 1 %(w/v) lauryl maltose neopentyl glycol (LMNG) (Anatrace). The buffer used post-solubilisation was supplemented with 0.02 %(w/v) LMNG. The gel filtration buffer included 50 mM Tris/HCl pH 7.5, 250 mM NaCl, 1 mM TCEP, and 0.02 %(w/v) LMNG. The purified protein was identified as Lgt by mass spectrometry.

Potential pre-prolipoprotein substrates were identified using the DOLOP database (*26*) and a PAO1 lipoprotein database of our own construction. ICP was chosen based on its overall size, signal sequence length and amino acid composition. The wild-type *icp* gene from *P. aeruginosa* PAO1 was synthesized (GeneScript) and sub-cloned into the pET28a plasmid using the restriction sites *NdeI* and *HindIII*. The resulting recombinant expression construct was the wild-type ICP gene with the same N-terminal modifications as described for LspA.

Pre-proICP was produced in *E. coli* C41 (DE3) (NEB) cells at 37 ^{o}C in TB broth (Sigma) following induction with 1 mM IPTG at an OD_{600} of 0.6. Expression was carried out for 2 h. Longer expression resulted in a lower yield presumably because the pre-proICP was processed by *E. coli* Lgt and LspA to a form no longer detectable by Western blotting with anti-His tag antibodies. The protocol for pre-proICP purification was similar to that used with LspA with the following modifications. The buffer was Hepes pH 7.1, 150 mM NaCl, 1 mM TCEP, and 10 %(v/v) glycerol. Pre-proICP was solubilised in 1 %(w/v) LMNG. The purified protein was identified as pre-proICP by mass spectrometry.

Enzyme assays

The proteolytic activity of LspA was assayed in what amounts to a coupled system. Thus, the proICP substrate was generated *in situ* in the reaction mix by the action of Lgt on pre-proICP using DOPG (sodium salt, Sigma) for the dagylation reaction. All assays were carried out in Buffer C (50 mM Tris/HCl pH 7.5, 150 mM NaCl, 1 mM TCEP, 0.02 %(w/v) LMNG) at 37 ^{o}C. DOPG was solubilised in Buffer C at a concentration of 1.5 mM. For time course assays, a 200 µL reaction was set up containing 12 µM pre-proICP, 250 µM DOPG, 1.5 µM Lgt and 0.5 µM LspA in Buffer C. The reaction was initiated by adding LspA. At timed intervals, 20 µL samples were removed and the reaction was stopped by adding SDS loading buffer (62.5 mM Tris/HCl pH 6.8, 2.5 %(w/v) SDS, 0.002 %(w/v) bromophenol blue, 0.5 M ß-mercaptoethanol, 10 %(v/v) glycerol). For LspA concentration-dependent activity assays, a 200 µL reaction mix was prepared containing 9 µM pre-proICP, 1.5 µM Lgt, and 250 µM DOPG in Buffer C. The reaction mix immediately divided into 20 µL samples to which LspA at a final concentration of 0 to 500 nM was added to start the reaction. The reaction was carried out for 20 min at 37 ^{o}C. For the globomycin inhibition assay, a 220 µL reaction mix was prepared containing 5 µM pre-proICP, 2.5 µM Lgt and 250 µM DOPG. After 30 min, the reaction mix, which now contains proICP following Lgt action, was divided into 20 µL samples. Globomycin was added to a final concentration of 0 to 20 µM followed immediately by 2 µM LspA. The reaction was allowed to proceed for 1 h at 37 ^{o}C. For the mutant assays a 280 µL reaction mix was prepared containing 10 µM pre-proICP, 1.2 µM Lgt, and 250 µM DOPG in Buffer C. 20 µL samples were taken from the reaction mix to which 2.5 µM of LspA (wild type, D115N, D115A, R116A, D124N, D143N) was added to start the reaction.

The reactions were allowed to proceed for 1 h at 37 °C and were stopped with SDS loading buffer. Samples were electrophoresed in Criterion™ Precast 10-20 % Tris-Tricine gels in Tris-Tricine buffer (Sigma) and stained using InstantBlue™ (Expedeon). Quantitation of protein on gels was by image analysis using ImageJ (*27*).

Molecular dynamics simulations and docking

All MDS were performed using GROMACS v5.0.5 (*28*). The Martini 2.2 force field (*29*) was used to run an initial 1 µs Coarse Grained (CG) MD simulation to permit the assembly and equilibration of a palmitoyloleolylphosphatidylglycerol (POPG) palmitoyloleoylphosphatidylethanolamine (POPE) (1:3 mole ratio) bilayer around LspA (*30*). The end-snapshot of the CG simulations was then converted to atomic detail with the crystal structure aligned with the CG protein within the assembled lipid bilayer (*31*). The systems were equilibrated for 1 ns with the protein restrained before 500 ns of unrestrained atomistic MD using the Gromos53a6 force field (*32*). The parameters for globomycin were manually created, modifying the standard amino acid force field parameters to generate the correct stereochemistry and to complete the cyclic structure of the antibiotic. The dagylated cysteine parameters were created from lipid parameters for diacylgylcerol and appended to the thiol side chain of cysteine. Systems were neutralised with 150 mM NaCl. *In silico* mutagenesis was performed and figures were prepared using PyMol (Schrödinger).

To model substrate binding into the apo-form of LspA (**Fig. 3B**), the helix of the signal peptide of pre-proICP was docked to the transmembrane groove in LspA using ZDOCK (*33*). The architecture of the signal peptide binding site is similar to the glove-like structure of the 6TM helix TatC protein, which forms an earlier step in the post-translational processing pathway for some substrates of LspA (*34*). The C-terminal end of the docked TM helix was aligned with the structure to permit the lipobox sequence to follow the coordinates of g·Leu, g·Ile and g·Ser of globomycin. The location of the neighboring MAG guided the position and orientation of the next residue, Cys*. This permits the flexible linker of the substrate to dock into the β-cradle, with the folded lipoprotein extending in the periplasm. The docking and simulations was repeated with LppL (PA5276), the shortest pre-prolipoprotein (46 residues total) from *P. aeruginosa* with similar outcomes as for ICP. The overall configuration for both was stable in the simulations, especially in the region of the TM helix and the lipobox. To study the membrane partitioning of globomycin by MDS, the antibiotic was placed in the solvent 50 Å from the membrane and allowed to freely sample the solvent and membrane.

Supplementary figures
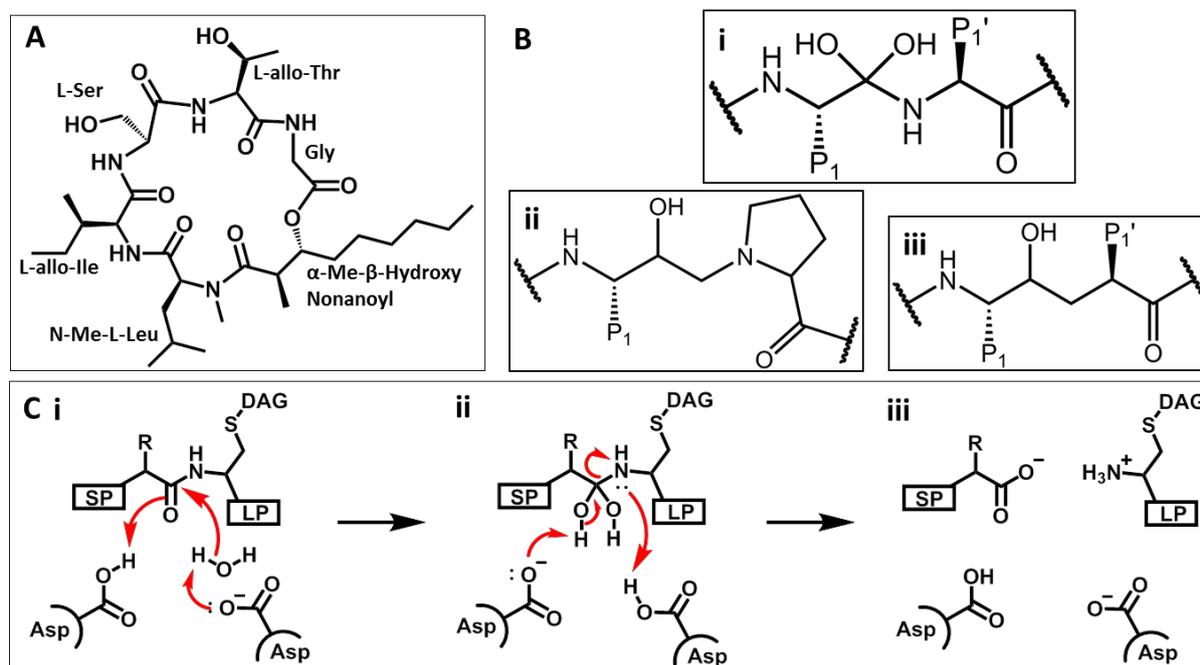
Fig. S1. Proposed mechanism for aspartyl protease catalysis and comparison of the tetrahedral intermediate to globomycin and non-cleavable transition state isosteres developed for the synthesis of protease inhibitors. (**A**) Chemical structure of globomycin (*5*). Component parts of this cyclic depsipeptide are labelled. Drawn in roughly the same orientation as globomycin in **Fig. 2**. Natural congeners come with L-Val in place of L-allo-Ile and with fatty acids that range in total number of carbons from 8 to12 (*5*). In the complex, globomycin was modelled with a 10 carbon acid. (**B.i**) Tetrahedral transition state proposed to form during the aspartyl protease reaction. (**B.ii, B.iii**). Non-cleavable transition state isosteres of the hydroxyethylamine and hydroxyethylene types, respectively (*9*). (**C**). Proposed catalytic mechanism for aspartyl proteases (*9, 12*) with parts of the substrate around the scissile bond labelled as found in a prolipoprotein: SP, signal peptide; LP, lipoprotein; DAG, diacylglycerol. (**C.i**) Michaelis complex. (**C.ii**) Tetrahedral intermediate. (**C.iii**) Product complex.
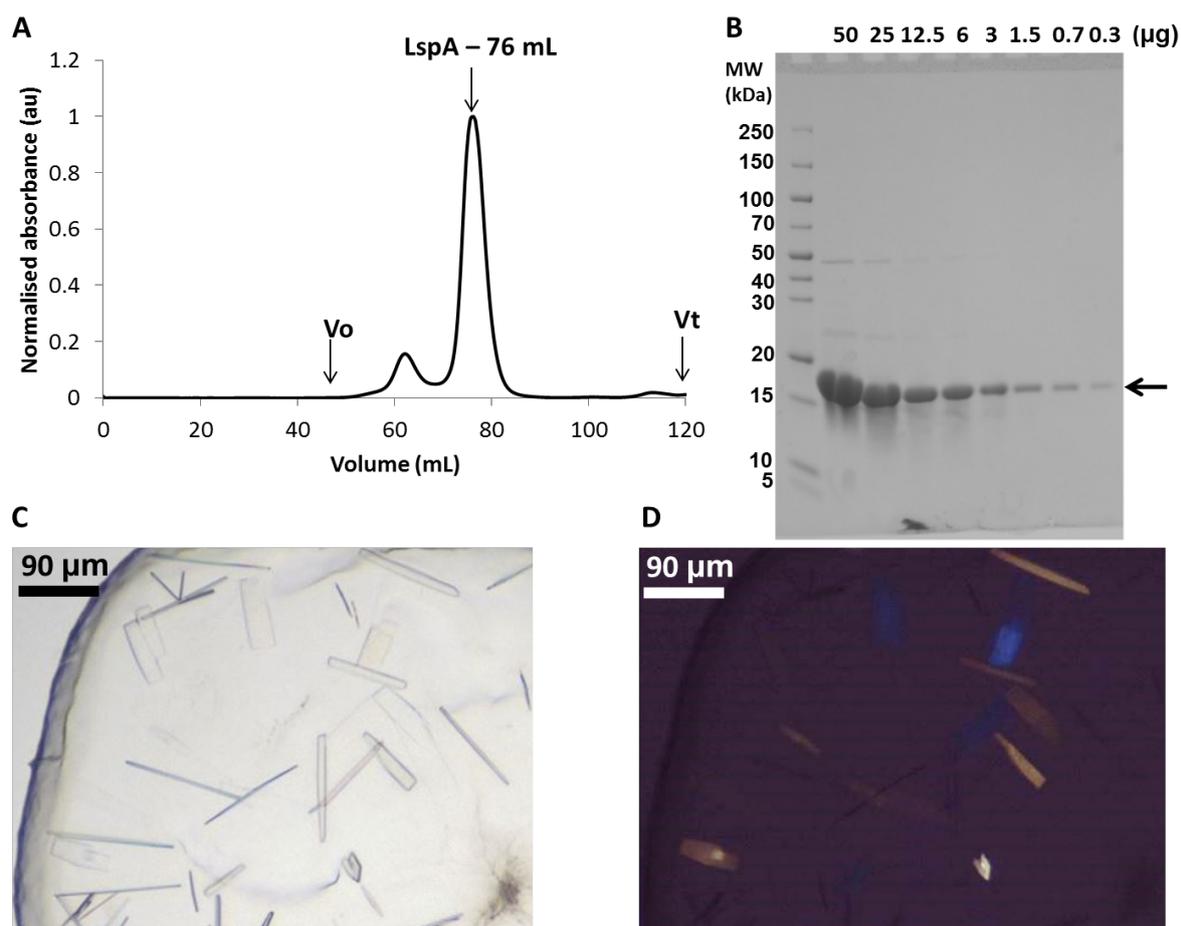
Fig. S2. Purification and crystallization of LspA. (A) LspA eluted from a Superdex200 size-exclusion column as a symmetric Gaussian peak at an elution volume of 76 mL. **(B)** An SDS PAGE loading series to estimate LspA purity after size-exclusion chromatography. The protein, which runs with an apparent molecular weight of ~17 kDa (arrow), is ~98 % pure as determined by ImageJ densitometry analysis. **(C)** Thin plate-shaped crystals of LspA growing in the 9.9 MAG mesophase imaged with a bright field microscope. **(D)** As in **(C)** viewed between crossed polarizers.
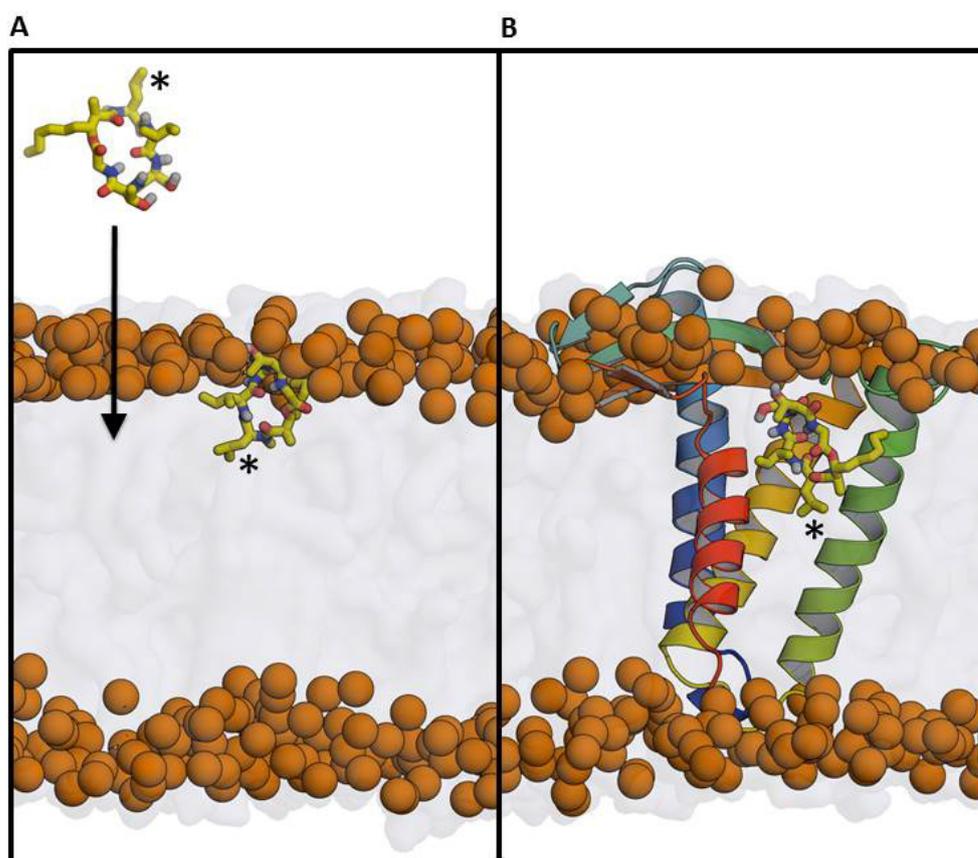
Fig. S3. Partitioning of globomycin into the membrane observed by MDS. (**A**) Globomycin partitions from the aqueous solution into the bulk membrane. In the membrane, the antibiotic is oriented at the interface as expected for an amphiphilic molecule. (**B**) A similar orientation is observed for the antibiotic in the LspA-globomycin complex structure. The g·Leu of globomycin is marked with an asterisk. Phosphorus atoms of the lipid headgroups are shown as orange spheres and globomycin as sticks. The lipid acyl chain compartment of the membrane along with headgroups in the background are shown as a light gray transparent surface.
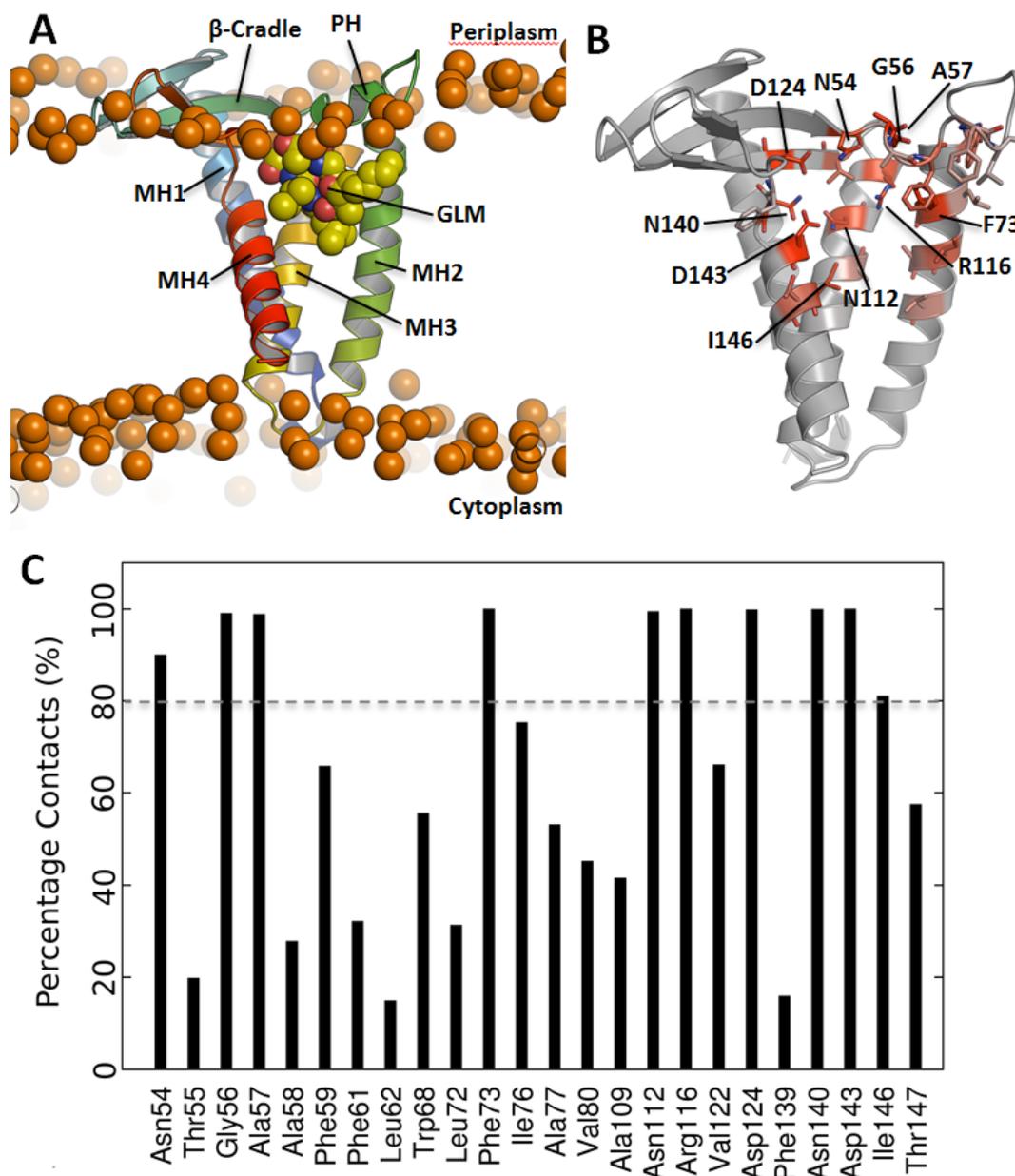
Fig. S4. Interactions between LspA and globomycin in a membrane as revealed by MDS. (A) The crystal structure of LspA (cartoon representation) embedded in a POPE:PG lipid bilayer, with globomycin (yellow spheres) bound. Phosphorus atoms of the lipid headgroups are shown as orange spheres. **(B)** Globomycin contacts with LspA over 500 ns MDS. Residues making contact with the globomycin are shown on a scale of grey to red, where grey designates no contacts and red corresponds to contacts throughout the 500 ns MDS. Residues that make contact with globomycin are shown as sticks, while residues that interact for greater than 80% of the simulation are labeled. **(C)** All contacts made by globomycin during the MDS. The 80% cut-off is marked by a dashed line.
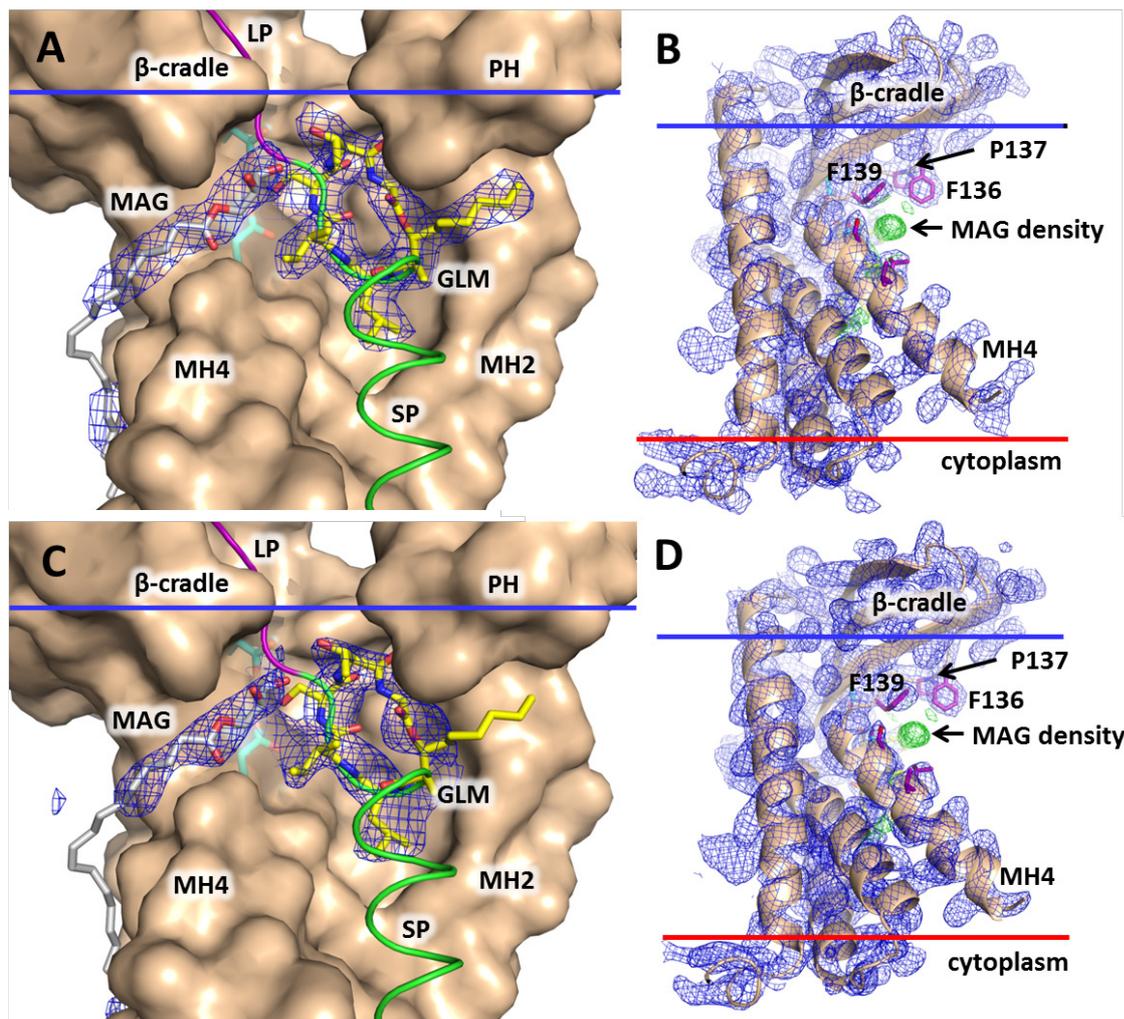
Fig. S5. Electron density for lipid extends into the active site of LspA. (**A**) The view is from the membrane into the active site where the catalytic dyad residues Asp124 and Asp143 are shown in stick representation with cyan carbons. Density for the lipid (9.9 MAG, the hosting lipid used for crystallization) and globomycin contoured at 1 σ is shown as a blue mesh. For clarity the map used in this figure is a *PHENIX* (*23*) feature enhanced map. LspA is shown in surface representation while lipid and globomycin are shown as sticks with white and yellow carbons, respectively. For reference, the positioning of a prolipoprotein chain according to our model (**Fig. 3B**) is shown in tube representation (green from N-terminus to cleavage site, magenta starting with Cys*). (**B**) View from within the membrane along the long axis of density near the active site into which the glycerol headgroup and the first seven methylenes of the acyl chain of a 9.9 MAG lipid were modelled. The view reveals the shape complementarity between the proposed prolipoprotein DAG binding site on LspA created where the β-cradle and MH4 come together and part of the DAG moiety. LspA density (blue) is shown as a *PHENIX* feature-enhanced map contoured at 1 σ. The lipid density (green) is a simulated annealing composite omit map contoured at 1 σ and carved at 2 Å around the modelled MAG. Residues forming the binding site are shown as sticks (purple carbons). Phe139 at the back of the binding site is strictly conserved. Residue 136 is strictly conserved as an aromatic residue and Pro137, which generates a kink in the fourth strand of the β-cradle that allows residue 136 to clamp down on the lipid, is 96% conserved. (**C**) and (**D**) As in (**A**) and (**B**), respectively, using bias-free, iterative-build composite OMIT maps contoured at 1 σ for all shown density. Despite the low-quality phases due to zero manual model building and the use of a ligand-free model, density for the two ligands (globomycin, GLM, and lipid, MAG) is clearly present and identifiable.
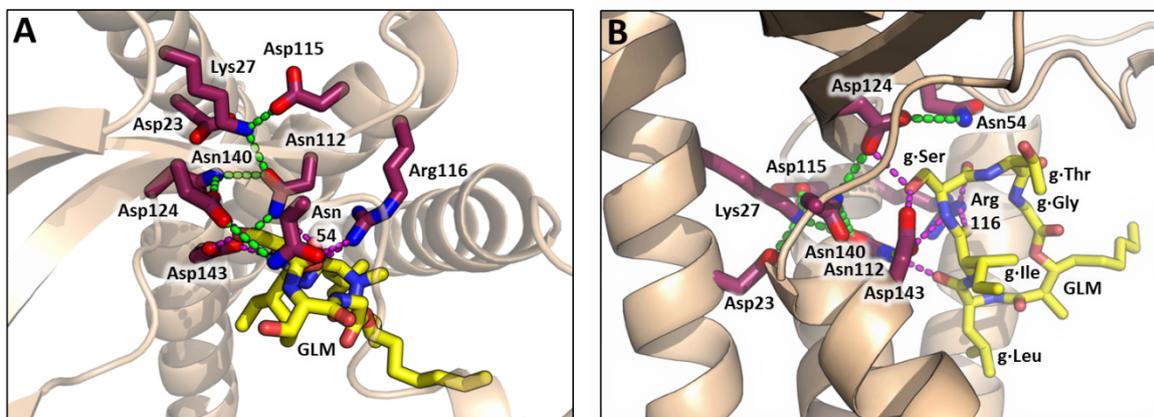
Fig. S6. Hydrogen bonded network involving all strictly conserved polar residues and Asp115 in the LspA-globomycin complex. (A) View from the periplasm between the β-cradle and the PH loop into the active and globomycin binding sites. Inter-side chain hydrogen bonds are shown as thick green dashed lines. Hydrogen bonds between side chains and globomycin are shown as thick magenta dashed lines. Side chains and globomycin have purple and yellow carbon atoms, respectively. Parts of the β-cradle that sit over the binding site are rendered transparent for clarity. **(B)** View of the hydrogen bonded network roughly along the proposed DAG binding groove.
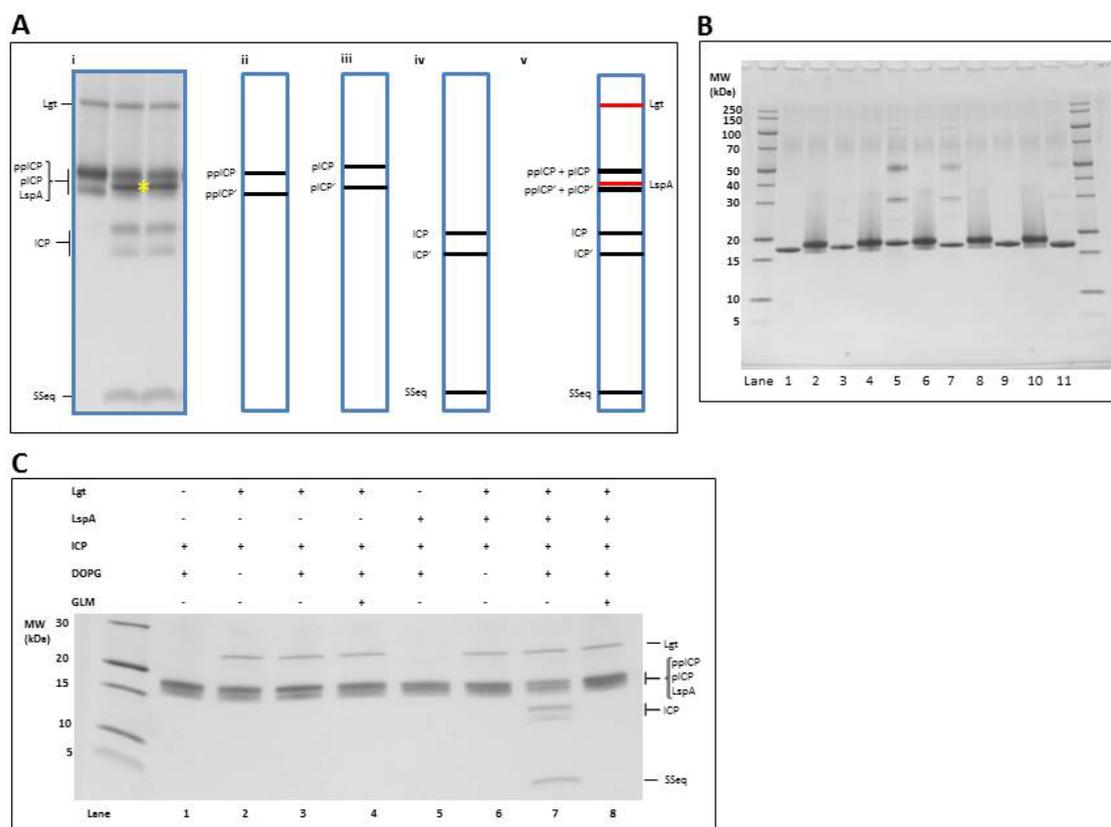
Fig. S7. SDS-PAGE gels used to assay LspA endopeptidase activity and to characterize LspA mutants. (**A**) Schematic representation of how proteins run on SDS-PAGE gels in assays of LspA endopeptidase activity with *in situ* produced proICP. **i**. Part of an actual gel from **Fig. 4G**. The yellow asterisk marks the position of wild-type LspA. **ii – iv**. Position on the gel of the pre-proICP (ppICP) substrate for Lgt, the proICP (pICP) substrate for LspA, and the products of the LspA reaction, ICP and signal sequence (SSeq), respectively. All forms of ICP (ppICP, pICP and ICP) run as two bands. The faster running of the two is less intense and is indicated by a prime. **v**. Position on the gel of proteins in a typical reaction mix containing the enzymes Lgt and LspA (red bands) and the substrates and products, ppICP, pICP, ICP and SSeq (black bands). Interpretation of the gels is complicated by the fact that all forms of ICP run as pairs of bands, the LspA, ppICP' and pICP' bands are closely spaced on the gel, and ppICP and pICP have similar mobilities. Thus, ppICP and pICP together appear as a single broad band (**v**). The same is true for ppICP' and pICP'. (**B**) Gel mobility of wild-type and mutant forms of LspA. Even numbered lanes contain pre-proICP. Lanes 1, 3, 5, 7, 9 and 11 contain the wild-type, D115A, D115N, R116A, D124N and D143N forms of LspA, respectively. (**C**) Activity assay measurement controls demonstrate that Lgt, LspA, DOPG and ICP must all be present in the reaction mix for the endopeptidase reaction to occur (Lane 7), that globomycin inhibits the reaction (Lane 8), and that LspA requires the dagylated form of ICP for activity (Lanes 5 and 6).
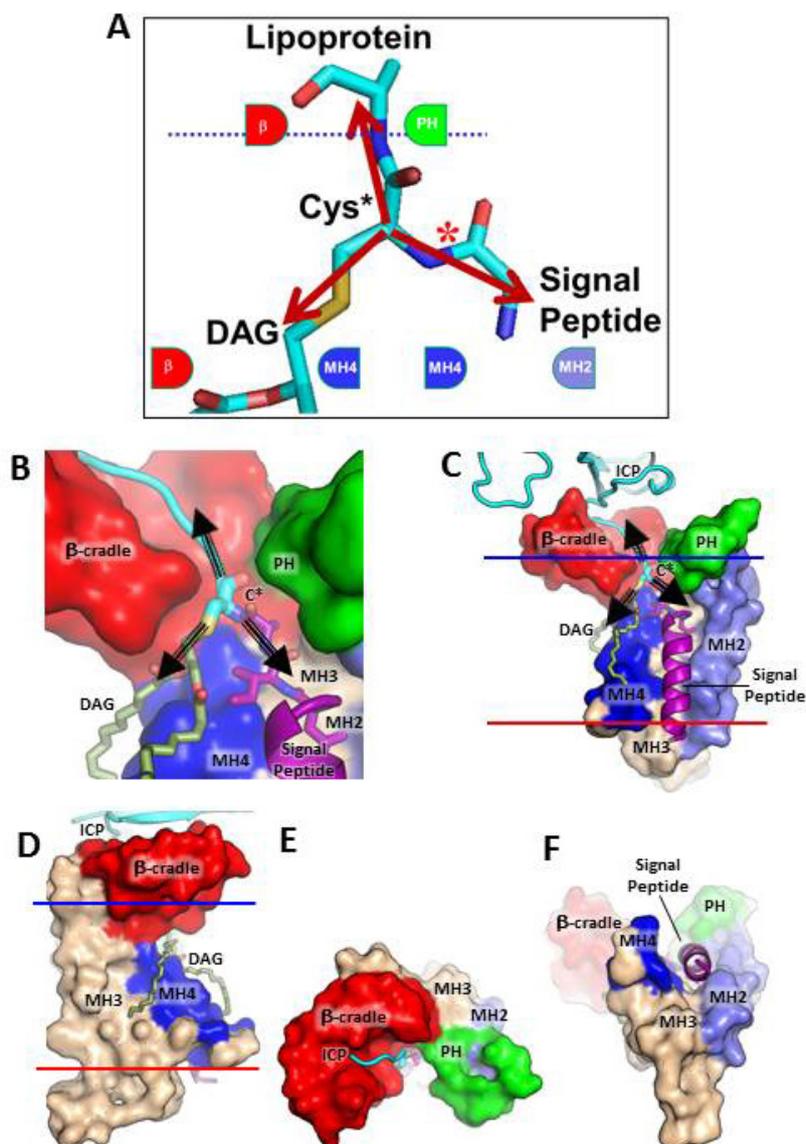
Fig. S8. Shape complementarity explains LspA's ability to function with a broad range of lipoprotein substrates. (**A**) The dagylated cysteine, Cys*, is common to all prolipoprotein substrates of LspA enabling the enzyme to proteolytically cleave in a highly specific way almost 200 different proteins. Its uniqueness derives from an extramembrane lipoprotein domain, a DAG and a signal peptide linked, respectively, to its carboxyl end, its side chain and its amino terminus. The Cys* is held in position in the active site by shape complementarity features (colored shapes) on LspA that include i) a cleft between the periplasmic β-cradle and PH loop sub-domains which secures the extended peptide leading into the lipoprotein domain, ii) a cleft where the β-cradle and MH4 meet at the membrane interface (dotted line) that holds the lipid in place, and iii) a long groove created by MH2, MH3 and MH4 which anchors the signal peptide helix. The net effect is to position the scissile bond (asterisk) of the prolipoprotein in the active site for cleavage. The bold red arrows are shown originating from the tetrahedral α-carbon of Cys* to convey the sense of how the components of this special residue are splayed apart by being anchored to disparate parts of the enzyme. The stereochemistry of the α-carbon dictates how the substrate is oriented for docking on the enzyme. The α-hydrogen of Cys* is out of view behind the α-carbon. Both α-hydrogen and α-carbon extend away from the viewer into a pocket on the protein surface where MH4 and the β-cradle meet. (**B**) Expanded view of the prolipoprotein substrate in the LspA active and binding sites. LspA is shown in surface representation, prolipoprotein in cartoon and DAG and Cys* (C*) as sticks. The example shown is the prolipoprotein ICP modelled in **Fig. 3B**. (**C**) Zoomed out view from

the membrane of the prolipoprotein bound to LspA as in (**B**). (**D**) As in (**C**) rotated ~90° to reveal the groove between the β-cradle and MH4 in which one of the DAG acyl chains sits. The hydrophobic cleft where the PH loop and MH2 meet accommodates the acyl chain of globomycin. Likewise, it may bind one of the prolipoprotein DAG acyl chains to assist in substrate docking for catalysis. (**E**) As in (**C**) viewed from the periplasm to reveal how the N-terminal end of the lipoprotein sits in the groove between the β-cradle and the PH loop. The lipoprotein domain has been removed for clarity. (**F**) As in (**C**) viewed from the cytoplasm to show how the signal peptide sits in the long groove created by MH2, MH3 and MH4.
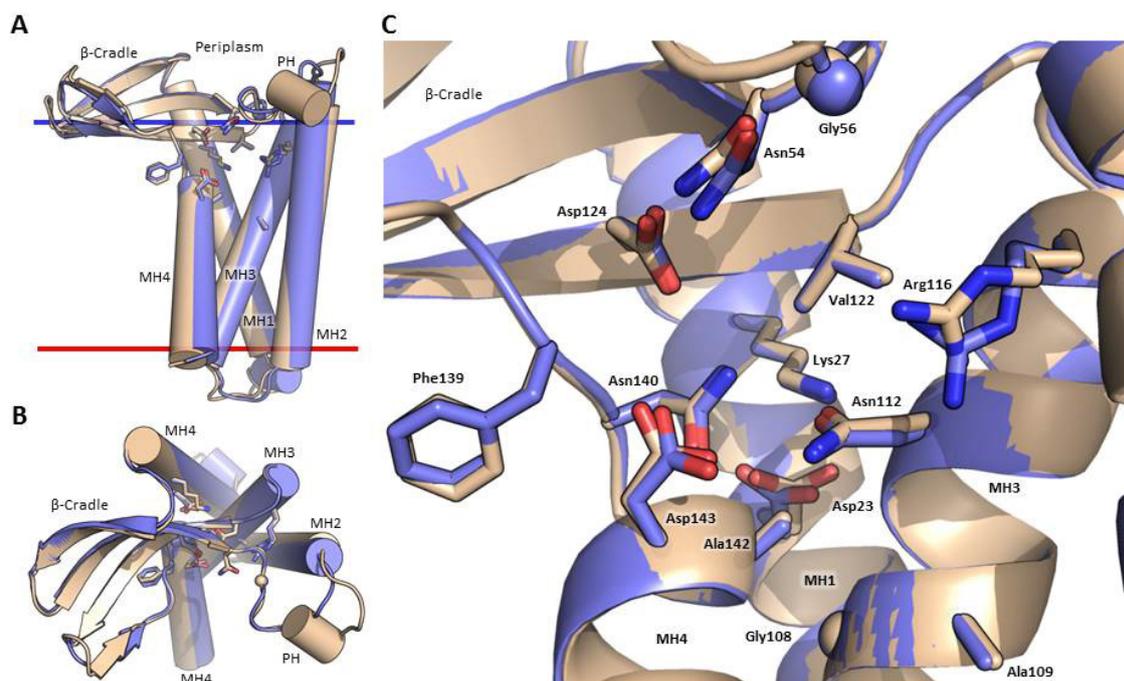
Fig. S9. A structure homology model of LspA from MRSA has a similar fold and arrangement of strictly conserved residues to those observed in LspA from *P. aeruginosa*. (**A, B**) Orthogonal views of LspA from within the membrane and from the periplasm, respectively. The protein is shown in the apo form in cartoon representation with cylindrical helices. With the exception of glycines, which are represented as spheres, strictly conserved residues are shown as sticks. Amino acids are numbered as in the *P. aeruginosa* sequence. *P. aeruginosa* and MRSA orthologs are colored cream and blue, respectively. The most notable differences between the two structures are at the PH loop and the N-terminus. (**C**) An expanded view from within the membrane of all fourteen strictly conserved residues in and around the active and binding sites of LspA.

Table S1. Data collection and refinement statistics

	Wild type LspA*	SeMet LspA**
Data collection		
Space group	*C*2	*C*2
Cell dimensions		
a, *b*, *c* (Å)	112.5, 105.9, 85.4	112.0, 106.1, 85.5
β (°)	96.9	96.4
Resolution (Å)	44.9-2.8 (2.9-2.8) ***	55.6-3.3 (3.39-3.30)
R_{merge}	0.151 (0.764)	0.318 (1.36)
$I/\sigma I$	7.4 (1.7)	11.5 (2.8)
Completeness (%)	98.1 (95.6)	99.9 (99.9)
Redundancy	3.6 (2.8)	8.1 (7.3)
CC_{anom}****		0.45
Refinement		
Resolution (Å)	2.8	
No. of reflections	23,968	
R_{work}/R_{free}	0.219/0.245	
No. of atoms	5,653	
Protein	4,853	
Ligand/ion	784	
Water	16	
Average B-factors (Å^2)		
Protein	66.0	
Ligand/ion	82.2	
Water	53.1	
R.m.s. deviations		
Bond lengths (Å)	0.004	
Bond angles (º)	0.92	
Validation		
Ramachandran favored (%)	97.3	
Ramachandran outliers (%)	0.2	
Clash Score	6.69	
MolProbity Score	1.50	

*   After merging 16 data sets from 9 crystals
**   After merging 8 data sets from 6 crystals
***   Highest resolution shell is shown in parenthesis
**** Anomalous correlation coefficient to a resolution of 5.5 Å (used for substructure determination)

Table S2. Hydrogen bonds in the active site hydrogen-bonded network of the LspA-globomycin complex*

Partner 1	Partner 2	Distance (Å)**
g.Leu:O	Asn112:ND2	3.0
g.Ile:O	Arg116:NH2	2.6
g.Ser:O	Arg116:NH2	3.0
g.Ser:OG	Asp143:OD2	2.4
g.Ser:OG	Asp124:OD1	2.9
g.Ser:N	Asp143:OD1	3.0
Asp23:OD2	Lys27:NZ	2.8
Asp23:OD1	Ala142:N	2.9
Lys27:NZ	Asp115:OD2	2.5
Lys27:NZ	Asn112:OD1	2.9
Asn54:ND2	Asp124:OD2	3.4
Asn54:OD1	Gly56:N	2.7
Asn112:OD1	Asn140:ND2	3.3
Asn112:ND2	Asp143:OD1	3.0
Asn140:ND2	Asp124:OD1	3.2
Asn140:OD1	Asp143:N	2.8
Asp143:OD2	Asn140:N	3.5

*With the exception of Asp115, all LspA residues identified in this table are strictly conserved.
**Distances were measured in molecule A, which has the best density of the four molecules in the asymmetric unit.

Movies:
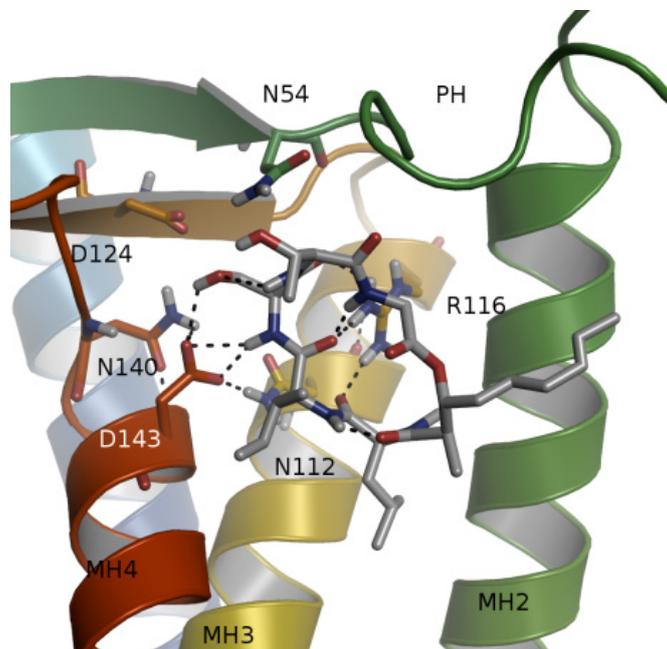
Movie S1. LspA and its globomycin (grey sticks) binding site are stable in a membrane bilayer through 500 ns of MDS. LspA is shown in cartoon representation, with residues involved in hydrogen bonds (dashed lines) shown as sticks.
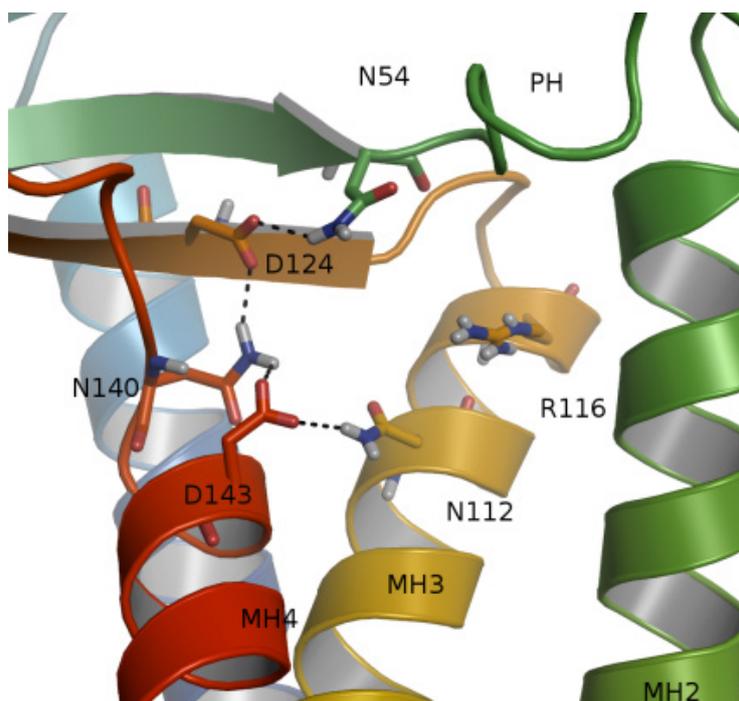
Movie S2. Apo-LspA is stable in a membrane bilayer through 500 ns of MDS. Removal of globomycin from the LspA-globomycin complex does not induce gross structural changes in 500 ns of MDS due, in part, to a tight network of hydrogen bonds between the conserved polar residues in and around the active site. LspA is shown in cartoon representation, with residues involved in hydrogen bonds (dashed lines) shown as sticks.
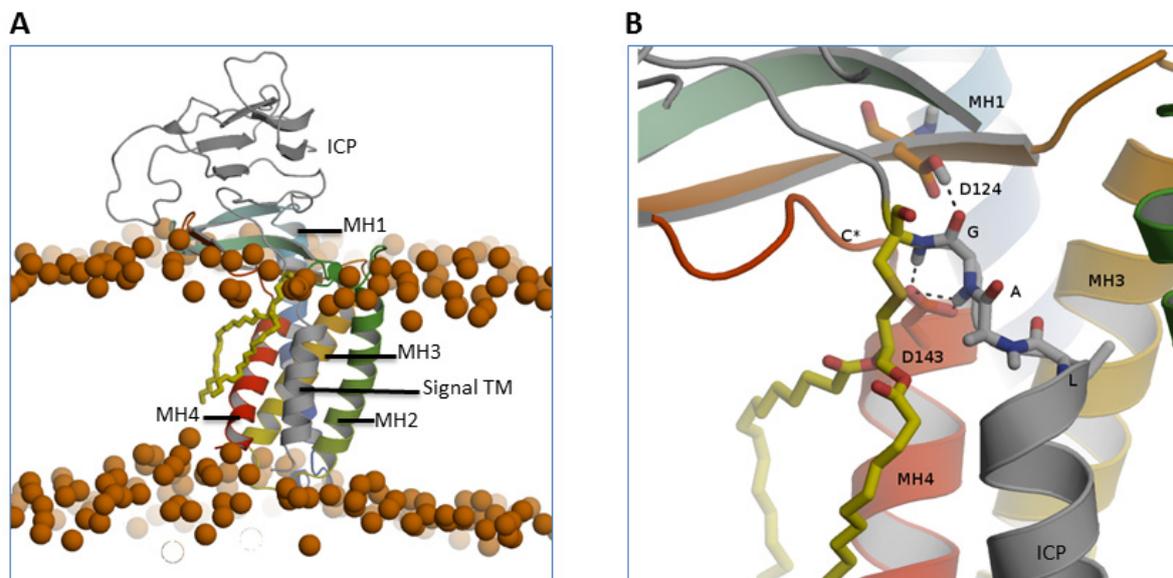
Movie S3. The prolipoprotein substrate, ProICP (grey), docked into apo-LspA is stable in a membrane bilayer through 100 ns of MDS. Full details of how proICP was modelled and docked into the apo-form of LspA are provided under Methods. **Movie S3A** is a view of the overall complex from the membrane plane. **Movie S3B** is a zoomed in view of the lipobox (shown as sticks) in the active site.
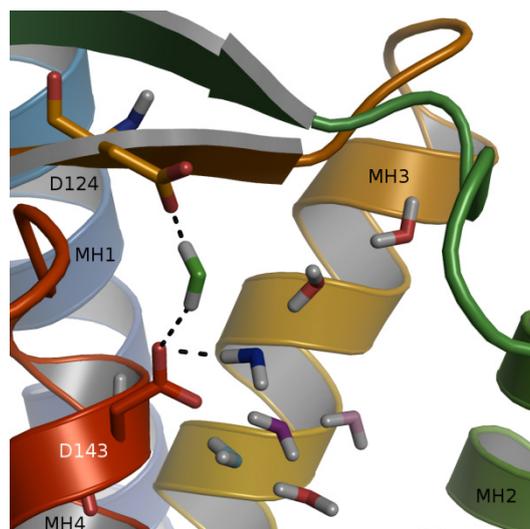
Movie S4. A water molecule coordinates between the catalytic dyad aspartates Asp124 and Asp143 of apo-LspA in MDS. With restraints imposed upon the protein to maintain the precise coordinates of the X-ray structure, the two catalytic aspartate residues dynamically coordinate waters individually that infiltrate the active site of the enzyme. Rather than coordinating a single water throughout the MDS, multiple waters (marked by different colors) engage with the catalytic aspartates. Waters also freely exchange between the active site and the periplasm, navigating through a portal formed by the PH, MH2 and MH3; observed at the top right of the screen.
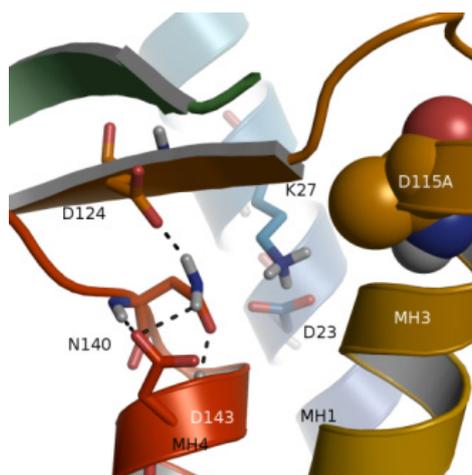
Movie S5. MDS of the mutant Asp115Ala indicates that Asp115 affects the catalytic dyad by way of an extensive hydrogen bonded network. Asp115 was expected to be a catalytic residue based on previous biochemical data (*8*). MDS of the mutant Asp115Ala (van der Waals spheres) indicate that it has a downstream effect on the catalytic site. The absence of the acidic side chain disrupts the side chain position of Lys27 enabling it to change conformation. This facilitates a flip in the side chain of Asn140, which is seen by MDS to initially coordinate Asp124 and Asp143 in the wild-type enzyme. The change in side chain rotamer of Asn140 destabilises the catalytic aspartates, enabling Asp124 to drift away from the active site thus lowering endopeptidase activity. By contrast, the Asp115Asn mutation (not shown) had a less adverse effect on LspA activity (Fig. 4). Its side chain carboxamide oxygen can stabilize Lys27 in position which in turn fixes Asn140 with less of an impact on the catalytic dyad.

Movies in dropbox

Movie 1: https://www.dropbox.com/s/u44o9m9vvtdqv78/movie%201%20LspA-globomycin.mov?dl=0

Movie 2: https://www.dropbox.com/s/i0cmdk03imxdqwk/movie%202%20LspA-apo.mov?dl=0

Movie 3: https://www.dropbox.com/s/tronnzo3s7jckni/movie%203%20ICP-combine.mov?dl=0

Movie 4: https://www.dropbox.com/s/frr6yj2kk012h0m/movie%204%20LspA-waters.mov?dl=0

Movie 5: https://www.dropbox.com/s/szrfuskcvhjb9se/movie%205%20LspA-D115A.mov?dl=0

References and Notes

1. H. Remaut, R. Fronzes, in *Bacterial Membranes: Structural and Molecular Biology* (Caister Academic Press, 2014), pp. 55-179.

2. W. R. Zückert, Secretion of bacterial lipoproteins: Through the cytoplasmic membrane, the periplasm and beyond. *Biochim. Biophys. Acta* **1843**, 1509–1516 (2014). [Medline](#) [doi:10.1016/j.bbamcr.2014.04.022](#)

3. N. Buddelmeijer, The molecular mechanism of bacterial lipoprotein modification—How, when and why? *FEMS Microbiol. Rev.* **39**, 246–261 (2015). [Medline](#) [doi:10.1093/femsre/fuu006](#)

4. K. Remans, K. Vercammen, J. Bodilis, P. Cornelis, Genome-wide analysis and literature-based survey of lipoproteins in *Pseudomonas aeruginosa*. *Microbiology* **156**, 2597–2607 (2010). [Medline](#) [doi:10.1099/mic.0.040659-0](#)

5. T. Kiho, M. Nakayama, K. Yasuda, S. Miyakoshi, M. Inukai, H. Kogen, Structure-activity relationships of globomycin analogues as antibiotics. *Bioorg. Med. Chem.* **12**, 337–361 (2004). [Medline](#) [doi:10.1016/j.bmc.2003.10.055](#)

6. T. El Arnaout, Structure-function studies of the Prolipoprotein Signal Peptidase, LspA, thesis, Trinity College Dublin (2013).

7. M. Caffrey, V. Cherezov, Crystallizing membrane proteins using lipidic mesophases. *Nat. Protoc.* **4**, 706–731 (2009). [Medline](#) [doi:10.1038/nprot.2009.31](#)

8. H. Tjalsma, G. Zanen, G. Venema, S. Bron, J. M. van Dijl, The potential active site of the lipoprotein-specific (type II) signal peptidase of *Bacillus subtilis*. *J. Biol. Chem.* **274**, 28191–28197 (1999). [Medline](#) [doi:10.1074/jbc.274.40.28191](#)

9. A. Brik, C. H. Wong, HIV-1 protease: Mechanism and drug discovery. *Org. Biomol. Chem.* **1**, 5–14 (2003). [Medline](#) [doi:10.1039/b208248a](#)

10. W. J. Greenlee, P. K. S. Siegl, Anglotensln/Renln modulators. *Annu. Rep. Med. Chem.* **27**, 59–68 (1992). [doi:10.1016/S0065-7743(08)60405-4](#)

11. S. S. Abdel-Meguid, Inhibitors of aspartyl proteinases. *Med. Res. Rev.* **13**, 731–778 (1993). [Medline](#) [doi:10.1002/med.2610130605](#)

12. A. Das, S. Mahale, V. Prashar, S. Bihani, J. L. Ferrer, M. V. Hosur, X-ray snapshot of HIV-1 protease in action: Observation of tetrahedral intermediate and short ionic hydrogen bond SIHB with catalytic aspartate. *J. Am. Chem. Soc.* **132**, 6366–6373 (2010). [Medline](#) [doi:10.1021/ja100002b](#)

13. Y. Fu, Structure and dynamics of *Pseudomonas aeruginosa* ICP, thesis, University of Glasgow (2009).

14. K. Lewis, Platforms for antibiotic discovery. *Nat. Rev. Drug Discov.* **12**, 371–387 (2013). [Medline](#) [doi:10.1038/nrd3975](#)

15. M. Landau, I. Mayrose, Y. Rosenberg, F. Glaser, E. Martz, T. Pupko, N. Ben-Tal, ConSurf 2005: The projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **33**, W299–W302 (2005). [doi:10.1093/nar/gki370](#)

16. C. Boland, D. Li, S. T. Shah, S. Haberstock, V. Dötsch, F. Bernhard, M. Caffrey, Cell-free expression and in meso crystallisation of an integral membrane kinase for structure determination. *Cell. Mol. Life Sci.* **71**, 4895–4910 (2014). Medline doi:10.1007/s00018-014-1655-7

17. M. Caffrey, C. Porter, Crystallizing membrane proteins for structure determination using lipidic mesophases. *J. Vis. Exp.* **45**, 1712 (2010). Medline

18. D. Li, C. Boland, K. Walsh, M. Caffrey, Use of a robot for high-throughput crystallization of membrane proteins in lipidic mesophases. *J. Vis. Exp.* **45**, e4000 (2012). Medline

19. W. Kabsch, XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010). Medline doi:10.1107/S0907444909047337

20. P. R. Evans, G. N. Murshudov, How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214 (2013). Medline doi:10.1107/S0907444913000061

21. G. M. Sheldrick, Experimental phasing with SHELXC/D/E: Combining chain tracing with density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 479–485 (2010). Medline doi:10.1107/S0907444909038360

22. T. Pape, T. R. Schneider, *HKL2MAP*: A graphical user interface for macromolecular phasing with *SHELX* programs. *J. Appl. Cryst.* **37**, 843–844 (2004). doi:10.1107/S0021889804018047

23. P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010). Medline

24. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010). Medline doi:10.1107/S0907444910007493

25. V. B. Chen, W. B. Arendall 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010). Medline doi:10.1107/S0907444909042073

26. M. M. Babu, M. L. Priya, A. T. Selvan, M. Madera, J. Gough, L. Aravind, K. Sankaran, A database of bacterial lipoproteins (DOLOP) with functional assignments to predicted lipoproteins. *J. Bacteriol.* **188**, 2761–2773 (2006). Medline doi:10.1128/JB.188.8.2761-2773.2006

27. C. A. Schneider, W. S. Rasband, K. W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012). Medline doi:10.1038/nmeth.2089

28. S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, E. Lindahl, GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 (2013). Medline doi:10.1093/bioinformatics/btt055

29. D. H. de Jong, N. Liguori, T. van den Berg, C. Arnarez, X. Periole, S. J. Marrink, Correction to "Atomistic and coarse grain topologies for the cofactors associated with the photosystem II core complex". *J. Phys. Chem. B* **119**, 10372–10372 (2015). Medline doi:10.1021/acs.jpcb.5b06925

30. P. J. Stansfeld, J. E. Goose, M. Caffrey, E. P. Carpenter, J. L. Parker, S. Newstead, M. S. Sansom, MemProtMD: Automated insertion of membrane protein structures into explicit lipid membranes. *Structure* **23**, 1350–1361 (2015). Medline

31. P. J. Stansfeld, M. S. P. Sansom, From Coarse grained to atomistic: A serial multiscale approach to membrane protein simulations. *J. Chem. Theory Comput.* **7**, 1157–1166 (2011). Medline doi:10.1021/ct100569y

32. C. Oostenbrink, A. Villa, A. E. Mark, W. F. van Gunsteren, A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **25**, 1656–1676 (2004). Medline doi:10.1002/jcc.20090

33. R. Chen, L. Li, Z. Weng, ZDOCK: An initial-stage protein-docking algorithm. *Protein* **52**, 80–87 (2003). doi:10.1002/prot.10389

34. S. E. Rollauer, M. J. Tarry, J. E. Graham, M. Jääskeläinen, F. Jäger, S. Johnson, M. Krehenbrink, S. M. Liu, M. J. Lukey, J. Marcoux, M. A. McDowell, F. Rodriguez, P. Roversi, P. J. Stansfeld, C. V. Robinson, M. S. Sansom, T. Palmer, M. Högbom, B. C. Berks, S. M. Lea, Structure of the TatC core of the twin-arginine protein transport system. *Nature* **492**, 210–214 (2012). Medline doi:10.1038/nature11683