LETTER

doi:10.1038/nature25985

Structure of the peptidoglycan polymerase RodA resolved by evolutionary coupling analysis

Megan Sjodt[1], Kelly Brock[2], Genevieve Dobihal[3], Patricia D. A. Rohs[3], Anna G. Green[2], Thomas A. Hopf[2], Alexander J. Meeske[3], Veerasak Srisuknimit[4], Daniel Kahne[4], Suzanne Walker[3], Debora S. Marks[2], Thomas G. Bernhardt[3], David Z. Rudner[3] & Andrew C. Kruse[1]

The shape, elongation, division and sporulation (SEDS) proteins are a large family of ubiquitous and essential transmembrane enzymes with critical roles in bacterial cell wall biology. The exact function of SEDS proteins was for a long time poorly understood, but recent work[1–3] has revealed that the prototypical SEDS family member RodA is a peptidoglycan polymerase—a role previously attributed exclusively to members of the penicillin-binding protein family[4]. This discovery has made RodA and other SEDS proteins promising targets for the development of next-generation antibiotics. However, little is known regarding the molecular basis of SEDS activity, and no structural data are available for RodA or any homologue thereof. Here we report the crystal structure of *Thermus thermophilus* RodA at a resolution of 2.9 Å, determined using evolutionary covariance-based fold prediction to enable molecular replacement. The structure reveals a ten-pass transmembrane fold with large extracellular loops, one of which is partially disordered. The protein contains a highly conserved cavity in the transmembrane domain, reminiscent of ligand-binding sites in transmembrane receptors. Mutagenesis experiments in *Bacillus subtilis* and *Escherichia coli* show that perturbation of this cavity abolishes RodA function both *in vitro* and *in vivo*, indicating that this cavity is catalytically essential. These results provide a framework for understanding bacterial cell wall synthesis and SEDS protein function.

The synthesis of a cell wall and maintenance of its integrity are essential processes for virtually all Eubacteria, and the targeted disruption of cell wall biogenesis is among the most effective of therapeutic strategies in the treatment of bacterial infections. A central step in cell wall synthesis is the concatenation of a lipid II disaccharide pentapeptide headgroup onto a peptidoglycan chain through a glycosyl transfer reaction. This reaction has long been known to be catalysed by the glycosyltransferase domains of class A penicillin-binding proteins[4] (aPBPs). However, deletion of all aPBPs is tolerated in *B. subtilis*[5] and bacteria that lack aPBPs are able to synthesize peptidoglycan[6], which implies the existence of non-aPBP glycosyltransferases. This was recently resolved with the discovery that the highly conserved SEDS membrane proteins comprise a second class of peptidoglycan polymerases[1–3] (Fig. 1a). In fact, SEDS proteins are even more widely distributed than aPBPs[1,7]. Despite their importance and broad phylogenetic distribution, however, SEDS protein function is not well understood and no SEDS protein has yet been characterized structurally.

To better understand SEDS protein function, we pursued structural studies of wild-type *T. thermophilus* RodA as well as a catalytically inactive RodA mutant (D255A). The proteins were expressed, purified and then crystallized using the lipidic cubic phase method. X-ray diffraction datasets were collected to a resolution of 2.9 Å and 3.2 Å for wild-type and D255A mutant proteins, respectively. Because RodA has no homologues of known structure, phase calculation by molecular replacement was impossible. A wide variety of heavy-atom phasing approaches were attempted without success. Having exhausted conventional methods, we sought to develop an approach that requires neither experimental phase data nor the structure of a homologous protein.

Recent methodological advances in molecular replacement have expanded the range of suitable templates[8], and evolutionary co-variation analysis now allows for fold prediction even in the absence of prior structural data[9–11]. This approach exploits the fact that residues that interact with one another structurally tend to co-evolve to maintain their interactions. The analysis of many sequences enables inference of spatial interactions between pairs of residues, providing restraints sufficient to define major features of protein structure. Analysis of RodA by this method showed extensive co-variation throughout the protein (Fig. 1b, c), and we reasoned that using evolutionary coupling restraints to build RodA models might provide suitable templates for molecular replacement phasing, an approach we call 'evolutionary coupling-enabled molecular replacement' (EC-MR).

In brief, our EC-MR approach consisted of the parallel construction and sampling of many independent evolutionary coupling-derived models of RodA to identify suitable templates, followed by phase calculation and model building. First, 100 models of RodA were constructed on the basis of evolutionary restraints. These were each tested as single templates for molecular replacement in Phaser[12]. Of these, 22 models yielded a cluster of solutions that were high-scoring and similar to one another (Fig. 1d). An ensemble search model was constructed from a subset of these, producing maps that were suitable for manual rebuilding followed by ROSETTA refinement in Phenix[13]. The final refined structure showed normal crystallographic statistics (Extended Data Fig. 1, Extended Data Table 1). While previous work has established that structural models constructed *ab initio* can in principle be suitable for molecular replacement phasing in select cases[14,15], these analyses were conducted on very short proteins (<100 amino acids) with high-resolution structural data (<2.1 Å). Other work has shown that at very high resolution, structures can be solved using even a single atom as a search model[16]. At lower resolution, symmetric α-helical proteins can be phased from helical fragments, although this relies on symmetry conditions that are met in only a small minority of cases[17]. Determination of the structure of RodA by EC-MR establishes that evolutionary covariance-derived models can be suitable for phase determination of even a large (359 amino acid) asymmetric protein at modest resolution.

The structures of wild-type RodA and the D255A mutant are virtually identical (Extended Data Fig. 2), and we focus here on the higher-resolution wild-type RodA structure. The overall structure shows ten well-resolved transmembrane helices (TM1–TM10) connected by loops, most of which are well ordered (Fig. 2). Searches for proteins of similar fold with the DALI server[18] yielded no hits, indicating that RodA possesses a unique overall fold. The transmembrane helices of RodA are largely straight and perpendicular to the membrane

[1]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA. [2]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. [3]Department of Microbiology and Immunobiology, Harvard Medical School, Boston, Massachusetts 02115, USA. [4]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA.

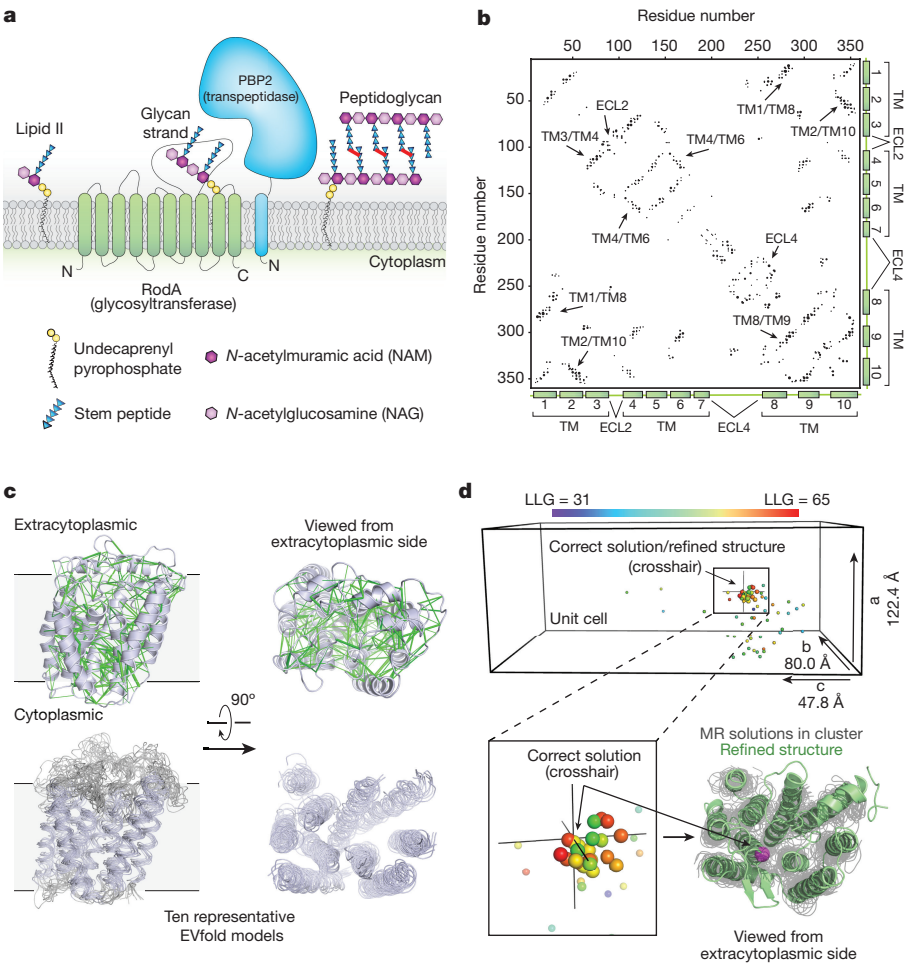


Figure 1 | Biological role of RodA and evolutionary co-variation fold prediction. **a**, RodA is a peptidoglycan polymerase. **b**, Evolutionary co-variation map showing 476 co-evolved residues in RodA. **c**, Evolutionary couplings (green lines) were used to generate models of RodA. Ten representative models are shown. **d**, Molecular replacement (MR) solutions obtained from 100 RodA models. As a reference point, the C_α atom of Glu108 is shown as a sphere coloured according to log-likelihood gain (LLG) score for each solution. The corresponding position for the final refined RodA structure is shown as a crosshair. The refined structure (green) and the C_α atom of Glu108 (magenta sphere) is overlaid with the top MR solutions shown in grey.

plane, with the exception of TM3 that runs diagonally through the membrane with a 45° kink at Pro71. Although some previous studies have suggested that the SEDS protein FtsW could function as a lipid II flippase[19], RodA lacks a transmembrane channel suitable for lipid II transport and bears no notable structural similarity to transporters or flippases.

The intracellular loops of RodA are structured but are very short, and little polar surface area is exposed on the intracellular face of the enzyme. By contrast, extracellular loops (ECLs) 2, 4 and 5 are large and contain many functionally essential residues[1], consistent with catalytic activity occurring at the extracytoplasmic face of the membrane. ECL2 includes a highly conserved β-hairpin, capped by Gly100 and Pro101. ECL4 is even larger at 80 amino acids in length, but is not resolved from residues 189 to 227 as well as from 237 to 251. These regions include essential residues as well as high-ranking evolutionary couplings between ECL4 and ECL1, ECL2 and ECL5, which suggests that they have functionally important roles despite not being resolved in the current structure (Extended Data Figs 3, 4). It is possible that these regions become ordered only upon substrate binding, or in complex with a peptidoglycan crosslinking enzyme. Unlike other ECLs, ECL5 is not exposed to the surface and instead is buried within the protein core.

Between TM2 and TM3 is a long hydrophobic groove containing electron density suggestive of a bound lipid molecule, which we tentatively modelled as monoolein owing to its high concentration in the crystallization conditions (Fig. 2b, c, Extended Data Figs 2, 5). This groove is adjacent to a collection of highly conserved residues (Fig. 2c), and may represent the binding site for the lipid-anchored substrates of RodA. Adjacent to this groove is a large water-filled cavity open to the extracellular face of the protein, flanked by Glu108, Met306, Leu307, Gln310 and Thr342 (Fig. 2d). On the edge of this cavity, Glu108 and Lys111 form an absolutely conserved salt bridge (Fig. 2e). Mapping the results of previous high-throughput mutagenesis in *B. subtilis* RodA onto the structure shows that the salt bridge and other residues in the central cavity are intolerant of substitution[1] (Fig. 3a).

To investigate the function of this central cavity in more detail, we turned to site-directed mutagenesis followed by phenotypic characterization in the representative model Gram-negative and Gram-positive organisms *E. coli* and *B. subtilis*. First, we assessed whether the conserved salt bridge between Glu108 and Lys111is essential, by monitoring the effect of the mutant enzyme when expressed in a wild-type background. Consistent with previous results, mutation of either of these residues to alanine resulted in a dominant-negative phenotype characteristic of Rod complex dysfunction, in which cells lose their elongated morphology and become enlarged and spherical before lysis (Fig. 3b). This suggests that the mutant proteins are properly folded and able to interact with other members of the Rod complex, but are unable to promote cell elongation.

To ascertain whether the salt bridge is catalytic or merely important for proper folding, we constructed a salt-bridge swap mutant (that is, an E108K and K111E double mutant). If the role of the salt bridge is purely structural, this swap should have a minimal effect on RodA function, because the salt bridge is maintained. However, if Glu108 has a role in catalysis, then swapping the two amino acids would be expected to abrogate function. Indeed, the mutant protein folds properly as measured by circular dichroism spectroscopy (Extended Data Fig. 6), but in cell assays the mutant protein shows a strong dominant-negative phenotype (Fig. 3b, c, Extended Data Fig. 7). Moreover, the mutant enzyme has no detectable peptidoglycan polymerization activity *in vitro*, confirming that its toxicity derives from a lack of catalytic activity (Fig. 3d, Supplementary Fig. 1). Other residues near the central cavity were similarly essential for RodA function, including Asp255

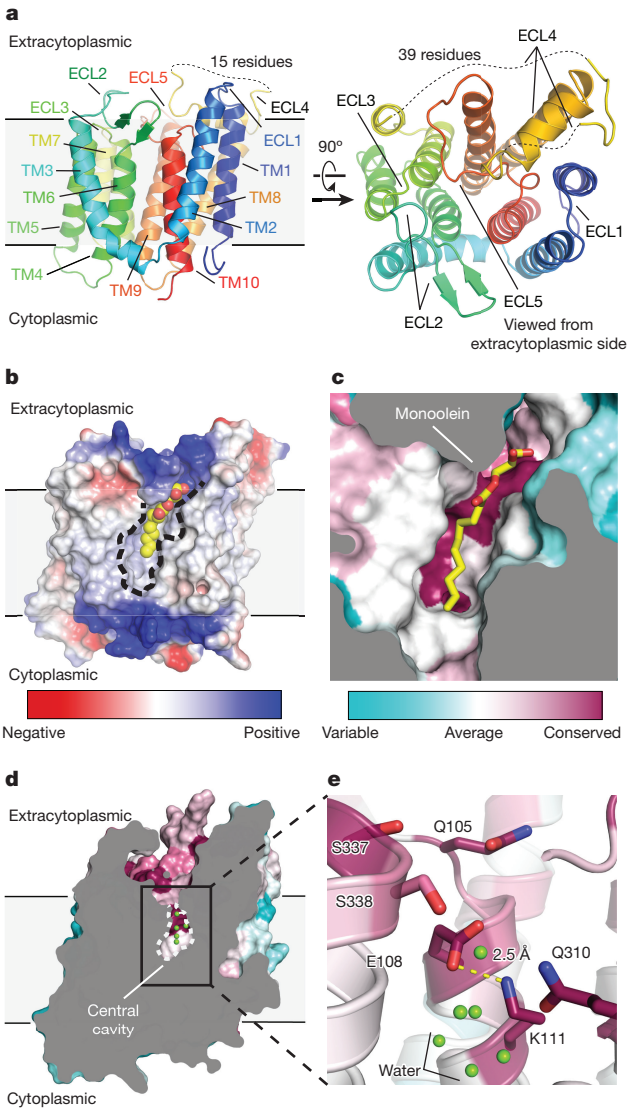


Figure 2 | Structure of RodA. a, Structure of RodA viewed parallel to the membrane plane (left) and from the extracytoplasmic side (right). **b**, Surface view of RodA showing electrostatic potential and the position of a bound lipid between TM2 and TM3 (orientation is identical to that in **a**. **c**, Close-up view of bound lipid (yellow). RodA surface is coloured by sequence conservation. **d**, In the centre of RodA there is a water-filled cavity open to the extracytoplasmic surface. **e**, The water-filled cavity of RodA is flanked by highly conserved polar residues and a salt bridge (yellow dashed line) between Glu108 and Lys111.

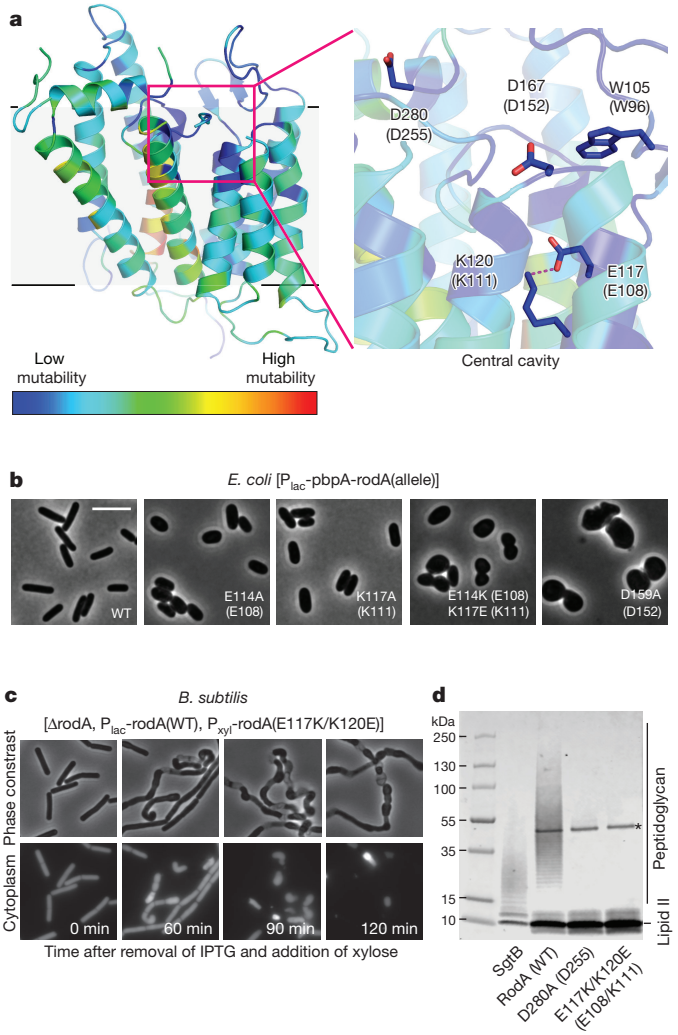


Figure 3 | The central cavity is essential for RodA function. a, Homology model of *B. subtilis* RodA with each residue coloured according to its tolerance for mutation (see Methods). Residues are numbered according to *B. subtilis*, with corresponding *T. thermophilus* numbering in parentheses. **b**, Wild-type *E. coli* cells harbouring a plasmid with indicated RodA mutants. Scale bar, 5 μm. WT, wild type. **c**, Expression of RodA charge-swap mutant is toxic in *B. subtilis*. Intracellular mCherry expression indicates cytosol. For both **b** and **c**, images are representative of three independent experiments. IPTG, isopropyl-β-D-thiogalactoside. **d**, Catalytic competence of *B. subtilis* RodA mutants, representative of two independent experiments. Asterisk indicates the PBP4 labelling enzyme used for detection; StgB, SgtB(Y181D) used as control.

as previously reported[1], as well as Asp152 (Fig. 3b, Extended Data Fig. 7). Analysis of evolutionary co-variation data using EVmutation[20] likewise predicts these residues, and the salt bridge, to be immutable. Taken together, the high degree of sequence conservation, intolerance to mutation and catalytic essentiality of residues surrounding the central cavity confirm that this portion of the protein has a critical role in peptidoglycan polymerization, which makes it a prime target for the development of antibiotics.

The glycan-strand polymerization process catalysed by RodA and aPBPs is essential but not sufficient to build a cell wall. A second key step is peptide crosslinking, catalysed by the penicillin-binding domains of aPBPs and class B penicillin-binding proteins (bPBPs). Cytological and protein–protein interaction studies indicate that SEDS proteins and bPBPs are likely to form a complex in cells[1,2], and evolutionary coupling analysis shows strong co-variation between bPBP and SEDS protein sequences. This is sufficient to map the binding site between bPBPs and RodA to TM8 and TM9 (Fig. 4a, b, Extended Data Fig. 8), corresponding to the proposed interaction site between the divisome SEDS protein FtsW and its corresponding bPBP, FtsI[10,21,22]. RodA mutants in this interface exhibit a dominant-negative effect in *E. coli* (Fig. 4c). However, mutation of this site does not prevent RodA-mediated peptidoglycan polymerization *in vitro* (Extended Data Fig. 7), confirming the functional importance of coordinated glycan strand elongation and peptide crosslinking in cells.

Complexes between SEDS proteins and bPBPs contain both glycan-strand polymerization and peptide crosslinking active sites, recapitulating the dual catalytic activities found in aPBPs (Fig. 4d). Understanding how SEDS proteins coordinate their activity with that of bPBPs to build a cell wall will be an important area for future investigation. The structure of RodA now provides a foundation for such work, serving as a framework for understanding the function of SEDS proteins.

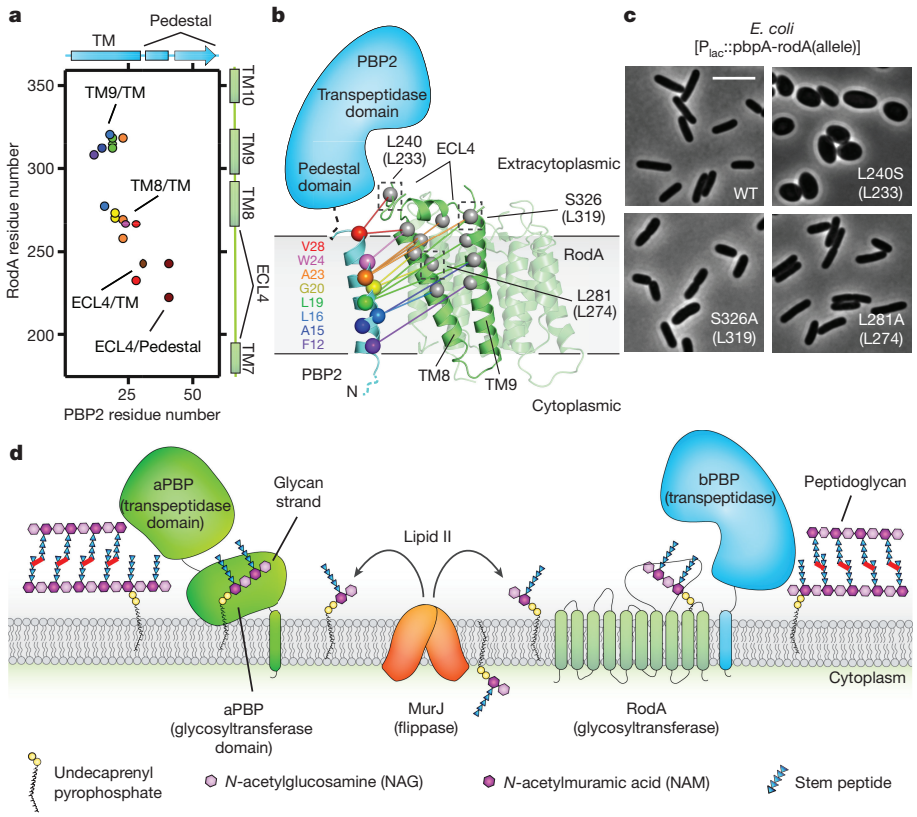


Figure 4 | Interaction between RodA and its class B pencillin-binding protein, PBP2. a, Evolutionary co-variation map showing 19 evolutionary couplings between RodA and PBP2. **b**, Representation of evolutionary co-variation. **c**, Mutations in RodA–bPBP interface result in morphological abnormalities. Mutations were made in *E. coli* RodA, with *T. thermophilus* numbering in parentheses. Data are representative of two independent experiments. Scale bar, 5 μm. **d**, Model of peptidoglycan biogenesis. Lipid II is flipped across the membrane by MurJ, and polymerized and crosslinked by a complex of a SEDS protein (RodA) and bPBP, or by a bifunctional aPBP.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 October 2017; accepted 8 February 2018.
Published online 28 March 2018.

1. Meeske, A. J. *et al.* SEDS proteins are a widespread family of bacterial cell wall polymerases. *Nature* **537,** 634–638 (2016).
2. Cho, H. *et al.* Bacterial cell wall biogenesis is mediated by SEDS and PBP polymerase families functioning semi-autonomously. *Nat. Microbiol.* **1,** 16172 (2016).
3. Emami, K. *et al.* RodA as the missing glycosyltransferase in *Bacillus subtilis* and antibiotic discovery for the peptidoglycan polymerase pathway. *Nat. Microbiol.* **2,** 16253 (2017).
4. Goffin, C. & Ghuysen, J.-M. Multimodular penicillin-binding proteins: an enigmatic family of orthologs and paralogs. *Microbiol. Mol. Biol. Rev.* **62,** 1079–1093 (1998).
5. McPherson, D. C. & Popham, D. L. Peptidoglycan synthesis in the absence of class A penicillin-binding proteins in *Bacillus subtilis*. *J. Bacteriol.* **185,** 1423–1431 (2003).
6. Packiam, M., Weinrick, B., Jacobs, W. R., Jr & Maurelli, A. T. Structural characterization of muropeptides from *Chlamydia trachomatis* peptidoglycan by mass spectrometry resolves 'chlamydial anomaly'. *Proc. Natl Acad. Sci. USA* **112,** 11660–11665 (2015).
7. Otten, C., Brilli, M., Vollmer, W., Viollier, P. H. & Salje, J. Peptidoglycan in obligate intracellular bacteria. *Mol. Microbiol.* **107,** 142–163 (2018).
8. DiMaio, F. *et al.* Improved molecular replacement by density- and energy-guided protein structure optimization. *Nature* **473,** 540–543 (2011).
9. Hopf, T. A. *et al.* Three-dimensional structures of membrane proteins from genomic sequencing. *Cell* **149,** 1607–1621 (2012).
10. Hopf, T. A. *et al.* Sequence co-evolution gives 3D contacts and structures of protein complexes. *eLife* **3,** e03430 (2014).
11. Marks, D. S. *et al.* Protein 3D structure computed from evolutionary sequence variation. *PLoS ONE* **6,** e28766 (2011).
12. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
13. DiMaio, F. *et al.* Improved low-resolution crystallographic refinement with Phenix and Rosetta. *Nat. Methods* **10,** 1102–1104 (2013).
14. Qian, B. *et al.* High-resolution structure prediction and the crystallographic phase problem. *Nature* **450,** 259–264 (2007).
15. Rigden, D. J., Keegan, R. M. & Winn, M. D. Molecular replacement using *ab initio* polyalanine models generated with ROSETTA. *Acta Crystallogr. D* **64,** 1288–1291 (2008).
16. McCoy, A. J. *et al. Ab initio* solution of macromolecular crystal structures without direct methods. *Proc. Natl Acad. Sci. USA* **114,** 3637–3641 (2017).
17. Strop, P., Brzustowicz, M. R. & Brunger, A. T. *Ab initio* molecular-replacement phasing for symmetric helical membrane proteins. *Acta Crystallogr. D* **63,** 188–196 (2007).
18. Holm, L. & Rosenström, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38,** W545–W549 (2010).
19. Mohammadi, T. *et al.* Identification of FtsW as a transporter of lipid-linked cell wall precursors across the membrane. *EMBO J.* **30,** 1425–1432 (2011).
20. Hopf, T. A. *et al.* Mutation effects predicted from sequence co-variation. *Nat. Biotechnol.* **35,** 128–135 (2017).
21. Ovchinnikov, S. *et al.* Large-scale determination of previously unsolved protein structures using evolutionary information. *eLife* **4,** e09248 (2015).
22. Leclercq, S. *et al.* Interplay between penicillin-binding proteins and SEDS proteins promotes bacterial cell wall synthesis. *Sci. Rep.* **7,** 43306 (2017).

Supplementary Information is available in the online version of the paper.

Acknowledgements Financial support for the work was provided by NIH grant U19AI109764 (A.C.K., D.Z.R., T.G.B., S.W. and D. K.), NIH grant R01GM106303 (D.S.M.) and a CIHR doctoral research award to P.D.A.R. We thank Advanced Photon Source GM/CA beamline staff for technical support during X-ray data collection, K. Arnett (Harvard Center for Macromolecular Interactions) for support of circular dichroism experiments and C. Sander for discussions.

Author Contributions M.S. and A.J.M. performed expression screening experiments, and M.S. performed large-scale purification and crystallization of RodA as well as enzyme assays and circular dichroism spectroscopy. Additional input regarding enzyme assays was provided by P.D.A.R, V.S., D.K. and S.W. The structure was solved and refined by M.S. and A.C.K. using evolutionary coupling-derived models developed by K.B., A.G.G., T.A.H. and D.S.M. Assessment of RodA mutant phenotypes was conducted by G.D. and P.D.A.R. with supervision from T.G.B. and D.Z.R. Overall project supervision was performed by A.C.K. with input from T.G.B. and D.Z.R. The manuscript was written by M.S. and A.C.K. with input from other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to A.C.K. (andrew_kruse@hms.harvard.edu).

Reviewer Information *Nature* thanks R. Read, K. Young and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

Protein purification. RodA from *T. thermophilus* was cloned into pAM172 plasmid[1] using EcoR1 and AvrII restriction enzymes (resulting in plasmids pMS211 and pMS224 for wild-type and D255A RodA, respectively). The expression plasmids contain an amino-terminal SUMO-fusion followed by a Flag epitope tag and a 3C protease cleavage site (SUMO–Flag–3C–RodA and SUMO–Flag–3C–RodA(D255A)), and were transformed into *E. coli* C43 derivative of BL21 (DE3) harbouring an arabinose-inducible Ulp1 protease plasmid (pAM174) under the selection for both plasmids. Five fresh transformants harbouring both plasmids were inoculated into 5 ml LB medium supplemented with 100 μg ml^{-1} ampicillin and 35 μg ml^{-1} chloramphenicol and allowed to grow overnight at 37 °C in a rolling shaker. The 5-ml overnight culture was then diluted in 1 litre of TB broth supplemented with 0.1% glucose, 2 mM $MgCl_2$, 100 μg ml^{-1} ampicillin, and 35 μg ml^{-1} chloramphenicol. Cultures were grown at 37 °C until an OD_{600} of 0.6 and shifted to 20 °C. At an OD_{600} of 0.8, protein expression was induced by addition of IPTG (1 mM final) and arabinose (0.2% final) for RodA and Ulp1, respectively. After a 16-h induction, cells were collected and frozen at −80 °C.

Cells were resuspended in lysis buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 20 mM $MgCl_2$, 100 μg ml^{-1} lysozyme, 1:100,000 (v/v) benzonase nuclease, and 2 mg ml^{-1} iodoacetamide), lysed by sonication and membranes were collected by ultracentrifugation at 100,000*g* for 1 h at 4 °C. Flag–3C–RodA was then extracted using a glass dounce tissue grinder in a solubilization buffer containing 20 mM HEPES pH 7.5, 500 mM NaCl, 20% (v/v) glycerol, 2 mg ml^{-1} iodoacetamide, and 1% (w/v) *N*-dodecyl β-D-maltoside (DDM; Anatrace). Samples were stirred for 2 h at 4 °C, then centrifuged as above for 1 h. The supernatant containing solubilized RodA was supplemented with 2 mM $CaCl_2$ and loaded by gravity flow onto 4 ml anti-Flag antibody affinity resin. The resin was washed extensively, first in 50 ml of buffer containing 20 mM HEPES pH 7.0, 500 mM NaCl, 2 mM $CaCl_2$, 20% glycerol, 0.1% DDM, and then in 50 ml of the same buffer supplemented with 10 mM adenosine 5′-triphosphate magnesium salt and 20 mM KCl to remove the bacterial chaperones, GroEL and DnaK. RodA was eluted in 20 mM HEPES pH 7.0, 500 mM NaCl, 20% glycerol, 0.1% DDM supplemented with 5 mM EDTA and 0.2 mg ml^{-1} Flag peptide. 3C protease was added (1:1,000 w/w) and incubated with RodA at 4 °C overnight. RodA was further purified by size exclusion chromatography (SEC) on a Sephadex S200 column (GE Healthcare) in buffer containing 20 mM HEPES pH 7.5, 500 mM NaCl and 0.1% DDM. After preparative SEC, the protein was concentrated to 30–40 mg ml^{-1} and flash-frozen with liquid nitrogen in aliquots of 8 μl. Samples were stored at −80 °C until use for crystallography. Purity and monodispersity of crystallographic samples was evaluated by SDS–PAGE and analytical SEC, respectively.

Crystallography and data collection. Purified *T. thermophilus* wild-type and D255A RodA were reconstituted into lipidic cubic phase by mixing with a 10:1 (w/w) mix of monoolein (Hampton Research) with cholesterol (Sigma Aldrich) at a ratio of 1.0:1.5 protein:lipid by mass, using the coupled syringe reconstitution method[23]. All samples were mixed at least 100 times before dispensing. The resulting phase was dispensed in 15–40 nl drops onto a glass plate and overlaid with 600 nl of precipitant solution using a Gryphon LCP robot (Art Robbins Instruments). Crystals grew in precipitant solution containing 35–50% PEG 200, 100 mM NaCl, 100 mM $MgCl_2$ and 100 mM Tris pH 7.6–8.2. Initial crystallization hits grew within 24 h, with diffraction-quality crystals reaching full size over the course of 2–4 weeks. Crystals were collected using mesh loops and stored in liquid nitrogen until data collection. Data collection was carried out at Advanced Photon Source GM/CA beamline 23ID-B. An initial grid raster with 80 × 30-μm beam dimensions was performed to locate crystals within the loop. Additional fine-tuning rasters were performed using a 10-μm beam diameter to optimize the position of the crystal for data collection. Data were collected using a 10-μm beam and 0.2°-oscillation width per frame at a wavelength of 1.033 Å and a fivefold attenuation factor. For both wild-type and D255A RodA, a complete dataset was obtained from a single crystal. Diffraction data were indexed and processed using XDS[24]. Both wild-type and D255A RodA crystallized in the C2 space group with one molecule in the asymmetric unit, with a solvent content of approximately 60%.

Generation of *ab initio* evolutionary coupling-derived models of RodA. Multiple sequence alignments (MSA) of the full length *T. thermophilus* RodA (Uniprot ID Q5SIX3) were generated using the iterative hidden Markov model-based sequence search tool jackhmmer[25] with five iterations. Alignments were built using the Uniref100 dataset[26] released in April 2017. An MSA was generated for nine different bitscores, a sequence inclusion threshold was normalized to length and expressed as the number of bits per residue, with values ranging from the most inclusive (0.1) to the least inclusive (0.9). The alignment depth was chosen to optimize the number of non-redundant sequences with the fewest gaps in the alignment as previously described[9,10], although the alignment choice was robust with respect to consistency of predicted evolutionary couplings over a wide range of alignment depths. Blindly optimizing the alignment choice resulted in two alignments, with the smaller one at 31,505 non-redundant sequences of length-346 residues (10–355) and an 'effective number' of 8,729 sequences after down-weighting sequences with more than 80% identity and no more than 30% gaps in any columns used for the EC model computation.

Model folding criteria. For the chosen MSAs, evolutionary couplings were determined using a pseudo-likelihood maximization (PLMC)[27–29]. A mixture model approach identified the 99% percentile probability of being in contact. We used a combination of commonly used prediction methods to determine secondary structure predictions (PSIPRED[30] and PolyPhobius[31]), together with secondary structure propensity computed directly from local evolutionary couplings[29]. This resulted in the identification of ten transmembrane helices and three smaller helices in ECL4, and two β-strands in ECL2. A total of 220 folded models for *T. thermophilus* RodA were generated for increasing numbers of evolutionary coupling restraints using the folding protocol in EVfold[9], which uses a distance geometry and simulated annealing protocol in CNS[32,33]. All models were ranked as previously described[11,34] and the 50 top-ranked models from each of the MSAs were used as molecular replacement search models (see below). To generate additional models used for further ensemble-based phasing runs, five additional models were generated using the top-ranked fold prediction from bitscore 0.8 with an additional round of dedicated simulated annealing in CNS. Cartesian dynamics were used for heating, torsion for cooling and 600 final minimization steps with 50 cycles were used with other parameters kept as defaults. In addition to evolutionary couplings scores, for folded models EVfold generates an evolutionary coupling enrichment score for each residue that reflects how constrained that residue is, which can be thought of as conservation of 'coupling'.

Phasing and refinement. One hundred evolutionary coupling-derived models of wild-type RodA were primed for molecular replacement using Sculptor[35], which included limited side-chain pruning and B-factor assignment based on accessible surface area[36]. Each model was tested as a single search template for molecular replacement in Phaser[12], resulting in 98 out of the 100 initial models providing candidate solutions. The solutions were sorted by translation function *Z*-score (TF-Z) and the top 32 solutions manually inspected in PyMOL[37]. A total of 22 solutions were highly similar and distinct from all other solutions. Sorting the solutions by log-likelihood gain (LLG) provided a similar result, in which all 22 solutions were among the top 40% of solutions. However, most of these were not well separated from other potential solutions in their respective searches, typically with two-to-four other candidate solutions giving LLG values at least 75% of that of the top solution. The models giving the top two solutions (as judged by both TF-Z and LLG metrics) were used as a single ensemble for phase calculation followed by manual building using COOT[38] and reciprocal space refinement using phenix.refine[39]. Manual refinement was complemented by the use of ROSETTA refinement in Phenix[13]. Subsequently, additional cycles of manual building and reciprocal space refinement led to the final refined structure.

Verification of sequence register was straightforward and unambiguous owing to the relatively high resolution and frequency of bulky amino acid side chains (RodA is roughly 10% tryptophan, tyrosine and phenylalanine). Representative composite omit map density is shown in Extended Data Fig. 1. The structure of RodA(D255A) was solved using wild-type RodA as the molecular replacement search model and the resulting refined structure is nearly identical to that of wild-type RodA (0.1 Å r.m.s.d. between all C_α atoms). After refinement, the quality of both structures was assessed using MolProbity to calculate Ramachandran statistics and other parameters[40], and figures were prepared in PyMOL. All crystallographic data processing, refinement and analysis software was compiled and supported by the SBGrid Consortium[41].

Predicting residue contacts between RodA and PBP2. EVComplex[10] was used to predict inter-protein contacts between *T. thermophilus* full-length RodA and full-length PBP2 (Uniprot ID Q5SJ23). We constructed alignments for RodA (33,670 sequences) and PBP2 (40,764 sequences) as described above for RodA alone, using the April 2017 Uniprot release[42] for clarity on species identifiers. We concatenated RodA and PBP2A sequences from each species when they were within 10,000 nucleotides of each other, based on European Nucleotide Archive data downloaded in February 2017[43], and evolutionary couplings were computed on the complex alignment as previously described[10]. A mixture model approach was used to identify top-scoring contacts, both within individual monomers and between RodA and PBP2, and to assign a probability to each contact of being within the tail of the distribution of coupling scores as previously described[29]. This scoring method is accurate for predicting whether proteins interact as well as which residues are in contact[10]. The distribution of evolutionary couplings is approximated by a Gaussian log-normal mixture model, and we defined the tail of the distribution as those scores that have >95% probability of belonging to

the log-normal component. Evolutionary couplings in this tail defined as high probability resulted in 26 residue pairs predicted as contacts between RodA and PBP2.

Sequence and structure conservation analysis. The sequence conservation analysis shown in Fig. 3 and Extended Data Fig. 5 was computed using the ConSurf server[44]. In brief, a multiple sequence alignment of *T. thermophilus* RodA to its closest 150 homologues was generated using the HHMER algorithm provided by ConSurf, with conservation scores plotted in PyMOL. Additionally, a multiple sequence alignment of RodA to 506 homologues from representative bacterial taxa was generated using a protein sequence BLAST search on the NCBI public database using *T. thermophilus* RodA protein sequence as query, and plotted in Extended Data Fig. 3.

Homology modelling and mutability index. A homology model of *B. subtilis* RodA based on the structure of *T. thermophilus* RodA was constructed using MODELLER[45]. In brief, owing to the low sequence identity (~26%) between the two enzymes, a multiple sequence alignment of 10 RodA homologues from diverse bacterial taxa was performed (*T. thermophilus*, *B. subtilis*, *E. coli*, *Staphylococcus aureus*, *Streptococcus pneumoniae*, *Shigella flexneri*, *Haemophilus influenzae*, *Deinococcus marmoris*, *Bacillus safensis* and *Lysinibacillus odysseyi*). The resulting alignment was used as the search template in MODELLER, in which disordered residues 204–256 and 266–276 were omitted from the model generation. A total of ten models were generated and the top-scoring model was chosen for further analysis.

The mutability index of each residue in the homology model of *B. subtilis* RodA was calculated based on previously reported MutSeq data[1]. First, for each residue the nonsynonymous mutation rate was defined as the number of mutations observed after excluding synonymous and nonsense mutations. To calculate a raw mutability index, the nonsynonymous mutation rate was divided by the summed rate of silent mutations observed for a given nucleotide position. Finally, the raw mutability index for each residue was normalized with respect to the expected silent mutation rate for each general class of nucleotide change, resulting in a per cent mutability index for each residue. A low mutability index (that is, low percentage) represents a residue that is relatively intolerant of mutation at that position and a high mutability index (high percentage) represents a residue that is relatively tolerant of changes at that position.

RodA mutant plasmid and strain construction. Mutants of *E. coli* RodA were introduced using QuikChange mutagenesis into a plasmid containing the allele P_{lac}-pbpA-rodA(WT) (pHC857)[1,2]. Wild-type *E.coli* cells (strain TB28) were then transformed with the mutant plasmids to generate strains that each harboured mutations in the central cavity (E114A, K117A, E114K/K117E and D159A) and at the RodA–PBP2 interface (L240S, S326A and L281A).

Mutations of *B. subtilis* RodA were introduced using QuikChange mutagenesis into the pER174a plasmid (*amyE::*P_{xyl}*-rodA(WT)-10×His*)[1]. The resulting mutagenic plasmids (*amyE::*P_{xyl}*-rodA(allele)-10×His*) was directly transformed into the *B. subtilis* strain (*rodA::kan*, P_{spank}*-rodA(WT)-10×His,* and P_{pen}*sacA::mCherry*) to generate strains each harbouring mutations in the central cavity (Q114A, E117A, K120A, E117K/K120E, D280N and S364A), the proposed lipid-binding cavities (I113A, I113S, and F118A and A171L), and the RodA-PBP2 interface (F292A). The *B. subtilis* strains harbouring *amyE::*P_{xyl}*-rodA-(D167A), amyE::*P_{xyl}*-rodA(D167N), and amyE::*P_{xyl}*-rodA(D280A)* have previously been described[1].

Mutational analysis *in vivo*. All *E. coli* cultures were grown at 37 °C in M9 minimal medium supplemented with 0.2% (w/v) maltose, 0.2% (w/v) casamino acids, and 25 μg ml^{-1} chloramphenicol. Overnight cultures were diluted to an OD_{600} of 0.05 and grown to OD_{600} = 0.25 in the absence of inducer. These cultures were then further diluted to an OD_{600} of 0.005 in medium containing 1 mM IPTG to induce expression of the pbpA–rodA construct. Cells were grown for five generations in the presence of inducer and fixed when the OD_{600} reached between 0.15 and 0.2. Fixative solution contained 0.04% glutaraldehyde, 4% formaldehyde, 32 mM sodium phosphate, pH 7.5. Wide-field phase-contrast microscopy was performed on a Nikon TE2000 microscope equipped with a 100× Plan Apo 1.4 NA objective and a CoolSNAP HQ2 monochrome camera.

B. subtilis strains were derived from the PY79 prototrophic strain[46]. Cells were grown in LB medium at 37 °C in the presence of 10 μM IPTG to express wild-type RodA. When cultures reached mid-log, cells were washed three times with plain LB medium and resuspended in 20 ml of LB medium to an OD_{600} of 0.02, then supplemented with 10 mM xylose to induce expression of RodA wild type and variants. Cells were analysed by phase-contrast microscopy 90 min later. Microscopy was performed on a Nikon Ti microscope equipped with Plan Apo 100×/1.4NA phase-contrast oil objective and a CoolSNAP HQ2 camera. Cells were immobilized using 2% agarose pads, containing growth medium. Images were cropped and adjusted using MetaMorph software (Molecular Devices).

Mutant analysis *in vitro*. Mutants of *B. subtilis* RodA were transformed into *E. coli* strain CAM333, a derivative of strain C43 with deletions in *ponB*, *pbpC* and *mgtA*, and were expressed and purified by immunoaffinity chromatography as previously described[1].

Before circular dichroism spectroscopy experiments, each RodA mutant was dialysed into a buffer consisting of 10 mM sodium phosphate pH 7.5, 500 mM potassium fluoride, 0.5% (w/v) CHAPS, and 0.05% (w/v) DDM. Spectra were acquired on a Jasco J-815 spectropolarimeter. Circular dichroic spectra were recorded between 200–260 nm using a quartz cuvette with a path length of 1 mm, a 50-nm/min scanning speed, and a bandwidth of 1 nm. Five spectra were measured at 25 °C, averaged and corrected for buffer contribution. Secondary structure assessment was not performed owing to high absorbance contributions at wavelengths less than 205 nm from the detergent mixture necessary for protein stability and function (0.5% CHAPS, 0.05% DDM).

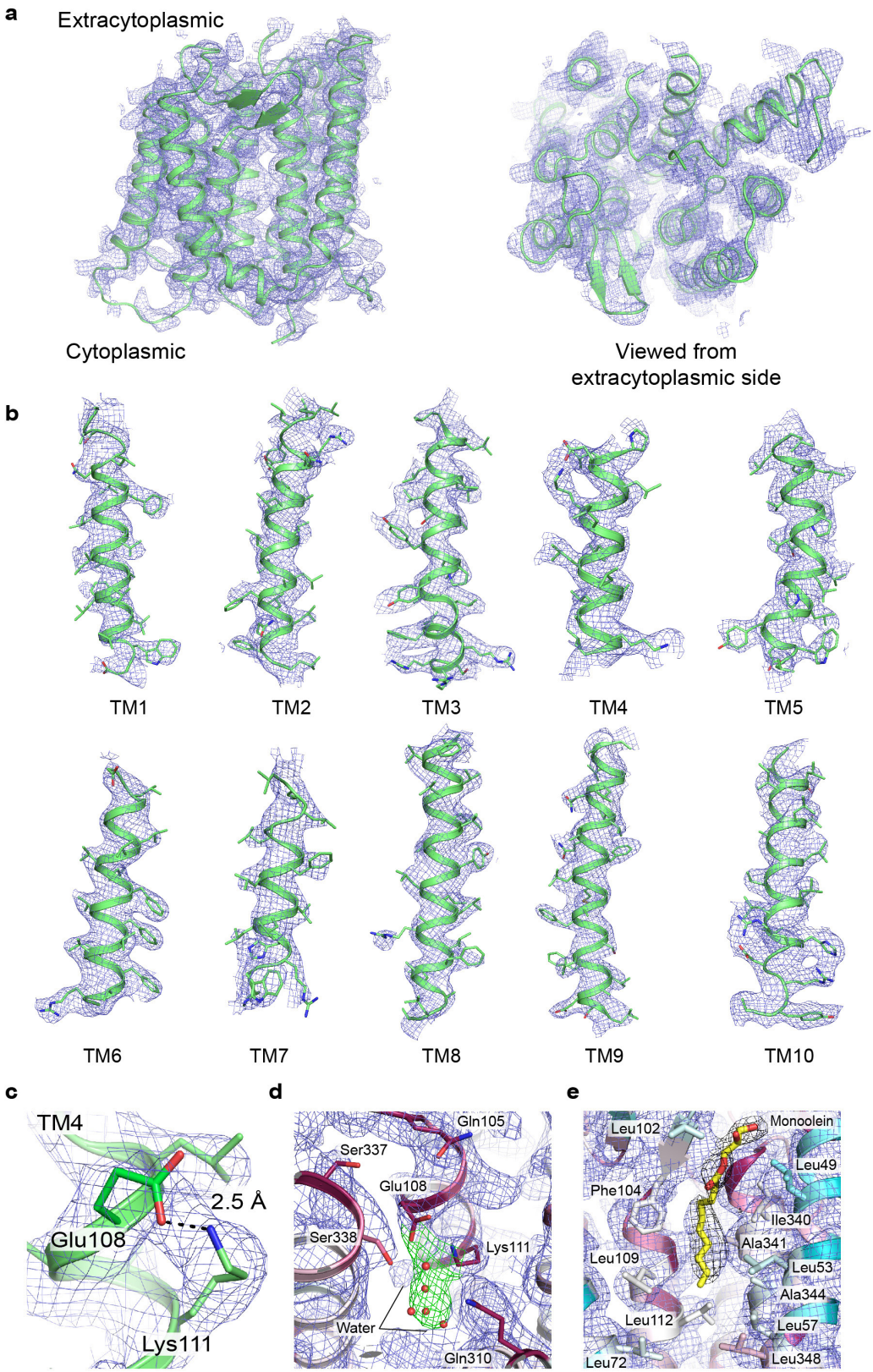
For assessment of enzymatic activity, lipid II substrate was purified from *E. coli* as described[47]. Peptidoglycan polymerization reactions were adapted from previously described methods[48]. In brief, purified *B. subtilis* wild-type and mutant RodA proteins were incubated with lipid II in reaction buffer containing 50 mM HEPES pH 7.0, 20 mM $CaCl_2$, 20 mM $MgCl_2$ and 20% DMSO. The working concentration of RodA and lipid II were 1 μM and 20 μM, respectively. RodA was purified in 0.5% CHAPS and 0.05% DDM, and therefore the working concentration for CHAPS and DDM was 0.05% and 0.005%, respectively. As a positive control lipid II was also polymerized with SgtB(Y181D) (1 μM) in reaction buffer containing 12.5 mM HEPES pH 7.5, 2 mM $MnCl_2$, 0.25 mM Tween-80 and 20% DMSO. All reactions were incubated at 25 °C for 1 h and quenched by incubation at 95 °C for 2 min. Peptidoglycan biotinylation of each reaction mixture was performed by addition of biotinylated D-lysine (1.5 mM, final concentration) and PBP4 (3.8 μM, final concentration) followed by incubation at 25 °C for 1 h. The biotinylation reaction was then quenched by addition of 13 μl 2× SDS loading dye. The samples were then loaded into a 4–20% gradient polyacrylamide gel and run at 180 V. The products were transferred onto a PVDF membrane (BioRad) and fixed in 0.4% paraformaldehyde diluted in PBS for 30 min at room temperature. The membrane was blocked with SuperBlock TBS blocking buffer (Thermo Fisher Scientific) for 1 h at room temperature and the biotinylated products were detected by incubation with fluorescently tagged streptavidin (IRDye 800-CW streptavidin (Li-Cor Biosciences, 1:5,000 in SuperBlock)) for an additional 30 min. Membranes were washed 4 × 10 min with TBST (0.01% Tween 20) and given a final 10-min wash in PBS. Blots were then visualized using an Odyssey CLx imaging system (LI-COR Biosciences).

Code availability. The full EVfold software package is available at https://github.com/debbiemarkslab/EVcouplings.

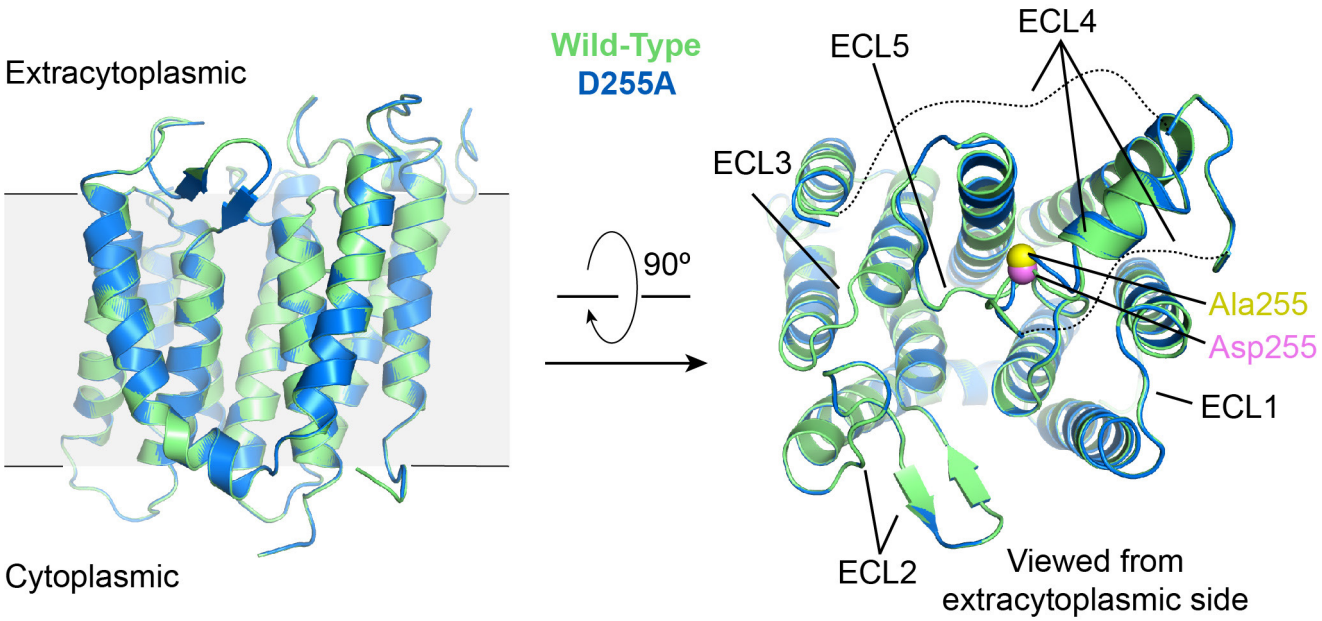
Data availability. Structure factors and refined atomic coordinates for RodA wild type and the RodA(D255A) mutant are deposited in the RCSB Protein Data Bank under accession codes 6BAR and 6BAS, respectively.

23. Caffrey, M. & Cherezov, V. Crystallizing membrane proteins using lipidic mesophases. *Nat. Protoc.* **4,** 706–731 (2009).
24. Kabsch, W. XDS. *Acta Crystallogr. D* **66,** 125–132 (2010).
25. Johnson, L. S., Eddy, S. R. & Portugaly, E. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* **11,** 431 (2010).
26. Suzek, B. E., Wang, Y., Huang, H., McGarvey, P. B. & Wu, C. H. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* **31,** 926–932 (2015).
27. Ekeberg, M., Lövkvist, C., Lan, Y., Weigt, M. & Aurell, E. Improved contact prediction in proteins: using pseudolikelihoods to infer Potts models. *Phys. Rev. E* **87,** 012707 (2013).
28. Balakrishnan, S., Kamisetty, H., Carbonell, J. G., Lee, S.-I. & Langmead, C. J. Learning generative models for protein fold families. *Proteins* **79,** 1061–1078 (2011).
29. Toth-Petroczy, A. *et al.* Structured states of disordered proteins from genomic sequences. *Cell* **167,** 158–170.e12 (2016).
30. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292,** 195–202 (1999).
31. Käll, L., Krogh, A. & Sonnhammer, E. L. L. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res.* **35,** W429–W432 (2007).
32. Brunger, A. T. Version 1.2 of the Crystallography and NMR system. *Nat. Protoc.* **2,** 2728–2733 (2007).
33. Brünger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).
34. Marks, D. S., Hopf, T. A. & Sander, C. Protein structure prediction from sequence variation. *Nat. Biotechnol.* **30,** 1072–1080 (2012).
35. Bunkóczi, G. & Read, R. J. Improvement of molecular-replacement models with Sculptor. *Acta Crystallogr. D* **67,** 303–312 (2011).
36. Schwarzenbacher, R., Godzik, A., Grzechnik, S. K. & Jaroszewski, L. The importance of alignment accuracy for molecular replacement. *Acta Crystallogr. D* **60,** 1229–1236 (2004).
37. Schrodinger. *The PyMOL Molecular Graphics System, Version 1.3r1* (Schrodinger, 2010).

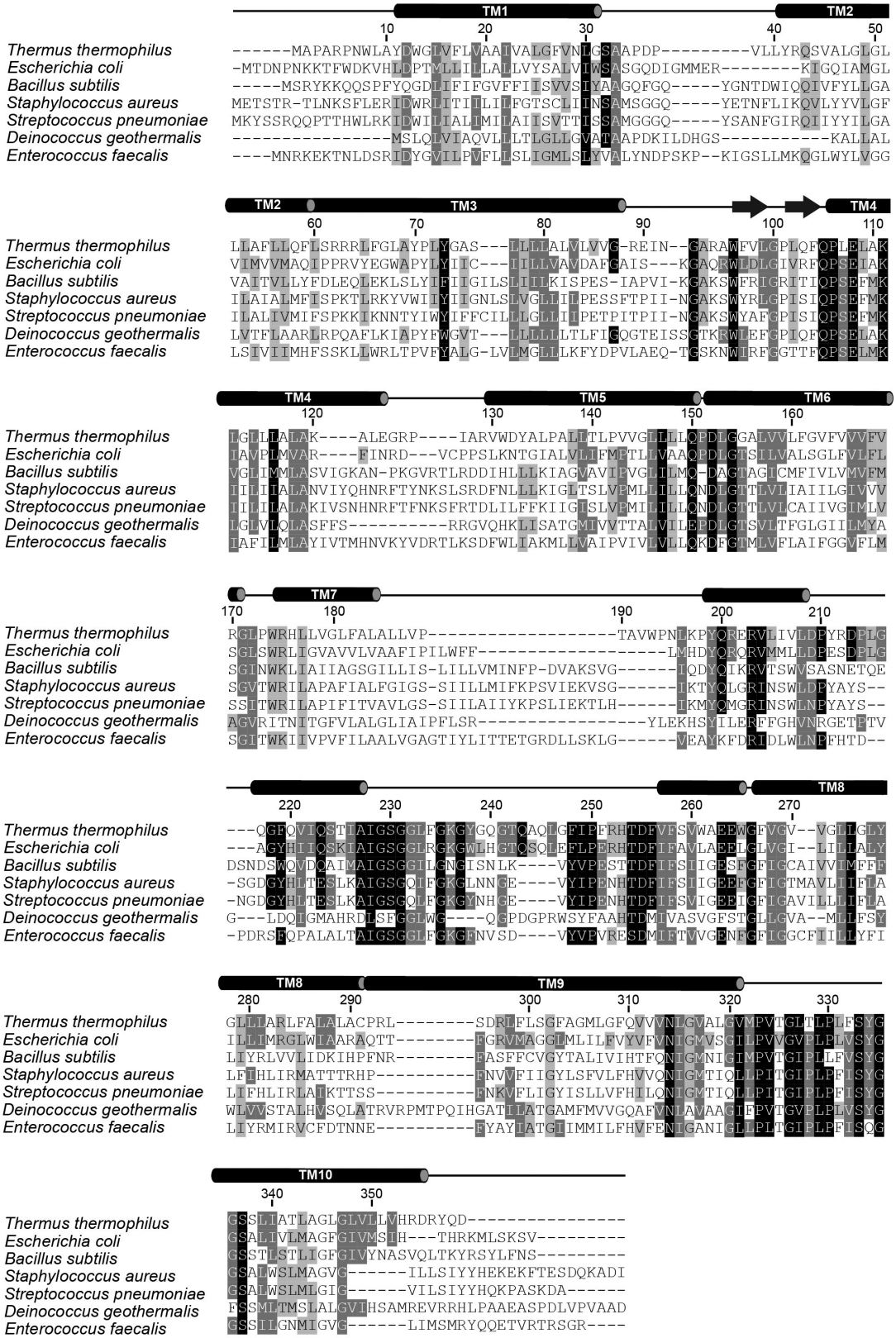
38. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
39. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* **68,** 352–367 (2012).
40. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66,** 12–21 (2010).
41. Morin, A. *et al.* Collaboration gets the most out of software. *eLife* **2,** e01456 (2013).
42. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45,** D158–D169 (2017).
43. Pakseresht, N. *et al.* Assembly information services in the European Nucleotide Archive. *Nucleic Acids Res.* **42,** D38–D43 (2014).
44. Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **38,** W529–W533 (2010).
45. Webb, B. & Sali, A. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein Sci.* **54,** 5.6.1–5.6.37 (2016).
46. Youngman, P. J., Perkins, J. B. & Losick, R. Genetic transposition and insertional mutagenesis in *Bacillus subtilis* with *Streptococcus faecalis* transposon Tn917. *Proc. Natl Acad. Sci. USA* **80,** 2305–2309 (1983).
47. Qiao, Y. *et al.* Lipid II overproduction allows direct assay of transpeptidase inhibition by β-lactams. *Nat. Chem. Biol.* **13,** 793–798 (2017).
48. Qiao, Y. *et al.* Detection of lipid-linked peptidoglycan precursors by exploiting an unexpected transpeptidase reaction. *J. Am. Chem. Soc.* **136,** 14678–14681 (2014).



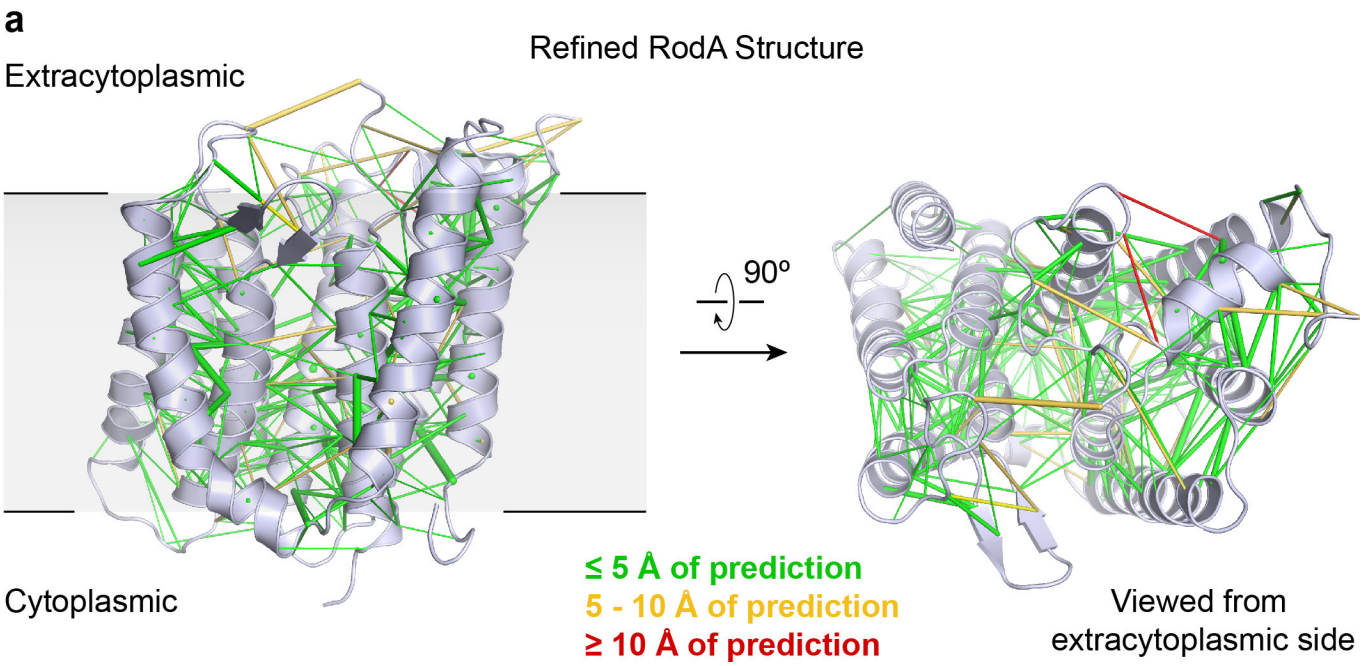
Extended Data Figure 1 | Representative electron density. a–c, Simulated annealing composite omit $2F_o - F_c$ electron density map of *T. thermophilus* RodA contoured at 1.0σ within a 2.0 Å radius of atoms shown. **d**, The same map contoured at 1.0σ and coloured blue and green for RodA and water molecules, respectively. The modelled water molecules are shown as red spheres. **e**, The same map contoured at 1.0σ within a 3.0 Å radius RodA (shown in blue) and the same map contoured at 1.0σ for monoolein (shown in black).

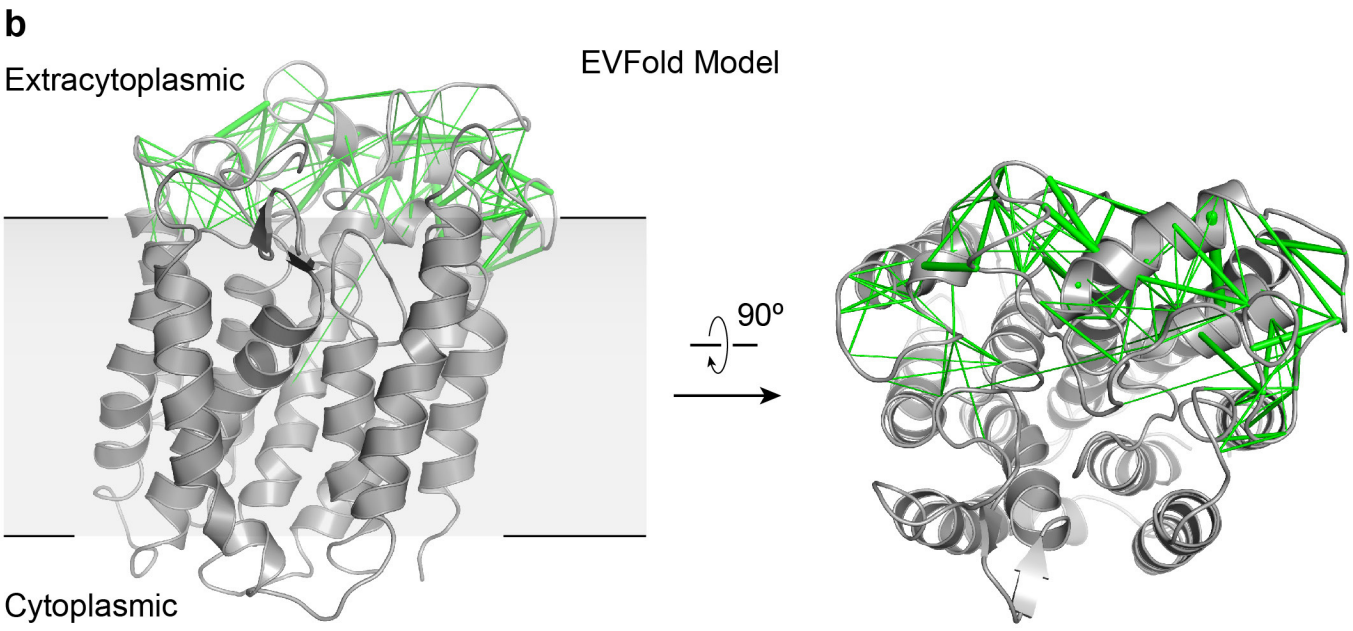


Extended Data Figure 2 | Comparison of RodA wild type and RodA(D255A) structures. Structures of wild-type (green) and D255A (blue) RodA are shown viewed parallel to the membrane (left) and from the extracytoplasmic side (right). The C_α atom of residue 255 for each structure is shown as a sphere, and coloured pink (wild type) or yellow (D255A). The dashed lines represent the disordered residues 189–227 and 237–251 in both structures. The two structures are essentially identical, with a C_α r.m.s.d. of 0.1 Å.

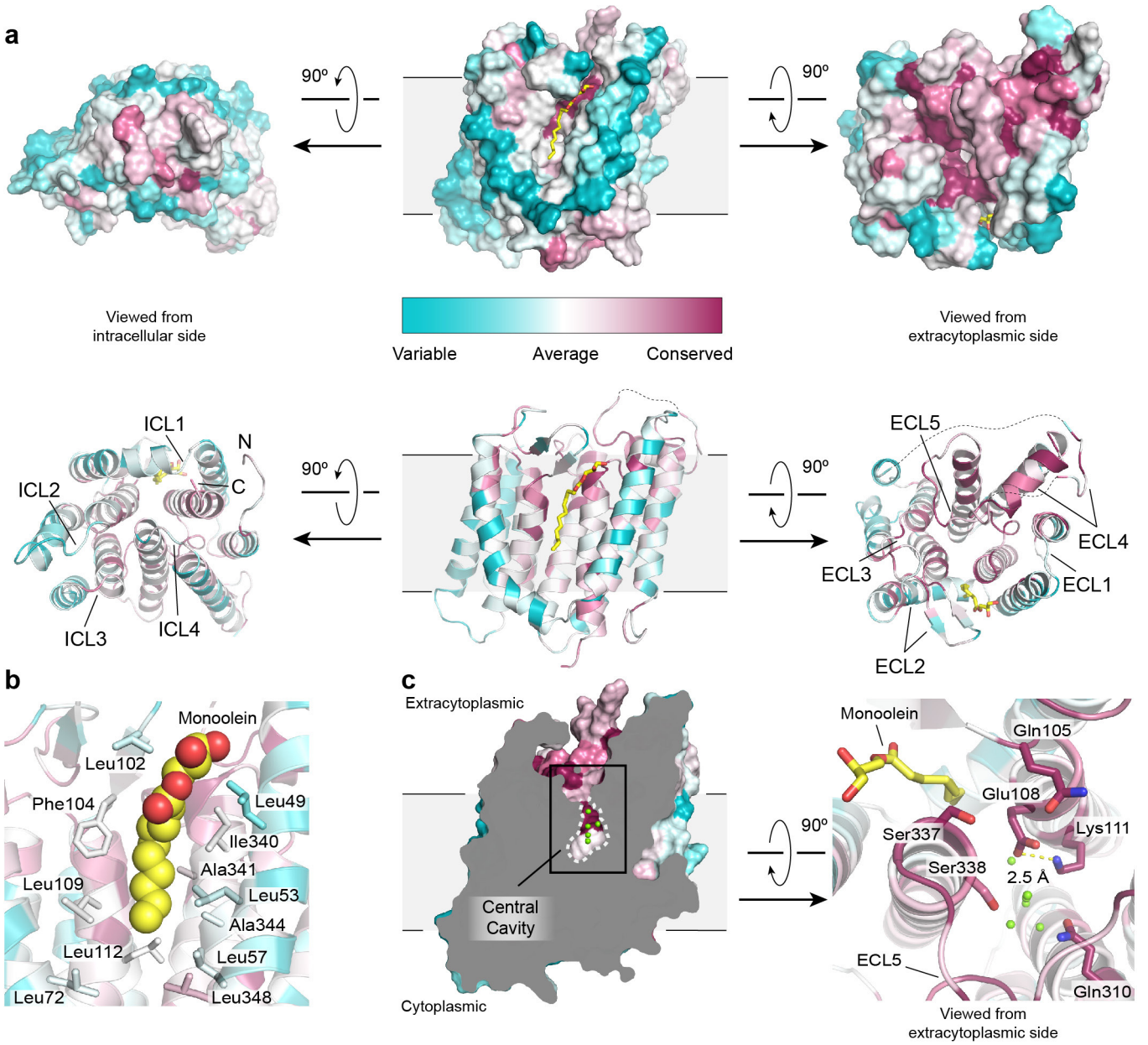


Extended Data Figure 3 | RodA sequence conservation. The results of an alignment of 506 RodA sequences from diverse bacterial taxa, with representative examples displayed. Residues with 98%, 80% and 60% similarity across all 506 sequences are shown in black, grey and light grey, respectively. Secondary structure elements are shown above the alignment on the basis of the *T. thermophilus* RodA crystal structure and JPRED analysis of the portions of ECL4 that were not modelled in the structure.

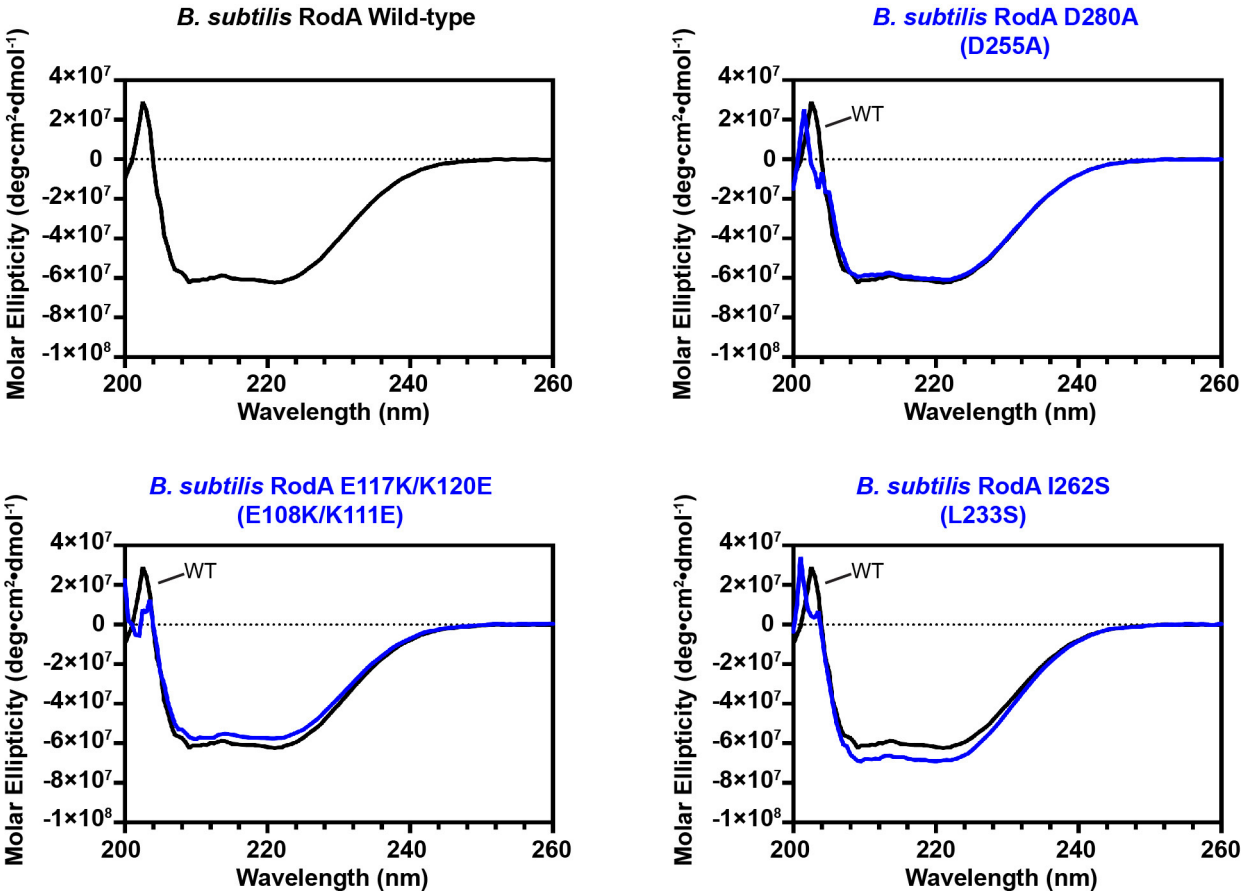




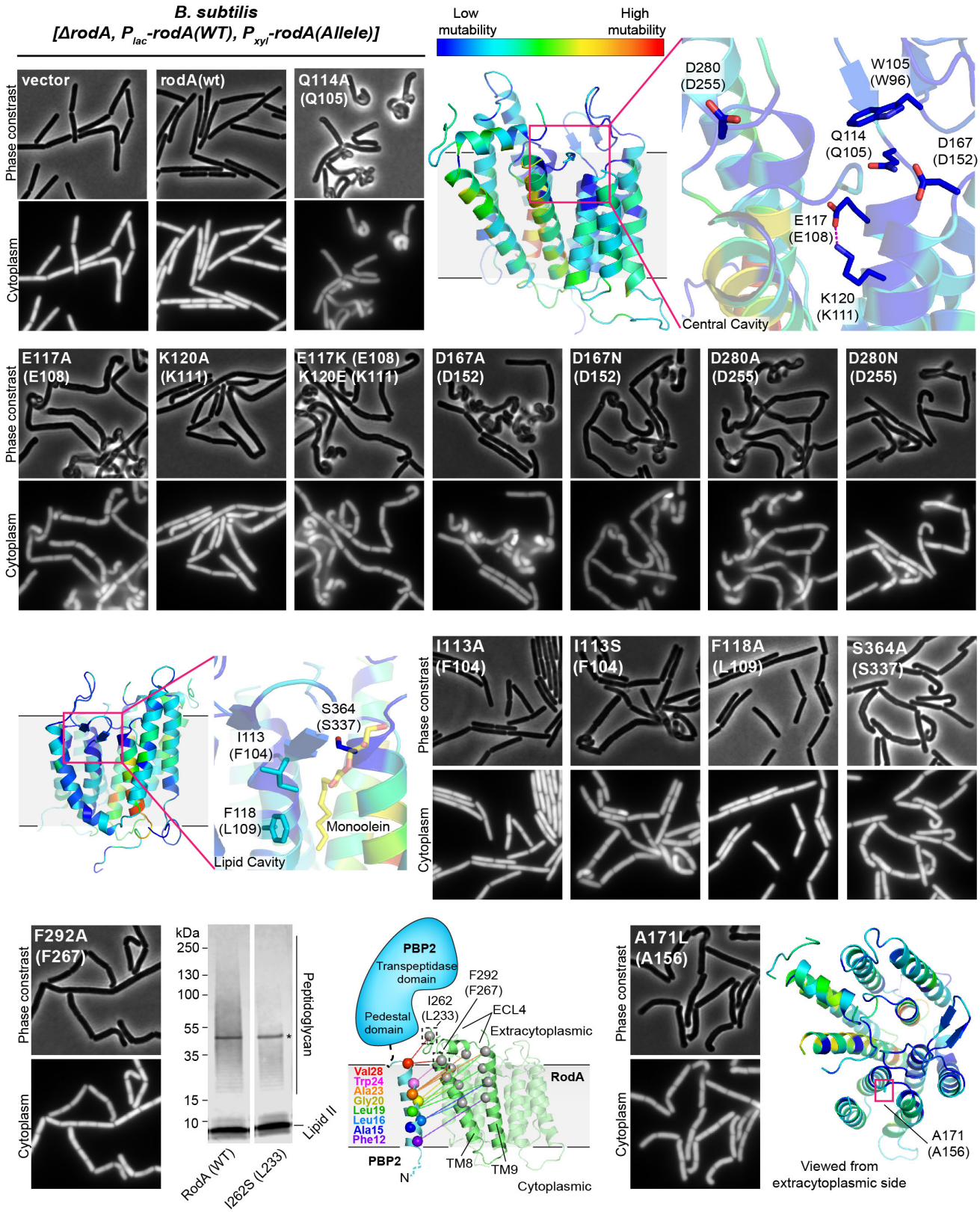
Extended Data Figure 4 | RodA evolutionary couplings. a, The refined crystal structure of *T. thermophilus* RodA (shown in light blue) is in close agreement with its evolutionary couplings. Green, yellow and red lines between residues represent regions of the structure that are less than 5 Å, between 5 and 10 Å, and greater than 10 Å of the predicted evolutionary couplings, respectively. **b**, The partially disordered ECL4 in the refined structure is strongly coupled evolutionarily to the transmembrane domain of RodA. A representation of predicted intra- and inter-domain evolutionary couplings for ECL4 is mapped onto a EVfold model (shown in grey).



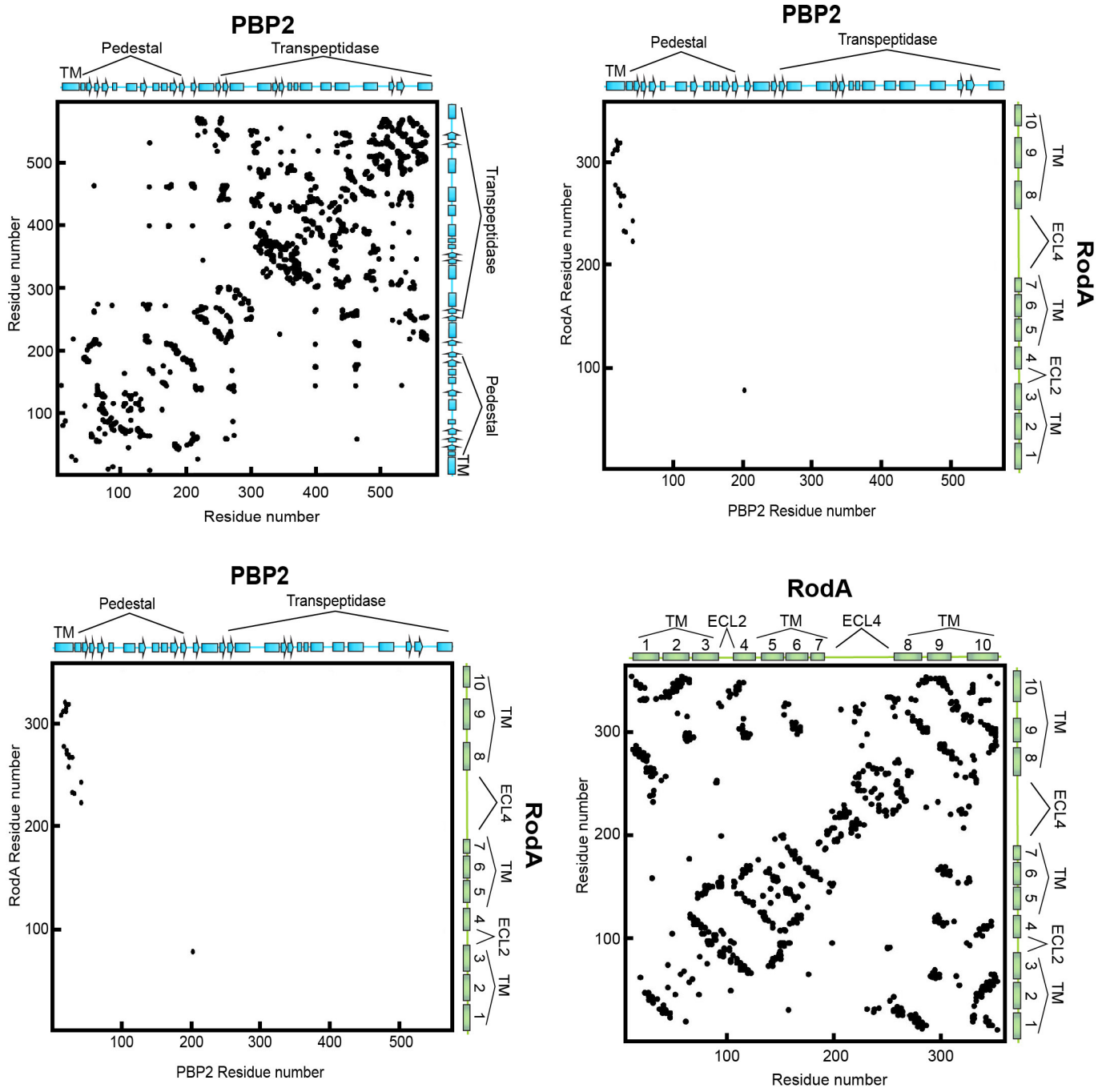
Extended Data Figure 5 | Sequence conservation of *T. thermophilus* RodA. a, Surface and ribbon representation of RodA (top and bottom, respectively). Analysis was performed using Consurf, coloured in a scale from teal (poorly conserved) to magenta (highly conserved). A bound lipid modelled as monoolein is shown in yellow sticks. The dashed lines represent disordered residues 189–227 and 237–251 in ECL4. **b**, The bound lipid (shown as spheres) is surrounded by many aliphatic amino acids (shown as sticks). **c**, Extracytoplasmic view of the water-filled central cavity and its proximity to the bound lipid molecule.



Extended Data Figure 6 | Circular dichroism spectroscopy analysis of purified *B. subtilis* RodA variants. Circular dichroism measurements of wild-type RodA as well as RodA(D280A), RodA(E117K/K120E) and RodA(I262S) indicate that the overall folds of all four forms are comparable and display characteristic α-helical peaks at 208 and 222 nm. For each panel, the circular dichroic spectra of wild-type and the indicated mutant RodA are shown in black and blue, respectively. The corresponding *T. thermophilus* numbering for each mutant is shown in parentheses. Experiments were repeated independently twice with similar results.



Extended Data Figure 7 | Mutagenic analysis of *B. subtilis* RodA function *in vivo*. Micrographs of *B. subtilis* strains harbouring an IPTG-inducible allele of wild-type *rodA*, and wild-type or mutant P_{xylA}-rodA. Expression was induced by growing cells in the presence of 10 μM IPTG and 10 mM xylose. The bacterial cytosol is indicated by intracellular mCherry expression (P_{pen}-mCherry). Mutants in the central cavity show particularly deleterious phenotypes in this dominant negative assay. Experiments were repeated independently 2–3 times with similar results. A mutation made in the predicted RodA–bPBP interface does not prevent peptidoglycan polymerization *in vitro* (lower left panel), representative of two independent experiments.



Extended Data Figure 8 | Evolutionary coupling analysis for the RodA–PBP2 complex. Evolutionary co-variation maps highlighting 693 couplings for PBP2 (top left panel), and 362 couplings for RodA (bottom right panel) using a 95% confidence threshold. These maps display good correlation to the crystal structure of RodA and homology models of PBP2. The top right and lower bottom panels depict the 19 predicted inter-protein contacts between RodA and PBP2 using the same 95% confidence threshold.

Extended Data Table 1 | Data collection and refinement statistics

	T. thermophilus RodA Wild-type	*T. thermophilus* RodA D255A
Data collection		
Space group	C2	C2
Cell dimensions		
a, *b*, *c* (Å)	122.4, 80.0, 47.8	121.2, 79.2, 47.4
α, β, γ (°)	90.0, 91.1, 90.0	90.0, 91.1, 90.0
Resolution (Å)	40.0 - 2.9 (3.1 - 2.9)*	40.0 - 3.2 (3.4 - 3.2)*
R_{sym} (%)	10.6 (147.9)	17.1 (204.9)
$\langle I/\sigma(I)\rangle$	6.69 (0.69)	4.08 (0.44)
$CC_{1/2}$ (%)	99.7 (45.8)	99.6 (41.7)
Completeness (%)	98.4 (98.8)	99.6 (99.5)
Multiplicity	3.02 (3.15)	3.43 (3.35)
Refinement		
Resolution (Å)	40.0 - 2.91 (2.99 - 2.91)	40.0 – 3.19 (3.31 – 3.19)
No. reflections	18684 (1851 in test set)	13756 (1360 in test set)
R_{work} / R_{free} (%)	22.9 / 27.4	27.7 / 30.2
No. atoms		
Protein	2229	2201
Solvent ions/lipids	44	1
Water	52	8
Average *B*-factors (Å^2)		
Protein	104.2	111.0
Lipids	109.6	-
Solvent ions	105.4	125.9
Waters	87.6	89.9
R.m.s. deviations		
Bond lengths (Å)	0.005	0.002
Bond angles (°)	0.825	0.751
Ramachandran Statistics		
Favored (%)	94.5	93.5
Allowed (%)	5.5	6.5
Outliers (%)	0.0	0.0

Each dataset was collected from a single crystal.
*Values in parentheses are for the highest-resolution shell.

nature research

Corresponding author(s): Andrew C. Kruse

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

   Describe how sample size was determined. — Not applicable

2. Data exclusions

   Describe any data exclusions. — No data were excluded.

3. Replication

   Describe whether the experimental findings were reliably reproduced. — All experiments were reproducible

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups. — Not applicable

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis. — No blinding was applied

   Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

   For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☒	☐	The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☒	☐	A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☒	☐	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	All software is referenced in the methods section of the manuscript. Custom software is available at GitHub and is linked in the manuscript.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	No restrictions on materials

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	Not applicable

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	Not applicable
b. Describe the method of cell line authentication used.	Not applicable
c. Report whether the cell lines were tested for mycoplasma contamination.	Not applicable
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	Not applicable

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	Not applicable

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	Not applicable