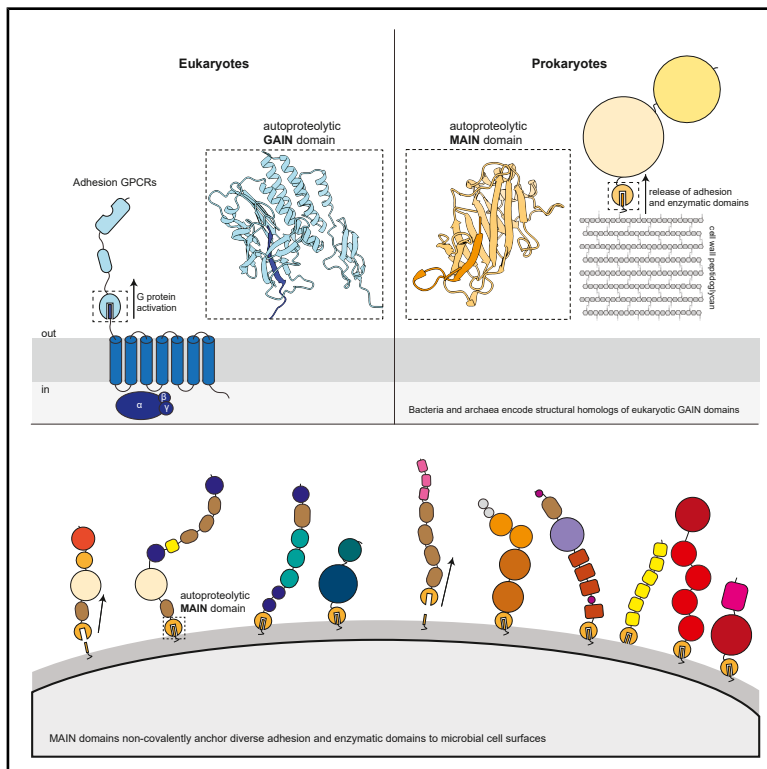


Current Biology

Microbial GAIN domains undergo autoproteolysis and enable release of diverse cell-surface-associated proteins

Graphical abstract



Authors

Anna P. Brogan, David Z. Rudner

Correspondence

rudner@hms.harvard.edu

In brief

Brogan and Rudner report the identification of a microbial autoproteolytic domain that is structurally homologous to the GAIN domain of eukaryotic adhesion GPCRs. Rather than participating in transmembrane signaling, these domains tether diverse domains to the cell surface and are proposed to release them in response to mechanical stimuli.

Highlights

- Structural homologs of GAIN domains are conserved across bacteria and archaea
- MAIN domains are autoproteolytic, and their cleavage products remain associated
- MAIN domains anchor diverse adhesins and enzymes to the bacterial cell surface
- MAIN-anchored domain release likely requires force, analogous to GPCR signaling

Article

Microbial GAIN domains undergo autoproteolysis and enable release of diverse cell-surface-associated proteins

Anna P. Brogan¹ and David Z. Rudner^{1,2,*}

¹Department of Microbiology, Harvard Medical School, Boston, MA 02115, USA

²Lead contact

*Correspondence: rudner@hms.harvard.edu

<https://doi.org/10.1016/j.cub.2026.07.052>

SUMMARY

Adhesion G protein-coupled receptors (aGPCRs) transduce mechanical stimuli across the cytoplasmic membrane in eukaryotes. These receptors contain extracellular GPCR autoproteolysis-inducing (GAIN) domains that undergo autoproteolysis but maintain stable associations of their cleavage products. A diverse set of adhesion domains appended to the GAIN domain binds surface ligands on neighboring cells or the extracellular matrix. Shear force is thought to disrupt the interaction between the cleavage products, exposing a tethered agonist that triggers GPCR signaling. Here, we report that proteins with structural homology to GAIN domains are broadly conserved among bacteria and archaea. The microbial domains lack strong sequence conservation to their eukaryotic counterparts but are predicted to adopt a similar fold. We demonstrate that these microbial autoproteolysis-inducing (MAIN) domains undergo autoproteolysis both *in vitro* and *in vivo*, using conserved catalytic residues. Furthermore, proteolysis occurs in a conserved β turn that allows stable non-covalent interaction between the cleavage products. MAIN domains are tethered to the cell envelope of bacteria and archaea and are fused to diverse sets of adhesion and enzymatic domains. Many of the same adhesion domains are appended to both MAIN and GAIN domains, suggesting these protein families share a common origin and function. We propose that MAIN domains allow microbes to release proteins from their cell surface in response to shear force, enabling broader nutrient scavenging, intoxication of neighboring cells, and dispersal through surface detachment.

INTRODUCTION

G protein-coupled receptors (GPCRs) are the largest family of cell-surface receptors in eukaryotes. In response to external stimuli, these receptors catalyze the activation of associated G proteins that mediate intracellular responses.¹ While most GPCRs respond to chemical stimuli, a subset, known as adhesion GPCRs (aGPCRs), respond to mechanical stimuli.^{2–4} Like all family members, aGPCRs contain seven transmembrane (7TM) segments that undergo a conformational change in response to stimulus to activate their associated G protein. The defining feature of the adhesion family is the presence of an extracellular GPCR autoproteolysis-inducing (GAIN) domain (Figure 1A) that is fused to the 7TM signaling domain. GAIN domains undergo constitutive autoproteolysis in a loop between two β strands,⁵ but the cleavage products remain non-covalently associated through an uninterrupted β sheet. Diverse adhesion domains that bind surface ligands on neighboring cells or the extracellular matrix (Figure 1A) are appended to the GAIN domain. The human genome encodes 33 aGPCRs. Each contains a GAIN domain fused to a large ectodomain with multiple adhesion domains and intrinsically disordered regions.⁶ Binding of these domains to surface ligands or to the matrix enables force sensing by the aGPCR.⁷ A pulling or shear force is thought

to be transduced to the GAIN domain, causing dissociation of its cleavage products. Removal of the ectodomain reveals a β strand known as the Stachel sequence or tethered peptide^{8,9} that is fused to the 7TM domain. The tethered peptide functions as an agonist, inducing GPCR signaling.² Thus, aGPCRs function as mechanosensors that signal in response to mechanical rather than chemical stimuli.

In addition to their role in aGPCR signaling, GAIN domains are found in a larger protein family of cell-surface receptors, including all five human polycystic kidney disease (PKD) proteins.¹⁰ This protein family is conserved in primitive eukaryotic ancestors,^{5,10} suggesting that the GAIN domain constitutes an ancient and broadly conserved mechanosensory domain.

We recently reported that bacteria encode an autoproteolytic domain called the SEAL domain (also known as the CRE and PD domain) that functions in a mechanotransduction pathway that is distinct from but resembles the role of the GAIN domain in aGPCR signaling.^{11–13} To uncover additional mechanosensory domains in prokaryotes, we searched for GAIN-like proteins in bacterial and archaeal genomes. We report that microbial autoproteolysis-inducing (MAIN) domains that are structurally homologous to GAIN domains are encoded in many archaea and bacteria. We show that MAIN domains undergo autoproteolysis in a manner similar to GAIN domains and remain non-covalently

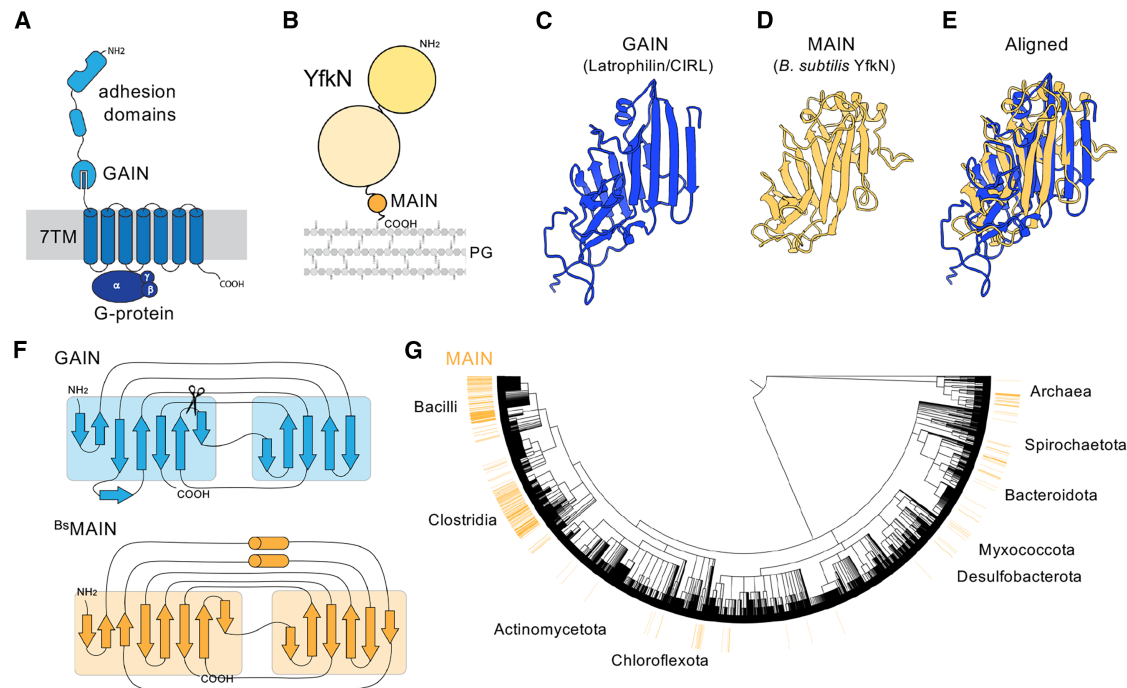


Figure 1. MAIN domains are widespread among bacteria and archaea

(A) Schematic of a prototypical adhesion GPCR. Gray represents the membrane.
(B) Schematic of *B. subtilis* YfkN. The protein is covalently anchored to the PG and contains a MAIN domain and two enzymatic domains.
(C) Crystal structure of the GAIN_B domain from latrophilin/CIRL (PDB: 4dlq).
(D) AlphaFold2-predicted structure of the MAIN domain from *B. subtilis* YfkN.
(E) Structural alignment by TM-align of the structures from (B) and (C).
(F) Protein topology maps of latrophilin/CIRL's GAIN_B domain and YfkN's MAIN domain as generated by PDMsum. Scissor depicts the site of autocleavage.
(G) Phylogenetic tree built from 5,767 representative taxa. The amino acid sequence of B^SMAIN was used to query the RefSeq database, and the resulting species containing a MAIN domain are highlighted in orange.
See also Figure S1.

associated after cleavage. Interestingly, MAIN domains are fused to diverse enzymatic and adhesion domains, several of which are also appended to GAIN domains. However, rather than functioning in transmembrane signaling, MAIN domains are tethered to the microbial cell surface with their enzymatic and adhesion domains facing the extracellular environment. Our findings support a model in which MAIN domains allow microbes to release hydrolytic enzymes, toxins, and adhesion domains from their cell surface in response to shear forces. We conclude that the GAIN domain fold, and likely its role in force sensing, are conserved throughout all domains of life, consistent with a common evolutionary origin.

RESULTS

GAIN domains are widespread throughout bacteria and archaea

The lack of reported microbial GAIN domains prompted us to investigate whether this domain or a structurally similar one exists in bacteria or archaea. Using the crystal structure of the GAIN domain from the latrophilin/CIRL1 aGPCR (Figure 1C),⁵ we queried the structural homology search server Foldseek.¹⁴ Strikingly, several bacterial proteins with a highly similar fold were identified, including a strong hit in a protein (YfkN) encoded

in *Bacillus subtilis* (Figure 1D). The GAIN domain consists of an α -helical rich GAIN_A and a β -sandwich GAIN_B domain, which is the site of autoproteolysis. The AlphaFold2-predicted structure of the *B. subtilis* domain aligned to the latrophilin/CIRL1 GAIN_B domain with a TM-align score¹⁵ of 0.63 and a root-mean-squared deviation of 3.42 Å (Figures 1E and S1), consistent with metrics expected from structurally homologous proteins. In addition, the topology map of the YfkN domain was nearly identical to GAIN_B, supporting the idea that they have a common fold and likely share a common ancestor (Figure 1F). Due to their structural similarity, we refer to these and homologous microbial domains as MAIN domains.

B. subtilis YfkN contains a signal peptide, predicted phosphodiesterase and nucleotidase domains, followed by a MAIN domain, and a C-terminal sortase recognition sequence. This sequence functions to covalently anchor the secreted protein to the cell wall via the activity of the enzyme sortase (Figure 1B).¹⁶ YfkN's enzymatic domains have been shown to cleave cyclic nucleotides *in vitro*,^{17,18} but the physiological function of the protein is unknown. Our sequence-based bioinformatic analysis indicates that MAIN domains are found throughout bacteria and archaea with various protein architectures. They are not only predominantly enriched in Bacillota, including Bacilli and Clostridia, but also found in Bacteroides,

Spirochetes, and Archaea, among other phyla (Figure 1G). The cell-surface anchoring of YfkN is characteristic of MAIN domain-containing proteins throughout bacteria and is discussed below.

MAIN domains are autoproteolytic

In addition to the predicted structural homology and fusion to large ectodomains, MAIN and GAIN domains share some sequence similarity. Specifically, BsMAIN and latrophilin/CIRL1 GAIN_B share 8.3% sequence identity and 16.6% sequence similarity. Importantly, the autocatalytic site and adjacent sequences are well conserved (Figures 2A, 2B, S3, and S4). GAIN domains undergo autoproteolysis using a conserved HX(S/T) motif, where X represents a hydrophobic amino acid. Briefly, the histidine deprotonates the serine or threonine residue, enabling attack on the adjacent amide bond to form an unstable tetrahedral intermediate. The protein then undergoes an N-O acyl shift, forming an ester intermediate, which is readily hydrolyzed by water to create two protein fragments.^{19,20} It has been shown that mutating the catalytic residue to alanine abolishes autoproteolysis and that it can be restored by mutating the serine or threonine interchangeably.²⁰ The HXS motif is among the most highly conserved regions of bacterial MAIN domains (Figures 2A and S4). To investigate whether *B. subtilis*' MAIN domain undergoes autoproteolysis, we fused a hexa-histidine tag to the N terminus of the MAIN domain and a SUMO protease tag to the C terminus (His-BsMAIN-SUMO). Upon expression in *Escherichia coli* and purification on Ni²⁺-agarose, we observed three polypeptides. The largest corresponded to the size of the full-length fusion protein. The sizes of the smaller two polypeptides were consistent with autocleavage products (Figure 2C). Importantly, similar amounts of the two cleavage products were co-purified despite the fact that only one of them had a His-tag, suggesting that they form a stable complex after autocleavage. Autoproteolysis was abrogated by mutation of serine-1346 to alanine (S1346A) and restored by mutation to threonine (S1356T) (Figure 2C). Size exclusion chromatography of the purified His-BsMAIN-SUMO and the S1346A variant revealed that the autocleavage products and full-length protein have identical retention times (Figures 2D and 2E). Collectively, these results argue that the mechanism of autoproteolysis is conserved between MAIN and GAIN domains and indicate that their cleavage products remain non-covalently associated.

To determine the autoproteolytic cleavage site, we used Edman degradation of the purified His-BsMAIN-SUMO protein. Cleavage occurred between the phenylalanine and serine residues of the HFS motif at the same position where GAIN domains undergo autoproteolysis²⁰ (Figures 2F and S5). Cleavage in this loop of the β sheet retains extensive hydrogen bonding in the undisturbed β sandwich, explaining how the cleavage products remain associated. Nearly all MAIN domains tested, including those from *Streptococcus agalactiae*, *Acetovibrio thermocellus*, *Oceanobacillus polygoni*, and *Streptococcus mitis* purified from *E. coli* as associated cleavage products, suggesting that they are also autoproteolytic (Figure 2G). Substitution of the putative catalytic serine to alanine in the *S. agalactiae* MAIN domain abolished autocleavage (Figure 2H). We conclude that bacterial MAIN domains undergo autoproteolysis identically to GAIN domains.

MAIN domains are cell surface anchored, surface exposed, and capable of release

To investigate whether YfkN's MAIN domain undergoes autoproteolysis *in vivo*, we inserted a hexa-histidine tag in the unstructured region that connects its two enzymatic domains (Figure 3A) and placed the sandwich fusion under the control of a strong IPTG-regulated promoter. To ensure that there were sufficient levels of sortase to attach His-YfkN to the cell wall, we fused the *yhcS* gene, encoding sortase, to a xylose-regulated promoter. Both expression constructs were inserted at non-essential loci in the *B. subtilis* genome. *B. subtilis* cells expressing both proteins were collected during growth in exponential and stationary phases, and lysates were prepared using lysozyme followed by the addition of SDS-sample buffer. Immunoblot analysis revealed that His-YfkN was more abundant in stationary phase than in exponential growth, likely because the protein was shed along with the peptidoglycan (PG) during growth (Figure S6). Accordingly, all experiments were performed 4 h after entry into stationary phase. The immunoblot in Figure 3B shows His-YfkN and the catalytic mutant His-YfkN^{S1346A}. In support of the model that the MAIN domain undergoes autoproteolysis *in vivo*, the catalytic mutant migrated more slowly than the wild-type protein (Figure 3B). A similar size shift was observed using ALFA-tagged YfkN (Figure S7). It is noteworthy that His-YfkN^{S1346A} migrated as a doublet, while His-YfkN was clearly a single band (Figure 3B). We wondered if the doublet corresponded to membrane-anchored and PG-anchored His-YfkN^{S1346A}. Prior to attachment to the cell wall, sortase substrates are anchored in the membrane by a C-terminal trans-membrane segment.²¹ Sortase cleaves off this TM segment and attaches the client protein to the stem peptide of the PG precursor lipid II. Polymerization of this modified precursor incorporates the substrate into the cell wall (Figure S8).²² Thus, sortase substrates often migrate as doublets: membrane-anchored and wall-anchored.²³ If His-YfkN undergoes constitutive autoproteolysis *in vivo*, we would not expect to see a doublet of the wild-type protein, since the C-terminal region where the modification occurs would not be covalently associated with the His-tagged YfkN cleavage product. To investigate this hypothesis, we treated cells with lysozyme in hypertonic medium and collected the soluble cell wall fraction and the protoplast pellet. The smaller His-YfkN^{S1346A} band was present in the cell wall fraction, while the larger band was present in the membrane fraction (Figure 3C). By contrast, the wild-type His-YfkN band was present in both the wall and membrane fractions, as hypothesized (Figure 3C). These results indicate that YfkN undergoes autoproteolysis *in vivo* prior to sortase-mediated attachment to the cell wall and that the cleavage products remain non-covalently associated and anchored to the PG meshwork.

We hypothesize that the enzymatic domains of YfkN are surface exposed but can be released in response to specific stimuli. To investigate whether YfkN can be released into the extracellular environment, we treated *B. subtilis* cells expressing His-YfkN or His-YfkN^{S1346A} with SDS and monitored the released proteins by immunoblot. As anticipated, the autoproteolyzed wild-type His-YfkN was released by SDS, while the membrane- and PG-anchored His-YfkN^{S1346A} proteins were not (Figure 3D). In addition, we found that exposing cells to 95°C for 5 min similarly resulted in the release of His-YfkN's ectodomain, while cells

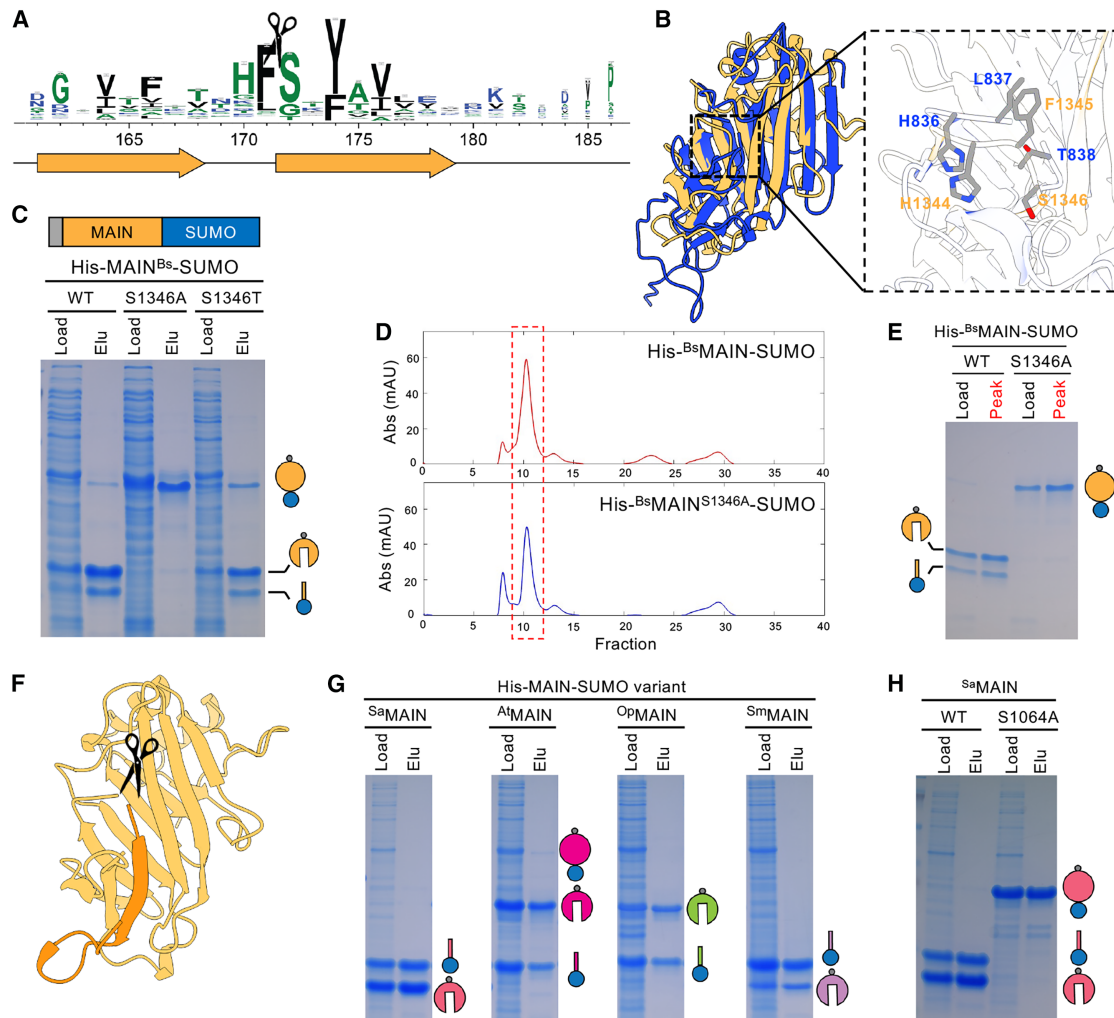


Figure 2. MAIN domains are autoproteolytic and remain non-covalently associated after cleavage

(A) Sequence logo determined from MAIN homologs in Bacillota. The autocleavage site is highlighted by a scissor. The approximate locations of beta strands (arrows) and loops (lines) are depicted below.

(B) Structural alignment (Tm-align) of latrophilin's GAIN domain (PDB: 4dlq) and the AlphaFold2 prediction of YfkN's GAIN domain. The inset shows the putative catalytic residues in the MAIN domain (orange) aligned with the conserved catalytic HX(S/T) residues in the GAIN domain (blue).

(C) Top: schematic of the His-MAIN-SUMO construct used for expression and purification. Bottom: representative Coomassie-stained SDS-polyacrylamide gel of His-MAIN-SUMO wild-type and the indicated variants expressed and purified from *E. coli*. The load and elution (Elu) from the Ni²⁺-NTA resin are shown. Schematics depict full-length and cleavage products.

(D) Size exclusion chromatograms depicting the A280 traces for His-BsMAIN-SUMO and His-BsMAIN^{S1346A}-SUMO.

(E) Coomassie-stained gel of the input (load) and peak fractions (peak) (boxed in red) from (D).

(F) AlphaFold2-predicted structure of the autocleaved fragments of BsMAIN. The C-terminal fragment (dark orange), the N-terminal fragment (light orange), and the cleavage site (scissors) are shown.

(G) Coomassie-stained gels of His-MAIN-SUMO variants from *Streptococcus agalactiae* (Sa), *Acetovibrio thermocellus* (At), *Oceanobacillus polygami* (Op), *Streptococcus mitis* (Sm), and *Pseudobacteroides cellulosolvens* (Pc). Load and elution are shown.

(H) Coomassie-stained gel of His-SaMAIN-SUMO wild-type and catalytic mutant.

See also Figures S2–S5.

expressing the autoproteolysis-deficient mutant did not (Figure S9). Together, these results provide further support for the MAIN domain autoproteolysis *in vivo* and indicate that YfkN's ectodomain can be released into the environment in these cases by non-physiological stimuli. To test whether YfkN is surface exposed, we inserted an ALFA tag between the signal sequence and the first enzymatic domain of YfkN and expressed the fusion protein under IPTG control. We monitored surface-associated

ALFA-YfkN in late stationary phase using a fluorophore-conjugated anti-ALFA nanobody that cannot cross the cell wall (Figure S10). As anticipated, cells expressing ALFA-YfkN and ALFA-YfkN^{S1346A} were readily labeled on their cell surface (Figure 3E).

PG is turned over as cells grow and divide. Accordingly, YfkN would be constantly released into the medium with the PG cleavage products if expressed throughout exponential growth,

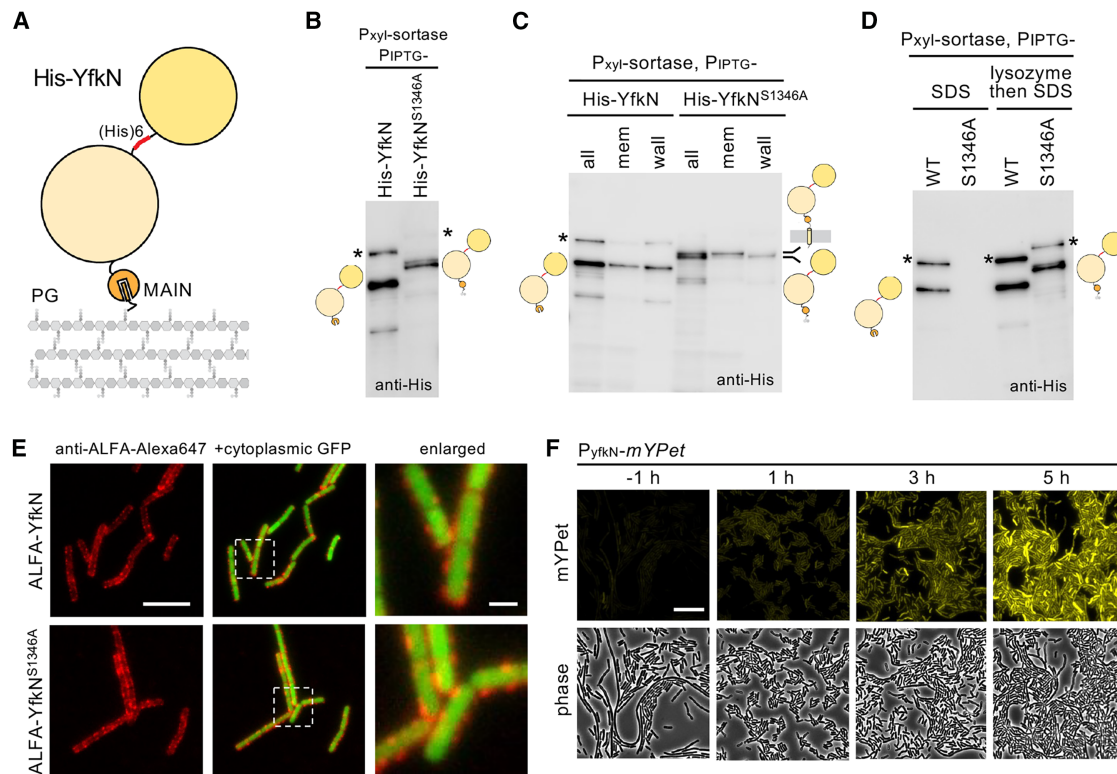


Figure 3. YfkN is anchored to the cell surface and undergoes autoproteolysis *in vivo*

(A) Schematic of His-YfkN. YfkN is anchored to the peptidoglycan (PG) at its C terminus by sortase, followed by its MAIN domain (orange), 5' nucleotidase (light yellow), and phosphodiesterase (dark yellow) domains. The (His)₆ tag was inserted between the two enzymatic domains.

(B) Immunoblot of stationary-phase *B. subtilis* cells constitutively expressing His-YfkN or the catalytic mutant S1346A. Schematics of His-YfkN variants are depicted adjacent to relevant bands on the blot. Anomalous migrating species are indicated (*).

(C) Immunoblot of His-YfkN fractionation. Cells were protoplasted, and membrane and wall fractions were analyzed.

(D) Immunoblot of SDS-released His-YfkN variants.

(E) Representative fluorescence micrographs of the indicated *B. subtilis* stationary-phase cells expressing cytoplasmic green fluorescent protein (GFP) probed with anti-ALFA-Alexa647. The boxed regions in the merged image were enlarged to more clearly visualize the surface-localized ALFA-YfkN. Far-left scale bar indicates 5 μ m, and the far-right scale bar indicates 1.3 μ m. *B. subtilis* YfkN was expressed with 500 μ M IPTG, and sortase was expressed with 20 mM xylose in all backgrounds (B–E).

(F) Fluorescence and phase-contrast micrographs of cells harboring a *PyfkN-mYPet* reporter. Time in hours (h) before and after entrance into the stationary phase is indicated. Scale bar indicates 10 μ m. The growth curve associated with this experiment is shown in Figure S11.

See also Figures S6–S11.

obviating the need for the MAIN domain. To determine when *yfkN* is transcribed, we fused the *yfkN* promoter to *mYPet* and inserted this transcriptional reporter at a non-essential locus in the *B. subtilis* genome. We assayed *yfkN* promoter activity during growth in LB and low-phosphate defined medium (LPDM), a medium that has previously been shown to increase production of YfkN.^{17,24} *mYPet* fluorescence increased in late stationary phase, consistent with the idea that YfkN is only expressed and anchored to the cell wall at a time when the cells have largely stopped growing (Figure 3G). As reported previously, we observed stronger induction in LPDM than LB (Figure S11).¹⁷ In sum, our data indicate that YfkN is produced when cells stop growing. The secreted protein undergoes autoproteolysis, but the cleavage products remain associated, and the protein is then attached to the cell wall and becomes surface exposed. We propose that the enzymatic domains can act on substrates adjacent to the surface and, if released by shear force, can access substrates further away from the cell.

MAIN domains are found in diverse protein architectures

To survey the breadth of domains associated with a MAIN domain, we compiled 15,000 MAIN-containing proteins across bacteria and archaea and annotated known Pfam domains within their protein sequences using PfamScan (Data S1).²⁵ In a complementary approach, we clustered the 15,000 homologs and used AlphaFold2 to predict the structure of 2,843 unique MAIN domain-containing proteins. We subsequently used domain parser for AlphaFold models (DPAMs)²⁶ to annotate known domains with structural similarity (Figure S12). Highlighting the incredible diversity of proteins that contain a MAIN domain, we identified 350 distinct Pfam domains from the sequences and 3,874 distinct ECODs (evolutionary classification of protein domains) domains from the predicted structures (Data S2). Several observations emerged from this analysis. First, virtually all the proteins contain a signal peptide as predicted by SignalP,²⁷ indicating that the proteins are secreted

across the cytoplasmic membrane. Second, the vast majority of MAIN-domain-containing proteins contain a sortase recognition sequence, cell-wall-anchoring domains, or an S-layer homology domain, which binds glycopolymers that are covalently attached to the cell wall.²⁸ The presence of these domains indicates that MAIN domain-containing proteins are typically anchored to the cell surface. Third, most MAIN domain-containing proteins are large by bacterial standards, and the average protein molecular weight among the 15,000 homologs was 158 kDa. Finally, the other domains associated with MAIN domains could be grouped into two categories: those that have predicted roles in adhesion and those that have enzymatic function. MAIN-domain-containing proteins often possess both types of domains, often with multiple copies of each (Figure 4A). Several of the adhesion domains that are associated with MAIN domains are also found on human aGPCRs, including Ig-like, cadherin-like, fibronectin-like, leucine-rich repeats, and lectins (Figures 4A and S13). Interestingly, a subset of MAIN-domain-containing proteins are fused to putative toxins in their ectodomains (Figure 4A).

A Gene Ontology (GO) enrichment analysis of the Pfam domains associated with MAIN domains revealed several enzymatic functions, the most prominent of which was hydrolase activity, particularly on glycosidic bonds (Figure S14). Furthermore, many of these proteins contain both glycoside hydrolase domains and carbohydrate-binding lectins (Figure 4A), a common arrangement for carbohydrate-degrading enzymes.²⁹ Notably, many of these MAIN-domain-containing proteins are encoded by bacterial species in the human and soil microbiomes (Data S1B and S2B). Finally, virtually all the proteins analyzed contained a similar domain arrangement (from N to C terminus): signal peptide, adhesion and enzymatic domains, MAIN domain, and cell-surface anchoring domain. These findings support the proposed role of MAIN domains in releasing cell-surface-associated enzymatic, adhesion, and toxin domains in response to shear to enable broader nutrient scavenging, dispersal, or intoxication of neighboring cells (Figure 4B).

Diverse MAIN domains are cell-surface-anchored and autoproteolytic

To extend our analysis to the diverse set of proteins containing autoproteolytic MAIN domains, we engineered cells to express and anchor heterologous substrates onto the *B. subtilis* cell wall. To do so, we introduced the N-terminal YfkN signal sequence and C-terminal cell-wall-anchoring sequence in place of those present on candidate proteins (Figure S15A). In addition, an ALFA tag was inserted between the signal sequence and the candidate proteins. These chimeras were expressed under the control of a strong IPTG-regulated promoter alongside the xylose-regulated sortase, YhcS. We analyzed a MAIN-containing protein from *Pseudomonas massiliensis* (WP_106763535) that contains a leucine-rich repeat and a putative starch-binding domain CBM26 and one from *Gracilbacillus dipsosauri* (WP_109985763) containing a cadherin-like domain, a glycosyl hydrolase 26 domain with predicted activity on xylans and mannans, and a putative mannan-binding CBM35 domain. Both MAIN domains were autoproteolytic, as assayed by purification of His-MAIN-SUMO wild-type and autoproteolysis-deficient mutants from *E. coli* (Figure S15B). In addition, immunoblot analysis of *B. subtilis* lysates revealed size shifts consistent with *in vivo* autoproteolysis for both

chimeras (Figure S15C). Finally, immunolabeling of *B. subtilis* cells expressing the sortase-anchored chimeras indicated that they are cell-surface exposed (Figure S15D). Of note, unlike native YfkN, both chimeras had to be expressed in a strain lacking extracellular proteases for efficient expression and cell-surface exposure, likely due to non-specific proteolysis of the non-native proteins. We conclude that diverse MAIN-domain-containing proteins are autoproteolytic *in vivo*.

DISCUSSION

Altogether, our findings define a new bacterial autoproteolytic domain, expand the phylogenetic diversity of GAIN-like domains, and provide a model for a new form of mechanosensing in bacteria. We show that diverse microbial proteins with MAIN domains are autoproteolytic, anchored to the cell surface, and capable of release. GAIN domains function in aGPCR signaling as mechanotransductive domains that are pulled apart in response to mechanical stimuli, allowing ectodomain release and GPCR signaling. By analogy, we propose that MAIN domain-containing proteins sense mechanical stimuli when bound to complex polysaccharides, neighboring cells, or surfaces and release their ectodomains in response to shear forces (Figure 4B).

Many gut and soil bacteria encode MAIN domains that tether large ectodomains consisting of repeated adhesion domains, lectins, and glycosyl hydrolases with predicted specificity for complex polysaccharides. We hypothesize that the lectin domains bind polysaccharides derived from plant cell walls in the gut lumen or the soil, and shear force results in the dissociation of the MAIN domain cleavage products. Release of the enzymatic domains enables degradation of polysaccharides and nutrient scavenging at a distance (Figure 4B). The force transduced upon polysaccharide binding is reminiscent of sigma/anti-sigma signaling by SigI/RsgI systems in Clostridial species. These RsgI anti-sigma factors contain extracytoplasmic SEAL-like domains called SEAL domains that undergo autoproteolysis, but the cleavage products remain stably associated. A carbohydrate-binding domain at the extreme C terminus of RsgI binds cellulose or other polysaccharides, and a resulting shear force is hypothesized to pull the SEAL domain apart.^{11,12} This likely triggers intramembrane proteolysis, sigma factor activation, and expression of the proteins that make up a multi-protein complex responsible for cellulose degradation, called the cellulosome.^{30,31}

A subset of MAIN domains are fused to toxin domains or domains at the bacteria-host interface. For example, the wall-anchored protein immunoglobulin A (IgA) FC receptor in Group B Streptococcus (*S. agalactiae*) contains a MAIN domain. The IgA FC receptor binds the Fc region of human IgA, enabling this bacterium to evade immune detection.^{32,33} Similar to YfkN, the MAIN domain is present at the extreme C terminus of the IgA FC receptor, adjacent to the sortase anchoring domain. Our findings raise the possibility that the MAIN domain enables release of the receptor from the *S. agalactiae* surface when it is bound by IgA. Some MAIN domains are also fused to RHS repeat-associated toxin domains. Interestingly, RHS toxins are often found on contact-dependent inhibition (CDI) systems.³⁴ In Gram-negative bacteria, some of these CDI toxins are

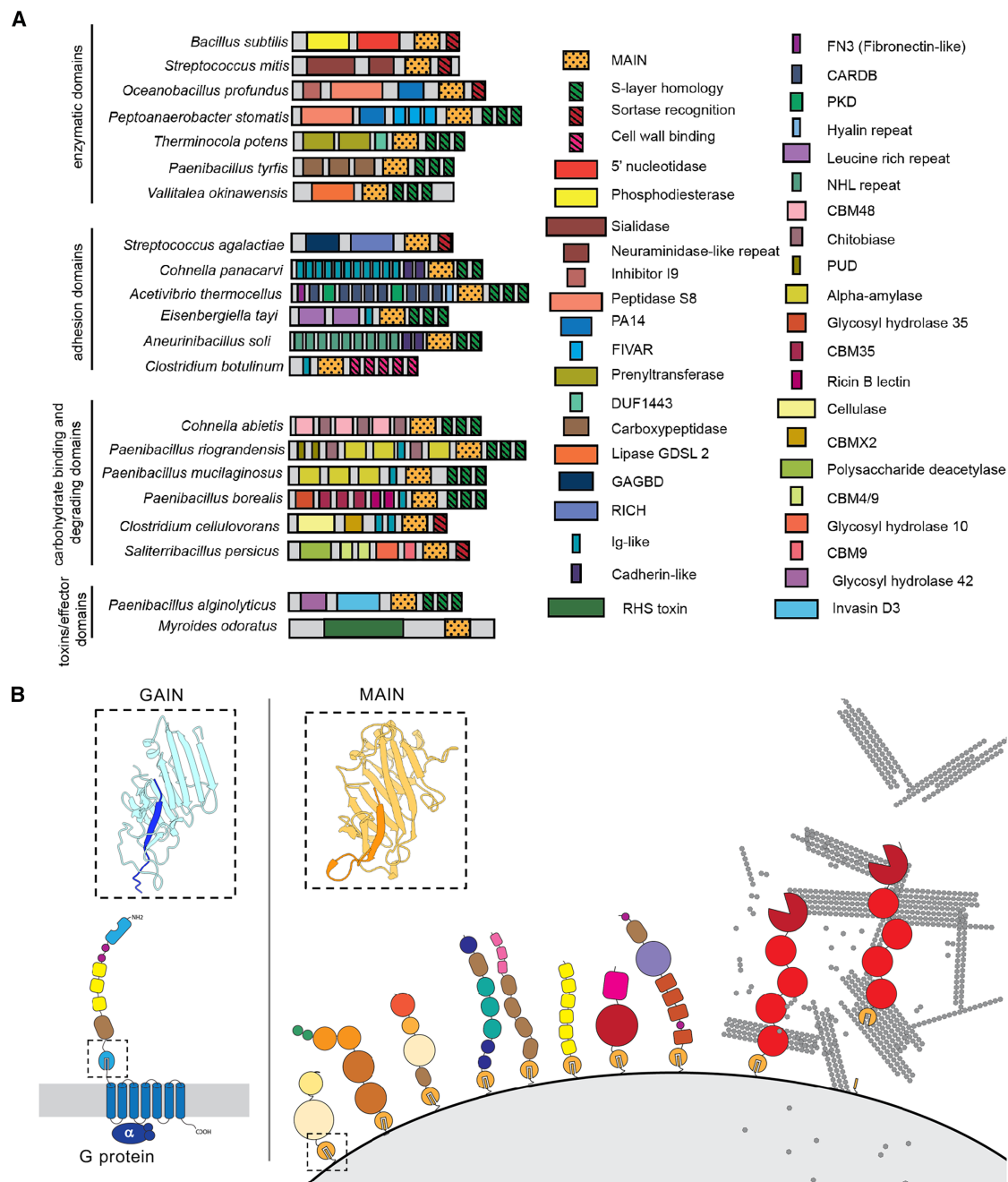


Figure 4. Schematic of MAIN and GAIN domain-containing proteins on the cell surface

(A) Schematic of representative MAIN domain-containing proteins from diverse bacteria.

(B) Schematic of a prototypical aGPCR and inset of the crystal structure of latrophilin/CIRL's GAIN_B domain with N-terminal (light blue) and C-terminal (dark blue) cleavage products. The schematic on the right depicts examples of MAIN domain-containing proteins. These proteins are anchored onto microbial cell surfaces and display large and diverse enzymatic and adhesion domains. The example on the far-right depicts a cellulose-binding and degrading protein proposed to be released by interaction with cellulose.

See also [Data S1](#) and [S2](#) and [Figures S12–S17](#).

anchored in the outer membrane by β barrels, and toxin release is mediated by physical contact.³⁵ Our findings suggest that the MAIN domain could mediate toxin delivery in response to cell-cell contact in a manner analogous to these Gram-negative CDI systems.

This proposed mode of toxin delivery shares striking similarities to type V secretion via autotransporters in Gram-negative bacteria.³⁶ Autotransporters assemble a C-terminal β -barrel translocator domain in the outer membrane. It is thought that after folding, the N-terminal domain, known as the passenger or α

domain, passes through the translocator and is released by proteolytic cleavage. This cleavage can be autoprotoleolytic or catalyzed by a membrane-anchored protease.^{37–39} In some cases, after cleavage, the passenger and β domains remain non-covalently associated.⁴⁰ Moreover, the passenger domains in these systems can be highly variable and have roles in adhesion and virulence.^{41,42} These similarities suggest that MAIN domains could be the Gram-positive counterpart to Gram-negative autotransporters.

Many of the same adhesion domains that are present on aGPCRs are also fused to MAIN domains in bacteria. In addition, aGPCR and MAIN-containing proteins have the same protein architecture: large N-terminal ectodomains, a MAIN or GAIN domain, and cell-surface anchoring domains. A key difference between the two is that release of the ectodomain is the regulated step of aGPCR signaling and alters intracellular physiology, whereas no signaling mechanisms have been associated with the C-terminal fragment of the MAIN domain that remains anchored to the cell wall. Whether these surface-anchored “tethered peptides” impact cellular physiology remains to be determined. However, it is noteworthy that aGPCRs are thought to have two modes of signaling: *cis* and *trans*. In the *cis* mode, the tethered agonist binds the 7TM domain to which it is fused, resulting in G-protein activation. In the *trans* mode, signaling occurs via interaction of the released ectodomain with an extracellular partner.⁴ This latter mode of signaling is analogous to the proposed release of the cleaved MAIN and associated domains from the microbial cell surface in response to force. We speculate that this represents the ancestral aGPCR signaling mode.

The finding of a conserved protein fold across all domains of life suggests an ancient function for the MAIN and GAIN domains. Both domains, as well as the mechanosensitive SEA and SEAL domains, undergo autoprotoleolysis in a loop between two β strands within a β sheet. This fold allows non-covalent association of the cleavage products via an extensive hydrogen bonding network. In response to mechanical stimuli, the cleavage products dissociate, allowing signaling in *cis* or *trans*.^{2,11–13,43,44} Like domains fused to MAIN and GAIN domains, SEAL and SEA domains are frequently linked to adhesion domains, often of the same fold across domains, suggesting the possibility of a similar signaling mechanism (Figures S16 and S17).¹¹ The convergent evolution of these force-sensing domains reveals a simple and robust solution to sensing and responding to mechanical stimuli.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, David Rudner (rudner@hms.harvard.edu).

Materials availability

All plasmids and strains generated in this study are available upon request.

Data and code availability

- Uncropped immunoblots and SDS-PAGE gels have been deposited at Mendeley Data at DOI: [10.17632/s92pz776ft.1](https://doi.org/10.17632/s92pz776ft.1) and are publicly available as of the date of publication.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

ACKNOWLEDGMENTS

We thank Shailab Shrestha and all members of the Bernhardt-Rudner supergroup, past and present, for helpful advice, feedback, discussions, and encouragement; the Tufts University Core Facility (TUCF) for Edman degradation analysis; and the Microscopy Resources on the North Quad (MicRoN) core at Harvard Medical School for help with microscopy. Portions of this research were conducted on the O2 High Performance Computing Cluster, which is supported by the Research Computing Group at Harvard Medical School. Support for this work comes from National Institutes of Health grants R35GM145299 and U19 AI158028 (D.Z.R.). A.P.B. was funded in part by NSF (DGE1745303) and NIH (F31AI181098).

AUTHOR CONTRIBUTIONS

A.P.B. and D.Z.R. conceived the study. A.P.B. performed the experiments. A.P.B. and D.Z.R. performed the analyses. D.Z.R. supervised the study. A.P.B. and D.Z.R. wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used ChatGPT in order to suggest new words and alternative sentence structures. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
 - Immunoblot analysis
 - Purification of His-MAIN-SUMO variants
 - Size Exclusion Chromatography
 - Edman Degradation
 - Bioinformatics Analysis
 - Multiple Sequence Alignment
 - AlphaFold2 Predictions
 - Structural model visualization
 - Fluorescence Microscopy
 - P_{yfkN} -mYPet expression
 - Anti-ALFA Immunolabeling
 - Plasmid Construction
 - pAB221 [His-yfkN(MAIN)-SUMO (amp)]
 - pAB222 [His-yfkN(MAIN)-S1346A-SUMO (amp)]
 - pAB225 [yvbJ-PxyIA-yhcS (kan) (amp)]
 - pAB227 [yhdG::Phyerspank-yfkN-His6 (spc) (amp)]
 - pAB228 [yhdG::Phyerspank-yfkN-S1346A-His6 (spc) (amp)]
 - pAB230 [His-yfkN(MAIN)-S1346T-SUMO (amp)]
 - pAB233 [His-SpneumoniaeR6_MAIN-SUMO (amp)]
 - pAB234 [His-Sagalactiae_MAIN-SUMO (amp)]
 - pAB235 [His-Athermococcus_MAIN-SUMO (amp)]
 - pAB238 [Opolygoni_MAIN-SUMO (amp)]
 - pAB239 [Smitis_MAIN-SUMO (amp)]
 - pAB240 [His-Sagalactiae_MAIN-S115A-SUMO (amp)]
 - pAB247 [yhdG::Phyerspank-ALFA-yfkN (spc) (amp)]
 - pAB248 [yhdG::Phyerspank-ALFA-yfkN-S1346A (spc) (amp)]
 - pAB259 [sacA::yfkN-mYPet (phleo) (amp)]
 - pAB261 [pminiMAD yfkN-S1346A]

- pAB263 [yhdG::Phy-spank-SS(yfkN)-His6-NotI-CWA(yfkN) (spc) (amp)]
- pAB264 [yhdG::Phy-spank-SS(yfkN)-PmCBM26-CWA(yfkN) (spc) (amp)]
- pAB265 [yhdG::Phy-spank-SS(yfkN)-PmCBM26-CWA(yfkN)-S623A (spc) (amp)]
- pAB269 [His-PmCBM26(MAIN)-SUMO (amp)]
- pAB270 [His-PmCBM26(MAIN)-S623A-SUMO (amp)]
- pAB285 [yhdG::Phy-spank-SS(yfkN)-ALFA-PmCBM26-CWA(yfkN) (spc) (amp)]
- pAB286 [yhdG::Phy-spank-SS(yfkN)-ALFA-PmCBM26-CWA(yfkN)-S623A (spc) (amp)]
- pAB288 [yhdG::Phy-spank-SS(yfkN)-ALFA-NotI-CWA(yfkN) (spc) (amp)]
- pAB289 [yhdG::Phy-spank-SS(yfkN)-ALFA-CBM35-CWA(yfkN) (spc) (amp)]
- pAB290 [yhdG::Phy-spank-SS(yfkN)-ALFA-CBM35-CWA(yfkN)-T770A (spc) (amp)]
- pAB295 [His-CBM35(MAIN)-SUMO (amp)]
- pAB296 [His-CBM35(MAIN)-T770A-SUMO (amp)]

● **QUANTIFICATION AND STATISTICAL ANALYSIS**

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cub.2026.07.052>.

Received: May 12, 2026

Revised: July 16, 2026

Accepted: July 22, 2026

REFERENCES

1. Rosenbaum, D.M., Rasmussen, S.G.F., and Kobilka, B.K. (2009). The structure and function of G-protein-coupled receptors. *Nature* 459, 356–363. <https://doi.org/10.1038/nature08144>.
2. Vizurraga, A., Adhikari, R., Yeung, J., Yu, M., and Tall, G.G. (2020). Mechanisms of adhesion G protein-coupled receptor activation. *J. Biol. Chem.* 295, 14065–14083. <https://doi.org/10.1074/jbc.rev120.007423>.
3. Langenhan, T. (2020). Adhesion G protein-coupled receptors—Candidate metabotropic mechanosensors and novel drug targets. *Basic Clin. Pharmacol. Toxicol.* 126, 5–16. <https://doi.org/10.1111/bcpt.13223>.
4. Hamann, J., Aust, G., Araç, D., Engel, F.B., Formstone, C., Fredriksson, R., Hall, R.A., Harty, B.L., Kirchhoff, C., Knapp, B., et al. (2015). International Union of Basic and Clinical Pharmacology. XCIV. Adhesion G protein-coupled receptors. *Pharmacol. Rev.* 67, 338–367. <https://doi.org/10.1124/pr.114.009647>.
5. Araç, D., Boucard, A.A., Bolliger, M.F., Nguyen, J., Soltis, S.M., Südhof, T.C., and Brunker, A.T. (2012). A novel evolutionarily conserved domain of cell-adhesion GPCRs mediates autoproteolysis. *EMBO J.* 31, 1364–1378. <https://doi.org/10.1038/emboj.2012.26>.
6. Langenhan, T., Aust, G., and Hamann, J. (2013). Sticky signaling—adhesion class GProtein-coupled receptors take the stage. *Sci. Signal.* 6. <https://doi.org/10.1126/scisignal.2003825>.
7. Lala, T., and Hall, R.A. (2022). Adhesion G protein-coupled receptors: structure, signaling, physiology, and pathophysiology. *Physiol. Rev.* 102, 1587–1624. <https://doi.org/10.1152/physrev.00027.2021>.
8. Liebscher, I., Schön, J., Petersen, S.C., Fischer, L., Auerbach, N., Demberg, L.M., Mogha, A., Cöster, M., Simon, K.-U., Rothmund, S., et al. (2014). A tethered agonist within the ectodomain activates the adhesion G protein-coupled receptors GPR126 and GPR133. *Cell Rep.* 9, 2018–2026. <https://doi.org/10.1016/j.celrep.2014.11.036>.
9. Stoveken, H.M., Hajduczuk, A.G., Xu, L., and Tall, G.G. (2015). Adhesion G protein-coupled receptors are activated by exposure of a cryptic tethered agonist. *Proc. Natl. Acad. Sci. USA* 112, 6194–6199. <https://doi.org/10.1073/pnas.1421785112>.
10. Liao, Y., Pei, J., Cheng, H., and Grishin, N.V. (2014). An ancient autoproteolytic domain found in GAIN, ZU5 and nucleoporin98. *J. Mol. Biol.* 426, 3935–3945. <https://doi.org/10.1016/j.jmb.2014.10.011>.
11. Brogan, A.P., Habib, C., Hobbs, S.J., Kranzusch, P.J., and Rudner, D.Z. (2023). Bacterial SEAL domains undergo autoproteolysis and function in regulated intramembrane proteolysis. *Proc. Natl. Acad. Sci. USA* 120, e2310862120. <https://doi.org/10.1073/pnas.2310862120>.
12. Chen, C., Dong, S., Yu, Z., Qiao, Y., Li, J., Ding, X., Li, R., Lin, J., Bayer, E.A., Liu, Y.-J., et al. (2023). Essential autoproteolysis of bacterial anti- σ factor RsgI for transmembrane signal transduction. *Sci. Adv.* 9, eadg4846. <https://doi.org/10.1126/sciadv.adg4846>.
13. Takayasu, A., Mahoney, B.J., Goring, A.K., Jessup, T., Ogorzalek Loo, R.R.O., Loo, J.A., and Clubb, R.T. (2024). Insight into the autoproteolysis mechanism of the RsgI9 anti- σ factor from *Clostridium thermocellum*. *Proteins* 92, 946–958. <https://doi.org/10.1002/prot.26690>.
14. van Kempen, M. van, Kim, S.S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C.L.M., Söding, J., and Steinegger, M. (2023). Fast and accurate protein structure search with Foldseek. *Nat. Biotechnol.* 42, 243–246. <https://doi.org/10.1038/s41587-023-01773-0>.
15. Zhang, Y., and Skolnick, J. (2005). TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* 33, 2302–2309. <https://doi.org/10.1093/nar/gki524>.
16. Fasehee, H., Westers, H., Bolhuis, A., Antelmann, H., Hecker, M., Quax, W.J., Mirolohi, A.F., van Dijk, J.M. van, and Ahmadian, G. (2011). Functional analysis of the sortase YhcS in *Bacillus subtilis*. *Proteomics* 11, 3905–3913. <https://doi.org/10.1002/pmic.201100174>.
17. Chambert, R., Pereira, Y., and Petit-Glatron, M.-F. (2003). Purification and characterization of YfkN, a trifunctional nucleotide phosphoesterase secreted by *Bacillus subtilis*. *J. Biochem.* 134, 655–660. <https://doi.org/10.1093/jb/mvg189>.
18. Andrade, W.A., Firon, A., Schmidt, T., Hornung, V., Fitzgerald, K.A., Kurt-Jones, E.A., Trieu-Cuot, P., Golenbock, D.T., and Kaminski, P.-A. (2016). Group B streptococcus degrades cyclic-di-AMP to modulate STING-dependent type I interferon production. *Cell Host Microbe* 20, 49–59. <https://doi.org/10.1016/j.chom.2016.06.003>.
19. Nieberler, M., Kittel, R.J., Petrenko, A.G., Lin, H.-H., and Langenhan, T. (2016). Control of adhesion GPCR function through proteolytic processing. *Handb. Exp. Pharmacol.* 234, 83–109. https://doi.org/10.1007/978-3-319-41523-9_5.
20. Lin, H.-H., Chang, G.-W., Davies, J.Q., Stacey, M., Harris, J., and Gordon, S. (2004). Autocatalytic cleavage of the EMR2 receptor occurs at a conserved G protein-coupled receptor proteolytic site motif. *J. Biol. Chem.* 279, 31823–31832. <https://doi.org/10.1074/jbc.m402974200>.
21. Novick, R.P. (2000). Sortase: the surface protein anchoring transpeptidase and the LPXTG motif. *Trends Microbiol.* 8, 148–151. [https://doi.org/10.1016/S0966-842X\(00\)01741-8](https://doi.org/10.1016/S0966-842X(00)01741-8).
22. Schneewind, O., Fowler, A., and Faull, K.F. (1995). Structure of the cell wall anchor of surface proteins in *Staphylococcus aureus*. *Science* 268, 103–106. <https://doi.org/10.1126/science.7701329>.
23. Navarre, W.W., and Schneewind, O. (1994). Proteolytic cleavage and cell wall anchoring at the LPXTG motif of surface proteins in Gram-positive bacteria. *Mol. Microbiol.* 14, 115–121. <https://doi.org/10.1111/j.1365-2958.1994.tb01271.x>.
24. Müller, J.P., An, Z., Merad, T., Hancock, I.C., and Harwood, C.R. (1997). Influence of *Bacillus subtilis* phoR on cell wall anionic polymers. *Microbiology (Reading)* 143, 947–956. <https://doi.org/10.1099/00221287-143-3-947>.
25. Finn, R.D., Bateman, A., Clements, J., Coghill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Misty, J., et al. (2014). Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D230. <https://doi.org/10.1093/nar/gkt1223>.

26. Zhang, J., Schaeffer, R.D., Durham, J., Cong, Q., and Grishin, N.V. (2023). DPAM: A domain parser for AlphaFold models. *Protein Sci.* 32, e4548. <https://doi.org/10.1002/pro.4548>.
27. Teufel, F., Almagro Armenteros, J.J.A., Johansen, A.R., Gislason, M.H., Pihl, S.I., Tsigiris, K.D., Winther, O., Brunak, S., von Heijne, G. von, and Nielsen, H. (2022). SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat. Biotechnol.* 40, 1023–1025. <https://doi.org/10.1038/s41587-021-01156-3>.
28. Sára, M. (2001). Conserved anchoring mechanisms between crystalline cell surface S-layer proteins and secondary cell wall polymers in Gram-positive bacteria? *Trends Microbiol.* 9, 47–49. [https://doi.org/10.1016/S0966-842X\(00\)01905-3](https://doi.org/10.1016/S0966-842X(00)01905-3).
29. Conway, J.M., Crosby, J.R., Hren, A.P., Southerland, R.T., Lee, L.L., Lunin, V.V., Alahuhta, P., Himmel, M.E., Bomble, Y.J., Adams, M.W.W., et al. (2018). Novel multidomain, multifunctional glycoside hydrolases from highly lignocellulolytic *Caldicellulosiruptor* species. *AIChE J.* 64, 4218–4228. <https://doi.org/10.1002/aic.16354>.
30. Kahel-Raifer, H., Jindou, S., Bahari, L., Nataf, Y., Shoham, Y., Bayer, E.A., Borovok, I., and Lamed, R. (2010). The unique set of putative membrane-associated anti- σ factors in *Clostridium thermocellum* suggests a novel extracellular carbohydrate-sensing mechanism involved in gene regulation. *FEMS Microbiol. Lett.* 308, 84–93. <https://doi.org/10.1111/j.1574-6968.2010.01997.x>.
31. Nataf, Y., Bahari, L., Kahel-Raifer, H., Borovok, I., Lamed, R., Bayer, E.A., Sonenshein, A.L., and Shoham, Y. (2010). *Clostridium thermocellum* cellulosomal genes are regulated by extracytoplasmic polysaccharides via alternative sigma factors. *Proc. Natl. Acad. Sci. USA* 107, 18646–18651. <https://doi.org/10.1073/pnas.1012175107>.
32. Michel, J.L., Madoff, L.C., Kling, D.E., Kasper, D.L., and Ausubel, F.M. (1991). Cloned alpha and beta C-protein antigens of group B streptococci elicit protective immunity. *Infect. Immun.* 59, 2023–2028. <https://doi.org/10.1128/iai.59.6.2023-2028.1991>.
33. Jerlström, P.G., Chhatwal, G.S., and Timmis, K.N. (1991). The IgaA-binding B antigen of the c protein complex of Group B streptococci: sequence determination of its gene and detection of two binding regions. *Mol. Microbiol.* 5, 843–849. <https://doi.org/10.1111/j.1365-2958.1991.tb00757.x>.
34. Koskiniemi, S., Lamoureux, J.G., Nikolakakis, K.C., t'Kint de Roodenbeke, C., Kaplan, M.D., Low, D.A., and Hayes, C.S. (2013). Rhs proteins from diverse bacteria mediate intercellular competition. *Proc. Natl. Acad. Sci. USA* 110, 7032–7037. <https://doi.org/10.1073/pnas.1300627110>.
35. Hayes, C.S., Koskiniemi, S., Ruhe, Z.C., Poole, S.J., and Low, D.A. (2014). Mechanisms and biological roles of contact-dependent growth inhibition systems. *Cold Spring Harb. Perspect. Med.* 4, a010025. <https://doi.org/10.1101/cshperspect.a010025>.
36. Henderson, I.R., Navarro-Garcia, F., Desvaux, M., Fernandez, R.C., and Ala'Aldeen, D. (2004). Type V protein secretion pathway: the autotransporter story. *Microbiol. Mol. Biol. Rev.* 68, 692–744. <https://doi.org/10.1128/MMBR.68.4.692-744.2004>.
37. Dautin, N., Barnard, T.J., Anderson, D.E., and Bernstein, H.D. (2007). Cleavage of a bacterial autotransporter by an evolutionarily convergent autocatalytic mechanism. *EMBO J.* 26, 1942–1952. <https://doi.org/10.1038/sj.emboj.7601638>.
38. Egile, C., D'Hauteville, H., Parsot, C., and Sansonetti, P.J. (1997). SopA, the outer membrane protease responsible for polar localization of IcsA in *Shigella flexneri*. *Mol. Microbiol.* 23, 1063–1073. <https://doi.org/10.1046/j.1365-2958.1997.2871652.x>.
39. Shere, K.D., Sallustio, S., Manessis, A., D'Aversa, T.G., and Goldberg, M.B. (1997). Disruption of IcsP, the major *Shigella* protease that cleaves IcsA, accelerates actin-based motility. *Mol. Microbiol.* 25, 451–462. <https://doi.org/10.1046/j.1365-2958.1997.4681827.x>.
40. Oliver, D.C., Huang, G., and Fernandez, R.C. (2003). Identification of secretion determinants of the *Bordetella pertussis* BrkA autotransporter. *J. Bacteriol.* 185, 489–495. <https://doi.org/10.1128/JB.185.2.489-495.2003>.
41. Vo, J.L., Martínez Ortiz, G.C.M., Subedi, P., Keerthikumar, S., Mathivanan, S., Paxman, J.J., and Heras, B. (2017). Autotransporter adhesins in *Escherichia coli* pathogenesis. *Proteomics* 17. <https://doi.org/10.1002/pmic.201600431>.
42. Benz, I., and Schmidt, M.A. (2011). Structures and functions of autotransporter proteins in microbial pathogens. *Int. J. Méd. Microbiol.* 301, 461–468. <https://doi.org/10.1016/j.ijmm.2011.03.003>.
43. Brunet, Y.R., Habib, C., Brogan, A.P., Artzi, L., and Rudner, D.Z. (2022). Intrinsically disordered protein regions are required for cell wall homeostasis in *Bacillus subtilis*. *Genes Dev.* 36, 970–984. <https://doi.org/10.1101/gad.349895.122>.
44. Wreschner, D.H., McGuckin, M.A., Williams, S.J., Baruch, A., Yoeli, M., Ziv, R., Okun, L., Zaretsky, J., Smorodinsky, N., Keydar, I., et al. (2002). Generation of ligand-receptor alliances by “SEA” module-mediated cleavage of membrane-associated mucin proteins. *Protein Sci.* 11, 698–706. <https://doi.org/10.1110/ps.16502>.
45. Fujita, M., and Sadaie, Y. (1998). Promoter selectivity of the *Bacillus subtilis* RNA polymerase 6k and CTH holoenzyme. *J. Biochem.* 124, 89–97. <https://doi.org/10.1093/oxfordjournals.jbchem.a022102>.
46. Youngman, P.J., Perkins, J.B., and Losick, R. (1983). Genetic transposition and insertional mutagenesis in *Bacillus subtilis* with *Streptococcus faecalis* transposon Tn917. *Proc. Natl. Acad. Sci. USA* 80, 2305–2309. <https://doi.org/10.1073/pnas.80.8.2305>.
47. Lee, S.J., Kim, D.M., Bae, K.H., Byun, S.M., and Chung, J.H. (2000). Enhancement of secretion and extracellular stability of staphylokinase in *Bacillus subtilis* by *wprA* gene disruption. *Appl. Environ. Microbiol.* 66, 476–480. <https://doi.org/10.1128/aem.66.2.476-480.2000>.
48. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539. <https://doi.org/10.1038/msb.2011.75>.
49. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nat. Methods* 19, 679–682. <https://doi.org/10.1038/s41592-022-01488-1>.
50. Steinegger, M., and Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* 35, 1026–1028. <https://doi.org/10.1038/nbt.3988>.
51. Laskowski, R.A., Jabłońska, J., Pravda, L., Vařeková, R.S., and Thornton, J.M. (2018). PDBsum: Structural summaries of PDB entries. *Protein Sci.* 27, 129–134. <https://doi.org/10.1002/pro.3289>.
52. Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. <https://doi.org/10.1093/nar/gkab301>.
53. Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345. <https://doi.org/10.1038/nmeth.1318>.
54. Sayers, E.W., Beck, J., Bolton, E.E., Brister, J.R., Chan, J., Comeau, D.C., Connor, R., DiCuccio, M., Farrell, C.M., Feldgarden, M., et al. (2024). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 52, D33–D43. <https://doi.org/10.1093/nar/gkad1044>.
55. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. <https://doi.org/10.1038/s41586-021-03819-2>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti-His	GenScript	Cat #A00186-100; RRID: AB_914704
Rabbit polyclonal anti-SigA	Fujita et al. ⁴⁵	N/A
anti-ALFA-rabbitFc nanobody	Nanotag	Cat # N1583; RRID: AB_3075998
anti-ALFA-Alexa647 nanobody	Nanotag	Cat #N1502-AF647-L; RRID: AB_3075981
HRP-anti-rabbit IgG	BioRad	Cat #1706515; RRID: AB_11125142
HRP-anti-mouse IgG	BioRad	Cat #1706516; RRID: AB_2921252
Bacterial and virus strains		
<i>Escherichia coli</i> DH5a	NEB	Cat #C2987
<i>Escherichia coli</i> BL21(DE3) Δ tonA	NEB	Cat #C2527
<i>Bacillus subtilis</i> PY79	Youngman et al. ⁴⁶	N/A
<i>Bacillus subtilis</i> WB600N	Lee et al. ⁴⁷	N/A
<i>Bacillus subtilis</i> PY79 yfkN::tet yhdG::Phyperspank-yfkN-His6 (spc) yvbJ::PxylA-yhcS (kan)	This Study	BAB1097
<i>Bacillus subtilis</i> PY79 yfkN::tet yhdG::Phyperspank-yfkN-S1346A-His6 (spc) yvbJ::PxylA-yhcS (kan)	This Study	BAB1098
<i>Bacillus subtilis</i> PY79 yhdG::Phy-spank-ALFA-yfkN (spc) yfkN::tet yvbJ::PxylA-yhcS (kan)	This Study	BAB1212
<i>Bacillus subtilis</i> PY79 yhdG::Phy-spank-ALFA-yfkN-S1346A (spc) yfkN::tet yvbJ::PxylA-yhcS (kan)	This Study	BAB1213
<i>Bacillus subtilis</i> PY79 yhdG::Phyperspank-ALFA-yfkN (spc) yfkN::tet yhcS::cat	This Study	BAB1271
<i>Bacillus subtilis</i> PY79 yhdG::Phyperspank-ALFA-yfkN-S1346A (spc) yfkN::tet yhcS::cat	This Study	BAB1272
<i>Bacillus subtilis</i> PY79 sacA::Pyfkn-mYPet	This Study	BAB1124
<i>Bacillus subtilis</i> PY79 yhdG::Phy-spank-ALFA-yfkN (spc) yfkN::tet yvbJ::PxylA-yhcS (kan) sacA::Pveg-gfp (Phleo)	This Study	BAB1731
<i>Bacillus subtilis</i> PY79 yhdG::Phy-spank-ALFA-yfkN-S1346A (spc) yfkN::tet yvbJ::PxylA-yhcS (kan) sacA::Pveg-gfp (Phleo)	This Study	BAB1732
<i>Bacillus subtilis</i> WB600N yvbJ::PxylA-yhcS (kan) yhdG-Phy-spank-SS(yfkN)-ALFA-PmCBM26-longerCWA(yfkN)	This Study	BAB1707
<i>Bacillus subtilis</i> WB600N yvbJ::PxylA-yhcS (kan) yhdG-Phy-spank-SS(yfkN)-ALFA-PmCBM26-longerCWA (yfkN)-S623A (spec)	This Study	BAB1708
<i>Bacillus subtilis</i> WB600N yvbJ::PxylA-yhcS (kan) yhdG::Phy-spank-SS(yfkN)-ALFA-CBM35-longerCWA(yfkN) (spc)	This Study	BAB1717

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<i>Bacillus subtilis</i> WB600N yvbJ::P _{xytA} -yhcS (kan) yhdG::Phy-spank-SS(yfkN)-ALFA-CBM35-longerCWA(yfkN)-T770A (spec)	This Study	BAB1718
Chemicals, peptides, and recombinant proteins		
LB Miller	Genesee	Cat #11-118
Bacto Tryptone	ThermoFisher	Cat #211705
Yeast extract	VWR	Cat #97064-372
PBS	Fisher	Cat # BP3991
Benzonase	Sigma	Cat #E1014-5KU
Scal-HF	NEB	Cat # R3122L
NheI-HF	NEB	Cat # R3131L
XhoI-HF	NEB	Cat # R0146L
BamHI-HF	NEB	Cat # R3136L
HindIII-HF	NEB	Cat # R3104L
SpeI-HF	NEB	Cat # R3133L
NotI-HF	NEB	Cat # R3189L
Pfu Turbo	Agilent	Cat #600250
cOmplete Protease Inhibitor Cocktail	Sigma	Cat #11697498001
IPTG	VWR	Cat #14213-263
D-(+)-Xylose	Sigma	Cat # X1500-500G
Instant Blue	Abcam	Cat #ab119211
β-mercaptoethanol	Chemcruz	Cat #SC-202966
NaCl	Sigma	Cat #S3014-1KG
MgCl ₂	Macron	Cat #5958-04
NaOH	VWR	Cat #0583
HEPES	Sigma	Cat #H3375-1KG
EDTA	VWR	Cat #BDH4616
Lysozyme	Sigma	Cat #L6876-10G
DNase I	NEB	Cat #M0303L
RNase	Promega	Cat #A797C
PMSF	USB	Cat #20203-2564
Glycerol	Sigma	Cat #G5516-1L
Nonfat milk	Nestle	Cat #12428935
Tween-20	USB	Cat #20605
Bovine Serum Albumin	Sigma	Cat #B4287
Tris Base	Fisher	Cat #BP152-1
Ampicillin	RPI	Cat #A40040-25.0
Sucrose	VWR	Cat #BDH9308
Maleic acid	USB	Cat #18685
Mutanolysin	Sigma	Cat #M9901-1KU
Imidazole	Alfa Aesar	Cat #47274
Methanol	Fisher	Cat #MK-3016-1
Acetic acid	Sigma	Cat #W200611
Sodium dodecylsulfate	RPI	Cat #L22010
(NH ₄) ₂ SO ₄	Sigma	Cat #A4418
trisodium citrate	EMD	Cat #SX0445-1
FeCl ₃	MP Biomedical	Cat # 02153500-CF
MnCl ₂	Sigma	Cat #M3634
MgSO ₄	EMD	Cat #MX0075-1
ZnCl ₂	Sigma	Cat #Z-3500

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Glucose	JT Baker	Cat #1919-05
Bacto casamino acids	BD	Cat #223050
L-arginine	USB	Cat #11490
KH ₂ PO ₄	BDH	Cat #BDH0268
K ₂ HPO ₄	Sigma	Cat #D3786
Coomassie brilliant blue	Sigma	Cat #B-0770
Critical commercial assays		
KLD	NEB	Cat #M0554S
Western lightning plus ECL reagent	PerkinElmer	Cat #NEL103E001EA
Ni-NTA resin	Qiagen	Cat #30210
Recombinant DNA		
Plasmid pAB221: His-yfkN(MAIN)-SUMO (amp)	This Study	pAB221
Plasmid pAB222: His-yfkN(MAIN)-S1346A-SUMO (amp)	This Study	pAB222
Plasmid pAB225: yvbJ-PxylA-yhcS (kan) (amp)	This Study	pAB225
Plasmid pAB227: yhdG::Phyderspank-yfkN-His6 (spc) (amp)	This Study	pAB227
Plasmid pAB228: yhdG::Phyderspank-yfkN-S1346A-His6 (spc) (amp)	This Study	pAB228
Plasmid pAB230: His-yfkN(MAIN)-S1346T-SUMO (amp)	This Study	pAB230
Plasmid pAB233: His-SpneumoniaeR6_MAIN-SUMO (amp)	This Study	pAB233
Plasmid pAB234: His-Sagalactiae_MAIN-SUMO (amp)	This Study	pAB234
Plasmid pAB235: His-Athermocellus_MAIN-SUMO (amp)	This Study	pAB235
Plasmid pAB238: Opolygoni_MAIN-SUMO (amp)	This Study	pAB238
Plasmid pAB239: Smitis_MAIN-SUMO (amp)	This Study	pAB239
Plasmid pAB240: His-Sagalactiae_MAIN-S115A-SUMO (amp)	This Study	pAB240
Plasmid pAB247: yhdG::Phyderspank-ALFA-yfkN (spc) (amp)	This Study	pAB247
Plasmid pAB248: yhdG::Phyderspank-ALFA-yfkN-S1346A (spc) (amp)	This Study	pAB248
Plasmid pAB259: sacA::yfkN-mYPet (phleo) (amp)	This Study	pAB259
Plasmid pAB261: pminiMAD yfkN-S1346A	This Study	pAB261
Plasmid pAB263: yhdG::Phy-spank-SS(yfkN)-His6-NotI-CWA(yfkN) (spc) (amp)	This Study	pAB263
Plasmid pAB264: yhdG::Phy-spank-SS(yfkN)-PmCBM26-CWA(yfkN) (spc) (amp)	This Study	pAB264
Plasmid pAB265: yhdG::Phy-spank-SS(yfkN)-PmCBM26-CWA(yfkN)-S623A (spc) (amp)	This Study	pAB265
Plasmid pAB269: His-PmCBM26(MAIN)-SUMO (amp)	This Study	pAB269
Plasmid pAB270: His-PmCBM26(MAIN)-S623A-SUMO (amp)	This Study	pAB270
Plasmid pAB285: yhdG::Phy-spank-SS(yfkN)-ALFA-PmCBM26-CWA(yfkN) (spc) (amp)	This Study	pAB285
Plasmid pAB286: yhdG::Phy-spank-SS(yfkN)-ALFA-PmCBM26-CWA(yfkN)-S623A (spc) (amp)	This Study	pAB286
Plasmid pAB288: yhdG::Phy-spank-SS(yfkN)-ALFA-NotI-CWA(yfkN) (spc) (amp)	This Study	pAB288

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Plasmid pAB289: yhdG::Phy-spank-SS(yfkN)-ALFA-CBM35-CWA(yfkN) (spc) (amp)	This Study	pAB289
Plasmid pAB290: yhdG::Phy-spank-SS(yfkN)-ALFA-CBM35-CWA(yfkN)-T770A (spc) (amp)	This Study	pAB290
pAB295: His-GdCBM35(MAIN)-SUMO (amp)	This Study	pAB295
pAB296: His-GdCBM35(MAIN)-T770A-SUMO (amp)	This Study	pAB296
Software and algorithms		
PSI-BLAST	NCBI	https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSeBlas&LINK_LOC=blablasth
Clustal Omega version 1.2.4	Sievers et al. ⁴⁸	http://www.clustal.org/omega/#Documentation
Fiji version 2.16.0	NIH	https://fiji.sc/
NIS Elements version 5.2	Nikon	https://www.microscope.healthcare.nikon.com/products/software/nis-elements
PfamScan version 1.6	Finn et al. ²⁵	https://github.com/aziele/pfam_scan
DPAM	Zhang et al. ²⁶	https://github.com/CongLabCode/DPAM
Colabfold version 1.5.5	Mirdita et al. ⁴⁹	https://github.com/sokrypton/colabfold
Mmseqs2	Steinegger et al. ⁵⁰	https://github.com/soedinglab/mmseqs2
ChimeraX1.3	UCSF	https://www.cgl.ucsf.edu/chimerax/
PDBsum	Laskowski et al. ⁵¹	https://github.com/RomanLas/PDBsum1
iTOL	Letunic et al. ⁵²	https://itol.embl.de/
MATLAB version R2020a	MathWorks	https://www.mathworks.com/products/matlab.html
Adobe Illustrator version 24.1.3	Adobe	https://www.adobe.com/products/illustrator.html
Other		
A6000 GPUs	NVIDIA	Cat # TS4-126533247 -

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

All *B. subtilis* strains were derived from the prototrophic strain PY79.⁴⁶ Unless otherwise indicated, cells were grown in LB at 37 °C. Insertion-deletion mutations were generated by isothermal assembly⁵³ of PCR products followed by direct transformation into *B. subtilis*. The chimeric MAIN domain-containing proteins were engineered in the strain WB600N, lacking seven extracellular proteases.⁴⁷ Insertion-deletion mutants in *B. subtilis* were constructed by transforming 3-piece isothermal assemblies containing 1000bp upstream homology, 1000bp downstream homology, and an antibiotic resistance cassette. They were generated using the following oligos: oAB27/28 (universal oligos for antibiotic cassettes), *yfkN* (oAB465/466 & oAB467/468). Ectopic integration was performed by transformation of plasmids linearized by digestion with Scal-HF. A table of oligonucleotides and geneblocks used in this study can be found in supplementary data (Tables S1 and S2).

METHOD DETAILS

Immunoblot analysis

For Immunoblot analysis, cells were subcultured in 3 mL LB for 2.5 hours, back-diluted into 25 mL LB containing 500 μM IPTG and 20 mM xylose at OD₆₀₀ = 0.05 and grown for 6 hours (4 hours into stationary phase). 1 mL of cells normalized to OD₆₀₀ = 1 was pelleted and resuspended in lysis buffer (20 mM Tris pH 7.0, 10 mM MgCl₂ and 1mM EDTA, 1 mg/mL lysozyme, 10 μg/mL DNase I, 100 μg/mL RNase A, 1 mM PMSF) to a final OD₆₀₀ of 20 for equivalent loading. The cells were incubated at 37 °C for 15 min followed by addition of an equal volume of Laemelli sample buffer (0.25 M Tris pH 6.8, 4% SDS, 20% glycerol, 10 mM EDTA) containing 10%. Samples were heated for 5 min at 95 °C prior to loading. Proteins were separated by SDS-PAGE on 7.5 % polyacrylamide gels, electroblotted onto Immobilon-P membranes (Millipore) and blocked in 5% nonfat milk in phosphate-buffered saline (PBS) with 0.5% Tween-20. The blocked membranes were probed with anti-SigA (1:10,000),⁴⁵ anti-His (1:4,000) (GenScript), anti-ALFA (1:1,000) (Nanotag) antibodies diluted into 3% BSA in PBS with 0.05% Tween-20. Primary antibodies were detected using horseradish peroxidase-conjugated goat anti-rabbit or anti-mouse IgG (BioRad) and the Super Signal chemiluminescence reagent as described by the manufacturer (Pierce). Signal was detected using a Bio-Techne FluorChem R System.

When cell wall and membrane fractions were separated, cell pellets were washed 2 times in 1XSMM (0.5 M sucrose, 20 mM MgCl₂, 20 mM maleic acid pH 6.5), resuspended in 50 μ L SMM + 0.5 mg/mL lysozyme + 5 U Mutanolysin and incubated at room temperature for 30 minutes. The cells were monitored by light microscopy until >95% had formed protoplasts. The protoplasts were pelleted by centrifugation at 3k rpm for 10 minutes. The supernatant was collected and mixed with equal volume 2X Laemmli sample buffer containing 10% β -mercaptoethanol. The protoplast pellet was resuspended in 100 μ L 1X Laemmli sample buffer containing 5% β -mercaptoethanol. Samples were further processed as described above.

For heat treatment, cell pellets were resuspended in 100 μ L PBS + 1 mM PMSF and heated at 95°C for 5 minutes. The cells were pelleted (4000g/4min), the supernatant combined with 2X Laemmli sample buffer containing 10% β -mercaptoethanol, and the cell pellet was lysed as described above.

For SDS treatment with and without lysozyme, 20 mL of culture was resuspended in 1 mL PBS + 1 mM PMSF + 10 μ g/mL DNase I + 100 μ g/mL RNase with or without 0.5 mg/mL lysozyme. Cells were lysed via sonication and 100 μ L of the sample combined with 100 μ L 2X Laemmli sample buffer containing 10% β -mercaptoethanol.

Purification of His-MAIN-SUMO variants

Expression plasmids containing His-MAIN-SUMO variants were transformed into *E. coli* BL21(DE3) Δ tonA cells. Transformants were sub-cultured in terrific broth (TB) + 100 μ g/mL ampicillin for 3 hours then back-diluted into 25 mL TB + 100 μ g/mL ampicillin to an OD₆₀₀ of 0.01 and grown at 37 °C to an OD₆₀₀ of 0.4. Terrific broth consists of: 1.2% tryptone, 2.4% yeast extract, 0.4% (v/v) glycerol, 0.17 M KH₂PO₄, 0.72 M K₂HPO₄. Cultures were induced with 500 μ M IPTG for 3 hours at 37 °C and then harvested by centrifugation at 4000 rpm for 15 minutes. Cells were concentrated 25-fold by resuspending the pellet in 1 mL Lysis Buffer (50 mM HEPES-NaOH, 300 mM NaCl, 25 mM Imidazole, 10% (v/v) glycerol) and frozen at -80 °C. Cell pellets were thawed on ice and lysed by sonication. 25 U Benzonase (Sigma) and 1X complete protease inhibitor (Roche) were added to the cell lysate and incubated on ice for 15 min. Lysates were clarified by ultracentrifugation at 40,000 rpm for 45 minutes at 4 °C. The supernatant was added to 25 μ L of Ni²⁺-NTA resin, washed with 120 bed volumes of wash buffer (20 mM HEPES-NaOH, 300 mM NaCl, 25 mM Imidazole, 10% (v/v) glycerol), and eluted in 100 μ L of elution buffer (20 mM HEPES-NaOH, 300 mM NaCl, 300 mM Imidazole, 10% (v/v) glycerol). Load and Elution fractions were resolved by 12.5% SDS-PAGE gels and stained with Instant Blue (Abcam).

Size Exclusion Chromatography

His-MAIN-SUMO and His-MAIN^{S1346A}-SUMO were analyzed by size-exclusion chromatography (SEC) on a Superdex S75 column (GE Healthcare) in buffer containing 20 mM of HEPES-NaOH, pH 7.5 and 300 mM of NaCl. The load and peak absorbance (A₂₈₀) fractions were diluted into equal volume 2X sample buffer containing 10% β -mercaptoethanol and resolved by SDS-PAGE using a 17.5% polyacrylamide gel and stained with Instant Blue (Abcam). Absorbance profiles were plotted using MATLAB.

Edman Degradation

Purified His-MAIN-SUMO as described above was separated by SDS-PAGE using a 17.5% polyacrylamide gel and transferred to a PVDF membrane at 90 V for 60 min. The membrane was then washed 5X with ddH₂O, stained with 0.02% Coomassie Brilliant blue in 40% methanol, 5% acetic acid for 30 sec, destained in 40% methanol, 5% acetic acid for 1 min, washed 3X with ddH₂O and allowed to air dry. The bands of interest were cut from the membrane and sent for 5 cycles of Edman Degradation analysis with an ABI 494 Protein Sequencer (Tufts University Core Facility).

Bioinformatics Analysis

A local PSI-BLAST run was separately performed on the amino acid sequences of *B. subtilis* YfkN's MAIN domain and *H. sapiens* MUC1's SEA domain against the RefSeq Select Protein database using an e-value of 0.05 and 5 iterations. Generally, MAIN homologs were found to have 8-100% sequence identity to *B. subtilis* MAIN and 5-13% sequence identity to GAIN domains. For the MAIN domain, eukaryotic proteins were removed from the results and the remaining protein sequences retrieved.⁵⁴ TaxIDs of all MAIN domain homologs were used to plot onto a phylogenetic tree of 5767 unique bacterial taxa. Tree visualization and manipulation was done using iTOL.⁵²

PfamScan

The ~15,000 proteins identified by PSI-BLAST were annotated using PfamScan against the Pfam database release 31.²⁵ The output domains were organized by RefSeq accession number and the resulting domain organizations built (Data S2; Figure S15). Counts of each domain found associated with the MAIN domain can be found in Data S1A.

DPAM

MAIN domain-containing proteins identified by PSI-BLAST were clustered using MMseqs2⁵⁰ using a minimum sequence identity of 50% and a coverage threshold of 80%. The structures of the resulting 2,843 unique proteins were predicted using colabfold version 1.5.5.^{49,55} with the following parameters: each prediction was run with five models, using three recycles, no dropout, no templates. These runs were performed using a local installation of Colabfold running on NVIDIA A6000 GPUs managed by the HMS O2 computing cluster. MSAs were generated using MMseqs2 API server.⁵⁰ The rank 1 prediction (based on pLDDT) from each AlphaFold2 run was used as an input into DPAM²⁶ (Figure S12). The resulting output domains were organized by RefSeq accession number and the domain organizations built (Data S2).

Multiple Sequence Alignment

Multiple sequence alignments were generated from the results of the previously described psi-blast searches using Clustal Omega.⁴⁸ NCBI taxID was used to filter the MSA for organisms from the Bacillota (Firmicute) phylum.

Alphafold2 Predictions

Protein structures were modeled using colabfold version 1.5.5.^{49,55} Each prediction was run with five models, using three recycles, no dropout, no templates. These runs were performed using a local installation of Colabfold running on NVIDIA A6000 GPUs managed by the HMS O2 computing cluster. MSAs were generated using MMseqs2 API server.⁵⁰ All predictions used paired and unpaired MSAs, and structures were unrelaxed. The highest ranked structure by pLDDT was used for structural annotation.

Structural model visualization

The crystal structure of CIRL1/Latrophilin's GAIN domain (PDB: 4dlq) was downloaded from the PDB. ChimeraX1.3 was used to visualize the structural models and generate images. PDBsum was used to generate protein topology maps.⁵¹

Fluorescence Microscopy

Fluorescence microscopy was performed on a Nikon Ti2-E inverted microscope equipped with Plan Apo 100x/1.45 phase contrast oil objective and a Hamamatsu ORCA-Flash4.0 V3 Digital CMOS camera. Light was delivered with the Lumencor Spectra III Light Engine containing the following filters: 390/22, 440/20, 475/28, 510/25, 555/28, 575/25, 637/12, 748/12. Images were acquired using the Nikon Elements 5.2 acquisition software. Cytoplasm was labeled with constitutively expressed GFP when indicated. Images were cropped and adjusted using Fiji. All images shown side by side were acquired, analyzed, and adjusted, identically for comparison. Images are shown at a magnification of 100x unless otherwise indicated.

P_{yfkN}-mYPet expression

Cells were subcultured in low phosphate defined media (LPDM) for 3 hours before back-dilution into 25 mL LPDM at OD₆₀₀ = 0.05 and grown for 9 hours. 1 mL of cells was sampled every 30 minutes, the OD₆₀₀ recorded, pelleted, resuspended in 50 μ L LPDM, spotted onto 2% LPDM agarose, and imaged as described in fluorescence microscopy. LPDM consists of: 50 mM Tris pH 7.1, 3.03 mM (NH₄)₂SO₄, 6.8 mM trisodium citrate, 3.04 mM FeCl₃, 1 mM MnCl₂, 3.5 mM MgSO₄, 0.01 mM ZnCl₂, 0.5% glucose, 0.05% Casamino acids, 10 mM L-arginine, 0.065 mM KH₂PO₄. Similar results were obtained using LB medium.

Anti-ALFA Immunolabeling

Cells expressing ALFA-tagged proteins were subcultured in 3 mL LB for 2.5 hours, back-diluted into 25 mL LB containing 500 μ M IPTG and 20 mM xylose at OD₆₀₀ = 0.05 and grown for 6 hours into stationary phase. 1 mL of cells normalized to OD₆₀₀ = 1 was pelleted, washed 1X in PBS, and labeled in PBS + 1:500 anti-ALFA-Alexa647 (Nanotag) for 15 minutes at room temperature. Cells were washed 1X in PBS and resuspended in 100 μ L PBS before spotting 6.5 μ L onto 2% PBS agarose pads and imaging as described in fluorescence microscopy.

Plasmid Construction

All plasmids were cloned into *Escherichia coli* DH5a and sequence verified using sanger or nanopore sequencing.

pAB221 [His-yfkN(MAIN)-SUMO (amp)]

pAB221 was generated in a 2-piece isothermal assembly reaction using a PCR product amplified from PY79 gDNA using oAB443 & oAB444 and pTD68 [His-SUMO (amp)] digested with NheI-HF.

pAB222 [His-yfkN(MAIN)-S1346A-SUMO (amp)]

pAB222 was generated in a KLD (NEB) reaction using a PCR product amplified from pAB221 using oAB436 & oAB437.

pAB225 [yvbJ-PxylA-yhcS (kan) (amp)]

pAB225 was generated in a 2-piece isothermal assembly reaction using a PCR product amplified from PY79 gDNA using oAB441 & oAB442 and pCB133 [yvbJ::PxylA (kan) (amp)] digested with XhoI-HF & BamHI-HF.

pAB227 [yhdG::Phyterspank-yfkN-His6 (spc) (amp)]

pAB227 was generated in a 3-piece isothermal assembly reaction using PCR products amplified from PY79 gDNA using oAB443 & 448 and oAB444 & 447, and pCB101 [yhdG::Phyterspank (spc) (amp)] digested with HindIII-HF & SpeI-HF.

pAB228 [yhdG::Phyterspank-yfkN-S1346A-His6 (spc) (amp)]

pAB228 was generated in a KLD (NEB) reaction using a PCR product amplified from pAB227 using oAB436 & oAB437.

pAB230 [His-yfkN(MAIN)-S1346T-SUMO (amp)]

pAB228 was generated in a KLD (NEB) reaction using a PCR product amplified from pAB221 using oAB437 & oAB450.

pAB233 [His-SpneumoniaeR6_MAIN-SUMO (amp)]

pAB233 was generated in a 2-piece isothermal assembly reaction using a PCR product amplified from the SpMAIN GBlock using oAB455/456 and pTD68 [His-SUMO (amp)] digested with NheI-HF.

pAB234 [His-Sagalactiae_MAIN-SUMO (amp)]

pAB234 was generated in a 2-piece isothermal assembly reaction using a PCR product amplified from the SaMAIN GBlock using oAB455/456 and pTD68 [His-SUMO (amp)] digested with NheI-HF.

pAB235 [His-Athermocellus_MAIN-SUMO (amp)]

pAB235 was generated in a 2-piece isothermal assembly reaction using a PCR product amplified from the AtMAIN GBlock using oAB455/456 and pTD68 [His-SUMO (amp)] digested with NheI-HF.

pAB238 [Opolygoni_MAIN-SUMO (amp)]

pAB238 was generated in a 2-piece isothermal assembly reaction using a PCR product amplified from the OpMAIN GBlock using oAB455/456 and pTD68 [His-SUMO (amp)] digested with NheI-HF.

pAB239 [Smitis_MAIN-SUMO (amp)]

pAB239 was generated in a 2-piece isothermal assembly reaction using a PCR product amplified from the SmMAIN GBlock using oAB455/456 and pTD68 [His-SUMO (amp)] digested with NheI-HF.

pAB240 [His-Sagalactiae_MAIN-S115A-SUMO (amp)]

pAB222 was generated in a KLD (NEB) reaction using a PCR product amplified from pAB239 using oAB463 & oAB464.

pAB247 [yhdG::Phyperspank-ALFA-yfkN (spc) (amp)]

pAB247 was generated in a 3-piece isothermal assembly reaction using a PCR product amplified from PY79 gDNA using oAB443 & 482, a PCR product amplified from PY79 gDNA using oAB444 & oAB483, and pCB101 [yhdG::Phyperspank (spc)] digested with HindIII-HF & SpeI-HF.

pAB248 [yhdG::Phyperspank-ALFA-yfkN-S1346A (spc) (amp)]

pAB247 was generated in a 3-piece isothermal assembly reaction using a PCR product amplified from PY79 gDNA using oAB443 & 482, a PCR product amplified from BAB1144 [yfkN-S1346A] gDNA using oAB444 & oAB483, and pCB101 [yhdG::Phyperspank (spc)] digested with HindIII-HF & SpeI-HF.

pAB259 [sacA::yfkN-mYPet (phleo) (amp)]

pAB259 was generated in a 3-piece isothermal assembly reaction using a PCR product amplified from Bs168 gDNA using oAB493 & oAB494, a PCR product amplified from pAB139 [ycgO::Pveg-ydaO-optRBS-mYPet (erm)] using oAB495 & oAB343, and pNC15 [sa-cA::(phleo)] digested with HindIII-HF & BamHI-HF.

pAB261 [pminiMAD yfkN-S1346A]

pAB261 was generated in a 3-piece isothermal assembly reaction using a PCR product amplified from PY79 gDNA using oAB501 & oAB502, a PCR product generated from PY79 gDNA using oAB503 & oAB504, and pminiMAD digested with HindIII-HF & EcoRI-HF.

pAB263 [yhdG::Phy-spank-SS(yfkN)-His6-NotI-CWA(yfkN) (spc) (amp)]

pAB263 was generated using a 3-piece isothermal assembly reaction using a PCR product amplified from PY79 gDNA using oAB443 & oAB512, a PCR product amplified from PY79 gDNA using oAB513 & oAB444, and pCB101 [yhdG::Phyperspank (spc) (amp)] digested with HindIII-HF & SpeI-HF.

pAB264 [yhdG::Phy-spank-SS(yfkN)-PmCBM26-CWA(yfkN) (spc) (amp)]

pAB264 was generated in a 2-piece isothermal assembly reaction using a PCR product amplified from the PmCBM26 GBlock using oAB514 & oAB515, and pAB263 digested with NotI-HF.

pAB265 [yhdG::Phy-spank-SS(yfkN)-PmCBM26-CWA(yfkN)-S623A (spc) (amp)]

pAB265 was generated in a 3-piece isothermal assembly reaction using a PCR product amplified from the PmCBM26 GBlock using oAB514 & oAB552, a PCR product amplified from pAB264 using oAB551 & oAB444, and pAB263 digested with NotI-HF.

pAB269 [His-PmCBM26(MAIN)-SUMO (amp)]

pAB269 was generated in a 2-piece isothermal assembly reaction using a PCR product amplified from the PmCBM26 GBlock using oAB525 & oAB526, and pTD68 [His-SUMO (amp)] digested with NheI-HF.

pAB270 [His-PmCBM26(MAIN)-S623A-SUMO (amp)]

pAB269 was generated in a 2-piece isothermal assembly reaction using a PCR product amplified from the PmCBM26 GBlock using oAB525 & oAB543, and pTD68 [His-SUMO (amp)] digested with NheI-HF.

pAB285 [yhdG::Phy-spank-SS(yfkN)-ALFA-PmCBM26-CWA(yfkN) (spc) (amp)]

pAB285 was generated in a 2-piece ITA using a PCR product amplified from PY79 gDNA using oAB443 & oAB482 and a PCR product amplified from pAB264a [yhdG-Phy-spank-SS(yfkN)-PmCBM26-CWA(yfkN) (spec) (amp)] using oAB562 & oAB444.

pAB286 [yhdG::Phy-spank-SS(yfkN)-ALFA-PmCBM26-CWA(yfkN)-S623A (spc) (amp)]

pAB286 was generated in a 2-piece ITA using a PCR product amplified from PY79 gDNA using oAB443 & oAB482 and a PCR product amplified from pAB265a [yhdG-Phy-spank-SS(yfkN)-PmCBM26-CWA(yfkN)-S623A (spec) (amp)] using oAB562 & oAB444.

pAB288 [yhdG::Phy-spank-SS(yfkN)-ALFA-NotI-CWA(yfkN) (spc) (amp)]

pAB288 was generated in a 3-piece isothermal assembly reaction using a PCR product amplified from PY79 gDNA using oAB443 & oAB482, a PCR product amplified from PY79 gDNA using oAB564 & oAB444, and pCB101 [yhdG::Phy-spank (spc) (amp)] digested with HindIII-HF & SpeI-HF.

pAB289 [yhdG::Phy-spank-SS(yfkN)-ALFA-CBM35-CWA(yfkN) (spc) (amp)]

pAB289 was generated in a 2-piece ITA using a PCR product amplified from the CBM35 GBlock (WP_109985763) using oAB565 & oAB566 and pAB288 [yhdG::Phy-spank-SS(yfkN)-ALFA-NotI-CWA(yfkN) (spc) (amp)] digested with NotI-HF.

pAB290 [yhdG::Phy-spank-SS(yfkN)-ALFA-CBM35-CWA(yfkN)-T770A (spc) (amp)]

pAB290 was generated in a Quikchange reaction (Agilent) using a PCR product amplified with Pfu turbo using oAB577 & oAB578.

pAB295 [His-CBM35(MAIN)-SUMO (amp)]

pAB295 was generated in a 2-piece isothermal assembly reaction using a PCR product amplified from the CBM35 GBlock (WP_109985763) using oAB571 & oAB572, and pTD68 [His-SUMO (amp)] digested with NheI-HF.

pAB296 [His-CBM35(MAIN)-T770A-SUMO (amp)]

pAB296 was generated in a 3-piece isothermal assembly reaction using a PCR product amplified from the CBM35 Gblock (WP_109985763) using oAB571 & oAB578, a PCR product amplified from the CBM35 Gblock (WP_109985763) using oAB572 & oAB577, and pTD68 [His-SUMO (amp)] digested with NheI-HF.

QUANTIFICATION AND STATISTICAL ANALYSIS

All experiments were carried out in biological triplicate unless otherwise indicated. For fluorescence microscopy, all experiments were carried out in biological triplicate, and at least 10 fields of view imaged and analyzed when quantified.

Current Biology, Volume 36

Supplemental Information

**Microbial GAIN domains undergo
autoproteolysis and enable release
of diverse cell-surface-associated proteins**

Anna P. Brogan and David Z. Rudner

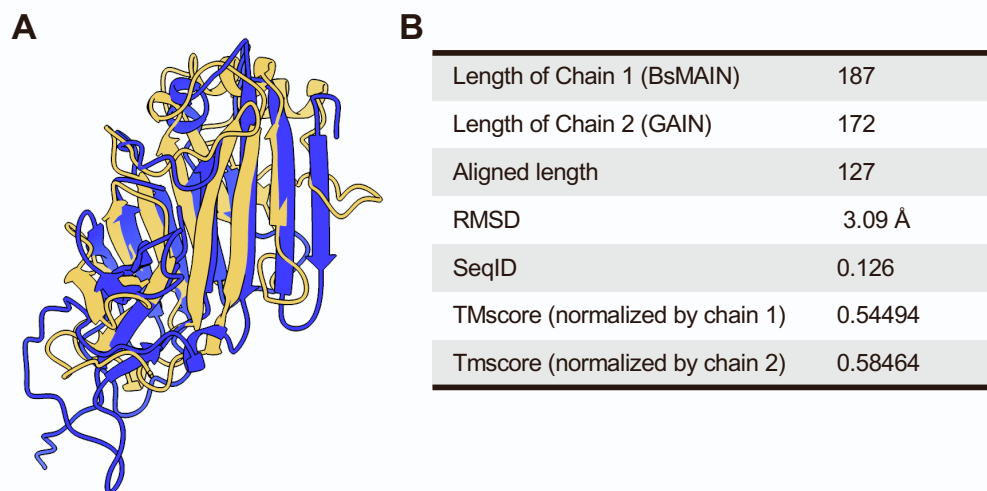


Figure S1. Alignment metrics for BsMAIN and Latrophilin/CIRL GAIN. Related to Figure 1. (A) Structural alignment by TM-align of the structures from Latrophilin/CIRL1's GAIN domain (blue) and YfkN's GAIN domain (orange). Confidence metrics for the AlphaFold2 prediction of *the B. subtilis* MAIN domain can be found in Figure S2. **(B)** Metrics from the structural alignment using TMalign.

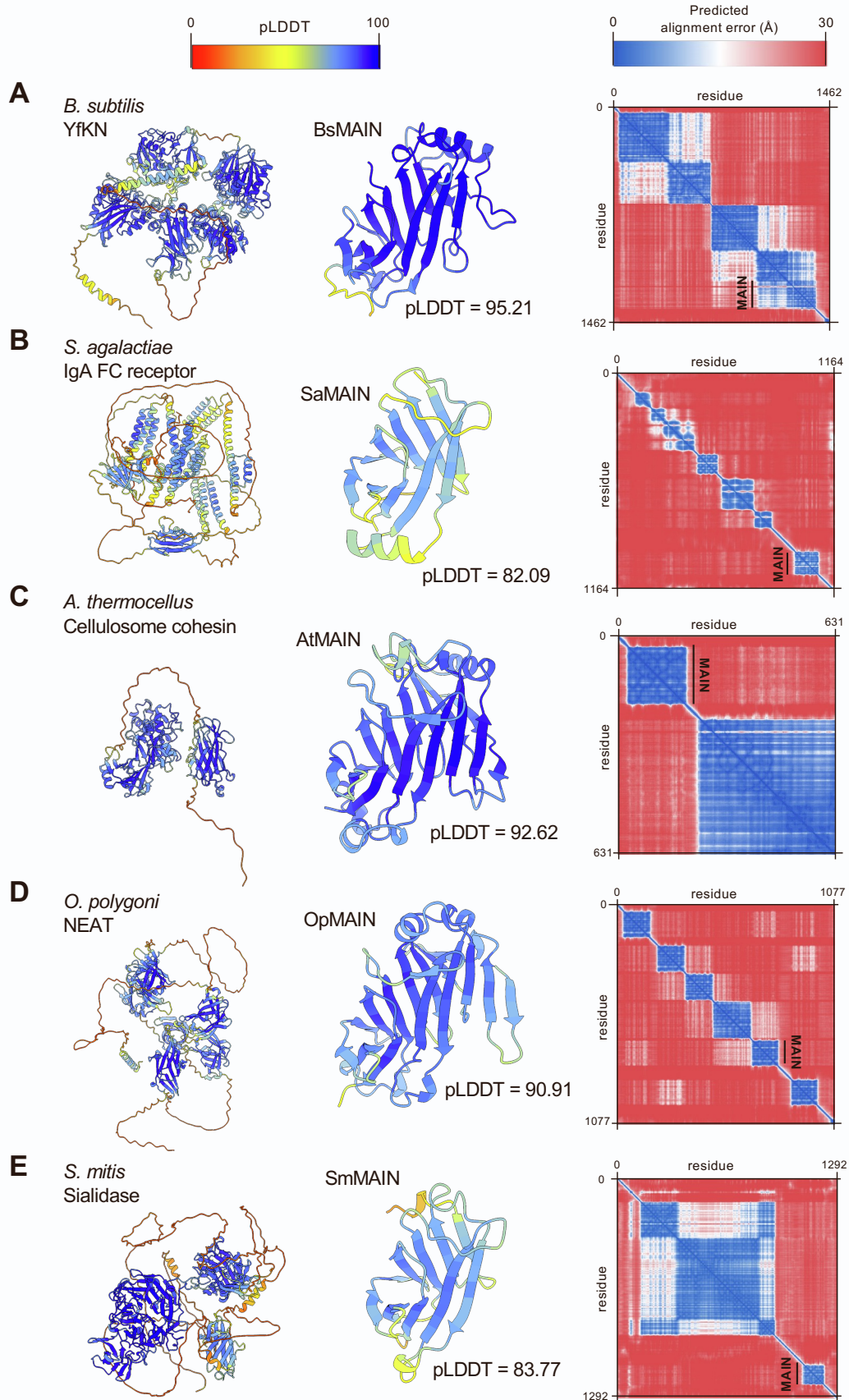


Figure S2. Metrics for AlphaFold2 predictions. Related to Figure 2. AlphaFold2 predictions of full-length MAIN domain containing proteins (left) and MAIN domains (center) colored by pLDDT and predicted alignment error (pAE) plots (right) for **(A)** *B. subtilis* YfKN, **(B)** *S. agalactiae* IgA FC receptor, **(C)** *A. thermocellus* cellulosome anchoring cohesin, **(D)** *O. polygoni* NEAT transporter, and **(E)** *S. mitis* sialidase.

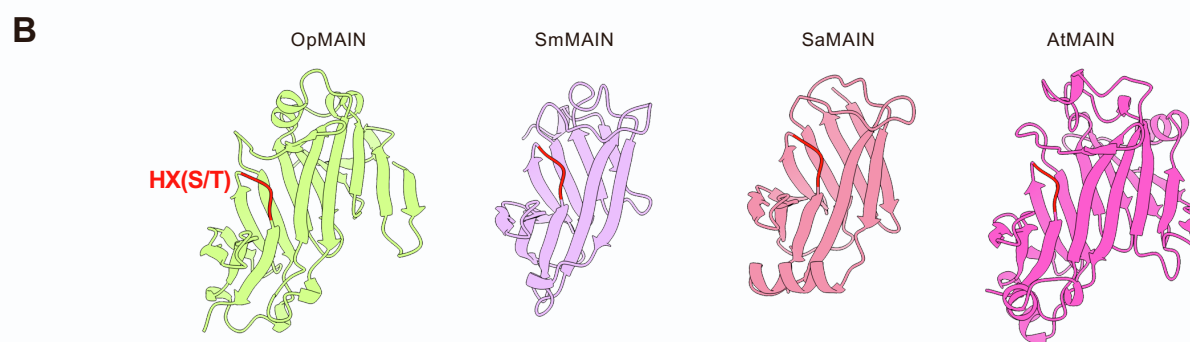


Figure S3. The site of autocleavage is conserved at the sequence and structural level. Related to Figure 2. (A) Multiple sequence alignment showing the alignment of the region proximal to the HX(S/T) autoproteolytic site between *H. sapiens* ADGRE3 and several bacterial MAIN domains. **(B)** AlphaFold predictions of MAIN domains from the indicated species with the HX(S/T) motif highlighted in red. Confidence metrics for the prediction can be found in Fig. S2.

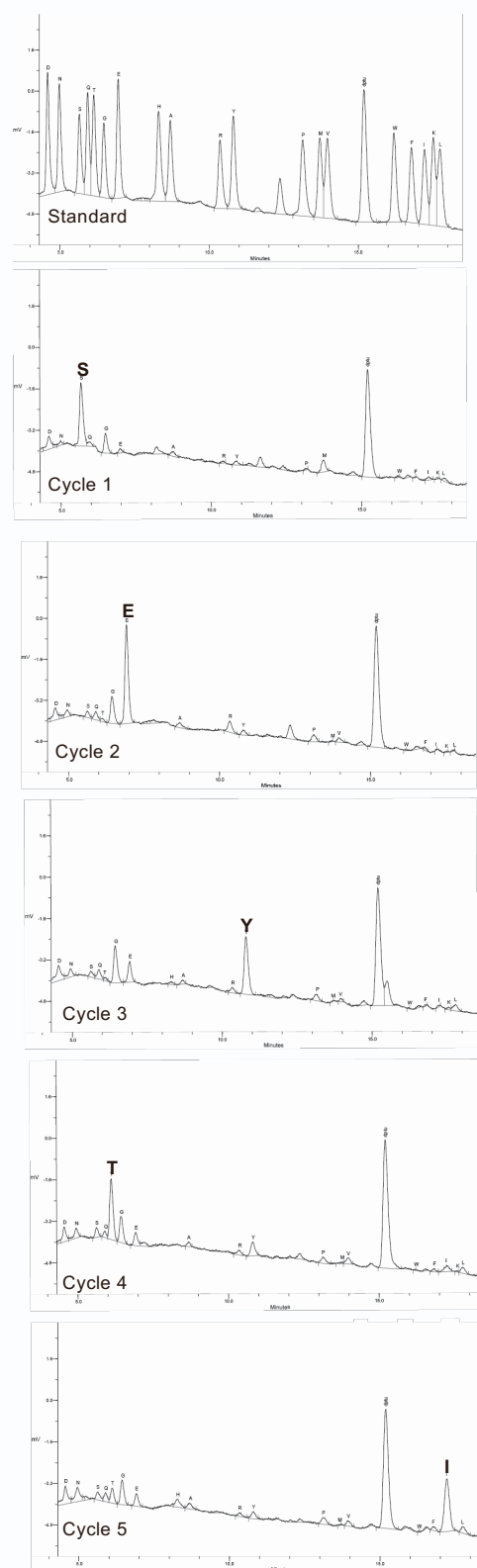


Figure S5. Edman degradation reveals that autocleavage occurs at the conserved HFS motif. Related to Figure 2. Chromatograms from 5 cycles of Edman degradation for the C-terminal cleavage product of BsMAIN as compared to standards of each amino acid.

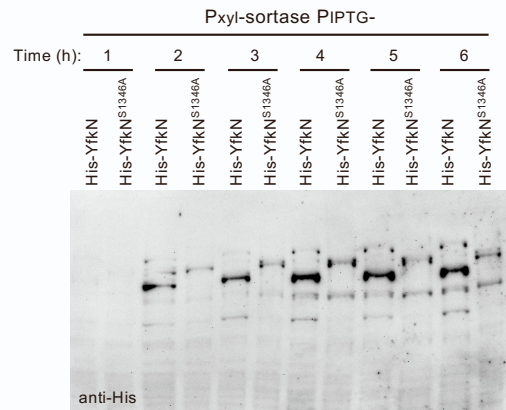


Figure S6. His-YfkN expression increases as cells enter stationary phase. Related to Figure 3. Immunoblot of the indicated *B. subtilis* strains expressing His-YfkN and His-YfkN (S1346A) from a strong IPTG regulated promoter induced with 500 μ M IPTG. Cells were subcultured to mid-log, back-diluted to OD600 = 0.05, and grown for 6 hours into stationary phase. Every hour, 1 mL of cells normalized to OD = 1 were harvested, lysed, and analyzed by immunoblot.

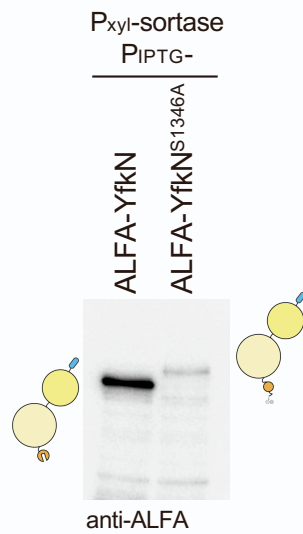


Figure S7. ALFA-tagged YfkN reveals a size shift between wild-type and an autoproteolysis-deficient mutant. Related to Figure 3. Immunoblot of *B. subtilis* stationary phase cells expressing ALFA-YfkN or ALFA-YfkN(S1346A) with 500 μ M IPTG. Cells express sortase with 20 mM xylose.

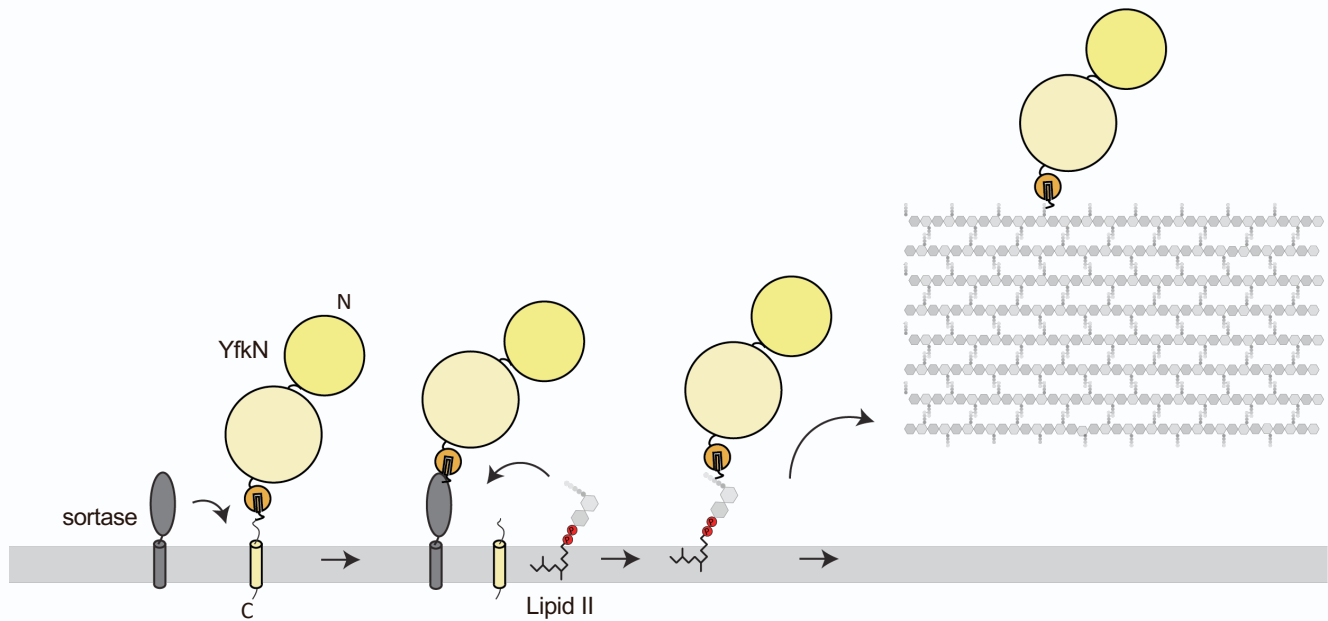


Figure S8. Mechanism of sortase-mediated cell wall anchoring. Related to Figure 3. Sortase substrates have recognition sequences at their C-termini. They are anchored in the membrane by this sequence, prior to transmembrane cleavage and crosslinking to the mDAP residue of the stem peptide of lipid II. Lipid II is then polymerized into the peptidoglycan cell wall.

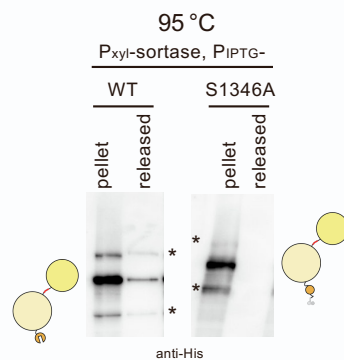


Figure S9. Heat treatment results in YfkN ectodomain release. Related to Figure 3. Immunoblot of *B. subtilis* cells expressing His-YfkN and His-YfkN(S1346A) from a strong IPTG regulated promoter induced with 500 μ M IPTG. Cells were subcultured to mid-log, back-diluted to OD600 = 0.05, and grown for 6 hours into stationary phase. Cells were pellet, resuspended in PBS and heated at 95°C for 5 minutes. Cell pellet and supernatant were separated by centrifugation. Anomalous migrating species are indicated (*).

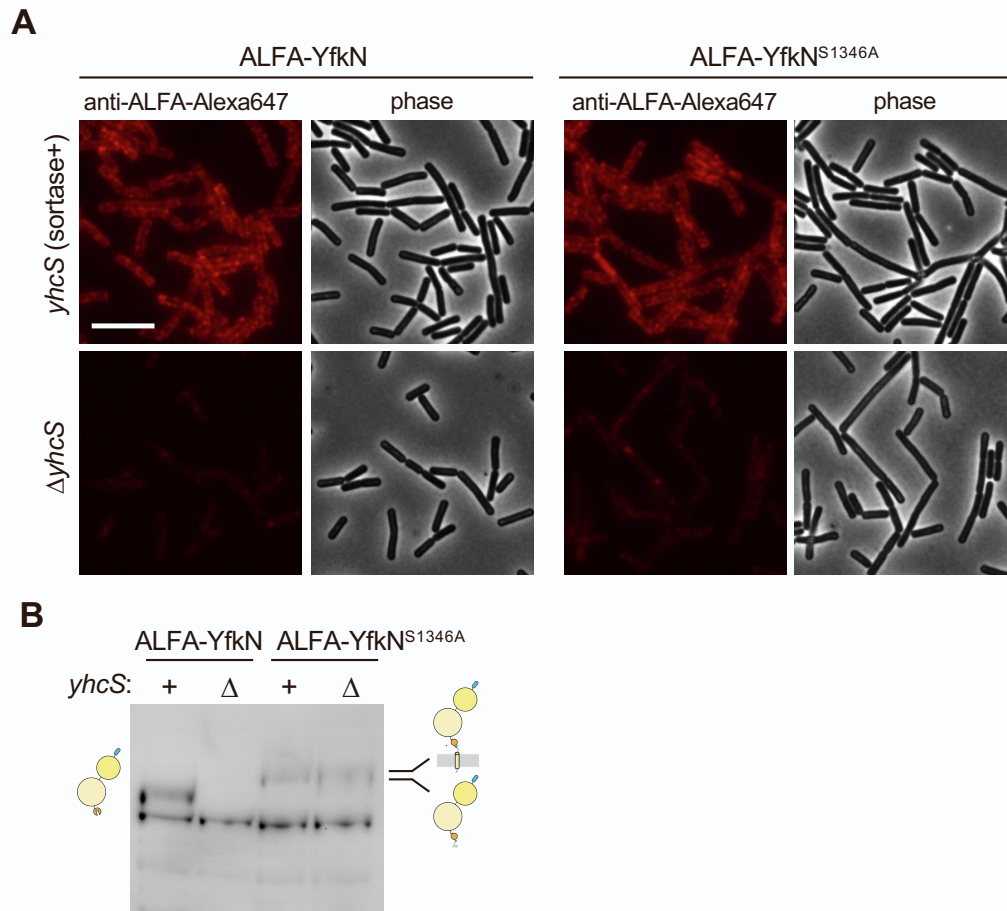


Figure S10. Anti-ALFA-Alexa647 does not label membrane-anchored ALFA-YhcN. Related to Figure 3. (A) Fluorescence and phase-contrast micrographs of stationary phase *B. subtilis* cells expressing ALFA-YfkN and ALFA-YfkN(S1346A) in the presence or absence of sortase (YhcS) immunolabeled with anti-ALFA-Alexa647. anti-ALFA-Alexa647 barely labels membrane anchored ALFA-YfkN(S1346A), even though it produced at levels similar to wall-anchored ALFA-YfkN(S1346A), as seen in Figure S12B. YhcN variants were expressed with 500 μ M IPTG. Scale bar indicates 5 μ m. **(B)** Immunoblot of cells expressing ALFA-YfkN and ALFA-YfkN(S1346A) in the presence (+) or absence (Δ) of sortase (*yhcS*). YfkN is unstable in the absence of sortase, but the autoproteolysis-deficient mutant is stable in the presence and absence of sortase.

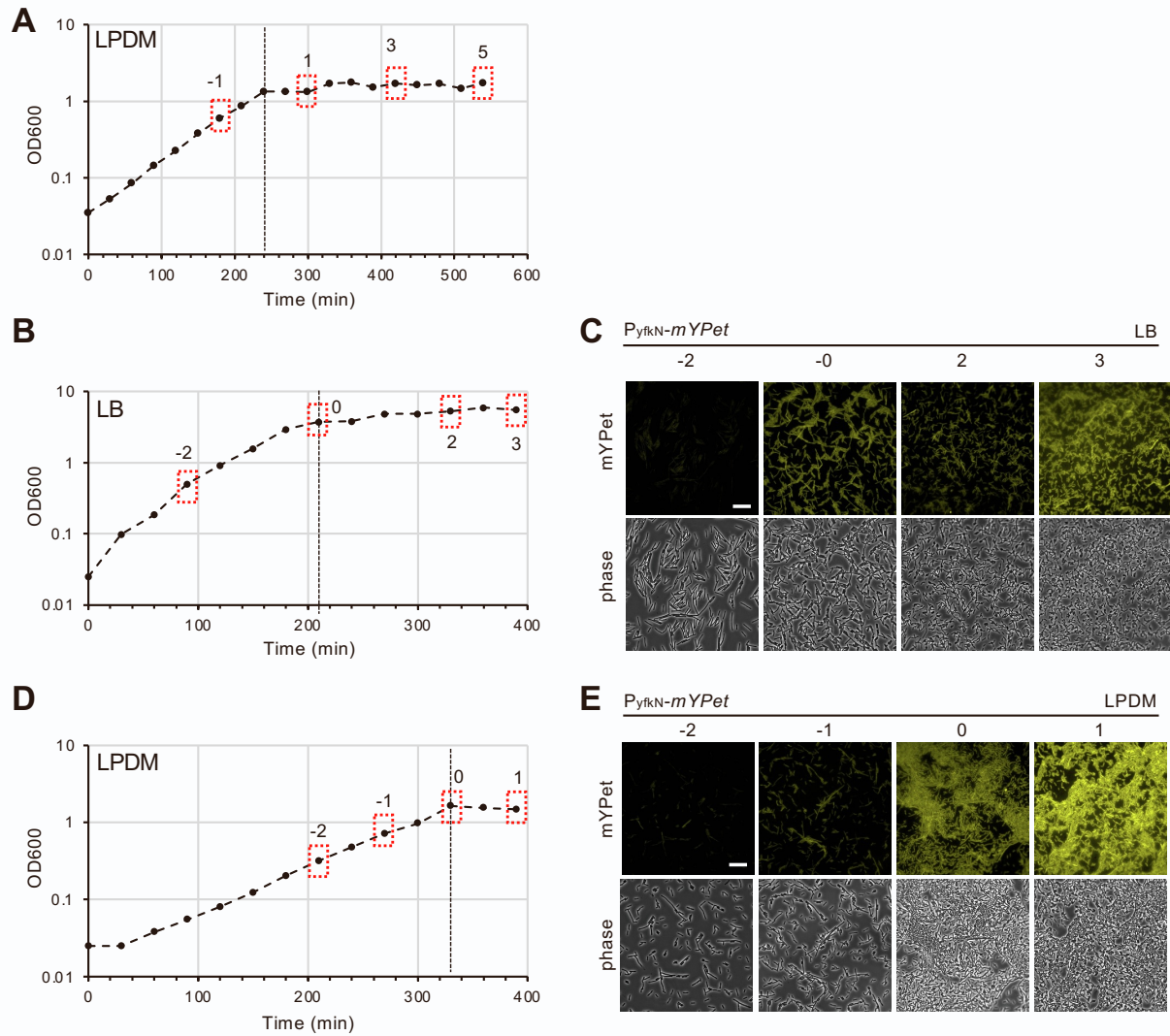


Figure S11. YfkN expression is induced in late stationary phase. Related to Figure 3. (A) Growth curve of the time course shown in Figure 3F. **(B)** Growth curve of *B. subtilis* cells harboring the P_{yfkN} -mYPet reporter grown in LB. **(C)** Fluorescence and phase-contrast micrographs of cells at the timepoints indicated in (B). **(D)** Growth curve of *B. subtilis* cells harboring the P_{yfkN} -mYPet reporter grown in LPDM. **(E)** Timepoints in (C) & (E) represent time in hours relative to entry into stationary phase. Horizontal dashed lines indicate entry into stationary phase. Scale bar indicates 10 μ m.

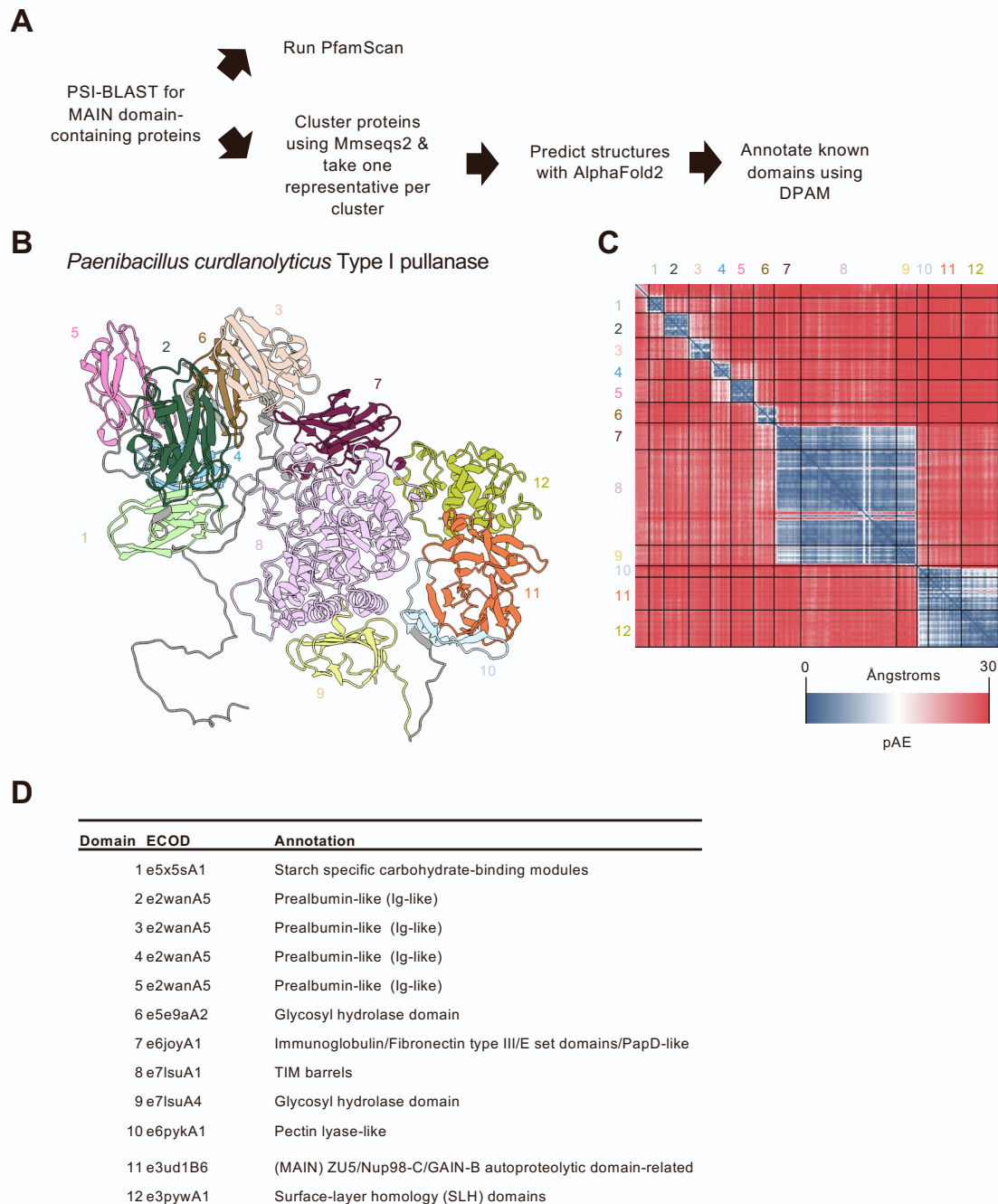


Figure S12. Bioinformatic analysis of MAIN domain-containing proteins. Related to Figure 4. (A) 15,000 MAIN domain containing proteins were identified by PSI-BLAST using *B. subtilis* YfkN's MAIN as a query. The results were used as input for PfamScan, and separately clustered using Mmseqs2 and one representative protein from each cluster used to predict the protein structure by AlphaFold2. Known domains were annotated using DPAM-AI. **(B-D)** DPAM-AI analysis of one representative MAIN domain-containing protein from *Paenibacillus curdlanolyticus* (WP_006036052). **(B)** AlphaFold2-predicted structure of WP_006036052 with domains annotated by DPAM-AI colored. **(C)** pAE plot of the predicted structure in **(B)**. Black lines indicate domain cutoffs identified by DPAM-AI. **(D)** Table listing the evolutionary classification of domains (ECOD) domains identified in WP_006036052.

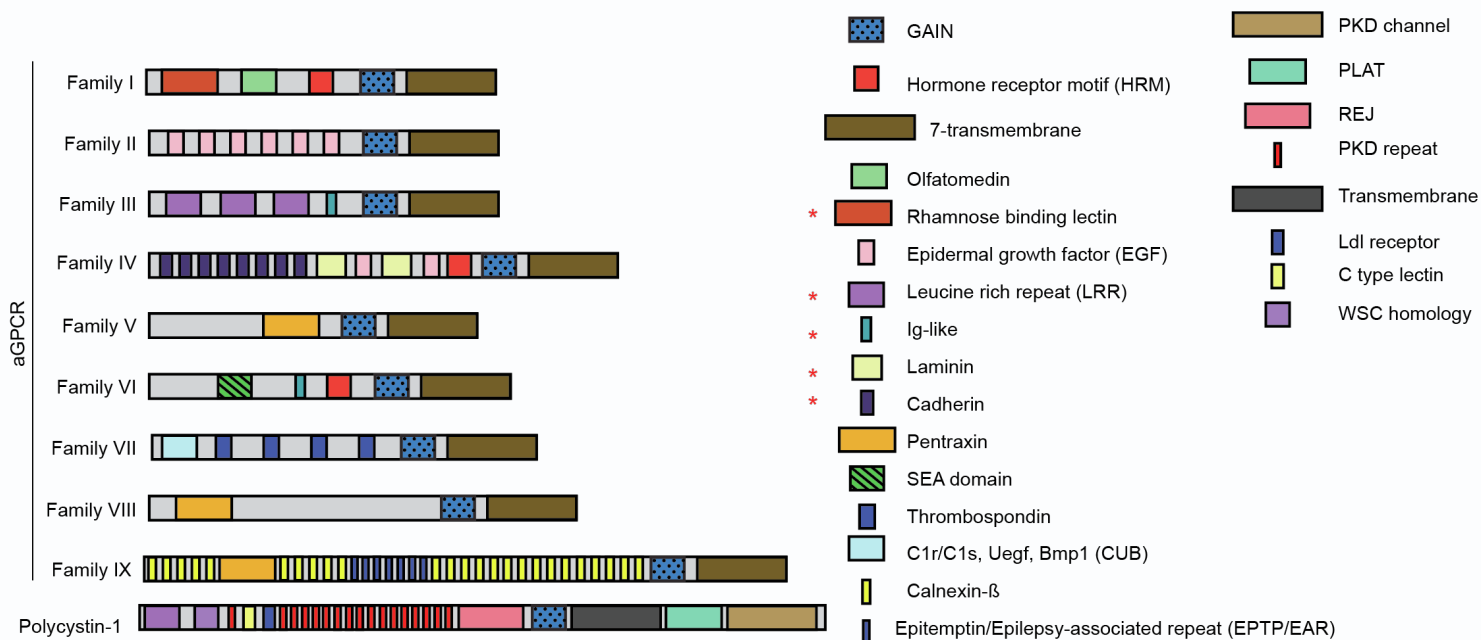


Figure S13. aGPCRs and MAIN-domain containing proteins have common domains. Related to Figure 4. Schematic of aGPCR protein architectures. The 33 aGPCRs can be subdivided into 9 families, which have similar domains. Red asterisks indicate domains that are found on both aGPCRs and MAIN domain containing proteins. Figure adapted from Langenhan et al (2013).

GO term	Z-score	P-value	FDR
hydrolase activity, acting on glycosyl bonds	17.97	1.37E-28	1.27E-26
serine hydrolase activity	9.33	9.39E-10	2.18E-08
glycosaminoglycan binding	6.58	2.66E-06	2.47E-05
transmembrane receptor protein kinase activity	6.67	8.89E-06	6.89E-05
collagen binding	8.27	4.74E-07	7.35E-06
calcium ion binding	6.39	2.05E-06	2.38E-05
endopeptidase activity	5.48	2.40E-05	1.60E-04
cargo receptor activity	6.46	8.62E-05	4.45E-04
phosphatase activity	4.85	1.62E-04	7.51E-04
sulfur compound binding	4.52	3.25E-04	1.31E-03
protein-lipid complex binding	5.84	4.34E-04	1.62E-03
extracellular matrix structural constituent	5.54	6.33E-04	2.27E-03
G-protein coupled receptor activity	4.99	7.15E-04	2.46E-03
exopeptidase activity	4.41	8.38E-04	2.52E-03
structural constituent of cytoskeleton	4.83	8.99E-04	2.61E-03
carboxylic ester hydrolase activity	4.33	9.63E-04	2.71E-03
axon guidance receptor activity	5.58	1.43E-03	3.60E-03
growth factor binding	4.14	1.80E-03	4.39E-03
antigen binding	4.61	2.08E-03	4.61E-03
protein tyrosine kinase activity	4.06	2.07E-03	4.61E-03
integrin binding	4.06	2.07E-03	4.61E-03
carbohydrate binding	3.86	3.86E-03	8.35E-03
apolipoprotein binding	4.49	4.54E-03	9.38E-03
semaphorin receptor activity	4.49	4.54E-03	9.38E-03

Figure S14. GO enrichment analysis indicates that carbohydrate hydrolase activity is enriched in MAIN-domain containing proteins. Related to Figure 4. A GO enrichment analysis was performed on the Pfam domains encoded on the same polypeptide as MAIN domains. FDR = false discovery rate.

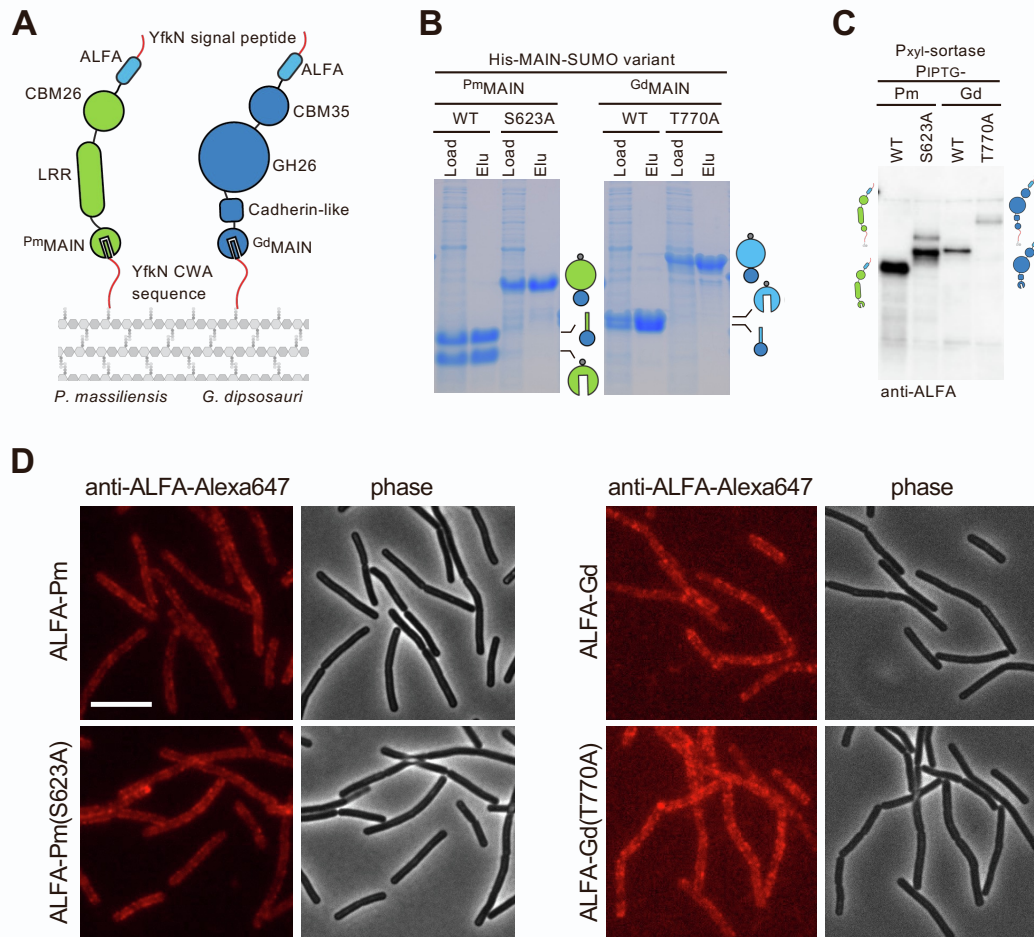


Figure S15. MAIN domain containing proteins from diverse bacteria are autoproteolytic and surface associated. Related to Figure 4. (A) Schematic of the system used to anchor chimeric proteins to the *B. subtilis* cell wall. The YfkN signal sequence was fused to the N-terminus and the YfkN cell wall anchoring (CWA) sequence and unstructured region were fused to the C-terminus. MAIN domain containing proteins from *Pseudoruminococcus massiliensis* (WP_106763535) and *Gracilibacillus dipsosauri* (WP_109985763). CBM26 is a putative starch binding fold, CBM35 is putative xylan-binding domain, GH26 is a putative xylan or mannan polysaccharide lyase. Both chimeras were ALFA-tagged at their N-termini just after the signal sequence. **(B)** Coomassie-stained polyacrylamide gel of His-MAIN-SUMO constructs of the indicated variants. Constructs were expressed in *E. coli* and clarified lysate (Load) were incubated with Ni²⁺-NTA resin. Proteins were eluted (Elu) with imidazole. **(C)** Immunoblot of *B. subtilis* cells expressing the indicated constructs and autoproteolysis deficient mutants from (A). **(D)** Fluorescence and phase micrographs of *B. subtilis* cells expressing ALFA-tagged variants of the indicated chimeras. Cells were grown 6 hours into stationary phase and labeled with anti-ALFA-Alexa647. Scale bar indicates 5 μ m.

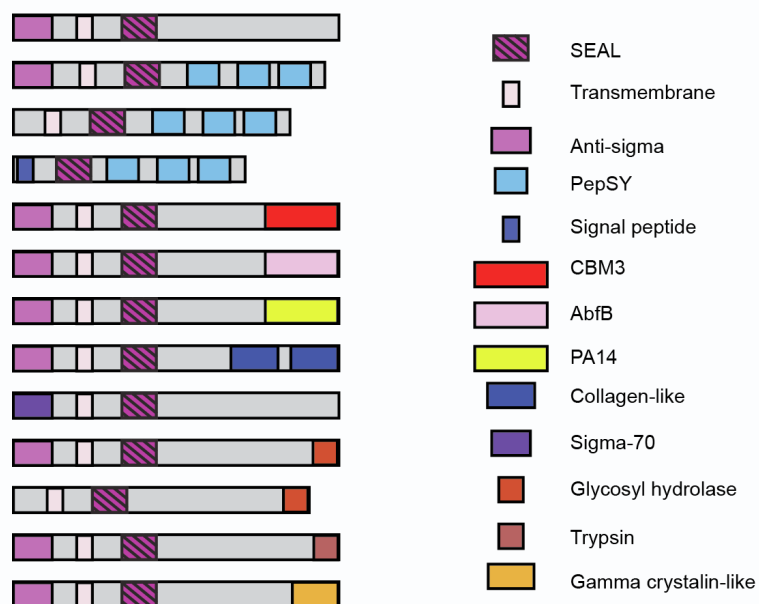


Figure S17. SEA domains are fused to lectins and adhesion domains. Related to Figure 4. Schematic depiction of SEAL domain-containing proteins from diverse bacteria. Adapted from Brogan et al.

Oligo	Sequence
oAB27	TTCTGCTCCCTCGCTCAG
oAB28	CAGGGAGCACTGGTCAAC
oAB436	GCGgaatacacgatcttaaattccgggtcaacaac
oAB437	gaagtggttgtatatcctgtcaccatatcatc
oAB441	atgAAGCTTgaGCTAGCatCTCGAGaaaaggagcgcctccagaacatg
oAB442	cctcaaatgggtcgctggGATCCttaagtcactcgtttccatatataatatagcgcttc
oAB443	TTGTGAGCGGATAACAATTAAGCTTcggtgaaagggtggagaactgatgagaatac
oAB444	CgaGCTAGCatCTGCAGttACTAGTttatgcctgattcgctctattctgtttctc
oAB447	gattcaaatCATCATCATCACCATCATccagatcctcctaaagacggactgtg
oAB448	gatctggATGATGGTGATGATGATGatttgaatcatcatcgaattgcagcttatac
oAB450	ACAgaatacacgatcttaaattccgggtcaacaac
oAB455	TCACCACCACCACCACCATATGGCT
oAB456	GATTGACTTCTGAGTCCGATCCGCT
oAB463	gcgTTATTTGCAATTAAGACCTTGTCCAAAAG
oAB464	GAAATGATTGGTTTTGAATACAACCTGCCCCG
oAB465	cccgagctacgcgcaaaag
oAB466	AACGGTACTGAGCGAGGGAGCAGAAtgattctcaccagttctccaccttcaac
oAB467	AACGGTAGTTGACCAAGTGCTCCCTGaatcaggcataaaaaagagctgctcgc
oAB468	gggctggagcctccgctc
oAB482	CGTCTCCGAAGCTCCTCCTCGAGGCGGGAagccgcgctttctctgc
oAB483	CCTCGAGGAGGAGCTTCGGAGACGGTTAACTGAAcctcaggtgcatctgagtatttggc
oAB493	caggatcgatcgaattcatAAGCTTtcccctgattcataacctcccg
oAB494	TTTGACATagtagttctctcttatgcagttctccaccttcaacgaaatttgtg
oAB495	tgaagggtggagaactgcataaggagggaactactATGTCAAAGGCGAAGAGCTGTTTAC
oAB501	ctatgaccatgattacgccaagcttgtggcgcttgacggcc
oAB502	accggaatttaagatcgtgtattcCGCgaagtggttgtatatcctgtcaccatatcatc
oAB503	gtgacaggatatacaaaccacttcGCGgaatacacgatcttaaattccgggtcaacaac
oAB504	ttgtaaaacgacggccagtgaaatcggcatgcgactaaggaatttgacc
oAB512	gaccgtctgttctgaGCGGCCGCATGATGGTGATGATGATGagccgcgctttctctgc
oAB513	aagaaagcgcggctCATCATCATCACCATCATGCGGCCGctcaggaacagacgggtccggc
oAB514	gcagaagaaagcgcggct
oAB515	gccggaccgtctgttctctg
oAB525	CACCACCACCACCACCATATGGCTAACTCTAAGGTGTTGACATCAAAGGTTA ATAAC
oAB526	GATTGACTTCTGAGTCCGATCCGCTCGTTGTGTCAGCTATCAACGCATAG GATTGACTTCTGAGTCCGATCCGCTCGTTGTGTCAGCTATCAACGCATAGATC
oAB543	GCGAGG
oAB551	CACCACCGATCACCTCGCGATCTATG

oAB552	CATAGATCGCGAGGTGATCGGTGGTG
oAB562	TCGAGGAGGAGCTTCGGAGACGGTAACTGAAAGCATCGACTTCGACTCTAA TAATACAG
oAB564	TCCCGCCTCGAGGAGGAGCTTCGGAGACGGTAACTGAAGCGGCCGCtcaggaa c
oAB565	GGAGCTTCGGAGACGGTAACTG
oAB566	cgccagccggaccgtctg
oAB571	TCACCACCACCACCACCATATGGCTTCTTATCTTTCACTGGACCTTGTAAGA AGTCAG
oAB572	GATTGACTTCTGAGTCCGATCCGCTATGATCGGATGCTTGAGACAATTCTTT TC
oAB577	GAGACGGATCATTTTGGCGATATTTTCCGTATTC
oAB578	GAATACGGAAAATATCGCAAAATGATCCGTCTC

Table S1. Oligonucleotides used in this study. Related to STAR Methods.

Name	Sequence
SpMAIN	TCACCACCACCACCACCATATGGCT AAA GAG GAT AAG AGT GCT GTT TTT TCA GAT TCC AAG ACC GGC GTG G AA GTC CAT TTT TCA AAC AAA GAA AAA ACG GTC ATT AAG GGA CTT AA A GTT GAG CGC GTG CAG GCT TCG GCA GAA GAA AAA AAG TAC TTT GC G GGC GAG GAT GCT CAC GTA TTT GAG ATT GAG GGG CTG GAT GAG AA G GGA CAA GAC GTA GAC CTG TCC TAC GCT TCG ATT GTC AAA ATT CCC ATT GAA AAA GAT AAA AAA GTC AAG AAA GTC TTT TTT CTG CCT GAA G GG AAG GGG GCC GTC GAG CTG GCA TTC GAA CAG ACC GAT AGT CAT G TG ATC TTT ACC GCC CCT CAC TTT ACG CAC TAT GCT TTT GTT TAT GAG TCC GCC GAG AAA AGCGGATCGGACTCAGAAGTCAATC
SaMAIN	TCACCACCACCACCACCATATGGCT GTA GGC CAA GCT GTG TTT ACT AGC ACA GAT GGG AAC ACT AAG GTC A CC GTC GTC TTC GAT AAA CCT ACG GAC GCC GAT AAG TTG CAT TTG AA A GAA GTA ACG ACC AAG GAG TTA GCC GAT AAA ATC GCG CAC AAA AC T GGA GGT GGC ACC GTC CGT GTC TTT GAT TTA AGT TTG TCA AAG GGA GGG AAG GAA ACC CAC GTC AAC GGA GAG CGC ACC GTT CGT CTG GCT TTA GGA CAG ACA GGA TCG GAC GTT CAT GTT TAC CAC GTT AAG GAA A AC GGG GAC TTG GAA CGT ATC CCA TCT AAA GTC GAG AAC GGG CAG G TT GTA TTC AAA ACC AAT CAT TTC TCC TTA TTT GCA ATT AAG ACC TTG TCC AAA AGCGGATCGGACTCAGAAGTCAATC
AtMAIN	TCACCACCACCACCACCATATGGCT AGT GCC AAA GTT GAA CTG GAG CCT GTT TTA GAT AAC GCT ACG GGC G AA GCT AAG GCC GCC ATC GAC GAG GAA AAA CTT AAT AAG GCA TTA G AC GAG GCA AAG AAA AGT GAG GAC GAT AAA CTC GTT GAA CTT AAC A TC AAA AAA GTT GAG AAC GCG GAT GCC TAC ATA CAA CAA TTG CCT GC T AAG TTT TTG ATC AAG TCC GAC GCA GAA TAC AAG TTA CGG ATA GCA ACC GAG CAA GGG ATA ATT GAG GTC CCA GCG AAC ATG CTG AAT ACT GCG GAT ATA TCT AAG TTG GTA AAG AAC GAC TCT GTT GTA GAG TTC G TA ATT AGA AAA GTC AAG GTA GAC GAG CTT GGA GCC GAG CTT AAA G AG AAG ATC GGA AAT CGG CCA GTC ATA GAC ATA AGT GTG GTT GTC G AT GGG AAG AAA GTG GAA TGG AGT AAT TAT AAG GCA AAA GTT AAG A TA AGT ATC CCT TAT AAG CCT GAT GCG AAG GAA CTC GAA AAC CAC GA G CAT ATA GTC GTT TTA CAT ATA GAC GAT GCG GGG AAA GCC GTA TCC GTG CCG AGC GGG AAG TAC GAA CCA TCT TTG GGA GTC GTT ACG TTT G AG ACA AAC CAC TTA TCT AAG TAT GCA GTC TCA TAC GTC TAC AGCGGATCGGACTCAGAAGTCAATC
OpMAIN	TCACCACCACCACCACCATATGGCT GAA CCA ATT ATG GAG GAA GAT GAG GAT GGA AAT ATA ACT GTG AAG GTG ACG GAC GTG AAC AAC TTC AAT CAC AAG GAA GCT GAC AAT TCT T AC AAA TTA GAC TAC GAG GCA AAG AAT GTT CAT ATT GGC AAG GCA A TC CTT AGC GAG CTG GAT GCA AAT GCT AAT CTT GTC ATC GCT GGA AA A GGT GGT GTG ACG GCG ACG TTT GCT ATA CAG GAG TTA TTG GAC CAG TTG GGG TCC TCA GAA ACG CTG GTA ATT GAC TTT ACT GAT ATA ACC T CC GAC TAT CCT GAT GCT TTA TCA AAC GTT TAT GAC ATA ACC TTA AA G GCG GAC GAC GAG GCT ATC CAC ACG TTT GAT GAG GCT GTA CGC TTA ACC TTT ACG GTA GAT ACG GAC AAA GTT GGC AAC TGG GAT AAC GTG CAT ATC GCG TAC ATA AAT GAC ACA GAC GAA AAA GAG GTC CTG CAG ACA ACA ATC GAC CGG GAC AAT GGT ATC GTT ACG GCG GAC GTG AGT

	CAC TTC TCT ATC TAC GGA GTT TTC GAG AAT ATA AGCGGATCGGACTCAGAAGTCAATC
SmMAIN	TCACCACCACCACCACCATATGGCT AAA CAT ACT TTG AAA AAT GCG ACT GGT AGC GTC CAA ATT CAC GCA A GC GAG GAA GTC CTG AAA AAC GTC AAA GAT GTG AAA ATT CAG GAA G TC AAA GTC TCT TCC TTG TCA AGC CTT AAT TAC AAG GCG TAT AAC ATT CAG CTC AAT GAC GCC AGC GGT AAA GCG GTA CAG CCT AAG GGG ACA GTT ATT GTG ACT TTT GTA GCT GAA CAA AGC GTC AAA AAT GTC TAT T AT GTC GAT TCA AAG GGC AAC TTG CAT ACA TTA GAA TTT TTG CAG AA A GAC GGT GAA GTG ACC TTC GAA ACC AAT CAC TTC AGT ATC TAC GCA ATG ACA TTT CAG TTG TCA AGCGGATCGGACTCAGAAGTCAATC
PmCBM26 (WP_11716 8634)	gcagaagaaagcggctCATCATCATCACCATCATAGCATCGACTTCGACTCTAATA ATACAGTATATTTTCGCAAAACCAAGTGACAAATTGCTCGAGAATCTCAACC TCGATTTCTGGGGGGACAACATTAAGGTTTCATGCTTGGAATGAATCCGGTA ATCTCAACGGCGACTGGCCAGGGGATGCGGCAACACACGTAGAAGGAAAC ATATATTCCTATACCTGGTATGGAGTGTTCGACGGGAATTATAATTAATA ACGGAGATAACGGCGTACAGACTCAAAACATCGACGGAGGCTATACAGTGC CTAAGAGTGATGAGACTCAGGCCAGAAAAAACCTGAACTGGAATACGGA AAAGTGTATAAAGCTACGGGGGAGAGCGTGGAAGATAGCAACGGCAATAC CAACATTGTTGTAGCAATCAACGGGGAGTCCACATATAATGATTTTCGGTAC AGATTAGAAGACGACGACACAGTGACGATTCTGGGTACTTTGGTAGCGAT GAATCAATAACGATTCCGGATGAAATACACGGTTTTCCGGTTAAGGAAATC TCCAACGAAGGCTTCGATAACTGCGAGTATATCACGGTTTCAAATAATAAC AACTTTTATTGCAGTGAACGAGTTCTCTTTAATAAAGACAAAACAAAA TTGATAAGATATCCAGGTAAGAAAGCGGACGCCAAGTACTCTATTCCGGAT AGCGTGTTACAATAGGAGAGTCCGCTTTTGCTGTAAATAAAAGCATTGAG TCTATCGAAATCCCGGACAGCGTTCAGGAAATATGTGATCGTGCATTTCTGT CCTGCACAAAACCTCAAAGATATCAATATTCCGGATTTCAGTCGTCTTGATAGA GGACGGCGACGGCGCAGGGATATTTTCATACTGCACAAGTTTAACTAATAT AACTATTCCTAACTCAATCACTACTATCCCGTGGTACACTTTTTCCGGATGT ACGGGCCTTACCGATGTTACAATACCGAACTCTGTTAATACTATCGGGTGTG GTGCTTTTAGCGGTTGCACTGGCTTGAAGAACATCACAATTCCTGACAGCGT TACCGAGATTGAAAGTAATGCTTTTCGCCGATTGTTTCAGGCTTGAGAGCATA AATATTCCGGACGGTGTACCTCCCTTGATAATTCAGTGTTTCGATGGCTGTA CCACATTAAGACATAAAGATACCAGACAATGTGACCCTGATTGGGAACT ATACCTTTAATGCATGCAAGTCCTTGACAAACATAGTTATCCCTAATAGCGT AGAAAAGATCGCGAGCGGTGCCTTCTCCAATTGCGACAATCTGAAAGATGT GACCGTTTTAACTCCGCCGCGGAGATAGTTGATCGCGCATTCGGAGATGTT CAGGATCTCACAATCTATGGATATGTGGGATCTACAGCCGAAACGTATGCC AAGGAAAACAATATACCATTTGTGATATCAACTCTAAGGTGTTGACATCA AAGGTTAATAACATTTCTGTCAAGGGGACCTTCGACGACGGTGTAACGCTT AACGTCGAGGAGGCTACAGTGAAGAACGCCGTGGCTGCATACGACATAACT CTCAAAGATGAAAACGGAGACAACATACAACCTAACGGTACTATAATTGTC AGTATTCCATCAGACAATGCTAACTGCAAGGTGTATTGGATTAAGGAGGAT GGGACCAAAGTTGACATGAATGCGACTTATAATAATGGAAATTACGAATTC ACCACCGATCACCTCAGCATCTATGCGTTGATAGCTGACACAACGGAGTCTT CATCTGAGGAAAGTAATACCACTTcaggaacagacgggtccggc
GdCBM35 (WP_10998 5763)	GGAGCTTCGGAGACGGTTAACTGAAACGGGGAAACGAAGTAGACGAACTCA AGCTGGAAGCAGAAGAAGCTACGTTGTCCGGATTGACCGTACAAGACGACA TGCCTGGTTATTCCGGAACAGGCTATGTGGGAAACTTCACTGAAGACGATG CGAAGGTAACCTTCACTATGGATGTTCCAGATAAAGCTTTATACTAATAAC

TGTCGGGTACGGCGGTATTTACGGTGGTGGAAGCACGCCAATATACAATT
 GAATGGCGAAGCATTTACGAGTTTCGAACTCGGGGAGGGCTTCGGCGAAAG
 TCCTGTAGGAAAAGTCTTGTTGAATGAAGGGTCTAATACAATTGCGTTTACT
 CCGAACTGGACTCATTTGCAATCGATTTTATAAAGCTTTCTTTGGCGCAGC
 CAAGTATTGACCATGAAGTCAAAAAGCAACTTATTAATCCAAACGCTACGA
 CCGAAACAAAAACCTTATTTTCTTATCTTGTAGATAATTTTGGGAAGAATAT
 CATCTCAGGACAACAAGATGACGCAAATAACGACTTGGCTGATGTTAAGTA
 CATTAAAGAGCTGACCGGGAAAACCTCCGGCGATATTGGGCCTTGATCTTAT
 GGATTACTCACCTTCAAGAGTCGAACGCGGAGCCGTCTCCACGATGTCGA
 TTTGGCTCTTGAATGGAGCGAAAAAGGCGGTATAGTCGCCTTCGCGTGGCA
 TTGGAACGCTCCTAAAGATCTTTTGGATACGGAAGAGCAACCATGGTGGTC
 TGGTTTCTACACAGATGCGACTACATTCGATGTAGAGTATGCCATGAACCAC
 CCAGAGTCTGAGGACTATCAATTATTGATAAGAGACATAGACGCCATTGCG
 GAGCAGTTACGGCGTCTGCAGGAGGAAAACGTGCCAGTATTATGGCGGCCA
 CTGCATGAGGCTGAGGGAGGGTGGTTCTGGTGGGGAGCAAAGGGCCCAGA
 GCCAACCATTAAGCTGTGGAAGGTCATGTATGATCGCTTGACAAATTACCA
 CGAGCTTAATAACCTCATCTGGGTTTGAATAGTGTTGACCCGGCTTGGTAC
 CCAGGCGACGATTATGTTGACATTGTCAGTTTGTAGATTATCCTGGTGCGT
 ACAACTACAGCCCACAAAACAATAATTACGAAGCGTTGGTTGACCTTAGTT
 CTAATAAAAAGTTGATTGCGATGACCGAGAATGGGCCAATTCCTGACCCTG
 ATATGTTGGACCGGTACCATAGTATGTATAGTTACTTCACCACCTGGATTGG
 GCTGATCACCGAAAATAATAGTAAAGAACATTTGCAAAAGGTGTACAATCA
 CGAAAAGGTCATAACGCTGGAGGAGCTGCCGGACCTCACTACCTACGGGCA
 AGAAAAAAGTTCAGACACGTCCTTACAGAACCCTACCGTTAGCACAGGTCA
 GTTAAGCCCAGGCTTCAGCAAGAATGTAAAGGAATACACGTTGGAGGTCGA
 GAAAACCGTCGACGAGATTATCCTTTCCCGGTGGCGAACTCACATAGCGC
 AACAGTGACAATTAACGGTAACACTGATTTAACCAAGGGAATAAACTGCA
 AAGTGGAGATAATATAATTCATATAGAGGTTACGGTTGAAGACGGGACGAA
 ATCCAAATATTTGATACGGGTCAAGAGATTGACATCTTATCTTTCACTGGAC
 CTTGTAAAGAAGTCAGACCACTATGTATTGCCTGACAAATCCATCGAGAAT
 CTTGATATAAACGGCTTACTCGAATTGAAAATTCCGAAAAACACGGACCAC
 TTGCAAATTGTCTTCAACCACGAACAAATGAAATCACTGAAAGAAAAGAAT
 AATAAGATCCACCTGCGCAAAGACGATCTTCGCATGAATTACAACATGAAA
 CTGTTTCTTCCGACAGCCCGCTTACACTGAGTATAGAGAGAATAGAAGAG
 AAGGCGTTAGAGCATGGTGAACATGCAACCACAAAATTGTATGATCTTCGT
 TTTGAGCAAAACGATCGTGTCTCATCTCACAATTTAGTTCCTCAGTGCAGTTAA
 GCTTCCTTATGAAGGAGCCGAATGATAACCAGAAGATCTACTACTGGAATA
 TACAGCAAAGCGAATGGATTGGAATTGGAGGGAAGATCGAAAACAAGTGG
 ATTACCGTAGAGACGGATCATTTTACTATATTTTCCGTATTCGACGAAAAAG
 AATTGTCTCAAGCATCCGATCATGATAAACAAGTAGAGCACAAAACCTGATG
 CTAATGAGCAACCTTCCACCGACCAAAAAAAGGACAAAATGGTCAAAGAA
 GTTATAGGAATGGAACAGGGTAAGCTGACCACAGACGATTCA

Atcaggaacagacgg
tccggctggcg

Table S2. Geneblocks used in this study. Related to STAR Method.