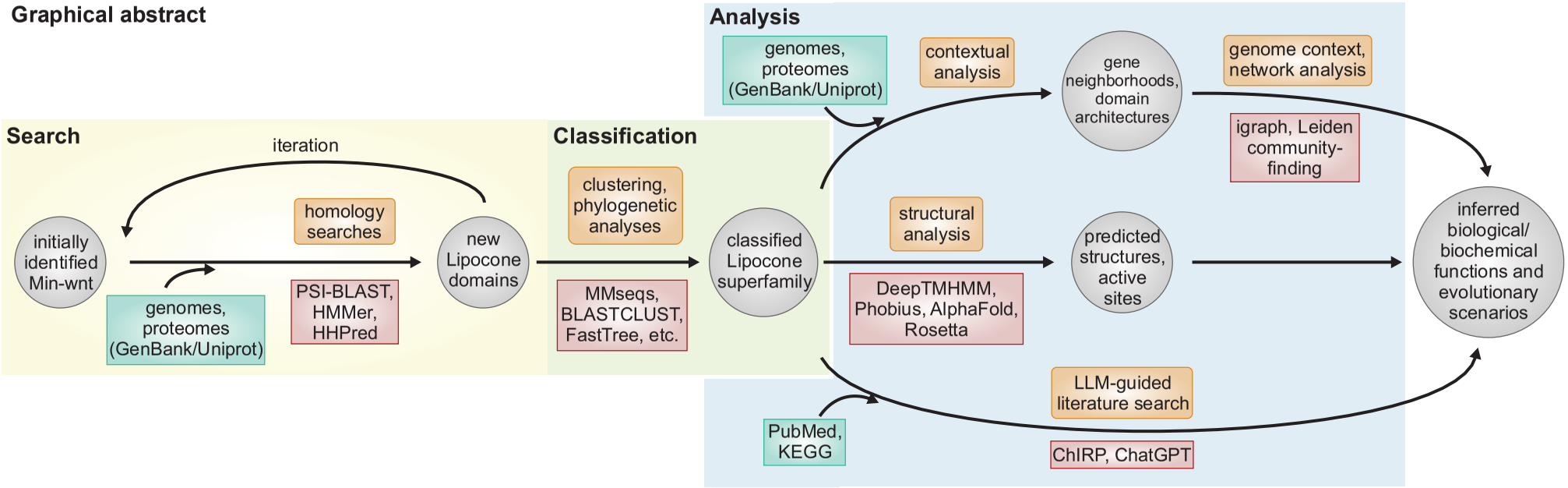
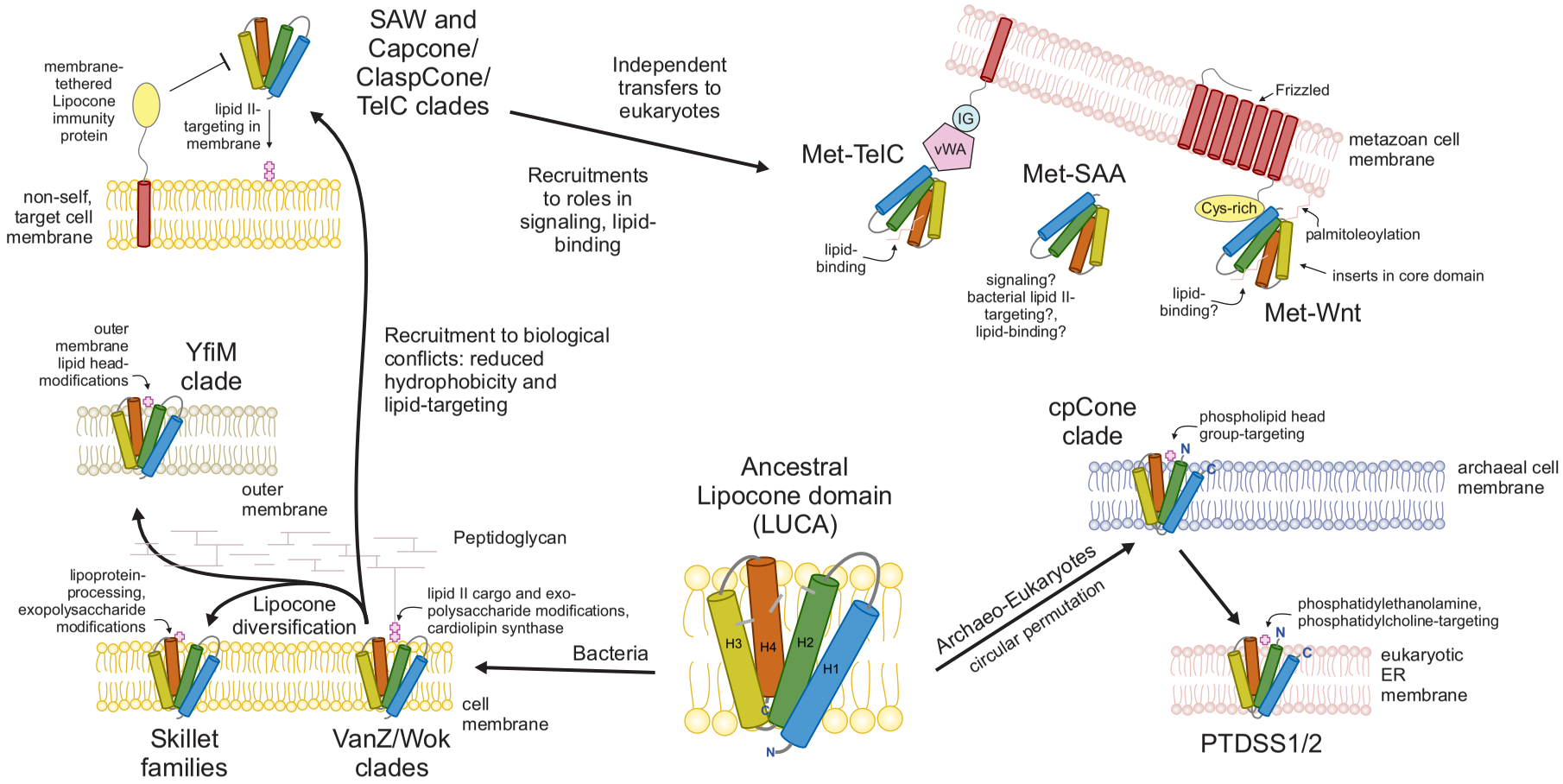
Graphical abstract





The Lipocone Superfamily: A Unifying Theme In Metabolism Of Lipids, Peptidoglycan And Exopolysaccharides, Inter-Organismal Conflicts And Immunity

A. Maxwell Burroughs†, Gianlucca G. Nicastro†, L Aravind*

Division of Intramural Research, National Library of Medicine, National Institutes of Health, USA

*To whom correspondence should be addressed. Tel: +1 (301)594-2445; Fax: +1(301)480-9241; Email: aravind@ncbi.nlm.nih.gov

†These authors contributed equally to this work

ABSTRACT

Wnt proteins are critical signaling molecules in developmental processes across animals. Despite intense study, their evolutionary roots have remained enigmatic. Using sensitive sequence analysis and structure modeling, we establish that the Wnts are part of a vast assemblage of domains, the Lipocone superfamily, defined here for the first time. It includes previously studied enzymatic domains like the phosphatidylserine synthases (PTDSS1/2) and the TelC toxin domain from *Streptococcus intermedius*, the enigmatic VanZ proteins, the animal Serum Amyloid A (SAA) and a further host of uncharacterized proteins in a total of 30 families. Although the metazoan Wnts are catalytically inactive, we present evidence for a conserved active site across this superfamily, versions of which are consistently predicted to operate on head groups of either phospholipids or polyisoprenoid lipids, catalyzing transesterification and phosphate-containing head group cleavage reactions. We argue that this superfamily originated as membrane proteins, with one branch (including Wnt and SAA) evolving into diffusible versions. By comprehensively analyzing contextual information networks derived from comparative genomics, we establish that they act in varied functional contexts, including regulation of membrane lipid composition, extracellular polysaccharide biosynthesis, and biogenesis of bacterial outer-membrane components, like lipopolysaccharides. On multiple occasions, members of this superfamily, including the bacterial progenitors of Wnt and SAA, have been recruited as effectors in biological conflicts spanning inter-organismal interactions and anti-viral immunity in both prokaryotes and eukaryotes. These findings establish a unifying theme in lipid biochemistry, explain the origins of Wnt signaling and provide new leads regarding immunity across the tree of life.

INTRODUCTION

The canonical Wnt signaling network is central to developmental decisions across animals relating to axis patterning, cell fate, cell migration and proliferation, and systems morphogenesis at many levels (1-7). Other crucial pathways, dubbed non-canonical Wnt signaling pathways, include those that regulate planar cell polarity and intracellular calcium levels (8-10). With these roles in development and homeostasis, dysfunction of Wnt signaling is causally associated with a range of diseases, including diverse cancer types and type II diabetes (11,12). Wnt signaling networks are centered on the secreted Wnt proteins acting as both paracrine and autocrine diffusible, extracellular messenger molecules (13). Wnt proteins are ligands for the N-terminal, cysteine-rich CBD/Fz domains of the Frizzled class of G-protein coupled receptors (GPCRs) (14,15). Modifications of the Wnt proteins via palmitoleoylation and glycosylation at internal sites are associated with their secretion (16). Palmitoleoylation of Wnt occurs at a conserved serine residue and is also required for recognition by the Frizzled receptors (17). Binding of the Frizzled receptor by Wnt recruits the Disheveled (Dsh) protein to its cytoplasmic face, in turn triggering a bevy of downstream responses, resulting in β-catenin stabilization in canonical pathways (18,19). When β-catenin concentrations cross a threshold, it is translocated into the nucleus, where it acts as a transcriptional coactivator, usually with an HMG domain transcription factor, to stimulate multiple transcriptional programs (20-22).

Despite its initial discovery over 40 years ago, the evolutionary origins of the Wnt protein have, until recently, been mysterious (23). In 2020, our group reported the discovery of the first prokaryotic versions of the Wnt domain (24). Using comparative genomics, we showed that these bacterial Wnt domains present contexts characteristic of toxins or effectors in biological conflict systems (24,25). Prompted by these initial observations, we set out to comprehensively identify and computationally characterize the evolutionary relationships of these newly identified Wnt homologs in an effort to understand their evolutionary history and predict their functions.

Consequently, we were able to unify the Wnt family with several other domains into a large superfamily described for the first time herein. These include two biochemically characterized families that were hitherto not known as Wnt homologs: the phosphatidylserine synthase (PTDSS1/2, EC: 2.7.8.29) (26-28) and the toxin domain of TelC from *Streptococcus intermedius* (29). However, the majority of the families we unify are either reported for the first time or are functionally poorly understood, including the animal Serum Amyloid A (SAA) (30) and the vancomycin resistance protein VanZ families (31,32). Our comparative genomics analyses, paired with existing experimental evidence, suggest that the superfamily is broadly comprised of enzymes operating on lipid head groups (e.g., transesterification reactions) in a diversity of biochemical contexts, notably including the regulation of membrane composition, extracellular biopolymer metabolism and as effectors in biological conflicts. Thus, we identify a unifying theme across diverse aspects of lipid metabolism.

RESULTS

Identification of the structural core of the Wnt domain

Although the structure of Wnt was described over a decade prior (33-35), its origins have been a mystery as it is phyletically restricted to Metazoa. Much attention has been focused on the three extended β-hairpins and a poorly structured loop extruding out of the core, their stabilizing cysteine residues, and the absolutely conserved serine residue, the site of palmitoleoylation (34,36) (Figure 1A). Our discovery of the first prokaryotic Wnt domains helped define its ancestral α-helical core, revealing the cysteine-rich extensions as Metazoa-specific insertions. Comparison of the core of the metazoan Wnt with AlphaFold structural models of the prokaryotic versions (24) revealed a shared globular domain composed of five α-helices (Figure 1A). As the prokaryotic homologs retained just the conserved core of the Wnt proteins, we named these the minimal Wnt (Min-Wnt) family. The core helices of the Min-Wnt family contained absolutely conserved sequence motifs (Figure 2), consistent with the enzymatic function we had earlier proposed for them (24) (see below).

Having defined this shared core, we initiated sequence-based homology searches in an effort to identify remote homologs. Iterative position-specific sequence matrix (PSSM)-based searches (see Methods) initially recovered animal and bacterial versions of the Serum Amyloid A (SAA) proteins, and further rounds of searching initiated from this set of sequences further recovered a vast collection of additional homologous families. As an example, a search initiated with a bacterial SAA-like sequence from *Bdellovibrio bacteriovorus* (Genbank acc: AHZ84906.1) retrieved a sequence overlapping with the Pfam models for "Domain of unknown function", DUF2279 (acc: WP_146898260.1, iteration: 5, e-value: 0.004), DUF4056 domain (acc: MBW8016507.1, iteration: 5, e-value: 0.005), and sequences automatically annotated as "YfiM" in the GenBank database (acc: WP_019077413.1, iteration: 4, e-value: 0.004). Sequence profile-profile searches with HHpred confirmed these relationships and captured more distant ones. For instance, a HHpred search initiated with the *Bacteriovorax stolpii* Min-Wnt domain (acc: WP_102242990.1, residues 1-109) recovered the Pfam Wnt profile (PF: PF00110.23, p-value: 1.5e-6) and the Pfam SAA profile (PF: PF00277.22, p-value: 3.7e-5). Similarly a HHpred search initiated with a Gemmatimonadetes sequence (acc: PYP94660.1, residues 75..170) recovered the DUF2279 Pfam profile (PF: PF10043.12, p-value: 5.4E-21), the DUF2238 profile (PF: PF09997.12, p-value: 1.3E-07), and the DUF4056 Pfam profile (PF: PF13265.9, p-value: 2.5E-05), among others.

Exhaustion of these searches, followed by clustering and manual inspection of the multiple sequence alignments of the retrieved sequences (see Methods), revealed a shared four-helix core across all of them, hereinafter referred to as H1 through H4 (Figure 1A). This 4-helix core of the domain was further confirmed by inspection of AlphaFold structural models constructed for representatives of the individual families, along with the rare instances of experimentally determined structures. These comparisons established that the above-mentioned fifth C-terminal helix in the Wnt core is a synapomorphy (shared derived character) restricted to the Wnts and closely related families like SAA (Figure 1A). In all, the results of our clustering analysis tallied 30 distinct families constituting a large superfamily. Remarkably, of these, 17 families had no pre-existing annotations. Phyletic analysis of individual families revealed a range of distributions, ranging from broad conservation in multiple superkingdoms of Life to those

118 restricted to a small number of lineages (see below, Figure S1). A relationship network for the
119 superfamily was constructed based on p-value and e-value scores using alignments of each
120 family as a query in HHalign profile-profile searches against the rest (see Methods, Figure 1B).
121 The Leiden community detection algorithm (37) was then applied to this network to identify
122 higher-order assemblages (see Methods). These groupings were also supported by structural
123 synapomorphies, such as a circular permutation and versions with a two-stranded 'handle' (see
124 below).
125
126 The four helices conserved across the superfamily constitute a cone-like structure (Figure 1A),
127 with the helices tending to coalesce on one end and opening out into a pocket on the other,
128 lined by the conserved sequence positions (Figures 1A, S2). The core is also marked by a linker
129 between H1 and H2, which adopts characteristic extended conformations in certain families and
130 higher-order groups. While the linkers joining H2 and H3 and H3 and H4 tend to be more
131 constrained, there are some exceptions; for example, the extended loop insert housing the
132 palmitoleoylated serine residue between H2 and H3 in the metazoan Wnt family (Figure 1A).
133
134 **Dramatic variability in hydrophobicity of the conserved core across the superfamily**
135 We observed that these Wnt-related families dramatically varied in their hydrophobicity. Using
136 an index for transmembrane propensity (38) (see Methods) and comparing that to known
137 transmembrane (TM) segments, we predict that the α-helices in 18 of the 30 families are
138 hydrophobic enough to qualify as TM domains, and show a statistically significant tendency to
139 group to the exclusion of the other families (Figure 1C, S3). Thus, these are predicted to be
140 integral membrane domains. Further, these 'hydrophobic families' often evince a broader and
141 deeper phyletic distribution pattern than the less-hydrophobic families (Figure S1, methods),
142 implying that the ancestral version of the superfamily was likely an integral membrane domain.
143 Thus, their association with the lipid membrane, combined with the cone-like shape of the
144 conserved core (Figure 1A), leads us to refer to the whole superfamily hereinafter as the
145 *Lipocone* superfamily.
146

147 Alphafold3-assisted transmembrane topology prediction (39) revealed that 14 of the 17 integral
148 membrane families are consistently oriented with the aperture of the cone-like structure
149 opening toward the outer face of the membrane. This predicted TM topology is also generally
150 consistent with the domain fusions when present: e.g., domains that are typically cytoplasmic
151 and those that have extracellular or periplasmic functions are respectively predicted as
152 projecting either inside or outside the membrane (see below). However, three families in the
153 cpCone clade (see below) did not yield consistent orientation predictions, potentially owing to
154 the diversity of structural variations observed in the clade, including a circular permutation event.
155

156 **A unified biochemistry for the Lipocone superfamily**

157 Of the 30 identified families, 26 display a striking conservation pattern of polar residues
158 associated with the pocket of the Lipocone domain (Figure 2, S2). Of these, a set of three
159 positions, one mapping to each of H2, H3, and H4, can be inferred as being ancestrally present
160 and were likely occupied by a histidine (H2), glutamate (H3), and aspartate (H4), though in some
161 families their identities have secondarily changed (Figures 2, S2). A fourth well-conserved polar
162 position is observed at or near the end of H3; while its ancestral identity is difficult to establish, it
163 is frequently an aspartate or glutamate (Figure 2). Two further well-conserved positions are
164 often seen in H4: a polar position downstream of the broadly conserved aspartate residue and a
165 glycine residue near the C-terminus of H4 (Figure 2) that likely caps the said helix. Although the
166 ancestral pattern is noticeably degraded in the metazoan Wnt (Met-Wnt) family, it is strongly
167 preserved in the prokaryotic Min-Wnt family (Figure 1A). In experimentally determined and
168 modeled structures, the above set of 4 conserved positions forms a predicted active site in the
169 aperture of the Lipocone domain. This, in turn, implies a shared biochemistry across the
170 superfamily, with secondary inactivation in some families like Met-Wnt (see below, Figure 2). At
171 the same time, the differences in the specific residues in the conserved positions between
172 different families point to a range of distinct but related activities across the superfamily (40-42).
173

174 Consistent with these observations, two of the families with intact active sites, the PTDSS1/2
175 (28,43) and TelC (29), which we identified in this work as members of the Lipocone superfamily,

176 have been characterized as active enzymes operating on different lipid substrates (Figure 3A).
177 The eukaryotic PTDSS1/2 localizes to the endoplasmic reticulum (ER) membrane and catalyzes a
178 reaction on the polar head group of phosphatidylcholine or phosphatidylethanolamine (44-46).
179 PTDSS1 and PTDSS2, respectively, exchange the phosphate-linked choline or ethanolamine head
180 groups with L-serine (28) (Figure 3A). The toxin domain of TelC acts on lipid II (29), the final
181 intermediate in peptidoglycan biosynthesis, which couples an undecaprenyl diphosphate tail to
182 a head group comprised of a N-acetylmuramic acid-N-acetylglucosamine disaccharide, with a
183 pentapeptide further linked to the former sugar (47,48). TelC cleaves the bond between the
184 undecaprenol and the diphosphate coupled to the head group (29) (Figure 3B). The reaction is
185 comparable to that catalyzed by PTDSS1/2, as both attack phosphate linkages in lipid head
186 groups. However, TelC apparently directs a water molecule for the attack in lieu of the hydroxyl
187 group of serine directed by PTDSS1/2 (Figure 3B).
188
189 Combining the above observations, we infer the unified biochemistry for the catalytically active
190 families thus: 1) They act on the head groups of lipids either by removing or swapping
191 phosphate-linked head groups (Figure 3A-B). These would be comparable to the phospholipase
192 D (PLD), transphosphatidylation or polyisoprenol phosphoesterase reactions (49). 2) Given the
193 cone-like cavity and the hydrophobicity of the helices, the lipid tail is predicted to be housed
194 within the lipocone with the head group positioned in the active site. 3) In the case of the
195 integral membrane versions, their orientation would predict the targeting of the head groups of
196 the outer leaf of the bilayer.
197
198 **Major clades of the Lipocone superfamily**
199 The extreme sequence divergence of the superfamily, coupled with the small size of the domain,
200 prevents the use of simple phylogenetic tree analyses to resolve its deep evolutionary history.
201 Hence, we combined community finding algorithms applied on profile-profile similarity
202 networks, comparison of structural features and motifs, and phyletic patterns (Figures 1B, 2, S1)
203 to reconstruct the most parsimonious evolutionary scenario for the diversification of the

204 Lipocone superfamily (Figure 4, see Methods). In the below sections, we survey the higher-order
205 clades, highlighting their specific features.
206
207 *SAW (SAA-Wnt) clade*
208 This clade consists of four families, with the two prokaryotic families (Min-Wnt and prok-SAA)
209 (24,50), respectively, giving rise to their counterpart eukaryotic families (Met-Wnt and Met-SAA;
210 Figures 1A,4). This clade is structurally unified by the presence of a fifth helix that stacks in the
211 space between the H2 and H4 helices (Figure 1A, S2). In the Wnt families, this helix is
212 comparable in length to the core helices, while in the SAA families it is usually shorter (Figure
213 1A). The clade is further unified by the pronounced conservation of a sNxxGR motif (where 's' is
214 a small residue) encompassing the conserved active site position in H4 (Figure 2). SAW clade
215 Lipocones show low overall hydrophobicity and are known or predicted to be soluble domains.
216 Outside of the clearly inactive eukaryotic Wnt family, the remaining three families largely
217 conserve the core active site residues (Figure 2).
218
219 *VanZ-Skillet clade*
220 This clade unites seven families: the two VanZ families, VanZ-1 and VanZ-2, prototyped by the
221 bacterial VanZ protein originally identified in the context of vancomycin resistance and the five
222 Skillet families, which form a distinct subclade. These are unified by a "handle"-like structure
223 (hence, "Skillet"), adopting a helical conformation in the H1-H2 linker (Figure 1A, S2). Strikingly,
224 a symmetric helical handle is present in the H3-H4 loop of the Skillet-DUF2809 and Skillet-3
225 families (Figure S2) of this clade. VanZ-1 features a conserved asparagine residue in the H2
226 position and a DxDDxxxN motif in H4, while VanZ-2 features RKxxH and DxxxD motifs in these
227 respective positions (Figure 2). The Skillet families are largely unifiable in their conservation of an
228 ExxQ motif in H3, an aspartate three positions upstream of the canonical H4 aspartate, and
229 another aspartate in the H2 contributing to the active site. These first two features specifically
230 ally them with the VanZ-1 family (Figure 2).
231

232 While the VanZ domain was previously reported as including a fifth TM helix, which is C-terminal
233 to the 4-helix Lipocone core defined here (51,52), our survey instead reveals a striking diversity
234 of configurations around the core 4-helix Lipocone domain (Supplemental Material). These
235 range from standalone Lipocone configurations to one or more TM-helices adorning the
236 domain at its N- and/or C-terminus. This variation is consistent with a further tendency for the
237 VanZ families to feature an extensive diversity of domain fusions to both soluble globular
238 domains and discrete TM modules (see below).
239
240 The VanZ families are deep-branching, as suggested by their wide phyletic spread (Figure S1).
241 VanZ-2 is the most widespread individual Lipocone family in bacteria, with several genomes
242 encoding multiple paralogs (Supplemental Material) (51,53). It is also found in certain eukaryotes,
243 including a pan-fungal presence and in some representatives of the SAR clade. Both VanZ-1 and
244 VanZ-2 are particularly well-represented in Gram-positive bacterial lineages like Actinomycetota
245 and Firmicutes, while VanZ-2 is nearly universally conserved in the Bacteriodetes/Chlorobi
246 lineage (Figure S1). In contrast, only one of the Skillet families, Skillet-DUF2809, is widely but
247 sporadically distributed, with the four others being more restricted (Figures 4, S1).
248
249 *YfiM clade*
250 This clade includes three families that are consistently centrally located in the profile-profile
251 similarity network (Figure 1B). This is likely due to their being close in sequence conservation to
252 the ancestral state of the superfamily (Figure 2). Consistent with this, the YfiM-1 family also
253 presents a structurally minimal Lipocone domain, comprised of just the 4-helix configuration
254 with no further elaborations. Notably, this also extends to a lack of domain fusions in this family.
255 In contrast, YfiM-DUF2279 and YfiM-Griddle (DUF3943) are structurally distinguished by an
256 unusual H1-H2 linker (Figure 1A), which wraps around the outside and stacks against the H3-H4
257 linker (Figure S2). The YfiM-Griddle family further features a unique 'flattened' surface around
258 the aperture of the Lipocone formed by protruding loops (hence, "Griddle"; Figure S2). This
259 leaves the active site pocket more accessible relative to families with more elaborately
260 structured inter-helix linkers. The Griddle family also features a C-terminal extension with a two-

261 helix hairpin (with a hhsP motif in the turn between the two helices, where 'h' is a hydrophobic
262 residue and 's' is a small residue) (Figure S2). The three YfiM families straddle the membrane-
263 propensity boundary in the plot (Figure 1C). Further, the YfiM-DUF2279 and Griddle families are
264 strikingly absent in Gram-positive bacterial lineages (Figure S1). Concurrent with these features,
265 they are often predicted by the deep-learning-based localization predictor deepTMHMM as
266 outer-membrane proteins, suggesting a role in this subcellular location (see below).
267
268 *ClaspCone-CapCone-TelC clade*
269 Members of this clade are unified by an elaborated H1-H2 linker that often contains one or
270 more helical segments that are typically predicted to guard the aperture of the Lipocone domain
271 (Figure S2). This linker ends in a "clasp"-like element, which forms a range of structures in
272 different families of the clade before leading into H2 (Figure S2). The clade is also unified by a
273 striking reduction of overall hydrophobicity, predicting that the members of this clade are
274 soluble domains (Figure 1C). Outside of the divergent TelC subclade, most of the families in this
275 clade conserve a serine residue three positions upstream of the active site aspartate in H4, often
276 preceded by an aromatic residue, which is typically phenylalanine. H4 also usually features a
277 conserved asparagine four positions downstream of the conserved aspartate active site position,
278 immediately preceded by a small residue (Figure 2). The second H3 active site position is
279 generally poorly conserved, though when present, it is usually an aspartate residue. Finally, H2
280 contains either a DK or xD motif four positions upstream of the canonical H2 active site histidine
281 residue (Figure 2).
282
283 The most rudimentary clasps are found in the ClaspCone-1, -2, and -3 families, where it is little
284 more than a rounded loop, though, in ClaspCone-1, a small β-hairpin emerges within it. The
285 three ClaspCone families are further unified by the presence of a two-helix insert leading into H2
286 that stacks against the Lipocone core (Figure S2). The three CapCone families, CapCone-
287 DUF4056, CapCone-1, and CapCone-2, are named so for an encasing structure over the active
288 site resembling a cap (Figure S2). They share a conserved glycine residue six positions upstream
289 of the active site H2 histidine and a S/GxxSxx motif upstream of the conserved H4 aspartate

290 (Figure 2). They are further unified by a pronounced β-hairpin clasp augmented by an additional
291 strand (Figure S2). They also display varying degrees of degeneration of H1, along with family-
292 specific structural elaborations.
293
294 The TelC group of this clade, prototyped by the streptococcal TelC toxin (29), is divided into two
295 families featuring prokaryotic (prok-TelC) (29) and metazoan versions (Met-TelC) (54). Both TelC
296 families feature a "cap" with contributions from inserts in the H1-H2 and H3-H4 loops (Figures
297 1A,S2). Unique to these families is the conservation of an aspartate residue located six positions
298 downstream of the canonical active site aspartate of H4 (Figure 2). This aspartate points away
299 from the center of the Lipocone and interacts with a conserved arginine from a synapomorphic
300 C-terminal helical extension.
301
302 *cpCone clade*
303 A widespread yet sporadically distributed clade of seven families emerging as a stable
304 community in the profile-profile similarity network (Figure 1B) is defined by a unique structural
305 synapomorphy: a circular permutation (55) (hence, cpCone) placing the normally N-terminal H1
306 at the C-terminus of H4 (Figure S2, S4). This clade is also united by unique sequence features,
307 viz., a polar residue (typically aspartate) six positions upstream of the conserved H2 histidine and
308 a second glutamate three positions downstream of the conserved H3 glutamate (Figure 2).
309 While the circular permutation is shared across the clade, several structural variations are seen,
310 often within the same family (Figure S4). These include: 1) versions containing a duplication of
311 the Lipocone domain. While the second copy in these versions is catalytically inactive, the H1'
312 from the second duplicate displaces the H1 from the first copy, suggestive of an intermediate to
313 the circular permutation. 2) Versions retaining a candidate H1 that has been displaced by H1' in
314 a five-helix arrangement. 3) Those containing just the circularly permuted core. 4) Versions
315 showing a degradation of the H1 helix, preserving just a 3-helix core (Figure S4). Despite this
316 propensity for structural variation, the active site residues are strongly conserved, with the
317 exception of the cpCONE-i family, which we infer to be catalytically inactive (Figure 2). The core
318 helices of the cpCone clade are strongly hydrophobic, and they are all predicted to be integral

319 membrane domains (Figure 1C). Consistent with this, the eukaryotic PTSSD1/2 domains reside in
320 the ER membrane (44,45).
321
322 *Wok family*
323 The Wok family (partly covered by the Pfam DUF2238 model) shows a higher order grouping
324 with the above circularly permuted clade (Figures 1B,4) but has a phyletic distribution only
325 rivaled by the VanZ-2 family (Figure S1), suggesting a deep-branching origin. The shape of this
326 family is reminiscent of a wok formed by two distinguishing structural synapomorphies: a 2-TM
327 helix N-terminal extension and a unique "handle" formed by the linker between the H3 and H4
328 (Figure S2). It additionally features a C-terminal, rapidly diversifying cytoplasmic tail. Despite
329 these elaborations, it retains the inferred ancestral active site configuration (Figure 2). The
330 strongly hydrophobic core helices of the Wok family predict it to be an integral membrane
331 enzyme (Figure 1C).
332
333 **Functional themes in the Lipocone superfamily**
334 Given our inference of shared general biochemistry across the Lipocone superfamily in targeting
335 phosphate-containing linkages in head groups of both classic phospholipids and polyisoprenoid
336 lipids, we next used contextual information from conserved gene-neighborhoods, domain
337 architectures and phyletic pattern vectors, a powerful means of deciphering gene function (56),
338 to narrow down the predictions for specific families (Figure 5, Table S1, Supplementary Data). To
339 this end, we constructed a graph (network) wherein the nodes are individual domains and edges
340 indicate adjacency in domain architectures or conserved gene-neighborhoods (Figure 6, see
341 Methods). We then identified cliques in these networks and merged the individual cliques
342 containing a particular Lipocone domain to define its dense subgraph (Figures S5-S7). We then
343 analyzed these subgraphs to identify statistically significant functional categories represented in
344 them (Table S2; see Methods). This data was combined with existing experimental results and
345 the sequence and structure analyses outlined above to arrive at the functional themes surveyed
346 in the below sections.
347

348 *Lipocone domains in membrane lipid, peptidoglycan and exopolysaccharide modifications*
349 Across different Lipocone families, we found statistically significant connections to roles in
350 modifying lipid head groups in various membranes and in lipids involved in the synthesis of
351 extracellular matrix polymers such as peptidoglycan and lipopolysaccharides (Figure 6, Table S2,
352 Supplementary Data).
353
354 *Archetypal lipid head group exchange reactions catalyzed by the cpCone clade*
355 One of the few experimentally characterized Lipocone families is the eukaryotic PTDSS1/2 family
356 of the cpCone clade, members of which exchange the head group of essential membrane
357 phospholipids to generate phosphatidylserine from phosphatidylethanolamine or
358 phosphatidylcholine (Figure 3A) (28,57). Given the pervasive presence of this clade in archaea
359 (Figure S1), it is thus tempting to speculate that these archaeal cpCones may play a role in the
360 modification of Archaea-specific lipids (58-60) through a comparable head group exchange
361 reaction (see below).
362
363 In bacteria, the related cpCone-1 family shows operonic association with a LolA-like lipoprotein
364 which shuttles lipoproteins to the outer membrane (61) and a novel 4TM protein (Figure 5A).
365 This raises the possibility that cpCone-1 might mediate the formation of membrane domains
366 featuring lipids with a modified head group that act as foci for the trafficking of lipoproteins.
367 Curiously, the cpCone-1 gene might also be inserted between the bacterial chromosome
368 segregation and condensation complex subunits the Kleisin ScpA and the wHTH ScpB (62-65).
369 The bacterial cpCONE-DUF2585 is operonically coupled to a GNAT family NH_2-group-
370 acetyltransferase and further linked to genes for the glycolate oxidase GlcE and GlcF (66) and
371 the bacterial proteasome subunits HslV and HslU (67) (Figure 5A, Supplementary Data). These
372 might point to the coupling of membrane lipid head group modifications with disparate
373 processes, such as chromosome segregation during cell division or different responses to stress
374 (68-70).
375
376 *The Wok and YfiM-1 families in cardiolipin and modified isoprenoid lipid pathways*

377 We observed a set of conserved gene neighborhoods displaying the mutually exclusive presence
378 of a synaptojanin-like phosphatase gene, with one encoding either a member of the Wok family
379 or a cardiolipin synthase of the HKD superfamily (71) (Figure 5B, Supplementary Data). This
380 suggested that the latter two are analogous enzymes catalyzing equivalent reactions. The
381 cardiolipin synthase utilizes two phosphatidyl glycerol molecules as substrates to generate
382 cardiolipin with the release of one of the glycerol head groups (72). This is comparable to the
383 head group exchange reaction catalyzed by PTDSS1/2 from the cpCone clade (Figure 3A). Hence,
384 we propose that these members of the Wok clade are cardiolipin synthases (Figure 3C). Distinct
385 phosphoesterases, namely the synaptojanin-like, calcineurin-like (73) and HAD (74) enzymes, are
386 also observed in gene-neighborhood associations with the Wok, suggesting that they might
387 together regulate membrane lipid composition by acting on the phospholipids or their
388 precursors (Figure 5C). In a distinct neighborhood, the Wok clade enzyme is coupled to
389 carotenoid biosynthesis genes (75,76). (Figures 5D,S5). This raises the possibility that these
390 members might also catalyze a comparable reaction to the above on isoprenoid lipids: for
391 instance, they could synthesize a carotenoid from two geranylgeranyl-diphosphate molecules
392 (77,78). In both these contexts, the actinobacterial operons often include genes for GT-A family
393 glycosyltransferases, suggesting the further synthesis of glycosylated derivatives of the lipids or
394 carotenoids (79) (Supplementary Data). In several bacteria, a YfiM-1 family Lipocone is
395 operonically coupled to a UbiA-like prenyltransferase (80). This gene neighborhood additionally
396 codes for a slew of enzymes, such as an amidophosphoribosyltransferase (81), a RidA-like
397 deaminase (82), and a pair of structurally distinct phosphoesterases, respectively, containing an
398 HD and a PHP domain (73,83) (Figures 5E, S5 Supplementary Data). This suggests a role for the
399 YfiM-1 Lipocone and the associated enzymes in generating a modified polyisoprenoid
400 metabolite.
401
402 *VanZ families modifying lipid head groups in peptidoglycan and exopolysaccharide metabolism*
403 The widespread VanZ-1 and VanZ-2 families (Figure 1A) frequently show either gene
404 neighborhood associations or direct domain fusions, with diverse genes involved in both
405 peptidoglycan and other extracellular polysaccharide pathways. Chief among these are the lipid

406 carrier flippase (Pfam: MviN_MATE clan) (84-86), the UDP-GlcNAc/MurNAc lipid transferases,
407 which generate the lipid-linked exopolysaccharide precursors (lipid I) (48,87), and UDP-N-
408 acetylglucosamine (UDP-GlcNAc) biosynthesis enzymes (88,89). Despite certain examples of
409 crossover in functional themes, the gene-neighborhood contexts of VanZ-1 and VanZ-2 suggest
410 a metabolic partitioning, with VanZ-2 significantly associating specifically with peptidoglycan-
411 related genes and VanZ-1 significantly linking with biosynthesis genes for other
412 exopolysaccharides (e.g., the outer-membrane-associated lipopolysaccharide) (90) (Figures
413 5F,6,S6, Table S2). The latter include WaaL-like lipid A transferase (91), the polysaccharide chain-
414 length determination domain Wzz (92), the Wzc kinase and the "extracellular antigen"-
415 regulating ElyC-like domain (Pfam: DUF218) (93), and numerous nucleotide-diphosphate sugar
416 biosynthesis and modification enzymes (94) (Figures 5F,6,S6).
417
418 The precursors of both peptidoglycan and exopolysaccharides are synthesized in the cytosol,
419 linked to lipid carriers via a diphosphate linkage, e.g., the polyisoprenoid lipid undecaprenol
420 (bactoprenol) (90,94-97). A key step in their maturation is the flipping by the flippase of the
421 lipid-linked intermediates associated with the inner membrane to the outer membrane. These
422 flipped units are then incorporated into the maturing chain (98,99) by the peptidoglycan
423 glycosyltransferase (GTase) (100) and the chain length determination protein, WzzE/polymerase
424 (WzyE) (92,101), in peptidoglycan and other exopolysaccharide maturation pathways,
425 respectively. Based on the precedence of the TelC-catalyzed reaction (Figure 3B), we propose
426 that VanZ-1 and VanZ-2 comparably act on the flipped lipid II head groups bearing the modified
427 sugar intermediates to release the undecaprenol via phosphoester cleavage (Figure 3F). Such
428 activity could modulate the concentration of available peptidoglycan intermediates and allow
429 formation of peptidoglycan with varying thickness and composition during different phases of
430 the life cycle, e.g., sporulation versus vegetative growth in Bacillota. Such a reaction could also
431 possibly modulate exopolysaccharide biosynthesis by comparably acting on their precursors.
432
433 The terminal transfer from the lipid carrier of the Gram-negative bacterial O-antigen (as well as
434 other exopolysaccharides attached to the lipid A carrier) has been attributed to the WaaL-like

435 enzymes (91,102). However, bacteria generate further lineage-specific polysaccharide
436 decorations, capsule structures, and other exopolysaccharides (e.g., xanthan, enterobacterial
437 common antigen (ECA), alginate, colonic acid), as well as teichoic acids (e.g., wall teichoic acids,
438 WTA) (103, 104). Notably, the analogs of WaaL, i.e., the terminal transferases for several
439 exopolysaccharides, including ECA and WTA, have to date escaped identification (93). Hence, it
440 is possible that, by analogy to the PTDSS1/2 reaction (Figure 3A), the VanZ families act on the
441 lipid carrier-linked sugar head groups to catalyze either the extension of the polysaccharide
442 chains through transesterification or the terminal release of the mature chain through
443 phosphoester cleavage (Figure 3E).
444
445 *Atypical VanZ domains in uncharacterized modifications of peptidoglycan and the outer*
446 *membrane*
447 Certain representatives of the two VanZ families also show operonic associations indicative of
448 outer membrane-associated or peptidoglycan modification functions distinct from those
449 described above (Figures 5G,6 Supplementary Data): 1) An operon in FCB group bacteria
450 couples a VanZ-2 gene with those coding for a SprA secretin-like channel protein (105), a
451 glycine cleavage H (GCVH)-like lipoyl-group carrier protein (106), a 2TM protein fused via a
452 proline-rich linker to a C-terminal TonB-C domain (107), and a secreted, second TonB-C domain
453 fused to a Wzi-like outer membrane protein (OMP) superfamily β-barrel (108) (Figures 5G, 6).
454 2) In betaproteobacteria, certain VanZ-1 domains are duplicated with the C-terminal copy being
455 inactive (VanZ-i) and found in an unusual four-gene operon with a thioredoxin-fold [2Fe-2S]
456 ferredoxin (109), a possible lipase of the α/β-hydrolase superfamily (110), and a metallo-β-
457 lactamase (MβL) fold D-Ala-D-Ala cross-linking transpeptidase (111,112). 3) A patescibacterial
458 operon encodes a VanZ-2 domain with an ABC ATPase transporter system, either of two
459 structurally distinct peptidases, namely a Papain-like or glycine-glycine peptidase (113,114),
460 fused to the same membrane-anchored N-terminal coiled-coil region, and a further TM protein
461 containing one or more external Lamin-Tail domains (LTDs) predicted to bind extracellular DNA
462 or polysaccharides (115) (Figures 5G,6,S6). The associations in the first of the above
463 neighborhoods point to a distinct outer membrane-associated lipid modification, while the

464 other two might be involved in lineage-specific decorations/modifications of peptidoglycan,
465 accompanied by peptide-crosslinking or cleavage activities.
466
467 *Lipocone domains operating in the outer membrane*
468 Contextual associations, phyletic patterns, and localization predictions support the action of two
469 Lipocone families directly in the outer membrane. Notably, the YfiM-Griddle and YfiM-DUF2279
470 families are found nearly obligately directly fused or operonically linked to several distinct OMP
471 β-barrels (116,117) (Figures 5H,6,S5). Up to three YfiM-Griddle Lipocones, usually with a cognate
472 OMP β-barrel, might be encoded next to each other in the genome. Additionally, YfiM-Griddle
473 family genes are often encoded in operons with several components of the outer membrane
474 lipid and protein trafficking apparatus, including the LolA-like chaperone (118), the POTRA
475 domain (119,120), the channel-blocking Plug domains (121), and the TolA-binding TolB-N
476 domain (122). Further, these operons might encode a Patatin-like lipase (123), GT-B family
477 glycosyltransferases (79), and a range of phosphoesterases (e.g., an integral membrane
478 phosphatidic acid phosphatase PAP2 (124), a lipobox-containing synaptojanin superfamily
479 phosphoesterase (125) and a secreted R-P phosphatase (126) (see Figures 5H,6, and
480 Supplementary Data)). In addition to the fusion to the OMP β-barrel, the YfiM-DUF2279 family
481 (Figure 5H) shows operonic associations with a secreted MltG-like peptidoglycan lytic
482 transglycosylase (127,128), a lipid-anchored cytochrome c heme-binding domain (129), a
483 phosphoglucomutase/phosphomannomutase enzyme (130), a GNAT acyltransferase (131), a
484 diaminopimelate (DAP) epimerase (132), and a lysozyme-like enzyme (133). In a distinct operon,
485 YfiM-DUF2279 is combined with a GT-A glycosyltransferase domain (79), a further OMP β-barrel,
486 and a secreted PDZ-like domain fused to a ClpP-like serine protease (134,135) (Figure 5H).
487
488 The strong linkage to the OMP β-barrel, together with their predicted localization, suggests that
489 these YfiM-Griddle and YfiM-DUF2279 Lipocone domains operate in the outer membrane,
490 potentially in concert with both cytoplasmic carbohydrate biosynthetic modules and periplasmic
491 lipid- and carbohydrate-processing enzymes. As with the inner membrane lipids, they could
492 potentially catalyze modifications of head groups through transesterification and/or

493 linkage/release of outer membrane-associated polysaccharide chains through action on lipid-
494 head group phosphoesters.
495
496 *Lipocone domains acting on lipids in transit to the outer membrane*
497 The ClaspCone-1 and ClaspCone-3 families lack the hydrophobicity indicative of direct residence
498 in the membrane (Figure 1C); instead, they are predicted to localize to the periplasmic space. In
499 the ClaspCone-1 family, the Lipocone domain is fused at the extreme N-terminus to either a
500 single TM or a 5TM domain predicted to anchor it to the cell membrane. Between this TM
501 element and the Lipocone domain, we detected a previously uncharacterized version of the
502 Tubular lipid binding protein (TULIP) domain (136,137) or an Ig-like and a Zincin-like
503 metallopeptidase (MPTase) domain (138) (Figures 5I, 6). These ClaspCone-1 genes may also
504 show operonic associations with genes encoding a lipase of the SGNH family (139) and a
505 membrane-bound O-acyltransferase (MBOAT; Figure 5I, Supplementary Data) (140). The TULIP
506 domain superfamily has recently been characterized as a lipid-binding domain (136,137), which
507 in proteobacteria functions in outer membrane lipid transport (141,142). Thus, we propose that
508 the ClaspCone-1 family is likely to act in the periplasmic space on the head groups of outer-
509 membrane targeted lipids bound to the TULIP or potentially to the Ig-like domains occupying
510 an equivalent position in the domain architecture.
511
512 *A Lipocone domain catalyzing a predicted lipoprotein lipid linkage reaction*
513 The Skillet-1 Lipocone is strongly coupled in an operon with a downstream gene coding for a
514 protein with an unusual lipobox-like sequence followed by one of several extracellular domains
515 (e.g., concanavalin, β-jelly roll, OB-fold, Ig-like, β-propeller) predicted to bind carbohydrates or
516 other ligands (143-147) (Figures 5J,6,S6 Supplementary Data). The lipobox-like sequence
517 features a conserved GS motif at its C-terminus instead of the usual GC of the classic lipobox of
518 bacterial lipoproteins (148) (Figure S8). In the canonical lipoprotein processing pathway, a
519 thioether linkage is formed between the sulfhydryl of the cysteine and a diacylglycerol lipid
520 embedded in the inner membrane by the lipoprotein diacylglyceryl transferase (lgt) enzyme,
521 followed by the cleavage of the signal peptide at the GC motif junction by the signal peptidase

522 (149,150). Given the serine in place of the cysteine in these lipobox-like sequences, we propose
523 that it undergoes non-canonical lipidation by the associated Skillet-1 Lipocone protein in lieu of
524 lgt. We propose that, comparable to PTDSS1/2, which act on free serine, the Skillet-1 family links
525 the conserved serine from the lipobox-like sequence to a phospholipid (Figure 3A,D).
526
527 ***Lipocone domains in predicted lipid-associated signaling systems***
528 *Systems defined by standalone proteins with Lipocone domains*
529 Several representatives of the two VanZ and Skillet-3 families are fused to a diverse array of
530 known or predicted extracellular ligand-binding domains (Figure 5K), where the architecture
531 takes the form of SP+X+TM+Lipocone or Lipocone+TM+X, where 'X' is the extracellular ligand-
532 binding domain and SP is a signal peptide. The ligand binding domains include: (i)
533 carbohydrate-binding lectin domains such as jelly-roll, concanavalin-like, NPCBM-like, CBD9-like,
534 and other β-sandwiches (143,144,151-153)); (ii) a lipid-binding helix-grip superfamily domain
535 (154)); (iii) those binding other potential ligands (e.g., Ig, OB-fold, YycI-like, DUF498-like, PepSY-
536 like, β-helix, TPR, MORN, and β-propeller repeats (145-147,153,155-158)) (Figure 5K,
537 Supplementary Data). We interpret these architectures as implying signaling, wherein the
538 binding of the cognate ligand by one of the above domains regulates the catalytic activity of the
539 associated Lipocone domain. Among these, the extracellular domains fused to the Skillet-3
540 family are particularly notable for their extreme variability (Figure 5K). This suggests their
541 diversification under an arms race scenario (also see below) in a biological conflict. Further, the
542 genes coding for the above are sporadically associated with exopolysaccharide metabolism
543 genes (Supplementary Data). Hence, it is conceivable that this signaling is associated with
544 exopolysaccharide variation (e.g., O-antigen phase-variation (159,160)), which might play a role
545 in evading bacteriophage attachment.
546
547 Additionally, VanZ-1 Lipocone domains are also fused to several known signaling domains
548 confidently predicted to reside in the cytoplasm, including the cyclic nucleotide-binding domain
549 (cNMPBD), phosphopeptide-binding FHA, and DNA-binding RHH and HTH domains (65,161-
550 163) (Figure 5K). These associations suggest potential VanZ regulation via a cytoplasmic cyclic

551 nucleotide (sensed by cNMPBD) or, conversely, VanZ acting as an allosteric regulator of a
552 transcriptional program via the HTH or RHH domain. One of the most common yet enigmatic
553 fusions to VanZ is with the integral membrane RDD domain (53). The role of this domain is
554 unknown; however, our analysis indicates that it contains a conserved intra-membrane binding
555 site oriented towards the cytoplasmic face of the membrane (Nicastro GN, Burroughs AM,
556 Aravind L, manuscript in preparation). The VanZ-RDD fusion is sometimes further fused to other
557 domains (Figure 5K), the most notable being a highly derived but active novel Histidine Kinase
558 domain (Figure 5K). Together, these associations point to the coupling of lipid modification with
559 a signaling event on the cytoplasmic face of the membrane, which might relate to the dynamic
560 regulation of lipid-carrier-bound exopolysaccharide precursors.
561
562 *Multi-component associations of the Lipocone proteins in signaling*
563 These systems resemble the above-discussed versions but are encoded by conserved gene
564 neighborhoods that separate the Lipocone and the signaling elements (typically predicted
565 transcription regulators) into distinct genes. Our analysis recovered at least three such systems:
566 1) A VanZ-1 Lipocone in the recently described HAAS/PadR-HTH two-component systems,
567 which sometimes replace classical Histidine kinase-Receiver two-component systems (164). In
568 these systems, the detection of an extracellular or intramembrane stimulus by a sensor domain
569 releases the PadR-HTH transcription regulator bound to the sensor-fused HAAS domain. Here,
570 VanZ-1 occupies the sensor position (Figure 5L). 2) A Skillet-2 Lipocone is coupled in a core
571 two-gene system to a conserved upstream gene (Figure 5L,S5). That gene encodes a single TM
572 protein with either a zinc ribbon (ZnR) fused to a conserved helix or an HTH domain fused to a
573 ClpS-ribosomal L7/L12 domain in its cytoplasmic region (165). These neighborhoods might also
574 code for an HMG-CoA reductase and GHMP kinase that catalyze successive reactions in the
575 production of phosphomevalonate, a precursor of isoprenoid lipids (166,167). 3) A Skillet-
576 DUF2809 Lipocone protein is operonically coupled with a 6TM protein and a further predicted
577 transcription factor with a wHTH protein. These operons are further elaborated via additional
578 embedded and flanking genes, either coding for components of isoprenoid lipid (e.g.,

579 undecaprenol) (168,169) or exopolysaccharide (e.g., ECA and related polysaccharides)
580 metabolism (94,97) (Figures 5L,6,S6 Supplementary Data).
581
582 The Lipocone domains in these systems are predicted to be active enzymes, which, together
583 with their operonic associations, point to functions involving the modification or
584 transesterification of isoprenoid lipid head groups, sometimes in the context of
585 exopolysaccharide biosynthesis. However, their associations with the intracellular HTH domains
586 suggest that the Lipocone enzymatic activity is potentially coupled with the transcriptional
587 regulation of the production of precursors of the lipids or exopolysaccharides. Given the high
588 variability in the associated genes related to exopolysaccharide/lipopolysaccharide biosynthesis,
589 we anticipate that the associated transcriptional regulation potentially relates to functional
590 categories showing high diversity across bacteria, such as responses to environmental stress,
591 phages, predatory bacteria attacks, or host immune response.
592
593 ***Lipocone domains as effectors in biological conflicts***
594 *Lipocone domains in antiviral immunity*
595 The Min-Wnt domains (Figure 1A) that we originally identified were predicted to play a role in
596 biological conflicts with invasive selfish elements, such as viruses (24). In this work, we better
597 explain their potential mechanism of action. These versions show no fusions to extracellular
598 domains or secretory signals, suggesting that they are deployed from within the bacterial cell
599 (Figure 1C). These Min-Wnts are typically fused to the DUF3892, which displays a fold
600 characterized by a three-stranded meander followed by a helix also seen in the dsRNA-binding
601 domain and the ribosome hibernation factors (HPF) (170,171) (Figure 5M). Hence, we propose
602 that these versions might potentially act to sense virally induced RNAs or modified ribosomes
603 (24) to trigger a dormancy or suicide response to limit viral infection via the Min-Wnt effector.
604 Specifically, the Min-Wnt might attack peptidoglycan precursors, such as lipid II, prior to their
605 'flipping' to restrict cell wall synthesis (85,172,173) or other such carrier lipids.
606

607 One other Min-Wnt domain, N-terminally fused to a three-stranded β-meander, is pervasive in
608 the Bacteroidetes clade. This is operonically coupled with genes encoding a TM-linked run of
609 pentapeptide repeats and two structurally distinct, secreted glycosyl hydrolase enzymes,
610 respectively, containing a TIM barrel domain and a run of β-helix repeats (Figure 5M,
611 Supplementary Data). Further, cyanobacteria show a standalone prok-TelC domain without any
612 secretory signals. These could again act as effectors targeting lipid-linked precursors of
613 peptidoglycan or expolysaccharides in response to intracellular invaders or stress (Figure 5M).
614 Interestingly, some tailed bacteriophages also code for intracellular Min-Wnt domains,
615 suggesting that they might also be deployed on the virus side in biological conflicts such
616 limiting superinfection (Supplementary Data).
617
618 *Lipocones as toxin domains in polymorphic and allied conflict systems*
619 Polymorphic toxins and related systems, widespread across bacteria and certain archaea, are
620 characterized by a highly variable C-terminal toxin domain ("toxin tip") that is preceded by a
621 range of more conserved domains typically required for autoproteolytic processing of the toxin,
622 its packaging and trafficking (e.g., RHS repeats), adhesion and secretion via one of several
623 secretory systems (174,175). The toxin might be delivered via one of the secretory systems into a
624 target cell or else via direct contact between interacting cells. Classical polymorphic toxins are
625 usually involved in kin discrimination and are accompanied by genomically linked cognate
626 immunity proteins that protect against self-intoxication (174,176). Keeping with the principle of
627 effector sharing between systems involved in distinct types of biological conflicts, we had
628 originally identified a Min-Wnt domain closely related to those described in the above
629 subsection as a toxin tip in polymorphic toxin systems (24). In the current work, we extend these
630 findings to show that several distinct Lipocone families have been independently recruited as
631 toxin tips of polymorphic toxins and related systems, namely Min-Wnt, prok-SAA, prok-TelC,
632 CapCone-1, CapCone-2, ClaspCone-2 and VanZ-1 (Figures 4,5N,6,S7).
633
634 Certain CapCone-2 and Min-Wnt toxins from Gram-positive bacteria define some of the
635 simplest of these toxin systems. Here, a standalone Lipocone domain is coupled to a signal

636 peptide or lipobox via a poorly structured linker. These are usually encoded in a two-gene
637 configuration with their cognate immunity protein (Figure 5O). More complex versions present,
638 in addition to adhesion, peptidoglycan-binding, lipid-binding and proteolytic processing
639 domains, multiple hallmarks of delivery through specific secretion systems. These include T4SS
640 (VirD4-binding domain), T6SS (PAAR domain), T7SS (WXG/LXG domain), T9SS, and MuF
641 domains (174,176) (Figure 5P, Supplementary Data). Additionally, we recovered standalone
642 CapCone-1 domains encoded in an operon with a PsbP/MOG1 superfamily domain diagnostic
643 of secretion via the T6SS (174,177) (Figure 5P). Further, we also found Min-Wnt domains fused
644 to the N-termini of RTX-like β-roll repeats, suggestive of T1SS-mediated export (178) (Figure 5P).
645
646 Our analysis also uncovered multiple, previously uncharacterized trafficking/packaging systems
647 associated with different Lipocone polymorphic toxins. Several Min-Wnt and CapCone-1
648 domains with lipoboxes are fused to an N-terminal Cystatin-like superfamily domain (179)
649 (Figure 5Q). The same domain is also comparably fused to several other C-terminal toxin
650 domains in related organisms, some of which are also predicted to target lipid head groups: (i) a
651 novel toxin domain we unified with the lipid-targeting Colicin M fold (180); (ii) a lipid-binding
652 START-domain-like helix-grip fold domain (154); (iii) a papain-like fold fatty acyltransferase
653 (181); (iv) a domain related to the VanY-like D-Ala-D-Ala carboxypeptidase (182) (Figure 5Q,
654 Supplementary Data). In all these cases, the toxins are coupled to a related immunity protein
655 (see below), suggesting that they define a distinct polymorphic toxin system. We propose that
656 this Cystatin-like domain specifies a novel packaging or deployment system upon secretion for
657 the C-terminal toxin domain, analogous to Cystatin domains in functioning with eukaryotic
658 proteases (183). The prok-TelC family Lipocones are found in distinctive architectures in two
659 poorly characterized, predicted polymorphic toxin systems. In one of them, they are fused to an
660 N-terminal glucan-binding GbpC β-sandwich domain (184) and repeats of MucBP-like Ig
661 domains (185), which might anchor them to exopolysaccharides (Figures 5Q,S7 Supplementary
662 Data). The second variant found in association with T9SS components (186) shows fusions to
663 one or more copies of a previously undetected TPM domain (Figure 5Q). While the domain has
664 been claimed to be a phosphatase (187), our recent analysis indicates that this is unlikely to be

665 the case (164). Instead, we propose that the TPM domain might assist in assembling membrane-
666 linked protein complexes, a role that might be relevant to the trafficking of these toxins (164).
667
668 To date, the only experimentally characterized Lipocone domain from polymorphic toxins is of
669 the prok-TelC family that are secreted via T7SS (29,188) (Figures 5P,6). Notably, prok-TelC has
670 been shown to be active only outside the cell and not in the cytoplasm (29). As noted above, it
671 attacks lipid II to cleave off the peptide-linked disaccharide pyrophosphate head group from the
672 undecaprenol tail (Figure 3B). Prok-TelC has also been speculated to similarly attack WTA-lipid II
673 linkages (29). These findings provide a template for other Lipocone superfamily effectors in
674 potentially targeting lipid carrier linkages in peptidoglycan and exopolysaccharide intermediates.
675 However, given the diversity within the family (Figure 3F), it is conceivable that they also target
676 other lipids.
677
678 *Immunity proteins of Lipocone polymorphic toxins indicate periplasmic/intramembrane action*
679 To date, only a single immunity protein has been reported for Lipocone toxins, viz., TipC, which
680 counters prok-TelC toxin in the periplasm (29,189) (Figure 5P,6). Here, we uncovered a range of
681 immunity proteins belonging to structurally distinct folds that counter the remaining Lipocone
682 toxins (Figures 5N-Q,S9, Supplementary Data). The most widespread of these is a rapidly
683 evolving, membrane-anchored member of the BamE-like superfamily that associates with not
684 only Min-Wnt and CapCone toxins but also other above-mentioned lipid-head-group targeting
685 toxins (e.g., the novel Colicin M-like domain). The BamE-like fold features a core two-helix
686 hairpin followed by a run of three β-strands (Figure S9). The classical BamE operates in a
687 pathway for the assembly of OMP β-barrels (190,191), suggesting that these immunity proteins
688 emerged from an ancestral BamE and, like it, function in the periplasm. Additional candidate
689 immunity proteins with more restricted phyletic spreads include (Figure S9): (i) a β-jelly-roll fold-
690 containing protein (144); (ii) an integral membrane protein with a 4-TM core. These two are
691 observed with Min-Wnt toxins. (iii) A novel domain combining an α-helix with a run of 4 β-
692 strands stabilized by four absolutely conserved cysteine residues. This is coupled to both Min-
693 Wnt and prok-SAA toxins; (iv) a protein with an OB-fold domain (145) (v) a protein with a β-

694 sandwich related to the eukaryotic centriolar assembly SAS-6 N-terminal domain (192). The last
695 two are coupled to CapCone-2 toxins (Supplementary Data). Notably, despite their structural
696 diversity, these immunity proteins are all TM or lipoproteins and, like TipC (29,189), are
697 predicted to operate at the membrane or in the periplasm (Figure 5N-Q, Supplementary Data).
698 This suggests that they intercept their cognate Lipocone toxin domain outside cells or in the
699 membrane rather than within the cell.
700
701 *Lipocone toxins in predator-prey and other interspecific conflicts*
702 In contrast to polymorphic toxins, which are typically deployed in intraspecific conflict between
703 competing strains of the same species, other toxin systems are deployed against more distantly
704 related target cells, such as prey and eukaryotic hosts (193). While some of these closely parallel
705 polymorphic toxins in their domain architecture, they are usually distinguished by the lack of an
706 accompanying immunity protein. The simplest of these systems are secreted Min-Wnt proteins
707 from bacteria and fungi. These present just a standalone Min-Wnt domain or one fused to a
708 novel domain with a half β-barrel wrapping around a helix (Figure 5R, Supplementary Data).
709 These are probably deployed as diffusible toxins that target rival organisms in the environment.
710
711 Another architectural theme is defined by Min-Wnt and prok-SAA Lipocones fused to an
712 enigmatic, novel, short C-terminal domain, which is comprised of a long β-hairpin with a
713 characteristic "break" in its central region, causing it to acquire an arch-like appearance (Figures
714 5S, S10). Hence, we refer to this domain as the broken-hairpin. We found the broken-hairpin
715 domain to be fused to a wide array of predicted toxin domains across the bacterial
716 superkingdom. These include effector domains otherwise found in polymorphic toxin and allied
717 systems that target peptidoglycan, carrier lipids and the membrane, such as members of the
718 Colicin M (180), Zeta toxin-kinase (194), lysozyme (195), an α/β-hydrolase superfamilies (110)
719 and nuclease toxins such as members of the HNH, HipA, SNase, and BECR superfamilies (174)
720 (Figures 5S, S10C). Remarkably, these proteins with the broken-hairpin tend to lack a signal
721 peptide or association with any other secretion system or immunity proteins (Figure 5S,
722 Supplemental Material). Hence, we propose that the broken-hairpin domain itself serves as a

723 trafficking mechanism for the externalization of these toxins in conflicts with rival environmental
724 organisms.
725
726 Some predicted secreted Lipocones are found predominantly in predatory bacteria. The first of
727 these are CapCone-2 domains from lineages like Bdellovibrionota, which are encoded in two-
728 gene systems, with the second gene coding for a further secreted effector such as an α/β-
729 hydrolase, Patatin, or acyltransferase or an OMP β-barrel domain (110,116,117,123) (Figures
730 5T,S7). Myxobacteria and some other lineages code for secreted prok-SAA domains fused to a
731 N-terminal Zincin-like metallopeptidase domain, and the first bacterial example of the von
732 Willebrand Factor D (vWD) and Ig domains at the C-terminus (196) (Figure 5T, Supplemental
733 Material). In the recently described predatory Patescibacterial branch of Omnitrophota species,
734 Skillet-clade Lipocone domains are found in gigantic proteins combined with several other
735 domains and TM segments. Domains found in these proteins include polysaccharide
736 biosynthesis enzymes (94,97), signaling proteins involved in Histidine kinase-Receiver relays
737 (197), peptidases of the MPTase and Papain-like superfamily (113,138), diverse methylases, and
738 extracellular ligand-binding domains like the peptidoglycan-binding LysM domain (198) (Figures
739 5T,6). Given the concentration of the above systems in predatory bacteria (Supplementary Data),
740 we posit that the above Lipocones might function as toxins targeting prey membranes alongside
741 a battery of effectors targeting other cellular components. In particular, the CapCone-2 systems
742 might play a role in the breaching of outer membranes by Bdellovibrionota. Animal vWD
743 domains are involved in adhesion (199); hence, the bacterial versions might play a similar role in
744 adhering to prey cells, while the MPTase in these proteins potentially releases the associated
745 Prok-SAA toxin through autoproteolysis. Finally, the giant proteins from the Patescibacteria are
746 likely to combine signaling prey presence with overcoming prey defenses and breaching prey
747 membranes.
748
749 Certain prok-TelC proteins are observed as part of several distinctive systems that could be
750 involved in as-yet-undiscovered predatory interactions or in targeting environmental
751 competitors. One such, defined by large proteins from spore-forming Bacillota, combines a

752 diversifying set of extracellular ligand-binding domains (e.g., Ig-like, Cell-wall-binding β-hairpins
753 and β-propellers (146,147,200)) with a two-enzyme core formed by a prok-TelC and a N-
754 acetylglucosamine (GlcNAc)-1-phosphodiester alpha-N-acetylglucosaminidase (NAGPA).
755 NAGPA catalyzes phosphoric-diester hydrolysis to release phosphodiester-linked sugars (Figures
756 5U,6,S7) (201). Some of these proteins feature an additional NlpC/p60 superfamily peptidase
757 domain predicted to target peptidoglycan (181). The recombinational diversity of ligand-binding
758 domains in this system, even among closely related Bacillota species, supports a possible arms
759 race and involvement in a biological conflict. Other TelC domains in some Bacillota,
760 Actinomycetota, and fungi are fused to peptidoglycan-binding domains (PGBD) (202) and an
761 Rv2525c-like TIM-barrel (203) (Figures 5U,S7). In Actinomycetota, this protein is further
762 combined in operons with either of two mutually exclusive genes coding for rapidly evolving
763 proteins (Figure 5U): (i) a secreted protein containing a pair of Ig domains (200); (ii) a 3-TM
764 protein (3TM-CCDN) with two conserved cysteines, an aspartate and asparagine residues
765 predicted to be located between the TM segments outside the cell. This version is further
766 coupled to a gene for a secreted VanY superfamily peptidase (182) (Figure 5U, Supplementary
767 Data). Common to these contexts are rapidly evolving and variable domains on the one hand
768 and peptidoglycan/exopolysaccharide binding or degrading domains on the other. Hence, we
769 interpret these as potential conflict systems that engage the cell wall and target it and
770 associated membranes in rival bacteria.
771
772 *Lipocone domains in resistance to antimicrobial agents*
773 VanZ-1 proteins (Figure 1A) were initially identified as encoded by a gene linked to that coding
774 for the VanY D-alanyl-D-alanine carboxypeptidase involved in resistance to glycopeptide
775 antibiotics like vancomycin and teicoplanin (31,204-206) (Figure 6). These antibiotics bind the
776 terminal D-Ala-D-Ala in the peptide moiety of peptidoglycan, preventing the transpeptidase
777 cross-linking reaction necessary for its maturation. Upon detection of these antibiotics, enzymes
778 encoded by the core vancomycin resistance operon re-engineer the exported peptidoglycan by
779 inserting a D-Ala-D-Lac in place of the D-Ala-D-Ala linkage, precluding antibiotic binding (53).
780 The VanY peptidase, while not strictly required for antibiotic resistance, acts as an accessory to

781 this system by cleaving any remaining D-Ala-D-Ala linkages generated via the canonical
782 pathway (53,205). However, the role of VanZ in this system has so far remained unknown. While
783 only a small fraction of the VanZ-1 genes are found in these antibiotic resistance contexts
784 (Supplementary Data), interestingly, other Lipocone genes, namely those of the VanZ-2 and the
785 Skillet families, might also be linked to VanY in lieu of VanZ-1. Further, VanY might be replaced
786 by a structurally unrelated secreted D-Ala-D-Ala carboxypeptidase of the metallo-beta-
787 lactamase fold (111) in operonic contexts with VanZ-1 (Figure 5V). Hence, given our above
788 prediction regarding VanZ acting in peptidoglycan and/or exopolysaccharide metabolism, VanZ-
789 1 and the Lipocones displacing it might indeed play an accessory role with VanY at the
790 membrane (204,205) in antibiotic resistance. We posit that, in these contexts, it likely acts on the
791 head group of Lipid II to recycle canonical peptidoglycan intermediates for their accelerated or
792 more thorough replacement with the resistant versions (Figure 3G).
793
794 We also identified a conserved five-gene operon featuring a YfiM-1 family Lipocone that might
795 play a role in resistance to antibacterial agents (Figures 5W,S5). Other than YfiM-1, this operon
796 contains genes for: (i) a thioredoxin domain protein (109); (ii) A DTW clade RNA modifying
797 enzyme of the SPOUT superfamily (207,208); (iii) a protein with acyl-CoA ligase, GNAT
798 superfamily N-acetyltransferase and ATP-grasp domains (131,209,210); (iv) a PssA-like
799 phosphatidylserine synthetase of the HKD superfamily (211) (Figure 5W, Supplementary Data).
800 Of these enzymes, the phosphatidylserine synthetase is predicted to act in its usual capacity to
801 generate a lipid with a serine head group (211). We propose that this would then function as a
802 substrate for the YfiM-1 Lipocone domain, which might exchange the serine for another moiety
803 via a reaction paralleling PTDSS1/2 (Figure 3A). This moiety could then be modified by
804 aminoacylation, further acylation and a redox modification by the third protein listed above,
805 together with the thioredoxin. Indeed, such peptide modifications of lipid head groups by lysine,
806 alanine, or arginine aminoacylation catalyzed by derived tRNA synthetases fused to GNATs have
807 been shown to be a key resistance mechanism against breaching of the membrane by
808 antibacterial peptides (212,213). Hence, we predict the modifications catalyzed by this system
809 might play a comparable role. The presence of a tRNA-modifying DTW domain suggests that in

810 parallel to the tRNA synthetases, the GNAT in this system might use a tRNA-linked acyl group as
811 a substrate, as seen in peptidoglycan biosynthesis (214,215).
812
813 *Eukaryotic recruitments of the Lipocone superfamily*
814 Lipocone domains have been transferred on several occasions from bacteria to eukaryotes
815 (Figures 4,S1). While there is predicted functional overlap with the above-described,
816 predominantly bacterial versions, we discuss these separately as the inferred biological contexts
817 of their deployment are often distinct from the above.
818
819 *Plant YfiM-1 and eukaryotic VanZ-2 proteins*
820 A conserved YfiM-1 family protein typified by the *Arabidopsis* AT1G15900 was acquired from
821 the bacteroidetes lineage of bacteria at the base of the plant lineage prior to the chlorophyte-
822 streptophyte (including land plants) split and is predicted to be catalytically active (Figure S2,
823 Supplementary Data). In *Arabidopsis*, this gene is widely expressed across different tissue types,
824 developmental stages, and other tested conditions (216,217). Given the above-predicted roles
825 for bacterial YfiM-1 proteins, it is conceivable that the plant version plays a comparable role in
826 the metabolism of a conserved plant-specific lipid. In a similar vein, a distinct clade of
827 standalone VanZ-2 domains typified by the *Saccharomyces cerevisiae* YJR112W-A was acquired
828 early in the fungal lineage. A similar transfer is also seen in the SAR clade of eukaryotes (Figure
829 S1). Since these eukaryotes lack peptidoglycan and other bacterial-type isoprenoid lipid-borne
830 exopolysaccharide intermediates, we suggest that this version was recruited for modifications of
831 a fungus-specific lipid (e.g., highly oxygenated isoprenoid lipids) (218).
832
833 *The Met-TelC proteins*
834 The Met-TelC clade is comprised of versions of the TelC family with a reconfigured active site
835 transferred from bacteria to Metazoa prior to the divergence of the cnidarians, and most
836 members are predicted to be catalytically inactive (Figure 2). In cnidarians and arthropods, the
837 Met-TelC domain is found in a secreted protein fused to C-terminal adhesion-related vWA (219)
838 and Ig domains, followed by a TM helix (Figure 5X, Supplementary Data). The chordate version,

839 typified by human PGLYRP2 (220), is also secreted and is fused to a C-terminal Amidase
840 targeting the N-acetylmuramoyl-L-alanine linkage (Figure 5X, Supplementary Data). PGLYRP2 is
841 a key innate immunity factor against bacterial pathogens degrading sugar-peptide linkages in
842 peptidoglycan via the Amidase domain (221-223). As most Met-TelC proteins lack the active site
843 residues but are modeled to retain the substrate-binding pocket, we propose that they
844 participate in anti-bacterial immunity as a Pathogen-Associated Molecular Pattern (PAMP)
845 receptor (224). Specifically, they could recognize polyisoprenoid pyrophosphate-linkage-
846 containing lipid intermediates of bacterial cell-surface molecules like peptidoglycan or
847 exopolysaccharides.
848
849 *Eukaryotic Wnt proteins*
850 Wnt family Lipocones were transferred on multiple occasions to eukaryotes. The best-known of
851 these are Met-Wnt proteins, which were acquired from bacteria at the base of Metazoa after
852 they had separated from their closest sister group, the choanoflagellates. These lost the
853 ancestral active site residues and function as well-studied secreted signaling molecules and will
854 not be detailed further in this work (for review, see (1,225)). Independently of the Met-Wnt
855 proteins, catalytically active, secreted versions closely related to the bacterial Min-Wnt proteins
856 were transferred to fungi and, within Metazoa, to the rotifers and the hemichordate acorn worm
857 *Saccoglossus kowalevskii*, where they are lineage-specifically expanded (Figure S1,
858 Supplementary Data). These versions are primarily standalone versions of the Min-Wnt domain,
859 lacking the large inserts typical of the Met-Wnt proteins (Figure 1A). We predict that these
860 eukaryotic Min-Wnt proteins retain their ancestral toxin role and might participate in anti-
861 bacterial immunity.
862
863 *Met-SAA proteins*
864 Met-SAA proteins (Figure 1A) were acquired from bacteria prior to the divergence of the
865 cnidarians from the rest of Metazoa. However, unlike the Met-Wnt and Met-TelC proteins, they
866 often conserve the ancestral active site residues, indicating that they are usually enzymatically
867 active (Figure 3). Human SAA has been recognized as a key immune marker that dramatically

868 increases in blood during the Acute Phase Response (226). It has been reported to bind the *E.*
869 *coli* outer membrane protein OmpA (227) and claimed to function as an opsonin in innate
870 immunity (228). Like Met-TelC, but in contrast to Met-Wnts, Met-SAAs appear to have been lost
871 or pseudogenized in several animal lineages (229-231) (Figure 3). This is consistent with an
872 arms-race scenario in immunity and the development of pathogen resistance against the Met-
873 SAAs, leading to loss. Keeping with an immune role for the Met-SAAs, we propose a catalytic
874 function for the active versions in severing lipid head groups of outer-membrane lipids or of
875 isoprenoid lipid carrier intermediates. Such action could also generate PAMPs that could explain
876 the activation of neutrophil- and macrophage-based immunity by SAA (228). Pertinent to these
877 observations, diverse OMP β-barrels have been linked to the translocation of polymorphic toxin
878 domains across the outer membrane of target cells (232-234). Given this and the origin of Met-
879 SAA from bacterial polymorphic toxin-related systems (Figure 4), its interaction with OmpA
880 might help it cross over into the periplasmic space and act on maturing peptidoglycan or
881 teichoic acid intermediates.
882
883 SAA was first reported as a component of secondary amyloid deposits (235), and its capacity to
884 form amyloid fibrils upon protease cleavage was theorized as a potential PAMP activating the
885 immune response (30). Indeed, bacteria produce their own secreted amyloids, such as Curli and
886 Fap, believed to contribute to biofilm formation (236,237), and might be PAMPs recognized by
887 animal immune systems (238). Further, other animal amyloids, such as the β-amyloid, have been
888 proposed to play a role as physical barriers in immunity against bacteria (239). Thus, amyloid
889 formation by protease cleavage (including potentially by bacterial proteases) may represent a
890 second line of defense mediated by Met-SAA proteins.
891
892 **DISCUSSION**
893 **Early Evolution of the Lipocone superfamily**
894 No single well-defined Lipocone clade is universally conserved across the three superkingdoms
895 of Life (Figures 4, S1). However, the VanZ and Wok clades are both found across all major
896 bacterial phyla (notwithstanding sporadic losses in certain lineages) and in some archaeal

897 lineages (Figure S1). At the same time, the cpCone clade is found across most major archaeal
898 lineages and is nearly universally conserved in the eukaryotes (absent in Ascomycota and some
899 choanoflagellates) (Figure S1). Notably, the cpCone and Wok clades tend to group together in
900 the profile-profile similarity network (Figure 1B). These observations suggest that at least a
901 single version of the Lipocone superfamily was likely present in the Last Universal Common
902 Ancestor (LUCA). The phyletic patterns suggest that the LUCA Lipocone gave rise to the
903 VanZ/Wok precursor in the bacterial lineage on the one hand and the cpCONE clade via a
904 circular permutation event in the archaeo-eukaryotic lineage on the other (Figure 4). Based on
905 the features of these deep-branching clades, the LUCA version is inferred to feature a
906 hydrophobic domain with a 4TM helix core, with the active site facing the outer leaf of the lipid
907 bilayer (Figure 1C). Given that extant versions operate both on classic phospholipids and
908 isoprenoid lipids, it is difficult to infer which of these might have been substrates for the LUCA
909 version. It is not impossible that this early version had a generic specificity that became
910 specialized in the descendant clades.
911
912 **Subsequent diversification of the Lipocone domain**
913 The early diversification of the Lipocone domain appears to have had different drivers in the two
914 prokaryotic superkingdoms. The presence of an extensive repertoire of exopolysaccharides in
915 the cell wall (peptidoglycan, teichoic acids), cell surface (e.g., ECA), and outer membrane (e.g.,
916 lipopolysaccharide), synthesized via isoprenoid lipid-linked intermediates, like lipid-II, was the
917 primary driver in the bacterial superkingdom (240). Here, this diversification yielded 4
918 monophyletic groups: the VanZs, Wok, YfiM and Skillet (Figure 4). The deeper VanZ and Wok
919 branches, which were likely recruited first for lipid-II-related functions, were probably the
920 predecessors of the more restricted bacterial families with specialized functions. For instance,
921 the emergence of the outer membrane in certain bacteria was potentially coupled with the
922 origin of the YfiM-like clade (Figure 4). Similarly, our predictions suggest that within these clades,
923 further diversification accompanied the acquisition of specialized functional roles in antibiotic
924 resistance, secondary sensor roles in single and multicomponent signaling and lipoprotein
925 processing. The interoperability of Lipocone domains on lipid carriers shared across different

926 biosynthetic pathways (see above, Figures 3,S5-S6) appears to have been a key factor leading to
927 this versatility.
928
929 In the ancestral archaeo-eukaryotic lineage, the absence of peptidoglycan and an apparently
930 lower diversity of structures with exopolysaccharides was reflected in the lesser diversification of
931 the Lipocone clades (Figure 4). There are open questions regarding the biochemical functions of
932 the primary archaeo-eukaryotic Lipocone clade, the cpCONE. Although the eukaryotic cpCone
933 PTDSS1/2 family has been shown to swap serine for ethanolamine or choline in lipid head
934 groups (28,57), their archaeal counterparts remain uncharacterized. Archaea have their own lipid
935 with a serine in the head group (archaeophosphatidyserine), but to date, its synthesis has been
936 shown to depend on a patchwork of different CDP-alcohol phosphatidyltransferase enzymes
937 (CaPs) in different archaeal species (58,241,242). While the CaPs are also integral membrane
938 enzymes with a 6TM helix core, catalyzing comparable reactions as the Lipocones on lipid head
939 groups in archaea and eukaryotes (242), they are evolutionarily unrelated. Nevertheless, we
940 suggest that the archaeal cpCones, like their eukaryotic counterparts, could contribute to
941 distinct, as yet uncharacterized, pathways for the generation of cell membrane phospholipids
942 like archaeophosphatidylserine or those with other head groups.
943
944 **Emergence of diffusible versions of the Lipocone domain and their repeated recruitment in**
945 **biological conflicts**
946 One of the remarkable aspects of the Lipocone superfamily is the loss of ancestral
947 hydrophobicity in several families (Figure 1C), transforming them from integral membrane
948 proteins to diffusible domains. While unexpected, such a transition in integral membrane
949 enzymes acting on lipid substrates is not unprecedented. The PAP2 superfamily of integral
950 membrane enzymes (e.g., diacylglycerol diphosphate phosphatase) (124,243) also contains
951 several soluble versions (244) that appear to have emerged from an integral membrane ancestor
952 (AMB and LA, unpublished observations). Most of the soluble Lipocone domains retain their
953 active site conservation (Figure 2) and, at least in one experimentally characterized case, catalyze
954 a comparable reaction as the TM version (29) (Figure 3B). The weight of the evidence presented

955 here, including the profile-profile similarity network (Figure 1B), phyletic patterns (Figure S1),
956 functional contexts (Figure 5-6), and the broadly shared structural features (Figures 1A, S2, S4),
957 suggests that the loss of hydrophobicity occurred on a single occasion in the Lipocone
958 superfamily, followed by diversification of these diffusible versions.
959
960 Our analysis of the diffusible Lipocone families reveals repeated recruitment as toxins/effectors
961 in anti-viral and polymorphic toxin and allied systems (174), suggesting that their diversification
962 was driven by the arms races arising from the biological conflicts where they are deployed.
963 Recruitment of a representative of the VanZ-1 family as a polymorphic toxin on rare occasions
964 (Figures 5N,6) suggests a possible evolutionary pathway for their recruitment as toxins: the
965 effector version of Lipocones attacking lipids in competing bacteria likely emerged from an
966 ancestral version that catalyzed endogenous lipid-head-group modifications on the same lipids
967 in metabolic pathways. Once versions with reduced hydrophobicity emerged, they could be
968 deployed as diffusible effectors that were shared across extracellular and intracellular conflict
969 systems, a trend previously recognized in many other effector domains (25).
970
971 **Repeated acquisition of Lipocones of bacterial origin by eukaryotes**
972 Unlike bacteria, eukaryotes as a whole do not possess a rich repertoire of Lipocone domains. The
973 PTDSS1/2 family, vertically inherited from the archaeal progenitor, is the only version that can be
974 inferred as being present in the Last Eukaryotic Common Ancestor (Figure 4). However, distinct
975 Lipocone families of ultimately bacterial provenance were acquired early and fixed in certain
976 eukaryotic lineages: (i) YfiM-1 in the plant lineage; (ii) the fungal VanZ-2 domains typified by the
977 *Saccharomyces cerevisiae* YJR112W-A; (iii) Met-Wnt (discussed further below) (Figure S1). The
978 early fixation of these versions in the eukaryotic lineages possessing them suggests that they
979 were recruited for definitive "housekeeping" or developmental in the respective lineages.
980 Beyond these, the fungal and metazoan lineages show more sporadically distributed versions,
981 which have all been acquired from bacterial secreted-toxin or antiviral systems: (i) Min-Wnt
982 independently in fungi and certain Metazoa; (ii) SAA; (iii) TelC; the latter two are absent in the
983 basal-most metazoans, the sponges, but are present in Cnidaria, suggesting a relatively early

984 acquisition (Figure 4). The weight of the evidence suggests that they have retained certain
985 aspects of the ancestral bacterial effector function for anti-pathogen immunity in eukaryotes.
986 This is consistent with both their episodic loss and lineage-specific expansion, the tendency to
987 show rapid sequence divergence and, in the Met-TelC family, loss of catalytic activity (Figures
988 2,S1, Supplementary Data).
989
990 This independent acquisition of at least 3 distinct Lipocone families in metazoan immunity from
991 polymorphic and allied effector systems of prokaryotes points to a persistent evolutionary trend.
992 Notably, the Lipocone domains participating in animal immunity have been drawn from secreted
993 effectors rather than the intracellular versions (bacterial intracellular Min-Wnts) predicted to
994 participate in bacterial anti-selfish element immunity. More generally, this adds to a growing list
995 of components drawn from secreted effector systems of prokaryotes in eukaryotic immune
996 systems (174,193,245). For example, this closely parallels another structurally unrelated effector
997 domain, the Zn-dependent deaminase (e.g., metazoan AID/APOBEC deaminases) (246). Hence,
998 these observations add further support to our hypothesis that the extensive expansion of
999 effectors in diverse prokaryotic inter-organismal conflict systems served as a reservoir from
1000 which eukaryotic immune systems repeatedly acquired components (193,245). We propose that
1001 symbiotic associations between the early animals and bacteria resulted in potential interactions
1002 via secreted effectors of the latter that aided the former against antagonistic bacteria. This
1003 probably led to their eventual acquisition by animals and incorporation into their immune
1004 processes.
1005
1006 **Origin of Wnt as a signaling molecule**
1007 Earlier considerations on the evolution of Wnt signaling indicated that it emerged at the base of
1008 the metazoan lineage and incorporated a wide range of components of different origins (e.g.,
1009 the HMG domain transcription factor TCF/LEF, the HEAT repeat protein β-catenin and the 7TM
1010 receptor Frizzled) (1). However, the provenance of Met-Wnt itself had been mysterious and was
1011 seen as a possible example of a metazoan innovation (225). While the Met-Wnt domains possess
1012 peculiar structural elaborations (34), its conserved core is a Lipocone domain (Figure 1A). We

1013 establish that the progenitor of Met-Wnt emerged as part of the radiation of Lipocone domains
1014 in bacteria as effectors deployed in both intracellular and inter-organismal conflict – the Min-
1015 Wnt proteins.
1016
1017 Whereas the Min-Wnt proteins are predicted to be secreted toxins, the Met-Wnts underwent an
1018 ancestral inactivation through loss of the catalytic residues (Figure 2). However, they retained
1019 their ancient involvement in cell-cell interactions as secreted agents. The Met-Wnt residues
1020 recognized as essential for the receptor (Frizzled) binding, including the absolutely conserved
1021 palmitoleoylated serine residue, are found in the aforementioned Metazoa-specific hairpins and
1022 loops (34,36). However, despite their inactivation, the Met-Wnts retain the ancestral substrate-
1023 binding pocket (Figures 1A, S2). This raises the possibility that they might be involved in as-yet
1024 unexplored interactions with ligands such as lipids.
1025
1026 Our tracing of the provenance of Wnt back to an effector in secreted bacterial toxin systems
1027 adds it to a growing list of components in metazoan signaling networks that have been acquired
1028 from such systems. For instance, this is also the case with components of the other key
1029 metazoan signaling pathway, Hedgehog (247). Here, the Hedgehog protein itself contains an
1030 autoproteolytic HINT peptidase domain that was likely drawn from a structurally and functionally
1031 cognate domain observed in polymorphic toxin systems (174,247). Further, an intracellular
1032 component of the same signaling pathway, Supressor of Fused (SuFu), was derived from a
1033 common immunity protein found in polymorphic toxin systems (247). Similarly, the
1034 Teneurin/Odd Oz proteins mediating signaling in cell migration, neuronal pathfinding, and
1035 fasciculation in Metazoa descended from a polymorphic toxin protein with a C-terminal HNH
1036 endonuclease toxin tip (174). In a similar vein, the immunity protein of certain CapCone toxins
1037 identified in this study might have given rise to the β-sandwich domain in the eukaryotic
1038 centriolar assembly factor SAS-6. These observations suggest that, in addition to immune
1039 system components, interactions with symbiotic bacteria also potentially furnished the
1040 progenitors of components of eukaryotic signaling and cytoskeletal networks that were central
1041 to the emergence of Metazoa as a clade of multicellular eukaryotes (248,249).

1042

1043 ## CONCLUSIONS

1044 Using sensitive sequence and structure analysis, we unify a large, hitherto unrecognized
1045 superfamily of enzymatic domains, the Lipocone. By combining analysis of the active site and
1046 the structure of the Lipocone domain with contextual information from conserved gene-
1047 neighborhoods and domain architectures, we present evidence that members of this superfamily
1048 target phosphate linkages in head groups of both classical phospholipids and polyisoprenoid
1049 lipids. Specifically, they catalyze reactions such as head group exchange or severing of the head
1050 group-diphosphate linkage from the polyisoprenol. We present evidence that these activities
1051 have been recruited in a wide range of biochemical contexts, including cell membrane lipid
1052 modification, metabolism of peptidoglycan and exopolysaccharide lipid-carrier linked
1053 intermediates, lipoprotein modifications, bacterial outer membrane modification, sensing of
1054 membrane-associated signals, effector activity in antiviral and inter-organismal conflicts and
1055 resistance to antimicrobials. Further, catalytically inactive versions like Met-Wnt have been
1056 recruited for signaling roles in Metazoa. We predict the catalytic activity and potential
1057 biochemical pathways of numerous representatives for the first time, including some proteins
1058 that have remained enigmatic for over two decades, like VanZ.

1059

1060 We identify three notable trends in Lipocone evolution. First, although we reconstruct the
1061 ancestral member of the superfamily as being a 4TM integral membrane domain, a large
1062 monophyletic subset underwent a dramatic loss of hydrophobicity, transforming them into
1063 diffusible versions, including the Wnts and the SAAs (Figure 1C). Second, the superfamily
1064 expanded in two major functional niches in bacteria, namely peptidoglycan/exopolysaccharide
1065 metabolism and effector domains of both secreted toxins and immune systems (Figure 4). Finally,
1066 members of the Lipocone superfamily were acquired on multiple occasions from bacteria by
1067 Metazoa and were reused in new functional contexts as signaling messengers and immune
1068 factors (Figure 4).

1069

1070 Importantly, our predictions in this regard underscore that much remains unexplored in terms of
1071 lineage-specific cell wall and membrane metabolism in prokaryotes. We present several testable
1072 biochemical, functional hypotheses for the many poorly understood branches of the superfamily,
1073 several of which are being recognized as enzymatic for the first time here. We hope this will also
1074 open new avenues of research to fill key gaps in our understanding of lipid metabolism.
1075

1076 METHODS

1077 Sequence analysis

1078 Sequence similarity searches were performed using PSI-BLAST (250) and JackHMMER (251)
1079 against the NCBI non-redundant protein database (nr) (252) or a version clustered down to 50%
1080 sequence identity (nr50). The searches were initiated using the previously identified prokaryotic
1081 Wnt (24), with multiple rounds of searches conducted, each using seeds collected from the
1082 preceding searches. Clusters based on sequence similarity (percentage identity or bit-score) were
1083 generated using MMseqs (253). The clustering parameters were adjusted according to specific
1084 goals, enabling redundancy removal, the definition of homologous groups, and the creation of
1085 new profiles. Multiple sequence alignments (MSA) were generated using the MAFFT program
1086 (254) with the local-pair algorithm, combined with the parameters –maxiterate 3000, –op 1.5,
1087 and –ep 0.2, and were manually refined based on structural superpositions and profile-profile
1088 comparisons.
1089

1090 Sequence similarity network analysis

1091 The HHalign program (255) was used to perform profile-profile comparisons, with the resulting
1092 p-value and e-value scores serving as edges for constructing a superfamily relationship network.
1093 This was then analyzed using the Leiden community finding algorithm (37) to detect sub-
1094 networks. Network analysis and visualization were performed using the R igraph (256) or Python
1095 networkX libraries (257).
1096

1097 Comparative genomics, domain identification, and phylogenetic analysis

Genomic neighborhoods were obtained from genomes available in the NCBI Genome database (252) using in-house scripts written in Perl and Python. Conservation analysis of these genomic neighborhoods was performed by clustering the protein products of neighboring genes. Domain identification was conducted using a collection of HMMs and PSSMs maintained by the Aravind lab, along with HMMs from the Pfam database (258), utilizing the RPSBLAST (259) and HMMSCAN (260) programs. To further refine detection, domain identification was extended through remote homology analysis using the HHpred (261) program, against profiles built from the Pfam (258) and PDB70 (262) databases. Phylogenetic analyses were performed using FastTree (263) and iqTREE2 (264). Experimental functional data for characterized members of the superfamily were collected with the assistance of the ChatGPT language model (https://chat.openai.com). Structural comparisons, along with shared genomic associations, were used to further refine the interrelationships within and between the groups of the superfamily.

Families with broader presence across multiple major lineages ("phyla") and deeper conservation within each of those lineages were inferred to be more ancient. In contrast, those with a more limited phyletic *spread* and/or limited *depth* of occurrence within each major lineage were likely later derivations (Figures 4,S1). We formalized this inference by calculating a phyletic metric for the Lipocone clades (Figure S1) comprised of both the phyletic spread and depth. The phyletic spread S_i of the *i*th Lipocone clade was computed thus:

$$S_i = \frac{m_i}{M},$$

Where m_i is the number of lineages with at least one representative of the Lipocone clade S_i, and M is the total number of lineages examined. The phyletic depth D_i of the *i*th Lipocone clade was computed as a weighted average of its occurrence within each lineage in the form of the mediant:

$$D_i = \frac{\sum_{j=1}^{M} n_j}{\sum_{j=1}^{M} N_j},$$

where n_j is the number of species in lineage *j* with a Lipocone domain of the *i*th Lipocone clade and N_j is the total number of species sampled in lineage *j*. S_i and D_i are plotted as a bar graph with S_i as its width and D_i its height.

1127 **Contextual network construction**

1128 Each domain architecture and conserved gene neighborhood was decomposed into its
1129 constituent domains. These domains were then labeled for their biochemical function and stored
1130 as a YAML file (Supplementary Data). The contextual connections were then rendered as a graph
1131 with the domains as its nodes and the adjacency relationships as its edges. Cliques containing a
1132 given Lipocone domain were detected in this graph and merged to constitute their respective
1133 dense subgraphs. These subgraphs were then examined for the statistically significant
1134 prevalence of particular labeled functions using the Fisher exact test. Network analysis was
1135 performed using the functions of the R igraph or Python networkX libraries.

1136

1137 **Structure analysis**

1138 Protein structures were modeled using Alphafold3 (265), with visualization and manipulation
1139 performed using either MOL* (266) or PyMOL. Structural similarity searches were conducted
1140 using the DALIlite (267) and FOLDSEEK (268) programs. DALIlite was also used to generate
1141 structural alignments.

1142

1143 **Hydrophobicity analysis**

1144 To create the membrane propensity plots, for each protein P_i in a given family, we compute the
1145 average TM-propensity of its amino acids using the TM tendency scale (38). This score H_i for P_i is
1146 calculated as:

1147

$$H_i = \frac{1}{n}\sum_{j=1}^{n} h_j$$

1148

1149 where h_j is the TM tendency of the j-th amino acid in the protein P_i, and n is its total length in
1150 amino acids. The Kruskal–Wallis nonparametric test was applied to assess whether TM
1151 propensity scores differed across the 30 groups. As the Kruskal–Wallis test indicated a significant
1152 difference ($p<0.05$), we performed post-hoc pairwise comparisons using Dunn's test with
1153 Bonferroni correction to control for multiple testing. Group-wise visualizations were presented

using critical difference diagrams, where groups not connected by horizontal bars are significantly different (adjusted $p<0.05$) (Figure S3).

FIGURE LEGENDS

Figure 1. (A) The four individual helices forming the core of the Lipocone superfamily are consistently colored across the illustrated representatives. The inter-helix linkers are colored gray, and lineage-specific synapomorphic insertions and extensions are colored light brown. Active site and other residues of interest are rendered as ball-and-stick. Protein Data Bank (PDB) IDs or Genbank accessions used to generate AF3 models are provided. (B) Relationship network of the Lipocone families. The thickness of the edges is scaled by the negative-log HHalign p-values. Families are colored according to the community identified by the Leiden algorithm (37) (see Methods). (C) Box plots displaying core helix transmembrane propensity scores of individual sequences within different Lipocone families. The horizontal divider represents the boundary between typical TM and soluble sequences.

Figure 2. Sequence logo of conserved core elements of the Lipocone families. These correspond to the core helices H2, H3, and H4. The three conserved active site residue positions are boxed in dotted lines with the inferred ancestral residue indicated at the top of the alignment. Families are grouped and labeled on the left in their higher-order clades.

Figure 3. Known and predicted Lipocone reaction mechanisms. Experimentally supported reactions are boxed in blue (A-B), while a predicted reaction based on genome displacement by a Lipocone domain of an experimentally characterized enzyme is boxed in orange (C). The remaining reactions (D-G) are suggested based on the contextual inferences in this work. Attacking and leaving groups are denoted by dashed green and red circles, respectively.

Figure 4. Reconstructed evolutionary scenario for the Lipocone superfamily. The relative temporal epochs are demarcated by vertical lines and labeled at the bottom. The clades are represented by colored lines indicating the maximum depth to which the families listed to the right can be traced. Colors track the superkingdom-level phyletic distribution of the family. Dashed-line circles indicate uncertainty in the origin of lineage(s). Inferred or experimentally characterized functions for families are indicated to the left of family names. Asterisks denote newly described families.

Figure 5. Representative contexts for the Lipocone superfamily, grouped by shared functional themes. Genes are depicted by box arrows, with the arrowhead indicating the 3′ end of genes. Genes encoding proteins with multiple domains are broken into labeled sections corresponding to them. Domain architectures are depicted by the individual domains represented by distinct shapes. TMs, lipoboxes (LPs), and SPs are depicted as unlabeled, narrow yellow, blue, and red rectangles, respectively. All Lipocone domains are consistently colored in orange. Genes marked with asterisks are labeled by the Genbank accession number below each context. Colored labels above genes denote well-known gene names or gene cluster modules. Abbreviations: PTase, peptidase; TFase, transferase; GlycosylTFase, Glycosyltransferase; MPTase, metallopeptidase; TGase, transglycosylase; SLP, serine-containing lipobox; cNMPBD, cNMP-binding domain; NCPBM, novel putative carbohydrate binding module; (w)HTH, (winged) helix-turn-helix; ZnR, Zinc ribbon; PPTs, pentapeptide repeats; Imm, immunity protein; βPs, β-propeller repeats;

Cystatin-FD, Cystatin fold domain; MTase, methylase; PGBD, peptidoglycan-binding domain; MβL, metallo-β-lactamase; L12-ClpS, ClpS-ribosomal L7/L12 domain; TA, teichoic acid.

Figure 6. Lipocone contextual network. The network represents the conserved contextual associations of Lipocone domains (hexagonal nodes). Nodes and edges are colored based on known or inferred functional categories of the domains. The nodes are scaled by their degree. Gray coloring indicates domains without specific functional assignments. Examples of conserved gene neighborhoods and domain architectures supplementing those in Figure 5 illustrate contexts that bridge functional themes. Here, individual domains are colored to match network coloring. Additional abbreviations to those in Figure 5: APH-Pkinase, aminoglycoside phosphotransferase-like kinase; HUP, HIGH, UspA and PP-ATPase superfamily-like domain; Alk-phosphatase, Alkaline phosphatase; dehyd, dehydrogenase; TPRs, tetratricopeptide repeats; PMM/PGM, phosphomannomutase/phosphoglucomutase; ZnF, zinc finger; APC-transporter, amino acid-polyamine-organocation transporter; LPS, lipopolysaccharide.

ACKNOWLEDGEMENTS

This research was supported by the Division of Intramural Research at the National Library of Medicine (NLM), National Institutes of Health (NIH). This research was supported in part by an appointment to the NLM Research Participation Program administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy (DOE) and the NLM.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest.

REFERENCES

1. Richards, G.S. and Degnan, B.M. (2009) The dawn of developmental signaling in the metazoa. *Cold Spring Harb Symp Quant Biol*, **74**, 81-90.
2. Jessen, S., Gu, B. and Dai, X. (2008) Pygopus and the Wnt signaling pathway: a diverse set of connections. *Bioessays*, **30**, 448-456.
3. Morata, G. and Lawrence, P.A. (1977) The development of wingless, a homeotic mutation of Drosophila. *Dev Biol*, **56**, 227-240.
4. Rijsewijk, F., Schuermann, M., Wagenaar, E., Parren, P., Weigel, D. and Nusse, R. (1987) The Drosophila homolog of the mouse mammary oncogene int-1 is identical to the segment polarity gene wingless. *Cell*, **50**, 649-657.
5. Sharma, R.P. (1973) Wingless a new mutant in Drosophila melanogaster. *Drosophila information service*, 134.
6. Nusse, R. and Varmus, H.E. (1982) Many tumors induced by the mouse mammary tumor virus contain a provirus integrated in the same region of the host genome. *Cell*, **31**, 99-109.
7. Nusse, R. and Varmus, H.E. (1992) Wnt genes. *Cell*, **69**, 1073-1087.
8. Komiya, Y. and Habas, R. (2008) Wnt signal transduction pathways. *Organogenesis*, **4**, 68-75.

1245 9. Segalen, M. and Bellaiche, Y. (2009) Cell division orientation and planar cell polarity
1246 pathways. *Semin Cell Dev Biol*, **20**, 972-977.
1247 10. Slusarski, D.C. and Pelegri, F. (2007) Calcium signaling in vertebrate embryonic patterning
1248 and morphogenesis. *Dev Biol*, **307**, 1-13.
1249 11. Logan, C.Y. and Nusse, R. (2004) The Wnt signaling pathway in development and disease.
1250 *Annu Rev Cell Dev Biol*, **20**, 781-810.
1251 12. Welters, H.J. and Kulkarni, R.N. (2008) Wnt signaling: relevance to beta-cell biology and
1252 diabetes. *Trends Endocrinol Metab*, **19**, 349-355.
1253 13. Christian, J.L. (2000) BMP, Wnt and Hedgehog signals: how far can they go? *Curr Opin*
1254 *Cell Biol*, **12**, 244-249.
1255 14. Schulte, G. and Bryja, V. (2007) The Frizzled family of unconventional G-protein-coupled
1256 receptors. *Trends Pharmacol Sci*, **28**, 518-525.
1257 15. Bhanot, P., Brink, M., Samos, C.H., Hsieh, J.C., Wang, Y., Macke, J.P., Andrew, D., Nathans, J.
1258 and Nusse, R. (1996) A new member of the frizzled family from Drosophila functions as a
1259 Wingless receptor. *Nature*, **382**, 225-230.
1260 16. Takada, R., Satomi, Y., Kurata, T., Ueno, N., Norioka, S., Kondoh, H., Takao, T. and Takada, S.
1261 (2006) Monounsaturated fatty acid modification of Wnt protein: its role in Wnt secretion.
1262 *Dev Cell*, **11**, 791-801.
1263 17. Kurayoshi, M., Yamamoto, H., Izumi, S. and Kikuchi, A. (2007) Post-translational
1264 palmitoylation and glycosylation of Wnt-5a are necessary for its signalling. *Biochem J*,
1265 **402**, 515-523.
1266 18. Gao, C. and Chen, Y.G. (2010) Dishevelled: The hub of Wnt signaling. *Cell Signal*, **22**, 717-
1267 727.
1268 19. Klingensmith, J., Nusse, R. and Perrimon, N. (1994) The Drosophila segment polarity gene
1269 dishevelled encodes a novel protein required for response to the wingless signal. *Genes*
1270 *Dev*, **8**, 118-130.
1271 20. Kramps, T., Peter, O., Brunner, E., Nellen, D., Froesch, B., Chatterjee, S., Murone, M., Zullig, S.
1272 and Basler, K. (2002) Wnt/wingless signaling requires BCL9/legless-mediated recruitment
1273 of pygopus to the nuclear beta-catenin-TCF complex. *Cell*, **109**, 47-60.
1274 21. van Tienen, L.M., Mieszczanek, J., Fiedler, M., Rutherford, T.J. and Bienz, M. (2017)
1275 Constitutive scaffolding of multiple Wnt enhanceosome components by Legless/BCL9.
1276 *Elife*, **6**.
1277 22. Archbold, H.C., Yang, Y.X., Chen, L. and Cadigan, K.M. (2012) How do they do Wnt they
1278 do?: regulation of transcription by the Wnt/beta-catenin pathway. *Acta Physiol (Oxf)*, **204**,
1279 74-109.
1280 23. Holstein, T.W. (2012) The evolution of the Wnt pathway. *Cold Spring Harb Perspect Biol*, **4**,
1281 a007922.
1282 24. Burroughs, A.M. and Aravind, L. (2020) Identification of Uncharacterized Components of
1283 Prokaryotic Immune Systems and Their Diverse Eukaryotic Reformulations. *J Bacteriol*,
1284 **202**.
1285 25. Aravind, L., Iyer, L.M. and Burroughs, A.M. (2022) Discovering Biological Conflict Systems
1286 Through Genome Analysis: Evolutionary Principles and Biochemical Novelty. *Annu Rev*
1287 *Biomed Data Sci*, **5**, 367-391.

1288 26. Kuge, O., Saito, K. and Nishijima, M. (1997) Cloning of a Chinese hamster ovary (CHO)
1289 cDNA encoding phosphatidylserine synthase (PSS) II, overexpression of which suppresses
1290 the phosphatidylserine biosynthetic defect of a PSS I-lacking mutant of CHO-K1 cells. *J*
1291 *Biol Chem*, **272**, 19133-19139.
1292 27. Kuge, O., Nishijima, M. and Akamatsu, Y. (1991) A Chinese hamster cDNA encoding a
1293 protein essential for phosphatidylserine synthase I activity. *J Biol Chem*, **266**, 24184-
1294 24189.
1295 28. Stone, S.J. and Vance, J.E. (1999) Cloning and expression of murine liver
1296 phosphatidylserine synthase (PSS)-2: differential regulation of phospholipid metabolism
1297 by PSS1 and PSS2. *Biochem J*, **342 (Pt 1)**, 57-64.
1298 29. Whitney, J.C., Peterson, S.B., Kim, J., Pazos, M., Verster, A.J., Radey, M.C., Kulasekara, H.D.,
1299 Ching, M.Q., Bullen, N.P., Bryant, D., Goo, Y.A., Surette, M.G., Borenstein, E., Vollmer, W. and
1300 Mougous, J.D. (2017) A broadly distributed toxin family mediates contact-dependent
1301 antagonism between gram-positive bacteria. *Elife*, **6**.
1302 30. Sack, G.H., Jr. (2018) Serum amyloid A - a review. *Mol Med*, **24**, 46.
1303 31. Arthur, M., Depardieu, F., Molinas, C., Reynolds, P. and Courvalin, P. (1995) The vanZ gene
1304 of Tn1546 from Enterococcus faecium BM4147 confers resistance to teicoplanin. *Gene*,
1305 **154**, 87-92.
1306 32. Arthur, M., Depardieu, F., Reynolds, P. and Courvalin, P. (1999) Moderate-level resistance
1307 to glycopeptide LY333328 mediated by genes of the vanA and vanB clusters in
1308 enterococci. *Antimicrob Agents Chemother*, **43**, 1875-1880.
1309 33. Bazan, J.F., Janda, C.Y. and Garcia, K.C. (2012) Structural architecture and functional
1310 evolution of Wnts. *Dev Cell*, **23**, 227-232.
1311 34. Janda, C.Y., Waghray, D., Levin, A.M., Thomas, C. and Garcia, K.C. (2012) Structural basis of
1312 Wnt recognition by Frizzled. *Science*, **337**, 59-64.
1313 35. Chu, M.L., Ahn, V.E., Choi, H.J., Daniels, D.L., Nusse, R. and Weis, W.I. (2013) structural
1314 Studies of Wnts and identification of an LRP6 binding site. *Structure*, **21**, 1235-1242.
1315 36. Zhong, Q., Zhao, Y., Ye, F., Xiao, Z., Huang, G., Xu, M., Zhang, Y., Zhan, X., Sun, K., Wang, Z.,
1316 Cheng, S., Feng, S., Zhao, X., Zhang, J., Lu, P., Xu, W., Zhou, Q. and Ma, D. (2021) Cryo-EM
1317 structure of human Wntless in complex with Wnt3a. *Nat Commun*, **12**, 4541.
1318 37. Traag, V.A., Waltman, L. and van Eck, N.J. (2019) From Louvain to Leiden: guaranteeing
1319 well-connected communities. *Sci Rep*, **9**, 5233.
1320 38. Zhao, G. and London, E. (2006) An amino acid "transmembrane tendency" scale that
1321 approaches the theoretical limit to accuracy for prediction of transmembrane helices:
1322 relationship to biological hydrophobicity. *Protein Sci*, **15**, 1987-2001.
1323 39. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool,
1324 K., Bates, R., Zidek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J.,
1325 Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman,
1326 D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S.,
1327 Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P. and Hassabis, D. (2021) Highly
1328 accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583-589.
1329 40. Bastard, K., Smith, A.A., Vergne-Vaxelaire, C., Perret, A., Zaparucha, A., De Melo-Minardi, R.,
1330 Mariage, A., Boutard, M., Debard, A., Lechaplais, C., Pelle, C., Pellouin, V., Perchat, N., Petit,
1331 J.L., Kreimeyer, A., Medigue, C., Weissenbach, J., Artiguenave, F., De Berardinis, V., Vallenet,

1332 D. and Salanoubat, M. (2014) Revealing the hidden functional diversity of an enzyme
1333 family. *Nat Chem Biol*, **10**, 42-49.
1334 41. Glasner, M.E., Gerlt, J.A. and Babbitt, P.C. (2006) Evolution of enzyme superfamilies. *Curr*
1335 *Opin Chem Biol*, **10**, 492-497.
1336 42. Zhang, D., Iyer, L.M., Burroughs, A.M. and Aravind, L. (2014) Resilience of biochemical
1337 activity in protein domains in the face of structural divergence. *Curr Opin Struct Biol*, **26**,
1338 92-103.
1339 43. Tomohiro, S., Kawaguti, A., Kawabe, Y., Kitada, S. and Kuge, O. (2009) Purification and
1340 characterization of human phosphatidylserine synthases 1 and 2. *Biochem J*, **418**, 421-
1341 429.
1342 44. Saito, K., Kuge, O., Akamatsu, Y. and Nishijima, M. (1996) Immunochemical identification
1343 of the pssA gene product as phosphatidylserine synthase I of Chinese hamster ovary cells.
1344 *FEBS Lett*, **395**, 262-266.
1345 45. Stone, S.J. and Vance, J.E. (2000) Phosphatidylserine synthase-1 and -2 are localized to
1346 mitochondria-associated membranes. *J Biol Chem*, **275**, 34534-34540.
1347 46. Miyata, N. and Kuge, O. (2021) Topology of phosphatidylserine synthase 1 in the
1348 endoplasmic reticulum membrane. *Protein Sci*, **30**, 2346-2353.
1349 47. Anderson, J.S., Matsuhashi, M., Haskin, M.A. and Strominger, J.L. (1967) Biosythesis of the
1350 peptidoglycan of bacterial cell walls. II. Phospholipid carriers in the reaction sequence. *J*
1351 *Biol Chem*, **242**, 3180-3190.
1352 48. Higashi, Y., Strominger, J.L. and Sweeley, C.C. (1967) Structure of a lipid intermediate in
1353 cell wall peptidoglycan synthesis: a derivative of a C55 isoprenoid alcohol. *Proc Natl Acad*
1354 *Sci U S A*, **57**, 1878-1884.
1355 49. English, D. (1996) Phosphatidic acid: a lipid messenger involved in intracellular and
1356 extracellular signalling. *Cell Signal*, **8**, 341-347.
1357 50. Zamocky, M. and Ferianc, P. (2023) Discovering the deep evolutionary roots of serum
1358 amyloid A protein family. *Int J Biol Macromol*, **252**, 126537.
1359 51. Woods, E.C., Wetzel, D., Mukerjee, M. and McBride, S.M. (2018) Examination of the
1360 Clostridioides (Clostridium) difficile VanZ ortholog, CD1240. *Anaerobe*, **53**, 108-115.
1361 52. Sur, V.P., Mazumdar, A., Vimberg, V., Stefani, T., Androvic, L., Kracikova, L., Laga, R.,
1362 Kamenik, Z. and Komrskova, K. (2021) Specific Inhibition of VanZ-Mediated Resistance to
1363 Lipoglycopeptide Antibiotics. *Int J Mol Sci*, **23**.
1364 53. Stogios, P.J. and Savchenko, A. (2020) Molecular mechanisms of vancomycin resistance.
1365 *Protein Sci*, **29**, 654-669.
1366 54. Dziarski, R. and Gupta, D. (2006) The peptidoglycan recognition proteins (PGRPs).
1367 *Genome Biol*, **7**, 232.
1368 55. Bliven, S. and Prlic, A. (2012) Circular permutation in proteins. *PLoS Comput Biol*, **8**,
1369 e1002445.
1370 56. Aravind, L. (2000) Guilt by association: contextual information in genome analysis.
1371 *Genome Res*, **10**, 1074-1077.
1372 57. Vance, J.E. (2018) Historical perspective: phosphatidylserine and
1373 phosphatidylethanolamine from the 1800s to the present. *J Lipid Res*, **59**, 923-944.

1374 58. Koga, Y. and Morii, H. (2005) Recent advances in structural research on ether lipids from
1375 archaea including comparative and physiological aspects. *Biosci Biotechnol Biochem*, **69**,
1376 2019-2034.
1377 59. Caforio, A. and Driessen, A.J.M. (2017) Archaeal phospholipids: Structural properties and
1378 biosynthesis. *Biochim Biophys Acta Mol Cell Biol Lipids*, **1862**, 1325-1339.
1379 60. Rezanka, T., Kyselova, L. and Murphy, D.J. (2023) Archaeal lipids. *Prog Lipid Res*, **91**,
1380 101237.
1381 61. Narita, S. and Tokuda, H. (2010) Sorting of bacterial lipoproteins to the outer membrane
1382 by the Lol system. *Methods Mol Biol*, **619**, 117-129.
1383 62. Kamada, K., Miyata, M. and Hirano, T. (2013) Molecular basis of SMC ATPase activation:
1384 role of internal structural changes of the regulatory subcomplex ScpAB. *Structure*, **21**,
1385 581-594.
1386 63. Schleiffer, A., Kaitna, S., Maurer-Stroh, S., Glotzer, M., Nasmyth, K. and Eisenhaber, F.
1387 (2003) Kleisins: a superfamily of bacterial and eukaryotic SMC protein partners. *Mol Cell*,
1388 **11**, 571-575.
1389 64. Soppa, J., Kobayashi, K., Noirot-Gros, M.F., Oesterhelt, D., Ehrlich, S.D., Dervyn, E.,
1390 Ogasawara, N. and Moriya, S. (2002) Discovery of two novel families of proteins that are
1391 proposed to interact with prokaryotic SMC proteins, and characterization of the Bacillus
1392 subtilis family members ScpA and ScpB. *Mol Microbiol*, **45**, 59-71.
1393 65. Aravind, L., Anantharaman, V., Balaji, S., Babu, M.M. and Iyer, L.M. (2005) The many faces
1394 of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol Rev*,
1395 **29**, 231-262.
1396 66. Pellicer, M.T., Badia, J., Aguilar, J. and Baldoma, L. (1996) glc locus of Escherichia coli:
1397 characterization of genes encoding the subunits of glycolate oxidase and the glc
1398 regulator protein. *J Bacteriol*, **178**, 2051-2059.
1399 67. Ramachandran, R., Hartmann, C., Song, H.K., Huber, R. and Bochtler, M. (2002) Functional
1400 interactions of HslV (ClpQ) with the ATPase HslU (ClpY). *Proc Natl Acad Sci U S A*, **99**,
1401 7396-7401.
1402 68. Storck, E.M., Ozbalci, C. and Eggert, U.S. (2018) Lipid Cell Biology: A Focus on Lipids in
1403 Cell Division. *Annu Rev Biochem*, **87**, 839-869.
1404 69. Barak, I. and Muchova, K. (2013) The role of lipid domains in bacterial cell processes. *Int J*
1405 *Mol Sci*, **14**, 4050-4065.
1406 70. Hauck, A.K. and Bernlohr, D.A. (2016) Oxidative stress and lipotoxicity. *J Lipid Res*, **57**,
1407 1976-1986.
1408 71. Guo, D. and Tropp, B.E. (2000) A second Escherichia coli protein with CL synthase activity.
1409 *Biochim Biophys Acta*, **1483**, 263-274.
1410 72. Tan, B.K., Bogdanov, M., Zhao, J., Dowhan, W., Raetz, C.R. and Guan, Z. (2012) Discovery of
1411 a cardiolipin synthase utilizing phosphatidylethanolamine and phosphatidylglycerol as
1412 substrates. *Proc Natl Acad Sci U S A*, **109**, 16504-16509.
1413 73. Aravind, L. and Koonin, E.V. (1998) Phosphoesterase domains associated with DNA
1414 polymerases of diverse origins. *Nucleic Acids Res*, **26**, 3746-3752.
1415 74. Burroughs, A.M., Allen, K.N., Dunaway-Mariano, D. and Aravind, L. (2006) Evolutionary
1416 genomics of the HAD superfamily: understanding the structural adaptations and catalytic

1417 diversity in a superfamily of phosphoesterases and allied enzymes. *J Mol Biol*, **361**, 1003-
1418 1034.
1419 75. Vershinin, A. (1999) Biological functions of carotenoids--diversity and evolution.
1420 *Biofactors*, **10**, 99-104.
1421 76. Šesták, Z., Britton, G., Liaaen-Jensen, S. and Pfander, H. (2004) Carotenoids. Handbook.
1422 *Photosynthetica*, **42**, 186.
1423 77. Sandmann, G. and Misawa, N. (1992) New functional assignment of the carotenogenic
1424 genes crtB and crtE with constructs of these genes from Erwinia species. *FEMS Microbiol*
1425 *Lett*, **69**, 253-257.
1426 78. Chamovitz, D., Misawa, N., Sandmann, G. and Hirschberg, J. (1992) Molecular cloning and
1427 expression in Escherichia coli of a cyanobacterial gene coding for phytoene synthase, a
1428 carotenoid biosynthesis enzyme. *FEBS Lett*, **296**, 305-310.
1429 79. Liu, J. and Mushegian, A. (2003) Three monophyletic superfamilies account for the
1430 majority of the known glycosyltransferases. *Protein Sci*, **12**, 1418-1431.
1431 80. Tran, U.C. and Clarke, C.F. (2007) Endogenous synthesis of coenzyme Q in eukaryotes.
1432 *Mitochondrion*, **7 Suppl**, S62-71.
1433 81. Massiere, F. and Badet-Denisot, M.A. (1998) The mechanism of glutamine-dependent
1434 amidotransferases. *Cell Mol Life Sci*, **54**, 205-222.
1435 82. Liu, X., Zeng, J., Chen, X. and Xie, W. (2016) Crystal structures of RidA, an important
1436 enzyme for the prevention of toxic side products. *Sci Rep*, **6**, 30494.
1437 83. Aravind, L. and Koonin, E.V. (1998) The HD domain defines a new superfamily of metal-
1438 dependent phosphohydrolases. *Trends Biochem Sci*, **23**, 469-472.
1439 84. Becker, A., Kleickmann, A., Kuster, H., Keller, M., Arnold, W. and Puhler, A. (1993) Analysis
1440 of the Rhizobium meliloti genes exoU, exoV, exoW, exoT, and exoI involved in
1441 exopolysaccharide biosynthesis and nodule invasion: exoU and exoW probably encode
1442 glucosyltransferases. *Mol Plant Microbe Interact*, **6**, 735-744.
1443 85. Ruiz, N. (2015) Lipid Flippases for Bacterial Peptidoglycan Biosynthesis. *Lipid Insights*, **8**,
1444 21-31.
1445 86. Ruiz, N. (2008) Bioinformatics identification of MurJ (MviN) as the peptidoglycan lipid II
1446 flippase in Escherichia coli. *Proc Natl Acad Sci U S A*, **105**, 15553-15557.
1447 87. Heydanek, M.G., Jr. and Neuhaus, F.C. (1969) The initial stage in peptidoglycan synthesis.
1448 IV. Solubilization of phospho-N-acetylmuramyl-pentapeptide translocase. *Biochemistry*, **8**,
1449 1474-1481.
1450 88. Mengin-Lecreulx, D. and van Heijenoort, J. (1993) Identification of the glmU gene
1451 encoding N-acetylglucosamine-1-phosphate uridyltransferase in Escherichia coli. *J*
1452 *Bacteriol*, **175**, 6150-6157.
1453 89. Mengin-Lecreulx, D. and van Heijenoort, J. (1996) Characterization of the essential gene
1454 glmM encoding phosphoglucosamine mutase in Escherichia coli. *J Biol Chem*, **271**, 32-39.
1455 90. van Heijenoort, J. (2007) Lipid intermediates in the biosynthesis of bacterial
1456 peptidoglycan. *Microbiol Mol Biol Rev*, **71**, 620-635.
1457 91. Ashraf, K.U., Nygaard, R., Vickery, O.N., Erramilli, S.K., Herrera, C.M., McConville, T.H.,
1458 Petrou, V.I., Giacometti, S.I., Dufrisne, M.B., Nosol, K., Zinkle, A.P., Graham, C.L.B., Loukeris,
1459 M., Kloss, B., Skorupinska-Tudek, K., Swiezewska, E., Roper, D.I., Clarke, O.B., Uhlemann,

1460 A.C., Kossiakoff, A.A., Trent, M.S., Stansfeld, P.J. and Mancia, F. (2022) Structural basis of
1461 lipopolysaccharide maturation by the O-antigen ligase. *Nature*, **604**, 371-376.
1462 92. Franco, A.V., Liu, D. and Reeves, P.R. (1998) The wzz (cld) protein in Escherichia coli: amino
1463 acid sequence variation determines O-antigen chain length specificity. *J Bacteriol*, **180**,
1464 2670-2675.
1465 93. Rai, A.K., Carr, J.F., Bautista, D.E., Wang, W. and Mitchell, A.M. (2021) ElyC and Cyclic
1466 Enterobacterial Common Antigen Regulate Synthesis of Phosphoglyceride-Linked
1467 Enterobacterial Common Antigen. *mBio*, **12**, e0284621.
1468 94. Kumar, A.S., Mody, K. and Jha, B. (2007) Bacterial exopolysaccharides--a perception. *J*
1469 *Basic Microbiol*, **47**, 103-117.
1470 95. van Heijenoort, J. (2001) Formation of the glycan chains in the synthesis of bacterial
1471 peptidoglycan. *Glycobiology*, **11**, 25R-36R.
1472 96. Hong, Y., Hu, D., Verderosa, A.D., Qin, J., Totsika, M. and Reeves, P.R. (2023) Repeat-Unit
1473 Elongations To Produce Bacterial Complex Long Polysaccharide Chains, an O-Antigen
1474 Perspective. *EcoSal Plus*, **11**, eesp00202022.
1475 97. Rai, A.K. and Mitchell, A.M. (2020) Enterobacterial Common Antigen: Synthesis and
1476 Function of an Enigmatic Molecule. *mBio*, **11**.
1477 98. Sham, L.T., Butler, E.K., Lebar, M.D., Kahne, D., Bernhardt, T.G. and Ruiz, N. (2014) Bacterial
1478 cell wall. MurJ is the flippase of lipid-linked precursors for peptidoglycan biogenesis.
1479 *Science*, **345**, 220-222.
1480 99. Kim, S., Pires, M.M. and Im, W. (2018) Insight into Elongation Stages of Peptidoglycan
1481 Processing in Bacterial Cytoplasmic Membranes. *Sci Rep*, **8**, 17704.
1482 100. Di Guilmi, A.M., Dessen, A., Dideberg, O. and Vernet, T. (2003) The glycosyltransferase
1483 domain of penicillin-binding protein 2a from Streptococcus pneumoniae catalyzes the
1484 polymerization of murein glycan chains. *J Bacteriol*, **185**, 4418-4423.
1485 101. Weckener, M., Woodward, L.S., Clarke, B.R., Liu, H., Ward, P.N., Le Bas, A., Bhella, D.,
1486 Whitfield, C. and Naismith, J.H. (2023) The lipid linked oligosaccharide polymerase Wzy
1487 and its regulating co-polymerase, Wzz, from enterobacterial common antigen
1488 biosynthesis form a complex. *Open Biol*, **13**, 220373.
1489 102. Han, W., Wu, B., Li, L., Zhao, G., Woodward, R., Pettit, N., Cai, L., Thon, V. and Wang, P.G.
1490 (2012) Defining function of lipopolysaccharide O-antigen ligase WaaL using
1491 chemoenzymatically synthesized substrates. *J Biol Chem*, **287**, 5357-5365.
1492 103. Imperiali, B. (2019) Bacterial carbohydrate diversity - a Brave New World. *Curr Opin Chem*
1493 *Biol*, **53**, 1-8.
1494 104. Mostowy, R.J. and Holt, K.E. (2018) Diversity-Generating Machines: Genetics of Bacterial
1495 Sugar-Coating. *Trends Microbiol*, **26**, 1008-1021.
1496 105. Saiki, K. and Konishi, K. (2007) Identification of a Porphyromonas gingivalis novel protein
1497 sov required for the secretion of gingipains. *Microbiol Immunol*, **51**, 483-491.
1498 106. Pares, S., Cohen-Addad, C., Sieker, L., Neuburger, M. and Douce, R. (1994) X-ray structure
1499 determination at 2.6-A resolution of a lipoate-containing protein: the H-protein of the
1500 glycine decarboxylase complex from pea leaves. *Proc Natl Acad Sci U S A*, **91**, 4850-4853.
1501 107. Shultis, D.D., Purdy, M.D., Banchs, C.N. and Wiener, M.C. (2006) Outer membrane active
1502 transport: structure of the BtuB:TonB complex. *Science*, **312**, 1396-1399.

1503 108. Rahn, A., Beis, K., Naismith, J.H. and Whitfield, C. (2003) A novel outer membrane protein,
1504 Wzi, is involved in surface assembly of the Escherichia coli K30 group 1 capsule. *J*
1505 *Bacteriol*, **185**, 5882-5890.
1506 109. Qi, Y. and Grishin, N.V. (2005) Structural classification of thioredoxin-like fold proteins.
1507 *Proteins*, **58**, 376-388.
1508 110. Suplatov, D.A., Besenmatter, W., Svedas, V.K. and Svendsen, A. (2012) Bioinformatic
1509 analysis of alpha/beta-hydrolase fold enzymes reveals subfamily-specific positions
1510 responsible for discrimination of amidase and lipase activities. *Protein Eng Des Sel*, **25**,
1511 689-697.
1512 111. Palomeque-Messia, P., Englebert, S., Leyh-Bouille, M., Nguyen-Disteche, M., Duez, C.,
1513 Houba, S., Dideberg, O., Van Beeumen, J. and Ghuysen, J.M. (1991) Amino acid sequence
1514 of the penicillin-binding protein/DD-peptidase of Streptomyces K15. Predicted
1515 secondary structures of the low Mr penicillin-binding proteins of class A. *Biochem J*, **279**
1516 **(Pt 1)**, 223-230.
1517 112. Aravind, L. (1999) An evolutionary classification of the metallo-beta-lactamase fold
1518 proteins. *In Silico Biol*, **1**, 69-91.
1519 113. Novinec, M. and Lenarcic, B. (2013) Papain-like peptidases: structure, function, and
1520 evolution. *Biomol Concepts*, **4**, 287-308.
1521 114. Razew, A., Schwarz, J.N., Mitkowski, P., Sabala, I. and Kaus-Drobek, M. (2022) One fold,
1522 many functions-M23 family of peptidoglycan hydrolases. *Front Microbiol*, **13**, 1036964.
1523 115. Mans, B.J., Anantharaman, V., Aravind, L. and Koonin, E.V. (2004) Comparative genomics,
1524 evolution and origins of the nuclear envelope and nuclear pore complex. *Cell Cycle*, **3**,
1525 1612-1637.
1526 116. Wimley, W.C. (2003) The versatile beta-barrel membrane protein. *Curr Opin Struct Biol*, **13**,
1527 404-411.
1528 117. Fairman, J.W., Noinaj, N. and Buchanan, S.K. (2011) The structural biology of beta-barrel
1529 membrane proteins: a summary of recent reports. *Curr Opin Struct Biol*, **21**, 523-531.
1530 118. Tokuda, H. and Matsuyama, S. (2004) Sorting of lipoproteins to the outer membrane in E.
1531 coli. *Biochim Biophys Acta*, **1693**, 5-13.
1532 119. Sanchez-Pulido, L., Devos, D., Genevrois, S., Vicente, M. and Valencia, A. (2003) POTRA: a
1533 conserved domain in the FtsQ family and a class of beta-barrel outer membrane proteins.
1534 *Trends Biochem Sci*, **28**, 523-526.
1535 120. Kim, S., Malinverni, J.C., Sliz, P., Silhavy, T.J., Harrison, S.C. and Kahne, D. (2007) Structure
1536 and function of an essential component of the outer membrane protein assembly
1537 machine. *Science*, **317**, 961-964.
1538 121. Oke, M., Sarra, R., Ghirlando, R., Farnaud, S., Gorringe, A.R., Evans, R.W. and Buchanan, S.K.
1539 (2004) The plug domain of a neisserial TonB-dependent transporter retains structural
1540 integrity in the absence of its transmembrane beta-barrel. *FEBS Lett*, **564**, 294-300.
1541 122. Carr, S., Penfold, C.N., Bamford, V., James, R. and Hemmings, A.M. (2000) The structure of
1542 TolB, an essential component of the tol-dependent translocation system, and its protein-
1543 protein interaction with the translocation domain of colicin E9. *Structure*, **8**, 57-66.
1544 123. Ghosh, M., Tucker, D.E., Burchett, S.A. and Leslie, C.C. (2006) Properties of the Group IV
1545 phospholipase A2 family. *Prog Lipid Res*, **45**, 487-510.

124. Stukey, J. and Carman, G.M. (1997) Identification of a novel phosphatase sequence motif. *Protein Sci*, **6**, 469-472.
125. Whisstock, J.C., Romero, S., Gurung, R., Nandurkar, H., Ooms, L.M., Bottomley, S.P. and Mitchell, C.A. (2000) The inositol polyphosphate 5-phosphatases and the apurinic/apyrimidinic base excision repair endonucleases share a common mechanism for catalysis. *J Biol Chem*, **275**, 37055-37061.
126. Burroughs, A.M. and Aravind, L. (2023) New biochemistry in the Rhodanese-phosphatase superfamily: emerging roles in diverse metabolic processes, nucleic acid modifications, and biological conflicts. *NAR Genom Bioinform*, **5**, lqad029.
127. Holtje, J.V., Mirelman, D., Sharon, N. and Schwarz, U. (1975) Novel type of murein transglycosylase in Escherichia coli. *J Bacteriol*, **124**, 1067-1076.
128. Yunck, R., Cho, H. and Bernhardt, T.G. (2016) Identification of MltG as a potential terminase for peptidoglycan polymerization in bacteria. *Mol Microbiol*, **99**, 700-718.
129. Einsle, O., Messerschmidt, A., Stach, P., Bourenkov, G.P., Bartunik, H.D., Huber, R. and Kroneck, P.M. (1999) Structure of cytochrome c nitrite reductase. *Nature*, **400**, 476-480.
130. Levin, S., Almo, S.C. and Satir, B.H. (1999) Functional diversity of the phosphoglucomutase superfamily: structural implications. *Protein Eng*, **12**, 737-746.
131. Dong, X., Kato-Murayama, M., Muramatsu, T., Mori, H., Shirouzu, M., Bessho, Y. and Yokoyama, S. (2007) The crystal structure of leucyl/phenylalanyl-tRNA-protein transferase from Escherichia coli. *Protein Sci*, **16**, 528-534.
132. Antia, M., Hoare, D.S. and Work, E. (1957) The stereoisomers of alpha epsilon-diaminopimelic acid. III. Properties and distribution of diaminopimelic acid racemase, an enzyme causing interconversion of the LL and meso isomers. *Biochem J*, **65**, 448-459.
133. Koraimann, G. (2003) Lytic transglycosylases in macromolecular transport systems of Gram-negative bacteria. *Cell Mol Life Sci*, **60**, 2371-2388.
134. Muley, V.Y., Akhter, Y. and Galande, S. (2019) PDZ Domains Across the Microbial World: Molecular Link to the Proteases, Stress Response, and Protein Synthesis. *Genome Biol Evol*, **11**, 644-659.
135. Hara, H., Yamamoto, Y., Higashitani, A., Suzuki, H. and Nishimura, Y. (1991) Cloning, mapping, and characterization of the Escherichia coli prc gene, which is involved in C-terminal processing of penicillin-binding protein 3. *J Bacteriol*, **173**, 4799-4813.
136. Wong, L.H. and Levine, T.P. (2017) Tubular lipid binding proteins (TULIPs) growing everywhere. *Biochim Biophys Acta Mol Cell Res*, **1864**, 1439-1449.
137. Levine, T.P. (2019) Remote homology searches identify bacterial homologues of eukaryotic lipid transfer proteins, including Chorein-N domains in TamB and AsmA and Mdm31p. *BMC Mol Cell Biol*, **20**, 43.
138. Dhanaraj, V., Ye, Q.Z., Johnson, L.L., Hupe, D.J., Ortwine, D.F., Dunbar, J.B., Jr., Rubin, J.R., Pavlovsky, A., Humblet, C. and Blundell, T.L. (1996) X-ray structure of a hydroxamate inhibitor complex of stromelysin catalytic domain and its comparison with members of the zinc metalloproteinase superfamily. *Structure*, **4**, 375-386.
139. Lo, Y.C., Lin, S.C., Shaw, J.F. and Liaw, Y.C. (2003) Crystal structure of Escherichia coli thioesterase I/protease I/lysophospholipase L1: consensus sequence blocks constitute the catalytic center of SGNH-hydrolases through a conserved hydrogen bond network. *J Mol Biol*, **330**, 539-551.

1590 140. Hofmann, K. (2000) A superfamily of membrane-bound O-acyltransferases with
1591 implications for wnt signaling. *Trends Biochem Sci*, **25**, 111-112.
1592 141. Rahlwes, K.C., Ha, S.A., Motooka, D., Mayfield, J.A., Baumoel, L.R., Strickland, J.N., Torres-
1593 Ocampo, A.P., Nakamura, S. and Morita, Y.S. (2017) The cell envelope-associated
1594 phospholipid-binding protein LmeA is required for mannan polymerization in
1595 mycobacteria. *J Biol Chem*, **292**, 17407-17417.
1596 142. Yeow, J. and Chng, S.S. (2022) Of zones, bridges and chaperones - phospholipid transport
1597 in bacterial outer membrane assembly and homeostasis. *Microbiology (Reading)*, **168**.
1598 143. Kadirvelraj, R., Foley, B.L., Dyekjaer, J.D. and Woods, R.J. (2008) Involvement of water in
1599 carbohydrate-protein binding: concanavalin A revisited. *J Am Chem Soc*, **130**, 16933-
1600 16942.
1601 144. Flint, J., Nurizzo, D., Harding, S.E., Longman, E., Davies, G.J., Gilbert, H.J. and Bolam, D.N.
1602 (2004) Ligand-mediated dimerization of a carbohydrate-binding molecule reveals a novel
1603 mechanism for protein-carbohydrate recognition. *J Mol Biol*, **337**, 417-426.
1604 145. Murzin, A.G. (1993) OB(oligonucleotide/oligosaccharide binding)-fold: common
1605 structural and functional solution for non-homologous sequences. *EMBO J*, **12**, 861-867.
1606 146. Williams, A.F. and Barclay, A.N. (1988) The immunoglobulin superfamily--domains for cell
1607 surface recognition. *Annu Rev Immunol*, **6**, 381-405.
1608 147. Chen, C.K., Chan, N.L. and Wang, A.H. (2011) The many blades of the beta-propeller
1609 proteins: conserved but versatile. *Trends Biochem Sci*, **36**, 553-561.
1610 148. Babu, M.M., Priya, M.L., Selvan, A.T., Madera, M., Gough, J., Aravind, L. and Sankaran, K.
1611 (2006) A database of bacterial lipoproteins (DOLOP) with functional assignments to
1612 predicted lipoproteins. *J Bacteriol*, **188**, 2761-2773.
1613 149. Sankaran, K. and Wu, H.C. (1994) Lipid modification of bacterial prolipoprotein. Transfer
1614 of diacylglyceryl moiety from phosphatidylglycerol. *J Biol Chem*, **269**, 19701-19706.
1615 150. Tjalsma, H., Zanen, G., Venema, G., Bron, S. and van Dijl, J.M. (1999) The potential active
1616 site of the lipoprotein-specific (type II) signal peptidase of Bacillus subtilis. *J Biol Chem*,
1617 **274**, 28191-28197.
1618 151. Notenboom, V., Boraston, A.B., Kilburn, D.G. and Rose, D.R. (2001) Crystal structures of the
1619 family 9 carbohydrate-binding module from Thermotoga maritima xylanase 10A in native
1620 and ligand-bound forms. *Biochemistry*, **40**, 6248-6256.
1621 152. Rigden, D.J. (2005) Analysis of glycoside hydrolase family 98: catalytic machinery,
1622 mechanism and a novel putative carbohydrate binding module. *FEBS Lett*, **579**, 5466-
1623 5472.
1624 153. Boraston, A.B., Bolam, D.N., Gilbert, H.J. and Davies, G.J. (2004) Carbohydrate-binding
1625 modules: fine-tuning polysaccharide recognition. *Biochem J*, **382**, 769-781.
1626 154. Iyer, L.M., Koonin, E.V. and Aravind, L. (2001) Adaptations of the helix-grip fold for ligand
1627 binding and catalysis in the START domain superfamily. *Proteins*, **43**, 134-144.
1628 155. Santelli, E., Liddington, R.C., Mohan, M.A., Hoch, J.A. and Szurmant, H. (2007) The crystal
1629 structure of Bacillus subtilis YycI reveals a common fold for two members of an unusual
1630 class of sensor histidine kinase regulatory proteins. *J Bacteriol*, **189**, 3290-3295.
1631 156. Das, K., Xiao, R., Wahlberg, E., Hsu, F., Arrowsmith, C.H., Montelione, G.T. and Arnold, E.
1632 (2001) X-ray crystal structure of MTH938 from Methanobacterium thermoautotrophicum
1633 at 2.2 A resolution reveals a novel tertiary protein fold. *Proteins*, **45**, 486-488.

1634 157. Bennett, H.J., Davenport, J.B., Collins, R.F., Trafford, A.W., Pinali, C. and Kitmitto, A. (2013)
1635 Human junctophilin-2 undergoes a structural rearrangement upon binding
1636 PtdIns(3,4,5)P3 and the S101R mutation identified in hypertrophic cardiomyopathy
1637 obviates this response. *Biochem J*, **456**, 205-217.
1638 158. Cortajarena, A.L. and Regan, L. (2006) Ligand binding by TPR domains. *Protein Sci*, **15**,
1639 1193-1198.
1640 159. Seed, K.D., Faruque, S.M., Mekalanos, J.J., Calderwood, S.B., Qadri, F. and Camilli, A. (2012)
1641 Phase variable O antigen biosynthetic genes control expression of the major protective
1642 antigen and bacteriophage receptor in Vibrio cholerae O1. *PLoS Pathog*, **8**, e1002917.
1643 160. Cai, R., Wang, G., Le, S., Wu, M., Cheng, M., Guo, Z., Ji, Y., Xi, H., Zhao, C., Wang, X., Xue, Y.,
1644 Wang, Z., Zhang, H., Fu, Y., Sun, C., Feng, X., Lei, L., Yang, Y., Ur Rahman, S., Liu, X., Han, W.
1645 and Gu, J. (2019) Three Capsular Polysaccharide Synthesis-Related Glucosyltransferases,
1646 GT-1, GT-2 and WcaJ, Are Associated With Virulence and Phage Sensitivity of Klebsiella
1647 pneumoniae. *Front Microbiol*, **10**, 1189.
1648 161. Yau, K.W. (1994) Cyclic nucleotide-gated channels: an expanding new family of ion
1649 channels. *Proc Natl Acad Sci U S A*, **91**, 3481-3483.
1650 162. Durocher, D., Henckel, J., Fersht, A.R. and Jackson, S.P. (1999) The FHA domain is a
1651 modular phosphopeptide recognition motif. *Mol Cell*, **4**, 387-394.
1652 163. Schreiter, E.R. and Drennan, C.L. (2007) Ribbon-helix-helix transcription factors: variations
1653 on a theme. *Nat Rev Microbiol*, **5**, 710-720.
1654 164. Ravi, J., Anantharaman, V., Chen, S.Z., Brenner, E.P., Datta, P., Aravind, L. and Gennaro, M.L.
1655 (2024) The phage shock protein (PSP) envelope stress response: discovery of novel
1656 partners and evolutionary history. *mSystems*, **9**, e0084723.
1657 165. AhYoung, A.P., Koehl, A., Vizcarra, C.L., Cascio, D. and Egea, P.F. (2016) Structure of a
1658 putative ClpS N-end rule adaptor protein from the malaria pathogen Plasmodium
1659 falciparum. *Protein Sci*, **25**, 689-701.
1660 166. Durr, I.F. and Rudney, H. (1960) The reduction of beta-hydroxy-beta-methyl-glutaryl
1661 coenzyme A to mevalonic acid. *J Biol Chem*, **235**, 2572-2578.
1662 167. Tchen, T.T. (1958) Mevalonic kinase: purification and properties. *J Biol Chem*, **233**, 1100-
1663 1103.
1664 168. Wolff, M., Seemann, M., Tse Sum Bui, B., Frapart, Y., Tritsch, D., Garcia Estrabot, A.,
1665 Rodriguez-Concepcion, M., Boronat, A., Marquet, A. and Rohmer, M. (2003) Isoprenoid
1666 biosynthesis via the methylerythritol phosphate pathway: the (E)-4-hydroxy-3-methylbut-
1667 2-enyl diphosphate reductase (LytB/IspH) from Escherichia coli is a [4Fe-4S] protein. *FEBS*
1668 *Lett*, **541**, 115-120.
1669 169. Guo, R.T., Ko, T.P., Chen, A.P., Kuo, C.J., Wang, A.H. and Liang, P.H. (2005) Crystal structures
1670 of undecaprenyl pyrophosphate synthase in complex with magnesium, isopentenyl
1671 pyrophosphate, and farnesyl thiopyrophosphate: roles of the metal ion and conserved
1672 residues in catalysis. *J Biol Chem*, **280**, 20762-20774.
1673 170. Burroughs, A.M. and Aravind, L. (2016) RNA damage in biological conflicts and the
1674 diversity of responding RNA repair systems. *Nucleic Acids Res*, **44**, 8525-8555.
1675 171. Ueta, M., Ohniwa, R.L., Yoshida, H., Maki, Y., Wada, C. and Wada, A. (2008) Role of HPF
1676 (hibernation promoting factor) in translational activity in Escherichia coli. *J Biochem*, **143**,
1677 425-433.

1678 172. Akanuma, G., Kazo, Y., Tagami, K., Hiraoka, H., Yano, K., Suzuki, S., Hanai, R., Nanamiya, H.,
1679 Kato-Yamada, Y. and Kawamura, F. (2016) Ribosome dimerization is essential for the
1680 efficient regrowth of Bacillus subtilis. *Microbiology (Reading)*, **162**, 448-458.
1681 173. Kuk, A.C.Y., Hao, A. and Lee, S.Y. (2022) Structure and Mechanism of the Lipid Flippase
1682 MurJ. *Annu Rev Biochem*, **91**, 705-729.
1683 174. Zhang, D., de Souza, R.F., Anantharaman, V., Iyer, L.M. and Aravind, L. (2012) Polymorphic
1684 toxin systems: Comprehensive characterization of trafficking modes, processing,
1685 mechanisms of action, immunity and ecology using comparative genomics. *Biol Direct*, **7**,
1686 18.
1687 175. Iyer, L.M., Zhang, D., Rogozin, I.B. and Aravind, L. (2011) Evolution of the deaminase fold
1688 and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from
1689 bacterial toxin systems. *Nucleic Acids Res*, **39**, 9473-9497.
1690 176. Ruhe, Z.C., Low, D.A. and Hayes, C.S. (2020) Polymorphic Toxins and Their Immunity
1691 Proteins: Diversity, Evolution, and Mechanisms of Delivery. *Annu Rev Microbiol*, **74**, 497-
1692 520.
1693 177. Alcoforado Diniz, J. and Coulthurst, S.J. (2015) Intraspecies Competition in Serratia
1694 marcescens Is Mediated by Type VI-Secreted Rhs Effectors and a Conserved Effector-
1695 Associated Accessory Protein. *J Bacteriol*, **197**, 2350-2360.
1696 178. Satchell, K.J. (2011) Structure and function of MARTX toxins and other large repetitive
1697 RTX proteins. *Annu Rev Microbiol*, **65**, 71-90.
1698 179. Kordis, D. and Turk, V. (2009) Phylogenomic analysis of the cystatin superfamily in
1699 eukaryotes and prokaryotes. *BMC Evol Biol*, **9**, 266.
1700 180. Cherier, D., Patin, D., Blanot, D., Touze, T. and Barreteau, H. (2021) The Biology of Colicin M
1701 and Its Orthologs. *Antibiotics (Basel)*, **10**.
1702 181. Anantharaman, V. and Aravind, L. (2003) Evolutionary history, structural features and
1703 biochemical diversity of the NlpC/P60 superfamily of enzymes. *Genome Biol*, **4**, R11.
1704 182. Arthur, M., Molinas, C. and Courvalin, P. (1992) Sequence of the vanY gene required for
1705 production of a vancomycin-inducible D,D-carboxypeptidase in Enterococcus faecium
1706 BM4147. *Gene*, **120**, 111-114.
1707 183. Turk, V. and Bode, W. (1991) The cystatins: protein inhibitors of cysteine proteinases. *FEBS*
1708 *Lett*, **285**, 213-219.
1709 184. Sato, Y., Yamamoto, Y. and Kizaki, H. (1997) Cloning and sequence analysis of the gbpC
1710 gene encoding a novel glucan-binding protein of Streptococcus mutans. *Infect Immun*,
1711 **65**, 668-675.
1712 185. MacKenzie, D.A., Tailford, L.E., Hemmings, A.M. and Juge, N. (2009) Crystal structure of a
1713 mucus-binding protein repeat reveals an unexpected functional immunoglobulin binding
1714 activity. *J Biol Chem*, **284**, 32444-32453.
1715 186. Chagnot, C., Zorgani, M.A., Astruc, T. and Desvaux, M. (2013) Proteinaceous determinants
1716 of surface colonization in bacteria: bacterial adhesion and biofilm formation from a
1717 protein secretion perspective. *Front Microbiol*, **4**, 303.
1718 187. Wu, H.Y., Liu, M.S., Lin, T.P. and Cheng, Y.S. (2011) Structural and functional assays of
1719 AtTLP18.3 identify its novel acid phosphatase activity in thylakoid lumen. *Plant Physiol*,
1720 **157**, 1015-1025.

188. Cao, Z., Casabona, M.G., Kneuper, H., Chalmers, J.D. and Palmer, T. (2016) The type VII secretion system of Staphylococcus aureus secretes a nuclease toxin that targets competitor bacteria. *Nat Microbiol*, **2**, 16183.
189. Klein, T.A., Pazos, M., Surette, M.G., Vollmer, W. and Whitney, J.C. (2018) Molecular Basis for Immunity Protein Recognition of a Type VII Secretion System Exported Antibacterial Toxin. *J Mol Biol*, **430**, 4344-4358.
190. Knowles, T.J., Browning, D.F., Jeeves, M., Maderbocus, R., Rajesh, S., Sridhar, P., Manoli, E., Emery, D., Sommer, U., Spencer, A., Leyton, D.L., Squire, D., Chaudhuri, R.R., Viant, M.R., Cunningham, A.F., Henderson, I.R. and Overduin, M. (2011) Structure and function of BamE within the outer membrane and the beta-barrel assembly machine. *EMBO Rep*, **12**, 123-128.
191. Hagan, C.L. and Kahne, D. (2011) The reconstituted Escherichia coli Bam complex catalyzes multiple rounds of beta-barrel assembly. *Biochemistry*, **50**, 7444-7446.
192. Kitagawa, D., Vakonakis, I., Olieric, N., Hilbert, M., Keller, D., Olieric, V., Bortfeld, M., Erat, M.C., Fluckiger, I., Gonczy, P. and Steinmetz, M.O. (2011) Structural basis of the 9-fold symmetry of centrioles. *Cell*, **144**, 364-375.
193. Aravind, L., Anantharaman, V., Zhang, D., de Souza, R.F. and Iyer, L.M. (2012) Gene flow and biological conflict systems in the origin and evolution of eukaryotes. *Front Cell Infect Microbiol*, **2**, 89.
194. Mutschler, H., Gebhardt, M., Shoeman, R.L. and Meinhart, A. (2011) A novel mechanism of programmed cell death in bacteria by toxin-antitoxin systems corrupts peptidoglycan synthesis. *PLoS Biol*, **9**, e1001033.
195. Monzingo, A.F., Marcotte, E.M., Hart, P.J. and Robertus, J.D. (1996) Chitinases, chitosanases, and lysozymes can be divided into procaryotic and eucaryotic families sharing a conserved core. *Nat Struct Biol*, **3**, 133-140.
196. Dong, X., Leksa, N.C., Chhabra, E.S., Arndt, J.W., Lu, Q., Knockenhauer, K.E., Peters, R.T. and Springer, T.A. (2019) The von Willebrand factor D'D3 assembly and structural principles for factor VIII binding and concatemer biogenesis. *Blood*, **133**, 1523-1533.
197. West, A.H. and Stock, A.M. (2001) Histidine kinases and response regulator proteins in two-component signaling systems. *Trends Biochem Sci*, **26**, 369-376.
198. Costa, T., Isidro, A.L., Moran, C.P., Jr. and Henriques, A.O. (2006) Interaction between coat morphogenetic proteins SafA and SpoVID. *J Bacteriol*, **188**, 7731-7741.
199. Zhang, C., Kelkar, A., Nasirikenari, M., Lau, J.T.Y., Sveinsson, M., Sharma, U.C., Pokharel, S. and Neelamegham, S. (2018) The physical spacing between the von Willebrand factor D'D3 and A1 domains regulates platelet adhesion in vitro and in vivo. *J Thromb Haemost*, **16**, 571-582.
200. Hermoso, J.A., Monterroso, B., Albert, A., Galan, B., Ahrazem, O., Garcia, P., Martinez-Ripoll, M., Garcia, J.L. and Menendez, M. (2003) Structural basis for selective recognition of pneumococcal cell wall by modular endolysin from phage Cp-1. *Structure*, **11**, 1239-1249.
201. Das, D., Lee, W.S., Grant, J.C., Chiu, H.J., Farr, C.L., Vance, J., Klock, H.E., Knuth, M.W., Miller, M.D., Elsliger, M.A., Deacon, A.M., Godzik, A., Lesley, S.A., Kornfeld, S. and Wilson, I.A. (2013) Structure and function of the DUF2233 domain in bacteria and in the human mannose 6-phosphate uncovering enzyme. *J Biol Chem*, **288**, 16789-16799.

1764 202. Dideberg, O., Charlier, P., Dive, G., Joris, B., Frere, J.M. and Ghuysen, J.M. (1982) Structure
1765 of a Zn2+-containing D-alanyl-D-alanine-cleaving carboxypeptidase at 2.5 A resolution.
1766 *Nature*, **299**, 469-470.
1767 203. Bellinzoni, M., Haouz, A., Miras, I., Magnet, S., Andre-Leroux, G., Mukherjee, R., Shepard,
1768 W., Cole, S.T. and Alzari, P.M. (2014) Structural studies suggest a peptidoglycan hydrolase
1769 function for the Mycobacterium tuberculosis Tat-secreted protein Rv2525c. *J Struct Biol*,
1770 **188**, 156-164.
1771 204. Wright, G.D., Molinas, C., Arthur, M., Courvalin, P. and Walsh, C.T. (1992) Characterization
1772 of vanY, a DD-carboxypeptidase from vancomycin-resistant Enterococcus faecium
1773 BM4147. *Antimicrob Agents Chemother*, **36**, 1514-1518.
1774 205. Arthur, M., Depardieu, F., Cabanie, L., Reynolds, P. and Courvalin, P. (1998) Requirement of
1775 the VanY and VanX D,D-peptidases for glycopeptide resistance in enterococci. *Mol*
1776 *Microbiol*, **30**, 819-830.
1777 206. Arthur, M., Depardieu, F., Snaith, H.A., Reynolds, P.E. and Courvalin, P. (1994) Contribution
1778 of VanY D,D-carboxypeptidase to glycopeptide resistance in Enterococcus faecalis by
1779 hydrolysis of peptidoglycan precursors. *Antimicrob Agents Chemother*, **38**, 1899-1903.
1780 207. Meyer, B., Wurm, J.P., Sharma, S., Immer, C., Pogoryelov, D., Kotter, P., Lafontaine, D.L.,
1781 Wohnert, J. and Entian, K.D. (2016) Ribosome biogenesis factor Tsr3 is the
1782 aminocarboxypropyl transferase responsible for 18S rRNA hypermodification in yeast
1783 and humans. *Nucleic Acids Res*, **44**, 4304-4316.
1784 208. Burroughs, A.M. and Aravind, L. (2014) Analysis of two domains with novel RNA-
1785 processing activities throws light on the complex evolution of ribosomal RNA biogenesis.
1786 *Front Genet*, **5**, 424.
1787 209. Fraser, M.E., Joyce, M.A., Ryan, D.G. and Wolodko, W.T. (2002) Two glutamate residues,
1788 Glu 208 alpha and Glu 197 beta, are crucial for phosphorylation and dephosphorylation
1789 of the active-site histidine residue in succinyl-CoA synthetase. *Biochemistry*, **41**, 537-546.
1790 210. Iyer, L.M., Abhiman, S., Maxwell Burroughs, A. and Aravind, L. (2009) Amidoligases with
1791 ATP-grasp, glutamine synthetase-like and acetyltransferase-like domains: synthesis of
1792 novel metabolites and peptide modifications of proteins. *Mol Biosyst*, **5**, 1636-1660.
1793 211. Raetz, C.R. and Kennedy, E.P. (1974) Partial purification and properties of
1794 phosphatidylserine synthetase from Escherichia coli. *J Biol Chem*, **249**, 5083-5045.
1795 212. Fields, R.N. and Roy, H. (2018) Deciphering the tRNA-dependent lipid aminoacylation
1796 systems in bacteria: Novel components and structural advances. *RNA Biol*, **15**, 480-491.
1797 213. Hancock, R.E. (1997) Peptide antibiotics. *Lancet*, **349**, 418-422.
1798 214. Benson, T.E., Prince, D.B., Mutchler, V.T., Curry, K.A., Ho, A.M., Sarver, R.W., Hagadorn, J.C.,
1799 Choi, G.H. and Garlick, R.L. (2002) X-ray crystal structure of Staphylococcus aureus FemA.
1800 *Structure*, **10**, 1107-1115.
1801 215. Schneider, T., Senn, M.M., Berger-Bachi, B., Tossi, A., Sahl, H.G. and Wiedemann, I. (2004)
1802 In vitro assembly of a complete, pentaglycine interpeptide bridge containing cell wall
1803 precursor (lipid II-Gly5) of Staphylococcus aureus. *Mol Microbiol*, **53**, 675-685.
1804 216. Schmid, M., Uhlenhaut, N.H., Godard, F., Demar, M., Bressan, R., Weigel, D. and Lohmann,
1805 J.U. (2003) Dissection of floral induction pathways using global expression analysis.
1806 *Development*, **130**, 6001-6012.

1807 217. Berardini, T.Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E. and Huala, E. (2015) The
1808 Arabidopsis information resource: Making and mining the "gold standard" annotated
1809 reference plant genome. *Genesis*, **53**, 474-485.
1810 218. Savidov, N., Gloriozova, T.A., Poroikov, V.V. and Dembitsky, V.M. (2018) Highly oxygenated
1811 isoprenoid lipids derived from fungi and fungal endophytes: Origin and biological
1812 activities. *Steroids*, **140**, 114-124.
1813 219. Lee, J.O., Rieu, P., Arnaout, M.A. and Liddington, R. (1995) Crystal structure of the A
1814 domain from the alpha subunit of integrin CR3 (CD11b/CD18). *Cell*, **80**, 631-638.
1815 220. Zhang, Y., van der Fits, L., Voerman, J.S., Melief, M.J., Laman, J.D., Wang, M., Wang, H.,
1816 Wang, M., Li, X., Walls, C.D., Gupta, D. and Dziarski, R. (2005) Identification of serum N-
1817 acetylmuramoyl-l-alanine amidase as liver peptidoglycan recognition protein 2. *Biochim*
1818 *Biophys Acta*, **1752**, 34-46.
1819 221. Lee, J., Geddes, K., Streutker, C., Philpott, D.J. and Girardin, S.E. (2012) Role of mouse
1820 peptidoglycan recognition protein PGLYRP2 in the innate immune response to
1821 Salmonella enterica serovar Typhimurium infection in vivo. *Infect Immun*, **80**, 2645-2654.
1822 222. Dziarski, R. and Gupta, D. (2010) Review: Mammalian peptidoglycan recognition proteins
1823 (PGRPs) in innate immunity. *Innate Immun*, **16**, 168-174.
1824 223. Dziarski, R. and Gupta, D. (2006) Mammalian PGRPs: novel antibacterial proteins. *Cell*
1825 *Microbiol*, **8**, 1059-1069.
1826 224. Jones, J.D. and Dangl, J.L. (2006) The plant immune system. *Nature*, **444**, 323-329.
1827 225. Holzem, M., Boutros, M. and Holstein, T.W. (2024) The origin and evolution of Wnt
1828 signalling. *Nat Rev Genet*, **25**, 500-512.
1829 226. Morrow, J.F., Stearman, R.S., Peltzman, C.G. and Potter, D.A. (1981) Induction of hepatic
1830 synthesis of serum amyloid A protein and actin. *Proc Natl Acad Sci U S A*, **78**, 4718-4722.
1831 227. Hari-Dass, R., Shah, C., Meyer, D.J. and Raynes, J.G. (2005) Serum amyloid A protein binds
1832 to outer membrane protein A of gram-negative bacteria. *J Biol Chem*, **280**, 18562-18567.
1833 228. Shah, C., Hari-Dass, R. and Raynes, J.G. (2006) Serum amyloid A is an innate immune
1834 opsonin for Gram-negative bacteria. *Blood*, **108**, 1751-1757.
1835 229. Sack, G.H., Jr., Talbot, C.C., Jr., Seuanez, H. and O'Brien, S.J. (1989) Molecular analysis of
1836 the human serum amyloid A (SAA) gene family. *Scand J Immunol*, **29**, 113-119.
1837 230. Uhlar, C.M., Burgess, C.J., Sharp, P.M. and Whitehead, A.S. (1994) Evolution of the serum
1838 amyloid A (SAA) protein superfamily. *Genomics*, **19**, 228-235.
1839 231. Sun, L. and Ye, R.D. (2016) Serum amyloid A1: Structure, function and gene polymorphism.
1840 *Gene*, **583**, 48-57.
1841 232. Aoki, S.K., Malinverni, J.C., Jacoby, K., Thomas, B., Pamma, R., Trinh, B.N., Remers, S., Webb,
1842 J., Braaten, B.A., Silhavy, T.J. and Low, D.A. (2008) Contact-dependent growth inhibition
1843 requires the essential outer membrane protein BamA (YaeT) as the receptor and the inner
1844 membrane transport protein AcrB. *Mol Microbiol*, **70**, 323-340.
1845 233. Virtanen, P., Waneskog, M. and Koskiniemi, S. (2019) Class II contact-dependent growth
1846 inhibition (CDI) systems allow for broad-range cross-species toxin delivery within the
1847 Enterobacteriaceae family. *Mol Microbiol*, **111**, 1109-1125.
1848 234. Ruhe, Z.C., Nguyen, J.Y., Xiong, J., Koskiniemi, S., Beck, C.M., Perkins, B.R., Low, D.A. and
1849 Hayes, C.S. (2017) CdiA Effectors Use Modular Receptor-Binding Domains To Recognize
1850 Target Bacteria. *mBio*, **8**.

1851 235. Levin, M., Franklin, E.C., Frangione, B. and Pras, M. (1972) The amino acid sequence of a
1852 major nonimmunoglobulin component of some amyloid fibrils. *J Clin Invest*, **51**, 2773-
1853 2776.
1854 236. Blanco, L.P., Evans, M.L., Smith, D.R., Badtke, M.P. and Chapman, M.R. (2012) Diversity,
1855 biogenesis and function of microbial amyloids. *Trends Microbiol*, **20**, 66-73.
1856 237. Rouse, S.L., Matthews, S.J. and Dueholm, M.S. (2018) Ecology and Biogenesis of
1857 Functional Amyloids in Pseudomonas. *J Mol Biol*, **430**, 3685-3695.
1858 238. Tukel, C., Wilson, R.P., Nishimori, J.H., Pezeshki, M., Chromy, B.A. and Baumler, A.J. (2009)
1859 Responses to amyloids of microbial and host origin are mediated through toll-like
1860 receptor 2. *Cell Host Microbe*, **6**, 45-53.
1861 239. Prosswimmer, T., Heng, A. and Daggett, V. (2024) Mechanistic insights into the role of
1862 amyloid-beta in innate immunity. *Sci Rep*, **14**, 5376.
1863 240. Egan, A.J.F., Errington, J. and Vollmer, W. (2020) Regulation of peptidoglycan synthesis
1864 and remodelling. *Nat Rev Microbiol*, **18**, 446-460.
1865 241. Koga, Y. and Morii, H. (2007) Biosynthesis of ether-type polar lipids in archaea and
1866 evolutionary considerations. *Microbiol Mol Biol Rev*, **71**, 97-120.
1867 242. Daiyasu, H., Kuma, K., Yokoi, T., Morii, H., Koga, Y. and Toh, H. (2005) A study of archaeal
1868 enzymes involved in polar lipid synthesis linking amino acid sequence information,
1869 genomic contexts and lipid composition. *Archaea*, **1**, 399-410.
1870 243. Carman, G.M. and Han, G.S. (2006) Roles of phosphatidate phosphatase enzymes in lipid
1871 metabolism. *Trends Biochem Sci*, **31**, 694-699.
1872 244. Neuwald, A.F. (1997) An unexpected structural relationship between integral membrane
1873 phosphatases and soluble haloperoxidases. *Protein Sci*, **6**, 1764-1767.
1874 245. Aravind, L., Nicastro, G.G., Iyer, L.M. and Burroughs, A.M. (2024) The Prokaryotic Roots of
1875 Eukaryotic Immune Systems. *Annu Rev Genet*, **58**, 365-389.
1876 246. Krishnan, A., Iyer, L.M., Holland, S.J., Boehm, T. and Aravind, L. (2018) Diversification of
1877 AID/APOBEC-like deaminases in metazoa: multiplicity of clades and widespread roles in
1878 immunity. *Proc Natl Acad Sci U S A*, **115**, E3201-E3210.
1879 247. Zhang, D., Iyer, L.M. and Aravind, L. (2011) A novel immunity system for bacterial nucleic
1880 acid degrading toxins and its recruitment in various eukaryotic and DNA viral systems.
1881 *Nucleic Acids Res*, **39**, 4532-4552.
1882 248. Kaur, G., Burroughs, A.M., Iyer, L.M. and Aravind, L. (2020) Highly regulated, diversifying
1883 NTP-dependent biological conflict systems with implications for the emergence of
1884 multicellularity. *Elife*, **9**.
1885 249. Kaur, G., Iyer, L.M., Burroughs, A.M. and Aravind, L. (2021) Bacterial death and TRADD-N
1886 domains help define novel apoptosis and immunity mechanisms shared by prokaryotes
1887 and metazoans. *Elife*, **10**.
1888 250. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J.
1889 (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search
1890 programs. *Nucleic Acids Res*, **25**, 3389-3402.
1891 251. Johnson, L.S., Eddy, S.R. and Portugaly, E. (2010) Hidden Markov model speed heuristic
1892 and iterative HMM search procedure. *BMC Bioinformatics*, **11**, 431.
1893 252. Sayers, E.W., Bolton, E.E., Brister, J.R., Canese, K., Chan, J., Comeau, D.C., Connor, R., Funk,
1894 K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C., Lathrop, S., Lu, Z.,

Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., Tse, T., Wang, J., Williams, R., Trawick, B.W., Pruitt, K.D. and Sherry, S.T. (2022) Database resources of the national center for biotechnology information. *Nucleic Acids Res*, **50**, D20-D26.

253. Hauser, M., Steinegger, M. and Soding, J. (2016) MMseqs software suite for fast and deep clustering and searching of large protein sequence sets. *Bioinformatics*, **32**, 1323-1330.

254. Katoh, K. and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*, **30**, 772-780.

255. Steinegger, M., Meier, M., Mirdita, M., Vohringer, H., Haunsberger, S.J. and Soding, J. (2019) HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics*, **20**, 473.

256. Csardi, G. and Nepusz, T. (2006) The Igraph Software Package for Complex Network Research. *InterJournal, Complex Systems*.

257. Hagberg, A.A., Schult, D.A., Swart, P. and Hagberg, J.M. (2008), *Proceedings of the Python in Science Conference*.

258. Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J. and Bateman, A. (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*, **44**, D279-285.

259. Schaffer, A.A., Wolf, Y.I., Ponting, C.P., Koonin, E.V., Aravind, L. and Altschul, S.F. (1999) IMPALA: matching a protein sequence against a collection of PSI-BLAST-constructed position-specific score matrices. *Bioinformatics*, **15**, 1000-1011.

260. Eddy, S.R. (2011) Accelerated Profile HMM Searches. *PLoS Comput Biol*, **7**, e1002195.

261. Soding, J., Biegert, A. and Lupas, A.N. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res*, **33**, W244-248.

262. Berman, H., Henrick, K., Nakamura, H. and Markley, J.L. (2007) The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res*, **35**, D301-303.

263. Price, M.N., Dehal, P.S. and Arkin, A.P. (2010) FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One*, **5**, e9490.

264. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A. and Lanfear, R. (2020) IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol*, **37**, 1530-1534.

265. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., Bodenstein, S.W., Evans, D.A., Hung, C.C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Zemgulyte, A., Arvaniti, E., Beattie, C., Bertolli, O., Bridgland, A., Cherepanov, A., Congreve, M., Cowen-Rivers, A.I., Cowie, A., Figurnov, M., Fuchs, F.B., Gladman, H., Jain, R., Khan, Y.A., Low, C.M.R., Perlin, K., Potapenko, A., Savy, P., Singh, S., Stecula, A., Thillaisundaram, A., Tong, C., Yakneen, S., Zhong, E.D., Zielinski, M., Zidek, A., Bapst, V., Kohli, P., Jaderberg, M., Hassabis, D. and Jumper, J.M. (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, **630**, 493-500.

266. Sehnal, D., Bittrich, S., Deshpande, M., Svobodova, R., Berka, K., Bazgier, V., Velankar, S., Burley, S.K., Koca, J. and Rose, A.S. (2021) Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Res*, **49**, W431-W437.

267. Holm, L. (2019) Benchmarking fold detection by DaliLite v.5. *Bioinformatics*, **35**, 5326-5327.
268. van Kempen, M., Kim, S.S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C.L.M., Soding, J. and Steinegger, M. (2024) Fast and accurate protein structure search with Foldseek. *Nat Biotechnol*, **42**, 243-246.

Figure 1

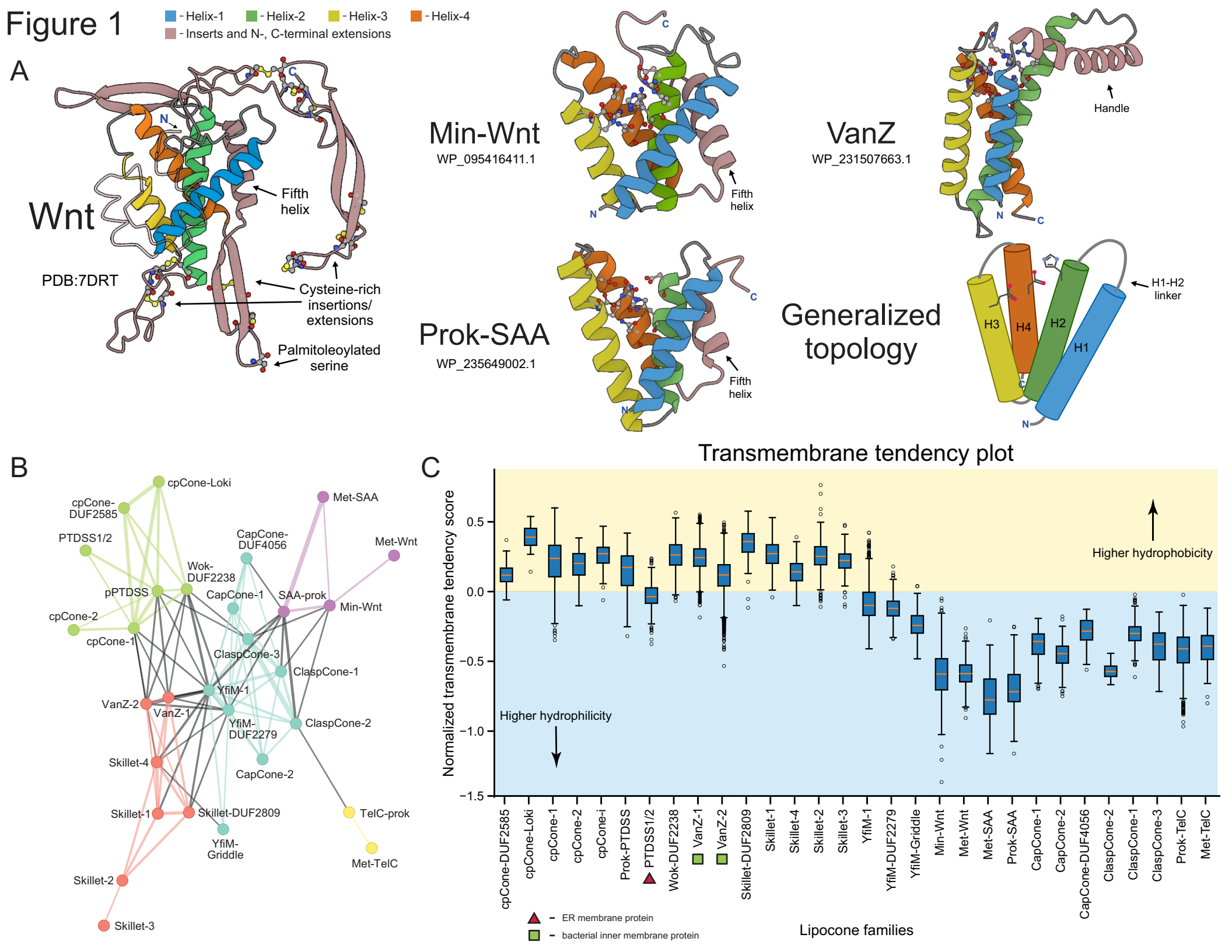
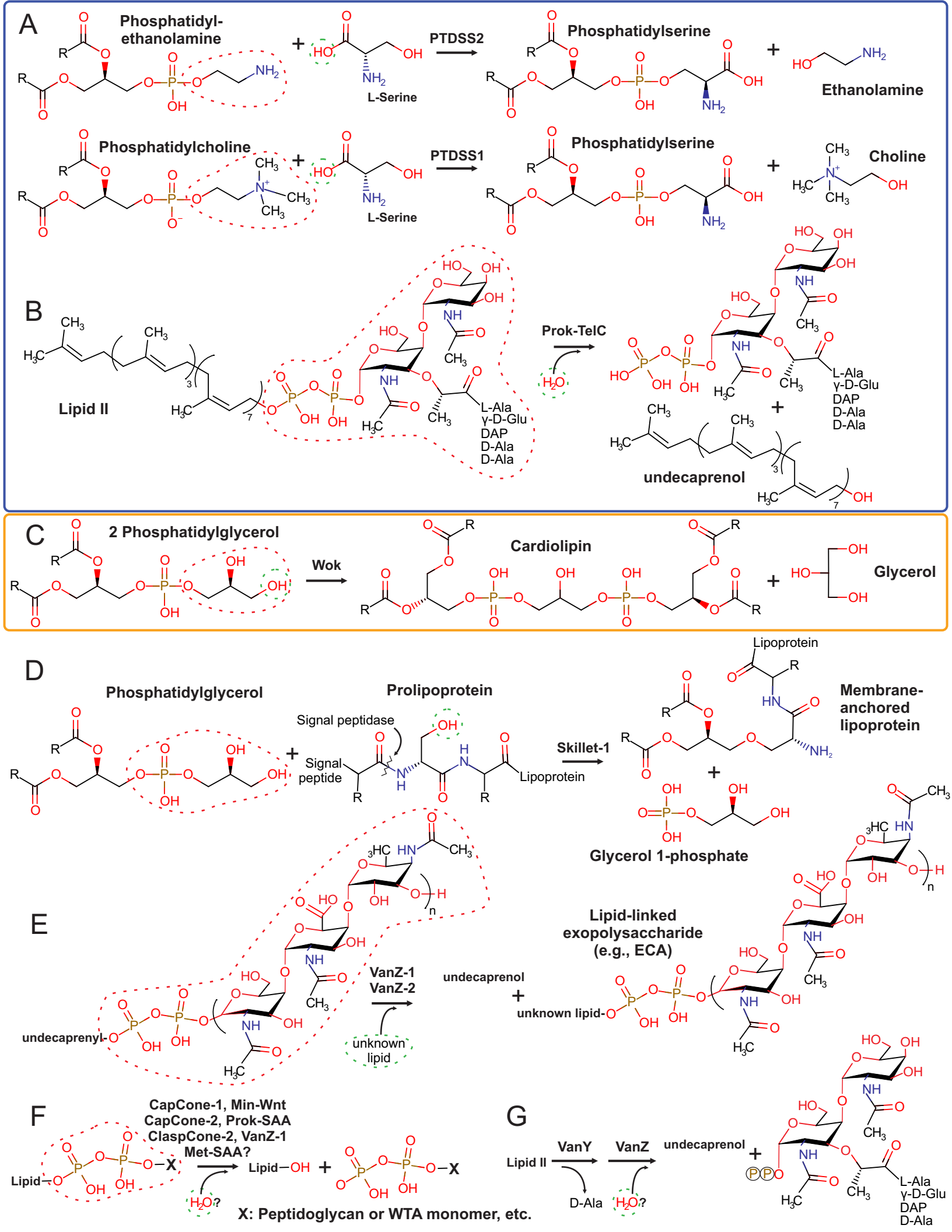
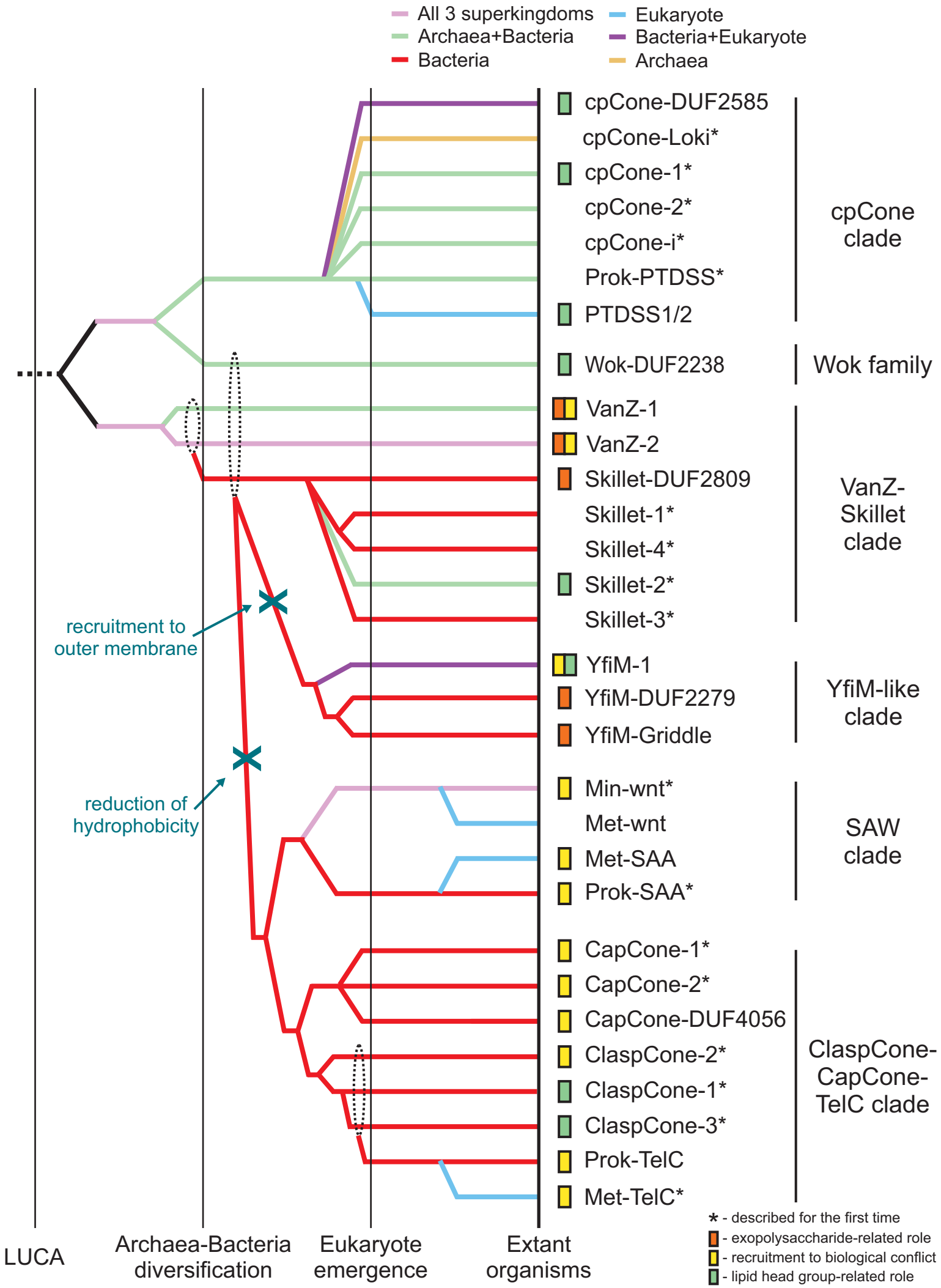

- Helix-1
- Helix-2
- Helix-3
- Helix-4
- Inserts and N-, C-terminal extensions
A
Wnt
PDB:7DRT
Fifth helix
Cysteine-rich insertions/ extensions
Palmitoleoylated serine
Min-Wnt
WP_095416411.1
Fifth helix
VanZ
WP_231507663.1
Handle
Prok-SAA
WP_235649002.1
Fifth helix
Generalized topology
H1-H2 linker
H3
H4
H2
H1
B
cpCone-Loki
cpCone-DUF2585
PTDSS1/2
Met-SAA
CapCone-DUF4056
Met-Wnt
Wok-DUF2238
pPTDSS
CapCone-1
SAA-prok
Min-Wnt
cpCone-2
cpCone-1
ClaspCone-3
ClaspCone-1
YfiM-1
VanZ-2
VanZ-1
YfiM-DUF2279
ClaspCone-2
Skillet-4
CapCone-2
Skillet-1
Skillet-DUF2809
TelC-prok
YfiM-Griddle
Met-TelC
Skillet-2
Skillet-3
C
Transmembrane tendency plot
Normalized transmembrane tendency score
0.5
0.0
−0.5
−1.0
−1.5
Higher hydrophobicity
Higher hydrophilicity
cpCone-DUF2585
cpCone-Loki
cpCone-1
cpCone-2
cpCone-i
Prok-PTDSS
PTDSS1/2
Wok-DUF2238
VanZ-1
VanZ-2
Skillet-DUF2809
Skillet-1
Skillet-4
Skillet-2
Skillet-3
YfiM-1
YfiM-DUF2279
YfiM-Griddle
Min-Wnt
Met-Wnt
Met-SAA
Prok-SAA
CapCone-1
CapCone-2
CapCone-DUF4056
ClaspCone-2
ClaspCone-1
ClaspCone-3
Prok-TelC
Met-TelC
ER membrane protein
bacterial inner membrane protein
Lipocone families

Figure 2

Helix 2
Helix 3
Helix 4
SAW (SAA-Wnt) clade
Met-Wnt
Min-Wnt
Met-SAA
SAA-prok
VanZ-Skillet clade
VanZ-1
VanZ-2
Skillet-DUF2809
Skillet-1
Skillet-2
Skillet-3
Skillet-4
YfiM clade
YfiM-1
YfiM-Griddle-DUF3943
YfiM-DUF2279
ClaspCone-TelC clade
ClaspCone-1
ClaspCone-2
ClaspCone-3
CapCone-DUF4056
CapCone-1
CapCone-2
TelC-prok
Met-TelC
cpCone clade
cpCone-DUF2585
cpCone-1
cpCone-2
cpCone-Loki
cpCone-i
pPTDSS
pTDSS1/2
Wok-DUF2238 family
Wok-DUF2238


A
Phosphatidyl-ethanolamine
L-Serine
PTDSS2
Phosphatidylserine
Ethanolamine
Phosphatidylcholine
L-Serine
PTDSS1
Phosphatidylserine
Choline
B
Lipid II
Prok-TelC
L-Ala
γ-D-Glu
DAP
D-Ala
D-Ala
undecaprenol
C
2 Phosphatidylglycerol
Wok
Cardiolipin
Glycerol
D
Phosphatidylglycerol
Prolipoprotein
Signal peptidase
Signal peptide
Lipoprotein
Skillet-1
Membrane-anchored lipoprotein
Glycerol 1-phosphate
E
undecaprenyl-
VanZ-1
VanZ-2
unknown lipid
undecaprenol
unknown lipid-
Lipid-linked exopolysaccharide (e.g., ECA)
F
CapCone-1, Min-Wnt
CapCone-2, Prok-SAA
ClaspCone-2, VanZ-1
Met-SAA?
Lipid
X: Peptidoglycan or WTA monomer, etc.
G
Lipid II
VanY
VanZ
D-Ala
undecaprenol
L-Ala
γ-D-Glu
DAP
D-Ala


All 3 superkingdoms
Archaea+Bacteria
Bacteria
Eukaryote
Bacteria+Eukaryote
Archaea
cpCone-DUF2585
cpCone-Loki*
cpCone-1*
cpCone-2*
cpCone-i*
Prok-PTDSS*
PTDSS1/2
cpCone clade
Wok-DUF2238
Wok family
VanZ-1
VanZ-2
Skillet-DUF2809
Skillet-1*
Skillet-4*
Skillet-2*
Skillet-3*
VanZ-Skillet clade
recruitment to outer membrane
YfiM-1
YfiM-DUF2279
YfiM-Griddle
YfiM-like clade
reduction of hydrophobicity
Min-wnt*
Met-wnt
Met-SAA
Prok-SAA*
SAW clade
CapCone-1*
CapCone-2*
CapCone-DUF4056
ClaspCone-2*
ClaspCone-1*
ClaspCone-3*
Prok-TelC
Met-TelC*
ClaspCone-CapCone-TelC clade
LUCA
Archaea-Bacteria diversification
Eukaryote emergence
Extant organisms
* - described for the first time
- exopolysaccharide-related role
- recruitment to biological conflict
- lipid head group-related role

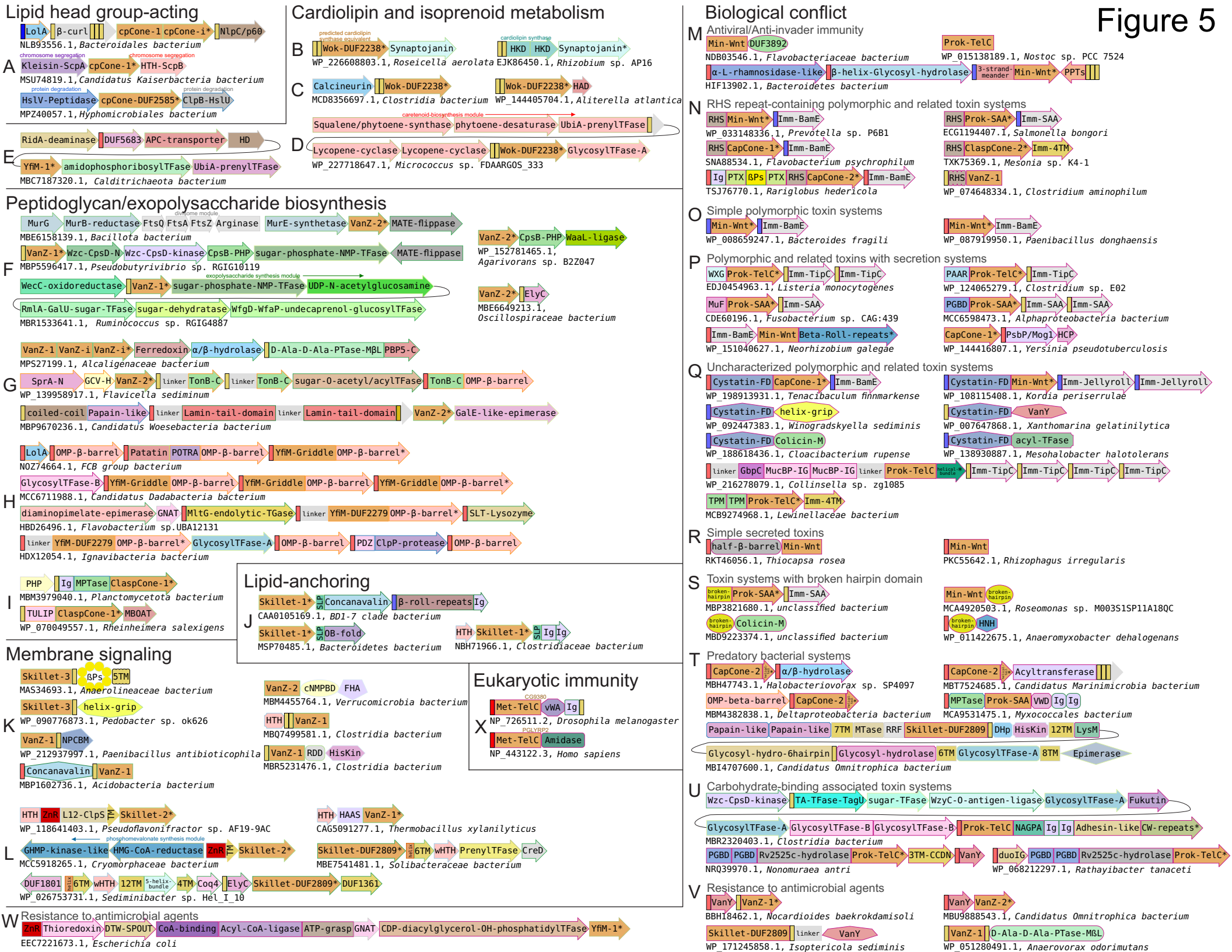


Figure 5
Lipid head group-acting
A
LolA β-curl cpCone-1 cpCone-1* NlpC/p60
NLB93556.1, Bacteroidales bacterium
chromosome segregation
chromosome segregation
Kleisin-ScpA cpCone-1* HTH-ScpB
MSU74819.1, Candidatus Kaiserbacteria bacterium
protein degradation
protein degradation
HslV-Peptidase cpCone-DUF2585* ClpB-HslU
MPZ40057.1, Hyphomicrobiales bacterium
Cardiolipin and isoprenoid metabolism
B
predicted cardiolipin synthase equivalent
Wok-DUF2238* Synaptojanin
WP_226608803.1, Roseicella aerolata
cardiolipin synthase
HKD HKD Synaptojanin*
EJK86450.1, Rhizobium sp. AP16
C
Calcineurin Wok-DUF2238*
MCD8356697.1, Clostridia bacterium
Wok-DUF2238* HAD
WP_144405704.1, Aliterella atlantica
caretenoid-biosynthesis module
D
Squalene/phytoene-synthase phytoene-desaturase UbiA-prenylTFase
Lycopene-cyclase Lycopene-cyclase Wok-DUF2238* GlycosylTFase-A
WP_227718647.1, Micrococcus sp. FDAARGOS_333
E
RidA-deaminase DUF5683 APC-transporter HD
YfiM-1* amidophosphoribosylTFase UbiA-prenylTFase
MBC7187320.1, Calditrichaeota bacterium
Peptidoglycan/exopolysaccharide biosynthesis
F
divisome module
MurG MurB-reductase FtsQ FtsA FtsZ Arginase MurE-synthetase VanZ-2* MATE-flippase
MBE6158139.1, Bacillota bacterium
VanZ-1* Wzc-CpsD-N Wzc-CpsD-kinase CpsB-PHP sugar-phosphate-NMP-TFase MATE-flippase
MBP5596417.1, Pseudobutyrivibrio sp. RGIG10119
exopolysaccharide synthesis module
WecC-oxidoreductase VanZ-1* sugar-phosphate-NMP-TFase UDP-N-acetylglucosamine
RmlA-GalU-sugar-TFase sugar-dehydratase WfgD-WfaP-undecaprenol-glucosylTFase
MBR1533641.1, Ruminococcus sp. RGIG4887
VanZ-2* CpsB-PHP WaaL-ligase
WP_152781465.1, Agarivorans sp. B2Z047
VanZ-2* ElyC
MBE6649213.1, Oscillospiraceae bacterium
G
VanZ-1 VanZ-i VanZ-i* Ferredoxin α/β-hydrolase D-Ala-D-Ala-PTase-MβL PBP5-C
MPS27199.1, Alcaligenaceae bacterium
SprA-N GCV-H VanZ-2* linker TonB-C linker TonB-C sugar-O-acetyl/acylTFase TonB-C OMP-β-barrel
WP_139958917.1, Flavicella sediminum
coiled-coil Papain-like linker Lamin-tail-domain linker Lamin-tail-domain VanZ-2* GalE-like-epimerase
MBP9670236.1, Candidatus Woesebacteria bacterium
H
LolA OMP-β-barrel Patatin POTRA OMP-β-barrel YfiM-Griddle OMP-β-barrel*
NOZ74664.1, FCB group bacterium
GlycosylTFase-B YfiM-Griddle OMP-β-barrel YfiM-Griddle OMP-β-barrel YfiM-Griddle OMP-β-barrel*
MCC6711988.1, Candidatus Dadabacteria bacterium
diaminopimelate-epimerase GNAT MltG-endolytic-TGase linker YfiM-DUF2279 OMP-β-barrel* SLT-Lysozyme
HBD26496.1, Flavobacterium sp.UBA12131
linker YfiM-DUF2279 OMP-β-barrel* GlycosylTFase-A OMP-β-barrel PDZ ClpP-protease OMP-β-barrel
HDX12054.1, Ignavibacteria bacterium
I
PHP Ig MPTase ClaspCone-1*
MBM3979040.1, Planctomycetota bacterium
TULIP ClaspCone-1* MBOAT
WP_070049557.1, Rheinheimera salexigens
Lipid-anchoring
J
Skillet-1* SLP Concanavalin β-roll-repeats Ig
CAA0105169.1, BD1-7 clade bacterium
Skillet-1* SLP OB-fold
MSP70485.1, Bacteroidetes bacterium
HTH Skillet-1* SLP Ig Ig
NBH71966.1, Clostridiaceae bacterium
Membrane signaling
K
Skillet-3 βPs 5TM
MAS34693.1, Anaerolineaceae bacterium
Skillet-3 helix-grip
WP_090776873.1, Pedobacter sp. ok626
VanZ-1 NPCBM
WP_212937997.1, Paenibacillus antibioticophila
Concanavalin VanZ-1
MBP1602736.1, Acidobacteria bacterium
VanZ-2 cNMPBD FHA
MBM4455764.1, Verrucomicrobia bacterium
HTH VanZ-1
MBQ7499581.1, Clostridia bacterium
VanZ-1 RDD HisKin
MBR5231476.1, Clostridia bacterium
Eukaryotic immunity
X
CG9380
Met-TelC vWA Ig
NP_726511.2, Drosophila melanogaster
PGLYRP2
Met-TelC Amidase
NP_443122.3, Homo sapiens
L
HTH ZnR L12-ClpS ⅀ Skillet-2*
WP_118641403.1, Pseudoflavonifractor sp. AF19-9AC
phosphomevalonate synthesis module
GHMP-kinase-like HMG-CoA-reductase ZnR ⅀ Skillet-2*
MCC5918265.1, Cryomorphaceae bacterium
HTH HAAS VanZ-1*
CAG5091277.1, Thermobacillus xylanilyticus
Skillet-DUF2809* HelX 6TM wHTH PrenylTFase CreD
MBE7541481.1, Solibacteraceae bacterium
DUF1801 HelX 6TM wHTH 12TM 5-helix-bundle 4TM Coq4 ElyC Skillet-DUF2809* DUF1361
WP_026753731.1, Sediminibacter sp. Hel_I_10
W
Resistance to antimicrobial agents
ZnR Thioredoxin DTW-SPOUT CoA-binding Acyl-CoA-ligase ATP-grasp GNAT CDP-diacylglycerol-OH-phosphatidylTFase YfiM-1*
EEC7221673.1, Escherichia coli
Biological conflict
M
Antiviral/Anti-invader immunity
Min-Wnt DUF3892
NDB03546.1, Flavobacteriaceae bacterium
Prok-TelC
WP_015138189.1, Nostoc sp. PCC 7524
α-L-rhamnosidase-like β-helix-Glycosyl-hydrolase 3-strand-meander Min-Wnt* PPTs
HIF13902.1, Bacteroidetes bacterium
N
RHS repeat-containing polymorphic and related toxin systems
RHS Min-Wnt* Imm-BamE
WP_033148336.1, Prevotella sp. P6B1
RHS Prok-SAA* Imm-SAA
ECG1194407.1, Salmonella bongori
RHS CapCone-1* Imm-BamE
SNA88534.1, Flavobacterium psychrophilum
RHS ClaspCone-2* Imm-4TM
TXK75369.1, Mesonia sp. K4-1
Ig PTX βPs PTX RHS CapCone-2* Imm-BamE
TSJ76770.1, Rariglobus hedericola
RHS VanZ-1
WP_074648334.1, Clostridium aminophilum
O
Simple polymorphic toxin systems
Min-Wnt* Imm-BamE
WP_008659247.1, Bacteroides fragili
Min-Wnt* Imm-BamE
WP_087919950.1, Paenibacillus donghaensis
P
Polymorphic and related toxins with secretion systems
WXG Prok-TelC* Imm-TipC Imm-TipC
EDJ0454963.1, Listeria monocytogenes
PAAR Prok-TelC* Imm-TipC
WP_124065279.1, Clostridium sp. E02
MuF Prok-SAA* Imm-SAA
CDE60196.1, Fusobacterium sp. CAG:439
PGBD Prok-SAA* Imm-SAA Imm-SAA
MCC6598473.1, Alphaproteobacteria bacterium
Imm-BamE Min-Wnt Beta-Roll-repeats*
WP_151040627.1, Neorhizobium galegae
CapCone-1* PsbP/Mog1 HCP
WP_144416807.1, Yersinia pseudotuberculosis
Q
Uncharacterized polymorphic and related toxin systems
Cystatin-FD CapCone-1* Imm-BamE
WP_198913931.1, Tenacibaculum finnmarkense
Cystatin-FD Min-Wnt* Imm-Jellyroll Imm-Jellyroll
WP_108115408.1, Kordia periserrulae
Cystatin-FD helix-grip
WP_092447383.1, Winogradskyella sediminis
Cystatin-FD VanY
WP_007647868.1, Xanthomarina gelatinilytica
Cystatin-FD Colicin-M
WP_188618436.1, Cloacibacterium rupense
Cystatin-FD acyl-TFase
WP_138930887.1, Mesohalobacter halotolerans
linker GbpC MucBP-IG MucBP-IG linker Prok-TelC helical-bundle* Imm-TipC Imm-TipC Imm-TipC Imm-TipC
WP_216278079.1, Collinsella sp. zg1085
TPM TPM Prok-TelC* Imm-4TM
MCB9274968.1, Lewinellaceae bacterium
R
Simple secreted toxins
half-β-barrel Min-Wnt
RKT46056.1, Thiocapsa rosea
Min-Wnt
PKC55642.1, Rhizophagus irregularis
S
Toxin systems with broken hairpin domain
broken-hairpin Prok-SAA* Imm-SAA
MBP3821680.1, unclassified bacterium
Min-Wnt broken-hairpin
MCA4920503.1, Roseomonas sp. M003S1SP11A18QC
broken-hairpin Colicin-M
MBD9223374.1, unclassified bacterium
broken-hairpin HNH
WP_011422675.1, Anaeromyxobacter dehalogenans
T
Predatory bacterial systems
CapCone-2 Prok-SAA* α/β-hydrolase
MBH47743.1, Halobacteriovorax sp. SP4097
CapCone-2 Prok-SAA* Acyltransferase
MBT7524685.1, Candidatus Marinimicrobia bacterium
OMP-beta-barrel CapCone-2 Prok-SAA*
MBM4382838.1, Deltaproteobacteria bacterium
MPTase Prok-SAA VWD Ig Ig
MCA9531475.1, Myxococcales bacterium
Papain-like Papain-like 7TM MTase RRF Skillet-DUF2809 DHp HisKin 12TM LysM
Glycosyl-hydro-6hairpin Glycosyl-hydrolase 6TM GlycosylTFase-A 8TM Epimerase
MBI4707600.1, Candidatus Omnitrophica bacterium
U
Carbohydrate-binding associated toxin systems
Wzc-CpsD-kinase TA-TFase-TagU sugar-TFase WzyC-O-antigen-ligase GlycosylTFase-A Fukutin
GlycosylTFase-A GlycosylTFase-B GlycosylTFase-B Prok-TelC NAGPA Ig Ig Adhesin-like CW-repeats*
MBR2320403.1, Clostridia bacterium
PGBD PGBD Rv2525c-hydrolase Prok-TelC* 3TM-CCDN VanY
NRQ39970.1, Nonomuraea antri
duoIG PGBD PGBD Rv2525c-hydrolase Prok-TelC*
WP_068212297.1, Rathayibacter tanaceti
V
Resistance to antimicrobial agents
VanY VanZ-1*
BBH18462.1, Nocardioides baekrokdamisoli
VanY VanZ-2*
MBU9888543.1, Candidatus Omnitrophica bacterium
Skillet-DUF2809 linker VanY
WP_171245858.1, Isoptericola sediminis
VanZ-1 D-Ala-D-Ala-PTase-MβL
WP_051280491.1, Anaerovorax odorimutans

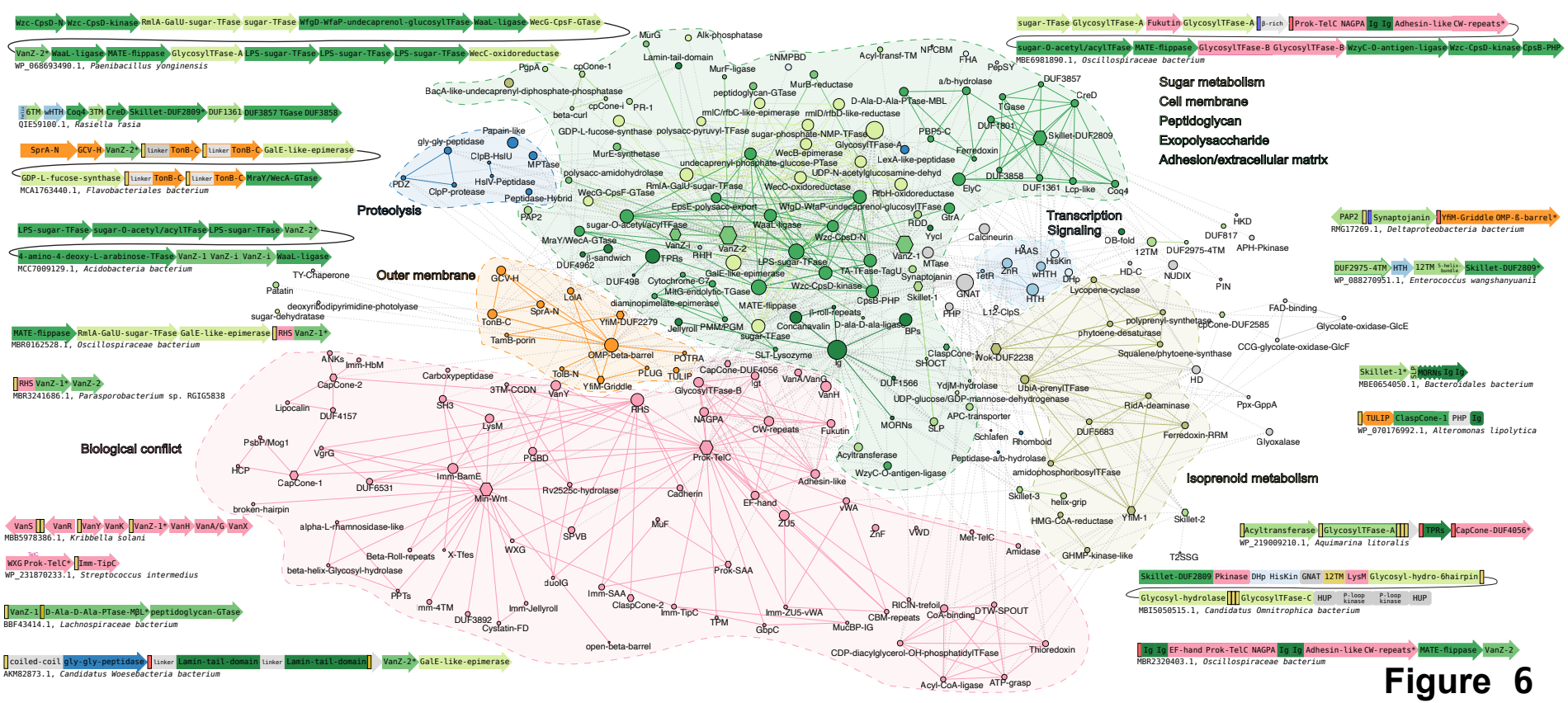

Sugar metabolism
Cell membrane
Peptidoglycan
Exopolysaccharide
Adhesion/extracellular matrix
Proteolysis
Transcription
Signaling
Outer membrane
Biological conflict
Isoprenoid metabolism
Wzc-CpsD-N Wzc-CpsD-kinase RmlA-GalU-sugar-TFase sugar-TFase WfgD-WfaP-undecaprenol-glucosyltFase WaaL-ligase WecG-CpsF-GTase
VanZ-2* WaaL-ligase MATE-flippase GlycosyltFase-A LPS-sugar-TFase LPS-sugar-TFase LPS-sugar-TFase WecC-oxidoreductase
WP_068693490.1, Paenibacillus yonginensis
6TM wHTH Coq4 3TM CreD Skillet-DUF2809 DUF1361 DUF3857 TGase DUF3858
QIE59100.1, Rasiella rasia
SprA-N GCV-H VanZ-2* linker TonB-C linker TonB-C GalE-like-epimerase
GDP-L-fucose-synthase linker TonB-C linker TonB-C MraY/WecA-GTase
MCA1763440.1, Flavobacteriales bacterium
LPS-sugar-TFase sugar-O-acetyl/acylTFase LPS-sugar-TFase VanZ-2*
4-amino-4-deoxy-L-arabinose-TFase VanZ-1 VanZ-i VanZ-i WaaL-ligase
MCC7009129.1, Acidobacteria bacterium
MATE-flippase RmlA-GalU-sugar-TFase GalE-like-epimerase RHS VanZ-1*
MBR0162528.1, Oscillospiraceae bacterium
RHS VanZ-1* VanZ-2
MBR3241686.1, Parasporobacterium sp. RGIG5838
VanS VanR VanY VanK VanZ-1* VanH VanA/G VanX
MBB5978386.1, Kribbella solani
WXG Prok-TelC* Imm-TipC
WP_231870233.1, Streptococcus intermedius
VanZ-i D-Ala-D-Ala-PTase-MBL peptidoglycan-GTase
BBF43414.1, Lachnospiraceae bacterium
coiled-coil gly-gly-peptidase linker Lamin-tail-domain linker Lamin-tail-domain VanZ-2* GalE-like-epimerase
AKM82873.1, Candidatus Woesebacteria bacterium
sugar-TFase GlycosyltFase-A Fukutin GlycosyltFase-A β-rich Prok-TelC NAGPA Ig Ig Adhesin-like CW-repeats*
sugar-O-acetyl/acylTFase MATE-flippase GlycosyltFase-B GlycosyltFase-B WzyC-O-antigen-ligase Wzc-CpsD-kinase CpsB-PHP
MBE6981890.1, Oscillospiraceae bacterium
PAP2 Synaptojanin YfbM-Griddle OMP-ß-barrel*
RMG17269.1, Deltaproteobacteria bacterium
DUF2975-4TM HTH 12TM 5-helix bundle Skillet-DUF2809*
WP_088270951.1, Enterococcus wangshanyuanii
Skillet-1* MORNs Ig Ig
MBE0654050.1, Bacteroidales bacterium
TULIP ClaspCone-1 PHP Ig
WP_070176992.1, Alteromonas lipolytica
Acyltransferase GlycosyltFase-A TPRs CapCone-DUF4056*
WP_219009210.1, Aquimarina litoralis
Skillet-DUF2809 Pkinase DHp HisKin GNAT 12TM LysM Glycosyl-hydro-6hairpin
Glycosyl-hydrolase GlycosyltFase-C HUP P-loop kinase P-loop kinase HUP
MBI5050515.1, Candidatus Omnitrophica bacterium
Ig Ig EF-hand Prok-TelC NAGPA Ig Ig Adhesin-like CW-repeats* MATE-flippase VanZ-2
MBR2320403.1, Oscillospiraceae bacterium

Figure 6