═══REVIEW COMMONS MANUSCRIPT═══

IMPORTANT:

- Manuscripts submitted to Review Commons are peer reviewed in a journal-agnostic way.
- Upon transfer of the peer reviewed preprint to a journal, the referee reports will be available in full to the handling editor.
- The identity of the referees will NOT be communicated to the authors unless the reviewers choose to sign their report.
- The identity of the referee will be confidentially disclosed to any affiliate journals to which the manuscript is transferred.

GUIDELINES:

- For reviewers: https://www.reviewcommons.org/reviewers
- For authors: https://www.reviewcommons.org/authors

CONTACT:

The Review Commons office can be contacted directly at: office@reviewcommons.org

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 Mycobacterium tuberculosis, the enigmatic VanZ proteins, the animal Serum Amyloid A (SAA) and a further host of uncharacterized proteins in a total of 30 families. Though 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 severance reactions. We argue that this superfamily originated as membrane proteins, with one branch (including Wnt and SAA) evolving into soluble 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). These paracrine and autocrine signaling networks are centered on the secreted Wnt proteins that act as 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 on its cytoplasm 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 *Mycobacterium tuberculosis* (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

While 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 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 other core helices of this family contained further absolutely conserved sequence motifs (Figure 2), consistent with the enzymatic function we had earlier proposed for the Min-Wnts (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 expansion 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 model of a "Domain of unknown function", DUF2279 (acc: WP_146898260.1, iteration: 5, e-value: 0.004), the Pfam DUF4056 domain (acc: MBW8016507.1, iteration: 5, e-value: 0.005), and sequences automatically annotated as "YfiM" (acc: WP_019077413.1, iteration: 4, e-value: 0.004). Sequence profile-profile searches with HHpred confirmed these relationships and captured more distant ones. A single sequence query from 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); while a single sequence query 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 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 structural models constructed for representatives of the individual families, along with the rare instances of solved experimental 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. Remarkably, of these, 17 had no pre-existing annotations. Phyletic analysis of individual families revealed a range of distributions, from families with broad conservation in multiple superkingdoms of Life to those restricted to a small number of lineages (see below, Figure S1).

A relationship network was constructed based on p-value and e-value scores using alignments of each family as a query in HHalign profile-profile searches against the rest (see Methods, Figure 1B). The Leiden community detection algorithm (37) was then applied to this network to identify higher-order assemblages. These groupings were also supported by structural synapomorphies, such as a circular permutation and versions with a two-stranded 'handle' (see below, Methods).

The four helices conserved across the superfamily constitute a cone-like structure (Figure 1A), with the helices tending to coalesce on one end and opening out into a pocket on the other, lined by the conserved sequence positions (Figures 1A, S2). The core is also marked by a linker between H1 and H2, which adopts characteristic extended conformations in certain families and higher-order groups. While the linkers joining H2 and H3 and H3 and H4 tend to be more constrained, there are some exceptions; for example, the extended loop insert housing the palmitoleoylated serine residue between H2 and H3 in the metazoan Wnt family (Figure 1A).

Dramatic variability in hydrophobicity of the conserved core across the superfamily

We observed that the Wnt-related families dramatically varied in their hydrophobicity. Using an index for transmembrane propensity (38) (see Methods) and comparing that to known transmembrane (TM) segments, we predict that the α-helices in 18 of the 30 families are hydrophobic enough to qualify as TM domains (Figure 1C). Thus, these are predicted to be integral membrane domains. Further, these 'hydrophobic families' often evince a deeper phyletic distribution pattern than the less-hydrophobic families (Figure S1), implying that the ancestral version of the superfamily was likely a TM domain. Thus, their association with the lipid membrane, combined with the cone-like shape of the conserved core (Figure 1A), leads us to refer to the whole superfamily hereinafter as the *Lipocone* superfamily.

Alphafold3-assisted transmembrane topology prediction (39) revealed that 16 of the 17 integral membrane families are consistently orientated with the aperture of the cone-like structure opening toward the outer face of the membrane. One family did not yield a consistent prediction for

orientation. The predicted TM topology is also generally consistent with the domain fusions when present: e.g., domains that are typically cytoplasmic and those that have extracellular or periplasmic functions are respectively predicted as projecting either inside or outside the membrane (see below).

A unified biochemistry for the Lipocone superfamily

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

Consistent with these observations, two of the families with intact active sites, the PTDSS1/2 (28,43) and TelC (29), which we identified in this work as members of the Lipocone superfamily, have been characterized as active enzymes operating on different lipid substrates (Figure 3A). The eukaryotic PTDSS1/2 localizes to the endoplasmic reticulum (ER) membrane and catalyzes a reaction on the polar head group of phosphatidylcholine or phosphatidylethanolamine (44-46). PTDSS1 and PTDSS2, respectively, exchange the phosphate-linked choline or ethanolamine head groups with L-serine (28)

(Figure 3A). The toxin domain of TelC acts on lipid II (29), the final intermediate in peptidoglycan biosynthesis, which couples an undecaprenyl diphosphate tail to a head group comprised of a N-acetylmuramic acid-N-acetylglucosamine disaccharide, with a pentapeptide further linked to the former sugar (47,48). TelC severs the link between the undecaprenol and the diphosphate coupled to the head group (29) (Figure 3B). The reaction is comparable to that catalyzed by PTDSS1/2, as both attack phosphate linkages in lipid head groups. However, TelC apparently directs a water molecule for the attack in lieu of the hydroxyl group of serine directed by PTDSS1/2 (Figure 3B).

Combining the above observations, we infer the unified biochemistry for the catalytically active families thus: 1) They act on the head groups of lipids either by removing or swapping phosphate-linked head groups (Figure 3A-B). These would be comparable to the phospholipase D (PLD), transphosphatidylation or polyisoprenol phosphoesterase reactions (49). 2) Given the cone-like cavity and the hydrophobicity of the helices, the lipid tail is predicted to be housed within the lipocone with the head group positioned in the active site. 3) In the case of the integral membrane versions, their orientation would predict the targeting of the head groups of the outer leaf of the bilayer.

Major clades of the Lipocone superfamily

The extreme sequence divergence of the superfamily, coupled with the small size of the domain, prevents the use of simple phylogenetic tree analyses to resolve its deep evolutionary history. Hence, we combined community finding algorithms applied on profile-profile similarity networks, comparison of structural features and motifs, and phyletic patterns (Figures 1B, 2, S1) to reconstruct the most parsimonious evolutionary scenario for the diversification of the Lipocone superfamily (Figure 4, see Methods). In the below sections, we survey the higher-order clades, highlighting their specific features.

SAW (SAA-Wnt) clade

This clade consists of four families, with the two prokaryotic families (Min-Wnt and prok-SAA) (24,50), respectively, giving rise to their counterpart eukaryotic families (Met-Wnt and Met-SAA; Figure 4). This

clade is structurally unified by the presence of a fifth helix that stacks in the space between the H2 and H4 helices (Figure 1A, S2). In the Wnt families, this helix is comparable in length to the core helices, while in the SAA families, the helix is shorter, usually containing only a few turns. The clade is further unified by the pronounced conservation of a sNxxGR motif (where 's' is a small residue) encompassing the conserved active site position in H4 (Figure 2). SAW clade Lipocones show low overall hydrophobicity and are known or predicted to be soluble domains. Outside of the clearly inactive eukaryotic Wnt family, the remaining three families largely conserve the core active site residues (Figure 2).

VanZ-Skillet clade

This clade unites seven families: the two VanZ families, VanZ-1 and VanZ-2, prototyped by the bacterial VanZ protein originally identified in the context of vancomycin resistance and the five Skillet families, which form a distinct subclade. These are unified by a "handle"-like structure (hence, "Skillet") adopting a helical conformation in the H1-H2 linker (Figure 1A, S2). Strikingly, a symmetric helical handle is present in the H3-H4 loop of the Skillet-DUF2809 and Skillet-3 families (Figure S2) of this clade. VanZ-1 features a conserved asparagine residue in the H2 position and a DxDDxxxN motif in H4, while VanZ-2 features RKxxH and DxxxD motifs in these respective positions (Figure 2). The Skillet families are largely unifiable in their conservation of an ExxQ motif in H3, an aspartate three positions upstream of the canonical H4 aspartate, and another aspartate in the H2 contributing to the active site. These first two features specifically ally them with the VanZ-1 family (Figure 2).

While the VanZ domain was previously reported as including a fifth TM helix which is C-terminal to the 4-helix Lipocone core defined here (51,52), our survey instead reveals a striking diversity of configurations around the core 4-helix Lipocone domain (Supplemental Material). These range from standalone Lipocone configurations to one or more TM-helices adorning the domain at its N- and/or C-terminus. This variation is consistent with a further tendency for the VanZ families to feature an

extensive diversity of domain fusions to both soluble globular domains and discrete TM modules (see below).

The VanZ families are deep-branching, as suggested by their wide phyletic spread (Figure S1). VanZ-2 is the most widespread individual Lipocone family in bacteria, with several genomes encoding multiple paralogs (Supplemental Material) (51,53). It is also found in certain eukaryotes, including a pan-fungal presence and in some representatives of the SAR clade. Both VanZ-1 and VanZ-2 are particularly well-represented in Gram-positive bacterial lineages like Actinomycetota and Firmicutes, while VanZ-2 is notably nearly universally conserved in the Bacteriodetes/Chlorobi lineage (Figure S1). In contrast, only one of the Skillet families, Skillet-DUF2809, is widely but sporadically distributed, with the four others being more restricted (Figures 4,S1).

YfiM clade

This clade includes three families that are consistently centrally located in the profile-profile similarity network (Figure 1B). This is likely due to their being close in sequence conservation to the ancestral state of the superfamily (Figure 2). Consistent with this, the YfiM-1 family also presents a structurally minimal Lipocone domain, comprised of just the 4-helix configuration with no further elaborations. Notably, this also extends to a lack of domain fusions in this family. In contrast, YfiM-DUF2279 and YfiM-Griddle (DUF3943) are structurally distinguished by an unusual H1-H2 linker, which wraps around the outside and stacks against the H3-H4 linker (Figure S2). The YfiM-Griddle family further features a unique 'flattened' surface around the aperture of the Lipocone formed by protruding loops (hence, "Griddle"; Figure S2). This leaves the active site pocket more accessible relative to families with more elaborately structured inter-helix linkers. The Griddle family also features a C-terminal extension with a two-helix hairpin (with a hhsP motif in the turn between the two helices, where 'h' is a hydrophobic residue and 's' is a small residue) (Figure S2). The three YfiM families straddle the membrane-propensity boundary in the plot (Figure 1C). Further, the YfiM-DUF2279 and Griddle families are strikingly absent in Gram-positive bacterial lineages (Figure S1). Concurrent with these features, they

are often predicted as outer-membrane proteins, suggesting a role in this subcellular location (see below).

ClaspCone-CapCone-TelC clade

Members of this clade are unified by an elaborated H1-H2 linker that often contains one or more helical segments that are typically predicted to guard the aperture of the Lipocone domain (Figure S2). This linker ends in a "clasp"-like element, which forms a range of structures in different families of the clade before leading into H2 (Figure S2). The clade is also unified by a striking reduction of overall hydrophobicity, predicting that the members of this clade are soluble domains (Figure 1C). Outside of the divergent TelC subclade, most of the families in this clade conserve a serine residue three positions upstream of the active site aspartate in H4, often preceded by an aromatic residue which is typically phenylalanine. H4 also usually features a conserved asparagine four positions downstream of the conserved aspartate active site position, immediately preceded by a small residue (Figure 2). The second H3 active site position is generally poorly conserved, though when present, it is usually an aspartate residue. Finally, H2 contains either a DK or xD motif four positions upstream of the canonical H2 active site histidine residue (Figure 2).

The most rudimentary clasps are found in the ClaspCone-1, -2, and -3 families, where it is little more than a rounded loop, though, in ClaspCone-1, a small β-hairpin emerges within it. The three ClaspCone families are further unified by the presence of a two-helix insert leading into H2 that stacks against the Lipocone core (Figure S2). The three CapCone families, CapCone-DUF4056, CapCone-1, and CapCone-2, are named so for an encasing structure over the active site resembling a cap (Figure S2). They share a conserved glycine residue six positions upstream of the active site H2 histidine and a striking S/GxxSxx motif upstream of the conserved H4 aspartate (Figure 2). They are further unified by a pronounced β-hairpin clasp augmented by an additional strand (Figure S2). They also display varying degrees of degeneration of H1, along with family-specific structural elaborations.

The TelC group of this clade, prototyped by the mycobacterial TelC toxin (29), is divided into two families featuring prokaryotic (prok-TelC) (29) and metazoan versions (Met-TelC) (54). Both TelC families feature a "cap" with contributions from inserts in the H1-H2 and H3-H4 loops (Figure S2). Unique to these families is the conservation of an aspartate residue located six positions downstream of the canonical active site aspartate of H4 (Figure 2). This aspartate points away from the center of the Lipocone and interacts with a conserved arginine from a synapomorphic C-terminal helical extension.

cpCone clade

A widespread yet sporadically distributed clade of seven families emerging as a stable community in the profile-profile similarity network (Figure 1B) is defined by a unique structural synapomorphy: a circular permutation (55) (hence, cpCone) placing the normally N-terminal H1 at the C-terminus of H4 (Figure S2-3). This clade is also united by unique sequence features, viz., a polar residue (typically aspartate) six positions upstream of the conserved H2 histidine and a second glutamate three positions downstream of the conserved H3 glutamate (Figure 2). While the circular permutation is shared across the clade, several structural variations are seen, often within the same family (Figure S3). These include: 1) versions containing a duplication of the Lipocone domain. While the second copy in these versions is catalytically inactive, the H1' from the second duplicate displaces the H1 from the first copy, suggestive of an intermediate to the circular permutation. 2) Versions retaining a candidate H1 that has been displaced by H1' in a five-helix arrangement. 3) Those containing just the circularly permuted core. 4) Versions showing a degradation of the H1 helix, preserving just a 3-helix core (Figure S3). Despite this propensity for structural variation, the active site residues are strongly conserved, with the exception of the cpCONE-i family, which we infer to be catalytically inactive (Figure 2). The core helices of the cpCone clade are strongly hydrophobic, and they are predicted to all be integral membrane domains (Figure 1C). Consistent with this, the eukaryotic PTSSD1/2 domains reside in the ER membrane (44,45).

Wok family

The Wok family (partly covered by the Pfam DUF2238 model) shows a higher order grouping with the above circularly permuted clade (Figures 1B,4) but has a phyletic distribution only rivaled by the VanZ-2 family (Figure S1), suggesting a deep-branching origin. This family has the general shape of a wok formed by two distinguishing structural synapomorphies: a 2-TM helix N-terminal extension and a unique "handle" formed by the linker between the H3 and H4 (Figure S2). It additionally features a C-terminal, rapidly diversifying cytoplasmic tail. Nevertheless, it retains the inferred ancestral active site configuration (Figure 2). The strongly hydrophobic core helices of the Wok family predict it to be an integral membrane enzyme (Figure 1C).

Functional themes in the Lipocone superfamily

Given our inference of shared general biochemistry across the Lipocone superfamily in targeting phosphate-containing linkages in head groups of both classic phospholipids and polyisoprenoid lipids, we next used contextual information from conserved gene-neighborhoods, domain architectures and phyletic pattern vectors, a powerful means of deciphering gene function (56), to narrow down the predictions for specific families (Figure 5, Table S1, Supplementary Data). To this end, we constructed a graph (network) wherein the nodes are individual domains and edges indicate adjacency in domain architectures or conserved gene-neighborhoods (Figure 6, see Methods). We then identified cliques in these networks and merged the individual cliques containing a particular Lipocone domain to define its dense subgraph (Figures S4-S6). We then analyzed these subgraphs to identify statistically significant functional categories (Table S2; see Methods). This data was combined with existing experimental results and the sequence and structure analysis outlined above to arrive at the functional themes surveyed in the below sections.

Lipocone domains in membrane lipid, peptidoglycan and exopolysaccharide modifications

Across different Lipocone families, we found statistically significant connections to roles in modifying lipid head groups in various membranes and in lipids involved in the synthesis of extracellular matrix polymers such as peptidoglycan and lipopolysaccharides (Figure 6, Table S2, Supplementary Data).

Archetypal lipid head group exchange reactions catalyzed by the cpCone clade

One of the few experimentally characterized Lipocone families is the eukaryotic PTDSS1/2 family of the cpCone clade, members of which exchange the head group of essential membrane phospholipids to generate phosphatidylserine from phosphatidylethanolamine or phosphatidylcholine (Figure 3A) (28,57). Given the pervasive presence of this clade in archaea (Figure S1), it is thus tempting to speculate that these archaeal cpCones may play a role in the modification of Archaea-specific lipids (58-60) through a comparable head group exchange reaction (see below).

In bacteria, the related cpCone-1 family shows operonic association with a LolA-like lipoprotein which shuttles lipoproteins to the outer membrane (61) and a novel 4TM protein (Figure 5A). This raises the possibility that cpCone-1 might mediate the formation of membrane domains featuring lipids with a modified head group that act as foci for the trafficking of lipoproteins. Curiously, the cpCone-1 gene might also be inserted between the bacterial chromosome segregation and condensation complex subunits the Kleisin ScpA and the wHTH ScpB (62-65). The bacterial cpCONE-DUF2585 is operonically coupled to a GNAT family NH_2-group-acetyltransferase and further linked to genes for the glycolate oxidase GlcE and GlcF (66) and the bacterial proteasome subunits HslV and HslU (67) (Figure 5A, Supplementary Data). These might point to the coupling of membrane lipid head group modifications with disparate processes, such as chromosome segregation during cell division or different responses to stress (68-70).

The Wok and YfiM-1 families in cardiolipin and modified isoprenoid lipid pathways

We observed a set of conserved gene neighborhoods displaying the mutually exclusive presence of a synaptojanin-like phosphatase gene with one either encoding a member of the Wok family or a cardiolipin synthase of the HKD superfamily (71) (Figure 5B, Supplementary Data). This suggested that the latter two are analogous enzymes catalyzing equivalent reactions. The cardiolipin synthase utilizes two phosphatidyl glycerol molecules as substrates to generate cardiolipin with the release of one of the

glycerol head groups (72). Effectively, this is comparable to the head group exchange reaction catalyzed by the cpCone clade (Figure 3A). Hence, we propose that these members of the Wok clade are cardiolipin synthases (Figure 3C). Distinct phosphoesterases from the synaptojanin-like enzyme, like calcineurin (73) and HAD (74), are also observed in association with the Wok, suggesting that they might regulate lipid composition by acting on the phospholipids or their precursors (Figure 5C). In a distinct neighborhood, the Wok clade enzyme is coupled to carotenoid biosynthesis genes (75,76). (Figures 5D,S4). This raises the possibility that these members might also catalyze a comparable reaction to the above on isoprenoid lipids: they could synthesize a carotenoid from two geranylgeranyl-diphosphate molecules (77,78). In both these contexts, the actinobacterial operons often include genes for GT-A family glycosyltransferases, suggesting the further synthesis of glycosylated derivatives of the lipids or carotenoids (79) (Supplementary Data). In several bacteria, a YfiM-1 family Lipocone is operonically coupled to a UbiA-like prenyltransferase (80). This gene neighborhood additionally codes for a slew of enzymes, such as an amidophosphoribosyltransferase (81), a RidA-like deaminase (82), and a pair of structurally distinct phosphoesterases, respectively, containing an HD and a PHP domain (73,83) (Figures 5E,S4 Supplementary Data). This suggests a role for the Lipocone and the associated enzymes in generating a modified polyisoprenoid metabolite.

VanZ families in lipid head group modifications in peptidoglycan and exopolysaccharide metabolism

The widespread VanZ-1 and VanZ-2 families frequently show either gene neighborhood associations or direct domain fusions, with diverse genes involved in both peptidoglycan and other extracellular polysaccharide pathways. Chief among these are the lipid carrier flippase (Pfam: MviN_MATE clan) (84-86), the UDP-GlcNAc/MurNAc lipid transferases, which generate the lipid-linked exopolysaccharide precursors (lipid I) (48,87), and UDP-N-acetylglucosamine (UDP-GlcNAc) biosynthesis enzymes (88,89). Despite multiple examples of crossover in functional themes, the gene-neighborhood contexts of VanZ-1 and VanZ-2 suggest a metabolic partitioning, with VanZ-2 significantly associating specifically with peptidoglycan-related genes and VanZ-1 significantly linking with biosynthesis genes for other exopolysaccharides (e.g., the outer-membrane-associated lipopolysaccharide) (90) (Figures 5F,6,S5,

Supplementary Data). The latter include WaaL-like lipid A transferase (91), the polysaccharide chain-length determination domain Wzz (92), the Wzc kinase and the “extracellular antigen”-regulating ElyC-like domain (Pfam: DUF218) (93), and numerous nucleotide-diphosphate sugar biosynthesis and modification enzymes (94) (Figures 5F,6,S5).

The precursors of both peptidoglycan and exopolysaccharides are synthesized in the cytosol, linked to lipid carriers via a diphosphate linkage, e.g., the polyisoprenoid lipid undecaprenol (bactoprenol) (90,94-97). A key step in their maturation is the flipping by the flippase of the lipid-linked intermediates associated with the inner membrane to the outer membrane. These flipped units are then incorporated into the maturing chain (98,99) by the peptidoglycan glycosyltransferase (GTase) (100) and the chain length determination protein, WzzE/polymerase, WzyE (92,101), in peptidoglycan and other exopolysaccharide maturation pathways, respectively. Based on the precedence of the PTDSS1/2 reaction (Figure 3A), we propose that VanZ-1 and VanZ-2 analogously act on the lipid carrier-linked sugar head groups to catalyze either the extension of the polysaccharide chains through transesterification or the terminal release of the mature chain through phosphoester cleavage (Figure 3E).

The above activities could be linked to distinct types of peptidoglycan synthesis during different phases of the life cycle, e.g., sporulation versus vegetative growth in Bacillota. Consistent with this, VanZ-1 is pervasive in sporulating Gram-positive bacteria (Figure S1). Terminal transfer from the lipid carrier of the Gram-negative bacterial O-antigen (as well as other exopolysaccharides attached to the lipid A carrier) has been attributed to the WaaL-like enzymes (91,102). However, bacteria generate further lineage-specific polysaccharide decorations, capsule structures, and other exopolysaccharides (e.g., xanthan, enterobacterial common antigen (ECA), alginate, colonic acid), as well as teichoic acids (e.g., wall teichoic acids, WTA) (103). Notably, the analogs of WaaL, i.e., the terminal transferases for several exopolysaccharides, including ECA and WTA, have to date escaped identification (93), making the VanZ-2 enzymes attractive candidates. Such a function would be consistent with its widespread yet

sporadic phyletic pattern (Figure S1), keeping with the rapid evolution of these exopolysaccharides due to strong selection for variation (104) (Figure 6).

Atypical VanZ domains in uncharacterized modifications of peptidoglycan and the outer membrane

Certain representatives of the two VanZ families also show operonic associations indicative of distinct outer membrane-associated or peptidoglycan modification functions distinct from those described above (Figures 5G,6 Supplementary Data): 1) An operon in FCB group bacteria couples a VanZ-2 gene with those coding for a SprA secretin-like channel protein (105), a glycine cleavage H (GCVH)-like lipoyl-group carrier protein (106), a 2TM protein fused via a proline-rich linker to a C-terminal TonB-C domain (107), and a secreted, second TonB-C domain fused to a Wzi-like outer membrane protein (OMP) superfamily β-barrel (108) (Figures 5G, 6). 2) In betaproteobacteria, certain VanZ-1 domains are duplicated with the C-terminal copy being inactive (VanZ-i) and found in an unusual four-gene operon with a thioredoxin-fold [2Fe-2S] ferredoxin (109), a possible lipase of the α/β-hydrolase superfamily (110), and a metallo-β-lactamase (MβL) fold D-Ala-D-Ala cross-linking transpeptidase (111,112). 3) A patescibacterial operon encodes a VanZ-2 domain with an ABC ATPase transporter system, either of two structurally distinct peptidases, namely a Papain-like or glycine-glycine peptidase (113,114), fused to the same membrane-anchored N-terminal coiled-coil region, and a further TM protein containing one or more external Lamin-Tail domains (LTDs) predicted to bind extracellular DNA or polysaccharides (115) (Figures 5G,6,S5). The associations in the first of the above neighborhoods point to a distinct outer membrane-associated lipid modification, while the other two might be involved in lineage-specific decorations/modifications of peptidoglycan associated with peptide-crosslinking or cleavage activities.

Lipocone domains operating in the outer membrane

Contextual associations, phyletic patterns, and localization predictions support the action of two Lipocone families directly in the outer membrane. Notably, the Griddle and YfiM-DUF2279 families are found nearly obligately directly fused or operonically linked to several distinct OMP β-barrels (116,117)

(Figures 5H,6,S4). Up to three Griddle Lipocones, usually with a cognate OMP β-barrel, might be encoded next to each other in the genome. Additionally, Griddle family genes are often encoded in operons with several components of the outer membrane lipid and protein trafficking apparatus, including the LolA-like chaperone (118), the POTRA domain (119,120), the channel-blocking Plug domains (121), and the TolA-binding TolB-N domain (122). Further, these operons might encode a Patatin-like lipase (123), GT-B family glycosyltransferases (79), and a range of phosphoesterases (e.g., an integral membrane phosphatidic acid phosphatase PAP2 (124), a lipobox-containing synaptojanin superfamily phosphoesterase (125) and a secreted R-P phosphatase (126) (see Figure 6, Supplementary Data)). In addition to the fusion to the OMP β-barrel, the YfiM-DUF2279 family (Figure 5H) shows operonic associations with a secreted MltG-like peptidoglycan lytic transglycosylase (127,128), a lipid-anchored cytochrome c heme-binding domain (129), a phosphoglucomutase/phosphomannomutase enzyme (130), a GNAT acyltransferase (131), a diaminopimelate (DAP) epimerase (132), and a lysozyme-like enzyme (133). In a distinct operon, YfiM-DUF2279 is combined with a GT-A glycosyltransferase domain (79), a further OMP β-barrel, and a secreted PDZ-like domain fused to a ClpP-like serine protease (134,135) (Figure 5H).

The strong linkage to the OMP β-barrel, together with their localization prediction, suggests that these Griddle and YfiM-DUF2279 Lipocone domains operate in the outer membrane. As with the inner membrane lipids, they could potentially catalyze modifications of head groups through transesterification and/or linkage/release of outer membrane-associated polysaccharide chains through action on lipid-head group phosphoesters.

Lipocone domains acting on lipids in transit to the outer membrane

The ClaspCone-1 and ClaspCone-3 families lack the hydrophobicity indicative of direct residence in the membrane (Figure 1C); instead, they are predicted to localize to the periplasmic space. In the ClaspCone-1 family, the Lipocone domain is fused at the extreme N-terminus to either a single TM or a 5TM domain predicted to anchor it to the cell membrane. Between this TM element and the Lipocone

domain, we detected a previously uncharacterized version of the Tubular lipid binding protein (TULIP) domain (136,137) or an Ig-like and a Zincin-like metallopeptidase (MPTase) domain (138) (Figures 5I, 6). These ClaspCone-1 genes may also show operonic associations with genes encoding a lipase of the SGNH family (139) and a membrane-bound O-acyltransferase (MBOAT; Figure 5I, Supplementary Data) (140). The TULIP domain superfamily has recently been characterized as a lipid-binding domain (136,137), which in proteobacteria functions in outer membrane lipid transport (141,142). Thus, we propose that the ClaspCone-1 family is likely to act in the periplasmic space on the head groups of outer-membrane targeted lipids bound to the TULIP or Ig-like domains.

A Lipocone domain catalyzing a predicted lipoprotein lipid linkage reaction

The Skillet-1 Lipocone is strongly coupled in an operon with a downstream gene coding for a protein with an unusual lipobox-like sequence followed by one of several extracellular domains (e.g., concanavalin, β-jelly roll, OB-fold, Ig-like, β-propeller) predicted to bind carbohydrates or other ligands (143-147) (Figures 5J,6,S5 Supplementary Data). The lipobox-like sequence features a conserved GS motif at its C-terminus instead of the usual GC of the classic lipobox of bacterial lipoproteins (148) (Figure S7). In the canonical lipoprotein processing pathway, a thioether linkage is formed between the sulfhydryl of the cysteine and a diacylglycerol lipid embedded in the inner membrane by the lipoprotein diacylglyceryl transferase (lgt) enzyme, followed by the cleavage of the signal peptide at the GC motif junction by the signal peptidase (149,150). Given the serine in place of the cysteine in these lipobox-like sequences, we propose that it undergoes non-canonical lipidation by the associated Skillet-1 Lipocone protein in lieu of lgt. We propose that comparable to PTDSS1/2, which act on free serine, the Skillet-1 family links the conserved serine from the lipobox-like sequence to a phospholipid (Figure 3A,D).

Lipocone domains in predicted lipid-associated signaling systems

Systems defined by standalone proteins with Lipocone domains

Several representatives of the two VanZ and Skillet-3 families are fused to a diverse array of known or predicted extracellular ligand-binding domains (Figure 5K), where the architecture takes the form of SP+X+TM+Lipocone or Lipocone+TM+X, where 'X' is the extracellular ligand-binding domain and SP is a signal peptide. The ligand binding domains include: (i) carbohydrate-binding lectin domains such as jelly-roll, concanavalin-like, NPCBM-like, CBD9-like, and other β-sandwiches (143,144,151-153)); (ii) a lipid-binding helix-grip superfamily domain (154)); (iii) those binding other potential ligands (e.g., Ig, OB-fold, YycI-like, DUF498-like, PepSY-like, β-helix, TPR, MORN, and β-propeller repeats (145-147,153,155-158)) (Figure 5K, Supplementary Data). We interpret these architectures as implying signaling, wherein the binding of the cognate ligand by one of the above domains regulates the catalytic activity of the associated Lipocone domain. Among these, the extracellular domains fused to the Skillet-3 family are particularly notable for their extreme variability (Figure 5K). This suggests their diversification under an arms race scenario (also see below) in a biological conflict. Further, the genes coding for the above are sporadically associated with exopolysaccharide metabolism genes (Supplementary Data). Hence, it is conceivable that this signaling is associated with exopolysaccharide variation (e.g., O-antigen phase-variation (159,160)), which might play a role in evading bacteriophage attachment.

Additionally, VanZ-1 Lipocone domains are also fused to several known signaling domains confidently predicted to reside in the cytoplasm, including the cyclic nucleotide-binding domain (cNMPBD), phosphopeptide-binding FHA, and DNA-binding RHH and HTH domains (65,161-163) (Figure 5K). These associations suggest potential VanZ regulation via a cytoplasmic cyclic nucleotide (sensed by cNMPBD) or, conversely, VanZ acting as an allosteric regulator of a transcriptional program via the HTH or RHH domain. One of the most common yet enigmatic fusions to VanZ is with the integral membrane RDD domain (53). The role of this domain is unknown; however, our analysis indicates a potential active site oriented towards the cytoplasmic face of the membrane (Nicastro GN, Burroughs AM, Aravind L, unpublished observations). The VanZ-RDD fusion is sometimes further fused to other domains (Figure 5K), the most notable being a highly derived but active novel Histidine Kinase domain (Figure 5K).

These associations again point to the coupling of lipid modification with a signaling event on the cytoplasmic face of the membrane, which might relate to the dynamic regulation of lipid-carrier-bound exopolysaccharide precursors.

Multi-component associations of the Lipocone proteins in signaling

These systems resemble the above-discussed versions but are encoded by conserved gene neighborhoods that separate the Lipocone and the signaling elements (typically predicted transcription regulators) into distinct components. Our analysis recovered at least three such systems: 1) A VanZ-1 Lipocone in the recently described HAAS/PadR-HTH two-component systems, which sometimes replace classical Histidine kinase-Receiver two-component systems (164). In these systems, the detection of an extracellular or intramembrane stimulus by a sensor domain releases the PadR-HTH transcription regulator bound to the sensor-fused HAAS domain. Here, VanZ-1 occupies the sensor position (Figure 5L). 2) A Skillet-2 Lipocone is coupled in a core two-gene system to a conserved upstream gene (Figure 5L,S4). That gene encodes a single TM protein with either a zinc ribbon (ZnR) fused to a conserved helix or an HTH domain fused to a ClpS-ribosomal L7/L12 domain in its cytoplasmic region (165). These neighborhoods might also code for an HMG-CoA reductase and GHMP kinase that catalyze successive reactions in the production of phosphomevalonate, a precursor of isoprenoid lipids (166,167). 3) A Skillet-DUF2809 Lipocone protein is operonically coupled with a 6TM protein and a further predicted transcription factor with a wHTH protein. These operons are further elaborated via additional embedded and flanking genes either coding for components of isoprenoid lipid (e.g., undecaprenol) (168,169) or exopolysaccharide (e.g., ECA and related polysaccharides) metabolism (94,97) (Figures 5L,6,S5 Supplementary Data).

The Lipocone domains in these systems are predicted to be active enzymes, which, together with their operonic associations, points to functions involving the modification or transesterification of isoprenoid lipid head groups, sometimes in the context of exopolysaccharide biosynthesis. However, their associations with the intracellular HTH domains suggest that the Lipocone enzymatic activity is

potentially coupled with the transcriptional regulation of the production of precursors of the lipids or exopolysaccharides. Given the high variability in the associated genes related to exopolysaccharide/lipopolysaccharide biosynthesis, we anticipate that the associated transcriptional regulation potentially responds to environmental stress, phages, predatory bacteria attacks, or host immune response.

Lipocone domains as effectors in biological conflicts

Lipocone domains in antiviral immunity

The Min-Wnt domains that we originally identified were predicted to play a role in biological conflicts with invasive selfish elements such as viruses (24). In this work, we better explain their potential mechanism of action. These versions show no fusions to extracellular domains or secretory signals, suggesting that they act as soluble versions within cells – a proposal also backed by their overall hydrophobicity (Figure 1C). These Min-Wnts are typically fused to the DUF3892, which displays a fold characterized by a three-stranded meander followed by a helix also seen in the dsRNA-binding domain and the ribosome hibernation factors (HPF) (170,171) (Figure 5M). Hence, we propose that these versions might potentially act to sense virally induced RNAs or modified ribosomes (24) to trigger a dormancy or suicide response to limit viral infection via the Min-Wnt effector. Specifically, the Min-Wnt might attack peptidoglycan precursors, such as lipid II, prior to their 'flipping' to restrict cell wall synthesis (85,172,173) or other carrier lipids.

One other Min-Wnt domain N-terminally fused to a three-stranded β-meander is pervasive in the Bacteroidetes clade. This is operonically coupled with genes encoding a TM-linked run of pentapeptide repeats and two structurally distinct, secreted glycosyl hydrolase enzymes, respectively, containing a TIM barrel domain and a run of β-helix repeats (Figure 5M, Supplementary Data). Further, cyanobacteria show a standalone prok-TelC domain without any secretory signals. These could again act as cell wall or membrane-lipid targeting effectors in response to intracellular invaders or stress (Figure 5M). Interestingly, some tailed bacteriophages also code for intracellular Min-Wnt domains,

suggesting that they might also be deployed on the virus side in biological conflicts (Supplementary Data).

Lipocones as toxin domains in polymorphic and allied conflict systems

Polymorphic toxins and related systems, widespread across bacteria and certain archaea, are characterized by a highly variable C-terminal toxin domain ("toxin tip") that is preceded by a range of more conserved domains typically required for autoproteolytic processing of the toxin, its packaging and trafficking (e.g., RHS repeats), adhesion and secretion via one of several secretory systems (174,175). The toxin might be delivered via one of the secretory systems into a target cell or else via direct contact between interacting cells. Classical polymorphic toxins are usually involved in kin discrimination and are accompanied by genomically linked cognate immunity proteins that protect against self-intoxication (174,176). Keeping with the principle of effector sharing between systems involved in distinct types of biological conflicts, we had originally identified the Min-Wnt domain closely related to those described in the above subsection as a toxin tip in polymorphic toxin systems (24). In the current work, we extend these findings to show that several distinct Lipocone families have been independently recruited as toxin tips of polymorphic toxins and related systems, namely Min-Wnt, pro-SAA, prok-TelC, CapCone-1, CapCone-2, ClaspCone-2 and VanZ-1 (Figures 4,5N,6,S6).

Certain CapCone-2 and Min-Wnt toxins define some of the simplest of these toxin systems that are common in Gram-positive bacteria. Here, a standalone Lipocone domain is coupled to a signal peptide or lipobox via a poorly structured linker. These are usually encoded in a two-gene configuration with their cognate immunity protein (Figure 5O). More complex versions present, in addition to adhesion, peptidoglycan-binding, lipid-binding and proteolytic processing domains, multiple hallmarks of specific secretion systems. These include T4SS (VirD4-binding domain), T6SS (PAAR domain), T7SS (WXG/LXG domain), T9SS, and MuF domains indicating that these toxins are delivered via diverse secretion systems (174,176) (Figure 5P, Supplementary Data). Additionally, we recovered standalone CapCone-1 domains encoded in an operon with a PsbP/MOG1-like domain diagnostic of the T6SS

(174,177) (Figure 5P), which predicts that these are likely toxins secreted via this system. Further, we also found Min-Wnt domains fused to the N-termini of RTX-like β-roll repeats, suggestive of a T1SS-mediated toxin export (178) (Figure 5P).

Our analysis also uncovered multiple, previously uncharacterized trafficking/packaging systems associated with different Lipocone polymorphic toxins. Several Min-Wnt and CapCone-1 domains with lipoboxes are fused to an N-terminal Cystatin-like superfamily domain (179) (Figure 5Q). The same domain is also comparably fused to several other C-terminal toxin domains in related organisms, some of which are also predicted to target lipid head groups: (i) a novel toxin domain we unified with the lipid-targeting Colicin M fold (180); (ii) a lipid-binding START-domain-like helix-grip fold domain (154); (iii) a papain-like fold fatty acyltransferase (181); (iv) a domain related to the VanY-like D-Ala-D-Ala carboxypeptidase (182) (Figure 5Q, Supplementary Data). In all these cases, the toxins are coupled to a related immunity protein (see below), suggesting that they define a distinct polymorphic toxin system. We propose that this Cystatin-like domain defines a novel packaging or deployment system upon secretion for the C-terminal toxin domain analogous to Cystatin domains in functioning with eukaryotic proteases (183). The prok-TelC family Lipocones are found in distinctive architectures in two poorly characterized, predicted polymorphic toxin systems. In one of them, they are fused to N-terminal glucan-binding GbpC β-sandwich domain (184) and repeats of MucBP-like Ig domains (185), which might anchor them to exopolysaccharides (Figures 5Q,S6 Supplementary Data). The second variant found in association with T9SS components (186) shows fusions to one or more copies of a previously undetected TPM domain (Figure 5Q). While the domain has been claimed to be a phosphatase (187), our recent analysis indicates that this is unlikely to be the case (164). Instead, we propose that the TPM domain might assist in the assembly of membrane-linked protein complexes, a role that might be relevant to the trafficking of these toxins (164).

To date, the only experimentally characterized Lipocone domain from polymorphic toxins is of the prok-TelC family, which are secreted via T7SS (29,188) (Figures 5P,6). Notably, prok-TelC has been

shown to be active only outside the cell and not in the cytoplasm (29). As noted above, it attacks lipid II to cleave off the peptide-linked disaccharide pyrophosphate head group from the undecaprenol tail (Figure 3B). Prok-TelC has also been speculated to similarly attack WTA-lipid II linkages (29). These findings provide a template for other Lipocone superfamily effectors in potentially targeting lipid carrier linkages in peptidoglycan and exopolysaccharide intermediates. However, given the diversity within the family, it is conceivable that they also target other lipids (Figure 3F).

Immunity proteins of Lipocone polymorphic toxins indicate periplasmic/intramembrane action

To date, only a single immunity protein has been reported for Lipocone toxins, viz., TipC, which counters the prok-TelC toxin in the periplasm (29,189) (Figure 5P,6). Here, we uncovered a range of immunity proteins belonging to structurally distinct folds that counter the remaining Lipocone toxins (Figures 5N-Q,S8, Supplementary Data). The most widespread of these is a rapidly evolving, membrane-anchored member of the BamE-like superfamily that associates with not only Min-Wnt and CapCone toxins but also other above-mentioned lipid-head-group targeting toxins (e.g., the novel Colicin M-like domain). The BamE-like fold features a core two-helix hairpin followed by a run of three β-strands (Figure S8). The classical BamE operates in a pathway for the assembly of OMP β-barrels (190,191), suggesting that these immunity proteins were recruited from an ancestral BamE and, like it, function in the periplasm. Additional candidate immunity proteins with more restricted phyletic spreads include (Figure S8): (i) a β-jelly-roll fold-containing protein (144); (ii) an integral membrane protein with a 4-TM core. These two are observed with Min-Wnt toxins. (iii) A novel domain combining an α-helix with a run of 4 β-strands stabilized by four absolutely conserved cysteine residues. This is coupled to both Min-Wnt and prok-SAA toxins; (iv) a protein with an OB-fold domain (145) (v) a protein with a β-sandwich related to the eukaryotic centriolar assembly SAS-6 N-terminal domain (192). The last two are coupled to CapCone-2 toxins (Supplementary Data). Notably, despite their structural diversity, these immunity proteins are all TM or lipoproteins and, like TipC (29,189), are predicted to operate at the membrane or in the periplasm (Figure 5N-Q, Supplementary Data). This suggests that they intercept their cognate Lipocone toxin domain outside cells or in the membrane rather than within the cell.

Lipocone toxins in predator-prey and other inter-specific conflicts

In contrast to polymorphic toxins, which are typically deployed in intra-specific conflict between competing strains of the same species, other toxin systems are deployed against more distantly related target cells, such as prey and eukaryotic hosts (193). While some of these closely parallel polymorphic toxins in their domain architecture, they are usually distinguished by the lack of an accompanying immunity protein. The simplest of these systems are secreted Min-Wnt proteins from bacteria and fungi. These present just a standalone Min-Wnt domain or one fused to a novel domain with a half β-barrel wrapping around a helix (Figure 5R, Supplementary Data). These are probably deployed as diffusible toxins that target rival organisms in the environment.

Another architectural theme is defined by Min-Wnt and prok-SAA Lipocones fused to an enigmatic, novel, short C-terminal domain, which is comprised of a long β-hairpin with a characteristic "break" in its central region, causing it to acquire an arch-like appearance (Figures 5S, S9). Hence, we refer to this domain as the broken-hairpin. We found the broken-hairpin domain to be fused to a wide array of predicted toxin domains across the bacterial superkingdom. These include effector domains found in polymorphic toxin and allied systems that target peptidoglycan, carrier lipids and the membrane, such as members of the Colicin M (180), Zeta toxin-kinase (194), lysozyme (195), an α/β-hydrolase superfamilies (110) and nuclease toxins such as members of the HNH, HipA, SNase, and BECR superfamilies (174) (Figures 5S, S9C). Remarkably, these proteins with the broken-hairpin tend to lack a signal peptide or association with any other secretion system or immunity proteins (Figure 5S, Supplemental Material). Hence, we propose that the broken-hairpin domain itself serves as a trafficking mechanism for the externalization of these toxins in conflicts with rival environmental organisms.

Some predicted secreted Lipocones are found predominantly in predatory bacteria. The first of these are CapCone-2 domains from lineages like Bdellovibrionota, which are encoded in two-gene systems, with the second gene coding for a further secreted effector such as an α/β-hydrolase, Patatin, or

acyltransferase or an OMP β-barrel domain (110,116,117,123) (Figures 5T,S6). Secreted prok-SAA domains from myxobacteria and some other lineages are fused to a Zincin-like metallopeptidase domain at the N-terminal, and the first bacterial example of the von Willebrand Factor D (vWD) and Ig domains at the C-terminus (196) (Figure 5T, Supplemental Material). In the recently described predatory Patescibacterial branch of Omnitrophota species, Skillet-clade Lipocone domains are found in gigantic proteins combined with several other domains and TM segments. Domains found in these proteins include polysaccharide biosynthesis enzymes (94,97), signaling proteins involved in Histidine kinase-Receiver relays (197), peptidases of the MPTase and Papain-like superfamily (113,138), diverse methylases, and extracellular ligand-binding domains like the peptidoglycan-binding LysM domain (198) (Figures 5T,6). Given the concentration of the above systems in predatory bacteria (Supplementary Data), we posit that the above Lipocones might function as toxins targeting prey membranes alongside a battery of effectors targeting other cell surface components. In particular, the CapCone-2 systems are likely to target outer membranes. In animals, the vWD domain is involved in adhesion (199) and might play a similar role in adhering to prey cells in the bacterial Prok-SAA toxins, while the MPTase in these proteins potentially releases the toxin through autoproteolysis. Finally, the giant proteins from the Patescibacteria are likely to combine signaling of prey presence with the overcoming of prey defenses and targeting of membranes.

Certain prok-TelC proteins are observed as part of several distinctive systems that could be involved in as-yet-undiscovered predatory interactions or in targeting environmental competitors. One such, defined by large proteins from spore-forming Bacillota, combines a diversifying set of extracellular ligand-binding domains (e.g., Ig-like, Cell-wall-binding β-hairpins and β-propellers (146,147,200)) with a two-enzyme core formed by a prok-TelC and a N-acetylglucosamine (GlcNAc)-1-phosphodiester alpha-N-acetylglucosaminidase (NAGPA). NAGPA catalyzes phosphoric-diester hydrolysis to release phosphodiester-linked sugars (Figures 5U,6,S6) (201). Some of these proteins feature an additional NlpC/p60 superfamily peptidase domain predicted to target peptidoglycan (181). The recombinational diversity of ligand-binding domains in this system, even among closely related Bacillota species,

supports a possible arms race and involvement in a biological conflict. Other TelC domains in some Bacillota, Actinomycetota, and fungi are fused to peptidoglycan-binding domains (PGBD) (202) and an Rv2525c-like TIM-barrel (203) (Figures 5U,S6). In Actinomycetota, this protein is further combined in operons with either of two mutually exclusive genes coding for rapidly evolving proteins (Figure 5U): (i) a secreted protein containing a pair of Ig domains (200); (ii) a 3-TM protein (3TM-CCDN) with two conserved cysteines, an aspartate and asparagine residues predicted to be located between the TM segments outside the cell. This version is further coupled to a gene for a secreted VanY superfamily peptidase (182) (Figure 5U, Supplementary Data). Common to these contexts are rapidly evolving and variable domains on the one hand and peptidoglycan/exopolysaccharide binding or degrading domains on the other. Hence, we interpret these as potential conflict systems that engage the cell wall and target it and associated membranes in rival bacteria.

Lipocone domains in resistance to antimicrobial agents

VanZ-1 proteins were initially identified as encoded by a gene linked to that coding for the VanY D-alanyl-D-alanine carboxypeptidase involved in resistance to glycopeptide antibiotics like vancomycin and teicoplanin (31,204-206) (Figure 6). These antibiotics bind the terminal D-Ala-D-Ala in the peptide moiety of peptidoglycan, preventing the transpeptidase cross-linking reaction necessary for its maturation. Upon detection of these antibiotics, enzymes encoded by the core vancomycin resistance operon re-engineer the exported peptidoglycan by inserting a D-Ala-D-Lac in place of the D-Ala-D-Ala linkage, precluding antibiotic binding (53). The VanY peptidase, while not strictly required for antibiotic resistance, acts as an accessory to this system by cleaving any remaining D-Ala-D-Ala linkages generated via the canonical pathway (53,205). However, the role of VanZ in this system has so far remained unknown. While only a small fraction of the VanZ-1 genes are found in these antibiotic resistance contexts (Supplementary Data), interestingly, other Lipocone genes, namely those of the VanZ-2 and the Skillet families, might also be linked to VanY in lieu of VanZ-1. Further, VanY might be replaced by a structurally unrelated secreted D-Ala-D-Ala carboxypeptidase of the metallo-beta-lactamase fold (111) in operonic contexts with VanZ-1 (Figure 5V). Hence, given our above prediction

regarding VanZ acting in peptidoglycan and/or exopolysaccharide metabolism, VanZ-1 and the Lipocones displacing it might indeed play an accessory role with VanY at the membrane (204,205) in antibiotic resistance. We posit that, in these contexts, it likely acts on the head group of Lipid II to recycle canonical peptidoglycan intermediates for their accelerated or more thorough replacement with the resistant versions (Figure 3G).

We also identified a conserved five-gene operon featuring a YfiM-1 family Lipocone that might play a role in resistance to antibacterial agents (Figures 5W,S4). Other than YfiM-1, this operon contains genes for: (i) a thioredoxin domain protein (109); (ii) A DTW clade RNA modifying enzymes of the SPOUT superfamily (207,208); (iii) a protein with acyl-CoA ligase, GNAT superfamily N-acetyltransferase and ATP-grasp domains (131,209,210); (iv) a PssA-like phosphatidylserine synthetase of the HKD superfamily (211) (Figure 5W, Supplementary Data). Of these enzymes, the phosphatidylserine synthetase is predicted to act in its usual capacity to generate a lipid with a serine head group (211). We propose that this would then function as a substrate for the YfiM-1 Lipocone domain, which might exchange the serine for another moiety via a reaction paralleling PTDSS1/2 (Figure 3A). This moiety could then be modified by aminoacylation, further acylation and a redox modification by the third protein listed above, together with the thioredoxin. Indeed, such peptide modifications of lipid head groups by lysine, alanine, or arginine aminoacylation catalyzed by derived tRNA synthetases fused to GNATs have been shown to be a key resistance mechanism against breaching of the membrane by antibacterial peptides (212,213). Hence, we predict the modifications catalyzed by this system might play a comparable role. The presence of a tRNA-modifying DTW domain suggests that in parallel to the tRNA synthetases, the GNAT in this system might use a tRNA-linked acyl group as a substrate, as seen in peptidoglycan biosynthesis (214,215).

Eukaryotic recruitments of the Lipocone superfamily

Lipocone domains have been transferred on several occasions from bacteria to eukaryotes (Figures 4,S1). While there is predicted functional overlap with the above-described, predominantly bacterial

versions, we discuss these separately as the inferred biological contexts of their deployment are often distinct from the above.

Plant YfiM-1 and eukaryotic VanZ-2 proteins

A conserved YfiM-1 family protein typified by the *Arabidopsis* AT1G15900 was acquired from the bacteroidetes lineage of bacteria at the base of the plant lineage prior to the chlorophyte-streptophyte (including land plants) split and is predicted to be catalytically active (Figure S2, Supplementary Data). In *Arabidopsis*, this gene is widely expressed across different tissue types, developmental stages, and other tested conditions (216,217). Given the above-predicted roles for bacterial YfiM-1 proteins, it is conceivable that the plant version plays a comparable role in the metabolism of a conserved plant-specific lipid. In a similar vein, a distinct clade of standalone VanZ-2 domains typified by the *Saccharomyces cerevisiae* YJR112W-A was acquired early in the fungal lineage. A similar transfer is also seen in the SAR clade of eukaryotes (Figure S1). Since these eukaryotes lack peptidoglycan and other bacterial-type isoprenoid lipid-borne exopolysaccharide intermediates, we suggest that this version was recruited for modifications of a fungus-specific lipid (e.g., highly oxygenated isoprenoid lipids) (218).

The Met-TelC proteins

The Met-TelC clade is comprised of versions of the TelC family with a reconfigured active site transferred from bacteria to Metazoa prior to the divergence of the cnidarians, and most members are predicted to be catalytically inactive (Figure 2). In cnidarians and arthropods, the Met-TelC domain is found in a secreted protein fused to C-terminal adhesion-related vWA (219) and Ig domains, followed by a TM helix (Figure 5X, Supplementary Data). The chordate version, typified by human PGLYRP2 (220), is also secreted and is fused to a C-terminal Amidase targeting the N-acetylmuramoyl-L-alanine linkage (Figure 5X, Supplementary Data). The chordate version is a key innate immunity factor against bacterial pathogens that degrades sugar-peptide linkages in peptidoglycan via the Amidase domain (221-223). As most Met-TelC proteins lack the active site residues but are modeled to retain the

783 substrate-binding pocket, we propose that they participate in anti-bacterial immunity as a Pathogen-
784 Associated Molecular Pattern (PAMP) receptor (224). Specifically, they could recognize polyisoprenoid
785 pyrophosphate-linkage-containing lipid intermediates of bacterial cell-surface molecules like
786 peptidoglycan or exopolysaccharides.
787
788 *Eukaryotic Wnt proteins*
789 Wnt family Lipocones were transferred on multiple occasions to eukaryotes. The best known of these
790 are Met-Wnt proteins, which were acquired from bacteria at the base of Metazoa after they had
791 separated from their closest sister group, the choanoflagellates. These lost the ancestral active site
792 residues and function as well-studied secreted signaling molecules and will not be detailed further in
793 this work (for review, see (1,225)). Independently of the Met-Wnt proteins, catalytically active, secreted
794 versions closely related to the bacterial Min-Wnt proteins were transferred to fungi and, within
795 Metazoa, to the rotifers and the hemichordate acorn worm *Saccoglossus kowalevskii*, where they are
796 lineage-specifically expanded (Figure S1, Supplementary Data). These versions are primarily standalone
797 versions of the Min-Wnt domain lacking the large inserts typical of the Met-Wnt proteins. We predict
798 that these eukaryotic Min-Wnt proteins retain their ancestral role and might participate in anti-
799 bacterial immunity in the lineages coding for them.
800
801 *Met-SAA proteins*
802 Met-SAA proteins were acquired from bacteria prior to the divergence of the cnidarians from the rest
803 of Metazoa. However, unlike the Met-Wnt and Met-TelC proteins, they often conserve the ancestral
804 active site residues indicating that they are usually enzymatically active (Figure 3). Human SAA has
805 been recognized as a key immune marker that dramatically increases in blood during the Acute Phase
806 Response (226). It has been reported to bind the *E. coli* outer membrane protein OmpA (227) and claimed
807 to function as an opsonin in innate immunity (228). Like Met-TelC, but in contrast to Met-Wnts, Met-
808 SAAs appear to have been lost or pseudogenized in several animal lineages (229-231) (Figure 3). This is
809 consistent with an arms-race scenario in immunity and the development of pathogen resistance

against the Met-SAAs, leading to loss. Keeping with an immune role for the Met-SAAs, we propose a catalytic function for the active versions in severing lipid head groups of outer-membrane lipids or of isoprenoid lipid carrier intermediates. Such action could also generate PAMPs that could explain the activation of neutrophil- and macrophage-based immunity by SAA (228). Pertinent to these observations, diverse OMP β-barrels have been linked to the translocation of polymorphic toxin domains across the outer membrane of target cells (232-234). Given this and the origin of Met-SAA from bacterial polymorphic toxin-related systems (Figure 4), its interaction with OmpA might help it cross over into the periplasmic space and act on maturing peptidoglycan or teichoic acid intermediates.

SAA was first reported as a component of secondary amyloid deposits (235), and its capacity to form amyloid fibrils upon protease cleavage was theorized as a potential PAMP activating the immune response (30). Indeed, bacteria produce their own secreted amyloids, such as Curli and Fap, believed to contribute to biofilm formation (236,237), and might be PAMPs recognized by animal immune systems (238). Further, other animal amyloids, such as the β-amyloid, have been proposed to play a role as physical barriers in immunity against bacteria (239). Thus, amyloid formation by protease cleavage (including potentially by bacterial proteases) may represent a second line of defense mediated by Met-SAA proteins.

DISCUSSION

Early Evolution of the Lipocone superfamily

No single well-defined Lipocone clade is universally conserved across the three superkingdoms of Life (Figures 4, S1). However, the VanZ and Wok clades are both found across all major bacterial phyla (notwithstanding sporadic losses in certain lineages) and in some archaeal lineages (Figure S1). At the same time, the cpCone clade is found across most major archaeal lineages and is nearly universally conserved in the eukaryotes (absent in Ascomycota and some choanoflagellates) (Figure S1). Notably, the cpCone and Wok clades tend to group together in the profile-profile similarity network (Figure 1B).

These observations suggest that at least a single version of the Lipocone superfamily was likely present in the Last Universal Common Ancestor (LUCA). The phyletic patterns suggest that the LUCA Lipocone gave rise to the VanZ/Wok precursor in the bacterial lineage on the one hand and the cpCONE clade via a circular permutation event in the archaeo-eukaryotic lineage on the other (Figure 4). Based on the features of these deep-branching clades, the LUCA version is inferred to feature a hydrophobic domain with a 4TM helix core, with the active site facing the outer leaf of the lipid bilayer (Figure 1C). Given that extant versions operate both on classic phospholipids and isoprenoid lipids, it is difficult to infer which of these might have been substrates for the LUCA version. It is not impossible that this early version had a generic specificity that became specialized in the descendant clades.

Subsequent diversification of the Lipocone domain

The early diversification of the Lipocone domain appears to have had different drivers in the two prokaryotic superkingdoms. The presence of an extensive repertoire of exopolysaccharides in the cell wall (peptidoglycan, teichoic acids), cell surface (e.g., ECA), and outer membrane (e.g., lipopolysaccharide), synthesized via isoprenoid lipid-linked intermediates, like lipid-II, was the primary driver in the bacterial superkingdom (240). Here, this diversification yielded 4 monophyletic groups: the VanZs, Wok, YfiM/Griddle-like and Skillet (Figure 4). The deeper VanZ and Wok branches, which were recruited first for lipid-II-related functions, were likely the predecessors of the more restricted bacterial families with specialized functions. For instance, the emergence of the outer membrane in certain bacteria was potentially coupled with the origin of the YfiM and Griddle clades (Figure 4). Similarly, our predictions suggest that within these clades, further diversification accompanied the acquisition of specialized functional roles in antibiotic resistance (Skillet-DUF2809, YfiM-1), secondary sensor roles in single and multicomponent signaling (VanZ and Skillet families) and lipoprotein processing (Skillet-1). The interoperability of Lipocone domains on lipid carriers shared across different biosynthetic pathways (see above, Figures 3,S4-S5) appears to have been a key factor leading to this versatility.

In the ancestral archaeo-eukaryotic lineage, the absence of peptidoglycan and an apparently lower diversity of structures with exopolysaccharides was reflected in the lesser diversification of the Lipocone clades (Figure 4). There are open questions regarding the biochemical functions of the primary archaeo-eukaryotic Lipocone clade, the cpCONE. Although the eukaryotic cpCone PTDSS1/2 family has been shown to swap serine for ethanolamine or choline in lipid head groups (28,57), their archaeal counterparts remain uncharacterized. Archaea have their own lipid with a serine in the head group (archaeophosphatidyserine), but to date, its synthesis has been shown to depend on a patchwork of different CDP-alcohol phosphatidyltransferase enzymes (CaPs) in different archaeal species (58,241,242). While the CaPs are also integral membrane enzymes with a 6TM helix core catalyzing similar reactions as the Lipocones on lipid head groups in archaea and eukaryotes (242), they are evolutionarily unrelated domains. Nevertheless, we suggest that the archaeal cpCones, like their eukaryotic counterparts, could contribute to distinct, as yet uncharacterized, pathways for the generation of cell membrane phospholipids like archaeophosphatidylserine or those with other head groups under special conditions.

Emergence of soluble versions of the Lipocone domain and their repeated recruitment in biological conflicts

One of the remarkable aspects of the Lipocone superfamily is the loss of ancestral hydrophobicity in several families (Figure 1C), transforming them from integral membrane proteins to soluble domains. While unexpected, such a transition in integral membrane enzymes acting on lipid substrates is not unprecedented. The PAP2 superfamily of integral membrane enzymes (e.g., diacylglycerol diphosphate phosphatase) (124,243) also contains several soluble versions (244) that appear to have emerged from an integral membrane ancestor (AMB and LA, unpublished observations). Most of the soluble Lipocone domains retain their active site conservation (Figure 2) and, at least in one experimentally characterized case, catalyze a comparable reaction as the TM version (29) (Figure 3B). The weight of the evidence presented here, including the profile-profile similarity network (Figure 1B), phyletic patterns (Figure S1), functional contexts (Figure 5-6), and the broadly shared structural features (Figures 1A, S2),

suggests that the loss of hydrophobicity occurred on a single occasion in the Lipocone superfamily, followed by diversification of these soluble versions.

Our analysis of the soluble Lipocone families reveals repeated recruitment as toxins/effectors in anti-viral and polymorphic toxin and allied systems (174), suggesting that their diversification was driven by the arms races arising from the biological conflicts where they are deployed. Recruitment of a representative of the VanZ-1 family as a polymorphic toxin on rare occasions (Figures 5N,6) suggests a possible evolutionary pathway by which this might have happened: the soluble effector version of Lipocones attacking lipids in competing bacteria likely emerged from an ancestral version that catalyzed endogenous lipid-head-group modifications on the same lipids in metabolic pathways. Once soluble versions emerged through reduced hydrophobicity, they could be deployed as diffusible effectors that were shared across extracellular and intracellular conflict systems, a trend previously recognized in many other effector domains (25).

Repeated acquisition of Lipocones of bacterial origin by eukaryotes

Unlike bacteria, eukaryotes as a whole do not possess a rich repertoire of Lipocone domains. The PTDSS1/2 family, vertically inherited from the archaeal progenitor, is the only version that can be inferred as being present in the Last Eukaryotic Common Ancestor (Figure 4). However, as noted above, particular eukaryotic lineages have acquired and fixed distinct Lipocone families of ultimately bacterial provenance: (i) YfiM-1 in the plant lineage; (ii) the fungal VanZ-2 domains typified by the *Saccharomyces cerevisiae* YJR112W-A; (iii) Met-Wnt (discussed further below) (Figure S1). The fixation of these versions early in the history of the eukaryotic lineages possessing them suggests that they were recruited for "housekeeping" or developmental roles definitive of the respective lineages.

Beyond these, the fungal and metazoan lineages show more sporadically distributed versions, which have all been acquired from bacterial secreted-toxin or antiviral systems: (i) Min-Wnt independently in fungi and certain Metazoa; (ii) SAA; (iii) TelC; the latter two are absent in the basal-most metazoans, the

918 sponges, but are present in Cnidaria, suggesting a relatively early acquisition (Figure 4). The weight of
919 the evidence suggests that they have retained certain aspects of the ancestral bacterial effector
920 function for anti-pathogen immunity in eukaryotes. This is consistent with both their episodic loss and
921 lineage-specific expansion, the tendency to show rapid sequence divergence and, in the latter family,
922 loss of catalytic activity (Figures 2,S1, Supplementary Data).
923
924 This independent acquisition of at least 3 distinct Lipocone families in metazoan immunity from
925 polymorphic and allied effector systems of prokaryotes points to a persistent evolutionary trend.
926 Notably, the Lipocone domains participating in animal immunity have been drawn from secreted
927 effectors rather than the intracellular versions (bacterial intracellular Min-Wnts) predicted to
928 participate in bacterial anti-selfish element immunity. More generally, this adds to a growing list of
929 components drawn from secreted effector systems of prokaryotes in eukaryotic immune systems
930 (174,193,245). For example, this closely parallels another structurally unrelated effector domain, the Zn-
931 dependent deaminase (e.g., metazoan AID/APOBEC deaminases) (246). Hence, these observations add
932 further support to our hypothesis that the extensive expansion of effectors in diverse prokaryotic
933 inter-organismal conflict systems served as a reservoir from which eukaryotic immune systems
934 repeatedly acquired components (193,245). We propose that symbiotic associations between the early
935 animals and bacteria resulted in potential interactions via secreted effectors of the latter that aided the
936 former against antagonistic bacteria. This probably led to their eventual acquisition by animals and
937 incorporation into their immune processes.
938
939 Origin of Wnt as a signaling molecule
940 Earlier considerations on the evolution of Wnt signaling indicated that it emerged at the base of the
941 metazoan lineage and incorporated a wide range of components of different origins (e.g., the HMG
942 domain transcription factor TCF/LEF, the HEAT repeat protein β-catenin and the 7TM receptor Frizzled)
943 (1). However, the provenance of Met-Wnt itself had been mysterious and was seen as a possible example
944 of a metazoan innovation (225). While the Met-Wnt domains show Metazoa-specific elaborations (34),

its conserved core is a Lipocone domain (Figure 1A). We establish that the progenitor of Met-Wnt emerged as part of the radiation of Lipocone domains in bacteria as effectors deployed in both intracellular and inter-organismal conflict – the Min-Wnt proteins.

Whereas the prokaryotic Min-Wnt proteins are predicted to be secreted toxins targeting phosphate-linked moieties in lipid head groups, the Met-Wnts underwent an ancestral inactivation through loss of the catalytic residues (Figure 2). However, they retained their ancient involvement in cell-cell interactions as secreted agents. The Met-Wnt residues recognized as essential for the receptor (Frizzled) binding, including the absolutely conserved palmitoleoylated serine residue, are found in the aforementioned Metazoa-specific hairpins and loops (34,36). However, despite their inactivation, the Met-Wnts retain the ancestral substrate-binding pocket (Figures 1A, S2). This raises the possibility that they might be involved in as-yet unexplored interactions with ligands such as lipids.

Our tracing of the provenance of Wnt back to an effector in secreted bacterial toxin systems adds it to a growing list of components in metazoan signaling networks that have been acquired from such systems. For instance, this is also the case with components of the other key metazoan signaling pathway, Hedgehog (247). Here the Hedgehog protein itself contains an autoproteolytic HINT peptidase domain that was likely drawn from a structurally and functionally cognate domain observed in polymorphic toxin systems (174,247). Further, an intracellular component of the same signaling pathway, Supressor of Fused (SuFu) was derived from a common immunity protein found in polymorphic toxin systems (247). Similarly, the Teneurin/Odd Oz proteins mediating signaling in cell migration, neuronal pathfinding, and fasciculation in Metazoa descended from a polymorphic toxin protein with a C-terminal HNH endonuclease toxin tip (174). In a similar vein, the immunity protein of certain CapCone toxins identified in this study might have given rise to the β-sandwich domain in the eukaryotic centriolar assembly factor SAS-6. These observations suggest that in addition to immune system components, interactions with symbiotic bacteria also potentially provided the progenitors of

971 components of eukaryotic signaling and cytoskeletal networks that were central to the emergence of
972 Metazoa as a clade of multicellular eukaryotes (248,249).
973
974 CONCLUSIONS
975 Using sensitive sequence and structure analysis, we unify a large, hitherto unrecognized superfamily of
976 enzymatic domains, the Lipocone. By combining analysis of the active site and the structure of the
977 Lipocone domain with contextual information from conserved gene-neighborhoods and domain
978 architectures, we present evidence that members of this superfamily target phosphate linkages in head
979 groups of both classical phospholipids and polyisoprenoid lipids. Specifically, they catalyze reactions
980 such as head group exchange or severing of the head group-diphosphate linkage from the
981 polyisoprenol. We present evidence that these activities have been recruited in a wide range of
982 biochemical contexts, including cell membrane lipid modification, metabolism of peptidoglycan and
983 exopolysaccharide lipid-carrier linked intermediates, lipoprotein modifications, bacterial outer
984 membrane modification, sensing of membrane-associated signals, effector activity in antiviral and
985 inter-organismal conflicts and resistance to antimicrobials. Further, catalytically inactive versions (e.g.,
986 Met-Wnt) have been recruited for signaling roles in Metazoa. We predict the catalytic activity and
987 potential biochemical pathways of numerous representatives for the first time, including some proteins
988 that have remained enigmatic for over two decades, like VanZ.
989
990 We identify three notable trends in Lipocone evolution. First, although we reconstruct the ancestral
991 member of the superfamily as being a 4TM integral membrane domain, a large monophyletic subset
992 underwent a dramatic loss of hydrophobicity, transforming them into diffusible versions, including the
993 Wnts and the SAAs (Figure 1C). Second, the superfamily expanded in two major functional niches in
994 bacteria, namely peptidoglycan/exopolysaccharide metabolism and effector domains of both secreted
995 toxins and immune systems (Figure 4). Finally, members of the Lipocone superfamily were acquired on
996 multiple occasions from bacteria by Metazoa and were reused in new functional contexts as signaling
997 messengers and immune factors (Figure 4).

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

METHODS

Sequence analysis

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

Sequence similarity network analysis

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

1025 **Comparative genomics, domain identification, and phylogenetic analysis**

1026 Genomic neighborhoods were obtained from genomes available in the NCBI Genome database (252)
1027 using in-house scripts written in Perl and Python. Conservation analysis of these genomic
1028 neighborhoods was performed by clustering the protein products of neighboring genes. Domain
1029 identification was conducted using a collection of HMMs and PSSMs maintained by the Aravind lab,
1030 along with HMMs from the Pfam database (258), utilizing the RPSBLAST (259) and HMMSCAN (260)
1031 programs. To further refine detection, domain identification was extended through remote homology
1032 analysis using the HHpred (261) program, against profiles built from the Pfam (258) and PDB70 (262)
1033 databases. Phylogenetic analyses were performed using FastTree (263) and iqTREE2 (264). Structural
1034 comparisons, along with shared genomic associations, were used to further refine the detailed
1035 interrelationships within and between the groups of the superfamily. Families with broader phyletic
1036 patterns and conservation across multiple major lineages ("phyla") were assumed to be more ancient in
1037 their emergence, and those with a more limited distribution were assumed to be later derivations
1038 (Figures 4,S1).

1039

1040 Contextual Network construction

1041 Each domain architecture and conserved gene neighborhood was decomposed into their constituent
1042 domains. These domains were then labeled for their biochemical function and stored as a YAML file
1043 (Supplementary Data). The contextual connections were then rendered as a graph with the domains as
1044 its nodes and the adjacency relationships as its edges. Cliques containing a given Lipocone domain were
1045 detected in this graph and merged to constitute their respective dense subgraphs. These subgraphs
1046 were then examined for the statistically significant prevalence of particular labeled functions using the
1047 Fisher exact test. Network analysis was performed using the functions of the R igraph or Python
1048 networkX libraries.

1049

1050 Structure analysis

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

Hydrophobicity analysis

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

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

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 amino acids.

FIGURE LEGENDS

Figure 1. (A) The four individual helices forming the core of the Lipocone superfamily are colored consistently across the structures from representative families. 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 are 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 conserved core elements of different Lipocone families, generally corresponding to the core H2, H3, and H4 helices. The three conserved core 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.

Figure 4. Reconstructed evolutionary scenario for the diversification of 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. Broken lines indicate uncertainty in the starting point of a lineage(s). Inferred or experimentally characterized functions for families are depicted by colored boxes to the left of family names.

Figure 5. Representative contexts for Lipocones, 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 geometric 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 (represented as hexagonal nodes). Nodes and edges are colored based on the known or inferred functional categories of the domains, with groups labeled in matching colors. The nodes are scaled by 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 protein kinase; HUP, HIGH-signature, 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.
9. Segalen, M. and Bellaiche, Y. (2009) Cell division orientation and planar cell polarity pathways. *Semin Cell Dev Biol*, **20**, 972-977.
10. Slusarski, D.C. and Pelegri, F. (2007) Calcium signaling in vertebrate embryonic patterning and morphogenesis. *Dev Biol*, **307**, 1-13.
11. Logan, C.Y. and Nusse, R. (2004) The Wnt signaling pathway in development and disease. *Annu Rev Cell Dev Biol*, **20**, 781-810.
12. Welters, H.J. and Kulkarni, R.N. (2008) Wnt signaling: relevance to beta-cell biology and diabetes. *Trends Endocrinol Metab*, **19**, 349-355.
13. Christian, J.L. (2000) BMP, Wnt and Hedgehog signals: how far can they go? *Curr Opin Cell Biol*, **12**, 244-249.
14. Schulte, G. and Bryja, V. (2007) The Frizzled family of unconventional G-protein-coupled receptors. *Trends Pharmacol Sci*, **28**, 518-525.
15. Bhanot, P., Brink, M., Samos, C.H., Hsieh, J.C., Wang, Y., Macke, J.P., Andrew, D., Nathans, J. and Nusse, R. (1996) A new member of the frizzled family from Drosophila functions as a Wingless receptor. *Nature*, **382**, 225-230.
16. Takada, R., Satomi, Y., Kurata, T., Ueno, N., Norioka, S., Kondoh, H., Takao, T. and Takada, S. (2006) Monounsaturated fatty acid modification of Wnt protein: its role in Wnt secretion. *Dev Cell*, **11**, 791-801.
17. Kurayoshi, M., Yamamoto, H., Izumi, S. and Kikuchi, A. (2007) Post-translational palmitoylation and glycosylation of Wnt-5a are necessary for its signalling. *Biochem J*, **402**, 515-523.
18. Gao, C. and Chen, Y.G. (2010) Dishevelled: The hub of Wnt signaling. *Cell Signal*, **22**, 717-727.
19. Klingensmith, J., Nusse, R. and Perrimon, N. (1994) The Drosophila segment polarity gene dishevelled encodes a novel protein required for response to the wingless signal. *Genes Dev*, **8**, 118-130.
20. Kramps, T., Peter, O., Brunner, E., Nellen, D., Froesch, B., Chatterjee, S., Murone, M., Zullig, S. and Basler, K. (2002) Wnt/wingless signaling requires BCL9/legless-mediated recruitment of pygopus to the nuclear beta-catenin-TCF complex. *Cell*, **109**, 47-60.
21. van Tienen, L.M., Mieszczanek, J., Fiedler, M., Rutherford, T.J. and Bienz, M. (2017) Constitutive scaffolding of multiple Wnt enhanceosome components by Legless/BCL9. *Elife*, **6**.
22. Archbold, H.C., Yang, Y.X., Chen, L. and Cadigan, K.M. (2012) How do they do Wnt they do?: regulation of transcription by the Wnt/beta-catenin pathway. *Acta Physiol (Oxf)*, **204**, 74-109.
23. Holstein, T.W. (2012) The evolution of the Wnt pathway. *Cold Spring Harb Perspect Biol*, **4**, a007922.
24. Burroughs, A.M. and Aravind, L. (2020) Identification of Uncharacterized Components of Prokaryotic Immune Systems and Their Diverse Eukaryotic Reformulations. *J Bacteriol*, **202**.

25. Aravind, L., Iyer, L.M. and Burroughs, A.M. (2022) Discovering Biological Conflict Systems Through Genome Analysis: Evolutionary Principles and Biochemical Novelty. *Annu Rev Biomed Data Sci*, **5**, 367-391.
26. Kuge, O., Saito, K. and Nishijima, M. (1997) Cloning of a Chinese hamster ovary (CHO) cDNA encoding phosphatidylserine synthase (PSS) II, overexpression of which suppresses the phosphatidylserine biosynthetic defect of a PSS I-lacking mutant of CHO-K1 cells. *J Biol Chem*, **272**, 19133-19139.
27. Kuge, O., Nishijima, M. and Akamatsu, Y. (1991) A Chinese hamster cDNA encoding a protein essential for phosphatidylserine synthase I activity. *J Biol Chem*, **266**, 24184-24189.
28. Stone, S.J. and Vance, J.E. (1999) Cloning and expression of murine liver phosphatidylserine synthase (PSS)-2: differential regulation of phospholipid metabolism by PSS1 and PSS2. *Biochem J*, **342 (Pt 1)**, 57-64.
29. Whitney, J.C., Peterson, S.B., Kim, J., Pazos, M., Verster, A.J., Radey, M.C., Kulasekara, H.D., Ching, M.Q., Bullen, N.P., Bryant, D., Goo, Y.A., Surette, M.G., Borenstein, E., Vollmer, W. and Mougous, J.D. (2017) A broadly distributed toxin family mediates contact-dependent antagonism between gram-positive bacteria. *Elife*, **6**.
30. Sack, G.H., Jr. (2018) Serum amyloid A - a review. *Mol Med*, **24**, 46.
31. Arthur, M., Depardieu, F., Molinas, C., Reynolds, P. and Courvalin, P. (1995) The vanZ gene of Tn1546 from Enterococcus faecium BM4147 confers resistance to teicoplanin. *Gene*, **154**, 87-92.
32. Arthur, M., Depardieu, F., Reynolds, P. and Courvalin, P. (1999) Moderate-level resistance to glycopeptide LY333328 mediated by genes of the vanA and vanB clusters in enterococci. *Antimicrob Agents Chemother*, **43**, 1875-1880.
33. Bazan, J.F., Janda, C.Y. and Garcia, K.C. (2012) Structural architecture and functional evolution of Wnts. *Dev Cell*, **23**, 227-232.
34. Janda, C.Y., Waghray, D., Levin, A.M., Thomas, C. and Garcia, K.C. (2012) Structural basis of Wnt recognition by Frizzled. *Science*, **337**, 59-64.
35. Chu, M.L., Ahn, V.E., Choi, H.J., Daniels, D.L., Nusse, R. and Weis, W.I. (2013) structural Studies of Wnts and identification of an LRP6 binding site. *Structure*, **21**, 1235-1242.
36. Zhong, Q., Zhao, Y., Ye, F., Xiao, Z., Huang, G., Xu, M., Zhang, Y., Zhan, X., Sun, K., Wang, Z., Cheng, S., Feng, S., Zhao, X., Zhang, J., Lu, P., Xu, W., Zhou, Q. and Ma, D. (2021) Cryo-EM structure of human Wntless in complex with Wnt3a. *Nat Commun*, **12**, 4541.
37. Traag, V.A., Waltman, L. and van Eck, N.J. (2019) From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep*, **9**, 5233.
38. Zhao, G. and London, E. (2006) An amino acid "transmembrane tendency" scale that approaches the theoretical limit to accuracy for prediction of transmembrane helices: relationship to biological hydrophobicity. *Protein Sci*, **15**, 1987-2001.
39. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E.,

Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P. and Hassabis, D. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583-589.

40. Bastard, K., Smith, A.A., Vergne-Vaxelaire, C., Perret, A., Zaparucha, A., De Melo-Minardi, R., Mariage, A., Boutard, M., Debard, A., Lechaplais, C., Pelle, C., Pellouin, V., Perchat, N., Petit, J.L., Kreimeyer, A., Medigue, C., Weissenbach, J., Artiguenave, F., De Berardinis, V., Vallenet, D. and Salanoubat, M. (2014) Revealing the hidden functional diversity of an enzyme family. *Nat Chem Biol*, **10**, 42-49.

41. Glasner, M.E., Gerlt, J.A. and Babbitt, P.C. (2006) Evolution of enzyme superfamilies. *Curr Opin Chem Biol*, **10**, 492-497.

42. Zhang, D., Iyer, L.M., Burroughs, A.M. and Aravind, L. (2014) Resilience of biochemical activity in protein domains in the face of structural divergence. *Curr Opin Struct Biol*, **26**, 92-103.

43. Tomohiro, S., Kawaguti, A., Kawabe, Y., Kitada, S. and Kuge, O. (2009) Purification and characterization of human phosphatidylserine synthases 1 and 2. *Biochem J*, **418**, 421-429.

44. Saito, K., Kuge, O., Akamatsu, Y. and Nishijima, M. (1996) Immunochemical identification of the pssA gene product as phosphatidylserine synthase I of Chinese hamster ovary cells. *FEBS Lett*, **395**, 262-266.

45. Stone, S.J. and Vance, J.E. (2000) Phosphatidylserine synthase-1 and -2 are localized to mitochondria-associated membranes. *J Biol Chem*, **275**, 34534-34540.

46. Miyata, N. and Kuge, O. (2021) Topology of phosphatidylserine synthase 1 in the endoplasmic reticulum membrane. *Protein Sci*, **30**, 2346-2353.

47. Anderson, J.S., Matsuhashi, M., Haskin, M.A. and Strominger, J.L. (1967) Biosythesis of the peptidoglycan of bacterial cell walls. II. Phospholipid carriers in the reaction sequence. *J Biol Chem*, **242**, 3180-3190.

48. Higashi, Y., Strominger, J.L. and Sweeley, C.C. (1967) Structure of a lipid intermediate in cell wall peptidoglycan synthesis: a derivative of a C55 isoprenoid alcohol. *Proc Natl Acad Sci U S A*, **57**, 1878-1884.

49. English, D. (1996) Phosphatidic acid: a lipid messenger involved in intracellular and extracellular signalling. *Cell Signal*, **8**, 341-347.

50. Zamocky, M. and Ferianc, P. (2023) Discovering the deep evolutionary roots of serum amyloid A protein family. *Int J Biol Macromol*, **252**, 126537.

51. Woods, E.C., Wetzel, D., Mukerjee, M. and McBride, S.M. (2018) Examination of the Clostridioides (Clostridium) difficile VanZ ortholog, CD1240. *Anaerobe*, **53**, 108-115.

52. Sur, V.P., Mazumdar, A., Vimberg, V., Stefani, T., Androvic, L., Kracikova, L., Laga, R., Kamenik, Z. and Komrskova, K. (2021) Specific Inhibition of VanZ-Mediated Resistance to Lipoglycopeptide Antibiotics. *Int J Mol Sci*, **23**.

53. Stogios, P.J. and Savchenko, A. (2020) Molecular mechanisms of vancomycin resistance. *Protein Sci*, **29**, 654-669.

54. Dziarski, R. and Gupta, D. (2006) The peptidoglycan recognition proteins (PGRPs). *Genome Biol*, **7**, 232.

1264 55. Bliven, S. and Prlic, A. (2012) Circular permutation in proteins. *PLoS Comput Biol*, **8**, e1002445.
1265 56. Aravind, L. (2000) Guilt by association: contextual information in genome analysis. *Genome*
1266 *Res*, **10**, 1074-1077.
1267 57. Vance, J.E. (2018) Historical perspective: phosphatidylserine and phosphatidylethanolamine
1268 from the 1800s to the present. *J Lipid Res*, **59**, 923-944.
1269 58. Koga, Y. and Morii, H. (2005) Recent advances in structural research on ether lipids from
1270 archaea including comparative and physiological aspects. *Biosci Biotechnol Biochem*, **69**, 2019-
1271 2034.
1272 59. Caforio, A. and Driessen, A.J.M. (2017) Archaeal phospholipids: Structural properties and
1273 biosynthesis. *Biochim Biophys Acta Mol Cell Biol Lipids*, **1862**, 1325-1339.
1274 60. Rezanka, T., Kyselova, L. and Murphy, D.J. (2023) Archaeal lipids. *Prog Lipid Res*, **91**, 101237.
1275 61. Narita, S. and Tokuda, H. (2010) Sorting of bacterial lipoproteins to the outer membrane by
1276 the Lol system. *Methods Mol Biol*, **619**, 117-129.
1277 62. Kamada, K., Miyata, M. and Hirano, T. (2013) Molecular basis of SMC ATPase activation: role of
1278 internal structural changes of the regulatory subcomplex ScpAB. *Structure*, **21**, 581-594.
1279 63. Schleiffer, A., Kaitna, S., Maurer-Stroh, S., Glotzer, M., Nasmyth, K. and Eisenhaber, F. (2003)
1280 Kleisins: a superfamily of bacterial and eukaryotic SMC protein partners. *Mol Cell*, **11**, 571-575.
1281 64. Soppa, J., Kobayashi, K., Noirot-Gros, M.F., Oesterhelt, D., Ehrlich, S.D., Dervyn, E., Ogasawara,
1282 N. and Moriya, S. (2002) Discovery of two novel families of proteins that are proposed to
1283 interact with prokaryotic SMC proteins, and characterization of the Bacillus subtilis family
1284 members ScpA and ScpB. *Mol Microbiol*, **45**, 59-71.
1285 65. Aravind, L., Anantharaman, V., Balaji, S., Babu, M.M. and Iyer, L.M. (2005) The many faces of
1286 the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol Rev*, **29**, 231-
1287 262.
1288 66. Pellicer, M.T., Badia, J., Aguilar, J. and Baldoma, L. (1996) glc locus of Escherichia coli:
1289 characterization of genes encoding the subunits of glycolate oxidase and the glc regulator
1290 protein. *J Bacteriol*, **178**, 2051-2059.
1291 67. Ramachandran, R., Hartmann, C., Song, H.K., Huber, R. and Bochtler, M. (2002) Functional
1292 interactions of HslV (ClpQ) with the ATPase HslU (ClpY). *Proc Natl Acad Sci U S A*, **99**, 7396-7401.
1293 68. Storck, E.M., Ozbalci, C. and Eggert, U.S. (2018) Lipid Cell Biology: A Focus on Lipids in Cell
1294 Division. *Annu Rev Biochem*, **87**, 839-869.
1295 69. Barak, I. and Muchova, K. (2013) The role of lipid domains in bacterial cell processes. *Int J Mol*
1296 *Sci*, **14**, 4050-4065.
1297 70. Hauck, A.K. and Bernlohr, D.A. (2016) Oxidative stress and lipotoxicity. *J Lipid Res*, **57**, 1976-
1298 1986.
1299 71. Guo, D. and Tropp, B.E. (2000) A second Escherichia coli protein with CL synthase activity.
1300 *Biochim Biophys Acta*, **1483**, 263-274.
1301 72. Tan, B.K., Bogdanov, M., Zhao, J., Dowhan, W., Raetz, C.R. and Guan, Z. (2012) Discovery of a
1302 cardiolipin synthase utilizing phosphatidylethanolamine and phosphatidylglycerol as
1303 substrates. *Proc Natl Acad Sci U S A*, **109**, 16504-16509.

73. Aravind, L. and Koonin, E.V. (1998) Phosphoesterase domains associated with DNA polymerases of diverse origins. *Nucleic Acids Res*, **26**, 3746-3752.
74. Burroughs, A.M., Allen, K.N., Dunaway-Mariano, D. and Aravind, L. (2006) Evolutionary genomics of the HAD superfamily: understanding the structural adaptations and catalytic diversity in a superfamily of phosphoesterases and allied enzymes. *J Mol Biol*, **361**, 1003-1034.
75. Vershinin, A. (1999) Biological functions of carotenoids--diversity and evolution. *Biofactors*, **10**, 99-104.
76. Šesták, Z., Britton, G., Liaaen-Jensen, S. and Pfander, H. (2004) Carotenoids. Handbook. *Photosynthetica*, **42**, 186.
77. Sandmann, G. and Misawa, N. (1992) New functional assignment of the carotenogenic genes crtB and crtE with constructs of these genes from Erwinia species. *FEMS Microbiol Lett*, **69**, 253-257.
78. Chamovitz, D., Misawa, N., Sandmann, G. and Hirschberg, J. (1992) Molecular cloning and expression in Escherichia coli of a cyanobacterial gene coding for phytoene synthase, a carotenoid biosynthesis enzyme. *FEBS Lett*, **296**, 305-310.
79. Liu, J. and Mushegian, A. (2003) Three monophyletic superfamilies account for the majority of the known glycosyltransferases. *Protein Sci*, **12**, 1418-1431.
80. Tran, U.C. and Clarke, C.F. (2007) Endogenous synthesis of coenzyme Q in eukaryotes. *Mitochondrion*, **7 Suppl**, S62-71.
81. Massiere, F. and Badet-Denisot, M.A. (1998) The mechanism of glutamine-dependent amidotransferases. *Cell Mol Life Sci*, **54**, 205-222.
82. Liu, X., Zeng, J., Chen, X. and Xie, W. (2016) Crystal structures of RidA, an important enzyme for the prevention of toxic side products. *Sci Rep*, **6**, 30494.
83. Aravind, L. and Koonin, E.V. (1998) The HD domain defines a new superfamily of metal-dependent phosphohydrolases. *Trends Biochem Sci*, **23**, 469-472.
84. Becker, A., Kleickmann, A., Kuster, H., Keller, M., Arnold, W. and Puhler, A. (1993) Analysis of the Rhizobium meliloti genes exoU, exoV, exoW, exoT, and exoI involved in exopolysaccharide biosynthesis and nodule invasion: exoU and exoW probably encode glucosyltransferases. *Mol Plant Microbe Interact*, **6**, 735-744.
85. Ruiz, N. (2015) Lipid Flippases for Bacterial Peptidoglycan Biosynthesis. *Lipid Insights*, **8**, 21-31.
86. Ruiz, N. (2008) Bioinformatics identification of MurJ (MviN) as the peptidoglycan lipid II flippase in Escherichia coli. *Proc Natl Acad Sci U S A*, **105**, 15553-15557.
87. Heydanek, M.G., Jr. and Neuhaus, F.C. (1969) The initial stage in peptidoglycan synthesis. IV. Solubilization of phospho-N-acetylmuramyl-pentapeptide translocase. *Biochemistry*, **8**, 1474-1481.
88. Mengin-Lecreulx, D. and van Heijenoort, J. (1993) Identification of the glmU gene encoding N-acetylglucosamine-1-phosphate uridyltransferase in Escherichia coli. *J Bacteriol*, **175**, 6150-6157.
89. Mengin-Lecreulx, D. and van Heijenoort, J. (1996) Characterization of the essential gene glmM encoding phosphoglucosamine mutase in Escherichia coli. *J Biol Chem*, **271**, 32-39.

90. van Heijenoort, J. (2007) Lipid intermediates in the biosynthesis of bacterial peptidoglycan. *Microbiol Mol Biol Rev*, **71**, 620-635.
91. Ashraf, K.U., Nygaard, R., Vickery, O.N., Erramilli, S.K., Herrera, C.M., McConville, T.H., Petrou, V.I., Giacometti, S.I., Dufrisne, M.B., Nosol, K., Zinkle, A.P., Graham, C.L.B., Loukeris, M., Kloss, B., Skorupinska-Tudek, K., Swiezewska, E., Roper, D.I., Clarke, O.B., Uhlemann, A.C., Kossiakoff, A.A., Trent, M.S., Stansfeld, P.J. and Mancia, F. (2022) Structural basis of lipopolysaccharide maturation by the O-antigen ligase. *Nature*, **604**, 371-376.
92. Franco, A.V., Liu, D. and Reeves, P.R. (1998) The wzz (cld) protein in Escherichia coli: amino acid sequence variation determines O-antigen chain length specificity. *J Bacteriol*, **180**, 2670-2675.
93. Rai, A.K., Carr, J.F., Bautista, D.E., Wang, W. and Mitchell, A.M. (2021) ElyC and Cyclic Enterobacterial Common Antigen Regulate Synthesis of Phosphoglyceride-Linked Enterobacterial Common Antigen. *mBio*, **12**, e0284621.
94. Kumar, A.S., Mody, K. and Jha, B. (2007) Bacterial exopolysaccharides--a perception. *J Basic Microbiol*, **47**, 103-117.
95. van Heijenoort, J. (2001) Formation of the glycan chains in the synthesis of bacterial peptidoglycan. *Glycobiology*, **11**, 25R-36R.
96. Hong, Y., Hu, D., Verderosa, A.D., Qin, J., Totsika, M. and Reeves, P.R. (2023) Repeat-Unit Elongations To Produce Bacterial Complex Long Polysaccharide Chains, an O-Antigen Perspective. *EcoSal Plus*, **11**, eesp00202022.
97. Rai, A.K. and Mitchell, A.M. (2020) Enterobacterial Common Antigen: Synthesis and Function of an Enigmatic Molecule. *mBio*, **11**.
98. Sham, L.T., Butler, E.K., Lebar, M.D., Kahne, D., Bernhardt, T.G. and Ruiz, N. (2014) Bacterial cell wall. MurJ is the flippase of lipid-linked precursors for peptidoglycan biogenesis. *Science*, **345**, 220-222.
99. Kim, S., Pires, M.M. and Im, W. (2018) Insight into Elongation Stages of Peptidoglycan Processing in Bacterial Cytoplasmic Membranes. *Sci Rep*, **8**, 17704.
100. Di Guilmi, A.M., Dessen, A., Dideberg, O. and Vernet, T. (2003) The glycosyltransferase domain of penicillin-binding protein 2a from Streptococcus pneumoniae catalyzes the polymerization of murein glycan chains. *J Bacteriol*, **185**, 4418-4423.
101. Weckener, M., Woodward, L.S., Clarke, B.R., Liu, H., Ward, P.N., Le Bas, A., Bhella, D., Whitfield, C. and Naismith, J.H. (2023) The lipid linked oligosaccharide polymerase Wzy and its regulating co-polymerase, Wzz, from enterobacterial common antigen biosynthesis form a complex. *Open Biol*, **13**, 220373.
102. Han, W., Wu, B., Li, L., Zhao, G., Woodward, R., Pettit, N., Cai, L., Thon, V. and Wang, P.G. (2012) Defining function of lipopolysaccharide O-antigen ligase WaaL using chemoenzymatically synthesized substrates. *J Biol Chem*, **287**, 5357-5365.
103. Imperiali, B. (2019) Bacterial carbohydrate diversity - a Brave New World. *Curr Opin Chem Biol*, **53**, 1-8.

104. Mostowy, R.J. and Holt, K.E. (2018) Diversity-Generating Machines: Genetics of Bacterial Sugar-Coating. *Trends Microbiol*, **26**, 1008-1021.
105. Saiki, K. and Konishi, K. (2007) Identification of a Porphyromonas gingivalis novel protein sov required for the secretion of gingipains. *Microbiol Immunol*, **51**, 483-491.
106. Pares, S., Cohen-Addad, C., Sieker, L., Neuburger, M. and Douce, R. (1994) X-ray structure determination at 2.6-A resolution of a lipoate-containing protein: the H-protein of the glycine decarboxylase complex from pea leaves. *Proc Natl Acad Sci U S A*, **91**, 4850-4853.
107. Shultis, D.D., Purdy, M.D., Banchs, C.N. and Wiener, M.C. (2006) Outer membrane active transport: structure of the BtuB:TonB complex. *Science*, **312**, 1396-1399.
108. Rahn, A., Beis, K., Naismith, J.H. and Whitfield, C. (2003) A novel outer membrane protein, Wzi, is involved in surface assembly of the Escherichia coli K30 group 1 capsule. *J Bacteriol*, **185**, 5882-5890.
109. Qi, Y. and Grishin, N.V. (2005) Structural classification of thioredoxin-like fold proteins. *Proteins*, **58**, 376-388.
110. Suplatov, D.A., Besenmatter, W., Svedas, V.K. and Svendsen, A. (2012) Bioinformatic analysis of alpha/beta-hydrolase fold enzymes reveals subfamily-specific positions responsible for discrimination of amidase and lipase activities. *Protein Eng Des Sel*, **25**, 689-697.
111. Palomeque-Messia, P., Englebert, S., Leyh-Bouille, M., Nguyen-Disteche, M., Duez, C., Houba, S., Dideberg, O., Van Beeumen, J. and Ghuysen, J.M. (1991) Amino acid sequence of the penicillin-binding protein/DD-peptidase of Streptomyces K15. Predicted secondary structures of the low Mr penicillin-binding proteins of class A. *Biochem J*, **279 (Pt 1)**, 223-230.
112. Aravind, L. (1999) An evolutionary classification of the metallo-beta-lactamase fold proteins. *In Silico Biol*, **1**, 69-91.
113. Novinec, M. and Lenarcic, B. (2013) Papain-like peptidases: structure, function, and evolution. *Biomol Concepts*, **4**, 287-308.
114. Razew, A., Schwarz, J.N., Mitkowski, P., Sabala, I. and Kaus-Drobek, M. (2022) One fold, many functions-M23 family of peptidoglycan hydrolases. *Front Microbiol*, **13**, 1036964.
115. Mans, B.J., Anantharaman, V., Aravind, L. and Koonin, E.V. (2004) Comparative genomics, evolution and origins of the nuclear envelope and nuclear pore complex. *Cell Cycle*, **3**, 1612-1637.
116. Wimley, W.C. (2003) The versatile beta-barrel membrane protein. *Curr Opin Struct Biol*, **13**, 404-411.
117. Fairman, J.W., Noinaj, N. and Buchanan, S.K. (2011) The structural biology of beta-barrel membrane proteins: a summary of recent reports. *Curr Opin Struct Biol*, **21**, 523-531.
118. Tokuda, H. and Matsuyama, S. (2004) Sorting of lipoproteins to the outer membrane in E. coli. *Biochim Biophys Acta*, **1693**, 5-13.
119. Sanchez-Pulido, L., Devos, D., Genevrois, S., Vicente, M. and Valencia, A. (2003) POTRA: a conserved domain in the FtsQ family and a class of beta-barrel outer membrane proteins. *Trends Biochem Sci*, **28**, 523-526.

120. Kim, S., Malinverni, J.C., Sliz, P., Silhavy, T.J., Harrison, S.C. and Kahne, D. (2007) Structure and function of an essential component of the outer membrane protein assembly machine. *Science*, **317**, 961-964.
121. Oke, M., Sarra, R., Ghirlando, R., Farnaud, S., Gorringe, A.R., Evans, R.W. and Buchanan, S.K. (2004) The plug domain of a neisserial TonB-dependent transporter retains structural integrity in the absence of its transmembrane beta-barrel. *FEBS Lett*, **564**, 294-300.
122. Carr, S., Penfold, C.N., Bamford, V., James, R. and Hemmings, A.M. (2000) The structure of TolB, an essential component of the tol-dependent translocation system, and its protein-protein interaction with the translocation domain of colicin E9. *Structure*, **8**, 57-66.
123. Ghosh, M., Tucker, D.E., Burchett, S.A. and Leslie, C.C. (2006) Properties of the Group IV 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.
140. Hofmann, K. (2000) A superfamily of membrane-bound O-acyltransferases with implications for wnt signaling. *Trends Biochem Sci*, **25**, 111-112.
141. Rahlwes, K.C., Ha, S.A., Motooka, D., Mayfield, J.A., Baumoel, L.R., Strickland, J.N., Torres-Ocampo, A.P., Nakamura, S. and Morita, Y.S. (2017) The cell envelope-associated phospholipid-binding protein LmeA is required for mannan polymerization in mycobacteria. *J Biol Chem*, **292**, 17407-17417.
142. Yeow, J. and Chng, S.S. (2022) Of zones, bridges and chaperones - phospholipid transport in bacterial outer membrane assembly and homeostasis. *Microbiology (Reading)*, **168**.
143. Kadirvelraj, R., Foley, B.L., Dyekjaer, J.D. and Woods, R.J. (2008) Involvement of water in carbohydrate-protein binding: concanavalin A revisited. *J Am Chem Soc*, **130**, 16933-16942.
144. Flint, J., Nurizzo, D., Harding, S.E., Longman, E., Davies, G.J., Gilbert, H.J. and Bolam, D.N. (2004) Ligand-mediated dimerization of a carbohydrate-binding molecule reveals a novel mechanism for protein-carbohydrate recognition. *J Mol Biol*, **337**, 417-426.
145. Murzin, A.G. (1993) OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *EMBO J*, **12**, 861-867.
146. Williams, A.F. and Barclay, A.N. (1988) The immunoglobulin superfamily--domains for cell surface recognition. *Annu Rev Immunol*, **6**, 381-405.
147. Chen, C.K., Chan, N.L. and Wang, A.H. (2011) The many blades of the beta-propeller proteins: conserved but versatile. *Trends Biochem Sci*, **36**, 553-561.
148. Babu, M.M., Priya, M.L., Selvan, A.T., Madera, M., Gough, J., Aravind, L. and Sankaran, K. (2006) A database of bacterial lipoproteins (DOLOP) with functional assignments to predicted lipoproteins. *J Bacteriol*, **188**, 2761-2773.
149. Sankaran, K. and Wu, H.C. (1994) Lipid modification of bacterial prolipoprotein. Transfer of diacylglyceryl moiety from phosphatidylglycerol. *J Biol Chem*, **269**, 19701-19706.

150. Tjalsma, H., Zanen, G., Venema, G., Bron, S. and van Dijl, J.M. (1999) The potential active site of the lipoprotein-specific (type II) signal peptidase of Bacillus subtilis. *J Biol Chem*, **274**, 28191-28197.
151. Notenboom, V., Boraston, A.B., Kilburn, D.G. and Rose, D.R. (2001) Crystal structures of the family 9 carbohydrate-binding module from Thermotoga maritima xylanase 10A in native and ligand-bound forms. *Biochemistry*, **40**, 6248-6256.
152. Rigden, D.J. (2005) Analysis of glycoside hydrolase family 98: catalytic machinery, mechanism and a novel putative carbohydrate binding module. *FEBS Lett*, **579**, 5466-5472.
153. Boraston, A.B., Bolam, D.N., Gilbert, H.J. and Davies, G.J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J*, **382**, 769-781.
154. Iyer, L.M., Koonin, E.V. and Aravind, L. (2001) Adaptations of the helix-grip fold for ligand binding and catalysis in the START domain superfamily. *Proteins*, **43**, 134-144.
155. Santelli, E., Liddington, R.C., Mohan, M.A., Hoch, J.A. and Szurmant, H. (2007) The crystal structure of Bacillus subtilis YycI reveals a common fold for two members of an unusual class of sensor histidine kinase regulatory proteins. *J Bacteriol*, **189**, 3290-3295.
156. Das, K., Xiao, R., Wahlberg, E., Hsu, F., Arrowsmith, C.H., Montelione, G.T. and Arnold, E. (2001) X-ray crystal structure of MTH938 from Methanobacterium thermoautotrophicum at 2.2 A resolution reveals a novel tertiary protein fold. *Proteins*, **45**, 486-488.
157. Bennett, H.J., Davenport, J.B., Collins, R.F., Trafford, A.W., Pinali, C. and Kitmitto, A. (2013) Human junctophilin-2 undergoes a structural rearrangement upon binding PtdIns(3,4,5)P3 and the S101R mutation identified in hypertrophic cardiomyopathy obviates this response. *Biochem J*, **456**, 205-217.
158. Cortajarena, A.L. and Regan, L. (2006) Ligand binding by TPR domains. *Protein Sci*, **15**, 1193-1198.
159. Seed, K.D., Faruque, S.M., Mekalanos, J.J., Calderwood, S.B., Qadri, F. and Camilli, A. (2012) Phase variable O antigen biosynthetic genes control expression of the major protective antigen and bacteriophage receptor in Vibrio cholerae O1. *PLoS Pathog*, **8**, e1002917.
160. Cai, R., Wang, G., Le, S., Wu, M., Cheng, M., Guo, Z., Ji, Y., Xi, H., Zhao, C., Wang, X., Xue, Y., Wang, Z., Zhang, H., Fu, Y., Sun, C., Feng, X., Lei, L., Yang, Y., Ur Rahman, S., Liu, X., Han, W. and Gu, J. (2019) Three Capsular Polysaccharide Synthesis-Related Glucosyltransferases, GT-1, GT-2 and WcaJ, Are Associated With Virulence and Phage Sensitivity of Klebsiella pneumoniae. *Front Microbiol*, **10**, 1189.
161. Yau, K.W. (1994) Cyclic nucleotide-gated channels: an expanding new family of ion channels. *Proc Natl Acad Sci U S A*, **91**, 3481-3483.
162. Durocher, D., Henckel, J., Fersht, A.R. and Jackson, S.P. (1999) The FHA domain is a modular phosphopeptide recognition motif. *Mol Cell*, **4**, 387-394.
163. Schreiter, E.R. and Drennan, C.L. (2007) Ribbon-helix-helix transcription factors: variations on a theme. *Nat Rev Microbiol*, **5**, 710-720.

164. Ravi, J., Anantharaman, V., Chen, S.Z., Brenner, E.P., Datta, P., Aravind, L. and Gennaro, M.L. (2024) The phage shock protein (PSP) envelope stress response: discovery of novel partners and evolutionary history. *mSystems*, **9**, e0084723.
165. AhYoung, A.P., Koehl, A., Vizcarra, C.L., Cascio, D. and Egea, P.F. (2016) Structure of a putative ClpS N-end rule adaptor protein from the malaria pathogen Plasmodium falciparum. *Protein Sci*, **25**, 689-701.
166. Durr, I.F. and Rudney, H. (1960) The reduction of beta-hydroxy-beta-methyl-glutaryl coenzyme A to mevalonic acid. *J Biol Chem*, **235**, 2572-2578.
167. Tchen, T.T. (1958) Mevalonic kinase: purification and properties. *J Biol Chem*, **233**, 1100-1103.
168. Wolff, M., Seemann, M., Tse Sum Bui, B., Frapart, Y., Tritsch, D., Garcia Estrabot, A., Rodriguez-Concepcion, M., Boronat, A., Marquet, A. and Rohmer, M. (2003) Isoprenoid biosynthesis via the methylerythritol phosphate pathway: the (E)-4-hydroxy-3-methylbut-2-enyl diphosphate reductase (LytB/IspH) from Escherichia coli is a [4Fe-4S] protein. *FEBS Lett*, **541**, 115-120.
169. Guo, R.T., Ko, T.P., Chen, A.P., Kuo, C.J., Wang, A.H. and Liang, P.H. (2005) Crystal structures of undecaprenyl pyrophosphate synthase in complex with magnesium, isopentenyl pyrophosphate, and farnesyl thiopyrophosphate: roles of the metal ion and conserved residues in catalysis. *J Biol Chem*, **280**, 20762-20774.
170. Burroughs, A.M. and Aravind, L. (2016) RNA damage in biological conflicts and the diversity of responding RNA repair systems. *Nucleic Acids Res*, **44**, 8525-8555.
171. Ueta, M., Ohniwa, R.L., Yoshida, H., Maki, Y., Wada, C. and Wada, A. (2008) Role of HPF (hibernation promoting factor) in translational activity in Escherichia coli. *J Biochem*, **143**, 425-433.
172. Akanuma, G., Kazo, Y., Tagami, K., Hiraoka, H., Yano, K., Suzuki, S., Hanai, R., Nanamiya, H., Kato-Yamada, Y. and Kawamura, F. (2016) Ribosome dimerization is essential for the efficient regrowth of Bacillus subtilis. *Microbiology (Reading)*, **162**, 448-458.
173. Kuk, A.C.Y., Hao, A. and Lee, S.Y. (2022) Structure and Mechanism of the Lipid Flippase MurJ. *Annu Rev Biochem*, **91**, 705-729.
174. Zhang, D., de Souza, R.F., Anantharaman, V., Iyer, L.M. and Aravind, L. (2012) Polymorphic toxin systems: Comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. *Biol Direct*, **7**, 18.
175. Iyer, L.M., Zhang, D., Rogozin, I.B. and Aravind, L. (2011) Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. *Nucleic Acids Res*, **39**, 9473-9497.
176. Ruhe, Z.C., Low, D.A. and Hayes, C.S. (2020) Polymorphic Toxins and Their Immunity Proteins: Diversity, Evolution, and Mechanisms of Delivery. *Annu Rev Microbiol*, **74**, 497-520.
177. Alcoforado Diniz, J. and Coulthurst, S.J. (2015) Intraspecies Competition in Serratia marcescens Is Mediated by Type VI-Secreted Rhs Effectors and a Conserved Effector-Associated Accessory Protein. *J Bacteriol*, **197**, 2350-2360.
178. Satchell, K.J. (2011) Structure and function of MARTX toxins and other large repetitive RTX proteins. *Annu Rev Microbiol*, **65**, 71-90.

179. Kordis, D. and Turk, V. (2009) Phylogenomic analysis of the cystatin superfamily in eukaryotes and prokaryotes. *BMC Evol Biol*, **9**, 266.
180. Cherier, D., Patin, D., Blanot, D., Touze, T. and Barreteau, H. (2021) The Biology of Colicin M and Its Orthologs. *Antibiotics (Basel)*, **10**.
181. Anantharaman, V. and Aravind, L. (2003) Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes. *Genome Biol*, **4**, R11.
182. Arthur, M., Molinas, C. and Courvalin, P. (1992) Sequence of the vanY gene required for production of a vancomycin-inducible D,D-carboxypeptidase in Enterococcus faecium BM4147. *Gene*, **120**, 111-114.
183. Turk, V. and Bode, W. (1991) The cystatins: protein inhibitors of cysteine proteinases. *FEBS Lett*, **285**, 213-219.
184. Sato, Y., Yamamoto, Y. and Kizaki, H. (1997) Cloning and sequence analysis of the gbpC gene encoding a novel glucan-binding protein of Streptococcus mutans. *Infect Immun*, **65**, 668-675.
185. MacKenzie, D.A., Tailford, L.E., Hemmings, A.M. and Juge, N. (2009) Crystal structure of a mucus-binding protein repeat reveals an unexpected functional immunoglobulin binding activity. *J Biol Chem*, **284**, 32444-32453.
186. Chagnot, C., Zorgani, M.A., Astruc, T. and Desvaux, M. (2013) Proteinaceous determinants of surface colonization in bacteria: bacterial adhesion and biofilm formation from a protein secretion perspective. *Front Microbiol*, **4**, 303.
187. Wu, H.Y., Liu, M.S., Lin, T.P. and Cheng, Y.S. (2011) Structural and functional assays of AtTLP18.3 identify its novel acid phosphatase activity in thylakoid lumen. *Plant Physiol*, **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.
202. Dideberg, O., Charlier, P., Dive, G., Joris, B., Frere, J.M. and Ghuysen, J.M. (1982) Structure of a Zn2+-containing D-alanyl-D-alanine-cleaving carboxypeptidase at 2.5 A resolution. *Nature*, **299**, 469-470.
203. Bellinzoni, M., Haouz, A., Miras, I., Magnet, S., Andre-Leroux, G., Mukherjee, R., Shepard, W., Cole, S.T. and Alzari, P.M. (2014) Structural studies suggest a peptidoglycan hydrolase function for the Mycobacterium tuberculosis Tat-secreted protein Rv2525c. *J Struct Biol*, **188**, 156-164.
204. Wright, G.D., Molinas, C., Arthur, M., Courvalin, P. and Walsh, C.T. (1992) Characterization of vanY, a DD-carboxypeptidase from vancomycin-resistant Enterococcus faecium BM4147. *Antimicrob Agents Chemother*, **36**, 1514-1518.
205. Arthur, M., Depardieu, F., Cabanie, L., Reynolds, P. and Courvalin, P. (1998) Requirement of the VanY and VanX D,D-peptidases for glycopeptide resistance in enterococci. *Mol Microbiol*, **30**, 819-830.
206. Arthur, M., Depardieu, F., Snaith, H.A., Reynolds, P.E. and Courvalin, P. (1994) Contribution of VanY D,D-carboxypeptidase to glycopeptide resistance in Enterococcus faecalis by hydrolysis of peptidoglycan precursors. *Antimicrob Agents Chemother*, **38**, 1899-1903.

207. Meyer, B., Wurm, J.P., Sharma, S., Immer, C., Pogoryelov, D., Kotter, P., Lafontaine, D.L., Wohnert, J. and Entian, K.D. (2016) Ribosome biogenesis factor Tsr3 is the aminocarboxypropyl transferase responsible for 18S rRNA hypermodification in yeast and humans. *Nucleic Acids Res*, **44**, 4304-4316.
208. Burroughs, A.M. and Aravind, L. (2014) Analysis of two domains with novel RNA-processing activities throws light on the complex evolution of ribosomal RNA biogenesis. *Front Genet*, **5**, 424.
209. Fraser, M.E., Joyce, M.A., Ryan, D.G. and Wolodko, W.T. (2002) Two glutamate residues, Glu 208 alpha and Glu 197 beta, are crucial for phosphorylation and dephosphorylation of the active-site histidine residue in succinyl-CoA synthetase. *Biochemistry*, **41**, 537-546.
210. Iyer, L.M., Abhiman, S., Maxwell Burroughs, A. and Aravind, L. (2009) Amidoligases with ATP-grasp, glutamine synthetase-like and acetyltransferase-like domains: synthesis of novel metabolites and peptide modifications of proteins. *Mol Biosyst*, **5**, 1636-1660.
211. Raetz, C.R. and Kennedy, E.P. (1974) Partial purification and properties of phosphatidylserine synthetase from Escherichia coli. *J Biol Chem*, **249**, 5083-5045.
212. Fields, R.N. and Roy, H. (2018) Deciphering the tRNA-dependent lipid aminoacylation systems in bacteria: Novel components and structural advances. *RNA Biol*, **15**, 480-491.
213. Hancock, R.E. (1997) Peptide antibiotics. *Lancet*, **349**, 418-422.
214. Benson, T.E., Prince, D.B., Mutchler, V.T., Curry, K.A., Ho, A.M., Sarver, R.W., Hagadorn, J.C., Choi, G.H. and Garlick, R.L. (2002) X-ray crystal structure of Staphylococcus aureus FemA. *Structure*, **10**, 1107-1115.
215. Schneider, T., Senn, M.M., Berger-Bachi, B., Tossi, A., Sahl, H.G. and Wiedemann, I. (2004) In vitro assembly of a complete, pentaglycine interpeptide bridge containing cell wall precursor (lipid II-Gly5) of Staphylococcus aureus. *Mol Microbiol*, **53**, 675-685.
216. Schmid, M., Uhlenhaut, N.H., Godard, F., Demar, M., Bressan, R., Weigel, D. and Lohmann, J.U. (2003) Dissection of floral induction pathways using global expression analysis. *Development*, **130**, 6001-6012.
217. Berardini, T.Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E. and Huala, E. (2015) The Arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. *Genesis*, **53**, 474-485.
218. Savidov, N., Gloriozova, T.A., Poroikov, V.V. and Dembitsky, V.M. (2018) Highly oxygenated isoprenoid lipids derived from fungi and fungal endophytes: Origin and biological activities. *Steroids*, **140**, 114-124.
219. Lee, J.O., Rieu, P., Arnaout, M.A. and Liddington, R. (1995) Crystal structure of the A domain from the alpha subunit of integrin CR3 (CD11b/CD18). *Cell*, **80**, 631-638.
220. Zhang, Y., van der Fits, L., Voerman, J.S., Melief, M.J., Laman, J.D., Wang, M., Wang, H., Wang, M., Li, X., Walls, C.D., Gupta, D. and Dziarski, R. (2005) Identification of serum N-acetylmuramoyl-l-alanine amidase as liver peptidoglycan recognition protein 2. *Biochim Biophys Acta*, **1752**, 34-46.

221. Lee, J., Geddes, K., Streutker, C., Philpott, D.J. and Girardin, S.E. (2012) Role of mouse peptidoglycan recognition protein PGLYRP2 in the innate immune response to Salmonella enterica serovar Typhimurium infection in vivo. *Infect Immun*, **80**, 2645-2654.
222. Dziarski, R. and Gupta, D. (2010) Review: Mammalian peptidoglycan recognition proteins (PGRPs) in innate immunity. *Innate Immun*, **16**, 168-174.
223. Dziarski, R. and Gupta, D. (2006) Mammalian PGRPs: novel antibacterial proteins. *Cell Microbiol*, **8**, 1059-1069.
224. Jones, J.D. and Dangl, J.L. (2006) The plant immune system. *Nature*, **444**, 323-329.
225. Holzem, M., Boutros, M. and Holstein, T.W. (2024) The origin and evolution of Wnt signalling. *Nat Rev Genet*, **25**, 500-512.
226. Morrow, J.F., Stearman, R.S., Peltzman, C.G. and Potter, D.A. (1981) Induction of hepatic synthesis of serum amyloid A protein and actin. *Proc Natl Acad Sci U S A*, **78**, 4718-4722.
227. Hari-Dass, R., Shah, C., Meyer, D.J. and Raynes, J.G. (2005) Serum amyloid A protein binds to outer membrane protein A of gram-negative bacteria. *J Biol Chem*, **280**, 18562-18567.
228. Shah, C., Hari-Dass, R. and Raynes, J.G. (2006) Serum amyloid A is an innate immune opsonin for Gram-negative bacteria. *Blood*, **108**, 1751-1757.
229. Sack, G.H., Jr., Talbot, C.C., Jr., Seuanez, H. and O'Brien, S.J. (1989) Molecular analysis of the human serum amyloid A (SAA) gene family. *Scand J Immunol*, **29**, 113-119.
230. Uhlar, C.M., Burgess, C.J., Sharp, P.M. and Whitehead, A.S. (1994) Evolution of the serum amyloid A (SAA) protein superfamily. *Genomics*, **19**, 228-235.
231. Sun, L. and Ye, R.D. (2016) Serum amyloid A1: Structure, function and gene polymorphism. *Gene*, **583**, 48-57.
232. Aoki, S.K., Malinverni, J.C., Jacoby, K., Thomas, B., Pamma, R., Trinh, B.N., Remers, S., Webb, J., Braaten, B.A., Silhavy, T.J. and Low, D.A. (2008) Contact-dependent growth inhibition requires the essential outer membrane protein BamA (YaeT) as the receptor and the inner membrane transport protein AcrB. *Mol Microbiol*, **70**, 323-340.
233. Virtanen, P., Waneskog, M. and Koskiniemi, S. (2019) Class II contact-dependent growth inhibition (CDI) systems allow for broad-range cross-species toxin delivery within the Enterobacteriaceae family. *Mol Microbiol*, **111**, 1109-1125.
234. Ruhe, Z.C., Nguyen, J.Y., Xiong, J., Koskiniemi, S., Beck, C.M., Perkins, B.R., Low, D.A. and Hayes, C.S. (2017) CdiA Effectors Use Modular Receptor-Binding Domains To Recognize Target Bacteria. *mBio*, **8**.
235. Levin, M., Franklin, E.C., Frangione, B. and Pras, M. (1972) The amino acid sequence of a major nonimmunoglobulin component of some amyloid fibrils. *J Clin Invest*, **51**, 2773-2776.
236. Blanco, L.P., Evans, M.L., Smith, D.R., Badtke, M.P. and Chapman, M.R. (2012) Diversity, biogenesis and function of microbial amyloids. *Trends Microbiol*, **20**, 66-73.
237. Rouse, S.L., Matthews, S.J. and Dueholm, M.S. (2018) Ecology and Biogenesis of Functional Amyloids in Pseudomonas. *J Mol Biol*, **430**, 3685-3695.

238. Tukel, C., Wilson, R.P., Nishimori, J.H., Pezeshki, M., Chromy, B.A. and Baumler, A.J. (2009) Responses to amyloids of microbial and host origin are mediated through toll-like receptor 2. *Cell Host Microbe*, **6**, 45-53.
239. Prosswimmer, T., Heng, A. and Daggett, V. (2024) Mechanistic insights into the role of amyloid-beta in innate immunity. *Sci Rep*, **14**, 5376.
240. Egan, A.J.F., Errington, J. and Vollmer, W. (2020) Regulation of peptidoglycan synthesis and remodelling. *Nat Rev Microbiol*, **18**, 446-460.
241. Koga, Y. and Morii, H. (2007) Biosynthesis of ether-type polar lipids in archaea and evolutionary considerations. *Microbiol Mol Biol Rev*, **71**, 97-120.
242. Daiyasu, H., Kuma, K., Yokoi, T., Morii, H., Koga, Y. and Toh, H. (2005) A study of archaeal enzymes involved in polar lipid synthesis linking amino acid sequence information, genomic contexts and lipid composition. *Archaea*, **1**, 399-410.
243. Carman, G.M. and Han, G.S. (2006) Roles of phosphatidate phosphatase enzymes in lipid metabolism. *Trends Biochem Sci*, **31**, 694-699.
244. Neuwald, A.F. (1997) An unexpected structural relationship between integral membrane phosphatases and soluble haloperoxidases. *Protein Sci*, **6**, 1764-1767.
245. Aravind, L., Nicastro, G.G., Iyer, L.M. and Burroughs, A.M. (2024) The Prokaryotic Roots of Eukaryotic Immune Systems. *Annu Rev Genet*, **58**, 365-389.
246. Krishnan, A., Iyer, L.M., Holland, S.J., Boehm, T. and Aravind, L. (2018) Diversification of AID/APOBEC-like deaminases in metazoa: multiplicity of clades and widespread roles in immunity. *Proc Natl Acad Sci U S A*, **115**, E3201-E3210.
247. Zhang, D., Iyer, L.M. and Aravind, L. (2011) A novel immunity system for bacterial nucleic acid degrading toxins and its recruitment in various eukaryotic and DNA viral systems. *Nucleic Acids Res*, **39**, 4532-4552.
248. Kaur, G., Burroughs, A.M., Iyer, L.M. and Aravind, L. (2020) Highly regulated, diversifying NTP-dependent biological conflict systems with implications for the emergence of multicellularity. *Elife*, **9**.
249. Kaur, G., Iyer, L.M., Burroughs, A.M. and Aravind, L. (2021) Bacterial death and TRADD-N domains help define novel apoptosis and immunity mechanisms shared by prokaryotes and metazoans. *Elife*, **10**.
250. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, **25**, 3389-3402.
251. Johnson, L.S., Eddy, S.R. and Portugaly, E. (2010) Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics*, **11**, 431.
252. Sayers, E.W., Bolton, E.E., Brister, J.R., Canese, K., Chan, J., Comeau, D.C., Connor, R., Funk, 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

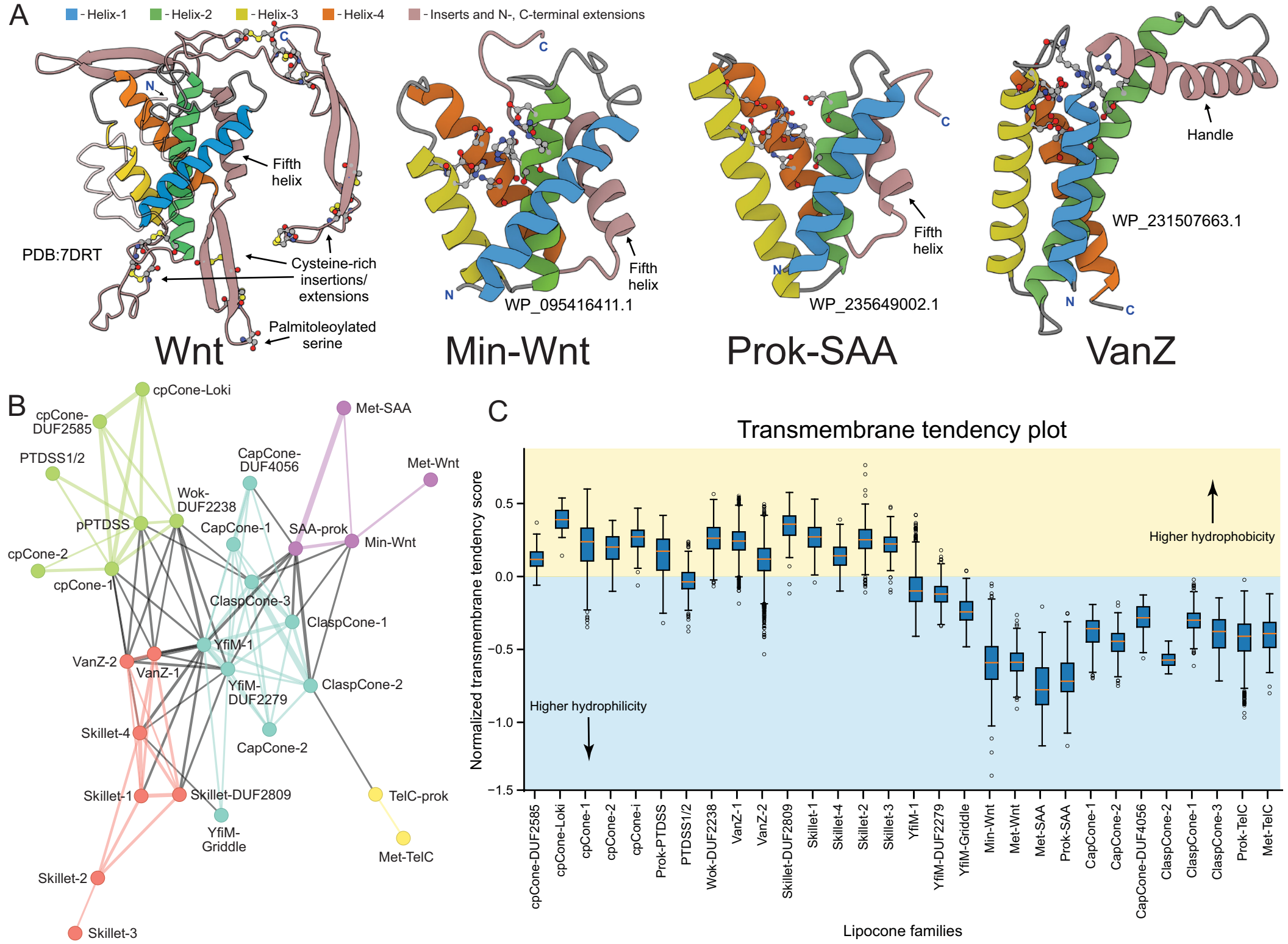
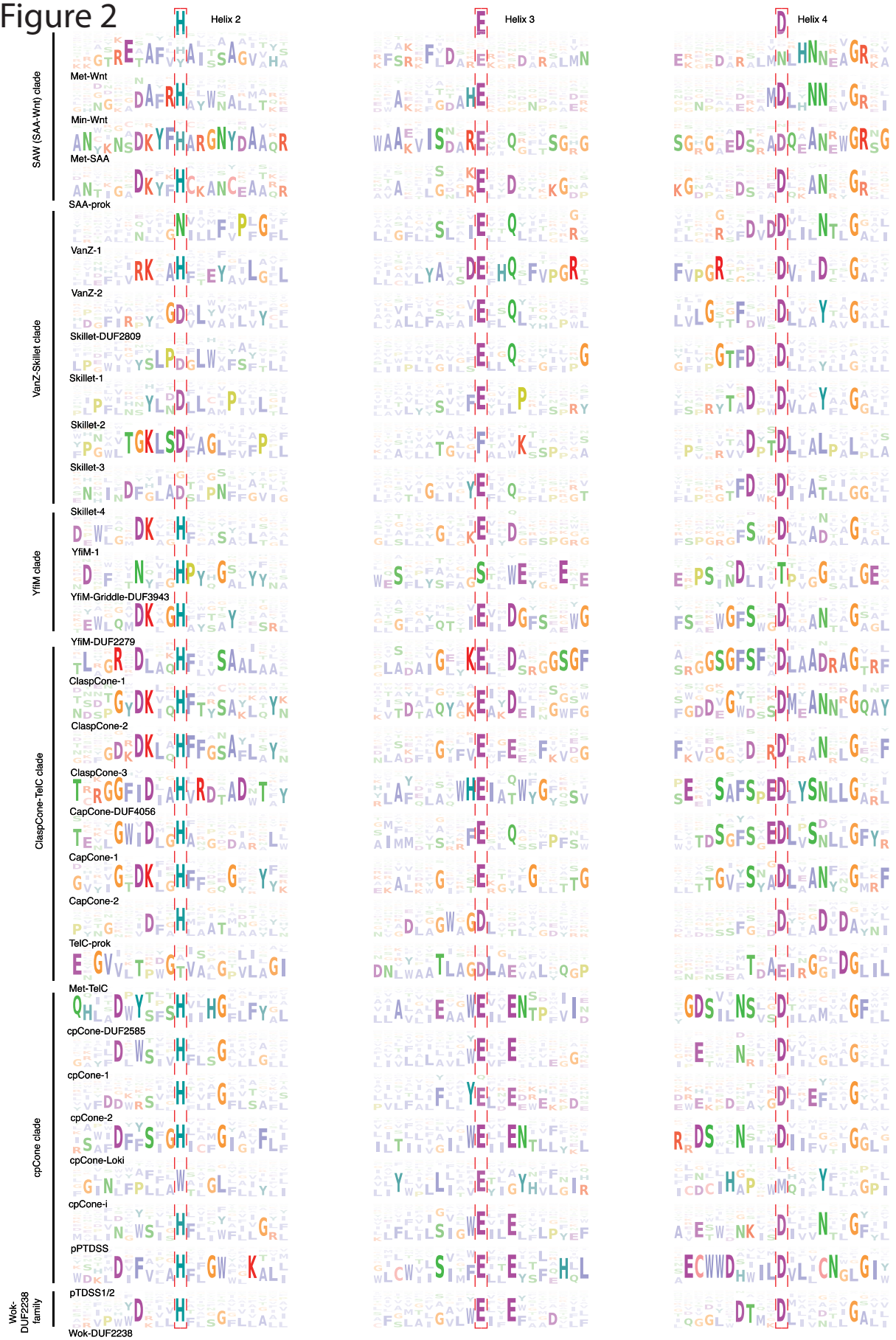

A
Helix-1
Helix-2
Helix-3
Helix-4
Inserts and N-, C-terminal extensions
PDB:7DRT
Fifth helix
Cysteine-rich insertions/ extensions
Palmitoleoylated serine
Wnt
WP_095416411.1
Fifth helix
Min-Wnt
WP_235649002.1
Fifth helix
Prok-SAA
Handle
WP_231507663.1
VanZ
B
cpCone-Loki
cpCone-DUF2585
PTDSS1/2
Met-SAA
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
Higher hydrophobicity
Higher hydrophilicity
Lipocone families
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

Figure 2


H
Helix 2
E
Helix 3
D
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

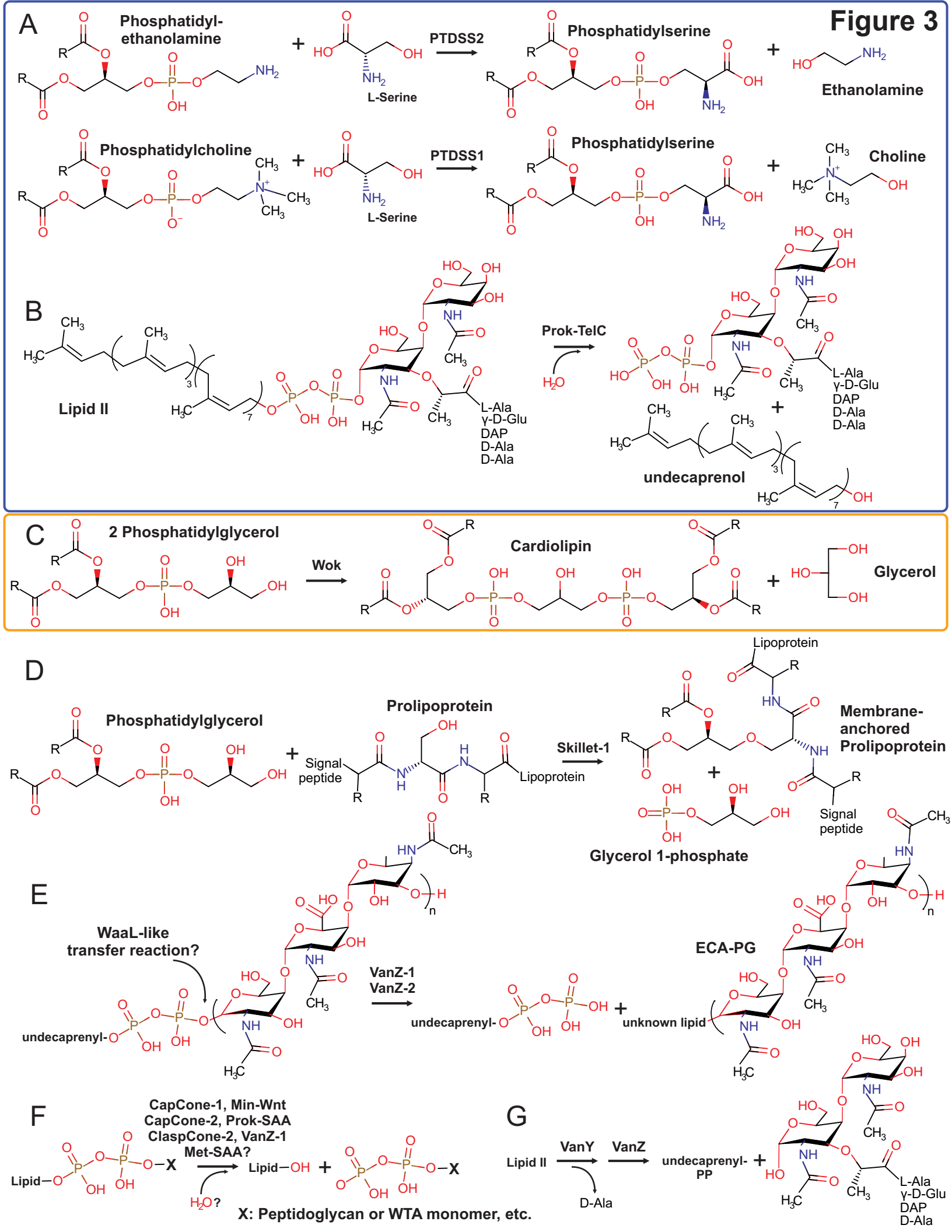
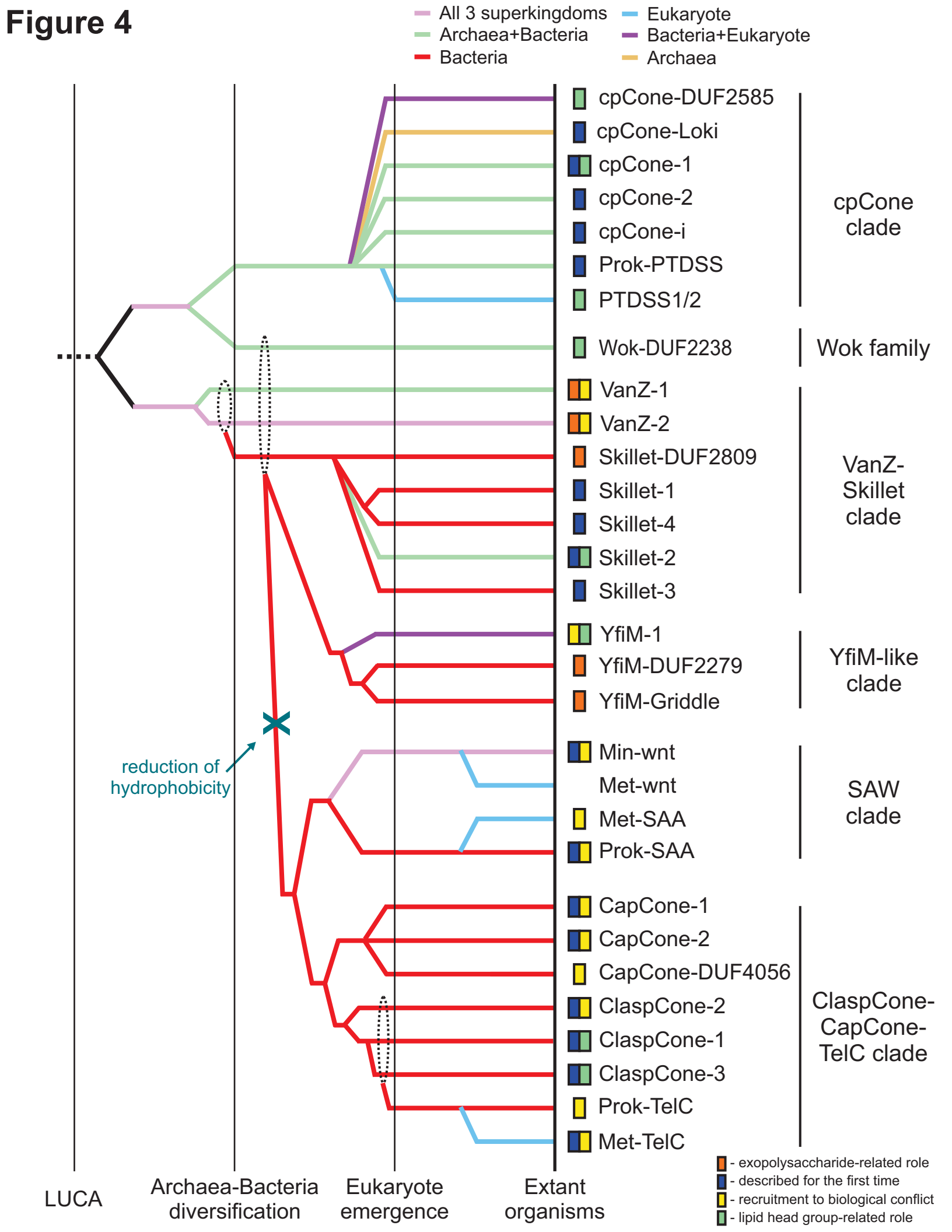


Figure 3
A
Phosphatidyl-ethanolamine
L-Serine
PTDSS2
Phosphatidylserine
Ethanolamine
Phosphatidylcholine
L-Serine
PTDSS1
Phosphatidylserine
Choline
B
Lipid II
Prok-TelC
L-Ala
γ-D-Glu
DAP
D-Ala
D-Ala
undecaprenol
C
2 Phosphatidylglycerol
Wok
Cardiolipin
Glycerol
D
Phosphatidylglycerol
Prolipoprotein
Signal peptide
Lipoprotein
Skillet-1
Membrane-anchored Prolipoprotein
Glycerol 1-phosphate
E
WaaL-like transfer reaction?
undecaprenyl-
VanZ-1
VanZ-2
undecaprenyl-
unknown lipid
ECA-PG
F
CapCone-1, Min-Wnt
CapCone-2, Prok-SAA
ClaspCone-2, VanZ-1
Met-SAA?
Lipid—
—X
Lipid—OH
X: Peptidoglycan or WTA monomer, etc.
G
Lipid II
VanY
VanZ
D-Ala
undecaprenyl-PP
L-Ala
γ-D-Glu
DAP
D-Ala

Figure 4


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
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
- exopolysaccharide-related role
- described for the first time
- recruitment to biological conflict
- lipid head group-related role

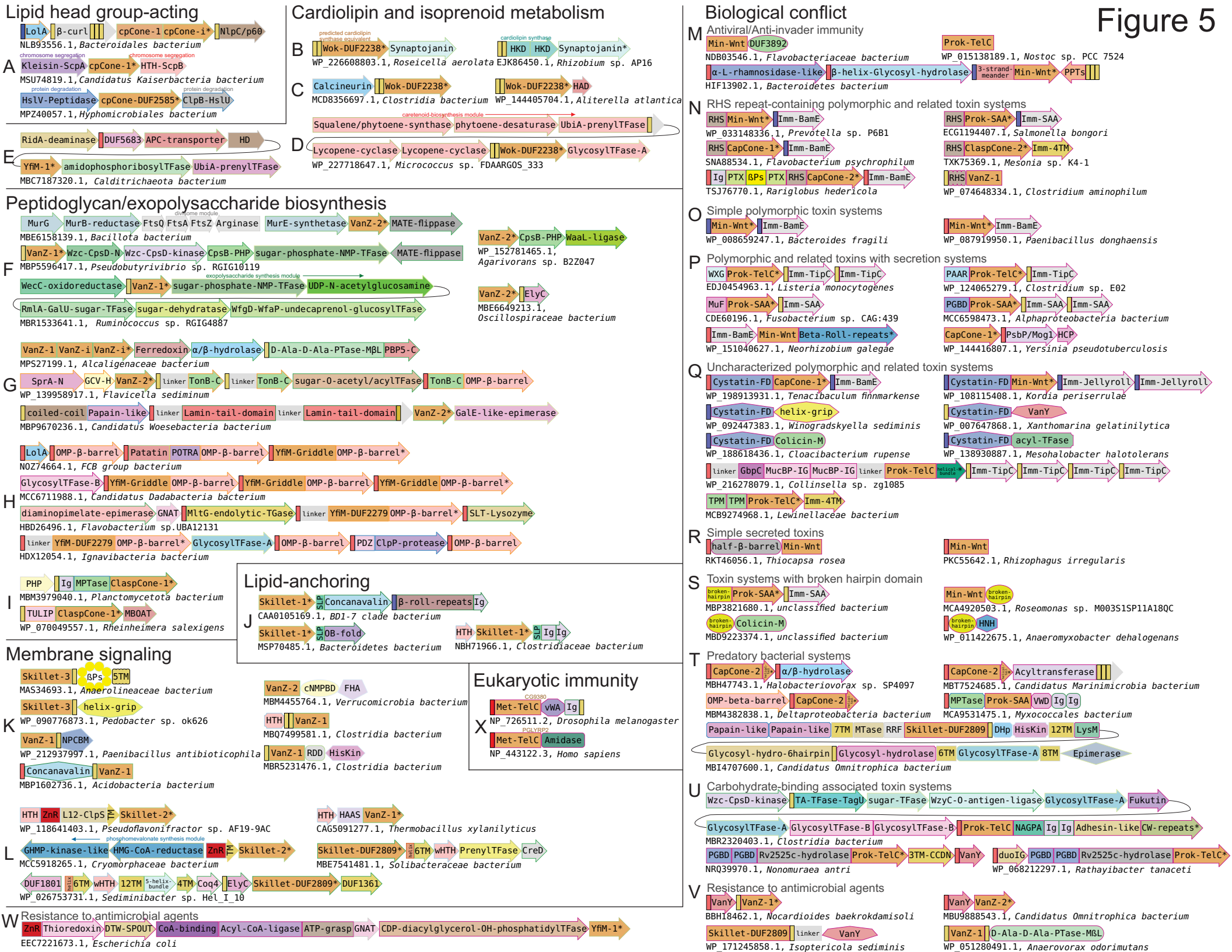


Figure 5
Lipid head group-acting
A
LolA β-curl cpCone-1 cpCone-i* NlpC/p60
NLB93556.1, Bacteroidales bacterium
chromosome segregation
Kleisin-ScpA cpCone-1* HTH-ScpB
MSU74819.1, Candidatus Kaiserbacteria bacterium
protein degradation
HslV-Peptidase cpCone-DUF2585* ClpB-HslU
MPZ40057.1, Hyphomicrobiales 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
D
caretenoid-biosynthesis module
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
VanZ-2* CpsB-PHP WaaL-ligase
WP_152781465.1, Agarivorans sp. B2Z047
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* 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
HTH HAAS VanZ-1*
CAG5091277.1, Thermobacillus xylanilyticus
phosphomevalonate synthesis module
GHMP-kinase-like HMG-CoA-reductase ZnR ⊠ Skillet-2*
MCC5918265.1, Cryomorphaceae bacterium
Skillet-DUF2809* HxHx 6TM wHTH PrenylTFase CreD
MBE7541481.1, Solibacteraceae bacterium
DUF1801 HxHx 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 α/β-hydrolase
MBH47743.1, Halobacteriovorax sp. SP4097
CapCone-2 Acyltransferase
MBT7524685.1, Candidatus Marinimicrobia bacterium
OMP-beta-barrel CapCone-2
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

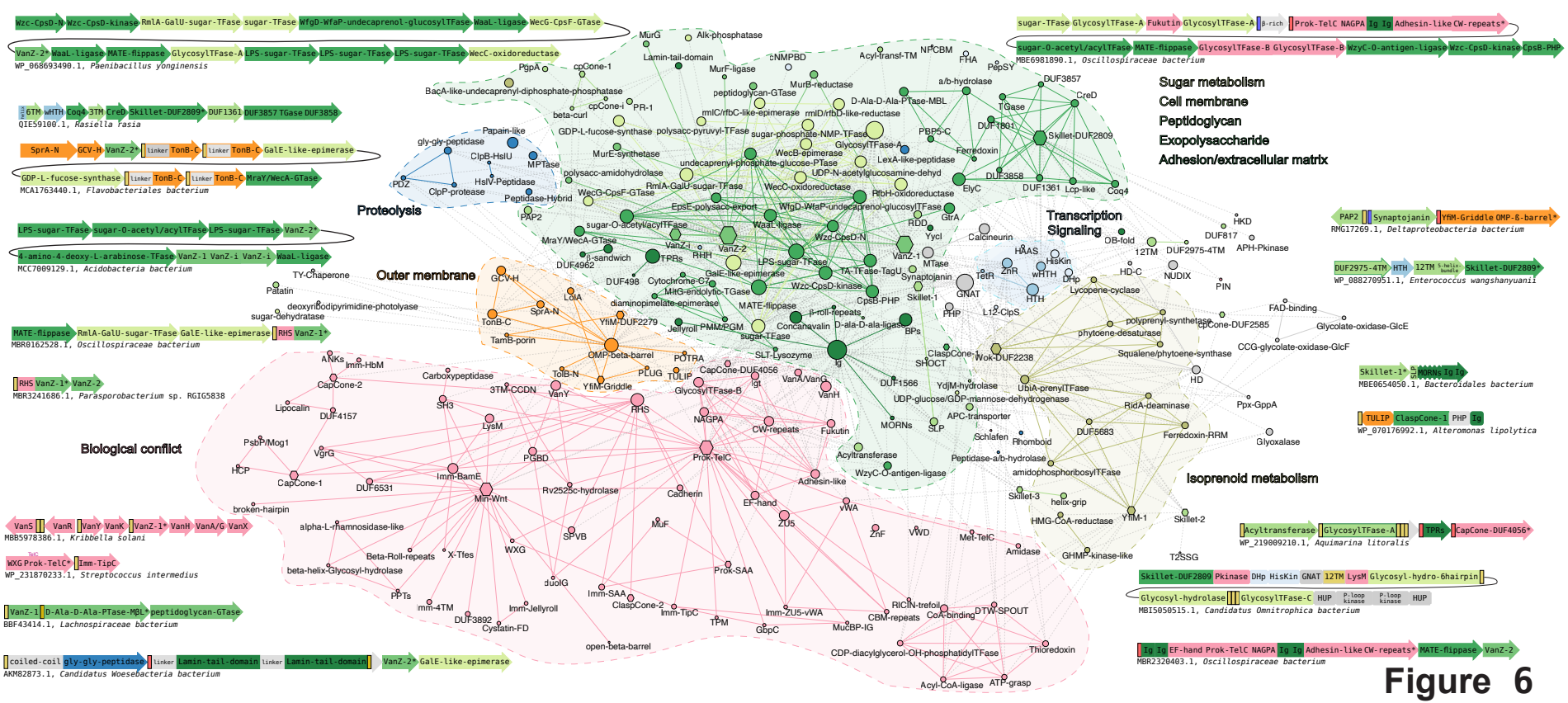

Sugar metabolism
Cell membrane
Peptidoglycan
Exopolysaccharide
Adhesion/extracellular matrix
Proteolysis
Transcription
Signaling
Outer membrane
Biological conflict
Isoprenoid metabolism
Wzc-CpsD-N Wzc-CpsD-kinase RmlA-GalU-sugar-TFase sugar-TFase WfgD-WfaP-undecaprenol-glucosyltFase WaaL-ligase WecG-CpsF-GTase
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 Adhesion-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 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 Adhesion-like CW-repeats* MATE-flippase VanZ-2
MBR2320403.1, Oscillospiraceae bacterium

Figure 6