PNAS RESEARCH ARTICLE | MICROBIOLOGY

Amine-recognizing domain in diverse receptors from bacteria and archaea evolved from the universal amino acid sensor

Jean Paul Cerna-Vargas[a,b,1], Vadim M. Gumerov[c,1], Tino Krell[a,2], and Igor B. Zhulin[c,2]

Edited by Eugene Koonin, NIH, Bethesda, MD; received April 11, 2023; accepted September 9, 2023

Bacteria possess various receptors that sense different signals and transmit information to enable an optimal adaptation to the environment. A major limitation in microbiology is the lack of information on the signal molecules that activate receptors. Signals recognized by sensor domains are poorly reflected in overall sequence identity, and therefore, the identification of signals from the amino acid sequence of the sensor alone presents a challenge. Biogenic amines are of great physiological importance for microorganisms and humans. They serve as substrates for aerobic and anaerobic growth and play a role of neurotransmitters and osmoprotectants. Here, we report the identification of a sequence motif that is specific for amine-sensing sensor domains that belong to the Cache superfamily of the most abundant extracellular sensors in prokaryotes. We identified approximately 13,000 sensor histidine kinases, chemoreceptors, receptors involved in second messenger homeostasis and Ser/Thr phosphatases from 8,000 bacterial and archaeal species that contain the amine-recognizing motif. The screening of compound libraries and microcalorimetric titrations of selected sensor domains confirmed their ability to specifically bind biogenic amines. Mutants in the amine-binding motif or domains that contain a single mismatch in the binding motif had either no or a largely reduced affinity for amines. We demonstrate that the amine-recognizing domain originated from the universal amino acid–sensing Cache domain, thus providing insight into receptor evolution. Our approach enables precise "wet"-lab experiments to define the function of regulatory systems and therefore holds a strong promise to enable the identification of signals stimulating numerous receptors.

bacterial signal transduction | evolution | biogenic amines | receptors

Significance

Cells possess numerous receptors that modulate signaling pathways. While hundreds of thousands of receptor sequences are available in databases, the signals that receptors recognize remain largely unknown. Here, we used a structure/sequence approach to identify thousands of microbial proteins with the predicted capability to detect biogenic amines. We selected several targets for experimental validation and demonstrated that all of them bind various amines. Amines are key nutrients and major neurotransmitters. Identification of amine sensors in bacteria inhabiting the human gut as well as in plant- and human pathogens opens research directions for those studying these important microbes. By revealing that amine-sensing receptors originated from amino acid–sensing receptors, we show how receptors can change their specificity during evolution.

Bacteria have evolved numerous receptors that sense environmental stimuli and modulate various signal transduction pathways to enable adaptation to changing conditions. Major receptor families include transcriptional regulators, sensor histidine kinases, chemoreceptors, cyclic (di)nucleotide cyclases and phosphodiesterases, serine/threonine protein kinases, and phosphatases (1, 2). The initial step of signal integration by these receptors involves ligand binding to sensor or ligand-binding domains (LBD). Hundreds of different sensor domains have evolved, although only a few of them are ubiquitous (1, 3), and the same type of a sensor domain is frequently found in different signal transduction systems (4). Signals that activate most of the signal transduction systems in bacteria and archaea are unknown, which presents a major bottleneck in microbiological research (5). Such knowledge is indispensable not only for understanding the physiological significance of regulatory circuits but also for the development of anti-infective therapies aimed at reducing bacterial virulence by interfering with signal transduction cascades. Revealing signals for thousands of unstudied receptors by extrapolation from a few well-characterized homologs is largely hampered by the fact that sensor domains are rapidly evolving thus displaying a large degree of sequence divergence (6).

We have recently reported the first study in which the type of signal molecules was successfully predicted and verified for a large sensor domain family. By combining sequence and structure information from a few known amino acid–sensing chemoreceptors of the ubiquitous dCache_1 domains (7), we derived the amino acid recognizing motif, which was used in database searches to identify thousands of motif-containing homologs, followed by experimental validation of selected targets (8). Subsequently, we defined a large subfamily of amino acid-sensing dCache_1 domains (termed dCache_1AA) containing more than ten thousand protein sequences from all major lineages of life (8). This iterative computational and experimental approach has an enormous potential to link many thousands of receptors to specific ligands, which is crucial for understanding the function of corresponding signal transduction circuits. dCache_1 domains are the predominant extracellular sensors found in all major receptor families in bacteria and archaea (7). In addition to binding amino acids, some dCache_1 domains bind organic acids (9, 10), sugars (11), quorum sensing signals (12), inorganic ions (13), purine derivatives (14), polyamines (15, 16), and quaternary amines (QA)

Author contributions: J.P.C.-V., V.M.G., T.K., and I.B.Z. designed research; J.P.C.-V., V.M.G., and T.K. performed research; J.P.C.-V., V.M.G., T.K., and I.B.Z. analyzed data; and J.P.C.-V., V.M.G., T.K., and I.B.Z. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]J.P.C.-V. and V.M.G. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: tino.krell@eez.csic.es or jouline.1@osu.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2305837120/-/DCSupplemental.

Published October 11, 2023.

(17, 18), suggesting that in addition to dCache_1AA, there might be other domain subfamilies with a well-defined ligand repertoire.

In this study, we focused on biogenic amines because of their important biological roles and the availability of three solved structures of amine-bound dCache_1 domains from bacterial receptors. Biogenic amines are products of amino acid metabolism and are characterized by a nitrogen atom that is covalently linked to two, three, or four alkyl substituents, resulting in secondary, tertiary, or quaternary amines. This compound family is present throughout the Tree of Life and its members possess a diverse range of biological functions. For example, choline is required for membrane phospholipid synthesis (19), acetylcholine is the major neurotransmitter (20), glycine-betaine and carnitine are important osmo- and cryoprotectants (21, 22), trimethylamine N-oxide (TMAO) is an electron acceptor for anaerobic respiration in bacteria (23), and methyl-, dimethyl-, and trimethylamines are important growth substrates for methanogenic archaea in habitats ranging from marine environments (24) to the human gut (25).

Here, we identify a sequence motif for biogenic amine binding in dCache_1 domains (termed dCache_1AM) and show that it evolved from the ubiquitous amino acid binding motif dCache_1AA, which is found throughout the Tree of Life. This study further demonstrates that our approach is applicable to characterizing other ligand-binding domain families thus leading to substantial gain in knowledge on signal transduction systems.

Results

Different Orientation of Ligands in Quaternary Amine-binding dCache_1 Domains. We analyzed several solved structures of dCache_1 sensory domains in complex with quaternary amines: McpX from *Sinorhizobium meliloti* in complex with proline betaine (PDB ID 6D8V) (26), PacA from *Pectobacterium atrosepticum* in complex with betaine (trimethyl glycine; PDB ID 7PSG) (17), and PctD from *Pseudomonas aeruginosa* in complexes with choline (PDB ID 7PRQ) and acetylcholine (PDB ID 7PRR) (17). Although the ligands are found in the same binding pocket in all three chemoreceptors, they are oriented in various directions with respect to their oxygen and nitrogen atoms (Fig. 1 *A*–*D*). This is in stark contrast with the dCache_1AA domains, where various

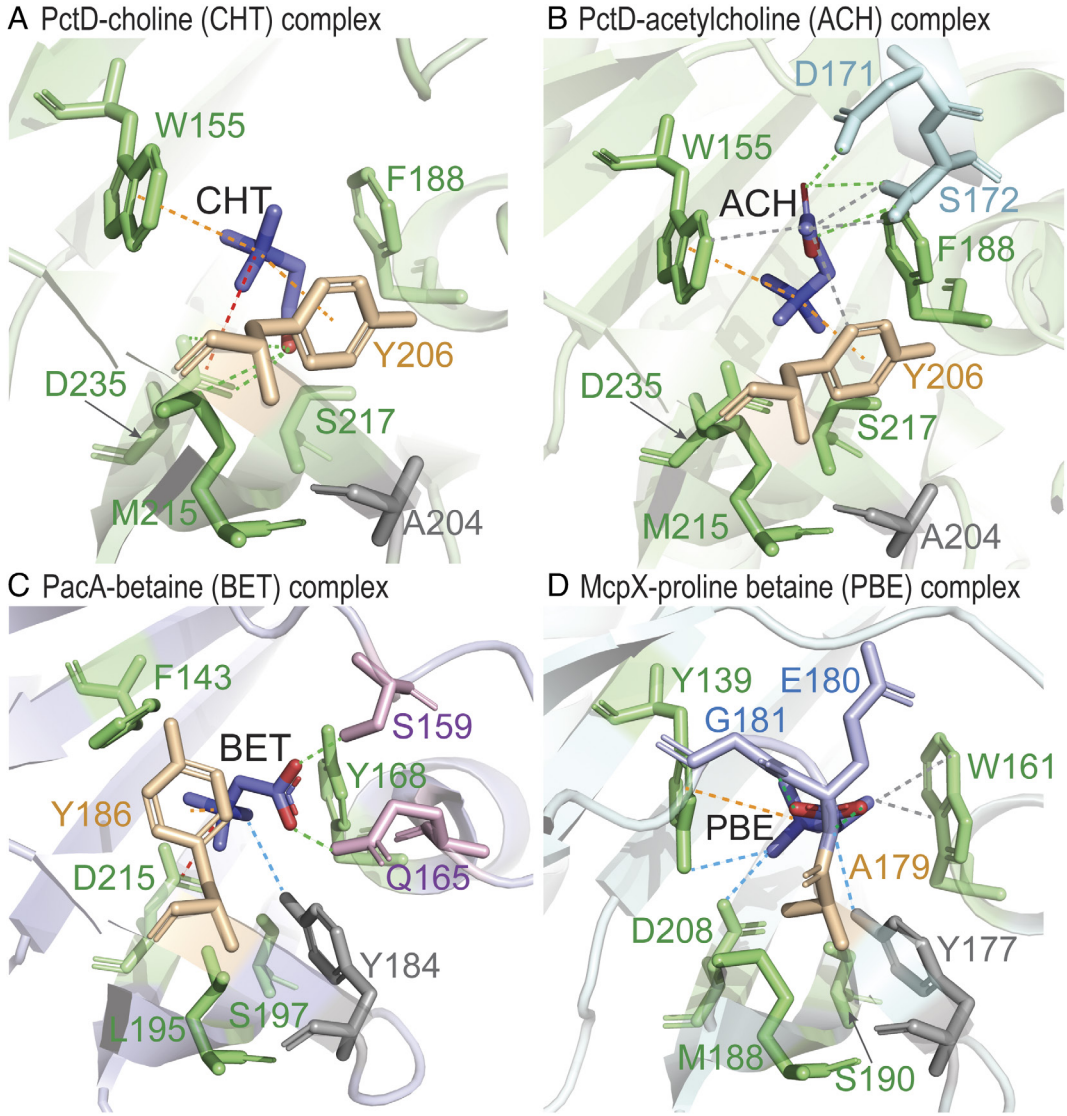

Fig. 1. Amine-binding receptors have a common ligand-binding interface. (*A*) PctD in complex with choline (CHT). (*B*) PctD in complex with acetylcholine (ACH). (*C*) PacA in complex with betaine (BET). (*D*) McpX in complex with proline betaine (PBE). Dashed lines: green – hydrogen bonds, blue – weak hydrogen bonds, orange – cation–π, red – ionic, gray – hydrophobic interactions. Residues shown in green indicate amino acids conserved across structures and analyzed protein sequences. The aromatic residue shown in yellow makes a cation–π bond with ligands. In McpX, A179 is in the corresponding position, and it cannot provide such a bond. Another cation–π bond is made with the aromatic residue corresponding to W155 in PctD.

amino acids bind in the same orientation (8). This phenomenon was observed not only between receptors but also in the same receptor with different quaternary amine ligands (Fig. 1 *A* and *B*). However, we noticed a common feature for all bound ligands—the cation–π interaction with the π-system of aromatic residues in the ligand-binding pocket. For example, in the PctD, ligand-binding pocket choline and acetylcholine are oriented in opposite directions, but both make cation–π bonds with Y206 and W155 (Fig. 1 *A* and *B*). In PacA, betaine interacts with Y186, also through the cation–π bond (Fig. 1*C*) and proline betaine in McpX interacts with the favorably positioned Y139 (Fig. 1*D*). Considering that bonding energy of the cation–π interaction can be significant (27, 28), this type of bond may be of special importance for quaternary amine-binding, acting as a "hinge" around which ligands can be oriented in various directions. In addition, several residues that form the ligand-binding interface and provide stabilization of bound ligands are conserved among the receptors (Fig. 1, residues in green). Specifically, proline betaine makes a weak hydrogen bond with D208 of McpX (Fig. 1*D*), betaine makes an ionic bond with the corresponding residue in PacA (Fig. 1*C*), while choline and acetylcholine make hydrogen bonds with the corresponding D235 in PctD (Fig. 1 *A* and *B*). In both McpX and PacA, quaternary amines also make two hydrogen bonds with two nonconserved residues (G181/E180 and S159/Q165, respectively) that "lift" the oxygen-containing end of the ligands (Fig. 1 *C* and *D*). Thus, it appears that all three quaternary amine-binding dCache_1 domains (PctD, PacA, and ad McpX) might share a conserved ligand-binding motif. Another unique feature of the three quaternary amine-binding receptors is the presence of an additional alpha-helical element above the binding pocket, where two charged residues (R103 and D133 in PctD) interact via two hydrogen bonds (Fig. 2*A*).

Amine-binding dCache_1 Domains Share a Conserved Sequence Motif. To find out whether the amine-binding site observed in PctD, PacA, and McpX is conserved in other homologous sequences, we performed sensitive profile searches against the NCBI RefSeq database (*Materials and Methods*) and collected 20,000 dCache_1 domain sequences most similar to PctD, PacA, and McpX. Next, we built a multiple sequence alignment of these sequences and tracked residues forming the amine-binding interface in the available structures of quaternary amine-receptor complexes. Through this analysis, we identified the most conserved ligand-binding residues, which we tentatively defined as a signature motif for this type of dCache_1 domains (Fig. 2). The motif has been identified in the dCache_1 domain of ~13,000 protein sequences (Dataset S1, multiple sequence alignments are available at https://github.com/ToshkaDev/Amine_motif). Interestingly, the structure of one of these domains, from a histidine kinase of an archaeon *Methanosarcina mazei*, is solved [PDB ID 3LIB (29)]. We superimposed this structure with that of the quaternary amine-binding sensory domain from *P. aeruginosa* PctD and found that overall structures and ligand-binding sites are remarkably similar (Fig. 2*A*). Furthermore, residues constituting the motif are located at very similar positions in the bacterial and archaeal proteins (Fig. 2*A*, close-up view, lower panel). We also established that charged residues above the ligand-binding module (R103 and D133 in PctD) are notably conserved (Fig. 2*A*, close-up view, upper panel). Based on these observations, we deduced a motif for amine-binding receptors (AM_motif) (Fig. 2*C*) and tested it in subsequent experiments.

To verify the contribution of individual conserved amino acid residues of the proposed sequence motif to ligand binding, we prepared alanine substitution mutants of PctD-dCache_1 and submitted them to microcalorimetric titrations with choline using the same experimental conditions that were used for the analysis of the wild-type protein (17) (Fig. 3). Because significant reduction in the binding affinity was observed for each mutant, experiments were repeated with a higher choline concentration to derive the dissociation constants (*SI Appendix*, Fig. S1). The replacement of the W155 and F188 residues that sandwich the bound ligand resulted in reductions in affinity by factors of 175 and 373, respectively (Table 1). Significant reductions were also observed for Asp235 and Arg103 substitutions (Table 1). Whereas the former residue is part of the binding pocket establishing a hydrogen bond with the bound ligand, Arg103 is outside the binding pocket and likely plays an important role in maintaining the correct geometry of the binding pocket. Replacement of S217 (interaction with choline via a water molecule) and M215 (hydrophobic interaction) had a more modest impact (Table 1). Upon experimental verification, we termed domains that share the motif dCache_1AM.

dCache_1AM Domains Are Widespread and Found in all Major Receptor Types. We identified dCache_1AM domains in more than 13,000 proteins from 8,000 species of bacteria and archaea

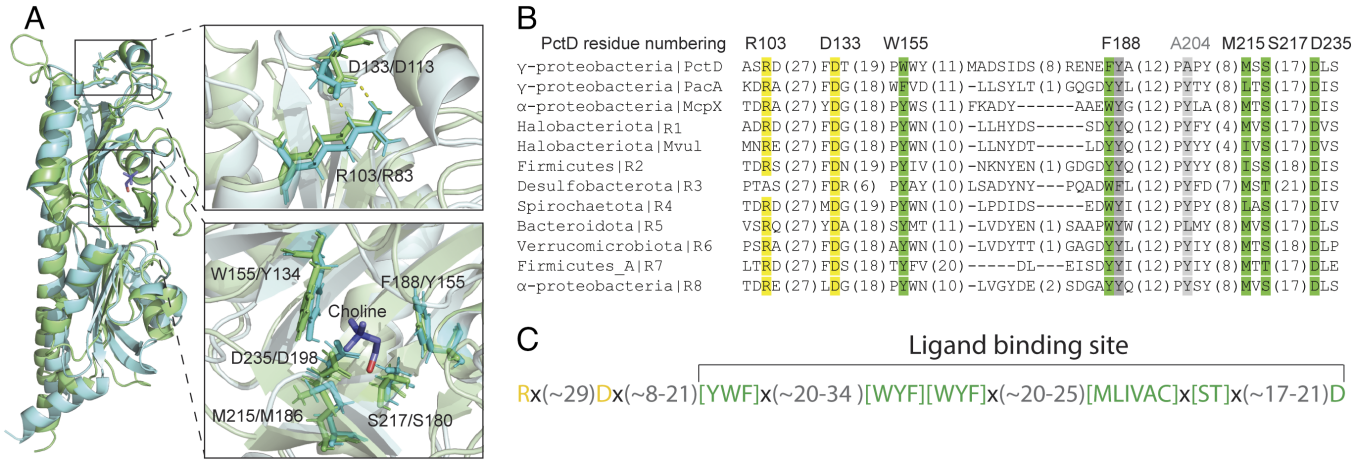

Fig. 2. Amine-binding sensors have a characteristic signature. (*A*) Structural superimposition of the PctD sensory domain with the sensory domain of the histidine kinase from the archaeon *Methanosarcina mazei*. PctD is shown in green, the archaeal protein – in cyan. In residue labels in close-up views, the first residue corresponds to PctD, the second—to the archaeal protein. (*B*) A multiple sequence alignment of amine receptor sensory domain sequences. Residues within the ligand-binding site are shown in green and gray, residues in the structure above the ligand-binding site are shown in yellow. A residue that is not part of the ligand-binding interface but which exhibits high conservation is shown in dark gray. (*C*) The amine-binding motif (AM_motif).

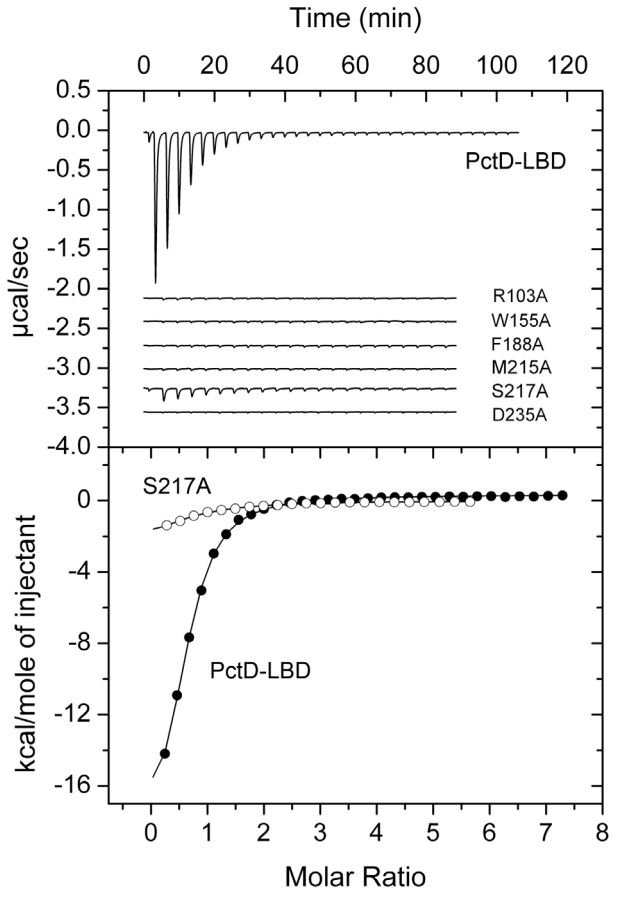

Fig. 3. Microcalorimetric titrations of the PctD sensor domain and site directed mutants in individual residues of the amine-binding motif with choline. *Upper* panel: Raw data for the titration of 16 µM of protein with 9.6 µL aliquots of 0.5 µM choline. *Lower* panel: Concentration-normalized and dilution heat–corrected integrated raw data. The line is the best fit using the "One binding site model" of the MicroCal version of ORIGIN. In case no binding heats were observed, experiments were repeated with a higher ligand concentration. The resulting curves are shown in *SI Appendix*, Fig. S1, and the derived dissociation constants are provided in Table 1.

(Datasets S1 and S2). We have extracted the corresponding isolation sources from the NCBI BioSample section of NCBI (Dataset S2) and observed that many dCache_1AM domains come from plant-associated bacteria, both beneficial, such as *S. meliloti* and *Azospirillum brasilense*, and pathogenic, such as *Xanthomonas campestris*. Similarly, we identified dCache_1AM domains in the human microbiome, including both beneficial human gut bacteria, for example, *Roseburia intestinalis* and *Ruminococcus lactaris*, and human pathogens, such as *Campylobacter jejuni* and *Aeromonas hydrophila* (Fig. 4 and Dataset S2). Analysis of dCache_1AM phyletic distribution revealed that receptors containing this sensor domain are found in many bacterial and in one archaeal phylum—Halobacteriota (Fig. 4*A* and Dataset S2).

Table 1. Microcalorimetric studies for the binding of choline to PctD-LBD and site directed mutants in amino acids of the QA binding motif

Protein	*K* (µM) D	Reduction with respect to wt protein (fold)
PctD-LBD*	2.6 ± 0.1	-
PctD-LBD_D235A	421 ± 17	162
PctD-LBD_F188A	970 ± 117	373
PctD-LBD_W155A	454 ± 20	175
PctD-LBD_M215A	150 ± 17	58
PctD-LBD_S217A	14.2 ± 1.4	5.4
PctD-LBD_R103A	943 ± 54	363

*Reported previously in ref. 17.
The corresponding titration curves are shown in *SI Appendix*, Fig. S1.

By performing domain analysis of all dCache_1AM-containing proteins, we revealed that they are exclusively found in signal transduction proteins or as stand-alone domains (Fig. 4*B* and Dataset S1). We found dCache_1AM domains in all four major types of bacterial and archaeal transmembrane receptors: chemoreceptors, sensor histidine kinases, cyclic (di)-nucleotide turnover enzymes, and serine/threonine phosphatases (Fig. 4*B* and Dataset S1). In bacteria, the vast majority of dCache_1AM domains are found in chemoreceptors, whereas in archaea, they are predominantly found in sensor histidine kinases. We also identified an unusual case of an intracellular dCache_1AM domain in Actinobacteria. In many actinobacterial genomes, including representatives of *Streptomyces*, *Nocardia*, *Rhodococcus*, *Arthrobacter, Mycobacterium*, and other genera, a stand-alone dCache_1AM domain protein is encoded in a two-gene operon with a transcription factor (Fig. 4*B* and Dataset S2). This is a rare example of repurposing an extracellular domain for intracellular sensing.

dCache_1AM Domains Bind Primary, Secondary, Tertiary, and Quaternary Amines. From the list of more than 13,000 dCache_1AM sequences (Dataset S1), we selected ten targets (R1 through R10) for experimental verification (Table 2). These domains were selected from i) four major receptor families, namely sensor histidine kinases, chemoreceptors, Ser/Thr phosphatases, and diguanylate cyclases/phosphodiesterases, ii) one archaeal and several bacterial phyla (Table 2), and iii) based on the amine motif variability. Two targets, each containing a single amino acid substitution in the AM_motif, were used as a negative control (R9 and R10, Table 2). The individual dCache_1AM domains from these ten receptors were overexpressed in *Escherichia coli* and purified by affinity chromatography. For nine of these proteins, a buffer system was identified that guaranteed protein solubility and stability, whereas the remaining protein, R2, was insoluble despite several solvent engineering efforts. Thermal unfolding studies revealed a transition for all other proteins, which is indicative of protein folding. Because the AM_motif was established based on the binding of quaternary amines, thermal shift unfolding studies and isothermal titration experiments were conducted in parallel to study the binding of various quaternary amines. We have conducted isothermal titration calorimetry (ITC) experiments for ligands that caused an increase in the midpoint of proteins unfolding (Tm) by at least 2 °C. We were able to observe the binding of choline and/or acetylcholine to four of the eight target proteins and one target protein, R5, bound trimethylamine N-oxide (TMAO) (Fig. 5, *SI Appendix*, Fig. S2, and Table 3).

Surprisingly, none of the other quaternary amines or polyamines putrescine and agmatine that were among compounds of the Biolog PM3B array caused any significant increases in Tm. Since thermal unfolding experiments at times give false-negative results (i.e., ligand binding that does not significantly increase the Tm), we conducted isothermal titration experiments with the maximal possible concentration (10 to 20 mM) of other biologically important quaternary amines, such as L-carnitine, betaine, proline, and trigonelline (Footnotes to Table 3). However, we observed no binding in any of these experiments, suggesting that choline and acetylcholine are the main quaternary amine ligands recognized by dCache_1AM-containing receptors. Computational docking experiments using structures and AlphaFold models revealed that an aromatic residue at the position corresponding to Y184 in PacA

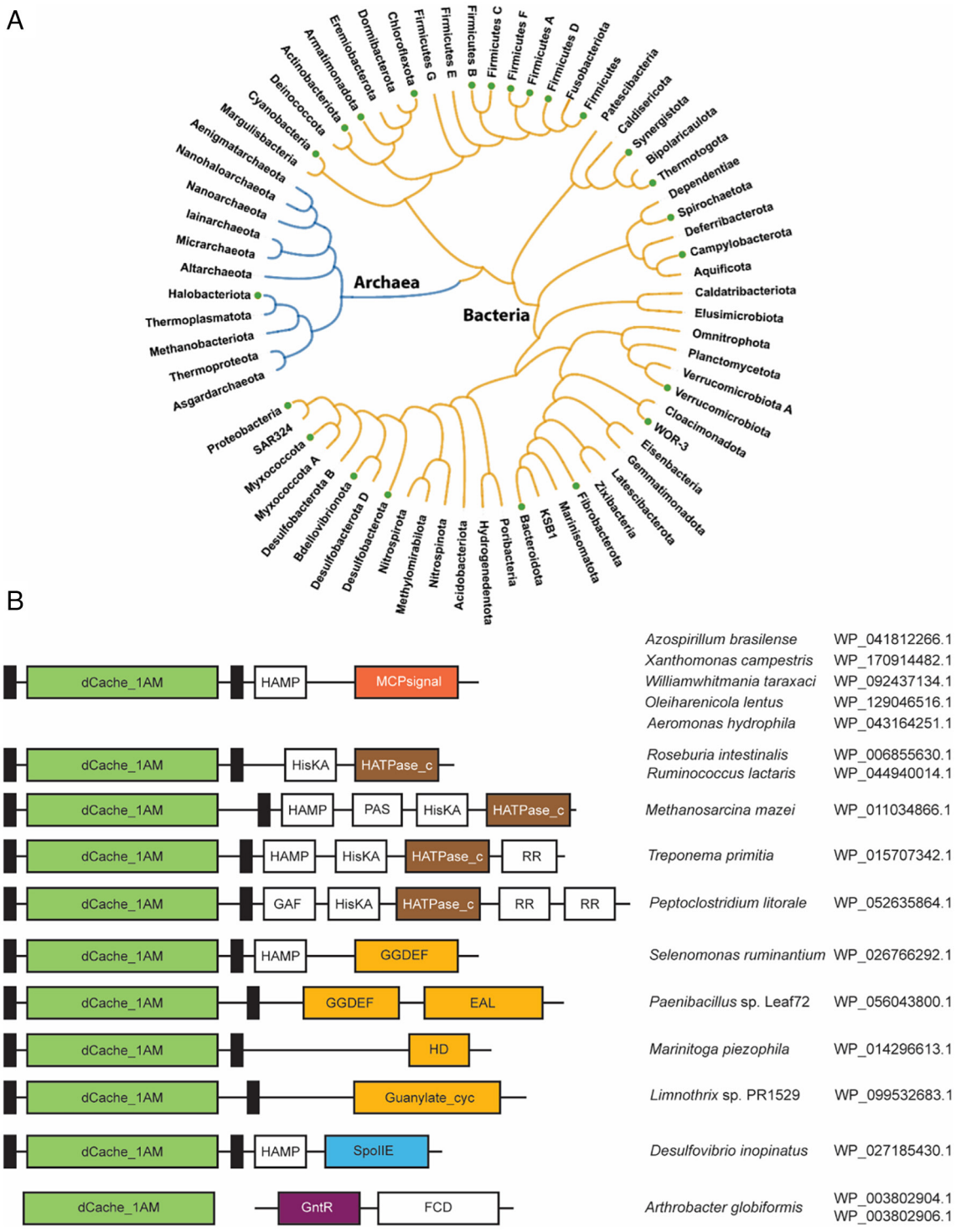

Fig. 4. Phyletic distribution (*A*) and prevalent domain architectures of amine receptors (*B*). Solid circles at the tips of the tree branches indicate that the AM_motif was found in the corresponding phylum. Black narrow rectangles depict transmembrane domains. Domain definitions follow the Pfam domain nomenclature (InterPro IDs are in parentheses): EAL (IPR035919), diguanylate phosphodiesterase; GGDEF (IPR000160), diguanylate cyclase; Guanylate_cyc (IPR001054), adenylate or guanylate cyclase; HATPase_c (IPR003594), histidine kinase; HD (IPR006674), phosphohydrolase; MCPsignal (IPR004089), methyl-accepting chemotaxis protein (chemoreceptor); SpoIIE (IPR001932), serine/threonine phosphatase; GntR (IPR000524), transcription regulator; HAMP (IPR003660), domain present in Histidine kinases, Adenyl cyclases, Methyl-accepting proteins and Phosphatases; HisKA (IPR003661), Histidine Kinase A (dimerization/phosphoacceptor) domain; RR, Response_reg (IPR001789), response regulator; GAF (IPR003018), domain found in proteins including cGMP-specific phosphodiesterases, adenylyl cyclases and FhlA; PAS (IPR000014), domain named after Per – period circadian protein, Arnt – Ah receptor nuclear translocator protein, and Sim – single-minded protein; FCD (IPR011711), the C-terminal ligand-binding domain of many members of the GntR family.

(Y177 in McpX) can interfere with ligand binding, unless the ligand is oriented in a more horizontal plane within the pocket (due to contacts with residues favorably positioned above and around the ligand). This "more horizontal" orientation is observed in PacA-betaine and McpX-proline betaine complexes (Fig. 1 *C* and *D*). On the other hand, in PctD-choline and PctD-acetylcholine complexes (Fig. 1 *A* and *B*) the ligands are oriented vertically, but at the position that could interfere with ligand binding, a smaller residue (A204) is present instead of a bulky aromatic residue. The presence of the "interfering" aromatic residue might be the reason

Table 2. Characteristics of the 10 receptor proteins analyzed in this study

Prot	Acc. code	Amine-binding prediction	Protein family*	Species/strain	Phylogenetic category	Lifestyle/isolation source (Ref.)
R1	WP_011034866.1	Yes†	HK	*Methanosarcina mazei*	Archaea	Anaerobic, produces methane from methylamines (30)
R2	WP_056043800.1	Yes†	GGDEEF/EAL	*Paenibacillus* sp. Leaf72	Firmicutes	Isolated from *Arabidopsis thaliana* leafs, BacDive:131858
R3	WP_027185430.1	Yes†	S/T phos	*Desulfovibrio inopinatus*	Desulfobacterota	Anaerobic, produces methane from choline (31)
R4	WP_015707342.1	Yes†	HK	*Treponema primitia*	Spirochaetota	Acetogenic, isolated from termite guts (32)
R5	WP_092437134.1	Yes†	CR	*Williamwhitmania taraxaci*	Bacteroidota	Anaerobic, does not grow in TMA (33)
R6	WP_129046516.1	Yes†	CR	*Oleiharenicola lentus*	Verrucomicrobiota	Aerobic, isolated from irrigation waters (34)
R7	WP_052635864.1	Yes†	HK	*Peptoclostridium littorale*	Firmicutes_A	Anaerobic, betaine is a carbon source (35)
R8	WP_041812266.1	Yes†	CR	*Azospirillum*	α-Proteobacteria	Plant-associated, nitrogen-fixing (36)
R9	WP_148526407.1	No‡	CR	*Vibrio cholerae*	γ-Proteobacteria	Facultative anaerobic, growth on TMAO (37)
R10	WP_083052832.1	No‡	CR	*Marispirochaeta aestuarii*	Spirochaetota	Anaerobic, isolated from coastal marine habitats (38)

*HK, serine histidine kinase; GGDEF/EAL, diguanylate cyclase/phosphodiesterase; S/T phos, serine/threonine phosphatase; CR, chemoreceptor.
†Possesses consensus motif.
‡Possesses consensus motif altered in one amino acid.
Proteins R1 to R8 contain the amine-binding motif and are predicted to bind amines. Proteins R9 and R10 possess slight modification of this binding motif (specified beneath the Table) and were predicted to not bind amines.

why four of the selected targets failed to bind quaternary amines, and in subsequent studies, we aimed at establishing their ligands. Target R1 came from an archaeon *Methanosarcina mazei*, which uses methylamines as carbon and energy sources (39). We used the available structure of R1 (PDB ID: 3lib) (29) to conduct in silico docking experiments with methylamine, dimethylamine, and trimethylamine and found that all ligands bind to the same ligand-binding pocket of the target; we observed that methylamine and dimethylamine make contacts with the conserved residues constituting the AM_motif (*SI Appendix*, Fig. S3). Subsequently, we have conducted ligand screening with the target proteins and various small biogenic amines. We observed significant TM increases for methylamine, ethylamine, ethylenediamine, dimethylamine, trimethylamine, and ethanolamine using the thermal shift assay (*SI Appendix*, Table S1). Six out of the seven target proteins showed a significant increase in Tm in the presence of small biogenic amines. We subsequently conducted ITC experiments to derive the corresponding dissociation constants (Fig. 6 and *SI Appendix*, Fig. S4). Three proteins, R1, R7, and R8, showed a wide ligand spectrum and recognized with high affinity four of the amine compounds. Proteins R1 and R8, that come from an archaeon and an alphaproteobacterium, had the same ligand profile. Both proteins had a preference for dimethylamine but also recognized with lower affinity methyl-, ethyl- and trimethylamine (Table 3). Other proteins showed a narrower ligand spectrum; for example, R5 recognized trimethylamine with high affinity, but its affinity for dimethylamine was reduced ~300-fold (Table 3). In addition, R5 bound choline with an affinity very similar to that of trimethylamine, indicative of plasticity in ligand recognition (Table 3).

The only predicted amine-binding dCache_1 domain for which no binding was observed in our experiments was R4. However, it cannot be ruled out that this protein binds other amine(s) that we have not tested. As mentioned above, targets R9 and R10 were used as a negative control: Both contained a single mismatch in the AM_motif. Thermal shift assays and ITC experiments with the maximal possible ligand concentration did not provide evidence for the binding of any of these ligands, further validating the conserved AM-motif. Based on these results, we wanted to verify whether the quaternary amine sensing chemoreceptor PacA (17) also binds methylamines that were not tested in the previous study (17). Thermal shift (*SI Appendix*, Table S1) and microcalorimetric titrations (Table 3) revealed that indeed PacA also binds trimethylamine, further suggesting that the capacity to bind various biogenic amines—primary, secondary, tertiary, and quaternary—is a general property of the dCache_1AM domain family.

dCache_1AM Domains Evolved from the Universal Amino Acid Sensor. To establish the evolutionary origins of the dCache_1AM domains, we performed sequence, structure, and phylogenetic analyses. Analysis of the multiple sequence alignment showed that AM- and AA motifs have several overlapping positions (Fig. 7*A*, the alignment in FASTA format is available at https://github.com/ToshkaDev/Amine_motif). Two aromatic residues (corresponding to F188 and Y189 in PctD) in the middle of the motifs and aspartate at the end of the motifs (corresponding to D235 in PctD; see Fig. 7*A*) are conserved in both motifs. Another position highly conserved in amino acid receptors as the aromatic position is shared with amine receptors (corresponds to A204 in PctD), although the position is more variable in dCache_1AM. In contrast, an aromatic and a positively charged positions at the

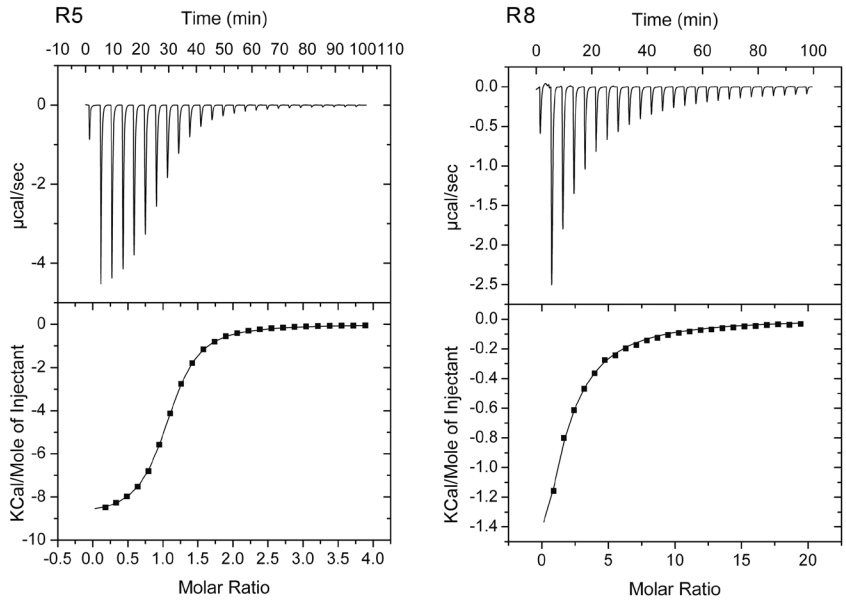

Fig. 5. Microcalorimetric titrations of predicted amine responsive dCache domains with choline. *Upper* panel: Raw data for the titration of 75 µM of protein with 8.0 µL aliquots of 2 mM (R5) or 10 mM (R8) choline. *Lower* panel: Concentration-normalized and dilution heat–corrected integrated raw data. The line is the best fit using the One binding site model of the MicroCal version of ORIGIN.

beginning of the AA motif are not conserved in the AM motif. dCache_1AM domains, in addition, have an insertion upstream of the ligand-binding site, which includes charged residues R103 and D133 that make hydrogen bonds with each other (Figs. 2*A* and 7*A*). Two positions of the AM motif (corresponding to M215 and S217 in PctD) are also found in amino acid receptors, although while in dCache_1AM the M215 position is exclusively hydrophobic, in dCache_1AA this position is more variable (Figs. 2*B* and 7*A*).

To explore structural similarities, we superimposed ligand-binding pockets of dCache_1AA and dCache_1AM and observed similar ligand-binding interfaces (Fig. 7*B*). Key residues constituting both the AM and AA motifs are in the same positions. One of the key residues of the AA_motif, corresponding to D173 in PctD, is well conserved in amine receptors; quaternary amines in PctD-CHT, PacA-BET, McpX-PBE complexes, as well as docked methylamines in the archaeal protein make contacts with this residue (Figs. 1 and 7 *A* and *B* and *SI Appendix*, Fig. S3). Two shared aromatic residues (Y168/W128 and Y184/Y144) are in favorable positions for ligand binding in both amine and amino acid receptors. A structurally important and highly conserved aromatic residue, which does not interact with the ligand (Y162/Y129) is found in a similar position in both amine-binding PacA and amino acid-binding PctA (Fig. 7*B*).

Table 3. Dissociation constants derived from microcalorimetric binding studies of different amines to sensor domains of the 10 receptors analyzed

	K_D (uM)								
	Quaternary amines			Methyl-/ethyl-/ethanol amines*					Polyamine
Protein	Acetylcholine	Choline	TMAO	Methyl-amine	Ethyl-amine	Dimethyl-amine	Trimethyl-amine	Ethanola-mine	Ethylene-diamine
R1†	1,845 ± 1,120	Nb‡	Nb	35.1 ± 3	19.6 ± 2	**0.53 ± 0.02**	10.0 ± 0.4	Nb	Nb
R2	Protein insoluble								
R3§	Nb	Nb	Nb	Nb	Nb	Nb	Nb	**2.3 ± 0.1**	29.6 ± 3
R4¶	Nb	Nb	Nb	Nb	Nb	Nb	Nb	Nb	Nb
R5†	581 ± 129	**3.6 ± 0.1**	1,742 ± 88	Nb	Nb	311 ± 79	**1.2 ± 0.1**	Nb	Nb
R6†	Nb	Nb	Nb	Nb	Nb	Nb	151 ± 4	Nb	Nb
R7§	Nb	2,950 ± 217	Nb	16.6 ± 2	Nb	**5.3 ± 0.4**	12.6 ± 1	**1.7 ± 0.02**	12.3 ± 1
R8†	735 ± 24	183 ± 4	Nb	25.1 ± 3	46.1 ± 7	**5.8 ± 1**	24.0 ± 4	Nb	Nb
R9§	Nb	Nb	Nb	Nb	Nb	Nb	Nb	Nb	Nb
R10§	Nb	Nb	Nb	Nb	Nb	Nb	Nb	Nb	Nb
PacA-LBD	Nb#	113 ± 16#	Nd‖	Nb	Nb	Nb	113 ± 6	Nb	Nb

*Volatile organic compound.
†No binding by ITC of L-carnitine, betaine, proline, and trigonelline.
‡No binding by ITC and/or thermal shift assay (increases above 2 °C were considered significant).
§No binding by ITC of L-carnitine and trigonelline.
¶No binding by ITC of L-carnitine, betaine, and trigonelline.
#Reported previously in ref. 17.
‖Not determined.
Dissociation constants below 10 µM are shown in bold face.

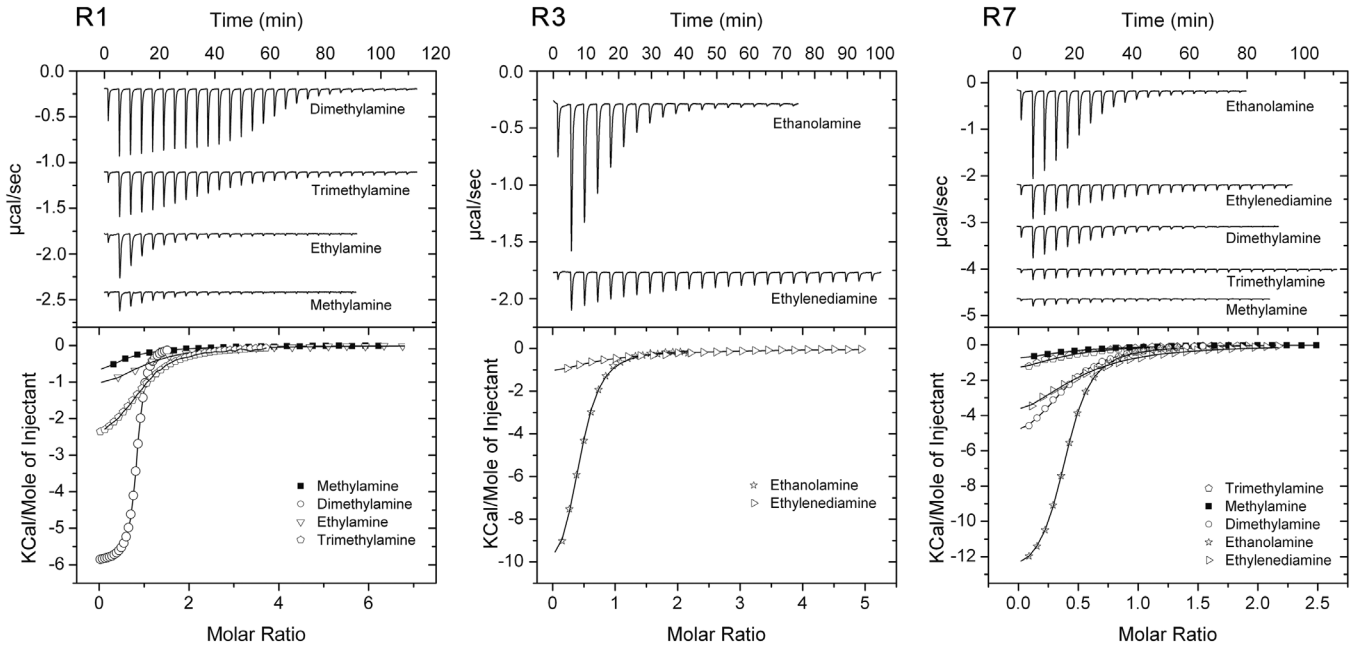

Fig. 6. Microcalorimetric titrations of predicted amine responsive dCache domains with small biogenic amines. *Upper* panel: Raw data for the titration of 30 to 50 µM of protein with 3.2 to 11.1 µL aliquots of 1 to 2 mM amine solutions. *Lower* panel: Concentration-normalized and dilution heat–corrected integrated raw data. The line is the best fit using the One binding site model of the MicroCal version of ORIGIN.

The main difference between AA and AM motifs was observed in three positions. A key position of the AA_motif corresponding to R126 in PctA makes a hydrogen bond with the carboxyl group of amino acid ligands. This position is not conserved in amine receptors and in its closest structural equivalent in PacA is Q165, which also makes a hydrogen bond with its designated ligand, betaine (Figs. 1*C*

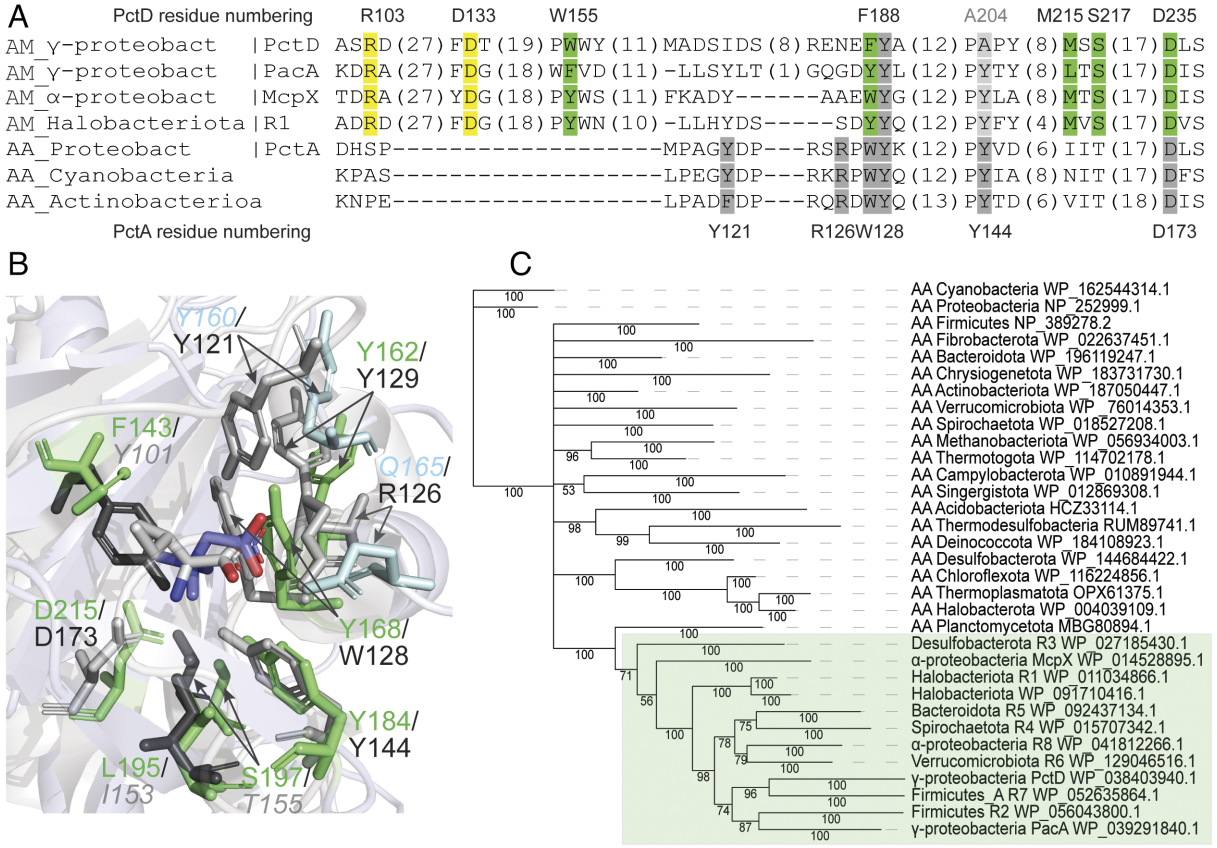

Fig. 7. Amine receptors evolved from amino acid receptors. (*A*) Multiple sequence alignment of amine and amino acid receptor sensory domains. (*B*) Structural superimposition of ligand-binding pockets of the amine receptor PacA from *P. atrosepticum* and the amino acid receptors PctA from *Pseudomonas aeruginosa*. Residues corresponding to the AM_motif are shown in green; residues in gray correspond to the AA_motif. Residues in the PacA pocket making hydrogen bonds with betaine are shown in blue. Residues that are not part of the corresponding motif but that are in positions equivalent to the residues in the other motif are shown with a faint italic font. (*C*) Bayesian phylogenetic tree of amine and amino acid receptor sensory domains. Amine receptors are shown in green background.

and 7*B*). An aromatic residue Y160 corresponding to Y121 in PctA is oriented in the opposite direction in PacA and does not contribute to ligand coordination (Fig. 7*B*). Conversely, an important aromatic residue making a cation–π bond with quaternary amine ligands is oriented downward in amino acid receptors and cannot interact with amino acid ligands (Fig. 7*B*). We then inferred a Bayesian phylogenetic tree using protein sequences of dCache_1AA and dCache_1AM from several bacterial and archaeal phyla (see *Materials and Methods*, the tree in NEXUS format is available at https://github.com/ToshkaDev/Amine_motif). The tree showed that all dCache_1AM sequences are found in a single branch derived from one of the branches of a more diverse set of dCache_1AA sequences (Fig. 7*C*). An amino acid receptor from Planctomycetota was the closest to amine-binding sensors among the ones used for the phylogenetic inference.

Discussion

Prior to this work, three bacterial chemoreceptors from closely related proteobacterial species were shown to bind and respond to quaternary amines and polyamines as signaling molecules (15–18). In this study, we identify thousands of bacterial and archaeal receptors containing dCache_1AM domains that bind various biogenic amines. We computed and experimentally verified a conserved sequence motif signature for this class of sensory domains. We show that dCache_1AM sensors bind not only quaternary but also primary, secondary, and tertiary amines as well as the polyamine ethylenediamine. Furthermore, based on their distribution and superior affinity, we conclude that small biogenic amines rather than quaternary amines constitute the primary ligand group for the dCache_1AM class (Table 3).

We show that amine sensors evolved from the universal dCache_1AA amino acid sensors (8) through a small insertion in the ligand-binding pocket and replacement of key ligand-binding residues. Although amine sensors are not as ubiquitous as amino acid sensors, they are still widespread. These domains are identified in all major types of bacterial transmembrane receptors—chemoreceptors, sensor histidine kinases, serine/threonine phosphatases, and cyclic (di) nucleotide turnover enzymes—from several major bacterial and one archaeal phyla. In archaea, we identified them in two classes of Halobacteriota—Methanomicrobia and Methanosarcinia—where they were most likely horizontally transferred from bacteria. Structural analyses of bacterial and archaeal receptors showed that the ligand-binding interface is well conserved in phylogenetically distant species.

Amine compounds sensed by dCache_1AM receptors are of enormous ecological importance. Many biogenic amines serve as energy, carbon, and nitrogen sources for bacteria (40, 41). Quaternary amines, such as acetylcholine, are important neurotransmitters and mediators of inter-kingdom and inter-bacterial interactions (42). Methyl-/ethyl-/ethanol amines are volatile organic compounds (voc). Voc are known to mediate plant-bacteria interactions and bacteria-derived voc were found to promote the growth, health, immunity, and stress resistance of plants (43). However, the molecular detail of voc signaling is frequently lacking. The presence of dCache_1AM domains in many plant-associated bacteria (Dataset S2) suggests a potential role of the corresponding receptors in mediating plant-bacteria interactions. The fact that the domain family identified responds to QAs as well as to methyl-/ethyl-/ethanol amines is the consequence of i) an important structural similarity of these ligands (*SI Appendix*, Fig. S5*A*) and ii) an intertwining of their metabolism (*SI Appendix*, Fig. S5*B*). For example, methanogenic archaea use methylamine, dimethylamine, trimethylamine, betaine, and choline as carbon and energy sources (44). Methylamine and trimethylamine are the result of betaine (24) and choline degradation (45), respectively. In humans, the production of TMAO from trimethylamine is linked to multiple diseases such as trimethylaminuria, atherosclerosis, or cardiovascular disease (46, 47). In marine environments, trimethylamine is produced from TMAO, glycine betaine, choline, and carnitine (48, 49). Dimethylamine and trimethylamine are produced in sewage water from choline and creatinine, which are present in urine as well as plant and animal tissues (50). Of all tested ligands and target sensor domains, the highest binding affinity was observed for the binding of dimethylamine to the dCache_1AM domain (target R1) that is found in a histidine kinase from *Methanosarcina mazei* (Table 3). Interestingly, R1 was among the first reported 3D structures of the Cache superfamily (29); however, ligands recognized by this domain remained unknown until now. *M. mazei* is a model methanogenic archaeon, found primarily in sewages and anoxygenic environments that are sources of quaternary and methylated amines (24). *M. mazei* has multiple genes encoding enzymes that demethylate trimethylamine (*mttB*), dimethylamine (*mtbB*), and methylamine (*mtmB*) (51), which is the initial event of a multi-step process leading to methane generation and energy production. We found that R1 binds all three substrates of the demethylation reactions with high affinity (Table 3). The fact that R1 also binds ethylamine (Table 3) suggests that this compound may also serve as a substrate for the demethylation reactions. The R1-containing sensor histidine kinase is encoded in a gene cluster with a dimethylamine methyltransferase (*SI Appendix*, Fig. S6) and, coincidently, dimethylamine is a preferred ligand for R1 (Table 3). Transcript levels of this methyltransferase are significantly upregulated in the presence of trimethylamine (51), which is the second-best substrate for R1, and it is likely that the R1-containing histidine kinase is involved in this regulatory process.

The ability to bind the same class of chemical signals evolved multiple times in various lineages of life. For example, amino acids are recognized by eukaryotic G-protein-coupled receptors (GPCRs) (52) and by nonhomologous bacterial domains, such as 4HB (four-helix bundle) (53), DAHL (double all-helical ligand-binding) (54), FIST (F-boxes and intracellular signal transduction proteins) (55), and dCache_1 (8). Similarly, amines can be recognized by eukaryotic GPCRs (56) and by dCache_1 domains reported here. There appear to be another class of dCache_1 domains, exemplified by McpU of *Pseudomonas putida* and TlpQ of *Pseudomonas aeruginosa* (15, 16) that specifically recognize polyamines, but it has not been characterized as yet. The LBD from the GacS histidine kinase (57), which appears to belong the DUF2222 family (7) of the Cache subfamily, is another example of polyamine-responsive sensor. Thus, we expect other types of bacterial and eukaryotic amine-recognizing sensor domains to be discovered.

The successful prediction of ligands for sensor domains that recognized their ligands through different types of bonding interactions, i.e., primarily H-bonds or hydrophobic interactions, indicates that this procedure appears to be generally applicable to identify ligands recognized by sensor domains. These approaches will allow the annotation of thousands of receptor proteins with a cognate ligand class. Such information will have a major impact in the field since it will enable precise wet-lab experiments to define the function of a given regulatory system. Thus, our approach holds the strong promise to address an important bottleneck in microbiology (5): identification of signal molecules that stimulate numerous signal transduction cascades in bacteria and archaea.

Materials and Methods

Strains and Plasmids. The bacterial strains and plasmids used are listed in *SI Appendix*, Table S2.

Identification of the Amine Motif Containing dCache Domains. Amine-binding protein sequences were identified in the following two steps. The NCBI RefSeq protein database was downloaded (July 2022) and searched with the dCache_1 domain profile hidden Markov model [the InterPro (58) identifier IPR033479] with the E-value threshold of 0.01 both for sequences and domains. Protein sequence regions corresponding to the dCache_1 domain were extracted from the identified sequences and divided into four separate datasets, and each was aligned on the local computational cluster using the FFT-NS-2 algorithm of the MAFFT package (59). Then, in each dataset, the positions corresponding to the defined AM_motif were tracked and corresponding protein sequences were extracted. In parallel, PSI-BLAST searches were initiated against the NCBI RefSeq database using the McpX, PacA, and PctD dCache_1 domain sequences with the maximal number of sequences set to 20,000. The obtained protein dCache_1 domain sequences were merged and aligned using the FFT-NS-2 algorithm of the MAFFT package. In the aligned protein set, positions corresponding to the AM_motif were tracked and corresponding protein sequences were extracted. At the final step, protein sequences extracted from the results of the HMM and PSI-BLAST searches were combined and realigned using the L-INS-i algorithm of the MAFFT package and again the AM_motif positions were tracked and verified. The GTDB taxonomy for the final sequence set was retrieved using the GTDB metadata tables (https://data.ace.uq.edu.au/public/gtdb/data/releases/release202/202.0/).

Multiple Sequence Alignment, Domain Identification. Jalview (60) was used to explore and edit the alignments. Domains were identified running TREND (61, 62) or CDVist (63) with the Pfam profile HMMs. The generated data were downloaded in JSON format from the website and processed programmatically to determine domain architecture variants and abundances. Additional sensitive profile-profile searches were carried out using HHpred (64).

Phylogeny Inference. The sequence alignment was edited using an alignment trimming tool, trimAl (65): Positions in the alignment with gaps in 10% or more of the sequences were removed unless this leaves less than 60%. In such case, the 60% best (with fewer gaps) positions were preserved. The amino acid replacement model for the set of protein sequences was determined running ProtTest (66) and based on Akaike and Bayesian information criteria. Using the determined amino acid replacement model, a phylogenetic tree was inferred using a Bayesian inference algorithm implemented in MrBayes (67). Metropolis-coupled Markov chain Monte Carlo simulation implemented in MrBayes was run with three heated and one cold chain and discarding the first 25% of samples from the cold chain at the "burn-in" phase. A total of 900,000 generations were run till the sufficient convergence was achieved (the average standard deviation of split frequencies equal to 0.01) with chain sampling every 500 generations.

Protein Structure Manipulations. Protein structures of the target proteins were modeled using AlphaFold 2 (68). Comparative analysis of solved and modeled protein structures was done using the PyMOL Molecular Graphics System (Version 20, Schrödinger, LLC) and Mol* Viewer (69).

In Silico Docking. AutoDock Vina (70) was used for computational docking experiments. The protein structure of the histidine kinase dCache_1 domain from the archaeon *Methanosarcina mazei* (PDB ID 3LIB) was prepared using MGLTools. For the experiments, we downloaded ligands from the ZINC database (71) in mol2 format and prepared them for the analysis using the Open Babel toolbox (72) and custom shell script. The docking was performed with the search exhaustiveness 8. Coordinates of the center of the simulation box (Angstroms): X: −25.388; Y: 45.395; Z: −2.575, b) the box dimensions (Angstroms): X: 20; Y: 24; Z: 20.

Construction of the Tree of Life. For the schematic representation of the Tree of Life (Fig. 4*A*), bacterial and archaeal phylogeny was retrieved from the GTDB taxonomy (73). Phyla with at least 10 genomes were depicted. Eukaryotic phylogeny was adapted from refs. 74 and 75. The overall tree topology is based on iTOL v4 (76).

Protein Overexpression and Purification. The transmembrane regions of proteins R1 to R10 were determined using TMHMM (77). pET28b(+) expression plasmids encoding the region between both transmembrane regions (i.e., the LBD) fused to an N-terminal His-tag were purchased from GeneScript Biotech (Netherlands). The corresponding protein sequences are provided in *SI Appendix*, Table S3. The site-directed mutants of PctD-LBD were purified like the native protein (17). The remaining proteins were purified as described previously (78) using the buffers specified in *SI Appendix*, Table S4. Freshly purified proteins were dialyzed overnight into the buffers specified in *SI Appendix*, Table S4 for immediate analysis.

Thermal Shift Assays. The detailed experimental protocol of the thermal shift assays has been reported in ref. 79. Briefly, assays were carried out using a MyIQ2 Real-Time PCR instrument (BioRad, Hercules, CA, USA). Experiments were conducted in 96-well plates and each assay mixture contained 20.5 μL of the dialyzed protein (10–70 μM), 2 μL of 5 X SYPRO orange (Life Technologies, Eugene, Oregon, USA) and 2.5 μL of the 20 mM ligand solution or the equivalent amount of buffer in the ligand-free control. Samples were heated from 23 °C to 85 °C at a scan rate of 1 °C/min. The protein unfolding curves were monitored by detecting changes in SYPRO Orange fluorescence. The Tm values correspond to the minima of the first derivatives of the raw fluorescence data.

ITC. Experiments were conducted on a VP-microcalorimeter (Microcal, Amherst, MA). The site-directed mutants of PctD-LBD were analyzed as reported previously (17). In the case no binding heats were observed, experiments were repeated with a higher choline concentration and the resulting curves (*SI Appendix*, Fig. S1) were used to derive the dissociation constants (Table 1). Proteins R1 to R10 were dialyzed into the analysis buffer, placed at a concentration of 10 to 75 μM into the sample cell and titrated with freshly made up amine solutions (0.5 to 20 mM). In the case no binding heats were observed for a titration with 14.42 μL aliquots of 10 to 20 mM ligand solution, it was concluded that there was no binding. The mean enthalpies measured from the injection of effectors into the buffer were subtracted from raw titration data prior to data analysis with the MicroCal version of ORIGIN. Data were fitted with the One binding site model of ORIGIN.

Data, Materials, and Software Availability. Alignment, tree, and model data have been deposited in GitHub (80). All other data are included in the manuscript and/or supporting information.

ACKNOWLEDGMENTS. This work was supported by the Spanish Ministry for Science and Innovation/*Agencia Estatal de Investigación* 10.13039/501100011033 (grant PID2020-112612GB-I00 to T.K.) and the Junta de Andalucía (grant P18-FR-1621 to T.K.) and by the US National Institutes of Health (grant R35GM131760 to I.B.Z.). J.P.C.-V. was supported by the grant Unión Europea-NextGenerationEU RD 289/2021 UPM-Recualifica Margarita Salas.

Author affiliations: [a]Department of Biotechnology and Environmental Protection, Estación Experimental del Zaidín, Consejo Superior de Investigaciones Científicas, Granada 18008, Spain; [b]Centro de Biotecnología y Genómica de Plantas, Universidad Politécnica de Madrid Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria/Consejo Superior de Investigaciones Científicas, Parque Científico y Tecnológico de la Universidad Politécnica de Madrid, Pozuelo de Alarcón, Madrid 28223, Spain; and [c]Department of Microbiology and Translational Data Analytics Institute, The Ohio State University, Columbus, OH 43210

1. V. M. Gumerov, D. R. Ortega, O. Adebali, L. E. Ulrich, I. B. Zhulin, MiST 3.0: An updated microbial signal transduction database with an emphasis on chemosensory systems. *Nucleic Acids Res.* **48**, D459–D464 (2020).
2. M. Y. Galperin, What bacteria want. *Environ. Microbiol.* **20**, 4221–4229 (2018).
3. A. Ortega, I. B. Zhulin, T. Krell, Sensory repertoire of bacterial chemoreceptors. *Microbiol. Mol. Biol. Rev.* **81**, e00033-17 (2017).
4. L. E. Ulrich, E. V. Koonin, I. B. Zhulin, One-component systems dominate signal transduction in prokaryotes. *Trends Microbiol.* **13**, 52–56 (2005).
5. T. Krell, Tackling the bottleneck in bacterial signal transduction research: High-throughput identification of signal molecules. *Mol. Microbiol.* **96**, 685–688 (2015).
6. J. A. Gavira *et al.*, How bacterial chemoreceptors evolve novel ligand specificities. *mBio* **11**, e03066-19 (2020).

7. A. A. Upadhyay, A. D. Fleetwood, O. Adebali, R. D. Finn, I. B. Zhulin, Cache domains that are homologous to, but different from PAS domains comprise the largest superfamily of extracellular sensors in prokaryotes. *PLoS Comput. Biol.* **12**, e1004862 (2016).
8. V. M. Gumerov *et al.*, Amino acid sensor conserved from bacteria to humans. *Proc. Natl. Acad. Sci. USA* **119**, e2110415119 (2022).
9. Y. F. Zhou *et al.*, C4-dicarboxylates sensing mechanism revealed by the crystal structures of DctB sensor domain. *J. Mol. Biol.* **383**, 49–61 (2008).
10. R. Wu *et al.*, Insight into the sporulation phosphorelay: Crystal structure of the sensor domain of *Bacillus subtilis* histidine kinase, KinD. *Protein Sci.* **22**, 564–576 (2013).
11. C. J. Day *et al.*, A direct-sensing galactose chemoreceptor recently evolved in invasive strains of *Campylobacter jejuni*. *Nat. Commun.* **7**, 13206 (2016).
12. L. Zhang *et al.*, Sensing of autoinducer-2 by functionally distinct receptors in prokaryotes. *Nat. Commun.* **11**, 5371 (2020).
13. A. Boyeldieu, A. Ali Chaouche, V. Mejean, C. Jourlin-Castelli, Combining two optimized and affordable methods to assign chemoreceptors to a specific signal. *Anal. Biochem.* **620**, 114139 (2021).
14. M. Fernandez, B. Morel, A. Corral-Lugo, T. Krell, Identification of a chemoreceptor that specifically mediates chemotaxis toward metabolizable purine derivatives. *Mol. Microbiol.* **99**, 34–42 (2016).
15. A. Corral-Lugo *et al.*, High-affinity chemotaxis to histamine mediated by the TlpQ chemoreceptor of the human pathogen *Pseudomonas aeruginosa*. *mBio* **9**, e01894–18 (2018).
16. J. A. Gavira *et al.*, Structural basis for polyamine binding at the dCACHE bomain of the McpU chemoreceptor from *Pseudomonas putida*. *J. Mol. Biol.* **430**, 1950–1963 (2018).
17. M. A. Matilla *et al.*, Chemotaxis of the human pathogen *Pseudomonas aeruginosa* to the neurotransmitter acetylcholine. *mBio* **13**, e0345821 (2022).
18. B. A. Webb *et al.*, *Sinorhizobium meliloti* chemotaxis to quaternary ammonium compounds is mediated by the chemoreceptor McpX. *Mol. Microbiol.* **103**, 333–346 (2017).
19. V. Michel, Z. Yuan, S. Ramsubir, M. Bakovic, Choline transport for phospholipid synthesis. *Exp. Biol. Med. (Maywood)* **231**, 490–504 (2006).
20. M. R. Picciotto, M. J. Higley, Y. S. Mineur, Acetylcholine as a neuromodulator: Cholinergic signaling shapes nervous system function and behavior. *Neuron* **76**, 116–129 (2012).
21. J. Giri, Glycinebetaine and abiotic stress tolerance in plants. *Plant Signal. Behav.* **6**, 1746–1751 (2011).
22. J. A. Meadows, M. J. Wargo, Carnitine in bacterial physiology and metabolism. *Microbiology (Reading)* **161**, 1161–1174 (2015).
23. A. R. Strom, J. A. Olafsen, H. Larsen, Trimethylamine oxide: A terminal electron acceptor in anaerobic respiration of bacteria. *J. Gen. Microbiol.* **112**, 315–320 (1979).
24. S. Schorn *et al.*, Diverse methylotrophic methanogenic archaea cause high methane emissions from seagrass meadows. *Proc. Natl. Acad. Sci. USA* **119**, e2106628119 (2022).
25. N. Gaci, G. Borrel, W. Tottey, P. W. O'Toole, J. F. Brugere, Archaea and the human gut: New beginning of an old story. *World J. Gastroenterol.* **20**, 16062–16078 (2014).
26. M. Shrestha *et al.*, Structure of the sensory domain of McpX from *Sinorhizobium meliloti*, the first known bacterial chemotactic sensor for quaternary ammonium compounds. *Biochem. J.* **475**, 3949–3962 (2018).
27. J. C. Ma, D. A. Dougherty, The cation-π interaction. *Chem. Rev.* **97**, 1303–1324 (1997).
28. D. T. Infield *et al.*, Cation-π interactions and their functional roles in membrane proteins. *J. Mol. Biol.* **433**, 167035 (2021).
29. Z. Zhang, W. A. Hendrickson, Structural characterization of the predominant family of histidine kinase sensor domains. *J. Mol. Biol.* **400**, 335–353 (2010).
30. P. Westermann, B. K. Ahring, R. A. Mah, Acetate production by methanogenic bacteria. *Appl. Environ. Microbiol.* **55**, 2257–2261 (1989).
31. K. Fiebig, G. Gottschalk, Methanogenesis from choline by a coculture of *Desulfovibrio* sp. and *Methanosarcina barkeri*. *Appl. Environ. Microbiol.* **45**, 161–168 (1983).
32. A. Z. Rosenthal, E. G. Matson, A. Eldar, J. R. Leadbetter, RNA-seq reveals cooperative metabolic interactions between two termite-gut spirochete species in co-culture. *ISME J.* **5**, 1133–1142 (2011).
33. E. V. Pikuta *et al.*, *Williamwhitmania taraxaci* gen. nov., sp. nov., a proteolytic anaerobe with a novel type of cytology from Lake Untersee in Antarctica, description of *Williamwhitmaniaceae* fam. nov., and emendation of the order *Bacteroidales* Krieg 2012. *Int. J. Syst. Evol. Microbiol.* **67**, 4132–4145 (2017).
34. W. M. Chen, T. Y. Chen, C. C. Yang, S. Y. Sheu, *Oleiharenicola lentus* sp. nov., isolated from irrigation water. *Int. J. Syst. Evol. Microbiol* **70**, 3440–3448 (2020).
35. M. Y. Galperin, V. Brover, I. Tolstoy, N. Yutin, Phylogenomic analysis of the family Peptostreptococcaceae (Clostridium cluster XI) and proposal for reclassification of Clostridium litorale (Fendrich et al. 1991) and Eubacterium acidaminophilum (Zindel et al. 1989) as Peptoclostridium litorale gen. nov. comb. nov. and Peptoclostridium acidaminophilum comb. nov. *Int. J. Syst. Evol. Microbiol.* **66**, 5506–5513 (2016).
36. M. A. Cruz-Hernandez, A. Mendoza-Herrera, V. Bocanegra-Garcia, G. Rivera, *Azospirillum* spp. from plant growth-promoting bacteria to their use in bioremediation. *Microorganisms* **10**, 1057 (2022).
37. K. M. Lee *et al.*, Activation of cholera toxin production by anaerobic respiration of trimethylamine N-oxide in *Vibrio cholerae*. *J. Biol. Chem.* **287**, 39742–39752 (2012).
38. Y. Shivani, Y. Subhash, C. Sasikala, C. V. Ramana, Characterisation of a newly isolated member of a candidatus lineage, *Marispirochaeta aestuarii* gen. nov., sp. nov. *Int. J. Syst. Evol. Microbiol.* **67**, 3929–3936 (2017).
39. K. Veit, C. Ehlers, R. A. Schmitz, Effects of nitrogen and carbon sources on transcription of soluble methyltransferases in *Methanosarcina mazei* strain Go1. *J. Bacteriol.* **187**, 6147–6154 (2005).
40. Y. Chen, N. A. Patel, A. Crombie, J. H. Scrivens, J. C. Murrell, Bacterial flavin-containing monooxygenase is trimethylamine monooxygenase. *Proc. Natl. Acad. Sci. USA* **108**, 17791–17796 (2011).
41. M. Taubert *et al.*, Methylamine as a nitrogen source for microorganisms from a coastal marine environment. *Environ. Microbiol.* **19**, 2246–2257 (2017).
42. V. V. Roshchina, New trends and perspectives in the evolution of neurotransmitters in microbial, plant, and animal Cells. *Adv. Exp. Med. Biol.* **874**, 25–77 (2016).
43. R. Sharifi, C. M. Ryu, Sniffing bacterial volatile compounds for healthier plants. *Curr. Opin. Plant Biol.* **44**, 88–97 (2018).
44. C. Welte, U. Deppenmeier, Bioenergetics and anaerobic respiratory chains of aceticlastic methanogens. *Biochim. Biophys. Acta* **1837**, 1130–1147 (2014).
45. G. M. King, Metabolism of trimethylamine, choline, and glycine betaine by sulfate-reducing and methanogenic bacteria in marine sediments. *Appl. Environ. Microbiol.* **48**, 719–725 (1984).
46. R. J. Mackay, C. J. McEntyre, C. Henderson, M. Lever, P. M. George, Trimethylaminuria: Causes and diagnosis of a socially distressing condition. *Clin. Biochem. Rev.* **32**, 33–43 (2011).
47. Z. Wang *et al.*, Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**, 57–63 (2011).
48. I. D. Lidbury, J. C. Murrell, Y. Chen, Trimethylamine and trimethylamine N-oxide are supplementary energy sources for a marine heterotrophic bacterium: Implications for marine carbon and nitrogen cycling. *ISME J.* **9**, 760–769 (2015).
49. S. Craciun, E. P. Balskus, Microbial conversion of choline to trimethylamine requires a glycyl radical enzyme. *Proc. Natl. Acad. Sci. USA* **109**, 21307–21312 (2012).
50. J. M. Thomas, M. Alexander, Microbial formation of secondary and tertiary amines in municipal sewage. *Appl. Environ. Microbiol.* **42**, 461–463 (1981).
51. C. Kratzer, P. Carini, R. Hovey, U. Deppenmeier, Transcriptional profiling of methyltransferase genes during growth of *Methanosarcina mazei* on trimethylamine. *J. Bacteriol.* **191**, 5108–5115 (2009).
52. A. Ellaithy, J. Gonzalez-Maeso, D. A. Logothetis, J. Levitz, Structural and biophysical mechanisms of class C G protein-coupled receptor function. *Trends Biochem. Sci.* **45**, 1049–1064 (2020).
53. L. E. Ulrich, I. B. Zhulin, Four-helix bundle: A ubiquitous sensory module in prokaryotic signal transduction. *Bioinformatics* **21** (Suppl. 3), iii45–iii48 (2005).
54. B. A. Elgamoudi *et al.*, The *Campylobacter jejuni* chemoreceptor Tlp10 has a bimodal ligand-binding domain and specificity for multiple classes of chemoeffectors. *Sci. Signal.* **14**, eabc8521 (2021).
55. K. Borziak, I. B. Zhulin, FIST: A sensory domain for diverse signal transduction pathways in prokaryotes and ubiquitin signaling in eukaryotes. *Bioinformatics* **23**, 2518–2521 (2007).
56. C. Izquierdo, J. C. Gomez-Tamayo, J. C. Nebel, L. Pardo, A. Gonzalez, Identifying human diamine sensors for death related putrescine and cadaverine molecules. *PLoS Comput. Biol.* **14**, e1005945 (2018).
57. C. D. de Mattos *et al.*, Polyamines and linear DNA mediate bacterial threat assessment of bacteriophage infection. *Proc. Natl. Acad. Sci. USA* **120**, e2216430120 (2023).
58. T. Paysan-Lafosse *et al.*, InterPro in 2022. *Nucleic Acids Res.* **51**, D418–D427 (2023).
59. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
60. A. M. Waterhouse, J. B. Procter, D. M. Martin, M. Clamp, G. J. Barton, Jalview Version 2–A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
61. V. M. Gumerov, I. B. Zhulin, TREND: A platform for exploring protein function in prokaryotes based on phylogenetic, domain architecture and gene neighborhood analyses. *Nucleic Acids Res.* **48**, W72–W76 (2020).
62. V. M. Gumerov, I. B. Zhulin, Correction to "TREND: A platform for exploring protein function in prokaryotes based on phylogenetic, domain architecture and gene neighborhood analyses". *Nucleic Acids Res.* **50**, 1795 (2022).
63. O. Adebali, D. R. Ortega, I. B. Zhulin, CDvist: A webserver for identification and visualization of conserved domains in protein sequences. *Bioinformatics* **31**, 1475–1477 (2015).
64. L. Zimmermann *et al.*, A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J. Mol. Biol.* **430**, 2237–2243 (2018).
65. S. Capella-Gutierrez, J. M. Silla-Martinez, T. Gabaldon, trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
66. D. Darriba, G. L. Taboada, R. Doallo, D. Posada, ProtTest 3: Fast selection of best-fit models of protein evolution. *Bioinformatics* **27**, 1164–1165 (2011).
67. F. Ronquist *et al.*, MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
68. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
69. D. Sehnal *et al.*, Mol* Viewer: Modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Res.* **49**, W431–W437 (2021).
70. O. Trott, A. J. Olson, AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).
71. J. J. Irwin, B. K. Shoichet, ZINC–A free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **45**, 177–182 (2005).
72. N. M. O'Boyle *et al.*, Open Babel: An open chemical toolbox. *J. Cheminform.* **3**, 33 (2011).
73. D. H. Parks *et al.*, A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* **36**, 996–1004 (2018).
74. F. Burki, N. Okamoto, J. F. Pombert, P. J. Keeling, The evolutionary history of haptophytes and cryptophytes: Phylogenomic evidence for separate origins. *Proc. Biol. Sci.* **279**, 2246–2254 (2012).
75. M. J. Telford, G. E. Budd, H. Philippe, Phylogenomic insights into animal evolution. *Curr. Biol.* **25**, R876–887 (2015).
76. I. Letunic, P. Bork, Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).
77. A. Krogh, B. Larsson, G. von Heijne, E. L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
78. M. Rico-Jimenez *et al.*, Paralogous chemoreceptors mediate chemotaxis towards protein amino acids and the non-protein amino acid gamma-aminobutyrate (GABA). *Mol. Microbiol.* **88**, 1230–1243 (2013).
79. M. Fernandez *et al.*, High-throughput screening to identify chemoreceptor ligands. *Methods Mol. Biol.* **1729**, 291–301 (2018).
80. V. M. Gumerov, Amine_motif. GitHub. https://github.com/ToshkaDev/Amine_motif. Deposited 24 March 2023.