



Amino acid sensor conserved from bacteria to humans

Vadim M. Gumerov^{a,b}, Ekaterina P. Andrianova^{a,b}, Miguel A. Matilla^c, Karen M. Page^d, Elizabeth Monteagudo-Cascales^c, Annette C. Dolphin^d, Tino Krell^{c,1}, and Igor V. Zhulin (Игорь Жулин)^{a,b,1}

^aDepartment of Microbiology, The Ohio State University, Columbus, OH 43210; ^bTranslational Data Analytics Institute, The Ohio State University, Columbus, OH 43210; ^cDepartment of Environmental Protection, Estación Experimental del Zaidín, Consejo Superior de Investigaciones Científicas, E-18008 Granada, Spain; and ^dDepartment of Neuroscience, Physiology and Pharmacology, University College London, London WC1E 6BT, United Kingdom

Edited by Eugene Koonin, National Center for Biotechnology Information, National Library of Medicine, NIH, Bethesda, MD; received June 4, 2021; accepted January 10, 2022

Amino acids are the building blocks of life, and they are also recognized as signals by various receptors in bacteria, archaea, and eukaryotes. Despite their common basic structure, no universal mechanism for amino acid recognition is currently known. Here, we show that a subclass of dCache_1 (double domain found in calcium channels and chemotaxis receptors, family 1), a ubiquitous extracellular sensory domain, contains a simple motif, which recognizes the amino and carboxyl groups of amino acid ligands. We found this motif throughout the Tree of Life. In bacteria and archaea, this motif exclusively binds amino acids, including γ -aminobutyric acid (GABA), and it is present in all major receptor types. In humans, this motif is found in $\alpha 2\delta$ -subunits of voltage-gated calcium channels that are implicated in neuropathic pain and neurodevelopmental disorders and in a recently characterized CACHD1 protein. Our findings suggest that GABA-derived drugs bind to the same motif in human $\alpha 2\delta$ -subunits that binds natural GABA ligands in bacterial chemoreceptors. The exact location on the target protein and the mechanism of binding may enable future improvements of drugs targeting pain and neurobiological disorders.

signal transduction | evolution | serine/threonine kinases | ion channels | gabapentin

Amino acids are involved in a variety of cellular processes, including signal transduction. They serve as signals for various pathways in both prokaryotes and eukaryotes (1). Extracellular amino acids and their derivatives are recognized by dedicated receptors, such as G protein-coupled receptors (GPCRs) and ligand-gated ion channels in eukaryotes (2, 3) and chemoreceptors in bacteria and archaea (4, 5). In eukaryotes, Class C GPCRs, including γ -aminobutyric acid (GABA) and metabotropic glutamate receptors, bind amino acid ligands at their Venus flytrap domain (6), whereas in ligand-gated ion channels, such as glycine and GABA receptors, amino acid ligands bind to an unrelated β -sandwich-like domain (7). In bacterial chemoreceptors, amino acids are also recognized by unrelated ligand binding domains [e.g., four-helix bundle (8), double all-helical ligand-binding domain (9), and dCache_1 (double domain found in Calcium channels and chemotaxis receptors, family 1) (10)]. No common mechanism of amino acid sensing that would be present in all domains of life is currently known. Here, we identify a simple conserved motif in the dCache_1 domain, which provides a common molecular mechanism for amino acid sensing for different types of receptors across the Tree of Life. The dCache_1 domain is the largest family of the Cache superfamily—ubiquitous extracellular ligand binding sensors in bacteria and archaea that are also found in eukaryotes (11, 12). dCache_1 domains serve as sensory modules in all major types of bacterial and archaeal signal transduction systems (e.g., chemoreceptors, histidine kinases, diguanylate cyclases and phosphodiesterases, serine/threonine kinases and phosphatases), and they are also present in eukaryotic voltage-gated calcium channel (VGCC) $\alpha 2\delta$ -subunits. In bacteria, dCache_1 domains bind various ligands, including amino acids,

sugars, organic acids, and nucleotides (5, 13). Ligands bind to membrane-distal (10, 14, 15) or membrane-proximal (16, 17) modules of the dCache_1 domain, and induced conformational changes travel via the adjacent transmembrane helix (always located C terminally to the dCache_1 domain) to the downstream cytoplasmic signaling domain, resulting in the activation of a kinase (4, 14, 18) or another type of signaling. Archaeal dCache_1 domains likely function similar to those in bacteria, as they are found in the same type of signal transduction proteins; however, their function in eukaryotes remains unknown.

Results and Discussion

Defining the AA motif in Bacterial dCache_1 Domains. In a previous study, we showed that amino acid residues that are involved in binding amino acid ligands by dCache_1 domains from PctABC (pseudomonas chemotaxis transducer-like proteins A, B, and C) chemoreceptors in the bacterium *Pseudomonas aeruginosa* are conserved in many homologous chemoreceptors from gammaproteobacteria (10). By analyzing several sequences of dCache_1 domains from other distantly related bacterial species that are known to bind amino acids, we found that the same positions are conserved in all of them, whereas this conservation is lost in dCache_1 domains that are known to bind ligands other than amino acids (Fig. 1 *A* and *B* and *SI Appendix, Table S1*). Based on this sequence analysis and the location of these residues on available three-dimensional (3D) structures, we propose the consensus amino acid binding motif

Significance

Amino acids are the building blocks of life and important signaling molecules. Despite their common structure, no universal mechanism for amino acid recognition by cellular receptors is currently known. We discovered a simple motif, which binds amino acids in various receptor proteins from all major life-forms. In humans, this motif is found in subunits of calcium channels that are implicated in pain and neurodevelopmental disorders. Our findings suggest that γ -aminobutyric acid-derived drugs bind to the same motif in human proteins that binds natural ligands in bacterial receptors, thus enabling future improvement of important drugs.

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

The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹To whom correspondence may be addressed. Email: tino.krell@eez.csic.es or joulina.1@osu.edu.

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2110415119/-DCSupplemental>.

Published March 1, 2022.

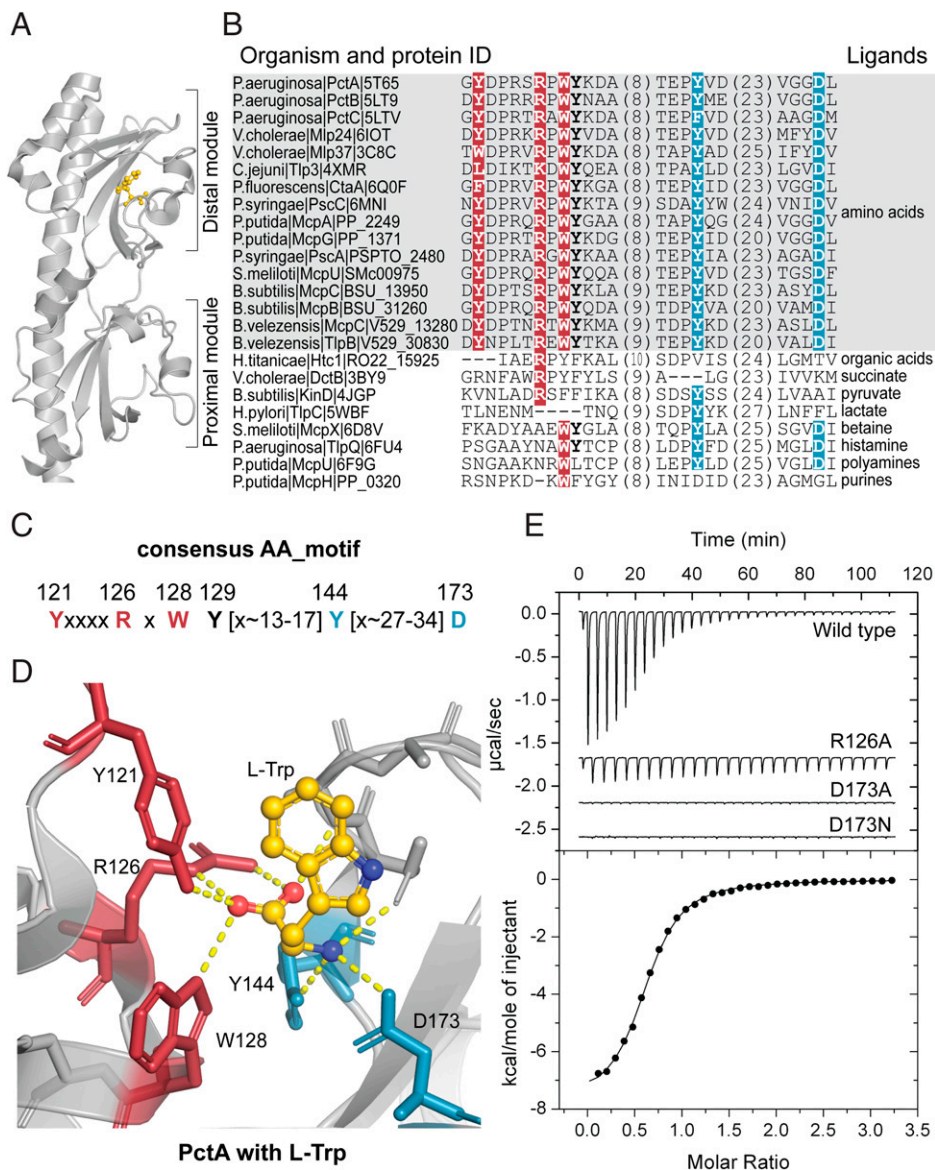


Fig. 1. AA_motif in the dCache_1 domain. (A) dCache_1 domain of the PctA chemoreceptor from *P. aeruginosa* PAO1 with bound L-Trp (gold; PDB ID code 5T7M). (B) Protein sequence alignment of experimentally studied bacterial dCache_1 domains with respective ligands. The AA_motif (in bold) is present in all amino acid binding dCache_1 domains (gray background). (C) Consensus AA_motif. Numbers above the motif correspond to positions in PctA. (D) L-Trp interactions with AA_motif residues in the ligand binding pocket of PctA. (B–D) Here and in all figures, red indicates residues that coordinate the carboxyl group of the ligand, and blue indicates residues that make contacts with the amino group. (E) Isothermal titration calorimetry study of L-Ala binding to the wild type and the mutated dCache_1 domain of PctA.

(AA_motif) in dCache_1 domains (Fig. 1C), which consists of two parts. Y121, R126, and W128 (from here and throughout the text, all motif positions are numbered according to *P. aeruginosa* chemoreceptor PctA, accession no. NP_252999.1) that make key contacts with the carboxyl group of the ligand comprise the N-terminal part of the motif, whereas Y144 and D173 that make key contacts with the amino group of the ligand comprise its C-terminal part (Fig. 1D), as demonstrated for chemoreceptors from *P. aeruginosa* (10), *Campylobacter jejuni* (14), and *Vibrio cholerae* (15). R126, W128, and Y129 were also proposed as conserved determinants of amino acid binding by others (14). The motif integrity is dictated by the structural arrangement of its residues in the folded ligand binding pocket, and in prokaryotes, the distance between the N-terminal and C-terminal parts in a primary sequence is fairly short—13 to 17 amino acid residues (Fig. 1B). To further verify the role of the AA_motif in amino acid binding, we mutated the key residues

in the dCache_1 domain of the *P. aeruginosa* chemoreceptor PctA. The R126A substitution led to a 61-fold decrease in the ligand binding affinity by PctA, whereas the D173A substitution completely abolished ligand binding (Fig. 1E and *SI Appendix, Fig. S1*). Similarly, mutations in these positions in the Tlp3 chemoreceptor in *C. jejuni* and in the Mlp37 chemoreceptor in *V. cholerae* significantly diminished amino acid binding (14, 15). Mutations in other positions of this motif also had a strong negative effect on amino acid binding in other bacterial receptors (*SI Appendix, Table S1*). Consequently, we renamed the AA_motif-containing dCache_1 domains as dCache_1AA.

Identification and Validation of dCache_1AA Domains in Diverse Signaling Proteins from Bacteria and Archaea. The list of dCache_1AA domains known to bind amino acids (Fig. 1B and *SI Appendix, Table S1*) was very short; it contains no archaea or eukaryotes and has representatives of only three bacterial phyla

of more than a hundred phyla defined by the latest bacterial taxonomy (19). Consequently, we searched for the presence of dCache_1AA domains throughout the Tree of Life. First, using the dCache_1 domain profile Hidden Markov Model (HMM; Protein Families Database [Pfam] accession no. PF02743), we scanned proteomes of 31,910 representative bacterial and archaeal genomes from the Genome Taxonomy Database (19) with the HMMER tool (20). This highly specific HMM detected dCache_1 domains in 86,346 protein sequences. Next, we built a multiple sequence alignment (MSA) of these sequences, tracked MSA positions corresponding to the generalized AA_motif definition, and selected sequences that had the preserved AA_motif. The detailed procedure for the identification of dCache_1AA containing proteins is available in *SI Appendix, SI Materials and Methods*. The final dataset contained 10,700 bacterial and 108 archaeal sequences with the dCache_1_AA domain (**Dataset S1**). Taxonomic analysis showed that dCache_1_AA-containing proteins come from representatives of most bacterial and archaeal phyla for which at least 10 high-quality genomes (>90% completeness) were available, indicating their broad phylogenetic distribution (**Dataset S1**). Certain variability within the AA_motif permitting amino acid binding was observed (**Dataset S1**). In addition, we found that ~6% of motif-containing sequences have a D173N substitution. To verify whether such substitution leads to the lack of amino acid binding, we introduced it in *P. aeruginosa* PctA and showed that this change leads to a loss of function (Fig. 1E). To summarize, in this part of the study, we identified dCache_1AA domains in thousands of proteins from the majority of bacterial and archaeal phyla, including important human pathogens, such as *Yersinia pestis*, *V. cholerae*, *Clostridium botulinum*, *Legionella pneumophila*, and *Treponema pallidum*. These domains were found not only in chemoreceptors but also, in all other major receptor proteins, such as sensor histidine kinases, diguanylate cyclases and phosphodiesterases, serine/threonine kinases, and phosphohydrolases (**Dataset S2**).

We then performed two types of validation analyses. First, we modeled structures of dCache_1AA domains from several previously unstudied proteins from our final dataset, representing diverse bacterial and archaeal phyla. All models revealed the presence of a typical dCache_1 fold and the characteristic spatial arrangement of the AA_motif residues (*SI Appendix, Fig. S2*). Second, we performed biochemical assays to demonstrate that dCache_1 domains implicated as amino acid sensors by our computational analyses indeed bind amino acids. To date, all but one known amino acid sensing dCache_1 domain were found in bacterial chemoreceptors that come from representatives of only three bacterial phyla (Fig. 1B and *SI Appendix, Table S1*). Therefore, we selected targets for experimental validation based on the following characteristics: 1) taxonomy, 2) the type of receptor, and 3) the species pathogenicity status (Fig. 2 and *SI Appendix, Table S2*). Ligand binding to recombinant dCache_1 domains was analyzed using differential scanning fluorimetry-based thermal shift assays followed by isothermal titration calorimetry, as previously described (21, 22) (*SI Appendix, SI Materials and Methods*). Satisfactorily, all ligands that bound to selected targets were amino acids (Fig. 2 and *SI Appendix, Figs. S3–S10 and Table S2*). The ligand binding affinity of the archaeal protein was low, whereas all bacterial proteins recognized their ligands with high affinity (K_D values in the nanomolar or lower micromolar concentration range) (Fig. 2) that are typical of functionally characterized bacterial sensor proteins (23). These experiments validated our computational predictions and confirmed that 1) dCache_1AA domains are amino acid sensors and that 2) they are present in major classes of bacterial and archaeal signal transduction proteins, including those from common pathogens. Furthermore, small molecule ligands for bacterial serine/threonine kinases (Fig. 2) were identified.

Search for the AA_motif in Eukaryotes. In eukaryotes, Cache domains were initially identified only in metazoan VGCC $\alpha 2\delta$ -subunits (11), where they were described as “unusual” and “circular permutations.” Later, these were reclassified as dCache_1 domains with “uncertain boundaries” and detected in some other eukaryotic signal transduction proteins (12); however, no ligands were known to bind to these domains. In humans, $\alpha 2\delta$ -subunits are widely expressed in both the central and peripheral nervous systems and are implicated in various disorders, including schizophrenia, bipolar disorder, autism spectrum disorders, epilepsies, and neuropathic pain (24, 25). The $\alpha 2\delta$ -1- and $\alpha 2\delta$ -2-subunits bind GABA-derived drugs gabapentin, pregabalin, and mirogabalin, which are of therapeutic benefit in neuropathic pain conditions (26). Coincidentally, GABA is a natural ligand for dCache_1AA domains of several bacterial chemoreceptors (*SI Appendix, Table S1*); however, it is unknown whether GABA-derived drugs bind to the dCache_1 domain, and the precise location of this domain in $\alpha 2\delta$ is also unknown. In order to find out whether eukaryotic dCache_1 domains might contain the AA_motif, we searched for eukaryotic dCache_1 proteins in several databases (*SI Appendix, SI Materials and Methods*) and tracked the AA_motif positions in corresponding domains, building MSAs. The analysis identified several hundred eukaryotic sequences with the AA_motif (**Dataset S3**), including $\alpha 2\delta$ -subunits and the recently characterized CACHD1 proteins that also modulate VGCCs and are highly expressed in the thalamus, hippocampus, and cerebellum (27, 28). We also found numerous previously unidentified dCache_1 proteins in protozoan lineages (**Dataset S3**). In CACHD1, the AA_motif was mapped to a C-terminal region, where our analysis revealed a eukaryotic version of the dCache_1 domain (Fig. 3 and *SI Appendix, Fig. S11*). No such motif was detected in the dCache_1 domain corresponding to VGCC_ $\alpha 2$ in $\alpha 2\delta$ -subunits (Fig. 3). Surprisingly, we found the N-terminal part of the AA_motif, YxxxRxWY, in the domain currently annotated as VWA_N (a domain located N terminally to the VWA [von Willebrand factor type A] domain; Pfam accession no. PF08399). Subsequent alignment of $\alpha 2\delta$ and CACHD1 with bacterial dCache_1_AA showed that bacterial sequences are well aligned with two regions of $\alpha 2\delta$ and CACHD1 proteins that are separated by the VWA domain (*SI Appendix, Fig. S11*). In the region located downstream of the VWA domain in the $\alpha 2\delta$ -sequence, we identified the C-terminal part of the AA_motif, Y[x~27 to 34]D (Fig. 3 and *SI Appendix, Fig. S11*).

AA_motif in Eukaryotes Is Split in Sequence but Preserved in 3D Structure. We took advantage of the recently published cryogenic electron microscopy (cryo-EM) structure of the rabbit VGCC and its $\alpha 2\delta$ -1-subunit (29) to scrutinize the $\alpha 2\delta$ -structure in light of our findings. A careful assessment of the $\alpha 2\delta$ -1-subunit structure showed the presence of two dCache_1 domains and one VWA domain (Fig. 3). Both dCache_1 domains have a long stalk α -helix, $\alpha 1$, followed by the upper distal and lower proximal modules. Usually in bacteria, the dCache_1 domain is flanked by two transmembrane regions, one preceding the stalk $\alpha 1$ -helix and another following the membrane-proximal module (10). However, the topology tracking of rabbit $\alpha 2\delta$ -1 and MSA analysis showed that the stalk $\alpha 1$ -helix of the C-terminal dCache_1 (termed here as second dCache_1) is not followed by distal and proximal modules but instead, by a stalk $\alpha 1$ -helix of the N-terminal dCache_1 (termed here as first dCache_1) and then, by its partial distal module (Fig. 3E). The next structural element is the VWA domain, which is followed by the remaining part of the first dCache_1 distal module and the proximal module, and then, the sequence proceeds to distal and proximal modules of the second dCache_1. This analysis indicates that the first dCache_1 is

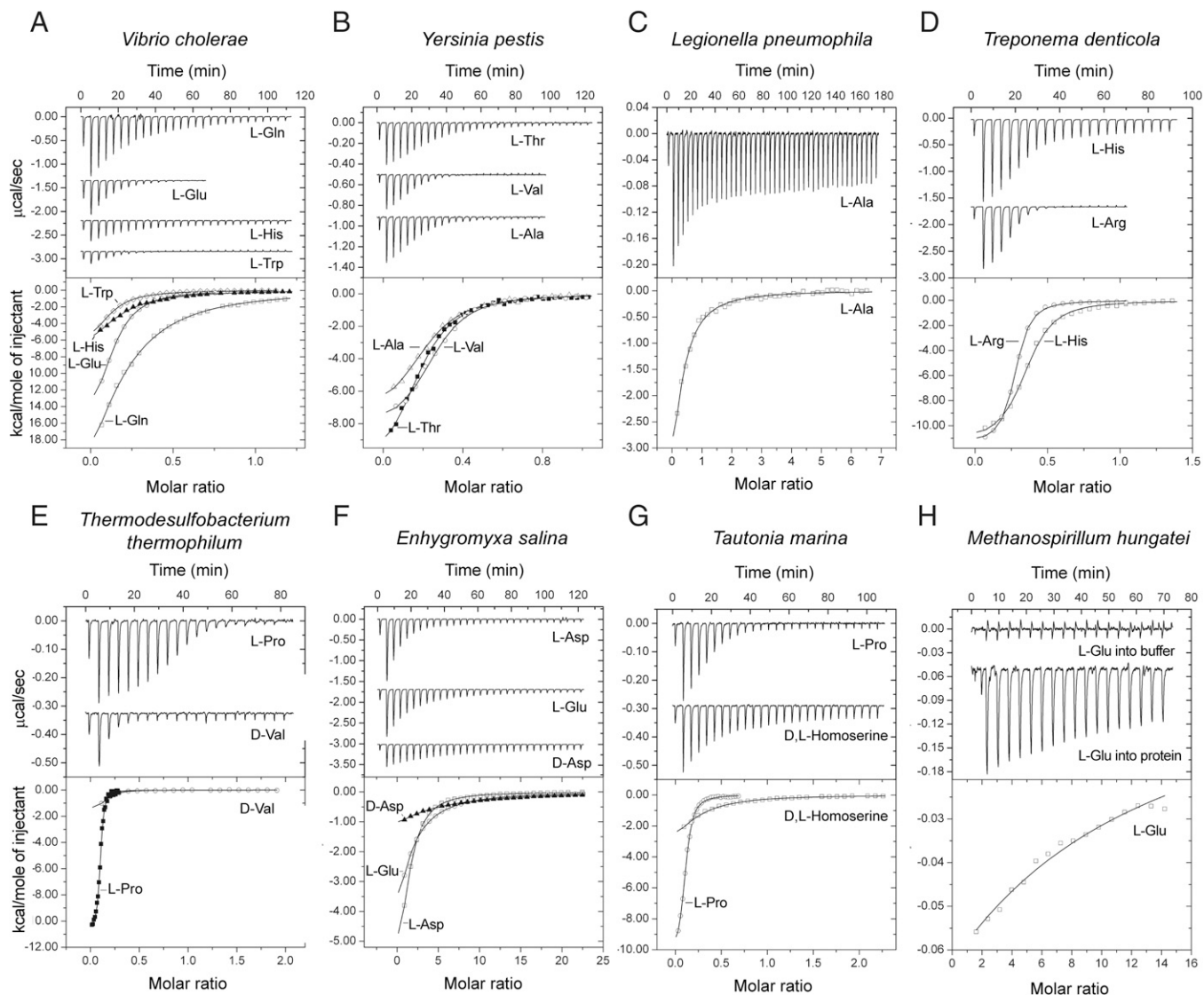


Fig. 2. Microcalorimetric titration of selected recombinant dCache_1AA domains with amino acids. In each panel, *Upper* shows raw titration data, and *Lower* shows integrated corrected peak areas of the titration data fit using the “one-binding site model.” Details of each experiment can be found in [SI Appendix, Table S2](#). (A) *V. cholerae* (gammaproteobacteria) c-di-GMP phosphodiesterase (NP_233280.1). (B) *Y. pestis* (gammaproteobacteria) chemoreceptor (WP_016674185.1). (C) *L. pneumophila* (gammaproteobacteria) guanylate/adenylate cyclase (WP_154766400.1). (D) *Treponema denticola* (spirochaetota) chemoreceptor (WP_002687321.1). (E) *Thermodesulfobacterium thermophilum* (desulfobacterota) c-di-GMP cyclase (WP_162138226.1). (F) *Enhygromyxa salina* (myxococcota) serine/threonine kinase (WP_106093935.1). (G) *Tautonia marina* (planctomycetota) serine/threonine phosphatase (WP_152054232.1). (H) *Methanospirillum hungatei* (archaea, halobacteriota) sensor histidine kinase (WP_011449640.1).

inserted into the loop between $\alpha 1$ -helix and the distal module of the second dCache_1. Furthermore, it confirmed that the VWA domain is inserted into the first dCache_1 domain between $\alpha 4$ and $\beta 4$ and splits the AA_motif. As a result, in a primary sequence, the distance between its N-terminal and C-terminal parts becomes much longer than that in prokaryotes. Remarkably, although the AA_motif in the first dCache_1 is split by the VWA domain insertion, the fold of the distal module is intact, and amino acid residues that constitute the AA_motif N-terminal and C-terminal parts come together in 3D and form the interface matching the one in the PctA chemoreceptor (Fig. 3D and [SI Appendix, Figs. S12 and S13](#)). Excised and concatenated dCache_1 domains of $\alpha 2\delta$ and CACHD1 proteins perfectly align with bacterial dCache_1 domains and find them in Basic Local Alignment Search Tool searches ([Dataset S4](#)). Next, we performed pairwise structural comparison of the dCache_1 domain from the *P. aeruginosa*

PAO1 chemoreceptor PctA with both dCache_1 domains of the rabbit $\alpha 2\delta$ -1-subunit. Remarkably, the PctA dCache_1 domain aligned very well with both dCache_1 domains of the $\alpha 2\delta$ -1-subunit. The alignment shows essentially identical topologies of the distal and proximal modules; however, additional secondary structure elements are present in $\alpha 2\delta$ -1, especially in the second dCache_1 ([SI Appendix, Fig. S12](#)).

$\alpha 2\delta$ -1-Subunit Binds Amino Acid Ligands through the AA_motif.

R241A substitutions in the murine and porcine $\alpha 2\delta$ -1 (corresponding to R126 in the PctA N-terminal part of the AA_motif) were shown to completely abolish the ability to bind pregabalin and gabapentin, respectively (30, 31). Furthermore, the R241A substitution in the murine $\alpha 2\delta$ -1 has been shown to result in a significant decrease in divalent cation current through $Ca_v2.2$ (N-type voltage gated calcium) channels by an effect on channel trafficking (31). Leucine and isoleucine were shown to bind to $\alpha 2\delta$ -1

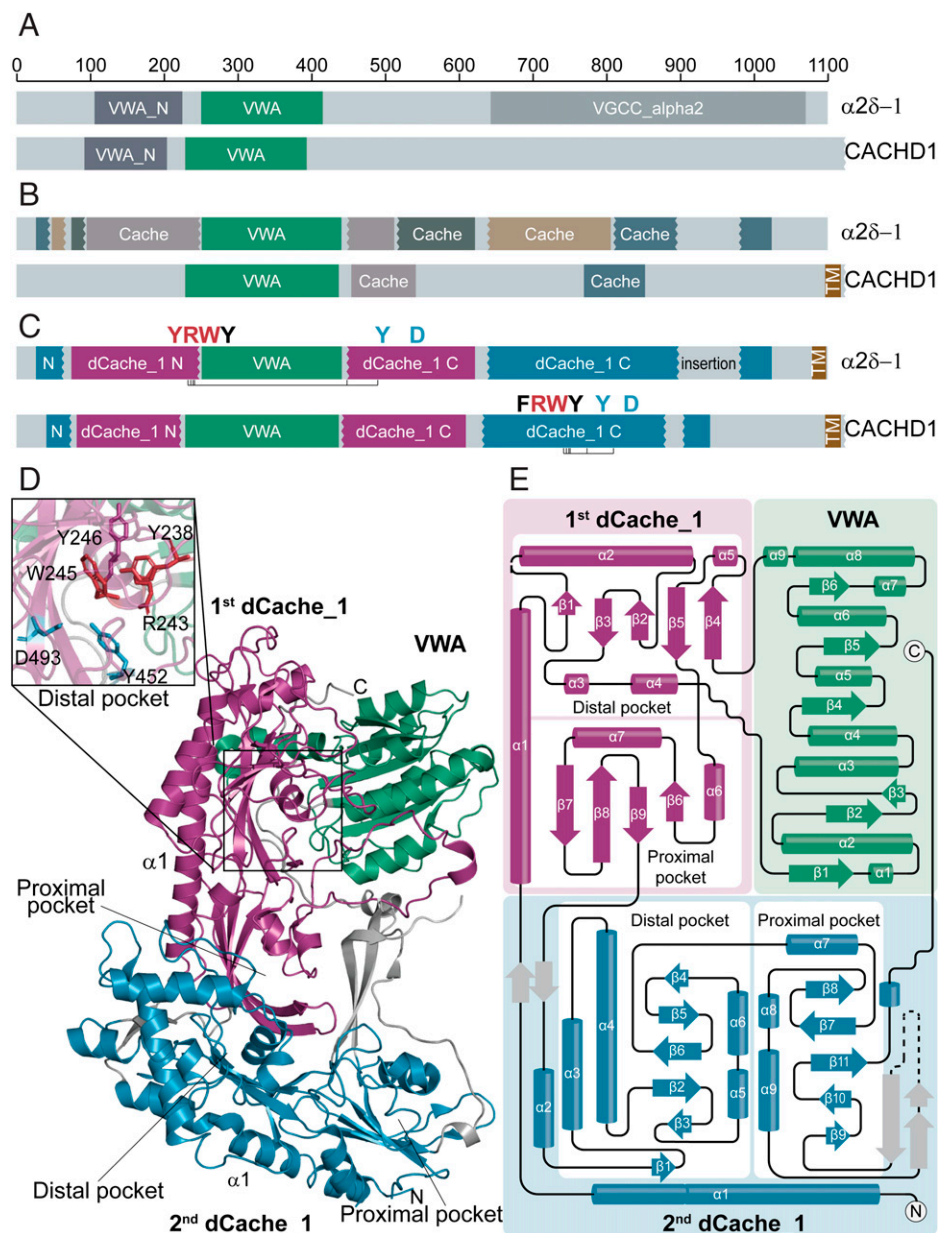


Fig. 3. dCache_1AA domains in $\alpha 2\delta$ -subunit and CACHD1 subunit of VGCC. (A and B) Domains that are currently recognized in $\alpha 2\delta$ -1 and CACHD1 proteins by the Pfam database (A) and experimental studies (27,29) (B). (C) Domain architectures of $\alpha 2\delta$ -1 and CACHD1 proteins revealed in the present study. The AA_motif is shown. (D) Structural composition of the $\alpha 2\delta$ -1-subunit uncovered in the present study shown on the solved structure [PDB ID code 6JPA (44)]. A close-up view of the dCache_1AA distal module (*Upper Left*) showing the spatial proximity of the AA_motif residues despite the VWA insertion. (E) The $\alpha 2\delta$ -1-subunit topology shows that the VWA domain is inserted into the first dCache_1 domain, which in turn, is inserted into the second dCache_1 domain.

(32) and inhibit gabapentin binding by the subunit (33). To further explore whether the AA_motif in $\alpha 2\delta$ -1 might serve as a site for binding GABA-derived drugs and amino acids, we performed computational docking experiments with the available structure of the rabbit $\alpha 2\delta$ -1-protein and 20 α -amino acids, GABA, and drug molecules gabapentin, pregabalin, and mirogabalin. Using the docking results, we calculated polar contacts made between these ligands and $\alpha 2\delta$ -1. We found that all these molecules made contacts with the AA_motif residues (Dataset S5) (the Protein Data Bank [PDB] file is available at <https://github.com/ToshkaDev/Motif>) (34). Most ligands made contacts with the Y and W residues of the N-terminal motif part and with the Y and D residues of its C-terminal part (Dataset S5 and SI Appendix, Fig. S13). The relative affinities agree with the available data. For example, the docking experiments demonstrated that leucine has higher affinity

to $\alpha 2\delta$ -1 than isoleucine (Dataset S5), which is consistent with published experimental data (32, 33). The order of affinities mirogabalin > gabapentin > pregabalin also agrees with recent experimental data (35).

To investigate the effect of mutation of another key position (Asp) in the C-terminal part of the AA_motif in the mammalian $\alpha 2\delta$ -1-protein, we replaced Asp491 (corresponding to D173 in PctA) by Ala in rat $\alpha 2\delta$ -1 and measured its expression at the plasma membrane of tsA-201 cells in the presence of gabapentin. We found that, whereas gabapentin inhibited the wild-type $\alpha 2\delta$ -1 trafficking in tsA-201 cells manifested by reduced expression at the plasma membrane by $43.2 \pm 1.8\%$ (100 μ M gabapentin) and $55.6 \pm 8.4\%$ (1 mM gabapentin), D491A mutation abolished this effect of gabapentin ($10.2 \pm 12.0\%$ inhibition by 1 mM gabapentin) (Fig. 4). Thus, this

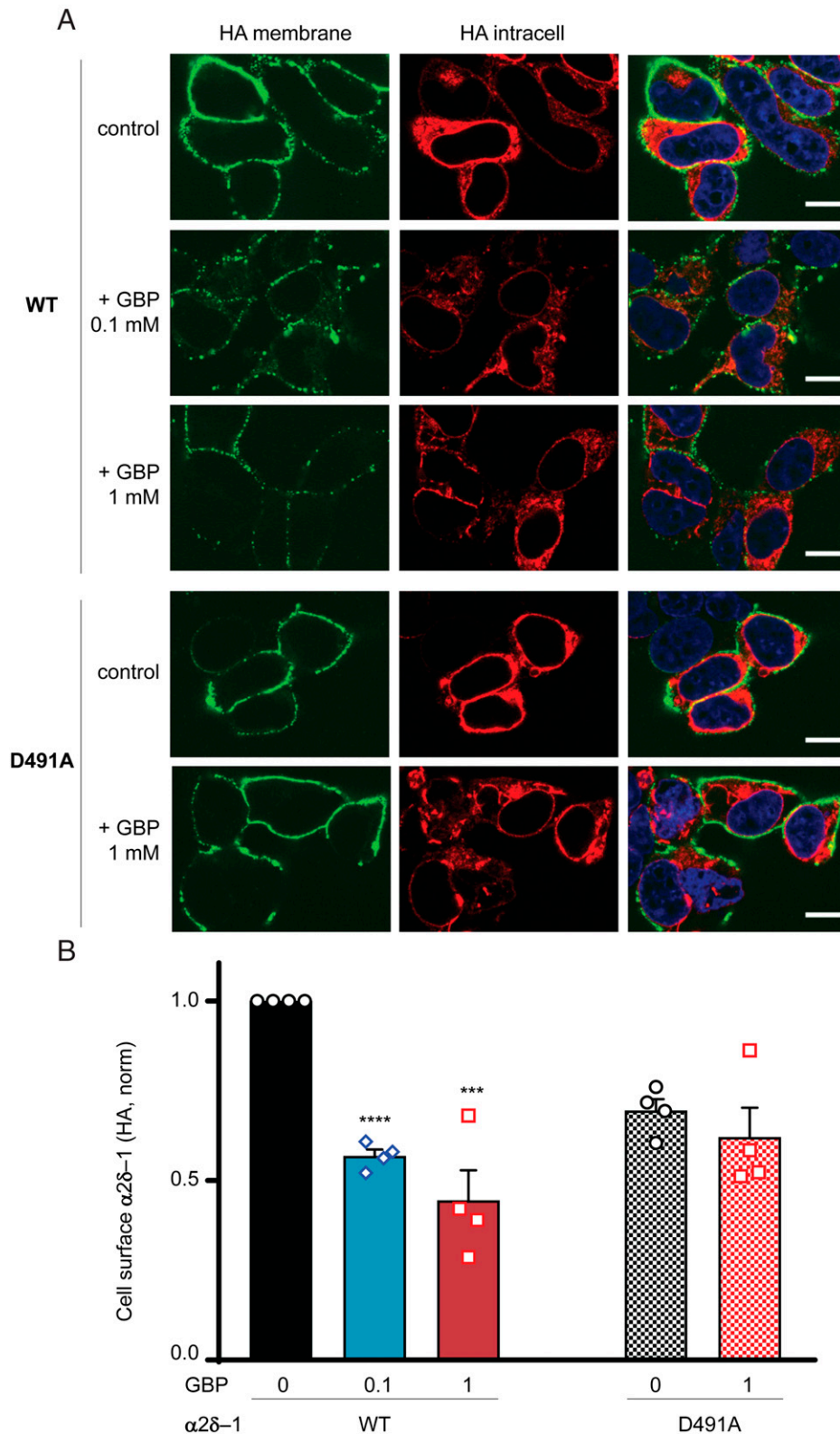


Fig. 4. Gabapentin impairs cell surface expression of wild-type (WT) $\alpha 2\delta - 1$ but not $\alpha 2\delta - 1^{D491A}$. (A) Representative images of tsA-201 cells expressing hemagglutinin (HA) tagged $\alpha 2\delta - 1$ -HA WT (rows 1 to 3) and $\alpha 2\delta - 1^{D491A}$ -HA (rows 4 and 5) in the absence (control; rows 1 and 4) or presence of gabapentin (GBP; 0.1 mM, row 2 [WT]; 1 mM, rows 3 [WT] and 5 [D491A]). *Left* shows cell surface HA staining in nonpermeabilized conditions (green). *Center* shows intracellular HA staining after permeabilization (red). Merged images with the nuclei stained with DAPI (blue) are shown in *Right*. (Scale bars: 10 μ m.) (B) Bar chart (mean \pm SEM for $n = 4$ independent experiments) for cell surface expression of $\alpha 2\delta - 1$ -HA for WT (left three bars) and D491A (right two bars) in the absence (0) or presence of 0.1 or 1 mM GBP. For each experiment, HA staining was measured for over 50 cells and normalized to that of the WT control. Points for individual experiments are shown. Total numbers of cells measured for each condition are as follows; WT control: 379 (black); WT + 0.1 mM GBP: 360 (blue); WT + 1 mM GBP: 449 (red); D491A control: 351 (black hatched); and D491A + 1 mM GBP: 395 (red hatched). Statistical significance of the effect of GBP on cell surface expression was determined using one-way ANOVA and Dunnett's post hoc test. *** $P = 0.0003$; **** $P < 0.0001$.

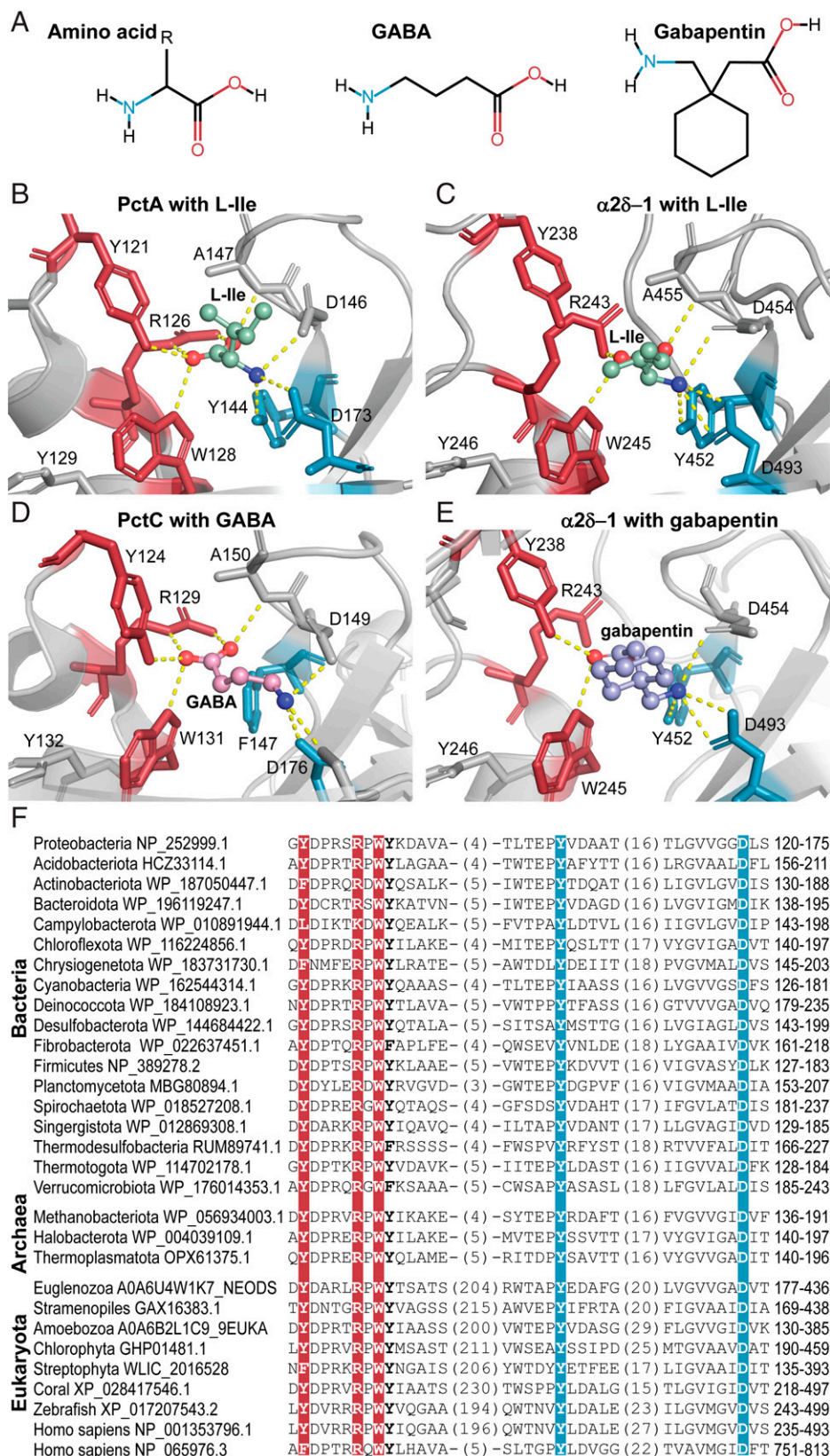


Fig. 5. Bacterial and mammalian receptors bind amino acid ligands through the conserved AA₁ motif. (A) Structural comparison of the ligands found to bind dCache_{1AA}. (B–E) Ligand binding modes of bacterial and eukaryotic dCache_{1AA}: PctA with L-Ile (B; PDB ID code 5T65), $\alpha 2\delta$ -1 with docked L-Ile (C), PctC in complex with GABA (D; PDB ID code 5LTV), and $\alpha 2\delta$ -1 with docked gabapentin (E). (F) Protein sequence alignment of the dCache_{1AA} from representatives of major phyla of Bacteria, Archaea, and Eukaryota.

residue plays a critical role in ligand binding by the eukaryotic $\alpha 2\delta$ -1-protein. We previously found (36) and confirmed here a similar result for the R241A mutation in the N-terminal part of the AA_motif ($4.2 \pm 7.1\%$ inhibition by 1 mM gabapentin).

Human and Bacterial Proteins Bind Ligands in a Similar Fashion. We structurally superimposed the ligand binding module of the first dCache_1 domain of the rabbit $\alpha 2\delta$ -1-subunit with that of the PctA dCache_1 domain (SI Appendix, Fig. S13A). The ligand binding pockets were similar in shape and size, which is astonishing considering the evolutionary time lapsed from bacteria to mammals and the presence of the VWA insertion. Furthermore, the AA_motif residues in the two structures are located at nearly the same positions. Next, we closely examined the rabbit $\alpha 2\delta$ -1-subunit ligand binding pocket with docked L-Ile in comparison with that of the PctA dCache_1 domain in complex with L-Ile (PDB ID code 5T65) (Fig. 5 and SI Appendix, Table S3). We observed that ligands made polar contacts with the AA_motif in these two molecules in a similar fashion. The amino group of L-Ile forms hydrogen bonds with the third Y and last D of the AA_motif, whereas the carboxyl group is bound by R and W in both first dCache_1 domain of $\alpha 2\delta$ -1 and dCache_1 domain of PctA (Fig. 5). The docking experiments also indicated that L-Leu, gabapentin, pregabalin, and mirogabalin are bound through the AA_motif of the $\alpha 2\delta$ -1-subunit following the same pattern (Fig. 5 and SI Appendix, Fig. S13 B–D). The first Y and W of the AA_motif coordinate the ligand carboxyl groups, and the third Y (except for the L-Leu ligand) and D interact with amino groups of the ligands. The carboxyl group of pregabalin is additionally stabilized by a hydrogen bond with R of the AA_motif.

To demonstrate that the motif is capable of binding amino acids and their derivatives in invertebrates, we have run docking experiments using the $\alpha 2\delta$ -protein structure from *Drosophila melanogaster* modeled by AlphaFold (37), amino acid ligands, and their derivatives. The above-described ligand binding pattern was again observed (Dataset S5 and SI Appendix, Fig. S14) (the PDB file is available at <https://github.com/ToshkaDev/Motif>) (34).

CACHD1 and $\alpha 2\delta$ -1 Have Similar Domain Architectures, but Their AA_motifs Are Arranged Differently. We have obtained a structural model of human CACHD1 protein from the AlphaFold Protein Structure Database (<https://alphafold.ebi.ac.uk/>). Structural alignment of the CACHD1 model with the rabbit $\alpha 2\delta$ -1 cryo-EM structure demonstrated that CACHD1 has the same structural composition as the $\alpha 2\delta$ -1-subunit, with a few differences (Fig. 3 and SI Appendix, Fig. S15). Similar to the $\alpha 2\delta$ -1-subunit, CACHD1 consists of two dCache_1 domains, one inserted into the other one and the VWA domain inserted into the first dCache_1 domain. The presence of these structural parts was also confirmed by the MSA (SI Appendix, Fig. S11). However, $\alpha 2\delta$ -proteins possess a distinctive long insertion inside the second dCache_1 between $\alpha 9$ -helix and $\beta 9$ -sheet, whereas it is significantly shorter in CACHD1 protein (SI Appendix, Fig. S11). Another essential difference between $\alpha 2\delta$ and CACHD1 is the AA_motif placement. Unlike $\alpha 2\delta$ -proteins that all carry the AA_motif in the first dCache_1, CACHD1 protein has an intact AA_motif in the distal module of the second dCache_1 (Fig. 3 and SI Appendix, Fig. S11). In order to obtain insights into ligand binding capabilities of CACHD1, we docked amino acids and GABA-derived drug molecules to the distal module of the second dCache_1 domain of the modeled human CACHD1 structure. The ligands made hydrogen bonds with the AA_motif residues following the mentioned pattern (Dataset S5 and SI Appendix, Fig. S14) (the PDB file is available at <https://github.com/ToshkaDev/Motif>); R and W of the N-terminal part of the AA_motif coordinate the ligand carboxyl group, whereas Y and D of the C-terminal part of the motif interact with the amino group of the ligands. Essentially, the same ligand binding pattern was

observed with the *D. melanogaster* CACHD1 protein structure modeled by AlphaFold (Dataset S5 and SI Appendix, Fig. S14) (the PDB file is available at <https://github.com/ToshkaDev/Motif>) (34).

Evolution of the Amino Acid Binding dCache_1 Domain in Eukaryotes. To establish the prevalence of the AA_motif in Eukaryota and its evolutionary history, we analyzed available eukaryotic genomes from several databases (SI Appendix, SI Materials and Methods). We found that dCache_1AA-containing proteins are present in almost all major eukaryotic groups (National Center for Biotechnology Information taxonomy): Euglenozoa, Heterolobosea, SAR (stramenopiles, alveolates, and Rhizaria supergroup), Haptista, Choanoflagellida, Archaeplastida, and Metazoa (Fig. 6 and Dataset S3). We could not detect dCache_1AA proteins in any genomes of angiosperms (flowering plants), fungi, and two protozoan lineages, where they were presumably lost. Domains of the Cache superfamily (Pfam CL0165), to which the dCache_1 belongs, have been shown to have bacterial origins, and their presence in archaea and eukaryotes was attributed to horizontal gene transfer (12). We found that in all eukaryotic proteins that contain dCache_1 domains, the VWA domain is inserted in one of them, which suggests that this insertion probably occurred in the last eukaryotic common ancestor (LECA). We have found that proteins containing two dCache_1 domains, where one (with the VWA insertion) is inserted into another, are present in all branches of eukaryotes. In addition, we identified proteins that contained only one dCache_1 (with the VWA domain insertion). All of these proteins were found in diverse eukaryotes, excluding vertebrates (Fig. 6 and SI Appendix, Fig. S16). Thus, two events happened around LECA: 1) insertion of the VWA domain into a dCache_1 domain and 2) insertion of this dCache_1-VWA module into another dCache_1 domain. Many protists and invertebrates have two types of dCache_1AA domain containing proteins: one with a single dCache_1 domain with the VWA insertion and one with two dCache_1 domains, one of which has the VWA insertion (SI Appendix, Fig. S16). Remarkably, in some members of the Streptophyta clade, the dCache_1AA domain with the VWA insertion is present in a serine/threonine kinase, resembling some of the bacterial dCache_1-containing proteins (Fig. 6).

To further understand evolutionary relationships between the proteins containing two dCache_1 domains, we inferred two phylogenetic trees using maximum likelihood estimation and Bayesian inference, respectively. The trees showed high agreement with each other (available at <https://github.com/ToshkaDev/Motif>) (34). We used the Bayesian tree for the subsequent analysis (SI Appendix, Fig. S17), which revealed two clusters: $\alpha 2\delta$ and CACHD1. The $\alpha 2\delta$ -cluster contains only metazoan sequences, including four $\alpha 2\delta$ -proteins from the human genome. In vertebrates, $\alpha 2\delta$ -1- and $\alpha 2\delta$ -2-sequences form one group, while $\alpha 2\delta$ -3 and $\alpha 2\delta$ -4 form another group. This suggests that a primordial $\alpha 2\delta$ -ancestor in vertebrates duplicated, giving rise to $\alpha 2\delta$ -1/ $\alpha 2\delta$ -2- and $\alpha 2\delta$ -3/ $\alpha 2\delta$ -4-ancestors, and subsequent duplications led to four current paralogs. $\alpha 2\delta$ -proteins in bony fishes have undergone additional duplications (SI Appendix, Figs. S16 and S17).

The CACHD1 cluster includes one protein encoded in the human genome, CACHD1, and proteins from metazoan organisms (SI Appendix, Fig. S17). Vertebrates have only one copy of CACHD1, while organisms preceding vertebrates have paralogous proteins in varying numbers (SI Appendix, Figs. S16 and S17). At the root of the tree and close to it, there are proteins that have the AA_motif preserved in the distal modules of both dCache_1 domains, which suggests that the ancestral proteins had the AA_motif in both dCache_1 domains (also supported by the phyletic distribution pattern) (SI Appendix, Fig. S16). In the course of evolution, the AA_motif was differentially lost in various groups of organisms (SI Appendix, Fig. S17). Eukaryotes prior to vertebrates

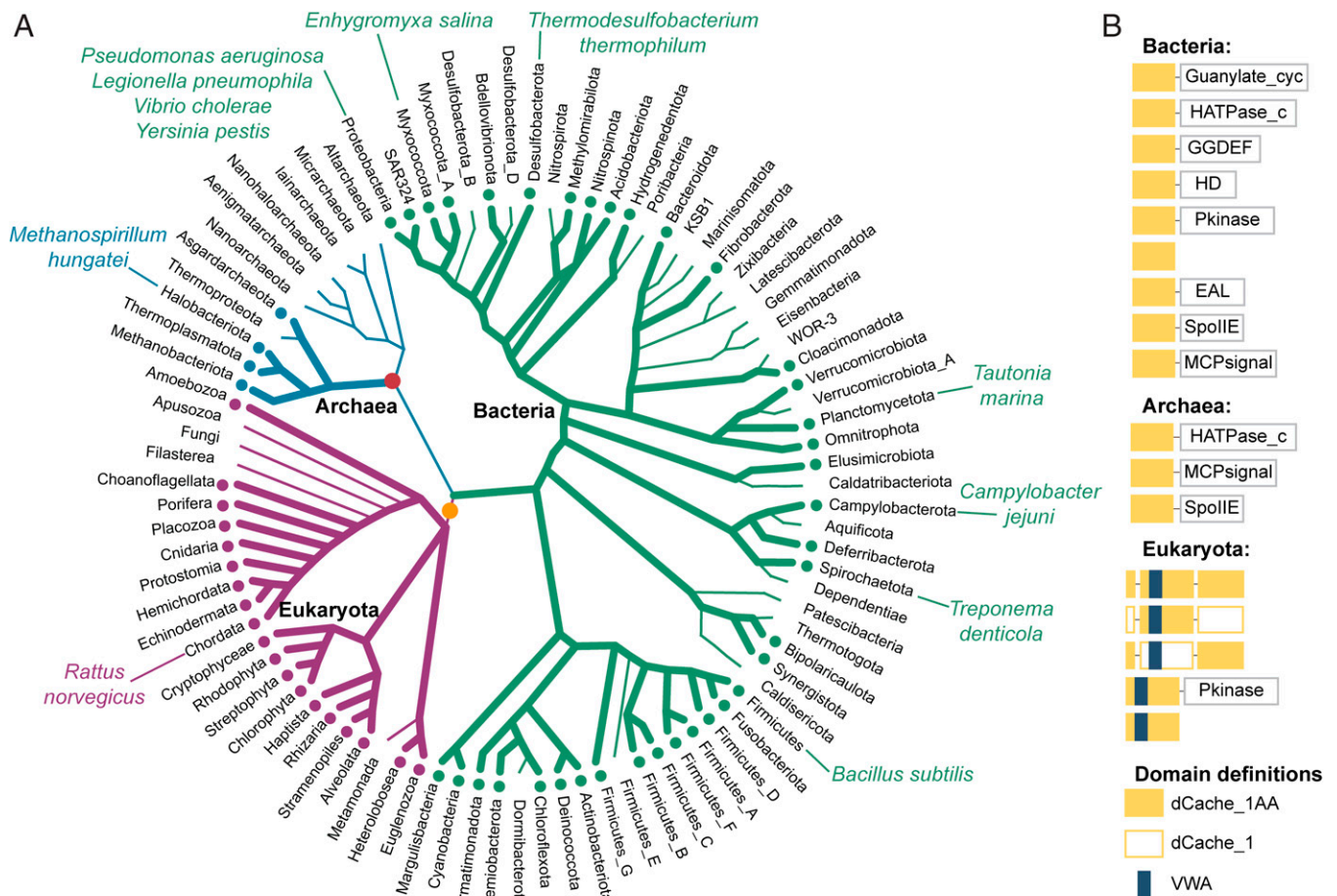


Fig. 6. The AA_motif across the Tree of Life. (A) Distribution of dCache_1AA across major lineages of life. Thick lines with dots at the tips denote the presence of the AA_motif. Positions of relevant organisms are shown. The red circle indicates horizontal gene transfer of the dCache_1AA to Archaea. The orange circle indicates three events that happened around the same time (LECA): 1) horizontal transfer of dCache_1AA from Bacteria to Eukaryota, 2) VWA domain insertion, and 3) insertion of the first dCache_1 into the second dCache_1 domain. (B) Prevalent domain architectures of the dCache_1AA containing proteins found in each domain of life are shown. Domain definitions are according to the Pfam domain nomenclature: EAL (PF00563), a diguanylate phosphodiesterase; GGDEF (PF00990), a diguanylate cyclase; Guanylate_cyc (PF00211), an adenylate or guanylate cyclase; HATPase_c (PF02518), a histidine kinase; HD (PF01966), phosphohydrolase; MCPsignal (PF00015), methyl-accepting chemotaxis protein (chemoreceptor); Pkinase (PF00069), serine/threonine kinase; SpoIIE (PF07228), serine/threonine phosphatase.

have dCache_1AA proteins from all of the described groups: single dCache_1AA proteins, proteins from $\alpha 2\delta$ and CACHD1 clusters, and “double-AA motif” proteins with the AA motif preserved in both domains of double-dCache_1 proteins (SI Appendix, Fig. S16). Phylogenetic reconstruction suggests that duplications of proteins originating from this early double-AA motif group gave rise to $\alpha 2\delta$ and CACHD1 clusters (SI Appendix, Figs. S16 and S17). In $\alpha 2\delta$ -proteins, the AA motif has been preserved predominantly in the first dCache_1 domain, while in CACHD1 proteins, it is in the second dCache_1 domain. Our analysis also demonstrated that the first dCache_1 domain of $\alpha 2\delta$ -subunits is under stronger selective pressure than the second dCache_1; in contrast, both dCache_1 domains of the CACHD1 protein are under strong selective pressure (SI Appendix, SI Materials and Methods). Phylogenetic analysis and sequence similarity searches (Dataset S4) did not allow us to definitively conclude which exact bacterial group could give rise to eukaryotic proteins.

Conclusions

In this work, we have described a universal amino acid binding sensor, which is present throughout the Tree of Life. We showed that these sensors bind amino acid ligands through a simple amino acid recognition motif that has been preserved over 3 billion years and used in all major cellular life-forms. We assign specific biological function—amino acid sensing—to thousands of receptors in

bacteria, archaea, and eukaryotes. It is especially important for human pathogens because amino acids are key mediators of pathogenicity (38). The vast majority of sensor proteins encoded in genomes of various organisms remain unstudied, and signals that they recognize are unknown. Sequence analysis alone does not permit their identification due to extreme sequence variation and complex evolutionary trajectories of sensory domains. On the other hand, structural studies and biochemical characterization can only be performed for a small fraction of sensor proteins identifiable in genome databases. Here, we show how combining these two strategies results in precise predictions for an important class of biological signals. It is likely that similar approaches can be utilized for functional annotation of other classes of sensor proteins. During the course of this study, we identified the AA motif in medically important CACHD1 proteins and $\alpha 2\delta$ -subunits of VGCCs and implicated it as the binding site for GABA-derived drugs in human $\alpha 2\delta$ -subunits. This finding provides opportunities for improving drugs targeting various neurobiological disorders.

Materials and Methods

Public databases and bioinformatics tools used throughout this study are described in detail in SI Appendix. Detailed procedures for identification of AA motif-containing proteins in prokaryotes and eukaryotes are also available in SI Appendix. Computational docking was carried out using AutoDock

Vina (39). MSAs were built using MAFFT (40). Phylogenetic inference was performed using RAxML (41) and MrBayes (42). Identification and analysis of protein domains were performed using TREND (21).

Protein expression, purification, thermal shift assays, isothermal titration calorimetry, and site-specific mutagenesis of bacterial and archaeal proteins were performed using published protocols (22, 43) and are described in detail in *SI Appendix*. Imaging cell surface expression of the calcium channel $\alpha_2\delta$ -1 subunit was carried out as previously described (36), and details are provided in *SI Appendix*.

1. L. Chantranupong, R. L. Wolfson, D. M. Sabatini, Nutrient-sensing mechanisms across evolution. *Cell* **161**, 67–83 (2015).
2. 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).
3. D. Lavery *et al.*, Cryo-EM structure of the human $\alpha 1\beta 3\gamma 2$ GABA_A receptor in a lipid bilayer. *Nature* **565**, 516–520 (2019).
4. J. S. Parkinson, G. L. Hazelbauer, J. J. Falke, Signaling and sensory adaptation in *Escherichia coli* chemoreceptors: 2015 update. *Trends Microbiol.* **23**, 257–266 (2015).
5. Á. Ortega, I. B. Zhulin, T. Krell, Sensory repertoire of bacterial chemoreceptors. *Microbiol. Mol. Biol. Rev.* **81**, e00033-17 (2017).
6. A. Koehl *et al.*, Structural insights into the activation of metabotropic glutamate receptors. *Nature* **566**, 79–84 (2019).
7. J. Du, W. Lü, S. Wu, Y. Cheng, E. Gouaux, Glycine receptor mechanism elucidated by electron cryo-microscopy. *Nature* **526**, 224–229 (2015).
8. M. V. Milburn *et al.*, Three-dimensional structures of the ligand-binding domain of the bacterial aspartate receptor with and without a ligand. *Science* **254**, 1342–1347 (1991).
9. 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).
10. J. A. Gavira *et al.*, How bacterial chemoreceptors evolve novel ligand specificities. *MBio* **11**, e03066-19 (2020).
11. V. Anantharaman, L. Aravind, Cache—a signaling domain common to animal Ca(2+)-channel subunits and a class of prokaryotic chemotaxis receptors. *Trends Biochem. Sci.* **25**, 535–537 (2000).
12. 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).
13. M. A. Matilla, F. Velando, D. Martín-Mora, E. Monteagudo-Cascales, T. Krell, A catalogue of signal molecules that interact with sensor kinases, chemoreceptors and transcriptional regulators. *FEMS Microbiol. Rev.* **46**, fuab043 (2022).
14. Y. C. Liu, M. A. Machuca, S. A. Beckham, M. J. Gunzburg, A. Roujeinikova, Structural basis for amino-acid recognition and transmembrane signalling by tandem Per-Arnt-Sim (tandem PAS) chemoreceptor sensory domains. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 2127–2136 (2015).
15. S. Nishiyama *et al.*, Identification of a *Vibrio cholerae* chemoreceptor that senses taurine and amino acids as attractants. *Sci. Rep.* **6**, 20866 (2016).
16. M. A. Machuca *et al.*, *Helicobacter pylori* chemoreceptor TlpC mediates chemotaxis to lactate. *Sci. Rep.* **7**, 14089 (2017).
17. K. S. Johnson *et al.*, The dCache chemoreceptor TlpA of *Helicobacter pylori* binds multiple attractant and antagonistic ligands via distinct sites. *MBio* **12**, e0181921 (2021).
18. I. Gushchin *et al.*, Mechanism of transmembrane signaling by sensor histidine kinases. *Science* **356**, eaah6345 (2017).
19. 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).
20. S. R. Eddy, Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).
21. 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).
22. M. Rico-Jiménez *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).
23. M. A. Matilla, D. Martín-Mora, T. Krell, The use of isothermal titration calorimetry to unravel chemotactic signalling mechanisms. *Environ. Microbiol.* **22**, 3005–3019 (2020).
24. A. C. Dolphin, Voltage-gated calcium channels and their auxiliary subunits: Physiology and pathophysiology and pharmacology. *J. Physiol.* **594**, 5369–5390 (2016).
25. A. C. Dolphin, Calcium channel auxiliary $\alpha_2\delta$ and β subunits: Trafficking and one step beyond. *Nat. Rev. Neurosci.* **13**, 542–555 (2012).
26. E. P. Calandre, F. Rico-Villademoros, M. Slim, Alpha₂delta ligands, gabapentin, pregabalin and mirogabalin: A review of their clinical pharmacology and therapeutic use. *Expert Rev. Neurother.* **16**, 1263–1277 (2016).
27. G. S. Cottrell *et al.*, CACHD1 is an $\alpha_2\delta$ -like protein that modulates Ca_v3 voltage-gated calcium channel activity. *J. Neurosci.* **38**, 9186–9201 (2018).
28. S. Dahmene *et al.*, The $\alpha_2\delta$ -like protein CACHD1 increases N-type calcium currents and cell surface expression and competes with $\alpha_2\delta$ -1. *Cell Rep.* **25**, 1610–1621.e5 (2018).
29. J. Wu *et al.*, Structure of the voltage-gated calcium channel Ca(v)1.1 at 3.6 Å resolution. *Nature* **537**, 191–196 (2016).
30. M. Wang, J. Offord, D. L. Oxender, T. Z. Su, Structural requirement of the calcium-channel subunit alpha2delta for gabapentin binding. *Biochem. J.* **342**, 313–320 (1999).
31. M. J. Field *et al.*, Identification of the alpha2-delta-1 subunit of voltage-dependent calcium channels as a molecular target for pain mediating the analgesic actions of pregabalin. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 17537–17542 (2006).
32. J. P. Brown, V. U. Dissanayake, A. R. Briggs, M. R. Milic, N. S. Gee, Isolation of the [3H]gabapentin-binding protein/alpha 2 delta Ca²⁺ channel subunit from porcine brain: Development of a radioligand binding assay for alpha 2 delta subunits using [3H]leucine. *Anal. Biochem.* **255**, 236–243 (1998).
33. D. J. Dooley, C. M. Donovan, W. P. Meder, S. Z. Whetzel, Preferential action of gabapentin and pregabalin at P/Q-type voltage-sensitive calcium channels: Inhibition of K⁺-evoked [3H]-norepinephrine release from rat neocortical slices. *Synapse* **45**, 171–190 (2002).
34. V. M. Gumerov *et al.*, “Motif.” GitHub. <https://github.com/ToshkaDev/Motif>. Deposited 18 February 2022.
35. J. Y. Kim, S. Abdi, B. Huh, K. H. Kim, Mirogabalin: Could it be the next generation gabapentin or pregabalin? *Korean J. Pain* **34**, 4–18 (2021).
36. J. S. Cassidy, L. Ferron, I. Kadurin, W. S. Pratt, A. C. Dolphin, Functional exofacially tagged N-type calcium channels elucidate the interaction with auxiliary $\alpha_2\delta$ -1 subunits. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 8979–8984 (2014).
37. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
38. W. Ren *et al.*, Amino acids as mediators of metabolic cross talk between host and pathogen. *Front. Immunol.* **9**, 319 (2018).
39. 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).
40. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
41. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
42. 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).
43. M. Fernández *et al.*, High-throughput screening to identify chemoreceptor ligands. *Methods Mol. Biol.* **1729**, 291–301 (2018).
44. Y. Zhao *et al.*, Molecular basis for ligand modulation of a mammalian voltage-gated Ca²⁺ channel. *Cell* **177**, 1495–1506.e12 (2019).