Amino acid sensor conserved from bacteria to humans

Vadim M. Gumerov[1], Ekaterina P. Andrianova[1], Miguel A. Matilla[2], Annette C. Dolphin[3], Tino Krell[2], and Igor B. Zhulin[1]

[1]Department of Microbiology, The Ohio State University, Columbus, Ohio, 43210, USA

[2]Department of Environmental Protection, Estación Experimental del Zaidín, Consejo Superior de Investigaciones Científicas, E-18008, Granada, Spain.

[3]Department of Neuroscience, Physiology and Pharmacology, University College London, London, WC1E 6BT, United Kingdom.

*Corresponding author. Email: jouline.1@osu.edu

ABSTRACT

Amino acids are recognized as signals by various receptors in bacteria, archaea, and eukaryotes. However, no common mechanism for amino acid recognition is currently known. Here we show that a subclass of a ubiquitous extracellular domain dCache_1 contains a simple amino acid recognition motif, and it is found throughout the Tree of Life. In bacteria, this motif exclusively binds amino acids, including GABA, and it is present in all major receptor types. In humans, this motif is found in α2δ subunits of voltage-gated calcium channels that are implicated in neuropathic pain and neurodevelopmental disorders. Our findings suggest that GABA-derived drugs bind to the same motif in human α2δ subunits that binds natural GABA ligands in bacterial chemoreceptors.

MAIN TEXT

Amino acids are the building blocks of life and 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 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 beta-sandwich-like domain (*7*). In bacterial chemoreceptors, amino acids are also recognized by unrelated ligand-binding domains, e.g. four-helix bundle (*8*) and dCache_1 (*9*). 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 (*10, 11*). 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) α2δ subunits. In bacteria, dCache_1 domains bind various ligands, including amino acids, sugars, organic acids, and nucleotides (*5*), but their function in archaea and eukaryotes is unknown.

In a previous study, we showed that amino acid residues that are involved in binding amino acid ligands by dCache_1 domains from PctABC chemoreceptors in a bacterium *Pseudomonas*

aeruginosa are conserved in many homologous chemoreceptors from gammaproteobacteria (*9*). In the present study we found that the same positions are conserved in all dCache_1 domains that are known to bind amino acids, whereas this conservation is lost in dCache_1 domains that bind ligands other than amino acids (Fig. 1A-B). Based on sequence and structure analysis, we propose the consensus amino acid binding motif (AA_motif) in dCache_1 domains (Fig. 1C, supplementary text, section 1), where Y121, R126 and W128 (from here and throughout the text, all motif positions are numbered according to *P. aeruginosa* chemoreceptor PctA, accession number NP_252999.1) make key contacts with the carboxyl group of the ligand and Y144 and D173 make key contacts with its amino group (Fig. 1D), as demonstrated for chemoreceptors from *P. aeruginosa* (*9*), *Campylobacter jejuni* (*12*), and *Vibrio cholerae* (*13*). R126, W128, and the Y129, were also proposed as conserved determinants of amino acid binding by others (*12*). 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 61-fold decrease in the ligand binding affinity by PctA, whereas the D173A substitution completely abolished ligand binding (Fig. 1E, 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 (*12, 13*). Mutations in other positions of this motif also had a strong negative effect on amino acid binding in other bacterial receptors (table S2). Consequently, we renamed the AA_motif-containing dCache_1 domains to dCache_1AA.

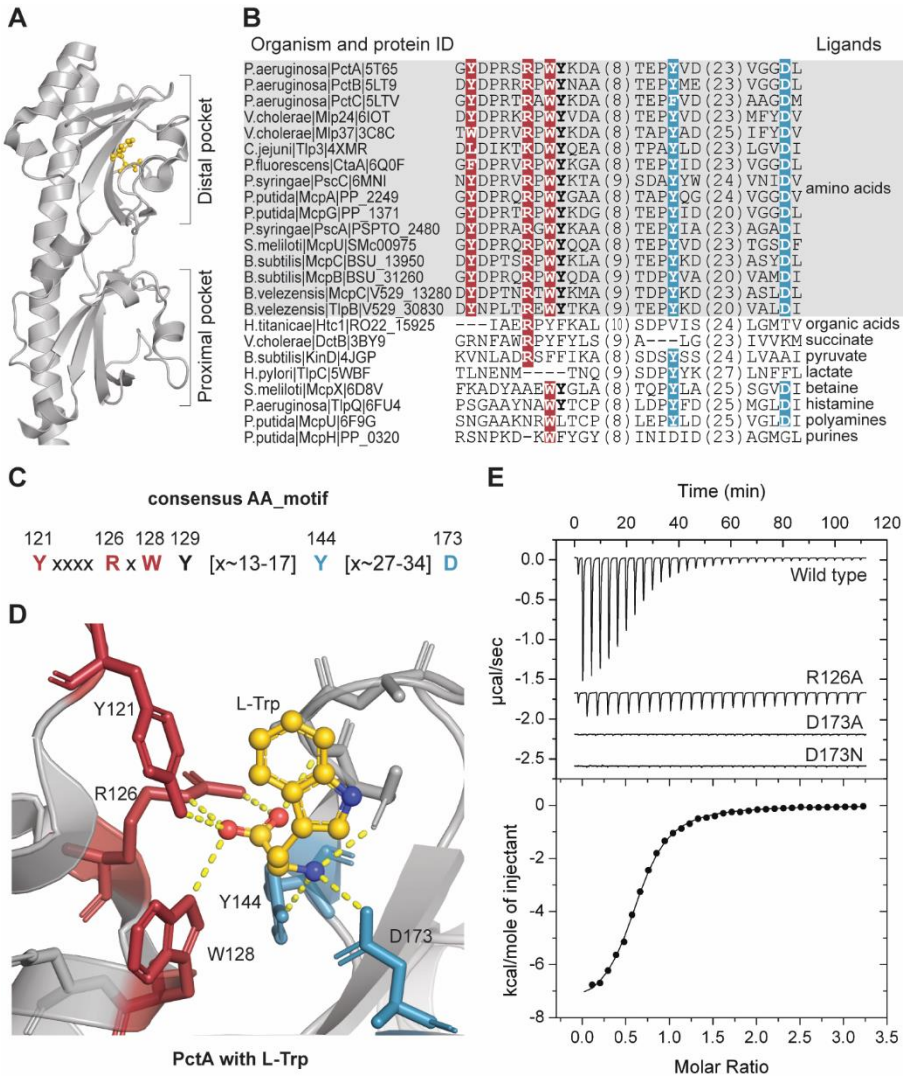

Fig. 1. Amino acid binding motif. (**A**) dCache_1 domain of PctA chemoreceptor from *P. aeruginosa* PAO1 with bound L-Trp (gold) (PDB ID: 5T7M). (**B**) Protein sequence alignment of experimentally studied bacterial dCache_1 domains with respective ligands. AA_motif (in bold) is present in all amino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

The list of dCache_1AA domains that are known to bind amino acids (Fig. 1B and table S2) contains no archaea or eukaryotes and has representatives of only three bacterial phyla

(Proteobacteria, Campylobacterota, and Firmicutes) out of 111 phyla defined by the latest bacterial taxonomy (*14*). Consequently, we searched for the presence of dCache_1AA domains throughout the Tree of Life. First, we used the generalized AA_motif definition to search a dataset of 31,910 representative bacterial and archaeal genomes from the Genome Taxonomy Database (release 95) (*14*) (see Materials & Methods). This search identified the AA_motif in most bacterial and archaeal phyla for which at least ten high-quality genomes (>90% completeness) were available, and in all instances, this motif was located within the dCache_1 domain (table S4). Certain variability within the AA_motif permitting amino acid binding (table S2) was observed primarily in paralogous sequences (supplementary text, section 1). 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, supplementary text, section 2). As a result, we identified dCache_1AA domains in ~11,000 proteins from the majority of bacterial and archaeal phyla (table S4), including important human pathogens, such as *Yersinia pestis*, *Vibrio cholerae, Clostridium botulinum, Legionella pneumophila,* and *Treponema pallidum.* dCache_1AA domains were found not only in chemoreceptors, but also in sensor histidine kinases, diguanylate cyclases and phosphodiesterases, serine/threonine kinases and phosphohydrolases and other proteins involved in signal transduction (table S5).

In the next step, we searched for amino acid sensing dCache_1 domains in eukaryotes. Cache domains were initially identified only in metazoan VGCC α2δ subunits (*10*), where they were described as “unusual”, “circular permutations”. Later, these were re-classified as dCache_1 domains with “uncertain boundaries” and detected in some other eukaryotic signal transduction proteins (*11*), however, no ligands were known to bind to these domains. In humans, α2δ 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 (*15, 16*). The α2δ-1 and α2δ-2 subunits bind GABA derived drugs gabapentin, pregabalin, and mirogabalin, of therapeutic benefit in neuropathic pain conditions (*17*). Coincidentally, GABA is a natural ligand for dCache_1AA domains of several bacterial chemoreceptors (table S2); however, it is unknown whether GABA derived drugs bind to the dCache_1 domain, and the precise location of this domain in α2δ is also unknown. In order to find out whether eukaryotic dCache_1 domains might contain the AA_motif, we performed a search initiated with this motif against several databases (see supplementary materials), which indeed retrieved several hundred eukaryotic sequences including α2δ subunits and the recently characterized CACHD1 proteins that also modulate VGCCs and are highly expressed in the thalamus, hippocampus, and cerebellum (*18, 19*).

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. 2, supplementary text, sections 3, 7). No such motif was detected in the dCache_1 domain corresponding to VGCC_alpha2 in α2δ subunits (Fig. 2). Surprisingly, we found the first part of the AA_motif, YxxxxRxWY, in the domain currently annotated as VWA_N (a domain located N-terminally to the von Willebrand factor type A domain, Pfam accession number PF08399). Subsequent alignment of α2δ and CACHD1 with bacterial dCache_1_AA showed that bacterial sequences are well aligned with two regions of α2δ and CACHD1 proteins that are separated by the VWA domain (fig. S2). In the region located downstream of the VWA domain, we identified the second part of the AA_motif, Y[x~27-34]D (Fig. 2, fig. S2). We took advantage of the recently published structure of the rabbit VGCC and its α2δ-1 subunit (*20*) to scrutinize the α2δ structure in light of these findings. Our structural analyses

and topology tracking (see supplementary text, sections 4 and 7) together with the sequence analysis presented above revealed that α2δ and CACHD1 proteins are comprised of two dCache_1 domains, one inserted into another (Fig. 2, fig. S3, S5). Furthermore, the VWA domain is inserted into the first dCache_1 domain resulting in splitting the AA_motif in α2δ subunits into two parts. Excised and concatenated VGCC dCache_1 domains perfectly align with bacterial dCache_1 domains and find them in simple BLAST searches (table S8). Remarkably, the two parts of the AA_motif that are separated by the VWA domain come together spatially inside the binding pocket of the folded protein (Fig. 2D; fig. S4).

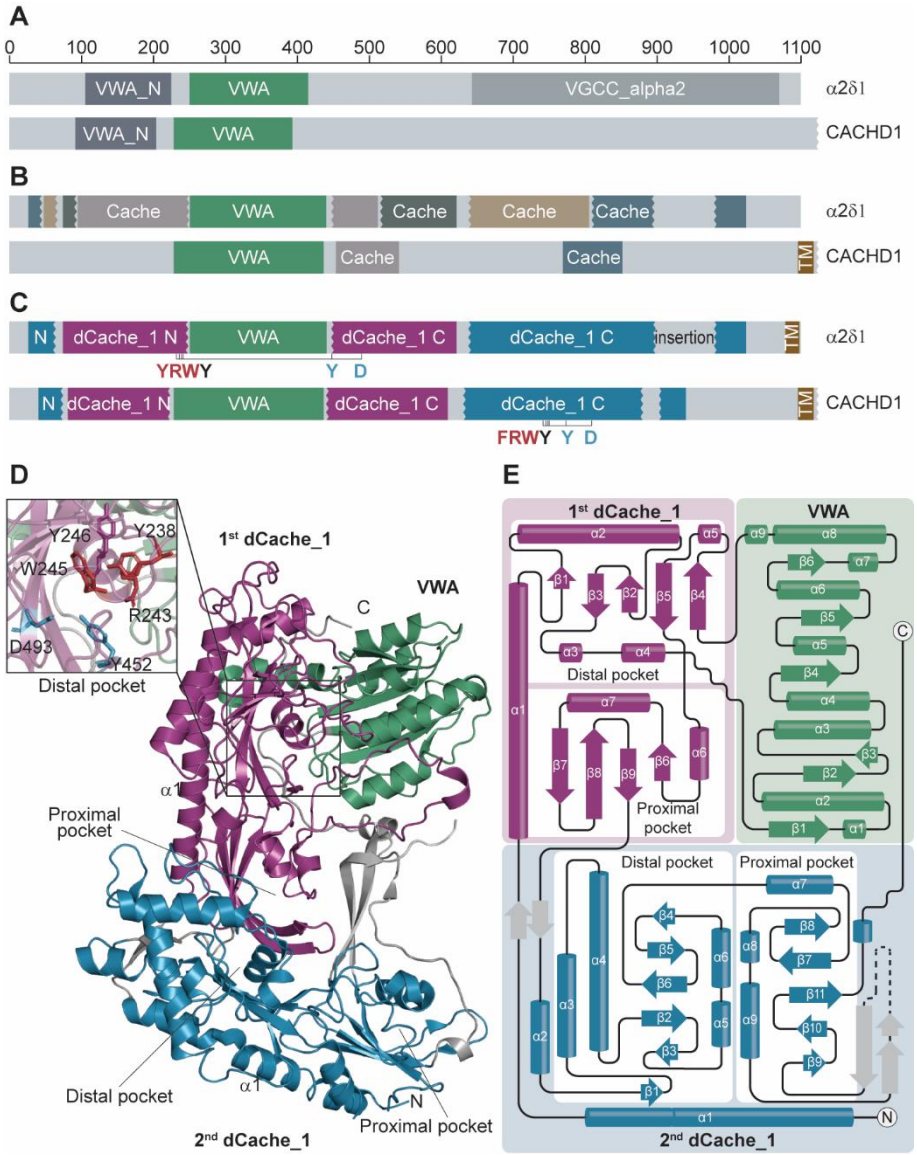

Fig. 2. dCache_1AA domains in α2δ and CACHD1 subunits of VGCC. (**A-B**) Domains that are currently recognized in α2δ-1 and CACHD1 subunits by Pfam database (**A**) and experimental studies (*21*) and (*18*) (**B**). (**C**) Domain architectures of α2δ-1 and CACHD1 subunits revealed in the present study. (**D**) Structural composition of α2δ-1 subunit uncovered in the present study shown on the solved structure (PDB ID: 6JPA (*20*)). A close-up view of the dCache_1AA distal pocket (top left) showing the spatial proximity of the AA_motif residues despite the VWA insertion. (**D**) The α2δ-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.

R241A substitutions in the murine and porcine α2δ-1 (corresponds to R126 in PctA) were shown to completely abolish the ability to bind pregabalin and gabapentin, respectively (*22*, *23*).

Furthermore, the R241A substitution in the murine α2δ-1 has been shown to result in a significant decrease in divalent cation current through CaV2.2 channels (*23*). Leucine and isoleucine were shown to bind to α2δ-1 (*24*) and inhibit gabapentin binding by the subunit (*25*). To further explore whether the AA_motif in α2δ-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 α2δ-1 protein (*20*). Results of computational experiments combined with structural superimpositions showed that α2δ-1 binds the drug molecules and amino acids by the AA_motif and the relative affinities agree with the available experimental data (Fig. 3 and supplementary text, sections 5-6). Position and orientation of the ligands in α2δ-1 and in PctA/PctC are very similar, and the ligands make polar contacts with the amino acid binding motif in a very similar fashion (Fig. 3, fig. S4), which is remarkable considering the evolutionary distance between bacteria and mammals.

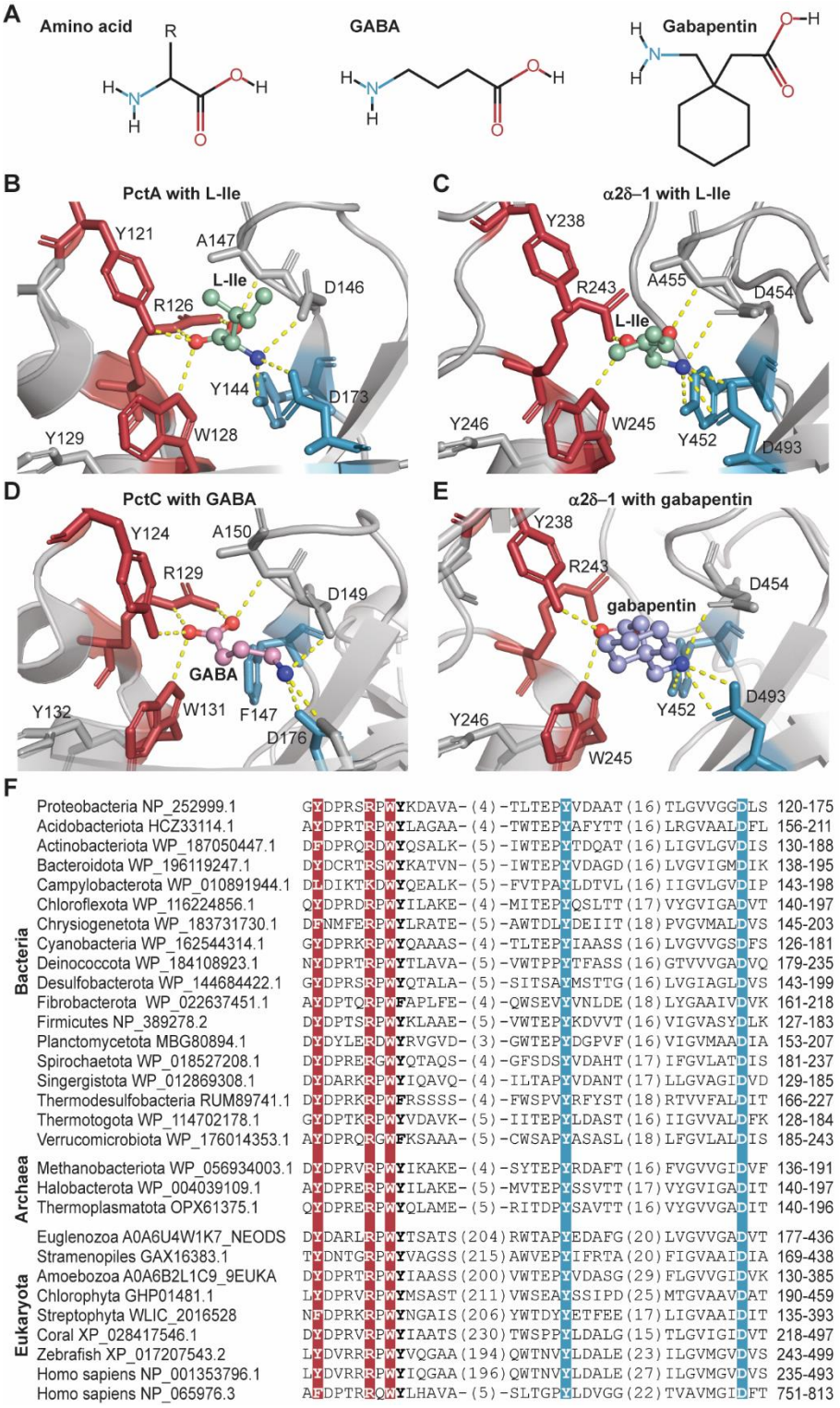

Fig. 3. Both bacterial and mammalian receptors bind amino acid ligands through the conserved AA_motif. (**A**) Structural comparison of the ligands found to bind dCache_1AA. (**B-E**) Ligand binding modes of bacterial and eukaryotic dCache_1AA: PctA with L-Ile (**B**, PDB ID: 5T65), α2δ-1 with docked L-Ile (**C**), PctC in complex with GABA (**D**, PDB ID: 5LTV), and α2δ-

1 with docked gabapentin (**E**). (**F**) Protein sequence alignment of the dCache_1AA from major phyla of Bacteria, Archaea and Eukaryota shows that AA_motif is conserved throughout the Tree of Life.

To establish the prevalence of the motif in Eukaryota and its evolutionary history we analyzed available eukaryotic genomes (supplementary text, section 8). We found that the dCache_1AA containing homologs of α2δ and CACHD1 proteins are universally present in eukaryotes except for flowering plants, fungi, and two protozoan lineages, in which the proteins were presumably lost (Fig. 4, supplementary text, section 8). A dCache_1AA domain containing protein was likely present in the Last Eukaryotic Common Ancestor (LECA) (fig. S7, table S7) conceivably as a result of horizontal gene transfer from bacteria. The VWA domain has been inserted into the first dCache_1 domain as early as LECA (see supplementary text, section 8), while the first dCache_1 domain has been inserted into the second domain in one eukaryotic branch prior to Choanoflagellata divergence. Following this insertion, the protein has undergone multiple duplications in Metazoa and has given rise to four α2δ paralogs and one CACHD1 protein in Vertebrata (fig. S6, S7). Our analysis also demonstrated that the first dCache_1 domain of α2δ subunits is under stronger selective pressure than the second one; in contrast, both dCache_1 domains of CACHD1 subunit are under strong selective pressure (supplementary text, section 9).

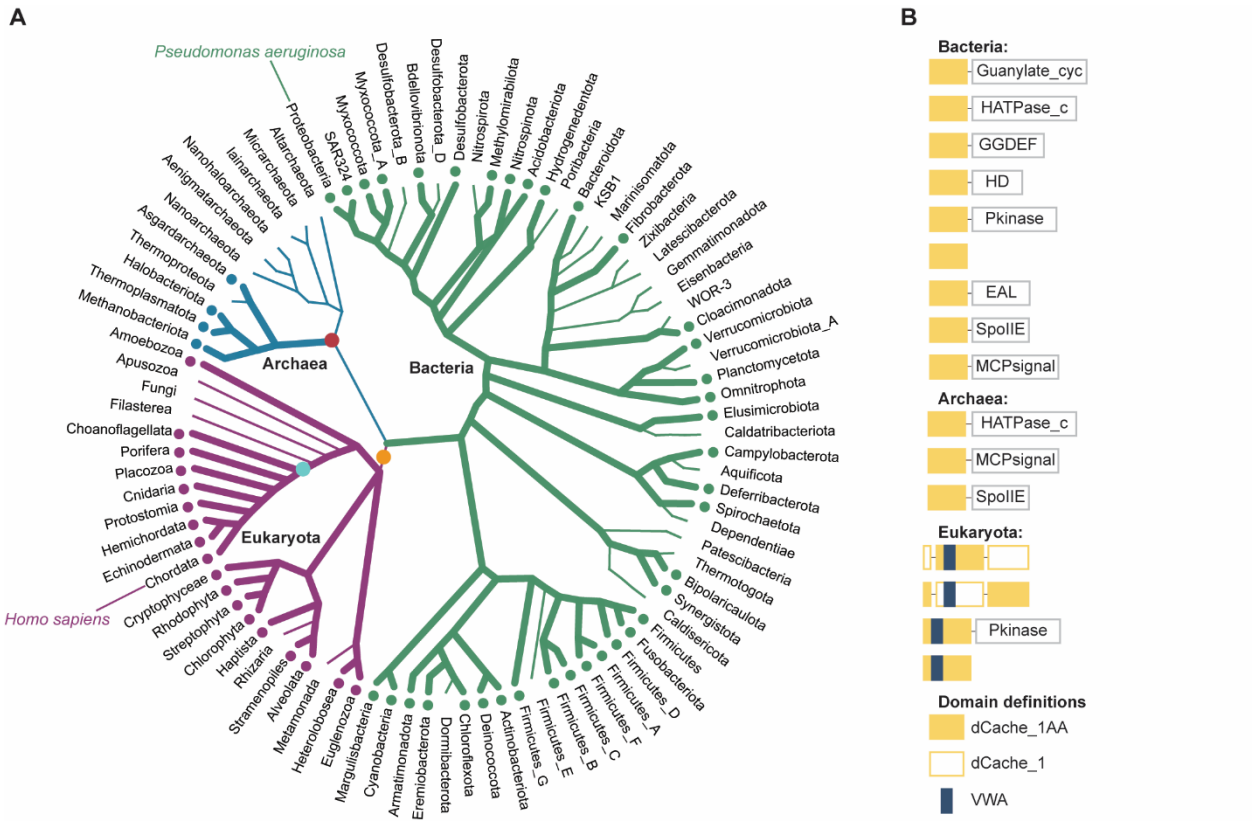

Fig. 4. The AA_motif across the Tree of Life. (**A**) Distribution of the dCache_1AA across major lineages of life. Thick lines with dots at the tips denote the presence of the AA_motif. Red circle indicates horizontal gene transfer of the dCache_1AA to Archaea. Orange circle shows probable event of the horizontal transfer of dCache_1AA to Eukaryota and VWA domain insertion. Circle in teal indicates an event of insertion of 2nd dCache_1 into the 1st dCache_1 domain in eukaryotes. (**B**) Prevalent domain architectures of the dCache_1AA containing proteins found in each domain of life are shown (the Pfam domain nomenclature is used). See supplementary materials for details.

In this work we have described a universal amino acid binding sensor, which is present throughout the Tree of Life. Consequently, we assign specific biological function – amino acid sensing – to thousands of receptors in bacteria and archaea. It is especially important for human pathogens because amino acids are key mediators of pathogenicity (*26*). We identified the amino acid binding motif in α2δ and CACHD1 subunits of voltage-gated calcium channels and implicated

it as the binding site for GABA-derived drugs in human α2δ subunits. This finding provides new opportunities for improving drugs targeting various neurobiological disorders.

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. Laverty *et al.*, Cryo-EM structure of the human alpha1beta3gamma2 GABAA 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 in microbiology*, (2015).
5. A. Ortega, I. B. Zhulin, T. Krell, Sensory Repertoire of Bacterial Chemoreceptors. *Microbiol Mol Biol Rev* **81**, (2017).
6. A. Koehl *et al.*, Structural insights into the activation of metabotropic glutamate receptors. *Nature* **566**, 79-84 (2019).
7. J. Du, W. Lu, 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. J. A. Gavira *et al.*, How Bacterial Chemoreceptors Evolve Novel Ligand Specificities. *mBio* **11**, (2020).
10. 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).
11. 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).
12. 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 crystallographica* **71**, 2127-2136 (2015).
13. S. Nishiyama *et al.*, Identification of a Vibrio cholerae chemoreceptor that senses taurine and amino acids as attractants. *Sci Rep* **6**, 20866 (2016).
14. 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).
15. A. C. Dolphin, Voltage-gated calcium channels and their auxiliary subunits: physiology and pathophysiology and pharmacology. *J Physiol* **594**, 5369-5390 (2016).
16. A. C. Dolphin, Calcium channel auxiliary alpha2delta and beta subunits: trafficking and one step beyond. *Nature reviews. Neuroscience* **13**, 542-555 (2012).
17. 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).
18. G. S. Cottrell *et al.*, CACHD1 is an alpha2delta-Like Protein That Modulates CaV3 Voltage-Gated Calcium Channel Activity. *J Neurosci* **38**, 9186-9201 (2018).

19. S. Dahimene *et al.*, The alpha2delta-like Protein Cachd1 Increases N-type Calcium Currents and Cell Surface Expression and Competes with alpha2delta-1. *Cell Rep* **25**, 1610-1621 e1615 (2018).
20. Y. Zhao *et al.*, Molecular Basis for Ligand Modulation of a Mammalian Voltage-Gated Ca(2+) Channel. *Cell* **177**, 1495-1506 e1412 (2019).
21. J. Wu *et al.*, Structure of the voltage-gated calcium channel Ca(v)1.1 at 3.6 A resolution. *Nature* **537**, 191-196 (2016).
22. M. Wang, J. Offord, D. L. Oxender, T. Z. Su, Structural requirement of the calcium-channel subunit alpha2delta for gabapentin binding. *The Biochemical journal* **342 (Pt 2)**, 313-320 (1999).
23. 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. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 17537-17542 (2006).
24. 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. *Analytical biochemistry* **255**, 236-243 (1998).
25. 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).
26. W. Ren *et al.*, Amino Acids As Mediators of Metabolic Cross Talk between Host and Pathogen. *Front Immunol* **9**, 319 (2018).