Supporting Information

Amino acid sensor conserved from bacteria to humans

Vadim M. Gumerov[1,2], Ekaterina P. Andrianova[1,2], Miguel A. Matilla[2], Karen M. Page[3], Elizabet Monteagudo-Cascales[2], Annette C. Dolphin[3], Tino Krell[2*], and Igor B. Zhulin[1,2*]

[1]Department of Microbiology, The Ohio State University, Columbus, Ohio, 43210, USA
[2]Translational Data Analytics Institute, The Ohio State University, Columbus, Ohio, 43210, USA
[3]Department of Environmental Protection, Estación Experimental del Zaidín, Consejo Superior de Investigaciones Científicas, E-18008, Granada, Spain.
[4]Department of Neuroscience, Physiology and Pharmacology, University College London, London, WC1E 6BT, United Kingdom.

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

This PDF file includes:

SI Materials and Methods
Figures S1 to S17
Tables S1 to S6
Legends for Datasets S1 to S5

Other Supplementary Materials for this manuscript include the following:

Datasets S1 to S5

SI Materials and Methods

Identification of prokaryotic AA_motif-containing proteins.

Protein sequences of representative bacteria and archaea were downloaded from the Genome Taxonomy database (GTDB, Release 95.0) (1). To identify dCache_1 domain containing proteins the retrieved dataset was scanned with the dCache_1 profile Hidden Markov Model (HMM) obtained from the Pfam database (Release 34.0) (2) with the E-value threshold of 0.01 both for sequences and domains. Next, protein sequence regions corresponding to the dCache_1 domain were extracted from the identified sequences and scanned for the presence of the AA_motif in two steps described below. To avoid the identification of false positive amino acid binding proteins we varied positions in the motif not within amino acid physicochemical groups but within those amino acids that were reported to permit amino acids binding (Table S1). The following AA_motif definition was used: [YFWL]….[RK].W[WYF](n1)[YF](n2)D. In this expression, any amino acid inside the given square brackets can be present in the corresponding position, n1 – a varying distance of approximately 13-17 amino acid residues in prokaryotic dCache_1 domains, n2 – a varying distance of approximately 27-34 amino acid residues in prokaryotic dCache_1 domains. In eukaryotic dCache_1 domains that have the von Willebrand factor type A domain (VWA) insertion, n1 is a varying distance of approximately 215-245 amino acid residues.

First, we scanned the dCache_1 domain protein sequences encoded in the genomes of the representative set of bacteria and archaea with the first part of the motif, [YFWL]….[RK].W[WYF], and extracted corresponding sequences. Using these sequences, we have built two MSAs: one – with the bacterial sequences and one – with the archaeal. Exploring the alignments, we identified positions corresponding to the amino acid binding motif. In the bacterial alignment coordinates corresponding to the motif are the following: Y[1160]….R[1278].W[1290]Y[1291](n1)Y[1431](n2)D[1647]. In the archaeal alignment the coordinates are the following: Y[163]….R[168].W[170]Y[171](n1)Y[189](n2)D[219]. Next, using the MSAs and the motif definitions and coordinates we calculated how many sequences have the second part of the motif ((n1)[YF](n2)D): 10700 bacterial and 108 archaeal sequences. Using these

sequences that have the motif in the full form we calculated the motif variant frequencies and abundances. Full protein sequence alignments can be found in https://github.com/ToshkaDev/Motif repository, “GTDB_Representative_Set” folder. Following this step, we retrieved full taxonomy information from the GTDB metadata tables (https://data.ace.uq.edu.au/public/gtdb/data/releases/release95/95.0/) for the genome assemblies in which we found the motif. Additionally, the NCBI RefSeq (Release 202) and Uniprot (Release 2021_02) databases were scanned for the AA_motif containing proteins. Full alignments of dCache_1 domains with the AA_motif from these datasets are in https://github.com/ToshkaDev/Motif repository. We identified 32395 dCache_1 domain containing protein sequences with the AA_motif in RefSeq database and 11330 sequences in Uniprot.

The AA_motif positions in the RefSeq MSA:

Y[2091]....R[2302].W[2385]Y[2414](n1)Y[2727] (n2)D[3326]

The AA_motif positions in the Uniprot MSA:

Y[308]....R[313].W[315]Y[316] (n1)Y[345] (n2)D[420]

We found dCache_1 domains with the AA_motif in 51 bacterial and 4 archaea phyla (Dataset S1, sheets 1 and 2) and counted the AA_motif variants (Table S4; Dataset S1, sheets 3 and 4). The AA_motif permits a certain variability within amino acid groups in each position (Table S1). Tracking the motif variability across genomes we established that the top three variants (Y....R.WY[n1]Y[n2]D, F....R.WY[n1]Y[n2]D, Y....R.WF[n1]Y[n2]D) are present almost universally in motif-containing genomes, while the most abundant consensus motif variant, Y....R.WY[n2]Y[n2]D, is present in 86% of the bacterial and 100% of the archaeal phyla (Dataset S1, sheets 7 and 8). Other variants coexist along with these top variants, mostly in paralogous proteins (Dataset S1, sheets 5 and 6).

Identification of eukaryotic AA_motif-containing proteins.

To identify dCache_1 domain containing proteins with the AA_motif in eukaryotes, human α2δ-1 and CACHD1 protein sequences (protein IDs are NP_001353796.1 and NP_065976.3, respectively) were used as queries for BLASTP and PSI-BLAST searches against NCBI RefSeq, NCBI

Nonredundant, Uniprot (3), and 1KP (4) databases. We verified the presence of the dCache_1 domain by scanning sequences with Pfam profile HMMs and Conserved Domain database position-specific score matrices in TREND (5) using HMMER. In cases when the domain was not recognized, we scanned the sequences using a more sensitive HMM profile-profile search implemented in HHpred (6). In addition, eukaryotic dCache_1 containing proteins were searched in EukProt database (7). The database was downloaded and scanned running HMMER (8) with the *ad hoc* prepared dCache_1AA profile HMM built using eukaryotic proteins. The presence of the motif was verified by constructing multiple sequence alignments with already identified motif-containing eukaryotic and bacterial sequences. Eukaryotic protein multiple sequence alignment can be found in https://github.com/ToshkaDev/Motif repository. Taxonomy information was retrieved from the NCBI Taxonomy database (https://www.ncbi.nlm.nih.gov/taxonomy).

N-terminal dCache_1 domain of α2δ subunits is under stronger selective pressure than the C-terminal dCache_1 domain

We performed BLASTP and PSI-BLAST searches against prokaryotic proteins in GenBank using several eukaryotic dCache_1 containing proteins with the preserved amino acid binding motif. We initiated separate searches with each of two dCache_1 domains of the proteins. In most cases the top hits were Firmicutes and Proteobacteria (Dataset S4). In almost all searches the coverages of identified proteins were limited to the ligand binding pocket of dCache_1 domains (see Dataset S4). Searches initiated with the 1st dCache_1 domain of α2δ proteins resulted in significant hits. Searches with the human protein easily identified bacterial proteins as significant hits with good E-values. In contrast, the 2nd dCache_1 domain of α2δ proteins from the majority of organisms could not identify significant hits when using BLASTP. Only using PSI-BLAST were we able to identify several significant hits.

Unlike α2δ subunit, both dCache_1 domains of CACHD1 proteins easily found prokaryotic proteins in BLASTP searches. Similarly, searches initiated with both dCache_1 domains of proteins from “double-AA_motif” group (Fig. S16) readily identified bacterial proteins with significant E-values (see Dataset S4).

Construction of the Tree of Life.

For the schematic representation of the Tree of Life (Fig. 6) bacterial and archaeal phylogeny was retrieved from the GTDB taxonomy. Phyla with at least 10 genomes were depicted. Eukaryotic phylogeny was adapted from (9) and (10). The overall tree topology is based on (11).

Multiple sequence alignment and domain identification

MSAs were constructed using L-INS-i algorithm of MAFFT (12). Jalview (13) was used to explore and edit the alignments. Domains were identified running TREND (5) 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 (6). Taxonomic trees were retrieved from Annotree (14) and the NCBI Taxonomy database and were edited in iTOL (11).

Phylogeny inference

The eukaryotic protein sequence alignment was edited using an alignment trimming tool, trimAl (15): 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 models for the set of eukaryotic protein sequences was determined running ProtTest (16) and based on Akaike (17) and Bayesian (18) information criteria. The best model was found to be WAG with empirical state frequencies and gamma distribution of rate variation across sites. WAG is an improved model compared to JTT estimated using a combination of counting and maximum likelihood approaches applied to a large database of globular proteins (19).

Using the determined amino acid replacement model phylogenetic trees were inferred using maximum likelihood estimation implemented in RaXML (20) with 500 bootstrap replicates and Bayesian inference implemented in MrBayes (21). Metropolis-coupled Markov chain Monte Carlo simulation implemented in MrBayes was run with 3 heated and 1 cold chain and discarding first 25% samples from

the cold chain at the “burn-in” phase. 4300000 generations were run till the convergence condition was satisfied with chain sampling every 500 generations.

Computational Docking.

AutoDock Vina (22) was used for computational docking experiments. Rabbit α2δ-1 subunit structure was obtained from the rabbit voltage gated calcium channel structure (PDB ID: 6JPA, (23)). The human CACHD1 protein (UniProt Id: Q5VU97, RefSeq Id: NP_0659976.3) and *D. melanogaster* α2δ (UniProt Id: A0A0B4K866, RefSeq Id: NP_001246303.1) and CACHD1 (UniProt Id: Q5BI42, RefSeq Id: NP_611469.1) protein structures were obtained from AlphaFold Protein Structure Database (https://alphafold.ebi.ac.uk/) (24). The protein structures were prepared using MGLTools (25). For the experiments we downloaded ligands from the Zink database (26) in mol2 format and prepared them for the analysis using the Open Babel toolbox (27) and custom shell script. The docking was performed with the search exhaustiveness 8.

Settings used to run the docking simulations with the **rabbit α2δ-1 protein**:

a) coordinates of the center of the simulation box (Angstroms):

X: 198.681; Y: 165.755; Z: 221.688

b) the box dimensions (Angstroms):

X: 22; Y: 22; Z: 24

Settings used to run the docking simulations with the **human CACHD1 protein**:

a) coordinates of the center of the simulation box (Angstroms):

X: -26.151; Y: -7.099; Z: 26.905

b) the box dimensions (Angstroms):

X: 26; Y: 26; Z: 24

Settings used to run the docking simulations with the ***D. melanogaster* α2δ-1 protein**:

a) coordinates of the center of the simulation box (Angstroms):

X: -16.816; Y: -3.349; Z: 2.921

b) the box dimensions (Angstroms):

X: 22; Y: 22; Z: 24

Settings used to run the docking simulations with the ***D. melanogaster*** **CACHD1 protein**:

a) coordinates of the center of the simulation box (Angstroms):

X: -17.011; Y: -5.243; Z: 9.296

b) the box dimensions (Angstroms):

X: 24; Y: 24; Z: 22

Protein Structure Manipulations.

The human CACHD1 protein and *D. melanogaster* α2δ and CACHD1 proteins were modeled based on the solved rabbit α2δ-1 structure by AlphaFold (24). In addition, multiple dCache_1AA domains from major bacterial and archaeal phyla were modeled using Phyre2 (28). Protein structures were explored in PyMoL (29) and Python Molecular Viewer from the MGLTools package (25).

Protein expression, purification, thermal shift assays, and isothermal titration calorimetry (ITC).

All proteins were overexpressed in *E. coli* BL21 (DE3) according to (55). PctA-LBD was purified according to (30). The remaining proteins were purified using the same procedure, except that the following buffers were used: buffer A (20 mM Tris, 0.5 M NaCl, 5 % glycerol (vol/vol), 10 mM imidazole, pH 8.0) and buffer B (20 mM Tris, 0.5 M NaCl, 5 % glycerol (vol/vol), 500 mM imidazole, pH 8.0). Freshly purified protein was dialyzed into buffers provided in the Table S5 for immediate thermal shift and ITC analyses.

The detailed experimental protocol of the thermal shift assays has been reported in (31). For these experiments the PM3B compound array from Biolog (https://www.biolog.com/) was used. The composition of this array can be found at https://www.biolog.com/wp-content/uploads/2020/04/00A-042-

Rev-C-Phenotype-MicroArrays-1-10-Plate-Maps.pdf. Briefly, assays were carried out using a MyIQ2 Real-Time PCR instrument (BioRad, Hercules, CA, USA). Ligand solutions were prepared by dissolving the array compounds in 50 µL of MilliQ water, which, according to the information provided by the manufacturer, corresponds to a concentration of 10–20 mM. Experiments were conducted in 96-well plates and each assay mixture contained 20.5 µL of the dialyzed protein (at 10-30 µM), 2 µL of 5X SYPRO orange (Life Technologies, Eugene, Oregon, USA) and 2.5 µL of the resuspended array compounds or the equivalent amount of buffer in the ligand-free control. Samples were heated from 23 ºC to 85ºC at a scan rate of 1 ºC/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.

Site-directed mutagenesis.

An overlapping PCR mutagenesis approach was employed to construct the substitution mutants of PctA-LBD using primers listed in Table S6. The resulting PCR fragments were cloned into the NdeI/BamHI sites of the expression vector pET28b(+). *Escherichia coli* DH5α was used as host for gene cloning. The resulting plasmids (Table S6) were verified by DNA sequencing. Mutant proteins were purified following the protocol for the wild-type protein.

We have repeated the isothermal titration calorimetry (ITC) experiments with the mutants of PctA chemoreceptor from *P. aeruginosa* PAO1 using a higher ligand concentration and larger injection volumes (Fig. S1). In the binding study with 3.3 mM L-Ala (12.8 µl injection volume) the D173A mutant has lost its affinity to the ligand (even at higher ligand concentration); the K_D value derived for the R126A mutant was of 202 ± 8 µM, corresponding to a 61-fold reduction as compared to the native protein (Fig. S1A). In the study with 3 mM L-Ser (12.8 µl injection volume) similar to L-Ala, D173A was devoid of L-Ser binding and a residual affinity has been determined for the R126A mutant with a K_D value of 258 ± 24 µM (Fig. S1B).

In the experimental set-up, the final ligand concentration in the protein containing sample cell is 175 µM. However, if binding occurred at lower affinity (i.e. K_D values in the mM range) one would not see

binding since it does not occur at this low ligand concentration. To drive complex formation and thus to monitor lower affinity binding, the concentration of both ligands must be increased. Therefore, in the next step we conducted the ITC experiments using 30 times more ligand concentration. The upper trace Fig. S1C corresponds to a titration of 140 µM PctA D173N mutant with 12.8 µl aliquots of 15 mM L-Ala. Data indicate binding and a K_D constant of approximately 2.1 mM was estimated (important: in Fig. S1C for K_D of 1 mM it is better to use the term "estimated" instead of "determined"). We have already shown that the D173A lacks binding activity. To confirm this the PctA D173A mutant was titrated using the same procedure as above (titration of 125 µM protein with 12.8 µl aliquots of 15 mM L-Ala). As shown in the lower trace in Fig. S1C no binding heats were observed with high ligand concentration confirming the absence of binding. Thus, replacement of D173 by A abolishes binding whereas substitution for N reduces the affinity by a factor of approximately 600. Experiemtnal conditions of ITC studies with eight new proteins described in the manuscript are presented in Table S2.

Imaging cell surface expression of α2δ-1

The calcium channel α2δ-1 subunit (rat, GenBank: M86621) (32), containing an HA tag (33) was used in the pcDNA3 vector and expressed in tsA-201 cells, as described (34). The D491A mutation was introduced using standard techniques, and verified by DNA sequencing. Cells were transfected with either α2δ-1 wild-type (WT), α2δ-1^{D491A} or α2δ-1^{R241A} (35) using PolyJet (Tebu-bio Ltd) according to the manufacturer's protocol. Media was replaced 16 h after transfection, with serum-free media in the absence (control) or presence of gabapentin (0.1 mM or 1 mM as indicated) and cells were incubated for a further 24 h. Cells were fixed and incubated in blocking buffer as described previously (35), before being incubated with anti-HA (rat, monoclonal, Sigma-Aldrich) antibody, washed and then incubated with the secondary antibody, anti-rat Alexa fluor 488 (Thermo Fisher). After extensive washing, cells were permeabilized with 0.1 % Triton in PBS for 5 min, then incubated again with anti-HA, followed by anti-rat Alexa fluor 594. All antibody incubations were at 20 ^{o}C for 1 h. The nuclei were stained with 0.5 µM 4',6'-diamidino-2-phenylindole (DAPI) for 10 min, before coverslips were mounted using VECTASHIELD® mounting medium (Vector Laboratories). Imaging was performed on a Zeiss LSM

780 confocal microscope as described previously (35). The tile function was used and expression was measured in all positively transfected cells to remove bias. Quantification of fluorescence intensity was carried out using ImageJ and data were analysed using GraphPad Prism 7.

SI References

1. Parks DH, et al. (2018) A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol 36(10):996-1004.

2. El-Gebali S, et al. (2019) The Pfam protein families database in 2019. Nucleic Acids Res 47(D1):D427-D432.

3. UniProt C (2019) UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 47(D1):D506-D515.

4. One Thousand Plant Transcriptomes I (2019) One thousand plant transcriptomes and the phylogenomics of green plants. Nature 574(7780):679-685.

5. Gumerov VM & Zhulin IB (2020) TREND: a platform for exploring protein function in prokaryotes based on phylogenetic, domain architecture and gene neighborhood analyses. Nucleic Acids Res 48(W1):W72-W76.

6. Soding J, Biegert A, & Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 33(Web Server issue):W244-248.

7. Richter DJ, Berney C, Strassert JFH, Burki F, & de Vargas C (2020).

8. Eddy SR (2011) Accelerated Profile HMM Searches. PLoS Comput Biol 7(10):e1002195.

9. Burki F, Okamoto N, Pombert JF, & Keeling PJ (2012) The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. Proc Biol Sci 279(1736):2246-2254.

10. Telford MJ, Budd GE, & Philippe H (2015) Phylogenomic Insights into Animal Evolution. Curr Biol 25(19):R876-887.

11. Letunic I & Bork P (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47(W1):W256-W259.

12. Katoh K & Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30(4):772-780.

13. Waterhouse AM, Procter JB, Martin DM, Clamp M, & Barton GJ (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics 25(9):1189-1191.

14. Mendler K, et al. (2019) AnnoTree: visualization and exploration of a functionally annotated microbial tree of life. Nucleic Acids Res 47(9):4442-4448.

15. Capella-Gutierrez S, Silla-Martinez JM, & Gabaldon T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25(15):1972-1973.

16. Darriba D, Taboada GL, Doallo R, & Posada D (2011) ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27(8):1164-1165.

17. Akaike H (1998) Information Theory and an Extension of the Maximum Likelihood Principle. Selected Papers of Hirotugu Akaike, Springer Series in Statistics), pp 199-213.

18. Schwarz G (1978) Estimating the Dimension of a Model. The Annals of Statistics 6(2).

19. Whelan S & Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol 18(5):691-699.

20. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9):1312-1313.

21. Ronquist F, et al. (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61(3):539-542.

22. Trott O & Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31(2):455-461.

23. Wu J, et al. (2016) Structure of the voltage-gated calcium channel Ca(v)1.1 at 3.6 A resolution. Nature 537(7619):191-196.

24. Jumper J, et al. (2021) Highly accurate protein structure prediction with AlphaFold. Nature 596(7873):583-589.

25. Sanner MF (1999) Python: a programming language for software integration and development. J Mol Graph Model 17(1):57-61.

26. Irwin JJ & Shoichet BK (2005) ZINC--a free database of commercially available compounds for virtual screening. Journal of chemical information and modeling 45(1):177-182.

27. O'Boyle NM, et al. (2011) Open Babel: An open chemical toolbox. J Cheminform 3:33.

28. Kelley LA, Mezulis S, Yates CM, Wass MN, & Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10(6):845-858.

29. Anonymous (2015) The PyMOL Molecular Graphics System, Version 2.0, Schrödinger, LLC.

30. Rico-Jimenez M, et al. (2013) Paralogous chemoreceptors mediate chemotaxis towards protein amino acids and the non-protein amino acid gamma-aminobutyrate (GABA). Molecular microbiology 88(6):1230-1243.

31. Fernandez M, et al. (2018) High-Throughput Screening to Identify Chemoreceptor Ligands. Methods Mol Biol 1729:291-301.

32. Kim HL, Kim H, Lee P, King RG, & Chin H (1992) Rat brain expresses an alternatively spliced form of the dihydropyridine-sensitive L-type calcium channel alpha 2 subunit. Proceedings of the National Academy of Sciences of the United States of America 89(8):3251-3255.

33. Davies A, et al. (2006) The calcium channel alpha2delta-2 subunit partitions with CaV2.1 into lipid rafts in cerebellum: implications for localization and function. J Neurosci 26(34):8748-8757.

34. Meyer JO & Dolphin AC (2021) Rab11-dependent recycling of calcium channels is mediated by auxiliary subunit alpha2delta-1 but not alpha2delta-3. Sci Rep 11(1):10256.

35. Cassidy JS, Ferron L, Kadurin I, Pratt WS, & Dolphin AC (2014) Functional exofacially tagged N-type calcium channels elucidate the interaction with auxiliary alpha2delta-1 subunits. Proceedings of the National Academy of Sciences of the United States of America 111(24):8979-8984.

36. Jeong H, et al. (2009) Genome sequences of Escherichia coli B strains REL606 and BL21(DE3). Journal of molecular biology 394(4):644-652.

37. Woodcock DM, et al. (1989) Quantitative evaluation of Escherichia coli host strains for tolerance to cytosine methylation in plasmid and phage recombinants. Nucleic Acids Res 17(9):3469-3478.

38. Taguchi K, Fukutomi H, Kuroda A, Kato J, & Ohtake H (1997) Genetic identification of chemotactic transducers for amino acids in Pseudomonas aeruginosa. Microbiology 143 (Pt 10):3223-3229.

39. Reyes-Darias JA, Yang Y, Sourjik V, & Krell T (2015) Correlation between signal input and output in PctA and PctB amino acid chemoreceptor of Pseudomonas aeruginosa. Molecular microbiology 96(3):513-525.

40. Gavira JA, et al. (2020) How Bacterial Chemoreceptors Evolve Novel Ligand Specificities. mBio 11(1).

41. Grist R, Croker A, Denne M, & Stallard P (2018) Technology Delivered Interventions for Depression and Anxiety in Children and Adolescents: A Systematic Review and Meta-analysis. Clin Child Fam Psychol Rev.

42. Reyes-Darias JA, et al. (2015) Specific gamma-aminobutyrate chemotaxis in pseudomonads with different lifestyle. Molecular microbiology 97:488-501.

43. Fernandez M, Morel B, Corral-Lugo A, & Krell T (2016) Identification of a chemoreceptor that specifically mediates chemotaxis toward metabolizable purine derivatives. Molecular microbiology 99(1):34-42.

44. Webb BA, Hildreth S, Helm RF, & Scharf BE (2014) Sinorhizobium meliloti chemoreceptor McpU mediates chemotaxis toward host plant exudates through direct proline sensing. Appl Environ Microbiol 80(11):3404-3415.

45. Webb BA, et al. (2017) Sinorhizobium meliloti Chemotaxis to Multiple Amino Acids Is Mediated by the Chemoreceptor McpU. Mol Plant Microbe Interact 30(10):770-777.

46. Ud-Din A, Khan MF, & Roujeinikova A (2020) Broad Specificity of Amino Acid Chemoreceptor CtaA of Pseudomonas fluorescens Is Afforded by Plasticity of Its Amphipathic Ligand-Binding Pocket. Mol Plant Microbe Interact 33(4):612-623.

47. Feng H, et al. (2018) Identification of Chemotaxis Compounds in Root Exudates and Their Sensing Chemoreceptors in Plant-Growth-Promoting Rhizobacteria Bacillus amyloliquefaciens SQR9. Mol Plant Microbe Interact 31(10):995-1005.

48. Glekas GD, et al. (2010) A PAS domain binds asparagine in the chemotaxis receptor McpB in Bacillus subtilis. J Biol Chem 285(3):1870-1878.

49. Glekas GD, et al. (2012) The Bacillus subtilis chemoreceptor McpC senses multiple ligands using two discrete mechanisms. J Biol Chem 287(47):39412-39418.

50. Nishiyama S, et al. (2016) Identification of a Vibrio cholerae chemoreceptor that senses taurine and amino acids as attractants. Sci Rep 6:20866.

51. Takahashi Y, Nishiyama SI, Sumita K, Kawagishi I, & Imada K (2019) Calcium Ions Modulate Amino Acid Sensing of the Chemoreceptor Mlp24 of Vibrio cholerae. Journal of bacteriology 201(9).

52. Ehrhardt MKG, Gerth ML, & Johnston JM (2021) Structure of a double CACHE chemoreceptor ligand-binding domain from Pseudomonas syringae provides insights into the basis of proline recognition. Biochemical and biophysical research communications 549:194-199.

53. Cerna-Vargas JP, et al. (2019) Chemoperception of Specific Amino Acids Controls Phytopathogenicity in Pseudomonas syringae pv. tomato. mBio 10(5).

54. McKellar JL, Minnell JJ, & Gerth ML (2015) A high-throughput screen for ligand binding reveals the specificities of three amino acid chemoreceptors from Pseudomonas syringae pv. actinidiae. Molecular microbiology 96(4):694-707.

55. Khan MF, et al. (2020) Structure-Activity Relationship Study Reveals the Molecular Basis for Specific Sensing of Hydrophobic Amino Acids by the Campylobacter jejuni Chemoreceptor Tlp3. Biomolecules 10(5).

56. Robinson CD, et al. (2021) Host-emitted amino acid cues regulate bacterial chemokinesis to enhance colonization. Cell host & microbe 29(8):1221-1234 e1228.

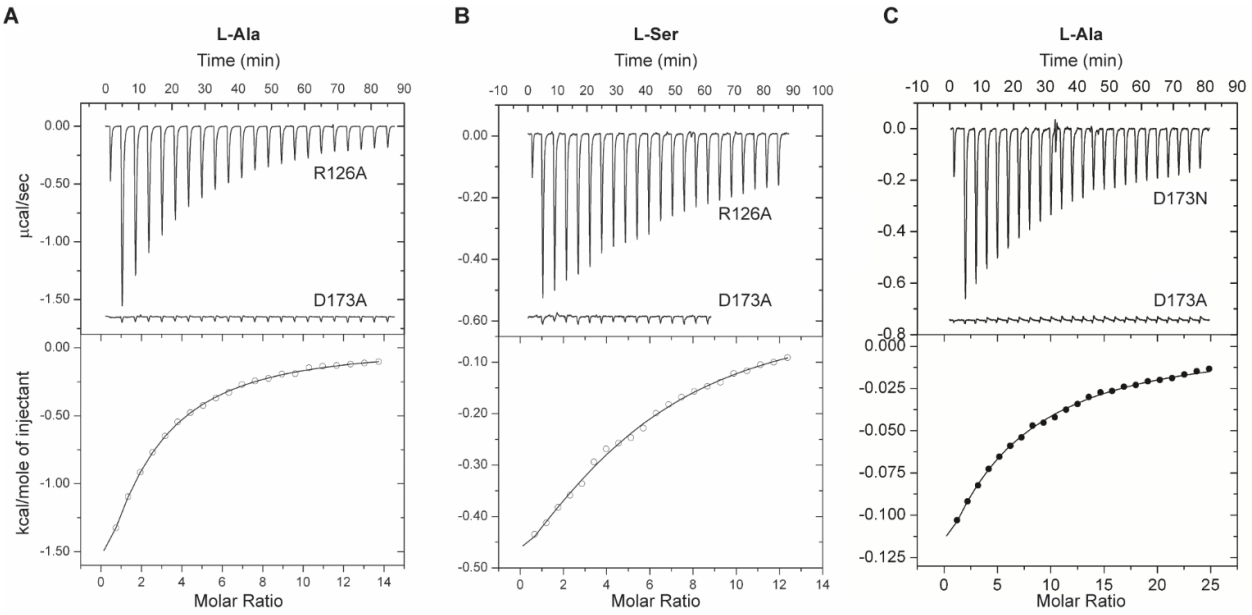


Fig. S1. Isothermal titration calorimetry binding studies of *P. aeruginosa* PctA dCache_1AA mutants. (**A**) 3.3 mM L-Ala (12.8 µl injection volume). (**B**) 3 mM L-Ser (12.8 µl injection volume). The protein concentration was 50 µM. R126A and D173A – mutations of two key residues of AA_motif to Ala. (**C**) 15 mM L-Ala (12.8 µl); the protein concentrations were 140 µM and 125 µM for D173N and D173A mutants, respectively. Upper panels: raw titration data. Lower panels: integrated, dilution heat-corrected and concentration-normalized peak areas of the titration data fitted with the 'One binding site' model of ORIGIN (ORIGINLAB CORPORATION, Northampton, MA, USA)

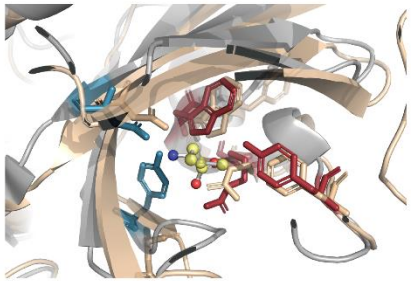
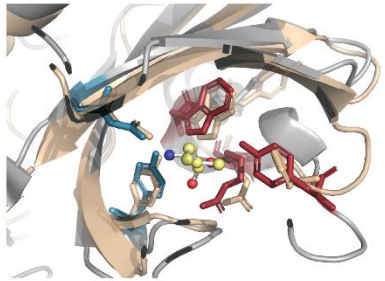
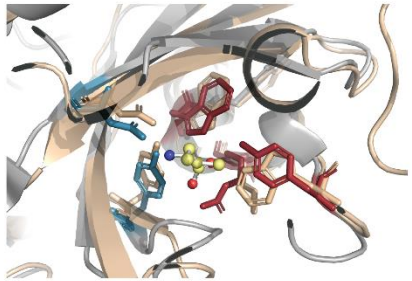
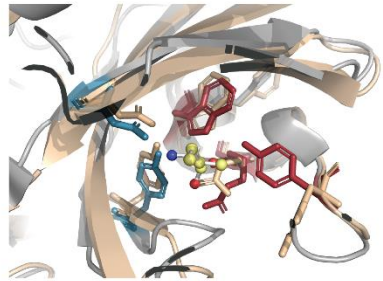
Fig. S2. **Structures of dCache_1AA domains of proteins from several major bacterial and archaeal phyla modeled using Phyre2.** The modeled structures (shown in gold) are superimposed with the solved structure of dCache_1AA domain of PctA chemoreceptor from *Pseudomonas aeruginosa* PAO1 (shown in gray; PDB Id 5T65). On each page the first panel shows ligand binding pockets with the AA_motif residues, the second panel – entire dCache_1AA domains. Protein identifiers and AA_motif variants are also shown.

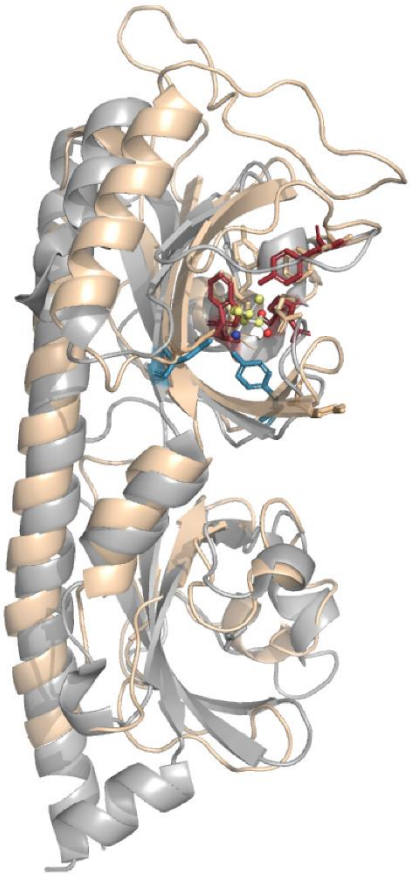
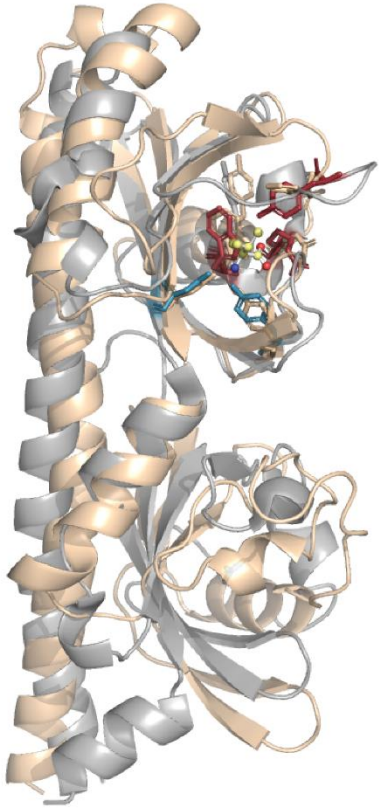
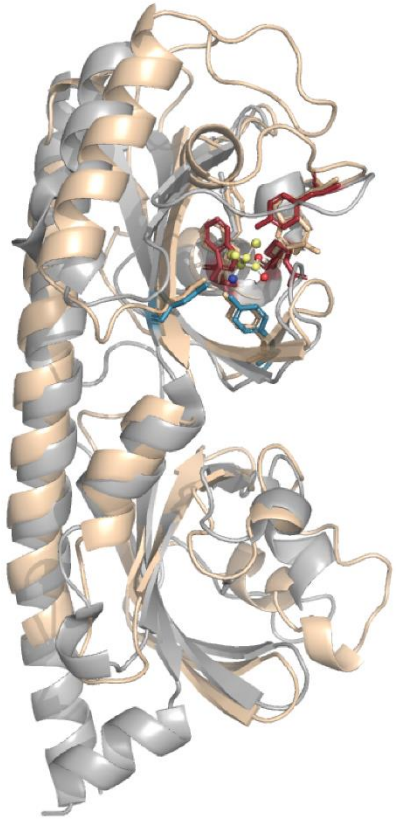
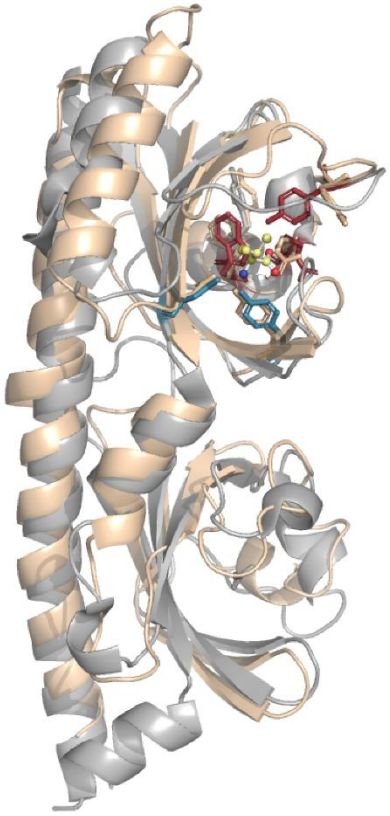
Fibrobacterota
Chitinivibrio alkaliphilus
WP_022636485.1
F....R.WY Y D

Chloroflexota
Chloroflexus islandicus
WP_066789187.1
F....R.WY Y D

Firmicutes_A
Clostridium botulinum
WP_021135996.1
Y....R.WY Y D

Firmicutes_B
Desulfofarcimen acetoxidans
WP_015757146.1
Y....R.WY Y D

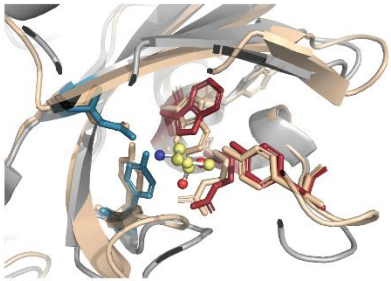
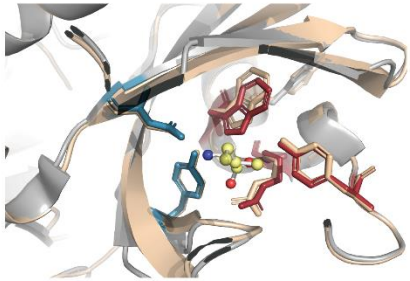
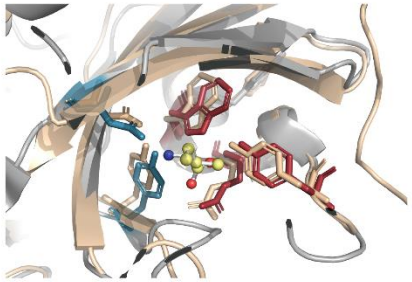
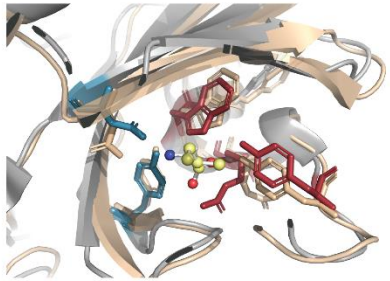
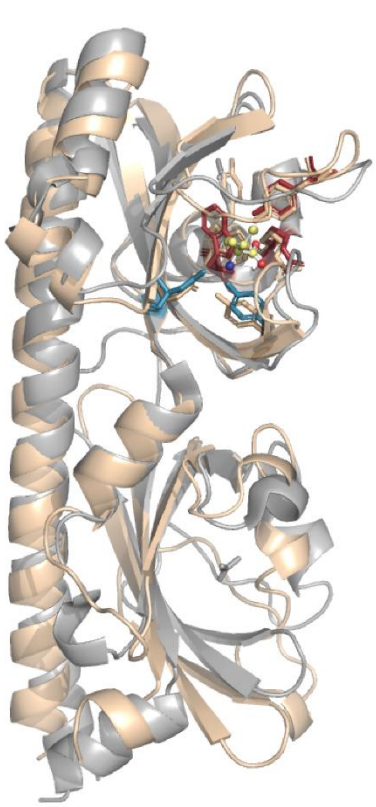
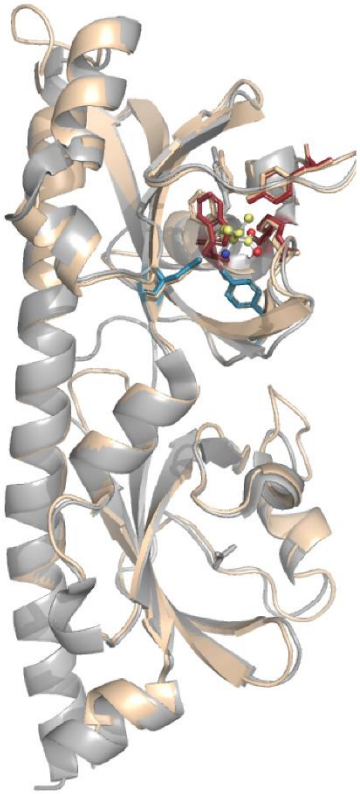
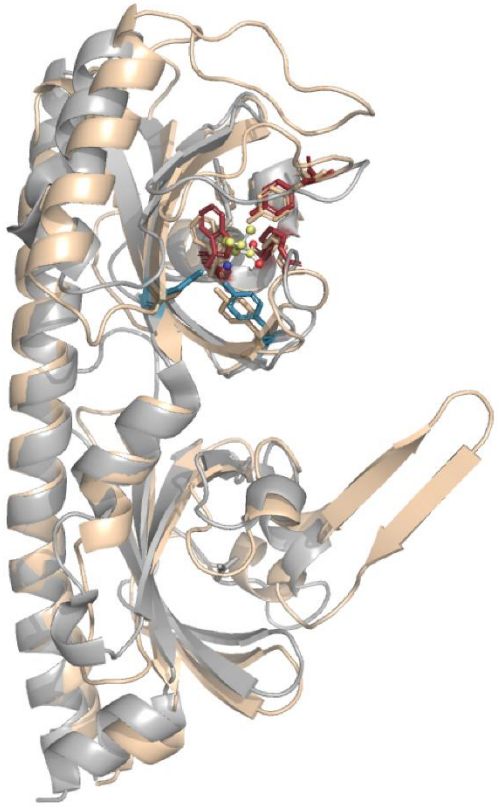
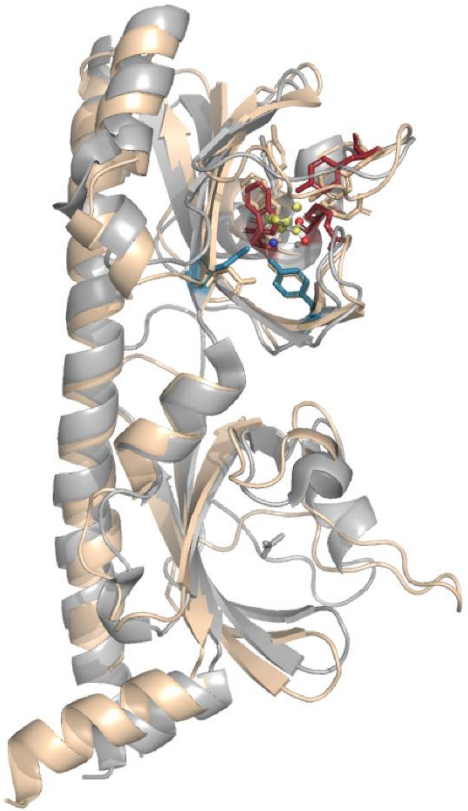





Cyanobacteria
Heliobacterium mobile
WP_155475420.1
Y....R.WY Y D

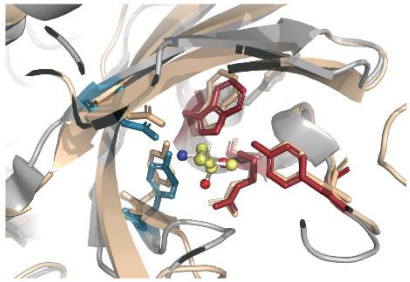
Fusobacteriota
Hypnocyclicus thermotrophus
WP_134111860.1
Y....R.WY Y D

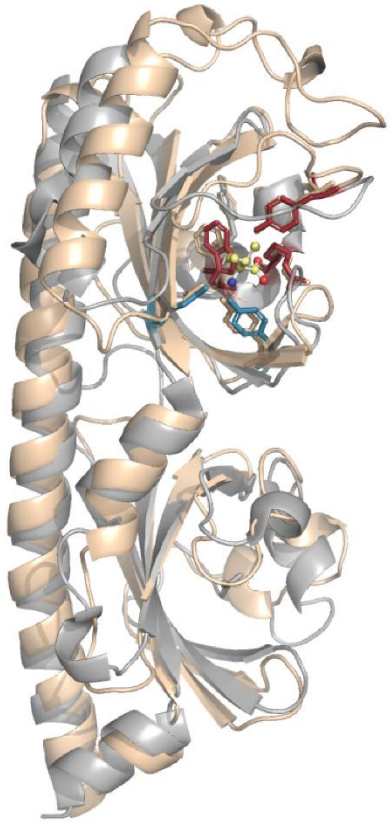
Verrucomicrobiota
Lentisphaeria bacterium
WP_176937576.1
Y....R.WY Y D

Proteobacteria
Marinicaulis flavus
WP_104830897.1
Y....R.WY Y D

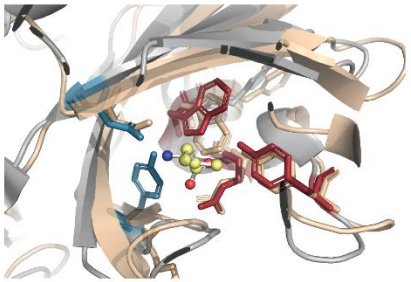









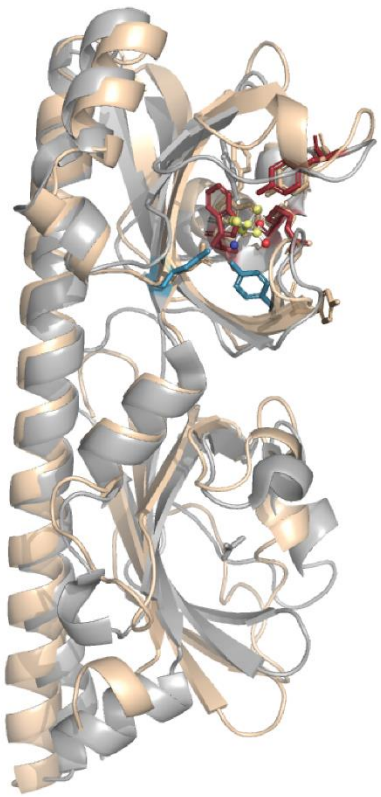
Archaea, Halobacteriota
Methanospirillum hungatei
WP_011449640.1
Y....R.WY Y D



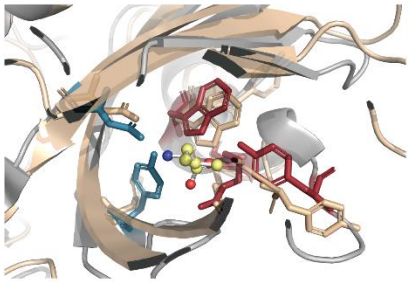


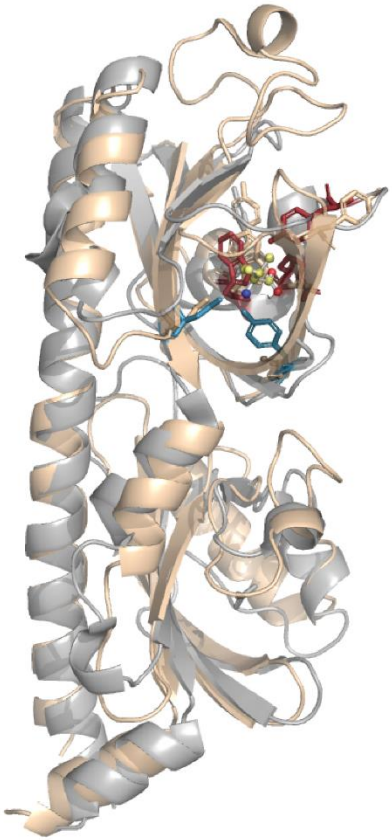
Myxococcota
Polyangium spumosum
WP_153822233.1
Y....R.WF Y D



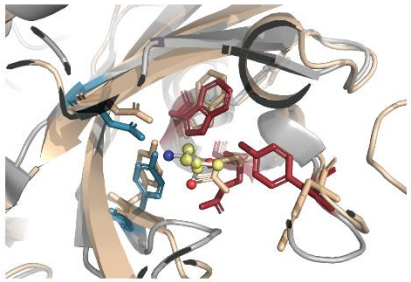


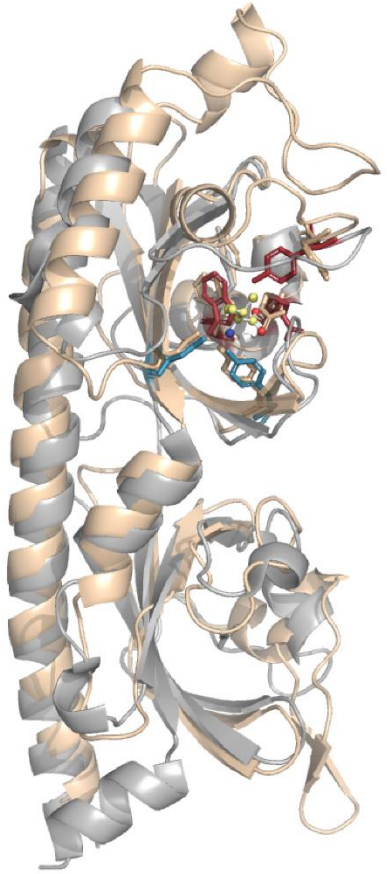
Bacteroidota
Prevotella ruminicola
WP_028908302.1
Y....K.WY Y D



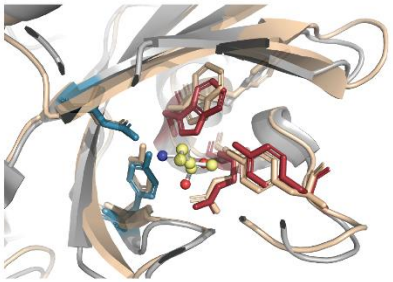


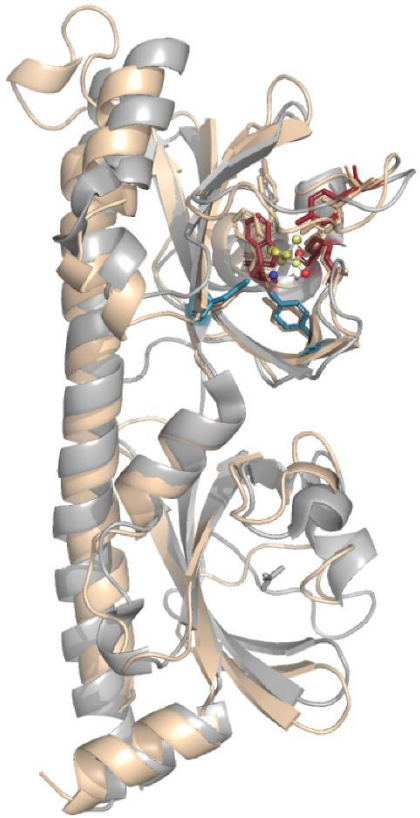
Firmicutes
Propionispora hippei
WP_149734257.1
Y....R.WY Y D



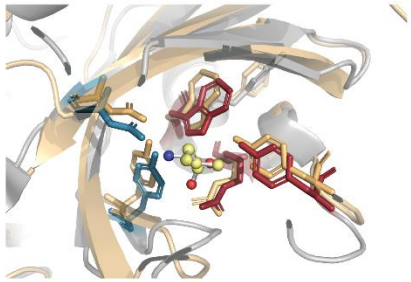


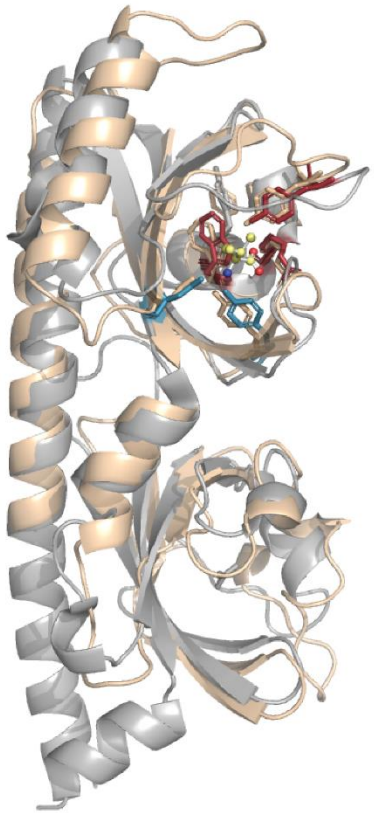
Synergistota
Thermanaerovibrio acidaminovorans
WP_012869308.1
Y...R.WY Y D



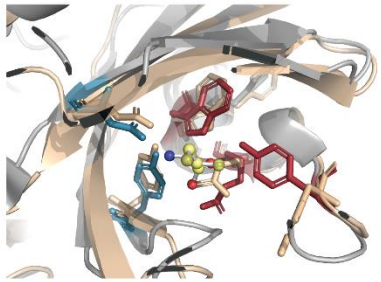


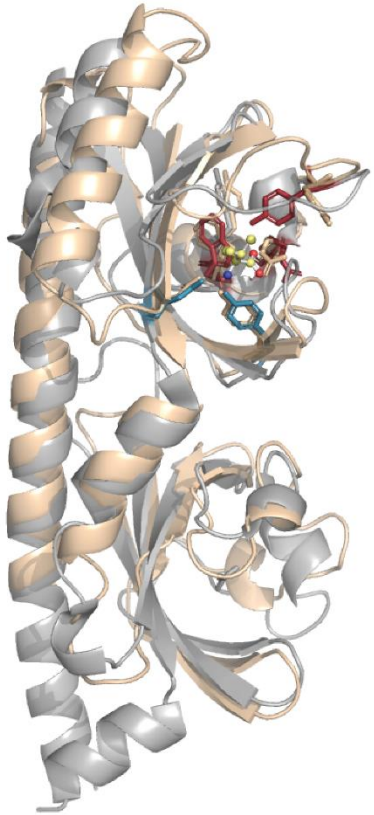
Archaea, Methanobacteriota
Thermococcus kodakarensis
WP_011251097.1
Y...R.WY Y D



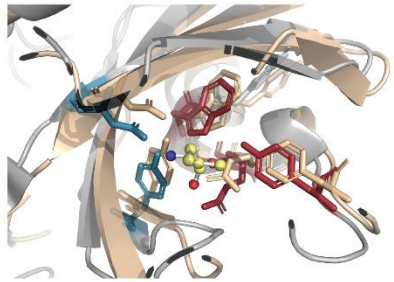


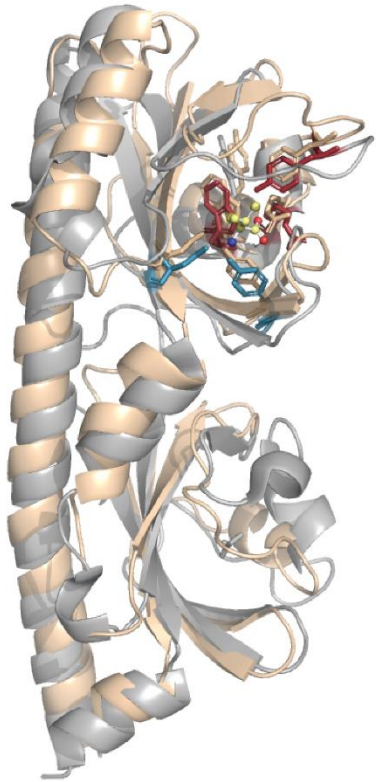
Desulfobacterota
Thermodesulfobacterium thermophilum
WP_162138226.1
Y...R.WY Y D



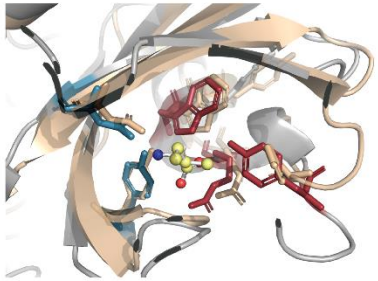


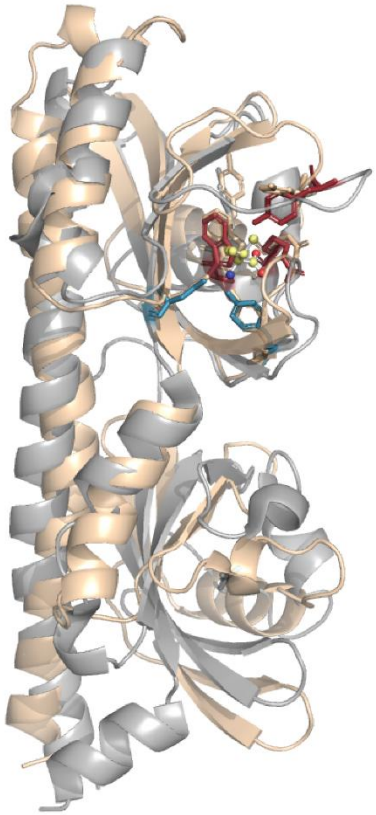
Thermosulfidibacterota
Thermosulfidibacter takai
WP_083498597.1
F....R.WY Y D



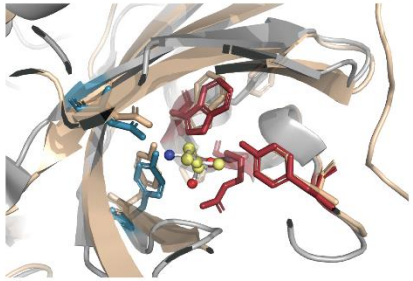


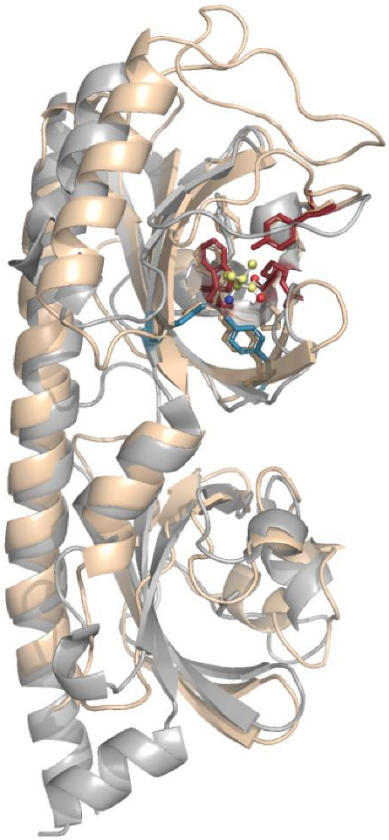
Thermotogota
Thermotoga petrophila
WP_011943173.1
Y...R.WY Y D



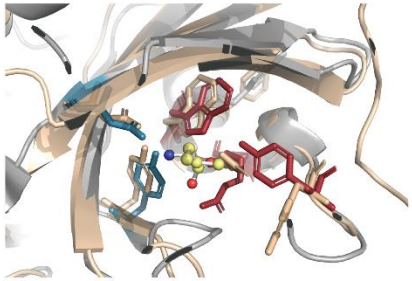


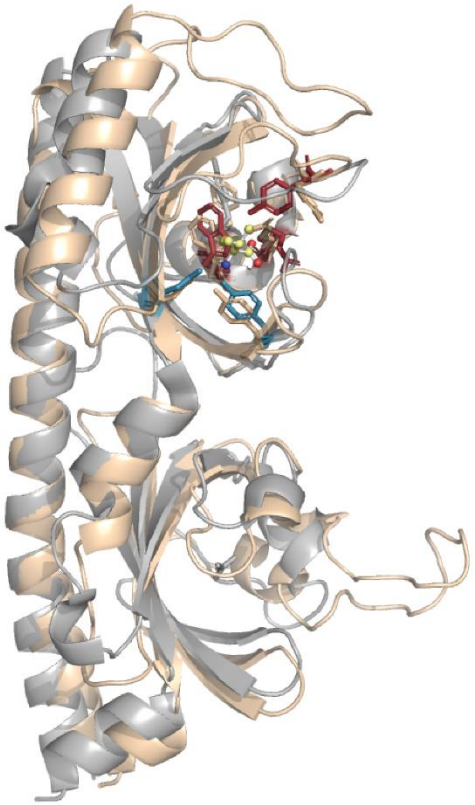
Spirochaeta
Treponema denticola
WP_002687321.1
F...R.WY Y D



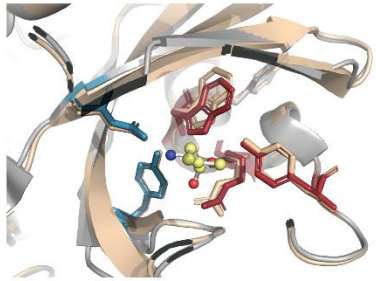


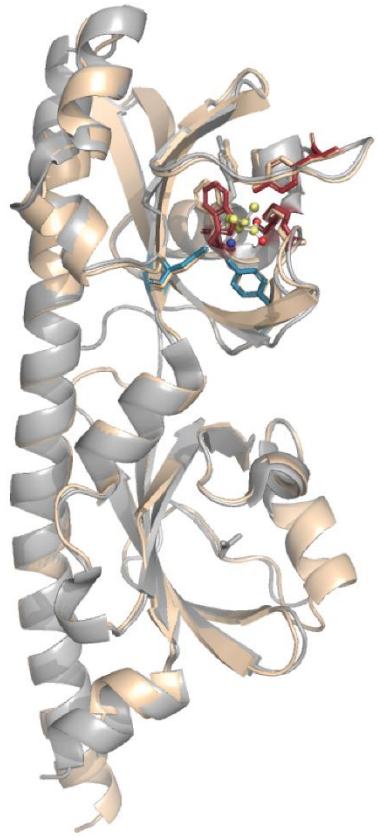
Verrucomicrobiota
Victivallis sp Marseille-Q1083
WP_176014353.1
Y...R.WF Y D





Proteobacteria
Yersinia pestis
WP_016674185.1
Y...R.WY Y D





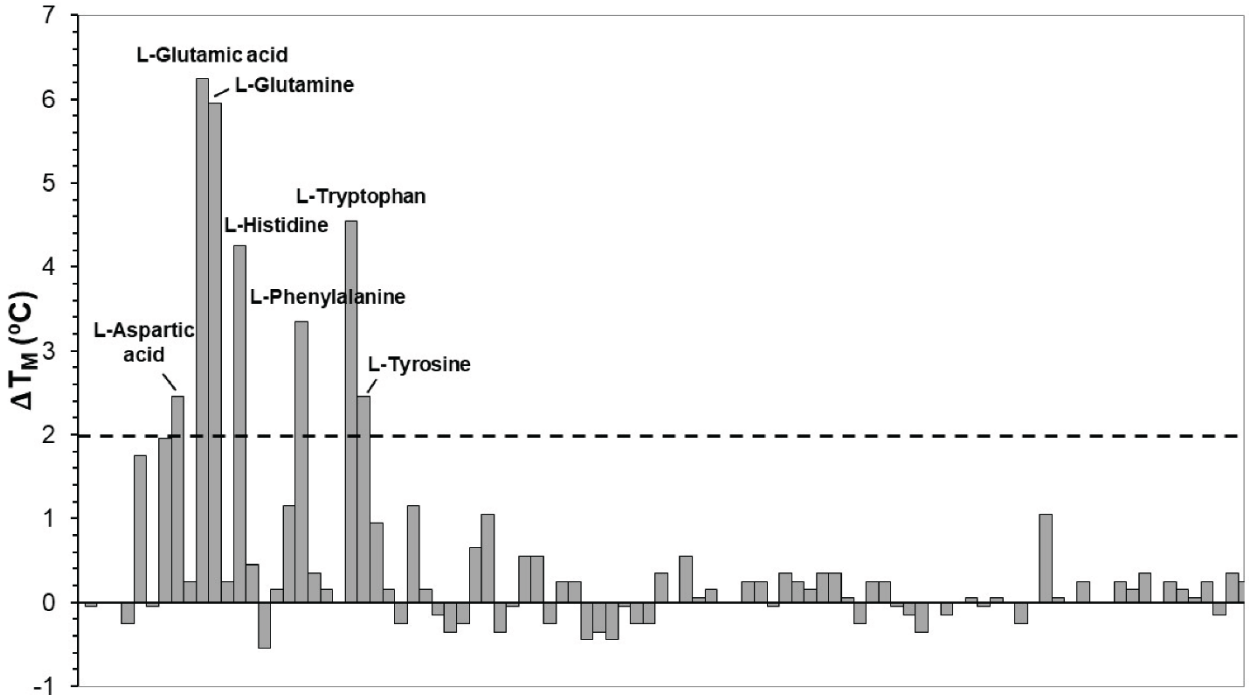


Fig. S3. **Thermal shift assays of the recombinant dCache_1AA domain of the NP_233280.1 metal-dependent phosphodiesterase of *Vibrio cholerae* in the presence of bacterial nitrogen sources from the Biolog screen plate PM3.** Shown are changes in T_M respective to the protein without ligand (49.2 ºC). Compounds that caused T_M shifts of at least 2 ºC are annotated.

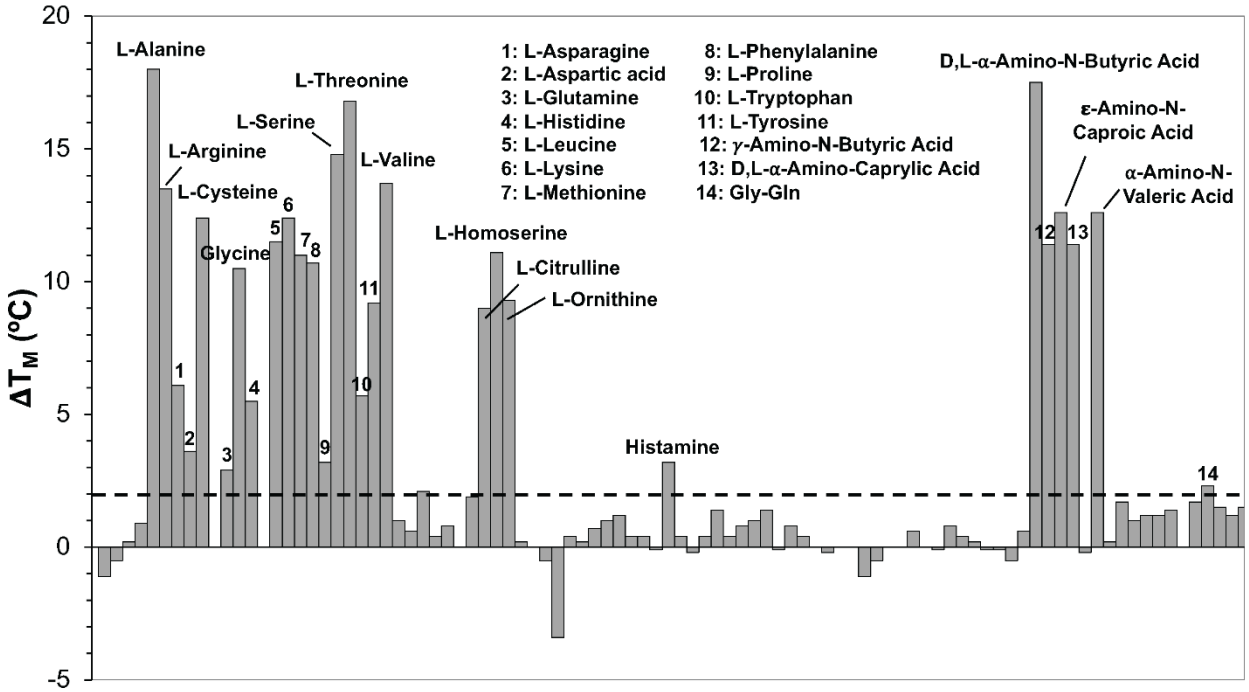


Fig. S4. Thermal shift assays of the recombinant dCache_1AA domain of the WP_016674185.1 chemoreceptor of *Yersinia pestis* in the presence of bacterial nitrogen sources from the Biolog screen plate PM3. Shown are changes in T_M respective to the protein without ligand (38.1 ºC). Compounds that caused T_M shifts of at least 2 ºC are annotated.

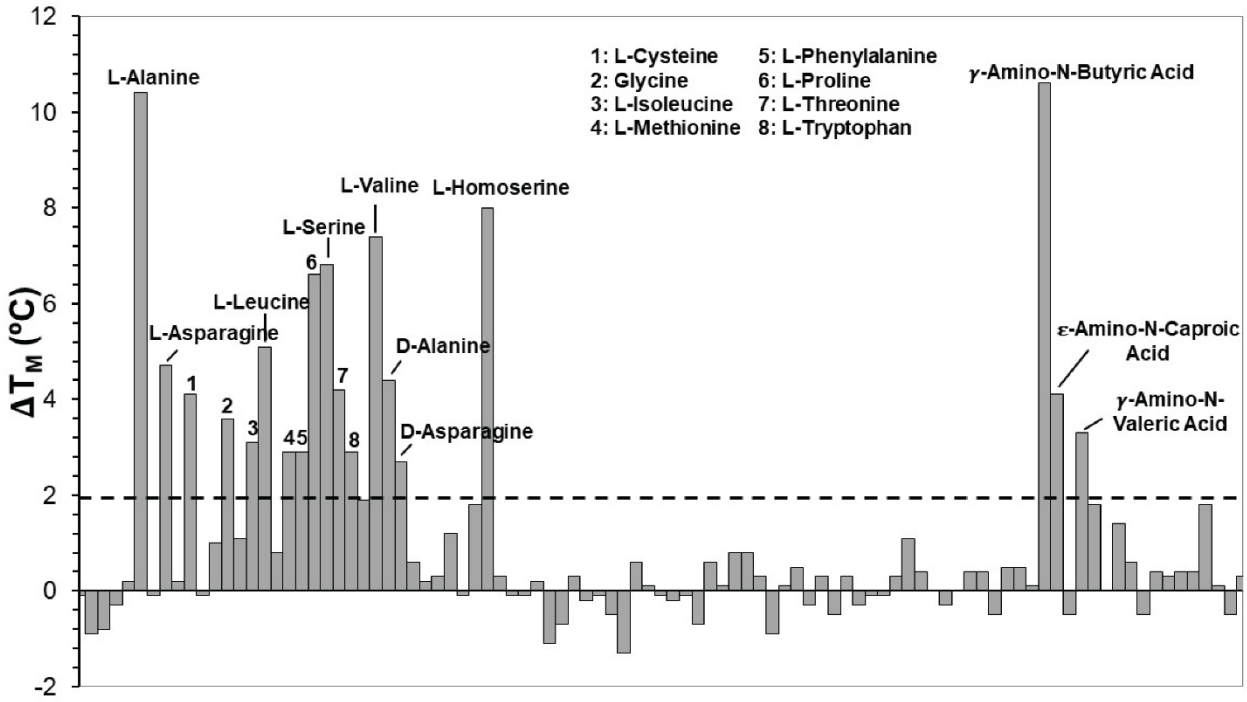


Fig. S5. Thermal shift assays of the recombinant dCache_1AA domain of the WP_154766400.1 guanylate/adenylate cyclase of *Legionella pneumophila* in the presence of bacterial nitrogen sources from the Biolog screen plate PM3. Shown are changes in TM respective to the protein without ligand (64.9 ºC). Compounds that caused TM shifts of at least 2 ºC are annotated.

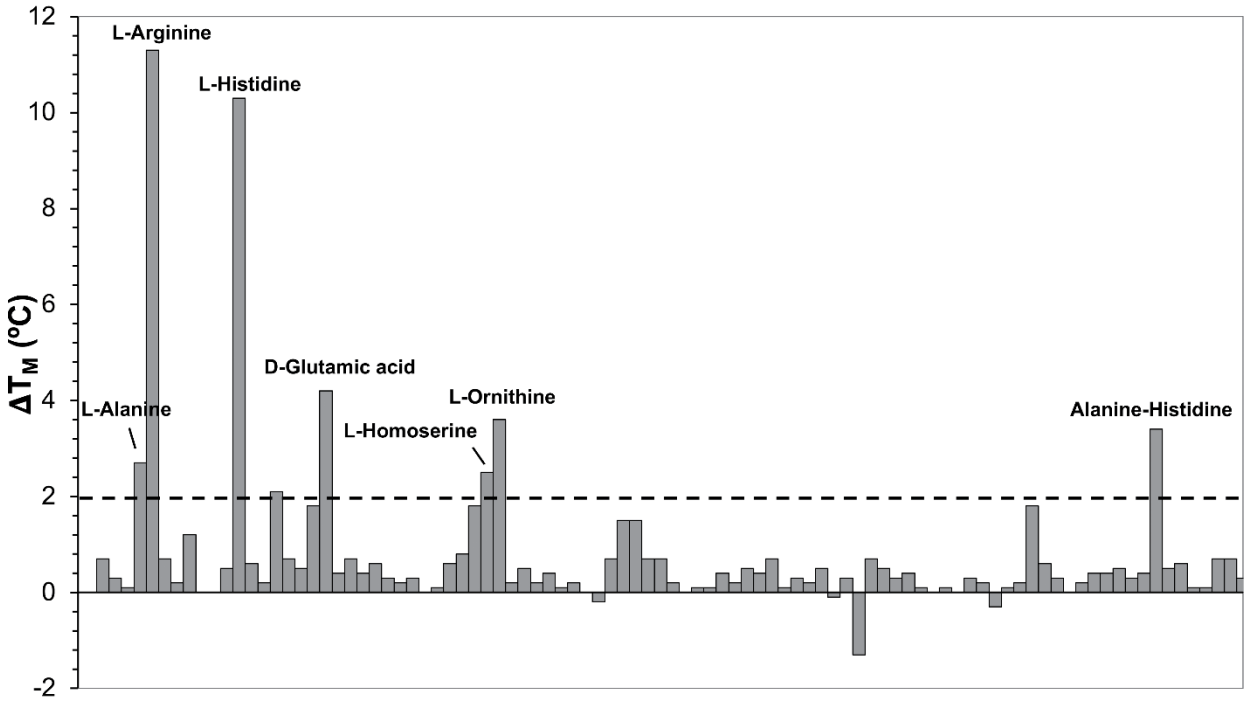


Fig. S6. Thermal shift assays of the recombinant dCache_1AA domain of the WP_002687321.1 chemoreceptor of *Treponema denticola* in the presence of bacterial nitrogen sources from the Biolog screen plate PM3. Shown are changes in TM respective to the protein without ligand (40.3 ºC). Compounds that caused TM shifts of at least 2 ºC are annotated.

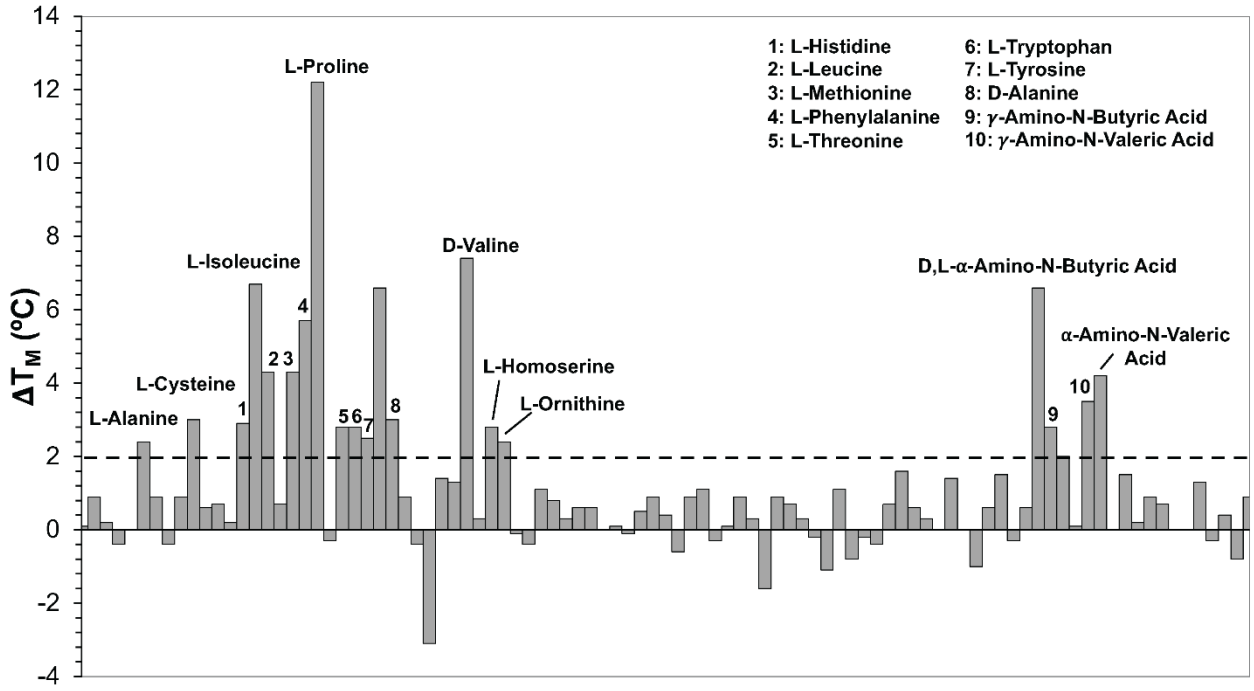


Fig. S7. Thermal shift assays of the recombinant dCache_1AA domain of the WP_162138226.1 diguanylate cyclase protein of *Thermodesulfobacterium thermophilum* in the presence of bacterial nitrogen sources from the Biolog screen plate PM3. Shown are changes in T_M respective to the protein without ligand (61.4 ºC). Compounds that caused T_M shifts of at least 2 ºC are annotated.

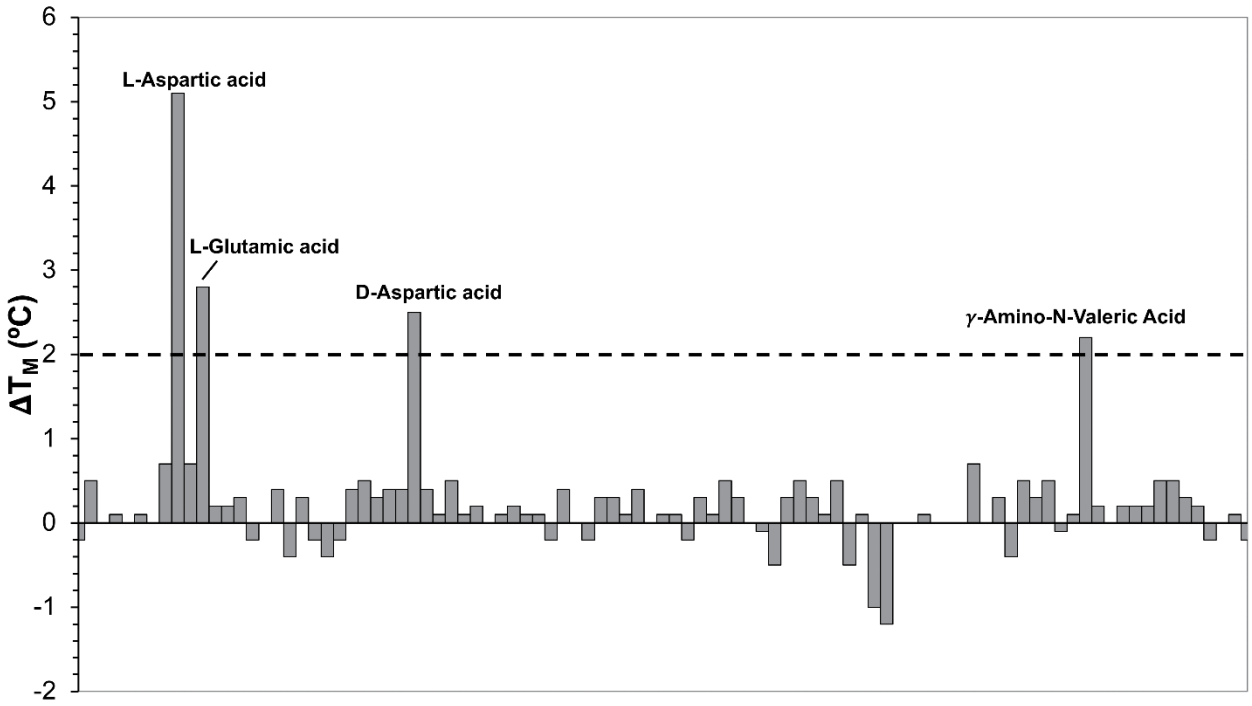


Fig. S8. Thermal shift assays of the recombinant dCache_1AA domain of the WP_106093935.1 serine/threonine kinase of *Enhygromyxa salina* in the presence of bacterial nitrogen sources from the Biolog screen plate PM3. Shown are changes in TM respective to the protein without ligand (50.5 ºC). Compounds that caused TM shifts of at least 2 ºC are annotated.

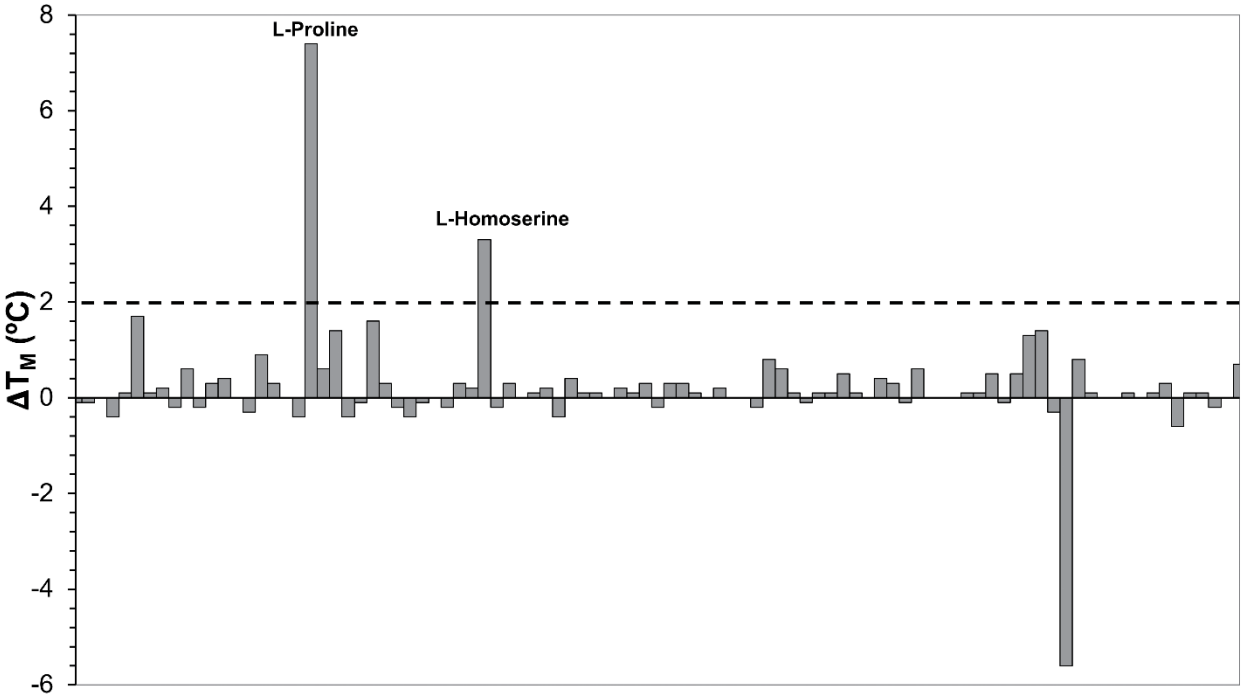


Fig. S9. Thermal shift assays of the recombinant dCache_1AA domain of the WP_152054232.1 serine/threonine phosphatase of *Tautonia marina* in the presence of bacterial nitrogen sources from the Biolog screen plate PM3. Shown are changes in TM respective to the protein without ligand (46.3 ºC). Compounds that caused TM shifts of at least 2 ºC are annotated.

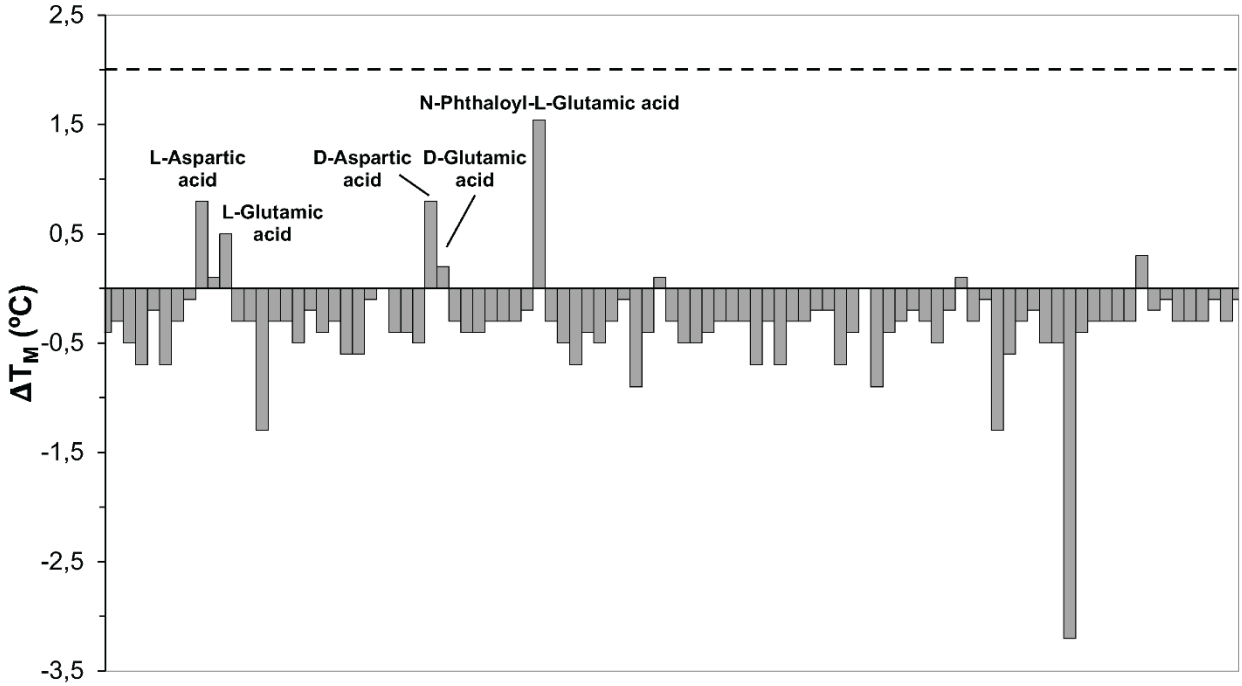


Fig. S10. Thermal shift assays of the recombinant dCache_1AA domain of the WP_011449640.1 hybrid histidine kinase of *Methanospirillum hungatei* in the presence of bacterial nitrogen sources from the Biolog screen plate PM3. Shown are changes in TM respective to the protein without ligand (44.9 ºC). Compounds that caused TM shifts of at least 0.5 ºC are annotated.

Fig. S11. Structure-based protein sequence alignment of PctA dCache_1 domain with two regions of the rabbit α2δ-1 and human α2δ and CACHD1 subunits. Topology is based on the protein structures of PctA (PDB ID 5T65) and α2δ-1 from rabbit (PDB ID 6JPA). 1st dCache_1 domain is shaded in purple, 2nd dCache_1 – in blue, and VWA domain – in green. The AA_motif amino acid positions are in bold and colored in accordance with Figure 1. P.aer – *P. aeruginosa* PAO1, PctA, XP_011514873.1; O.cun – *Oryctolagus cuniculus,* α2δ-1, NP_001075745.1; H.sap – *Homo sapiens*: α2δ-1 –

NP_001353796.1, α2δ-2 – NP_006021.2, α2δ-3 – NP_060868.2, α2δ-4 – NP_758952.4, CACHD1 – NP_065976.3. The MSA showed that the bacterial dCache_1 domain can be aligned with two regions of human α2δ and CACHD1 proteins: N-terminally and C-terminally located, respectively. The secondary protein structure elements mapped on the alignment showed that the aligned regions of the bacterial and eukaryotic sequences are in good agreement with each other. These data suggested that α2δ and CACHD1 proteins contain two dCache_1 domains.

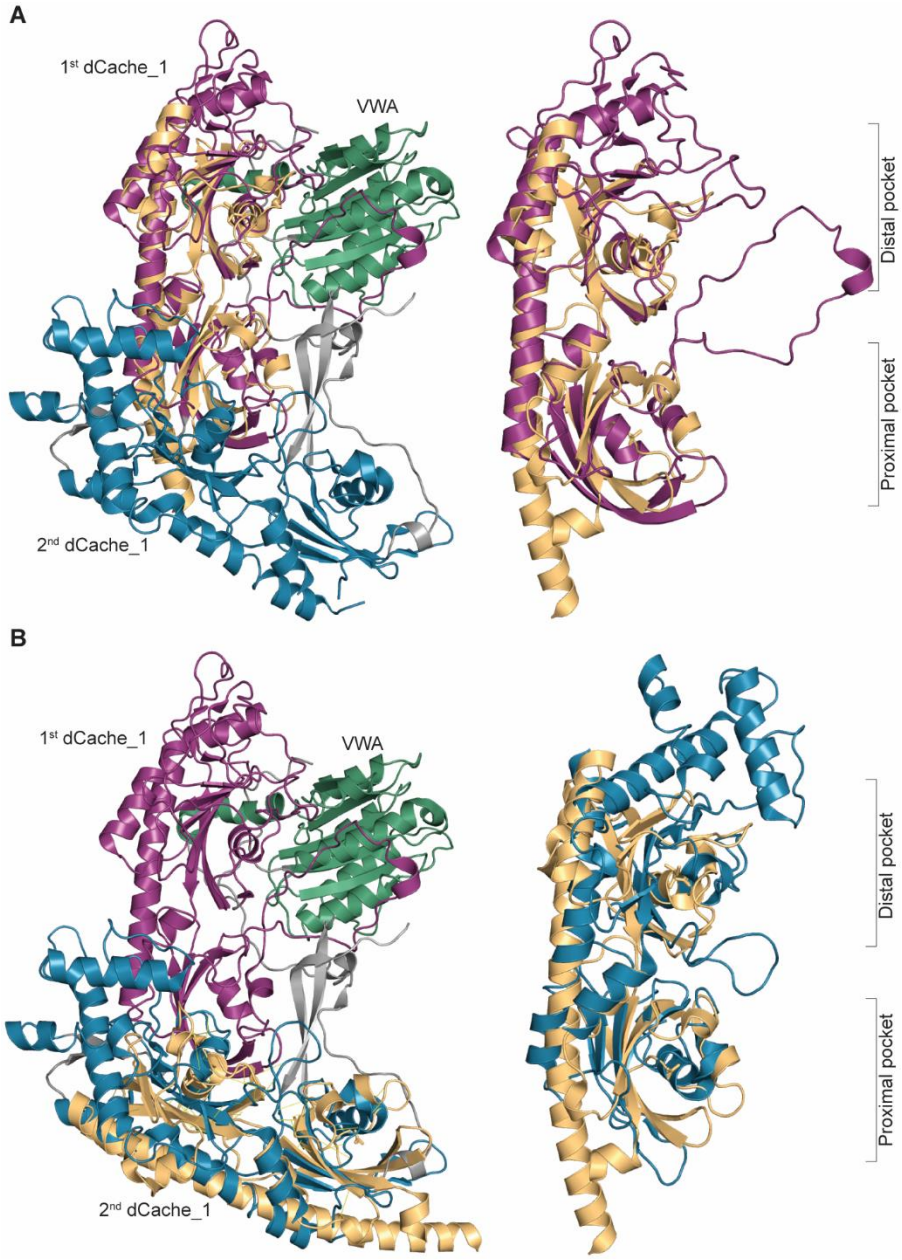


Fig. S12. Superimposition of the *P. aeruginosa* PAO1 PctA dCache_1 (PDB ID 5T65; gold) and two dCache_1 domains of the rabbit α2δ-1 subunit: (A) with the 1st (N-terminally located) dCache_1 domain (purple) and (B) with the 2nd (C-terminally located) dCache_1 domain (blue). On the right corresponding superimpositions are enlarged.

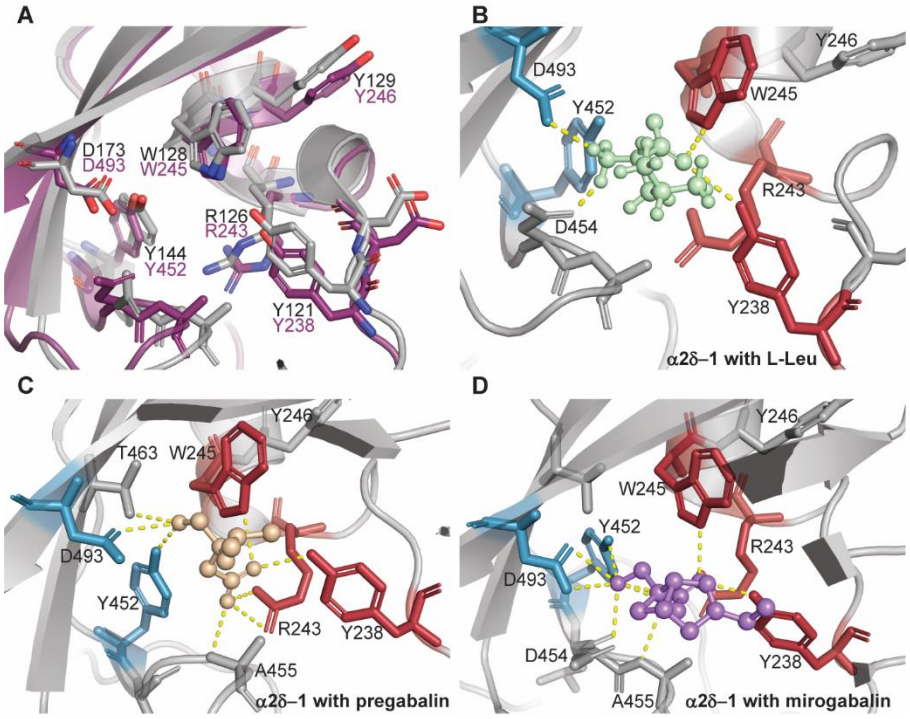


Fig. S13. Ligand binding pocket (LBP) of the α2δ-1 1st dCache_1 domain. (**A**) Structural overlay of the AA_motifs from α2δ-1 1st dCache_1 (purple) with *P. aeruginosa* PctA dCache_1 (grey, PDB ID: 5T65). (**B-D**) LBP of the α2δ-1 1st dCache_1 domain docked with L-Leu (**B,** light green), pregabalin (**C,** beige), and mirogabalin (**D,** violet). The first Y, R and W of the AA_motif are shown in red, and third Y and D are shown in blue.

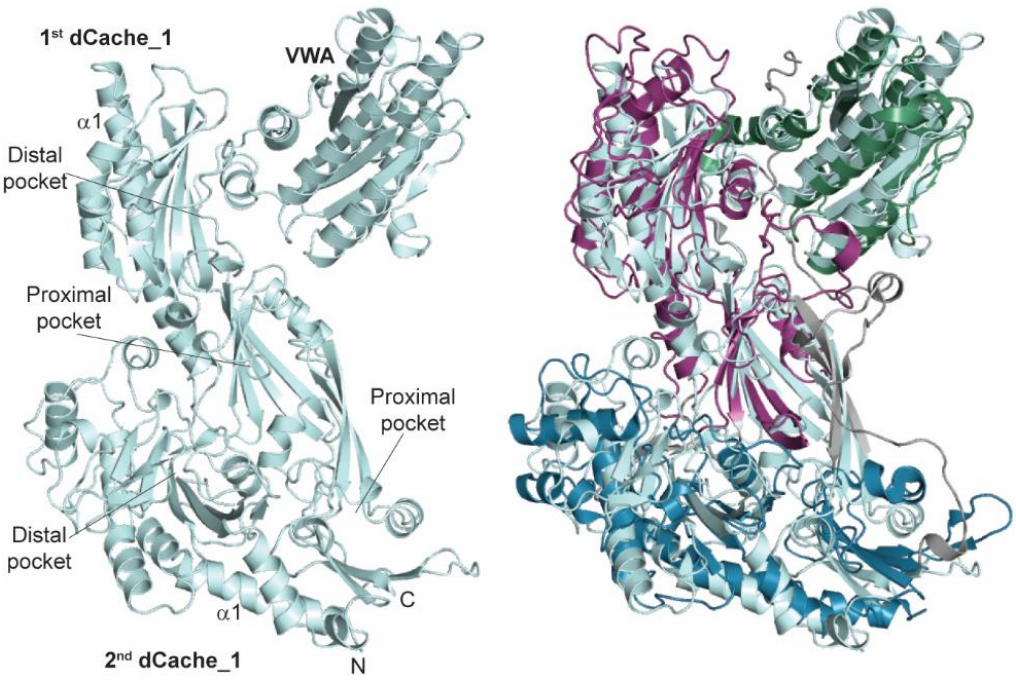


Fig. S14. **Structural overlay of human CACHD1 protein modeled by AlphaFold (pale cyan) with rabbit α2δ-1 protein** (colored as in Fig. 3).

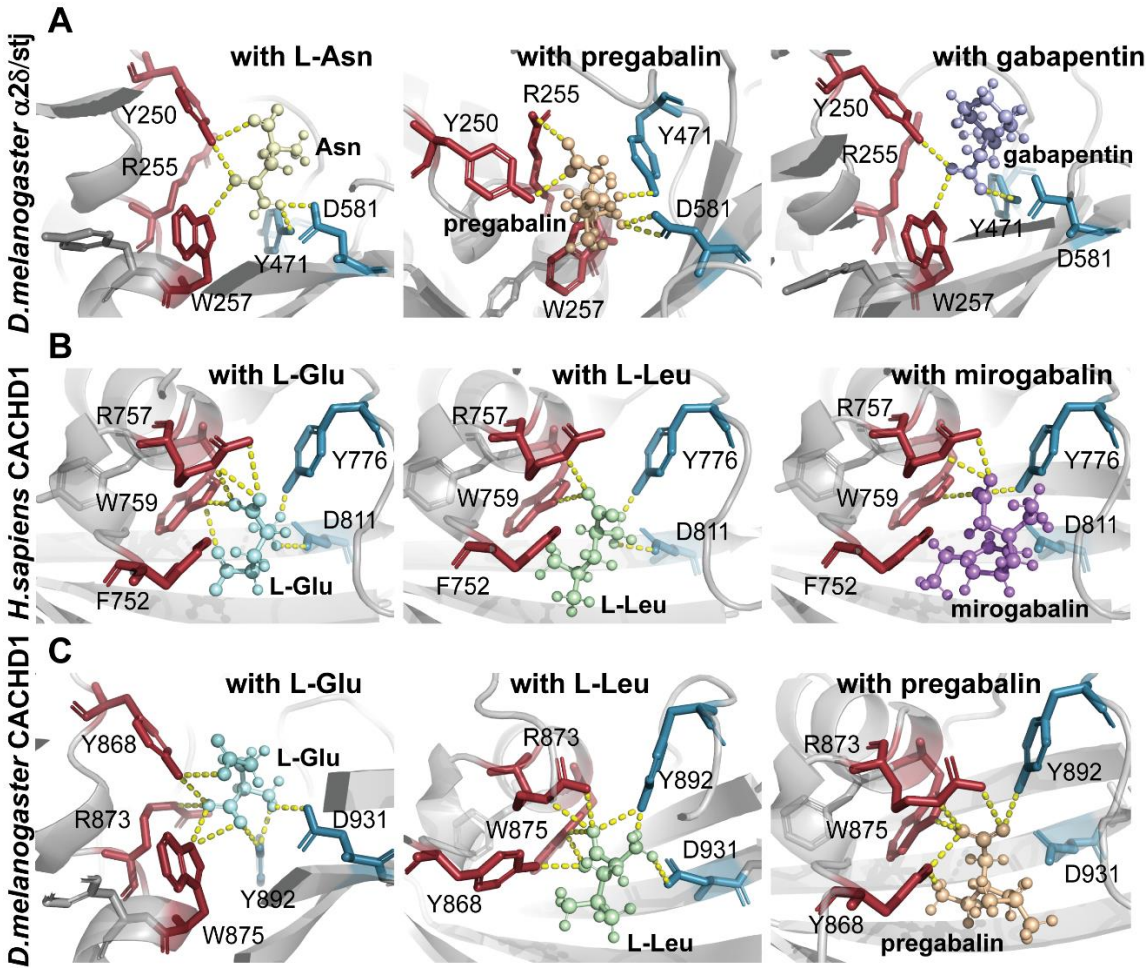


Fig. S15. Ligand-binding pockets of modelled human CACHD1 and fly CACHD1 and α2δ proteins docked with amino acid ligands and their derivative drugs. The proteins were modelled by AlphaFold. In all cases ligands made hydrogen bonds with the AA_motif. (**A**) Human CACHD1 protein, (**B**) *D. melanogaster* α2δ protein (annotated in NCBI database as “straightjacket”), (**C**) *D. melanogaster* CACHD1 protein.

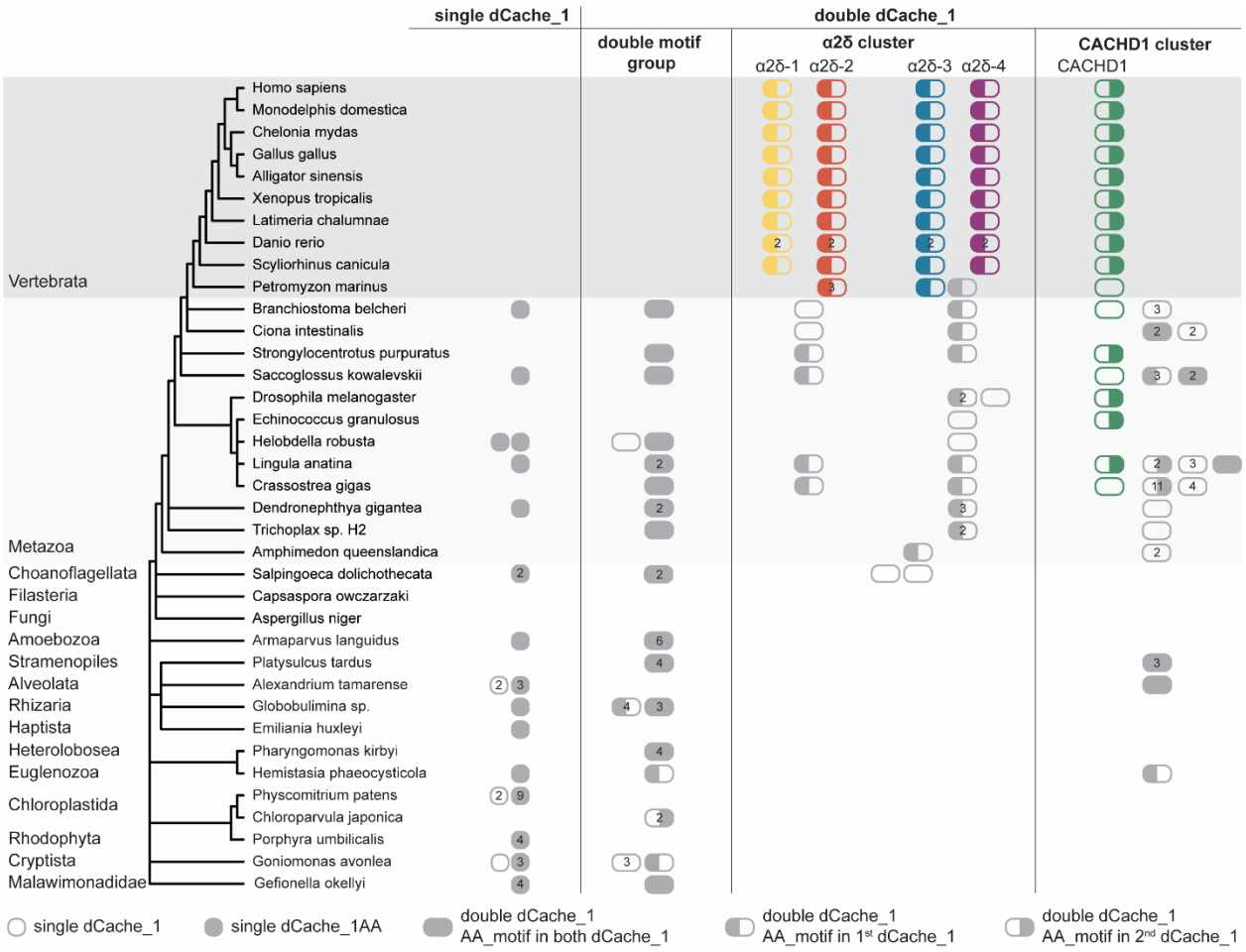


Fig. S16. Evolutionary history of dCache_1 domain containing proteins across major eukaryotic lineages. Presence of the intact AA_motif is denoted by color. When the AA_motif is intact in both dCache_1 domains the entire circle or ellipse is filled. Empty circles and ellipses denote proteins with no AA_motif in either of the dCache_1 domains. Numbers indicate that the genomes of corresponding organisms encode the specified number of paralogous proteins.

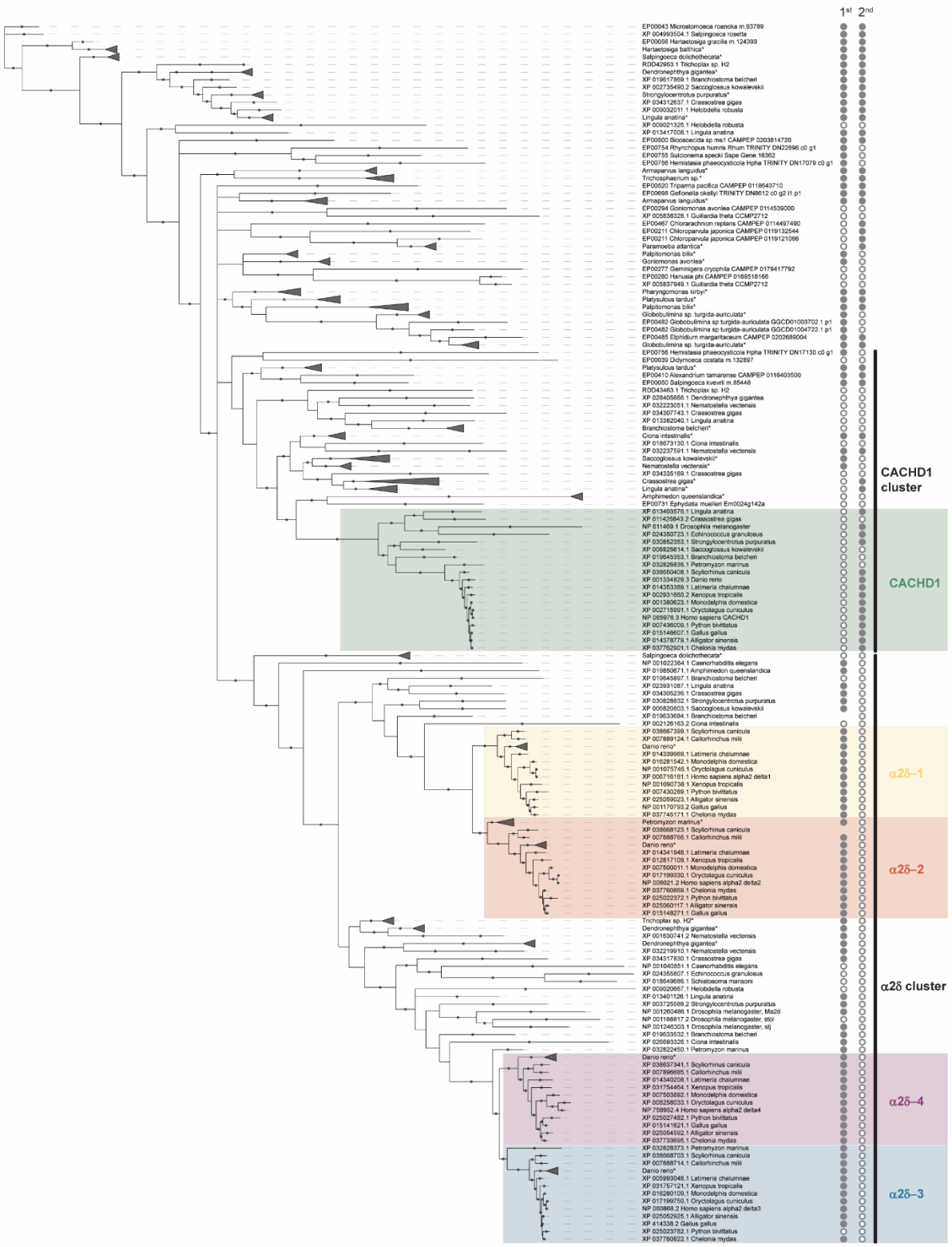


Fig. S17. Phylogenetic tree of eukaryotic double dCache_1 domain containing proteins from representative set of eukaryotic species. The tree was built using Bayesian inference. Grey circles indicate the presence (filled circle) or the absence (empty circle) of the AA_motif in the 1st and 2nd

dCache_1 domains. Black dots on the tree branches mark the probability values that are ≥0.9 (probabilities are in the range 0 to 1). Full-size figure link:

https://github.com/ToshkaDev/Motif/blob/main/Eukaryotes/S6_tree.tif?raw=true

Table S1. Experimentally studied amino acid binding proteins containing dCache_1 domain with AA_motif. For each protein its identifier, ligand repertoire, affinities, and amino acid binding motif variant is shown. For proteins with solved structure contacts with the ligand carboxyl and amino group are shown.

Organism	Protein Name	Protein ID	Reference	PDB ID	Ligand	K_D (μM)	AA_motif; hydrogen bonds with the ligand carboxyl group are in **red** and amino group are in **blue**	Mutagenic analysis
Pseudomonas aeruginosa PAO1	PctA	NP_252999.1	(30, 38-40)		L-Arg	1.8 ± 0.4	**Y....R.WY Y D**	Present study: Mutation of Y....R.WY Y **D** to A abolished L-Ala and L-Ser binding Mutation of Y....R.WY Y **D** to N strongly reduced L-Ala binding Mutation of Y....**R**.WY Y D to A reduced L-Ala and L-Ser binding by 61- and 78-fold, respectively.
					L-Lys	1.6 ± 0.1		
					L-Tyr	5.2 ± 1.7		
				5T7M	L-Trp	2.3 ± 0.9		
					L-Phe	1.1 ± 0.2		
					L-Ala	0.72 ± 0.1		
				5T65	L-Ile	24 ± 5.6		
					L-Leu	116 ± 10		
				5LTX	L-Met	0.91 ± 0.2		
					L-Asn	2.0 ± 0.2		
					L-Ser	1.2 ± 0.2		
					L-Cys	0.79 ± 0.1		
					L-Thr	0.28 ± 0.1		
					L-His	28 ± 5		
					L-Pro	0.60 ± 0.2		
					L-Gly	21 ± 7.7		
					L-Val	0.34 ± 0.2		
P. aeruginosa PAO1	PctB	NP_253000.1	(30, 38, 39, 41)	5LT9	L-Arg	64 ± 4	**Y....R.WY Y D**	
					L-Lys	1096 ± 88		
					L-Ala	641 ± 50		
					L-Met	46 ± 2		
				5LTO	L-Gln	1.2 ± 0.1		
P. aeruginosa PAO1	PctC	NP_252997.1	(30, 38, 39, 41)	5LTV	GABA	1.2 ± 0.3	**Y....R.WY F D**	
					L-His	17 ± 3		
					L-Pro	80 ± 6		
P. putida KT2440	McpG	WP_010952482.1	(42)		GABA	0.175	**Y....R.WY Y D**	
P. putida KT2440	McpA		(43)		Gly	35 ± 2	**Y....R.WY Y D**	

Organism	Protein Name	Protein ID	Reference	PDB ID	Ligand	K_D (μM)	**AA_motif;** hydrogen bonds with the ligand carboxyl group are in **red** and amino group are in **blue**	Mutagenic analysis
		WP_010953225.1			L-Ala	13 ± 0.6		
					L-Cys	0.6 ± 0.1		
					L-Ser	43 ± 2		
					L-Asn	4.3 ± 0.2		
					L-Gln	5.5 ± 0.2		
					L-Phe	2.3 ± 0.1		
					L-Tyr	12.1 ± 0.9		
					L-Val	373 ± 42		
					L-Ile	85 ± 5		
					L-Met	5.8 ± 0.6		
					L-Arg	1.2 ± 0.1		
Sinorhizobium meliloti RU11/001	McpU	WP_010968886.1	(44)		Pro	104	**Y....R.WY Y D**	mutation of Y....R.WY Y **D** to A abolished taxis toward Pro; mutation of Y....R.WY Y **D** to E showed intermediary response
Sinorhizobium meliloti RU11/001	McpU	WP_010968886.1	(45)		Arg	350	**Y....R.WY Y D**	
					Phe	53		
					Trp	34		
					Pro	42		
P. fluorescens Pf0-1	CtaA	WP_011335661.1	(46)	6Q0F	L-Val	4.7 ± 1.2	**F....R.WY Y D**	
				6PXY	L-Ala	5.2 ± 0.3		
				6PY5	L-Ser	9.5 ± 1.1		
				6PY4	L-Leu	11.9 ± 1.8		
				6Q0G	L-Pro	13.5 ± 0.3		
				6PY3	L-Ile	27.4 ± 0.9		
					L-Arg	446 ± 17.6		
Bacillus velezensis SQR9	McpC	AHZ15354.1	(47)		L-Leu	3.6 ± 0.78	**Y....R.WY Y D**	
					L-Pro	3.6 ± 0.78		
B. velezensis SQR9	TlpB	AHZ17109.1	(47)		L-Phe	3.17± 0.89	**Y....R.WY Y D**	
B. subtilis OI1085	McpB	WP_003243461.1	(48)		L-Asn	14	**Y....R.WY Y D**	mutation of **Y**....R.WY Y D to A showed defects in taxis

Organism	Protein Name	Protein ID	Reference	PDB ID	Ligand	K_D (μM)	**AA_motif;** hydrogen bonds with the ligand carboxyl group are in **red** and amino group are in **blue**	Mutagenic analysis
								toward Asp; mutation of **Y**....R.WY Y D to F had no effect
B. subtilis OI1085	McpC	WP_003245443.1	(49)		L-Pro	14	**Y....R.WY Y D**	
					L-Thr	21		
					L-Ser	90		
					L-Val	154		
					L-Ala	18		
					L-Tyr	360		
					L-Phe	492		
					L-Leu	72		
					L-His	320		
Vibrio cholerae O395N1	Mlp37	AAF96820.1	(50)	5AVF	taurine	3.2	**W....R.WY Y D**	mutations of: **W**....**R**.**W**Y **Y D** to A showed defects in taxis toward Ala, Ser and taurine
				5AVE	L-Ser	3.6		
				3C8C	L-Ala	2.7		
				6IOV	L-Arg	5.6		
V. cholerae O395N1	Mlp24	WP_001212589.1	(51)		L_Met	4.7	**Y....R.WY Y D**	
				6IOT	L-Arg	4.8		
					L-Ala	5.1		
					L-Gln	5.5		
				6IOR	L-Asn	6.3		
					L-Lys	6.8		
				6IOU	L-Ser	18.3		
					L_Val	22.1		
				6IOS	L-Pro	23.8		
				6IOQ	Gly	9		
P. syringae pv. actinidiae	PscC	WP_017684350.1	(52)	6MNI	Pro	5 ± 0.9	**Y....R.WY Y D**	
P. syringae pv. Tomato DC3000	PscA	WP_011104030.1	(53)		L-Asp	1.2 ± 0.1	**Y....R.WY Y D**	
					L-Glu	3.4 ± 0.1		
P. syringae pv. Tomato DC3000	PscA	WP_011104030.1	(54)		L-Asp	6.1	**Y....R.WY Y D**	
					L-Glu	27		
Campylobacter jejuni subsp. jejuni ATCC 700819	Tlp3/ CcmL	YP_002344933.1	(55)	6W3V	L-Phe	730 ± 55	**L....K.WY Y D**	mutations of: L....**K**.**W**Y **Y D** to A abolished Ile binding
				4XMR	L-Ile	86 ± 10		
					L-Leu	105 ± 6		

Organism	**Protein Name**	**Protein ID**	**Reference**	**PDB ID**	**Ligand**	***K*$_D$ (μM)**	**AA_motif;** hydrogen bonds with the ligand carboxyl group are in **red** and amino group are in **blue**	**Mutagenic analysis**
					L-Val	405 ± 27		
Aeromonas caviae isolate ZOR0002	SpdE	WP_052815311.1	(56)	7K5N	L-Pro		**F....R.WY Y D**	Mutations of: F....R.**W**Y **Y** D to A reduced thermal stabilization by ligands
					L-Val			
					L-Ile			

Table S2. Microcalorimetric studies of recombinant dCache_1AA domains from selected bacterial and archaeal species.

Organism, Phylum or Class	Protein ID	Cellular function	Motif	Compound	Protein concentration (μM)	Compound concentration (mM)	K_D (μM)	ΔH (kcal · mol^{-1})
Methanospirillum hungatei Archaea, Halobacteriota	WP_011449640.1	Sensor histidine kinase	Y….R.WY Y D	L-Glutamic acid	73	5	2816 ± 3450	-2.40 ± 280.30
Thermodesulfobacterium thermophilum Desulfobacterota	WP_162138226.1	c-di-GMP cyclase	Y….R.WY Y D	L-Proline	54	0.25	0.21 ± 0.01	-10.72 ± 0.10
				D-Valine	54	1	1.56 ± 0.24	-1.66 ± 0.12
				L-Isoleucine	54	1	2.20 ± 0.21	-1.29 ± 0.07
				L-Valine	54	1	2.25 ± 0.33	-2.76 ± 0.35
Yersinia pestis Gammaproteobacteria pathogen	WP_016674185.1	Chemoreceptor	Y….R.WY Y D	L-Alanine	42	0.5	2.61 ± 0.19	-7.98 ± 0.25
				L-Threonine	42	0.5	2.32 ± 0.10	-11.44 ± 0.22
				L-Serine	42	0.5	3.20 ± 0.16	-14.58 ± 0.46
				L-Valine	42	0.35	0.96 ± 0.08	-8.54 ± 0.21
				L-Arginine	42	0.35	1.37 ± 0.12	-11.39 ± 0.51
Tautonia marina Planctomycetota	WP_152054232.1	Serine/threonine phosphatase	Y….R.WY Y D	L-Proline	30	0.5	0.52 ± 0.02	-10.72 ± 0.14
				D,L-Homoserine	30	0.5	9.80 ± 0.79	-6.81 ± 1.05
Enhygromyxa salina Myxococcota	WP_106093935.1	Serine/threonine kinase	Y….R.WY Y D	L-Glutamic acid	15	2	48.31 ± 0.73	-20.16 ± 1.46
				L-Aspartic acid	15	2	15.26 ± 0.41	-9.13 ± 0.27
				D-Aspartic acid	15	2	162.34 ± 6.00	-12.91 ± 4.58
Treponema denticola Spirochaeta pathogen	WP_002687321.1	Chemoreceptor	F….R.WY Y D	L-Arginine	48	0.5	0.52 ± 0.01	-11.46 ± 0.07
				L-Histidine	48	0.5	1.35 ± 0.11	-11.41 ± 0.18
Vibrio cholerae Gammaproteobacteria pathogen	NP_233280.1	Metal-dependent phosphodiesterase	F….R.WY Y D	L-Glutamic acid	37	0.75	2.17 ± 0.09	-19.11 ± 0.57
				L-Glutamine	37	0.75	8.54 ± 0.32	-43.55 ± 2.54
				L-Histidine	42	0.75	6.62 ± 0.44	-11.60 ± 0.94
				L-Tryptophan	42	0.75	3.33 ± 0.23	-8.93 ± 0.62
Legionella pneumophila	WP_154766400.1	guanylate/	Y….R.WY Y D	L-Alanine	11.5	0.35	5.74 ± 0.59	-9.74 ± 2.55

Gammaproteobacteria pathogen		adenylate cyclase		L-Serine	11.5	10	23.04 ± 8.81	0.35 ± 0.28
				L-Valine	11.5	10	211.86 ± 65.53	3.60 ± 44.67

Table S3. Corresponding amino acid residues of the AA_motif in the rabbit α2δ-1 and in chemoreceptors of *P. aeruginosa* PAO1.

AA_motif residue	**α2δ-1** (rabbit)	**PctA/PctB** (bacterium)	**PctC** (bacterium)
Y….R.WY[n1]Y[n2]D	Y238	Y121	Y124
Y….**R**.WY[n1]Y[n2]D	R243	R126	R129
Y....R.**W**Y[n1]Y[n2]D	W245	W128	W131
Y....R.WY[n1]**Y**[n2]D	Y452	Y144	F147
Y....R.WY[n1]Y[n2]**D**	D493	D173	D176

Table S4. Frequencies and variations of the AA_motif found in Bacteria and Archaea in the GTDB representative dataset. For Bacteria top ten variants are shown. Complete data are in Table S3, sheets 3 and 4. NF – not found.

Motif variant	Protein count in Bacteria	Protein count in Archaea
Y....R.WY[n1]Y[n2]D	62.8%	58.3%
F....R.WY[n1]Y[n2]D	17.2%	11.3%
Y....R.WF[n1]Y[n2]D	7%	7.0%
W....R.WY[n1]Y[n2]D	3.7%	NF
F....R.WF[n1]Y[n2]D	2.9%	23.5%
Y....K.WY[n1]Y[n2]D	1.5%	NF
Y....R.WY[n1]F[n2]D	0.9%	NF
F....R.WY[n1]F[n2]D	0.7%	NF
L....R.WY[n1]Y[n2]D	0.6%	NF
W....R.WF[n1]Y[n2]D	0.5%	NF

Table S5. Analysis buffers compositions.

Protein ID	Analysis buffer composition
NP_252999.1 (PctA-LBD)	5 mM Tris, 5 mM PIPES, 5 mM MES, 0.15 M NaCl, 10% glycerol (vol/vol), pH 7.0
WP_011449640	3 mM Tris, 3 mM PIPES, 3 mM MES, 0.15 M NaCl, 10% glycerol (vol/vol), pH 7.0
WP_154766400	3 mM Tris, 3 mM PIPES, 3 mM MES, 0.15 M NaCl, 10% glycerol (vol/vol), pH 7.0
WP_152054232	3 mM Tris, 3 mM PIPES, 3 mM MES, 0.15 M NaCl, 10% glycerol (vol/vol), pH 6.0
WP_162138226	3 mM Tris, 3 mM PIPES, 3 mM MES, 0.15 M NaCl, 10% glycerol (vol/vol), pH 7.5
WP_106093935	3 mM Tris, 3 mM PIPES, 3 mM MES, 0.15 M NaCl, 10% glycerol (vol/vol), pH 7.0
WP_002687321	3 mM Tris, 3 mM PIPES, 3 mM MES, 0.15 M NaCl, 10% glycerol (vol/vol), pH 7.0
NP_233280	3 mM Tris, 3 mM PIPES, 3 mM MES, 0.15 M NaCl, 10% glycerol (vol/vol), pH 8.0
WP_016674185	3 mM Tris, 3 mM PIPES, 3 mM MES, 0.15 M NaCl, 10% glycerol (vol/vol), pH 7.0

Table S6. Strains, plasmids and oligonucleotides used in this study.

	Relevant characteristics		**Reference or source**
Strains			
E. coli BL21 (DE3)	F^-, *ompI, hsdS*$_B$ (r^-_B m^-_B) *gal*, *dam*, *met*		(36)
E. coli DH5α	*supE44 lacU169* (*Ø80lacZ*ΔM15) *hsdR17* ($r_K^-m_K^-$), *recA1 endA1 gyrA96 thi-1 relA1*		(37)
Plasmids			
pET28b(+)	Km^R; Protein expression plasmid		Novagen
pET28-PctA-LBD	Km^R; pET28b(+) derivative containing DNA fragment encoding PctA-LBD		(30)
pMAMV359	Km^R; pET28b(+) derivative containing DNA fragment encoding PctA-LBD (R126A)		Present study
pMAMV360	Km^R; pET28b(+) derivative containing DNA fragment encoding PctA-LBD (D173A)		Present study
pMAMV376	Km^R; pET28b(+) derivative containing DNA fragment encoding PctA-LBD (D173N)		Present study
Oligonucleotides			
Name	**Sequence (5´-3´)**	**Purpose**	**Source**
PctA-LBD-F	GTTCACCCATATGAACGATTACCTGCAGCGCAACG	*pctA-LBD* into pET28b(+)	(30)
PctA-LBD-R	GCGGATCCTCAGGCCGAGACGCGGAACTTG	*pctA-LBD* into pET28b(+)	(30)
R126A-F	TCCGCGCAGCGCGCCCTGGTACAAG	Mutant R126A	Present study
R126A-R	CTTGTACCAGGGCGCGCTGCGCGGA	Mutant R126A	Present study
D173A-F	GTAGGCGGCGCCCTCAGCCTGAAG	Mutant D173A	Present study
D173A-R	CTTCAGGCTGAGGGCGCCGCCTAC	Mutant D173A	Present study
D173N-F	GTAGGCGGCAACCTCAGCCTGAAG	Mutant D173N	Present study
D173N-R	CTTCAGGCTGAGGTTGCCGCCTAC	Mutant D173N	Present study

Dataset S1. The AA_motif variants across Bacteria and Archaea.

Dataset S2. Domain compositions of prokaryotic dCache_1AA containing proteins in the GTDB representative set.

Dataset S3. Eukaryotic dCache_1AA containing proteins.

Dataset S4. Results of BLASTP and PSI-BLAST searches using eukaryotic dCache_1AA against prokaryotic protein database.

Dataset S5. Molecular docking of rabbit α2δ-1, modeled human CACHD1, and modeled fly α2δ-1 and CACHD1 proteins with α-amino acids, GABA, gabapentin, pregabalin, and mirogabalin.