Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Environ Microbiol. Author manuscript; available in PMC 2024 January 01.

Published in final edited form as:

Environ Microbiol. 2023 January ; 25(1): 11–16. doi:10.1111/1462-2920.16143.

Is it possible to predict signal molecules that are recognised by bacterial receptors?

Miguel A. Matilla[1], **Elizabet Monteagudo-Cascales**[1], **Jean Paul Cerna-Vargas**[1], **Vadim M. Gumerov**[2], **Igor B. Zhulin**[2,#], **Tino Krell**[1,#]

[1]Department of Environmental Protection, Estación Experimental del Zaidín, Consejo Superior de Investigaciones Científicas, Prof. Albareda 1, 18008 Granada, Spain

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

The bottleneck:

The ability to adapt to changing environmental conditions is essential for bacterial survival. Bacteria have evolved a wide array of signal transduction systems that sense signals and generate cellular responses. Major families of signal transduction proteins include transcriptional regulators, sensor histidine kinases, chemoreceptors, (di)nucleotide cyclases, cyclic (di)nucleotide phosphodiesterases, extracytoplasmic function sigma factors, Ser/Thr/Tyr protein kinases and phosphoprotein phosphatases (Galperin, 2018; Gumerov *et al.*, 2020). The regulatory outputs of these systems are diverse and include regulation of gene expression, chemotaxis, and modulation of second messenger levels.

Although signal transduction systems differ in their composition and molecular mechanisms, in the canonical activation pathway a signal (for example, a small molecule) interacts with a sensor domain of the receptor protein, which leads to modulation of the activity of enzymatic domains such as the autokinase domain of sensor kinases or the GGDEF and EAL domains of diguanylate cyclases and phosphodiesterases, respectively. Hundreds of different sensor domains have evolved (Ortega *et al.*, 2017; Matilla *et al.*, 2022a) and novel domains are discovered regularly (Elgamoudi *et al.*, 2021; Martin-Rodriguez *et al.*, 2022). Majority of sensor domains are ligand-binding modules that contain all the determinants necessary for ligand recognition, as demonstrated in studies showing that the ligand affinities to full-length receptors and the individual sensor domains are comparable (Foster *et al.*, 1985; Milligan and Koshland, 1993). The same type of sensor domain is frequently found in different signal transduction systems (Ulrich *et al.*, 2005), indicating that these modules have been exchanged and recombined with different sensor proteins during evolution.

The phenotypic analysis of bacterial mutants of signal transduction proteins provides valuable information on the function of the corresponding regulatory circuits. However, these systems are frequently expressed and activated in the presence of specific environmental stimuli, which often hinders the phenotypic characterisation of mutants as

[#]Address correspondence to Igor Zhulin, jouline.1@osu.edu, or Tino Krell, tino.krell@eez.csic.es.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

for the vast majority of signal transduction systems the signal molecule(s) is unknown. Therefore, knowledge about the signals detected by receptors is indispensable for understanding the physiological significance of regulatory circuits and development of anti-infective approaches aimed at reducing bacterial virulence by interfering with signal transduction cascades (Krell and Matilla, 2022).

There are several problems that hamper the identification of signal molecules, including: i) ligand screening is frequently highly labour intensive, ligand libraries are costly and may not contain the relevant compounds; ii) sensor domains of the same family often show a high degree of sequence divergence, impeding an extrapolation of the ligand recognized from characterized systems; iii) there are a number of non-canonical sensing mechanisms that are not based on a direct ligand interaction with sensor domains.

The question:

Can signals recognized by sensor domains be predicted from their protein sequences?

The answers

1) Analysing the overall sequence similarities and sensor domain types

There are millions of sensor domain sequences available in public databases. In the first approach, we wanted to establish to what degree ligand specificity correlates with individual sensor domain types. We compiled a catalogue of signal molecules that were demonstrated to directly bind to sensor domains of transcriptional regulators, chemoreceptors and sensor kinases (Matilla *et al.*, 2022a). These domains were subsequently classified according to their Pfam families (Mistry *et al.*, 2021). Whereas canonical transcriptional regulators recognize their cognate signals in the cytosol, chemoreceptors and sensor kinases possess frequently extracytosolic sensor domains. As for the extracytosolic sensor domains, no clear pattern emerged relating a given signal type with a sensor domain family (Matilla *et al.*, 2022a). This may be exemplified by the two most abundant extracytosolic sensor domain families, dCache and the four-helix bundle domains (Ulrich and Zhulin, 2005; Upadhyay *et al.*, 2016; Sanchis-Lopez *et al.*, 2021). dCache domains were shown to bind a wide range of structurally different signal molecules including amino acids, polyamines, purines, quaternary amines, organic acids, sugars or metal oxanions (Matilla *et al.*, 2022a). Similar observations were made for the four-helix bundle domains that recognize different amino and organic acids, aromatic hydrocarbons, benzoate derivatives or borate (Matilla *et al.*, 2022a). No such relationships were observed for the remaining, less abundant extracytosolic sensor domains analysed (Matilla *et al.*, 2022a).

However, the conclusions drawn from the analysis of 87 families of sensor domains present in transcriptional regulators and 16 families of single-domain transcriptional regulators were somewhat different (Matilla *et al.*, 2022a). In analogy to the extracytosolic sensor domains, most domain families of these regulators respond to diverse types of signals. For example, the sensor domain of the highly abundant LysR type transcriptional regulators was shown to bind structurally very diverse compounds, including amino and organic acids, sugar phosphates, flavonoids, aromatic compounds, peptides, NADPH, ATP, c-di-

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

GMP, ppGpp, HOCl, H_2O_2 or fatty acid CoA indicating the absence of a signal type – domain type relationship (Matilla *et al.*, 2022a). However, some other sensor domains or single-domain regulators were found to be highly specific for a given signal type (Fig. 1). Next to a significant number of domains/proteins that recognized specifically metal ions and sugars (or sugar derivatives), there were three well populated families, namely AsnC_trans_reg (25 characterised proteins), CodY (14 characterised proteins) and Arg_repressor_C (11 characterised proteins), that showed very strong preference for amino acids. Furthermore, the sensor domains Aminotran_1_2 and Autoind_bind appear to have evolved to specifically recognize pyridoxal-5'-phosphate and acyl homoserine lactones, respectively. The information shown in Fig. 1, detailing the signal domain-signal type relationships, provides valuable information for the design of experiments aimed at establishing the signals recognized by a given signal transduction system.

2) Defining ligand binding amino acid motifs

As mentioned above, the signal type recognized by extracytosolic sensor domains is not reflected in overall sequence similarity. However, recent advances in structural and computational biology have permitted the prediction of ligands that are recognised by sensor domains, regardless of their overall sequence identity with the characterised domains. In a previous study we reported the 3D structures of the dCache_1 domains of the *Pseudomonas aeruginosa* chemoreceptors, PctA, PctB and PctC, that bind different amino acids (Gavira *et al.*, 2020). The comparison of the amino acid residues involved in ligand binding enabled the identification of a conserved sequence motif in these three dCache_1 domains (Gavira *et al.*, 2020). In a subsequent study (Gumerov *et al.*, 2022), we showed that this motif was also present in a number of other amino acid responsive dCache domains from phylogenetically diverse species such as chemoreceptors Mlp24 and MLP37 of *Vibrio cholerae* (Takahashi *et al.*, 2020), Tlp3 of *Campylobacter jejuni* (Liu *et al.*, 2015), McpU of *Sinorhizobium meliloti* (Webb *et al.*, 2017) or McpC and McpB of *Bacillus subtilis* (Glekas *et al.*, 2010; Glekas *et al.*, 2012). In marked contrast, this motif could not be detected in dCache domains that bind compounds other than amino acids, including the quaternary amine receptors McpX (Shrestha *et al.*, 2018) and PctD (Matilla *et al.*, 2022b), the polyamine responsive chemoreceptors McpU (Gavira *et al.*, 2018) and TlpQ (Corral-Lugo *et al.*, 2018), the purine chemoreceptor McpH (Fernandez *et al.*, 2016) and the organic acid binding KinD (Wu *et al.*, 2013), DctB (Cheung and Hendrickson, 2008) and Htc1 (Gasperotti *et al.*, 2020).

Thus, the study demonstrated the existence of a sequence motif specific for amino acid responsive dCache_1 domains. This motif consists of three amino acids, Y121, R126 and W128 (PctA numbering), that interact with the carboxylic moiety of the bound amino acid, whereas Y144 and D173 coordinate the amino group of the ligand (Fig. 2A, B) (Gumerov *et al.*, 2022). Replacement of these amino acids with alanine resulted in either no or strongly reduced amino acid binding (Fig. 2C). Sequence database searches of dCache_1 domains containing this sequence motif resulted in the identification of more than 10 000 bacterial and archaeal proteins (Gumerov *et al.*, 2022). Interestingly, sensor domains with this sequence motif were also detected in eukaryotes. Although the Pfam profile Hidden Markov models did not recognize these domain in eukaryotic proteins as member of the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

dCache family, structural analysis and computational modelling clearly indicated that they have the typical dCache_1 fold (Gumerov *et al.*, 2022).

A number of experiments were conducted to verify whether the identified domains indeed bind amino acids (Gumerov *et al.*, 2022). As for the eukaryotic proteins, the site-directed mutagenesis of the residues of this conserved motif in one of the proteins, the α2δ−1-subunit of voltage-gated calcium channels, resulted in a significant reduction in biological activity. A different strategy was used to study ligand binding to prokaryotic and archaeal amino acid binding dCache_1 containing proteins. The predicted amino-acid responsive dCache_1 sensor domains were generated as individual, purified proteins that were then subjected to differential scanning fluorimetry based thermal shift assays followed by isothermal titration calorimetry ligand binding studies (Fernandez *et al.*, 2018; Matilla *et al.*, 2020). As shown in Fig. 3, proteins from phylogenetically diverse microorganisms, including bacteria belonging to different phyla (e.g. γ-Proteobacteria, Spirochaeta, Desulfobacterota, Myxococcota and Planctomycetota) and Archaea, were selected as potential targets. These proteins were also selected to cover the major families of bacterial transmembrane receptors (Galperin, 2018), namely chemoreceptors, sensor histidine kinases, c-di-GMP cyclases and phosphodiesterases, serine/threonine kinases and phosphatases as well as guanylate/adenylate cyclases (Fig. 3). Importantly, amino acid binding was detected for all of the selected proteins (Gumerov *et al.*, 2022). Ligand screening showed that in most cases these domains recognize proteinogenic amino acids, whereas some domains bound alternative amino acids such as D-Val, D-Asp, and D,L-homoserine. For most of the proteins analysed, amino acids showed tight binding, with dissociation constants in the nanomolar or lower micromolar range (Gumerov *et al.*, 2022), indicating that the corresponding receptors mediate high-sensitivity responses to amino acids. Thus, we showed that amino acid responsive dCache domains are found throughout the Tree of life. The primary physiological relevance of chemotaxis consists in accessing nutrients (Colin *et al.*, 2021) and the observation that there are many amino acid binding chemoreceptors, permitting chemoattraction to amino acids, may highlight the nutritional values of these ligands. However, the fact that amino acid-binding sensor domains are also found in many other types of transmembrane receptors supports the notion that amino acids are key signal molecules that provide the bacterium with important information about their environment.

Conclusions and future outlook

The Pfam database (Mistry *et al.*, 2021) contains hundreds of different sensor domain families. Particularly over the last decade, there has been a significant increase in the number of deposited three-dimensional structures of sensor domains in complex with their respective ligands, permitting the identification of key residues involved in signal recognition. Such information forms the basis for analogous studies to computationally predict and experimentally verify ligands recognized by sensor domains of unknown function. The determination of three-dimensional structures remains labour-intensive and challenging for certain proteins, but the recent development of computational approaches for highly accurate protein structure prediction (Jumper *et al.*, 2021) and deep neural networks to predict protein functions (Sanderson *et al.*, 2022) are alternative approaches to identify

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

amino acids involved in signal binding. Scarcity of information on the signals that stimulate bacterial receptors is currently a major bottleneck that limits our understanding of many regulatory circuits, but novel *in vivo*, *in vitro* and *in silico* approaches have a significant potential to advance our knowledge of bacterial and archaeal signal transduction.

Acknowledgements

This work was supported by grants PID2019–103972GA-I00 (to MAM) and PID2020–112612GB-I00 (to TK) from the Spanish Ministry for Science and Innovation/*Agencia Estatal de Investigación* 10.13039/501100011033, grant P18-FR-1621 (to TK) from the Junta de Andalucía, and NIH grant R35GM131760 (to IBZ).

References

Cheung J, and Hendrickson WA (2008) Crystal structures of C4-dicarboxylate ligand complexes with sensor domains of histidine kinases DcuS and DctB. J Biol Chem 283: 30256–30265. [PubMed: 18701447]

Colin R, Ni B, Laganenka L, and Sourjik V (2021) Multiple functions of flagellar motility and chemotaxis in bacterial physiology. FEMS Microbiol Rev 45: fuab038.

Corral-Lugo A, Matilla MA, Martin-Mora D, Silva Jimenez H, Mesa Torres N, Kato J et al. (2018) High-Affinity Chemotaxis to Histamine Mediated by the TlpQ Chemoreceptor of the Human Pathogen *Pseudomonas aeruginosa*. MBio 9: e01894–01818. [PubMed: 30425146]

Elgamoudi BA, Andrianova EP, Shewell LK, Day CJ, King RM, Taha et al. (2021) The *Campylobacter jejuni* chemoreceptor Tlp10 has a bimodal ligand-binding domain and specificity for multiple classes of chemoeffectors. Sci Signal 14: eabc8521.

Fernandez M, Morel B, Corral-Lugo A, and Krell T (2016) Identification of a chemoreceptor that specifically mediates chemotaxis toward metabolizable purine derivatives. Mol Microbiol 99: 34–42. [PubMed: 26355499]

Fernandez M, Ortega A, Rico-Jimenez M, Martin-Mora D, Daddaoua A, Matilla MA, and Krell T (2018) High-Throughput Screening to Identify Chemoreceptor Ligands. Methods Mol Biol 1729: 291–301.

Foster DL, Mowbray SL, Jap BK, and Koshland DE Jr. (1985) Purification and characterization of the aspartate chemoreceptor. J Biol Chem 260: 11706–11710. [PubMed: 2995346]

Galperin MY (2018) What bacteria want. Environ Microbiol 20: 4221–4229. [PubMed: 30187651]

Gasperotti AF, Karina Herrera Seitz M, Balmaceda RS, Prosa LM, Jung K, and Studdert CA (2020) Direct binding of benzoate derivatives to two chemoreceptors with Cache sensor domains in *Halomonas titanicae* KHS3. Mol Microbiol 115: 672–683. [PubMed: 33098326]

Gavira JA, Ortega A, Martin-Mora D, Conejero-Muriel MT, Corral-Lugo A, Morel B et al. (2018) Structural Basis for Polyamine Binding at the dCACHE Domain of the McpU Chemoreceptor from *Pseudomonas putida*. J Mol Biol 430: 1950–1963. [PubMed: 29758259]

Gavira JA, Gumerov VM, Rico-Jimenez M, Petukh M, Upadhyay AA, Ortega A et al. (2020) How Bacterial Chemoreceptors Evolve Novel Ligand Specificities. mBio 11: e03066–03019. [PubMed: 31964737]

Glekas GD, Foster RM, Cates JR, Estrella JA, Wawrzyniak MJ, Rao CV, and Ordal GW (2010) A PAS domain binds asparagine in the chemotaxis receptor McpB in *Bacillus subtilis*. J Biol Chem 285: 1870–1878. [PubMed: 19864420]

Glekas GD, Mulhern BJ, Kroc A, Duelfer KA, Lei V, Rao CV, and Ordal GW (2012) The *Bacillus subtilis* chemoreceptor McpC senses multiple ligands using two discrete mechanisms. J Biol Chem 287: 39412–39418. [PubMed: 23038252]

Gumerov VM, Ortega DR, Adebali O, Ulrich LE, and Zhulin IB (2020) MiST 3.0: an updated microbial signal transduction database with an emphasis on chemosensory systems. Nucleic Acids Res 48: D459–D464. [PubMed: 31754718]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Gumerov VM, Andrianova EP, Matilla MA, Page KM, Monteagudo-Cascales E, Dolphin AC et al. (2022) Amino acid sensor conserved from bacteria to humans. Proc Natl Acad Sci U S A 119: e2110415119. [PubMed: 35238638]

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O et al. (2021) Highly accurate protein structure prediction with AlphaFold. Nature 596: 583–589. [PubMed: 34265844]

Krell T, and Matilla MA (2022) Antimicrobial resistance: progress and challenges in antibiotic discovery and anti-infective therapy. Microb Biotechnol 15: 70–78. [PubMed: 34610194]

Liu YC, Machuca MA, Beckham SA, Gunzburg MJ, and Roujeinikova A (2015) 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. [PubMed: 26457436]

Martin-Rodriguez AJ, Higdon SM, Thorell K, Tellgren-Roth C, Sjoling A, Galperin MY et al. (2022) Comparative Genomics of Cyclic di-GMP Metabolism and Chemosensory Pathways in *Shewanella algae* Strains: Novel Bacterial Sensory Domains and Functional Insights into Lifestyle Regulation. mSystems 7: e0151821. [PubMed: 35311563]

Matilla MA, Mora DM, and Krell T (2020) The use of Isothermal Titration Calorimetry to unravel chemotactic signaling mechanisms. Environ Microbiol 22: 3005–3019. [PubMed: 32329116]

Matilla MA, Velando F, Martin-Mora D, Monteagudo-Cascales E, and Krell T (2022a) A catalogue of signal molecules that interact with sensor kinases, chemoreceptors and transcriptional regulators. FEMS Microbiol Rev 46: fuab043.

Matilla MA, Velando F, Tajuelo A, Martín-Mora D, Xu W, Sourjik V et al. (2022b) Chemotaxis of the Human Pathogen Pseudomonas aeruginosa to the Neurotransmitter Acetylcholine. mBio 13: e0345821. [PubMed: 35254130]

Milligan DL, and Koshland DE Jr. (1993) Purification and characterization of the periplasmic domain of the aspartate chemoreceptor. J Biol Chem 268: 19991–19997. [PubMed: 8397194]

Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL et al. (2021) Pfam: The protein families database in 2021. Nucleic Acids Res 49: D412–D419. [PubMed: 33125078]

Ortega A, Zhulin IB, and Krell T (2017) Sensory Repertoire of Bacterial Chemoreceptors. Microbiol Mol Biol Rev 81: e00033–00017.

Sanchis-Lopez C, Cerna-Vargas JP, Santamaria-Hernando S, Ramos C, Krell T, Rodriguez-Palenzuela P et al. (2021) Prevalence and Specificity of Chemoreceptor Profiles in Plant-Associated Bacteria. mSystems: e0095121. [PubMed: 34546073]

Sanderson T, Bileschi ML, Belanger D, and Colwell LJ (2022) ProteInfer: deep networks for protein functional inference. bioRxiv: 10.1101/2021.1109.1120.461077.

Shrestha M, Compton KK, Mancl JM, Webb BA, Brown AM, Scharf BE, and Schubot FD (2018) Structure of the sensory domain of McpX from *Sinorhizobium meliloti*, the first known bacterial chemotactic sensor for quaternary ammonium compounds. Biochem J 475: 3949–3962. [PubMed: 30442721]

Takahashi Y, Nishiyama SI, Kawagishi I, and Imada K (2020) Structural basis of the binding affinity of chemoreceptors Mlp24p and Mlp37p for various amino acids. Biochem Biophys Res Commun 523: 233–238. [PubMed: 31862138]

Ulrich LE, and Zhulin IB (2005) Four-helix bundle: a ubiquitous sensory module in prokaryotic signal transduction. Bioinformatics 21 Suppl 3: iii45–48. [PubMed: 16306392]

Ulrich LE, Koonin EV, and Zhulin IB (2005) One-component systems dominate signal transduction in prokaryotes. Trends Microbiol 13: 52–56. [PubMed: 15680762]

Upadhyay AA, Fleetwood AD, Adebali O, Finn RD, and Zhulin IB (2016) 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. [PubMed: 27049771]

Webb BA, Compton KK, Del Campo JSM, Taylor D, Sobrado P, and Scharf BE (2017) *Sinorhizobium meliloti* Chemotaxis to Multiple Amino Acids Is Mediated by the Chemoreceptor McpU. Mol Plant Microbe Interact 30: 770–777. [PubMed: 28745538]

Wu R, Gu M, Wilton R, Babnigg G, Kim Y, Pokkuluri PR et al. (2013) Insight into the sporulation phosphorelay: crystal structure of the sensor domain of *Bacillus subtilis* histidine kinase, KinD. Protein Sci 22: 564–576. [PubMed: 23436677]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

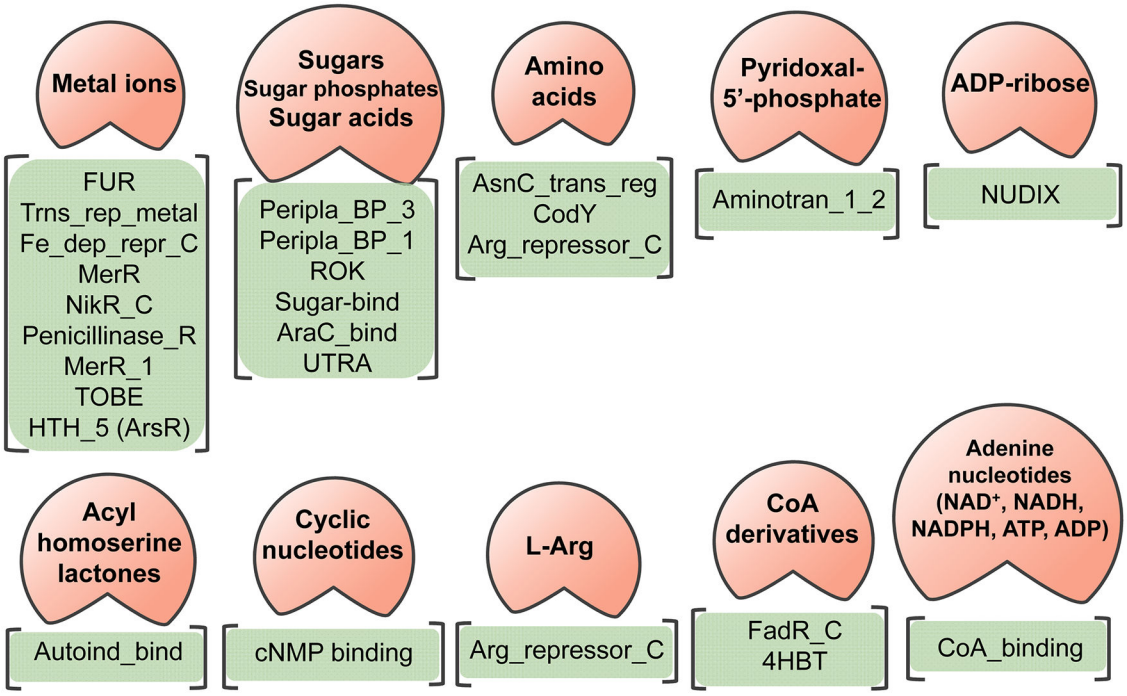

Fig. 1). Sensor domains that preferentially recognize a single molecule or families of closely related molecules.

Reproduced with permission from (Matilla *et al.*, 2022a).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

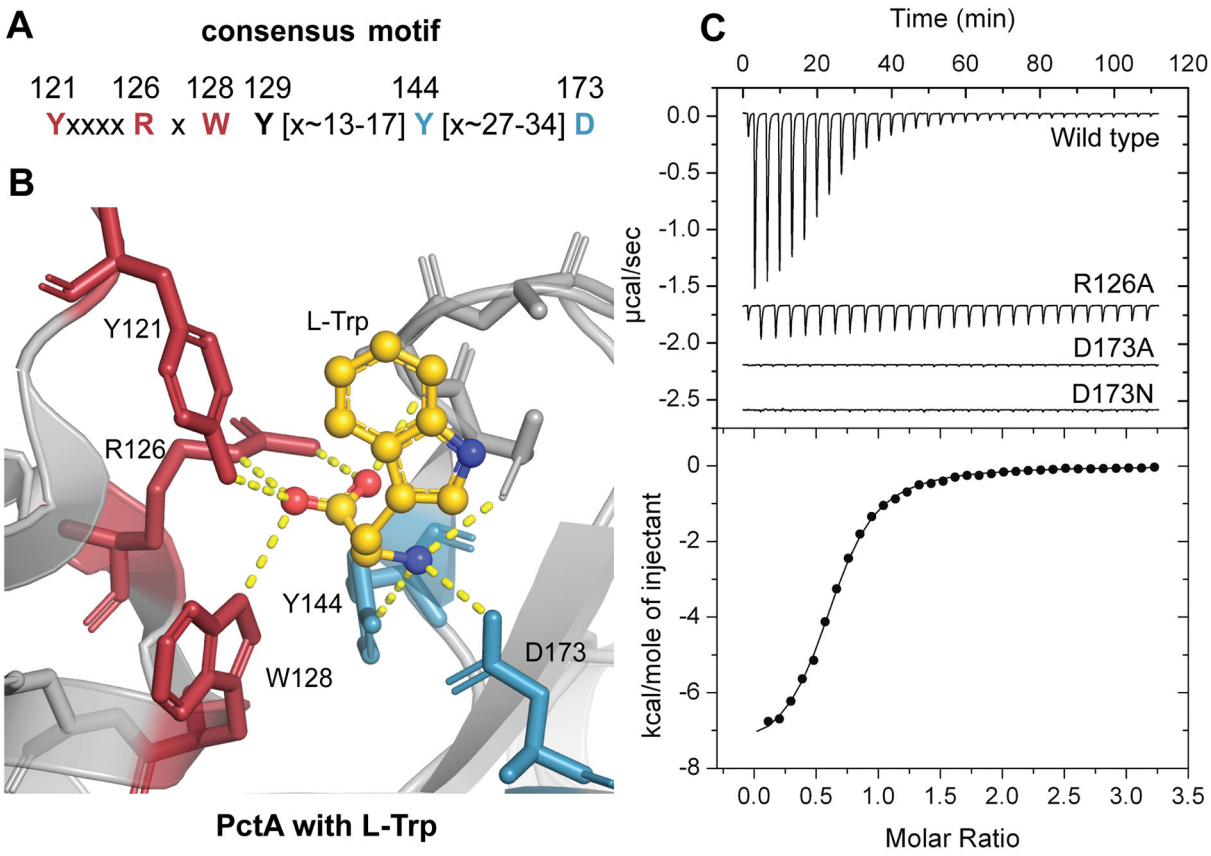

Fig. 2). Conserved sequence motif in the ligand binding pocket of amino acid-binding dCache domains.

A) The consensus motif. Numbers above the motif correspond to positions in the *P. aeruginosa* PctA chemoreceptor. B) Zoom at the binding pocket of the sensor domain of the PctA chemoreceptor in complex with bound L-Trp. The amino acids that interact with L-Trp are shown in the same colour mode as in panel A. C) Isothermal titration calorimetry study of L-Ala binding to the PctA sensor domain and mutants in individual amino acids of the motif. Upper panel: raw titration data. Lower panel: Best fit of binding data for the wild type protein. Modified figure reproduced with permission from (Gumerov *et al.*, 2022).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

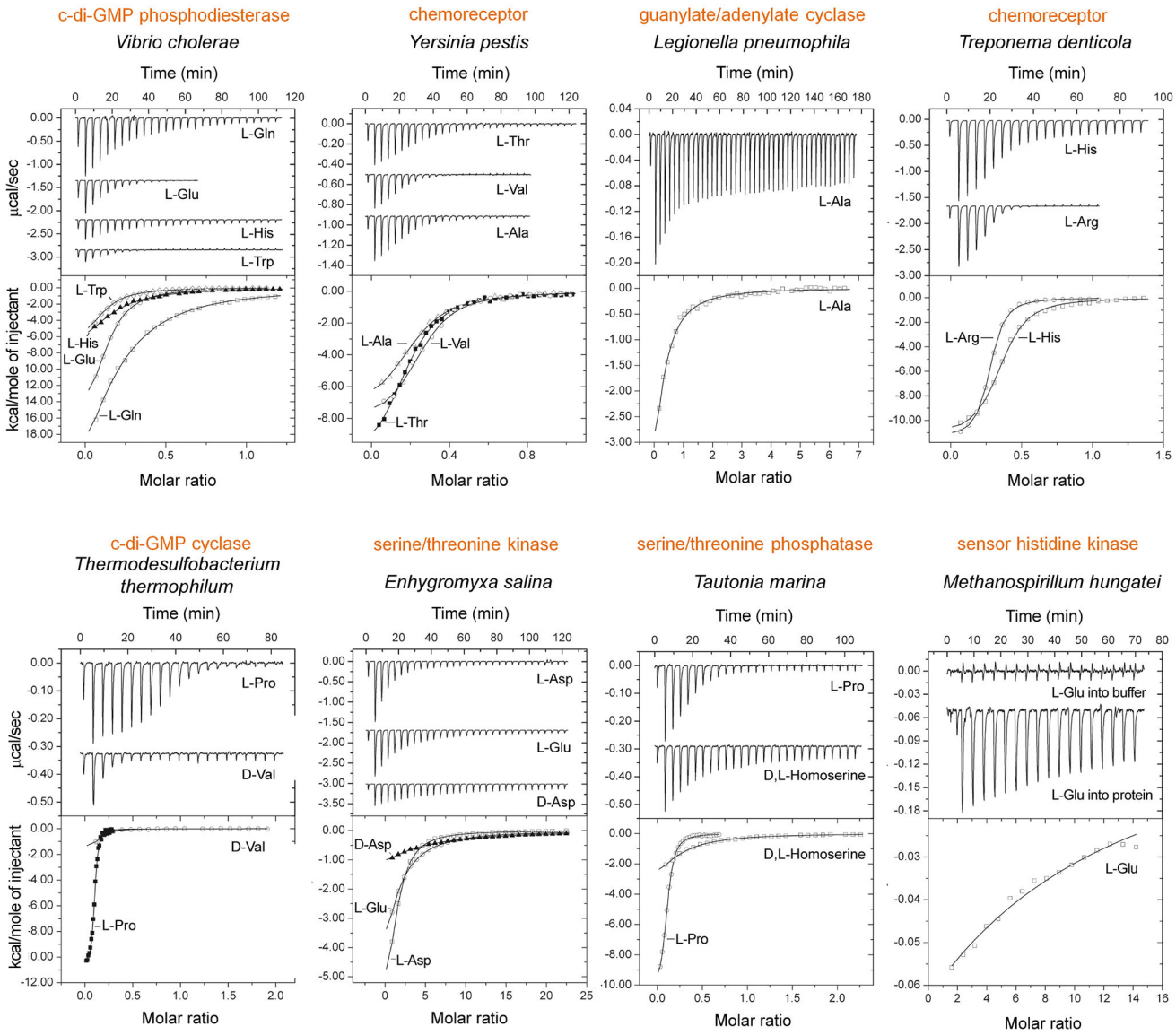

Fig. 3). Experimental verification of ligand binding to sensor domains predicted to recognize amino acids.

Isothermal titration calorimetry studies of individual sensor domains with different amino acids. The receptor family and corresponding bacterial species are indicated. Modified figure reproduced with permission from (Gumerov *et al.*, 2022).