

The HBM domain: Introducing bimodularity to bacterial sensing

Álvaro Ortega* and Tino Krell

Department of Environmental Protection, Estación Experimental del Zaidín, Consejo Superior de Investigaciones Científicas, 18008 Granada, Spain

Received 17 October 2013; Accepted 11 December 2013

DOI: 10.1002/pro.2410

Published online 17 December 2013 proteinscience.org

Abstract: We have recently reported the three dimensional structure of the McpS chemoreceptor sensor domain in complex with its cognate ligands. The domain was characterized by a bimodular architecture, where ligand binding to each module caused a chemotactic response. This is a novel small molecule binding domain, which, however, is un-annotated in relevant databases. We report here the domain signature of the family of McpS-like sensor domains, which was termed helical bimodular (HBM) domain. The HBM domain was identified in *Bacteria* and *Archaea* and forms part of chemoreceptors and histidine kinases. The conservation of amino acids in the ligand binding sites of both modules suggests that HBM family members recognize similar ligands.

Keywords: domain profile; chemotaxis receptor; histidine kinase; sequence analysis

Introduction

Bacteria respond to changing environmental conditions through various signal transduction mechanisms. Genome analyses suggest that responses are primarily mediated by one-component systems, two-component systems, and chemoreceptor-based signaling pathways.¹ These different systems share sensor domains for signal recognition. Chemoreceptors are typically transmembrane proteins and the molecular stimulus caused by ligand binding to the sensor domain is transmitted across the membrane causing a modulation of the Chemotaxis protein A

(CheA) autophosphorylation activity and consequently transphosphorylation activity to the Chemotaxis protein Y (CheY).² Chemosensory systems mediate flagellum-mediated chemotaxis, but are also involved in Type IV pili mediated taxis or in the regulation of alternative cellular processes.^{3,4}

The study of chemosensory pathways from different organisms has revealed a high degree of diversity.^{4,5} This diversity is also reflected in the architecture of chemoreceptors, which were found to differ in their topologies and in the type of sensor domains.⁶ Chemoreceptor sensor domains can be classified according to their size: cluster I sensor domains are of approximately 150 amino acids whereas cluster II domains harbor approximately 250 amino acids.⁶ Cluster I domains were identified as PAS, GAF, CHASE, or TarH domains.⁶ In contrast, around 40% of chemoreceptors possess larger cluster II domains which are mostly un-annotated and in cases an annotation is available it was found to be unreliable.⁶

The methyl-accepting chemotaxis protein S (McpS) of *Pseudomonas putida* binds and mediates

Grant sponsor: Fondo Social Europeo and Fondos FEDER from the European Union through grants from the Junta de Andalucía; Grant numbers: P09-RNM-4509, CVI-7335. Grant sponsor: Spanish Ministry of Economy and Competitiveness; Grant number: Bio2010-16937 awarded to CSIC; Grant sponsor: Jae-Doc contract from CSIC.

*Correspondence to: Álvaro Ortega, Department of Environmental Protection, Estación Experimental del Zaidín, Consejo Superior de Investigaciones Científicas, C/ Prof. Albareda, 1, 18008 Granada, Spain. E-mail: alvaro.ortega@csic.es

chemotaxis toward six Krebs cycle intermediates, butyrate, and acetate.^{7,8} McpS has a cluster II sensor domain and its secondary structure prediction (two long and four short helices) was incompatible with structures of known sensor domains.⁷ We have recently reported the 3D structure of the McpS sensor domain which is composed of two structural modules each composed of a 4-helix bundle.⁸ Interestingly, malate and succinate were bound to the membrane proximal module whereas acetate was present on the membrane distal module. In a superimposition of both modules the ligand binding sites were found to overlap. Interestingly, ligand binding to each module causes a chemotactic response and we have proposed a molecular mechanism that is based on a piston-like movement of the final, long helix H6 that forms part of both ligand binding sites. Although the 3D structure of McpS is composed of two 4-helical bundles, its sequence is not recognized by the corresponding domain signature 4HB_MCP.⁹ We report here the generation of the domain signature of the family McpS-like sensor domains, which we have termed Helical Bi-Modular (HBM) domain.

Results and Discussion

An alignment of a subset of HBM domain sequences representing the major phylogenetic categories is shown in Figure 1(A) and the corresponding secondary structure prediction is provided in Figure 1(B). The HBM domains are predicted to form two short helices followed by a long helix, another pair of small helices, and a final extended helix [Fig. 1(B)], which is in full agreement with the 3D structure of the McpS sensor domain [Fig. 2(A)].⁸ The profile Hidden Markov Model (HMM) generated as explained in the Materials and Methods section was then used to perform a search in the UniprotKB database, which resulted in the retrieval of the approximately 1200 sequences. The majority of proteins were chemoreceptors, but some proteins contained a Histidine kinase phosphoacceptor domain (HisKA, PF00512) and Histidine kinase like ATPase domain (HATPase_c, PF02518), indicating that HBM domains are also found in histidine kinases. The secondary structure prediction and domain size of HBM domains of histidine kinases was entirely comparable to those of chemoreceptors. The sensor domain of the TorS histidine kinase has a similar structure as compared to McpS,¹⁰ but has not been retrieved by the UniprotKB search. TorS binds the protein TorT in complex with the signal molecule. We show below that the conserved residues of the HBM domain are primarily in the two small ligand binding sites. The fact that TorS does not bind small signal molecules is responsible for the lacking sequence conservation which explains why it is not recognized by the HBM domain signature. Using

BLAST we have retrieved homologues of the TorS sensor domain and constructed a domain signature, which, however, was incompatible with the HBM signature described here. This is consistent with the notion that there are two protein families, the HBM; domain and the TorS-like sensor domains, which do not share significant sequence similarities but possess similar structures. The family of TorS-like domains was found to be significantly less populated than the HBM family. A sequence alignment of the McpS and TorS sensor domain gave rise to an identity of less than 12%, which is a value very close to the random identity of two unrelated proteins.

HBM domains are found in bacteria and archaea but are absent from eukaryotes. HBM domain containing chemoreceptors are primarily found in *Proteobacteria* (98%) and in particular in *Alpha-*(16%) and *Gammaproteobacteria* (76%) [Fig. 2(B)]. This may not be surprising since approximately 75% of all chemoreceptor sequences are found in this phylum. Although most of the remaining chemoreceptors are found in *Firmicutes*, HBM domains were not detected in this phylum. Apart from *Proteobacteria* HBM domain containing chemoreceptors were also detected in species like *Methanosarcina* or *Methanobolbus* that belong to the phylum of *Euryarchaeota* of the *Archaea* superkingdom. In addition, receptors with the HBM domain are found in species that belong to the phyla *Deferribacterae*, *Bacteroidia*, *Thermotogae*, or *Planctomycetia* ($\leq 1\%$).

Significant differences were identified in the relative abundance of HBM containing chemoreceptors which is illustrated by the color coding in Figure 2(B). HBM containing receptors were found to represent more than 10% of total receptors in a number of families including *Desulfomonadaceae*, *Bacteriovoracaceae*, and *Pseudoalteromonadaceae* and 5–10% of total chemoreceptors in orders like *Rhodospirillales*, *Thiotrichales*, and *Pseudomonadales*.

A major challenge in signal transduction reside in the identification of the signal molecules that are recognized by a given sensor domain. However, the sequence alignments derived from the selected subset of sequences (Fig. 1) as well as from the complete set of 1200 sequences provide interesting clues. We have noted the conservation of residues R41, Y47, and Y99 in the proximal bundle of the McpS sensor domain as well as R176 and Y235 of its membrane distal bundle (numbering according to Fig. 1). These residues are located close to the bound chemo-attractants in the structure of the McpS sensor domain [Fig. 2(A)] and establish in the case of Y235, R176, and R41 direct interactions with bound ligand. The involvement of some of these residues in ligand binding to McpS has been determined experimentally since the mutation of R41 (corresponding to R60 in McpS) abolished malate binding whereas mutation of R176 (R183 in McpS) reduced significantly acetate binding.⁸ The

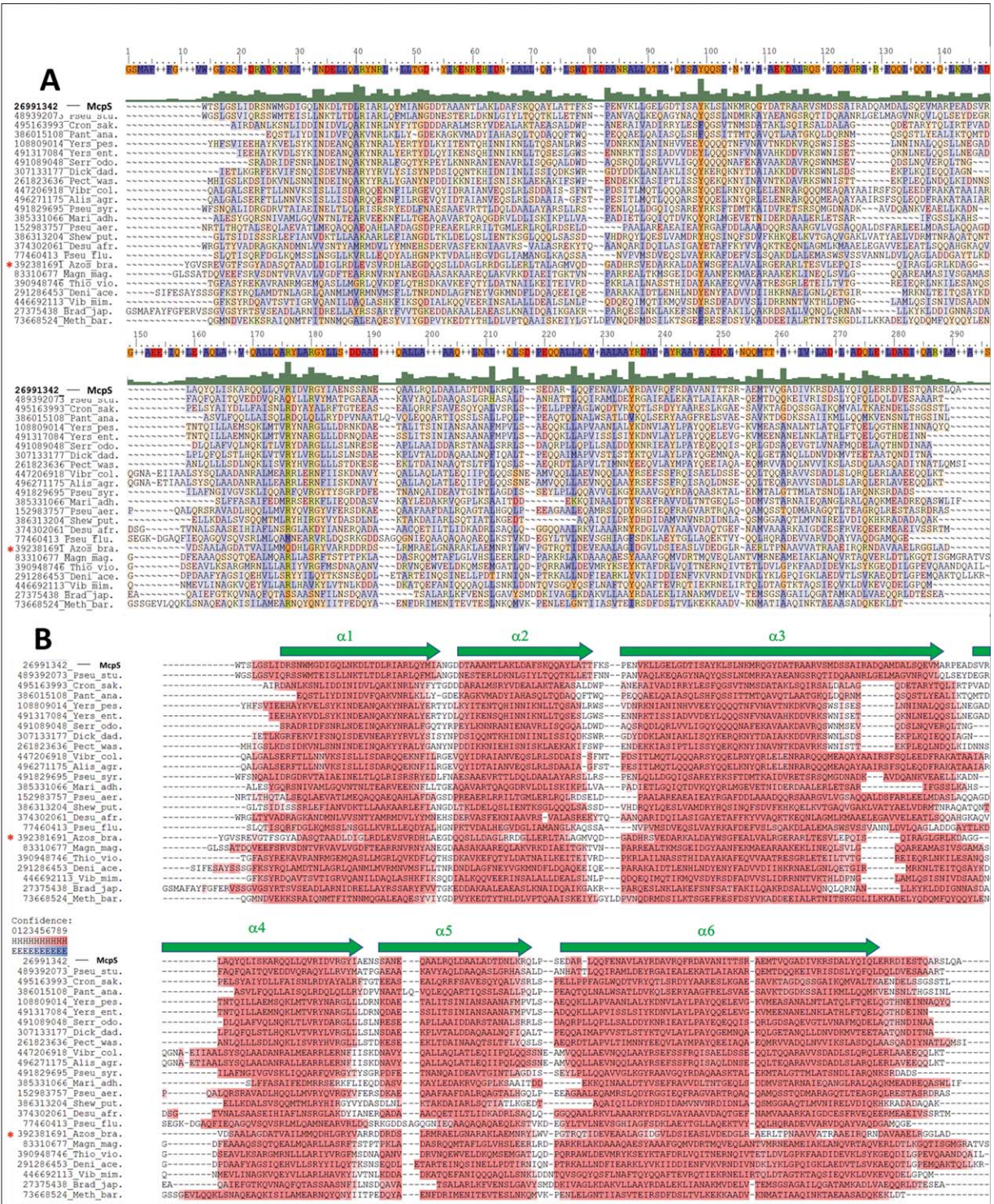


Figure 1. Alignment of a representative subset of HBM domain sequences from chemoreceptors and a sensor kinase. The asterisk marks the sensor kinase, the remaining sequences are from chemoreceptors. Sequences were selected to cover the phylogenetic distribution of species with HBM domains. The GI numbers from the NCBI database are indicated. *Alis_agr*: *Alishewanella agri*; *Azos_lip*: *Azospirillum lipoferum*; *Brad_jap*: *Bradyrhizobium japonicum*; *Cron_sak*: *Cronobacter sakazakii*; *Deni_ace*: *Denitrovibrio acetiphilus*; *Desu_afr*: *Desulfovibrio africanus*; *Dick_dad*: *Dickeya dadantii*; *Magn_mag*: *Magnetospirillum magneticum*; *Mari_adh*: *Marinobacter adhaerens*; *Meth_bar*: *Methanosarcina barkeri*; *Pan_ana*: *Pantoea ananas*; *Pect_was*: *Pectobacterium wasabiae*; *Pseu_aer*: *Pseudomonas aeruginosa*; *Pseu_flu*: *Pseudomonas fluorescens*; *Pseu_stu*: *Pseudomonas stutzeri*; *Pseu_syr*: *Pseudomonas syringae*; *Serr_odo*: *Serratia odorifera*; *Shew_put*: *Shewanella putrefasciens*; *Thio_vio*: *Thiocystis violascens*; *Vibri_cho*: *Vibrio cholerae*; *Vibri_mim*: *Vibrio mimicus*; *Yers_ent*: *Yersinia enterocolitica*; *Yers_pes*: *Yersinia pestis* (A). ClustalX alignment highlighting conserved residues: blue for hydrophobic residues (ACFILMPVV), red for acidic residues (DE), green-yellow for basic amino acids (HKR) and orange for the remaining residues (GNQSTY). The consensus sequence is shown on top of the alignment. (B). Secondary structure prediction for the same set of sequences, red color indicates alpha-helical regions in ascending intensity according to confidence. Green arrows represent the position of the α helices in the 3D structure of the McpS sensor domain.

3. Hickman JW, Tifrea DF, Harwood CS (2005) A chemosensory system that regulates biofilm formation through modulation of cyclic diguanylate levels. *Proc Natl Acad Sci USA* 102:14422–14427.
4. Wuichet K, Zhulin IB (2010) Origins and diversification of a complex signal transduction system in prokaryotes. *Sci Signal* 3:ra50.
5. Hamer R, Chen PY, Armitage JP, Reinert G, Deane CM (2010) Deciphering chemotaxis pathways using cross species comparisons. *BMC Syst Biol* 4:3.
6. Lacal J, Garcia-Fontana C, Munoz-Martinez F, Ramos JL, Krell T (2010) Sensing of environmental signals: classification of chemoreceptors according to the size of their ligand binding regions. *Environ Microbiol* 12: 2873–2884.
7. Lacal J, Alfonso C, Liu X, Parales RE, Morel B, Conejero-Lara F, Rivas G, Duque E, Ramos JL, Krell T (2010) Identification of a chemoreceptor for tricarboxylic acid cycle intermediates: differential chemotactic response towards receptor ligands. *J Biol Chem* 285: 23126–23136.
8. Pineda-Molina E, Reyes-Darias JA, Lacal J, Ramos JL, García-Ruiz JM, Gavira JA, Krell T (2012) Evidence for chemoreceptors with bimodular ligand binding regions harboring two signal-binding sites. *Proc. Acad. Natl. Sci. USA*. 109:18926–18931.
9. Ulrich LE, Zhulin IB (2005) Four-helix bundle: a ubiquitous sensory module in prokaryotic signal transduction. *Bioinformatics* 21:iii45–iii48.
10. Moore JO, Hendrickson WA (2012) An asymmetry-to-symmetry switch in signal transmission by the histidine kinase receptor for TMAO. *Structure* 20:729–741.
11. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.
12. Cserzo M, Wallin E, Simon I, von Heijne G, Elofsson A (1997) Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein Eng* 10:673–676.
13. Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 292:195–202.
14. Huang Y, Niu B, Gao Y, Fu L, Li W (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 26:680–682.
15. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948.
16. Biegert A, Mayer C, Remmert M, Soding J, Lupas AN (2006) The MPI Bioinformatics Toolkit for protein sequence analysis. *Nucleic Acids Res* 34:W335–W339.
17. Eddy SR (2009) A new generation of homology search tools based on probabilistic inference. *Genome Inform* 23:205–211.
18. Neron B, Menager H, Maufrais C, Joly N, Maupetit J, Letort S, Carrere S, Tuffery P, Letondal C (2009) Mobylye: a new full web bioinformatics framework. *Bioinformatics* 25:3005–3011.
19. Magrane M, Consortium tU. UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford)* 2011;2011:bar009.