

# MASE1 and MASE2: Two Novel Integral Membrane Sensory Domains

Anastasia N. Nikolskaya<sup>a</sup> Armen Y. Mulkidjanian<sup>b</sup> Iwona B. Beech<sup>c</sup>  
Michael Y. Galperin<sup>a</sup>

<sup>a</sup>National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Md., USA; <sup>b</sup>Division of Biophysics, Department of Biology and Chemistry, Universität Osnabrück, Osnabrück, Germany; <sup>c</sup>University of Portsmouth School of Pharmacy and Biomedical Sciences, Portsmouth, UK

## Key Words

Signal transduction · Genome analysis · *Escherichia coli* · Adenylate cyclase · Histidine kinase · Membrane topology

## Abstract

*Escherichia coli* proteins YegE and YaiC contain N-terminal integral membrane regions, followed by the putative diguanylate cyclase (GGDEF, DUF1) domains. The membrane domains of these proteins, named MASE1 (membrane-associated sensor) and MASE2, respectively, were found in other bacterial signaling proteins, such as histidine kinases (MASE1) and an adenylate cyclase (MASE2). Although the nature of the signals sensed by MASE1 and MASE2 is still unknown, MASE1-containing receptors appear to play important roles in bacteria, including iron and/or oxygen sensing by hemerythrin-containing proteins in the sulfate-reducing bacterium *Desulfovibrio vulgaris*.

Copyright © 2003 S. Karger AG, Basel

Studies of the bacterial membrane receptors revealed a number of conserved extracellular domains, such as the Cache (Ca<sup>2+</sup> channels, chemotaxis receptors) domain [Anantharaman and Aravind, 2000; Zhulin, 2001],

CHASE (cyclases/histidine kinases associated sensing extracellular) domain [Anantharaman and Aravind, 2001; Mougél and Zhulin, 2001], and several others [Zhulin et al., 2003]. An important feature of all those domains is their propensity to associate with more than one type of signal output domains (histidine kinases, adenylate cyclases, chemotaxis transducers), which made possible their recognition as conserved domains. In addition, these domains are often found in association with the GGDEF and EAL domains [Tal et al., 1998], which have not been directly biochemically characterized so far, but most likely have diguanylate cyclase [Ausmees et al., 2001; Pei and Grishin, 2001] and phosphodiesterase [Galperin et al., 1999] activities, respectively [see Galperin et al., 2001a, for a review].

Recently, the first integral membrane sensory domain was described, named MHYT after its conserved Met, His, and Tyr residues [Galperin et al., 2001b]. In addition to serving as the sensory domain in several histidine kinases, MHYT was found in a combination with the GGDEF and EAL domains, as well as with the heme- and flavin-binding PAS domain and the DNA-binding LytTR domain [Galperin et al., 2001b; Nikolskaya and Galperin, 2002]. Here we describe two more integral membrane sensory domains that are also shared between different classes of bacterial receptor proteins. The MASE1 (membrane-associated sensor) domain, consisting of 8 pre-

## KARGER

Fax +41 61 306 12 34  
E-Mail karger@karger.ch  
www.karger.com

© 2003 S. Karger AG, Basel  
1464–1801/03/0051–0011\$19.50/0

Accessible online at:  
www.karger.com/mmb

Michael Y. Galperin  
National Center for Biotechnology Information  
National Library of Medicine, National Institutes of Health  
Bethesda, MD 20894 (USA)  
Tel. +1 301 435 5910, Fax +1 301 435 7794, E-Mail galperin@ncbi.nlm.nih.gov

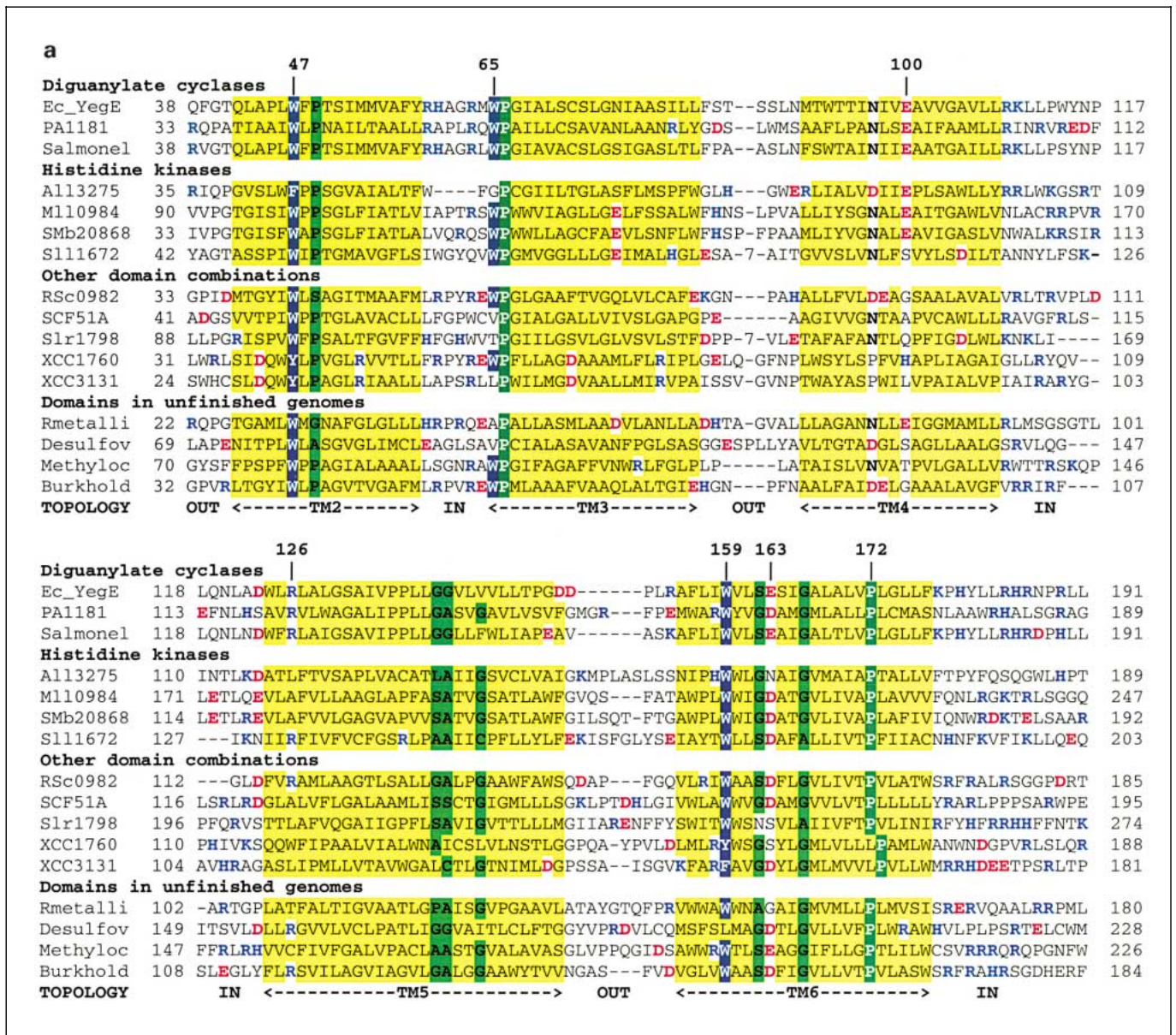


Fig. 1. Sequence alignments of the MASE1 and MASE2 domains. Protein names, species abbreviations, and the accession numbers are as in table I. The most conserved amino acid residues are shown in bold and/or shaded. Yellow shading indicates uncharged amino acid residues located in the predicted transmembrane segments. Positively charged amino acid residues (Arg, Lys, His) are shown in blue, acidic residues (Asp and Glu) are shown in red. Small amino acid (Gly, Ala, Ser, Cys) and Pro residues are shaded green. The most conserved Trp, Tyr and Pro residues are shown in reverse colors. a MASE1 domain. The alignment shows only five transmembrane segments out of eight, starting from the second one. 'Salmonel' indicates two identical domains from *S. enterica* subsp. *typhimurium* protein STM2123 and *S. enterica* subsp. *typhi* protein STY2336. 'Burkhold' indicates two identical domains encoded in the unfinished genomes of *B. mallei* (sequenced at TIGR) and *B. pseudomallei* (sequenced at the Sanger Institute).



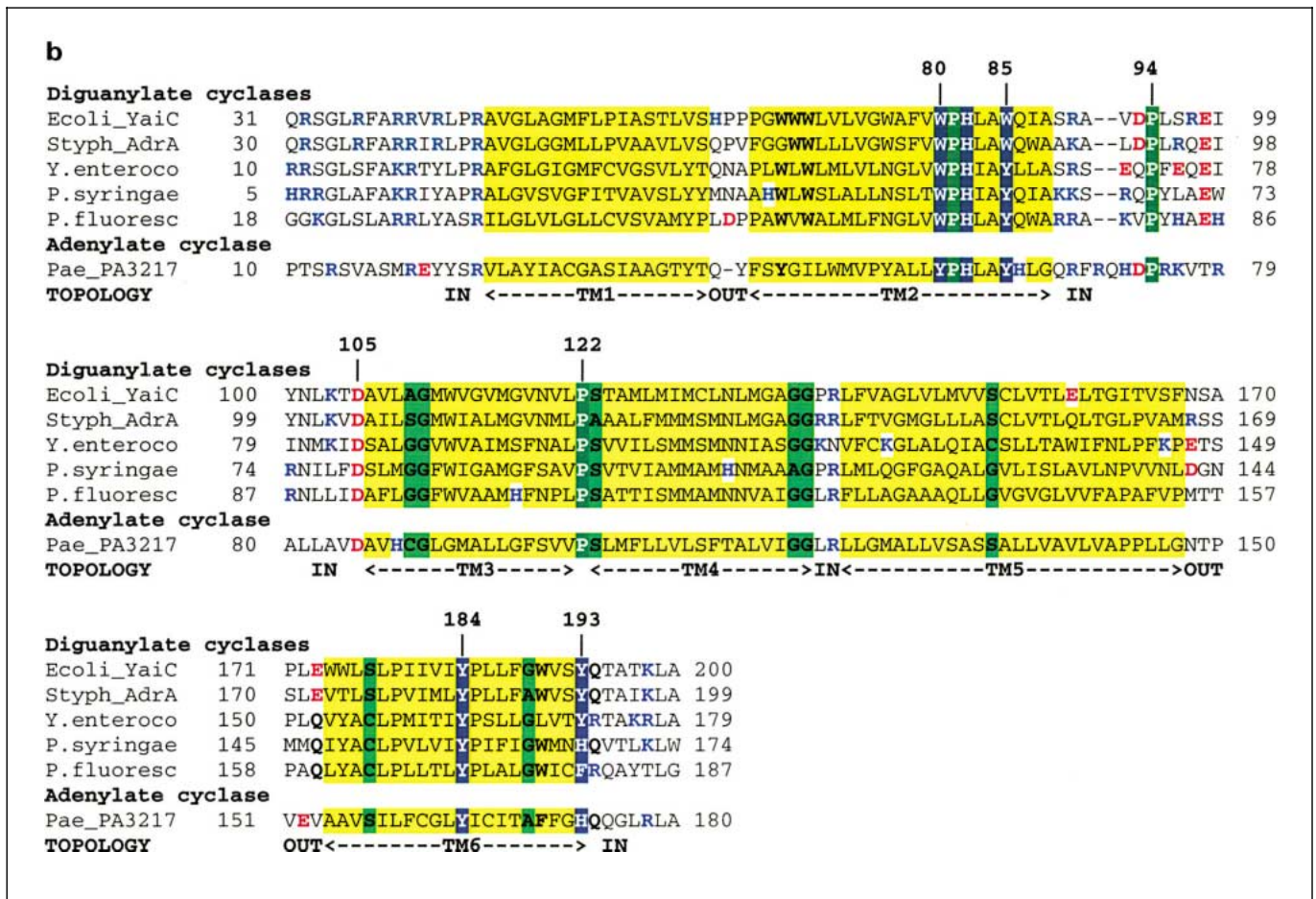


Fig. 1. b MASE2 domain. 'Styph-AdrA' indicates almost identical domains from *S. enterica* subsp. *typhimurium* protein STM0385 and *S. enterica* subsp. *typhi* protein STY0418.

dicted transmembrane segments, was found in association with histidine kinase, GGDEF, GGDEF-EAL, and PAS domains, as well as in the stand-alone form. The much less widespread MASE2 domain consists of six predicted transmembrane segments and has been found so far in association with adenylate cyclase and GGDEF-EAL domains.

The MASE1 and MASE2 domains were identified in the course of assigning putative signaling proteins from newly sequenced bacterial genomes to the Clusters of Orthologous Groups of proteins (COG) database [Tatusov et al., 2000] as conserved N-terminal domains fused to different C-terminal domains. They have been further defined using PSI-BLAST and gapped BLAST searches against, respectively, the nonredundant protein database

and the database of finished and unfinished microbial genome sequences at the National Center for Biotechnology Information (NCBI, Bethesda, Md.). Domain architectures were defined by comparing the proteins against the CDD [Marchler-Bauer et al., 2002], COG [Tatusov et al., 2000], Pfam [Bateman et al., 2002], and SMART [Letunic et al., 2002] databases. Transmembrane regions and membrane topology assignments were based on combining predictions, generated by PHDhtm [Rost et al., 1996], HMMTOP [Tusnady and Simon, 2001], TMpred [Hofmann and Stoffel, 1993] and TopPred2 [Claros and von Heijne, 1994] programs.

Sequence alignment of the MASE1 domains (fig. 1a) shows three well-conserved Trp residues, two of which are located close to the outer face of the cytoplasmic mem-

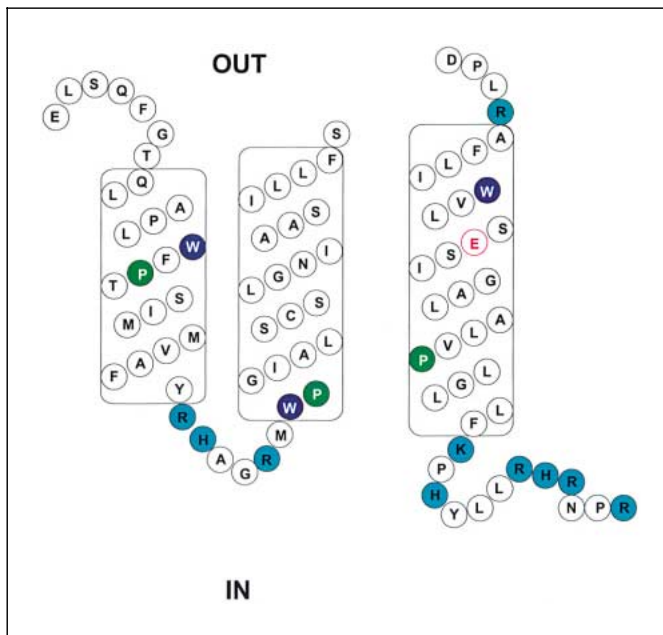


Fig. 2. A model of the membrane orientation of the MASE1 domain. The model shows the predicted localization of the conserved residues in the second, third and sixth transmembrane segments of the MASE1 domain from *E. coli* YegE protein. Residue coloring is as in figure 1.

brane (Trp-47 and Trp-159 in *E. coli* YegE) and one (Trp-65) is located on the inner face of the membrane. These three residues, together with conserved prolines (Pro-66 and Pro-172) and a conserved acidic residue (Glu-163), form the MASE1 domain signature (fig. 2). MASE1 is an integral membrane domain with short hydrophilic loops between the transmembrane segments. Another interesting trait of this domain is the abundance of Gly, Ser, and Ala residues, indicating that it has a flexible structure [Popot and Engelman, 2000] that could potentially respond to a variety of membrane parameters, such as osmotic pressure, transmembrane electric gradient, or the presence of cations. This suggestion is supported by the abundance of intrahelical proline residues, which, in the case of membrane proteins, are specific for enzymes that alternate between different conformations [Ubarretxena-Belandia and Engelman, 2001]. The conservation of carboxyl residues in the central part of the hydrophobic region (Glu-100, Glu-163) might also be related to conformational changes of the MASE1 domain accompanied (or triggered) by the changes in the charge state of the carboxyls, as it happens in energy-transducing enzymes [Ubar-

retxena-Belandia and Engelman, 2001]. The stable structure of the whole domain is probably maintained by the salt bridges between terminal parts of different helices, formed by poorly conserved charged residues (e.g., Asp-123, Arg-109).

Domain architectures of the MASE1-containing proteins show significant diversity. Cyanobacteria and  $\alpha$ -proteobacteria encode MASE1-containing histidine kinases, while  $\gamma$ -proteobacteria mostly encode MASE1 combinations with the PAS, GGDEF, and EAL domains (table 1). In addition, MASE1 can have only PAS as the output domain or even be present in the stand-alone form. A particularly interesting domain combination has been found in the unfinished genome of the sulfate-reducing bacterium *Desulfovibrio vulgaris*. Although this organism is classified as an obligate anaerobe, it reportedly can survive low levels of oxygen. *D. vulgaris* encodes a MASE1-GGDEF-containing sensor with an additional hemerythrin-like domain (table 1). An alignment of the latter domain (not shown) reveals good conservation of its iron-binding residues, suggesting involvement of this sensor in the response to iron and/or oxygen. Remarkably, *D. vulgaris* also contains a C-terminal hemerythrin domain in the recently described chemotaxis sensor DcrH, suggesting its participation in oxygen sensing [Xiong et al., 2000]. In addition, *D. vulgaris* encodes a heme-containing chemotaxis sensor [Fu et al., 1994]. Sulfate-reducing bacteria readily form biofilms on surfaces of ferrous materials often leading to their severe deterioration [the phenomenon referred to as ‘microbially influenced corrosion’, see Beech and Gaylarde, 1999; Beech, 2002, and references therein]. Given that GGDEF-containing proteins often regulate secretion of extracellular proteins and polysaccharides [Galperin et al., 2001a] that are essential for the biofilm development [Beech, 2002; Beech et al., 2000, and references therein], delineation of the exact signal perceived by the *D. vulgaris* MASE1-GGDEF-containing sensor protein would be of great interest.

An alignment of the MASE2 domains (fig. 1b) also shows a number of well-conserved aromatic residues [Trp-80, His-82, Trp-85, Tyr-184, and Tyr-193, *E. coli* YaiC numbering], prolines and tiny residues. Unfortunately, because all known MASE2 domains were found in  $\gamma$ -proteobacteria, an assessment of the degree of sequence conservation among these domains could have been biased by their phylogenetic proximity. The few known MASE2-containing proteins reveal only two domain architectures with adenylate cyclase and GGDEF-EAL output domains (table 1). The MASE2-containing AdrA protein of *Salmonella typhimurium* is required for the bio-

Table 1. Domain architectures of MASE1- and MASE2-containing proteins

Domain architecture <sup>a</sup>	Organism, protein name (accession code)
MASE1-PAS-HisK-CheY	<i>Anabaena</i> sp. PCC 7120 All3275 (BAB74974)
MASE1-PAS-PAS-PAS-HisK	<i>Mesorhizobium loti</i> Mll0984 (BAB48455)
MASE1-PAS-PAS-PAS-PAS-HisK	<i>Sinorhizobium meliloti</i> SMb20868 (CAC49562)
MASE1-HisK-CheY	<i>Synechocystis</i> sp. PCC6803 Sl11672 (BAA16806) <i>Caulobacter crescentus</i> CC2632 (AAK24599)
MASE1-GAF-HisK	<i>Synechocystis</i> sp. PCC6803 Slr1805 (BAA17733)
MASE1-PAS-PAS-PAS-GGDEF-EAL	<i>Escherichia coli</i> YegE (P38097), <i>Pseudomonas aeruginosa</i> PA1181 (AAG04570)
MASE1-PAS-PAS-GGDEF-EAL	<i>Salmonella typhimurium</i> STM2123 (AAL21027), <i>Salmonella typhi</i> STY2336 (CAD02486)
MASE1-PAS-GGDEF-EAL	<i>Corynebacterium glutamicum</i> Cgl1011 (gi19552261)
Stand-alone MASE1	<i>Streptomyces coelicolor</i> SCF51A.28c (CAB56680) <i>Novosphingobium aromaticivorans</i> Saro_p_452 (gi23107141) <sup>b</sup>
MASE1-GGDEF	<i>Synechocystis</i> sp. PCC6803 Slr1798 (BAA16824), <i>Burkholderia fungorum</i> Bcep_p_7208 (gi22989260) <sup>b</sup> <i>Microbulbifer degradans</i> Mdeg_p_1234 (gi23027413) <sup>b</sup>
MASE1-GGDEF-Hemerythrine	<i>Desulfovibrio vulgaris</i> <sup>b</sup>
MASE1-PAS	<i>Ralstonia solanacearum</i> RSc0982 (CAD14684), <i>Burkholderia mallei</i> <sup>b</sup> , <i>Burkholderia pseudomallei</i> <sup>b</sup>
MASE1-PAS-PAS-PAS	<i>Ralstonia metallidurans</i> Reut_p_1540 (gi22976760) <sup>b</sup>
MASE1-Unk	<i>Xanthomonas campestris</i> XCC1760 (AAM41051), XCC3131 (AAM42402)
MASE2-AdeCyc	<i>Pseudomonas aeruginosa</i> PA3217 (AAG06605) <i>Azotobacter vinelandii</i> Avin_p_4245 (gi23106061) <sup>b</sup> <i>Microbulbifer degradans</i> Mdeg_p_3734 (gi23030048) <sup>b</sup>
MASE2-GGDEF	<i>Escherichia coli</i> YaiC (AAC73488), <i>Salmonella typhi</i> AdrA (CAD08840), <i>Salmonella typhimurium</i> YaiC (CAB77394), <i>Pseudomonas fluorescens</i> Pflu_p_3166 (gi23061031) <sup>b</sup> , <i>Pseudomonas syringae</i> <sup>b</sup> , <i>Yersinia enterocolitica</i> <sup>b</sup>

<sup>a</sup> Domain name abbreviations are as follows: AdeCyc, adenylate cyclase; HisK, histidine kinase; CheY, CheY-like phosphoacceptor domain; PAS (Per-Arnt-Sim), heme-, flavin-, ATP-, and cinnamic acid-binding domain [Taylor and Zhulin, 1999]; GAF (cGMP-AdeCyc-FhlA), cGMP- and phytochrome-binding domain [Aravind and Ponting, 1997]; GGDEF, GGDEF domain, putative diguanylate cyclase [Galperin et al., 2001a; Tal et al., 1998]; EAL, EAL domain, putative cyclic diguanylate phosphodiesterase [Galperin et al., 2001a; Tal et al., 1998]; hemerythrine, heme-binding domain [Xiong et al., 2000]; Unk indicates previously uncharacterized domain that will be described elsewhere.

<sup>b</sup> The protein from an unfinished genome project at the Institute of Genomic Research (*Burkholderia mallei*, *Desulfovibrio vulgaris*, *Methylococcus capsulatus*, *Pseudomonas syringae*), the Sanger Institute (*Burkholderia pseudomallei*, *Yersinia enterocolitica*), or the DOE Joint Genome Institute (*Azotobacter vinelandii*, *Burkholderia fungorum*, *Microbulbifer degradans*, *Novosphingobium aromaticivorans*, *Pseudomonas fluorescens*, *Ralstonia metallidurans*).

synthesis of extracellular cellulose, which is important for intercellular adhesion [Romling et al., 2000; Zogaj et al., 2001].

Although most organisms encode just one MASE1- or MASE2-containing sensor, *Synechocystis* sp. PCC6803 encodes MASE1-containing receptors of two classes, a putative diguanylate cyclase and two histidine kinases (table 1). As has been noted earlier [Zhulin et al., 2003], shar-

ing of sensory domains between different membrane signal transducers suggests that microorganisms use similar conserved domains to sense similar environmental signals and transmit this information via different signal transduction pathways to spawn different regulatory circuits: transcriptional regulation (histidine kinases), catabolite repression (adenylate cyclases), and production of extracellular factors (diguanylate cyclase and phosphodiester-

ase). The diversity of signaling pathways fed by the MASE1 and MASE2 domains indicates that these domains sense some critically important extracellular signals and are attractive candidates for detailed experimental studies.

## Acknowledgements

This study was supported in part by a grant from the NHS to I.B.B. and by a travel grant of the Deutsche Forschungsgemeinschaft to A.Y.M. We acknowledge the availability of unfinished genome sequences from the Institute for Genomic Research, the Sanger Institute, and the DOE Joint Genome Institute.

## References

- Anantharaman, V. and Aravind, L. 2000. Cache – a signaling domain common to animal Ca<sup>2+</sup>-channel subunits and a class of prokaryotic chemotaxis receptors. *Trends Biochem. Sci.* 25: 535–537.
- Anantharaman, V. and Aravind, L. 2001. The CHASE domain: a predicted ligand-binding module in plant cytokinin receptors and other eukaryotic and bacterial receptors. *Trends Biochem. Sci.* 26:579–582.
- Aravind, L. and Ponting, C.P. 1997. The GAF domain: an evolutionary link between diverse phototransducing proteins. *Trends Biochem. Sci.* 22:458–459.
- Ausmees, N., Mayer, R., Weinhouse, H., Volman, G., Amikam, D., Benziman, M., and Lindberg, M. 2001. Genetic data indicate that proteins containing the GGDEF domain possess diguanylate cyclase activity. *FEMS Microbiol. Lett.* 204:163–167.
- Bateman, A., Birney, E., Cerruti, L., Durbin, R., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M., and Sonnhammer, E.L. 2002. The Pfam protein families database. *Nucleic Acids Res.* 30:276–280.
- Beech, I.B. and Gaylarde, C.C. 1999. Recent advances in the study of biocorrosion – An overview. *Rev. Microbiol.* 30:177–190.
- Beech, I.B., Tapper, R., and Gubner, R. 2000. Microscopy methods for studying biofilms. In Evans, L.V. (Ed.), *Biofilms: Recent Advances in Their Study and Control*. London: Harwood, pp. 51–70.
- Beech, I.B. 2002. Biocorrosion: role of sulfate-reducing bacteria. In Bitton, G. (Ed.), *Encyclopedia of Environmental Microbiology*. New York: Wiley, pp. 465–475.
- Claros, M.G. and von Heijne, G. 1994. TopPred II: an improved software for membrane protein structure predictions. *Comput. Appl. Biosci.* 10:685–686.
- Fu, R., Wall, J.D., and Voordouw, G. 1994. DcrA, a c-type heme-containing methyl-accepting protein from *Desulfovibrio vulgaris* Hildenborough, senses the oxygen concentration or redox potential of the environment. *J. Bacteriol.* 176: 344–350.
- Galperin, M.Y., Natale, D.A., Aravind, L., and Koonin, E.V. 1999. A specialized version of the HD hydrolase domain implicated in signal transduction. *J. Mol. Microbiol. Biotechnol.* 1: 303–305.
- Galperin, M.Y., Nikolskaya, A.N., and Koonin, E.V. 2001a. Novel domains of the prokaryotic two-component signal transduction system. *FEMS Microbiol. Lett.* 203:11–21.
- Galperin, M.Y., Gaidenko, T.A., Mulkidjanian, A.Y., Nakano, M., and Price, C.W. 2001b. MHYT, a new integral membrane sensor domain. *FEMS Microbiol. Lett.* 205:17–23.
- Hofmann, K. and Stoffel, W. 1993. TMbase – A database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler* 374:166.
- Letunic, I., Goodstadt, L., Dickens, N.J., Doerks, T., Schultz, J., Mott, R., Ciccarelli, F., Copley, R.R., Ponting, C.P., and Bork, P. 2002. Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Res.* 30:242–244.
- Marchler-Bauer, A., Panchenko, A.R., Shoemaker, B.A., Thiessen, P.A., Geer, L.Y., and Bryant, S.H. 2002. CDD: A database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Res.* 30: 281–283.
- Mougel, C. and Zhulin, I.B. 2001. CHASE: an extracellular sensing domain common to transmembrane receptors from prokaryotes, lower eukaryotes and plants. *Trends Biochem. Sci.* 26:582–584.
- Nikolskaya, A.N. and Galperin, M.Y. 2002. A novel type of conserved DNA-binding domain in the transcriptional regulators of the AlgR/AgrA/LytR family. *Nucleic Acids Res.* 30: 2453–2459.
- Pei, J. and Grishin, N.V. 2001. GGDEF domain is homologous to adenylyl cyclase. *Proteins.* 42: 210–216.
- Popot, J.L. and Engelman, D.M. 2000. Helical membrane protein folding, stability, and evolution. *Annu. Rev. Biochem.* 69:881–922.
- Romling, U., Rohde, M., Olsen, A., Normark, S., and Reinkoster, J. 2000. AgfD, the checkpoint of multicellular and aggregative behaviour in *Salmonella typhimurium* regulates at least two independent pathways. *Mol. Microbiol.* 36:10–23.
- Rost, B., Fariselli, P., and Casadio, R. 1996. Topology prediction for helical transmembrane proteins at 86% accuracy. *Protein Sci.* 5:1704–1718.
- Tal, R., Wong, H.C., Calhoon, R., Gelfand, D., Fear, A.L., Volman, G., Mayer, R., Ross, P., Amikam, D., Weinhouse, H., et al. 1998. Three *cdg* operons control cellular turnover of cyclic di-GMP in *Acetobacter xylinum*: Genetic organization and occurrence of conserved domains in isoenzymes. *J. Bacteriol.* 180:4416–4425.
- Tatusov, R.L., Galperin, M.Y., Natale, D.A., and Koonin, E.V. 2000. The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 28:33–36.
- Taylor, B.L. and Zhulin, I.B. 1999. PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiol. Mol. Biol. Rev.* 63:479–506.
- Tusnady, G.E. and Simon, I. 2001. The HMMTOP transmembrane topology prediction server. *Bioinformatics* 17:849–850.
- Ubarretxena-Belandia, I. and Engelman, D.M. 2001. Helical membrane proteins: Diversity of functions in the context of simple architecture. *Curr. Opin. Struct. Biol.* 11:370–376.
- Xiong, J., Kurtz, D.M., Jr., Ai, J., and Sanders-Loehr, J. 2000. A hemerythrin-like domain in a bacterial chemotaxis protein. *Biochemistry.* 39:5117–5125.
- Zhulin, I.B. 2001. The superfamily of chemotaxis transducers: From physiology to genomics and back. *Adv. Microb. Physiol.* 45:157–198.
- Zhulin, I.B., Nikolskaya, A.N., and Galperin, M.Y. 2003. Common sensory domains in transmembrane receptors for diverse signal transduction pathways in Bacteria and Archaea. *J. Bacteriol.* 185:285–294.
- Zogaj, X., Nitz, M., Rohde, M., Bokranz, W., and Romling, U. 2001. The multicellular morphotypes of *Salmonella typhimurium* and *Escherichia coli* produce cellulose as the second component of the extracellular matrix. *Mol. Microbiol.* 39:1452–1463.