

Common Extracellular Sensory Domains in Transmembrane Receptors for Diverse Signal Transduction Pathways in *Bacteria* and *Archaea*

Igor B. Zhulin,^{1*} Anastasia N. Nikolskaya,² and Michael Y. Galperin²

School of Biology, Georgia Institute of Technology, Atlanta, Georgia 30332-0230,¹ and National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894²

Received 29 July 2002/Accepted 11 October 2002

Transmembrane receptors in microorganisms, such as sensory histidine kinases and methyl-accepting chemotaxis proteins, are molecular devices for monitoring environmental changes. We report here that sensory domain sharing is widespread among different classes of transmembrane receptors. We have identified two novel conserved extracellular sensory domains, named CHASE2 and CHASE3, that are found in at least four classes of transmembrane receptors: histidine kinases, adenylate cyclases, predicted diguanylate cyclases, and either serine/threonine protein kinases (CHASE2) or methyl-accepting chemotaxis proteins (CHASE3). Three other extracellular sensory domains were shared by at least two different classes of transmembrane receptors: histidine kinases and either diguanylate cyclases, adenylate cyclases, or phosphodiesterases. These observations suggest that microorganisms use similar conserved domains to sense similar environmental signals and transmit this information via different signal transduction pathways to different regulatory circuits: transcriptional regulation (histidine kinases), chemotaxis (methyl-accepting proteins), catabolite repression (adenylate cyclases), and modulation of enzyme activity (diguanylate cyclases and phosphodiesterases). The variety of signaling pathways using the CHASE-type domains indicates that these domains sense some critically important extracellular signals.

Signal transduction networks are information-processing pathways that recognize various physicochemical stimuli, amplify and process signals, and trigger adaptive responses in a cell. In prokaryotes, the best-studied and most frequently found signal transduction pathways are centered around two components, a sensor histidine kinase and a response regulator (for recent reviews, see references 28, 29, 56, and 62). Most of the characterized two-component systems regulate gene expression and utilize sensor (class I) histidine kinases for both stimulus detection and signal transmission to a response regulator (18). A specialized version of a two-component system that regulates microbial motility (chemotaxis) utilizes chemotaxis transducers (methyl-accepting chemotaxis proteins [MCPs]) that detect stimuli and transmit the signal to a nonsensor (class II) histidine kinase, CheA, which transmits it to its cognate response regulator, CheY (10, 11). Recent experimental data and genomic studies revealed the existence of additional sensory pathways in microorganisms that combine different elements of two-component signaling with various output modules, including adenylate cyclases and putative diguanylate cyclases and phosphodiesterases (8, 22, 25, 40, 48, 58).

The similarity between systems governed by sensor histidine kinases and by chemotaxis histidine kinases was first recognized in conserved but differently organized signaling domains involved in the phosphotransfer between the kinase and response regulator (5, 47). The advent of genomics resulted in detecting less-pronounced similarities between sensor histidine kinases, MCPs, and sensor cyclases. All three major types of

sensor proteins share conserved intracellular sensory domains, PAS (49, 60, 70) and GAF (7, 68), as well as a specialized intracellular module for transmembrane signaling, the HAMP domain (6, 63).

A majority of sensor histidine kinases and MCPs recognize extracellular stimuli. The transmembrane MCPs of *Escherichia coli* serve as the model for studying transmembrane signaling (19). Comparative genomic analysis detected hundreds of MCPs in dozens of sequenced microbial genomes that have the same membrane topology as the *E. coli* proteins (69) and therefore are predicted to have a similar signaling mechanism. However, the extracellular ligand-binding domains of various MCPs and other sensor proteins that serve as input modules in microbial signal transduction appeared to comprise an extremely diverse group. Experimental studies seem to support this view. For example, the extracellular sensory domains of MCPs from *E. coli* (19), *Bacillus subtilis* (1), *Pseudomonas aeruginosa* (64), and *Halobacterium salinarum* (36) have quite different repertoires of chemical ligands, ranging from ions to polypeptides.

The recent explosion of microbial genomic data and the improvement in bioinformatic tools helped identify two types of extracellular sensing domains that are present in various transmembrane sensor proteins. The Cache (Ca²⁺ channels, chemotaxis receptors) domain (3, 69) is common to three major classes of sensor proteins: class I histidine kinases, MCPs, and adenylate cyclases. It is found in several characterized proteins: McpB (26) and McpC (23) of *B. subtilis*, which serve as amino acid and carbohydrate sensors, respectively, and the DctB histidine kinase of *Rhizobium leguminosarum*, which is a dicarboxylate sensor (51). In addition, the Cache domain has been found in association with the GGDEF and HD-GYP domains, which have not yet been characterized biochemically

* Corresponding author. Mailing address: School of Biology, Georgia Institute of Technology, Atlanta, GA 30332-0230. Phone: (404) 385-2224. Fax: (404) 894-0519. E-mail: igor.zhulin@biology.gatech.edu.

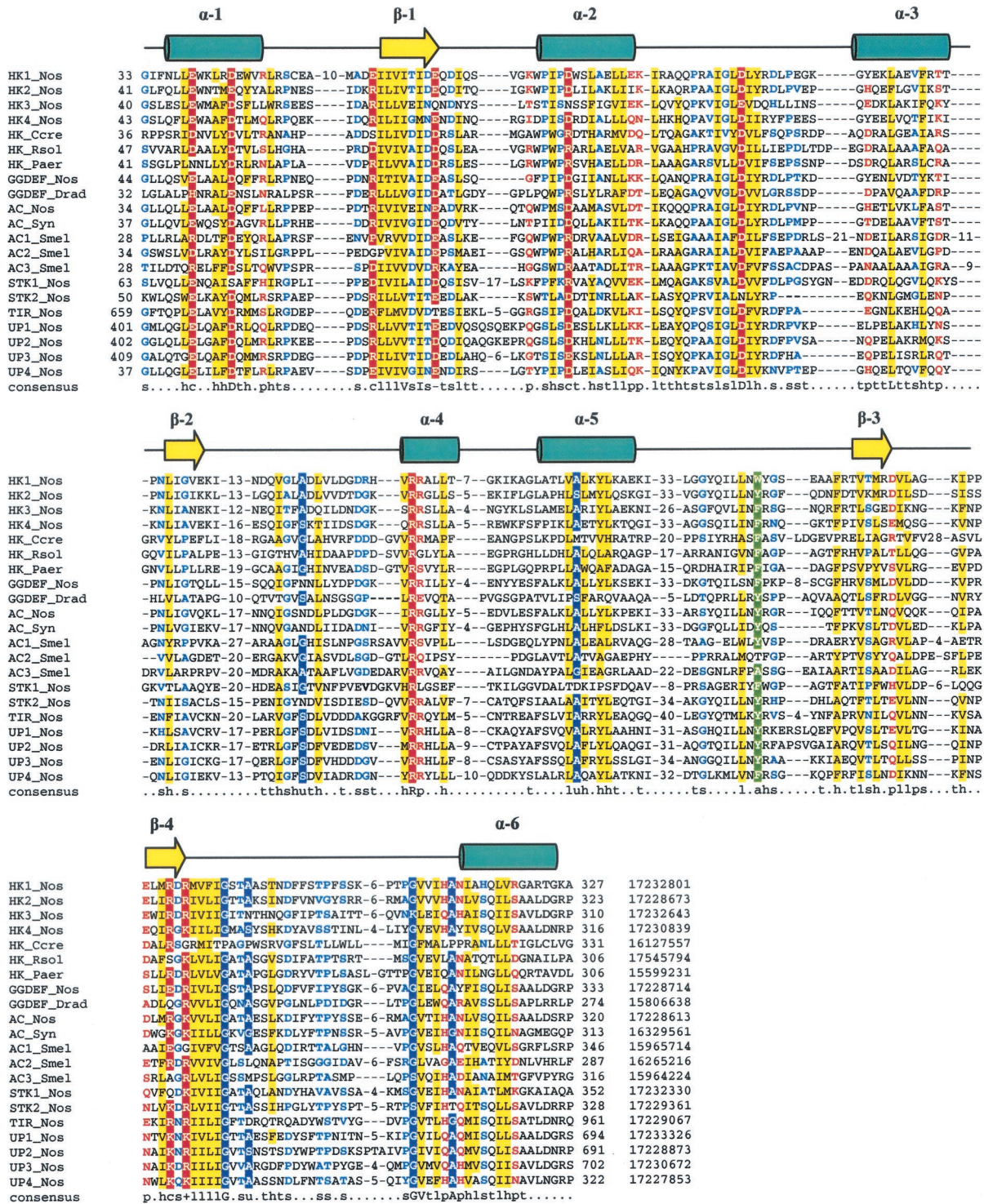
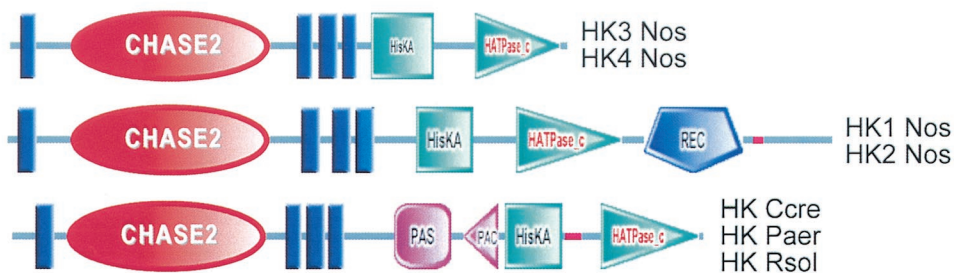
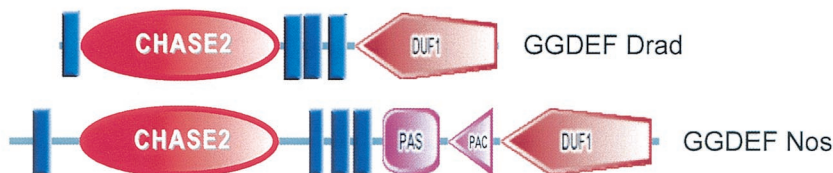


FIG. 1. Multiple alignment of CHASE2 domains. The predicted secondary-structure elements are shown above the alignment: cylinders indicate α -helices, and arrows indicate β -strands. Amino acid residues are colored according to an 80% consensus calculated for an extended alignment, which included sequences from unfinished microbial genomes (data not shown): polar (p; KRHEDQNST) in red; bulky hydrophobic (h; ALIVMYFW) and the aliphatic subset of these (l; ILMV) with yellow background; aromatic (a; FHWY) in white with green background; tiny (t; AGS) in white with blue background; small (s; ACDGNPSTV) in blue; and charged (c; DEHKR) and the positively (+; HKR) or negatively (-; DE) subset of these in white with red background. Abbreviations: AC, adenylate cyclase; GGDEF, diguanylate cyclase; HK, histidine kinase; STK, serine/threonine kinase; TIR, the TIR domain-containing protein; UP, unknown protein. Species abbreviations: Ccre, *Caulobacter crescentus*; Drad, *Deinococcus radiodurans*; Nos, *Nostoc (Anabaena)* sp.; Paer, *Pseudomonas aeruginosa*; Rsol, *Ralstonia solanacearum*; Smel, *Sinorhizobium meliloti*; Syn, *Synechocystis* sp. The last column shows the NCBI protein database gene identification (gi) numbers.

histidine kinases



diguanylate cyclases



adenylate cyclases



serine/threonine kinases



unknown proteins

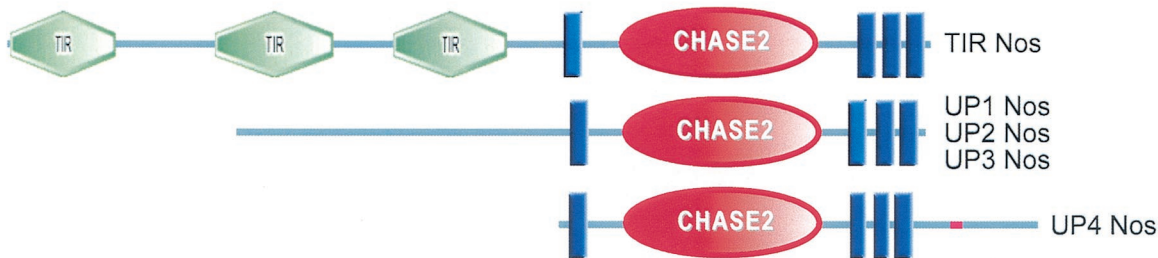


FIG. 2. Domain architectures of CHASE2-containing proteins. Abbreviations are described in the legend to Fig. 1. Domain architectures are from SMART (39), with modifications.

but are believed to have diguanylate cyclase (8, 48) and phosphodiesterase (21) activity, respectively (see reference 22 for a review). A recent survey detected Cache domains in more than 100 transmembrane sensors from various species, including the pathogens *Borrelia burgdorferi*, *Campylobacter jejuni*, *Helicobacter pylori*, *Pseudomonas aeruginosa*, *Treponema denticola*, and *Vibrio cholerae* (C. J. Shu, N. Lanka, and I. B. Zhulin, unpublished observations).

The CHASE (cyclase/histidine kinase-associated sensing extracellular) domain (4, 42) is another extracellular module that

is present in more than one type of sensory proteins. It is less common than the Cache domain and was detected in sensor histidine kinases, adenylate cyclases, and predicted diguanylate cyclases/phosphodiesterases, but not in chemotaxis transducers. It appears to recognize such stimuli as cytokines and short peptides that are important for the development program of an organism.

In this study, we analyzed the extracellular domains that are found in different classes of sensory proteins encoded in recently sequenced bacterial and archaeal genomes. We identi-

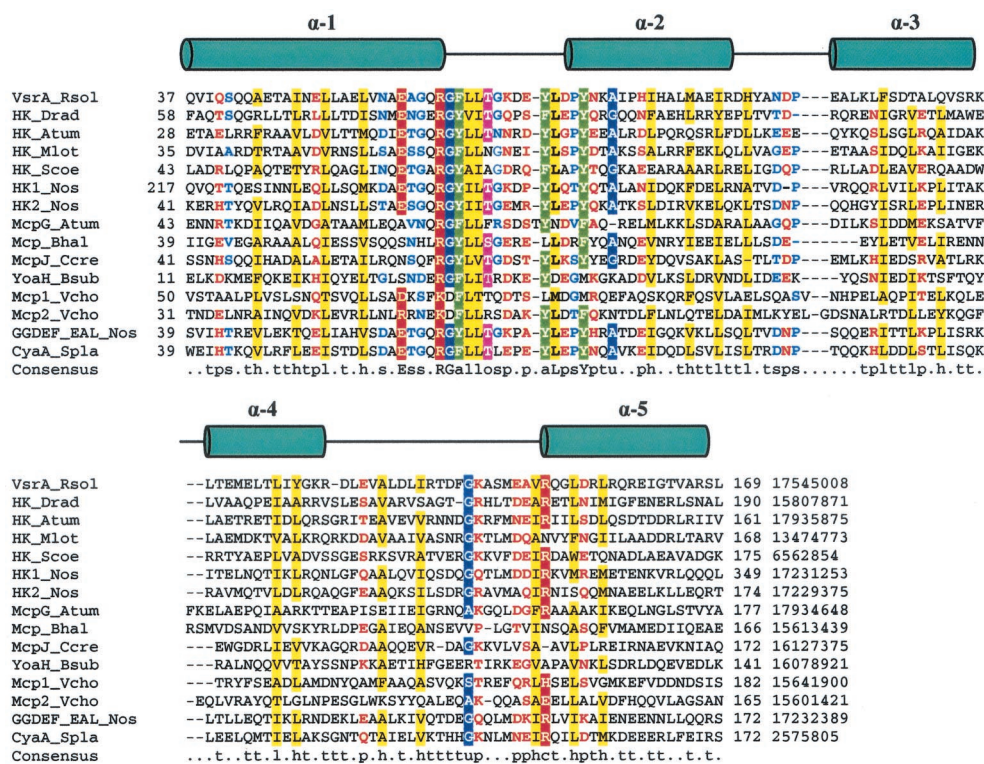


FIG. 3. Multiple alignment of the CHASE3 domain. Secondary-structure elements and consensus amino acids are described in the legend to Fig. 1. Species abbreviations: Atum, *Agrobacterium tumefaciens*; Bhal, *Bacillus halodurans*; Bsub, *Bacillus subtilis*; Ccre, *Caulobacter crescentus*; Drad, *Deinococcus radiodurans*; Mlot, *Mezorhizobium lotii*; Nos, *Nostoc (Anabaena) sp.*; Rsol, *Ralstonia solanacearum*; Scoe, *Streptomyces coelicolor*; Spla, *Spirulina platensis*; Vcho, *Vibrio cholerae*.

fied five such domains, including two domains that were found in as many as four classes of transmembrane receptors. We named these novel domains CHASE2 through CHASE6 in order to maintain a uniform nomenclature for the extracellular sensory domains and underscore their occurrence in histidine kinases and various other membrane sensors.

MATERIALS AND METHODS

Database searches. Similarity searches were performed against the nonredundant protein database and the database of finished and unfinished microbial genome sequences at the National Center for Biotechnology Information (NCBI, Bethesda, Md.). Standard BLAST searches (2) were performed with default parameters. Position-specific iterative (PSI) BLAST searches (2) were performed with the inclusion threshold (E value) of 0.01. Domain architecture analyses were carried out by individually comparing the protein sequences against the Conserved Domain Database at NCBI (<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>), using reverse-position-specific (RPS) BLAST, against the Clusters of Orthologous Groups (COG) database at NCBI (<http://www.ncbi.nlm.nih.gov/COG>) using its COGnitor tool (59) and by searching the SMART (39) and Pfam (9) databases with HMMer with default parameters. The ProDom database (13, 14) was searched for automatically generated alignments of protein sequence regions of interest.

Membrane topology prediction. Transmembrane regions were predicted by the TMHMM2 program (38) and refined with the Dense Alignment Surface algorithm (15). Signal peptides were predicted by the SignalP program (43).

Multiple alignments. Multiple alignments were constructed with Clustal (61) and T-Coffee (44) programs and manually adjusted based on the PSI-BLAST outputs. The consensus for each multiple alignment was calculated with a custom Perl script by Nigel Brown and Jianmei Lai (available at <http://www.bork.embl-heidelberg.de/Alignment/consensus.html>).

Structure prediction. Secondary-structure predictions were performed with the JPRED2 server (16), which utilizes several independent algorithms, includ-

ing the PHD program (52). Fold recognition was performed with the 3D-PSSM (position-specific scoring matrix) program (32).

Availability of results. The domains identified in this work will comprise new COGs in the upcoming release of the COG database (59). The corresponding alignments have been submitted to the SMART (39) and Pfam (9) domain databases and are available from the authors.

RESULTS

All novel domains were identified in the course of systematic genomic analysis of microbial signal transduction proteins (22, 69) and assigning putative signaling proteins from newly sequenced bacterial genomes to the COG database (59). They were initially found as conserved N-terminal domains fused to different C-terminal domains, such as histidine kinase, adenylate cyclase, predicted diguanylate cyclase (with or without the putative phosphodiesterase domain), and/or MCP domains and further defined with exhaustive, reciprocal PSI-BLAST searches initiated with predicted extracellular regions of each signal transduction protein.

Several conserved regions appeared to define novel domains which were not detected in current domain databases, such as SMART, Pfam, or COG. Threading of these regions of conservation against the database of known three-dimensional structures failed to identify a recognizable fold (no hits with more than 50% certainty based on the 3D-PSSM E value), suggesting that each of these domains had a novel structure.

CHASE2 domain. PSI-BLAST searches initiated with a predicted periplasmic region (residues 44 to 309) of a sensor

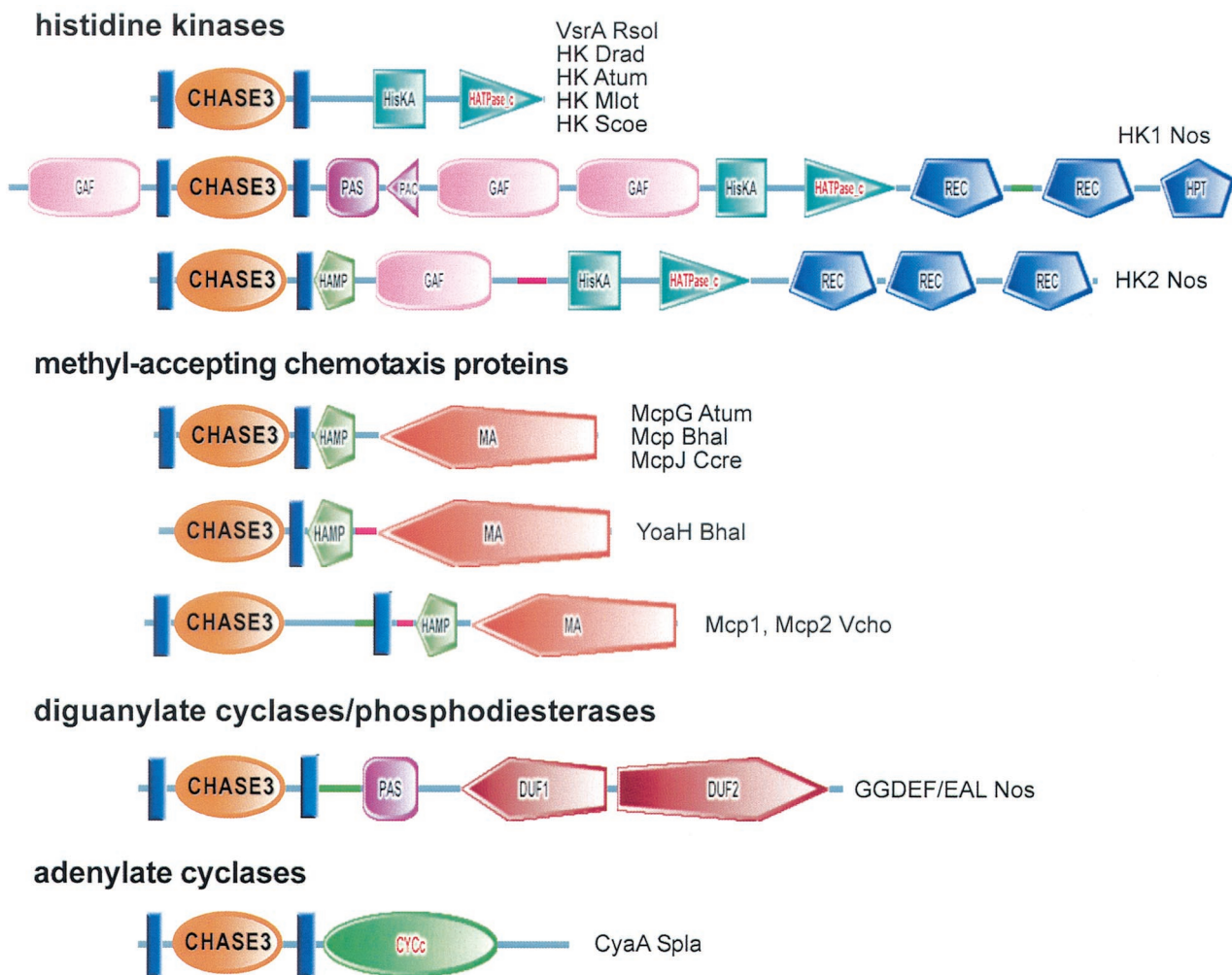


FIG. 4. Domain architectures of CHASE3-containing proteins. Abbreviations are described in the legends to Fig. 1 and 3. Domain architectures are from SMART (39), with modifications.

histidine kinase from *Pseudomonas aeruginosa* (gi 15599231; accession no. AAG07423) identified more than 30 proteins that contained similar regions. Figure 1 shows a multiple alignment of this domain, termed CHASE2 based on its extracellular location and occurrence in sensor cyclases and histidine kinases. CHASE2 comprises a 250- to 300-amino-acid-long domain that consists of six predicted α -helices and eight predicted β -strands. Conserved charged residues are signature elements of the CHASE2 domain: two in the α -1 helix, two bordering the β -1 strand, one in the α -2 helix, one in the α -2- α -3 loop, one in the α -4 helix, and two preceding the β -4 strand.

Analysis of the domain architecture of CHASE2-containing proteins revealed that they are present in four classes of sensory proteins from a variety of bacterial signal transduction pathways (Fig. 2): class I histidine kinases, adenylylase cyclases, predicted diguanylate cyclases, and serine/threonine kinases. The latter combination, found in two *Nostoc* sp. strain PCC 7120 (also referred to as *Anabaena* sp. strain PCC 7120) proteins (Fig. 2), differs from the cytoplasmically located serine/threonine kinases described earlier in *Nostoc* spp. (24, 67) and

in some other bacteria (33, 50, 55), as these proteins have a typical transmembrane receptor organization, indicating that they respond to the extracytoplasmic signals. In all cases, the CHASE domain is found N-terminally to the signaling domain and is flanked by predicted transmembrane regions, indicating an extracytoplasmic (periplasmic, thylakoidal) localization. The CHASE2 domain was always followed by three transmembrane regions (Fig. 2) but not the HAMP domain, which has been implicated in transmembrane signaling (6, 63).

CHASE3 domain. The CHASE3 domain was identified in PSI-BLAST searches initiated with a predicted periplasmic region (residues 29 to 183) of the adenylylase cyclase CyaA from a cyanobacterium, *Spirulina platensis* (gi 2575805; accession no. BAA22996) (66). Similar regions were identified in more than 20 proteins. Figure 3 shows a multiple alignment of CHASE3 domains. CHASE3 comprises a 130- to 150-amino-acid-long domain that appears to be entirely α -helical (four to six putative α -helices). No coiled-coil regions were predicted in this domain. The signature motif for the CHASE3 domain (Arg-Gly-aromatic-aliphatic-aliphatic-alcohol residue) is located in a highly conserved loop connecting the α -1 and α -2 helices

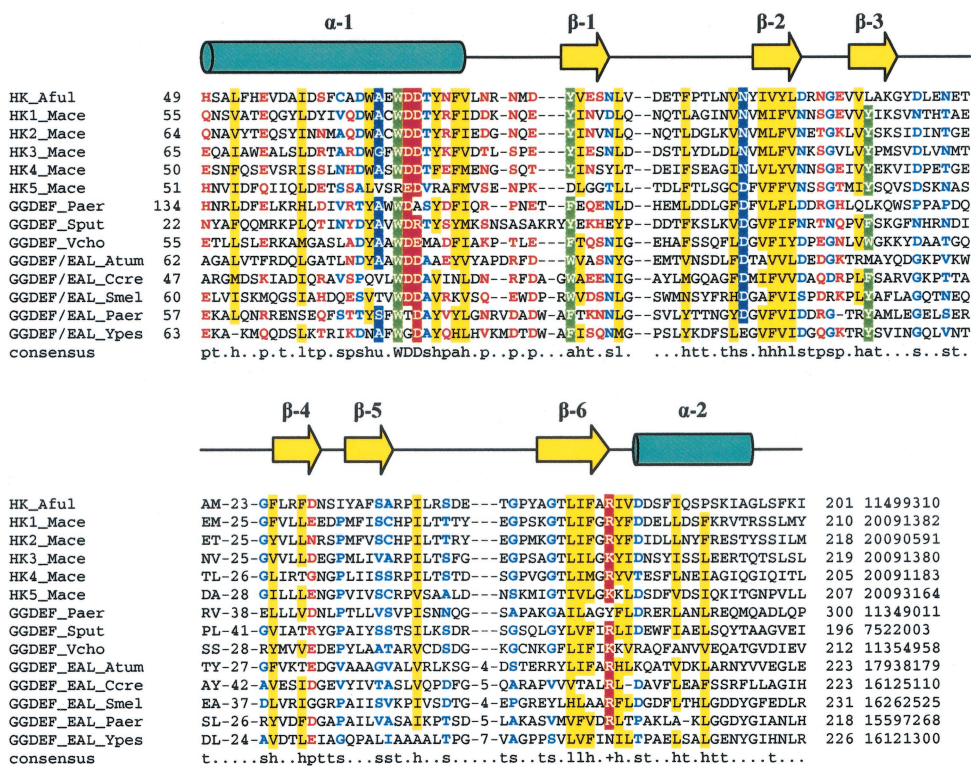


FIG. 5. Multiple alignment of CHASE4 domains. Secondary-structure elements and consensus amino acids are described in the legend to Fig. 1. Species abbreviations: Aful, *Archaeoglobus fulgidus*; Atum, *Agrobacterium tumefaciens*; Ccre, *Caulobacter crescentus*; Mace, *Methanosarcina acetivorans*; Paer, *Pseudomonas aeruginosa*; Smel, *Sinorhizobium meliloti*; Sput, *Shewanella putrefaciens*; Vcho, *Vibrio cholerae*; Ypes, *Yersinia pestis*.

(Fig. 3). Other noticeable elements of the CHASE3 domain are a conserved tyrosine in the α -2 helix and a conserved charged residue in the α -5 helix. Although no statistically significant hits (more than 50% certainty based on the 3D-PSSM E value) were obtained by threading of the CHASE3 domain against the database of known three-dimensional structures, the CHASE3 domain showed remote similarity to the periplasmic ligand-binding domain of the aspartate chemoreceptor Tar from both *Escherichia coli* (3D-PSSM E value, 2.71) and *Salmonella enterica* serovar Typhimurium (3D-PSSM E value, 3.69). This similarity may reflect an antiparallel helical bundle characteristic of the Tar protein (41).

Analysis of the domain architecture of CHASE3-containing proteins revealed that it occurs in four classes of sensory proteins (Fig. 4). CHASE3 domains are present in class I histidine kinases, methyl-accepting chemotaxis proteins, adenylate cyclases, and diguanylate cyclases. In multidomain, receptor-like proteins, the CHASE3 domain is found N-terminal to the signaling domain and is flanked by two predicted transmembrane regions, indicating the extracytoplasmic localization. The CHASE3 domain is often found in combination with the intracellular sensory domains PAS and GAF. Some CHASE3-containing proteins have the HAMP domain, whereas others do not (Fig. 4).

CHASE4 domain. The CHASE4 domain was identified in PSI-BLAST searches initiated with a predicted periplasmic region (residues 47 to 306) of a predicted sensor histidine kinase from the hyperthermophilic archaeon *Archaeoglobus fulgidis* (gi 11499310; accession no. AAB89530). The novel domain was identified in 19 proteins. Figure 5 shows a multiple

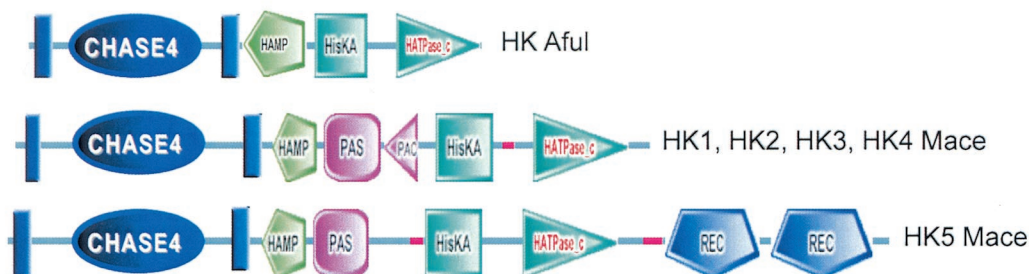
alignment of CHASE4 domains. This is a 150- to 160-amino-acid-long domain with an apparent $\alpha\beta$ fold: six β -strands flanked by two α -helices. An extensive (20 to 45 amino acid residues) loop is predicted in the middle portion of the CHASE4 domain (omitted from the alignment). The signature motif of the novel domain, WDD (Trp-Asp-Asp), is located in the first long α -helix. Similar to CHASE2 and CHASE3, the CHASE4 domain is found exclusively in signal transduction proteins. We detected CHASE4 in only two classes, sensor histidine kinases and diguanylate cyclases/phosphodiesterases (Fig. 6). Some CHASE4-containing proteins also have HAMP and PAS domains. Like other extracellular sensory domains, CHASE4 is always found between two transmembrane regions N-terminal to known signaling domains.

Identification of CHASE5 and CHASE6 sensory domains. Careful analysis of domain architectures of the transmembrane receptors encoded in complete microbial genomes identified two more extracellular domains that were shared by two classes of transmembrane receptors. The extracellular domain common to histidine kinases and predicted diguanylate cyclases was named CHASE5, and another extracellular domain found in histidine kinases and the HDc phosphodiesterases was named CHASE6 (Fig. 7 and Table 1).

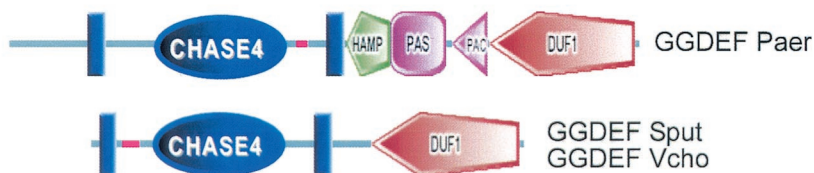
DISCUSSION

The ability to sense extracellular stimuli is a fundamental property of living cells. However, little is known about the diversity of extracellular sensor elements of biological receptors. By using sensitive database searches combined with de-

histidine kinases



diguanylate cyclases



diguanylate cyclases/phosphodiesterases

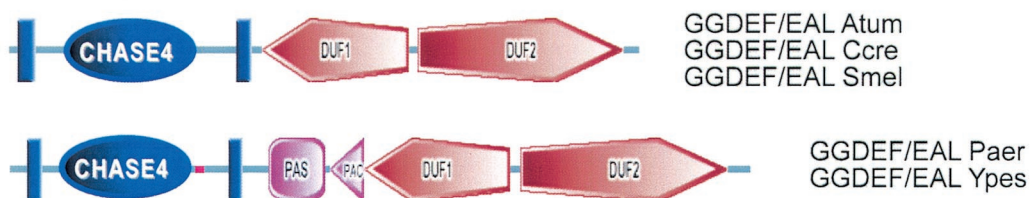


FIG. 6. Domain architectures of CHASE4-containing proteins. Abbreviations are described in the legends to Fig. 1 and 5. Domain architectures are from SMART (39), with modifications.

tailed protein sequence analysis, we have identified several novel extracellular sensory domains that are found in various bacterial transmembrane receptors. All of the domains are found in sensor histidine kinases and nucleotide (adenylate and diguanylate) cyclases, similar to the distribution of two previously described extracellular sensory domains, Cache (3, 69) and CHASE (4, 42). CHASE2 was recently described as being an exclusively cyanobacterial (CMS, cyanobacterial membrane sensor) domain (45). Indeed, in the cyanobacterium *Nostoc* sp. strain PCC 7120, the CHASE2 domain is present in 13 proteins. However, our results demonstrate that CHASE2 is found in transmembrane receptors in many species outside the cyanobacterial lineage. These include *Deinococcus radiodurans*, spirochetes, and various proteobacterial species (Fig. 1 and 2). Similarly, the CHASE3 domain is present in several proteins in cyanobacteria but is also found in *D. radiodurans* and various gram-positive bacteria and proteobacteria (Fig. 3 and 4).

While neither CHASE2 nor CHASE3 has been found in Archaea, CHASE4 and CHASE6 domains are encoded in archaeal species as well as in representatives of gram-positive bacteria and proteobacteria (CHASE4) and cyanobacteria (CHASE6). Thus, novel sensory domains have a wide phyletic

distribution. Two-component regulatory systems are found in only few species of *Archaea* and could have been acquired from *Bacteria* through lateral gene transfer (34, 37). Thus, the prevalence of CHASE-like sensory domains in *Bacteria* is not surprising.

Some of the proteins that contain newly defined domains have been studied experimentally, for example, the CHASE3-containing adenylate cyclase CyaA from the cyanobacteria *Spirulina platense* and *Nostoc* sp. strain PCC 7120 (31, 66) and histidine kinase VsrA from *Ralstonia* (formerly *Pseudomonas solanacearum*) (54). The histidine kinase VsrA was identified as a critical sensor required for expression of virulence factors, both polysaccharides and proteins, in this wilt-inducing phytopathogen. CHASE3-containing sensors are also found in other important pathogens of plants and mammals, such as *Pseudomonas syringae*, *Burkholderia fungorum*, *Legionella pneumophila*, and *Bacillus anthracis* (unpublished genomes; data not shown). A recent study of the adenylate cyclase CyaA from *Myxococcus xanthus* demonstrated that this CHASE2 domain-containing protein participates in signal transduction during osmotic stress and might function as an osmosensor (35). Thus, it is an attractive hypothesis that the CHASE2 domain is the

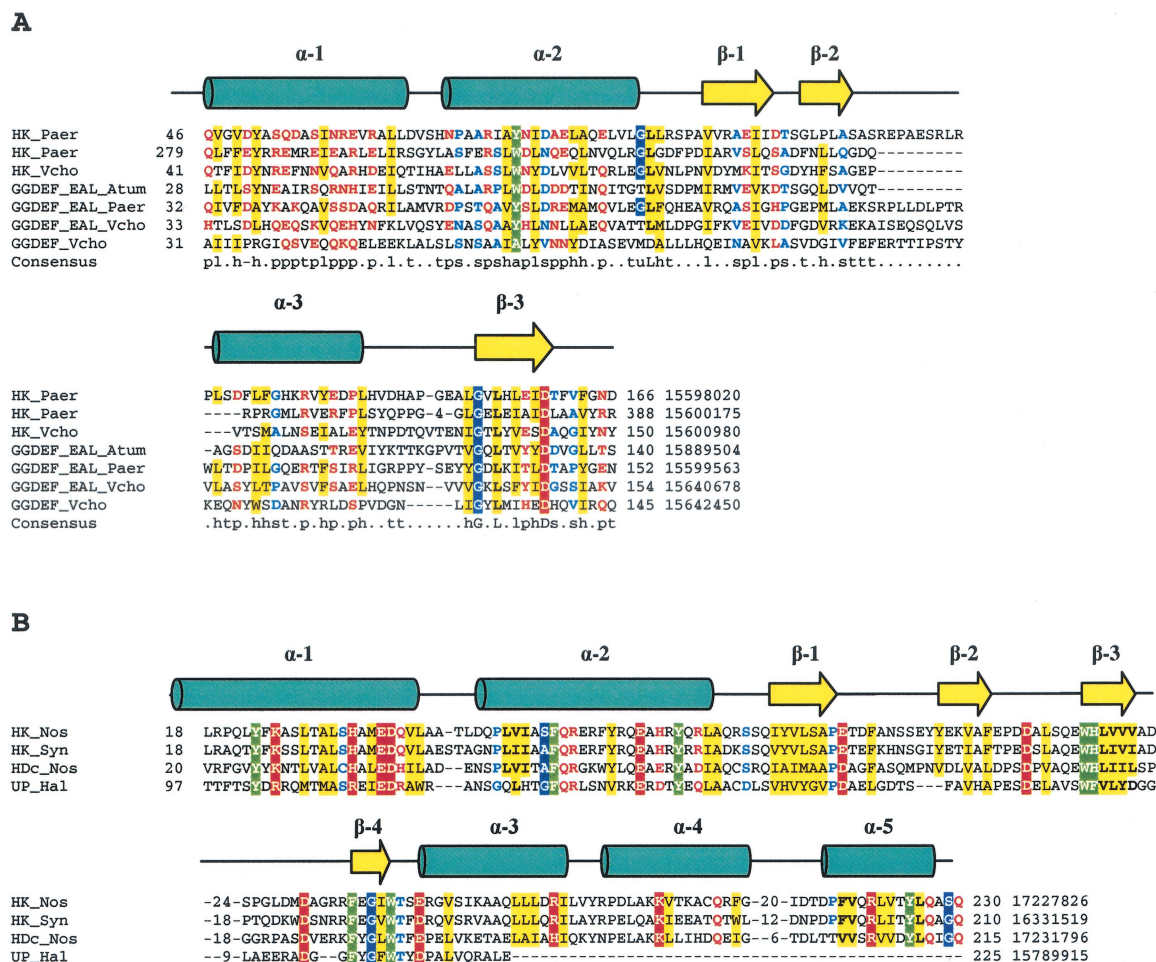


FIG. 7. Multiple alignments of CHASE5 (A) and CHASE6 (B) domains. Secondary-structure elements and domain abbreviations are described in the legend to Fig. 1 except for Hdc, the phosphohydrolase (phosphodiesterase) domain. Species abbreviations: Atum, *Agrobacterium tumefaciens*; Hal, *Halobacterium* sp.; Nos, *Nostoc (Anabaena)* sp.; Paer, *Pseudomonas aeruginosa*; Syn, *Synechocystis* sp.; Vcho, *Vibrio cholerae*.

osmosensing module in all other types of transmembrane receptors in which it is found.

At this time, the stimuli that are recognized by other novel domains are unknown. Identification of conserved residues in

CHASE-like domains presented in this work, such as the RG-FLLT motif in the VsrA protein (Fig. 3), may assist investigators in future studies of the ligand specificity of various transmembrane receptors.

TABLE 1. CHASE-type sensors in various signal transduction proteins from prokaryotes

Sensory domain	Function, based on presence of a conserved signaling domain ^a								Reference(s)
	Histidine kinase (HATPase _c)	MCP (MA)	Adenylate cyclase (CYC _c)	Diguanylate cyclase/phosphodiesterase (GGDEF/EAL)	Hydrolase/phosphodiesterase (HDC)	Ser/Thr kinase (STYK _c)	Protein phosphatase (PP2C)	Enzyme (cupin)	
Cache	X	X	X	X	X			X	3, 69
CHASE	X		X	X	X				4, 42
CHASE2	X		X	X		X			This work
CHASE3	X	X	X	X					This work
CHASE4	X			X					This work
CHASE5	X			X			X ^b		This work
CHASE6	X				X				This work

^a Domain abbreviations are from the SMART database (39) or from the Pfam data base (9). Cupin, a double-stranded β -helix domain, is found in mannose-6-phosphate isomerase and many other proteins (17). An X indicates the presence of the given domain combination in the NCBI protein database and in the database of unfinished genomes as of 1 June 2002.

^b The CHASE5-protein phosphatase 2C fusion has been found only in the unfinished genome of *Magnetospirillum magnetotacticum* (data not shown). The protein phosphatase 2C family of bacterial protein phosphatases is exemplified by the RsbU protein (65).

The CHASE2 and CHASE4 domains have not been found in MCPs. However, as sensor elements of histidine kinases, they are present in proteobacterial species that have large numbers of MCPs, for example, *Pseudomonas aeruginosa* (57), *Ralstonia solanacearum* (53), and *Vibrio cholerae* (27). This observation suggests that CHASE2 and CHASE4 recognize stimuli that are not important for the immediate motility response but might be critical for regulation of metabolism or cell development. It is remarkable that in archaea, CHASE4 appears to be associated exclusively with histidine kinases, while in bacteria it is associated exclusively with diguanylate cyclases (Fig. 6). Uncovering the reasons for this dichotomy should help in understanding the principles of signal transduction in these two groups of prokaryotes.

It is important to note that the complex multidomain organization of transmembrane sensors seriously complicates functional annotation of these proteins. In the course of microbial genome sequencing projects, signaling proteins containing the CHASE2, CHASE3, and CHASE4 domains have been repeatedly misannotated and deposited in GenBank under obscure or erroneous names. For example, the histidine kinase (All3347, gi 17230839) and an uncharacterized protein (Alr0357, gi 17227853) from *Nostoc* sp. strain PCC 7120 (Fig. 2) were both annotated as adenylate cyclases (30), although their similarity to adenylate cyclases was limited to the common N-terminal CHASE2 domain. Curiously, three virtually identical CHASE2-containing proteins were annotated as adenylate cyclase (All7310, gi 17233326), similar to adenylate cyclase (All3180, gi 17230672), and a hypothetical protein (Alr1378, gi 17228873), although none of them contained the cyclase domain (Fig. 2). In fact, annotation of transmembrane sensors as unknown, hypothetical, or conserved hypothetical proteins is quite common.

As we have argued earlier, short of a systematic mistake in sequencing, a protein that is conserved across diverse phylogenetic lineages is not hypothetical anymore (20). We hope that delineation of the novel sensor domains in this work would help simplify the recognition of transmembrane receptor proteins and eventually improve their annotation. It is clear, however, that fully consistent annotation of the CHASE-type domain-containing sensors will be impossible without experimentally identifying their ligands and understanding the full set of environmental signals sensed by these domains.

Although most of the novel domains described here are found in transmembrane receptors, they might utilize different mechanisms of signal transduction across the membrane. CHASE3 and CHASE4 are often found in proteins that also contain the HAMP domain (6, 63). Moreover, the CHASE3 domain shows remote (likely topological) similarity to the periplasmic-ligand binding domain of the HAMP domain-containing Tar chemoreceptor (41). Therefore, in CHASE3-containing receptors, signal translocation across the membrane might be similar to the piston model suggested for Tar (12, 46). However, the CHASE2 domain is never found in combination with the HAMP domain but is always followed by three transmembrane helices. This unusual motif occurred without exception in all proteins in which CHASE2 was detected (Fig. 2) and might be an attractive subject for future studies on novel mechanisms of transmembrane signaling.

In summary, our results indicate that various transmem-

brane receptors that transduce signals into diverse regulatory pathways may utilize similar sensory (input) domains. The pervasiveness of a particular sensor domain in receptors from diverse regulatory networks within a given organism suggests that the stimuli recognized by this domain might be important for the biology of the organism.

ACKNOWLEDGMENTS

This work was supported, in part, by National Science Foundation grant EIA-0219079 (to I.B.Z.).

We thank Darren Natale for pointing out the cupin output domain. We acknowledge the availability of unfinished genome sequences from the Institute for Genomic Research, the Sanger Institute, and the DOE Joint Genome Institute.

REFERENCES

- Aizawa, S.-I., I. B. Zhulin, L. Marquez-Magana, and G. W. Ordal. 2002. Chemotaxis and motility, p. 437–452. In A. L. Sonenshein, J. A. Hoch, and R. Losik (ed.), *Bacillus subtilis* and its closest relatives: from genes to cells. ASM Press, Washington, D.C.
- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zheng, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST—a new generation of protein database search programs. *Nucleic Acids Res.* **25**:3389–3402.
- Anantharaman, V., and L. Aravind. 2000. Cache—a signaling domain common to animal Ca²⁺-channel subunits and a class of prokaryotic chemotaxis receptors. *Trends Biochem. Sci.* **25**:535–537.
- Anantharaman, V., and L. Aravind. 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.
- Appleby, J. L., J. S. Parkinson, and R. B. Bourret. 1996. Signal transduction via the multistep phosphorelay: not necessarily a road less traveled. *Cell* **86**:845–848.
- Aravind, L., and C. P. Ponting. 1999. The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins. *FEMS Microbiol. Lett.* **176**:111–116.
- Aravind, L., and C. P. Ponting. 1997. The GAF domain: an evolutionary link between diverse phototransducing proteins. *Trends Biochem. Sci.* **22**:458–459.
- Ausmees, N., R. Mayer, H. Weinhouse, G. Volman, D. Amikam, M. Ben-Ziman, and M. Lindberg. 2001. Genetic data indicate that proteins containing the GGDEF domain possess diguanylate cyclase activity. *FEMS Microbiol. Lett.* **204**:163–167.
- Bateman, A., E. Birney, L. Cerruti, R. Durbin, L. Etwiler, S. R. Eddy, S. Griffiths-Jones, K. L. Howe, M. Marshall, and E. L. Sonnhammer. 2002. The Pfam protein families database. *Nucleic Acids Res.* **30**:276–280.
- Bilwes, A. M., L. A. Alex, B. R. Crane, and M. I. Simon. 1999. Structure of CheA, a signal-transducing histidine kinase. *Cell* **96**:131–141.
- Bourret, R. B., and A. M. Stock. 2002. Molecular information processing: lessons from bacterial chemotaxis. *J. Biol. Chem.* **277**:9625–9628.
- Chervitz, S. A., and J. J. Falke. 1996. Molecular mechanism of transmembrane signaling by the aspartate receptor: a model. *Proc. Natl. Acad. Sci. USA* **93**:2545–2550.
- Corpet, F., J. Gouzy, and D. Kahn. 1999. Recent improvements of the ProDom database of protein domain families. *Nucleic Acids Res.* **27**:263–267.
- Corpet, F., F. Servant, J. Gouzy, and D. Kahn. 2000. ProDom and ProDomCG: tools for protein domain analysis and whole genome comparisons. *Nucleic Acids Res.* **28**:267–269.
- Cserzo, M., E. Wallin, I. Simon, G. von Heijne, and A. Elofsson. 1997. Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein Eng.* **10**:673–676.
- Cuff, J. A., and G. J. Barton. 2000. Application of multiple sequence alignment profiles to improve protein secondary structure prediction. *Proteins* **40**:502–511.
- Dunwell, J. M., S. Khuri, and P. J. Gane. 2000. Microbial relatives of the seed storage proteins of higher plants: conservation of structure and diversification of function during evolution of the cupin superfamily. *Microbiol. Mol. Biol. Rev.* **64**:153–179.
- Dutta, R., L. Qin, and M. Inouye. 1999. Histidine kinases: diversity of domain organization. *Mol. Microbiol.* **34**:633–640.
- Falke, J. J., and G. L. Hazelbauer. 2001. Transmembrane signaling in bacterial chemoreceptors. *Trends Biochem. Sci.* **26**:257–265.
- Galperin, M. Y. 2001. Conserved “hypothetical” proteins: new hints and new puzzles. *Comp. Funct. Genomics* **2**:1. 4–18.
- Galperin, M. Y., D. A. Natale, L. Aravind, and E. V. Koonin. 1999. A specialized version of the HD hydrolase domain implicated in signal transduction. *J. Mol. Microbiol. Biotechnol.* **1**:303–305.

22. Galperin, M. Y., A. N. Nikolskaya, and E. V. Koonin. 2001. Novel domains of the prokaryotic two-component signal transduction system. *FEMS Microbiol. Lett.* **203**:11–21.
23. Garrity, L. F., S. L. Schiel, R. Merrill, J. Reizer, M. H. Saier, Jr., and G. W. Ordal. 1998. Unique regulation of carbohydrate chemotaxis in *Bacillus subtilis* by the phosphoenolpyruvate-dependent phosphotransferase system and the methyl-accepting chemotaxis protein McpC. *J. Bacteriol.* **180**:4475–4480.
24. Gonzalez, L., V. Phalip, and C. C. Zhang. 2001. Characterization of PknC, a Ser/Thr kinase with broad substrate specificity from the cyanobacterium *Anabaena* sp. strain PCC 7120. *Eur. J. Biochem.* **268**:1869–1875.
25. Guo, Y. L., T. Seebacher, U. Kurz, J. U. Linder, and J. E. Schultz. 2001. Adenylyl cyclase Rv1625c of *Mycobacterium tuberculosis*: a progenitor of mammalian adenylyl cyclases. *EMBO J.* **20**:3667–3675.
26. Hanlon, D. W., and G. W. Ordal. 1994. Cloning and characterization of genes encoding methyl-accepting chemotaxis proteins in *Bacillus subtilis*. *J. Biol. Chem.* **269**:14038–14046.
27. Heidelberg, J. F., J. A. Eisen, W. C. Nelson, R. A. Clayton, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, J. D. Peterson, L. Umayam, S. R. Gill, K. E. Nelson, T. D. Read, H. Tetelin, D. Richardson, M. D. Ermolaeva, J. Vamathevan, S. Bass, H. Qin, I. Dragoi, P. Sellers, L. McDonald, T. Utterback, R. D. Fleischmann, W. C. Nierman, and O. White. 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* **406**:477–483.
28. Hoch, J. A. 2000. Two-component and phosphorelay signal transduction. *Curr. Opin. Microbiol.* **3**:165–170.
29. Hoch, J. A., and K. I. Varughese. 2001. Keeping signals straight in phosphorelay signal transduction. *J. Bacteriol.* **183**:4941–4949.
30. Kaneko, T., Y. Nakamura, C. P. Wolk, T. Kuritz, S. Sasamoto, A. Watanabe, M. Iriguchi, A. Ishikawa, K. Kawashima, T. Kimura, Y. Kishida, M. Kohara, M. Matsumoto, A. Matsuno, A. Muraki, N. Nakazaki, S. Shimpo, M. Sugimoto, M. Takazawa, M. Yamada, M. Yasuda, and S. Tabata. 2001. Complete genome sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. *DNA Res.* **8**:205–213.
31. Katayama, M., and M. Ohmori. 1997. Isolation and characterization of multiple adenylyl cyclase genes from the cyanobacterium *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **179**:3588–3593.
32. Kelley, L. A., R. M. MacCallum, and M. J. Sternberg. 2000. Enhanced genome annotation with structural profiles in the program 3D-PSSM. *J. Mol. Biol.* **299**:499–520.
33. Kennelly, P. J., and M. Potts. 1996. Fancy meeting you here! A fresh look at “prokaryotic” protein phosphorylation. *J. Bacteriol.* **178**:4759–4764.
34. Kim, D., and S. Forst. 2001. Genomic analysis of the histidine kinase family in bacteria and archaea. *Microbiology* **147**:1197–1212.
35. Kimura, Y., Y. Mishima, H. Nakano, and K. Takegawa. 2002. An adenylyl cyclase, CyaA, of *Myxococcus xanthus* functions in signal transduction during osmotic stress. *J. Bacteriol.* **184**:3578–3585.
36. Kokoeva, M. V., and D. Oesterhelt. 2000. BasT, a membrane-bound transducer protein for amino acid detection in *Halobacterium salinarum*. *Mol. Microbiol.* **35**:647–656.
37. Koretke, K. K., A. N. Lupas, P. V. Warren, M. Rosenberg, and J. R. Brown. 2000. Evolution of two-component signal transduction. *Mol. Biol. Evol.* **17**:1956–1970.
38. Krogh, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**:567–580.
39. Letunic, I., L. Goodstadt, N. J. Dickens, T. Doerks, J. Schultz, R. Mott, F. Ciccarelli, R. R. Copley, C. P. Ponting, and P. Bork. 2002. Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Res.* **30**:242–244.
40. McCue, L. A., K. A. McDonough, and C. E. Lawrence. 2000. Functional classification of cNMP-binding proteins and nucleotide cyclases with implications for novel regulatory pathways in *Mycobacterium tuberculosis*. *Genome Res.* **10**:204–219.
41. Milburn, M. V., G. G. Prive, D. L. Milligan, W. G. Scott, J. Yeh, J. Jancarik, D. E. Koshland, Jr., and S. H. Kim. 1991. Three-dimensional structures of the ligand-binding domain of the bacterial aspartate receptor with and without a ligand. *Science* **254**:1342–1347.
42. Mougel, C., and I. B. Zhulin. 2001. CHASE: an extracellular sensing domain common to transmembrane receptors from prokaryotes, lower eukaryotes and plants. *Trends Biochem. Sci.* **26**:582–584.
43. Nielsen, H., J. Engelbrecht, S. Brunak, and G. von Heijne. 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10**:1–6.
44. Notredame, C., D. G. Higgins, and J. Heringa. 2000. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**:205–217.
45. Ohmori, M., M. Ikeuchi, N. Sato, P. Wolk, T. Kaneko, T. Ogawa, M. Kanehisa, S. Goto, S. Kawashima, S. Okamoto, H. Yoshimura, H. Katoh, T. Fujisawa, S. Ehira, A. Kamei, S. Yoshihara, R. Narikawa, and S. Tabata. 2001. Characterization of genes encoding multidomain proteins in the genome of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. *DNA Res.* **8**:271–284.
46. Ottemann, K. M., W. Xiao, Y. K. Shin, and D. E. Koshland, Jr. 1999. A piston model for transmembrane signaling of the aspartate receptor. *Science* **285**:1751–1754.
47. Parkinson, J. S., and E. C. Kofoid. 1992. Communication modules in bacterial signaling proteins. *Annu. Rev. Genet.* **26**:71–112.
48. Pei, J., and N. V. Grishin. 2001. GGDEF domain is homologous to adenylyl cyclase. *Proteins* **42**:210–216.
49. Ponting, C. P., and L. Aravind. 1997. PAS: a multifunctional domain family comes to light. *Curr. Biol.* **7**:R674–R677.
50. Ponting, C. P., L. Aravind, J. Schultz, P. Bork, and E. V. Koonin. 1999. Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer. *J. Mol. Biol.* **289**:729–745.
51. Reid, C. J., and P. S. Poole. 1998. Roles of DctA and DctB in signal detection by the dicarboxylic acid transport system of *Rhizobium leguminosarum*. *J. Bacteriol.* **180**:2660–2669.
52. Rost, B. 1996. PHD: predicting one-dimensional protein structure by profile-based neural networks. *Methods Enzymol.* **266**:525–539.
53. Salanoubat, M., S. Genin, F. Artiguenave, J. Gouzy, S. Mangenot, M. Arlat, A. Billault, P. Brottier, J. C. Camus, L. Cattolico, M. Chandler, N. Choinsne, C. Claudel-Renard, S. Cunnac, N. Demange, C. Gaspin, M. Lavie, A. Moisan, C. Robert, W. Saurin, T. Schiex, P. Signier, P. Thebault, M. Whalen, P. Wincker, M. Levy, J. Weissenbach, and C. A. Boucher. 2002. Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature* **415**:497–502.
54. Schell, M. A., T. P. Denny, and J. Huang. 1994. VsrA, a second two-component sensor regulating virulence genes of *Pseudomonas solanacearum*. *Mol. Microbiol.* **11**:489–500.
55. Shi, L., M. Potts, and P. J. Kennelly. 1998. The serine, threonine, and/or tyrosine-specific protein kinases and protein phosphatases of prokaryotic organisms: a family portrait. *FEMS Microbiol. Rev.* **22**:229–253.
56. Stock, A. M., V. L. Robinson, and P. N. Goudreau. 2000. Two-component signal transduction. *Annu. Rev. Biochem.* **69**:183–215.
57. Stover, C. K., X. Q. Pham, A. L. Erwin, S. D. Mizoguchi, P. Warren, M. J. Hickey, F. S. Brinkman, W. O. Hufnagle, D. J. Kowalik, M. Lagrou, R. L. Garber, L. Goltry, E. Tolentino, S. Westbrook-Wadman, Y. Yuan, L. L. Broder, S. N. Coulter, K. R. Folger, A. Kas, K. Larbig, R. Lim, K. Smith, D. Spencer, G. K. Wong, Z. Wu, I. T. Paulsen, J. Reizer, M. H. Saier, R. E. Hancock, S. Lory, and M. V. Olson. 2000. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* **406**:959–964.
58. Tal, R., H. C. Wong, R. Calhoun, D. Gelfand, A. L. Fear, G. Volman, R. Mayer, P. Ross, D. Amikam, H. Weinhouse, A. Cohen, S. Sapir, P. Ohana, and M. Benziman. 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.
59. Tatusov, R. L., M. Y. Galperin, D. A. Natale, and E. V. Koonin. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28**:33–36.
60. Taylor, B. L., and I. B. Zhulin. 1999. PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiol. Mol. Biol. Rev.* **63**:479–506.
61. Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673–4680.
62. West, A. H., and A. M. Stock. 2001. Histidine kinases and response regulator proteins in two-component signaling systems. *Trends Biochem. Sci.* **26**:369–376.
63. Williams, S. B., and V. Stewart. 1999. Functional similarities among two-component sensors and methyl-accepting chemotaxis proteins suggest a role for linker region amphipathic helices in transmembrane signal transduction. *Mol. Microbiol.* **33**:1093–1102.
64. Wu, H., J. Kato, A. Kuroda, T. Ikeda, N. Takiguchi, and H. Ohtake. 2000. Identification and characterization of two chemotactic transducers for inorganic phosphate in *Pseudomonas aeruginosa*. *J. Bacteriol.* **182**:3400–3404.
65. Yang, X., C. M. Kang, M. S. Brody, and C. W. Price. 1996. Opposing pairs of serine protein kinases and phosphatases transmit signals of environmental stress to activate a bacterial transcription factor. *Genes Dev.* **10**:2265–2275.
66. Yashiro, K., T. Sakamoto, and M. Ohmori. 1996. Molecular characterization of an adenylyl cyclase gene of the cyanobacterium *Spirulina platensis*. *Plant Mol. Biol.* **31**:175–181.
67. Zhang, C. C. 1993. A gene encoding a protein related to eukaryotic protein kinases from the filamentous heterocystous cyanobacterium *Anabaena* PCC 7120. *Proc. Natl. Acad. Sci. USA* **90**:11840–11844.
68. Zhulin, I. B. 2000. A novel phototaxis receptor hidden in the cyanobacterial genome. *J. Mol. Microbiol. Biotechnol.* **2**:491–493.
69. Zhulin, I. B. 2001. The superfamily of chemotaxis transducers: from physiology to genomics and back. *Adv. Microb. Physiol.* **45**:157–198.
70. Zhulin, I. B., B. L. Taylor, and R. Dixon. 1997. PAS domain S-boxes in Archaea, Bacteria and sensors for oxygen and redox. *Trends Biochem. Sci.* **22**:331–333.