Bacterial signal transduction network in a genomic perspective†

Michael Y. Galperin

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, USA.

Summary

Bacterial signalling network includes an array of numerous interacting components that monitor environmental and intracellular parameters and effect cellular response to changes in these parameters. The complexity of bacterial signalling systems makes comparative genome analysis a particularly valuable tool for their studies. Comparative studies revealed certain general trends in the organization of diverse signalling systems. These include (i) modular structure of signalling proteins; (ii) common organization of signalling components with the flow of information from N-terminal sensory domains to the C-terminal transmitter or signal output domains (N-to-C flow); (iii) use of common conserved sensory domains by different membrane receptors; (iv) ability of some organisms to respond to one environmental signal by activating several regulatory circuits; (v) abundance of intracellular signalling proteins, typically consisting of a PAS or GAF sensor domains and various output domains; (vi) importance of secondary messengers, cAMP and cyclic diguanylate; and (vii) crosstalk between components of different signalling pathways. Experimental characterization of the novel domains and domain combinations would be needed for achieving a better understanding of the mechanisms of signalling response and the intracellular hierarchy of different signalling pathways.

Introduction: from sequence gazing to genome grazing

Bacterial signal transduction systems provide a fascinating array of numerous interacting components that sense changes in a variety of environmental and intracellular parameters and transmit these signals to various cellular mechanisms to cause adaptive changes in metabolism, physiology and/or behaviour (for reviews, see Hoch and Silhavy, 1995; Grebe and Stock, 1999; Stock *et al*., 2000; Inouye and Dutta, 2003). The complexity of signalling networks in model organisms, such as *Escherichia coli* and *Bacillus subtilis*, has long hindered their systematic analysis. The first description of a two-component system by Ninfa and Magasanik (1986) was quickly followed by the discovery of crosstalk between nitrogen assimilation and chemotaxis (Ninfa *et al*., 1988), suggesting complex interactions between different regulatory systems and signal integration. However, various components of the bacterial signalling machinery – histidine kinases and response regulators, cyclic AMP (cAMP)-dependent systems, phosphotransferase system components and chemotaxis proteins – were traditionally viewed and studied as separate entities. Certain components of the signal transduction circuits,

such as adenylate cyclases, diguanylate cyclases and phosphodiesterases, and serine/threonine protein kinases and phosphatases, were identified only recently and many remain poorly characterized. The availability of complete genomic sequences from numerous bacterial species allowed the researchers for the first time to evaluate the total number and composition of the signal transduction proteins encoded in each particular genome and finally appreciate the complexity of the whole system. In a way, instead of describing legs, trunk and tail of the same elephant, we can now take a look at the entire elephant. Studies of signal transduction offer one of the first examples where genomics has actually lead to important biological insights that would have been otherwise impossible. It has become clear that, in addition to sensory histidine kinases and methyl-accepting chemotaxis proteins, bacteria have other receptor proteins with similar overall organization, namely with an N-terminal periplasmic or extracytoplasmic sensory domain, followed by one or more transmembrane segments and a cytoplasmically located signal transduction domain. This kind of organization has been described for membrane-anchored adenylate cyclases, putative diguanylate cyclases and phosphodiesterases, serine/threonine protein kinases and phosphatases, revealing a much more complex signalling network than has been generally assumed before genomics (Galperin *et al.*, 2001a; Kennelly, 2002). Another interesting result was gleaned from the absence of certain genes in a genome, namely the absence of membrane-associated components (EIIB and EIIC) of the phosphoenolpyruvate-dependent sugar:phosphotransferase system (PTS) in *Xylella fastidiosa* and several other bacteria that still encode the soluble components of PTS (EI, HPr and EIIA). This observation, which could not have been made without knowledge of the complete genome, suggested that soluble PTS components in these organisms are involved solely in signalling. In other words, studies of signal transduction have grown from 'sequence gazing' (a term coined by Henikoff, 1991), to 'genome grazing' when comparative genomics became an integral part of discovery and analysis. The rapid rate of genome sequencing (148 complete prokaryotic genomes available in GenBank® at the end of 2003) is contributing to the progress in comparative genome analysis.

When the first bacterial genomes were sequenced, the first order of business, of course, was to enumerate the genes encoding signal transduction proteins in each organism and to perform cross-species comparisons to determine which systems are common and which are specific for a given species or genus. Several independent 'censuses' of bacterial signal transduction proteins (Mizuno *et al.*, 1996; Mizuno, 1997; Koretke *et al.*, 2000; Galperin *et al.*, 2001a; Ashby, 2004) brought very similar results, leading to a general consensus on the distribution of various signalling systems in various microorganisms (see Table 1). It has been found, for example, that parasitic bacteria usually encode fewer signalling proteins than free-living bacteria, even if one takes into account their smaller genome sizes. Gram-positive bacteria and archaea turned out to have fewer signal transduction proteins than proteobacteria or cyanobacteria of the same genome size (Galperin *et al.*, 2001a).

In the beginning of the genome era, we have predicted that the new paradigm of genome-based microbiology would eventually replace the old paradigm of gene-by-gene approach (Koonin and Galperin, 1997). Now, seven years later, this process of paradigm shift can be seen in earnest. The ever-improving coverage of microbial diversity by complete genome sequences allows increasingly accurate reconstructions of the metabolic pathways in poorly studied organisms (Koonin and Galperin, 2002; Osterman and Overbeek, 2003) and even prediction of their nutritional requirements (Lemos *et al.*, 2003). In a similar fashion, one could hope that some day it would be possible to reconstruct microbial signalling pathways and predict responses of a given microorganism to various environmental factors, based solely on its genome content. The first glimpses of such approaches are already evident from the Table 1. Indeed, two α-proteobacteria, *Mesorhizobium loti* and *Caulobacter crescentus*, encode the same number of histidine kinases but differ dramatically in the number of encoded adenylate cyclases and methyl-carrier proteins, suggesting the importance of chemotactic response for

the latter, but not the former, organism. Here I briefly discuss the recent insights into microbial signal transduction that originated from comparative genome analyses and list some unresolved problems.

Defining the parts set

The most important feature of signal transduction proteins is their modular organization, presciently noted by Parkinson and Kofoid (1992) 12 years ago. Modular organization accounts for enormous diversity of components of the bacterial signal transduction systems, but it also makes possible their systematic analysis. Most signal transduction proteins consist of two or more domains – evolutionarily conserved individually folding compact protein units that have more or less same functions regardless of genomic context (Table 2). If this definition sounds somewhat fuzzy, so are boundaries of many domains. Nevertheless, most domains can be relatively easily recognized and associated with particular biochemical functions. Thus, response regulators typically contain a phosphate-accepting receiver domain CheY – similar to the chemotaxis transducer protein of the same name, often referred to as 'chemotaxis response regulator' – and a DNA-binding signal output domain of the helix-turn-helix (HTH), winged helix, SAPR, LytTR, Fis, or some other family. This means that the CheY domain, like many other signalling domains, is 'promiscuous', i.e. can be found in a variety of distinct proteins associated with distinct signalling domains. In sequence similarity searches using BLAST, FASTA or other algorithms such promiscuous domains readily align with each other, which results in convincingly high similarity scores between otherwise unrelated proteins and significantly complicates sequence analysis (Fedorova *et al*., 2003). For example, in a BLASTP search, two different proteins sharing only a common CheY domain would nonetheless be aligned over 100–120 residues with reported probability of such a hit solely by chance of 10^{-4} or even lower. This makes splitting a multidomain signal transduction protein into individual domains the necessary first step in its sequence analysis. Often enough, it is also the easiest way to get an insight into potential functions of this protein.

The most straightforward way to delineate the domain composition of a signal transduction protein is to compare it to a protein domain database, such as Pfam (http://www.sanger.ac.uk/Software/Pfam), SMART (http://smart.embl-heidelberg.de), or COG (http://www.ncbi.nlm.nih.gov/COG) (see Bateman *et al*., 2004; Letunic *et al*., 2004; Tatusov *et al*., 2000 respectively). The European Bioinformatics Institute (EBI) and the National Center for Biotechnology Information (NCBI) maintain integrated domain databases, respectively, InterPro (http://www.ebi.ac.uk/interpro), which unifies, among others, SMART and Pfam entries (Mulder *et al*., 2003), and the Conserved Domains Database (CDD, http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml), which unifies SMART, Pfam and COG entries (Marchler-Bauer *et al*., 2003). Each of these databases provides a brief description of every domain and a variety of external links. SMART lets the user to list all the proteins with given domain organization (or domain composition). The Conserved Domains Database (CDD) offers a useful CDART tool that allows one to look for various domain combinations involving the given domain (Geer *et al*., 2002). Importantly, these databases rely on different software tools and different default parameters for domain identification (hidden Markov models in SMART, Pfam, and InterPro, COGnitor in COGs and Reverse Position-Specific BLAST in CDD), so the results obtained are not necessarily identical and it is always better to use more than one database (see Koonin and Galperin, 2002; for a discussion of the advantages of each particular database). For example, the recently characterized HWE family of histidine kinases (Karniol and Vierstra, 2004) had been properly annotated in COGs and CDD, but not in SMART or Pfam.

Although novel signalling domains continue to be discovered, it is safe to assume that domains with the widest phylogenetic distribution have already been identified through cross-genome

comparisons. It has become clear that certain sensory domains have very narrow specificity towards their ligands (e.g. citrate or nitrate binding domains, see below). Some domains, however, are rather indiscriminate in their affinities and interact with a wide variety of ligands. The PAS domain, for example, binds flat heterocyclic molecules from haeme to flavin to cinnamic acid and, reportedly, even the adenine moiety of ATP (see Table 2). The exact functions of many signalling domains remain obscure, which opens new avenues for future experimental studies.

Recently described signalling domains

For the purposes of this review, components of the signal transduction system can be subdivided into sensory (usually, ligand-binding) domains, signal transduction (phosphorylation, methylation, homodimerization) domains, and signal output (DNA-binding, heterodimerization or enzymatic) domains.

Sensory domains

Recent studies have greatly expanded the diversity of known sensory domains, adding many novel periplasmic (or extracytoplasmic) domains, as well as cytoplasmically located and integral membrane domains (see Table 3). Some recently described sensory domains have well-defined and narrow substrate specificity, e.g. nitrate-binding NIT domain (Shu *et al*., 2003) and citrate-binding CitAP domain (Gerharz *et al*., 2003;Pappalardo *et al*., 2003). For many other domains, the nature of the sensed signal(s) remains unknown and their roles as sensors are deduced solely from their predicted location as periplasmic domains of different transmembrane receptors: histidine kinases, methyl-accepting proteins, adenylate or diguanylate cyclases and phosphodiesterases.

Periplasmic solute-binding proteins have long been known to function as ligand-binding domains of sensor histidine kinases, for example, in the EvgS protein from *Escherichia coli*. Recently, however, a protein of the same family, *Pseudomonas aeruginosa* AmiC (PA33664), was found localized in the cytoplasm. AmiC serves as the receptor and negative regulator for amide-inducible aliphatic amidase operon *amiEBCRS*. Together with the RNA-binding response regulator AmiR, AmiC regulates expression of the AmiE amidase, as well as expression of its own gene, *amiC*, in response to amides (Wilson *et al*., 1996).

Another cytoplasmically located sensor domain is the N-terminal turgor-sensing domain of the K^+-transport regulatory kinase KdpD from *E. coli* and other bacteria. Recent data indicate that the transmembrane segments of this protein are needed only for proper positioning of the sensory domain with respect to the histidine kinase domain (Heermann *et al*., 2003). Remarkably, in *Bacillus cereus, Deinococcus radiodurans* and several other bacteria this sensory domain is encoded in a stand-alone form.

The most common cytoplasmic signalling domains are PAS and GAF domains (Aravind and Ponting, 1997; Ho *et al*., 2000; reviewed in Taylor and Zhulin, 1999; Galperin *et al*., 2001a; Hurley, 2003). Originally recognized as cytoplasmic domains of histidine kinases, these domains have now been found in combination with a great variety of other signalling domains (see Table 4). The PAS and GAF domains were shown to have similar structures, characterized by a presence of a ligand-binding pocket that can accommodate a variety of small-molecule ligands, from haeme to flavin to adenine and guanine (Ho *et al*., 2000; Crosson and Moffat, 2001). Presence of oxygen was shown to affect the position of the PAS-bound haeme molecule, causing a change in the general conformation of the PAS domain and thereby allowing the sensing of oxygen to effect signal transmission to the C-terminally located signal transduction domains (Delgado-Nixon *et al*., 2000; Chang *et al*., 2001). Due to the availability of several comprehensive reviews (Taylor and Zhulin, 1999; Zhulin, 2001; Hurley, 2003), these relatively

well-studied domains will not be discussed here in any detail. We will also leave aside small-molecule-binding domains (ACT, 4VR, 3H and others) that are involved in allosteric regulation of metabolic enzymes and modulation of the activity of transcriptional regulators (Anantharaman *et al.*, 2001).

Integral membrane domains can also serve as sensors. Although certain histidine kinases were long known to contain multiple transmembrane segments (Kadner, 1995), it was not clear whether these segments actually worked as sensors or just anchored the enzyme in the membrane. The first membrane domain with proven sensory function was the ethylene-binding domain of the *Arabidopsis thaliana* ETR1 protein (Bleecker, 1999). This domain was found later in cyanobacteria and several proteobacteria (Mount and Chang, 2002). Histidine kinases from *P. aeruginosa* (PA3271), *Vibrio cholerae* (VC0303), and several other bacteria contain proline permease-like N-terminal domains, which led to a suggestion that these proteins might serve as sensors of sodium-motive force (Häse *et al.*, 2001). The final evidence that integral membrane segments may represent evolutionary mobile conserved domains has come when several such domains (MHYT, MASE1 and MASE2) were found in association with two or more different signal output domains (see below), suggesting their involvement in signalling (Galperin *et al.*, 2001b; Nikolskaya *et al.*, 2003). Recently, four new families of membrane signalling domains with seven, seven, five and eight transmembrane segments, respectively, were identified (Anantharaman and Aravind, 2003). Although sequence conservation in these domains allows one to make educated guesses about their ligands (metal binding for MHYT and MHYE, aromatic compound binding for MASE1 and MASE2, carbohydrate binding for 7TMR-DISMED1), their exact nature is still obscure and needs to be explored in experimental studies.

Signal transduction domains

Environmental changes, sensed by the periplasmic (extracytoplasmic) or membrane-embedded sensory domains of transmembrane receptors, affect cytoplasmically located domains of these receptors to trigger appropriate cellular responses. These responses include increased gene expression, changes in motility (chemotaxis), changes in secretion and many other processes. Exactly how the signal is transmitted across the membrane from sensory to cytoplasmic domains is still not completely understood. Dimerization (oligomerization) events appear to be important in some cases. A dimerization domain, HAMP (Aravind and Ponting, 1999; Williams and Stewart, 1999), has been found in many transmembrane receptors, but certainly not in all of them (Zhulin *et al.*, 2003). Several potential mechanisms of transmembrane signalling are currently being considered and breakthroughs in this area are expected in the near future (see Falke and Hazelbauer, 2001).

An important recent development coming from comparative genomics was the realization that histidine kinases and methyl-accepting proteins are the major but by no means the only receptor molecules capable of sensing extracellular signals. Fusions of periplasmic sensory domains to adenylate cyclase, serine/threonine protein kinase, diguanylate cyclase and phosphodiesterase domains have been described, reinforcing the notion that bacterial signalling is even more complex than previously thought. The first of these, the receptor adenylate cyclase, was originally recognized in cyanobacteria *Spirulina platense* and *Anabaena* sp. PCC7120 as a fusion of type 3 adenylate cyclase domain to a periplasmic sensor domain and experimentally verified to have adenylate cyclase activity (Yashiro *et al.*, 1996; Katayama and Ohmori, 1997). Sensory adenylate cyclases were soon described in a number of other organisms, including proteobacteria *Stigmatella aurantiaca* and *Myxococcus xanthus* (Coudart-Cavalli *et al.*, 1997; Kimura *et al.*, 2002).

Serine/threonine protein kinases and phosphatases have been known in prokaryotes for quite some time (Yang *et al.*, 1996; Shi *et al.*, 1998), but most of them either were soluble enzymes,

or contained cytoplasmic N-terminal kinase domains and were anchored in the membrane by their C-terminal fragments. Recently, however, serine/threonine protein kinase domains were found in membrane receptor proteins, fused to N-terminal periplasmic sensory domains (Zhulin *et al*., 2003; see Table 2). In addition, transmembrane receptors were found whose cytoplasmic domain was similar to RsbU, a protein phosphatase of the PP2C family that is involved in an environmental stress signalling pathway (Yang *et al*., 1996; Zhulin *et al*., 2003). Unfortunately, the signals sensed by these proteins are still obscure, as are their phosphorylation/dephosphorylation targets. Because of their apparent similarity to the eukaryotic signalling systems, bacterial protein kinases and phosphatases are attractive targets for further experimental studies of the signal transduction mechanisms.

The list of bacterial membrane receptors also includes a group of proteins that combine periplasmic sensory domains with the cytoplasmic GGDEF, EAL and HD-GYP domains, whose enzymatic functions are still somewhat uncertain (reviewed in Galperin *et al*., 2001a). The GGDEF domain is often paired with the EAL domain, forming a diguanylate cyclase/phosphodiesterase combination that catalyses synthesis and hydrolysis of cyclic diguanylate (c-diGMP, Fig. 1). In *Gluconobacter xylinum*, c-diGMP regulates formation of extracellular cellulose (Tal *et al*., 1998). It has recently been implicated in regulation of extracellular polysaccharide formation in a number of other bacteria (Rashid *et al*., 2003). In *Caulobacter crescentus,* a GGDEF domain-containing response regulator PleD is involved in regulation of the cell development programme, offering a convenient model to study its potential functions (Ausmees *et al*., 2001; Aldridge *et al*., 2003). Studies of this protein strongly suggested that the GGDEF domain acts as a diguanylate cyclase that combines two GTP molecules to form c-diGMP (Ausmees *et al*., 2001; Pei and Grishin, 2001). Although an unequivocal biochemical proof that the purified GGDEF domain indeed carries this activity has not been published so far, preliminary data indicate that this is in fact the case (M. Gomelsky, pers. comm.). The EAL domain must therefore be responsible for the complementary phosphodiesterase activity that degrades c-diGMP, either by itself or in combination with the GGDEF domain. Judging from its sequence motifs, HD-GYP is also a phosphodiesterase domain (Galperin *et al*., 1999). Its natural phosphoester substrate remains unidentified; it too could be c-diGMP. The biochemical characterization of these domains and processes that they regulate is still in the very early stages. In any case, the sheer abundance of genes encoding GGDEF, EAL, and HD-GYP domains in diverse bacterial genomes (Table 2) shows that they represent a major signalling system with c-diGMP most likely functioning as a secondary messenger in signal transduction. The potential importance of this novel signalling pathway came to light only thanks to the availability of complete genome sequences.

Signal transduction downstream of sensory histidine kinases and methyl-accepting proteins often involves intermediate domains, including HPt and CheY domains, which are relatively well characterized and described in detail elsewhere (Stock *et al*., 2000; Inouye and Dutta, 2003). In contrast, signal transduction downstream of receptor adenylate cyclases does not involve any proteins or domains beside the cAMP receptor protein (CAP). The cAMP-CAP complex has been shown to activate transcription of many genes in diverse bacterial species. Although the mechanisms of c-diGMP action are still obscure, signal transduction from transmembrane receptors containing GGDEF, EAL or HD-GYP signalling domains likewise do not seem to involve any intermediate domains, at least judging from the domain numerology (Table 2). In the only experimentally characterized model, activation of the *G. xylinum* cellulose synthase by c-diGMP was mediated by a membrane-bound c-diGMP-binding protein (Weinhouse *et al*., 1997).

In a sense, adenylate cyclase, diguanylate cyclase and phosphodiesterase can be considered output domains, so that input and output modules of these receptors co-exist on the same polypeptide chain. In an even further deviation from the 'two-component' paradigm, certain

transmembrane sensors were found to contain C-terminal DNA-binding domains (Nikolskaya and Galperin, 2002), following the classical example of the lysine-sensing transcriptional regulator CadC (Dell *et al.*, 1994). These extreme cases clearly show that there are no strict limits on the number of components in the signal transduction chain, which may vary from one to three or more. It should also be noted that certain signalling systems include a stand-alone sensor protein that interacts with a histidine kinase (Kadner, 1995) while other transmit the signal directly to the transcription regulation apparatus (Braun, 1997).

Output domains

Response regulators of the two-component system typically consist of an N-terminal phosphoacceptor CheY domain and a C-terminal DNA-binding output domain that activates or represses transcription of specific target genes (Martinez-Hackert and Stock, 1997; Grebe and Stock, 1999; Stock *et al.*, 2000). These DNA-binding domains are quite diverse: a majority belongs to the winged helix family, exemplified by the well-known OmpR and PhoB proteins, but there are several families of helix-turn-helix (HTH) domains, such as NarL/FixJ, AraC/XylS and Spo0A domains. In addition, certain response regulators contain non-HTH DNA-binding domains of SAPR, LytTR or Fis families. Although the operons that these response regulators activate or repress are often unknown, SAPR family proteins are typically involved in the regulation of secondary metabolism (Wietzorrek and Bibb, 1997), whereas LytTR family proteins often regulate production of virulence factors (Nikolskaya and Galperin, 2002). In transcriptional regulators of the NtrC family, the N-terminal CheY domains and the C-terminal DNA-binding Fis-like domains are separated by the central AAA-type ATP-binding domains, whose ATPase activity is required for the DNA-binding (Hwang *et al.*, 1999).

In addition to the DNA-binding response regulators, an RNA-binding output domain has been described in *Pseudomonas aeruginosa* response regulator AmiR and related proteins (Shu and Zhulin, 2002).

In certain response regulators, the output domains are enzymatic and do not necessarily regulate transcription. Such response regulators combine the CheY domain with CheB-type methylesterase domain, GGDEF, EAL, HD-GYP or PP2C domains, mentioned above, or with other, sometimes unknown, enzymatic domains.

Intracellular signalling network

In addition to transmembrane receptors, there are several well-studied histidine kinases that have no transmembrane segments, such as chemotaxis histidine kinase CheA and nitrogen regulation protein NtrB (GlnL) from *E. coli*, sporulation kinase KinA from *B. subtilis,* or rhizobial oxygen sensor FixL (see Hoch and Silhavy, 1995; for reviews). Receptor census (Table 1) shows that free-living bacteria typically encode a significant number of intracellular histidine kinases, adenylate cyclases, diguanylate cyclases and phosphodiesterases. In fact, their cytoplasmic signalling network may be as complex as transmembrane signal transduction system. The genome of *M. loti*, for example, encodes 13 copies of the adenylate cyclase domain (Table 1). Of these, only one appears to be fused to a periplasmic sensor domain, and another one is fused to an integral membrane sensor domain. All the rest are found in predicted cytoplasmic proteins, fused to poorly characterized N-terminal or C-terminal domains, most of which are likely involved in signalling. Of the 32 copies of the GGDEF domain, encoded in *M. loti*, 18 belong to transmembrane sensors and 14 are found in intracellular signal transduction proteins and response regulators (Table 1).

Intracellular signalling proteins typically combine N-terminal cytoplasmic sensor domains, usually PAS or GAF, with a variety of signal transduction or output domains (Table 4). Some of these proteins contain N-terminal CheY domains and can be considered *bona fide* response

regulators. Indeed, phosphorylation of the CheY domain was shown to affect adenylate cyclase activity of the C-terminal ACyc domain, just as it affects DNA-binding properties of classical response regulators (Coudart-Cavalli *et al*., 1997). However, many intracellular signalling proteins lack the CheY domains. Such proteins should not be confused with response regulators, despite certain parallelism in their domain architectures (see Fig. 2 and Table 4). For example, in addition to four NtrC-type response regulators of the CheY-AAA-Fis domain architecture (AtoC, GlnG, HydG and YfhA, see COG2204), *E. coli* K12 encodes three intracellular signalling proteins with GAF-AAA-Fis domain structure (FhlA, HyfA, and NorR, see COG3604) and one more protein (YgeV) with GAF-PAS-AAA-Fis domain structure. Whereas the exact nature of the ligands of most of these proteins remains obscure, there is little doubt that they are directly involved in monitoring levels of NO and other intracellular parameters and regulating transcription in response to changes in these parameters (Gardner *et al*., 2003).

Several pioneering studies have provided experimental evidence of the involvement of cytoplasmic signalling proteins in intracellular signalling. An *E. coli* protein with the PAS-GGDEF-EAL domain combination has been named a 'direct oxygen sensor' (DOS), based on the effect oxygen binding has on the conformation of its N-terminal domain (Delgado-Nixon *et al*., 2000). Further, oxygen binding has been shown to activate the phosphodiesterase activity of a *G. xylinum* protein with the same domain organization (Chang *et al*., 2001). Likewise, cGMP binding to the GAF domain of human phosphodiesterase PDE5 was shown to stimulate the activity of its C-terminal enzymatic domain (Rybalkin *et al*., 2003). Besides oxygen, the DOS protein could also bind NO and CO, indicating that PAS- or GAF-containing molecules could be used for sensing a variety of intracellular parameters and effecting a variety of cellular responses. Finally, the NorR protein of GAF-AAA-Fis domain architecture has been shown to regulate transcription in response to nitric oxide and reactive nitrogen species (Pohlmann *et al*., 2000; Gardner *et al*., 2003; Mukhopadhyay *et al*., 2004). These results clearly demonstrate that ligand binding to the N-terminal PAS and/or GAF domains can modulate the activities of the downstream output domains. Thus, the similarity between CheY-containing response regulators and PAS- or GAF-containing signallers apparently extends to their regulation mechanisms: both phosphorylation of the CheY domain in response to the extracellular signal and ligand binding to PAS or GAF, comprising an intracellular signal, induce conformational changes in these domains. In turn, these conformational changes activate (rather, cause a relief of inhibition) the downstream output domains, allowing them to perform their functions, be that binding DNA or RNA, catalysing synthesis or hydrolysis of cAMP or c-diGMP, demethylation of MCPs, and so on.

Unfortunately, most intracellular signalling proteins are still poorly studied and remain to be recognized as legitimate members of the bacterial signalling network.

Cross-talk between different signalling systems

In the extreme diversity of signalling domain combinations encoded in different microbial genomes (see Table 4 for examples), several domain architectures stand out, clearly demonstrating the possibility of cross-talk between different signalling systems. It is well known that many histidine kinases and at least some methyl-carrier proteins contain PAS and GAF domains, which appear to modulate the activity of these proteins (Aravind and Ponting, 1997;Taylor and Zhulin, 1999). Likewise, the existence of PAS-PP2C, GAF-PP2C and PAS-GAF-PP2C combinations in many bacteria, particularly Gram-positive bacteria and actinobacteria, indicates a link between energy stress and σ^B-dependent transcription (Vijay *et al*., 2000). The existence of the GAF-PtsI combination in many proteobacteria provides a way for GAF-sensed signals to affect PTS-dependent processes of catabolite repression and inducer exclusion. Because the GAF domain can bind cAMP, this might be a feedback

mechanism of maintaining cAMP levels. Fusions of the phosphothreonine-binding FHA (forkhead-associated) domain with ACyc and GGDEF domains, found in several cyanobacteria, suggest that protein phosphorylation could affect the activities of the respective cyclases. Finally, two recently sequenced genomes, *Thermosynechococcus elongatus* and *Pirellula* sp., encode a FHA-GAF-HisKin domain combination that ties together three different signalling mechanisms. An even more vivid demonstration of the principle that any two signalling pathways can affect each other is the coexistence of the Ser, Thr-kinase and HisKin domains on the same polypeptide chain in several proteins from *Anabaena* sp. (Ohmori *et al*., 2001).

'Bacterial' signal transduction in eukaryotic cells and vice versa

For many years, two-component systems were believed to be specifically bacterial, whereas serine-, threonine- and tyrosine-dependent kinases were seen exclusively in eukaryotes. Sequencing of complex bacterial genomes, as well as genomes of plants, animals and lower eukaryotes, revealed the presence of unexpected signalling domains in many of them. How exactly these systems appeared in these genomes is still a matter of controversy. Some of them most likely have come from a common ancestor or were appropriated by first eukaryotes from their pro-mitochondrial or pro-chloroplast symbionts. In other cases, a relatively recent horizontal gene transfer seems to be a plausible explanation. Anyway, one should not be shocked to find in a free-living bacterium a signalling domain seen previously only in eukaryotes. Conversely, many eukaryotic signalling domains appear to have roots in bacteria (Ponting *et al*., 1999; Koretke *et al*., 2000; Aravind *et al*., 2003). This is an unexpected but promising development, as data on eukaryotic signal transduction could help in deciphering the functions of bacterial proteins and vice versa. Of course, one has to be cautious, as, for example, most, if not all, the genes encoding the GGDEF, EAL and HisKA domains in the mosquito *Anopheles gambiae* genome (according to the SMART database, nine, six and 10 copies respectively) probably have come from bacterial contamination.

Functional annotation of signal transduction proteins

Studies of domain architectures of metabolic enzymes revealed a very limited number of possible domain architectures, making the annotation relatively straightforward even in cases like human *CAD* protein (PYR1_HUMAN, P27708) that consists of three different domains. In signalling systems, however, associations of various sensory, signal transduction and output domains produce almost an infinite number of domain combinations with a three-domain protein looking fairly mundane. The complexity of domain organization makes correct functional annotation of signal transduction proteins anything but trivial (Fedorova *et al*., 2003). Annotation is further complicated by the fact that many conserved domains have poorly understood or unknown enzymatic activities and/or binding specificities. Even the best domain analysis tools, employed in domain databases like SMART, Pfam and CDD, are not designed to provide correct annotation for the whole protein. The best they can do is to (i) uncover the domain composition of the given protein and (ii) show the annotations of other proteins with the same domain composition, if available (Geer *et al*., 2002). In many cases these annotations are inconsistent; at least some of them are likely to be wrong. We have noted, for example, that three virtually identical proteins from *Anabaena* sp. PCC7120 were originally annotated, respectively, as 'adenylate cyclase' (All7310, TrEMBL entry Q8YKI7), 'similar to adenylate cyclase' (All3180, Q8YSA9) and a 'hypothetical protein' (Alr1378, Q8YX39), although none of them actually contained the adenylate cyclase domain (Zhulin *et al*., 2003). In fact, annotation of signal transduction proteins as 'unknown', 'hypothetical', or 'conserved hypothetical' proteins is quite common and is generally considered to be appropriate. As we have argued earlier, short of a systematic mistake in sequencing, a protein that is conserved across diverse phylogenetic lineages should not be considered hypothetical (Galperin, 2001).

So what could be an acceptable annotation of a novel signal transduction protein in a newly sequenced genome? First of all, it would be helpful to include the word 'signalling', as in 'signalling protein'. If domain analysis shows the presence of a well-characterized enzymatic domain, the protein should be annotated based on its enzymatic activity, as a histidine kinase, adenylate cyclase, and so on. Presence of a periplasmic or membrane-bound sensory domain should also be reflected in the name, making it, for example, 'sensory Ser/Thr-protein kinase'. Of course, if the ligand specificity of the sensory domain is known, that, too, should be reflected in the name, for example, in 'osmosensory cAMP phosphodiesterase', or 'pH-sensing histidine kinase'. Finally, if one can say nothing besides the domain composition, annotation of the novel protein should probably look as follows: 'Predicted signal transduction protein, containing PAS, GAF and HD-GYP domains'. This would still be much better than 'similar to hypothetical protein'.

Concluding remarks: some unresolved problems

Despite significant progress made in the last several years, we are still far from understanding many key aspects of bacterial signal transduction. First of all, it is often not clear which particular parameters are measured by many sensor domains (Table 3). Osmolarity of the medium, for example, affects the intracellular K^+ concentration, the ionic strength in the cytoplasm, water content of the cell and a number of other physicochemical parameters. The effects of changes in the external pH values or temperature are equally dramatic. Therefore, even for the relatively well characterized pH- and osmosensors, the exact nature of the signal often remains elusive (see Heermann and Jung, 2004; for a recent review).

Second, the list of signalling domains is probably far from complete and new domains of poorly defined function are still being described. For example, a predicted hydrolase of HD superfamily (COG1639), which is found mostly in stand-alone form, serves as the output domain of the *P. aeruginosa* response regulator PA0267 and several related proteins. However, it is not clear what is the substrate of this predicted hydrolase and what (if any) is the function of its inactivated variant in the *V. cholerae* response regulator VC1081.

Third, there are many domains that are likely involved in signalling but whose functions are still enigmatic. One of the most conspicuous examples is the tryptophan-rich sensory protein TspO/CrtK/MBR, an integral membrane protein found in representatives of all domains of life, from archaea to human (PF03073, COG3476). This protein, often referred to as peripheral-type mitochondrial benzodiazepine receptor, contains five predicted transmembrane segments with 12–14 well conserved aromatic amino acid residues, including seven Trp residues (Yeliseev and Kaplan, 2000). It has been shown to regulate photosynthesis gene expression in *Rhodobacter sphaeroides*, nutrient stress in *Sinorhizobium meliloti*, and to bind various benzodiazepins, tetrapyrrols and steroids, including cholesterol, protoporphyrin IX, and many others (Gavish *et al.*, 1999; Davey and de Bruijn, 2000; Yeliseev and Kaplan, 2000; Lacapere and Papadopoulos, 2003). None of these functions, however, readily explains the role of this domain in cells of *B. subtilis* or *Archaeoglobus fulgidus*, which do not carry out photosynthesis and have no known affinity to benzodiazepines or steroids.

Another important but still uncharacterized signalling domain is PfoR, a predicted membrane protein found in many bacteria and distantly related to the membrane components of fructose- and sucrose-specific PTS (EIIC[Fru], see COG1299). In *Clostridium* spp., *pfoR* genes are located upstream of the genes encoding thiol-activated cytotoxins perfringolysin O, tetanolysin O and septicolysin, but PfoR does not seem to regulate toxin expression (Awad and Rood, 2002). In *Streptococcus pyogenes*, however, the *pfoR*-like *sloR* gene was shown to affect streptolysin O expression, despite the fact that these genes are not adjacent in the genome (Savic *et al.*,

2002). These examples clearly demonstrate that comparative sequence analysis is but a first step, rather than a panacea, and has to rely upon and be followed up by experimental studies.

Besides delineating the domain 'parts set', understanding of the signal transduction will require answering many critical questions. The following is a selection of questions that I consider most interesting and experimentally tractable within the next several years:

> What are the exact biochemical activities of the GGDEF and EAL domains?
>
> What are the principal targets and mechanisms of the c-diGMP action?
>
> What is the function of the C-terminal CheY domains in hybrid histidine kinases?
>
> Are the effects of tandem sensory domains (e.g. PAS, GAF) in a single protein additive or hierarchical (or both)?
>
> What is the extent of crosstalk between different histidine kinases and their response regulators *in vivo*?
>
> Is there an order or hierarchy in signal transduction from different membrane receptors sharing the same sensory domain?

Although this selection primarily reflects personal bias, I strongly believe that the time has come when these and other critical questions about bacterial signalling can be finally addressed. This would require combining a variety of experimental and computational approaches, of which genome analysis will be a significant part.

References

Aldridge P, Paul R, Goymer P, Rainey P, Jenal U. Role of the GGDEF regulator PleD in polar development of *Caulobacter crescentus.* Mol Microbiol 2003;47:1695–1708. [PubMed: 12622822]

Anantharaman V, Aravind L. Cache – a signalling domain common to animal Ca^{2+}-channel subunits and a class of prokaryotic chemotaxis receptors. Trends Biochem Sci 2000;25:535–537. [PubMed: 11084361]

Anantharaman V, Aravind L. The CHASE domain: a predicted ligand-binding module in plant cytokinin receptors and other eukaryotic and bacterial receptors. Trends Biochem Sci 2001;26:579–582. [PubMed: 11590000]

Anantharaman V, Aravind L. Application of comparative genomics in the identification and analysis of novel families of membrane-associated receptors in bacteria. BMC Genomics 2003;4:34. [PubMed: 12914674]

Anantharaman V, Koonin EV, Aravind L. Regulatory potential, phyletic distribution and evolution of ancient, intracellular small-molecule-binding domains. J Mol Biol 2001;307:1271–1292. [PubMed: 11292341]

Aravind L, Ponting CP. The GAF domain: an evolutionary link between diverse phototransducing proteins. Trends Biochem Sci 1997;22:458–459. [PubMed: 9433123]

Aravind L, Ponting CP. The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins. FEMS Microbiol Lett 1999;176:111–116. [PubMed: 10418137]

Aravind L, Anantharaman V, Iyer LM. Evolutionary connections between bacterial and eukaryotic signalling systems: a genomic perspective. Curr Opin Microbiol 2003;6:490–497. [PubMed: 14572542]

Ashby MK. Survey of the number of two-component response regulator genes in the complete and annotated genome sequences of prokaryotes. FEMS Microbiol Lett 2004;231:277–281. [PubMed: 14987775]

Ausmees N, Mayer R, Weinhouse H, Volman G, Amikam D, Benziman M, Lindberg M. Genetic data indicate that proteins containing the GGDEF domain possess diguanylate cyclase activity. FEMS Microbiol Lett 2001;204:163–167. [PubMed: 11682196]

Awad MM, Rood JI. Perfringolysin O expression in *Clostridium perfringens* is independent of the upstream *pfoR* gene. J Bacteriol 2002;184:2034–2038. [PubMed: 11889112]

Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, et al. The Pfam protein families database. Nucleic Acids Res 2004;32:D138–D141. [PubMed: 14681378]

Bleecker AB. Ethylene perception and signalling: an evolutionary perspective. Trends Plant Sci 1999;4:269–274. [PubMed: 10407443]

Braun V. Surface signalling: novel transcription initiation mechanism starting from the cell surface. Arch Microbiol 1997;167:325–331. [PubMed: 9148773]

Chang AL, Tuckerman JR, Gonzalez G, Mayer R, Weinhouse H, Volman G, et al. Phosphodiesterase A1, a regulator of cellulose synthesis in *Acetobacter xylinum*, is a heme-based sensor. Biochemistry 2001;40:3420–3426. [PubMed: 11297407]

Coudart-Cavalli MP, Sismeiro O, Danchin A. Bifunctional structure of two adenylyl cyclases from the myxobacterium *Stigmatella aurantiaca*. Biochimie 1997;79:757–767. [PubMed: 9523018]

Crosson S, Moffat K. Structure of a flavin-binding plant photoreceptor domain: insights into light-mediated signal transduction. Proc Natl Acad Sci USA 2001;98:2995–3000. [PubMed: 11248020]

Davey ME, de Bruijn FJ. A homologue of the tryptophan-rich sensory protein TspO and FixL regulate a novel nutrient deprivation-induced *Sinorhizobium meliloti* locus. Appl Environ Microbiol 2000;66:5353–5359. [PubMed: 11097914]

Davis SJ, Vener AV, Vierstra RD. Bacterio-phytochromes: phytochrome-like photoreceptors from non-photosynthetic eubacteria. Science 1999;286:2517–2520. [PubMed: 10617469]

Delgado-Nixon VM, Gonzalez G, Gilles-Gonzalez MA. Dos, a heme-binding PAS protein from *Escherichia coli*, is a direct oxygen sensor. Biochemistry 2000;39:2685–2691. [PubMed: 10704219]

Dell CL, Neely MN, Olson ER. Altered pH and lysine signalling mutants of *cadC*, a gene encoding a membrane-bound transcriptional activator of the *Escherichia coli cadBA* operon. Mol Microbiol 1994;14:7–16. [PubMed: 7830562]

Falke JJ, Hazelbauer GL. Transmembrane signalling in bacterial chemoreceptors. Trends Biochem Sci 2001;26:257–265. [PubMed: 11295559]

Falke JJ, Bass RB, Butler SL, Chervitz SA, Danielson MA. The two-component signalling pathway of bacterial chemotaxis: a molecular view of signal transduction by receptors, kinases, and adaptation enzymes. Annu Rev Cell Dev Biol 1997;13:457–512. [PubMed: 9442881]

Fedorova, N.D., Nikolskaya, A.N., and Galperin, M.Y. (2003) Using genomic context in the analysis of multi-component systems. In *Frontiers in Computational Genomics* Galperin, M.Y., and Koonin, E.V. (eds). Wymondham, UK: Caister Academic Press, pp. 167–194.

Galperin MY. Conserved 'hypothetical' proteins: new hints and new puzzles. Comp Funct Genomics 2001;2:14–18.

Galperin MY, Natale DA, Aravind L, Koonin EV. A specialized version of the HD hydrolase domain implicated in signal transduction. J Mol Microbiol Biotechnol 1999;1:303–305. [PubMed: 10943560]

Galperin MY, Nikolskaya AN, Koonin EV. Novel domains of the prokaryotic two-component signal transduction systems. FEMS Microbiol Lett 2001a;203:11–21. [PubMed: 11557134]

Galperin MY, Gaidenko TA, Mulkidjanian AY, Nakano M, Price CW. MHYT, a new integral membrane sensor domain. FEMS Microbiol Lett 2001b;205:17–23. [PubMed: 11728710]

Gardner AM, Gessner CR, Gardner PR. Regulation of the nitric oxide reduction operon (*norRVW*) in *Escherichia coli*. Role of NorR and sigma54 in the nitric oxide stress response. J Biol Chem 2003;278:10081–10086. [PubMed: 12529359]

Gavish M, Bachman I, Shoukrun R, Katz Y, Veenman L, Weisinger G, Weizman A. Enigma of the peripheral benzodiazepine receptor. Pharmacol Rev 1999;51:629–650. [PubMed: 10581326]

Geer LY, Domrachev M, Lipman DJ, Bryant SH. CDART: protein homology by domain architecture. Genome Res 2002;12:1619–1623. [PubMed: 12368255]

Gerharz T, Reinelt S, Kaspar S, Scapozza L, Bott M. Identification of basic amino acid residues important for citrate binding by the periplasmic receptor domain of the sensor kinase CitA. Biochemistry 2003;42:5917–5924. [PubMed: 12741850]

Gomelsky M, Klug G. BLUF: a novel FAD-binding domain involved in sensory transduction in microorganisms. Trends Biochem Sci 2002;27:497–500. [PubMed: 12368079]

Grebe TW, Stock JB. The histidine protein kinase superfamily. Adv Microb Physiol 1999;41:139–227. [PubMed: 10500846]

Häse CC, Fedorova ND, Galperin MY, Dibrov PA. Sodium ion cycle in bacterial pathogens: evidence from cross-genome comparisons. Microbiol Mol Biol Rev 2001;65:353–370. [PubMed: 11528000]

Heermann R, Jung K. Structural features and mechanisms for sensing high osmolarity in microorganisms. Curr Opin Microbiol 2004;7:168–174. [PubMed: 15063855]

Heermann R, Altendorf K, Jung K. The hydrophilic N-terminal domain complements the membrane-anchored C-terminal domain of the sensor kinase KdpD of *Escherichia coli.* J Biol Chem 2000;275:17080–17085. [PubMed: 10747873]

Heermann R, Fohrmann A, Altendorf K, Jung K. The transmembrane domains of the sensor kinase KdpD of *Escherichia coli* are not essential for sensing K^+ limitation. Mol Microbiol 2003;47:839–848. [PubMed: 12535080]

Henikoff S. Playing with blocks: some pitfalls of forcing multiple alignments. New Biol 1991;3:1148–1154. [PubMed: 1669287]

Ho YS, Burden LM, Hurley JH. Structure of the GAF domain, a ubiquitous signalling motif and a new class of cyclic GMP receptor. EMBO J 2000;19:5288–5299. [PubMed: 11032796]

Hoch, J.A., and Silhavy, T.J., (eds). (1995) *Two-Component Signal Transduction* Washington, DC: American Society for Microbiology Press.

Hurley, J.H. (2003) GAF domains: cyclic nucleotides come full circle. *Science STKE* 2003: PE1.

Hwang I, Thorgeirsson T, Lee J, Kustu S, Shin YK. Physical evidence for a phosphorylation-dependent conformational change in the enhancer-binding protein NtrC. Proc Natl Acad Sci USA 1999;96:4880–4885. [PubMed: 10220387]

Inouye, M., and Dutta, R., (eds). (2003) *Histidine Kinases in Signal Transduction* New York: Academic Press.

Iyer LM, Anantharaman V, Aravind L. Ancient conserved domains shared by animal soluble guanylyl cyclases and bacterial signalling proteins. BMC Genomics 2003;4:5. [PubMed: 12590654]

Kadner, R.J. (1995) Expression of the Uhp sugar-phosphate transport system of *Escherichia coli* In *Two-Component Signal Transduction* Hoch, J.A, and Silhavy, T.J. (eds). Washington, DC: American Society for Microbiology Press, p. 488.

Karniol B, Vierstra RD. The HWE histidine kinases, a new family of bacterial two-component sensor kinases with potentially diverse roles in environmental signalling. J Bacteriol 2004;186:445–453. [PubMed: 14702314]

Katayama M, Ohmori M. Isolation and characterization of multiple adenylate cyclase genes from the cyanobacterium *Anabaena* sp. strain PCC 7120. J Bacteriol 1997;179:3588–3593. [PubMed: 9171404]

Kennelly PJ. Protein kinases and protein phosphatases in prokaryotes: a genomic perspective. FEMS Microbiol Lett 2002;206:1–8. [PubMed: 11786249]

Kimura Y, Mishima Y, Nakano H, Takegawa K. An adenylyl cyclase, CyaA, of *Myxococcus xanthus* functions in signal transduction during osmotic stress. J Bacteriol 2002;184:3578–3585. [PubMed: 12057952]

Koonin EV, Galperin MY. Prokaryotic genomes: the emerging paradigm of genome-based microbiology. Curr Opin Genet Dev 1997;7:757–763. [PubMed: 9468784]

Koonin, E.V., and Galperin, M.Y. (2002) *Sequence – Evolution – Function. Computational Approaches in Comparative Genomics* Boston: Kluwer Academic Publishers.

Koretke KK, Lupas AN, Warren PV, Rosenberg M, Brown JR. Evolution of two-component signal transduction. Mol Biol Evol 2000;17:1956–1970. [PubMed: 11110912]

Lacapere JJ, Papadopoulos V. Peripheral-type benzodiazepine receptor: structure and function of a cholesterol-binding protein in steroid and bile acid biosynthesis. Steroids 2003;68:569–585. [PubMed: 12957662]

Lemos EG, Alves LM, Campanharo JC. Genomics-based design of defined growth media for the plant pathogen *Xylella fastidiosa.* FEMS Microbiol Lett 2003;219:39–45. [PubMed: 12594021]

Letunic I, Copley RR, Schmidt S, Ciccarelli FD, Doerks T, Schultz J, et al. SMART 4.0: towards genomic data integration. Nucleic Acids Res 2004;32:D142–D144. [PubMed: 14681379]

Marchler-Bauer A, Anderson JB, DeWeese-Scott C, Fedorova ND, Geer LY, He S, et al. CDD: a curated Entrez database of conserved domain alignments. Nucleic Acids Res 2003;31:383–387. [PubMed: 12520028]

Martinez-Hackert E, Stock AM. Structural relationships in the OmpR family of winged-helix transcription factors. J Mol Biol 1997;269:301–312. [PubMed: 9199401]

Matsushika A, Mizuno T. The structure and function of the histidine-containing phosphotransfer (HPt) signalling domain of the *Escherichia coli* ArcB sensor. J Biochem 1998;124:440–445. [PubMed: 9685739]

Mizuno T. Compilation of all genes encoding two-component phosphotransfer signal transducers in the genome of *Escherichia coli.* DNA Res 1997;4:161–168. [PubMed: 9205844]

Mizuno T, Kaneko T, Tabata S. Compilation of all genes encoding bacterial two-component signal transducers in the genome of the cyanobacterium, *Synechocystis* sp. strain PCC 6803. DNA Res 1996;3:407–414. [PubMed: 9097043]

Mougel C, Zhulin IB. CHASE: an extracellular sensing domain common to transmembrane receptors from prokaryotes, lower eukaryotes and plants. Trends Biochem Sci 2001;26:582–584. [PubMed: 11590001]

Mount SM, Chang C. Evidence for a plastid origin of plant ethylene receptor genes. Plant Physiol 2002;130:10–14. [PubMed: 12226482]

Mukhopadhyay P, Zheng M, Bedzyk LA, LaRossa RA, Storz G. Prominent roles of the NorR and Fur regulators in the *Escherichia coli* transcriptional response to reactive nitrogen species. Proc Natl Acad Sci USA 2004;101:745–750. [PubMed: 14718666]

Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Barrell D, Bateman A, et al. The InterPro Database, 2003 brings increased coverage and new features. Nucleic Acids Res 2003;31:315–318. [PubMed: 12520011]

Nikolskaya AN, Galperin MY. A novel type of conserved DNA-binding domain in the transcriptional regulators of the AlgR/AgrA/LytR family. Nucleic Acids Res 2002;30:2453–2459. [PubMed: 12034833]

Nikolskaya AN, Mulkidjanian AY, Beech IB, Galperin MY. MASE1 and MASE2: two novel integral membrane sensory domains. J Mol Microbiol Biotechnol 2003;5:11–16. [PubMed: 12673057]

Ninfa AJ, Magasanik B. Covalent modification of the *glnG* product, NRI, by the *glnL* product, NRII, regulates the transcription of the *glnALG* operon in *Escherichia coli.* Proc Natl Acad Sci USA 1986;83:5909–5913. [PubMed: 2874557]

Ninfa AJ, Ninfa EG, Lupas AN, Stock A, Magasanik B, Stock J. Crosstalk between bacterial chemotaxis signal transduction proteins and regulators of transcription of the Ntr regulon: evidence that nitrogen assimilation and chemotaxis are controlled by a common phosphotransfer mechanism. Proc Natl Acad Sci USA 1988;85:5492–5496. [PubMed: 3041412]

Ohmori M, Ikeuchi M, Sato N, Wolk P, Kaneko T, Ogawa T, et al. Characterization of genes encoding multi-domain proteins in the genome of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. DNA Res 2001;8:271–284. [PubMed: 11858227]

Osterman A, Overbeek R. Missing genes in metabolic pathways: a comparative genomics approach. Curr Opin Chem Biol 2003;7:238–251. [PubMed: 12714058]

Pappalardo L, Janausch IG, Vijayan V, Zientz E, Junker J, Peti W, et al. The NMR structure of the sensory domain of the membranous two-component fumarate sensor (histidine protein kinase) DcuS of *Escherichia coli.* J Biol Chem 2003;278:39185–39188. [PubMed: 12907689]

Parkinson JS, Kofoid EC. Communication modules in bacterial signalling proteins. Annu Rev Genet 1992;26:71–112. [PubMed: 1482126]

Pei J, Grishin NV. GGDEF domain is homologous to adenylyl cyclase. Proteins 2001;42:210–216. [PubMed: 11119645]

Pohlmann A, Cramm R, Schmelz K, Friedrich B. A novel NO-responding regulator controls the reduction of nitric oxide in *Ralstonia eutropha.* Mol Microbiol 2000;38:626–638. [PubMed: 11069685]

Ponting CP, Aravind L, Schultz J, Bork P, Koonin EV. Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer. J Mol Biol 1999;289:729–745. [PubMed: 10369758]

Rashid MH, Rajanna C, Ali A, Karaolis DK. Identification of genes involved in the switch between the smooth and rugose phenotypes of *Vibrio cholerae.* FEMS Microbiol Lett 2003;227:113–119. [PubMed: 14568156]

Rybalkin SD, Rybalkina IG, Shimizu-Albergine M, Tang XB, Beavo JA. PDE5 is converted to an activated state upon cGMP binding to the GAF A domain. EMBO J 2003;22:469–478. [PubMed: 12554648]

Savic DJ, McShan WM, Ferretti JJ. Autonomous expression of the *slo* gene of the bicistronic *nga-slo* operon of *Streptococcus pyogenes.* Infect Immun 2002;70:2730–2733. [PubMed: 11953421]

Shi L, Potts M, Kennelly PJ. The serine, threonine, and/or tyrosine-specific protein kinases and protein phosphatases of prokaryotic organisms: a family portrait. FEMS Microbiol Rev 1998;22:229–253. [PubMed: 9862122]

Shu CJ, Zhulin IB. ANTAR: an RNA-binding domain in transcription antitermination regulatory proteins. Trends Biochem Sci 2002;27:3–5. [PubMed: 11796212]

Shu CJ, Ulrich LE, Zhulin IB. The NIT domain: a predicted nitrate-responsive module in bacterial sensory receptors. Trends Biochem Sci 2003;28:121–124. [PubMed: 12633990]

Stock AM, Robinson VL, Goudreau PN. Two-component signal transduction. Annu Rev Biochem 2000;69:183–215. [PubMed: 10966457]

Tal R, Wong HC, Calhoon R, Gelfand D, Fear AL, Volman G, et al. Three *cdg* operons control cellular turnover of cyclic di-GMP in *Acetobacter xylinum*: genetic organization and occurrence of conserved domains in isoenzymes. J Bacteriol 1998;180:4416–4425. [PubMed: 9721278]

Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res 2000;28:33–36. [PubMed: 10592175]

Taylor BL, Zhulin IB. PAS domains: internal sensors of oxygen, redox potential, and light. Microbiol Mol Biol Rev 1999;63:479–506. [PubMed: 10357859]

Vijay K, Brody MS, Fredlund E, Price CW. A PP2C phosphatase containing a PAS domain is required to convey signals of energy stress to the sigmaB transcription factor of *Bacillus subtilis.* Mol Microbiol 2000;35:180–188. [PubMed: 10632888]

Volz K. Structural conservation in the CheY superfamily. Biochemistry 1993;32:11741–11753. [PubMed: 8218244]

Weinhouse H, Sapir S, Amikam D, Shilo Y, Volman G, Ohana P, Benziman M. c-di-GMP-binding protein, a new factor regulating cellulose synthesis in *Acetobacter xylinum.* FEBS Lett 1997;416:207–211. [PubMed: 9369216]

Wietzorrek A, Bibb M. A novel family of proteins that regulates antibiotic production in streptomycetes appears to contain an OmpR-like DNA-binding fold. Mol Microbiol 1997;25:1181–1184. [PubMed: 9350875]

Williams SB, Stewart V. 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 1999;33:1093–1102. [PubMed: 10510225]

Wilson SA, Wachira SJ, Norman RA, Pearl LH, Drew RE. Transcription antitermination regulation of the *Pseudomonas aeruginosa* amidase operon. EMBO J 1996;15:5907–5916. [PubMed: 8918468]

Yang X, Kang CM, Brody MS, Price CW. Opposing pairs of serine protein kinases and phosphatases transmit signals of environmental stress to activate a bacterial transcription factor. Genes Dev 1996;10:2265–2275. [PubMed: 8824586]

Yashiro K, Sakamoto T, Ohmori M. Molecular characterization of an adenylate cyclase gene of the cyanobacterium *Spirulina platensis.* Plant Mol Biol 1996;31:175–181. [PubMed: 8704152]

Yeats C, Finn RD, Bateman A. The PASTA domain: a beta-lactam-binding domain. Trends Biochem Sci 2002;27:438. [PubMed: 12217513]

Yeliseev AA, Kaplan S. TspO of *Rhodobacter sphaeroides*. A structural and functional model for the mammalian peripheral benzodiazepine receptor. J Biol Chem 2000;275:5657–5667. [PubMed: 10681549]

Zhang CC, Gonzalez L, Phalip V. Survey, analysis and genetic organization of genes encoding eukaryotic-like signalling proteins on a cyanobacterial genome. Nucleic Acids Res 1998;26:3619–3625. [PubMed: 9685474]

Zhulin IB. The superfamily of chemotaxis transducers: from physiology to genomics and back. Adv Microb Physiol 2001;45:157–198. [PubMed: 11450109]

Zhulin IB, Nikolskaya AN, Galperin MY. Common extracellular sensory domains in transmembrane receptors for diverse signal transduction pathways in bacteria and archaea. J Bacteriol 2003;185:285–294. [PubMed: 12486065]
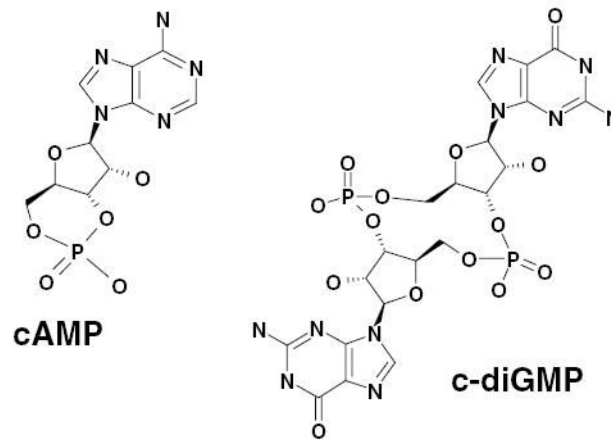
Fig. 1.
Structures of cyclic AMP (cAMP) and cyclic diguanylate (c-diGMP).
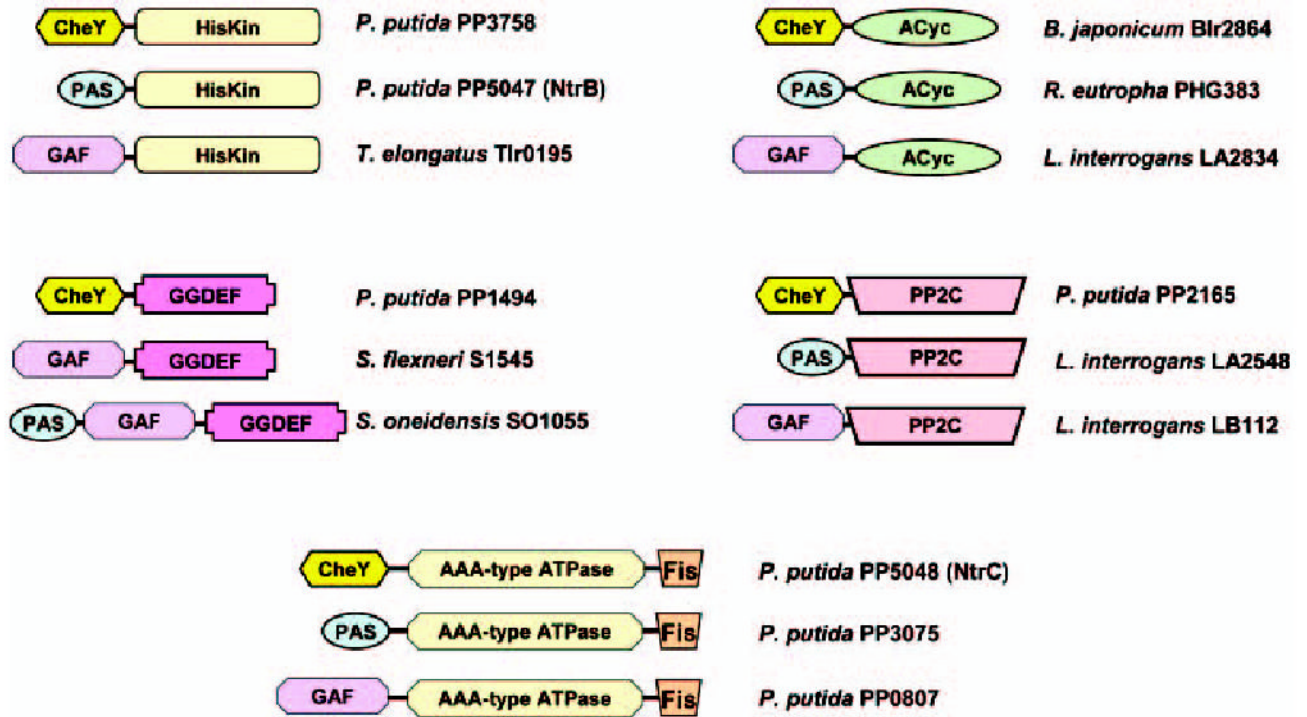
Fig. 2.
Parallelism in structures of response regulators and intracellular signalling proteins.
Transmission of environmental signals from sensory kinases to response regulators involves
phosphorylation of Asp residues in their CheY-like receiver domains, which changes the
conformation of this domain, liberating the downstream DNA-binding or enzymatic domains.
Likewise, ligand (e.g. oxygen) binding by the cytoplasmic PAS and/or GAF domains
(intracellular signal) changes the conformation of these domains, liberating the downstream
DNA-binding or enzymatic domains. Domain names are as Table 2, with the exception of
HisKin, indicating a histidine kinase, and Fis, a DNA-binding domain. The examples are taken
from the recently sequenced genomes of *Bradyrhizobium japonicum*, *Leptospira
interrogans*, *Pseudomonas putida*, *Ralstonia eutropha*, *Shigella flexneri*, *Shewanella
oneidensis* and *Thermosynechococcus elongatus*. See Table 4 for more examples.

Table 1

Distribution of environmental signalling systems in free-living and parasitic bacteria[a]

| Group, species[b] | Total No. of proteins | Transmembrane sensor molecules[c] | | | | | | All sensors | All DNA-binding response regulators[c] | | | | | | | | All CheY-X response regulators |
		HisK	ACyc	MCP	GGDEF	HD-GYP	S_TKc		AraC	NarL	OmpR	NtrC	LytTR	Fis		
α-Proteobacteria																
Mesorhizobium loti	6752	33 (61)	2 (13)	1	18 (32)	0 (1)	1 (3)	111	0	11	23	5	0	2		45
Caulobacter crescentus	3737	38 (62)	1 (2)	19	3 (11)	0	1	95	0	2	12	4	0	2		26
*Brucella melitensis	3197	13 (22)	0	0	7 (9)	0	1	32	0	5	10	3	0	1		19
*Rickettsia prowzekii	834	4	0	0	0 (1)	0	0 (1)	6	0	0	2	1	0	0		4
γ-Proteobacteria																
Pseudomonas aeruginosa	5565	44 (62)	1	26	19 (33)	1 (3)	2 (9)	134	0	17	24	10	1	1		66
Shewanella oneidensis	4931	34 (45)	1	26	26 (52)	1 (9)	1 (7)	140	0	8	18	5	1	1		47
Escherichia coli	4289	27 (30)	0	5	12 (19)	0	1 (2)	56	0	12	14	4	2	0		32
Vibrio cholerae	3827	30 (41)	0	45	26 (41)	2 (9)	1 (2)	138	0	5	12	8	2	0		
*Coxiella burnetii	2134	2 (7)	0 (1)	0	0 (1)	0	1 (4)	13	0	4	3	0	0	0		7
*Pasturella multocida	2014	9	0	0	0	0	1 (2)	11	0	3	5	1	0	0		9
*Haemophilus influenzae	1709	4	0	0	0	0	0 (2)	6	0	1	4	0	0	0		5
Gram-positive bacteria																
Bacillus subtilis	4100	30 (36)	0	10	3 (4)	0	2	52	1	9	13	0	2	0		33
Bacillus halodurans	4066	35 (42)	0	12	2 (4)	0 (2)	3 (5)	67	7	7	17	2	0	0		40
Oceanobacillus iheyensis	3496	16 (19)	0	9	2 (4)	0 (2)	2 (4)	38	1	5	8	1	1	1		19
*Listeria innocua	2973	13 (14)	0	3	5	0	2	21	1	2	10	0	2	0		16
*Streptococcus pneumoniae	2236	13	0	0	1	0	1 (4)	18	2	2	8	0	2	0		14

[a]. The numbers are from the COG and SMART databases or from the results of iterative PSI-BLAST searches of domain-specific profiles against the protein sets encoded in each organism (see Galperin et al., 2001a; for details). Certain diverged and/or inactivated proteins were not counted. A variant of this table with appropriate WWW links is available at the author's web site: ftp://ftp.ncbi.nih.gov/pub/galperin/SignalCensus.html

b. The names of obligate parasites marked by an asterisk. The numbers of encoded proteins are from the latest entries on the NCBI's Entrez Genomes web site, http://www.ncbi.nlm.nih.gov/genomes/ MICROBES/Complete.html.

c. Each figure represents the number of predicted transmembrane sensor proteins (with at least one transmembrane segment) or response regulators with the given type of the DNA-binding domain encoded in each particular genome. The total number of proteins containing the given signalling domain, if different, is shown in parentheses. No attempt was made to account for membrane receptors with non-sensory membrane domains, such as UhpB (Kadner, 1995). Many Ser, Thr-specific protein kinases listed here have N-terminal S_TKc domains and C-terminal membrane anchors. The last column gives the count of all two-domain response regulators, including Spo0A-type, CheY-CheB, CheY-GGDEF, etc.

NIH-PA Author Manuscript

Table 2

Most common signalling domains. Protein domains are usually identified as conserved sequence modules shared by two or more different proteins. Therefore, domain names are either based on conserved sequenced patterns or on the names of the proteins where these domains were first described. In the rare cases when the enzymatic activity or binding specificity of a new domain is already known, the name usually reflects this information. In most other cases, the domain name has mostly historical significance and spelling it out is unlikely to help in understanding its function. For better clues, one could look up the original domain description or the comments for the particular entry in the domain databases, such as SMART.

Domain name	Name origin	Function	Length, aa	Structure (PDB entry)	Copies per genome[a]				Reference
					Ecoli	Bsub	Syne	Mace	
ACyc	Abbreviation: Adenylyl/ guanylyl cyclase type 3 catalytic domain	Formation of cAMP or cGMP	~190	1CUL	0	0	3	1	Yashiro et al. (1996); Katayama and Ohmori (1997)
CheY	Similarity to the CheY chemotaxis protein	Phosphoacceptor domain with an active site Asp	~120	2CHE	37	36	58	20	Volz (1993); Grebe and Stock (1999)
EAL	Conserved sequence motif	c-diGMP phosphodiesterase (?)	~250	—	17	3	13	0	Tal et al. (1998)
GAF	Common domain in cGMP phosphodiesterases, adenylyl cyclases and FhlA protein	cGMP, cAMP and photopigment binding	~150	1F5M	9	3	28	15	Aravind and Ponting (1997); Ho et al. (2000)
GGDEF	Conserved sequence motif	Diguanylate cyclase (?)	~180	—	19	4	23	0	Tal et al. (1998)
HAMP	Common domain in histidine kinases, adenylyl cyclases and methyl carrier proteins	Dimerization (?)	~50	1JOY	19	17	1	7	Aravind and Ponting (1999); Williams and Stewart (1999)
HATPase	Abbreviation: histidine kinase-type ATPase catalytic domain	Phosphoryl transfer from ATP to the HisK domain	~110	1BXD	30	36	42	53	Grebe and Stock (1999); Stock et al. (2000)
HD-GYP	Conserved sequence motifs	Phosphodiesterase	~170	—	0	0	2	0	Galperin et al. (1999)
HisKA	Histidine kinase A domain	Dimerization/ phosphoacceptor domain with an active site His	~120	1JOY	28	33	42	51	Grebe and Stock (1999); Stock et al. (2000)
HPt	Abbreviation: Histidine phosphotransfer domain	Phosphoacceptor domain with an active site His	~100	2A0B	6	1	7	3	Matsushika and Mizuno (1998)
MCP	Abbreviation: Methyl carrier protein	Methylated domain with an active site Glu	~350	1QU7	5	10	4	2	Falke et al. (1997)
PAS	Common domain in Period circadian protein (Per), Aryl hydrocarbon receptor nuclear translocator (ARNT), and Single-minded protein (Sim)	Haeme, flavin, adenine, cinnamic acid binding, sensing light, oxygen and other parameters	~100	1DRM	14	14	28	54	Taylor and Zhulin (1999)
PP2A	Protein phosphatase family 2 A	Hydrolysis of pSer and pThr	~180	1TCO	3	1	1	1	Shi et al. (1998)
PP2C	Protein phosphatase family 2C	Hydrolysis of pSer and pThr	~250	1A6Q	0	5	8	0	Shi et al. (1998)
PTP	Low molecular weight protein tyrosine phosphatase	Hydrolysis of pSer, pThr, pTyr	~150	1D1P	2	3	2	1	Shi et al. (1998)
S_TKc	Abbreviation: Ser, Thr-specific kinase, catalytic domain	Phosphorylation of Ser and Thr residues in target proteins	~250	1JWH	2	3	12	5	Shi et al. (1998)

a. The data are taken from SMART and COG databases (corrected for redundant entries) or from the results of iterative psi-blast searches of domain-specific profiles against the protein set of each given organism. Each figure represents the number of proteins in a given genome that contain the corresponding domain; multiple occurrences of the same domain on a single polypeptide chain are counted as one, see (Galperin et al., 2001a) for more details. Ecoli, Escherichia coli K12; Bsub, Bacillus subtilis, Syne, Synechocystis sp. PCC6803; Mace, Methanosarcina acetivorans. The

discrepancies with the previously reported *Synechocystis* data (Zhang *et al.*, 1998) are due to the inclusion of ABC1-like predicted protein kinases and PTC1-like protein phosphatases. Complete lists of proteins that contain each particular domain are available at ftp://ncbi.nlm.nih.gov/pub/galperin/SignalCensus.html.

Table 3

Diverse transmembrane receptors with recently described sensory domains.

Sensor domain	Binding specificity	Pfam	CDD	COG	InterPro	Signal output domains[a]	Examples (Name, accession no.)[b]	Domain description
Extracytoplasmic sensor domains								
Cache	Small molecules	PF02743	3250	—	IPR004010	HisKin ACyc GGDEF MCP PP2C	VC1831 (AAF94979) CyaA (CAC41696) PP2413 (AAN68025) McpB (CAB15104) Mll6700 (BAB52943)	Anantharaman and Aravind (2000)
CHASE	Unknown	PF03924	12938	COG3614	IPR005620	HisKin ACyc	PA4112 (AAG07499) CyaG_Dicdi (Q03101)	Anantharaman and Aravind (2001); Mougel and Zhulin (2001)
CHASE2	Unknown (osmotic stress)	PF05226	13500	COG4252	IPR007890	GGDEF HisKin	VC1376 (AAF94534) PA4036 (AAG07423)	Kimura et al. (2002); Zhulin et al. (2003)
CHASE3	Unknown	PF05227	14400	COG5278	IPR007891	ACyc GGDEF ST_Kc HisKin ACyc GGDEF MCP	All1118 (BAB73075) All1219 (BAB73176) All4838 (BAB76537) Alr3761 (BAB75460) CyaA (BAA22996) All4897 (BAB76596) McpG (AAL41754)	Zhulin et al. (2003)
CHASE4	Unknown	PF05228	12656	COG3322	IPR007892	HisKin GGDEF	MA1739 (AAM05146) PA0847 (AAG04236)	Zhulin et al. (2003)
CHASE5	Unknown	—	—	—	—	HisKin GGDEF	PA2824 (AAG06212) SO1208 (AAN54277)	Zhulin et al. (2003)
CHASE6	Unknown	— HD-GYP	13498	COG4250 All4304 (BAB76003)	—	HisKin	Slr0484 (BAA10317)	Zhulin et al. (2003)
PASTA	β-Lactams	PF03793	9525	COG2815	IPR005543	S_TKc	Pkn2_Bacsu (O34507)	Yeats et al. (2002)
Membrane-bound sensor domains								
ETR1	Ethylene	—	—	—	—	HisKin	Slr1212 (P73363)	Mount and Chang (2002)
MHYT	Metal (?)	PF03707	12634	COG3300	IPR00530	HisKin GGDEF LytTR LytTR	Slr 2098 (BAA17992) PA1727 (AAG05116) CoxC (CAB76246) CC1610 (AAK23589)	Galperin et al. (2001b)
MHYE	Metal (?)	—	—	—	—			Nikolskaya and Galperin (2002)
MASE1	Unknown	PF05231	12775	COG3447	IPR007895	HisKin	Sll1672 (BAA16806)	Nikolskaya et al. (2003)
MASE2	Unknown	PF05230	—	—	IPR007894	GGDEF ACyc GGDEF HisKin	Slr 1798(BAA16824) PA3217 (AAG06605) AdrA (CAD08840) PA3271 (AAG06659)	Häse et al. (2001)
PutP-like	Unknown	—	10461	COG0591	—			
Cytoplasmic sensor domains								
BLUF	FAD	PF04940	16235	—	IPR007024	ACyc	PAC (BAB85620)	Gomelsky and Klug (2002)

Sensor domain	Binding specificity	Domain database entry				Signal output domains[a]	Examples (Name, accession no.)[b]	Domain description
		Pfam	COG	CDD	InterPro			
HNOB	Haeme, NO	—	—	21748	—	EAL None MCP GuCyc None	YcgF_Ecoli (P75990) AppA (AAA81591) CAC3243 (AAK81177) Alr2278 (BAB73977) SO2144 (AAN55189)	Iyer et al. (2003)
KdpD	Sensor for K+, osmolarity	PF02702	COG2205	23365	IPR003852	HisKin	KdpD_Ecoli (P21865)	Heermann et al. (2000, 2003)
Phytochrome	Tetrapyrroles (light-sensing)	PF00360	—	16708	IPR001294	HisKin S_TKc	BphY_Deira (Q9RZA4) Phy1_Cerpu (P25848)	Davis et al. (1999)

[a]. The domain names are as in Table 2. HisKin denotes a histidine kinase (a combination of HisKA and HATPase or their variants), LytTR indicates a DNA-binding domain, GuCyc stands for soluble guanylate cyclase, 'none' indicates a stand-alone sensor domain.

[b]. Protein names and accession numbers are from Swiss-Prot and NCBI protein databases.

Table 4

Intracellular signalling proteins.

Domain organization	Examples (Organism, protein name, accession no.)
Histidine kinases	
CheY-HisKin	*Synechocystis* sp. Sll1400 (BAA16953)
PAS-HisKin	*Escherichia coli* NtrC (P06712)
GAF-HisKin	*Synechocystis* sp. Slr1147 (BAA17304)
CheY-PAS-PAS-HisKin	*Methanosarcina acetivorans* MA1267 (AAM04686)
PAS-PAS-PAS-HisKin	*Methanosarcina acetivorans* MA3543 (AAM06905)
PAS-GAF-PAS-HisKin	*Methanosarcina acetivorans* MA2266 (AAM05659)
GAF-PAS-GAF-HisKin	*Methanosarcina acetivorans* MA0863 (AAM04302)
Adenylate cyclases	
CheY-Acyc	*Stigmatella aurantiaca* CyaB (P40138)
PAS-Acyc	*Ralstonia eutropha* PHG383 (AAP86132)
GAF-Acyc	*Thiocapsa roseopersicina* CyaA (AAK56849)
GAF-GAF-PAS-Acyc	*Anabaena* sp. CyaB1 (BAA13998)
CheY-GAF-GAF-HisK-CheY-ACyc	*Spirulina platensis* CyaC (BAA22997)
FHA-Acyc	*Synechocystis* sp. Slr 1991 (BAA17880)
FHA-GAF-Acyc	*Anabaena* sp. All0743 (BAB72700)
ACyc-TPR	*Xanthomonas campestris* XCC0240 (AAM39559)
Diguanylate cyclases/phosphodiesterases	
CheY-GGDEF	*Pseudomonas fluorescens* WspR (AAL71852)
GAF-GGDEF	*Esherichia coli* YeaP (P76245)
CheY-GGDEF-EAL	*Shewanella oneidensis* SO0545 (AAN53626)
PAS-GGDEF-EAL	*Acetobacter xylinum* PHE1 (AAC61683)
	Escherichia coli Dos (BAA15160)
GAF-GGDEF-EAL	*Vibrio cholerae* VCA0080 (AAF95994)
PAS-GAF-GGDEF	*Shewanella oneidensis* SO1055 (AAN54127)
GGDEF-EAL-CheY	*Xanthomonas campestris* XCC1958 (AAM41247)
GAF-PAS-GGDEF-EAL	*Shewanella oneidensis* SO2498 (AAN55529)
GAF-GAF-GGDEF-EAL-GAF-GGDEF	*Synechocystis* sp. Cph2 (BAA10536)
EAL-GGDEF	*Shewanella oneidensis* SO0555 (AAN53636)
HD-GYP-GGDEF	*Aquifex aeolicus* aq_2027 (O67821)
TPR-TPR-GGDEF	*Vibrio cholerae* VC1367 (AAF94525)
Protein phosphatases	
CheY-PP2C	*Bacillus cereus* BC1006 (AAP07993)
CheY-PAS-PP2C	*Geobacter sulfurreducens* GSU0700 (AAR34030)
PAS-PP2C	*Bacillus subtilis* RsbP (O07014)
GAF-PP2C	*Streptomyces coelicolor* SCO3723 (CAD55330)
PAS-GAF-PP2C	*Streptomyces coelicolor* SC5C11.05 (CAB76311)
ANTAR-PP2C	*Streptomyces coelicolor* SCO7326 (CAB92872)
Fusions to PTS proteins	
GAF-PtsI	*E. coli* PtsP (AAB40476)
CelB-EAL	*Vibrio cholerae* VC1211 (AAF94370)

The domain names are as in Table 2; FHA (forkhead-associated) is a phosphothreonine-binding domain, TPR (tetratricopeptide repeat) is involved in protein–protein interactions. PtsI is Enzyme I and CelB is a cellobiose transporter, respectively, of the PEP-dependent sugar:phosphotransferase system. Only a small selection of known domains combinations for intracellular signalling proteins is listed. For other common domain combinations, see SMART and CDD databases.