Published in final edited form as:
Annu Rev Microbiol. 2011 ; 65: 261–286. doi:10.1146/annurev-micro-121809-151631.

Ligand binding PAS domains in a genomic, cellular, and structural context

Jonathan T. Henry[1] and **Sean Crosson**[1,2]

[1]The Committee on Microbiology, The University of Chicago, Chicago, IL 60637

[2]Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL 60637

Abstract

Per-Arnt-Sim (PAS) domains occur in proteins from all kingdoms of life. In the bacterial kingdom, PAS domains are commonly positioned at the amino terminus of signaling proteins such as sensor histidine kinases, cyclic-di-GMP synthases/hydrolases, and methyl-accepting chemotaxis proteins. Although these domains are highly divergent at the primary sequence level, the structures of dozens of PAS domains across a broad section of sequence space have been solved, revealing a conserved three-dimensional architecture. An all-versus-all alignment of 63 PAS structures demonstrates that the PAS domain family forms structural clades on the basis of two principal variables: (a) topological location inside or outside the plasma membrane and (b) the class of small molecule that they bind. The binding of a chemically diverse range of small-molecule metabolites is a hallmark of the PAS domain family. PAS ligand binding either functions as a primary cue to initiate a cellular signaling response or provides the domain with the capacity to respond to secondary physical or chemical signals such as gas molecules, redox potential, or photons. This review synthesizes the current state of knowledge of the structural foundations and evolution of ligand recognition and binding by PAS domains.

Keywords

PAS; sensor; ligand binding; signal transduction

Introduction

The fitness of a bacterial cell is dependent on its ability to sense and adapt to changes in the physicochemical makeup of its environment. To this end, bacteria express a variety of sensory and signal transduction proteins, among which the PAS domain is widely utilized

*Corresponding Author: Sean Crosson, Department of Biochemistry and Molecular Biology, The University of Chicago, 929 East 57th Street, GCIS W138, Chicago, IL 60637, Phone: (773) 834-1926, Fax: (773) 702-0439, scrosson@uchicago.edu .

Annotated references:

Taylor BL, Zhulin IB. 1999. PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiol Mol Biol Rev* 63:479-506. This classic review presents an encyclopedic study of early and foundational PAS domain research.

Hefti MH, Francoijs K-J, de Vries SC, Dixon R, Vervoort J. 2004. The PAS fold. A redefinition of the PAS domain based upon structural prediction. *Eur J Biochem* 271:1198-208. PAS domain nomenclature as well as definition as a conserved structural form are put forth in this publication.

Möglich A, Ayers RA, Moffat K. 2009. Structure and signaling mechanism of Per-ARNT-Sim domains. *Structure* 17:1282-94. A thorough distillation of current structural knowledge of PAS domain form and function is found in this review.

Krell T, Lacal J, Busch A, Silva-Jimenez H, Guazzaroni M-E, Ramos JL. 2010. Bacterial sensor kinases: diversity in the recognition of environmental signals. *Annu Rev Microbiol* 64:539-59. This review of bacterial sensor kinases focuses in part on PAS domains, the domainant signal input in two-component systems.

(110). The PAS domain consists of ≈100 amino acids and is found coupled to a wide range of enzymatic and non-enzymatic effector domains that function within multiple classes of cellular signaling systems. These domains are annotated in all but the smallest bacterial genomes, and have important sensory functions in archaea and eukaryotes as well (105). PAS domains perform a variety of functions within sensory proteins by promoting protein/protein interaction (68; 73; 81) or signal transfer (82), and also by directly sensing perceived stimuli (105). The utility of PAS domains in performing this last function, that of a direct cellular sensor, is in part due to their plasticity in binding different substrates (79). Indeed, over the course of evolution, the PAS domain has been selected to bind a remarkable array of cofactors and ligands. Depending on the system, the binding of small molecules or ions to a PAS domain can either serve as a direct signal (20; 94; 112), or provide a cofactor that enables perception of signals such as dissolved gases (41; 107), redox potential (90; 92; 100; 108), or visible light (22; 91; 103).

Although several examples of PAS sensor/signal pairs have been experimentally characterized over the past two decades, only a small fraction of PAS domains annotated within protein domain databases have been paired with a signal or ligand (79). While it is unlikely that all or even most PAS domains bind a small molecule, it can be assumed that the full signal recognition capacity of this domain remains largely underdetermined. Two major stumbling blocks make PAS ligand identification difficult. First, though a conserved three-dimensional fold characterizes PAS domains (79), these domains often share little primary sequence identity. Thus, only a few clues are available to the investigator about substrate recognition based on protein sequence data alone. Secondly, PAS domains appear to be almost completely unrestricted in the class of substrate that can be bound. Investigators studying PAS sensors are faced with infinite possibilities and few hints as to the signal sensed. Herein, we first provide an overview of PAS signaling protein diversity across prokaryotic taxa. We then analyze select classes and specific examples of PAS sensors, in an effort to define both commonalities and peculiarities of PAS domains and the signals with which they interact.

A Note on PAS Nomenclature

The PAS domain was first noted as a conserved entity by sequence comparison between the fly clock protein PERIOD, the vertebrate aryl hydrocarbon nuclear translocator (ARNT), and the fly developmental regulator single-minded (SIM), hence **P**er-**A**rnt-**S**im or PAS (53). Excellent early and recent reviews describe the evolution of modern understanding of the PAS domain; briefly, what was once described as a PAS repeat, PAS/PAC motif or S1/S2 box is now know to form a single coherent protein fold of approximately 100 residues in length (46; 75; 105). This fold is composed of a single anti-parallel, five-stranded β-sheet with a 2-1-5-4-3 strand order. Intervening α-helices create a pocket upon the β-sheet, within which ligand binding often occurs. The structure of *Halorhodospira halophila* photoactive yellow protein (PYP) (13) was first put forth as the prototype of the PAS fold in 1998 (86). Since this time, crystal structures of dozens of PAS domains have been solved and deposited in the Protein Data Bank (PDB) (**see Supplemental Material**). Sequences encoding PAS domains are found singly, in tandem, and sometimes in many copies within a single protein. In this review, we use the term “PAS domain” to refer to single protein regions whose known or predicted core β-sheet resembles PYP.

PAS Domain Prediction and Representation in the Prokaryotic Kingdom

Known PAS domains share less than 20% average pairwise sequence identity (34). Although PAS domains were originally identified using basic sequence alignment algorithms (53), statistical scoring metrics such as Position Specific Scoring Matrices

(PSSMs) and profile Hidden Markov Models (HMMs) have been used to identify the great majority of PAS domains currently annotated in the Conserved Domain Database (74). To date, the Pfam database maintains seven different profile-HMMs which, when applied to the entire non-redundant set of publically available protein sequences, apprehend over 33,000 PAS domains. 87% of predicted PAS domains are encoded from prokaryotic genomes (34). Within the set of bacteria encoding at least one PAS domain, ~27,500 PAS domains are annotated among ~1500 organisms, giving a higher ratio of PAS domains per prokaryote (18.3) than eukaryotes (5.0), but a lower ratio than archaea (45.5).

An Overview of the PAS Domain in Bacterial Sensory Biology

Tens of thousands of PAS domains have been annotated, and though only a subset of these domains likely serve as sensors, most appear to play a role in signal transduction. Examples of signal transduction proteins include histidine kinases (HK), methyl-accepting chemotaxis proteins (MCP), and nucleotide cyclases/phosphodiesterases (often containing GGDEF/EAL domains), each of which mediate downstream responses to environmental stimuli (37). Several previous studies have cataloged the combinatorial pairing of bacterial signaling input and output domains (64; 67; 110). Our analysis of the full set of PAS-containing proteins annotated in Pfam (Figure 1) shows that the top twelve domains found in PAS proteins are related to sensory input (GAF), transduction (HAMP), or output (HK, GGDEF, MCP, *et cetera*). Remarkably, almost half of all PAS proteins are histidine kinases. Tandem and multiple PAS domains are common in individual proteins: about one third of PAS proteins contain two or more PAS domains. Thus PAS domains most often function within the context of signal transduction proteins. We further characterized the spatial relationship between each individual PAS domain and other domains on the same protein (Figure 2). PAS domains commonly neighbor one another and are often situated N-terminal to known signal transduction outputs (37). This domain architecture follows a common paradigm among sensor proteins: sensory input domains such as PAS perceive a signal, often via ligand binding or cofactor modification, and information is transmitted to C-terminal output domains (85).

A Structure-based PAS Domain Classification Provides Insight into Ligand Binding and Cellular Localization

To understand the evolution of PAS domains in signal transduction proteins, we sought to expand upon previous phylogenetic analyses of PAS domains. The body of known PAS sequences resides in the “twilight zone” of sequence homology, possessing low pairwise identity with no universally conserved residues (50; 113). Basic multiple sequence alignment is often unreliable for sequence sets this diverse, confounding attempts at sequence-based phylogenetics (113). Knowledge of protein structure can improve multiple sequence alignments by confirming homologous residues; furthermore, structure-based analysis may provide sequence-independent information on protein relatedness (96). At the time of publication, high-resolution structures of over sixty individual PAS domains had been deposited in the Protein Data Bank (10). Various attempts have been made to integrate this body of information using pairwise comparisons to create structure-based trees and sequence alignments (54; 79). We performed an independent analysis of PAS domain relatedness using multiple structure alignment. The resultant all-verses-all root mean square deviation (RMSD) values allowed clustering analysis and production of a structural relatedness tree (Figure 3). In agreement with previous analyses, PAS domains cluster with those that bind the same or similar ligands or cofactors (54; 79). For example, domains binding flavin mononucleotide (FMN) are more closely related in structure to others that bind FMN, and are also related to those binding flavin adenine dinucleotide (FAD). Subsequent structure-based sequence comparison as well as investigation of ligand binding

confirms the relatedness of these clades, suggesting a specific common ancestor for most classes of ligand- or cofactor-binding PAS domain within our set of structures. Heme binding poses one interesting exception to this observation; structures have been solved for cytoplasmic PAS domains that bind heme *b*, as well as extracytoplasmic PAS domains that bind heme *c* in a distinct manner and location (44; 88). Given such divergence in sequence and structure space, it appears that heme binding has evolved at least twice over the course of PAS domain evolution.

In contrast to observed clustering based on ligand or cofactor binding, PAS domains do not appear to cluster strongly with signaling output modality. Most clusters are composed of PAS domains coupled to diverse outputs, most commonly HK, GGDEF and MCP domains. However, PAS domains do fall into structural clades related to their cellular topological location, in either the cytoplasm or outside the cell in the periplasmic or extracellular space. Differences between cytoplasmic PAS domains and extracytoplasmic PAS domains have been noted: though both possess a conserved β-sheet core, extracytoplasmic PAS domains are often anchored to the membrane by a long N-terminal α-helical spine, and most have a reduced complement of helices between the second and third β-strand (18; 19; 46; 79). There is some debate as to whether extracytoplasmic domains should be considered canonical PAS domains or PAS-like “PDC domains,” named for members PhoQ, DcuS, and CitA (16; 18; 19; 79). Since all domains under consideration are structurally similar, often serve as signal recognition modules with and without cofactors, and transduce signals to the same output domains (e.g. HK, GGDEF, and MCP), we will use the terms cytoplasmic and extracytoplasmic PAS domains for the purposes of this review. The collection of aligned PAS domains is available in the supplementary material (Supplementary File 1), allowing readers of this review to easily compare and contrast PAS domain structures.

Several protein domains for which high-resolution structures exist have been described as “PAS-like,” superimposing well with known PAS domains but possessing an altered complement of β-strands, for example (3; 45; 48). We have included in our analysis the structures of two cytoplasmic PAS-like domains, one found at the C-terminus of the *B. subtilis* phosphodiesterase YcuI, and a second found within the *D. deserti* DNA damage response protein IrrE (77; 114). Each contains an extra β-strand but superposes well with the cores of recognized PAS domains. Interestingly, these cytoplasmic PAS-like proteins cluster with extracytoplasmic PAS domains and possess some of their defining features. The IrrE domain, whose similarity to GAF domains and CitA was noted by its discoverers, is missing the same helical elements absent in other extracytoplasmic PAS domains (114). Similarly, the investigators who solved the structure of YcuI note a long N-terminal α-helical spine similar to those found anchoring extracytoplasmic PAS domains to the outer surface of the membrane, though this protein is not known to be membrane-associated (77). Future structural studies will be necessary to place these domains within the PAS evolution of form and may provide a link between cytoplasmic and extracytoplasmic PAS domains, as well as to the seemingly related GAF sensor domain (5; 48). Nevertheless, it appears clear that our knowledge of PAS domain structure and function, particularly in the realm of extracytoplasmic PAS domains, is rapidly expanding.

PAS Domains are a Dominant Extracytoplasmic Sensor Domain

Recent crystallographic and bioinformatic analyses provide evidence that PAS domains are highly underpredicted within the extracytoplamsic sensory region of histidine kinases. Using a combination of structural biology and *in silico* protein modeling, Chang and colleagues (16) have predicted that 11 of 13 *Bacillus subtilis* sensor histidine kinases containing large extracytoplasmic domains (>40 non-membrane spanning amino acids) have a predicted PAS-like fold. Based on their analysis of *B. subtilis* and *E. coli*, PAS is predicted to be the

dominant extracytoplamsic sensor domain in histidine kinases. Such structure-based domain predictions promise to expand the annotated set of PAS domains across bacterial genomes. Indeed, our simple application of the structure/sequence profile-profile matching algorithm, Phyre (57), identified 4 additional high probability PAS sensor histidine kinases in the genome of *Caulobacter crescentus*, that are not predicted using purely sequenced-based HMM or PSSM prediction algorithms. These findings likely not only apply to histidine kinases, but also to the extracytoplasmic sensor domains of MCP proteins. A notable example of this is McpB of *Bacillus subtilis*, which has a tandem pair of extracytoplasmic PAS domains predicted by profile-profile homology modeling. One of these two putative PAS domains binds asparagine with low micromolar affinity and is required for regulation of asparagine chemotaxis (42). As will be detailed below, structurally-characterized signaling proteins encoding multiple extracytoplamsmic PAS domains (e.g. NifL, MmoS, DcuS, CitA, DctB) generally appear to bind ligand through their amino-terminal PAS domains, though ligand-binding carboxy-terminal PAS domains have been described (98; 115).

Individual PAS Domains Play Different Roles in Signal Transduction Proteins, and Many Function as Sensory Inputs

PAS domains have been shown to perform a variety of functions within signal transduction proteins. PAS domains are often important mediators of protein-protein interactions; examples include the PAS domains of sporulation kinase KinA and quorum sensing protein LuxQ (68; 81). PAS domains have also been implicated in signal transfer and subcellular localization. For example, in plant phytochrome B, mutations in an N-terminal PAS domain block signaling from a chromophore-containing GAF domain (82), while a C-terminal PAS domain provides a nuclear localization signal (17). Unfortunately, PAS domain sequences appear to hold few clues that would suggest these roles to the investigator of a novel protein. However, a wealth of molecular, biochemical, and structural data have been amassed in the study of PAS sensors that directly bind a sensed ligand or accessory cofactor, and careful comparison of novel PAS domains to well-characterized ones has proven a fruitful strategy in identifying PAS sensors and functions. Examples include references (115) and (103). The remainder of this review will focus on this final role, highlighting the interplay between the PAS sensor and the signal sensed. We hope this comparative analysis of PAS sensors and ligands will aid those investigators studying novel PAS proteins, and help differentiate classes of ligand-binding PAS sensors from PAS domains playing myriad other roles in microbial signal transduction.

THE STRUCTURAL BASIS OF LIGAND BINDING IN THE PAS DOMAIN FAMILY

The structure-based phylogenetic groupings described in Figure 3 provide the foundation for our analysis of PAS protein-ligand interactions. Presuming that all PAS domains share a common ancestor, we note the remarkable evolutionary plasticity of these domains, as they have been selected to specifically bind a chemically diverse array of small molecules while maintaining the conserved PAS fold. Indeed, the high “evolvability” of PAS has been experimentally characterized in the model PAS domain, photoactive yellow protein (PYP) (87).

Our discussion of ligand binding is by no means comprehensive; recent biochemical and biophysical studies of PAS domains suggest that new classes of functional ligands are certain to be discovered in the future (2; 60). Our analysis is largely limited to those PAS domains for which there is a known functional ligand, and for which high-resolution

structural information in the ligand-bound state exists. This group includes PAS domains that bind heme, flavin mononucleotide (FMN), flavin adenine dinucleotide (FAD), 4-hydroxycinnamic acid (4-HCA), C3-C4 carboxylic acids (malonate, malate and succinate), C6 carboxylic acids (citrate), and divalent metal cations. Recent structural and biochemical data supporting the possibility that fatty acids can serve as a PAS ligand will also be briefly discussed. Our description of ligand recognition/binding centers on defining the PAS domain residues that directly contact the bound ligand. The identification of such ligand binding consensus sequences may inform future predictions of cofactor binding in PAS domains that otherwise have low overall sequence identity.

1) b-type heme binding in a PAS-histidine kinase and a PAS-phosphodiesterase

FixL is a sensor histidine kinase conserved in α-proteobacteria that controls, in a species-dependent manner, the expression of a number of proteins that function under microaerobic or anaerobic conditions (23; 26; 35; 95). FixL was known to sense oxygen (27) via a heme *b* cofactor (40) several years before it was determined that the heme-binding sensory domain is a member of the PAS structural superfamily (44). A related class of *b*-type heme binding PAS domain is encoded by the *Escherichia coli* direct oxygen sensor, DosP (28) and its orthologs. This sensor enzyme functions as an oxygen-regulated phosphodiesterase that catalyzes the cleavage of cyclic-di-GMP into pGpG (104; 107).

FixL-PAS and DosP-PAS1 share a common site of heme ligation via a conserved histidine residue (Figure 4). The conservation of this ligation site suggests a common evolutionary origin for this gas-sensing PAS domain. However, an analysis of the molecular details of interaction between PAS and unliganded ferrous heme *b* in these two domains reveals that other direct interactions between the protein and heme are not conserved. In particular, DosP-PAS1 (66; 84) and FixL-PAS (44; 59; 78) interact with the propionate side chains of heme *b* through a different set of residues. FixL-PAS employs arginine and histidine side chain, and peptide backbone interactions to bind the negatively-charged propionates (Figure 4B). Heme propionates of DosP-PAS have more extensive interactions with solvent and are complexed by asparagine and peptide backbone interactions. Moreover, DosP-PAS heme is ligated by methionine on the distal side in the absence of bound oxygen (Figure 4D). These structures thus present evidence for evolutionary divergence in the molecular details of heme *b* binding from a common heme-binding ancestor.

2) c-type heme binding in PAS chemoreceptors

The identification of PAS domains that bind *c*-type heme was first described in chemoreceptor proteins of *Desulfovibrio vulgaris* (116) and *Geobacter sulfurreducens* (71). Although a cellular function has not been assigned to *G. sulfurreducens* GSU0582 and GSU0935, the structures of the heme *c*-binding extracytoplasmic PAS domains (88) of these chemoreceptors merit discussion as they provide evidence for independent evolution of a second class of heme-binding PAS domains. Structurally, heme *c* is covalently bound via a conserved bi-cysteine ligation site outside of the PAS domain core (Figure 5). The heme cofactor is contained on its opposite face by a helical motif from a second PAS domain, which forms a swapped dimer within the crystal (Figure 5A). Thus, these proteins present a case of a cofactor that is bound and presented on the surface of the PAS domain.

Spectroscopic analysis of these PAS domains has shown that heme *c* iron binds nitric oxide in both its ferrous (Fe^{2+}) and ferric (Fe^{3+}) forms, but is only able to bind carbon monoxide in the ferrous state (15). Future cell physiology studies are necessary to define the physiological ligand sensed by heme *c*, and the regulatory function of these heme *c*-binding chemoreceptors.

3) Visible light perception by flavin-binding PAS domains

A forward genetic screen for regulators of the phototropic response in plants identified a new class of PAS domain that binds a flavin cofactor (21) and forms a cysteinyl-flavin C4(a) covalent adduct in response to blue light absorption (97). The crystal structure of this photosensory PAS domain (24), more specifically termed a LOV domain (25; 52), revealed a conserved set of amino acids that form a network of polar interactions with the isoalloxazine ring, ribtyl chain, and phosphate of the flavin (Figure 6). The presence of this flavin-binding consensus sequence and the conserved photoresponsive cysteine has been used to predict putative photosensory LOV domains in other sequenced genomes. This bioinformatic analysis has shown that the LOV domain is broadly conserved in prokaryotes: the percentage of bacterial genomes encoding a LOV photoreceptor domain is estimated at 3.5% (63) to over 10% (72). It is notable that many bacterial species encoding LOV domains are neither photosynthetic, phototactic, nor pigmented; there are few hypotheses regarding a physiological role for a visible light photosensor in such species. However, recent work in both Gram-positive and Gram-negative bacteria has defined unanticipated regulatory roles for these proteins.

Early functional data for bacterial LOV proteins emerged from the genetic analysis of the σ^B general stress pathway in *Bacillus subtilis*. Specifically the gene *ytvA*, which encodes and N-terminal LOV domain and a C-terminal **s**ulfate **t**ransporter **a**nti-**s**igma factor antagonist (STAS) domain (see Figure 6A), was shown to function as a positive regulator of σ^B (1). It was later revealed that blue light modulates σ^B signaling in *B. subtilis* via YtvA (6). Subsequent work on Gram-negative species in the α-proteobacteria clade has identified LOV histidine kinases that exhibit light regulated autokinase activity and regulate virulence in the mammalian pathogen, *Brucella abortus* (103), and cell adhesion in the freshwater bacterium, *C. crescentus* (91). Recent in vitro studies on the *C. crescentus* LOV histidine kinase, LovK, provide evidence that cellular redox state can provide an additional signaling input into LOV signaling systems (90).

4) Cellular redox/energy regulation by FAD-binding PAS domains

In addition to the LOV domains, there are a number of other PAS domains that bind flavin cofactors. Here, we focus on PAS proteins that function as cellular redox sensors via a bound flavin adenine dinucleotide (FAD) cofactor. Among the most well-studied of these systems is the transmembrane aerotaxis receptor (Aer) of *E. coli* (106), which contains a cytoplasmic, FAD-binding PAS domain that monitors the energy state of the cell by sensing electron transport (12; 93). Although no high-resolution structural data exist for the Aer PAS domain, mutagenesis studies have identified amino acids required for FAD binding (11; 47). The identity of these residues is generally consistent with high-resolution crystal structures of the related FAD-binding, redox-sensing PAS domains of NifL (58) and MmoS (109). In NifL, MmoS, and Aer the isoalloxazine ring moiety and phosphate of FAD is bound within the PAS domain though a series of conserved polar interactions involving asparagine and lysine/arginine residues. Additionally, a conserved tryptophan forms aromatic stacking interactions with the FAD adenine (Figure 7). These data provide evidence for a common evolutionary origin of the FAD-PAS redox sensor.

As in the other domains discussed above, the modularity of this PAS redox sensor is clear. In the case of Aer, the PAS domain is encoded on the same polypeptide as a methyl-accepting chemotaxis domain and ultimately controls the aerotaxis response by modulating interaction with the taxis kinase, CheA, in trans. In NifL, the FAD binding PAS domain is encoded at the N-terminus of a histidine kinase domain that controls expression of nitrogen fixation genes though its redox-dependent interaction with the DNA-binding protein, NifA (70). The PAS domain of the MmoS sensor kinase likely has an analogous role as NifL in

redox-dependent regulation of gene expression, where it is proposed to control the expression of soluble methane monooxygenase (108).

5) Regulation of di- and tricarboxylate transport and metabolism by PAS sensor kinases

Carboxylic acids can serve as a carbon and energy source during aerobic metabolism and can be fermented or serve as terminal electron acceptors during anaerobiosis (14; 29; 55). Bacteria that utilize these carbon compounds have sensory systems that regulate genes required for their uptake and metabolism. Among these are the related DcuS and CitA sensor histidine kinases, which encode extracytoplasmic PAS sensor domains that directly sense C4 and C6 carboxylate-containing substrates (62; 83; 94; 117) (Figure 8). Activation of these sensor kinases by substrate binding controls expression of genes required for C4 and C6 transport and utilization (14; 43). In the rhizobia, C4 dicarboxylates serve as the primary energy and carbon source during nitrogen fixation within the legume root nodule (102). The dicarboxylate transport system in the rhizobiaceae (and other) species is regulated by the DctB sensor kinase, which binds C3 and C4 substrates in a similar region of the PAS domain pocket, but using a set of residues that are distinct from DcuS and CitA (18; 117) (Figure 9). Indeed, a primary sequence alignment of these three sensor PAS domains (DcuS, CitA, and DctB) shows that DctB is distinct from DcuS/CitA (18). Thus, the PAS domain structural scaffold has been selected to bind a whole range of carboxylic acids using different backbone and side chain interactions. However, it is difficult to clearly discern whether the capacity of PAS domains to bind di- and tricarboxylic acids has evolved independently in the DctB and DcuS/CitA systems or if there has been divergence from a common ancestor.

6) Divalent metal binding by the PAS sensor domain of histidine kinase PhoQ

The transmembrane sensor kinase PhoQ and its cognate receiver PhoP are conserved pleiotropic regulators that control processes including type III secretion, Mg^{2+} transport, acid stress resistance, and LPS remodeling (31; 32). PhoQ activates PhoP-dependent transcription under low divalent metal ion concentrations (39); this response is mediated by direct metal binding to the extracytoplasmic PAS domain of PhoQ (38). In pathogenic species PhoQ is known to regulate virulence, and senses a complex array of signals that are particular to the host environment including low pH and low divalent metal ions, and may sense antimicrobial peptides (89).

Structural data on the PhoQ PAS domain have revealed the molecular basis of metal binding. In particular, an acidic cluster of residues on a surface of the PAS domain that is proximal to the outer leaflet of the inner membrane is required for sensing of divalent cations and maintaining the PhoQ kinase in a repressed state (20) (Figure 10). While the first reported structure of PhoQ PAS revealed Ca^{2+} binding to this acidic patch (20), an independently determined crystal structure has shown binding to Ni^{2+} (18). Thus, this molecular surface has promiscuous metal binding properties. The range of divalent metal ligands that are biologically relevant regulators of PhoQ is underdetermined; it may be the case that Mg^{2+}, Ca^{2+}, and Mn^{2+} are all relevant physiological signals under particular environmental conditions.

7) 4-hydroxycinnamic acid binding in photoactive yellow protein confers sensitivity to visible light signals

Photoactive yellow protein (PYP) has long served as the structural prototype of PAS domains (86). This bacterial photosensor consists of a single PAS domain that covalently binds the secondary metabolite 4-hydroxycinnamic acid (*p*-coumaric acid) via a conserved cysteine residue (7; 49); this covalent cofactor confers sensitivity to blue/near-UV light and is stabilized within the PAS pocket by a conserved set of polar interactions (4; 13) (Figure

11). PYP was originally discovered in extracts of the halophilic bacterium *Halorhodospira halophila* (76), where it has been implicated in the regulation of negative phototaxis (101). Multidomain proteins encoding a PYP domain have been discovered in a number of other bacteria (111), the most well-characterized being Ppr of *Rhodospirillum centenum*. This hybrid photoreceptor encodes a blue-absorbing PYP domain coupled to a red-absorbing bacteriophytochrome domain (56), and regulates the development of starvation-resistant cysts (9). Recent characterization of Ppr from *Rhodocista centenaria* provides evidence that this hybrid photoreceptor also functions in the chemotaxis signaling pathway (65).

8) A fatty acid-binding PAS domain in a σ regulatory system

The cofactors discussed to this point are bound to the PAS domain largely through polar side chain and backbone interactions. The recent discovery of bound palmitic acid (C16:0) in the PAS domain crystal structure of *Mycobacterium tuberculosis* σ^F regulatory protein, RV1364c, provides evidence that lipids/fatty acids can potentially function as PAS domain ligands (PDB ID: 3K3C) (61). Although palmitic acid binding to RV1364c-PAS expressed in *E. coli* was confirmed by mass spectrometry, fluorescence binding assays revealed 10-fold tighter binding of palmitoleic (C16:1; 5 nM) and oleic acid (C18:0; 7.5 nM) to this protein. The functional relevance of fatty acids as a regulatory cofactor has not been established in this system. However, lipid signaling between host and pathogen is an important determinant of tuberculosis disease (51), and it is known that the transcription of *Mycobacterium bovis rv1364c* is upregulated in host macrophages and co-regulated with genes involved in fatty acid metabolism (69).

Conclusions

Since the initial classification of the PAS domain nearly 20 years ago, significant strides have been made in understanding function, structure, and regulation of PAS-containing proteins. The utility of PAS as a cytoplasmic and extracytoplasmic sensor is in part due to its capacity to bind a chemically diverse range of small molecule ligands. Indeed, the PAS structural scaffold has been selected to bind a broad range of molecules and ions including hemes, flavins, di- and tricarboxylic acids, amino acids, divalent metal cations, coumaric acid, and fatty acids. This list is by no means comprehensive, and is certain to expand. The remarkable evolutionary plasticity of the PAS domain has likely driven its expansion and diversification in signaling proteins across all kingdoms of life. In short, it seems that the PAS fold can be selected to perform many functions in a variety of protein structural contexts. In the context of the emerging field of synthetic biology, the extraordinary flexibility/modularity of sensory PAS domains has been exploited to engineer artificial proteins with new activities (reviewed in (80)). Future studies promise to expand our view of the range of functions/activities that can be regulated by PAS domains in both natural and synthetic systems.

Mini-glossary

LOV domain	a PAS subfamily, generally classified as a blue-light sensor
GGDEF/EAL	amino acid motifs found in enzymes that either synthesize the bacterial second messenger cyclic-di-GMP (nucleotide cyclase activity), or hydrolyze it (phosphodiesterase activity) respectively. GGDEF and EAL-containing domains are found separately or in tandem.

Response regulator	part of a two-component signal transduction system; typically a histidine kinase phosphorylates a response regulator protein, changing its functionality (promoting DNA binding, for example).
HisKA	a Pfam protein domain known to dimerize and accept a phosphate group, part of a full histidine kinase protein.
H_ATPase_c	a Pfam protein domain associated with ATP hydrolysis in histidine kinases, found C-terminal to the HisKA domain.

Important acronyms

PYP	**p**hotoactive **y**ellow **p**rotein
HK	**h**istidine **k**inase
MCP	**m**ethyl-accepting **c**hemotaxis **p**rotein
GAF	sensory domain first characterized in c**G**MP-dependent phosphodiesterase, **a**denylyl cyclases, and E. Coli **F**hlA
STAS	**s**ulfate **t**ransporter **a**nti-**s**igma factor
REC	**REC**eiver domain of a response regulator
HPt	**h**istidine-containing **p**hospho**t**ransfer domain
FMN/FAD	**f**lavin **m**ono**n**ucleotide / **f**lavin **a**denine **d**inucleotide
4-HCA	the chromophore **4**-**h**ydroxy**c**innamic **a**cid

Literature Cited

1. Akbar S, Gaidenko TA, Kang CM, O'Reilly M, Devine KM, Price CW. New family of regulators in the environmental signaling pathway which activates the general stress transcription factor sigma(B) of Bacillus subtilis. J Bacteriol. 2001; 183:1329–38. [PubMed: 11157946]
2. Amezcua CA, Harper SM, Rutter J, Gardner KH. Structure and interactions of PAS kinase N-terminal PAS domain: model for intramolecular kinase regulation. Structure. 2002; 10:1349–61. [PubMed: 12377121]
3. Anantharaman V, Koonin EV, Aravind L. Regulatory potential, phyletic distribution and evolution of ancient, intracellular small-molecule-binding domains. J Mol Biol. 2001; 307:1271–92. [PubMed: 11292341]
4. Anderson S, Crosson S, Moffat K. Short hydrogen bonds in photoactive yellow protein. Acta Crystallogr D Biol Crystallogr. 2004; 60:1008–16. [PubMed: 15159559]
5. Aravind L, Ponting CP. The GAF domain: an evolutionary link between diverse phototransducing proteins. Trends Biochem Sci. 1997; 22:458–9. [PubMed: 9433123]
6. Avila-Perez M, Hellingwerf KJ, Kort R. Blue light activates the sigmaBdependent stress response of Bacillus subtilis via YtvA. J Bacteriol. 2006; 188:6411–4. [PubMed: 16923909]
7. Baca M, Borgstahl GE, Boissinot M, Burke PM, Williams DR, et al. Complete chemical structure of photoactive yellow protein: novel thioester-linked 4- hydroxycinnamyl chromophore and photocycle chemistry. Biochemistry. 1994; 33:14369–77. [PubMed: 7981196]
8. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. GenBank. Nucleic Acids Res. 2008; 36:25–30.
9. Berleman JE, Hasselbring BM, Bauer CE. Hypercyst mutants in Rhodospirillum centenum identify regulatory loci involved in cyst cell differentiation. J Bacteriol. 2004; 186:5834–41. [PubMed: 15317789]

10. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. The Protein Data Bank. Nucleic Acids Res. 2000; 28:235–42. [PubMed: 10592235]
11. Bibikov SI, Barnes LA, Gitin Y, Parkinson JS. Domain organization and flavin adenine dinucleotide-binding determinants in the aerotaxis signal transducer Aer of Escherichia coli. Proc Natl Acad Sci U S A. 2000; 97:5830–5. [PubMed: 10811894]
12. Bibikov SI, Biran R, Rudd KE, Parkinson JS. A signal transducer for aerotaxis in Escherichia coli. J Bacteriol. 1997; 179:4075–9. [PubMed: 9190831]
13. Borgstahl GE, Williams DR, Getzoff ED. 1.4 A structure of photoactive yellow protein, a cytosolic photoreceptor: unusual fold, active site, and chromophore. Biochemistry. 1995; 34:6278–87. [PubMed: 7756254]
14. Bott M. Anaerobic citrate metabolism and its regulation in enterobacteria. Arch Microbiol. 1997; 167:78–88.
15. Catarino T, Pessanha M, De Candia AG, Gouveia Z, Fernandes AP, et al. Probing the chemotaxis periplasmic sensor domains from Geobacter sulfurreducens by combined resonance Raman and molecular dynamic approaches: NO and CO sensing. J Phys Chem B. 2010; 114:11251–60. [PubMed: 20690670]
16. Chang C, Tesar C, Gu M, Babnigg G, Joachimiak A, et al. Extracytoplasmic PASlike domains are common in signal transduction proteins. J Bacteriol. 2010; 192:1156–9. [PubMed: 20008068]
17. Chen M, Tao Y, Lim J, Shaw A, Chory J. Regulation of phytochrome B nuclear localization through light-dependent unmasking of nuclear-localization signals. Curr Biol. 2005; 15:637–42. [PubMed: 15823535]
18. Cheung J, Bingman CA, Reyngold M, Hendrickson WA, Waldburger CD. Crystal structure of a functional dimer of the PhoQ sensor domain. J Biol Chem. 2008; 283:13762–70. [PubMed: 18348979]
19. Cheung J, Hendrickson WA. Sensor domains of two-component regulatory systems. Curr Opin Microbiol. 2010; 13:116–23. [PubMed: 20223701]
20. Cho US, Bader MW, Amaya MF, Daley ME, Klevit RE, et al. Metal Bridges between the PhoQ Sensor Domain and the Membrane Regulate Transmembrane Signaling. J Mol Biol. 2006; 356:1193–206. [PubMed: 16406409]
21. Christie JM, Reymond P, Powell GK, Bernasconi P, Raibekas AA, et al. Arabidopsis NPH1: a flavoprotein with the properties of a photoreceptor for phototropism. Science. 1998; 282:1698–701. [PubMed: 9831559]
22. Christie JM, Salomon M, Nozue K, Wada M, Briggs WR. LOV (light, oxygen, or voltage) domains of the blue-light photoreceptor phototropin (nph1): binding sites for the chromophore flavin mononucleotide. Proc Natl Acad Sci U S A. 1999; 96:8779–83. [PubMed: 10411952]
23. Crosson S, McGrath PT, Stephens C, McAdams HH, Shapiro L. Conserved modular design of an oxygen sensory/signaling network with species-specific output. Proc Natl Acad Sci U S A. 2005; 102:8018–23. [PubMed: 15911751]
24. Crosson S, Moffat K. Structure of a flavin-binding plant photoreceptor domain: insights into light-mediated signal transduction. Proc Natl Acad Sci U S A. 2001; 98:2995–3000. [PubMed: 11248020]
25. Crosson S, Rajagopal S, Moffat K. The LOV domain family: photoresponsive signaling modules coupled to diverse output domains. Biochemistry. 2003; 42:2–10. [PubMed: 12515534]
26. David M, Daveran ML, Batut J, Dedieu A, Domergue O, et al. Cascade regulation of nif gene expression in Rhizobium meliloti. Cell. 1988; 54:671–83. [PubMed: 2842062]
27. de Philip P, Batut J, Boistard P. Rhizobium meliloti FixL is an oxygen sensor and regulates R. meliloti nifA and fixK genes differently in Escherichia coli. J Bacteriol. 1990; 172:4255–62. [PubMed: 2115865]
28. Delgado-Nixon VM, Gonzalez G, yGonzalez MA. Dos, a heme-binding PAS protein from Escherichia coli, is a direct oxygen sensor. Biochemistry. 2000; 39:2685–91. [PubMed: 10704219]
29. Dimroth P, Schink B. Energy conservation in the decarboxylation of dicarboxylic acids by fermenting bacteria. Arch Microbiol. 1998; 170:69–77. [PubMed: 9683642]
30. Eargle J, Wright D, Luthey-Schulten Z. Multiple Alignment of protein structures and sequences for VMD. Bioinformatics. 2006; 22:504–6. [PubMed: 16339280]

31. Ernst RK, Guina T, Miller SI. Salmonella typhimurium outer membrane remodeling: role in resistance to host innate immunity. Microbes Infect. 2001; 3:1327–34. [PubMed: 11755422]
32. Fass E, Groisman EA. Control of Salmonella pathogenicity island-2 gene expression. Curr Opin Microbiol. 2009; 12:199–204. [PubMed: 19264535]
33. Felsenstein, J. PHYLIP (Phylogeny Inference Package) version 3.6. Department of Genome Sciences, University of Washington; Seattle: 2005.
34. Finn RD, Mistry J, Tate J, Coggill P, Heger A, et al. The Pfam protein families database. Nucleic Acids Res. 2010; 38:D211–22. [PubMed: 19920124]
35. Fischer HM. Genetic regulation of nitrogen fixation in rhizobia. Microbiol Rev. 1994; 58:352–86. [PubMed: 7968919]
36. Fitch WM, Margoliash E. Construction of phylogenetic trees. Science. 1967; 155:279–84. [PubMed: 5334057]
37. Galperin MY. Bacterial signal transduction network in a genomic perspective. Environ Microbiol. 2004; 6:552–67. [PubMed: 15142243]
38. Garcia-Vescovi E, Ayala YM, Di Cera E, Groisman EA. Characterization of the bacterial sensor protein PhoQ. Evidence for distinct binding sites for Mg2+ and Ca2+ J Biol Chem. 1997; 272:1440–3. [PubMed: 8999810]
39. Garcia-Vescovi E, Soncini FC, Groisman EA. Mg2+ as an extracellular signal: environmental regulation of Salmonella virulence. Cell. 1996; 84:165–74. [PubMed: 8548821]
40. Gilles-Gonzalez MA, Ditta GS, Helinski DR. A haemoprotein with kinase activity encoded by the oxygen sensor of Rhizobium meliloti. Nature. 1991; 350:170–2. [PubMed: 1848683]
41. Gilles-Gonzalez MA, Gonzalez G, Perutz MF, Kiger L, Marden MC, Poyart C. Heme-based sensors, exemplified by the kinase FixL, are a new class of heme protein with distinctive ligand binding and autoxidation. Biochemistry. 1994; 33:8067–73. [PubMed: 8025112]
42. Glekas GD, Foster RM, Cates JR, Estrella JA, Wawrzyniak MJ, et al. A PAS domain binds asparagine in the chemotaxis receptor McpB in Bacillus subtilis. J Biol Chem. 2010; 285:1870–8. [PubMed: 19864420]
43. Golby P, Davies S, Kelly DJ, Guest JR, Andrews SC. Identification and characterization of a two-component sensor-kinase and response-regulator system (DcuS-DcuR) controlling gene expression in response to C4-dicarboxylates in Escherichia coli. J Bacteriol. 1999; 181:1238–48. [PubMed: 9973351]
44. Gong W, Hao B, Mansy SS, Gonzalez G, Gilles-Gonzalez MA, Chan MK. Structure of a biological oxygen sensor: a new mechanism for heme-driven signal transduction. Proc Natl Acad Sci U S A. 1998; 95:15177–82. [PubMed: 9860942]
45. Gonzalez LC, Weis WI, Scheller RH. A novel snare N-terminal domain revealed by the crystal structure of Sec22b. J Biol Chem. 2001; 276:24203–11. [PubMed: 11309394]
46. Hefti MH, Francoijs K-J, de Vries SC, Dixon R, Vervoort J. The PAS fold. A redefinition of the PAS domain based upon structural prediction. Eur J Biochem. 2004; 271:1198–208. [PubMed: 15009198]
47. Herrmann S, Ma Q, Johnson MS, Repik AV, Taylor BL. PAS domain of the Aer redox sensor requires C-terminal residues for native-fold formation and flavin adenine dinucleotide binding. J Bacteriol. 2004; 186:6782–91. [PubMed: 15466030]
48. Ho YS, Burden LM, Hurley JH. Structure of the GAF domain, a ubiquitous signaling motif and a new class of cyclic GMP receptor. EMBO J. 2000; 19:5288–99. [PubMed: 11032796]
49. Hoff WD, Dux P, Hard K, Devreese B, Nugteren-Roodzant IM, et al. Thiol esterlinked p-coumaric acid as a new photoactive prosthetic group in a protein with rhodopsin-like photochemistry. Biochemistry. 1994; 33:13959–62. [PubMed: 7947803]
50. Holm L, Ouzounis C, Sander C, Tuparev G, Vriend G. A database of protein structure families with common folding motifs. Protein Sci. 1992; 1:1691–8. [PubMed: 1304898]
51. Houben EN, Nguyen L, Pieters J. Interaction of pathogenic mycobacteria with the host immune system. Curr Opin Microbiol. 2006; 9:76–85. [PubMed: 16406837]
52. Huala E, Oeller PW, Liscum E, Han IS, Larsen E, Briggs WR. Arabidopsis NPH1: a protein kinase with a putative redox-sensing domain. Science. 1997; 278:2120–3. [PubMed: 9405347]

53. Huang ZJ, Edery I, Rosbash M. PAS is a dimerization domain common to Drosophila period and several transcription factors. Nature. 1993; 364:259–62. [PubMed: 8391649]
54. Jaiswal RK, Manjeera G, Gopal B. Role of a PAS sensor domain in the Mycobacterium tuberculosis transcription regulator Rv1364c. Biochem Biophys Res Commun. 2010; 398:342–9. [PubMed: 20541534]
55. Janausch IG, Zientz E, Tran QH, Kroger A, Unden G. C4-dicarboxylate carriers and sensors in bacteria. Biochim Biophys Acta. 2002; 1553:39–56. [PubMed: 11803016]
56. Jiang Z, Swem LR, Rushing BG, Devanathan S, Tollin G, Bauer CE. Bacterial photoreceptor with similarity to photoactive yellow protein and plant phytochromes. Science. 1999; 285:406–9. [PubMed: 10411503]
57. Kelley LA, Sternberg MJ. Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc. 2009; 4:363–71. [PubMed: 19247286]
58. Key J, Hefti M, Purcell EB, Moffat K. Structure of the redox sensor domain of Azotobacter vinelandii NifL at atomic resolution: signaling, dimerization, and mechanism. Biochemistry. 2007; 46:3614–23. [PubMed: 17319691]
59. Key J, Moffat K. Crystal structures of deoxy and CO-bound bjFixLH reveal details of ligand recognition and signaling. Biochemistry. 2005; 44:4627–35. [PubMed: 15779889]
60. Key J, Scheuermann TH, Anderson PC, Daggett V, Gardner KH. Principles of ligand binding within a completely buried cavity in HIF2alpha PAS-B. J Am Chem Soc. 2009; 131:17647–54. [PubMed: 19950993]
61. King-Scott J, Konarev PV, Panjikar S, Jordanova R, Svergun DI, Tucker PA. Structural characterization of the multi-domain regulatory protein Rv1364c from Mycobacterium tuberculosis. Structure. 2010 in press.
62. Kramer J, Fischer JD, Zientz E, Vijayan V, Griesinger C, et al. Citrate sensing by the C4-dicarboxylate/citrate sensor kinase DcuS of Escherichia coli: binding site and conversion of DcuS to a C4-dicarboxylate- or citrate-specific sensor. J Bacteriol. 2007; 189:4290–8. [PubMed: 17416661]
63. Krauss U, Minh BQ, Losi A, Gärtner W, Eggert T, et al. Distribution and phylogeny of light-oxygen-voltage-blue-light-signaling proteins in the three kingdoms of life. J Bacteriol. 2009; 191:7234–42. [PubMed: 19783626]
64. Krell T, Lacal J, Busch A, Silva-Jimenez H, Guazzaroni M-E, Ramos JL. Bacterial sensor kinases: diversity in the recognition of environmental signals. Annu Rev Microbiol. 2010; 64:539–59. [PubMed: 20825354]
65. Kreutel S, Kuhn A, Kiefer D. The photosensor protein Ppr of Rhodocista centenaria is linked to the chemotaxis signalling pathway. BMC Microbiol. 2010; 10:281. [PubMed: 21062468]
66. Kurokawa H, Lee D-S, Watanabe M, Sagami I, Mikami B, et al. A redoxcontrolled molecular switch revealed by the crystal structure of a bacterial heme PAS sensor. J Biol Chem. 2004; 279:20186–93. [PubMed: 14982921]
67. Lacal J, Garcia-Fontana C, Munoz-Martinez F, Ramos J-L, Krell T. Sensing of environmental signals: classification of chemoreceptors according to the size of their ligand binding regions. Environ Microbiol. 2010
68. Lee J, Tomchick DR, Brautigam CA, Machius M, Kort R, et al. Changes at the KinA PAS-A dimerization interface influence histidine kinase function. Biochemistry. 2008; 47:4051–64. [PubMed: 18324779]
69. Li MS, Waddell SJ, Monahan IM, Mangan JA, Martin SL, et al. Increased transcription of a potential sigma factor regulatory gene Rv1364c in Mycobacterium bovis BCG while residing in macrophages indicates use of alternative promoters. FEMS Microbiol Lett. 2004; 233:333–9. [PubMed: 15063504]
70. Little R, Martinez-Argudo I, Dixon R. Role of the central region of NifL in conformational switches that regulate nitrogen fixation. Biochem Soc Trans. 2006; 34:162–4. [PubMed: 16417511]
71. Londer YY, Dementieva IS, D'Ausilio CA, Pokkuluri PR, Schiffer M. Characterization of a c-type heme-containing PAS sensor domain from Geobacter sulfurreducens representing a novel family

of periplasmic sensors in Geobacteraceae and other bacteria. FEMS Microbiol Lett. 2006; 258:173–81. [PubMed: 16640569]

72. Losi A, Gärtner W. Bacterial bilin- and flavin-binding photoreceptors. Photochem Photobiol Sci. 2008; 7:1168–78. [PubMed: 18846280]
73. Ma X, Sayed N, Baskaran P, Beuve A, van den Akker F. PAS-mediated dimerization of soluble guanylyl cyclase revealed by signal transduction histidine kinase domain crystal structure. J Biol Chem. 2008; 283:1167–78. [PubMed: 18006497]
74. Marchler-Bauer A, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, et al. CDD: specific functional annotation with the Conserved Domain Database. Nucleic Acids Res. 2009; 37:D205–10. [PubMed: 18984618]
75. McIntosh BE, Hogenesch JB, Bradfield CA. Mammalian Per-Arnt-Sim proteins in environmental adaptation. Annu Rev Physiol. 2010; 72:625–45. [PubMed: 20148691]
76. Meyer TE. Isolation and characterization of soluble cytochromes, ferredoxins and other chromophoric proteins from the halophilic phototrophic bacterium Ectothiorhodospira halophila. Biochim Biophys Acta. 1985; 806:175–83. [PubMed: 2981543]
77. Minasov G, Padavattan S, Shuvalova L, Brunzelle JS, Miller DJ, et al. Crystal structures of YkuI and its complex with second messenger cyclic Di-GMP suggest catalytic mechanism of phosphodiester bond cleavage by EAL domains. J Biol Chem. 2009; 284:13174–84. [PubMed: 19244251]
78. Miyatake H, Mukai M, Park SY, Adachi S, Tamura K, et al. Sensory mechanism of oxygen sensor FixL from Rhizobium meliloti: crystallographic, mutagenesis and resonance Raman spectroscopic studies. J Mol Biol. 2000; 301:415–31. [PubMed: 10926518]
79. Möglich A, Ayers RA, Moffat K. Structure and signaling mechanism of Per-ARNT-Sim domains. Structure. 2009; 17:1282–94. [PubMed: 19836329]
80. Möglich A, Moffat K. Engineered photoreceptors as novel optogenetic tools. Photochem Photobiol Sci. 2010; 9:1286–300. [PubMed: 20835487]
81. Neiditch MB, Federle MJ, Pompeani AJ, Kelly RC, Swem DL, et al. Ligandinduced asymmetry in histidine sensor kinase complex regulates quorum sensing. Cell. 2006; 126:1095–108. [PubMed: 16990134]
82. Oka Y, Matsushita T, Mochizuki N, Quail PH, Nagatani A. Mutant screen distinguishes between residues necessary for light-signal perception and signal transfer by phytochrome B. PLoS Genet. 2008; 4
83. Pappalardo L, Janausch IG, Vijayan V, Zientz E, Junker J, 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; 278:39185–8. [PubMed: 12907689]
84. Park H, Suquet C, Satterlee JD, Kang C. Insights into signal transduction involving PAS domain oxygen-sensing heme proteins from the X-ray crystal structure of Escherichia coli Dos heme domain (Ec DosH). Biochemistry. 2004; 43:2738–46. [PubMed: 15005609]
85. Parkinson JS. Signal transduction schemes of bacteria. Cell. 1993; 73:857–71. [PubMed: 8098993]
86. Pellequer JL, Wager-Smith KA, Kay SA, Getzoff ED. Photoactive yellow protein: a structural prototype for the three-dimensional fold of the PAS domain superfamily. Proc Natl Acad Sci U S A. 1998; 95:5884–90. [PubMed: 9600888]
87. Philip AF, Kumauchi M, Hoff WD. Robustness and evolvability in the functional anatomy of a PER-ARNT-SIM (PAS) domain. Proc Natl Acad Sci U S A. 2010; 107:17986–91. [PubMed: 20889915]
88. Pokkuluri PR, Pessanha M, Londer YY, Wood SJ, Duke NE, et al. Structures and solution properties of two novel periplasmic sensor domains with c-type heme from chemotaxis proteins of Geobacter sulfurreducens: implications for signal transduction. J Mol Biol. 2008; 377:1498–517. [PubMed: 18329666]
89. Prost LR, Miller SI. The Salmonellae PhoQ sensor: mechanisms of detection of phagosome signals. Cell Microbiol. 2008; 10:576–82. [PubMed: 18182085]
90. Purcell EB, McDonald CA, Palfey BA, Crosson S. An analysis of the solution structure and signaling mechanism of LovK, a sensor histidine kinase integrating light and redox signals. Biochemistry. 2010; 49:6761–70. [PubMed: 20593779]

91. Purcell EB, Siegal-Gaskins D, Rawling DC, Fiebig A, Crosson S. A photosensory two-component system regulates bacterial cell attachment. Proc Natl Acad Sci U S A. 2007; 104:18241–6. [PubMed: 17986614]
92. Qi Y, Rao F, Luo Z, Liang Z-X. A flavin cofactor-binding PAS domain regulates cdi-GMP synthesis in AxDGC2 from Acetobacter xylinum. Biochemistry. 2009; 48:10275–85. [PubMed: 19785462]
93. Rebbapragada A, Johnson MS, Harding GP, Zuccarelli AJ, Fletcher HM, et al. The Aer protein and the serine chemoreceptor Tsr independently sense intracellular energy levels and transduce oxygen, redox, and energy signals for Escherichia coli behavior. Proc Natl Acad Sci U S A. 1997; 94:10541–6. [PubMed: 9380671]
94. Reinelt S, Hofmann E, Gerharz T, Bott M, Madden DR. The structure of the periplasmic ligand-binding domain of the sensor kinase CitA reveals the first extracellular PAS domain. J Biol Chem. 2003; 278:39189–96. [PubMed: 12867417]
95. Rey FE, Harwood CS. FixK, a global regulator of microaerobic growth, controls photosynthesis in Rhodopseudomonas palustris. Mol Microbiol. 2010
96. Russell RB, Breed J, Barton GJ. Conservation analysis and structure prediction of the SH2 family of phosphotyrosine binding domains. FEBS Lett. 1992; 304:15–20. [PubMed: 1377638]
97. Salomon M, Christie JM, Knieb E, Lempert U, Briggs WR. Photochemical and mutational analysis of the FMN-binding domains of the plant blue light receptor, phototropin. Biochemistry. 2000; 39:9401–10. [PubMed: 10924135]
98. Scheuermann TH, Tomchick DR, Machius M, Guo Y, Bruick RK, Gardner KH. Artificial ligand binding within the HIF2alpha PAS-B domain of the HIF2 transcription factor. Proc Natl Acad Sci U S A. 2009; 106:450–5. [PubMed: 19129502]
99. Schrodinger LLC. The PyMOL Molecular Graphics System. Version 1.3r1. 2010
100. Soderback E, Reyes-Ramirez F, Eydmann T, Austin S, Hill S, Dixon R. The redox- and fixed nitrogen-responsive regulatory protein NIFL from Azotobacter vinelandii comprises discrete flavin and nucleotide-binding domains. Mol Microbiol. 1998; 28:179–92. [PubMed: 9593306]
101. Sprenger WW, Hoff WD, Armitage JP, Hellingwerf KJ. The eubacterium Ectothiorhodospira halophila is negatively phototactic, with a wavelength dependence that fits the absorption spectrum of the photoactive yellow protein. J Bacteriol. 1993; 175:3096–104. [PubMed: 8491725]
102. Streeter JG. Recent developments in carbon transport and metabolism in symbiotic systems. Symbiosis. 1995; 19
103. Swartz TE, Tseng T-S, Frederickson MA, Paris G, Comerci DJ, et al. Blue-lightactivated histidine kinases: two-component sensors in bacteria. Science. 2007; 317:1090–3. [PubMed: 17717187]
104. Tanaka A, Takahashi H, Shimizu T. Critical role of the heme axial ligand, Met95, in locking catalysis of the phosphodiesterase from Escherichia coli (Ec DOS) toward Cyclic diGMP. J Biol Chem. 2007; 282:21301–7. [PubMed: 17535805]
105. Taylor BL, Zhulin IB. PAS domains: internal sensors of oxygen, redox potential, and light. Microbiol Mol Biol Rev. 1999; 63:479–506. [PubMed: 10357859]
106. Taylor BL, Zhulin IB, Johnson MS. Aerotaxis and other energy-sensing behavior in bacteria. Annu Rev Microbiol. 1999; 53:103–28. [PubMed: 10547687]
107. Tuckerman JR, Gonzalez G, Sousa EH, Wan X, Saito JA, et al. An oxygen-sensing diguanylate cyclase and phosphodiesterase couple for c-di-GMP control. Biochemistry. 2009; 48:9764–74. [PubMed: 19764732]
108. Ukaegbu UE, Henery S, Rosenzweig AC. Biochemical characterization of MmoS, a sensor protein involved in copper-dependent regulation of soluble methane monooxygenase. Biochemistry. 2006; 45:10191–8. [PubMed: 16922494]
109. Ukaegbu UE, Rosenzweig AC. Structure of the redox sensor domain of Methylococcus capsulatus (Bath) MmoS. Biochemistry. 2009; 48:2207–15. [PubMed: 19271777]
110. Ulrich LE, Koonin EV, Zhulin IB. One-component systems dominate signal transduction in prokaryotes. Trends Microbiol. 2005; 13:52–6. [PubMed: 15680762]

111. van der Horst MA, Key J, Hellingwerf KJ. Photosensing in chemotrophic, nonphototrophic bacteria: let there be light sensing too. Trends Microbiol. 2007; 15:554–62. [PubMed: 18024131]

112. Vescovi EG, Ayala YM, Di Cera E, Groisman EA. Characterization of the bacterial sensor protein PhoQ. Evidence for distinct binding sites for Mg2+ and Ca2+ J Biol Chem. 1997; 272:1440–3. [PubMed: 8999810]

113. Vogt G, Etzold T, Argos P. An Assessment of Amino Acid Exchange Matrices in Aligning Protein Sequences: The Twilight Zone Revisited. J Mol Biol. 1995; 249:816–31. [PubMed: 7602593]

114. Vujicic-Zagar A, Dulermo R, Le Gorrec M, Vannier F, Servant P, et al. Crystal structure of the IrrE protein, a central regulator of DNA damage repair in deinococcaceae. J Mol Biol. 2009; 386:704–16. [PubMed: 19150362]

115. Xie Z, Ulrich LE, Zhulin IB, Alexandre G. PAS domain containing chemoreceptor couples dynamic changes in metabolism with chemotaxis. Proc Natl Acad Sci U S A. 2010; 107:2235–40. [PubMed: 20133866]

116. Yoshioka S, Kobayashi K, Yoshimura H, Uchida T, Kitagawa T, Aono S. Biophysical properties of a c-type heme in chemotaxis signal transducer protein DcrA. Biochemistry. 2005; 44:15406–13. [PubMed: 16285745]

117. Zhou Y-F, Nan B, Nan J, Ma Q, Panjikar S, et al. C4-dicarboxylates sensing mechanism revealed by the crystal structures of DctB sensor domain. J Mol Biol. 2008; 383:49–61. [PubMed: 18725229]

Summary list

- PAS domains are ubiquitous in signal transduction proteins and often serve as direct sensors internal or external stimuli.
- These domains have been selected to sense small molecules, ions, gases, light, and redox state; we expect new PAS ligands will be discovered in the future.
- Structurally, PAS domains form clades based on their cellular location: intra- or extracytoplasmic, as well as by the class of ligand they bind.
- Increasing numbers of extracytoplasmic PAS domains have been characterized, and they are likely underpredicted in microbial genomes.
- Though PAS domains are highly divergent at the primary sequence level, detailed comparison of novel PAS domains to well-characterized ones has been a successful strategy for matching sensors to ligands.

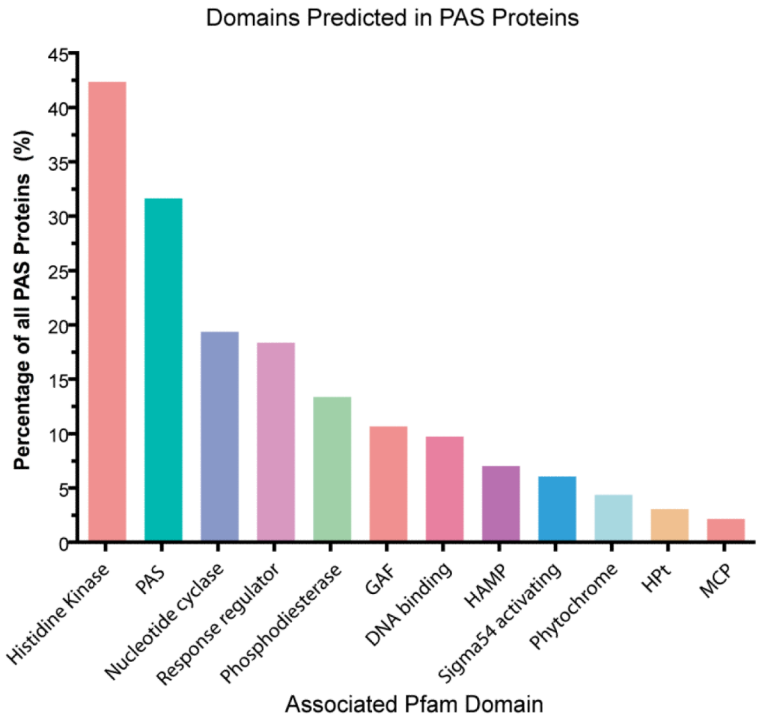

Figure 1.

Predicted Domains contained in PAS Proteins. We calculated the proportions of all PAS proteins (annotated by one of the seven PAS Clan CL0183 HMMs in Pfam 24.0) containing other Pfam-annotated domains (34). For the special case of the PAS domain itself, we calculate the proportion of PAS proteins containing two or more PAS domains. Counts for domains with similar functions were combined, and the highest scoring twelve domain types are presented here. Counts for each individual domain are tabulated in Supplementary Table 1.

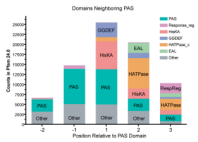

Figure 2.

Domains that are neighbors of PAS, by position. For each individual PAS domain annotated in Pfam (34), we identified other annotated domains at its N- and C-terminus in the same protein. The PAS domain occupies position zero, the −1 position represents the nearest domain annotated N-terminal to the PAS domain, and the +1 position represents the nearest domain annotated C-terminal. Subsequent positions are separated from the PAS domain in question by intervening domains. A full count of all domains neighboring PAS domains by position is found in Supplementary Table 2.

Figure 3.
Structural relatedness between PAS and PAS-like domains. PDB coordinates for representatives of what we believe to be all known PAS structures at the time of writing were trimmed to include only the region between the beginning of the first β-strand (Aβ) and the end of the final β-strand (Iβ); cofactors, ligands, and flanking sequences were discarded. The aligned region is illustrated in the secondary structure diagram, in which arrows represent β-strands and boxes represent α-helices present in most structures. We note that some sequences corresponding to the included structures are annotated as Cache domains in Genbank (8) or unrecognized by current HMM-based annotation; others are erroneously annotated as possessing a "heme pocket." Generally, the 'A' chains of multi-chain files were used unless other chains were more fully resolved; for structures in which multiple PAS domains are present on a single chain, the N- and C-terminal domains are referred to as 'A' and 'B' respectively. All structures were further annotated according to their organism of origin, given name and/or predicted or known signaling output domain. Structures were oriented in PyMOL (99), then multiply aligned using the Structural Alignment of Multiple Proteins (STAMP) algorithm (96) with the MultiSeq extension of the VMD software package (30). For all structures for which reliable multiple alignment was obtained, a distance matrix of RMSD values was exported to FITCH (Phylip) for clustering by the Fitch-Margoliash method (33; 36). Black text indicates cytoplasmic localization, blue indicates extracytoplasmic localization, and purple denotes cytoplasmic PAS-like structures included in the alignment. Physiological ligands and cofactors, but not fortuitous ligands of crystallization, are indicated on the right.

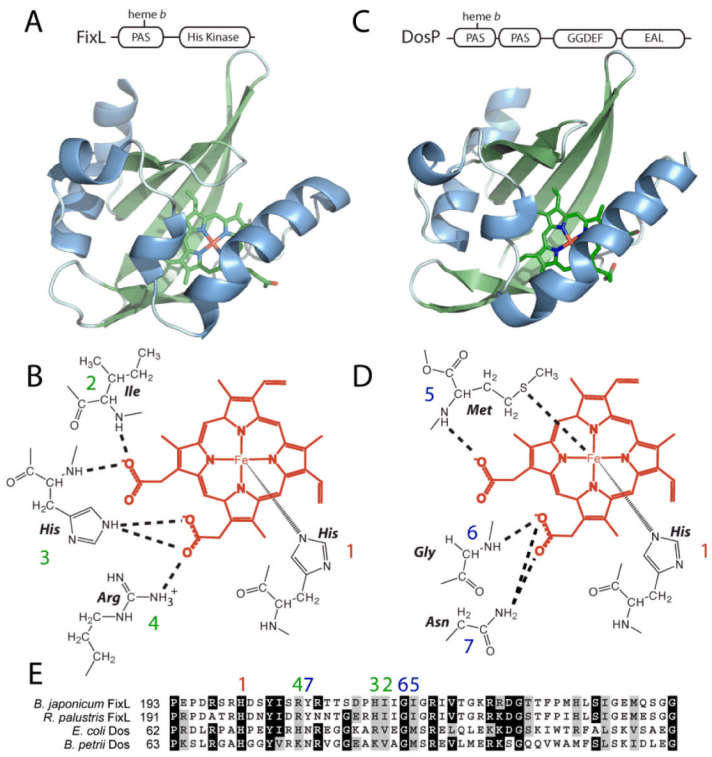

Figure 4.

Structure and ligand binding in the heme *b* PAS sensor domains of FixL and Dos. (A) Overall domain architecture and ribbon rendering of the *B. japonicum* FixL PAS domain bound to heme *b* (PDB ID: 1DRM). β-strands are rendered in light green, α-helices in blue. Atoms of the heme cofactor bound to the PAS domain are colored by element: carbon-green; nitrogen-blue; oxygen-red; iron-orange. (B) Drawing of protein interactions with the heme *b* cofactor (in red) in FixL-PAS marked with dotted lines or wedge lines. Interacting residues are numbered on the line drawing and sequence alignment with red numbers indicating conserved interactions between FixL and *E. coli* Dos, and green numbers indicating heme interactions that are specific to FixL-PAS. (C) Overall domain architecture and ribbon rendering of *E. coli* Dos PAS domain bound to heme *b* (PDB ID: 1S66) colored as in panel A. (D) Drawing of side chain and backbone interactions with the heme *b* cofactor (in red) in Dos-PAS marked with dotted lines or wedge lines. Blue numbers indicate heme interactions specific to Dos-PAS. (E) Clustal sequence alignment of FixL and Dos heme *b* PAS domains with the position of interacting residues numbered as described above. FixL of *Rhodopseudomonas palustris* (gb|ACF03217) and the DosP-like protein of *Bordatella petrii* (emb|CAP41703) are shown for comparison.

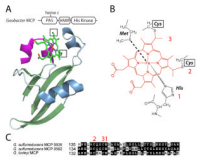

Figure 5.
Structure and ligand binding in the heme *c* PAS sensor domains of chemoreceptor proteins. (A) Overall domain architecture and ribbon rendering of *G. sulfurreducens* MCP GSU0935 bound to heme *c* (PDB ID: 3B42). β-strands are rendered in light green, α-helices in blue. Atoms of the heme cofactor bound to the PAS domain are colored by element: carbon-green; nitrogen-blue; oxygen-red; iron-orange. (B) Drawing of protein interactions with the heme *c* cofactor (in red) in GSU0935-PAS marked with dotted lines or wedge lines. Interacting residues are numbered on the line drawing and sequence alignment with red numbers indicating conserved interactions between heme and MCPs of *G. sulfurreducens*. (C) Clustal sequence alignment of chemoreceptor heme *c* PAS domains with the position of interacting residues numbered as described above.

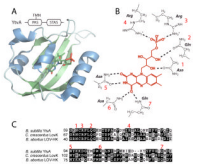

Figure 6.

Structure and flavin binding in the class of PAS blue light photosensors known as LOV domains. (A) Overall domain architecture and ribbon rendering of *B. subtilis* YtvA bound to flavin mononucleotide (FMN) (PDB ID: 2PR5). β-strands are rendered in light green, α-helices in blue. Atoms of the flavin cofactor bound to the PAS domain are colored by element: carbon-green; nitrogen-blue; oxygen-red. (B) Drawing of side chain interactions with the flavin cofactor (in red) in LOV domains. Interacting residues are numbered on the line drawing and sequence alignment with red numbers indicating conserved interactions between protein and the flavin cofactor. (C) Clustal sequence alignment of YtvA and two LOV histidine kinases of *C. crescentus* and *B. abortus*, with the position of conserved interacting residues numbered as described above.

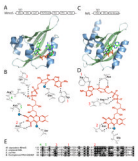

Figure 7.
Structure and ligand binding in flavin adenine dinucleotide (FAD)-binding PAS redox sensor domains of MmoS and NifL. (A) Overall domain architecture and ribbon rendering of the *M. capsulatus* MmoS PAS1 domain bound to FAD (PDB ID: 3EWK). β-strands are rendered in light green, α-helices in blue. Atoms of the flavin cofactor bound to the PAS1 domain are colored by element: carbon-green; nitrogen-blue; oxygen-red. (B) Drawing of side chain and backbone interactions with the flavin cofactor (in red) in MmoS-PAS1 marked with dotted lines. Bridging water molecules are shown as blue circles. Interacting residues are numbered on the line drawing and sequence alignment with red numbers indicating conserved interactions between MmoS and *A. vinelandii* NifL-PAS1, and green numbers indicating flavin interactions that are specific to MmoS-PAS1. (C) Overall domain architecture and ribbon rendering of *A. vinelandii* NifL-PAS1 domain bound to FAD (PDB ID: 2GJ3) colored as in panel A. (D) Drawing of side chain and backbone interactions with the FAD (in red) in NifL-PAS1 marked with dotted lines. Bridging water molecules are shown as blue circles. (E) Clustal sequence alignment of MmoS and NifL PAS1 domains with the position of FAD interacting residues numbered as described above. Related PAS domains of the *E. coli* aerotaxis sensor, Aer (NP_417543), and *B. thuringiensis* PAS-GGDEF (ZP_04100494) are shown for comparison.

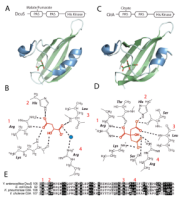

Figure 8.

Structure and ligand binding in the related di-/tricarboxylate-binding PAS domains of the DcuS and CitA sensor histidine kinases. (A) Overall domain architecture and ribbon rendering of *E. coli* DcuS PAS1 domain bound to a malate ion (PDB ID: 3BY8). β-strands are rendered in light green, α-helices in blue. Atoms of the malate cofactor bound to the PAS1 domain are colored by element: carbon-green; oxygen-red. (B) Drawing of side chain and backbone interactions of DcuS-PAS1 with malate (in red) marked with dotted lines. Interacting residues are numbered on the line drawing and sequence alignment with red numbers indicating conserved interactions between DcuS and *Klebsiella pneumoniae* CitA-PAS1. (C) Overall domain architecture and ribbon rendering of *K. pneumoniae* CitA-PAS1 domain bound to a citrate ion (PDB ID: 1P0Z) colored as in panel A. (D) Drawing of side chain and backbone interactions with citrate (in red) in CitA-PAS1 marked with dotted lines. (E) Clustal sequence alignment of DcuS and CitA PAS1 domains with the position of ligand interacting residues numbered as described above. Related PAS domains of *Yersinia enterocolitica*, DcuS (YE2505), and *Vibrio cholerae* CitA (VC0791) are shown for comparison.

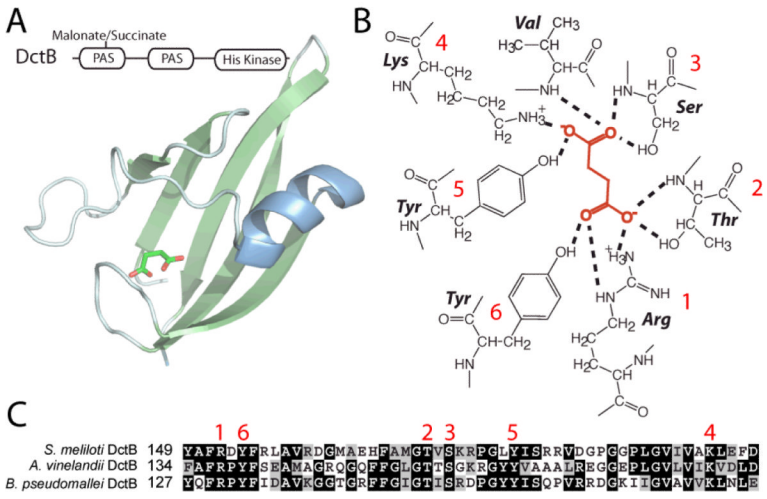

Figure 9.

Structure and ligand binding in the dicarboxylate-binding PAS1 domain of DctB. (A) Overall domain architecture and ribbon rendering of *Sinorhizobium meliloti* DctB PAS1 domain bound to a succinate ion (PDB ID: 3E4O). β-strands are rendered in light green, α-helices in blue. Atoms of the succinate cofactor bound to the PAS1 domain are colored by element: carbon-green; oxygen-red. (B) Drawing of side chain and backbone interactions of DctB-PAS1 with succinate (in red) marked with dotted lines. Interacting residues are numbered on the line drawing and sequence alignment with red numbers indicating conserved interactions with DctB orthologs of *A. vinelandii* and *Burkholderia pseudomallei*. (C) Clustal sequence alignment of DctB PAS1 domains with the position of ligand interacting residues numbered as described above. Related PAS domains of *A. vinelandii*, DctB, and *Burkholderia pseudomallei* DctB are shown for comparison.

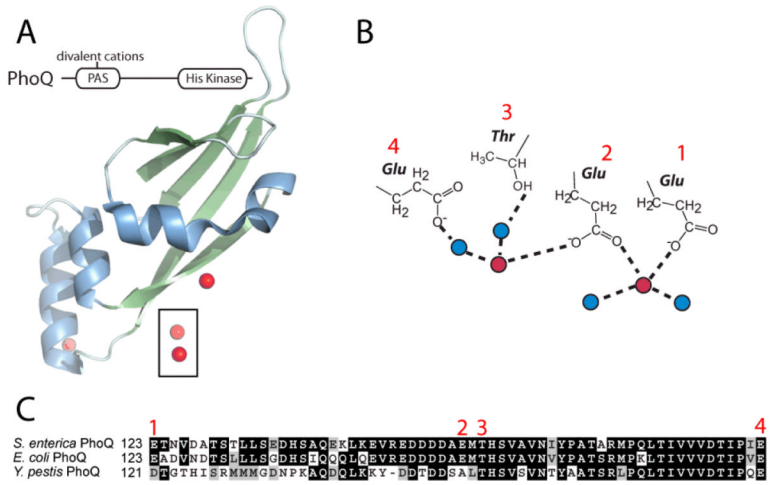

Figure 10.

Structure and metal binding in the PAS domain of PhoQ. (A) Overall domain architecture and ribbon rendering of *Salmonella typhimurium* PhoQ PAS domain bound to Ca^{2+} ions (PDB ID: 1YAX). β-strands are rendered in light green, α-helices in blue; calcium ions colored in red. (B) Drawing of protein interactions of PhoQ-PAS with calcium cations (in red) marked with dotted lines. Ion interactions pictured are boxed in panel A. Bridging water molecules are colored as blue circles. Interacting residues are numbered on the line drawing and sequence alignment with red numbers indicating conserved interactions with PhoQ orthologs of *E. coli* and *Yersinia pestis*. (C) Clustal sequence alignment of PhoQ PAS domains with the position of ligand interacting residues numbered as described above. Related PAS domains of *E. coli*, PhoQ, and *Y. pestis* PhoQ are shown for comparison.

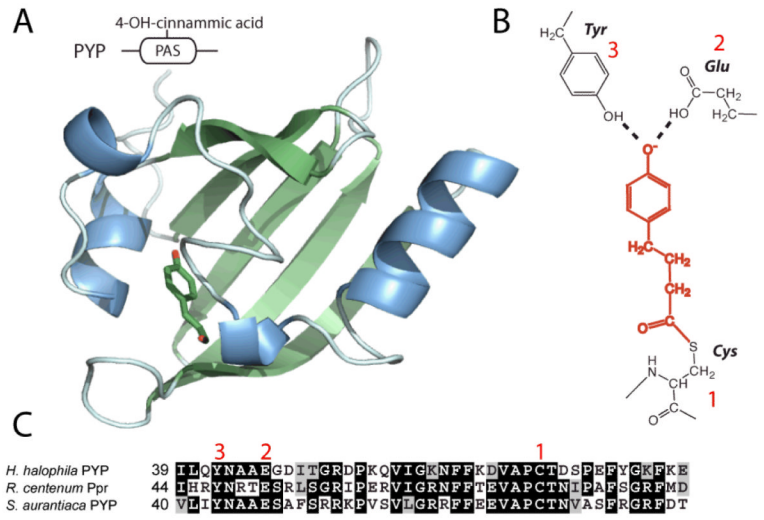

Figure 11.

Structure and ligand binding in the PAS domain, photoactive yellow protein (PYP). (A) Overall domain architecture and ribbon rendering of *Halorhodospira halophila* PYP bound to 4-hydroxycinnamic acid (4-HCA) (PDB ID: 2PHY). β-strands are rendered in light green, α-helices in blue. Atoms of the 4-HCA cofactor bound to PYP are colored by element: carbon-green; oxygen-red. (B) Drawing of side chains of PYP that form the covalent linkage and direct polar interactions with 4-HCA (in red). Interacting residues are numbered on the line drawing and sequence alignment with red numbers indicating conserved interactions across related PYP domains of *Stigmatella aurantiaca* and *Rhodospirillum centenum*. (C) Clustal sequence alignment of PYP domains with the position of ligand interacting residues numbered as described above. Related PYPs of *S. aurantiaca* and *R. centenum* are shown for comparison.