Phylogenetic Analysis with Prediction of Cofactor or Ligand Binding for *Pseudomonas aeruginosa* PAS and Cache Domains

Andrew Hutchin, Charlotte Cordery, Martin A. Walsh, Jeremy S. Webb, Ivo Tews

Abstract

PAS domains are omnipresent building blocks of multidomain proteins in all domains of life. Bacteria possess a variety of PAS domains in intracellular proteins and the related Cache domains in periplasmic or extracellular proteins. PAS and Cache domains are predominant in sensory systems, often carry cofactors or bind ligands, and serve as dimerization domains in protein association. To aid our understanding of the wide distribution of these domains, we analyzed the proteome of the opportunistic human pathogen *Pseudomonas aeruginosa* PAO1 in silico. The ability of this bacterium to survive under different environmental conditions, to switch between planktonic and sessile/biofilm lifestyle, or to evade stresses, notably involves c-di-GMP regulatory proteins or depends on sensory pathways involving multidomain proteins that possess PAS or Cache domains. Maximum likelihood phylogeny was used to group PAS and Cache domains on the basis of amino acid sequence. Conservation of cofactor- or ligand-coordinating amino acids aided by structure-based comparison was used to inform function. The resulting classification presented here includes PAS domains that are candidate binders of carboxylic acids, amino acids, fatty acids, flavin adenine dinucleotide (FAD), 4-hydroxycinnamic acid, and heme. These predictions are put in context to previously described phenotypic data, often generated from deletion mutants. The analysis predicts novel functions for sensory proteins and sheds light on functional diversification in a large set of proteins with similar architecture.

Importance

To adjust to a variety of life conditions, bacteria typically use multidomain proteins, where the modular structure allows functional differentiation. Proteins responding to environmental cues and regulating physiological responses are found in chemotaxis pathways that respond to a wide range of stimuli to affect movement. Environmental cues also regulate intracellular levels of cyclic-di-GMP, a universal bacterial secondary messenger that is a key determinant of bacterial lifestyle and virulence. We study *Pseudomonas aeruginosa*, an organism known to colonize a broad range of environments that can switch lifestyle between the sessile biofilm and the planktonic swimming form. We have investigated the PAS and Cache domains, of which we identified 101 in 70 *Pseudomonas aeruginosa* PAO1 proteins, and have grouped these by phylogeny with domains of known structure. The resulting data set integrates sequence analysis and structure prediction to infer ligand or cofactor binding. With this data set, functional predictions for PAS and Cache domain-containing proteins are made.

Keywords

Cache domain, PAS domain, phylogeny, *Pseudomonas*, cofactors, phylogenetic analysis, sensory transduction processes
The Gram-negative bacterium *Pseudomonas aeruginosa* is capable of growth in a wide range of different conditions, including soil and coastal marine habitats or plant and animal tissues (1, 2). *P. aeruginosa* is also a significant opportunistic human pathogen recently described as a species urgently requiring development of novel antibiotics for treatment of disease due to the emergence of multidrug-resistant strains (3). *P. aeruginosa* is able to infect patients suffering from burns, immunosuppression, and cystic fibrosis (CF); reduced pulmonary function caused by chronic *P. aeruginosa* infection is the largest cause of mortality in cystic fibrosis patients (1, 2, 4).

Diversity in cultivation of habitats is likely underpinned by adaptation mechanisms of *P. aeruginosa* to alter phenotypic behavior. This marked pleiotropy identifies a broad array of environmental cues and a number of archetypal bacterial responses. These resulting bacterial responses might include movement away from or toward a specific chemical stimulus, also known as chemotaxis (5). Alteration of gene expression may also be directly induced by a stimulus, often as part of a two-component regulatory system (6). Finally, transition from a planktonic phenotype to a sessile biofilm lifestyle makes *P. aeruginosa* an important human pathogen causing chronic infection (7, 8). This transition, and with it, bacterial virulence, is critically regulated by intracellular c-di-GMP levels (9). *P. aeruginosa*, and particularly the reference strain PAO1, is an extensively studied model organism in biofilm formation (2, 10, 11).

Adaptive responses that require mechanisms of signal perception are often directly transmitted through sensory proteins. A classic, versatile, and very widespread protein architecture used in many sensory proteins is the Per-Arnt-Sim or PAS domain (12, 13). The domain was first identified to be conserved between the fly clock protein PERIOD, the vertebrate aryl hydrocarbon nuclear translocator (ARNT), and the fly developmental regulator single-minded (SIM) (14, 15), and PAS domains are found across all biological kingdoms. They are frequently found among bacterial sensory systems (15) and play crucial roles in environmental responses of *P. aeruginosa* (16, 17). They are also widespread in regulators of intracellular c-di-GMP, where they are suggested to play a role in regulation of virulence (18), as well as motility and biofilm phenotype (19).

Cache domains are the extracellular relatives of the intracellular PAS domains (20–22). Typically acting as signal receptors, they bind small ligands and propagate signals into the cell interior, suggested to be mediated by the C-terminal helix that crosses the membrane (21–24). They are often classified into sCache or dCache domains with one or two PAS-like domains, respectively (22). While Cache domains are the predominant superfamily of extracellular receptors in prokaryotes, they are also found as extracellular domains ubiquitous across all kingdoms (22).

Adaptation of PAS and Cache domains to a variety of signals is achieved by cofactor or ligand binding. Further, they can typically dimerize in response to physiological change and, in doing so, alter the activity of the effector outputs within PAS domain-containing proteins (14, 25). Many proteins contain several PAS domains that may be common in structure but different in function, and hence study of cofactor or ligand binding is a systems approach that is essential to determine how responses are regulated.

Here, we update the list of known PAS and Cache domains in *P. aeruginosa* PAO1, provide a phylogenetic analysis, and add functional insights using a sequence/structure-based approach, with prediction of cofactor or ligand binding. Building upon earlier identification through Hidden Markov Model analysis (22, 26), the analysis includes a total of 90 PAS domains, 2 sCache domains, and 9 dCache domains that were identified from a total of 70 genes. Phylogeny and physiology are put in context, as illustrated here for carboxylic acid-binding PAS domains.

**RESULTS**

**Selection of *P. aeruginosa* PAO1 PAS and Cache domains.** The PAS domain fold is a highly conserved yet versatile structure. Sequence identity of PAS domains is typically below 20 percent (14, 15, 27), making the identification of all PAS domains in an organism difficult. Hidden Markov Models (HMM) were used as a sensitive method for...
homology detection, and these methods are typically employed for cases with low sequence identity (28). Application of HMM methods through HMM-to-HMM comparison (implemented through HHblits (29)) has initially identified 106 domains from 70 PAO1 proteins (26), of which 18 form 9 dCache domains and 2 were later classified as sCache, thus leaving 86 bona fide PAS domains (22).

The search used here employed domain boundaries predicted from the earlier studies. Additionally, the 70 PAO1 proteins were queried with the SMART domain prediction server (30, 31). The SMART analysis uses different criteria for minimum sequence length compared with the more stringent HMM-to-HMM analysis (26). SMART identifies the shorter PAS and PAC sequence motifs separately (32) and was thus able to identify the N-terminal PAS motif for PA4021, PA4112, PA4147, and PA4197. Sequences were extended at the C terminus to a length of ~120 amino acids to facilitate further analysis. The final list of 101 PAS and Cache domains selected from this analysis is given in Table 1.

**Phylogenetic analysis and grouping of PAS and Cache domains.** We performed a grouping of sequences by maximum likelihood phylogeny to understand the relationships between sequences (see Materials and Methods). Neighborhood within this phylogenetic analysis might infer similar properties with regard to cofactor or ligand binding. The relationship between proteins grouped within the same clade can thus be used as an indicator toward a functional assignment of the individual domains, leading to experimentally testable hypotheses.

The sequence alignment was performed against a reference data set made up from PAS and Cache domains with known ligand or cofactor obtained from structural analysis. Table 2 indicates cofactor or ligand identified in these structures and size of the ligand- or cofactor-binding pocket. The reference data set also included 37 sequences for ligand- or cofactor-free PAS or Cache structures, selected based on physiological roles not requiring ligand or cofactor, e.g., mediating dimerization or downstream signaling in response to a conformational change in a multidomain protein or as a result of binding to another protein. This ties in with the observation that a significant number of PAS or Cache domains were reported as structures without an associated cofactor or ligand (15).

The maximum likelihood phylogeny analysis with a 100-replicate bootstrap consensus tree is shown in Fig. 1. For the PAS and Cache domains analyzed here, we found that maximum likelihood grouped PAS or Cache domains from the reference data set into clades of similar cofactor or ligand binding across the largest number of bootstrap replicates, in comparison to other ways of constructing phylogenetic trees (see supplemental material for further detail). The phylogenetic tree identifies a number of clades and groups PAO1 sequences together with structurally known PAS and Cache domains. The grouping is based solely on the phylogenetic analysis and is thus unbiased by ligand or cofactor binding or structural knowledge. A number of PAO1 PAS and Cache domains have been characterized previously with respect to ligand or cofactor binding, and the fact that these sequences cluster in the tree with the reference sequence from structures with similar ligand or cofactor validates the approach taken here.

Assignments were made based on the basis of clustering in more than 15 bootstrap replicates, as this threshold provides unambiguous clustering of the reference PAS and Cache domains in almost all cases while retaining the clustering of known homologues. For previously uncharacterized PAO1 PAS or Cache domains, inference suggests that grouping of *P. aeruginosa* sequences with structural representatives from the reference indicates similar ligand or cofactor binding. Alignments of individual clades are presented in the supplemental material, and we give a few examples in the following section.

**Inferences from example clades and grouping of PAS and Cache domains.** A prominent clade, marked with arrow 1 in Fig. 1 (alignment found in the supplemental material), places the PAO1 PA1336, PA5165, and PA5512 dCache domains with the reference structure sequences of the two DctB dCache domains of *Vibrio cholerae* and...
### TABLE 1  P. aeruginosa PAO1 proteins with PAS or Cache domains

| Gene  | Protein   | Domain boundary       | PAS1     | PAS2     | PAS3     | PAS4     |
|-------|-----------|-----------------------|----------|----------|----------|----------|
| PA0172| SiaA      | dCache 102–304        |          |          |          |          |
| PA0176| Aer2/TlpG/McpB | 166–287            |          |          |          |          |
| PA0285| 79–198    |                       |          |          |          |          |
| PA0290| 31–151    |                       |          |          |          |          |
| PA0338| 50–170    |                       |          |          |          |          |
| PA0464| CreC      | sCache35–179          |          |          |          |          |
| PA0533| 12–135    |                       | 137–255  | 265–379  |          |          |
| PA0575| 310–426   |                       | 438–550  | 562–675  | 682–797  |          |
| PA0600| AgtS      |                       | 323–436  | 446–568  |          |          |
| PA0847| 142–284   |                       | 444–560  |          |          |          |
| PA0861| RbdA      | 243–363               |          |          |          |          |
| PA0873| PhhR      | 82–187                |          |          |          |          |
| PA0928| GacS      | 43–161                |          |          |          |          |
| PA1098| FleS      | 74–164                |          |          |          |          |
| PA1120| TpbB/YfnN | 46–152                |          |          |          |          |
| PA1180| PhoQ      | 33–161                |          |          |          |          |
| PA1181| YegE      | 298–415               | 427–542  | 553–674  |          |          |
| PA1196| DdaR      | 20–132                |          |          |          |          |
| PA1243| 57–169    |                       | 343–456  |          |          |          |
| PA1261| IhpR      | 1–107                 |          |          |          |          |
| PA1336| AauS      | dCache51–346          | dCache51–346 |          |          |          |
| PA1347| 23–129    |                       |          |          |          |          |
| PA1423| BdiA      | 3–112                 | 116–234  |          |          |          |
| PA1438| MmnS      | 41–166                |          |          |          |          |
| PA1561| Aer/ TlpC | 8–121                 |          |          |          |          |
| PA1611| 38–169    |                       |          |          |          |          |
| PA1976| ErcS      | 97–207                | 226–338  | 339–454  |          |          |
| PA1992| ErcS      | 41–157                |          |          |          |          |
| PA2005| HbcR      | 17–123                |          |          |          |          |
| PA2072| 301–414   |                       |          |          |          |          |
| PA2177| 62–180    |                       | 190–308  |          |          |          |
| PA2449| 79–182    |                       |          |          |          |          |
| PA2480| 30–148    |                       |          |          |          |          |
| PA2524| CzcS      | 34–171                |          |          |          |          |
| PA2652| CtpM      | sCache42–198          |          |          |          |          |
| PA2654| TlpQ      | dCache50–346          | dCache50–346 |          |          |          |
| PA2824| SagS      | 56–169                |          |          |          |          |
| PA2870| 97–211    |                       | 241–348  |          |          |          |
| PA3044| RocS2     | 110–225               |          |          |          |          |
| PA3271| 636–751   |                       |          |          |          |          |
| PA3946| RocS1     | 573–687               |          |          |          |          |
| PA4021| EatR      | 80–185                | 225–344  |          |          |          |
| PA4036| 432–537   |                       |          |          |          |          |
| PA4112| 343–460   |                       | 491–614  | 626–744  |          |          |
| PA4117| BphP      | 23–123                |          |          |          |          |
| PA4147| AcoR      | 82–191                | 225–344  |          |          |          |
| PA4197| Bfs       | 158–265               | 266–383  | 389–504  |          |          |
| PA4290| 411–520   |                       |          |          |          |          |
| PA4293| PprA      | 303–421               | 431–549  | 560–675  |          |          |
| PA4307| PctC      | dCache34–275          | dCache34–275 |          |          |          |
| PA4309| PctA      | dCache35–273          | dCache35–273 |          |          |          |
| PA4310| PctB      | dCache35–274          | dCache35–274 |          |          |          |
| PA4398| 50–154    |                       | 286–395  |          |          |          |
| PA4546| PiIS      | 195–296               |          |          |          |          |
| PA4581| RtcR      | 52–165                |          |          |          |          |
| PA4601| MorA      | 290–411               | 582–705  | 717–845  | 825–967  |          |
| PA4633| PprA      | dCache51–346          | dCache51–346 |          |          |          |
| PA4725| CbrA      | 630–739               |          |          |          |          |
| PA4886| 69–166    |                       |          |          |          |          |
| PA4959| FimX      | 142–254               |          |          |          |          |

(Continued on next page)
**Sinorhizobium melliloti** in 29 out of 100 bootstrap replicates. While PA5165 has previously been identified as a DctB homologue within *P. aeruginosa* (33, 34), this clade gives new insight, as it implies coevolution with PA1336 and PA5512. It may further predict the potential for binding similar ligands in all five domains.

A close relationship is detected between PA5124 PAS1 and the PAS domain from *Escherichia coli* DhaR, identified in 93 out of 100 bootstrap replicates, marked with arrow 2 in Fig. 1. The *E. coli* DhaR protein is a regulator of transcription. The PAS domain of DhaR contains a very small cavity that precludes binding of larger organic cofactors (Table 2), and the PAS domain is instead thought to be involved in signal transmission through dimerization (35). Conformational changes of the entire protein would be induced by binding of a number of different known interaction partners (35). By inference, the PA5124 PAS1 domain may also not bind any cofactor. This finding may be surprising, as the two proteins have vastly different domain architectures: DhaR consists of a GAF domain, a PAS domain, and a C-terminal domain involved in interaction with σ70 (35), while PA5124 is predicted to have a single PAS domain, as well as a histidine kinase and an accompanying phospho-transfer domain known from two-component signaling pathways (6, 30, 31).

The phylogenetic tree in Fig. 1 shows many PAO1 PAS and Cache domains that do not cluster to the chosen reference data set, guided by choice of our reference cofactor- and ligand-binding domain data set. However, the phylogenetic analysis performed here still provides insight into the evolutional origin of several of these domains. An example cluster identified in 96% of replicates is marked with arrow 3 in Fig. 1 and contains the two PAS domains of PA1423 (BdlA) and the two PAS domains of PA1930 (McpS). The protein architectures are similar, as both proteins consist of two PAS domains N-terminal to a methyl-accepting chemotaxis domain. PA1423 and PA1930 are unique within PAO1, as they possess methyl-accepting chemotaxis domains shorter than those of any other chemoreceptors (36). The similar architecture suggests functional differentiation for these proteins.

**Assignment of cofactor or ligand binding based on sequence motif.** The ability of a PAS or Cache domain to bind cofactor or ligand should be reflected in conservation of cofactor- or ligand-interacting amino acids. PAS and Cache domains are structurally homologous, albeit with overall rather low sequence identity. To add additional information, the alignment can therefore be constrained by predicted secondary structure. Using the combined primary and predicted secondary information then gives sufficient confidence for modeling of the 3D localization of conserved cofactor- or ligand-coordinating amino acids.

We use this approach here to inspect the ligand- or cofactor-binding environment. The PAO1 test data set was aligned to the different cofactor- and ligand-binding subsets of the reference data set, using secondary structure constraints through use of PROMALS3D (37). Detection of conservation or conservative substitution of amino acids known to form sidechain-mediated interactions within the resulting alignments

| Gene     | Protein | Domain boundary |
|----------|---------|-----------------|
| PA4961   |         | 53–166          |
| PA4982   | AruS    | 288–388         |
| PA5017   | DipA    | 9–130           |
| PA5124   | NtrB    | 3–116           |
| PA5165   | DctB    | dCache44–291    |
| PA5361   | PhoR    | 101–201         |
| PA5442   |         | 275–393         |
| PA5484   | KinB    | 257–369         |
| PA5512   | MifS    | dCache31–298    |

Of the 70 genes listed, several encode more than one PAS domain. Domain boundaries were identified by HMM analysis in previous studies (22, 26) or with the SMART domain web server (30, 31).
*TABLE 2* The reference data set contains sequences from PAS or Cache domain structures, grouped by physiological cofactor or ligand and by protein and species name, as well as references to the structural database and literature.

| Cofactor or ligand | M, of cofactor or ligand (g/mol) | Protein | Organism | PDB | PAS domain boundary from PDB RCSB | Pocket MS vol (Å³) | Comment |
|--------------------|---------------------------------|---------|----------|-----|----------------------------------|-------------------|---------|
| 4’-Hydroxycinnamic acid | 164.16 | Ppr | *Rhodospirillum centenum* | 1MZU (70) | 25–129 | 397.0 | |
| Autoinducers | 124.14 | VqmA | *Vibrio cholerae* | 6DE (39) | 16–121 | 318.0 | |
| Aromatics | 92.14 | TodS | *Pseudomonas putida* | SHWV (71) | 5–133 | 216.8 | |
| FAD | 785.55 | MmoS (PAS A) | *Methyllococcus capsulatus* | 3EWK (72) | 1–100 | 763.2 | |
| Fatty acids | 228.37 | Caur,2278'/ MtrR | *Chloroflexus aurantiacus* | 3XP (74) | 111–292 | 965.3 | |
| FMN | 456.34 | Aureochrome 1a LOV | *Phaeodactylum tricornutum* | 5A8B (77) | 34–138 | 552.8 | |
| Heme-B | 616.49 | Aer2 | *Pseudomonas aeruginosa* | 3VOL (93) | 32–135 | 814.2 | |
| | 616.49 | Aer2 | *Vibrio cholerae* | 6CEQ (94) | 170–280 | 1,007.5 | |
| | 616.49 | DosP | *Escherichia coli* | 1V9Y (95) | 30–132 | 565.1 | |
| | 616.49 | HODM | *Pseudomonas mendocina* | 5LTE (96) | 155–290 | 1,869.9 | |
| | 616.49 | FixL | *Bradyrhizobium japonicum* | 1DRM (97) | 13–117 | 984.8 | |
| | 616.49 | FixL | *Rhizobium melliloti* | 1D06 (98) | 26–130 | 907.4 | |

(Continued on next page)
| Cofactor or ligand | M<sub>ω</sub> of cofactor or ligand (g/mol) | Protein | Organism | PDB | PAS domain boundary from PDB RCSB | Pocket MS vol (Å<sup>3</sup>) | Comment |
|-------------------|------------------------------------------|---------|----------|-----|-----------------------------|---------------------------|---------|
| Heme-C            |                                          |         |          |     |                            |                           |         |
|                   | 616.49                                   | GSU0582 | Geobacter sulfurreducens | 3B47 (99) | 45–131                      | 24.8                      | Non-classical heme |
|                   | 616.49                                   | GSU0935 | Geobacter sulfurreducens | 3B42 (99) | 45–127                      | 12.1                      | cofactor binding |
|                   | 618.50                                   | TI0287  | Thermosynechococcus elongatus | 5B82 (100) | 26–186                      | 1,196.8                   | Extended pocket |
| Metals            |                                          |         |          |     |                            |                           |         |
|                   | 107.87                                   | CusS    | Escherichia coli          | 5KUS (101) | 38–185                      | 58.1                      |                     |
|                   | 65.39                                    | CzcS    | Pseudomonas aeruginosa    | 5GPO (102) | 38–161                      | 91.5                      |                     |
| No cofactor or ligand binding | NA | Agp1 (Atu1990) | Agrobacterium fabrum | 5HSQ (103) | 20–108                      | 52.7                      |                     |
|                   | NA                                       | Agp2 (Atu2165) | Agrobacterium fabrum | 6GY1 (104) | 21–119                      | 175.1                     | Pocket open to solvent |
|                   | NA                                       | AhR     | Homo sapiens             | 5NJ8 (105) | 106–253                     | 67.5                      |                     |
|                   | NA                                       | AhR     | Mus musculus             | 4M4X (106) | 41–186                      | 42.3                      |                     |
|                   | NA                                       | AhRR    | Homo sapiens             | 5Y7Y (107) | A102–A256                   | 133.0                     |                     |
|                   | NA                                       | ARNT (PAS A) | Bos taurus           | 5Y7Y (107) | 89–189                      | 1,188.7                   |                     |
|                   | NA                                       | ARNT (PAS B) | Bos taurus           | 5Y7Y (107) | B208–B311                   | 128.2                     |                     |
|                   | NA                                       | ARNT (PAS B) | Homo sapiens             | 1XXO (108) | 1–119                       | 38.9                      |                     |
|                   | NA                                       | ARNT (PAS A) | Mus musculus           | 4ZP4 (109) | 92–263                      | 48.1                      |                     |
|                   | NA                                       | ARNT (PAS B) | Mus musculus           | 4ZP4 (109) | 282–384                     | 137.3                     |                     |
|                   | NA                                       | BMAL1/ARNTL (PASB) | Mus musculus | 4F3L (110) | 277–382                     | 234.3                     |                     |
|                   | NA                                       | CLOCK (PAS B) | Mus musculus           | 4F3L (110) | 250–353                     | 146.5                     |                     |
|                   | NA                                       | Cph1    | Synechocystis sp.       | 2VEA (111) | 29–126                      | 65.8                      |                     |
|                   | NA                                       | DhaR/ YcgU | Escherichia coli         | 4LRX (35) | 14–206                      | 90.7                      |                     |
|                   | NA                                       | BhpH    | Deinococcus radiodurans | 1ZTU (112) | 122–194                     | 79.7                      |                     |
|                   | NA                                       | EAG/ Kcnh1 | Mus musculus           | 4LLO (113) | 127–200                     | 112.5                     |                     |
|                   | NA                                       | EAG/ Kcnh1 | Rattus norvegicus      | 5K7L (114) | 131–240                     | 186.6                     |                     |
|                   | NA                                       | PadC    | Idiomarina species A28L | 5LWL (115) | 123–128                     | 86.7                      |                     |
|                   | NA                                       | MmoA (PAS B) | Methylomonas capsulatus | 3DWK (72) | 122–227                     | 88.0                      |                     |
|                   | NA                                       | NcoA1 PAS B | Homo sapiens             | 5NWM (116) | 245–345                     | 410.7                     |                     |
|                   | NA                                       | NcoA1-1/ SRC-1 | Mus musculus           | 1OJS (117) | 249–367                     | 204.5                     | Pocket w/o occupancy |
|                   | NA                                       | BhpH    | Pseudomonas aeruginosa  | 3C2W (118) | 25–114                      | 136.1                     |                     |
|                   | NA                                       | PhyB    | Arabidopsis thaliana    | 4OUR (119) | 123–223                     | 41.1                      |                     |
|                   | NA                                       | PpsR (N-PAS) | Rhodobacter sphaeroides | 4HH2 (120) | 124–223                     | 25.7                      |                     |
|                   | NA                                       | PpsR (PAS1) | Rhodobacter sphaeroides | 4HH2 (120) | 124–223                     | 25.7                      |                     |
|                   | NA                                       | PpsR (PAS2) | Rhodobacter sphaeroides | 4HH2 (120) | 124–223                     | 56.3                      |                     |
|                   | NA                                       | BhpP1 PAS1 | Rhodopseudomonas palustris | 4GW9 (121) | 54–145                      | 53.2                      |                     |
|                   | NA                                       | BhpP2    | Rhodopseudomonas palustris | 4GW9 (121) | 54–145                      | 53.2                      |                     |
|                   | NA                                       | BhpP2    | Rhodopseudomonas palustris | 4EO4 (122) | 29–121                      | 21.0                      |                     |
|                   | NA                                       | BhpP3    | Rhodopseudomonas palustris | 2OOL (123) | 42–138                      | 100.1                     |                     |
|                   | NA                                       | BhpP     | Stigmatella aurantiaca | 6BAF (124) | 17–112                      | 12.3                      |                     |
|                   | NA                                       | BhpP2    | Stigmatella aurantiaca | 6PTQ (125) | 19–103                      | 31.4                      |                     |

(Continued on next page)
| Cofactor or ligand | $M_w$ of cofactor or ligand (g/mol) | Protein | Organism | PDB | PAS domain boundary from PDB RCSB | Pocket MS vol (Å³) | Comment |
|-------------------|-----------------------------------|---------|----------|-----|---------------------------------|-------------------|---------|
| NA                | Soluble guanylate cyclase (sGC) α domain | Manduca sexta | 4GJ4 (126) | 10–110 | 31.4 |
| NA                | Soluble guanylate cyclase (sGC) α subunit | Homo sapiens | 6JTO (127) | A288–A386 | 47.1 |
| NA                | Soluble guanylate cyclase (sGC) β subunit | Homo sapiens | 6JTO (127) | B217–B326 | 72.5 |
| NA                | XccBphP (N-terminal PAS domain) | Xanthomonas campestris | 5AKP (128) | 33–128 | 97.4 |
| NA                | XccBphP (C-terminal PAS domain) | Xanthomonas campestris | 5AKP (128) | 534–637 | 852.6 |
| dCache - amino acids | 89.09 | CtaA | Pseudomonas fluorescens | 6PXY (129) | 41–269 | 147.3 |
| dCache - amino acids | 89.09 | Mlp24/ McpX/ VC_A0923 | Vibrio cholerae | 3C8C (21) | 1–226 | 138.5 |
| dCache - amino acids | 89.09 | Mlp37 | Vibrio cholerae | 5AVE (130) | 5–234 | 127.5 |
| dCache - amino acids | 105.09 | PctA | P. aeruginosa | 5LTX (44) | 29–256 | 334.5 |
| dCache - amino acids | 149.21 | PctB | P. aeruginosa | 5LT9 (44) | 33–256 | 227.8 |
| dCache - amino acids | 175.21 | PctC | P. aeruginosa | 5LTV (44) | 33–257 | 191.8 |
| dCache - amino acids | 151.3 | PscC | Pseudomonas syringae | 6MNI | 23–275 | 192.5 |
| dCache - amino acids | 131.17 | Tlp3 | Campylobacter jejuni | 4XMR (131) | 37–285 | 248.9 |
| dCache - amino acids | 111.14 | TlpQ | P. aeruginosa | 6FU4 (132) | 39–323 | 753.7 |
| dCache - cytosine | 111.10 | Dret_0059 | Desulfobulbacter retbaense | 5ERE (133) | 322–562 | 349.1 |
| dCache - phosphate | 94.97 | VP0354 (vpHK1S-Z8) | Vibrio parahaemolyticus | 3LID (21) | 8–269 | 224.1 |
| dCache - polyamines | 88.15 | McpU | Pseudomonas putida | 6F9G (134) | 41–300 | 615.0 |
| Cache - no cofactor or ligand binding | 120.15 | LuxQ | Vibrio cholerae | 3C38 (21) | 21–240 | 37.0 |
| Cache - no cofactor or ligand binding | 120.15 | LuxQ | Vibrio harveyi | 2HJE (135) | 2–221 | 23.7 |
| dCache - QACs | 144.19 | McpX | Rhizobium melliloti | 6D8V (136) | 38–306 | 229.7 |
| dCache - cytokinins | 203.24 | AHK4 | Arabidopsis thaliana | 3T4J (137) | 126–393 | 528.2 |
| dCache - carboxylic acids | 118.09 | DctB | Rhizobium melliloti | 3E4O (138) | 48–301 | 134.1 |
| dCache - carboxylic acids | 118.09 | DctB | Vibrio cholerae | 3BY9 (24) | 27–285 | 130.3 |
| dCache - carboxylic acids | 88.06 | KinD | Bacillus subtilis | 4JGO (139) | 6–204 | 156.2 |
| dCache - carboxylic acids | 90.08 | TlpC | Helicobacter pylori | 5WBF (140) | 3–261 | 263.2 |
| sCache - acetate sensing | 59.04 | Adeh_3718 | Anaeromxobacter dehalogenans | 4K08 (47) | 57–144 | 84.9 |
| sCache - carboxylic acids | 189.10 | CitA | Klebsiella pneumoniae | 1P0Z (50) | 50–126 | 356.9 |
| sCache - carboxylic acids | 134.09 | DcuS | Escherichia coli | 3BY8 (24) | 56–130 | 174.9 |

(Continued on next page)
| Cofactor or ligand | M_w of cofactor or ligand (g/mol) | Protein | Organism | PDB | PAS domain boundary from PDB RCSB | Pocket MS vol (Å³) | Comment |
|-------------------|---------------------------------|---------|----------|-----|----------------------------------|-------------------|---------|
| VP0183            | 88.06                           | Vibrio parahaemolyticus | 4EXO (45) | 56–146 | 95.9                             |                   |         |
| PscD-SD           | 73.07                           | Pseudomonas syringae     | 5G4Y (46) | 32–178 | 98.7                             |                   |         |
| PhoQ              | 58.69                           | Escherichia coli         | 3BQ8 (59) | 41–138 | 108.5                            |                   |         |
| sCache - metals   |                                 | PhoQ                | Salmonella enterica serovar Typhimurium | 1YAX (23) | 39–138 | 32.8                             |                   |         |
| sCache - urea     | 60.05                           | TlpB                | Helicobacter pylori     | 3UB6 (45) | 70–156 | 170.9                            |                   |         |

*PAS or Cache domain boundaries are indicated. The pocket or cavity volume is presented along with the molecular weight (M_w) of the cofactor or ligand in the pocket/cavity, where present. MS, pocket volume based on the molecular surface; QAC, quaternary ammonium compound.
was taken to be indicative of the capacity to bind cofactor or ligand. Cofactor or ligand binding capacity was thus inferred from phylogenetic analysis and conservation of key amino acid residues.

The dCache domains of PA1336, PA5165, and PA5512 that were identified to group with sequences of carboxylic acid-binding dCache domain structures from the reference data set (arrow 1 in Fig. 1) were investigated to confirm conservation of ligand-coordinating residues. Amino acids responsible for substrate coordination within the carboxylic acid-binding

FIG 1 Maximum likelihood phylogenetic analysis of *Pseudomonas aeruginosa* PA01 PAS or Cache domains with the reference set of structurally characterized domains. The percentage of bootstrap replicates that reproduced each branch is given, with branches corresponding to less than 15% of bootstrap replicates collapsed and rearranged for clarity. PAS, dCache, and sCache domains are labeled with a circle, square, or triangle, respectively. The nature of ligand or cofactor is given in the key and denoted by color, and individual alignments of these groups are found in the supplemental material. Groups discussed in the text are marked with a numbered arrow. The supplement to this article contains an evaluation of different phylogenetic analyses and alignments of individual clades shown in Fig. 1 and discussed in the text.
The dCache domain of DctB are shown in Fig. 2. It is worth pointing out that these residues vary between KinD carboxylic acid-binding dCache domains, where substrates (succinate, malonate, pyruvate, and lactate) adopt a different binding pose, and DcuS sCache domains of different organisms that display different substrate specificity (citrate, malate, pyruvate, and propionate). PA5512, PA1336, and PA5165 use the same repertoire of ligand-coordinating amino acid residues as DctB. However, the slight variation needs to be discussed with respect to substrate specificity.

A interesting example evaluates fatty acid-binding PAS domains, of which a number of different binding poses and interactions have been structurally characterized (alignment found in supplemental material). The reference protein Rv1364c from *Mycobacterium tuberculosis* contains a binding motif used to coordinate palmitic acid in reference 38. PA0847 PA52, PA1196 PAS1, PA1976 PAS3, and PA4112 PAS2 all display conservation of the two relevant amino acids responsible for side chain-specific ligand interaction. In PA1196, there is a conservative exchange (aspartate to glutamate)
in one of these two recognition amino acids. The analysis of conservation performed here adds significantly to the phylogenetic analysis, as the majority of fatty acid-binding PAS domains were assigned on the basis of conserved ligand-binding amino acids.

Indeed, a number of novel assignments can be made based on conserved binding motif. For example, PA2005 PAS1 could not be placed into a clade through phylogenetic analysis but is assigned here as autoinducer binding based on conservation of the two residues critical for side chain-specific coordination of the autoinducer DPO (3,5-dimethylpyrazin-2-ol), seen in *V. cholerae* VqmA (6IDE) (39). The alignment reveals that one amino acid is conserved while the other one is a conservative exchange from lysine to arginine. The data generated are summarized in Fig. 3 and Table 3.

**DISCUSSION**

Individual domains are the building blocks of modular proteins and are required for functional diversification of the proteome. Understanding of protein function is crucially dependent on our grasp of physiological and functional roles of these constituting domains. For the omnipresent PAS and Cache domains, analysis is generally hampered by failure to predict cofactor- or ligand-binding state from sequence. Even identification of these domains proves to be difficult, due to low sequence conservation. While HMM sequence searches are a sensitive method to detect homology in cases of low sequence identity, we also made use of the SMART domain prediction server (30, 31) and used structure-guided analyses here.

We have studied the bacterial model organism *Pseudomonas aeruginosa* that has the ability to adapt to various environmental conditions, a survival strategy and an underlying property important in the clinical setting. We sought to identify the nature of cofactors and ligands that bind to PAS or Cache domains within *P. aeruginosa* PAO1 using phylogeny and structural conservation analyses. A number of differences exist
### TABLE 3 PAS or Cache domains and predicted cofactors or ligands assigned on the basis of combined phylogeny and sequence-structure alignment

| Cofactor or ligand | Protein | Domain | Known physiological role |
|--------------------|---------|--------|--------------------------|
| **Amino acids**    |         |        |                          |
|                    | PA2654 (TlpQ) | dCache | Chemotaxis toward ethylene and histamine (132, 141, 142) |
|                    | PA4307 (PctC) | dCache | Chemotaxis toward amino acids (44, 143, 144) |
|                    | PA4309 (PctA) | dCache | Chemotaxis toward amino acids (44, 143–145) |
|                    | PA4310 (PctB) | dCache | Chemotaxis toward amino acids (44, 143, 144) |
|                    | PA4633     | dCache | Unknown (146) |
|                    | PA4886     | PAS1   | Unknown (147, 148) |
| **Autoinducers**   | PA1261 (LhpR) | PAS1   | Transcriptional regulator (149) |
|                    | PA2005 (HbcR) | PAS1   | Regulation of (R)-3-hydroxybutyrate catabolism (150) |
| **Carboxylic acids -dCache like** |         |        |                          |
|                    | PA1336 (AauS) | dCache | Regulation of genes involved in aspartate, glutamate, and glutamine uptake and catabolism (43) |
|                    | PA1417 (AcoR) | PAS1   | Regulation of α-ketoglutarate transport and utilization (41, 42) |
|                    | PA5165 (DctB) | PAS1   | Regulation of C4-dicarboxylic acid transport systems (34) |
|                    | PA5512 (MftS) | dCache | Regulation of β-ketoglutarate transport and utilization (41, 42) |
| **Carboxylic acids -sCache like** |         |        |                          |
|                    | PA2652 (CtpM) | sCache | Chemotaxis toward malate (48, 49, 151) |
|                    | PA4021 (EutR) | PAS1   | Regulation of ethanolamine catabolism (51) |
|                    | PA4147 (AcoR) | PAS1   | Regulation of 2,3-butanediol and acetoin metabolism (52, 53) |
| **Cytokinins**     | PA1976 (ErcS') | PAS3   | Regulation of ethanol oxidation (152, 153) |
| **FAD**            | PA0285     | PAS2   | Regulation of biofilm formation (154) |
|                    | PA0575     | PAS4   | Regulation of biofilm formation in response to L-arginine |
|                    | PA4123 (BdiA) | PAS1   | Regulation of biofilm dispersal (17, 155, 156) |
|                    | PA4123 (BdiA) | PAS2   | Regulation of biofilm dispersal (17, 155, 156) |
|                    | PA1561 (Aer/TlpC) | PAS1   | Aerotaxis (157, 158) |
|                    | PA1930 (McpS) | PAS1   | Regulation of chemotaxis (40) |
|                    | PA1930 (McpS) | PAS2   | Regulation of chemotaxis (40) |
|                    | PA4601 (MolA) | PAS4   | Regulation of flagellar development and protease secretion (159–162) |
|                    | PA5442     | PAS2   | Unknown |
| **Fatty acids**    | PA0290     | PAS1   | Regulation of biofilm formation and Psl production (154, 163–165) |
|                    | PA0847     | PAS2   | Regulation of motility in response to a no. of stimuli (165, 166) |
|                    | PA1196 (DdaR) | PAS1   | Regulation of methylarginine metabolism, role in quorum-sensing (167, 168) |
|                    | PA1243     | PAS1   | Regulation of swimming and biofilm formation (169) |
|                    | PA1438 (MmnS) | PAS1   | Regulation of efflux pump expression (170) |
|                    | PA1976 (ErcS') | PAS2   | Regulates ethanol oxidation (152, 153) |
|                    | PA4112     | PAS2   | Histidine kinase of unknown pathway |
|                    | PA4197 (BfS) | PAS2   | Regulation of biofilm formation (171–174) |
|                    | PA4293 (PprA) | PAS2   | Regulation of outer membrane permeability of biofilm formation (175–177) |
|                    | PA4581 (RtcR) | PAS1   | Homologous to E. coli regulator of RNA 3'-terminal phosphate cyclase expression (178–180) |
|                    | PA5017 (DipA) | PAS1   | Biofilm regulation, chemotaxis, motility, maintenance of c-di-GMP heterogeneity (19, 181–183) |
|                    | PA5442     | PAS1   | Unknown |
| **Heme-b**         | PA0176 (Aer2/TlpG/McpB) | PAS1   | Aerotaxis and virulence (93, 184, 185) |
|                    | PA1976 (ErcS') | PAS1   | Regulates ethanol oxidation (152, 153) |
|                    | PA2177     | PAS1   | Unknown |
|                    | PA3271 (MxtR) | PAS1   | Redox sensing and interbacterial signaling (186, 187) |
|                    | PA5442     | PAS2   | Unknown |
| **Metals**         | PA0464 (CreC) | sCache | Regulation of carbon source catabolism (188, 189) |
|                    | PA2524 (CzcS) | PAS1   | Regulation of metal detoxification and resistance to carbapenem antibiotics (102, 190–192) |
|                    | PA2870     | PAS1   | Diguanylate cyclase involved in biofilm production, Psl production, regulation of swimming motility (165) |
| **No cofactor or ligand binding** |         |        |                          |
|                    | PA0285     | PAS1   | Regulation of biofilm formation (154) |
|                    | PA0338     | PAS1   | Regulation of biofilm formation, Psl production, and swimming motility (165) |
|                    | PA1181 (YegE) | PAS2   | Biofilm dispersal (18, 193) |
|                    | PA4112     | PAS3   | Histidine kinase of unknown pathway |
|                    | PA4117 (BphP) | PAS1   | Quorum sensing (118, 194, 195) |
|                    | PA5124 (NtrB) | PAS1   | Regulation of nitrogen metabolism, rhamnolipid production, biofilm formation, expression of virulence genes, and swarming (196–200) |
|                    | PA5442     | PAS1   | Unknown |
between predictions based on maximum likelihood phylogeny and the individual alignment and inspection of conservation of cofactor- or ligand-interacting amino acids. The PAS domains of PA0873 and PA2449 provide an example in which phylogenetic analysis places them with the reference 4-hydroxycinnamic acid-binding PAS domains. However, when conserved ligand- or cofactor-interacting sidechains were assessed, this classification did not hold. Therefore, analysis based on one method alone may be indicative but not conclusive. The results of our combined analysis and predictions are summarized in Tables 1 and 3, and Fig. 3 highlights differences in assignment from the two different approaches used here.

Our analysis revealed a number of relationships and provides new insight. An example are the four PAS domains marked with the black arrow 3 in Fig. 1 that mark the PAS domains of PA1423 (BdlA) and PA1930 (McpS). Both proteins possess the same architecture, with two N-terminal PAS domains coupled to a methyl-accepting chemotaxis domain. Though they are clearly related, the question of functional diversification arises. Indeed, PA1930 has been reported to have a negative effect on chemotaxis (40), while PA1423 is involved in biofilm dispersal (17). It is therefore likely that the two proteins respond to different triggers and, in doing so, lead to a different biological output. We have experimentally characterized a similar example previously with the proteins PA2072 and RbdA that share an architecture but are responsible for two almost orthogonal functions (19). Thus, there are examples where gene duplication allows proteins to diversify to functionally evolve.

An interesting observation is made here with flavin binders. The distinct clade with sequences of the flavin mononucleotide (FMN)-binding PAS domain structures does not reveal PAO1 PAS domain relatives. However, FAD-binding PAS domains are identified in PAO1. Whether, indeed, FMN is not used as a cofactor in PAO1 remains to be seen. It might turn out that PAO1 has some remarkable and truly distinct PAS domains, and further structural analyses rather than predictions will in time reveal this.

The strength of the combined approach to analyze both phylogeny and conservation of cofactor- or ligand-specifying amino acids is exemplified here with the analysis of carboxylic acid-binding PAS or Cache domains. When the in silico results are placed into physiological context, additional insight is gained. We identified carboxylic acid-binding domains in two distinct classes for both sCache and dCache sensory architectures. Variation in substrates and their coordination is detectable between the two classes (24), and consequently, we identify different clades likely to present different substrate interaction and selectivity. The dCache domains illustrate the approach taken by combining sequence, phylogenetic, and structural information. As such, the dCache domains of PA1336, PAS165, and PA5512 are all inferred here to be able to bind carboxylic acids (Fig. 2).

The dCache domain of PAS165 (DctB) was assigned as carboxylic acid binding and, within P. aeruginosa, DctB acts as sensor of a two-component pathway involved in regulating the expression of C₄-dicarboxylic acid transport systems (34). It therefore follows that binding of carboxylic acids to the dCache domain of DctB could directly couple levels of C₄-dicarboxylic acids to a signaling cascade responsible for the expression of transport systems used in their uptake (Fig. 4).

Similarly, a carboxylic acid-binding dCache domain within PA5512 (MifS) could also directly link signal perception to a known phenotype (Fig. 4). MifS is required for the transport and utilization of the C₅-dicarboxylic acid α-ketoglutarate (41, 42). It can therefore be hypothesized that this might reveal a potential substrate not previously recognized for dCache domains. As carboxylic acid-binding dCache domains are known to bind C₃ and C₄ substrates with at least one carboxylic acid (15), it is conceivable that the dCache domain in MifS may also be able to bind α-ketoglutarate and act as a sensor. The variation of amino acids identified in the binding pocket from sequence alignments may reflect the required level of flexibility to accept various substrates or binding poses across the variety of PAS and Cache domains in these proteins.

The assignment of PA1336-dCache as carboxylic acid binding could help to identify a source of selectivity within dCache domains. Studies of a protein orthologous to
PA1336 (AauS) within Pseudomonas putida demonstrate a role in utilization of the amino acids aspartate, glutamate, and glutamine (43), all of which contain side chain carboxylic acid or C=O groups. As PA1336-dCache was classified here as carboxylic acid binding and not amino acid binding, it could be speculated that a similar role for PA1336 within PAO1 to its orthologue in P. putida could be accommodated through interaction between the side chains of these amino acids and the PA1136-dCache domain, which then induces a conformational change that activates the two-component system partner of PA1336 to alter gene expression (Fig. 4) (43). Indeed, the highly conserved region identified for amino acid-binding dCache domains (44) is different in PA1336-dCache and instead shows similarity to carboxylic acid-binding dCache domains, with carboxyl groups likely neutralizing the charge through conservation of positively charged side chains of the amino acids labeled R1 and K6 in Fig. 2. These observations may guide future predictions of amino acid selectivity.

Another example of a straightforward link between our analysis and a previously defined physiological function would be the coupling of malate binding to PA2652 (CtpM) with chemotaxis (Fig. 4). The sCache domain of CtpM is assigned as carboxylic acid binding on the basis of its phylogenetic relationship with Vibrio paraheamolyticus VP0183 (4EXO) (45), P. syringae PscD (5G4Y) (46), and Anaeromyxobacter dehalogenans Adeh_3718 (4K08) (47) (Fig. 1) and conservation of five ligand-coordinating amino acids. A function in carboxylic acid binding aligns well with previous reports that CtpM is involved in chemotaxis and has substrate specificity toward malate, which is a known substrate for carboxylic acid-binding dCache domains (24, 48, 49). It may therefore follow from our analysis that the binding of malate to an sCache within CtpM directly couples malate concentration to associated chemotaxis signaling.

Interestingly, phylogeny analysis groups the first PAS domain of PA4021 (EatR) and the first PAS domain of PA4147 (AcoR) with the carboxylic acid-binding sCache domains of E. coli DcuS (3BY8) (24) and Klebsiella pneumoniae CitA (1POZ) (50) and the...
phosphate-binding dCache domain of Vibrio parahaemolyticus VP0354 (3LID) (21). Both PA4021 and PA4147 have known functions as transcriptional regulators for the metabolism of small, hydroxyl-containing, organic compounds and are proposed to perform those functions in response to acetaldehyde (51–53). As acetaldehyde is, to some extent, similar in structure to the carboxylic acids detected by the reference structures in these clades, it is possible that acetaldehyde binding directly to the PAS domains present in PA4021 and PA4147 could form a concise way to induce these changes in transcription and could be the basis of a novel class of PAS domain ligand (Fig. 4).

In conclusion, this study uses protein sequence comparison, phylogeny, and structure-based prediction of ligand or cofactor binding for PAO1 PAS and Cache domains. Although just predictions, the classifications presented give insight from comparison with similar proteins, leading to experimentally testable hypotheses to gain functional insights.

MATERIALS AND METHODS

Selection of P. aeruginosa PAO1 PAS and Cache domains. HMM-to-HMM comparisons have previously identified 70 proteins within P. aeruginosa PAO1 that contain PAS and Cache domains (22, 26). Protein sequences of these proteins were retrieved from the Pseudomonas genome database (11). Selection of the final data set of 101 sequences, containing 91 PAS domains, 9 dCache domains, and 2 sCache domains, is described in Results. These are listed in Table 1.

Generation of the reference data set with 3D structures of PAS and Cache domains. The DALI webserver (54) was used for an exhaustive search of PAS and Cache domains within the Protein Data Bank (PDB; March 2020). Search models were chosen to represent different cofactor- or ligand-binding architectures. PAS domains from Bradyrhizobium japonicum FixL (heme-b binder, PDB: 1xj2 [55]), Brucella abortus LOV-HK (FMN binder, 3S50 [56]), Azotobacter vinelandii NifL (FAD binder, 2qj3 [57]), and H. halophila PYP (’4’-hydroxycinnamoyl acid binder, 2phy [58]) were used. Further, sCache domains PhoQ (cation binder, 3bq8 [59]) and DcuS (carboxylic acid binder, 3by8 [24]) from E. coli were used. Finally, the dCache domains DctB (carboxylic acid binder, 3by9 [24]) from V. cholerae and PctB (amino acid binder, 5It9 [44]) from P. aeruginosa were used. The structures were submitted individually and together retrieved a total of 7,513 matches, corresponding to 986 individual PDB entries. The results included structures not classified as either PAS or Cache domains that were discarded, for example, structurally related GAF domains. Retained were structures with a functional cofactor or ligand bound as well as those with a reported signaling function, referred to hereafter as “no cofactor or ligand” binding. The final reference data set contained a total of 106 PAS and Cache domains trimmed down to the PAS and Cache domain boundaries and included 78 PAS domains, 20 dCache domains, and 8 sCache domains.

Maximum likelihood phylogeny. The 106 sequences of the reference data set and the 101 sequences from P. aeruginosa PAO1 were aligned using CLUSTALW, as implemented in MEGA7 (60, 61). This alignment was then subjected to molecular phylogenetic analysis by maximum likelihood methods within MEGA7 (61, 62). Initial phylogenetic trees were obtained by applying Neighbor-Join (63) and BioNJ (64) algorithms to a matrix of pairwise distances estimated using the JTT-matrix based model (65). The trees were scored and automatically selected based on log-likelihood scores. The bootstrap consensus tree is inferred from 100 replicates and taken to represent the evolutionary history of taxa analyzed (66).

Sequence-structure analysis. The P. aeruginosa sCache and dCache domains were aligned to the equivalent subsets in the reference data set using PROMALS3D (37) to determine conservation of ligand-coordinating amino acids residues. Conservation was used to suggest their potential ligand-binding class. For the larger set of PAS domain sequences, the reference data set was divided up according to ligand or cofactor (see Table 2) and then aligned against the P. aeruginosa PAS sequences.

Determination of binding pocket or cavity size. The sizes of enclosed cavities or of binding pockets that are open to the surrounding environment allow different classes of PAS or Cache domains to be distinguished. To map their size, the coordinates of the reference structures were uploaded to the CASTp server (67), with the PAS or Cache domain boundaries as given in Table 2. CASTp returns multiple pockets and cavities, which were inspected using UCSF Chimera (68); where a cofactor or ligand was present, this pocket was chosen, but when no cofactor or ligand was identified, the one closest to the center of the PAS or Cache domain was reported, ensuring all cavities/pockets reported here were in a similar position. Table 2 reports the solvent-excluded volume calculated with a probe sphere radius of 1.4 Å, based on Connolly’s molecular surface calculation (69), as this parameter was able to discriminate ligand- or cofactor-binding pockets/cavities most clearly. Where the open pockets reported volumes that include not only the actual ligand or cofactor cavity but also the access to the cavity, this is noted in Table 2.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 7.4 MB.
ACKNOWLEDGMENTS
We acknowledge funding by Diamond Light Source and the University of Southampton to A.H. and C.C. Conceptualization, I.T., M.A.W.; investigation, A.H., C.C.; formal analysis, data curation, and validation, A.H., C.C., I.T.; visualization, A.H., I.T.; supervision, M.A.W., J.S.W., I.T.; writing – review and editing, A.H., I.T.; writing – original draft, A.H.; funding acquisition, M.A.W., J.S.W., I.T.

REFERENCES
1. Hardalco C, Edberg SC. 1997. *Pseudomonas aeruginosa* assessment of risk from drinking water. Crit Rev Microbiol 23:47–75. https://doi.org/10.3109/104084970911530
2. Stover CK, Pham QX, Erwin al, Mizoguchi SD, Warrener P, Hickey MJ, Brinkman FS, Hufnagle WD, Kowalik DJ, Lagrou M, Garber RL, Golty L, Tolentino E, Westbrook-Wadam S, Yuan Y, Brody LL, Coulter SN, Folger KR, Kars A, Larbig K, Lim R, Smith K, Spencer D, Wong GK, Wu Z, Paulsen IT, Reizer J, Saier MH, Hancock RE, Lory S, Olson MV. 2000. Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. Nature 406:959–964. https://doi.org/10.1038/35023079
3. Bateman A. 2010. The Pfam protein families databases. Nucleic Acids Res 43:D211–222. https://doi.org/10.1093/nar/gkp985
4. Pausch S, Rocha E, Delport WM, Redfield CR, Spiering CA, Jones MW, Southgate DJ, Telford SD, Wilm LE, Wilkinson KS, Simpson MR, Martens SG, St併derup E, Westbrock-Wadman S, Yuan Y, Brody LL, Coulter SN, Folger KR, Kars A, Larbig K, Lim R, Smith K, Spencer D, Wong GK, Wu Z, Paulsen IT, Reizer J, Saier MH, Hancock RE, Lory S, Olson MV. 2000. Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. Nature 406:959–964. https://doi.org/10.1038/35023079
5. Cai Y-m, Hutchin A, Craddock J, Walsh MA, Webb JS, Tews I. 2020. Differential impact on motility and biofilm dispersal of closely related phosphodiesterases in *Pseudomonas aeruginosa*. Sci Rep 10:6232. https://doi.org/10.1038/s41598-020-63008-5
6. Zhang Z, Hendrickson WA. 2010. Structural characterization of the predominant family of histidine kinase sensor domains. J Mol Biol 400:335–333. https://doi.org/10.1016/j.jmb.2010.04.049
7. Waddell F, Whitchurch CB, Mattick JS. 2003. FimX, a multidomain protein connecting environmental signals to twitching motility in *Pseudomonas aeruginosa*. Microbiol Mol Biol Rev 67:167–193. https://doi.org/10.1128/MMBR.00043-12
8. Zhang Z, Hendrickson WA. 2010. Structural characterization of the predominant family of histidine kinase sensor domains. J Mol Biol 400:335–333. https://doi.org/10.1016/j.jmb.2010.04.049
9. Anantharaman V, Aravind L. 2000. Cache – A signaling domain common to archaeal two-component systems and status in 2005. Nucleic Acids Res 33:D257–261. https://doi.org/10.1093/nar/gkf087
10. Klockgether J, Munder A, Neugebauer J, Davenport CF, Stanke F, Larbig K, Lim R, Smith K, Spenc...
49. Lacal J, Alfonso C, Liu X, Parales RE, Morel B, Conejero-Lara F, Rivas G, Alvarez-Ortega C, Harwood CS. 2007. Identification of a functional dimer of the PhoQ sensor domain. J Biol Chem 282:13762–13768. https://doi.org/10.1074/jbc.M305868200.

50. Lindsey, O., Williams DR, Getzoff ED. 1995. 1.4 Å structure of photoactive yellow protein of the periplasmic ligand-binding domain of the sensor kinase CitA. J Bacteriol 178:2543–2551. https://doi.org/10.1128/JB.178.13.2543-2551.1996.

51. Donahoe, DM, Maddock JR. 2005. Polar localization of a soluble methyl-accepting chemotaxis protein of Pseudomonas aeruginosa. J Bacteriol 187:7840–7844. https://doi.org/10.1128/JB.187.13.7840-7844.2005.

52. Lundgren BR, Villegas-Parianara LR, Harris JR, Mottern AM, Dunn DM, Boddy CN, Nomura CT. 2014. Genetic analysis of the assimilation of CS-dicarboxylic acids in Pseudomonas aeruginosa PA01. J Bacteriol 196:2543–2551. https://doi.org/10.1128/JB.00823-14.

53. Tatske, G, Kumar H, Silva-Hertzog E, Ramirez L, Mathee K. 2015. Pseudomonas aeruginosa a MifS-MifR two-component system is specific for α-ketoglutarate utilization. PLoS One 10:e0129629. https://doi.org/10.1371/journal.pone.0129629.

54. Sonawane, AM, Singh B, Röhm KH. 2006. The AauR-AauS two-component system regulates uptake and metabolism of acidic amino acids in Pseudomonas putida. Appl Environ Microbiol 72:6569–6577. https://doi.org/10.1128/AEM.00830-06.

55. Gavira, JA, Gumerov VM, Rico-Jiménez M, Petukh M, Upadhyay AA, Alfonso C, Liu X, Parales RE, Morel B, Conejero-Lara F, Rivas G, Alvarez-Ortega C, Harwood CS. 2007. Identification of a functional dimer of the PhoQ sensor domain. J Biol Chem 282:13762–13768. https://doi.org/10.1074/jbc.M305868200.

56. Brewster, JL, McKellar JLO, Finn TJ, Newman J, Peat TS, Gerth ML. 2016. Structural basis for ligand recognition by a Cache chemosensory domain protein of Pseudomonas aeruginosa. Structure 24:693–703. https://doi.org/10.1016/j.str.2016.01.006.

57. Parales, RE, Morel B, Conejero-Lara F, Rivas G, Alvarez-Ortega C, Harwood CS. 2007. Identification of a functional dimer of the PhoQ sensor domain. J Biol Chem 282:13762–13768. https://doi.org/10.1074/jbc.M305868200.

58. Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791. https://doi.org/10.1111/j.1558-5646.1985.tb02102.x.

59. Zhang, X, Wu, C, Lei, X, Zhao, X, Jiao, L. 2018. CASTP 3.0: computed atlas of surface topography of proteins. Nucleic Acids Res 46:W363–W367. https://doi.org/10.1093/nar/gky473.

60. Pettersen, EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612. https://doi.org/10.1002/jcc.20004.

61. Connolly, ML. 1983. Space-filling-accessible surfaces of proteins and nucleic acids. Science 221:709–713. https://doi.org/10.1126/science.6871970.

62. Zhang, O, Tian, W, Cheng, C, Lei, X, Zhao, X, Jiao, L. 2018. CASTP 3.0: computed atlas of surface topography of proteins. Nucleic Acids Res 46:W363–W367. https://doi.org/10.1093/nar/gky473.

63. Johannsen, O, Felsenstein, S, Neuhaus, J, Janssen, C, Stroh, C, Sturmer, E, Guarente, L. 2009. Analysis of periplasmic sensor domains from Aeromonas hydrophila. J Biol Chem 285:17369–17380. https://doi.org/10.1074/jbc.M900043200.

64. Riedel, S, Hofmann-Kiefer, T, Bott, M, Madden, DR. 2003. The structure of the periplasmic ligand-binding domain of the sensor kinase CiaT reveals the first extracellular PAS domain. J Biol Chem 278:39189–39196. https://doi.org/10.1074/jbc.M305864200.

65. Lundgren, BR, Sarwar Z, Pinto A, Ganley JG, Nomura CT. 2016. Ethanolamine catabolism in Pseudomonas aeruginosa PA01 is regulated by the enhancer-binding protein Eal7 (PA0421) and the alternative sigma factor RpoN. J Bacteriol 198:2318–2329. https://doi.org/10.1128/JB.00357-16.

66. Liu, Y, Liu, Y, Wang, Z, Xiao, D, Gao, C, Yu, P, Ma, C. 2018. 2.3-Butanediol catabolism in Pseudomonas aeruginosa PA01. Environ Microbiol 20:3927–3940. https://doi.org/10.1111/1462-2920.14332.

67. Camus, L, Briaud P, Bastien S, Elsen S, Doleáns-Jordheim A, Vandenesch F, Moreau K. 2020. Trophic cooperation promotes bacterial survival of Staphylococcus aureus and Pseudomonas aeruginosa. ISME J 14:3093–3105. https://doi.org/10.1038/s41396-020-00741-9.

68. Holtm., L, Elsöfsn, A. 2019. Benchmarking fold detection by Dalilite v5. Bioinformatics 35:S326–S337. https://doi.org/10.1093/bioinformatics/btz356.

69. Key, J, Moffat K. 2005. Crystall structures of deoxyx and CO-bound BjFxnL reveal details of ligand recognition and signaling. Biochemistry 44: 4627–4635. https://doi.org/10.1021/bi047942r.
synthase. PfPL. PLoS Biol 17:e1000123. https://doi.org/10.1371/journal.pbio.3000013.

77. Baneejee A, Herman E, Kottke E, Tessen LO. 2016. Structure of a native-like aureochrome 1a LOV domain dimer from Phaeodactylum tricornutum. Structure 24:171–178. https://doi.org/10.1016/j.str.2015.10.022.

78. Nazareno VV, Remeeva A, Yudenko A, Kovalev K, Dubenko A, Goncharov IM, Kuzmichev P, Rogachev AV, Buslaev P, Borschhevsky V, Mishin A, Dhoke GV, Schwaneberg U, Davari MD, Jaeger K-E, Krauss U, Gordel' D, Gushchin I. 2019. A thermostable flavin-based fluorescent protein from Chloroflexus aggregans: a framework for ultra-high resolution structural studies. Photochem Photobiol Sci 18:1793–1805. https://doi.org/10.1016/j.aphysci.2019.e00067d.

79. Fettweis T, Röllen K, Granzin J, Reiners O, Endres S, Drepper T, Willbold D, Jaeger K-E, Batra-Safferling R, Krauss U. 2018. Mechanistic basis of the fast dark recovery of the short LOV protein DsLOV from Dinorobacter shiboe. Biochemistry 57:4833–4847. https://doi.org/10.1021/acs.biochem.7b00645.

80. Nakasato M, McNulty R, Shillito ME, Swartz TE, Bogomolni RA, Luecke H. 2013. Structural basis of the LOV1 dimerization of the fungal LOV domain photoreceptor reveals an evolutionarily conserved pathway for light oxygen-sensor FixL from Rhizobium melliloti: crystallographic, mutagenesis and resonance Raman spectroscopic studies. J Mol Biol 301:415–431. https://doi.org/10.1016/j.jmb.2000.3954.

81. Pokkuluri PR, Pessanha M, Londer Y, Wood SJ, Duke NEC, Wilton R, Catarino T, Salgueiro CA, Schiffer M. 2008. 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 377:1498–1517. https://doi.org/10.1016/j.jmb.2008.01.087.

82. Lokhandwala J, Hopkins HC, Rodriguez-Iglesias A, Dattenböck C, Bittl R, Mayer G, Möglich A. 2019. A blue light receptor that mediates host cell immunity. J Biol Chem 294:1092–1109. https://doi.org/10.1074/jbc.A119.222294.

83. Schmidt A, Sauthof L, Szczepek M, Lopez MF, Escobar FV, Qureshi BM, Michael N, Buhrke D, Stevens T, Kwiatkowski D, von Stetten D, Mrojski MS, Weidbauer N, Lamparter T, Michael T, Hildebrandt P, Scheerer P, Epple P, Krauß N. 2016. Structural characterization of the periplasmic sensor domain of the histidine kinase CusS shows unusual metal ion coordination at the dimeric interface. Biochemistry 55:5296–5306. https://doi.org/10.1021/acs.biochem.6b00707.

84. Fedorov R, Schlütting I, Hartmann E, Domratcheva T, Fuhrmann M, Hartmann E, Zikihara K, Matsuoka D, Tokutomi S. 2008. Structure of a biological oxygen sensor: a new mechanism for heme-driven signal transduction. Proc Natl Acad Sci U S A 105:15189–15194. https://doi.org/10.1073/pnas.0805159105.

85. Zoltowski BD. 2017. Kinetics of the LOV domain of ZEITLUPE determine its circadian function in Arabidopsis. Elife 6:e21646. https://doi.org/10.7554/eLife.21646.

86. Weber AM, Kaiser J, Ziegler T, Pilsl S, Dattenböck C, Bittl R, Mayer G, Möglich A. 2019. A blue light receptor that mediates host cell immunity. J Biol Chem 294:1092–1109. https://doi.org/10.1074/jbc.A119.222294.

87. Sakurai S, Shimizu T, Ohto U. 2017. The crystal structure of the AhRR-PAS-A domain of the aryl hydrocarbon receptor reveals insights into its functional roles. Mol Cell 67:1470–1479. https://doi.org/10.1016/j.molcel.2017.06.028.

88. Kurokawa H, Lee D-S, Watanabe M, Sagami I, Mikami B, Raman CS, Shizuma T. 2004. A redox-controlled molecular switch revealed by the crystal structure of a bacterial heme PAS sensor. J Mol Biol 297:2016–2019. https://doi.org/10.1016/j.jmb.2004.02.008.

89. Nishida N, Takahashi JS. 2012. Crystal structure of the heterodimeric transcription factor HIF-1α/HIF-2α-DNA complex. Structure 20:698–706. https://doi.org/10.1016/j.str.2012.02.016.

90. Pudasaini A, Shim JS, Song YH, Shi H, Kiba T, Somers DE, Imazumi T, Zolothwitz BD. 2017. Kinetics of the LOV domain of ZEITLUPE determine its circadian function in Arabidopsis. Elife 6:e21646. https://doi.org/10.7554/eLife.21646.

91. Sawai H, Sugimoto H, Shiro Y, Ishikawa H, Mizutani Y, Aono S. 2012. Structural basis for oxygen sensing and signal transduction of the heme-based sensor protein Aer2 from Pseudomonas aeruginosa. Chem Commun (Camb) 48:6523-5. https://doi.org/10.1039/c2cc32549g.

92. Greer-Phillips SE, Sukomun N, Chua TK, Johnson MS, Crane BR, Watts KJ. 2018. The Aer2 receptor from Vibrio cholerae is a dual PAS-heme oxygen sensor. Mol Microbiol 109:209–224. https://doi.org/10.1111/mmi.13978.

93. Kurokawa H, Lee D-S, Watanabe M, Sagami I, Mikami B, Raman CS, Shizuma T. 2004. A redox-controlled molecular switch revealed by the crystal structure of a bacterial heme PAS sensor. J Mol Biol 297:2016–2019. https://doi.org/10.1016/j.jmb.2004.02.008.

94. Ortmayer M, Laflite P, Menon BRK, Traul T, Fisher K, Denkhaus L, Scrutton NS, Rigby SEJ, Munro AW, Hay S, Leys D. 2016. An oxidative N-demethylation reaction from ultra-high resolution structure to enzyme to mimic. Nature 539:593–597. https://doi.org/10.1038/nature20159.

95. Gung W, Hao B, Mans S, Gonzales Z, Gillez-Monaco MA, Chan MK. 1998. Structure of a biological oxygen sensor: a new mechanism for heme-driven signal transduction. Proc Natl Acad Sci U S A 95:15177–15182. https://doi.org/10.1073/pnas.95.26.15177.

96. Miyatake H, Mukai M, Park SY, Adachi S, Tamura K, Nakamura H, Nakamura K, Tsuchiya T, Izuka T, Shiro Y. 2000. Sensory mechanism of oxygen sensor FixL from Rhizobium melliloti: crystallographic, mutagenesis and resonance Raman spectroscopic studies. J Mol Biol 301:415–431. https://doi.org/10.1016/j.jmb.2000.3954.

97. Nakasato M, McNulty R, Shillito ME, Swartz TE, Bogomolni RA, Luecke H. 2013. Structural basis of the LOV1 dimerization of the fungal LOV domain photoreceptor reveals an evolutionarily conserved pathway for light oxygen-sensor FixL from Rhizobium melliloti: crystallographic, mutagenesis and resonance Raman spectroscopic studies. J Mol Biol 301:415–431. https://doi.org/10.1016/j.jmb.2000.3954.

98. Miyatake H, Mukai M, Park SY, Adachi S, Tamura K, Nakamura H, Nakamura K, Tsuchiya T, Izuka T, Shiro Y. 2000. Sensory mechanism of oxygen sensor FixL from Rhizobium melliloti: crystallographic, mutagenesis and resonance Raman spectroscopic studies. J Mol Biol 301:415–431. https://doi.org/10.1016/j.jmb.2000.3954.

99. Nakasato M, McNulty R, Shillito ME, Swartz TE, Bogomolni RA, Luecke H. 2013. Structural basis of the LOV1 dimerization of the fungal LOV domain photoreceptor reveals an evolutionarily conserved pathway for light oxygen-sensor FixL from Rhizobium melliloti: crystallographic, mutagenesis and resonance Raman spectroscopic studies. J Mol Biol 301:415–431. https://doi.org/10.1016/j.jmb.2000.3954.
plasticity of its amphiaphagic ligand-binding pocket. Mol Plant Microbe Interact 33:612–623. https://doi.org/10.1094/MPMI-10-19-0277-R.

113. Nishiyama S-i, Takahashi Y, Yamamoto K, Suzuki D, Itoh Y, Sumita K, Uchida Y, Homma M, Imada K, Kagawishi I. 2016. Identification of a Vibrio cholerae chemoreceptor that senses taurine and amino acids as attractants. Sci Rep 6:20866. https://doi.org/10.1038/srep20866.

114. Liu YC, Machuca MA, Beckham SA, Gunzburg MJ, Roujeinikova A. 2015. Structural basis for amino-acid recognition and transmembrane signaling by tandem Per-Arnt-Sim (tandem PAS) chemoreceptor sensory domains. Acta Crystallographica D Biol Crystallography 71:2127–2136. https://doi.org/10.1107/S139900471501384X.

115. Corral-Lugo A, Matilla MA, Martin-Mora D, Silva Jiménez H, Mesa Torres N, Kato J, Hida A, Oku S, Conejero-Muriel M, Gavira JA, Krell T. 2018. High-affinity chemotaxis to histamine mediated by the TlpQ chemoreceptor of the human pathogen Pseudomonas aeruginosa. mBio 9: e01894–18. https://doi.org/10.1128/mBio.01894-18.

116. Wu R, Wilton R, Cuff ME, Endres M, Babnigg G, Edrisingsjn HEN, Henry, CS, Cho, M, Flakas, F, Kim, H, Chen, S, Lu, G, O’Toole, G, Perrie, P, Wittenburg, J, Huang, B, Li, T, Krell, T. 2017. A novel signal transduction mechanism: combination of solute binding and tandem PAS-like sensor domains in one polypeptide chain. Protein Sci 26:857–869. https://doi.org/10.1002/pro.3134.

117. Gavira JA, Ortega A, Martin-Mora D, Conejero-Muriel MT, Corral-Lugo A, Morel B, Matilla MA, Krell T. 2018. Structural basis for polyamine binding at the dCACH domain of the McpU chemoreceptor from Pseudomonas putida. J Mol Biol 430:1950–1963. https://doi.org/10.1016/j.jmb.2018.05.008.

118. Neidtich MB, Federle MJ, Pompeani AJ, Kelly RC, Jeffery PD, Bassler BL, Hugheson FM. 2006. Ligand-induced asymmetry in histidine sensor kinase complex regulates quorum sensing. Cell 126:1095–1108. https://doi.org/10.1016/j.cell.2006.07.032.

119. Shrestha M, Compton KK, Mancl JM, Webb BA, Brown AM, Scharf BE, Schobut FD. 2018. Structure of the sensory domain of McpX from Sinorhizobium meliloti, the first known bacterial chemotactic sensor for quaternary ammonium compounds. Biochem J 475:3949–3962. https://doi.org/10.1042/BCJ20180769.

120. Hothorn M, Dabi T, Chory J. 2011. Structural basis for cytokinin recognition by Arabidopsis thaliana histidine kinase 4. Nat Chem Biol 7:766–768. https://doi.org/10.1038/nchembio.667.

121. Zhou YF, Nan B, Nan J, Ma P, Panjikar S, Liang YH, Wang Y, Su XD. 2008. C4-Dicarboxylates sensing mechanism revealed by the crystal structures of DctB sensor domain. J Mol Biol 383:49–61. https://doi.org/10.1016/j.jmb.2008.08.010.

122. Wu R, Gu M, Wilton R, Babnigg G, Kim Y, Pukkoli PR, Szurmant H, Joachimiak A, Schiffer M. 2013. Insight into the sporulation phosphate-lipid: crystal structure of the sensor domain of Bacillus subtilis histidine kinase, KinD. Protein Sci 22:564–576. https://doi.org/10.1002/pro.2237.

123. Kato J, Kim H-E, Takiguchi N, Kuroda A, Ohtake H. 2008. Structural basis for chemotaxis to histamine mediated by the TlpQ chemoreceptor of the human pathogen Pseudomonas aeruginosa. Science 323:1867. https://doi.org/10.1126/science.1148799.

124. Kato J, Kim H-E, Takiguchi N, Kuroda A, Ohtake H. 2008. Structural basis for histidine kinase 4. Nat Chem Biol 7:766–768. https://doi.org/10.1038/nchembio.667.

125. Morel B, Ortega A, Ramos JL, Krell T. 2013. Paralogous chemoreceptors provide clues to its evolution in Bacillus subtilis. Mol Microbiol 96:1298–1310. https://doi.org/10.1111/mmi.12964.

126. Hothorn M, Dabi T, Chory J. 2011. Structural basis for cytokinin recognition by Arabidopsis thaliana histidine kinase 4. Nat Chem Biol 7:766–768. https://doi.org/10.1038/nchembio.667.

127. Corral-Lugo A, Matilla MA, Martin-Mora D, Silva Jiménez H, Mesa Torres N, Kato J, Hida A, Oku S, Conejero-Muriel M, Gavira JA, Krell T. 2018. High-affinity chemotaxis to histamine mediated by the TlpQ chemoreceptor of the human pathogen Pseudomonas aeruginosa. mBio 9: e01894–18. https://doi.org/10.1128/mBio.01894-18.

128. Wu R, Wilton R, Cuff ME, Endres M, Babnigg G, Edrisingsjn HEN, Henry, CS, Cho, M, Flakas, F, Kim, H, Chen, S, Lu, G, O’Toole, G, Perrie, P, Wittenburg, J, Huang, B, Li, T, Krell, T. 2017. A novel signal transduction mechanism: combination of solute binding and tandem PAS-like sensor domains in one polypeptide chain. Protein Sci 26:857–869. https://doi.org/10.1002/pro.3134.

129. Gavira JA, Ortega A, Martin-Mora D, Conejero-Muriel MT, Corral-Lugo A, Morel B, Matilla MA, Krell T. 2018. Structural basis for polyamine binding at the dCACH domain of the McpU chemoreceptor from Pseudomonas putida. J Mol Biol 430:1950–1963. https://doi.org/10.1016/j.jmb.2018.05.008.

130. Neidtich MB, Federle MJ, Pompeani AJ, Kelly RC, Jeffery PD, Bassler BL, Hugheson FM. 2006. Ligand-induced asymmetry in histidine sensor kinase complex regulates quorum sensing. Cell 126:1095–1108. https://doi.org/10.1016/j.cell.2006.07.032.

131. Shrestha M, Compton KK, Mancl JM, Webb BA, Brown AM, Scharf BE, Schobut FD. 2018. Structure of the sensory domain of McpX from Sinorhizobium meliloti, the first known bacterial chemotactic sensor for quaternary ammonium compounds. Biochem J 475:3949–3962. https://doi.org/10.1042/BCJ20180769.

132. Hothorn M, Dabi T, Chory J. 2011. Structural basis for cytokinin recognition by Arabidopsis thaliana histidine kinase 4. Nat Chem Biol 7:766–768. https://doi.org/10.1038/nchembio.667.

133. Zhou YF, Nan B, Nan J, Ma P, Panjikar S, Liang YH, Wang Y, Su XD. 2008. C4-Dicarboxylates sensing mechanism revealed by the crystal structures of DctB sensor domain. J Mol Biol 383:49–61. https://doi.org/10.1016/j.jmb.2008.08.010.

134. Wu R, Gu M, Wilton R, Babnigg G, Kim Y, Pukkoli PR, Szurmant H, Joachimiak A, Schiffer M. 2013. Insight into the sporulation phosphate-lipid: crystal structure of the sensor domain of Bacillus subtilis histidine kinase, KinD. Protein Sci 22:564–576. https://doi.org/10.1002/pro.2237.

135. Morel B, Ortega A, Ramos JL, Krell T. 2013. Paralogous chemoreceptors provide clues to its evolution in Bacillus subtilis. Mol Microbiol 96:1298–1310. https://doi.org/10.1111/mmi.12964.
147. Gallagher LA, McKnight SL, Kuznetzova MS, Pesci EC, Manoil C. 2002. Functions required for extracellular quinolone signaling by Pseudomonas aeruginosa. J Bacteriol 184:6472–6480. https://doi.org/10.1128/JB.184.23.6472-6480.2002.

148. Francis VI, Stevenson EC, Porter SL. 2017. Two-component systems required for virulence in Pseudomonas aeruginosa. FEMS Microbiology Lett 364. https://doi.org/10.1093/femsle/fnx104.

149. Li G, Lu C-D. 2016. Molecular characterization of LhpR in control of hydroxyproline catabolism and transport in Pseudomonas aeruginosa PA01. Microbiology (Reading) 162:1232–1242. https://doi.org/10.1099/mic.0.003000.

150. Lundgren BR, Norris HC, Sarwar Z, Scheel RA, Nomura CT. 2015. The metabolism of (R)-3-hydroxybutyrate is regulated by the enhancer-binding protein PA2005 and the alternative sigma factor RpoN in Pseudomonas aeruginosa PA01. Microbiology (Reading) 161:2222–2242. https://doi.org/10.1099/mic.0.000163.

151. Martin-Mora D, Ortega Á, Pérez-Maldonado FJ, Krell T, Matilla MA. 2018. PAS Domains in Pseudomonas aeruginosa are controlled by chemoattractants and antagonists. Sci Rep 8:1–16. https://doi.org/10.1038/s41598-018-20283-7.

152. Beaudoin T, Zhang L, Hinz AJ, Parr CJ, Mah T-F. 2012. The biofilm-specific antibiotic resistance gene ndvB is important for expression of ethanol oxidation genes in Pseudomonas aeruginosa biofilms. J Bacteriol 194:3128–3136. https://doi.org/10.1128/JB.00671-11.

153. Petrova OE, Sauer K. 2009. A novel signaling network essential for regulation of the anaerobic regulator ANR in Pseudomonas aeruginosa. FEMS Microbiol Lett 287:247–247. https://doi.org/10.1016/j.femsle.2009.04.009.

154. Petrova OE, Sauer K. 2011. SagS contributes to the motile-sessile switch of Pseudomonas aeruginosa by regulating cupD expression. Environ Microbiol 13:666–668. https://doi.org/10.1111/j.1462-2920.2010.02372.x.

155. Giraud C, Bernard CS, Calderon V, Yang L, Filloux A, Molin S, Fichant G. 2002. The biofilm-specific antibiotic resistance gene ndvB is important for expression of ethanol oxidation genes in Pseudomonas aeruginosa biofilms. J Bacteriol 184:6472–6480. https://doi.org/10.1128/JB.184.23.6472-6480.2002.

156. Münzen M, Di Fiore S, Gruber R, Häussler R. 2010. Genetic determinants of Pseudomonas aeruginosa biofilm establishment. Microbiology (Reading) 156:431–441. https://doi.org/10.1099/mic.0.003290-0.

157. Bhamse P, Wei Q, Xu A, Naqvi STA, Wang D, Ma LZ. 2020. Evaluation and characterization of the predicted diguanylate cyclase-encoding genes in Pseudomonas aeruginosa. MicrobiologyOpen 9:e9975. https://doi.org/10.1002/mbo3.975.

158. Zhang Y, Guo J, Zhang N, Yuan W, Lin Z, Huang W. 2019. Characterization and analysis of a novel diguanylate cyclase PA0847 from Pseudomonas aeruginosa PA01. Infect Drug Resist 12:655–665. https://doi.org/10.2147/IDR.S194462.

159. Liang H, Li L, Kong W, Shen L, Duan K. 2009. Identification of a novel regulator of the quorum-sensing systems in Pseudomonas aeruginosa. FEMS Microbiol Lett 293:196–204. https://doi.org/10.1111/j.1574-6968.2009.01544.x.

160. Lundgren BR, Bailey FI, Moley G, Nomura CT. 2017. DdaR (PA1196) regulates expression of dimethylarginine dimethylaminohydrolase for the metabolism of methylarginines in Pseudomonas aeruginosa PA01. J Bacteriol 199:e00001-17. https://doi.org/10.1128/JB.00001-17.

161. Badal D, Jayarani AV, Kollaran MA, Kumar A, Singh V. 2020. Pseudomonas aeruginosa biofilm formation on endotracheal tubes requires multiple two-component systems. J Med Microbiol 69:906–919. https://doi.org/10.1099/jmm.0.001199.

162. Ranjttkar S, Jones AK, Mostafavi M, Zvirko Z, Ichtouch O, Whitney Barnes S, Walker JR, Willis TW, Lee PS, Dean CR. 2019. Target (MexB) and efflux-based mechanisms decreasing the effectiveness of the efflux pump inhibitor D13-9001 in Pseudomonas aeruginosa PA01: uncovering a new role for MexM-OprM in efflux of β-lactams and a novel regulatory circuit (MmmRS) controlling Me. Antimicrob Agents Chemother 63:e01718-18. https://doi.org/10.1128/AAC.01718-18.

163. Petrova OE, Sauer K. 2009. A novel signaling network essential for regulating Pseudomonas aeruginosa biofilm development. PLoS Pathog 5:e1000668. https://doi.org/10.1371/journal.ppat.1000668.

164. Petrova OE, Sauer K. 2010. The novel two-component regulatory system BfsR regulates biofilm development by controlling the small RNA rsmZ through Ca²⁺. J Bacteriol 192:5275–5288. https://doi.org/10.1128/JB.00305-11.

165. Petrova OE, Sauer K. 2011. SagS contributes to the motile-sessile switch and acts in concert with BfsR to enable Pseudomonas aeruginosa biofilm formation. J Bacteriol 193:6614–6628. https://doi.org/10.1128/JB.00305-11.

166. Giraud C, Bernard CS, Calderon V, Yang L, Filloux A, Molin S, Fichant G, Bordi C, de Bentzmann S. 2011. The PprR-PprP two-component system activates CufE, the first non-archetypal Pseudomonas aeruginosa chaperone– usher pathway system assembling fimbriae. Environ Microbiol 13:666–683. https://doi.org/10.1111/j.1462-2920.2010.02372.x.

167. de Bentzmann S, Giraud C, Bernard CS, Calderon V, Ewald F, Pléiat P, Nguyen C, Grunwald D, Attiret I, Jeannot K, Fauvarque M-O, Bordi C. 2012. Unique biofilm signature, drug susceptibility and decreased virulence in Drosophila through the Pseudomonas aeruginosa two-component system PprPAB. PLoS Pathog 8:e1003052. https://doi.org/10.1371/journal.ppat.1003052.

168. Genschik P, Drabikowski K, Filipowicz W. 1998. Characterization of the Escherichia coli RNA 3′-terminal phosphate cyclase and its sigma54-regulated operon. J Biol Chem 273:25516–25526. https://doi.org/10.1074/jbc.273.39.25516.

169. Winsor GL, Lam DKW, Fleming L, Lo R, Whiteside MD, Yu NY, Hancock REW, Brinkman FSL. 2011. Pseudomonas Genome Database: improved annotation and analysis of a novel diguanylate cyclase PA0847 from Pseudomonas aeruginosa. J Bacteriol 193:273.39.25516.39.25526.

170. Choy W-K, Zhou L, Syn CK-C, Zhang L-H, Swarup S. 2004. MorA deactivates virulence-associated protease secretion in Pseudomonas aeruginosa PA01. PLoS One 10:e0123805. https://doi.org/10.1371/journal.pone.0123805.

171. Waite RD, Papakonstantinopoulou A, Littler E, Curtis MA. 2005. Transcriptome analysis of Pseudomonas aeruginosa growth: comparison of gene expression in planktonic cultures and developing and mature biofilms. J Bacteriol 187:6571–6576. https://doi.org/10.1128/JB.187.18.6571-6576.2005.
fibrosis airway isolates TBCF10839 and TBCF121838b, distinct signatures of transcriptome, proteome, metabolome, adhesion and pathogenicity despite an almost identical genome sequence. Environ Microbiol 15:191–210. https://doi.org/10.1111/1462-2920.2012.02842.x

181. Li Y, Xia H, Bai F, Xu H, Yang L, Yao H, Zhang L, Zhang X, Bai Y, Saris PEJ, Miller SI. 2013. c-di-GMP heterogeneity is generated by the chemotaxis RalB protein and the chemotaxis response regulator CzcS, a two-component system involved in heavy metal and carbapenem resistance in Pseudomonas aeruginosa. J Biol Chem 279:8761–8768. https://doi.org/10.1074/jbc.M312080200.

182. Roy AB, Petrova OE, Sauer K. 2012. The phosphodiesterase DipA (PA5017) is essential for Pseudomonas aeruginosa biofilm dispersion. J Bacteriol 194:2904–2915. https://doi.org/10.1128/JB.00346-11.

183. Barkovits K, Schubert B, Heine S, Scheer M, Frankenberg-Dinkel N. 2011. The absence of the Pseudomonas aeruginosa OprF protein leads to increased biofilm formation through variation in c-di-GMP level. Front Microbiol 6:630. https://doi.org/10.3389/fmicb.2015.00630.

184. Hassan MT, van der Lelie D, Springael D, Römling U, Ahmed N, Mergeay M, Barkovits K, Schubert B, Heine S, Scheer M, Frankenberg-Dinkel N, Barkovits K, Schubert B, Heine S, Scheer M, Frankenberg-Dinkel N. 2011. The absence of the Pseudomonas aeruginosa OprF protein leads to increased biofilm formation through variation in c-di-GMP level. Front Microbiol 6:630. https://doi.org/10.3389/fmicb.2015.00630.

185. Luu RA, Schneider BJ, Ho CC, Netheryuk V, Ngyesse SE, Liu X, Paraes JVL, Ditty JL, Parales RE. 2013. Taxis of Pseudomonas putida F1 toward phenylacetic acid is mediated by the energy taxis receptor AER2. Appl Environ Microbiol 79:2416–2423. https://doi.org/10.1128/AEM.03895-12.

186. Caille O, Rossier C, Perron K. 2007. A copper-activated two-component system interacts with zinc and impenem resistance in Pseudomonas aeruginosa. J Bacteriol 189:4561–4568. https://doi.org/10.1128/JB.00095-07.

187. Yeung ATY, Torfs ECW, Jamshidi F, Bains M, Wiegand I, Hancock REW, Merrick MJ, Edwards RA. 1995. Nitrogen control in bacteria. Microbiol Rev 59:604–622. https://doi.org/10.1128/mr.59.4.604-622.1995.

188. Alford MA, Baghela A, Yeung ATY, Pletzer D, Hancock REW. 2020. NtrBC and NtrRBC two-component regulatory system controls the utilization of multiple carbon and nitrogen sources in Pseudomonas aeruginosa. Mol Microbiol 40:917–931. https://doi.org/10.1111/1365-2958.2020.02435.x.

189. Li W, Lu CD. 2007. Regulation of carbon and nitrogen utilization by the CrbS/R two-component system in Pseudomonas aeruginosa. FEBS J 272:1927–1936. https://doi.org/10.1111/j.1742-4658.2005.04623.x.

190. Yeung ATY, Torfs ECW, Jamshidi F, Bains M, Wiegand I, Hancock REW, Merrick MJ, Edwards RA. 1995. Nitrogen control in bacteria. Microbiol Rev 59:604–622. https://doi.org/10.1128/mr.59.4.604-622.1995.

191. Li W, Lu CD. 2007. Regulation of carbon and nitrogen utilization by the CrbS/R two-component system in Pseudomonas aeruginosa. FEBS J 272:1927–1936. https://doi.org/10.1111/j.1742-4658.2005.04623.x.

192. Caille O, Rossier C, Perron K. 2007. A copper-activated two-component system interacts with zinc and impenem resistance in Pseudomonas aeruginosa. J Bacteriol 189:4561–4568. https://doi.org/10.1128/JB.00095-07.

193. Barkovits K, Schubert B, Heine S, Scheer M, Frankenberg-Dinkel N. 2011. Function of the bacteriophytochrome Bhp in the RpoS/Las quorum-sensing network of Pseudomonas aeruginosa. Microbiology (Reading) 157:1651–1664. https://doi.org/10.1099/mic.0.049007-0.

194. Roy AB, Petrova OE, Sauer K. 2012. The phosphodiesterase DipA (PA5017) is essential for Pseudomonas aeruginosa biofilm dispersion. J Bacteriol 194:2904–2915. https://doi.org/10.1128/JB.00346-11.

195. Luu RA, Schneider BJ, Ho CC, Netheryuk V, Ngyesse SE, Liu X, Paraes JVL, Ditty JL, Parales RE. 2013. Taxis of Pseudomonas putida F1 toward phenylacetic acid is mediated by the energy taxis receptor AER2. Appl Environ Microbiol 79:2416–2423. https://doi.org/10.1128/AEM.03895-12.

196. Merrick MJ, Edwards RA. 1995. Nitrogen control in bacteria. Microbiol Rev 59:604–622. https://doi.org/10.1128/mr.59.4.604-622.1995.

197. Nishiyoi T, Haas D, Itoh Y. 2001. The CbrA-CbrB two-component regulatory system controls the utilization of multiple carbon and nitrogen sources in Pseudomonas aeruginosa. Mol Microbiol 40:917–931. https://doi.org/10.1111/1365-2958.2001.02435.x.

198. Li W, Lu CD. 2007. Regulation of carbon and nitrogen utilization by the CrbS/R two-component system in Pseudomonas aeruginosa. FEBS J 272:1927–1936. https://doi.org/10.1111/j.1742-4658.2005.04623.x.

199. Yeung ATY, Torfs ECW, Jamshidi F, Bains M, Wiegand I, Hancock REW, Merrick MJ, Edwards RA. 1995. Nitrogen control in bacteria. Microbiol Rev 59:604–622. https://doi.org/10.1128/mr.59.4.604-622.1995.

200. Alford MA, Baghela A, Yeung ATY, Pletzer D, Hancock REW. 2020. NtrRC regulates invasiveness and virulence of Pseudomonas aeruginosa during high-density infection. Front Microbiol 11:773. https://doi.org/10.3389/fmicb.2020.00773.