JOURNAL OF BACTERIOLOGY, Feb. 2010, p. 1156–1159 Vol. 192, No. 4
0021-9193/10/$12.00 doi:10.1128/JB.01508-09

Extracytoplasmic PAS-Like Domains Are Common in Signal Transduction Proteins▽†

Changsoo Chang,[1] Christine Tesar,[1] Minyi Gu,[1] Gyorgy Babnigg,[1] Andrzej Joachimiak,[1] P. Raj Pokkuluri,[1] Hendrik Szurmant,[2]* and Marianne Schiffer[1]*

Biosciences Division, Argonne National Laboratory, Argonne, Illinois 60439,[1] and Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, California 92037[2]

Received 18 November 2009/Accepted 19 November 2009

We present the crystal structure of the extracytoplasmic domain of the *Bacillus subtilis* PhoR sensor histidine kinase, part of a two-component system involved in adaptation to low environmental phosphate concentrations. In addition to the PhoR structure, we predict that the majority of the extracytoplasmic domains of *B. subtilis* sensor kinases will adopt a fold similar to the ubiquitous PAS domain.

Transmembrane signal transduction in bacteria is primarily mediated by the so-called two-component signal transduction system. To generate an appropriate response to an environmental signal, a sensor histidine kinase (SK) adjusts its intrinsic autokinase activity upon detection of ligands. Subsequent transfer of the phosphoryl group alters the propensity of the associated response regulator/transcription factor (RR) to generate the appropriate response (11).

SK proteins are modular multidomain proteins featuring a structurally conserved C-terminal catalytic core and N-terminal signal-sensing and activity-modulating domains (31). The C-terminal catalytic core consists of two domains, the phosphorylatable histidine containing a four-helix bundle domain (called DHp or HisKA domain) and an ATP-binding domain (called HATPase or CA domain). The autokinase activity of the catalytic core is controlled by the N terminus, which is variable in length and domain architecture. Since the majority of SKs are transmembrane proteins, they commonly feature signal detection and activity-modulating domains localized to both the extracytoplasmic space and the cytoplasm.

The cytoplasmic input and activity-modulating domains of the SKs are readily identifiable by sequence analysis, and the dominant folds are the HAMP, GAF, and PAS domains (reviewed in reference 31). The PAS fold is a particularly widespread and versatile protein domain fold, is commonly found in signaling proteins, and has been implicated in small-molecule binding as well as protein-protein interaction (21, 32). The Pfam database (7) reveals over 20,000 examples in more than 1,300 organisms.

The extracytoplasmic domains in contrast to cytoplasmic portions of transmembrane-embedded SKs show very low sequence conservation. With a few exceptions, the extracytoplasmic sensing domains cannot be categorized by sequence analysis (11, 31, 35). This hampers progress in the identification of signaling ligands, which for most SKs still remain unknown.

Here, we present a structure prediction analysis of the complete set of extracytoplasmic sensor domains of the *Bacillus subtilis* SK proteins. This analysis suggests that the PAS-like fold (10, 21) is the dominant sensing module adopted in *B. subtilis* SKs. To back up this prediction, we present the crystal structure of the *B. subtilis* PhoR extracytoplasmic domain, which adopts the expected PAS-like fold.

B. subtilis two-component systems have been the subject of extensive studies in the last 3 decades. Some of these studies have identified growth conditions under which individual systems are activated (e.g., cell envelope stress, phosphate limitation, nutrient deprivation) (6, 11). However, the molecular ligands sensed by these kinases remain unknown for all but the MalK and CitS proteins, malate and citrate sensors, respectively (5, 33). The complete set of *B. subtilis* SKs was previously analyzed by Fabret and colleagues (6). Of the 35 SKs, five were found to be soluble, whereas 30 contained at least one identifiable transmembrane helix. In the original work, the extracytoplasmic portions of the transmembrane histidine kinases did not receive any further attention.

We recently determined the structures of two periplasmic sensor domains of chemotaxis receptor proteins (Mcp's) from the Gram-negative bacterium *Geobacter sulfurreducens* (26) and found that they adopted a PAS-like fold, as we had previously predicted (20), by using the structure prediction program 3DPSSM (17). This program combines one- and three-dimensional amino acid sequence profiles with secondary structure predictions and solvation potential. The structural folds could not be identified using pure sequence comparison methods, such as BLAST. This prompted us to revisit the analysis of the *B. subtilis* SK sequences. We identified 13 sensor kinases that have an extracytoplasmic domain containing more than 40 amino acids located between two transmembrane helices near the N terminus. The exact boundaries of the transmembrane helices were predicted by using the program TMHMM (19). These domains were analyzed by applying a newer fold recognition program, FFAS03 (15). FFAS03 uses an amino acid sequence profile-profile comparison method for

* Corresponding author. Mailing address for M. Schiffer: 9700 S. Cass Avenue, Bldg. 202, A-145, Argonne, IL 60439. Phone: (630) 252-3883. Fax: (630) 252-3387. E-mail: mschiffer@anl.gov. Mailing address for H. Szurmant: MEM-116, 10550 N. Torrey Pines, La Jolla, CA 92037. Phone: (858) 784-7945. Fax: (858) 784-7966. E-mail: Szurmant@scripps.edu.

† Supplemental material for this article may be found at http://jb.asm.org/.

▽ Published ahead of print on 11 December 2009.

TABLE 1. Structure predictions for extracytoplasmic domains of *B. subtilis* SKs

Protein	Residues[a]	Function	Reference	Top structure hit[b]	Score (% identity)[c]	Predicted fold
PhoR	32–150	Phosphate limitation	13	*K. pneumoniae* CitA	**−16.2** (12)	PAS-like
ResE	34–172	Oxygen limitation	23	*B. subtilis* PhoR	**−12.0** (16)	PAS-like
YycG	35–182	Cell wall/cell division	8	*E. coli* DcuS	**−37.9** (9)	PAS-like
CitS	38–173	Citrate uptake/metabolism	33	*E. coli* DcuS	**−49.0** (18)	PAS-like
MalK	33–172	Malate uptake/metabolism	5	*E. coli* DcuS	**−59.0** (31)	PAS-like
KinD	37–250	Sporulation	16	*V. cholerae* Mcp	**−55.1** (13)	2× PAS-like[d,e]
CssS	30–166	Secretion stress response	3	*B. subtilis* PhoR	**−11.4** (8)	PAS-like
DctS	36–173	Dicarboxylic acid uptake	14	*E. coli* DcuS	**−60.2** (28)	PAS-like
YkoH	30–153	Unknown	6	*E. coli* DcuS	**−11.8** (13)	PAS-like
YclK	33–162	Unknown	6	*B. subtilis* PhoR	**−40.2** (11)	PAS-like
YesM	42–284	Unknown	6	*V. cholerae* Mcp	**−49.7** (10)	2× PAS-like[d]
YvrG	27–249	Cell surface maintenance	28	*E. coli* RNAP β-chain	−7.16	?
YrkQ	35–144	Unknown	6	*H. sapiens* cathepsin D	−6.62	?

[a] Extracytoplasmic residue stretch as predicted by the TMHMM program (19).
[b] Top structure hit identified by the Fold and Function Assignment System (FFAS) (15).
[c] Score and sequence identity as identified by the FFAS03 program. A score lower than −9.5 in boldface type is considered significant, with <3% false positive hits.
[d] The "double PAS-like" (2× PAS-like) fold has been observed in the LuxQ and DctB SKs (2, 24), and a number of other such structures have been deposited in the PDB.
[e] The structure has been recently determined (PDB code, 3fos) and is as predicted.

protein families. The program predicted that 11 out of the 13 sensors would have PAS-like domain folds (Table 1), whereas the structures of the other two domains were not reliably predicted. Some low-scoring hits for these two proteins, however, were PAS-like domains, not excluding the possibility that all 13 extracytoplasmic sensor domains adopt the PAS-like fold. An analogous analysis conducted for other organisms reveals that at least two out of five *Streptococcus pneumoniae*, at least three out of six *Staphylococcus aureus*, and 11 out of 21 *Escherichia coli* SK extracytoplasmic domains are likely to adopt the PAS-like fold as well (see Table S1 in the supplemental material).

To verify the above-described predictions, we have undertaken a study to determine the structures of the extracytoplasmic domains of the 13 proteins from *B. subtilis* by X-ray diffraction. Diffraction quality crystals could be obtained for one of these extracytoplasmic domains from the phosphate limitation sensor, PhoR. The PhoR extracytoplasmic domain, residues 32 to 150 from the sensor histidine kinase (GeneID no. 16079962), was expressed in *E. coli* by using vector pMCSG7 (30). The protein was purified using previously described procedures (18) and was crystallized from Index Screen, condition G4 (Hampton Research). The crystals were orthorhombic with space group of $P2_12_12_1$ and have unit cell dimensions of $a = 55.50$, $b = 67.63$, and $c = 77.69$ Å. The crystals diffracted to 2.2-Å resolution. Crystallographic data were collected at the Structural Biology Center (SBC) beamline 19BM at the Advanced Photon Source (APS). The structure was determined using anomalous scattering from bromide ions by using the quick-halide soak method (4). The asymmetric unit contains two molecules that form a dimer. Electron density is observed for residues 33 to 141 in both monomers. The structure was refined with the program REFMAC (22), and the final coordinates were deposited in the Protein Data Bank (PDB accession code 3cwf). The *R* factor and the free *R* factor are 19.4% and 24.4%, respectively. The Ramachandran plot shows that 90.3% of the residues are in the most favored regions, while none are in the disallowed regions.

The structure comparison program, DaliLite (12), identified the sensor domain of CitA, a citrate sensor from *Klebsiella pneumoniae* (27), as the closest structural homologue of PhoR in the PDB. After structural alignment, the two domains share only 12% sequence identity (Fig. 1a). The most significant structural differences (Fig. 1b and c) are in the elements surrounding the citrate-binding pocket in CitA. In particular, the loop connecting β-strands 3 and 4, which is utilized in the CitA sensor as a lid for the binding pocket, is shorter in sequence and is observed to be in an open conformation in the structure of PhoR sensor. Similarly, an extended loop-helix-loop motif between β-strands 2 and 3 covers the ligand-binding pocket in the CitA sensor and differs in both length and positioning between the two structures (Fig. 1b and c). This suggests that PhoR either might bind a ligand of very different nature or perhaps might not serve a small-molecule-binding purpose at all. Consistent with this notion, Shi and Hulett have shown that this domain is not required for the phosphate limitation response elicited by PhoR (29). It might be argued then that the extracytoplasmic domains in most SKs have no function. This is unlikely in the light that despite astounding sequence variability, so many SK sensor domains have retained a conserved structural fold. It is conceivable that new functions for these sensor domains beyond sensing small-molecule ligands will come to light.

Currently available structures of the extracytoplasmic PAS-like sensor domains share a central five-stranded β-sheet of identical topology with other PAS domains (Fig. 1d). Differences observed between these domains have recently been pointed out by Cheung et al. (1). They suggested that these domains comprise a fold "distinct from the PAS domain structures with which they share a β-sheet topology" and named them "PDC" fold based on previously determined structures (PhoQ, DcuS, CitA). Given the similarity of the folds of the extracytoplasmic PAS-like domains and other PAS domains and given their conserved function in regulating the activity of SKs and other signal transduction proteins, often in response to small-molecule ligands, it is more than likely that they share an evolutionary origin. As such, we do not see the necessity to define them as a separate fold and feel that they should be

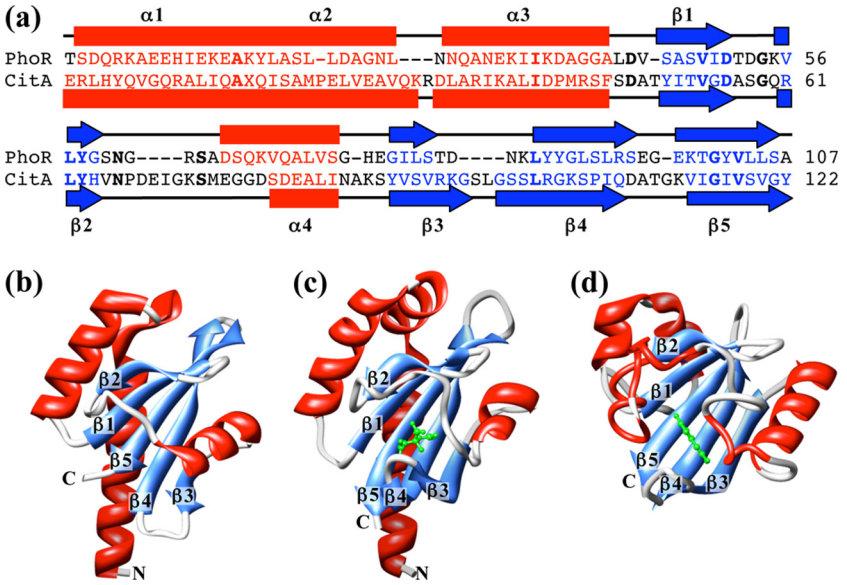

FIG. 1. The extracytoplasmic domain of PhoR forms a PAS-like fold. (a) Structural alignment of the *B. subtilis* PhoR extracytoplasmic domain with that of *K. pneumoniae* CitA. Despite a very low number of identical residues in the alignment (in boldface type), secondary structure elements are very similar in position and length (red for α-helix, blue for β-strand). (b) Structure of PhoR sensor domain (PDB code 3cwf). (c) Structure of CitA sensor domain (PDB code 2j80). Main structural differences are found surrounding the ligand-binding pocket of CitA (citrate in green), suggesting that PhoR binds a ligand of very different nature, or might not serve a ligand binding purpose at all. (d) For comparison, the PAS domain of photoactive yellow protein (PYP) as a representative of cytoplasmic PAS domains is presented (PDB code 2qj5). It features an identical strand topology and very similar structure of the central β-sheet. The location of the ligand-binding pocket (citrate and hydroxycinnamic acid molecules in green) is also conserved between the CitA PAS-like and PYP PAS domains. The extensive loop and helix elements inserted between strands β2 and β3, flanking the ligand-binding pocket, differ in length and position. The long N-terminal helix in the extracytoplasmic PAS-like domains, an extension of the N-terminal transmembrane helix, is absent in cytoplasmic PAS domains. β-Strands are shown in blue and α-helices in red. For clarity, some PYP helices in the region connecting strands β2 and β3 are shown in red but not depicted as helices.

considered to be part of the clan of PAS folds. Previous publications about this domain have also considered it a PAS domain (10, 21).

Naturally, localization to different cellular compartments as well as response to differing ligands dictates diverse requirements that account for the major differences observed for the extracytoplasmic PAS domains. Case in point is one of the most obvious structural differences found at the N terminus, preceding the first strand of the central β-sheet. In the extracytoplasmic PAS-like domains, the N terminus forms a long α-helix, which may be envisioned as an extension of the N-terminal transmembrane helix. Physically, such an element is not required for cytoplasmic PAS domains, which are commonly found in the center of modular SK proteins in between other N- and C-terminal domains. The other major structural difference is observed around the elements that form the ligand-binding pockets. The number of residues between β-strands 2 and 3 is smaller in the extracytoplasmic PAS-like domain (<40 residues) than in the Pfam-annotated PAS domains (~60 residues). This is perhaps not surprising, since diversity in this region could merely reflect the capacity of this domain fold to bind and respond to a variety of different ligands.

It is expected that there are large numbers of extracytoplasmic PAS-like domains present in bacterial SK signal transduction proteins that are not yet identified due to the inherent low sequence identity and the fact that they are consequently not defined as distinct domains in the Pfam database. This is supported, not only by our current analysis, but also by an increasing number of such structures, part of SK proteins from various organisms, available in the PDB (e.g., see references 24, 25, 27, and 34). Some recent publications and new structures deposited into the PDB demonstrate that extracytoplasmic PAS-like domains are not confined to SK proteins but are also part of Mcp proteins we studied previously (26), other Mcp proteins (PDB codes 3c38 and 2qhk), and a GGDEF signaling protein (PDB code 2p7j). Homology modeling has revealed two tandem PAS domains for the extracytoplasmic sensing domain of *B. subtilis* asparagine chemotaxis receptor McpB (9). We conclude that the prevalence of PAS-like sensor domains in signal transduction proteins is highly underestimated using search algorithms based on sequence alone. Our limited analysis suggests that at least in the Gram-positive model organism *B. subtilis* and the Gram-negative model organism *E. coli*, the PAS-like domain is the dominant extracytoplasmic signaling detection module for SK proteins. We hypothesize that this is likely to be true across the microbial kingdom.

This work was supported by the following: National Institutes of Health (grant GM62414-01) and the U.S. Department of Energy, Office of Science, Biological and Environmental Research Genomics: GTL program, under contract no. DE-AC02-06CH11357. H.S. was supported by National Institutes of Health grant GM019416 awarded to J. A. Hoch. The use of SBC beamlines and APS are supported by the DOE under contract no. DE-AC02-06CH11357.

The manuscript was created by UChicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne"). Argonne, a U.S. De-

partment of Energy Office of Science laboratory, is operated under contract no. DE-AC02-06CH11357.

REFERENCES

1. **Cheung, J., C. A. Bingman, M. Reyngold, W. A. Hendrickson, and C. D. Waldburger.** 2008. Crystal structure of a functional dimer of the PhoQ sensor domain. J. Biol. Chem. **283:**13762–13770.
2. **Cheung, J., and W. A. Hendrickson.** 2008. Crystal structures of C4-dicarboxylate ligand complexes with sensor domains of histidine kinases DcuS and DctB. J. Biol. Chem. **283:**30256–30265.
3. **Darmon, E., D. Noone, A. Masson, S. Bron, O. P. Kuipers, K. M. Devine, and J. M. van Dijl.** 2002. A novel class of heat and secretion stress-responsive genes is controlled by the autoregulated CssRS two-component system of *Bacillus subtilis*. J. Bacteriol. **184:**5661–5671.
4. **Dauter, Z., M. Dauter, and K. R. Rajashankar.** 2000. Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. Acta Crystallogr. D Biol. Crystallogr. **56:**232–237.
5. **Doan, T., P. Servant, S. Tojo, H. Yamaguchi, G. Lerondel, K. Yoshida, Y. Fujita, and S. Aymerich.** 2003. The Bacillus subtilis *ywkA* gene encodes a malic enzyme and its transcription is activated by the YufL/YufM two-component system in response to malate. Microbiology **149:**2331–2343.
6. **Fabret, C., V. A. Feher, and J. A. Hoch.** 1999. Two-component signal transduction in *Bacillus subtilis*: how one organism sees its world. J. Bacteriol. **181:**1975–1983.
7. **Finn, R. D., J. Mistry, B. Schuster-Bockler, S. Griffiths-Jones, V. Hollich, T. Lassmann, S. Moxon, M. Marshall, A. Khanna, R. Durbin, S. R. Eddy, E. L. Sonnhammer, and A. Bateman.** 2006. Pfam: clans, web tools and services. Nucleic Acids Res. **34:**D247–D251.
8. **Fukushima, T., H. Szurmant, E. J. Kim, M. Perego, and J. A. Hoch.** 2008. A sensor histidine kinase co-ordinates cell wall architecture with cell division in Bacillus subtilis. Mol. Microbiol. **69:**621–632.
9. **Glekas, G. D., R. M. Foster, J. R. Cates, J. A. Estrella, M. J. Wawrzyniak, C. V. Rao, and G. W. Ordal.** 2009. A pas domain binds asparagine in the chemotaxis receptor MCPB in Bacillus subtilis. J. Biol. Chem. [Epub ahead of print.] doi:10.1074/jbc.M109.072108.
10. **Hefti, M. H., K.-J. Francoijs, S. C. de Vries, R. Dixon, and J. Vervoort.** 2004. The PAS fold. A redefinition of the PAS domain based upon structural prediction. Eur. J. Biochem. **271:**1198–1208.
11. **Hoch, J. A.** 2000. Two-component and phosphorelay signal transduction. Curr. Opin. Microbiol. **3:**165–170.
12. **Holm, L., S. Kaariainen, P. Rosenstrom, and A. Schenkel.** 2008. Searching protein structure databases with DaliLite v. 3. Bioinformatics **24:**2780–2781.
13. **Hulett, F. M.** 1996. The signal-transduction network for Pho regulation in Bacillus subtilis. Mol. Microbiol. **19:**933–939.
14. **Janausch, I. G., E. Zientz, Q. H. Tran, A. Kroger, and G. Unden.** 2002. C4-dicarboxylate carriers and sensors in bacteria. Biochim. Biophys. Acta **1553:**39–56.
15. **Jaroszewski, L., L. Rychlewski, Z. Li, W. Li, and A. Godzik.** 2005. FFAS03: a server for profile–profile sequence alignments. Nucleic Acids Res. **33:** W284–W288.
16. **Jiang, M., W. Shao, M. Perego, and J. A. Hoch.** 2000. Multiple histidine kinases regulate entry into stationary phase and sporulation in Bacillus subtilis. Mol. Microbiol. **38:**535–542.
17. **Kelley, L. A., R. M. MacCallum, and M. J. Sternberg.** 2000. Enhanced genome annotation using structural profiles in the program 3D-PSSM. J. Mol. Biol. **299:**499–520.
18. **Kim, Y., I. Dementieva, M. Zhou, R. Wu, L. Lezondra, P. Quartey, G. Joachimiak, O. Korolev, H. Li, and A. Joachimiak.** 2004. Automation of protein purification for structural genomics. J. Struct. Funct. Genomics **5:**111–118.
19. **Krogh, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer.** 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. **305:**567–580.
20. **Londer, Y. Y., I. S. Dementieva, C. A. D'Ausilio, P. R. Pokkuluri, and M. Schiffer.** 2006. 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. **258:**173–181.
21. **Möglich, A., R. A. Ayers, and K. Moffat.** 2009. Structure and signaling mechanism of Per-ARNT-Sim domains. Structure **17:**1282–1294.
22. **Murshudov, G. N., A. A. Vagin, and E. J. Dodson.** 1997. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. **53:**240–255.
23. **Nakano, M. M., and F. M. Hulett.** 1997. Adaptation of Bacillus subtilis to oxygen limitation. FEMS Microbiol. Lett. **157:**1–7.
24. **Neiditch, M. B., M. J. Federle, A. J. Pompeani, R. C. Kelly, D. L. Swem, P. D. Jeffrey, B. L. Bassler, and F. M. Hughson.** 2006. Ligand-induced asymmetry in histidine sensor kinase complex regulates quorum sensing. Cell **126:**1095–1108.
25. **Pappalardo, L., I. G. Janausch, V. Vijayan, E. Zientz, J. Junker, W. Peti, M. Zweckstetter, G. Unden, and C. Griesinger.** 2003. The NMR structure of the sensory domain of the membranous two-component fumarate sensor (histidine protein kinase) DcuS of Escherichia coli. J. Biol. Chem. **278:**39185–39188.
26. **Pokkuluri, P. R., M. Pessanha, Y. Y. Londer, S. J. Wood, N. E. Duke, R. Wilton, T. Catarino, C. A. Salgueiro, and M. Schiffer.** 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.
27. **Reinelt, S., E. Hofmann, T. Gerharz, M. Bott, and D. R. Madden.** 2003. The structure of the periplasmic ligand-binding domain of the sensor kinase CitA reveals the first extracellular PAS domain. J. Biol. Chem. **278:**39189–39196.
28. **Salzberg, L. I., and J. D. Helmann.** 2008. Phenotypic and transcriptomic characterization of *Bacillus subtilis* mutants with grossly altered membrane composition. J. Bacteriol. **190:**7797–7807.
29. **Shi, L., and F. M. Hulett.** 1999. The cytoplasmic kinase domain of PhoR is sufficient for the low phosphate-inducible expression of pho regulon genes in Bacillus subtilis. Mol. Microbiol. **31:**211–222.
30. **Stols, L., M. Gu, L. Dieckman, R. Raffen, F. R. Collart, and M. I. Donnelly.** 2002. A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. Protein Expr. Purif. **25:**8–15.
31. **Szurmant, H., R. A. White, and J. A. Hoch.** 2007. Sensor complexes regulating two-component signal transduction. Curr. Opin. Struct. Biol. **17:**706–715.
32. **Taylor, B. L., and I. B. Zhulin.** 1999. PAS domains: internal sensors of oxygen, redox potential, and light. Microbiol. Mol. Biol. Rev. **63:**479–506.
33. **Yamamoto, H., M. Murata, and J. Sekiguchi.** 2000. The CitST two-component system regulates the expression of the Mg-citrate transporter in Bacillus subtilis. Mol. Microbiol. **37:**898–912.
34. **Zhou, Y. F., B. Nan, J. Nan, Q. Ma, S. Panjikar, Y. H. Liang, Y. Wang, and X. D. Su.** 2008. C4-dicarboxylates sensing mechanism revealed by the crystal structures of DctB sensor domain. J. Mol. Biol. **383:**49–61.
35. **Zhulin, I. B., A. N. Nikolskaya, and M. Y. Galperin.** 2003. Common extracellular sensory domains in transmembrane receptors for diverse signal transduction pathways in *Bacteria* and *Archaea*. J. Bacteriol. **185:**285–294.