Published in final edited form as:

Curr Opin Microbiol. 2010 April ; 13(2): 150–159. doi:10.1016/j.mib.2010.01.005.

Diversity of Structure and Function of Response Regulator Output Domains

Michael Y. Galperin

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

Summary

Response regulators (RRs) within two-component signal transduction systems control a variety of cellular processes. Most RRs contain DNA-binding output domains and serve as transcriptional regulators. Other RR types contain RNA-binding, ligand-binding, protein-binding or transporter output domains and exert regulation at the transcriptional, post-transcriptional or post-translational levels. In a significant fraction of RRs, output domains are enzymes that themselves participate in signal transduction: methyltransferases, adenylate or diguanylate cyclases, c-di-GMP-specific phosphodiesterases, histidine kinases, serine/threonine protein kinases and protein phosphatases. In addition, there remain output domains whose functions are still unknown. Patterns of the distribution of various RR families are generally conserved within key microbial lineages and can be used to trace adaptations of various species to their unique ecological niches.

Keywords

protein domains; transcriptional regulation; protein phosphorylation; signal transduction; genome annotation; protein structure

Introduction

Bacterial adaptation to changing environmental conditions can be thought of as occurring on several different levels: 1) the level of individual genes and proteins (changes in gene expression, allosteric regulation of enzyme activity), 2) the level of global regulons (expression of multiple operons, stress response), 3) the whole-cell level (cellular motility, sporulation), and 4) the multicellular level (cell aggregation, biofilm formation). Two-component signal transduction systems (TCSs) affect processes at all these levels, primarily through transcriptional, post-transcriptional and post-translational regulation of gene expression, but also through a variety of protein-protein interactions. These regulatory processes are performed by various response regulators (RRs) that all share the common phosphoacceptor (receiver, REC) domain but differ in their output domains. The sheer number and diversity of RRs mirror the diversity of functions that they perform. At the time of this writing, public databases include more than 70,000 protein sequences that contain the REC domain; in the latest release of the Pfam database [•1] these sequences are classified into 1716 domain architectures. Although the latter figure primarily reflects the diversity of hybrid histidine kinases, it does not include combinations of REC with uncharacterized domains. Thus, the easiest way to investigate the diversity of RRs is by browsing various genomic databases, including those specifically dedicated to signal transduction. Such databases include Microbial Signal Transduction database (MiST,

*Corresponding author: Galperin, Michael (galperin@ncbi.nlm.nih.gov).

<http://genomics.ornl.gov/mist/>, [2]), whose updated version was recently released at a new web site <http://mistdb.com> [3], and P2CS (<http://www.p2cs.org/>), a database of prokaryotic two-component systems [4]. The author of this review maintains a manually curated listing of the RRs encoded in a non-redundant collection of prokaryotic genomes and classified based on their domain architectures,

http://www.ncbi.nlm.nih.gov/Complete_Genomes/RRcensus.html. This listing has been recently updated and now covers all bacterial and archaeal species whose genomes were released by the end of 2009. Compared to the original 200-genome set [5], the current list provides a much better coverage of microbial diversity, particularly for such bacterial phyla as Aquificae, Bacteroidetes, Chlorobi, Chloflexi, Thermotogae, and Verrucomicrobia, as well as for delta and epsilon subdivisions of Proteobacteria. However, the key trends deduced from the smaller genome sample [5] still appear to hold true. Here, I briefly review the key rationale for RR classification [5,6] and discuss the growing list of RR output domains, as well as the recent progress in their structural and functional characterization.

Classification in lieu of annotation

Although most RRs combine the REC domain with one or more output domains, the term “response regulator” was coined in 1977 by Daniel Koshland to describe the chemotaxis protein CheY that consists solely of the REC domain [7]. Such single-domain RRs comprise ~17% of bacterial RRs and almost half of all RRs in Archaea (Figure 1). These proteins control bacterial motility (e.g., CheY) and participate in signaling phosphorelays (e.g., Spo0F) and in protein-protein interactions governing cell development and division (e.g., DivK, CpdR, see [8] for a recent review).

A key problem in studying RRs using the standard genome analysis techniques is that sequence similarity of RRs does not necessarily indicate functional identity. For example, sequences of *Bacillus subtilis* Spo0F and *Caulobacter crescentus* DivK share 30% identity, which is higher than between chorismate mutases from these two organisms. Still, these proteins have dramatically different functions. Transcriptional regulators OmpR and PhoB from *Escherichia coli* have 37% identical amino acid residues but regulate entirely different processes. As a result, assigning functions to newly sequenced RRs based solely on sequence similarity is problematic, and many such annotations in the current genome databases appear to be wrong. In contrast, family assignments of RRs based on their domain architectures are relatively robust, easy to make, and immediately indicate function. Indeed, it is often impossible to say whether a newly sequenced REC-only domain protein controls chemotaxis, cell development, or any other chemosensory pathway [9]. Still, such a protein can be confidently assigned to the single-domain RR family, meaning that it is definitely not a transcriptional regulator. Likewise, although it is usually hard to predict the target operon(s) for an RR that combines a REC domain with a winged helix domain, such RR can be confidently assigned to the OmpR/PhoB family, meaning that it is definitely a transcriptional regulator that likely shares the key structural features and the regulatory mechanism with the well-studied eponymous members of that family (see [6]). Thus, functional assignment of an RR typically relies on the recognition of its output domain and its functional assignment to the DNA-binding, RNA-binding, ligand-binding, protein-binding, enzyme or transporter protein category.

DNA-binding output domains

A great majority of bacterial RRs contain DNA-binding output domains and serve as transcriptional regulators (this does not appear to be the case in Archaea, see Figure 1b). The most common types of DNA-binding output domains are listed in Table 1. Most of them have known three-dimensional (3D) structures that feature different variants of the helix-

turn-helix (HTH) DNA-binding structural motif [10]. However, crystal structures of full-length RRs are currently available only for the OmpR/PhoB and NarL/FixJ families [11,12].

The last major gap in the list of output domain structures has been filled last year when Ann Stock and colleagues reported the crystal structure of the LytTR domain, a unique DNA-binding domain that lacks the HTH motif and consists mostly of β -strands [•13].

Transcriptional regulators with the LytTR-type output domains are widespread in bacteria and control production of virulence factors in several important bacterial pathogens [14,15]. The complex of the LytTR domain from *Staphylococcus aureus* AgrA with its 15-bp DNA target fragment revealed a previously unknown mode of protein-DNA interaction, which involves side chains of amino acid residues that are located in the loops between the β -strands [•13]. The variability of these residues among the members of the LytR/AgrA family could explain the diversity of the DNA targets of these RRs. Another interesting observation from this structure is the significant bending of the target DNA fragment site upon binding of the LytTR domain [•13]. This DNA bending is likely to increase the activity of the respective promoters and could explain the mechanism of transcriptional activation by the RRs of the LytR/AgrA family. A similar structure was recently submitted to the Protein Data Bank (PDB) by the Midwest Center for Structural Genomics (PDB: 3d6w, Osipiuk *et al.*, LytTR DNA-binding domain of putative methyl-accepting/DNA response regulator from *Bacillus cereus*, unpublished), confirming the key observations of [•13].

For other transcriptional regulators, recent efforts have been targeted towards structural characterization of full-length RRs, analysis of the conformational changes brought by phosphorylation of the REC domain, and analysis of the precise mechanisms of DNA binding and target specificity. For the OmpR/PhoB family, structures of a full-length RR and a DNA-bound complex of the winged-helix domain have been available since 2002 [11,16] and the conformational changes caused by phosphorylation have been studied in detail [6]. Remarkably, even very similar RRs appear to differ in their activation mechanisms [17]. A recent paper studied the interaction of the C-terminal DNA-binding domain of OmpR with its DNA target sequences and concluded that there are principal differences in how OmpR and PhoB are affected by phosphorylation of their REC domains [•18].

The NarL/FixJ family is the second most abundant family of bacterial RRs. Its members have a typical HTH DNA-binding output domain (referred to as GerE in Pfam) that is similar to the one in the transcriptional regulator LuxR. For that reason, RRs of this family are often annotated as members of the LuxR family. Sometimes, the NarL/FixJ family is further subdivided into TetR, IclR and other families. The recent work from Valley Stewart's group revealed the fine-tuned regulation of assimilation of nitrate and nitrite in *E. coli* by closely related paralogous RRs NarL and NarP [19,•20].

For the RRs of the NtrC/DctD family, no full-length structure has been reported so far, although structural models have been built based on the structures of the three individual constituent domains and domain pairs [21,22]. Structural data indicated that inactive (non-phosphorylated) NtrC molecule forms dimers in solution, and its activation requires further oligomerization with the central σ^{54} -interacting ATPase domain forming a ring-like structure. A recent work clarified this issue by showing that the active form of NtrC is a hexamer [•23].

Enzymatic output domains

Although RRs are often assumed to serve as transcriptional regulators, a significant fraction of bacterial RRs do not regulate transcription, at least not in a direct way. In addition to

single-domain RRs, such RRs include those with output domains that have enzymatic activity and themselves participate in signal transduction: methyltransferases, adenylate cyclases, diguanylate cyclases, c-di-GMP-specific phosphodiesterases, histidine kinases, serine/threonine protein kinases and protein phosphatases (Table 2). Some of these RRs have been extensively characterized. These include the widespread CheB family of chemotaxis proteins that combine the REC domain with a C-terminal methyltransferase domain [24,25] and PleD-like RRs that contain two REC domains followed by the diguanylate cyclase (GGDEF, PF01590) domain [26–28]. Two recent papers provided structural characterization of WspR, an RR with the REC-GGDEF domain architecture, whose activation mechanism is substantially different from that of PleD and appears to involve formation of a WspR tetramer [•29,•30]. The GGDEF-containing RRs of *Anaplasma phagocytophilum* (PleD family) and *Borrelia burgdorferi* (WspR-family) have been shown to serve as global regulators of cell metabolism in these important human pathogens [•31,•32].

In many other RRs, output domains have well-characterized structures and functions but the structures of full-length RRs still remain to be solved. These include RRs of PvrR and RpfG families, so named after their first characterized representatives [33,34], whose output domains are c-di-GMP-specific phosphodiesterases of two classes, represented, respectively, by the EAL and HD-GYP domains. Several 3D structures of the EAL domain have been solved, suggesting a catalytic mechanism for the c-di-GMP hydrolysis [•35, •36], whereas for the HD-GYP domain, only the structure of generic HD-type phosphodiesterase is available at this time [28]. Similarly, although the structures of class III adenylate cyclases, Ser/Thr protein kinases and CheC-type protein phosphatases are known, structures of full-length RRs with these output domains are still unavailable. The structure of a full-length RR with a PP2C-type protein phosphatase output domain has been recently released but has not yet been formally described (PDB: 3eq2, Levchenko *et al.*, Structure of hexagonal crystal form of *Pseudomonas aeruginosa* RssB, unpublished). Based on this work, RRs with the PP2C-type output domain can now be referred to as the RssB family (to avoid confusion, it should be noted that the *E. coli* RssB protein, also referred to as Hnr or SprE [37], has a truncated PP2C-type domain that is apparently devoid of phosphatase activity and functions solely in protein-protein interactions [5]).

Protein-binding output domains

Recent studies revealed another interesting RR that functions through protein-protein interactions. This RR, encoded in the genomes of most alpha-proteobacteria but so far not seen outside that lineage, combines a REC domain with an N-terminal domain that is very similar to the σ^E (RpoE, σ^{24}) subunit of RNA polymerase. Based on the DNA-binding properties of σ^E , we and others initially speculated that these RRs were yet another group of DNA-binding transcriptional regulators [5, 38], a suggestion that now appears to be incorrect. The actual story proved to be much more complex. First, this RR, named PhyR, was shown to regulate plant colonization by *Methylobacterium extorquens* [38]. Subsequent studies demonstrated that PhyR is involved in regulation of the general stress response, including resistance to heat shock and desiccation in both *M. extorquens* and *Bradyrhizobium japonicum* [•39,•40]. Most importantly, PhyR did not appear to be involved in DNA binding or interaction with RNA polymerase, as might be expected of its sigma factor domain [••41]. Instead, it appeared to act through a partner-switching mechanism, somewhat similar to that regulating the activity of σ^B in *B. subtilis* [42]. The σ^E domain of PhyR was shown to bind an anti-sigma factor, a newly characterized small protein NepR. This led to the model of PhyR action that includes sequestering of NepR and thus freeing up a genuine sigma factor σ^{EcfG} , which allows σ^{EcfG} to interact with the RNA polymerase thereby stimulating transcription of the stress-related genes. The environmental control of

this signaling mechanism is achieved through phosphorylation of the PhyR REC domain, which could unlock the σ^E domain and allow its interaction with NepR [••41].

An expanding list of output domains

Continued genome sequencing of phylogenetically and ecologically diverse microorganisms results in rapid growth of the number of RR sequences and continuously introduces previously unknown RR domain architectures. In many cases, these architectures include well-characterized domains that just have not been previously associated with two-component signaling. For example, the recently sequenced genomes of the alkane-degrading sulfate-reducing delta-proteobacteria *Desulfatibacillum alkenivorans* and *Desulfococcus oleovorans* encode RRs whose output domains are membrane transporters for nitrate, sulfate, and dicarboxylates. Several other bacteria encode fusions of the REC domain with membrane-bound and/or ATPase components of the ABC-type transporters (see Supplementary Table 1). Thus, the list of functions controlled by RRs still keeps growing and now includes membrane transport of various ions.

Another clear trend among recently discovered RR sequences is the growing number of metabolic enzymes that are found fused to the REC domain in at least some bacteria. Such enzymes include P-loop-type ATPases of the MinD/ParA and PilB families, threonine synthase, nucleoside phosphorylase, sugar transferase, dolichyl-phosphate glucosyltransferase, NAD(P)-dependent glutamate dehydrogenase, potential metal-dependent hydrolase, and other enzymes (MYG, manuscript in preparation, see Supplementary Table 1). As discussed previously [5], such fusions appear to be evolutionarily harmless – but potentially advantageous – events that put these metabolic enzymes under environmental control.

Phylogenetic conservation of response regulator family profiles

Representatives of different bacterial lineages show clear differences in the families of RRs that they encode (Figure 1b). However, closely related species typically have very similar RR family profiles, i.e. they encode similar amounts of certain RRs and do not encode the same RRs (Figure 2). This trend is well preserved at the genus level and can often be followed up to the phylum level, even though the absolute numbers of encoded RRs differ greatly, in accordance with the organism's genome size. For example, Figure 2 shows that genome compaction in the course of adaptation of *Anoxybacillus flavithermus* to its unique silica-saturated ecological niche [43] was accompanied by a massive loss of RR genes. However, the relative numbers of RRs of each kind did not change much, meaning that different RR families were equally affected by this process. In addition to its potential use in taxonomy, this conservation of RR family profiles could have predictive power, eventually allowing us to define the ecological preferences of a given organism based solely on its signaling capacity deduced from the genome sequence (see Box 1).

Box 1

Future directions

The current knowledge of TCS comes from studies of only a few RRs in a handful of model organisms. The growing list of sequenced genomes and metagenomes provides opportunities for expanding this set ([••13,•31,•32,•40,••41] are good examples) and gaining a much better understanding of signaling mechanisms in a wide variety of microorganisms from diverse environments. In the near future, to make better use of the genomic data, it would be necessary to:

- Solve 3D structures of all kinds of full-length RRs and their output domains in complex with their DNA or protein targets and/or ligands and identify the specificity-determining residues
- Create an exhaustive list of all protein-protein interactions that involve prokaryotic RRs
- Identify the environmental signals and cellular targets for the c-di-GMP-based regulation and Ser/Thr protein phosphorylation
- For DNA-binding RRs, compile the lists of target genes (operons), define the target binding sites and delineate the entire regulons for each RR (i.e. convert family assignments into true annotation)
- Correlate the sets of encoded RRs and their respective regulons with the ecological niches of the host organisms

In the long run, accomplishing these goals should allow us to deduce the behavior and metabolism of each organism based solely on its genome sequence and compile a unified picture of microbial life in each microcosm.

Response regulators in Archaea and Eukaryotes

Although TCSs are found in all three domains of life, Bacteria, Archaea and Eukaryotes, the presence and abundance of particular RR classes varies between the lineages. Archaea (at least the ones with sequenced genomes) encode very few RRs with DNA-binding output domains (Figure 1b). Almost half of all archaeal RRs are single-domain RRs (stand-alone REC domains); in some organisms they comprise 90–100% of all RRs [8,44]. Many archaea also encode RRs that combine a REC domain with one or more PAS and/or GAF domains [5]. The ligands, if any, bound by these RRs remain uncharacterized and the pathways that they regulate remain obscure. These RRs, similarly to some single-domain RRs, likely participate in allosteric regulation of their cognate histidine kinases. Other widespread archaeal RRs include the chemotaxis regulator CheB (Figure 1b) and RRs with the HalX (PF08663) output domain, whose function still remains unknown.

Among eukaryotes, TCSs are found primarily in protozoa, fungi, algae and green plants. The few RR genes apparently found in metazoan genomes most likely come from bacterial contamination. In yeast and in other fungi, TCSs are involved in cellular response to osmotic, oxidative, and other environmental stresses [45,46,47]. In plants, TCSs are additionally involved in circadian rhythms, adaptation to salinity, cytokinin signaling [48,49,50] and in redox regulation of chloroplast gene expression [51]. These pathways involve single-domain RRs, as well as RRs with various output domains, such as SANT (Myb_DNA-binding, PF00249). Studies of RRs from eukaryotic microorganisms are only beginning but hold great promise for understanding signal transduction mechanisms operating in higher organisms.

Conclusions

Bacterial response regulators are extremely diverse: they include almost all known types of DNA-binding domains, many ligand-binding and protein-binding domains, membrane transporters and a variety of metabolic and signaling enzymes. There appears to be no restriction on the variety of domains that could be fused to the REC domain and thereby put under environmental control. Obviously, the principal control mechanism is regulation of gene expression at the transcriptional level. However, widespread RRs with methyltransferase, diguanylate cyclase, c-di-GMP-specific phosphodiesterase or Ser/Thr protein phosphatase

output domains provide additional means of control that work at post-transcriptional levels. Such RRs allow TCSs to interfere with the functioning of other signal transduction systems and appear to put TCSs at the top of the bacterial signaling hierarchy.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

I thank Armen Mulikidjanian and Sergei Mekhedov for helpful suggestions and many other colleagues for critical comments. This study was supported by the Intramural Research Program of the National Library of Medicine at the U.S. National Institutes of Health.

References

- 1. Finn RD, Mistry J, Tate J, Coghill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, et al. The Pfam protein families database. *Nucleic Acids Res.* 2010; 38:D211–D222. The new release of the Pfam domain database (<http://pfam.sanger.ac.uk/>) contains an expanded list of RR output domains. It offers the possibility of viewing the taxonomic distribution of all proteins containing any given domain or domain architecture and of downloading protein sequences with the selected domain from any given taxonomic node. [PubMed: 19920124]
2. Ulrich LE, Zhulin IB. MiST: a microbial signal transduction database. *Nucleic Acids Res.* 2007; 35:D386–D390. [PubMed: 17135192]
- 3. Ulrich, LE.; Zhulin, IB. The MiST2 database: a comprehensive genomics resource on microbial signal transduction; *Nucleic Acids Res.* 2010. p. D401-D407. The new release of the popular Microbial Signal Transduction database (MiST, <http://mistdb.com>) comes with a host of new features, including an enhanced taxonomy browser, a new classification of signaling proteins, and a list of adjacent histidine kinase and RR genes
- 4. Barakat M, Ortet P, Jourlin-Castelli C, Ansaldi M, Mejean V, Whitworth DE. P2CS: a two-component system resource for prokaryotic signal transduction research. *BMC Genomics.* 2009; 10:315. A new database provides a detailed classification of TCSs encoded in complete genomes and metagenomes, subdividing histidine kinases into classic, hybrid, CheA-type, or unorthodox ones and classifying the RRs into families according to their output domains. [PubMed: 19604365]
5. Galperin MY. Structural classification of bacterial response regulators: diversity of output domains and domain combinations. *J Bacteriol.* 2006; 188:4169–4182. [PubMed: 16740923]
6. Gao R, Mack TR, Stock AM. Bacterial response regulators: versatile regulatory strategies from common domains. *Trends Biochem Sci.* 2007; 32:225–234. [PubMed: 17433693]
7. Koshland DE Jr. A response regulator model in a simple sensory system. *Science.* 1977; 196:1055–1063. [PubMed: 870969]
8. Jenal U, Galperin MY. Single domain response regulators: molecular switches with emerging roles in cell organization and dynamics. *Curr Opin Microbiol.* 2009; 12:152–160. [PubMed: 19246239]
9. Kirby JR. Chemotaxis-like regulatory systems: unique roles in diverse bacteria. *Annu Rev Microbiol.* 2009; 63:45–59. [PubMed: 19379070]
10. Aravind L, Anantharaman V, Balaji S, Babu MM, Iyer LM. The many faces of the helix-turn-helix domain: Transcription regulation and beyond. *FEMS Microbiol Rev.* 2005; 29:231–262. [PubMed: 15808743]
11. Buckler DR, Zhou Y, Stock AM. Evidence of intradomain and interdomain flexibility in an OmpR/PhoB homolog from *Thermotoga maritima*. *Structure.* 2002; 10:153–164. [PubMed: 11839301]
12. Baikalov I, Schroder I, Kaczor-Grzeskowiak M, Grzeskowiak K, Gunsalus RP, Dickerson RE. Structure of the *Escherichia coli* response regulator NarL. *Biochemistry.* 1996; 35:11053–11056. [PubMed: 8780507]
- 13. Sidote DJ, Barbieri CM, Wu T, Stock AM. Structure of the *Staphylococcus aureus* AgrA LytTR domain bound to DNA reveals a beta fold with an unusual mode of binding. *Structure.* 2008;

16:727–735. The first structural description of the widespread LytTR DNA-binding domain, which is found in RRs that control alginate production in *Pseudomonas aeruginosa*, bacteriocin production in various lactobacilli, and expression of virulence factors in streptococci and staphylococci. The structure of the LytTR domain in complex with its target DNA fragment reveals an unusual mode of protein-DNA interaction and provides the first clues to the specificity determinants in the RRs of the LytR/AgrA family. [PubMed: 18462677]

14. Nikolskaya AN, Galperin MY. A novel type of conserved DNA-binding domain in the transcriptional regulators of the AlgR/AgrA/LytR family. *Nucleic Acids Res.* 2002; 30:2453–2459. [PubMed: 12034833]
15. Galperin MY. Telling bacteria: do not LytTR. *Structure.* 2008; 16:657–659. [PubMed: 18462668]
16. Blanco AG, Sola M, Gomis-Ruth FX, Coll M. Tandem DNA recognition by PhoB, a two-component signal transduction transcriptional activator. *Structure.* 2002; 10:701–713. [PubMed: 12015152]
17. Gao R, Stock AM. Molecular strategies for phosphorylation mediated regulation of response regulator activity. *Curr Opin Microbiol.* 2010; 13:160–167. [PubMed: 20080056]
- 18. Rhee JE, Sheng W, Morgan LK, Nolet R, Liao X, Kenney LJ. Amino acids important for DNA recognition by the response regulator OmpR. *J Biol Chem.* 2008; 283:8664–8677. A NMR solution structure of the DNA-binding domain of OmpR is used to analyze the residues involved in DNA binding and to compare the binding modes of OmpR and PhoB. There appear to be surprising differences between these two RRs with respect to the roles of phosphorylation and dimerization in DNA binding. While in PhoB phosphorylation precedes dimerization and DNA binding, the authors suggest that OmpR binds to DNA as a monomer and only after that forms a DNA-bound dimer. [PubMed: 18195018]
19. Noriega CE, Schmidt R, Gray MJ, Chen LL, Stewart V. Autophosphorylation and dephosphorylation by soluble forms of the nitrate-responsive sensors NarX and NarQ from *Escherichia coli* K-12. *J Bacteriol.* 2008; 190:3869–3876. [PubMed: 18375557]
- 20. Noriega CE, Lin HY, Chen LL, Williams SB, Stewart V. Asymmetric cross regulation between the nitrate-responsive NarX-NarL and NarQ-NarP two-component regulatory systems from *Escherichia coli* K-12. *Mol Microbiol.* 2010; 75:394–412. A careful analysis of the interplay of the two closely related paralogous TCSs that control utilization of nitrate and nitrite by *E. coli*. The study shows that even very similar RRs may have distinct functions and complement each other in fine-tuning the cellular response to their common environmental signal. [PubMed: 19968795]
21. De Carlo S, Chen B, Hoover TR, Kondrashkina E, Nogales E, Nixon BT. The structural basis for regulated assembly and function of the transcriptional activator NtrC. *Genes Dev.* 2006; 20:1485–1495. [PubMed: 16751184]
22. Batchelor JD, Doucleff M, Lee CJ, Matsubara K, De Carlo S, Heideker J, Lamers MH, Pelton JG, Wemmer DE. Structure and regulatory mechanism of *Aquifex aeolicus* NtrC4: variability and evolution in bacterial transcriptional regulation. *J Mol Biol.* 2008; 384:1058–1075. [PubMed: 18955063]
- 23. Batchelor JD, Sterling HJ, Hong E, Williams ER, Wemmer DE. Receiver domains control the active-state stoichiometry of *Aquifex aeolicus* σ^{54} activator NtrC4, as revealed by electrospray ionization mass spectrometry. *J Mol Biol.* 2009; 393:634–643. This study shows that the central σ^{54} -interacting domain of the NtrC family response regulator, like other members of the AAA+ ATPase family, is active in a hexameric form. [PubMed: 19699748]
24. West AH, Martinez-Hackert E, Stock AM. Crystal structure of the catalytic domain of the chemotaxis receptor methylesterase, CheB. *J Mol Biol.* 1995; 250:276–290. [PubMed: 7608974]
25. Djordjevic S, Goudreau PN, Xu Q, Stock AM, West AH. Structural basis for methylesterase CheB regulation by a phosphorylation-activated domain. *Proc Natl Acad Sci USA.* 1998; 95:1381–1386. [PubMed: 9465023]
26. Chan C, Paul R, Samoray D, Amiot NC, Giese B, Jenal U, Schirmer T. Structural basis of activity and allosteric control of diguanylate cyclase. *Proc Natl Acad Sci USA.* 2004; 101:17084–17089. [PubMed: 15569936]

27. Wassmann P, Chan C, Paul R, Beck A, Heerklotz H, Jenal U, Schirmer T. Structure of BeF_3^- -modified response regulator PleD: implications for diguanylate cyclase activation, catalysis, and feedback inhibition. *Structure*. 2007; 15:915–927. [PubMed: 17697997]
28. Schirmer T, Jenal U. Structural and mechanistic determinants of c-di-GMP signalling. *Nat Rev Microbiol*. 2009; 7:724–735. [PubMed: 19756011]
- 29. De N, Pirruccello M, Krasteva PV, Bae N, Raghavan RV, Sondermann H. Phosphorylation-independent regulation of the diguanylate cyclase WspR. *PLoS Biol*. 2008; 6:e67. [PubMed: 18366254]
- 30. De N, Navarro MV, Raghavan RV, Sondermann H. Determinants for the activation and autoinhibition of the diguanylate cyclase response regulator WspR. *J Mol Biol*. 2009; 393:619–633. The above two papers present a detailed structural analysis of the founding member of the WspR family of RRs with the REC-GGDEF domain architecture. The authors argue that the active form of WspR is a tetramer whose formation is favored by phosphorylation of its REC domain, whereas the WspR dimer can exist in several different inactive states. [PubMed: 19695263]
- 31. Lai TH, Kumagai Y, Hyodo M, Hayakawa Y, Rikihisa Y. *Anaplasma phagocytophilum* PleC histidine kinase and PleD diguanylate cyclase two-component system and role of cyclic di-GMP in host-cell infection. *J Bacteriol*. 2008; 191:693–700. [PubMed: 18978058]
- 32. Rogers EA, Terekhova D, Zhang HM, Hovis KM, Schwartz I, Marconi RT. Rrp1, a cyclic-di-GMP-producing response regulator, is an important regulator of *Borrelia burgdorferi* core cellular functions. *Mol Microbiol*. 2009; 71:1551–1573. These two papers demonstrate the global role of GGDEF-containing RRs in two important human pathogens with dramatically reduced genome sizes that encode, respectively, only two (*A. marginale*) and four (*B. burgdorferi*) other RRs. [PubMed: 19210621]
33. Drenkard E, Ausubel FM. *Pseudomonas* biofilm formation and antibiotic resistance are linked to phenotypic variation. *Nature*. 2002; 416:740–743. [PubMed: 11961556]
34. Slater H, Alvarez-Morales A, Barber CE, Daniels MJ, Dow JM. A two-component system involving an HD-GYP domain protein links cell-cell signalling to pathogenicity gene expression in *Xanthomonas campestris*. *Mol Microbiol*. 2000; 38:986–1003. [PubMed: 11123673]
35. Minasov G, Padavattan S, Shuvalova L, Brunzelle JS, Miller DJ, Basle A, Massa C, Collart FR, Schirmer T, Anderson WF. Crystal structures of YkuI and its complex with second messenger cyclic di-GMP suggest catalytic mechanism of phosphodiester bond cleavage by EAL domains. *J Biol Chem*. 2009; 284:13174–13184. [PubMed: 19244251]
- 36. Barends TR, Hartmann E, Griesse JJ, Beitlich T, Kirienko NV, Ryjenkov DA, Reinstein J, Shoeman RL, Gomelsky M, Schlichting I. Structure and mechanism of a bacterial light-regulated cyclic nucleotide phosphodiesterase. *Nature*. 2009; 459:1015–1018. These two papers describe the structures of two different EAL-type c-di-GMP-specific phosphodiesterases in complex with the c-di-GMP substrate and propose detailed catalytic mechanisms of the enzyme. [PubMed: 19536266]
37. Hengge R. The two-component network and the general stress sigma factor RpoS (σ^S) in *Escherichia coli*. *Adv Exp Med Biol*. 2008; 631:40–53. [PubMed: 18792681]
38. Gourion B, Rossignol M, Vorholt JA. A proteomic study of *Methylobacterium extorquens* reveals a response regulator essential for epiphytic growth. *Proc Natl Acad Sci USA*. 2006; 103:13186–13191. [PubMed: 16926146]
- 39. Gourion B, Francez-Charlot A, Vorholt JA. PhyR is involved in the general stress response of *Methylobacterium extorquens* AM1. *J Bacteriol*. 2008; 190:1027–1035. [PubMed: 18024517]
- 40. Gourion B, Sulser S, Frunzke J, Francez-Charlot A, Stiefel P, Pessi G, Vorholt JA, Fischer HM. The PhyR- σ^{EcfG} signalling cascade is involved in stress response and symbiotic efficiency in *Bradyrhizobium japonicum*. *Mol Microbiol*. 2009; 73:291–305. [PubMed: 19555458]
- 41. Francez-Charlot A, Frunzke J, Reichen C, Ebnetter JZ, Gourion B, Vorholt JA. Sigma factor mimicry involved in regulation of general stress response. *Proc Natl Acad Sci USA*. 2009; 106:3467–3472. The above four papers uncover a remarkable story of a new type of global response regulator that contains a σ^{E} (RpoE)-like output domain and functions as an anti-anti-sigma factor in a partner-switching mechanism, promoting expression of stress-related genes by sequestering a sigma factor antagonist. [PubMed: 19218445]

42. Yang X, Kang CM, Brody MS, Price CW. Opposing pairs of serine protein kinases and phosphatases transmit signals of environmental stress to activate a bacterial transcription factor. *Genes Dev.* 1996; 10:2265–2275. [PubMed: 8824586]
43. Saw JH, Mountain BW, Feng L, Omelchenko MV, Hou S, Saito JA, Stott MB, Li D, Zhao G, Wu J, et al. Encapsulated in silica: genome, proteome and physiology of the thermophilic bacterium *Anoxybacillus flavithermus* WK1. *Genome Biol.* 2008; 9:R161. [PubMed: 19014707]
44. Ashby MK. Distribution, structure and diversity of “bacterial” genes encoding two-component proteins in the *Euryarchaeota*. *Archaea.* 2006; 2:11–30. [PubMed: 16877318]
45. Jones CA, Greer-Phillips SE, Borkovich KA. The response regulator RRG-1 functions upstream of a mitogen-activated protein kinase pathway impacting asexual development, female fertility, osmotic stress, and fungicide resistance in *Neurospora crassa*. *Mol Biol Cell.* 2007; 18:2123–2136. [PubMed: 17392518]
46. Azuma N, Kanamaru K, Matsushika A, Yamashino T, Mizuno T, Kato M, Kobayashi T. In vitro analysis of His-Asp phosphorelays in *Aspergillus nidulans*: the first direct biochemical evidence for the existence of His-Asp phosphotransfer systems in filamentous fungi. *Biosci Biotechnol Biochem.* 2007; 71:2493–2502. [PubMed: 17928704]
- 47. Kaserer AO, Andi B, Cook PF, West AH. Effects of osmolytes on the SLN1-YPD1-SSK1 phosphorelay system from *Saccharomyces cerevisiae*. *Biochemistry.* 2009; 48:8044–8050. A recent analysis of the roles of two yeast RRs, SSK1 and SKN7, in cellular response to hyperosmotic stress. The wide conservation of these RRs suggests that similar mechanisms of stress response might be operative in other fungi. [PubMed: 19618914]
48. Mizuno T. Two-component phosphorelay signal transduction systems in plants: from hormone responses to circadian rhythms. *Biosci Biotechnol Biochem.* 2005; 69:2263–2276. [PubMed: 16377883]
49. Ishida K, Yamashino T, Yokoyama A, Mizuno T. Three type-B response regulators, ARR1, ARR10 and ARR12, play essential but redundant roles in cytokinin signal transduction throughout the life cycle of *Arabidopsis thaliana*. *Plant Cell Physiol.* 2008; 49:47–57. [PubMed: 18037673]
- 50. Ishida K, Niwa Y, Yamashino T, Mizuno T. A genome-wide compilation of the two-component systems in *Lotus japonicus*. *DNA Res.* 2009; 16:237–247. A census of all TCSs encoded in the *Lotus japonicus* genome in comparison with the TCS sets of *Arabidopsis thaliana* and rice. A useful review of the current state of knowledge of the roles of TCSs in plants in general. [PubMed: 19675111]
- 51. Puthiyaveetil S, Kavanagh TA, Cain P, Sullivan JA, Newell CA, Gray JC, Robinson C, van der Giezen M, Rogers MB, Allen JF. The ancestral symbiont sensor kinase CSK links photosynthesis with gene expression in chloroplasts. *Proc Natl Acad Sci USA.* 2008; 105:10061–10066. Sequence analysis reveals a histidine kinase that is conserved in cyanobacteria and in chloroplasts of modern plants. In *Arabidopsis thaliana*, this sensor kinase is shown to be encoded in the nucleus and be imported into the chloroplast, where it regulates expression of the photosystem I reaction center protein A (PsaA), apparently in response to the redox status of the plastoquinone. In algae, it appears to interact with RRs of the OmpR family, in higher plants its cognate RRs remain unknown. [PubMed: 18632566]
52. Velankar S, Best C, Beuth B, Boutselakis CH, Cogley N, Sousa Da Silva AW, Dimitropoulos D, Golovin A, Hirshberg M, John M, et al. PDBe: Protein Data Bank in Europe. *Nucleic Acids Res.* 2010; 38:D308–D317. [PubMed: 19858099]

Figure 1. Distribution of common output domains of bacterial response regulators

A. Aggregate data for all bacterial species. The sector for Fis domain includes response regulators of the NtrC and ActR families (no structure of a Fis-DNA complex is currently available). Unlabelled output domains (clockwise from AraC) are as follows: Spo0A (dark blue), ANTAR (cyan), CheW (pink), CheC (white), GGDEF+EAL (orange), HD-GYP (yellow). The domain structures are as in Tables 1 and 2 and are taken from Protein Data Bank in Europe [52].

B. Response regulator family profiles for selected bacterial phyla, Archaea, and α , β , γ and δ subdivisions of Proteobacteria. The columns indicate the relative fractions of the RRs of single-domain (white), OmpR (maroon), NarL (bisque), NtrC (blue), combined LytR, ActR, YesN/AraC, and Spo0A (purple), CheB (magenta), combined c-di-GMP signaling (WspR, PleD, PvrR, FimX, and RpfG, orange), and enzymatic (HisK and PP2C output domains, green) families within each taxonomic group. The grey bars indicate combined fractions of all other RR families.

Figure 2. Response regulator family profiles in representatives of the genus *Shewanella* (A) and the family *Bacillaceae* (B)

The columns indicate the number of response regulators of each family encoded in each genome. The family names as in Tables 1 and 2; the REC column shows the number of single-domain RRs (no output domains), the WspR column shows combined data for WspR and PleD families (the GGDEF output domain), the FimX label indicates response regulators that contain fused GGDEF and EAL output domains. The organisms (from front to back) are as follows (only every other row is labeled):

A: *Shewanella denitrificans* OS217, *S. frigidimarina* NCIMB 400, *S. halifaxensis* HAW-EB4, *S. pealeana* ATCC 700345, *S. baltica* OS155, *S. loihica* PV-4, *S. piezotolerans* WP3, *S. oneidensis* MR-1, *S. amazonensis* SB2B, *S. woodyi* ATCC 51908, and *S. sediminis* HAW-EB3.

B: *Anoxybacillus flavithermus* WK1, *Geobacillus kaustophilus* HTA426 G, *thermodenitrificans* NG80-2, *Bacillus pumilus* SAFR-032, *B. subtilis* subsp. *subtilis* str. 168,

B. amyloliquefaciens FZB42, *B. licheniformis* ATCC 14580, *B. clausii* KSM-K16, *B. cereus* ATCC 14579, *B. anthracis* str. Ames, *B. halodurans* C-125, *B. thuringiensis* serovar *konkukian* str. 97–27, and *B. weihenstephanensis* KBAB4.

Table 1

Widespread DNA-binding output domains in bacterial response regulators

Output domain	RR family name	Pfam entry ^a	3D structure ^b		% Total ^c	RR phylogenetic distribution
			Output domain	Whole RR		
Two-domain RRs						
Winged helix (Trans_reg_C)	OmpR/PhoB	PF00486	1gxp	1kgs	30.1%	All bacterial phyla
Helix-turn-helix (LuxR_C_like or GerE)	NarL/FixJ	PF00196	1je8	1a04	16.9%	Most bacterial phyla except for Chlamydiae, Thermotogae
LytTR	LytR/AgrA	PF04397	3bs1	n/a	3.0%	Not in Chlamydiae, Chlorobi, Chloroflexi, Cyanobacteria
HTH_AraC	YesN/AraC	PF00165	1bl0	n/a	1.6%	Mostly in Firmicutes
Fis (HTH_8)	ActR/PtrA	PF02954	3e7l	n/a	1.0%	Mostly in Proteobacteria
HTH_ArsR	DctR/CitB	PF08279	2pn6	n/a	0.4%	Various bacteria, some archaea
Spo0A_C	Spo0A	PF08769	1fc3	n/a	0.3%	Firmicutes
YcbB	YcbB (GlnL)	PF08664	n/a	n/a	0.1%	Firmicutes
MerR	n/a ^d	PF00376	1r8e	n/a	<0.1%	Various bacteria
Three-domain RRs ^e						
AAA-Fis ^e	NtrC/DctD	PF00158 +PF02954 ^e	1ojl	1ny5 ^f	9.1%	Most bacterial phyla, except for Actinobacteria, Cyanobacteria
wHTH-BTAD ^e	SARP	PF00486 +PF03704 ^e	2ff4	n/a	<0.1%	Actinobacteria, Firmicutes
cNMPbinding-CRP ^e	n/a	PF00027 +PF00325 ^e	1hw5	n/a	<0.1%	Bacteroidetes

^aThe Pfam database [51] entry, can be retrieved using the <http://pfam.sanger.ac.uk/family?PF0xxxxx> format

^bThe Protein Data Bank [51] entry, where available, can be retrieved using the <http://www.rcsb.org/pdb/explore/explore.do?structureId=1gxp> format; n/a – not available

^cThe fraction of the given domain architecture among all RRs from a non-redundant set of completely sequenced bacterial genomes, as listed at http://www.ncbi.nlm.nih.gov/Complete_Genomes/RRcensus.html

^dRRs of this family have the REC domain at their C-termini

^eThese RR families contain two distinct output domains, represented by distinct Pfam entries

^fThis structure lacks the C-terminal Fis domain

Table 2

Response regulators with non-DNA-binding output domains^a

Output domain	Pfam entry	Function	RR family	3D structure		% Total	RR phylogenetic distribution
				Output domain	Whole RR		
RNA-binding							
ANTAR	PF03861	Transcriptional antiterminator	AmiR	1s8n	1qo0	1.1%	Actinobacteria, Chloroflexi, Firmicutes, Proteobacteria
CsrA ^b	PF02599	Unknown	n/a	1vpz	n/a	<0.1%	Planctomycetes
Chemotaxis							
CheB	PF01339	Methyltransferase	CheB	1chd	1a2o	2.4%	Many bacteria, archaea
CheW ^b	PF01584	Scaffolding protein	CheV ^b	1k0s	n/a	1.3%	Firmicutes, Proteobacteria
CheC	PF04509	Asp-P phosphatase	n/a	1xkr	n/a	<0.1%	Firmicutes, Proteobacteria
CheR	PF01739	Methyltransferase	n/a	1af7	n/a	<0.1%	Delta-proteobacteria
Two-component phosphorelay							
HisK (HisKA +HATPase)	PF00512 +PF02518	Phosphoacceptor +phosphotransferase	n/a	2c2a	3dgc ^c	1.8%	Most bacterial phyla
Hpt	PF01627	Phosphocarrier	n/a	1a0b	1bdf ^c	<0.1%	Mostly Proteobacteria
c-di-GMP signaling							
GGDEF	PF01590	Diguanylate cyclase	WspR PleD	3ign	3bre 1w25	2.5% ^d	Most bacterial phyla
EAL	PF00563	C-di-GMP-specific phosphodiesterase	PvrR	2r60	n/a	0.4% ^d	Cyanobacteria, Proteobacteria, Spirochaetes
HD-GYP	PF01966 ^e	C-di-GMP-specific phosphodiesterase	RpfG	1yoy ^e	n/a	1.8%	Most bacterial phyla
Plz	PF07238	C-di-GMP binding	n/a	1ywu	n/a	<0.1%	Delta-proteobacteria
cAMP signaling							
CYCc	PF00211	Adenylyate cyclase	n/a	1cul	n/a	0.2%	Actinobacteria, Chloroflexi, Proteobacteria, Spirochaetes
CAP_ED	PF00027	cAMP binding	n/a	1hw5	n/a	<0.1%	Bdellovibrio, Cytophaga
Protein Ser/Thr phosphorylation							
PP2Cc (SpoIIE)	PF07228	Ser/Thr protein phosphatase	RssB	1a6q	3es2	1.1%	Most bacterial phyla
RpoE ^b	PF04542 +PF04545	Sigma factor (anti-anti-sigma)	PhyR ^e	1or7	n/a	0.3%	Alpha-proteobacteria
S_TKc (RsbW or Pkinase)	PF00069	Anti-sigma (Ser/Thr protein kinase)	n/a	1jwh	n/a	<0.1%	Delta-proteobacteria
STAS (RsbY)	PF01740	Anti-anti-sigma	n/a	3f43	n/a	<0.1%	Delta-proteobacteria

^aExplanations of column headings are as in Table 1; n/a – 3D structure is not available or family name not assigned

^bRRs of this family have the REC domain at their C-termini

^cStructure of a co-crystallized complex, rather than of a single RR polypeptide chain

^dThese numbers do not include RRs of the FimX family that contain both GGDEF and EAL output domains and comprise ~1.1% of the total set.

^eThis Pfam domain and the respective PDB entry represent only part of the HD-GYP domain structure