GUEST COMMENTARY

Census of Prokaryotic Senses

Robert B. Bourret*

Department of Microbiology and Immunology, University of North Carolina, Chapel Hill, North Carolina 27599-7290

The overwhelming majority of life on our planet is microbial, both in terms of phylogenetic diversity (15, 19) and sheer numbers of organisms (25). Virtually every conceivable environmental niche harbors microorganisms capable of growing there. This evolutionary success likely depends in part on signal transduction, the ability to sense changing environmental conditions and then implement appropriate responses. Bacterial, archaeal, and eukaryotic microorganisms utilize two-component regulatory systems for the purpose of signal transduction. In the prototypical case (Fig. 1A), the first component is a sensor kinase, in which detection of environmental stimuli by an input domain is represented by autophosphorylation of a conserved transmitter domain. Phosphoryl groups are then transferred from the sensor kinase to the second component, a response regulator, in which the phosphorylation status of the conserved receiver domain regulates activity of an output domain. In this issue of the *Journal of Bacteriology*, Michael Galperin describes a census of output domains in response regulators encoded by 200 different prokaryotic species (6). The results give us a much more detailed and comprehensive look at the diversity of responses implemented by two-component regulatory systems than has previously been available.

Some history. The existence of two-component regulatory systems was recognized because researchers independently studying a variety of bacterial regulatory networks sequenced the genes they were investigating and deposited the results in appropriate databanks. When a critical mass of relevant sequences was achieved in the mid-1980s, scientists realized that numerous apparently disparate processes were actually different manifestations of the same phenomenon, namely, two-component regulatory systems (12, 18). During the pre-genome sequencing era, our view of the diversity of two-component regulatory systems was strongly biased by the available sample of what people happened to be studying. In 1993, a comprehensive analysis of response regulator sequences contained 79 entries, which could be subdivided into five groups on the basis of their output domains (24). Most (68%) were transcriptional regulators, with 29% of the total in the OmpR class, 23% in the FixJ/NarL class, and 16% in the NtrC class. Response regulators consisting of a receiver domain alone, with no accompanying output domain, comprised a fourth distinct group, with 9% of the total. Finally, the remaining 23% were lumped together in a miscellaneous category that defied any further rational classification. The conclusion, then, was that most

response regulators are transcriptional regulators, of which there are three major classes.

A postgenomic perspective. Galperin's analysis (6) is based on more than 4,600 response regulator sequences and is remarkable for both its similarities and differences to the results of smaller surveys. Again, most (66%) response regulator output domains appeared to bind DNA and regulate transcription, and the same three classes predominated—OmpR-like (33%), FixJ/NarL-like (19%), and NtrC-like (9%)—although at somewhat different percentages than indicated by smaller samples. Also as before, a significant fraction (14%) of response regulators contained only a receiver domain.

Due to the large sample size, the primary impact of the present study was to reveal the rich diversity among the minority classes. In addition to the three main types of transcriptional regulators, there were response regulators with LytR-like (3%), Fis-like (1%), and AraC-like (1%) DNA-binding output domains. The old "miscellaneous" category can now be resolved into multiple classes: ~1% of response regulators contained RNA-binding output domains, ~2% contained protein-binding output domains, and ~12% had output domains with enzymatic activity. The enzymatic output domains fell into six main groups: CheB-like methylesterase (2%), GGDEF domain diguanylate cyclase (3%), EAL domain c-di-GMP [bis-(3′→5′)-cyclic diguanosine] phosphodiesterase (2%), HD-GYP domain-predicted c-di-GMP phosphodiesterase (2%), PP2C protein phosphatase (1%), and histidine kinase (2%). The "histidine kinase" category was comprised of a unique set of proteins that contain only receiver and transmitter domains, with the transmitter domain in the position normally occupied by the output domain (Fig. 1B). A plausible interpretation of this domain architecture, yet to be tested experimentally, is that phosphorylation of the receiver domain regulates the activity of the transmitter domain, which functions as the output domain. (Note that hybrid kinases, which contain input, transmitter, and receiver domains [Fig. 1C], were excluded from the census because they lack output domains.) Finally, 6% of response regulators remained in a miscellaneous category, because their output domains failed to reach the categorization threshold of belonging to a group with at least 40 representatives in two different phyla. However, a wide range of output domain types (DNA-binding, protein-binding, enzymatic, etc.) were observed in the miscellaneous group and presumably provide a glimpse of what is likely to be found as additional genomic sequences accumulate.

Modular modules. Modules that carry out distinct functions, such as signal transduction, are believed to represent an important organizational unit within living cells (9). In turn, the

* Mailing address: Department of Microbiology and Immunology, University of North Carolina, Chapel Hill, NC 27599-7290. Phone: (919) 966-2679. Fax: (919) 962-8103. E-mail: bourret@med.unc.edu.
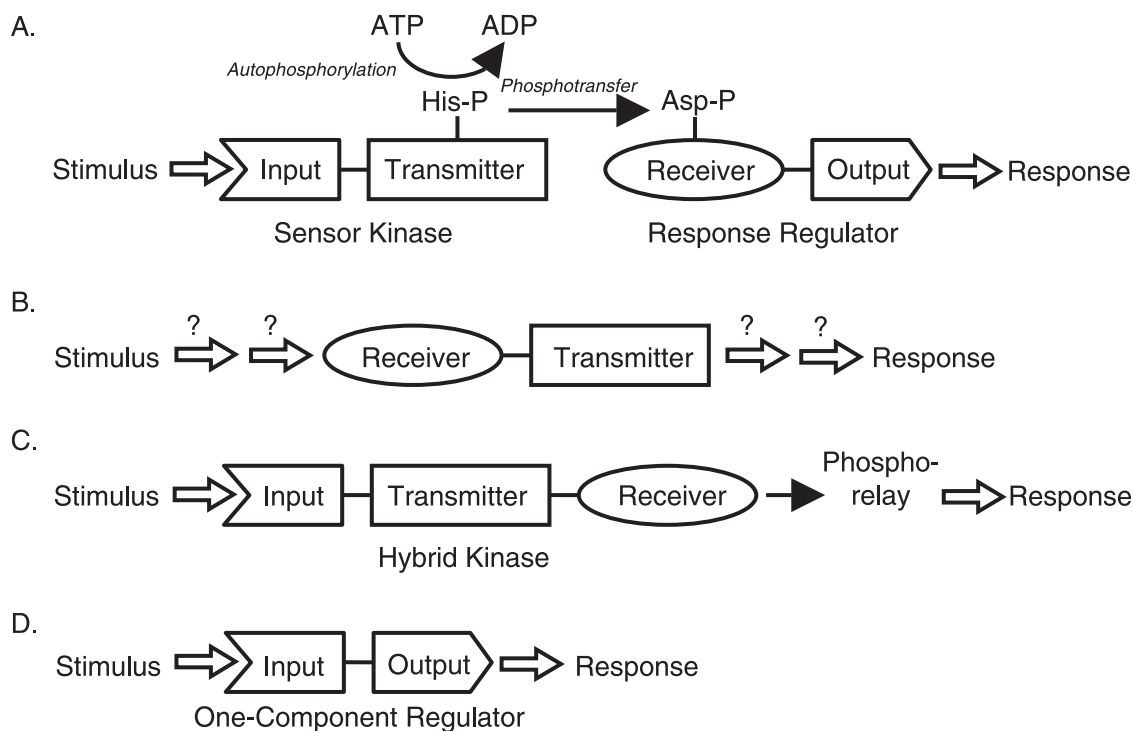
FIG. 1. Domain architecture of signaling proteins. (A) Prototypical example of a two-component regulatory system. Stimulus detection by the input domain regulates autophosphorylation of the transmitter domain. Phosphotransfer to the receiver domain regulates activity of the output domain, which implements the adaptive response. (B) Response regulators that appear to have histidine kinase enzymatic output domains. In addition to the receiver-transmitter architecture depicted, examples of receiver-transmitter-receiver and receiver-receiver-transmitter were also observed (6). (C) A common permutation of two-component regulatory systems is a four-step phosphorelay, which is often initiated by a hybrid kinase, in which a receiver domain is fused to a sensor kinase. Transmitter autophosphorylation and phosphotransfer to the receiver occur within the hybrid kinase. The phosphoryl group is subsequently transferred to a histidine phosphotransferase domain and, finally, to the receiver domain of a standard response regulator (not shown). (D) One-component systems contain direct fusions of input and output domains (21).

proteins that comprise functional modules are themselves modular in design. Structural domains that perform discrete functions have been shuffled during evolution to create proteins containing different combinations of domains (14, 16). Although a staggering number of domain arrangements must have been created and discarded, the ability of multiple domains to productively interact within a protein in a manner that provides a benefit during natural selection is astonishing. Again, the general perspective of Galperin's census is familiar. Previous surveys of two-component regulatory systems revealed much of the wide scope of input and output domains found in sensor kinases and response regulators, respectively (2, 7, 21, 26). The diversity of domains and combinations is even higher in one-component systems (Fig. 1D), in which output and input domains are directly connected in a single protein, without intervening transmitter and receiver domains (21).

However, the new census, which surveys the most genomes to date and focuses exclusively on response regulators, provides some enlightening details. Prokaryotic genomes do not appear to have much excess coding capacity, which suggests that if a gene is present it is likely functional. This line of reasoning implies that phosphorylation of a receiver domain can be used to regulate the activity of 14 major types of output domains (each comprising 1% or more of the sample), plus at least twice as many minor types of output domains (which do

not meet the 1% threshold). Thus, receiver domains have an amazing capacity to adapt their regulatory role to a multitude of output domains.

World Wide Web resources. Happily, for those of us who are bioinformatics challenged, the professionals have done the sequence crunching. Even better, the results are posted on the World Wide Web for all to see. Galperin has compiled all 4,600 response regulators included in his census into a database (5) listing them by species and output domain type, with links to each sequence as well as corresponding entries in various domain analysis databases. This is an incredibly valuable and easy to use resource for anyone who wishes to peruse the response regulator content of a particular species or examine the distribution of output domains across species. It is also a potential gold mine of raw material for future analyses, such as phylogenetic trees to investigate coevolution of receiver and output domains. Such a study could provide insight into the nuances of how apparently similar receiver domains can regulate the activity of so many obviously different types of output domains. Another helpful summary comes from Zhulin and colleagues, who surveyed 145 prokaryotic genomes for one- and two-component regulatory systems (21). Their website (22) contains lists of input and output domains, their prevalence in *Bacteria* and *Archaea*, and the numbers of one- and two-component systems in each species.

The domains of life. Phylogenetic studies suggest that two-component regulatory systems originated in *Bacteria* and later radiated into *Archaea* and *Eukarya* (10). The present census (6) included 177 bacterial genomes and 23 archaeal genomes; eukaryotes were excluded. Thus, the census results summarized so far primarily reflect the situation in *Bacteria*. Examination of Galperin's website (5) quickly reveals a variety of interesting tidbits. Response regulator genes were found in 162 (92%) of the bacterial genomes examined. The presence of response regulators in *Bacteria* correlated with genome size. All bacterial genomes of >1,100 genes contained response regulators, whereas only 50% of the bacterial genomes with <1,100 genes contained response regulators.

Only 97 response regulator genes (~2% of the sample) were found in *Archaea*, and these were confined to only 11 (48%) of the archaeal genomes examined (6). There was no obvious relationship between archaeal genome size and the presence or absence of response regulators. The distribution of response regulator output domains also was markedly different in *Archaea* and *Bacteria*. None of the archaeal output domains belonged to any of the six classes of DNA-binding domains that make up two-thirds of bacterial response regulator output domains. This is particularly surprising because transcriptional regulators in *Archaea* are believed to be largely bacterial in character (8). The RNA-binding and protein-binding categories observed in *Bacteria* were also absent from the archaeal census. Most (57%) archaeal response regulators contained only a receiver domain, without an output domain. Two of the six enzymatic classes of bacterial output domains were observed in *Archaea*: 12% were CheB-like methylesterase, and 2% were histidine kinases. The remaining 29% of output domains in archaeal response regulators were relegated to the miscellaneous category.

Among eukaryotes, response regulators have been found in various microorganisms (amoeba, fungi, and yeasts) and plants. I am not aware of a comprehensive census of eukaryotic response regulators, although several publications survey the field (3, 13, 17). Indeed, an insufficient number of genomes have been sequenced to establish the phylogenetic distribution of two-component regulatory systems among eukaryotes. Some eukaryotic response regulators contain transcriptional activation output domains, even though transcription is very different in eukaryotic cells than in prokaryotes. Many eukaryotic response regulators use protein interaction output domains to feed into other signal transduction pathways such as mitogen-activated protein kinase cascades. There are also examples of solo receiver domains.

A lingering bias. Although a vast improvement in comparison to the pregenomic era, it is important to recognize that currently available genome sequences nevertheless still constitute a biased sample. Obviously, *Archaea* and *Eukarya* are underrepresented in comparison to *Bacteria*. More subtly, a substantial fraction of the initially sequenced microbial genomes belong to human pathogens, which occupy a narrow environmental niche that dictates chemo-organoheterotrophic metabolism at 37°C. Known pathogens represent an infinitesimal fraction of microbial life. More recent genome sequencing efforts have substantially broadened the phylogenetic representation of sequenced genomes. However, genome sequencing has traditionally required laboratory growth of the subject organism in pure culture for the purpose of DNA isolation, which is a problem because only a small fraction of the microorganisms present in a given environmental sample can be successfully cultured with established techniques (1). Fortunately, there has been steady improvement in culture techniques (11), and recent innovations raise the possibility of obtaining complete genome sequences from mixtures of organisms taken directly from the environment (20, 23).

More is better. Microbial genome sequences are being completed at an accelerating rate (4). Future response regulator censuses based on these data will likely cause some reapportionment of the population among existing categories, as well as discover previously unrecognized output domains.

REFERENCES

1. **Amann, R. L., W. Ludwig, and K.-H. Schleifer.** 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol. Rev. **59:**143–169.
2. **Ashby, M. K.** 2004. Survey of the number of two-component response regulator genes in the complete and annotated genome sequences of prokaryotes. FEMS Microbiol. Lett. **231:**277–281.
3. **Catlett, N. L., O. C. Yoder, and B. G. Turgeon.** 2003. Whole-genome analysis of two-component signal transduction genes in fungal pathogens. Eukaryot. Cell **2:**1151–1161.
4. **Fraser-Liggett, C. M.** 2005. Insights on biology and evolution from microbial genome sequencing. Genome Res. **15:**1603–1610.
5. **Galperin, M. Y.** February 2006, posting date. Response regulators encoded in bacterial and archaeal genomes. [Online.] http://www.ncbi.nlm.nih.gov /Complete_Genomes/RRcensus.html.
6. **Galperin, M. Y.** 2006. Structural classification of bacterial response regulators: diversity of output domains and domain combinations. J. Bacteriol. **188:**4169–4182.
7. **Galperin, M. Y., A. N. Nikolskaya, and E. V. Koonin.** 2001. Novel domains of the prokaryotic two-component signal transduction systems. FEMS Microbiol. Lett. **203:**11–21.
8. **Geiduschek, E. P., and M. Ouhammouch.** 2005. Archaeal transcription and its regulators. Mol. Microbiol. **56:**1397–1407.
9. **Hartwell, L. H., J. J. Hopfield, S. Leibler, and A. W. Murray.** 1999. From molecular to modular cell biology. Nature **402:**C47–C52.
10. **Koretke, K. K., A. N. Lupas, P. V. Warren, M. Rosenberg, and J. R. Brown.** 2000. Evolution of two-component signal transduction. Mol. Biol. Evol. **17:**1956–1970.
11. **Leadbetter, J. R.** 2003. Cultivation of recalcitrant microbes: cells are alive, well and revealing their secrets in the 21st century laboratory. Curr. Opin. Microbiol. **6:**274–281.
12. **Nixon, B. T., C. W. Ronson, and F. M. Ausubel.** 1986. Two-component regulatory systems responsive to environmental stimuli share strongly conserved domains with the nitrogen assimilation regulatory genes *ntrB* and *ntrC*. Proc. Natl. Acad. Sci. USA **83:**7850–7854.
13. **Oka, A., H. Sakai, and S. Iwakoshi.** 2002. His-Asp phosphorelay signal transduction in higher plants: receptors and response regulators for cytokinin signaling in *Arabidopsis thaliana*. Genes Genet. Syst. **77:**383–391.
14. **Orengo, C. A., and J. M. Thornton.** 2005. Protein families and their evolution-a structural perspective. Annu. Rev. Biochem. **74:**867–900.
15. **Pace, N. R.** 1997. A molecular view of microbial diversity and the biosphere. Science **276:**734–740.
16. **Ponting, C. P., and R. R. Russell.** 2002. The natural history of protein domains. Annu. Rev. Biophys. Biomol. Struct. **31:**45–71.
17. **Saito, H.** 2001. Histidine phosphorylation and two-component signaling in eukaryotic cells. Chem. Rev. **101:**2497–2509.
18. **Stock, A., D. E. Koshland, Jr., and J. Stock.** 1985. Homologies between the *Salmonella typhimurium* CheY protein and proteins involved in the regulation of chemotaxis, membrane protein synthesis, and sporulation. Proc. Natl. Acad. Sci. USA **82:**7989–7993.
19. **Torsvik, V., L. Ovreas, and T. F. Thingstad.** 2002. Prokaryotic diversity—magnitude, dynamics, and controlling factors. Science **296:**1064–1066.
20. **Tyson, G. W., J. Chapman, P. Hugenholtz, E. E. Allen, R. J. Ram, P. M. Richardson, V. V. Solovyev, E. M. Rubin, D. S. Rokhsar, and J. F. Banfield.** 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature **428:**37–43.

21. **Ulrich, L. E., E. V. Koonin, and I. B. Zhulin.** 2005. One-component systems dominate signal transduction in prokaryotes. Trends Microbiol. **13:**52–56.

22. **Ulrich, L. E., E. V. Koonin, and I. B. Zhulin.** 22 December 2004, posting date. One-component systems dominate signal transduction in prokaryotes. [Online.] http://genomics.biology.gatech.edu/research/TIM/.

23. **Venter, J. C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Wu, I. Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Levy, A. H. Knap, M. W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y. H. Rogers, and H. O. Smith.** 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science **304:**66–74.

24. **Volz, K.** 1993. Structural conservation in the CheY superfamily. Biochemistry **32:**11741–11753.

25. **Whitman, W. B., D. C. Coleman, and W. J. Wiebe.** 1998. Prokaryotes: the unseen majority. Proc. Natl. Acad. Sci. USA **95:**6578–6583.

26. **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.

The views expressed in this Commentary do not necessarily reflect the views of the journal or of ASM.