Published in final edited form as:

Trends Microbiol. 2005 February ; 13(2): 52–56. doi:10.1016/j.tim.2004.12.006.

One-component systems dominate signal transduction in prokaryotes

Luke E. Ulrich¹, Eugene V. Koonin², and Igor B. Zhulin¹

¹Center for Bioinformatics and Computational Biology, School of Biology, Georgia Institute of Technology, Atlanta, Georgia 30332, GA, USA

²National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, MD, USA

Abstract

Two-component systems that link environmental signals to cellular responses are viewed as the primary mode of signal transduction in prokaryotes. By analyzing information encoded by 145 prokaryotic genomes, we found that the majority of signal transduction systems consist of a single protein that contains input and output domains but lacks phosphotransfer domains typical of two-component systems. One-component systems are evolutionarily older, more widely distributed among bacteria and archaea, and display a greater diversity of domains than two-component systems.

Introduction

Signal transduction pathways in prokaryotes regulate cellular functions in response to environmental cues. According to the current view, prokaryotic signal transduction is conducted mostly by two-component regulatory systems that function as a result of phosphotransfer between two key proteins: a sensor histidine kinase and a response regulator [1-6]. Most histidine kinases that have been experimentally studied are membrane-bound and have an extracellular input domain, whereas all response regulators are cytosolic. In a typical two-component system, the input domain of the sensor histidine kinase detects the environmental signal (usually a small molecule ligand) resulting in the activation of the histidine kinase domain, which autophosphorylates at a specific histidine residue (H; Figure 1a). The phosphoryl group (P) is then transferred to a specific aspartate residue (D) in the receiver domain of a response regulator. Phosphorylation of the response regulator activates the output domain, which initiates the corresponding cellular response. The majority of experimentally characterized two-component systems regulate gene expression at the level of transcription using the DNA-binding helix-turn-helix (HTH) output domains of the response regulators [4,5]. In addition, some response regulators contain enzymatic output domains, such as the di-guanylate cyclase [7].

What is a one-component system?

The modular design of two-component systems is elegant, but does not appear to be the simplest possible solution to the signal transduction challenge - linking environmental stimuli to adaptive responses. A much simpler design for a signal transducer is the direct fusion of an input domain to an output domain in a single protein molecule (Figure 1a). It is well known

that transcription of prokaryotic operons is typically regulated by single-molecule repressors and activators that contain ligand-binding and DNA-binding HTH domains (this is also the case for most two-component systems). The LacI lactose operon repressor [8] and the catabolite activator protein (CAP) [9] of *E. coli* are classic examples of such transcriptional regulators. Although these transcriptional regulators are not normally described as signal transduction systems, it has been noted that they contain some of the same input and output domains that are typical of two-component signal transduction systems [10-15]. For example, PAS (found in period clock protein, aryl hydrocarbon receptor, and single-minded protein) [10] and HTH are input and output domains in the two-component system NtrB-NtrC [16] and in the single-molecule transcriptional regulator RocR [17], respectively (Figure 1b). Using domain database searches, many other combinations of input and output domains have been identified as direct fusions in known or predicted regulatory proteins (Figure 1b). Thus, one-component systems, which are defined as proteins that contain known or predicted input and output domains but lack histidine kinase and receiver domains, appear to have a repertoire of input and output domains similar to that of two-component systems and therefore might detect similar stimuli and elicit similar cellular responses. Sensory and regulatory properties of some of the one-component systems have been well-documented experimentally [18-20].

One-component versus two-component systems: a survey of bacterial and archaeal genomes

The observation that the input and output domains of two-component systems are found in one-component systems prompted us to perform an exhaustive database search and analysis (Box 1), which yielded detailed information on the distribution and co-occurrence of input and output domains in 145 complete and draft prokaryotic genomes (for complete results, please visit our website at <http://genomics.biology.gatech.edu/research/TIM>). Strikingly, this analysis detected many more one-component systems (w17 000) than two-component systems (w4000). More-over, one-component regulators show much greater diversity with regards to the input and output domain repertoire than two-component systems (Figure 2). Many domains are found exclusively in one-component systems, whereas there are no unique input or output domains in two-component systems. The principal type of output activity in both classes of signal transduction systems is regulation of gene expression at the level of transcription: 87% of the known output domains in two-component systems and 84% in one-component regulators are DNA-binding HTH domains. The rest of the output domains are enzymes regulating the level of cyclic nucleotides and protein phosphorylation. The major input activity in both classes is small-molecule-binding: 96% of the known input domains in two-component systems and 93% in one-component regulators are various small-molecule-binding domains. The rest of the input domains are enzymatic and cofactor-containing (mostly redox-responsive) domains and domains involved in protein-protein interactions. Distinct input domains were detected in many but not all of the one-component regulators. However, it is important to note that current computational tools have been unable to detect input domains within protein sequences of several one-component regulators that are known to carry out specific sensory functions. For example, in the CueR

Box 1. Detection of signal transduction proteins in sequenced genomes

The identification of signal transduction proteins is based on the computational domain analysis of protein sequences. We obtained 145 complete and draft prokaryotic genomes from the National Center for Biotechnology Information [a complete list of genomes is available on our website (<http://genomics.biology.gatech.edu/research/TIM>)]. Protein sequences encoded in each genome were searched against the Pfam [28] and SMART [29] domain libraries (hidden Markov models) using the HMMER software package

(<http://hmmer.wustl.edu/>) on a parallel Linux cluster. The resulting domain architectures were stored in a MySQL database. Custom Perl scripts were developed to query the database for domains and domain combinations using regular expressions.

Definitions of input and output domains of prokaryotic signal transduction and gene regulation categories are based on curated assignments in the Pfam-A [28], SMART [29] and clusters of orthologous groups (COGs) [30] resources, and recent genomic surveys [11,12,15]. A complete list of input and output domains used in this study can be found on our website (<http://genomics.biology.gatech.edu/research/TIM>). Two-component systems were identified by the presence of the histidine kinase (HATPase_c) and response regulator (response_reg) domains. One-component regulators were identified by the presence of one or more known output domains and the absence of histidine kinase and response regulator domains. Because many input domains involved in prokaryotic signal transduction also participate in other cellular processes (e.g. ligand-binding in transport and metabolism), they were counted only when found in a combination with the histidine kinase (two-component systems) or a known output domain (one-component regulators). By contrast, output domains typically have a single, specific regulatory function (e.g. transcriptional regulation via binding to specific promoters) and therefore all output domains were counted in the domain analysis.

Cu⁺ sensor-transcriptional activator, the HTH output domain is fused directly to a simple helix-loop-helix element (undetectable by current computational domain searches), which serves as a metal-sensing (input) domain [21]. Therefore, we hypothesize that most, if not all, one-component regulators detected in our genomic analysis via the identification of an output domain participate in various forms of prokaryotic signal transduction.

Genome size, lifestyle and environment contribute to the complexity of signal transduction

The number of one-component and two-component systems per genome positively correlates with the genome size and, in both cases, is roughly proportional to the square of the total number of genes (Figure 3). As shown recently, signal transduction and regulation of gene expression stand out among all functional categories of proteins in showing the steepest dependence on the total number of genes [22,23]. This appears to reflect the disproportionate increase in the hierarchical complexity of gene regulation with the increase in genome size, which might ultimately control the maximum achievable genome size, at least in prokaryotes. Both one-component and two-component systems contribute to this increase in biological complexity, but given the similar exponents of the plots (Figure 3), the contribution of the more abundant one-component regulators is greater. Significant deviations from this general trend appear to reflect particular biological phenomena as well as environmental conditions in a microbial habitat. For example, the genomes of the marine cyanobacterium *Trichodesmium erythraeum* and the soil α -proteobacterium *Sinorhizobium meliloti* are comparable in size: 7.7 and 6.7 Mb, respectively. However, there are 69 one-component regulators encoded in the former (unusually few) versus 390 in the latter (unusually many). The difference in the number of two-component systems in the two genomes is much less noticeable and, in fact, not significant: 35 and 40, respectively. Both bacterial species have a versatile metabolism that is reflected by their large genome size. However, *S. meliloti* has a complex developmental program [24] and experiences significant fluctuation of various physico-chemical parameters in its microenvironments (soil, rhizosphere and plant root interior). By contrast, *T. erythraeum* does not undergo developmental changes typical of other nitrogen-fixing cyanobacteria (heterocyst formation) and lives under more or less constant environmental conditions (upper levels of tropical oceans) [25].

One-component systems as the primordial form of prokaryotic signal transduction

Three lines of evidence suggest that one-component regulators are evolutionary precursors of the two-component systems. First, the modular design of one-component systems is obviously simpler than that of two-component systems. Second, as shown above, the domain repertoire of one-component regulators is considerably more diverse than that of two-component systems. Finally, one-component regulators are more widely distributed among prokaryotes than two-component systems; with the exception of some parasites with highly degraded genomes, such as mycoplasmas, all prokaryotes encode a substantial diversity of one-component regulators. By contrast, two-component systems are missing in many species, particularly, among archaea (Figure 2). Furthermore, archaeal two-component systems have probably been acquired from bacteria via horizontal gene transfer, which has been suggested previously [26]. Therefore, it is possible that the last common ancestor of archaea and bacteria (i.e. the last common ancestor of all modern life forms) did not have two-component systems, but encoded several one-component regulators. Two-component systems appear to be a subsequent bacterial innovation that emerged as a result of insertion of histidine kinase domains and receiver domains into one-component regulators. If one-component regulatory systems comprise such a straightforward solution to the requirements of prokaryotes for signal transduction, then, what is the advantage of two-component systems? We believe that this has to do with the fact that one-component regulators detect stimuli (including environmental cues, such as gases, light and various small molecules) almost exclusively in the cytosol. All 25 303 protein sequences that had been identified as components of signal transduction systems in 145 prokaryotic genomes were scanned for the presence of transmembrane regions using the DAS (dense alignment surface) program [27]. It was found that 97% of the one-component regulators that contain an HTH domain do not have transmembrane regions and therefore are predicted to be cytosolic proteins. By contrast, more than 73% of the sensor histidine kinases were predicted to be membrane-associated on the basis of the presence of one or more transmembrane regions. Thus, the fundamental difference in the sensing mode between the one-component and two-component systems is intracellular versus extracellular detection of stimuli, respectively. Extracellular sensing provides a microbe with an obvious advantage compared with exclusive intracellular sensing. However, because more than 80% of signal transduction pathways involve DNA-binding, arrangement of single-molecule regulators in the membrane would place major constraints on their ability to interact with their targets in genomic DNA. A straightforward and efficient solution to this problem is dividing the signal transduction system into two proteins, a membrane-bound sensor and a soluble cytosolic DNA-binding regulator, which are linked via a phosphotransfer relay. Hence the emergence of the two-component signals transduction systems.

Concluding remarks

The availability of a large number of sequenced prokaryotic genomes has enabled the dominance of one-component signal transduction systems to be revealed and the apparent ancestor-descendant relationship between them and the two-component systems in prokaryotes to be proposed. It had been reported previously that some of the transcriptional regulators in prokaryotes possess sensory properties. However, to our knowledge, it has not yet been recognized that two-component signal transduction systems use a subset of the input and output domains that are present in one-component regulators, or that one-component regulators have extraordinary combinatorial diversity.

Acknowledgements

The literature on signal transduction in prokaryotes is vast. We extend our apologies and appreciation to all colleagues whose work is not cited here solely owing to space limitations. We thank Yuri Wolf for assistance with the genome size dependence analysis and Michael Galperin, Susan Golden and Sydney Kustu for helpful discussions. We also acknowledge the valuable work of the Pfam, SMART and COG developers. This work was supported by research grants GM72285 from National Institutes of Health and EIA-0219079 from National Science Foundation (to I.B.Z.). L.E.U. was supported by IGERT-0221600 grant from National Science Foundation.

References

1. Parkinson JS, Kofoid EC. Communication modules in bacterial signaling proteins. *Annu. Rev. Genet.* 1992;26:71–112. [PubMed: 1482126]
2. Parkinson JS. Signal transduction schemes of bacteria. *Cell* 1993;73:857–871. [PubMed: 8098993]
3. Hoch, JA.; Silhavy, TJ., editors. Two-component signal transduction. ASM Press; 1995.
4. Stock AM, et al. Two-component signal transduction. *Annu. Rev. Biochem* 2000;69:183–215. [PubMed: 10966457]
5. Hoch JA. Two-component and phosphorelay signal transduction. *Curr. Opin. Microbiol* 2000;3:165–170. [PubMed: 10745001]
6. Inouye, M.; Dutta, R., editors. Histidine kinases in signal transduction. Academic Press; 2003.
7. Paul R, et al. Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. *Genes Dev* 2004;18:715–727. [PubMed: 15075296]
8. Lewis M, et al. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* 1996;271:1247–1254. [PubMed: 8638105]
9. Kolb A, et al. Transcriptional regulation by cyclic AMP and its receptor protein. *Annu. Rev. Biochem* 1993;62:749–795. [PubMed: 8394684]
10. Taylor BL, Zhulin IB. PAS domains: internal sensors of oxygen, redox potential and light. *Microbiol. Mol. Biol. Rev* 1999;63:479–506. [PubMed: 10357859]
11. Anantharaman V, et al. Regulatory potential, phyletic distribution and evolution of ancient, intracellular small-molecule-binding domains. *J. Mol. Biol* 2001;307:1271–1292. [PubMed: 11292341]
12. Galperin MY, et al. Novel domains of the prokaryotic two-component signal transduction systems. *FEMS Microbiol. Lett* 2001;203:11–21. [PubMed: 11557134]
13. Shu CJ, Zhulin IB. ANTAR: an RNA-binding domain in transcription antitermination regulatory proteins. *Trends Biochem. Sci* 2002;27:3–5. [PubMed: 11796212]
14. Shu CJ, et al. The NIT domain: a predicted nitrate-responsive module in bacterial sensory receptors. *Trends Biochem. Sci* 2003;28:121–124. [PubMed: 12633990]
15. Zhulin IB, et al. Common extracellular sensory domains in transmembrane receptors for diverse signal transduction pathways in Bacteria and Archaea. *J. Bacteriol* 2003;185:285–294. [PubMed: 12486065]
16. Weiss V, et al. Mechanism of regulation of the bifunctional histidine kinase NtrB in *Escherichia coli*. *J. Mol. Microbiol. Biotechnol* 2002;4:229–233. [PubMed: 11931552]
17. Calogero S, et al. RocR, a novel regulatory protein controlling arginine utilization in *Bacillus subtilis*, belongs to the NtrC/NifA family of transcriptional activators. *J. Bacteriol* 1994;176:1234–1241. [PubMed: 8113162]
18. Spiro S, Guest JR. FNR and its role in oxygen-regulated gene expression in *Escherichia coli*. *FEMS Microbiol. Rev* 1990;6:399–428. [PubMed: 2248796]
19. Shelper D, et al. CoxA, a CO-sensing transcriptional factor from *Rhodospirillum rubrum*, is a CO-binding heme protein. *Proc. Natl. Acad. Sci. U. S. A* 1997;94:11216–11220. [PubMed: 9326589]
20. Vannini A, et al. The crystal structure of the quorum sensing protein TraR bound to its autoinducer and target DNA. *EMBO J* 2002;21:4393–4401. [PubMed: 12198141]
21. Changela A, et al. Molecular basis of metal-ion selectivity and zeptomolar sensitivity by CueR. *Science* 2003;301:1383–1387. [PubMed: 12958362]

22. Van Nimwegen E. Scaling laws in the functional content of genomes. *Trends Genet* 2003;19:479–484. [PubMed: 12957540]
23. Konstantinidis KT, Tiedje JM. Trends between gene content and genome size in prokaryotic species with larger genomes. *Proc. Natl. Acad. Sci. U. S. A* 2004;101:3160–3165. [PubMed: 14973198]
24. Galibert F, et al. The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science* 2001;293:668–672. [PubMed: 11474104]
25. Staal M, et al. Temperature excludes N₂-fixing heterocystous cyanobacteria in the tropical oceans. *Nature* 2003;425:504–507. [PubMed: 14523445]
26. Koretke KK. Evolution of two-component signal transduction. *Mol. Biol. Evol* 2000;17:1956–1970. [PubMed: 11110912]
27. Cserzo M, et al. TM or not TM: transmembrane protein prediction with low false positive rate using DAS-TMfilter. *Bioinformatics* 2004;20:136–137. [PubMed: 14693825]
28. Bateman A, et al. The Pfam protein families database. *Nucleic Acids Res* 2004;32:D138–D141. [PubMed: 14681378]
29. Letunic I, et al. SMART 4.0: towards genomic data integration. *Nucleic Acids Res* 2004;32:D142–D144. [PubMed: 14681379]
30. Tatusov RL, et al. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 2003;4:41. [PubMed: 12969510]

Figure 1.

Two-component and one-component signal transduction. (a) A prototypical two-component signal transduction system contains input (colored red) and output (colored yellow) domains in two different proteins that communicate via a His-Asp phosphotransfer. A one-component system is a protein that contains input and output domains but lacks His-Asp phosphotransfer domains (colored grey and white). (b) Examples of two-component and one-component systems that use the same type of input and output domains. Experimentally studied proteins are identified by name, whereas proteins predicted from genome sequences are identified by their GenBank ID number: FixJ (GenBank accession number 120202), FixL (GenBank accession number 120208), NtrB (GenBank accession number 128594), NtrC (GenBank accession number 417388), RocR (GenBank accession number 34395962), TraR (GenBank accession number 17743220). Organism abbreviations: *A. tumefaciens*, *Agrobacterium tumefaciens*; *B. anthracis*, *Bacillus anthracis*; *B. japonicum*, *Bradyrhizobium japonicum*; *B. pertussis*, *Bordetella pertussis*; *B. subtilis*, *Bacillus subtilis*; *C. violaceum*, *Chromobacterium violaceum*; *E. coli*, *Escherichia coli*; *N. punctiforme*, *Nostoc punctiforme*; *P. putida*, *Pseudomonas putida*; *S. meliloti*, *Sinorhizobium meliloti*; *S. oneidensis*, *Shewanella oneidensis*; *T. maritima*, *Thermotoga maritima*.

Figure 2. Distribution of input and output domains in bacterial and archaeal signal transduction systems. The counts of the 25 most abundant input and output domains in bacterial and archaeal one-component and two-component systems are shown. Domain names are from the curated Pfam-A database (<http://www.sanger.ac.uk/Software/Pfam/>). Detailed information on domain distribution is available at the following website (<http://genomics.biology.gatech.edu/research/TIM>).

Figure 3.

Dependence of the number of one-component and two-component signal transduction systems on the genome size. The plot is in a double logarithmic scale. One hundred forty-five genomes were ranked by size and split into 16 size classes. Each point indicates the average number of genes for one-component or two-component signal-transduction systems in the respective class.