AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Bacteriology®

COMMENTARY

Exploring the (Almost) Unknown: Archaeal Two-Component Systems

Tino Krell[a]

[a]Estación Experimental del Zaidín, Department of Environmental Protection, Consejo Superior de Investigaciones Científicas, Granada, Spain

ABSTRACT Two-component systems (TCS) exist in bacteria and archaea. In contrast to the knowledge of bacterial TCSs, little information is available on their archaeal counterparts. In the current issue of *Journal of Bacteriology*, Galperin and coworkers present a bioinformatics analysis of TCS genes from archaeal genome sequences (M. Y. Galperin, K. S. Makarova, Y. I. Wolf, and E. V. Koonin, J Bacteriol 200:e00681-17, 2018, https://doi.org/10.1128/JB.00681-17). This study identifies different aspects in which TCS-mediated signaling differs in bacteria and archaea and forms a sound basis for the experimental design of studies to increase our knowledge of this poorly investigated protein family.

KEYWORDS *Archaea*, two-component regulatory systems

The capacity to respond to different signals is an essential feature of bacteria and is primarily based on the action of one-component systems (OCS), two-component systems (TCS), and chemosensory pathways. Many OCSs are transcriptional regulators that are typically composed of an input domain for signal sensing and an output domain. TCSs are composed of at least a sensor kinase (SK) and a response regulator (RR). Interestingly, TCSs and transcriptional regulators possess the same types of input and output domains, which has prompted the concept that TCSs have evolved from transcriptional regulators by the insertion of an autokinase and receiver (REC) domain, followed by division into two individual proteins, namely, an SK and an RR (1). In a canonical TCS, the recognition of signal molecules at the extracytoplasmic SK sensor domain creates a molecular stimulus that is transmitted across the membrane, where it modulates the activity of the SK autokinase domain, which in turn alters the transphosphorylation of the RR REC domain. TCS activity is based on the fact that REC domain phosphorylation alters the functional properties of the RR. Chemosensory pathways are more sophisticated versions of TCSs that are composed of six core proteins and, depending on the bacterial species, a set of auxiliary proteins (2). Signals are sensed by chemoreceptors, leading to a modulation of the CheA autokinase activity. The concerted action of the phosphorylation-based excitatory pathway and chemoreceptor methylation-based adaptation mechanisms permits the sensing of signal gradients and, thus, chemotaxis.

OCSs dominate signal transduction processes in bacteria and archaea since they are more abundant, show a wider phylogenetic distribution, and are more diverse in their input and output domains than TCSs (1). Intriguingly, initial studies indicated that archaeal TCSs do not possess the output domains typically found in their bacterial counterparts (1, 3). The predominant physiological role of bacterial TCSs is transcriptional control in response to signal molecules, a function also carried out by transcriptional regulators. However, TCS-based regulation requires greater metabolic and genetic cellular efforts than do transcriptional regulators since two proteins need to be made, and TCS function is energetically costly since it is dependent on ATP hydrolysis.

Accepted manuscript posted online 16 January 2018

Citation Krell T. 2018. Exploring the (almost) unknown: archaeal two-component systems. J Bacteriol 200:e00774-17. https://doi.org/10.1128/JB.00774-17.

Editor Ann M. Stock, Rutgers University-Robert Wood Johnson Medical School

Address correspondence to tino.krell@eez.csic.es.

For the article discussed, see https://doi.org/10.1128/JB.00681-17.

This raises the question of what the advantages of TCS-based regulatory mechanisms are. Most bacterial SKs are membrane-bound receptors that sense their signals in the extracytoplasmic space, while only some are soluble proteins that bind their ligands in the cytosol (4). The capacity to sense signals in the extracytoplasmic space is a major advantage of TCSs. However, this may be only one of several advantages, since further, as-yet-unidentified factors may exist that have influenced the evolutionary choice to control gene expression by either transcriptional regulators or TCSs. For example, six different toluene degradation pathways have been identified. While transcriptional regulators control three of them, the remaining pathways are controlled by TCSs (5). Interestingly, these TCSs are all located in the cytosol, and the advantages of these systems are unknown.

The phylogenetic distribution of archaeal TCSs is irregular, which supports the notion that TCSs are bacterial inventions that have been acquired by archaea via horizontal gene transfer (1, 6). This in turn raises the questions of what the cellular advantages that sustained the TCS acquisition by archaea are and whether the primary physiological roles of bacterial and archaeal TCSs are the same.

However, inspection of the literature on archaeal TCSs reveals an extreme scarceness of experimental data. For example, a PubMed search retrieved about 80 times as many publications if the term "two-component system" was combined with "bacteria" than if it was combined with "archaea." In the current issue of *Journal of Bacteriology*, Galperin and coworkers publish an excellent bioinformatics study in which they have analyzed TCS genes from 218 complete archaeal genomes, as well as from unfinished genomes of metagenomic data sets (7). Like all good scientific articles, the study of Galperin et al. prompts many questions, which in turn serve to define experimental strategies to provide the answers.

LARGE DIFFERENCES IN THE ABSOLUTE AND RELATIVE NUMBERS OF SKs AND RRs

Galperin et al. show that there are large differences in the absolute numbers of SK and RR genes in different archaeal genomes, confirming the initial studies (1). No TCS genes were found in genomes from the *Cren-*, *Kor-*, and *Nanoarchaeota*; however, this finding is to be considered with caution due to the low number of genome sequences available from the latter two phyla. In contrast, the idea that TCSs are absent from members of the *Crenarchaeota* is based on 86 completed genomes, a finding that confirms the initial observations (4). However, TCS genes were detected in 5 draft crenarchaeal genomes, a fact that underlines the value of such sequence information. TCSs were most abundant in genomes from the *Halobacteria* and *Methanomicrobia*, which possess more than 40 TCS genes on average, based on more than 30 genome sequences for each class. TCSs were most abundant in the methanogen *Methanospirillum hungatei* strain JF-1, with 133 SK and RR genes. In bacteria, there was a good correlation between the numbers of SK and RR genes per genome. This ratio was close to 1, suggesting that bacterial TCSs function primarily in pairs. In contrast, archaeal SK/RR ratios were more scattered, and in many genomes, the SK genes outnumbered the RR genes. This may suggest that fewer TCSs operate as strict pairs in archaea and may point to mechanisms where a single RR is phosphorylated by multiple SKs.

THE IMPORTANCE OF CYTOSOLIC SIGNAL SENSING IN ARCHAEA

The capacity to sense signals in the extracytoplasmic space was proposed as a major force that has led to the evolution of TCSs from OCSs (1). Several studies have calculated the percentage of bacterial SKs with transmembrane regions, which was found to be between 73 and 88% (1, 4, 7, 8). Most of these proteins are suspected to operate by a mechanism involving extracytoplasmic signal sensing. Cyanobacteria were identified as the only bacterial phylum that had a comparatively low percentage of membrane-bound sensor proteins (4). In accordance with their previous observations that were based on a reduced number of sequences (4), Galperin et al. (7) now show that the majority of archaeal SKs, 62% on average, lack transmembrane regions,

indicative of the cytosolic location of the SKs. In this context, striking parallels to chemoreceptors, the other large family of prokaryotic sensor proteins, exist (9). In bacteria, the large majority (86%) of chemoreceptors are transmembrane proteins, whereas in archaea, the share of these proteins is only 57% (9). These data suggest that SKs and chemoreceptors play more important roles in the sensing of cytosolic signals in archaea than in bacteria.

This notion is further supported by the analysis of the sensor domain types in both protein families. Galperin et al. (7) report that 72% of archaeal SKs carried a PAS and/or a GAF domain, whereas analyses by Zhang and Hendrickson (10) indicate that the dCACHE (formerly dPDC) domain is the predominant sensor domain in the bacterial SKs. The relative abundances of sensor domains in archaeal and bacterial chemoreceptors are again comparable to the situation in SKs. The PAS domain, at 47%, is by far the most abundant sensor domain in archaeal chemoreceptors (9), whereas 4-helix-bundle (4HB) and dCACHE domains are the most abundant domains in bacterial chemoreceptors (11). Interestingly, the currently available data indicate that 4HB and dCACHE are exclusive to extracellular sensing modules (12, 13), whereas PAS and GAF domains are restricted to cytosolic sensing modules (13–16). Taken together, the elevated percentages of archaeal SKs with PAS/GAF domains underline the importance of cytosolic sensing in archaea.

TRANSCRIPTIONAL REGULATION MAY NOT BE THE PRIMARY FUNCTION OF ARCHAEAL TCSs

The REC domain of RRs exists either as a stand-alone protein or fused to other domains that, in general, mediate the signaling output (17, 18). Inspection of the RR domain arrangement can therefore provide initial clues as to the function of a given TCS. The large majority of bacterial RR output domains are DNA-binding domains, most of which contain the helix-turn-helix (HTH) DNA-binding motif (18). This suggests that the primary function of bacterial RRs is transcriptional regulation, an idea that is documented by a very large body of experimental evidence. Previous studies have indicated that archaea contain a low number of RRs with a DNA-binding output domain (3, 18). The present study of Galperin et al. (7), in which they have listed REC-associated domains, confirms this observation. Although some of these domains are novel (see below) or of unknown function, it is obvious that only relatively few possess DNA-binding activity. For example, HTH motif-containing DNA-binding domains amount to only 6% of the total archaeal REC-associated domains. These analyses indicate that transcriptional regulation may not be the primary function of archaeal TCSs.

ELEVATED NUMBER OF STAND-ALONE RECEIVER DOMAINS

Confirming earlier observations (3, 17, 18), Galperin et al. (7) show that archaea contain an elevated number (40%) of response regulators that are solely composed of an REC domain and are referred to as stand-alone REC domains. Multiple functions have so far been identified for this class of proteins. CheY is a stand-alone REC protein that, in its phosphorylated form, binds to the flagellar motor, mediating chemotaxis. Other stand-alone RRs exert regulatory functions by acting as phosphate sinks, both in the context of chemotaxis (19) and transcriptional regulation (20). These proteins compete with other RRs for phosphoryl groups from a single histidine kinase. In addition, stand-alone RRs were found to form part of more-complex multiprotein phosphorelay systems, of which the best-characterized example is the sporulation system of *Bacillus subtilis* (21). Galperin et al. (7) report the conservation of amino acids essential for phosphorylation in the archaeal stand-alone RRs, suggesting that these are active proteins. Since only a small fraction of these proteins are encoded in chemosensory gene clusters, chemotaxis may not be their predominant function. In addition, no obvious clues concerning the function of stand-alone RRs could be obtained from the inspection of their genetic environment, making laboratory experimentation indispensable to identify their function.

THE UNKNOWN FUNCTION OF REC-(PAS)$_n$ AND REC-PAS-GAF PROTEINS

The canonical RR is composed of an REC input domain and an output domain. Bacterial output domains are primarily DNA-binding domains, but a series of other output domains have been identified, such as those that possess enzyme activities or bind to RNA (18). PAS and GAF domains are typical input domains, omnipresent in bacterial signal transduction systems that bind a chemically diverse range of small molecules (22, 23). Previous work has provided some initial evidence of fusions of REC with PAS and GAF domains (18). In archaea, Galperin et al. (7) have now observed that a large share of archaeal RRs possess an REC-(PAS)$_n$ or REC-PAS-GAF domain arrangement. Such proteins are thus RRs that are entirely composed of domains that are considered to mediate the signal input. The function of these proteins is unknown, but the authors speculate that phosphorylation or ligand binding may cause protein dimerization, which in turn may trigger an output. However, experimental studies are required to unveil the secret of these proteins.

NOVEL OUTPUT DOMAINS IN ARCHAEA

Galperin et al. (7) established a repertoire of archaeal REC-associated domains. They identified 14 different known domains, the most abundant of which were PAS, GAF, CheB, and HalX. In contrast to the first three domains, which are well studied, the function of the HalX domain is unknown. Based on secondary structure predictions indicating that this domain consists of three α-helices, the authors hypothesize that it may form a structure similar to those of HTH motif-containing DNA-binding domains. However, no functional studies are available for this domain. Among the archaeal REC-associated domains was the KaiC-like ATPase domain, known to be a key component of the circadian clock in cyanobacteria (24). In another recent article (25), the authors show that this domain is abundant and is present in most archaeal genomes, in contrast to its patchy phylogenetic distribution in bacteria. KaiC domain-containing proteins were frequently found to harbor characterized or potential input and output domains or domains known to participate in signaling processes. The authors therefore suggest that there are extensive KaiC-centered signal transduction networks in archaea that are predicted to play major roles in their physiology.

In addition, the authors (7) identify eight novel REC-associated domains and provide the Pfam entry codes to facilitate their identification. Most of these domains showed a narrow phylogenetic distribution and were found exclusively in genomes from the *Halobacteria*, *Methanobacteria*, or *Thaumarchaeota*. No information is available as to the function of these domains. Based on the secondary structure prediction and the presence of certain sequence motifs, the authors hypothesize that the function of these domains may consist of DNA/RNA or small ligand binding or of different enzymatic activities. Although Galperin et al. (7) referred to these novel domains as output domains, future research will have to show whether this indeed is the case. In analogy to the above-mentioned REC-(PAS)$_n$ and REC-PAS-GAF proteins, which are exclusively composed of domains considered to mediate input, it cannot be excluded that some of these novel domains have input functions.

CONCLUDING REMARKS

The analysis of archaeal sequences indicates that archaeal TCSs differ in many respects from the bacterial sequences. The main differences are related to the primary physiological role, the types of signal molecules recognized, and the cellular compartment of sensing, as well as mechanistic differences mainly related to novel domains or atypical domain fusions. Experimental research is now required to precisely identify and characterize these differences, which is necessary information to understand the pressures that have led to the acquisition and evolution of archaeal TCSs.

ACKNOWLEDGMENTS

This work was supported by a grant held by T. Krell from the Spanish Ministry for Economy and Competitiveness (grant number BIO2016-76779-P).

REFERENCES

1. Ulrich LE, Koonin EV, Zhulin IB. 2005. One-component systems dominate signal transduction in prokaryotes. Trends Microbiol 13:52–56. https://doi.org/10.1016/j.tim.2004.12.006.
2. Wuichet K, Zhulin IB. 2010. Origins and diversification of a complex signal transduction system in prokaryotes. Sci Signal 3:ra50. https://doi.org/10.1126/scisignal.2000724.
3. Galperin MY. 2006. Structural classification of bacterial response regulators: diversity of output domains and domain combinations. J Bacteriol 188:4169–4182. https://doi.org/10.1128/JB.01887-05.
4. Galperin MY. 2005. A census of membrane-bound and intracellular signal transduction proteins in bacteria: bacterial IQ, extroverts and introverts. BMC Microbiol 5:35. https://doi.org/10.1186/1471-2180-5-35.
5. Parales RE, Parales JV, Pelletier DA, Ditty JL. 2008. Diversity of microbial toluene degradation pathways. Adv Appl Microbiol 64:1–73. https://doi.org/10.1016/S0065-2164(08)00401-2.
6. Koretke KK, Lupas AN, Warren PV, Rosenberg M, Brown JR. 2000. Evolution of two-component signal transduction. Mol Biol Evol 17:1956–1970. https://doi.org/10.1093/oxfordjournals.molbev.a026297.
7. Galperin MY, Makarova KS, Wolf YI, Koonin EV. 2018. Phyletic distribution and lineage-specific domain architectures of archaeal two-component signal transduction systems. J Bacteriol 200:e00681-17. https://doi.org/10.1128/JB.00681-17.
8. Cock PJ, Whitworth DE. 2007. Evolution of prokaryotic two-component system signaling pathways: gene fusions and fissions. Mol Biol Evol 24:2355–2357. https://doi.org/10.1093/molbev/msm170.
9. Collins KD, Lacal J, Ottemann KM. 2014. Internal sense of direction: sensing and signaling from cytoplasmic chemoreceptors. Microbiol Mol Biol Rev 78:672–684. https://doi.org/10.1128/MMBR.00033-14.
10. Zhang Z, Hendrickson WA. 2010. Structural characterization of the predominant family of histidine kinase sensor domains. J Mol Biol 400:335–353. https://doi.org/10.1016/j.jmb.2010.04.049.
11. Ortega A, Zhulin IB, Krell T. 2017. Sensory repertoire of bacterial chemoreceptors. Microbiol Mol Biol Rev 81:e00033-17. https://doi.org/10.1128/MMBR.00033-17.
12. Ulrich LE, Zhulin IB. 2005. Four-helix bundle: a ubiquitous sensory module in prokaryotic signal transduction. Bioinformatics 21(Suppl 3):iii45–iii48. https://doi.org/10.1093/bioinformatics/bti1204.
13. Upadhyay AA, Fleetwood AD, Adebali O, Finn RD, Zhulin IB. 2016. Cache domains that are homologous to, but different from PAS domains comprise the largest superfamily of extracellular sensors in prokaryotes. PLoS Comput Biol 12:e1004862. https://doi.org/10.1371/journal.pcbi.1004862.
14. Ponting CP, Aravind L. 1997. PAS: a multifunctional domain family comes to light. Curr Biol 7:R674–R677. https://doi.org/10.1016/S0960-9822(06)00352-6.
15. Zhulin IB, Taylor BL, Dixon R. 1997. PAS domain S-boxes in Archaea, Bacteria and sensors for oxygen and redox. Trends Biochem Sci 22:331–333. https://doi.org/10.1016/S0968-0004(97)01110-9.
16. Aravind L, Ponting CP. 1997. The GAF domain: an evolutionary link between diverse phototransducing proteins. Trends Biochem Sci 22:458–459. https://doi.org/10.1016/S0968-0004(97)01148-1.
17. Jenal U, Galperin MY. 2009. Single domain response regulators: molecular switches with emerging roles in cell organization and dynamics. Curr Opin Microbiol 12:152–160. https://doi.org/10.1016/j.mib.2009.01.010.
18. Galperin MY. 2010. Diversity of structure and function of response regulator output domains. Curr Opin Microbiol 13:150–159. https://doi.org/10.1016/j.mib.2010.01.005.
19. Sourjik V, Schmitt R. 1998. Phosphotransfer between CheA, CheY1, and CheY2 in the chemotaxis signal transduction chain of *Rhizobium meliloti*. Biochemistry 37:2327–2335. https://doi.org/10.1021/bi972330a.
20. Crosson S, McGrath PT, Stephens C, McAdams HH, Shapiro L. 2005. Conserved modular design of an oxygen sensory/signaling network with species-specific output. Proc Natl Acad Sci U S A 102:8018–8023. https://doi.org/10.1073/pnas.0503022102.
21. Hoch JA. 2017. A life in *Bacillus subtilis* signal transduction. Annu Rev Microbiol 71:1–19. https://doi.org/10.1146/annurev-micro-030117-020355.
22. Henry JT, Crosson S. 2011. Ligand-binding PAS domains in a genomic, cellular, and structural context. Annu Rev Microbiol 65:261–286. https://doi.org/10.1146/annurev-micro-121809-151631.
23. Martinez SE, Beavo JA, Hol WG. 2002. GAF domains: two-billion-year-old molecular switches that bind cyclic nucleotides. Mol Interv 2:317–323. https://doi.org/10.1124/mi.2.5.317.
24. Cohen SE, Golden SS. 2015. Circadian rhythms in cyanobacteria. Microbiol Mol Biol Rev 79:373–385. https://doi.org/10.1128/MMBR.00036-15.
25. Makarova KS, Galperin MY, Koonin EV. 2017. Proposed role for KaiC-like ATPases as major signal transduction hubs in Archaea. mBio 8:e01959-17. https://doi.org/10.1128/mBio.01959-17.