Amino Acids (2009) 37:459–466
DOI 10.1007/s00726-009-0259-2

REVIEW ARTICLE

Evolution of prokaryotic two-component systems: insights from comparative genomics

David E. Whitworth · Peter J. A. Cock

Received: 22 January 2009 / Accepted: 2 February 2009 / Published online: 25 February 2009

Abstract Two-component systems (TCSs) are diverse and abundant signal transduction pathways found predominantly in prokaryotes. This review focuses on insights into TCS evolution made possible by the sequencing of whole prokaryotic genomes. Typical TCSs comprise an autophosphorylating protein (a histidine kinase), which transfers a phosphoryl group onto an effector protein (a response regulator), thus modulating its activity. Histidine kinases and response regulators are usually found encoded as pairs of adjacent genes within a genome, with multiple examples in most prokaryotes. Recent studies have shed light on major themes of TCS evolution, including gene duplication, gene gain/loss, gene fusion/fission, domain gain/loss, domain shuffling and the emergence of complexity. Coupled with an understanding of the structural and biophysical properties of many TCS proteins, it has become increasingly possible to draw inferences regarding the functional consequences of such evolutionary changes. In turn, this increase in understanding has the potential to enhance both our ability to rationally engineer TCSs, and also allow us to more powerfully correlate TCS evolution with behavioural phenotypes and ecological niche occupancy.

D. E. Whitworth (✉)
Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Cledwyn Building, Penglais Campus, Ceredigion SY23 3DD, UK
e-mail: dew@aber.ac.uk

P. J. A. Cock
MOAC Doctoral Training Centre, Coventry House, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK

Keywords Response regulator · Histidine kinase · Hybrid kinase · Gene fusion · Gene fission · Evolution · Recombination · Duplication · Domain

Two-component systems are abundant prokaryotic signalling pathways

All organisms must respond to a changing environment. However, environmental change can occur over many different geographical and temporal scales, and there are consequently several types of biological responses to environmental change, including migration, extinction, adaptation and evolution. Evolution, migration and extinction are often not viable options for an individual organism when responding to environmental change; however, historical changes in the environment have led to the evolution of signalling pathways, which allow rapid and robust responses to short-term changes of the environment, through the process of adaptation. This review focuses on the evolution of an abundant and virtually ubiquitous adaptive mechanism, the two-component system (TCS).

TCSs are the dominant signal transduction pathways of the prokaryotes (Galperin 2005; Ulrich et al. 2005). In addition, TCSs and their variants are often found in eukaryotes (for example regulating circadian control in plants, and osmotic adaptation in yeast), although they have not yet been found in animals, making them attractive targets for novel bioactives (Stephenson and Hoch 2004; Watanabe et al. 2008). For virtually all bacterial behaviours, examples have been described in the literature that are regulated by TCSs (including virulence, pathogenicity, motility, nodulation, nutrient uptake, secondary metabolite production, metabolic regulation, cell division, etc.).

Typical TCSs comprise a receptor protein (a histidine protein kinase, or HK), and an effector protein (response regulator, or RR), usually encoded by a pair of adjacent genes. Both the HK and the RR are multi-domain proteins. HKs are defined by the possession of a transmitter domain, which is a combination of two sub-domains (HisKA/Hpt and HATPase), typically at the C-terminus of the HK protein. The N-termini of HKs are diverse, and usually contain sensory or 'input' domains, which respond to changes in environmental stimuli. RRs typically comprise an N-terminal receiver domain, with diverse C-terminal effector or 'output' domains. Upon stimulus perception by an input domain, the transmitter domain of an HK is activated, leading to autophosphorylation of the HK at a conserved His residue within the transmitter domain. Phosphorylated transmitter domains are able to interact with the receiver domains of their partner (cognate) RRs. Formation of a phosphotransmitter–receiver domain complex allows transfer of the phosphoryl group from the transmitter His residue onto a conserved Asp residue within the RR receiver domain. Phosphorylation of the RR receiver domain then alters its interaction with the RR output domain, up- or down-regulating effector activity (Gao et al. 2007).

A common and expanded form of the basic TCS is known as a phosphorelay. In phosphorelays, a phosphoryl group is shuttled alternately between the conserved His and Asp residues found in successive transmitter, receiver, histidine-containing phosphotransfer domain (Hpt) and receiver domains. Often these domains are found together on single polypeptides, with the most common example comprising a fusion of the first three domains (transmitter, receiver, Hpt), with the terminal phospho-accepting receiver domain encoded as a separate RR protein (Appleby et al. 1996; Zhang and Shi 2005).

Whilst the basic scheme of TCS and phosphorelay action presented above appears to be the situation for most systems, more complex/atypical examples are commonplace. For instance, many HKs and RRs have multiple partners, whilst some systems posses 'extra' receiver and/or transmitter domains whose phosphorylation state affects phosphoryl group transfer to effector RRs. Transmitter domains can often also act as phosphatases towards phosphorylated forms of their partner RRs. A discussion of all facets of TCS signalling is far beyond the scope of this review, and we would point the interested reader towards the following reviews for details of TCS structure, function and dynamics (Stock et al. 1989; Robinson et al. 2000; Bijlsma and Groisman 2003; Stephenson and Lewis 2005). Instead, we focus on recent insights into the evolutionary pressures acting on prokaryotic TCSs, made possible by whole-genome sequencing.

Most TCSs can be categorised into a handful of large families

As a consequence of whole-genome sequencing and metagenomics, there are already staggering numbers of TCS protein and gene sequences in the public databases. Analysing the domains found in the predicted proteomes of 316 bacteria showed that receiver and transmitter (HATPase) domains are the second and third most numerous PFAM domains in bacteria, surpassed only by ABC transporters (Whitworth and Cock 2008a). The volume of publicly available genomic sequence is so large that devoted databases have had to be developed, including SENTRA, MiST and P2CS (http://compbio.mcs.anl.gov/sentra; http://genomics.ornl.gov/mist/; http://www.p2cs.org; D'Souza et al. 2007; Ulrich and Zhulin 2007; Barakat et al. 2009). In the current version of the P2CS database, there can be found 53,233 TCS proteins encoded within 755 prokaryotic replicons, and 39 metagenomes. The average bacterial genome is around 3.5 Mbp and encodes about 50 TCS proteins (Whitworth 2008). Cyanobacteria and the myxobacteria appear exceptional, with some members of each group encoding >240 TCS proteins (Whitworth and Cock 2008a). As might be expected, large numbers of TCSs within a genome seems to correlate with complexity of life-style and the changeability of an organism's environment (Ashby 2004; Galperin 2005).

With such large numbers of entries, TCS databases have needed to classify TCS proteins in order to usefully catalogue them. There have generally been two approaches adopted, leading to a large number of classification schemes. In one approach, phylogenetic trees are constructed based on the sequences of transmitter and receiver domains (Grebe and Stock 1999; Koretke et al. 2000; Kim and Forst 2001). The second approach classifies proteins on the basis of which input/output domains are found within the TCS (Mizuno 1997; Galperin 2006; Whitworth and Cock 2008a). Although these two approaches have birthed a large number of overlapping classification schemes (and secondary approaches), the results of most classifications agree to a surprising extent, certainly at a gross level. It seems that the majority of TCS proteins belong to a relatively small number of major 'families', which share common ancestry, and gene/domain architecture.

Early phylogenetic analyses suggested the existence of 5–11 major families of TCS (Grebe and Stock 1999; Koretke et al. 2000; Kim and Forst 2001). With more TCS sequences available, it now appears that by far the most common families are the Che, Ntr, Omp and Nar families (named here after archetypal family members in *Escherichia coli*) (Galperin 2006; Barakat et al. 2009). These families have expanded in size during evolution, with the average bacterium containing multiple examples of each

TCS family. However, many organisms have disproportionately expanded some TCS families, at the expense of others. For instance, *Myxococcus xanthus* encodes 30 Ntr family TCSs (Whitworth and Cock 2008a), whilst *Streptomyces coelicolor* possesses no Ntr family TCSs, but does have 41 Nar family TCSs (Hutchings et al. 2004).

Members of a major family tend almost exclusively to possess the same output domains (and input domains to a lesser extent), and within an organism will largely have the same gene order/architecture. Additionally, all domains within such systems (particularly output, transmitter, and receiver domains) will generally exhibit congruent phylogeny, implying linear descent from the same common ancestor (Pao and Saier 1995; Fabret et al. 1999; Koretke et al. 2000; Qian et al. 2008), although examples of recruitment (paired HK and RR genes displaying incongruent phylogeny) are to be found (Koretke et al. 2000; Chen et al. 2004). This observation also provides an explanation of why classification schemes based on phylogenetics and those based on domain architecture give largely similar results.

Presumably, there was a single original TCS, from which the major family ancestors duplicated and diverged. Such a process of duplication and divergence, creating new families, appears to have continued to occur throughout bacterial evolution, with the emergence of minor families of TCS proteins. Contemporary members of such minor families usually possess TCS domains that phylogenetically fall within clades of the major family members, implying adoption of new input/output functions by domain shuffling (see for example Pao and Saier 1995; Stephenson and Hoch 2002; Qian et al. 2008). Analysis of incongruence in the results of phylogenetic and domain architecture-based classification schemes has the potential to yield interesting insights into the frequency and rate of domain shuffling during TCS evolution, a consensus on which has yet to be reached, particularly for HKs, which possess exceptionally diverse (combinations of) input domains. Certainly, some well-studied systems show extensive domain shuffling, for instance the input domains of the sporulation HKs of *Bacillus* species (Stephenson and Hoch 2002), and HK input domains of organisms exhibiting large lineage specific expansions (Alm et al. 2006).

TCS gene duplication

Although significant numbers of TCSs in some organisms are acquired by lateral transfer (Alm et al. 2006), for most organisms the contemporary set of TCSs belonging to the major families can be most easily explained by gene duplication (Pao and Saier 1995; Koretke et al. 2000; Qian et al. 2008; Whitworth and Cock 2008a). The evolutionary duplication of TCSs has some interesting theoretical repercussions. Presumably, immediately upon duplication the genome will encode two identical 'daughter' TCSs. Not only will these daughter systems be redundant, but the components of each will be indistinguishable from one another, enabling cross-communication between the two systems. In some cases, one of the daughters will be lost from the genome without further consequences. However, in other cases there will be retention and sequence divergence of the two daughter TCSs. During this process the ability to cross-communicate may be lost or retained, depending on the nature of any changes at the transmitter–receiver interfaces of the two systems (Bijlsma and Groisman 2003). In *M. xanthus*, the two most similar TCSs (Pho2 and Pho3) appear to have arisen during a duplication in the lineage forming the Cystobacterinae. Both Pho2 and Pho3 remain members of the *M. xanthus* Pho regulon, however, they give overlapping phenotypes upon gene disruption (Moraleda-Muñoz et al. 2003) and are differentially regulated (Whitworth et al. 2008a), although contemporary proteins of the two systems apparently remain able to cross-communicate (Whitworth et al. 2008b; Cock and Whitworth 2009). Similarly, the Nar system in *E. coli* regulates nitrate and nitrite metabolism, and is an example of a small TCS network. Encoded by neighbouring genes, NarX and NarL are an ordinary pair of HK and RR proteins. However, two further locations in the genome encode another HK, NarQ, and another RR, NarP. The TCS domains of these genes are more similar to each other than any other TCS genes in *E. coli*. Both NarX and NarQ will phosphorylate both NarL and NarP, although the interactions between these proteins are not fully symmetric, and they are also regulated differently (Bijlsma and Groisman 2003). However, a comprehensive experimental study suggests that cross-communication between the same-family TCSs does not occur significantly in vivo, primarily due to kinetic preferences (Skerker et al. 2005), implying an evolutionary insulation of daughter systems.

There are many ways in which daughter TCSs can diverge, giving a selective advantage for the retention of both systems. For instance, daughters can become differentially regulated, their catalytic activities can alter, and the specificity of their input and output domains can change, subtly altering the signal-response coupling of the system. Such changes will have implications for whether the daughter systems can (or must) retain or lose their ability to cross-communicate. It is difficult to imagine that global trends regarding the mechanisms leading to divergence of duplicated TCSs will emerge, due to the diversity of TCSs, and our current inability to correlate

TCS gene sequence with in vivo properties. However, it is possible that experimental reconstructions of 'ancestral' TCSs may clarify the selective pressures operating on duplicated TCSs. It is also possible that careful comparative genomic analysis will be able to shed some light on some specific aspects of the problem. For instance, if considering only recently duplicated TCSs, what are the relative selective pressures acting on input and output domains? How do they relate to the selective pressure acting on the interfacial amino acid residues governing the specificity of the HK-RR interaction? And how do these pressures compare to selective pressures acting on unduplicated TCSs? Is divergence a slow and gradual process or does it happens in large jumps, for instance through domain shuffling? Recently, Qian et al. (2008) have suggested that accumulation of point mutations has not been the dominant factor driving evolution of TCSs in *Xanthomonas* spp., instead recombination, gene fusion/fission and insertion/deletion have been more significant drivers of divergence. In any case it will be important to discriminate between those TCSs that have diverged and gained new functions conferring a selective advantage, from those that have merely accumulated mutations and are non-functional, conferring no advantage on the host organism and awaiting loss from the genome. The relatively high frequency of TCS pseudogenes, and TCSs entirely lacking one of the two components observed in many genomes, suggests that this is not a trivial problem (Tong et al. 2005; Martiny et al. 2006; Hinchliffe et al. 2008; Qian et al. 2008).

TCS gene organisation

It is generally accepted that there is a selective pressure in prokaryotic genomes for the co-localisation of genes of related function (Lawrence 1999). Perhaps unsurprisingly, typical TCSs are usually encoded by a pair of adjacent, co-directional genes, and these genes frequently overlap, allowing for their co-ordinated expression and regulation (Cock and Whitworth 2007b). The order of the two genes does not seem to matter for TCS function (Cock and Whitworth 2007a), although gene order is often largely conserved within the major TCS family members of an organism (Fabret et al. 1999; Whitworth and Cock 2008b). Additionally, the genes of TCSs are often found adjacent to genes whose expression they regulate, or adjacent to genes whose products they control the activity of, for example *che* gene clusters and Pho regulons (Martiny et al. 2006; Zusman et al. 2007).

However, in some cases HK and RR genes can be found in isolation (orphans), or in clusters of TCS genes. TCSs encoded by complex gene clusters are generally found to participate in the same signalling pathway (see for example Higgs et al. 2005; Wegener-Feldbrügge and Søgaard-Andersen 2009), however, the presence of multiple HKs and RRs precludes trivial inference of pathway action from genome organisation. If an HK or RR is encoded by an orphan gene, that protein can often be shown to belong to an atypical (relatively complex) TCS, with more than one signalling partner. That the orphan has more than one partner seems to prevent its genetic localisation with either partner gene, though why this should be the case is uncertain (for well-studied examples, see the Spo pathway of *Bacillus subtilis*, and the Nar systems of *E. coli*, reviewed by Bijlsma and Groisman 2003).

Hybrid kinases

Paired genes encoding a typical TCS, can often be found fused together, forming a hybrid kinase that encodes a protein with both a transmitter and receiver domain (and often also input and output domains). In such cases, phosphotransfer between transmitter and receiver domains occurs within the single protein. Several features of a TCS appear to affect whether such gene fusions result in a functional hybrid kinase (Cock and Whitworth 2007a). Relative order of the parent HK and RR genes dictates a minimal distance between encoded transmitter and receiver domains (presumably manifested as a linker domain large enough to allow an appropriate geometry for the transmitter–receiver complex). Additionally, an important determinant of successful fusion appears to be the nature of any input and output domains, particularly regarding any localisation requirements of such domains. Transmitter and receiver domains reside in the bacterial cytoplasm. Many input domains contain transmembrane (TM) helices, implying detection of extracellular stimuli (Galperin 2005). Conversely, many output domains have DNA-binding activity and thus will exert their regulatory effects within the bacterial nucleoid. TCSs that possess both TM helices and DNA-binding domain appear virtually incapable of fusing to form a functional hybrid kinase, presumably because their membrane localisation precludes DNA-binding and effector function upon stimulation (Ulrich et al. 2005; Cock and Whitworth 2007a).

Early phylogenetic analyses suggested that all hybrid kinases were monophyletic, belonging to a 'hybrid' clade (Grebe and Stock 1999; Koretke et al. 2000). This observation suggested that there had been very occasional TCS gene fusion events early in TCS evolution, from which all contemporary hybrid kinases had emerged by duplication. It now seems that the initial observation of 'hybrid' clades was an artefact. Successful hybrid kinase formation relies predominantly upon having a particular HK/RR gene order, and having input/output domains of particular functional

classes. These features of a TCS are conserved within a major TCS family, and therefore hybrids will seem to appear preferentially within a subset of TCS phylogenetic clades, and when dealing with small numbers of sequences, often apparently monophyletically. Phylogenetic analysis using large numbers of hybrid kinase and two-gene TCS sequences (>3,000), suggests that TCS gene fusion (and indeed fission of hybrid kinases back into two-gene TCSs) has occurred sporadically, but throughout TCS evolution, followed by selective gene duplication (Zhang and Shi 2005; Cock and Whitworth 2009). Contemporary examples of hybrid kinases can now be found within most TCS phylogenetic clades (Cock and Whitworth 2009). Interestingly, in *Xanthomonas* spp., hybrid kinases were found to exhibit significantly greater levels of polymorphisms than HK or RRs, suggesting relatively rapid evolution of hybrid kinases (Qian et al. 2008).

Mutational events affecting gene organisation

It would seem reasonable to assume that the fusion/fission of adjacent TCS genes would have occurred predominantly through point mutations. However, recent data (Qian et al. 2008; Cock and Whitworth 2009) suggest that fusion/fission events are generally (our unpublished data suggests perhaps as often as 50% of cases) associated with insertion/deletion mutations, which often add or remove potential input/output domains from the TCS. Additionally, phylogenetic incongruence of partnered transmitter and receiver domains suggests that many fusion/fission events have arisen as a consequence of recombination between TCSs from different TCS families (Zhang and Shi 2005; Qian et al. 2008; Cock and Whitworth 2009).

Whilst the majority of hybrid kinases consist of single transmitter and receiver domains, many hybrid kinases have multiple transmitter and/or receiver domains (complex hybrids), particularly in cyanobacterial and deltaproteobacterial genomes (Zhang and Shi 2005; Whitworth and Cock 2008a). It seems that these complex hybrids have generally been constructed piecemeal through the serial addition of individual TCS domains (Zhang and Shi 2005), although we have preliminary data that at least some myxobacterial complex hybrids may have arisen as a consequence of recombination between pre-existing hybrid kinases (Cock and Whitworth 2009).

We have also obtained preliminary phylogenetic evidence that during evolution the order of paired HK and RR genes can invert, with an apparent frequency even greater than that of fusion/fission (Cock and Whitworth 2009). Such gene inversion events can occur within a TCS (around two-thirds of cases), or as a consequence of recombination between TCSs, presumably through illegitimate recombination (Michel 1999).

TCS domain architectures

During TCS evolution most HKs and RRs have undergone changes to their domain architecture. Whilst many such changes can be readily explained by fusion/fission/recombination between extant TCSs, changes involving novel domain acquisition and domain loss cannot be so easily explained. Throughout their evolution, TCSs have acquired novel input and output domains. Presumably, these domain acquisitions have occurred by gene fusion or recombination with non-TCS proteins. Evidence of such events can be observed in phylogenetic trees, when clades of TCS proteins with receiver/transmitter domains of shared heritage contain a protein whose input/output domain does not match other members of its clade (see supporting information in Whitworth and Cock 2008a for examples). TCS input/output domain loss can also be commonly seen using phylogenetic approaches. For instance, many organisms contain RRs that lack output domains. The archetype of such proteins is the CheY protein of enteric bacteria, which regulates flagella rotation (Armitage 1999). Whilst most bacteria that exhibit chemotaxis typically encode one CheY homologue, some organisms also possess many 'CheY-like' proteins that consist of just a receiver domain (Mizuno et al. 1996; Whitworth and Cock 2008a). Genes encoding CheY proteins tend to lie in clusters of genes involved with motility, including methyl-accepting chemotaxis proteins (MCPs) and MCP methylases and demethylases (Zusman et al. 2007). Such CheY proteins are monophyletic, however, CheY-like proteins are found distributed throughout receiver domain phylogenetic trees. This observation implies that such proteins have arisen by the deletion of the output domain from another TCS, and usually phylogenetic approaches can suggest what output domain the protein's ancestor originally contained. In many cases, truncated output domains can be found C-terminal to the receiver domain of CheY-like proteins, supporting such suggestions (Cock and Whitworth 2009).

How do TCSs retain/regain functionality after domain gain/loss? Receiver domains appear to function by engaging in phosphorylation state-dependent protein–protein interactions with effector domains (Gao et al. 2007). Such receiver–effector domain interactions usually occur within a RR, however, many RRs (for example CheY) act through inter-protein interactions. Similarly, some HKs sense their environment through protein–protein interactions with non-TCS sensory proteins (for example CheA). Therefore, addition or loss of input/output domains would seem to merely require adoption of new interaction partners for the TCS transmitter and receiver domains in order to regain or initiate functionality—evolutionary changes which might potentially require mutation of just a small number of amino acid residues.

HK–RR interactions

The transmitter–receiver domain interaction (with subsequent phosphotransfer) is another protein–protein interaction that can occur either within or between proteins. It is of paramount importance as most prokaryotes possess multiple TCSs and the specificity of the transmitter–receiver interaction dictates the level of cross-communication occurring between systems, as discussed above. It has been proposed that most cross-communication will occur between TCSs that belong to the same phylogenetic group (Fabret et al. 1999), which is to be expected as members of a group have diverged from a common ancestor. However, in some notable cases communication occurs between proteins from different TCS families. For instance the CheA HKs typically interact with both CheY and CheB RRs, which are phylogenetically distinct groups.

Some signature amino acid residues are particularly conserved within a TCS family, and this observation has been used as the basis for TCS classification schemes (Fabret et al. 1999; Hutchings et al. 2004). It seems that these family specific residues are important for maintenance of the geometry of the transmitter–receiver interaction, and help to prevent unwanted cross-communication between TCSs of different families (Fabret et al. 1999), though they seem to contribute less in defining within-family interaction specificity. Early work focussing on the TCSs of *Bacillus subtilis* (Fabret et al. 1999) suggested that residues could be classified as catalytic (invariant), anchoring (conserved within a TCS family and preventing inter-family binding) or recognition residues (mediating within-family specificity), through analysis of multiple sequence alignments (Fabret et al. 1999; Hoch and Varughese 2001).

With the advent of genomics, the abundance of TCS gene sequences available has allowed the development of computational approaches that can accurately identify a subset of amino acid residues that are responsible for defining transmitter–receiver interaction specificity, as a consequence of covariation between HK and RR residues (Skerker et al. 2008; Weigt et al. 2009; Cock 2009). The validity of such approaches has been confirmed experimentally by Laub and co-workers (Skerker et al. 2008). In theory it is now possible to investigate the evolutionary pressures acting on that subset of residues which dictate the specificity of HK–RR interaction, in comparison to residues that have a role in maintaining protein structure, or other features of TCS function, such as effector binding, propagation of conformational change upon stimulus perception, flexibility of protein structure, or phosphoprotein stability (Thomas et al. 2008). Additionally, knowledge regarding the co-evolution of interfacial residues has the potential for the generation of predictive models that can suggest partnerships between TCS proteins based solely on their amino acid sequence (Burger and van Nimwegen 2008; Cock 2009). Such predictive tools will hopefully be able to investigate reconstructed ancestral genomes, potentially shedding light onto the evolution of system cross-communication/insulation.

Whilst these predictive algorithms are still under development, one current method has demonstrated that sequence-based predictions of transmitter–receiver binding specificity work less well for hybrid kinases than for two-gene TCSs (Cock 2009). In addition to highlighting the importance of considering all features of TCSs in computational analyses, such an observation suggests that interaction specificity is 'relaxed' within hybrid kinases compared to the between-protein interactions of two-gene TCSs. Presumably, the close proximity of transmitter and receiver domains within a hybrid kinase allows a reduction in interaction specificity/affinity, without compromising formation of a transmitter–receiver complex. Such an interpretation is supported by recent experimental work (Wegener-Feldbrügge and Søgaard-Andersen 2009) which demonstrates preferential within-protein phosphotransfer. It therefore seems that TCS protein fusion/fission has important consequences for subsequent evolution and for signal transduction. As fusion/fission events are major features of TCS evolution, leading to gain/loss of input/output domains and formation/fission of hybrid kinases, the subtleties and functional consequences of such events are potentially very interesting from an evolutionary viewpoint.

Summary and perspectives

The modular nature of the typical TCS has clearly provided a solid platform on which evolution can act, as demonstrated by the contemporary abundance and diversity of TCSs. The simplicity of early TCSs (inferred as the ancestral members of the current major families) means that most contemporary systems have evolved largely unchanged, however, at the global level there has been a general increase in TCS complexity during evolution. Such increases in complexity have allowed the evolution of TCSs which integrate multiple signals, and TCSs with sophisticated dynamic properties. An important challenge is now to look at such systems in detail, to define stages in their evolutionary origin, and to relate the selective pressures which acted on ancestral systems and the functional innovations which consequently evolved.

Recently, the evolutionary events underlying gross changes in TCS architecture have begun to be dissected, although the functional consequences for a TCS of such evolutionary changes are usually not clear. It therefore

remains a future goal to understand TCS evolution in terms of selective pressures and/or niche adaptation. In order to gain such an understanding it will be necessary to integrate functional, structural and ecological information. However, progress is being made in filling such gaps in our knowledge, and some of the broad 'rules' by which TCS evolution occurs are being defined. Hopefully, if we continue to explore the natural history of TCS evolution, and truly understand the rules (subtle as well as gross) by which TCSs evolve, then we can start to 'rationally evolve' TCS signalling pathways of our own devising, with potential applications limited only by our imagination.

References

Alm E, Huang K, Arkin A (2006) The evolution of two-component systems in bacteria reveals different strategies for niche adaptation. PLoS Comp Biol 2:e143

Appleby JL, Parkinson JS, Bourret RB (1996) Signal transduction via the multi-step phosphorelay: not necessarily a road less traveled. Cell 86:845–848

Armitage JP (1999) Bacterial tactic responses. Adv Microb Physiol 41:229–289

Ashby MK (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

Barakat M, Ortet P, Jourlin-Castelli C, Ansaldi M, Mejean V, Whitworth DE (2009) P2CS: a two-component system resource for prokaryotic signal transduction research (submitted)

Bijlsma JJE, Groisman E (2003) Making informed decisions: regulatory interactions between two-component systems. Trends Microbiol 11:359–366

Burger L, van Nimwegen E (2008) Accurate prediction of protein–protein interactions from sequence alignments using a Bayesian method. Mol Syst Biol 4:165

Chen Y-T, Chang HY, Lu CL, Peng HL (2004) Evolutionary analysis of the two-component systems in *Pseudomonas aeruginosa* PAO1. J Mol Evol 59:725–737

Cock PJA (2009) Two-component regulation: modelling, predicting and identifying protein–protein interactions and assessing signalling networks of bacteria. Ph.D. Thesis, University of Warwick, UK

Cock PJA, Whitworth DE (2007a) Evolution of prokaryotic two-component system signaling pathways: gene fusions and fissions. Mol Biol Evol 24:2355–2357

Cock PJA, Whitworth DE (2007b) Evolution of gene overlaps: relative reading frame bias in prokaryotic two-component system genes. J Mol Evol 64:457–462

Cock PJA, Whitworth DE (2009) Unpublished data

D'Souza M, Glass EM, Syed MH, Zhang Y, Rodriguez A, Maltsev N, Galperin MY (2007) Sentra: a database of signal transduction proteins for comparative genome analysis. Nucleic Acids Res 35:D271–D273

Fabret C, Feher VA, Hoch JA (1999) Two-component signal transduction in *Bacillus subtilis*: how one organism sees its world. J Bacteriol 181:1975–1983

Galperin MY (2005) A census of membrane-bound and intracellular signal transduction proteins in bacteria: bacterial IQ, extroverts and introverts. BMC Microbiol 5:35

Galperin MY (2006) Structural classification of bacterial response regulators: diversity of output domains and domain combinations. J Bacteriol 188:4169–4182

Gao R, Mack TR, Stock AM (2007) Bacterial response regulators: versatile regulatory strategies from common domains. Trends Biochem Sci 32:225–234

Grebe TW, Stock JB (1999) The histidine protein kinase superfamily. Adv Microb Physiol 41:139–227

Higgs PI, Cho K, Evans LS, Whitworth DE, Zusman DR (2005) Four unusual two-component signal transduction homologs, RedC–RedF, are necessary for timely development in *Myxococcus xanthus*. J Bacteriol 187:8191–8195

Hinchliffe SJ, Howard SL, Huang YH, Clarke DJ, Wren BW (2008) The importance of the Rcs phosphorelay in the survival and pathogenesis of the enteropathogenic yersiniae. Microbiology 154:1117–1131

Hoch JA, Varughese KI (2001) Keeping signals straight in phosphorelay signal transduction. J Bacteriol 183:4941–4949

Hutchings MI, Hoskisson PA, Chandra G, Buttner MJ (2004) Sensing and responding to diverse extracellular signals? Analysis of the histidine kinases and response regulators of *Streptomyces coelicolor* A3(2). Microbiology 150:2795–2806

Kim D-J, Forst S (2001) Genomic analysis of the histidine kinase family in bacteria and archaea. Microbiology 147:1197–1212

Koretke KK, Lupas AN, Warren PV, Rosenberg M, Brown JR (2000) Evolution of two-component signal transduction. Mol Biol Evol 17:1956–1970

Lawrence J (1999) Selfish operons: the evolutionary impact of gene clustering in prokaryotes and eukaryotes. Curr Opin Genet Dev 9:642–648

Martiny AC, Coleman ML, Chisholm SW (2006) Phosphate acquisition genes in *Prochlorococcus* ecotypes: evidence for genome-wide adaptation. Proc Natl Acad Sci USA 103:12552–12557

Michel B (1999) Illegitimate recombination in bacteria. In: Charlebois RL (ed) Organization of the prokaryotic genome. ASM Press, Washington, pp 129–150

Mizuno T (1997) Compilation of all genes encoding two-component phosphotransfer signal transducers in the genome of *Escherichia coli*. DNA Res 4:161–168

Mizuno T, Kaneko T, Tabata S (1996) Compilation of all genes encoding bacterial two-component signal transducers in the genome of the cyanobacterium, *Synechocystis* sp. strain PCC 6803. DNA Res 3:407–414

Moraleda-Muñoz A, Carrero-Lérida J, Pérez J, Muñoz-Dorado J (2003) Role of two novel two-component regulatory systems in development and phosphatase expression in *Myxococcus xanthus*. J Bacteriol 185:1376–1383

Pao GM, Saier MH Jr (1995) Response regulators of bacterial signal transduction systems: selective domain shuffling during evolution. J Mol Evol 40:136–154

Qian W, Han Z-J, He C (2008) Two-component signal transduction systems of *Xanthomonas* spp.: a lesson from genomics. Mol Plant–Microbe Interact 21:151–161

Robinson VL, Buckler DR, Stock AM (2000) A tale of two components: a novel kinase and a regulatory switch. Nat Struct Biol 7:626–633

Skerker JM, Prasol MS, Perchuk BS, Biondi EG, Laub MT (2005) Two-component signal transduction pathways regulating growth and cell cycle progression in a bacterium: a system-level analysis. PLoS Biol 3:e334

Skerker JM, Perchuk BS, Siryaporn A et al (2008) Rewiring the specificity of two-component signal transduction systems. Cell 133:1043–1054

Stephenson K, Hoch JA (2002) Evolution of signalling in the sporulation phosphorelay. Mol Microbiol 46:297–304

Stephenson K, Hoch JA (2004) Developing inhibitors to selectively target two-component and phosphorelay signal transduction systems of pathogenic microorganisms. Curr Med Chem 11:765–773

Stephenson K, Lewis RJ (2005) Molecular insights into the initiation of sporulation in Gram-positive bacteria: new technologies for an old phenomenon. FEMS Microbiol Rev 29:281–301

Stock JB, Ninfa AJ, Stock AM (1989) Protein phosphorylation and regulation of adaptive responses in bacteria. Microbiol Rev 53:450–490

Thomas SA, Brewster JA, Bourret RB (2008) Two variable active site residues modulate response regulator phosphoryl group stability. Mol Microbiol 69:453–465

Tong ZZ, Zhou DS, Song YJ et al (2005) Genetic variations in the *pgm* locus among natural isolates of *Yersinia pestis*. J Gen Appl Microbiol 51:11–19

Ulrich LE, Zhulin IB (2007) MiST: a microbial signal transduction database. Nucleic Acids Res 35:D386–D390

Ulrich LE, Koonin EV, Zhulin IB (2005) One-component systems dominate signal transduction in prokaryotes. Trends Microbiol 31:52–56

Watanabe T, Okada A, Gotoh Y, Utsumi R (2008) Inhibitors targeting two-component signal transduction. Adv Exp Med Biol 631:229–236

Wegener-Feldbrügge S, Søgaard-Andersen L (2009) The atypical hybrid histidine protein kinase RodK in *Myxococcus xanthus*: spatial proximity supersedes kinetic preference in phosphotransfer reactions. J Bacteriol (in press) doi:10.1128/JB.01405-08

Weigt M, White RA, Szurmant H, Hoch JA, Hwa T (2009) Identification of direct residue contacts in protein–protein interaction by message passing. Proc Natl Acad Sci USA 106:67–72

Whitworth DE (2008) Genomes and knowledge—a questionable relationship? Trends Microbiol 16:512–519

Whitworth DE, Cock PJA (2008a) Two-component systems of the myxobacteria: structure, diversity and evolutionary relationships. Microbiology 154:360–372

Whitworth DE, Cock PJA (2008b) Myxobacterial two-component systems. In: Whitworth DE (ed) Myxobacteria: multicellularity and differentiation. ASM Press, Washington, pp 169–189

Whitworth DE, Holmes AB, Irvine AG, Hodgson DA, Scanlan DJ (2008a) P-acquisition components of the *Myxococcus xanthus* Pho regulon are regulated by both P-availability and development. J Bacteriol 190:1997–2003

Whitworth DE, Millard A, Hodgson DA, Hawkins PF (2008b) Protein–protein interactions between two-component system transmitter and receiver domains of *Myxococcus xanthus*. Proteomics 8:1839–1842

Zhang W, Shi L (2005) Distribution and evolution of multiple-step phosphorelay in prokaryotes: lateral domain recruitment involved in the formation of hybrid-type histidine kinases. Microbiology 151:2159–2173

Zusman DR, Scott AE, Yang Z, Kirby JR (2007) Chemosensory pathways, motility and development in *Myxococcus xanthus*. Nat Rev Microbiol 5:862–872