Published in final edited form as:

Curr Opin Microbiol. 2010 April ; 13(2): 190–197. doi:10.1016/j.mib.2010.01.007.

Interaction Fidelity in Two-Component Signaling

Hendrik Szurmant* and James A. Hoch*

Department of Molecular and Experimental Medicine, The Scripps Research Institute, 10550 N Torrey Pines Road, La Jolla, CA 92037

Summary

Two component signal transduction systems and phosphorelays have been adapted and amplified by bacteria to respond to a multitude of environmental, metabolic and cell cycle signals while maintaining essentially identical structures for the domains responsible for recognition and phosphotransfer between the sensor histidine kinase and the response regulator. Co-crystal structures of these domains have revealed the variable residues at the interaction surface of the two components responsible for interaction specificity in signal transfer. This information has formed the basis for the development and validation of statistical methods to identify interaction residues and surfaces from compiled databases of interacting proteins and holds forth the promise of determining structures of multi-protein complexes and signaling networks.

Introduction

Genome sequencing has revealed that a multitude of different adaptation processes in the Bacteria are regulated by distinct sets of two signaling proteins with conserved sequence characteristics. This class of signal transduction system was originally termed the two-component signaling system (TCS) although it is now clear that many TCS have additional regulatory partners [1]. The two individual proteins that form the core of a TCS serve both signal detection and response generation roles. Signal detection is achieved via the sensor histidine kinase (SK), a modular multi domain protein, which adjusts its autokinase activity in response to a signaling molecule [2]. The output of the system is generated by the response regulator (RR) protein, which most commonly serves as a DNA-binding transcription factor [3]. The two proteins communicate via a transphosphorylation reaction by passing a phosphoryl group from a histidine on the SK to an aspartate residue on the phosphoacceptor domain (also referred to as receiver domain) of the RR protein. This reaction requires the formation of a precise but transient complex between the individual phosphorylatable domain of the SK [the HisKA (also referred to as DHP) domain] and the phosphoacceptor domain of the RR [4**].

By the year 1996, three high resolution structures of RR phosphoacceptor domains had been solved, namely those of *Salmonella enterica* serovar *Typhimurium* chemotaxis RR CheY, *Bacillus subtilis* RR Spo0F and the *Escherichia coli* RR NarL [5–7]. These three phosphoacceptor domain structures revealed a remarkably similar fold featuring a central parallel five-stranded β -sheet surrounded by five α -helices. The structural similarity of phosphoacceptor domains in diverse RR suggested that all such domains are very similar in

*Corresponding Authors, H.S.: Tel: 858-784-7945 Fax: 858-784-7966 Szurmant@scripps.edu, J.A.H: Tel: 858-784-7905 Fax: 858-784-7966 hoch@scripps.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

structure. This has since been confirmed by the availability of over 75 structures of distinct RR proteins that have been deposited in the protein database (PDB).

Early observations derived from a few model organisms suggested that the TCS is a common and highly amplified signaling system in the Bacteria. Since the onset of the genomic revolution, more than 1000 bacterial genomes have been sequenced [8]. It is now clear that TCS are almost ubiquitous in the Bacteria and are also found in the Archaea and some Eukaryotes. According to the microbial signal transduction database MiST2, 1087 sequenced bacterial genomes encode for 63259 TCS proteins and hence the average genome features roughly 29 individual signaling systems [9]. Most histidine kinase proteins span the membrane, potentially making the TCS a dominant signal transduction system for the detection of extra-cellular stimuli [10]. A typical TCS signaling network of the Gram-positive model organism *B. subtilis* is shown in Figure 1.

The observed amplification of TCS to regulate diverse processes while retaining strong conservation of structural folds raises an important question: How is signaling fidelity between the SK and RR proteins achieved to assure that an input signal generates only the appropriate response? A few instances of in vivo crosstalk between TCS systems in unmutated bacterial strains have been reported (e.g. [11,12] and for a comprehensive review, [13]) but most systems appear monogamously mated [14,15]. It is clear that fidelity has to be mainly achieved on the primary sequence level because variations in secondary and tertiary structure are minimal.

This review emphasizes advancements in our understanding of the molecular determinants that confer fidelity in TCS signaling made between 2007 and 2009, and also recounts some of the early structural and mutagenesis studies, which not only laid the foundation for the more recent studies but have stood the test of time and are important pieces in unraveling the puzzle of interaction specificity in two-component signaling proteins. Experimental approaches, such as site directed mutagenesis and determination of crystal and NMR structures of the involved proteins have formed the basis for the development of computational approaches to infer interaction residues from primary sequence alone. This technology has applications beyond the problem of TCS signaling fidelity and should become amenable to understand and predict protein interactions on a more global level.

Structures of core catalytic domains of TCS are identical

Because signaling fidelity is in large part a question of variations in the protein interaction surfaces, structural studies of the individual proteins as well as the complexes of RR and SK have contributed greatly to our understanding of signaling fidelity [4th,16]. As previously mentioned, the fidelity question arose as a consequence of the observed conservation of the structural fold of the RR phosphoacceptor domain. These domains have been very amenable to X-ray crystallographic studies and the ever-increasing number of available structures in the PDB has underscored the notion that all of these domains are highly similar in structure.

The SK proved to be a more challenging target for structural studies. The first high resolution X-ray structure of the two domain catalytic portion of a SK, HK853 from *Thermatoga maritima*, did not become available until 2005, revealing an N-terminal dimeric four-helix bundle HisKA domain and a C-terminal ATP-binding domain [17]. This structure was very similar to the *B. subtilis* Spo0B structure, which had been solved in 1998 [18]. Spo0B serves a phosphotransfer function in the sporulation phosphorelay of the bacilli [19]. The phosphorelay is an extended version of the prototypical TCS, in which phosphoryl groups from SK proteins are shuttled to the terminal RR protein via two intermediary proteins, a second RR (Spo0F in the sporulation phospho-relay) and a second His containing protein [20]. Most phosphorelays feature a phosphotransferase protein formed by a monomeric four-helix bundle Hpt domain, which is structurally distinct from the SK HisKA domain but is found also in the

atypical chemotaxis kinase CheA [21]. In contrast, the Spo0B protein is clearly an evolutionary divergent SK protein featuring both the dimeric four-helix bundle HisKA domain and a degenerate ATP-binding domain [4^{**},22]. As such, the protein has lost its ability to autophosphorylate but retained the ability to form a phosphotransfer complex with the phosphoacceptor domain of its partners Spo0F and Spo0A. The sporulation phosphorelay can thus be envisioned as two individual TCS arranged in tandem.

Structures of SK/RR complexes reveal contact residues at the interaction surface

While the individual structures of the SK and RR proteins revealed those residues that are surface exposed, the identification of the contact residues at the interaction surface required the structure of a SK/RR complex. Given the difficulty of crystallizing SK proteins, it comes as no surprise that a high-resolution structure of this inherently transient complex eluded the signal transduction field for many years. In contrast, the Spo0B/Spo0F phosphotransfer complex was determined in 2000 and has since served as a model for SK/RR mating [4^{**}]. The determination of the SK HK853 structure in 2005 [17] revealed that Spo0B was in fact a close structural relative of the SK and thus gave strong evidence that the Spo0B/Spo0F complex is indeed an excellent structural example of the SK/RR interaction, as was suggested in the original study [4^{**},23].

Two structures of SK/RR complexes from *T. maritima*, (HK853/RR468 and the ThkA/TrrA) were published in 2009 [24^{*},25]. These can now be compared with the Spo0B/Spo0F structure (Fig. 2). All three available structures conserve an identical overall orientation between SK and RR. The β 4- α 4-loop and the β 5- α 5-loop regions of the RR phosphoacceptor domain are nestled at the interface formed by helices α 1 and α 2' of the dimeric four-helix bundle HisKA domain. Multiple contacts are observed along the α 1-helices of both, SK and RR, which show a parallel arrangement. The new SK/RR structures reveal some additional contacts between RR helix α 1 and SK helix α 2. These contacts are not realized in the Spo0B/Spo0F structure because of a different orientation of helix α 2 with respect to helix α 1 in the Spo0B four-helix bundle domain. Remarkably, the orientation of helix α 1 to helix α 2' in the dimer, however remains unchanged in all three structural complexes, suggesting that the contacts made by the RR proteins with that interface are very important for function. In summary, these structures support the idea that the observed interaction mode is generic to all SK/RR protein pairs. Furthermore, it is clear that a subset of residues in the above-described contact elements must be a main contributor to SK/RR fidelity.

Computational approaches to infer signaling fidelity

Whereas the structures of the SK/RR complexes were able to reveal contact surface residues, a sequence analysis was necessary to distinguish those contact residues that are conserved across different SK/RR pairs from those that are variable and hence could contribute to interaction specificity and fidelity. Guided by the Spo0B/Spo0F structure, a small-scale analysis based on the limited sequencing information available at the time revealed that variable residue positions likely to confer interaction specificity are located to the RR helix α 1, as well as the β 4- α 4- and β 5- α 5-loop regions [23]. In the past two years more comprehensive mathematical approaches to identify specificity conferring residue positions have been published relying on sequence rather than structural information [26^{*}–28^{**}]. Nevertheless, the existing structures proved crucial in validating the results of these computational approaches.

Co-variance based computational approaches to identify interacting residues

A simple, yet powerful computational approach to extract specificity determining residue positions from protein sequence databases is the co-variance based approach, which has been applied extensively to individual proteins to gain insights into tertiary structure [29–33]. When applied to the TCS, the basic premise of such an approach is that specificity determining residues positions between SK and RR need to co-evolve and hence are correlated, i.e. a residue choice at a specificity determining position within the SK dictates what residues are permissible at the interacting RR position. To infer statistically significant correlations requires a large dataset of functional SK/RR pairs. This requirement can be easily met utilizing the fact that a large fraction of putatively functional SK/RR pairs may be identified when their genes are found chromosomally adjacent, organized in operons. Correlation of residue positions can be measured as mutual information (MI) [34]. Applied to interacting proteins, MI compares the frequency of occurrence for individual aminoacids at two residue positions in the multiple sequence alignment to the frequency of the co-occurrence of these aminoacids at the particular residue positions (Fig. 3a).

When such a co-variance based approach was first applied to TCS protein databases by White and colleagues¹, it revealed residue positions in the α 1-helix of the RR phosphoacceptor domain that were highly correlated with residue positions in the α 1 and α 2 helices of the SK HisKA domain [26*]. Highly correlated residue positions were also identified in the β 4- α 4-loop/ α 4-helix region of the phosphoacceptor domain that co-varied with two buried residues at the base of the HisKA four-helix bundle (Fig. 3). Two later reports applying co-variance based methods to identify SK/RR specificity identified roughly the same set of residue pairings [27*,36*].

Experimental evidence exists to support the notion that the mathematically inferred residue pairings contribute to SK/RR fidelity. The inferred pairings connecting RR phosphoacceptor helix α 1 with SK HisKA helices α 1 and α 2 are found in contact at the interaction surface in the three SK/RR co-crystal structures discussed above [4**,24*,25]. A systematic alanine scanning mutagenesis of surface residues of the RR Spo0F had previously revealed the importance of these contact pairings for SK/RR interaction [16]. Finally, Skerker et al. were successful in switching SK specificity from its partner RR to other RR proteins by substituting the SK α 1 and α 2 contact residues [36*]. In this study mutagenesis efforts were guided by a co-variance analysis analogous to the earlier one by White et al. [26*] that identified the same set of residue pairings.

Whereas the correlations for the contact residue pairings are easily explained due to the availability of the complexed structures, the observed correlations between the RR β 4- α 4-loop/ α 4-helix region with two buried SK residues could not be explained as direct interaction contacts. Skerker et al. suggested that the origin of the observed co-variance might be due to phylogenetic noise and applied a phylogenetic correction algorithm by Dunn et al. [37]. This analysis proved ineffective in eliminating co-variance between these residue pairings, likely because they are not of phylogenetic origin. This is consistent with results from previous experimental approaches that identified substitutions in these RR residues with altered in vivo SK/RR specificity [38,39], suggesting that the co-variance analysis captured correlations of

¹The described approach is distinct from the earlier phylogenetic approach by Li et al. [35], which compares sequence conservation of residues across orthologs to those in all SK or RR proteins separately, with the assumption that residue positions conserved only in orthologs contribute to interaction specificity. Li et al. use a co-variance measure to determine statistical relevance of ortholog specific conservation. As such the approach investigates SK and RR separately and does not reveal any connectivity between SK and RR proteins without previous structural knowledge.

importance for signaling fidelity. An NMR study on mutants of the RR Spo0F with altered kinase preference was undertaken and revealed that substitutions of these residues strongly affect the overall β 4- α 4-loop and α 4-helix orientation with respect to the remainder of the RR fold [40]. Consistent with the co-crystal structures and mutagenesis on the RR Spo0F is the hypothesis that the β 4- α 4-loop serves an important function in sealing off the phospho-transfer active site from water access [16]. A perfect fit of the β 4- α 4-loop with the SK four-helix bundle is necessary to perform this task explaining the correlation between the involved residues and their role in signaling fidelity.

From sequence to contact residues: direct coupling analysis

It is now well established from TCS studies that co-variance analyses identify those residues of interacting pairs of proteins essential for direct interaction as well as indirect interactions required for function. These indirect interactions impede the application of co-variance to other protein pairs, because in the absence of TCS co-crystal data it would be impossible to infer the interaction structure using all of the co-variance information. If co-variance analysis were to only identify direct interaction contact residues and not indirect, such information could be utilized to assemble structures of protein complexes that are not easily amenable to X-ray crystallography.

Weigt et al. improved the co-variance based approach by addition of a statistical inference step, aimed at disentangling direct correlations from those that arise by correlation chains, i.e. via joined correlations with additional residue positions in the protein sequences [28^{**},41]. As such, this approach induces a measure called ‘Direct Information’, which is the fraction of the MI that is due to a direct correlation between two residue positions. Applied to TCS, this two-step approach referred to as direct coupling analysis (DCA) performs beautifully in eliminating indirect correlations. Unlike the purely co-variance based approach for which only 35% of highly correlated residue pairings are found at the protein-protein interface, DCA identifies as its top 10 hits residue pairings in contact at the protein interface [28^{**}]. DCA proved equally successful in identifying the RR-RR homodimerization surface, demonstrating that this type of analysis is not limited to the interaction of SK/RR and might in fact be applicable to study a wide range of protein complexes [28^{**}].

Whether these sequence-derived contacts are sufficient to assemble the high-resolution structure of a protein complex has been answered by Schug et al. [42^{*},43]. Utilizing DCA inferred interactions along with the individual protein structures as input in molecular dynamics simulations the authors were able to generate a Spo0B/Spo0F complex that overlaid with the X-ray diffraction structure to a root mean square deviation of 2.5Å. Furthermore a structural prediction of the HK853/RR468 complex was made blindly, which turned out to be highly comparable to the concurrently published crystal structure [24^{*},42^{*}]. Indeed the structural prediction captured some significant structural changes evident in the HK853 crystal structure upon RR468 docking, demonstrating the power of such an approach.

Conclusions

Structural, computational and mutagenesis approaches have contributed to our understanding of the signaling fidelity problem in the highly amplified yet structurally conserved TCS. These advances now await application towards a number of demanding challenges (box 1). Future applications of this knowledge will aim at predicting the signaling networks of complex microbial organisms towards a system level understanding. This is a particularly challenging task for organisms such as *Myxococcus xanthus*, which features more than 200 TCS proteins, with many of these proteins encoded as orphans, i.e. not in an operon with its partner [44]. Defining the signaling network of this and equally complex organisms experimentally is

challenging. To date it is not clear how many of these orphan SK and RR proteins are part of branched pathways similar to the sporulation or chemotaxis pathways, an aspect that requires further investigation. A computational approach to signaling fidelity will allow predictions as to who interacts with whom, based on primary sequence alone.

Box 1

Future Directions

Structures of the sensor kinase/response regulator complex along with computations and mutagenesis experiments have identified the residue positions that confer specificity to the interaction, to avoid unwanted cross-talk between systems.

The major challenge awaiting the signal transduction community will be to utilize this information to generate algorithms that predict the two-component signaling networks of entire organisms, including identification of branched pathways and extended phosphorelays.

Such information will be useful in isolated cases, to identify individual interaction partners. More importantly, such efforts will be a step towards a systems level understanding of the signaling events that regulate bacterial behavior and adaptation.

Some of the computational tools developed to study two-component signaling fidelity show promise for broad applicability to the study of transient protein complexes. These tools will need to be further developed and broad applicability needs to be demonstrated.

First attempts to generate partner predictions for SK/RR pairings have been made [26*,27*]. In the approach by White et al. the chromosomally adjacent SK/RR pairs are used as a training set to infer a co-variance based scoring function for the prediction of SK-RR interaction partners. The approach by Burger et al. applies a Bayesian network method to infer interaction partners from multiple sequence alignments. Burger et al. observed that predictions for the orphan SK and RR proteins were much less reliable than those made for paired proteins [27*]. This suggests that these approaches are not yet able to capture all parameters that contribute to interaction specificity or alternatively that orphan proteins are much more promiscuous than chromosomally paired proteins. There is hope that the recent structures of the SK/RR complex will generate additional clues on how to improve scoring functions towards complete and accurate signaling network predictions. For completeness such approaches should eventually be expanded to cover all TCS and phosphorelay proteins including the structurally distinct Hpt-type phosphotransferases. Exemplary structures of the Hpt/RR complex exist for the yeast YPD1/SLN1 phosphorelay complex [45,46*], but computational specificity analysis, similar to the one for HisKA-type SK proteins has not yet been performed.

Other applications utilizing the knowledge on signaling fidelity are found in the field of synthetic biology. In particular the experiments by Skerker et al. demonstrate the feasibility of designing de novo bacterial signaling circuits by rewiring the specificity of TCS [36*].

Perhaps most intriguingly, computational studies on the TCS have given a glimpse into the future by demonstrating the type of statistically significant data that can be extracted from large protein sequence databases derived from genomic sequences. With the availability of more than a thousand bacterial genomes, computational analysis such as the DCA approach described above should become amenable to study interactions between bacterial proteins that are not as highly amplified as the TCS proteins [28**].

Acknowledgments

This work was supported in part by Grant GM19416 from the Institute of General Medical Sciences, and Grant AI055860 from the National Institute of Allergy and Infectious Diseases, National Institutes of Health.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Hoch, JA.; Silhavy, TJ. Two-component signal transduction. Washington, D.C: ASM Press; 1995.
2. Szurmant H, White RA, Hoch JA. Sensor complexes regulating two-component signal transduction. *Curr Opin Struct Biol* 2007;17:706–715. [PubMed: 17913492]
3. Galperin MY. Structural classification of bacterial response regulators: diversity of output domains and domain combinations. *J Bacteriol* 2006;188:4169–4182. [PubMed: 16740923]
- 4••. Zapf J, Sen U, Madhusudan Hoch JA, Varughese KI. A transient interaction between two phosphorelay proteins trapped in a crystal lattice reveals the mechanism of molecular recognition and phosphotransfer in signal transduction. *Structure* 2000;8:851–862. This manuscript presents the first high-resolution insight into the SK/RR interaction specificity in the form of the Spo0B/Spo0F co-crystal structure. Although Spo0B is a phosphotransfer protein in the sporulation phosphorelay rather than a SK, the structure demonstrates that it shares fold and function with the SK. As such, this manuscript proved to be the major milestone in the understanding of signaling fidelity by revealing the contact interface between SK and RR. [PubMed: 10997904]
5. Stock AM, Mottonen JM, Stock JB, Schutt CE. Three-dimensional structure of CheY, the response regulator of bacterial chemotaxis. *Nature* 1989;337:745–749. [PubMed: 2645526]
6. Baikalov I, Schroder I, Kaczor-Grzeskowiak M, Grzeskowiak K, Gunsalus RP, Dickerson RE. Structure of the *Escherichia coli* response regulator NarL. *Biochemistry* 1996;35:11053–11061. [PubMed: 8780507]
7. Madhusudan, Zapf J, Whiteley JM, Hoch JA, Xuong NH, Varughese KI. Crystal structure of a phosphatase-resistant mutant of sporulation response regulator Spo0F from *Bacillus subtilis*. *Structure* 1996;4:679–690. [PubMed: 8805550]
8. Liolios K, Mavromatis K, Tavernarakis N, Kyrpides NC. The Genomes On Line Database (GOLD) in 2007: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 2008;36:D475–479. [PubMed: 17981842]
9. Ulrich LE, Zhulin IB. The MiST2 database: a comprehensive genomics resource on microbial signal transduction. *Nucleic Acids Res*. 2009
10. Galperin MY. A census of membrane-bound and intracellular signal transduction proteins in bacteria: bacterial IQ, extroverts and introverts. *BMC Microbiol* 2005;5:35. [PubMed: 15955239]
11. Rietkotter E, Hoyer D, Mascher T. Bacitracin sensing in *Bacillus subtilis*. *Mol Microbiol* 2008;68:768–785. [PubMed: 18394148]
12. Howell A, Dubrac S, Noone D, Varughese KI, Devine K. Interactions between the YycFG and PhoPR two-component systems in *Bacillus subtilis*: the PhoR kinase phosphorylates the non-cognate YycF response regulator upon phosphate limitation. *Mol Microbiol* 2006;59:1199–1215. [PubMed: 16430694]
13. Laub MT, Goulian M. Specificity in two-component signal transduction pathways. *Annu Rev Genet* 2007;41:121–145. [PubMed: 18076326]
14. Yamamoto K, Hirao K, Oshima T, Aiba H, Utsumi R, Ishihama A. Functional characterization in vitro of all two-component signal transduction systems from *Escherichia coli*. *J Biol Chem* 2005;280:1448–1456. [PubMed: 15522865]
15. Skerker JM, Prasol MS, Perchuk BS, Biondi EG, Laub MT. Two-component signal transduction pathways regulating growth and cell cycle progression in a bacterium: a system-level analysis. *PLoS Biol* 2005;3:e334. [PubMed: 16176121]

16. Tzeng YL, Hoch JA. Molecular recognition in signal transduction: the interaction surfaces of the Spo0F response regulator with its cognate phosphorelay proteins revealed by alanine scanning mutagenesis. *J Mol Biol* 1997;272:200–212. [PubMed: 9299348]
17. Marina A, Waldburger CD, Hendrickson WA. Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein. *Embo J* 2005;24:4247–4259. [PubMed: 16319927]
18. Varughese KI, Madhusudan, Zhou XZ, Whiteley JM, Hoch JA. Formation of a novel four-helix bundle and molecular recognition sites by dimerization of a response regulator phosphotransferase. *Mol Cell* 1998;2:485–493. [PubMed: 9809070]
19. Burbulys D, Trach KA, Hoch JA. Initiation of sporulation in *Bacillus subtilis* is controlled by a multicomponent phosphorelay. *Cell* 1991;64:545–552. [PubMed: 1846779]
20. Hoch JA. Two-component and phosphorelay signal transduction. *Curr Opin Microbiol* 2000;3:165–170. [PubMed: 10745001]
21. Gao R, Stock AM. Biological insights from structures of two-component proteins. *Annu Rev Microbiol* 2009;63:133–154. [PubMed: 19575571]
22. Varughese KI, Tsigelny I, Zhao H. The crystal structure of beryll fluoride Spo0F in complex with the phosphotransferase Spo0B represents a phosphotransfer pretransition state. *J Bacteriol* 2006;188:4970–4977. [PubMed: 16788205]
23. Hoch JA, Varughese KI. Keeping signals straight in phosphorelay signal transduction. *J Bacteriol* 2001;183:4941–4949. [PubMed: 11489844]
24. Casino P, Rubio V, Marina A. Structural insight into partner specificity and phosphoryl transfer in two-component signal transduction. *Cell* 2009;139:325–336. The first structure of the prototypical SK HK853 in complex with its RR RR468 is presented. The structure reveals significant similarity with the Spo0B/Spo0F complex and contributes to our understanding of interaction fidelity in TCS signaling by revealing a few additional contact pairings. [PubMed: 19800110]
25. Yamada S, Sugimoto H, Kobayashi M, Ohno A, Nakamura H, Shiro Y. Structure of PAS-linked histidine kinase and the response regulator complex. *Structure* 2009;17:1333–1344. [PubMed: 19836334]
26. White RA, Szurmant H, Hoch JA, Hwa T. Features of protein-protein interactions in two-component signaling deduced from genomic libraries. *Methods Enzymol* 2007;422:75–101. This study is the first to apply co-variance analysis between SK and RR proteins to gain insights into the SK/RR contact residue pairs and introduces a scoring function to predict SK/RR interaction partners. The residue pairings predicted to confer interaction specificity have been experimentally validated in other studies. [PubMed: 17628135]
27. Burger L, van Nimwegen E. Accurate prediction of protein-protein interactions from sequence alignments using a Bayesian method. *Mol Syst Biol* 2008;4:165. In this manuscript a Bayesian network approach aimed at identifying protein interaction partners is presented and applied to two-component systems. The approach is mathematically distinct from the co-variance analysis by White et al. Interestingly both methods extract largely the same specificity conferring residue positions. [PubMed: 18277381]
28. Weigt M, White RA, Szurmant H, Hoch JA, Hwa T. Identification of direct residue contacts in protein-protein interaction by message passing. *Proc Natl Acad Sci U S A* 2009;106:67–72. In this study the co-variance approach introduced by White et al. is amended by a statistical inference step, which disentangles direct and indirect contributions to the observed co-variation of residue pairings. The new method referred to as direct coupling analysis strongly improves the predictive power of co-variance in identifying interaction surface contact residues and might find broad applicability to study elusive protein complexes. [PubMed: 19116270]
29. Göbel U, Sander C, Schneider R, Valencia A. Correlated mutations and residue contacts in proteins. *Proteins* 1994;18:309–317. [PubMed: 8208723]
30. Atchley WR, Wollenberg KR, Fitch WM, Terhalle W, Dress AW. Correlations among amino acid sites in bHLH protein domains: an information theoretic analysis. *Mol Biol Evol* 2000;17:164–178. [PubMed: 10666716]
31. Altschuh D, Lesk AM, Bloomer AC, Klug A. Correlation of co-ordinated amino acid substitutions with function in viruses related to tobacco mosaic virus. *J Mol Biol* 1987;193:693–707. [PubMed: 3612789]

32. Lockless SW, Ranganathan R. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* 1999;286:295–299. [PubMed: 10514373]
33. Suel GM, Lockless SW, Wall MA, Ranganathan R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat Struct Biol* 2003;10:59–69. [PubMed: 12483203]
34. Shannon CE. A mathematical theory of communication. *Bell Syst Tech J* 1948;27:379–423.
35. Li L, Shakhnovich EI, Mirny LA. Amino acids determining enzyme-substrate specificity in prokaryotic and eukaryotic protein kinases. *Proc Natl Acad Sci U S A* 2003;100:4463–4468. [PubMed: 12679523]
36. Skerker JM, Perchuk BS, Siryaporn A, Lubin EA, Ashenberg O, Goulian M, Laub MT. Rewiring the specificity of two-component signal transduction systems. *Cell* 2008;133:1043–1054. A covariance approach analogous to the earlier one by White et al. is presented. The authors utilize this information to rewire several SK proteins by site-directed mutagenesis to phosphorylate non-native RR targets. [PubMed: 18555780]
37. Dunn SD, Wahl LM, Gloor GB. Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction. *Bioinformatics* 2008;24:333–340. [PubMed: 18057019]
38. Haldimann A, Prahalad MK, Fisher SL, Kim SK, Walsh CT, Wanner BL. Altered recognition mutants of the response regulator PhoB: a new genetic strategy for studying protein-protein interactions. *Proc Natl Acad Sci U S A* 1996;93:14361–14366. [PubMed: 8962056]
39. Jiang M, Tzeng YL, Feher VA, Perego M, Hoch JA. Alanine mutants of the Spo0F response regulator modifying specificity for sensor kinases in sporulation initiation. *Mol Microbiol* 1999;33:389–395. [PubMed: 10411754]
40. Szurmant H, Bobay BG, White RA, Sullivan DM, Thompson RJ, Hwa T, Hoch JA, Cavanagh J. Co-evolving motions at protein-protein interfaces of two-component signaling systems identified by covariance analysis. *Biochemistry* 2008;47:7782–7784. [PubMed: 18588317]
41. Lunt B, Szurmant H, Procaccini A, Hoch JA, Hwa T, Weigt M. Inference of direct residue contacts in two-component signaling. *Methods Enzymol* 2010;471:17–41.
42. Schug A, Weigt M, Onuchic JN, Hwa T, Szurmant H. High resolution protein complexes from integrating genomic information with molecular simulation. *Proc Natl Acad Sci U S A* 2009;106:22124–22129. This study presents a molecular dynamics approach to assemble high-resolution structures of protein complexes that utilizes individual proteins structures and protein-protein contacts derived from the DCA approach by Weigt et al. [PubMed: 20018738]
43. Schug A, Weigt M, Hoch JA, Onuchic JN, Hwa T, Szurmant H. Computational modeling of phosphotransfer complexes in two-component signaling. *Methods Enzymol* 2010;471:43–58.
44. Shi X, Wegener-Feldbrugge S, Huntley S, Hamann N, Hedderich R, Sogaard-Andersen L. Bioinformatics and experimental analysis of proteins of two-component systems in *Myxococcus xanthus*. *J Bacteriol* 2008;190:613–624. [PubMed: 17993514]
45. Zhao X, Copeland DM, Soares AS, West AH. Crystal structure of a complex between the phosphorelay protein YPD1 and the response regulator domain of SLN1 bound to a phosphoryl analog. *J Mol Biol* 2008;375:1141–1151. [PubMed: 18076904]
46. Xu Q, Porter SW, West AH. The yeast YPD1/SLN1 complex: insights into molecular recognition in two-component signaling systems. *Structure* 2003;11:1569–1581. Here the first high-resolution complex between a phosphorelay Hpt-type phosphotransferase and a RR is presented. The structure serves as the only representative model of the Hpt/RR complex, which is distinct from the HisKA-type SK/RR complex. [PubMed: 14656441]

Figure 1.

The *Bacillus subtilis* set of two component signaling pathways. The Gram-positive model organism *B. subtilis* reveals a set of 29 TCS signaling pathways, about the average for bacterial genomes. The modular SK proteins feature variable sets of signal detection domains (in blue) and most are transmembrane spanning proteins. Most RR proteins feature a DNA-binding transcription factor domain (green). The SK is defined by a structurally conserved autokinase catalytic core, which features a dimeric four-helix bundle HisKA (also referred to as DHp) domain (red rectangles), subject to phosphorylation by the ATP-binding (also referred to as CA) domain (red diamonds). The RR proteins are defined by a structurally conserved phosphoacceptor domain (red circle), which is subject to phosphorylation by the HisKA domain of the SK. Despite the conservation of the structural elements almost all pathways appear monogamously mated. A special case is the sporulation phosphorelay, in which five SK converge to phosphorylate a single RR Spo0F, which transfers its phosphoryl group to a second SK like protein Spo0B. The complex of Spo0B and Spo0F was the first structurally resolved example of the SK/RR interaction and has long served as a model to explain SK/RR fidelity. For clarity the chemotaxis pathway, which feature a structurally distinct kinase CheA and three RR proteins, CheY, CheB and CheV are excluded from the figure.

Figure 2.

Structures of the SK/RR complex. (a) Top view and (b) side view of the three available representative structures of the SK/RR complex demonstrate identical structural orientation of RR phosphoacceptor domain (blue) and SK HisKA domain (red). The orientation of the ATPase domain differs in each case (green). The structures are from left to right the Spo0B/Spo0F complex from *B. subtilis* (PDBID 1f51), the HK853/RR468 complex (PDBID 3dge) and the ThkA/TrrA complex (PDBID 3a0r), both from *T. maritima*. (c) When overlaying the RR proteins Spo0F (blue) and RR468 (beige) the four-helix bundles of Spo0B (red) and HK853 (yellow) also overlay well, demonstrating the conservation of the structural organization of the complex between RR phosphoacceptor and SK HisKA domains. Significant contacts in the individual structures are made between the individual secondary structure elements as indicated and described in the text.

Figure 3.

Co-variance analysis identifies specificity determining residues for the SK/RR interaction. (a) From multiple sequence alignments of all chromosomally adjacent SK/RR pairs the frequency $f_i(A_i)$ and $f_j(A_j)$ of individual amino acid choice A (which can be any of the 20 proteinogenic aminoacids or a gap) can be calculated at positions i in the SK HisKA domain and positions j in the RR phosphoacceptor (PA) domain. Mutual information (MI) calculation compares these individual frequencies to the co-frequency of amino acid choice $f_{i,j}(A_i, A_j)$ for two residue positions. If two positions are unlinked, the co-frequency is equal to the product of the individual frequencies. This results in an $MI_{i,j}$ close to 0. If two positions are linked, $f_{i,j}(A_i, A_j)$ is larger than the product of the individual frequencies and $MI_{i,j}$ is larger than 0. (b) Shown is the network of SK/RR position pairings with above background MI values (green and red). Some of the high MI values are due to direct correlations between two residue positions (red), whereas others show MI values that are inflated by correlation chains, due to the residue choice of neighboring residues (green) and are hence indirect. These different correlations can be disentangled when amending co-variance analysis by a statistical inference step [28]. Residue numbers are as in HK853 for the SK and as in RR468 for the RR for easy identification of residue pairings in the HK853/RR468 structure, but in contrast to the numbering used in [28]. (c) Mapped to the newly available HK853/RR468 co-crystal structure (PDBID 3dge) it is obvious that the directly correlated residue position pairings (red) are at the interaction surface, whereas the indirectly correlated ones are not. Both, indirect and direct correlations have been shown to contribute to fidelity of the SK/RR phosphotransfer reaction that occurs between the conserved His and Asp residues (blue).