Co-Evolving Motions at Protein−Protein Interfaces of Two-Component Signaling Systems Identified by Covariance Analysis[†]

Hendrik Szurmant,[‡] Benjamin G. Bobay,[§] Robert A. White,[‡,||] Daniel M. Sullivan,[§] Richele J. Thompson,[§] Terence Hwa,[||] James A. Hoch,[‡] and John Cavanagh*[,§]

Division of Cellular Biology, Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, California 92037, Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, North Carolina 27695, and Centre for Theoretical Biological Physics, University of California at San Diego, La Jolla, California 92037

Received May 23, 2008; Revised Manuscript Received June 17, 2008

ABSTRACT: Short-lived protein interactions determine signal transduction specificity among genetically amplified, structurally identical two-component signaling systems. Interacting protein pairs evolve recognition precision by varying residues at specific positions in the interaction surface consistent with constraints of charge, size, and chemical properties. Such positions can be detected by covariance analyses of two-component protein databases. Here, covariance is shown to identify a cluster of co-evolving dynamic residues in two-component proteins. NMR dynamics and structural studies of both wild-type and mutant proteins in this cluster suggest that motions serve to precisely arrange the site of phosphoryl transfer within the complex.

Transient protein−protein interactions are the basis for signal propagation in signal transduction and for assembly of metastable protein complexes. Residues at interaction surfaces determine the recognition specificity required to correctly route signals. An important example of this is found in bacterial two-component signal transduction systems. As a consequence of gene amplification in the same cell, the unique function of each two-component system requires that evolution maintain strict monogamous interactions between protein partners. Moreover, in two-component systems, the phosphoryl transfer between a His residue on one protein and an Asp residue on its mate requires precise positioning of the two residues. This necessitates that the partner proteins interact in the same way regardless of the signaling pathway. These constraints confine the evolutionary choices of amino acid changes in both partners at any position in the interaction surface. Limited choices are revealed as high-covariance positions between partners in statistical analyses of large two-component system databases (*1, 2*).

The prototypical two-component system for these studies originates from the phosphorelay signal transduction pathway controlling *Bacillus subtilis* sporulation (*3*). The Spo0F response regulator (RR) is typical of the "receiver domain" of two-component systems that accepts a phosphoryl group from a histidine kinase domain (HisKA) of a specific sensor kinase (SK). The structures of a variety of RR receiver domains, including Spo0F, have been determined and the folds found to be essentially identical (*4*). Spo0F interacts with a sporulation specific SK and with the Spo0B phosphotransferase. The HisKA domain of histidine kinases interacts with RR receiver domains. HisKA domains are structurally similar to the prototypical Spo0B, consisting of a four-helix bundle formed by two associating helical hairpins in the dimer of the kinase or in Spo0B (*2, 5*).

A cocrystal structure of the Spo0B−Spo0F complex was determined and is currently the only exemplary structural representative of SK−RR recognition (*2*). It showed that helix α1 of the RR makes significant contacts with the helical region just C-terminal to the phosphorylatable His in the HisKA domain. These and other studies reinforce the conclusion that specific protein−protein interactions govern phosphotransfer in two-component signaling (*6*). The high level of amplification of these systems in the bacterial genomes along with common adjacent organization into operons of interacting SK and RR genes means that, through the wealth of genomic sequencing data, several thousand functional pairs can be immediately recognized without further experimental validation (*3, 7*). The basic premise underlying covariance methods is that specific positions in interacting surfaces of protein partners have to co-evolve to allow for continued interaction.

We recently developed and applied a covariance method based on mutual information calculations to a database of more than 2000 two-component protein pairs (*1*). The method measures the mutual information between every residue *i* within the SK multiple-sequence alignment (MSA) and every residue *j* in the RR MSA, resulting in a three-dimensional matrix of 7200 SK−RR residue combinations. The mutual information for each residue pair was expressed as a score $S_{i,j}$. The determination of significantly large $S_{i,j}$ values is achieved through the analysis of the total distribution of all possible position pairings between the interacting domains, which shows an overrepresentation of $S_{i,j}$ values of >0.26, thereby providing a natural cutoff. This resulted in a so-called "best friend network" of interactions [see the Supporting Information (*1*)].

The best friend network of interactions contains six independent clusters, which were mapped on the individual exemplary structures of the HisKA domain of sensor kinase

[†] This work is supported by NIH Grants GM055769 (J.C.), GM019416 (J.A.H.), AI055860 (J.A.H.) and GM077298 (T.H.) and by NSF Grants PHY-0216576 (T.H.), PHY-0225630 (T.H.), and DBI-0532925 (R.A.W.).

[‡] The Scripps Research Institute.
[§] North Carolina State University.
[||] University of California at San Diego.

10.1021/bi8009604 CCC: $40.75
Published on Web 06/28/2008

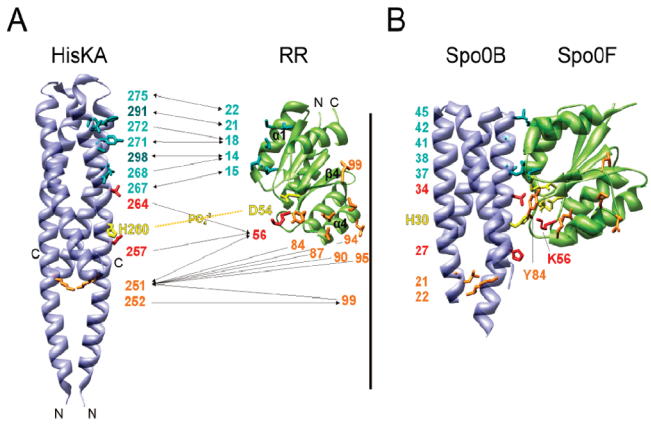

FIGURE 1: Two-component signaling best friend network. (A) Residues with significant covariance scores (i.e., $S_{i,j} > 0.26$) are mapped onto exemplary structures of a HisKA domain and a RR (see the text). Clusters 1−5 are colored cyan; cluster 6 is colored red or orange, and the phosphotransfer residues H260 and D54 are colored yellow. (B) Best friend network mapped onto a HisKA−RR cocrystal structure (Spo0B−Spo0F; see the text). Residues involved in the phosphotransfer reaction and metal ion orientation are colored yellow.

HK853 of *Thermotoga maritima* (*5*) and onto Spo0F of *B. subtilis* (*4*) (Figure 1A).

The first five clusters include HK853 residues T267, A268, A271, Y272, T275, F291, and Q298 and Spo0F residues G14, I15, L18, E21, and V22. These clusters define an interaction between helix α1 of the HisKA and helix α1 of the RR. This is the exact mode seen in the Spo0B−Spo0F cocrystal structure (Figure 1B) (*2, 8*). Thus, covariance clusters 1−5 are the positions interacting in the cocrystal that were predicted to be responsible for specificity (*6*). These residues covary to maintain the spatial configuration of the active site.

The residues identified by covariance in cluster 6 are more difficult to explain on the basis of direct interaction. Cluster 6 includes K251, T252, N257, and T264 from HK853 and K56, Y84, L87, I90, K94, E95, and L99 from Spo0F. A subset of this cluster, HK853 residues N257 and T264 and Spo0F residue K56, is located adjacent to the His for Asp phosphotransfer residues. Covariance between these latter residues is reasonable on the basis of the Spo0B−Spo0F structure. The explanation for the covariance between the remainder of cluster 6 is not immediately apparent from the static Spo0B−Spo0F structure. HK853 residues K251 and T252 are in the core of the HisKA four-helix bundle and are only partially solvent-exposed. Y84, L87, I90, K94, and E95 from Spo0F are in the β4−α4 loop/helix α4 region. It is impossible for all covarying residues to form a direct interaction without one or both of the proteins undergoing significant conformational change.

What causes cluster 6 covariance? The easiest interpretation is that cluster 6 is a false positive phylogenetic artifact with no functional constraint (*9*). However, experimental evidence strongly suggests an alternate explanation. The β4−α4 loop/helix α4 region of PhoB is known to be involved in kinase specificity (*10*). Furthermore, alanine scanning mutagenesis of Spo0F identified several cluster 6 residues (*11*). K56A, Y84A, and L87A Spo0F mutations caused a decrease in sporulation frequency in vivo, reflecting a reduction in the rate of phosphotransfer from the KinA

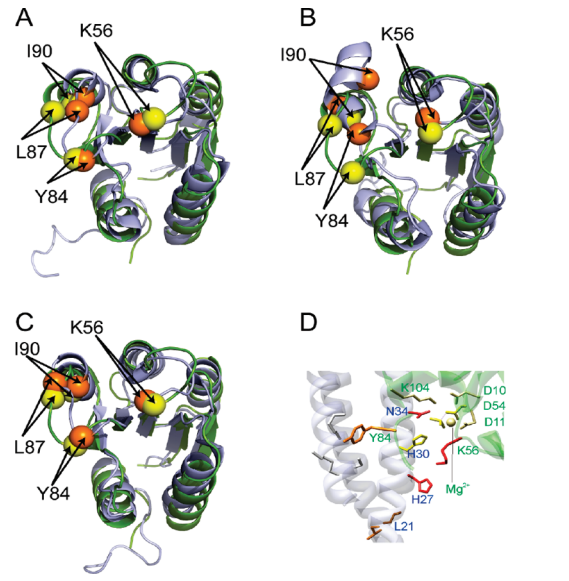

FIGURE 2: Lowest-energy NMR structures of Spo0F mutants and their role in active site conformation. Superpositions of wild-type Spo0F (light blue) with (A) L66A, (B) I90A, and (C) H101A mutations (light green). Residues K56, Y84, L87, and I90 (marked by their Cα atoms) were picked up in the covariance analysis as being important for SK interaction specificity: Spo0F in yellow and mutant proteins in orange. (D) Active site of Spo0B−Spo0F phosphotransfer shown in close-up.

kinase observed in vitro (*11*). The I90A mutation—along with two other mutations not identified by covariance analysis, H101A and L66A—caused an increase in sporulation frequency (*12*). This is a direct result of a relaxed kinase specificity, allowing at least two other kinases to induce sporulation via Spo0F phosphorylation (*12*).

With the importance of cluster 6 residues established, it appears that the covariance method captured specificity parameters not apparent from the static Spo0B−Spo0F structure. However, these parameters were found in the mutagenesis studies. To improve our understanding of covariance in cluster 6, we turn to NMR studies on Spo0F and some of its mutants. Previous NMR structural and dynamics data showed that the β4−α4 loop/helix α4 region of Spo0F, which contains cluster 6 residues, undergoes significant, functional motions (*13*). Unphosphorylated Spo0F adopts two distinct subfamilies of structures in the β4−α4 loop/helix α4 region, one of which is similar to that of phosphorylated Spo0F. Phosphorylation of Spo0F shifts conformation in this region (Figure S1). This demonstrates that Spo0F, in its inactive state, samples the conformation that it will adopt when it becomes activated. This is only possible because the β4−α4 loop/helix α4 region possesses predetermined functional dynamic character. To supplement these studies, we have determined the NMR structures of the three Spo0F mutants (L66A, H101A, and I90A) with significantly relaxed sensor kinase specificity (*12*) (Figure 2).

Comparing mutant and wild-type structures (Figure 2A−C), we see that the β4−α4 loop/helix α4 region undergoes the largest perturbation. Also, for cluster 6 residues in the Spo0F mutants, $^{1}H-^{15}N$ NMR line widths are on average ~20% broader than their wild-type counterparts (Figure S2). For Spo0F residues K56 and K94 in mutants L66A and I90A, and residue L87 in H101A, resonances broaden and disap-

pear. These data indicate enhanced flexibility. It appears that altered kinase specificity of the Spo0F mutants correlates with changes in conformation and motion of the RR residues that make up covariance cluster 6.

Overall, there is considerable evidence that the $\beta4-\alpha4$ loop/helix $\alpha4$ region of Spo0F is functionally flexible. It is also proposed that flexibility of this region affects function in other response regulators (*14, 15*). Of the covarying RR residues within cluster 6, two are special, residues K56 and Y84. As mentioned, residue K56 is in the proximity of the site of phosphotransfer whereas residue Y84 has a proposed crucial role in sealing off the active site from solvent access (Figure 2D). It has also been suggested that Y84 stabilizes the HisKA−RR interaction (*6*). Triangulating the position of residues Y84 and K56 with conserved residue D11, responsible for metal−phosphate coordination, reveals a significant change in the NMR structure of the three Spo0F mutants with respect to the wild type. In wild-type Spo0F, this angle is 20.3°, while it is 27.6°, 29.4°, and 29.3° for the L66A, H101A, and I90A mutants, respectively. Evidently, the positioning of the $\beta4-\alpha4$ loop/helix $\alpha4$ region, and specifically residue Y84, affects the formation of an efficient phosphotransfer active site.

The two mostly buried residues of the four-helix bundle, positions K251 and T252 in the SK, show covariance with the RR $\beta4-\alpha4$ loop/helix $\alpha4$ region. Their role in this process remains unexplained. There are two possibilities. Since these core SK positions exhibit some exposed surface area, it is possible that they can directly contact the RR residues in the $\beta4-\alpha4$ loop/helix $\alpha4$ region. However, since the exposed surface area is limited, we currently favor a model in which the nature of these residues dictates the overall arrangement of the individual helices within the four-helix bundle (i.e., via charge repulsion or hydrophobic interactions). The basis of the covariance of cluster 6 would then be to ensure perfect positioning of the active site residues in both the RR and the SK to allow for efficient phosphotransfer. In addition, the SK core residues K251 and T252 are ideally located to serve as a potential amplifier of conformational changes dictated by signal input domains upon ligand binding. Such conformational changes would then alter not only the rate of autophosphorylation but also likely the efficiency of phosphotransfer to the paired RR.

Our covariance approach readily detects spatial coupling across the interaction HisKA−RR interface. The majority of the coupling clusters were rationalized without difficulty since they connected residues known to be proximal from earlier X-ray studies. However, the basis of the coupling of a significant subset of residues from another cluster could not be interpreted with merely static structural information. Examining both the structural and dynamical character of the RR, Spo0F, along with several of its mutants with altered kinase specificity, enabled us to determine that the modified covariance approach recognizes statistical coupling for residues undergoing functional conformational exchange. These residues, though not in suitable orientations in the X-ray complex, can reorient themselves to play a critical role in dictating the correct arrangement of the phosphotransfer active site (Figure 2D). There is a surfeit of evidence to suggest that the $\beta4-$-$\alpha4$ loop/helix $\alpha4$ region in RRs is generally dynamic (*13, 14*). Select residues in this area may occupy functional positions at the SK−RR interface. Even a small population of such conformations at the interface could influence specificity. Indeed, this is an appealing mechanism for fine-tuning specificity. Diverse dynamic signatures among RRs in this region may assist in precise targeting. RRs also use different surface subsets for regulatory interactions, and while their phosphorylation-activated switch domains are architecturally similar, there is versatility in the way these domains regulate activity (*15*).

We propose that covariance clusters 1−5 represent a "landing strip" that performs a major role by docking the SK and its paired RR together. Covariance cluster 6 provides additional, subtle, contributions to this specificity and is required for an efficient phosphotransfer reaction. This study demonstrates that a covariance method can identify regions with functionally important dynamics that contribute to intermolecular surfaces. It has long been recognized that pairs of interacting proteins demand precise steric positioning of amino acids across the interface and that the contributing proteins co-evolve with that requirement in mind. This work provides strong evidence to support the idea that interacting proteins also co-evolve dynamic characteristics across interfaces to perform their biological roles.

SUPPORTING INFORMATION AVAILABLE

Experimental material and methods. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. White, R. A., Szurmant, H., Hoch, J. A., and Hwa, T. (2007) *Methods Enzymol. 422*, 75–101.
2. Zapf, J., Sen, U., Madhusudan, Hoch, J. A., and Varughese, K. I. (2000) *Structure 8*, 851–862.
3. Hoch, J. A. (2000) *Curr. Opin. Microbiol. 3*, 165–170.
4. Feher, V. A., Zapf, J. W., Hoch, J. A., Whiteley, J. M., McIntosh, L. P., Rance, M., Skelton, N. J., Dahlquist, F. W., and Cavanagh, J. (1997) *Biochemistry 36*, 10015–10025.
5. Marina, A., Waldburger, C. D., and Hendrickson, W. A. (2005) *EMBO J. 24*, 4247–4259.
6. Hoch, J. A., and Varughese, K. I. (2001) *J. Bacteriol. 183*, 4941–4949.
7. Galperin, M. Y. (2006) *J. Bacteriol. 188*, 4169–4182.
8. Varughese, K. I., Tsigelny, I., and Zhao, H. (2006) *J. Bacteriol. 188*, 4970–4977.
9. Skerker, J. M., Perchuk, B. S., Siryaporn, A., Lubin, E. A., Ashenberg, O., Goulian, M., and Laub, M. T. (2008) *Cell 133*, 1043–1054.
10. Haldimann, A., Prahalad, M. K., Fisher, S. L., Kim, S. K., Walsh, C. T., and Wanner, B. L. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 14361–14366.
11. Tzeng, Y. L., and Hoch, J. A. (1997) *J. Mol. Biol. 272*, 200–212.
12. Jiang, M., Tzeng, Y. L., Feher, V. A., Perego, M., and Hoch, J. A. (1999) *Mol. Microbiol. 33*, 389–395.
13. Feher, V. A., and Cavanagh, J. (1999) *Nature 400*, 289–293.
14. Gardino, A. K., and Kern, D. (2007) *Methods Enzymol. 423*, 149–165.
15. West, A. H., and Stock, A. M. (2001) *Trends Biochem. Sci. 26*, 369–376.

BI8009604