Hypothesis

Sub-classification of response regulators using the surface characteristics of their receiver domains

Douglas J. Kojetin, Richele J. Thompson, John Cavanagh*

Department of Molecular and Structural Biochemistry, North Carolina State University, 128 Polk Hall, Campus Box 7622, Raleigh, NC 27695, USA

Received 15 July 2003; revised 11 August 2003; accepted 12 August 2003

First published online 20 October 2003

Edited by Thomas L. James

Abstract The omnipresent bacterial switch known as a two-component system is comprised of a response regulator and a sensor kinase with which it interacts. Sensor kinases have been classified and further sub-classified into groups based on their sequence similarity, loop lengths and domain organization. Response regulators have been classified predominantly by the identity and function of their output domains. Here, comparative based homology modeling of the receiver domains of the OmpR sub-family of response regulators in *Bacillus subtilis* and *Escherichia coli* suggests further sub-classification is possible. A color-coded scale is used to show trends in surface hydrophobicity. For the OmpR receiver domains modeled these trends allow further sub-classification. The specific surface regions used for this sub-classification procedure correlate with clusters of residues that are important for interaction with cognate four helix bundle HisKA/Hpt domains.

Key words: OmpR sub-families; Homology modeling

1. Introduction

Bacteria are remarkably successful in surviving under disparate ecological conditions. They respond to continually fluctuating surroundings by efficiently sensing the environment and translating this information into a calculated cellular response. This process involves the ubiquitous communication element referred to as a *two-component system* [1–3]. Two-component systems consist of a sensor kinase and a response regulator. In response to a specific environmental signal, the sensor kinase first autophosphorylates in an ATP-dependent reaction on a conserved histidine residue. A subsequent step finds the phosphoryl group then being transferred into the conserved aspartic acid binding pocket of a response regulator protein. The meticulous nature of the recognition process between the response regulator and its kinase cannot be overstated. An imprecise interaction may result in the dissemination of an incorrect message and such flawed recognition may compromise the cells ability to adapt and survive.

However, such crucial specificity is not trivial to accomplish. The more well-characterized bacteria such as *Bacillus subtilis* and *Escherichia coli* possess more than 40 different two-component systems, each one responding to a unique signal [1,4–6]. More recently, the draft genome of the bacterium *Nostoc punctiforme* revealed the presence of well over 100 two-component systems [7]. The complications of recognition afforded by this abundance of systems are further amplified by the similarities across both the response regulator and sensor kinase families. In nearly all cases response regulators are two-domain proteins with an output domain and a receiver domain. The carboxy-terminal output domain is most commonly a transcriptional regulator possessing a helix-turn-helix type DNA binding motif. The activity of each unique output domain is carefully controlled by the receiver domain's level of phosphorylation.

Each receiver domain consists of approximately 120 amino acids and there is a high level of sequence similarity [1,6]. It is not unexpected, therefore, that they adopt a conserved three-dimensional structure. This structure consists of five α-helices assembled around a central β-sheet comprised of five parallel β-strands. In all cases the conserved aspartic acid phosphorylation pocket is located at the 'top' of the protein and is surrounded, and somewhat buried, by five β-α loops. Other critical and invariant amino acids are apparent in this region, including a threonine/serine from the β4-α4 loop and a lysine from the β5-α5 loop. Both of these residues are instrumental in relaying information of the phosphorylation state of the binding pocket to the rest of the protein. With the structural characteristics of the receiver domains being preserved, it is no surprise that the kinase domains with which they interact also share structural similarities [8,9]. The response regulator interaction domain in the sensor kinase consists of a four helix bundle motif referred to as either a HisKA or an Hpt domain. Current nomenclature has the HisKA four helix bundle formed from a dimeric assembly, while the Hpt four helix bundle is formed from a single, continuous polypeptide chain.

With all this similarity in mind, several studies have endeavored to categorize response regulators and sensor kinases into a variety of classes and sub-classes [8–12]. Such undertakings are not restricted to two-component systems of course, and there have been many studies performed to classify and sub-classify protein families [13–15]. The majority of these studies have relied upon predicted and actual sequential information (primary, secondary and tertiary), domain identity, or biological function. Overall these efforts have attempted to provide a global genomic linking of proteins and their respective biochemical relevance.

*Corresponding author. Fax: (1)-919-515 2047.
E-mail address: john_cavanagh@ncsu.edu (J. Cavanagh).

0014-5793/03/$22.00
doi:10.1016/S0014-5793(03)01167-0

A broad classification of sensor kinase–response regulator pairs has been made, to some degree, by looking at genomic location and cellular function [8,9]. In some cases such linkages have been experimentally verified using biological assays. More detailed classification of sensor kinases has relied upon domain organization characteristics and sequence similarity. In addition, sub-families of sensor kinases have been classified based upon features such as homology around the conserved phospho-histidine residue and the lengths of intra-/extracellular loops and domains. Response regulators have been classified into sub-families using sequence similarity and function of the output domain [8,10–12]. However, to this point, they have been only *broadly* classified using sequence similarity and alignment of the receiver domain. Using merely the sequence similarity of the receiver domain, no accurate sets of sub-families had been proposed. We considered that it may be possible to sub-classify response regulators based on the surface characteristics of the receiver domains alone. As far as we are aware, response regulators have not been previously sub-classified in this way.

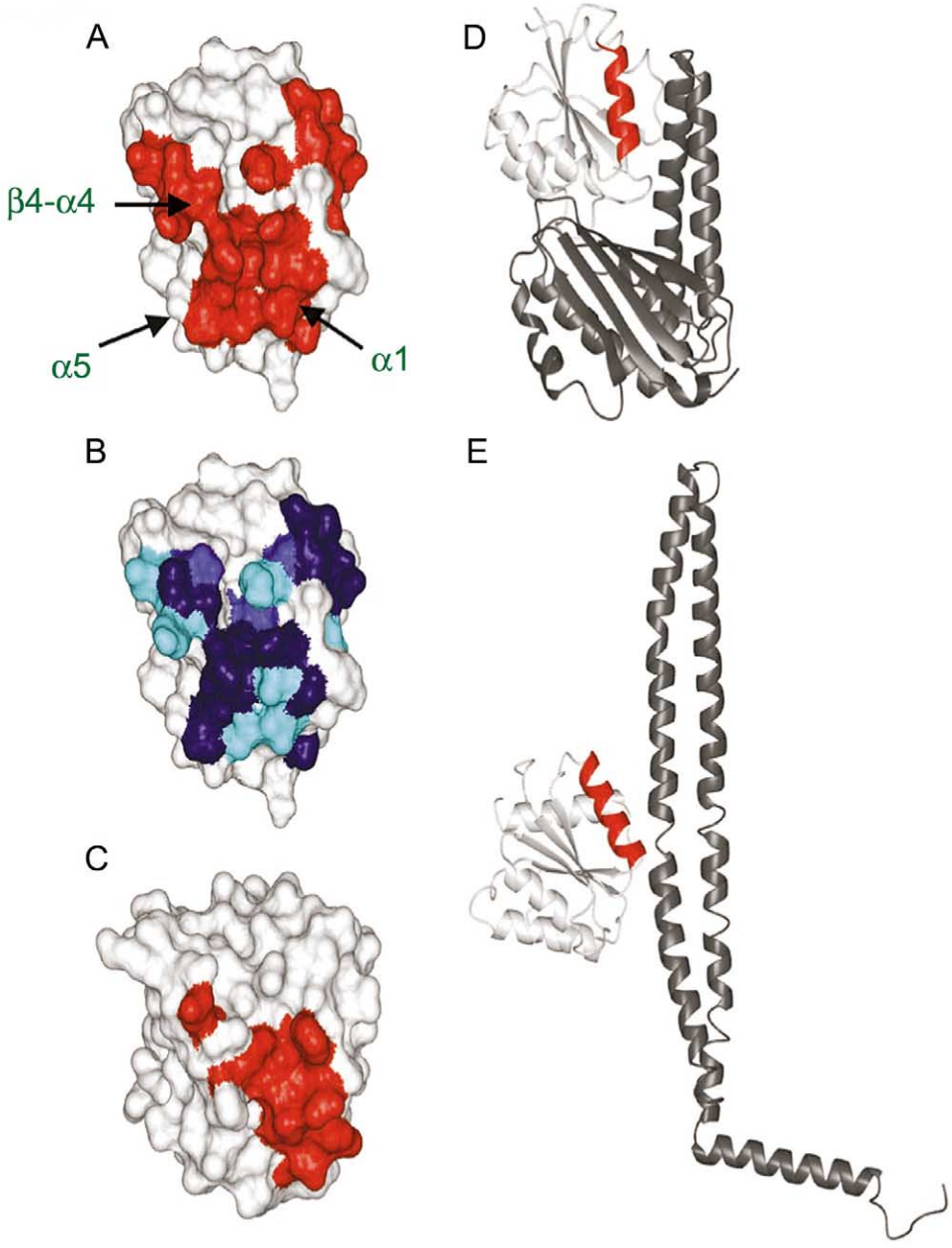

Fig. 1. Structural and biochemical data from *E. coli* CheY and *B. subtilis* Spo0F. A: Contact surface plot showing regions of Spo0F that contact Spo0B in the co-crystal structure. Residues of Spo0F that contact the Spo0B are highlighted in red. B: Contact surface plot revealing regions of Spo0F that, when mutated to alanine, give rise to reduced interaction kinetics with residues affecting KinA interaction in blue, Spo0B in dark blue, or both in light blue. C: Contact surface plot revealing regions of CheY that contact CheZ in the co-crystal structure. Residues of CheY that contact CheZ are highlighted in red. D: Co-crystal structure of Spo0F:Spo0B. The ribbon diagram of Spo0F is colored white, α1-helix of Spo0F is highlighted in red and Spo0B is colored in dark gray. E: Co-crystal structure of CheY:CheZ. CheY is colored white, α1-helix of CheY is highlighted in red and CheZ is colored dark gray. Contact surface plots are viewed from directly above the aspartic acid phosphorylation pocket.

A variety of structural, mutagenesis and modeling studies show that receiver domains primarily use the surface comprising the N-terminus of α1-helix, the α1-helix/α5-helix interface and, to a lesser degree, the β4-α4 loop, for interactions with the HisKA/Hpt domain targets [16–21]. Examples of such interactions are shown in Fig. 1. With this in mind, we theorized that the surface characteristics of these regions could be used as a means for identifying functional surface trends and thereby determining possible sub-classes of response regulators. This is in contrast to the current broad classification of response regulators via their receiver domains which is based on primary sequence alignments, looking at overall sequence similarity [8]. Our hypothesis, therefore, is that sub-classification of response regulators may be possible based solely on the surface characteristics of the N-terminus of α1-helix, the α1-helix/α5-helix interface and, to a less significant extent, the β4-α4 loop, of their receiver domains.

In order to test this hypothesis, we have employed comparative based homology modeling to develop sufficient structural models of the receiver domains of response regulators in the OmpR sub-family from *B. subtilis* and *E. coli*. Analysis of trends in surface characteristics between these proteins in the regions described above has allowed us to sub-classify these proteins into more related groups. Furthermore, such receiver domain-based sub-classification of response regulators will allow questions concerning the nature of sensor kinase–response regulator recognition and discrimination to be more readily addressed.

2. Modeled receiver domains of the OmpR sub-family of response regulators

In order to visualize surface characteristics of the receiver domains in the OmpR sub-family of response regulators from *B. subtilis* [4] and *E. coli* [5], structural models were developed for the sequences listed in Table 1 using MODELLER [22,23]. Sequences were modeled to the four most homologous proteins with solved structures, determined using PSI-BLAST [24], with default MODELLER parameters and the refine variable 'refine_1'. Fifty models per target were generated, and final models were analyzed and chosen through analysis using PROCHECK-NMR [25], ERRAT [26], VERYIFY3D [27], C_α RMSD to the most homologous template and visual inspection. Models were visualized using the graphics program MOLMOL [28]. So as not to bias the modeling process with respect to proteins with no known structure, structures solved of proteins within this study were not used as templates for the development of the respective model (PhoB, PhoP). In addition, proteins with abnormal structural elements compared to 'conventional', monomeric response regulator domains were not used (Spo0A, 1DZ3).

Superimposition of models generated indicates that the backbones of the models are extremely similar in all areas (see Fig. 2). The one minor exception is the positioning of the α4-helix, which displays structural deviation in previously solved structures [29]. This difference in helix positioning makes it difficult to obtain a 'correct' orientation in the mod-

Table 1
Response regulator proteins (RR), histidine protein kinases (HPK) and sub-classes in the 'OmpR' two-component system sub-family from *B. subtilis* and *E. coli*

RR	Sequence	Accession number	RR class[a]	RR sub-class[b]	HPK	HPK class[a]	HPK class[c]
B. subtilis							
CssR	1–120	O32192		B	CssS		3
PhoP	1–118	P13792	R A1	A	PhoR	HPK 1a	4
ResD	8–121	P35163	R A1	A	ResE	HPK 1a	4
SpaR	1–120	P33112	R A1	B	SpaK	HPK 3c	
YbdJ	1–116	O31432		C	YbdK		2
YcbL	1–112	P42244		C	YcbM		1
YccH	1–119	P70955		E	YccG		5
YclJ	1–118	P94413	R A1	C	YclK	HPK 1a	3
YkoG	1–122	O34903	R A1	B	YkoH	HPK 1a	3
YrkP	1–116	P54443		B	YrkQ	HPK 4	3
YtsA	1–119	O34951	R A1	D	YtsB	HPK 3i	2
YvcP	1–119	O06978	R A1	D	YvcQ	HPK 3i	2
YvrH	133–252	P94504	R A1	C	YvrG	HPK 1a	3
YxdJ	1–116	P42421	R A1	D	YxdK	HPK 3i	2
YycF	1–117	P37478	R A1	A	YycG	HPK 1a	4
E. coli							
ArcA	1–118	P03026	R A1	C	ArcB	HPK 1b, hybrid, H2 domain	
BaeR	12–125	P30846	R A1	B	BaeS	HPK 1a	
BasR	1–116	P30843	R A1	E	BasS	HPK 2a	
CpxR	1–116	P16244	R A1	E	CpxA	HPK 2b	
CreB	1–119	P08368	R A1	A	CreC	HPK 3c	
CusR	1–116	P77380		B	CusS		
KdpE	1–116	P21866	R A1	D	KdpD	HPK 1a	
OmpR	1–120	P03025	R A1	E	EnvZ	HPK 2b	
PcoR	1–117	Q47456	R A1	B	PcoS	HPK 2a	
PhoB	1–120	P08402	R A1	D	PhoR	HPK 1a	
PhoP	1–116	P23836	R A1	E	PhoQ	HPK 3a	
QseB	1–116	P52076	R A1	E	QseC	HPK 2a	
RstA	6–119	P52108	R A2	E	RstB	HPK 2b	
TorR	1–117	P38684	R A1	C	TorS	HPK 1b, hybrid, H2 domain	
YedW	1–115	P76340		A	YedV		

[a]Reference [8].
[b]This work.
[c]Reference [9].

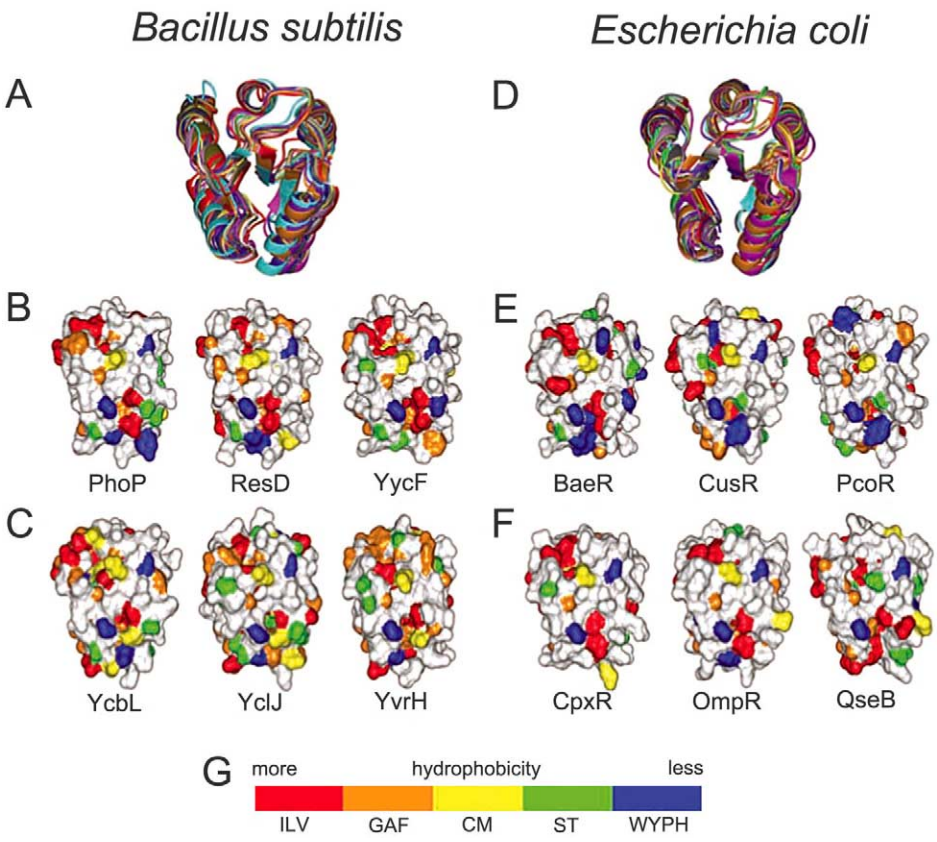

Fig. 2. Examples of sub-classifications of receiver domains in the 'OmpR' two-component system sub-family from *B. subtilis* and *E. coli*. These sub-classes were delineated using surface characteristics as discussed in the text. Shown for *B. subtilis* are (A) backbone superimposition of all models developed for receiver domains in *B. subtilis*, (B) three members of sub-class A, and (C) three members of sub-class C. Shown for *E. coli* are (D) backbone superimposition of all models developed for receiver domains in *E. coli*, (E) three members of sub-class E, and (F) two members of sub-class F. (G) Color-coded and numerically weighted hydrophobic scale (see text).

els generated. However, the emphasis of this work is on the more structurally conserved region consisting of the α1-helix, α1-helix/α5-helix interface, which shows a very high degree of similarity with respect to the surface characteristics of receiver domains with solved structures.

In order to detect subtle changes in surface side chain hydrophobicity, a color gradient scheme based on hydrophobic characteristics of the amino acid side chains was used to represent different levels of hydrophobicity rather than the traditional single color plots. Two available 'hydrophobic scales' were combined and statistically weighted to provide the color gradient scheme shown in Fig. 2 [30,31]. This methodology allows the relative hydrophobic strengths and composition to be compared and contrasted quickly when contrasting similar proteins. In the studies described here, it is precisely this variability in hydrophobic surface composition that is critical in the sub-classification of the models and in the protein target discrimination process.

3. Response regulator–HisKA/Hpt domain interactions: a method for classification

It is clear that many delicate contributions shape the precision of recognition between each response regulator and its cognate kinase. It has been noted that the interaction surface of a response regulator with its sensor kinase is a mosaic of amino acids and that the mosaic pattern is similar in all members of a related family of response regulators. Hoch and Varughese [32] have suggested that this surface montage consists of three general types of amino acids: (i) essential invariant catalytic residues directly involved in the phosphotransfer mechanism, (ii) anchor residues that establish broad orientational contacts for catalysis and (iii) recognition residues that ensure the correct two proteins come together. To ensure that unproductive interactions between a sensor kinase and the wrong response regulator are prevented, a means of discrimination must be discernible. This discrimination most likely comes in the form of subtle surface variability.

If this is the case, then it is plausible that a specific region on the receiver domain surface shows heightened variability across the family and that this surface may act as the primary contributor in HisKA/Hpt domain interactions. In addition, there may also be similar surface patterns of amino acids that have been evolutionarily conserved or modified. The results from our receiver domain modeling study show that the highest degree of both hydrophobic (and indeed electrostatic surface) variation is along the α1-helix and the α1-helix/α5-helix interface. The hydrophobic variability is particularly evident in this region. We suggest that the highly variable surface composition in this region allows for discrimination between different four helix bundles containing specific HisKA/Hpt domains and is the primary determinant in this recognition process. As we have commented, supporting studies show that approximately 75% of all receiver domain–HisKA/Hpt do-

main contacts so far determined come from the α1-helix and the α1/α5-helix interface regions (see Fig. 1).

4. Sub-classification of the receiver domains within the OmpR sub-family of response regulators

The general sub-classification procedure consists of several stages. Initially a surface square of ~225 Å^2 (15 Å × 15 Å) is centered about the α1/α5-helix interface. This area also encompasses the HisKA/Hpt domain recognition residues in α1-helix and α5-helix. This square is then divided into three vertical strips (~5 Å wide each). Within each strip the overall hydrophobic content is evaluated using a sliding scale for each amino acid type (see Fig. 2). Averaged hydrophobic scores are given to each strip. This represents the initial classification step. Further sub-classification is achieved by considering the characteristics of the individual amino acids and includes: side chain types, clustering of residues and relative amino acid positions. An example of this process is provided in the Supplementary Materials at http://www4.ncsu.edu/~jcavana/.

Response regulators have been previously sub-classified primarily on the basis of the output function and organization of the domain C-terminal to the receiver domain [8–12]. Analysis of the hydrophobic surface characteristics of models for receiver domains within the OmpR sub-family in *B. subtilis* and *E. coli*, in conjunction with information concerning sensor kinase discrimination and recognition, has allowed us to sub-classify the receiver domains into related groups. Table 1 shows this information along with previous classification studies [8,9]. This analysis reveals hydrophobic trends that are more readily identifiable when compared to primary sequence alignment methods. Accordingly, this has enabled us to take this sub-family of proteins and further divide it into more closely related sub-classes. Furthermore, whereas multiple sequence alignments examine overall sequence similarity, the comparative based modeling approach used in this study allows conserved surface 'patches' of residues, within which there is variability, to be easily studied. Importantly, these surface patches correlate to a region known to interact with the four helix bundle motif present in HisKA/Hpt domains [16–21].

The discrete sub-classes within this sub-family disclose extra detail into the surface and recognition relatedness of different receiver domains and may aid in identifying groups of response regulators that have potential for crosstalk between pathways. Sub-classes were generated by analyzing the hydrophobic content of the conserved patch of hydrophobic residues comprising the α1-helix, α1-helix/α5-helix interface, as well as in the β4-α4 loop region. Shown in Fig. 2 are very clear examples from the sub-classes of proteins discerned within each of the OmpR sub-families from *B. subtilis* and *E. coli*. It is evident from the surface hydrophobic views that sub-sets of proteins within the OmpR sub-family are clearly more related to some proteins than others. These sub-classes were arbitrarily assigned within the OmpR sub-family from *B. subtilis* and *E. coli*, respectively. Specifically, sub-class A in *B. subtilis* is not the 'same' as sub-class A in *E. coli* (see Table 1). We found through the process of our sub-classification studies that, although the OmpR sub-family in general is a highly homologous group of proteins across different species, the proteins are clearly more sub-related within the same species.

Using a matrix-based sequence alignment procedure, which aligns each respective protein sequence listed in Table 1 to all others listed, we were able to correlate our sub-classes, found using surface hydrophobic analysis, to groups of related proteins within the sequence alignment with moderate agreement (available in the Supplementary Materials at http://www4.ncsu.edu/~jcavana/). In the case of PhoP, ResD and YycF in *B. subtilis* (Fig. 2B), there was complete agreement between our surface method and the sequence alignment. However, in the case of YcbL, YclJ and YvrH in *B. subtilis* (Fig. 2C), and generally amongst all sequences analyzed, sequence alignment analysis revealed that these proteins were generally more related to the proteins in sub-class A. However, our surface hydrophobic analysis clearly shows that these proteins are more related with regard to α1-helix, α1-helix/α5-helix interface and the β4-α4 look region. This is an example of how using the comparative based modeling approach to study the degree of hydrophobic variation within 'conserved surface patches' can complement sequence alignment analysis in the understanding of protein relatedness.

Interestingly, there is a significant correlation between the receiver domain sub-classes generated from our comparative surface analysis and that of the *kinase* sub-classes grouped using multiple sequence alignment and domain architecture [8,9]. Specifically in the case of the *B. subtilis* OmpR sub-family, our receiver domain groupings correlate to greater than 60% to the group IIIA kinase sub-classes that are classified according to domain architecture and configuration [9]. The kinase groups developed for the *E. coli* OmpR sub-family based on primary sequence alignment show significant correlation to our groupings also [8]. This is a notable finding in that classification of two-component regulatory systems has been primarily generated using information about the specific kinase or output domain/activity of the response regulator. Our study reveals that classification and correlations based on both sequence *and* surface characteristics of the receiver domain are plausible and fit well with existing data detailing the relatedness of these communication modules.

5. Concluding remarks

For the first time we have been able to sub-classify response regulators based on the surface characteristics of their receiver domains, primarily in the region comprised of α1-helix and the α1-helix/α5-helix interface. This type of analysis complements and extends existing classification data concerning two-component systems and further supports the idea that different sub-classes of His-Asp phosphorelay systems evolved independently as noted by Grebe and Stock [8]. While the OmpR sub-family is very close in sequence, domain architecture and output domain function compared to other response regulator families, it is clear that there are proteins within this sub-family that are more homologous to some than to others. It is plausible that this approach is somewhat general, and that in many cases, sub-classes within a broad grouping of proteins can be revealed using knowledge of protein–protein interaction surfaces.

Acknowledgements: This work was supported by NIH RO1 GM55769 (J.C.). We are grateful to Dr. James Hoch (Scripps) for encouragement.

References

[1] Robinson, V.L., Buckler, D.R. and Stock, A.M. (2000) Nat. Struct. Biol. 7, 626–633.
[2] Mizuno, T. (1998) J. Biochem. (Tokyo) 123, 555–563.
[3] Ronson, C.W., Nixon, B.T. and Ausubel, F.M. (1987) Cell 49, 579–581.
[4] Kunst, F. et al. (1997) Nature 390, 249–256.
[5] Blattner, F.R. et al. (1997) Science 277, 1453–1474.
[6] West, A.H. and Stock, A.M. (2001) Trends Biochem. Sci. 26, 369–376.
[7] Meeks, J.C., Elhai, J., Thiel, T., Potts, M., Larimer, F., Lamerdin, J., Predki, P. and Atlas, R. (2001) Photosynth. Res. 70, 85–106.
[8] Grebe, T.W. and Stock, J.B. (1999) Adv. Microb. Physiol. 41, 139–227.
[9] Fabret, C., Feher, V.A. and Hoch, J.A. (1999) J. Bacteriol. 181, 1975–1983.
[10] Volz, K. (1993) Biochemistry 32, 11741–11753.
[11] Pao, G.M. and Saier Jr., M.H. (1995) J. Mol. Evol. 40, 136–154.
[12] Hakenbeck, R. and Stock, J.B. (1996) Methods Enzymol. 273, 281–300.
[13] Russell, R.B. (2002) Mol. Biotechnol. 20, 17–28.
[14] Ponting, C.P. and Russell, R.R. (2002) Annu. Rev. Biophys. Biomol. Struct. 31, 45–71.
[15] Kriventseva, E.V., Biswas, M. and Apweiler, R. (2001) Curr. Opin. Struct. Biol. 11, 334–339.
[16] Tzeng, Y.L. and Hoch, J.A. (1997) J. Mol. Biol. 272, 200–212.
[17] Tzeng, Y.L., Feher, V.A., Cavanagh, J., Perego, M. and Hoch, J.A. (1998) Biochemistry 37, 16538–16545.
[18] Zapf, J., Sen, U., Madhusudan, Hoch, J.A. and Varughese, K.I. (2000) Struct. Fold. Des. 8, 851–862.
[19] Zhao, R., Collins, E.J., Bourret, R.B. and Silversmith, R.E. (2002) Nat. Struct. Biol. 9, 570–575.
[20] McEvoy, M.M., Hausrath, A.C., Randolph, G.B., Remington, S.J. and Dahlquist, F.W. (1998) Proc. Natl. Acad. Sci. USA 95, 7333–7338.
[21] Zhu, X., Volz, K. and Matsumura, P. (1997) J. Biol. Chem. 272, 23758–23764.
[22] Sanchez, R. and Sali, A. (2000) Methods Mol. Biol. 143, 97–129.
[23] Sali, A. and Blundell, T.L. (1993) J. Mol. Biol. 234, 779–815.
[24] Altschul, S.F. and Koonin, E.V. (1998) Trends Biochem. Sci. 23, 444–447.
[25] Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R. and Thornton, J.M. (1996) J. Biomol. NMR 8, 477–486.
[26] Colovos, C. and Yeates, T.O. (1993) Protein Sci. 2, 1511–1519.
[27] Eisenberg, D., Luthy, R. and Bowie, J.U. (1997) Methods Enzymol. 277, 396–404.
[28] Koradi, R., Billeter, M. and Wuthrich, K. (1996) J. Mol. Graph. 14, 29–32.
[29] Birck, C., Chen, Y., Hulett, F.M. and Samama, J.P. (2003) J. Bacteriol. 185, 254–261.
[30] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.
[31] Wolfenden, R., Andersson, L., Cullis, P.M. and Southgate, C.C. (1981) Biochemistry 20, 849–855.
[32] Hoch, J.A. and Varughese, K.I. (2001) J. Bacteriol. 183, 4941–4949.