

On Evolutionary Conservation of Thermodynamic Coupling in Proteins*

Received for publication, March 8, 2004
Published, JBC Papers in Press, March 15, 2004, DOI 10.1074/jbc.M402560200

Anthony A. Fodor and Richard W. Aldrich‡

From the Department of Molecular and Cellular Physiology, and Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, California 94305-5345

The inherent complexity of thermodynamic coupling in proteins presents a major challenge in understanding and engineering protein function. Recent work has argued that the study of proteins can be simplified by the use of correlated mutations in the evolutionary record to locate a small subset of thermodynamically coupled residues that participate in functionally important, evolutionarily conserved energetic pathways. To test this hypothesis, we examined the predictions of correlated mutation algorithms for a number of proteins for which coupling between residues has been determined by analysis of double mutant cycles. We find that correlated mutation algorithms can find residue pairs that are physically close and that physically close residue pairs tend to be thermodynamically coupled. We find little evidence, however, for the hypothesis that thermodynamic coupling is limited to the subset of evolutionarily constrained residue positions.

When residues in a protein are thermodynamically coupled, changing one residue affects the energetics of all coupled residues obscuring the interpretation of site-directed mutagenesis experiments and complicating our ability to understand and engineer protein function (1–3). A common technique for understanding thermodynamic coupling is the double mutant cycle, in which the energetic independence of two residue positions is established if the sum of the free energy changes of two independent mutations is equal to the free energy change of the double mutant (4, 5). Even a modestly sized protein, however, has hundreds of millions of possible double mutations that render systematic laboratory study of double mutant cycles impractical. How can an appropriate subset of residues be chosen for experimental study?

An obvious place to look for guidance is the evolutionary record. One approach to relating information in protein sequence databases to protein energetics involves the detection of correlated mutations. If every time a given residue in a column of an alignment changes, there is a corresponding change in another column of the alignment, then the two corresponding residue positions may be energetically linked and under selective pressure. In 1994, Horovitz *et al.* (6) proposed that coordinated changes between two columns in a protein sequence

alignment predicted non-additivity in the results of the corresponding double mutant cycle. In a series of papers, Ranganathan and colleagues (7–10) have expanded on this idea to argue that a correlated mutation algorithm could be used to find “pathways of energetic connectivity” that “have emerged early in the evolution of the protein folds and, much like the atomic structure, are fundamentally conserved features of the domain families” (7). If this hypothesis were generally true, it would have profound consequences for experimental protein science. Rather than attempting to perform experiments on unmanageably complex networks of coupled residues, computational screens of aligned sequences could lead investigators to the small subset of important residues that form “evolutionarily conserved sparse networks” (8) of thermodynamically linked energetic pathways.

In this study, we wanted to test the hypothesis that information in multiple sequence alignments can generally be used to find conserved energetic pathways in proteins. Toward that end, we examined four published double mutant cycle data sets. We found that correlated mutation algorithms predict the results of only one of these four data sets. Our results argue against the general principle that there are isolated pathways of evolutionarily conserved energetic connectivity in proteins.

MATERIALS AND METHODS

The Gobel covariance algorithm (11) was implemented as previously described (12). The SCA¹ algorithm was implemented as previously described (7, 13) except that when performing calculations for the PDZ family, the (*i*, *j*) ordering was constrained so that the residue at the equivalent position to position 76 reported by Lockless and Ranganathan (7) was always in the *i* position. This was done so that our results in Fig. 1*D* would correspond as closely as possible to Fig. 3*C* of Lockless and Ranganathan (7). There are still small differences between our Fig. 1*D* and their Fig. 3*C*, which we believe can be explained by differences in the alignments that were used. Lockless and Ranganathan (7) created their own alignment with PSI-BLAST, while we used the Pfam PDZ alignment. In addition, there are significant changes in the description of the SCA algorithm and its implementation in the Windows binaries distributed by the Ranganathan laboratory (see the “Methods” section of Dekker *et al.* (13)), which may be a cause of some of the differences between the two figures.

For all $\Delta\Delta G$ values, the absolute value of the reported $\Delta\Delta G$ was taken. For the PDZ data set, the Pfam (14) PDZ alignment was downloaded, and the DLG4_MOUSE sequence was aligned using CLUSTALW (15) to the 1BE9 Protein Data Bank (16) structure. For the Barnase-Barnstar complex the 1BRS Protein Data Bank crystal structure was used. Distances were calculated as the average $C\beta$ - $C\beta$ distance for each pair of residues of the AD, BE, and CF complexes. For the Staphylococcal Nuclease data sets, the SNase Pfam alignment was downloaded, and the NUC_STAAU sequence was aligned to the 2SNS Protein Data Bank structure. The SNase alignment did not extend to residue 7 of 2SNS. The 10 double-mutant cycle data points from Green and Shortle (17), which involved residue 7, were therefore excluded from Fig. 2, *D* and *E*, but not from Fig. 3. (Note that what we call $\Delta\Delta G$

* This work was supported by the Mathers Foundation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ An investigator with the Howard Hughes Medical Institute. To whom correspondence and requests for materials should be addressed: Dept. of Molecular and Cellular Physiology, and Howard Hughes Medical Inst., Stanford University School of Medicine, Beckman Center, 279 Campus Dr., Stanford, CA 94305-5345. Tel.: 650-723-6531; Fax: 650-725-4463; E-mail: raldrich@stanford.edu.

¹ The abbreviation used is: SCA, statistical coupling analysis.

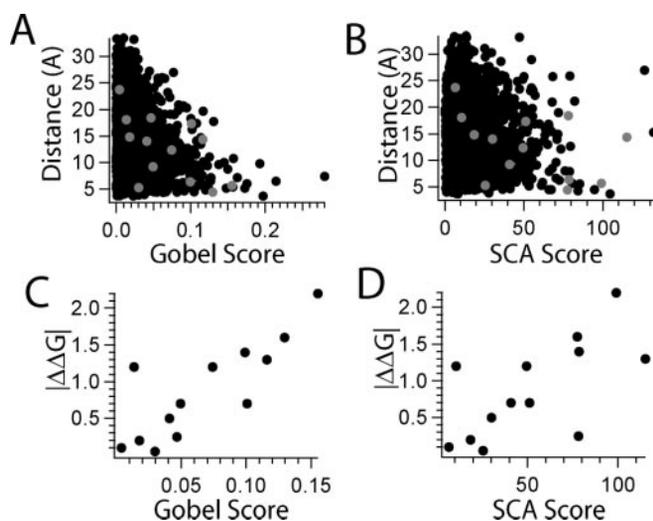


FIG. 1. Distance and energy as a function of SCA and Gobel covariance for the PDZ binding domain. A and B, C β -C β distance as a function of Gobel (A) and SCA (B) covariance. The probability that a random pairing algorithm could do as well in finding residues below the 50th pair distance percentile for the top 75 pairs of residues is $p < 10^{-11}$ for Gobel covariance and $p = 0.005$ for SCA covariance. Gray symbols are the residue pairs for which there are double mutant cycle data. C and D, the Lockless and Ranganathan (7) double mutant cycle data set as a function of Gobel (C) and SCA (D) covariance. Spearman rank correlation scores are given in Table I.

was referred to as $\Delta\Delta\Delta G$ in the Green and Shortle paper (17)). For the potassium channel data set, the Q9YDF8 sequence in the ion_trans Pfam family was aligned to the 1ORQ crystal structure. Because the Shaker sequence (CIKS_DROME) in the ion_trans Pfam alignment did not extend to Shaker residue 366, the single data point for Leu³⁶⁶ \rightarrow Val⁴⁷⁶ was excluded from Fig. 2F. For all Pfam families, redundant sequences with (>90% identity) were removed. Because we wanted to include as much energy data as possible, we filtered the Pfam alignments in a slightly less rigorous way than in our previous study (12); specifically, we considered lowercase symbols to be valid Pfam residues, and we only excluded a column from the analysis that had greater than 90% gapped residues. These changes from our previous methods had virtually no effect on the overall power of the covariance algorithms (data not shown).

Note that we excluded one residue pair from the PDZ double mutant data set. In their paper, Lockless and Ranganathan (7) report data for a double mutant V32I/H76Y. However, in the 1BFE and 1BE9 PDZ crystal structures (from Doyle *et al.* (18) referenced as footnote 21 in the Lockless and Ranganathan (7) paper), the equivalent position to Val³² is already an isoleucine. The nearest valine is 13 residues upstream. Data from V32I/H76Y were therefore excluded from all our analyses.

RESULTS

Correlated Mutations as Predictors of Distance and Energy—

Before they were used in an attempt to map protein energetics (7), it was recognized that correlated mutation algorithms find pairs of residues that are close in physical space in a protein structure (11, 12, 19–22). In a previous study (12), we observed that a correlated mutation algorithm first described by Gobel *et al.* (11) was able to use alignments from the Pfam (14) data base to correctly predict physically close residue pairs in the corresponding PDB structures at a statistically significant level ($p < 0.05$) for over 91% of the 224 protein families we studied. Fig. 1A shows pair distance as a function of the Gobel covariance algorithm for a member of the PDZ binding domain family. For each point in Fig. 1A, the value on the x axis represents the degree of correlation between residue changes in two columns of the PDZ multiple sequence alignment downloaded from the Pfam data base. The corresponding value on the y axis is the C β -C β distance between the corresponding residues in the 1BE9 crystal structure. As we would expect from our previous work on other protein families (12), the most

highly covarying pairs of residues (to the right on the x axis) tend to be physically close (to the bottom of the y axis).

With unlimited resources, an evaluation of the hypothesis that the results of correlated mutation algorithms can predict the results of double mutant cycle experiments would measure the non-additivity of double mutant cycles for every possible residue pair in a protein. Such an experiment is, of course, not feasible. To evaluate their hypothesis that “pathways of energetic connectivity” are “fundamentally conserved features of the domain families”, Lockless and Ranganathan (7) had to choose a subset of residue pairs for which to create double mutants for the PDZ family. In their study, Lockless and Ranganathan (7) introduced a novel correlated mutation algorithm, which they called SCA for statistical coupling analysis. The relationship between SCA covariance and pair distance for the PDZ alignment and 1BE9 crystal structure is shown in Fig. 1B. Because they were explicitly testing the power of a correlated mutation algorithm, they chose a subset of residue pairs that had a wide range of covariance scores under the SCA algorithm (Fig. 1B, gray symbols).

In a previous study of 224 protein families (12), we found that the SCA covariance algorithm tends to have less power than the Gobel covariance algorithm in finding physically close residue pairs. The PDZ family in Fig. 1, A and B, is no exception to this trend, as the most highly covarying residue pairs under Gobel covariance (Fig. 1A) are closer together than the most highly covarying pairs under SCA (Fig. 1B). Nonetheless, despite these differences, the subset of residue pairs chosen for experimental analysis by Lockless and Ranganathan (7) (Fig. 1, A and B, gray symbols) have a wide range of scores under both algorithms. The relationship between $\Delta\Delta G$, the non-additivity of double mutant cycle experiments, and covariance is shown for the Gobel algorithm in Fig. 1C and for SCA in Fig. 1D. We see that under both of these algorithms, these two values are nicely correlated. These correlations provide evidence for the hypothesis that energetic coupling is under tight evolutionary control.

We were curious as to how well this relationship between correlated mutations and the non-additivity of double mutant cycle experiments would hold up when applied to other double mutant cycle data sets. The labor-intensive nature of performing double mutant cycles has led to the creation of few data sets large enough to have the potential to generate a meaningful correlation. Moreover, many double mutant cycle data sets are between proteins and their peptide or protein binding partners (23–25), which makes them unsuitable for a correlated mutation analysis that needs to be performed on a single protein alignment. There are, however, two appropriately large sets of double mutant cycles based on the folding energetics of Staphylococcal Nuclease (1, 17). In these data sets, free energy differences between the folded and unfolded state of Staphylococcal Nuclease were estimated by monitoring fluorescent changes as the protein was denatured using guanidine hydrochloride. In addition, there is a double mutant cycle data set based on the energetics of pore opening of the Shaker K⁺ channel in which currents were recorded across a range of voltages, and free energy differences between the open and closed state were generated from a model that assumed that mutations effect a single opening transition (26). Fig. 2, A and B, show that, as we would expect, the most highly covarying residue pairs under Gobel covariance tend to have small C β -C β distances in a corresponding crystal structure for Staphylococcal Nuclease. While there is no crystal structure of the Shaker potassium channel, Fig. 2C shows that the most highly covarying residue pairs under Gobel covariance are physically close in the 1ORQ crystal structure, a prokaryotic homologue (KvAP) of the

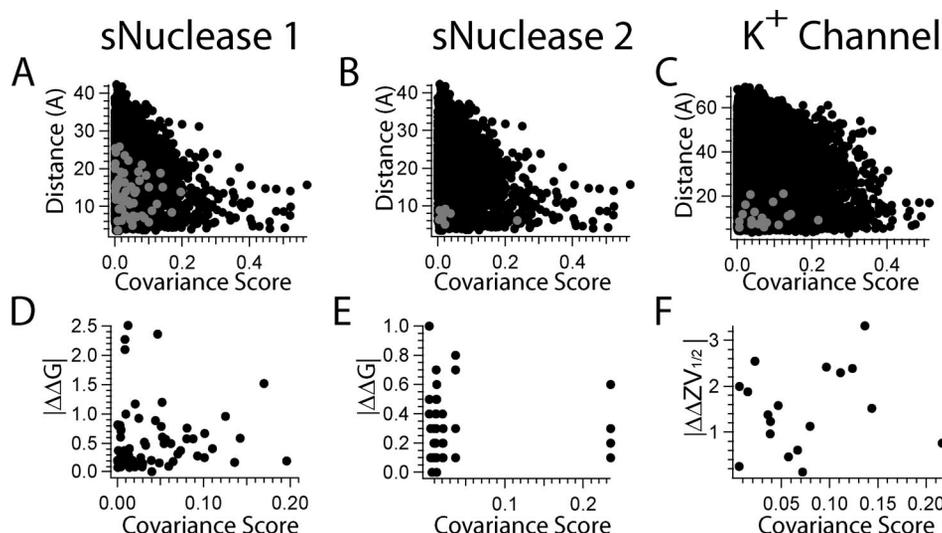


FIG. 2. The relationship between non-additivity in free energy, Gobel covariance, and $C\beta$ - $C\beta$ distance for the Green and Shortle (17) *Staphylococcal Nuclease* data set (A and D), the Chen and Stites (1) *Staphylococcal Nuclease* data (B, E) set, and the Yifrach and MacKinnon Shaker (26) *potassium channel* data set (C, F). A–C, pair distance as a function of Gobel covariance score. Gray symbols are the residue pairs for which there are double mutant cycle data. The probability that a random pairing algorithm could do as well in finding residues below the 50th pair distance percentile for the top 75 pairs of residues is $p < 10^{-13}$ for *Staphylococcal Nuclease* (A and B) and $p < 10^{-7}$ for the *K⁺ channel KvAP* (C). See the “Methods” section of Fodor and Aldrich (12) for the details on calculating these probabilities. D–F, non-additivity in free energy as a function of Gobel covariance scores. Spearman rank correlation scores are given in Table I.

Shaker potassium channel. For all three data sets, however, there is no correlation between the non-additivity of free energy and the degree of correlated mutations (Fig. 2, D–F; Table I). This lack of correlation argues against the idea that double mutant cycle experiments reliably lead us to energetic pathways that are fundamentally conserved features (7) of domain families.

The results shown in Fig. 2 are for the Gobel covariance algorithm, which in a previous survey of four covariance algorithms displayed the most power (12). In addition to the Gobel covariance method, we examined the ability of other covariance algorithms, including the SCA algorithm (7, 8), to predict the results of double mutant cycle experiments. We found a similar level of performance from all of the algorithms that we examined (Table I). It therefore seems unlikely that our results can be explained by our choice of a particular covariance algorithm.

A potential source of error in these methods is the multiple sequence alignments fed to the covariance algorithms. If there were biases or errors in the construction of the Pfam alignments that we used, it might explain why correlated mutation algorithms failed to predict energetic coupling. However, as can be seen in the *top panels* of Figs. 1 and 2, the Gobel correlated mutation algorithm was able to use the Pfam alignments to successfully predict physically close residue pairs. The probability that a random pairing algorithm could do as well as Gobel covariance in finding residues below the 50th pair distance percentile for the top 75 pairs of residues is $p < 10^{-7}$ or smaller for the families shown in Figs. 1A and 2, A–C. The fact that the information in the alignments can successfully be used to discover close residue pairs suggests that the use of Pfam alignments is appropriate for the problem at hand and argues against errors or biases in these alignments causing the poor correlations observed in Fig. 2, D–F.

Distance as a Predictor of Energy—Why is there such a striking correlation for the PDZ family (Fig. 1, C and D), while the other three data sets show no correlation (Fig. 2, D–F)? Because the PDZ data set was created to explicitly test the results of a correlated mutation algorithm, some of the PDZ residue pairs that were chosen have high covariance ranks (Fig. 1, A and B, gray symbols). For the other protein domains, none of the chosen residue pairs ranked among the most highly

covarying residue pairs (Fig. 2, A–C, gray symbols). As we see in the *top panels* of Figs. 1 and 2, and has been shown repeatedly in the literature, pairs of residues with the highest scores under correlated mutation algorithms tend to be physically close (11, 12, 19–21). The forces that can generate thermodynamic coupling in proteins also tend to work over fairly short distances (1). For example, a recent study estimates that helix interactions are more likely to occur when α -carbons are less than 12 Å apart (27). It therefore seems likely that at least part of the reason that a covariance algorithm is able to find coupled residues is that highly covarying residue pairs tend to be physically close.

To investigate the relationship between distance and thermodynamic coupling, we looked for double mutant cycle data sets for which there was a known structure (>90% sequence identity) and for which energetic coupling data were reported across a wide range of distances (up to at least 15 Å). Fig. 3 shows the results for three such double mutant cycle data sets. The blue symbols are from a Barnase-Barnstar double mutant cycle data set (25). While we could not perform a correlated mutation analysis of the complexed Barnase-Barnstar proteins, since such an analysis requires an alignment from a single protein family, there is a crystal structure of the Barnase-Barnstar complex available that allows for a plot of distance *versus* non-additivity in energy. The red symbols in Fig. 3 are from the PDZ study (7) shown in Fig. 1. The black symbols are from the *Staphylococcal Nuclease* study (17) shown in Fig. 2D. There is a consistent trend across all three data sets in Fig. 3 with pronounced thermodynamic coupling much more likely to occur if the $C\beta$ - $C\beta$ distance of the residue pair is less than ~12 Å.

Our observation that thermodynamic coupling is largely limited to physically close residue pairs is not surprising. Indeed, a number of double mutant cycle studies (23, 24) have argued that thermodynamic coupling can be used to find physically close residue pairs in protein-peptide complexes for which there is no available crystal structure. Nonetheless, the interpretation of the relationship between energetic non-additivity and distance can vary quite a bit from study to study. For example, the Barnase-Barnstar (25) study (data plotted as blue symbols in Fig. 3) concluded that “coupling energy between two

TABLE I
The performance of different covariance algorithms in predicting thermodynamic coupling for the four data sets shown in Figs. 1, C and D, and 2, D–F

r and p values shown are from the Spearman rank correlation test. See the “Methods” section of Fodor and Aldrich (12) for implementation details of each algorithm.

Algorithm	PDZ	sNuclease 1	sNuclease 2	Shaker
Gobel <i>et al.</i> (McBASC)	$p = 0.001$ $r = -0.80$ (Fig. 1C)	$p = 0.67$ $r = 0.06$ (Fig. 2D)	$p = 0.99$ $r = 0.00$ (Fig. 2E)	$p = 0.56$ $r = 0.15$ (Fig. 2F)
SCA	$p = 0.01$ $r = 0.68$ (Fig. 1D)	$p = 0.94$ $r = 0.00$	$p = 0.67$ $r = 0.06$	$p = 1.0$ $r = 0.0$
OMES	$p = 0.004$ $r = 0.74$	$p = 0.84$ $r = -0.03$	$p = 0.34$ $r = 0.15$	$p = 0.93$ $r = -0.02$

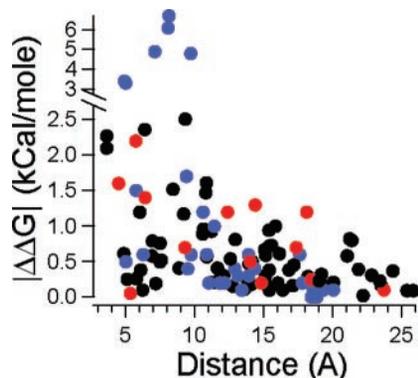


FIG. 3. Combined $\Delta\Delta G$ versus distance plots for the PDZ binding domain (7) (red symbols), the Green and Shortle Staphylococcal Nuclease data set (17) (black symbols), and the Barnase-Barnstar complex (25) (blue symbols). Spearman rank for the combined data sets $p = 4.0 \times 10^{-6}$, $r = -0.41$

residues was found to decrease with the distance between them.” The Staphylococcal Nuclease study (17) (data plotted as black symbols in Fig. 3), on the other hand, came to the opposite conclusion that there were patterns of interactions at “widely separated positions.” It is apparent from Fig. 3, however, that despite the different conclusions, the data from these two studies overlap a good deal.

Part of the evidence from Lockless and Ranganathan (7) for the idea that correlated mutation algorithms can predict the results of thermodynamic coupling is that the SCA algorithm predicted a “set of energetically coupled positions . . . that includes unexpected long range interactions” and that “mutational studies confirm these predictions” (7). This line of argument suggests that correlated mutation algorithms can find energetically connected residues that could not be discovered by examining close residues in crystal structures. However, in the PDZ data set, the residue pairs that showed non-additivity > 1.5 kcal/mol were within 8 Å of each other and the only residue pair > 20 Å apart showed very little energetic connectivity (Fig. 3, red symbols). Moreover, the PDZ distance versus energy data (Fig. 3, red symbols) can be roughly superimposed with the two other double mutant cycle data sets (Fig. 3, blue and black symbols). This suggests that the “long range” thermodynamic coupling observed in the PDZ experiment was not unusual and that across all three data sets pronounced thermodynamic coupling is more likely to occur among close residues. Correlated mutation algorithms reliably find physically close residue pairs (12) but do not reliably find energetically coupled pairs (Fig. 2, D–F). The most parsimonious explanation, therefore, of the ability of correlated mutation algorithms to discover energetically linked residues is that correlated mutation algorithms discover physically close residues. The close residues discovered by correlated mutation algorithms, however, are not part of a “sparse network” (8) of energetically

linked residues. Rather, any algorithm that chooses a set of physically close residues might detect thermodynamic coupling. In particular, it has been demonstrated repeatedly (12, 28–30) that highly conserved residues also tend to cluster together in the protein core. We would expect, therefore, that these clustered sets of conserved residues, like the clustered sets of covarying residues, would tend to be thermodynamically coupled.

DISCUSSION

Conserved columns in multiple sequence alignments have successfully been used to find buried residues in the protein core (30). It has also been found that mutation of conserved residues is often associated with disruption of protein function and disease causing phenotypes (31). Despite these successful applications of sequence information to predicting protein structure and function, attempts to relate the evolutionary record to protein energetics have been controversial. For example, it has been proposed that conserved residues are more likely to participate in the structure of the folding-unfolding transition state (32, 33). However, attempts to relate the impact of a mutation on the stability of the transition state to the degree of conservation of that residue position has found little significant correlation (34, 35). In discussing these results, Plaxco *et al.* (35) noted that while “sequence alone encodes the three-dimensional structure of a protein and the rate with which that structure is formed . . . protein folding kinetics are relatively insensitive to fine details of sequence and are, instead, defined by sequence through its effects on more global parameters such as topology . . . and, to a lesser extent, stability.” In our study we found that the results of correlated mutation algorithms predicted the degree of thermodynamic coupling for only one of four data sets. It therefore seems reasonable to take a similar line of arguments regarding thermodynamic coupling. We argue that, as appears to be the case for protein kinetics, thermodynamic coupling may be insensitive to the fine details of sequence. Rather, the global protein fold moves some residues close to other residues and a subset of those residues within ~ 12 Å of one another are coupled (Fig. 3). Conservation and covariance algorithms reliably find physically close sets of residues (12, 20), and we would expect, therefore, that because they tend to be close, conserved and covarying residue positions would also tend to be thermodynamically coupled. Thermodynamic coupling, however, does not appear to be limited to the subset of highly covarying residue positions.

Acknowledgments—We thank W. Li, R. Piskowski, and J. Sack for critically reading an earlier version of this manuscript.

REFERENCES

- Chen, J. & Stites, W. E. (2001) *Biochemistry* **40**, 14004–14011
- Green, S. M., Meeker, A. K. & Shortle, D. (1992) *Biochemistry* **31**, 5717–5728
- Skinner, M. M. & Terwilliger, T. C. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93**, 10753–10757

4. Carter, P. J., Winter, G., Wilkinson, A. J. & Fersht, A. R. (1984) *Cell* **38**, 835–840
5. Horowitz, A. (1996) *Fold Des.* **1**, R121–R126
6. Horowitz, A., Bochkareva, E. S., Yifrach, O. & Girshovich, A. S. (1994) *J. Mol. Biol.* **238**, 133–138
7. Lockless, S. W. & Ranganathan, R. (1999) *Science* **286**, 295–299
8. Suel, G. M., Lockless, S. W., Wall, M. A. & Ranganathan, R. (2003) *Nat. Struct. Biol.* **10**, 59–69
9. Shulman, A., Larson, C., Mangelsdorf, D. & Ranganathan, R. (2004) *Cell* **116**, 417–429
10. Hatley, M. E., Lockless, S. W., Gibson, S. K., Gilman, A. G. & Ranganathan, R. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100**, 14445–14450
11. Gobel, U., Sander, C., Schneider, R. & Valencia, A. (1994) *Proteins* **18**, 309–317
12. Fodor, A. & Aldrich, R. (2004) *Proteins*, in press
13. Dekker, J., Fodor, A., Aldrich, R. & Yellen, G. (2004) *Bioinformatics*, in press
14. Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M. & Sonnhammer, E. L. (2002) *Nucleic Acids Res.* **30**, 276–280
15. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673–4680
16. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000) *Nucleic Acids Res.* **28**, 235–242
17. Green, S. M. & Shortle, D. (1993) *Biochemistry* **32**, 10131–10139
18. Doyle, D. A., Lee, A., Lewis, J., Kim, E., Sheng, M. & MacKinnon, R. (1996) *Cell* **85**, 1067–1076
19. Altschuh, D., Lesk, A. M., Bloomer, A. C. & Klug, A. (1987) *J. Mol. Biol.* **193**, 693–707
20. Olmea, O., Rost, B. & Valencia, A. (1999) *J. Mol. Biol.* **293**, 1221–1239
21. Larson, S. M., Di Nardo, A. A. & Davidson, A. R. (2000) *J. Mol. Biol.* **303**, 433–446
22. Shindyalov, I. N., Kolchanov, N. A. & Sander, C. (1994) *Protein Eng.* **7**, 349–358
23. Ranganathan, R., Lewis, J. H. & MacKinnon, R. (1996) *Neuron* **16**, 131–139
24. Hidalgo, P. & MacKinnon, R. (1995) *Science* **268**, 307–310
25. Schreiber, G. & Fersht, A. R. (1995) *J. Mol. Biol.* **248**, 478–486
26. Yifrach, O. & MacKinnon, R. (2002) *Cell* **111**, 231–239
27. Lee, S. & Chirikjian, G. S. (2004) *Biophys. J.* **86**, 1105–1117
28. Taylor, W. R. & Hatrick, K. (1994) *Protein Eng.* **7**, 341–348
29. Olmea, O. & Valencia, A. (1997) *Fold Des.* **2**, S25–S32
30. Schueler-Furman, O. & Baker, D. (2003) *Proteins* **52**, 225–235
31. Mooney, S. D. & Klein, T. E. (2002) *BMC Bioinformatics* <http://biomedcentral.com/1471-2105/3/24>
32. Shakhnovich, E., Abkevich, V. & Puitsyn, O. (1996) *Nature* **379**, 96–98
33. Mirny, L. & Shakhnovich, E. (2001) *J. Mol. Biol.* **308**, 123–129
34. Larson, S. M., Ruczinski, I., Davidson, A. R., Baker, D. & Plaxco, K. W. (2002) *J. Mol. Biol.* **316**, 225–233
35. Plaxco, K. W., Larson, S., Ruczinski, I., Riddle, D. S., Thayer, E. C., Buchwitz, B., Davidson, A. R. & Baker, D. (2000) *J. Mol. Biol.* **298**, 303–312