features with maxima higher than five times the rms deviation from the mean density (Fig. 4). The only feature that might have represented a sixth ligand had a maximum density similar to the rms deviation. In the final ($F_o - F_c$) map, a bean-shaped feature at three times the rms deviation was observed 2.4 Å from the Fe and along the line between Tyr^{250}-O^{η} and water-3012. However, when a water molecule was added at this site, refinement resulted in displacement to 2.6 Å away from the Fe and a *B* factor of 64 $Å^2$; in contrast, the *B* factors for the ligands and Fe atom are 11 to 14 $Å^2$. Thus, the addition of a water molecule as a sixth ligand is not warranted.
21. R. C. Scarrow *et al.*, *Biochemistry* **33**, 15023 (1994).
22. To assess the reliability of the Fe-ligand distances, we investigated the influence of refinement restraints, the consequences of variations in the resolution and source of the diffraction data, and tested for bias in the final model by systematically changing Fe-ligand distances or randomly perturbing the entire model. In brief, we found that the use of weak Fe-ligand bond distance restraints (for example, 10 kcal mol^{-1} $Å^{-2}$) produced consistent results for all ligands given a variety of starting models and diffraction data sets. None of the perturbations caused variations as large as 0.1 Å for any distance. The final structure is also entirely consistent with omit-refine maps (*30*) obtained by omitting all atoms within 6 Å of the Fe.
23. A *t*-butanol molecule was assigned to a large-density feature 6 Å from the Fe. In the view shown in Fig. 4, the *t*-butanol is located along the line-of-sight to Tyr^{250}-O^{η} and approximately at the depth of the side chain of His^{209}; it is not illustrated because it obscures the Fe. The *t*-butanol blocks the most direct route to the Fe via the large opening of the funnel. This is consistent with the observation that *t*-butanol, in concentrations similar to that in the crystallization solution, inhibits and stabilizes the enzyme.
24. J. Singh and J. M. Thornton, *Atlas of Protein Side-Chain Interactions* (Oxford Univ. Press, New York, 1992).
25. A. W. E. Chan, E. G. Hutchinson, D. Harris, J. M. Thornton, *Protein Sci.* **2**, 1574 (1993).
26. C-domain ligands His^{146}, His^{195}, and His^{210} correspond to N-domain residues Tyr^{8}, Trp^{48}, and Phe^{62}. Glu^{115}, the residue corresponding to ligand Glu^{260}, forms a salt bridge with Arg^{98}, which extends from across the funnel. Thus, neither the ligands nor the free space necessary for Fe and substrate binding are available. However, a second Fe is found at the surface of the C-domain for crystals that were transferred to a stabilizing solution containing 1 mM $(NH_4)_2Fe(SO_4)_2$ before diffraction measurements. The additional Fe is bound only by His^{189} and, according to its anomalous difference density, is present at partial (~50%) occupancy. Anomalous difference density is observed only for crystals treated with excess Fe.
27. K. Hori, T. Hashimoto, M. Nozaki, *J. Biochem. (Tokyo)* **74**, 375 (1973).
28. Strongly conserved residues, including Val^{148}, Phe^{187}, Ala^{198}, His^{209}, and Tyr^{250}, border on this larger pocket, consistent with its assignment as the distal ring binding site.
29. Binding to the Fe(II) center may activate dioxygen by converting it to an Fe-bound, superoxide-like species (*17*), which attacks the substrate at C1, the point of attachment for the distal ring. For either binding mode, conserved residues are available to promote the formation of bound superoxide by stabilizing a partial negative charge on its distal oxygen atom. For mode A, dioxygen would likely bind in the vacant sixth site of the free enzyme. This position is near the hydrogen bonded His^{241}-Tyr^{250} pair and especially close to O^{η} of Tyr^{250}, which might stabilize the superoxide by an interaction involving the partial positive charge associated with the O^{η} proton. In the case of mode B, dioxygen would likely bind in the water-3012 site near His^{195}. If His^{195} is protonated on substrate binding, it would provide a positive charge to promote formation of the superoxide.
30. A. Hodel, S.-H. Kim, A. T. Brünger, *Acta Crystallogr. Sect. A* **48**, 851 (1992).
31. P. J. Kraulis, *J. Appl. Crystallogr.* **24**, 946 (1991).
32. W. Kabsch, *ibid.* **21**, 916 (1988).
33. Z. Otwinowski, in *Proceedings of the CCP4 Study Weekend: Data Collection and Processing*, L. Sawyer, N. Issacs, S. Bailey, Eds. [Science and Engineering Research Council (SERC) Daresbury Laboratory, Warrington, UK, 1991)], pp. 56–62; W. Minor, *XDISPLAYF* (Purdue University, West Lafayette, IN, 1993).
34. *CCP4: A Suite of Programs for Protein Crystallography* (SERC Daresbury Laboratory, Warrington WA4 4AD, UK, 1979).
35. Z. Otwinowski, in *Proceedings of the CCP4 Study Weekend: Isomorphous Replacement and Anomalous Scattering*, W. Wolf, P. R. Evans, A. G. W. Leslie, Eds. (SERC Daresbury Laboratory, Warrington, UK, 1991), pp. 80–86.
36. J. T. Bolin, J. L. Smith, S. W. Muchmore, paper presented at the Annual Meeting of the American Crystallographic Association, Albuquerque, NM, 26 May 1993.
37. T. A. Jones, J.-Y. Zou, S. W. Cowan, M. Kjelgaard, *Acta Crystallogr. Sect. A* **47**, 110 (1991).
38. R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, *J. Appl. Crystallogr.* **26**, 283 (1993).
39. A. T. Brünger, *X-PLOR, Version 3.1* (Yale Univ. Press, New Haven, CT, 1992).
40. We acknowledge the contributions of S. Harayama, B. Hofmann, and H.-J. Hecht to preliminary crystallization and diffraction studies. The assistance of R. K. Thauer and R. Hedderich in the initial anaerobic purification of DHBD is also gratefully acknowledged. Research in Braunschweig was supported in part by the German Ministry of Research and Technology. K.N.T. also thanks the Fonds der Chemischen Industrie for generous support. J.T.B. thanks L. Holm and G. Vriend for assistance with a preliminary evaluation of structural homology, and the Lucille P. Markey Foundation and the National Institutes of Health (NIH) (GM 52831) for support. Shared diffraction and computing facilities at Purdue University have been supported by grants from NIH and the National Science Foundation.

31 May 1995; accepted 23 August 1995

Guidelines for Protein Design: The Energetics of β Sheet Side Chain Interactions

Catherine K. Smith and Lynne Regan*

To determine the interaction energy between cross-strand pairs of side chains on an antiparallel β sheet, pairwise amino acid substitutions were made on the solvent-exposed face of the B1 domain of streptococcal protein G. The measured interaction energies were substantial (1.8 kilocalories per mole) and comparable to the magnitude of the β sheet propensities. The experimental results paralleled the statistical frequency with which the residue pairs are found in β sheets of known structure.

A complete understanding of the construction of α helices and β sheets is essential for the manipulation of the structure and properties of proteins. A quantitative energetic description of many factors that determine the stability of α-helical secondary structure is available (*1–4*). Our understanding of the interactions that determine β sheet stability is much less advanced. Different amino acids do have measurably different intrinsic propensities for forming β sheets (*5*, *6*), and these intrinsic β sheet–forming propensities differ between central and edge strand positions in a β sheet (*7*). Statistical surveys of proteins of known structure reveal that a nonrandom pairwise distribution of amino acids occurs in adjacent positions across two β strands (*8–10*). This observation implies that in addition to the underlying differences in β sheet propensity, side chain interactions contribute to the local structure and stability of β sheets. Our study here measures the extent to which these side chain interactions contribute to β sheet stability.

A variant of the immunoglobulin G binding domain of streptococcal protein G (referred to as β1 and illustrated in Fig. 1A) was originally used to quantify the β sheet–forming propensities of the amino acids (*6*). Here, β1 was used to measure the energetics of side chain interactions between pairs of amino acids located in adjacent β strands. Pairwise amino acid substitutions were made on the solvent-exposed face of β1 in an antiparallel β sheet. The investigated site, positions 44 and 53, was selected to provide maximal solvent exposure for the studied side chains in the most homogeneous sheet environment available. Because the backbone atoms of amino acids 44 and 53 are hydrogen-bonded to one another, the site of mutation occurs in a hydrogen-bonded position (*8*), also called a narrow H-bond ring (*11*). This type of site has statistical pairwise preferences distinct from those of a non–H-bonded pair such as 53-6 (*8*). Pairwise mutants are referred to as β1XY, where X is the amino acid at position 44 and Y is the residue at position 53.

The side chain interactions between pairs of the best β sheet–forming residues (Trp, Phe, Tyr, Thr, Ile, and Val) and between pairs of charged residues were examined (*5–7*). The charged residues (Arg, Lys, and Glu) are among the intrinsically poorest β sheet formers, yet these residues would be useful for engineering solubility

Department of Molecular Biophysics and Biochemistry, Yale University, 266 Whitney Avenue, New Haven, CT 06520, USA.

*To whom correspondence should be addressed.

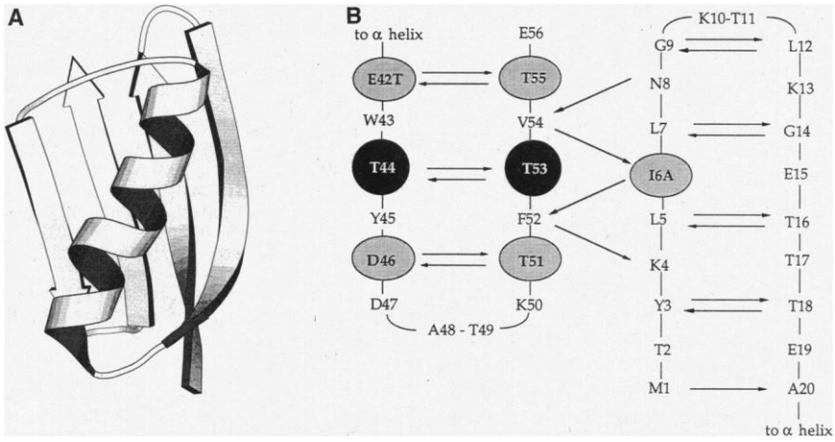

Fig. 1. (**A**) Ribbon diagram illustrating the overall protein fold of β1 (PDB filecode, 2GB1) (*17*). (**B**) Illustration of the location of the pairwise substitutions in the solvent-exposed β sheet region of β1 (*12*). Hydrogen bonds are indicated by arrows. The guest sites, positions 44 and 53, are circled and highlighted in black. The nearest neighbors on the same face of the β sheet are also circled but highlighted in gray. To maximize both solvent exposure of the guest residues and freedom from other interactions, nearest neighbors were mutated (I6A and E42T) on the basis of previous modeling studies (*6*). Because Asp^{46} is involved in hydrogen bonding to Ala^{48} in the loop, it was not altered to Thr as was Glu^{42} (*17*). We introduced the mutations at position 44 by polymerase chain reaction, using the original position 53 mutant DNAs as the template, and the proteins were purified as described (*6*).

and functionality in a new design. Measuring the interactions between these side chains provides the most useful guide for β sheet design and engineering.

We studied 32 pairwise mutants at positions 44 and 53. The stability of each β1 protein containing a pairwise mutation was determined from its thermal denaturation transition, as monitored by circular dichroism. The thermal stabilities of all double mutants were greater than those of single mutants containing the studied residue paired with Ala (illustrated for β1EK, β1EA, and β1AK) (*12*) (Fig. 2). From such stability curves, the free energy of folding (ΔG) for each protein was calculated (*13*). The free energy associated with the pairing of amino acid side chains, Δ(ΔΔG), was also calculated for each pairwise mutant β1XY with the following equation:

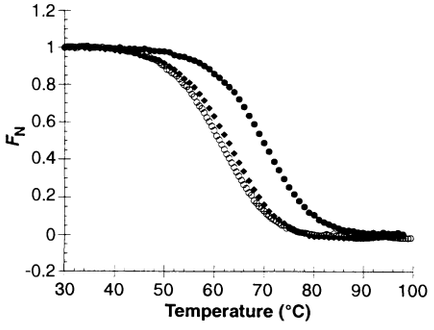

Fig. 2. Typical thermal denaturation curves of a pairwise mutant compared with two single mutants each containing one member of the pair. β1EA, open circles; β1AK, closed diamonds; β1EK, closed circles.

$$\Delta(\Delta\Delta G)^{XY}_{interaction} = \Delta\Delta G^{XY}_{paired} - (\Delta\Delta G^{XA}_{intrinsic} + \Delta\Delta G^{AY}_{intrinsic})$$

where $\Delta\Delta G^{XY}_{paired}$ is the stability of the double mutant containing X and Y amino acids relative to the variant β1AA; $\Delta\Delta G^{XA}_{intrinsic}$ is the β sheet–forming propensity of residue X at the edge position; and $\Delta\Delta G^{AY}_{intrinsic}$ is the β sheet–forming propensity of residue Y at the central strand position (*14*). A favorable side chain interaction is indicated if the stability of the double mutant is greater than that predicted by the sum of the intrinsic propensities.

From an examination of Δ(ΔΔG), it is clear that not all pairs participate in favorable side chain interactions (Table 1). The interaction energy of pairs such as β1TI fell short of the stability predicted by the sum of the individual propensities by 0.75 kcal/mol, whereas the interaction energy of pairs such as β1FF and β1EK exceeded the predicted stability by nearly 1 kcal/mol. The range of free energies associated with side chain interactions is approximately 1.8 kcal/mol. This magnitude is comparable to the range of the intrinsic β sheet–forming propensities of the residues involved in making the pairs (1.4 kcal/mol) as well as to other factors stabilizing secondary structure, such as helix N-capping (*4*). Clearly, both pairwise side chain interactions and β sheet–forming propensities are important in determining β sheet stability.

The pairs containing exclusively aliphatic and aromatic residues ranked consistently higher in the Δ(ΔΔG) scale than the pairs containing Thr. The different pairs of aliphatic and aromatic residues do not all interact equally well, which presumably reflects both side chain volume and packing requirements. An inspection of the sum of side chain volumes in different aliphatic-aromatic pairs suggested that there may be an optimal volume requirement for maximal stability (300 to 375 $Å^3$). A similar stabilization between two appropriately paired solvent-exposed hydrophobic groups has been reported in α helices (*2*). The physical basis for such stabilization may be related to the exclusion of solvent from hydrophobic surfaces of the side chains.

The three pairs of amino acids that were constructed in both orientations (TI, IT; FI, IF; and FT, TF) did not exhibit identical interaction energies. Intrinsic positional effects of the edge and central strand locations were taken into account because the β sheet propensities used in the calculation were specific for these positions. The differences in measured interaction energies therefore may reflect the asymmetry of the site as a whole. In this regard, no single site can be representative of all antiparallel β sheet pairs because there will be differences in each pair's specific local geometry and between H-bonded and non–H-bonded positions.

A correlation between the results of statistical surveys and the experimental data was also clear, as demonstrated by the synergistic and antagonistic effects observed in

Table 1. The amino acid pairs (*12*) are listed in the order of decreasing values of Δ(Δ*ΔG*). Also shown are the statistical correlations for the occurrence of each pair in an antiparallel β sheet in an H-bonded position as determined from statistical surveys of proteins of known structure (*8*). The correlation value is the ratio between the number of times a pair of residues is found together in the data set and the expected number of times that pair would occur randomly. The pair correlations that were determined at a confidence level of at least 90% are underlined.

Mutant	Δ(ΔΔ*G*) (kcal/mol)	Pair correlation	Confidence level (%)
ER	−0.96	3.4	99.5
EK	−0.95	3.4	99.5
FF	−0.91	2.4	99.5
IY	−0.61	1.4	90
FY	−0.61	1.2	50
IF	−0.59	1.0	50
FI	−0.44	1.0	50
II	−0.36	1.0	50
IW	−0.34	1.4	75
FW	−0.31	1.1	50
IV	−0.27	1.0	50
FV	−0.20	1.4	97.5
FT	−0.19	0.7	75
IT	−0.19	0.7	75
TT	0.21	1.6	95
TW	0.24	0.3	95
TF	0.36	0.7	75
TV	0.47	0.7	95
TY	0.54	0.8	75
TI	0.75	0.7	75

both studies (Table 1). In general, a high interaction energy was associated with pairs that are statistically likely to be found together in H-bonded positions of antiparallel β strands. By contrast, a destabilizing interaction energy was exhibited by pairs that are statistically unlikely to be found together. Only the pairs for which the statistical data are known above a 90% confidence level are discussed here.

Statistically, the charged pairs (β1EK and β1ER) and the aromatic pair (β1FF) are the most likely to be paired. In the experimental system, each charged residue individually is a poor β sheet–forming amino acid (6). When each charged residue was paired with another charged cross-strand neighbor, however, a substantial stabilization energy was observed (Table 1). This stabilization is probably due to the complementary electrostatics of the pair. Phenylalanine is one of the best β sheet formers. In proteins of known structure, cross-strand pairs of Phe side chains are often oriented in an edge-to-face fashion, which optimizes energetically favorable electrostatic π-π interactions between the aromatic rings (*8*, *15*). When these strands were paired with a second Phe residue, a large synergistic stabilization was observed in our model system, which suggests that the two Phe side chains may be interacting in this way.

The amino acid pairs TV and TW are two of the least likely to be found together across a β sheet, and the interaction energy measured was destabilizing (Table 1). The poor interaction between these pairs may reflect a tendency of the Thr side chain hydroxyl to form a hydrogen bond with the solvent. As a result, the number of possible side chain orientations and the ability of Thr to interact favorably with its cross-strand neighbor may be restricted.

The pair TT was anomalous because it exhibited low side chain interaction energy but a high statistical correlation (Table 1). However, the statistics do not distinguish between the exposed and buried faces of the studied β sheets. The high correlation of the TT pair may reflect a preference for a partner of similar hydrophilicity rather than a specific side chain interaction (*10*).

Table 2 lists the measured pairs in the order of their relative free energy, illustrating which pairs were most favorable for β sheet structure overall. Although the charged pairs exhibited the highest interaction energy, they did not form the most stable proteins. This was not surprising given the low intrinsic propensity of the charged residues for β sheet structure (6). The pairs containing a Thr residue showed no particular distribution with regard to ΔΔG, whereas they consistently showed the least interaction energy. As expected for pairing residues with high intrinsic propensities, all the proteins containing pairs of the best β sheet formers were quite stable, and the melting temperature (T_m) ranged between approximately 72° and 80°C (Table 2). From a comparison of Tables 1 and 2, it was evident that stability and side chain interaction energy are not necessarily correlated: Side chain interaction energy and the intrinsic β sheet–forming propensities work in conjunction to make a stable protein.

Our studies present a systematic investigation of a subset of pairwise interactions between amino acid residues in an H-bonded position across two antiparallel β strands. Previously, intrinsic β sheet–forming propensities were the only guidelines available for the design or engineering of a β sheet (*5–7*, *16*). The propensities could provide information only in terms of which residues should be included in the design without consideration of sequence or context. Side chain interactions contribute to β sheet stability on a scale comparable to that of the propensity to form β sheets and other factors stabilizing secondary structure. Thus, it should now be possible to base protein designs on the secondary and tertiary contacts that are unique to β sheet structure.

Table 2. The residue pairs (*12*) are listed in order of decreasing value of ΔΔ*G*, which was calculated relative to β1AA (*14*).

Mutant	ΔΔ*G* (kcal/mol)	T_m(°C)
FY	−2.37	79.5
FF	−2.26	78.3
IY	−2.16	77.8
FI	−1.93	76.3
TT	−1.91	75.3
TY	−1.91	77.4
TW	−1.80	75.1
IF	−1.75	74.9
TF	−1.71	75.2
II	−1.67	75.8
FW	−1.63	75.7
FT	−1.59	75.3
IW	−1.47	73.6
TV	−1.47	73.3
TI	−1.47	74.3
FV	−1.38	73.2
IT	−1.37	72.7
IV	−1.30	72.3
ER	−1.20	70.5
EK	−1.09	70.0

REFERENCES AND NOTES

1. K. T. O'Neil and W. F. DeGrado, *Science* **250**, 646 (1990); P. C. Lyu, M. I. Liff, L. A. Marky, N. R. Kallenbach, *ibid.*, p. 669; S. Padmanabhan, S. Marqusee, T. Ridgeway, T. M. Laue, R. L. Baldwin, *Nature* **344**, 268 (1990); M. Blaber, X.-j. Zhang, B. W. Matthews, *Science* **260**, 1637 (1993).
2. L. Serrano, M. Bycroft, A. R. Fersht, *J. Mol. Biol.* **218**, 465 (1991); S. Padmanabhan and R. L. Baldwin, *ibid.* **241**, 706 (1994); *Protein Sci.* **3**, 1992 (1995).
3. S. Marqusee and R. L. Baldwin, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8898 (1987).
4. A. Chakrabartty, A. J. Doig, R. L. Baldwin, *ibid.* **90**, 11337 (1993); J. S. Richardson and D. C. Richardson, *Science* **240**, 1648 (1988); L. G. Presta and G. D. Rose, *ibid.*, p. 1632.
5. C. A. Kim and J. M. Berg, *Nature* **362**, 267 (1993); D. L. Minor Jr. and P. S. Kim, *ibid.* **367**, 660 (1994).
6. C. K. Smith, J. M. Withka, L. Regan, *Biochemistry* **33**, 5510 (1994).
7. D. L. Minor Jr. and P. S. Kim, *Nature* **371**, 264 (1994); R. C. Garratt, J. M. Thornton, W. R. Taylor, *FEBS Lett.* **280 (1)**, 141 (1991).
8. M. A. Wouters and P. M. G. Curmi, *Proteins Struct. Funct. Genet.* **22**, 119 (1995).
9. G. von Heijne and C. Blomberg, *J. Mol. Biol.* **117**, 821 (1977).
10. S. Lifson and C. Sander, *ibid.* **139**, 627 (1980).
11. F. R. Salemme, *Prog. Biophys. Mol. Biol.* **42**, 95 (1983).
12. The one-letter amino acid code is as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
13. We monitored thermal denaturation by following the ellipticity at 222 nm as a function of temperature on an Aviv circular dichroism spectrophotometer model 62DS at a protein concentration of 0.5 mg/ml in 50 mM sodium acetate, pH 5.2. Experiments were performed in triplicate and the results averaged. The error in calculating ΔG was approximately 5%. Under these conditions, the thermal denaturation transition was reversible, and the calorimetric enthalpy was equal to the van't Hoff enthalpy for the variant β1FI. The melting temperature of this protein was independent of protein concentration, which confirms that no dimerization nor aggregation occurred. All double and single mutants eluted with the same retention time in gel filtration chromatography, which further indicates that the proteins were monomeric. The thermal denaturation curves were analyzed according to a two-state transition model. Linear extrapolation of the pre- and post-transition baselines was performed with the program Kaleidagraph (Synergy Software). The fraction that was native (F_n) at each temperature was calculated with the equation $F_n = (Ø_{obs} - Ø_u)/(Ø_f - Ø_u)$, where $Ø_{obs}$ is the observed ellipticity at 222 nm and $Ø_f$ and $Ø_u$ are the folded and unfolded ellipticities, respectively, derived from the extrapolated base lines. The fraction unfolded (F_u) is given by $F_u = 1 - F_n$. We determined the free energy, Δ*G*, by calculating the enthalpy, Δ*H*, in the transition region using the van't Hoff equation and then using this value to calculate the entropy, Δ*S*, at the T_m. This treatment assumes that within the transition region Δ*H* is independent of temperature. Accordingly, Δ*G* values were calculated at a temperature that is within the transition region for all the mutants (346 K).
14. The ΔΔ*G* values (kcal/mol) at 346 K of the single mutants represent the intrinsic β sheet–forming propensities relative to $\beta1AA_{E42T}$ and were as follows: EA, 0.27; AA, 0.0; IA, −0.07; FA, −0.26; and TA, −0.98. The following ΔΔG values at 346 K were derived from our original intrinsic propensity study at position 53 (*6*) and were corrected to allow for the stabilization contributed by the additional E42T mutation used in this pairwise study: AR, −0.38; AK, −0.40; AV, −0.95; AW, −1.06; AF, −1.09; AT, −1.14; AI, −1.23; AY, −1.47. The additional stabilization (−0.54 kcal/mol) was determined from the difference between the stability of the $\beta1AA_{E42T}$ and $\beta1AA_{E42}$ mutants calculated at 346 K. The ΔG at 346 K for the standard $\beta1AA_{E42T}$ mutant is 1.41 kcal/mol.
15. C. A. Hunter, J. Singh, J. M. Thornton, *J. Mol. Biol.* **218**, 837 (1991).
16. P. Y. Chou and G. D. Fasman, *Biochemistry* **13**, 211 (1973).
17. A. M. Gronenborn *et al.*, *Science* **253**, 657 (1991).

18. We thank C. Sander for bringing his initial statistical studies to our attention; M. Wouters and P. Curmi for sharing their unpublished data with us; J. Sturtevant for performing the differential scanning calorimetry experiments; and G. Schoenhals, T. Ragoczy, and L.R. lab members for their support. Supported by National Science Foundation (NSF) grant MCB-9316863. L.R. is an NSF National Young Investigator and a Dreyfus Teacher-Scholar.

15 May 1995; accepted 21 August 1995