PROTEINS: Structure, Function, and Bioinformatics 57:400–413 (2004)

Large-Scale Prediction of Protein Geometry and Stability Changes for Arbitrary Single Point Mutations

A. J. Bordner[1,2]* and R. A. Abagyan[1,2]
[1]*The Scripps Research Institute, 10550 North Torrey Pines Rd., Mail TPC-28, San Diego, California*
[2]*Molsoft LLC, 3366 North Torrey Pines Court, Suite 300, San Diego, California*

ABSTRACT **We have developed a method to both predict the geometry and the relative stability of point mutants that may be used for arbitrary mutations. The geometry optimization procedure was first tested on a new benchmark of 2141 ordered pairs of X-ray crystal structures of proteins that differ by a single point mutation, the largest data set to date. An empirical energy function, which includes terms representing the energy contributions of the folded and denatured proteins and uses the predicted mutant side chain conformation, was fit to a training set consisting of half of a diverse set of 1816 experimental stability values for single point mutations in 81 different proteins. The data included a substantial number of small to large residue mutations not considered by previous prediction studies. After removing 22 (≈2%) outliers, the stability calculation gave a standard deviation of 1.08 kcal/mol with a correlation coefficient of 0.82. The prediction method was then tested on the remaining half of the experimental data, giving a standard deviation of 1.10 kcal/mol and covariance of 0.66 for 97% of the test set. A regression fit of the energy function to a subset of 137 mutants, for which both native and mutant structures were available, gave a prediction error comparable to that for the complete training set with predicted side chain conformations. We found that about half of the variation is due to conformation-independent residue contributions. Finally, a fit to the experimental stability data using these residue parameters exclusively suggests guidelines for improving protein stability in the absence of detailed structure information. Proteins 2004;57:400–413.**

Key words: empirical potential; Monte Carlo optimization; denatured state; conformation; free energy

INTRODUCTION

The theoretical prediction of the structure and stability of proteins is a fundamental goal in molecular biology. A more tractable version of this problem is to predict changes in structure and stability induced by point mutations. Even this more modest goal has immediate application in computational protein design.[1–7] The energy functions and methods used for these predictions also have applications in protein–protein docking,[8] structure prediction[9] and structure validation,[10] as well as providing insight into the energetic factors contributing to protein folding.

Although free energy simulations have been used to provide accurate predictions of the relative stabilities of point mutants,[11–15] they are presently too computationally intensive to test the large number of mutations studied in protein design applications. Optimization of a physical force field energy is also used to estimate the stability. One study[16] used a simplified energy function with only van der Waals (vdW) and side chain torsion potentials to predict the stabilities of the λ repressor protein for mutations involving only hydrophobic residues. A subsequent work[17] with an improved optimization method also demonstrated better prediction accuracy for continuously flexible side chain angles as compared to discrete side chain angles from a rotamer library.

Another approach is to use statistical potentials derived from geometric and environmental propensities and correlations of residues in X-ray crystal structures. A statistical potential for contacting residues was used in Ota et al.[17] to calculate the stabilities of ribonuclease H1 mutants. Zhou and Zhou[19] introduced an improved reference state for the statistical potential that yielded more accurate stability predictions for 895 mutants and demonstrated that the correlation between predicted and experimental stability is significantly lower for surface residue mutations. Potentials derived from substitution and occurrence frequencies for amino acids in different structural environment classes, such as main chain conformations and solvent accessibilities, have also been used to calculate the stability differences induced by point mutations.[20,21] Finally, statistical potentials that use higher order residue correlations have been applied to stability predictions[22] in order to overcome possible limitations in pairwise residue statistical potentials.[23,24]

Empirical potentials, in which free energy differences are calculated using a combination of physical energy terms, statistical energy terms and structural descriptors with weight factors scaled to fit experimental data, is an approach that will be used in this paper. A weighted sum of

The Supplementary Materials referred to in this article can be found at http://www.interscience.wiley.com/jpages/0887-3585/suppmat/index.html

*Correspondence to: Andrew Bordner, Molsoft LLC, 3366 North Torrey Pines Court, Suite 300, San Diego, CA 92037. Email: bordner@molsoft.com.

Received 27 December 2003; Revised 10 February 2004; Accepted 29 March 2004

Published online 17 June 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20185

physical energy terms, local secondary structure and cavity volume with parameters derived from human lysozyme mutants was used in Funashi et al.[25] to predict stability changes in T4 lysozyme mutants. Only mutants with available X-ray crystal structures were considered in this analysis. Another study in Guerois et al.[26] used a weighted combination of physical energy terms, including van der Waals, solvation, hydrogen bonding, electrostatic, entropic and water bonding terms to predict the stabilities of a large set of 1088 mutants. The mutant structures were modelled by either deleting atoms or changing atom types while inheriting the atomic coordinates of the native structure, followed by hydrogen bond network optimization. This simple modelling procedure limited the applicability of the method to only large to small residue mutations.

We have developed a method to calculate the relative stabilities of mutants, which employs global optimization of a realistic energy function to predict the mutant structure based on the native structure and so allows predictions for arbitrary mutations. The only previous study using a detailed empirical energy function and tested on a data set of comparable size could only be used with large to small residue mutations.[26] Our method uses an empirical energy function that includes a parameter for each amino acid in order to account for the free energy difference of the denatured states, a contribution that is often neglected. We have also collected a data set of 2006 pairs of available X-ray crystal structures of proteins that differ by a single point mutation and used this database to optimize and test the modelling procedure for the mutant structures. A data set of 1816 experimental $\Delta\Delta G$ values was compiled for single point mutations with available X-ray crystal structures of the native proteins. The optimization method was then used to predict the mutant protein conformation based on the native protein structure, and an energy function was fit to a training set consisting of half of the data and tested on the remaining data. A subset of 137 mutants also had X-ray structures for the mutant protein so the same energy function was fit to this data in order to determine the effects of the mutant geometry prediction on calculated $\Delta\Delta G$ values.

We also fit a simple empirical energy function that includes only conformation-independent residue energy parameters to the experimental $\Delta\Delta \boldsymbol{G}$ data (see "Predicting the Stabilization Effect of Mutations Without Structure"). Although this function is less accurate in predicting protein stability changes than the one mentioned above, it may be used to suggest stabilizing mutations in the absence of detailed structure information.

Protein Stability Measurements

The change in stability of a protein after site-directed mutagenesis is quantitatively described by $\Delta\Delta \boldsymbol{G}$ with

$$\Delta\Delta G = \Delta G^F_{\text{mutant}} - \Delta G^F_{\text{native}} \quad (1)$$

and the Gibbs free energy of folding defined by

$$\Delta G^F = G^{\text{folded}} - G^{\text{unfolded}} \quad (2)$$

Reversible denaturation of the protein is usually accomplished by either increasing the temperature or adding urea or guanidinium hydrochloride (GmCl) as a denaturant. In the latter case, a linear dependence of ΔG^F_{meas} on denaturant concentration,

$$\Delta G^F_{\text{meas}} = \Delta G^F_{H_2O} - m\ [\text{denaturant}] \quad (3)$$

is often assumed. The extrapolated value at zero denaturant concentration, $\Delta G^{\text{F}}_{\text{H}_2\text{O}}$, is then used in Eq. (1). The free energy of folding is calculated from the temperature or denaturant concentration at which half of the protein is denatured, assuming a two-state model of unfolding, with no stable intermediates.[27] The most common methods of measuring the denatured fraction are circular dichroism,[28–30] intrinsic fluorescence, and differential scanning calorimetry.[31]

Because of the lack of adequate experimental or theoretical information about the ensemble of denatured states, its contribution to $\Delta \boldsymbol{G}$ is usually either ignored or included implicitly by fitting only properties of the folded state to experimental free energy differences. The characterization of the denatured state may be further complicated by residual local structure.[32–38] Because of these difficulties, most calculation methods do not explicitly include the free energy difference between the unfolded states of the native and mutant proteins in the calculation of $\Delta\Delta \boldsymbol{G}$. One exception is the study by Lee,[39] in which the denatured state was modelled as an extended peptide chain. Here we include an explicit term for the denatured state in the expression for $\Delta\Delta \boldsymbol{G}$ that is based on a simple energetic model in which the contribution of the unfolded states to the free energy difference is represented by $\boldsymbol{E}^{\boldsymbol{U}}_{\boldsymbol{X}}$–$\boldsymbol{E}^{\boldsymbol{U}}_{\boldsymbol{X}'}$ for the mutation $\boldsymbol{X} \rightarrow \boldsymbol{X}'$, where $\boldsymbol{E}^{\boldsymbol{U}}_{\boldsymbol{X}}$ and $\boldsymbol{E}^{\boldsymbol{U}}_{\boldsymbol{X}'}$ are adjustable empirical parameters.

MATERIALS AND METHODS

Mutant Geometry Prediction

Single Mutant Structure Database of 2006 Pairs

We believe that an accurate calculation of $\Delta\Delta \boldsymbol{G}$ in the general case requires an accurate prediction of the mutant protein conformation. In order to study and optimize the method for predicting the mutant conformation using the native conformation, a set of all protein pairs, that differ by a single amino acid substitution and have structures in the Protein Data Bank[40] (PDB) was compiled. The best resolution structures for each protein were used, and only proteins with more than 20 residues were included. Unmatched end residues were ignored, and only X-ray structures with resolutions of less than 3.0 Å were included. In addition, for pairs in which one of the structures had missing residues nearby (within two residues along the chain) any of the residues sampled in the Monte Carlo (MC) simulation were removed, because the absence of neighboring residues may affect the energy function and hence the simulation results. This initial set contained 2006 structure pairs with an average resolution of 1.9 Å.

Selection of Structures for Modelling and Comparison

Simulations were performed starting with both structures in each pair, because the identity of the wild type protein is unimportant for the purpose of studying the conformation prediction accuracy. This effectively doubled the number of predictions to compare to the X-ray structures. The appropriate residue was modified, and biased-probability Monte Carlo simulations[41] (BPMC) were performed, as described in the following section.

In order to compare the MC prediction to the corresponding X-ray structure, the C_α C_O and N atoms of the backbone for all residues within the 4.0 Å region around the mutated residue, defined in the following section, were aligned in order to minimize their root mean square (RMSD) distance. The positions of the amino acid side chains were then compared.

One difficulty of the structural comparison arises when one protein in the pair has undergone large structural changes, due to the binding of a small molecule ligand, the binding of another protein, different crystal contacts or a change in chemical conditions such as pH and not due to the effect of the mutation itself. For example, there are X-ray structures of maltodextrin in both the open conformation, with either no bound ligand[42] or bound β-cyclodextrin[43] as well as a structure in the closed conformation with bound maltose.[44] We also removed those protein pairs in which the RMSD deviation of the backbone atoms of neighboring residues (within 5 Å), after structural alignment, was shown to be greater than 0.5 Å. This is because, using only structural information it is difficult to automatically determine whether a large structural movement is due to the mutation or to other factors. This eliminates only about 12% of the pairs. The RMS backbone deviation of the remaining pairs was as low as 0.21 Å.

Since we considered each protein as an isolated monomer, and modelled them as such, X-ray structures in which nearby protein molecules interact with the mutated residue should be not be used to compare with simulation results. Thus, cases in which the X-ray structure used for comparison had at least one atom of a different protein molecule within 3.0 Å of any atom in the mutated residue were not considered in the analysis. Because molecules that are crystallographic symmetry partners may also strongly interact with the mutated residue and thus affect their conformation, symmetry-related molecules in surrounding unit cells were included in this calculation by applying the appropriate crystallographic symmetry transformations. Structures that had missing atoms within 3 Å of any atom in the mutated residue were also removed. This left 2798 ordered structure pairs that differed by a single point mutation. The mutated residue in the first structure was changed to that in the second one, an MC simulation was performed, and the lowest energy conformation was compared to the second structure, after the backbone superposition procedure described above.

Uncertainties in the atomic positions due to thermal oscillations or disorder within the crystal are reflected in the crystallographic B-factors. The prediction of a unique conformation of a residue with a high average B-factor is expected to be difficult, since presumably there are many similar conformations of similar energy. Therefore all cases in which the mutated residue in the X-ray structure used for comparison had a normalized average B-factor $\boldsymbol{B}_{\mathrm{norm}} = (\boldsymbol{B}_{\mathrm{res}} - B_{\mathrm{ave}})/B_{\mathrm{ave}}$, > 0.1, with B_{ave} the average B-factor for the structure, were also removed from the analysis. There were then 2141 ordered structure pairs remaining for conformation prediction and analysis (see Table I for a summary of the selection criteria). This benchmark is currently the largest mutation geometry benchmark and is provided as Supplementary Material.

TABLE I. Criteria for Selection of 2141 Ordered X-ray Structure Pairs in Single Mutant Structure Database

Resolution < 3.0 Å
Number of residues > 20
Neighboring residue backbone atom RMSD between native and mutant structures < 0.5 Å
No other protein molecule within 3.0 Å of mutated residue (including symmetry-related molecules)
$B_{norm} < 0.1$
No missing atoms within 3.0 Å of mutated residue side chain
No missing side chains within two residues along the main chain from neighboring residues

Structure Prediction Using BPMC Optimization

The X-ray structures were first regularized in order to make a model with idealized covalent geometry, with bond lengths and angles defined by the ECEPP/3[45–47] force field. This procedure consists of constructing a peptide structure with the same amino acid sequence and idealized covalent geometry, defining a quadratic constraint potential between equivalent atoms in the X-ray structure, and iteratively performing local minimization of the constraint potential in torsion angle space while gradually reducing the constraint potential strength to zero. All protein modelling calculations were performed using the ICM 3.0 program.[48]

The mutated residue was changed in the regularized native protein structure, followed by conjugate gradient minimization of the side chain χ angles for all residues with side chain non-hydrogen atoms within 4.0 Å of any side chain non-hydrogen atoms in the the mutated residue. An energy function in internal coordinate (torsion angle) space that combines electrostatic, vdW, hydrogen bonding and torsional terms from the ECEPP/3 force field with solvation and entropic terms was used.[41] The electrostatic term was calculated using the Coulomb energy with a distance dependent dielectric $\varepsilon = 4r$ to roughly include dielectric screening. The 6–12 vdW energy term

$$E_{vw}^{0} = \sum_{j} -\frac{A_{ij}}{R_{ij}^{6}} + \frac{B_{ij}}{R_{ij}^{12}} \tag{4}$$

was smoothed by introducing a cutoff value[49] $E_{vw}^{\max}$ of 7.0 kcal/mol, such that

$$E_{vw} = \begin{cases} E^0_{vw} & \text{if } E^0_{vw} \le 0 \\ \dfrac{E^0_{vw}E^{\max}_{vw}}{E^0_{vw} + E^{\max}_{vw}} & \text{otherwise} \end{cases} \quad (5)$$

A cutoff somewhat reduces the sensitivity of $\boldsymbol{E}_{\mathrm{vw}}$ to the molecular conformation and also speeds up convergence of local optimization of the energy. The solvation term

$$\boldsymbol{E}_{\mathrm{solv}} = \sum_{\boldsymbol{i}} \sigma_{\boldsymbol{i}} \boldsymbol{A}_{\boldsymbol{i}} \quad (6)$$

is a sum of terms proportional to the solvent accessible surface area (SASA) $\boldsymbol{A}_{\boldsymbol{i}}$ of atom type $\boldsymbol{i}$ with solvation parameters σ_i from Wesson and Eisenberg.[50] Finally, the side chain entropic free energy was $-\boldsymbol{TS}$ with $\boldsymbol{T} = 300K$ and entropy ($\boldsymbol{S}$) proportional to the SASA of reference atoms in the residue, $S = S_{\max} \times A/A_{\max}$, with the maximum SASA, $A_{\max}$, calculated with the extended residue between two glycine residues and the reference entropy, $S_{\max}$, calculated using approximate rotamer distributions.[41]

The BPMC method[41] is used to optimize the energy function in internal coordinates. Each iteration of the procedure consists of a random MC move that is biased to sample conformations that are are prevalent in experimental protein structures. If the move is accepted according to a constant temperature Metropolis criterion, it is followed by local minimization. A stack of lowest energy conformations within different conformation space regions is used to guide sampling away from highly visited or unfavorably high energy regions.[51]

If the mutated residue is not alanine, glycine or proline, a BPMC simulation with temperature 700K was then used to sample its χ angles. Each accepted Monte Carlo move was followed by up to 2000 steps of conjugate gradient minimization of the neighboring residues, as described above. A stack of 50 of the lowest energy conformations that differ by at least 15° RMSD in torsion angle space was used to prevent oversampling certain regions of conformational space. The simulation was terminated after 10^5 function calls, and only the lowest energy conformation was used for subsequent calculations. The average computation time for a simulation was 12 min on a 1.3 GHz Athlon processor.

Mutant Stability Prediction

Database of Experimental Stability Changes for 1816 Single Point Mutants

A total of 3793 experimental $\Delta\Delta\boldsymbol{G}$ values of single point mutants for which X-ray crystal structures are available in the PDB and the pH lies in the range $5.0 < \mathrm{pH} < 9.5$ was collected from the PROTHERM database[52] and refs. 26,53–62. Database errors, such as missing or incorrect PDB entry names and $\Delta\Delta\boldsymbol{G}$ values and incorrect residue numbers, were corrected. The pH limits were chosen to be outside the pK_a values for isolated glutamic acid (4.1) and lysine (10.5) side chains; however, no attempts were made to calculate the pK_a shifts or predominant protonation states at the experimental pH value, which are particularly important for histidine and cysteine as well as buried residues that are charged in the denatured state. Data for proteins that form stable multimers, such as ketosteroid isomerase, or membrane proteins, such as FepA, were also not included in the data set. In cases where more than one measurement was available for the same protein and mutation, the most recent one was used. Next, the amino acid sequences for the proteins from the original PDB entries were used to search for other proteins in the PDB with identical sequences in the overlap region of the alignment. The best resolution structures from this search were then used for all subsequent modelling. All data for which there was a break within two residues along the chain in the X-ray structure were eliminated, as in the geometry predictions described in 'Single Mutant Structure Database of 2006 Pairs.' The final data set contained $\Delta\Delta\boldsymbol{G}$ values for 1816 mutants and 81 proteins and is provided as Supplementary Material. Half of these entries (908) were then randomly selected for the training set to be used for parameter fitting, and the remaining half comprised the test set to be used for validation.

TABLE II. Prediction Accuracy of Mutant and Surrounding Residue Conformation for Entire Set of Structure Pairs and Sets in which Mutated Residue has a Normalized B-factor Less Than 0.1 or a Fractional SASA Less Than 0.55

Set	Number of Structures	Mutant Residue RMSD (Å)	Neighbor Residues RMSD (Å)	Fractional SASA
All	2798	0.87	0.72	0.26
$B_{\mathrm{norm}} < 0.1$	2141	0.76	0.66	0.20
Fractional SASA < 0.55	2436	0.84	0.70	0.20

The fractional SASA cutoff value was chosen such that the average fractional SASA was the same as for the set with the B-factor cutoff.

RESULTS AND DISCUSSION

Mutant Geometry Prediction

Analysis of Conformation Prediction Results

The average RMSD values of the side chain atoms of the mutated residue and surrounding residues, whose conformations were sampled in the simulation, are given in Table II. The average fractional SASA is also shown. This quantity is calculated by dividing the SASA for the mutated residue by the SASA of the same residue in an extended conformation and surrounded by glycine residues. The SASA is defined by the center of a sphere of radius 1.4 Å, the vdW radius of a water oxygen atom, touching the vdW surface of the molecule. The averages in Table II show that comparing the predicted geometry only to X-ray crystal structures in which the mutated residue has a low normalized B-factor, $\boldsymbol{B}_{\mathrm{norm}} < 0.1$, significantly improves the agreement between the conformations. Because the B-factors of surface residues are, on average, larger than buried residues, this cut also reduces the average fractional SASA. One might conclude that the improvement in the prediction is due to the fact that there are more buried residues in the set. However, if a cut is

TABLE III. Prediction Accuracy of Mutant and Surrounding Residue Conformation

Set	Surface or Buried	Number of Structures	Mutant Residue RMSD (Å)	Neighbor Residues RMSD (Å)
All	buried	795	0.66	0.56
	surface	1346	0.82	0.71
Large → small (non-AG → AG)	buried	159	-	0.46
	surface	322	-	0.54
Small → large (AG → non-AG)	buried	138	0.89	0.63
	surface	195	0.97	0.74

Data are divided by size of the mutated residues and the surface accessibility of the original residue. Alanine and glycine are classified as small residues, and residues with fractional SASA < 0.05 are denoted as buried.

made by requiring that the fractional SASA be less than 0.55, a value chosen so that the average SASA is the same as that for the B-factor cut set, the RMSD for both the mutated residues and neighbor residues is higher. Therefore the improvement in the agreement between the MC geometry prediction and the X-ray crystal structure may be attributed to the removal of structures for which the mutated residue does not have a well-defined unique conformation.

Next we compare the prediction accuracy for small to large residue mutations and large to small mutations for both buried and surface residues. Predicting conformational changes for small to large residue mutations was expected to be more difficult. The analysis was performed on the set with $\boldsymbol{B}_{\text{norm}} < 0.1$. Mutated residues in the structure compared to MC simulation results that have fractional SASA of less than 0.05 were classified as buried, and the remainder were classified as surface residues. These results are shown in Table III. By this criterion, the set with all mutations has almost twice as many surface exposed residues as buried ones, and, as expected, the prediction results are more accurate for the buried residues. In other words, the structures in the set are biased toward the more difficult to predict surface residue mutations. Also large to small residue mutations are found to be more accurate than small to large ones, based on the neighbor residues RMSD. It is not meaningful to compare the mutated residue RMSD, since alanine and glycine do not have any conformational degrees of freedom that are sampled in the MC calculation. However, the larger conformational space available to large residues combined with inherent errors in the energy function lead to reduced prediction accuracy for the surrounding residues' conformation.

Mutant Stability Prediction

Empirical Energy Function Fit to Stability Data Using Predicted Mutant Conformations

Starting from the X-ray crystal structure of the native protein, the mutated residue side chain was changed and the same BPMC procedure described above for conformation prediction was used. Again, only the lowest energy conformation was used to model the mutant protein.

Linear regression analysis without a constant term was then used to fit the following empirical energy function to the experimental data:

$$\Delta\Delta G_{\text{calc}}(X \rightarrow X') = w_{el}\Delta E_{el} + w_{vwto}(\Delta E_{vw} + \Delta E_{to}) + w_{hb}\Delta E_{hb} + w_{en}\Delta E_{en} + w_{hp}\Delta E_{hp} - E^{U}_{X'} + E^{U}_{X} \quad (7)$$

with

$$\Delta E_i = E_{i,X'} - E_{i,X}, \quad i \in \{el, vw, hb, en, hp, to\} \quad (8)$$

representing the difference in the corresponding energy terms between the predicted mutant structure and native structure and $\boldsymbol{X}$ and $\boldsymbol{X}'$ the mutated residue types in the native and mutant proteins, respectively. The electrostatic ($\boldsymbol{E_{el}}$), van der Waals ($\boldsymbol{E_{vw}}$), hydrogen bonding ($\boldsymbol{E_{hb}}$) and torsional ($\boldsymbol{E_{to}}$) energies were calculated with the ECEPP/3 force field in internal coordinates using the boundary element method[63] with an internal dielectric constant of 4 and solvent dielectric constant of 78.5 for $\boldsymbol{E_{el}}$. The truncated van der Waals potential of eq. (2.5), with $\boldsymbol{E}^{\max}_{\boldsymbol{vw}} = 4.0$ kcal/mol for $\boldsymbol{E_{vw}}$, constant surface tension $\sigma = 12$ cal mol^{-1} A^{-2} hydrophobic term ($\boldsymbol{E_{hp}}$) and entropic term ($\boldsymbol{E_{en}}$) from Abagyan and Totrov,[41] as described in 'Structure Prediction using BPHC Optimization' was employed. $\boldsymbol{E_{en}}$ was calculated using the experimental temperature or 300K if unspecified. Only the interaction energy between the mutated residue and the whole protein molecule was calculated for the pairwise energy terms, $\boldsymbol{E_{el}}$, $\boldsymbol{E_{vw}}$, and $\boldsymbol{E_{hb}}$, and only the contribution from the mutated residue itself was included in $\boldsymbol{E_{en}}$, $\boldsymbol{E_{hp}}$, and $\boldsymbol{E_{to}}$. $\boldsymbol{E^{U}_{X}}$ and $\boldsymbol{E^{U}_{X'}}$ are parameters that represent the unfolded state free energy contributions of the native and mutated residues, respectively. They depend only on residue type, and these 20 parameters are fit, along with the weight parameters $\boldsymbol{w_{el}}$, $\boldsymbol{w_{vwto}}$, $\boldsymbol{w_{hb}}$, $\boldsymbol{w_{en}}$ and $\boldsymbol{w_{hp}}$, in the regression analysis.

The energy function in eq. (7) was calculated using the X-ray crystal structures for the native protein and the conformations predicted by the BPMC simulations for the mutant structures. Linear regression analysis was then used to determine the parameters in eq. (7) that minimize the RMSD between the calculated and experimental $\Delta\Delta\boldsymbol{G}$ values. An arbitrary constant shift of the residue energies, $\boldsymbol{E^{U}_{X}}$, in the empirical energy function was chosen so that the lowest value was zero. The initial fit to the data yielded a correlation coefficient of 0.79. Next, 22 outliers with $|\text{residual}| > 3.1$ kcal/mol ($\approx 2.5\sigma$) were removed, and the linear regression fit was repeated. The resulting fit had a correlation coefficient of 0.82 and a standard deviation of 1.08 kcal/mol. The values and standard errors for these best fit parameters are given in Table IV, a list of outliers is given in Table V, and a plot of experimental versus calculated $\Delta\Delta$G values for the training set is shown in Figure 1(a). These best fit parameters were then used to calculate $\Delta\Delta\boldsymbol{G}$ for the remaining 908 mutants in the test set and compared to the experimental values. The covariance between $\Delta\Delta\boldsymbol{G}_{\text{calc}}$ and $\Delta\Delta\boldsymbol{G}_{\text{exp}}$ was 0.59, with a standard deviation of 1.28 kcal/mol. After removing 26 outliers

TABLE IV. Best Fit Parameters and Standard Errors for the Empirical Potential of Eq. (7)

Parameter	Value
w_{el}	0.00297 ± 0.00221
w_{vwto}	0.105 ± 0.00604
w_{en}	1.33 ± 0.206
w_{hb}	0.153 ± 0.0271
w_{hp}	3.08 ± 0.237
E_A^U	0.958 ± 0.0734
E_C^U	0.859 ± 0.255
E_D^U	0.695 ± 0.161
E_E^U	1.10 ± 0.172
E_F^U	2.44 ± 0.140
E_G^U	0.0 ± 0.104
E_H^U	1.41 ± 0.173
E_I^U	3.06 ± 0.111
E_K^U	1.11 ± 0.198
E_L^U	2.65 ± 0.117
E_M^U	2.52 ± 0.206
E_N^U	1.21 ± 0.172
E_P^U	2.38 ± 0.180
E_Q^U	0.676 ± 0.214
E_R^U	2.19 ± 0.239
E_S^U	0.261 ± 0.136
E_T^U	1.18 ± 0.117
E_V^U	2.34 ± 0.0913
E_W^U	2.27 ± 0.297
E_Y^U	3.16 ± 0.164

The parameters were calculated using a linear regression fit to 886 experimental $\Delta\Delta G$ values in the training set remaining after removing the 22 outliers in Table V.

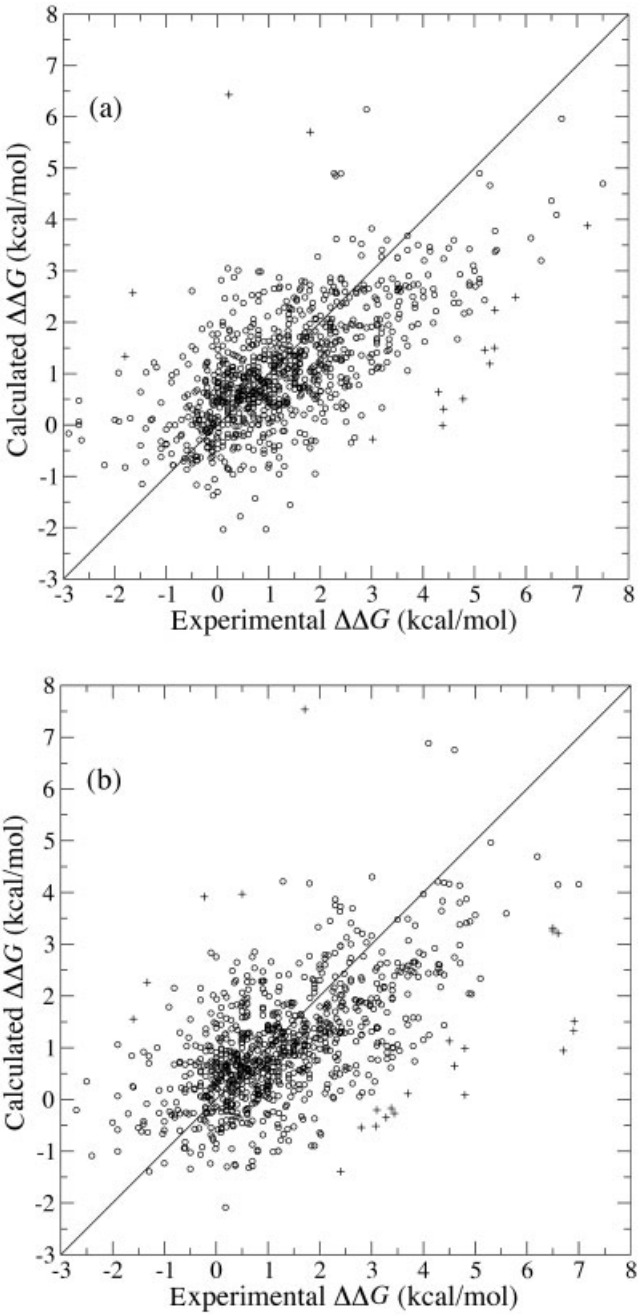

Fig. 1. Experimental $\Delta\Delta G$ versus values calculated using eq. (7) with the parameters in Table IV for (a) the training set of 908 mutants and (b) the test set of 908 mutants. Outliers are shown with + symbols. The experimental $\Delta\Delta G$ values shown in (a) were used to fit the parameters in eq. (7) (correlation coefficient of 0.82). Even though the experimental data in (b) were not used in fitting the empirical energy function, they still had a high covariance (0.66) with the predicted $\Delta\Delta G$ values. A surprising yet correctly predicted point with $(\Delta\Delta G_{exp}, \Delta\Delta G_{calc}) = (10.98, 9.33)$ is not shown in (b).

with |residual| > 3.1 kcal/mol, the covariance increased to 0.66 and the standard deviation was 1.10 kcal/mol. A list of the test set outliers is shown in Table VI and a plot of experimental versus calculated ΔΔG values for the test set is shown in Figure 1(b).

Training and Test Set Outliers

Many of the outliers in Tables V and VI whose $\Delta\Delta\boldsymbol{G}$ values were measured using a chemical denaturant have large differences of 11–151% between the slopes $\boldsymbol{m}$ appearing in eq. (3) for the mutant and the native compounds. Only values of $\boldsymbol{m}$ that differ by more than 10% are shown in Tables V and VI. Theoretical and experimental evidence suggests that such differences in $\boldsymbol{m}$ are due to differing degrees of residual structure in the denatured state ensembles for the mutant and native proteins [32,64–66] which may cause $\Delta\Delta\boldsymbol{G}_{calc}$ to deviate from the experimental value, since the former only accounts for the denatured state implicitly through structure-independent parameters fit to experimental data.

A higher proportion of outliers had native proteins with disulfide bonds (21%) compared to the entire data set (11%). In one case, that of the training set outlier 3LZT, with mutation C94A, the mutated residue forms a disulfide bond in the native protein. Unless reducing agents are added, disulfide bonds in the native structure remain intact in the reversible denaturation experiments used to measure $\Delta\Delta\boldsymbol{G}$. This may cause the average environment of nearby residues in the denatured protein to differ from

TABLE V. Training Set Outliers with |Residual| > 3.1 kcal/mol

PDB Entry	Mutation	$\Delta\Delta G_{exp}$ [Reference]	$\Delta\Delta G_{calc}$ (Residual)	Comments
2TRX	D26A	−3.7 [79]	0.127 (3.83)	disulfide bond present; electrostatic destabilization of native protein through pK_a shift of Asp75
3LZT	C94A	4.78[80]	0.506 (−4.27)	residue forms disulfide bond in native protein; no quantitative data in reference
1FVK	H32Y	−6.8 [82]	−0.335 (6.47)	nearby disulfide bond; possible electrostatic interactions in native or mutant
1ART	C401A	−3.2 [82]	0.630 (3.83)	$m_{mutant} = 3.87$, $m_{native} = 2.56$
5PTI	G12D	4.3 [83]	0.636 (−3.66)	3 disulfide bonds present; possible buried charged residue in mutant
1DDS	G121Y	0.2 [84]	−3.26 (−3.46)	$m_{mutant} = 2.16$, $m_{native} = 1.96$; probable large backbone movement
1FRD	H42R	−4.0 [85]	0.797 (4.80)	$m_{mutant} = 2.7$, $m_{native} = 2.0$
1HMK	T29I	−4.4 [86]	−1.24 (3.16)	4 disulfide bonds present; $m_{mutant} = 2.6$, $m_{native} = 2.2$
1HFZ	Y103P	0.22[87]	6.43 (6.21)	4 disulfide bonds present; possible structural changes
1LVE	V27bL	−1.8 [88]	1.34 (3.14)	disulfide bond present
1SHF	E107L	3.02[55]	−0.280 (−3.30)	$m_{mutant} = 1.83$, $m_{native} = 1.50$
1STN	D95F	5.3 [89]	1.19 (−4.11)	in loop; $m_{mutant} = 3.89$ (lowest value), $m_{native} = 6.60$
1STN	G107A	4.4 [90]	0.309 (−4.09)	
1STN	G55V	1.8 [91]	5.70 (3.90)	$m_{mutant} = 5.89$, $m_{native} = 6.85$
1STN	L108A	5.8 [92]	2.49 (−3.31)	$m_{mutant} = 5.27$, $m_{native} = 6.85$
1STN	L108G	7.2 [92]	3.88 (−3.32)	unstable, ammonium sulfate added
1STN	N100A	5.2 [91]	1.46 (−3.74)	$m_{mutant} = 5.48$, $m_{native} = 6.85$; unstable, ammonium sulfate added
1STN	V74S	5.4 [93]	2.24 (−3.16)	unstable, ammonium sulfate added
1WEJ	P30A	5.39[94]	1.50 (−3.89)	$m_{mutant} = 0.70$, $m_{native} = 1.21$
1HMT	T40Q	4.38[95]	−0.0120 (−4.39)	$m_{mutant} = 5.3$ (lowest value), $m_{native} = 7.6$
4LZM	L66P	12.23[96]	6.88 (−5.35)	$m_{mutant} = 1.708$, $m_{native} = 3.065$
5PTI	L29A	−1.65[97]	2.57 (4.22)	3 disulfide bonds present; $m_{mutant} = 2.77$, $m_{native} = 1.11$

Units for $\Delta\Delta G$ and m are kcal/mol and kcal/(mol M), respectively.

residues of the same type in proteins without disulfide bonds and thus cause large errors in the calculated values of $\Delta\Delta \boldsymbol{G}$, since the denatured state contribution to the empirical energy function depends on residue type only.

Several of the outliers had either confirmed or probable large changes in the backbone conformation of the mutant protein. The predicted value of $\Delta\Delta \boldsymbol{G}$ would then be inaccurate, since the calculation assumes that the backbone conformations of the native and mutant proteins are identical. Also, several of the mutant proteins in the outliers were so unstable that the unfolded fraction was more than 50% even in the absence of denaturant so ammonium sulfate was added as a renaturant. It may be difficult to extrapolate $\Delta\Delta \boldsymbol{G}$ to its value in pure water for these measurements. In addition, some of the outliers had large electrostatic interactions involving the mutated residue and in one case a large $\mathrm{p}\boldsymbol{K_a}$ shift of a nearby side chain. The large errors for these mutants is likely due to the small value of $\boldsymbol{w_{el}}$, as discussed above. Finally two of the test set outliers had $\Delta\Delta \boldsymbol{G}$ values that were calculated using different relations than eq. (3), which may have given different values.

Empirical Energy Function Fit to Stability Data Using Both Native and Mutant X-ray Structures

X-ray crystal structures are available for both the native and mutant structures for 137 (~ 7.5%) of the experimental $\Delta\Delta \boldsymbol{G}$ values. A linear regression fit of eq. (7) was performed using these structures in order to investigate how much errors in the predicted mutant protein conformation affect the accuracy of $\Delta\Delta \boldsymbol{G}_{calc}$. Both the native and mutant structures were first subjected to 10^4 steps of local minimization using the same energy function as for the geometry prediction. Next the energy terms described in the previous section were calculated for the native and mutant structures, and the parameters in eq. (7) were fit to the experimental $\Delta\Delta \boldsymbol{G}$ values using linear regression analysis. $\boldsymbol{E_W^U}$ was not included, since there were not any mutations involving tryptophan in the data set. The correlation coefficient for the fit was $\boldsymbol{r} = 0.78$, and the standard deviation was 1.25 kcal/mol. These fit statistics are essentially equal to those obtained in the fit to the complete training set before removing outliers. This result implies that, in the usual situation, in which X-ray crystal structures for both the native and mutant proteins are unavailable, the prediction accuracy is reasonably good, even with a BPMC simulation prediction of the mutant protein conformation.

Empirical Energy Function Fit to Stability Data for Large Residue to Alanine Mutations

We also performed a fit to the empirical energy function of eq. (7) to only $\Delta\Delta \boldsymbol{G}$ data for large residue (excluding glycine) to alanine mutations. A fit to all such mutants in both the

TABLE VI. Test Set Outliers with |Residual| > 3.1 kcal/mol

PDB Entry	Mutation	$\Delta\Delta G_{exp}$ [Reference]	$\Delta\Delta G_{calc}$ (Residual)	Comments
3LZT	K13D	6.7 [80]	0.942 (−5.76)	4 disulfide bonds present; no quantitative data in reference
1AKE	G85V	2.4 [98]	−1.39 (−3.79)	probable large backbone movement and intermolecular interactions
1B8E	W19Y	6.91 [99]	1.51 (−5.40)	disulfide bond present; nonstandard calculation of ΔG^{H_2O}
1A2P	D75N	4.8 [100]	0.0845 (−4.72)	buried salt bridge (Asp75-Arg83) in native; $m_{mutant} = 1.69$, $m_{native} = 1.93$
1A2P	F7L	4.6 [101]	0.646 (−3.95)	
1A2P	T6P	3.08[102]	−0.516 (−3.60)	
3CHY	D57A	−3.3 [103]	0.879 (4.18)	buried salt bridge Asp57-Lys109 in native; m unknown
1FMK	Y136A	−0.23 [58]	3.92 (4.15)	$\Delta\Delta G$ assumed independent of [GuHCl]
1LW9	M102K	6.9 [104]	1.32 (−5.58)	buried charged residue Lys102 in mutant
1PUC	P90G	−1.33[105]	2.26 (3.59)	probably large structural change
1QLP	F51C	−3.01[106]	2.57 (5.58)	2 disulfide bonds present; native and mutant may have residual structure; m unknown
1SHF	E107P	1.71 [55]	7.53 (5.82)	$m_{mutant} = 1.77$, $m_{native} = 1.50$
1SHF	E107R	3.38 [55]	−0.173 (−3.55)	$m_{mutant} = 1.88$, $m_{native} = 1.50$
1STN	A132V	4.8 [91]	0.985 (−3.82)	$m_{mutant} = 5.89$, $m_{native} = 6.85$; unstable, ammonium sulfate added
1STN	A60V	2.8 [91]	−0.542 (−3.34)	$m_{mutant} = 6.03$, $m_{native} = 6.85$
1STN	G96V	3.7 [91]	0.114 (−3.59)	
1STN	L36I	3.1 [107]	−0.202 (3.30)	
1STN	L38G	0.5 [90]	3.97 (3.47)	
1STN	P117G	−1.6 [108]	1.55 (3.15)	
1STN	V104G	6.5 [92]	3.32 (−3.18)	$m_{mutant} = 5.62$, $m_{native} = 6.85$
1STN	V74G	6.6 [92]	3.21 (−3.39)	
1STN	Y93A	6.5 [92]	3.26 (−3.24)	
1STN	Y93L	4.5 [93]	1.13 (−3.37)	unstable, ammonium sulfate added
1YCC	C102T	−4.09[109]	0.0799 (4.17)	$m_{mutant} = 1.79$, $m_{native} = 0.908$
1HMT	F16Y	3.44 [95]	−0.267 (−3.71)	$m_{mutant} = 1.51$, $m_{native} = 1.82$
1HMT	T40E	3.28 [95]	−0.350 (−3.63)	$m_{mutant} = 1.51$, $m_{native} = 1.82$

Units for $\Delta\Delta G$ and m are kcal/mol and kcal/(mol M), respectively.

training and test sets (635 values) yielded a correlation coefficient of 0.85, which increased to 0.88 with a standard deviation of 0.92 kcal/mol after removing 17 outliers with |residual| > 2.7 kcal/mol (~ 2.5σ). A plot of the experimental versus calculated $\Delta\Delta G$ values is shown in Figure 2.

The quality of this fit is significantly better than for the fit to the training set data without outliers. This cannot be explained by differences in the average relative SASA of the mutated residue, since it is approximately the same for the data with mutations to alanine as for the complete data (0.28). One possible explanation for the increased prediction accuracy is the reduced error in the coordinate-dependent energy terms for the small alanine side chain. As seen in the last section, even small errors in the X-ray crystal structure atomic coordinates apparently leads to errors in the energy function comparable to those for the BPMC-predicted structures. However, there is no torsional freedom, beyond a methyl group rotation, since alanine and electrostatic and hydrogen bonding interactions are essentially absent, thus reducing the error in these terms.

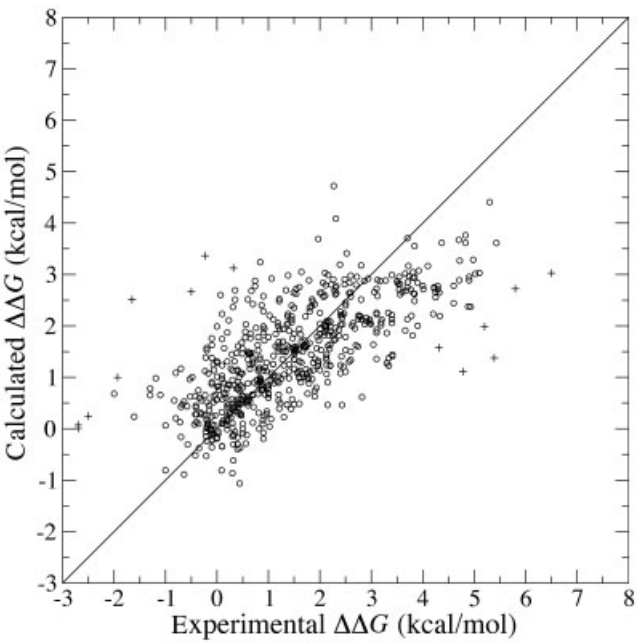

Fig. 2. Experimental $\Delta\Delta G$ versus values calculated using eq. (7) for 625 non-AG → A mutants. Outliers are shown with + symbols. The fit correlation coefficient was 0.88 after removing outliers.

Energy Component Weights in the Empirical Energy Function

The values and uncertainties of the best fit parameters in Table IV may be explained in terms of physical energy

components and uncertainties in their values. The small values of w_{el} and w_{vwto}, and to a lesser extent w_{hb} are likely due to their sensitivity to small conformational changes and the absence of Boltzmann-weighted conformational averaging. The electrostatic weight term, w_{el}, is so small compared to its standard error that it does not appreciably contribute to the fit. The neglect of protonation equilibria for ionizable residues and the simple electrostatics model also contribute to this error. However, the electrostatic energy appears to make a small but significant contribution to $\Delta\Delta G$ in the fit to data using mutant protein structures described previously. In that case, $w_{el} = 0.061$, with a T-statistic p-value of only 6.2×10^{-7}, indicating that this value is significant. This is confirmed by the fact that removing the electrostatics term causes the correlation coefficient to decrease to 0.71 and the standard deviation to increase to 1.40 kcal/mol. Even though there is a cutoff on pairwise vdW energies, E_{vw} is large for many native and mutant conformations, so this term's weight must be decreased in order to reduce the large errors in $\Delta\Delta G_{\text{calc}}$ that would result with a larger weight. The weight w_{hp} effectively redefines the surface tension parameter to be 37 kcal mol^{-1} Å^{-1}.

Effects of Varying Experimental Conditions

Any differences in the experimental conditions for measuring $\Delta\Delta G$, such as pH, temperature, and denaturation method, may affect the measured free energy values. The predominant charges of ionizable residues and hence the electrostatic contribution to $\Delta\Delta G$ depends on the pH. This is unaccounted for in the energy function of eq. (7), since residue charges are fixed at their value for an isolated amino acid at pH = 7.0. Although the side chain entropic term E_{en} is evaluated at the experimental temperature, the surface tension in E_{hp} is assumed to be temperature independent. Including temperature dependence of the microscopic surface tension would probably not significantly improve the accuracy of this simple model for hydrophobic free energy. Evidence from several studies[67–69] also indicates that free energy of folding measurements may depend on whether thermal, urea, or GmCl denaturation is used. This may be due to alteration of the thermodynamic properties of the folded and denatured states in the presence of high denaturant concentrations, which affects the folding free energy extrapolated to zero concentration using eq. (3). For example, GmCl, but not urea, has a strong charge screening effect.

Contribution of Structure-Dependent Parameters to the Stability Change Prediction

It is interesting to examine the relative contributions of the residue energy parameters $E^U_{X'}$, which are independent of the folded protein structures, and the remaining structure dependent terms to the total $\Delta\Delta G_{\text{calc}}$. If the structure-dependent terms were unimportant, then the stability changes could be accurately predicted without any structural information. The RMS ratio of the contribution of the structure-independent terms to the contribution of the structure-dependent terms for the training set fit was 1.09, indicating that each set of terms contributes approximately equally. A fit to the training set without outliers, using only the residue energy parameters, gave a considerably lower correlation coefficient of 0.61 and a higher standard deviation of 1.40 kcal/mol. This further confirms the importance of including physical energy terms and quantities depending on the mutated residue's structural environment in the empirical energy function in order to obtain an accurate stability change prediction.

Predicting the Stabilization Effect of Mutations Without Structure

It is useful to also predict protein stability changes even when atomic level structural information is unavailable for the protein under study. A simple energy function, $\Delta\Delta G = E_{X'} - E_X$ for the mutation $X \rightarrow X'$, was fit to the entire data set excluding outliers (1768 values) for this purpose. This data set was then divided into those with buried mutated residues (native residue fractional SASA $<$ 0.05; 399 values) and the remainder with surface exposed mutated residues. The results of fitting the residue energy parameters E_X to these data sets are given in Table VII. A graph of the best fit parameters for the buried and surface mutated residues is also shown in Figure 3.

The correlation coefficients (standard deviations) were 0.71 (1.21 kcal/mol), 0.55 (1.15 kcal/mol), and 0.57 (1.18 kcal/mol) for the buried residue, surface residue and combined sets, respectively. As expected, these fits have larger errors than those using the empirical energy function of eq. (7) that includes conformation dependent terms. Because $\Delta\Delta G$ values are, on the average, of lower magnitude for surface residues, the standard deviation of the fit to surface residue data is smaller than that for the buried residue data, even though the correlation coefficient is lower. The standard errors in energy parameters for hydrophilic residues such as R, K, and Q are large in the fit to the buried residue set because of their relative scarcity in protein interiors (only 2, 3, and 4 $\Delta\Delta G$ values, respectively).

The stability change of a point mutation may be predicted using the parameters in Table VII when no detailed structural information is available. Of course, if an X-ray crystal structure is available, the empirical energy function of eq. (7) will yield a more accurate prediction. The relative values of the residue parameters are qualitatively similar for buried and surface residues, except for isoleucine, leucine, and methionine, which are more stabilizing on the surface. This is probably due to steric constraints on buried residues with large side chains. Residues with small side chains, glycine, serine, and alanine, are the most destabilizing. Comparison with the E^U parameters in Table IV reveals that the destabilizing effect of alanine depends on its structural context since E^U_A is considerably higher than E^U_G and E^U_S. The most stabilizing amino acids are tyrosine, isoleucine and leucine. This is in agreement with their high frequency of occurrence in β sheets[70]. This same paper also found glycine, the most destabilizing residue in our study, to be one of the most destabilizing residues in β-sheets. Stabilization of the denatured state

TABLE VII. Structure Independent Residue Stability Parameters and Their Standard Errors From a Fit to $\Delta\Delta G$ Data Without Conformation Dependent Energy Terms

	Residues		
Amino Acid	buried	surface	all
A	4.28 ± 0.112	1.75 ± 0.0455	2.21 ± 0.0439
C	3.17 ± 0.426	1.52 ± 0.277	1.75 ± 0.235
D	3.75 ± 0.462	1.26 ± 0.140	1.67 ± 0.142
E	1.79 ± 0.682	0.647 ± 0.115	0.962 ± 0.124
F	1.25 ± 0.294	0.711 ± 0.120	0.862 ± 0.115
G	5.28 ± 0.250	2.15 ± 0.0762	2.62 ± 0.0774
H	3.66 ± 0.766	0.972 ± 0.162	1.38 ± 0.170
I	1.58 ± 0.172	0.0616 ± 0.132	0.251 ± 0.107
K	2.27 ± 1.07	1.06 ± 0.107	1.44 ± 0.117
L	1.48 ± 0.184	0.160 ± 0.129	0.300 ± 0.107
M	2.54 ± 0.488	0.131 ± 0.247	0.670 ± 0.227
N	2.80 ± 0.631	1.12 ± 0.161	1.47 ± 0.167
P	2.40 ± 0.505	1.33 ± 0.167	1.58 ± 0.167
Q	3.15 ± 0.956	1.27 ± 0.152	1.63 ± 0.163
R	1.56 ± 1.30	0.512 ± 0.163	0.867 ± 0.175
S	3.73 ± 0.264	1.78 ± 0.140	2.18 ± 0.127
T	3.23 ± 0.239	0.862 ± 0.107	1.34 ± 0.101
V	2.10 ± 0.124	0.680 ± 0.102	0.800 ± 0.0812
W	0.738 ± 0.610	0.662 ± 0.296	0.697 ± 0.274
Y	0.0 ± 0.326	0.0 ± 0.139	0.0 ± 0.132

Parameters were fit to three data sets: all data without outliers (1768 values), data with buried mutated residues (399 values) defined by fractional SASA < 0.05, and data with surface mutated residues (1369 values). The correlation coefficient for the fit to all $\Delta\Delta G$ data, 0.57, is worse than that presented in Table IV, 0.82. The arbitrary constant shift of the parameters is chosen so that the lowest value is zero.

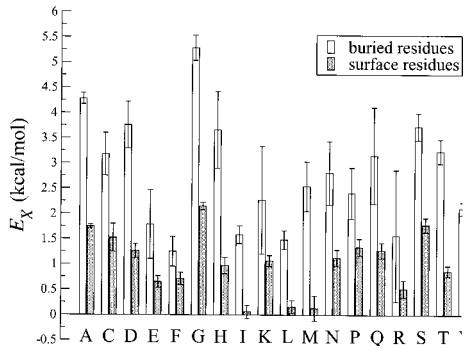

Fig. 3. Residue stability parameters, E_X, for buried and surface mutated residues. Parameter values are shown in Table VII. The change in the folding free energy, $\Delta\Delta G_{X\rightarrow X'}$, can be calculated as $E_{X'} - E_X$.

from increased backbone solvation[71] and larger backbone entropy for glycine as compared to other residues[72,73] may explain its destabilizing effect.

Zhou and Zhou[74] describe a similar least squares fit of conformation-independent residue energy parameters to 1023 experimental $\Delta\Delta G$ values, but only for large to small residue mutations. Although the seven most destabilizing residue energy parameters agree in rank with our results, there are considerable differences in the order of the most stabilizing residue parameters. For example, the most stabilizing residue in Zhou and Zhou[74] is tryptophan, whereas it is the fifth most stabilizing residue in our fit. This difference may be due to the large variance in $\boldsymbol{E_W}$ because of the small number of $\Delta\Delta G$ values for this residue. Also, the most stabilizing residue from our analysis, tyrosine, is only the fifth most stabilizing in their study, and the second most stabilizing residue in their study, phenylalanine, is only the seventh most stabilizing residue in our fit. The magnitudes of residue energy parameters in their paper are also generally larger than for our parameters. As the values in Table VII demonstrate, residue energy parameters are larger in magnitude for buried residues, so this difference may be caused by a larger bias in their set towards buried residues, perhaps because of the neglect of small to large mutations. Zhou and Zhou[74] also demonstrated a strong correlation between residue energies and the corresponding octanol-water transfer free energies for hydrophobic but not hydrophilic residues. We also find a strong correlation for hydrophobic residues ($\boldsymbol{r} = 0.96$) and effectively no correlation for hydrophilic residues ($\boldsymbol{r} = 0.048$). Furthermore, the correlation is slightly higher for buried residues ($\boldsymbol{r} = 0.99$) than for surface residues ($\boldsymbol{r} = 0.96$), and the magnitude of the slope is closer to one for buried residues (0.77) than for surface residues (0.37). Because surface residues are not completely desolvated, the contribution of the solvation free energy should be smaller, and consequently the regression slope should be lower. Almost complete solvation of the mutated residue in the denatured state, combined with the relatively homogeneous environment of the protein interior for a hydrophobic residue, after thermal averaging, may explain the high correlation of buried

hydrophobic residue energies with octanol-water transfer free energies. On the contrary, the interior environment for a buried hydrophilic residue is likely heterogeneous, even after thermal averaging, due to conformation-dependent short-range electrostatic and hydrogen bonding interactions with nearby atoms.

The frequencies of residue substitutions in thermophile proteins versus their mesophile homologues is also correlated with residue stability differences.[75–78] This is because evolutionary selection is thought to favor stabilizing residue substitutions in order for thermophile proteins to function at higher temperatures than their mesophile counterparts. The most common substitution in mesophiles compared to thermophiles,[77] K → R, supports this, since it is also stabilizing according to the parameters in Table IV. It is interesting that McDonald et al.[77] also found that serine, one of the most destabilizing residues found in our study, had the lowest substitution asymmetry. Since both lysine and arginine are charged residues that usually occur on protein surfaces, the effects of structural context may be small, thus making K → R substitutions a good choice for engineering a more stable protein.

Future Directions

A straightforward extension of this method would be to apply it to calculating the binding free energy of protein–protein complexes, as was done in Zhou and Zhou[74] and Guerois et al.[26] It would also be interesting to use our method to calculate the stability changes due to multiple point mutations, particularly for neighboring mutation sites in which the residues strongly interact and an all-atom representation of the surrounding side chains is important. However, as the number of mutation sites increases, the prediction error would increase correspondingly. Although we validated the method using only monomeric proteins, it could also be applied to multimeric proteins by using the structures of the native protein complexes.

A number of improvements to the method are possible to increase both its speed and accuracy. First, rather than using a fixed number of MC steps for each simulation, the relation between the number of free torsion variables and the number of MC steps necessary to achieve convergence could be estimated and used to reduce the average computational time. Also p$\boldsymbol{K_a}$ shifts of ionizable side chains could be calculated and the contribution of protonation equilibria included in the free energy calculation. This, along with conformational averaging, may increase the accuracy of the electrostatic and vdW terms and thus give a more accurate $\Delta\Delta\boldsymbol{G}_{\mathrm{calc}}$. In addition, other terms representing, for example, secondary structure type or cavity size, could be added to the energy function to improve its accuracy, although care must be taken to maintain the manifest symmetry of the empirical energy function under exchange of the native and mutant residues. Finally, allowing limited backbone flexibility in certain regions, such as in loops, may improve the mutant conformation prediction.

CONCLUSIONS

In this paper, we (1) presented an algorithm to predict mutant geometry, (2) showed how use this geometry for improved stability prediction and (3) derived the rules to predict the stabilization effects of mutations without any structures.

Comparison of predicted structures for 2141 single point mutants to X-ray crystal structures demonstrated that short BPMC simulations yield reasonably accurate conformations. As expected, predicted geometry was less accurate for small to large mutations, surface residues and residues with high B-factors. The prediction of the mutant geometry is important, since an X-ray crystal structure is usually unavailable, and we found that about half of the contribution to the predicted $\Delta\Delta\boldsymbol{G}_{\mathrm{calc}}$ is from structure-dependent energy terms.

An empirical energy function containing both physical energy terms for the folded state conformation and residue energy terms for the unfolded state was fit to half of a large set of 1816 experimental $\Delta\Delta\boldsymbol{G}_{\mathrm{calc}}$ values using the X-ray crystal structure of the native protein and the predicted structure of the mutant protein. Even with the inclusion of the more difficult small to large mutation data, the prediction accuracy was approximately 1 kcal/mol. Furthermore predictions for the remaining half of the data using this energy function demonstrated its transferability. Most outliers had either large differences in the slope, m, of eq. (3), attributable to differences in the native and mutant protein denatured states, disulfide bonds that constrain the denatured state conformations, or large backbone conformational changes. The strongly conformation-dependent electrostatic energy term, and to a lesser extent the vdW and hydrogen bonding terms, had small weights, probably due to the large variance in their predicted values.

A simple, conformation-independent energy function consisting of residue energies was also fit to the entire set of experimental $\Delta\Delta\boldsymbol{G}_{\mathrm{calc}}$ values without the outliers. Although the fit accuracy is lower than for the complete empirical energy function, these values should be useful in suggesting stabilizing mutations when no structure is available. Residues with small side chains, glycine, alanine and serine, were the most destabilizing, and tyrosine, isoleucine, and leucine were the most stabilizing. We also found that the residue energies for hydrophobic residues are strongly correlated with octanol-water transfer free energies, whereas the residue energies for hydrophilic residues are not, in accordance with the results of Zhou and Zhou[74]. However, the dependence is weaker (i.e. the magnitude of the slope smaller) for surface residues. The residue energies also predict that the residue change K → R is stabilizing, in agreement with its high frequency of occurrence in aligned mesophile and thermophile protein homologues.

ACKNOWLEDGMENTS

We would like to thank M. Totrov for assistance with ICM and V. Katritch for useful discussions.

REFERENCES

1. Dahiyat BI. In silico design for protein stabilization. Curr Opin Biotech 1999;10:387–390.

2. DeGrado WF. De novo design and structural characterization of proteins and metalloproteins. Ann Rev Biochem 1999;68:779–819.
3. Street AG, Mayo SL. Computational protein design. Struct Fold Des 1999;7:R105–R109.
4. Saven J. Combinatorial protein design. Curr Opin Struct Biol 2002;12:453–458.
5. Mendes J, Guerois R, Serrano L. Energy estimation in protein design. Curr Opin Struct Biol 2002;12:441–446.
6. Bolon DN, Marcus JS, Ross SA, Mayo SL. Prudent modeling of core polar residues in computational protein design. J Mol Biol 2003;329:611–622.
7. Looger LL, Dwyer MA, Smith JJ, Hellinga HW. Computational design of receptor and sensor proteins with novel functions. Nature 2003;423:185–190.
8. Vajda S, Sippl M, Novotny J. Empirical potentials and functions for protein folding and binding. Curr Opin Struct Biol 1997;7:222–228.
9. Lazaridis T, Karplus M. Effective energy functions for protein structure prediction. Curr Opin Struct Biol 2000;10:139–145.
10. Lu H, Skolnick J. A distance-dependent atomic knowledge-based potential for improved protein structure selection. Prot Str Funct Gen 2001;44:223–232.
11. Bash PA, Singh UC, Langridge R, Kollman PA. Free-energy calculations by computer simulation. Science 1987;256:564–568.
12. Dang LX, Merz KM, Kollman PA. Free-energy calculations on protein stability: Thr-157→Val-157 mutation of T4 lysozyme. J Am Chem Soc 1989;111:8505–8508.
13. Tidor B, Karplus M. Simulation analysis of the stability mutant R96H of T4 lysozyme. Biochemistry 1991;30:3217–3228.
14. Prévost M, Wodak SJ, Tidor B, Karplus M. Contribution of the hydrophobic effect to protein stability: analysis based on simulations of the Ile-96-Ala mutation in barnase. J Am Chem Soc 1991;88:10880–10884.
15. Sneddon SF, Tobias DJ. The role of packing interactions in stabilizing folded proteins. Biochemistry 1992;31:2842–2846.
16. Lee C, Levitt M. Accurate prediction of the stability and activity effects of site- directed mutagenesis on a protein core. Nature 1991;352:448–451.
17. Lee C. Testing homology modeling on mutant proteins: predicting structural and thermodynamic effects in the Ala98-Val mutants of T4 lysozyme. Fold Des 1995;1:1–12.
18. Ota M, Kanaya S, Nishikawa K. Desktop analysis of the structural stability of various point mutations introduced into ribonuclease H. J Mol Biol 1995;248:733–738.
19. Zhou H, Zhou Y. Distance-scaled, finite ideal-gas reference state improves structure- derived potentials of mean force for structure selection and stability prediction. Prot Sci 2002;11:2714–2726.
20. Topham CM, Srinivasan N, Blundell TL. Prediction of the stability of protein mutants based on structural environment-dependent amino acid substitution and propensity tables. Prot Eng 1997;10:7–21.
21. Gilis D, Rooman M. Prediction of stability changes upon single-site mutations using database-derived potentials. Theor Chem Acc 1999;101:46–50.
22. Carter CW, LeFebvre BC, Cammer SA, Torpsha A, Edgell MH. Four-body potentials reveal protein-specific correlations to stability changes caused by hydrophobic core mutations. J Mol Biol 2001;311:625–638.
23. Betancourt MR, Thirumalai D. Pair potentials for protein folding: choice of reference states and sensitivity of predicted native states to variations in the interaction schemes. Prot Sci 1999;8:361–369.
24. Thomas PD, Dill KA. Statistical potentials extracted from protein structures: how accurate are they? J Mol Biol 1996;257:457–469.
25. Funahashi J, Takano K, Yutani K. Are the parameters of various stabilization factors estimated from mutant human lysozymes compatible with other proteins? Prot Eng 2001;14:127–134.
26. Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. J Mol Biol 2002;320:369–387.
27. Pace CN. Determination and analysis of urea and guanidine hydrochloride denaturation curves. Methods Enzymol 1986;131:266–280.
28. Woody RW. Aromatic side-chain contributions to far ultraviolet circular-dichroism of peptides and proteins. Biopolymers 1978;17:1451–1467.
29. Vuilleumier S, Sancho J, Loewenthal R, Fersht AR. Circular-dichroism studies of barnase and its mutants: characterization of the contribution of aromatic side-chains. Biochemistry 1993;32:10303–10313.
30. Snatzke G. Circular dichroism: An introduction. Nakonishi K, Berova N, Woody RW, eds. New York: VCH, 1994: p. 1–38.
31. Privalov PL, Potekhin SA. Scanning microcalorimetry in studying temperature-induced changes in proteins. Methods Enzymol 1986;131:4.
32. Dill KA, Shortle D. Denatured states of proteins. Ann Rev Biochem 1991;60:795–825.
33. Evans PA, Topping KD, Woolfson DN, Dobson CM. Hydrophobic clustering in nonnative states of a protein - interpretation of chemical-shifts in NMR-spectra of denatured states of lysozyme. Prot Struct Funct Gen 1991;9:248–266.
34. Neri D, Billeter M, Wider G, Wuthrich K. NMR determination of residual structure in a urea-denatured protein, the 434-repressor. Science 1992;257:1559–1563.
35. Pace CN, Laurents DV, Erickson RE. Urea denaturation of barnase: pH dependence and characterization of the unfolded state. Biochemistry 1992;31:2728–2734.
36. Bond CJ, Wong K, Clarke J, Fersht AR, Daggett V. Characterization of residual structure in the thermally denatured state of barnase by simulation and experiment: description of the folding pathway. Proc Natl Acad Sci USA 1997;94:13409–13413.
37. Nöting B, Golbik R, Soler-González AS, Fersht AR. Circular dichroism of denatured barstar suggests residual structure. Biochemistry 1997;36:9899–9905.
38. Yi Q, Scalley-Kim ML, Alm EJ, Baker D. NMR characterization of residual structure in the denatured state of protein L. J Mol Biol 2000;299:2000.
39. Lee C. Predicting protein mutant energetics by self-consistent ensemble optimization. J Mol Biol 1994;236:918–939.
40. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
41. Abagyan RA, Totrov MM. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. J Mol Biol 1994;235:983–1002.
42. Sharff AJ, Rodseth LE, Spurlino JC, Quiocho FA. Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport and chemotaxis. Biochemistry 1992;31:10657–10663.
43. Sharff AJ, Rodseth LE, Quiocho FA. Refined 1.8-Å structure reveals the mode of binding of beta-cyclodextrin to the maltodextrin binding protein. Biochemistry 1993; 32:10553–10559.
44. Spurlino JC, Lu G, Quiocho FA. The 2.3-Å resolution structure of the maltose- or maltodextrin- binding protein, a primary receptor of bacterial active transport and chemotaxis. J Biol Chem 1991;266:5202–5219.
45. Momany FA, McGuire RF, Burgess AW, Scheraga HA. Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids. J Phys Chem 1975;79:2361–2381.
46. Némethy George N, Pottle Marcia S, Scheraga HA. Energy parameters in polypeptides. 9. Updating of geometrical parameters, nonbonded interactions and hydrogen bond interactions for the naturally occurring amino acids. J Phys Chem 1983;87:1883– 1887.
47. Némethy GN, Gibson KD, Palmer KA, Yoon CN, Pater- lini G, Zagari A, Rumsey S, Scheraga HA. Energy parameters in polypeptides. 10. Improved geometrical parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with application to proline-containing peptides. J Phys Chem 1992;96:6472–6484.
48. Molsoft LLC. ICM software manual. Version 3.0. 2002.
49. Fernandez-Recio J, Totrov M, Abagyan RA. Soft protein-protein docking in internal coordinates. Protein Sci 2002;11:280–91.
50. Wesson L, Eisenberg D. Atomic solvation parameters applied to molecular dynamics of proteins in solution. Protein Science 1992;1:227–235.
51. Abagyan RA, Argos P. Optimal protocol and trajectory visualization for conformational searches of peptides and proteins. J Mol Biol 1992;225:519–532.
52. Gromiha MM, An JH, Kono H, Oobatake M, Uedaira H, Prabaka-

ran P, Sarai A. ProTherm, version 2.0: thermodynamic database for proteins and mutants. Nuc Acids Res 2000;28:283–285.
53. Itzhaki IS, Otzen DE, Fersht AR. The structure of the transition-state for folding of chymotrypsin inhibitor-2 analyzed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein-folding. J Mol Biol 1995;254:260–288.
54. Lopez-Hernandez E, Serrano L. Structure of the transition state for folding of the 129 aa protein CheY resembles that of a smaller protein CI-2. Fold Des 1996;1:43–55.
55. Maxwell KL, Davidson AR. Mutagenesis of a buried polar interaction in an SH3 domain: sequence conservation provides the best prediction of stability effects. Bio- chemistry 1998;37: 16172–16182.
56. Villegas V, Martinez JC, Aviles FX, Serrano L. Structure of the transition state in the folding process of human procarboxypeptidase A2 activation domain. J Mol Biol 1998;283:1027–1036.
57. Fulton KF, Main ERG, Daggett V, Jackson SE. Mapping the interactions present in the transition state for unfolding/folding of FKBP12. J Mol Biol 1999;291:445–461.
58. Riddle DS, Grantcharova VP, Santiago JV, Alm E, Ruczinski I, Baker D. Experiment and theory highlight role of native state topology in SH3 folding. Nature Struct Biol 1999;6:1016–1024.
59. Guerois R, Serrano L. The SH3-fold family: experimental evidence and prediction of variations in the folding pathway. J Mol Biol 2000;304:967–982.
60. Hamill SJ, Steward A, Clarke J. The folding of an immunoglobulin-like Greek key protein is defined by a common-core nucleus and regions constrained by topology. J Mol Biol 2000;297:165–178.
61. Kim DE, Fisher C, Baker D. A breakdown of symmetry in the folding transition state of protein L. J Mol Biol 2000;298:971–984.
62. McCallister EI, Alm E, Baker D. Critical role of beta-hairpin formation in protein G folding. Nature Struct Biol 2000;7:669–673.
63. Totrov M, Abagyan R. Rapid boundary element solvation electrostatics calculations in folding simulations: successful folding of a 23-residue peptide. Biopolymers 2001; 60:124–133.
64. Tanford C. Protein denaturation. Adv Protein Chem 1970;24:1–95.
65. Schellman JA. Solvent denaturation. Biopolymers 1978;17:1305–1322.
66. Pace CN, Laurents DV, Thomson JA. pH-dependence of the urea and guanidine hydrochloride denaturation of ribonuclease A and ribonuclease T1. Biochemistry 1990; 29:2964–2572.
67. Monera OD, Kay CM, Hodges RS. Protein denaturation with guanidine hydrochloride or urea provides a different estimate of stability depending on the contributions of electrostatic interactions. Protein Sci 1994;3:1984–1991.
68. Ibarra-Molero B, Loladze VV, Makhatadze GI, Sanchez-Ruiz JM. Thermal versus guanidine-induced unfolding of ubiquitin. an analysis in terms of the contributions from charge-charge interactions to protein stability. Biochemistry 1999;38:8138–8149.
69. Phelan P, Gorfe AA, Jelesarov I, Marti DN, Warwicker J, Bosshard HR. Salt bridges destabilize a leucine zipper designed for maximized ion pairing between helices. Biochemistry 2002;41: 2998–3008.
70. Smith CK, Withka JM, Regan L. A thermodynamic scale for the beta-sheet forming tendencies of the amino-acids. Biochemistry 1994;33:5510–5517.
71. Avbelj F, Baldwin RL. Role of backbone solvation in determining thermodynamic beta propensities of the amino acids. Proc Natl Acad Sci USA 2002;99:1309–1313.
72. D'Aquino JA, Gomez J, Hilser VJ, Lee KH, Amzel LM, Freire E. The magnitude of the backbone conformational entropy change in protein folding. Proteins 1996; 25:143–156.
73. Street AG, Mayo SL. Intrinsic beta-sheet propensities result from van der Waals interactions between side chains and the local backbone. Proc Natl Acad Sci USA 1999;96:9074–9076.
74. Zhou HY, Zhou YQ. Stability scale and atomic solvation parameters extracted from 1023 mutation experiments. Prot Str Funct Gen 2002;49:483–492.
75. Argos P, Rossman MG, Grau UM, Zuber H, Frank G, Tratschin JD. Thermal stability and protein structure. Biochemistry 1979; 18:5698–5703.
76. Menendez-Arias L, Argos P. Engineering protein thermal stability. Sequence statistics point to residue substitutions in alpha-helices. J Mol Biol 1989;206:397–406.
77. McDonald JH, Grasso AM, Rejto LK. Patterns of temperature adaptation in proteins from methanococcus and bacillus. Mol Biol Evol 1999;16:1785–1790.
78. Nishio Y, Nakamura Y, Kawarabayasi Y, Usuda Y, Kimura E, Sugimoto S, Matsui K, Yamagishi A, Kikuchi H, Ikeo K, Gojobori T Comparative complete genome sequence analysis of the amino acid replacements responsible for the thermostability of *Corynebacterium efficiens* Genome Res 2003;13:1572–1579.
79. Langsetmo K, Fuchs JA, Woodward C. The conserved, buried aspartic acid in oxidized *Escherichia coli* thioredoxin has a pK_a of 7.5. Its titration produces a related shift in global stability. Biochemistry 1991;30:7603–7609.
80. Kato A, Shimizu T, Saga S. Conformational changes in mutant lysozymes detected with monoclonal antibody. FEBS Lett 1995; 371:17–20.
81. Guddat LW, Bardwell JCA, Glockshuber R, Huber-Wunderlich M, Zander T, Martin JL. Structural analysis of three His32 mutants of DsbA: support for an electrostatic role of His32 in DsbA stability. Prot Sci 1997;6:1893–1900.
82. Gloss LM, Planas A, Kirsch JF. Contribution to catalysis and stability of the five cysteines in *Escherichia coli* aspartate aminotransferase. Preparation and properties of a cysteine-free enzyme. Biochemistry 1992;31:32–39.
83. Goldenberg DP, Bekeart LS, Laheru DA, Zhou JD. Probing the determinants of disulfide stability in native pancreatic trypsin inhibitor. Biochemistry 1993;32:2835– 2844.
84. Gekko K, Kunori Y, Takeuchi H, Ichihara S, Kodama M. Point mutations at glycine- 121 of *Escherichia coli* dihydrofolate reductase: important roles of a flexible loop in the stability and function. J Biochem 1994;116:34–41.
85. Hurley JK, Caffrey MS, Markley JL, Cheng H, Xia B, Chae YK, Holden HM, Tollin G. Mutations of surface residues in Anabaena vegetative and heterocyst ferredoxin that affect thermodynamic stability as determined by guanidine hydrochloride denaturation. Prot Sci 1995;4:58–64.
86. Yoda T, Saito M, Arai M, Horii K, Tsumoto K, Matsushima M, Kumagai I, Kuwajima K. Folding-unfolding of goat alpha-lactalbumin studied by stopped-flow circular dichroism and molecular dynamics simulations. Prot Str Funct Gen 2001; 42:49–65.
87. Greene LH, Grobler JA, Malinovskii VA, Tian J, Acharya KR, Brew K. Stability, activity and flexibility in alpha-lactalbumin. Prot Eng 1999;12:581–587.
88. Raffen R, Dieckman LJ, Szpunar M, Wunschl C, Pokkuluri PR, Dave P, Stevens PW, Cai X, Schiffer M, Stevens FJ. Physicochemical consequences of amino acid variations that contribute to fibril formation by immunoglobulin light chains. Prot Sci 1999;8: 509–517.
89. Schwehm JM, Kristyanne ES, Biggers CC, Stites WE. Stability effects of increasing the hydrophobicity of solvent-exposed side chains in staphylococcal nuclease. Biochemistry 1998;37:6939–6948.
90. Stites WE, Meeker AK, Shortle D. Evidence for strained interactions between side- chains and the polypeptide backbone. J Mol Biol 1994;235:27–32.
91. Green SM, Meeker AK, Shortle D. Contributions of the polar, uncharged amino acids to the stability of staphylococcal nuclease: evidence for mutational effects on the free energy of the denatured state. Biochemistry 1992;31:5717–5728.
92. Shortle S, Stites WE, Meeker AK. Contributions of the large hydrophobic amino acids to the stability of staphylococcal nuclease. Biochemistry 1990;29:8033–8041.
93. Byrne MP, Manuel RL, Lowe LG, Stites WE. Energetic contribution of side chain hydrogen bonding to the stability of staphylococcal nuclease. Biochemistry 1995; 34:13949–13960.
94. Schejter A, Luntz TI, Koshy TI, Margoliash E. Relationship between local and global stabilities of proteins: site-directed mutants and chemically-modified derivatives of cytochrome c. Biochemistry 1992;31:8336–8343.
95. Prinsen CFM, Veerkamp JH. Fatty acid binding and conformational stability of mutants of human muscle fatty acid-binding protein. Biochem J 1996;314:253–260.
96. Gray TM, Arnoys EJ, Blankespoor S, Born T, Jagar R, Everman R, Plowman D, Stair A, Zhang D. Destabilizing effect of proline substitutions in two helical regions of T4 lysozyme: leucine 66 to proline and leucine 91 to proline. Prot Sci 1996;5:742–751.
97. Konno T, Morishima I. Transition-state of an unfolding step in

human cyanomet myoglobin. Biochim Biophys ACTA 1993;1162: 93–98.
98. Rose T, Glaser P, Surewicz WK, Mantsch HH, Reinstein J, Blay KL, Gilles A, Bârzu O. Structural and functional consequences of amino acid substitutions in the second conserved loop of *Escherichia coli* adenylate kinase. J Biol Chem 1991; 266:23654–23659.
99. Katakura Y, Totsuka M, Ametani A, Kaminogawa W. Tryptophan-19 of beta- lactoglobulin, the only residue completely conserved in the lipocalin superfamily, is not essential for binding retinol, but relevant to stabilizing bound retinol and maintaining its structure. Biochim Biophys ACTA 1994;1207:58–67.
100. Tissot AC, Vuilleumier S, Fersht AR. Importance of two buried salt bridges in the stability and folding pathway of barnase. Biochemistry 1996;35:6786–6794.
101. Kellis JT, Nyberg K, Sail D, Fersht AR. Contribution of hydrophobic interactions to protein stability. Nature 1988;333: 784–786.
102. Serrano L, Sancho J, Hirshberg M, Fersht AR. Alpha-helix stability in proteins 1. Empirical correlations concerning substitution of side-chains at the N and C-caps and the replacement of alanine by glycine or serine at solvent-exposed surfaces. J Mol Biol 1992;227:544–559.
103. Solà M, López-Hernández E, Cronet P, Lacrois E, Serrano L, Coll M, Fárraga A. Towards understanding a molecular switch mechanism: thermodynamic and crystallographic studies of the signal transduction protein CheY. J Mol Biol 2000;303:213–225.
104. Dao-pin D, Anderson DE, Baase WA, Dahlquist FW, Matthews BW. Structural and thermodynamic consequences of burying a charged residue within the hydrophobic core of T4 lysozyme. Biochemistry 1991;30:11521–11529.
105. Schymkowitz JWH, Rousseau F, Itzhaki LS. Sequence conservation provides the best prediction of the role of proline residues in p13sucl. J Mol Biol 2000;301:199–204.
106. Kwon KS, Kin J, Shin HS, Yu MH. Single amino acid substitutions of alpha 1- antitrypsin that confer enhancement in thermal stability. J Biol Chem 1994;269:9627– 9631.
107. Holder JB, Bennett AF, Chen J, Spencer DS, Byrne MP, Stites WE. Energetics of side chain packing in staphylococcal nuclease assessed by exchange of valines, isoleucines, and leucines. Biochemistry 2001;40:13998–134003.
108. Nakano T, Antonino LC, Fox RO, Fink AL. Effect of proline mutations on the stability and kinetics of folding of staphylococcal nuclease. Biochemistry 1993;32:2534– 2541.
109. Huang Y, Beeser S, Guillemette JG, Storms RK, Kornblatt JA. Mutations of iso-1-cytochrome-c at position-13 and position-90. Separate effects on physical and functional properties. Eur J Biochem 1994;223:155–160.