Modeling Backbone Flexibility Improves Protein Stability Estimation

Shuangye Yin,[1] Feng Ding,[1] and Nikolay V. Dokholyan[1,*]
[1]Department of Biochemistry and Biophysics, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
*Correspondence: dokh@med.unc.edu
DOI 10.1016/j.str.2007.09.024

SUMMARY

In designing mutagenesis experiments, it is often crucial to know how certain mutations will affect the structure and thermodynamic stability of the protein. Here, we present a methodology, Eris, to efficiently and accurately compute the stability changes of proteins upon mutations using our protein-modeling suite, Medusa. We evaluate the stability changes upon mutations for 595 mutants from five structurally unrelated proteins, and find significant correlations between the predicted and experimental results. For cases when the high-resolution protein structure is not available, we find that better predictions are obtained by backbone structure prerelaxation. The advantage of our approach is that it is based on physical descriptions of atomic interactions, and does not rely on parameter training with available experimental protein stability data. Unlike other methods, Eris also models the backbone flexibility, thereby allowing for determination of the mutation-induced backbone conformational changes. Eris is freely available via the web server at http://eris.dokhlab.org.

INTRODUCTION

Protein engineering is an invaluable tool for molecular biologists. Mutagenesis is used to probe functional (Obara et al., 1988; Fersht et al., 1985; Pakula and Sauer, 1989), structural (Matouschek et al., 1990; Serrano et al., 1992), and folding kinetic (Fersht et al., 1992; Jackson and Fersht, 1991) roles of specific protein sites. However, the extent to which a given set of mutations affects protein stability is difficult to estimate due to the complex nature of the physical interactions. A measure of protein stability is the difference between the free energies of the folded and unfolded states (ΔG). Change in this free energy difference ($\Delta \Delta G = \Delta G^{mutant} - \Delta G^{wild-type}$) upon mutation is a measure of protein (de)stabilization by mutations. Due to a large number of degrees of freedom associated with proteins and solvent(s), computational estimation of free

energies is an extremely challenging problem (Beveridge and Dicapua, 1989).

The $\Delta \Delta G$ values can be calculated from simulations of proteins with detailed atomic models with physical force fields (Bash et al., 1987; Dang et al., 1989; Duan and Kollman, 1998; Kollman et al., 2000; Vorobjev and Hermans, 1999; Khare et al., 2006). Such methods, although rigorous, are computationally too intense to be applied to a large number of mutations in a course of protein engineering. Alternatively, knowledge-based potentials derived from known protein structure databases have been used to estimate $\Delta \Delta G$ with reasonable accuracy (Gilis and Rooman, 1996, 1997, 2000; Ota et al., 2001; Zhou and Zhou, 2002; Hoppe and Schomburg, 2005). Gilis and Rooman first applied database-derived backbone potentials to study the change of thermodynamic stability upon point mutations (Gilis and Rooman, 1996, 1997, 2000). They found that torsion angle-based potentials predict $\Delta \Delta G$ accurately for mutations of solvent-exposing residues, and distance-dependent statistical potentials are necessary for buried residues. They obtained correlation coefficients of 0.55–0.87 for their whole dataset of 238 mutations. Zhou and Zhou (2002) developed distance-scaled finite ideal-gas reference states based on statistical potentials and calculated $\Delta \Delta G$s for 895 mutants, which agree reasonably well with experimental measurements ($r = 0.67$). Using direction- and distance-dependent statistical potentials, Hoppe and Schomburg (2005) trained their parameters on 646 mutants, and were able to predict $\Delta \Delta G$ of 747 mutants in the test set, with a correlation of 0.46. Guerois et al. (2002) developed FOLD-X energy function to study the stabilities of 1088 mutants, and obtained a correlation of 0.64 for the blind test set after training their parameters on 339 mutants. Kortemme and Baker (2002) successfully used an energy function similar to ours to estimate the $\Delta \Delta G$ for both protein stabilities and protein binding affinities. However, they did not model the backbone flexibility, thereby limiting their study to mutations of smaller or the same side-chain sizes. Khatun et al. (2004) utilized contact potentials to predict $\Delta \Delta G$ of three sets of 303, 658, and 1356 mutants, and their prediction correlations varied between 0.45 and 0.78. Bordner and Abagyan (2004) used a combination of physical energy terms, statistical energy terms, and structural descriptor with weight factors scaled to experimental data for $\Delta \Delta G$ predictions. Saraboji et al. classified the available thermal denaturing data on mutations according to

substitution types, secondary structures, and the area of solvent accessibilities, and used the average value from each category for the prediction, and obtained a correlation of 0.64 (Saraboji et al., 2006). Capriotti et al. (2005a) introduced machine-learning techniques for $\Delta\Delta G$ predictions. They trained a support vector machine (SVM) using temperature, pH, mutations, nearby residues, and relative solvent-accessible area as input vectors. The SVM, when applied on a test set, gives a prediction correlation of 0.71.

However, there are two significant drawbacks in training-based studies (Yin et al., 2007). First, the improvement of the prediction accuracy relies on the available experimental stability data for parameter trainings. It is questionable whether parameters obtained from these trainings are transferable to other protein studies (Khatun et al., 2004), since the experimentally available data may be biased toward mutations from large to small residues, such as in alanine scanning experiments. Second, some mutations introduce strains in proteins' backbones. In order to properly estimate the $\Delta\Delta G$ values, it is necessary to simulate the structure relaxation that a protein may undergo to release the strains. To our knowledge, protein dynamics and flexibility have not been explicitly modeled in previous $\Delta\Delta G$ prediction methods. Ignoring protein flexibility limits the application of current prediction methods to a small range of mutations (Guerois et al., 2002; Zhou and Zhou, 2002).

Here, in order to address both of these caveats, we propose a novel method, Eris, for accurate and rapid evaluation of the $\Delta\Delta G$ values with the recently developed Medusa modeling suite (Ding and Dokholyan, 2006). Eris features an all-atom force field, a fast side-chain packing algorithm, and a backbone relaxation method. The force field parameters are independently trained with high-resolution protein structures (see Experimental Procedures). The $\Delta\Delta G$ values of 595 mutants from 5 structurally unrelated proteins are calculated and compared with the experimental data from the Protherm database and other sources (Guerois et al., 2002; Edgell et al., 2003; Bava et al., 2004; Khatun et al., 2004) (Table S1, see the Supplemental Data available with this article online). We find significant correlations between the calculations and the experimental measurements. The Pearson linear regression coefficient between actual and predicted $\Delta\Delta G$ values is ≈ 0.65 over the whole dataset without disregarding outliers. After taking into account corrections to the reference energies, the correlation coefficient further improves to ≈ 0.75. Given the difficulty in the $\Delta\Delta G$ estimation, such an agreement between $\Delta\Delta G$ estimations and experimental measurements is reasonable for guiding experimental efforts in delineating effects of point mutations on protein structure, stability, and function. Additionally, Eris automatically identifies and efficiently relaxes the strains in the backbone when clashes and backbone strains are introduced by a small-to-large amino acid substitution. Importantly, when high-resolution structures are not available, Eris adopts a prerelaxation procedure to relax possible strains in the structure, which yields higher estimation accuracy.

We built a web-based Eris server (Yin et al., 2007) for on-line $\Delta\Delta G$ estimations. The server is freely accessible at http://eris.dokhlab.org. We expect applications of Eris to a wide range of protein engineering problems.

RESULTS

We test our protocol on a dataset that consists of 595 mutations and compare the results with experimental measurements. Wild-type structures of these proteins, which are required in our $\Delta\Delta G$ calculations, are retrieved from the Protein Data Bank (PDB; Berman et al., 2000). The mutants are of five structurally unrelated proteins: FK506 binding protein (PDB ID: 1fkj), apomyoglobin (PDB ID: 1bvc), staphylococcal nuclease (PDB ID: 1stn), chymotrypsin inhibitor 2 (PDB ID: 2ci2), and α spectrin domain repeat 16 (PDB ID: 1u5p). In all proteins, the mutations are distributed throughout primary sequences and secondary structures (Figure 1). These proteins also have diverse secondary structures: FK506 binding protein is mainly composed of β sheets (Figure 1C); apomyoglobin (Figure 1F) and α spectrin (Figure 1O) are mainly α helical; and staphylococcal nuclease (Figure 1I), chymotrypsin inhibitor 2 (Figure 1L) have mixtures of different secondary structure elements. In order to assess the importance of backbone-flexibility modeling in $\Delta\Delta G$ evaluation, we compute the $\Delta\Delta G$ values of mutants using both fixed-backbone and flexible-backbone methods (see Experimental Procedures).

Eris Predicts ΔΔG in Agreement with the Experiments

The calculated $\Delta\Delta G$ values for all the five proteins significantly correlate with the experimental measurements (Figure 2 and Table 1). The Pearson linear regression correlation coefficients between the calculated and experimental $\Delta\Delta G$ values for the aggregated 595 mutants are 0.64 and 0.66 by fixed and flexible backbone methods, respectively. The null hypothesis probabilities (the probabilities of observing the correlation by chance) of the correlations are 3.1×10^{-71} and 1.2×10^{-75}, respectively. These correlations between the predicted and the measured $\Delta\Delta G$ values is comparable to those of previous methods (Gilis and Rooman, 1996, 1997, 2000; Ota et al., 2001; Guerois et al., 2002; Zhou and Zhou, 2002; Bordner and Abagyan, 2004; Khatun et al., 2004; Capriotti et al., 2005a, 2005b; Hoppe and Schomburg, 2005; Cheng et al., 2006; Saraboji et al., 2006). The slopes from the linear regression fit are 0.91 (± 0.04) and 0.95 (± 0.04) for fixed and flexible backbone methods, respectively. For each individual protein, the correlation coefficients vary from 0.47 to 0.73, and the slopes from the linear regression fit range from 0.7 to 1.4 (Figure 1 and Table 1).

To test if Eris is limited to single-site mutations (i.e., with one amino acid substitution), we separate all the multiple-site mutations (with more than one amino acid substitutions) from the dataset and compare the performance of
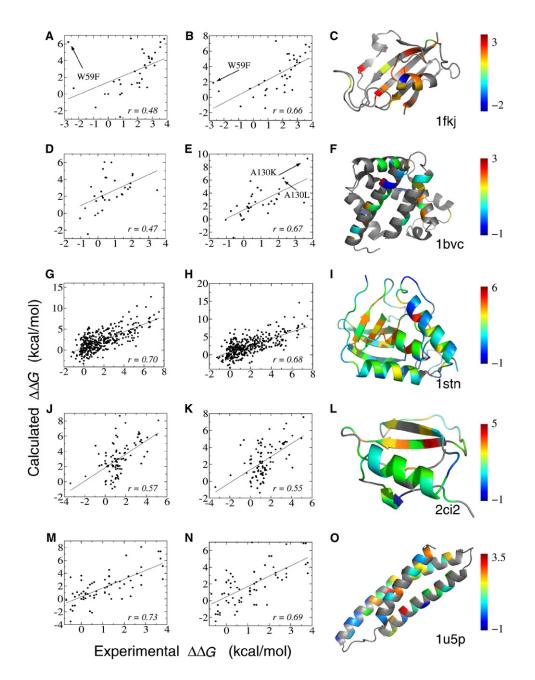
Figure 1. Correlations between the Calculated and Measured ΔΔG For 595 Mutations on 5 Proteins by Both Fixed-and Flexible-Backbone Prediction Methods

(A–O) Fixed-backbone method is presented in the first column and flexible-backbone in the second column. The corresponding protein structures are shown in the third column. The mutation sites are color-coded according to the average experimental stability changes of the mutations at these sites (color bars on the right; units are kcal/mol). The straight lines correspond to the linear regression fits of the data. Some mutants are marked where significant differences are found in the ΔΔG calculations by fixed- and flexible-backbone methods.

Eris protocol on the divided datasets (Table 2). We do not observe loss of prediction accuracy for multiple-site mutations. The correlation coefficients for all single-site mutations using fixed and flexible backbone methods are 0.65 and 0.66, respectively. For multiple-site mutations, the correlations are 0.69 and 0.64, respectively.

The Flexible-Backbone versus Fixed-Backbone Method

From the correlation coefficients over all datasets, there is no significant difference between the flexible- and fixed-backbone modeling methods. For individual protein datasets, the flexible- and fixed-backbone methods also
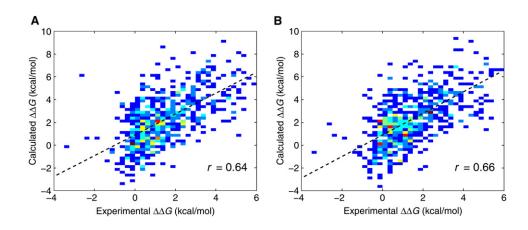
Figure 2. The Calculated and Experimental ΔΔGs for All 595 Mutations by Fixed- and Flexible-Backbone Methods
(A) Fixed- and (B) flexible-backbone method. The color of the point corresponds to the number of data points in the 0.25 kcal/mol × 0.25 kcal/mol bin, which ranges from 1 (blue color) to 10 (red color). The lines correspond to linear regression fits to the data points. The overall correlation coefficients are 0.64 and 0.66 for fixed- and flexible-backbone estimations, respectively.

perform similarly. For example, the flexible-backbone method performs better for FK506 binding protein (1fkj) and apomyoglobin (1bvc), while the fixed-backbone method results in higher correlations for the other four proteins (see Table 1). However, a closer examination suggests that the flexible-backbone method is superior to the fixed-backbone method in certain cases. For instance, in FK506 binding protein (1fkj), mutant W59F is experimentally determined to be more stable than the wild-type, and the measured ΔΔG is −2.72 kcal/mol (Fulton et al., 2003). While neither of the prediction methods estimates this stabilization, the flexible-backbone method yields better prediction (Figures 1A and 1B). Additionally, for apomyoglobin (1bvc), mutants A130K and A130L destabilize the protein by 3.7 and 2.3 kcal/mol, respectively (Hughson et al., 1991). Our fixed-backbone calculations for these mutations yield unrealistic destabilizing values (which are screened out), because there is no space to accommodate the larger lysine or leucine side chains between the two helices in the wild-type structure. In contrast, by the flexible backbone method, one of the helix

bends outward slightly to fit these larger mutation side chains (Figure 3C), and the calculated stability changes are in agreement with the experimental measurements (Figures 1D and 1E). Therefore, allowing for backbone flexibility during calculations effectively resolves atomic clashes and offers more accurate evaluation of the ΔΔG values.

Backbone Prerelaxation Improves Prediction Accuracy
Eris protocol evaluates a protein's free energy by its three-dimensional structural information. When a protein structure is of low resolution or has some poorly resolved fragments, the accuracy of ΔΔG calculation could be affected. We counter this problem by providing a prerelaxation step in the Eris protocol, in which the protein's backbone structure is optimized to have the least strain throughout the whole protein (see Experimental Procedures). In our test on the protein α-spectrin domain R16, we find that, if a nuclear-magnetic-resonance (NMR)-derived structure (PDB ID: 1aj3) is used in the energy evaluation, the overall

Table 1. The Correlation Coefficients between the Calculated and Experimental ΔΔG for Various Proteins by Fixed- and Flexible-Backbone Methods

Protein	N_{res}	ΔG_{wt} (kcal/mol)	N_{mut}	Fixed Backbone			Flexible Backbone		
				r	p Value	Slope	r	p Value	Slope
1fkj	107	6.11	34	0.48	3.9×10^{-3}	0.7	0.66	1.3×10^{-5}	1.0
1bvc	153	5.80	31	0.47	8.6×10^{-3}	0.9	0.67	2.7×10^{-5}	1.3
1stn	136	5.40	371	0.70	7.7×10^{-57}	0.9	0.68	4.2×10^{-52}	0.9
2ci2	65	7.53	91	0.57	2.3×10^{-9}	0.9	0.55	1.1×10^{-8}	0.8
1u5p	98	6.30	68	0.73	6.5×10^{-13}	1.4	0.69	8.4×10^{-11}	1.2

N_{res} is the number of residues in each protein. ΔG_{wt} is the measured ΔG of the wild type. N_{mut} is the number of mutations of each protein in our dataset. r is the correlation between the calculated and measured ΔΔG. The p value is the null hypothesis probability of the correlation (the probability of observing the correlation by chance). Slopes are of the linear regression fittings lines shown in Figure 1.

Table 2. The Correlation Coefficients between the Calculated and Experimental ΔΔG for Single-Site and Multiple-Site Mutations

Mutation type	N_{mut}	Fixed Backbone		Flexible Backbone	
		r	p Value	r	p Value
All	595	0.64	3.1×10^{-71}	0.66	1.2×10^{-75}
Single-site	573	0.65	1.7×10^{-69}	0.66	3.6×10^{-73}
Multiple-site	22	0.69	3.9×10^{-4}	0.64	1.5×10^{-3}

correlation with experiments is only 0.09. When an NMR-derived structure has multiple models, we select the first model, which is often the lowest energy conformation. A closer examination shows that the disagreement is mainly due to mutations located in the protein core (Figure 3A). However, if we calculate ΔΔG using the prerelaxed backbone structure, the overall correlation increases to 0.70. In this case, the calculated ΔΔG for both surface and core mutations are in line with experiments (Figure 3B). We find that the prerelaxed backbone has an rmsd of about 0.77 Å from the original NMR-derived structure. Such a prerelaxation step significantly improves the side-chain packing of the protein core, which, consequently, leads to more accurate ΔΔG evaluations. Hence, by relieving backbone strains, the prerelaxation step enables Eris to tolerate imperfections in the input protein structures and improves the accuracy of stability evaluation.

DISCUSSION

Using the Medusa force field and modeling suite, we estimate protein stabilities in agreement with previous experiments (Guerois et al., 2002; Bava et al., 2004; Khatun

et al., 2004) (Table S1). Our approach does not rely on fitting or training force field parameters using stability measurements, yet it yields comparable accuracy with heuristic methods. This success, in return, validates the Medusa force field.

Further Medusa Force Field Improvement

In the original Medusa force field, there is a reference energy term for each amino acid, which is optimized to recapitulate native amino acid sequences during protein sequence search (design) for the optimal one for a given backbone structure (see Experimental Procedures). Thus, the optimized reference energy of each amino acid is comprised of two components: its free energy in the unfolded state, and the effective energy that corresponds to the natural occurrence of this amino acid. The second component is irrelevant to protein stability, but is required for molecular evolutionary studies (Saunders and Baker, 2005). In order to accurately predict the protein stability, we need to separate the second component from the reference energies. Since the relative contributions of the above components are not known, it is not straight-forward how to separate them. We propose to adjust the amino acids' reference energies in the force field by reca-pitulating the experimentally measured ΔΔGs. When we adjust the reference free energies of the 20 amino acids, we indeed find that the correlation coefficient significantly improves from 0.64 (0.66) to 0.74 (0.76) for fixed (flexible) backbone ΔΔG calculations (Figure 4). As we expected, we find that the recapitulation rate of the native sequence decreases at the same time. This observation is in agreement with the current understanding that protein stabilities are not necessarily evolutionary optimized (Kuhlman and Baker, 2004). To show that these reference energy adjustments are indeed related to protein stability, we apply
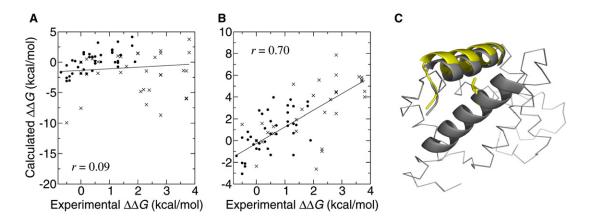

Figure 3. Flexible-Backbone Modeling Helps to Relieve Backbone Strains
(A) The ΔΔG values of α-spectrin mutants are obtained with a protein structure from NMR experiments (PDB ID: 1aj3). The correlation is 0.66 for surface mutations (solid circle) and only 0.28 for core mutations (cross).
(B) The ΔΔG values of α-spectrin mutants are calculated with a backbone prerelaxed structure. Significant correlations are obtains for both surface (solid circle) and core mutations (cross).
(C) Backbone structures of the apomyoglobin protein and its mutant, A130L, from a flexible-backbone calculation. Part of the backbone of A130L bends slightly outward to fit the larger mutated side chain, which is highlighted in yellow. Other parts of the A130L and the wild-type structure are shown in gray.
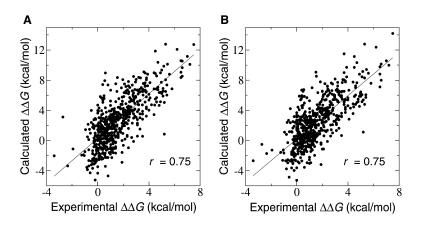
Figure 4. Scatter Plots of the Calculated and Experimental ΔΔG for All Mutations after Adjusting Reference Energies
The new ΔΔG values from (A) fixed- and (B) flexible-backbone calculations are plotted against experimental values. The straight lines are the linear regression fits; the line corresponds to (A) y = 1.5x + 0.30 and (B) y = 1.4x + 0.21.

these parameters to each individual protein, and find that the correlation coefficients are all significantly improved (Table 3). Additionally, we randomly divide the mutation dataset into a training set and a test set. The parameters obtained from the training set are nearly identical to those from the whole dataset. The new parameters also significantly improve the prediction accuracy when they are applied to the test set. These crossvalidation tests indicate that the adjusted parameters are indeed relevant to protein stabilities. In the following discussions, and also on the Eris server, we use the new set of reference energies.
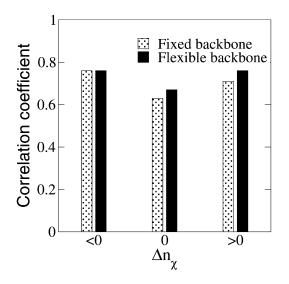
Fixed versus Flexible Backbone
We demonstrate that the flexible-backbone method can successfully resolve atomic clashes by relaxing the backbone conformations. Why, then, is the average performance of the flexible-backbone method not better than that of the fixed-backbone approach? We notice that most available mutation experiments are often biased toward large-to-small mutations to prevent incurring severe backbone perturbations. In order to assess the advantage of backbone flexibility modeling in an unbiased manner, we divide all the single mutations into three classes based

on the change of number of side-chain χ angles (Δn_χ) of the mutation, and compare the performance of the fixed- and flexible-backbone methods on these three classes separately (Figure 5). The mutations with $\Delta n_\chi < 0$ are associated with large-to-small mutations; therefore, the protein backbone is expected to be minimally altered—those with $\Delta n_\chi \geq 0$ correspond to mutation to residues of the same or larger sizes, and backbone adjustment is expected if the mutation site is buried.

Among the 573 single-site mutations, 439 mutations have decreasing n_χ, 103 mutations have the same n_χ, and 31 mutations have increasing n_χ. The uneven distribution of the dataset over Δn_χ clearly reflects the bias of the currently available experimental data toward large-to-small mutations. The fixed and flexible backbone prediction methods perform equally well for $\Delta n_\chi < 0$. However, the flexible-backbone ΔΔG prediction method correlates

Table 3. Comparison between the Correlation Coefficients of ΔΔG Estimations with the Original and the Improved Reference Energies

Protein	Fixed Backbone		Flexible Backbone	
	r	r^*	r	r^*
All	0.64	0.75	0.66	0.75
1fkj	0.48	0.70	0.66	0.77
1bvc	0.47	0.55	0.67	0.69
1stn	0.70	0.78	0.68	0.77
2ci2	0.57	0.69	0.55	0.70
1u5p	0.73	0.83	0.69	0.76

r and r^* are the correlation coefficients between the calculated and experimental ΔΔG with and without using the trained parameters. Significant improvements are found for all the proteins using the new reference energies.

Figure 5. Correlation Coefficients between the Calculated and Experimental ΔΔGs for All Single-Site Mutations of Different Classes of Δn_χ
The flexible-backbone method predicts with higher accuracy than the fixed-backbone method for mutations of $\Delta n_\chi \geq 0$.

better with experiments for $\Delta n_\chi \geq 0$ cases. We believe that the superior performance of the flexible-backbone method in the latter cases is due to its ability to resolve possible side-chain clashes. Hence, unlike previous methods that are often limited to large-to-small mutations (Guerois et al., 2002; Zhou and Zhou, 2002), our flexible backbone approach is applicable to any type of mutations, thereby opening wide opportunities in protein engineering applications.

Does the Eris estimation accuracy depend on the solvent exposure of a mutated site? We compute the correlation coefficients between computationally estimated and experimentally measured $\Delta\Delta G$ based on the extent of solvent exposure of mutated residues. A residue is defined as buried if its solvent-accessible surface area (SASA) is less than half of its total surface area, and a residue is exposed if its SASA is larger than half of its total surface area. Interestingly, we find that Eris predicts the $\Delta\Delta G$ more accurately for the buried residues than for the solvent-exposed residues (Table S2). We believe such a difference in the $\Delta\Delta G$ estimation accuracy is attributable to solvent-exposed residues having higher side-chain entropies in the folded state than buried residues, which is not explicitly considered in Eris. We also compute the correlation coefficients between computationally estimated and experimentally measured $\Delta\Delta G$ values based on different types of the secondary structures where the mutated residues locate. However, we do not observe any significant trends of the estimation accuracies for mutations at different types of secondary structures, except loops and isolated β bridges (Table S3). These two types of secondary structures are often solvent exposed. Therefore, we expect the improvement of Eris prediction by including entropic effect in future studies.

Backbone Prerelaxation

We find that the Eris protocol is more accurate when high-resolution protein structures are used for $\Delta\Delta G$ evaluation. Guerois et al. (2002) also found that more reliable prediction is obtained using high-resolution X-ray protein structures (resolution < 1.5 Å). They performed the prerelaxation of the initial protein structure, but failed to improve the prediction accuracy, which is possibly due to the fact that they used different force fields in structure relaxation and in stability estimations. On the contrary, our structure relaxation seems to improve the prediction accuracy for most proteins, one extreme example of which has been demonstrated earlier for α-spectrin domain R16 protein. Here, the advantage of the Eris protocol is that the same force field is used for both structure modeling and for free energy evaluations. Such force field transferability is usually not available for statistical potential- or empirical parameter-based $\Delta\Delta G$ evaluation methods.

Comparison with Other Methods

Compared with other $\Delta\Delta G$ prediction methods, the Eris protocol is a unique approach that combines physical and statistical energies with effective atomic modeling, resulting in fast and accurate side-chain packing and

Table 4. Comparison of Eris $\Delta\Delta G$ Predictions with Other Servers

Server (Reference)	Correlation
Eris flexible backbone	0.77
Eris fixed backbone	0.70
FoldX (Guerois et al., 2002)	0.41
Dmutant (Zhou and Zhou, 2002)	0.46
I-Mutant (Capriotti et al., 2005b)	0.58
MutPro (Cheng et al., 2006)	0.64

The calculations of $\Delta\Delta G$s for all mutants of protein 1fkj are submitted to some publicly accessible servers and the resulting correlation coefficients are compared with those from the Eris server.

backbone optimization. We compare our Eris server with other available online prediction servers for protein 1fkj. As shown in Table 4, all the servers we tested show impressive prediction accuracy; the correlations range from 0.41 to 0.64 for all the mutants of 1fkj. Our Eris flexible-backbone protocol outperforms all other servers mainly, due to its ability to model backbone conformational changes induced by some mutations.

Strengths and Limitations of the Eris Method

Due to a large number of degrees of freedom associated with protein and solvent molecules, precise estimations of $\Delta\Delta G$ values through "brute force" simulations are virtually impossible, unless certain assumptions are made to simplify the problem. For example, direct simulation approaches often assume that structural changes associated with mutations are small, so that short simulations ($\sim$10 ps) can provide sufficient sampling of the protein's conformational space (Dang et al., 1989). Statistical potential-based approaches rely on the assumption that total free energy can be decomposed into additive pairwise functions and/or dihedral angle-dependent potentials, which are obtained from statistical analysis of solved protein structures. In Eris, we use a physical approach, in which we reduce the complexity of the sampling problem by following several approximations. First, we avoid sampling of the solvent by utilizing an implicit solvent model, so that the free energy contribution from the solvent is replaced by an average hydration effect (Lazaridis and Karplus, 1999). Second, we postulate that the protein conformational entropy is not significantly changed upon mutation. Third, we use a rotamer library (Dunbrack and Cohen, 1997) to efficiently model side-chain conformations, and we use statistical potentials associated with the backbone-dependent rotamer distributions to model the internal interactions within amino acids. With such simplifications, the sampling efficiency is significantly improved, which allows Eris to model structural changes associated with both side-chain and backbone motions. When the structural change upon mutations is modest, the structural perturbation is accounted for by the side-chain repacking and the backbone relaxation. The statistically significant

agreement between Eris estimations and the experimental measurements suggests that these assumptions are reasonable and the method can capture major factors that govern the mutation-induced stability changes, $\Delta\Delta G$. In the future, the Eris method will benefit from the continuing development of the physical force field and the implicit solvent model. Better estimations of the protein conformational entropy may dramatically improve the $\Delta\Delta G$ calculation accuracy.

Is the Eris method restricted to particular protein families? The five proteins we tested belong to different fold families. The mutations are distributed throughout both solvent-exposed and buried sites, and are located on distinct secondary structure fragments (Figure 1). Since our method is based on physical interactions between atoms, it is expected that it does not significantly depend on the protein choice. We also find that the Eris method does not have decreased accuracy for small-to-large mutations. The consistent performance of the Eris method on such a structurally diverse test set is sufficient to verify that both the force field and the sampling algorithm are functioning properly. Therefore, we believe that the Eris method is valid for any other proteins at the claimed accuracy.

Does the accuracy of the prediction depend on protein size? For the proteins we study (length ranges from 65 to 153 residues), we do not observe any length dependence of the correlation between $\Delta\Delta G$ estimations and experimental measurements. As the protein size increases (e.g., to over 1000 residues), the Eris' sampling may become insufficient. In such cases, the side-chain packing and backbone relaxation algorithms may not be able to efficiently find the minimal energy structure. Therefore, larger uncertainty is possible for large proteins. In certain cases, when the structural change upon mutations is limited only to residues in the vicinity of the mutation site, the Eris method can be applied to the fragment of the structure that contains only neighboring residues, thereby circumventing the sampling problem.

In Eris, we do not consider posttranslationally modified proteins. Eris uses a statistically derived rotamer library to account for the internal energy of an amino acid at a given backbone conformation, and to increase the side-chain conformational sampling efficiency. Unfortunately, there are no statistically sufficient experimental data to evaluate a rotamer library for modified residues. Hence, we only include the 20 natural amino acids in the current Eris protocol. In principle, as more experimental data become available, Eris could be extended to include posttranslationally modified and nonnatural amino acids.

It is important that Eris is statistically correct and should be used as a statistically predictive tool. The overall agreement between $\Delta\Delta G$ estimations and experimental measurements indicates that Eris results can provide $\Delta\Delta G$ estimation in the proximity of the actual stability change with a high probability. Hence, the Eris results will help decide on a set of candidate mutations for experimentalists who aim to probe protein stability, structure, and function with mutations (Yin et al., 2007).

EXPERIMENTAL PROCEDURES

Medusa Force Field

The Medusa force field is based on a united atom model that includes all the heavy atoms and polar hydrogen atoms in the protein. We use the backbone-dependent rotamer library (Dunbrack and Cohen, 1997; Ding and Dokholyan, 2006) to model the local interactions and express the free energy of the protein as a weighted sum of van der Waals (VDW), solvation, hydrogen-bonding, and backbone-dependent statistical energies (Dunbrack and Cohen, 1997). Mathematically, the ΔG of a protein is calculated as:

$$\Delta G = W_{vdw\_attr}E_{vdw\_attr} + W_{vdw\_rep}E_{vdw\_rep} + W_{solv}E_{solv}$$
$$+ W_{bb\_hbond}E_{bb\_hbond} + W_{sc\_hbond}E_{sc\_hbond} + W_{bb\_sc\_hbond}E_{bb\_sc\_hbond} \quad (1)$$
$$+ W_{aa|\phi,\psi}E_{aa|\phi,\psi} + W_{rot|\phi,\psi,aa}E_{rot|\phi,\psi,aa} - E_{ref}$$

Here E_{vdw_attr} and E_{vdw_rep} are the attractive and repulsive parts of the VDW interaction, respectively; E_{solv} is the solvation energy; E_{bb_hbond}, E_{sc_hbond}, and $E_{bb\_sc\_hbond}$ are the hydrogen bond energies among backbones, among side chains, and between backbones and side chains, respectively. $E_{aa|\phi,\psi}$ and $E_{rot|\phi,\psi,aa}$ correspond to the internal energy for an amino acid (aa) in its rotamer state (rot) given the backbone dihedrals, phi (ϕ) and psi (ψ). E_{ref} is the reference energy of the unfolded state, which is calculated as a linear sum over the reference energies of all the amino acids (Ding and Dokholyan, 2006). Since the parameters for the energy terms come from different sources, we use weight coefficients to balance their contribution to the total free energy. The weight coefficients and the reference energies are trained on 34 high-resolution X-ray protein structures so that the native amino acid sequences will have the lowest free energy (Ding and Dokholyan, 2006; Kuhlman and Baker, 2000).

Fixed-Backbone Method

In a fixed backbone calculation, we read the native structure and mutate the specified amino acids. After the mutation, the amino acids' rotamer states are first randomized and are subsequently minimized using a Monte Carlo (MC) simulated annealing procedure. In each MC simulation step, one amino acid is randomly selected and transformed to another rotamer state. This trial step is either accepted or rejected according to the Metropolis criterion, depending on the change of total free energy. To sample subrotamer space, for each step, deviations from the average values of the rotamer states are allowed within the tabulated standard deviations in the rotamer library (Dunbrack and Cohen, 1997). We start the simulation at high temperature and gradually decrease the temperature. Additionally, at the last annealing temperature, we apply a quench procedure, where each rotamer state change is followed by a conjugate-gradient minimization on the subrotamer state, and the trial step is accepted only if the new energy is lower. Because of its stochastic nature, we run the simulation 20 times and take the average free energy as the final calculated ΔG for the protein. We run the ΔG calculation for both mutant and wild-type proteins, and calculate the $\Delta\Delta G$ value as the difference between the ΔG of the mutant and that of the wild type ($\Delta\Delta G = \Delta G^{mutant} - \Delta G^{wild\ type}$).

Flexible-Backbone Method

The flexible-backbone method is similar to the fixed-backbone one, except that we allow the backbone dihedrals to relax if backbone strains are detected. This is implemented by allowing a conjugate-gradient minimization of the total free energy with respect to all backbone dihedral angles if the total acceptance rate during an MC loop is below a predefined threshold. Such a low acceptance rate usually indicates strains on the backbone induced by mutations. In the Eris flexible-backbone protocol, the threshold is taken to be 0.05.

Backbone Prerelaxation

The backbone prerelaxation is implemented by performing the flexible-backbone calculation 20 times with the wide-type protein structure and sequence. The output structure with the minimum free energy is selected as the prerelaxed structure.

Correlation Coefficient

We evaluate the $\Delta\Delta G$ estimation accuracy by comparing the Pearson linear regression correlation coefficients between the calculated and experimental $\Delta\Delta G$ values. The Pearson linear regression correlation r between two datasets, x_i and y_i, is calculated as,

$$r = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}}. \qquad (2)$$

Supplemental Data

Supplemental Data include three tables and are available online at http://www.structure.org/cgi/content/full/15/12/1567/DC1/.

REFERENCES

Bash, P.A., Singh, U.C., Langridge, R., and Kollman, P.A. (1987). Free-energy calculations by computer-simulation. Science 236, 564–568.

Bava, K.A., Gromiha, M.M., Uedaira, H., Kitajima, K., and Sarai, A. (2004). ProTherm, version 4.0: thermodynamic database for proteins and mutants. Nucleic Acids Res. 32, D120–D121.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Res. 28, 235–242.

Beveridge, D.L., and Dicapua, F.M. (1989). Free-energy via molecular simulation—applications to chemical and biomolecular systems. Annu. Rev. Biophys. Biophys. Chem. 18, 431–492.

Bordner, A.J., and Abagyan, R.A. (2004). Large-scale prediction of protein geometry and stability changes for arbitrary single point mutations. Proteins 57, 400–413.

Capriotti, E., Fariselli, P., Calabrese, R., and Casadio, R. (2005a). Predicting protein stability changes from sequences using support vector machines. Bioinformatics 21 (Suppl 2), ii54–ii58.

Capriotti, E., Fariselli, P., and Casadio, R. (2005b). I-Mutant 2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res. 33, W306–W310.

Cheng, J.L., Randall, A., and Baldi, P. (2006). Prediction of protein stability changes for single-site mutations using support vector machines. Proteins 62, 1125–1132.

Dang, L.X., Merz, M.M., and Kollman, P.A. (1989). Free energy calculations on protein stability: Thr-157 → Val-157 mutation of T4 lysozyme. J. Am. Chem. Soc. 111, 8505–8508.

Ding, F., and Dokholyan, N.V. (2006). Emergence of protein fold families through rational design. PLoS. Comput. Biol. 2, e85. Published online July 7, 2006. 10.1371/journal.pcbi.0020085.

Duan, Y., and Kollman, P.A. (1998). Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. Science 282, 740–744.

Dunbrack, R.L., and Cohen, F.E. (1997). Bayesian statistical analysis of protein side-chain rotamer preferences. Protein Sci. 6, 1661–1681.

Edgell, M.H., Sims, D.A., Pielak, G.J., and Yi, F. (2003). High-precision, high-throughput stability determinations facilitated by robotics and a semiautomated titrating fluorometer. Biochemistry 42, 7587–7593.

Fersht, A.R., Shi, J.P., Knilljones, J., Lowe, D.M., Wilkinson, A.J., Blow, D.M., Brick, P., Carter, P., Waye, M.M.Y., and Winter, G. (1985). Hydrogen-bonding and biological specificity analyzed by protein engineering. Nature 314, 235–238.

Fersht, A.R., Matouschek, A., and Serrano, L. (1992). The folding of an enzyme. 1. Theory of protein engineering analysis of stability and pathway of protein folding. J. Mol. Biol. 224, 771–782.

Fulton, K.F., Jackson, S.E., and Buckle, A.M. (2003). Energetic and structural analysis of the role of tryptophan 59 in FKBP12. Biochemistry 42, 2364–2372.

Gilis, D., and Rooman, M. (1996). Stability changes upon mutation of solvent-accessible residues in proteins evaluated by database-derived potentials. J. Mol. Biol. 257, 1112–1126.

Gilis, D., and Rooman, M. (1997). Predicting protein stability changes upon mutation using database-derived potentials: solvent accessibility determines the importance of local versus non-local interactions along the sequence. J. Mol. Biol. 272, 276–290.

Gilis, D., and Rooman, M. (2000). PoPMuSiC, an algorithm for predicting protein mutant stability changes: application to prion proteins. Protein Eng. 13, 849–856.

Guerois, R., Nielsen, J.E., and Serrano, L. (2002). Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. J. Mol. Biol. 320, 369–387.

Hoppe, C., and Schomburg, D. (2005). Prediction of protein thermostability with a direction- and distance-dependent knowledge-based potential. Protein Sci. 14, 2682–2692.

Hughson, F.M., Barrick, D., and Baldwin, R.L. (1991). Probing the stability of a partly folded apomyoglobin intermediate by site-directed mutagenesis. Biochemistry 30, 4113–4118.

Jackson, S.E., and Fersht, A.R. (1991). Folding of chymotrypsin inhibitor-2. 1. Evidence for a 2-state transition. Biochemistry 30, 10428–10435.

Khare, S.D., Caplow, M., and Dokholyan, N.V. (2006). FALS mutations in Cu, Zn superoxide dismutase destabilize the dimer and increase dimer dissociation propensity: a large-scale thermodynamic analysis. Amyloid 13, 226–235.

Khatun, J., Khare, S.D., and Dokholyan, N.V. (2004). Can contact potentials reliably predict stability of proteins? J. Mol. Biol. 336, 1223–1238.

Kollman, P.A., Massova, I., Reyes, C., Kuhn, B., Huo, S.H., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., et al. (2000). Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Acc. Chem. Res. 33, 889–897.

Kortemme, T., and Baker, D. (2002). A simple physical model for binding energy hot spots in protein-protein complexes. Proc. Natl. Acad. Sci. USA 99, 14116–14121.

Kuhlman, B., and Baker, D. (2000). Native protein sequences are close to optimal for their structures. Proc. Natl. Acad. Sci. USA 97, 10383–10388.

Kuhlman, B., and Baker, D. (2004). Exploring folding free energy landscapes using computational protein design. Curr. Opin. Struct. Biol. 14, 89–95.

Lazaridis, T., and Karplus, M. (1999). Effective energy function for proteins in solution. Proteins Struct. Funct. Genet. 35, 133–152.

Matouschek, A., Kellis, J.T., Serrano, L., Bycroft, M., and Fersht, A.R. (1990). Transient folding intermediates characterized by protein engineering. Nature 346, 440–445.

Obara, M., Kang, M.S., and Yamada, K.M. (1988). Site-directed mutagenesis of the cell-binding domain of human fibronectin: separable, synergistic sites mediate adhesive function. Cell 53, 649–657.

Ota, M., Isogai, Y., and Nishikawa, K. (2001). Knowledge-based potential defined for a rotamer library to design protein sequences. Protein Eng. 14, 557–564.

Pakula, A.A., and Sauer, R.T. (1989). Genetic-analysis of protein stability and function. Annu. Rev. Genet. 23, 289–310.

Saraboji, K., Gromiha, M.M., and Ponnuswamy, M.N. (2006). Average assignment method for predicting the stability of protein mutants. Biopolymers 82, 80–92.

Saunders, C.T., and Baker, D. (2005). Recapitulation of protein family divergence using flexible backbone protein design. J. Mol. Biol. 346, 631–644.

Serrano, L., Kellis, J.T., Cann, P., Matouschek, A., and Fersht, A.R. (1992). The folding of an enzyme. 2. Substructure of barnase and the contribution of different interactions to protein stability. J. Mol. Biol. *224*, 783–804.

Vorobjev, Y.N., and Hermans, J. (1999). ES/IS: estimation of conformational free energy by combining dynamics simulations with explicit solvent with an implicit solvent continuum model. Biophys. Chem. *78*, 195–205.

Yin, S., Ding, F., and Dokholyan, N.V. (2007). Eris: an automated estimator of protein stability. Nat. Methods *4*, 466–467.

Zhou, H.Y., and Zhou, Y.Q. (2002). Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. Protein Sci. *11*, 2714–2726.