**Protein Engineering** vol.7 no.10 pp.1209–1220, 1994

# Protein stability for single substitution mutants and the extent of local compactness in the denatured state

**Sanzo Miyazawa and Robert L.Jernigan[1]**

Gunma University, Faculty of Technology, Kiryu, Gunma 376, Japan and [1]Laboratory of Mathematical Biology, DCBDC, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA

**The stability changes caused by single amino acid substitutions are studied by a simple, empirical method which takes account of the free energy change in the compact denatured state as well as in the native state. The conformational free energy is estimated from effective inter-residue contact energies, as evaluated in our previous study. When this method is applied, with a simple assumption about the compactness of the denatured state, for single amino acid replacements at Glu49 of the tryptophan synthase α subunit and at Ile3 of bacteriophage T4 lysozyme, the estimates of the unfolding Gibbs free energy changes correlate well with observed values, especially for hydrophobic amino acids, and it also yields the same magnitudes of energy as the observed values for both proteins. When it is also applied for amino acid replacements at various positions to estimate the average number of contacts at each position in the denatured state from the observed value of unfolding free energy change, those values for replacements with Gly and Ala at the same residue position in staphylococcal nuclease correlate well with each other. The estimated numbers of contacts indicate that the protein is not fully expanded in the denatured state and also that the compact denatured state may have a substantially native-like topology, like the molten globule state, in that there is a weak correlation between the estimated average number of contacts at each residue position in the denatured state and the number of contacts in the native structure. These results provide some further evidence that the inter-residue contact energies as applied here (i) properly reflect actual inter-residue interactions and (ii) can be considered to be a pairwise hydrophobicity scale. Also, the results indicate that characterization of the denatured state is critical to understanding the folding process.**

*Key words:* hydrophobic energy/inter-residue contact energy/native-like compact denatured state/protein folding/protein stability

## Introduction

Recent progress in molecular biology has made it possible to modify the gene for any protein, virtually without limits. However, intelligent protein design requires a quantitative understanding of the folding energetics and mechanisms of proteins. As originally pointed out by Kauzmann (1959), hydrophobic interactions are a principal force in leading to a condensed protein molecule. Hydrophobic energies have been evaluated by a wide variety of methods: the transfer energy of amino acids from organic solvents to water (Tanford, 1962; Nozaki and Tanford, 1971), the hydrophobic energy per interfacial area of residues exposed to water from the transfer energy of liquid hydrocarbons into water (Herman, 1972), the hydrophobic energy per surface area from the compressibilities of proteins (Lee, 1983), methods based on the accessible surface areas of each atom (Eisenberg and McLachlan, 1986; Ooi *et al.*, 1987; Oobatake and Ooi, 1989) and on the contact frequencies of amino acid pairs in protein structures (Miyazawa and Jernigan, 1985). Also, there have been many hydrophobic scales of amino acids compiled from protein structural data (Sweet and Eisenberg, 1983; Rose *et al.*, 1985; Cornette *et al.*, 1987). The contributions of hydrophobic energies to the energetics of protein folding have been discussed on the basis of those estimates of hydrophobic interactions (Chothia, 1976; Finney *et al.*, 1980; Oobatake and Ooi, 1989; Dill, 1990). In early discussions of folding energetics, the denatured state was simply presumed to be fully expanded. However, recently there has been additional evidence that there is substantial residual structure in both the denatured state and folding intermediates upon thermal, pH and even solvent denaturation [see Dill and Shortle (1991) and Ptitsyn (1994) for reviews]. Therefore, conformational characteristics and energetics of the denatured state must be studied to understand protein folding better. Hydrophobic interactions seem to be a dominant force not only in the native structure but also in the compact denatured state (Ptitsyn, 1987; Dill, 1990; Dill and Shortle, 1991).

Recent studies of protein stability with amino acid replacements strongly support the conclusion that hydrophobic interactions play a major role. Yutani *et al.* (1984, 1987) found that the stabilities of the mutant proteins of tryptophan synthase α subunit with single amino acid replacements at Glu49 tended to increase linearly with the hydrophobicity of the substituting residues. Such a correlation was also confirmed for single amino acid replacements at Ile3 in T4 lysozyme by Matsumura *et al.* (1988). However, unlike Ile3 replacements in T4 lysozyme, in Glu49 replacements of tryptophan synthase α subunit the unfolding Gibbs free energies of mutant proteins changed far more than expected from their estimates of hydrophobic energy changes caused by the single amino acid substitutions. In other words, although their data showed that the stabilities of the mutant proteins were related to the change of hydrophobic interactions, their estimates of the hydrophobic interactions were not sufficiently large to explain the observed changes of unfolding Gibbs free energy in the mutant proteins. Shortle *et al.* (1990) also concluded that when the average stability loss caused by Gly replacements for each of six types of residue (Leu, Val, Tyr, Ile, Met and Phe) was compared with the transfer free energy of the side chain of that residue, the value of the unfolding free energy change was two to three times larger than expected if hydrophobicity were the sole source of stability. Kellis *et al.* (1989) suggested that cavity formation caused by the replacement of a bulky side chain by a smaller side chain would destabilize the native state relative to the denatured state. Such unaccounted factors require explanation, even though it is not easy to estimate the stability of protein native structures with calculations.

To determine the stability of proteins, it is necessary to know the free energies of both the native and denatured states, because stability is defined as their difference. If the native structure of a protein were known then, in principle, semi-empirical computational methods (Dang *et al.*, 1989; Tidor and Karplus, 1991) could be employed to estimate the change to the conformational free energy of the protein structure in water caused by single amino acid substitutions. On the other hand, it is extremely difficult to estimate the free energy change of the denatured state because an estimate of the change in the conformation entropy is needed. Usually the free energy of the denatured state is assumed to remain constant whenever single amino acid substitutions are considered. Otherwise, the denatured state was assumed to be fully solvated (Sneddon and Tobias, 1992). However, such an assumption may introduce a substantial uncertainty into any attempted calculation of the stability change of the native structure. Recent data indicate that the denatured state is rather compact (Ahmad and Bigelow, 1986; Pace *et al.*, 1990) and also replacing single amino acid residues can affect the denatured states of proteins (Shortle and Meeker, 1989; Shortle *et al.*, 1990; Dill and Shortle, 1991).

In this paper we present a simple, empirical method to estimate the stability change of protein structures by single amino acid substitutions. In a previous study (Miyazawa and Jernigan, 1985) we estimated the effective inter-residue contact energies for proteins in solution from the numbers of residue–residue contacts observed in crystal structures of globular proteins by means of the quasi-chemical approximation with an approximate treatment of the effects of chain connectivity. This empirical energy function includes solvent effects, and provides an estimate of the long-range component of conformational energies. From the contact energies for residues in contact with a substituted residue, we can evaluate the energy increment for the native structure and then the change in the stability of the native structure together with a simple approximation for the free energy change of the denatured state. Three cases are considered, corresponding to the available data. The dependence of protein stability on residue type at residue position 49 in tryptophan synthase α subunit (Yutani *et al.*, 1987) and also at residue position 3 in T4 lysozyme (Matsumura *et al.*, 1988) is estimated with an assumption about the compactness of the denatured state for these proteins; the average number of contacts at these residue positions in the denatured state is approximated by that value which is expected in a hypothetical case of no inter-residue interactions. Because these analyses support the present approximation, the average numbers of contacts at some residue positions in the denatured state of staphylococcal nuclease are estimated from the experimental data (Shortle *et al.*, 1990) of stability changes for single amino acid replacements at various residue positions in this protein; here it should be noted that the denatured state means the non-native state for which the unfolding Gibbs free energies are measured. Results are consistent with the idea of the compact denatured state and especially the 'molten globule' state (Ptitsyn, 1987; Kuwajima, 1989) which has a significantly native-like configuration.

## Materials and methods

*Contact energy*

Previously we estimated the effective inter-residue contact energies for proteins in solution from the numbers of residue–residue contacts observed in crystal structures of globular proteins by means of the quasi-chemical approximation. The basic assumption was that for a large enough sample the average characteristics of residue–residue contacts formed in a large number of protein crystal structures would reflect the actual differences in the interactions among residues, as if there were no significant contributions from the specific amino acid sequence in each protein as in intra-residue and short-range interactions. By employing a lattice model, each residue of a protein was assumed to occupy a site in a lattice, and vacant sites were regarded to be occupied by an effective solvent molecule whose size equals the average size of a residue. Then, taking account of the effects of the chain connectivity as imposing a limit to the size of the system, i.e. the number of lattice sites, the system was regarded as a mixture of unconnected residues and effective solvent molecules. The quasi-chemical approximation, that contact pair formation resembles a chemical reaction, is applied to this system to obtain formulae to relate the statistical averages of the numbers of contacts to contact energies. Each residue was represented by the center of its side-chain position, and contacts among residues were defined to be those within 6.5 Å. Following the notation in the previous study, $e_{i,j}$ represents the contact energy between $i$ and $j$ types of residue previously calculated between all 20 types of amino acid, $n_i$ is the number of $i$ types of residue, and $n_{ii}$ and $2n_{ij}$ ($= n_{ij} + n_{ji}$) are the number of contacts between two residues of the same $i$ type and between $i$ and $j$ types of residue. The values of $n_i$ and $n_{ij}$ were derived from 42 crystal structures of globular proteins, including 30 monomeric proteins. $N_i$ and $N_{ij}$ represent the sums of $n_i$ and $n_{ij}$ over all proteins, respectively. The total number of residue–residues contacts, $N_{rr}$ ($= \Sigma N_{ij}$), was 18 192 and the total number of residues, $N_r$, was 9040; in the case of oligomeric molecules, the numbers of contacts were calculated in the oligomeric state and averaged over subunits. The estimates of contact energies reflected well the expected physico-chemical properties of residues. Further, the estimates of transfer energies for hydrophobic residues from the outside to the interior of a protein showed a linear relationship with the transfer energies from water to non-polar solvent estimated by Nozaki and Tanford (1971). However, the estimated values were twice as large as those of Nozaki and Tanford.

*Unfolding Gibbs free energy change*

The change $[\Delta(\Delta_d G)_{i\to j}]$ of the unfolding Gibbs free energy, which is needed to unfold the native structure, expected when the $i$ type of residue at the specific position $p$ is substituted with the $j$ type of residue in a protein, can be approximated in terms of these contact energies as follows:

$$\Delta(\Delta_d G)_{i\to j} \approx -\Sigma_{kl}[e_{kl}(\Delta n_{kl} - \Delta n^d_{kl})]. \quad (1)$$

[See Miyazawa and Jernigan (1985), especially their equation 4a, for details of this formulation.] The sums over $k$ and $l$ are each over all 20 types of residue and $\Delta n_{kl}$ and $\Delta n^d_{kl}$ are the changes in the number of $k$–$l$ contacts caused by the substitution in the native and denatured states, respectively. Here it is assumed that the conformational entropy change by a single amino acid replacement is negligible in both the denatured and native states. If the conformation of the native structure was not changed significantly by a single amino acid replacement, then the energy change of the native state could be more simply calculated by:

$$\Sigma_{kl} e_{kl}(\Delta n_{kl})) \approx 2\Sigma_k (e_{jk} - e_{ik}) n_{p;ik}, \quad (2)$$

where $2n_{p;ik}$ is the number in the native structure of $k$ type

residues surrounding the $i$ type residue at sequence position $p$. Of course, if unfavorable energies were to arise through such a replacement, local or large-scale conformational changes would presumably occur insofar as possible to remove unfavorable interactions (Eriksson *et al.*, 1992). Also, even though favorable contact energies are expected, unfavorable interactions which are not taken into account, such as steric hindrance, may decrease the favorable effects; such effects may be alleviated substantially by small shifts below the resolution of the method.

The second term of Equation 1 is approximated in the simplest possible way as follows. The denatured state of proteins is a large ensemble of relatively expanded conformations, compared with their unique native structures. Therefore, contacting residue pairs in the denatured state are not unique, but each residue pair has instead a certain probability for contacts. Let us represent the average number of contacts at residue position $p$ in the denatured state by $2n^{d}_{p;jk}$, where $j$ is the residue type at position $p$ and $k$ is the type of residue in contact. Here we assume that single amino acid mutations do not significantly affect this overall distribution except for that at residue position $p$. With this approximation, the second term in Equation 1 becomes:

$$\Sigma_{kl} e_{kl}(\Delta n^{d}_{kl}) \approx 2\Sigma_{k}(e_{jk} n^{d}_{p;jk} - e_{ik} n^{d}_{p;ik}). \quad (3)$$

Next, the distribution of contacts over residue pairs in the denatured state is taken to be the same as that observed in the native structures, although interior residues must have fewer contacts and exterior residues tend to have more contacts in the denatured state than in the native structure. That is,

$$n^{d}_{p;ik} \approx (N_{ik}/N_{rr})(N_r/N_i) n^{d}_{p} \quad (4)$$

is assumed, where $N_{ij}$ is taken to be the total number of contacts observed in the set of 42 protein crystals (Miyazawa and Jernigan, 1985). $N_i$ is the number of residues of type $i$, and $N_{ir}$ and $N_{rr}$ are defined as:

$$\begin{aligned} N_{ir} &= \Sigma_j N_{ij} \\ N_{rr} &= \Sigma_i N_{ir} \\ N_r &= \Sigma_i N_i, \end{aligned} \quad (5)$$

where the subscript $r$ represents any residue. $2n^{d}_{p}$ is the average number of contacts between the residue at position $p$ and other residues in the denatured state. Then Equation 3 becomes:

$$\Sigma_{kl} e_{kl}(\Delta n^{d}_{kl}) \approx 2(f_j - f_i) n^{d}_{p} \quad (6)$$

where

$$\begin{aligned} f_i &= \Sigma_k [e_{ik}(N_{ik}/N_{ir})](N_{ir}/N_{rr})(N_r/N_i) \quad (7) \\ &= e_i(N_{ir}/N_{rr})(N_r/N_i), \end{aligned}$$

where $e_i$ is the average contact energy for the $i$ type of amino acid. The values of $e_i$, $N_{ij}$ and $N_i$ are shown in Tables IV and V in Miyazawa and Jernigan (1985). As a result, the actual change of the unfolding Gibbs free energy, with this assumption that the distribution of contact pairs in the denatured state resembles that of the native state, will then be:

$$\Delta(\Delta_d G)_{i \to j} \approx -2[\Sigma_k(e_{jk} - e_{ik}) n_{p;ik} - (f_j - f_i) n^{d}_{p}]. \quad (8)$$

To estimate the unfolding free energy changes from Equation 8, we must know $2n^{d}_{p}$, the average number of contacts at each site in the denatured state. A lower boundary for $n^{d}_{p}$ corresponds to having no inter-residue interactions, i.e. $e_{jk} = 0$. The average number of contacts per residue in such a hypothetical state was estimated previously as 0.817–0.870 for proteins of size $100 < n_r < 300$ (Miyazawa and Jernigan, 1985). Therefore, we take a rough estimate of $n^{d}_{p}$ to be 0.85. Here it may be useful to know that the number of contacts per residue in the native structure ranges from ~1.6 to 2.2 on average, depending on sequence length and protein shape. Conversely, it is possible to estimate $n^{d}_{p}$ as a value which best fits Equation 8, i.e. $n^{d}_{p}$ may be estimated from the observed values of unfolding free energies together with the predicted values of the free energy changes for the native structure. In this paper, $n^{d}_{p}$ is assumed to be 0.85 in predicting the dependence of unfolding free energy changes on residue type, but in discussions of their dependence on residue position it will be treated instead as a parameter to be estimated.

Tanford (1970) developed a model of solvent denaturation in which the difference between unfolding free energy changes in protein denaturation with and without denaturant is proportional to the fractional change in solvent exposure on protein denaturation. Based on this model, Ahmad and Bigelow (1986) and also Pace *et al.* (1990) determined the average fractional increases, $\Delta\alpha$, in solvent exposure of residues for guanidinium hydrochloride (GuHCl) and urea denaturations for several proteins. These values range from 0.14 to 0.53, depending on the protein and experimental conditions. From these values, together with the average exposure of residues in protein structures, the average value of $n^{d}_{p}$ over residue positions may be estimated.

The first term in Equation 8 can be calculated easily if the native structure of a protein is known. In case the native structure of a protein is unknown, the best estimate of unfolding energy change can be obtained by replacing $n_{p;ik}$ by its average. The average of $n_{p;ik}$ over all occurrences of $i$ type residues in a given set of proteins is:

$$<n_{p;ik}> = N_{ik}/N_i. \quad (9)$$

$N_{ik}/N_i$ represents the distribution of the types of residues surrounding an $i$ type residue, that is, it represents the mean field of residues surrounding a specific type of amino acid in the native structures of proteins. In the case of Equation 2, the values of $n_{p;ik}$ reflect the immediate environment of residues centered around an $i$ type residue at position $p$; whereas the average in Equation 9 can be a last resort, if no details of a specific structure are known, to crudely estimate the typical environment of a residue of type $i$.

*Statistical tests*

In this manuscript, the significance of a correlation coefficient is tested by calculating a $t$ value as follows:

$$t = r[(n' - 2)/(1 - r^2)]^{\frac{1}{2}}, \quad (10)$$

where $n'$ is the number of samples and $r$ is the value of the correlation coefficient. The probability $P$ that better correlations are obtained from uncorrelated populations is equal to the probability of falling in a range greater than the value of $t$ above in the Student's $t$ distribution for the degrees of freedom $n = n' - 2$ (see Fisher, 1970).

A $t$ value for the significance of the difference between the estimated value $b$ and hypothesis $\beta$ of the slope of a regression line, $Y = a + b(x - <x>)$, is:

$$t = (b - \beta)[\Sigma(x_i - <x>)^2]^{\frac{1}{2}}/s, \quad (11)$$

and that for the significance of the difference between the estimated value $a$ and hypothesis $\alpha$ is:

$$t = (a - \alpha)(n')^{\frac{1}{2}}/s, \quad (12)$$

**Table IA.** The average contact energy change by substituting the $i$ type (columns) by the $j$ type (rows) of amino acid: $2\{\Sigma_k[(e_{kj} - e_{ki})N_{ki}/N_i] - (f_j - f_i)0.85\}$[a]

| | Cys | Met | Phe | Ile | Leu | Val | Trp | Tyr | Ala | Gly | Thr | Ser | Gln | Asn | Glu | Asp | His | Arg | Lys | Pro |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cys | 0.00 | 3.09 | 3.41 | 2.83 | 2.18 | 1.22 | 2.46 | 1.26 | 0.67 | 1.04 | 1.03 | 1.11 | 1.83 | 1.77 | 2.81 | 2.34 | 1.14 | 1.68 | 3.37 | 2.33 |
| Met | –1.45 | 0.00 | 0.80 | 0.08 | –0.51 | –1.18 | 0.12 | –0.45 | –0.61 | 0.24 | 0.19 | 0.49 | 1.34 | 1.34 | 2.36 | 2.02 | 0.07 | 1.08 | 3.62 | 2.03 |
| Phe | –2.41 | –0.46 | 0.00 | –0.64 | –1.21 | –1.83 | –0.51 | –0.92 | –0.99 | –0.05 | –0.14 | 0.17 | 1.18 | 1.11 | 2.23 | 1.85 | –0.43 | 0.76 | 3.45 | 1.80 |
| Ile | –1.35 | 0.30 | 0.81 | 0.00 | –0.53 | –1.22 | 0.21 | –0.26 | –0.63 | 0.27 | 0.21 | 0.49 | 1.45 | 1.46 | 2.48 | 2.11 | 0.16 | 1.14 | 3.56 | 2.06 |
| Leu | –0.96 | 0.86 | 1.35 | 0.62 | 0.00 | –0.70 | 0.72 | 0.06 | –0.39 | 0.43 | 0.37 | 0.59 | 1.48 | 1.46 | 2.39 | 2.06 | 0.36 | 1.26 | 3.42 | 2.02 |
| Val | –0.04 | 1.77 | 2.21 | 1.49 | 0.87 | 0.00 | 1.40 | 0.49 | –0.18 | 0.48 | 0.42 | 0.54 | 1.29 | 1.31 | 2.18 | 1.87 | 0.54 | 1.14 | 2.86 | 1.77 |
| Trp | –1.45 | 0.43 | 0.94 | 0.29 | –0.31 | –1.12 | 0.00 | –0.81 | –1.03 | –0.41 | –0.52 | –0.27 | 0.51 | 0.47 | 1.42 | 1.08 | –0.53 | 0.23 | 2.41 | 1.08 |
| Tyr | 0.31 | 2.42 | 2.90 | 2.35 | 1.62 | 0.58 | 1.61 | 0.00 | –0.56 | –0.34 | –0.48 | –0.48 | 0.03 | –0.05 | 0.63 | 0.38 | –0.27 | –0.06 | 1.14 | 0.48 |
| Ala | 1.78 | 3.88 | 4.28 | 3.59 | 2.87 | 1.71 | 3.04 | 1.28 | 0.00 | 0.17 | 0.14 | –0.02 | 0.41 | 0.42 | 1.04 | 0.78 | 0.66 | 0.52 | 1.03 | 0.75 |
| Gly | 2.27 | 4.79 | 5.19 | 4.54 | 3.80 | 2.49 | 3.60 | 1.46 | 0.21 | 0.00 | 0.00 | –0.20 | 0.11 | 0.06 | 0.68 | 0.35 | 0.66 | 0.25 | 0.53 | 0.46 |
| Thr | 2.54 | 4.84 | 5.20 | 4.60 | 3.83 | 2.55 | 3.61 | 1.44 | 0.26 | 0.09 | 0.00 | –0.18 | 0.11 | 0.05 | 0.56 | 0.30 | 0.63 | 0.24 | 0.47 | 0.45 |
| Ser | 3.08 | 5.55 | 5.85 | 5.28 | 4.49 | 3.15 | 4.26 | 1.93 | 0.59 | 0.37 | 0.31 | 0.00 | 0.27 | 0.19 | 0.68 | 0.45 | 0.95 | 0.42 | 0.35 | 0.56 |
| Gln | 3.20 | 5.54 | 5.98 | 5.42 | 4.62 | 3.24 | 4.23 | 1.80 | 0.56 | 0.27 | 0.18 | –0.10 | 0.00 | –0.02 | 0.36 | 0.17 | 0.79 | 0.27 | 0.18 | 0.37 |
| Asn | 3.50 | 5.89 | 6.26 | 5.83 | 4.99 | 3.62 | 4.52 | 1.98 | 0.81 | 0.43 | 0.33 | 0.01 | 0.14 | 0.00 | 0.47 | 0.26 | 0.90 | 0.31 | 0.15 | 0.43 |
| Glu | 3.71 | 5.59 | 6.03 | 5.56 | 4.67 | 3.34 | 4.27 | 1.70 | 0.58 | 0.30 | 0.11 | –0.17 | –0.09 | –0.16 | 0.00 | –0.06 | 0.71 | 0.23 | 0.04 | 0.24 |
| Asp | 3.59 | 6.20 | 6.62 | 6.12 | 5.26 | 3.86 | 4.74 | 2.03 | 0.83 | 0.35 | 0.25 | –0.05 | 0.03 | –0.05 | 0.23 | 0.00 | 0.96 | 0.37 | 0.15 | 0.37 |
| His | 1.37 | 3.91 | 4.25 | 3.84 | 3.05 | 1.83 | 2.90 | 0.82 | –0.01 | –0.04 | –0.14 | –0.31 | 0.02 | –0.08 | 0.62 | 0.33 | 0.00 | –0.08 | 0.61 | 0.44 |
| Arg | 3.11 | 5.51 | 5.85 | 5.37 | 4.60 | 3.20 | 4.15 | 1.75 | 0.64 | 0.33 | 0.26 | –0.02 | 0.20 | 0.08 | 0.81 | 0.49 | 0.67 | 0.00 | 0.20 | 0.46 |
| Lys | 5.62 | 7.64 | 7.87 | 7.37 | 6.57 | 5.07 | 6.10 | 3.32 | 1.69 | 1.28 | 1.18 | 0.65 | 0.76 | 0.60 | 1.06 | 0.92 | 1.86 | 0.81 | 0.00 | 0.90 |
| Pro | 2.45 | 4.82 | 5.17 | 4.65 | 3.85 | 2.53 | 3.52 | 1.28 | 0.12 | –0.08 | –0.17 | –0.41 | –0.19 | –0.27 | 0.26 | 0.00 | 0.42 | –0.07 | 0.03 | 0.00 |

**Table IB.** The average number of residues surrounding the $i$ type amino acid, $2N_{ir}/N_i$, and the effective contact energy, $f_i$, in $RT$ units for the denatured state

| | Cys | Met | Phe | Ile | Leu | Val | Trp | Tyr | Ala | Gly | Thr | Ser | Gln | Asn | Glu | Asp | His | Arg | Lys | Pro |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $2N_{ir}/N_i$ | 5.54 | 5.58 | 5.47 | 5.54 | 5.44 | 5.20 | 5.21 | 4.64 | 4.05 | 3.76 | 3.76 | 3.50 | 3.21 | 3.23 | 2.98 | 3.08 | 4.06 | 3.48 | 2.34 | 3.02 |
| $f_i$ | –5.51 | –6.81 | –6.95 | –6.72 | –6.28 | –5.39 | –5.64 | –3.73 | –2.84 | –2.18 | –2.15 | –1.80 | –1.58 | –1.53 | –1.43 | –1.39 | –2.78 | –1.88 | –0.87 | –1.66 |

[a]These values are represented in dimensionless form in $RT$ units.

**Table II.** The unfolding Gibbs free energy changes for 18 single amino acid replacements at Glu49 of tryptophan synthase α subunit

| | $\Delta\Delta_dG$ in water[a] (25°C) | | $-\Delta G_{tr}$[b] | $-\Delta$(contact energy) at Glu49[c] | | $-\Delta\Delta_d$(contact energy)[c] | |
|---|---|---|---|---|---|---|---|
| | pH 7.0 | pH 9.0 | | native[d] | denatured[e] | net[f] | typical Glu[g] |
| Cys | 2.2 | 3.4 | | 8.0 | 4.1 | 3.9 | –1.7 |
| Met | 4.5 | 3.5 | 1.3 | 11.8 | 5.4 | 6.4 | –1.4 |
| Phe | 2.4 | 3.4 | 2.5 | 12.2 | 5.5 | 6.7 | –1.3 |
| Ile | 8.0 | 5.1 | | 11.5 | 5.3 | 6.2 | –1.5 |
| Leu | 6.2 | 7.3 | 1.8 | 10.4 | 4.9 | 5.6 | –1.4 |
| Val | 3.2 | 4.5 | 1.5 | 8.8 | 4.0 | 4.8 | –1.3 |
| Trp | 1.1 | 0.8 | 3.4 | 9.7 | 4.2 | 5.5 | –0.8 |
| Tyr | 0.0 | 1.9 | 2.3 | 5.4 | 2.3 | 3.1 | –0.4 |
| Ala | –0.3 | 1.9 | 0.5 | 3.8 | 1.4 | 2.4 | –0.6 |
| Gly | –1.7 | 1.5 | 0.0 | 2.2 | 0.8 | 1.4 | –0.4 |
| Thr | 0.0 | 2.1 | 0.4 | 2.1 | 0.7 | 1.3 | –0.3 |
| Ser | –1.4 | 3.1 | –0.3 | 0.9 | 0.4 | 0.6 | –0.4 |
| Gln | –2.5 | 3.6 | | 0.5 | 0.2 | 0.4 | –0.2 |
| Asn | –0.6 | 1.3 | | 0.4 | 0.1 | 0.3 | –0.3 |
| Glu | 0.0 | 0.0 | | 0.0 | 0.0 | 0.0 | 0.0 |
| Asp | –0.3 | 2.1 | | –0.3 | 0.0 | –0.3 | –0.1 |
| His | 1.3 | 4.3 | 0.5 | 3.4 | 1.4 | 2.1 | –0.4 |
| Lys | –0.9 | 2.6 | | –1.7 | –0.6 | –1.2 | –0.6 |
| Pro | –0.6 | 2.0 | | 1.3 | 0.2 | 1.0 | –0.2 |

[a] In kcal/mol; taken from Yutani *et al.* (1987); the unfolding free energy of the wild type (Glu) is 8.8 kcal/mol.
[b]The free energy in kcal/mol of transfer of a residue from water to organic solvent taken from Nozaki and Tanford (1971).
[c]In kcal/mol; $RT$ is assumed to be 0.59 kcal/mol.
[d]The first term in Equation 8; Glu49 is in contact with the Phe22, Gly98, Leu100, Leu127, Tyr175, Ile232 and Gly234; the coordinate set 1WSY in the Protein Data Bank is used.
[e]The second term in Equation 8 with $n^d_p = 0.85$.
[f]The expected values of unfolding Gibbs free energy change.
[g]Calculated from Table I, as if there was no structure.

where

$$s^2 = \Sigma(y_i - Y_i)^2/(n' - 2).$$

These $t$ values are for the Student's $t$ distribution of the degrees of freedom $n = n' - 2$ (see Fisher, 1970).

## Results

Table IA shows the average contact energy changes expected for single amino acid replacements which are calculated by Equation 8 with Equation 9 and $n^d_p = 0.85$; the dimensionless

energy units are in units of $RT$, where $R$ is the gas constant and $T$ is the absolute temperature. Values reflect the net change occurring in both the denatured and native states. The average numbers of residues surrounding an $i$ type amino acid and the values of $f_i$ are listed in Table IB. On average, even substituting a hydrophilic residue by a hydrophobic one is expected to decrease the total contact energy of the native structure. However, it is possible that such a substitution could also diminish the free energy of the denatured state if there were more contacts formed in the denatured state. Table IA shows that the net effect of such amino acid replacements tends to make a protein unstable, because a hydrophobic residue replacing a hydrophilic one tends to be more exposed at the surface of the native structure than it is in the denatured state. In an extreme case, replacing hydrophilic residues by hydrophobic ones could of course result in an aggregation of proteins or a rearrangement of the hydrophobic core, although this is not so likely for single amino acid replacements. Such effects are of course outside the range of the present approach, and we are limiting considerations to small changes only.

Yutani *et al.* (1987) measured the stabilities of mutant proteins of tryptophan synthase α subunit from *Escherichia coli* in which the Glu49 is replaced by each of 18 other amino acids. (The mutant protein with Arg could not be obtained in large enough quantities for analysis.) Table II shows the changes of the unfolding Gibbs free energy in water caused by those single amino acid substitutions. A positive correlation was found between the unfolding energy changes and hydrophobicities of the substituted residue, and it was claimed that the stabilities of those mutant proteins tended to increase linearly with increasing hydrophobicity of the substituted residue, unless the volume of the substituted residue exceeded a certain limit. They employed two scales of hydrophobicity, the transfer energy of a residue from water to organic solvents taken from Nozaki and Tanford (1971) and the OMH empirical scale devised by Sweet and Eisenberg (1983). As Yutani *et al.* (1987) pointed out, the correlation coefficient between the unfolding energy changes at pH 7.0 and the Nozaki–Tanford transfer energy values is 0.95 if the aromatic residues Phe, Tyr and Trp are excluded, and only Met, Leu, Val, Ala, Gly, Thr, Ser and His are taken into account (see Figure 1A). This correlation is significant, with the probability $P = 0.0002$ for the degrees of freedom $n = 6$. Thus, it indicates that the unfolding energy changes of those mutant proteins result from changes in the hydrophobic interactions. However, the unfolding energy changes were much larger than the hydrophobic energy changes expected from the transfer energy of each amino acid; the slope of the regression line is 3.7. In other words, the Nozaki–Tanford hydrophobicity scale accounts for <30% of the unfolding energy changes. Here it should be noted that the hydrophobic energy changes estimated by Yutani *et al.* (1987) are those expected for completely buried residues and therefore correspond to the maximum hydrophobic energy changes expected for single amino acid replacements.

The contact energy changes expected for single amino acid replacements are given in Table II and are plotted against the observed values of the unfolding Gibbs free energy changes in Figure 1B. The protein structure of tryptophan synthase α subunit is known, so that the contact energy change at the 49th position is calculated directly from Equation 2. On average, Glu is surrounded by about three residues within 6.5 Å of the center of its side chain, because Glu tends to be located on the surface of proteins (Table IB). However, the

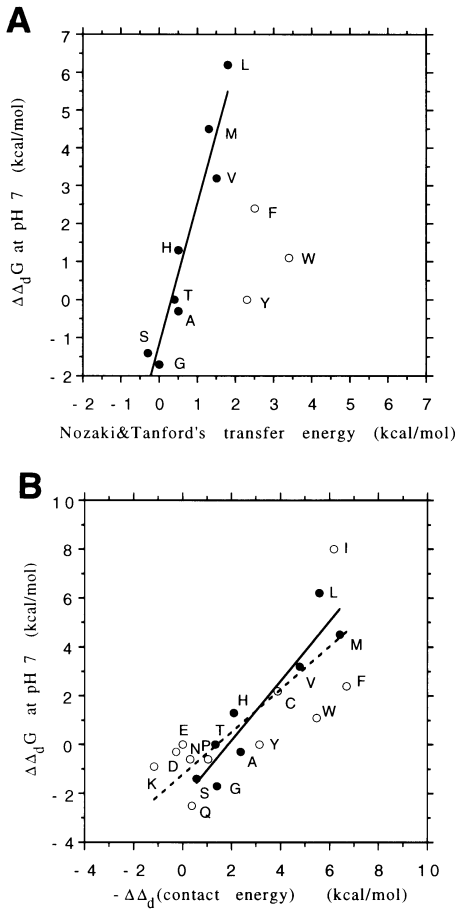


**Fig. 1.** (**A**) Unfolding Gibbs free energy changes at pH 7 of the mutant tryptophan synthase α subunit with single amino acid substitutions at Glu49 (Yutani *et al.*, 1987) are plotted against the transfer energies of residues from water to organic solvent (Nozaki and Tanford, 1971). The solid line indicates a regression line, $y = -1.2 + 3.7x$, for the eight amino acids shown by solid circles (Leu, Met, Val, His, Thr, Ala, Ser and Gly). The correlation coefficient is 0.95; $P = 0.0002$. Amino acids omitted in this correlation analysis are the aromatic residues Phe, Trp and Tyr. (**B**) The same unfolding Gibbs free energy changes shown in (A) are plotted against the contact energy changes due to the replacements calculated by Equation 8 with $n^d_p = 0.85$ (Table II, next to last column). The solid line indicates a regression line, $y = -2.3 + 1.2x$, for the eight amino acids employed in (A) and shown by solid circles. The same eight amino acids only are chosen for comparison with (A). The correlation is as good as in (A); the correlation coefficient is 0.93; $P = 0.0008$. Even in the case of including all 19 amino acids shown by both solid and open circles, the correlation is still good, with a correlation coefficient of 0.81; $P = 0.00003$. The regression line, $y = -1.2 + 0.88x$, for this case is shown by the dotted line. The single-letter codes in the figures indicate the amino acids substituted at position 49 in the mutant proteins. The numerical values of the quantities displayed in these figures are all given in Table II.

Glu49 of interest is uncharacteristically buried in the interior of protein and is surrounded by seven residues. Therefore, the contact energy changes expected for the replacements of Glu49 are completely different from the average contact energy changes expected for a typical Glu (Table II, last two columns). The replacement of Glu49 by a hydrophobic residue would make the protein structure more stable. This experimental fact is correctly predicted in the present estimates of unfolding Gibbs free energy change. The correlation coefficient is 0.93 with $P = 0.0008$ if only the eight amino acids Met, Leu, Val, Ala, Gly, Thr, Ser and His are included. Even if all 19 amino

**Table III.** The unfolding Gibbs free energy changes for 14 single amino acid replacements at Ile3 of bacteriophage T4 lysozyme

| | $\Delta\Delta_dG$ in water[a] | | $-\Delta G_{tr}$[b] | $-\Delta$(contact energy) at Ile3[c] | | $-\Delta\Delta_d$(contact energy)[c] | |
|---|---|---|---|---|---|---|---|
| | pH 2.0 | pH 6.5 | | native[d] | denatured[e] | net[f] | typical Ile[g] |
| Cys(S–S) | 1.0 | 1.2 | 1.00 | –2.5 | –1.2 | –1.3 | –1.7 |
| Cys(SH) | –0.4 | –1.2 | 1.00 | –2.5 | –1.2 | –1.3 | –1.7 |
| Met | –0.3 | –0.9 | 1.30 | –0.2 | 0.1 | –0.3 | 0.0 |
| Phe | –1.0 | –1.1 | 2.65 | 0.8 | 0.2 | 0.6 | 0.4 |
| Ile | 0.0 | 0.0 | 2.97 | 0.0 | 0.0 | 0.0 | 0.0 |
| Leu | 0.9 | 0.4 | 2.42 | –0.6 | –0.4 | –0.1 | –0.4 |
| Val | –0.6 | –0.4 | 1.69 | –2.0 | –1.3 | –0.7 | –0.9 |
| Trp | –3.6 | –2.8 | 3.00 | –1.2 | –1.1 | –0.1 | –0.2 |
| Tyr | –2.7 | –2.3 | 2.87 | –4.5 | –3.0 | –1.5 | –1.4 |
| Ala | –1.1 | –0.7 | 0.73 | –5.9 | –3.9 | –2.0 | –2.1 |
| Gly | –1.8 | –2.1 | 0.00 | –7.2 | –4.6 | –2.7 | –2.7 |
| Thr | –1.7 | –2.3 | 0.44 | –7.4 | –4.6 | –2.9 | –2.7 |
| Ser | –1.9 | –1.7 | 0.04 | –8.0 | –4.9 | –3.1 | –3.1 |
| Glu | –1.1 | –2.0 | 0.55 | –9.1 | –5.3 | –3.8 | –3.3 |
| Asp | –1.8 | –3.2 | 0.54 | –9.5 | –5.3 | –4.1 | –3.6 |

[a]In kcal/mol; taken from Matsumura *et al.* (1988); the values of $\Delta\Delta_dG$ were calculated from changes in melting temperature.
[b] The free energy of transfer in kcal/mol of a residue from water to organic solvent taken from Tanford (1962), except Cys(S–S) which is from Cantor *et al.* (1980).
[c]In kcal/mol; *RT* is assumed to be 0.59 kcal/mol.
[d]The first term in Equation 8; Ile3 is in contact with the Met6, Leu7, Val71, Cys97 and Ile100; the coordinate sets 2LZM and 3LZM in the Protein Data Bank are used.
[e]The second term in Equation 8 with $n^d_p = 0.85$.
[f]The expected value of unfolding Gibbs free energy change.
[g]Calculated from Table I as if there was no structure.

acids are considered, the correlation is still good, with a correlation coefficient of 0.81 and $P = 0.00003$. This is not a surprising fact because the contact energies evaluated in Miyazawa and Jernigan (1985) are effective empirical energies that include solvent effects, hydrophobic interactions and all other interactions occurring in a representative set of protein structures. On the other hand, these empirical energies may be good for the crystal pH. It may be one reason why the correlation is not so good between the unfolding Gibbs energy changes at pH 9.0 and the contact energy changes.

It is interesting (Figure 1B) that almost all of the unfolding Gibbs free energy changes can be explained by the change of contact energy. A regression line is $y = -2.3 + 1.2x$ for the eight amino acids, and $y = -1.2 + 0.88x$ for all amino acids. On the contrary, the Nozaki–Tanford hydrophobic energies can account for <30% of the magnitude of the observed values. The comparison of our inter-residue contact energies with the Nozaki–Tanford hydrophobic energies indicates that the former are about twice as large as the latter (see Miyazawa and Jernigan, 1985). Figure 1B suggests that our estimates of inter-residue contact energies are more appropriate.

A similar experiment on bacteriophage T4 lysozyme was performed by Matsumura *et al.* (1988). They measured the unfolding Gibbs free energies of mutant proteins in which Ile at position 3 was replaced by each of 14 other amino acids. Table III shows their data together with the transfer energies (Tanford, 1962) of amino acids from water to ethanol. Ile3 is a rather typical Ile because it is surrounded within 6.5 Å by five residues, which is near the average of 5.54 for Ile.

A comparison of the changes of the unfolding energy with the Tanford transfer energies is shown in Figure 2A, and again they are compared with our contact energy changes in Figure 2B. In both figures, the points for cysteine [Cys(S–S)] and Trp fall far off the linear correlation. The exception of cysteine is understandable because neither the transfer energy nor the contact energy can properly reflect the large effect of cysteine with a S–S bridge in increasing the stability of proteins. However, why is the value for Trp an exception? Some conformational change may occur in this mutant protein. As Matsumura *et al.* (1988) pointed out, there is a good correlation in Figure 2A if the aromatic residues Phe, Tyr and Trp and the ionizable residues Glu and Asp are excluded as well as Cys(S–S). However, even if Phe, Tyr, Glu and Asp are included, the correlation is still good with a correlation coefficient of 0.79 and $P = 0.001$, as shown in Figure 2B. An important fact in Figure 2 is that unlike the Glu49 replacements in the tryptophan synthase α subunit, which have a large value of 3.7 for the slope of the regression line, the slope of the regression line between the unfolding Gibbs free energy and the transfer energy is 0.80 in the case of Ile3 replacements in T4 lysozyme. It must be explained why the slopes of the regression lines differ so much in the two cases. On the other hand, the present approach gives a narrower range of values for the slopes of the regression lines of the observed values of unfolding Gibbs free energies with their crude estimates: 0.66 for the nine amino acids shown by solid circles or 0.53 for the 13 amino acids shown by the solid and open circles in Figure 2B, and 1.2 or 0.88 for the Glu49 replacements in the tryptophan synthase α subunit, as shown in Figure 1B. This narrower range of values for the slopes may result from the fact that the amino acid types and the number of residues surrounding a residue of interest in a protein structure are taken into account in estimating the hydrophobic energy changes in the native structure accompanied with single amino acid replacements, and also the free energy change in the denatured state is crudely estimated and included in Equation 8. The result is a more consistent view.

So far we have assumed an approximate value for the average number of contacts, $n^d_p$, in the denatured state. This does not imply that $n^d_p$ cannot depend on residue position. The slopes and intercepts of the regression lines in Figures 1B and 2B depend on the value of $n^d_p$. We did not vary it to

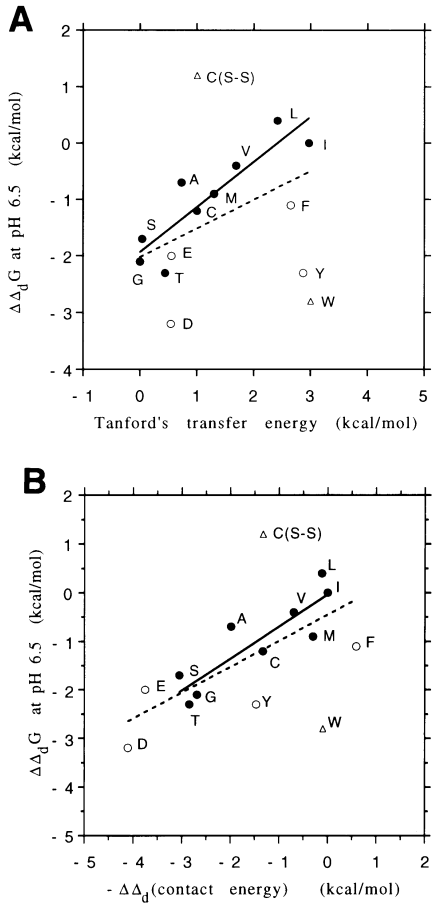


**Fig. 2.** (**A**) Unfolding Gibbs free energy changes of the mutant bacteriophage T4 lysozymes with single amino acid substitutions at position Ile3 at pH 6.5 (Matsumura *et al.*, 1988) are plotted against the transfer energies of residues from water to organic solvent (Tanford, 1967). The solid line indicates a regression line, $y = -1.9 + 0.80x$, for the nine amino acids shown by solid circles (Leu, Ile, Met, Val, Cys, Thr, Ala, Ser and Gly). The correlation coefficient is 0.89; $P = 0.001$. Amino acids omitted in this correlation analysis are the aromatic residues, the acidic residues and cysteine. (**B**) The same unfolding Gibbs free energy changes as shown in (A) are plotted against the contact energy changes due to the replacements calculated by Equation 8 with $n^d_p = 0.85$ (Table III, next to last column). The solid line indicates a regression line, $y = -0.03 + 0.66x$, for the nine amino acids employed in (A) and shown by solid circles. The same nine amino acids are chosen only for comparison with (A). The correlation is as good as in (A); the correlation coefficient is 0.88 with $P = 0.002$. Even in the case of including 13 amino acids shown by solid and open circles, the correlation is still good, with a correlation coefficient of 0.79; $P = 0.001$. The regression line, $y = -0.45 + 0.53x$, for this case is shown by the dotted line. The single-letter codes in the figures indicate the amino acids at position 49 in the mutant proteins. The numerical values of the quantities displayed in these figures are all given in Table III.

optimize $n^d_p$ for these data, but rather we used a fixed approximate value because the correlation coefficient does not depend strongly on $n^d_p$ in these cases in which all amino acid substitutions occur at the same sequence positions, and therefore $n^d_p$ ought to be similar for all replacements. However, the dependence of $n^d_p$ on sequence position should be explicitly taken into account in treating cases that include single amino acid replacements at different positions.

Shortle *et al.* (1990) studied extensively the effects of single replacements with Ala and Gly on protein stability for the different positions 11 Leu, 9 Val, 7 Tyr, 5 Ile, 4 Met and 3 Phe in staphylococcal nuclease. Unfolding free energy changes observed for these replacements are listed in Table IV. Figure 3A shows the observed values of unfolding free energy changes and the contact energy changes ($\Delta G_n$) of the native structure for those replacements. Squares and circles show replacements with Ala and Gly, respectively. Open squares and circles give seven replacements of Tyr with Ala and seven replacements of Tyr with Gly, respectively; other replacements are shown by solid squares and circles. As Shortle *et al.* (1990) showed in their original analyses, there is a significant correlation between those two quantities but points are significantly scattered; the correlation coefficient for all replacements is 0.59 with $P = 10^{-8}$. In Figure 3B the abscissas are taken to be the expected values of unfolding Gibbs free energy changes if the average number of contacts in the denatured state $n^d_p$ is assumed to be 0.85, irrespective of the residue position $p$. Taking account of the energy change of the denatured state makes the predicted values the same magnitude of energy as the observed values of unfolding Gibbs free energy changes. However, points are still scattered. The correlation between the observed and expected values is not improved but remains highly significant; the correlation coefficient for all replacements is 0.61 with $P = 2\times10^{-9}$; see figure legends for details. The points for replacements with Ala tend to be located higher in the figure than those for replacements with Gly. This scatter of points may result from the inappropriate assumption that $n^d_p$ is constant, regardless of the residue position.

Alternatively, for these data it is possible to treat the average number of contact pairs formed with residue $p$ in the denatured state $n^d_p$ as an adjustable parameter. Table IV lists the observed values of unfolding free energy changes for those replacements, the expected free energy changes of the native structure and the values of the position-dependent factor, $n^d_p$, which best fit Equation 8 with the condition that they can have only non-negative values. A significant correlation is found between the estimated values of unfolding free energy changes for replacements with Gly and for replacements with Ala, as shown in Figure 4A; the correlation coefficient is 0.83 with $P = 3\times10^{-9}$, and the regression line is $y = 0.18 + 0.93x$. The expected value, 1, of the slope can be accepted with $P = 0.52$ for the degree of freedom $n = 37$. This supports the present method in Equation 8 considering that the $n^d_p$ varies from position to position. Figure 3C shows the observed and expected values of unfolding free energy changes in which $n^d_p$ is taken to be the simple average of the estimated values from replacements with Gly and Ala. The regression line for all data points is $y = 0.19 + 1.04x$ and the correlation coefficient is 0.88 with $P = 2\times10^{-13}$ for the degrees of freedom $n = 37$. The unit slope of the expected regression line can be accepted with $P = 0.65$ for $n = 37$. However, points for Tyr replacements, which are shown by open squares and circles, deviate significantly from the regression line. Tyr can have both a hydrophilic and/or a hydrophobic character, and it is often difficult to assess simply the role played by Tyr. The contact energies for Tyr are estimated as the average of such amphipathic characteristics. The contact energies for Tyr, which are buried inside the protein and probably play the role of hydrophobic residues may be underestimated by averaging over all Tyr occurrences. The estimates of contact energies for aromatic and polar, charged amino acids appear to be less accurate than those for hydrophobic residues (Figures 1–3). If Tyr replacements are excluded in the correlation analysis of Figure 3C, the correlation coefficient will increase from 0.88 to 0.95 and $P$ will become $10^{-16}$. In this case, the

**Table IV.** The unfolding Gibbs free energy changes and the estimated values of $n^d p$ for mutant staphylococcal nucleases with a single amino acid replacement

| Mutant | $\Delta\Delta_d G^a$ | $\Delta G_n{}^b$ | $-2(f_A - f_i)^c$ | $n^d{}_p{}^d$ | Mutant | $\Delta\Delta_d G^a$ | $\Delta G_n{}^b$ | $-2(f_G - f_i)^c$ | $n^d{}_p{}^d$ | $2n_p{}^e$ |
|---|---|---|---|---|---|---|---|---|---|---|
| L7A | –1.6 | –4.0 | –4.0 | 0.61 | L7G | –1.5 | –4.5 | –4.8 | 0.63 | 5 |
| L14A | –2.3 | –4.8 | –4.0 | 0.62 | L14G | –3.7 | –6.2 | –4.8 | 0.53 | 5 |
| I15A | –2.7 | –3.2 | –4.5 | 0.12 | I15G | –3.3 | –3.6 | –5.3 | 0.06 | 3 |
| I18A | –2.5 | –3.7 | –4.5 | 0.27 | I18G | –2.5 | –4.1 | –5.3 | 0.31 | 4 |
| V23A | –2.9 | –4.2 | –3.0 | 0.44 | V23G | –5.6 | –5.9 | –3.7 | 0.07 | 6 |
| L25A | –2.7 | –7.1 | –4.0 | 1.10 | L25G | –4.5 | –9.5 | –4.8 | 1.05 | 7 |
| M26A | –1.5 | –4.3 | –4.6 | 0.61 | M26G | –2.2 | –4.8 | –5.4 | 0.49 | 4 |
| Y27A | –4.2 | –2.2 | –1.0 | 0.00 | Y27G | –5.8 | –3.8 | –1.8 | 0.00 | 6 |
| M32A | –1.7 | –4.3 | –4.6 | 0.57 | M32G | –2.4 | –5.0 | –5.4 | 0.48 | 4 |
| F34A | –3.7 | –9.9 | –4.8 | 1.30 | F34G | –6.2 | –11.9 | –5.5 | 1.03 | 8 |
| L36A | –3.5 | –6.6 | –4.0 | 0.79 | L36G | –5.3 | –8.3 | –4.8 | 0.62 | 7 |
| L37A | –1.7 | –4.2 | –4.0 | 0.63 | L37G | –3.8 | –4.8 | –4.8 | 0.22 | 5 |
| L38A | –1.7 | –5.6 | –4.0 | 0.98 | L38G | –0.6 | –6.9 | –4.8 | 1.32 | 6 |
| V39A | –2.2 | –6.0 | –3.0 | 1.30 | V39G | –4.7 | –7.8 | –3.7 | 0.83 | 9 |
| V51A | –0.3 | –1.7 | –3.0 | 0.46 | V51G | –0.4 | –1.8 | –3.7 | 0.37 | 3 |
| Y54A | –2.2 | –1.5 | –1.0 | 0.00 | Y54G | –1.9 | –2.1 | –1.8 | 0.13 | 4 |
| F61A | –2.3 | –6.2 | –4.8 | 0.82 | F61G | –4.8 | –7.3 | –5.5 | 0.45 | 5 |
| M65A | –2.0 | –5.3 | –4.6 | 0.72 | M65G | –4.6 | –6.6 | –5.4 | 0.37 | 5 |
| V66A | –2.2 | –5.4 | –3.0 | 1.09 | V66G | –4.4 | –7.5 | –3.7 | 0.84 | 8 |
| I72A | –5.1 | –8.1 | –4.5 | 0.66 | I72G | –6.5 | –10.1 | –5.3 | 0.68 | 7 |
| V74A | –3.1 | –5.7 | –3.0 | 0.88 | V74G | –6.6 | –8.0 | –3.7 | 0.37 | 8 |
| F76A | –4.0 | –7.1 | –4.8 | 0.64 | F76G | –4.7 | –8.1 | –5.5 | 0.61 | 6 |
| Y85A | –0.4 | –1.1 | –1.0 | 0.69 | Y85G | –1.0 | –1.0 | –1.8 | 0.00 | 2 |
| L89A | –2.6 | –2.6 | –4.0 | 0.00 | L89G | –3.2 | –2.8 | –4.8 | 0.00 | 3 |
| Y91A | –5.3 | –3.3 | –1.0 | 0.00 | Y91G | –6.7 | –4.3 | –1.8 | 0.00 | 8 |
| I92A | –4.0 | –10.2 | –4.5 | 1.38 | I92G | –6.6 | –12.7 | –5.3 | 1.17 | 9 |
| Y93A | –6.5 | –2.3 | –1.0 | 0.00 | Y93G | –7.5 | –2.9 | –1.8 | 0.00 | 5 |
| M98A | –4.6 | –6.5 | –4.6 | 0.41 | M98G | –4.5 | –7.4 | –5.4 | 0.53 | 6 |
| V99A | –3.2 | –5.7 | –3.0 | 0.86 | V99G | –5.0 | –7.7 | –3.7 | 0.74 | 8 |
| L103A | –4.6 | –9.2 | –4.0 | 1.15 | L103G | –6.6 | –11.4 | –4.8 | 1.00 | 10 |
| V104A | –2.9 | –5.4 | –3.0 | 0.84 | V104G | –6.5 | –7.2 | –3.7 | 0.19 | 8 |
| L108A | –5.8 | –4.7 | –4.0 | 0.00 | L108G | –7.2 | –5.6 | –4.8 | 0.00 | 5 |
| V111A | –4.7 | –5.9 | –3.0 | 0.41 | V111G | –4.9 | –7.7 | –3.7 | 0.76 | 8 |
| Y113A | 0.0 | –1.5 | –1.0 | 1.47 | Y113G | –0.3 | –1.6 | –1.8 | 0.72 | 3 |
| V114A | 0.0 | –3.5 | –3.0 | 1.18 | V114G | –0.2 | –4.1 | –3.7 | 1.05 | 5 |
| Y115A | –0.3 | –0.4 | –1.0 | 0.10 | Y115G | –0.7 | –0.6 | –1.8 | 0.00 | 1 |
| L125A | –4.9 | –7.7 | –4.0 | 0.71 | L125G | –7.0 | –9.4 | –4.8 | 0.50 | 8 |
| L137A | –2.3 | –3.2 | –4.0 | 0.23 | L137G | –4.6 | –3.7 | –4.8 | 0.00 | 4 |
| I139A | –3.5 | –3.8 | –4.5 | 0.07 | I139G | –4.4 | –4.2 | –5.3 | 0.00 | 4 |

[a]20°C and pH 7.0; taken from Shortle *et al.* (1990).
[b]The first term in Equation 8 in kcal/mol; 0.58 kcal/mol for *RT*; the protein coordinate set 1SNC is used.
[c]In kcal/mol; 0.58 kcal/mol for *RT*.
[d]A value for $n^d{}_p$ which best fits Equation 8.
[e]The number of contacts at each residue position in the native structure.

expected regression line, $y = x$, can be accepted with $P > 0.18$ for $n = 30$. In the correlation analyses below, data for Tyr replacements are excluded simply for this reason.

Now, the average numbers of contacts in the denatured state $n^d{}_p$ have been estimated for various residue positions of staphylococcal nuclease as shown in Table IV. With these values of $n^d{}_p$ we can estimate the contribution of hydrophobic interactions to the unfolding free energy change for any amino acid replacement at these sites. The expected values, if steric effects are ignored, of the unfolding Gibbs free energy changes for five replacements of Ile with Val in staphylococcal nuclease are listed in Table V and compared with their observed values (Shortle *et al.*, 1990) in Figure 5. Except for the replacement at position 92, there is virtually no difference between the observed and expected values of the unfolding Gibbs free energy changes within the resolution of the present approximation. Actually, the expected regression line, shown by a dotted line in Figure 5, can be accepted with $P > 0.25$ for $n = 3$ or $P > 0.35$ for $n = 2$ if I92V is excluded. This result supports the present model of Equation 8.

The average number of contacts at each residue position in the denatured state, which is equal to $2n^d{}_p$, ranges from 0 to ~3, as shown in Figure 4A. These values are much smaller than the numbers of contacts at the same positions in the native structure, as listed in Table IV, but they are significantly larger than those expected from the simplistic, extreme view that protein conformations are fully expanded in the denatured state. To understand the underlying mechanism, correlations are examined between the average values, $<n^d{}_p>$, for replacements with Ala and Gly, and local or non-local environments of residues along a chain. Local, mean hydrophobicities of residues along a chain are calculated with the window sizes of one, seven, 11, 15 and 19 residues; the hydrophobicity of each residue type has been evaluated in terms of the average contact energy of each residue, $e_i$. The best but extremely weak correlation is found with a correlation coefficient of 0.31 between $<n^d{}_p>$ and the mean hydrophobicity in a segment from $p$ –7 to $p$ +7 residues. By contrast, an extremely interesting and significant correlation with a correlation coefficient of 0.69 is found between $2<n^d{}_p>$ and the number of contacts, $2n_p$, at each residue in the native structure, indicating that residues buried in the native state tend to be less exposed

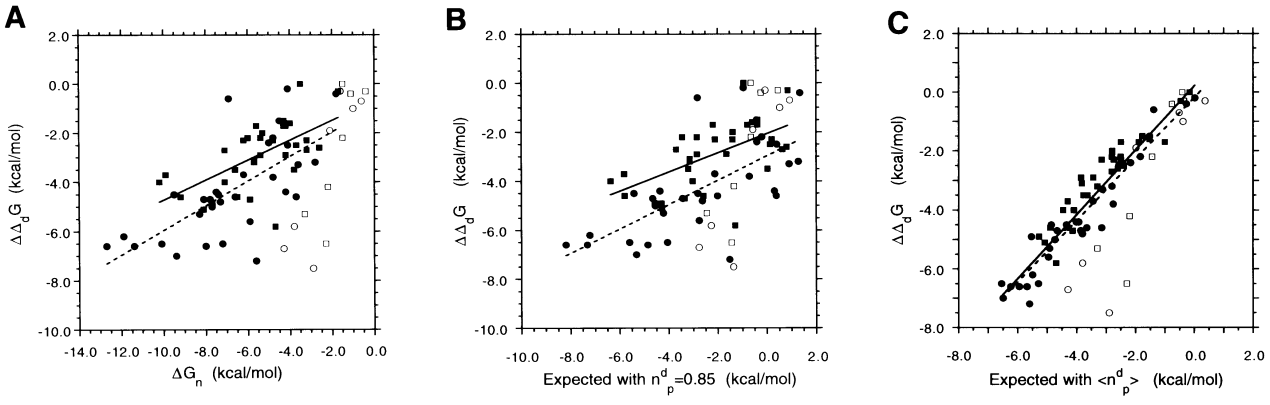


**Fig. 3.** The observed values (Shortle *et al.*, 1990) listed in Table IV are plotted against the expected values for unfolding Gibbs free energy changes of the mutant staphylococcal nucleases with single amino acid substitutions at the different positions 11 Leu, 9 Val, 7 Tyr, 5 Ile, 4 Met and 3 Phe. Squares and circles show replacements with Ala and Gly, respectively. Open squares and circles show seven replacements of Tyr with Ala or Gly, respectively. Solid squares and circles are for other substitutions. (**A**) The abscissa is taken to be the contact energy change ($\Delta G_n$) of the native structure. The regression line for 32 solid squares shown by a solid line is $y = -0.68 + 0.40x$, and the correlation coefficient is 0.61 with $P = 0.0002$. A dotted line shows the regression line $y = -0.97 + 0.50x$ for 32 solid circles; the correlation coefficient is 0.68 with $P = 0.00002$. (**B**) The expected values are calculated from Equation 8 by assuming a constant value 0.85 for $n^d_p$. The regression line for 32 solid squares shown by a solid line is $y = -2.06 + 0.39x$, and the correlation coefficient is 0.57 with $P = 0.0007$. A dotted line shows the regression line $y = -2.97 + 0.50x$ for 32 solid circles; the correlation coefficient is 0.68 with $P = 0.00002$. (**C**) Data are calculated by taking $n^d_p$ to be the mean of those values estimated for replacements with Ala and Gly and listed in Table IV. The regression line for solid squares and circles shown by a solid line is $y = -0.21 + 1.09x$, and the correlation coefficient is 0.95 with $P = 10^{-16}$ for the degrees of freedom $n = 30$. The expected regression line, $y = x$, can be accepted with $P > 0.18$ for $n = 30$. A dotted line shows the regression line $y = -0.19 + 1.04x$ for all points; the correlation coefficient is 0.88 with $P = 2\times10^{-13}$ for $n = 37$. In this case, the expected slope, one, can be accepted with $P = 0.65$ for $n = 37$.

to solvent even in the denatured state (Figure 4B). This correlation is not so strong but is significant with $P = 0.00001$ for the degrees of freedom $n = 30$. If the total contact energy between each residue position and the surrounding residues is used instead of the number of contacts, we find a slightly better correlation with a correlation coefficient of 0.72. Here for simplicity the correlation for the number of contacts is shown in Figure 4B. From this, it is clear that long-range interactions along a chain may be significant even in the denatured state. In other words, this compact denatured state is indicated to have a residual native-like structure.

## Discussion

The contact energies employed here include hydrophobic interactions between amino acids as the main contribution. Rose *et al.* (1985) pointed out that there is a good correlation between the values of the characteristic fractional area loss and our values of the average energy change for an *i* type residue, $e_{ir}$, upon contact formation with a residue of average type; the correlation coefficient is 0.94 [see Miyazawa and Jernigan (1985) for $e_{ir}$]. The $e_{ir}$ also correlate well with the OMH scale of Sweet and Eisenberg (1983); the correlation coefficient is 0.89. Here it should be noted that all three of these scales are based on statistical data of protein structures, and so they reflect not only hydrophobic energies but also all other types of interaction. However, unlike those hydrophobic scales, these contact energies are pairwise potentials in energy units.

These contact energies were demonstrated to discriminate successfully between native-like and incorrectly folded conformations. In a study of five small proteins (Covell and Jernigan, 1990), lattice points were fit to $C^\alpha$ positions and these points were used for the generation of large numbers of diverse conformations, as reflected by the frequent occurrences of almost all possible non-native contact pairs. Contact energies from Miyazawa and Jernigan (1985) were used to calculate average contact energies between all residue pairs. Native contact pairs then proved to be highly favorable. Also, the native conformation was always found among the best 2% of the thousands to tens of thousands of conformations when they were ranked by their total contact energies. Ranking by hydrophobicity alone proved to be substantially less successful, with the rank of the native form determined in this way only at the 12% level. These residue–residue contact energies can clearly play a useful role in screening conformations prior to more detailed atomic conformational calculations.

As for the values of contact energies, comparison of the contact energies with the Nozaki–Tanford transfer energies indicates that on average the contact energies yield about twice the energy gain as do the Nozaki–Tanford transfer energies (Miyazawa and Jernigan, 1985). The changes of unfolding energy in the case of tryptophan synthase α subunit are much larger than expected from those transfer energies for replacing residues; the slope of the regression line between the unfolding energy changes and the transfer energies is 3.7. Yutani *et al.* (1987) thought that factors other than hydrophobicity must be important in determining the unfolding Gibbs free energy changes. Shortle *et al.* (1990) also concluded that the average stability loss caused by Gly substitutions for each of the six types of residue, such as Leu, Val, Tyr, Ile, Met and Phe, is two to three times larger than expected from the transfer free energy of that residue's side chain if hydrophobicity were the sole source of stability. Taking account of the free energy cost due to cavity formation in the native structure, by the replacement of a bulky side chain by a smaller side chain, could not completely resolve this discrepancy because for several positions the stability loss that accompanied removal of the $C^\beta$ atom exceeded the 2.0–2.5 kcal/mol upper limit estimated for the contribution of methyl groups to the cohesion of hydrocarbon solids (Shortle *et al.*, 1990; Dill and Shortle, 1991). Shortle *et al.* (1990) proposed that some of these unexpectedly large variations in unfolding free energy changes caused by single amino acid mutations could be a consequence

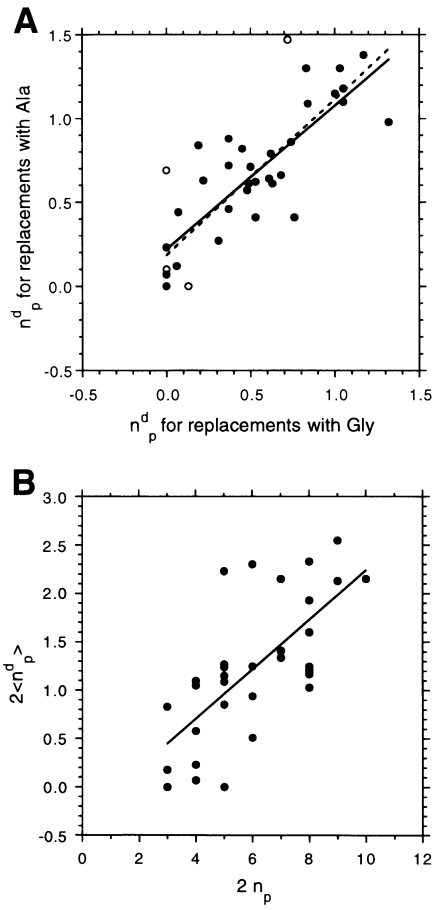


**Fig. 4.** (**A**) Half of the average number of contacts, $n^d_p$, at each residue position, $p$, in the denatured state, estimated by Equation 8 from unfolding Gibbs free energy changes for replacements with Ala and Gly in staphylococcal nuclease. Open circles show seven replacements of Tyr with Ala or Gly. Solid circles are for 32 other substitutions. The regression line for solid circles shown by a solid line is $y = 0.22 + 0.86x$, and the correlation coefficient is 0.83 with $P = 3\times10^{-9}$. The expected unit slope can be accepted with $P = 0.19$ for $n = 30$. A dotted line shows the regression line, $y = 0.18 + 0.93x$, for solid and open circles; the correlation coefficient is 0.83 with $P = 7\times10^{-11}$. In this case, the expected value, one, of the slope can be accepted with $P = 0.52$ for $n = 37$. The values of $n^d_p$ are listed in Table IV. (**B**) The average number of contacts, $2<n^d_p>$, at residue position $p$ in the denatured state, is plotted against the number of contacts, $2n_p$, at the same position in the native structure; $2<n^d_p>$ is the mean of those estimated from replacements with Ala and Gly. The values of $n^d_p$ and $2n_p$ are listed in Table IV. Solid circles show the different positions of 11 Leu, 9 Val, 5 Ile, 4 Met and 3 Phe. A solid line shows the regression line $y = -0.32 + 0.26x$ for solid circles; the correlation coefficient is 0.69 with $P = 0.00001$ for $n = 30$.

of disruption of the hydrophobic clustering in the denatured state. For mutant nuclease with a less structured denatured state, part of the stability loss would then be accounted for by the large value of entropy gain on denaturation (Dill and Shortle, 1991). On the other hand, the disruption of the hydrophobic clustering in the denatured state would also reduce the hydrophobic energy of the denatured state. Therefore, both of these unaccounted factors must be estimated quantitatively to explain the large unfolding free energy changes for these replacements.

By contrast, the contact energy changes of the protein native structures are large enough to account for the observed free energy changes as shown in Tables II–IV. The large variation of unfolding free energy changes among residue positions for

**Table V.** The unfolding Gibbs free energy changes and their expected values, neglecting steric effects for mutant staphylococcal nucleases with a single amino acid replacement

| Mutant | $\Delta\Delta_dG^a$ | $\Delta G_n{}^b$ | $-2(f_A - f_I)^c$ | $<n^d_p>^d$ | Expected $\Delta\Delta_dG^e$ | $2n_p{}^f$ |
|---|---|---|---|---|---|---|
| I15V | –0.8 | –1.2 | –1.5 | 0.09 | –1.1 | 3 |
| I18V | –1.1 | –1.8 | –1.5 | 0.29 | –1.3 | 4 |
| I72V | –1.8 | –3.0 | –1.5 | 0.67 | –2.0 | 7 |
| I92V | –0.5 | –3.6 | –1.5 | 1.28 | –1.7 | 9 |
| I139V | –1.5 | –1.4 | –1.5 | 0.04 | –1.4 | 4 |

[a]20°C and pH 7.0; taken from Shortle *et al.* (1990).
[b]The first term in Equation 8 in kcal/mol; 0.58 kcal/mol for *RT*; the protein coordinate set 1SNC is used.
[c]In kcal/mol; 0.58 kcal/mol for *RT*.
[d]The simple average of the estimated values, listed in the fifth and tenth columns of Table IV, of $n^d_p$ from Gly and Ala replacements.
[e]Calculated from Equation 8.
[f]The number of contacts at each residue position in the native structure.

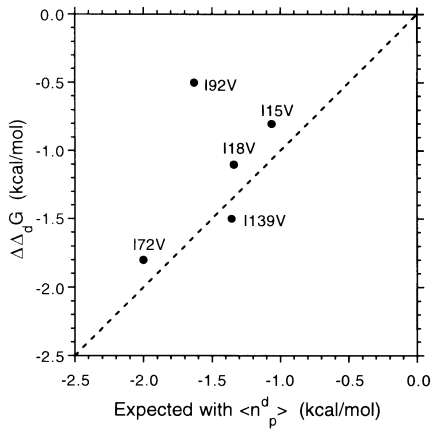


**Fig. 5.** Comparison between the observed and expected values of the unfolding Gibbs free energy changes for five replacements of Ile with Val in staphylococcal nuclease. These expected values are calculated from Equation 8 with $<n^d_p>$ which are the means of the estimated values of $n^d_p$ from Ala and Gly replacements (Table V). The observed values are taken from Shortle *et al.* (1990). The dotted line shows the expected regression line where both axes take identical values. This expected regression line can be accepted with $P > 0.25$ for $n = 3$ or $P > 0.35$ for $n = 2$ if I92V is excluded. Point identifications in the figure indicate the position and type of replacement.

replacements with Ala and Gly, observed by Shortle *et al.* (1990), could be accounted for if free energy changes in the denatured state are explicitly taken into account. Here, we take account of the hydrophobic energy change in the denatured state caused by single amino acid replacements, while conformational entropy changes in the denatured state are not taken into account but are assumed to be insignificant in single amino acid replacements. The large variation may be due to different environments surrounding each residue in the native structure and the denatured state. However, even in the present method, unfolding free energy changes for replacements of Tyr with Ala and Gly are significantly larger than expected (see Figure 3C in which the adjustable parameter $n^d_p$ is taken to be the mean of those values estimated for replacements with Gly and Ala). If Tyr, which has amphipathic characteristics, plays the role of a hydrophobic amino acid at these residue positions, the contact energies for Tyr residues will be underestimated, predicting relatively small changes in the unfolding free energy. As Shortle *et al.* (1990) stated, the

disruption of a hydrophobic cluster in the denatured state could also cause a large entropy gain for the denatured state. Such an entropy change is completely ignored in the present analysis.

The estimated values of the average number of contacts, $2n^{d}_{p}$, at each of those residues in the denatured state of staphylococcal nuclease are consistent with the indication of the existence of two alternative denatured states of the same protein, one representing a compact, structured, but non-native state, and another representing an expanded, less structured state (Shortle and Meeker, 1989). This result supports the models that protein conformations are not fully expanded but are still rather compact in the denatured state (Dill, 1990; Alonso and Dill, 1991). The compact denatured state is now found in pH and thermal denaturations of many proteins, and even for solvent denaturation, residual structures are observed in the denatured state [see Dill and Shortle (1991) for a review]. A compact denatured state was first referred to as the molten globule state in acid denaturation by Ohgushi and Wada (1983).

One of the most interesting features found here is that $n^{d}_{p}$ does not depend much on the local hydrophobicity along the sequence, but is better correlated with the hydrophobic environment surrounding each residue in the native structure. This indicates that the long-range interactions along a chain, which are critical for stabilizing the native structure, are present even in the denatured state, implying that denatured conformations are not so expanded. In other words, these results indicate the presence of a native-like compact denatured state in staphylococcal nuclease, which might resemble the molten globular state [described as the acid-denatured state of α-lactalbumin by Kuwajima (1989), or proposed as a folding intermediate by Ptitsyn (1987), Shakhnovich and Finkelstein (1989) and Finkelstein and Shakhnovich (1989)], that has much native-like topology.

However, the estimated values of $n^{d}_{p}$ range from 0 to 1.47, with an average value of 0.54. This average number of contacts in the denatured state was expected to be larger, but is $<0.85$ which is the value estimated previously for a hypothetical case of no inter-residue interactions; the discrepancy is probably due to both estimates being based on crude approximations. Actual values may not be so accurately specified by this procedure.

Taking account of the free energy change in the denatured state is critical in the analysis of stability changes due to amino acid replacements. This is true, especially in analyzing the effects on stability of the replacement of hydrophilic residues on protein surface by hydrophobic ones, because such hydrophobic residues tend to be less exposed to solvent in the denatured state than they would be on the surface of the native structure. Even in the case of amino acid replacements in a protein interior, the free energy change of the denatured state is not negligible in comparison with that for the native state (Tables II and III). Also, it must be considered explicitly if amino acid replacements at different positions are discussed in terms of stability changes, because the solvent exposure of each residue position in the denatured state also depends on its position.

Here, unfavorable energies from steric hindrance, caused by size and shape differences in an amino acid replacement, in the native structure are not taken into account. Such an effect has been considered explicitly by Lee and Levitt (1991). They calculated stabilizing energies, which include only van der Waals interactions and a simple torsional energy in the 'molten zone' in which a limited number of side-chains are allowed to move, for 78 triple-site sequence variants of λ repressor characterized experimentally by Lim and Sauer (1991). The calculated energies correlate well with the measured activities of the mutants and also directly with the melting temperatures of the mutant proteins. We estimated contact energy changes for these mutant proteins with measured melting temperatures. The estimates of contact energy changes are all similar and fall into a narrow range; whereas the sums of contact energy changes and stability energies calculated by Lee and Levitt (1991) correlate well with the melting temperatures of the mutant proteins, indicating that the stability changes in this case are mainly due to steric factors. However, in this case amino acid substitutions are those among Ile, Leu, Met, Val and Phe, so that large changes in contact energies do not occur. Eriksson *et al.* (1992) reported that there was a good correlation between unfolding free energy changes and the increases in the sizes of cavities in the protein structures of mutant phage T4 lysozymes. They pointed out that the variety of sizes of cavity introduced by mutations can usually account for the wide range of unfolding Gibbs free energy changes caused by an identical amino acid replacement. However, from this point of view an explanation of the unfolding Gibbs free energy changes for the case of the buried Glu49 replacements in tryptophan synthase α subunit would require a substantial decrease in cavity size. The magnitude of cavity size in the case of the Glu49 replacements is unknown. With our approach, these Gibbs free energy changes were explained on the basis of contact energy changes. These calculations and observations indicate that any realistic estimate of stability change may require accounting for both steric factors and hydrophobic interactions.

The present approximation is crude but useful as a simple, empirical method to estimate the contribution of hydrophobic energies to protein stability. For staphylococcal nuclease, the average numbers of contacts in the denatured state $n^{d}_{p}$ have been estimated for various residue positions as shown in Table IV. With these values of $n^{d}_{p}$, we can estimate the contribution of hydrophobic interactions to the unfolding free energy change for any amino acid replacement at these sites. A comparison between the observed and expected values, by neglecting steric effects, of the unfolding Gibbs free energy changes for five replacements of Ile with Val in staphylococcal nuclease (Shortle *et al.*, 1990) shows that there is virtually no difference within the resolution of the present approximation except for the replacement at position 92. These results clearly show that the inter-residue contact energies as applied here properly reflect actual inter-residue interactions, including hydrophobic energies.

## References


Ahmad,F. and Bigelow,C.C. (1986) *Biopolymers*, **25**, 1623–1633.

Alonso,D.O.V. and Dill,K.A. (1991) *Biochemistry*, **30**, 5974–5985.

Chothia,C. (1976) *J. Mol. Biol.*, **105**, 1–14.

Cornette,J.L., Cease,K.B., Margalit,H., Spouge,J.L., Berzofsky,J.A. and DeLisi,C. (1987) *J. Mol. Biol.*, **195**, 659–685.

Covell,D.G. and Jernigan,R.L. (1990) *Biochemistry*, **29**, 3287–3294.

Dang,L.X., Merz,K.M.,Jr and Kollman,P.A. (1989) *J. Am. Chem. Soc.*, **111**, 8505–8508.

Dill,K.A. (1990) *Biochemistry*, **29**, 7133–7155.

Dill,K.A. and Shortle,D. (1991) *Annu. Rev. Biochem.*, **60**, 795–825.

Eisenberg,D. and McLachlan,A.D. (1986) *Nature*, **319**, 199–203.

Eriksson,A.E., Baase,W.A., Zhang,X.-J., Heinz,D.W., Blaber,M., Baldwin,E.P. and Matthews,B.W. (1992) *Science*, **255**, 178–183.

Finkelstein,A.V. and Shakhnovich,E.I. (1989) *Biopolymers*, **28**, 1681–1694.

Finney,J.L., Gellatly,B.J., Golton,I.C. and Goodfellow,J. (1980) *Biophys. J.*, **32**, 17–33.
Fisher,R.A. (1970) *Statistical Methods for Research Workers*. Hafner Publishing Company, Darien, CT.
Herman,R.B. (1972) *J. Phys. Chem.*, **76**, 2754–2759.
Kauzmann,W. (1959) *Adv. Protein Chem.*, **14**, 1–63.
Kellis,J.T., Nyberg,K. and Fersht,A.R. (1989) *Biochemistry*, **28**, 4914–4922.
Kuwajima,K. (1989) *Proteins: Struct. Funct. Genet.*, **6**, 87–103.
Lee,B. (1983) *Proc. Natl Acad. Sci. USA*, **80**, 622–626.
Lee,C. and Levitt,M. (1991) *Nature*, **352**, 448–451.
Lim,W.A. and Sauer,R.T. (1991) *J. Mol. Biol.*, **219**, 359–376.
Matsumura,M., Becktel,W.J. and Matthews,B.W. (1988) *Nature*, **334**, 406–410.
Miyazawa,S. and Jernigan,R.L. (1985) *Macromolecules*, **18**, 534–552.
Nozaki,Y. and Tanford,C. (1971) *J. Biol. Chem.*, **246**, 2211–2217.
Ohgushi,M. and Wada,A. (1983) *FEBS Lett.*, **164**, 21–24.
Oobatake,M. and Ooi,T. (1989) *Bull. Inst. Chem. Res. Kyoto Univ.*, **66**, 433–445.
Ooi,T., Oobatake,M., Nemethy,G. and Scheraga,H.A. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 3086–3090.
Pace,C.N., Laurents,D.V. and Thomson,J.A. (1990) *Biochemistry*, **29**, 2564–2572.
Ptitsyn,O.B. (1987) *J. Protein Chem.*, **6**, 273–293.
Ptitsyn,O.B. (1994) *Adv. Protein Chem.*, in press.
Rose,G.D., Geselowitz,A.R., Lesser,G.J., Lee,R.H. and Zehfus,M.H. (1985) *Science*, **229**, 834–838.
Shakhnovich,E.I. and Finkelstein,A.V. (1989) *Biopolymers*, **28**, 1667–1680.
Sharp,K.A., Nicholls,A., Friedman,R. and Honig,B. (1991) *Biochemistry*, **30**, 9686–9697.
Shortle,D. and Meeker,A.K. (1989) *Biochemistry*, **28**, 936–944.
Shortle,D., Sites,W.E. and Meeker,A.K. (1990) *Biochemistry*, **29**, 8033–8041.
Sneddon,S.F. and Tobias,D.J. (1992) *Biochemistry*, **31**, 2842–2846.
Sweet,R.M. and Eisenberg,D. (1983) *J. Mol. Biol.*, **171**, 479–488.
Tanford,C. (1962) *J. Am. Chem. Soc.*, **84**, 4240–4247.
Tanford,C. (1970) *Adv. Protein Chem.*, **24**, 1–95.
Tidor,B. and Karplus,M. (1991) *Biochemistry*, **30**, 3217–3228.
Yutani,K., Ogasahara,K., Aoki,K., Kakuno,T. and Sugino,Y. (1984) *J. Biol. Chem.*, **259**, 14076–14081.
Yutani,K., Ogasahara,K., Tsujita,T. and Sugino,Y. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 4441–4444.

*Received August 7, 1993; revised June 15, 1994; accepted July 2, 1994*

## Note added in proof

Sharp *et al.* (1991) argued that previous estimates of the hydrophobic effect derived from an analysis of solute partition data did not fully account for changes in volume entropy and provided new estimates by including a volume correction term. Their estimates were almost two times larger than the previous ones and of the same magnitudes as our estimates [$-0.6q_ie_i/2$ in Miyazawa and Jernigan (1986)] from the contact energies of the transfer energy of side chains from the outside to the interior of a protein; a comparison of their data for nonpolar side chains (Phe, Ile, Leu, Val, Ala) with ours yields the regression line $y = 0.09 + 1.1x$ with the correlation coefficient 0.99.