Protein Engineering vol.12 no.7 pp.549–555, 1999

Role of structural and sequence information in the prediction of protein stability changes: comparison between buried and partially buried mutations

M.Michael Gromiha, Motohisa Oobatake[1], Hidetoshi Kono, Hatsuho Uedaira and Akinori Sarai[2]

Tsukuba Life Science Center, The Institute of Physical and Chemical Research (RIKEN), 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

[1]Present address: Faculty of Science and Technology, Meijo University, 1-501 Shiogamaguchi, Tenpaku-ku, Nagoya 468-8502, Japan

[2]To whom correspondence should be addressed

Predicting mutation-induced changes in protein stability is one of the greatest challenges in molecular biology. In this work, we analyzed the correlation between stability changes caused by buried and partially buried mutations and changes in 48 physicochemical, energetic and conformational properties. We found that properties reflecting hydrophobicity strongly correlated with stability of buried mutations, and there was a direct relation between the property values and the number of carbon atoms. Classification of mutations based on their location within helix, strand, turn or coil segments improved the correlation of mutations with stability. Buried mutations within β-strand segments correlated better than did those in α-helical segments, suggesting stronger hydrophobicity of the β-strands. The stability changes caused by partially buried mutations in ordered structures (helix, strand and turn) correlated most strongly and were mainly governed by hydrophobicity. Due to the disordered nature of coils, the mechanism underlying their stability differed from that of the other secondary structures: the stability changes due to mutations within the coil were mainly influenced by the effects of entropy. Further classification of mutations within coils, based on their hydrogen-bond forming capability, led to much stronger correlations. Hydrophobicity was the major factor in determining the stability of buried mutations, whereas hydrogen bonds, other polar interactions and hydrophobic interactions were all important determinants of the stability of partially buried mutations. Information about local sequence and structural effects were more important for the prediction of stability changes caused by partially buried mutations than for buried mutations; they strengthened correlations by an average of 27% among all data sets.

Keywords: amino acid properties/buried and partially buried mutants/local sequence and structural effects/multiple regression technique/protein stability/unfolding free energy change

Introduction

The major non-covalent interactions—hydrophobic, van der Waals and electrostatic interactions and hydrogen bonds—play an important role in the stabilization of protein structures (Dill, 1990; Ponnuswamy, 1993; Rose and Wolfenden, 1993; Ponnuswamy and Gromiha, 1994; Pace *et al.*, 1996). The importance of such interactions for protein stability has been revealed by site-directed mutagenesis (Alber *et al.*, 1987; Yutani *et al.*, 1987; Wetzel *et al.*, 1988; Shortle *et al.*, 1990; Chen *et al.*, 1993; Matthews, 1995; Takano *et al.*, 1995, 1997; Tissot *et al.*, 1996; Akasako *et al.*, 1997). The effect of buried and partially buried mutations on stability has been extensively studied for proteins such as barnase, T4 lysozyme, ribonuclease HI, tryptophan synthase, bovine pancreatic trypsin inhibitor, arc repressor, chymotrypsin inhibitor and staphylococcal nuclease (Table I).

Several theoretical methods have been proposed to predict the stability changes induced by mutation. These methods are based on detailed atomic models coupled with semi-empirical force fields (Bash *et al.*, 1987; Dang *et al.*, 1989; Tidor and Karplus, 1991; Simonson and Brunger, 1992), simplified energy criteria (Lee and Levitt, 1991; van Gunsteren and Mark, 1992), an empirical method that takes account of the free energy change between the denatured state and the compact native state (Miyazawa and Jernigan, 1994), database-derived potentials (Gills and Rooman, 1996, 1997) and the structural environment-dependent amino-acid substitution tables (Topham *et al.*, 1997).

It has been shown that changes in stability caused by mutations in the buried and partially buried regions of proteins are affected by different types of interactions. The contribution made by hydrophobicity to changes in stability of proteins caused by buried mutations has been emphasized by several investigators (Yutani *et al.*, 1987; Matsumura *et al.*, 1988; Shortle *et al.*, 1990; Takano *et al.*, 1995, 1997; Akasako *et al.*, 1997; Xu *et al.*, 1998). Moreover, the contributions of buried hydrogen bonds (H bonds) and salt bridges to protein stability have also been elucidated (Chen *et al.*, 1993; Tissot *et al.*, 1996). In the case of partially buried and surface mutations, the contribution made by H bonds to the stability of T4 lysozyme was extensively studied by Alber *et al.* (1987), who replaced Thr157 with 13 other amino acids. Electrostatic interactions (surface salt bridges and helix–dipole interactions) also play important roles in protein stability changes caused by partially buried and surface mutations (Akke and Forsen, 1990; Serrano *et al.*, 1990; Nicholson *et al.*, 1991; Sun *et al.*, 1991).

On the other hand, Gills and Rooman (1997) derived torsion and distance potentials and then applied them to predict the stability of seven proteins with mutations in the buried and partially buried regions of the molecules. They found that distance potentials, which are dominated by hydrophobic interactions, represent the strongest interactions stabilizing the protein core; whereas torsion potentials, which are defined by local interactions along the chain, represent the strongest interactions at the protein surface. Combining the distance potentials with the torsion potentials, weighted by 0.4 and 0.7, respectively, for buried and partially buried mutations, yielded correlations between computed and measured changes in folding free energy with correlation coefficients (r) of 0.80 and 0.71, respectively.

In spite of these studies, the relationship between the specific

Table I. List of buried and partially buried mutations

Protein	PDB code	Number of mutations	
		Buried	Partially buried
Adenylate kinase	1ANK	1	1
Arc repressor	1ARQ	7	24
Barnase	1BNI	16	37
β-Lactamase	1BTL	2	0
Bovine pancreatic trypsin inhibitor	1BPI	1	16
Calbindin D9K	4ICB	0	1
Chemotactic protein	1CEY	0	3
Chicken lysozyme	4LYZ	16	2
Chymotrypsin inhibitor	2CI2	8	24
Cytochrome C2	1C2R	0	2
Dihydrofolate reductase	1DYJ	1	12
Growth hormone	1HGU	0	10
Histidine containing protein HPR	1POH	0	3
Interleukin 1β	1IOB	0	3
Iso-1-cytochrome c	1YCC	2	0
Iso-2-cytochrome c	1YEA	1	0
λ Cro protein	1CRO	0	8
λ repressor	1LRP	0	8
P22 Tailspike protein	1TYU	0	3
Phosphoglycerate kinase	3PGK	1	0
Ribonuclease HI	2RN2	11	17
Ribonuclease T1	1RN1	1	4
ROP protein	1ROP	0	2
Staphylococcal nuclease	1STN	55	90
Subtilisin BPN′	1SUP	1	3
Subtilisin inhibitor	2SSI	5	6
T4 Lysozyme	2LZM	29	33
Tryptophan synthase	1WSY	45	8
Ubiquitin	1UBQ	4	0

properties of amino acid residues and the stability (in terms of melting temperature and free energy) of protein mutants has yet to be fully explored. In the present work, we analyzed the correlation between stability changes of proteins caused by buried and partially buried mutations and the changes in 48 physicochemical, energetic and conformational properties. Multiple regression analysis was performed with selected groups of three properties obtained from among all possible combinations of the total 48. Finally, inclusion of information about local sequence and structural effects increased the *r* value of partially buried mutations to 0.97, which reflects a higher level of accuracy than has been obtained using other published methods. We observed that hydrophobicity is the major factor with respect to buried mutations, but hydrogen bonds, other polar interactions and hydrophobic interactions are all important determinants of the stability changes caused by partially buried mutations.

Materials and methods

Database

A database was set up for protein mutants from 103 different proteins. It includes the thermodynamic data for more than 4000 mutants (Gromiha *et al.*, 1999). In the present study, we selected data for different solvent accessibility between: (i) 0 and 2% for completely buried mutations; (ii) 2 and 20% for intermediate between buried and partially buried mutations; (iii) 20 and 50% for partially buried mutations and (iv) more than 50% for exposed mutations. We considered the data for all pH values in the native state.

Furthermore, we have used a subset of data obtained for pH values between 5 and 9 in our analysis and the number of buried and partially buried mutations for each protein is given in Table I. The complete information is available at http://www.rtc.riken.go.jp/~gromiha/table.html and the experimental data (ΔT_m, $\Delta\Delta G$ and $\Delta\Delta G^{H_2O}$) can be obtained at the URL, http://www.rtc.riken.go.jp/protherm.html (Gromiha *et al.*, 1999). We then subdivided the data set into hydrophobic mutations and mutations in selected secondary structural regions (helix, strand, coil or turn). Mutations were defined as hydrophobic when both the mutant and wild-type residues were hydrophobic; Ala, Cys, Phe, Gly, Ile, Leu, Met, Val, Trp and Tyr were considered to be hydrophobic (nonpolar), and all other residues were considered to be polar.

Computation of accessibility

Solvent accessibility (%) was defined as the ratio of the solvent accessible surface area (ASA) of a residue in the native state and that of the residue in an extended tripeptide (Ala-X-Ala) conformation. The solvent accessible surface areas of all atoms were computed using the program ASC (Eisenhaber and Argos, 1993; Eisenhaber *et al.*, 1995) with the van der Waals radii of the atoms given by Ooi *et al.* (1987). The extended state coordinates were computed using the ECEPP/2 algorithm (Momany *et al.*, 1975) with the dihedral angles of Oobatake and Ooi (1993).

Amino acid properties

It has been shown that there are specific cooperativities among amino acid residues, which due to similarities in their various physical, chemical, energetic and conformational properties, enable them to preserve their specific, preferred environments and spatial positions in the folded conformation of proteins (Prabakaran and Ponnuswamy, 1979). We considered a set of 48 diverse amino acid properties, which fall into various clusters analyzed by Tomii and Kanehisa (1996). Detailed descriptions of the properties have already been published (Gromiha and Ponnuswamy, 1993; Oobatake and Ooi, 1993) and the numerical values are available at http://www.rtc.riken.go.jp/~gromiha/table.html. The selected 48 amino acid properties were normalized between 0 and 1.

Computation procedure

We computed the mutation-induced changes in property values, $\Delta P(i)$, using the equation,

$$\Delta P(i) = P_{mut}(i) - P_{WT}(i) \quad (1)$$

where $P_{mut}(i)$ and $P_{WT}(i)$ are, respectively, the normalized property value of the *i*th mutant and wild-type residues; *i* varies from 1 to *N*; *N* being the total number of mutants. The computed differences in property values (ΔP) were related to the changes in experimental stability values (ΔT_m, $\Delta\Delta G$ or $\Delta\Delta G^{H_2O}$) using the correlation coefficient.

Local sequence and structural effects

The effect of local sequence, $P_{seq}(i)$, was included using the equation

$$P_{seq}^{(i)} = \left[\sum_{j=i-k}^{j=i+k} P_j(i)\right] - P_{mut}^{(i)}, \quad (2)$$

where $P_{mut}(i)$ is the property value of the *i*th mutant residue, and $\Sigma P_j(i)$ is the total property value of a segment of $(2k + 1)$ residues, ranging from $i - k$ to $i + k$ about the *i*th wild-type residue. We used windows of 3 and 9 ($k = 1, 4$) residues to

include the influence of short and medium-range interactions (Ponnuswamy *et al.*, 1973, 1980; Gromiha and Selvaraj, 1997).

The structural information was included as follows. Each residue in the protein was represented by its C_α atom. Using the C_α coordinates, a volume with a radius of 8 Å was fixed around the mutant residue, and the residues occurring within this volume were identified. From the analysis of a set of three-dimensional structures of globular proteins, Manavalan and Ponnuswamy (1977, 1978) showed that the influence of each residue over the surrounding medium extends effectively up to 8 Å; hence, we used an 8 Å radius rather than 6 or 10 Å, as was previously used for barnase (Serrano *et al.*, 1992) and staphylococcal nuclease (Shortle *et al.*, 1990), respectively. The residues surrounding the mutant residue were assigned their property values, and the total property value, $P_{sur}(i)$, was computed using the equation

$$P_{sur}^{(i)} = \sum_j n_{ij} \cdot P_j \qquad (3)$$

where n_{ij} is the total number of type j residues surrounding the ith residue of the protein, and P_j is the property value of the type j residue. The structural information, $P_{str}(i)$, was included using the equation

$$P_{str}(i) = P_{sur}(i) - P_{mut}(i), \qquad (4)$$

where $P_{mut}(i)$ is the property value of the ith mutant residue. The computed $P_{seq}(i)$ and $P_{str}(i)$ were related to the changes in experimental stability values (ΔT_m, $\Delta\Delta G$ or $\Delta\Delta G^{H_2O}$) using correlation coefficients.

Results

The investigation was carried out on four groups of data for different ASA, (i) 0–2, (ii) 2–20, (iii) 20–50 and (iv) 50–100%, and three measures of stability: (i) changes in thermal stability, ΔT_m; (ii) unfolding Gibbs free energy changes ($\Delta\Delta G$) based on thermal denaturation; and (iii) unfolding Gibbs free energy changes ($\Delta\Delta G^{H_2O}$) based on denaturant denaturation. Analyses of the correlations between amino acid properties and ΔT_m, $\Delta\Delta G$ and $\Delta\Delta G^{H_2O}$ were carried out for hydrophobic mutations and subsets based on the secondary structures, helix, strand, coil and turn. Considering group (ii), we found that for ΔT_m, the highest correlation coefficients obtained were similar to group (i) for hydrophobic and strand mutations; the levels were similar to group (iii) for the whole dataset and mutations in helical segments. Hence, the results were intermediate between groups (i) and (iii). Furthermore, no significant correlation was obtained for all subsets of group (iv). The results obtained for groups (i) (buried; 0–2%) and (iii) (partially buried; 20–50%) are distinct and highly significant for the understanding of the mechanism of stability and, hence, we present the results for these two groups of mutations in detail. We have performed the analysis using the dataset of all pH and a subset of data for pH values between 5 and 9. We found that the subset of data improved the correlation up to 0.1 and the major results and conclusions are similar. Hence, we discuss the results pertinent to pH values from 5 to 9.

Correlation of stability with properties of amino acids

We observed a significant correlation for several properties of amino acids and stability. The results are similar for each of the three measures of stability—ΔT_m, $\Delta\Delta G$ and $\Delta\Delta G^{H_2O}$. Hence, we only discuss the results obtained for ΔT_m.

The correlation coefficients (r), relating changes in each of the selected 13 properties (Eqn 1) to experimental ΔT_m values, are given in Table II. In the case of buried mutations, the chromatographic index (R_f, the characteristic migration rate in a solvent-absorbent system) has the highest value ($r = 0.59$). Surrounding hydrophobicity (H_p), bulkiness (B_l, the ratio of side chain volume to length), normalized consensus hydrophobicity (H_{nc}), long range non-bonded energy (E_l), mean r.m.s. fluctuation displacement (F), solvent accessible reduction ratio (R_a), average number of surrounding residues (N_s), unfolding solvent accessible surface area change (ΔASA), unfolding entropy change of hydration ($-T\Delta S_h$), and unfolding hydration heat capacity change (ΔC_{ph}) are the other 10 properties that correlate significantly ($|r| > 0.5$) with ΔT_m (Table II).

In the case of partially buried mutations, the correlation was poor when the analysis was carried out with the complete data set. Hence, we divided the data set into hydrophobic mutations and several subgroups based on different secondary structures to perform the analysis. The computed single-property correlations for buried and partially buried mutations for selected properties are included in Table II.

When we generated 48 sets of random numbers, and calculated the correlation with the experimental stability changes (ΔT_m, $\Delta\Delta G$ and $\Delta\Delta G^{H_2O}$), the average r value fell within a range between 0.07 ± 0.05 and 0.39 ± 0.18 for all data sets, which emphasizes the validity of selecting various amino acid properties.

Mutation in helical segments

Considering the mutations in helical segments of partially buried mutations, the strongest positive correlation was observed for ΔASA, whereas the strongest negative correlation was for $-T\Delta S_c$ (Table II). Three other properties, $-T\Delta S_h$, ΔC_{ph} and V^0, also showed significant correlations ($r > 0.5$). We noted that all of these properties were related to thermodynamic parameters, and that the conformational parameter, P_α, was not correlated with stability, as was the case for buried mutations.

Mutations in strand segments

When we analyzed the effects of partially buried mutations within strand segments, β-strand tendency (P_β) showed the strongest single-property correlation with ΔT_m ($r = 0.67$). Interestingly, properties responsible for accessibility and long range interactions (R_a, H_p, B_l and N_s) also correlated significantly ($r > 0.5$) with experimental thermal stability, which demonstrates the importance of long range interactions for the stability of globular proteins (Gromiha and Selvaraj, 1997, 1998, 1999a,b; Selvaraj and Gromiha, 1998). We found that F, P_c, P_t and α_c had strong negative correlations with ΔT_m. Thus, strand (P_β) and coil (P_c) tendency (Fasman, 1989) have opposing influences on stability. In buried mutations, long-range non-bonded energy, E_l, correlated most strongly with ΔT_m.

Mutations in coil regions

No significant correlation was observed between any of the amino acid properties and ΔT_m when partially buried mutations within coil regions were treated as a single group. Mutations within regions of a coil were, therefore, further subdivided based on the H-bond forming capability of the amino acids involved: set 1 included mutations from a residue with H-bond forming capability (hb) to another hb residue (hb→hb); set 2 included mutations from a residue without H-bond forming capability (nhb) to another nhb residue (nhb→nhb);

Table II. Single property correlation with stability of buried and partially buried protein mutants

No	Property	Correlation coefficient, *r*											
		ΔT_m				$\Delta\Delta G$				$\Delta\Delta G^{H_2O}$			
		All	Hyd	Helix	Strand	All	Hyd	Helix	Strand	All	Hyd	Helix	Strand
Buried mutants													
1	H_p	0.50	0.55	0.55	0.54	0.52	0.51	0.61	0.46	0.68	0.55	0.32	0.74
2	M_w	0.32	0.51	0.31	0.44	0.36	0.56	0.40	**0.81**	0.30	0.56	0.48	0.39
3	B_1	0.52	0.65	0.55	0.61	0.55	0.64	0.58	0.78	0.56	0.64	0.58	0.68
4	R_f	**0.59**	0.61	0.59	0.57	**0.66**	0.64	0.68	0.57	**0.72**	**0.65**	**0.60**	0.78
5	H_{nc}	0.50	0.51	0.52	0.49	0.54	0.48	0.56	0.37	0.66	0.48	0.16	0.72
6	E_1	0.57	0.60	0.59	**0.68**	0.54	0.51	0.64	0.53	0.67	0.46	0.16	0.75
7	P_α	0.03	–0.12	0.01	–0.23	0.25	0.12	0.26	–0.01	–0.17	0.32	0.37	–0.33
8	F	–0.53	–0.58	–0.61	–0.55	–0.61	–0.62	–0.73	–0.58	–0.67	–0.63	–0.61	–0.74
9	R_a	0.56	0.57	0.54	0.56	0.58	0.52	0.59	0.54	0.54	0.50	0.23	0.61
10	N_s	0.50	0.54	0.55	0.63	0.44	0.42	0.54	0.48	0.67	0.51	0.23	0.73
11	ΔASA	0.50	0.61	0.57	0.53	0.58	0.66	0.66	0.67	0.65	0.61	0.59	0.74
12	$-T\Delta S_h$	0.52	0.66	0.54	0.57	0.60	**0.72**	0.61	0.72	0.64	0.61	0.57	0.74
13	ΔC_{ph}	0.57	**0.68**	**0.63**	0.53	0.65	0.70	**0.69**	0.65	0.71	0.62	0.55	**0.80**
Partially buried mutants		Hyd	Helix	Strand	Coil	Hyd	Helix	Strand	Coil	Hyd	Helix	Strand	Coil	Turn
1	Hp	0.50	0.43	0.60	–0.19	0.64	0.45	0.52	–0.29	0.40	0.48	0.40	–0.09	0.60
2	pH_i	–0.18	0.23	0.13	0.23	0.18	0.38	**0.56**	0.12	–0.34	–0.07	–0.15	0.07	–0.29
3	B_l	0.41	0.45	0.59	–0.11	0.54	0.48	0.45	–0.41	0.44	0.49	0.53	–0.11	0.28
4	P_α	0.09	–0.09	0.17	–0.14	0.21	–0.01	0.18	–0.30	0.28	0.28	0.56	–0.06	0.22
5	P_β	0.41	0.42	**0.67**	–0.29	0.50	0.48	0.43	–0.39	0.31	0.28	0.23	–0.18	0.60
6	P_t	–0.33	–0.21	–0.53	0.19	–0.51	–0.32	–0.54	0.36	–0.39	–0.44	–0.53	0.15	–0.33
7	P_c	–0.30	–0.16	–0.54	0.26	–0.48	–0.26	–0.45	0.48	–0.38	–0.43	–0.62	0.12	–0.40
8	F	–0.45	–0.27	–0.71	0.03	–0.52	–0.30	–0.55	0.13	–0.47	–0.58	–0.53	0.16	–0.40
9	R_a	0.44	0.32	0.65	0.06	0.55	0.32	0.40	–0.01	0.33	0.44	0.50	–0.15	**0.63**
10	N_s	0.50	0.46	0.50	–0.06	0.66	0.47	0.44	–0.04	0.27	0.31	0.23	–0.10	0.57
11	α_c	–0.26	–0.40	–0.50	–0.03	–0.16	–0.49	–0.38	–0.07	–0.04	0.03	0.08	0.02	–0.29
12	V^0	0.29	0.52	0.31	–0.09	0.12	0.57	0.41	–0.43	**0.54**	0.44	0.51	–0.05	0.25
13	H_{gm}	**0.54**	0.33	0.45	–0.12	**0.66**	0.34	0.40	–0.14	0.26	0.32	0.18	–0.01	0.46
14	ΔASA	0.35	**0.60**	0.48	–0.15	0.23	0.58	0.43	–0.52	0.54	**0.61**	**0.60**	–0.09	0.46
15	$-T\Delta S_h$	0.27	0.60	0.41	–0.22	0.12	**0.63**	0.47	–0.54	0.51	0.47	0.49	–0.08	0.38
16	ΔC_{ph}	0.37	0.52	0.48	–0.20	0.35	0.50	0.43	–0.41	0.52	0.51	0.41	–0.15	0.40
17	$-T\Delta S_c$	–0.23	–0.52	–0.26	**0.41**	0.07	–0.30	0.04	**0.63**	–0.55	–0.61	–0.54	**0.26**	–0.40

The largest correlation coefficients are shown in boldface; hyd, hydrophobic mutation; mutation in turn segments has significant number of data only for $\Delta\Delta G^{H_2O}$ of partially buried mutations.

H_p, Surrounding hydrophobicity (Ponnuswamy, 1993); M_w, molecular weight (Sober, 1970); pH_i, isoelectric point; B_l, bulkiness; R_f, chromatographic index (Zimmerman *et al*., 1968); H_{nc}, normalized consensus hydrophobicity (Eisenberg *et al*., 1982); E_l, long range non-bonded energy (Oobatake and Ooi, 1977); P_α, P_β, P_t and P_c are, respectively, α-helical, β-structure, turn and coil tendencies (Chou and Fasman, 1978); F, mean r.m.s. fluctuational displacement (Bhaskaran and Ponnuswamy, 1984); R_a, solvent accessible reduction ratio (Rose *et al*., 1985); N_s, average number of surrounding residues (Manavalan and Ponnuswamy, 1978); α_c, power to be at the C-terminal of α-helix (Chou and Fasman, 1978); V^0, partial specific volume (Iqbal and Verrall, 1988); H_{gm}, combined surrounding hydrophobicity (globular and membrane; Ponnuswamy and Gromiha, 1993); ΔASA, solvent accessible surface area for unfolding; $-T\Delta S_h$, unfolding entropy change of hydration; ΔC_{ph}, unfolding hydration heat capacity change; $-T\Delta S_c$, unfolding entropy changes of chain (tables 7 and 8 in Oobatake and Ooi, 1993).

set 3 contained mutations from an hb residue to a nhb residue (hb→nhb); and set 4 contained mutations from a nhb residue to an hb residue (nhb→hb). We considered Asp, Cys, Glu, His, Lys, Met, Asn, Gln, Arg, Ser, Thr, Trp and Tyr to belong to the hb group, as suggested by Vogt *et al.* (1997).

Within set 1, the strongest positive and negative correlations were with $-T\Delta S_c$ ($r = 0.72$) and C_a ($r = -0.91$), respectively. ΔH_h, E_{sm} and ΔG_h also showed significant positive correlations with ΔT_m, while ΔH_c showed a strong negative correlation ($r = -0.83$). Thus stability is favored by the chain entropic contribution ($-T\Delta S_c$) and disfavored by chain enthalpic terms (ΔH_c). Multiple regression analysis strengthened the correlation ($r = 0.95$), and inclusion of local sequence information yielded a still stronger correlation ($r = 0.97$).

The single-property correlations for set 2 were weak, but significant single-property correlations were obtained for sets 3 and 4. Inclusion of local sequence information increased r values to 0.97, 0.78 and 0.95, respectively, for sets 2, 3 and 4. A similar strengthening of the correlation with stability was obtained when structural information was included.

Discussions

Role of hydrophobicity in stability of buried mutation

A perusal of the top four properties (R_f, E_l, R_a and ΔC_{ph}) that showed the strongest correlation with ΔT_m and $\Delta\Delta G$ of buried mutation shows that for all of the properties, the nonpolar residues (A, V, M, Y, I, W, L and F) had higher values; charged residues had lower values; and polar residues had intermediate values. Also, we observed a direct relation between property values and the number of carbon atoms. For instance, considering the negatively charged residues, Asp and Glu, the Glu residue, which has more carbon atoms than Asp, acquires a higher value. Similar trends were observed between Lys and

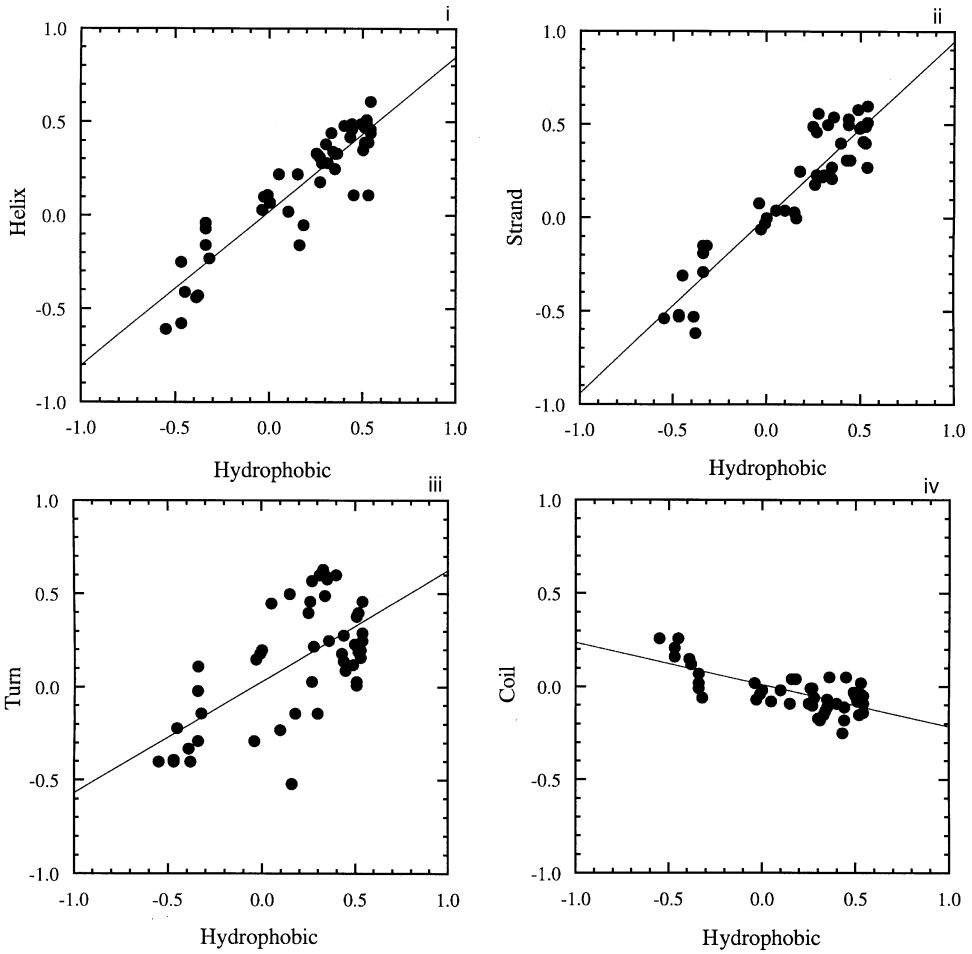

Fig. 1. The relationship among subgroups for all 48 properties with respect to $\Delta\Delta G^{H_2O}$. The relation between *r* values obtained for partially buried hydrophobic mutations and for partially buried mutations within secondary structures: (i) helix; (ii) strand; (iii) turn; and (iv) coil.

Arg [P(Arg) > P(Lys)] and between Asn and Gln [P(Gln) > P(Asn)], among others. An in-depth analysis of nonpolar residues shows a similar pattern: P(Ala) > P(Gly); P(Leu) > P(Val); P(Ile) > P(Val); and P(Trp) > P(Tyr). Thus, hydrophobic residues play a major role in explaining the changes in stability caused by buried mutations, which is consistent with the well-known fact that hydrophobicity, defined as a measure of transfer free energy change, varies linearly with the number of methylene carbons (Tanford, 1980).

With respect to $\Delta\Delta G^{H_2O}$, we observed a better single-property correlation for buried mutations in β-strand segments than for those in α-helical segments (Table II), which may be due to the stronger hydrophobicity of β-strands (Ponnuswamy *et al.*, 1980; Gromiha and Ponnuswamy, 1995). These authors showed that β-strand segments have higher average H_p values than do α-helical segments, which explains the stronger hydrophobicity of β-strands and the stronger correlation between mutations within β-strand regions and stability.

Subgrouping secondary structures strengthens correlation for partially buried mutations

The analysis carried out based on the secondary structure (helix, strand, coil and turn) yielded good correlations between amino acid properties and protein stability changes caused by partially buried mutations (Table II); moreover, multiple regression analysis using three properties substantially strengthened these correlations. When the analysis was carried out using the complete data set, the correlation was poor (data not shown), which may have been due to combining the effects of irregular structures (coil) with those of ordered structures (helix, strand and turn): the two structure types have opposing roles in the folding and stability of protein molecules, and hence their contributions to stability might offset one another. The present study, therefore, suggests that mutations in ordered and disordered structures should be considered separately, and that classification based on secondary structures is necessary for better understanding of the relationship between the structure and stability of proteins.

Coil has different tendency

The *r* values obtained for the 17 selected properties in all of the respective subgroups are given in Table II. It is evident from this table that the mechanism responsible for changes in stability caused by partially buried mutations within coil segments is the opposite of that in helical and strand segments. Within coil segments, $-T\Delta S_c$ was the property most strongly correlated with stability (Table II), which was indicative of the importance of entropic effects in the stability of mutations within coil regions. Interestingly, we note that the signs for the properties governing stability of coil regions are the opposite of those observed in all the other segments.

The relationship among the subgroups for all 48 properties with respect to $\Delta\Delta G^{H_2O}$ is shown in Figure 1. In this figure, we show the relation between the *r* values obtained for hydrophobic mutations and those for mutations in secondary

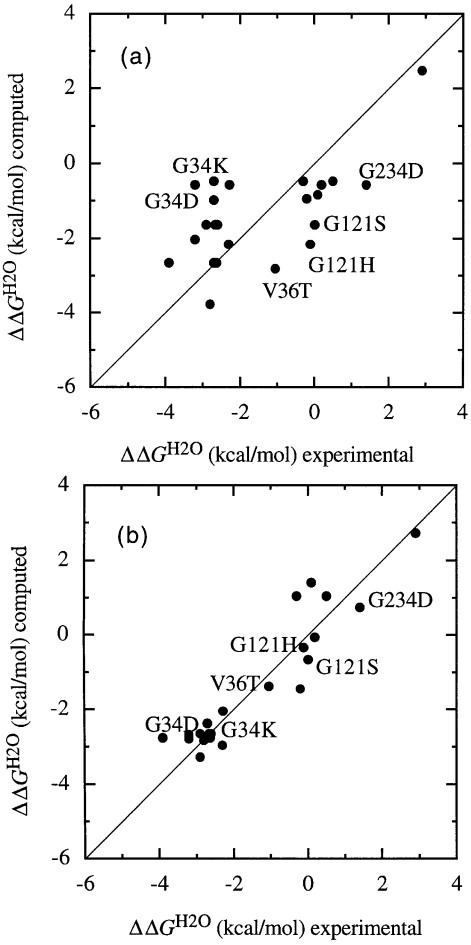

Fig. 2. Relationship between experimental and computed stability ($\Delta\Delta G^{H_2O}$) for a subset of partially buried coil mutations (set 4; nhb→hb): (**a**) without sequence and structural information; and (**b**) with structural information.

structures. We found a direct relationship between hydrophobicity and ordered structures, indicating that hydrophobicity plays a role in stabilizing the structures of mutations in ordered structures. On the other hand, the inverse was found for coil regions, which are destabilized by hydrophobicity. We also observed that a coil behaves in a manner opposite to that of a helix, strand or turn; *r* values are –0.82, –0.68 and –0.71, respectively.

Importance of hydrogen bonds for protein stability changes

We found no significant correlation with stability when all mutations in coil regions were considered simultaneously. Therefore, mutations in coil regions were subdivided based on their capacity to form H-bonds: set 1, hb→hb; set 2, nhb→nhb; set 3, hb→nhb; and set 4, nhb→hb. Without considering sequence and structural information, we obtained significant single-property and multiple correlations for sets 1, 3 and 4; *r* values ranged from 0.54 to 0.95. In contrast, set 2 (nhb→nhb) showed a poor correlation. Most of the residues involved in the mutations with set 2 were hydrophobic, and hydrophobicity apparently did not contribute to the stability of coil. For the other three sets, polar and charged residues dominated the mutations, which yielded stronger correlations. These results illustrate the contribution made by H bonds and other polar interactions to the protein stability changes caused by partially buried mutations. Inclusion of hydrogen bonds in helical and strand segments improved the correlation to some extent but the data set is insufficient to carry out an extensive analysis.

Importance of sequence and structural information

Information about the local sequence and/or the surrounding residues strengthened the correlation in several subsets of partially buried mutations, by an average increase of 27%. This may be due to either (i) additional information about short and medium-range interactions with neighboring residues or (ii) information about polar–polar interactions and/or ion pairs and salt bridges surrounding the mutant residues. In the case of partially buried mutations, the residues are surrounded not only by hydrophobic residues, but also by polar residues, and hence the real situation is better expressed by the inclusion of structural information. In the case of buried mutations, the mutant residues are surrounded primarily by nonpolar residues, and hence the effect is comparatively less. We illustrate the effect of sequence and structure with the following example.

In Figure 2, the relationship between the experimental and computed $\Delta\Delta G^{H_2O}$ values for mutations in coil regions (set 4, nhb→hb) are depicted; Figure 2a shows the correlation without sequence and structural information, while Figure 2b shows the correlation when structural information is included. The *r* values obtained for the two sets of data are 0.71 and 0.95, respectively. Note that in Figure 2a some of the mutants are substantially displaced from the diagonal line (G234D of 1WSY; G34D, G34K and V36T of 1BNI; and G121H and G121S of 1DYJ). Inspection of residues surrounding the mutant residue showed that along the sequence, more than 50% of both the nearby and more distant residues were polar. Furthermore, we noted the presence of at least one aromatic residue nearby most of the mutants. The inclusion of the information about the polar and aromatic residues (structural effect) strengthened the correlation, and all six mutants moved toward the diagonal line (Figure 2b). This finding emphasizes the importance of structural effects for improving the correspondence between experimental and computed stabilities. We observed a similar improvement using information about sequence. Information about the surrounding residues and three-dimensional structures yields better results with respect to partially buried mutations than to buried ones. This may be due to the fact that in partially buried regions, mutants are more likely to be surrounded by both hydrophobic and polar residues.

For buried mutations, we found that there was no significant improvement in the correlation coefficients obtained based on sequence and structural information. It increased the correlation in only two subgroups, and the improvement (0.02) was not statistically significant. This may be due to the fact that buried mutant residues are surrounded mainly by nonpolar residues, which means that non-specific interactions dominate.

Comparison among different measures of stability

We observed basically the same conclusions for all three measures of stability, ΔT_m, $\Delta\Delta G$ and $\Delta\Delta G^{H_2O}$. Specifically, hydrophobicity (B_l, R_f, ΔASA, $-T\Delta S_h$ and ΔC_{ph}) and entropy ($-T\Delta S_c$) are the dominant factors, respectively, in all data sets of buried mutations and in coil regions of partially buried mutations, for these three measures of stability. Interestingly, for ΔT_m and $\Delta\Delta G$ (the data obtained from thermal denaturation experiments), thermodynamic parameters, $-T\Delta S_h$ and ΔC_{ph} have more influence than other hydrophobic properties, whereas other properties reflecting hydrophobicity (B_l and R_f)

are dominants for $\Delta\Delta G^{H_2O}$ to the stability of hydrophobic buried mutations. The cross correlation between single property correlation of subgroups in partially buried mutations shows that the r values obtained for $\Delta\Delta G^{H_2O}$ are better than those of ΔT_m and $\Delta\Delta G$. This may be due to the fact that most of the amino acid properties were derived at room temperature.

Conclusions

Hydrophobicity was the major factor determining the stability of buried mutations, whereas hydrogen bonds, long-range interactions and hydrophobic interactions were all important determinants of stability for partially buried mutations. Subgroups based on secondary structure yielded significantly better results. The strongest correlations between partially buried mutations within helix, strand and turn and changes in protein stability are governed by hydrophobicity, whereas a coil has the opposite (destabilizing) tendency with respect to hydrophobicity. As compared to helix, strand and turn, regions of coil were inversely related to stability, indicating that the mechanism underlying coil stability is different from those of other secondary structures; a coil is strongly influenced by the effects of entropy. Information about sequence and structure was very important for predicting stability changes caused by partially buried mutations, but did not significantly strengthen the correlation for buried mutations.

Acknowledgements

We would like to thank Dr J.An for his help in preparing the data and Dr R.L.Jernigan for critically reading the manuscript. M.M.G. wishes to thank the Institute of Physical and Chemical Research (RIKEN) and JISTEC for providing financial support in the form of a Post Doctoral Fellowship to carry out this work. H.K. is supported by the Special Postdoctoral Researchers Program from RIKEN.

References

Akasako,A., Haruki,M., Oobatake,M. and Kanaya,S. (1997) *J. Biol. Chem.*, **272**, 18686–18693.
Akke,M. and Forsen,S. (1990) *Proteins Struct. Funct. Genet.*, **8**, 23–29.
Alber,T., Dao-pin,S., Wilson,K., Wozniak,J.A., Cook,S.P. and Matthews,B.W. (1987) *Nature*, **330**, 41–46.
Bash,P.A., Singh,U.C., Langridge,R. and Kollman,P.A. (1987) *Science*, **236**, 564–568.
Bhaskaran,R. and Ponnuswamy,P.K. (1984) *Int. J. Pept. Protein Res.*, **24**, 180–191.
Chen,Y.W., Fersht,A.R. and Henrick,K. (1993) *J. Mol. Biol.*, **234**, 1158–1170.
Chou,P.Y. and Fasman,G.D. (1978) *Adv. Enzymol.*, **47**, 45–148.
Dang,L., Merz,K.M. and Kollman,P.A. (1989) *J. Am. Chem. Soc.*, **111**, 8505–8508.
Dill,K.A. (1990) *Biochemistry*, **29**, 7133–7155.
Eisenberg,D., Weiss,R.M., Terwilliger,T.C. and Wilcox,W. (1982) *Faraday Symp. Chem. Soc.*, **17**, 109–120.
Eisenhaber,F. and Argos,P. (1993) *J. Comp. Chem.*, **14**, 1272–1280.
Eisenhaber,F., Lijnzaad,P., Argos,P., Sander,C. and Scharf,M. (1995) *J. Comp. Chem.*, **16**, 273–284.
Fasman,G.D. (1989) In Fasman,G.D. (ed.), *Prediction of Protein Structure and Principles of Protein Conformation*. Plenum press, New York, pp. 193–316.
Gills,D. and Rooman,M. (1996) *J. Mol. Biol.*, **257**, 1112–1126.
Gills,D. and Rooman,M. (1997) *J. Mol. Biol.*, **272**, 276–290.
Gromiha,M.M. and Ponnuswamy,P.K. (1993) *J. Theor. Biol.*, **165**, 87–100.
Gromiha,M.M. and Ponnuswamy,P.K. (1995) *Int. J. Pept. Protein Res.*, **45**, 225–240.
Gromiha,M.M. and Selvaraj,S. (1997) *J. Biol. Phys.*, **23**, 151–162.
Gromiha,M.M. and Selvaraj,S. (1998) *Protein Engng*, **11**, 249–251.
Gromiha,M.M. and Selvaraj,S. (1999a) *Biophys Chem.*, **77**, 49–68.
Gromiha,M.M. and Selvaraj,S. (1999b) *Prep. Biochem. Biotech.*, in press.
Gromiha,M.M., An,J., Kono,H., Oobatake,M., Uedaira,H. and Sarai,A. (1999) *Nucleic Acids Res.*, **27**, 286–288.
Iqbal,M. and Verrall,R.E. (1988) *J. Biol. Chem.*, **263**, 4159–4165.
Lee,C. and Levitt,M. (1991) *Nature*, **352**, 448–451.
Manavalan,P. and Ponnuswamy,P.K. (1977) *Arch. Biochem. Biophys*, **184**, 476–487.
Manavalan,P. and Ponnuswamy,P.K. (1978) *Nature*, **275**, 673–674.
Matsumura,M., Becktel,W.J. and Matthews,B.W. (1988) *Nature*, **334**, 406–410.
Matthews,B.W. (1995) *Adv. Protein Chem.*, **46**, 249–278.
Miyazawa,S. and Jernigan,R.L. (1994) *Protein Engng*, **7**, 1209–1220.
Momany,F.A., McGuire,R.F., Burgess,A.W. and Scheraga,H.A. (1975) *J. Phys. Chem.*, **79**, 2361–2381.
Nicholson,H., Anderson,D.E., Dao-pin,S. and Matthews,B.W. (1991) *Biochemistry*, **30**, 9816–9828.
Oobatake,M. and Ooi,T. (1977) *J. Theor. Biol.*, **67**, 567–584.
Oobatake,M. and Ooi,T. (1993) *Prog. Biophys Mol. Biol.*, **59**, 237–284.
Ooi,T., Oobatake,M., Némethy,G. and Scheraga,H.A. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 3086–3090.
Pace,C.N., Shirely,B.A., McNutt,M. and Gajiwala,K. (1996) *FASEB J.*, **10**, 75–83.
Ponnuswamy,P.K. (1993) *Prog. Biophys Mol. Biol.*, **59**, 57–103.
Ponnuswamy,P.K. and Gromiha,M.M. (1993) *Int. J. Pept. Protein Res.*, **42**, 326–341.
Ponnuswamy,P.K. and Gromiha,M.M. (1994) *J. Theor. Biol.*, **166**, 63–74.
Ponnuswamy,P.K., Warme,P.K. and Scheraga,H.A. (1973) *Proc. Natl Acad. Sci. USA*, **70**, 830–833.
Ponnuswamy,P.K., Prabhakaran,M. and Manavalan,P. (1980) *Biochim. Biophys. Acta*, **623**, 301–316.
Prabhakaran,M. and Ponnuswamy,P.K. (1979) *J. Theor. Biol.*, **80**, 485–504.
Rose,G.D. and Wolfenden,R. (1993) *Annu. Rev. Biophys Biomol. Str.*, **22**, 381–415.
Rose,G.D., Geselowitz,A.R., Lesser,G.J., Lee,R.H. and Zehfus,M.H. (1985) *Science*, **229**, 834–838.
Selvaraj,S. and Gromiha,M.M. (1998) *J. Protein Chem.*, **17**, 691–697.
Serrano,L., Horovitz,A., Avron,B., Bycroft,M. and Fersht,A.R. (1990) *Biochemistry*, **29**, 9343–9352.
Serrano,L., Kellis,J.T.,Jr., Cann,P., Matouschek,A. and Fersht,A.R. (1992) *J. Mol. Biol.*, **224**, 783–804.
Shortle,D., Stites,W.E. and Meeker,A.K. (1990) *Biochemistry*, **29**, 8033–8041.
Simonson,T. and Brunger,A.T. (1992) *Biochemistry*, **31**, 8661–8674.
Sober,H.A. (1970) *Handbook of Biochemistry: Selected Data for Molecular Biology*, 2nd edn. The Chemical Rubber Co., Cleveland, Ohio.
Sun,D.P., Sauer,U., Nicholson,H. and Matthews,B.W. (1991) *Biochemistry*, **30**, 7142–7153.
Takano,K., Ogasahara,K., Kaneda,H., Yamagata,Y., Fujii,S., Kanaya,E., Kikuchi,M., Oobatake,M. and Yutani,K. (1995) *J. Mol. Biol.*, **254**, 62–76.
Takano,K., Yamagata,Y., Fujii,S. and Yutani,K. (1997) *Biochemistry*, **36**, 688–698.
Tanford,C. (1980) *The Hydrophobic Effect*. Wiley & Sons, New York.
Tidor,B. and Karplus,M. (1991) *Biochemistry*, **30**, 3217–3228.
Tissot,A.C., Vuilleumier,S. and Fersht,A.R. (1996) *Biochemistry*, **35**, 6786–6794.
Tomii,K. and Kanehisa,M. (1996) *Protein Engng*, **9**, 27–36.
Topham,C.M., Srinivasan,N. and Blundell,T.L. (1997) *Protein Engng*, **10**, 7–21.
van Gunsteren,W.F. and Mark,A.E. (1992) *J. Mol. Biol.*, **227**, 389–395.
Vogt,G., Woell,S. and Argos,P. (1997) *J. Mol. Biol.*, **269**, 631–643.
Wetzel,R., Perry,L.J., Baase,W.A. and Becktel,W.J. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 401–405.
Xu,J., Baase,W.A., Baldwin,E. and Matthews,B.W. (1998) *Protein Sci.*, **7**, 158–177.
Yutani,K., Ogasahara,K., Tsujita,T. and Sugino,Y. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 4441–4444.
Zimmerman,J.M. Eliezer,N. and Simha,R. (1968) *J. Theor. Biol.*, **21**, 170–201.

Received December 22, 1998; revised March 24, 1999; accepted March 30, 1999