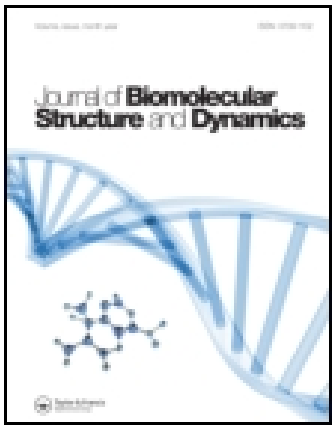

Journal of Biomolecular Structure and Dynamics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tbsd20

# Importance of Surrounding Residues for Protein Stability of Partially Buried Mutations

M. Michael Gromiha [a], Motohisa Oobatake [a], Hidetoshi Kono [a], Hatsuho Uedaira [a] & Akinori Sarai [a]

[a] RIKEN Tsukuba Institute, The Institute of Physical and Chemical Research (RIKEN), 3-1-1 Koyadai, Tsukuba, Ibaraki, 305-0074, Japan

Published online: 15 May 2012.

To cite this article: M. Michael Gromiha, Motohisa Oobatake, Hidetoshi Kono, Hatsuho Uedaira & Akinori Sarai (2000) Importance of Surrounding Residues for Protein Stability of Partially Buried Mutations, Journal of Biomolecular Structure and Dynamics, 18:2, 281-295, DOI: 10.1080/07391102.2000.10506666

To link to this article: http://dx.doi.org/10.1080/07391102.2000.10506666

*Journal of Biomolecular Structure & Dynamics, ISSN 0739-1102*
*Volume 18, Issue Number 2, (2000)*


# Importance of Surrounding Residues for Protein Stability of Partially Buried Mutations

http//www.adeninepress.com

**M. Michael Gromiha, Motohisa Oobatake†, Hidetoshi Kono, Hatsuho Uedaira and Akinori Sarai***

RIKEN Tsukuba Institute,
The Institute of Physical
and Chemical Research (RIKEN),
3-1-1 Koyadai, Tsukuba,
Ibaraki 305-0074, Japan
†Present address:
Faculty of Science and Technology,
Meijo University,
1-501 Shiogamaguchi, Tenpaku-ku,
Nagoya 468-8502, Japan

***Abstract***

For understanding the factors influencing protein stability, we have analyzed the relationship between changes in protein stability caused by partially buried mutations and changes in 48 physico-chemical, energetic and conformational properties of amino acid residues. Multiple regression equations were derived to predict the stability of protein mutants and the efficiency of the method has been verified with both back-check and jack-knife tests. We observed a good agreement between experimental and computed stabilities. Further, we have analyzed the effect of sequence window length from 1 to 12 residues on each side of the mutated residue to include the sequence information for predicting protein stability and we found that the preferred window length for obtaining the highest correlation is different for each secondary structure; the preferred window length for helical, strand and coil mutations are, respectively, 0, 9 and 4 residues on both sides of the mutant residues. However, all the secondary structures have significant correlation for a window length of one residue on each side of the mutant position, implying the role of short-range interactions. Extraction of surrounding residue information for various distances (3 to 20Å) around the mutant position showed the highest correlation at 8Å, 6Å and 7Å, respectively, for mutations in helical, strand and coil segments. Overall, the information about the surrounding residues within the sphere of 7 to 8Å, may explain better the stability in all subsets of partially buried mutations implying that this distance is sufficient to accommodate the residues influenced by major intramolecular interactions for the stability of protein structures.



## *Introduction*

The major energies, including hydrophobic, van der Waals and electrostatic interactions and hydrogen bonds (H bonds) are important to the stability of protein structures (1-5). The decisive roles of such interactions have been revealed by site directed mutagenesis experiments (6-11). The proteins such as barnase, T4 lysozyme, bovine pancreatic trypsin inhibitor, arc repressor, chymotrypsin inhibitor and staphylococcal nuclease (see Table I) have been used as a main source to study the stability changes caused by partially buried mutations.

Akke and Forsen (12) studied the effect of substituting charged residues in protein surface and observed that electrostatic interactions can have significant effects on protein stability. Also, the importance of such interactions has been stressed by other workers (13-15). Su et al. (16) found that the long-range effects along with local perturbation influence the stability of protein mutants at the surface. The contribution of H bonds to the stability of T4 lysozyme was extensively studied by Alber et al. (6), who replaced Thr157 with 13 other amino acids. Recent investigations show that the degree of which hydrophobic surface substitution destabilizes (17) or stabilizes (18) a globular protein is highly context dependent (19); moreover, steric and packing considerations are also important on the surface of a globular protein and are not as important as in the interior (19).

In our previous work, we analyzed the correlation between stability changes of proteins

*Phone: +81-298-36-9082;
Fax: +81-298-36-9080;
E-mail: sarai@rtc.riken.go.jp

caused by buried mutation and the changes in various physicochemical, energetic and conformational properties; a maximum *r* value of 0.88 was obtained for a group of mutations in β-strand segments (20). Further, we observed that hydrophobicity is the major factor for buried mutations, but hydrogen bonds, other polar interactions and hydrophobic interactions are important determinants of the stability changes caused by partially buried mutations (22). The sequence and/or structural information was found to be very important to understand the stability of mutations which are highly accessible to solvent (22). In this work, we analyzed the relationship between stability changes caused by partially buried mutations and changes in 48 amino acid properties in detail and derived multiple regression equations to predict the stability of protein mutants. We observed a good agreement between experimental and computed stabilities. Further, we analyzed the effect of sequence window length from 1 to 12 residues on each side of the mutated residue and distance from 3 to 20Å around the mutated residue for the inclusion of sequence and structural information, respectively. We found that the window length of one reside on each side of the mutant position and a distance of 7 to 8Å explain better the stability in all subsets of partially buried mutations.

## *Materials and Methods*

### *Database*

A database was set up for protein mutants from 194 proteins (21). For the present study, we selected data according to the following criteria:
**1**. Data based on thermal denaturation: (i) the solvent accessibility of the residues was between 20.0 and 50.0% for partially buried mutations and 0-2% for buried mutations (20); (ii) the pH was between 5.0 and 9.0 (around neutral pH).
**2**. Data based on denaturant (urea or GdnHCl) denaturation: in addition to the above two conditions, we restricted the temperature to between 15 and 30°C.

The resultant data set for partially buried mutations contained 161 values of $\Delta T_m$ (transition temperature change due to mutation) and 109 values of $\Delta\Delta G$ (unfolding Gibbs free energy change due to mutation), respectively, based on thermal denaturation and 251 values for $\Delta\Delta G^{H2O}$ (unfolding Gibbs free energy change due to mutation in the absence of denaturant) based on denaturant denaturation. The complete set of 320 mutations, located at 180 different sites in 24 proteins, are given in Table I. The experimental data ($\Delta T_m$, $\Delta\Delta G$ and $\Delta\Delta G^{H2O}$) can be obtained at the URL, http://www.rtc.riken.go.jp/jouhou/Protherm/protherm.html (21). We then subdivided the data set into hydrophobic mutations and mutations in selected secondary structural regions (helix, strand, coil and turn). Within the data set, 34% of the mutations were in helical segments, 20% were in β-strands, and 33% were in coil regions; while 28% of the mutations were from nonpolar to nonpolar, 8% were from nonpolar to polar, 47% were from polar to nonpolar, and 17% were from polar to polar.

### *Computational Procedure*

A detailed description of the computation of accessibility, mutation-induced changes in property values, and single and multiple correlation are described in our previous articles (20,22). A description of the computation of local sequence and surrounding environment (structure) is provided in the following sections. For this work, we used the same set of 48 properties described in the previous article (20; the correct $G_{hD}$ value for Glu is -7.37). The numerical values of all the properties are available as supplementary material.

### *Local Sequence Effect*

The effect of local sequence, $P_{seq}(i)$, was included using the equation (22),

$$P_{seq}(i) = [\sum_{j=i-k}^{j=i+k} P_j(i)] - P_{mut}(i) \quad [1]$$

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 (23,24). The computed $P_{seq}$(i) were related to the changes in experimental stability values ($\Delta T_m$, $\Delta\Delta G$ and $\Delta\Delta G^{H2O}$) using correlation coefficient. Further, we examined the effect of sequence window length ranging from 1 to 12 residues on each side of the central residue.

**Table I**

A set of 320 mutations considered in the present study.
(Table I continued on following pages).

| Protein[a] | Sequence position | Mutated residue | Mutant residue | Secondary[b] structure | *ASA* (%) | Reference |
|---|---|---|---|---|---|---|
| Adenylate kinase (1ANK) | 88 | R | G | C | 29.34 | [1] |
| Arc repressor (1ARQ) | 1 | M | A | C | 39.68 | [2] |
| | 3 | G | A | C | 49.43 | [2] |
| | 7 | M | A | C | 38.63 | [2] |
| | 8 | P | A,L | S | 32.63 | [2,3] |
| | 10 | F | A | S | 20.33 | [2] |
| | 11 | N | A | S | 27.94 | [2] |
| | 13 | R | A | S | 36.86 | [2] |
| | 15 | P | A | C | 35.10 | [2] |
| | 20 | D | A | H | 36.22 | [2] |
| | 21 | L | A | H | 41.23 | [2] |
| | 23 | R | A | H | 38.73 | [2] |
| | 24 | K | A,T | H | 48.52 | [2,4] |
| | 28 | E | A | H | 47.57 | [2] |
| | 31 | R | A | C | 35.07 | [2] |
| | 35 | S | A | H | 37.54 | [2] |
| | 38 | Y | A | H | 24.19 | [2] |
| | 39 | Q | A | H | 24.92 | [2] |
| | 42 | M | A | H | 33.13 | [2] |
| | 43 | E | A | H | 28.51 | [2] |
| | 48 | E | A | C | 47.34 | [2] |
| | 50 | R | A | C | 48.95 | [2] |
| | 52 | G | A | C | 39.77 | [2] |
| Barnase (1BNI) | 4 | I | A,V | C | 34.05 | [5] |
| | 15 | Q | I | H | 38.63 | [6] |
| | 19 | K | R | C | 35.00 | [6] |
| | 26 | T | A,D,G | H | 28.48 | [5,7] |
| | 27 | K | A,G | H | 32.96 | [5,8] |
| | 29 | E | A,G,Q,S | H | 48.20 | [5,6,7] |
| | 34 | G | A,D,H,K,N,R,S,T | C | 48.92 | [7] |
| | 36 | V | A,T | C | 46.71 | [5] |
| | 45 | V | A,T | H | 27.91 | [5] |
| | 62 | K | R | C | 42.68 | [5] |
| | 77 | N | A | C | 48.70 | [5] |
| | 84 | N | A | C | 34.70 | [5] |
| | 85 | S | A | S | 23.89 | [6] |
| | 92 | S | A | T | 47.27 | [5] |
| | 94 | W | F,L,Y | C | 29.48 | [9] |
| | 103 | Y | F | C | 28.22 | [5] |
| | 109 | I | A,V | C | 24.32 | [5] |
| | 110 | R | A | C | 31.31 | [5] |
| Bovine pancreatic trypsin inhibitor (1BPI) | 2 | P | A | H | 47.78 | [10] |
| | 7 | E | A | H | 26.79 | [10] |
| | 9 | P | A | C | 34.60 | [10] |
| | 10 | Y | A | C | 36.59 | [10] |
| | 11 | T | A | C | 47.44 | [10] |
| | 12 | G | A | C | 21.73 | [10] |
| | 18 | I | A | S | 36.97 | [10] |
| | 24 | N | A | S | 23.84 | [10] |
| | 31 | Q | A | S | 36.39 | [10] |
| | 34 | V | A | S | 40.27 | [10] |
| | 37 | G | A | C | 43.71 | [10] |
| | 41 | K | A | C | 47.59 | [10] |
| | 47 | S | A | H | 30.46 | [10] |
| | 50 | D | A | H | 36.02 | [10] |

| | | | | | | |
|---|---|---|---|---|---|---|
| | 54 | T | A | H | 36.12 | [10] |
| | 56 | G | A | H | 23.89 | [10] |
| Calbindin D9K (4ICB) | 17 | E | Q | C | 46.88 | [11] |
| Chemotactic | 14 | F | N | C | 40.01 | [12,13] |
| protein | 88 | A | G | C | 47.28 | [12] |
| (1CEY) | 90 | A | G | C | 36.39 | [12] |
| Chicken lysozyme | 15 | H | L | H | 24.30 | [14] |
| (4LYZ) | 34 | F | Y | H | 35.77 | [14] |
| Chymotrypsin | 21 | K | A,M | C | 41.37 | [15] |
| inhibitor (2CI2) | 30 | K | A | T | 41.23 | [15] |
| | 31 | S | A,G | H | 20.48 | [15] |
| | 36 | L | A,G | H | 43.90 | [15] |
| | 38 | V | A | H | 34.03 | [16] |
| | 40 | L | A,G | H | 43.53 | [15] |
| | 42 | D | A | H | 41.78 | [15] |
| | 51 | L | A,I,V | S | 21.25 | [15] |
| | 55 | T | A,S,V | C | 41.17 | [15] |
| | 57 | V | A | C | 40.27 | [15] |
| | 62 | R | A | T | 49.69 | [15] |
| | 64 | D | A | S | 21.24 | [15] |
| | 71 | D | A | S | 40.94 | [15] |
| | 72 | D | N | T | 28.41 | [15] |
| | 75 | N | A,D | C | 28.35 | [15] |
| Cytochrome C2 | 12 | K | D | H | 39.38 | [17] |
| (1C2R) | 13 | K | E | H | 47.16 | [17] |
| Dihydrofolate | 44 | R | L | H | 29.26 | [18] |
| reductase (1DYJ) | 88 | V | A,I | C | 33.77 | [19] |
| | 121 | G | A,D,C,H,L,S,V,Y | C | 45.62 | [20,21] |
| | 139 | E | K | S | 42.36 | [18] |
| Growth hormone | 71 | S | A,Q,T,V | C | 29.61 | [22] |
| (1HGU) | 74 | E | A,D,L,Q,S,T | H | 25.59 | [22] |
| Histidine containing protein HPR (1POH) | 46 | S | A,D,N | H | 27.13 | [23] |
| Interleukin 1β (1IOB) | 97 | K | G,R,V | H | 44.48 | [24] |
| λ Cro protein (1CRO) | 26 | Y | D,C,F,H,L,Q,V,W | C | 23.30 | [25] |
| λ repressor | 22 | Y | H | H | 21.71 | [26] |
| (1LRP) | 33 | Q | Y | H | 37.96 | [26] |
| | 36 | V | I | H | 30.25 | [27] |
| | 40 | M | A | C | 34.98 | [27] |
| | 44 | Q | Y | H | 45.24 | [26] |
| | 49 | A | V | H | 32.30 | [26] |
| | 78 | P | A | C | 45.17 | [28] |
| | 84 | I | S | H | 32.38 | [26] |
| P22 Tailspike | 177 | G | R | S | 36.09 | [29] |
| protein | 285 | R | K | S | 35.20 | [29] |
| (1TYU) | 382 | R | S | S | 25.94 | [29] |
| Ribonuclease HI | 70 | D | A,E,N | C | 37.54 | [30] |
| (2RN2) | 76 | Q | L | H | 32.08 | [31] |
| | 91 | K | R | C | 39.13 | [32] |
| | 119 | E | V | S | 32.63 | [31] |
| | 127 | H | A | C | 38.48 | [33] |
| | 134 | D | A,E,H,I,L,N,Q,S,T,V | H | 34.70 | [30,34,35] |
| Ribonuclease T1 | 9 | N | A | S | 30.05 | [36] |
| (1RN1) | 12 | S | A | S | 37.46 | [36] |
| | 36 | N | A | T | 46.31 | [36] |
| | 44 | N | A | T | 33.95 | [36] |
| ROP protein (1ROP) | 41 | L | A,V | H | 46.53 | [37] |
| Staphylococcal | 7 | L | A,G | S | 22.94 | [38,39] |
| nuclease (1STN) | 9 | K | A,G | S | 45.11 | [40] |
| | 14 | L | A,G | S | 22.39 | [38] |
| | 15 | I | A,G,V | S | 37.30 | [38] |
| | 18 | I | A,G,M,V | S | 26.54 | [38,41] |
| | 24 | K | A,G | S | 35.44 | [40,41] |
| | 26 | M | A,G | S | 35.48 | [38] |
| | 32 | M | A,G | S | 21.49 | [38] |
| | 33 | T | A,G,S | S | 24.59 | [41,42] |
| | 40 | D | A,G | C | 45.59 | [40] |

| | | | | | | |
|---|---|---|---|---|---|---|
| | 43 | E | A,G | C | 23.76 | [40] |
| | 46 | H | A,G | T | 21.06 | [40] |
| | 50 | G | A,V | C | 39.90 | [42] |
| | 52 | E | A,G | T | 21.98 | [40] |
| | 54 | Y | A,G | H | 20.59 | [38] |
| | 56 | P | A,G | H | 46.16 | [42] |
| | 61 | F | A,G | H | 27.90 | [38] |
| | 63 | K | A,G | H | 45.75 | [40] |
| | 67 | E | A,G | H | 49.57 | [40] |
| | 69 | A | G,T,V | C | 22.50 | [39,41,42,43] |
| | 73 | E | A,G | S | 26.22 | [40] |
| | 79 | G | A,D,S,V | C | 23.63 | [41,42,44] |
| | 83 | D | A,G | T | 24.71 | [40] |
| | 87 | R | A,G | C | 25.90 | [40] |
| | 95 | D | A,G | S | 39.42 | [40] |
| | 101 | E | A,G | H | 20.38 | [40] |
| | 105 | R | A,G | H | 32.05 | [40] |
| | 106 | Q | A,G | H | 39.42 | [42] |
| | 110 | K | A,G | C | 48.61 | [40] |
| | 117 | P | A,G,L,T | T | 47.50 | [41,42,45] |
| | 119 | N | A,G | T | 20.97 | [42] |
| | 120 | T | A,G | T | 37.56 | [42] |
| | 121 | H | A,G | H | 20.62 | [40] |
| | 126 | R | A,G | H | 40.09 | [40] |
| | 131 | Q | A,G | H | 33.09 | [42] |
| | 133 | K | A,G | H | 43.12 | [40] |
| | 135 | E | A,G | H | 44.82 | [40] |
| | 137 | L | A,G | T | 31.13 | [38,39] |
| | 138 | N | A,G | T | 30.74 | [42] |
| | 139 | I | A,G,V | T | 23.78 | [38] |
| Subtilisin BPN' | 87 | S | C | C | 33.36 | [46] |
| (1SUP) | 206 | Q | C | S | 36.17 | [47] |
| | 217 | Y | K | S | 32.52 | [47] |
| Subtilisin inhibitor | 13 | V | A, F,G, I, L, M | S | 36.56 | [48] |
| (2SSI) | | | | | | |
| T4 Lysozyme | 20 | D | A,N,S,T | S | 32.06 | [49] |
| (2LZM) | 22 | E | K | T | 47.97 | [50] |
| | 30 | G | A,F | T | 21.98 | [49] |
| | 38 | S | D,N | S | 44.45 | [51] |
| | 39 | L | A | H | 44.89 | [52] |
| | 41 | A | D,V | H | 35.39 | [50,53] |
| | 43 | K | A | H | 23.33 | [54] |
| | 45 | E | A | H | 30.57 | [54] |
| | 47 | D | A | H | 24.71 | [54] |
| | 72 | D | P | H | 47.88 | [55] |
| | 86 | P | D,H,R | H | 42.49 | [56] |
| | 90 | S | H | H | 32.42 | [57] |
| | 92 | D | N | C | 43.86 | [51] |
| | 96 | R | H | H | 30.74 | [58] |
| | 105 | Q | A,E,G | H | 32.53 | [59] |
| | 132 | N | F,I,M | H | 28.01 | [49] |
| | 147 | K | E | H | 43.90 | [60] |
| | 151 | T | S | H | 22.71 | [50] |
| | 154 | R | E | H | 34.69 | [60] |
| | 156 | G | D | C | 22.87 | [61] |
| | 157 | T | I | C | 41.82 | [62] |
| Tryptophan | 28 | P | A,G,L,S | C | 24.59 | [63,64] |
| synthase (1WSY) | 60 | D | G | C | 42.68 | [64] |
| | 207 | P | A | C | 40.73 | [63] |
| | 234 | G | D,K | C | 28.34 | [65] |

[a] The codes in parantheses correspond to the PDB codes (37) of proteins.

[b] H - helix; S - strand; T - turn and C - coil.

[1] Reinstein et al. (1989) *J. Biol. Chem. 264*, 8107-8112; [2] Milla et al. (1994) *Nat. Str. Biol. 1*, 518-523;
[3] Schildbach et al. (1995) *Biochemistry 34*, 1405-1412; [4] Bowie & Sauer (1989) *Proc. Natl. Acad. Sci. 86*, 2152-2156;
[5] Serrano et al. (1992) *J. Mol. Biol. 224*, 783-804; [6] Serrano et al. (1993) *J. Mol. Biol. 233*, 305-312;
[7] Serrano et al. (1992) *J. Mol. Biol. 227*, 544-559; [8] Meiring et al. (1992) *J. Mol. Biol. 225*, 585-589;
[9] Lowenthal et al. (1992) *J. Mol. Biol. 224*, 759-770; [10] Yu et al. (1995) *J. Mol. Biol. 249*, 388-397;
[11] Wendt et al. (1988) *Eur. J. Biochem. 175*, 439-455; [12] Lopez-hernandez & Serrano (1995) *Proteins 22*, 340-349;
[13] Munoz et al. (1994) *Biochemistry 33*, 5858-5866; [14] Shih & Kirsch (1995) *Protein Sci. 4*, 2063-2072;

[15] Itzhaki et al. (1995) *J. Mol. Biol. 254*, 260-288; [16] Jackson et al. (1993) *Biochemistry 32*, 11259-11269;
[17] Caffrey & Cusanovich (1991) *Biochemistry 30*, 9238-9241; [18] Perry et al. (1987) *Biochemistry 26*, 2674-2682;
[19] Ahrweiler & Frieden (1991) *Biochemistry 30*, 7801-7809; [20] Gekko et al. (1993) *J. Biochem. 113*, 74-80;
[21] Gekko et al. (1994) *J. Biochem. 116*, 34-41; [22] Zhukovsky et al. (1994) *Biochemistry 33*, 9856-9864;
[23] Thapar et al. (1996) *Biochemistry 35*, 11268-11277; [24] Chrunyk et al. (1993) *J. Biol. Chem. 268*, 18053-18061;
[25] Pakula & Sauer (1990) *Nature 344*, 363-364; [26] Hecht et al. (1984) *Proc. Natl. Acad. Sci. 81*, 5685-5689;
[27] Lim et al. (1992) *Biochemistry 31*, 4324-4333; [28] Reidhaar-olson et al. (1990) *Biochemistry 29*, 7563-7571;
[29] Sturtevant et al. (1989) *J. Biol. Chem. 264*, 10693-10698; [30] Kanaya et al. (1996) *J. Biol. Chem. 271*, 32729-32736;
[31] Haruki et al. (1994) *J. Biol. Chem. 269*, 26904-26911; [32] Kimura et al. (1992) *J. Biol. Chem. 267*, 22014-22017;
[33] Kanaya et al. (1991) *Eur. J. Biochem. 198*, 437-440; [34] Akasako et al. (1995) *Biochemistry 34*, 8115-8122;
[35] Haruki et al. (1994) *Eur. J. Biochem. 220*, 623-631; [36] Shirley et al. (1992) *Bichemistry 31*, 725-732;
[37] Steif et al. (1995) *Proteins 23*, 83-96; [38] Shortle et al. (1990) *Biochemistry 29*, 8033-8041;
[39] Carra & Privalov (1995) *Biochemistry 34*, 2034-2041; [40] Meeker et al. (1996) *Biochemistry 35*, 6443-6449;
[41] Green & Shortle (1993) *Biochemistry 32*, 10131-10139; [42] Green et al. (1992) *Biochemistry 31*, 5717-5728;
[43] Shortle & Meeker (1988) *Biochemistry 27*, 4761-4768; [44] Tanaka et al. (1993) *Protein Sci. 2*, 567-576;
[45] Ettink et al. (1991) *Biochemistry 30*, 1193-1199; [46] Pantoliano et al. (1987) *Biochemistry 26*, 2077-2082;
[47] Pantoliano et al. (1989) *Biochemistry 28*, 7205-7213; [48] Tamura et al. (1995) *J. Mol. Biol. 249*, 636-645;
[49] Shoichet et al. (1995) *Proc. Natl. Acad. Sci. 92*, 452-456; [50] Pjura et al. (1993) *Protein Sci. 2*, 2217-2225;
[51] Nicholson et al. (1991) *Biochemistry 30*, 9816-9828; [52] Heinz et al. (1994) *J. Mol. Biol. 236*, 869-886;
[53] Zhang et al. (1995) *Protein Eng. 8*, 1017-1022; [54] Heinz et al. (1992) *Proc. Natl. Acad. Sci. 89*, 3751-3755;
[55] Sauer et al. (1992) *J. Biol. Chem. 267*, 2393-2399; [56] Alber et al. (1988) *Nature 239*, 631-635;
[57] Dao-pin et al. (1991) *Biochemistry 30*, 7142-7153; [58] Wetzel et al. (1988) *Proc. Natl. Acad. Sci. 85*, 401-405;
[59] Pjura et al. (1993) *Proteins 15*, 401-412; [60] Dao-pin et al. (1991) *J. Mol. Biol. 221*, 873-887;
[61] Gray & Matthews (1987) *J. Biol. Chem. 262*, 16858-16864; [62] Gruetter et al. (1987) *J. Mol. Biol. 197*, 315-329;
[63] Yutani et al. (1991) *Proteins 9*, 90-98; [64] Lim et al. (1992) *Arch. Biochem. Biophys. 292*, 34-41;
[65] Beasty et al. (1986) *Biochemistry 25*, 2965-2974.



*Structural Effect*

The structural information, $P_{str}(i)$, was included using the equation (22),

$$P_{str}(i) = P_{sur}(i) - P_{mut}(i) \quad [2]$$

where $P_{mut}(i)$ is the property value of the $i^{th}$ mutant residue, and

$$P_{sur}(i) = \sum_{j} n_{ij} \cdot P_j \quad [3]$$

where $n_{ij}$ is the total number of type j residues surrounding the $i^{th}$ residue of the protein within a volume of radius 8Å, and $P_j$ is the property value of the type j residue. More details about the computation of surrounding residues have been described in our earlier articles (25-30). Further, we analyzed the effect of surrounding residues within a volume of radius ranging from 3 to 20Å in space around the central residue. These radii, 3 to 20Å is termed as distances from the mutant residue in the present analysis.

## *Results*

The investigation was carried out on three sets of data: (i) changes in thermal stability, $\Delta T_m$; (ii) unfolding Gibbs free energy changes ($\Delta\Delta G$) based on thermal denaturation; and (iii) unfolding Gibbs free energy changes ($\Delta\Delta G^{H2O}$) based on denaturant denaturation. Analyses of the correlations between amino acid properties and $\Delta T_m$, $\Delta\Delta G$ and $\Delta\Delta G^{H2O}$ were carried out after dividing the data set into subsets based on the secondary structures, helix, strand, coil and turn. With respect to $\Delta T_m$ and $\Delta\Delta G$, however, the turn regions contained an insufficient number of mutations to be used in the analysis.

### *I. Correlation of $\Delta T_m$ with properties of amino acids*

*Mutations in Helical Segments*

The *r* values obtained by relating the changes in each of the 48 properties to the experimental values of $\Delta T_m$ reveals that there was a wide distribution of *r* values with both pos-

**Table II**
Multiple Correlation of Properties with the Stability of Protein Mutants.

| Method | Data set | No. of data | Highest absolute single property correlation | Average random number correlation | Properties: back check | Properties: jack-knife test | Multiple correlation coefficient using sequence: back check | Multiple correlation coefficient using sequence: jack-knife test | Multiple correlation coefficient using structure: back check | Multiple correlation coefficient using structure: jack-knife test | Eqn |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Thermal** | | | | | | | | | | | |
| $\Delta T_m$ | hydrophobic | 44 | 0.54 | 0.16±0.12 | 0.56 | 0.42 | 0.66 (0.60) | 0.57 (0.50) | **0.66** | 0.54 | (i) |
| | helix | 68 | 0.60 | 0.14±0.11 | 0.70 | 0.65 | 0.67 (0.63) | 0.61 (0.56) | **0.70** | 0.65 | (ii) |
| | strand | 30 | 0.71 | 0.21±0.16 | **0.82** | 0.74 | 0.71 (0.77) | 0.61 (0.69) | 0.75 | 0.64 | (iii) |
| | coil (all) | 54 | 0.41 | 0.22±0.13 | 0.63 | 0.54 | 0.68 (**0.75**) | 0.59 (0.69) | 0.68 | 0.62 | (iv) |
| | (set 1) | 12 | 0.91 | 0.27±0.20 | 0.95 | 0.88 | 0.97 (**0.97**) | 0.90 (0.91) | 0.95 | 0.88 | (v) |
| | (set 2) | 14 | 0.32 | 0.15±0.09 | 0.44 | 0.34 | **0.97** (0.88) | 0.93 (0.69) | 0.94 | 0.89 | (vi) |
| | (set 3) | 17 | 0.56 | 0.30±0.14 | 0.69 | 0.43 | 0.78 (**0.81**) | 0.54 (0.68) | 0.75 | 0.59 | (vii) |
| | (set 4) | 11 | 0.80 | 0.39±0.18 | 0.89 | 0.80 | 0.95 (**0.96**) | 0.90 (0.93) | 0.93 | 0.86 | (viii) |
| $\Delta\Delta G$ | hydrophobic | 25 | 0.66 | 0.24±0.18 | 0.79 | 0.68 | 0.73 (**0.85**) | 0.59 (0.77) | 0.81 | 0.70 | (ix) |
| | helix | 52 | 0.63 | 0.16±0.11 | 0.71 | 0.65 | 0.75 (0.68) | 0.69 (0.59) | **0.80** | 0.75 | (x) |
| | strand | 24 | 0.56 | 0.21±0.16 | 0.80 | 0.58 | 0.86 (0.79) | 0.75 (0.59) | **0.89** | 0.81 | (xi) |
| **Denaturant (Urea or GdnHCl)** | | | | | | | | | | | |
| $\Delta\Delta G^{H_2O}$ | hydrophobic | 76 | 0.55 | 0.21±0.11 | 0.64 | 0.60 | 0.58 (0.60) | 0.52 (0.54) | 0.56 | 0.49 | |
| | helix | 73 | 0.61 | 0.14±0.10 | **0.69** | 0.63 | 0.46 (0.67) | 0.36 (0.62) | 0.43 | 0.32 | (xii) |
| | strand | 45 | 0.62 | 0.18±0.11 | **0.81** | 0.76 | 0.69 (0.49) | 0.62 (0.36) | 0.59 | 0.46 | (xiii) |
| | coil (all) | 110 | 0.26 | 0.07±0.05 | 0.38 | 0.27 | 0.50 (0.62) | 0.45 (0.57) | 0.61 | 0.57 | |
| | (set 1) | 12 | 0.74 | 0.26±0.17 | 0.91 | 0.84 | 0.95 (0.95) | 0.91 (0.84) | **0.95** | 0.89 | (xiv) |
| | (set 2) | 41 | 0.27 | 0.13±0.08 | 0.35 | 0.10 | 0.51 (0.58) | 0.36 (0.47) | 0.52 | 0.37 | |
| | (set 3) | 30 | 0.42 | 0.14±0.10 | 0.54 | 0.34 | **0.84** (0.77) | 0.78 (0.69) | 0.71 | 0.60 | (xv) |
| | (set 4) | 27 | 0.56 | 0.23±0.17 | 0.71 | 0.61 | 0.94 (0.93) | 0.92 (0.90) | **0.95** | 0.92 | (xvi) |
| | turn | 23 | 0.63 | 0.21±0.13 | **0.83** | 0.76 | 0.74 (0.71) | 0.51 (0.53) | 0.77 | 0.67 | (xvii) |

The multiple correlation coefficients obtained with 3-residue window are given in brackets. The highest multiple correlation coefficients among each subgroups are shown in bold. The regession equations are given for subgroups of which the multiple correlation coefficient is greater than 0.65.

(i) -7.51 $B_l$ + 1.67 $F$ + 4.91 $\alpha_n$ + 2.62; (ii) 3.14 $K^0$ - 4.24 $H_t$ + 1.33 $s$ - 6.84; (iii) 16.02 $B_l$ - 27.51 $F$ - 23.34 $\Delta C_{ph}$ - 3.42; (iv) -10.55 p$H_i$ - 5.51 $P_c$ + 5.78 -$T\Delta S_h$ + 8.30;
(v) -8.48 $B_l$ + 29.98 $E_l$ - 18.63 $B_r$ + 4.94; (vi) -12.00 p$H_i$ + 2.52 $R_f$ - 4.09 $\Delta H_h$ + 44.98; (vii) 6.85 $H_{nc}$ + 25.08 $P_\beta$ - 25.86 $N_l$ - 13.11;
(viii) 12.73 $E_l$ + 3.48 $\alpha_n$ - 13.20 $H_{gm}$ - 9.93; (ix) 1.90 p$K$' - 3.73 $B_l$ - 4.87 $E_{sm}$ + 6.64; (x) 1.40 p$K'$ - 0.78 $H_{nc}$ - 0.66 $\alpha_n$ - 2.31;
(xi) -4.08 $B_l$ + 10.37 $ASA_D$ - 10.73 $f$ + 2.98; (xii) -1.05 $P_\beta$ + 3.62 $\Delta ASA$ - 1.75 -$T\Delta S$ - 0.55; (xiii) 2.83 $P$ - 3.81 $\mu$ - 4.81 $F$ - 0.58;
(xiv) 1.06 $F$ + 0.30 $\Delta G_h$ - 1.27 -$T\Delta S_c$ - 0.46; (xv) -1.30 $H$p - 0.81 p$H_i$ + 1.05 $\Delta C_{ph}$ + 1.64; (xvi) -4.32 $\mu$ + 2.18 $C_a$ + 3.33 $\Delta H_c$ - 1.85;
(xvii) 5.02 $H_{nc}$ + 18.08 $C_a$ - 14.27 $ASA_D$ - 3.13.

itive and negative signs. The strongest positive correlation was observed for $\Delta ASA$ (solvent accessible surface area for unfolding), whereas the strongest negative correlation was for -$T\Delta S_c$ (unfolding entropy change of chain). Multiple regression analysis, which incorporated 3 properties, increased the $r$ value to 0.70 (Table II).

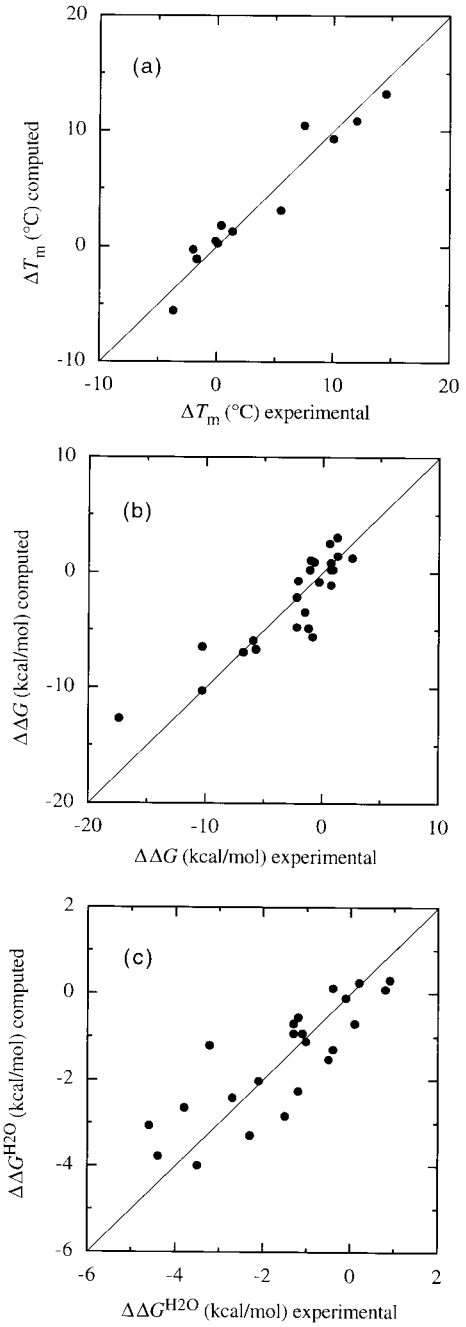


**Figure 1:** Relationship between experimental and computed stabilities for three typical subset of mutations.
(a) $\Delta T_m$; in coil regions (set 1, hb -> hb) using sequence information;
(b) $\Delta\Delta G$; in β-strand segments using structural information;
(c) $\Delta\Delta G^{H2O}$: in turn segments in the absence of sequence or structural information.

*Mutations in Strand Segments*

The analysis on the effects of mutations within strand segments showed that the β-strand tendency ($P_\beta$), followed by the properties responsible for accessibility and long range interactions ($R_a$, solvent accessible reduction ratio; $H_p$, surrounding hydrophobicity, an empirical parameter determined for each amino acid residue in protein environment (2); $B_l$, bulkiness and $N_s$, average number of surrounding residues for each amino acid residue within the distance of 8Å in protein structures) correlated significantly with experimental thermal stability. This result demonstrates the importance of long range interactions to the stability of globular proteins (24,28,29). When three properties were combined in a multiple regression analysis, the correlation was substantially strengthened ($r$ = 0.82).

*Mutations in Coil Region*

We observed no significant correlation between any amino acid properties and $\Delta T_m$ when mutations within the coil regions were treated as a single group. Mutations within regions of coil were, therefore, further subdivided based on the H-bond forming capability of the amino acids involved (22): 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); set 3 contained mutations from a hb residue to a nhb residue (hb -> nhb); and set 4 contained mutations from a nhb residue to a 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. (31).

We observed significant single-property correlations for sets 1, 3 and 4 whereas set 2 showed weak correlation. The inclusion of local sequence information increased the $r$ values to 0.97, 0.97, 0.78 and 0.95, respectively, for sets 1, 2, 3 and 4. A similar strengthening of the correlation with stability was obtained when structural information was included. Figure 1a shows the good agreement between experimental and computed $\Delta T_m$ values for set 1.

*Hydrophobic Mutations*

When only the hydrophobic mutations were considered, the strongest correlation ($r$ = 0.54) was with $H_{gm}$, combination of surrounding hydrophobicity scales obtained for globular and membrane proteins (25); other properties related to hydrophobicity ($H_p$ and $N_s$) were also significantly correlated ($r$ > 0.5) with stability.

***II. Correlation of ΔΔG with Properties of Amino Acids***

*Mutations in Helical Segments*

The strongest correlation was observed for $-T\Delta S_h$, unfolding entropy change of hydration ($r$ = 0.63), while $\Delta ASA$, $V^0$ (molar specific volume), $C_a$ (helical contact area), $ASA_D$ (solvent accessible surface area for denatured protein) and $\Delta C_{ph}$ (unfolding hydration heat capacity change) were also significantly correlated with $\Delta\Delta G$. Multiple regression analysis substantially strengthened the correlation ($r$ = 0.71). Inclusion of sequence and structural information strengthened the correlation further ($r$ = 0.75 and 0.80, respectively), demonstrating the importance of the surrounding residues in predicting the stability of proteins following partially buried mutations.

*Mutations in Strand Segments*

The analysis of mutations in β-strand segments revealed that $pH_i$ (isoelectric point) and $H_p$ were significantly correlated ($r$ > 0.5) with $\Delta\Delta G$. Multiple regression analysis strengthened the correlation ($r$ = 0.80), and sequence and structure information

strengthened it still further ($r$ = 0.86 and 0.89, respectively). A comparison of experimental and computed $\Delta\Delta G$ values using structural information is shown in Figure 1b; note the good agreement between them.

*Mutations in Coil Regions*

With respect to the mutations in set 1 (hb -> hb), $\Delta H_h$ (unfolding enthalpy change of hydration), $E_{sm}$ (short and medium-range nonbonded energy), $\Delta G_h$, (Gibbs free energy change of hydration for unfolding), $-T\Delta S_c$ and $G_{hD}$ (Gibbs free energy change of hydration for denatured protein) were all significantly correlated with $\Delta\Delta G$ ($r > 0.5$); it is noteworthy that the thermodynamic properties ($\Delta H_h$, $\Delta G_h$, $-T\Delta S_c$ and $G_{hD}$) were the most strongly correlated. The strongest negative correlation was observed for $V^0$ ($r$ = -0.80), and 18 other properties also showed significant negative correlations ($r < -0.5$). Data sets 2,3 and 4 did not contain sufficient numbers of mutations for analysis.

*Hydrophobic Mutations*

For this subset of mutations, we obtained a strong correlation between $\Delta\Delta G$ and $H_{gm}$ ($r$ = 0.66), as was the case with $\Delta T_m$. It was surprising to us that the properties influenced by long range interactions ($N_s$, $H_p$, $E_t$, total nonbonded energy, $R_a$, $B_l$, $B_r$, buriedness and $N_l$, average long-range contacts) were the most strongly correlated with stability. Multiple regression analysis increased the $r$ value to 0.79.

***III. Correlation of $\Delta\Delta G^{H2O}$ with Amino Acid Properties***

*Mutations in Helical Segments*

We obtained the strongest correlation with $\Delta ASA$ ($r$ = 0.61), and $\Delta C_{ph}$ was also significantly correlated ($r > 0.5$) with $\Delta\Delta G^{H2O}$. $-T\Delta S_c$ and $F$ (mean r.m.s fluctuational displacement) had the strongest negative correlations. Multiple regression analysis increased the $r$ value to 0.69.

*Mutations in Strand Segments*

Within strand segments, $\Delta ASA$ correlated most strongly with $\Delta\Delta G^{H2O}$ ($r$ = 0.60). In addition, 6 other properties were significantly correlated ($r > 0.5$). $P_c$ (coil tendency) had the strongest negative correlation ($r$ = -0.62). Multiple regression analysis yielded an $r$ value of 0.81.

*Mutations in Coil Regions*

For set 1, the property that correlated most strongly with $\Delta\Delta G^{H2O}$ was $K°$, compressibility ($r$ = 0.74). Multiple regression analysis strengthened the correlation ($r$ = 0.91), and inclusion of the sequence information strengthened it further ($r$ = 0.95), indicating the importance of local sequence effects. A poor correlation was observed with set 2, but sets 3 and 4 were significantly correlated. Inclusion of sequence effects yielded $r$ values of 0.58, 0.84 and 0.94, respectively; a similar increase was observed when structural information was included in the analysis.

*Mutations in Turn Regions*

Because in this case we observed a set of 23 mutations within turn regions, we analyzed that group separately. We found that $R_a$ was most strongly correlated with $\Delta\Delta G^{H2O}$ ($r$ = 0.63), while $H_p$, $P_\beta$, $B_r$, $H_{nc}$ (normalized consensus hydrophobicity) and $N_s$ were all significantly correlated ($r > 0.5$). Interestingly, for some of these properties, mutations in β-strand segments correlated more strongly with stability, and they were substantially influenced by long range interactions. Combining three

properties in a multiple regression analysis increased the *r* value to 0.83. Figure 1c shows the relationship between experimental and computed $\Delta\Delta G^{H2O}$; again there was good agreement.

*Hydrophobic Mutations*

Eleven properties correlated significantly with $\Delta\Delta G^{H2O}$ ($r > 0.5$); of those, $V^0$, $R_f$, refractive index and $\Delta ASA$ correlated most strongly ($r = 0.55$).

***IV. Validation of the Method***

*Back-Check Test*

A back-check test was carried out to verify the self consistency of the present analysis. It entailed calculating coefficients of multiple regression using all of the mutants and computing their stability by re-substituting the values. The respective *r* values obtained for all of the subgroups are given in Table II. There was good agreement between the computed and experimental stabilities: three examples are shown in Figures 1 a-c.

*Jack-Knife Test*

The validity of the present method was further tested by determining the coefficients of multiple regression using (n-1) data entries (i.e., omitting one mutant at a time) and then computing the stability of the omitted mutant. This test, termed the jack-knife test, was repeated for each of the mutants, and the *r* values were computed (Table II). We observed a good agreement between experimental and computed stabilities from the jack-knife test, which verifies the statistical significance of the experimental results.

*Amino Acid Properties and Random Numbers*

The correlations between each of the individual amino acid properties and experimental stability changes were significantly correlated in most of the subgroups ($r = 0.6$ to 0.9). In contrast, when we generated 48 sets of random numbers, normalized the values to those of the amino acid properties, and then calculated the correlation between the random numbers and the experimental stability changes ($\Delta T_m$, $\Delta\Delta G$ and $\Delta\Delta G^{H2O}$), the average *r* value fell within a range between 0.07±0.05 and 0.39±0.18 for all data sets (Table II). This verifies that we could clearly discriminate between amino acid properties and random numbers and emphasizes the validity in selecting various amino acid properties.

***Discussion***

The analysis carried out using all data yielded poor correlation due to the combined effects of irregular (coil) and ordered structures (helix, strand and turn), those have opposing roles in the folding and stability of protein molecules (22). Hence, we have analyzed the stability of protein mutants for different subsets based on secondary structures and we observed a good correlation (Table II). We have derived multiple regression equations for each subset of mutations. This method has been verified with both back-check and jack-knife tests and we observed a good agreement between experimental and computed stabilities (Table II). Hence, we suggest that these equations can be used to predict the stability of proteins caused by partially buried mutations. In buried mutations, the analysis made with all data set provided significant correlation and the subsets of data further improved the correlation.

The property, $-T\Delta S_c$ most strongly correlated with stability of coil mutations, showing the importance of entropic effects to the stability. We observed an inverse cor-

relation between coil and other secondary structures, and the correlation coefficients are given in Table III. The main contribution of stabilizing ordered structures is enthalpy that is related with hydrophobicity and the coil segments are stabilized by entropy, which has opposite relationship with hydrophobicity (22). Figure 2 shows that coil behaves in a manner opposite to that of helix, strand or turn. In buried mutations, the properties reflecting hydrophobicity are strongly correlated with the stability of protein mutants.

**Table III**

Cross correlation between single property correlations of subgroups.

| *r* | hydrophobic | helix | strand | turn* | coil |
|---|---|---|---|---|---|
| (i) $\Delta T_m$ | | | | | -0.40 |
| hydrophobic | 1.00 | 0.75 | 0.90 | | -0.62 |
| helix | | 1.00 | 0.79 | | -0.42 |
| strand | | | 1.00 | | 1.00 |
| coil | | | | | |
| (ii) $\Delta\Delta G^{H2O}$ | | | | | |
| hydrophobic | 1.00 | 0.91 | 0.94 | 0.64 | -0.71 |
| helix | | 1.00 | 0.91 | 0.76 | -0.82 |
| strand | | | 1.00 | 0.63 | -0.68 |
| turn | | | | 1.00 | -0.71 |
| coil | | | | | 1.00 |

*mutation in turn segments has significant number of data only for $\Delta\Delta G^{H2O}$.

The local sequence effect and surrounding residue information play a key role in improving the correlation and explaining the stability of (i) partially buried mutations and (ii) intermediate mutations with accessibility between buried and partially buried mutations (2-20%). This may be due to the inclusion of information from nearby polar/charged residues along the sequence and/or from the surrounding residues that are closer in structure and far in the sequence around the mutant residue (22). The importance of including structural information has been well understood from the example of coil mutations (set 4; nhb -> hb) obtained for $\Delta\Delta G^{H2O}$ in which the correlation was increased from 0.71 to 0.95 as shown in Table II. Further, we observed the improvement of correlation from 0.54 to 0.84 in another subset of coil mutations (set 3; hb -> nhb) due to the inclusion of sequence effect. Detailed analysis on the neighboring and surrounding residues showed the presence of more than 50% polar and charged residues and a few aromatic residues. These results illustrate the importance of sequence and/or structural information for understanding the stability of partially buried mutations. In buried mutations, consideration of sequence and structural information did not significantly strengthen the correlations, suggesting that nonspecific interactions dominate in the interior of proteins. The improvement was 9% for mutations with intermediate accessibility and 27% for partially buried mutations. This result suggests that inclusion of sequence and/or structural information are important for mutations that are highly accessible.

As we observed a remarkable improvement in the correlation of partially buried mutations due to the inclusion of sequence and/or structural information, we analyzed the effect of sequence window length from 1 to 12 residues on each side of the mutated residue and distance from 4 to 20Å around the mutated residue. The correlation coefficients obtained using sequence information for different window lengths are shown in Figure 3a. In this figure, the highest correlation coefficient obtained for each set of mutations are marked by filled circles. We observed that each secondary structure preferred a specific window length for obtaining the highest correlation with stability. The preferred window lengths for the secondary structures, helix, strand and coil are, respectively, 0, 9 and 4 residues on both sides of the mutant residue; hydrophobic mutations preferred the window length of 2 neighboring residues. The preference of 9 residues on both sides of the mutants in strand segments includes the information from the neighboring strands, which are close to each other in three-dimensional structures. The information from 4 residues on both sides of the mutants in coil region includes the short and medium-range interactions.

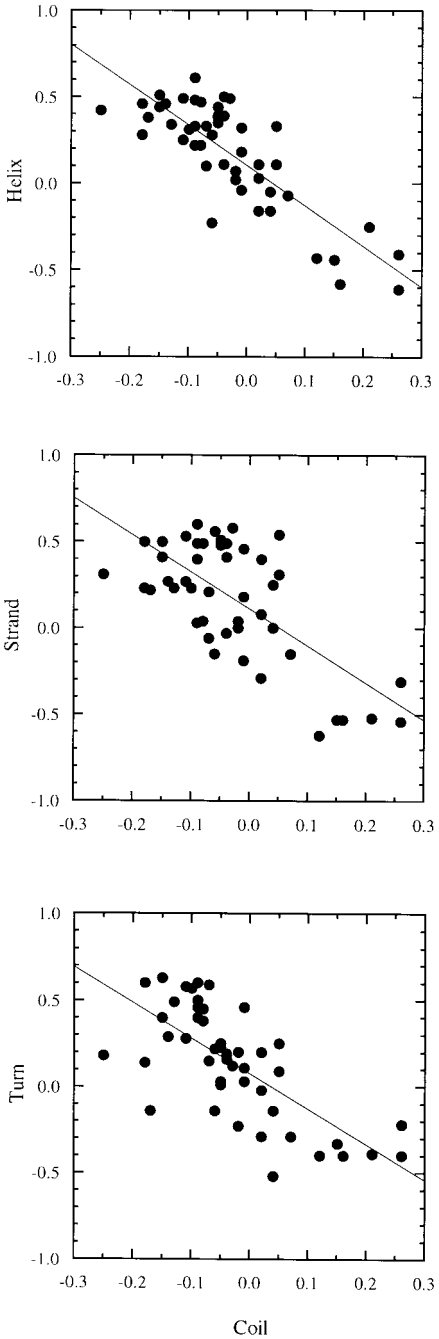


**Figure 2:** The inverse relations between *r* values obtained for mutations within coil and the other secondary structures, (i) helix, (ii) strand and (iii) turn for single-property correlations of the 48 properties under study and $\Delta\Delta G^{H2O}$.

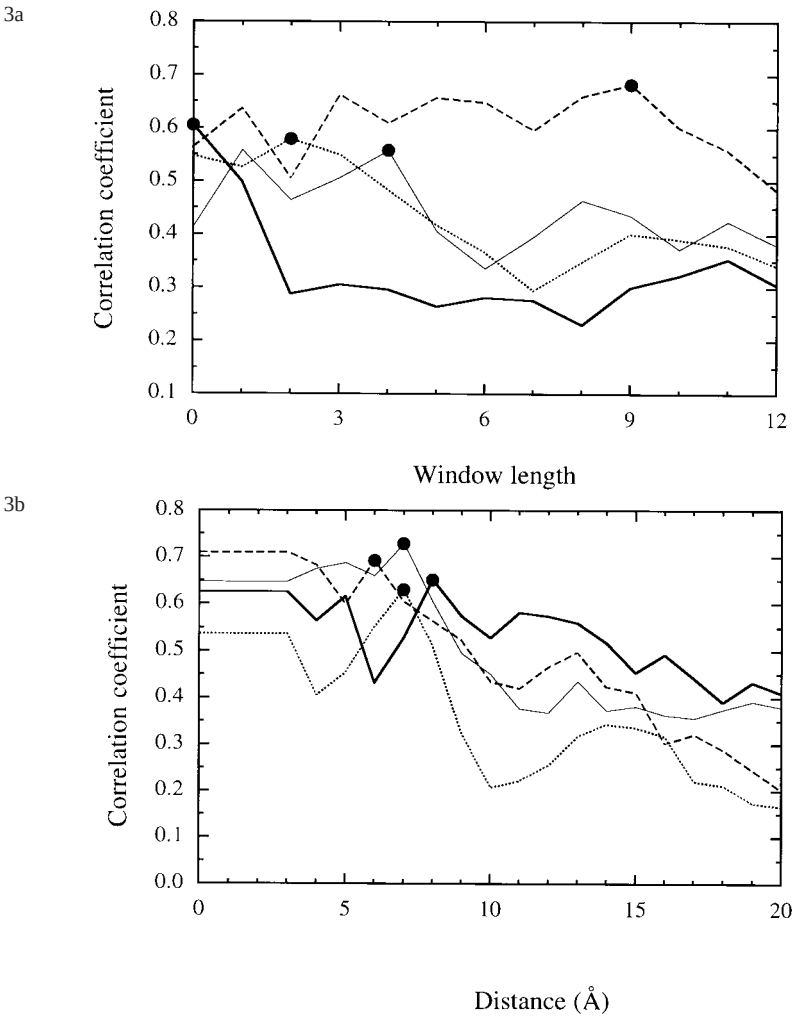


**Figure 3:** The single correlation coefficients obtained for different (a) sequence window lengths (number of residues from the mutant residue in the amino acid sequence) and (b) distances from the mutant residue (radii of the spheres around the mutant residue to accommodate the surrounding residues). thick line: helix; dash line: strand; thin line: coil; dot line: hydrophobic mutations. The highest correlation coefficient for each set of mutations are shown in filled circles.

Although, the preference of 4 residues was expected for helical mutations, we observed the highest correlation only from the information about the residues at mutant position. However, the overall behavior points out that all the mutations have significant correlation when the sequence information of one residue on each side of the mutant position is taken into account. This results indicate that the short-range interaction is effective for predicting the stability of partially buried mutations.

The correlation coefficients obtained using structural information for different distances are shown in Figure 3b. The observed highest correlation coefficient for each set of mutations showed that all the mutations prefer the distance of 6 to 8Å and specifically, helical mutations prefer 8Å; strand mutations, 6Å and coil mutations, 7Å. The distance of 6 to 8Å is sufficient to accommodate the nearest neighboring residues and several residues that are far in sequence level. The analysis on the distribution of distance between $C^{\alpha}$ atoms in 35 proteins showed that the first peak appears at 5-6Å and the second peak near 10Å, corresponding to the residues of first and second packing shells, respectively, around the central residue (32). It has been also shown that the inter-residue interactions (short, medium and long-range) obtained within the limit of 8Å can be helpful to understand the mechanism of protein folding and stability for different folds (29). Further, the radius of 6 to 8Å is widely used in the analysis of mutant stability (33), to delineate amino acid side chain clusters (34) and as effective length to include the influence of surrounding residues (35,36).

### *Conclusions*

The analysis of the relation between amino acid properties and stability changes

caused by mutations at the partially buried leads us to the following conclusions:
(i) The following properties strongly correlated with protein stability in different secondary structures: $\Delta ASA$ and $\Delta C_{ph}$ were strongly correlated in helical segments; $R_a$, $H_p$ and $B_l$ in strand segments; and $P_c$, $P_t$, $-T\Delta S_c$, $\Delta H_h$ and $\Delta G_h$ in coil regions.
(ii) Information about sequence and structure are very important to improve the predictive accuracy of partially buried mutations and the average improvement is 27%.
(iii) Multiple regression equations have been set up for each subset of partially buried mutations and can be used for predicting the stability of protein mutants.
(iv) The window length of one residue on each side of the mutant position along the sequence provides better correlation for mutation in all secondary structures. In particular, helical, strand and coil mutations preferred 0, 9 and 4 residues, respectively, on both sides of the mutant residue.
(v) The surrounding residue information obtained with the distance of 6 to 8 Å provides better correlation; specifically, 8 Å for helix, 6 Å for strand and 7 Å for coil.

***Acknowledgments***

We thank Dr. J. An for his help in preparing the data. MMG wishes to thank the Institute of Physical and Chemical Research (RIKEN) and JISTEC for providing financial support in the form of Post Doctoral Fellowship to carry out this work.



***References and Footnotes***

1. K.A. Dill, *Biochemistry 29*, 7133-7155 (1990).
2. P.K. Ponnuswamy, *Prog. Biophys. Mol. Biol. 59*, 57-103 (1993).
3. G.D. Rose and R. Wolfenden, *Ann. Rev. Biophys. Biomol. Str. 22*, 381-415 (1993).
4. P.K. Ponnuswamy and M.M. Gromiha, *J. Theor. Biol. 166*, 63-74 (1994).
5. C.N. Pace, B.A. Shirely, M. McNutt and K. Gajiwala, *FASEB J. 10*, 75-83 (1996).
6. T. Alber, S. Dao-pin, K. Wilson, J.A. Wozniak, S.P. Cook and B.W. Matthews, *Nature 330*, 41-46 (1987).
7. K. Yutani, K. Ogasahara, T. Tsujita and Y. Sugino, *Proc. Natl. Acad. Sci. USA 84*, 4441-4444 (1987).
8. B.W. Matthews, *Adv. Protein Chem. 46*, 249-278 (1995).
9. K. Takano, K. Ogasahara, H. Kaneda, Y. Yamagata, S. Fujii, E. Kanaya, M. Kikuchi, M. Oobatake and K. Yutani, *J. Mol. Biol. 254*, 62-76 (1995).
10. K. Takano, Y. Yamagata, S. Fujii and K. Yutani, *Biochemistry 36*, 688-698 (1997).
11. A. Akasako, M. Haruki, M. Oobatake and S. Kanaya, *J. Biol. Chem. 272*, 18686-18693 (1997).
12. M. Akke and S. Forsen, *Proteins: Struct. Funct. Genet. 8*, 23-29 (1990).
13. L. Serrano, A. Horovitz, B. Avron, M. Bycroft and A.R. Fersht,*Biochemistry 29*, 9343-9352 (1990).
14. S. Dao-pin, H. Nicholson, W.A. Basse, X.J. Zhang, J.A. Wozniak and B.W. Matthews, *Ciba Found Symp. 161*, 52-62 (1991).
15. H. Nicholson, D.E. Anderson, S. Dao-pin and B.W. Matthews, *Biochemistry 30*, 9816-9828 (1991).
16. S. Su, Y.G. Gao, H. Zhang, T.C. Terwilliger and A.H. Wang, *Protein Sci. 6*, 771-780 (1997).
17. B.E. Bowler, K. May, T. Zaragoza, P. York, A. Dong and W.S. Caughey,*Biochemistry 32*, 183-190 (1993).
18. B.V. Burg, B.W. Dijkstra, G. Vriend, B.V. Vinne, G. Venema and V.G. Eijsink, *Eur. J. Biochem. 220*, 981-985 (1994).
19. J.M. Schwehm, E.S. Kristyanne, C.C. Biggers and W.E. Stites, *Biochemistry 37*, 6939-6948 (1998).
20. M.M. Gromiha, M. Oobatake, H. Kono, H. Uedaira. and A. Sarai, *J. Prot. Chem. 18*, 565-578 (1999).
21. M.M. Gromiha, J. An, H. Kono, M. Oobatake, H. Uedaira and A. Sarai, *Nucl. Acids Res. 27*, 286-288 (1999).
22. M.M. Gromiha, M. Oobatake, H. Kono, H. Uedaira and A. Sarai, *Protein Engng. 12*, 549-555 (1999).
23. P.K. Ponnuswamy, P.K. Warme and H.A. Scheraga, *Proc. Natl. Acad. Sci. USA 70*, 830-833 (1973).
24. M.M. Gromiha and S. Selvaraj, *J. Biol. Phys. 23*, 151-162 (1997).
25. P.K. Ponnuswamy and M.M. Gromiha, *Int. J. Peptide Protein Res. 42*, 326- 341 (1993).
26. M.M. Gromiha and S. Selvaraj, *Protein Engng. 11*, 249-251 (1998).
27. S. Selvaraj and M.M. Gromiha, *J. Prot. Chem. 17*, 407-415 (1998).
28. S. Selvaraj and M.M. Gromiha, *J. Prot. Chem. 17*, 691-697 (1998).
29. M.M. Gromiha and S. Selvaraj, *Biophys. Chem. 77*, 49-68 (1999).
30. M.M. Gromiha and S. Selvaraj, *Prep. Biochem. Biotech. 29*, 339-351 (1999).

31. G. Vogt, S. Woell and P. Argos, *J. Mol. Biol. 269*, 631-643 (1997).
32. M. Oobatake and G.M. Crippen, *J. Phys. Chem. 85*, 1187-1197 (1981).
33. L. Serrano, J.T. Kellis, Jr, P. Cann, A. Matouschek and A.R. Fersht,*J. Mol. Biol. 224*, 783-804 (1992).
34. N. Kannan and S. Vishveshwara, *J. Mol. Biol. 292*, 441-464 (1999).
35. P. Manavalan and P.K. Ponnuswamy, *Arch. Biochem. Biophys. 184*, 476-487 (1977).
36. P. Manavalan and P.K. Ponnuswamy, *Nature 275*, 673-674 (1978).
37. H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig,I.N. Shindyalov and P.E. Bourne, *Nucl. Acids Res. 28*, 235-242 (2000).

*Date Received: May 31, 2000*

***Communicated by the Editor Ramaswamy H. Sarma***

**Supplementary Material**

Numerical values of 48 selected Physico-chemical, Energetic and Conformational Properties of the 20 amino acids/residues.

| No | Property | Ala | Asp | Cys | Glu | Phe | Gly | His | Ile | Lys | Leu | Met | Asn | Pro | Gln | Arg | Ser | Thr | Val | Trp | Tyr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | $K^0$ | -25.50 | -33.12 | -32.82 | -36.17 | -34.54 | -27.00 | -31.84 | -31.78 | -32.40 | -31.78 | -31.18 | -30.90 | -23.25 | -32.60 | -26.62 | -29.88 | -31.23 | -30.62 | -30.24 | -35.01 |
| 2 | $H_t$ | 0.87 | 0.66 | 1.52 | 0.67 | 2.87 | 0.10 | 0.87 | 3.15 | 1.64 | 2.17 | 1.67 | 0.09 | 2.77 | 0.00 | 0.85 | 0.07 | 0.07 | 1.87 | 3.77 | 2.67 |
| 3 | $Hp$ | 13.05 | 11.10 | 14.30 | 11.41 | 13.89 | 12.20 | 12.42 | 15.34 | 11.01 | 14.19 | 13.62 | 11.72 | 11.06 | 11.78 | 12.40 | 11.68 | 12.12 | 14.73 | 13.96 | 13.57 |
| 4 | $P$ | 0.00 | 49.70 | 1.48 | 49.90 | 0.35 | 0.00 | 51.60 | 0.10 | 49.50 | 0.13 | 1.43 | 3.38 | 1.58 | 3.53 | 52.00 | 1.67 | 1.66 | 0.13 | 2.10 | 1.61 |
| 5 | $pH_i$ | 6.00 | 2.77 | 5.05 | 5.22 | 5.48 | 5.97 | 7.59 | 6.02 | 9.74 | 5.98 | 5.74 | 5.41 | 6.30 | 5.65 | 10.76 | 5.68 | 5.66 | 5.96 | 5.89 | 5.66 |
| 6 | $pK'$ | 2.34 | 2.01 | 1.65 | 2.19 | 1.89 | 2.34 | 1.82 | 1.36 | 2.18 | 2.36 | 2.28 | 2.02 | 1.99 | 2.17 | 1.81 | 2.21 | 2.10 | 2.32 | 2.38 | 2.20 |
| 7 | $M_w$ | 89.00 | 133.00 | 121.00 | 147.00 | 165.00 | 75.00 | 155.00 | 131.00 | 146.00 | 131.00 | 149.00 | 132.00 | 115.00 | 146.00 | 174.00 | 105.00 | 119.00 | 117.00 | 204.00 | 181.00 |
| 8 | $B_l$ | 11.50 | 11.68 | 13.46 | 13.57 | 19.80 | 3.40 | 13.67 | 21.40 | 15.71 | 21.40 | 16.25 | 12.82 | 17.43 | 14.45 | 14.28 | 9.47 | 15.77 | 21.57 | 21.61 | 18.03 |
| 9 | $R_f$ | 9.90 | 2.80 | 2.80 | 3.20 | 18.80 | 5.60 | 8.20 | 17.10 | 3.50 | 17.60 | 14.70 | 5.40 | 14.80 | 9.00 | 4.60 | 6.90 | 9.50 | 14.30 | 17.00 | 15.00 |
| 10 | $\mu$ | 14.34 | 12.00 | 35.77 | 17.26 | 29.40 | 0.00 | 21.81 | 19.06 | 21.29 | 18.78 | 21.64 | 13.28 | 10.93 | 17.56 | 26.66 | 6.35 | 11.01 | 13.92 | 42.53 | 31.55 |
| 11 | $H_{nc}$ | 0.62 | 0.90 | 0.29 | -0.74 | 1.19 | 0.48 | -0.40 | 1.38 | -1.50 | 1.06 | 0.64 | -0.78 | 0.12 | -0.85 | -2.53 | -0.18 | -0.05 | 1.08 | 0.81 | 0.26 |
| 12 | $E_{sm}$ | 1.40 | 1.16 | 1.37 | 1.16 | 1.14 | 1.36 | 1.22 | 1.19 | 1.07 | 1.32 | 1.30 | 1.18 | 1.24 | 1.12 | 0.92 | 1.30 | 1.25 | 1.25 | 1.03 | 1.03 |
| 13 | $E_l$ | 0.49 | 0.35 | 0.67 | 0.37 | 0.72 | 0.53 | 0.54 | 0.76 | 0.30 | 0.65 | 0.65 | 0.38 | 0.46 | 0.40 | 0.55 | 0.45 | 0.52 | 0.73 | 0.83 | 0.65 |
| 14 | $E_t$ | 1.90 | 1.52 | 2.04 | 1.54 | 1.86 | 1.90 | 1.76 | 1.95 | 1.37 | 1.97 | 1.96 | 1.56 | 1.70 | 1.52 | 1.48 | 1.75 | 1.77 | 1.98 | 1.87 | 1.69 |
| 15 | $P_\alpha$ | 1.42 | 1.01 | 0.70 | 1.51 | 1.13 | 0.57 | 1.00 | 1.08 | 1.16 | 1.21 | 1.45 | 0.67 | 0.57 | 1.11 | 0.98 | 0.77 | 0.83 | 1.06 | 1.08 | 0.69 |
| 16 | $P_\beta$ | 0.83 | 0.54 | 1.19 | 0.37 | 1.38 | 0.75 | 0.87 | 1.60 | 0.74 | 1.30 | 1.05 | 0.89 | 0.55 | 1.10 | 0.93 | 0.75 | 1.19 | 1.70 | 1.37 | 1.47 |
| 17 | $P_t$ | 0.66 | 1.46 | 1.19 | 0.74 | 0.60 | 1.56 | 0.95 | 0.47 | 1.01 | 0.59 | 0.60 | 1.56 | 1.52 | 0.98 | 0.95 | 1.43 | 0.96 | 0.50 | 0.96 | 1.14 |
| 18 | $P_c$ | 0.71 | 1.21 | 1.19 | 0.84 | 0.71 | 1.52 | 1.07 | 0.66 | 0.99 | 0.69 | 0.59 | 1.37 | 1.61 | 0.87 | 1.07 | 1.34 | 1.08 | 0.63 | 0.76 | 1.07 |
| 19 | $C_a$ | 20.00 | 26.00 | 25.00 | 33.00 | 46.00 | 13.00 | 37.00 | 39.00 | 46.00 | 35.00 | 43.00 | 28.00 | 22.00 | 36.00 | 55.00 | 20.00 | 28.00 | 33.00 | 61.00 | 46.00 |
| 20 | $F$ | 0.96 | 1.14 | 0.87 | 1.07 | 0.69 | 1.16 | 0.80 | 0.76 | 1.14 | 0.79 | 0.78 | 1.04 | 1.16 | 1.07 | 1.05 | 1.13 | 0.96 | 0.79 | 0.77 | 1.01 |
| 21 | $B_r$ | 0.38 | 0.14 | 0.57 | 0.09 | 0.51 | 0.38 | 0.31 | 0.56 | 0.04 | 0.50 | 0.42 | 0.15 | 0.18 | 0.11 | 0.07 | 0.23 | 0.23 | 0.48 | 0.40 | 0.26 |
| 22 | $R_a$ | 3.70 | 2.60 | 3.03 | 3.30 | 6.60 | 3.13 | 3.57 | 7.69 | 1.79 | 5.88 | 5.21 | 2.12 | 2.12 | 2.70 | 2.53 | 2.43 | 2.60 | 7.14 | 6.25 | 3.03 |
| 23 | $N_s$ | 6.05 | 4.95 | 7.86 | 5.10 | 6.62 | 6.16 | 5.80 | 7.51 | 4.88 | 7.37 | 6.39 | 5.04 | 5.65 | 5.45 | 5.70 | 5.53 | 5.81 | 7.62 | 6.98 | 6.73 |
| 24 | $\alpha_n$ | 1.59 | 0.53 | 0.33 | 1.45 | 1.14 | 0.53 | 0.89 | 1.22 | 1.13 | 1.91 | 1.25 | 0.53 | 0.00 | 0.98 | 0.67 | 0.70 | 0.75 | 1.42 | 1.33 | 0.58 |
| 25 | $\alpha_c$ | 1.44 | 2.13 | 0.76 | 2.01 | 1.01 | 0.62 | 0.56 | 0.68 | 0.59 | 0.58 | 0.73 | 0.93 | 2.19 | 1.20 | 0.39 | 0.81 | 1.25 | 0.63 | 1.40 | 0.72 |
| 26 | $\alpha_m$ | 1.22 | 0.56 | 1.53 | 1.28 | 1.13 | 0.40 | 2.23 | 0.77 | 1.65 | 1.05 | 1.47 | 0.93 | 0.00 | 1.63 | 1.59 | 0.87 | 0.46 | 1.20 | 0.46 | 0.52 |
| 27 | $V^0$ | 60.46 | 73.83 | 67.70 | 85.88 | 121.48 | 43.25 | 98.79 | 107.72 | 108.50 | 107.75 | 105.35 | 78.01 | 82.83 | 93.90 | 127.34 | 60.62 | 76.83 | 90.78 | 143.91 | 123.60 |
| 28 | $N_m$ | 2.11 | 1.80 | 1.88 | 2.09 | 1.98 | 1.53 | 1.98 | 1.77 | 1.96 | 2.19 | 2.27 | 1.84 | 1.32 | 2.03 | 1.94 | 1.57 | 1.57 | 1.63 | 1.90 | 1.67 |
| 29 | $N_l$ | 3.92 | 2.85 | 5.55 | 2.72 | 4.53 | 4.31 | 3.77 | 5.58 | 2.79 | 4.59 | 4.14 | 3.64 | 3.57 | 3.06 | 3.78 | 3.75 | 4.09 | 5.43 | 4.83 | 4.93 |
| 30 | $H_{gm}$ | 13.85 | 11.61 | 15.37 | 11.38 | 13.93 | 13.34 | 13.82 | 15.28 | 11.58 | 14.13 | 13.86 | 13.02 | 12.35 | 12.61 | 13.10 | 13.39 | 12.70 | 14.56 | 15.48 | 13.88 |
| 31 | $ASA_D$ | 104.00 | 132.20 | 132.50 | 161.90 | 182.00 | 73.40 | 165.80 | 171.50 | 195.20 | 161.40 | 189.80 | 134.90 | 135.10 | 164.90 | 210.20 | 111.40 | 130.40 | 143.90 | 208.80 | 196.40 |
| 32 | $ASA_N$ | 33.20 | 62.40 | 17.90 | 81.00 | 33.10 | 29.20 | 57.70 | 28.30 | 107.50 | 31.10 | 41.30 | 60.50 | 60.70 | 71.50 | 94.50 | 48.70 | 52.00 | 28.10 | 39.50 | 50.40 |
| 33 | $\Delta ASA$ | 70.90 | 69.60 | 114.30 | 80.50 | 148.40 | 44.00 | 107.90 | 142.70 | 87.50 | 129.80 | 147.90 | 74.00 | 73.50 | 93.30 | 116.00 | 62.80 | 78.00 | 115.60 | 167.80 | 145.90 |
| 34 | $\Delta G_h$ | -0.54 | -2.97 | -1.64 | -3.71 | -1.06 | -0.59 | -3.38 | 0.32 | -2.19 | 0.27 | -0.60 | -3.55 | 0.32 | -3.92 | -5.96 | -3.82 | -1.97 | 0.13 | -3.80 | -5.64 |
| 35 | $G_{hD}$ | -0.58 | -6.10 | -1.91 | -7.37 | -1.35 | -0.82 | -5.57 | 0.40 | -5.97 | 0.35 | -0.71 | -6.63 | 0.56 | -7.12 | -12.78 | -6.18 | -3.66 | 0.18 | -4.71 | -8.45 |
| 36 | $G_{hN}$ | -0.06 | -3.11 | -0.27 | -3.62 | -0.28 | -0.23 | -2.18 | 0.07 | -1.70 | 0.07 | -0.10 | -3.03 | 0.23 | -3.15 | -6.85 | -2.36 | -1.69 | 0.04 | -0.88 | -2.82 |
| 37 | $\Delta H_h$ | -2.24 | -4.54 | -3.43 | -5.63 | -5.11 | -1.46 | -6.83 | -3.84 | -5.02 | -3.52 | -4.16 | -5.68 | -1.95 | -6.23 | -10.43 | -5.94 | -4.39 | -3.15 | -8.99 | -10.67 |
| 38 | $-T\Delta S_h$ | 1.70 | 1.57 | 1.79 | 1.92 | 4.05 | 0.87 | 3.45 | 4.16 | 2.83 | 3.79 | 3.56 | 2.13 | 2.27 | 2.31 | 4.47 | 2.12 | 2.42 | 3.28 | 5.19 | 5.03 |
| 39 | $\Delta C_{ph}$ | 14.22 | 2.73 | 9.41 | 3.17 | 39.06 | 4.88 | 20.05 | 41.98 | 17.68 | 38.26 | 31.67 | 3.91 | 23.69 | 3.74 | 16.66 | 6.14 | 16.11 | 32.58 | 37.69 | 30.54 |
| 40 | $\Delta G_c$ | 0.51 | 2.89 | 2.71 | 3.58 | 3.22 | 0.68 | 3.95 | -0.40 | 1.87 | -0.35 | 1.13 | 3.26 | -0.39 | 3.69 | 5.25 | 3.42 | 1.74 | -0.19 | 5.59 | 6.56 |
| 41 | $\Delta H_c$ | 2.77 | 4.72 | 8.64 | 5.69 | 11.93 | 1.23 | 7.64 | 4.03 | 3.57 | 3.69 | 7.06 | 3.64 | 1.97 | 4.47 | 6.03 | 5.80 | 4.42 | 3.45 | 13.46 | 14.41 |
| 42 | $-T\Delta S_c$ | -2.25 | -1.83 | -5.92 | -2.11 | -8.71 | -0.55 | -3.69 | -4.42 | -1.70 | -4.04 | -5.93 | -0.39 | -2.36 | -0.78 | -0.78 | -2.38 | -2.68 | -3.64 | -7.87 | -7.95 |
| 43 | $\Delta G$ | -0.02 | -0.08 | 1.08 | -0.13 | 2.16 | 0.09 | 0.56 | -0.08 | -0.32 | -0.08 | 0.53 | -0.30 | -0.06 | -0.23 | -0.71 | -0.40 | -0.24 | -0.06 | 1.78 | 0.91 |
| 44 | $\Delta H$ | 0.51 | 0.18 | 5.21 | 0.05 | 6.82 | -0.23 | 0.79 | 0.19 | -1.45 | 0.17 | 2.89 | -2.03 | 0.02 | -1.76 | -4.40 | -0.16 | 0.04 | 0.30 | 4.47 | 3.73 |
| 45 | $-T\Delta S$ | -0.54 | -0.26 | -4.14 | -0.19 | -4.66 | 0.31 | -0.23 | -0.27 | 1.13 | -0.24 | -2.36 | 1.74 | -0.08 | 1.53 | 3.69 | -0.24 | -0.28 | -0.36 | -2.69 | -2.82 |
| 46 | $v$ | 1.00 | 4.00 | 2.00 | 5.00 | 7.00 | 0.00 | 6.00 | 4.00 | 5.00 | 4.00 | 4.00 | 4.00 | 3.00 | 5.00 | 7.00 | 2.00 | 3.00 | 3.00 | 10.00 | 8.00 |
| 47 | $s$ | 0.00 | 2.00 | 0.00 | 3.00 | 2.00 | 0.00 | 2.00 | 1.00 | 0.00 | 2.00 | 0.00 | 2.00 | 0.00 | 3.00 | 5.00 | 0.00 | 1.00 | 1.00 | 2.00 | 2.00 |
| 48 | $f$ | 0.00 | 2.00 | 1.00 | 3.00 | 2.00 | 0.00 | 2.00 | 2.00 | 4.00 | 2.00 | 3.00 | 2.00 | 0.00 | 3.00 | 5.00 | 1.00 | 1.00 | 1.00 | 2.00 | 2.00 |

$K^0$, compressibility: $H_t$, thermodynamic transfer hydrophobicity; $H_p$, surrounding hydrophobicity; $P$, polarity; $pH_i$, isoelectric point; $pK'$, equilibrium constant with reference to the ionization property of COOH group; $M_w$, molecular weight; $B_l$, bulkiness; $R_f$, chromatographic index; $\mu$, refractive index; $H_{nc}$, normalized consensus hydrophobicity; $E_{sm}$, short and medium range non-bonded energy; $E_l$, long range non-bonded energy; $E_t$, total non-bonded energy ($E_{sm}$+$E_l$); $P_\alpha$, $P_\beta$, $P_t$, and $P_c$ are, respectively, α-helical, β-structure, turn and coil tendencies; $C_a$, helical contact area; $F$, mean r.m.s. fluctuational displacement; $B_r$, buriedness; $R_a$, solvent accessible reduction ratio; $N_s$, average number of surrounding residues; $\alpha_n$, $\alpha_c$ and $\alpha_m$ are, respectively, power to be at the N-terminal, C-terminal and middle of α-helix; $V^0$, partial molar volume; $N_m$ and $N_l$ are, respectively, average medium and long range contacts; $H_{gm}$, combined surrounding hydrophobicity (globular and membrane); $ASA_D$, $ASA_N$ and ΔASA are, respectively, sol-

vent accessible surface area for denatured, native and unfolding; $\Delta G_h$, $G_{hD}$ and $G_{hN}$ are, respectively, Gibbs free energy change of hydration for unfolding, denatured and native protein; $\Delta H_h$, unfolding enthalpy change of hydration; $-T\Delta S_h$, unfolding entropy change of hydration; $\Delta C_{ph}$, unfolding hydration heat capacity change; $\Delta G_c$, $\Delta H_c$ and $-T\Delta S_c$ are, respectively, unfolding Gibbs free energy, unfolding enthalpy and unfolding entropy changes of chain; $\Delta G$, $\Delta H$ and $-T\Delta S$ are respectively, unfolding Gibbs free energy change, unfolding enthalpy change and unfolding entropy change; *v*, volume (number of non-hydrogen side-chain atoms); *s*, shape (position of branch point in a side-chain); *f*, flexibility (number of side-chain dihedral angles).

$K^0$ in $m^3$/mol/Pa ($\times 10^{-15}$); $H_t$, $H_p$, $H_{nc}$, $H_{gm}$, $\Delta G_h$, $G_{hD}$, $G_{hN}$, $\Delta H_h$, $-T\Delta S_h$, $\Delta G_c$, $\Delta H_c$, $-T\Delta S_c$, $\Delta G$, $\Delta H$ and $-T\Delta S$ in kcal/mol; *P* in Debye; $pH_i$ and $pK'$ in pH units; $E_{sm}$, $E_l$ and $E_t$ in kcal/mol/atom; $B_l$, $C_a$, $ASA_D$, $ASA_N$ and $\Delta ASA$ in $Å^2$; *F* in Å; $V^0$ in $m^3$/mol ($\times 10^{-6}$); $\Delta C_{ph}$, in cal/mol/K and the rest are dimensionless quantities.