ELSEVIER

Biophysical Chemistry 107 (2004) 263–272

Biophysical Chemistry

www.elsevier.com/locate/bpc

Role of non-covalent interactions for determining the folding rate of two-state proteins

M. Michael Gromiha[a,*], K. Saraboji[b], Shandar Ahmad[c], M.N. Ponnuswamy[b], Makiko Suwa[a]

[a]*Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), Aomi Frontier Building 17F, 2-43 Aomi, Koto-ku, Tokyo 135-0064, Japan*
[b]*Department of Crystallography and Biophysics, University of Madras, Chennai 600 025, India*
[c]*Department of Biochemical Engineering and Science, Kyushu Institute of Technology, Iizuka, Kyushu, Japan*

Received 24 July 2003; received in revised form 8 September 2003; accepted 17 September 2003

Abstract

Understanding the factors influencing the folding rate of proteins is a challenging problem. In this work, we have analyzed the role of non-covalent interactions for the folding rate of two-state proteins by free-energy approach. We have computed the free-energy terms, hydrophobic, electrostatic, hydrogen-bonding and van der Waals free energies. The hydrophobic free energy has been divided into the contributions from different atoms, carbon, neutral nitrogen and oxygen, charged nitrogen and oxygen, and sulfur. All the free-energy terms have been related with the folding rates of 28 two-state proteins with single and multiple correlation coefficients. We found that the hydrophobic free energy due to carbon atoms and hydrogen-bonding free energy play important roles to determine the folding rate in combination with other free energies. The normalized energies with total number of residues showed better results than the total energy of the protein. The comparison of amino acid properties with free-energy terms indicates that the energetic terms explain better the folding rate than amino acid properties. Further, the combination of free energies with topological parameters yielded the correlation of 0.91. The present study demonstrates the importance of topology for determining the folding rate of two-state proteins.

Keywords: Free-energy approach; Hydrophobic; Hydrogen bond; Protein-folding rates; Amino acid properties; Correlation coefficient; Multiple regression technique

1. Introduction

Understanding the mechanisms for the fast folding of proteins is a challenging task. Several approaches have been put forward to reveal the major determinants of protein-folding rates. As an advance to this problem, Plaxco et al. [1] proposed the concept of contact order (CO) using the information about the average sequence separation of all contacting residues in the native state of two-state proteins, and found a significant correlation between CO and folding rates of two-state

*Corresponding author. Tel.: +81-3-3599-8046; fax: +81-3-3599-8081.
E-mail address: michael-gromiha@aist.go.jp (M. Michael Gromiha).

0301-4622/04/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2003.09.008

proteins. Debe and Goddard [2] applied a first-principles approach, based on a nucleation–condensation folding mechanism, for predicting the experimentally determined folding rates. Further, an elementary statistical mechanical model has been used to calculate the protein-folding rates [3]. Dinner and Karplus [4] performed a statistical analysis to predict the protein-folding rates and reported that both CO and stability play important roles in determining the folding rate.

Gromiha and Selvaraj [5] defined a novel parameter, long-range order (LRO) from the knowledge of long-range contacts (contact between two residues that are close in space and far in the sequence) in protein structure, and established a simple statistical model for predicting the protein-folding rates. Miller et al. [6] experimentally demonstrated that LRO is one of the best parameters that correlate with protein-refolding rates including circular permutations of ribosomal proteins S6 from *Thermus thermophilus*. Zhou and Zhou [7] combined the two parameters, CO and LRO, and proposed total contact distance for predicting the protein-folding rates. Further, neural-networks-based models have been suggested to relate folding rates of proteins from the topological parameters, CO and LRO, and the combination of these terms, total contact distance [8].

Recently, we have analyzed the relationship between amino acid properties and protein-folding rates, and found that the amino acid sequence alone is inadequate for explaining the protein-folding rates. Further, we reported that the secondary structure content and solvent accessibility play a marginal role for determining the folding rates of two-state proteins [9]. In this work, we have studied the role of non-covalent interactions for the folding rates of two-state proteins. We have computed different free-energy terms, hydrophobic, electrostatic, hydrogen-bonding and van der Waals free energies. We found that the combination of hydrophobic and hydrogen bonding is one of the major determinants for folding rates. The combination of free-energy terms with total contact distance raised the correlation up to 0.91. Our results also emphasize the importance of topology to understand the folding rates of two-state proteins.

2. Materials and methods

2.1. Data set

We have used a data set of 28 two-state proteins for which the protein-folding rates are available [7,8]. The PDB codes of the proteins are 1LMB, 2ABD, 1IMQ, 2PDD, 1NYF, 1PKS, 1SHG, 1SRL, 1FNF_9, 1FNF_10, 1HNG, 1TEN, 1TIT, 1WIT, 1CSP, 1MJC, 2AIT, 1APS, 1HDN, 1URN, 2HQI, 1PBA, 1UBQ, 2PTL, 1FKB, 1COA, 1DIV and 2VIK. The structural information for all the proteins was obtained from the Protein Data Bank [10].

2.2. Computational procedure

We followed our previous method [11] to compute various free-energy terms, such as, hydrophobic, electrostatic, hydrogen-bonding and van der Waals free energies.

2.3. Hydrophobic free energy

The hydrophobic free energy (G_{hy}) of protein folding has been computed using the expression [12]

$$G_{hy} = \sum \Delta\sigma_i [A_i \text{ (folded)} - A_i \text{ (unfolded)}] \qquad (1)$$

where A_i (folded) and A_i (unfolded) represent, respectively, the accessible surface area of each atom in the folded and unfolded (extended) states of the protein. The solvent accessible surface areas of all the atoms in the folded state were computed using the program NACCESS [13]. The extended state ASA of the atoms was taken from Eisenberg et al. [14]; they are for the amino acid X in a Gly–X–Gly sequence in a typical extended conformation. $\Delta\sigma_i$ are atomic solvation parameters for the five classes of atoms, namely, carbon (C), neutral oxygen and nitrogen (N/O), charged oxygen (O^-), charged nitrogen (N^+) and sulfur (S). The $\Delta\sigma$ values are C: 12.02, N/O: −5.86, O^-: −34.98, N^+: −19.46 and S: 35.51 cal/mol/Å^2 [11]. We have also analyzed the contribution of G_{hy} due to each atom for the folding rate of two-state proteins.

2.4. Electrostatic free energy

The electrostatic free energy (G_{el}) has been computed with the two types of interactions, namely, ion pairs and helix–dipole interactions. On the basis of the experimental observations, we have used the energy of 1 kcal/mol for exposed ion pair [15,16], 3 kcal/mol for buried ion pair [17] and 1.6 kcal/mol for charge–helix dipole interactions [18,19]. Accordingly, G_{el} is computed using the expression

$$G_{el1} = 1N_{si} + 3N_{bi} + 1.6N_{ch} \quad (2)$$

where N_{si} and N_{bi} are the number of surface and buried ion pairs in a given protein and N_{ch} are the number of charge–helix dipole interactions.

As an alternate way, we have used AMBER force field [20] to compute electrostatic free energy. In this method, G_{el} is computed using the equation

$$G_{el2} = \sum q_i q_j / \varepsilon r_{ij}, \quad (3)$$

where q_i and q_j are, respectively, the charges for the atoms i and j, and r_{ij} is the distance between them. We have used the distant-dependent dielectric constant ($\varepsilon = r_{ij}$) to take account of the dielectric damping effect of the Coulomb interactions with solvent [21].

2.5. Hydrogen-bonding free energy

The hydrogen-bonding free energy (G_{hb}) has been computed from the information about the number of hydrogen bonds in a protein. We have used the program HBPLUS [22] to calculate the number of hydrogen bonds (N_{hb}). However, as the interaction between charged residues has already been considered as ion pairs, the number of hydrogen bonds between these residues has to be excluded from the total number of hydrogen bonds in a protein. Hence, the actual number of hydrogen bonds to be included in the free-energy computation is given by $N_{hb} = N_{hb} - (N_{si} + N_{bi})$. It has been reported that the free energy due to hydrogen bond is approximately 1 kcal/mol [11], and hence, the G_{hb} is taken to be

$$G_{hb} = 1N_{hb} \quad (4)$$

2.6. Contribution from disulfide bridges

Recent site-directed mutagenesis experiments showed that the contribution of a disulfide bond to protein stability is in the range of 1.5–3.5 kcal/mol [23]. Further, from the analysis of disulfide bonds in protein structures, Thornton [24] reported that the contribution of a disulfide bond is approximately 2.3 kcal/mol. Accordingly, in this work we have used the value of 2.3 kcal/mol for each disulfide bond, and the free energy due to disulfide bond (G_{ss}) is given by

$$G_{ss} = 2.3N_{ss} \quad (5)$$

where N_{ss} is the number of disulfide bonds in a protein. We observed that only three proteins in the data set contain disulfide bonds.

2.7. van der Waals free energy

We have used the AMBER 6 force field [20] to compute the contribution of van der Waals free energy (G_{vw}) in a protein. It is given by

$$G_{vw} = 4\varepsilon_{ij} \, (A_{ij}/r_{ij}^{12} - B_{ij}/r_{ij}^{6}) \quad (6)$$

where $A_{ij} = \varepsilon_{ij}{}^*(R_{ij}{}^*)^{12}$ and $B_{ij} = 2\varepsilon_{ij}{}^*(R_{ij}{}^*)^6$; $R_{ij}{}^* = (R_i{}^* + R_j{}^*)$ and $\varepsilon_{ij}{}^* = (\varepsilon_i{}^* \varepsilon_j{}^*)^{1/2}$; R^* and ε^* are, respectively, the van der Waals radius and well depth, and these parameters are obtained from Cornell et al. [20].

2.8. Amino acid properties

We have started with a set of 49 diverse amino acid properties [25–28] and we noticed that several amino acid properties are inter-related with each other. Hence, we retained only one property if it shows the correlation more than 0.65 with any of the other properties. This procedure has resulted in a set of 15 properties. Further, we realized the presence of two properties with similar behavior.

This subjective analysis helped to omit six more properties and we derived the final set of nine properties (K^0, compressibility; H_t, thermodynamic transfer hydrophobicity; P, polarity; pH_i, isoelectric point; μ, refractive index; E_{sm}, short- and medium-range non-bonded energy; P_α, α-helical tendency; N_l, average long-range contacts and $-T\Delta S$, unfolding entropy change). In this work, we have used all the three sets of properties. However, the sets with 9 and 15 properties did not show any significant difference between the results.

2.9. Multiple regression technique

We have combined these various free-energy terms (or amino acid properties) with the aid of multiple regression technique with four different sets of input parameters. These inputs are basically from two aspects: (i) amino acid properties and (ii) free-energy terms. For each protein, we have computed the average amino acid property using the expression

$$P_{ave}(i) = \sum p_{i,j}/N \quad (7)$$

where $P_{ave}(i)$ is the average property value of ith protein and $\sum p_{ij}$ is the total property value of ith protein, which can be obtained as the sum of the property values from the respective amino acid residues (j) in a protein. Considering the first property, we have computed the average property value for all the 28 proteins. The computation has been repeated with all the 49 amino acid properties and hence we obtained 49 sets of data. We have divided these sets of properties into three groups, namely (i) all 49 properties, (ii) selected 15 properties and (iii) finally derived 9 properties. The deduction of 9 properties from the set of 49 properties has been explained above.

The last set of data is based on free-energy contributions. We have derived the input parameters for multiple regression analysis with 16 sets of data: 6 sets from the hydrophobic free energy due to C, N, N^+, O, O^- and total G_{hy}, which is the total contribution from all the atoms in a protein, 6 from the folded state G_{hy}, 2 electrostatic free energies, 1 hydrogen-bonding and van der Waals free energy each.

We have computed the multiple correlation coefficients for all possible combinations of two free energies and picked up the highest one. Further, the computations were repeated with the combination of 3–6 free-energy terms. On the other hand, we have used the set of 49, 15 and 9 properties and computed the multiple correlation coefficients for all possible combinations of 2–6 properties. The results obtained with free-energy terms and amino acid properties have been analyzed. The multiple correlation coefficients were determined using standard procedures [29].

3. Results and discussions

3.1. Contribution of hydrophobic free energy

We have computed the hydrophobic free energy for each atom in all the 28 proteins and the results are presented in Table 1. We noticed a poor correlation between G_{hy} due to the contribution from each atom (C, N/O, O^-, N^+ and S) and protein-folding rates. Our result reveals that the role of hydrophobic free energy and the burial of atoms in the protein interior are minimal for determining the folding rate of two-state proteins. This observation is consistent with the previous analysis that the role of solvent accessibility is marginal for understanding the folding rate of two-state proteins [9]. However, the combination of hydrophobic free energy with other parameters shows a good relationship with protein-folding rates (see below). Nevertheless, the burial of side chain and main chain atoms in the interior, free energy of hydration and packing play a dominant role to the stability of proteins [30,31].

3.2. Electrostatic, hydrogen-bonding and van der Waals free energies

The computed free energies due to electrostatic interactions, hydrogen-bonding and van der Waals contacts have been given in Table 2. We have repeated the calculations by perturbing the PDB coordinates and there is no significant difference in the energy values. We observed a weak correlation between any of the free-energy terms and protein-folding rates. On the other hand, it has

Table 1
Contribution of hydrophobic free energy in 28 two-state proteins

No.	PDB	N_{res}	G_{hy} (kcal/mol)						
			C	N	N^+	O	O^-	S	Total
1	1LMB	92	−74.44	10.27	5.87	6.14	6.06	−1.93	−48.04
2	2ABD	86	−75.45	8.89	12.15	5.19	12.33	−0.87	−37.76
3	1IMQ	86	−74.47	9.29	3.93	4.11	5.98	−1.51	−52.67
4	2PDD	43	−27.84	4.98	5.62	1.91	2.72	0.30	−12.33
5	1NYF	58	−50.50	5.91	2.44	2.59	3.43	0.00	−36.12
6	1PKS	77	−67.85	9.13	5.25	4.25	9.30	0.00	−39.92
7	1SHG	57	−52.05	5.71	5.96	4.29	7.85	−1.21	−29.44
8	1SRL	56	−46.58	5.80	2.16	2.99	2.13	0.00	−33.50
9	1FNF_9	89	−79.56	12.37	9.44	7.17	4.30	0.00	−46.28
10	1FNF_10	94	−80.14	11.04	6.79	6.10	3.06	0.00	−53.15
11	1HNG	175	−162.40	22.08	17.46	13.15	12.23	−6.18	−103.67
12	1TEN	96	−84.47	10.01	5.79	6.62	10.60	−0.97	−52.42
13	1TIT	89	−77.32	7.98	3.35	4.43	5.48	−3.07	−59.16
14	1WIT	93	−82.24	8.61	4.24	5.22	6.67	0.00	−57.50
15	1CSP	67	−54.35	6.06	3.81	3.31	2.89	−1.02	−39.31
16	1MJC	69	−56.08	6.24	3.08	3.72	4.77	−1.06	−39.33
17	2AIT	74	−57.72	8.20	4.95	4.17	1.55	−1.34	−40.20
18	1APS	98	−83.80	13.01	12.08	8.92	7.60	−1.39	−43.58
19	1HDN	85	−74.49	8.73	5.25	7.73	4.49	−1.82	−50.10
20	1URN	96	−88.98	11.43	9.31	7.13	4.69	−3.71	−60.13
21	2HQI	72	−58.02	7.07	5.02	5.22	4.87	−0.34	−36.18
22	1PBA	81	−67.17	9.87	7.06	5.44	14.34	0.00	−30.44
23	1UBQ	76	−69.86	9.08	7.99	4.78	3.52	−0.96	−45.46
24	2PTL	78	−56.49	7.26	2.90	3.18	6.02	0.00	−37.13
25	1FKB	107	−101.27	13.16	9.15	7.87	10.65	−2.84	−63.27
26	1COA	64	−55.79	7.64	8.36	3.70	7.15	0.11	−28.82
27	1DIV	149	−127.26	16.22	17.06	9.27	9.96	−1.21	−75.96
28	2VIK	126	−112.32	14.13	7.18	8.96	10.99	−4.03	−75.09

N_{res}, number of residues in a protein.

been reported that the hydrogen bonds and ion pairs are very important for the stability of proteins [30,32,33]. From the relationship between individual free-energy terms and protein-folding rates, we propose that (i) the free-energy terms are mainly responsible for the stability of proteins and the individual effect of these interactions to initiate and expedite the folding is minimal, (ii) the topology of proteins is more important for determining the folding rates of two-state proteins rather than the non-covalent interactions and (iii) the combination of free energies may explain better the folding rates than each term individually (see below). It has been reported that the combination of free energy with CO shows a good relationship with protein-folding rates ($r=0.79$), whereas the free energy alone has a poor correlation ($r=0.29$) with folding rates of two-state proteins [4].

3.3. Multiple regression analysis

We have started with single parameter in the fit (single correlation coefficient) and then increased the number of fitting parameters. The multiple correlation coefficients obtained for each set of data (i) 49 amino acid properties, (ii) 9 amino acid properties (there is no significant difference between the multiple correlation coefficients obtained with 15 and 9 properties), (iii) free-energy terms and (iv) free-energy terms normalized by dividing with the number of residues in each protein are displayed in Fig. 1.

Table 2
Electrostatic, hydrogen-bonding and van der Waals free energies

No.	PDB	$-G_{el1}$ (kcal/mol)	$-G_{el2}$ (kcal/mol)	$-G_{hb}$ (kcal/mol)	$-G_{ss}$ (kcal/mol)	G_{vw} (kcal/mol)
1	1LMB	6.4	23.50	86	0	461.21
2	2ABD	17.0	74.17	83	0	523.37
3	1IMQ	4.2	87.85	58	0	602.90
4	2PDD	3.2	15.85	24	0	240.05
5	1NYF	0.0	58.23	28	0	353.55
6	1PKS	21.2	30.93	31	0	489.96
7	1SHG	1.0	36.20	46	0	333.09
8	1SRL	0.0	51.75	19	0	339.84
9	1FNF_9	5.0	0.00	61	0	407.73
10	1FNF_10	0.0	0.00	58	0	425.19
11	1HNG	1.0	10.54	135	4.6	929.32
12	1TEN	0.0	0.00	72	0	569.46
13	1TIT	0.0	139.00	19	0	566.75
14	1WIT	4.0	129.94	51	0	543.29
15	1CSP	1.0	0.00	44	0	287.12
16	1MJC	1.0	71.58	51	0	295.47
17	2AIT	0.0	0.00	34	4.6	405.39
18	1APS	16.2	0.00	71	0	358.27
19	1HDN	8.6	0.00	60	0	274.50
20	1URN	6.4	63.38	78	0	450.54
21	2HQI	1.0	89.99	40	2.3	447.42
22	1PBA	4.0	0.00	62	0	461.64
23	1UBQ	4.2	33.66	61	0	388.83
24	2PTL	4.2	96.53	36	0	492.63
25	1FKB	11.0	0.00	72	0	470.98
26	1COA	1.0	30.49	46	0	348.34
27	1DIV	7.4	0.00	126	0	725.37
28	2VIK	6.4	116.76	78	0	759.46

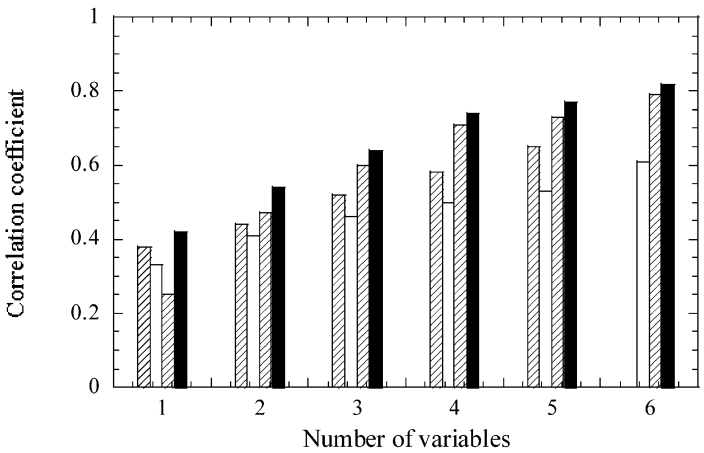

Fig. 1. Variation of correlation coefficients with number of variables: slant column, 49 amino acid properties; empty column, 9 amino acid properties; crossed column, total free-energy terms and filled column, normalized free-energy terms.

We found that the minimum number of three parameters yielded the correlation coefficients of 0.52, 0.46, 0.60 and 0.64, respectively, for 49 amino acid properties, 9 amino acid properties, 16 free-energy terms and normalized free-energy terms. Similar observation is noticed by the combination of different number of parameters (Fig. 1). This result indicated that the free-energy contributions carry more information than amino acid properties. This might be due to the knowledge observed from the three-dimensional structure of proteins. On the other hand, the combination of three amino acid properties shows an excellent correlation with the stability of proteins upon mutations [25–28].

The combination of six variables shows the correlation of 0.82 with free-energy terms while the *r*-value is 0.61 with amino acid properties. The difference of approximately 0.20 illustrates the influence of free-energy contributions to the folding rates of two-state proteins. This result is consistent with previous studies that conformational stability is one of the determinants of protein-folding rates [34]. Further studies on the free-energy contributions for high correlation revealed that the hydrophobic free energy and hydrogen-bonding free energy are the major contributors similar to the stability of globular proteins [11]. This observation indicates that the hydrophobic and hydrogen-bonding free energies are important for initiating the folding process and for maintaining the stability of proteins.

We have also analyzed the role of normalization for relating free-energy terms with folding rates of two-state proteins. As observed for CO and LRO, the normalization improved the correlation significantly for the combination of up to six parameters. Recent analysis on the effect of chain length also demonstrates the importance of normalization for understanding the protein-folding rates [35].

3.4. Combination of free energies with secondary structure content, solvent accessibility and topological parameters

We have combined the different free-energy terms with other structural parameters, such as (i) the content of amino acid residues at various secondary structures, helix, strand, turn and coil, (ii) the solvent accessibility of the protein at different secondary structures and (iii) the topological parameters, CO, LRO and TCD. The numerical values of secondary structure content and solvent accessibility information for all the 28 two-state proteins have been taken from Gromiha [9] and that of CO, LRO and TCD were taken from Zhang et al. [8].

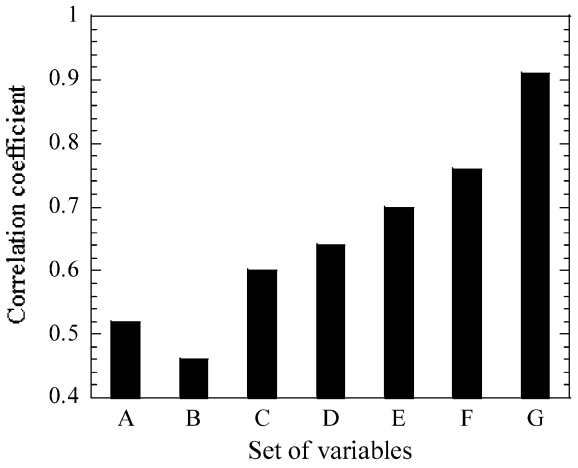

Fig. 2. Correlation coefficients obtained with three parameters for different sets of variables: *A*, 49 amino acid properties; *B*, 9 amino acid properties; *C*, total free-energy terms; *D*, normalized free-energy terms; *E*, combination of free-energy terms and solvent accessibility; *F*, combination of free-energy terms and secondary structure and *G*, combination of free-energy terms and topological parameters.

The highest correlation obtained from the combination of three parameters among (i) amino acid properties, (ii) free-energy terms, (iii) free-energy terms and secondary structure, (iv) free-energy terms and solvent accessibility and (v) free-energy terms and topological parameters is shown in Fig. 2. We observed that the highest correlation is 0.52 from the combination among amino acid properties and 0.64 with free-energy terms.

The combination of free energies with solvent accessibility improved the correlation to 0.70, and that with secondary structure raised the *r*-value up to 0.77 as seen in Fig. 2. On the other hand, the combination of free-energy terms with CO, LRO and TCD increased the correlations up to 0.86,

0.86 and 0.91, respectively. This result demonstrates that the topological parameters are the major determinants for the folding rates of two-state proteins.

We have repeated the computation with 100 random permutations of experimental folding rates. We observed that the random shuffling of ln(k) values decreased the average correlation coefficient from 0.91 to 0.32 ± 0.10. The high correlation obtained with experimental folding rates and poor correlation with random folding rates verify the validity of the present analysis.

3.5. Non-linear information and neural networks

In the present work, we have used the experimental folding rates of 28 proteins and different sets of parameters (free-energy terms, amino acid properties, secondary structure content, solvent accessibility and topological parameters), which have been linearly combined (regressions analysis) to fit with the folding rate. The linear combination provides the information about the important parameters to be combined together for the best fit. On the other hand, a non-linear model will inevitably bring in new set of terms to be optimized and it may be possible to make the highest possible correct fit. However, the non-linear model is not applicable due to the higher number of variables to be trained or optimized than the number of data points. Further, it is difficult to extract the information about the important determinants of protein-folding rates. Therefore, we have used a linear regression analysis in this study.

3.6. Implications for protein folding

The present analysis on protein-folding rates reveals the following implications for protein folding. The contribution of free-energy terms indicates a poor correlation between each of the individual free energies and protein-folding rates, which illustrates that the secondary forces are important for the stability and the contribution is less for determining the folding rates of two-state proteins. The combination of free-energy terms shows the relative importance of hydrophobic free energy and hydrogen-bonding free energy for understanding the protein-folding rates. It has been reported that the hydrophobic content of a protein is an important determinant to the folding rate [36]. The hydrogen bonds are mainly due to the formation of secondary structures and hence it stresses the importance of topology to relate with protein-folding rates. Further, the combination of free-energy terms with secondary structural content of proteins improved the correlation up to 0.77 (an increase of 20%) using three parameters (folded state G_{hy} due to O atoms, G_{el1} and α-helical content), which reveals the fact that the local secondary structure content is one of the major factors for predicting the folding rates of two-state proteins [37]. The inclusion of topological parameters raised the correlation to 0.91. This result demonstrates that the combination of topological parameters along with free-energy terms explains the folding rates of two-state proteins, and the topology is the major determinant for understanding the protein-folding rates [1–3,9,38,39].

4. Conclusions

Elucidating the factors influencing the folding rates of two-state proteins is one of the important tasks similar to protein-folding problem. We have systematically analyzed the relationship between various free-energy contributions and protein-folding rates in a set of 28 proteins. We found that the combination of free energies shows a good correlation with folding rates of two-state proteins. The relative correlation coefficients are better than the one obtained with amino acid properties. The inclusions of secondary structural content improved the correlation significantly. Further, the accommodation of topological parameters raised the correlation up to 0.91. Our results demonstrate that the topological parameters are the major determinants for the folding rates of two-state proteins.

Acknowledgments

M.M.G. wishes to thank Dr Yutaka Akiyama for encouragement.

References

[1] K.W. Plaxco, K.T. Simons, D. Baker, Contact order, transition state placement and the refolding rates of

single domain proteins, J. Mol. Biol. 277 (1998) 985–994.
[2] D.A. Debe, W.A. Goddard, First principles prediction of protein folding rates, J. Mol. Biol. 294 (1999) 619–625.
[3] V. Munoz, W.A. Eaton, A simple model for calculating the kinetics of protein folding from three-dimensional structures, Proc. Natl. Acad. Sci. USA 96 (1999) 11311–11316.
[4] A.R. Dinner, M. Karplus, The roles of stability and contact order in determining protein folding rates, Nat. Struct. Biol. 8 (2001) 21–22.
[5] M.M. Gromiha, S. Selvaraj, Comparison between long-range interactions and contact order in determining the folding rate of two-state proteins: application of long-range order to folding rate prediction, J. Mol. Biol. 310 (2001) 27–32.
[6] E.J. Miller, K.F. Fischer, S. Marqusee, Experimental evaluation of topological parameters determining protein-folding rates, Proc. Natl. Acad. Sci. USA 99 (2002) 10359–10363.
[7] H. Zhou, Y. Zhou, Folding rate prediction using total contact distance, Biophys. J. 82 (2002) 458–463.
[8] L. Zhang, J. Li, Z. Jiang, A. Zia, Folding rate prediction on neural network model, Polymer 44 (2003) 1751–1756.
[9] M.M. Gromiha, Importance of native-state topology for determining the folding rates of two-state proteins', J. Chem. Inf. Comput. Sci. 43 (2003) 1481–1485.
[10] H.M. Berman, J. Westbrook, Z. Feng, et al., The Protein Data Bank, Nucleic Acids Res. 28 (2000) 235–242.
[11] P.K. Ponnuswamy, M.M. Gromiha, On the conformational stability of folded proteins, J. Theor. Biol. 166 (1994) 63–74.
[12] D. Eisenberg, A.D. McLachlan, Solvation energy in protein folding and binding, Nature 319 (1986) 199–203.
[13] S.J. Hubbard, J.M. Thornton, NACCESS, Computer program, Department of Biochemistry and Molecular Biology, University College London, 1993. Available from: http://www.wolf.bi.umist.ac.uk/unix/naccess.html.
[14] D. Eisenberg, M. Wesson, M. Yamashita, Interpretation of protein folding and binding with atomic solvation parameters, Chem. Scripta Part A 29 (1989) 217–221.
[15] M.F. Perutz, H. Raidt, Stereochemical basis of heat stability in bacterial ferredoxins and in haemoglobin A2, Nature 255 (1975) 256–259.
[16] L.R. Brown, A. De Marco, R. Richarz, G. Wagner, K. Wuthrich, The influence of a single salt bridge on static and dynamic features of the globular solution conformation of the basic pancreatic trypsin inhibitor. 1H and 13C nuclear-magnetic-resonance studies of the native and the transaminated inhibitor, Eur. J. Biochem. 88 (1978) 87–95.
[17] A.R. Fersht, Conformational equilibria in chymotrypsin. The energetics and importance of the salt bridge, J. Mol. Biol. 64 (1972) 497–509.
[18] D. Sali, M. Bycroft, A.R. Fersht, Stabilization of protein structure by interaction of alpha-helix dipole with a charged side chain, Nature 335 (1988) 740–743.
[19] H. Nicholson, W.J. Becktel, B.W. Matthews, Enhanced protein thermostability from designed mutations that interact with alpha-helix dipoles, Nature 336 (1988) 651–656.
[20] W.D. Cornell, P. Cieplak, C.I. Bayly, et al., A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, J. Am. Chem. Soc. 117 (1995) 5179–5197.
[21] S.H. Northrup, M.R. Pear, J.D. Morgan, J.A. McCammon, M. Karplus, Molecular dynamics of ferrocytochrome *c*. Magnitude and anisotropy of atomic displacements, J. Mol. Biol. 153 (1981) 1087–1109.
[22] I.K. McDonald, J.M. Thornton, Satisfying hydrogen bonding potential in proteins, J. Mol. Biol. 238 (1994) 777–793.
[23] M.M. Gromiha, H. Uedaira, J. An, S. Selvaraj, P. Prabakaran, A. Sarai, ProTherm, thermodynamic database for proteins and mutants: developments in version 3.0, Nucleic Acids Res. 30 (2002) 301–302.
[24] J.M. Thornton, Disulphide bridges in globular proteins, J. Mol. Biol. 151 (1981) 261–287.
[25] M.M. Gromiha, J. An, H. Kono, M. Oobatake, H. Uedaira, A. Sarai, Role of structural and sequence information in the prediction of protein stability changes: comparison between buried and partially buried mutations, Protein Eng. 12 (1999) 549–555.
[26] M.M. Gromiha, J. An, H. Kono, M. Oobatake, H. Uedaira, A. Sarai, Relationship between amino acid properties and protein stability: buried mutations, J. Protein Chem. 18 (1999) 565–578.
[27] M.M. Gromiha, J. An, H. Kono, M. Oobatake, H. Uedaira, A. Sarai, Importance of surrounding residues for protein stability of partially buried mutations, J. Biomol. Struct. Dyn. 18 (2000) 281–295.
[28] M.M. Gromiha, J. An, H. Kono, M. Oobatake, H. Uedaira, A. Sarai, Importance of mutant position in Ramachandran plot for predicting protein stability of surface mutations, Biopolymers 64 (2002) 210–220.
[29] P.S. Grewal, Numerical Methods of Statistical Analysis, Sterling Publishers, New Delhi, 1987.
[30] G. Vogt, S. Woell, P. Argos, Protein thermal stability, hydrogen bonds, and ion pairs, J. Mol. Biol. 269 (1997) 631–643.
[31] M.M. Gromiha, M. Oobatake, A. Sarai, Important amino acid properties for enhanced thermostability from mesophilic to thermophilic proteins, Biophys. Chem. 82 (1999) 51–67.
[32] L. Xiao, B. Honig, Electrostatic contributions to the stability of hyperthermophilic proteins, J. Mol. Biol. 289 (1999) 1435–1444.
[33] S. Kumar, C.J. Tsai, R. Nussinov, Factors enhancing protein thermostability, Protein Eng. 13 (2000) 179–191.

[34] K.W. Plaxco, K.T. Simons, I. Ruczinski, D. Baker, Topology, stability, sequence, and length: defining the determinants of two-state protein folding kinetics, Biochemistry 39 (2000) 11177–11183.

[35] O.V. Galzitskaya, S.O. Garbuzynskiy, D.N. Ivankov, A.V. Finkelstein, Chain length is the main determinant of the folding rate for proteins with three-state folding kinetics, Proteins 51 (2003) 162–166.

[36] G. Calloni, N. Taddei, K.W. Plaxco, G. Ramponi, M. Stefani, F. Chiti, Comparison of the folding processes of distantly related proteins. Importance of hydrophobic content in folding, J. Mol. Biol. 330 (2003) 577–591.

[37] H. Gong, D.G. Isom, R. Srinivasan, G.D. Rose, Local secondary structure content predicts folding rates for simple, two-state proteins, J. Mol. Biol. 327 (5) (2003) 1149–1154.

[38] D.E. Makarov, K.W. Plaxco, The topomer search model: a simple, quantitative theory of two-state protein folding kinetics, Protein Sci. 12 (2003) 17–26.

[39] A.I. Jewett, V.S. Pande, K.W. Plaxco, Cooperativity, smooth energy landscapes and the origins of topology-dependent protein folding rates, J. Mol. Biol. 326 (2003) 247–253.