Important amino acid properties for determining the transition state structures of two-state protein mutants

M. Michael Gromiha[a,*], S. Selvaraj[b]

[a] *Computational Biology Research Center (CBRC), AIST, 2-41-6 Aomi, Koto-ku, Tokyo 135-0064, Japan*
[b] *Department of Physics, Bharathidasan University, Tiruchirapalli 620 024 Tamil Nadu, India*

Received 26 March 2002; revised 10 July 2002; accepted 10 July 2002

First published online 25 July 2002

Edited by Thomas L. James

Abstract **Understanding the mechanism in the folding pathways of proteins is an important problem in molecular biology. The Φ-value analysis provides insight into the transition state structures during protein folding. In this work, we have analyzed the relationship between the observed Φ values upon mutations in two-state proteins (FK506 binding protein, chymotrypsin inhibitor and src SH3 domain) and the changes in 48 various physico-chemical, energetic and conformational properties. We found that the classification of mutations based on solvent accessibility improved the correlation significantly. The relationship between conformational properties and Φ values determines the presence/absence of secondary structures in the transition state. In buried mutations, the physical properties volume, shape and flexibility, and the thermodynamic properties enthalpy, entropy and free-energy change have significant correlation with Φ. The short and medium-range non-bonded energy in partially buried mutations and average long-range contacts in exposed mutations showed a strong correlation with Φ values. Multiple regression analysis incorporating combinations of three properties from among all possible combinations of the 48 properties increased the correlation coefficient up to 0.99, by an average rise of 20% for all the data sets. Information about local sequence and structure is more important in surface mutations than those in buried mutations for explaining the transition state structures of two-state proteins. Further, the implications of our results for understanding the process of protein folding have been discussed.** © **2002 Published by Elsevier Science B.V. on behalf of the Federation of European Biochemical Societies.**

Key words: Amino acid property; Solvent accessibility; Φ Value; Transition state structure; Two-state protein

1. Introduction

The process of protein folding has to follow a specific pathway or set of pathways in order to fold in a finite time [1]. Elucidating the factors responsible for the folding pathways of proteins is a challenging task. Hence, several experiments have been performed using protein engineering techniques and Φ-value analysis [2] to characterize the transition state structures of two-state proteins, such as chymotrypsin inhibitor (CI2) [3], src SH3 domain (SH3) [4], FK506 binding protein [5], acyl-coenzyme binding protein [6] etc. These studies reveal that the transition state is consistent with hydrophobic collapse [7,8] and nucleation–condensation [3] mechanisms.

Recently, Nolting and Andert [9] proposed that the speed of protein folding could be understood by the formation of residue clusters which have high preference for the early formation of regular secondary structure in the presence of significant amounts of tertiary structure interactions. It has also been observed that the residue clusters play an important role in the folding and stability of globular proteins [10–12]. These clusters are formed by the residues of similar physico-chemical properties and hence the amino acid properties are one of the major determinants to understand the folding and stability of proteins [13–16].

The change in amino acid residues in a protein through protein engineering alters the physico-chemical properties and hence its interaction with other residues, which is reflected in the Φ values. In our earlier work, we have related the changes in amino acid properties with changes in stability upon buried, partially buried and surface mutations [13,16–18]. Hence, it will be interesting and important to analyze the relationship between amino acid properties and Φ values. We have extensively studied the FK506 binding protein (FKBP12) to understand the important amino acid properties responsible for determining its transition state structure and examined the results with two other proteins, CI2 and SH3.

Fulton et al. [5] characterized the transition state structure of FKBP12 by a combination of protein engineering technique, unfolding kinetics and molecular dynamics simulation. The structure of the transition state for CI2 has been analyzed by Itzhaki et al. [3] using protein engineering methods. Further, Grantcharova et al. [4] reported the role of hydrogen bonds in the structurally polarized transition state for the folding of the SH3 domain. The common feature of all these three proteins is that they fold via two-state kinetics. Although a set of Φ values are available for several mutants in these two-state proteins, the relationship between specific properties of amino acid residues and Φ values has yet to be completely explored.

In view of these facts, in the present work we have analyzed the relationship between experimental Φ values of FKBP12, CI2 and SH3 mutants and the changes in various physico-chemical, energetic and conformational properties. We observed that the physical and thermodynamic properties are the major determinants in the transition state structure of buried mutants. The medium and long-range interactions show a good relationship with Φ values in the partially buried and exposed mutations, respectively. A multiple regression analysis has been performed with different groups comprised of three properties, obtained from all possible combinations among the 48 properties, which improved the correlation co-

*Corresponding author. Fax: (81)-3-3599 8081.
E-mail address: michael-gromiha@aist.go.jp (M.M. Gromiha).

0014-5793/02/$22.00
PII: S0014-5793(02)03122-8

efficient up to 0.99. Further, the local sequence information is very important for a better understanding of transition state structures in partially buried and surface mutations than that of buried mutations.

2. Materials and methods

2.1. Experimental data and amino acid properties

The experimental Φ values of two-state proteins and a set of 48 amino acid properties form the basis of the present study. The Φ values of all the 117 mutants in FKBP12, CI2 and SH3 were taken from the literature [3–5]. The detailed information about the mutants along with their corresponding solvent accessibility (ASA) and Φ values are given in Table 1. The ASA of residues at the mutant site has been computed using the program ASC as described in our earlier article [13]. The three-dimensional structures of FKBP12 (PDB code: 1FKJ), CI2 (PDB code: 2 CI2) and SH3 (PDB code: 1SRL) have been taken from the Protein Data Bank [19]. We have divided the data set into three groups based on ASA, (i) buried (ASA < 2%), (ii) partially buried (2 < ASA < 20%) and (iii) exposed (ASA > 20%). The cutoff value of ASA has been slightly varied for a few cases in order to have a sufficient number of data in each data set.

In our earlier works [17,18], we have used a set of 48 diverse amino acid properties that fall into various clusters analyzed by Tomii and Kanehisa [20] to understand the stability of proteins upon mutations. The same set of properties has been used in the present analysis. The property values have been normalized between 0 and 1.

2.2. Computational procedure

We compute the changes in property values due to mutations, $\Delta P(i)$, using the equation [18]

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

where $P_{\mathrm{mut}}(i)$ and $P_{\mathrm{wild}}(i)$ are, respectively, the property values of the ith mutant and wild type residue. i varies from 1 to N, N being the total number of mutants. The computed differences of property values (ΔP) are related to the changes in experimental Φ values using correlation coefficients. The correlation coefficients and regression equations were determined by standard procedures [21] as described in our previous article [17].

2.3. Local sequence and structural effects

The effect of local sequence, $P_{\mathrm{seq}}(i)$, was included using the equation [18]

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

where $P_{\mathrm{mut}}(i)$ is the property value of the ith mutant residue and $\Sigma P_j(i)$ is the total property value of the segment of $(2k+1)$ residues ranging from $i-k$ to $i+k$ about the ith residue of wild type. We use a window of three and nine ($k = 1, 4$) residues to include the influence of short and medium-range interactions [22,23].

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

$$P_{\mathrm{str}}(i) = P_{\mathrm{sur}}(i) - P_{\mathrm{mut}}(i) \quad (3)$$

where $P_{\mathrm{mut}}(i)$ is the property value of the ith mutant residue, and

$$P_{\mathrm{sur}}(i) = \sum_j n_{ij} P_j \quad (4)$$

where n_{ij} is the total number of type j residues surrounding the ith residue of the protein within a volume of 8 Å and P_j is the property value of residue type j. Further details about the computation of surrounding residues have been described in our earlier articles [24,25].

Table 1
A set of 117 mutations considered in the present study

Mutation	ASA	Φ	Mutation	ASA	Φ
FK506 binding protein (5)					
V 2 A	18.09	0.55	R 57 A	58.38	0.42
V 4 A	25.63	0.39	R 57 G	58.38	0.09
I 7 V	46.70	0.16	E 60 A	10.53	0.13
T 21 A	17.09	0.44	E 60 G	10.53	0.06
T 21 S	17.09	0.55	E 61 A	35.78	0.09
T 21 V	17.09	0.61	E 61 G	35.78	0.25
V 23 A	1.43	0.55	V 63 A	0.00	0.49
V 24 A	0.00	0.44	T 75 A	13.27	0.34
T 27 A	13.27	0.38	T 75 V	17.38	0.70
T 27 S	13.27	0.63	I 76 A	0.00	0.51
T 27 V	13.27	0.31	I 76 V	0.00	0.56
F 36 A	25.66	−0.08	I 91 A	7.24	0.00
L 50 A	4.97	0.46	I 91 V	7.24	0.80
V 55 A	19.32	0.12	L 97 A	0.16	0.16
I 56 A	19.84	0.21	V 98 A	12.30	0.31
I 56 D	19.84	0.08	V 101 A	0.00	0.61
I 56 T	19.84	0.17	L 106 A	15.57	0.34
Chymotrypsin inhibitor (3)					
K 21 A	41.37	−0.19	I 49 G	40.70	0.26
K 21 M	41.37	0.03	I 49 T	40.70	0.35
T 22 A	38.57	0.13	L 51 A	21.25	0.19
T 22 G	38.57	0.05	L 51 I	21.25	−0.31
T 22 V	38.57	0.52	L 51 V	21.25	−0.04
P 25 A	50.18	0.07	V 53 A	50.68	−0.01
E 26A	50.20	0.40	V 53 G	50.68	0.16
L 27 A	0.49	0.15	V 53 T	50.68	0.23
K 30 A	41.23	−0.49	T 55 V	41.17	0.19
S 31 A	20.48	0.43	V 57 A	40.27	0.12
S 31 G	20.48	0.29	T 58 A	76.78	0.19
E 33 D	56.84	0.20	E 60 A	52.83	0.32
E 33 N	56.84	0.75	Y 61 G	49.69	0.07
E 33 Q	56.84	1.23	R 62 A	49.69	0.09
E 34 D	50.72	0.22	V 66 A	0.00	0.21
E 34 N	50.72	0.53	L 68 A	0.05	0.53
E 34 Q	50.72	0.53	F 69 A	17.19	0.30
A 35 G	0.00	1.06	F 69 L	17.19	0.28
K 36 A	43.90	0.28	F 69 V	17.19	0.25
K 36 G	43.90	0.38	V 70 A	7.87	0.25
K 37 G	67.57	0.70	D 71 A	40.94	0.12
V 38 A	34.03	−0.26	N 75 A	28.35	0.09
I 39 V	0.00	0.40	N 75 D	28.35	0.19
L 40 A	43.53	0.25	I 76 A	0.54	0.08
L 40 G	43.53	0.35	A 77 G	40.02	0.11
Q 41 G	78.16	0.12	V 79 A	25.63	−0.03
D 42 A	41.78	−0.25	V 79 G	25.63	0.04
K 43 A	4.38	−0.35	V 79 T	25.63	0.51
K 43 G	4.38	0.10	P 80 A	2.61	0.02
P 44 A	77.17	0.20	V 82 A	11.58	0.03
E 45 A	54.49	0.42	V 82 G	11.58	0.03
I 48 A	18.27	0.25	V 82 T	11.58	0.08
I 48 V	18.27	0.17	D 83 A	11.58	0.44
I 49 A	40.70	0.31			
src SH3 domain (4)					
F 10 I	19.37	0.05	L 44 A	16.54	0.37
D 15 A	48.92	0.01	A 45 G	1.08	0.72
Y 16 A	15.44	0.10	S 47 A	18.51	0.87
L 24 A	31.45	0.21	T 50 A	80.96	0.73
G 29 A	77.76	0.33	G 51 A	49.04	0.77
E 30 A	46.53	0.79	Y 55 A	36.59	0.39
I 34 A	18.24	5.50	I 56 A	0.21	0.58
W 42 A	50.68	0.07	V 61 A	3.32	−0.01

3. Results and discussion

We have carried out the investigations for three different sets of data, such as (i) buried, (ii) partially buried and (iii) exposed mutations, based on solvent accessibility. The computations have been performed using the three methods, (i) the property alone (Eq. 1), (ii) using sequence (Eq. 2) and (iii) using structural (Eq. 3) information.

3.1. Buried mutations

The correlation coefficient, r, obtained by relating the changes in each of the selected 17 properties (Eq. 1) to the

Table 2
Single property correlation with Φ values in the mutants of two-state proteins

No.	Property	Correlation coefficient, r		
		buried	partially buried	exposed[a]
FK506 binding protein				
1	Hp	−0.22	−0.30	0.69
2	E_{sm}	0.36	**0.72**	0.05
3	E_l	−0.11	−0.36	0.67
4	P_α	0.11	−0.39	0.13
5	P_β	−0.12	−0.15	0.76
6	N_s	0.13	−0.21	0.76
7	α_n	**0.87**	−0.43	0.31
8	α_c	−0.30	0.04	−0.06
9	α_m	0.00	−0.28	−0.14
10	N_l	−0.30	−0.13	**0.79**
11	H_{gm}	−0.54	−0.15	0.64
12	ΔG	−0.64	0.06	−0.29
13	ΔH	−0.59	0.18	−0.53
14	$-T\Delta S$	0.64	−0.21	−0.37
15	v	0.59	−0.63	−0.11
16	s	0.78	−0.60	0.18
17	f	0.64	−0.58	−0.27
Chymotrypsin inhibitor				
1	Hp	0.59	0.26	0.80
2	M_w	0.61	−0.14	0.19
3	E_{sm}	−0.88	**0.72**	−0.11
4	P_α	**0.93**	0.56	0.31
5	N_s	0.71	0.15	0.79
6	V^0	0.60	0.18	0.55
7	N_l	0.83	0.09	0.76
8	H_{gm}	0.43	0.28	**0.81**
9	ΔG	−0.76	−0.42	0.16
10	ΔH	−0.90	0.23	0.42
11	$-T\Delta S$	0.89	−0.29	−0.36
12	v	0.62	−0.14	0.21
13	f	0.75	−0.03	0.21
src SH3 domain				
1	Hp		0.73	**0.76**
2	E_{sm}		0.66	−0.05
3	P_α		0.37	0.30
4	P_β		**0.87**	0.63
5	N_s		0.77	0.74
6	N_l		0.85	0.66
7	H_{gm}		0.72	0.74

The highest correlation coefficient in each set of data is shown in bold.
[a]Using sequence information.

experimental Φ values of FKBP12 are given in Table 2. A perusal of this table indicates that there is a wide distribution of r values. The conformational parameter, α_n (tendency to be at the N-terminal of the α-helix) has the highest correlation ($r = 0.87$) with Φ. Interestingly, the three properties volume, shape and flexibility introduced by van Gunsteren and Mark [26] for predicting the stability of protein mutants show a very good correlation with Φ values. This result indicates that the transition state is more compact than the unfolded state, which is consistent with experimental observations [27]. Further, the properties ΔG ($r = -0.64$), ΔH ($r = -0.59$) and $-T\Delta S$ ($r = 0.64$) have significant correlation with Φ, showing the influence of thermodynamic parameters to determine the transition state structures of FKBP12, as evidenced from kinetic experiments [27]. The inclusion of sequence and/or structural information (Table 3) did not improve the single property correlation as in the case of protein stability upon buried mutations [17]. The multiple regression fit increases the correlation up to 0.99.

In CI2, the conformational parameter, P_α (α-helical tendency) shows the highest correlation ($r = 0.93$) with experimental Φ values (Table 2). This suggests that the tendency for forming an α-helix is very important in the transition state structures. Experimental studies on CI2 support our observation that α-helices are formed in the hydrophobic core of transition state structures [3]. Further, the physical and thermodynamic properties show an appreciable correlation with Φ ($|r| > 0.6$), indicating that the mutations due to the reduction of methyl groups decrease the number of contacts in the hydrophobic core, which is reflected in the Φ values, in agreement with experiments [3]. As observed in FKBP12, the inclusion of sequence/structural information did not improve the correlation (Table 3). The buried mutants of CH3 are insufficient to be used in the analysis.

3.2. Partially buried mutations

The changes in each of the property values upon partially buried mutations have been computed with Eq. 1 and the effect of sequence and structure has been included using Eqs. 2 and 3, respectively. We found that the inclusion of nearest neighboring residue information (window length of three residues in Eq. 2) improves the correlation between

Table 3
Highest single property and multiple correlation coefficients obtained for the three methods in different data sets of two-state proteins

Method	Highest absolute single correlation (multiple correlation)		
	buried	partially buried	exposed
FK506 binding protein			
Property	**0.87** (0.99)	0.64 (0.77)	0.66 (0.82)
Sequence ($w=3$)	0.84 (0.98)	**0.72** (0.80)	0.68 (0.83)
Sequence ($w=9$)	0.76 (0.98)	0.52 (0.78)	**0.79** (0.83)
Structure	0.63 (0.98)	0.43 (0.71)	0.65 (0.81)
Chymotrypsin inhibitor			
Property	**0.92** (0.97)	0.44 (0.76)	0.79 (0.92)
Sequence ($w=3$)	0.87 (0.95)	0.72 (0.92)	0.84 (0.94)
Sequence ($w=9$)	0.66 (0.93)	0.60 (0.92)	**0.84** (0.94)
Structure	0.44 (0.88)	**0.88** (0.93)	0.79 (0.92)
src SH3 domain			
Property		0.82 (0.98)	0.83 (0.96)
Sequence ($w=3$)		0.85 (0.99)	**0.90** (0.99)
Sequence ($w=9$)		0.87 (0.99)	0.87 (0.99)
Structure		**0.94** (0.99)	0.77 (0.92)

Multiple correlation coefficients are given in parentheses; the highest single property correlation coefficient among different methods for each data set is shown in bold.

$P_{seq}(i)$ and Φ values with an increase of 13% in FKBP12 (Table 3). Further increase of window length did not improve the correlation. This result is consistent with our previous analysis that the knowledge about the nearest neighboring residue itself improves the predictive accuracy of solvent accessibility [28] and protein structural class [29]. In Table 2, we present the correlation coefficients for the selected 17 properties obtained with the inclusion of sequence information in FKBP12. We found that the short and medium-range energy (E_{sm}) strongly correlated with Φ values, expressing the importance of medium-range interactions for the partially buried mutants to influence the transition state structures. A good relationship is observed between E_{sm} and Φ, as seen in Fig. 1 ($r=0.72$).

The mutation of Glu60 to Ala and Gly at the partially buried regions suggests that the α-helix is primarily unstructured in the transition state [5]. We found an insignificant correlation between helical propensities and Φ values and the correlation is 0.04 for α_c (tendency to be at the C-terminal of the α-helix). Our observation indicates that the helical propensities are not important in partially buried regions of the transition state structures of FKBP12, suggesting that the tendency to form an α-helix is very minimal, consistent with experimental results [5].

Our further analysis on CI2 and SH3 showed that the inclusion of nearest neighboring residue information improves the correlation of property values with Φ and the correlation coefficients for selected properties are presented in Table 2. As observed in FKBP12, the property E_{sm} has the highest correlation ($r=0.72$) with Φ in the helical mutants of CI2, which reveals the importance of medium-range interactions in the formation of transition state structures. Further, an appreciable correlation ($r=0.66$) was observed between E_{sm} and Φ in SH3. The relationship between E_{sm} and Φ in CI2 and SH3 is shown in Fig. 1. The multiple regression fit improved the correlation to 0.92 and 0.99, respectively, in CI2 and SH3 (Table 3).

In SH3, high Φ values are observed for the partially buried mutants in strand segments [4]. Our analysis showed that β-strand tendency (P_β) has the highest correlation ($r=0.87$) with Φ values (Table 2), demonstrating its importance in transition state structures. Our findings are supported by experimental observations that the elements in β-strands could serve as a folding nucleus during the process of SH3 folding [4].

3.3. Exposed mutations

In exposed mutations, we found that the inclusion of sequence information using a nine-residue window length remarkably improved the correlation (20%) with Φ values in FKBP12 (Table 3). On the other hand, methods such as (i) property itself and (ii) structural information showed a similar level of correlation (r lies between 0.65 and 0.68). Interestingly, the long-range interactions show a very good correlation with Φ values, similar to the stability of proteins upon

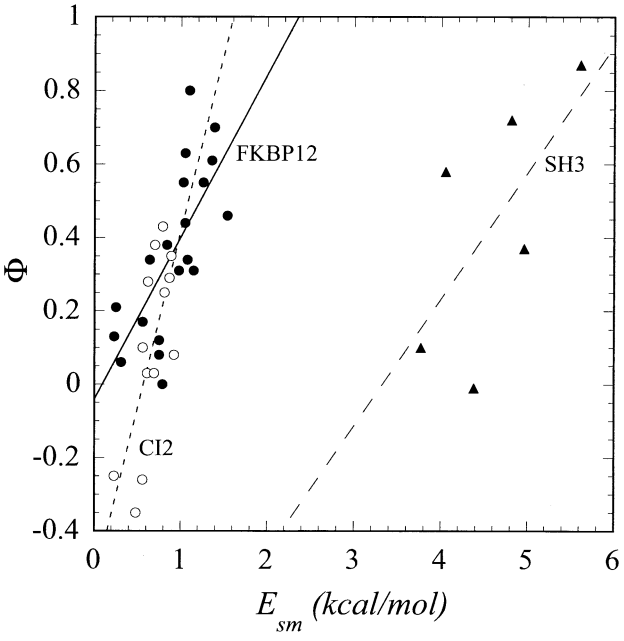

Fig. 1. Relationship between experimental Φ values and E_{sm} in partially buried mutations of two-state proteins. The symbols ●, ○ and ▲ represent FKBP12, CI2 and SH3 mutants and their respective r values are 0.72, 0.72 and 0.66.

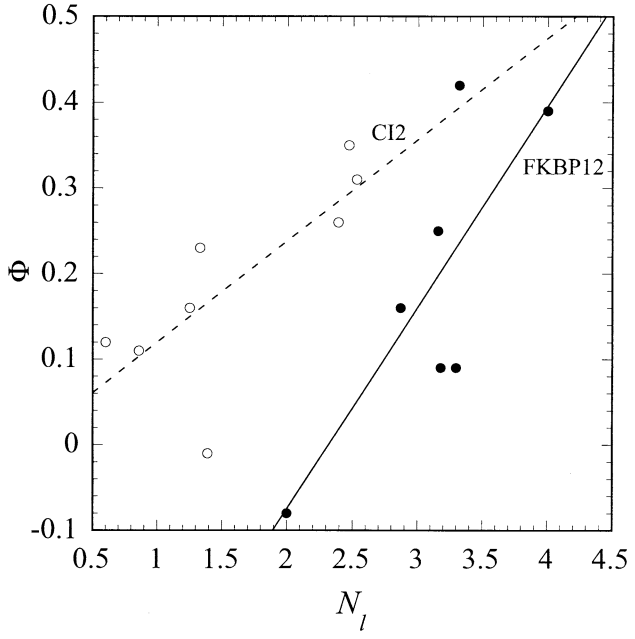

Fig. 2. Relationship between experimental Φ values and N_l in surface mutations of FKBP12 (●) and CI2 (○). The correlation coefficient for FKBP12 and CI2 is 0.79 and 0.76, respectively.

surface mutations [16] and protein folding rates [30,31], which is believed to play an important role in protein folding [24].

The correlation coefficient obtained for the selected amino acid properties by the inclusion of information from a nine-residue window in the sequence with Φ values of FKBP12 is included in Table 2. We found that the average long-range contact (N_l) has the strongest correlation with Φ ($r = 0.79$) and the relationship between them is shown in Fig. 2. Other properties reflecting long-range interactions, such as long-range non-bonded energy, average number of surrounding residues and β-strand tendency, show significant correlation with Φ values. A good correlation between P_β and Φ reflects the tendency for forming β-strands in the transition state structures of FKBP12, which is consistent with experimental observations that β-strand is weakly formed in the transition state [5]. Interestingly, there is no correlation between helical parameters and Φ values, indicating that these properties are not important in the formation of transition state structures, as observed in experiments [5]. It is noteworthy that the hydrophobicity scales including the effect of surrounding residues have a good correlation, ranging from 0.65 to 0.75. The multiple regression analysis increased the correlation up to the level of 0.83.

Similar analysis has been carried out for CI2 and SH3 and the r values are presented in Tables 2 and 3. We found that the properties reflecting long-range interactions have a strong correlation with Φ ($r > 0.75$) in both CI2 and SH3. This result is consistent with experimental observations that the long-range contacts play an important role in the formation of folding nucleus [3,4]. The relationship between N_l and Φ in exposed strand mutants of CI2 is shown in Fig. 2. Further, the multiple regression analysis raised the correlation up to 0.94 and 0.99, respectively, for CI2 and SH3.

3.4. Amino acid properties and random numbers

The correlation between each of the individual amino acid properties and experimental Φ values was significant in all sets of data ($r = 0.6$–0.9). In contrast, when we generated 48 sets of random numbers, normalized the values to those of amino acid properties, and then calculated the correlation between the random numbers and the experimental Φ values, the average r value fell within a range between 0.18 ± 0.11 and 0.37 ± 0.17 for all data sets. This verifies that we could clearly discriminate between amino acid properties and random numbers and emphasizes the validity in selecting various amino acid properties.

3.5. Implications for protein folding

The results obtained in the present study reveal the following implications for protein folding. In order to interpret the fractional Φ values upon mutations, the location of amino acid residues based on solvent accessibility seems to be important. The correlation between any of the considered 48 amino acid properties and Φ values is not significant when all the mutants were considered as a single data set. However, a remarkable improvement was observed when the mutants are grouped into buried, partially buried and exposed according to their accessibility. Similar classification also improved the stability of protein mutants [13], indicating the necessity of grouping the mutants for a better understanding of protein mutant stability as well as the transition state structures of protein mutants.

In buried mutations, the physico-chemical properties shape, volume and flexibility [26] are the major determinants for the transition state structures. This implies that there will be a decrease in number of native state contacts with the reduction of methyl groups, which decreases the Φ values. This is consistent with the previous observation of an approximately linear relationship between Φ values and nonpolar contacts [3,32,33]. In the case of partially buried mutations, the Φ values are mainly influenced by short and medium-range interaction energies rather than those due to long-range contacts. This indicates that the transition state structures are dominated by local interactions. For exposed mutations, the long-range interactions play a crucial role for determining the transition state structures. The information carried by the individual residue may not be sufficient and a cluster of residues that interacts through long-range contacts may be necessary to determine the transition state structures. Experimental studies support our suggestion that the long-range contacts play an important role in the formation of folding nucleus [3–5]. Further, the formation of secondary structures in the transition state is dictated by the relationship between conformational properties and Φ values and the results obtained in the present study are in excellent agreement with experimental observations.

4. Conclusions

We have systematically analyzed the influence of amino acid properties in the transition state structures of two-state proteins for various ranges of solvent accessibility. We found that the formation of secondary structures (α-helix or β-strand) in the transition state are reflected by a good relationship between helical or strand tendency with Φ values. The medium and long-range interactions play an important role in understanding the transition state structures of partially buried and surface mutants, respectively. The Φ values in

buried mutants are mainly explained with physical and thermodynamic properties. Further, we examined the effect of sequence and structural information to determine the transition state structures. We observed that the knowledge from amino acid sequences is sufficient for better understanding the transition state structures of partially buried and surface mutants. The results obtained in the present study could be helpful for understanding the folding mechanism of two-state proteins.

Acknowledgements: We thank the referee for constructive comments. M.M.G. wishes to thank Dr. Yutaka Akiyama and Dr. Makiko Suwa for encouragement.

References

[1] Levinthal, C. (1968) J. Chem. Phys. 85, 44–45.
[2] Fersht, A.R., Matouscheck, A. and Serrano, L. (1992) J. Mol. Biol. 224, 771–782.
[3] Itzhaki, L.S., Otzen, D.E. and Fersht, A.R. (1995) J. Mol. Biol. 254, 260–288.
[4] Grantcharova, V.P., Riddle, D.S., Santiago, J.V. and Baker, D. (1998) Nat. Struct. Biol. 5, 714–720.
[5] Fulton, K.F., Main, E.R., Daggett, V. and Jackson, S.E. (1999) J. Mol. Biol. 291, 445–461.
[6] Kragelund, B.B., Osmark, P., Neergaard, T.B., Schiodt, J., Kristiansen, K., Knudsen, J. and Poulsen, F.M. (1999) Nat. Struct. Biol. 6, 594–601.
[7] Dill, K.A., Fiebig, K.M. and Chan, H.S. (1993) Proc. Natl. Acad. Sci. USA 90, 1942–1946.
[8] Viguera, A.R., Serrano, L. and Wilmanns, M. (1996) Nat. Struct. Biol. 3, 874–880.
[9] Nolting, B. and Andert, K. (2000) Proteins 41, 288–298.
[10] Karlin, S., Zuker, M. and Brocchieri, L. (1994) J. Mol. Biol. 239, 227–248.
[11] Selvaraj, S. and Gromiha, M.M. (1998) J. Protein Chem. 17, 407–415.
[12] Kannan, N. and Vishveshwara, S. (1999) J. Mol. Biol. 292, 441–464.
[13] Gromiha, M.M., Oobatake, M., Kono, H., Uedaira, H. and Sarai, A. (1999) Protein Eng. 12, 549–555.
[14] Koehl, P. and Levitt, M. (1999) Proc. Natl. Acad. Sci. USA 96, 12524–12529.
[15] Gromiha, M.M., Oobatake, M. and Sarai, A. (1999) Biophys. Chem. 82, 51–67.
[16] Gromiha, M.M., Oobatake, M., Kono, H., Uedaira, H. and Sarai, A. (2002) Biopolymers 64, 210–220.
[17] Gromiha, M.M., Oobatake, M., Kono, H., Uedaira, H. and Sarai, A. (1999) J. Protein Chem. 18, 565–578.
[18] Gromiha, M.M., Oobatake, M., Kono, H., Uedaira, H. and Sarai, A. (2000) J. Biomol. Struct. Dyn. 18, 281–295.
[19] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) Nucleic Acids Res. 28, 235–242.
[20] Tomii, K. and Kanehisa, M. (1996) Protein Eng. 9, 27–36.
[21] Grewal, P.S. (1987) Numerical Methods of Statistical Analysis, Sterling Publ., New Delhi.
[22] Ponnuswamy, P.K. (1993) Prog. Biophys. Mol. Biol. 59, 57–103.
[23] Gromiha, M.M. and Selvaraj, S. (1997) J. Biol. Phys. 23, 151–162.
[24] Gromiha, M.M. and Selvaraj, S. (1999) Biophys. Chem. 77, 49–68.
[25] Gromiha, M.M. and Selvaraj, S. (2000) Recent Res. Devel. Biophys. Chem. 1, 1–14.
[26] van Gunsteren, W.F. and Mark, A.E. (1992) J. Mol. Biol. 227, 389–395.
[27] Main, E.R., Fulton, K.F. and Jackson, S.E. (1999) J. Mol. Biol. 291, 429–444.
[28] Ahmad, S. and Gromiha, M.M. (2002) Bioinformatics 18, 819–824.
[29] Kumarevel, T.S., Gromiha, M.M. and Ponnuswamy, M.N. (2000) Biophys. Chem. 88, 81–101.
[30] Gromiha, M.M. and Selvaraj, S. (2001) J. Mol. Biol. 310, 27–32.
[31] Zhou, H. and Zhou, Y. (2002) Biophys. J. 82, 458–463.
[32] Jackson, S.E., Moracci, M., Masry, N.E., Johnson, C.M. and Fersht, A.R. (1993) Biochemistry 32, 11259–11269.
[33] Li, A. and Daggett, V. (1996) J. Mol. Biol. 257, 412–429.