Importance of Native-State Topology for Determining the Folding Rate of Two-State Proteins†

M. Michael Gromiha*

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

Received February 14, 2003

Understanding the relationship between amino acid sequences and folding rate of proteins is a challenging task similar to protein folding problem. In this work, we have analyzed the relative importance of protein sequence and structure for predicting the protein folding rates in terms of amino acid properties and contact distances, respectively. We found that the parameters derived with protein sequence (physical-chemical, energetic, and conformational properties of amino acid residues) show very weak correlation ($|r| < 0.39$) with folding rates of 28 two-state proteins, indicating that the sequence information alone is not sufficient to understand the folding rates of two-state proteins. However, the maximum positive correlation obtained for the properties, number of medium-range contacts, and α-helical tendency reveals the importance of local interactions to initiate protein folding. On the other hand, a remarkable correlation (r varies from -0.74 to -0.88) has been obtained between structural parameters (contact order, long-range order, and total contact distance) and protein folding rates. Further, we found that the secondary structure content and solvent accessibility play a marginal role in determining the folding rates of two-state proteins. Multiple regression analysis carried out with the combination of three properties, β-strand tendency, enthalpy change, and total contact distance improved the correlation to 0.92 with protein folding rates. The relative importance of existing methods along with multiple-regression model proposed in this work will be discussed. Our results demonstrate that the native-state topology is the major determinant for the folding rates of two-state proteins.

INTRODUCTION

Predicting the three-dimensional structure of a protein from its amino acid sequence is a challenging problem. A related interesting and important task is to understand the relationship between sequences and folding rates of proteins, which is a measure of slow/fast folding of proteins from the unfolded state to the native three-dimensional structure. Recently, three structural parameters have been developed to predict the protein folding rates from the information of inter-residue contacts. These parameters emphasize the importance of topology of the folded state to understand the protein folding rates. 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 proteins. Two-state proteins are the ones, which follow the simplest two-state transition upon folding; i.e., a reaction that proceeds directly from the unfolded state, U, to the folded state, F, (U ⇌ F) without the occurrence of any intermediate states. Gromiha and Selvaraj[2] 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. Recently, these two parameters, CO and LRO, are incorporated into a new parameter, total contact distance (TCD), which shows a good relationship with protein folding rates.[3] Miller et al.[4] experimentally demonstrated that long-range order is one of the best parameters that correlate with protein-refolding rates including circular permutations of ribosomal proteins S6 from *Thermus thermophilus*. Further, it has been reported that CO, LRO, and TCD are equally important for understanding the folding rates of two-state proteins.[5]

The parameters CO, LRO, and TCD were obtained from the three-dimensional structure of proteins. It will be interesting to analyze the influence of amino acid sequence for determining the folding rate of two-state proteins. In this work, we have analyzed the relationship between 49 various amino acid properties and folding rates of two-state proteins. We found that none of the properties show a correlation of more than 0.39, indicating that the amino acid sequence alone is inadequate for explaining the protein folding rates. The data obtained from protein secondary and tertiary structure show a moderate and good correlation, respectively, with the folding rates of two-state proteins. The multiple regression technique performed with the combination of β-strand tendency, enthalpy change, and TCD raised the correlation up to 0.92 with protein folding rates.

MATERIALS AND METHODS

Data Set. We have used a data set of 28 two-state proteins for which the protein folding rates are available.[3,5] The PDB

* Corresponding author phone: +81-3-3599-8046; fax: +81-3-3599-8081; e-mail: michael-gromiha@aist.go.jp.

† Abbreviations: CO, contact order; LRO, long-range order; TCD, total contact distance; NN, neural network; MR, multiple regression; $\ln k_f$, logarithms of folding rate.

10.1021/ci0340308 CCC: $25.00
Published on Web 08/09/2003

codes of the proteins are as follows: 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 were obtained from the Protein Data Bank.[6]

Amino Acid Properties. We used a set of 49 diverse amino acid properties (physical-chemical, energetic, and conformational), which fall into various clusters analyzed by Tommi and Kanehisa[7] in the present study. The amino acid properties are normalized between 0 and 1 using the expression, $P_{norm}(i) = [P(i) - P_{min}]/ [P_{max} - P_{min}]$, where $P(i)$, $P_{norm}(i)$ are, respectively, the original and normalized values of amino acid i for a particular property, and P_{min} and P_{max} are, respectively, the minimum and maximum values. The numerical values of all the 49 properties and the brief descriptions about them are available in our earlier articles.[8−11]

Computational Procedure. The average amino acid property for each protein, $P_{av}(i)$, was computed using the standard formula

$$P_{av}(i) = \sum_{j=1}^{N} P(j)/N \qquad (1)$$

where $P(j)$ is the property value of jth residue and the summation is over N, the total number of residues in a protein. The computed property value $P_{av}(i)$ for the 28 two-state proteins was related with experimental folding rate $\ln k_f(i)$ using single correlation coefficient.

Secondary Structure Content, Solvent Accessibility, and Topological Parameters. To understand the role of secondary structure, we have obtained the information about the percentage of α-helix, β-strand, turn, and coil in each protein with the aid of DSSP.[12] We have analyzed the relationship between the content of each secondary structure (α-helix, β-strand, turn, and coil) and protein folding rates using correlation coefficient.

Further, we have examined the influence of accessible surface areas (ASA) of each protein to understand folding rates. We have calculated the r-value between ASA (total and at different secondary structures) and folding rates of two-state proteins. The program DSSP[12] was used for computation of ASA.

The topological parameters, CO, LRO, and TCD, have been computed using the three-dimensional structures of proteins as explained elsewhere.[1−3]

The complete details about the amino acid sequences of all the 28 proteins, numerical values of 49 amino acid properties, and the data for secondary structural content, ASA, and topological parameters are available at http://www.cbrc.jp/~gromiha/support.html.

Multiple Regression Analysis. We have selected three properties by searching all possible combinations of the 61 variables (49 amino acid properties, 4 secondary structural contents, 5 ASA, total and at different secondary structures and 3 topological parameters) and computed the multiple correlation coefficients. The multiple correlation coefficients and regression equations were determined using standard procedures.[13]

RESULTS AND DISCUSSIONS

Role of Amino Acid Sequence for Understanding Protein Folding Rates. We have computed the average amino acid property for all 28 two-state proteins, as explained in the Methods section. In Table 1, we have presented the correlation coefficients that were obtained by relating each of the amino acid properties and protein folding rates. We observed that none of the properties show a correlation of more than 0.39. We have repeated the calculations using total amino acid property of proteins, and we found that the highest correlation ($|r| = 0.25$) is less than that obtained with average property values, indicating the necessity of normalization with sequence length for understanding protein folding rates.[1,2,14] This result shows that the sequence information alone is not sufficient to understand the folding rate of two-state proteins. However, the properties showing high correlations reveal few implications for protein folding. The average medium-range contact has the highest correlation of 0.39 followed by α-helical tendency ($r = 0.35$). It is interesting to note that the properties reflecting medium-range interactions show a positive correlation to protein folding rates. This might be due to the fact that the short- and medium-range interactions may predominate and initiate protein folding with the formation of α-helices.[15]

Secondary Structure Content and Folding Rate. We have computed the content of α-helix, β-strand, turn, and coil in each protein using DSSP.[12] The relationship between secondary structure content and protein folding rate shows that α-helical content has moderate positive correlation of 0.52 (Table 1). This result also indicates the formation of α-helices prior to β-strands during the process of protein folding, as suggested by Sheinerman and Brooks.[16] The content of coil regions has negative correlation with protein folding rates. Our analysis reveals that the local secondary structure content may also influence the folding rates of two-state proteins, which is consistent with the recent suggestion of Gong et al.[17] that the folding rate of simple, two-state proteins is a function of their local secondary structure content.

Solvent Accessibility and Protein Folding Rate. We have computed the solvent accessibility for the whole protein and that at different secondary structures, helix, strand, turn, and coil. The resultant solvent accessibility is not showing good correlation with protein folding rate, and the highest correlation is 0.43 for the ASA at coil region (Table 1). We have further examined the effect of solvent accessibility by relating the number of residues at different ASA cutoff (0−100% with an interval of 1%) and protein folding rates. We observed no significant correlation at any specific cutoff ASA, which indicate that the number of residues in the interior/exterior of protein structures is not a major determinant for protein folding rates.

Topological Parameters and Folding Rates of Proteins. We have analyzed the relationship between topological parameters (CO, LRO, and TCD) and protein folding rates and found a good correlation between them, as reported earlier.[1−3] The correlation coefficients are −0.74 with CO, −0.81 with LRO, and −0.88 with TCD in a set of 28 two-state proteins. The regression equation used to compute the folding rates ($\ln k_f$) are presented in Table 2. One

Table 1. Correlation Coefficient between Amino Acid Properties, Secondary Structure Content, ASA, and Topological Parameters with Protein Folding Rates[a]

no.	variable	correlation coefficient, *r*
	Amino Acid Properties	
1.	K^0, compressibility	−0.17
2.	H_t, thermodynamic transfer hydrophobicity	0.05
3.	H_p, surrounding hydrophobicity	0.08
4.	P, polarity	0.13
5.	pH_i, isoelectric point	0.07
6.	pK', equilibrium constant with reference to the ionization property of COOH group	0.08
7.	M_w, molecular weight	0.00
8.	B_l, bulkiness	−0.10
9.	R_f, chromatographic index	−0.09
10.	μ, refractive index	0.09
11.	H_{nc}, normalized consensus hydrophobicity	0.02
12.	E_{sm}, short and medium range nonbonded energy	0.05
13.	E_l, long-range nonbonded energy	−0.16
14.	E_t, total nonbonded energy ($E_{sm}+E_l$)	−0.08
15.	P_α, α-helical tendency	0.35
16.	P_β, β-strand tendency	−0.15
17.	P_t, turn tendency	−0.27
18.	P_c, coil tendency	−0.33
19.	C_a, helical contact area	0.09
20.	F, mean rms fluctuational displacement	−0.10
21.	B_r, buriedness	0.08
22.	R_a, solvent accessible reduction ratio	0.20
23.	N_s, average number of surrounding residues	−0.09
24.	α_n, power to be at the N-terminal of α-helix	0.34
25.	α_c, power to be at the C-terminal of α-helix	−0.12
26.	α_m, power to be at the middle of α-helix	0.35
27.	V^0, partial specific volume	0.04
28.	N_m, average medium-range contacts	**0.39**
29.	N_l, average long-range contacts	−0.19
30.	H_{gm}, combined surrounding hydrophobicity (globular and membrane)	−0.05
31.	ASA_D, solvent accessible surface area for denatured protein	0.10
32.	ASA_N, solvent accessible surface area for native protein	0.08
33.	ΔASA, solvent accessible surface area for protein unfolding	0.08
34.	ΔG_h, Gibbs free energy change of hydration for protein unfolding	0.13
35,	G_{hD}, Gibbs free energy change of hydration for denatured protein	0.10
36.	G_{hN}, Gibbs free energy change of hydration for native protein	0.14
37.	ΔH_h, unfolding enthalpy change of hydration	0.14
38.	$-T\Delta S_h$, unfolding entropy change of hydration	−0.10
39.	ΔC_{ph}, unfolding hydration heat capacity change	−0.04
40.	ΔG_c, unfolding Gibbs free energy changes of chain	−0.09
41.	ΔH_c, unfolding enthalpy changes of chain	−0.10
42.	$-T\Delta S_c$, unfolding entropy change of chain	0.09
43.	ΔG, unfolding Gibbs free energy change	0.11
44.	ΔH unfolding enthalpy change	0.33
45.	$-T\Delta S$, unfolding entropy change	0.00
46.	υ, volume (number of non-hydrogen side-chain atoms)	−0.01
47.	s, shape (position of branch point in a side-chain)	−0.04
48.	f, flexibility (number of side-chain dihedral angles)	0.19
49.	$P_{\Phi-\psi}$, backbone dihedral probability	0.02
	Secondary Structure Content	
50.	α-helix	**0.52**
51.	β-strand	0.00
52.	turn	−0.03
53.	coil	−0.52
	Accessible Surface Area (ASA)	
54.	average	0.28
55.	at α-helix	0.32
56.	at β-strand	0.00
57.	at turn	−0.26
58.	at coil	**0.43**
	Topological Parameters	
59.	CO	−0.73
60.	LRO	−0.81
61.	TCD	**−0.88**

[a] The highest correlation coefficient among amino acid properties, secondary structure content, ASA, and topological parameters is shown in bold.

can also use these equations to get the actual values of CO, LRO, and TCD.

Multiple Regression Technique. We have combined a set of three properties and performed multiple regression

Table 2. Calculated Protein Folding Rates by Five Different Methods[a]

PDB	P_β	ΔH	lnk_f (calculated) CO[b]	LRO[b]	TCD[b]	NN	MR[b]	lnk_f expt
1lmb	0.450	0.398	10.01	8.92	9.83	**8.49**	9.55	8.50
2abd	0.409	0.422	7.01	6.16	6.00	**6.45**	5.52	6.55
1imq	0.435	0.424	8.21	7.70	**7.45**	8.12	7.79	7.31
2pdd	0.480	0.373	7.26	9.54	**9.83**		9.06	9.80
1nyf	0.473	0.442	2.76	4.89	3.63	**4.81**	5.19	4.54
1pks	0.442	0.407	3.06	2.23	1.51	**−0.52**	0.56	−1.05
1shg	0.481	0.401	1.81	4.32	1.91	1.60	**1.58**	1.41
1srl	0.489	0.425	1.96	**4.12**	3.23	4.18	4.32	4.04
1fnf_9	0.479	0.387	2.26	1.87	2.57		**1.62**	−0.91
1fnf_10	0.503	0.401	4.16	2.48	4.68	4.79	**5.13**	5.48
1hng	0.479	0.395	2.56	4.07	3.23	3.10	**2.72**	2.89
1ten	0.455	0.402	2.81	**2.23**	3.36		2.67	1.06
1tit	0.457	0.421	2.61	1.46	3.10	2.85	**3.29**	3.47
1wit	0.458	0.413	1.66	−0.63	**0.19**	0.18	−0.28	0.41
1csp	0.450	0.440	3.76	4.27	5.21	**6.95**	6.36	6.98
1mjc	0.457	0.439	3.81	4.43	**4.68**		5.88	5.24
2ait	0.502	0.433	2.06	1.46	0.99	2.04	**2.48**	4.20
1aps	0.481	0.398	1.76	1.36	−0.33		**−1.05**	−1.48
1hdn	0.469	0.407	3.66	**3.20**	1.91		1.59	2.70
1urn	0.487	0.406	4.01	**4.58**	3.89		4.13	5.73
2hqi	0.491	0.414	3.66	1.05	**0.19**		0.49	0.18
1pba	0.454	0.412	4.31	5.29	**5.47**		5.45	6.80
1ubo	0.479	0.374	4.66	6.01	5.61	**7.29**	4.39	7.33
2ptl	0.413	0.422	**3.66**	5.04	3.49		2.82	4.10
1fkb	0.465	0.411	3.41	1.92	2.57	**1.56**	2.42	1.46
1coa	0.507	0.387	**3.71**	4.78	4.68	4.62	4.57	3.87
1div	0.446	0.398	**7.01**	7.75	8.11		7.56	6.58
2vik	0.486	0.421	8.36	3.51	6.93	**6.69**	8.17	6.80

[a] P_β, β-strand tendency; ΔH, enthalpy change; CO, contact order; LRO, long-range order; TCD, total contact distance; NN, neural network; MR, multiple regression; lnk_f: logarithms of folding rate. The calculated lnk_f value, which is close to experimental lnk_f, is shown in bold. [b]The following equations have been used to obtain the calculated lnk_f values with CO, LRO, TCD, and MR. Values with NN was taken from Zhang et al.[5] lnk_f = −0.50 (± 0.07) CO + 19.21 (± 2.03) (r = −0.74) (ref 1). lnk_f = −5.11 (± 1.06) LRO + 12.04 (± 1.78) (r = −0.81) (ref 2). lnk_f = −13.2 (± 1.18) TCD + 19.73 (± 1.43) (r = −0.88) (ref 3). lnk_f = 21.75 (± 2.55) P_β + 45.96 (± 2.99) ΔH − 14.64 (± 1.04) TCD − 7.54 (± 1.19) (r = 0.92) (this work).

analysis for all possible combinations (a total of 35 990 times). We found that the combination of β-strand tendency, enthalpy change, and total contact distance improved the correlation up to 0.92 with protein folding rates. The relationship between experimental and calculated lnk_f are shown in Figure 1. This result demonstrates that the topology of a protein in native state is a major determinant to the folding rate of two-state proteins, as observed in experiments.[18−20]

Comparative Analysis of Different Methods. We have carried out an extensive analysis to compare the existing methods along with the present multiple-regression (MR) model for calculating the folding rates of two state proteins. The results obtained with CO,[1] LRO,[2] TCD,[3] neural network,[5] and MR model are presented in Table 2. From this table, we observed that the best fit is achieved by different methods in all proteins. Zhang et al.[5] devised a neural network method with three descriptors (CO, LRO, and TCD) and 28 optimized parameters (the connection weights between three units in the input layer, seven units in hidden layer, and one unit in the output layer) for predicting the folding rate of two-state proteins. This method has been trained with 25 proteins and tested with three proteins, and the test results have been reported for a random set of 17

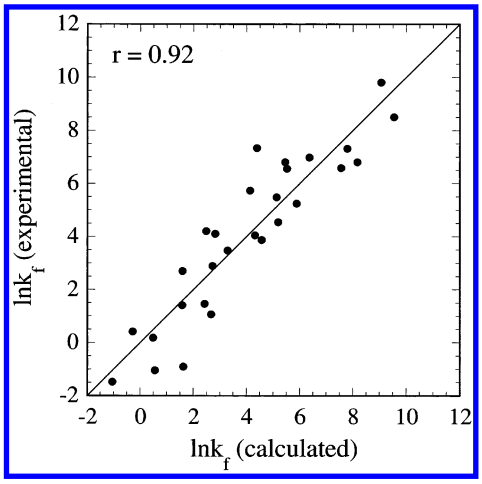

Figure 1. Relationship between experimental and calculated lnk_f values using multiple regression model in 28 two-state proteins.

proteins. However, they have not used any validation data and a certain amount of training may be present in the test data.[21] NN method predicted the folding rate of eight proteins close to experimental value. On the other hand, the methods CO, LRO, TCD, and MR use only a limited number of variables (1−3 descriptors and 2−4 optimized parameters).

The methods proposed with CO, LRO, and TCD contain one descriptor and two optimized parameters (coefficients). The deviation between experimental and computed lnk_f is minimum for three, four, and six proteins, respectively, using CO, LRO, and TCD. The MR model uses three descriptors and four optimized parameters, and it predicts the folding rate of seven proteins better than other methods. Further, the correlation coefficient depends on the number of parameters used in the model; r = −0.74, −0.81, and −0.88, respectively, using CO, LRO, and TCD and 0.92 using MR.

It is noteworthy that the methods TCD, NN, and MR are mainly focusing on the prediction/calculation of folding rates. On the other hand, CO and LRO demonstrate the importance of native-state topology for determining the folding rate of two-state proteins. Further, the weak correlation between amino acid properties and protein folding rates and the moderate correlation between secondary structure content and lnk_f emphasize the role of native-state topology for the folding rate of two-state proteins. In essence, the present study shows that all methods are comparable to each other for understanding the protein folding rates of two-state proteins.

CONCLUSIONS

We have systematically compared the information available in primary, secondary, and tertiary structures of 28 two-state proteins for understanding their folding rates. We found that the amino acid sequence alone is inadequate to explain the folding rates of two-state proteins. The secondary structure content shows a moderate correlation with folding rates. The parameters derived from the interresidue contacts in protein structures (CO, LRO, and TCD) show a very good correlation with protein folding rates, emphasizing the importance of topology in determining the folded state of globular proteins. The multiple regression technique performed with β-strand tendency, enthalpy change, and total contact distance improved the correlation up to 0.92. Our present study reveals that the parameters CO, LRO, and TCD

are equally important to understand the folding rates of two-state proteins.

Note added in proof: During the process of this manuscript, we observed that the classification of proteins into different structural classes shows a good correlation between protein folding rates and amino acid properties (Gromiha, under preparation).

ACKNOWLEDGMENT

The author thanks the referee for constructive comments and wishes to acknowledge Dr. Yutaka Akiyama and Dr. Makiko Suwa for encouragement.

REFERENCES AND NOTES

(1) Plaxco, K. W.; Simons, K. T.; Baker, D. Contact Order, Transition State Placement and the Refolding Rates of Single Domain Proteins. *J. Mol. Biol.* **1998**, *277*, 985−994.

(2) Gromiha, M. M.; Selvaraj, S. 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.* **2001**, *310*, 27−32.

(3) Zhou, H.; Zhou, Y. Folding Rate Prediction Using Total Contact Distance. *Biophys. J.* **2002**, *82*, 458−463.

(4) Miller, E. J.; Fischer, K. F.; Marqusee, S. Experimental Evaluation of Topological Parameters Determining Protein-Folding Rates. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 10359−10363.

(5) Zhang, L.; Li, J.; Jiang, Z.; Zia, A. Folding Rate Prediction on Neural Network Model. *Polymer* **2003**, *44*, 1751−1756.

(6) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000** *28*, 235−242.

(7) Tomii, K.; Kanehisa, M. Analysis of Amino Acid Indices and Mutation Matrices for Sequence Comparison and Structure Prediction of Proteins. *Protein Eng.* **1996**, *9*, 27−36.

(8) Gromiha, M. M.; Oobatake, M.; Kono, H.; Uedaira H.; Sarai, A. Role of Structural and Sequence Information in the Prediction of Protein Stability Changes: Comparison Between Buried and Partially Buried Mutations. *Protein Eng.* **1999**, *12*, 549−555

(9) Gromiha, M. M.; Oobatake, M.; Kono, H.; Uedaira H.; Sarai, A. Importance of Surrounding Residues for Protein Stability of Partially Buried Mutations. *J. Biomol. Str. Dyn.* **2000**, *18*, 281−295.

(10) Gromiha, M. M.; Oobatake, M.; Kono, H.; Uedaira H.; Sarai, A. Importance of Mutant Position in Ramachandran Plot for Predicting Protein Stability of Surface Mutations. *Biopolymers* **2002**, *64*, 210−220.

(11) Gromiha, M. M. Factors Influencing the Thermal Stability of Buried Protein Mutants. *Polymer* **2003**, *44*, 4061−4066.

(12) Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-bonded and Geometrical Features. *Biopolymers* **1983**, *22*, 2577−2637.

(13) Grewal, P. S. *Numerical Methods of Statistical Analysis*; Sterling Publishers: New Delhi, 1987.

(14) Galzitskaya, O. V.; Garbuzynskiy, S. O.; Ivankov, D. N.; Finkelstein, A. V. Chain Length is the Main Determinant of the Folding Rate for Proteins with Three-state Folding Kinetics. *Proteins* **2003**, *51*, 162−166.

(15) Unger, R.; Moult, J. Local Interactions Dominate Folding in a Simple Protein Model. *J. Mol. Biol.* **1996**, *259*, 988−994.

(16) Sheinerman, F. B.; Brooks, C. L. 3rd. Molecular Picture of Folding of a Small alpha/beta Protein. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 1562−1567.

(17) Gong, H.; Isom, D. G.; Srinivasan, R.; Rose, G. D. Local Secondary Structure Content Predicts Folding Rates for Simple, Two-state Proteins. *J. Mol. Biol.* **2003**, *327*, 1149−1154.

(18) Baker, D. A Surprising Simplicity to Protein Folding. *Nature* **2000**, *405*, 39−42.

(19) Makarov, D. E.; Plaxco, K. W. The Topomer Search Model: A Simple, Quantitative Theory of Two-state Protein Folding Kinetics. *Protein Sci.* **2003**, *12*, 17−26.

(20) Jewett, A. I.; Pande, V. S.; Plaxco, K. W. Cooperativity, Smooth Energy Landscapes and the Origins of Topology-dependent Protein Folding Rates. *J. Mol. Biol.* **2003**, *326*, 247−253.

(21) Ahmad, S.; Gromiha, M. M.; Sarai, A. Real Value Prediction of Solvent Accessibility From Amino Acid Sequence. *Proteins* **2003**, *50*, 629−635.

CI0340308