Protein and Peptide Letters, Vol. 10, No. 3, pp. 277-280, 2003
Bentham Science Publishers Ltd.
0929-8665/03 $ 41.00 + .00

A SIMPLE PARAMETER RELATING SEQUENCES WITH FOLDING RATES OF SMALL α HELICAL PROTEINS

Hui Shao, Yi Peng and Zong-Hao Zeng[*]

Centre of Molecular Biology, Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Chaoyang District, Beijing 100101, China.

ABSTRACT: It is found that the helix parameter (HP), which favors clustering of non-polar residues, is linearly correlated with the logarithms of rate constants of folding of small two-state α-helical proteins. The definition is HP = $N_H^{-1} \sum [f_i + (f_{i-1}+f_{i+1})/2]$, where f_i=1 or −1, if the i'th residue is hydrophobic or hydrophilic, respectively, N_H is the number of hydrophobic residues and the summation is taken over the hydrophobic residues.

INTRODUCTION

In recent years, an advance has been made on understanding the relation between protein native structures and folding rates [1-4]. Significant correlations were found to exist between folding rates of a set of two-state folding proteins and the contact order [1], long-range order [2], total contact distance [3], and the number of native contacts [4]. The success of these works is impressive, since in each of the methods folding rates of two dozens, or more, up to 28, of proteins can be predicted from just one parameter determined by the native structures. Some of the parameters, i.e. the long-range order [2] and the number of native contacts [4], are calculated from just the distances between Cα atoms. The involved proteins include α, β and α/β proteins. The folding rate constants range over 4-5 orders of magnitude. The linear correlation coefficients were found in the range from more than 0.8 to less than 0.9.

We call such parameters determined by native structure as structural parameters. Here we present a sequence parameter, determined by the amino acid sequences, which correlates well with the folding rate constants of 5 α proteins. As α proteins are located at the fast folding end with rate constants ranging from 279 s^{-1} to 4900 s^{-1} in all the previously analyzed datasets, the new parameter can be regarded as a supplement to the structural parameters and reveals a different aspect of protein folding.

To define our sequence parameter, i.e. the helix parameter (HP), a number f_i is assigned to the amino acid residue at site i, such that f_i=1, if that residue is hydrophobic (Ala, Val, Leu, Ile, Phe, Tyr, Trp, Cys and Met); or fi=−1, otherwise, with i=1, 2, …, N, and N is the total number of residues in the sequence. Then the parameter HP is defined as

$$HP = N_H^{-1} \sum [f_i + (f_{i-1}+f_{i+1})/2], \qquad (1)$$

where N_H is the number of hydrophobic residues and the summation is taken over all the hydrophobic residues. In addition, f_i=0, when i<0 or i>N.

Table 1. HP and experimental $\ln k_f$ of 5 α-proteins.

Protein ID	Length	HP	$\ln k_f$	Predicted $\ln k_f$
1LMB [5]	80	0.794	8.50	7.73
2ABD [6]	86	0.697	6.55	6.76
1HRC [7, 11]	104	0.859	7.93	8.37
1IMQ [8]	86	0.766	7.31	7.45
2PDD [9, 10]	43	1.000	9.80	9.78

We calculated HP for 5 α proteins. The rate constants of folding are from reference [5-10] and [11]. The calculated HP and the logarithms of rate constants are listed in Table 1 with the protein's ID in Protein Data Bank. The correlation coefficient is 0.927 by linear fitting. As there are only 5 structures included in the fitting, this correlation is not as significant as the fitting by structural parameters. But as far as only α proteins are concerned, the result is better than the fitting by long-range order, where only 4 α proteins were involved and the correlation coefficient is 0.77. The low correlation there may due to the involvement of the yeast cytochrome c (1YCC), which was found to have a complex folding kinetics [12]. Compared to the fitting by total contact distance, which used 4 α proteins, too, similar correlation coefficients are obtained, but one more protein is involved here. Therefore HP is a parameter that correlates folding rates of α proteins with their sequences.

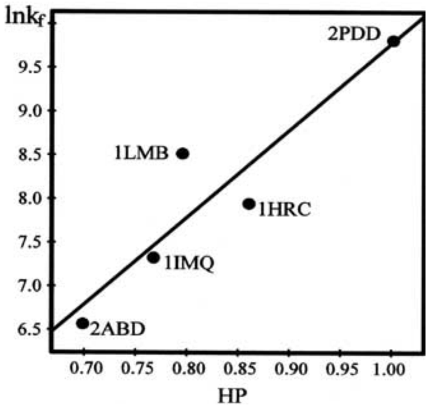

Figure 1. Correlation between the experimental observed $\ln k_f$ and HP.

The summation in Eq. (1) takes over only the non-polar residues. To release the summation from this restriction, a factor, $(1+f_i)/2$, is multiplied to each term in Eq. (1). Then HP can be re-expressed as

$$HP = [N + 2(N_H - N_P) + \Sigma f_i f_{i+1}]/(2N_H), \qquad (2)$$

where N is the total number of residues; N_H and N_P are the number of hydrophobic and polar residues, respectively, and the summation in the third term is taken over all the residues. The meaning of the first and the second terms in Eq. (2) are obvious, they are the total number of residues and the twice of the difference of non-polar with polar residues, respectively. The third term in Eq. (2) is the correlation function between neighboring residues in the sequences. Only local correlation (between direct neighbors) of the amino acid sequence is involved in HP. Obviously from this new expression, one can see that more hydrophobic residues and more concentrated distribution (clustering) of these non-polar residues will increase the value of HP. These characters of the sequences will increase the phase separation tendency of the proteins, and therefore accelerate their folding.

As earlier works had emphasized the importance of clustering of non-polar residues [13-17], e.g. at positions 1-2-4 and 1-4-5, correlation functions involving $f_i f_{i+j}$, with j>1, had been incorporated into HP, but, till now, we did not find the right way to improve the correlation between folding rates and the modified HP. It has also been tested that HP is not suitable for β-proteins. At the present stage, we can only say that the major feature favored by HP is the clustering of non-polar residues. We suggest that HP is a parameter not for prediction of a sequence’s helical preference, but for prediction of relative folding rates of short sequences known to fold into α-proteins. These small α-proteins have simple topologies, i.e. packing of two helices does not strongly prevent stereo-chemically others to incorporate into the packing. The chemical potential of phase segregation, which gathers the non-polar residues by forming helices and packing them together, is the rate limiting force. This sequence parameter is not suitable for β-proteins. One of the reasons for the failure of HP used on β-proteins may due to the lack of long-range correlation in HP. Much work has to be done to get a sequence parameter suitable for predicting folding rates of all classes of proteins.

ACKNOWLEDGEMENTS

Grant sponsor: National Natural Science Foundation of China; Grant number 30080004.

REFERENCES

[1] Plaxco, K.W. Simons, K.T. and Baker, D. (**1998**) *J. Mol. Biol., 277*, 985-994.

[2] Gromiha, M.M. and Selvaraj, S. (**2001**) *J. Mol. Biol., 310*, 27-32.

[3] Zhou, H.Y. and Zhou, Y.Q. (**2002**) *Biophysical. J., 82*, 458-463.

[4] Makarov, D.E. Keller, C. A. Plaxco, K.W. and Metiu, H. (**2002**) *Proc. Natl. Acad. Sci. USA, 99*, 3535-3539.

[5] Burton, R.E. Huang, G. S. Daugherty, M.A. Fullbrigth, P. W. and Oas, T. G. (**1996**) *J. Mol. Biol., 263*, 311-322.

[6] Kragelund, B. B. Hojrup, P. Jensen, M. S. Schjerling, C. K. Juul, E. Knudsen, J. and Poulsen, F. M. (**1996**) *J. Mol. Biol., 256*, 187-200.

[7] Chan, C. K. Hu, Y. Takahashi, S. Rousseau, D. L. Eaton, W.A. and Hofrichter, J. (**1997**) *Proc. Natl. Acad. Sci. USA, 94*, 1779-1784.

[8] Ferguson, N., Capaldi, A. P., James, R., Kleanthous, C. and Radford, S. E. (**1999**) *J. Mol. Biol., 286*, 1597-1608.

[9] Spector, S. Kuhlman, B. Fairman, R. Wong, E. Boice, J.A. and Raleigh, D. P. (**1998**) *J. Mol. Biol., 276*, 479-489.

[10] Spector, S. and Raleigh, D. P. (**1999**) *J. Mol. Biol., 293*, 763-768.

[11] Jackson, S.E. (**1998**) *Folding Des., 3,* R81-R91.

[12] Pierce, M.M. and NaII, B.T. (**2000**) *J. Mol. Biol., 298*, 955-969.

[13] Schifer, M. and Edmundson, A. B. (**1967**) *Biophys. J., 7*, 121-135.

[14] Schifer, M. and Edmundson, A. B. (**1968**) *Biophys. J., 8*, 29-39.

[15] Palau, J. and Puigdomenech, P. (**1974**) *J. Mol. Biol., 88*, 457-469.

[16] Lim, V. I. (**1974**) *J. Mol. Biol., 88*, 857-872.

[17] Lim, V. I. (**1974**) *J. Mol. Biol., 88*, 873-894.

Received on December 11, 2002, accepted on February 18, 2003.