

2 结果与讨论

2.1 氨基酸紧密接触对的相互作用能

从蛋白质数据库(www.rcsb.org)中取得 60 种球形蛋白质结构数据,它们分别是 1BB3,1BVW,1CN4-A,1DAF,1BEF,1BCD,1DE8,1BCZ,1EOR-B,1DZ4-A,1A5C-A,1E6K,1QLH-A, H8D-H, 1ADD, 1AFO-A, 8ADH, 2TS1, 3PGK, 4PFK, 3CHY, 5PTP, 2CNA, 5CPA, 4CLA, 5RXN, 2RHE, 5P21, 2CDV, 2SOD-B, 5FD1, 2SNS, 9PAP, 8TLN-D, 3PGK, 2TAA, 4DFR, 1GOB, 5CPV, 2GCH, 8CAT-A, 4ENL, 2PGD, 2AAA, 1GLA-G, 6XIA, 5ABP, 4TLN, 3LDH, 3APP, 2LIV, 2ER7-E, 4BLM-A, 3CNA, 2SBT, 2DR1, 2ACT, 1BKS,1CA2,1COL. 20 种氨基酸之间可以形成 $\frac{1}{2} \times 20 \times 20(200)$ 种不同的相互作用能.图 1 中仅给出了氨基酸 Ser, Asn, Gln, Asp, Glu, Arg, Lys, Pro 之间的相互作用能.根据相互作用能的大小,把氨基酸 Cys, Met, Phe, Ile, Leu, Val, Trp, Tyr 称为疏水(H)基团,氨基酸 Ala, Gly, Thr, His 为中性基团(N)和氨基酸 Ser, Asn, Gln, Asp, Glu, Arg, Lys, Pro 为亲水(P)基团.在表 1 中给出了 H、N、P 之间的相互作用能(氨基酸之间相互作用能的平均值).同时还给出了不同类型的氨基酸出现的几率,这里的几率是通过 60 种蛋白质统计得到的.发现 H—H 之间的相互作用能最小,只有 $-5.57RT$,而 P—P 之间的相互作用能最大,有 $-1.37RT$.同时 N—N 之间形成的氨基酸紧密接触对出现的几率最小,只有 6.90%.在表 2 中给出了在 60 种蛋白质中不同氨基酸出现的几率.这里发现 Gln, Ala 和 Gly 出现的几率比较大,而 Trp, Met 和 Cys 出现的几率较小.同时还给出了 3 类氨基酸(H、N、P)出现的几率,疏水基团出现的可能性最大(44.49%).

2.2 HNP 格点模型

通过上面的计算,可以得到蛋白质分子的 Hamiltonian 能量为^[4-10]

$$H = - \sum_{i < j} E_{S_i - S_j} \delta(r_{ij}, l)$$

$(S_i, S_j = H, N, P)$

(5)

这里 $E_{H-H} = -5.59RT$, $E_{H-N} = -3.68RT$, $E_{H-P} = -3.07RT$, $E_{N-N} = -2.23RT$, $E_{N-P} = -1.78RT$, $E_{P-P} = -1.37RT$, r_{ij} 是氨基酸与氨基酸之间的距离,假定蛋白质分子的键长为 l .我们采用 HNP

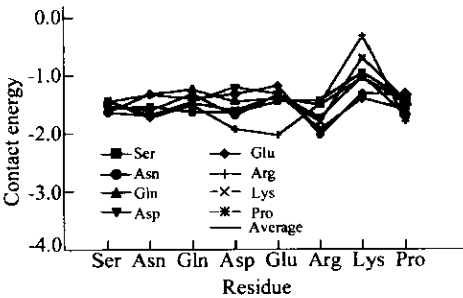


Fig. 1 Contact energy of Hydrophilic-Hydrophilic (P—P) in RT units
The average value is shown by line (—).

Table 1 Contact energy and percentage of contact pairs in protein sample

Contact type *	Contact energy in RT units	Percentage of contact pairs $\xi\eta$
H—H	- 5.57	20.5
H—N	- 3.68	21.2
H—P	- 3.07	27.0
N—N	- 2.23	6.90
N—P	- 1.78	14.2
P—P	- 1.37	10.1

* H:Hydrophobic; N:Neutral; P:Hydrophilic

Table 2 Distributions of amino acid residues in real proteins from the present analyses

Amino acid	Percentage(%)	Amino acid	Total percentage(%)
Cys	1.68	Hydrophobic(H)	30.93
Met	1.55		
Phe	3.60		
Ile	5.61		
Leu	7.40		
Val	6.50		
Trp	1.53	Neutral(N)	24.36
Tyr	3.06		
Ala	8.89		
Gly	8.26		
Thr	4.94	Hydrophilic(H)	44.49
His	2.27		
Ser	6.27		
Gln	8.91		
Asn	4.03		
Glu	5.27		
Asp	6.03		
Arg	3.72		
Lys	6.13		
Pro	4.13		

格点模型用完全算法计算了链长从 $n = 7$ 到 $n = 14$ 的二维蛋白质分子的所有构象.在 HNP 模型中对于链长为 n 的蛋白质分子,有 3^{n-1} 种蛋白质序列.在图 2 中给出了在“自然态”时不同氨基酸类型出现的可能性.发现 H—H, H—N, 和 H—P 形成紧

- 5 Zhang C, Kim S H. Proc Natl Acad Sci USA, 2000, 97: 2550 ~ 2555
- 6 Jiang Z T, Zhang L X, Chen J, Xia A G. Polymer, 2002, 43: 6037 ~ 6047
- 7 Jiang Zhouting(姜舟婷), Chen Jin(陈进), Zhang Linxi(章林溪), Zhao Delu(赵得禄). Acta Polym Sin(高分子学报), 2003, (2): 207 ~ 210
- 8 Chen Jin(陈进), Li Jing(李靖), Zhang Linxi(章林溪), Jiang Zhouting(姜舟婷), Xia Agen(夏阿根), Zhao Delu(赵得禄). Acta Polym. Sinica(高分子学报), 2003, (2): 171 ~ 174
- 9 Chan H S, Dill K A. J Chem Phys, 1993, 99: 2116 ~ 2127
- 10 Dill K A, Bromberg S, Yue S, Fiebig K M, Yee K M, Thomas D P, Chan H S. Protein Sci, 1995, 4: 561 ~ 602
- 11 Zhang Linxi(章林溪), Zhao Delu(赵得禄). Chem J Chinese Univ(高等学校化学学报)[J], 2002, 22(7): 1241 ~ 1243
- 12 Zhang L X, Li J, Jiang Z T, Zhao D L. Polymer, 2003, 44: 1751 ~ 1755

A HNP LATTICE MODEL OF PROTEINS

WANG Xianghong¹, ZHANG Linxi¹, ZHAO Delu²

(¹ Department of Physics, Zhejiang University, Hangzhou 310027)

(² Center of Molecular Science, Institute of Chemistry, Chinese Academy of Sciences, Beijing 100080)

Abstract Effective interresidue contact energies for proteins in solutions are estimated from the number of residue-residue contacts observed in crystal structures of globular proteins (<http://www.rcsb.org/>) by means of the Miyazawa-Jerniganm (MJ) interaction matrix. It is found that residues can be classified into three groups according to the contact energies: hydrophobic residues (H), neutral residues (N), and hydrophilic residues (P). Here we develop a modified HP lattice model, *i. e.* a HNP lattice model. In this model, the contact Hamiltonian is simply written as

$$H = - \sum_{i < j} E_{S_i S_j} \delta(r_{ij}, l) \quad (S_i = \text{H, or N, or P}) \quad (4)$$

Where $E_{\text{H-H}} = -5.59 RT$, $E_{\text{H-N}} = -3.68 RT$, $E_{\text{H-P}} = -3.07 RT$, $E_{\text{N-N}} = -2.23 RT$, $E_{\text{N-P}} = -1.78 RT$, $E_{\text{P-P}} = -1.37 RT$, and r_{ij} is the distance between residues i and j . We first calculate the conformational properties of the native states of short protein chains using 2-dimensional HNP lattice model and enumeration calculation method. The distributions of various types of residue-residue contacts and three types of residues of the native states are discussed. We also investigate the folding of the protein with a sequence of HHNHNPNHPPHPNPPHPHPHPPHNNH using the modified HP lattice model. Here the sequence is selected at random. The folded conformation is obtained in square lattice, and the folding energy is $-64.89 RT$. Our calculation shows that the HNP lattice model may be more reasonable for simulation of the folding behavior of real proteins.

Key words Amino acid, Interaction energy, HNP lattice model, Conformation, Folding