

$(\alpha/\beta)_8$ 桶状蛋白质中紧密接触对的统计性质研究*

陈 进¹ 成 军¹ 姜舟婷¹ 章林溪^{1**} 赵得禄²

(¹浙江大学物理系 杭州 310027) (²中国科学院化学研究所高分子物理及化学国家重点实验室 北京 100080)

摘 要 在计算蛋白质分子中的氨基酸紧密接触对时采用以氨基酸的重心代替 C^α 原子的新方法,通过计算 $(\alpha/\beta)_8$ 桶状蛋白质中的中程和远程紧密接触对的数目,研究了不同氨基酸在形成紧密接触对时所具有的不同能力以及在蛋白质结构稳定性中所起的不同作用.发现在 $(\alpha/\beta)_8$ 桶状蛋白质中氨基酸形成的远程紧密接触对数目与它的 Fauchere-Pliska 疏水性实验数值 (FPH) 存在着非常好的线性关系,而对于氨基酸所形成的中程紧密接触对数目不存在这种关系.同时还研究了 $(\alpha/\beta)_8$ 桶状蛋白质分子大小,发现其回转半径大小与表示远程接触对数目的远程次序值 LRO 存在着关系.表明这一类 $(\alpha/\beta)_8$ 桶状蛋白质具有相似的结构,分子的远程接触对数目越多,分子越紧密,分子尺寸就越小.同时还研究了不同氨基酸之间形成紧密接触对的能力.

关键词 $(\alpha/\beta)_8$ 桶状蛋白质, 中程和远程紧密接触对, 远程次序值, Fauchere-Pliska 疏水性值

蛋白质分子是一类特殊的高分子,其内部作用主要体现为氨基酸之间的紧密接触对相互作用,因此紧密接触对在蛋白质分子的折叠过程和结构稳定性中起着十分重要的作用.人们根据已知的蛋白质结构数据出发,研究了氨基酸紧密接触对的许多统计性质^[1,2].根据蛋白质结构域的不同,可以把蛋白质分成 α 螺旋, β 折叠, $\alpha + \beta$ 和 α/β 这样四类二级结构. α 螺旋是通过反平行或者近乎于互相垂直的状态而连接,排列成层,再平行叠起,从而构成一定的空间构象. β 折叠是指由 β 折叠片层排列而形成一定的三维构象的蛋白质结构. α/β 是指 α 螺旋与 β 折叠交替存在, β 折叠在内, α 螺旋在外卷曲组成一定的蛋白质结构. $\alpha + \beta$ 类型也是由 α 螺旋与 β 折叠混合而成,但不存在象 α/β 类型在空间上的拓扑结构.不同类型的蛋白质由于其结构不同,紧密接触对体现出来的统计性质也不一样. $(\alpha/\beta)_8$ 桶状蛋白质 ($(\alpha/\beta)_8$ barrel protein) 是一类 α/β 蛋白质,它的特点是一组 8 个平行 β 折叠片层在中央,相互连成氢键网,排列成内桶,与各 β 折叠连接的 8 条 α 螺旋则在其间沿着相同的方向排列在 β 桶的外侧,成为外桶.由于结构相似,不同大小的桶状蛋白质分子的紧密接触对会具有相似的性质,因此人们就从结构数据出发,研究了这一类蛋白质的许多紧密接触对的统计性质^[3,4].这些统计性质一定程度上取

决于紧密接触对的定义.习惯上采用比较简单的方法,即如果两个氨基酸的 C^α 原子之间的距离小于 R_c 就定义形成一个紧密接触对^[5-10].但应该看到在构成蛋白质分子的 20 种氨基酸其结构是不一样的,特别是侧基,如 Gly 甘氨酸侧基只有一个氢原子,而精氨酸侧基有 18 个原子,因此有可能发生侧基之间形成紧密接触对这样一种情况.为了克服这个缺点,本文采用氨基酸的重心代替 C^α 原子的新方法,这样可以考虑不同氨基酸具有不同结构的特点,从而来研究 $(\alpha/\beta)_8$ 桶状蛋白质中紧密接触对的统计性质,揭示结构与性质的关系.

1 计算方法

研究了 90 种 $(\alpha/\beta)_8$ 桶状蛋白质分子的紧密接触对统计性质.这 90 种蛋白质分别是 1ADD, 1ADS, 1AGL-O, 1AGL-T, 1ALD, 1AZ1, 1AZ2, 1BFN, 1BTC, 1BTM-A, 1BTM-B, 1BYA, 1BYB, 1BYC, 1BYD, 1CDG, 1CGT, 1CTN, 1DID-A, 1DID-B, 1DIE-A, 1DIE-B, 1DIK, 1DKW-A, 1DKW-B, 1EBH-A, 1EBH-B, 1EF3-A, 1EF3-B, 1ELS, 1FCB-A, 1FCB-B, 1FKW, 1FKX, 1GHR, 1GOX, 1HEI-A, 1HG-A, 1HG-B, 1HH-A, 1HH-B, 1KCL-A, 1KV5-A, 1KV5-B, 1MNS, 1MSS-A, 1MSS-B, 1MUC-A, 1MUC-B, 1NAR, 1NEL, 1ONE-

* 2003-08-08 收稿, 2003-09-27 修稿; 国家自然科学基金 (基金号 29874021, 20174036, 20274040) 和浙江省自然科学基金 (基金号 101002) 资助项目; ** 通讯联系人

A, 1ONE-B, 1PHI, 1TIM-A, 1TIM-B, 1TPD-A, 1TPD-B, 1TPE, 1TPF-A, 1TPF-B, 1TPU-A, 1TPU-B, 1TRI, 1TTI, 1TTJ, 1UIO, 1UIP, 1YPI-A, 1YPI-B, 2ACQ, 2ACR, 2ACS, 2ACU, 2ADA, 2ALR, 2CMD, 3CBH, 3ICD, 4ENL, 4XIA-A, 4XIA-B, 5ENL, 5TIM-A, 5TIM-B, 5XIA-A, 5XIA-B, 6TIM-A, 6TIM-B, 6XIA. 它们的分类是根据 CATH^[1] 方法得到, 同时它们的结构数据是从 Protein Data Bank(PDB, <http://www.rcsb.org/>)数据库获得的.

图 1 给出了紧密接触对的计算方法. 首先决定第 i 个氨基酸的重心 x_c , 然后再计算第 i 个氨基酸与第 j 个氨基酸的重心之间的距离. 如果小于或者等于 R_s , 那么就形成一个紧密接触对. 在蛋白质分子中, 氨基酸侧基之间经常可以形成氢键等的相互作用, 这样使得二个氨基酸侧基之间的距离要小于二个 C_α 原子之间的距离, 有的时候差别还很大 (主要是在包含大侧基的氨基酸之间). 如果仅仅考虑了二个 C_α 原子之间的距离, 那么很多这样的氢键相互作用就不包含在这紧密接触对里面, 导致了统计性质与真实的存在着较大的误差^[12]. 在这里我们取 $R_s = 0.80\text{ nm}$ 和 1.00 nm , 这个与以前计算相似⁵⁻¹⁰. 这里还定义中程紧密接触对和远程紧密接触对. 如果形成紧密接触对的二个氨基酸相距 ± 3 , 或者 ± 4 个氨基酸, 就定义它为中程紧密接触对. 如果相距大于 ± 4 个氨基酸, 则定义它为远程紧密接触对^[3,4].

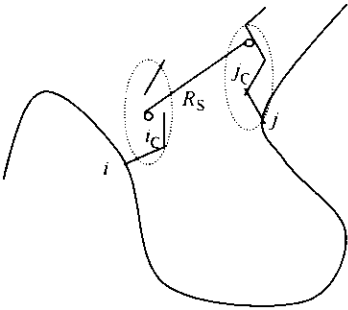


Fig. 1 A contact between residues i and j with the distance between their geometrical centers i_c and j_c less than or equal to R_s

2 结果与讨论

2.1 蛋白质分子形成紧密接触对的平均能力

首先计算了蛋白质分子中每个氨基酸形成紧密接触对的数目, 同时计算了蛋白质分子中能够形成较大数目紧密接触对的氨基酸的比例, 定义 P_L 为

$$P_L = \frac{N_{P_L}}{N} \tag{1}$$

这里 N_{P_L} 为蛋白质分子中能够形成至少有 N_{long} 个紧密接触对的氨基酸数目, N 为蛋白质分子中氨基酸的长度. 图 2 给出了不同蛋白质的不同 P_L 数值与蛋白质中氨基酸形成的平均紧密接触对数目的关系, 这里取 $N_{long} = 6$. 从图 2 可知, 如果蛋白质中氨基酸形成的平均紧密接触对数目越大, 其 P_L 也就越大. 分子越紧密, 蛋白质形成紧密接触对的能力就越强, 蛋白质分子中能够形成较大数目紧密接触对的氨基酸的比例就越高, 更多的氨基酸就在分子内部.

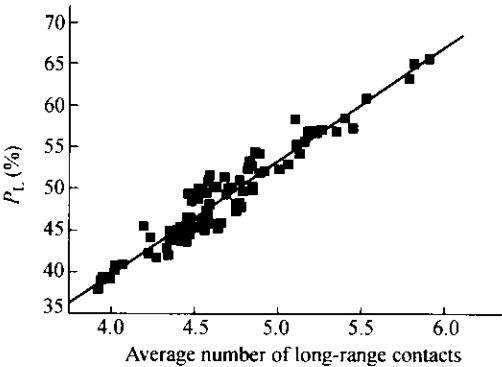


Fig. 2 Percentage of residue with having the number of long-range contacts greater than or equal to N_{long} versus average residue contacts for 90 $(\alpha/\beta)_8$ barrel proteins

Here $N_{long} = 6$

2.2 不同氨基酸形成紧密接触对的能力

由于不同氨基酸的结构不同, 不同氨基酸形成紧密接触对的能力也就不一样. 定义氨基酸形成中程紧密接触对平均数目和远程紧密接触对平均数目 $N_{L,\alpha}$ 为

$$N_{M,\alpha} = \frac{\sum_{\beta = Leu, Val, Ile, \dots, Pro} N_{M,\alpha-\beta}}{N_\alpha}$$

$(\alpha = Leu, Val, Ile, \dots, Pro)$ (2)

$$N_{L,\alpha} = \frac{\sum_{\beta = Leu, Val, Ile, \dots, Pro} N_{L,\alpha-\beta}}{N_\alpha}$$

$(\alpha = Leu, Val, Ile, \dots, Pro)$ (3)

这里 $N_{M,\alpha-\beta}$ 和 $N_{L,\alpha-\beta}$ 分别表示在 90 种 $(\alpha/\beta)_8$ 桶状蛋白质中 α 类氨基酸与 β 类氨基酸之间形成的中程紧密接触对数目和远程紧密接触对数目, N_α 为在 90 种 $(\alpha/\beta)_8$ 桶状蛋白质中 α 类氨基酸的个数. $N_{M,\alpha}$ 和 $N_{L,\alpha}$ 分别表示 α 类氨基酸形成中程紧密接触对和远程紧密接触对的平均能力, 最后结

果见图 3,在这里取 $R_s = 1.00\text{ nm}$,而横坐标是表示氨基酸疏水性实验数值,即 Fauchere-Pliska 疏水性值(FPH)^[13],图中的英文字母分别表记 20 种不同氨基酸.从图 3(a)中可以看出,氨基酸远程紧密接触对平均数目 N_L 与 FPH 值存在非常好的线型关系,即

$$N_L = a + b \times FPH \quad (4)$$

这里 $b(=2.17)$ 为大于零的常数,表明氨基酸的疏水性数值越大,它形成的远程紧密接触对能力越强,而对于中程紧密接触对数目不存在这种关系,(见图 3b),可见氨基酸形成中程紧密接触对能力仅仅与蛋白质中氨基酸的序列有关.从这里可知,氨基酸形成远程紧密接触对平均能力与氨基酸的疏水性密切相关.我们同样计算就氨基酸与氨基酸之间形成紧密接触对的能力大小,表 1 中给出了最容易形成远程紧密接触对的氨基酸对和最不容易形成远程紧密接触对的 10 个氨基酸对,为了比较,同样给出了最容易形成中程紧密接触对的氨基酸对和最不容易形成中程紧密接触对的 10 个氨基酸对.发现最容易形成远程紧密接触对的氨基酸对是缬氨酸-缬氨酸,缬氨酸-丙氨酸和异亮氨酸-异亮氨酸,最不容易形成远程紧密

接触对的氨基酸对是巯基丙氨酸-巯基丙氨酸,巯基丙氨酸-甲硫氨酸和色氨酸-精氨酸,这个与氨基酸的亲水性没有特别大的联系.

Table 1 Topmost and badmost 10 residue-residue contacts in all residue contacts of $(\alpha/\beta)_8$ barrel proteins ($R_s = 1.0\text{ nm}$)

No.	Medium-range contacts *		Long-range contacts *	
	Topmost	Badmost	Topmost	Badmost
1	A-A	C-R	V-V	C-C
2	L-L	C-M	V-A	C-M
3	A-G	C-F	I-I	W-R
4	A-K	M-M	L-V	C-K
5	V-G	M-W	L-L	H-H
6	L-A	C-P	V-I	C-D
7	V-A	W-R	L-C	W-N
8	A-E	C-E	L-W	C-N
9	L-G	W-P	I-A	E-M
10	A-D	C-D	A-A	C-E

* Amino acid residues expressed in one letter symbol

2.3 蛋白质的分子大小与紧密接触对的关系

蛋白质分子中如果氨基酸形成远程紧密接触对的能力越强,分子就越紧密,那么蛋白质的分子尺寸就越小.为了找到这种联系,计算了 90 种 $(\alpha/\beta)_8$ 桶状蛋白质分子的回转半径大小 $\langle S^2 \rangle$,同时还计算了蛋白质分子中表示氨基酸形成远程紧密接触对的平均能力大小的远程次序值(Long-range order, LRO),结果见图 4.这里 LRO 定义为^[2,7]

$$LRO = \frac{\sum n_{ij}}{N} \quad n_{ij} = \begin{cases} 1 & |i-j| \geq 12 \\ 0 & \text{其他情况} \end{cases} \quad (5)$$

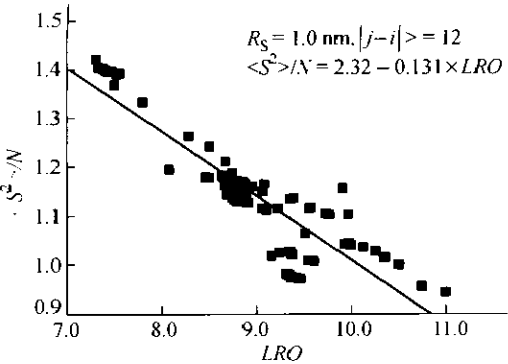


Fig. 4 The ratio $\langle S^2 \rangle / N$ versus long-range order (LRO)
 $\langle S^2 \rangle$ is the mean-square radius of gyration and N is residue length of proteins

在研究蛋白质的折叠速度与氨基酸紧密接触对的关系时,发现满足 $|i-j| \geq 12$ 这个条件的远

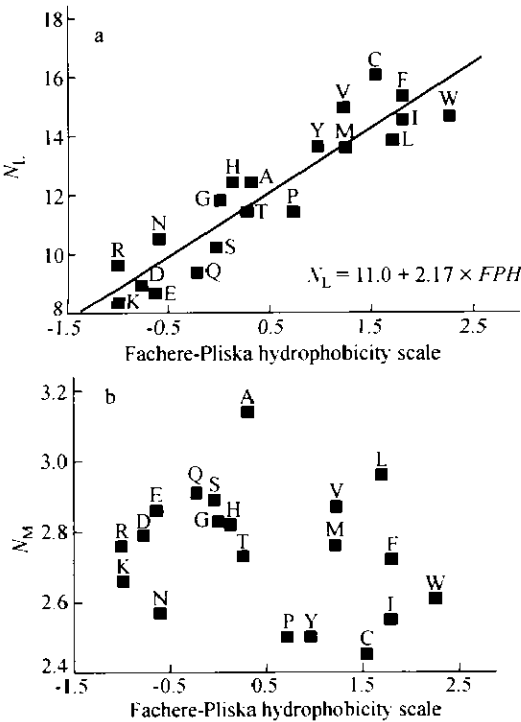


Fig. 3 The relationship between the average number of long- and medium-range contacts per residue and Fauchere-Pliska hydrophobicity scale(FPH) for 20 amino acid residues in 90 $(\alpha/\beta)_8$ barrel proteins
a) Long-range contacts;b) Medium-range contacts, and $R_s = 1.0\text{ nm}$

程紧密接触对对蛋白质的折叠速度影响最大^[2], 因此人们就采用了 $|i - j| \geq 12$ 这个条件来定义远程次序值. 在图 4 中, 发现这 90 种 $(\alpha/\beta)_8$ 桶状蛋白质的分子的回转半径 $\langle S^2 \rangle$ 与 LRO 值存在比较好的线型关系

$$\frac{\langle S^2 \rangle}{N} = a' + b' \times LRO \quad (6)$$

这里 $b' (= -0.131)$ 为小于零的常数, 表明这一类 $(\alpha/\beta)_8$ 桶状蛋白质具有相似的结构. 分子越紧密, 分子的远程接触对数目越多, 分子尺寸就越小. 通过这些研究, 能够帮助了解对于这一类 $(\alpha/\beta)_8$ 桶状蛋白质氨基酸的序列与结构之间的联系. 有关更复杂的联系如热力学性质与稳定性等的研究正在进行之中.

REFERENCES

- 1 Tanaka S, Scheraga H A. Proc Natl Acad Sci USA, 1975, 72: 3802 ~ 3805
- 2 Gromiha M M, Selvaraj S. J Mol Biol, 2001, 310: 27 ~ 32
- 3 Gromiha M M, Selvaraj S. J Biol Phys, 1997, 23: 209 ~ 217
- 4 Selvaraj S, Gromiha M M. J Biol Phys, 1997, 23: 151 ~ 162
- 5 Jiang Z T, Zhang L X, Chen J, Xia A G, Zhao D L. Polymer, 2002, 43: 6037 ~ 6047
- 6 Zhu Z Y, Karlin S. Proc Natl Acad Sci USA, 1996, 93: 8350 ~ 8355
- 7 Zhang L X, Li J, Jiang Z T, Chen J, Zhao D L. Polymer, 2003, 44: 1751 ~ 1756
- 8 Zhang Linxi(章林溪), Jin Jinsheng(金进生), Wang Xianghong(王向红), Ye Gaoxiang(叶高翔), Zhao Delu(赵得禄). Acta Polymerica Sinica(高分子学报), 2002, (4): 497 ~ 501
- 9 Chen Jin(陈进), Li Jing(李靖), Zhang Linxi(章林溪), Jiang Zhouting(姜舟婷), Xia Agen(夏阿根), Ye Gaoxiang(叶高翔), Zhao Delu(赵得禄), Acta Polymerica Sinica(高分子学报), 2003, (2): 171 ~ 174
- 10 Chen J, Zhang L X, Jiang Z T, Xia A G, Zhao D L. Biophys Chem, 2003, 105: 11 ~ 22
- 11 Orengo C A, Pearl F M G, Bray J E, Todd A E, Martin A C, Loconte L, Thornton J M. Nucl Acids Res, 1999, 27: 275 ~ 279
- 12 Jiang Z T, Zhang L X, Chen J, Xia A G, Zhao D L. Polymer, 2004, 45: 609 ~ 621
- 13 Fauchere J L, Ptiska V. Eur J Media Chem, 1983, 18: 369 ~ 375

CONFIGURATION-DEPENDENT PROPERTIES OF CONTACTS IN $(\alpha/\beta)_8$ BARREL PROTEINS

CHEN Jin¹, CHEN Jun¹, JIANG Zhouting¹, ZHANG Linxi¹, ZHAO Delu²

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

(² Polymer Physics Laboratory, Institute of Chemistry, Chinese Academy of Sciences, Beijing 100080)

Abstract The residue-residue contacts contribute more function to the protein folding and play an active role in the stability of proteins. We calculate the number of medium- and long-range contacts for 90 $(\alpha/\beta)_8$ barrel proteins with considering the distance between the centers of two residues in defining residue-residue contacts. The ability to forming contacts in each $(\alpha/\beta)_8$ barrel protein was discussed through calculating the number of long- and medium-ranges contacts for different residues in proteins. The relationship between the average numbers of medium- and long-range contacts with Fauchere-Pliska hydrophobicity value (FPH) for 20 amino acid residues was studied, and it was found that the average number of long-range contacts had a linear relation with the Fauchere-Pliska hydrophobicity value (FPH), however, it had not valid for medium-range contacts. The dimension of $(\alpha/\beta)_8$ barrel proteins has a good relation to the long-range contact of proteins, and the characteristic ratio of mean-square radius of gyration $\langle S^2 \rangle$ can be expressed as $\langle S^2 \rangle / N = a' + b' \times LRO$; where LRO is the long-range order, N is the number of residues, and $a' = 2.32$, $b' = -0.131$. When a protein has a large LRO value, the structure is compact and it has a small dimension value. These investigations can provide some insights into structure of $(\alpha/\beta)_8$ the barrel proteins.

Key words $(\alpha/\beta)_8$ barrel protein, Medium- and long-range contacts, LRO , Fauchere-Pliska hydrophobicity value (FPH)