

类蛋白质分子形成紧密接触对的速率研究*

成军 章林溪**

(温州师范学院物理系 温州 325027)

摘要 采用 Kolinski 等建立类蛋白质分子的格点模型,通过计算类蛋白质分子的末端距分布函数 $P(r)$ 来研究类蛋白质分子形成紧密接触对的速率 k .发现不同的氨基酸序列,其分布函数 $P(r)$ 不同.对于序列 $(H)_x$ 和 $(P)_x$, 分布函数 $P(r)$ 有二个峰值;而对于序列 $(HP)_x$, 分布函数 $P(r)$ 只有一个峰值.对于类蛋白质分子形成紧密接触对的速率 k ,当链长 $N < 11$,随着 N 的增加而增加;当 $N > 11$,形成紧密接触对的速率 k 随着 N 的增加而减少,这个趋势与实验结果一致,并存在关系 $k \sim N^{-\alpha}$ ($N > 11$),系数 α 与氨基酸序列有关.这些研究能够帮助我们加深对蛋白质结构形成的了解.

关键词 类蛋白质分子, 形成紧密接触对的速率, 末端距的分布函数

蛋白质分子从一般的状态折叠到具有一定生物功能特性的三维结构,在这个过程中,分子之间的相互作用起着十分重要的作用,其主要体现为氨基酸之间形成紧密接触对,因此研究氨基酸之间形成紧密接触对的过程对于了解蛋白质的折叠和生物功能特性具有十分重要的意义.最近已经有人用实验的方法测定了蛋白质分子内部形成紧密接触对的速率^[1,2],发现其速率一般随着链长的增加而减少;而对于较短的链,随着链长的增加而增加,但幅度不大.同时 Makarov 等用计算机模拟的方法研究了蛋白质形成紧密接触对的速率问题^[3],但在他们的计算中存在着一个非常大的问题就是没有考虑氨基酸之间的相互作用,而实际上氨基酸之间的相互作用在蛋白质分子的折叠过程和形成氨基酸的紧密接触对过程中起着十分重要的作用.可以这样说,如果没有氨基酸之间的相互作用,就不可能形成蛋白质的紧密结构,因此在研究蛋白质形成紧密接触对时必须考虑氨基酸之间的相互作用.我们已经用完全计数法研究了二维紧密高分子形成紧密接触对的速率问题,但是用完全计数法研究紧密高分子的性质时遇到的最大问题是链长比较短^[4].这里将采用 Kolinski 等建立类蛋白质分子的格点模型,通过计算类蛋白质分子的末端距的分布函数 $P(r)$ 来研究类蛋白质分子形成紧密接触对的速率 k ,并讨论不同的序列对形成紧密接触对的速率 k 的影响.

1 模型和计算方法

在 Kolinski 等建立类蛋白质分子的格点模型中^[5,6],键长是涨落的,其键矢量分别可取 $(3,1)$, $(3,1,0)$, $(3,0,0)$, $(2,2,1)$, $(2,2,0)$, 一共存在 90 种的可能.同时把 20 种氨基酸分成两类,即疏水基团(H)和亲水基团(P),它们之间的相互作用为

$$E_{ij} = \begin{cases} E_{\text{rep}} & r_{ij} < 4 \\ \epsilon_{ij} & 4 \leq r_{ij} < 6 \\ 0 & r_{ij} \geq 6 \end{cases} \quad (1)$$

其中 $|i-j| > 3$, $E_{\text{rep}} = 5k_B T$, 而 ϵ_{ij} 分别取 ϵ_{HH} (H 与 H 之间作用), 或者 ϵ_{PP} (P 与 P 之间作用), 或者 ϵ_{HP} (H 与 P 之间作用), 同时其值分别为 $\epsilon_{HH} = -2k_B T$, $\epsilon_{PP} = -1k_B T$, $\epsilon_{HP} = 0$ ^[5,6].在模拟过程中,类似于键长涨落模型^[7],即随机选择要运动的基团,然后从(100)和(110)构成的 18 个方向中选择其中一个方向.同时采用 Metropolis 方法,即如果 $\Delta E = E_{\text{new}} - E_{\text{old}} < 0$,那么新的构象被接受.如果 $\Delta E > 0$,新的构象被接受的几率为 $\exp(-\Delta E/k_B T)$,不被接受的几率为 $1 - \exp(-\Delta E/k_B T)$,这里同时考虑排斥体积效应.这样经过足够长的时间,体系就能够达到平衡状态.

在图 1 中给出了某一次的模拟过程.发现经过一段时间,体系的能量已经达到平衡了.在模拟中,每次都经过 5×10^5 (Monte Carlo Step, MCS) 以

* 2004-01-05 收稿,2004-02-10 修稿;国家自然科学基金(基金号 20174036,20274040)和浙江省自然科学基金(基金号 R404047)资助项目;

** 通讯联系人

后,每隔 10000 (MCS) 取一个状态,一共取了 100000 个样品.这里 1MCS 表示一条链中每个氨基酸基团都运动一次.

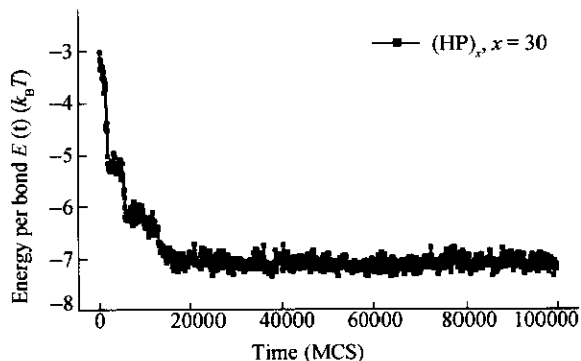


Fig. 1 Average energy per bond ($k_B T$) versus time (MCS) for protein-like chain with the sequence of $(HP)_x$

章林溪在计算蛋白质分子形成紧密接触对的速率 k 时候,采用了 Szabo 建立的理论^[3,4,8]

$$k^{-1} = D^{-1} \int_a^\infty \frac{dx}{p(x)x^2} \left[\int_x^\infty p(y)y^2 dy \right]^2 / \int_a^\infty p(x)x^2 dx \quad (2)$$

这里 D 是扩散常数, $P(x)$ 是末端距为 x 的分布函数, a 表示当二个氨基酸距离小于 a 就形成一个紧密接触对.

2 结果与讨论

2.1 类蛋白质分子的末端距分布密度函数 $P(r)$

由于采用的是格点模型,因此末端距是不连续的,所以类蛋白质分子的末端距分布函数 $P(r)$ 也是不连续的.如果我们取的总样品数为 N_s ,而末端距为 r 的样品数为 N_r ,因而有 $P(r) = N_r / N_s$.但对于较长的类蛋白质分子,末端距可以取的数值还是比较大的,因此这里研究一定区间内的分布函数.定义 $S_{r+\delta/2}$ 为

$$S_{r+\delta/2} = \int_r^{r+\Delta r} P(r) dr = \frac{N_\Delta}{N_s} \quad (3)$$

这里 N_Δ 表示末端距在 $r \sim r + \Delta r$ 这个区间的样品数,而 $S_{r+\delta/2}$ 反映的是在该区间内的样品分数.在图 2 中,给出了相同链长 ($N = 29$)、不同序列时的 $S_{r+\delta/2}$ 值与末端距 r 的关系.对于序列 $(HP)_x$,发现 $S_{r+\delta/2}$ 值随末端距出现大小的交替变化,这是格点模拟中经常出现的‘奇偶效应’,同时发现它只存在一个极值点.而对于序列 $(H)_x$ 和 $(P)_x$, (见图 2, b 和 c) 发现存在两个极值点,这个与我们以

前在回转半径的分布函数研究中发现的结果相似^[9].同时还计算了不同链长的类蛋白质分子的末端距分布函数,都有相似的结果.对于序列为 $(H)_x$ 和 $(P)_x$ 的类蛋白质分子,由于在一条链中氨基酸的性质相同(全部为疏水基团或者亲水基团),它们之间只有 H 与 H(或者 P 与 P)这样一种相互这作用能量,因此它们体现了自避行走链的一部分性质,出现了两个极值点;而对于序列为 $(HP)_x$ 的类蛋白质分子,由于 H 与 H, P 与 P, 和 H 与 P 之间的相互作用能量不一样,因此就可能存在着与 $(H)_x$ 或者 $(P)_x$ 不一样的分布函数 $P(r)$.

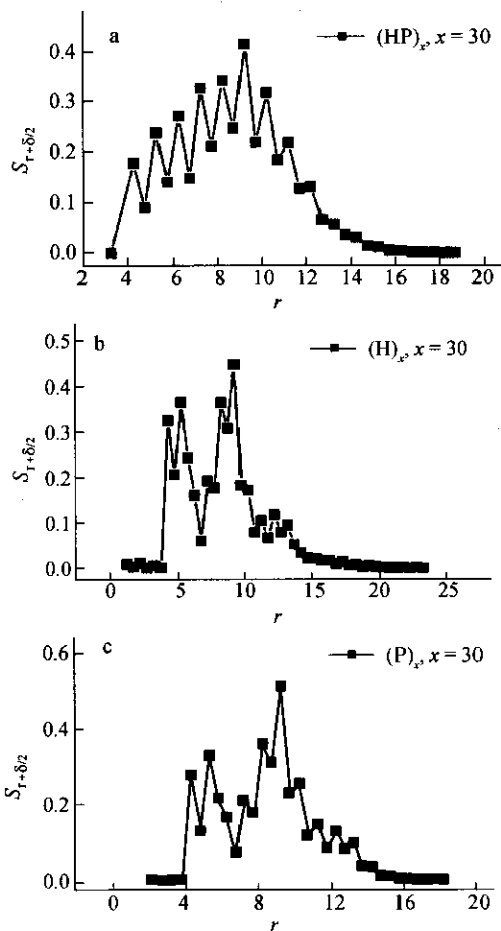


Fig. 2 $S_{r+\delta/2}$ versus end-to-end distance r for three sequences
a) $(HP)_x$ with $x = 15$; b) $(H)_x$ with $x = 30$; c) $(P)_x$ with $x = 30$

2.2 类蛋白质分子形成紧密接触对的速率 k

由于采用的是类蛋白质分子的格点模型,因此等式(2)中的积分可以用求和来代替.在等式(2)中, a 表示形成一个紧密接触对的条件.在具体的计算中,一般都取 $a = 0.50 \sim 0.65$ nm 之间^[10,11],而真实的两个氨基酸之间的 C^α 与 C^α 距

离为 $l = 0.36 \sim 0.39$ nm, 因此二者之比为 $1.28 \sim 1.81$. 在本文的计算中, 类蛋白质分子的链长为 $l = 2.828 \sim 3.316$, 因此 a 可以取 $2.92 \sim 6.00$. 在图 3 中仅给出了 $a = 4.0$ 时在不同链长、不同序列下

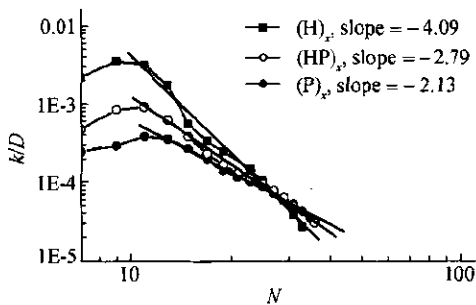


Fig. 3 Ratio of k/D versus chain length N for three sequences

$x = (N - 1)/2$ for $(HP)_x$; $x = (N - 1)$ for $(H)_x$ and $(P)_x$; $a = 4.0$

的类蛋白质分子形成紧密接触对的速率 k 值. 在图中发现, 当 $N < 11$ 时, k 随着 N 的增加而增加; 当 $N > 11$, k 随着 N 的增加而减少. 这个趋势与实验结果非常相似^[1,2]. 同时还存在着关系

$$k \sim DN^{-\alpha} \quad (4)$$

这里假定扩散常数 D 为常数^[3,4]. 指数 α 值分别为 2.13 , 2.79 和 4.09 , 发现这个指数值大于实验值 ($1.36 \sim 1.50$)^[1,2]. 其原因一方面是真实的蛋白质氨基酸序列不可能全部为 $(H)_x$ 基团、或者全部为 $(P)_x$ 基团, 或者全部为 $(HP)_x$ 基团. 另一方面, 这里采用的类蛋白质分子格点模型与真实的蛋白质分子还是有差别的. 比如蛋白质分子一般在水溶液里, 这样疏水基团和亲水基团与水分子的作用完全不一样, 而这里研究的蛋白质是在真空中, 但我们这个方法和计算结果对于了解蛋白质氨基酸之间形成紧密接触对的本质是有帮助的.

REFERENCES

- 1 Lapidus L J, Eaton W A, Hofrichter J. Proc Natl Acad Sci USA, 2000, 97: 7220 ~ 7225
- 2 Bieri O, Wirz J, Hollrang B, Schulkowski M, Dnewello M, Keifhaber T. Proc Natl Acad Sci USA, 1999, 96: 9597 ~ 9601
- 3 Wang Z S, Makarov D E. J Chem Phys, 2002, 117: 4591 ~ 4593
- 4 Chen Jin(陈进), Dong Zhujun(董志君), Zhang Linxi(章林溪), Jiang Zhouting(姜舟婷), Zhao Delu(赵得禄). Acta Polymerica Sinica(高分子学报), 2003, (4): 577 ~ 580
- 5 Kolinski A, Skolnick J. Proteins, 1994, 18: 353 ~ 366
- 6 Kolinski A, Madziar P. Biopolymers, 1997, 42: 537 ~ 548
- 7 Carmesin I, Kremer K. Macromolecules, 1988, 21: 2819 ~ 2822
- 8 Szabo A, Schulten K, Schulten Z. J Chem Phys, 1980, 72: 4350 ~ 4357
- 9 Zhang L X, Xia A G. Polym J, 1996, 28: 548 ~ 550
- 10 Sun T T, Zhang L X, Chen J. Polymer, 2004, 45: 1045 ~ 1053
- 11 Jiang Z T, Zhang L X, Chen J, Xia A G, Zhao D L. Polymer, 2004, 45: 609 ~ 622

A STUDY OF THE RATE OF INTRA-MOLECULAR CONTACT FORMATION IN PROTEIN-LIKE POLYMERS

CHENG Jun, ZHANG Linxi

(Department of Physics, Wenzhou Normal College, Wenzhou 325027)

Abstract The rate of intra-molecular contact formation in protein-like polymers was studied through calculating the probability distribution $P(r)$ of the end-to-end distance of molecules using a (310) lattice model of Kolinski. It is found that the probability distribution $P(r)$ is different for various amino acid sequences. For the sequences of $(H)_x$ and $(P)_x$, the probability distribution $P(r)$ has two peak values, however, for the sequence of $(HP)_x$, it has only one peak value. The rate of intra-molecular contact formation, k , increases with increasing chain length N when $N < 11$. However, the value of k decreases with chain length N when $N > 11$, and there exists a relation as follows: $k \sim N^{-\alpha}$ where α is a parameter depends on the amino acid sequence of molecules. Our results also agree with the experimental ones, and this investigation can help us to understand the structure of globular proteins.

Key words Protein-like polymer, Rate of intra-molecular contact formation, $P(r)$