

·研究简报·

类蛋白质分子扩散系数的模拟计算*

孙婷婷 章林溪** 夏阿根

(浙江大学物理系 杭州 310027)

关键词 类蛋白质分子, 扩散系数, Monte Carlo 模拟

蛋白质分子是一类比较特殊的高分子, 一方面它的性质与氨基酸序列密切相关, 另一方面其二级结构(如 α 螺旋以及 β 折叠等)极大地影响着它的性质. 另外它的内部结构又非常紧密, 与一般的高分子链完全不同. 同时随着计算机技术的发展, 计算机模拟日益作为一种有效的手段广泛应用于高分子科学的研究. 但是由于蛋白质分子比一般的高分子结构更加复杂, 如氨基酸之间有复杂的互相作用等, 因此对蛋白质性质的研究往往建立在简化模型的基础上, 如基于格点的 HP 紧密高分子模型^[1-5]等. 虽然建立在格点模型上的蛋白质分子和真实的蛋白质分子存在着一定的差异, 但基本上能体现蛋白质分子的主要特征, 因此把这一类分子称为类蛋白质分子. 本文采用 Monte Carlo 方法对基于三维格点的 ODI (Orientation-Dependent Monomer-Monomer Interactions) 模型^[6]上的类蛋白质分子进行动力学行为研究, 研究目的是为了了解类蛋白质分子与一般高分子在动力学行为的差异, 帮助我们对蛋白质分子性质的深入了解.

1 计算方法

本文所采用的模型是基于 Zhdanov 和 Kasemo 所提出的 ODI 模型^[6]的改进. 此改进的模型曾被应用于研究类蛋白质分子的构象以及折叠行为^[7,8]. 在此模型中, 同时考虑了蛋白质的二级结构, 包括 α 螺旋及 β 折叠. 类蛋白质分子链的原子位于简立方格点上, 链上相邻的原子是最近邻的, 并且每个格点只能被一个单体所占据. 考虑到蛋白质分子的二级结构, 类蛋白质分子的能量由 4 个部分构成^[6-8], 即

$$E = \sum_{|i-j| \geq 3} \epsilon^{is} \delta(r_{ij} - a) + \sum_{|i-j| \geq 3} \epsilon^{ss} \delta(r_{ij} - a) +$$

$$\sum_{i=1}^{N-1} A(\vec{S}_i \vec{S}_j) \epsilon_{hel} n_{hel} \quad (1)$$

其中 r_{ij} 是单体 i 和 j 之间的距离, a 是格子的间距. (实际上, 在我们的计算中 $a = 1$). 当 $x = 0$, $\delta(x) = 1$; 当 $x \neq 0$, $\delta(x) = 0$. 单位矢量 $\vec{S}_i$ 代表单体的侧基方向, $(\vec{S}_i \vec{S}_j)$ 是方向矢量的标量积. 这里等式 (1) 的第一项代表了紧密接触对的相互作用能量. 第二项中的 ϵ^{ss} 对应于类蛋白质分子的氢键相互作用能. 考虑到氢键形成的必要条件, 认为要存在 ϵ^{ss} 值必须同时满足以下两个条件, (1) 单体 i 和另一个单体 j 满足条件 $\vec{r}_i + \vec{S}_i = \vec{r}_j$, (2) 同时矢量 $\vec{S}_i$ 和 $\vec{S}_j$ 必须是平行的; 其他情况下 $\epsilon^{ss} = 0$. 因此这里的第二项是 β 折叠相互作用能量^[6-8]. 第三项 ($A > 0$) 描述了类蛋白质分子链上相邻单体的相互作用能对方向性的依赖, 其目的是为了产生一个相邻单体反方向的能量. 在第四项中, ϵ_{hel} 表示形成一个 α 螺旋的能量, n_{hel} 表示 α 螺旋的总个数, 因此 $\epsilon_{hel} n_{hel}$ 就表示了类蛋白质分子内的 α 螺旋的总能量. 在模拟中, 我们取 $\epsilon^{is} = -3.0$, $\epsilon^{ss} = -1.5$, $A = 0.5$ 以及 $\epsilon_{hel} = -1.5$ (以 $k_B T$ 为单位)^[7,8].

本文在研究类蛋白质分子的动力学行为时采用的模拟方法是经典的 Monte Carlo 随机取样算法, 即 Metropolis 方法. 首先让类蛋白质分子运动 3×10^5 MCS, 这里 1 MCS 表示每个单体平均都运动一次, 此时类蛋白质分子就处于平衡状态. 从这个时刻开始进行动力学模拟研究, 因此把这个时刻定义为 $t = 0$ 时刻, 然后再让类蛋白质分子运动 2×10^5 MCS. 每隔 1×10^4 MCS 取一构象, 分别计算 X , Y 以及 Z 方向上的质心的平均位移, 扩散系数的定义^[9]为:

* 2005-01-24 收稿, 2005-03-23 修稿; 国家自然科学基金(基金号 20174036, 20274040, 20574052)和浙江省青年人才基金(基金号 R404047)资助项目, ** 通讯联系人, E-mail: Lxzhang@hzcnc.com

$$D = \frac{[<(\Delta x)^2> + <(\Delta y)^2> + <(\Delta z)^2>]}{6\Delta t} \quad (2)$$

这里, $\Delta x = \sum_i \Delta x_i / N = (x_c(t) - x_c(0))$, $\Delta y = \sum_i \Delta y_i / N = (y_c(t) - y_c(0))$, $\Delta z = \sum_i \Delta z_i / N = (z_c(t) - z_c(0))$. Δt 是扩散时间, 这里 Δt 等于 2×10^5 MCS, 同时样品数为 2000.

2 计算结果与讨论

图 1 给出的是某一次模拟过程中单体数为 40 的类蛋白质分子的体系能量随模拟时间的变化关系. 从图中可以看出, 在模拟开始时, 体系的能量降得非常迅速, 接着能量下降的幅度有所减小. 大约经过 5×10^4 MCS, 能量基本上达到了平衡, 数值在 -170 左右浮动. 能量达到平衡以后就可以进行动力学行为的研究. 在本文的动力学行为研究中, 定义运动了 3×10^5 MCS 这点作为起始时刻, 也就是 $t = 0$, 然后再运动 2×10^5 MCS, 从而来讨论这个过程中的动力学特性.

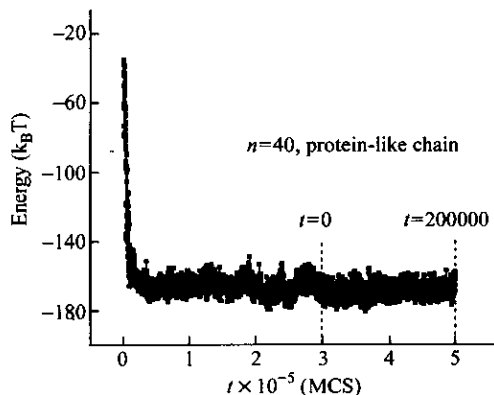


Fig. 1 Total energy as a function of time for 40-monomer protein-like chains

在动力学行为的研究中最关键就是研究扩散系数的变化情况. 在等式(2)中给出了扩散系数的计算方法. 这里定义 $g(t) = <(\Delta x)^2> + <(\Delta y)^2> + <(\Delta z)^2>$. $g(t)$ 称之为均方位移. 可以先求出每一个时刻对应的 $g(t)$ 值, 然后作出均方位移 $g(t)$ 和时间 t 的关系图, 如图 2 所示. 在图 2 中发现均方位移 $g(t)$ 和时间 t 有着很好的线性关系. 根据公式(2)就可以得到扩散系数 D .

本文研究的主要目的是能够了解类蛋白质分子的动力学行为与一般的高分子的区别和联系. 因此在图 2 中同时给出了单体数为 40 的自避链 (SAW chain) 的均方位移 $g(t)$ 与时间 t 的关系. 由

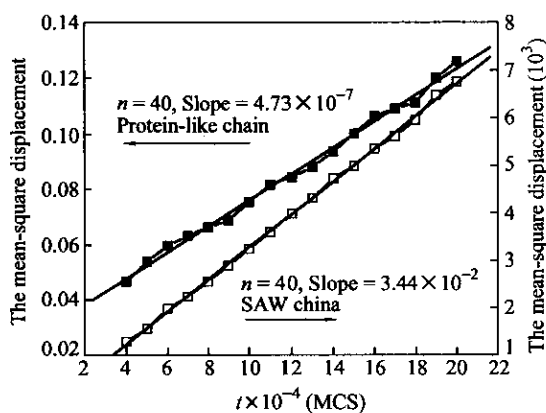


Fig. 2 The mean-square displacement as a function of t for 40-monomer protein-like chains and self-avoiding walk (SAW) chains

图中可以看出, 其均方位移 $g(t)$ 和 t 同样有着非常好的线性关系. 通过对数据的拟合可以得到类蛋白质及 SAW 链的斜率分别为 4.73×10^{-7} 和 3.44×10^{-2} . 由此可计算得到单体数为 40 的类蛋白质链的扩散系数为 7.88×10^{-8} , 而相同长度的 SAW 链的扩散系数则为 5.7×10^{-3} .

分别计算了不同链长的类蛋白质分子的扩散

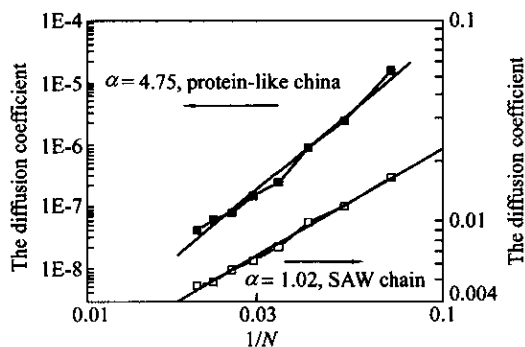


Fig. 3 The diffusion coefficient as a function of $1/N$ (N is the chain length) for protein-like chains and SAW chains

系数. 图 3 给出了类蛋白质分子和自避链的扩散系数与链长的关系. 发现在双对数坐标下, 扩散系数和 $1/N$ 存在着很好的线性关系, 即

$$D \sim N^{-\alpha} \quad (3)$$

通过拟合, 可以得到对于自避链, α 为 1.02; 但对于类蛋白质分子, α 为 4.75, 其值约是自避链的 4~5 倍. 指数的明显不同主要原因在于类蛋白质分子和自避链的内部结构存在着巨大差异. 自避链内部没有任何二级结构, 但类蛋白质分子内部存在着 α 螺旋及 β 折叠等这样一类二级结构, 存在着不同类型的相互作用, 这些就导致了类蛋白质分子的结构比一般的高分子链要紧密得多,

因此就存在着与一般高分子链完全不同的动力学 有着重要的参考意义。
行为.这个结论对今后蛋白质其他特性的研究也

REFERENCES

- 1 Lau K F, Dill K A. *Macromolecules*, 1989, 22; 3986 ~ 3997
- 2 Lau K F, Dill K A. *Proc Natl Acad Sci U. S. A.*, 1990, 87: 638 ~ 642
- 3 Kolinski A, Madzlar P. *Biopolymers*, 1997, 42: 537 ~ 548
- 4 Zhang Linxi(章林溪), Jin Jinsheng(金进生), Wang Xianghong(王向红), Xia Agen(夏阿根), Zhao Delu(赵得禄). *Acta Polymerica Sinica(高分子学报)*, 2002, (4): 515 ~ 519
- 5 Chen Jin(陈进), 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
- 6 Zhdanov V P, Kasemo B. *Proteins: Structure, Function, Genetics*, 1997, 29: 508 ~ 516
- 7 Sun T T, Zhang L X. *Polymer*, 2004, 45: 7759 ~ 7766
- 8 Sun T T, Zhang L X. *Polymer*, 2005, 46: 5714 ~ 5722
- 9 Terzaka I. *Polymer Solutions*. New York: John Wiley & Sons, Inc, 2002. 104 ~ 168

SIMULATION STUDIES ON THE DIFFUSION COEFFICIENT OF PROTEIN-LIKE CHAINS

SUN Tingting, ZHANG Linxi, XIA Agen

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

Abstract Dynamics of protein-like chains are investigated by using Monte Carlo simulation method based on the modified orientation-dependent monomer-monomer interactions model (ODI). The diffusion coefficients for different chain length are calculated, and some comparisons with self-avoiding chains are also made. There exists a relationship between the diffusion coefficient D and chain length N for both protein-like chains and SAW chains, *i. e.* $D \sim N^{-\alpha}$. Here the value of α for SAW chains is 1.02, however, the value of α for protein-like chains is 4.75, which is 4 ~ 5 times larger than that of SAW chains. Therefore, dynamics of protein-like chains is different from SAW chains completely. These investigations may provide some insights into dynamics of proteins.

Key words Protein-like chain, Diffusion coefficient, Monte Carlo simulation