

二维 HP 格点模型中的紧密高分子链构象及其热力学性质的研究*

章林溪¹ 金进生¹ 王向红¹ 叶高翔¹ 赵得禄²

(¹浙江大学物理系 杭州 310028)

(²中国科学院化学研究所分子科学中心高分子物理实验室 北京 100080)

摘 要 采用二维 HP 模型用精确计数法和 Monte Carlo 方法研究了链长为 N (≤ 22) 的紧密高分子链的构象和热力学性质. 发现不同 HP 序列的紧密高分子链的平均自由能和平均配分函数与链长 N 存在关系: $\langle F \rangle = aN + b$, $\ln \langle Z \rangle = a'N + b'$. 同时发现对于可折叠成基态且简并度为 1 的紧密高分子链, 其平均自由能和平均配分函数与链长 N 也存在相似的关系. 在 HP 模型中对于链长为 N 的紧密高分子链, 存在着 2^{N+1} 个不同的 HP 序列. 我们发现可以折叠成基态且简并度为 1 的蛋白质分子的 HP 序列数目 N_s 为 $N_s = a \times 2^{N+1}$ ($a = 0.025$), 对应的 HP 序列中, 疏水基团(H)数目的含量为 40% ~ 60% 的序列出现的几率最大. 同时在这些紧密高分子链中有些具有相同的结构, 发现结构的‘简并度’为 3.3 ~ 4.0 ($10 \leq N \leq 16$). 在紧密高分子链折叠过程中, 折叠的初期能量下降比较快, 折叠的中期能量下降比较缓慢, 折叠的后期能量下降也是比较快.

关键词 紧密高分子链, 二维 HP 模型, 精确计数法, Monte Carlo 方法, 折叠

紧密高分子链(Compact polymer)是特殊的高分子链, 它的特点是在一条高分子链中存在许多紧密接触对(Contact), 体系的能量来源于紧密接触对的相互作用^[1,2]. 对它的研究非常重要, 因为它是蛋白质分子的基态构象. Dill 等首先对紧密高分子链的构象进行了初步研究^[1~3], 但还有许多问题有待于深入研究. 对紧密高分子链的热力学性质的研究, 能加深我们对蛋白质分子稳定性的了解^[4]. 由于生物大分子的构象由氨基酸序列决定, 人们就开始研究不同氨基酸序列时的基态构象^[3,5]. 这种基态构象不仅与大分子在空间的位置有关, 更重要还是与氨基酸序列有关, 因此对它的研究比一般高分子的研究要困难的多, 目前还存在许多问题. 紧密高分子链的折叠又是一个重要的内容, 它能使我们了解氨基酸序列是怎样决定蛋白质分子结构, 预测其三维结构及结构所表现出来的蛋白质分子的性能^[6]. 蛋白质分子不仅要有非常低的能量(基态), 还应具备能量简并度为 1 这个特点, 这样才能保证其结构的稳定. 本文将研究这些简并度为 1 的基态构象, 通过分析紧密高分子链在折叠过程中的能量变化, 来讨论其折叠过程.

1 计算方法

本文采用二维的正方形格点模型. 在这个模型中, 紧密高分子链的键角分别可取 $\pm \pi/2, \pi$, 考虑排斥体积效应, 并排除两个完全相同的构象及链段与链段不能相交. 在考虑相互作用时采用紧密高分子链的 HP 模型, 体系的哈密顿(Hamiltonian)能量为^[6~10]:

$$E = - \sum_{i < j} S_i S_j \delta(r_{ij}, 1) \quad (1)$$

第 i 个基团如是亲水基团(P), 则 $S_i = 0$; 如是疏水基团(H), 则 $S_i = 1$; r_{ij} 表示基团 i 和 j 的距离, 如果 $r_{ij} = 1$, 则 $\delta(r_{ij}, 1) = 1$, 即形成一个紧密接触对(Contact); 如果 $r_{ij} \neq 1$, 则 $\delta(r_{ij}, 1) = 0$. 这里假定形成一个紧密接触对的两个亲水基团之间的作用能为 -1. 对于链长为 N 的紧密高分子链, 存在着 2^{N+1} 个不同的 HP 序列. 对于每一个 HP 序列, 我们计算其配分函数和自由能, 其中配分函数和自由能定义为:

$$Z = \sum_i \exp(-E_i/kT) \quad (2)$$

$$F = -kT \ln Z \quad (3)$$

这里 $\sum_i$ 表示对某个 HP 序列中所有可能的构

* 2001-10-08 收稿, 2001-11-15 修稿; 国家自然科学基金(基金号 29874021 和 20174036)、国家重点基础研究专项经费(批准号 95-预-12, G1999064800)和浙江省自然科学基金(基金号 101002)资助项目

象求和.在讨论基态构象时,我们计算了在给定的 HP 序列中的全部构象,找出简并度为 1 的基态构象并记录其结构,通过分析这些结构,找出相同的.在紧密高分子链的折叠过程中,基团的运动方式非常重要,如果选择不合适,有时很难或不能折叠到基态.在本文计算中,除了采用‘Three-bead flips’和‘Four-bead crankshaft’^[3]运动方式以外,还引入新的运动方式,见图 1.这种方式对于相邻两个键角为 π 的分子运动非常有效,如对于图 1 中的(■)分子,在以前的运动方式是无法运动.同时采用 Metropolis 运动方法.

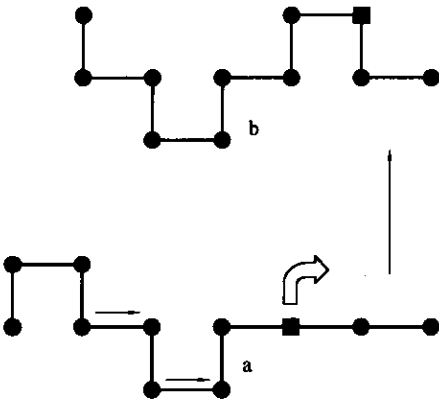


Fig. 1 A type of movement in our calculation

2 结果与讨论

2.1 热力学性质

我们计算了链长从 $N = 4$ 到 $N = 21$ 的紧密高分子链的配分函数和自由能.对于链长为 N 的紧密高分子链(氨基酸数目为 $N + 1$),存在着 2^{N+1} 个不同的 HP 序列,不同的 HP 序列有不同的热力学性质^[4],因此这里讨论平均的配分函数和平均的自由能,最后的结果见图 2 和图 3.在图 2 和图 3 中,对于链长 $N = 4$ 到 $N = 15$ 的紧密高分子链,计算时采用精确计数法.而对于链长 $N = 16$ 到 $N = 21$ 的紧密高分子链,我们采用了 Monte Carlo 方法^[11],随机产生 HP 序列,序列的数目占总的序列的 30% ~ 10%.这里温度分别取 $kT = 0.333$ 、0.50、1.00 和 2.00(以一个紧密接触对的作用能为单位).发现平均的配分函数和平均的自由能与链长 N 存在着以下关系:

$$\langle F \rangle = aN + b \tag{4}$$

$$\ln \langle Z \rangle = a'N + b' \tag{5}$$

系数 a, b 和 a', b' 仅与温度有关.这个关系与一

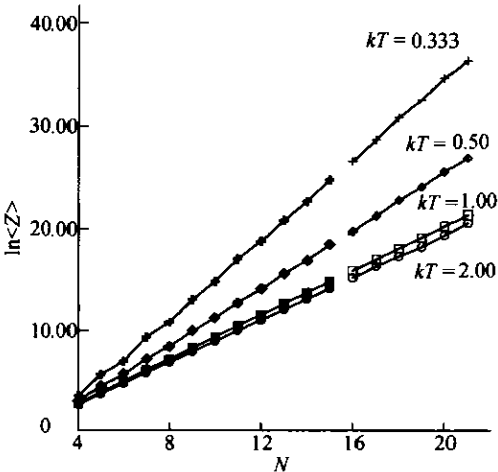


Fig. 2 Average partition function $\langle Z \rangle$ as a function of chain length N at different temperature

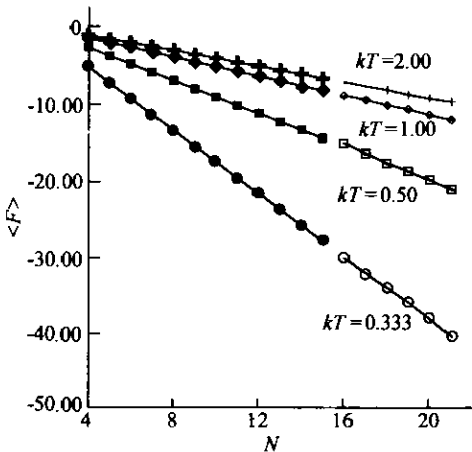


Fig. 3 Average Helmholtz free energy $\langle F \rangle$ as a function of chain length N at different temperature

般的高分子非常相似,在旋转异构态模型中^[12],配分函数 $Z \sim \lambda_1^N$,这说明紧密高分子链也有一般高分子链的特点.

同样计算了基态简并度等于 1 的紧密高分子链的热力学性质,结果见图 4 和图 5,发现其平均配分函数和平均自由能与链长 N 同样存在着与等式 4 和 5 相似的关系,不同的地方在于系数 a, b 和 a', b' 的不同.在较低温度时,二者差异不大;在较高温度时,二者存在一定的差别.

2.2 基态构象

在 HP 模型中,链长为 N 的紧密高分子链,存在着 2^{N+1} 个 HP 序列.不同的 HP 序列有不同的折叠过程,有的 HP 序列可以折叠,有的 HP 序列不能折叠^[1].对于真实的蛋白质分子,由于它的结构是

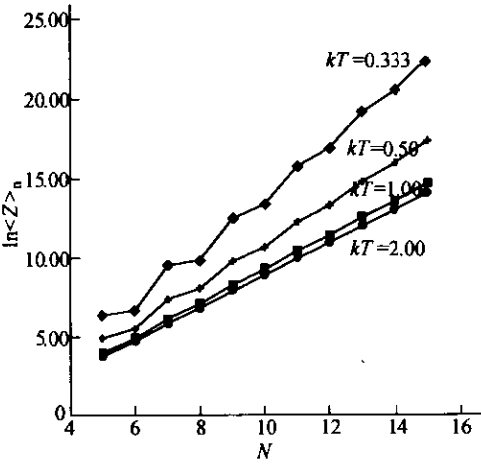


Fig. 4 Average partition function of compact chain with unique ground state as a function of chain length N at different temperature

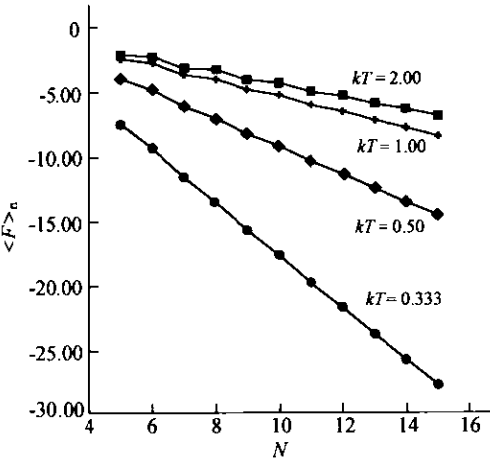


Fig. 5 Average Helmholtz free energy of compact chain with unique ground state as a function of chain length N at different temperature

非常稳定的,因此它的基态简并度必定等于 1.因此研究这种 HP 序列非常重要^[5,6].用精确计数法研究了链长为 $N(\leq 16)$ 的紧密高分子链,发现大约有 1% ~ 3% 的序列可以有基态简并度等于 1 的构象,结果见图 6.对于链长 $N > 16$ 的蛋白质分子,其构象数是非常大的(如 $N = 17$,总的构象数为 1.6×10^{12}),因此采用 Monte Carlo 方法,序列的样品数目为总序列数的 5% ~ 10%,结果也见图 6.发现具有这种性质的 HP 序列数目 N_s 可以表示为

$$N_s = a \times 2^{N+1} \quad (a = 0.025, \quad N > 9) \quad (2)$$

同时此结果小于三维的情况^[4].

同样计算了在紧密高分子链中存在基态且简

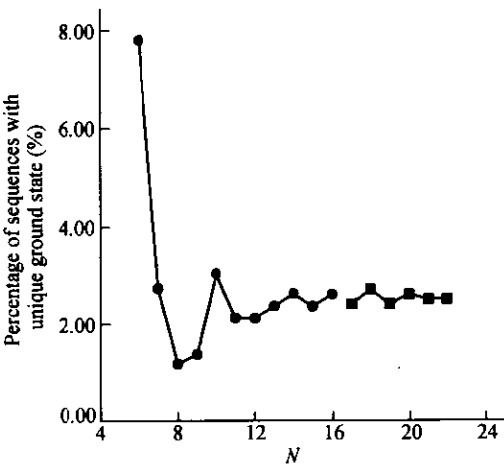


Fig. 6 Percentage of sequences with unique ground state ($N_s/2^{N+1}$) as a function of chain length N
● Enumeration calculation method; ■ Monte Carlo simulation method

并度等于 1 的构象中所含疏水基团(H)数目的情况,结果见图 7.发现在紧密高分子链中,并不是想象中那样,疏水基团(H)数目越多结构就越稳定,而是疏水基团(H)数目的含量为 40% ~ 60% 这种序列出现的可能性最大.这与在蛋白质分子中不可能是同一类(同一类疏水基团)氨基酸排列而成这个特点是一致的.

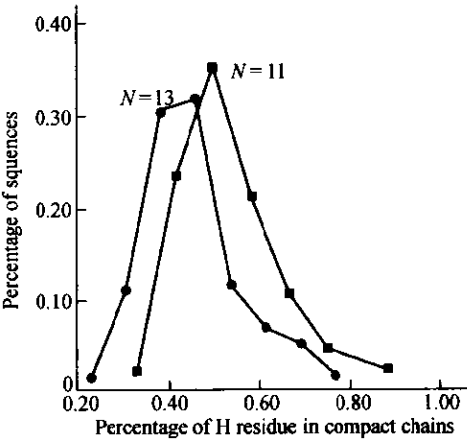


Fig. 7 Percentage of sequences as a function of percentage of H residue in compact chains

对于具有基态简并度为 1 的序列所对应的结构,发现有许多结构是相同的,对这些结构进行分析,得到了结构“简并度” g_c ,即具有基态简并度为 1 的总结构数与不重复的结构数之比,结果在图 8,对于 $10 \leq N \leq 16, g_c = 3.3 \sim 4.0$. 对于 $N < 10, g_c \sim 1.5$.

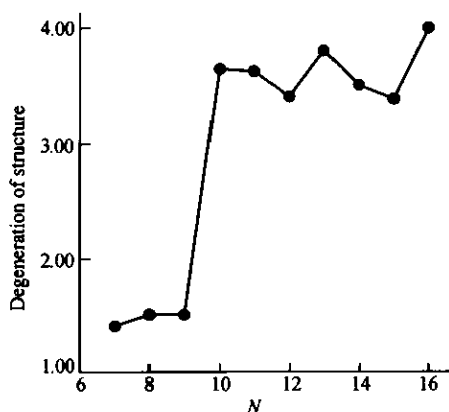


Fig. 8 Degeneration of structure g_s as a function of chain length N

2.3 折叠过程

分析了链长为 49, 序列为 HPHHHHPHPHPH-PPHHHHHPHPPPHHHPHHHHHPHPHHHHHPHHH-PHHH 紧密高分子链的折叠过程. 在图 9 中给出了其中某 2 次的折叠过程(这里假定 $kT = 0.1$), 发现在折叠过程的初期, 能量下降比较快, 运动到 10000 MCS (1 MCS 等于每个分子平均‘运动’一

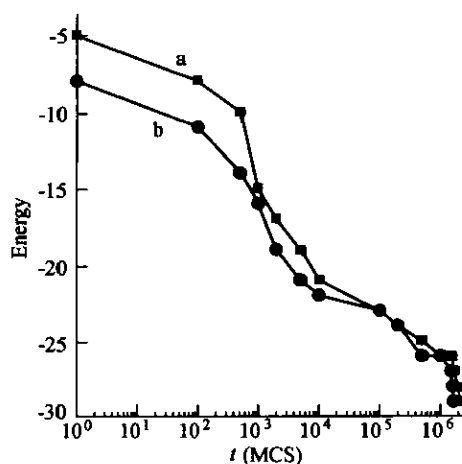


Fig. 9 Energy as a function of time (MCS) in the process of model chain folding for two random runs (a and b)

次), 能量几乎下降到基态能量的 70%. 折叠过程的中期, 能量下降比较缓慢, 几乎用了 2000000 MCS 才下降到基态能量的 90%. 到了折叠过程的后期, 能量下降又比较快. 对于其它几次模拟也发现同样的趋势. 生物体本身折叠是非常快又非常复杂, 人们对它的研究目前也是非常初步的, 有待进一步深入研究.

REFERENCES

- 1 Chan H S, Dill K A. *Macromolecules*, 1989, 22: 4 559 ~ 4 573; 22: 3 988 ~ 3 997
- 2 Chan H S, Dill K A. *J Chem Phys*, 1989, 90: 492 ~ 509
- 3 Chan H S, Dill K A. *J Chem Phys*, 1993, 99: 2 116 ~ 2 127
- 4 Reith D, Huber T, Florian M, Torda A. *J Chem Phys*, 2001, 114: 4 998 ~ 5 002
- 5 Li H, Helling R, Tang C, Wingreen N. *Science*, 1996, 273: 666 ~ 669
- 6 Yue K, Dill K A. *Proc Natl Acad Sci*, 1995, 92: 146 ~ 150
- 7 Irback A, Peterson C, Potthast F, Sandelin E. *J Chem Phys*, 1997, 107: 273 ~ 282
- 8 Yue K, Fiebig K M, Thomas P D, Dill K A. *Proc Natl Acad Sci*, 1995, 92: 325 ~ 329
- 9 Zhang Linxi (章林溪), Zhao Delu (赵得禄). *Chem J Chinese Universities (高等学校化学学报)*, 2001, 22: 1241 ~ 1243
- 10 Zhang Linxi, Zhao Delu. *Macromol Theory Simul*, 2001, 10: 513 ~ 518
- 11 Zhang Linxi, Xia Agen, Xu Jianmin. *Chin J Polym Sci*, 1999, 17: 347 ~ 351
- 12 Flory P J. *Statistical Mechanics of Chain Molecules*. New York: Interscience, 1969. 41 ~ 129

STUDY OF NATIVE CONFORMATIONS AND THERMODYNAMIC PROPERTIES OF TWO-DIMENSIONAL HP COMPACT POLYMERS

ZHANG Linxi¹, JIN Jinsheng¹, WANG Xianghong¹, YE Gaoxiang¹, ZHAO Delu²

(¹ Department of Physics, Zhejiang University, Hangzhou 310028)

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

Abstract The native conformations and thermodynamic properties of compact polymers with chain length N ($N \leq 22$) are investigated by using exact enumeration calculation method and Monte Carlo simulation method employing the HP model. The average Helmholtz free energy $\langle F \rangle$ and average partition function $\langle Z \rangle$ can be expressed in the form of $\langle F \rangle = aN + b$, $\ln \langle Z \rangle = a'N + b'$, where N is chain length of compact chains a, b and a', b' only depend on temperature. We also find that the compact chains with unique ground states have similar relationship in the thermodynamic property and the chain length. In the HP model, the number of sequences of conformations for a protein of chain length N is 2^{N+1} , and the number of sequences with unique ground state may be expressed as N_s : $N_s = a \times 2^{N+1}$ ($a = 0.025$). In these HP sequences, the sequences with 40% ~ 60% in the percentage of H residue have a large probability of emergence. We find that many structures are the same in those unique ground state, and the values of degeneration of structure are 3.3 ~ 4.0 ($10 \leq N \leq 16$). In the processing of folding, the change of energy is fast in the beginning of folding, slow in the middle of folding and also fast in the end of folding of proteins.

Key words Compact polymer, HP model, Enumeration calculation method, Monte Carlo method, Folding