

弹性竿模型下超螺旋 DNA 分子的构象研究*

姜舟婷 章林溪** 陈进 夏阿根 叶高翔

(浙江大学物理系 杭州 310028)

赵得禄

(中国科学院化学研究所分子科学中心 北京 100080)

摘 要 采用弹性竿模型(Elastic rod model),用 Monte Carlo 方法对 DNA 分子的构象进行研究.通过计算发现, DNA 分子的能量是由弯曲势能 E_b 和扭转势能 E_t 两部分组成,通常 E_b 比 E_t 大一至两个数量级.同时给出了均方回转半径与链长之间的关系为 $\langle R_g^2 \rangle = 1169.5 - 3.5 \times n + 0.025 \times n^2$,它体现了 DNA 分子结构的特点.验证了公式 $Lk = Wr + Tw$,得出 Lk 与 Wr 比较接近的结论,考虑 DNA 分子的构型,意味着 DNA 分子容易被弯曲而不易被扭转,但随着连接系数的增加, DNA 被扭转的几率也在增加.这为分析 DNA 分子的结构特征提供了一种新方法.

关键词 弹性竿模型, Monte Carlo 模拟, DNA 分子

DNA 是一种生物高分子,它在传递遗传信息起了主要的作用.通常认为 DNA 分子是右旋的双螺旋结构,两条螺旋链之间的距离很小,只有 1 nm,但长度至少有 10^3 nm,因此 DNA 分子常常看成是线状的或形成封闭的圈.其中后者的空间结构非常复杂,人们把这种结构的 DNA 分子叫做超螺旋 DNA 分子(Supercoiled DNA).大部分的 DNA 分子在它生命的某一阶段都能形成这种超螺旋的结构,因此研究这种结构特性对进一步认识 DNA 分子有很重要的意义.

生物学中的一个重要组成部分就是使用分子模拟的方法来分析生物分子的性质.对 DNA 来说更是如此,几乎所有已知的 DNA 的结构特性都依赖于某些水平上的分子模拟.它对考察 DNA 分子的物理化学性质或分析其实验结果都起着非常重要的作用.超螺旋 DNA 常用双螺旋线所形成的带状结构(Ribbon)在均匀的、各向同性的弹性竿上环绕的模型来研究它的弹性作用和构象性质.在 Kirchhoff^[1]的弹性竿模型(Elastic rod model)中认为通过对竿子的变形和双链所形成“带”以陀螺似的旋转进行精密的计算和分析,是能够模拟出 DNA 分子的三维平衡状态.本文在弹性竿模型的基础上采用 Zhang^[2]等的傅立叶系数展开方法对双螺旋结构的 DNA 的结构特性和热力学性质作进一

步的分析和研究.

1 计算方法

为了使计算机能模拟出闭合环状 DNA 分子的构型,用有限的傅立叶级数来表示光滑闭合没有交点且假定长度为 L 的弹性竿的中轴方程,其表达式如下^[2]:

$$\mathbf{r}(t) = \sum_{n=1}^N \mathbf{a}_n \sin(2\pi n t) + \sum_{n=1}^N \mathbf{b}_n \cos(2\pi n t) \quad (1)$$

t 的取值区间是 $[0, 1]$. 当 t 在这个区间变化,通过给定傅立叶系数 $\mathbf{a}_n$ 和 $\mathbf{b}_n$, ($n = 1, 2, \dots, N$), 方程(1)能给出一条光滑的三维曲线.从标准的圆开始,即取 $\mathbf{a}_1 = (0, r, 0)$, $\mathbf{b}_1 = (r, 0, 0)$, $\mathbf{a}_n = \mathbf{b}_n = 0$, ($n = 2, 3, \dots, N$).任意改变傅立叶系数三维方向中的一维,这样能量可以从 $E(\mathbf{r})$ 变化到 $E(\mathbf{r}')$.如果能量降低,新的构型就被接收,作为下一次模拟的标准能量;若能量升高,我们取一个 0 到 1 之间的随机数与 $\exp\left(\frac{-\Delta E}{k_B T}\right)$ 比较.若随机数小,仍然接收新的构型,反之则不接收.这就是人们常用的 Metropolis Monte Carlo 模拟方法,由此产生一条 MC 轨迹.从理论上说,只要经过足够多的轨迹,最后得到的构型与初始构型的选择无关.

* 2002-04-18 收稿, 2002-05-28 修稿; 国家自然科学基金(基金号 29874021, 20174036, 20274040)、国家重点基础研究专项经费(项目号 95-预-12 和 G1999064800)和浙江省自然科学基金(基金号 101002)资助项目; ** 通讯联系人

根据 Tanaka 的推导^[3,4], 在弹性竿模型中, DNA 分子的能量可表示为两项:

$$E_B(r) = \frac{A}{2} \int_0^L \kappa^2 ds \quad (2)$$

和
$$E_T(r) = \frac{2\pi^2 C}{L} (\Delta Lk - Wr)^2 \quad (3)$$

(2)式可表示为:

$$\int_0^L \kappa^2 ds = \sum_{i=2}^{n+1} \int_0^1 \frac{|\mathbf{r}'(t) \times \mathbf{r}''(t)|^2}{|\mathbf{r}'(t)|^3} dt \quad (4)$$

(3)式也可表示为:

$$Wr = \frac{1}{4\pi} \oint \oint \left[\mathbf{r}'(t_1) \times \mathbf{r}'(t_2) \cdot \frac{\mathbf{r}(t_1) - \mathbf{r}(t_2)}{|\mathbf{r}(t_1) - \mathbf{r}(t_2)|^3} \right] dt_1 dt_2 \quad (5)$$

s 是弧长, κ 表示弹性竿轴的曲率. 系数 A 、 C 是弹性竿的弯曲和扭转常数, 在 NaCl 溶液中, 所有 DNA 的弯曲和扭转常数为^[5]

$$A = 2.70 \times 10^{-28} \text{ J} \cdot \text{m}; C = 2.04 \times 10^{-28} \text{ J} \cdot \text{m} \quad (6)$$

引入连接系数(Linking number) Lk , 它由闭合 DNA 最初的状态所决定的,

$$Lk = Wr + Tw \quad (7)$$

其中 Wr (Writhing number) 体现的是 DNA 形成的封闭带状结构中轴 $\mathbf{r}(s)$ 的性质. Tw (Twist number) 是指带状物扭转的次数. 这三个物理量的具体含义在 Crick 的文章中有详细的说明^[6]. 公式(3)中用到的是连接系数的差值 $\Delta Lk = Lk - Lk_0$, 在 B-DNA 中 $Lk_0 = L/3.4$. (2)式表示弹性竿弯曲的能量, (3)式表示弹性竿扭转的能量, 体系总能量是两者之和.

另外, 考虑到排斥体积效应, 只计算两点间距离大于 2 nm 的情况. 如取 $L = 340 \text{ nm}$ (相当于 1000 基对连成的 DNA 链), t 在 $[0, 1]$ 分成 700 等分, $t_i = i/700$, ($i = 1, 2, \dots, 700$), 我们只取 $\mathbf{r}(t_{1,s_k})$ 和 $\mathbf{r}(t_{5k+26,s_j})$, ($k = 0, 1, \dots, 134; j = 0, 1, \dots, 134 - k$) 这些点作计算.

2 计算结果分析

2.1 体系的能量

图 1 给出了用 Metropolis Monte Carlo^[7] 模拟方法得出体系能量的平衡过程. 对链长为 340 nm 的 DNA 分子的构型做了 5000 步模拟, 得到一条 MC 的能量轨迹. 这里给出的是 200 条这样的轨迹在每一步上取平均值所构成的曲线. 从图上可以清

楚的看到, 随着模拟次数的增加, 能量逐渐减少, 经过 2000 步的模拟后能量基本上固定在某一数值, 仅在有限的范围内涨落. 认为这就是体系的平衡能量. 为了进一步减少误差, 在实际的计算中通过 10000 步的模拟来得到一条 MC 的轨迹, 对 1000 条 MC 轨迹取平均值得出该构型的能量.

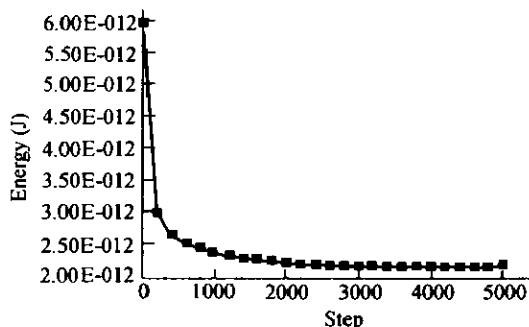


Fig. 1 Energy versus time(step)

计算了链长从 200 基对到 2000 基对 (68 nm 到 680 nm) 的 DNA 分子链的能量, 结果见图 2. 发现随着链长的增加, 能量随之减少. 在较短的 DNA 分子链, 这种趋势更加明显. 为了证明这种趋势, 又计算了 E_B 和 E_T (结果见图 3 和图 4). 在图 3 中, E_B 下降比较明显. 而在图 4 中, 除了特别短的链, E_T 随着链长的增加并没有很明显的下降趋势. 更重要的是 E_B 比 E_T 大一个数量级左右. 因此, E_B 对分子链的能量起主要作用.

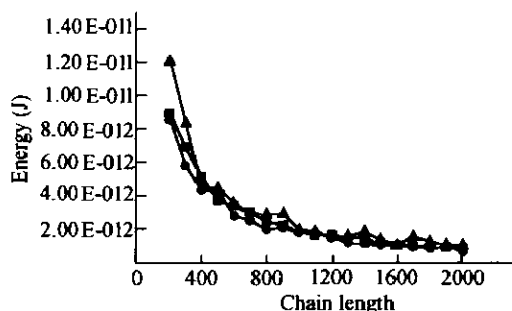


Fig. 2 Total energy versus chain length of DNA chain

●, ■ and ▲ represent the chain of $\Delta Lk = 5, 7, 10$, respectively.

在理想状态下, 公式(2)可改写成

$$E_B(r) = \frac{A}{2} \int_0^L \frac{ds}{R^2} = \frac{A}{2} \int_0^L \frac{d\theta}{R} \propto \frac{1}{L} \quad (8)$$

得到 E_B 与链长 L 成反比的结论, 但在事实上, E_B 与链长的关系极为复杂, 但总的趋势与理想状态是一致的.

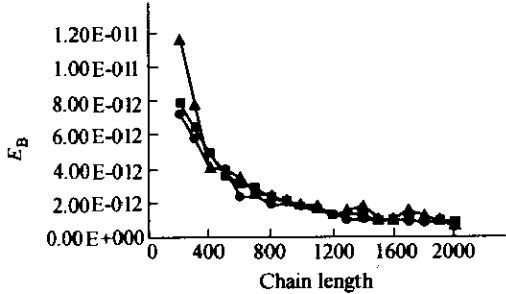


Fig. 3 E_B versus length of DNA chain
●, ■ and ▲ represent the chain of $\Delta Lk = 5, 7, 10$, respectively.

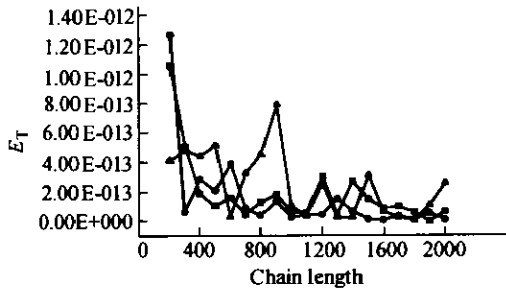


Fig. 4 E_T versus length of DNA chain
●, ■ and ▲ represent the chain of $\Delta Lk = 5, 7, 10$, respectively.

2.2 DNA 分子链的构象

为了考虑 ΔLk 对 DNA 分子链能量和构型的影响,计算了 $\Delta Lk = 5, 7, 10$ 时的能量.考察分子链的总能量,发现随着 ΔLk 的增加,总能量略微有所增加(见图 2),但总体上变化不大,能量的变化范围在 $1.2 \times 10^{-11} \sim 7.8 \times 10^{-13} \text{ J}$ 之间.前面曾提到 Wr 和 Tw 分别是弹性竿中轴的弯曲和扭转的量度.对(3)式中的另一个主要的参量 Wr 进行分析,从图 5 上看到,在链长变化的过程中, Wr 呈现无规的震荡,但相对于不同 ΔLk 的情况, Wr 还是固定在某一值附近.当 ΔLk 分别为 5, 7, 10 时,相对应 Wr 的平均值为 4.98, 6.56, 9.32.这个值约等于 ΔLk , 但比它稍小,它们之间的差值就是体现分子链扭转的参量 Tw .这说明在 DNA 分子中,连接系数 Lk 由 Wr 起决定作用.但随着连接系数的增加, Tw 的值越来越大, DNA 分子链被扭转的趋势就越明显.

2.3 DNA 分子链的大小

计算了 DNA 分子在不同 ΔLk 值时的均方回转半径 $\langle R_g^2 \rangle$, 结果见图 6. 当 ΔLk 分别为 5, 7, 10 时, $\langle R_g^2 \rangle$ 与链长的曲线分别用 ●, ■, ▲ 标识表示

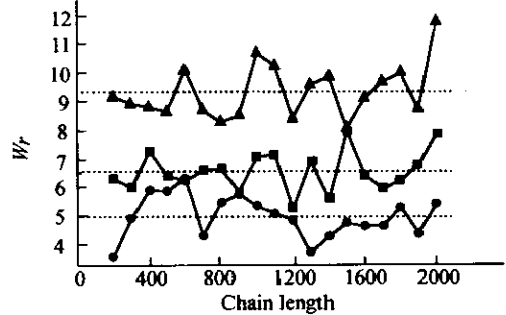


Fig. 5 Wr versus length of DNA chain
●, ■ and ▲ represent chain of $\Delta Lk = 5, 7, 10$, respectively.
..... Represents the average value of Wr .

出来.从图上可以看到,这 3 条曲线并不随着 ΔLk 的改变而改变,尤其是在链长小于 1000 基对(长度小于 340 nm)的时候,其取值范围在 $0.8 \times 10^3 \sim 9.8 \times 10^4 \text{ nm}^2$.对 3 条曲线的平均值作拟合,得到公式为

$$\begin{aligned} \langle R_g^2 \rangle &= 1169.5 - 3.5 \times n + 0.025 \times n^2 \\ &= 1169.5 - 10.4 \times L + 0.218 \times L^2 \quad (9) \end{aligned}$$

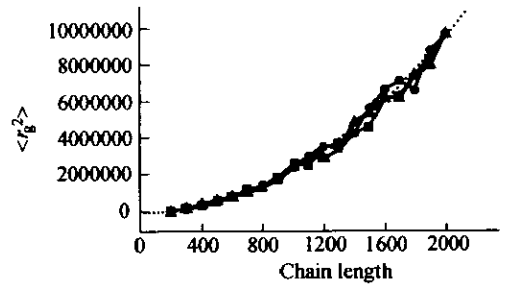


Fig. 6 $\langle R_g^2 \rangle$ versus length of DNA chain
●, ■ and ▲ represent the chain of $\Delta Lk = 5, 7, 10$, respectively.
..... Represents the polynomial fit line.

其中, n 表示这条链所含的键的数目, $L = n \times 0.34$.这个方程与一般的高分子链不同.我们知道在一般的高分子链中 $\langle R_g^2 \rangle \propto n^\alpha$ (考虑排斥体积 $\alpha = 1.15 \sim 1.20$),而在 DNA 分子中, $\langle R_g^2 \rangle$ 与 n 的关系比较复杂,与 n 的最大指数是 2,这主要与 DNA 分子的形状有关.如果把 DNA 分子看作一条长为 L 的棒,那么棒的回转半径为 $L^2/12$.当然 DNA 分子不是一条理想的棒,因此,也不是简单的于链的长度成二次方的关系.

这些结论能进一步加深对 DNA 分子的了解,更加复杂的 DNA 计算机模拟正在研究之中.

REFERENCES

- 1 Kirchhoff G. *J Math (Crelle)*, 1859, 50: 285 ~ 313
- 2 Peisen Z, Wilma K O, Irwin T. *Computational Polymer Science*, 1991, 1: 3 ~ 17
- 3 Fuller F B. *Proc Natl Acad Sci USA*, 1971, 68: 815 ~ 819
- 4 Tanaka F, Takahashi H. *J Chem Phys*, 1985, 83: 6017 ~ 6034
- 5 Hagerman P. *J Ann Rev Biophys Chem*, 1988, 17: 265 ~ 273
- 6 Crick F H C. *Proc Natl Acad Sci USA*, 1976, 73: 2639 ~ 2643
- 7 Zhang Linxi (章林溪), Jin Jinsheng (金进生), Wang Xianghong (王向红), Xia Agen (夏阿根), Ye Gaoxiang (叶高翔). *Acta Polymerica Sinica (高分子学报)*, 2002, (4): 515 ~ 519

A STUDY ON CONFORMATIONS OF DNA CHAIN UNDER THE ELASTIC ROD MODEL

JIANG Zhouting, ZHANG Linxi, CHEN Jin, XIA Agen, YE Gaoxiang

(Department of Physics, Zhejiang University, Hangzhou 310028)

ZHAO Delu

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

Abstract The conformations of DNA chains are studied under the elastic rod model by Metropolis Monte Carlo simulation. The energy of DNA chains includes the bending energy E_b and the twisting energy E_t , and they can be expressed in the form of $E_b(r) = \frac{A}{2} \int_0^L \kappa^2 ds$ and $E_t(r) = \frac{2\pi^2 C}{L} (\Delta Lk - Wr)^2$. The energy decreases with increasing the chain length of DNA. We find that E_b is greater than E_t by $10 \sim 10^2$ times. In the same time, the relationship between $\langle R_g^2 \rangle$ and chain length n is $\langle R_g^2 \rangle = 1169.5 - 3.5 \times n + 0.025 \times n^2$. The equation of $Lk = Wr + Tw$ is verified, and the values of Lk and Wr are found to be very close to each other. It shows that DNA chain is easier to be bent than to be twisted. But for the large value of Lk , the tendency of torsion becomes large. Our simulation can provide a new method to analyze the conformations of DNA chains.

Key words Elastic rod model, Monte Carlo simulation, DNA chain