

核小体纤维结构特性及环境影响因素的 Monte Carlo 研究

目录

第一章 绪言

- | | |
|----------------------|---|
| 1.1 Monte Carlo 方法介绍 | 4 |
| 1.2 DNA 的多级结构 | 5 |

第二章 研究方法及模型的建立

- | | |
|-------------|----|
| 2.1 刚球模型 | 9 |
| 2.2 系统能量的计算 | 10 |
| 2.3 模拟步骤及优化 | 12 |
| 2.4 模拟参数及设定 | 16 |

第三章 模拟结果与讨论

- | | |
|---------------------------|----|
| 3.1 生理溶液浓度下核小体纤维的模拟过程 | 17 |
| 3.2 不同溶液浓度下核小体纤维的终态模拟结果 | 19 |
| 3.3 不同溶液浓度下核小体纤维结构参数的比较分析 | 22 |

第四章 结论

23

Monte Carlo Simulation of Chromatin Fiber Conformation

Affected by Ionic Strength

Contents

Chapter 1 General Introduction

1.1 Monte Carlo method 4

1.2 Multi-level structure of DNA 5

Chapter 2 Modeling and Arithmetic

2.1 Charged hard sphere model 9

2.2 System energy 10

2.3 Arithmetic and optimization 12

2.4 Parameters in simulation 16

Chapter 3 Simulation Results and Analysis

3.1 Fibers in physiological solution 17

3.2 Fibers in solutions with different ionic strengths 19

3.3 Analysis of the structure parameters 22

Chapter 4 Conclusion 23

摘要：

我们通过建立核小体纤维模型，采用 Monte Carlo 方法模拟得出的核小体结构参数能与实验结果较好地符合，例如我们通过得出温度为 300K 的生理溶液浓度下核小体纤维的平均直径为 24.3nm, 与实验值 32nm 接近。而且我们模型还定性地显示了从低浓度到高浓度的过程中，核小体纤维从锯齿状到超螺旋结构的变化过程，这也与一系列地实验结果相吻合。我们还分析了静电能对核小体纤维结构形成的贡献，较好的解释了溶液浓度对核小体纤维状态的影响问题。可见此模型和方法对研究核小体纤维的形成机制有一定的合理性和适用性，可以得出某些定量的结果。

关键词：核小体纤维，超螺旋结构，Monte Carlo 模拟，溶液浓度

Abstract

Using Monte Carlo simulation, we model chromatin fiber and gain the structural parameters which well agree with the experiment results. For example, the average fiber diameter 24.3nm in the physiological ionic solution from our simulation is close to the experimental value 32nm. Furthermore, the simulation predicts the fiber transits from a breadlike state to a super helix state when we increase the ionic strength. We also analyze the contribution of static electric energy to the final state of the fiber. The model we propose here is useful in explaining the conformational change of chromatin fiber in different ionic solutions.

Key words: chromatin fiber,super heix structure, Monte Carlo simulation, ionic strength

第一章 绪言

DNA 是地球上所有生物遗传信息的载体,自 1953 年 Watson 和 Crick [1] 第一次提出 DNA 的双螺旋结构被以来,人们就开始尝试用各种物理模型来解释它的各种力学性质和结构特性,因为这是进一步描述或预言 DNA 行为的基础。在理论计算方面,较多地应用了平均场理论和朗道相变理论,而在计算机模拟方面,则主要有分子动力学方法和 Monte Carlo[3,4]方法。

近年来,DNA 的高级结构引起了研究者的兴趣。实验表明 DNA 链在细胞核中的是以压缩的形式存在的[5]。一条长 1.5 米,直径为 2nm 的 DNA 链大约只占几立方微米的空间。人们提出了各种模型来解释 DNA 链的压缩形式。得到大多数人认同的模型是七十年代 Deson 提出的模型,并为 X 射线衍射实验所证实[6]。该模型认为 DNA 链是缠绕在一种蛋白质核上,该蛋白质核由于 DNA 链的弹性和电荷相互作用以及核间的相互作用容易形成一种超螺旋结构。本文主要运用 Monte Carlo 方法来研究核小体纤维结构特性及环境影响因素(溶液浓度)。以下简要介绍一下 Monte Carlo 方法和 DNA 链的各级结构:

1.1 Monte Carlo 方法在统计物理中

Monte Carlo 方法基于对上式的求和近似:

即仅在相空间中选取 M 个具有代表性的相点 $\{r\}_1, \{r\}_2, \dots, \{r\}_M$ 为统计样本来计算物理量 A 的平均值。Metropolis 算法:

随机选取系统的初始状态，计算该状态的相应的系统能量 H_1 。

随机改变系统的构型，计算新状态的系统能量 H_2 。

若 H_2 小于 H_1 ，则接受新状态；若 H_2 大于 H_1 ，则产生一个 $[0, 1]$ 区间的随机数 r ，若 $r \leq \exp[-(H_2 - H_1)/kT]$ ，则接受新状态，否则放弃仍取原来的状态。

反复上述步骤直至达到所需的精度。

1.2 DNA 的各级结构

(a) 一级结构（信息结构）：DNA 的一级结构是指 DNA 分子中核苷酸的排列顺序，DNA 顺序（或序列）是这一概念的简称。由于核苷酸之间的差异仅仅是碱基的不同，故可称为碱基顺序。包含在组成 DNA 的 A, G, C, T 这四种核苷酸的排列顺序之中[7]。

(b) 二级结构（双螺旋结构）：

DNA 主链是由脱氧核糖和磷酸基通过酯键交替连接而成[8]。主链有二条，它们绕一共同轴心以右手方向盘旋，相互平行而走向相反形成双螺旋构型。主链处于螺旋的外侧，所谓双螺旋就是针对二条主链的形状而言的。碱基位于螺旋的内侧，它们以垂直于螺旋轴的取向通过糖苷键与主链糖基相连。同一平面的碱基在二条主链间形成碱基对。碱基对具有二次旋转对称性的特征，即碱基旋转 180° 并不影响双螺旋的对称性。DNA 的螺旋直径 2nm；螺旋周期包含 10 对碱基；相邻碱基对平面的间距 0.34nm。

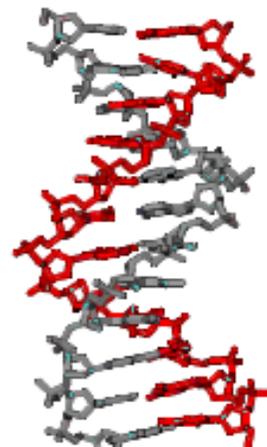
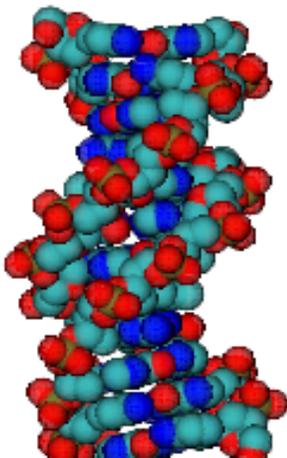


图 1.1 DNA 的双链结构[9]

(c) 三级结构(即染色体纤维结构)

染色质纤维之基本结构是由核小体串连而成[10]。它通常含有 200 个碱基对的脱氧核糖核酸 (DNA) 和 9 个组蛋白分子。由核小体核心 (nucleosome core) 和一条含有 H_1 组蛋白的连接区 DNA (Linker DNA) 所组成。核小体核心由四种组蛋白 (H_2A 、 H_2B 、 H_3 、 H_4 各 2 个分子) 组成, 上边缠绕了 1.75 圈的 DNA, 约有 177 个碱基对。核小体核心结合上 H_1 组蛋白后, 165 个碱基对的 DNA 分子在核小体上缠绕 2 圈。 H_1 组蛋白和 DNA 双链的进出口端相结合。连接区 DNA 连接在染色质小体之间, 长度随不同细胞而异, 通常为 35 个碱基对。核小体核心直径约 70 埃, 高约 55 埃, 呈扁圆柱体形状。串接在一起的核小体在一定的溶液环境下, 由于弹性能及静电能又会形成螺旋状结构, 这种结构称为超螺旋结构[5, 6, 9]。见下图:

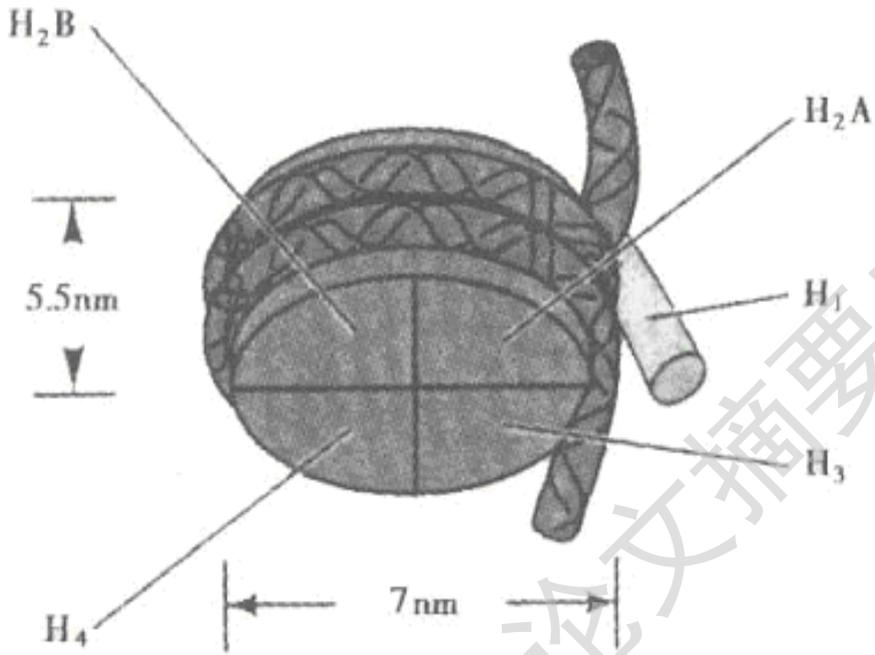


图 1.2 核小体模型

DNA 更高级的压缩形式见下图。

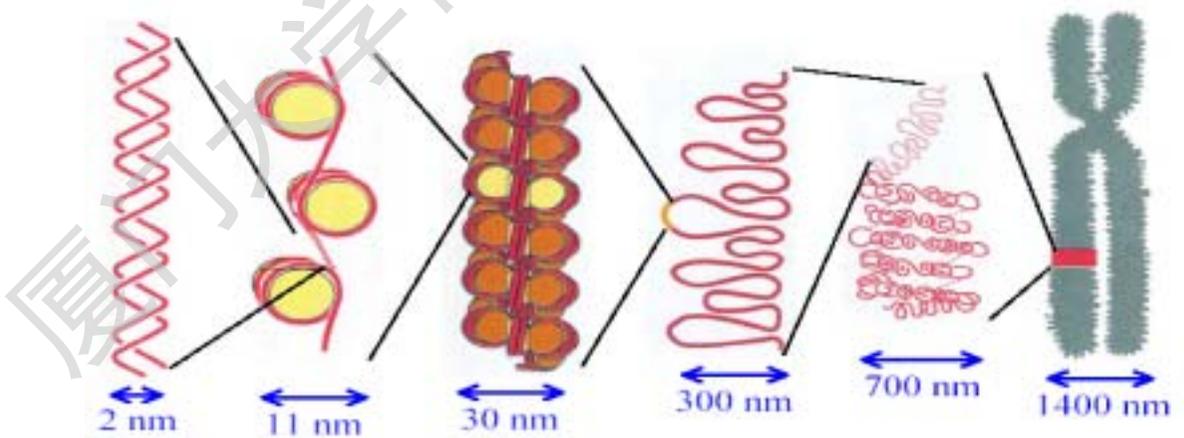


图 1.3 DNA 的各级结构

本文主要研究核小体纤维结构。下图是在生理溶液中观测到的染色质纤维。但是这种纤维的形成机制还没有完全明确，如它的结构参数（diameter, tilt angle, persistent length, linear density...）与 DNA 链的弹性及电荷，蛋白质核间相互作用有何联系。此外，实验表明这种被 DNA 链缠绕的蛋白质核并不总排列成超螺旋结构。如在低浓度的溶液中它是呈简单的锯齿状排列的，在生理溶液浓度中会形成直径 30nm 的超螺旋结构[11]，更高的浓度会导致这种超螺旋结构一系列参数的变化，如直径变小等。对于这种超螺旋结构是如何形成的，其中那种形式的能量占了主导地位以及溶液浓度是如何影响其形态的等等问题目前还没一个明确的了解。因此，定量地研究核小体纤维结构特性即环境影响因素是极具意义和挑战性的。

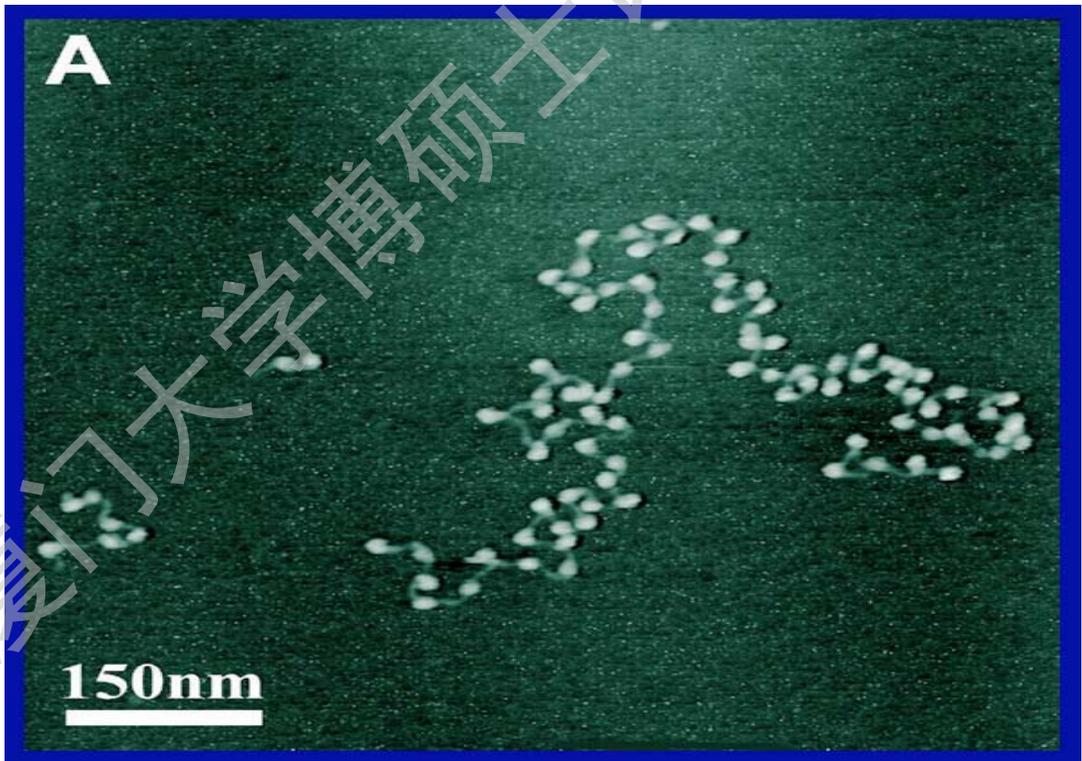


图 1.4 实验中观测到的核小体纤维

第二章 研究方法及模型的建立

鉴于系统的复杂性,我们建立了一个简化的力学模型,再用 Monte Carlo 方法进行模拟,模拟中所用的参数均来自实验或 Poisson 方程数值求解的结果。

2.1 均匀带电的刚球模型

和实体相对应,我们的模型分为两部分,即蛋白质核和 Linker DNA。

蛋白质核可近似为一个刚性球,其半径大约为 6nm。在球体的一侧是没有任何外力作用下 LD 的出口方向,即自然出口方向。图中用箭头表示。自然出口方向和球体的位置关系可用一下参量来描述:

- 1 - 出口距离球心的距离
- 2 - 两个出口方向间的夹角
- 3 - 两个出口方向的轴向距离

模拟中以上三个参数的取值均来自实验数据。

由于 Linker DNA (LD)的长度只有 11bp,所以可以近似为一段均匀弹性绳。

蛋白质核和 LD 的位置关系可用 L , α , β 表示。其中 L 为相邻蛋白质核间 LD 的自然长度, α 为接入端 LD 与其自然方向的夹角, β 为相邻蛋白质核间 LD 的扭角:

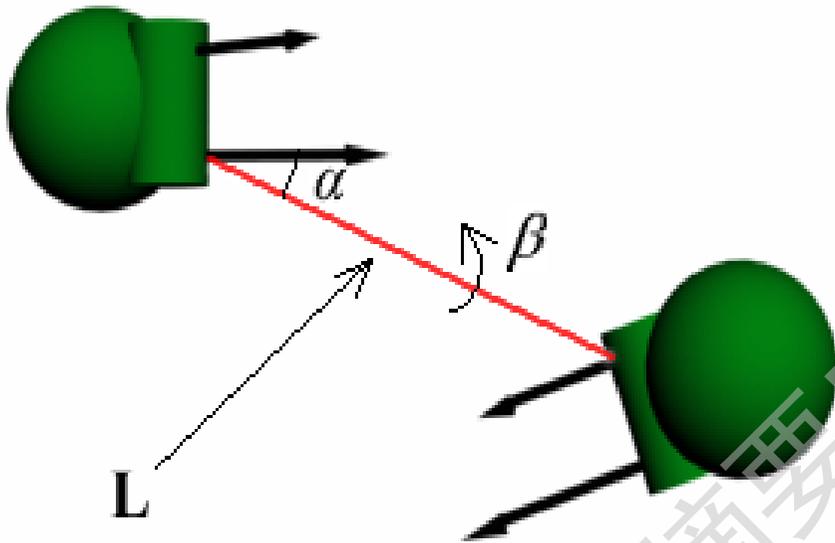


图 2.1 核小体的刚球模型示意

2.2 系统能量的计算

放置在溶液中的核小体纤维，系统能量可以分为三部分：弹性能，静电能，和水分子的极化能。由于水分子的极化能远小于弹性能和静电能，因此在我们的模型中忽略不计，即系统的哈密顿量可以写为：

$$H = H_{ela} + H_{ele}$$

2.2.1 弹性能的计算

弹性能主要由系统各部分的相对位置所决定，因此它可以由系统的位置参数描述。由于 Linker DNA 可以看成是均匀弹性绳，因此它的弹性能可以写为：

$$E = K_0(l - l_0)^2 + K_2 \times (\beta - \beta_0)^2 + K_1 \times (\alpha_1^2 + \alpha_2^2)$$

各个参数的含义为：

l - Linker DNA 的长度 l_0 - Linker DNA 的自然长度

β - Linker DNA 的扭角 β_0 - Linker DNA 的自然扭角

α_i - Linker DNA 与其自然出口方向的夹角

K_i 为相应的弹性系数。对相邻两个核小体，设其位置为 \vec{R}_1, \vec{R}_2 ，对应的 Linker DNA 的出口位置为 \vec{r}_1, \vec{r}_2 ，Linker DNA 的自然出口方向为 \vec{n}_1, \vec{n}_2 ，核小体的主轴为 \vec{z}_1, \vec{z}_2 ，则 l, α_i, β 可以由一下公式计算：

设 $\vec{l} = \vec{r}_2 - \vec{r}_1$ ，则有： $l = |\vec{l}|$ ，

记 $\vec{n} = \frac{\vec{l}}{|\vec{l}|}$ ，经过简单的矢量运算可得

$$\alpha_1 = \arccos(\vec{n} \cdot \vec{n}_1) \quad , \quad \alpha_2 = \arccos(-\vec{n} \cdot \vec{n}_2)$$

$$\beta = \arccos(\vec{z}_1 \times \vec{z}_2)$$

2.2.2 静电能 H_{ele} 的计算

系统的静电能可以写为：

$$E_{ele} = E_{dc} + E_{dd} + E_{cc}$$

其中 E_{dd} 为 LD 与 LD 之间的静电能， E_{dc} 为 LD 与蛋白质核之间的静电能， E_{cc} 为蛋白质核之间的静电能。所以我们只考虑蛋白质核间相互作用。

根据实验数据[13-15]，一个蛋白质核的净电荷为 164e, B-Form DNA 的电荷线密度为 $(-2e)/bp$, LD 的长度为 11bp, 缠绕在蛋白质核上的 DNA 长度为 177bp. 蛋白质核与核上 DNA 的净电量为： $(-2) \times 177 + 164 = -190e$

LD 的电量为：

$$-2 \times 11 = -22e$$

所以粗略估计相同的平均距离下 $E_{dc} : E_{dd} : E_{cc} \approx 1:10:100$ ，因此我们只考虑蛋白质核间的相互作用，即假设 $E_{ele} = E_{cc}$ 。在我们的模型中蛋白质核可以看成均匀带电刚性球壳，我们可以采用连续介质模型[19]求出它在溶液中的电势。模型的关键是求解溶液中的 Poisson-Boltzmann 方程：

$$\nabla^2 \psi(\vec{r}) = -4\pi\rho(\vec{r})/\epsilon$$

Debye-Huckel 理论核心是运用 Boltzmann 分布率计算溶液中各点离子浓度差异[18]，即：

$$\rho(\vec{r}) = \sum_i (Z_i e) n_i^0 \exp(-W_i(\vec{r})/kT)$$

其中 $W_i(\vec{r}) = (Z_i e)\psi(\vec{r})$

在较低溶液浓度下，代入方程得：

$$V(\vec{r}) = Z^2 e^2 [\epsilon(1 + \frac{k\sigma}{2})^2]^{-1} \frac{e^{-r(k-\sigma)}}{r}$$

$$\text{其中 } \kappa = (8\pi n e^2 / \epsilon k_B T)^{1/2}, \quad n = 1/2 \sum_i n_i^0 Z_i^2$$

对所有的核小体求和则可得到整个核小体纤维的静电 H_{ele} [18]。2.3 模拟步

骤及优化：

(a) 本模型中运用 Monte Carlo 方法进行模拟的步骤为：

1. 赋予核小体初始排布。
2. 计算当前位置状态下系统的哈密顿量 H_1
3. 分别让每个核小体沿随机方向移动一个步长。

4. 分别让每个核小体沿随机轴移动一个角步长。
5. 计算新的哈密顿量 H_2
6. 如果 $H_2 < H_1$, 则采用新的位置和角度; 如果 $H_2 > H_1$, 则以 $e^{-(H_2-H_1)/k_bT}$ 的几率采用新的位置和角度。
7. 继续执行第 2 步直到能量变化达到指定的精度或达到指定的步数为止。

(b) 静电能计算的优化。

模拟中我们考察了 $C=0.001, 0.01, 0.02, 0.1, 0.15, 0.2$ 共 6 种浓度下均匀带电球壳的静电相互作用, 结果如下图。可见溶液浓度对球壳间的静电能有有着很大的影响, 从而也影响了核小体纤维的结构参数(见模拟结果)。

由于静电能的计算较为复杂, 开销较大, 我们预先计算了一个函数值表, 在数值表节点间的函数值则采用线性插值. 下表为不同浓度下静电能, 单位为 $10^{-17} J$.

距离(nm)	0.001M	0.01M	0.02M	0.1M	0.15M	2.0M
6.0	8.17819	3.60207	2.4859	0.84587	0.61855	0.49124
6.1	7.96245	3.43052	2.33607	0.7513	0.53693	0.41825
6.2	7.75449	3.26802	2.19587	0.66747	0.46621	0.35621
6.3	7.55393	3.11402	2.06461	0.59315	0.40491	0.30345
6.4	7.36041	2.96803	1.94169	0.52724	0.35176	0.25857
6.5	7.17359	2.82957	1.82653	0.46877	0.30566	0.22038
6.6	6.99318	2.69821	1.71861	0.41688	0.26566	0.18787
6.7	6.81886	2.57354	1.61744	0.37082	0.23095	0.1602
6.8	6.65038	2.45518	1.52256	0.32993	0.20082	0.13663
6.9	6.48746	2.34276	1.43356	0.2936	0.17466	0.11656
7.0	6.32986	2.23597	1.35005	0.26133	0.15194	0.09945
7.1	6.17735	2.13448	1.27166	0.23266	0.1322	0.08488
7.2	6.02971	2.038	1.19805	0.20717	0.11505	0.07245
7.3	5.88673	1.94625	1.12893	0.18451	0.10014	0.06186

硕士论文	核小体纤维结构特性及环境影响因素的 Monte Carlo 研究					关治强	厦门大学
7.4	5.74823	1.85899	1.064	0.16436	0.08718	0.05282	
7.5	5.614	1.77596	1.00298	0.14644	0.07591	0.04511	
7.6	5.48389	1.69694	0.94563	0.13049	0.06611	0.03854	
7.7	5.35772	1.62172	0.89171	0.1163	0.05759	0.03292	
7.8	5.23533	1.55009	0.84101	0.10367	0.05017	0.02813	
7.9	5.11659	1.48187	0.79332	0.09243	0.04372	0.02405	
8.0	5.00133	1.41688	0.74846	0.08242	0.0381	0.02055	
8.1	4.88944	1.35495	0.70624	0.07351	0.03321	0.01757	
8.2	4.78078	1.29593	0.66651	0.06557	0.02895	0.01503	
8.3	4.67523	1.23966	0.62911	0.05849	0.02524	0.01285	
8.4	4.57267	1.18601	0.59389	0.05219	0.02201	0.01099	
8.5	4.473	1.13484	0.56072	0.04657	0.0192	0.0094	
8.6	4.3761	1.08603	0.52948	0.04157	0.01674	0.00804	
8.7	4.28189	1.03946	0.50004	0.0371	0.01461	0.00688	
8.8	4.19025	0.99501	0.47231	0.03312	0.01274	0.00589	
8.9	4.10111	0.95259	0.44617	0.02957	0.01112	0.00504	

从上表可以看出，溶液浓度对球壳间的静电能有很大的影响。在浓度为 0.001M 溶液中，球心距离为 8nm 的两个均匀带电球壳的静电能为 $5.001 \times 10^{-17} J$ ，而在 0.1M 的溶液中则为 $8.2 \times 10^{-19} J$ ，相差了约 50 倍。以下图形更为直观地显示了这种差异：

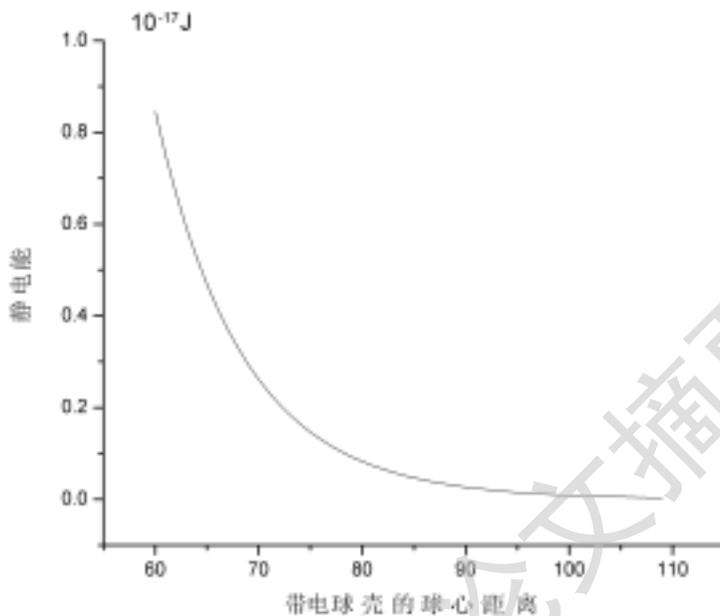


图 2.2 C=0.1M 的势能曲线

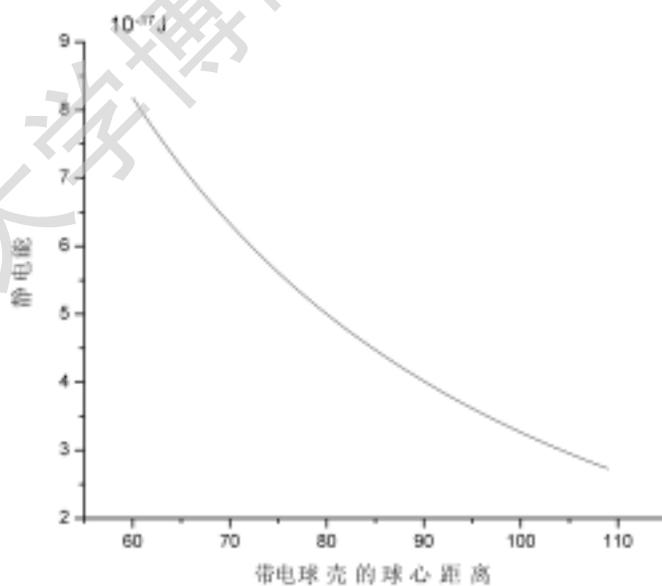


图 2.2 C=0.001M 的势能曲线

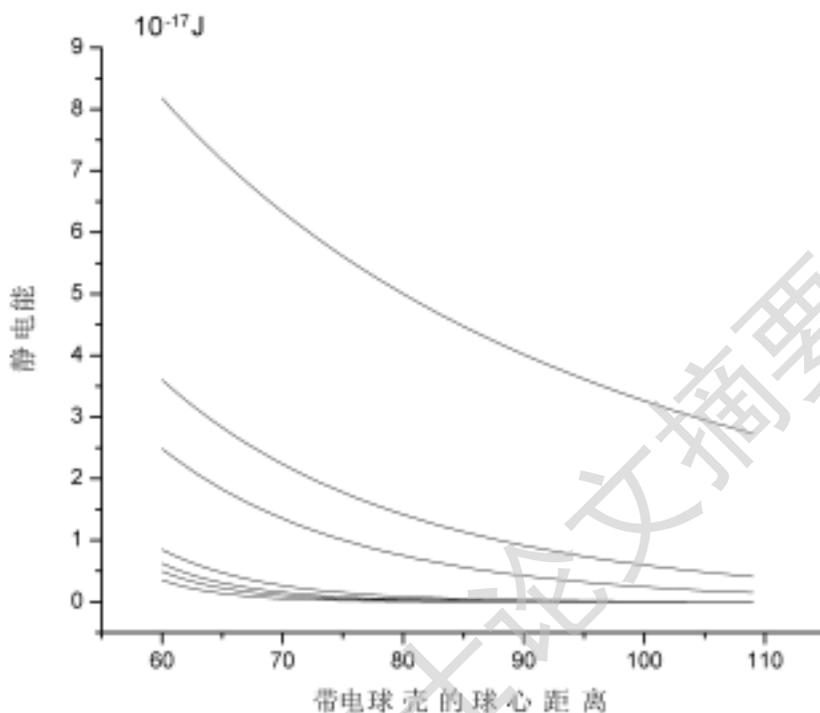


图 2.3 各种溶液浓度下的势能曲线对比

(从上到下依次为 $C=0.001M, 0.01M, 0.02M, 0.1M, 0.15M, 0.2M$ 对应的势能曲线)

2.4 模拟参数及设定：

Stretching module DNA[14]: $1.10 \times 10^{-18}J/nm^{-2}$

Bending module DNA[14]: $2.06 \times 10^{-19}J$

Torsion module DNA[14]: $2.67 \times 10^{-19}J$

Temperature : 300K

核小体半径 : 3.0nm

核小体净带电量 : -190e

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士学位论文摘要库