

# 碱基序列对 DNA 分子电子结构的影响<sup>\*</sup>

徐 慧 郭爱敏<sup>†</sup> 马松山

(中南大学物理科学与技术学院,长沙 410083)

(2006 年 4 月 3 日收到,2006 年 6 月 30 日收到修改稿)

将 DNA 分子看成一维二元随机序列,利用负本征值理论计算其态密度,针对碱基对分布和相对含量等参量,讨论了 DNA 分子的电子结构.结果表明碱基对分布和相对含量都对电子能态结构影响较大,说明碱基序列对 DNA 分子的电子结构影响很大.

关键词:DNA 分子,电子态密度,碱基对

PACC:8715,7120H,6320K

## 1. 引言

DNA 是生物体中储存和传递遗传信息的重要物质和分子信息材料.在 DNA 芯片和 DNA 计算机、DNA 分子器件、超微电子器件、纳米机械、纳米导体、DNA 传感器和 DNA 探针、DNA 分子光开关及 DNA 疫苗等材料与信息科学领域有广阔的应用前景.

早在 1962 年, Eley 和 Spivey<sup>[1]</sup> 提出碱基对中  $\pi$ - $\pi$  的相互作用可以使 DNA 具有导电行为.而近来,人们分别从理论和实验上研究 DNA 分子导电行为及其影响因素,其结果使人吃惊. Barton 等<sup>[2-6]</sup> 利用荧光猝灭法得出电子在 DNA 内高速传递而能成为分子导体.有些人则直接将 DNA 分子搭在两个间距为 8 nm 至 1.5  $\mu$ m 的电极之间,用伏安法测量电阻,结果也大相径庭,包括绝缘体<sup>[7]</sup>、半导体<sup>[8]</sup>、导体<sup>[9]</sup>甚至超导体<sup>[10]</sup>.在理论方面,人们把 DNA 分子近似地当作一维周期结构,并采用不同的近似模型来讨论 DNA 分子的导电性.得到的结果也差别极大,可能是半导体<sup>[11]</sup>,也可能是绝缘体<sup>[12]</sup>.最近, Roche<sup>[13]</sup> 根据转移矩阵方法得出电荷的传播距离和效率强烈地依赖于 DNA 分子的  $\pi$ - $\pi$  堆积势与碱基序列.宋骏等<sup>[14]</sup> 则从 DNA 分子链内的巡游电子数出发,研究其能带结构与电子态.同时,高绪团等<sup>[15]</sup> 和马松山等<sup>[16]</sup> 分别讨论了位置涨落和碱基对组分、格点能量无序度对 DNA 分子电子结构的影响.

DNA 作为一种生物大分子,是由腺嘌呤(A)、鸟嘌呤(G)、胞嘧啶(C)和胸腺嘧啶(T)四种碱基组成的,它们通过碱基互补配对原则形成碱基对.而碱基对间存在着堆积相互势和其他相互作用,由于这些吸引势的存在,相邻碱基对平面之间形成旋转而常常呈现右手双螺旋结构,可见 DNA 分子内部结构非常复杂.在最有代表性的 B 型 DNA 分子中,一个螺旋周期包含约 10 个碱基对,其平面基本垂直于 DNA 分子的中心轴线,而相邻碱基对的轴向距离约为 0.34 nm.为便于理论研究,将 DNA 分子简化为一维二元随机序列模型,该模型是适合研究 DNA 分子的<sup>[17,18]</sup>.我们在考虑最近邻相互作用下,计算了包含 10000 个碱基对的 DNA 分子的电子态密度,同时考虑碱基对分布和碱基对相对含量对态密度的影响,发现通过调整碱基对的分布, DNA 分子将可能呈现出绝缘体、半导体和导体行为.

## 2. 模型和方法

在此,我们分别将 A = T 和 C = G 两个碱基对看成一个整体,则 DNA 分子就成为由 A = T, C = G 构成的一维二元随机序列.进行这样的简化之后,就可以利用一维紧束缚模型下的哈密顿量来描述<sup>[18]</sup>,从而利用负本征值理论来求解它的态密度.为方便起见,以 A 来表示碱基对 A = T,其位能为  $\epsilon_A$ ,相应地,以 C 表示碱基对 C = G,其位能为  $\epsilon_C$ .而各种 DNA

<sup>\*</sup> 高等学校博士学科点专项科研基金(批准号 20020533001)资助的课题.

<sup>†</sup> 通讯联系人. E-mail: guoaimin1982@163.com

分子, 包括  $\lambda$ -phage DNA<sup>[19]</sup>, 它们并非完全无序, 而是存在长程相关性<sup>[18]</sup>; 同时, 一些 DNA 分子片段富含 C≡G 碱基对, 另一些则富含 A = T 碱基对, 于是我们采取如下的方法来确定 DNA 分子的序列: 首先给出一组满足一定分布的数  $\{\eta_i\}$ , 规定一个阈值  $P$ , 当  $\eta_i > P$  时, 在模型的第  $i_R$  个位点连续取  $R$  个碱基对 C, 反之连续取  $R$  个碱基对 A, 则通过调节  $P$  可以改变 DNA 分子中两种碱基对的相对含量, 其中  $i_R = R_1 + R_2 + \dots + R_{i-1} + 1$ ,  $R_k$  表示第  $k$  次操作时的  $R$  值, 而  $R$  可取为一个固定的数或者在一个区间里均匀取值, 表示连续取相同碱基对的个数.

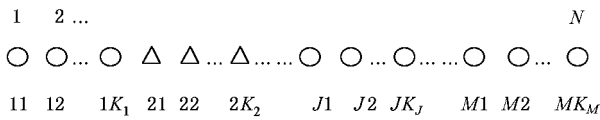


图 1 DNA 一维随机碱基序列模型

图 1 给出了 DNA 分子的模型, 其中  $\circ$  代表碱基对 A (C) 相应  $\triangle$  代表碱基对 C (A). 此模型含有  $N$  个碱基对, 而  $K_i (i = 1, 2, \dots, M)$  为大于 0, 小于  $N$  的整数, 满足

$$\sum_{i=1}^M K_i = N, \quad (1)$$

则该模型的哈密顿量可表示为

$$H = \sum_{i=1}^N \varepsilon_i |i\rangle\langle i| + \sum_{i=1}^N \sum_{j \neq i}^N t_{ij} |i\rangle\langle j|, \quad (2)$$

其中  $|i\rangle$  为一组已知的正交完备基函数,  $N$  为碱基对个数, 取为 10000,  $\varepsilon_i$  为第  $i$  个碱基对的位能,  $t_{ij}$  为第  $i$  个碱基对和第  $j$  个碱基对之间的电子交叠势. 当 DNA 分子处于不同的环境中时,  $\varepsilon_i$  取不同的值, 相应的  $t_{ij}$  也取不同的值. 在此, 我们采用如下的取值<sup>[20, 21]</sup>:

$$\varepsilon_A = 0.2, \quad \varepsilon_C = 0.28, \quad (3)$$

$$t_{ij} = \begin{cases} t_{AA} = -0.0695, \\ t_{AC} = -0.12905, & |i-j| = 1 \\ t_{CC} = -0.1409, \\ 0, & |i-j| > 1 \end{cases} \quad (4)$$

可以将 (2) 式写成  $N$  阶的三对角矩阵形式

$$M = \begin{pmatrix} \varepsilon_1 & t_{12} & 0 & \dots & 0 \\ t_{21} & \varepsilon_2 & t_{23} & \ddots & \\ 0 & t_{32} & \ddots & \ddots & 0 \\ \ddots & \ddots & \ddots & \ddots & t_{N-1N} \\ 0 & \dots & 0 & t_{NN-1} & \varepsilon_N \end{pmatrix}, \quad (5)$$

其本征值可根据 Dean 的负本征值理论<sup>[22]</sup>精确求得, 态密度也可知道.

### 3. 结果与分析

这里我们分别给出了  $\{\eta_i\}$  在随机均匀取值和随机正态取值条件下 DNA 分子的态密度, 如图 2, 3 所示. 在随机均匀取值条件下,  $\eta_i$  在区间 0 到 1 之间, 而在随机正态取值条件下, 随机数均值为 0.5, 方差为 2.25,  $\eta_i$  也在区间 0 到 1 之间. 同时, 我们定义同碱基对分子团, 表示 DNA 分子中连着的相同碱基对通过堆积作用而形成的分子团簇.

#### 3.1. 不同分布的影响

分别比较图 2 和图 3 对应的图, 得知  $R = 1$  时, 在随机均匀分布下, DNA 分子的能带没有发生明显的分裂, 其能量比较连续; 而在随机正态分布下, DNA 分子态密度变成分立的条状或者带状, 其能带发生了明显的分裂, 此时, 电子将从一个局域态跃迁到另一个局域态. 当  $R$  取值较大时, 两种分布下 DNA 分子态密度图将趋于一致. 究其原因, 我们认为在其他条件都相同的情况下, 正态分布随机数主要分布在均值附近, 当给定阈值  $P$  时,  $\eta_i$  跨越  $P$  的概率高, 即连续取的随机数  $\eta_i$  在  $P$  的同一侧的个数少, 则  $R = 1$  时的碱基对排列更加没有规律, 无序度更大, 意味着同碱基对分子团更加小, 也越容易受到邻近碱基对分子团的影响, 从而产生能隙, 同时由于长程关联性的缺失<sup>[23]</sup>和长的序列长度 ( $L = 3.4 \mu\text{m}$ ) 而能成为绝缘体. 从分析两种分布下 DNA 分子碱基对序列也可说明这一点, 图 3(a)(c)(e) 中, 同碱基对分子团个数的最大值分别为 35, 12, 29; 而在图 3(a)(b)(c) 中, 同碱基对分子团个数的最大值分别为 9, 6, 10. 当  $R$  比较大时, 碱基对排列也越有序, 同碱基对分子团也较大, 则受邻近碱基对分子团的影响将变小, 不同分布对态密度的影响也将变小.

#### 3.2. 碱基对相对含量的影响

显而易见, 完全由碱基对 C 构成的理想 DNA 分子为一维周期结构, 其态密度图将在能带两端出现两个峰<sup>[24]</sup>, 很明显它们的本征值主要集中在两个峰上, 说明此时 DNA 分子中电子的简并度大, 系统的对称性好, 这反映了有序情况; 而随着碱基对 A

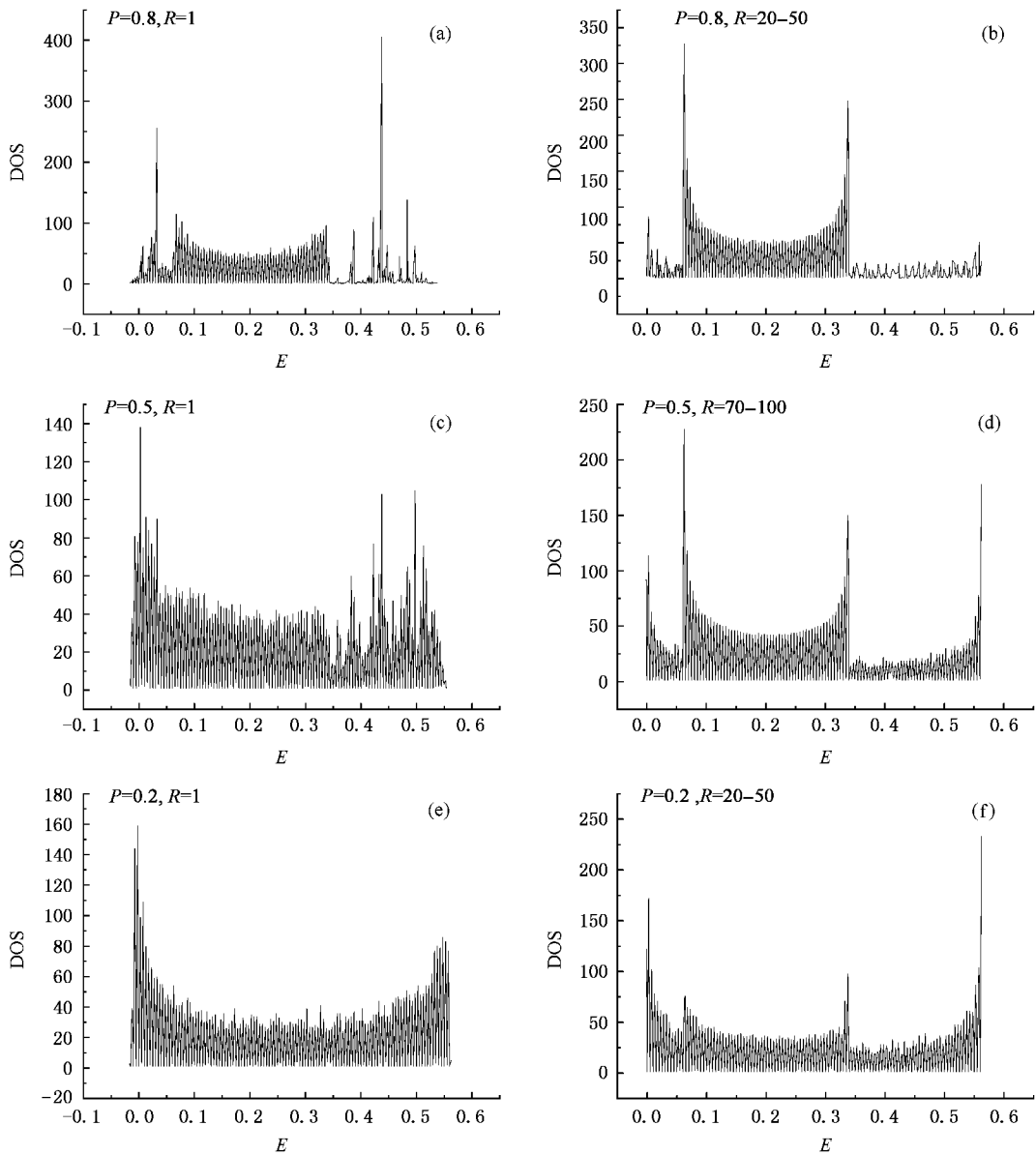


图 2 随机均匀分布态密度

的掺入,系统的对称性遭到破坏,峰值下降,能带宽度也有所展宽,这都体现了系统由有序变成无序所发生的变化.分析和比较图 2(a)(c)(e),可得:此时 DNA 分子是由碱基对 A、C 共同组成的,在碱基对 C 占据主导地位时,系统仍具有很好的对称性,如图 2(e),此时电子态将在整个系统中扩展而能成为导体,随着位能较高的碱基对 C 相对含量的减少, DNA 分子的能带宽度也慢慢变窄,右边的峰逐渐劈裂成几个峰,如图 2(c),电子态变为局域态,同时,在  $E = 0.0625$  eV 和  $E = 0.3375$  eV 会出现另外两个峰;随着碱基对 C 的进一步减少,直到碱基对 A 在 DNA 分子中占据主导地位时,则上述两个峰更

加明显且峰值更大,并在能带两端出现两个很明显的能隙(图 2(a)),此时 DNA 分子将具有半导体行为;当碱基对 C 的相对含量为 0 时,则 DNA 分子是完全由碱基对 A 构成的理想生物大分子,也会在能带两端出现两个峰,但能带宽度明显变小.再比较图 2(b)(d)(f),可知:对 R 较大的情况,随着碱基对 C 相对含量的逐渐减少,能带两端的峰值将变小,而中间的峰值将增大.同样地,在随机正态分布下, DNA 分子的态密度图随着碱基对 C 相对含量的减少也会出现上述情况.

产生如上现象跟我们即位能和相邻碱基对之间交叠势的取值有很大关系.由两种元素构成的一维

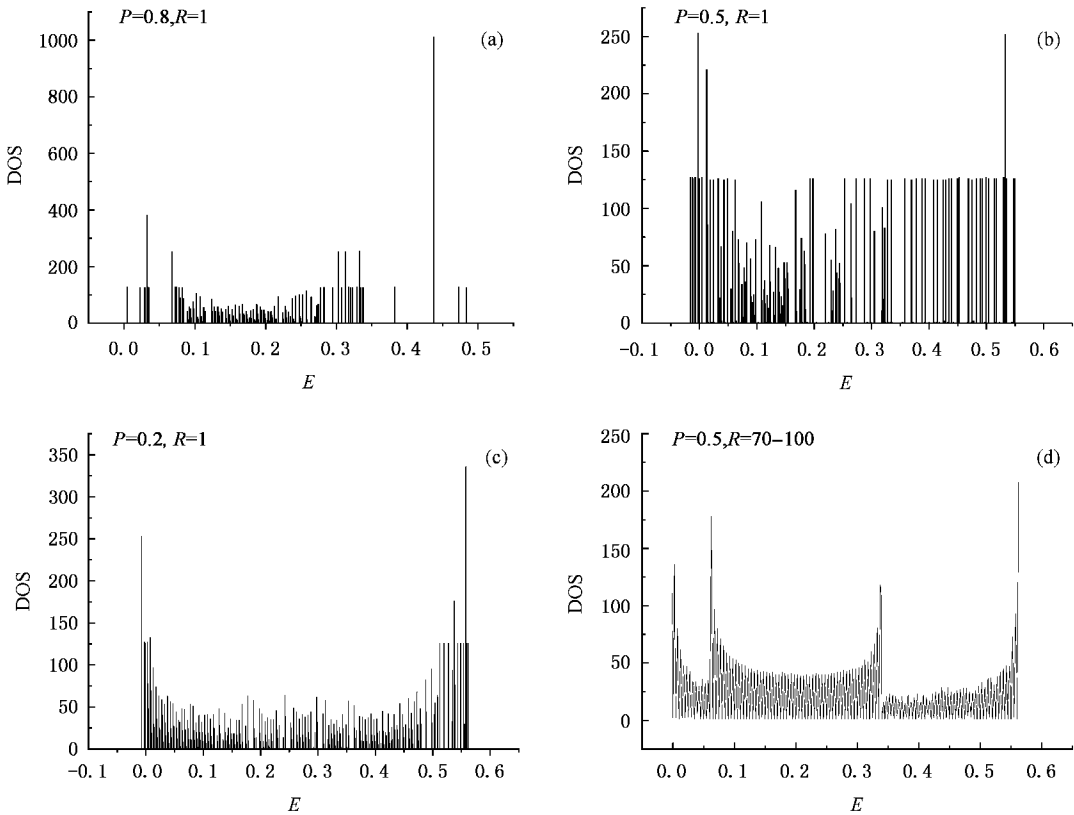


图3 随机正态分布态密度

二元随机序列, 一般而言, 只取两个不同的位能更加符合此模型, 而当两元素按一定序列排列后, 则相邻元素之间的交叠势就完全确定了. 对于 DNA 分子, 相邻碱基对通过  $\pi$  相互作用形成有利于载流子传导的“通路”, 若碱基对间的交叠势越大, 则意味着此“通路”能更好地传导载流子. 由于碱基对 C 之间具有最大的交叠势, 则在碱基对 C 占据主导的 DNA 分子中, 将能提供更好的“通路”, 因此此时 DNA 分子具有导体行为; 而在碱基对 A 占支配地位的 DNA 分子中, 同样能提供“通路”, 但由于碱基对 A 间的交叠势最小, 因此 DNA 分子趋于半导体行为.

### 3.3. 分子团的影响

显然, 分子团的大小与  $R$  的大小成正比例关系. 从图 2(a)(c)(e) 和图 3(a)(b)(c) 看到, 在  $R$  比较小的时候, 不同分布下的态密度有很大区别, 而随着  $R$  的增大, 不同分布对态密度的影响将变小. 再分别比较图 2(a) 和 (b)(c) 和 (d)(e) 和 (f), 得知在随机均匀分布下, 随着  $R$  的增大, 当碱基对 A 占主导地位时, 在  $E = 0.0625\text{eV}$  和  $E = 0.3375\text{eV}$

处的峰值将升高, 而能带两端的峰值则下降, 能隙宽度变小甚至能隙消失; 当碱基对 C 占主导地位时, 在  $E = 0.0625\text{eV}$  和  $E = 0.3375\text{eV}$  处会出现两个峰; 当碱基对 A 和 C 的相对含量相当时, 则在能带中出现四个高峰. 当  $R$  足够大时, 碱基对相对含量相同的 DNA 分子态密度随  $R$  将基本保持不变, 如图 2(d) 和图 3(d) 所示, 说明其电子结构已趋于稳定. 另外, 在正态分布下,  $R$  的增大同样使得 DNA 分子在其能带中间产生两个峰值, 最终也趋于稳定. 同时, 我们还注意到, 在  $R$  足够大时, 任何一种 DNA 分子的态密度都可以看成完全由 A 构成的 DNA 分子态密度和完全由 C 构成的 DNA 分子态密度的线性叠加, 说明此时同碱基对分子团已很稳定, 基本可以忽略邻近碱基对分子团的影响, 也意味着电子结构的稳定性, 此时电子态将在整个系统中扩展而趋于导体行为.

总之,  $R$  的增加使得碱基对排列更加有序, 同碱基对分子团变得更大, 以至来自邻近碱基对分子团的影响可以忽略而趋于稳定, DNA 分子的电子结构也趋于稳定而能成为导体.

## 4. 结 论

根据以上分析,得知 DNA 分子中碱基序列对其态密度有很大的影响,同时也会影响 DNA 分子的导电行为,主要表现在如下几个方面:

(1)在  $R$  比较小的时候,不同分布下的态密度有很大区别;而随着  $R$  的增大,不同分布对态密度的影响将变小.当  $R = 1$  时,随机正态分布下的 DNA 分子将有可能成为绝缘体.

(2)由于邻近碱基对间的堆积相互作用及其差别,碱基对  $C$  占主导地位的 DNA 分子将倾向于导体,而碱基对  $A$  占主导地位的 DNA 分子将倾向于半导体.

(3)随着同碱基对分子团的增大( $R$  的增大),由碱基对  $A$  和  $C$  构成的 DNA 分子将在其态密度图中出现四个峰.同时碱基对分子团的增大意味着 DNA 分子更加有序性,使得其电子结构趋于稳定,电子态在整个系统中扩展而能成为导体.

- 
- [ 1 ] Eley D D , Spivey D I 1962 *Trans. Faraday Soc.* **58** 411
- [ 2 ] Arkin M R , Stemp E D A , Holmlin R E , Barton J K , Hörmann A , Olson E J C , Barbara P F 1996 *Science* **273** 475
- [ 3 ] Hall D B , Holmlin R E , Barton J K 1996 *Nature* **382** 731
- [ 4 ] Dandliker P J , Holmlin R E , Barton J K 1997 *Science* **275** 1465
- [ 5 ] Kelley S O , Barton J K 1999 *Science* **283** 375
- [ 6 ] Treadway C R , Hill M G , Barton J K 2002 *Chem. Phys.* **281** 409
- [ 7 ] De Pablo P J , Moreno-H F , Colchero J , Herrero J G , Herrero P , Baró A M , Ordejón P , Soler J M , Artacho E 2000 *Phys. Rev. Lett.* **85** 4992
- [ 8 ] Porath D , Bezryadin A , De Vries S , Dekker C 2000 *Nature* **403** 635
- [ 9 ] Fink H W , Schönenberger C 1999 *Nature* **398** 407
- [ 10 ] Kasumov A Y , Kociak M , Guéron S , Reulet B , Volkov V T , Klinov D V , Bouchiat H 2001 *Science* **291** 280
- [ 11 ] Pullman B , Pullman A 1961 *Nature* **189** 725
- [ 12 ] Dee D , Baur M E 1971 *J. Chem. Phys.* **60** 541
- [ 13 ] Roche S 2003 *Phys. Rev. Lett.* **91** 108101
- [ 14 ] Song J , Chen L , Liu D S , Xie S J 2004 *Acta Phys. Sin.* **53** 2792 ( in Chinese ) [ 宋 骏、陈 雷、刘德胜、解士杰 2004 物理学报 **53** 2792 ]
- [ 15 ] Gao X T , Fu X , Song J , Liu D S , Xie S J 2006 *Acta Phys. Sin.* **55** 952 ( in Chinese ) [ 高绪团、傅 雪、宋 骏、刘德胜、解士杰 2006 物理学报 **55** 952 ]
- [ 16 ] Ma S S , Xu H , Liu X L , Guo A M 2006 *Acta Phys. Sin.* **55** 3170 ( in Chinese ) [ 马松山、徐 慧、刘小良、郭爱敏 2006 物理学报 **55** 3170 ]
- [ 17 ] Yu Z G , Song X 2001 *Phys. Rev. Lett.* **86** 6018
- [ 18 ] Carpena P , Bernaola-G P , Ivanov P C , Stanley H E 2002 *Nature* **418** 955
- [ 19 ] Sanger F , Coulson A R , Hong G F *et al* 1982 *J. Mol. Biol.* **162** 729
- [ 20 ] Endres R G , Cox D L , Singh R R P 2002 *Cond-mat/0201404*
- [ 21 ] Zhang W , Ulloa S E 2004 *Microelectron. J.* **35** 23
- [ 22 ] Dean P 1972 *Rev. Mod. Phys.* **44** 127
- [ 23 ] Roche S , Bicout D , Maciá E , Kats E 2003 *Phys. Rev. Lett.* **91** 228101
- [ 24 ] Song Z Q , Xu H , Li Y F , Liu X L 2004 *Acta Phys. Sin.* **54** 2198 ( in Chinese ) [ 宋招权、徐 慧、李燕峰、刘小良 2004 物理学报 **54** 2198 ]

# The influence of base pair sequence on electronic structure of DNA molecules<sup>\*</sup>

Xu Hui Guo Ai-Min<sup>†</sup> Ma Song-Shan

( *College of Physical Science and Technology, Central South University, Changsha 410083, China* )

( Received 3 April 2006 ; revised manuscript received 30 June 2006 )

## Abstract

The DNA molecule is considered as one-dimensional disordered binary system. Its electronic structure is discussed in terms of base pairs' relative percentage and distribution using the density of electronic states which are calculated by the negative eigenvalue theory. The result shows that the effect of the base pairs' relative percentage and distribution on the electronic energy state structure is great. In other words, the sequence of the base pairs plays an important role in determining the electronic structure.

**Keywords** : DNA molecules , density of electronic state , base pairs

**PACC** : 8715 , 7120H , 6320K

---

<sup>\*</sup> Project supported by the Specialized Research Found for the Doctoral Program of High Education of China( Grant No. 20020533001 ).

<sup>†</sup> Corresponding author. E-mail : guoaimin1982@163.com