一种优化 DNA 计算模板性能的新方法

刘文斌[①] 朱翔鸥[①] 王向红[①] 张 强[②] 马润年[③]
[①](温州大学计算机科学与工程学院 温州 325035)
[②](大连大学信息科学与工程重点实验室 大连 116622)
[③](空军工程大学电讯工程学院 西安 710077)

摘 要： 编码问题是目前 DNA 计算中的重点和难点之一，该文介绍了影响编码的各种因素及模板编码的基本思想。在此基础上分析了移位杂交出现的原因，提出了提高模板结合移位距离的一种新算法。该算法一方面降低了搜索空间，另一方面筛选了那些自身移位距离性质差的序列因而提高了算法的效率。计算结果表明模板集合的性能明显提高。此外，在保持 01 含量基本不变的情况下，适当扩展模板集合的搜索范围可以增加模板的数量。

关键词： DNA 计算；编码问题；模板编码方法
中图分类号： TP18 **文献标识码：** A **文章编号：** 1009-5896(2008)05-1131-05

A New Method to Optimize the Template Set in DNA Computing

Liu Wen-bing[①] Zhu Xiang-ou[①] Wang Wiang-hong[①] Zhang Qiang[②] Ma Run-nian[③]
[①](*College of Computer Science and Engineering*, *Wenzhou University*, *Wenzhou* 325035, *China*)
[②](*University Key Lab of Information Science & Engineering*, *Dalian University*, *Dalian* 116622, *China*)
[③](*Telecommunication Engineering Institute*, *Air Force Engineering University*, *Xi'an* 710077, *China*)

Abstract: The encoding issue is a most fundamental one in DNA based computing. In this paper, the various factors that influence the encoding problem and the general idea of the template encoding method are first introduced. Then the reason of the shift hybridisation occurred in DNA computing is presented. And a new method is proposed to search template set with high shift distance. Additionally, to increase the search space can also increase the number of template string.

Key words: DNA computation; Encoding issue; Template method

1 引言

DNA 计算是近年来计算机研究领域的一个热点方向，在近几年分子生物计算机的研究中备受学者们的关注，其标志是 1994 年 Adlema 在 Science 上发表的文章——Molecular computation of solution to combinatorial problems[1]。在这种新型计算方式中，信息是通过 DNA 分子的 4 种碱基来编码的，并通过 DNA 分子间的特异性杂交来实现的。

由于 DNA 计算中的核心操作——杂交反应在不完全互补的情况下也有可能发生，从而形成各种不希望的两级结构(如图 1(b), 1(c), 1(d), 1(e))，从而导致错误的计算结果。在聚合酶链式反应 (Polymerase Chain Reaction, PCR)扩增过程中，引物与引物之间同样会出现上述不希望的两级结构，以致扩增失败。因此，如何通过有效的编码来提高 DNA 计算过程中的“信噪比”，是 DNA 计算研究中的一个重点和难点问题。

编码研究的目的就是希望能够在实际的生化反应过程中，编码每一个信息元的 DNA 序列能够被最大限度地唯一识别，从而使得计算过程能够按照计算模型所设计的方向进行。目前有关编码的研究主要集中在如何降低编码之间的相似度。Garzon 给出了 DNA 计算中的编码问题定义[2]，他还借鉴二进制超立方体的理论对编码进行研究[3]。Baum 提出降低 DNA 序列间的相似度假设[4]。Feldkamp 等给出了另一种定义序列间相似度的方法[5]。Suyama 等在基于 DNA 计算

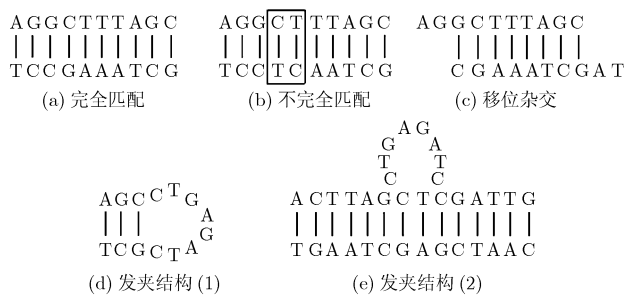

图 1 几种可能的杂交形式

2006-10-25 收到，2007-03-23 改回
国家自然科学基金(60403002，60403001，30670486)，中国博士后科学基金(2004036130)和浙江省自然科学基金(Y106654，Y405553)资助课题

的基因表达分析的 DNA 编码数概念[6]。有的学者还提出采用三字母表的编码策略[7]，来降低 DNA 分子生产两级结构。Braich 等提出了 DNA 序列编码的约束条件，用来解决可满足性问题的编码问题，并在实验中取得了良好的效果[8]。Frutos 等提出的模板编码方法[9]。本文提出了一个随机搜索算法，对模板编码方法做了进一步的研究。由于满足条件的模板集合和映射集合并不唯一，于是应用 Garzon 提出的移位距离对模板集合进行优化，计算结果表明，通过优化后模板生成的 DNA 编码的质量明显提高[10]。本文则提出了一种优化搜索空间的方法来提高模板集合移位距离。

2 编码问题及其约束条件

2.1 编码问题[2]

DNA 计算中的编码问题可以表述为：以构成 DNA 分子的 4 个碱基为字母表 $\Sigma=\{\mathrm{A,T,G,C}\}$，存在一个长度为 l 的 DNA 分子的基础编码集合 Z, $Z=\Sigma^{l}=\{<b_1,b_2,\cdots,b_l>|b_i\in\Sigma, i=1,2,\cdots,l\}$，显然$|Z|=4^{l}$。求 Z 的一个子集 W 使得

$$\forall x_i,x_j\in W,\ \tau(x_i,x_j)\geq k \tag{1}$$

其中 k 为正整数，τ 是评价编码的期望准则，如汉明距离、GC 含量、最大相同子序列长度等。另外，在编码问题中的还要考虑集合 W 的大小$|W|$。因为$|W|$越大，可供选择的满足条件的编码越多。显然评价准则 τ 越严格，可供选择的编码数量$|W|$就越小。

2.2 约束条件

2.2.1 化学自由能变化 ΔG 任何一个化学反应都是一个热平衡的过程，由化学热力学可知，杂交反应的方向为自由能减小的方向。对于长度为 l 的 DNA 分子，自由能变化 ΔG 可用下式近似计算[11]

$$\Delta G=\theta+\sum_{i=1}^{l-1}w(b_i,b_{i+1}) \tag{2}$$

其中 θ 为一个修正值，w 是长度为 2bp 的序列 b_ib_{i+1} 的一个负的权值。在其他条件相同的情况下，杂交配对程度越高，w 的绝对值越大，其自由能的下降值也越大，双链的热力学稳定性也越好。

2.2.2 解链温度 T_m 解链温度 T_m 是指双链 DNA 分子在变性过程中，有 50%的碱基对变为单链时的温度。它是评价 DNA 分子的热力学稳定性的一个重要的参数。解链温度随着“GC 含量”的增加而升高，解链温度还与“GC”在双链中位置分布的均匀程度有关，“GC”在双链中位置分布不均匀将导致解链温度的下降。解链温度 T_m 值的计算公式为

$$T_m=\Delta H^{\circ}/(\Delta S^{\circ}+R\ln C_t) \tag{3}$$

其中 ΔH° 和 ΔS° 分别为杂交反应的标焓变和熵变，其计算方法参见文献[11]，R 为气体常数 1.987cal/kmol，C_t 为 DNA 分子的摩尔浓度。

2.2.3 DNA 分子的组成 在 DNA 计算中，由于有大量 DNA 分子参加杂交反应，因此，我们希望所有参加反应的 DNA 分子在完全杂交时的解链温度 T_m 和自由能变化 ΔG 都能够保持在一个比较小的区间。这样就能够通过控制生化反应的各种参数来提高反应的可靠性。因为在双链 DNA 分子中 A-T 间有两个氢键，而 G-C 间有三个氢键，因此 GC 含量对 DNA 分子的解链温度 T_m 和自由能变化 ΔG 有很大的影响。由于在 PCR 反应中的引物设计中，一般要求 GC 含量约为 50%，因此，在编码问题中通常都取此值。

2.2.4 生物酶 酶是生化反应中的重要工具，不同的酶相当于不同的算子。在各种 DNA 计算模型中，经常都要借助于特定的生物酶来完成特定的目标。因此，在编码中就要考虑在 DNA 序列的特定位置设计酶的识别序列。同时，为了保证生物酶作用的可靠性，就必须保证其识别序列只能唯一出现在特定的位置。

2.2.5 编码距离 任何计算模式的实质都可以归结为对信息的传输和处理过程，编码距离实际上就是描述任意两个编码间 “相似度”常用的一个参数。编码距离越大其“相似度”越小。信息论中纠错码方法有效地解决了以 0，1 编码的电子计算机中的编码问题，其数学基础是用汉明距离来度量二进制超立方体空间中的两个顶点间的距离。由于 DNA 计算的特殊性，又引伸出其它几种扩展形式。下面简要介绍一下它们的定义[12]：

(1)汉明距离 $H(x_i,x_j)$：序列 x_i 和 x_j 上所有对应位置上字符不同的总和；

(2)汉明反距离 $H^{r}(x_i,x_j)$：序列 x_i 和序列 x_j 的反序列 x_j^{r} 之间的汉明距离；

(3)汉明补距离 $H^{c}(x_i,x_j)$：序列 x_i 和序列 x_j 的补序列 x_j^{c} 之间的汉明距离，对于二进制序列，其补序列是将所有的“0”变为“1”，所有的“1”变为“0”后得到的。对于 DNA 序列，则是将所有字母变为与其配对的碱基字母；同时，其方向将发生变化，即原来序列方向为 $5'\rightarrow 3'$，则其补序列变为 $3'\rightarrow 5'$；

(4)汉明反补距离 $H^{rc}(x_i,x_j)$：序列 x_i 的补序列 x_i^{c} 和序列 x_j 的反序列 x_j^{r} 间的汉明距离；

(5)H 测度 H_G：序列 x_j 相对序列 x_i 移动 $k(-n<k<n)$ 个位置后所得的汉明距离的最小值

$$H_G(x_i,x_j)=\min_{-n<k<n}H(x_i,\rho^{k}(x_j))=n-c_{ij} \tag{4}$$

其中 ρ^{k} 表示偏移 k 个位置，c_{ij} 为序列 x_i 和 x_j 偏移 k 个位置后的最大相同字符之和。显然，H 测度可以更为准确地描述两个序列间的相似度。该定义最早是由 Garzon 提出[13]，在本文中，要求当 $x_i=x_j$ 时，$k\neq 0$。这样，该定义就可以用来计算序列 x_i 中出现的最大重复子序列的长度。

H 测度 H_G 是实质上表示了两个序列移动了 k 个位置后得到的最小的 Hamming 距离。由于 DNA 分子之间可能发生移位杂交，因此 H 测度 H_G 可以用来很好地度量两个序列间的相似性。H 测度 H_G 越小，表示两个序列之间的相似度

大，容易发生移位杂交；H 测度 H_G 越大，表示两个序列之间的相似度小，越不易发生移位杂交。因此，H 测度 H_G 更为准确地衡量 DNA 序列间的距离性质。

3　模板编码方法

模板方法(Template method)最初是由 Wisconsin 大学的 Frutos 等提出的，其主要思想是：通过一个二进制序列的模板(template)集合 T 和另一个二进制序列的映射(Map)集合 M 生成满足要求的 DNA 序列集合 S

$$T \times M \to S \tag{5}$$

其生成规则为 $1\times1 \to T$，$1\times0 \to A$，$0\times1 \to G$，$0\times0 \to C$。图 2 给出了一个由模板方法生成的长度为 8 的 DNA 序列的示意图。

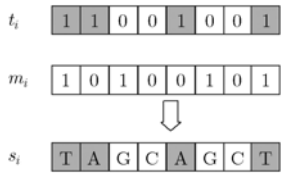

图 2　DNA 序列的生成示意图

Frutos等最初对编码的要求为：(1)编码的GC含量约50%；(2)对任意两个编码 s_i, s_j，要同时满足 $H(s_i,s_j) \ge d$ 和 $H^{rc}(s_i,s_j) \ge d$，其中 d 编码长度的一半。对于条件(1)，按照模板的思想只要模板集合 T 中的序列0，1含量相等即可。由式(5)可知，在模板方法中任意一个DNA序列 s 及其反序列 s^r 是由如下规则产生的：

$$(1)\ \ t\times m \to s \quad (5' \to 3') \tag{6}$$

$$(2)\ \ t^r \times m^{rc} \to s^c \quad (3' \to 5') \tag{7}$$

$$(3)\ \ t^r \times m^r \to s^r \ (3' \to 5') \tag{8}$$

其中 $t \in T$，$m \in M$，$s \in S$。由式(5)和式(7)可以看出，当一个模板序列 t 为对称序列即 $t = t^r$ 时，同时有两个映射序列 m_i, m_j 满足 $m_i^r = m_j^c$，则由它们生成的DNA序列 s_i, s_j 互补。因此，将模板集合划分为两个独立的子集 T_n 和 T_s

$$T = T_n \cup T_s \tag{9}$$

其中 T_n 为非对称模板序列(不回文序列)集合，T_s 为对称模板序列(回文序列)集合。相应的将与它们对应的映射集合称为 M_n 和 M_s。在模板理论下可以得到如下结论：

结论1　当模板集合 T_n 满足汉明距离及汉明反距离时，映射集合 M_n 只需满足汉明距离，由 T_n 和 M_n 生成的序列集 S_n 满足汉明距离及汉明反补距离。

结论2　当模板集合 T_s 仅满足汉明距离，映射集合 M_s 同时满足汉明距离及汉明反补距离时，由 T_s 和 M_s 生成的序列集 S_s 满足汉明距离及汉明反补距离。

结论3　当模板集合 T_n 和 T_n 满足汉明距离时，$S = S_n \cup S_s$ 同时满足汉明距离及汉明反补距离。

模板编码方法的数学基础：当模板 t_i，t_j 或映射 m_i，m_j 具有某种性质时，其生成的两个DNA序列 s_i，s_j 也具有此性质。其实质是将 $\sum_{\mathrm{DNA}} = \{\mathrm{A,G,C,T}\}$ 上的编码问题转化为 $\sum_{01} = \{0,1\}$ 上的编码问题，即在 n 维二进制超立方体空间中搜索满足条件的最大模板集合 T 和最大映射集合 M。这实际上是一个NP-完全问题，即二进制超立方体中的最大独立集问题(independent set problem)。模板方法所能得到的编码集合的大小可以表示为

$$T = |T_n| \times |M_n| + |T_s| \times |M_s| \tag{10}$$

由于在模板编码方法中，模板集合的性质对于最终的编码性质影响最大，而映射集合，特别是映射集合 M_n 已经在传统的信息论中有现成的算法，其大小随编码长度的增加线性增加。因此，本文将主要研究模板集合的优化算法。

4　模板集合的搜索算法

由于在模板编码方法中，编码性能的好坏取决于得到的模板集合，为了得到较好性能的 DNA 序列就要优化模板集合。显然，任意两个 DNA 序列不外乎是由图 3 所示的 3 种方式生成：(1)同一个模板序列和两个不同的映射序列；(2)两个不同的模板序列和同一个的映射序列；(3)两个不同的模板序列和两个不同的映射序列。显然，两个 DNA 序列间的 H 测度是由它们之间的相同子序列决定的。而相同子序列的出现必须具备两个条件：(1)两个模板序列间有一个相同的子序列 $t' = a_1, a_2, \cdots, a_p$，两个映射序列间有一个相同的子序列 $m' = b_1, b_2, \cdots, b_p$ $(1 < p < n)$；(2) t' 在两个模板序列中的位置必须与 m' 在相应的映射序列中的位置相同。为了最大限度地避免 DNA 序列间的移位杂交，必须尽可能减小模板集合 T 和映射集合 M 中的各序列间的相同子序列的最大长度及其出现频率。如“00000000”，“11111111”，“01010101”等，其自身存在各种长度的重复子序列。例如，在“00000000”中有 2 个“0000000” 子序列，3 个“000000”子序列和 4 个“000000”子序列等。它们的存在增加了 DNA 序列中特定长度的相同子序列出现的概率。因此，模板集合的优化问题应该考虑两个方面：(1)单个模板序列自身的 H 测度性质；(2)模板序列间的 H 测度性质。

模板编码方法的实质是在二进制超立方体中搜索满足条件的一个集合，n 维超立方体可以用 2^n 不同的点来表示。

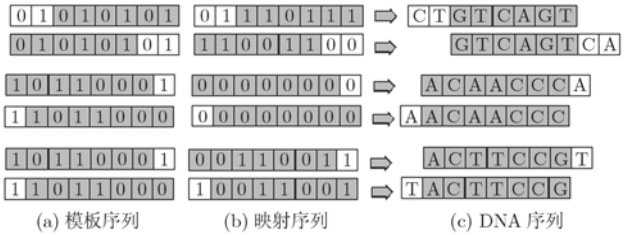

图 3　DNA 序列的 3 种生成方式

在超立方体图中所有的编码按照 01 的含量将被分成不同的列，如果两个点之间的 Hamming 距离为 1 则用一条边将它们连接起来。由于 01 含量系统的序列间的距离为偶数，因此，在同一列中所有的顶点之间不会有边。通常长度为 n 二进制串有 $n+1$ 列，且每一列的 01 含量相同(如图 4)。由于要求编码满足 GC 含量约为 50%，因此，模板集合应该在超立方体中 01 含量相等的列中搜索，从图 4 可以看出，中间列的顶点数量最多。同时随着编码的长度 n 增加，可以发现在超立方体的中间列 $\lceil n/2 \rceil$ 的两边的列 $\lceil n/2 \rceil -1$ 和 $\lceil n/2 \rceil +1$ 的 01 含量也满足约束条件。因此，可以将模板集合的搜索空间由中间一列 $\lceil n/2 \rceil$ 扩展为 3 列 $\lceil n/2 \rceil$，$\lceil n/2 \rceil -1$ 和 $\lceil n/2 \rceil +1$，从而适当增加模板编码的数量。

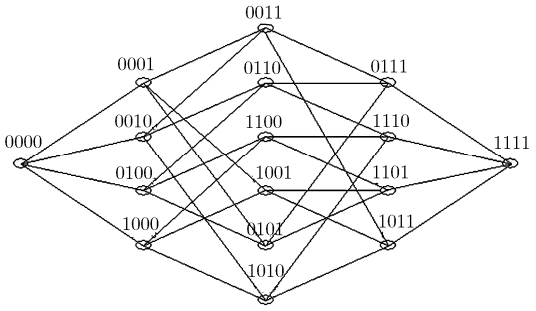

图 4 4 维超立方体图

4.1 算法

从数学的角度，DNA 计算中的编码搜索问题实质上是一个困难的 NP-完全问题[12]。即使将模板集合的限定为特定的一列或三列，其搜索空间也将随编码长度的增加而指数增加。为了得到满足一定要求的最大模板集合，只能采用随机搜索的方法。由于 DNA 序列的方向性，我们不仅要保证两个编码在同一方向时保持足够的移位 H 测度，还要保证两个编码间相互杂交的可能性很小。因此在模板集合的搜索过程中采用

$$h(t_i,t_j)=\min(H_G(t_i,t_j^r),H_G(t_i,t_j)) \tag{11}$$

来计算模板序列间的移位距离，当 $t_i=t_j$ 时表示单个模板自身的移位距离。

当确定了模板集合的搜索空间，即直接特定 01 含量二进制序列后，显然其中包含很多自身移位距离 $h(t,t)$ 很小的二进制串，它们的存在，一方面容易导致随机搜索算法挑选那些自身移位距离性质差的二进制序列进入候选的模板集合，从而严重影响了模板集合的移位距离性质；另一方面，也增加了搜索算法的计算工作量。因此，将模板集合的搜索分两个方面：(1)筛选初始搜索空间(4.2 节的(1), (2), (3))；(2)在筛选后的空间上进行随机搜索。

4.2 算法步骤

(1)产生特定 01 含量的二进制串集合 S；

(2)选取一个二进制串 t_i 并从 S 中将其删除，若 $h(t_i,t_i) \geq d$ 则将其加入到集合 Space 中；

(3)重复步骤(2)，直到集合 S 为空；

(4)由 Space 生成一个关系矩阵 Matrix，当 $h(t_i,t_j) \geq d'$ 的为 1，否则为 0(由于 $d \geq d'$ 因此矩阵的正对角线上都是 1)；

(5)选择矩阵 Matrix 中 1 的个数最多的行和对应的列(若有多行，随机抽取一行)，将其对应的二进制串加入模板集合 T；删除该行及该行中为 0 的元素对应的行和列；

(6)重复(5)，直到 Matrix 为空；

(7)最后所得的 T 即为模板集合 T。

4.3 结果分析

本文主要计算了编码长度 n 为 8，12，16，20，24 等几种情况，为了比较模板序列之间的相似度，利用移位距离来评价一个编码集合的性能，模板的平均移位距离就是模板 t 和模板集合 T 中其他序列移位距离之和除以比较次数，计算结果如表 1 所示。

4.3.1 移位距离 移位距离是衡量模板集合性能的一个有力的指标。结果表明，在编码长度为 8，12，16 长度的时候，本文提出的算法所得到的模板序列的最小移位距离和平均移位距离均优于文献[10，12]的结果。因此，通过优化搜索空间，可以明显提高模板集合的移位距离性质。

4.3.2 模板数量 模板集合的大小直接影响最终的编码数量，显然编码数量与搜索过程中的约束强度密切相关。对于编码长度为 20，24 的情况，本文分别计算了两种不同约束强度时的模板数量，显然，约束强度越强，模板数量越少。因此，可以通过适当地调整约束强度，增加满足条件的模板数量，通常认为模板间移位距离约为模板长度的 1/3 比较合适。此外，在长度不变地情况下，通过改变模板的权值(1 的

表 1 计算结果

	$l=8$		$l=12$		$l=16$		$l=20$		$l=24$	
	$(3,2)^*$	T_{liu}^{**}	(5,4)	T_{liu}	(7,6)	T_{liu}	(9,8)	(9,7)	(11,10)	(12,9)
最小移位距离	2	1	4	1	6	2	8	7	10	9
最大移位距离	3	4	5	6	7	8	8	8	10	10
平均移位距离	2.4	2.8	4.1	4.4	6.1	6	8	7.4	10	9.1
模板总数	11	8	11	8	8	16	5	11	5	7

注：* (3,2)表示 $d=3,d'=2$；** T_{liu} 文献[10]中得到的模板序列

个数)以扩大搜索范围，也有可能增加模板的数量。在长度16的情况，将搜索的空间扩展到权值为7，8，9的所有模板序列，得到的结果如表2所示。可以看出，扩展后得到的模板的性质和未扩展时得到的模板的性质相近，而数量却也所增加。

表2 L=16 扩展搜索空间对模板数量的影响

T (d=9)	总数	W=7	W=8	W=9	最小移位距离	最大移位距离	平均移位距离
$\boldsymbol{A}$($d'=6$)	8	3	1	4	6	7	6.1
$\boldsymbol{B}$($d'=6$)	7	0	7	0	6	7	6.1

注：W为序列中1的个数，d'为模板间的移位距离

5 结束语

编码问题是DNA 计算中的一个基本问题，它直接影响到DNA 计算的可靠性和计算效率。模板编码方法是目前编码问题研究最有前景的一种方法，DNA 计算中的编码问题的难度主要在于DNA 序列间发生的各种移位杂交。由于模板集合的性质直接影响最终编码的性质，如何提高模板集合的鲁棒性就是一个急需解决的问题。本文在分析了移位杂交产生的原因后，提出通过两个方面来提高模板序列的移位距离特性：(1)模板序列自身的移位距离性质；(2)模板间的移位距离性质。与文献[10]相比，本文的算法具有以下优点：

(1)由于移位杂交是导致DNA 计算编码问题复杂的主要原因，因此直接采用移位距离约束来搜索模板序列更为合理，计算结果表明模板集合的移位距离性质得到明显提高；

(2)由于首先采用 $h(t_i,t_i)\geq d$ 筛选掉那些自身移位距离性质差的二进制序列，在相同的01含量下，一方面减少搜索空间；另一方面，提高了算法的效率；

(3)在基本保持01含量基本不变的情况下，适当扩展模板集合的搜索范围，可以在基本保持移位距离性质不变的情况下适当增加模板的数量。

参 考 文 献

[1] Adleman L. Molecular computation of solution to combinatorial problems [J]. *Science*, 1994, 266: 1021–1024.

[2] Garzon M, *et al.*. A new metric for DNA computing [C]. Proceedings of the 2nd Annual Genetic Programming Conference GP–97, Morgan Kaufmann, Stanford University, 1997: 472–487.

[3] Garzon M, Deaton R, Nino L F, Stevens S E , and Wittner M. Genome encoding for DNA computing [C]. The ThirdDIMACS Workshop on DNA-based Computing, Uinversity of Pennsylvania, 1997: 230–237.

[4] Baum E B. DNA sequences useful for computation [C]. Proc. Second Annual Meeting on DNABased Computers, American Mathematical Society, Princton University, 1996: 235–242.

[5] Feldkamp, *et al.*. A DNA sequence compile[C]. Proceedings of 6th DIMACS Workshop on DNA Based Computers, Netherlands, 2000: 253–257.

[6] Suyama A, *et al.*. DNA chips-integrated chemical circuits for DNA diagnosis and DNA computers [C]. Proc. 3rd International Micromachine Symp., Tokyo, 1997: 7–12.

[7] Morey J. Encoding Choices for Error Resistant DNA Computers[OL]. www.csd.uwo.ca /~morey /dnatalk/kevin/dna/dnaerror.html.

[8] Braich R, Johnson C, Rothemund P, and Adleman L. Solution of a satisfiability problem on a Gel-based DNA computer [C]. DNA 2000, 2001, LCNS 2054: 27–42.

[9] Frutos A, *et al.*. Demonstration of a word design strategy for DNA computing on surface [J]. *Nucleic Acids Research*, 1997, 25(23): 4748–4757.

[10] Liu Wenbin, Wang Shudong, Gao Lin，and Xu Jin. DNA sequence design based on template strategy [J]. *Chem. Info. Comput. Sci*, 2003, 43(6): 2014–2018.

[11] SantaLucia J, Allawi H, and Seneviratne P. Improved nearest-neighbor parameters for predicting DNA duplex stability [J]. *Biochemistry*, 1996, 35(11): 3555–3562.

[12] 刘文斌. DNA 计算中的编码问题及模型研究. [博士论文], 武汉：华中科技大学, 2004.1.
Liu Wen-bin. Research on the Enconding Problem and Algorithms of DNA Computing. [Dissertation(Doctor)], Huazhong University of Science & Technology, Wuhan, China, 2004.1.(in Chinese)

[13] Garzon M, Neathery P, and Deaton P. A new metric for DNA computing [C]. In Proc. of 2nd Annual Genetic Programming Conference, Morgan Kaufmann, 1997: 472–478.

刘文斌： 男，1969年生，博士后，副教授，主要从事DNA 计算、神经网络、遗传算法及计算生物学等方面的研究.

朱翔鸥： 男，1969年生，硕士，副教授，主要从事智能计算及生物信息学方向的研究.

王向红： 男，1969年生，教授，主要从事蛋白质结构预测、计算生物学等方面的研究.

张 强： 男，1976年生，博士后，教授，主要从事DNA 计算、神经网络、遗传算法等方面的研究.

马润年： 男，1963年生，博士后，教授，主要从事DNA 计算、神经网络、图与组合优化等方面的研究.