

# DNA 计算机原理、进展及难点(Ⅳ):论 DNA 计算机模型

许 进<sup>1),2)</sup> 谭钢军<sup>2)</sup> 范月科<sup>2)</sup> 郭养安<sup>3)</sup>

<sup>1)</sup>(北京大学信息科学技术学院 北京 100871)

<sup>2)</sup>(华中科技大学分子生物计算机研究所 武汉 430074)

<sup>3)</sup>(西安老年大学 西安 710014)

**摘 要** 在 DNA 计算机研究中,所建模型的好坏直接影响着 DNA 计算中诸多问题,如编码的难易程度、整个生物操作或生化反应的设计、解空间的大小、计算时间多少、应用范围以及通用性的程度等. 如何建立快速的、功能强的、具有一定通用性的 DNA 计算机模型,是从事 DNA 计算机研究者一直关注与感兴趣的难题. 为此,该文将主要围绕着 DNA 计算机的模型建立展开讨论,重点讨论 10 年来所建立起来的一些主要模型. 共分为三种类型:第一种是利用 DNA 分子结构与特性所建立起来的几种主要模型;第二种是利用生物操作方式所建立的三种模型:试管型、表面型与芯片型;第三种是所谓的 DNA 计算机模型. 文中讨论了这些模型的基本原理、功能、优缺点以及应用的研究进展等. 最后,对 DNA 计算机模型研究中的难点进行了分析,并给出了相应的解决思路.

**关键词** DNA 计算; DNA 计算机; 模型

**中图法分类号** TP301

## DNA Computer Principle, Advances and Difficulties (Ⅳ): On the Models of DNA Computer

XU Jin<sup>1),2)</sup> TAN Gang-Jun<sup>2)</sup> FAN Yue-Ke<sup>2)</sup> GUO Yang-An<sup>3)</sup>

<sup>1)</sup>(School of Electronic Engineering and Computer Science, Peking University, Beijing 100871)

<sup>2)</sup>(Institute of Biomolecular Computer, Huazhong University of Science and Technology, Wuhan 430074)

<sup>3)</sup>(Xi'an Agedness University, Xi'an 710014)

**Abstract** In the DNA computer research, the quality of the model set up in DNA computing have a effect directly on all it's aspects, such as the extent of difficulty of encoding, the complete design of bio-operation and biochemistry reaction, size of solution space and computing time. Especially, the models have a effect directly on the application scope of computer as well as extent of universality. How to establish a high speed and powerful DNA Computer model with a certain universality, it is a difficulty problem that computer researcher has been interesting. For this, this paper will make a discussion around the modelings of DNA Computer. It is important that discuss some models built in recent 10 years. They are divided into 3 types: The first one is several main models built by using DNA molecular structure and characteristics; The second one is to make use of the bio-operation method build up of three kinds of model: Tube type, the surface type and chip type; The third is the so-called DNA computer model build up, working properly; This paper discusses the basic principle, function, merit and shortcoming of these models and research progress of the applications, the difficutties of the DNA computer model research analyzes, and gives solving way correspondingly.

**Keywords** the DNA computing; DNA computer; model

收稿日期:2007-03-24. 本课题得到国家自然科学基金重点、面上资助项目(60533010,30670540)和国家“八六三”高技术研究发展计划项目基金(2006AA01Z104)资助. 许 进,男,1959年生,教授,博士生导师,主要研究领域为生物计算机、图论与系统工程、生物信息学等. E-mail: jxu@pku.edu.cn; jxu@mail.hust.edu.cn. 谭钢军,男,1962年生,博士研究生,教授,主要研究领域为生物计算机与系统工程. 范月科,男,1959年生,博士研究生,副教授,主要研究方向为 DNA 计算机、图论等. 郭养安,男,1960年生,副教授,主要研究方向为计算机科学与管理科学等.

# 1 引言

Adleman 博士 1994 年创立 DNA 计算模型<sup>[1]</sup>至今已有 10 多年了. 10 多年来, DNA 计算无论是在理论上, 应用模型的建立上, 还是实验与检测技术上都取得了惊人的进步. 在 DNA 计算机的研究中, 其模型研究是主要内容之一.

在目前所获得的 DNA 计算模型中, 最多引起学者们兴趣与关注的模型有如下几种: (1) 粘贴模型<sup>[2-9]</sup>; (2) 剪接系统模型<sup>[10-16]</sup>; (3) 表面与芯片 DNA 计算模型<sup>[17-25]</sup>. 其中表面与芯片 DNA 计算模型是 DNA 计算在实验技术上的模型, 而前两个模型则是依据 DNA 分子特性机理产生出的两个模型. 另外, 还有发夹状 DNA 计算模型<sup>[26]</sup>、质粒 DNA 计算模型<sup>[27-29]</sup>、基于  $k$ -臂 DNA 分子的计算模型<sup>[30-33]</sup>、自组装 DNA 模型<sup>[34-36]</sup>、插入-删除系统模型<sup>[37-39]</sup>、试管型的 DNA 计算模型<sup>[1, 40-46]</sup>、图灵机 DNA 计算机模型<sup>[10, 47]</sup>、布尔 DNA 计算机模型<sup>[48-52]</sup>等等. 备受学者关注的是将 DNA 计算的思想与模型应用于基因分析与疾病诊断与治疗相结合的研究中<sup>[24-53]</sup>.

在这篇文章中, 我们将讨论 DNA 计算机研究中的主要模型. 限于篇幅, 我们只就一些概念、原理、实验手段以及要点等给予讨论, 并且较全面地给出相关的文献. 下面, 先给出文中所需要的一些基本概念与记号.

字母表是一个有穷的符号集合, 本文中用  $\Sigma$  来表示. 在计算理论中最常用的是二进制字母表  $\{0, 1\}$ . 字母表上的字符串表示该字母表中的符号的有穷序列. 把一个没有符号的字符串称为空串, 记做  $e$ . 通常用  $u, v, w, x, y, z$  和希腊字母表示字符串. 字母表  $\Sigma$  上所有的字符串 (包括空串在内) 的集合记作  $\Sigma^*$ . 一个字符串的长度是指它作为序列的长度. 字符串  $w$  的长度用  $|w|$  表示. 显然  $|e| = 0$ . 我们把一个字符串  $w \in \Sigma^*$  也作为一个函数

$$w: \{1, 2, \dots, |w|\} \rightarrow \Sigma,$$

如  $w = baaccddrr$ , 有  $w(1) = b, w(2) = w(3) = a, w(4) = w(5) = c, w(6) = w(7) = d$ .

字符串  $x$  和字符串  $y$  的连接运算是指在字符串  $x$  的后面接上字符串  $y$ , 记做  $x \circ y$ , 或者简记为  $xy$ . 字符串  $v$  是字符串  $w$  的子串当且仅当存在字符串  $x, y$  使得  $w = xvy$ . 如果对于某个  $x$ , 有  $w = xv$ , 则称  $v$  是  $w$  的后缀; 如果对于某个  $y$ , 有  $w = vy$ , 则

称  $v$  是  $w$  的前缀. 对于每一个字符串  $w$  和每一个非负整数  $n$ , 字符串  $w^n$  的定义为

$$w^0 = e,$$

$$w^{n+1} = w^n \circ w, \quad \text{对于每个 } n \geq 0.$$

设  $\Sigma$  是一个字母表, 我们把  $\Sigma$  上的字符串集合, 即  $\Sigma^*$  的子集称为语言. 设  $A \in \Sigma^*$  是一个语言,  $A$  的补用  $\bar{A}$  来表示, 即  $\bar{A} = \Sigma^* - A$ . 设  $X, Y \in \Sigma^*$ , 即  $X, Y$  是  $\Sigma$  上的两个语言,  $X, Y$  的连接, 记作  $L = A \circ B$ , 定义为

$L = \{w \in \Sigma^* : w = x \circ y, \text{ 其中 } x \in X, y \in Y\}$ , 这就是语言的连接运算. 语言的另一个重要运算是所谓的语言  $L$  的 Kleene 星号, 记作  $L^*$ ,  $L^*$  是连接  $L$  中 0 个或者多个字符串得到的所有字符串的集合. 易知, 0 个字符串是  $e$ , 一个字符串是它本身, 而对一般的情况有

$$L^* = \{w \in \Sigma^* : w = w_1 \circ w_2 \circ \dots \circ w_k,$$

$$\text{其中 } k \geq 0 \text{ 且 } w_1, w_2, \dots, w_k \in L\}.$$

我们用  $G = (V, E)$  表示一个无向图, 其中  $V$  称为图  $G$  的顶点集,  $E$  称为图  $G$  的边集. 我们用  $D = (V, A)$  表示一个有向图, 其中  $V$  称为顶点集,  $A$  称为弧集. 文中所用到的有关 DNA 计算的概念可参见文献[54-55]. 文中所用到的有关形式语言中的术语可参见文献[56]. 有关图论方面的概念与记号可参见文献[57].

## 2 基于 DNA 分子结构特性的 DNA 计算模型

本节将主要讨论已经获得的基于 DNA 分子结构特性的几种 DNA 计算模型, 给出这些模型在图与组合优化等问题上的研究进展并指出存在的有关问题.

### 2.1 粘贴 DNA 计算模型

粘贴模型是 Roweis 等人在 1996 年提出来的一个 DNA 计算模型<sup>[3]</sup>, 该模型是当前 DNA 计算中备受学者关注的一个模型. 粘贴模型的创新点是采用单、双链混合型 DNA 分子进行编码, 这种编码显然可形成一个 2-进制数序列: 其中单链的编码用来表示 2-进制数中的“0”, 而双链用来表示 2-进制数中的“1”. 这样, 一个具有单双链的 DNA 分子的编码序列唯一地对应于一个 2-进制数 (见图 1). 由此可见粘贴 DNA 计算模型具有一定的通用性. 粘贴模型目前已有较多的应用, 诸如图的最大团与最大独立集问题、图的顶点着色问题、图的顶点覆盖问题、

子图问题、可满足性问题、图的同构问题等<sup>[2-9]</sup>，我们在文献[8-9]中对粘贴 DNA 计算模型从理论与应用两个方面进行了较为系统的研究. 故在此省略，有兴趣的读者可参见相关文献.

AAAAA|TTTCC|GGGGG|TTTTT|CCCCC

AAAAA|TTTCC|GGGGG|TTTTT|CCCCC

TTTTT                  CCCCC

AAAAA|TTTCC|GGGGG|TTTTT|CCCCC

TTTTT|AAAGG|                  |AAAAA|GGGGG

AAAAA|TTTCC|GGGGG|TTTTT|CCCCC

TTTTT|AAAGG|CCCCC|AAAAA|GGGGG

图 1 分别代表 4 个 2-进制数 00000、10100、11011、11111 的单、双混合型编码的 DNA 分子

2.2 剪接系统模型

剪接系统模型是当前 DNA 计算研究模型中的另一个主要模型. 顾名思义, 所谓“剪接”, 意指首先将 DNA 分子通过相关的“限制内切酶”从某个位置“剪开”, 然后将“剪开”后的 DNA 片断再重新组合, 用连接酶将它们重新连接起来. 通过研究剪接模型, 我们发现剪接模型很像代数学中的“群、环或者域”的概念. 所谓群是指一个非空集合中给出一种运算,

5'AAAAATCGATT TTT3'                  5'CCCCGCGCGGGGG3'

3'TTTT TAGCTAAAAA5'                  3'GGGGGCGCGCCCC5'

及两个限制内切酶: Taq I 和 SciN I, 它们识别的 DNA 序列是 T↓CGA 和 G↓CGC. 这两种酶分别

5'AAAAAT      CGATT TTT3'                  5'CCCCG      CGCGGGGG3'

3'TTTT TAGC      TAAAAA5'                  3'GGGGGCGC      GCCCC5'

注意到上面的第 1 个 DNA 片断的粘性末端与最后一个 DNA 片断的粘性末端是互补的, 第 2 个与第 3 个 DNA 片断也是互补的, 于是通过 DNA 连接酶,

这个运算必须满足诸如封闭性、结合律、单位元的存在, 以及每个元素的逆元存在等; 而环和域是指在一个非空集合上定义了两种运算, 并且这两种运算满足类似于群定义中的几条规则. 对于剪接模型而言, 也是对于一个非空的集合, 给出两种生物操作, 这两种生物操作必须满足一些必须的规则. 这里的非空集合是指含有 DNA 双链分子构成的集合, 而两种生物操作分别是所谓的切割和连接. 完成这两种生物操作主要采用核算内切限制酶和连接酶. 显然, 这两种生物操作是封闭的, 因为不管是切割后还是连接后所得到的还是 DNA 分子. 通过图 2, 我们可以比较清楚地看到剪接 DNA 计算模型的基本原理. 在图 2 中, 所用的双链 DNA 分子是:

5'GAATTCAAAACTTT3'

3'CTTAAGTTTGTGAA5'

所用的切割酶是 EcoR I. EcoR I 识别的 DNA 序列分别是 GA↓ATTC. 为了对剪接模型有一个较为深入的了解, 我们再举一例, 此例中有两个 DNA 分子和两个切割酶, 其中的两个 DNA 分子是:

作用在相应的两个 DNA 片断上, 形成如下 4 个新的 DNA 片断:

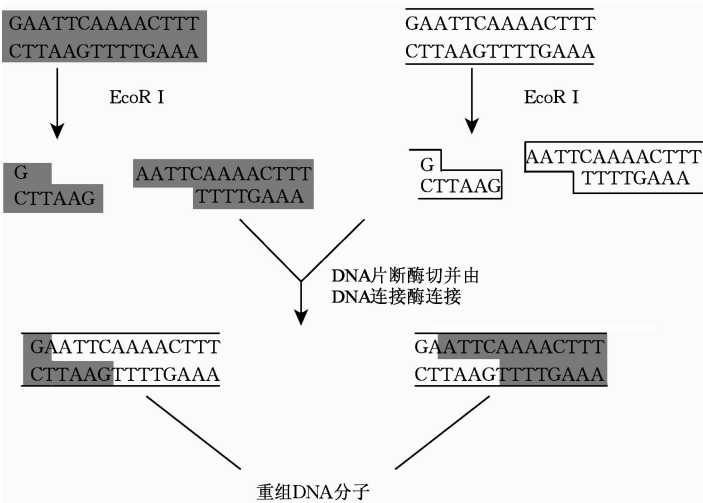


图 2 剪接模型的基本原理图

5'AAAAATCGCGGGGG3'	5'CCCCGCGATTTT3'
3'TTTTTAGCGCCCC5'	3'GGGGGCGCTAAAAA5'

Head 首先引入剪接系统的概念,此概念就是在基因工程中常用的两个基本操作:切割和连接,即在剪接模型的基础上,以形式语言为工具抽象出来的一种数学模型.在该模型中,(1)字符串不再是双链结构,但 A-T,C-G 的互补性这样的特性得以保留;(2)字符集合不限于 4 个字符;酶切序列的长度及同时作用的酶的数量不受限制<sup>[11]</sup>.这里,我们只给出 Head 引入的剪接系统中最重要的概念——剪接规则,此概念是模仿核酸内切限制酶和连接酶而引入的.

设  $V$  是一个有穷字母表,其中字符  $\#$ ,  $\$ \notin V$ .  $V$

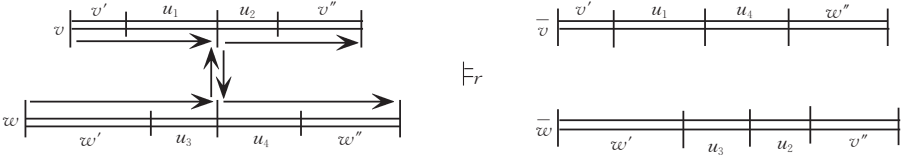


图 3 剪接运算

字母表  $V$  上的剪接系统是一个三元组  $S=(V, I, R)$ ,其中  $I \subseteq V^*$  是初始语言,  $R \subseteq V^* \# V^* \$ V^* \# V^*$  是剪接规则集.将剪接规则  $u_1 \# u_2 \$ u_3 \# u_4$  记为  $(u_1, u_2; u_3, u_4)$ .

1992 年,Head 又利用形式语言技术(Lindenmayer 系统)给出了生物分子计算的抽象模型——剪接模型<sup>[12]</sup>.Head 的剪接模型是将 DNA 链表示成一个有穷字母表的串,并将这些串的编辑操作集合指定为诸如切割和追加的 DNA 重组操作模型.从 Head 的工作之后,关于剪接系统研究的主要方向是在实验技术基础上的形式语言理论.

1995 年,Csuhaj-Varju, Freund, Kari 和 Paun 在夏威夷举行的第一次生物计算讨论会上做了题为“剪接系统的 DNA 计算”的报告<sup>[13]</sup>,1996 年,Paun 和 Salomaa 等从递归可数语言的特点出发,进一步论证了基于剪接操作上 DNA 计算的可能性<sup>[14]</sup>.1999 年,Freund, Kari 和 Paun 提出了剪接系统上的 DNA 计算,并证明了通用 DNA 计算机的存在性<sup>[15]</sup>.1999 年,Sakakibara 和 Ferrettiy 等提出了在拟-树结构上的剪接运算.该文提出了一种增强剪接系统能力的方法.文中在树上引入了剪接系统,树是作为部分退火单链而设计的,它是一个非常类似的概念,且是在串上的剪接系统的一种自然的扩充<sup>[16]</sup>.Benenson 等人在 2001 年所给出的一台可编程的有穷自动机<sup>[10]</sup>以及在 2004 年研究的“用于基

上的剪接规则是形为  $r=u_1 \# u_2 \$ u_3 \# u_4$  的词,其中  $u_1, u_2, u_3, u_4 \in V^*$ .对于剪接规则  $r=u_1 \# u_2 \$ u_3 \# u_4$  和词  $v, w, \bar{v}, \bar{w} \in V^*$ ,如果存在词  $v', v'', w', w'' \in V^*$  使得  $v=v'u_1u_2v'', w=w'u_3u_4w''$ ,那么可以将规则  $r$  应用到  $v, w$  上,并产生两个新词  $\bar{v}=v'u_1u_4w''$  和  $\bar{w}=w'u_3u_2v''$ ,将这种剪接过程记为  $(v, w) \vdash_r (\bar{v}, \bar{w})$ ,词  $v, w$  称为剪接项.在已知上下文的情况下可以省略  $r$ ,用  $\vdash$  代替  $\vdash_r$ .通过剪接规则  $r$ ,从  $v, w$  到  $\bar{v}, \bar{w}$  的传递过程如图 3 所示.我们将系统中不能进行进一步剪接运算的词称为垃圾词.

因表达逻辑控制的自律式分子计算机”模型<sup>[24]</sup>,均含有剪接 DNA 计算模型中的思想.这充分说明了剪接系统是计算完备的,也就是说每个 PASCAL 程序都可以通过合适的剪接系统来模拟,反之亦然.另外,他们还证明了通用剪接系统的存在性,即剪接操作技术上的通用可编程的 DNA 计算机的存在性<sup>[10]</sup>.

2.3 发夹 DNA 计算模型

DNA 计算中最主要最核心的反应是杂交反应.在 DNA 计算的有关生物操作上,诸如提取、PCR 扩增、DNA 重组以及检测等,其可靠性都依赖于 DNA 分子间的特异性杂交.因此,杂交反应的效率和精度都直接影响到 DNA 计算的效率和最终输出结果的可靠性.通常而言,DNA 分子间的杂交中不希望发夹 DNA 构型发生,所以,在 DNA 计算的编码研究中,设计要尽可能避免这种情况的发生.

然而,日本学者 Sakatomo 等人却巧妙地应用 DNA 分子的这种发夹结构建立了一种新颖的、用于求解一个可满足性问题的 DNA 计算模型<sup>[26]</sup>.为了阅读方便,这里先给出几个必要的概念与记号.设  $A=\{a_1, a_2, \dots\}$  是一组 Boole 变量的集合.一个子句是一个公式:

$$C=b_1 \vee b_2 \vee \dots \vee b_k,$$

其中符号  $\vee$  表示逻辑“或”,每个  $b_i$  是  $A$  中的一个变量或者  $A$  中变量的补.  $A$  中的元素  $a$  的补用符号  $\bar{a}$

来表示. 一个逻辑式是指由若干个子句的“与”构成的式子:

$$\alpha = C_1 \wedge C_2 \wedge \cdots \wedge C_r,$$

其中每个  $C_i$  是一个子句, 而符号  $\wedge$  表示逻辑“与”.

可满足性问题是要求求出满足逻辑式  $\alpha$  中满足  $\alpha=1$  的解的集合. 这个问题显然是一个 NP-完全问题.

由 Sakatomo 等人提出的基于 DNA 分子发夹结构的 DNA 计算模型的基本思想是: 通过编码技术, 使得非可行解在生化反应过程中形成发夹构型, 而对于发夹构型, 在每个编码的 DNA 分子中设计了内切酶 BstN I 的识别位 CCAGG, 以便于清除这些发夹构型. 为方便, 我们把 Sakatomo 等人这种以发夹构型的 DNA 分子为非可行解方法设计的 DNA 计算模型称为发夹 DNA 计算模型. Sakatomo 等人所建立的发夹 DNA 计算模型的计算方法和步骤如下.

1. 按照给定的 Boole 公式产生字符串, 其具体的方法是: 首先将 Boole 公式转化成字符串. 例如, 对于公式  $(a \vee b \vee c) \wedge (\bar{a} \vee c \vee \bar{d})$ , 它可转化成 9 个字符串:  $a\bar{a}, ac, a\bar{d}, b\bar{a}, bc, b\bar{d}, c\bar{a}, c\bar{c}, c\bar{d}$ . 然后通用连接反应将这些字符串连接起来.
2. 通过 ssDNA 分子, 每个代表一个字符串来形成发夹 DNA 分子结构; 这一步骤仅仅通过调解温度来执行计算逻辑, 甚至不需要生物酶;
3. 删除形成发夹状 DNA 分子. 保留的分子代表了可满足的字符串, 这些字符串代表了问题的解.

文献[4]对下列 Boole 公式通过上述方法步骤进行了生化仿真操作:

$$\begin{aligned} F = & (a \vee b \vee \bar{c}) \wedge (a \vee c \vee d) \wedge (a \vee \bar{c} \vee \bar{d}) \wedge \\ & (\bar{a} \vee \bar{c} \vee d) \wedge (a \vee \bar{c} \vee e) \wedge (a \vee d \vee \bar{f}) \wedge \\ & (\bar{a} \vee c \vee d) \wedge (a \vee c \vee \bar{d}) \wedge (\bar{a} \vee \bar{c} \vee \bar{d}) \wedge \\ & (\bar{a} \vee c \vee \bar{d}). \end{aligned}$$

到目前为止, 我们所看到的关于发夹 DNA 计算模型方面的研究很少. 但这并不意味发夹 DNA 计算用途不大. 我们以为, 发夹 DNA 计算模型是一个在 DNA 计算研究中很有前途的研究方向, 若与诸如粘贴模型、剪接系统模型等其它模型相结合, 可使得 DNA 计算模型得到进一步完善和实用.

我们利用 Sakatomo 等人思想, 给出了用于图的顶点着色问题的发夹 DNA 计算模型, 其基本思想是: 若两个相邻顶点着有相同的颜色, 则必形成 DNA 分子的发夹构型, 并在每个发夹构型的 DNA 分子上设计了内切酶的识别位点, 然后通过杂交反应后清除这些非解.

2.4 质粒 DNA 计算模型

基因工程是当前人类科学研究领域内核心的领域之一, 而质粒是基因工程中主要的一种载体, 无论从理论研究上还是生物操作的技术上, 关于质粒 DNA 分子的研究都必有很大的投入与发展, 因而, 也会对基于质粒 DNA 分子的所谓的质粒 DNA 计算模型的研究起积极的促进作用. 质粒 DNA 计算模型是由 Head 等人于 2000 年提出来的<sup>[27]</sup>. 这种模型是利用质粒 DNA 分子的环状结构进行 DNA 计算的. 在双链环状的质粒 DNA 分子上, 含有 DNA 分子复制点, 还含有一些别的基因. 这些基因通常

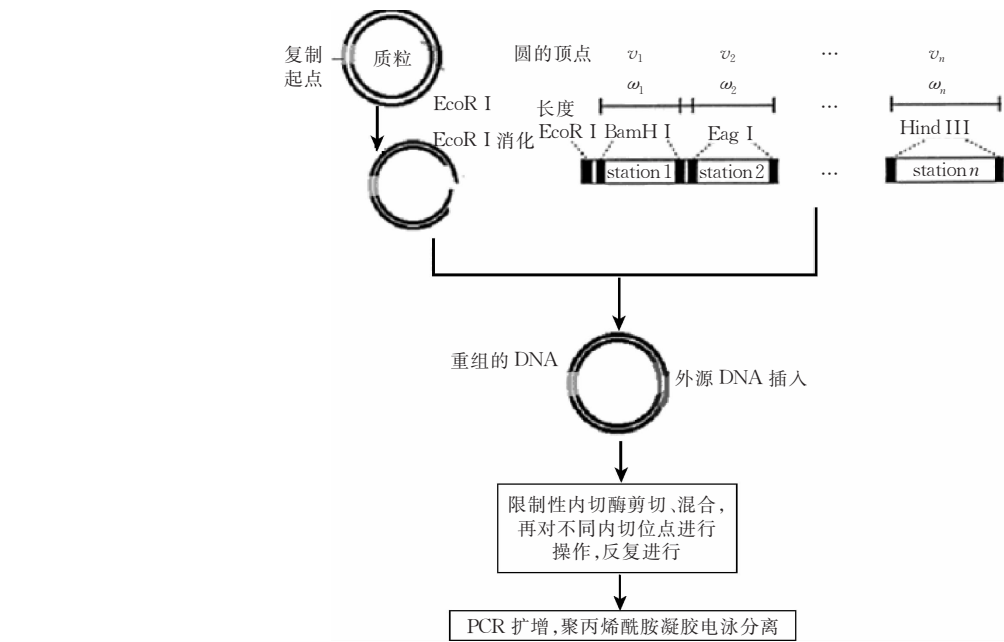


图 4 质粒 DNA 计算原理图示, 其中以图的独立集问题为例



是:①产生抗菌素;②抵抗抗菌素;③降解有机化合物;④产生大肠菌素;⑤产生内毒素;产生限制或修饰性的酶等。

正是由于质粒 DNA 分子的上述特性,才使得 DNA 计算得以实施.其质粒 DNA 计算模型充分利用质粒 DNA 分子上的基因位,通过核酸内切酶与连接酶的作用,可产生两种状态,分别用 0 和 1 来表示这两种状态,这样,每个质粒体与传统计算机中  $k$ -位数据寄存器作用相同.这就是质粒 DNA 计算的基本思想.我们在这里给出了质粒 DNA 计算的一般原理图示,如图 4.

Head 本人建立了基于质粒 DNA 计算模型的图的最大独立集问题的计算模型<sup>[27]</sup>,并且抽象成一般的质粒 DNA 计算的概念如下:设  $P$  是一个质粒体,正整数  $k$  和  $s_1, s_2, \dots, s_k$  是  $P$  的  $k$  个不重叠的子段,即假设对每个  $i, 1 \leq i \leq k$ ,核苷酸序列  $s_i$  在质粒体  $P$  的其它位置不出现.我们把这样选择的子段称为质粒体的位.整个计算从含有大量且相同的  $k$ -位质粒体试管开始.为了具有可读性,计算中质粒按下述方法修正.其修正仅在位上进行.在计算的任何时刻,位  $s_i$  处于两种状态之一,分别用 0 和 1 来表示这两种状态.这样,每个质粒与传统计算机中的  $k$ -位数据存储器作用相同.质粒 DNA 计算中的存储器

是带有质粒溶质的水溶液.在整个计算中,质粒 DNA 分子均保持双链结构.

近几年我们也相继建立了基于质粒 DNA 计算的其它几种模型,包括图的最大匹配问题的质粒 DNA 计算模型<sup>[28]</sup>、最大权图问题的质粒 DNA 计算模型<sup>[29]</sup>以及背包问题的计算模型等.

## 2.5 $k$ -臂 DNA 计算模型

$k$ -臂 DNA 计算模型是 1999 年由 Jonoska 等人提出的一种 DNA 计算模型<sup>[30]</sup>.他们在原文中称之为“三维图 DNA 结构”,并用于求解诸如 3-可满足性和图的顶点 3-着色等问题.而三维 DNA 结构实际上是由所谓的  $k$ -臂 DNA 分子构成,故我们把这种模型称为  $k$ -臂 DNA 计算模型.

Jonoska 等人提出的  $k$ -臂 DNA 计算模型的生物机理是:3-臂和 4-臂 DNA 分子是稳定的.这项研究成果是由 Seeman 等人获得的<sup>[31-32]</sup>.在图 5 中,我们给出了 2-臂、3-臂、4-臂等 DNA 分子.这里需要说明三点:(1)类似于  $k$ -臂的 DNA 分子在自然界已经存在,如 Holliday 中间体;(2)Seemans 等人的研究成果表明:3-臂和 4-臂 DNA 分子是相当稳定的;(3) $k$ -臂 DNA 分子的 3'端一般为单链的延伸,且延伸部分一般为 30bp~45bp.

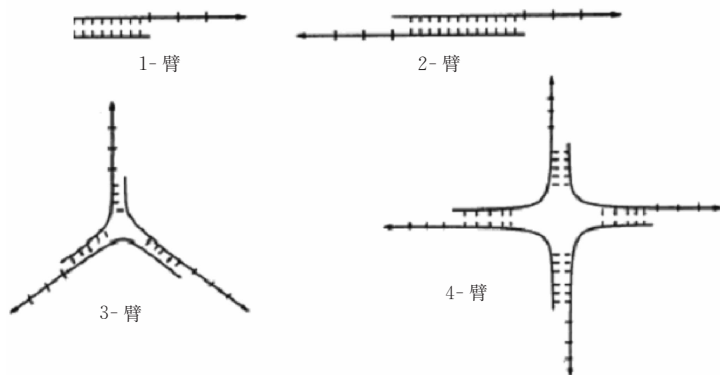


图 5 1~4-臂 DNA 分子

从图 5 中自然得出结论:通过连接酶,可以用  $k$  ( $=1, 2, 3, 4$ )-臂 DNA 分子构造出任意顶点的最大度 $\leq 4$ 的图来. Jonoska 等人也正是利用这一思想,首先将一个给定的可满足性问题转换成一个完全等价的接触网络图,而此网络图的最大度是 4. 如逻辑式  $\alpha = (x+y) \cdot (\bar{x}+y+z) \cdot (\bar{x}+y+\bar{z})$  转化成接触网络图,如图 6 所示.然后通过核酸内切限制酶与 PCR 扩增给出可满足性问题的解.文献<sup>[33]</sup>以及我们其他的文章中分别建立了基于  $k$ -臂 DNA 计算的图的同构问题和求解图的连通度的 DNA 计算模型.

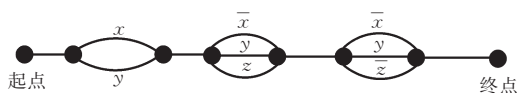


图 6 逻辑式  $\alpha$  所对应的接触网络图

遗憾的是,上述结果仅仅是理论上的,没有给出相应的生化实验.需要指出, $k$ -臂 DNA 计算模型实际上是一种图信息处理的 DNA 计算机模型,但仅在图的顶点度数小于或等于 4 的条件下可以直接应用,但对于图的顶点度数大于 4 时,则不能直接应用,对于有些问题,可能通过一些等价转换来应用.

Asiewicz 和 Lee 等则提出了利用多臂 DNA 分子来实现谓词逻辑中的定理证明的 DNA 计算模型<sup>[58-59]</sup>.

2.6 插入-删除系统

插入-删除系统是 Kari 和 Thierrin 在 1996 引入的一种建立在上下文删除基础上的 DNA 计算模型. 其生物机理源于在一个单链的 DNA 序列中, 通过生物酶以及加热与退火的等生物操作技术, 插入一段 DNA 片断, 或删除一段单链 DNA 片断. 下面, 我们先对插入和删除的生物操作方法步骤作简要讨论.

**插入方法.** 设在一个试管  $N$  里含有一个 DNA 序列, 它由 5 个片断  $x_1, u, v, x_2$  及  $z$  构成 (如图 7(a)所示). 欲在该序列中插入一个新的 DNA 片断  $y$ , 其方法步骤如下:

1. 给试管  $N$  里加入单链 DNA 序列  $3'\overline{u}\overline{y}\overline{v}5'$ , 如图 7(a)所示), 其中  $\overline{x}$  表示  $x$  的补链;
2. 对试管  $N$  施行退火操作, 则  $\overline{u}$  与  $u$ ,  $\overline{v}$  与  $v$  结合, 并且  $\overline{y}$  被折叠, 如图 7(b)所示的 DNA 结构;
3. 使用核算内切限制酶, 将试管  $N$  中的双链切断, 得到如图 7(c)所示的 DNA 分子结构;
4. 在试管  $N$  中加入 DNA 片断  $\overline{z}$  和有关的聚合酶,  $\overline{z}$  是引物. 施行 PCR 扩增, 得到如图 7(d)所示的完整的双链 DNA 分子;
5. 施行解链, 得到两条单链, 其中一条是  $x_1 u y v x_2 z$ , 即已经将 DNA 片断  $y$  插入在  $u$  与  $v$  之间.

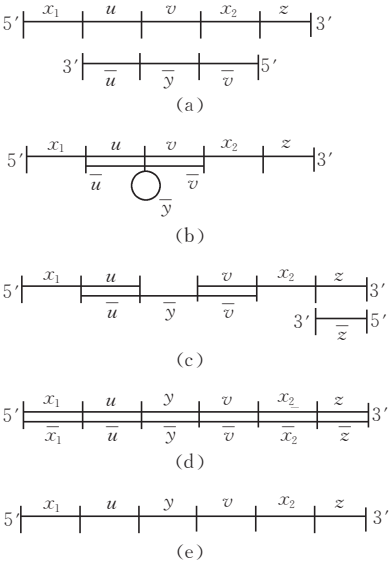


图 7 插入生物操作过程图示

**删除方法.** 设在一个试管  $N$  里含有一个 DNA 序列, 它由 6 个片断  $x_1, u, y, v, x_2$  及  $z$  构成 (如图 8(a)所示). 欲在该序列中删除其中的 DNA 片断  $y$ , 其方法步骤如下:

1. 给试管  $N$  里加入单链 DNA 序列  $3'\overline{u}\overline{v}5'$ , 如图 8(a)所示);
2. 对试管  $N$  施行退火操作, 则  $\overline{u}$  与  $u$ ,  $\overline{v}$  与  $v$  结合, 并且  $\overline{y}$  被折叠, 如图 8(b)所示的 DNA 结构;
3. 在试管  $N$  中加入 DNA 片断  $\overline{z}$  和有关的聚合酶,  $\overline{z}$  是引物. 施行 PCR 扩增, 得到如图 8(c)所示的完整的双链 DNA 分子;
4. 施行解链, 得到两条单链, 其中一条是  $x_1 u v x_2 z$ , 即已经将 DNA 片断  $y$  在  $u$  与  $v$  之间删除.

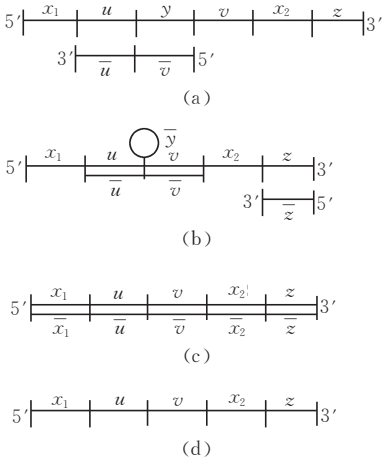


图 8 删除生物操作过程图示

插入-删除系统就是根据上述插入和删除的生物操作的机理以及形式语言中的有关上下文无关等理论而建立的, 其中上下文插入和上下文删除都可以在实验室用分子生物标准技术来实现.

确切地讲, 插入-删除系统是一个五元族  $ID = (\Sigma, T, A, I, D)$ , 其中  $\Sigma$  是一个字母表,  $T \subseteq \Sigma$  是  $ID$  的终结字符集,  $A \subseteq \Sigma^*$  是公理集,  $I \subseteq \Sigma^* \times \Sigma^* \times \Sigma^*$  是插入规则集,  $D \subseteq \Sigma^* \times \Sigma^* \times \Sigma^*$  是删除规则集.

1998 年, Paun 利用上下文插入文法刻画了递归可列语言<sup>[37]</sup>, 1999 年, Kari, Paun, Thierrin 和 Yu 等利用建立在插入-删除系统基础上的递归可数语言特征, 研究了形式语言和 DNA 计算的交叉性. 他们证明了单个字母的插入-删除系统是计算完备的, 并且存在通用的插入-删除系统<sup>[38]</sup>. 2002 年, 在日本札幌市举行的第 8 届国际 DNA 计算机会议上, Takahara 阐述了插入-删除系统的计算能力<sup>[39]</sup>.

3 基于生物操作与实现的 DNA 计算模型

从生物操作的方式, 可将 DNA 分为试管型、表面型和芯片型. 在这一节里, 我们将简要介绍这三种 DNA 计算模型的研究概况.

### 3.1 试管型 DNA 计算模型

试管型 DNA 计算模型是从实验角度被称作 DNA 计算模型类的. DNA 计算机的研制过程可分为三个阶段:可行性论证阶段、提高速度阶段和实用性阶段. 这三个不同的阶段,反映在实验手段上应为试管阶段、表面实验阶段和芯片阶段. 试管阶段的主要任务是研究 DNA 计算基本原理的可行性问题;表面阶段的主要任务是在试管阶段的基础上,进一步想方设法提高 DNA 计算的速度,使得 DNA 计算逐步走向实用化和芯片阶段(这里的芯片阶段是一个广义的概念);芯片阶段是 DNA 计算机研制的最终阶段,芯片阶段标志着 DNA 计算机研制趋于成熟. 目前在国际上关于前两个阶段的研究已经展开. 我们把试管阶段,即在试管里 DNA 分子和有关生物酶等进行生化反应的 DNA 计算,称为试管型 DNA 计算模型. 试管阶段始于 Adleman 关于 DNA 计算首篇文章<sup>[1]</sup>. 此文给出了用于求解有向 Hamilton 路问题的 DNA 计算模型. 其后,建立图与组合优化问题的试管型的 DNA 计算模型很多,这里仅给出其中的一些主要结果. Adleman 等人给出了自组装问题的试管型 DNA 计算模型<sup>[34-35]</sup>;Lipton 在 1995 年给出了可满足性问题 DNA 计算模型<sup>[40]</sup>;Ouyang 等人在 1997 年给出了基于试管模型的图的最多团与最大独立集问题的 DNA 计算模型<sup>[42]</sup>;我们在文献[43-44]中给出了图的着色问题两种不同的试管型 DNA 计算模型;在文献[45]中首次给出了中国邮递员问题的试管型 DNA 计算模型;在文献[46]中首次给出了 0-1 规划问题的试管型的 DNA 计算机模型等.

### 3.2 表面与芯片 DNA 计算模型

表面与芯片 DNA 计算模型是近几年备受学者们关注的 DNA 计算模型. 试管型的 DNA 计算有两个很大的缺点:(1)DNA 分子漂浮在溶液中给样品操作带来困难;(2)在某些生物操作过程中容易丢失 DNA 链. 这个缺点也是人们提出表面计算的一种原因. 在表面 DNA 计算模型中,DNA 链被固定在硅片或玻璃等经过严格化学处理后的表面上,而不是将 DNA 分子漂浮在溶液中. 表面计算简化了对 DNA 链的操作,并减少了 DNA 链的丢失,使 DNA 计算朝芯片型和实用型方向发展. 我们在此以文献[3]中介绍的一种基于表面的 SAT 问题的解决方法为例,来说明表面计算的基本方法和步骤. 该文解决的 SAT 问题如下:

$$F = (w \vee x \vee y) \wedge (w \vee \bar{y} \vee z) \wedge (\bar{x} \vee y) \wedge (\bar{w} \vee \bar{y}),$$

对于上面的 4 个变量  $w, x, y, z$  共有  $2^4$  个组合,文中用下面的 DNA 分子形式编码了每一个组合情况:



$\text{T}_{15}$  序列是处于载体表面和要进行杂交的 DNA 分子之间的一段序列,它是用来起间隔作用的;GCT-TvvvvvvvvTTCG 称作一个“字”,其中两边的 4 个大写的碱基称作“字标”,中间的 8 个 v 代表信息位. 表面计算主要是利用核酸外切酶 E. coli 能够特异性地识别 DNA 单链分子并将其水解,而对 DNA 双链分子却不起作用的性质. 其最基本的操作步骤如下:

(1) 合成:按照如上的形式合成代表解空间的 DNA 分子(关于 DNA 分子的合成方法与技术可参见文献[60]);

(2) 固定:将合成好的 DNA 分子固定在经过化学处理的载体表面上;

(3) 标记:放入满足第一个子句分子的补序列,使其与表面上的 DNA 分子杂交;

(4) 破坏:加入核酸外切酶 E. coli 水解没有杂交的单链分子并清洗表面;

(5) 去标记:加入一种碱性溶液使 DNA 分子变性并清洗掉解链的 DNA 分子;

(6) 读结果:对剩下的 3 个子句重复进行步骤(3)(4)(5),最后在表面上剩下 DNA 分子即为问题的解,经过 PCR 扩增后读出结果(参见图 9 所示).

表面 DNA 计算模型的研究始于 1997 年,这方面的主要创始人是 Frutos, Smith 等人<sup>[19-20]</sup>. 目前表面 DNA 计算模型主要应用于下列几个图与组合优化中的模型:多人给出了基于表面计算的可满足性问题的 DNA 计算模型<sup>[17,19-20]</sup>;我们在文献[21]中给出了基于表面计算的图的最大匹配问题的 DNA 计算模型;在文献[22]中给出了基于表面计算的图的最大团问题的计算模型;在文献[23]中建立了基于表面计算的 0-1 规划问题的 DNA 计算模型,其中文献[18]是文献[23]的进一步的完善.

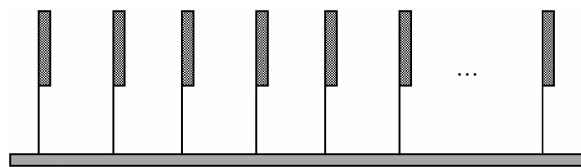


图 9 表面 DNA 分子排列示意图(阴影部分表示进行杂交反应的“字”)

什么是芯片型 DNA 计算机?目前尚未有人给出确切的定义,原因是当前 DNA 计算机的研究基



本上处于“实验室”阶段,还不可能应用于市场,因而,没有人花费较多的时间来考虑研制所谓的“全自动化的 DNA 计算机”.所谓芯片型的 DNA 计算机,应该是 IT 与 BT 相结合的“产物”,即利用现有的电子计算机,或者电子信息处理中尽可能用到的工具,特别是在输入、输出、信息转换等方面利用电子计算机技术来实现.其中的“芯片”概念实际上是借用电子计算机中的芯片概念.芯片型 DNA 计算机中的“芯片”是一个较为广义的概念,不仅仅是限于“生物芯片”,在芯片上可能集成了诸如“DNA 分子的合成、生化操作、生化反应、荧光标记库、由 DNA 分子形成的解空间库、解的检测、PCR 扩增、甚至 DNA 序列的次序与阶段及存储库”等.

## 4 DNA 计算机模型

本节将给出业已建立的 DNA 计算机模型,或者接近的 DNA 计算机模型,这些模型包括图灵 DNA 计算机、布尔 DNA 计算机、自组装 DNA 计算机以及用于基因分析的 DNA 计算机模型等.

### 4.1 图灵 DNA 计算机模型

图灵机型 DNA 计算机模型是 Rothmund 在 1995 年通过形式语言的形式引入的. Rothmund 称该模型为 DNA 分子图灵机<sup>[47]</sup>,或简称为图灵 DNA 计算机.我们在本文中称为图灵机型 DNA 计算机模型,或简称为图灵 DNA 计算机模型.图灵 DNA 计算机的研究在于设计一种通用的分子计算装置.图灵 DNA 计算机定义了一个分子计算系统,该计算系统总处于一定的状态,具有一定的存储能力和执行无限数量的状态转换功能.其定义如下:

一个具有确定的图灵 DNA 计算机定义为  $M = (Q, \Sigma, \delta, q_0, q_1, q_2)$ , 其中:

- (1)  $\Sigma$  是有穷字母表;
- (2)  $Q$  是有穷状态集;
- (3)  $q_0 \in Q$  为初始状态;
- (4)  $q_1 \in Q$  为接受状态;
- (5)  $q_2 \in Q$  为拒绝状态,这里  $q_2 \neq q_1$ ; 接受和拒绝状态都是停机状态;
- (6)  $\delta$  是转移函数,是从  $Q \times \Sigma$  到  $Q \times \Sigma \times \{L, 0, R\}$  的映射,但对某些变元,允许  $\delta$  没有定义.

图灵 DNA 计算机的计算过程可用一系列连续的格局来表示,每一个格局代表一步计算变换.设格局  $C = x_1 \cdots x_{i-1} q x_i x_{i+1} \cdots x_n$ , 其中,  $x_1 \cdots x_n$  表示带上的内容,  $q$  是当前状态,  $i$  是读写头的位置.

虽然 Rothmund 对图灵 DNA 计算机进行了初步的研究,但遗憾的是, Rothmund 所提出的 DNA 计算机模型并没有在分子水平上实现自动计算装置. 2001 年,由 Shapiro 教授领导的 DNA 计算机研究组对 Rothmund 提出的图灵机型 DNA 计算机模型进行了深化和实现.他们提出了一种可编程的有穷自动机,该有穷自动机是由 DNA 分子和在 DNA 分子上进行各种操作的生物酶构成.该自动机的硬件是由限制性核酸酶和连接酶以及 ATP 等组成的,软件和输入用双链 DNA 来编码,而程序则是通过选择适当的软件分子来构成.一旦在溶液中混合了各个组成部分,自动机则通过一系列的限制性内切酶的切割、杂交和连接对输入分子进行处理,产生可检测的输出分子,这些分子对自动机的最后状态进行了编码,从而给出计算问题的结果.在他们的实现过程中,共有  $10^{12}$  个自动机共享相同的软件,并独立并行地运行<sup>[10]</sup>.

该自动机有两个内部状态 ( $s_0$  和  $s_1$ ) 和一个字母表,字母表是由输入符号  $a$  和  $b$  组成.该自动机有 8 个可能的转化规则.程序是选择某些转化规则并确定那些内部状态是可接受的.可选择的传递规则共有 255 个,可选择的接受状态有 3 个(要么  $s_0$ , 要么  $s_1$  或两个都选择),这样就有 765 个语法上不同的程序.

### 4.2 布尔 DNA 计算机模型

布尔 DNA 计算机模型是利用 DNA 分子的特性来模拟布尔运算(或逻辑电路)的一种 DNA 计算机模型.传统的电子计算机实际上是由大量简单的逻辑门电路集成而构成的,并在此基础上由逻辑电路实现各种计算.为了探索 DNA 计算新的潜在的应用,1996 年, Ogihara 等首次提出了基于 DNA 的模拟布尔电路<sup>[48]</sup>(Ogihara 在 1996 年已经完成该文,并在他们学校进行了报告,在 1999 年正式发表).

布尔变量集可视为  $\{0, 1\}$ , 操作在该集合上基本运算有或(OR)、与(AND)、非(NOT)这三种.另外,还有在这三种基本运算基础上几个常用的运算,如 XOR 运算、NAND 运算等. Ogihara 等提出的布尔 DNA 计算机模型中设计了两种编码器,一种称为操作符编码器,另一种称作操作数编码器,其中,每一个操作数及操作运算符均编码成单链 DNA 分子.设计这两种编码器的目的在于避免 DNA 链在杂交过程中发生错配和发夹状的发生.其中设计进行运算的 DNA 链在结构上由三个部分构成:操作运算符编码、操作数的编码、操作运算符编码的回文,同

时在组成编码的碱基上加以限制,要求组成操作运算符的编码只能为 C、G 两种碱基,另外两种碱基 A、T 组成操作数的编码。

随后,Amos 等提出了一种改进的逻辑与非门的 DNA 模型<sup>[49-50]</sup>。Mulawka 提出了一个基于表面方式的逻辑与非门的 DNA 计算模型<sup>[51]</sup>。Wasiewicz 提出了数据流逻辑操作实现的 DNA 计算模型<sup>[52]</sup>。上面几种模型的缺点是在经过一次运算后有 DNA 分子编码的逻辑门也随着被限制性内切酶水解,因此不能重复使用。这里需要指出的是,布尔 DNA 计算机模型目前主要是在理论上进行探索,而在生化实验与操作上的研究较少。我们以为,布尔 DNA 计算机模型是 DNA 计算中一个具有发展潜力的方向,随着 DNA 合成技术以及生物芯片技术的发展,必将促进这一研究领域的进一步发展。

### 4.3 自组装计算模型

自组装计算模型是由 Winfree 等首先提出来的。一种 DNA 计算模型<sup>[36]</sup>,它类似于细胞自动机(Cellular Automata),通过 DNA 分子键的相互作用形成特定的构型来完成计算过程<sup>[34-35]</sup>。DNA 自组装是对分子量级结构进行构造的一种方法。在这个方法中,人工合成单链 DNA 成为 DNA 杂交分子(“瓦片”)。这些 DNA 瓦片有粘性末端,它可优先与其它 DNA 瓦片的粘性末端匹配,促使进一步组装成瓦格。

### 4.4 基因表达分析的 DNA 计算机模型

基因表达分析的 DNA 计算机模型是日本学者 Suyama 和 Nishida 等于 2001 年提出来的<sup>[24]</sup>。他们将 DNA 芯片、DNA 编码数(DNA Code Number)及 DNA 计算技术结合起来,设计了一种智能化的基因表达分析方法。该方法成功地完成了有一个有 7 个小鼠表达基因的分析实例<sup>[25]</sup>。此外,他们结合在生物芯片中日趋发展成熟的微流控制技术(将微泵、微阀门、检测装置、加热器及微型毛细管通等集成在一块不大的半导体芯片上),提出了一种 Hamilton 有向路问题计算芯片的原理图<sup>[24]</sup>。

2004 年,由 Shapiro 教授领导的 DNA 计算机研究组提出了一种新颖的“用于基因表达逻辑控制的自律式分子计算机”模型<sup>[53]</sup>。该模型由三个可编程模块组成:一个计算模块,由一个随机分子自动机构成;一个输入模块,通过特定 mRNA 水平或点突变调整软设备分子浓度,从而实现自动机转换可能性;一个输出模块,可控制对一个短单链 DNA 分子进行释放操作。该计算机模型可应用于试管生化传感器、基因工程甚至医学诊断和治疗。作为实例,文

章中报道了他们应用所设计的计算机用于诊疗小细胞肺癌和前列腺癌这两种疾病的具体模型。

## 5 结论与展望

通过上面的讨论我们已经看到,不管是利用 DNA 分子特性所建立的 DNA 计算模型,还是通过利用生物实验或生化操作上所建立的 DNA 计算模型,还是目前从理论上所建立的 DNA 计算机模型,均已取得了相当丰富的成果。然而,与科学家拟研制的 DNA 计算机相比,还有相当大的差距。

(1) 目前所建立的每一种 DNA 计算模型只能解决一类,或者极少数的一些问题(主要是 NP-完全问题),没有像电子计算机那样的统一处理问题的模型;

(2) 虽然已经建立了不少的基于 DNA 计算的图与组合优化问题上的计算模型,但目前建立的求解图问题的 DNA 计算模型仍很贫乏,已获的 DNA 计算模型还不能解决众多图与组合优化中的 NP-完全问题,更谈不上给出通用型的 DNA 计算机模型;

(3) 在已经建立的 DNA 计算模型中,大多数问题存在着所谓的“解空间指数爆炸问题”,即随着问题规模的增大,构成问题的解空间成指数阶地增大,也就是说,所需要的 DNA 分子数量成指数阶地剧增!如文献<sup>[61]</sup>等。

(4) 从 Adleman 建立的 DNA 计算模型至今,其计算步骤主要分为两步:第一步,通过生化杂交或者生物操作技术,产生问题的所有可能的解,即问题的解空间,这一步是 DNA 计算的“特长”,易于实现;第二步,从解空间中寻找出问题所需要的解,即就是 DNA 计算中解的检测问题。这一步是影响 DNA 计算发展的最困难的一个问题。目前主要采用一些常规的生化操作方法,诸如各种电泳技术、PCR 扩增、荧光标记(如我们也研究了采用分子信标检测技术等)、生物酶方法、生物素标记、磁珠分离、层析技术、Acrydite<sup>TM</sup>等方法。这些方法确实为 DNA 计算机的发展起到了一定的促进作用,但与实用性的 DNA 计算机相差很远。

综上所述,目前在 DNA 计算机模型的研究上存在的重点问题是:通用性差、指数爆炸问题、解的检测问题。如何在未来的 DNA 计算机研究中克服这些困难?我们以为,应主要从下面两个方面来考虑:

(1) 编码。在所有的 DNA 计算中,首当其冲的

问题是编码问题. DNA 序列的编码实际上是数学上的一个很困难的组合优化问题. 在 DNA 计算中, 作为“数据”的 DNA 分子不能随机地产生, 原因是由于诸如氢键引力的作用有可能导致不希望出现的发夹构型的 DNA 分子; 还有问题的规模与 DNA 序列长度的选择问题等. 所以, 采用什么样的编码(单链的? 双链的? 环状的? 发夹状的? 混合型的?) 是 DNA 计算中的一个基本问题. 进而要考虑的是编码的长度问题, 若过长, 则解空间“膨胀”, 不仅产生巨大的资源浪费, 而且给生物操作, 或者生化反应带来不必要的麻烦, 使问题求解变得复杂, 甚至难于求解. 因此, 如何根据具体的待解问题, 给出尽可能优化的编码是 DNA 计算的另一个基本问题.

在克服“解空间指数爆炸问题”方面, 应与编码问题结合处理. 其基本思想是: 在上述关于编码的两个基本问题的基础上, 将编码与生物酶、或者与带有“酶切位点”的发夹 DNA 分子构型、或者与分子信标等结合起来进行解决, 使得问题的“解”、或者“非解”对应生物酶的酶切位点, 或者发夹分子 DNA 分子构型, 或者分子信标等.

在处理 DNA 计算中关于解的检测问题上与解决“解空间指挥所爆炸问题”上有类似的思想.

(2) 与 IT 技术的结合. 可以利用 IT 技术, 特别是电子计算机技术来帮助处理 DNA 计算机研究过程中的许多问题, 诸如 DNA 分子的合成与生化反应相结合的自动化问题、DNA 计算机模型的通用性问题等. 我们以为, DNA 计算机研究过程中某些的不可缺少的辅助性的技术(如显示问题等)也可通过 IT 技术来完成.

## 参 考 文 献

- [1] Adleman L. Molecular computation of solution to combinatorial problems. *Science*, 1994, 66(11): 1021-1024
- [2] Braich R S, Chelyapov N, Johnson C, Rothmund P W K, Adleman L. Solution of a 20-variable 3-SAT problem on a DNA computer. *Science*, 2002, 296(19): 499-502
- [3] Roweis S, Winfree E, Burgoyne R, Chelyapov N, Goodman M, Rothmund P, Adleman L. A sticker based architecture for DNA computation//Baum E B et al eds. *DNA Based Computers*, Proceedings of the 2nd Annual Meeting. Princeton, 1999: 1-27
- [4] Kari L, Paun Gh, Rozenberg G, Salomaa A, Yu S. DNA computing, sticker systems, and universality. *Acta Informatica*, 1998, 35(5): 401-420
- [5] Paun G, Rozenberg G. Sticker systems. *Theoretical Computer Science*, 1998, 204(2): 183-203
- [6] Gao Lin, Xu Jin. DNA solution of vertex cover problem based on sticker model. *Chinese Journal of Electronics*, 2002, 11(2): 280-284
- [7] Zimmermann K-H. Efficient DNA sticker algorithms for NP-complete graph problems. *Computer Physics Communications*, 2002, 144(3): 297-309
- [8] Xu Jin et al. Sticker DNA computer model (I): Theory. *Chinese Science Bulletin*, 2004, 49(7): 1-8
- [9] Xu Jin et al. Sticker DNA computer model (II): Applications. *Chinese Science Bulletin*, 2004, 49(8): 1-9
- [10] Yaakov Benenson, Tamar Paz-Elizur, Rivka Adar, Ehud Keinan, Zvi Livneh, Ehud Shapiro. Programmable and autonomous computing machine made of biomolecules. *Nature*, 2001, 414(22): 430-434
- [11] Head T. Formal language theory and DNA: An analysis of the generative capacity of specific recombinant behaviors. *Bull. Math. Biology*, 1987, 49(6): 737-759
- [12] Head T. Splicing schemes and DNA//Rozenberg G, Salomaa A eds. *Lindenmayer Systems: Impacts on Theoretical Computer Science, Computer Graphics and Developmental Biology*. Springer-Verlag, 1992: 371-383
- [13] Csuhaj-Varju E, Freund L, Kari L, Paun Gh. DNA computing based on splicing: Universality results//Proceedings of the 1st Annual Pacific Symposium on Biocomputing. Hawaii, 1996
- [14] Paun Gh, Salomaa A. DNA computing based on splicing operation. *Math. Japonica*, 1996, 43(3): 607-632
- [15] Freund Gh P R, Kari L. DNA computing based on splicing: The existence of universal computers. *Theory of Computing Systems*, 1999, 32(1): 69-112
- [16] Sakakaibara Y, Ferretti C. Splicing on tree-like structures. *Theoretical Computer Science*, 1999, 210(2): 227-243
- [17] Liu Qing-Hua, Wang Li-Man, Frutos Anthony G, Condon Anne E, Corn Robert M, Smith Lloyd M. DNA computing on surfaces. *Nature*, 2000, 403(13): 175-178
- [18] Zhang Feng-Yue, Yin Zhi-Xiang, Liu Bo, Xu Jin. DNA computation model to solve 0-1 programming problem. *Bio-systems*, 2004, 74(1-3): 9-14
- [19] Frutos A G et al. Demonstration of a word design strategy for DNA computing on surfaces. *Nucleic Acids Research*, 1997, 25: 4748-4757
- [20] Smith L M. A surface-based approach to DNA computation. *Journal of Computational Biology*, 1998, 5: 255-267
- [21] Liu Wen-Bin, Gao Lin, Wang Shu-Dong, Xu Jin. DNA computation model of maximal matching problem based on surface. *Acta Electronica Sinica*, 2003, 31(10): 1496-1499 (in Chinese)
- [22] Pan Lin-Qiang, Xu Jin, Liu Ya-Chun. A surface-based DNA algorithm for the maximal clique problem. *Chinese Journal of Electronics*, 2002, 11(4): 469-471

(刘文斌, 高琳, 王淑栋, 刘向荣, 许进. 最大匹配问题的 DNA 表面计算模型. 电子学报, 2003, 31(10): 1496-1499)

- [23] Zhang Feng-Yue, Yin Zhi-Xiang, Xu Jin. Application of DNA chip on 0-1 planning problem. *Progress in Biochemistry and Biophysics*, 2003, 30(3): 412-415(in Chinese)  
(张风月,殷志祥,许进. DNA 芯片在 0-1 规划问题中的应用. *生物化学与生物物理进展*, 2003, 30(3): 412-415)
- [24] Suyama A et al. DNA chips- intergrated chemical circuits for DNA diagnosis and DNA computers//*Proceedings of the 3rd International Micromachine Symposium*. Tokyo, 1998: 7-12
- [25] Sakakibara Y et al. Intelligent DNA chips: Logical operation of gene expression profiles on DNA computers. *Genome Informatics*, 2002, 11: 33-42
- [26] Sakamoto K, Gouzu H, Komiya K et al. Molecular computation by DNA hairpin formation. *Science*, 2000, 288: 1223-1226
- [27] Head T, Rozenberg G, Bladergroen R B, Breek C K D, Lommerse P H M, Spaink H P. Computing with DNA by operating on plasmids. *BioSystems*, 2000, 57: 87-93
- [28] Gao Lin, Ma Run-Nian, Xu Jin. The molecular algorithm of the matching problem based on plasmid DNA. *Progress in Biochemistry and Biophysics*, 2002, 29(5): 820-823(in Chinese)  
(高琳,马润年,许进. 基于质粒求解最大匹配问题的 DNA 算法. *生物化学与生物物理学进展*, 2002, 29(5): 820-823)
- [29] Ma Run-Nian, Zhang Qiang, Gao Lin, Xu Jin. Using DNA to solve the maximum weight clique of graphs. *Acta Electronica Sinica*, 2004, 32(1): 13-16(in Chinese)  
(马润年,张强,高琳,许进. 图的最大权团的 DNA 计算. *电子学报*, 2004, 32(1): 13-16)
- [30] Jonoska N, Karl S A, Saito M. Three dimensional DNA structures in computing. *Biosystems*, 1999, 52: 143-153
- [31] Seeman N C, Chen J. Synthesis from DNA of a molecule with the connectivity of a cube. *Nature*, 1991, 350: 631-633
- [32] Seeman N C, Zhang Y, Du S M, Chen J. Construction of DNA polyhedra and knots through symmetry minimization//*Siegel J S ed. Supramolecular Stereochemistry*. Dordrecht: Kluwer, 1995: 27-32
- [33] Liu G, Yin Z, Xu J. Algorithm of graph isomorphism with three dimensional DNA graph structures. *Progress in Natural Science*, 2004
- [34] Adleman L, Cheng Q, Goel A, Huang M. Running time and program size for self-assembled squares. *ACM Symposium on Theory of Computing (STOC)*, 2001: 740-748
- [35] Adleman L, Cheng Q, Goel A, Huang M et al. Combinatorial optimization problems in self-assembly. *ACM Symposium on Theory of Computing (STOC)*, 2002: 1-8
- [36] Winfree E. Design and self-assembly of two-dimensional DNA crystals. *Nature*, 1998, 394: 1223-1226
- [37] Martin-Vide C, Păun G, Salomaa A. Characterizations of recursively enumerable languages by means of insertion grammars. *Theoretical Computer Science*, 1998, 205(1-2): 195-205
- [38] Kari L, Păun G, Thierrin G, Yu S. At the crossroads of DNA computing and formal languages: Characterizing recursively enumerable languages by insertion-deletion systems//*Rubin H, Wood D eds. Proceedings of the DNA Based Computers III, DIMACS Series*, AMS Press, 1999, 48: 329-347
- [39] Takahara A. On the computational power of insertion-deletion systems//*Proceedings of the 8th International Meeting on DNA Based Computers*. Sapporo, 2002: 139-150
- [40] Lipton R J. DNA solution of hard computational problems. *Science*, 1995, 268(28): 542-545
- [41] Yin Zhi-Xiang, Zhang Feng-Yue, Xu Jin. The general form of 0-1 programming problem based on DNA computing. *Bio-systems*, 2003, 70(1): 73-79
- [42] Qi Ou-Yang, Kaplan P D, Liu Shu-Mao, Libchaber A. DNA solution of the maximal clique problem. *Science*, 1997, 278(17): 446-449
- [43] Liu Ya-Chun, Xu Jin, Pan Lin-Qiang, Wang Shi-Ying. DNA solution of a graph coloring problem. *Journal of Chemical Information and Computer Science*, 2002, 42(3): 524-528
- [44] Liu Wen-Bin, Xu Jin. A DNA algorithm for the graph coloring problem. *Journal of Chemical Information and Computers*, 2002, 42: 1176-1178
- [45] Yin Zhi-Xiang, Zhang feng-Yue, Xu Jin. A Chinese postman problem based on DNA computing. *Journal of Chemical Information and Computer Sciences*, 2002, 42(2): 222-224
- [46] Yin Zhi-Xiang, Zhang Feng-Yue, Xu Jin. The general form of 0-1 programming problem based on DNA computing. *Bio-systems*, 2003, 70(1): 73-79
- [47] Rothemund P. A DNA and restriction enzyme implementation of turing machine//*Lipton R J, Baum E B eds. DNA based Computers; Proceedings of the DIMACS Workshop*. Princeton University, 1995: 75-119(American Mathematical Society, Providence, Rhode Island, 1996)
- [48] Ogihara M, Ray A. Simulating Boolean circuits on a DNA computer. *Algorithmica*, 1999, 25(2,3): 239-250
- [49] Amos M, Dunne P E, Gibbons A. DNA simulation of boolean circuits//*Proceedings of the Third Annual Conference*. University of Wisconsin, Madison, Wisconsin, San Francisco, CA, 1998: 679-683
- [50] Dunne P E, Amos M, Gibbons A. Boolean transitive closure in DNA//*George Paun ed. Proceedings of the Computing with Bio-Molecules: Theory and Experiments*. Springer-Verlag, Singapore, 1998: 127-137
- [51] Mulawka J J, Wasiewicz P, Plucienniczak A. Another logical molecular NAND gate system//*Proceedings of the Seventh International Conference on Microelectronics for Neural, Fuzzy and Bio-Inspired Systems*. Granada, Spain, 1999: 340-346
- [52] Wasiewicz P. DNA computing: Implementation of data flow logical operations. *Future Generation Computer Systems*, 2001, 17(4): 361-378
- [53] Benenson Y, Gil B, Ben-Dor Uri, Adar R, Shapiro E. An autonomous molecular computer for logical control of gene expression. *Nature*, 2004: 1-6

[54] Praun G, Rozenberg G, Salomaa A. DNA Computing-New Computing Paradigms. Berlin: Springer,1998(in Chinese)  
(许进等译. DNA 计算——一种新的计算模式. 北京:清华大学出版社, 2004)

[55] Xu Jin, Zhang Lei. DNA computer principle, advances and difficulties (I): Biological computing system and its applications to graph theory. Chinese Journal of Computers, 2003, 26(1): 1-11(in Chinese)  
(许进等. DNA 计算机:原理、进展及难点(I). 计算机学报, 2003, 26(1): 1-11)

[56] Lewis H R, Papadimitriou C H. Zhang Li-An, Liu Tian translate. Elements of the Theory of Computation. Beijing: Tsinghua University Press, 2000(in Chinese)  
(Lewis H R, Papadimitriou C H. 张立昂,刘田译. 计算理论基础. 北京:清华大学出版社, 2000)

[57] Bondy J A, Murty U S R. Graph Theory with Applications. London: Macmillan Press, 1976

[58] Asiewicz P W, Janczak T, Mulawka J J. The inference based on molecular computing. Cybernetics and Systems, 2000, 31: 283-315

[59] Lee I H, Park J Y, Jang H M, Chai Y G, Zhang B T. DNA Implementation of Theorem Proving with Resolution Refutation in Propositional Logic. Berlin Heidelberg: Springer-Verlag, 2003: 156-167

[60] Xu Jin, Huang Bu-Yi. DNA computer principle, advances and difficulties (II): Setting up the database of DNA computer by the synthesis of DNA molecules. Chinese Journal of Computers, 2005, 28(10): 1583-1591(in Chinese)  
(许进,黄布毅. DNA 计算机:原理、进展及难点(II)——计算机“数据库”的形成——DNA 分子的合成问题. 计算机学报, 2005, 28(10): 1583-1591)

[61] Xu Jin, Qiang Xiao-Li, Fang Gang, Zhou Kang. A DNA computer model for solving vertex coloring problem. Chinese Science Bulletin, 2006, 51(4): 480-487(in Chinese)  
(许进,强小利,方刚,周康. 一种图顶点着色 DNA 计算机. 科学通报, 2006, 51(4): 480-487)

[62] Feynman R P. There's plenty of room at the bottom//Minaturization D H, Gilbert E D. New York: Reinhold, 1961: 282-296

[63] Bennett C H. On constructing a molecular computer. IBM Journal of Research and Development, 1973, 17: 525-532

[64] Conrad M. On design principles for a molecular computer. Communication of the ACM, 1985, 28: 464-480

[65] Ellenbogen J C. Architecture for molecular electronic computers. Proceeding of the IEEE, 2000, 88(3): 386-426

[66] Yin Zhi-Xiang, Zhang Feng-Yue, Xu Jin. DNA computing based on molecular beacons. Journal of Biomathematics, 2003, 18(4): 497-501(in Chinese)  
(殷志祥,张凤月,许进. 基于分子信标的 DNA 计算. 生物数学学报, 2003, 18(4): 497-501)



**XU Jin**, born in 1959, professor, Ph. D. supervisor. His research interests include bio-computer, graph theory and system engineering, bio-informatics, neural networks, genetic algorithms etc.

Background

This research is supported by the National Natural Science Foundation of China (60533010, 30670540), and the National High Technology Research and Development Program(863 Program) of China (2006AA01Z104); Research on the Theory, Model and Method of DNA Computer etc. The projects mainly focus on DNA computer models for processing graphical messages, including encoding DNA se-

**TAN Gang-Jun**, born in 1962, Ph. D. candidate, professor. His research interest include bio-computer and system engineering.

**FAN Yue-Ke**, born in 1959, associate professor. His research interest include DNA computing and graph theory.

**GUO Yang-An**, born in 1960, associate professor. His research interesting includes computer science and management science.

quences, synthesizing DNA molecules, setting up the model, detecting solutions, etc. The research group has been working on many aspects of DNA computing since 1996. They have published a monograph and more than 100 papers on DNA computing and DNA computer. In this paper, the authors discuss some basic models set up in DNA Computing. These models is the foundations of DNA computer.