

图顶点着色问题的 DNA 计算模型

强小利 赵东明 张 凯

(北京大学信息科学技术学院高可信软件技术教育部重点实验室 北京 100871)

摘 要 DNA 计算是以 DNA 分子作为数据的一种新型计算模式. 为了减少 DNA 计算中编码的数量, 不降低生化实验操作的可靠性, 文中建立了一种基于酶切技术和 PCR 技术的图顶点着色 DNA 计算模型, 给出了实现该模型的双编码的编码方案. 分析表明, 利用酶切技术和 PCR 技术能够有效删除非解并读取真解. 该模型的解的检测方法类似于 DNA 测序技术, 使得该模型更容易实现自动化操作.

关键词 DNA 计算; 图顶点着色问题; 编码

中图法分类号 TP301 DOI 号: 10.3724/SP.J.1016.2009.02332

A DNA Computing Model for Graph Vertex Coloring Problem

QIANG Xiao-Li ZHAO Dong-Ming ZHANG Kai

(Key Laboratory of High Confidence Software Technologies of Ministry of Education,
School of Electronics Engineering and Computer Science, Peking University, Beijing 100871)

Abstract DNA computing is a novel computation paradigm with DNA molecules as “data”, and biochemistry trials as “information processing instruments”. In this paper, a DNA computing model to solve graph vertex 3-coloring problem is proposed based on enzyme digestion reactions. The graph vertex coloring problem is encoded by double encoding method and the false solutions deletion and the true solutions detection are updated and automatized partly after enzyme digestion reactions and polymerase chain reaction. This method could be easier and faster to read out the solution. Especially, the procedure of solution detection is similar to DNA sequencing technology.

Keywords DNA computing; graph vertex coloring problem; encoding

1 引 言

1994 年, Adleman^[1]开拓性地给出了应用 DNA 计算方法求一个给定有向图的有向 Hamilton 路问题的算法, 他给出了 DNA 计算的具体算法步骤, 每个算法步骤都对应着特定的生物操作. 这一成果表明了采用 DNA 进行特定目的计算是可行的. 之后,

十多年的研究表明, DNA 计算无论在理论上, 还是在实验技术上都取得了很大的进展. 1995 年, Lip-ton^[2]将此思想进行推广, 并解决了 SAT 问题(Satisfiability Problem). 1997 年, Ouyang 等人^[3]利用 POA(Parallel Overlap Assemble) 技术合成 DNA 分子来设计 DNA 计算中所需要的数据库, 提出求一个图的最大团问题的 DNA 算法模型, 并利用图中的团在其图的补图中对应的是空图这一简单的结

收稿日期: 2009-07-06; 最终修改稿收到日期: 2009-10-10. 本课题得到国家自然科学基金(60974112, 60910002, 60971085, 30970969)、国家“八六三”高技术研究发展计划项目基金(2009AA012413)、中国教育部博士点基金(20070001020)和中国博士后科学基金(20080440257)资助. 强小利, 女, 1979 年生, 博士, 目前主要从事分子计算模型的理论和应用、生物信息学等方面的研究. E-mail: xliqiang@pku.edu.cn. 赵东明, 男, 1977 年生, 博士, 主要研究方向为 DNA 计算. 张 凯, 男, 1979 年生, 博士, 主要研究方向为 DNA 计算.

论对 6 个顶点的图进行了实验仿真. 2000 年 Sakamoto 等人^[4]巧妙地采用发夹形式的 DNA 分子计算方法解决了 6 变量 10 子句的 CNF-SAT 问题,这使得 DNA 计算向自动的方向迈进了一步. Benenson 等^[5]实现了由 DNA 分子和酶组成的可编程半自动化试管中运算的具有状态转移功能的分子有限状态自动机,是 DNA 计算中的一个突破性的进展,并提出了一种“用于基因表达逻辑控制的自律式分子计算机”模型,该模型可用于疾病的诊疗^[5-6]. 此外,Winfree 利用人工合成的线性和带有分支的 DNA 分子作为自组装的基本元件(“瓦片”),提出了自组装 DNA 计算模型^[7]. 在 2002 年, Braich 等人^[8]应用粘贴模型给出了求解具有 20 个变量的 3 可满足性问题的半自动化自组装 DNA 计算模型,该模型的搜索次数可达 104 万次以上.

在 DNA 计算中,首先要解决的问题就是编码问题,它也是 DNA 计算中的一个关键问题. 影响编码的主要因素有自由能变化(ΔG)、解链温度(T_m)、DNA 分子的组成、酶以及汉明距离等^[9],目前已经有许多用于 DNA 计算编码的算法^[10-13]. 然而几乎所有的编码理论和可用程序的研究,都是基于降低杂交时的错配而展开的,都是将生物实验中的约束条件,用数学表达式表达,进而寻找合适的算法. 在 DNA 计算中,往往为了得到足够数量的 DNA 分子,而降低一些生物约束的要求^[14],进而导致实验中出现假阳性或是假阴性.

正是围绕上述问题,本文提出了一种编码方法,并在这个编码方法的基础上,建立一种用于求解图顶点着色问题的 DNA 计算模型. 利用该模型,不但实验所需的 DNA 链的数量减少,而且可以提高生化操作的可行性和有效性. 更重要的是,该模型中的读解的过程类似于 DNA 测序,这可以使得 DNA 计算易于实现解的自动化检测.

2 图顶点着色问题

本文所言之图皆指有限、无自环、无圈的无向简单图,通常用 G 表示图. 我们用 $V(G)$ 、 $E(G)$ 分别表示图 G 的顶点集和边集. 一个图 G 的着色是指对 G 中的每个顶点分配一种颜色,使得相邻的顶点染不同的颜色. 换言之,是指对图 G 中的顶点集 $V(G)$ 的剖分: $V(G) = V_1 \cup V_2 \cup \dots \cup V_k$, $V_i \neq \emptyset$ (空集), $V_i \cap V_j = \emptyset$, $i=1, 2, \dots, k$. 图 G 的一个 k -正常顶点着色,简称为图的 k -顶点着色,是指用 k 种颜色对图

G 进行着色.

由于本文仅考虑图的 3-着色问题,故我们总是假定颜色集为 $\{r, b, y\}$,其中 r 表示红色, b 表示蓝色, y 表示黄色. 所以,求解图的 3-着色问题可以认为是寻找从图的顶点集 $V(G)$ 到颜色集 $\{r, b, y\}$ 的一种映射:

$$f: V(G) \rightarrow \{r, b, y\}, \forall uv \in E(G), f(u) \neq f(v),$$

使得对于 $\forall uv \in E(G), f(u) \neq f(v)$. 图的 3-着色问题是一个 NP-完全问题.

众所周知,图的着色问题是一个困难的组合优化问题,具有良好的应用背景,诸如工序问题、排课表问题以及存储问题等均有直接的应用^[15-17]. 目前,已经有不少的算法用于研究图的着色问题,诸如常规的算法^[18-20]、神经网络算法^[21-22]、遗传算法^[23]等.

近年来,已经提出了几个基于 DNA 计算的图的着色问题的理论计算模型^[24-26]. 本文所提出的这个新的图顶点着色 DNA 计算机模型,也正是在上述研究工作的基础上提出来的.

3 图顶点着色问题的 DNA 计算模型

本文所言的图均指简单无向图,通常用 $V(G)$ 、 $E(G)$ 分别表示图 G 的顶点集和边集. 下面针对图 1 给出的具体的图,我们给出具体的算法. 该算法所采用的主要的分子生物学实验技术是酶切技术和 PCR 技术.

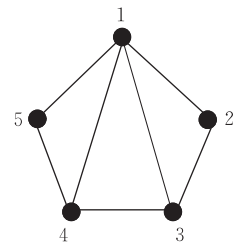


图 1 一个 5 阶图

3.1 模型的算法步骤

下面我们给出该图顶点 3-着色问题的 DNA 计算模型的算法步骤:

1. 编码. 首先根据现有的限制性内切酶,确定代表每个顶点不同颜色的序列,以及在现有编码理论的基础上根据本实验所涉及到的生物实验的约束条件,确定数学约束条件,并进行编程,获得连接序列.

2. 构建初始解空间:将代表每个顶点的限制性内切酶序列和代表不同颜色组合的 DNA 序列,按照一定的合成方法正确地连接在一起,构建代表待求解问题的所有的可能的

解的 DNA 链,即初始解空间,记为 H .

3. 删除非解,利用限制性内切酶切割消化代表不满足图正常着色的 DNA 链.
4. 利用琼脂糖电泳,分离得到未被切割消化的 DNA 链.
5. 根据待求解图的相邻矩阵,重复步 3、步 4,继续排除代表不满足图正常着色的 DNA 链,得到代表待求解图的正正常着色的 DNA 链.
6. 利用 PCR 方法,确定连接序列的位置,并确定图的正正常着色方案.

3.2 模型的实现

3.2.1 编 码

文章中提出的算法模型是基于双编码法,即在实验中采用两套编码,具体设计如下:

从现有的限制内切酶的识别序列选取一些序列来代表图的不同定点的可能的颜色,作为第一套编码.这套编码的设计不需要借助计算机,只需要从现有的限制性内切酶中选出限制性内切酶,用其酶切位点的碱基序列即可.选择原则条件主要是:(1)每种酶的酶切效率要高,(2)不能出现同位酶,(3)各种酶的缓冲体系尽可能一致.一般地限制性内切酶的识别序列的长度为 4~8 个碱基,这里,选取序列长度为 6 个碱基的限制性内切酶,定义该长度为 l .

根据生物公司提供的产品说明,选择出 15 种限制性内切酶的酶切位点的序列来代表图 G 每个顶点的不同着色.我们令每个顶点都与 3 个固定的限制性内切酶的酶切位点的碱基序列对应,并记为 $Ex_i, x \in \{r, b, y\}, i = 1, 2, \dots, 5$,分别表示顶点 i 着色为红色、蓝色和黄色.因此对于 n 个顶点的图来讲,则需要有 $3n$ 不同的限制性内切酶. Ex_i 及其对应的限制性内切酶及其识别序列见表 1.

表 1 Ex_i 对应的限制性内切酶及其识别序列

顶点颜色	限制性内切酶	DNA 序列
r_1	SmaI	5'-CCCGGG-3'
b_1	AatII	5'-GACGTC-3'
y_1	Alw44I	5'-GTGCAC-3'
r_2	BauI	5'-CACGAG-3'
b_2	Bsu15I	5'-ATCGAT-3'
y_2	Bsp1407I	5'-TGTACA-3'
r_3	EheI	5'-GGCGCC-3'
b_3	Eco72I	5'-CACGTG-3'
y_3	Eco105I	5'-TACGTA-3'
r_4	MbiI	5'-GAGCGG-3'
b_4	MunI	5'-CAATTG-3'
y_4	NcoI	5'-CCATGG-3'
r_5	NheI	5'-GCTAGC-3'
b_5	NsbI	5'-TGCGCA-3'
y_5	PdiI	5'-GCCGGC-3'

第二套编码是 13 条全长为 l' 基的寡核苷酸片段,其中 R, B 和 Y 代表首序列,尾序列用 T 来表示.还有 9 条寡聚核苷酸序列则用于编码不同顶点的各种可能的颜色的组合,用 $L_j (j = 1, 2, \dots, 9)$ 来表示,可作为连接序列,位与代表不同顶点的相应颜色的序列之间,即与第一套编码交替排列. L_j 所代表的组合如表 2 所示,第二条编码对应的寡核苷酸序列如表 3 所示.

表 2 L_j 对应的颜色组合

	r	b	y
r	L_1	L_2	L_3
b	L_4	L_5	L_6
y	L_7	L_8	L_9

表 2 中左边第 1 列代表 L_j 前一个顶点的着色,而上边第 1 行则代表 L_j 后面一个顶点的着色.

表 3 第二套编码对应的寡核苷酸序列

连接序列	DNA 序列
R	5'-CTCTTCCATAATCC-3'
B	5'-ATTCTCATCCCATTA-3'
Y	5'-TCTCTATCCATCCAA-3'
L_1	5'-ACTACCTCCATAACC-3'
L_2	5'-ACCATTTATCTCCAA-3'
L_3	5'-ATACTCCTCCTCTTC-3'
L_4	5'-AAACCACAAATACCA-3'
L_5	5'-TATTCTCCTTATCC-3'
L_6	5'-CTTTCTATCATTCCA-3'
L_7	5'-CTTCATCCACTCAAA-3'
L_8	5'-TACCAATCCACATCA-3'
L_9	5'-TTACCTTCCATTCA-3'
T	5'-CAACACCCTACTACTA-3'

3.2.2 解空间的合成

在确定好编码之后,按照顶点的顺序连接编码,就可以形成代表图的每个可能的着色方案的 DNA 链,这样就得到了许多条以 R, B 或者 Y 开始,以 T 结尾,中间为 $Ex_i, x \in \{r, b, y\}, i = 1, 2, \dots, 5$ 和 $L_j, j = 1, 2, \dots, 9$, 交替排列的,全长为 $5l + 6l'$ 个碱基 DNA 链.这里 $l = 6, l' = 15$, 则 DNA 链中含有 120 个碱基.这些 DNA 链的集合即为

$$H = \{XEx_1L_jEx_2L_j \cdots Ex_5T; X \in \{R, B, Y\}; x \in \{r, b, y\}; j = 1, 2, \dots, 9\}.$$

这个集合中 DNA 链的数目,即为需要合成代表图 G 的、所有可能顶点 3-着色方案的初始解空间中 DNA 序列的数目,是 3^5 条.

H 为 3-着色问题的所有的解的集合,其中 L_j 和代表酶切序列的 Ex_i 是相互交替的,重要的是每个 Ex_i 后的 L_j 不是任意的,而是与顶点的颜色 x ,

$x \in \{r, b, y\}$ 相对应的. 例如 L_1 代表的是颜色 rr 这个组合, 那么 L_1 所连接的两个点的颜色就必须都是红色, 即只有 $Er_i L_1 Er_{i+1}$ 被合成, 而 $Er_i L_1 Eb_{i+1}$ 或者 $Er_i L_1 Ey_{i+1}$ 则不能被合成. 如何在两个顶点之间加入正确的 L_j 是 DNA 合成的关键. DNA 序列合成方法如下:

(1) 先在 9 个试管中(分别标记为 1~9 号)分别合成 3 条链, 其中 1~3 号中的序列相同为 REr_1 , 管 4~6 中的序列相同为 BEb_1 , 管 7~9 中的序列相同为 YEy_1 .

(2) 给管 1~9 的序列后分别添加序列 $L_1 \sim L_9$, 即管 1 的序列后添加 L_1 , 管 2 的序列后添加 L_2 , 以此类推. 这样就保证了每个顶点的不同颜色后加上了相应的连接序列. 而且在 9 个试管中形成了 9 条不同的链: 管 1~3 中的序列为 $REr_1 L_j, j=1, 2, 3$; 管 4~6 中的序列为 $BEb_1 L_j, j=4, 5, 6$; 管 7~9 中的序列为 $YEy_1 L_j, j=7, 8, 9$.

(3) 因为管 1, 4, 7 中连接序列所代表的颜色组合为 xr , 所以给这 3 管中的序列后, 加 Er_2 ; 管 2, 5, 8 中的连接序列所代表的颜色组合为 xb , 则给这些序列后加 Eb_2 ; 管 3, 6, 9 中的连接序列所代表的颜色组合为 xy , 则给这些序列后加 Ey_2 . 这样在完成此步后, 已经将顶点 1 和 2 按顺序相连接, 并且在相应的颜色间加入了正确的连接序列, 并且形成 3^2 条不同的 DNA 链. 此时, 管 1~9 中的 DNA 序列分别为 $REr_1 L_1 Er_2, REr_1 L_2 Eb_2, REr_1 L_3 Ey_2, BEb_1 L_4 Er_2, BEb_1 L_5 Eb_2, BEb_1 L_6 Ey_2, YEy_1 L_7 Er_2, YEy_1 L_8 Eb_2, YEy_1 L_9 Eb_2$.

(4) 因为在管 1, 4, 7 中的序列的最后都是 Er_2 , 混合管 1, 4, 7 中的序列, 并再平均分成 3 份, 重新标记为 1, 2, 3, 每个试管中都有 3 条不同的 DNA 链. 同样, 因为原来 2, 5, 8 号管中的序列的最后都是 Eb_2 , 混合管 2, 5, 8 中的序列, 并再平均分成 3 份, 重新标记为 4, 5, 6; 而原来 3, 6, 9 号管中的序列的最后均为 Ey_2 , 混合后再分为 3 份, 重新标记为 7, 8, 9.

(5) 重复步骤(2)~(4), 依次连接顶点 3, 4, 5. 这样每重复完一次, 即连接上 1 个顶点后, DNA 链的数量就会增加 3 倍, 合成完成后, DNA 链的数量即变为 3^5 条.

(6) 当顶点 5 被连接上后, 给所有试管中的序列后加尾, 即在顶点 5 后加序列 T . 合并所有试管中的序列, 即为代表图的 3-着色问题的所有的可能解的 DNA 链的集合, 此时为单链 DNA 分子.

(7) 用引物对 $\langle X, \bar{T} \rangle, X = \{R, B, Y\}$ 为引物, 进行 PCR 扩增, 得到 3^5 条双链 DNA 分子, 即为所有

的可能的顶点 3 着色的解.

代表图 1 所有可能的解的 DNA 序列可以送到生物公司去合成. DNA 的合成都是在 DNA 分子合成仪上, 按照给定的 DNA 序列和上述给定步骤自动合成的. 以这些合成好的 DNA 链为模板, 用 $\langle R, \bar{T} \rangle$ 为引物, 多所合成存储库进行扩增, 所得到的产物则为如图 2 所示的双链 DNA 数据链.

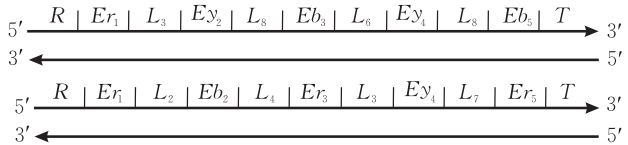


图 2 双链 DNA 数据链

3.2.3 删除非解

根据顶点颜色的序列, 采用酶切反应来删除非解. 具体以边 $e = \{1, 3\}$ 为例来讲, 为了消除顶点 1 和顶点 3 同时染红色的非解, 先将初始的 DNA 链平均分成两份分别置于试管 t_1 和 t_2 中. 给 t_1 试管中加入 Er_1 对应的限制性内切酶 $SmaI$, 并进行酶切反应. 同时, 给 t_2 试管中加入 Er_3 的限制内切酶 $EheI$ 进行酶切. 酶切反应后, t_1 试管带有 Er_1 的所有的 DNA 链被切断, 而 t_2 试管带有 Er_3 的所有的 DNA 链被切断, 再将这两个试管中的溶液合并, 并进行琼脂糖凝胶电泳进行分离, 回收片段大小为 120bp 的 DNA 链, 并保存于新试管 t 中. 此时, 这些 DNA 链中将不存在 r_1 和 r_3 同时出现的情况.

之后, 再将试管 t 重新分为两个试管 t_1 和 t_2 , 用于下一次酶切反应, 来删除其它边对应的非解. 重复此操作步骤, 直至完成所有边. 酶切反应全部完成后, 剩下的全长为 120bp 的 DNA 序列则代表了该图的真解.

执行非解的删除操作这一步骤的主要生物技术是酶切技术, 通过第 1 套编码来实现. 对于每条边来讲, 为了删除非解, 需要执行 6 次酶切操作. 因此, 对整个图来讲, 需要执行的酶切操作次数为 $t \leq 3 \times 2 \times m$, 其中 m 为图的边数. 当用单酶切反应时, $t = 6m$; 当用多酶切反应时, $t < 6m$.

3.2.4 解的检测

在所有的酶切反应完全后, 即在完全删除非解后, 剩下的全长为 120bp 的 DNA 序列则代表了待求解图的可以正常 3-着色的方案. 为了检测结果, 以 $\langle X, \bar{L}_j \rangle, X = \{R, B, Y\}, j = 1, 2, \dots, 9$ 作为引物对, 进行 PCR 反应. 所需要的反应次数为 27 次.

正如前面所讲到的, 代表图 G 的可能的 3-着色方案的每条双链 DNA 分子都是由 15bp 和 6bp 的寡核苷酸间隔排列所形成的全长为 120bp 的 DNA

分子,当以 $\langle X, \overline{L}_j \rangle$ 引物对时,经过 PCR 扩增后,所得到的 DNA 片段是一系列大小差异为 21bp 的 DNA,其中最短的片段长为 36bp,最长的片段为 99bp. 这样,我们就可以通过引物对和 PCR 扩增产物的片段大小来确定每个顶点的颜色.

在这里我们定义当 PCR 结果中没有 DNA 片段被扩增时,则记为 \emptyset , F_i 为以 X 和顶点 i 后面的 \overline{L}_j 为引物对时扩增出来的 DNA 片段. 通过分析 DNA 片段的大小,便可以读出最后的真解. 例如,当顶点 1 着色为红色时,完成 PCR 后,结果如表 4 所示.

表 4 解的检测

引物对	PCR 产物	顶点的着色
$\langle R, \overline{L}_1 \rangle$	\emptyset	
$\langle R, \overline{L}_2 \rangle$	36bp	F_1 $r_1 b_2$
$\langle R, \overline{L}_3 \rangle$	36bp	F_1 $r_1 y_2$
$\langle R, \overline{L}_4 \rangle$	\emptyset	
$\langle R, \overline{L}_5 \rangle$	\emptyset	
$\langle R, \overline{L}_6 \rangle$	57bp, 78bp, 99bp	F_2, F_3, F_4 $b_2 y_3, b_3 y_4, b_4 y_5$
$\langle R, \overline{L}_7 \rangle$	\emptyset	
$\langle R, \overline{L}_8 \rangle$	57bp, 78bp, 99bp	F_2, F_3, F_4 $y_2 b_3, y_3 b_4, y_4 b_5$
$\langle R, \overline{L}_9 \rangle$	\emptyset	

通过这些 PCR 反应,可以得到的结果为 $r_1 b_2$ 、 $r_1 y_2$ 、 $b_2 y_3$ 、 $b_3 y_4$ 、 $b_4 y_5$ 、 $y_2 b_3$ 、 $y_3 b_4$ 和 $y_4 b_5$. 通过重复的部分,我们就可以将所有的 DNA 片段重新拼接成全长为 120bp 的序列,就可以直接得到以顶点 1 为 r 的所有的解. 如图 3 所示.

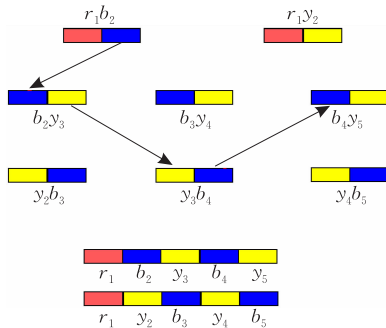


图 3 顶点 1 染红色时的解

从图 3 中,我们不难读出当顶点 1 着色为 r 时,满足图 G 顶点 3 着色的解为 $r_1 b_2 y_3 b_4 y_5$ 和 $r_1 y_2 b_3 y_4 b_5$. 用同样的方法也很容易得到当顶点 1 着色为 b 和 y 时的解.

4 结论和讨论

DNA 计算的最大特点是大规模的并行运算及巨大的信息存储能力,但是编码问题确实是目前困扰 DNA 计算的一个障碍. 本文建立了一种基于酶切反应的 DNA 计算模型,给出了实现该模型的双

编码的编码方案. 并且基于该编码方案,提出构建初始解空间的方法、删除非解以及解的检测方法. 该模型的核心思想是双编码方法,利用该编码方法,在某种程度上可实现解检测和非解删除的自动化操作. 分析论证表明,该模型具有编码量少、求解过程简单、容易实现自动化等优点,可用于解决一定规模的图顶点着色问题.

本文提出了一种双编码法,使得所需的编码数量减少,进而增加编码条件的约束力,提高编码的质量. 该编码方法不仅可以应用于图顶点着色问题的 DNA 计算,同时也应用于其它问题的 DNA 计算.

本文提出的建立在双编码基础上的解的检测过程也是利用 PCR 反应,对于任意一个图的顶点 3-着色问题来讲,PCR 反应管数最多仅为 27 管,不会随着顶点的增加而增加. 而且这种读解的思想类似于最初的基因组测序思想. 对于酶切求解的过程,可以通过微流控制等操作. 因此,本文提出的方法易于实现自动化操作.

当然用此种方法,需要寻找合适的足够多的限制性内切酶. 当有 n 个顶点时,需要有 $3n$ 种限制性内切酶参与反应. 这是制约解决更大规模的顶点 3-着色问题的很重要的一个因素. 另外,对于限制性内切酶的反应效率不能达到 100% 来讲,可以通过增加反应中的酶浓度、酶切反应的次数和时间等来减少误差. 如何在实验中保证酶切反应的有效性等问题都需要做进一步的研究.

参 考 文 献

- [1] Adleman L M. Molecular computation of solutions to combinatorial problems. *Science*, 1994, 266: 1021-1024
- [2] Lipton R J. DNA solution of hard computational problems. *Science*, 1995, 268: 542-544
- [3] Ouyang Q, Kaplan P D, Liu S et al. DNA solution of the maximal clique problem. *Science*, 1997, 278: 446-449
- [4] Sakamoto K, Gouzu H, Komiya K et al. Molecular computation by DNA hairpin formation. *Science*, 2000, 288(5469): 1223-1226
- [5] Benenson Y, Paz-Elizur T, Adar R et al. Programmable and autonomous computing machine made of biomolecules. *Nature*, 2001, 414(22): 430-434
- [6] Benenson Y, Gil B, Ben-Dor U et al. An autonomous molecular computer for logical control of gene expression. *Nature*, 2004, 429(27): 1-6
- [7] Winfree E. Design and self-assembly of two-dimensional DNA crystals. *Nature*, 1998, 394(6693): 539-544
- [8] Braich R S, Chelyapov N, Johnson C et al. Solution of a 20-variable 3-SAT problem on a DNA computer. *Science*, 2002, 296: 499-502

- [9] Tulpan D, Andronescu M, Chang S B et al. Thermodynamically based DNA strand design. *Nucleic Acids Research*, 2005, 33(15): 4951-4964
- [10] Kobayashi S, Kondo T, Arita M. On template method for DNA sequence design//*Letter Notes in Computer Science* 2568. Springer, 2003: 205-214
- [11] Błazewicz J, Formanowicz P, Kasprzak M et al. DNA sequencing, eulerian graphs, and the exact perfect matching problem//*Lecture Notes in Computer Science* 2573. Springer, 2002: 13-24
- [12] Pancoska P, Moravek Z, Moll U M. Rational design of DNA sequences for nanotechnology, microarrays and molecular computers using Eulerian graphs. *Nucleic Acids Research*, 2004, 32(15): 4630-4645
- [13] Tanaka F, Kameda A, Yamamoto M et al. Design of nucleic acid sequences for DNA computing based on a thermodynamic approach. *Nucleic Acids Research*, 2005, 33(3): 903-911
- [14] Aoi Y, Yoshinobu T, Tanizawa K et al. Ligation errors in DNA computing. *BioSystems*, 1999, 52(1-3): 181-187
- [15] Bondy J A, Murty U S R. *Graph Theory with Applications*. London, Basingtoke and New York; The Macmillan Press LTD, 1976
- [16] Gibbons A. *Algorithmic Graph Theory*. Cambridge, London, New York; Cambridge University Press, 1985
- [17] Jensen T R, Toft B. *Graph Coloring Problems*. New York; John Wiley & Sons, 1995
- [18] Wood D C. A technique for coloring a graph applicable to large-scale optimization problems. *The Computer Journal*, 1969, 12(317): 1
- [19] Karger D, Motwani R, Sudan M. Approximate graph coloring by semidefinite programming. *Journal of ACM*, 1998, 45(2): 246-265
- [20] Halldorsson M M. A still better performance guarantee for approximate graph coloring. *Information Processing Letters*, 1994, 45(1): 19-23
- [21] Xu Jin, Bao Zheng. Neural network and graph theory. *Science in China, Series A*, 2001, 31(6): 533-555(in Chinese) (许进, 保铮. 神经网络与图论. *中国科学 E 辑*, 2001, 31(6): 533-555)
- [22] Blas A D, Jagota A, Hughey R. Energy function-based approaches to graph coloring. *IEEE Transactions on Neural Networks*, 2002, 13(1): 81-90
- [23] Goldberg D. *Genetic Algorithm in Search, Optimization and Machine Learning*. Addison-Wesley Publishing, 1989
- [24] Liu Wenbin, Xu Jin. A DNA algorithm for the graph coloring problem. *Journal of Chemical Information and Computers*, 2002, 42(5): 1176-1178
- [25] Gao Lin, Xu Jin. A DNA algorithm for graph vertex coloring problem. *Acta Electronica Sinic*, 2003, 31(4): 494-496(in Chinese) (高琳, 许进. 图的顶点着色问题的 DNA 算法. *电子学报*, 2003, 31(4): 494-496)
- [26] Xu Jin, Qiang Xiaoli, Fang Gang, Zhou Kang. A DNA computer model for solving vertex coloring problem. *Chinese Science Bulletin*, 2006, 51(20): 2541-2549



QIANG Xiao-Li, born in 1979, Ph. D.. Her main research interests include bioinformatics, and DNA computing.

ZHAO Dong-Ming, born in 1977, Ph. D.. His main research interests focus on DNA computing.

ZHANG Kai, born in 1979, Ph. D.. His main interests focus on DNA computing.

Background

DNA computing is a novel computation paradigm with DNA molecules as “data”, and biochemistry trials as “information processing instruments”. Because of its massive parallelism, high-density storage and energy efficiency, DNA computing attracts the concern of many scientists with different backgrounds. Up to now, many accomplishments have been achieved to improve its performance and increase its reliability. The results of these assays indicate that DNA computing is superior to electronic computer in solving some NP-complete problem. However, many issues still exist in DNA computing. It is important and interesting for further study on DNA computing.

In this paper, a DNA computing model to solve graph

vertex 3-coloring problem was proposed based on enzyme digestion reactions. The graph vertex coloring problem is encoded by double encoding method, and the false solutions deletion and the true solutions detection are updated and automated partly after enzyme digestion reactions and polymerase chain reaction. By using this model the proper color of any graph will be obtained after times of digestion reaction and times of PCR operation, where is the number of edges in a graph, and is the chromatic number of a graph.

Compared with other DNA computing algorithms, this method could be easier and faster to read out the solution. Especially, the procedure of solution detection similar to DNA sequencing technology.