

# 最小支配集问题的活体分子计算模型

刘向荣<sup>1,2)</sup> 王淑栋<sup>3)</sup> 郝方<sup>1)</sup> 陈梅<sup>1)</sup>

<sup>1)</sup>(北京大学信息科学技术学院软件研究所 北京 100871)

<sup>2)</sup>(厦门大学信息科学与技术学院计算科学系 福建 厦门 361005)

<sup>3)</sup>(山东科技大学信息科学与工程学院 山东 青岛 266510)

**摘要** 生物体内分子网络中信息的传输、储存、放大、整合等大量任务可以看成是一种生物分子计算过程.文中提出了一种活体分子计算模型,借助 RNA 干扰技术和乳糖操纵子调控模型,在细胞内构建了一个基因网络,用于求解图的最小支配集.该模型展示了利用生物体自身的信息处理能力进行计算的能力,在生物体内建立具有一定智能的分子机器,这将在计算科学、生物学、医学上有着深远的应用前景.

**关键词** 活体分子计算;基因网络;RNA 干扰;最小支配集问题

中图法分类号 TP301 DOI号: 10.3724/SP.J.1016.2009.02325

## A Biomolecular Computing Model in Vivo for Minimum Dominating Set Problem

LIU Xiang-Rong<sup>1,2)</sup> WANG Shu-Dong<sup>3)</sup> XI Fang<sup>1)</sup> CHEN Mei<sup>1)</sup>

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

<sup>2)</sup>(Department of Computer Science, School of Information Science and Technology, Xiamen University, Xiamen, Fujian 361005)

<sup>3)</sup>(College of Information Science and Engineering, Shandong University of Science and Technology, Qingdao Shandong 266510)

**Abstract** Biomolecular computing models in vivo are an emerging computing model inspired from the biological phenomena that the biochemical molecular in living perform computation, communications, and signal processing collaboratively. In this paper, a biomolecular computing model in vivo for minimum dominating set problem is presented, a synthetic gene network is constructed by RNAi and lactose operon in living cell. This model explores further the ability to solve hard problems based on organism processing signal, and try to construct an intelligent molecule machine in cell. It may be widely and further used in computing science, biology, and medicine.

**Keywords** biomolecular computing in vivo; gene network; RNA interference; minimum dominating set problem

## 1 引言

DNA 计算是以 DNA 分子为“数据”、生物酶或者生化操作为“工具”的一种新的信息处理计算机模

型.自 Adleman 博士 1994 年提出的 DNA 计算模型<sup>[1]</sup> 至今已有 10 多年的历史.10 多年来, DNA 计算无论是在理论与应用模型的建立上,还是在实验与检测技术上都取得了惊人的进展.许多学者在此领域内作出了杰出的工作,如 Lipton 建立了 SAT

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(60910002, 60974112, 60971085, 30970969)、国家“八六三”高技术研究发展计划项目基金(2009AA012413)、教育部博士点基金(20070001020)和中国博士后基金(20080440257)资助.刘向荣,男,1978年生,博士,副教授,目前主要研究方向为分子计算、系统生物学等. E-mail: liuxr@pku.edu.cn. 王淑栋,女,1973年生,博士,教授,目前主要从事 DNA 计算模型的理论和应用、图与组合最优化等方面的研究.郝方,男,1986年生,博士研究生,主要研究方向为 DNA 计算.陈梅,女,1986年生,博士研究生,主要研究方向为 DNA 计算.

问题的 DNA 计算模型<sup>[2]</sup>, Ouyang 等建立了求解图的最大团问题的 DNA 计算模型<sup>[3]</sup>, Sakamoto 等建立的求解 SAT 问题的发卡 DNA 计算模型<sup>[4]</sup>, Frutos 和 Liu 等给出了求解 SAT 问题的表面 DNA 计算模型<sup>[5-6]</sup>, Rothmund 等建立了图灵 DNA 计算机模型<sup>[7]</sup>, Winfree 等提出了自组装 DNA 计算模型<sup>[8]</sup>, Roweis 等提出的粘贴 DNA 计算模型<sup>[9]</sup>. 2002 年, Braich 和 Adleman 等建立了具有 20 个变量 3-可满足性问题的 DNA 计算机模型<sup>[10]</sup>. 文献[11-12]分别建立了不同类型的图顶点着色 DNA 计算模型, 文献[13]给出了一种枚举型的图顶点着色 DNA 计算模型.

上述计算模型和实验都是在体外试管溶液中或者固体表面上进行的. 但是这些方法在利用生物体自身的信息处理能力方面显得有些不足. 在生物分子自动机研究中, 人们已经开始探讨利用细胞内部的信息处理能力. 2001 年, Shapiro 等提出了一种可编程的有穷自动机模型<sup>[14]</sup>, 2004 年, 他领导的小组提出了一种用于基因表达逻辑控制的分子自动机模型<sup>[15]</sup>, 该模型被用于肺癌和前列腺癌的诊疗, 但其生物实验还是在体外试管中进行仿真的. 2004 年, Kramer 等在哺乳动物细胞内实现了一个用于转录控制的生物分子逻辑门<sup>[16]</sup>. 2007 年, Benenson 等利用 RNA 干扰技术, 在细胞内构建了一个分子计算机来执行一般的布尔逻辑运算<sup>[17]</sup>. 本文借助 RNA 干扰技术和乳糖操作子调控模型, 给出了图最小支配集问题的一种活体分子计算模型. 在细胞内构建了一个基因网络, 通过细胞的自身生理机制转录、翻译表达, 实现了图最小支配集问题的求解. 其中, 输入的是 RNA 双链分子及目标 DNA 分子, 目标 DNA 分子通过编码设计、载体构建、细胞转染等生物操作来实现运算, 最后输出为蛋白质. 这是直接利用生物体自身强大的信息处理能力解决一些计算问题的尝试, 是一种全新的计算模式.

## 2 模型设计

### 2.1 模型设计理论

#### (1) 最小支配集问题

设  $G$  是一个有限、无向、简单图, 分别用  $V(G)$ 、 $E(G)$  和  $N(v)$  表示图  $G$  的顶点集、边集和顶点  $v$  相邻顶点集. 设  $D \subseteq V(G)$ , 若对任意顶点  $v \in V(G)$ , 要么  $v \in D$ , 要么  $v$  与  $D$  内一个顶点相邻, 则称  $D$  为图  $G$  的一个支配集; 若  $D$  的任何真子集皆非支配集,

则  $D$  为图  $G$  的一个极小支配集. 设  $D_0$  是图  $G$  的一个支配集, 但已无支配集  $D_1$ , 使得  $|D_1| < |D_0|$ , 则称  $D_0$  为图  $G$  的最小支配集. 最小支配集  $D_0$  中包含的顶点个数称为  $G$  的支配数, 用  $\gamma(G)$  表示. 图最小支配集问题是图与组合优化中的一个 NP 难题, 它与图的最大独立集、最小覆盖以及最大团问题息息相关, 在网络资源优化分配等问题中有着广泛的应用. 从支配集的定义出发, 式(1)可以求出图的所有极小支配集:

$$\varphi = \prod_{i=1}^n (v_i + \sum_{u \in N(v_i), v_i \in V} u) \quad (1)$$

其中,  $n = |V|$ , 公式中的运算是逻辑运算. 若顶点  $v_i$  在支配集  $D$  中, 则式(1)中逻辑变量  $v_i$  的取值为“1”, 否则为“0”. 所有满足  $\varphi = 1$  的  $D = \{v_i | v_i = 1, i = 1, 2, \dots, n\}$  都为图  $G$  的极小支配集.

#### (2) RNA 干扰与基因调控网络

RNA 干扰(RNAi)是由双链 RNA (dsRNA) 引发的转录后基因沉默机制(PTGS). RNAi 主要是指 dsRNA 分子进入细胞内, 特异性降解与之同源的 mRNA, 从而特异高效地抑制靶基因表达活性的现象. 一般可分为两个阶段:

第 1 步(启动阶段). 较长 dsRNA 在 ATP 参与下被类似 RNase II 的特异核酸酶切割加工成 21 ~ 23nt 的由正义和反义链组成的小干扰 RNA (siRNA).

第 2 步(效应阶段). siRNA 在 ATP 参与下被 RNA 解旋酶解旋成单链, 并由其中反义链指导形成 RNA 诱导的沉默复合体(RISC). RISC 由 siRNA、解旋酶、ATP、核酸内切酶、核酸外切酶等多种成分组成. 活化的 RISC 在单链 siRNA 引导下识别互补的 mRNA, 并在 RISC 中的核酸内切酶作用下从 siRNA 引导链中心所对应的靶基因位置切割靶 mRNA, 最后再被核酸外切酶进一步降解, 从而干扰基因表达. 而被降解的 mRNA 进入 RdRP 机制扩增合成 siRNA, 从而再次执行 RNA 干扰.

基因调控网络由基因、蛋白质、小分子以及他们之间相互作用、相互影响的关系组成, 是一个复杂的生物信息处理系统, 精密的调节控制着生物体的生长发育等重要的生命活动. 基因调控过程可以发生在染色体、染色质、DNA、转录、翻译以及细胞周期等水平上. 其中最重要的部分是转录水平的调控, 通过 RNA 聚合酶 II、启动子、增强子、阻遏蛋白等分子调节着以 DNA 为模板合成 mRNA 的转录过程.

乳糖操纵子调控模型是目前转录水平最清楚的调控, 而 RNAi 是转录后水平的调控, 它们的可控特

异性阻断目标基因表达的能力, 是研究基因功能的有力工具, 而且也可以实现计算功能.

### 2.2 最小支配集问题的活体 DNA 计算模型

本文在细胞体内建立了一个求解图最小支配集的 DNA 计算模型, 具体实例如图 1 所示的图 G.

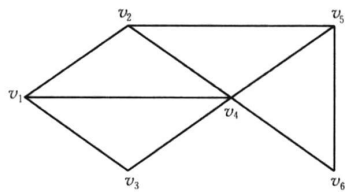


图 1 6 个顶点 9 条边的图

下式求出:

$$\varphi = (v_1 + v_2 + v_3 + v_4)(v_2 + v_1 + v_4 + v_5) \cdot (v_3 + v_1 + v_4)(v_4 + v_1 + v_2 + v_3 + v_5 + v_6) \cdot (v_5 + v_2 + v_4 + v_6)(v_6 + v_4 + v_5) \quad (2)$$

每个括号为一个子句, 子句内部“+”是顶点间的逻辑或操作, 子句间是逻辑与操作, 逻辑变量  $v_i$  取值为“1”时, 表示该逻辑变量对应的顶点属于某个极小支配集, 反之不属于.  $\varphi = 1$  时, 取值为“1”的顶点组成的集合为图的一个极小支配集.

结合 RNAi 和乳糖操纵子调控模型, 我们在细胞内构建了一个基因调控网络实现上述计算 (如图 2 所示). 该活体 DNA 计算模型在功能上可分为以下 3 个模块:

根据式(1)可知: 图 G 所有的极小支配集可由

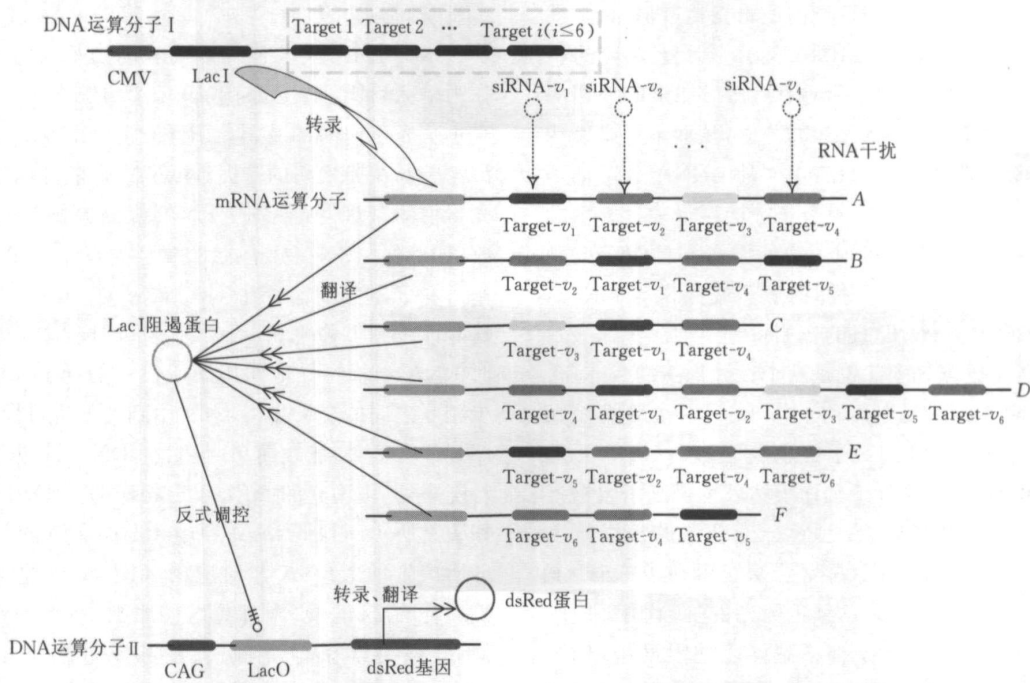


图 2 极小支配集活体 DNA 计算设计

(1) 输入模块. 主要包含代表图顶点的 siRNA 分子.

分别用分子  $siRNA - v_i$  的“出现”或“缺失”表示图 G 中顶点  $v_i$  对应逻辑变量取值为“1”或“0”, 即顶点“属于”或“不属于”某个极小支配集,  $i = 1, 2, \dots, 6$ . 针对图 1 中极小支配集  $D = \{v_1, v_5\}$ , 我们列出顶点、对应逻辑变量取值以及其对应分子的状态, 如表 1 所示.

(2) 计算模块. 主要包含 6 种 DNA 运算分子 I、转录生成的 6 种 mRNA 运算分子以及 1 种 DNA 运算分子 II. 计算功能主要由 RNA 干扰和乳糖操纵子的转录调控实现.

表 1 支配集  $D = \{v_1, v_5\}$  中顶点及其对应分子的状态表示

顶点	对应逻辑变量的取值	对应分子	分子状态
$v_1$	1	$siRNA - v_1$	出现
$v_2$	0	$siRNA - v_2$	缺失
$v_3$	0	$siRNA - v_3$	缺失
$v_4$	0	$siRNA - v_4$	缺失
$v_5$	1	$siRNA - v_5$	出现
$v_6$	0	$siRNA - v_6$	缺失

DNA 运算分子 I 是一个构建的载体, 含有一个巨细胞病毒增强子 (CMV)、乳糖操纵子中的阻遏蛋白基因 LacI 以及编码 mRNA 运算分子上的 RNA 干扰靶 mRNA 的若干 DNA 片段 (Target  $i, i = 1, 2, \dots, 6$ , 图 2 中虚线框部分) 组成. 根据子句的不同,

DNA 运算分子 I 所含的 Target  $i$  的组成和个数也不同. 这些 DNA 片段位于 LacI 基因的 3' 端非编码区.

mRNA 运算分子, 由相应的 DNA 运算分子 I 转录而来, 对应于式 (2) 中每一个子句. 图 2 中的 mRNA 运算分子-A, 由虚线框部分只含有 Target 1、Target 2、Target 3、Target 4 片断的 DNA 运算分子 I 转录而来. 该分子左部分是乳糖操纵子阻遏蛋白 (lacI repressor) 的 mRNA, 右部分是式 (2) 中第一个子句 ( $v_1 + v_2 + v_3 + v_4$ ) 对应的 mRNA: Target- $v_1$ , Target- $v_2$ , Target- $v_3$ , Target- $v_4$ , 分别与输入的干扰分子 siRNA- $v_1$ , siRNA- $v_2$ , siRNA- $v_3$ , siRNA- $v_4$  互补配对. 这 4 种 siRNA 分子中的任意一种或多种“出现” (即该分子对应顶点的逻辑变量取值“1”), 就可以引导 RISC 切割 mRNA 运算分子上对应的靶 mRNA (即互补配对的 Target), 然后在核酸外切酶的作用下降解该 mRNA 运算分子, 从而不翻译表达乳糖操纵子阻遏蛋白. 只有当 4 种 siRNA 分子都不“出现”时, 该 mRNA 运算分子才会翻译表达乳糖操纵子阻遏蛋白, 这就实现了第一个子句的逻辑或操作. 式 (2) 中所有子句内部的逻辑或操作与 mRNA 运算分子-B、C、D、E、F 的作用过程类似.

DNA 运算分子 II 也是一个设计构造的载体, 含有一个鸡  $\beta$  肌动蛋白启动子 (CAG)、一个乳糖操纵子操纵基因 LacO 和一个红色荧光蛋白 (dsRed) 基因. 当 mRNA 运算分子翻译表达乳糖操纵子阻遏蛋白时, 该蛋白将与 DNA 运算分子 II 上的操纵基因 LacO 结合, 并阻止 RNA 聚合酶在启动子 CAG 上的转录开始, 从而不表达红色荧光蛋白. 当图 2 中所有的 mRNA 运算分子都不翻译表达阻遏蛋白时, DNA 运算分子 II 将通过转录、翻译表达出红色荧光蛋白. 从而实现了多个 mRNA 运算分子间的一个逻辑与运算, 即实现了式 (2) 中 6 个子句间的与操作.

(3) 输出模块. 主要包含 DNA 运算分子 II 和 siRNA 分子. DNA 运算分子 II 通过转录、翻译表达的红色荧光蛋白的“出现”或“缺失”, 分别对应于  $\varphi$  取值“1”或“0”. 整个计算的输出结果由 siRNA 分子和红色荧光蛋白的“出现”和“缺失”组成. 真解是红色荧光蛋白“出现”时, 处于“出现”状态的 siRNA 分子组成的集合. 即  $\varphi=1$  时,  $D = \{v_i \mid v_i = 1, i = 1, 2, \dots, |V|\}$  为极小支配集, 比较所得的极小支配集的顶点个数, 含顶点个数最少的极小支配集就是图 G 的最小支配集.

上述 3 个模块组成了求解图最小支配集的活体

DNA 计算模型. 除了上述构建的基因网络, 计算过程还涉及到细胞本身含有的各种酶、转录因子等分子.

### 3 生物计算步骤

#### (1) 编码与合成

编码主要是指 siRNA 的编码, 涉及到目标 DNA 以及它表达的靶 mRNA. 由于 RNA 干扰机制中 siRNA 与靶 mRNA 互补, 所以三者是一致的. 在生物技术上, siRNA 编码有自身比较成熟的原则, 如 G/C 含量、siRNA 与靶 mRNA 的热力学特征、mRNA 二级折叠结构以及确保序列与细胞内基因组没有同源性等.

本文采用现有文献中非哺乳动物基因的 siRNA 作为输入, siRNA- $v_1$ 、siRNA- $v_4$ 、siRNA- $v_5$  来自萤火虫荧光蛋白酶的基因, siRNA- $v_2$ 、siRNA- $v_3$  来自海洋腔肠荧光素酶的基因, siRNA- $v_6$  来自增强绿荧光蛋白的基因. 并根据 siRNA 编码原则进行了修改, 如确保 siRNA 分子的 5' 端至少有一个 A/U 碱基对和 3' 端至少有一个 G/C 碱基对. 另外, 由于计算时涉及到多种靶 mRNA 的干扰调控, 需确保计算时 RNA 干扰特异地作用于所设计的靶 mRNA, 减少相互之间的交叉作用. 本文用到的编码目标 DNA 序列只有 6 种, 比较简单. 但目标 DNA 链包含一段特殊序列, 因为它的 mRNA 与相应的 siRNA 互补. 如表 2 所示, 目标 DNA Target 1 反义链的斜体部分所产生的 mRNA 与对应的 siRNA- $v_1$  的反义链是互补的. 借助 DNA 计算编码原则, 针对序列的相似性、热力学条件、编码长度和条数来优化目标 DNA 的编码. 具体编码的 siRNA- $v_i$  和目标 DNA Target  $i$  ( $i=1, 2, \dots, 6$ ) 见表 2, siRNA 两端有两个碱基的粘性末端, 目标 DNA 有 4 个碱基的粘性末端, 它们可由商业合成.

#### (2) DNA 重组

在载体 pZsYellow-C1 的基础上, 重组合成分别与式 (2) 中 6 个子句对应的 6 种 DNA 运算分子 I 和 1 种 DNA 运算分子 II. 载体 pZsYellow-C1 限制性图谱及其多克隆位点如图 3 所示, 它含有启动子 CMV, 用限制性酶 BamHI 和 XhoI 处理, 将目标 DNA 链插入载体的多克隆位点. 用限制性酶 BglIII 和 NheI 处理载体, 将 ZsYellow 基因替换成 LacI 基因, 这些都可以通过标准的克隆规程完成. 类似构建如图 2 所示的 6 种 DNA 运算分子 I:

表 2 6 种 siRNA 和目标 DNA 的编码

编码	链	序列(5'-3')
siRNA-v <sub>1</sub> <sup>[18]</sup>	正义链	GUCGAAG UACUCAGCGUAAGUG
	反义链	CU UACGCUGAGUACUUCGACAA
siRNA-v <sub>2</sub> <sup>[19]</sup>	正义链	CGUACGCGGAA UACUUCGAAU U
	反义链	UUCGAAGUA UUCGCGUACGUG
siRNA-v <sub>3</sub> <sup>[20]</sup>	正义链	CGUUGCUGAGUACCA ACCCUAU U
	反义链	UAGGG UUGG UACUAGCAACGCU
siRNA-v <sub>4</sub> <sup>[21]</sup>	正义链	GGCAAGCUGACCCUGAAGU UUG
	反义链	AACU UCAGGGUCAGCU UGCCGU
siRNA-v <sub>5</sub> <sup>[22]</sup>	正义链	CGAU AUGGCUGAAUACAAAU U
	反义链	UU UG UAUUCAGCCCA UAUCGUU
siRNA-v <sub>6</sub> <sup>[22]</sup>	正义链	GCUUGAAGUCU UUAUU UAAAAU
	反义链	UU UAAUUAAAAGACUUCAAGCGG
目标 DNA Target 1	正义链	TCCGACTTAACAAGCTTCGATTGTCG AAG TACTCAGCGTAAGG
	反义链	GATCCCTTACGCTGAGTACTTCGACAATCGAAGCTTGTTAAGC
目标 DNA Target 2	正义链	TCCGACTTAACAAGCTTCGACACGTACGCGGAATACTTCGAAG
	反义链	GATCCCTCGAAGTATTCGCGTACGTGT CGAAGCTTGTTAAGC
目标 DNA Target 3	正义链	TCCGACTTAACAAGCTTCGAAGCGTTGCTAGTACCAACCCTAG
	反义链	GATCCTAGGGTTGGTACTAGCAACGCTTCGAAGCTTGTTAAGC
目标 DNA Target 4	正义链	TCCGACTTAACAAGCTTCGAACGGCAAGCTGACCCTGAAGTTG
	反义链	GATCCAACTTCAGGGTCAGCTTCCGTT CGAAGCTTGTTAAGC
目标 DNA Target 5	正义链	TCCGACTTAACAAGCTTCGAAACGATATGGGTGAATACAAAG
	反义链	GATCCCTTGTATTAGCCATATCGTTTCGAAGCTTGTTAAGC
目标 DNA Target 6	正义链	TCCGACTTAACAAGCTTCGACCGCTTGAAGCTTTAAATTAAAG
	反义链	GATCCCTTAAATTAAGAATTCAAGCGGTCGAAGCTTGTTAAGC

- ① -CMV-LacI-Target 1-Target 2-Target 3-Target 4-;
- ② -CMV-LacI-Target 2-Target 1-Target 4-Target 5-;
- ③ -CMV-LacI-Target 3-Target 1-Target 4-;
- ④ -CMV-LacI-Target 4-Target 1-Target 2-Target 3-Target 5-Target 6-;
- ⑤ -CMV-LacI-Target 5-Target 2-Target 4-Target 6-;
- ⑥ -CMV-LacI-Target 6-Target 4-Target 5.

经转录分别对应生成 mRNA 运算分子-A、B、C、D、E、F.

类似, 可得 DNA 运算分子 II: -CAG-LacO-dsRED-.

(3) 细胞培养、转染与结果检测

本文使用商业化的细胞系 293-H 细胞, 按照常规转染操作规程, 将上述所有的目标 DNA 和 siRNA 转染进细胞, 经培养(48h 左右)后, 用 488nm

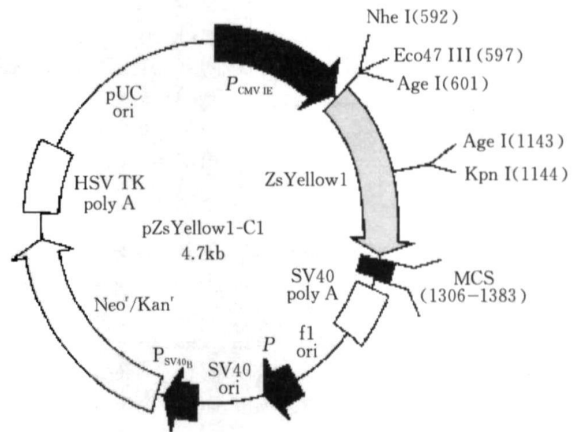


图 3 载体 pZsYellow-C1 限制性图谱

激光激发红色荧光蛋白, 在 575nm 左右波段检测荧光信号, 以确定是否有 deRed 蛋白“出现”, 即输入的 siRNA 是否是极小支配集. 可得到图 1 的所有极小支配集如表 3 所示, 其中“1”表示“出现”, “0”表示“缺失”.

表 3 图 G 的所有极小支配集对应的分子结果

极小支配集	输入输出						dsRed 蛋白
	siRNA-v <sub>1</sub>	siRNA-v <sub>2</sub>	siRNA-v <sub>3</sub>	siRNA-v <sub>4</sub>	siRNA-v <sub>5</sub>	siRNA-v <sub>6</sub>	
D <sub>1</sub>	1	0	0	0	1	0	1
D <sub>2</sub>	1	0	0	0	0	1	1
D <sub>3</sub>	0	0	0	1	0	0	1
D <sub>4</sub>	0	1	1	0	1	0	1
D <sub>5</sub>	0	1	1	0	0	1	1

即图 G 的所有极小支配集是

$$D_4 = \{v_2, v_3, v_5\}, D_5 = \{v_2, v_3, v_6\};$$

$$D_1 = \{v_1, v_5\}, D_2 = \{v_1, v_6\}, D_3 = \{v_4\},$$

$$\text{最小支配集是 } D_0 = D_3 = \{v_4\};$$

支配数:  $\gamma(G) = |D_0|$ .

## 4 结 论

本文在一些常规的生物操作(如核酸序列合成、载体构建、细胞转染等)和基因调控网络中比较成熟的理论(如乳糖操纵子调控模型、RNA 干扰技术等)的基础上,提出了一个求解图最小支配集的活体 DNA 计算理论模型.通过构建一个基因调控网络,经过编码、合成、DNA 重组、细胞培养和转染等一系列的生物操作,得到代表图的最小支配集的分子集合.这是利用生物体信息处理能力解决数学问题的一个尝试.

## 参 考 文 献

- [1] Adleman L M. Molecular computation of solution to combinatorial problems. *Science*, 1994, 266: 1021-1024
- [2] Lipton R J. DNA solution of hard computational problems. *Science*, 1995, 268(28): 542-545
- [3] Ouyang Q, Kaplan P D, Liu Shumao, Libchaber A. DNA solution of the maximal clique problem. *Science*, 1997, 278(17): 446-449
- [4] Sakamoto K, Gouzu H, Komiya K, Kiga D, Yokoyama S, Yokomori T, Hagiya M. Molecular computation by DNA hairpin formation. *Science*, 2000, 288(5): 1223-1226
- [5] Frutos A G, Liu Q, Thiel A J, Sanner A M, Condon A E, Smith L M, Corn R M. Demonstration of a word design strategy for DNA computing on surfaces. *Nucleic Acids Research*, 1997, 25(23): 4748-4757
- [6] Liu Qinghua, Wang Liman, Frutos A G, Condon A E, Corn R M, Smith L M. DNA computing on surfaces. *Nature*, 2000, 403(13): 175-178
- [7] Rothmund P. A DNA and restriction enzyme implementation of Turing machine, in DNA based computers//Proceedings of the DIMACS Workshop, Princeton University, 1995, 4: 75-119
- [8] Winfree E. Algorithmic self-assembly of DNA. Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy. California Institute of Technology Pasadena, California, 1998
- [9] Roweis S, Winfree E, Burgoyne R, Chelyapov N, Goodman M, Rothmund P, Adleman L M. 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
- [10] Braich R S, Chelyapov N, Johnson C, Rothmund P W K, Adleman L M. Solution of a 20-variable 3-SAT problem on a DNA computer. *Science*, 2002, 296(19): 499-502
- [11] Liu Yachun, Xu Jin, Pan Linqiang, Wang Shiyong. DNA solution of a graph coloring problem. *Journal of Chemical Information and Computer Science*, 2002, 42(3): 524-528
- [12] Liu Wenbin, Xu Jin. A DNA algorithm for the graph coloring problem. *Journal of Chemical Information and Computers*, 2002, 42(5): 1176-1178
- [13] Xu Jin, Qiang Xiao-Li, Fang Gang, Zhou Kang. A DNA computer model for solving vertex coloring problem. *Chinese Science Bulletin*, 2006, 51(20): 2541-2549
- [14] Benenson Y, Paz-Elizur T, Adar R, Keinan E, Livneh Z, Shapiro E. Programmable and autonomous computing machine made of biomolecules. *Nature*, 2001, 414(22): 430-434
- [15] Benenson Y, Gil B, Ben-Dor U, Adar R, Shapiro E. An autonomous molecular computer for logical control of gene expression. *Nature*, 2004, 429: 423-429
- [16] Kramer B B, Fischer C, Fussenegger M. Biologic gates enable logical transcription control in mammalian cells. *Biotechnology and Bioengineering*, 2004, 87(4): 478-484
- [17] Rinaldo K, Bleris L, Maddamsetti R, Subramanian S, Weiss R, Benenson Y. A universal RNAi-based logic evaluator that operates in mammalian cells. *Nature Biotechnology Research*, 2007, 25(7): 795-801
- [18] Paddison P J, Caudy A A, Bernstein E, Hannon G J, Conklin D S. Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes & Development*, 2002, 16(8): 948-958
- [19] Elbashir S M, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature*, 2001, 411(6836): 494-498
- [20] Kobayashi N, Matsui Y, Kawase A, Hirata K, Miyagishi M, Taira K, Nishikawa M, Takakura Y. Vector-based in vivo RNA interference: Dose- and time-dependent suppression of transgene expression. *Journal of Pharmacology and Experimental Therapeutics*, 2004, 308(2): 688-693
- [21] Sullivan C S, Ganem D. A virus-encoded inhibitor that blocks RNA interference in mammalian cells. *Journal of Virology*, 2005, 79(12): 7371-7379
- [22] Reynolds A, Leake D, Boese Q, Scaringe S, Marshall WS, Khvorovva A. Rational siRNA design for RNA interference. *Nature Biotechnology*, 2004, 22(3): 326-330



**LIU Xiang-Rong** born in 1978, Ph. D., associate professor. His main research fields focuses on biomolecular computing and system biology.

**WANG Shu-Dong** born in 1973. Ph. D., professor. Her research interests include biomolecular computing graph theory and its application etc. Her current research focuses on computing models and applications.

**XI Fang** born in 1986, Ph. D. candidate. His main research interests focus on DNA computing.

**CHEN Mei** born in 1986, Ph. D. candidate. Her main research interests focus on DNA computing.

## Background

Computation is a map or a transform process of symbol strings based on the rules. Symbol is the special physics state. It can be the electronic circuit behavior based on transistor technology. The machine is successfully realized by digital computer with a vast number of states. If the interaction between the bimolecular can be viewed as a finite number of components and the components can take on a few states. The particular biochemical reaction is biomolecular computing.

In 1994, Adleman was the first to demonstrate by a DNA experiment for solving a small instance of traveling salesman problem that biomolecular computations are feasible. Features such as the massive parallelism are very interesting from the computational point of view. A number of experiments solving hard computational problems have been presented. Researchers realized some of the obstacles related to this incipient technology soon thereafter, efficiency and precision of biochemical reaction, and exponential explosion of solution space with respect to problem size.

Therefore, progress in biomolecular computing will depend on both novel computing concepts and biological operation technique. The goal of researchers is to find biomolecular computing paradigms capable more than Turing machines or a new application platform.

The architecture of gene regulatory networks is reminiscent of electronic circuits. Increasing knowledge on how cell behavior is shaped by the complex regulatory transcription networks of different sets of genes, which show response characteristics as electric circuits. Precise and programmed control of gene activity can be accomplished by incorporating synthetic biochemical logic circuits into the cells. Such biochemical circuits perform a variety of simple computational tasks including amplification, integration and information storage.

Bacterial illustrates other common features of biomolecular computing *in vivo*, such as their ability to integrate multiple inputs. Plasmid is a circular, double-stranded DNA molecule, which contain an origin for replication and allows the production in bacteria. A method of computing using DNA plasmids is introduced by Head in 2000. It illustrated by reporting a laboratory computation of an instance of the NP-complete algorithmic problem of computing the cardinal number of a maximal independent subset of the vertex set of a graph. Henkel extended plasmid computing with protein expression by the construction of the whole computing region of a plasmid as part of an open reading frame. In 2007, several instances of knapsack problems using plasmids were presented.

Finite state automata operating autonomously at the molecular scale can be used conceptually for applications in the living cell. Shapiro built a small finite state automaton from DNA strands and enzymes. This approach might be applied *in vivo* to biochemical sensing, genetic engineering and even medical diagnosis and treatment. The first autonomous finite state machine working in a living cell was proposed by Sakakibara and

Coworkers.

Genetic circuits are collections of basic elements that interact to produce a particular behavior. By constructing biochemical logic circuits and embedding them in cells, one can extend or modify the behavior of cells. To date, several small synthetic gene networks have been built that accomplish specific genetic regulatory functions *in vivo*. The auto regulatory is a repressor regulates its own production to reduce noise in gene expression. Savageau proposed but not demonstrated that the negative feedback loops in gene circuits provide stability. Becskei and Serrano have designed and constructed simple gene circuits consisting of a regulator and transcriptional repressor modules in *Escherichia coli* and show the gain of stability produced by negative feedback. To test the role of negative feedback in the stability of gene networks, they first designed simple gene circuits based on simple control systems. Gardner presented the construction of a genetic toggle switch, a synthetic bistable gene-regulatory network, in *Escherichia coli* and provided a simple theory that predicts the conditions necessary for bistability. The toggle is constructed from any two repressible promoters arranged in a mutually inhibitory network. Elowitz present the design and construction of a synthetic network to implement a particular function, three transcriptional repressor systems that are not part of any natural biological clock to build an oscillating network in *Escherichia coli*. To overcome the shortcoming and challenge for the design and construction of more sophisticated genetic circuitry in the future. A combined rational and evolutionary design strategy for constructing genetic regulatory circuits proposed. Weiss presented a genetic component library and a gene circuit design methodology for assembling these components into compound circuits.

In 2004, Kramer pioneer a variety of different two- and three-input biologic gates in mammalian cells by combining several compatible heterogenous gene control units responsive to tetracycline, streptogramin, macrolide, and butyrolactose. In 2007, Weiss and Benenson used RNA interference (RNAi) in human kidney cells to construct a molecular computing core that implements general Boolean logic to make decisions based on endogenous molecular inputs. Computational gene model, a molecular finite state machine which can be implemented in molecular diagnostic and therapy of disease.

Here the authors propose a theoretic model *in vivo* for minimum dominating set problem. In the model, a synthetic gene network constituting a lactose operon and RNAi is constructed in living cell. The problem is solved by gene regulation in transcription and translation process. The inputs are RNA double strands and target DNA molecules, target DNA encoded the information, i.e. the logic values of the vertex of a graph. The computation is realized by vector construction, cell culture and transfection, and detection. The outputs are fluorescence protein.