

# DNA 计算方法

李 燕<sup>1</sup> 王秀峰<sup>2</sup>

(浙江师范大学计算机科学与工程学院 金华321004)<sup>1</sup> (南开大学信息工程学院 天津300071)<sup>2</sup>

**摘 要** DNA 计算是应用分子生物技术进行计算的新方法。本文主要介绍了 DNA 计算的基本思想及在解决 NP 完全问题中的应用。

**关键词** NP 完全问题, DNA 计算, 最大团问题

## DNA Computation

LI Yan<sup>1</sup> Wang Xiu-Feng<sup>2</sup>

(College of Computer Science and Engineering, Zhejiang Normal University, Jinhua, 321004)<sup>1</sup>

(College of Information Engineering, Nankai University, Tianjin, 300071)<sup>2</sup>

**Abstract** DNA algorithm is a new method for computation using the technology in molecular biology. This article first introduces the basic idea of DNA computation, then presents DNA computing method for solving NP-complete problems.

**Keywords** NP-complete problem, DNA computing, Max clique

## 1 引言

在计算复杂性理论中,把可以通过确定的算法在多项式时间内解决的问题称为 P 类。只能通过不确定的算法在多项式时间内解决的问题称为 NP 类。NP 问题随着问题维数的增多,算法的计算量也大大增加,以至于超过常规电子计算机的计算能力,所以 NP 问题被认为是难以解决的。20世纪70年代初斯蒂芬·库克和列奥尼德·列文在 P 与 NP 问题的研究中取得了重大的进展,他们发现 NP 类问题中的某些问题的复杂性与整个类的复杂性相关,这些问题中的任何一个问题如果存在多项式时间算法,那么所有 NP 问题都是多项式时间可解的,这些问题被称为 NP 完全的问题<sup>[1]</sup>。NP 完全问题是国内外算法研究领域的热点和难点,该类问题的解决不仅具有重要的理论意义,而且在国民经济各行各业中有非常重要的应用前景。P 与 NP 的问题已被公认为数学和理论计算机科学至今尚未解决的重要问题之一。多少年来, P=?NP 问题的研究受到科学家们的高度重视,但对于这个问题至今没有得到明确的解答。因此,突破传统的计算方法,开辟模拟生物智能的新路已成为人们研究 NP 问题的新方向。

## 2 DNA 分子计算方法的产生

在智能化领域的探索过程中,人们意识到研究智能的最好方式是模仿生物特性向人类自身学习。心理学家 W. McCulloch 和数理逻辑学家 W. Pitts 于1943年首先提出了 M-P 神经网络模型<sup>[2]</sup>。1985年,美国加州工学院物理学家 Hopfield 采用 Hopfield 神经网络模型,对 TSP 旅行商最优路径问题进行了研究。美国 Michigan 大学的 John Holland 教授,以达尔文的生物进化论和孟德尔的遗传变异理论为基础,于1975提出了遗传算法,在思路突破了原有最优化方法的框架<sup>[3]</sup>。这对 NP 问题的研究起到了极大的促进作用,很多学

者纷纷采用这种模拟生物系统的方法解决 NP 问题,但对于 NP 完全问题仍然没有有效的通用解决的办法。

1994年,美国加利福尼亚大学的 Aldeman 教授第一次利用现代分子生物技术,提出了哈密尔顿有向路径(Hamilton Path)问题的 DNA 分子生物计算方法,并成功地在 DNA 溶液的试管中进行了实验,在国际上引起了巨大的反响<sup>[4]</sup>。在 Aldeman 实验成功后不久, Lipton 提出了用于求解 SAT 问题的 DNA 分子计算模型<sup>[5]</sup>。根据 NP 完全理论,解决 SAT 问题的有效算法,自然地也是解决任意 NP 类问题的有效算法。随后,研究人员相继提出一些不同的 DNA 计算模型。这些模型有的是对 Adleman 的方案进行改进,使之操作更简便、通用、计算速度快、能容错等<sup>[6~12]</sup>;有的是基于表面化学的 DNA 计算模型,与 Adleman 在液相体系内完成 DNA 计算的模型不同,如 Q. Liu 等人在金属表面实现了 DNA 计算<sup>[13,14]</sup>。

## 3 DNA 分子的计算机理

科学家从 DNA 分子的特性与计算的本质中发现了它们的共性:(1)由 DNA 序列表示的原始信息,能够在酶的作用下执行简单的操作(复制、切割、粘接等),生物体表现出的千变万化、异常复杂的结构就是这些操作的结果;(2)可计算函数  $f(w)$  的结果,可以通过在  $w$  上执行一系列基本的简单函数而获得。DNA 计算本质上就是利用大量不同的核酸分子杂交,运用分子生物技术产生类似某种数学过程的运算结果。一个 DNA 单链可以看作是由四个不同的元素 A、T、C、G 连成的一个串。这四个元素能天然地形成数学上的一种良好的代数结构—域<sup>[14]</sup>。四元集  $\Sigma = \{A, T, C, G\}$ ,按照图1中所指定的两种代数运算“+”和乘法“\*”,构成了一个四元域。如图1所示。可以使用四个字母的集合  $\Sigma = \{A, T, C, G\}$ ,按照 DNA 的双螺旋结构对问题进行编码。如果给组成 DNA 分子

李 燕 博士,研究方向:计算智能、DNA 分子计算理论。王秀峰 博士导师。

的四个字母分配如下一个二进制或者四进制编码： $T \equiv 00 \equiv 0$ ， $C \equiv 01 \equiv 1$ ， $G \equiv 10 \equiv 2$ ， $A \equiv 11 \equiv 3$ 。它们能将一个 DNA 单链转换为熟悉的二进制编码或者四进制编码。

加法运算					乘法运算				
+	T	A	C	G	*	T	A	C	G
T	T	A	C	G	T	T	T	T	T
A	A	T	G	C	A	T	A	C	G
C	C	G	T	A	C	T	C	G	A
G	G	C	A	T	G	T	G	A	C

图1 DNA 四元域

完整的 DNA 计算过程包括数据输入、运算以及结果的输出。首先，将问题空间的对象映射成 DNA 分子链，在试管中构建出所有可能答案的原始数据。第二步，利用 DNA 分子链的高度并行运算的特点，用可控的生化操作完成正确答案的筛选工作。第三步，结果的输出。这一步通过测定来判断是否有解。如果有，将 DNA 链进行译码，就是所求问题的解。以上过程需要用到的分子生物技术有：聚合酶链式反应 PCR (Polymerize chain reaction)、凝胶电泳(使 DNA 链按长度分离)、亲和层析(过滤析取)、加热/退火(合成 DNA 双链及其逆过程解链)、克隆、磁珠分离、标记、合并、连接、切割、检测等。在生物试验中，可以利用酶对 DNA 分子序列进行操作，不同的酶相当于 DNA 串上的不同算子。如限制内切酶(restriction endonucleases)可作为分离算子，它识别链中特定的短序列，并在这个位上进行“切割”；连接酶(ligase)可作为连接算子，它把刚切过的 DNA 的黏端与其它链连接在一起；聚合酶(polymerases)可作为复制算子，进行 DNA 链的复制；外切核酸酶(exonuclease)可作为删除算子；还有转移酶(terminal transferase)、修饰酶(modifying enzymes)等，根据不同的过程可以采用不同的酶。

#### 4 最大团问题(Max clique)

一个图 G 的团 (clique) 是图 G 的一个顶点子集 V，使得 V 中任意两顶点之间都有边相连，若对 G 的任意其它团 V' 都有  $|V| \geq |V'|$ ，则称 V 是 G 的最大团<sup>[16]</sup>。最大团问题是一个 NP 完全问题。

分子计算步骤如下

(1) 初始化试管。对于具有 n 个顶点的图 G 进行编码。将每个可能的团表示为 n 位二进制数字串(按顶点顺序进行编号)。若二进制串的位为 1，则对应的顶点在团里，若为 0，则对应的顶点不在团里。例如如图 2(a) 的团 (1, 2, 4)，可用二进制串 011010 来表示。依照二进制位与 DNA 分子的映射关系，将图中所有表示团的 n 位二进制串分别映射为相应的 DNA 分子链，这样就形成了初始试管中的数据。

(2) 寻找图 G 的补图

(3) 从初始数据中删除在补图中具有相邻顶点的 DNA 分子链。例如如图 2(b)。相继重复使用相应的 DNA 生物操作方法(亲和层析技术)，从试管中去除表示二进制串的形式为 1x1xxx1xxxx1x, x1xxxx1xx1x1xx 的 DNA 分子链。

(4) 对试管中剩余的数据进行检测，检测是否有 DNA 链存在。如果有，从中选择出包含顶点数最多的 DNA 分子链，

通过译码就可得到图 G 的最大团。

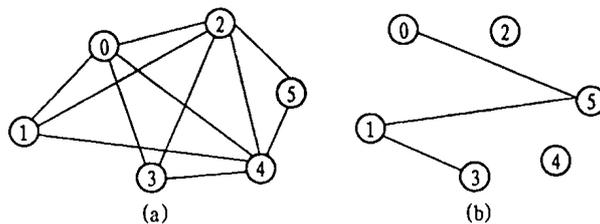


图2 一个6阶图与它所对应的补图

**结论** DNA 计算的出现已经打开了人们畅想未来计算方式的视窗，推动了计算科学的发展。从遗传进化、人工神经网络和 DNA 分子生物技术对智能的模拟过程看，它们分别对应生物群体、生物神经元和生物分子三个截然不同的层次。由此可以看到，基于对 DNA 分子的模拟和研究不仅是解决复杂类问题的突破口，而且有可能更深刻地揭示智能形成的本质。

#### 参考文献

- 1 加里 M R, 约翰逊 D S. 计算机和难解性[M]. 科学出版社, 1990
- 2 史忠植. 神经计算[M]. 电子工业出版社, 1993
- 3 Holland J H. Adaptation in natural and artificial system: An Introduction Analysis with Application to Biology[A]. Control and Artificial Intelligence. USA, The University of Michigan press, 1975
- 4 Adleman L M. Molecular computation of solutions to combinatorial problems[J]. Science, 1994, 226: 1021~1024
- 5 Lipton R J. DNA solution of hard computational problems [J]. Science, 1995, 268(5210): 542~545
- 6 Rooß D, Wagner K W. On the Power of DNA-Computing [J]. Information and Computation, 1996, 131(2): 95~109
- 7 Roweis, Sam, Erik W, et al. A sticker based model for DNA computation [J]. Journal of Computational Biology, 1998, 5(4): 615~629
- 8 Roweis S, Winfree E. On the reduction off errors in DNA computation [J]. Journal of Computational Biology, 1999, 6(1): 65~75
- 9 Maley C C. DNA computation: Theory, practice and prospects [J]. Evolutionary Computation, 1998, 6(3): 201~229
- 10 Rozen D E, et al. Molecular Computing: Does DNA Compute? [J]. Current Biology, 1996, 6(3): 254~257
- 11 Sanchez E. DNA biosoft computing [A]. In: Proc 4th Intl. Conf. on Soft Computing, 1996, 1: 30~37
- 12 Cox J C, Cohen D S, Ellington A D. The complexities of DNA computation [J]. Trends in Biotechnology, 1999, 17(4): 151~154
- 13 Liu Q, Liman W. DNA computing on surfaces [J]. Nature, 2000 (403): 175~179
- 14 Wasiewicz P. DNA computing: implementation of data flow logical operations [J]. Future Generation Computer Systems, 2001, 17(4): 361~378
- 15 李人厚, 余文. 关于 DNA 计算的基本原理与探讨 [J]. 计算机学报, 2001, 24(9): 972~978
- 16 卢开澄, 卢华明. 图论及其应用 [M]. 北京: 清华大学出版社, 1995