

# 基于生物芯片的背包问题 DNA 算法<sup>\*</sup>

朱莹<sup>1</sup> 丁永生<sup>1,2</sup> 李汪根<sup>1</sup>

(东华大学信息科学与技术学院 上海 201620)<sup>1</sup>

(数字化纺织服装技术教育部工程研究中心 上海 201620)<sup>2</sup>

**摘要** 通过生物芯片上的 DNA 算法求解背包问题。先将给定问题的约束条件进行分解,然后将物品重量映射为 DNA 序列,再依次在设计好的生物芯片上进行链接反应、凝胶电泳、探针检测和放射自显影,最后得到问题的解。本文的工作是在生物芯片上实现 DNA 算法,求解优化问题的一次有益尝试。

**关键词** DNA 计算,生物芯片,背包问题,链接反应,凝胶电泳,DNA 探针,放射自显影

## Knapsack Problem with DNA Algorithm Based on Biochips

ZHU Ying<sup>1</sup> DING Yong-sheng<sup>1,2</sup> LI Wang-gen<sup>1</sup>

(College of Information Sciences and Technology, Shanghai 201620, China)<sup>1</sup>

(Engineering Research Center of Digitized Textile & Fashion Technology, Ministry of Education Donghua University, Shanghai 201620, China)<sup>2</sup>

**Abstract** This paper focuses on the lab-on-a-chip DNA computation to solve the Knapsack Problem (KP). Firstly, we partition the constraint of KP into several different constraints. Secondly, we design some oligonucleotides corresponding to the weight of the items and other ligation splints, with which the combinations of the items can happen randomly. Thirdly, we design a biochip, on which there happens the ligation, Gel Electrophoresis (GE), probe detection and autoradiograph. Finally, we get the optimal solutions to the given problem. We take the initiative work in realizing DNA algorithm for optimal research on biochips.

**Keywords** DNA computing, Biochips, Knapsack problem (KP), Ligation, Gel electrophoresis (GE), DNA probe, Autoradiograph

## 1 引言

DNA 计算,是基于生化反应的一种全新的计算模式。DNA 计算方法以及计算理论的研究对 DNA 计算机的实现具有重要意义,自 Adleman<sup>[1]</sup>运用 DNA 的生化反应特性成功地设计并实现了第一个分子水平计算之后,利用 DNA 分子实现分子水平计算的研究便成为一个热点课题<sup>[2-5]</sup>。近年来,由于生命科学的发展以及人类基因组计划的推动,对 DNA 序列研究、DNA 分子计算的研究而言,由于新的实验方法、检测技术的不断发展(如表面化学方法的引入),促使 DNA 分子计算的研究不断取得令人瞩目的新进展<sup>[6-9]</sup>。

背包问题(Knapsack Problem, KP)是运筹学中一个典型的优化难题,在预算控制、项目选择、材料切割、货物装载等实践中有重要应用,并且还常常作为其他问题的子问题加以研究,例如:卫星通信;共享计算机系统设计,其中的每个进程导致附加的运行开销;购物中心的规划,其中售货点的类型与规模的设计均应权衡<sup>[8]</sup>。随着网络技术的不断发展,背包公钥密码在电子商务中的公钥设计中也起着重要的作用。KP 是一个 NP 困难问题,关于其在电子计算机上的算法已有很多人做过研究<sup>[9]</sup>。最近,文献<sup>[12]</sup>设计了一种求解标准背包问题的 DNA 算法。本文在文献<sup>[12]</sup>算法的基础上提出了一种

基于生物芯片的求解背包问题的 DNA 算法,通过基本的生物操作求出问题的解。

## 2 背包问题的数学模型

背包问题的数学模型为:

$$\max f(x_1, x_2, \dots, x_n) = \sum_{j=1}^n c_j x_j \quad j=1, 2, \dots, n$$

$$\text{s. t.} \quad \sum_{j=1}^n a_{ij} x_j \leq b_i \quad i=1, 2, \dots, m; x_j \in \{0, 1\}$$

式中,  $n$  为物品的编号;  $m$  为资源的编号;  $c_j$  为第  $j$  个物品的收益量;  $b_i$  成为第  $i$  种资源的预算;  $a_{ij}$  为第  $j$  个物品占用第  $i$  种资源的量;  $x_j$  为 0-1 决策变量(当物品  $j$  被选择时  $x_j = 1$ , 否则  $x_j = 0$ )。

背包问题的语言描述可以这样:现有  $j(j=1, 2, \dots, n)$  个物品,每个物品将会消耗  $m$  种资源  $a_{ij}(i=1, 2, \dots, m)$ , 如果将物品  $j$  装入背包将会获益  $c_j$ , 与此同时,要求所有装入背包的物品消耗的资源  $I$  不能超过  $b_i$ 。

背包问题在实践中有广泛的应用背景。许多简单结构的有机组合构成了复杂结构,对简单问题的深入探索也使复杂问题的解决变得相对容易。在设计解决大量的复杂组合优化问题算法时,背包问题往往作为子问题出现。背包问题算法改进,对复杂组合优化问题算法的改良是十分有益的。

<sup>\*</sup>国家自然科学基金重点项目(60534020);教育部新世纪优秀人才支持计划(NCET-04-415);教育部科技创新工程重大项目培育资金项目(706024);上海市国际科技合作基金项目(061307041)。朱莹 博士研究生,从事 DNA 计算机等研究;丁永生 博士,教授,博士生导师,从事智能系统、网络智能、DNA 计算、人工免疫系统、生物网络结构、生物信息学、数字化纺织服装、智能决策与分析等研究。

### 3 基于生物芯片的背包问题的 DNA 算法

根据上节给出的背包问题的模型,在文献[12]的基础上,我们提出求解背包问题的 DNA 计算方法,下面结合背包问题的具体案例来说明。

有一组物品  $S$ , 共有  $n=9$  件, 其中第  $i$  件重  $w_i$ , 价值  $v_i$ , 从  $S$  中取出一些物品出来装背包, 使总价值最大, 而不超过总重量的给定上限 30 千克, 如表 1 所示。

表 1 背包物品的重量和价格

第 $i$ 件物品	1	2	3	4	5	6	7	8	9
$w_i$ (千克)	2	1	1	2.5	10	6	5	4	3
$v_i$ (元)	10	45	30	100	150	90	200	180	300

具体设计分步的 DNA 计算方法如下:

步骤 1 对运算对象编码, 建立完全数据池;

对于物品  $i$ , 对应地给出一条寡聚核苷酸片断  $D_i$ , 其长度对应物品  $i$  的重量  $w_i$ 。为了在生物芯片上实现 DNA 算法, 我们将其编码设定为如图 1 所示, 其中  $w_i$  对应物品的重量, 值为重量的 20 倍减去 10, 是双链 DNA 片段, 两边的粘性末端是 5bp 长度的碱基, 用以区分每个物品 (5bp 原则上可以表示  $4^5$  种物品, 如果物品数量很多, 可增加粘性末端的长度)。

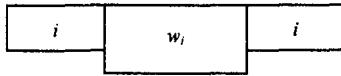


图 1 物品  $i$  的寡聚核苷酸片断  $D_i$

以  $D_i$  粘性末端的补链和  $D_j$  的粘性末端的补链相结合形成的寡聚核苷酸链片段作为链接反应的链接单元, 记为  $\overline{D_{i \rightarrow j}}$ 。

这样  $D_i$  与  $\overline{D_{i \rightarrow j}}$  在连接酶的作用下进行链接反应时, 代表各种物品的  $D_i$  将以  $\overline{D_{i \rightarrow j}}$  为媒介随机连接。为确保每件物品最多只能装入背包一次,  $\overline{D_{i \rightarrow j}}$  应满足  $\{i, j \in [1, 2, 3, 4, 5, 6, 7,$

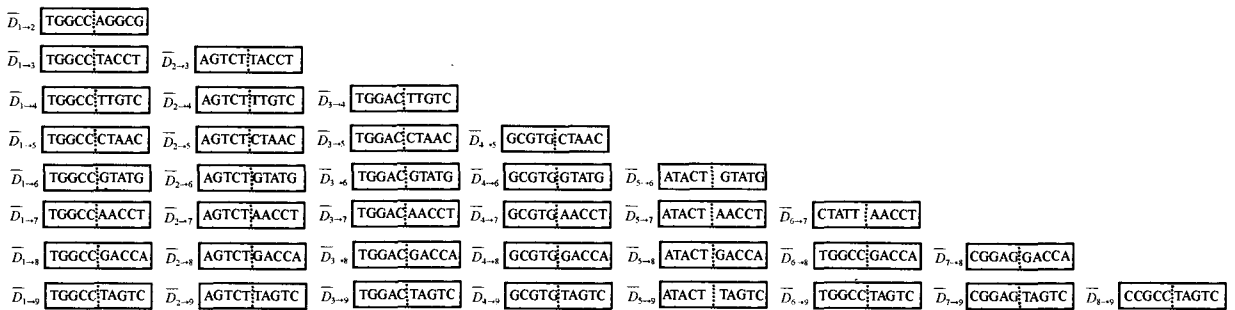


图 3 背包中任意物品组合的单链 DNA

步骤 2 通过芯片电泳来确定解对应的 DNA 片断, 进行电泳分离;

上步反应的结果为线状 DNA, 由于线状 DNA 在凝胶基质中的迁移率与其碱基对数量成对数反比, 因此只要回收迁移率小于或等于 600 的所有 DNA 片段即可。对每种 DNA 片段的回收方法为: 切下含有该种 DNA 片段的琼脂凝胶块, 在酚存在的条件下将其冻结, 使凝胶变性, 离心后, 含有 DNA 片段的电泳缓冲液可从变性胶中析出, 用酚、氯仿、异戊醇再抽提一次, DNA 片段用乙醇沉淀并经离心回收, 然后经过分离、提纯作为下一步的计算元素。

8, 9}], 其中  $i < j$ 。

具体地, 我们将  $D_1, D_2, D_3, D_4, D_5, D_6, D_7, D_8, D_9$  设计为如图 2 所示的形式。

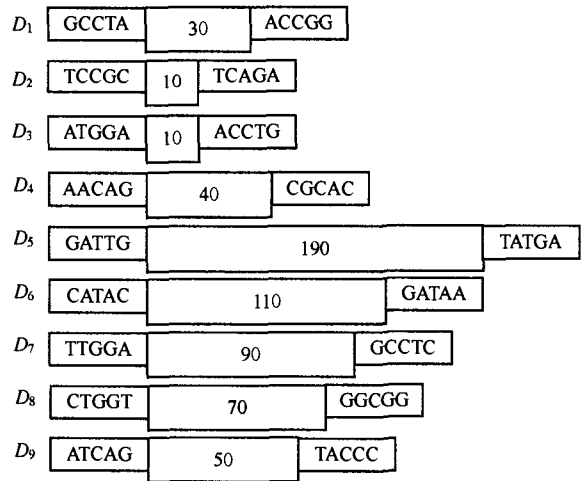


图 2  $D_i$  的表示形式

同时合成 36 种互不相同的连接媒介片段:

- $\overline{D_{1 \rightarrow 2}}, \overline{D_{1 \rightarrow 3}}, \overline{D_{1 \rightarrow 4}}, \overline{D_{1 \rightarrow 5}}, \overline{D_{1 \rightarrow 6}}, \overline{D_{1 \rightarrow 7}}, \overline{D_{1 \rightarrow 8}}, \overline{D_{1 \rightarrow 9}},$
- $\overline{D_{2 \rightarrow 3}}, \overline{D_{2 \rightarrow 4}}, \overline{D_{2 \rightarrow 5}}, \overline{D_{2 \rightarrow 6}}, \overline{D_{2 \rightarrow 7}}, \overline{D_{2 \rightarrow 8}}, \overline{D_{2 \rightarrow 9}},$
- $\overline{D_{3 \rightarrow 4}}, \overline{D_{3 \rightarrow 5}}, \overline{D_{3 \rightarrow 6}}, \overline{D_{3 \rightarrow 7}}, \overline{D_{3 \rightarrow 8}}, \overline{D_{3 \rightarrow 9}},$
- $\overline{D_{4 \rightarrow 5}}, \overline{D_{4 \rightarrow 6}}, \overline{D_{4 \rightarrow 7}}, \overline{D_{4 \rightarrow 8}}, \overline{D_{4 \rightarrow 9}},$
- $\overline{D_{5 \rightarrow 6}}, \overline{D_{5 \rightarrow 7}}, \overline{D_{5 \rightarrow 8}}, \overline{D_{5 \rightarrow 9}},$
- $\overline{D_{6 \rightarrow 7}}, \overline{D_{6 \rightarrow 8}}, \overline{D_{6 \rightarrow 9}},$
- $\overline{D_{7 \rightarrow 8}}, \overline{D_{7 \rightarrow 9}},$
- $\overline{D_{8 \rightarrow 9}}$

(注意,  $i < j$ , 且为上面 5bp 的补链, 组合在一起)。将这些寡聚核苷酸链片段放在一起进行链接反应即可生成代表任意物品组合的 DNA 链。这 36 个单链 DNA 片段如图 3 所示。

在给出芯片反应图之前, 先说明求解方法实验操作的流程图, 如图 4 所示。

步骤 3 通过生化检测求得可行解;

分别以  $D_1, D_2, D_3, D_4, D_5, D_6, D_7, D_8, D_9$  的补链为 DNA 探针模板合成 9 种带有放射性同位素标记的探针, 记为  $\overline{D_1}, \overline{D_2}, \overline{D_3}, \overline{D_4}, \overline{D_5}, \overline{D_6}, \overline{D_7}, \overline{D_8}, \overline{D_9}$ 。同时利用  $D_1, D_2, D_3, D_4, D_5, D_6, D_7, D_8, D_9$  设计探针, 并将它们分别固定于芯片上的不同的储液池。将上步中得到的每一种 DNA 分子加热解链形成单链分子, 再分别同 9 种探针  $\overline{D_1}, \overline{D_2}, \overline{D_3}, \overline{D_4}, \overline{D_5}, \overline{D_6}, \overline{D_7}, \overline{D_8}, \overline{D_9}$  反应, 则探针  $\overline{D_i}$  将与含有  $D_i$  的单链分子杂交。

为了将多余的探针分子除去,我们将与探针 $\bar{D}_i$ 反应后的溶液通过刻蚀的通道进入反应池,探针 $D_i$ 则通过高精度机械手开启的微泵流入反应池,经过充分反应后,多余的探针 $\bar{D}_i$ 将与 $D_i$ 杂交而固定在通道壁上,经过冲洗后,再将反应池中的溶液驱动进入检测通道,过一段时间后若该通道内有自显影现象发生,则所检测的DNA分子中含有 $D_i$ 。如果某种DNA分子含有寡聚核苷酸片断 $D_i$ ,则其对应的解的第 $i$ 个分量 $x_i=1$ ,否则 $x_i=0$ 。由于对所有DNA分子的检测可以同时进行,因此计算具有极大的可并行性。

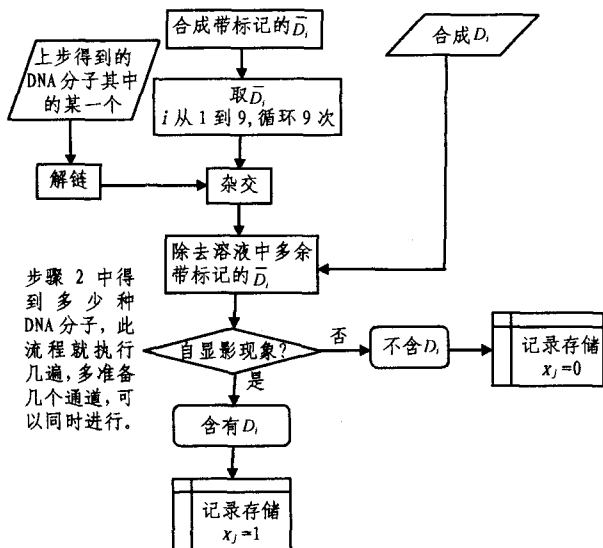


图4 生物芯片上求解方法流程图

步骤4 取得背包问题的可行解集S;  
 步骤5 比较S中各个可行解对应的目标函数值,进而得到最优解为(1,1,1,1,1,0,1,1,1),其对应的目标函数值为570。

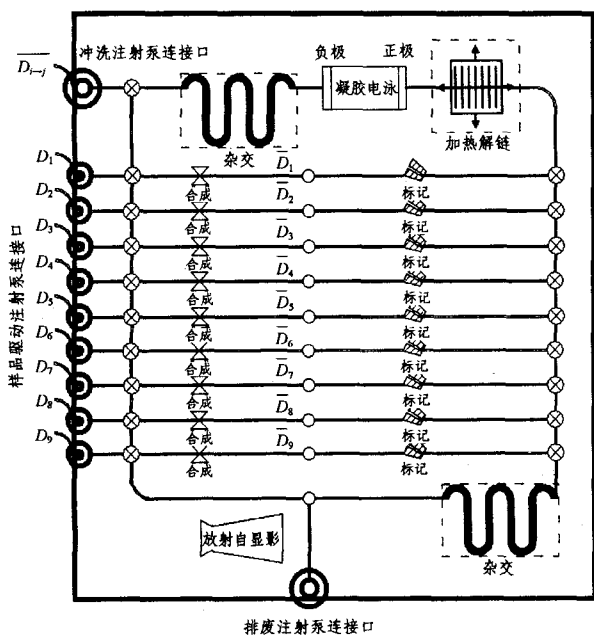


图5 生物芯片上背包问题DNA算法设计示意图

结果为:取其中的第1,2,3,4,5,7,8,9件物品放入背包中,总质量等于28.5千克。

芯片上进行的反应如图5所示。

**结束语** 本文提出了基于生物芯片的求解背包问题的DNA算法。先设计一系列与物品重量相对应的寡聚核苷酸片段及其连接模板,在连接酶的作用下将它们进行链接反应,生成代表任意物品组合的DNA链;再通过芯片电泳、探针检测以及放射自显影技术筛选出可行解;最后比较各个可行解对应的目标函数值,进而得到最优解。

为了说明问题方便,文中的示例规模较小,对于规模较大的问题其编码方法和算法的求解过程完全是一致的。但随着问题规模的增大,下列问题将值得进一步研究:

(1)可行解怎样与其它解分离,怎样输出,这是一个技术性极强的问题。尽管现代分子生物学提供了像PCR、高效电泳、亲和层析、探针检测等技术,但所消耗的时间和空间复杂性远比在此前所进行的反应过程复杂得多,特别是随着求解问题规模的增大,“输出技术”瓶颈可能成为DNA计算实现的主要障碍;

(2)运算过程中的错误发生与传播,会导致一些“伪解”出现,并在整个解空间中传播,这样就增大了可行解输出的难度。

参考文献

- [1] Adleman L. Molecular Computation of Solutions to Combinatorial Problems [J]. Science, 1994, 266(5187): 1021-1023
- [2] Lipton R J. DNA Solution of Hard Computational Problems [J]. Science, 1995, 268(5210): 542-545
- [3] Ouyang Q. DNA Solution of the Maximal Clique Problem [J]. Science, 1997, 278: 446-449
- [4] Liu Qinghua, Wang Liman, Anthony G, et al. DNA computing on surfaces [J]. Nature, 2000, 403: 175-179
- [5] Lila Kari, Greg Gloor, Sheng Yu. Using DNA to Solve the Bounded Post Correspondence Problem [J]. Theoretical Computer Science, 2000, 231: 193-203
- [6] Bernard Y, Mills Allen P Jr, Cheng Siulai. DNA Implementation of Addition in Which the Input Strands are Separate from the Operator Strands [J]. Biosystems, 1999, 52: 165-174
- [7] Landweber L F, Lila K. The Evolution of Cellular Computing: Nature's Solution to a Computational Problem [J]. Biosystems, 1999, 52: 3-13
- [8] Ji S. The Cell as the Smallest DNA-Based Molecular Computer [J]. Biosystems, 1999, 52: 123-133
- [9] Shin S Y, et al. Multi-objective Evolutionary Optimization of DNA Sequences for Reliable DNA Computing [J]. IEEE Transaction on Evolutionary Computation, 2005, 9(2): 143-158
- [10] Pferschy U, Pisinger D, Woeginger G J. Simple but efficient approaches for the collapsing knapsack problem [J]. Discrete Applied Mathematics, 1997, 77(3): 271-280
- [11] Martello S, Toth P. A new algorithm for the 0-1 knapsack problem [J]. Management Sci, 1988, 34: 633-644
- [12] Shi Xiaolong, Xu Jin. A Model of DNA Computation with Knapsack Problem [J]. Computer Engineering and Application, 2003, 27: 44-52