

文章编号:1004-1478(2011)04-0094-04

基于二维 DNA 分子 tiler 自组装求解最大团问题

冯婷¹, 张倩¹, 董亚非^{1,2}

(1. 陕西师范大学 生命科学学院, 陕西 西安 710062;

2. 陕西师范大学 计算机科学学院, 陕西 西安 710062)

摘要:针对常用算法在求解完全 NP 问题中最大团问题时,存在实验操作步骤过多、活体内不易操作以及环化效率不高等问题,设计了一种用二维 DNA(k -臂 DNA 分子)结构来解决最大团问题的方法.该方法将二维 DNA 分子设计为分子 tiler,通过二维 DNA 分子构建三维 DNA 图结构并建立计算模型,以减少解决问题所需的时间和步骤.该算法是求解最大团问题的一种可以降低复杂度的新算法,对 DNA 计算和 DNA 计算机的研究是一次有意义的实践.

关键词:DNA 计算;最大团问题; k -臂 DNA 分子;纳米金 DNA 探针

中图分类号:TP301

文献标志码:A

The solution to the maximum clique problem based on two-dimensional DNA tiles self-assemble

FENG Ting¹, ZHANG Qian¹, DONG Ya-Fei^{1,2}

(1. College of Life Sci., Shaanxi Normal Univ., Xi'an 710062, China;

2. College of Comp. Sci., Shaanxi Normal Univ., Xi'an 710062, China)

Abstract: Because there are some disadvantages of traditional algorithms in solving maximum clique problem—one of the NP complete problems, such as too many experimental steps, uneasiness to operate in vivo, and low cyclization efficiency, a 2D DNA(k -arm DNA molecule) structure was designed to solve the maximum clique problem. In order to reduce the time and steps required in resolving this problem, the method is that the 2D DNA molecular was designed as molecular tiles and the computation model was obtained by using the 3D DNA tiles to build the 3D DNA structure chart. This algorithm reduces the complexity of calculation for solving the maximum clique problem, and it is also a meaningful practice for DNA computing and DNA computer research.

Key words: DNA computing; maximum clique problem; k -arm DNA molecule; nanometer gold DNA probe

收稿日期:2011-05-06

基金项目:国家自然科学基金项目(60970005);陕西师范大学2008年重点项目

作者简介:冯婷(1987—),女,宁夏回族自治区中卫市人,陕西师范大学硕士研究生,主要研究方向为生物计算和生物信息学.

通信作者:董亚非(1963—),男,陕西省西安市人,陕西师范大学副教授,主要研究方向为生物计算、优化算法、生物信息学.

0 引言

作为自然选择的遗传物质, DNA 具有纳米级、超强并行操作能力以及特异性杂交识别等天然特性,这些特性使 DNA 计算在解决困难数学问题方面具有潜在的优势. 尤其是在海量信息存储方面,相比硅材料, DNA 分子具有显著的优势^[1-2].

自 1994 年美国计算机科学家 L. M. Adleman^[3]成功地运用线性 DNA 分子退火和链接方法解决了一个有向 Hamilton 路径问题后,研究者们将 DNA 和其他生物分子应用于计算领域,并进行了不断的探索.

最大团问题是一个困难的计算问题,它是 R. M. Karp^[4]提出的 21 个 NP 完全问题之一. 当前常用的近似求解最大团的启发式算法有遗传算法、神经网络、模拟退火算法、禁忌搜索算法等^[5]. 近年来,研究者对最大团问题的分子求解方法进行了很多有益的尝试,如基于粘贴模型的最大团问题算法、质粒 DNA 算法、闭环求解最大团问题算法等,但 these 方法均存在实验操作步骤过多、活体内不易操作以及环化效率不高等弊端. 鉴于此,本文拟通过 k -臂 DNA 分子构造空间结构的方法来解决这一 NP 问题.

1 最大团问题

给定一个无向图 $G = (V, E)$, 其中 $V = \{1, 2, \dots, n\}$ 是顶点的集合, $E \subseteq V \times V$ 是它的边的集合. 一个团是顶点集 V 的一个子集, 记为 $C, C \subseteq V$, 要求 C 中任意 2 个顶点之间有边相连. 也就是说, 由团 C 中顶点及其边组成的子图是完全图. 最大团问题就是要找到具有最多元素数的子集 C , 而最大团问题的目标就是要找到给定图的最大团^[6].

设 $t: (0, 1)^n \rightarrow 2^n, t(x) = \{i \in V: x_i = 1\}, \forall x \in (0, 1)^n, \forall s \in 2^n$, 则

$$x = t^{-1}(S) = x_i \quad i = \{1, 2, \dots, n\}$$

其中, $x_i = \begin{cases} 1 & i \in S \\ 0 & i \notin S \end{cases}, n$ 为图的顶点数.

$$\min f(x) = - \sum_{i=1}^n x_i \quad (1)$$

$$\text{St } x_i + x_j \leq 1, \forall (i, j) \in \bar{E}, x \in (0, 1)^n$$

如果 x^* 是式①的最优解, 那么集合 $C = t(x^*)$ 是图 G 的一个最大团, 且 $|C| = -f(x^*)$.

2 k -臂 DNA 分子及其 DNA 编码

k -臂 DNA 分子是 DNA 二维结构的一种. 用 k -臂 DNA 分子和一般的线性双螺旋 DNA 分子构造图^[7]的结构自然界已存在, 如在大肠杆菌同源重组过程中有一种叫做 Holliday 中间体的结构^[8]. 而图结构的思想由 N. C. Seeman 等^[9]提出, 他们描述了不同三维 DNA 图结构的构造. 研究表明, 他们已获得的 3-臂和 4-臂 DNA 分子是相当稳定的. k -臂 DNA 分子的 3' 端为单螺旋延伸, 延伸部分一般为 30—45 对碱基. 如果某顶点的度为 k , 则用 k -臂分子作为其顶点构造块, 这种顶点构造块称为分子 tiler, 如图 1 所示. 图中以 2-臂、3-臂和 4-臂分子为例, 分别称为 2-tiler, 3-tiler, 4-tiler.

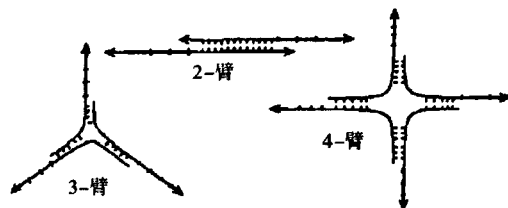


图 1 分子 tiler

经连接酶作用将这些 2-tiler, 3-tiler 和 4-tiler 分子连接在一起, 即形成图结构.

1997 年, M. Garzon 等^[10]认为生物分子计算的核心是对 DNA 序列进行编码, 从而将现实问题映射到 DNA 分子上, 再通过生物实验获得代表问题解的 DNA 分子. 因此, DNA 编码质量的优劣直接决定了 DNA 计算的效率. 图 1 中的每个顶点可以映射为一定长度的 DNA 链, 如前所述, 3' 端为单螺旋延伸, 延伸部分一般为 30—45 对碱基, 且与相关联顶点的一个臂的单链互补. 参照文献^[11], 在设计顶点时, 为了检测, 对每个顶点设计相应的探针, 本文选用金属纳米粒子标记的探针. 纳米粒子的激发光谱宽, 且连续分布, 而发射光谱呈对称分布且宽度窄, 颜色可调, 即不同大小的纳米粒子能被单一波长的光激发而发出不同颜色的光, 并且光稳定性高, 不易降解, 所以可用荧光分析法对纳米粒子标记物进行测定^[12]. 每个顶点采取不同大小的纳米粒子设计探针.

3 算法及生物计算过程

3.1 算法

步骤 1: 对具有 N 个顶点的图 G 进行编码, 建立

一个完全数据池,并设计相应的纳米金探针;

步骤 2:生成图中满足条件的所有匹配;

步骤 3:找出其中符合条件的匹配并输出结果.

3.2 生物计算过程

1)模型的 DNA 编码与合成. 对图中各个顶点进行编码,进行 DNA 的合成. 按顶点度数编码 k -臂的分子 tiler,且 k -臂具有黏性末端. 文中提出的顶点构造块中 k -臂 DNA 分子的 3'端以单链延伸终止,并且和与它相关联顶点的构造块相应的一个臂的单链延伸互补^[9]. 制备纳米金探针^[13-14].

2)形成解空间. 将足量的顶点 tiler 分子混合在一个试管中,并加入连接酶及相关缓冲溶液,使之随机反应得到完整的三维 DNA 结构.

3)筛选可行解. 用核酸外切酶去除部分形成的、不完全匹配的、带有开口端的三维 DNA 结构.

4)筛选最大团. 使用凝胶电泳去除部分形成的覆盖图,从形成的三维结构可以看到,在这个三维结构里有几个环就说明图中共有几个团. 将形成的三维结构经过变性处理,会形成长度不一的单链以及单链环状 DNA. 为了检测单链环状 DNA 的顶点构成,将单链环状 DNA 进行 PCR 扩增,将扩增后的产物与 DNA 纳米金探针在合适条件下于玻片上发生杂交反应. 因为不同的顶点其相应探针的纳米粒子尺寸不同,最后采用光纤照明器照亮以识别不同尺寸纳米颗粒散射出的颜色来确定顶点的组成,从而求出最优解.

需要指出的是,在应用 k -臂 DNA 分子构建大规模的图时,可通过顶点的摄动来实现^[15],如图 2 所示. 图 2a)表示一个度为 5 的顶点 v ,在图 2b)中,该顶点分为 v, v', v'' 这 3 个顶点,每个顶点的度均为 3,同时也引入了连接 v 和 v', v' 和 v'' 的 2 条边. 原来与 v 关联的 5 条边在顶点 v 和新顶点 v', v'' 之间被分割开.

4 生物案例

笔者以图 3 所示的图 G 为例进行具体讨论.

图 3 中,顶点 1 的度为 3,与顶点 3,4,6 相连接,笔者把它构造具有三臂的分子瓦片 3-tiler,把与顶点 1 有关联的边构造一个三臂的 DNA 分子 tiler. 同样,把顶点 2,3,4,5,6 按照同样的方法构造 6 种分子 tiler.

图 4 为反应时顶点 1 形成的结构. 由图 4 可以

看到,顶点 1 与顶点 3,4,6 有链接,因此顶点 3,4,6 的分子 tiler 会和顶点 1 的分子 tiler 相连. 为了区分,图中用不同的线条分别表示顶点 1,4 和 6. 它们互补配对以后会形成一个团. 这样直到所有的团组合形成. 形成的三维结构如图 5 所示.

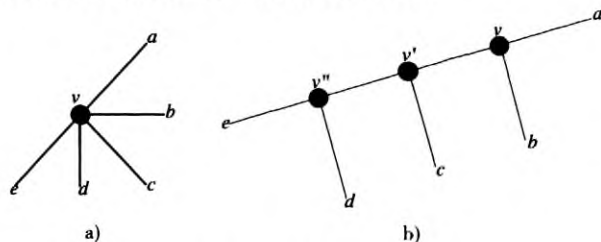


图 2 顶点的摄动

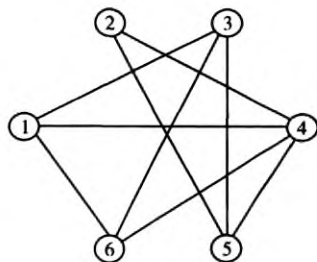


图 3 图 G



图 4 反应时顶点 1 形成的结构

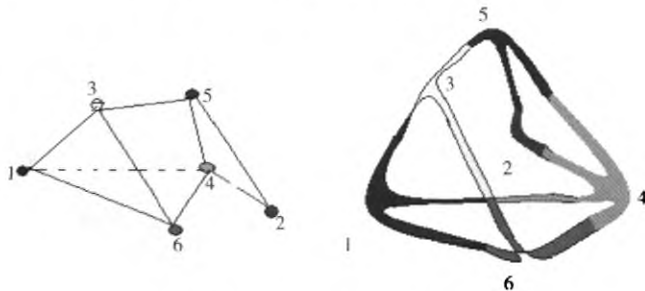


图 5 形成的三维立体结构

从图 5 可以看到,在这个三维结构里总共形成了 3 个环. 将形成的三维结构经过变性处理,会形成长度不一的单链以及单链环状 DNA. 为了检测单链环状 DNA 的顶点构成,将单链环状 DNA 进行 PCR 扩增,将扩增后的产物与 DNA 纳米金探针在合适条件下于玻片上发生杂交反应. 因为不同的顶点其相应探针的纳米粒子尺寸不同,最后采用光纤照明器

照亮以识别不同尺寸纳米颗粒散射出的颜色来确定顶点的组成,我们最后得到的最优解即为顶点为1.3.6,1.4.6以及2.4.5的团。

5 结论

目前DNA计算模型的DNA编码已经从一维结构发展到了三维结构,相对于其他的计算模型,从理论上已表明,使用三维DNA图结构可以明显减少解决问题所需的时间和步骤,而且由于它能直观地反映图结构,用它建立计算模型可能会使一些传统的难解问题得以解决。本文主要从理论上探讨使用三维DNA结构图来解决最大团问题的可行性,该算法对最大团问题来说是一种可以降低复杂度的新算法,对DNA计算和DNA计算机的研究来说,是一次有意义的实践。

但是在技术层面上,将这种二维DNA分子组装成三维DNA图结构模型还存在很多未知的领域。例如目前对复杂三维DNA结构进行普通的实验操作还少有研究,且与它相关联的酶处理操作也容易发生错误。随着科学技术的不断进步,有理由相信这种三维结构图的结构模型将得到广泛的应用,三维DNA计算也将会有更好的发展。

参考文献:

- [1] 张成,杨静,许进,等. 缩短法计算模型求解最大独立集问题[J]. 科学通报,2009,54(24):3913.
- [2] 崔光照,周君和,王延峰. DNA计算中的编码序列设计问题[J]. 郑州轻工业学院学报:自然科学版,2007,22(2/3):77.
- [3] Adleman L M. Molecular computation of solutions to combination problems[J]. Science,1994,266(11):1021.
- [4] Karp R M. Reducibility among combinatorial problems [C]//Proc of Complexity of Comp Comp, New York: Plenum Press,1972:85-103.
- [5] Flor E N. A spatiotemporal data model for incorporating time in geographic information systems[D]. South Florida: University of South Florida,2001.
- [6] 周晓晓,白杨. 关于最大团问题的一种新算法[J]. 电脑知识与技术,2008(8):708.
- [7] 张社民,方刚. 连通度问题的三维DNA结构进化算法[J]. 计算机工程与应用,2007,43(7):41.
- [8] Holliday R. Induced mitotic crossing-over in relation to genetic replication in synchronously dividing cells of *Ustilago Maydis*[J]. Genet Res,1965(10):104.
- [9] Seeman N C, Wang H, Liu B, et al. The perils of polynucleotides: the experimental gap between the design and assembly of unusual DNA structures[C]//Proc of 2nd Annual Meeting on DNA Based Computers. Washington DC: American Mathematics Society,1996:215-233.
- [10] Garzon M, Deaton R, Nino L F. et al. Genome encoding problem for DNA computing[C]//Proc of the 3rd DIMACS Workshop on DNA-based Computing. Washington DC: American Mathematics Society,1997:230-237.
- [11] 张凯,耿修堂,肖建华,等. DNA编码问题及其复杂性研究[J]. 计算机应用研究,2008(11):3264.
- [12] 孙伟,尤加宇,江宏,等. 纳米粒子标记DNA探针的制备与检测应用[J]. 中国卫生检验杂志,2005,15(8):1008.
- [13] Elghanian R, Storhoff J J, Mirkin C A. Selective colorimetric detection of polynucleotides based on the distance-dependent optical properties of gold nanoparticles[J]. Science,1997,277(5329):1078.
- [14] Alivisatos A P, Johnsson K P, Peng X G, et al. Organization of nanocrystal molecules' using DNA[J]. Nature,1996,382(6592):609.
- [15] 方刚,张社民,许进. 连通度问题的三维DNA图结构解法[J]. 系统工程与电子技术,2006,28(1):1001.