基于DNA折纸术求解图的顶点着色问题的方法

麻晶晶*1 许进2

①(山西财经大学统计学院 太原 030000)
②(北京大学信息科学技术学院 北京 100871)

摘 要: 该文基于DNA折纸术,设计了一个通过DNA折纸结构的自组装求解图的顶点着色问题的方法。利用 DNA折纸术可以构建出具有特定形状的DNA折纸结构。这些结构可以用来编码图的顶点和边,由于这些结构具 有粘性末端,因此可以通过特异的分子杂交组装成为代表了不同的图的顶点着色方案的高级结构。利用DNA-纳 米颗粒共聚体的属性和电泳等实验方法,可以筛选出正确的符合条件的图的顶点着色方案。该方法是一种高度并 行的方法,可以极大地降低求解图的顶点着色问题的复杂度。

关键词: DNA计算; DNA折纸术; DNA-纳米颗粒共聚体; 自组装; 顶点着色问题

中图分类号: O157.6 文献标识码: A 文章编号: 1009-5896(2021)06-1750-06 **DOI**: 10.11999/JEIT200203

A Method for the Graph Vertex Coloring Problem Based on DNA Origami

MA Jingjing¹ XU Jin²

^①(School of Statistics, Shanxi University of Finance and Economy, Taiyuan 030000, China)

⁽²⁾(School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China)

Abstract: Based on the DNA origami technique, a method for the graph vertex coloring problem is proposed via the self-assembly of DNA origami structures. Utilizing the DNA origami technique different DNA origami structures with specific shapes are constructed. These structures are utilized to encode the information of a graph's vertices and edges, and because these structures have sticky ends, so they can assemble to advanced structures which stands for different answers of the graph vertex coloring problem via specific molecular hybridization. Utilizing the property of DNA nanoparticle conjugation and electrophoresis as well as other experimental methods, the correct answer of the graph vertex coloring problem can be detected. This method is highly parallel, and can greatly reduce the complexity of the graph vertex coloring problem.

Key words: DNA computing; DNA origami; DNA-nanoparticle conjugation; Self-assembly; Graph vertex coloring problem

1 引言

分子计算的设想是由Feynman^[1]在20世纪60年 代初首先提出的,他指出单个分子或原子能够被用 来构建计算机的组成元件。1994年,Adleman^[2] 第1次通过生物化学实验求解了一个7个顶点的哈密 尔顿路问题,说明了DNA计算在解决复杂的数学 问题方面的能力。随着生物学和纳米技术的快速发 展,分子计算也得到了极大的发展。许多不同的方

基金项目: 国家自然科学基金(61801279)

法已经证明生物分子能够作为开发更好的计算系统 和提高计算能力的新工具。各种不同的DNA计算 模型被设计出来,如粘贴模型^[3]、剪接系统模型^[4]、 表面与芯片DNA计算模型^[5]、发夹状DNA计算模 型^[6]、质粒DNA计算模型^[7]、基于k-臂的DNA计算 模型^[8]、自组装DNA计算模型^[9]、插入一删除系统 模型^[10]、试管型DNA计算模型^[11]、图灵机DNA计 算机模型^[12]、布尔DNA计算机模型^[13]等,以及将 DNA计算的思想应用于基因分析与疾病诊断与治 疗的模型等^[14]。2016年,Xu^[15]提出了一种探针机 模型,该模型能够求解当今电子计算机无法处理的 NP完全问题。近些年来的研究表明,DNA计算在 理论、模型构建、实验方法和检测技术上都取得了 很大的进展。

收稿日期: 2020-03-24; 改回日期: 2020-08-20; 网络出版: 2020-08-27 *通信作者: 麻晶晶 casy@pku.edu.cn

Foundation Item: The National Natural Science Foundation of China (61801279)

与此同时, DNA自组装技术的研究也取得了 巨大的进展。1998年Winfree等人^[16]在"简单自组 装瓦片结构"的基础上,构建了具有足够刚性的 5种DX元件。2003年,LaBean教授的研究组构造 出一种十字形的支架结构(即四星元件)^[17],利用这 种模块组装成方形DNA自组装芯片。2006年, Rothemund等人^[18]成功地应用了一条长为7000个碱 基左右的M13噬菌体的单链,借助于200多条短的 寡核苷酸链(钉书针链),通过碱基互补,在特定部 位折叠成复杂的2维DNA自组装结构。依照 Rothemund的方法, DNA分子可以构造人们想要 的任何平面图形,因此,他的这一方法被形象地称 为"DNA折纸术"。不满足于仅仅构造2维平面的 简单图形,科学家们成功地在其它类型的DNA分 子基础上设计出了新的DNA折纸方法,构造出了 更为复杂高级3维DNA自组装结构。2011年, Han Dongran等人^[19]利用DNA折纸的折叠技术,在3维 空间中设计和构造了具有复杂的空间曲面的自组装 DNA纳米结构,如球面、椭圆球面、瓶子等。之 后,DNA折纸术的相关研究取得了很大的进展^[20-30]。

图的顶点着色问题是著名的NP完全问题。2018 年,Xu等人^[31]提出了一种新颖的图顶点着色 DNA计算模型, 该模型以一个3-着色的61个顶点的 图为例,设计了一种并行型聚合酶链反应(Polymerase Chain Reaction, PCR)操作技术,应用这种技术一 次可以对图中多条边来删除非解, 使得生物操作次 数大大减少,极大地提高了运行速度。实验表明, 99%的非可行解在构建初始解空间时就被删除,并 利用DNA自组装和并行PCR方法,通过识别、拼 接以及组装等技术得到解。该模型还通过如下3种 方法来克服解空间指数爆炸问题:顶点颜色集的确 定方法; 子图分解方法以及子图中的顶点优化排序 方法。本文提出了一种基于DNA折纸术求解图的 顶点着色问题的方法,利用DNA折纸术可以构建 出具有特定形状的DNA折纸结构。这些结构可以 用来编码图的顶点和边,由于这些结构具有粘性末 端,因此可以通过特异的分子杂交组装成代表不同 的图的顶点着色方案的高级结构。利用DNA-纳米 颗粒共聚体的属性和电泳等实验方法,可以筛选出 正确的符合条件的图的顶点着色方案。这种方法是 一种高度并行的方法,可以极大地降低求解图的顶 点着色问题的复杂度。

2 图的顶点着色问题

本文所言之图皆指有限、无自环、无圈的简单 无向图,通常用G表示图。本文用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,使得相邻的顶点着 不同的颜色。换言之,是指对图G中的顶点集 V(G)的划分: $V(G)=V_1 \cup V_2 \cup \cdots V_k, V_i \neq \varphi$ (空集), $V_i \cap V_i = \varphi, i = 1, 2, \dots, k$ 。满足图G正常着色的最小 的颜色数称为色数,记为 $\chi(G)$ 。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[k≥ $\chi(G)$]种颜色对图G进行着色。通常用 $C_k(G)$ 表示图 G的全体k着色构成的集合。由于本文仅考虑图的 3-着色问题, 故k=3, 我们总是假定颜色集为 $C_3(G) = \{r, b, y\}$,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: $f: V(G) \rightarrow \{r, b, y\}$, 使得对于∀ $uv \in E(G)$, $f(u) \neq f(v)$ 。图的3-着色问题是一个NP-完全问 题。本文以一个具有6个顶点的简单图(图1)为例说 明所设计的基于DNA折纸术求解图的顶点着色问 题的算法。

3 基于DNA折纸术求解图的顶点着色问题 的算法

DNA折纸术最先由Rothmund提出,这种新的 DNA自组装策略是DNA纳米技术的巨大进步,依 据这种自组装方法,不仅2维平面的不同图形,如 正方形、圆形、笑脸等形状可以被构造出,而且 3维立体的高级结构也可以被构造出。本文利用DNA 折纸术构造的不同形状的结构对图G的每个顶点进 行编码,通过DNA分子的杂交和一系列实验方法 获得图G顶点着色的正确解,由于DNA折纸结构可 以在原子力显微镜下被观测出,所以图G的可行的 顶点着色方案可以通过观察原子力显微镜而得到。 本文设计的基于DNA折纸术求解图的顶点着色问 题的算法包含以下4个步骤: (1)用DNA折纸结构来 编码图G的信息; (2)将DNA折纸结构进行退火反 应,通过自组装过程生成问题的所有可能的解; (3) 删除非解,并提取出正确的解:(4) 识别结果。 以下是具体方法的描述。

3.1 编码

本文所使用的图如图1所示,它包括6个顶点、



9条边。图2中以顶点1和顶点2为例展示了本文用 DNA折纸结构编码图G的顶点和边的具体方法。由 于本文求解的是图的3-着色问题,所以每个顶点都 有可能染3种颜色,即红色、黄色和蓝色。为此以 DNA折纸结构的不同形状来代表这3种颜色,即如 图2所示,正方形的DNA折纸结构代表顶点被染成黄色,三角 形的DNA折纸结构代表顶点被染成黄色,三角 形的DNA折纸结构代表顶点被染成蓝色。每个 DNA折纸结构上还具有突出的单链DNA,即粘性 末端,这些粘性末端能依照图G中顶点的邻接关系 相互杂交,粘性末端杂交的原理即DNA的两条单 链在合适的实验室条件下依照碱基互补配对的规则 (A与T 配对,G与C 配对)杂交形成双螺旋结构。

例如顶点1与顶点2、顶点3之间有边相连,所 以顶点1有两条突出的粘性末端,分别能与顶点2、 顶点3上的粘性末端杂交。按照这种方法,每个顶 点都要设计正方形、圆形和三角形3种不同的DNA 折纸结构,并且需要设计好需要的粘性末端序列, 每个顶点所需要设计的粘性末端的数目恰好等于 该顶点的度。同时,DNA折纸结构上还需要设计 代表顶点序号的标记,以便在原子力显微镜下分辨 顶点。

3.2 退火

设计好DNA折纸结构之后,就需要将所有代 表顶点和边的信息的DNA折纸结构进行退火反应, 这个过程由DNA粘性末端序列中碱基的互补配对 完成,这是一个随机自由组合的反应,也是一个高 度并行的过程,这种高度的并行性能在非常短的时 间内形成问题的所有可能的解。图3中以顶点1和顶 点3为例展示了DNA折纸结构的杂交方式。因为顶 点1和顶点3有边相连,所以代表顶点1和顶点3的 DNA折纸结构的其中两条突出的粘性末端可以互 相配对杂交,杂交的方式如图中形成的双链所示。 值得注意的是,由于顶点被赋予了颜色,所以两个 顶点结构杂交之后可能是颜色相同的,也可能是颜 色不同的,如图3(a)所示的杂交结构为顶点1和顶 点3都是红色,如图3(b)所示的杂交结构是顶点1为 黄色,顶点3为红色。有边相连的顶点被染成相同 的颜色是不符合图的顶点染色问题的要求的,所以 这种错误的解需要被删除掉。为了实现这个目的, 本文在设计DNA折纸结构时预先给每一种颜色的 DNA折纸结构设计了一段代表其颜色的特异序列, 如图3中所示,在两个顶点结构的双链杂交部分的 两侧即为本文预先设计的代表顶点颜色的特异序 列,如图3(a)所示的顶点杂交结构中,顶点1和顶 点3都被染成了红色,所以特异序列也是代表红色 的序列,图中用红线表示;而如图3(b)所示的顶点 杂交结构中,顶点1染成了黄色,顶点3为红色,所 以特异序列也为相应的黄色和红色序列,图中分别 用黄线和红线表示。可以想象,当顶点结构的粘性 末端没有杂交形成双螺旋结构之前,这些特异序列 (图3中的短的红色和黄色线段)是游离的而且靠近 的几率很小,只有当顶点结构的粘性末端杂交形成 双螺旋之后这些特异序列才会被拉近距离并且靠得 很近无法分开。这样的设计将在后面删除非解的步 骤中发挥作用。

3.3 删除非解并提取正确的解

经过DNA折纸结构的退火之后,图G的顶点三



图 3 DNA折纸结构的退火方式

着色问题的所有可能的解都已经形成,接下来的操 作就是删除所有的非解,即删除掉退火结构中所有 的顶点有边相连并且被染成相同颜色的染色方案, 这个过程通过与DNA-纳米颗粒共聚体的杂交和琼 脂糖凝胶电泳分离来实现。图4展示了用DNA-纳米 颗粒共聚体与退火结构进行杂交的方法。如图4所 示,金纳米颗粒上修饰了能与代表顶点颜色的两段 特异序列杂交的互补序列,值得注意的是这段互补 序列正好是特异序列的两倍,当且仅当两个杂交的 顶点结构为染相同的颜色时,代表顶点颜色的两段 特异序列才能被拉近距离,这样就能与DNA-纳米 颗粒共聚体上的互补序列杂交。图4中以顶点1和顶 点3为例展示了DNA-纳米颗粒共聚体与退火结构进 行杂交的方法。顶点1和顶点3有边相连,它们有 3种被染成相同颜色的情况,即同为红色、同为黄 色和同为蓝色。如果退火结构中任意两个有边相连 的点都染异色,那么DNA-纳米颗粒共聚体就不会 与其杂交,这是因为本文设计的特异序列比较短, 依照双链DNA分子的特性,在实验所需的温度条 件下,单独一段长度较短的特异序列无法与互补序 列杂交形成双链,只有两段特异序列组成的长序列 才能与互补序列杂交形成双链。



图 4 DNA-纳米颗粒共聚体与退火结构的杂交

经过与DNA-纳米颗粒共聚体杂交之后,就要 将产物进行琼脂糖凝胶电泳。在琼脂糖凝胶电泳中 杂交了DNA-纳米颗粒共聚体的退火结构由于分子 量较大在电泳的凝胶中处于落后的位置,没有杂交 DNA-纳米颗粒共聚体的退火结构由于分子量较小 处于领先的位置,就可以被回收回来,即为正确的 解,将会用于下一步的识别和分析。

3.4 识别结果

由于DNA折纸结构可以通过原子力显微镜观 察到,因此对于上一步回收的退火结构,可以用原 子力显微镜进行观察,以确定最终的图G的顶点三 着色方案。由于红色、黄色、蓝色分别由正方形、 圆形和三角形DNA折纸结构表示,因此在原子力 显微镜下就可以通过辨认折纸结构的形状和顶点的 标记确认正确答案。图5显示了一种杂交结构,这 是一种正确的图G的顶点三着色方案。

4 结论

图的顶点着色问题有着非常广泛的理论与应用



图 5 图G的顶点三着色方案的正确解

研究意义,各种DNA计算模型都将其作为研究的 目标。DNA自组装技术为DNA计算提供了新的方 法,各种DNA自组装计算模型也被用于解决各种 不同的NP完全问题。本文中,我们利用DNA折纸 结构编码图的信息,通过DNA折纸结构上粘性末 端的自组装过程,构建了一种图的顶点着色问题的 非确定性算法。通过构建数以亿计的参与计算的 DNA折纸结构,该算法可以并行地测试每种可能 的顶点着色方案。对于给定的n个顶点和m条边的 图,顶点着色方案存在与否以及存在几种顶点着色 方案都可以通过相应的DNA折纸结构的自组装过 程和一系列实验方法计算出来,计算结果可以直接 通过原子力显微镜观察到,这种方法具有高度的并 行性,类似的基于DNA折纸术的方法在求解NP完 全问题方面具有巨大的潜力。

参考文献

- FEYNMAN R P. There's plenty of Room at the Bottom[M]. GILBERT D H. Minaturization. New York: Reinhold Publishing Corporation, 1961: 282–296.
- [2] ADLEMAN L M. Molecular computation of solutions to combinatorial problems[J]. Science, 1994, 266(5187): 1021–1024. doi: 10.1126/science.7973651.
- [3] BRAICH R S, CHELYAPOV N, JOHNSON C, et al. Solution of a 20-variable 3-SAT problem on a DNA computer[J]. Science, 2002, 296(5567): 499–502. doi: 10.1126/science.1069528.
- [4] BENENSON Y, PAZ-ELIZUR T, ADAR R, et al. Programmable and autonomous computing machine made of biomolecules[J]. Nature, 2001, 414(6862): 430–434. doi: 10.1038/35106533.
- [5] LIU Qinghua, WANG Liman, FRUTOS A G, et al. DNA computing on surfaces[J]. Nature, 2000, 403(6766): 175–179. doi: 10.1038/35003155.
- [6] SAKAMOTO K, GOUZU H, KOMIYA K, et al. Molecular computation by DNA hairpin formation[J]. Science, 2000, 288(5469): 1223–1226. doi: 10.1126/science.288.5469.1223.
- HEAD T, ROZENBERG G, BLADERGROEN R S, et al. Computing with DNA by operating on plasmids[J]. Biosystems, 2000, 57(2): 87–93. doi: 10.1016/S0303-2647 (00)00091-5.
- [8] JONOSKA N, KARL S A, and SAITO M. Three dimensional DNA structures in computing[J]. *Biosystems*, 1999, 52(1/3): 143–153. doi: 10.1016/S0303-2647(99)00041-6.
- [9] ADLEMAN L, CHENG QI, GOEL A, et al. Running time and program size for self-assembled squares[C]. ACM Symposium on Theory of Computing (STOC), New York, USA, 2001. 740–748. doi: 10.1145/380752.380881.
- [10] MARTIN-VIDE C, PĂUN G, and SALOMAA A. Characterizations of recursively enumerable languages by means of insertion grammars[J]. *Theoretical Computer Science*, 1998, 205(1/2): 195-205. doi: 10.1016/S0304-3975(97)00079-0.
- [11] LIPTON R J. DNA solution of hard computational problems[J]. Science, 1995, 268(5210): 542–545. doi: 10.1126/ science.7725098.
- [12] ROTHEMUND P. A DNA and restriction enzyme

implementation of Turing machines[C]. Proceedings of the DIMACS Workshop, Princeton, USA, 1995.

- OGIHARA M and RAY A. Simulating Boolean circuits on a DNA computer[J]. Algorithmica, 1999, 25(2/3): 239–250. doi: 10.1007/PL00008276.
- [14] BENENSON Y, GIL B, BEN-DOR U et al. An autonomous molecular computer for logical control of gene expression[J]. *Nature*, 2004, 429(6990): 423–429. doi: 10.1038/nature02551.
- [15] XU Jin. Probe machine[J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 27(7): 1405–1416. doi: 10.1109/TNNLS.2016.2555845.
- [16] WINFREE E, LIU Furong, WENZLER L A, et al. Design and self-assembly of two-dimensional DNA crystals[J]. *Nature*, 1998, 394(6693): 539–544. doi: 10.1038/28998.
- [17] YAN Hao, PARK S H, FINKELSTEIN G et al. DNAtemplated self-assembly of protein arrays and highly conductive nanowires[J]. Science, 2003, 301(5641): 1882–1884. doi: 10.1126/science.1089389.
- [18] ROTHEMUND P W K. Folding DNA to create nanoscale shapes and patterns[J]. Nature, 2006, 440(7082): 297–302. doi: 10.1038/nature04586.
- [19] HAN Dongran, PAL S, NANGREAVE J, et al. DNA origami with complex curvatures in Three-Dimensional space[J]. Science, 2011, 332(6027): 342–346. doi: 10.1126/ science.1202998.
- [20] ISHIKAWA D, SUZUKI Y, KUROKAWA C, et al. DNA origami Nanoplate-based emulsion with Nanopore function[J]. Angewandte Chemie-International Edition, 2019, 58(43): 15299-15303. doi: 10.1002/anie.201908392.
- [21] JOURNOT C M A, RAMAKRISHNA V, WALLACE M I, et al. Modifying membrane morphology and interactions with DNA origami clathrin-mimic networks[J]. ACS Nano, 2019, 13(9): 9973–9979. doi: 10.1021/acsnano.8b07734.
- [22] LI Na, SHANG Yingxu, XU Rui et al. Precise organization of metal and metal Oxide Nanoclusters into arbitrary patterns on DNA origami[J]. Journal of the American Chemical Society, 2019, 141(45): 17968-17972. doi: 10.1021/jacs.9b09308.
- [23] JOHNSON J A, DEHANKAR A, WINTER J O, et al. Reciprocal control of hierarchical DNA origami-Nanoparticle assemblies[J]. NANO Letters, 2019, 19(12): 8469–8475. doi: 10.1021/acs.nanolett.9b02786.
- [24] KLEIN W P, THOMSEN R P, TURNER K B, et al. Enhanced catalysis from multienzyme cascades assembled on a DNA origami triangle[J]. ACS Nano, 2019, 13(12): 13677–13689. doi: 10.1021/acsnano.9b05746.
- [25] BERENGUT J F, BERENGUT J C, DOYE J P K, et al. Design and synthesis of pleated DNA origami nanotubes with adjustable diameters[J]. Nucleic Acids Research, 2019, 47(22): 11963–11975. doi: 10.1093/nar/gkz1056.

- [26] HAHN J, CHOU L Y T, SØRENSEN R S, et al. Extrusion of RNA from a DNA-origami-based Nanofactory[J]. ACS Nano, 2020, 14(2): 1550–1559. doi: 10.1021/acsnano. 9b06466.
- [27] ANASTASSACOS F M, ZHAO Zhao, ZENG Yang, et al. Glutaraldehyde cross-linking of Oligolysines coating DNA origami greatly reduces susceptibility to nuclease degradation[J]. Journal of the American Chemical Society, 2020, 142(7): 3311-3315. doi: 10.1021/jacs.9b11698.
- [28] BARTNIK K, BARTH A, PILO-PAIS M, et al. A DNA origami platform for single-pair Förster resonance energy transfer investigation of DNA-DNA interactions and ligation[J]. Journal of the American Chemical Society, 2020, 142(2): 815–825. doi: 10.1021/jacs.9b09093.
- [29] 殷志祥,唐震,张强,等.基于DNA折纸基底的与非门计算模型[J].电子与信息学报,2020,42(6):1355-1364. doi: 10.11999/ JEIT190825.

YIN Zhixiang, TANG Zhen, ZHANG Qiang, *et al.* NAND gate computational model based on the DNA origami

template[J]. Journal of Electronics & Information Technology, 2020, 42(6): 1355-1364. doi: 10.11999/JEIT 190825.

 [30] 王君珂,印珏,牛人杰,等. DNA计算与DNA纳米技术[J]. 电子与信息学报,2020,42(6):1313-1325. doi: 10.11999/ JEIT190826.
WANG Junke, YIN Jue, NIU Renjie, *et al.* DNA computing

and DNA Nanotechnology[J]. Journal of Electronics & Information Technology, 2020, 42(6): 1313–1325. doi: 10.11999/JEIT190826.

[31] XU Jin, QIANG Xiaoli, ZHANG Kai, et al. A DNA computing model for the Graph Vertex Coloring problem based on a probe graph[J]. Engineering, 2018, 4(1): 61–77. doi: 10.1016/j.eng.2018.02.011.

麻晶晶: 女, 1983年生, 讲师, 从事生物计算等研究. 许 进: 男, 1959年生, 教授, 从事生物计算等研究.

责任编辑:余 蓉