

关于DNA计算的报告

欧阳筱羿^{1*}

摘要

DNA计算是当今计算机科学和分子生物学交叉领域中的一个前沿方向，它是分子计算的一门分支，旨在利用DNA信息储存的大容量、低能耗、高并行实现计算。DNA计算自诞生之初便被用来探索求解NP完全问题的可能性，在长达20多年的发展中，还在密码学、人工智能上取得多项突破。本文从DNA的分子结构入手，直观上解释DNA能被用作计算的原因，进一步按照历史发展综述了DNA计算的理论发展和实验成果，总结DNA计算的优势和其局限性，并对其发展前景进行展望。

关键词

DNA计算

¹北京大学物理学院，1600011385

*e-mail: ipsumo@pku.edu.cn

1. 引言

DNA计算是利用生物分子实现区别传统硅基计算技术的一门分支，其研究起始于上世纪九十年代，美国南加州大学的Leonard Adleman教授首次提出并实现了用DNA分子求解七点情况下的哈密顿路径问题[1]。从Adleman的实验开始，又相继有多种图灵机被证明为可以由DNA计算实现，时至今日，也有了越来越多的实验（包括体外实验和体内实验）证据来证实DNA计算在诸多情况下的可行性。

2. DNA的分子结构和计算原理

DNA全称为脱氧核糖核酸，是几乎所有生命的遗传物质。DNA是一种生物大分子，其基本单位是脱氧核苷酸，由一分子磷酸、一分子脱氧核糖和一分子含氮碱基组成。含氮碱基有四种，分别是腺嘌呤（A）、鸟嘌呤（G）、胞嘧啶（C）和胸腺嘧啶（T）。两条脱氧核糖核酸会以右旋的方式交互缠绕成双螺旋结构，磷酸联结形成的骨架位于外侧，碱基通过氢键连接起来位于内侧，碱基对的组成遵循互补配对原则，A、T碱基对由两个氢键相连，C、G碱基对由三个氢键相连[2]。

正如同信息可以储存在一串0、1构成的二进制码里，它也同样可以编码在一串A、T、C、G构成的DNA分子中。DNA分子不仅信息存储密度大、能量耗散小、具有高度并行性，而且DNA的双链结构保证两条链严格互补，在扩繁的时候可以减少DNA复制错误带来的影响，一定程度上可以弥补DNA复制出错率远高于使用Reed-Solomon纠错码时硬盘数据拷贝出错率这一缺陷[3]。更多有

关DNA计算的优势可参见第4章。

除了DNA可以存储信息之外，还有一系列DNA酶可以使DNA变性、复制、退火等，这些酶可以视为作用在DNA序列上的算子，进行剪切、复制、粘贴等操作，例如限制性核酸内切酶相当于分离算子、连接酶相当于连接算子、聚合酶相当于复制算子、限制性核酸外切酶相当于删除算子[4]。

因此，DNA具备计算机的两个关键要素：存储和操作。理论上来说，我们可以把运算的对象映射到DNA分子链上，再将其与按特定比例混合的DNA酶相互作用，生成各种数据池，然后按照一定的规则将原始问题的数据高度并行地映射为DNA分子链的可控生化反应，最后利用分子生物技术如聚合酶链式反应等来获得运算结果。简单地说可以认为一台DNA计算机以DNA的形式输入数据，经过特定生化反应，再以DNA的形式输出运算结果。

3. DNA计算的理论发展和实验成果

1994年，Adleman教授首次提出DNA计算的概念并通过实验演示其可行性[5]。作为概念验证，他选择用哈密顿路径问题来描述自己的理论。首先，生成一系列不同且能被识别的DNA片段，每个片段代表一座城市（一个节点），在酶的作用下，不同片段可以形成连接；将这些DNA片段和酶按一定比例混合在试管中，小的DNA片段很快就开始连接形成成长的DNA链，代表不同的路径，这一步只需要数秒；最终，通过一系列操作如亲和层析、磁珠分离、电泳分离等筛选出分子量最小者，也即最短路径，

但这个分离过程会持续数天甚至一周。

受到Adleman研究的启发，人们开始认识到DNA计算具有非常高的并行度，相继有多种组合问题的DNA解法被提出：例如图的最大团问题、可满足性问题、象棋问题等等。以可满足性问题为例，1995年，Lipton提出所有NP问题都能被有效化简为哈密顿路径问题并提出了一种可满足性问题的DNA解法[6]；2000年，Sakamoto等人发现用DNA单链的发卡结构可用来探索分子自治计算的可能性，并提出可满足性问题可通过该方法得到求解[7]；同年Liu等人利用DNA计算仅用了91步就解决了可满足性问题，同样的问题传统的串行计算机则需要花160万步[8]。值得一提的是，现在任职于北京大学物理学院的欧阳颀教授曾经在1997年给出了解决图的最大团问题的一个DNA解[9]。

在实验上，DNA计算也取得了可观的进展，除了解决NP问题，还在其他方面有诸多应用：2004年，魏茨曼科学研究所的Shapiro等人构建出了一个具有输入-输出模块的DNA计算机，理论上也可以被应用在细胞内实时诊断癌症并释放抗癌药[10]；2013年，研究者们把JPEG格式的图片、莎士比亚的十四行诗集、马丁路德金的演讲词存储在DNA中[1]；2016年，哈佛医学院George Church组等人利用CRISPR基因编辑技术把一张骑手骑马飞驰的动图储存在活体细菌的DNA中，DNA数字存储技术逐渐成为研究的热点[11]；后续的研究还包括可逆DNA计算，尤其是杜克大学John Reif组分别提出利用dsDNA门和DNA发卡结构复合体的两种实现方式，都是DNA计算领域的重大突破[12][13]。

4. DNA计算的优势和限制

DNA双链上相邻核苷酸仅仅间隔0.35 nm，1 g DNA理论上能存储艾字节量级（ 10^{18} ）的信息，具有极大的信息存储密度；另外，DNA计算所消耗的能量只有一台电子计算机完成同样计算所消耗的能量十亿分之一；最重要的是DNA分子生物算法具有高度的并行性：在Adleman的实验中，据估计DNA串的并行操作数目可达 10^{14} ，尽管对于七节点数的哈密顿路径问题实验大约用了一个星期，而电子计算机只需数秒就可以完成，但如果把节点数推广到50及更高时，电子计算机却需要花上数年的时间去计算求解，此时DNA计算高并行度的优越性就凸显了出来。

但另一方面，DNA计算也有受限于自身原理的硬伤：由于运算过程中必然借助生化反应，其响应时间尺度往往在分钟、小时甚至数天的量级，显著高于电子计算机毫秒级的计算时间；另外DNA复制过程中的出错率也高于硬盘数据拷贝的出错率。尽管以上两点缺陷能一定程度上被DNA计算的高并行度和DNA双链的互补性弥补，但DNA计算仍然在数据分析、验证上比传统的电子计算机困难很多。

5. DNA计算的前景

尽管DNA计算仍然有着自身的局限性，但随着科学发现和技术革新，当时的一些瓶颈现在可能已经不复存在。二十世纪初就有DNA计算与人工智能结合的呼声，例如智能DNA芯片、基于DNA计算实现遗传算法、群体智能等等，现如今人工智能领域兴兴向荣，我们也看到了DNA计算和机器学习的实验产物，例如加州理工大学的Lulu Qian组成功构架出基于DNA计算的人工神经网络，能够识别100比特的手写数字[14][15]；另外，近年来DNA计算在密码学[16]、数据存储技术[17]上也取得了一定进展。

总而言之，DNA计算是当今计算机科学和分子生物学交叉领域中一个新颖而令人着迷的方向，它不单单是一项信息处理的新方法，还能促进各领域间的知识的交流，甚至能潜在地改变我们对计算科学理论和实践的认知。

参考文献

- [1] Wikipedia contributors. DNA computing — Wikipedia, The Free Encyclopedia, 2020. [Online; accessed 26-February-2020].
- [2] Wikipedia contributors. DNA — Wikipedia, The Free Encyclopedia, 2020. [Online; accessed 26-February-2020].
- [3] Z Ezziane. DNA computing: applications and challenges. *Nanotechnology*, 17(2):R27–R39, dec 2005.
- [4] 张军英高琳, 许进. DNA计算的研究进展与展望, volume 29. 电子学报, 2001.
- [5] Leonard M Adleman. Molecular computation of solutions to combinatorial problems. *Science*, pages 1021–1024, 1994.
- [6] Richard J Lipton. Dna solution of hard computational problems. *science*, 268(5210):542–545, 1995.
- [7] Kensaku Sakamoto, Hidetaka Gouzu, Ken Komiya, Daisuke Kiga, Shigeyuki Yokoyama, Takashi Yokomori, and Masami Hagiya. Molecular computation

- by dna hairpin formation. *Science*, 288(5469):1223–1226, 2000.
- [8] Qinghua Liu, Liman Wang, Anthony G Frutos, Anne E Condon, Robert M Corn, and Lloyd M Smith. Dna computing on surfaces. *Nature*, 403(6766):175–179, 2000.
 - [9] Qi Ouyang, Peter D Kaplan, Shumao Liu, and Albert Libchaber. Dna solution of the maximal clique problem. *Science*, 278(5337):446–449, 1997.
 - [10] Yaakov Benenson, Binyamin Gil, Uri Ben-Dor, Rivka Adar, and Ehud Shapiro. An autonomous molecular computer for logical control of gene expression. *Nature*, 429(6990):423–429, 2004.
 - [11] Seth L Shipman, Jeff Nivala, Jeffrey D Macklis, and George M Church. Crispr-cas encoding of a digital movie into the genomes of a population of living bacteria. *Nature*, 547(7663):345–349, 2017.
 - [12] Sudhanshu Garg, Shalin Shah, Hieu Bui, Tianqi Song, Reem Mokhtar, and John Reif. Renewable time-responsive dna circuits. *Small*, 14(33):1801470, 2018.
 - [13] Abeer Eshra, Shalin Shah, Tianqi Song, and John Reif. Renewable dna hairpin-based logic circuits. *IEEE Transactions on Nanotechnology*, 18:252–259, 2019.
 - [14] Lulu Qian, Erik Winfree, and Jehoshua Bruck. Neural network computation with dna strand displacement cascades. *Nature*, 475(7356):368–372, 2011.
 - [15] Kevin M Cherry and Lulu Qian. Scaling up molecular pattern recognition with dna-based winner-take-all neural networks. *Nature*, 559(7714):370–376, 2018.
 - [16] Yinan Zhang, Fei Wang, Jie Chao, Mo Xie, Huajie Liu, Muchen Pan, Enzo Kopperger, Xiaoguo Liu, Qian Li, Jiye Shi, et al. Dna origami cryptography for secure communication. *Nature communications*, 10(1):1–8, 2019.
 - [17] Luis Ceze, Jeff Nivala, and Karin Strauss. Molecular digital data storage using dna. *Nature Reviews Genetics*, 20(8):456–466, 2019.