

DNA 计算与 DNA 计算机研究进展与展望^{*}

丁建立 陈增强 袁著祉

(南开大学信息技术科学学院 天津300071)

A Survey of DNA Computing and DNA Computer

DING Jian-Li CHEN Zeng-Qiang YUAN Zhu-Zhi

(College of Information Technology and Science, Nankai University, Tianjin 300071)

Abstract DNA computing is a new method based on biochemical reactions and molecular biology technology. The paper first introduces the basic principle and advantages of DNA computing, and then surveys DNA computing and DNA computer, finally, points out current existing problems and future search directions of DNA computing and DNA computer.

Keywords DNA computing, DNA computer

1 引言

1994年,Adleman在《Science》杂志上发表开创性文章^[1],提出了DNA计算概念,并成功地解决了著名的Hamilton路径问题,还指出了DNA计算潜在的巨大并行性及待研究的问题。1995年,Lipton也在《Science》杂志上发表文章^[2],进一步论证了DNA计算可解决完全性问题。自此,DNA计算引起了国际上许多学者的关注与兴趣,并开展研究。

DNA算法是一种模拟生物分子DNA的结构并借助于分子生物技术进行计算的新方法,它开创了以化学反应作为计算工具的先例。目前研究主要涉及方面有^[3~6]:DNA计算的能力与求解某些NP问题、计算模型和算法、数据海量存储与数据加密和解密、智能算法的融合与智能控制、基于生化技术的独特应用与DNA芯片合成、Boolean电路和数据逻辑运算的仿真和实现等。

DNA计算的研究虽然刚刚开始,但进展很快。DNA计算机的国际会议从1995年开始每年一次,美国与日本的DNA计算机研究非常活跃。第一本DNA计算的专著已于1998年问世^[7]。研究表明,目前科学家已提出有多种不同的DNA计算方法,这为DNA计算机研制开辟了广阔的途径。

2 DNA 计算的原理与模型

2.1 DNA 分子结构特征

控制生物性状遗传的主要物质是脱氧核糖核酸DNA(Deoxyribonucleic Acid),DNA的基本元素是核苷酸,由于化学结构的不同,核苷酸划分为四类碱基:腺嘌呤(A)、鸟嘌呤(G)、胞嘧啶(C)和胸腺嘧啶(T)。DNA由两条极长的核苷酸键组成,这两条极长的核苷酸键利用碱基之间的氢键而结合在一起,形成一条双股的螺旋结构,且一股中的碱基序列与另一股中的碱基序列互补,A和T配对,C和G配对,每个染色体是一段双股螺旋的DNA,遗传信息以A、T、C、G在核苷酸中的排列次序而体现,其排列序列的多样性构成了丰富的遗传信息。图1是DNA分子碱基互补配对结构示意图。

通过DNA的复制而保留遗传信息,重组或交叉是DNA交换遗传信息的过程。两条DNA链通过交叉互相交换键上的核苷酸。通过变异生成新的遗传信息,变异是DNA序列中的变化,最常见的变异是基因中代码序列变化。如:转换变异:嘌呤代替嘌呤,嘧啶代替嘧啶。颠换变异:嘌呤被嘌呤或嘧啶替代。框架转移变异:一个或多个碱基对丢失或插入,再重新组合。

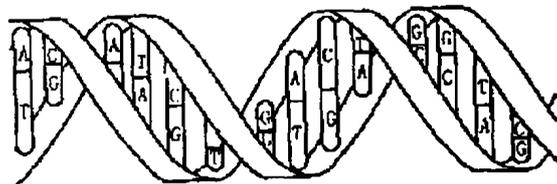


图1 DNA分子结构图

2.2 DNA 计算的原理

DNA计算的基本思想:

①利用DNA特性的双螺旋结构和碱基互补配对原则对问题进行编码,把要运算的对象映射成DNA分子链。

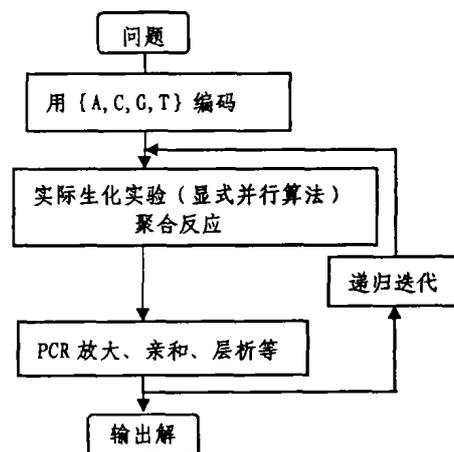


图2 DNA计算框图

^{*} 基金项目:国家自然科学基金(60174021);天津自然科学基金重点项目(013800711);河南科技攻关项目(0124140141)。丁建立 博士,副教授,从事智能仿生算法、智能网络优化等研究。陈增强 教授,博士生导师,从事智能预测控制、智能网络管理等研究。袁著祉 教授,博士生导师,从事自适应控制、智能控制等研究。

②在 DNA 溶液的试管里,在生物酶的作用下,不同的酶相当于作用在 DNA 串上的不同算子,如限制内核酸酶可作为分离算子,连接酶可作为连接算子,聚合酶可作为复制算子,外核酸酶可作为删除算子,生成各种数据池(Data Pool)。然后按照一定的规则将原始问题的数据运算高度并行地影射成 DNA 分子链的可控的生化过程。

③利用分子生物技术聚合酶链反应 PCR (Polymerize Chain Reaction)、聚合重叠放大技术 POA (Parallel Overlap Assembly)、超声波降解、亲和层析、克隆、诱变、分子纯化、电泳、磁珠分离等破获运算结果。其基本运算框架如图2所示^[7]。

2.3 DNA 的数学机理

DNA 的四类核苷酸碱基天然形成数学上的一个良好的代数结构——域。四元素机 $\Sigma = \{A, G, C, T\}$,可按图3的两种代数运算^[8]加法“+”和乘法“*”构成一个四元域。

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3 DNA 的域运算

显然,加法运算表实际上是四元素 $\{A, G, C, T\}$ 构成的双螺旋链的阵列。

另外,如果给这四元素分配一个四进制编码:

$$T=00=0_4, C=01=1_4,$$

$$G=10=2_4, A=11=3_4.$$

就可以将 DNA 的单链转换成熟悉的二进制编码。

2.4 DNA 计算的主要生物操作^[9]

- (1)合成:将核苷酸按一定次序合成一条所需长度的链;
- (2)溶合:把二个试管的溶液倒入另一个试管以达到溶合;
- (3)复合:把二条单链互补的 DNA 序列通过冷却来重新结合;
- (4)溶解:双链 DNA 通过加热分解成单链互补部分;
- (5)连接或修复:在双链 DNA 中,若其中一条单链中有不连续部分,可以通过 DNA 连接酶来修复;
- (6)复制:通过聚合酶链反应(PCR)将 DNA 串复制;
- (7)分离:使用凝胶电泳使 DNA 串按长度排序的重要技术;
- (8)提取:使用亲和力和提纯提取包含特定短序列的所有单链;
- (9)聚合:DNA 聚合可实现 DNA 的修复和复制等功能;
- (10)切割:利用切割酶对 DNA 双链进行切割;
- (11)切断:利用限制酶在某一特定的位置切断 DNA 双链;
- (12)标识:设置标识链,标识两条完整序列合成的双链;
- (13)替换和删除:利用 PCR 操作来插入、替换或删除 DNA 序列;
- (14)绑结:使用 DNA 结合酶来粘贴 DNA 链;
- (15)破损及切除:使用外核酸酶、限制内核酸酶及凝胶电泳来完成;
- (16)检测及读取:若一个给定试管中至少包含一个 DNA 链,则说“是”,反之则说“不”。

3 DNA 计算的特点与芯片制作方法

3.1 DNA 计算的特点

DNA 计算的核心问题是经过编码后的 DNA 链作为输入,在试管内经过一定时间完成控制的生物化学反应,以此来完成运算,使得从反应后的产物及溶液中能够得到全部的解空间。DNA 计算具有如下几方面的显著特点:

①具有高度的并行性。Adleman 实验中,通过适当估计, DNA 串的并行操作数目可达 10^{14} ,许多学者认为,当前技术最多可达 $10^{15} \sim 10^{20}$ 个串的并行操作,而目前最快巨型机 10^{12} 个操作。

②海量存储。DNA 作为信息的载体其储存容量非常之大,1立方米的 DNA 溶液可存储1万亿亿的二进制数据,远远超过目前所有电子计算机总储存量。

③低能耗。DNA 分子生物计算所消耗的能量只有一台电子计算机完成同样所消耗的能量十亿分之一。

④资源丰富。从植物或动物中提取 DNA 的成熟技术为 DNA 计算机的开发提供了丰富的核心资源。

3.2 DNA 芯片的制作方法

DNA 芯片俗称基因芯片,目前采用的制作方法有^[10]:

(1)点接触法:是由 Shalon 和 Browse 等人建立的。它的优点是实验样机很容易做,可以在 3.6cm^2 面积上点上10000个 cDNAs。不足之处是每个样品都必须合成好的,经过钝化的、事先保存的。

(2)喷墨法:它是定滴供给的方式通过压电晶体或其他推进形式从最小的喷嘴内把生物样品喷射到玻璃载体上。它需要的样品和点接触法一样,需要合成好的纯样品,这些样品可以是 cDNAs、染色体 DNAs 等,在 1cm^2 上可以喷射10000个点。与点接触法的差别是喷嘴不与芯片接触。

(3)光刻合成法:Fodor 和 Colleagues 建立的这种方法。它把半导体工业中的光刻术和 DNA 的化学合成方法相结合,把光不稳定保护的四种 DNA 模块固定在玻片上,通过光脱保护,用少量的保护寡酸和试剂按照设计的序列进行 DNA 合成。合成的密度和精度优于前两种方法,能在 1.6cm^2 上合成400000组寡核苷酸。该方法的缺点是需要花费大量的时间去设计和制造价格很高的照相掩护网。

4 DNA 计算与 DNA 计算机研究进展

美国加利福尼亚大学的 Adleman 博士第一次利用现代分子生物技术,提出了哈密尔顿有向路径问题的 DNA 分子生物计算方法,并成功地在 DNA 溶液试管中进行实验。Adleman 的 DNA 计算完全是一种新的概念,它突破了传统计算机体系结构的束缚,采用迄今为止还没有作为计算机硬件的生物工业技术来实现,为未来基因芯片计算机的研制奠定了开创性的基础。

Adleman 的哈密尔顿路径问题^[1](如图4所示)计算步骤如下:

(1)问题的编码及输入,每个节点的编码 $O_i (i=0, 1, \dots, 6)$ (或 A, B, C, D, E, F, G)长度为20bp (base pair),保证编码是可识别和唯一的;

(2)生成一个通过图的随机路径集,通过温度控制, DNA 链接酶,溶液来实现;

(3)搜索出以 V_0 开始 V_6 结束的路径集,通过以 O_0 和 O_6 作引物的聚合酶链反应来实现;

(4) 搜索出具有6个边的路径集,层析分离,3~5%琼脂糖凝胶;

(5) 搜索出不重复边的路径集,生物素亲和层析磁珠分离;

(6) 选择出最短路径,电泳分离,取分子量最小者。

Adleman 用实际的 DNA 序列做了以上的实验大约用了一个星期,而在一张纸上只需一个小时就可以解决同样的问题,使用数字计算机甚至数秒就可以完成,但是当城市的数目增长到70时,使用一台速度到达1000Mips 的超级计算机也难以解决。但利用 DNA 计算,这个问题却变得非常容易。这种良好的可扩展性吸引人们去研究 DNA 计算机。

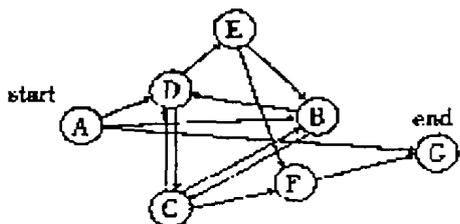


图4 七个节点的有向图

受 Adleman 思路的启发,很多人投入了研究 DNA 算法之中^[1,2,11~19],表1总结了近年来各国学者 DNA 计算研究问题及研究成果。

2001年11月,第一台全自动运行的 DNA 计算机(每秒钟能够进行10亿次运算)被以色列 Weizmann 科学研究所研制成功,这种新型计算机 DNA 和酶的混合物无需人的干涉即可运行。虽然它仍然比最好的硅计算机慢得多,但对于由 DNA 和酶组成的“汤”来说已经不错了,它极大地提高了今后几年 DNA 的筛选速度。

表1 近年来 DNA 计算问题及研究成果

时间	研究者	求解问题及研究成果
1994	L. Adleman	有向 Hamilton 路径问题
1995	R. Lipton	可满足性问题 连接网的可满足性问题
1995	L. Adleman	3-可着色问题
1995	Beaver	大因数分解问题
1995	Dan Boneh	电路的可满足性问题 最大电路可满足性问题 正则电路可满足性问题
1995	John Reif	进行 DNA 计算的联想记忆研究
1997	P. Nadia	提出 DNA 计算的两种图灵机模型
1999	L. Adleman	应用 DNA 链数据解密
2001	以色列 Weizmann 科学研究所	第一台全自动运行的 DNA 计算机研制成功(每秒钟能够进行10亿次运算)
2002	R S. Braich L. Adleman	20个变量的3-SAT 问题

该项研究发表在2001年11月22日出版的《自然》周刊上。作者之一、以色列 Weizmann 科学研究所的 Ehud Shapiro 指出,早先的 DNA 计算机需要研究人员对其作一点推动,如改变温度或者添加化学物。而他的研究小组所精心设计的 DNA 和酶的混合物使这种新型计算机无需人的干涉即可运行。这部机器的“硬件”由限制酶组成,这类酶可在特定点切断特定的 DNA 序列。8种 DNA 片段作为“软件”轮流连接到作为输

入数据而添加的 DNA 上,以确保这些酶正确地进行切割。为了组装这台机器,研究人员将数以万亿计的酶和 DNA 分子与水混合起来。一旦这台计算机完成了对输入 DNA 的读取和切割,溶液中作为特定探测器的 DNA 分子就锁上去,研究人员可以通过电泳测出计算结果。迄今为止,研究小组仅能对作为输入数据添加到溶液中人造 DNA 的24个碱基同时进行处理。但是 Shapiro 表示,将来的改进也许可使计算机能够直接分析真实 DNA 的集合(例如测定一个基因的哪个版本包括在某一特定的序列中),或者甚至游弋于人的血流中,发现疾病并释放出适当的药物。

Masami Hagiya 是东京大学的一位计算机学家,2001年曾协助制造了几乎能自动运行的 DNA 计算机。他表示:“这种想法非常聪明。它只包含了十分简单的反应,但已足以执行任何有限的自动操作。”

5 未来研究与展望

DNA 计算机的核心思想在于将经过编码后的 DNA 链作为输入,在试管内经过一定时间控制的生化反应,以此完成运算,反应的产物及溶液给出了全部的解空间。但是最优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^[20~30]。

(1)“输出技术”瓶颈。尽管现代分子生物学提供了像 PCR、高效电泳、亲和层析等选择以及放大纯化技术,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 计算机实现的主要障碍。

(2)“伪解”与“误差”。PCR 扩增是一种循环过程,耐热聚合酶有较高的碱基错配率;循环迭代次数的增加使 DNA 双链中不含任何错误碱基的拷贝比例在产物中会越来越小;另外,把大量的 DNA 链放在一起通过几百个过程步骤,偶尔会有一些非酶的非受控制的支路反应发生,甚至包括 DNA 链的动力分解。这些错误会导致一些“伪解”和“误差”的出现。为此必须增加系统结构及程式中相应的程序来纠错,从而又加大了求解的难度。

(3)人机界面。怎样使生化反应的 DNA 计算系统转化成自动控制运行的计算机系统,而且使得 DNA 计算机适应广阔的问题面,并具有实用性、自动翻译和良好的用户界面。从目前研究来看,人们正致力于 DNA 计算机与传统计算机结合构成良好的人机界面。因而有人认为,DNA 计算机至多只能起到一个运算器的作用。

展望未来,DNA 计算与 DNA 计算机具有十分光明的研究的前途。

(1)DNA 计算机体系结构。目前 DNA 计算与 DNA 计算机研究刚刚起步,也只是仅仅对某些特定问题和对照现有的计算机系统进行了初步探索,真正建立起 DNA 计算机体系结构还有很多理论与技术问题需要研究。

(2)DNA 计算机与电子计算机的融合。现有电子计算机经过几十年的发展,无论从硬件还是软件都日臻成熟,因此,DNA 计算机如能与电子计算机有机融合,将发挥出难以估量的潜力。

(3)智能 DNA 计算机。计算智能是未来计算机发展的方向。因此,DNA 计算与各种智能算法(如神经网络、遗传算法、蚂蚁算法、模糊系统、免疫系统、混沌系统、专家系统等)的集成必将成为未来 DNA 计算机研究的方向。

(4)开发 DNA 海量存储器。DNA 具有天然的海量存储功能,因此,如果据此开发海量存储器,将在数据加密、解密以及 NP 等复杂性问题取得突破性进展。

(5)开发各种用途的 DNA 芯片。针对计算机、半导体、光学合成等微加工技术以及复杂系统等开发特殊用途的 DNA 芯片必将引起研究者与应用者的极大兴趣。

(6)开发分子操作的高级语言。DNA 计算是一种新的计算方法,它以切割、粘贴、插入和删除等操作为基本运算,因此,开发以此基本运算为基础的分子操作的高级语言必将推动 DNA 计算机的形成和推广使用起巨大作用。

参考文献

- 1 Alderman L. M. Molecular Computation of Solution to Combinatorial Problems [J]. Science, 1994, 266(11): 1021~1024
- 2 Richard L. J. DNA Solution of Hard Computation Problems [J]. Science, 1995, 268(28): 542~545
- 3 Deaton R, et al. A DNA Based Implementation of an Evolutionary Search for Good Encoding for DNA Computation [A]. In: Proc. of 1997 IEEE Intl. Conf. on Evolutionary Computation, 1997, 13~16: 267~271
- 4 Manganaro G, et al. DNA Computing Based on Chaos [A]. In: Proc. of 1997 IEEE Intl. Conf. on Evolutionary Computation, 1997, 13~16: 255~260
- 5 Russo M F, et al. An Improved DNA Encoding Scheme for Neural Network Modeling [A]. 1994 Intl. Neural networks Society Annual meeting, 1994, 5~9: 354~359
- 6 Yoshikawa T, et al. Emergence of effective fuzzy rules for controlling mobile robots using DNA coding method [J]. In: Proc. 1996 IEEE Intl. Conf. on Evolutionary Computation, 1996: 581~586
- 7 Paun G, Rozenberg G, Saloman A. DNA Computing [M]. Springer, 1998. 10~41
- 8 邓少平,等. DNA 计算的一些基本问题. 科学(中文版), 1996, (5): 51~54
- 9 Kari L. DNA Computing: Arrival of Biological Mathematics [J]. The Mathematical Intelligencer, 1997, 19(2): 9~22
- 10 马立人,等. 生物芯片 [M]. 北京: 化学工业出版社, 2000
- 11 Adleman L. M. On Constructing a Molecular Computing [R]: [Technical Report TR 79-387]. 1995
- 12 Boneh D, et al. On the Computational Power of DNA [R]: [Technical Report]. 1995
- 13 Quang Q, et al. DNA Solution of the Maximal Clique Problem [J]. Science, 1997, 278(17): 446~449
- 14 Adleman L. M. Computing with DNA. Scientific American, 1998, 279(2): 54
- 15 Benenson Y, et al. Programmable and autonomous computing machine made of biomolecules [J]. Nature [J], 2001, 414(11): 430~434
- 16 Ravinderjit S, et al. Solution of a 20-Variable 3-SAT Problem on a DNA Computer [J]. Science, 2002, 296(4): 499~502
- 17 Liman W, et al. A DNA Computing readout operation based on structure-specific cleavage [J]. Nature Biotechnology, 2001, 19(11): 1053~1059
- 18 Dennis N. DNA-Based Computer Takes Aim at Genes [J]. Science, 2002, 295(2): 951
- 19 Adrian C. Hairpins Trigger an Automatic Solution [J]. Science, 2000, 288(5): 1152~1153
- 20 Leier A, Richter C, Banzhaf W, Rauhe H. Cryptography with DNA binary strands [J]. Biosystems, 2000, 57(1): 13~22
- 21 Bach E, et al. DNA models and algorithms for NP-compute problems [J]. Journal of Computer and System Sciences, 1998, 57(2): 172~186
- 22 Sakakibara Y. DNA computers: A new computing paradigm. Journal of Photopolymer Science and Technology [J], 1998, 11(4): 681~686
- 23 Yoshikawa T, et al. Effects of new mechanism of development from artificial DNA and discovery of fuzzy control rules [J]. In: Proc. 4th Intl. Conf. Soft Computing, 1996, 2: 498~501
- 24 Ogihara M, et al. Simulating Boolean circuits on a DNA computer. Algorithmica [J], 1999, 25(2-3): 239~250
- 25 Faulhammer D, et al. Counting DNA: Estimating the complexity of a test tube of DNA [J]. Biosystems, 1999, 52(1-3): 193~196
- 26 Garzon M H, et al. The bounded complexity of DNA computing [J]. Biosystems, 1999, 52(1-3): 63~72
- 27 Kari L. DNA computing in vitro and vivo [J]. Future Generation Computer Systems, 2001, 17(7): 823~834
- 28 Wasiewicz P. DNA computing: Implementation of data flow logical operations [J]. Future Generation Computer System, 2001, 17(4): 361~378
- 29 Deaton R. DNA computing: A review [J]. Fundamenta Informaticae, 1998, 35(1~4): 301~307
- 30 Garzon M H. Bimolecular computing and programming [J]. IEEE Trans. On Evolutionary Computation, 1999, 3(3): 236~250

(上接第18页)

的算法,并且探索开发异构性即发挥异构性,以及屏蔽异构性即抑制异构性的场合和方法。

参考文献

- 1 Ekmecic I, Tartalja I. A Survey of Heterogeneous Computing: Concepts and Systems. Proceeding of the IEEE, 1996, 84(8)
- 2 Freund R F, Siegel H J. Heterogeneous processing. Computer, 1993, 26(6): 18~27
- 3 曾国荪,陆鑫达. 异构计算中的负载共享. 软件学报, 2000, 11(4): 551~556
- 4 黄伟民,陆鑫达,曾国荪. 异构 BSP 模型及其通信协议. 电子学报, 2000, 28(8): 72~75
- 5 Zeng Guosun. Exploiting Maximum Parallelism in Loop Using Heterogeneous Computing. Chinese Journal of Electronics, 2001, 10(3): 340~344
- 6 Merrick I, Wood A M. Coordination with Scopes. In: Proc. ACM Symposium of Applied Computing, May 2000. 210~217
- 7 Zhang X, Yan Y. A Framework of Performance Prediction of Parallel Computing on Nondedicated Heterogeneous Networks of Workstations. In: Proc. of the 1995 Intl. Conf. of Parallel Processing, Aug. 1995. 334~338
- 8 曾国荪. 自动提取程序的异构特征实现异构计算: [博士学位论文]. 上海交通大学计算机科学与工程系, 2000
- 9 Zeng Guosun, Lu Xinda. Structured-Based Automatic Extraction of the Program Heterogeneity. In: Proc. of Conf. On Software: Theory and Practice, 16th World Computer Congress 2000, Aug. 2000. 985~986