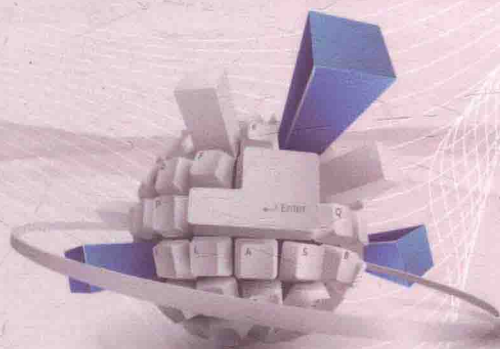


计算机算法的 实际运用与实践研究

李凤麟 著



WUHAN UNIVERSITY PRESS

武汉大学出版社

计算机算法的 实际运用与实践研究

李凤麟 著



WUHAN UNIVERSITY PRESS

武汉大学出版社

图书在版编目(CIP)数据

计算机算法的实际运用与实践研究/李凤麟著. —武汉: 武汉大学出版社, 2018. 4

ISBN 978-7-307-20081-4

I. 计… II. 李… III. 计算机算法—研究 IV. TP301.6

中国版本图书馆 CIP 数据核字(2018)第 055082 号

责任编辑: 林 莉 辛 凯

责任校对: 李孟潇

版式设计: 马 佳

出版发行: 武汉大学出版社 (430072 武昌 珞珈山)

(电子邮件: cbs22@whu.edu.cn 网址: www.wdp.com.cn)

印刷: 虎彩印艺股份有限公司

开本: 720 × 1000 1/16 印张: 19 字数: 271 千字 插页: 1

版次: 2018 年 4 月第 1 版 2018 年 4 月第 1 次印刷

ISBN 978-7-307-20081-4 定价: 58.00 元

版权所有, 不得翻印; 凡购我社的图书, 如有质量问题, 请与当地图书销售部门联系调换。

前 言

计算机强大而准确的计算能力是建立在其算法基础上的。在计算机科学中，计算机工作的原理就是执行计算机算法，计算机算法指的是为了解决一个具体问题而精心组织并且能够准确进行定义的一系列规则或者指令。计算机算法种类繁多，因此算法分析设计成为用户利用计算机程序设计实现解答问题的重要前提，面对可能存在的多个计算机算法，如何对计算机算法进行分析设计，选择最优算法，是用户寻求问题解答最佳方案的关键。计算机算法是计算机程序的基础和先导，计算机算法和数据结构构成了计算机程序。计算机算法是解决某一实际问题的具体运算序列，是问题解答过程的具体运算描述。计算机算法可以分为数值运算算法和非数值运算算法这两大种类。在进行实际问题的解答时，需要根据具体情况设计准确高效的计算机算法或者对已经存在的计算机算法进行分析判断，选择最优计算机算法，从而快速准确地解决实际问题。

在进行实际问题的解答时，需要根据具体情况设计准确高效的计算机算法或者对已经存在的计算机算法进行分析判断，选择最优计算机算法，从而快速准确地解答问题。在对计算机算法设计与分析时，必须要对计算机算法多种要素全面地进行分析考虑，从而实现科学合理的判断。计算机算法复杂性问题有算法空间复杂性和时间复杂性两项。算法空间复杂性指的是执行计算机算法需要占用的存储空间大小。算法时间复杂性指的是执行计算机算法需要花费的时间代价，即执行该计算机算法需要的工作量。一个计算机算法的工作效率主要受执行该算法占用空间和工作量这两方面资源的约束限制，我们要求计算机算法的空间复杂性与时间复杂性尽可能的低。在计算机的实际计算中，进行处理的数值经常是近似的，可能

是因为观测误差或估算问题比较困难，起始值不一定是最精确的；而且在计算机的计算过程中，还会受到数字有效数位限制。所以在确定某一计算机算法时要考虑该计算机算法在每一个计算过程中的具体行为，对在计算机运算中每一步骤产生的误差进行有效控制，使计算机计算结果具有实际意义。通过对计算机算法的有效分析，可能辨别判断出该问题存不存在最优的计算机算法。辨别判断一个计算机算法是否是最优的计算机算法，其依据一般情况下采用这个计算机算法平均性能的分析或者改计算机最坏情况的分析，如在某一问题的计算机算法类中不存在一个计算机算法比现存的计算机算法限于比较简单的场合，对于大中型的问题要真正准确寻找一个最优计算机算法是非常困难的。本书通过 DNA 计算机算法、网络图计算机算法、基于 RIFD 技术的室内定位系统研究、基于 Alpha-Beta 搜索算法研究计算机博弈以及计算机视觉关键算法对计算机算法进行深入分析，揭示了计算机算法的实践运用的神奇世界。

本书为武汉工商学院李凤麟老师的辛勤劳动成果。本书在编写的过程中参考和引用了许多学者的研究和实践成果，在此由衷地表达本人的谢意。

本书疏漏在所难免，希望得到同行和读者的批评指正。

作 者

2017 年 9 月

目 录

第一章 DNA 计算机算法研究	1
第一节 图论问题的 DNA 计算机算法研究概述	3
第二节 预备知识	20
第三节 求解 Ramsey 数的 DNA 计算机算法	31
第四节 基于分治的 Ramsey 数的 DNA 计算机算法	47
第五节 图同构问题的 DNA 计算机算法	64
第六节 最小生成树问题的 DNA 计算机算法	77
第二章 网络图的计算机算法和显示方法研究	86
第一节 背景及意义	86
第二节 基础知识部分	89
第三节 网络图的显示	97
第四节 图的连通性	110
第五节 最短路径问题	120
第三章 基于 RIFD 技术的室内定位系统研究	138
第一节 基于 RIFD 技术概述	139
第二节 室内定位系统概述	144
第三节 经典 RFID 室内定位系统	154
第四节 VIRE 算法的优化和实现	168
第四章 基于 Alpha-Beta 搜索算法的计算机博弈研究	178
第一节 人工智能	179
第二节 六子棋的状态表示和走法生成	191

第三节	基于 Alpha-Beta 的搜索算法.....	198
第四节	评价函数的详细设计.....	215
第五节	开局库和残局库.....	228
第六节	六子棋的制作与实现.....	233
第五章	计算机视觉关键算法的并行化研究.....	242
第一节	计算机视觉算法概述.....	243
第二节	同时多线程阵列机 SMT-PAAG	247
第三节	计算机视觉关键算法.....	253
第四节	关键算法的并行化实现.....	263
第六章	结语.....	293
参考文献		295

第一章 DNA 计算机算法研究

Ramsey 数问题、图同构问题、最小生成树问题等图论问题在当今科学研究的多个领域都有着广泛的应用，随着其应用范围的扩大，这些问题的求解规模日益庞大，但由于求解这些问题算法的复杂度太高，极大地影响了此类问题的深层次应用研究。

20 世纪 90 年代，随着 Adelman 在 DNA 计算领域的开拓性工作的展开，DNA 计算凭借着其海量的存储空间与可高度并行运算的能力，在理论上克服了传统电子计算机存储与运算速度方面的不足，已成为求解图论中 NP 完全问题及其他难解问题的潜在解决方案之一。随着生化技术的不断成熟，理论和实验上能够求解的 NP 完全问题的规模也越来越大。

由于目前的 DNA 计算机尚不似传统计算机般通用，求解一个问题的 DNA 计算机算法很难不作修改的应用于其他类似问题，相应的几乎所有基于 DNA 超级计算的算法均采用完全穷举方式。这种方式的直接后果造成了目前 DNA 计算中的“指数爆炸”问题，即基于穷举方法的 DNA 计算算法中所需的 DNA 链的数目随问题规模的增大而呈指数数量级增长，该问题已成为限制 DNA 超级计算机算法应用和发展的瓶颈。因此，既具有多项式求解时间又可克服 DNA 链数呈指数爆炸的 DNA 计算机新算法和模型的研究日显重要，已成为理论计算机科学的重要研究内容之一，具有相当的理论和实践意义。

本章对 DNA 计算的“指数爆炸”问题展开了一定的探索，通过将传统电子计算机并行算法设计的策略和方法引入 DNA 超级计算中，设计了求解 Ramsey 数问题、图同构问题、最小生成树问题三种图论问题的 DNA 计算机新算法，并对其 DNA 生化计算过程进行

了仿真模拟验证。

Ramsey 理论是图论中一个庞大而又丰富的领域,在集合论、逻辑学、分析,以及代数学上具有极重要的应用。Ramsey 数的求解是当前科学极难解决的问题之一。将 Adleman-Lipton 模型生物操作与粘贴模型解空间相结合的 DNA 计算模型进行扩展,在许进等提出来的位序列编码方法的基础上,提出了一种用于求解 Ramsey 数的 DNA 计算模型与算法。算法从下界开始,直到上界,每次产生问题的解空间,然后根据 Ramsey 数的定义,删除满足特定条件的解,最后检测最终的试管,以确定当前值是否为所要求的 Ramsey 数,从而得到具体的 Ramsey 数值,算法性能的理论分析和模拟实验结果表明了本算法在求解 Ramsey 数上的理论可能性,同时,由于使用了错误率更低的 DNA 计算模型,和同类算法相比,新算法具有更低的误解率,生物操作也更为简单。

在上述算法的基础上,利用分治法这一算法设计技术,设计了一种基于分治的求解 Ramsey 数 DNA 计算机算法,和前述算法相比,新算法的操作时间基本维持不变,

但显著地减少了算法所需的 DNA 链数,从而扩大了 DNA 计算理论上所能求解 Ramsey 数的问题的规模。

图的同构问题属于经典的 NP 完全问题之一,在 Sun 基于粘贴模型提出的 DNA 分子计算算法的基础上,对图同构问题的 DNA 计算机算法进行进一步研究,提出了一种基于粘贴模型和 Adleman-Lipton 解空间的图同构问题 DNA 计算机算法,算法中利用了节点的度序列概念,算法操作简单,在最坏情况下仅需 $O(2^n)$ DNA 链数,其中, n 是图的顶点数,而且保持了算法的生化操作次数仍为多项式量级。

最小生成树是图论中被广为研究的问题之一,具有重要的应用背景。本章基于 Adleman 模型生物操作与粘贴模型的解空间,提出的一种求解最小生成树问题的 DNA 计算机新算法。新算法由解空间生成器、边导出子图搜索器、生成树搜索器及最小生成树搜索四部分组成,算法求解具有 m 条边, n 个顶点的最小生成树问题所用到的生物操作数为 $O(n^2)$,测试试管数为 $O(n)$,最大链长度为

$O(m+n)$, DNA 链数为 $O(2^m)$ 。由于使用了具有更低杂交错误率的 DNA 计算机模型, 该算法提高了基于 DNA 计算解决生成树问题算法的容错性与精确性。

第一节 图论问题的 DNA 计算机算法研究概述

一、背景和目的

计算机科学与数学的迅速发展, 使图论研究应用到了当今社会的各个领域, 包括物理学、化学、通信科学、计算机技术、土木工程、建筑学、运筹学、生物遗传学、心理学、社会学、经济学、人类学和语言学, 等等。图论可以为任何一个包含二元关系的系统提供一个数学模型, 利用图本身所具有的直观、漂亮的表现特性, 使人们对现实的系统有更清晰的了解。在现实世界中许多问题的数学抽象形式也可以用图来描述, 如互联网、交通网、通信网、集成电路、分子结构等。图论已经成为了人们研究自然科学以及社会科学的一个重要工具。

在图论研究中, 有很大一部分的应用研究属于 NP 问题, 而随着图论应用领域的不断扩展和电子计算机在人类社会中所占据地位的不断增强, 该类问题也广泛应用于不断出现的新工程领域的复杂系统中, 在计算机理论科学的研究中占据了十分重要的地位。如 Hamilton 圈问题就是图论最古老的研究课题之一, 是至今未解决的世界难题, 在许多领域有着重要应用, 其应用如在一个城市, 特别是像美国的大城市, 有很多取款机, 那么, 如何制定出一个最优的路线, 使运钞车过每个提款机一次就能运完? 这时取款机是点, 路线是圈, 这个理论告诉你在某种条件下, 路线图就能形成一个圈。这中间就涉及进行图与组合优化方面的优化计算。但不管是一些常规的图论算法还是人工神经网络方法、遗传算法、模拟退火算法等, 在电子计算机这个平台上, 用某种方法对某个 NP 问题在一些特定条件下, 或者在规模比较小的条件下能很好地解决, 而在一般情况下大规模的 NP 问题, 用这些算法使用电子计算机便难以

解决。NP 类问题，有时需要问题的精确解，如多数公钥系统中的难解问题：背包问题、大数分解问题，椭圆曲线上的离散对数问题等，由于要对基于其上的密钥系统的安全性进行分析，就必须找出问题的精确解；或者对于个别 NP 问题，找其近似最优解也需要问题规模的指数时间，如团问题，此时电子计算机因运算速度和存储容量太小，便无法对这些 NP 类问题进行有效求解。近些年来，随着并行处理技术在超级计算机体系结构和算法研究上的日臻成熟，这类问题在电子计算机平台上的处理困境稍有缓解。而事实上，通过利用超级计算机提供的并行计算资源，采用资源或空间重叠的方式来求解这些困难问题也确实取得了重大进展，主要体现于可以将待求解的 NP 问题的规模成倍增加。但是，尽管超级计算机技术的研究仍在不断发展中，利用传统并行处理技术来求解这些 NP 问题的能力仍然不够。因此，对于这类难解问题，需要寻找新的解决方法。

现代自然科学研究已进入一个学科高度交叉的时期，随着现代分子生物学和计算机科学的飞速发展，生命科学与信息科学得以更加紧密地联系在一起。由此产生了现代分子生物学理论和技术与计算机科学相交叉的新兴领域——DNA 计算，而它的出现对诸多科学领域研究均产生了巨大的影响，也为我们求解图论中 NP 类问题提供了新的尝试方向与解决方式。DNA 即核酸，它是生命的遗传物质，从 1868 年由瑞士外科医生 Friedrich Miescher 在脓细胞核中首次分离出，到 1944 年 Oswald Avery 利用致病肺炎球菌中提取的 DNA 使另一非致病性肺炎球菌的遗传性状发生改变而成为致病菌得以证实 DNA 是遗传的物质基础，再到 1953 年美国生物学家 James Watson 和英国生物物理学家 Francis Crick 首先建立 DNA 双螺旋(double helix)结构模型，标志着现代分子生物学的确立，最后到 1958 年 Francis Crick 提出了蛋白质合成的“中心法则”(Central Dogma)，阐明了在生命活动中核酸的功能是储存和转移遗传信息，指导和控制蛋白质的合成；蛋白质的主要功能是作为生物体的结构成分和调节新陈代谢活动，使遗传信息得到表达，以 DNA 为核心研究的分子生物学历经百余年的发展后已经成为生命科学范围最为

重要的一个前沿领域。它不仅对我们的日常生活产生了巨大的影响,而且也影响了包括信息科学在内的许多其他领域学科。随着 DNA 重组、“聚合酶链反应法”(PCR)等现代生物技术的日臻完善,分子生物学与信息科学相交叉催生了一种新型的计算方法——分子计算。

我们都知道计算机技术作为 20 世纪最伟大的科学革命之一,其为社会的发展起到了巨大的促进作用,它的出现极大地改变了人类社会的发展并深刻地影响着我们的日常生活。在计算科学领域,我们将计算的问题划分为:容易、困难和不可计算三类。随着社会和技术的深入发展,许多工程领域中的复杂巨系统不断涌现,这些复杂巨系统的研究领域内,充满着各种各样的非线性问题,形形色色棘手的通常被称为 NP 类问题的困难问题处处可见,而现代电子计算机面对这一类 NP 问题,则显得无能为力,难以解决。其原因主要有两方面:一方面,电子计算机的运算速度太慢、存储容量太小;另一方面,量子物理学已经成功地预测出芯片微处理能力的增长不能长期地保持下去,计算机芯片制造的物理极限正日益临近。按照目前的发展速度,有人预言 2010 年芯片制造技术将达到极限尺寸——0.08 微米。基于上述两点原因以及制造技术的极限,一系列新的计算机模型应运而生,人工神经网络计算机模型、量子计算机模型、光学计算机以及基于分子水平计算基础的 DNA 计算机模型等,其中,在分子水平上进行计算的概念最早于 20 世纪 60 年代早期由 Richard Feynman 提出的,他提出构建一种“亚微观尺度的计算机”。1973 年, Bennett 也曾设想制造一种用酶来催化的图灵机,它以极低的能量代价在遗传物质如 DNA 上执行计算过程。但是由于当时尚缺乏合适的材料、工具与方法,这些想法只能是一种超前的、美好的愿望。直到 1994 年,美国南加州大学的 Adleman 博士在 Science 杂志上发表文章 *Molecular Computation of Solutions to Combinatorial Problems*, 利用 DNA 分子首先在基于分子生化反应的基础上成功求解了 7 个顶点的有向 Hamilton 路径问题(简称 HPP),开创性地提出了 DNA 计算模型之后, DNA 计算与 DNA 计算模型的研究形成了理论计算机科学、数学和生物学领域的一个新的研究

热点,吸引了众多计算机科学家和生物学家们的研究兴趣,而可用 DNA 计算解决的数学问题的种类也迅速增长。DNA 计算以其具有的海量存储和并行运算能力从理论上可克服电子计算机存储量小与运算速度慢的不足。而且,只要未来关于 DNA 计算的生物技术走向成熟(无错码、链长适中、操作自动化等),其超级计算的成本将远低于现有基于 VLSI 结构的超级计算机的成本:目前为止,一个测试试管已可产生 10^{18} 个 DNA 链,它可使 10^{18} 位数据以数据并行的方式并行运行。因此, DNA 计算可提供相当于 10^{18} 个处理单元的并行性和 $O(10^{18})$ 的存储空间。到 2008 年,最快的超级计算机在 1000s 内大约能并发处理 128×10^{17} 位的信息,而 DNA 计算中耗时最长的“抽取”操作在 1000s 内可在试管中同时处理 10^{18} 位的数据单元; DNA 计算的存储密度大约为磁带的 10^{12} 倍。尽管目前科学界还没有给出 DNA 计算将来的确切前景,但由于它在科学计算、制药和医疗等方面所展现的巨大潜力,其未来研究仍然相当可期。

尽管 DNA 计算的出现为困难问题的解决提供了希望,其未来发展也极富潜力,但要使该计算在实际应用中完全取代或者部分取代传统电子计算机工作,还面临许多实质性困难,主要体现在 DNA 计算模型的选取与 DNA 计算机算法扩展两个方面:

1. DNA 计算模型的选取

DNA 计算是伴随着分子生物学的兴起和发展而出现的。1994 年,美国加利福尼亚大学的 Adleman 博士在《科学》期刊上首次发表了关于 DNA 分子生物计算方法的开创性文章,他通过生化方法求解了 7 个顶点的哈密顿回路问题,显示了用 DNA 进行特定目的计算的可行性,其新颖性不仅仅在于算法,也不仅仅在于速度,而在于采用了迄今为止还没有作为计算机硬件的生物技术来实现。这篇文章引起了许多学者尤其是计算机科学家的兴趣,随后, Lipton 等学者也很快地提出了基于 DNA 模型的 DNA 算法,近年来该领域更是吸引了众多学者的目光。

目前, DNA 计算研究已涉及许多方面,如 DNA 计算的能力、模型和算法等。最近也有学者开始将 DNA 计算与遗传算法、神经网络、模糊系统和混沌系统等智能计算方法相结合。DNA 计算的

许多研究等待着各个学科的合作研究,如生物学、化学、计算机科学、数学和工程等,目前许多领域的科学家正在协调合作将这一理想变为现实。

就 DNA 计算而言,用 DNA 计算求解 NP 完全的图和组合优化问题的一个关键问题是 DNA 计算模型的选择。虽然 DNA 计算模型一直在不断地发展,从最初的剪接系统模型发展到目前常用的剪接系统模型、粘贴模型、表面计算模型、基于质粒的 DNA 计算模型和基于图的自组模型以及最近的禁止-强迫模型(forbidding-enforcing, FE)等,这些模型在生物操作中所出现的伪解等方面较最初的模型均有较大改进。但现有的这些模型均存在一定不足:基于粘贴的 DNA 计算模型能实现二进制的算术运算,理论上可解决许多 NP 完全问题,而且在运算过程中不需要 DNA 链的延伸和酶的作用,但算法必须通过杂交、退火和探针的设计来完成。而由于杂交及退火的不完整性,会导致大量的伪解出现。表面计算可解决一些如最佳匹配问题等 NP 完全问题,但基于该模型的算法中仍然存在杂交和酶切等生物操作,因而依然无法避免大量伪解的产生。基于质粒的 DNA 计算模型成功解决了最大权团等 NP 完全问题,该模型虽没有复杂的自身退火单链 DNA 或 PCR 扩增步骤造成的麻烦,可确保计算过程免受其他酶的干扰,从而保证计算的准确性和稳定性,但受其自身特性的影响,该模型实现算术减法、乘法和除法操作目前均有难度。基于图的自组模型和 fe 模型尽管可成功求解小规模 SAT 问题,但满足条件的图的建造有时却相当困难。

2. 完全穷举方法导致的 DNA 计算算法可扩展性过低

由于 DNA 计算尚不似传统计算机通用,求解一个问题的 DNA 计算或算法通常很难不作修改地用于其他类似问题的求解,因此,不管基于 DNA 计算的何种模型,目前几乎所有的基于 DNA 超级计算的算法均使用完全穷举方式。这种方法的直接后果是 DNA 计算算法中需要问题规模纯指数量级的 DNA 链数,若用 DNA 计算算法破解一个 n 个变量的背包公钥,则算法所需的初始 DNA 链数为 $O(2^n)$ 。尽管最新生物技术能够使一个试管中 DNA 链的浓度达到 10^{18} ,但即使这样,算法理论上在试管内仅可破解的背包密钥的变

量数也只能达到 60。而这已可在通用的 PC 机上利用二表法破解,若要破解 100 维的密钥,则需一个游泳池的 DNA 容量,但这样的密钥却可在一般的并行平台上成功破解。与现有 DNA 计算算法低可扩展性的另一个不利结果是求解问题的 DNA 分子链过长,目前生物技术在制备长链时尚存在实质性困难。因此在基于穷举方法的 DNA 计算算法中,算法的可扩展性已成为限制 DNA 超级计算在大型问题上应用的瓶颈因素。

基于上述原因, DNA 计算模型的选取和算法可扩展性的研究已成为 DNA 分子生物计算和理论计算机科学领域极具有重要性和迫切性的研究内容。本书研究以图论中 Ramsey 数、图同构问题和图的最小生成树三个经典问题为目标,将新的计算模型和处理技术引入 DNA 计算中,以显著降低 DNA 计算求解问题的 DNA 链数和链长为目标,在深入分析现有 DNA 主要计算模型的基础上,通过改进现有模型使其具有高可扩展性,以发挥 DNA 计算的内在并行性,提高求解这些图论问题的 DNA 计算算法的可扩展性,为拓宽 DNA 计算的应用广度奠定基础。

二、DNA 计算研究的基本思想,国内外现状与未来发展趋势

DNA 分子生物计算的研究包括: DNA 计算模型、使用具体模型求解问题时的 DNA 计算算法和包括编码、分子合成、输入输出技术等在内的 DNA 计算实现技术等三个方面。DNA 计算模型和 DNA 计算算法设计两者密切相关,从 DNA 计算首次求解 NP 问题开始,就备受人们关注,近十年来已涌现有许多重要的成果。

(一) DNA 计算的基本思想和特性

随着科学的进步,人们除利用有形资源外,开始对大自然这种无形的存在更加重视,因为任何生命或存在的东西都有其合理性,根据这种思想,人们揭开了生物遗传物质的面纱,并在应用于科技、军事方面取得了很多成果。

要了解 DNA 计算的本质,首先就要了解 DNA 分子结构。生物的各种生命活动都有其的物质基础,生物的遗传和变异也是这样。

根据现代细胞学和遗传学的研究得知,控制生物性状的主要物质是脱氧核糖核酸即 DNA。DNA 是一种高分子化合物,组成它的基本单位是脱氧核苷酸,每一分子的脱氧核苷酸是由一分子的磷酸,一分子脱氧核糖和一分子的含氮碱基组成的。含氮碱基共有四种,分别为腺嘌呤 (Adenine): A、鸟嘌呤 (Guanine): G、胞嘧啶 (Cytosine): C 和胸腺嘧啶 (Thymine): T,这是 DNA 的化学组成。而 DNA 的组成结构非常特殊,它是由两条双向平行的脱氧核苷酸长链围绕一个共同的纤维轴旋转而成的, DNA 分子中的脱氧核糖和磷酸交替连接,排列在外侧。其内部是通过碱基配对形成的有规则的双螺旋结构,腺嘌呤 A 与胸腺嘧啶 T 配对 (A-T),鸟嘌呤 G 与胞嘧啶 C 配对 (C-G)。组成 DNA 分子的碱基虽然只有四种,配对方式只有两种,但由于碱基对可重复排列的,所以 DNA 分子具有多样性。在分子生物计算中通常都采用 DNA 这种高分子化合物,这不仅因为 DNA 是生命信息的载体,而且还因具有在遗传工程实验室中的易操作性(见图 1.1)。

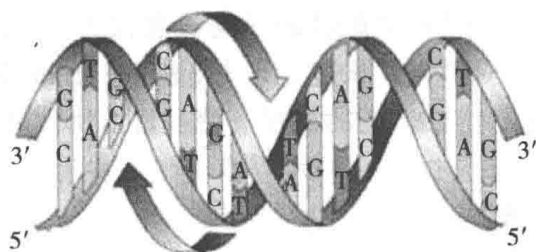


图 1.1 DNA 碱基链结构

DNA 计算就是以 DNA 与相关的某些生物酶等作为最基本材料的、基于某些生化反应原理的一种新型的分子生物计算方法。DNA 计算的最大优点是充分利用海量的 DNA 分子中的遗传密码,以及巨量的并行性。因而,以 DNA 计算模型为背景而产生的所谓的新一代计算机,称为 DNA 计算机,必有海量的存储以及极快的运行速度等特点。在 DNA 计算系统中, DNA 分子中的密码作为存储的数据,当 DNA 分子间在某种酶的作用下瞬间完成某种生物化学

反应时,可以从一种基因代码变为另一种基因代码。如果将反应前的基因代码作为输入数据,那么反应后的基因代码就可以作为运算结果。简言之,其基本原理是:首先利用 DNA 分子的双螺旋结构和碱基互补配对原则对规律进行信息编码,将所要处理的问题映射为特定的 DNA 分子片段;然后在生物酶的作用下,通过可控的生化反应生成问题的解空间;最后利用各种现代分子生物技术如聚合酶链反应 PCR、聚合重叠放大技术 POA、超声波降解、亲和层析、分子纯化、克隆、诱变、电泳、磁珠分离、序列测定等手段检测并获得运算结果。这样,通过对 DNA 双螺旋进行丰富的精确可控的化学反应,包括标记、扩增或者破坏原有链来完成各种不同的运算过程,就可能研制成一种以生物分子作为硬件的 DNA 计算。DNA 计算机(脱氧核糖核酸计算机)是近十几年来最有独创性和出乎意料的发现之一,也是生命科学与计算机科学相结合的产物。理论上 DNA 计算机不但具有现代电子计算机同样的计算能力,而且它还具有传统电子计算机不可比拟的巨大潜力和功能:

1. 海量的存储能力

组成 DNA 分子的 4 个碱基的平均长度仅为 0.35nm,按每条 DNA 链为 1000bp 计算,其长度也仅有 350nm。 1dm^3 的 DNA 溶液可以存储 1 万亿亿位二进制的数,超过目前所有计算机的存储容量。

2. 高度并行性

在十几个小时里的运算量相当于所有计算机从问世以来的运算总量。这是因为参加生化反应的每一个 DNA 分子都相当于一个纳米级处理器,在目前技术条件下,常规生化试验所能处理的 DNA 分子的数目大约为 10^{18} ,即使考虑到单个生化反应操作的时间延迟,其计算速度也将比现有的超级计算机快至少 10^5 倍。

3. 耗能低

据估计,电子计算机每焦耳耗能可执行 10^9 个操作,而相同的能量可以完成 2×10^{19} 个 DNA 连接反应,也就是说它的能量消耗只有一台普通计算机的十亿分之一。

4. 密集度高