



智能

科/学/技/术/著/作/丛/书

基于生物芯片的DNA计算 ——模型、算法及应用

丁永生 李汪根 任立红 著



科学出版社

智能科学技术著作丛书

基于生物芯片的 DNA 计算

——模型、算法及应用

丁永生 李汪根 任立红 著

科学出版社

北 京

内 容 简 介

本书面向生物计算的学科前沿,讨论了基于生物芯片的DNA计算模型、算法和应用的若干方面,包括基于生物芯片的DNA计算机的体系结构,DNA计算的运算系统、存储系统、通信机制、数据结构、一些典型的算法设计与实现,以及DNA计算在生物、医学等领域中的应用。本书取材新颖,内容深入浅出,材料丰富,理论密切结合实际,具有较高的学术水平和参考价值。

本书可作为高等院校有关专业高年级本科生或研究生的教材及参考书,也可供从事生物计算、计算机科学、自动控制、智能科学、系统科学、应用数学等领域研究的教师和科技工作者参考使用。

图书在版编目(CIP)数据

基于生物芯片的DNA计算:模型、算法及应用 / 丁永生,李汪根,任立红著.
—北京:科学出版社,2011

(智能科学技术著作丛书)

ISBN 978-7-03-032616-4

I. ①基… II. ①丁…②李…③任… III. ①脱氧核糖核酸-应用-
计算机辅助计算 IV. ①TP391.75

中国版本图书馆CIP数据核字(2011)第216300号

责任编辑:耿建业 / 责任校对:朱光兰

责任印制:赵博 / 封面设计:耕者设计工作室

科学出版社出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

新蕾印刷厂印刷

科学出版社发行 各地新华书店经销

*

2011年11月第一版 开本:B5(720×1000)

2011年11月第一次印刷 印张:14

字数:263 000

定价:70.00元

(如有印装质量问题,我社负责调换)

《智能科学技术著作丛书》编委会

名誉主编：吴文俊

主 编：涂序彦

副 主 编：钟义信 史忠植 何华灿 何新贵 李德毅 蔡自兴 孙增圻
谭 民 韩力群 黄河燕

秘 书 长：黄河燕

编 委：（按姓氏汉语拼音排序）

蔡庆生（中国科学技术大学）

蔡自兴（中南大学）

杜军平（北京邮电大学）

韩力群（北京工商大学）

何华灿（西北工业大学）

何 清（中国科学院计算技术研究所）

何新贵（北京大学）

黄河燕（北京理工大学）

黄心汉（华中科技大学）

焦李成（西安电子科技大学）

李德毅（中国人民解放军总参谋部第六十一研究所）

李祖枢（重庆大学）

刘 宏（北京大学）

刘 清（南昌大学）

秦世引（北京航空航天大学）

邱玉辉（西南师范大学）

阮秋琦（北京交通大学）

史忠植（中国科学院计算技术研究所）

孙增圻（清华大学）

谭 民（中国科学院自动化研究所）

谭铁牛（中国科学院自动化研究所）

涂序彦（北京科技大学）

王国胤（重庆邮电学院）

王家钦（清华大学）

王万森（首都师范大学）

吴文俊（中国科学院数学与系统科学研究院）

杨义先（北京邮电大学）

于洪珍（中国矿业大学）

张琴珠（华东师范大学）

赵沁平（北京航空航天大学）

钟义信（北京邮电大学）

庄越挺（浙江大学）

《智能科学技术著作丛书》序

“智能”是“信息”的精彩结晶,“智能科学技术”是“信息科学技术”的辉煌篇章,“智能化”是“信息化”发展的新动向、新阶段。

“智能科学技术”(intelligence science&technology, IST)是关于“广义智能”的理论方法和应用技术的综合性科学技术领域,其研究对象包括:

- “自然智能”(natural intelligence, NI),包括“人的智能”(human intelligence, HI)及其他“生物智能”(biological intelligence, BI)。
- “人工智能”(artificial intelligence, AI),包括“机器智能”(machine intelligence, MI)与“智能机器”(intelligent machine, IM)。
- “集成智能”(integrated intelligence, II),即“人的智能”与“机器智能”人机互补的集成智能。
- “协同智能”(cooperative intelligence, CI),指“个体智能”相互协调共生的群体协同智能。
- “分布智能”(distributed intelligence, DI),如广域信息网、分散大系统的分布式智能。

“人工智能”学科自1956年诞生的,五十余年来,在起伏、曲折的科学征途上不断前进、发展,从狭义人工智能走向广义人工智能,从个体人工智能到群体人工智能,从集中式人工智能到分布式人工智能,在理论方法研究和应用技术开发方面都取得了重大进展。如果说当年“人工智能”学科的诞生是生物科学技术与信息科学技术、系统科学技术的一次成功的结合,那么可以认为,现在“智能科学技术”领域的兴起是在信息化、网络化时代又一次新的多学科交融。

1981年,“中国人工智能学会”(Chinese Association for Artificial Intelligence, CAAI)正式成立,25年来,从艰苦创业到成长壮大,从学习跟踪到自主研发,团结我国广大学者,在“人工智能”的研究开发及应用方面取得了显著的进展,促进了“智能科学技术”的发展。在华夏文化与东方哲学影响下,我国智能科学技术的研究、开发及应用,在学术思想与科学方法上,具有综合性、整体性、协调性的特色,在理论方法研究与应用技术开发方面,取得了具有创新性、开拓性的成果。“智能化”已成为当前新技术、新产品的发展方向 and 显著标志。

为了适时总结、交流、宣传我国学者在“智能科学技术”领域的研究开发及应用成果,中国人工智能学会与科学出版社合作编辑出版《智能科学技术著作丛书》。需要强调的是,这套丛书将优先出版那些有助于将科学技术转化为生产力以及对社会和国民经济建设有重大作用和应用前景的著作。

我们相信,有广大智能科学技术工作者的积极参与和大力支持,以及编委们的共同努力,《智能科学技术著作丛书》将为繁荣我国智能科学技术事业、增强自主创

新能力、建设创新型国家做出应有的贡献。

祝《智能科学技术著作丛书》出版,特赋贺诗一首:

智能科技领域广

人机集成智能强

群体智能协同好

智能创新更辉煌

涂序彦

中国人工智能学会荣誉理事长

2005 年 12 月 18 日

前 言

DNA 计算是以 DNA 分子及生化酶为物质基础,施以适当的生化操作来解决计算问题的一种新型计算模式。DNA 计算机有望实现现有电子计算机无法真正实现的大规模并行处理和组合运算功能,是解决包括 NP 等复杂问题的突破口之一。尤为重要的是,DNA 计算的研究已经超越了纯粹计算机研究的范畴,在研究生命的本质过程、疾病诊断、新材料开发等领域也有重要应用。

目前,对各种计算问题缺乏一种直接的、统一的翻译方式来完成 DNA 计算的程式设计;在实现运算的 DNA 生化反应的途径、鉴别和输出最优解的技术路线等方面也很薄弱。换言之,DNA 计算系统的框架还未形成,DNA 计算的实用性还远远达不到要求。

生物芯片是 DNA 计算机发展的必然选择。提出基于生物芯片的 DNA 计算模型,从计算机体系结构的角度系统地讨论模型的硬件基础、指令系统、运算系统、数据结构及其应用非常有必要。基于生物芯片的 DNA 计算模型将先进的生物芯片技术和电子计算机技术结合在一起,为 DNA 计算提供了一个崭新的计算平台。该模型的提出为 DNA 计算机构建了一个相对统一的框架,提高了 DNA 计算过程的可靠性、可控性,拓展了 DNA 计算机的应用。

东华大学智能系统与网络智能研究所近年来相继得到了国家自然科学基金(60474037,60004006,69874038,61070060)、教育部高等学校博士学科点专项科研基金(20060255006,20030255009,1999025509)、教育部高等学校骨干教师资助计划、上海市科学技术委员会重点基础研究(08JC1400100)和上海市曙光计划(99SG20)等项目的资助,系统地研究了 DNA 计算的理论、技术及其应用。为了推动该领域的研究与应用,本书结合我们近年来在该领域的研究成果,对基于生物芯片的 DNA 计算模型、算法和应用的若干方面进行了较系统地阐述和讨论。全书共分 15 章:第 1 章介绍了 DNA 计算的研究现状和所面临的困难和挑战;第 2 章阐述了 DNA 计算的生物学基础;第 3 章设计了一种面向 DNA 计算的智能芯片反应装置;第 4~7 章给出了基于生物芯片的 DNA 计算模型、通信机制、存储系统和运算系统;第 8 章讨论了基于生物芯片的 DNA 计算模型的数据结构及其算法;第 9 章和第 10 章分别探讨了随机 DNA 计算模型和可信 DNA 计算模型;第 11~15 章分别给出了基于生物芯片的 DNA 计算模型在整数规划、模式识别、背包问题、医学和蛋白质组学上的应用。

本书的学术思想先进,内容新颖,材料丰富,其研究内容涉及多学科交叉领域,

且理论密切结合实际。本书结构安排合理,既照顾到面,又照顾到点,有深度和广度。读者既可以了解到这一领域的前沿研究进展,又可以深入到某一较深的研究方向。

感谢美国 Wayne State University 电气及计算机工程系的 Hao Ying 教授、东华大学信息科学与技术学院的邵世煌教授,他们多年来一直支持我们的研究工作,并提出了许多宝贵建议。感谢东华大学提供的科研和工作条件,使我们能顺利地完成任务。另外,还要感谢东华大学信息科学与技术学院的同仁对我们工作的一贯支持。我感到非常欣慰的是东华大学智能系统与网络智能研究所这个研究群体,努力工作,协同进步,取得了许多具有重要意义的研究成果。尤其要感谢朱莹等博士生和周鹏等硕士生为本书贡献了重要的研究成果。

由于本书内容涉及多个学科前沿,知识面广,而学识有限,书中难免有不妥之处,恳请广大同行、读者给予批评指正。

作 者

东华大学信息科学与技术学院

Email: ysding@dhu.edu.cn

2011 年 8 月

目 录

《智能科学技术著作丛书》序

前言

第 1 章 绪论	1
1.1 引言	1
1.2 DNA 计算概述	2
1.2.1 DNA 计算的原理	2
1.2.2 DNA 计算的特点	3
1.3 DNA 计算的实现形式	3
1.3.1 基于试管的 DNA 计算	4
1.3.2 基于表面的 DNA 计算	4
1.3.3 基于芯片的 DNA 计算	4
1.4 DNA 计算的研究现状	5
1.4.1 DNA 计算的研究成果	5
1.4.2 DNA 计算的重要性	8
1.4.3 DNA 计算所面临的困难和挑战	8
1.5 小结	10
参考文献	10
第 2 章 DNA 计算的生物学基础	13
2.1 引言	13
2.2 核酸的生物基础	14
2.3 DNA 的结构	15
2.3.1 DNA 的一级结构	15
2.3.2 DNA 的二级结构	16
2.3.3 DNA 的三级结构	17
2.4 DNA 计算的分子生物操作	17
2.4.1 计算工具——生化酶	17
2.4.2 计算方法——分子操纵手段	20
2.5 结论	24
参考文献	24

第 3 章 面向 DNA 计算的智能芯片反应装置	26
3.1 引言	26
3.2 生物芯片	27
3.2.1 生物芯片的概念	27
3.2.2 生物芯片的类型	27
3.3 芯片扫描系统	28
3.3.1 芯片扫描系统概述	29
3.3.2 自动定位芯片扫描系统	33
3.4 芯片电泳系统	36
3.4.1 芯片电泳的工作原理	36
3.4.2 芯片电泳高压装置	37
3.5 芯片 PCR 系统	38
3.5.1 芯片 PCR 概述	38
3.5.2 智能芯片 PCR 系统	39
3.6 智能芯片反应装置的设计	40
3.7 芯片微管道自动定位	42
3.7.1 图像细化算法概述	42
3.7.2 基于图像处理技术的微管道自动定位方法	46
3.7.3 自动化分析改造的实验结果	50
3.8 结论	52
参考文献	52
第 4 章 基于生物芯片的 DNA 计算模型	55
4.1 引言	55
4.2 模型的基本指令集	56
4.3 基于生物芯片的 DNA 计算模型分析	58
4.4 结论	65
参考文献	65
第 5 章 模块化 DNA 计算机的分层通信机制	68
5.1 引言	68
5.2 微流控芯片	68
5.3 基于电子计算机的 DNA 计算反应器	69
5.4 DNA 计算机与电子计算机之间的分层通信模型	71
5.5 层次通信模型中指令封装/解释层的研究	73
5.5.1 指令封装/解释的编码规则	73
5.5.2 指令封装/解释的编码实现	74

5.5.3 指令封装/解释的通用性问题	77
5.6 层次通信模型中反馈/接口层的研究	81
5.7 层次通信模型中的算法实例	84
5.7.1 选择操作	84
5.7.2 Hamilton 路径问题	87
5.8 结论	93
参考文献	93
第 6 章 基于生物芯片的 DNA 计算模型的存储系统	95
6.1 引言	95
6.2 DNA 计算模型上存储系统的设计	95
6.3 DNA 计算模型上栈式存储模块的实现	96
6.4 DNA 计算模型上链式存储器的实现	98
6.4.1 存储载体	98
6.4.2 信息编码	99
6.5 小结	100
参考文献	100
第 7 章 基于生物芯片的 DNA 计算模型的运算系统	102
7.1 引言	102
7.2 具有栈式存储结构的 DNA 自动机	102
7.3 基于 DNA 自动机的二进制串行加法	103
7.3.1 基于 DNA 自动机的一位二进制进位加法	103
7.3.2 基于 DNA 自动机的 n 位二进制串行加法	106
7.4 编码示例	106
7.5 DNA 计算机中奇偶校验的实现	107
7.5.1 奇偶校验算法	108
7.5.2 实现奇偶校验算法的有限状态自动机	109
7.5.3 有限状态自动机的核苷酸编码	109
7.5.4 算法的仿真	110
7.6 结论	112
参考文献	112
第 8 章 基于生物芯片的 DNA 计算模型的数据结构	113
8.1 引言	113
8.2 DNA 计算模型上堆栈数据结构	114
8.2.1 DNA 计算模型中堆栈的存储结构	114
8.2.2 DNA 计算模型中堆栈的生物操作	115

8.2.3 DNA 计算模型中堆栈的算法实例	117
8.3 DNA 计算模型上队列数据结构	120
8.3.1 DNA 计算模型中队列的存储结构	120
8.3.2 DNA 计算模型中队列的生物操作	120
8.3.3 DNA 计算模型中队列的算法实例	124
8.4 DNA 计算模型上广义表数据结构	126
8.4.1 广义表的存储结构	126
8.4.2 k -臂 DNA 分子	126
8.4.3 广义表链表结点的 DNA 分子表示	127
8.4.4 DNA 计算机中广义表的操作	128
8.5 结论	131
参考文献	131
第 9 章 随机 DNA 计算的研究	132
9.1 引言	132
9.2 确定 DNA 有限状态自动机	132
9.3 不确定 DNA 有限状态自动机	133
9.4 DNA 下推自动机在回文识别中的应用	135
9.4.1 接受回文语言的下推自动机	135
9.4.2 可自治 DNA 下推自动机实现	136
9.5 不确定 DNA 有限状态自动机在基因网络中的应用	139
9.5.1 基因表达的不确定有限状态自动机模型	140
9.5.2 不确定 DNA 有限状态自动机的实现	141
9.6 结论	141
参考文献	142
第 10 章 可信 DNA 计算的研究	144
10.1 引言	144
10.2 DNA 计算的自复制性	145
10.2.1 DNA 片段自组装	145
10.2.2 二维 DNA 分子元胞自动机	145
10.3 DNA 计算的可逆性	147
10.3.1 基于 DNA 计算的布尔电路	147
10.3.2 可逆容错门电路	147
10.3.3 基于 DNA 计算的可逆容错门电路	149
10.4 结论	150
参考文献	150

第 11 章 0-1 整数规划问题的 DNA 求解	152
11.1 引言	152
11.2 0-1 整数规划问题及其解法	152
11.2.1 0-1 整数规划问题的定义	152
11.2.2 0-1 整数规划问题的电子计算机算法	153
11.2.3 0-1 整数规划问题的通用 DNA 算法	153
11.3 DNA 计算模型上 0-1 整数规划问题的算法	154
11.3.1 基于芯片电泳的在线 DNA 片段分离	154
11.3.2 用于实现算法的生物芯片	155
11.3.3 算法描述	156
11.4 仿真实例	156
11.5 结论	159
参考文献	159
第 12 章 图像模式识别的 DNA 算法	161
12.1 引言	161
12.2 基于句法的图像识别及其在等腰三角形识别中应用	162
12.3 等腰三角形识别的 DNA 算法	163
12.3.1 算法设计	163
12.3.2 计算机仿真实例	165
12.3.3 生物实验	166
12.4 人脸识别的 DNA 算法	171
12.4.1 人脸识别的研究现状	171
12.4.2 一种基于奇异值分解的人脸识别方法	171
12.4.3 结合 DNA 算法和奇异值分解的人脸识别算法	171
12.5 结论	173
参考文献	174
第 13 章 基于生物芯片的背包问题 DNA 算法	175
13.1 引言	175
13.2 反应设计和编码实现	175
13.2.1 背包问题的数学模型	175
13.2.2 DNA 反应链的设计	176
13.2.3 计算和结果检测	177
13.3 材料和方法	178
13.3.1 实验材料准备	178
13.3.2 实验操作步骤	180

13.3.3 克隆测序检测	181
13.4 实验结果	183
13.4.1 反应产物 PCR 结果	183
13.4.2 克隆测序结果	184
13.5 结论	184
参考文献	185
第 14 章 DNA 计算在医学上的应用	187
14.1 引言	187
14.2 败血症基因芯片检测模型	187
14.2.1 方法	188
14.2.2 实验步骤	188
14.3 基于 DNA 计算的疾病基因诊疗模型	189
14.4 结论	192
参考文献	192
第 15 章 DNA 计算在基因注释以及蛋白质组学上的应用	194
15.1 引言	194
15.2 基于 Apollo 平台的基因注释	195
15.2.1 为牛啤的基因作注释	195
15.2.2 Apollo 软件的使用	196
15.3 用蛋白质组的方法研究牛啤感染免疫疫苗	197
15.4 基于分层通信模型的蛋白质测量	198
15.4.1 蛋白质液相分离	198
15.4.2 2D 凝胶电泳	199
15.5 牛啤胃液蛋白质组分析	205
15.6 结论	208
参考文献	209

第 1 章 绪 论

1.1 引 言

进入 21 世纪,随着社会和科学技术的发展,许多复杂系统不断出现,如 NP 完全问题、蛋白质结构预测、药物筛选等。电子计算机在解决这类复杂系统时常常显得无能为力,主要体现在两个方面:一方面,与要解决的实际问题相比,电子计算机的运算速度太慢,无法在可行时间内解决这些实际问题;另一方面,目前的电子计算机存储容量太小,在现有计算机体系结构和算法下,其内存远远不能满足解决这些复杂问题的实际需要。虽然电子计算机也正向高速度、大容量、小体积方向飞速发展,但是集成电路的复杂性、硅芯片的存储极限以及传统计算机本身计算方法的局限性,使得计算机在实现超微结构、超大存储容量、超高速运算等方面存在很大困难。电子计算机是否可能走入穷途末路? 摩尔定律是否还能继续有效? 目前的现状是,越来越高的集成度要求向传统的集成电路工艺提出了严峻的挑战,而集成电路的发展已经越来越接近技术所能容许的极限。首先,电子通道的布线因越来越密而变得越来越难,因为晶体管之间连接导线的厚度已被蚀刻到 $0.18\mu\text{m}$,电路线宽在 $0.1\mu\text{m}$ 以下将不可避免地达到仅有单个分子大小的物理学极限,现在指甲盖大小的面积上已经能安装上百万个晶体管,再增加晶体管数量是很困难的,必须考虑别的途径;其次,由于电流效应,在更小的硅芯片上布设更密的电路将导致过热的高温而使硅芯片难以承受。多数观点认为,随着半导体晶体管的尺寸接近纳米级,不仅芯片发热等副作用逐渐显现,电子的运行也难以控制,半导体晶体管将不再可靠。基于这些原因,科学家一直在寻求新的取代电子计算机的计算技术,以满足这些新的需要。

生命的发展是一个漫长的过程,生命体在遗传、变异和选择的作用下不断地向前发展和进化。生命体是一个非常复杂的系统。生命为了维持远离平衡的耗散结构,必须能够进行自组织、自适应。任何生命在其存在的每一瞬间,都在不断地调节自己内部各种机能的状况,调整自身与外界环境的关系。高等生物的自我调节是多层次的,其中包括分子的、细胞的、整体的调节。即使是原核生物也有自我调节,而且它也是通过多种途径实现的,例如,细菌有能力合成许多自身所需要的分子,可是某一分子是否合成,合成的速度如何,则随自身内部状态与环境的不同而不同。细菌内部所需要的分子,既不过多地产生,也不感到缺乏,而是靠自身的调

节机制完成的。生命的基本组成——脱氧核糖核酸 (deoxyribonucleic acid, DNA)、酶、蛋白质,能够完成这些复杂的生物任务。从这个意义上说,生命系统可看成一个高级信息处理系统,生命自开始就进行着各种复杂的计算。计算机科学中的很多技术都是受到生命信息系统的启发而发展的,如元胞自动机、神经网络、进化计算、免疫计算等。

我们知道,蝙蝠是用超声波来进行定向的,然而人类制成一台这样的超声定向仪,其体积却比蝙蝠大上许多倍。生物体的这种高效能和超小型使科学家获得启发:能否也用有机物来制造计算机呢?生物分子的大小在纳米尺度,同时生物分子具有良好的可操作性与强大的信号转导能力,这就为利用生物分子元件组装成生物计算机的研究提供了可能。在生物计算领域,DNA 计算机的诞生和发展引起了研究者的广泛关注^[1]。

1.2 DNA 计算概述

DNA 是生物遗传的物质基础。DNA 由 4 种碱基:腺嘌呤、鸟嘌呤、胞嘧啶、胸腺嘧啶(分别简称为 A,G,C 和 T)组合而成。碱基的排列组合存储着生物遗传信息。DNA 的一个重要特性是 DNA 链可以通过碱基互补配对作用形成双链,而且这种配对具有高度的特异性:A 只能与 T,G 只能与 C 配对。同一平面的碱基在两条骨架之间形成碱基对;G 和 C 之间有三个氢键,而 A 和 T 之间只有两个氢键。DNA 计算就是通过 DNA 分子的这些特点而构建的,即将运算信息编码在 DNA 链上,并通过 DNA 片段之间特定的生化操作来得出运算结果。

1.2.1 DNA 计算的原理

DNA 计算的基本思想是以 DNA 链作为信息载体,将原始问题映射成为 DNA 分子链(单链、双链或带有黏性末端的混合链);按照一定的规则将原始问题的数据运算高度并行地映射成对 DNA 分子链的分子生物操作;在分子生物实验室完成这一系列的分子生物操作;利用分子生物技术检测运算结果。因此,DNA 计算可分为四个要素:

(1) 运算载体。从目前生物计算的研究成果来看,主要以 DNA 分子作为运算载体^[2]。在 DNA 分子中,有学者应用诸如质粒 DNA 分子、单链 DNA 分子、双链 DNA 分子、发卡型 DNA 分子等各种类型的 DNA 分子为材料来进行计算,也有一些学者应用核糖核酸(ribonucleic acid, RNA)^[3]和蛋白质^[4]等为材料来进行分子计算。

(2) 运算工具。DNA 计算的操作是利用各种生化酶来完成的,这些生化酶包括连接酶、核酸内切酶、DNA 聚合酶、核酸外切酶等。

(3) 生化反应系统。DNA 计算中常用的生化反应包括退火、杂交、连接等。

(4) 检测系统。用于检测 DNA 计算的最后结果,常用的手段包括电泳技术、层析分析技术、分子纯化技术、同位素技术、荧光技术、激光技术等。

随着人们对 DNA 计算研究的不断深入,用于 DNA 计算所对应的可控生化反应以及 PCR 扩增技术,特别是关于检测技术的不断提高,生物芯片技术的不断成熟,必将改进如 Adleman 提出的试管实验方法与步骤。

1.2.2 DNA 计算的特点

DNA 计算的核心问题是将经过编码后的 DNA 链作为输入,经过一定的生物化学反应来完成运算,使得从反应后的产物及溶液中能得到全部的解空间。和传统的电子计算机相比,DNA 计算具有如下三个显著特点^[1]:

(1) 高密度的存储容量。组成 DNA 的 4 个碱基的平均长度是 0.35nm,因此利用其存储 1bit(比特)信息只占用 1nm³ 的空间,而通常的磁带需要用 10¹² 立方纳米来存储 1bit 信息。平均每个碱基的分子质量是 33Da^①,这样 1g DNA 所能存储的信息量相当于 10000 亿张普通 CD 光盘,存储密度是通常使用磁盘存储器的 1000 亿~100000 亿倍,远远超过目前所有计算机的总存储容量。

(2) 高度的并行性。参加反应的每一个 DNA 分子都可以看作一个独立的处理器。每摩尔的溶液中含有 DNA 分子的数量是 10²³,即使在当前的实验条件下生化反应的效率不是 100%,并且生化操作的时间延迟,其计算速度也将是当前超级计算机的 10 万倍以上。在 Adleman 的实验中,通过适当估计,DNA 串的并行操作数目可达 10¹⁴。许多研究者认为,用当前技术实现 10¹⁵~10²⁰ 个串的并行操作是可以达到的^[5,6],而目前最快的巨型机每秒能执行 10¹² 个操作。虽然 DNA 计算每个操作本身与电子实现相比非常缓慢,但对于当前更强的计算挑战,DNA 反应的巨大并行性足以补偿这一缺点。

(3) 极低的能耗。DNA 计算机所消耗的能量仅仅是当前电子计算机的十亿分之一^[1]。电子计算机执行 10⁹ 次操作需要 1J 能量,而用于实现 DNA 计算操作的酶是在进化中产生的,具有很高的能量效率,1J 能量足以执行 2×10^{19} 次操作。以色列科学家甚至提出了不需要外界提供能量的 DNA 计算模型^[7]。

DNA 计算的上述特性,即运算的高度并行性、大容量、低能耗是目前电子计算机所无法比拟和替代的。从这个意义上说,1994 年由 Adleman 所开创的分子生物计算技术具有划时代的意义,正因为如此,DNA 计算机成为人们所追求的目标。

1.3 DNA 计算的实现形式

根据完成 DNA 计算的生化操作的实现形式可将目前的 DNA 计算划分为试

① Da 指单位道尔顿,1Da=1.660 54×10⁻²⁷kg。