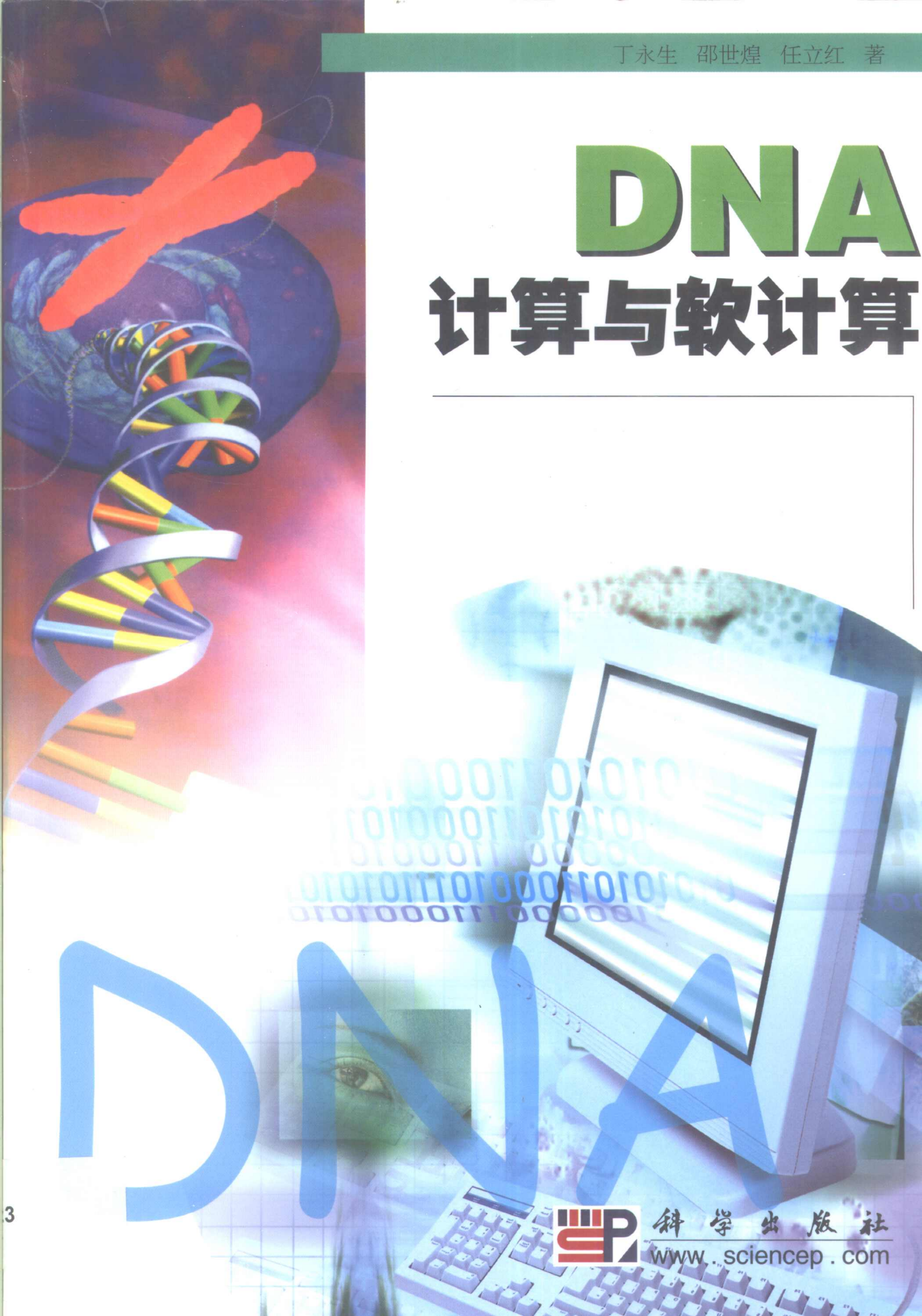


丁永生 邵世煌 任立红 著

DNA 计算与软计算



科学出版社

www.sciencep.com

内 容 简 介

DNA 计算的可能性引起了人们的广泛兴趣,其研究已涉及许多方面,如 DNA 计算的能力、模型、算法及与软计算中的智能技术相集成等。本书集中了现有绝大多数有关的主要研究成果,如从生物 DNA、计算机和数学角度研究其计算模型、生物实现及并行性功能等。另外,本书还将近年来作者在 DNA 计算与软计算集成方面做出的成果介绍给读者。全书共分 12 章,较系统地介绍了与 DNA 计算相关的生物学知识、DNA 计算的生物工具,探讨了 DNA 计算的算法实现、DNA 计算及其模型、DNA 计算机等理论研究及实验实现,给出了 DNA 计算与进化计算、模糊系统、神经网络、人工免疫系统、专家系统和混沌系统进行集成的技术、方法与应用。

本书是目前国内外首本全面而系统论述 DNA 计算与软计算工作的专著。本书可作为相关学科的研究生教材,也可供有关科研人员和工程技术人员参考。

图书在版编目(CIP)数据

DNA 计算与软计算/丁永生著. —北京:科学出版社,2002
ISBN 7-03-010898-1

I. D... II. 丁... III. 脱氧核糖核酸—计算方法
IV. Q523

中国版本图书馆 CIP 数据核字(2002)第 080477 号

责任编辑:张 臻 / 责任校对:连秉亮
责任印制:刘 学 / 封面设计:一 明

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>

苏州市吴中区教育印刷厂印刷
科学出版社发行 各地新华书店经销

*

2002 年 10 月第 一 版 开本:787×1092 1/16

2002 年 10 月第一次印刷 印张:10 1/2

印数:1—2200

字数:250 000

定价:21.00 元

前 言

在20 世纪的科学史上,有两件事也许很类似:一件是计算机芯片的出现,集成了电子器件,从而导致了计算机技术的发展及信息产业的诞生;另一件是在生物技术方面,生命信息芯片的出现,集成了生命信息,也必将导致新的产业革命。生命信息芯片主要分为4 类: DNA(基因)芯片、蛋白质芯片、免疫芯片和药物芯片。其中,DNA 芯片技术是近年来在高科技领域内出现的最具时代特征的重大进展,将为 DNA 计算和基于 DNA 计算的系统应用到许多实际问题中奠定硬件基础。

1994 年,L. M. Adleman 在《Science》杂志上首先发表开创性论文“Molecular Computation of Solutions to Combinatorial Problems”,提出了 DNA 计算的概念,并成功地解决了著名的 Hamilton 路径问题,指出了 DNA 计算潜在的巨大并行性及待研究的问题。1995 年,R. Lipton 也在《Science》杂志上发表论文,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,并掀起了 DNA 计算的研究热潮,成为一个引人注目的研究方向。自 1995 年起,以 DNA 计算为主题的国际会议每年召开一次。目前,其研究已涉及 DNA 计算的能力、模型和算法等许多方面,现有成果多集中在从生物 DNA、计算机和数学角度研究其计算模型、生物实现及并行性功能等。

DNA 生物科学技术的发展使我们能使用 DNA 作为存贮媒体,且能使用这些技术在这种内存上进行操作,以实现算法。DNA 这种媒体有强大的并行性、较高的存贮效率和极低的能耗,这就为发展 DNA 计算机开辟了前景。为了给基于 DNA 计算的系统和 DNA 智能计算机做一些基础性工作,近年来作者及其他一些学者探讨了 DNA 计算与软计算的集成及其应用,尤其是与进化算法(包括遗传算法)、模糊系统、神经网络、人工免疫系统、专家系统及混沌系统等集成。

到目前为止,国内还没有一本关于 DNA 计算的专著,而国外也没有专著全面地论述 DNA 计算的各个方面。为了推动该领域的研究与应用,尤其是让国内科学家迎头赶上这一新的学科增长点,我们特撰拙著,希望有更多的学者加入该领域,为人类社会的发展做出贡献。本书将结合我们在该领域的研究成果,从 DNA 计算的生物工具、算法设计、理论分析、DNA 计算机与软计算集成及其应用等方面对这一新的技术进行系统而全面地阐述和讨论。

全书共分 12 章。第一章介绍了 DNA 计算与软计算的背景、研究目的和意义。第二章介绍了 DNA 计算的生物基础和生物工具等。第三章给出了 DNA 计算的数学机理、算法实现、优点和缺点。第四章详细地阐述了目前 DNA 计算的主要计算模型。第五章介绍了 DNA 计算机的各个方面。第六章简述了 DNA 计算与软计算集成的概念和方法。第七章至第十二章讨论了 DNA 计算与进化算法(包括遗传算法)、模糊系统、神经网络、人工免疫系统、专家系统和混沌系统等集成及其应用。

DNA 计算的研究是一个很广的范畴,涉及分子生物信息学、遗传学、神经科学、控制论、智能科学、计算机科学、数学和非线性科学等许多领域。虽然 DNA 计算的研究还处于起步阶段,但已显示出非常大的潜力,在信息科学、生物、化学和医学中有许多潜在的应用。DNA 计算与软计算的集成是一个非常新颖的想法,其研究范围很广、理论上有很

多方面有待深入研究,在得到实际应用之前还有许多问题需要解决。

基于人工 DNA 模型的计算智能理论与技术有待于进一步的研究。本书主要讨论基于人工 DNA 模型的遗传算法、进化算法、免疫算法及其在函数寻优、模糊系统和模糊神经网络优化中的应用。为进一步与计算智能中的其他技术进行集成,需要研究基于 DNA 机理的各种智能学习方法,并将其用于智能系统的生成和优化设计。同时,在现有人工 DNA 计算模型的基础上,借助其他领域的研究成果如模糊随机搜索、模糊马尔可夫、混沌和形式语言等方法来进一步发展和优化人工 DNA 计算模型。

虽然现有成果在 DNA 计算模型的基础上引入了一些基因级操作,但研究还远远不够。DNA 蕴藏着丰富的遗传信息,其许多机理和功能有待于引入人工 DNA 模型,如模拟 DNA 修补系统功能和多层递阶结构机理。在现有工作的基础上,应深入讨论其他基因级操作和 DNA 遗传机理。进一步地模仿 DNA(基因)的调控作用,将 DNA(基因)调控机理应用于智能系统的研究中,从而使基于 DNA 计算模型的进化方法更丰富和有效,为复杂系统的控制、建模和优化设计等问题提供更丰富的功能与方法。还可从分子计算的生物智能研究中不断地得到启示,探讨更高层次计算智能的新模型,为新型智能信息处理系统逐步实现生物智能的目标作出贡献。另外,虽然已有一些关于用基因材料来体外(*in vitro*)实现 DNA 进化计算的研究,但研究的结果很简单,需要进一步地与遗传学家合作,实现更复杂和功能更强的 DNA 进化算法。

在研究 DNA 软计算的同时,加速研制 DNA(基因)芯片、DNA 计算机和生物计算机,使 DNA 计算的软件和硬件的研究同步,也应加强 DNA 智能计算机理论基础的研究。尽管 DNA 计算机的研究已取得一些进展,但还有许多实际问题和理论挑战有待解决,如物理上如何处理大规模系统和复制时的误差、逻辑上如何适应广泛种类计算问题的多用性和有效性等。目前在 DNA 计算的研究中,几乎还没有对如何获得 DNA 串之间通信的研究成果。另外 DNA 计算机与电子计算机相结合的研究还没人涉及,高级分子编程语言也有待开发。而 DNA 芯片也可能向以下几个方面发展:① 如何制造更高密度的 DNA 芯片;② 如何制造更方便、更灵活的 DNA 芯片;③ 如何减低信号之间的差异及灵敏探测;④ 如何用 DNA 芯片分析大量的信息;⑤ 如何发展 DNA 运算芯片和大规模集成芯片等。

由于时间仓促,本书中的有些观点和提法难免有不妥之处,恳请读者指正。

本书主要根据作者完成的国家自然科学基金(69874038)、国家自然科学基金青年基金(60004006)、国家教育部高校博士点专项基金(1999025509)、国家教育部骨干教师计划、上海市曙光计划(99SG20)和上海市青年科技启明星计划(00QD14038)等的资助项目的研究成果撰写的。

感谢徐元森院士在百忙之中为本书作序。感谢美国 Wayne State University 电气及计算机工程系的 Hao Ying 教授提供给我们这方面的大量信息,并对我们的研究工作提出了许多宝贵建议。感谢国外该领域的各位专家所做的研究工作,为本书提供了部分素材。另外,我们也要感谢东华大学提供的科研和工作条件,使我们能顺利地、完成其中的研究工作。最后我们还要感谢东华大学信息学院的同仁对我们工作的支持,以及我们的研究生们对部分书稿的录入和校对。

丁永生 邵世煌 任立红

2002 年 5 月于东华大学信息科学与技术学院

目 录

序

前言

第一章 绪论	(1)
1.1 生物背景	(1)
1.2 DNA 计算的产生	(2)
1.3 DNA 计算与软计算的集成	(3)
1.4 研究内容及其前景	(3)
第二章 预备知识	(5)
2.1 DNA 的结构	(5)
2.2 遗传信息流程及操作方法	(6)
2.3 DNA 计算的生物工具	(8)
2.4 结论	(9)
第三章 DNA 计算——一个新的计算领域	(10)
3.1 引言	(10)
3.2 DNA 计算的数学机理	(10)
3.3 DNA 计算的研究进展	(11)
3.3.1 DNA 通用计算机	(11)
3.3.2 DNA 计算模型及其算法	(12)
3.4 DNA 计算的算法实现实例	(12)
3.4.1 Adleman 实验:旅行商问题	(13)
3.4.2 SAT 问题	(14)
3.5 DNA 计算的优点及目前存在的问题	(17)
3.5.1 DNA 计算的优点	(17)
3.5.2 DNA 计算目前存在的问题	(17)
3.6 结论	(18)
第四章 DNA 计算的模型	(19)
4.1 引言	(19)
4.2 拼接系统	(19)
4.2.1 Chomsky 体系	(19)
4.2.2 拼接操作	(19)
4.2.3 非迭代的拼接操作	(20)
4.2.4 迭代拼接操作	(21)
4.3 图灵机的实现	(21)
4.4 DNA 计算的最小模型	(22)

4.5	DNA-EC: 等量校验模型	(23)
4.5.1	等量机器	(24)
4.5.2	操作	(24)
4.5.3	采用 DNA-EC 的计算	(24)
4.5.4	DNA-EC 的讨论	(25)
4.6	图灵机的 DNA 和限制酶的实现	(25)
4.6.1	图灵磁带	(26)
4.6.2	传输表	(26)
4.6.3	运行图灵机	(26)
4.6.4	停止	(27)
4.6.5	讨论	(27)
4.7	Sticker 模型	(27)
4.7.1	数据表示	(27)
4.7.2	操作	(28)
4.7.3	Sticker 机器	(28)
4.7.4	例子	(29)
4.7.5	讨论	(30)
4.8	Sticker 系统	(30)
4.8.1	Sticker 操作	(30)
4.8.2	Sticker 系统	(31)
4.8.3	计算能力	(32)
4.8.4	讨论	(32)
4.9	结论	(32)

第五章 DNA 计算机 (34)

5.1	引言	(34)
5.2	DNA 计算机的算术和逻辑运算	(34)
5.2.1	文本的插入、缺失和替代	(34)
5.2.2	逻辑运算	(35)
5.2.3	算术运算	(37)
5.2.4	小结	(39)
5.3	数字 DNA 分子	(39)
5.3.1	数字 DNA 分子的语法规则	(40)
5.3.2	数字 DNA 分子的应用	(43)
5.3.3	讨论	(44)
5.4	DNA 计算机上的布尔电路仿真	(44)
5.4.1	半无界扇入电路	(44)
5.4.2	半无界扇入电路的仿真	(45)
5.4.3	半无界扇入电路的实验结果	(47)
5.4.4	小结	(48)
5.5	生物分子计算与编程	(48)
5.6	DNA 计算机的应用: 运行动态规划算法	(51)

5.6.1	动态规划	(51)
5.6.2	用 DNA 解决图的可达性	(51)
5.6.3	用 DNA 解决背包问题	(53)
5.6.4	小结	(54)
5.7	结论	(55)
第六章	DNA 计算与软计算的集成	(56)
6.1	软计算	(56)
6.2	DNA 计算与软计算的集成	(56)
6.2.1	DNA 计算与进化计算的集成	(56)
6.2.2	DNA 计算与模糊系统的集成	(57)
6.2.3	DNA 计算与神经网络的集成	(58)
6.2.4	DNA 计算与人工免疫系统的集成	(58)
6.2.5	DNA 计算与专家系统的集成	(59)
6.2.6	DNA 计算与混沌系统的集成	(59)
6.3	结论	(59)
第七章	DNA 计算与进化计算	(61)
7.1	引言	(61)
7.2	进化计算的主要分支	(61)
7.2.1	遗传算法	(61)
7.2.2	进化策略	(62)
7.2.3	进化规划	(62)
7.2.4	遗传程序设计	(62)
7.3	DNA 遗传算法的理论及其应用	(63)
7.3.1	基本概念和术语	(63)
7.3.2	DNA 遗传算法的假设	(64)
7.3.3	DNA 遗传算法的结构	(64)
7.3.4	DNA 遗传算法在函数寻优中的应用	(67)
7.3.5	DNA 遗传算法与常规遗传算法的比较	(68)
7.4	DNA 进化算法	(68)
7.4.1	一类典型广义隶属度型 TS 模糊控制器的构成	(69)
7.4.2	基于基因转移操作的 DNA 进化算法	(70)
7.4.3	仿真研究	(73)
7.5	结论	(74)
第八章	DNA 计算与模糊系统	(76)
8.1	引言	(76)
8.2	模糊系统的基本原理	(77)
8.2.1	MISO 广义 Mamdani 型模糊系统的构成	(78)
8.2.2	一般化 TS 模糊控制器的构成	(79)

8.2.3	模糊系统是万能函数逼近器	(80)
8.3	基于重叠 DNA 编码方法的计算模型及其模糊系统的优化设计	(81)
8.3.1	DNA 遗传操作算子	(81)
8.3.2	模糊规则的 DNA 编码表达	(83)
8.3.3	DNA 编码方法获取机器人避碰的模糊控制规则	(86)
8.4	基于 DNA - GA 优化设计的广义模糊系统	(87)
8.4.1	Mamdani 模糊系统的设计	(87)
8.4.2	仿真结果	(89)
8.5	基于 DNA - GA 优化设计的 TS 模糊控制系统	(90)
8.5.1	TS 模糊系统的设计	(90)
8.5.2	仿真结果	(91)
8.6	结论	(93)
第九章	DNA 计算与神经网络	(94)
9.1	引言	(94)
9.2	神经网络基本理论	(94)
9.2.1	神经元模型	(94)
9.2.2	神经网络	(95)
9.2.3	基本理论	(95)
9.3	神经网络的 DNA 编码框架	(98)
9.4	采用重叠基因编码的遗传算法优化设计 TS 模糊神经网络控制器	(98)
9.4.1	广义隶属度型 TS 模糊神经网络控制器的构成	(99)
9.4.2	基于重叠基因的 DNA 遗传算法	(100)
9.5	结论	(104)
第十章	DNA 计算与人工免疫系统	(105)
10.1	引言	(105)
10.2	生物免疫机理	(105)
10.2.1	生物免疫系统的组成与特征	(105)
10.2.2	免疫系统的反馈机理	(106)
10.2.3	与免疫遗传算法相关的生物机理	(107)
10.2.4	免疫系统的自己-非己识别机理	(107)
10.3	基于免疫反馈机理的控制系统	(107)
10.3.1	免疫反馈规律	(107)
10.3.2	免疫反馈控制器的设计	(108)
10.3.3	免疫反馈规律的模糊自调整	(109)
10.3.4	热疗法中组织温度控制的应用	(110)
10.4	基于人工 DNA 计算模型的免疫遗传算法	(111)
10.4.1	免疫遗传算法的原理	(112)
10.4.2	基于 DNA 计算的免疫遗传算法及其用于 TS 模糊控制系统的优化设计	(113)

10.4.3	采用 DNA - IGA 优化设计 TS 模糊控制系统的仿真结果	(114)
10.5	基于 DNA 计算的自己-非己识别系统	(115)
10.5.1	自己-非己的异常数据检测算法	(116)
10.5.2	自己-非己判别的 DNA 实现	(117)
10.6	结论	(118)
第十一章	DNA 计算与专家系统	(119)
11.1	引言	(119)
11.2	专家系统的基本原理	(119)
11.2.1	专家系统的定义	(119)
11.2.2	专家系统的结构	(120)
11.3	基于分子计算的推理	(121)
11.3.1	DNA 链的符号描述	(121)
11.3.2	推理机制	(123)
11.3.3	基于环状 DNA 分子的典型推理系统的实现	(125)
11.3.4	实验验证	(127)
11.4	未来推理系统的发展	(129)
11.4.1	多条件规则的推理系统的 DNA 分子表达	(129)
11.4.2	DNA 分子组成的推理系统的范例	(132)
11.5	结论	(134)
第十二章	DNA 计算与混沌	(135)
12.1	引言	(135)
12.2	混沌系统	(135)
12.2.1	混沌系统的特性	(135)
12.2.2	洛伦兹(Lorenz)系统	(136)
12.2.3	Poincare 映射和符号动力	(137)
12.2.4	符号动力和外部干扰	(137)
12.3	基于混沌的 DNA 计算	(138)
12.3.1	信息的混沌编码	(138)
12.3.2	混沌与试管	(138)
12.3.3	有向的 Hamilton 路径问题	(139)
12.3.4	仿真结果	(140)
12.4	基于 DNA 编码的推理网络	(141)
12.4.1	具有类核苷酸相互作用的 Logistic 网	(141)
12.4.2	仿真结果	(143)
12.5	结论	(145)
参考文献		(146)

第一章 绪 论

1.1 生物背景

生物的发展是一个漫长的进化过程。19 世纪 50 年代,英国生物学家达尔文(C. R. Darwin)根据对世界各地生物的考察资料和人工选择的实验提出了生物进化论。1889 年达尔文出版《物种起源》巨著,提出了以自然选择为基础的生物进化论学说。达尔文进化论的中心内容是自然选择学说,生物发展进化主要有遗传、变异和选择三个原因。遗传性是一切生物所共有的特性,正是这种遗传性,使得生物能够把它的特性、性状传给后代,在后代中保持相似。遗传是生物进化的基础。变异是一切生物所具有的共同特征,是生物个体之间相互区别的基础。生物的变异性为生物的进化和发展创造了条件。选择决定了生物进化的方向。选择可分为人工选择和自然选择。人工选择是在人为环境下,把对人有利的生物特性保留下来,淘汰不利的生物特性。自然选择是指生物在自然界的生存环境中适者生存、劣者淘汰的过程。生物就是在遗传、变异与选择三者综合作用下,不断地向前发展和进化。生物进化论揭示了生物长期自然选择的进化规律,自然选择过程蕴含着一种搜索和优化的先进思想。科学家们从中得到启迪,将这种思想用于工程和科学领域,发展了多种进化学习算法,为许多传统优化方法难以解决的优化问题提供了新的途径。

随着人类对生命的进一步研究,遗传学的研究从细胞水平深入到分子水平,计算也相应地深入到细胞计算^[1, 2]和分子计算^[3~6]水平。在遗传学方面,人体信息处理系统如神经系统和免疫系统中起关键控制作用的 DNA 遗传密码也被破译。在自然界中,生物体表现出的性状多种多样,而每个物种又保持其相似性。这一切都是由生物体中的遗传物质脱氧核糖核酸(Deoxyribonucleic acid,缩写为 DNA)决定的。DNA 中有 4 种碱基:腺嘌呤(Adenine,缩写为 A)、鸟嘌呤(Guanine,缩写为 G)、胞嘧啶(Cytosine,缩写为 C)和胸腺嘧啶(Thymine,缩写为 T)。各种碱基间的不同组合就构成了异常丰富的遗传信息。DNA 含有大量的遗传密码,通过生化反应传递遗传信息。这一过程是生命现象的基本特征之一。DNA 分子的遗传信息可通过 DNA 的复制传递给后代,也可通过细胞中称为“核糖核酸”的信息分子(RNA)的信息,指导蛋白质信息大分子的合成,由此决定生物各种形态和生理特征,还可调控生物的生长、发育、生命活动和功能作用等。另外,生物的神经等组织是细胞的一种分化,其生成和功能也受 DNA 遗传信息的作用。DNA(基因)在免疫系统也起关键控制作用,基因复合体不仅对人体免疫系统能否识别外来抗原起关键作用,而且也决定着为什么不同人的组织和器官会存在差别。近 10 多年来,DNA 科学的发展非常迅速,对生命科学、医学等方面带来了巨大影响。

1.2 DNA 计算的产生

在 20 世纪的科学史上,有两件事也许很类似:一件是计算机芯片的出现,集成了电子器件,从而导致计算机技术的大发展及信息产业的诞生;另一件是生命信息芯片的出现,集成了生命信息,也必将导致新的产业革命。生命信息芯片分为 4 类:DNA (基因)芯片、蛋白质芯片、免疫芯片和药物芯片。其中,DNA 芯片技术是近年来在高科技领域内出现的具有时代特征的重大进展。目前有许多公司已制造出 DNA 芯片或研制出 DNA 芯片与系统的制作方法。DNA 芯片作为一种核心且交叉性极强的技术,不仅影响基础科学的研究,而且直接涉及信息产业、人类健康、环境保护、生态农业、制造业等受人关注的领域。可以说,DNA 芯片将为 DNA 计算和基于 DNA 计算的系统应用到许多实际问题中奠定了一定的硬件基础。

DNA 能做成芯片,那么 DNA 能“计算”吗?DNA 是否有通用计算模型呢?考虑到现有的计算机是建立在二进制编码基础之上,两者存在相似性。那么,是否可以找到一种方法,在 DNA 的 4 字母表{A, T, C, G}的基础上进行编码,对问题进行描述,并运用生物工具和技术进行处理达到计算的目的呢?1994 年,L. M. Adleman 在《Science》杂志上首先发表开创性文章“Molecular Computation of Solutions to Combinatorial Problems”^[3],提出了 DNA 计算概念,并成功地解决了著名的 Hamilton 路径问题,指出了 DNA 计算潜在的巨大并行性及待研究的问题,从而开创了 DNA 计算机的新纪元。1995 年 R. Lipton 也在《Science》杂志上发表文章^[7],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。美国很多大学如 Princeton 大学、Stanford 大学、加州理工大学等纷纷开展研究,并得到了美国国家科学基金会和五角大楼国防高级研究项目局的支持。1995 年,来自世界各国及地区的 200 多名科技界人士共同探讨并肯定了 DNA 计算机的可行性,认为一旦 DNA 计算机研制成功,其运算量是目前传统计算机所望尘莫及的。这无疑是一个极具开发价值的研究领域。目前国际学术机构每年举行一次关于 DNA 计算机的学术会议,还定期召开一些研讨会。

DNA 计算的研究目前已涉及 DNA 计算的能力、计算模型和算法等许多方面。现有成果多集中在从生物 DNA、计算机和数学角度研究其计算模型、生物实现及并行性功能等。DNA 生物科学技术的发展使我们能使用 DNA 作为存贮媒体,且能使用这些技术在这种内存上进行操作,以实现算法。DNA 这种媒体有强大的并行性、较高的存贮效率和极低的能耗,为发展 DNA 计算机开辟了前景。

DNA 计算是一种关于计算的新的思维方式,同时也是关于化学和生物的一种新的思维方式。尽管生物和数学的过程有各自的复杂性,但它们有一个重要的共性,即生物体所具有的复杂的结构实际上是编码在 DNA 序列中的原始信息经过一些简单的处理后得到的。求一个含变量的可计算函数的值也可以通过求一系列含变量的简单函数的复合来实现。这便是研究 DNA 计算机的出发点。DNA 计算的本质就是利用大量不同的核酸分子杂交,产生类似某种数学过程的一种组合的结果并对其进行筛选的。DNA 本身有两个不同的末端即 5' 端和 3' 端,因此整个分子有方向性。当大量随机的

DNA 链相互杂交后,每个 DNA 链原来携带的信息就会产生一种数学组合的结果。而对一种特定的运算,这种结果的获得是通过对 DNA 进行一系列连续操作来实现的。

DNA 计算机的原理是:DNA 分子中的密码相当于存储的数据,DNA 分子之间可以在某种酶的作用下瞬间完成生物化学反应,从一种基因代码变为另一种基因代码;反应前的基因代码可以作为输入的数据,反应后的基因代码可以作为运算结果;如果控制得当,那么就可以利用这一过程制成一种新型计算机。目前,人们已开始考虑将分子生物学技术用于计算问题的一整套方案,从破解密码到构造一台能进行任何逻辑和代数运算的“通用计算机”。

1.3 DNA 计算与软计算的集成

生命现象的奥秘和生物的智能行为一直被科学家们所关注。生物体是一个非常复杂的大规模系统。生物为维持个体的非平衡耗散结构,获得了根据自组织化构造信息处理系统的能力。生物信息系统可看成一个高级信息处理系统,能使人类获得启发而提供给工程领域各种富有成效的智能技术和方法。生物信息处理系统有功能与时空尺寸相异的脑神经系统、免疫系统和内分泌系统三个子系统,并通过能分泌荷尔蒙的内分泌系统发生相互作用。这些子系统以及遗传信息系统向生物智能方向发展可导向“计算智能”。这些计算智能技术用于计算领域可看成“软计算”。现有的计算智能主要包括神经网络、进化计算、免疫计算及模拟人类大脑思维方式的模糊系统等。这些计算智能技术一直是当前智能科学领域的研究热点,且已被广泛应用于各个领域。但到目前为止,计算智能的理论研究只是停留在对生物系统的简单模拟,对生物系统的研究结果也仅局限于理解生物过程、仿真生命、计算模型、分布计算及智能机器人等方面。如遗传算法,虽然在不可微函数、复杂及非线性系统寻优等问题上显示出突出的作用,但其实质是对生物进化的一种简单抽象,即基于“0”和“1”编码的信息模型,不能表达丰富的遗传信息,且也没有在其计算模型中反映出遗传信息对生物体生长、发育的调控作用,尤其是起关键控制作用的 DNA 编码机理和调控作用在现有的计算智能中没有体现出来。

DNA 的生物背景使我们思考,能否进一步地分析和模仿遗传信息调控系统的自生成、自组织功能,引入“基因工程”机理改造系统结构和参数,实现基于遗传机理的在线优化,从而建立分子水平上的基于 DNA 控制机理的遗传信息模型呢?作者即是基于以上思考,近年来开始研究基于 DNA(基因)在遗传处理系统中的控制作用,开发新型的基于人工 DNA 模型的计算智能来解决一些科学与实践问题。同时,生物计算机是人们多年来的梦想。DNA 这种媒体有强大的并行性、较高的存贮效率和极低的能耗等特点,可彻底实现现有计算机所无法真正实现的模糊推理功能和神经网络运算功能,是人工智能的又一个突破口。

1.4 研究内容及其前景

DNA 计算的研究属于生物、遗传、化学、数学、物理、计算机、控制论和智能科学等

学科的交叉领域,其研究内容很深,涉及范围很广。目前,DNA 计算的研究内容主要集中在以下 5 个方面。

(1) DNA 计算的生物工具和算法实现技术。

(2) DNA 计算的模型:主要考虑各种 DNA 计算的理论模型,讨论它们的计算能力和数学实现等。

(3) DNA 计算机的基本计算:主要研究 DNA 的布尔电路运算、数字 DNA、算术运算、分子乘、分子编程和应用等方面。

(4) DNA 计算与软计算的集成:基于 DNA 机理,阐述 DNA 计算与软计算集成的技术和方法,主要有面向智能系统的人工 DNA 模型、基于 DNA 机理的智能系统和基于人工 DNA 模型的计算智能理论及其应用方法,并将其应用于智能系统的在线学习、优化和控制。具体研究时,以开发新型的智能系统方法为目标,进一步分析和模仿生物遗传机理和 DNA 遗传系统的编码和控制机理,研究基于 DNA 模型的各种学习算法及其应用方法来解决智能系统的在线优化和设计问题,从而在分子水平上促进基于计算和知识相结合的智能理论、模型和方法的综合集成,为发展 DNA 计算机和生物智能计算机起一定的促进作用,为智能控制及其相应的理论与方法开辟一个更为广阔的研究领域,也为今后不断地从生物智能研究中得到启示,探讨更高层次计算智能的新模型,为新型智能信息处理系统逐步实现生物智能的目标提供一些思路和方法。

(5) DNA 智能计算机:国外许多学者认为,DNA 计算的研究为人类研制分子计算机奠定基础。基于 DNA 计算的智能系统将会架设 DNA 计算与智能系统研究的桥梁,在 DNA 智能计算机等相关研究中起一些垫脚石作用。

随着 DNA 计算研究热潮的兴起,DNA 计算机为并行计算提供了一种强大的选择。DNA 计算与软计算的集成将为 DNA 智能计算机提供一条很好的实现途径。DNA 智能计算机可彻底解决现有计算机所无法实现的智能运算功能。相信随着生物技术和计算技术的发展,DNA 计算将会在许多领域发挥重要作用。

第二章 预备知识

2.1 DNA 的结构

DNA 是在 DNA 计算中起中心作用的分子,是重要的基因物质,携带着生物的遗传信息^[8]。DNA 的基本元素是核苷酸。由于化学结构的不同,核苷酸划分为腺嘌呤(A)、鸟嘌呤(G)、胞嘧啶(C)和胸腺嘧啶(T)4 类碱基。

多核苷酸链(Polynucleotide chain)是核酸(DNA 和 RNA)结构的基础,是由核苷酸通过 3',5'-磷酸二酯键连接成的线性大分子。核苷酸是多核苷酸链的结构单位。DNA 链主要是由一个脱氧核苷酸上的 5'-磷酸基和另一个脱氧核苷酸核糖上的 3'-羟基共价连接而成。DNA 由两条极长的核苷酸链利用碱基之间的氢键结合在一起,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(图 2.1)。A 和 T 配对,C 和 G 配对。碱基的上述配对关系称为 Watson-Crick(WC)配对。

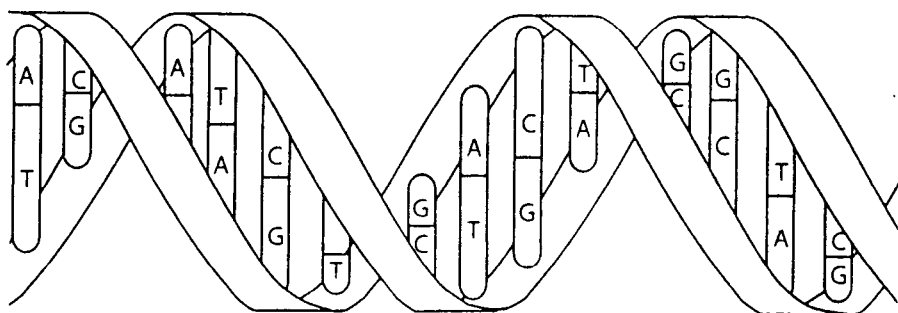


图 2.1 DNA 双螺旋结构

DNA 有两个最主要的功能:第一个功能是 DNA 携带遗传信息,能转录成 RNA, RNA 再转译成蛋白质;第二个功能自我复制, DNA 以本身作为模板,复制出另一个相同的 DNA。DNA 一般为长而无分支的双股线型分子,但有些为环形,也有少些为单股环形。每个染色体是一段双股螺旋的 DNA。遗传信息以 A、T、C 和 G 在核苷酸中的排列顺序而体现,其排列序列的多样性体现了丰富的遗传信息。

从生物 DNA 到蛋白质的形成过程如下(图 2.2):首先,通过转录作用将 DNA 中携带的遗传信息转录到信使 RNA(mRNA)中;在 mRNA 中排列着由 3 个连续的碱基组成的密码子,这些密码子是合成蛋白质的密码,64 种密码子对应 20 种氨基酸(表 2.1);通过翻译作用,将 mRNA 中携带的遗传信息转译成含特定氨基酸序列的蛋白质,蛋白质则构成了细胞。在生物 DNA 中,基因是储存遗传信息的基本单位,一个基因开始于起始密码子 ATG,终止于终止密码子 TAA、TAG 或 TGA。

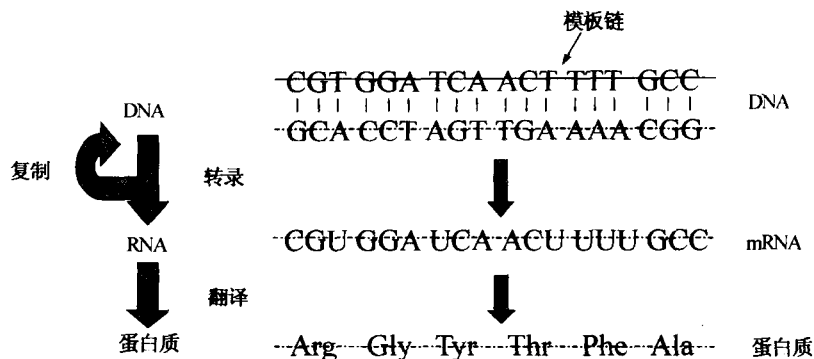


图 2.2 从 DNA 到蛋白质的形成过程

表 2.1 遗传密码表

第一个核苷酸	第二个核苷酸				第三个核苷酸
	T(U)	C	A	G	
T(U)	Phe	Ser	Tyr	Cys	T(U)
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	终止	终止	A
	Leu	Ser	终止	Trp	G
C	Leu	Pro	His	Arg	T(U)
	Leu	Pro	His	Arg	C
	Leu	Pro	Glu	Arg	A
	Leu	Pro	Glu	Arg	G
A	Ile	Thr	Asp	Ser	T(U)
	Ile	Thr	Asp	Ser	C
	Ile (Met)	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	T(U)
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

2.2 遗传信息流程及操作方法

在细胞利用遗传信息的过程中,有几种非常重要的操作^[9]。

(1) 复制: 通过 DNA 的复制而保留遗传信息。在复制过程中,以亲代 DNA 的两条链为模板,在每条链上各复制出一条互补的 DNA 链。

(2) 转录: 把遗传信息从细胞核传送到细胞质的第一步就是从 DNA 到 mRNA 的转录过程。在该过程中,DNA 双链解开,其中一条链作为转录的模板,合成一条互补的 RNA 链(mRNA)。

(3) 翻译: 用 mRNA 编码的遗传信息翻译成由一个特定序列氨基酸连成的蛋白质。在翻译过程中,mRNA 和转移 RNA(tRNA)这两种 RNA 起着非常重要的作用。mRNA

中携带着合成氨基酸的密码。tRNA 充当“接头”，起调节作用，识别密码子，将氨基酸插入到多缩氨基酸的合适位置。

(4) 重组(交叉): DNA 交换遗传信息的过程,两条 DNA 链通过交叉相互交换链上的核苷酸。根据对 DNA 序列和所需蛋白质因子的要求,可将重组分为同源重组、位点特异性重组、异常重组和转座作用等。遗传重组为产生遗传性变异的原因之一,对育种和进化具有重要的表现意义。

(5) DNA 变异: DNA 序列中的变化。最常见的变异是基因中代码序列的变化。这种改变可以分为两类。一种是碱基的取代,即 DNA 序列中某个碱基被另一个碱基所取代。碱基的取代又有两种:① 同类型碱基转换(transition)变异(图 2.3(a)),即嘌呤替代嘌呤、嘧啶替代嘧啶;② 异类型碱基颠换(transversion)变异(图 2.3(b)),即嘌呤被嘌呤或嘧啶替代。另一种是框构转移(frameshift)变异,包括碱基的丢失(deletion)(图 2.4(a))和碱基的嵌入(insertion)(图 2.4(b)),即一个或多个碱基对丢失或嵌入,再重新组合。例如,由于酶操作而导致的丢失操作,在酶的起始密码子和终止密码子之间的碱基被丢失。又如由于病毒操作而导致的嵌入操作,一段碱基序列被嵌入到染色体中。

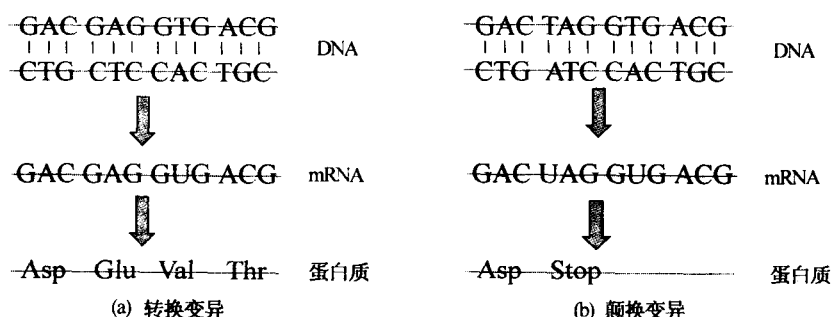


图 2.3 碱基的取代

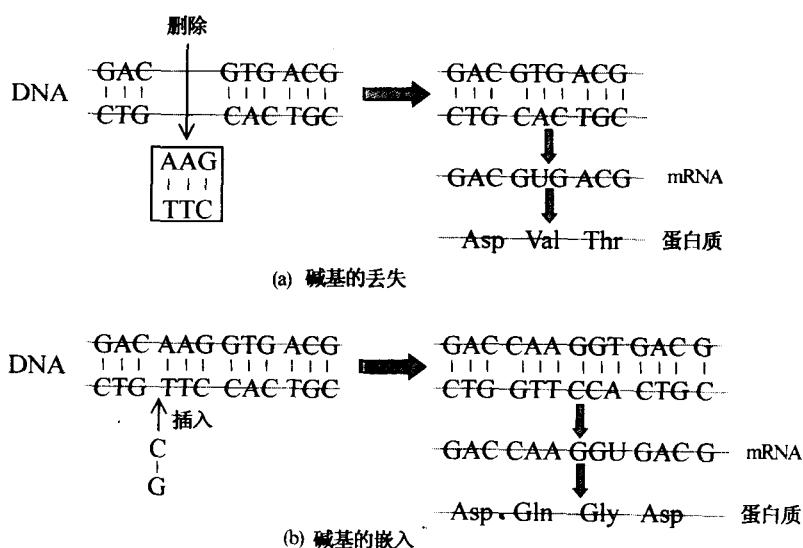


图 2.4 框构转移变异

2.3 DNA 计算的生物工具

DNA 计算是通过结合各种生物操作来实现的。目前, DNA 计算的生物工具主要有^[10]以下几种。

(1) 合成: 可以用一台机器将寡核苷酸按一定次序合成, 得到某一期望长度的链。合成器提供 4 种碱基, 它们按用户输入的序列结合在一起。在该机器中生成上万个所需寡核苷酸的复制, 然后将它们放入到一小瓶溶液中。例如, 核苷酸 A 装在一只瓶子中, 将含有核苷酸 C 的溶液倒进去, 于是 A 和 C 发生反应后形成 2 个核苷酸链 AC, 将剩余的 C 从溶液中清洗后, 接着, 可以将链 AC 与 T 合成为 3 个核苷酸链, 以此类推。

(2) 混合: 把 2 支试管中含不同核苷酸的溶液倒入另一支试管, 获得混合体。

(3) 熔解: 通过加热溶液到某一特定温度, 将双链 DNA 分解成两条互补的单链 DNA。加热破坏了互补链之间的氢键。

(4) 复合(退火): 是熔解的逆过程。把含有单链的溶液冷却, 将两条完整的单链 DNA 序列重新连接起来, 形成双链 DNA。

(5) 连接: 用 DNA 连接的方式将 DNA 互补链首尾相连。

(6) 修复: 在双链 DNA 中, 如果其中一条单链中有不连续部分, 可以通过 DNA 连接酶来修复。

(7) 杂交分离: 是早期 DNA 计算模型的极为重要的操作, 可以提取包含特定短序列的所有单链。

(8) 复制(放大): 通过聚合酶链反应(PCR)将 DNA 串复制。PCR 是基因工程中的一种 *in vitro* 手段, 利用 DNA 聚合酶, 将给定溶液中的特定 DNA 序列快速地复制或放大。PCR 包含了一系列温度循环的重复, 每一循环由三个阶段组成: 模板 DNA 的变性分解成链; 然后冷却使引物与寡聚核苷酸的模板进行复合, 以此来设计与感兴趣的 DNA 区域相接; 最后由 DNA 聚合酶使引物延伸。每一个反应循环都使目标 DNA 分子的数量加倍, 因此该反应呈现指数式增长。

(9) 分离(凝胶电泳): 是一种使 DNA 按大小排序的重要技术。将 DNA 链放在湿的凝胶顶部并对凝胶施加一电场, 由于 DNA 带负电, 在电场的牵引下, 使 DNA 链到达凝胶底部。由于短 DNA 链的压力较长 DNA 链的要小, 其移动速度快。经过一段时间以后, DNA 链就按各自的大小排列在不同的区域。

(10) 聚合: 给出一段短的寡核苷酸引物 *P*, 引物 *P* 可以是 DNA, 也可以是 RNA, 但是引物的 3' 末端核苷酸或脱氧核糖核苷酸的核糖的第三碳上必须是羟基(3'-OH 末端)。当 *P* 和一个长寡核苷酸模板结合后, 在 DNA 聚合酶的作用下, 将脱氧核苷酸一个一个地往引物上添加, 直到形成一个完整的互补双链结构为止。DNA 聚合可实现 DNA 的修复和复制等功能。

(11) 切割: 利用限制酶在某一特定的位置处切断 DNA 双链。

(12) 提纯: 使用亲和力, 通过吸引将那些包含给定模式的 DNA 链作为子链(串)提取出来。

(13) 标识: 设置标识链, 标识两条完整序列合成的双链。其逆操作是将双链解除标