

基因工程

——从基础研究到技术原理



Genetic Engineering

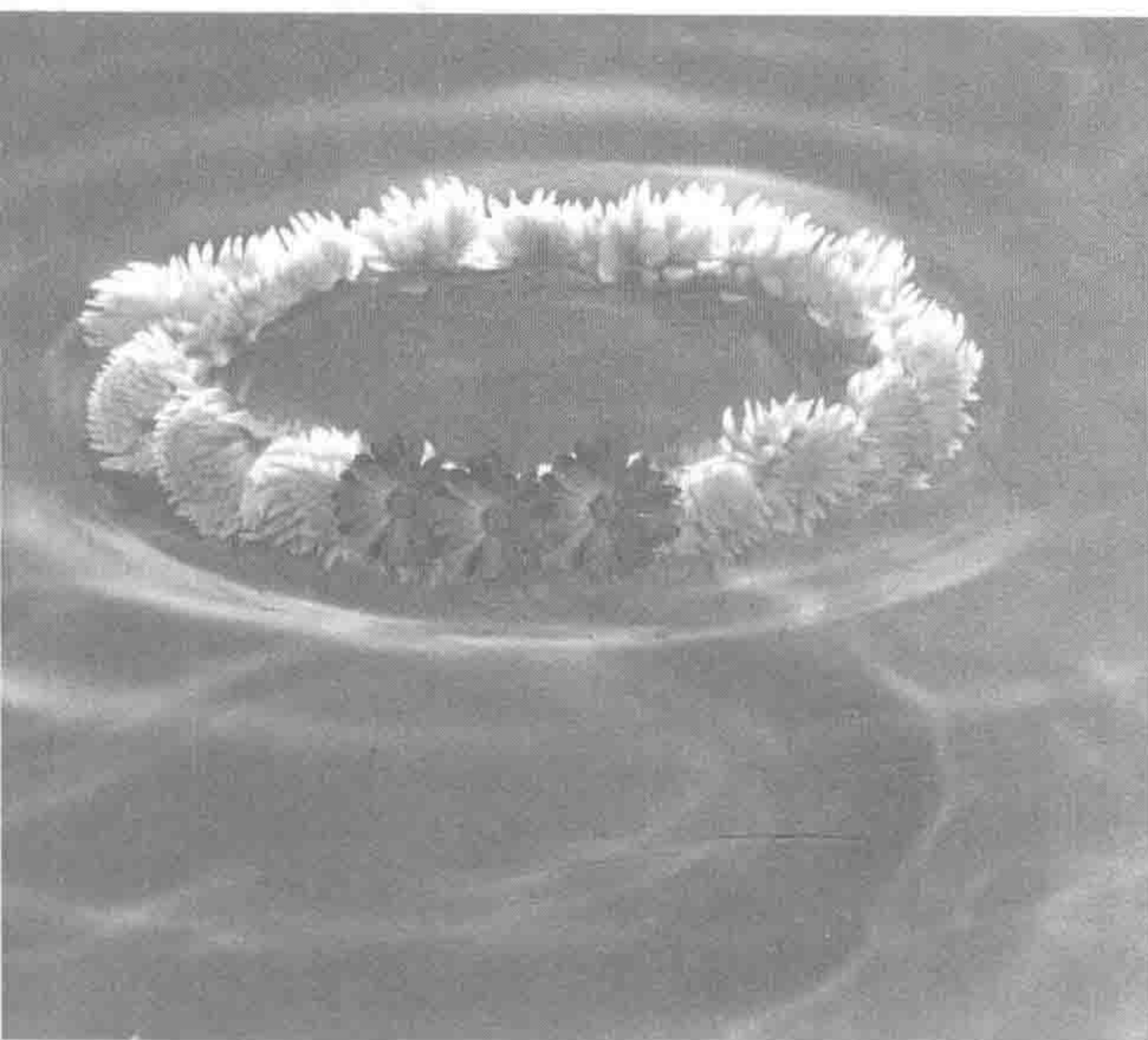
From Basic Research to
Technical Principles

邢万金 编著

高等教育出版社

基因工程

——从基础研究到技术原理



Genetic Engineering

From Basic Research to
Technical Principles

邢万金 编著

高等教育出版社·北京

内容简介

本书共8章，内容包括绪论、基因工程的基本技术与原理、基因工程的酶学基础、目的基因的获取、基因工程载体、基因的体外重组和转移、重组克隆的筛选与鉴定、克隆基因的表达。本书强调数据、实验原理的原始来源、改进和发展历程，讲述的结论和技术发明都力求有学术上的出处，并按照科学发现和技术发明的时间顺序将基因工程的来龙去脉交代清楚，使读者不仅学习到基因工程的基本理论和技术原理，更重要的是了解学术和技术发展的过程。

本书具有以下特点：①以“科学史引导教学改革”的理念为指导，尝试科学史引导的教材编写方式，以历史的视角梳理原创，展示科学发展过程中的创新思维。②正文中标注参考文献，强调学术性，有利于反映科学的发展过程，也便于读者延伸阅读或自行考证。③在DNA序列水平上展示酶切位点、启动子、多克隆位点、终止子、复制起点等重要载体元件的结构与功能，精准解读原理和技术细节。④配套丰富的数字课程资源，包括作者原创的精美动画和教学课件、全程授课录像和1500多篇参考文献。

本书适合作为各类院校的基因工程教学用书，书中提供的详尽学科发展沿革和参考文献也为翻转课堂和讨论式教学改革提供了阅读资源和研讨话题。

图书在版编目（CIP）数据

基因工程：从基础研究到技术原理 / 邢万金编著. -- 北京：高等教育出版社，2018.2
ISBN 978-7-04-049007-7

I. ①基… II. ①邢… III. ①基因工程－高等学校－教学参考资料 IV. ①Q78

中国版本图书馆CIP数据核字（2018）第032334号

JIYIN GONGCHENG

策划编辑 王 莉 责任编辑 王 莉 封面设计 张志奇 责任印制 毛斯璐

出版发行	高等教育出版社	网 址	http://www.hep.edu.cn
社 址	北京市西城区德外大街4号		http://www.hep.com.cn
邮政编码	100120	网上订购	http://www.hepmall.com.cn
印 刷	高教社（天津）印务有限公司		http://www.hepmall.com
开 本	850 mm × 1168 mm 1/16		http://www.hepmall.cn
印 张	24	版 次	2018年2月第1版
字 数	670千字	印 次	2018年2月第1次印刷
购书热线	010-58581118	定 价	58.00元
咨询电话	400-810-0598		

本书如有缺页、倒页、脱页等质量问题，请到所购图书销售部门联系调换
版权所有 侵权必究
物 料 号 49007-00

数字课程（基础版）

基因工程

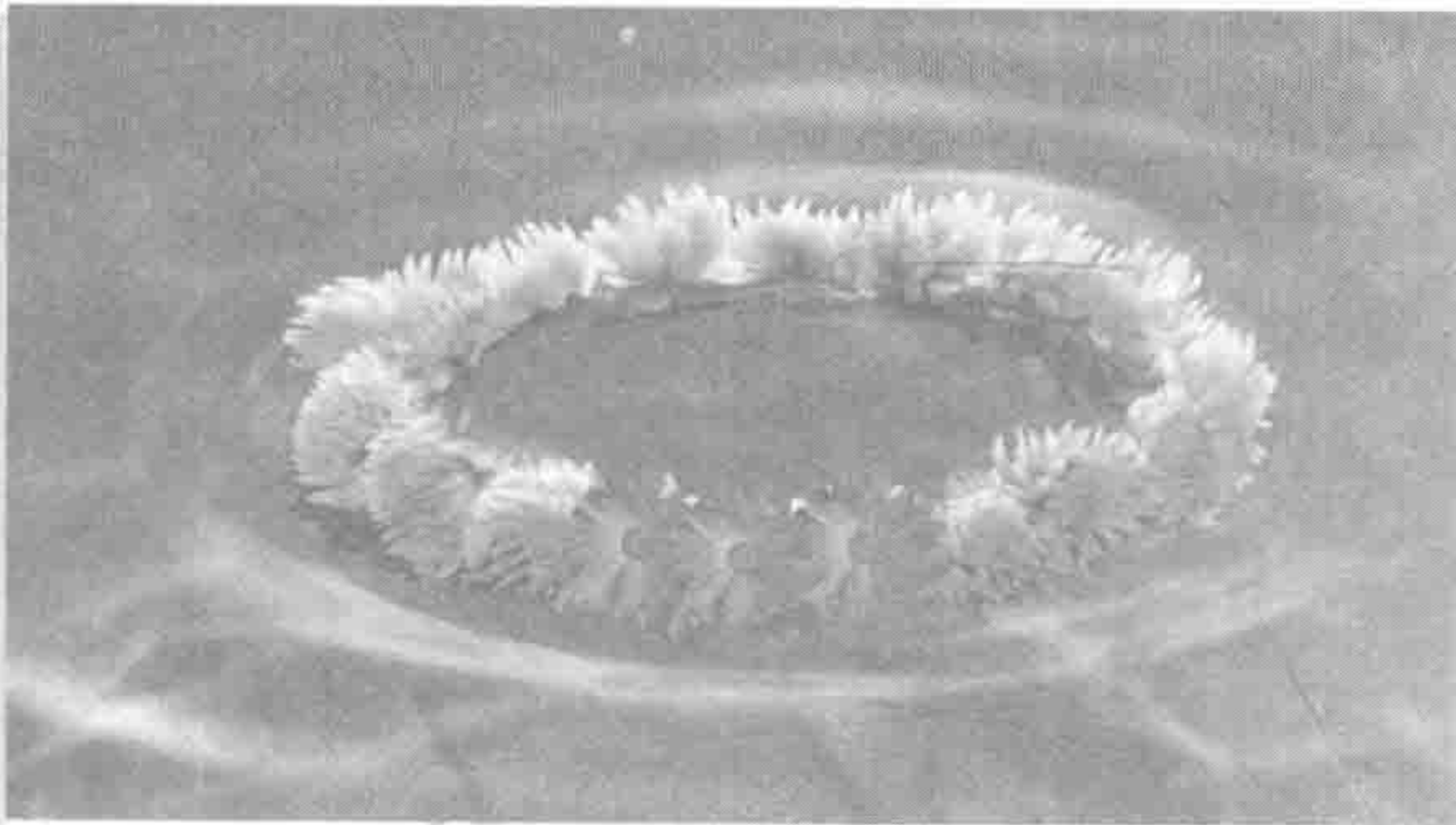
邢万金 编著



登录方法：

- 1. 电脑访问 <http://abook.hep.com.cn/49007>，或手机扫描下方二维码、下载并安装 Abook 应用。
- 2. 注册并登录，进入“我的课程”。
- 3. 输入封底数字课程账号（20 位密码，刮开涂层可见），或通过 Abook 应用扫描封底数字课程账号二维码，完成课程绑定。
- 4. 点击“进入学习”，开始本数字课程的学习。

课程绑定后一年为数字课程使用有效期。如有使用问题，请发邮件至：
lifescience@pub.hep.cn



基因工程

基因工程数字课程与纸质教材一体化设计，紧密配合。立足全面展现课程知识体系并反映学科快速发展的趋势和成果，数字课程涵盖了全课程教学录像和制作精美的动画、教学课件，以及 1500 余篇相关原始文献等资源，充分运用多种形式的媒体资源，丰富知识的呈现形式，更加贴合课程教学的实际需要，在提升课程教学效果的同时，为学生学习提供了更多思考和探索的空间。

用户名：

密码：

验证码：

5360 忘记密码？

登录

注册

<http://abook.hep.com.cn/49007>



扫描二维码，下载 Abook 应用

教师服务

本书另配有专供教师使用的PPT。请选用本书作为学生教材的授课教师发邮件至 lifescience@pub.hep.cn，我们会有专人与您联系，协助您免费获取教师资源。

前言

每每学完一门专业课，笔者皆感叹科学大师之星光灿烂，科学成果之精彩纷呈。前辈大师们一个个奇思妙想和精妙绝伦的实验设计以及令人信服的实验结果，搭建成了本门学科的框架。面对前辈科学家们令人叹为观止的伟大成就，笔者常不禁好奇：他们是怎么想到的？为什么是他们而不是别人？我能做出什么？我该如何做？

基因工程就是这类引人入胜、令人深思的专业课之一。自从Paul Berg团队在斯坦福大学医学中心把SV40病毒DNA片段与携带大肠杆菌半乳糖操纵子的 λ 噬菌体DNA片段在试管中连接起来建立了人工重组DNA技术以后，以基因操作为核心的基因工程就成了生物学家的“魔术”。它不仅愉悦了科学家同行，启迪了莘莘学子，而且已渗透到了生活的各个领域，彻底改变了人类生活和社会发展。无法想象，如果没有基因工程技术，成千上万的糖尿病患者如何能便宜地得到他们日日必须注射的胰岛素，人类如何才能从棉铃虫嘴里夺回棉花，如何才能继续吃上甜绵润口的木瓜，甚至屡遭金融危机的各国如何才能养活如此众多的人口并满足其日益挑剔的口味。正是凭借基因工程，如今我们能按照人类的需求和意愿改造微生物、植物、动物甚至人工合成生命；正是凭借基因工程，生命科学要实实在在地兑现“只有你想不到的，没有我做不到的”的艺术广告。如今这门新兴的“科学魔术”正在不断推陈出新，为人类生活服务，推动社会快速发展。相信，有朝一日，基因工程将会上演魔术师最神奇的“分身术”和“大变活人”等戏法。

毫无疑问，21世纪是生命科学的世纪。改善和解决目前困扰人们的环境、生态、健康、寿命、发展等诸多问题，都将在很大程度上依赖于基因工程。鉴于基因工程已经取得的辉煌成就，我们有理由相信，它是可持续地解决这些问题的不二之选。因此，原本只是理学生物学的一门普通专业课的基因工程已被列为工学、农学、医学、生态学和环境科学等多类别专业的选修课或必修课，以适应不同学科领域的需求。

为不同专业学生开设的基因工程课程自然也打上了各自学科的烙印，具有该学科的特色，甚至具有主讲团队的研究特色。因此，尽管为“国家生命科学与技术人才培养基地”班和“国家理科基础科学研究和教学人才培养基地”班本科生全程讲授了10多年基因工程课程，笔者始终未能从市面上已有的数十个版本的基因工程教材中筛选到反映自身教学思想和满足自身教学目标的课本，而是一直围绕基因工程的基础理论和教学框架自行收集和阅读原始文献并编写教案。经过10多年的持续积累，笔者逐渐形成了自己的课程特色和教学理念，即“学习思想，启迪思维”，主要讲授发现问题和解决问题的研究性思想。学生的反馈和进入研究生阶段的科研体会都表明这种教学理念和与之配套的教学材料更适合培养创新性人才。于是，笔者决定把从1500多篇原始研究文献中凝炼而成的教案整理成书，与广大同行分享教学理念和教学资源，为提高基因工程教学质量奉献微薄之力，也为从事基因操作的研究人员提供一本全面反映基因工程起源、发展和技术原理的专业读物。

本书有别于一般意义上的高等院校专业课教材，具有鲜明的特色，即科学史的视角、基因操作的框架、综述的风格和课本的要求。

(1) 受内蒙古大学莫日根教授提出的“以科学史引导教学”的教改理念的启发，作者尝试以科

学史引导教材建设,沿历史的脉络梳理原创,展示科学发展过程中的创新思维。每项技术都从基础研究或原创写起,包括科学家的姓名、工作单位及当时的科研背景。

(2)以“作者(年代)^[序号]”格式及时标注参考文献,强调学术性,有利于反映学科的发展过程,也便于读者延伸阅读或自行考证。

(3)在DNA序列水平上理解酶切位点、启动子、多克隆位点、终止子、复制起点等重要载体元件的结构与功能,精准解读基因操作原理和技术细节。

(4)学者的工作单位和专业名词在每章中首次出现时给出对应英文,便于科研工作者了解其原始名称。

(5)内容只限于基因工程技术基础和原理,不偏重农学、医学、化工等某一领域的应用。聚焦于基因工程课堂教学内容的少而精,不谋求基因工程所有相关问题的大而全。

(6)配套丰富的数字资源。笔者亲自制作了超炫动画和教学课件、全程教学录像,并提供1 500多篇参考文献。

以上特点和资源使本书适合综合性和师范类高校的基因工程教学,以及工、农、医等院校基因工程基本原理的教学,也适合读者自行阅读和配合教学录像的慕课教学。书中详尽的学科发展沿革和参考文献也为翻转课堂和讨论式教学改革提供了阅读资源和研讨话题。

基因工程尽管看似魔术,但来源于真正的科学,而且是一门高深的学问。基因工程魔术师帽子里的“机关”很多,也极为精巧,但其枢纽皆是基因。因此,为了学懂这门高深的学问,读者需要有扎实的生命科学知识基础,特别是生物化学、细胞生物学和遗传学基础。

那么,基因工程魔术师那顶神奇的帽子里究竟暗藏什么玄机呢?本书将带你穿越基因工程的发展历程,领略每位做出创新成果的基因工程魔术师的奇思妙想,并从专业的角度详细解读他们的手法。这里需要强调的是,科学家需要魔术师的奇思妙想和勤学苦练,也可以用魔术的形式展示科学成果,但不能用魔术的手段和方法开展研究。因此,笔者希望广大读者不仅仅是了解几项耳熟能详的基因工程操作技术,在实验室里能做几个分子生物学实验,而是能透过每项成果和技术去学习前辈科学家的科研思想和方法。化技能为思想,这才是写作本书的初衷。

基因工程是一门飞速发展的学科。本书采用了综述的写作风格,只是为了揭示本学科基础概念的来源和基本技术的建立与发展过程,并无意于写成研究进展,所以仅仅列举了基因工程发展史上承上启下的关键事件,并从时间和内容上指出它们的学术传承,试图勾画出完整的学科发展历程,而有意省略了基因工程技术应用于各领域的最新成果。读者可以从文献报道和相关方向的研究进展中了解基因工程应用的最新成果。叙事风格的写法也不允许笔者在某个段落武断地给出终极的名词解释或“标准化”的实验操作方案,而是留给读者从概念的诞生与发展过程中自行思考和归纳甚至画出操作流程圖,总结出自己理解的概念和原理。这虽是本书的瑕疵,却也算作特点。这样的写法需要读者的想象与思考,适合于把基因工程当做一门学问来研读的读者。

笔者欲尽一己之微能,图穷浩瀚之经书,妄窥科学之精髓,拟绘学科之全貌。虽百省以求无过,百密而防一疏,然限于学识、局于见识、苦于精力、疲于作息,错漏终将难免。但求同仁不吝赐教,感激不尽。精益求精,方不貽笑大方。

邢万金

2017年4月7日于呼和浩特

目 录

绪论	1
第一节 基因工程的诞生.....	2
一、基因工程诞生的理论基础.....	2
二、基因工程诞生的技术突破.....	12
三、基因工程诞生的标志.....	14
四、基因工程的主要操作内容.....	17
第二节 基因工程的应用.....	19
一、基因工程在农业中的应用.....	19
二、基因工程在工业中的应用.....	23
三、基因工程在医药领域的应用.....	23
第三节 基因工程的安全性.....	27
一、基因工程操作的安全性.....	28
二、基因工程产品的安全性.....	28
第一章 基因工程的基本技术与原理	30
第一节 DNA的提取与纯化.....	31
一、质粒DNA的提取.....	31
二、基因组或其他DNA的提取.....	35
三、DNA的浓度和纯度测定.....	35
四、DNA相对分子质量的估计.....	36
第二节 凝胶电泳.....	36
一、自由溶液电泳.....	37
二、固相支持物电泳.....	37
第三节 基因扩增技术.....	44
一、PCR扩增.....	44
二、PCR技术的扩展	49
第四节 核酸序列分析.....	61
一、RNA序列分析.....	61
二、DNA序列分析.....	61
三、DNA序列分析的自动化.....	68
第五节 核酸的分子杂交.....	72
一、核酸杂交探针及其检测方法.....	73

二、核酸分子杂交.....	78
三、核酸杂交技术的拓展.....	81
第二章 基因工程的酶学基础	85
第一节 限制性内切核酸酶.....	86
一、限制修饰的发现.....	86
二、限制性内切酶的发现.....	88
三、限制酶的分类与命名.....	90
四、限制酶的酶活性.....	95
第二节 内切单链DNA酶.....	96
一、稻谷曲霉S1核酸酶.....	96
二、Bal-31核酸酶.....	97
第三节 外切核酸酶.....	97
一、外切单链DNA酶.....	98
二、外切双链DNA酶.....	98
第四节 DNA连接酶.....	100
一、细菌DNA连接酶.....	101
二、T4 DNA连接酶.....	104
三、真核生物的DNA连接酶.....	105
四、影响DNA连接反应的主要因素.....	106
第五节 DNA聚合酶.....	106
一、大肠杆菌DNA聚合酶.....	106
二、T4 DNA聚合酶.....	110
三、T7 DNA聚合酶.....	110
四、逆转录酶.....	111
第六节 DNA末端修饰酶.....	113
一、末端脱氧核苷酸转移酶.....	113
二、T4多核苷酸激酶.....	114
三、碱性磷酸酶.....	116
第三章 目的基因的获取	118
第一节 基因组DNA片段化.....	119
第二节 目的基因的分离.....	119
一、构建基因组文库.....	120
二、构建cDNA文库.....	123
三、富集靶DNA序列.....	131
第三节 PCR扩增获得目的基因.....	137
第四节 化学合成目的基因.....	143
第四章 基因工程载体.....	148
第一节 自然界中细菌之间的DNA交换.....	149
一、转化.....	149

二、接合	149
三、转导	150
第二节 质粒及其分配机制	151
一、质粒的发现	151
二、质粒的一般性质	152
三、天然质粒的分配机制	153
第三节 质粒载体	162
一、天然质粒载体	162
二、人工构建的质粒载体	166
第四节 噬菌体载体	182
一、单链噬菌体载体	182
二、噬菌粒载体	192
三、双链噬菌体载体	195
四、黏粒载体	210
第五节 大片段DNA克隆载体	212
一、酵母人工染色体	213
二、P1来源的克隆载体及P1人工染色体	224
三、细菌人工染色体	234
第五章 基因的体外重组和转移	239
第一节 DNA片段的体外重组	240
一、黏性末端的重组连接	240
二、齐平末端的重组连接	243
第二节 把重组DNA导入宿主细胞	244
一、导入大肠杆菌	244
二、导入酵母细胞	247
三、导入动物细胞	250
四、导入植物细胞	253
第六章 重组克隆的筛选与鉴定	261
第一节 基于载体基因的筛选与鉴定	262
一、抗生素抗性筛选	262
二、互补选择法	266
三、载体DNA电泳检测法	271
第二节 基于目的基因的筛选与鉴定	272
一、载体插入片段的长度鉴定	272
二、载体插入片段的序列鉴定	272
三、目的基因的转录产物鉴定	273
四、目的基因的翻译产物鉴定	274
第三节 用报告基因鉴定	284
一、氯霉素乙酰转移酶基因	285
二、萤火虫萤光素酶基因	286

三、 β -葡萄糖醛酸糖苷酶基因	286
四、荧光蛋白基因	287
第七章 克隆基因的表达	295
第一节 外源基因在原核细胞中表达	296
一、原核表达载体的基本调控元件	297
二、外源基因在大肠杆菌中的表达方式	315
第二节 外源基因在酵母细胞中表达	322
一、酵母表达载体的基本元件	322
二、酵母表达载体的类型	329
三、异源蛋白质在酵母中的分泌表达	333
第三节 外源基因在植物细胞中表达	335
一、植物表达载体的基本元件	336
二、植物表达载体	348
第四节 外源基因在动物细胞中表达	358
一、哺乳动物病毒载体	358
二、昆虫杆状病毒载体	364
三、质粒载体	367
后记	373



绪 论



本章学习目标

了解关于基因结构与功能的若干遗传学基础问题的解决过程，以及在基础研究过程中建立的各种操作 DNA 的技术及其安全性。从历史的高度俯瞰基因工程建立的过程、必然性和其基本操作内容。

就在学术界对 DNA 和蛋白质到底谁是遗传物质存在争议的时候，想象力丰富且不需要提供实验依据的科幻小说作家们早就勾画出了他们心目中的人造生物（artificial organism）甚至制造人类的远景。美国著名科幻小说作家 John Stewart Williamson 在他 1951 年出版的科幻小说 *Dragon's Island*（又名 *The Not-Men*）中提出了“genetic engineering”（遗传工程）这个名词（见 Wikipedia 词条）。

可见，遗传工程这个概念从提出之时即表示人工改造或创造生物。从遗传学的角度看，小说家们创作的遗传工程生物其实是性状的变异。遗传学家们明白，性状是由基因决定的，要改造性状，必须先改造基因，要改造基因，必须先弄清基因是什么。科幻作家可以在想象中任意改造生物，但科学家们却要在现实中一步一个脚印地探索改造生物的可能性。庆幸的是，在众多科学家的努力下，随着基因的实质及其结构与功能逐渐明朗，科学意义上的遗传工程不久就变成了现实。

本节简要介绍遗传工程（我国多称为基因工程，本书也用基因工程这个名称）的孕育和诞生过程，带领读者沿着科学发展的轨迹了解遗传工程的整体内容及其与母体遗传学的联系。这也是本书各章节的写作特色，务求揭示应用技术与基础研究的联系，而不是孤立地叙述某项技术的原理和用途。

第一节 基因工程的诞生

基因工程是遗传学发展到一定程度的必然产物。当遗传学研究确认 DNA 是遗传物质，揭示了 DNA 分子结构，破译了 DNA 中的遗传密码并发现了能切割和连接双链 DNA 的酶以后，尝试体外切割和重新连接 DNA（甚至来自不同生物的 DNA）片段乃是自然而然的进一步学术探究，并在研究基因结构与功能的过程中，建立了操作 DNA 的技术。

一、基因工程诞生的理论基础

1. DNA是遗传物质

自从开始人工栽培植物和驯化野生动物为自己获取稳定的食物来源，人类就一直对部分动植物进行选择繁育（selective breeding），从杂交后代中筛选更适合需求的性状，一代一代地持续改变农作物和驯化动物的性状。英国科学家 Francis Galton 1883 年还提出优生学（eugenics），用于指导对人类自身进行选择繁殖。部分国家的政府甚至实施了优生学行动，引发了严重的社会问题。

早期的人们并不清楚动植物的“好”性状和“坏”性状是由什么决定的。直到奥地利（现捷克共和国）摩拉维亚总督管辖区布尔诺的圣托马斯大教堂（St. Thomas' Abbey in Brno, Margraviate of Moravia）的神父 Gregor Johann Mendel（孟德尔）1866 年〔见 William Bateson（1901）^[1]* 翻译的英文版孟德尔论文〕发表了豌豆杂交实验结果，人们才知道性状是由遗传因子决定的。丹麦遗传学家 Wilhelm Johannsen 在其 1903 年的著作 *Arvelighedslærens elementer*（*The Elements of Heredity*）中使用了“gene”（基因）一词来称呼孟德尔的遗传因子。William Bateson 和 E. R. Saunders（1902）^[2] 在重申孟德尔遗传学的时候用“allelomorphs”区分孟德尔的显性和隐性遗传因子〔后来简化为 allele（等位基因）〕，同时把由携带相反等位基因的两个配子形成的合子称为“heterozygote”（杂合体），携带相似等位基因的两个配子形成的合子称为“homozygote”（纯合子）。但此后将近半个多世纪，人们尽管用大量的动植物杂交实验进一步夯实了等位基因学说，哥伦比亚大学（Columbia University）的 Thomas H. Morgan（摩尔根）1910 年^[3]甚至还用化学试剂诱发果蝇（*Drosophila melanogaster*）突变，筛选到眼和翅的变异性状，并通过杂交实验推论白眼基因位于 X 染色体上，但仍不清楚控制生物性状的基因的理化属性、结构及其生物学功能。

对染色体（chromosome）化学组成的分析结果显示，染色体主要由 DNA 和蛋白质组成，两种物质都是生物大分子（macrobimolecule），DNA 相对分子质量更大。最早发现 DNA 的瑞士医生 Friedrich Miescher 曾一度认为 DNA 在遗传性状（hereditary trait）的传递中扮演了角色，但后来又放弃了这种想法^[4]。洛克菲勒医学研究所（The Rockefeller Institute for Medical Research）的 Phoebus A. Levene（1909）^[5]测定了 DNA 的化学组分，提出四核苷酸假说（tetranucleotide hypothesis），认为 DNA 由 4 种等量的核苷酸互相连接而成，显示 DNA 的化学结构过于简单，不足以储存遗传信息。而蛋白质由 20 种氨基酸组成，20 种氨基酸的排列组合方式显然远远超过 DNA 中 4 种核苷酸的排列组合方式，因而人们认为蛋白质更有可能承担储存遗传信息的重任。

证明 DNA 才是改变生物性状的大分子的直接实验证据来自于洛克菲勒医学研究所的 Oswald Avery、Colin MacLeod 和 Maclyn McCarty（1944）^[6]的细菌转化（transformation）实验（史称 Avery-Macleod-McCarty experiment）。他们从Ⅲ型肺炎链球菌（*Pneumococcus*）SA66 菌株中提纯了 DNA，转

* 全书上角序号对应的文献详见数字课程 ⑥。

化Ⅱ型的R36A菌株，结果明确地显示提纯的DNA能转化肺炎链球菌，而RNA和蛋白质不能。用结晶的胰蛋白酶（trypsin）、结晶的糜蛋白酶（chymotrypsin）和结晶的核酸酶（ribonuclease）处理所提纯的DNA样品都不影响样品的转化活性；但如果用几种能降解脱氧核糖核酸的酶处理就能使样品失活，说明DNA是导致肺炎链球菌转化的物质，即DNA是肺炎链球菌的遗传物质。尽管Avery等的研究成果遭到了同事的强烈质疑，但还是激发了研究分子结构的欧洲科学家们解析DNA的分子结构。

哥伦比亚大学的Erwin Chargaff（1950，1951）^[7, 8]对多种生物DNA的嘌呤（purine）和嘧啶（pyrimidine）含量精细测定后指出，DNA中的4种核苷酸含量并不相同，并发现嘌呤与嘧啶的摩尔数相等，且腺嘌呤（adenine）与胸腺嘧啶（thymine）的摩尔数相等，鸟嘌呤（guanine）与胞嘧啶（cytosine）的摩尔数相等，暗示嘌呤和嘧啶对于维持DNA的某种特殊结构有重要意义。随后，华盛顿卡内基研究所（Carnegie Institution of Washington）的Alfred D. Hershey和他的助手Martha Chase（1952）^[9]在试图揭示T2噬菌体侵染细菌后的繁殖细节时发现亲本T2噬菌体的DNA（³²P标记）传给了子代T2噬菌体，而亲本T2噬菌体的外壳蛋白质（³⁵S标记）并不传给子代T2噬菌体，说明DNA是T2噬菌体的遗传物质（史称Hershey-Chase experiment）。至此，DNA（而非蛋白质）是遗传物质终于成为共识。

2. DNA双螺旋结构的解析

一旦确认DNA是遗传物质，解析DNA分子结构自然成为理解遗传物质功能的迫切要求。遗传物质具有自我复制、承载遗传信息、控制性状、突变和修复等特性。什么样的分子结构才能具备如此复杂而又独特的功能呢？

莫斯科实验生物研究所（Institute of Experimental Biology Moscow）的Nikolai Koltsov（1927）曾惊人地预言遗传物质是一种大分子，由镜像对称的两条链组成，互为模板，以半保留（semi-conservative）方式复制^[10]。尽管DNA和蛋白质都是生物大分子，其单分子的大小仍然无法用显微镜直接观察到，即使通过电子显微镜也无法直接在原子水平上看清其空间结构。当时用于解析物质的原子排列结构的方法是英国利兹大学卡文迪什物理学教授（Cavendish Professor of Physics in the University of Leeds）William Henry Bragg和剑桥三一学院（Trinity College）的William Lawrence Bragg父子（1913）^[11]建立的X线晶体衍射技术（两人获1915年诺贝尔物理学奖）。

X线晶体衍射（X-ray crystallography）技术根据X线穿过结晶样品过程中被样品的原子遮挡和折射后投影在胶片上的图案推算样品中的原子空间排列方式。英国利兹大学的William Thomas Astbury和Florence O. Bell（1938）^[12]用X线晶体衍射技术解析了蛋白质和核酸的分子结构，根据DNA的衍射图案判断DNA是纤维状螺旋长链。伦敦国王学院（King's College, London）的Maurice Wilkins团队和Rosalind Franklin团队也都致力于用X线晶体衍射技术解析DNA的分子结构。随后剑桥卡文迪什实验室（Cavendish Laboratory）的Francis Crick和James Watson也在实验室主任William Lawrence Bragg的支持下开展了解析DNA分子结构的研究。大洋彼岸，对蛋白质分子结构颇有研究的化学键理论权威Linus Pauling也在美国加州理工学院（California Institute of Technology）紧锣密鼓地解析DNA的分子结构，并在*PNAS*杂志上抢先发表了他的3条 α 螺旋（three-chain helix）结构模型^[13]。两个月后，Watson和Crick（1953）^[14]、Wilkins团队（1953）^[15]及Franklin团队（1953）^[16]在同一期*Nature*杂志上以连续页码的形式发表了DNA双螺旋（double helix）结构模型。其中Watson和Crick模型由于强调了A-T和G-C的碱基配对原则（后被称为Watson-Crick配对原则），能很好地解释长期困扰遗传学家们的两个基本问题：DNA如何控制自身合成（复制）和DNA如何控制其他物质合成（转录），从而成为广为接受的DNA双螺旋结构模型。这一成果标志着分子生物学（molecular biology）诞生了！

Watson、Crick和Wilkins由此获得了1962年诺贝尔化学奖（Franklin因卵巢癌于1958年6月辞世）。Pauling虽然未能揭示正确的DNA分子结构，但凭借其化学键理论获得了1954年诺贝尔化学奖。

3. 中心法则的建立

位于 DNA 双螺旋中的基因在细胞内如何控制由蛋白质和酶促反应形成的表型呢？早期对噬菌体（bacteriophage 或 phage）的组成成分的测定为解决这个问题提供了线索。人们测定各种噬菌体的化学成分均显示有蛋白质和高浓度的核酸，也有人报道有脂质（lipid）。杜克大学医学院（Duke University School of Medicine）的 A. R. Taylor（1946）^[17] 仔细测定了 T2 噬菌体的化学成分，发现主要是蛋白质和核酸，还有少量脂质。其中核酸主要是 DNA（40% ~ 46%）和少量 RNA（1.3% ~ 6.6%）。但从细菌培养基中提取的噬菌体可能被细菌细胞渣滓污染，测定子代噬菌体形成前细菌细胞内合成什么物质可能为解决这个问题提供更多的线索。

那么，当大肠杆菌（*Escherichia coli*）被噬菌体感染后，细胞内最初有什么化学物质开始合成呢？宾夕法尼亚大学医学院（School of Medicine, University of Pennsylvania）的 Seymour S. Cohen（1948）^[18] 测定了 T2 噬菌体感染大肠杆菌后的蛋白质和核酸合成情况。正常细菌合成的 RNA 远多于 DNA，但被 T2 噬菌体感染的大肠杆菌合成大量的 DNA 和蛋白质；未感染的细胞内主要合成 RNA（3 倍于 DNA），而被感染的大肠杆菌细胞只合成 T2 的 DNA。如果先用紫外线照射 T2 噬菌体，再感染大肠杆菌，则并不引起上述合成。据此 Cohen 认为 DNA 和蛋白质组成 T2 噬菌体，RNA 不太可能是 T2 噬菌体的组成成分。

Hershey 和 Chase 用放射性同位素 ^{32}P 标记的核酸证明了 T2 噬菌体的核酸才是进入子代噬菌体的遗传物质后，又进一步研究了被 T2 噬菌体感染的大肠杆菌细胞内 ^{32}P 标记的核酸合成与子代噬菌体出现的关系。Hershey（1953）^[19] 发现，大肠杆菌被 T2 感染 10 min 后，合成的主要磷酸是 T2 噬菌体或噬菌体前体（phage precursor）的 DNA。但他也注意到，与未受感染的大肠杆菌不同，被感染的大肠杆菌内会极为迅速地出现 ^{32}P 标记的 RNA，被感染 5 min 就达到每个细胞吸收 25 个单位的磷（每单位定义为 $2 \times 10^{-11} \mu\text{g}$ ）；但当出现 ^{32}P 标记的 DNA 后，RNA 就很快减少，即出现了有代谢活性的 RNA（metabolically active RNA）。

为了确定 T2 噬菌体感染大肠杆菌后迅速合成的 RNA 是噬菌体的 RNA 还是宿主菌的 RNA，美国橡树岭国家实验室（Oak Ridge National Laboratory）的 Elliot Volkin 和 L. Astrachan（1956）^[20] 分析了 T2 噬菌体感染大肠杆菌后细胞内合成的 RNA 的化学组成，确认 T2 噬菌体侵染大肠杆菌后会合成一些 ^{32}P 标记的 RNA，这些 RNA 的碱基组成比例与 T2 噬菌体 DNA 类似，而与大肠杆菌 DNA 的碱基组成比例不同。鉴于 T7 噬菌体的 DNA 与 T2 噬菌体 DNA 的碱基组成区别很大，Volkin 等（1958）^[21] 又测定了 T7 噬菌体感染大肠杆菌后合成的 RNA 上的放射性同位素标记模式，发现与 T2 噬菌体感染后出现的 RNA 的标记模式区别很大，即噬菌体感染后引发的 RNA 似乎与噬菌体有关，而与宿主 DNA 无关。他们甚至猜测这些 RNA 的功能可能是噬菌体蛋白质合成的功能单位——模板（template），存在 RNA-DNA 相互依赖关系。

Volkin（1960）^[22] 随后在培养基中加入 RNA 合成的抑制剂 6-氮尿嘧啶（6-azauracil），发现能抑制子代噬菌体形成，证明噬菌体 T2 侵染大肠杆菌后如果要合成子代 T2 噬菌体颗粒，必须先在大肠杆菌细胞内合成 RNA。即 T2 噬菌体感染大肠杆菌需要先合成自己的“T2 特异性 RNA”，然后才能合成子代噬菌体。后来伊利诺伊大学（University of Illinois）的 Benjamin D. Hall 和 S. Spiegelman（1961）^[23] 用 ^3H 标记的 T2 DNA 作为探针与从被 T2 噬菌体感染的大肠杆菌细胞中提取的 ^{32}P 标记 RNA 进行杂交，证明能形成 RNA/DNA 杂合双链，说明“T2 特异性 RNA”可能是以 T2 DNA 为模板合成的，即 T2 DNA 中的遗传信息通过 RNA 传递给子代 T2 噬菌体的蛋白质外壳。那么，RNA 与蛋白质是如何建立关系的呢？

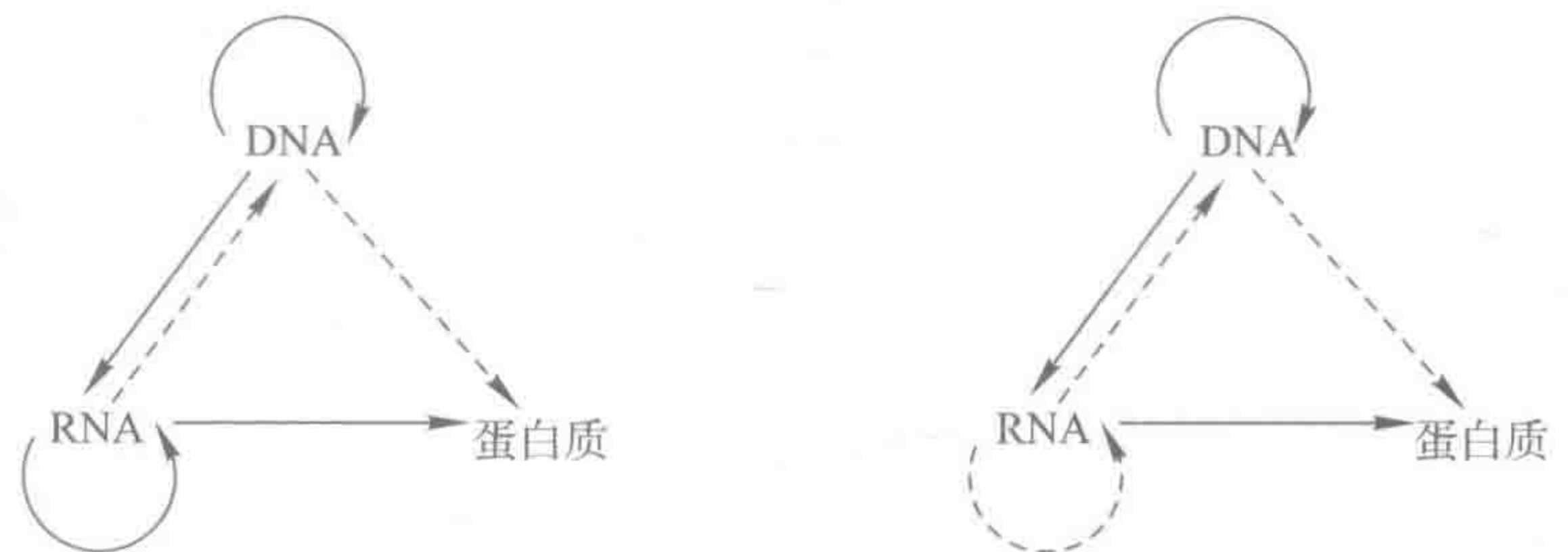
哈佛大学马萨诸塞州总医院（Harvard University at the Massachusetts General Hospital）的 Paul C. Zamecnik 和 Elizabeth B. Keller（1954）^[24] 用大鼠肝提取液的无细胞系统（cell-free system）研究 ^{14}C 标记的亮氨酸（leucine）掺入蛋白质时，发现提取液里的微粒体（microsome）是 ^{14}C 标记的亮氨酸掺入蛋白质的必需组分。后来 Zamecnik 团队（1958）^[25] 在用大鼠肝无细胞系统研究蛋白质合成时发现

一种相对分子质量约 20 000 的小的可溶性 RNA (soluble RNA, 后被称为 tRNA) 与 ^{14}C 标记的亮氨酸偶联, 并把氨基酸转移到微粒体上, 他们称之为 pH5 RNA。洛克菲勒医学研究所的 George E. Palade (1955)^[26] 用电镜观察了细胞内的这种微粒体成分, 发现是规范的颗粒结构。后来 Richard B. Roberts (1958) 提议把 microsome 改称为 ribosome (核糖体) [见 Kenneth Mcquillen 等 (1959)^[27]]。

DNA 位于细胞核内, 而蛋白质是在细胞质里的核糖体上合成, 因此 DNA 并不能与单个氨基酸直接作用。那么, 由 4 种脱氧核糖核苷酸聚合成的 DNA 长链怎么转化成由 20 种氨基酸聚合成的多肽链呢? Francis Crick (1955) 在一篇未公开发表的 RNA 领带俱乐部内部通讯 (A Note for the RNA Tie Club: “On Degenerate Templates and the Adaptor Hypothesis”) 中反驳了 George Gamow 的钻石密码 (diamond code) 理论, 提出了简并 (degeneracy) 密码子假说, 认为 DNA 中的 3 个核苷酸对应一个氨基酸, 4 种核苷酸虽然可以组成 64 种密码子, 但有些密码子代表同一种氨基酸。在这篇文章中, Crick 还提出了接头假说 (adaptor hypothesis), 认为需要 20 种接头 (adaptor) 分子携带 20 种氨基酸到模板上合成蛋白质, 每种接头只能识别双链 DNA 上一条链的三联体密码子。那么什么分子能胜任接头的角色呢?

与被局限在细胞核里的 DNA 不同, RNA 分子在细胞核和细胞质里都有身影。Crick (1958)^[28] 提出了 “central dogma” (中心法则) (图 1 左), 认为 DNA 的遗传信息先传到 RNA, 然后再传到蛋白质里, 即 DNA → RNA → 蛋白质。Howard Martin Temin (1970)^[29] 和 David Baltimore (1970)^[30] 在 RNA 病毒中发现了逆转录酶 (reverse transcriptase) 后, Crick (1970)^[31] 重新解释并进一步强调了中心法则 (图 1 右)。

□ 图 1 Crick 提出的中心法则示意图。
左: 1958 年提出的中心法则。实线箭头表示可能的信息传递方向, 虚线箭头表示不可能的传递方向。右: 1970 年修改的中心法则。实线箭头表示普遍的信息传递方向, 虚线箭头表示特殊的传递方向。



核糖体含有 RNA 和蛋白质, 但不同物种的核糖体 RNA 相对分子质量均一, 组成成分相似, 不大可能是合成种类繁多的蛋白质的模板。那么, 能指导合成蛋白质的模板 RNA 是什么 RNA 呢? 巴斯德研究所 (Institut Pasteur) 的 François Jacob 和 Jacques Monod (1961)^[32] 在全面综述基因表达与调控的研究成果时提出了 “messenger RNA” (mRNA, 信使 RNA) 的概念, 认为一些不稳定的 mRNA 作为中间体, 到细胞质里与核糖体结合合成蛋白质; 核糖体是个非特异性结构, 它只是按照 mRNA 带来的基因信息来合成特定的蛋白质。他们同时也设想, 遗传信息从 DNA 传到蛋白质, 除了需要 mRNA 外, 可能还需要另外一些 RNA 来负责正确翻译 (translation) mRNA 所携带的信息, 如 M. B. Hoagland 和 Paul C. Zamecnik 等^[33] 发现的具有转运氨基酸功能的 pH5 RNA, 当时人们称之为 sRNA (soluble RNA), Jacob 和 Monod 称之为 “transfer RNA”, 后来京都大学 (Kyoto University) 的 Toshiteru Morita 团队 (1964)^[34] 称之为 “tRNA” (transfer RNA, 转移 RNA)。Jacob 和 Monod 认为这些 tRNA 的相对分子质量应该远小于 mRNA, 碱基组成也不反映 DNA 的碱基组成。

4. 遗传密码的破译

Watson 和 Crick 提出了 DNA 双螺旋结构模型后, 关于基因结构与功能的最明显问题之一就是组成 DNA 的 4 种脱氧核糖核苷酸与组成蛋白质的 20 种氨基酸之间的对应关系问题。华盛顿大学 (University of Washington) 理论物理学家兼天文学家 George Gamow 与提出 DNA 双螺旋结构模型的 Watson 决定群策群力来破解这个迷。他们联合物理学、化学和生物学等多个领域的科学家于 1954 年成立了一个由 20 位知名学者组成的俱乐部, 名为 “RNA Tie Club”。俱乐部的每位成员都领到一条领

带，负责破译一个氨基酸的密码。此外还聘请了4位荣誉会员代表4种核苷酸。俱乐部的口号是“Do or die; or don't try.”（见Wikipedia词条）。

Gamow (1954)^[35]根据伸展的DNA中两个核苷酸之间的空间距离与肽链中两个氨基酸残基之间的距离相当，认为这就是个钥匙与锁（key-and-lock）的关系。他认为在Watson和Crick的双链DNA模型中，两条链上的两对配对碱基连线可形成一个菱形，只要有其中3个顶点的核苷酸不同，第四个顶点与另外3个之中的一个相同，就能形成20种不同的菱形“锁孔”，以此对应一种氨基酸“钥匙”。他甚至逐一画出了对应于20种氨基酸的菱形锁孔，后被称为“钻石密码”（diamond code）（见原文图2）。钻石密码模型意味着相邻的2个菱形的4个顶点需要共用2个或3个，即，密码子将会重叠。“RNA Tie Club”的成员Crick在俱乐部内部通讯中反驳过Gamow的钻石密码理论。

另一个成员Sydney Brenner (1957)^[36]也认为密码子不可能重叠。因为假设4个核苷酸A、B、C、D中每3个形成一个氨基酸密码，理论上可形成64（ 4^3 ）种密码子。而如果密码子需要重叠，则ABCD四连体核苷酸就包含2个氨基酸的密码子（ABC和BCD）。而这样的ABCD四连体组合在理论上最多只有256（ 4^4 ）种，即最多能编码256种二肽（dipeptide）。然而，自然界20种氨基酸却可形成400（ 20^2 ）种二肽。因此，重叠密码理论反而导致64种密码子不足以编码多样化的多肽。此外，Crick认为如果ABCD四连体包含2个互相重叠的氨基酸密码子（ABC和BCD），则其中B或C改变将同时导致相邻的2个氨基酸密码子都改变，即如果基因内的某一个核苷酸发生了突变，应该同时引起相邻2个氨基酸突变。这意味着如果我们能比较野生型和突变型基因的产物蛋白质的氨基酸序列，应该发现绝大部分是两个氨基酸的区别，而且总是相邻的。这种现象也与观察到的情况不符。例如，卡文迪什实验室的Vernon M. Ingram^[37]测定了镰状细胞贫血（sickle cell anemia）的血红蛋白HbS与正常人血红蛋白HbA的部分氨基酸序列，发现只有一个氨基酸残基的差异，HbA的 β 链N端第6个氨基酸为谷氨酸（Glu），而HbS为缬氨酸（Val），与之相邻的氨基酸（第5或7）均相同。

关于密码子组成的理论争执需要实验结果来平息。当时研究基因结构与功能的主要手段是借助突变进行基因作图分析。哥伦比亚大学摩尔根团队的Alfred Henry Sturtevant (1913)^[38]根据摩尔根关于染色体上基因线形排列的理论首次绘制了果蝇染色体上若干基因的排列图。Sturtevant的染色体三点作图（three-point mapping）法后来被广泛地应用于真核、原核生物甚至噬菌体染色体作图。尽管果蝇的唾腺染色体很大，但研究基因突变和染色体作图，作为复杂的真核生物的果蝇远不如原核生物大肠杆菌及其噬菌体方便。

普渡大学（Purdue University）的Seymour Benzer对T4噬菌体*rII*基因的精细结构感兴趣，但无意于解析遗传密码子。然而其研究成果却对Crick用实验研究和验证其密码子理论提供了思路 and 材料。T4噬菌体*rII*基因是控制噬菌体在大肠杆菌K株中生长的裂解基因，*rII*发生突变将不能在大肠杆菌K株中生长（但可以裂解大肠杆菌B株）。Benzer (1955)^[39]制备了很多T4噬菌体*rII*基因的突变体，通过互补杂交发现这些突变属于两个相邻且能互补的连锁群（linkage group），即两个功能单位，*rIIA*和*rIIB*，他称之为顺反子（cistron）。

但两个顺反子内部的那些不能互补的突变位点又是如何排列的呢？也是线形排列的吗？彼此相距多远？当时人们对基因内部的结构毫无知晓。Benzer决定继续用杂交作图策略来测定*rII*位点的两个顺反子内部突变位点之间的图距。当时人们认为基因是染色体上的单个不能继续分割的核苷酸串。他这种要用杂交重组（需要分割基因）的策略测定基因内部突变位点排列方式的想法听起来甚为荒诞，以至于他的博士后导师Max Delbrück都觉得离谱。但是，执著的Benzer (1959, 1961)^[40, 41]发明了缺失作图（deletion mapping）法，与三点杂交作图法相结合，硬是绘制了*rII*基因内2400个突变位点的排列图谱。这些突变位点分别位于*rIIA*和*rIIB*顺反子中，两个位点之间最小的重组率是0.02 cM，相当于1.6 bp（1~2个核苷酸）的距离。Delbrück后来感叹道，Benzer真的做到了“Run the genetic map into the ground”。

尽管 Benzer 未被邀请加入“RNA Tie Club”，也并未参与破解遗传密码子，但他的实验结果说明基因内部的核苷酸序列也是连续线形排列的，而非两条链上扭曲的菱形排列，也不大可能有分支。单个核苷酸可能就是一个影响顺反子功能的突变位点。这在当时达到了解析基因结构与功能最精细的境界。“RNA Tie Club”的成员、剑桥卡文迪什实验室的 Crick 和 Brenner 立刻意识到 Benzer 的 T4 噬菌体 *rII* 基因突变实验可能有助于阐明核苷酸与氨基酸之间的关系，于是他们（1961）^[42] 也对 *rII* 基因进行了突变实验，发现吖啶（acridine）类染料，如原黄素（proflavin）处理引起的突变几乎总是导致基因功能彻底丧失的负突变，而不出现还保留部分功能的渗漏突变，而用碱基类似物（base analog）引起的突变则一般是渗漏突变。

Crick 等（1961）^[43] 还发现原黄素处理引起的突变也可再用原黄素处理实现回复突变（reverse mutation），但这种回复突变并非把原有的突变部位本身回复，而是在原突变部位附近的另外一处引发一个新的突变，即同一个基因内部的双突变反而回复成野生型。上述奇怪的突变特点说明原黄素引起 *rII* 基因突变的分子机制不是改变某个核苷酸，而是添加或删除一个或几个核苷酸，因为如果密码子是以固定的位置和核苷酸数目线形排列，则某个部位添加或删除一个核苷酸引起的突变效果就可以被相邻部位删除或添加一个核苷酸来矫正。由此，他们认为 *rII* B 顺反子可能编码一条有功能的多肽，其密码子信息为相邻的 3 个核苷酸组成一个密码子，从某个固定的核苷酸开始向一个方向连续阅读。相邻密码子互不重叠，但也不间隔。如果某个位点插入一个核苷酸，将导致下游的密码全部错读。这些三联体密码（triplet code）可能简并，即多个不同的密码子代表同一种氨基酸。

虽然密码子的雏形呼之欲出，但怎么才能精确鉴定 DNA 中可能形成的 64 种三联体密码的具体核苷酸组成与蛋白质里的 20 种氨基酸的对应关系呢？根据中心法则理论，DNA 并不直接指导合成蛋白质，而是先合成 mRNA，由 mRNA 中的密码子指导合成多肽链。如果能在试管里人工合成 mRNA，再令其在细胞提取液（无细胞蛋白合成系统）中合成多肽链，然后测定肽链中的氨基酸序列，对比 RNA 中的核苷酸序列和蛋白质中的氨基酸序列，即可识别编码氨基酸的核苷酸密码序列。但如何才能合成 RNA 呢？

纽约大学医学院（New York University College of Medicine）的 Severo Ochoa 在研究氧化磷酸化（oxidative phosphorylation）的分子机制时想找一种能把 ADP 转化为 ATP 的酶，以便能测定放射性同位素 ³²P 掺入 ATP 的化学反应。Ochoa 认为棕色固氮菌（*Azotobacter vinelandii*）的氧化活性高，可能具有这样的酶。他的博士后 Grunberg-Manago（1955）^[44, 45] 果然从棕色固氮菌提取物中发现了一种能把 5'-二磷酸核苷酸（5'-nucleoside diphosphate）通过 3',5'-磷酸二酯键连接成多核苷酸（RNA）并释放出磷酸根的酶，他们称之为多核苷酸磷酸化酶（polynucleotide phosphorylase）。他们用这种酶在体外合成了单一核苷酸多聚物（single polymer），如多聚 AMP（polyA）和多聚 IMP（polyI），也用 2 种或多种不同比例的核苷酸混合合成了混合多聚物（mixed polymer），如用等摩尔量的 ADP 和 UDP 合成多聚 A-U [poly(AU)]，或用 ADP : GDP : UDP : CDP（1 : 0.5 : 1 : 1）合成 A-G-U-C 多聚物等，而且合成的多聚物的理化性质与天然 RNA 并无区别。这是第一个能在体外合成 RNA 的酶。Ochoa 因此与发现 DNA 聚合酶 I（1958 年）的 Arthur Kornberg 分享了 1959 年诺贝尔生理学或医学奖。

就在 Ochoa 发现多核苷酸磷酸化酶的同时，哈佛大学马萨诸塞州总医院的 Paul C. Zamecnik 和 Keller（1954）^[24] 用大鼠肝提取液建立了无细胞蛋白质合成系统，用于体外研究蛋白质的合成情况。受其启发，西储大学医学院（Western Reserve University School of Medicine）的 J. A. DeMoss 和 David Novelli（1955）^[46] 也用多种细菌提取液建立了细菌无细胞蛋白质合成系统。借助无细胞蛋白质合成系统，人们就可以测试人工合成的 RNA 片段能否作为模板指导蛋白质合成，并进而破译遗传密码。Ochoa 发现的多核苷酸磷酸化酶旋即成了破解 RNA 中密码子的“罗塞塔石碑”（Rosetta Stone）。

拿到诺贝尔奖的 Ochoa 本人却并未意识到他手里的多核苷酸磷酸化酶的巨大潜力。第一个看到无细胞蛋白质合成系统与多核苷酸磷酸化酶这两大利器结合的威力的学者是美国国立卫生研究院