



PRINCIPLES OF GENETIC ENGINEERING

# 基因工程原理

徐晋麟 陈 淳 徐 沁/编著



 科学出版社  
www.sciencep.com

# 基因工程原理

徐晋麟 陈 淳 徐 沁 编著

科学出版社

北 京

## 内 容 简 介

本书是作者在多年教学经验的基础上精心编写的,力求内容新颖、系统明了、原理清楚、方法先进。本书共分9章,基因的分子生物学基础、原核生物分子克隆的宿主和载体系统、基因文库的构建、基因组DNA的分析、cDNA文库的构建和筛选、聚合酶链反应、培养细胞中克隆基因的表达、转基因动植物和基因治疗。在每章的后面都附有一些习题,供读者练习,书后附有缩略语表和索引。

本书适合高等院校生物科学、生物技术专业和农林、医药院校相关专业学生作为教科书使用,也可供相关科学研究人员参考。

### 图书在版编目(CIP)数据

基因工程原理/徐晋麟,陈淳,徐沁编著. —北京:科学出版社,2007  
ISBN 978-7-03-019217-2

I. 基… II. ①徐…②陈…③徐… III. 基因-遗传工程-高等学校-教材 IV. Q78

中国版本图书馆CIP数据核字(2007)第125372号

责任编辑:周辉 彭克里 席慧/责任校对:张琪  
责任印制:张克忠/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>

陈海印刷有限责任公司印刷

科学出版社发行 各地新华书店经销

2007年8月第 一 版 开本:787×1092 1/16

2007年8月第一次印刷 印张:16 1/2 插页1

印数:1—3 000 字数:373 000

定价:28.00元

(如有印装质量问题,我社负责调换<长虹>)

## 前 言

自从 20 世纪 70 年代初美国斯坦福大学的伯格 (P. Berg)、加州大学的博耶 (H. W. Boyer) 和斯坦福大学的科恩 (S. N. Cohen) 开创了基因工程技术以来, 生命科学像是装上了巨大的引擎, 迅猛向前发展, 遗传学迅速地进入了一个崭新的“基因组时代”。正是基因工程技术提供了很多新的克隆策略和技术, 使得人类完成了两大创举, 一是体细胞克隆, 二是人类基因组计划的完成。

基因工程技术不仅是一门应用学科, 也是研究生命科学极为重要的工具, 不懂得生物工程技术也就很难进入生命科学实验室。目前基因工程实验手册之类的书已出版了很多, 最有影响的可能是 J. 萨姆布鲁克、E. F. 弗理奇和 T. 曼尼阿蒂斯编著的《分子克隆实验指南》。该书近乎成为与现代生命科学研究有关的实验室中的必备参考手册。1993 年出版的卢圣栋主编的《现代分子生物学实验技术》(高等教育出版社), 是国内出版的第一本分子生物学方面的实验指南, 深受读者的欢迎。尽管相关书籍陆续问世, 但有关基因工程的教材出版得很少。1987 年科学出版社出版的黄翠芬院士主编的《遗传工程理论与方法》是我国第一本基因工程教材, 刚一出版就一售而空。同年, 科学出版社组织翻译出版了由 J. D. 沃森、J. 图泽, D. T. 库尔茨著 (沈孝宇等译, 王培楠校) 的《重组 DNA 简明教程》。这是沃森所编著的四本名著之一, 弄得一时洛阳纸贵。接着 1988 年四川大学出版社出版了齐义鹏等编著的《基因工程原理与方法》, 这是国内第一本系统介绍基因工程原理和方法的教科书, 从事基因工程教学的老师们如获至宝。1989 年吴乃虎主编《基因工程原理》(第一版) 问世, 掀起了一股绿色的冲击波 (书的封面为绿色), 几乎占据了全国的教材市场, 并在专业领域连续多年被采用。1998 年吴乃虎先生出版了《基因工程原理》(第二版) 上册, 2001 年出版了下册, 和第一版相比内容更为丰富, 装帧更为精美。可惜的是各高校本科生的“基因工程原理与方法”这门课一般都只安排 36 个学时, 也就无法再用这上、下两册精装书作为教材。

多年来我一直从事遗传学、分子生物学和基因工程原理与方法的的教学, 前两门课都有很多国外著名教材可参考, 亦有较多的国内编写的教材可选用, 唯独感到困难的就是“基因工程原理与方法”。为了适应教学的需要, 我们参考了 Roger L. Miesfeld 编著的 *Applied Molecular Genetics* 等书和近期的一些文献资料, 编写了本书, 共分 9 章, 适合于 36~54 个学时的教学。我们在编写过程中, 力求内容新颖, 系统明了, 原理清楚, 方法先进, 深入浅出, 有一定的启发性。在每章的后面都附有习题, 供读者练习, 以便于读者把握重点, 加深理解。希望这本教材对读者在学习基因工程原理和方法方面有所帮助。

徐晋麟

于上海交通大学

2007 年 6 月 15 日

# 目 录

## 前言

第一章 基因工程的分子生物学基础	1
第一节 遗传信息流的中心法则 (central dogma): DNA→RNA→蛋白质	2
一、DNA 合成	3
二、RNA 合成	3
三、蛋白质合成	3
第二节 核酸的生物化学	4
一、核酸大分子的结构	4
二、DNA 双螺旋的变性与复性	4
第三节 基因工程中采用的主要酶类	7
一、序列特异的 DNA 限制性内切核酸酶	7
二、连接酶和激酶	9
三、DNA 聚合酶和 RNA 聚合酶	9
第四节 研究 DNA 和 RNA 的方法	10
一、核酸的分离纯化	10
二、凝胶电泳	10
三、DNA 测序	11
第五节 膜印迹和核酸杂交	16
第六节 实例 1: 鉴别特定基因转录物的转录起始点	18
一、研究目的	18
二、可利用的信息和资源	18
三、基本策略	19
四、注释	21
五、可供参考的思路	21
习题	21
第二章 原核生物分子克隆的宿主和载体系统	24
第一节 应用广泛的细菌宿主	24
一、 <i>E. coli</i> K-12 的生物学	24
二、乳糖操纵子	25
三、细菌限制与修饰系统	27
四、大肠杆菌的克隆载体	29
第二节 质粒载体	29
一、质粒生物学	29
二、质粒上常编码抗生素-抗性基因	30

三、质粒 DNA 克隆载体 .....	32
四、通过转化将 DNA 转移到大肠杆菌中 .....	33
第三节 噬菌体克隆载体 .....	35
一、 $\lambda$ 噬菌体的生物学 .....	35
二、 $\lambda$ 噬菌体克隆载体 .....	36
三、M13 噬菌体克隆载体 .....	39
第四节 外源 DNA 和载体的连接 .....	41
一、外源 DNA 和载体连接的方法 .....	41
二、定向克隆 .....	42
第五节 实例 2: 一个基因在 <i>E. coli</i> 中的调节表达 .....	44
一、研究目的 .....	44
二、可利用的信息和资源 .....	44
三、基本策略 .....	44
四、注释 .....	47
五、可供参考的思路 .....	47
习题 .....	47
<b>第三章 基因文库的构建 .....</b>	<b>49</b>
第一节 DNA 基因文库 .....	50
一、基因组文库 .....	50
二、DNA 文库的构建 .....	50
三、用于构建 DNA 文库的克隆载体 .....	53
第二节 亚克隆 .....	54
第三节 克隆 DNA 序列的体外诱变 .....	56
一、缺失诱变 .....	57
二、寡核苷酸定向诱变 .....	58
三、随机诱变 .....	60
四、PCR 诱变法 .....	60
五、大引物诱变法 .....	61
第四节 实例 3: 基因编码序列的丙氨酸扫描诱变 .....	62
一、研究目的 .....	62
二、可利用的信息和资源 .....	62
三、基本策略 .....	62
四、注释 .....	63
五、可供参考的思路 .....	63
习题 .....	65
<b>第四章 基因组 DNA 的分析 .....</b>	<b>67</b>
第一节 基因组的组成 .....	68
一、大肠杆菌 K-12 基因组的组成 .....	68
二、酿酒酵母基因组的组成 .....	69

三、人类基因组的组成 .....	69
第二节 基因组作图 .....	71
一、用减数分裂的重组频率进行连锁作图 .....	72
二、用放射诱导染色体重排进行基因组作图 .....	72
三、用 DNA 序列多态性作为遗传标记进行基因组作图 .....	74
四、用遗传资料、DNA 标记和克隆片段来构建连接群图谱 .....	76
第三节 长的基因组 DNA 片段的操作 .....	77
一、用荧光激活细胞分选法来分选染色体 .....	78
二、标签克隆法 .....	78
三、染色体的显微切割 .....	79
四、脉冲电场凝胶电泳 .....	79
第四节 基因组 DNA 克隆载体 .....	81
一、黏粒载体 .....	82
二、酵母人工染色体 (YAC) 载体 .....	83
三、细菌人工染色体 (BAC) 载体 .....	84
四、P1 噬菌体载体和 P1 人工染色体 .....	85
五、哺乳动物人工染色体 (MAC) 载体 .....	85
六、用 DNA 池来筛选基因组文库 .....	86
第五节 基因调节序列的作图 .....	88
一、用体外表达进行基因调节序列作图 .....	88
二、定位蛋白质结合位点的分析方法 .....	89
三、电泳迁移率变动分析 .....	91
第六节 实例 4: 通过定点克隆分离一个人类致病基因 .....	92
一、研究目的 .....	92
二、相关信息和资源 .....	92
三、基本策略 .....	92
四、注释 .....	92
五、可供参考的思路 .....	93
习题 .....	94
第五章 cDNA 文库的构建和筛选 .....	102
第一节 将 RNA 转变成 cDNA .....	102
一、mRNA 的分离纯化 .....	102
二、cDNA 合成 .....	104
三、cDNA 克隆载体 .....	106
第二节 cDNA 文库的筛选 .....	108
一、简并寡核苷酸探针 .....	109
二、抗体探针 .....	110
三、差示杂交 .....	112
四、扣除杂交 .....	113

第三节	cDNA 表达文库的功能筛选	114
一、	蛋白质活性分析	115
二、	酵母双杂交系统	115
三、	cDNA 噬菌体展示	117
第四节	用克隆 cDNA 作为反应物	119
一、	RNA 印迹	119
二、	RNase 保护实验	119
三、	核连缀转录分析	120
第五节	实例 5: 体外克隆编码细胞内信号蛋白的基因	122
一、	研究目的	122
二、	可利用的信息和资源	122
三、	基本策略	122
四、	注释	122
五、	可供参考的思路	124
习题		124
第六章	聚合酶链反应	126
第一节	基本的 PCR 方法	126
一、	PCR 扩增循环	127
二、	PCR 需要一种热稳定 DNA 聚合酶	129
三、	PCR 引物的设计	129
四、	两个引物之间的最佳距离	131
五、	PCR 实验的最优化	131
第二节	反转录 PCR	132
一、	cDNA 末端快速扩增 (RACE)	132
二、	定量 RT-PCR	134
三、	差异表达基因的扩增	136
四、	肠道细菌基因间重复保守序列-聚合酶链反应	139
第三节	PCR 用于诊断	139
一、	在组织样本中检测病原体	139
二、	鉴别遗传突变	140
三、	用标签序列位点进行基因型分型	141
第四节	PCR 在实验室中的应用	142
一、	用 PCR 亚克隆靶 DNA	142
二、	PCR 介导产生融合基因	144
第五节	解链酶复制法	145
第六节	实例 6: 用 RT-PCR 克隆细胞特异转录物	145
一、	研究目的	145
二、	可利用的信息和资源	145
三、	基本策略	146



四、注释 .....	148
五、可供参考的思路 .....	148
习题 .....	148
<b>第七章 培养细胞中克隆基因的表达</b> .....	<b>149</b>
<b>第一节 基因调控的研究</b> .....	<b>149</b>
一、表达载体 .....	149
二、报道基因 .....	151
三、反义技术 .....	156
四、RNA 干扰 .....	158
<b>第二节 基因在细胞系中表达的方法</b> .....	<b>160</b>
一、基因转染的策略 .....	161
二、瞬时转染 .....	164
三、稳定转染 .....	165
<b>第三节 用酵母作为模式真核细胞</b> .....	<b>167</b>
<b>第四节 蛋白质在培养细胞中的表达</b> .....	<b>168</b>
一、 <i>E. coli</i> 表达系统 .....	169
二、杆状病毒 (baculovirus) 表达系统 .....	170
<b>第五节 实例 7: 启动子选择转录因子的研究</b> .....	<b>171</b>
一、研究目的 .....	171
二、可利用的信息和资源 .....	172
三、基本策略 .....	172
四、注释 .....	173
五、可供参考的思路 .....	175
习题 .....	175
<b>第八章 转基因动植物</b> .....	<b>176</b>
<b>第一节 果蝇发育的分子生物学</b> .....	<b>176</b>
一、P 因子介导的转化 .....	177
二、P 因子增强子捕获载体 .....	177
三、可介导转化的其他真核转座因子 .....	180
<b>第二节 构建转基因的农作物</b> .....	<b>181</b>
一、根癌农杆菌介导的基因转移 .....	181
二、用基因枪制备转基因水稻和玉米 .....	183
<b>第三节 转基因小鼠</b> .....	<b>186</b>
一、用受精卵细胞制备转基因小鼠 .....	186
二、通过干细胞的同源重组进行基因敲除 .....	189
三、人类疾病的转基因小鼠模型 .....	193
<b>第四节 转基因家畜</b> .....	<b>194</b>
一、用转基因动物制备药物 .....	194
二、动物的体细胞克隆 .....	195

第五节 实例 8: 敲除转基因鼠产生特异基因 .....	199
一、研究目的 .....	199
二、可利用的信息和资源 .....	199
三、基本策略 .....	199
四、注释 .....	200
五、可供参考的思路 .....	201
习题 .....	201
<b>第九章 基因治疗</b> .....	<b>205</b>
<b>第一节 基因治疗的策略</b> .....	<b>206</b>
一、体细胞基因治疗的策略 .....	206
二、基因治疗的关键步骤 .....	208
三、基因治疗的靶细胞的选择 .....	209
四、体细胞基因治疗的途径 .....	209
<b>第二节 基因转移的方法</b> .....	<b>210</b>
一、基因转移的非生物方法 .....	211
二、基因转移的病毒方法 .....	212
<b>第三节 基因治疗药物</b> .....	<b>218</b>
一、p53 抑癌基因 .....	218
二、siRNA .....	219
三、反义 RNA .....	221
四、核酶 .....	221
五、肽核酸 .....	224
六、三链 DNA .....	225
七、自杀基因 .....	227
<b>第四节 基因治疗的安全性</b> .....	<b>228</b>
<b>第五节 实例 9: 重症联合免疫缺陷的基因治疗</b> .....	<b>229</b>
一、研究目的 .....	229
二、可利用的信息和资源 .....	229
三、基本策略 .....	230
习题 .....	231
<b>主要参考文献</b> .....	<b>233</b>
<b>附录 缩略语表</b> .....	<b>234</b>
<b>索引</b> .....	<b>245</b>
<b>彩版</b>	

## 第一章 基因工程的分子生物学基础

**基因工程** (genetic engineering) 是指用分离纯化或人工合成的外源基因在体外插入到载体分子中, 成为重组 DNA, 再导入宿主细胞内, 进行扩增和表达。基因工程也称为遗传工程、分子克隆 (molecular cloning)、基因操作 (gene manipulation) 或重组 DNA 技术 (recombinant DNA technology)。

在 20 世纪 60 年代后期, 一些遗传学家们认为遗传学已经发展到了巅峰状态, 似乎很难有新的重大发现, 因此有的遗传学家便转到生命科学的其他领域去寻找新的战机。例如, 1955 年建立了染色体的精细作图方法和互补测验, 并提出了“顺反子” (cistron) 概念的美国著名遗传学家 S. Benzer 于 1960 年改变研究方向, 从事昆虫神经生理的研究。此时遗传学的发展处在一个“山穷水复疑无路”的状态。

1962 年瑞士的 W. Arber 提出限制与修饰假说来解释宿主对噬菌体生长的限制。根据这个模型, 细菌的 DNA 中含特异的核苷酸序列可被细胞中的一种限制性内切核酸酶识别和剪切。这些细菌同时还含有一种甲基化酶, 这种酶可以甲基化上述特异的核苷酸序列。经过甲基化的化学修饰保护了细菌的 DNA, 使它们不会被本身的限制性内切核酸酶所降解。但其可以降解外来未经过甲基化的噬菌体 DNA。因此他预言了限制性内切核酸酶的存在。1968 年, H. O. Smith 等分离出第一个限制性内切核酸酶。1966 年 B. Weiss 和 C. C. Richardson 就已经分离了 DNA 连接酶, 这样就可以对 DNA 进行体外加工了。

1972 年美国斯坦福大学的 P. Berg 等将猿猴病毒 (SV40) 的 DNA 和  $\lambda$  噬菌体的 DNA 进行体外重组, 开创了体外重组的先河。他因此项杰出贡献而获得了 1980 年的诺贝尔化学奖。Berg 用限制酶 *EcoR* I 分别切割两种 DNA 分子, 然后再用外切核酸酶进行处理, 将两种 DNA 的末端切成平端, 再用连接酶在 SV40 的片段的 3' 端接上寡聚 A, 而在  $\lambda$  DNA 片段 3' 端接上寡聚 T, 使二者互补结合, 再用连接酶封闭缺口, 将它们连接成异源嵌合的环状 DNA。当时市场上并没有限制酶 *EcoR* I 出售, Berg 是从加州大学 H. W. Boyer 处获取的。1972 年 Boyer 等就在研究限制酶 *EcoR* I, 并鉴别了  $\lambda$  噬菌体 DNA 中被特异的内切核酸酶所识别的核苷酸序列。斯坦福大学的 S. N. Cohen 等证明大肠杆菌能接纳环状质粒 DNA 分子, 并利用质粒所携带的抗生素抗性基因来筛选细菌群落中的转化株。1973 年, Coher 和 Boyer 等人, 用限制酶 *EcoR* I 切割含有抗四环素基因的质粒 pSC101 和来自鼠伤寒沙门氏菌的质粒 RSF1010 (带有抗链霉素和磺胺基因), 并用连接酶进行拼接, 构建一个新的 DNA 分子, 即杂种质粒。用这一杂种质粒转化大肠杆菌后, 转化子既抗四环素, 又抗链霉素。从而第一个建立了基因工程技术。1974 年他们又将非洲爪蟾的 DNA 切成片段, 连接到质粒 pSC101 中并转移到大肠杆菌中, 使沉寂的遗传学进入了“柳暗花明又一村”。一个新的“基因工程”时代展现在人们的眼前。它既给人们带来了惊奇、兴奋和希望, 同时也带来了惶恐、争议和不安。

基因工程所依赖的基础理论主要是: ①1953 年 J. D. Watson 和 F. H. C. Crick 建立

的 DNA 的双螺旋模型。②1946 年 J. L. Lderberg 等相继发现在细菌和噬菌体中遗传物质横向传递的一系列现象和规律。③1961 年法国巴斯德研究所的 F. Jacob 和 J. Monod 建立的操纵子模型。④美国国立卫生研究院 M. W. Nirenberg (1964) 和威斯康星大学的 H. G. Khorana (1967) 等分别采用三联体结合实验和合成重复共聚物破译了全部有义密码子。

基因工程所依赖的技术之多，发展速度之快也是令人咋舌的，但使得基因工程得以问世的基本技术是一些重要的酶的发现，如 1956 年美国的 A. Kornberg 发现了 DNA 聚合酶；1966 年 B. Weiss 和 C. C. Richardson 分离出 DNA 连接酶；1968 年美国的 H. O. Smith 发现了 II 类限制酶并阐明了其性质；1970 年 D. Baltimore、H. M. Temin 和 S. Mizutani 两个组分别报道在两种癌基因的 RNA 病毒中发现了反转录酶。而 1987 年由美国的 K. B. Mullis 和 K. Saiki 等建立的 PCR 体外扩增技术更使得基因工程如虎添翼。

在分子遗传学中对于基因工程中所涉及的基础理论已有了系统地描述，但为了本书的系统性和便于读者阅读，我们也在下面简明扼要地介绍一些相关的知识。

## 第一节 遗传信息流的中心法则 (central dogma): DNA→RNA→蛋白质

DNA 和 RNA 分别由很多的单脱氧核苷酸 (deoxynucleotide) 和单核苷酸 (ribonucleotide) 聚合而成。除部分病毒以外，大部分生物的细胞中储存着双链 DNA 分子，作为其生命的蓝图，它和组蛋白结合，形成染色体。真核细胞在细胞分裂期中，染色体进行复制，在减数分裂中染色体进行交换和重组。所谓**基因表达** (gene expression)，

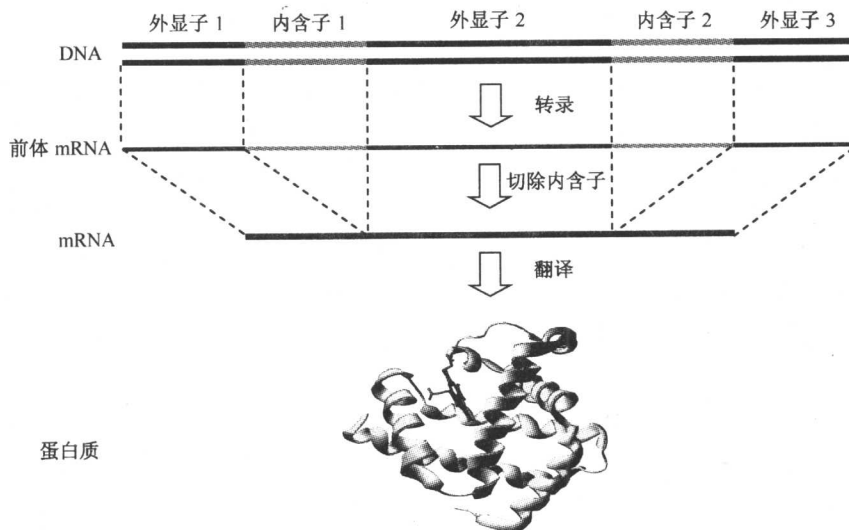


图 1-1 细胞中遗传信息传递的过程。基因是遗传信息的储存单位，包括编码序列和转录调节序列。在真核细胞中，大部分基因编码序列含有外显子，它们被不编码的内含子所隔开。转录是从基因 5' 端启动子的起始区 (I) 开始的，然后延伸直到终止，产生初始 mRNA。经过剪切和加上 3' 多聚腺苷尾巴形成成熟的 mRNA。mRNA 的翻译是在真核细胞的细胞质中进行的。在原核细胞中  
转录和翻译是偶联进行的

就是指 DNA 分子经转录产生互补的 RNA 分子 (图 1-1)。

作为基因编码序列的 DNA 分子携带着遗传信息,通过转录出初始 mRNA,经过加工产生成熟的 mRNA,也称为**转录物** (transcript),再翻译成相关的蛋白质。DNA 是信息的载体,只有蛋白质才能行使功能。DNA 的复制是负责遗传信息的传递;转录和翻译负责遗传信息的形式转换。中心法则看起来比较简单,但却是分子生物学的基本规则。

为了了解基本的分子遗传学原理,认识单核苷酸的聚合物这一化学信息是有帮助的。首先让我们回忆一下 DNA 中的 4 种碱基:G、A、C、T;其次计算 10 个核苷酸所能排列出多少种寡核苷酸序列。由于在寡核苷酸序列中每一个位点可能是 4 种碱基中的一种,那么就可能有  $4^{10}$  或 1 046 576 条不同的排列的序列。在人类细胞中约有  $3 \times 10^9$  以上的核苷,不难想像染色体 DNA 的储存空间有多大,储存的信息量有多大。这使得几乎每一个个体的 DNA 序列都不会相同 (除了同卵双生子)。

## 一、DNA 合成

细胞分裂依赖于 DNA 的复制从而使染色体复制。关于 DNA 复制的两个主要的概念是:①两条互补的反向平行的 DNA 单链之间由众多的氢键连接 (G-C 碱基配对, A-T 碱基配对) 形成稳定的 DNA 双链分子;②DNA 合成的起始需要引物提供 3' 羟基, DNA 聚合酶按照  $5' \rightarrow 3'$  方向合成 DNA。

## 二、RNA 合成

RNA 合成也称为 DNA 的**转录** (transcription),此过程需要 RNA 聚合酶的催化,即以单链 DNA 为模板合成一条互补的 RNA 序列,将遗传信息从 DNA 传递到 RNA。**RNA 聚合酶** (RNA polymerase) 从  $5' \rightarrow 3'$  方向进行 RNA 合成,此与 DNA 聚合酶相同。RNA 合成和 DNA 合成二者之间的主要差别是:①转录时只有一条 DNA 链为模板,而复制时两条链都作为模板;②DNA-RNA 杂合双链不稳定, RNA 合成后释放,而 DNA 复制叉形成后一直打开,新链和母链结合在一起,形成子链;③RNA 合成不需引物,而 DNA 复制需引物;④转录的底物是 rNTP,复制的底物是 dNTP;⑤所用的聚合酶系不同。

关于 RNA 的合成应注意:①RNA 的合成产生基因的转录物,它含有该基因的遗传信息。RNA 的合成起始于基因上游 (upstream) 5' 端的一个特异位点,即转录起始点,而终止于下游 (downstream) 3' 端,在原核生物中为**终止子** (terminator),而在真核细胞中,通常分离到的“初始”RNA 其 3' 端已经加上多聚腺苷,因此不清楚是否有终止子的存在。②一个 DNA 模板可以合成多个拷贝的 RNA 分子, RNA 分子的半衰期短,极易被 RNA 酶所降解。单位时间内 RNA 转录物合成的数量是取决于 RNA 聚合酶转录起始的速率。

## 三、蛋白质合成

蛋白质的合成涉及细胞中的翻译装置——**核糖体** (ribosome),核糖体是由大量的 rRNA 和核糖体蛋白质组成的。

在蛋白质合成中有三个主要的概念：①DNA 基因序列忠实地转录成 mRNA，其三联密码子含有蛋白质合成的信息，称为**遗传密码** (genetic code)。一共有 64 个密码子，其中有 61 个有义密码子编码 20 种氨基酸，3 个无义密码子作为翻译的终止密码子。61 个有义密码子中有一个为翻译的起始密码子 (AUG)，它编码甲硫氨酸。②核糖体结合到 mRNA 的 5' 端，开始蛋白质的合成。第一个氨基酸通常为甲硫氨酸。mRNA 的 5' 端的密码子对应于合成的蛋白质的 N 端。核糖体以 5'→3' 的方向阅读 mRNA 直到终止密码子前结束，此是末端氨基酸的羧基端，在终止密码子处，没有相应的氨酰-tRNA 进入核糖体的 A 位点，核糖体暂停下来，**释放因子 (RF)** 识别终止密码子并与之结合，激活肽基转移酶，水解 P 位点上多肽与 tRNA 之间的肽键，然后释放多肽和 tRNA。③在任何 DNA 序列中，理论上能合成一个蛋白质或一个多肽的连续的密码子系列称为**可读框** (open reading frame)。

## 第二节 核酸的生物化学

要了解基因工程的基本原理就要掌握溶液中 DNA 和 RNA 的两个重要的特点：①影响到核酸分子结构稳定性的各种力；②决定互补异源双链变性和复性的动力学参数。由于在基因工程中要应用各种分子杂交技术，它们都是以复性和变性动力学为理论基础的。

### 一、核酸大分子的结构

在图 1-2 中表明一个 DNA-RNA 杂种双链核酸分子的化学结构。其关键的特点是：①同一条链中相邻的核苷之间以磷酸二酯键连接。②DNA-RNA 异源双链核酸分子的反向平行极性。5'→3' 的 DNA 链与 3'→5' 的 RNA 链进行碱基配对。③通过 T-A、G-C 和 A-U 之间形成氢键，DNA 链和 RNA 链进行互补连接。在 DNA 中存在 G、C、A、T 4 种碱基，在 RNA 中存在 G、C、A、U 4 种碱基。在嘌呤 (G 和 A) 和相对的嘧啶 (C、T 和 U) 之间形成氢键。在 G-C 间形成 3 个氢键，在 A-T 和 A-U 之间形成 2 个氢键。不同的氢键对双链 DNA 分子的热稳定性的贡献是不同的。维持 DNA 双螺旋的三维结构的主要作用力是反向平行链之间的氢键和相邻碱基之间的疏水作用以及**范德华力** (Van der Waals force)。虽然双链 DNA 的分子模型显示其是刚性的，但实际上 DNA 和蛋白质结合时是十分柔软的，可以弯曲成 100° 以上。DNA 链在原位的解离和重新聚合的能力可作为评价其功能的分子资料。DNA 的复制、重组和转录都需要双链的解开，这一过程在体内是由一些特殊的蛋白质——**解旋酶** (helicase)、单链结合蛋白等来完成的。在体外，提纯的 DNA 的解链是可以通过提高温度或者加入变性剂来实现的。

### 二、DNA 双螺旋的变性与复性

**变性** (denaturation) 是指在一定条件下 (高温或变性剂等) DNA 双螺旋解链的过程。对 260nm 紫外光的吸收率——**光密度值** (optical density, OD) 可用来检测溶液中 DNA 单链或双链的浓度。和单链 DNA 相比，在双链中由于作为发色基团的碱基藏在

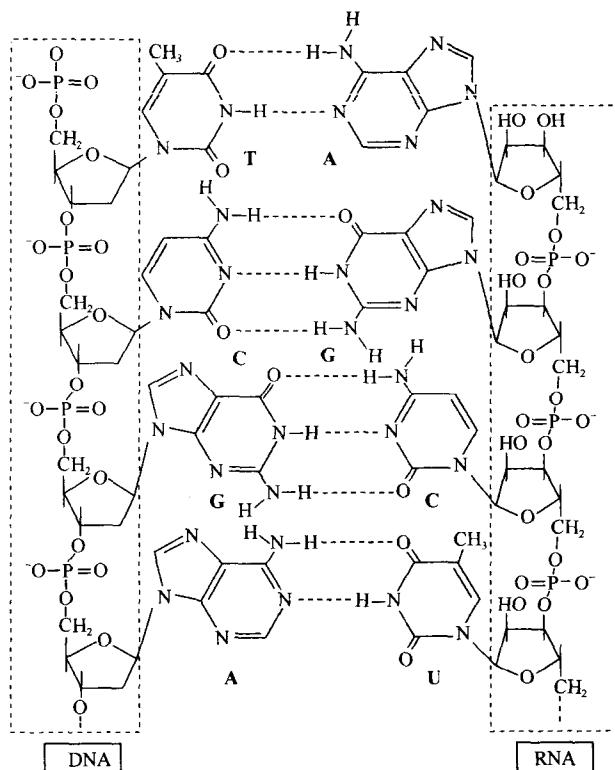


图 1-2 DNA-RNA 杂种双链核酸分子的化学结构。它们以反向平行，互补结合形成双链

DNA 的内部，不易受到紫外光的照射，故其  $OD$  值 ( $A_{260} = 1.00$ ) 低于单链 DNA 的  $OD$  值 ( $A_{260} = 1.37$ )。因此在  $55 \sim 90^\circ\text{C}$  的范围内，测定  $OD$  值的差异可以衡量温度对 DNA 结构的影响。图 1-3 表示双链 DNA 的解链曲线，从中可以看出变性 DNA 的量在一个很窄的温度范围内，这就意味着 DNA 变性是一个协同反应的过程。协同性 (cooperativity) 是指杂合双链 DNA 一旦在原位解链，它仅需要很少的外在能量来分开两条互补链。在 DNA 变性进行到一半时的温度称为解链温度或熔点温度 (melting temperature)，以  $T_m$  来表示。双链 DNA 的  $T_m$  受到 3 种因素的影响：①碱基的成分；②双链的长短；③溶液中离子的浓度。碱基的成分对  $T_m$  的影响很大，这是由于 G-C 碱基对之间比 A-T 碱基对之间多一个氢键，因此 G-C 含量高的双螺旋与 A-T 丰富的双螺旋相比，前者的  $T_m$  较高，双链也比较稳定。杂种双链的长度也影响着  $T_m$  值，这是因为双链分子的稳定性直接和其碱基对的数目呈正比。当双链分子的长度少于 150bp 时，这个现象特别明显。在基因工程中，这两种因素是短片段寡核苷酸进行分子杂交和不完全互补的双链形成时的重要反应条件，常用于同源而并不完全相同的 DNA 分子的杂交反应中。例如，不同物种中的两个同源基因或基因家族的不同成员之间的分子杂交。在异源双链中每 1% 的错配会使  $T_m$  减低  $1^\circ\text{C}$ 。第三个影响  $T_m$  的因素是离子浓度。在高浓度  $\text{Na}^+$  (1mol/L) 存在的条件下，由于正电荷中和了 DNA 磷酸骨架上的负电荷，减少了双链间的互斥作用，使此  $T_m$  升高。相反，在低浓度  $\text{Na}^+$  (0.1mol/L) 时  $T_m$  减低。当互补的短寡核苷酸链 (10~20bp) 在  $\text{Na}^+$  为 1mol/L 时，用下面的经验公式能计算出

$T_m$  的近似值。

$$T_m(^\circ\text{C}) = 2(A+T) + 4(G+C)$$

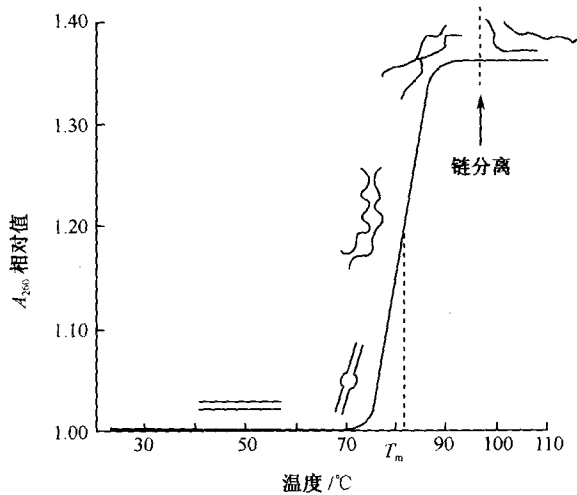


图 1-3 表示双链 DNA 的解链曲线。DNA 在一定温度范围内的 OD 值可以反映双链的变性状况。 $T_m$  是 DNA 变性进行到一半时的温度。在  $T_m$  以下双链占的比例大，而在  $T_m$  以上单链占的比例大。随着温度的增加单链的比例越来越大，最终双链完全解链，变成单链（引自 Freifelder, 1983）

DNA 的复性 (renaturation) 又称为退火 (annealing)，是指 DNA 分子由单链状态恢复到双链结构的过程。复性与变性情况相反，复性过程受到温度、离子浓度、两条互补链的摩尔浓度及反应时间的影响最大。另外两个影响变性的因素包括：①变性剂，如甲酰胺和尿素，能降低  $T_m$ ；②在杂交反应中加入的右旋糖苷硫酸盐 (dextran sulfate)，它可以使反应速率增加 10 倍。在最佳的温度 (比  $T_m$  低 10~15°C) 和最佳离子浓度 (0.2mol/L  $\text{Na}^+$ ) 条件下使核酸浓度在分子杂交反应中成为限速步骤 (rate-limiting step)。 $C_0t$  是指反应初始的浓度和反应时间的乘积。这个术语是用来描述溶液中两条 DNA 单链分子杂交动力学的。当溶液中两种互补链的浓度高时，完成分子杂交反应所需的时间比其中一条单链或两条单链的浓度低时要短。

$C_0t$  曲线以复性的双链 DNA 浓度为纵坐标，以  $C_0t_{1/2}$  的对数为横坐标绘制出 DNA 复性动力学曲线 (图 1-4)；基因组小的生物其 DNA 序列的复杂性低，在复性反应中其  $C_0t$  值比 DNA 序列复杂性高的生物要低。 $C_0t$  曲线表明  $C_0t$  值对复性动力学的影响以及  $C_0t$  值和序列的复杂程度的相关性。例如，相同浓度的 *E. coli* DNA 片段的  $C_0t$  值要比  $\lambda$  噬菌体的  $C_0t$  值大 1000 倍以上。分析人类而获得的  $C_0t$  曲线是一个混合的曲线 (图 1-4)。这是由于人类具有丰富的重复序列 (repetitive sequence)，这些重复序列的拷贝数多，即浓度高，复性反应所需的时间就短，所以  $C_0t$  值比较低；人类基因组中的单一序列 (unique sequence)，拷贝数很少，浓度很低，且比细菌的单一序列要长得多，所以复性反应所需的时间较长，即  $C_0t$  值较高。

表 1-1 概括了各种影响 DNA 的复性和变性的因素。



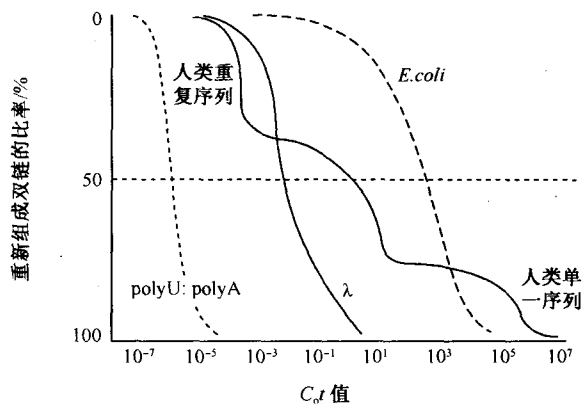


图 1-4  $C_0t$  曲线表示 DNA 序列的复杂性和杂交动力学之间的相关性。在这个曲线中，polyU: polyA 的序列复杂性低，其  $C_0t$  值也比较低。人类的基因组含有重复序列和单一序列，重复序列的  $C_0t$  值低，而单一序列的  $C_0t$  值高（引自 Miesfeld, 1999）

表 1-1 概括各种影响 DNA 的复性和变性的因素

参数	对 $T_m$ 的影响	对复性速率的影响
碱基的成分	$T_m$ 随 G-C 含量的增加而增加	无
序列的长度	$T_m$ 随序列长度的增加而增加, 当 >500bp 时 $T_m$ 值不受影响	随序列长度的增加而上升
离子浓度	$T_m$ 随 $Na^+$ 浓度的增加而增加	最理想的 $Na^+$ 浓度为 1.5mol/L
碱基错配比例	最理想的值随碱基错配的比例(%)的增加而下降	随碱基错配比例的上升而下降
DNA 浓度	不影响	$T_m$ 随 $Na^+$ 浓度的增加而增加
变性剂	$T_m$ 随甲酰胺和尿素浓度的增加而减少	最理想的甲酰胺浓度为 50%
温度	不适用	最理想的温度是低于 $T_m$ 20°C

### 第三节 基因工程中采用的主要酶类

#### 一、序列特异的 DNA 限制性内切核酸酶

基因工程中主要采用的酶类很多（表 1-2），现着重介绍其中的几种酶。

1962 年瑞士的 W. Arber 提出限制与修饰假说来解释宿主对噬菌体生长的限制。根据这个模型，推测细菌的 DNA 中含有特异的核苷酸序列，它可被细胞中的一种**限制性内切核酸酶**（restriction endonuclease，简称限制酶）识别和剪切。这些细菌同时还含有一种 DNA **甲基化酶**（DNA methylase），这种酶可以甲基化上述特异的核苷酸序列（图 1-5）。经过甲基化的化学修饰保护了细菌的 DNA，使它们不会被本身的限制性内切核酸酶所降解。但可以降解外来未经过甲基化的噬菌体 DNA。1968 年 H. O. Smith 等分离了 II 类限制酶，并对其性质进行了阐明。1972 年 U. Kuhnlein 和 W. Arber 分离了大肠杆菌噬菌体的识别位点突变体，支持了限制与修饰学说。“限制”这个术语是描述这些限制性内切核酸酶限制了噬菌体感染的宿主范围。根据细菌的各种类型的限制性系统的作用机制将限制酶分为 3 类，其中仅 II 类限制酶可用于基因工程。这是因为它在体外能识别和剪切 4~8 个碱基的 DNA 特异序列。