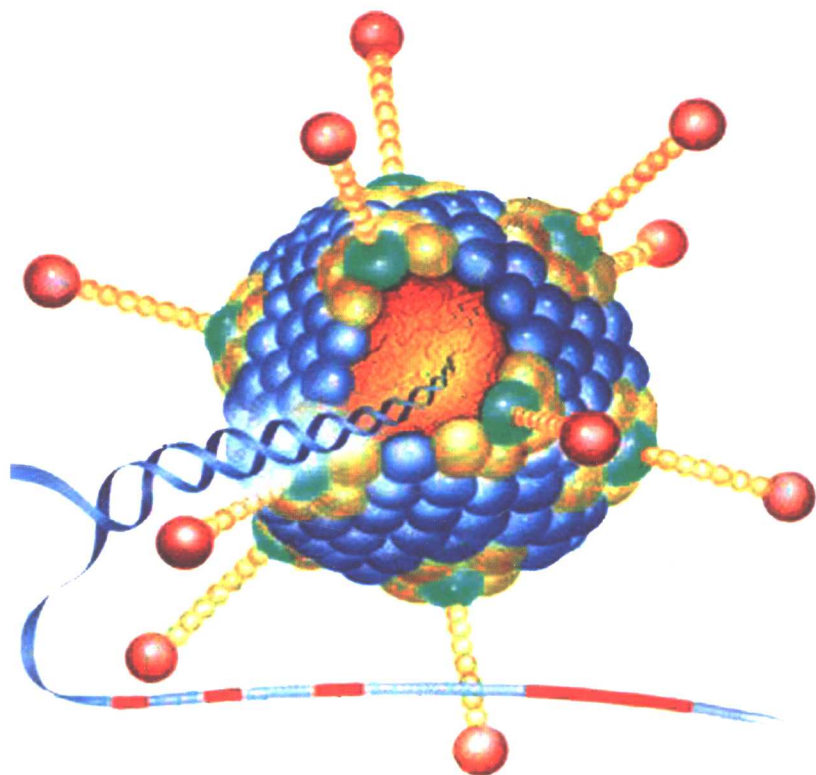


赵晓瑜 李继刚 编著

实用分子生物学技术



Chemical Industry Press



化学工业出版社

实用分子生物学技术

赵晓瑜 李继刚 编著



化学工业出版社

· 北 京 ·

本书以基因重组技术为基础和核心,介绍了基本的分子生物学实验技术,包括基因克隆操作、DNA文库的构建、多聚酶链式反应(PCR)技术、DNA序列的测定、分子标记与杂交等;同时介绍了分子系统学研究中遗传多样性分析技术、真核基因表达调控、蛋白质组学的基本研究方法和有关生物信息的基本知识。

本书注重实用性,特别是针对实验操作中容易忽视的一些问题提出需要注意的事项;同时以一些著名的分子生物学试剂公司产品为实例,介绍它们的工作原理,为开拓读者思路提供参考。

本书适用于生命科学各个专业的研究生阅读,也可供从事相关生物技术工作的教师和科研人员参考。

图书在版编目(CIP)数据

实用分子生物学技术/赵晓瑜,李继刚编著. —北京:
化学工业出版社, 2006.4
ISBN 7-5025-8596-6

I. 实… II. ①赵…②李… III. 分子生物学
IV. Q7

中国版本图书馆 CIP 数据核字 (2006) 第 040783 号

实用分子生物学技术

赵晓瑜 李继刚 编著

责任编辑:赵玉清

文字编辑:袁海燕

责任校对:陶燕华

封面设计:胡艳玮

*

化学工业出版社出版发行

(北京市朝阳区惠新里3号 邮政编码 100029)

购书咨询:(010)64982530

(010)64918013

购书传真:(010)64982630

[http:// www. cip. com. cn](http://www.cip.com.cn)

*

新华书店北京发行所经销

大厂聚鑫印刷有限责任公司印刷

三河市延风装订厂装订

开本 720mm×1000mm 1/16 印张 11½ 字数 241 千字

2006 年 6 月第 1 版 2006 年 6 月北京第 1 次印刷

ISBN 7-5025-8596-6

定 价: 28.00 元

版权所有 违者必究

该书如有缺页、倒页、脱页者,本社发行部负责退换

前 言

分子生物学是一门从 20 世纪 50 年代开始,并于 80 年代迅速发展起来的新型学科。它从分子水平、基因结构和功能以及调控机制上解释生命现象的本质。纵观它的发展史,基因克隆技术的发明和日益成熟,以及随之其他各种分子生物学技术的大量涌现,对于分子生物学的发展起到了巨大的推动作用,同时发展出 21 世纪的支柱产业——生物技术产业。

作者多年从事分子生物学教学和科研工作,积累了一些成功经验和更多的失败教训,深感掌握实验技术和提高实验技能对从事科学研究人员的重要性。分子生物学实验技术的飞速发展有力地推动了整个生命学科的发展,因此分子生物学实验技术也已成为生命科学各个研究领域的基础,也是从事生命科学研究的硕士、博士研究生的必修课。作为 21 世纪的支柱产业,生物技术需要一些具有开拓性精神的技术人才;同时,随着基因组学和蛋白质组学研究的不断深入,生物技术正在以日新月异的速度发展,要求从事生物技术工作的科研人员必须及时跟踪和掌握技术的最新进展,才能适应 21 世纪生命学科发展的需要。因此,如何培养既有扎实的理论基础、又有灵活和宽阔思路的科研能力,是我们教师所肩负的重要责任。希望本书能为此目的提供微薄之力。本书是以河北大学生命学院硕士研究生和博士研究生的“分子生物学实验技术”、“动物分子生物学”讲义为蓝本所撰写的一部教学参考书。

本书内容以基因重组为主线,结合基因组学、蛋白质组学,以及分子系统学的基本研究方法,介绍了相关的基本原理和一些具体实验方案,特别强调了实验操作中的一些容易忽视的问题。尽管目前很多有实力的公司开发了一系列的试剂盒,给具体操作的研究者带来了方便,但随之而来的负面影响往往是出现问题无从下手和分析,尤其对初学者来说,造成过分依赖于试剂公司的局面,这对于专门从事研究的科研人员来说,不是一件好事。从另一角度来说,能够清楚地了解所运用方法和技术的原理,将会有助于操作者更加正确地使用试剂盒。实际上,分子生物学实验是一个经验性很强的工作,因此积累和吸取实验经验是非常必要的。本书还引用了一些比较典型的研究论文作为参考,同时介绍一些著名的分子生物学试剂公司开发新产品的特殊构思,为开拓读者思路提供具体实例,并附以必要的常用数据,以增强该书的实用性。为了帮助读者理解,在每个章节均列出一些实例或文献,其中文献以中文为主(均可以从中国期刊全文数据库中获得),便于读者理解和查寻。如想更深入了解,可以根据这些论文中引用的参考文献进一步搜寻原始资料。

本书共 12 章,内容以基因重组技术为基础和核心,介绍了基本的分子生物学实验技术,包括基因克隆操作、DNA 文库的构建、多聚酶链式反应(PCR)技术、DNA 序列的测定、分子标记与杂交等;同时介绍了分子系统学研究中遗传多样性分析技术、真核基因表达调控和蛋白质组学的基本研究方法,以及生物信息学中最

主要的比对分析技术、常用数据库和软件。

本书适用于生命科学各个专业的研究生阅读，也可供从事相关生物技术工作的教师和科研人员参考。

书中第1~8章由赵晓瑜教授完成，第9~11章由李继刚博士完成，第12章则由上述二人共同完成。全书由静天玉教授帮助校正。

本书在编写过程中，得到河北省生物工程重点学科的大力支持，在此表示由衷地感谢。本书强调实用性，希望能够给读者一些具体的经验、提供设计实验和分析结果的思路，但由于作者研究工作涉及面有限，因此会有许多不足或错误之处，欢迎读者批评指正。

编者

2006年1月

目 录

第一章 绪论	1
第二章 基因克隆的四大要素	3
一、工具酶	3
二、载体	7
三、基因	17
四、受体细胞	19
五、实验过程中菌种的保存和鉴定的重要性	20
参考文献	20
第三章 DNA 文库的构建	21
一、基因组文库的构建	21
二、cDNA 文库的构建	23
三、DNA 和 RNA 纯度鉴定和保存	27
四、反(逆)转录酶	28
五、两种文库的优缺点及用途	29
六、基因文库的筛选及应用	30
参考文献	31
第四章 聚合酶链式反应技术	33
一、PCR 技术的原理	33
二、PCR 体系	34
三、PCR 条件的选择	36
四、PCR 的自动化——PCR 仪	37
五、PCR 引物设计	38
六、耐热 DNA 聚合酶	39
七、提高 PCR 特异性的措施	41
八、防止污染问题	44
九、几种特殊的 PCR	44
十、PCR 技术的应用	48
参考文献	50
第五章 基因重组(克隆)方法	52
一、DNA 连接酶	52
二、克隆方案	54
三、重组 DNA 导入受体细胞	61
四、重组子的筛选和鉴定	64

参考文献	65
第六章 DNA 序列测定	67
一、DNA 序列分析的基本步骤	67
二、Sanger 双脱氧链终止法	67
三、Maxam-Gilbert 化学测序法	71
四、焦磷酸测序技术	72
五、DNA 序列测定技术的发展	73
参考文献	75
第七章 重组产物的表达、分析和纯化	76
一、重组产物的表达系统	76
二、利用大肠杆菌表达系统表达	77
三、真核表达系统的必要性及优势	81
四、表达蛋白的分离和鉴定	81
五、克隆的前沿技术	84
参考文献	88
第八章 示踪和杂交技术	89
一、分子杂交技术的基本原理	89
二、探针的标记物和标记方法	90
三、膜的选择	96
四、杂交技术的应用	97
参考文献	97
第九章 遗传多样性研究中的分子生物学方法	99
一、DNA 指纹图谱	99
二、物理图谱	102
三、限制性片段长度多态性分析	104
四、随机扩增多态性 DNA 分析	105
五、核糖体 rRNA 基因分析	107
六、微卫星标记	108
参考文献	110
第十章 真核基因表达调控的研究方法	112
一、DNase I 超敏感性分析	112
二、凝胶滞留法	113
三、足迹法	114
四、酵母杂交技术	116
五、免疫沉淀和染色质免疫沉淀	119
参考文献	120
第十一章 基因组学与蛋白质组学中的生物技术	122
一、基因芯片技术	122

二、蛋白质双向电泳·····	124
三、质谱技术·····	128
四、基因突变技术·····	130
五、噬菌体展示技术·····	132
六、基因表达系列分析·····	135
参考文献·····	136
第十二章 生物技术中生物信息学的应用 ·····	138
一、生物信息学·····	138
二、生物信息数据库·····	144
三、数据库的搜索、序列比对及分析·····	150
四、生物信息学软件·····	155
参考文献·····	158
附录：常用数据表 ·····	160
一、蛋白质、核酸的换算·····	160
二、常用核酸的长度与分子量·····	160
三、抗生素的储存溶液配制·····	161
四、常用核酸研究的工具酶·····	161
五、常用测序引物·····	165
六、离心机转速与相对离心力的换算·····	165
七、密码子表·····	167
八、常见市售酸碱的浓度·····	168
九、常用两性离子缓冲液的物性·····	168
十、常用克隆宿主菌株的基因型·····	169
十一、常用宿主菌的遗传标记·····	169
十二、生物信息学的相关术语·····	172
参考文献·····	176

第一章 绪 论

Introduction

分子生物学是一门实验性学科，它是以实验结果为主要依据建立起来的理论系统，因此新的实验技术的发明或改进，有力地推动了该门学科的发展。分子生物学的发展史很清楚地证明了这一点。

正像美国康乃尔大学著名华裔分子生物学家吴瑞先生所说：“分子生物学的研究依赖于先进的实验方法，而最好的方法又来源于反复的修改与验证。”因此实验技术和方法不是一成不变的，而是总在不断地改进、简化和优化，从而推动科学研究的深入和发展。因此对于从事分子生物学的科研工作者来说，实践经验和实际操作是极其重要的，特别是分子生物学实验材料用量均为微量，所以手法（手感）很重要，好的手法就是理论加经验的结果。

分子生物学技术是以 DNA 重组（又称基因克隆或基因重组）技术为核心发展起来的，随着 DNA 合成、DNA 序列分析、单克隆抗体技术、微量纯化技术、计算机技术、PCR 技术、杂交技术、生物芯片，以及网络技术的产生和发展，分子生物学技术以惊人的速度覆盖了生命科学领域的方方面面，分子生物学技术不仅发展成为一门独立的学科，同时促进并发展了许多边缘学科，特别是带动了生物技术整体产业的发展。

克隆（clone, clon）一词源于希腊文 klon，原意为树木的枝条。在生物学中其名词含义系指一个细胞或个体以无性繁殖的方式产生一群细胞或一群个体，在不发生突变的情况下，具有完全相同的遗传性状，常称无性繁殖（细胞）系。其动词（clone, cloned, cloning）含义是指在生物体外用重组技术将特定基因插入载体分子中，即基因（分子）克隆技术。

该技术实际上是在分子水平上提供一种纯化和扩增特定 DNA 片段的方法，即将目的基因用体外重组方法将它们插入克隆载体，形成重组克隆载体，通过转化或转导的方式，引入合适的受体（宿主）细胞内得到复制与扩增，然后经过筛选从阳性受体细胞内分离提纯所需的克隆载体，这样就可以得到插入 DNA 的许多拷贝，从而获得目的基因的扩增。如果构建的载体具有高表达元件，则克隆的目的基因同时可以通过受体细胞直接进行表达，而获得大量目的蛋白。简言之，基因克隆（即分子克隆，molecular cloning）是一种 DNA 体外重组技术，通过将一段目的 DNA 经切割、连接插入适当载体，并导入受体细胞，扩增形成大量子代分子的过程。

基因克隆（DNA 重组）技术的过程如图 1.1 所示。

正是在基因克隆技术的基础上，人类才能按照自己的

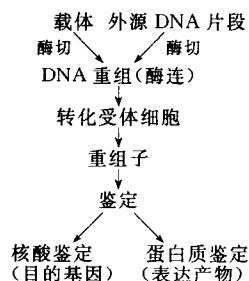


图 1.1 基因克隆流程

意愿，将不同来源的基因进行重组，并获得大量所需要的表达产物；对一些基因进行突变或改造，以便更多地满足人类的需要，或治疗人类的许多疑难病和癌症等疾病；甚至可以制造人类所需的活生生的器官、个体——克隆植物、动物。因此基因克隆技术发展至此就不仅仅是一项技术，而是一个具有广阔应用前景的基因工程，或基因工程产业。

事实上，经过近二十多年的发展，尤其 PCR 技术、杂交瘤（细胞融合）技术的发明，特别是人类基因组计划的实施，大量模式生物基因组测序工作的完成，从分子生物学派生出基因组学、蛋白质组学、生物信息学等，各种新的实验技术也同时应运而生，并在精度、细度、自动化程度和高通量上不断地改进和发展。也正是由于这些新技术的出现，进一步推动了新型学科以及相关产业的迅猛发展。

第二章 基因克隆的四大要素

Four Elements for Gene Cloning

将外源基因通过体外重组后导入受体细胞，使该基因能在受体细胞内复制、转录、翻译和表达，整个操作称为基因重组技术。要实施该技术必须具备四大要素：工具酶、载体、基因和受体（宿主）细胞。

一、工具酶

基因工程的基本技术是人工进行基因的剪切、拼接、组合。基因是一段具有一定功能的 DNA 分子，要把不同基因的 DNA 线形分子片段准确地切出来，需要各种限制性核酸内切酶（restriction endonuclease）；要把不同片段连接起来，需要 DNA 连接酶（DNA ligase）；要合成基因或其中的一个片段，需要 DNA 聚合酶（DNA polymerase）等。因此，酶是 DNA 重组技术中必不可少的工具，基因工程中所用的酶统称为工具酶。

工具酶就其用途而言可分为三大类：限制性内切酶、连接酶和修饰酶，其中限制性内切酶为一大类酶（达上千种）。基因重组正是利用了这些工具酶对 DNA 分子进行一系列的酶催化反应，才得以在体外实现 DNA 分子的切割和连接。因此，工具酶的发现为基因操作提供了十分重要的技术基础。首先重点介绍限制性内切酶（restriction endonucleases 也称 restriction enzyme），其他酶在相关内容中再一一介绍。

从分子生物学发展历史看，核酸限制性内切酶的发现和应用对该学科发展所起的作用是难以估量的。使外源基因在大肠杆菌中克隆的实验是在 1973 年完成的，Stanley Cohen, Herbert Boyer 正是利用了限制性内切酶这一分子手术刀才得以实现。

核酸限制性内切酶是原核生物中的一类能识别双链 DNA 中特定碱基顺序的核酸水解酶。原核生物的限制和修饰系统犹如高等动物的免疫系统，依靠一对识别相同序列的核酸限制性内切酶和甲基化酶活性来对抗外来 DNA 的入侵：当自身的基因组在复制完成而下轮 DNA 复制尚未开始前就被甲基化酶修饰（使某特定序列甲基化），避免了被对应的限制性内切酶的识别和水解，而入侵的噬菌体由于未来得及修饰而被破坏，从而保护细菌不受噬菌体的感染。各种细菌都能合成一种或几种顺序专一的核酸内切酶。这些酶的功能就是通过特异性序列的识别后进行 DNA 的切割，来限制外源性 DNA 侵入自身的细胞内，所以称这种核酸内切酶为限制酶。

根据酶的识别切割序列的特性、催化条件以及是否具有修饰酶的活性而分成三类：I 类、II 类、III 类。

第Ⅰ类限制性内切酶是双功能酶，具有修饰活性（甲基化）和内切酶活性，作用时需要消耗 ATP，能识别专一的核苷酸顺序，并在距离识别点大约 1000 个核苷酸对处切割 DNA 分子中的双链，但是切割的核苷酸顺序没有专一性，是随机的。

第Ⅱ类限制性内切酶只具有核酸内切酶活性，能识别专一的具有回文结构的核苷酸顺序，并在该顺序内的固定位置上切割双链，作用时不需要水解 ATP 提供能量。

第Ⅲ类限制性内切酶也同时具有修饰活性和内切酶活性，具有专一的识别顺序，但不是对称的回文顺序。它在识别顺序旁边 24~26 个核苷酸对的固定位置上切割双链，但这几个核苷酸对是任意的。

其中第Ⅱ类酶在基因重组中最有应用价值。因此以下内容均以第Ⅱ类限制性内切酶为主。

1. 限制性内切酶的命名和书写

限制性内切酶主要是从原核生物中提取的。现在通用的命名原则是：第一个字母是细菌属名的第一个字母，第二、三个字母是细菌种名的前两个字母，这些字母都用斜体字书写；如果同一生物种内又分为不同的血清型和菌株，其菌株名称的第一个字母用正体字书写，并放在限制酶名称的第三个字母后面。比如限制酶 *Hinc*Ⅱ和 *Hind*Ⅲ是分别来自流感嗜血菌（*Haemophilus influenzae*）的 c 和 d 血清型菌株。如果同一菌株中有几种不同的内切酶时，则分别用罗马数字Ⅰ、Ⅱ、Ⅲ……来代表，如表 2.1 所示。

表 2.1 几种限制性内切酶的来源及名称

细菌属名	细菌种名	菌株名称	限制酶名称和书写格式
<i>Bacillus</i>	<i>amyloliquefaciens</i>	H	<i>Bam</i> HⅠ
<i>Escherichia</i>	<i>coli</i>	RY13	<i>Eco</i> RⅠ
<i>Haemophilus</i>	<i>influezae</i>	Rd	<i>Hind</i> Ⅲ

2. 限制性内切酶的特点

(1) 识别特定的核苷酸序列 长度一般为 4~6bp 并具有回文结构的顺序 (palindromes，一段自我互补的 DNA 顺序，即上下链从 5'→3'方向所读的顺序完全一样)。

例如：

EcoRⅠGAATTC

CTTAAG

pstⅠCTGCAG

GACGTC

SmaⅠCCCGGG

GGGCCC

(2) 具有特定的酶切位点 即在识别序列的特定位点对双链 DNA 进行切割，由此产生特定的酶切末端。通常双链 DNA 被酶切后可出现三种形式的末端：5'突出的黏末端、3'突出的黏末端、平末端。

(3) 由两种酶分子组成的二元系统 一种是限制性内切酶, 另一种是甲基化酶, 二者识别位点相同, 但后者不是切割, 而是对识别序列中的一个碱基进行甲基化修饰, 该甲基基团伸入到双螺旋的大沟中去, 阻碍了限制性内切酶的作用, 使之不受对应的限制性内切酶的切割。对于原核生物来说, 甲基化酶对其自身 DNA 序列进行甲基化, 使其免受限制性酶的作用, 因此甲基化酶是细菌体内的一种保护机制。换言之, 正是由于限制性内切酶与甲基化酶, 组成了原核生物的一个完整的免疫系统。

3. 应用限制性内切酶注意事项

酶单位 (U) 的定义: 在 $50\mu\text{L}$ 反应液中, 37°C 保温 1h, 使 $1\mu\text{g}$ 的特定 DNA 完全分解所需的酶量为 1 个活性单位。由于酶的活性与其所处的反应条件有很大的关系, 因此在使用限制性内切酶需注意以下几点。

① 反应条件: 每种酶所需的最适条件各不相同, 包括温度、不同的离子浓度、pH、还原剂等, 因此为了保证酶处于最佳反应条件, 每种酶必须备有配套的缓冲系统 (buffer)。

② DNA 的纯度是影响反应效率的主要因素, 杂质包括蛋白质、酚、氯仿、乙醇、EDTA、SDS 以及高浓度的盐离子。降低污染的办法有适当增加酶的用量、扩大反应体积 (通过稀释降低污染物的浓度) 或延长反应时间。

③ 注意甘油的浓度 (酶储液), 所提供的缓冲液均为 10 倍浓缩液, 目的是保证稀释的酶保存液中甘油的含量不会超过 5%。但较小的反应体积更容易受到移液器误差的影响, 因此酶切反应体系不宜在体积过小 (如小于 $20\mu\text{L}$) 范围中进行。

④ 注意反应液的充分混合, 但不可振荡。反应液充分混合是为了保证反应完全, 推荐用取液器反复吸取混合, 或是用手指轻弹管壁混合, 然后再快速离心即可。

上述几点只是一些基本原则, 实际操作中需要综合考虑。一些大的试剂公司均会提供各种限制性内切酶的详细资料。

特别强调的是, 并不是酶量越多越好, 反应时间越长越好, 因为限制性内切酶都有非特异性 (Star) 活性, 即发生非特异性的反应 (所谓 Star activity, 是在一些特定的条件下, 酶对底物 DNA 的特异性可能降低, 以致在识别序列以外的位点进行切割。在甘油含量高、酶量大, 有机溶剂以及盐浓度低时容易发生)。同时酶在保温过程中, 活性也会发生变化。

为了能够正确地利用限制性内切酶, 应该注意阅读产品说明书以及相关的介绍。例如, 各种限制性内切酶在保温过程中的活性变化表; 双酶切反应时的通用缓冲液使用表; 不同缓冲液中各种限制性内切酶的活性变化表等等。另外需要注意的是, 不同公司产品的酶和缓冲液不要混用。(见附录四中 1, 常用限制性内切酶表)

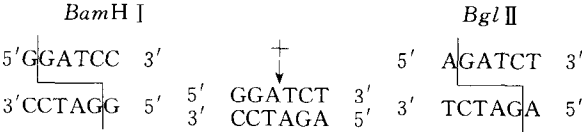
在多数情况下, 同一种限制性内切酶所产生的 DNA 双链末端结构总是相同的, 因而用同一种限制性内切酶酶切同一个或两个不同来源的 DNA 分子所产生的末端都可以相互配对, 在 DNA 连接酶的作用下, $3'$ 和 $5'$ 末端重新形成磷酸二酯

键，从而成为一个重组的 DNA 分子。

由于这些限制性内切酶的识别序列都是对称的，因而两个 DNA 片段可以从两个不同的方向连接起来。这种外源 DNA 片段的插入没有一定的方向，这对插入片段的方向并不显得重要时，如在构建基因文库时适用。但在研究基因表达时，考虑外源 DNA 的插入方向则是十分重要的，需要采取不同的方式进行控制，以达到定向连接的目的。

虽然不同种的限制性内切酶产生的 DNA 双链末端结构在大多数情况下是不相同的，但有些不同种限制性内切酶产生的末端却是相同的，即具有相同类型的突出末端（都是 5' 端或 3' 端突出），突出的核苷酸数目相等且序列也相同，因而可以互相连接起来。这种由不同种的限制酶产生的能相互连接的末端常称之为相容性末端，这些酶则称之为同尾酶（isocaudamer）。

例如，限制酶 *Bam*H I 和 *Bgl* II 都能识别各自的 6 核苷酸序列，且切点都在同一位置，依次为 GGATCC 和 AGATCT，因此产生的 DNA 片段都产生一个相同的单键 5' 端突出（突出序列是 GATC）。当这些酶作用 DNA 分子后，它们都能相互连接，但是新形成的 DNA 分子将同时失去这两种酶的识别序列。如：



通常，不同的限制酶具有不同的识别序列，但有些不同来源的酶能识别相同的序列，这些酶被称为同裂酶（isoschizomer），不过其中有些酶具有相同的切点，而有些酶的切点并不相同，如 *Acc* III 与 *Bsp*E I、*Bse*A I、*Bsi*M I、*Bsp* 131、*Kpn* 21、*Mro* I 识别序列都是 TCCGGA，它们的切点都在 T 和 C 之间，即 T/CCGGA；但 *Bbe* I 与 *Kas* I、*Nar* I、*Sfo* I 识别序列都是 GGCGCC，但它们的切点分别为 GGCGC/C、G/GCGCC、GG/CGCC、GGC/GCC。

现在已从各种微生物中发现三千余种限制酶，它们识别序列的长度最短的是 4 个核苷酸，最长的为 8 个核苷酸。基因克隆过程中使用频率最高的是识别 6 个核苷酸的限制酶。这是因为由 4 个核苷酸组成的识别序列在 DNA 分子中出现的频率很高，如果按完全随机分布的原则，每 $4^4 = 256$ 个核苷酸可出现一个相同的 4 核苷酸识别序列。如果是由 6 个核苷酸序列组成的限制酶识别位点，那么就应该有 $4^6 = 4096$ bp 才可能出现一次。识别 8 个核苷酸的限制酶识别位点就应该有 $4^8 = 65536$ bp 才重复一次。因此，在一个 DNA 分子中识别 4 个核苷酸的限制酶位点太多，识别 8 核苷酸的限制酶位点又太少。换句话说，识别 4 核苷酸的限制酶将 DNA 切得太短，识别 8 个核苷酸的限制酶将 DNA 切得太长，而识别 6 个核苷酸的限制酶则比较适中，切下的 DNA 片段平均长度为 4.1kb 左右。除此之外，片段过长或过短，从技术角度讲，操作起来也不太方便。

实际上，任何一种生物基因组的 DNA 分子中的核苷酸分布并不是完全随机的。也就是不同物种的基因组 DNA 中所含的限制性内切酶位点的种类和数量各不

相同。也正由于此，可以通过绘制限制性内切酶图谱来描述某一个物种的基因组特征，即物理图谱。

比如，在SV40 (*Simian vacuolating virus*，猿猴空泡病毒)的DNA分子中(基因组长为5243bp)仅有一个Hpa II位点(识别序列为C↓CGG)，而Tha I位点(识别序列为CG↓CG)则根本没有。但是，在大小相似的原核生物DNA分子中，如最常见的大肠杆菌质粒载体pBR322(总长4363bp)则有26个Hpa II位点和23个Tha I位点。在一个含有75% AT和25% GC的DNA分子中，长达262144bp中只有一个Sma I位点(识别序列CCC↓GGG)。一种裸藻(*Euglena gracilis*)的叶绿体DNA分子(130000bp)中因没有Sma I位点，而无法用Sma I将其切为两段。

因此使用DNA限制性内切酶的过程中，需要根据不同的研究目的来选择。比如在建立基因组文库时，DNA片段要求尽可能长一些以便使这些片段含有完整的基因，这就需要选择一些限制酶，它们的识别序列在该生物DNA分子中应当是一些不常见的核苷酸序列；而在绘制限制酶图谱时，选择的限制酶则应该识别较为常见的识别序列。表2.2列出一些限制酶消化不同物种的基因组DNA获得的平均DNA片段的长度。

表 2.2 一些限制酶消化不同物种的基因组DNA^①获得的平均DNA片段的长度/bp

限制性酶	识别序列	大肠杆菌 (<i>E. coli</i>)	酿酒酵母 (<i>S. cerevisiae</i>)	秀丽隐杆线虫 (<i>C. elegans</i>)	黑腹果蝇 (<i>D. melanogaster</i>)	人 (<i>H. sapiens</i>)
Avr II	CCTAGG	290000	18000	17000	25000	5000
BamH I	GGATCC	9000	7000	9000	7000	7000
BssH II	GCGCGC	2000	27000	14000	12000	33000
Eag I	CGGCCG	16000	27000	17000	11000	32000
Sal I	GTCGAC	9000	11000	10000	8000	83000
Spe I	ACTAGT	59000	5000	8000	10000	8000
Xba I	TCTAGA	120000	4000	4000	8000	4000

① 基因组DNA的总长度：*E. coli*，4.6Mb；*S. cerevisiae*，12.1Mb；*C. elegans*，100Mb；*D. melanogaster*，3.8Mb；*H. sapiens*，22Mb

二、载体

一种能够结合外源DNA片段形成重组DNA，并能进行自我复制的DNA分子称为载体(vectors)。

最主要的载体是质粒、噬菌体和病毒，分别应用于原核细胞或真核细胞。从功能来分，有克隆载体(构建基因组文库或cDNA文库)、测序载体、表达载体以及能在两个不同物种中存在的穿梭质粒。随着生物技术和科学研究的深入，不断出现具有不同功能的新载体，如可以容纳几百万碱基的酵母人工染色体。载体的具体种类有以下几种。

质粒(plasmids)	10kb 以下
噬菌体(phages)	λ 噬菌体(λphages); 12~20kb M13 噬菌体: 单链 DNA 柯斯质粒(cosmids): 兼有质粒和噬菌体的行为 phagemids: 单链, 同样兼有质粒和噬菌体的行为
病毒(virus)	Ti 质粒(tumor inducing plasmid) DNA 病毒 RNA(反转录)病毒 昆虫杆状病毒(baculovirus)
人工染色体(artificial chromosome)	YAC(酵母人工染色体) BAC(细菌人工染色体) PAC(P1 衍生人工染色体) MAC(哺乳动物人工染色体)

这些载体均有各自的特点, 表 2.3 中列出它们的主要特性。

表 2.3 各种克隆载体特性的比较

载 体	受体细胞	结 构	插入大小/kb
λ 噬菌体	<i>E. coli</i>	线形载体	~24
柯斯质粒	<i>E. coli</i>	环形质粒	35~45
BAC	<i>E. coli</i>	环形质粒	~300
PAC	<i>E. coli</i>	环形质粒	100~300
YAC	酵母细胞	线形染色体	100~2000
MAC	哺乳类细胞	线形染色体	>1000

1. 质粒

质粒(plasmids)是细菌内的共生型遗传因子。作为染色体外小型双链环状 DNA 复制子, 它能在细菌中垂直遗传并且赋予宿主(受体菌)某些代谢活动和抗药表型, 是比病毒更简单的原始生命。质粒由 Lederburg 发现并于 1952 年正式命名。由于质粒不仅可以携带外源基因进入细菌中进行扩增和表达, 而且在添加真核复制信号和启动子后, 构建出能在原核或真核细胞中均可复制的穿梭质粒, 使外源基因在真核细胞中得以表达, 因而成为基因的运载工具(即载体), 在基因工程中具有极广泛的应用价值。质粒并不是宿主染色体的永久组成部分, 而是一种自主复制的遗传单元。不同质粒 DNA 的复制, 其机理不尽相同, 根据质粒拷贝数的不同, 分为严紧型和松弛型, 前者与宿主的繁殖结合, 致使每个细胞仅有一个或几个拷贝质粒, 而后者却由质粒自己编码的基因合成功能蛋白所调节, 因此在宿主染色体复制停止的情况下, 质粒可以继续扩增, 因而具有 10~200 个拷贝数。对于克隆载体, 松弛型质粒更为适用。

天然质粒不足以用为载体, 必须根据基因克隆的要求进行体外修饰改造。因此

可以用作载体的质粒需具备一定的基本元件后才有相应的功能。

(1) 质粒（载体）具备的基本元件

- ① 复制起始点 (ori)：使质粒可以自我复制和扩增。
- ② 抗生素抗性基因：一般有两个，以便为受体菌提供易于检测的表型性状的选择记号，而且在外源基因插入后形成的重组质粒中，至少仍保留一个强选择记号。常用的抗性有抗氨苄青霉素基因 *Amp^r*、抗四环素基因 *Tet^r*、抗卡那霉素基因 *Kan^r*、抗氯霉素基因 *Cml^r*、抗链霉素基因 *Str^r*。
- ③ 若干限制酶单一识别位点（多克隆位点，polylinker）：满足各种基因克隆的需要，而且外源基因插入后不影响质粒的复制功能。
- ④ 较小分子量和较高拷贝数：便于进行分子操作和质粒扩增。

典型质粒 pBR322 的物理图谱见图 2.1，pUC18/19 的物理图谱及多克隆位点见图 2.2。

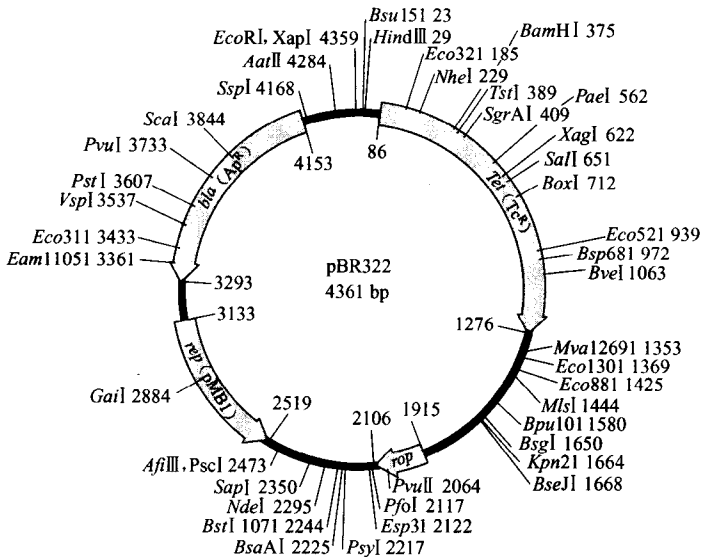


图 2.1 pBR322 质粒物理图谱

(2) 质粒的功能

① 运载外源 DNA 片段：质粒通常用于运载小于 15kb 的外源基因片段到宿主中。为了便于操作，人们又构建了穿梭质粒 (shuttle plasmids)。所谓穿梭质粒是指一类人工构建的具有两种不同复制起点和选择标记，因而可以在两种不同类群宿主中存活和复制的质粒载体^[1]。此概念不仅用于不同的微生物菌群之间，也可以推广到真核生物表达载体的构建，如用于枯草杆菌的 pBE2、酵母的 pPIC9K、哺乳动物表达载体 pMT2 和用于植物细胞的 Ti 质粒。这些穿梭质粒不仅可以在大肠杆菌中复制扩增，也可以在相应的枯草杆菌、酵母、动物或植物细胞中扩增和表达。由于大肠杆菌的培养、基因重组、扩增、测序等工作易于进行，技术也相当成