



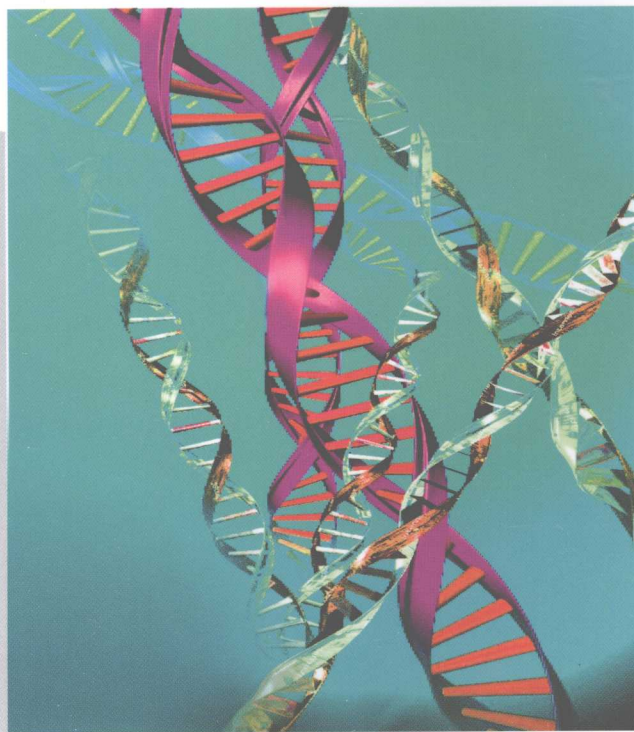
高等院校生命科学专业基础课教材

基因工程及其分子生物学基础

基因工程分册

(第2版)

静国忠 编著



北京大学出版社
PEKING UNIVERSITY PRESS



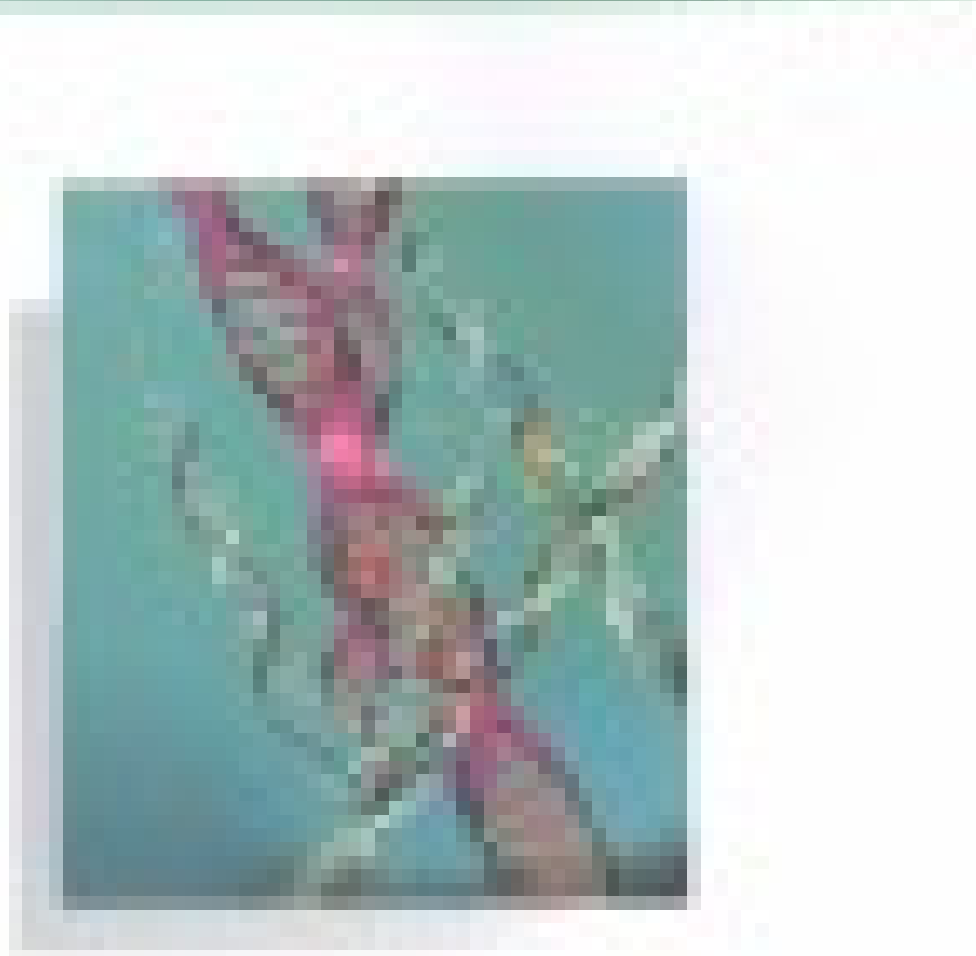
中国科学院知识工程出版工程

中国科学院知识工程出版工程

基因工程分册

王德成 主编

科学出版社



ISBN 7-03-010000-0

高等院校生命科学专业基础课教材

基因工程及其分子生物学基础

——基因工程分册

(第2版)

静国忠 编著



北京大学出版社
PEKING UNIVERSITY PRESS

图书在版编目(CIP)数据

基因工程及其分子生物学基础: 基因工程分册/静国忠编著. —2版. —北京: 北京大学出版社, 2009. 7

(高等院校生命科学专业基础课教材)

ISBN 978-7-301-15546-2

I. 基… II. 静… III. ①基因—遗传工程—高等学校—教材②分子生物学—高等学校—教材 IV. Q7

中国版本图书馆 CIP 数据核字(2009)第 121160 号

书 名: 基因工程及其分子生物学基础——基因工程分册(第 2 版)

著作责任者: 静国忠 编著

责任编辑: 黄 炜

封面设计: 张 虹

标准书号: ISBN 978-7-301-15546-2/Q·0120

出版发行: 北京大学出版社

地 址: 北京市海淀区成府路 205 号 100871

网 址: <http://www.pup.cn> 电子信箱: zpup@pup.pku.edu.cn

电 话: 邮购部 62752015 发行部 62750672 编辑部 62752038 出版部 62754962

印 刷 者: 北京大学印刷厂

经 销 者: 新华书店

787 毫米×1092 毫米 16 开本 16.75 印张 410 千字

1999 年 8 月第 1 版

2009 年 7 月第 2 版 2009 年 7 月第 1 次印刷

定 价: 28.00 元

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容

版权所有,侵权必究

举报电话: (010)62752024 电子信箱: fd@pup.pku.edu.cn

目 录

1 基因工程的四大要素及实施要点	(1)
1.1 基因工程操作中常用的工具酶	(1)
1.2 基因的分离	(6)
1.3 基因工程载体	(14)
1.4 受体细胞和重组基因的导入	(30)
1.5 基因重组的方法	(33)
1.6 基因重组体的筛选	(37)
2 外源基因在宿主细胞中的高效表达	(41)
2.1 有效的转录起始与基因的高效表达	(41)
2.2 mRNA 的有效延伸和转录终止与基因的高效表达	(42)
2.3 mRNA 的稳定性与基因的高效表达	(43)
2.4 有效的翻译起始与基因的高效表达	(43)
2.5 遗传密码应用的偏倚性与基因的高效表达	(45)
2.6 mRNA 的二级结构与基因的高效表达	(50)
2.7 RNA 的加工与基因的高效表达	(50)
2.8 mRNA 序列上终止密码的选择	(50)
2.9 表达质粒(或载体)的拷贝数及稳定性与基因的高效表达	(51)
2.10 外源蛋白的稳定性与基因的高效表达	(51)
3 基因的融合和融合蛋白的表达	(53)
3.1 利用基因融合技术表达外源基因的缘故	(53)
3.2 基因融合的策略	(53)
3.3 基因融合和重组蛋白的产生	(55)
3.4 基因融合和展示筛选	(56)
3.5 基因融合和蛋白分泌	(56)
3.6 融合蛋白的纯化	(57)
3.7 融合蛋白的位点特异性切割	(57)
附录 重组蛋白表达和纯化中常用的融合标签	(58)
4 外源基因的分泌表达	(63)
4.1 外源基因在 <i>E. coli</i> 细胞中的分泌表达	(63)
4.2 外源基因在枯草杆菌中的分泌表达	(65)
4.3 α -因子前导序列介导的酵母细胞分泌系统	(66)
4.4 外源基因在哺乳动物细胞中的分泌表达	(69)
5 重组蛋白的正确折叠及修饰	(71)
5.1 重组蛋白的可溶性表达和折叠	(71)
5.2 重组蛋白的重折叠	(74)

5.3	外源蛋白在翻译后的修饰	(77)
6	几种真核细胞表达系统	(78)
6.1	哺乳动物细胞表达系统	(78)
6.2	外源基因在哺乳动物细胞中的组成性表达和诱导性表达	(86)
6.3	核型多角体病毒为载体的昆虫(细胞)表达系统	(87)
6.4	转基因动物及其应用	(95)
6.5	转基因植物及其应用	(97)
6.6	DNA 疫苗	(98)
7	分子杂交技术	(100)
7.1	探针与目标核酸相互作用的原理	(100)
7.2	探针的选择和特异性	(102)
7.3	杂交的速率与探针长度、浓度及杂交加速剂的关系	(103)
7.4	杂交的最适条件	(105)
7.5	放射性探针的制备	(106)
7.6	非放射性探针的制备	(112)
7.7	放射性和非放射性标记探针的应用范围	(115)
7.8	DNA 微阵列——基因组芯片	(116)
8	粒子轰击和基因转移	(119)
8.1	粒子轰击技术简介	(119)
8.2	粒子轰击技术的应用范围	(119)
9	聚合酶链反应及其应用	(122)
9.1	聚合酶链反应的原理	(122)
9.2	标准的 PCR 扩增方案	(122)
9.3	PCR 引物及其设计	(123)
9.4	关于热稳定的 DNA 聚合酶	(133)
9.5	PCR 对模板质量的要求	(134)
9.6	PCR 反应缓冲液和循环(周期)数	(135)
9.7	PCR 相关技术的原理及其应用	(136)
10	各种生物学展示技术	(152)
10.1	噬菌体展示技术	(152)
10.2	细菌展示技术	(157)
10.3	酵母展示技术	(159)
10.4	核糖体展示技术	(160)
10.5	mRNA 展示技术	(162)
11	基因打靶技术及其应用	(166)
11.1	同源重组与基因打靶	(166)
11.2	组织特异性的基因打靶	(171)
11.3	转座子介导的基因打靶	(174)
11.4	RNA 干涉与基因敲除	(178)

11.5	反义核酸技术	(180)
11.6	细菌基因敲除方法	(180)
11.7	基因打靶技术的应用	(181)
12	DNA 序列分析	(183)
12.1	Sanger 的双脱氧链终止测序法	(183)
12.2	Maxam-Gilbert DNA 化学降解法	(186)
12.3	关于变性聚丙烯酰胺测序凝胶	(187)
12.4	全基因组 DNA 序列的分析	(187)
13	基因突变	(190)
13.1	寡核苷酸介导的基因突变	(190)
13.2	盒式突变法	(200)
13.3	利用 PCR 进行 DNA 序列的突变	(201)
13.4	关于随机突变	(205)
14	寡核苷酸的化学合成	(210)
14.1	寡核苷酸片段固相合成的原理及步骤	(210)
14.2	寡核苷酸片段的纯化及鉴定	(214)
15	蛋白质相互作用及其分析方法	(218)
15.1	蛋白质相互作用的重要性	(218)
15.2	蛋白质相互作用所产生的结果	(218)
15.3	蛋白质相互作用的类型	(219)
15.4	酵母双杂交系统	(219)
15.5	利用绿色荧光蛋白(GFP)片段重组装研究蛋白质 相互作用——细菌双杂交法	(222)
15.6	蛋白质相互作用分析的 <i>in vitro</i> 方法	(224)
16	蛋白质-核酸相互作用	(237)
16.1	研究蛋白质-核酸相互作用的重要性	(237)
16.2	蛋白质-核酸相互作用的类型	(237)
16.3	蛋白质-核酸相互作用分析的 <i>in vitro</i> 方法	(237)
17	蛋白质工程概述	(244)
17.1	蛋白质分子的结构分析	(244)
17.2	蛋白质的结构预测与分子设计	(245)
17.3	基因工程是实现蛋白质工程的关键技术	(246)
	参考文献	(248)



1 基因工程的四大要素及实施要点

将外源基因(编码蛋白质和 RNA 的基因)通过体外重组后导入受体细胞内,使这个基因能在受体细胞内复制、转录、翻译表达的操作叫做基因工程。一个完整的基因工程包括基因的分、重组、转移,基因在受体细胞中的保持、转录、翻译表达等全过程。基因工程又称重组 DNA 技术,其实施至少要有四个必要的条件:工具酶、基因、载体、受体细胞。本章侧重介绍这些基因工程的基本要素,对具体操作请查阅有关工具书(Sambrook J, et al, 1989; Wu R, et al, 1989; Goeddel DV, 1991; Ausubel FM, et al, 1995)。

1.1 基因工程操作中常用的工具酶

基因工程工具酶就其用途可分为三大类,即限制性内切酶、连接酶及修饰酶。

1.1.1 限制性内切酶、甲基化酶和限制修饰系统

限制性内切酶属于 DNA 内切酶,其通过识别 DNA 分子中的特定的识别位点,即一段短的核苷酸序列,也称限制位点(restriction site),对双链 DNA 进行切割。在真细菌和古菌中发现的限制性内切酶被认为是通过进化而产生的抗病毒(噬菌体)入侵的一种防御机制(Arber W, Linn S, 1969; Krieger DH, Bickle TA, 1983)。在细菌宿主细胞内存在着一个称为“限制”(restriction)的过程,在此过程中限制性内切酶选择性地切割外来的 DNA (foreign DNA),而宿主 DNA 由于被一个称为甲基化酶(methylase)的修饰酶甲基化,使其免遭限制性内切酶的切割。这两个过程一起形成了所谓的限制修饰系统(restriction modification system)(Kobayashi I, 2001)。限制性内切酶通过水解每条链中糖-磷酸骨架间的磷酸二酯键对双螺旋 DNA 进行切割。

基于限制性内切酶的亚基组成,酶活性对辅助因子的需求,识别序列(靶序列)的性质以及它们的 DNA 切割位点与识别序列之间的相对位置等特点,将限制性内切酶大致分为三大类,分别称为 I 型、II 型和 III 型限制性内切酶。

简言之, I 型限制性内切酶(EC:3.1.21.3)的切割位点远离其识别位点,其间至少相隔 1000 个碱基对。识别位点的碱基序列不对称,一端含有 3~4 个碱基,而另一端含 4~5 个碱基,它们之间由 6~8 个碱基组成的间隔区相隔。其活性需要 ATP(水解)、S-腺苷甲硫氨酸(AdoMet)和 Mg^{2+} 。I 型限制性内切酶由 HsdR、HsdM 和 HsdS 等三个亚基组成。HsdR 为限制酶活性所必需;HsdM 具甲基化酶活性;HsdS 除具甲基转移酶活性外,其对切点识别的特异性也是重要的。因此, I 型限制性内切酶是多功能的蛋白质,既有限制性内切酶活性又具有甲基化酶活性。

II 型限制性内切酶(EC:3.1.21.4)在如下几方面与 I 型酶不同: II 型酶由一个亚基组成,

它们的识别位点通常是连续的、由4~8个碱基对组成的回文结构(呈二重对称),其识别和切割DNA的位点相同,即识别DNA分子上的特定序列并在特定的位点进行切割,其活性只需要 Mg^{2+} ,并不需要ATP或S-腺苷甲硫氨酸。需要指出的是,在20世纪90年代至21世纪初所发现的某些II型限制性内切酶并不完全符合上述II型酶的标准,分别作为II型酶的亚型,用Type II B、II E、II G表示,此地不赘述。II型酶无甲基化酶活性。

III型限制性内切酶(EC:3.1.21.5),其识别位点与切割位点不同,但二者之间相距不远;其活性需要ATP(但并不水解ATP),S-腺苷甲硫氨酸能促进其酶解反应(但并不必需);其以与甲基化酶(EC:2.1.1.72)形成复合体的形式存在。

由于I型和III型限制性内切酶很难形成用于DNA重组的稳定的特异性切割末端,故I、III型酶在基因工程中基本不用。II型限制性内切酶是基因工程中的主要工具酶,其有如下特点:

(1) 识别特定的碱基(核苷酸)序列。

(2) 识别序列和切割DNA的位点相同,即II型酶在其识别序列的特定位点对双链DNA进行切割,由此产生出特定的酶切末端。双链DNA被酶切后可出现三种形式的末端:

① 5'突出黏性末端,如限制性内切酶在二重对称轴的5'侧切割双链DNA的每条链时,双链DNA交错断开,产生带5'突出的黏性末端片段。如EcoR I识别序列为

$$\begin{array}{c} 5'-GAATTC-3' \\ 3'-CTTAAG-5' \end{array}$$

双链DNA被切割后产生

$$\begin{array}{c} 5'-G \qquad \qquad \qquad 5'AATTC-3' \\ 3'-CTTAA5' \qquad \qquad \qquad G-5' \end{array}$$

的5'突出黏性末端。

② 3'突出黏性末端,如限制性内切酶在二重对称轴的3'侧切割双链DNA的每条链时,双链DNA交错断开,产生3'突出的黏性末端片段。如Pst I识别序列为

$$\begin{array}{c} 5'-CTGCAG-3' \\ 3'-GACGTC-5' \end{array}$$

DNA被切割后产生

$$\begin{array}{c} 5'-CTGCA3' \qquad \qquad \qquad G-3' \\ 3'-G \qquad \qquad \qquad 3'ACGTC-5' \end{array}$$

的3'突出黏性末端。

③ 平端,如限制性内切酶在二重对称轴上同时切割DNA的两条链,则产生带平端的片段。例如Sma I识别序列为

$$\begin{array}{c} 5'-CCC\downarrow GGG-3' \\ 3'-GGG\uparrow CCC-5' \end{array}$$

经此酶切过的双链DNA产生

$$\begin{array}{c} 5'-CCC \quad GGG-3' \\ 3'-GGG \quad CCC-5' \end{array}$$

带平端的DNA片段。

(3) II类限制-修饰系统是由两种酶分子组成的二元系统。一种为限制性内切酶,另一种为独立的甲基化酶,其修饰与限制性内切酶识别位点相重叠的序列。

限制性内切酶在基因工程中的主要用处是通过切割DNA分子,对含有特定基因的片段进行分离、分析。现在几乎在基因工程中所用的所有限制性内切酶都已商品化,注意查阅各公司的样本,就可以找到各种酶反应条件。在使用限制性内切酶时要注意如下问题:

① 仔细查阅酶的识别序列和切割位点。不同来源但识别同样顺序和相同切割位点的限制性内切酶叫做同裂酶(isoschizomer)。有时两种限制性内切酶识别不同的核苷酸序列,但切割后DNA分子所产生的黏性末端却相同,如BamH I、Mbo I、Bcl I、Sau3A、Xho I、Bgl II就属于这类酶,它们的识别序列虽不同,但都产生5'-GATC-3'的5'黏性末端。这类酶为具不同酶切位点的DNA片段的重组提供了方便。但要注意,往往相容的黏性末端在用DNA连接酶连接后,都不能再被相应的内切酶切割。

② 注意酶反应条件,相同反应条件的酶可以同时酶解DNA样品,特别是在需要快速知道酶切结果时,就显得更有用。

③ 注意说明书中所列出的注释的内容,这里会向你提供有关限制性内切酶识别序列中位

点特异性甲基化的信息以及引起酶产生星活性(star activity)的原因。对于有些限制性内切酶,当反应体系中甘油的质量分数>12%、酶:DNA>25 U/μg 或缺少 NaCl 和存在 Mn²⁺ 等条件下都可能出现星活性。所谓限制性内切酶的星活性,是指当酶的反应条件改变时,其在识别序列以外的位点进行切割。

④ 注意酶的浓度,酶切时不是酶加得越多越好。一般以在 20 μL 反应体积中,37°C 条件下每小时水解 1 μg DNA 的酶量定义为一个酶单位。根据 DNA 的用量选择适当的酶量不但可以避免不必要的浪费,也可以保证酶解的质量。

⑤ 注意酶在保温过程中活性的变化,从而确定合适的酶解时间(图 1-1)。

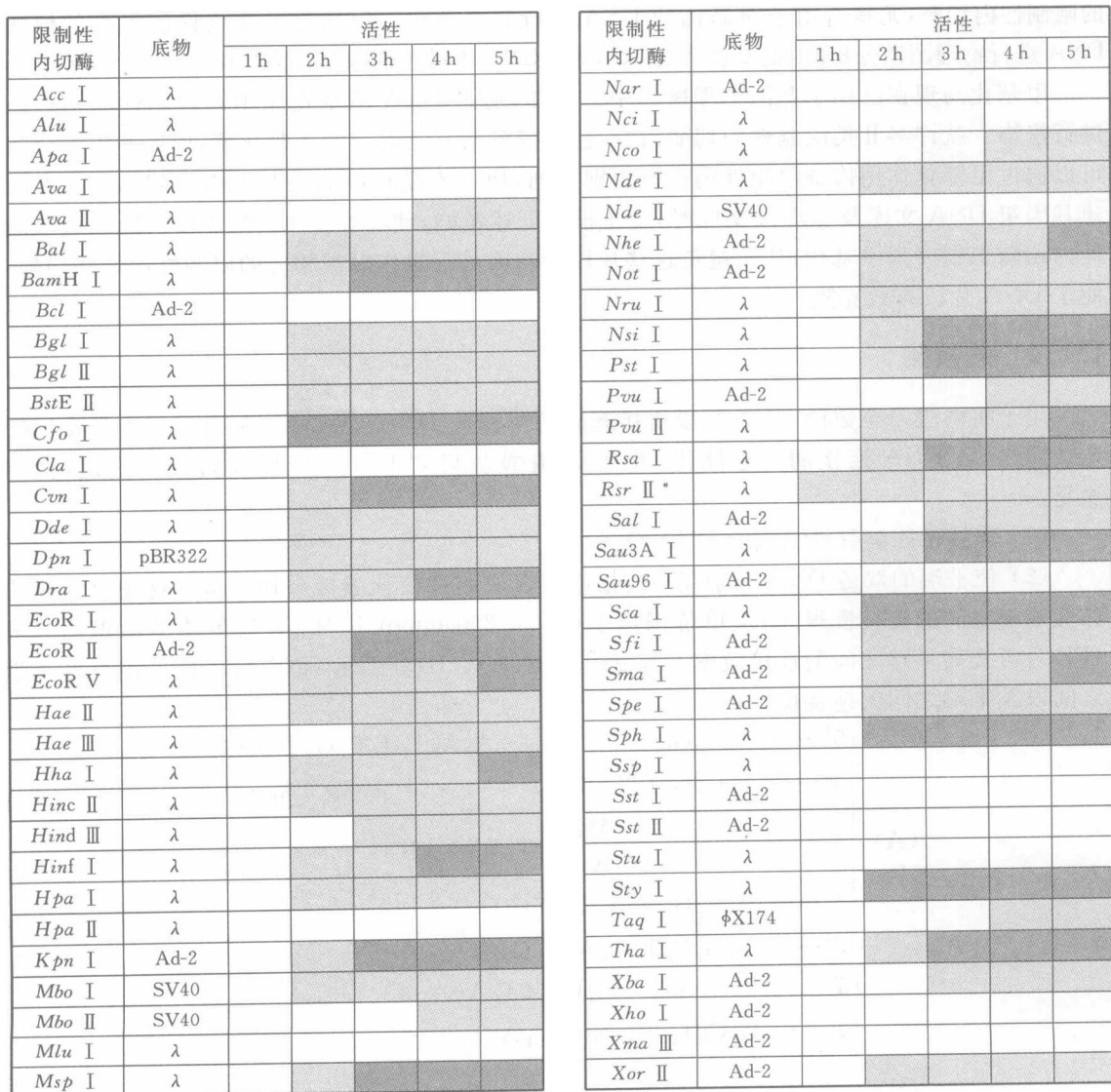


图 1-1 各种限制性内切酶在保温过程中的活性变化

无活性, 部分活性, 完全活性

* 1 U 的 Rsr II 在 18 h 中酶解 1 μg λDNA。

(引自 GIBCO BRL, Catalogue)

E. coli 绝大多数菌株含有两种 DNA 甲基化酶,即 *dam* 和 *dcm* 甲基化酶。*dam* 可在 5'-GATC-3' 序列中的腺嘌呤 N⁶ 位置上引入甲基,而 *dcm* 则在 5'-CCAGG-3' 或 5'-CCTGG-3' 中的胞嘧啶 C⁵ 位置上引入甲基。很多较高等的真核生物含有在特定位点 C_FG 和 C_PN_FG 的 DNA 甲基化酶,可使上述序列的胞嘧啶 C⁵ 位上甲基化。在哺乳类的基因组中,甲基化主要发生在 CG 序列,而植物基因组中甲基化可以发生在 CG 和 CNG 序列。在选用限制性内切酶时要注意:

① 所选用的限制酶对甲基化的敏感性。

② 根据 DNA 的不同来源选择不同的限制性内切酶。对某些位点特异性甲基化作用不敏感的限制性内切酶,尤其适用于对修饰的 DNA 进行完全消化。例如对甲基化程度很高的植物 DNA 进行物理图谱分析时,就要选择对^{m5}CG 和^{m5}CNG 不敏感的限制性内切酶。

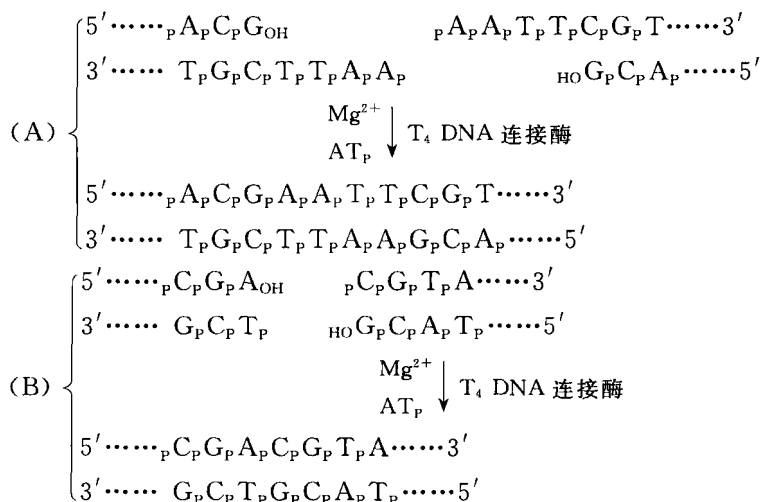
甲基化酶现在已用于基因工程操作中。例如,通过 DNA 甲基化作用对限制性酶切位点进行修饰。就许多 II 类限制性内切酶而言,已分离到与其相对应的甲基化酶,而这些甲基化酶可以通过甲基化作用修饰限制性内切酶识别序列,使之不受相应的限制性内切酶切割。在构建基因组 DNA 文库及 cDNA 文库时,正是根据上述策略,使存在于 DNA 片段内的一些限制性内切酶识别序列甲基化,从而避免这些片段在基因工程操作时被相应的限制性内切酶切碎,便于 DNA 片段有效克隆。

1.1.2 连接酶

用于将两段乃至数段 DNA 片段拼接起来的酶称为 DNA 连接酶。基因工程中最常用的连接酶是 T₄ DNA 连接酶。它催化 DNA 5'-磷酸基与 3'-OH 之间形成磷酸二酯键,其用途是:

(1) 连接带匹配黏性末端的 DNA 分子。

(2) 使平端的双链 DNA 分子互相连接或与合成的寡核苷酸接头相连接。这类反应要比黏性末端之间的连接慢得多,但单价阳离子(150~200 mmol/L NaCl)或低浓度的聚乙二醇(PEG)可提高平端之间的连接效率。其反应如下(A)、(B)分别示 DNA 片段间匹配黏性末端之间以及平端之间的连接反应:



除 T₄ DNA 连接酶外,还有 *E. coli* 的 DNA 连接酶,它所行使的催化反应基本同 T₄ DNA

连接酶相同,只是催化反应需要 NAD^+ 参与。第三种连接酶是 T_4 RNA 连接酶,其催化单链 DNA 或 RNA 的 5'-磷酸与另一单链 DNA 或 RNA 的 3'-OH 之间形成共价连接。

1.1.3 用于基因工程的修饰酶

1. DNA 聚合酶

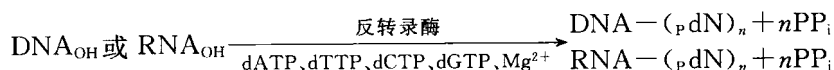
目前常用的 DNA 聚合酶是 *E. coli* DNA 聚合酶 I、*E. coli* DNA 聚合酶 I 大片段 (Klenow fragment)、 T_4 噬菌体 DNA 聚合酶、 T_7 噬菌体 DNA 聚合酶以及耐高温 DNA 聚合酶(如 *Taq* DNA 聚合酶)等。不同来源的 DNA 聚合酶具有各自的酶学特性。*E. coli* DNA 聚合酶 I 具有 5'→3' DNA 聚合酶活性、5'→3' 及 3'→5' 外切核酸酶活性,其大片段是经枯草杆菌蛋白酶裂解完整的 DNA 聚合酶 I,或用基因工程手段去除全酶中 5'→3' 外切酶活性片段而得到的具有 5'→3' DNA 聚合酶活性及 3'→5' 外切酶活性的全酶大片段。 T_4 噬菌体 DNA 聚合酶的酶活性与上述大片段酶活性相似,然而其 3'→5' 的外切核酸酶活力比 Klenow 大片段强近 200 倍。与 Klenow 大片段相比, T_4 噬菌体 DNA 聚合酶的最大特点是不从单链 DNA 模板上置换寡核苷酸引物,因此,在体外突变反应中, T_4 DNA 聚合酶比大片段具更大的效率。 T_7 噬菌体 DNA 聚合酶所催化合成的 DNA 的平均长度要比其他 DNA 聚合酶催化合成的 DNA 的平均长度大得多。耐高温的 DNA 聚合酶(如 *Taq* DNA 聚合酶)由于其最佳作用温度为 75~80℃,目前广泛用于多聚酶链反应(polymerase chain reaction, PCR)及 DNA 测序。

无论哪种 DNA 聚合酶,其催化的反应是:在存在单链 DNA 模板及带 3'-OH 的引物时,其反应结果为:



它们在基因工程中的应用是多方面的。例如, DNA 分子的体外合成、体外突变、DNA 片段探针的标记、DNA 的序列分析、DNA 分子的修复、多聚酶链反应(PCR)等。具体的操作和注意事项请参阅有关工具书。

反转录酶是一种依赖于 RNA 的 DNA 聚合酶。商品化的两种反转录酶分别来源于鼠反转录病毒和禽反转录病毒。反转录酶具有 5'→3' 的 DNA 聚合酶活性,在存在 RNA 或 DNA 模板及带 3'-OH 的 RNA 或 DNA 引物时,其催化如下反应:



反转录酶还具有 RNase H 的活性,即 5'→3' 及 3'→5' 外切核酸酶活性。利用反转录酶所具有的 RNase H 的活性,可使 RNA-DNA 杂交体中的 RNA 被持续地、特异性地降解,从而免去对反转录后的 RNA 模板用 NaOH 水解的步骤。

值得指出的是,目前所用的反转录酶都不具备 3'→5' 的脱氧核糖核酸的外切酶活性,故不具备如 Klenow 大片段那样的校对功能,所以在进行延伸反应时容易出现碱基错误掺入。

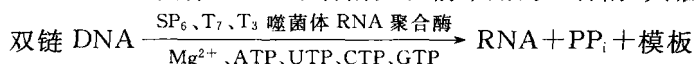
反转录酶在基因工程中的用途主要是以真核 mRNA 为模板合成 cDNA,用以组建 cDNA 文库,进而分离为特定蛋白质编码的基因。将 mRNA 反转录与 PCR 偶联建立起来的反转录 PCR(RT-PCR),使真核基因的分离更加高效。

另一类型的 DNA 聚合酶是末端脱氧核苷酸转移酶。在二价阳离子(Co^{2+})存在下,末端转移酶催化 dNTP 加于 DNA 分子的 3'-OH 端。主要用于给载体或 cDNA 加上互补的同聚

尾,便于外源基因重组;标记 DNA 片段的 3' 端以及制备寡核苷酸的分子长度的标准(Jing GZ, et al, 1986)。此酶不需 DNA 模板。

2. 依赖于 DNA 的 RNA 聚合酶

SP₆ 噬菌体及 T₇ 和 T₃ 噬菌体 RNA 聚合酶是目前常用的三种酶,其催化的反应如下:



这三种酶的共同特点是只识别各自噬菌体的 DNA 序列中特定的启动子区。用这三种噬菌体启动子组建的各种类型的载体已经面世(Promega Corporation, 1996)。这三种酶的主要用途是在离体条件下,合成单链 RNA 作为杂交探针,完成体外翻译系统中功能性 mRNA 或体外剪接反应的底物的合成。

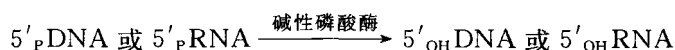
3. T₄ 噬菌体多核苷酸激酶

它催化 ATP 的 γ -磷酸基转移至 DNA 或 RNA 片段的 5' 末端。基因工程中主要用于:

- (1) 标记 DNA 片段的 5' 端,制备杂交探针;
- (2) 基因化学合成中,寡核苷酸片段 5' 磷酸化;
- (3) 测序引物的 5'-磷酸标记。

4. 碱性磷酸酶

市售的产品主要来源于细菌和牛小肠的碱性磷酸酶,其功能是去除 DNA 或 RNA 5' 末端的磷酸反应:



碱性磷酸酶用于:

(1) 去除 DNA 片段 5'-磷酸,以防止在重组中 DNA 分子(如载体 DNA)的自身环化,从而提高重组效率;

(2) 在以 [γ -³²P]-ATP 标记 DNA 或 RNA 的 5' 末端前,去除 DNA 或 RNA 片段的非标记的 5'-磷酸。

除上面介绍的一系列工具酶之外,还有其他一些工具酶在基因工程的操作中也被广泛应用。例如核酸酶 BAL31、S₁ 核酸酶、绿豆核酸酶、核糖核酸酶 A(RNase A)、脱氧核糖核酸酶 I(DNase I)及外切核酸酶 III 等。这些核酸酶在可控的条件下,用于核酸分子的修饰或降解。

基因工程的工具酶为基因的分离、重组、修饰、突变提供了必要的手段,有关它们在应用中的反应条件、注意事项可以参阅各种有关“分子克隆”实验指南及各生物工程公司所提供的样品手册。仔细地阅读这些样品手册将为实验工作的成功创造条件。

1.2 基因的分离

基因的分离是基因工程研究中最主要的要素,目的基因的成功分离是基因工程操作的关键。由于每种基因,特别是单拷贝基因占整个生物基因组很少一部分,且 DNA 的化学结构相似,都是由 A、T、G、C 四种碱基组成,具有极相似的理化性质,这给分离特定的目的基因带来很大困难。尽管如此,人们仍可以通过各种方法有效地分离出所要的基因。

1.2.1 基因分离的物理化学方法

这是基因工程在初始阶段所用的方法,目前已不用。介绍此方法对于了解基因分离手段

的进步是有益的。DNA 分子的两条链存在着 G≡C、A=T(分别代表 GC 间的 3 个氢键、AT 间 2 个氢键)碱基配对。如果 DNA 分子中某段的 G≡C 碱基对含量高,则其热稳定性就高,即其溶解温度(T_m)值就高。这样,人们可以通过控制溶解温度使富 A=T 区解链变性,而富 G≡C 区仍维持双链,再利用单链核酸酶 S_1 去除解开的单链部分,得到富 G≡C 区的 DNA 片段。海胆 rDNA 基因的分离就是一例。海胆 rDNA 分子内 G≡C 含量可以达 63%,通过热变性和 S_1 酶解处理可得到提纯 50 倍的 rDNA,最后经氯化铯平衡梯度离心,得到相对分子质量为 1.9×10^7 的高纯 rDNA。

1.2.2 鸟枪法分离基因

鸟枪法(shot gun)又叫霰弹法。这一方法是绕过特定基因分离这一关口,用生物化学方法,如用限制性内切酶将基因组 DNA 进行切割,得到很多在长度上同一般基因大小相当的 DNA 片段(相对分子质量为 $0.8 \times 10^6 \sim 9 \times 10^6$)。然后,将这些片段混合物随机地重组入适当的载体,转化后在受体菌(如 *E. coli*)中进行扩增,再用适当的筛选方法筛选出所需要的基因。此方法在基因工程发展的初级阶段曾起过很大的作用。用鸟枪法分离基因要求有简便的筛选方法,如利用特定基因缺陷型(如营养缺陷型等)的受体菌,或特定的寡核苷酸或 DNA 片段探针以及特定基因产物的抗体,可通过对表型的筛选或用分子杂交技术、免疫筛选技术检出所需要的基因。

1.2.3 cDNA 文库的建立与基因的分离

cDNA 的克隆对于特定基因的分离和特性研究是极有价值的工具。cDNA 文库(cDNA library)最关键的特征是它只包括在特定的组织或细胞类型中已经被转录成 mRNA 的那些基因序列。这样使得 cDNA 文库的复杂性要比基因组文库(genomic DNA library)要低得多(图 1-2)。由于不同的细胞类型、发育阶段以及细胞所处的特定状态是由特定基因的表达所决定的,因此各自的 mRNA 的种类就不同,由此而产生独特的 cDNA 文库。在建立 cDNA 文库时,如果选择的细胞或组织类型得当,就容易从 cDNA 文库中筛选出所需要的基因序列。

一个 cDNA 文库的组建应包括如下步骤:

1. 分离表达目的基因的组织或细胞

为了最大限度地得到目的基因,避免来源于其他组织或细胞的基因的混杂,要尽量避免其他组织和细胞类型的污染。为了得到全长的 cDNA,所用的材料要尽量新鲜。

2. 从组织和细胞中制备总体 RNA 或 mRNA

从有限量的起始材料去组建 cDNA 文库的主要限制是能否得到足够量的模板 RNA。为增加 RNA 的回收率,保证有足够量的模板 RNA 用于第一条互补链 DNA 的合成,利用硫氰胍或酸性硫氰胍-酚-氯仿等 RNA 微量分离法以及载体 RNA 共沉淀法来制备 RNA。应该注意到所用的载体 RNA 一定要同目的 RNA 有明显区别,不作为模板参与 cDNA 的合成。如果起始物质不受限制,一般用 oligo(dT)纤维素柱亲和层析分离出 poly(A)⁺ RNA 作为 cDNA 合成模板。如果材料有限,应直接用总体 RNA 进行 cDNA 文库的组建。利用 poly(A)⁺ mRNA 可使文库的复杂性降低,大部分 mRNA 序列可以富集 50 倍左右。由于 mRNA 的相对含量增加,这样对于在细胞内丰度较低的 mRNA cDNA 的合成有利。然而,在 poly(A)⁺ mRNA 的纯化过程中,也可能造成某些基因序列的丢失。

基于利用 oligo(dT)亲和柱层析,通过 oligo(dT)与 mRNA 3'端 poly(A)配对可分离纯化 poly(A)⁺ mRNA 的原理,现已开发出更简便、高效地分离纯化 poly(A)⁺ mRNA 的方法:利

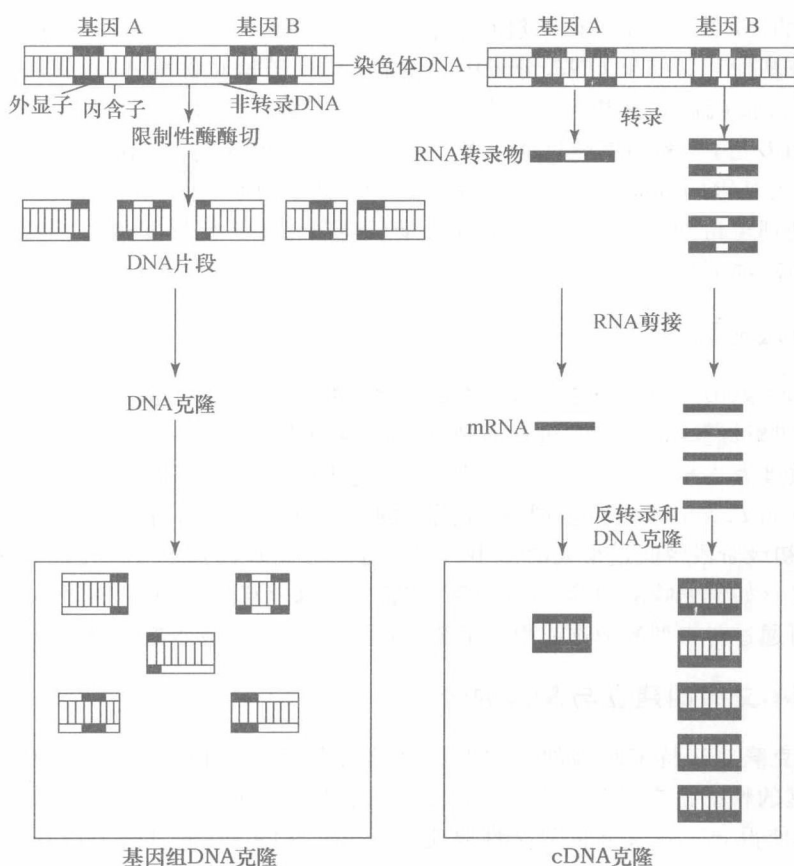


图 1-2 基因组 DNA 文库及 cDNA 文库组建示意图

用生物素化的 oligo(dT) (biotinylated oligo(dT)) 作为亲和介质, 从细胞或组织抽提物中富集 poly(A)⁺ mRNA, 然后, 将所得到的生物素 oligo(dT)-poly(A)⁺ mRNA 混合物与链霉菌亲和素 (抗生物素蛋白) 磁球 (streptavidin MagneSphere) 保温, 通过生物素-抗生物素蛋白之间有效的相互作用, 达到简便、高效分离纯化 poly(A)⁺ mRNA 的目的。此方法的优点是 mRNA 产率高, 其回收量与所用细胞或组织的量成正比, 不用在纯化前纯化总体 RNA, 减少 mRNA 在分离纯化过程中发生降解的机会。此类 mRNA 纯化系统可从各厂家获得。

3. 第一条 cDNA 链的合成

第一条互补 DNA 链 (cDNA) 的合成需要 RNA 模板、cDNA 合成引物、反转录酶、四种脱氧核苷三磷酸以及相应的缓冲液 (Mg²⁺) 等。第一条链的合成可参见“分子生物学基础分册”中图 4-50。如前所述, RNA 模板可以用总体 RNA 或 poly(A)⁺ RNA。引物通常用 oligo(dT) (12~18 个) 或由小牛胸腺 DNA 制备的随机引物 (5~6 个核苷酸长, 商品有售)。利用 oligo(dT) 作为引物的优点是可以最大限度地减少 poly(A)⁻ 模板合成 cDNA 的可能性; 缺点是 cDNA 的合成总是从 mRNA 的 3' 端开始, 在最后的 cDNA 文库中, 代表 mRNA 的 3' 区域的组分所占比例会高, 而且对于一些较长的 mRNA 分子来说, 由于反转录酶在 cDNA 合成过程中易从 mRNA 分子上脱开, 从而造成第一条链合成不完全。利用随机引物的好处是, 合成的 cDNA 文库可能更好地代表最初 RNA 模板所代表的组分; 缺点是, 由于 cDNA 在 RNA 模板上的合成是随机的, 易产生大量的非全长的 RNA 模板转录物, 从而影响到全长 cDNA 分子

的获得率。需指出的是,如果想使 cDNA 文库最大限度地代表 poly(A)⁺mRNA 的序列,在合成 cDNA 的反应中,用总体 RNA 和 oligo(dT)引物是最有效的组合。第一条链的合成质量可用碱性琼脂糖凝胶电泳来检测。

4. 第二条 cDNA 链的合成

目前通用的 cDNA 第二条链的合成的方法有三种:

(1) 自身引导法。即利用第一条链上的 3' 端序列的折回或发夹来自引导法来合成(见“分子生物学基础分册”图 4-50)。虽然过去组建的大多数 cDNA 文库都用此法,但由于其效率低,目前已不多用。

(2) cDNA 第二条链的置换合成法(图 1-3)。此方法的特点是,作为第一条链合成反应产物的 cDNA-mRNA 杂交分子充当切口平移的模板。RNase H 在杂交分子的 mRNA 链上造成切口或缺口,产生一系列 RNA 引物,它们被 *E. coli* 的 DNA 聚合酶 I 用以合成第二条链。

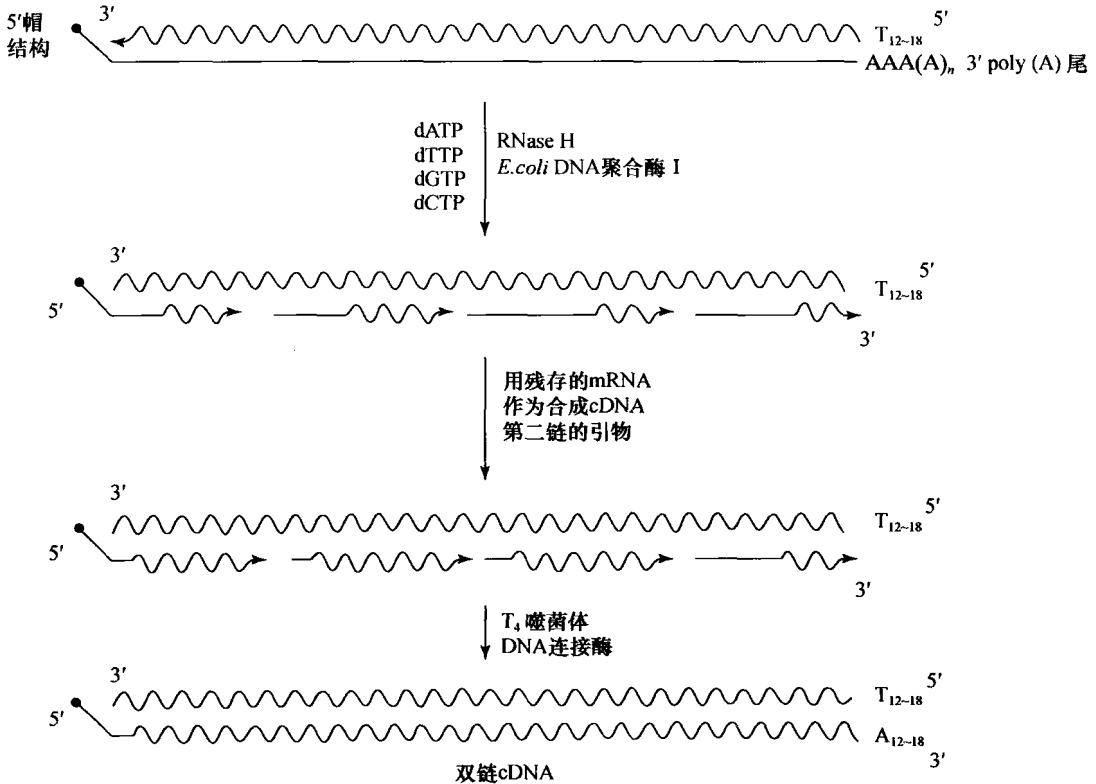


图 1-3 双链 cDNA 的置换合成

(引自 Sambrook J, et al, 1989)

该方法的优点是:效率高;直接利用第一条链反应产物,无需进一步处理和纯化;省去 S₁ 酶处理,改善了 cDNA 的质量。近来 Ray 等人报道了用 RNase H/*E. coli* DNA 聚合酶 I /Klenow 大片段不同配比合成第二条链的方法。此方法使产生全长 cDNA 的概率大大提高(McCarrey JR, et al, 1994)。

(3) 引物-衔接头法合成双链 cDNA。其原理如图 1-4 所示。此方法的特点是通过将 cDNA 两端加上限制性核酸内切酶位点,使其较方便地克隆入相应的载体。cDNA 文库的建

立方法仍在不断地改进,其基本原则是使 cDNA 文库具有最大的代表性。多聚酶链反应(PCR)的出现给 cDNA 文库的组建带来新的活力。

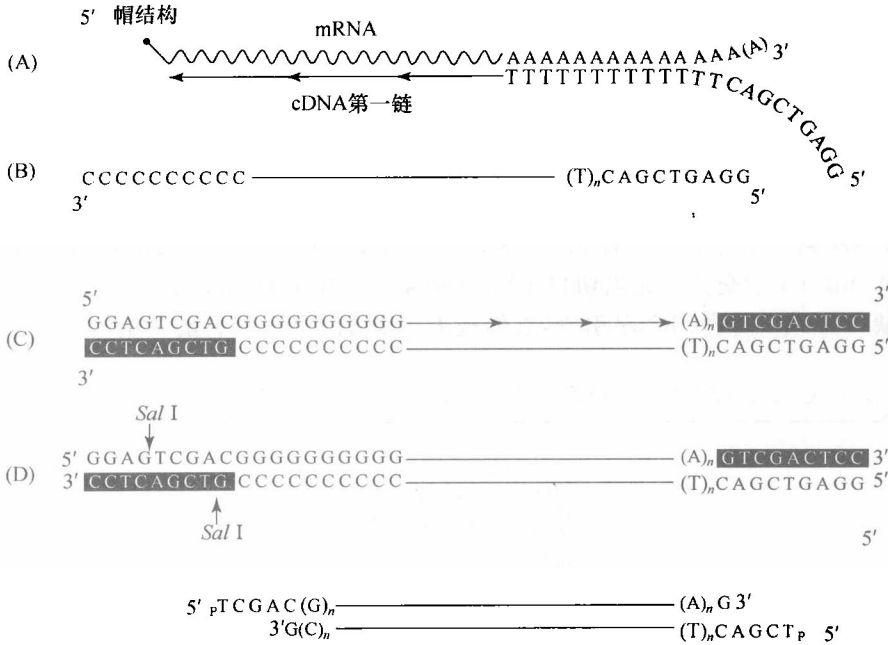


图 1-4 引物-衔接头法合成双链 cDNA

(A) 用单链引物-衔接头引导 cDNA 第一链的合成;(B) 水解 RNA 并用末端转移酶将同聚尾加到 cDNA 第一链的 3'端;(C) 用单链引物-衔接头引导 cDNA 第二链的合成(为尽量使 cDNA 末端的限制酶切位点再生,合成时通常加入与衔接头互补的单链寡核苷酸);(D) 用适当限制酶切割双链 cDNA 并连接到载体上。(引自 Sambrook J, et al, 1989)

5. cDNA 的甲基化和接头的加入

一旦双链 cDNA 被合成,可以用标准法将 cDNA 甲基化,并在其 5'和 3'端加上适当的接头。当用有限材料制备时,甲基化这步可以省略,用预先切好的接头连在 cDNA 片段的两端,这样就可以在不甲基化的条件下进行克隆。接头除了提供相容的末端以便将 cDNA 进行克隆外,也可以作为双链 cDNA PCR 扩增引物位点。

6. 双链 cDNA 同载体的连接

双链 cDNA 同载体的有效连接是 cDNA 文库建立好坏的关键,特别是起始材料很少时更是如此。为保证 cDNA 文库具代表性,要选择尽可能有效的载体。通常用 λgt10/λgt11 及与其相当的载体来组建非表达和表达的 cDNA 文库,而一般不用转化效率较低的质粒作载体。插入的顺序(cDNA)与载体 DNA 的比率按下式计算:

$$\frac{\text{插入顺序}}{\text{载体}} = \frac{\text{插入顺序质量}}{\text{插入顺序碱基对}} \times \frac{\text{载体碱基对}}{\text{载体质量}}$$

式中插入顺序和载体的质量以 ng 计。一般,插入顺序和载体的摩尔比在 1:1 和 3:1 之间为好。

cDNA 文库的建立为我们分离特定的有用基因提供了来源,也为研究特定细胞中基因表达的相对水平开辟了道路。当然,如果分离 cDNA 克隆的目的就是为分离少数几个基因,cDNA 文库数量上的代表性就不必太多考虑,但必须选对所用的细胞类型。cDNA 文库为从