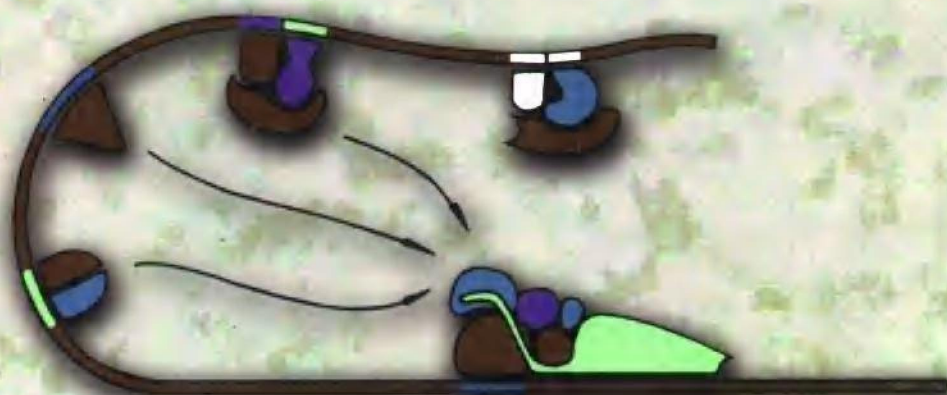


高等学校教学用书

现代生物技术概论

何忠效 静国忠 许佐良 孙万儒

编著



北京师范大学出版社

高等学校教学用书

现代生物技术概论

何忠效 静国忠 编著
许佐良 孙万儒

2025/22

北京师范大学出版社

· 北京 ·

图书在版编目(CIP)数据

现代生物技术概论/何忠效等编著. —北京: 北京师范大学出版社, 1999. 5
ISBN 7-303-04831-6

I. 现… II. 何… III. 生物工程-概论 IV. Q81

中国版本图书馆 CIP 数据核字(98)第 32296 号

北京师范大学出版社出版发行

(北京新街口外大街 19 号 邮政编码:100875)

出版人:常汝吉

北京师范大学印刷厂印刷 全国新华书店经销

开本:787mm×1 092mm 1/16 印张:24 字数:614 千字

1999 年 5 月第 1 版 1999 年 5 月第 1 次印刷

印数:1~3 000 册 定价:32.00 元

前 言

分子生物学、结构生物学等生物学前沿学科的不断深入发展,使现代生物技术作为一门新兴的高技术学科得到迅猛发展,并作为使二十一世纪成为生命科学世纪的主要因素之一,而受到了世界各国的高度重视。我国教育、科研部门也予以极大的关注。许多高等院校都纷纷设立了“生物技术”专业;在科研方面,从 863 计划到 973 计划,国家都以极大的投入将生物高技术列为各类高技术发展之首。

在这样的背景下,国内迫切需要一部从基本理论到技术特点等对现代生物高技术四大分支领域(基因工程和蛋白质工程、酶工程、细胞工程、发酵工程)进行深入介绍的教材。本书作者均为中国科学院研究生院的兼职教授,已在各相关领域为研究生们作了多年课堂讲授。本书的撰写基本上是以我们的讲授提纲为骨架,并结合每人在各自领域多年的科研经验,再参考国内外有关最新进展而完成的,在此同时,也照顾到讲授时的条理性和系统性。由于本书旨在使读者对各主要技术的基本原理及实施要点有深刻的了解,而不在于对诸分支领域做面面俱到的全面描述,故本书以“概论”的形式出现。

本书可供综合大学、师范院校、医学及农林轻工等院校相关专业本科生、研究生作为教材或教学参考书,也可供相关专业的教师及科研人员参考。

由于我们都是尚站在科研、教学第一线上的年近六旬之人,工作任务十分繁重,尽管大家都倾力想使本书成为一部现代生物技术方面的力著,但由于我们专业背景的局限,难免有复盖度不够等问题。同时,由于时间紧迫及我们的水平所限,书中可能会出现这样那样的问题,在此衷心欢迎各位专家学者、广大本科生、研究生对本书提出批评指正,以便再版时,使它更臻完善。

作 者

1998 年 12 月于北京

目 录

第 1 篇 基因工程和蛋白质工程

第 1 章 基因工程的分子生物学基础	(1)
1.1 遗传信息传递的方向——中心法则	(1)
1.2 DNA 的复制	(2)
1.3 DNA 聚合酶	(3)
1.4 原核和真核生物基因结构的特征	(6)
1.5 RNA 的转录和 RNA 的加工	(7)
1.6 逆转录和逆转录酶	(18)
1.7 翻译——蛋白质的生物合成	(18)
1.8 基因表达的调控	(33)
第 2 章 基因工程的四大要素及实施要点	(49)
2.1 基因工程操作中常用的工具酶	(49)
2.2 基因的分离	(53)
2.3 基因工程载体	(60)
2.4 受体细胞和重组基因的导入	(70)
2.5 基因重组的方法	(71)
2.6 基因重组体的筛选	(74)
第 3 章 与基因工程相关的一些理论和技术问题	(79)
3.1 外源基因在宿主细胞中的高效表达	(79)
3.2 基因的融合和融合蛋白的表达	(83)
3.3 外源蛋白的分泌表达	(86)
3.4 关于重组蛋白的正确折叠及修饰	(90)
3.5 几种产生基因突变的方法	(93)
3.6 DNA 序列分析	(98)
3.7 基因的修饰及其对表达的影响	(101)
第 4 章 蛋白质工程概述	(110)
第 5 章 基因工程和蛋白质工程的进展	(113)
5.1 COS 细胞瞬时表达系统及其应用	(113)
5.2 哺乳动物细胞稳定表达系统	(116)
5.3 核型多角体病毒为载体的昆虫表达系统	(119)
5.4 噬菌体显示技术 (phage display)	(124)
5.5 粒子轰击和基因转移	(126)
5.6 同源重组与基因寻靶	(128)

5.7 关于转基因动物	(131)
5.8 聚合酶链式反应及其应用进展	(132)
5.9 几种有开发前景的研究	(135)

第2篇 酶工程

第1章 酶工程概述	(139)
1.1 定义及特点	(139)
1.2 范围	(139)
第2章 酶的来源及酶反应的特点	(140)
2.1 酶的来源及微生物作为酶源的优越性	(140)
2.2 酶的分类及其基本作用原理	(146)
第3章 酶的分离纯化及纯度鉴定技术	(150)
3.1 酶的抽提	(150)
3.2 用于生化物质分离的沉淀技术	(153)
3.3 各种分离纯化技术的基本理论依据	(156)
3.4 凝胶过滤	(156)
3.5 吸附层析	(166)
3.6 离子交换层析	(168)
3.7 高效液相色谱 (HPLC)	(175)
3.8 亲和层析	(180)
3.9 聚丙烯酰胺凝胶电泳 (PAGE)	(188)
3.10 等电聚焦 (IEF)	(194)
3.11 几种其它的聚丙烯酰胺凝胶电泳	(218)
3.12 印迹转移电泳	(220)
3.13 各种分离纯化技术的顺序及衔接	(222)
第4章 固定化技术	(223)
4.1 固定化技术的发展史	(223)
4.2 固定化技术	(223)
4.3 固定化技术的新进展	(225)
4.4 生物传感器为酶工程开拓了新的发展领域	(227)
第5章 多酶反应器	(232)
5.1 意义	(232)
5.2 多酶反应器的设计目标和要求	(232)
5.3 使用多酶反应器时的一些注意事项	(232)
第6章 化学修饰及蛋白质工程等新技术为酶的开发和改造提供的新途径	(234)
6.1 酶的化学修饰	(234)
6.2 基因工程技术为酶的生产提供新途径	(234)
6.3 蛋白质工程技术为酶的定向改造提供新手段	(235)
第7章 酶工程发展的现状和展望	(236)

第3篇 细胞工程

第1章 抗体产生的机理	(239)
1.1 抗体的概念	(239)
1.2 抗体形成的理论	(239)
1.3 免疫球蛋白的分子结构	(241)
1.4 免疫球蛋白的分子片段	(243)
1.5 免疫球蛋白的基因结构	(243)
1.6 抗体多样性的基因基础	(245)
1.7 抗体多样性的细胞学基础	(248)
1.8 抗体制备技术的发展	(249)
第2章 细胞工程抗体	(262)
2.1 小鼠体系	(262)
2.2 融合前的准备工作	(263)
2.3 细胞融合	(268)
2.4 融合后的管理	(280)
第3章 人单克隆抗体的制备	(288)
3.1 制备人单抗的技术路线	(289)
3.2 人源抗体制备中的难点及新技术方法	(297)
第4章 单克隆抗体在肿瘤治疗中的应用	(306)
4.1 单克隆抗体在体外的应用	(306)
4.2 单克隆抗体在体内的应用	(308)
附录：杂交瘤单克隆抗体技术	(311)
(一) 融合前的准备工作.....	(311)
(二) B细胞融合技术	(315)
(三) 融合后的管理工作.....	(316)

第4篇 发酵工程及其进展

第1章 概述	(325)
1.1 历史的回顾	(325)
1.2 生物技术的发展使发酵工业进入了新时期	(326)
第2章 优良菌种的选育	(328)
2.1 菌种分离、筛选的原则与步骤	(328)
2.2 产生特殊物质的微生物筛选	(331)
2.3 自然选育	(334)
2.4 诱变育种	(335)
第3章 代谢调控和代谢工程	(340)
3.1 代谢调控	(340)
3.2 代谢工程	(348)

第 4 章 发酵与产物分离偶联	(354)
4.1 发酵与产物分离偶联的基础	(354)
4.2 偶联技术的应用	(356)
第 5 章 发酵过程的优化与控制	(366)
5.1 发酵过程的测量	(366)
5.2 发酵过程模型化	(369)
5.3 发酵过程控制	(372)

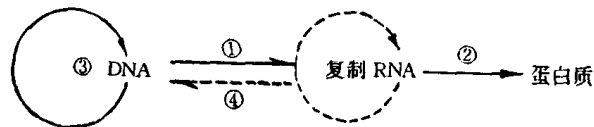
第1篇 基因工程和蛋白质工程

第1章 基因工程的分子生物学基础

自从 Watson 和 Crick 于 1953 年提出 DNA 的双螺旋结构模型以来, 分子生物学经过 20 年来的研究和探索, 终于在 70 年代初期取得了决定性的突破, 产生了一门全新的学科——基因工程。基因工程的理论基础是分子生物学, 它的出现不仅带动了现代生物技术产业的发展, 也使生命科学的研究产生了革命性的变化。本章不是全面地介绍分子生物学的内容, 仅就与基因工程有关的分子生物学基础知识作一简明的提示。

1.1 遗传信息传递的方向——中心法则

生命界除了某些病毒是以 RNA 作为其遗传信息的载体外, 绝大多数生物是将其遗传信息贮存在 DNA 分子中, 而其功能的实现则是通过蛋白质分子。1958 年 Crick 提出遗传信息传递的中心法则:



DNA 通过以自身为模板进行复制而使遗传信息代代相传, 并通过 RNA 最终将遗传信息传给蛋白质分子。利用 DNA 为模板合成 RNA 的过程叫转录; 以 RNA 为模板合成蛋白质的过程叫翻译。遗传信息这种单向不可逆传递的方式, 由于逆转录酶的发现而受到挑战。1970 年 Temin 和 Baltimore 在 RNA 肿瘤病毒中发现逆转录酶, 在细胞侵染期间, RNA 通过逆转录过程转换为单链 DNA, 然后再转换为双链 DNA, 并可插入到细胞染色体 DNA 中一代代遗传下去。逆转录酶可以以 RNA 为模板将贮存在 RNA 肿瘤病毒中的遗传信息转给 DNA 分子。为此, Crick 于 1971 年对中心法则作了补充, 提出三角形中心法则 (图 1-1)。中心法则阐明了贮存在核酸中的遗传信息的连续性和传递的方向。然而 Crick 的中心法则只告诉我们一个蛋白质的氨基酸顺序是由为其编码的基因的核苷酸顺序所决定的。近来的研究指出, 蛋白质的生物活性不仅依赖于其氨基酸顺序, 而且同其特定的三维结构的完整性密切相关。很明显, 遗传信息从基因转变成具生物活性的功能蛋白质是通过两种“密码”介导的: 一种就是遗传密码, 通过遗传密码将 mRNA 多核苷酸顺序翻译成线性多肽链; 另一种是折叠密码 (folding code), 其决定存在于氨基酸顺序中的一维结构信息向蛋白质特定的三维结构的转变。鉴于此, 中心法则中 RNA → 蛋白质的方向应写为 RNA → 多肽链 → 蛋白质。折叠

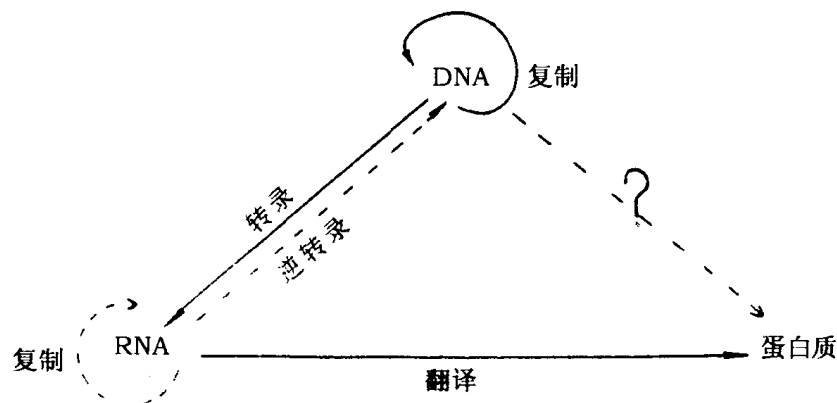


图 1-1 遗传信息传递的中心法则

细胞中的过程用实线表示，仅在病毒侵染时才发生的过程用虚线表示。

密码到底是什么？它是如何指导从多肽链到蛋白质的折叠？这正是当今关于新生肽链折叠研究中关键的问题。

1.2 DNA 的复制

对遗传物质的关键性要求是它必须能够准确地复制。图 1-2A 给出 DNA 分子复制的基本

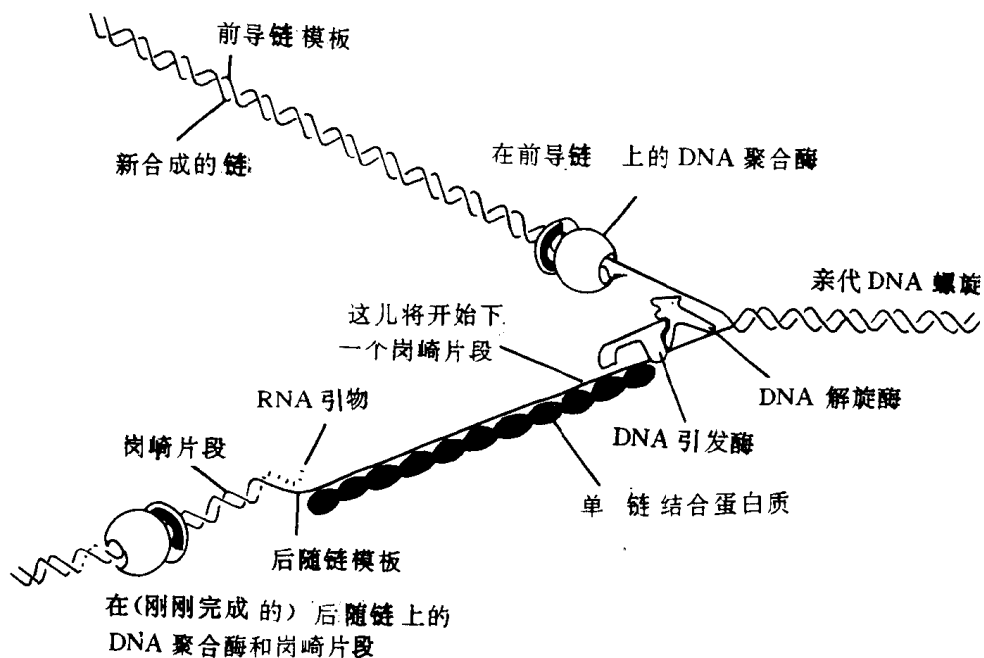


图 1-2A DNA 分子复制的基本模式

示半不连续、半保留复制及作用于 DNA 复制叉上的主要蛋白质类型
(Molecular Biology of the Cell P257, 1994, B. Alberts et al.)

模式。DNA 复制有如下特点：

(1) DNA 复制从特定的位点开始，这个特定位点就叫做复制起始点。在复制起始点双链 DNA 解旋，形成所谓的复制叉。现在已知最小的大肠杆菌复制起始点由大约 245bp 组成并起

始双向复制。从酵母中分离出的复制起始点叫做自主复制序列 (autonomously replicating sequence, ARS), 是一个由 11 个碱基对组成的共有序列, 这个短的共有序列几乎完全由 A · T 碱基对组成, 而第一个和最后一个 A · T 碱基对的方向可能是相反的 (图 1-2B)。



图 1-2B 酵母自主复制序列
(Genes IV P336, 1990, B. Lewin)

(2) 半保留复制, 即双链 DNA 分子在复制过程中, DNA 的两条链各自作为新链合成时的模板。复制后, 每一双链体都是由一条亲链和一条新合成的子链组成。

(3) 半不连续复制, 即 DNA 分子在复制过程中, 复制叉是不对称的。两条新合成的链中, 一条链是连续合成的, 叫做前导链 (leading strand); 另一条是在模板 DNA 的指导下, 通过 DNA 引发酶 (DNA primase) 在特定的间隔区合成由大约 10 个核苷酸组成的 RNA 引物 (RNA primer), 为 DNA 聚合酶提供 3'-OH, 然后合成一个个称之为冈崎片段 (Okazaki fragments) 的不连续的 DNA 片段, 最后通过专门的 DNA 修复系统, 快速去除 RNA 引物而代之以 DNA, 再经 DNA 连接酶通过 3', 5'-磷酸二酯键将其连接起来。这一条链叫做后随链 (lagging strand)。无论是前导链还是后随链, 其合成的方向都是从 5' 到 3'。由于 DNA 聚合酶在没有引物的情况下, 不能从头合成 DNA, 前导链的合成是如何起始的呢? 一种可能是同后随链一样, 由 RNA 引物提供 3'-OH; 一种可能是模板链上的起始点产生缺口, 暴露出 3'-OH 作为引物; 第三种可能是由后随链提供 3'-OH, 这将在下面介绍。

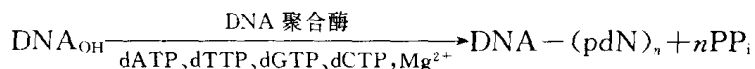
(4) DNA 复制具有高度的忠实性, 其复制的忠实性同 DNA 聚合酶所具有的自我校正功能密不可分。DNA 聚合酶不能通过将两个核苷三磷酸连接起来开始新的多核苷酸链的合成, 绝对需要碱基配对的引物提供 3'-OH。如在其引物链上存在一个碱基错配的 3'-OH 末端, 此 DNA 分子就不能作为有效的模板。DNA 聚合酶本身所具有的 3'→5' 的外切核酸酶活性, 使其能对错配的碱基进行有效的校正, 从而保证了 DNA 复制的高度忠实性。

(5) 虽然 DNA 聚合酶是 DNA 复制的主酶, 然而 DNA 复制是多种酶和蛋白因子协同有序工作的结果。如螺旋失稳蛋白 (helix-destabilizing proteins) 或叫单链 DNA 结合蛋白 (single strand DNA binding protein, SSB), 其功能是同单链 DNA 结合, 使相关的 DNA 双螺旋失掉稳定性, 有利于 DNA 双螺旋的解旋; DNA 解旋酶 (DNA helicases) 的功能是在复制叉处使双螺旋 DNA 解旋; DNA 拓扑异构酶 (DNA topoisomerases) 目前知道有两种 (I、II 型) 是 DNA 复制所必需的, 它们通过分别在 DNA 双螺旋的一条链和两条链的不同位点产生断裂和重新连接的方式, 帮助 DNA 双螺旋有效地旋转、解旋, 并能在复制完成后在 DNA 双链中引入超螺旋, 帮助 DNA 缠绕、折叠。DNA 连接酶是将 DNA 片段通过 3', 5'-磷酸二酯键连接起来。下面将专门介绍 DNA 聚合酶。

1.3 DNA 聚合酶

DNA 聚合酶是指以脱氧核苷三磷酸作为底物催化合成 DNA 的一类酶, 其催化的反应

为:



所有的原核和真核细胞中都含有几种 DNA 聚合酶, 这些酶协同作用负责染色体 DNA 的复制、DNA 分子损伤的修复、核 DNA 的重组以及染色体外 DNA 的复制等。

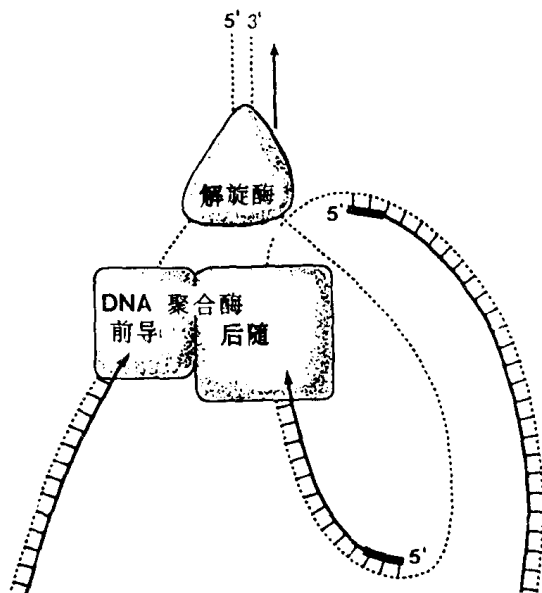


图 1-3 DNA 复制模型 (TIBS, 17, 1992)

示不对称 DNA 复制叉。大多数生物体的两条 DNA 链复制时, 需要不对称型的 DNA 复制叉。点线示亲本链, 实线示新合成的链; 短粗线示引物。

原核细胞(如细菌)中有三种 DNA 聚合酶即聚合酶 I、II、III, 其中聚合酶 III 是 DNA 复制的主酶, 由 α 、 β 、 δ 、 γ 、 ϵ 、 θ 和 τ 等 7 个亚基组成。DNA 复制过程中, 前导链和后随链是由具不同聚合酶活性的酶进行复制的, 然而原核生物中只有 DNA 聚合酶 III 为复制酶, 如何解决这一问题呢? 现在的研究指出, DNA 聚合酶 III 核心酶的三个亚基 (α 、 ϵ 、 θ), 无论在合成前导链还是在合成后随链的 DNA 聚合酶中都是相同的, 此核心酶通过与不同的亚基结合装配成特性不同的全酶复合物, 以使它们具有分别复制前导链和后随链的能力。图 1-3 给出这种 DNA 的复制模型。按照这个模型, 在复制叉处上述两种特性的 DNA 聚合酶在 τ 亚基作用下组成不对称的双体, 一个负责前导链的合成, 另一个负责后随链的合成。在复制过程中, 后随链在复制叉处转 180 度的弯形成一个环(loop), 这种结构的形成使双体 DNA 聚合酶沿两条模板链以同复制叉相一致的方向往前移动, 而又不破坏 DNA 复制是从 5'→3' 进行的原则。

真核细胞中目前已知有五种 DNA 聚合酶, 分别叫做聚合酶 α 、 β 、 γ 、 δ 、 ϵ 。聚合酶 β 是核中主要的修复酶并参与 DNA 重组; 聚合酶 γ 负责线粒体基因组的复制; 聚合酶 α 、 δ 、 ϵ 是复制主酶。聚合酶 α 是核 DNA 复制的重要酶之一, 由于它缺少 3'→5' 外切核酸酶的活性且同引发酶密切结合, 认为其负责复制叉后随链合成的起始。聚合酶 δ 和 ϵ 具有 3'→5' 外切核酸酶的活性, 这表示二者都具有校对功能。 δ 和 ϵ 之间的差别是, 在催化持续的 DNA 合成时 δ 需要一个辅助蛋白因子: 增殖细胞核抗原 (proliferating cell nuclear antigen, PCNA), 而 ϵ 则不需要。然而, 二者能与相同的辅助蛋白 (PCNA、RF-A、RF-C 等) 形成具有复制功能的全酶结构, 且在细胞内都是重要的基因产物, 说明 δ 和 ϵ 是染色体 DNA 复制中的关键酶。

目前认为 DNA 聚合酶 α /引发酶起始后随链 DNA 的合成。当一段短的 DNA 片段已被合成时, 在有 ATP 存在的条件下, 辅助蛋白 PCNA 和复制因子 C (RF-C) 在正在延伸的 DNA 链的 3'-OH 末端形成一紧密结合的复合物, 这种结合阻止了聚合酶 α /引发酶再同 3'-OH 末端结合, 而代之以聚合酶 δ 结合, 开始了前导链的合成。解离下来的聚合酶 α /引发酶则转到后随链上的下一个引发位点重新开始不连续的复制。图 1-4 给出真核细胞 DNA 复制的工作模型。聚合酶 α 、 δ 、 ϵ 在 DNA 复制的延伸过程中可能共同完成复制的任务。DNA 双链在给定

复制起始点解旋后，负责起始的 DNA 聚合酶（DNA 聚合酶 α /引发酶）在复制叉处合成 RNA 引物及一小段短的 DNA 片段从而起始后随链合成。同聚合酶 α 紧密结合的引发酶是后随链引发过程中不可缺少的酶。在 DNA 合成起始后，DNA 聚合酶 α /引发酶被负责延伸的 DNA 聚合酶所代替，而进入 DNA 复制的延伸过程，前导链被聚合酶 δ 或 ϵ 进行复制。与此同时，负责 DNA 复制起始的 DNA 多聚酶 α /引发酶在后随链继续合成 RNA 引物和 DNA 短片段，而各个冈崎片段随后由 DNA 聚合酶 ϵ 或 δ 来完成。

到底什么因素决定了哪种 DNA 聚合酶（ δ 还是 ϵ ）负责前导链还是后随链的延伸合成？一个可以接受的说法是：由于同各 DNA 聚合酶结合的各种辅助因子（如 PCNA、RF—A、RF—C 以及 DNA 解旋酶等）的不同的组成而决定了 DNA 聚合酶各自不同的功能。由于聚合酶 δ 和 ϵ 都具有 $3' \rightarrow 5'$ 的外切核酸酶的活性，所以上述的这种 DNA 复制的工作模型也解释了两条 DNA 链在复制过程中可以分别通过 δ 和 ϵ 进行校对，从而保证了 DNA 复制的忠实性。从图 1-4 中也可以看出，在双向复制的过程中，随着后随链的延伸合成，从不对称的双体聚合酶之一穿过后去起始前导链的合成。

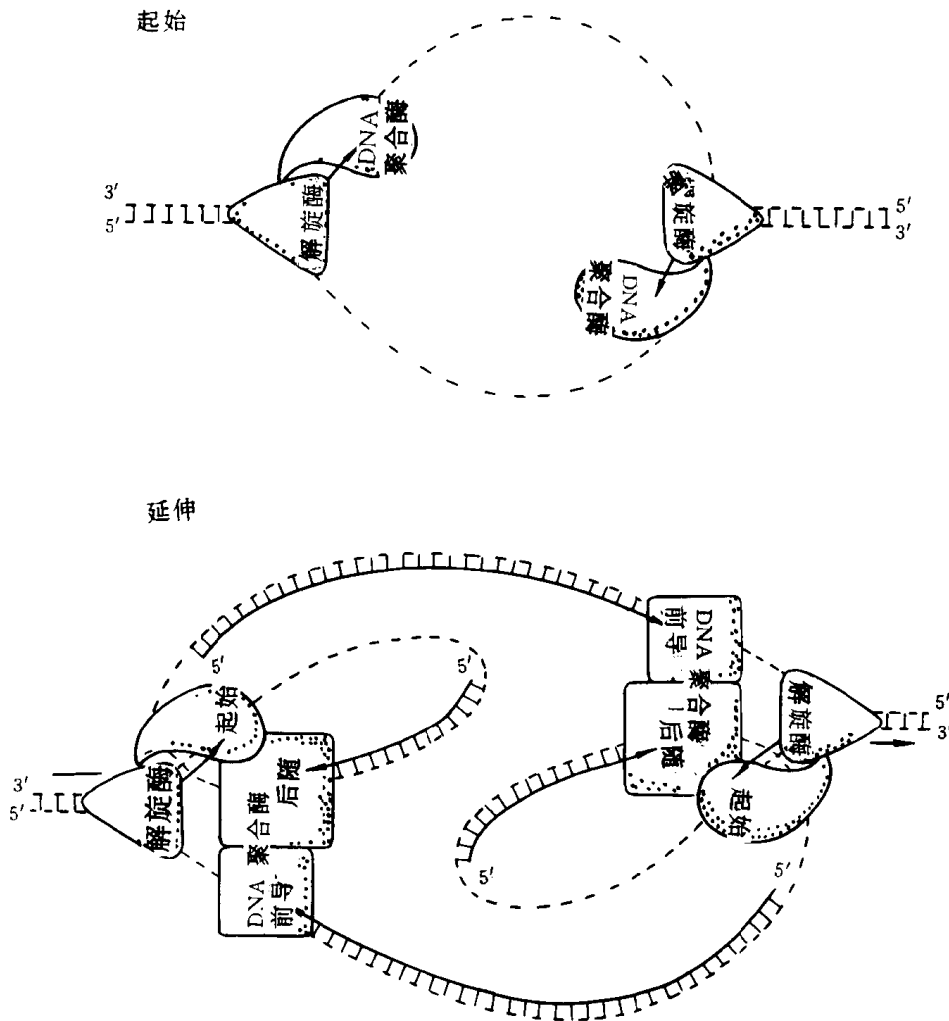


图 1-4 真核细胞 DNA 复制模型 (TIBS, 17, 1992)

点线示亲本链，实线示新合成的链；短粗线示引物。

1.4 原核和真核生物基因结构的特征

人们通常根据细胞内是否存在真正的细胞核膜,将有细胞的生物分为原核和真核生物。原核生物一般只有由一个核酸分子组成的染色体,以DNA双螺旋结构为主。一般而言,原核基因具有如下的结构特征:(1)功能相关的基因大多数以操纵子(operon)结构出现,如大肠杆菌中的乳糖操纵子、色氨酸操纵子等。操纵子是细菌基因表达和调控的一个完整单位,包括结构基因、调控基因和被调控基因产物所识别的DNA调控元件,如启动区(promoter)以及操纵区(operator)等。(2)蛋白质基因通常以单拷贝的形式存在。一般而言,为蛋白质编码的核苷酸顺序是连续的,中间不被非编码顺序所打断。(3)编码RNA的基因通常是多拷贝的。值得指出的是,位于DNA上的各种调控元件的DNA顺序是多种多样的,这为基因表达调控的多样性和精确性提供了结构基础。

真核生物的基因结构要比原核生物的基因结构复杂得多,其主要的结构特征是:(1)真核生物具有复杂的染色体结构,染色体由染色质构成。染色质是真核生物细胞核中的遗传物质在细胞分裂间期的存在形式,由DNA、组蛋白、非组蛋白以及RNA四种大分子组成,其基本结构物质是DNA和组蛋白。核小体是染色质的基本单位,包括200个碱基对左右的DNA和由H_{2A}、H_{2B}、H₃、H₄组成的组蛋白八聚体以及一个分子的组蛋白H₁三部分。染色质的这种复杂的组织结构显示出真核基因表达调控的复杂性。这里需要指出的,真核染色体上有二个特殊的DNA序列,这就是着丝点(centromere)和端粒(telomere),它们同DNA复制起始点一起构成染色体不可缺少的三要素,目前通用的酵母人工染色体(YAC)以及正在研制的哺乳动物细胞人工染色体(MAC)就是以此为基础组建的。(2)根据DNA复性动力学的研究,真核生物DNA顺序可分为非重复序列、轻度重复序列、中度重复序列以及高度重复序列。非重复序列指的是一个基因组中只有一个拷贝,真核生物的大多数基因在单倍体细胞中都是单拷贝的;轻度重复序列是指在一个基因组中有2~10个拷贝的序列,如组蛋白基因和酵母tRNA基因,此组序列有时并入非重复序列;中度重复序列是指在一个基因组中有10至几百个拷贝的序列,其平均长度大约为300bp,中度重复序列一般是不编码的序列,推测在基因表达调控中起重要作用;高度重复序列,其在基因组中的拷贝数可以从几百到几百万个,通常说的卫星DNA就属于高度重复序列。值得指出,重复序列的分类标准并非严格确定,但重复序列的存在是真核生物DNA区别于原核生物DNA的一个重要特征。(3)真核基因具不连续性。绝大多数真核生物的基因都是不连续基因(interrupted gene)。所谓不连续基因就是基因的编码序列在DNA分子上是不连续的,为不编码的序列所隔开,不连续基因是通过mRNA和DNA杂交试验发现的。就为蛋白质编码的DNA序列而言,那些在成熟mRNA中仍存在的序列称为外显子(exon);而将那些存在于不连续基因中,但在成熟的mRNA中不存在的序列称为内含子(intron)。值得指出,外显子和内含子的概念与是否编码氨基酸的概念并不相对应。从不连续基因到成熟mRNA之间存在着一个基因转录的中间体,叫做初级转录物(primary transcripts),对于编码蛋白质的基因而言又叫做不均一核RNA(heterogeneous nuclear RNA, hnRNA),这个基因的初级转录物既含有外显子又含有内含子序列,从不均一核RNA到成熟mRNA要经过一个所谓的转录后加工的拼接过程。真核生物基因的不连续性和转录后加工是真核基因有别于原核基因的又一重要特征。当然也应该指出,原核生物中基因的连续性和真核生物基因的不连续性并不是绝对的,在真核生物中也有些基因是不

含内含子的,如组蛋白基因、大多数酵母蛋白基因等。(4)真核生物的基因组中有许多来源相同、结构相似、功能相关的基因组成单一的基因簇或称基因家族(gene family),如血红蛋白基因家族。(5)真核生物中存在串联重复基因,其特点是各成员之间有高度的序列一致性甚至完全相同,拷贝数高、非转录的间隔区短而一致。组蛋白基因、rRNA基因、tRNA基因都是串联重复基因,这些基因的产物在细胞中都是大量需要的。

上面我们概括地介绍了原核基因和真核基因的结构特征,应该说这些结果是对有限基因的研究而得到的,随着研究工作的不断深入,对于基因结构特征将会有更透彻的了解,这将为基因工程设计提供更可靠的理论基础。

1.5 RNA的转录和RNA的加工

转录是指在DNA模板上合成RNA的过程,转录是基因表达的关键一步,DNA分子中所贮存的遗传信息,必须转录成信使RNA(mRNA)才能通过蛋白质生物合成的过程转变成具有生物活性的蛋白质。在本节我们主要集中讲述为蛋白质编码的基因的转录和转录后加工。

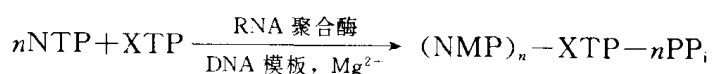
1.5.1 RNA合成的基本特征

1. RNA聚合酶是RNA合成的关键酶。在原核生物中只有一种RNA聚合酶,催化所有种类RNA的合成;在真核生物中有三种不同的RNA聚合酶,分别称为RNA聚合酶I、II、III,在RNA合成中分别催化5.8S rRNA、18S rRNA、28S rRNA, hnRNA(mRNA前体)、snRNA,以及tRNA、5S rRNA的合成。RNA聚合酶能在DNA模板上起始一条新链的合成,起始的核苷酸一般为嘌呤核苷三磷酸,而且在RNA链的5'末端保持这一三磷酸基团。

2. RNA合成是以四种核糖核苷三磷酸:ATP、GTP、CTP、UTP为底物。

3. RNA转录是以一条DNA链为模板,按照碱基互补的原则($A=U, G\equiv C$)进行转录。RNA聚合酶通过识别DNA上的特定的启动子序列开始转录, RNA聚合酶按3'→5'的方向沿着DNA模板移动,而RNA新链的合成则以5'→3'的方向进行,随着RNA链的增长和RNA聚合酶的向前移动,原来解链部分的DNA又恢复双螺旋结构,而RNA链逐步游离于DNA分子之外,当RNA聚合酶遇到在DNA上的转录终止信号(序列),其释放了DNA模板和新合成的RNA链完成了转录。

4. RNA的合成的整个过程如图1-5所示,包括(1)RNA聚合酶结合于DNA分子上的特定位点;(2)使DNA双链解旋,起始RNA合成;(3)RNA链的延伸;(4)RNA合成的终止和释放。其整个的反应过程可以下式表示:



其中NTP代表四种核糖核苷三磷酸;XTP代表RNA 5'末端的核苷三磷酸;PP_i代表合成过程中释放出的焦磷酸。在37℃,RNA聚合酶的合成速率大约是30核苷酸/秒。

1.5.2 RNA聚合酶

原核生物的RNA聚合酶以大肠杆菌的RNA聚合酶为代表加以介绍。大肠杆菌的RNA聚合酶通常认为由五个亚基组成,即 $\alpha_2\beta'\sigma$,其分子量分别为36.5kDa、150kDa、160kDa和80kDa。 $\alpha_2\beta'\sigma$ 组成的酶叫做全酶,而只由 $\alpha_2\beta'$ 四个亚基组成的酶叫做核心酶。此外,还可以看到一个分子量在10kDa左右的 ω 亚基,其功能尚不清楚;后来又发现一个69kDa的酸性蛋白质,其功能可能与RNA转录终止有关,叫做NusA亚基。RNA聚合酶五个主要亚基的

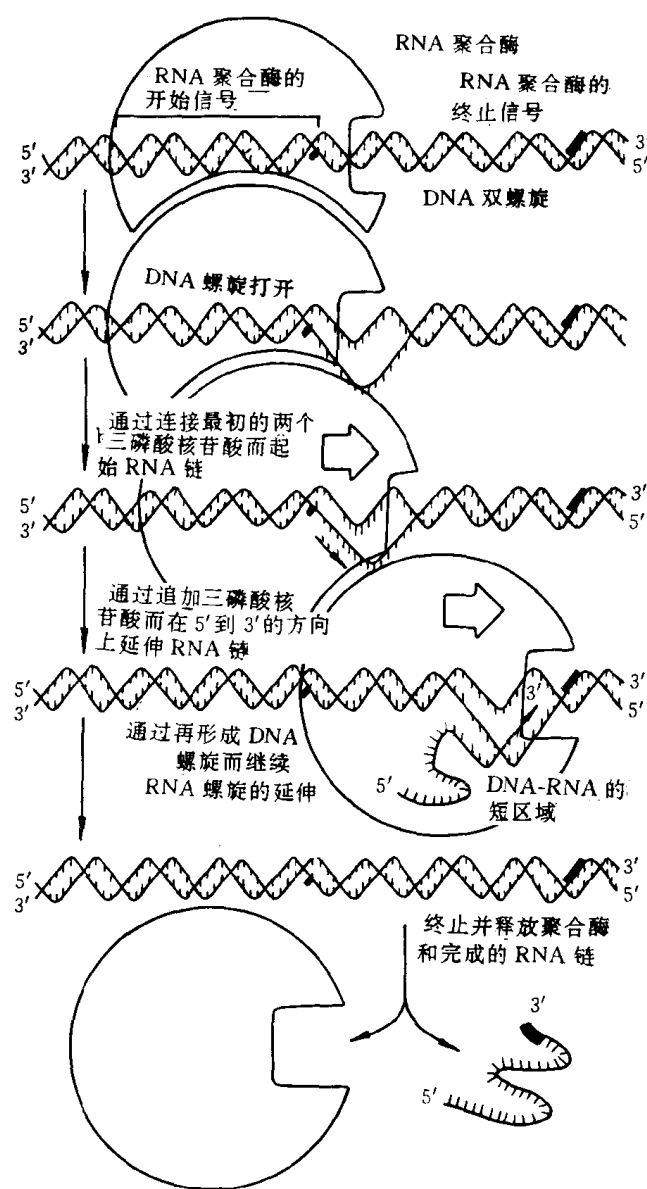


图 1-5 RNA 合成过程的示意图

RNA 在称为启动子的 DNA 上的特定位点开始合成；在特定的终止位点完成合成过程，并释放出 RNA 聚合酶的 RNA 链。(Molecular Biology of the cell, P201, 1984)

种具不同生命功能的、成熟 RNA。真核细胞中这三种不同的 RNA 聚合酶可以用它们各自对 α -鹅膏蕈碱 (α -amanitin) 的敏感性不同来进行区别。RNA 聚合酶 I 对 α -鹅膏蕈碱不敏感，在 $>10^{-3}$ mol/L 才表现出轻微的抑制作用；RNA 聚合酶 II 对 α -鹅膏蕈碱最敏感，在 $10^{-9} \sim 10^{-8}$ mol/L 浓度下就会被抑制；RNA 聚合酶 III 对 α -鹅膏蕈碱的敏感度介于 I、II 之间，在浓度为 $10^{-5} \sim 10^{-4}$ mol/L 时表现出抑制。

值得指出的是，纯化了的原核细胞来源的 RNA 聚合酶在一个较为简单的 in vitro 转录系统中（如 1.5.1 所示）可以产生忠实的转录，而纯化了的真核 RNA 聚合酶则需要一系列转录因子的参与才能在 in vitro 条件下进行转录，这说明真核生物中有效的转录调控要比原核生

功能尚不完全清楚， α 亚基可能参与全酶同启动子的牢固结合。这一牢固结合需要 DNA 双螺旋的局部解链，当核心酶沿着 DNA 模板移动进行 RNA 链的延伸时，需要不断地在前面解开双螺旋，在后面恢复双螺旋，这些作用可能与 α 亚基的功能有关。 β 亚基可能参与底物的结合以及 RNA 合成时磷酸二酯键的形成。 β' 亚基的碱性最强，它适于担任与 DNA 模板结合的角色。 σ 亚基与核心酶结合，使核心酶产生构象上的变化，使全酶对启动子的结合力大大提高。因此， σ 亚基的主要功能是识别启动子，特别是启动子上的初始结合位点，DNA 双链中哪条链被转录（即转录的方向）及转录的起点选择都与 σ 亚基有关。

真核生物的 RNA 聚合酶与原核生物不同，有三种位于不同地点的 RNA 聚合酶，其亚基数远多于原核 RNA 聚合酶。RNA 聚合酶 I 位于核仁中，负责转录 5.8S、18S 和 28S rRNA。RNA 聚合酶 II 位于核质中，负责核不均一 RNA 的转录，即 mRNA 前体的转录以及与 RNA 拼接有关的核小 RNA (snRNA) 的转录。RNA 聚合酶 III 也位于核质中，负责 tRNA 及 5S rRNA 的转录。实际上在真核生物中所有的 RNA 聚合酶都是负责各自的 RNA 前体的转录，而这些前体转录物都要经过转录后加工才最后形成各种

物复杂得多,我们将在有关真核细胞中基因表达调控的章节进行介绍。

1.5.3 与转录调控有关的 DNA 序列

在本节开始之前,首先介绍几个具体的概念:(1)有义链或编码链(sense or coding strand),是指在DNA的两条链中不作模板转录的那条链。这样,有义链的方向和核苷酸序列都与从这一区段转录出来的RNA序列相一致,其唯一的差别是在RNA序列中U代替了T。(2)反义链或模板链(antisense or template strand),是指作为RNA转录的模板链,其方向同转录出来的RNA相反,核苷酸序列同RNA相互补。(3)转录起始点及有关位置表示法:转录起始点是指mRNA开始转录的第一个碱基,此点通常用+1表示,由此与转录方向相一致的下游序列的碱基位置用正值表示;与转录方向相反的上游序列的碱基位置用负值表示。

转录的起始是基因表达的关键所在,因此对转录起始相关的DNA序列的了解在设计高效表达载体时是很有意义的。原核生物中,很多功能上相关的基因都以操纵子的形式存在,其转录由一共同的控制区及调控蛋白因子调控。原核基因调控序列通常由称为启动子(RNA聚合酶的识别和结合位点)以及正、负调控蛋白因子的结合位点所组成。从168个对大肠杆菌RNA聚合酶结合的启动区DNA序列可以得出一个共有序列,即称为-10区的pribnow框(TATAAT)和被称为-35区的Sextama框(TTGACA)。 -10 区是RNA聚合酶的牢固结合位点, -35 区是RNA聚合酶的识别位点。这两个序列是决定启动子强度的重要因素。在 -10 区和 -35 区之间的碱基序列的组成并不特别重要,然而这两个序列之间的距离却十分重要。实验表明,两个序列之间为17bp时转录效率最高。天然启动子中这段距离大多为15~20bp。 -10 区和 -35 区间距的大小也可能是决定启动子强度的因素之一。一般说来,对于给定的启动子,其特异性序列趋向于启动子共有序列(consensus sequence)时 -10 与 -35 区之间的间距趋于17bp时,此启动子转录效率可能就高。基于此,可以通过突变来构建高效表达外源基因的强启动子。

除了 -10 及 -35 区之外,在结构基因的上游区还存在与转录调控有关的负调控区和正调控区。以乳糖操纵子为例(图1-6),在紧靠结构基因LacZ的上游区有一段DNA序列叫做操纵区(operator简称为O),它是调节基因(i基因)的产物——阻遏蛋白的结合位点。阻遏蛋白本身有两个结合位点,分别与操纵区和诱导物结合,前者称操纵位点,后者称诱导物位点。当细胞中不存在诱导物时,阻遏蛋白与操纵区相结合,使RNA聚合酶的转录被阻断。当足够量诱导物存在时,诱导物同阻遏蛋白结合,改变了阻遏蛋白的构象,导致阻遏蛋白从操纵区解离,使RNA聚合酶转录得以进行。这就是由调节基因通过其产物与操纵区相结合以及诱导物之间关系所构成的乳糖操纵子的基因负调控机制。

环化AMP受体蛋白(cAMP-CAP)的结合位点是基因转录的正调控区。乳糖操纵子上有两个cAMP-CAP结合位点,分别处于 $-70 \sim -50$ 和 $-50 \sim -40$ 区段,称为结合位点I和II(图1-6)。位点I包含一个反向重复序列,这似乎是大多强结合位点的特征。位点II是一个很弱的结合位点,但当cAMP-CAP复合物同位点I结合后,使位点II同cAMP-CAP复合物的结合力显著提高,一旦位点II被cAMP-CAP复合物结合,RNA聚合酶就很快与 -35 区及 -10 区结合而起始转录。cAMP-CAP作为一个正调控蛋白因子,对那些CAP依赖性启动区(如乳糖操纵子)的起始转录是必需的。对于不同的操纵子,CAP在其启动区的结合部位并不相同,如在阿拉伯糖操纵子(ara)中,CAP的结合位点位于 -107 到 -78 区段。除上述有关调控序列外,在原核细胞中也发现类增强子(enhancer)的序列,如从大肠杆菌trpE基因3'