



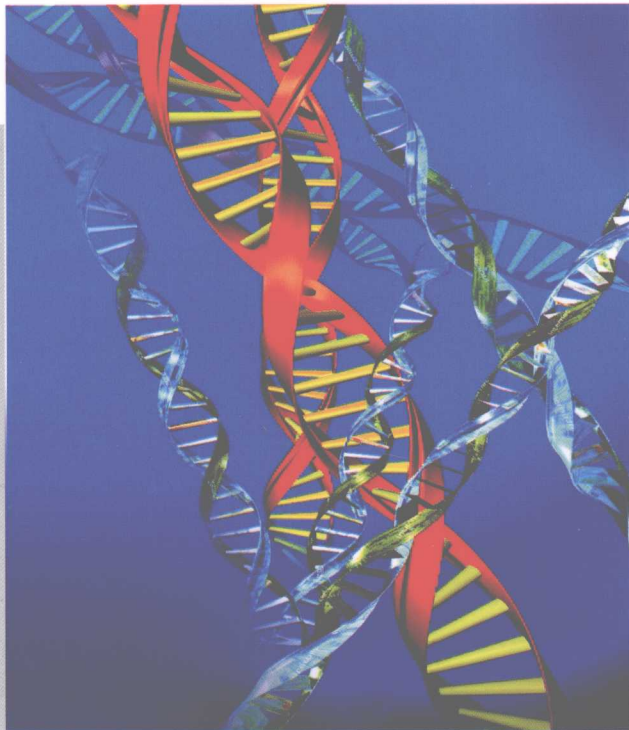
高等院校生命科学专业基础课教材

# 基因工程及其分子生物学基础

# 分子生物学基础分册

(第2版)

静国忠 编著



北京大学出版社  
PEKING UNIVERSITY PRESS



# 现代生物科学前沿与生物技术应用 分子生物学基础分卷

【第 1 版】

ISBN 7-309-05411-9



9 787309054119 >

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

——分子生物学基础分册

(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：分子生物学基础分册/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 978-7-301-15527-1

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 121159 号

书 名：基因工程及其分子生物学基础——分子生物学基础分册(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 978-7-301-15527-1/Q·0119

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](mailto: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 12.75 印张 310 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：22.00 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](mailto:fd@pup.pku.edu.cn)

## 第 2 版前言

作为生物学科的基础课,“基因工程”课程使学生既要了解什么是基因工程,又要了解什么是它的理论基础。写这本书的目的是希望给读者,尤其是想要了解此领域的大同行或小同行,提供一本较精炼的、基本概念清楚且有一定深度的分子生物学基础及基因工程的读本或教材。第 2 版在原版的基础上做了较大的补充和修改,并以分子生物学基础分册和基因工程分册的形式出版。

在分子生物学基础分册中,加强了对基因工程的分子生物学基础内容的介绍。对于原核和真核基因在复制、转录、翻译等水平的表达调控机制进行了较为系统的介绍,并对其在基因工程操作中的意义做了必要的提示。对蛋白质的折叠和错误折叠机制过程加以较精炼的介绍,还对蛋白质的剪接、蛋白质的结构及其测定方法,特别是对蛋白质溶液构象在研究蛋白质结构与功能中的应用,以及在基因工程中对蛋白质产物分析的意义进行了介绍。在上述基础上,对基因表达调控的其他方面,如转录衰减作用与基因表达调控、信号转导与基因表达调控以及 RNA 干涉的分子机制与基因沉默等方面作了概述。

在基因工程分册中,加强了对基因工程原理、外源基因在受体细胞内表达过程中所遇到的问题和解决办法以及相关的技术方法的论述:

(1) 对外源基因在宿主细胞中高效表达、分泌表达,以及基因的融合和融合蛋白的表达等内容进行了系统化和补充。

(2) 对实现重组蛋白正确折叠的方法和包涵体变性——复性的方法作了较为系统的介绍。

(3) 在“几种真核细胞表达系统”和“分子杂交技术”的相关章节中分别加入了对转基因动、植物,DNA 疫苗和 DNA 微阵列——基因组芯片等内容的介绍。

(4) 在“聚合酶链反应及其应用”章节中加大了对常用 PCR 方法的原理及应用的论述。

(5) 将“噬菌体展示技术”一章改为“各种生物学展示技术”。内容包括噬菌体展示技术、细菌展示技术、酵母展示技术、核糖体展示技术和 mRNA 展示技术等基本原理及其应用。

(6) “在基因打靶技术及其应用”章节中,较系统地介绍了基因打靶技术的基本原理、组织特异性基因打靶、转座子和 RNA 干涉与基因敲除等内容;并对常用的细菌基因敲除方法进行了概述。

(7) 在“基因突变”章节中补充了易错 PCR 和 DNA 改组的原理和应用。

(8) 蛋白质相互作用和蛋白质-核酸相互作用原理及其分析方法是分子生物学的重要内容之一,是研究基因表达调控的重要手段。由于这方面的内容,如酵母双杂交系统和细菌双杂交方法等,涉及应用基因工程方法的问题,因此,我们将其放在“基因工程分册”中第 15 章和第 16 章加以介绍。

本书共分 26 章,为了便于读者理解文字内容,第 2 版本共加入 321 个插图。文中的关键词后仍给出相对应的英文,以便读者进行网上查寻。

对书中可能出现的不当和疏漏之处切望科技同行指正。

在此,我要感谢我们实验室的同仁以及我的学生们对我的极大的鼓励、支持和关心。感激我家人的无微不至的关爱。感谢北京大学出版社黄炜老师辛勤、细致的工作,使第2版书以最快的速度同读者见面。

静国忠

于中国科学院生物物理研究所

生物大分子国家重点实验室

# 第 1 版前言

《基因工程及其分子生物学基础》是以我在中国科学院及中国科技大学研究生院授课的讲义为基础并结合实验室的工作写成的。作为研究生的基础课教材,应该是注重基础兼顾提高。为此,本书加大了关于基因工程的分子生物学基础的内容,尽量将近年来国内外在分子生物学研究中的新进展写入,并将蛋白质折叠、分子伴侣和折叠酶以及蛋白质剪接等分专门章节加以介绍,同时针对教学中发现的问题,加强了基本概念的论述。希望来自不同高等院校的学生,通过本书的学习,得到更好的基础训练,开阔思路,有所提高,有所前进。

关于基因工程章节的内容,本书有别于基因工程或 DNA 重组技术操作手册,重点介绍基因工程的原理,并对外源基因在宿主细胞中的表达、分泌、折叠等相关问题,进行了较系统的分析和综合归纳。考虑到国内多数大专院校对基因工程课的教授内容多偏于原核细胞,本书以专门章节,较系统地介绍了包括哺乳动物细胞、昆虫细胞以及整体动物等在内的真核细胞表达体系。为了较全面地将基因工程技术介绍给学生,本书除了对基因重组、基因突变、基因的化学合成、分子杂交、DNA 序列分析以及聚合酶链反应(PCR)等相关技术的原理和应用进行介绍外,对近年来发展起来的粒子轰击和基因转移、噬菌体显示技术、基因打靶及基因剔除技术、反义核酸技术、DNA(或基因)疫苗等也作了必要的介绍和论述,以期给学生一个较为全面的训练。

基因工程是一个包括上游工程和下游后处理工艺在一起的现代生物技术。基因工程的下游后处理工艺已发展成为生物工程技术的专门领域,本书只限于介绍基因工程的上游技术。众所周知,分子生物学及以其为基础的基因工程学是现今发展最快的学科之一,在编写本书时我们虽然尽量注意到将新的进展写入本书,但由于本学科的飞速发展和作者能力之所限,在此书完成之际,实有挂一漏万之感。书中出现的不当甚至错误之处,切望科技同行指正,便于在教学中及时修正,以免以讹传讹。

在本书写作过程中得到科技同仁的多方帮助。中国科学院生物物理研究所的周波同志、李昭洁同志从头至尾地通读了几乎整个教材,并提出了不少中肯的意见;北京大学出版社的周月梅老师在本书尚未完成之时,就欣然答应编辑出版此书;北京大学生命科学学院吴鹤龄老师在百忙中对本书进行了审阅,给了我很大的支持。这里需要特别提及的是中国科学院上海生物化学研究所的吴祥甫教授,他是从事昆虫-杆状病毒基因表达体系研究的专家,在本书写作过程中慨然提供给我有关这方面的研究成果和资料,加速了我的写作过程,也为本书增色匪浅。在本书出版之际,在此深表谢意。

静国忠

于中国科学院生物物理所  
生物大分子国家重点实验室

# 目 录

<b>1 遗传信息的传递和分子生物学的中心法则</b> .....	(1)
1.1 DNA 是遗传信息的主要载体 .....	(1)
1.2 RNA 是某些噬菌体和病毒的遗传信息载体 .....	(1)
1.3 RNA 反转录酶的发现改变了对遗传信息单向传递的认识 .....	(2)
1.4 在高等生物中 RNA 作为信息的载体可从亲代传给子代 .....	(4)
1.5 多肽链如何折叠成为功能蛋白质仍然是一个没有解决的问题 .....	(5)
<b>2 DNA 的复制</b> .....	(6)
2.1 DNA 结构的特征 .....	(6)
2.2 DNA 复制的一般特点 .....	(20)
2.3 原核细胞 DNA 的复制机器 .....	(30)
2.4 真核细胞 DNA 的复制机器 .....	(48)
2.5 DNA 重组 .....	(62)
<b>3 原核、真核生物染色体结构和基因结构的特征</b> .....	(64)
3.1 原核生物染色体结构 .....	(64)
3.2 原核生物基因结构特征 .....	(65)
3.3 真核生物染色体结构 .....	(66)
3.4 真核生物基因结构特征 .....	(69)
3.5 真核基因组中 DNA 序列复杂性分析 .....	(71)
<b>4 RNA 的转录和转录后的加工</b> .....	(72)
4.1 RNA 合成的基本特征 .....	(72)
4.2 与原核生物基因转录相关的序列 .....	(73)
4.3 原核生物基因转录起始及调控 .....	(75)
4.4 原核生物基因转录的延伸和终止 .....	(79)
4.5 真核生物基因转录起始及调控 .....	(82)
4.6 真核生物基因转录的延伸和终止 .....	(94)
4.7 在真核细胞中 mRNA 转录后加工 .....	(97)
4.8 RNA 编辑 .....	(113)
4.9 mRNA 功能的质量控制和 mRNA 转运 .....	(116)
4.10 反转录和反转录酶 .....	(117)
<b>5 翻译及翻译过程中的调控</b> .....	(119)
5.1 遗传密码 .....	(119)
5.2 参与蛋白质生物合成的生物大分子及其功能 .....	(125)
5.3 蛋白质生物合成的过程 .....	(132)
5.4 翻译效率的调控 .....	(144)
5.5 硒代半胱氨酸：是否是蛋白质中的第 21 个氨基酸 .....	(147)



5.6	蛋白质翻译后的修饰和加工	(149)
<b>6</b>	<b>蛋白质的折叠和错误折叠</b>	<b>(152)</b>
6.1	一个蛋白质的氨基酸序列决定其三维空间结构,即氨基酸序列为蛋白质的结构编码	(152)
6.2	分子伴侣和折叠酶	(153)
6.3	蛋白质质量控制,蛋白质错误折叠和折叠病	(160)
<b>7</b>	<b>蛋白质的剪接</b>	<b>(162)</b>
7.1	蛋白质剪接的发现	(162)
7.2	蛋白质剪接的机制	(163)
7.3	蛋白质剪接的应用	(166)
<b>8</b>	<b>蛋白质的结构及其测定方法概述</b>	<b>(168)</b>
8.1	蛋白质分子的一、二、三、四级结构	(168)
8.2	蛋白质各级结构的测定	(170)
<b>9</b>	<b>基因表达调控的其他方面</b>	<b>(180)</b>
9.1	转录衰减作用与基因表达调控	(180)
9.2	信号转导与基因表达调控	(183)
9.3	RNA 干涉与基因沉默	(185)
	<b>参考文献</b>	<b>(193)</b>



# 1 遗传信息的传递和分子生物学的中心法则

基因工程(genetic engineering)或称重组 DNA 技术(recombinant DNA technology)是 20 世纪 70 年代发展起来的一门全新的学科,是分子生物学研究理论和实践的结晶。因此,系统地掌握分子生物学的基本原理,特别是基因表达及其调控的分子机制,对于学习、领会和贯通基因工程学是十分重要的,是一个知其然也知其所以然的必经之路。

## 1.1 DNA 是遗传信息的主要载体

在生物进化的长河中,绝大多数生物选择了将 DNA 作为它们的遗传信息载体。绝大多数生物的基因组 DNA 为双链,而一些病毒(如细菌噬菌体  $\phi$ X174, M13)则以单链 DNA 作为其基因组。在后面的章节中我们会看到,无论是(+)单链 DNA(ss DNA(+)),还是(-)单链 DNA(ss DNA(-)),其复制的中间体(复制型)都是双链 DNA。值得指出的是,人们对什么是遗传物质的认识经历了从蛋白质到 DNA 的认识过程。如果从 Friedrich Miescher 在 1869 年发现 DNA 算起,到 DNA 最终被证明为遗传物质为止,用了 80 多年的时间。DNA 之所以能成为生物体遗传信息的载体,是由其独特的结构特性所决定的。正是这些独特的结构特性使得 DNA 分子能够更稳定地储存遗传信息,精确地传递遗传信息,通过突变、遗传和自然选择使生物体得以进化。与其说生物体选择了 DNA 作为其遗传信息载体,不如说是生物体的进化造就了 DNA。

## 1.2 RNA 是某些噬菌体和病毒的遗传信息载体

虽然绝大多数生物体以 DNA 作为它们的遗传信息载体,可是一些噬菌体(如 *E. coli* 的 MS2 噬菌体)和动、植物病毒以 RNA 作为其遗传信息载体。根据 RNA 病毒基因组复制的特点,可将 RNA 病毒分为两大类:一类是在其复制过程中不存在 DNA 阶段(DNA phase),另一类是在其复制时需要首先将它们的 RNA 基因组反转录成 DNA。前者就是我们平时说的 RNA 病毒,而后者就是反转录病毒(retroviruses)。进而,根据 RNA 病毒基因组的有义性或极性,将 RNA 病毒分为如下三类:正单链 RNA 病毒(positive-sense ss RNA viruses),如甲、丙、戊型肝炎病毒,SARS 病毒以及烟草花叶病毒等;负单链 RNA 病毒(negative-sense ss RNA viruses),如流感病毒、麻疹病毒、狂犬病病毒等;双链 RNA 病毒(double stranded RNA viruses),如轮状病毒。正链 RNA 基因组可直接作为 mRNA,通过翻译产生蛋白质,而负链 RNA 基因组必须经 RNA 聚合酶“转录”成正链 RNA,才能进行蛋白质合成。双链 RNA 病毒基因组也必须经 RNA 聚合酶“转录”后产生有功能的 mRNA。无论是单链 RNA 病毒基因组还是双链 RNA 病毒基因组的复制都是由病毒自身的 RNA 聚合酶或 RNA 复制酶(RNA

replicase)来完成的。就是说,这些 RNA 病毒基因组具有自我复制的能力。图 1-1 以 RNA 噬菌体 MS2 为例,给出正单链 RNA 病毒在增殖过程中所发生的一系列事件。

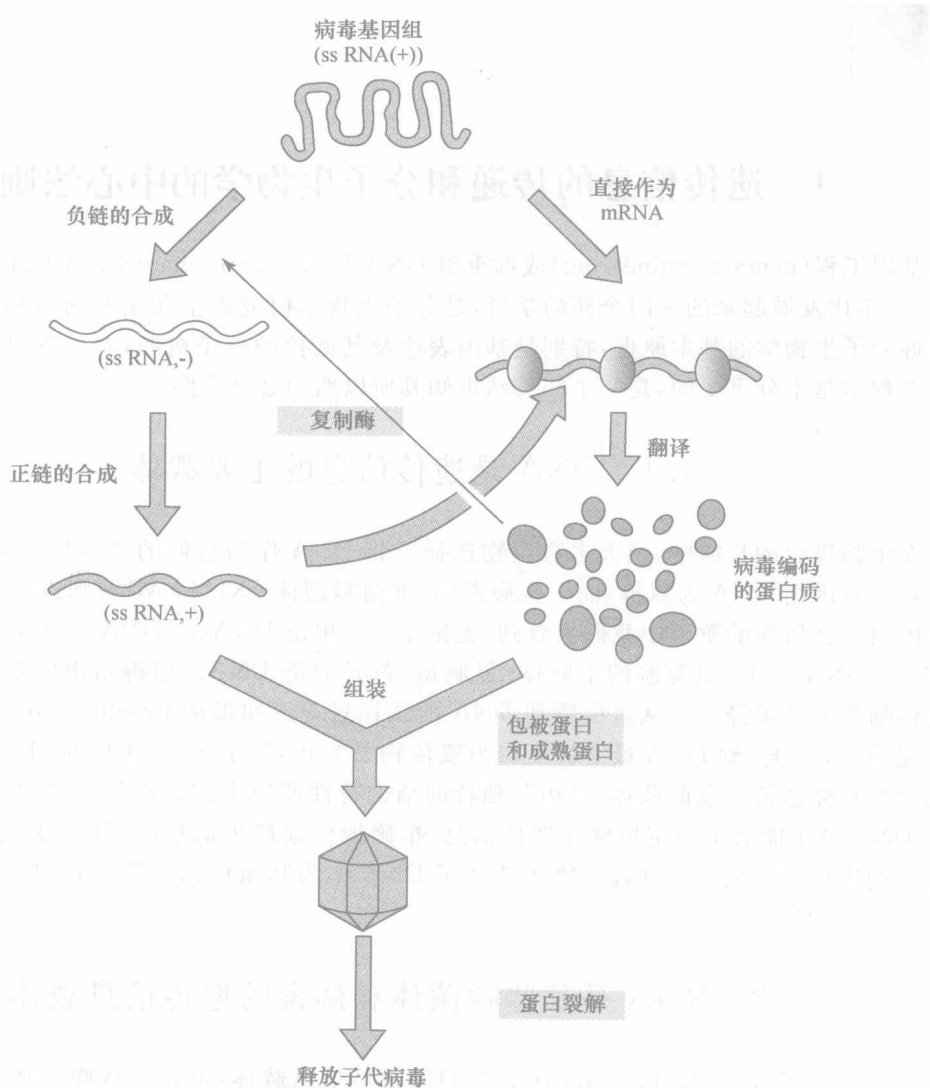


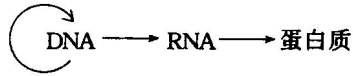
图 1-1 正链 RNA 病毒在增殖过程中所发生的事件

图中表明 ss RNA(+)可直接作为 mRNA 进行蛋白质合成。其翻译产物中的 RNA 复制酶将 ss RNA(+)“转录”成 ss RNA(-),进而以此为模板合成 ss RNA(+),而成熟蛋白和包被蛋白与 ss RNA(+)基因组一起组装成新的病毒颗粒。(Murray P, et al, 1998)

值得指出的是,所有 RNA 病毒都有着很高的突变速率,这是因为它们缺少像 DNA 聚合酶那样能够发现和修复错误的 RNA 聚合酶,所以不能对受损的遗传物质进行修复。这也说明为什么在长期的进化过程中,绝大多数生物体选择 DNA 作为其遗传信息载体的原因。

### 1.3 RNA 反转录酶的发现改变了对遗传信息单向传递的认识

1956 年 Crick 提出遗传信息传递的中心法则,也称分子生物学中心法则(central dogma):



按照这个法则, DNA 分子以自身为模板进行复制, 并通过 RNA 分子将遗传信息传递给蛋白质分子。DNA 分子以自身为模板进行自我拷贝的过程叫做复制(replication); 利用 DNA 分子为模板合成 RNA 分子的过程叫转录(transcription); 以 RNA 分子为模板合成蛋白质分子的过程叫翻译(translation)。复制、转录、翻译作为分子生物学中心法则中的三个关键步骤, 一直是分子生物学的核心问题。

遗传信息的这种单向不可逆传递的方式, 由于 RNA 反转录酶(reverse transcriptase)的发现而发生改变。1970 年 Temin 和 Baltimore 在 RNA 肿瘤病毒中发现 RNA 反转录酶。这类 RNA 病毒就是 1.2 中提到的反转录病毒。与其他单链 RNA 病毒不同, 当它侵染细胞时, 病毒 RNA 分子通过其编码的反转录酶将病毒 RNA 分子转换成与之互补的 DNA 链, 进而拷贝成双链的病毒 DNA 分子并整合到细胞染色体 DNA 中, 这样, 病毒 DNA 可随细胞染色体 DNA 的复制一代代地遗传下去。储存在这个 DNA 中的信息指导产生新一代的反转录病毒(图 1-2)。这种以 RNA 为模板在反转录酶作用下合成互补 DNA 分子的过程叫做反转录(reverse transcription)。RNA 反转录酶又叫做依赖于 RNA 的 DNA 聚合酶(RNA-dependent DNA polymerase), 它可以 RNA 为模板将储存在 RNA 病毒中的遗传信息传给 DNA 分子。为此, Crick 于 1971 年对他所提出的遗传信息单向不可逆传递方式的中心法则作了补充, 提出了中心法则的三角形表示法(图 1-3)。中心法则的三角形表示法进一步完善了生物体遗传信息传递方向的描述, 它不但包括了反转录, 而且对以 RNA 作为遗传信息载体的复制过程(1.2)也

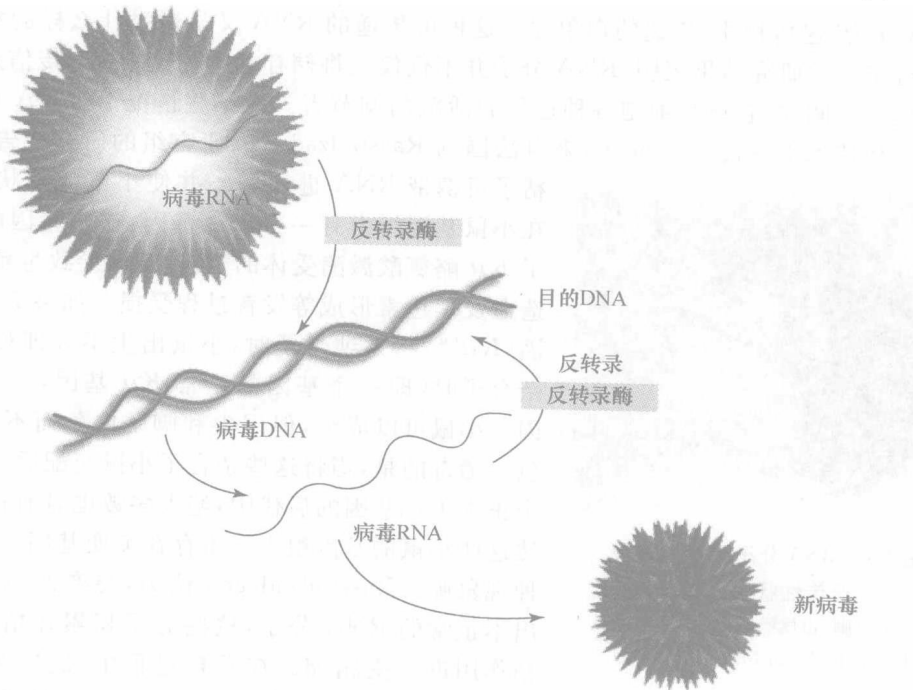


图 1-2 示反转录病毒和反转录酶的作用

<http://www.molecularstation.com/molecular-biology-images/506-molecular-biology-pictures/>

在图中标出。应该指出的是,这种表示法是从蛋白质表达的角度来考虑的。从这点出发,在生物有机体中从 DNA → RNA → 蛋白质是遗传信息传递的主要方向。这也正是基因工程得以进行的基本依据。

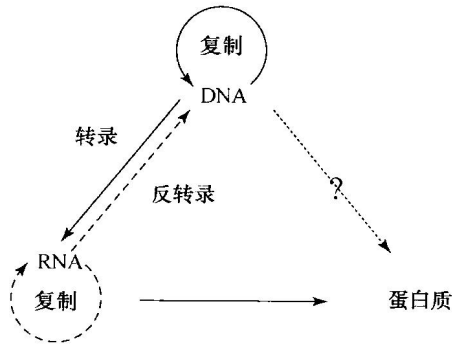


图 1-3 中心法则的三角形表示法

细胞中正常条件下进行的过程用实线表示;仅在 RNA 病毒感染时才发生的过程用虚线表示;从 DNA → 蛋白质并未发现确切证据。

## 1.4 在高等生物中 RNA 作为信息的载体可从亲代传给子代

在 1.2 中我们讲到细菌和动、植物的某些病毒以 RNA 作为其遗传信息的载体,在宿主细胞内可以自我复制并以 RNA(+)链为模板进行蛋白质的合成。那么,在高等生物中是否存在 RNA 作为“遗传物质”传递的现象呢?这种可传递的 RNA 又是通过什么样的机制起作用的呢?近年来的研究表明, RNA 分子并不仅仅是将储存在 DNA 中的遗传信息被动地传给蛋白质的中间体,它还是细胞各种过程的活跃的调节者。在某些情况下, RNA 可作为信息载体从一代传到下一代。2006 年,来自法国的 Rassoulzadegan 研究组的研究报告指出,小鼠

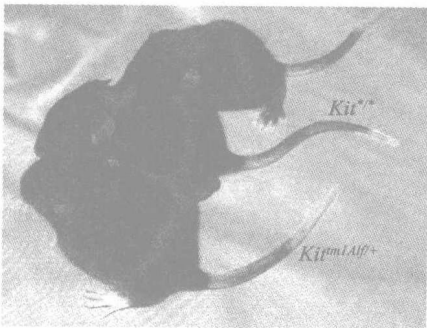


图 1-4 RNA 分子介导的表现遗传性状的变化

示白脚、白尾类小鼠的产生。

(Rassoulzadegan M, et al, 2006)

精子可携带 RNA 进入卵子并使小鼠的性状发生改变:在小鼠基因组中有一个 *Kit* 基因,这个基因的突变破坏了 *Kit* 酪氨酸激酶受体的合成,从而导致生殖细胞分化、造血及黑色素形成等发育过程受损。如果突变的 *Kit* 基因(*Kit<sup>m1 Aif</sup>*)是纯合子时,小鼠出生不久即死亡;如果是杂合子时(即一个基因是正常 *Kit* 基因,一个是突变基因),小鼠可以成活,但尾尖和脚是白色而不是正常的灰色。稀奇的是,当将这些杂合子小鼠交配后,在遗传有两个正常 *Kit* 基因的后代中,绝大多数也具有白色尾尖,虽然这些小鼠的基因组中已不存在突变基因。如何解释这种现象呢? Rassoulzadegan 认为,突变的 *Kit* 基因制造出不正常的 RNA 分子,这些分子积累在精子中并随受精作用进入受精卵。在发育过程中,这些突变的 RNA

分子使子代的正常的 *Kit* 基因沉默,也就是说改变了正常 *Kit* 基因的表达。她的这一看法被以下实验所证实:当将这些突变的 RNA 注射到处在一个细胞期的正常小鼠胚胎后,50%的小鼠后代也具有白脚和白尾尖的表现型(图 1-4)。这一发现至少指出:携带有特定“遗传信息”的

RNA 也可以在高等生物中传代;从目前的研究结果看,这类 RNA 参与了特定基因的表达调控。这些研究结果为进一步了解 RNA 在表观遗传(epigenetic inheritance)中所起的重要作用开辟了一条新路。所谓表观遗传,是指 DNA 序列不发生变化但基因表达却发生了可遗传的改变。这种改变是细胞内 DNA 序列以外的其他遗传物质发生的改变,且在发育和细胞增殖过程中通过减数分裂能稳定传递。应该说,这也提示我们应给予在真核染色质中存在的 RNA 成分的功能研究以足够的重视。

## 1.5 多肽链如何折叠成为功能蛋白质仍然是一个没有解决的问题

中心法则告诉我们,一个蛋白质分子中的氨基酸序列是由为其编码的核苷酸序列(遗传密码)所决定,通过翻译过程,只是将储存在 mRNA 中的遗传信息转换成由特定氨基酸序列所组成的多肽链,并没有涉及多肽链如何折叠成为具有特定三维结构的功能蛋白质的问题。然而,蛋白质的正确三维结构的形成是产生具有完全生物活性的蛋白质所必需的。20 世纪 60~70 年代,Anfinsen 根据他对核糖核酸酶在变性和还原后可以自发地氧化折叠恢复天然结构和全部生物活性的实验结果,提出蛋白质中一定的氨基酸序列决定其一定的空间结构,即蛋白质的一级结构决定其三级结构的假说。这一假说虽然已被科学界广泛接受,但蛋白质的一级结构(即多肽链中的氨基酸序列)如何决定蛋白质的三维空间结构的问题仍然是分子生物学研究中的大难题。20 世纪 80 年代中期,有一些科学家提出关于存在所谓折叠密码的假设,认为折叠密码决定了存在于多肽链氨基酸序列中的一级结构信息向蛋白质特定的三维空间结构的转变。然而现在看来,“折叠密码”应是相当复杂,绝不会像遗传密码那样用一种简单的关系,即三个核苷酸决定一个氨基酸所能概括。现在的研究指出,一个新合成的多肽链转化成完全折叠的蛋白质的方式既依赖于氨基酸序列内在的特性,又依赖于来自拥挤细胞环境的多种因素(如分子伴侣、折叠酶、各种蛋白因子和膜结构等)的作用,是一个非常复杂的过程。在细胞内蛋白质的折叠和去折叠是调控其生物活性和蛋白质定位于不同的细胞部位的关键途径。可以预见,对蛋白质折叠机制研究的突破,必然对蛋白质的结构形成和功能表达、全新功能蛋白质的设计和合成,以及重组蛋白质的复性和正确折叠等理论和实际应用的研究带来革命性的变化。所以,一个完整的分子生物学中心法则应该包括从多肽链到蛋白质的折叠这一步(图1-5)。

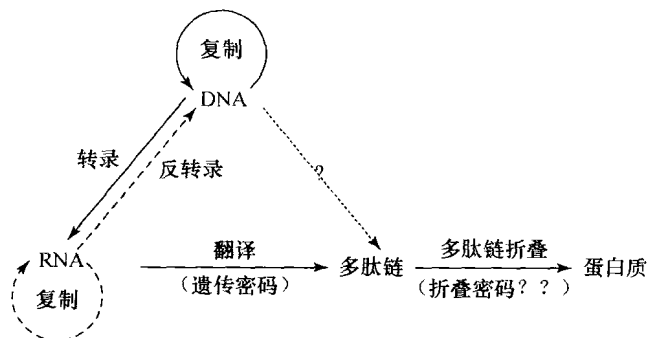
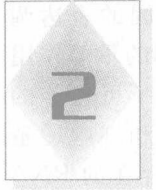


图 1-5 中心法则的另一种表述形式



## 2 DNA 的复制

如前所述, DNA 是生物体遗传信息的主要载体, 它包含了生物有机体生长发育、繁殖遗传所需的全部信息。DNA 在生命活动过程中准确有序的复制是生物体得以稳定延续的分子基础。

### 2.1 DNA 结构的特征

#### 2.1.1 DNA 是四种脱氧核糖核苷酸的多聚体

DNA 分子由四种脱氧核糖核苷酸聚合而成, 每个核苷酸由一个含氮的碱基、脱氧核糖和一个磷酸基团组成。这四种脱氧核苷酸是: 脱氧腺嘌呤核苷 5'-单磷酸(deoxy adenosine 5'-monophosphate, dAMP), 脱氧鸟嘌呤核苷 5'-单磷酸(deoxy guanosine 5'-monophosphate, dGMP), 脱氧胞嘧啶核苷 5'-单磷酸(deoxy cytosine 5'-monophosphate, dCMP)和脱氧胸腺嘧啶核苷 5'-单磷酸(deoxy thymidine 5'-monophosphate, dTMP)。脱氧核糖的 C-1 分别与嘌呤碱基的 N-9 和嘧啶碱基的 N-1 相连, 而磷酸基团与脱氧核糖的 C-5 相连。成千上万个脱氧核糖核苷酸单体通过 3', 5'-磷酸二酯键相连形成一个线性的高分子聚合物, 如 *E. coli* 的基因组 DNA (一条链) 就由  $4.6 \times 10^6$  个脱氧核糖核苷酸组成。DNA 的初级结构是由交替的糖-磷酸基团构成的骨架组成, 而嘌呤和嘧啶碱基形成侧链。图 2-1 示出由 dCMP-dGMP-dAMP-dAMP-

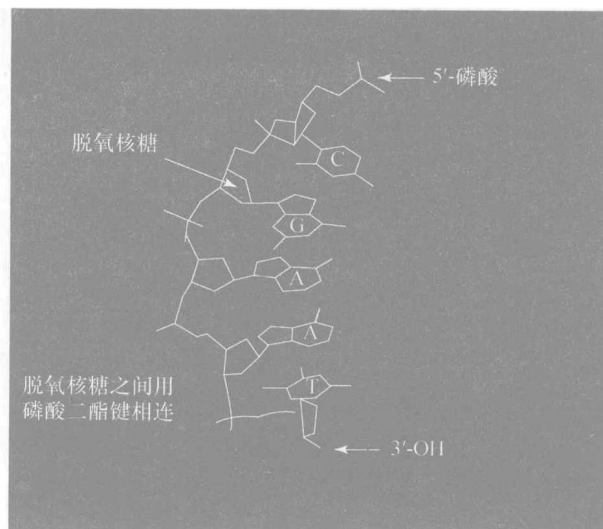


图 2-1 5'-d(CGAAT)五核苷酸形成的 DNA 主链骨架的示意图

dTMP五个核苷酸组成的五核苷酸主链。从图中可以看出：交替的脱氧核糖和磷酸二酯基团形成骨架主链；核苷酸链的方向从上到下是以  $5' \rightarrow 3'$  方向表示，写成  $5'-P-dCGAAT-OH-3'$  或简称为  $5'-d(CG AAT)$ ，即  $5'-P$  末端写在左边， $3'-OH$  写在右边；磷酸的氧原子带负电荷；A、G、C、T 碱基伸出链外，通过疏水相互作用彼此堆积。

### 2.1.2 双螺旋是 DNA 的基本结构

除了少数病毒 DNA 外，绝大多数生物有机体的 DNA 是以双链形式存在。即使单链的病毒 DNA，其复制型(replication form)也是双链 DNA。两条 DNA 链围绕着分子主轴相互盘绕形成一个双螺旋结构(double stranded helix)，双螺旋结构作为 DNA 的二级结构有如下特性：

(1) 组成 DNA 双螺旋的两条链通过腺嘌呤(A)与胸腺嘧啶(T)之间和鸟嘌呤(G)与胞嘧啶(C)碱基之间的氢键和碱基间的堆积力(base stacking interactions)来维系。DNA 两条链之间的这种非共价键结合便于 DNA 复制和转录过程中两条链的分开。

(2) DNA 双螺旋两条链的碱基严格配对，即一条链上的 A 总是与另一条链上的 T 配对，而 G 总是与 C 相配对，这就是 Watson-Crick 配对规则。A-T 和 G-C 之间的碱基配对是由它们自身的立体形状和氢键的互补性(steric and hydrogen bonding factors)所决定的。这样的配对使得 A-T 之间形成两个氢键，而 G-C 之间形成三个氢键(图 2-2)。碱基配对是 DNA 结构的最重要特性之一，它保证了 DNA 两条链的碱基序列完全互补，赋予 DNA 具有自我编码的特性。DNA 双链的彼此互补为 DNA 自我复制的机制提供了结构基础。值得指出的是，DNA 链之间的碱基互补使双链 DNA 中 A 的数目总是等于 T，而 G 的数目总是等于 C。然而，对于不同来源的 DNA，其碱基组成是不同的，通常用含 G+C 的百分数，即  $((G+C)/全部碱基数) \times 100\%$  来表示。由于 G-C 之间是三个氢键，所以 G+C 含量高的 DNA 其热稳定性就高。

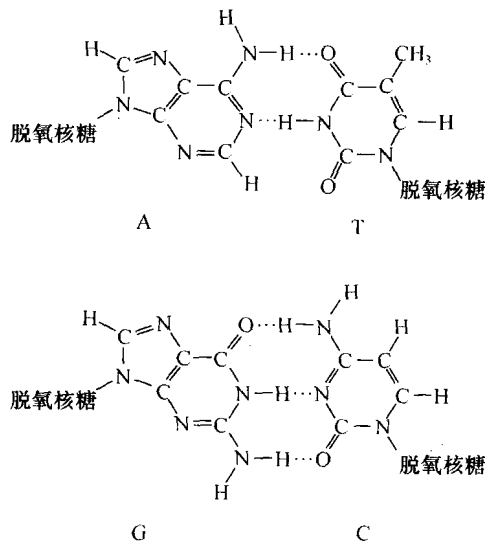


图 2-2 A-T 和 G-C 碱基配对及碱基间的氢键

图示 A-T 之间为两个氢键，G-C 之间为三个氢键(氢键用点线表示)。



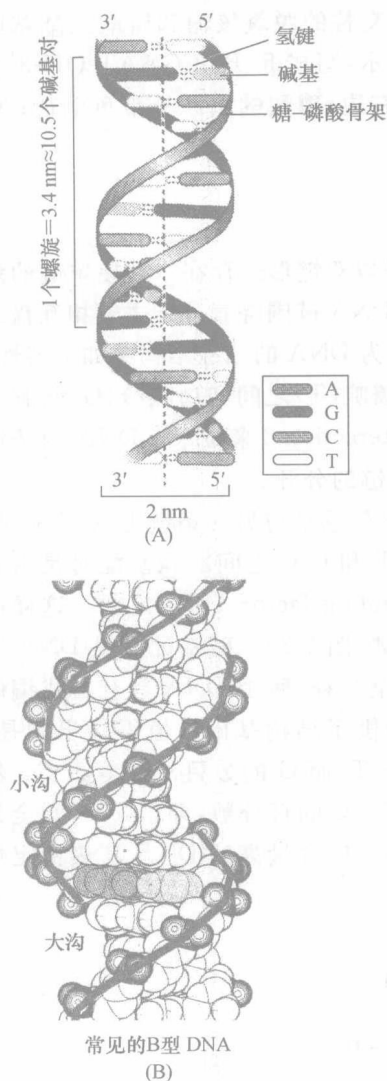


图 2-3 DNA 的双螺旋结构

(A) DNA 双螺旋结构的示意图, 示 DNA 双链的反平行走向。(B) B 型 DNA 的空间填充模型。B 型 DNA 是细胞中 DNA 常见的结构形式。图中显示每条 DNA 链中的糖和磷酸基团形成 DNA 双螺旋骨架(用深色球和曲线标出)并成螺旋缠绕。碱基朝向双螺旋内部, 处于大沟和小沟处的碱基可与细胞内其他分子(如蛋白质等)相接触并产生相互作用。

(引自 Watson J, et al, 2004; Lodish H, et al, 2000)

(3) 形成 DNA 双螺旋的两条互补链的走向正好相反。这是由于 DNA 一条链上 5' 端的碱基总是与其互补链上 3' 端的碱基配对, 这种碱基间互相配对的立体化学结果使得 DNA 的两条链呈反平行(antiparallel)的走向。DNA 互补链的反平行走向使得一条链是 5' → 3', 而另一条链则是 3' → 5' (图 2-3)。牢记住 DNA 的这一特点对后面我们要讲到的 DNA 复制、转录机制以及引物(primers)设计都是很重要的。

(4) 在 DNA 双螺旋结构中, 糖-磷酸主链沿着分子的外侧形成一个螺旋, 碱基在分子的内部形成一个螺旋。一般而言, 每个碱基对的两个碱基处在同一个平面上且与螺旋主轴垂直。碱基对平面之间的距离为 0.34 nm, 两个相邻碱基对间的夹角约为 36°, 双螺旋的直径约为 2.0 nm。这样, 螺旋的每一圈含有 10 个碱基对, 一个螺旋的长度为 3.4 nm (图 2-3)。在生物有机体中的 DNA 螺旋通常是右手螺旋, 也即双螺旋的两条链以右手螺旋的方式互相缠绕。值得指出的是, 碱基处于分子内部使其受到外界损伤的机会减小, 这正是 DNA 作为稳定的遗传信息载体所需要的。与碱基不同, 围绕在分子外围的糖-磷酸主链, 由于磷酸基团的存在而带负电荷。

(5) DNA 双螺旋所具有的大沟和小沟为其同蛋白质的相互作用提供了可及的位点。由于碱基对的空间几何结构使得 DNA 螺旋的主链在螺旋的一侧比在另一侧靠得更近些, 这样在主链靠得近的一侧就形成小沟(minor groove), 而在其对侧, 主链分开较远的一侧就形成大沟(major groove)。由于在 DNA 双螺旋的外部出现了大沟和小沟, 使每个碱基对的边缘在这些螺旋沟处暴露出来, 可允许蛋白质分子在无需解开继而破坏双螺旋结构的情况下, 识别碱基序列并参与 DNA 复制或转录的调控。绝大多数研究结果表明, DNA 结合蛋白与 DNA 分子内部碱基的相互作用发生在大沟侧, 但也不排除位于大沟的蛋白质与小沟的特定碱基产生相互作用的可能。值得指出的是, 蛋白质(或酶)与 DNA 的相互作用不仅仅可在螺旋沟处进行, 有时 DNA 中的单个碱基通过与酶蛋白的相互作用会从双螺旋中翻转出来, 人们将这种现象叫做碱基翻出(base flipping)。一些酶蛋白就是通过这种机制对碱基进行修饰或对受损碱基进行修复。图 2-4 显示甲基化酶将胞嘧啶碱基从螺旋中完全翻出, 使其进入酶的活性口袋而被甲基化; 然后, 这个碱基又回到螺旋中的正常位置。整个过程并不需要外来能源。