

生命科学名著

[英] T.A. 布朗 著  
袁建刚 等 译



# 基因组 3

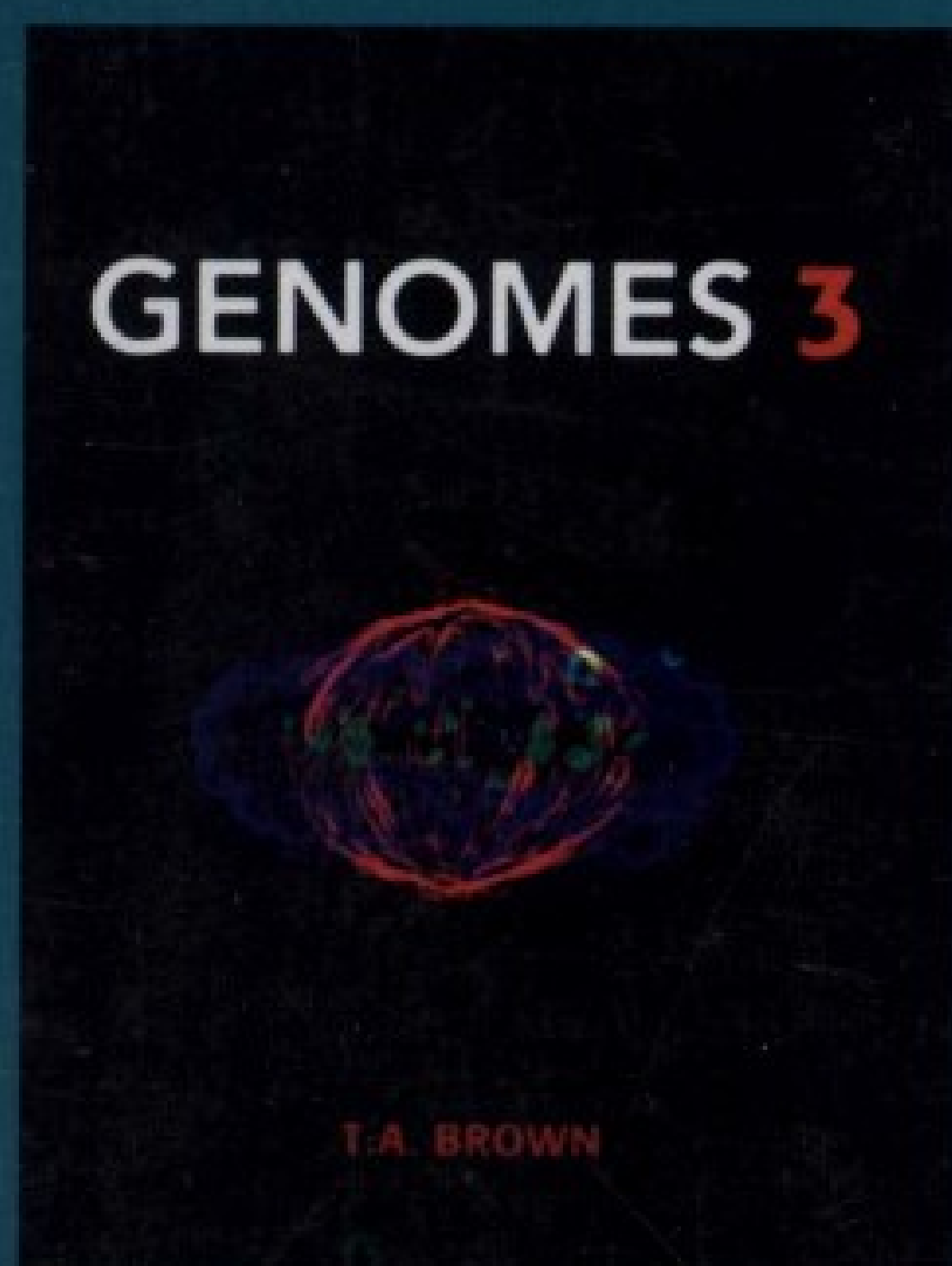
## GENOMES 3



科学出版社  
[www.sciencep.com](http://www.sciencep.com)







- 3章全新的内容，介绍对基因组的剖析以及最新的基因组测序成就。
- 含介绍基因组功能的新章节，描述了应用在转录组学及蛋白质组学研究中的最新技术。
- 含对重组的分子基础进行更深入、全面介绍的新章节。
- 最新设计及更加丰富的节后练习题，包含选择题、简答题、论述题及图形测试题等多种形式。
- 超过550幅全新设计的插图。为读者理解复杂概念提供了最有力的帮助。
- 随书附赠彩图光盘。

*“Terry Brown’s Genomes 3...everything from the index to page numbers to colour figures is fully accessible. Certainly, for any biology student interested in molecular genetics, Brown cuts mustard.”*

—THE TIMES HIGHER EDUCATION SUPPLEMENT



生物分社  
联系电话: 010-64012501  
<http://www.lifescience.com.cn>  
e-mail: lifescience@mail.sciencep.com



销售分类建议: 生物/分子生物学

定价: 98.00 元(含光盘)

# 生命科学名著

本书是生命科学名著丛书之一，由美国著名遗传学家、细胞生物学家、诺贝尔奖金获得者 T. A. 布朗著，袁建刚等译。本书主要介绍基因组学的发展，包括基因组的概念、基因组的结构、基因组的复制、基因组的表达、基因组的变异、基因组的进化、基因组的比较、基因组的分析、基因组的工程、基因组的医学、基因组的农业、基因组的工业、基因组的能源、基因组的环保、基因组的伦理、基因组的法律、基因组的政策、基因组的未来。本书内容丰富，图文并茂，是生命科学领域的一本重要著作。

〔英〕T. A. 布朗 著  
袁建刚 等 译

科学出版社

科学出版社

(总发行) 元 10.00

北京 北京 北京 北京 北京 北京 北京 北京 北京 北京





图字:01-2007-1854 号

## 内 容 简 介

《基因组 3》在前两版的基础上对原有内容进行了大量的更新和扩充,并对部分章节和内容进行了重排,使背景资料更充实,层次更清晰,行文更流畅。本书共包含四大部分内容,分别为研究基因组、基因组结构、基因组功能和基因组的复制及进化。本书以清新而简明的写作风格将基因组学的概念、观点和内容与传统的基因分子生物学和分子遗传学研究方法相结合,为基因组作为生命蓝图所起的作用提供了全新的视角。

本书内容翔实,深入浅出,引人入胜,根据内容需要采用大量图表,形象而简洁,是一部适合作为教材的基因组学读物。本书非常适合作为生物、医学、农学、林学等相关学科本科生和研究生的基因组学课程教材,也可供专业科技人员阅读。

T. A. Brown

Genomes 3

Copyright ©2007 by Garland Science Publishing.

All rights reserved. Authorized translation from English language edition published by Garland Science, a member of Taylor & Francis Group.

本书中彩色插图见随书附赠光盘。

## 图书在版编目(CIP)数据

基因组 3/(英)布朗(Brown, T. A.)著;袁建刚等译. —北京:科学出版社, 2009

(生命科学名著)

ISBN 978-7-03-023347-9

I. 基… II. ①布…②袁… III. 基因组-教材 IV. Q343.2

中国版本图书馆 CIP 数据核字(2008)第 172809 号

责任编辑:李 悦 彭克里 刘 晶 陈珊珊/责任校对:包志虹

责任印制:钱玉芬/封面设计:陈 敬

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>

新蕾印刷厂印刷

科学出版社发行 各地新华书店经销

\*

2009 年 3 月第 一 版 开本:787×1092 1/16

2009 年 3 月第一次印刷 印张:43 1/4

印数:1—4 000 字数:988 000

定价:98.00 元(含光盘)

(如有印装质量问题,我社负责调换(环伟))



## 译者序



基因组，集生命之大成者，有人称之为生命的上帝。这个上帝的手中掌握着人类生老病死和所有生命活动的奥秘。20 世纪中后叶以来，科技的进步使探讨基因组奥秘成为可能。因此，上帝之手终于一点点松开了，我们终于开始从中窥视到绚丽多彩的生命之源。T. A. 布朗在总结了基因组研究主要成果的基础上，于 1999 年编写了《基因组》第一版；为了反映基因组研究的最新进展，又于 2002 年出版了《基因组 2》。

作为一部优秀的分子生物学和生命科学的教材，它为我们提供了独特的思路、崭新的知识、新颖的风格和灵活多样、生动活泼的表达方式。本书既是一部教科书，又像是一部专业词典和技术工具书，兼容并包，颇具美感和动感效果。

选择“基因组”作为一个命题撰写一部教科书，既是科学发展之大势所趋，又是一项艰巨的任务。当今，在大学、研究院（所）和企业里，涉及生命科学的研究、开发和应用，大多受到基因组科学的影响。这是因为基因组科学给我们带来了观念上和技术方法上的飞跃。在观念上，使我们把基因组的活动和功能作为一个整体来看待，它由个别基因的活动和功能的网络作用综合而成，这更加接近生命活动的真实情况。在技术方法上，为了达到综合和集成，产生大量数据和信息，发展了高通量技术平台。因此，发展基因组科学是功在当代利在千秋的大业。但是，毕竟基因组科学属于宝塔尖里的科学，太前沿，以致大学分子生物学和生命科学类教科书极少涉及。在以往此类教科书中，即使含有有关基因的描述，也都是针对单个基因的，基本处于零敲碎打、坐井观天状态。因此，欲在经典的分子生物学领域里，围绕基因组这个核心把多种知识整合起来，并且做到条理化、系统化，实属不易。但是，我们看到，《基因组 3》一书做到了。

本书为生命科学和分子生物学提供了一个全新的视角，避免了沿袭旧教材的套路，独树一帜，挑选了基因组科学中最基本和最受关注的四个方面作为主要内容，它们是：研究基因组、基因组结构、基因组功能、基因组的复制与进化。这些内容在现有教科书中是找不到的，必须依靠作者从现刊中搜集，还需要丰富的知识和经验。正如原书作者在致谢中所言，书中素材来源于最近出版的 *Nature*、*Science*、*Trends in Genetics* 和 *Trends in Biochemical Sciences* 等著名杂志，由此可见其内容之新、之前沿。



新字当头，注重系统性及与传统知识的衔接，是该书的另一特色。作者深知，作为一本教科书，新旧知识和概念恰到好处的衔接和系统化，是其性质所决定的，也是一本教科书成败至关重要的标志。读者只要浏览一下目录，对全书各部分要介绍的内容即有一个系统完整的印象，求知欲油然而生。

本书的另一个独特之处是表达方式生动、活泼、多样、图文并茂。全书每个篇章除正文外，还包括学习要点、图例、技术注解、问答题、难题和图形测试、拓展阅读、词汇表等，它们单独归类编排，内容翔实、深入浅出、实用，引人入胜，与正文光彩互映。

本书“图例”堪称一绝，不仅简要明了，而且好临摹、易理解、有助记忆，真可谓一图可抵千言。

非常感谢原书作者为我们奉献了这么优秀的教科书。为了促进我国高等院校有关基因组科学的的教学和研究工作向前发展，在科学出版社的大力支持下，我们翻译出版了《基因组 3》一书。参加本书翻译的人员都是在科研第一线工作的科研工作者和博士研究生，他们承担着国家多项重大科研任务，工作十分繁忙，为了此书中译版的顺利出版，全体参译人员都不惜利用节假日和业余时间勤奋工作。然而，由于此书涉及的领域十分广泛，内容非常丰富，新概念、新技术贯穿全书，而参译人员的知识背景又有一定的局限性，虽然我们尽了最大努力争取中译版尽量忠于原著，但也认识到译文的不准确或不贴切之处还是难以完全避免。因此，衷心希望广大读者对本书的谬误之处多提宝贵意见，以便重印时使此书更加完善。

译者

2008年11月



## 第三版前言



从《基因组 2》的出版到现在，四个激动人心的年头已然过去。

不时有新完成的人类染色体序列被公布，黑猩猩基因组的测序也已经完成。真核生物中部分或全部序列已明确的物种数目在迅猛增加，而且新的原核序列每周都在实时地更新着。转录组和蛋白质组的实验技术为研究基因组表达提供了新的视点，同时系统生物学这一新学科正在把基因组研究与细胞生物化学联系在一起。所有这些进展都已经被加入到这本《基因组 3》之中。特别是以前单独一章的“基因组结构”被分解成了三个章节，而且，通过撰写如何分析序列及如何研究转录组及蛋白质组的独立章节，本书极大地丰富了后基因组学方面的材料。我也抓住时机，更加深入地描述了基因组的表达、复制和重组。

上述变化已经使《基因组》一书篇幅增加不少，作为补偿，我努力令这本书更便于使用。现在，“技术注解”只用于描述技术。这样，正文作为一个整体就不那么零散，从而有了更加连续的叙述。插图已经彻底重新设计，图表变得更加明了，更加引人入胜。阅读条目和章节结尾处的习题也已进行了相应全面的再评估。

在做这些修改的过程中，我已经考虑到一些来自世界各地的教师和学生们的反馈。这些人“不胜枚举”，我只好向大家一起道声谢谢。有一个我要单独感谢的人，他就是曼彻斯特大学的 Daniela Delneri，他关于后基因组以及分子进化章节的注释是如此全面，以致我自己几乎不需要在这些领域再做任何调研了。我还非常感谢纽约州立大学 Fredonia 校区生物系的 Ted Lee，他承担了令人畏惧的（至少对我是如此）任务，为书中每一章撰写综合问答及习题，这些学习辅助材料极大地提升了本书的价值。我还要感谢 Garland Science 出版社的 Dominic Holdsworth 和 Jackie Harbor，他们在我准备《基因组 3》的过程中给予了巨大的支持。感谢 Matthew McClements 重新设计了精彩插图。在第一版中我就写过“如果这本书对你有帮助，那就请你谢谢 Keri，而不是谢谢我，因为是她促成了本书的完成”，我很高兴能有这样的人促使我致力于本书的编写。

T. A. 布朗  
于曼彻斯特



## 第二版前言



《基因组》问世后又过去了三个鼓舞人心的年头，果蝇、拟南芥和人类的基因组草图绘制工作已经完成，现在原核生物基因组序列正以平均每个月两种或三种的速度得到公布。研究转录组和蛋白质组的实验方法正在逐步成熟，它们给基因组表达的研究带来了全新的视野。在此同时，基因组表达和复制的过程得到了更为详尽的阐释。所有这些进展都纳入到了《基因组 2》。本书第 1 章的中心是人类基因组，随后在第 1 篇的后续章节中笼统概括了基因组的物理和遗传结构，并以对转录组学和蛋白质组学的展望作为结束。在第 2 篇，即研究基因组的方法学这一部分中，补充了关于克隆和 PCR 技术的全新的一章，这些内容在第一版中仅仅分散插述，并不令人十分满意。同时对测序和功能研究的章节进行了更新，力求反映 1999 年以来的技术进展。对基因组表达的第 3 篇，以及有关基因组复制和进化的第 4 篇也进行了全面更新。许多读者称赞第一版《基因组》与时俱进，我希望第二版能够继续保持这一优点。

为了使这本书更便于读者阅读，我们做了相应的一些更改。第 1 篇的内容进行重新编排后，给那些初次接触分子生物学的学生一个更加深入浅出的介绍。每一章节的后面都补充了一系列的辅助内容，希望这些内容不但能够引导读者阅读本书，而且是一个直接的补充辅导资料。在每一章的起始部分我都列出了学习纲要，这是近年来最实用的教学改革成果，目前是英国许多大学教学改革质量测评中的必需项目。

衷心感谢所有热心批评和对《基因组 2》提出建议的人们，你们在阅读的过程中会发现，作为对你们的意见和建议的反馈，新的版本中做出了大大小小的修改。我非常感谢 BIOS 的 Jonathan Ray 和 Simon Watkins 在我编写本书的过程中所给予的全力支持，同时我衷心感谢 Sarah Carlson 和 Helen Barham，正是他们的通力协作，才使得这本书的面世过程不令人感到压力。最后，我要感谢我的妻子 Keri，没有她的支持，就不会有本书。在第一版《基因组》中我写道：“如果这本书对你有帮助，那就请你谢谢 Keri，而不是谢谢我，因为是她促成了本书的完成”，我很高兴能有这样的人促使我致力于本书的编写。

T. A. 布朗  
于曼彻斯特



## 第一版前言



《基因组》这本书希望以一种全新的方式来进行分子生物学本科生教学。它的编写鉴于这样一种理念：大学分子生物学教学大纲必须反映新千年的主要研究方向，而不应再是那些 20 世纪 70 年代或者 80 年代流行过的主题。因此这本书重在讨论基因组而非基因，因为我们认识到今日的分子生物学所取得的进展更多地来自于基因组测序和功能研究而非单个基因的功能分析。绝大多数在校分子生物学本科生毕业后将从事基因组方面的研究，所有这些人都会发现，他们的工作将会或多或少地受到基因组研究的影响。如果对大学生授课的目的是为他们将来的研究生涯作准备的话，那么就应该向他们教授基因组方面的知识！

当然如果就此断定单个基因研究不再重要，这将是非常愚蠢的做法。在编著《基因组》的过程中我所面临的主要挑战是如何将传统分子生物学教学大纲与基因组学的新进展有机结合起来。当然，如果试图完全用“从基因组到蛋白质组”的术语来笼统描述从 DNA 到蛋白质的过程，这是行不通的，因此《基因组》中相当一部分内容仍然着重于单个基因的表达途径。这本书的与众不同之处在于，它将具体基因的表达途径包含在基因组的功能和活性这样一个整体的概念之中。同样，对 DNA 的复制、突变和重组的叙述也是讲述它们对整个基因组的影响，而不只是对基因复制和变化过程的影响。

分子生物学教学必须以基因组为中心，这一理念在我编写这本书的过程中不断得到巩固和强化，与传统的教学大纲相比，这种新的教学理念显得更加完美，信息量更大。许多过去对我来说属于边缘性的主题现在逐渐凸现出来，并显出其新的重要性。希望我能将在撰写本书过程中所感受到的兴奋之情传递给读者。

T. A. 布朗  
于曼彻斯特



## 内容介绍

为尽可能使《基因组 3》便于读者使用，本书包含了一些试图帮助读者并使本书成为有效的教学辅助资料。

### 本书的结构

本书分为四部分。

- **第 1 篇“研究基因组” (Studying Genomes)** 以导读性章节开篇，向读者介绍基因组、转录组和蛋白质组，然后转到方法学，重点介绍了那些在前基因组时代用于研究单个基因的克隆和 PCR 方法 (第 2 章)。随后按基因组计划的研究顺序介绍了专用于基因组的研究方法：构建遗传图谱和物理图谱的方法 (第 3 章)；DNA 测序方法学和组装基因组邻接序列的策略 (第 4 章)；确定基因组序列中所包含的基因和这些基因在细胞中的功能的方法 (第 5 章)；蛋白质组和转录组的研究 (第 6 章) 人类基因组计划是整个第一篇的主线，但并非唯一内容，与此同时我努力全面论及这一过程中已经使用的和正在使用的研究策略，以便有助于理解其他生物的基因组。
- **第 2 篇“基因组结构” (Genomes Anatomies)** 纵览了我们星球上所发现的不同类型的基因组的解析。第 7 章涵盖真核生物的核基因组，其首要的重点是人类基因组；第 8 章探讨原核生物及真核生物的细胞器基因组，后者被纳入是因为它们具有原核生物的起源；第 9 章描述病毒基因组和移动的遗传元件，这些被放在一起是因为一些移动的元件与病毒基因组有关。
- **第 3 篇“基因组如何行使功能” (How Genomes Function)** 涵盖了以前被不适当地描述为“从 DNA 到 RNA 到蛋白质”的内容。第 10 章着重介绍染色质结构如何影响基因组表达；第 11 章阐述原核和真核生物转录起始复合物的组装，其中详尽讨论了 DNA 结合蛋白，它们在基因组表达的起始阶段发挥极其重要的作用；第 12 章和第 13 章全面阐述转录组和蛋白质组合成；第 14 章讨论基因组活性的调节。因为许多不同的主题都与基因组活性的调节相关，因此，要想将第 14 章的篇幅掌握在一定的范围之内并非易事，我希望通过具体范例来讨论一般规律，从而达到简洁而全面的目的。
- **第 4 篇“基因组如何复制及进化” (How Genomes Replicate and Evolve)** 将 DNA 的复制、突变和重组与基因组随时间的逐渐进化联系起来。第 15 章到第 17 章描述了复制、突变、修复和重组相关的分子过程；第 18 章则讨论了这些过程对基因组结构及遗传内容的形成发生影响的可能途径；最后，第 19 章致力于阐述在推断 DNA 序列间进化关系时分子系统发生学日益增加的资料使用。



## 章节编排

### 学习要点

每一章的起始处都有一套学习要点，这些要点的词句都经过仔细推敲。它们不仅仅是该章的内容提要，而且指明了学生阅读该章节后应该掌握的知识水平和层次。因此，学习要点用简洁的词汇精确地陈述了学生所应该达到的预期学习目的：哪些内容他们应该能够独立阐述、描绘、讨论、解释和评价。这样做是为了使学生明白他们学习每一章的目标，从而确切无疑地了解他们是否已经满意地掌握了这些材料。

### 图例

好图可抵千言，糟糕的图让人晕头转向，而多余的图则让人心烦意乱。因此我尽力保证每一幅图都必不可少并且都能达到其目的，而不仅仅是用来将正文分段从而使之更加美观。同时我尽力使图示易于临摹，因为这在我看来更有助于学生学习。我一直不能理解那些把教科书的图示艺术化的做法，因为如果不能临摹，那样的图示也仅仅只是一张图而已，并不能帮助学生理解其所欲传递的信息。《基因组 3》中的插图尽可能清晰，简洁而不零乱。

### 技术注解

每一章的正文都有一系列的技术注解加以支持和补充，包含在独立的注释框中。每一个技术注解都是基因组研究中一个或一组重要技术的较完备的解释。技术注解被安排与正文相连一起阅读，都位于正文中首次提到该技术应用之处。

### 问答题、难题和图形测试

每个章节的末尾给出了四种不同类型的习题。

- **选择题** 涵盖了章节中关键点，测试学生对材料的基本理解。传统上人们有时候对选择题在正式评估中的价值有所保留，不过该题型作为复习辅助的价值是毫无疑问的：如果一个学生能够准确地回答出其中的每一个问题，那么他们基本上是对该章节的实质内容有了透彻的理解。
- **简答题** 要求一个 50~300 个字的回答，有个别题目要求给出一个标注的图或表格。问题涉及章节中全部的内容，提问方式相当直接，并且大多数问题可以简单地通过检索正文的相应部分来解答。学生可以使用简答题来系统地读完该章节，也可以挑选个别问题来评估他们对于特殊专题的回答能力。简答题可以用于闭卷测验。
- **论述题** 要求更加详细地回答。它们的性质和难度各异，最简单的仅比文献评述要求高一点点，这些问题的目的是从《基因组 3》达到的高度展开，使学生将他们的学识再提高几步。其他问题要求学生基于他们对书中材料的理解（也可能提供了相关专题的补充阅读物）来评估某一段陈述或假设。这部分问题在理想情况下将形成一些思想和批判意识。少数问题是很难的，在某些情况下所提的问题没有固定答案。



设计这部分问题的用意在于激励辩论与推测，这将拓展学生的知识并促使他们对自己的陈述严加思考。这些论述题学生可以单独解答或者作为一个小组讨论的起点。

- **图形测试** 类似于简答题，但是使用了选自相应章节中的插图作为练习的核心。该测试的价值在于它是一种把从阅读正文获得的确实信息与插图所显示的结构及过程联系在一起的方法。好图可抵千言，但其前提是要仔细地研读并充分地理解该图。图形测试有助于提供这种类型的理解。

## 拓展阅读

每章后面的阅读书目包括我认为非常有用的文献、综述和书籍等辅助材料。在《基因组 3》一书中，我自始至终的目的是当学生要针对特别专题写大型论文或研究报告的时候应该能够使用该阅读书目来获得进一步的信息。所以研究类型的文献也被列入，但仅包括那些内容可以被本书平均水平读者所理解的。重点放在了能获取的综述文献上，比如，*Science* 的展望栏目、*Nature* 的新闻与视点栏目以及 *Trends* 系列杂志中的文章。这些综合性文章的优点在于为某一方面的工作提供了历史背景及关联性。绝大多数的阅读书目根据章节中信息的结构进行了分组，在某些情况下我给每组的特殊价值附加了少量总结性的语句，以帮助读者来决定他们到底该搜寻哪些读物。该书目不是包罗万象的，我建议读者花一些时间研究一下他们自己图书馆中书架上的其他书籍和文献。浏览是发现你以前未曾发现的兴趣点的绝好方法。

## 词汇表

我非常喜欢词汇表作为学习的帮手，在《基因组 3》中我提供了一个更加丰富的词汇表。正文中的每一个黑体词都在词汇表中有定义，表中一些额外词条是读者在参考阅读书目中的书籍和文献时可能遇到的。词汇表中的每一个词也在索引列出，这样读者可以快速翻阅关于该词在相应页面内更详细的介绍。

## 《基因组 3》的影像材料

本书附带的光盘内含有书中的彩图文件。JPEG 图像已经被优化过，以使其适合于打印以及网络浏览。

## 给教师的建议

Garland Science Classwire™（Garland Science 出版社的网络课堂）的网址是 <http://www.classwire.com/garlandscience>，它为使用者提供了指导性材料以及课程管理的工具，其中包括《基因组 3》中 JPEG 和幻灯片形式的图像。那些在附录中没有答案或辅导的选择题、简答题、论述题和图形测试可用于布置家庭作业或当作测验题。如果需要，这些练习题的答案和辅导将通过 Classwire™ 提供给导师使用。采用《基因组 3》的教师能够从其他教材中获取额外的资源。Classwire™ 还是一种灵活易用的课程管理工具，它允许导师们为他们的班级建立网页，并具备如下特色工具：课程表编排工具、课程进度表、信息中心、课程计划工具、虚拟办公时间以及资源管理，不需要使用者具有编程以及技术方面的技巧。



## 审稿人名单

《基因组 3》的作者和出版商诚挚感谢下述审稿人在本版书稿出版中的贡献。

Dean Danner, Emory University School of Medicine  
Daniela Delneri, University of Manchester  
Yuri Dubrova, University of Leicester  
Bart Eggen, University of Groningen  
Robert Fowler, San Jose State University  
Adrian Hall, Sheffield Hallam University  
Glyn Jenkins, University of Aberystwyth  
Torsten Kristensen, University of Aarhus  
Mike McPherson, University of Leeds  
Andrew Read, University of Manchester  
Darcy Russell, Baker College  
Amal Shervington, University of Central Lancashire  
Robert Slater, University of Hertfordshire  
Klaas Swart, Wageningen University  
John Taylor, University of Newcastle  
Guido van den Ackerveken, Utrecht University  
Matthew Upton, University of Manchester  
Vassie Ware, Lehigh University



## 缩略语

$\mu\text{m}$	微米	cGMP	环 GMP
5-bU	5-溴尿嘧啶	CHEF	等高加压均匀电场
A	腺嘌呤; 丙氨酸	CJD	克雅 (氏) 病
ABF	ARS 结合因子	Col	大肠杆菌素
Ac/Ds	激活子/解离	CPSF	切割与多聚腺苷酸化特异性因子
ADAR	作用于 RNA 的腺嘌呤脱氨基酶	CRM	染色质重建机器
ADP	5'-二磷酸腺嘌呤	CstF	切割刺激因子
AIDS	获得性免疫缺陷综合征	CTAB	溴化十六烷基三甲基铵
Ala	丙氨酸	CTD	C 端
AMP	5'-单磷酸腺嘌呤	CTP	胞苷三磷酸
ANT-C	控制触角的基因复合体	Cys	半胱氨酸
AP	无嘌呤/无嘧啶	D	天冬氨酸
Arg	精氨酸	DAG	甘油二酯
ARMS	耐扩增突变系统	Dam	DNA 腺嘌呤甲基化酶
ARS	自主复制序列	DAPI	二脒基苯基吡啶
A cite	接受位点	DASH	动态等位基因特异性杂交
Asn	天冬酰胺	dATP	脱氧腺苷三磷酸
ASO	等位基因特异的寡核苷酸	DBS	双链 DNA 结合位点
Asp	天冬氨酸	Dcm	DNA 胞嘧啶甲基化酶
ATP	腺苷三磷酸	dCTP	脱氧胞苷三磷酸
ATPase	腺苷三磷酸酶	ddATP	双脱氧腺苷三磷酸
BAC	细菌人工染色体	ddCTP	双脱氧胞苷三磷酸
bis	双丙烯酰胺	ddGTP	双脱氧鸟苷三磷酸
BLAST	基本逻辑比对搜索工具	ddNTP	双脱氧核苷三磷酸
bp	碱基对	ddTTP	双脱氧胸苷三磷酸
BSE	牛海绵状脑病	Dfd	畸形的
BX-C	双胸复合体	dGTP	脱氧鸟苷三磷酸
C	半胱氨酸; 胞嘧啶	DMSO	二甲基亚砷
cAMP	环 AMP	DNA	脱氧核糖核酸
CAP	分解代谢物激活蛋白质	DNase	脱氧核糖核酸酶
CASP	CTD 相关的 SR 样蛋白质	Dnmt	DNA 甲基转移酶
cDNA	互补 DNA	dNTP	脱氧核苷三磷酸
CEPH	巴黎人类多态性中心	DPE	下游启动子元件



<b>DSB</b>	双链断裂	<b>H</b>	组氨酸
<b>DSP1</b>	背侧转换蛋白 1	<b>HAT</b>	次黄嘌呤、氨蝶呤与胸苷
<b>dsRAD</b>	双链 RNA 腺苷脱氨酶	<b>HBS</b>	异源双链结合位点
<b>dsRBD</b>	双链 RNA 结合结构域	<b>HDAC</b>	组蛋白去乙酰酶
<b>dTTP</b>	脱氧胸苷三磷酸	<b>His</b>	组氨酸
<b>E</b>	谷氨酸	<b>HIV</b>	人免疫缺陷病毒
<b>EDTA</b>	乙二胺四乙酸	<b>HLA</b>	人白细胞抗原
<b>eEF</b>	真核延伸因子	<b>HMG</b>	高迁移率族
<b>EEO</b>	电内渗值	<b>HNPCC</b>	遗传性非息肉性结肠直肠癌
<b>EF</b>	延伸因子	<b>hnRNA</b>	核异质 RNA
<b>eIF</b>	真核起始因子	<b>HOM-C</b>	同源复合体
<b>EMS</b>	乙基甲烷磺酸盐	<b>HPLC</b>	高效液相色谱
<b>eRF</b>	真核释放因子	<b>HPRT</b>	次黄嘌呤磷酸核糖转移酶
<b>ERV</b>	内源性逆转录病毒	<b>HTH</b>	螺旋-转角-螺旋
<b>ES</b>	胚胎干细胞	<b>I</b>	异亮氨酸
<b>ESE</b>	外显子剪接增强子	<b>ICAT</b>	同位素亲和标签
<b>E site</b>	退出位点	<b>ICF</b>	免疫缺陷性着丝粒不稳定性 面部异常综合征
<b>ESS</b>	外显子剪接沉默子	<b>IF</b>	起始因子
<b>EST</b>	表达序列标签	<b>Ig</b>	免疫球蛋白
<b>F</b>	生育力; 苯丙氨酸	<b>IHF</b>	整合宿主因子
<b>FEN</b>	侧翼内切核酸酶	<b>Ile</b>	异亮氨酸
<b>FIGE</b>	倒转电场凝胶电泳	<b>Inr</b>	起始物
<b>FISH</b>	荧光原位杂交	<b>Ins(1,4,5)P<sub>3</sub></b>	1, 4, 5-三磷酸肌醇
<b>FRAP</b>	荧光漂白恢复	<b>IPTG</b>	异丙基-β-D-硫代半乳糖苷
<b>G</b>	甘氨酸; 鸟嘌呤	<b>IRE-PCR</b>	分散重复序列 PCR
<b>G<sub>1</sub></b>	间期 1	<b>IRES</b>	内部核糖体进入位点
<b>G<sub>2</sub></b>	间期 2	<b>IS</b>	插入序列
<b>GABA</b>	氨基丁酸	<b>ITF</b>	整合宿主因子
<b>GAP</b>	GTP 酶活化蛋白	<b>ITR</b>	末端反向重复
<b>Gb</b>	十亿个碱基对	<b>JAK</b>	Janus 激酶
<b>GDP</b>	5'-二磷酸鸟嘌呤	<b>K</b>	赖氨酸
<b>GFP</b>	绿色荧光蛋白	<b>kb</b>	千碱基对
<b>Gln</b>	谷氨酰胺	<b>kcal</b>	千卡
<b>Glu</b>	谷氨酸	<b>kDa</b>	千道尔顿
<b>Gly</b>	甘氨酸	<b>L</b>	亮氨酸
<b>GMP</b>	5'-磷酸鸟苷	<b>LCR</b>	基因座控制区
<b>GMRP</b>	鸟嘌呤核苷酸释放蛋白	<b>Leu</b>	亮氨酸
<b>GTF</b>	一般转录因子	<b>LINE</b>	长散布核元件
<b>GTP</b>	鸟苷三磷酸		



lod	优势对数
LTR	长末端重复
Lys	赖氨酸
M	甲硫氨酸; 有丝分裂期
MALDI-TOF	基质辅助激光解吸电离飞行时间
MAP	促分裂原活化蛋白
MAR	基质相关区域
Mb	兆碱基对
MeCP	甲基化 CpG 结合蛋白
Met	甲硫氨酸
MGMT	O <sup>6</sup> -甲基鸟嘌呤-DNA 甲基转移酶
miRNA	微小 RNA
mol	摩尔
mRNA	信使 RNA
MudPIT	多维蛋白质鉴定技术
MULE	突变子样转座元件
Myr	百万年
N	天(门)冬酰胺
NAD	二氢尿嘧啶脱氢酶
NADH	还原型烟酰胺腺嘌呤二核苷酸
ng	纳克
NHEJ	非同源末端连接
NJ	相邻连接
nm	纳米
NMD	无义突变介导的 mRNA 降解
NMR	核磁共振
NTP	三磷酸核苷
OFAGE	正交变电场凝胶电泳
OLA	寡核苷酸连接试验
Omp	膜外蛋白质
ORC	复制起始识别复合物
ORF	可读框
OTU	分类操作单元
P	脯氨酸
PAC	噬菌体 P1 衍生的人工染色体
PADP	多聚腺苷酸结合蛋白
PAUP	利用简约分析进行系统发生

	分析的软件包
PCNA	增殖细胞核抗原
PCR	聚合酶链反应
pg	皮克
Phe	苯丙氨酸
PHYLIP	一个系统发育推断软件
PIC	前起始复合物
PNA	肽核苷酸
PNPase	多聚核苷酸磷酸化酶
Pro	脯氨酸
PSE	近端序列元件
PSI-BLAST	位置相关的迭代 BLAST
P site	肽基位点
PtdIns(4,5)P <sub>2</sub>	4, 5-二磷酸肌醇
PTRF	聚合酶 I 与转录释放因子
Pu	嘌呤
Py	嘧啶
Q	谷氨酰胺
R	精氨酸; 嘌呤
RACE	cDNA 末端快速扩增
RAM	随机存取存储器
RBS	核糖体结合序列
RC	复制复合物
RF	释放因子
RFC	复制因子 C
RFLP	限制性片段长度多态性
RHB	Rel 同源区域
RISC	RNA 诱导沉默复合物
RLF	复制许可因子 RNA 核糖核酸
RMP	复制介导蛋白
RNA	核糖核酸
RNAi	RNA 干扰
RNase	核糖核酸酶
RNP	核糖核蛋白
RPA	复制蛋白 A
RRF	核糖体循环因子
rRNA	核糖体 RNA
RT-PCR	逆转录聚合酶链反应
RTVL	逆转录病毒样元件



<b>S</b>	丝氨酸; 合成期	<b>TEMED</b>	四甲基乙二胺
<b>SAGE</b>	基因表达系列分析	<b>TF</b>	转录因子
<b>SAP</b>	应激激活蛋白	<b>TGF</b>	转化生长因子
<b>SAR</b>	支架附着区	<b>Thr</b>	苏氨酸
<b>SCAF</b>	SR 样 CTD 结合因子	<b>Ti</b>	肿瘤诱发
<b>scRNA</b>	小细胞质 RNA	<b>TIC</b>	TAF 与起始子依赖的辅助因子
<b>SCS</b>	特异染色质结构	<b>TK</b>	胸苷激酶
<b>SDS</b>	十二烷基磺酸钠	<b>T<sub>m</sub></b>	解链温度
<b>SeCys</b>	硒代半胱氨酸	<b>tmRNA</b>	转运信使 RNA
<b>Ser</b>	丝氨酸	<b>Tn</b>	转座子
<b>SINE</b>	短散布核元件	<b>TOL</b>	甲苯
<b>siRNA</b>	小干扰 RNA	<b>TPA</b>	组织型纤溶酶原激活剂
<b>SIV</b>	猴免疫缺陷病毒	<b>TRAP</b>	色氨酸 RNA 结合型衰减蛋白
<b>SL RNA</b>	剪接前导 RNA	<b>tRNA</b>	转运 RNA
<b>SMAD</b>	SMA/MAD 相关的	<b>Trp</b>	色氨酸
<b>snoRNA</b>	小核仁 RNA	<b>Try</b>	酪氨酸
<b>SNP</b>	单核苷酸多态性	<b>U</b>	尿嘧啶
<b>snRNA</b>	小核 RNA	<b>UCE</b>	上游控制元件
<b>snRNP</b>	小核核糖核蛋白	<b>UTP</b>	尿苷三磷酸
<b>SRF</b>	血清反应因子	<b>UTR</b>	非翻译区
<b>SSB</b>	单链结合蛋白	<b>UV</b>	紫外线
<b>SSLP</b>	简单重复序列长度多态性	<b>Val</b>	缬氨酸
<b>STAT</b>	信号转导与转录激活	<b>VNTR</b>	可变数目串联重复序列
<b>STR</b>	短串联重复	<b>W</b>	腺嘌呤或胸腺嘧啶; 色氨酸
<b>STS</b>	序列标签位点	<b>X-gal</b>	5-溴-4-氯-3-吲哚-β-D-乳糖苷
<b>T</b>	苏氨酸; 胸腺嘧啶	<b>Y</b>	嘧啶; 酪氨酸
<b>TAF</b>	TBP 相关因子	<b>YAC</b>	酵母人工染色体
<b>TAP</b>	串联亲和纯化	<b>YIp</b>	酵母整合型质粒
<b>TBP</b>	TATA 框结合蛋白		



# 目 录

译者序

第三版前言

第二版前言

第一版前言

内容介绍

缩略语

## 第 1 篇 研究基因组

第 1 章 基因组、转录组和蛋白质组	3
1.1 DNA	5
1.2 RNA 和转录组	14
1.3 蛋白质和蛋白质组	17
总结	24
第 2 章 研究 DNA	30
2.1 用于 DNA 操作的酶	33
2.2 DNA 克隆	44
2.3 聚合酶链反应 (PCR)	56
总结	59
第 3 章 基因组作图	66
3.1 遗传图谱和物理图谱	68
3.2 遗传作图	69
3.3 物理作图	88
总结	101
第 4 章 基因组测序	107
4.1 DNA 测序方法学	108
4.2 连续 DNA 序列的组装	116
4.3 人类基因组计划	125
总结	128
第 5 章 解读基因组序列	134
5.1 在基因组序列中定位基因	135
5.2 确定单个基因的功能	147
5.3 一个例研究: 标注酿酒酵母基因组序列	160
总结	163
第 6 章 理解基因组是如何行使功能的	169



6.1 转录组研究 .....	170
6.2 蛋白质组研究 .....	176
6.3 蛋白质组之外 .....	188
总结 .....	190

## 第2篇 基因组结构

第7章 真核生物核基因组 .....	199
7.1 核基因组包含于染色体当中 .....	200
7.2 真核生物核基因组的遗传特征 .....	206
总结 .....	221
第8章 原核生物基因组和真核生物细胞器基因组 .....	227
8.1 原核生物基因组的物理特征 .....	228
8.2 原核生物基因组的遗传学特征 .....	232
8.3 真核生物细胞器基因组 .....	239
总结 .....	243
第9章 病毒基因组和可移动的遗传元件 .....	249
9.1 噬菌体和真核生物病毒的基因组 .....	250
9.2 可移动的遗传元件 .....	257
总结 .....	264

## 第3篇 基因组如何行使功能

第10章 接近基因组 .....	273
10.1 细胞核内部 .....	274
10.2 染色质修饰和基因组表达 .....	281
10.3 DNA 修饰和基因组表达 .....	287
总结 .....	291
第11章 转录起始复合物的组装 .....	297
11.1 DNA 结合蛋白及其结合位点 .....	299
11.2 转录起始中 DNA-蛋白质的相互作用 .....	310
11.3 转录起始的调控 .....	317
总结 .....	326
第12章 RNA 的合成和加工 .....	333
12.1 细菌 RNA 的合成和加工 .....	334
12.2 真核细胞 RNA 的合成和加工 .....	346
总结 .....	373
第13章 蛋白质组的合成与加工 .....	381
13.1 tRNA 在蛋白质合成中的作用 .....	382
13.2 核糖体在蛋白质合成中的作用 .....	388
13.3 蛋白质翻译后加工 .....	401



13.4 蛋白质降解·····	409
总结·····	410
<b>第 14 章 基因组活性的调控</b> ·····	417
14.1 基因组活性的瞬时变化·····	419
14.2 基因组活性的永久性和半永久性变化·····	430
14.3 发育过程中基因组活性的调节·····	436
总结·····	448
 <b>第 4 篇 基因组如何复制及进化</b> 	
<b>第 15 章 基因组复制</b> ·····	457
15.1 拓扑学问题·····	458
15.2 复制过程·····	464
15.3 真核生物基因组复制的调控·····	481
总结·····	486
<b>第 16 章 突变和 DNA 修复</b> ·····	493
16.1 突变·····	494
16.2 DNA 修复 ·····	510
总结·····	519
<b>第 17 章 重组</b> ·····	526
17.1 同源重组·····	527
17.2 位点特异性重组·····	534
17.3 转座·····	536
总结·····	540
<b>第 18 章 基因组如何进化</b> ·····	546
18.1 基因组: 最初的 100 亿年·····	547
18.2 新基因的获得·····	551
18.3 非编码 DNA 与基因组进化 ·····	563
18.4 人类基因组: 最近的 500 万年·····	567
总结·····	568
<b>第 19 章 分子系统发生学</b> ·····	575
19.1 从分类学到分子系统发生学·····	576
19.2 基于 DNA 的系统发生树的重建 ·····	579
19.3 分子系统发生学的应用·····	589
总结·····	598
<b>附录</b> ·····	604
<b>词汇表</b> ·····	626
<b>索引</b> ·····	660



## 此为试读版本

试读版本有缺页，仅供参考，如需完整版，请至“爱读书”淘宝店购买。付款时请注明所需书籍名称作者等以及您的收件邮箱，最好是QQ邮箱。通常付款后12小时内完整版PDF书籍会发送至您的邮箱，最晚三天，请注意查收。

付费方式：使用支付宝购买读书卡付费。

付费标准：按照书籍页数收费，百页一元，不计零头不议价。即1-199页的书1元，200-299页的书两元，以此类推。请购买相应面值的读书卡。

爱读书淘宝店：

<http://shop58129009.taobao.com/>

如有问题请及时联系：

QQ: 896578748 电子邮箱: 896578748@qq.com

更多试读书籍请到新浪共享资料下载阅读，地址：

<http://iask.sina.com.cn/u/ish?uid=1706535084>



PART

# 第 1 篇

## 研究基因组

---

第 1 章 基因组、转录组和蛋白质组

第 2 章 研究 DNA

第 3 章 基因组作图

第 4 章 基因组测序

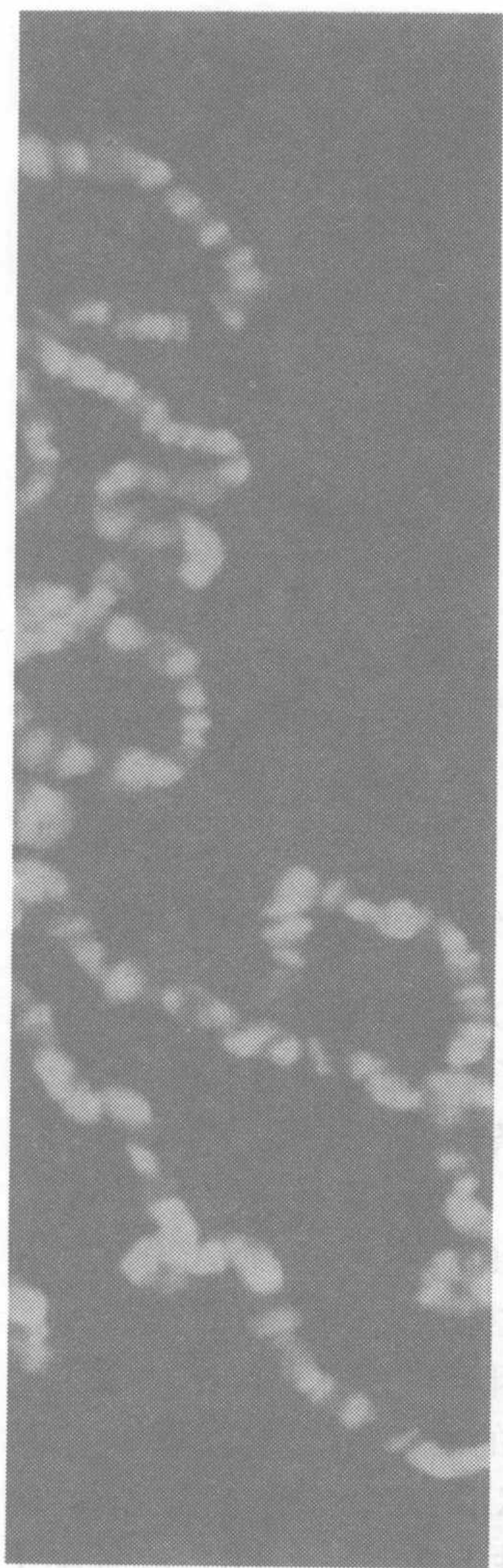
第 5 章 解读基因组序列

第 6 章 理解基因组是如何行使功能的









## 1.1 DNA

## 1.2 DNA 和转录组

## 1.3 蛋白质和蛋白质组

## 学习要点

当你阅读完第 1 章之后，应该能够：

- 为名词 “genome”、“transcriptome”、“proteome” 下定义并且阐明它们在基因组表达过程中是如何联系在一起的。
- 描述分子生物学家推断基因由 DNA 组成的两个实验，以及每一个实验的局限性。
- 详细描述多聚核苷酸的结构，总结 DNA 和 RNA 化学性质的差别。
- 讨论 Watson 和 Crick 演绎 DNA 双螺旋结构的依据，列举这一结构的关键特征。
- 区分编码 RNA 和功能性 RNA，并分别举出一些例子。
- 概述细胞中 RNA 是如何被合成与加工的。
- 详细描述蛋白质结构的不同层次，并解释为何氨基酸差异决定蛋白质差异。
- 描述遗传密码的关键特征。
- 解释为何一个蛋白质的功能依赖于它的氨基酸序列。
- 列举活体生物中蛋白质的各种作用并将其多样性与基因组功能联系起来。



众所周知，“生命”由地球上与我们共存的无数生物的**基因组**（genome）决定。每一种生物的基因组都包含着相应的**生物学信息**（biological information），这些信息是每个活生生的个体建立和维持其生物学特征所必需的。绝大多数基因组，包括人类和其他细胞生命形式的基因组，都由 **DNA**（脱氧核糖核酸）组成，但是，也有一些病毒基因组是由 **RNA**（核糖核酸）组成的。DNA 和 RNA 是由**核苷酸**（nucleotide）单体组成的多聚分子。

作为所有多细胞动物代表的人类基因组由两个独立的部分组成（图 1.1）。

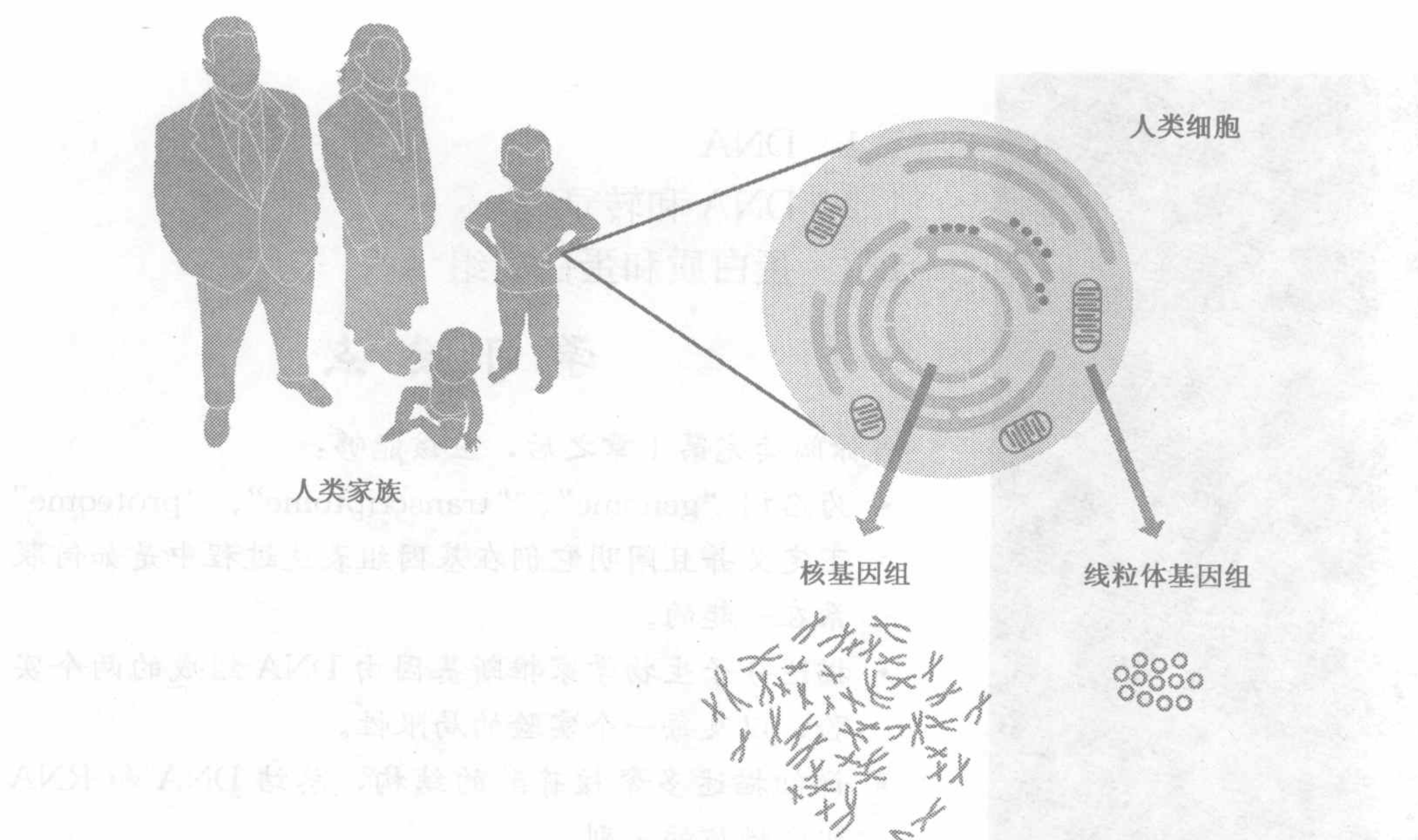


图 1.1 人类基因组的细胞核和线粒体成分

- **核基因组**（nuclear genome）由包含约 32 亿个核苷酸的 DNA 组成，分为 24 个线性分子，最短的 50Mb，最长的 260Mb，每一个分子包含在不同的染色体（chromosome）中。24 条染色体中有 22 条常染色体，2 条性染色体 X 和 Y。加在一起，在人类核基因组中约存在 35 000 个基因。
- **线粒体基因组**（mitochondrial genome）是一个长为 16 569bp 的环形 DNA 分子，在线粒体这个生产能量的细胞器中有许多拷贝。人类的线粒体基因组仅含 37 个基因。

成人体内约  $10^{13}$  个细胞中均含有单拷贝或多拷贝的基因组，只有少数细胞类型例外，如红细胞，在其完全分化状态下缺乏细胞核。绝大多数的细胞是**二倍体**（diploid），含有两个拷贝的常染色体和两条性染色体，雌性个体中是 XX，雄性个体中是 XY，总共 46 条染色体。这些细胞叫做**体细胞**（somatic cell），而**性细胞**（sex cell）或者**配子**（gamete）是**单倍体**（haploid），只有 23 条染色体，包括一套常染色体和一条性染色体。这两类细胞中都含有 8000 份拷贝的线粒体基因组，每个线粒体中有大约 10 个拷贝。

基因组是一个生物信息库，但是，仅仅靠其自身还不能将这些信息传递给细胞。基因组所包含的生物信息的利用需要酶及其他参与**基因组表达**（genome expression）过程中一系列复杂生化反应的蛋白质的协同活性（图 1.2）。基因组表达的最初产物是**转录组**



(transcriptome), 即那些含有细胞在特定时间所需生物信息、编码蛋白质的基因衍生而来的 RNA 分子的集合。转录组由称为**转录** (transcription) 的过程来维持, 在这个过程中各个基因被复制为 RNA 分子。基因组表达的第二个产物是**蛋白质组** (proteome), 即细胞中那些决定细胞能够进行生化反应的所有蛋白质组分。组成蛋白质组的蛋白质是通过**翻译** (translation) 转录组中所包含的各个 RNA 分子来合成的。

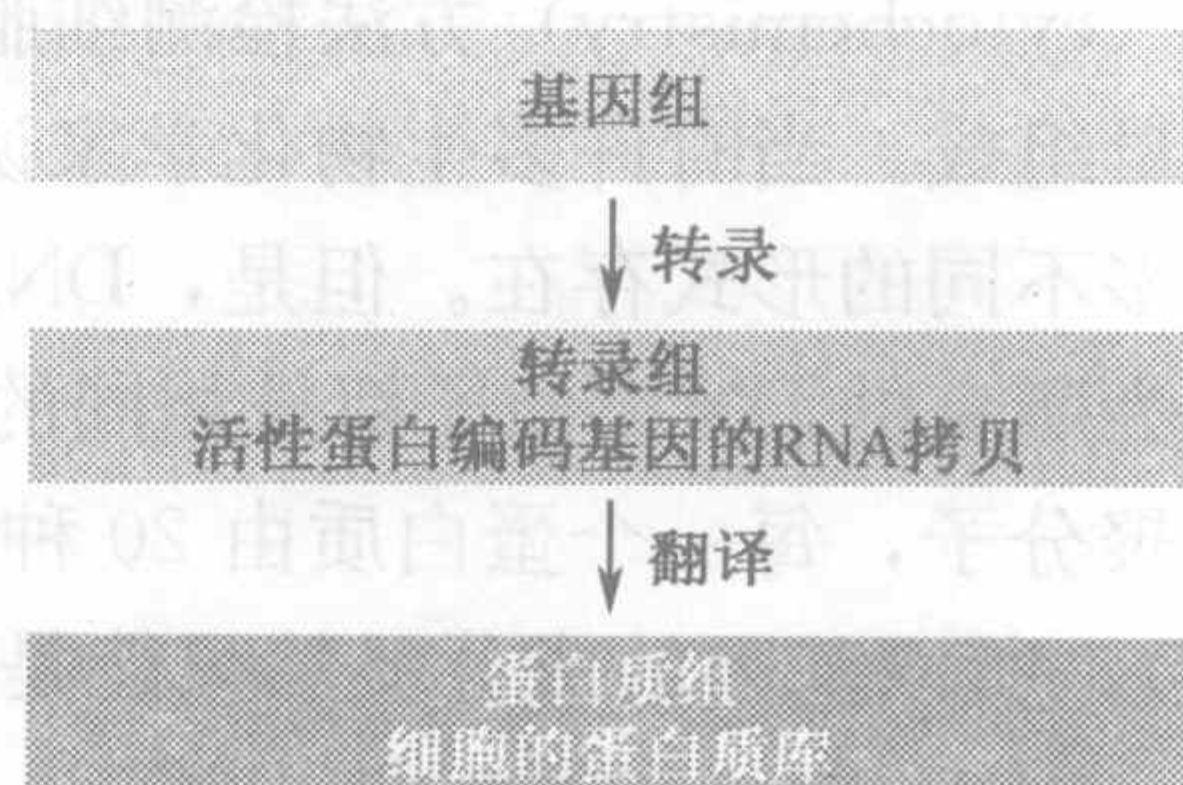


图 1.2 基因组、转录组和蛋白质组

本书是一部关于基因组及基因组表达的书, 将阐述如何研究基因组 (第 1 篇)、基因组是如何组织的 (第 2 篇)、基因组是如何发挥功能的 (第 3 篇) 以及它们是如何复制和进化的 (第 4 篇)。直到最近, 本书才能成稿。虽然从 20 世纪 50 年代以来, 分子生物学家已经研究了单个的或者小批的基因, 并在这些研究中积累了关于基因如何发挥作用的丰富知识。然而, 直到最近 10 年, 技术才发展到允许整个基因组水平的检测。与此同时, 单个的基因还在集中研究, 但关于单个基因的信息现在已经是放在基因组这一整体背景下来解释了。不仅仅在基因组研究中, 还在所有的生物化学和细胞生物学中都体现了这一新的、更宽泛的着重点。理解单个的生化通路或亚细胞过程是不够的, 现在的挑战由**系统生物学** (systems biology) 提出, 该学科试图把这些通路和过程连接起来成为描述活细胞和活生物体的所有功能的网络。

本书将带你在我们所掌握的基因组知识中遨游, 并向你展示这个激动人心的研究领域是如何正在加深我们对生物系统的不断了解。首先, 我们要回顾参与基因组和基因组表达的三种生物分子 (DNA、RNA 和蛋白质) 所具有的关键特征, 以此来关注分子生物学的基本原则。

## 1.1 DNA

DNA 由 Johann Friedrich Miescher 在 1869 年发现, 这位瑞士生物学家当时在德国蒂宾根 (Tübingen) 工作。Miescher 第一次从人白细胞中提取的抽提物是 DNA 和染色体蛋白的粗制混合物, 但他后来到了瑞士的巴塞尔 (Basel, 这家研究所后来以他的名字命名), 从鲑鱼精子中得到了纯的**核酸** (nucleic acid) 抽提物。Miescher 通过化学实验证明 DNA 是酸性的, 富含磷, 同时他发现单个 DNA 分子很大。不过, 直到 20 世纪 30 年代生物物理技术应用于 DNA 研究后, 人们才真正意识到 DNA 多聚链的长度是如此之大。

### 1.1.1 基因由 DNA 组成

基因由 DNA 组成的事实今天已广为人知, 但是, 很难理解的是在 DNA 被发现后的第一个 75 年内它的真正作用一直被怀疑。回溯到 1903 年, W. S. Sutton 发现细胞分裂过程中基因的遗传模式与染色体的行为类似, 从而得出基因位于染色体上的**染色体理论** (chromosome theory) 推论。用只和某一种化学成分特异结合的染料染色的细胞化



学 (cytochemistry) 方法检测细胞后, 人们发现染色体由 DNA 和蛋白质按大致相当的含量组成。当时许多生物化学家认为一定存在数十亿个基因, 而且遗传物质必定能够以许多不同的形式存在。但是, DNA 似乎不能满足这个要求, 因为在 20 世纪前叶, 人们认为所有的 DNA 分子都是相同的。另一方面, 大家正确地认识到蛋白质是高度变异的多聚分子, 每一个蛋白质由 20 种化学性质各异的氨基酸单体经过不同组合构成 (1.3.1 节)。所以基因的成分显然只能是蛋白质而不是 DNA。

对于 DNA 结构的错误认识一直持续着, 直到 20 世纪 30 年代后期人们才逐渐接受了 DNA 分子和蛋白质分子一样具有高度可变性。最初, 认为蛋白质是遗传物质的观念还是很强势的, 但两个重要实验的结果彻底地扭转了这种观念。

- Oswald Avery、Colin MacLeod 和 Maclyn McCarty 发现 DNA 是转化要素 (transforming principle) 的活性组分。转化要素是细菌细胞的抽提物, 当它与无毒的肺炎链球菌 (*Streptococcus pneumoniae*) 混合以后注射到小鼠体内, 即可将这些

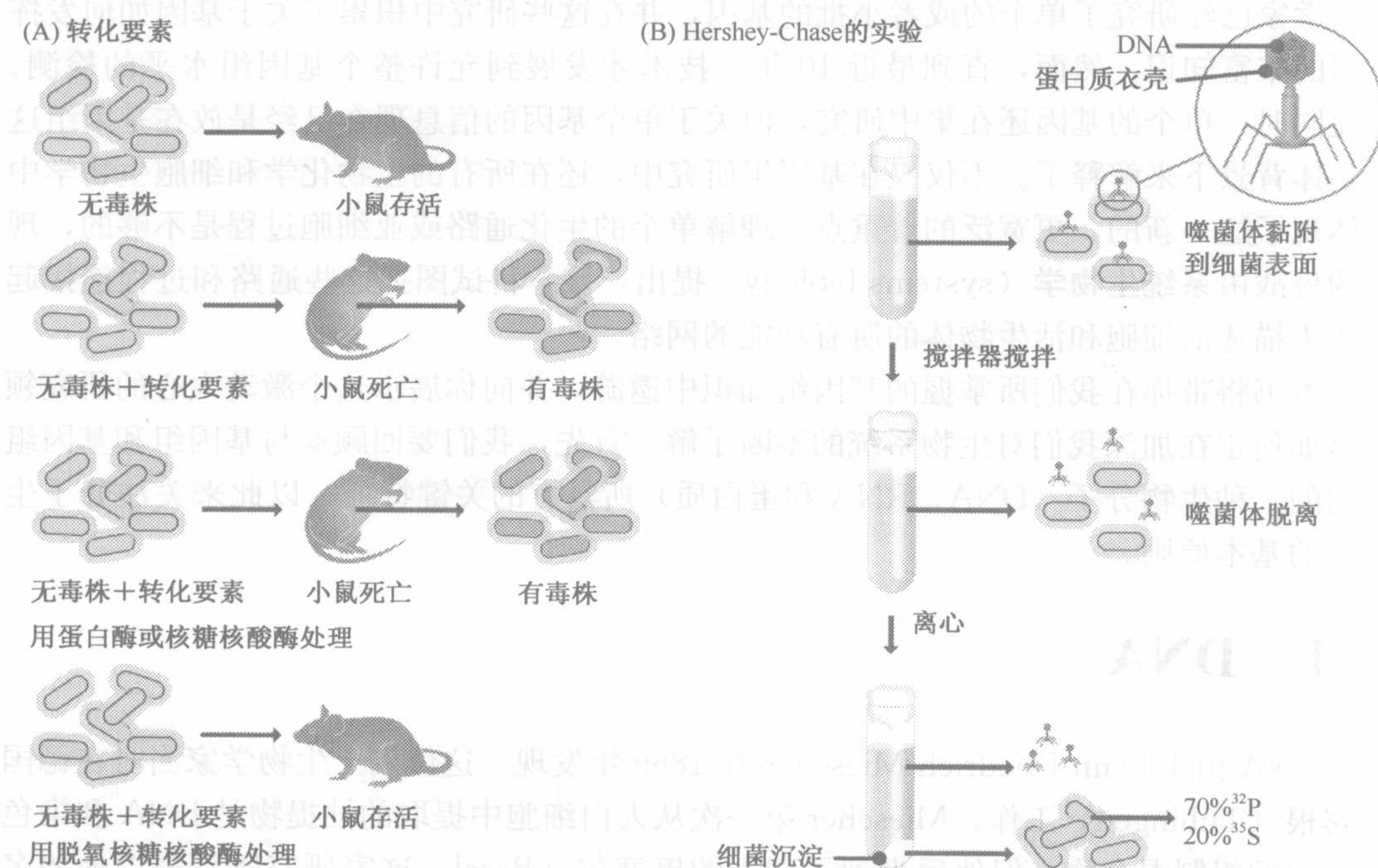


图 1.3 两个提示基因由 DNA 组成的实验

(A) Avery 和同事发现转化要素是 DNA。最上面的两行显示的是当小鼠被注射了加或不加转化要素 (从致病菌 *S. pneumoniae* 获得的细胞抽提物) 的无毒肺炎链球菌后所发生的情况。当加转化要素的时候, 转化要素中的基因将无毒株转化为致病株从而导致小鼠死亡, 随后这些致病株从死亡的小鼠肺部被分离出来。图的后两行显示的是用蛋白酶或核糖核酸酶处理对转化要素没有影响, 但是用脱氧核糖核酸酶却能使之失活。(B) Hershey-Chase 的实验使用了 T2 噬菌体, 每一个 T2 噬菌体由一个 DNA 分子及头-尾结构的蛋白质衣壳组成, 后者能使噬菌体黏附到细菌并注入基因。噬菌体的 DNA 用  $^{32}\text{P}$  标记, 而其蛋白质则用  $^{35}\text{S}$  标记。在噬菌体感染细菌几分钟以后, 振摇培养物使噬菌体的空壳从细菌表面脱落, 然后离心培养物, 收集离心管底部含有噬菌体基因的细菌沉淀, 而相对较轻的噬菌体颗粒被保留在上清液中。Hershey 和 Chase 发现细菌沉淀含有绝大部分  $^{32}\text{P}$  标记的噬菌体组分 (DNA), 但仅仅包含了 20% 的  $^{35}\text{S}$  标记物质 (噬菌体蛋白质)。在第二个实验中, 他们发现在感染周期晚期产生的新的噬菌体仅含有 1% 的亲代噬菌体蛋白质。关于噬菌体感染周期的更多细节请参见图 2.19



无毒的菌株转化成能够引起肺炎的致病菌株 [图 1.3 (A)]。当这个实验结果在 1944 年发表的时候, 只有少数几个微生物学家能理解转化过程涉及将基因从细胞抽提物中传递到活的细菌中。不过, 一旦这一点被接受, “Avery 实验”的精髓就显露了: 细菌的基因肯定是由 DNA 构成的。

- Alfred Hershey 和 Martha Chase Hershey 使用放射性标记 (radiolabeling) 发现当细菌培养物被噬菌体 (bacteriophage) (即感染细菌的病毒) 感染时, 是噬菌体的主要成分 DNA 进入了细菌体内 [图 1.3 (B)]。这个结果非常关键, 因为大家知道在感染周期中, 感染性噬菌体的基因是用来指导在细菌体内合成新的噬菌体的。如果只有感染性噬菌体的 DNA 进入了细菌体内, 那么这些噬菌体的基因就肯定是由 DNA 构成的。

虽然, 在我们看来这两个实验提供了关键的结果向我们表明基因是由 DNA 构成的, 但是, 当时的生物学家可不是那么容易被说服的。况且两个实验都有局限性, 从而给怀疑者留下了蛋白质也可以是遗传物质的置疑空间。比如, 有人担心 Avery 及其同事用来使转化要素失活的脱氧核糖核酸酶 (deoxyribonuclease) 的特异性, 其中哪怕混有痕量的蛋白酶 (protease) 都可能降解蛋白质, 那么这个作为转化要素是 DNA 的核心证据的实验结果就可能是错误的。噬菌体实验也不是结论性的, 因为 Hershey 和 Chase 在发表他们的结果的时候强调: “我们的实验显示, 将 T2 噬菌体物理的分离为遗传和非遗传组分是可能的……遗传组分的化学成分还有待鉴定, 但是直到一些问题……被解决以后”。回顾起来, 这两个实验之所以重要, 并不在于它们告诉了我们什么, 而在于它们促使生物学家重视 DNA 作为遗传物质的可能性, 并由此发现研究 DNA 的价值。正是这两个实验影响着 Watson 和 Crick 研究 DNA, 如后面所述, 他们的双螺旋 (double helix) 结构的发现解开了基因复制的谜团, 也真正使科学界相信基因是由 DNA 组成的。

### 1.1.2 DNA 的结构

James Watson 和 Francis Crick 这两个名字与 DNA 的联系如此之紧密, 以至于人们常常忽略其实当他们在 1951 年 10 月开始合作时, DNA 多聚体的详细结构已为人们所知。他们的贡献不在于发现 DNA 本身的结构, 而在于发现活细胞内两条 DNA 链相互缠绕形成的双螺旋结构。因此, 首先我们要检视一下在 Watson 和 Crick 开始工作之前他们都知道些什么。

#### 核苷酸和多聚核苷酸

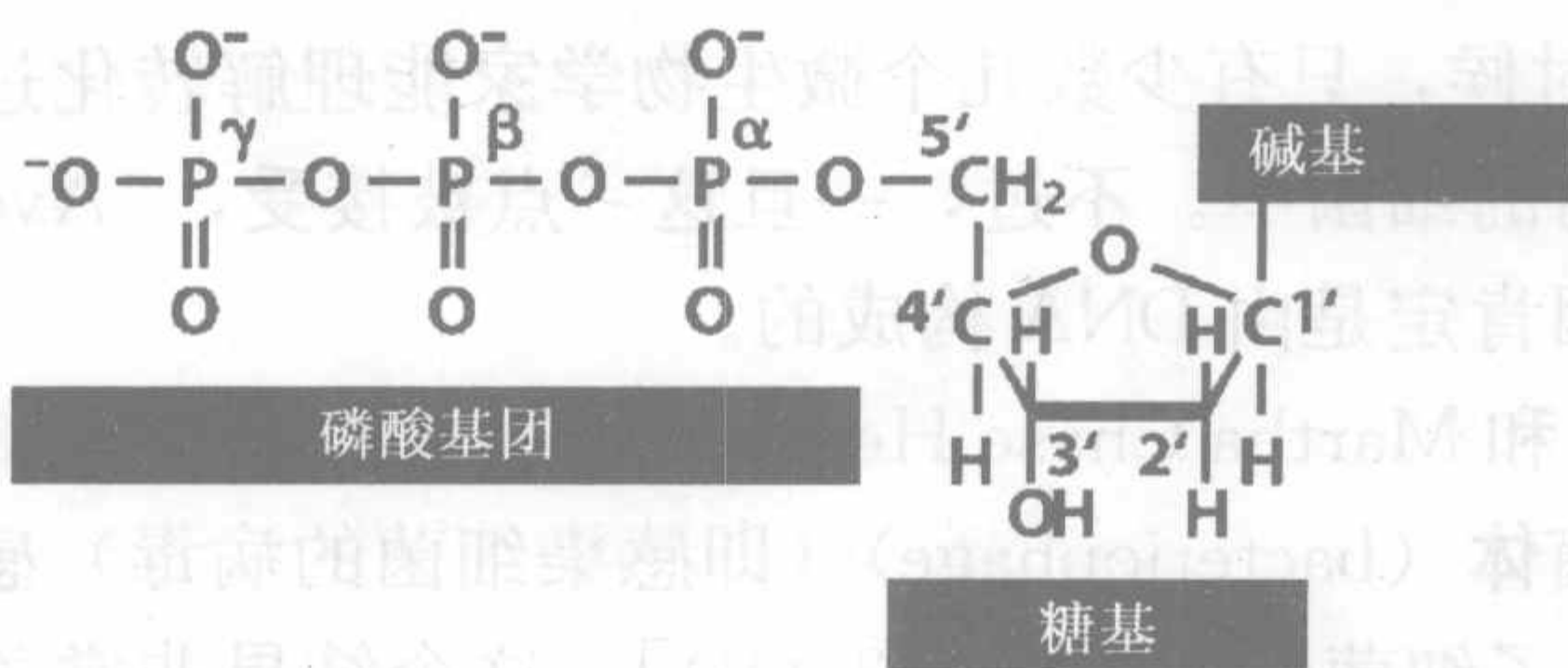
DNA 是线性、无分支的多聚分子, 由四种化学性质不同的核苷酸单体按任意顺序连接成链, DNA 链中核苷酸的数目可以达到几百、几千, 甚至几百万。DNA 多聚分子中的所有核苷酸都由三种成分组成 (图 1.4)。

(1) **2'-脱氧核糖** (2'-deoxyribose), 这是一种戊糖 (pentose), 即含有五个碳原子的糖。这些碳原子分别编为 1' (读作“1 撇”), 2' 等。“2'-脱氧核糖”意味着这种特殊的糖由核糖衍生而来, 其中与核糖的 2' 碳相连的羟基 (—OH) 被氢 (—H) 所替代。

(2) 一个**含氮碱基** (nitrogenous base), 为**胞嘧啶** (cytosine)、**胸腺嘧啶**



(A) 一个核苷酸



(B) DNA 中的四种碱基

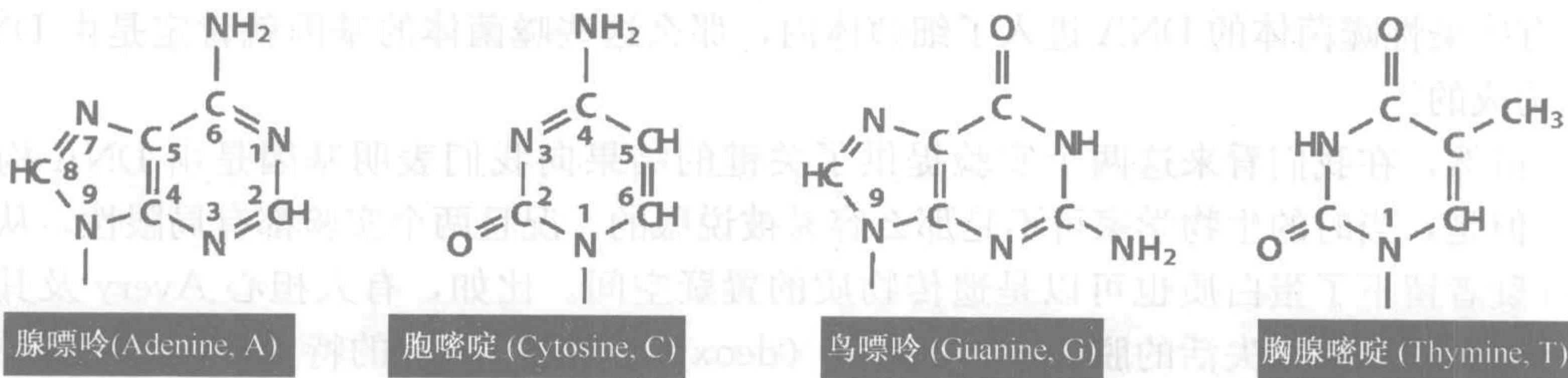


图 1.4 核苷酸的结构

(A) 脱氧核糖核苷酸的一般结构，即 DNA 中的核苷酸的共同结构。(B) 脱氧核糖核苷酸中的四种碱基 (thymine) (单环嘧啶)、腺嘌呤 (adenine)、鸟嘌呤 (guanine) (双环嘌呤) 当中的一种。碱基通过嘧啶的 1 位氮或嘌呤的 9 位氮与糖的 1' 碳以  $\beta$ -N-糖苷键 ( $\beta$ -N-glycosidic bond) 相连。

(3) 一个磷酸基团 (phosphate group)，由一个、两个或者三个相连的磷酸基与 2'-脱氧核糖的 5'-碳原子结合。磷酸基团编码为  $\alpha$ 、 $\beta$ 、 $\gamma$ ， $\alpha$  磷酸为直接与糖相连的磷酸基团。

5' 磷酸末端

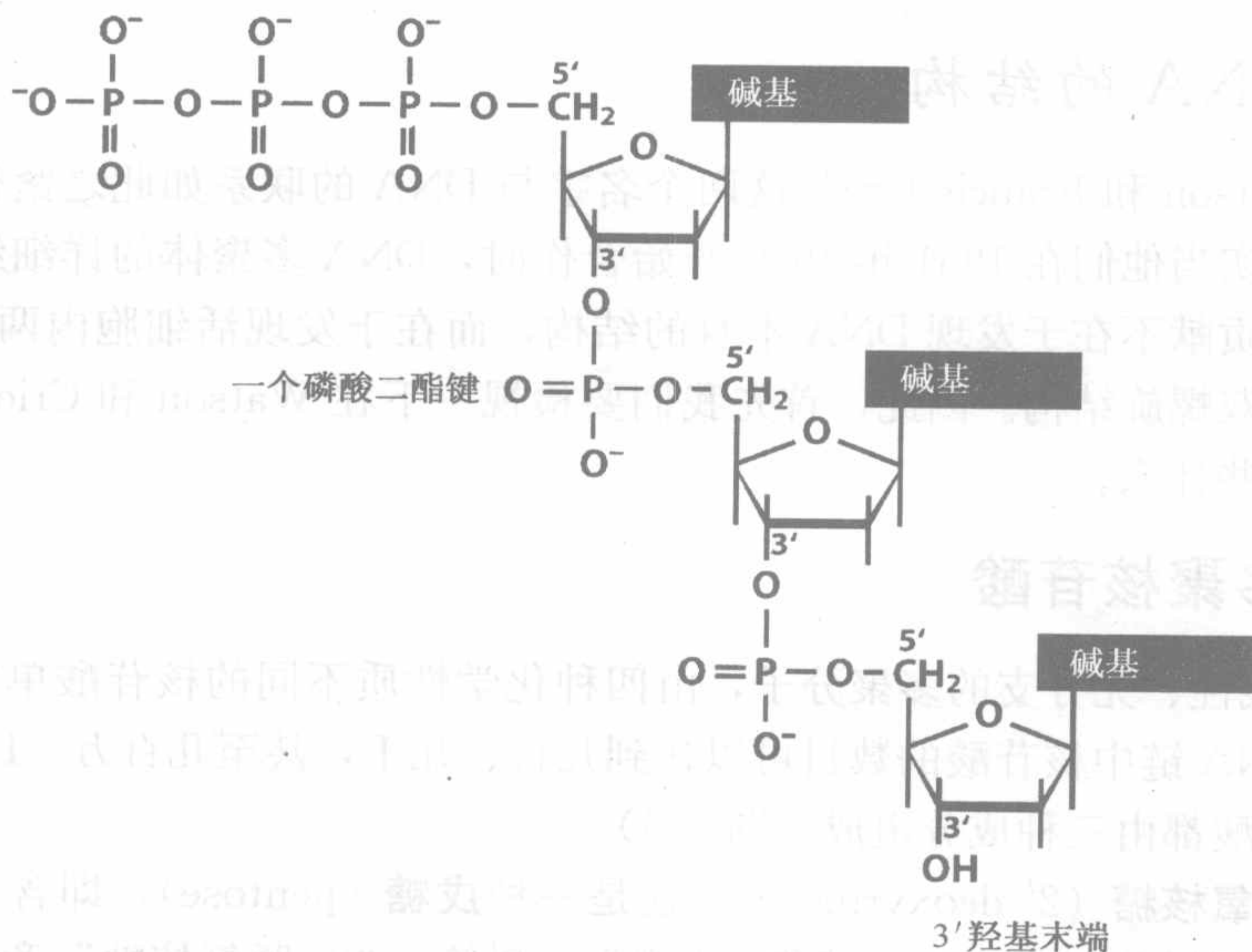


图 1.5 一段短的 DNA 多聚核苷酸序列来说明磷酸二酯键的结构

请注意多核苷酸两个末端的化学性质不同



仅由糖和碱基组成的分子称为核苷 (nucleoside)；加入磷酸基团后成为核苷酸。虽然细胞中的核苷酸可以含有一个、两个或者三个磷酸基团，但是，只有含有三个磷酸基团的核苷酸是 DNA 合成的底物。四种参与 DNA 合成的核苷酸的化学全称是：

- 2'-脱氧腺苷 5'-三磷酸
- 2'-脱氧胞苷 5'-三磷酸
- 2'-脱氧鸟苷 5'-三磷酸
- 2'-脱氧胸苷 5'-三磷酸

它们的缩写分别是 dATP、dCTP、dGTP 和 dTTP，在 DNA 序列中分别简写为 A、C、G 和 T。

在多核苷酸中，单个核苷酸由 5' 和 3' 碳原子之间形成的磷酸二酯键 (phosphodiester)

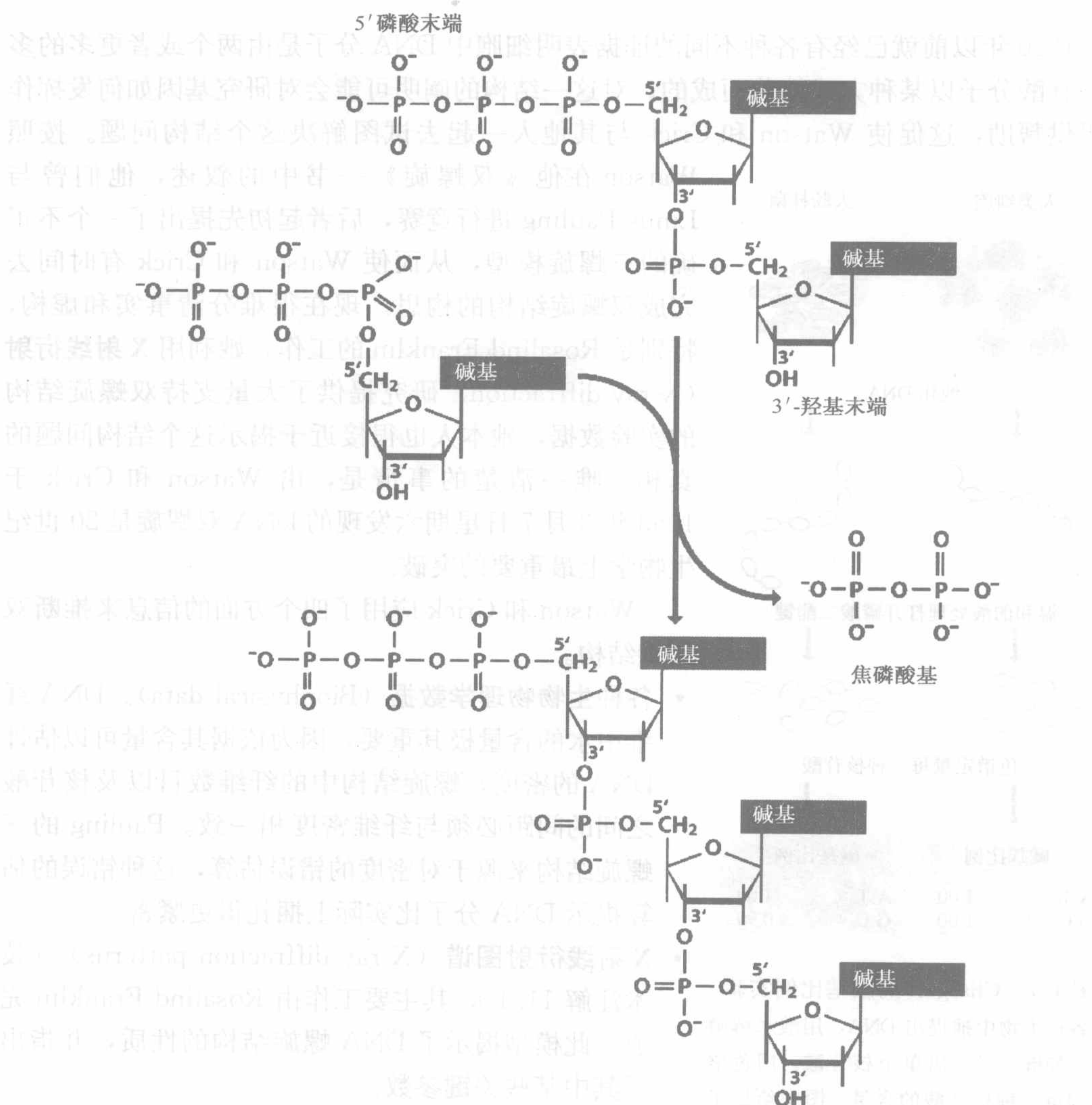


图 1.6 DNA 多聚核苷酸合成的聚合反应

合成从 5' → 3' 方向进行，新的核苷酸被加到多核苷酸链的 3' 端。核苷酸的 β 和 γ 磷酸基团以焦磷酸盐的形式被去除



bond) 连接在一起 (图 1.5)。通过这个连接的结构我们可以看到聚合反应 (图 1.6) 包括一个核苷酸的外面两个磷酸基团 ( $\beta$  和  $\gamma$  磷酸) 的去除和另一个核苷酸上与 3' 碳原子相连的羟基被取代。注意多核苷酸两端的化学性质不同, 一端有未反应的三磷酸基团连接于 5' 碳原子上 [5' 或 5'-P terminus (5'-磷酸末端)], 另一端有一个未反应的羟基连接于 3' 碳原子上 [3' 或 3'-OH terminus (3'-羟基末端)]。这意味着多聚核苷酸有化学上的方向性, 可表述为 5'→3' (图 1.5 由上至下) 或 3'→5' (图 1.5 由下至上)。磷酸二酯键极性的一个重要结果是 5'→3' 方向延伸 DNA 多聚体所需的化学反应不同于 3'→5' 方向的延伸反应。所有天然的 DNA 聚合酶 (DNA polymerase) 仅能进行 5'→3' 方向的合成, 这明显地增加了双链 DNA 复制过程的复杂性 (15.2 节)。

## 证明双螺旋的证据

1950 年以前就已经有各种不同的证据表明细胞中 DNA 分子是由两个或者更多的多聚核苷酸分子以某种方式组装而成的。对这一结构的阐明可能会对研究基因如何发挥作用提供帮助, 这促使 Watson 和 Crick 与其他人一起去试图解决这个结构问题。按照

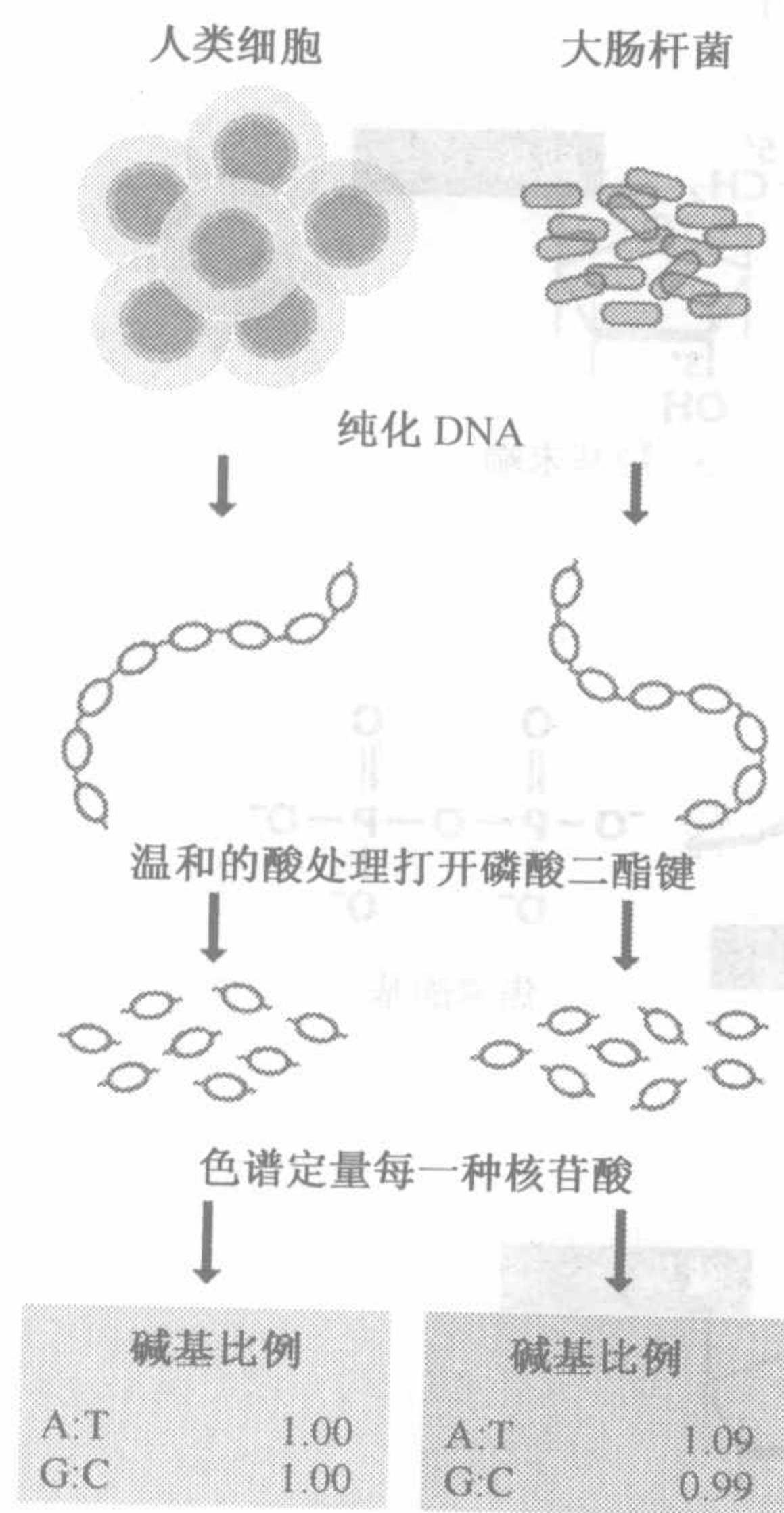


图 1.7 Chargaff 的碱基比例实验  
从各种生物中抽提出 DNA, 用酸水解磷酸二酯键, 释放出单个核苷酸。用色谱检测每一种核苷酸的含量。图中给出了 Chargaff 的某些实验结果。可以看到包括实验误差在内, 腺嘌呤和胸腺嘧啶含量相等, 鸟嘌呤和胞嘧啶含量相等

Watson 在他《双螺旋》一书中的叙述, 他们曾与 Linus Pauling 进行竞赛, 后者起初先提出了一个不正确的三螺旋模型, 从而使 Watson 和 Crick 有时间去完成双螺旋结构的构思。现在很难分清事实和虚构, 特别是 Rosalind Franklin 的工作, 她利用 X 射线衍射 (X-ray diffraction) 研究提供了大量支持双螺旋结构的实验数据, 她本人也很接近于揭示这个结构问题的真相。唯一清楚的事情是, 由 Watson 和 Crick 于 1953 年 3 月 7 日星期六发现的 DNA 双螺旋是 20 世纪生物学上最重要的突破。

Watson 和 Crick 应用了四个方面的信息来推断双螺旋结构。

- 各种生物物理学数据 (Biophysical data)。DNA 纤维中水的含量极其重要, 因为依据其含量可以估计 DNA 的密度。螺旋结构中的纤维数目以及核苷酸之间的间距必须与纤维密度相一致。Pauling 的三螺旋结构来源于对密度的错误估算, 这种错误的估算提示 DNA 分子比实际上捆扎得更紧密。
- X 射线衍射图谱 (X-ray diffraction patterns) (技术注解 11.1), 其主要工作由 Rosalind Franklin 完成, 此模型揭示了 DNA 螺旋结构的性质, 并指出了其中某些关键参数。
- 碱基比例 (base ratios), 由纽约哥伦比亚大学的 Erwin Chargaff 发现。Chargaff 对不同来源的 DNA 样品进行了极其详尽的分析, 他发现虽然不



同生物的 DNA 中碱基的含量不同，但是腺嘌呤的含量总是与胸腺嘧啶的含量相等，而鸟嘌呤的含量则总是与胞嘧啶的含量相等（图 1.7）。从碱基比例引申出了**碱基配对**（base-pairing）定律，它是最终促成发现双螺旋结构的关键所在。

- **模型构建**（model building），这是由 Watson 和 Crick 自己真正完成的内容。构建比例模型来描述 DNA 的可能结构，由此可以计算各种原子之间的相对位置，保证成键的基团不会相距太远，而且其他基团也不会因为彼此间距太近而相互干扰。

## 双螺旋的主要特征

双螺旋是右手螺旋，这就好像在爬一个螺旋楼梯的时候，楼梯外侧的扶手在你的右手侧。双螺旋由两条多核苷酸链相互缠绕在一起组成，两条链方向相反 [图 1.8 (A)]。

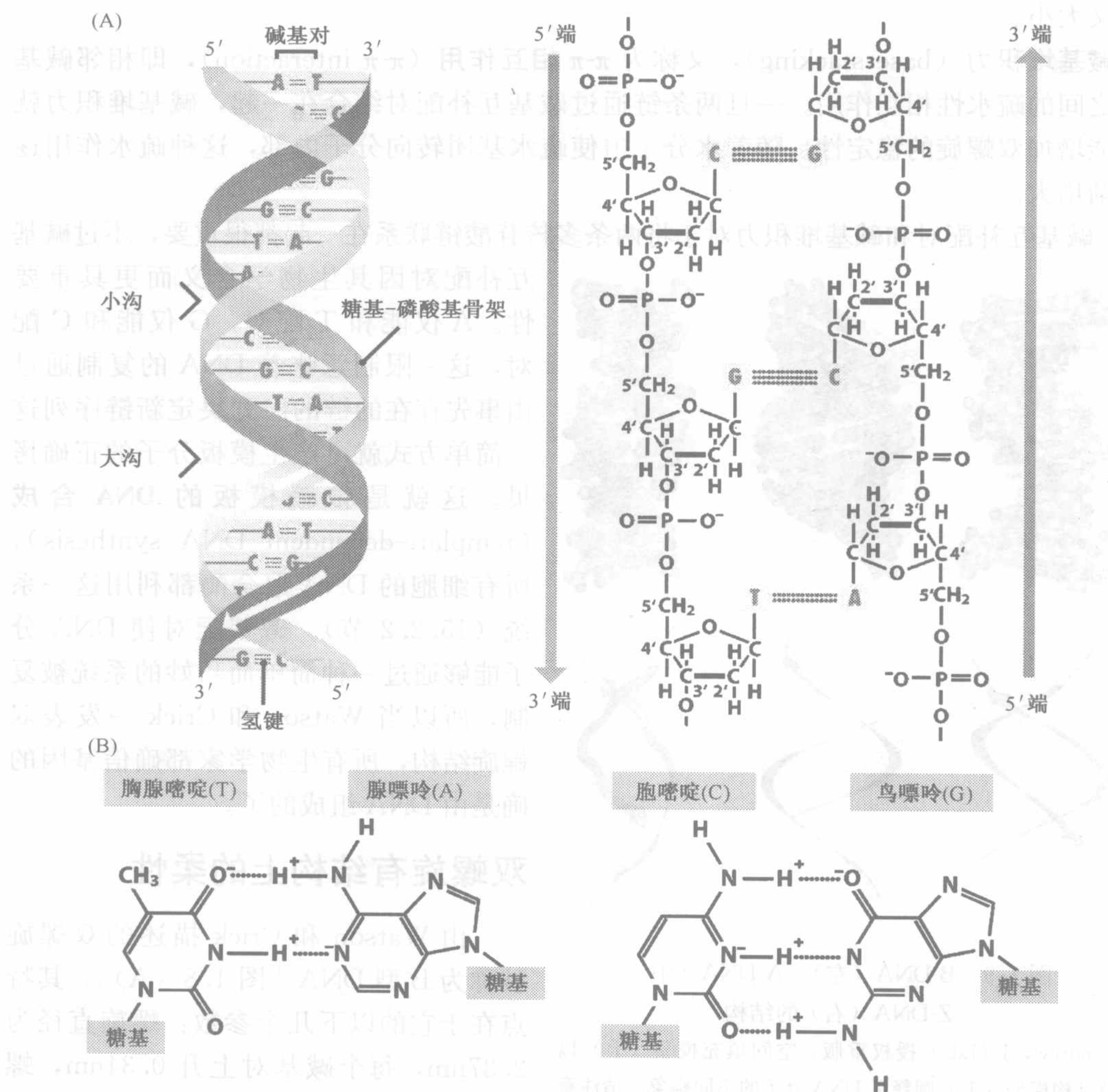


图 1.8 DNA 双螺旋结构

(A) 双螺旋结构的两种表示形式。左边的结构图中，用条带表示每个核苷酸的糖基-磷酸骨架，用黑色标记碱基对。右边的结构图中给出了三对碱基的化学结构。(B) A 和 T 配对，G 和 C 配对。用轮廓图表示碱基，用虚线表示氢键。注意 G—C 碱基对有三个氢键，而 A—T 碱基对只有两个



螺旋由两种化学相互作用来稳定。

- 两条链间的**碱基配对** (base pairing), 包括一条链上的腺嘌呤与另一条链上的胸腺嘧啶之间或胞嘧啶与鸟嘌呤之间配对形成**氢键** (hydrogen bond) [图 1.8 (B)]。氢键是一个负电性原子 (如氧或者氮) 与结合在另外一个负电性原子上的氢之间的弱静电引力。氢键比共价键长, 作用弱得多, 氢键在 25°C 时的键能为 1~10kcal/mol, 而共价键的键能为 90kcal/mol。和在 DNA 双螺旋中所起的作用一样, 氢键亦能稳定蛋白质的二级结构。A 和 T 配对、G 和 C 配对这两种碱基配对组合解释了 Chargaff 发现的碱基比例。它们是唯一被允许的配对, 一方面这是由核苷酸碱基的几何结构和参与形成化学键基团的相对位置决定的, 另外还因为碱基配对必须在嘌呤和嘧啶之间, 因为嘌呤-嘌呤的配对太大, 位于螺旋内不合适, 而嘧啶-嘧啶配对又太小。
- **碱基堆积力** (base stacking), 又称为  $\pi$ - $\pi$  相互作用 ( $\pi$ - $\pi$  interaction), 即相邻碱基之间的疏水性相互作用。一旦两条链通过碱基互补配对结合在一起, 碱基堆积力就能增加双螺旋的稳定性。随着水分子迫使疏水基团转向分子内部, 这种疏水作用逐渐增大。

碱基互补配对和碱基堆积力对于将两条多核苷酸链联系在一起都很重要, 不过碱基

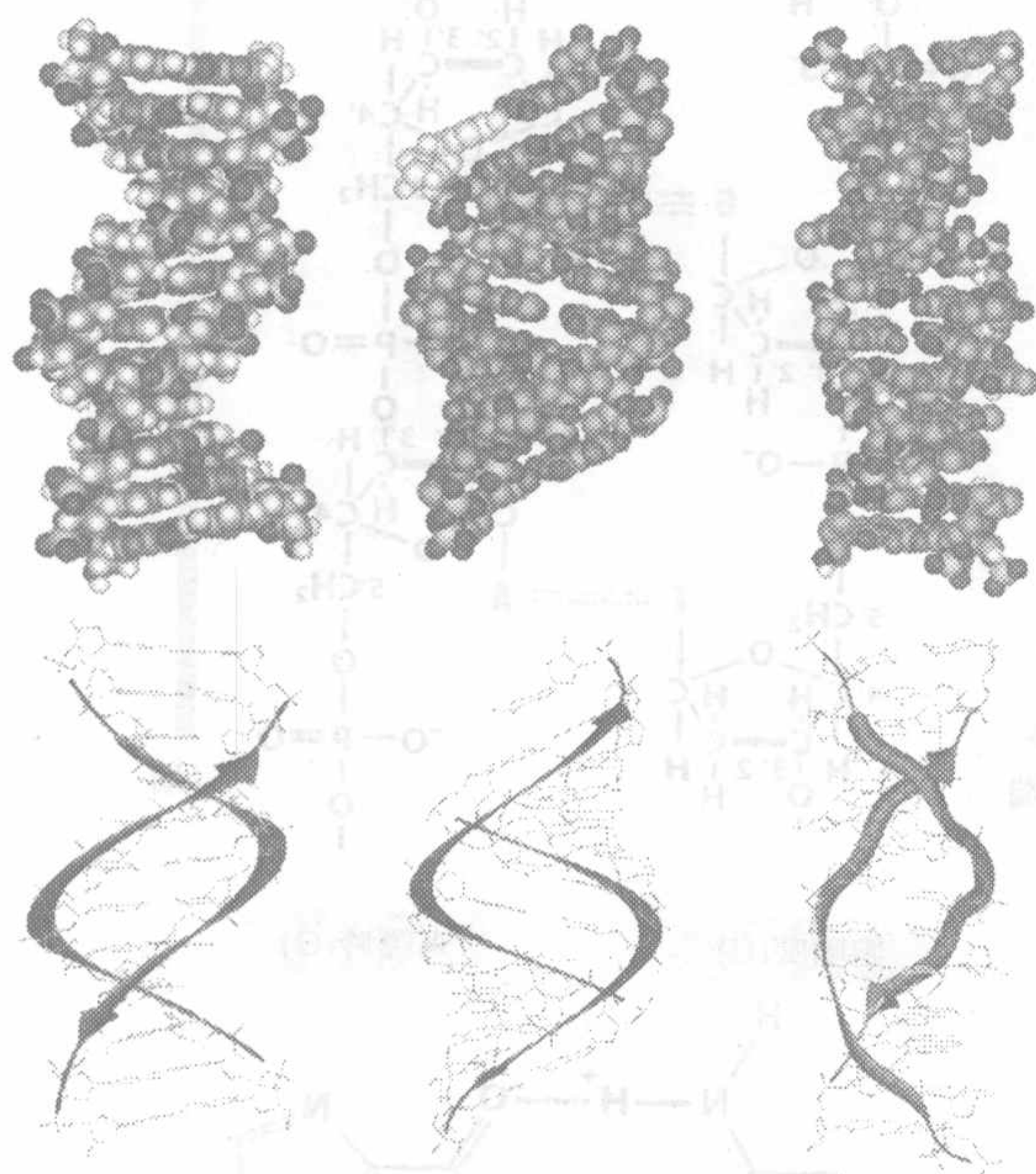


图 1.9 B-DNA (左)、A-DNA (中)、Z-DNA (右) 的结构

Kendrew, J (Ed.) 授权重版。空间填充模型 (上) 以及结构模型 (下) 阐释了 DNA 分子的不同构象。请注意螺旋的直径、所有完整螺旋所包含的碱基对的数目以及在这些分子间大沟、小沟的拓扑学方面的差异

(The Encyclopaedia of Molecular Biology. Copyright 1994 Blackwell Science)

互补配对因其生物学含义而更具重要性。A 仅能和 T 配对, G 仅能和 C 配对, 这一限制意味着 DNA 的复制通过由事先存在的链的序列决定新链序列这一简单方式就可产生模板分子的正确拷贝。这就是**依赖模板的 DNA 合成** (template-dependent DNA synthesis), 所有细胞的 DNA 聚合酶都利用这一系统 (15.2.2 节)。碱基配对使 DNA 分子能够通过一种简单而巧妙的系统被复制, 所以当 Watson 和 Crick 一发表双螺旋结构, 所有生物学家都确信基因的确是由 DNA 组成的了。

## 双螺旋有结构上的柔性

由 Watson 和 Crick 描述的双螺旋被称为 B 型 DNA [图 1.8 (A)]。其特点在于它的以下几个参数: 螺旋直径为 2.37nm, 每个碱基对上升 0.34nm, 螺距 (即一圈完整的螺旋的距离) 为 3.4nm, 这相当于每圈 10 个碱基对。活体细胞中的 DNA 被认为主要是 B 型, 但现在我们知道基因组 DNA 分子在结



构上不是完全一致的，这主要是由于螺旋中的每个核苷酸都有柔性，能采取略为不同的分子形状。为采用这些不同的构象，核苷酸中原子的相对位置必须略为改变。构象变化有很多种可能，但最重要的是围绕着  $\beta$ -N-糖苷键的旋转，这改变了碱基相对于糖的方向以及围绕 3'-与 4'-碳之间的键的旋转。两种旋转对双螺旋都有很显著的影响：改变碱基方向会影响两个多核苷酸的相对位置，围绕 3'-4' 键的旋转会影响糖-磷酸骨架的构象。

因此，单个核苷酸的内部旋转可导致螺旋整体结构发生较大的变化。20 世纪 50 年代以后，人们已经认识到 DNA 分子纤维暴露于不同的相对湿度下时，双螺旋的尺度会发生变化。比如，双螺旋可改变为 A 型（图 1.9），其直径为 2.55nm，每个碱基对上升 0.29nm，螺距为 3.2nm，相当于每圈 11 个碱基对（表 1.1）。其他的变构类型包括 B'-、C'-、C'''-、D-、E-和 T-型 DNA。所有这些都和 B 型一样是右手螺旋，但也有可能发生更为剧烈的重构而产生左手螺旋的 Z-DNA（图 1.9）。Z-DNA 是双螺旋的一种更为细长的形式，其直径仅为 1.84nm。

表 1.1 DNA 双螺旋不同构象的特征

特征	B-DNA	A-DNA	Z-DNA
螺旋类型	右手螺旋	右手螺旋	左手螺旋
螺旋直径/nm	2.37	2.55	1.84
每个碱基上升高度/nm	0.34	0.29	0.37
完整螺旋间的距离（螺距）/nm	3.4	3.2	4.5
每转螺旋的碱基对数	10	11	12
大沟形状	宽，深	窄，深	平坦
小沟形状	窄，浅	宽，浅	窄，深

仅从各种形式双螺旋的碱基参数不能揭示它们之间最显著的差别，这些差别与直径和螺距无关，而是从结构的表面到螺旋内部区域的可接近程度。如图 1.8 和图 1.9 所示，B 型 DNA 没有十分光滑的表面，而是沿着螺旋有两条沟，其中一条沟相对宽且深，称为大沟（major groove），另一个窄而较浅，称为小沟（minor groove）。A-DNA 也有两个沟（图 1.9），但这种构象的大沟更深，小沟更浅而宽。Z-DNA 又有不同，一个沟基本上不存在，另一个则非常窄而深。对于 DNA 的每种构象，至少都有一个沟的部分内表面是由与核苷酸碱基相连的化学基团形成的。在第 11 章中我们将讨论基因组中所含生物信息的表达是通过 DNA 结合蛋白介导的，这些蛋白质与双螺旋结合并调节其中基因的活性。DNA 结合蛋白必须结合在那些受其影响的基因附近的特异位点才能发挥其功能。这样，DNA 结合蛋白可仅仅通过进入沟的底部来“读取”核苷酸序列，而不必破坏碱基对将螺旋打开。正因为如此，能够被一种 DNA 结合蛋白的结构识别的 B-DNA 内一段特异的核苷酸序列如果形成其他的构象，该 DNA 结合蛋白将不能识别它。在第 11 章我们会看到，DNA 分子的构象变化和核苷酸序列所导致的其他结构的多态性，对决定基因组与它的 DNA 结合蛋白之间相互作用的特异性非常重要。



## 1.2 RNA 和转录组

基因组表达的最初产物是转录组 (图 1.2)，即由那些含有细胞在特定时间所需生物信息的、编码蛋白质的基因衍生而来的 RNA 分子的集合。转录组中的 RNA 分子以及其他来自非编码基因的 RNA 都由叫做转录的过程产生。在下面这一章节中我们将先揭示 RNA 的结构，然后将更多地关注活细胞中存在的各种类型的 RNA 分子。

### 1.2.1 RNA 的结构

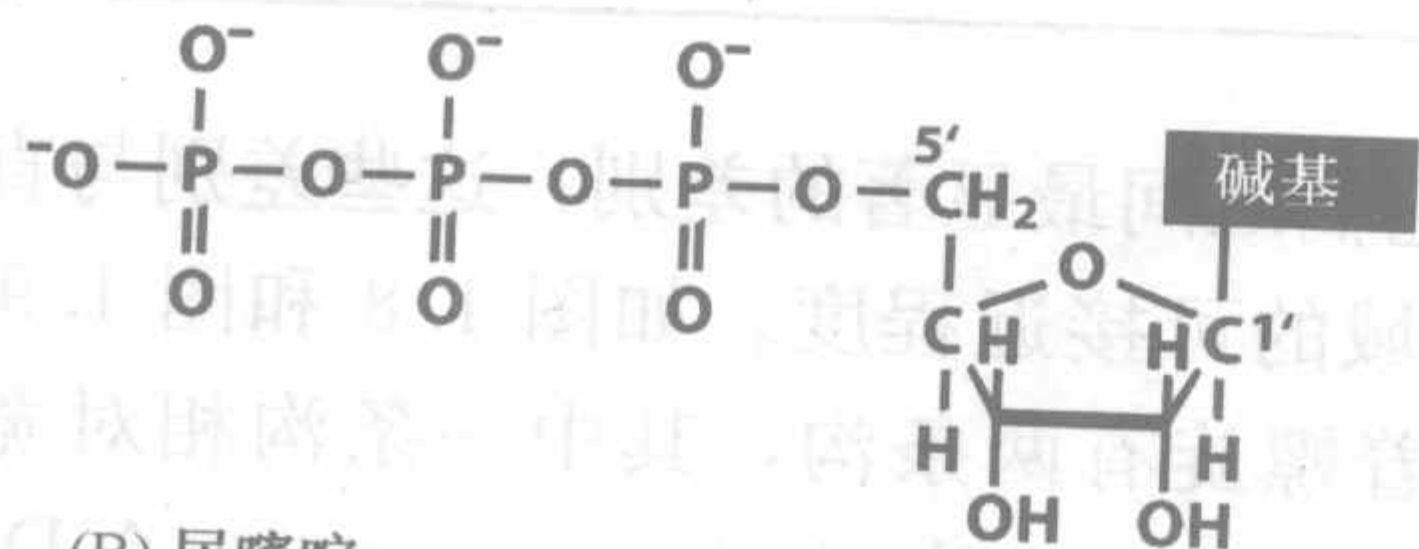
RNA 也是一个多聚核苷酸，但和 DNA 相比有两点不同 (图 1.10)。首先，RNA 核苷酸中的糖是核糖 (ribose)；其次，RNA 含有尿嘧啶 (uracil) 而没有胸腺嘧啶。因此合成 RNA 的四种核苷酸底物是：

- 腺苷 5'-三磷酸
- 胞苷 5'-三磷酸
- 鸟苷 5'-三磷酸
- 尿苷 5'-三磷酸

简写成 ATP、CTP、GTP 和 UTP 或者 A、C、G 和 U。

和 DNA 一样，RNA 多核苷酸含有 3'-5' 磷酸二酯键，但这些磷酸二酯键比 DNA 多核苷酸中的稳定性差，这是由于糖的 2' 位羟基的间接作用。RNA 分子在长度上很少超过几千个核苷酸，而且虽然许多 RNA 形成分子内的碱基配对 (图 13.2)，但是大多数 RNA 还是单链形式而不是双链。

(A) 一个核苷酸



(B) 尿嘧啶

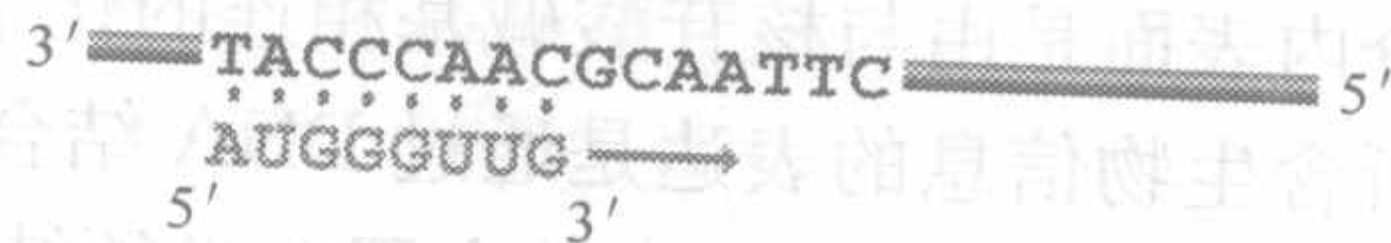
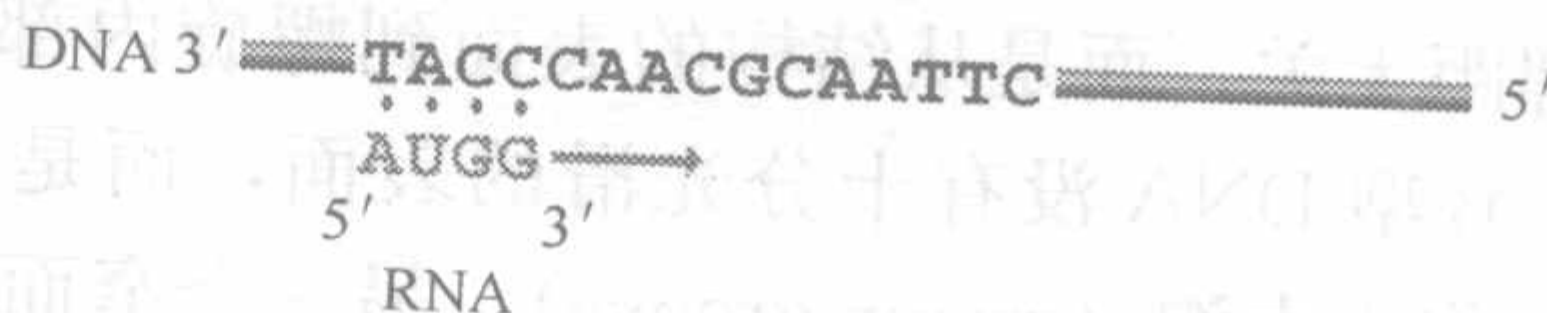
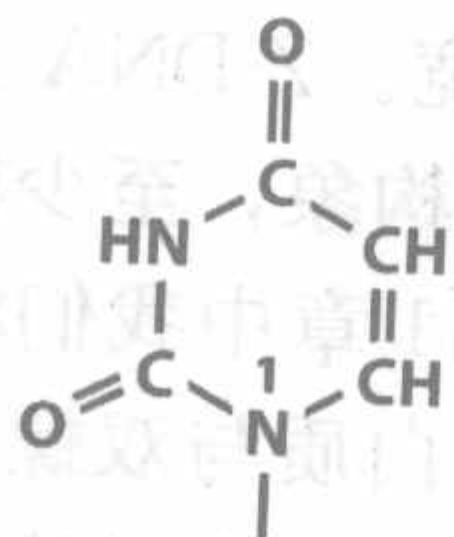


图 1.10 DNA 和 RNA 化学性质的区别

(A) RNA 由核糖核苷酸组成，其中所含的糖是核糖，而不是 2'-脱氧核糖。其差异在于与 2'-碳原子结合的是羟基而非氢原子。(B) RNA 含尿嘧啶，不含胸腺嘧啶

图 1.11 依赖模板的 RNA 合成

RNA 转录物以 5'→3' 方向合成，沿 3'→5' 方向阅读 DNA，转录物的序列由 DNA 模板通过碱基配对原则决定

负责将 DNA 转录为 RNA 的酶叫做依赖于 DNA 的 RNA 聚合酶 (DNA-dependent RNA polymerases)。该命名暗示了它们以依赖于 DNA 的形式催化的酶反应产物是来自



核苷酸的多聚 RNA 链，意味 DNA 模板中的核苷酸序列指导着产生的 RNA 中核苷酸的序列（图 1.11）。还可以将这个命名缩略为 **RNA 聚合酶**（RNA polymerase），因为在使用该命名的时候很少会与涉及某些病毒基因组复制及表达的**依赖 RNA 的 RNA 聚合酶**（RNA-dependent RNA polymerases）混淆。依赖模板的 RNA 合成的化学基础与图 1.6 中显示的 DNA 合成相似。核苷酸被一个接一个地加到 RNA 转录物的 3' 端，根据碱基配对原则决定加入哪一种核苷酸：A 与 T 或 U 配对；G 与 C 配对。在加入每个核苷酸的过程中，加进来的核苷酸上的  $\beta$ -和  $\gamma$ -磷酸基团被去除，多聚链 3' 端的核苷酸上 3' 碳上的羟基被去除，这些完全与 DNA 多聚反应一致。

## 1.2.2 细胞内的 RNA 组分

一个典型的细菌细胞含有 0.05~0.10pg 的 RNA，约占其总质量的 6%。一个哺乳动物细胞，比细菌大得多，含有更多的 RNA，共有 20~30pg，但是只占细胞总质量的 1%。理解细胞 RNA 组成的最好方法是根据功能将它们分为类和亚类。分类的方法有许多种，最富含信息的方案如图 1.12 所示。初步的划分是**编码 RNA**（coding RNA）和**非编码 RNA**（noncoding RNA）。编码 RNA 构成转录组，仅由**信使 RNA**（messenger RNA, mRNA）这一类分子组成，它们是蛋白质编码基因的转录物，因而在基因组表达的较晚阶段被翻译为蛋白质。信使 RNA 很少占总 RNA 的 4% 以上，并且寿命很短，合成后不久就降解。细菌 mRNA 的半衰期不到几分钟，而真核生物中大部分 mRNA 合成后数小时即降解。这种快速的周转意味着转录组并不是固定的，而是可以通过改变个别 mRNA 的合成速率而重建。

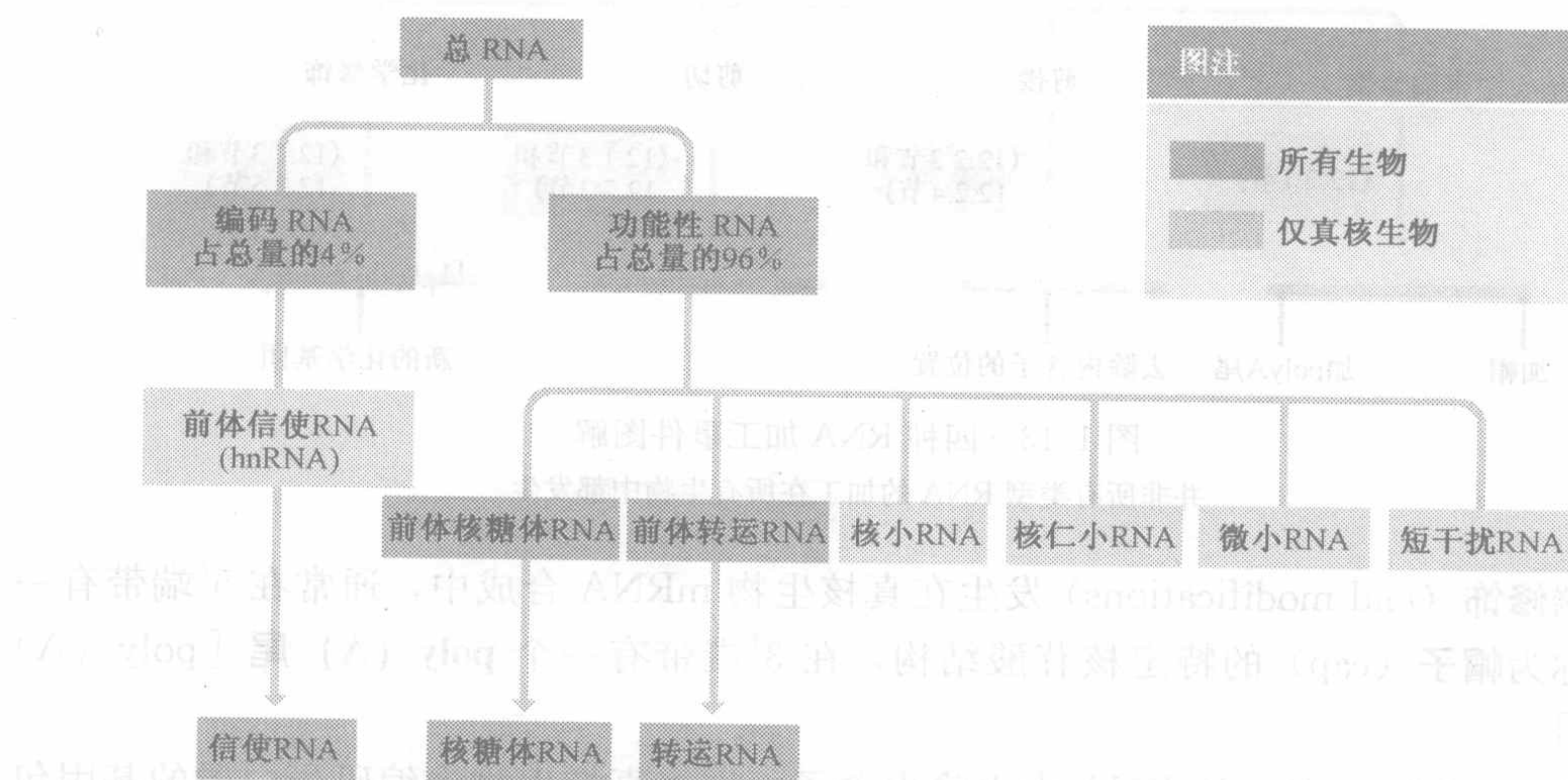


图 1.12 细胞的 RNA 组成

本图显示了所有生物中都有的 RNA 类型和仅在真核细胞中发现的类型

第二种 RNA 因为不被翻译成蛋白质而被叫做“非编码”RNA。不过**功能性 RNA**（functional RNA）应该是一个更好的命名，因为这就强调了虽然非编码 RNA 不是转录组的成分，但是它们在细胞内仍起关键的作用。有多种功能性 RNA，最重要的如下。

- **核糖体 RNA**（ribosomal RNA, rRNA）存在于所有生物中，是细胞中最丰富的



RNA，在活跃分裂的细菌中占80%以上。这些分子是核糖体的组分，而蛋白质合成就发生在核糖体上（13.2节）。

- **转运 RNA** (transfer RNA, tRNA) 是和 rRNA 一同参与蛋白质合成过程的小分子，存在于所有生物中。它们的功能是将氨基酸携带至核糖体，并确保它们按照被翻译的 mRNA 的核苷酸序列所指定的顺序进行连接（13.1节）。
- **核小 RNA** (small nuclear RNA, snRNA; 又叫 U-RNA, 因为这些分子富含尿嘧啶) 发现于真核细胞核中。它们涉及剪接 (splicing)——一个将蛋白质编码基因的原始转录物加工为 mRNA 过程中的关键步骤（12.2.2节）。
- **核仁小 RNA** (small nucleolar RNA, snoRNA) 发现于真核细胞核的核仁区。它们在 rRNA 分子加工过程中通过指导酶来修饰特定的核苷酸位点（例如，必须加上一个甲基）起到核心作用（12.2.5节）。
- **微小 RNA** (micro RNA, miRNA) 和 **短干扰 RNA** (short interfering RNA, siRNA) 是调控个别基因表达的小 RNA（12.2.6节）。

### 1.2.3 前体 RNA 的加工

除以上所述的成熟 RNA 外，细胞中还含有前体分子。许多 RNA，尤其是真核生物中，起初是以前体或前体 RNA (pre-RNA) 的形式合成的，它们在执行功能前必须经过加工。在第12章将详述各种加工事件，包括以下过程（图1.13）。

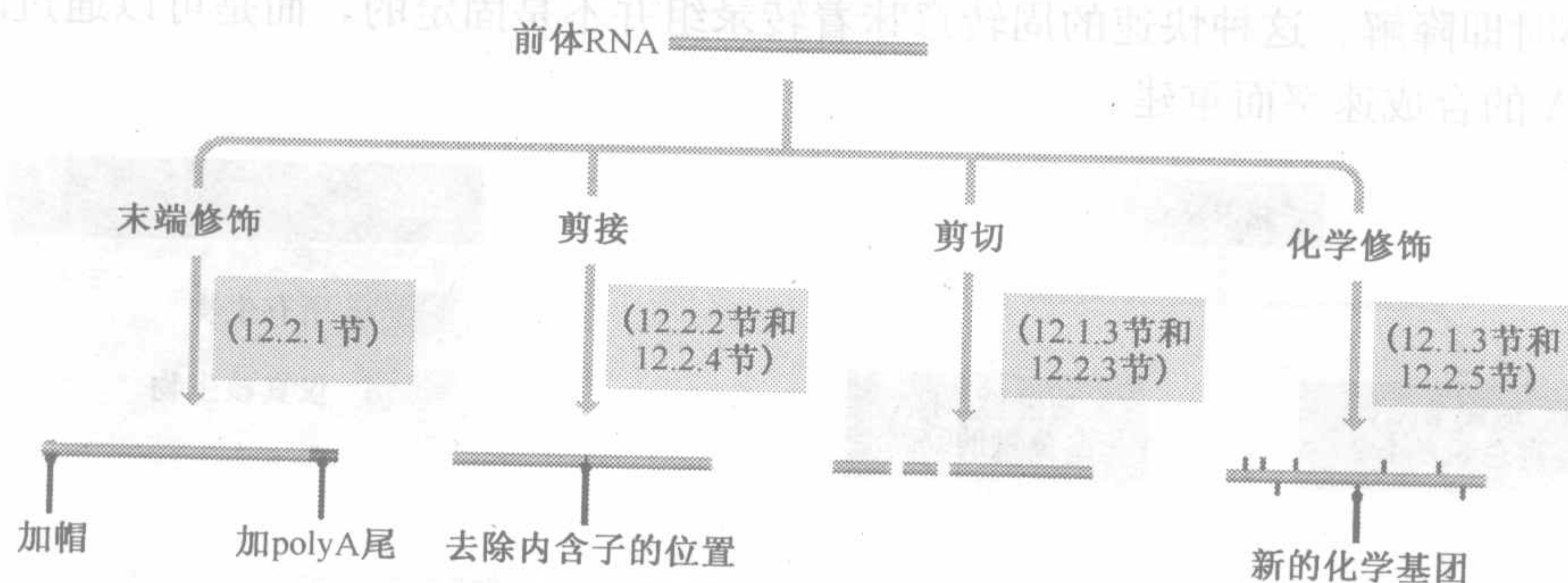


图 1.13 四种 RNA 加工事件图解

并非所有类型 RNA 的加工在所有生物中都发生

- **末端修饰** (end-modifications) 发生在真核生物 mRNA 合成中，通常在 5' 端带有一个称为帽子 (cap) 的特定核苷酸结构，在 3' 端带有一个 **poly (A) 尾** [poly (A) tail]。
- **剪接** (splicing) 从前体 RNA 中去除内含子。许多真核生物中编码蛋白质的基因包含没有生物学信息的内部片段，这些片段叫做**内含子** (intron)，它们在基因转录时与含有编码信息的外显子 (exon) 一起被拷贝。内含子通过剪切和连接反应从前体信使 RNA (pre-mRNA) 中去除。未经剪接的前体 RNA 形成核 RNA 组分，称为核不均一 RNA (heterogenous nuclear RNA, hnRNA)。
- **切割** (cutting events) 在 rRNA 和 tRNA 的加工过程中尤为重要，许多这类 rRNA 和 tRNA 首先从可产生一种以上分子的转录单位合成，这些前体-rRNA (pre-



rRNA) 和**前体-tRNA** (pre-tRNA) 必须经过切割以产生成熟的 RNA。这种类型的加工在原核和真核细胞中都存在。

- **化学修饰** (chemical modification) 在 rRNA、tRNA 和 mRNA 中都存在。所有生物体的 rRNA 和 tRNA 都可以通过加上新的化学基团而被修饰, 这些基团被加到 RNA 的特定核苷酸上。mRNA 的化学修饰, 称作 **RNA 编辑** (RNA editing), 在多种真核细胞中存在。

## 1.2.4 转录组

转录组虽然不到细胞总 RNA 的 4%, 却是细胞最重要的组分, 因为它包含了基因组表达的下一个阶段中所要使用的编码 RNA。值得注意且非常重要的一点是, 转录组从不从头 (*de novo*) 合成。一个细胞通过细胞分裂诞生时就接受了其上一代的部分转录组, 并维持一生。细菌孢子和植物种子这样的静息细胞也有转录组, 不过其转录组到蛋白质的翻译可能被完全关闭。所以, 各蛋白质编码基因的转录过程并不是导致转录组的合成, 而是通过替换被降解的 mRNA 来维持转录组, 并通过开闭不同的基因来改变转录组的组成。

即使在细菌和酵母这些最简单的生物体中, 在同一时间也存在许多活跃的基因。因而转录组十分复杂, 包含成百或上千种不同的 mRNA。通常, 每种 mRNA 只占 mRNA 总量的一小部分, 即使最常见的类型也仅占不到总 mRNA 1%。但有些细胞是特例, 它们有高度特化的生化机能, 转录组中有一种或几种 mRNA 占显著优势。例如, 发育中的小麦种子, 合成大量的麦醇溶蛋白聚积在休眠的谷粒中, 为发芽提供氨基酸来源。发育的种子中麸朊蛋白 mRNA 可占细胞转录物组的 30%。

## 1.3 蛋白质和蛋白质组

基因组表达的第二个产物是蛋白质组——细胞内那些决定细胞所能进行的生化反应性质的所有蛋白质组成成分 (图 1.2)。这些蛋白质是通过**翻译** (translation) 那些组成转录组的 mRNA 分子而合成的。

### 1.3.1 蛋白质的结构

蛋白质和 DNA 分子一样, 是一个线性的无分支的多聚体。蛋白质中的单体亚单位称为**氨基酸** (amino acid, 图 1.14)。氨基酸形成的多聚体或**多肽** (polypeptide) 在长度上很少超过 2000 个单位。和 DNA 一样, 蛋白质结构的主要特征在 20 世纪头 50 年已经被确定, 20 世纪 40 年代和 50 年代初 Pauling 和 Corey 对多肽主要构象或**二级结构** (secondary structure) 的阐明是该时期蛋白质生化研究的代表。近年来的研究热点集中在这些二级结构是如何组合产生蛋白质复杂的三维结构的。

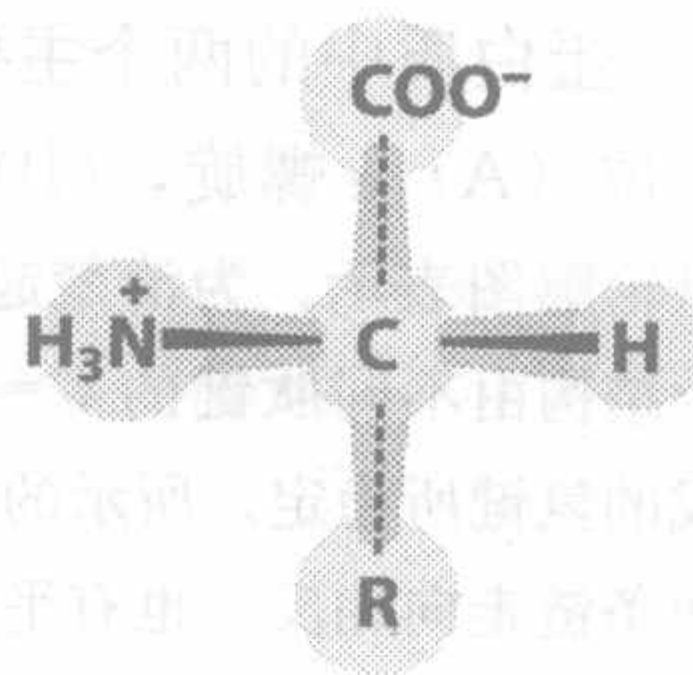


图 1.14 氨基酸的一般结构  
所有氨基酸大体的结构都相同, 由一个中心  $\alpha$  碳与一个氢原子、一个羧基、一个氨基和一个 R 基相连。R 基在各种氨基酸中是不同的 (图 1.18)



## 蛋白质结构的四个水平

传统上认为蛋白质有四个不同水平的结构。这些水平具有层级性，逐级地组装成蛋白质，每一个水平的结构都取决于低一等级的结构。

- **蛋白质的一级结构** (primary structure) 是由氨基酸通过**肽键** (peptide bond) 连接成一条多肽链形成，由一个氨基酸的羧基和另一个氨基酸的氨基进行缩合反应产生 (图 1.15)。注意，与多核苷酸一样，多肽链两端的化学性质不同：一端有一个自由氨基基团，称为**氨基** (amino、 $-\text{NH}_2$ ) 或 **N 端** (N terminus)；另一端有一个自由羧基基团，称为**羧基** (carboxyl、 $-\text{COOH}$ ) 或 **C 端** (C terminus)。故多肽链的方向可表示为  $\text{N} \rightarrow \text{C}$  (图 1.15 中从左到右) 或  $\text{C} \rightarrow \text{N}$  (图 1.15 中从右到左)。
- **二级结构** (secondary structure) 指多肽采取的不同构象。二级结构的两种主要形式是  **$\alpha$  螺旋** ( $\alpha$ -helix) 和  **$\beta$  片层** ( $\beta$ -sheet, 图 1.16)，二者都由多肽的不同氨基酸之间形成的氢键所稳定。大部分多肽的长度足以折叠成一系列依次相连的二级结构。

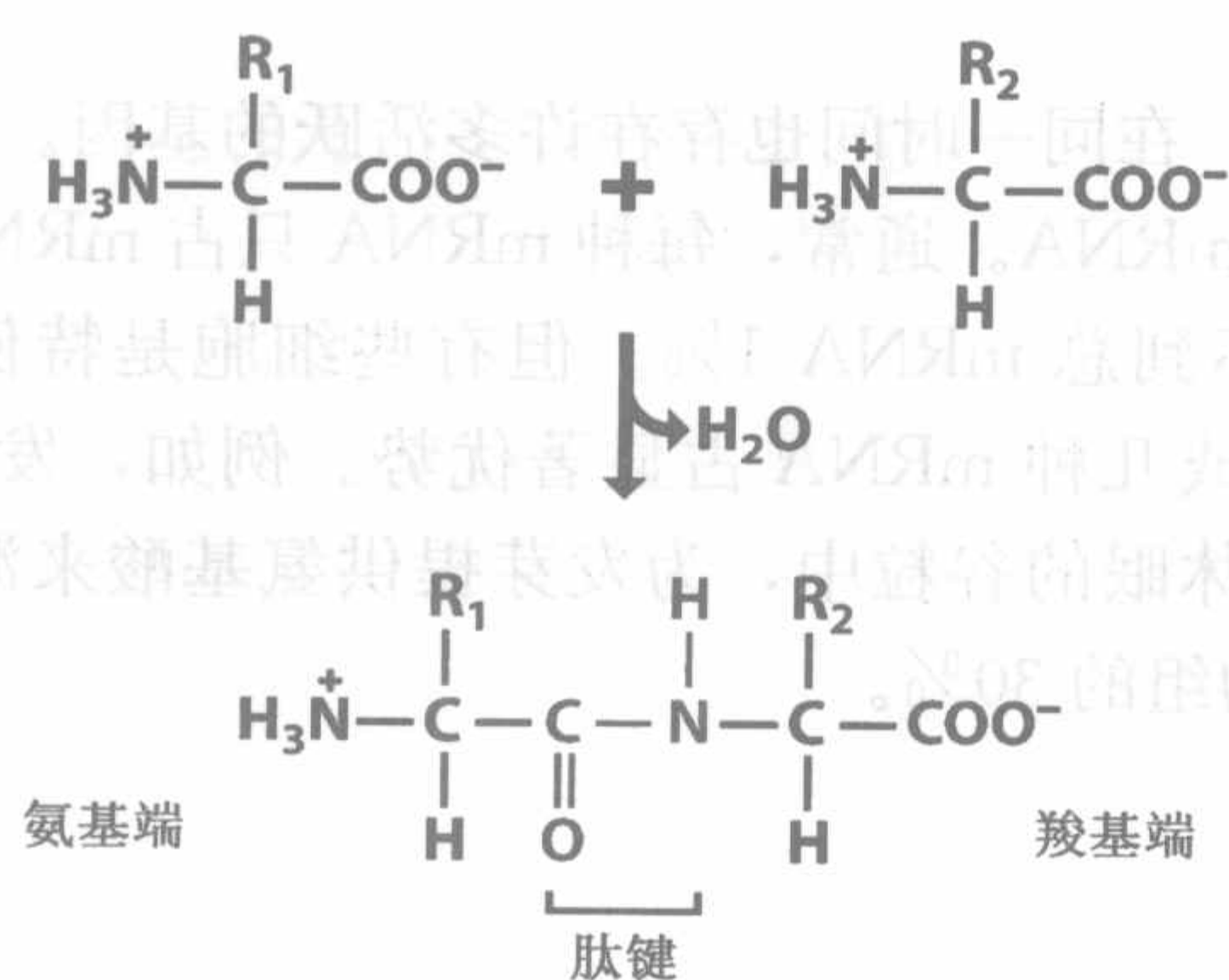
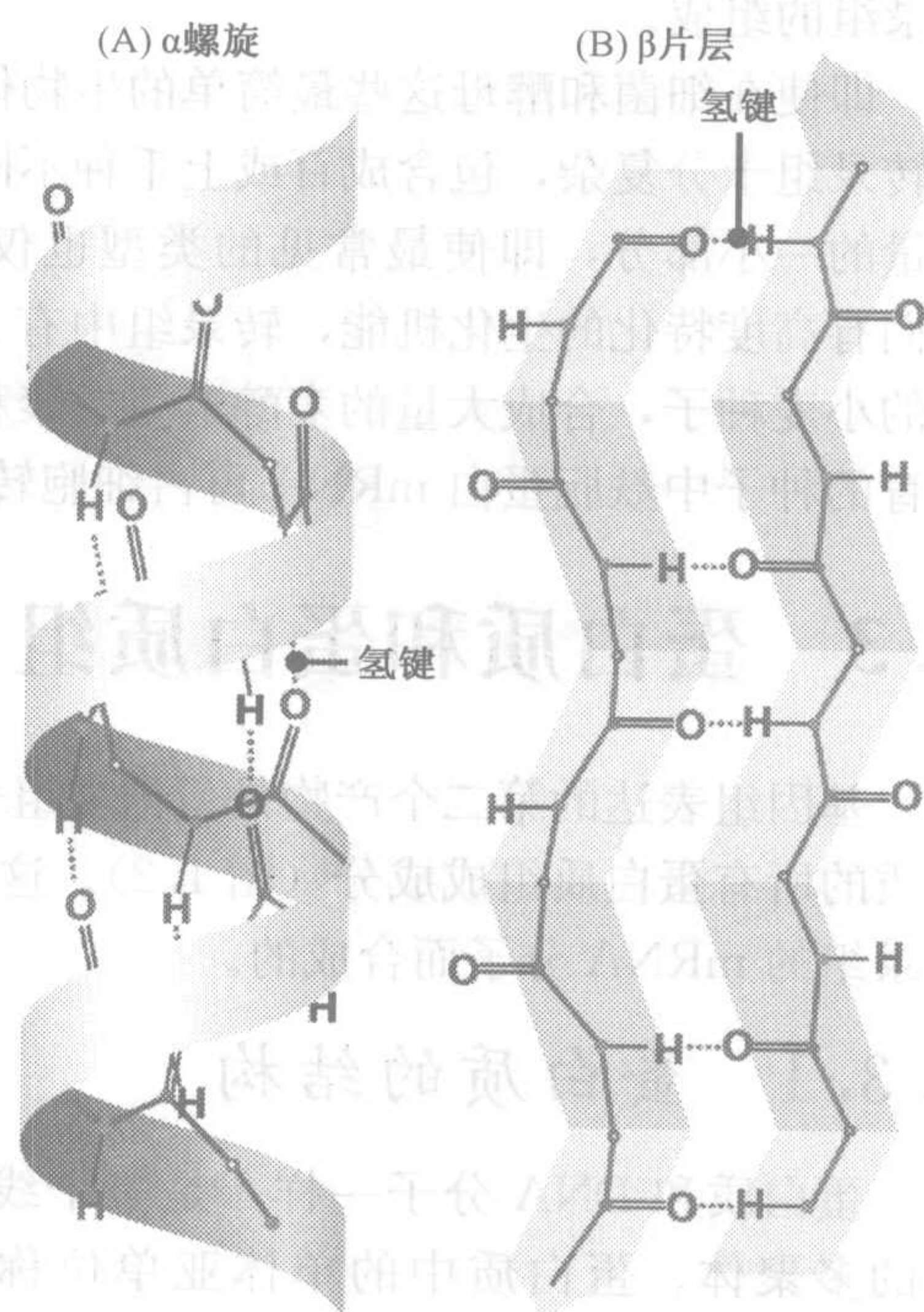


图 1.15 多肽中，氨基酸由肽键相连

图示的化学反应中两个氨基酸由一个肽键连接起来。此反应去除了一分子的水，故为缩合反应

图 1.16 蛋白质中的两个主要的二级结构单位 (A)  $\alpha$  螺旋, (B)  $\beta$  片层

多肽链以轮廓图表示。为清楚起见，将 R 基团略去。每个结构由不同肽键的  $\text{C}=\text{O}$  和  $\text{N}-\text{H}$  基团之间形成的氢键所稳定。所示的  $\beta$  片层构象是反平行的，两条链走向相反。也有平行  $\beta$  片层结构存在



- **三级结构** (tertiary structure) 是将多肽链的二级结构组分折叠成为三维构型而形成的 (图 1.17)。它被各种化学力所稳定，主要是氨基酸残基之间的氢键和带电荷的氨基酸 R 基团之间的静电相互作用 (图 1.18)，以及疏水相互作用，即带有非极性侧链基团的氨基酸 (即“疏水”) 必须埋在蛋白质内部以便与水隔绝。另外多肽链不同位置上的半胱氨酸残基之间可以产生共价连接，即**二硫键** (disulfide bridge)。



• **四级结构** (quaternary structure) 即两条或更多已形成三级结构的多肽链组合在一起形成一个多亚基蛋白质。不是所有蛋白质都有四级结构，但它是多种具有复杂功能的蛋白质的特征，包括几种参与基因组表达的蛋白质。一些四级结构通过不同多肽间的二硫键结合在一起，形成稳定的不会被轻易解离的多亚基蛋白质。但很多蛋白质由氢键和疏水作用来稳定，亚基间较松散地结合在一起，从而可以依据细胞功能的需要复原为组成它们的肽链，或者是改变其亚基的组成。

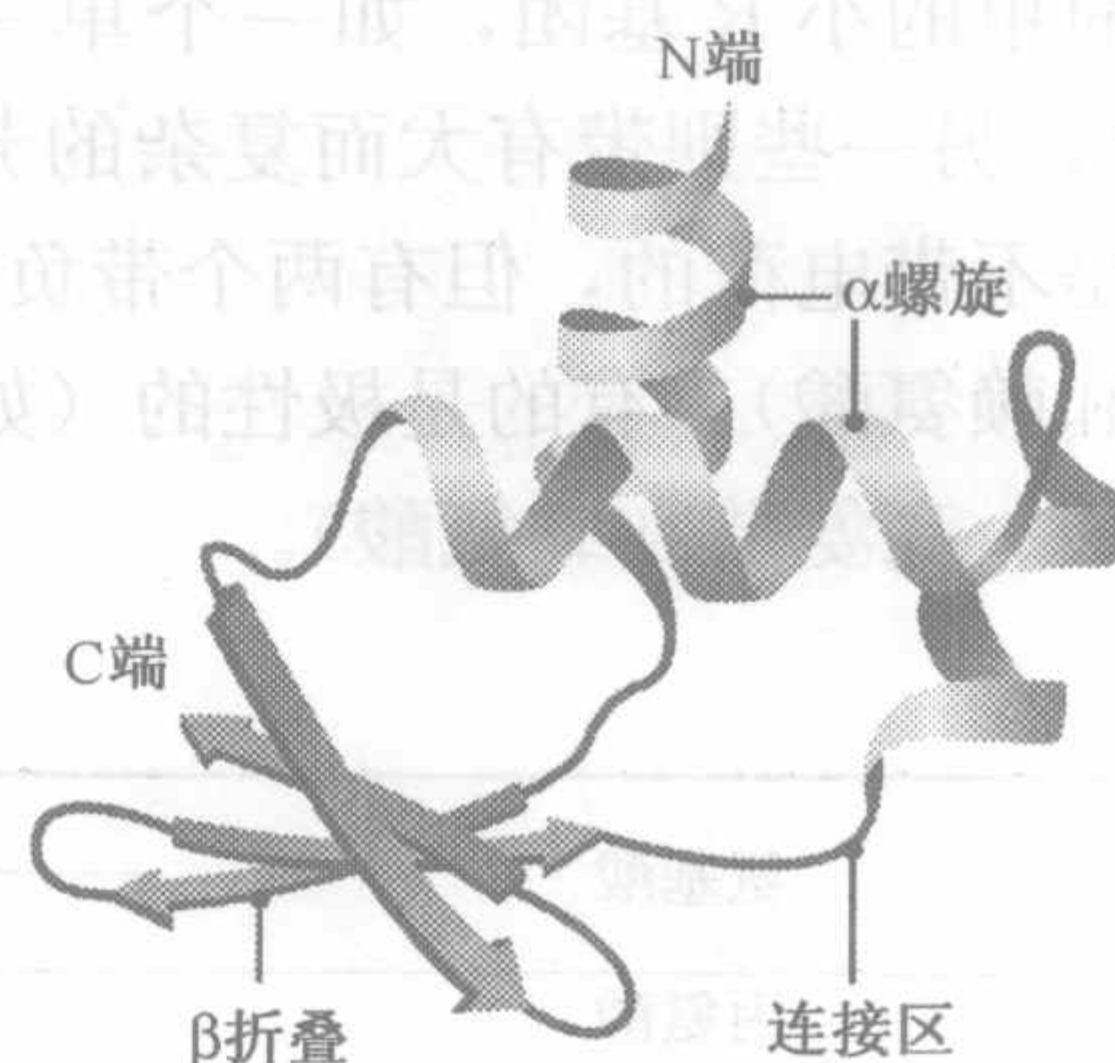


图 1.17 蛋白质的三级结构

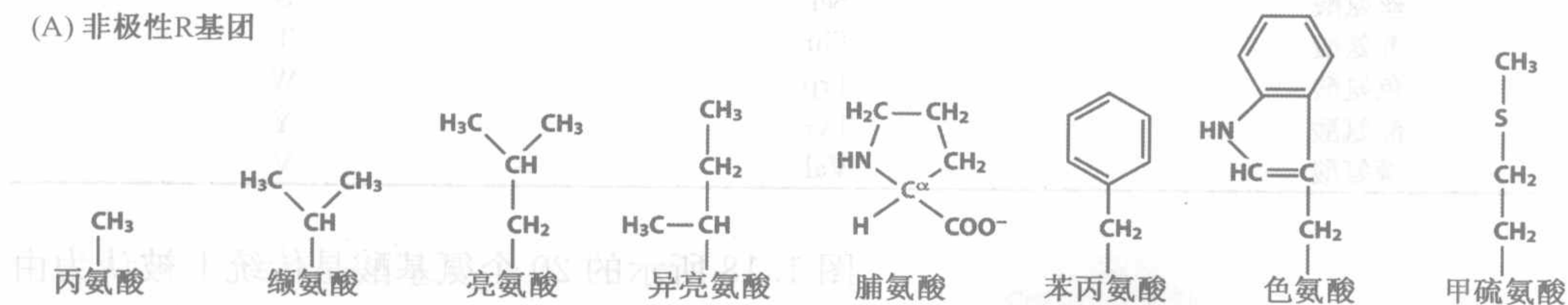
这个假想的蛋白质由三个  $\alpha$  螺旋（以卷曲表示）和一个四条链的  $\beta$  片层（以箭头表示）组成

## 蛋白质的多样性取决于氨基酸的多样性

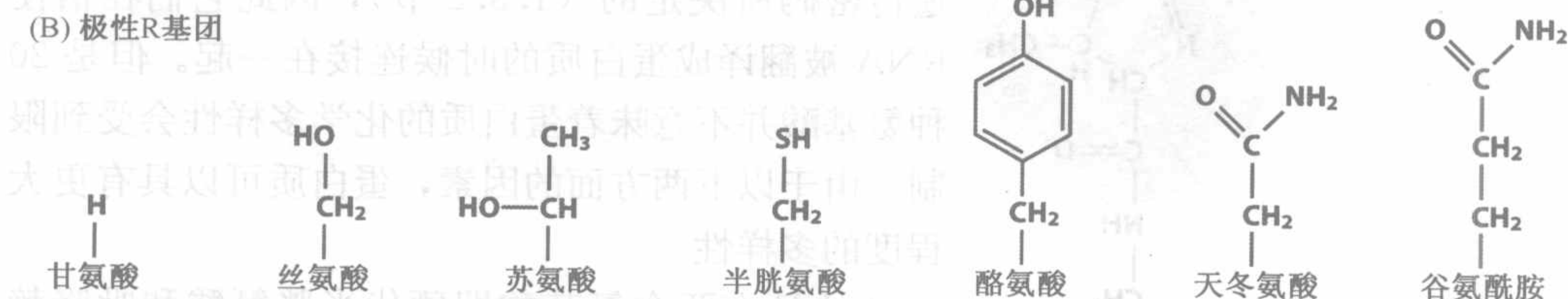
组成蛋白质的氨基酸在化学性质上有多多样性，因此蛋白质的功能也是多种多样的。不同的氨基酸顺序导致了化学活性的不同组合，这些组合不仅决定着产生的蛋白质的整体结构，而且也指导了该蛋白质的表面上决定其化学性质的活性基团的定位。

氨基酸的多样性源于 R 基团，因为这部分在每一种氨基酸中是不同的，结构上差异也很大。蛋白质由 20 种氨基酸组成（图 1.18，表 1.2）。其中一些氨基酸含有结构相

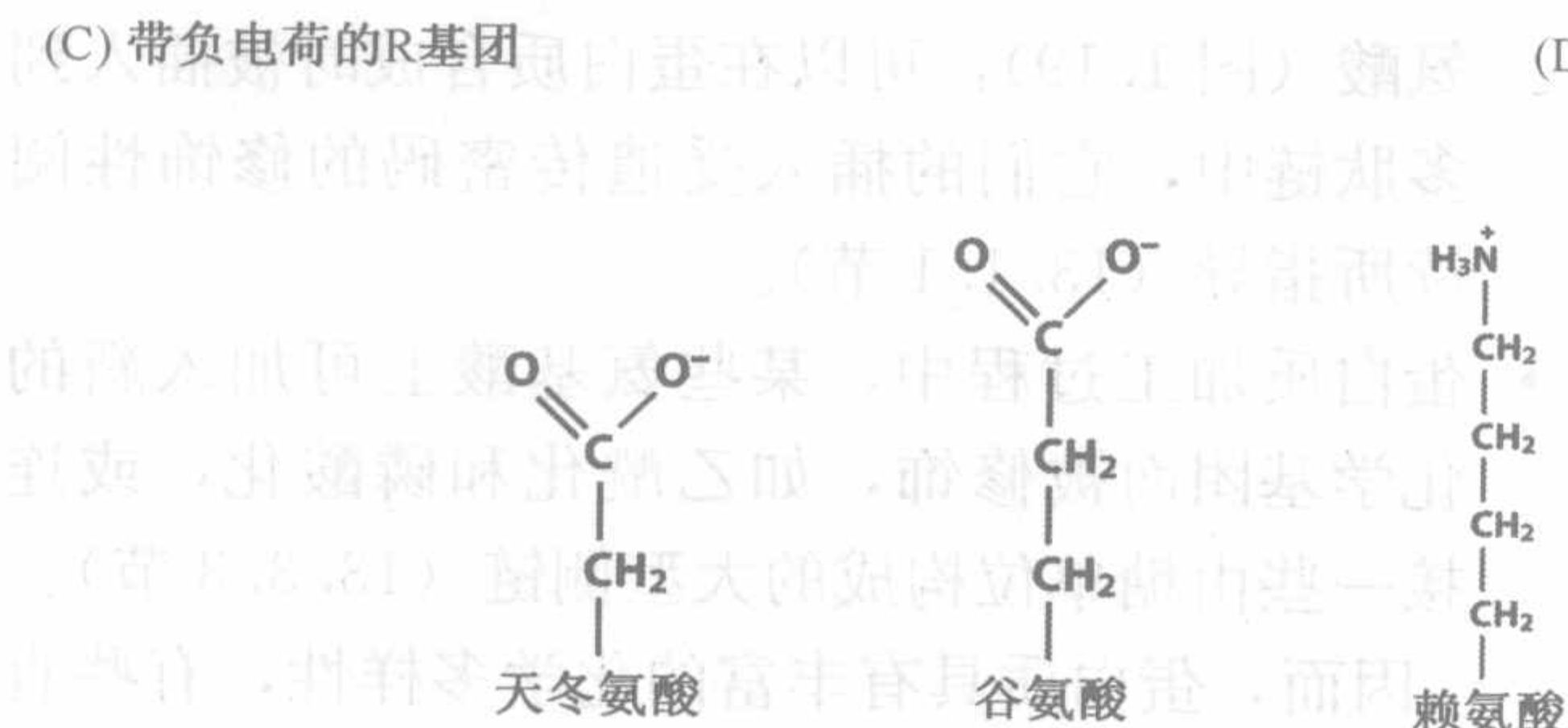
(A) 非极性R基团



(B) 极性R基团



(C) 带负电荷的R基团



(D) 带正电荷的R基团

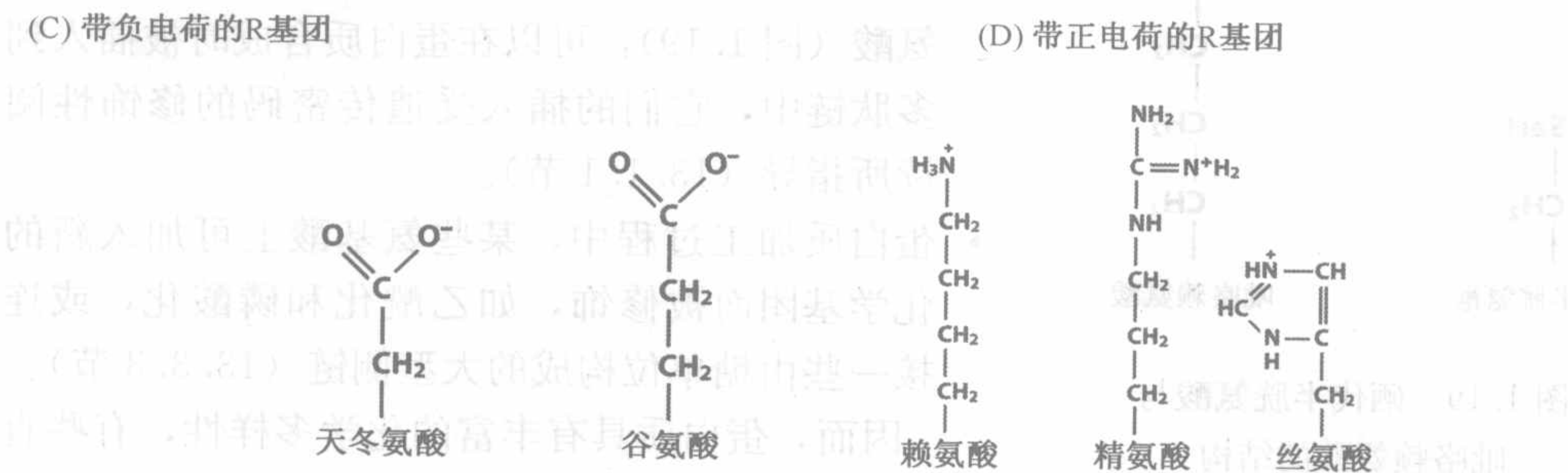


图 1.18 氨基酸的 R 基团

这 20 种氨基酸是传统上认为的由遗传密码决定的氨基酸



对简单的小 R 基团，如一个单一的氢原子（其氨基酸为甘氨酸）或一个甲基（丙氨酸），另一些则带有大而复杂的芳香族侧链（苯丙氨酸、色氨酸和酪氨酸）。多数 R 基团是不带电荷的，但有两个带负电（天冬氨酸和谷氨酸），三个带正电（精氨酸、组氨酸和赖氨酸）。有的是极性的（如甘氨酸、丝氨酸和苏氨酸），另一些是非极性的（如丙氨酸、亮氨酸和缬氨酸）。

表 1.2 氨基酸缩写

氨基酸	缩写	
	三个字母	一个字母
丙氨酸	Ala	A
精氨酸	Arg	R
天冬酰胺	Asn	N
天冬氨酸	Asp	D
半胱氨酸	Cys	C
谷氨酸	Glu	E
谷氨酰胺	Gln	Q
甘氨酸	Gly	G
组氨酸	His	H
异亮氨酸	Ile	I
亮氨酸	Leu	L
赖氨酸	Lys	K
甲硫氨酸	Met	M
苯丙氨酸	Phe	F
脯氨酸	Pro	P
丝氨酸	Ser	S
苏氨酸	Thr	T
色氨酸	Trp	W
酪氨酸	Tyr	Y
缬氨酸	Val	V

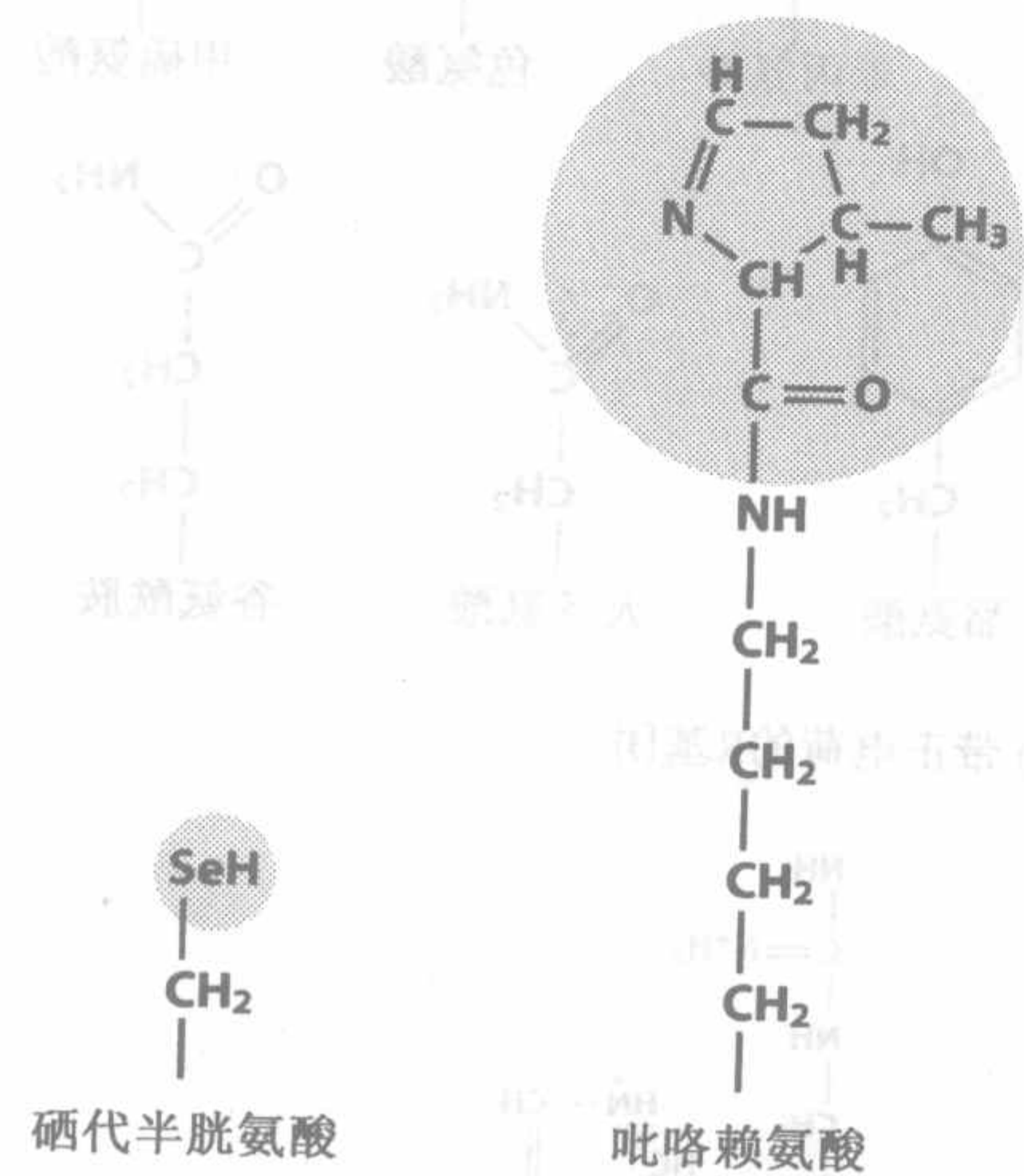


图 1.19 硒代半胱氨酸与吡咯赖氨酸的结构  
圆圈部分显示的分别是这两个氨基酸与半胱氨酸和赖氨酸的差异

图 1.18 所示的 20 个氨基酸是传统上被认为由遗传密码所决定的（1.3.2 节），因此它们在信使 RNA 被翻译成蛋白质的时候连接在一起。但是 20 种氨基酸并不意味着蛋白质的化学多样性会受到限制。由于以下两方面的因素，蛋白质可以具有更大程度的多样性。

- 至少还有两个氨基酸即硒代半胱氨酸和吡咯赖氨酸（图 1.19），可以在蛋白质合成时被插入到多肽链中，它们的插入受遗传密码的修饰性阅读所指导（13.1.1 节）。
- 蛋白质加工过程中，某些氨基酸上可加入新的化学基团而被修饰，如乙酰化和磷酸化，或连接一些由糖单位构成的大型侧链（13.3.3 节）。

因而，蛋白质具有丰富的化学多样性，有些直接由基因组决定，其余的通过蛋白质的加工产生。



### 1.3.2 蛋白质组

蛋白质组包括了在特定时间存在于细胞中的所有蛋白质。一个“典型”的哺乳动物细胞，如肝细胞中，含有 10 000~20 000 种不同的蛋白质，共有约  $8 \times 10^9$  个蛋白质分子，重约 0.5ng，占细胞总重量的 18%~20%。各种蛋白质的拷贝数差别很大，最少的每个细胞中不足 20 000 个，最多的可达 1 亿个。每个细胞中拷贝数大于 50 000 的蛋白质被认为其含量较丰富，通常哺乳动物细胞中约有 2000 种蛋白质属于此类。研究不同类型的哺乳动物细胞的蛋白质组后发现，这些丰富蛋白质的差异很小，说明它们大多属于管家 (housekeeping) 蛋白，执行所有细胞中普遍的生化活动。而那些使细胞具有特化功能的蛋白质含量通常极少。不过也有特例，如只出现在红细胞中的血红蛋白含量很高。

### 转录组和蛋白质组的联系

信息通过转录从 DNA 到 RNA 在概念上并不难理解。DNA 和 RNA 结构相似，一个基因的 RNA 拷贝通过模板依赖的方式采用我们所熟知的碱基配对原则合成是容易理解的。但在基因组表达的第二阶段，转录组的 mRNA 分子指导蛋白质的合成，如果仅考虑分子结构就不那么容易理解。在 20 世纪 50 年代早期，DNA 的双螺旋结构发现后不久，几个分子生物学家试图解释氨基酸按顺序结合到 mRNA 分子上的机制，但所有这些方案中至少有一些键比实际的物理化学定律所允许的键长或短，因而不久就销声匿迹了。直到 1957 年，Francis Crick 预测 mRNA 及其合成的多肽间存在着将二者联系起来的接头分子才打破了这种僵局。此后不久，人们意识到 tRNA 就是这种接头分子，这个事实一旦被确立，关于蛋白质合成机制的详细了解就建立起来了。我们将在 13.1 节探讨这一过程。

20 世纪 50 年代分子生物学家所感兴趣的蛋白质合成的另一方面是信息问题 (informational problem)，它反映了在转录组和蛋白质组之间联系的第二个重要成分——遗传密码 (genetic code)，它特异地决定一个 mRNA 核苷酸序列如何被翻译成一个蛋白质的氨基酸序列。20 世纪 50 年代人们就认识到需要用三联体遗传密码，每个密码子 (codon) 由三个核苷酸组成，来解读蛋白质中的所有 20 种氨基酸。2 个字母的密码只有  $4^2=16$  个密码子，不足以解读 20 种氨基酸；而 3 个字母的密码会产生  $4^3=64$  种密码子。遗传密码是在 20 世纪 60 年代被解开的，部分工作是通过分析已知的或可预知其序列的人工 mRNA 在无细胞蛋白质合成体系中翻译出的肽链来确定的。另一部分是通过分析在纯化的核糖体中 (细胞内负责蛋白质合成的蛋白质-RNA 复合体) 哪个氨基酸与哪个 mRNA 序列结合是确定的。遗传密码的工作完成后人们认识到，64 个密码子可分为不同的组，每组成员编码同一种氨基酸 (图 1.20)。色氨酸和甲硫氨酸各仅有一个密码子，其他氨基酸都有 2、3、4 或 6 个密码子。密码的这一特性叫做简并性 (degeneracy)。密码中有 4 个标点密码子 (punctuation codon)，用于表明 mRNA 中核苷酸序列翻译起始和终止的位置 (图 1.21)。起始密码子 (initiation codon) 通常为 5'-AUG-3'，它也是甲硫氨酸的密码子 (所以多数新合成的多肽以甲硫氨酸开始)，尽管在有的 mRNA 中也使用其他起始密码子，如 5'-GUG-3' 和 5'-UUG-3'。三个终止密码子



(termination codon) 分别为 5'-UAG-3'、5'-UAA-3'和 5'-UGA-3'。

UUU	phe	UCU	ser	UAU	tyr	UGU	cys
UUC		UCC		UAC		UGC	
UUA	leu	UCA		UAA	stop	UGA	stop
UUG		UCG		UAG		UGG	trp
CUU		CCU	pro	CAU	his	CGU	arg
CUC	leu	CCC		CAC		CGC	
CUA		CCA		CAA	gln	CGA	
CUG		CCG		CAG		CGG	
AUU	ile	ACU	thr	AAU	asn	AGU	ser
AUC		ACC		AAC		AGC	
AUA		ACA		AAA	lys	AGA	arg
AUG	met	ACG		AAG		AGG	
GUU	val	GCU	ala	GAU	asp	GGU	gly
GUC		GCC		GAC		GGC	
GUA		GCA		GAA	glu	GGA	
GUG		GCG		GAG		GGG	

图 1.20 遗传密码

氨基酸以标准的三个字母缩写表示，请参见表 1.2

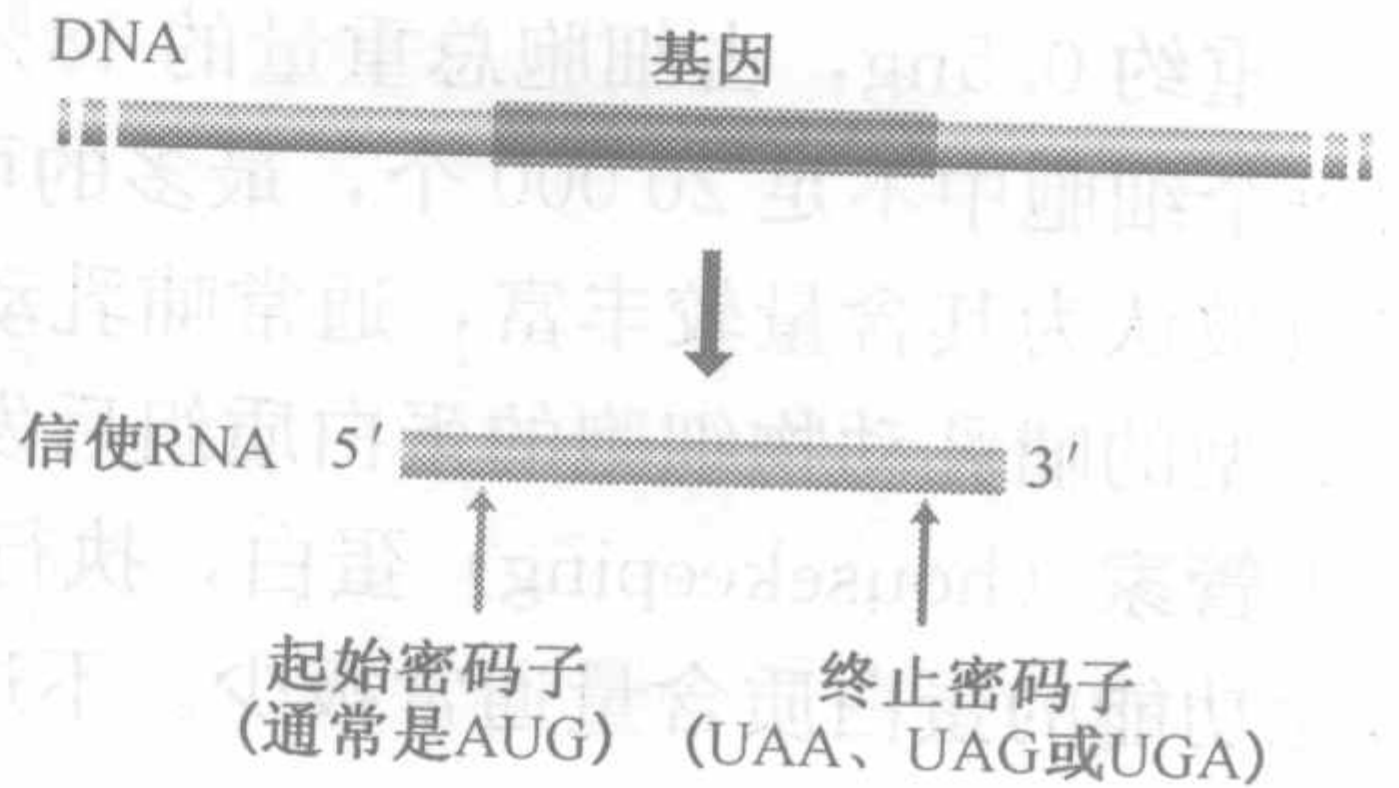


图 1.21 mRNA 中标点密码子的位置

### 遗传密码并不通用

起初认为遗传密码在所有生物中都相同。该观点认为密码子一旦建立就不会改变，因为给任何一个密码子赋予新的含义都会导致蛋白质氨基酸序列的全面混乱。这种推论看似有道理，所以发现事实上遗传密码并不通用是令人吃惊的。图 1.20 中所示的密码子适用于大多数生物体中的大多数基因，但偏差也广泛存在。尤其是线粒体基因组，通常使用非标准的密码（表 1.3）。这种现象最初于 1979 年由英国剑桥大学 Frederick Sanger 小组发现。他们发现若干人线粒体基因在并不终止翻译的内在位点上有 5'-UGA-3'密码子，而 5'-UGA-3'通常为终止密码子。比较这些基因编码的蛋白质的氨基酸顺序表明，5'-UGA-3'在人线粒体中是一个色氨酸密码子，这仅是这种特殊遗传系统中的四种密码偏差之一。其他生物体的线粒体基因也表现出密码偏差，虽然至少其中一种，即植物线粒体用 5'-CGG-3'作为色氨酸密码子，很可能在翻译之前由 RNA 编辑（12.2.5 节）校正。

表 1.3 标准遗传密码的偏离示例

生物	密码子	应该编码	实际编码
<b>线粒体基因组</b>			
哺乳动物	UGA	终止	Trp
	AGA, AGG	Arg	终止
	AUA	Ile	Met
果蝇 ( <i>Drosophila</i> )	UGA	终止	Trp
	AGA	Arg	Ser
	AUA	Ile	Met
啤酒酵母 ( <i>Saccharomyces cerevisiae</i> )	UGA	终止	Trp
	CUN	Leu	Thr



续表

生物	密码子	应该编码	实际编码
真菌	AUA	Ile	Met
	UGA	终止	Trp
	CGG	Arg	Trp
玉米			
核和原核基因组	UAA, UAG	终止	Gln
一些原生动物	CUG	Leu	Ser
<i>Candida cylindracea</i>	AGA	Arg	终止
<i>Micrococcus</i> sp.	AUA	Ile	终止
	UGA	终止	Cys
<i>Euplotes</i> sp.	UGA	终止	Trp
<i>Mycoplasma</i> sp.	CGG	Arg	终止
上下文依赖性密码子重分配	UGA	终止	硒代半胱氨酸
多种古生菌	UAG	终止	吡咯赖氨酸

缩写：N，代表任意核苷酸

非标准密码子也存在于低等真核生物的核基因组中。通常密码子的修饰仅限于一小部分生物体中，而且往往包括终止密码子的重新分配（表 1.3）。虽然在原核生物中修饰并不多见，但支原体就是一个已知的例子。一类更重要的密码子变异是上下文依赖的**密码子重分配**（context-dependent codon reassignment），出现在所要合成的蛋白质含有硒代半胱氨酸或吡咯赖氨酸时。含有吡咯赖氨酸的蛋白质很罕见，可能只存在于叫做古生菌的原核生物中（第 8 章），但是硒蛋白质在许多生物中普遍存在，一个例子是能保护人类或哺乳动物细胞免受氧化损伤的谷胱甘肽过氧化物酶。硒代半胱氨酸由 5'-UGA-3' 编码而吡咯赖氨酸由 5'-UAG-3' 编码，因为它们还被同一生物用作终止密码子，因而这些密码子有双重含义（表 1.3）。硒代半胱氨酸的密码子 5'-UGA-3' 与真正的终止密码子的区别在于其 mRNA 中有一个发夹环结构，该结构位于紧靠原核生物的硒代半胱氨酸密码子的下游，或真核生物的 3' 非翻译区（即终止密码子后的 mRNA 部分）。对这一密码子的识别需要发夹结构与涉及这些 mRNA 翻译的特殊蛋白质之间的相互作用。类似的系统可能在翻译吡咯赖氨酸时起效。

## 蛋白质组和细胞生化功能之间的联系

基因组编码的生物学信息最终由蛋白质表现，而蛋白质的生物学性质由折叠结构及表面的化学基团的空间分布决定。通过决定不同类型的蛋白质合成，基因组得以构建和维持蛋白质组，蛋白质组整体的生物学性质组成了生命的基础。蛋白质组能扮演这样的角色是因为其极大的可形成蛋白质结构的多样性，使蛋白质能够执行多种生物学功能，包括：

- 生化催化作用（biochemical catalysis）是一种叫做酶的特殊蛋白质所具有的功能。给细胞提供能量的核心代谢途径，以及合成核酸、蛋白质、碳水化合物和脂类的生物



合成过程都是由酶催化的。生化催化作用还通过酶的活动，如 RNA 聚合酶，驱使基因组的表达。

- 结构 (structure) 在细胞水平是由组成细胞骨架的蛋白质决定的，某些胞外的蛋白质也主要起结构功能，如胶原，是骨骼和肌腱的重要组分。
- 运动 (movement) 通过收缩蛋白实现，最著名的例子是细胞骨架纤维中的肌动蛋白和肌球蛋白。
- 运输 (transport) 身体周围的物质是一种重要的蛋白质活动，如血红蛋白运输血流中的氧，血清白蛋白运输脂肪酸。
- 细胞进程的调节是通过诸如 STAT (signal transducers and activators of transcription) 这样的信号蛋白 (14.1.2 节)，以及像活化调节因子 (activator) 这样能结合到基因组上并影响单个或一群基因表达水平的蛋白质 (11.3 节) 介导的。调节和协调细胞群活动的细胞外激素和细胞因子很多都是蛋白质 (如控制血糖水平的胰岛素和白细胞介素调节细胞分裂和分化的一组细胞因子)。
- 保护身体和细胞个体是一定范围的蛋白质的功能，包括抗体以及参与凝血反应的蛋白质。
- 储存功能是由蛋白质执行的，如铁蛋白在肝脏内储铁，麦醇溶蛋白在休眠的小麦种子中储存氨基酸。

蛋白质功能的多样性使蛋白质组有能力将基因组中包含的信息蓝图转变成生命过程中实质的要素。

## 总结

基因组是地球上每一物种具有的生物学信息的存储库。绝大部分的基因组是由 DNA 构成的，极少的例外是那些具有 RNA 基因组的病毒。基因组的表达是将其包含的信息释放到细胞中的过程。基因组表达的第一个产物是转录组，即由那些含有细胞在特定时间所需生物信息的、编码蛋白质的基因衍生而来的 RNA 分子的集合。第二个产物是蛋白质组，即细胞中那些决定细胞能够进行的生化反应性质的所有蛋白质的组分。证明基因由 DNA 构成的实验学证据最初是在 1945 年和 1952 年取得的，但是 1953 年 Watson 和 Crick 发现的 DNA 双螺旋结构才使得生物学家确信 DNA 就是遗传物质。DNA 多聚核苷酸链是一种没有分支的多聚体，由多拷贝的四种化学性质不同的核苷酸组成。在双螺旋中，两条多聚核苷酸链互相缠绕，核苷酸的碱基位于分子的内部。多聚核苷酸通过碱基间的氢键连接，且通常是 A 与 T 配对，G 与 C 配对。RNA 也是一种多聚核苷酸，但是其中的单个核苷酸相对于 DNA 中的核苷酸结构不同，而且通常是单链。依赖 DNA 的 RNA 聚合酶通过叫做转录的过程负责将基因拷贝为 RNA，从而不仅产生转录组，而且也产生一批功能性 RNA (即一些不编码蛋白质但是在细胞中仍起到重要作用的 RNA)。许多 RNA 最初是以前体分子的形式被合成的，之后通过剪切、连接反应以及化学修饰而变为成熟的形式。蛋白质也是没有分支的多聚体，但是其组成单元是通过肽键连接的氨基酸。氨基酸序列构成蛋白质的一级结构。更高级的结构水平 (二级、三级和四级) 是通过一级结构折叠为三维构象以及单个多肽间的相互作用形成



多蛋白质结构来形成。蛋白质功能上是多样化的,因为每一个氨基酸都具有不同的化学性质,当它们以不同方式被组合在一起的时候所产生的蛋白质就具有各种化学特性。蛋白质是依照某三个核苷酸特异地编码某个氨基酸的遗传密码原则来翻译信使 RNA 而产生的。遗传密码并不通用,在线粒体中和低等真核生物中有变动,而且某些密码子在单个基因中可能有两种不同的含义。

## 选择题

\* 奇数问题的答案见附录

- 1.1\* 下面关于生物的基因组的哪一项陈述是错误的?
- 基因组含有构建和维持活体生物的遗传学信息。
  - 多细胞生物的基因组是由 DNA 组成的。
  - 基因组不需要酶和蛋白质的活性就可以表达它的信息。
  - 真核基因组是由核及线粒体 DNA 构成的。
- 1.2 体细胞是指:
- 含有染色体的单倍体。
  - 生成配子。
  - 缺乏线粒体。
  - 含有双倍体染色体并构成绝大多数人类细胞。
- 1.3\* 细胞内的遗传信息流向是下面的哪一个?
- DNA 是被转录为 RNA,后者再被翻译为蛋白质。
  - DNA 被翻译为蛋白质,后者被转录为 RNA。
  - RNA 被转录为 DNA,后者被翻译为蛋白质。
  - 蛋白质被翻译为 RNA,后者转录为 DNA。
- 1.4 在 20 世纪早期人们认为蛋白质可能带有遗传信息。这是因为下面的哪一个原因?
- 染色体大致由等量的 DNA 和蛋白质组成。
  - 已知蛋白质是由 20 种不同的氨基酸构成而 DNA 仅仅由 4 种核苷酸构成。
  - 已知不同的蛋白质含有不同的序列,而所有 DNA 被认为都是一样的序列。
  - 以上全部。
- 1.5\* DNA 中是下面的哪种键连接单个核苷酸的?
- 糖苷键
  - 肽键
  - 磷酸二酯键
  - 静电
- 1.6 在解决 DNA 的结构问题时, Watson 和 Crick 灵活地使用了下述哪一个技术?
- 构建 DNA 分子模型以确信原子的准确定位。
  - DNA 的 X 射线晶体学。
  - 色谱研究来决定来自不同物种的核苷酸的相应组成。
  - 遗传学研究证实 DNA 是遗传物质。



- 1.7\* Erwin Chargaff 研究来自不同物种的 DNA 并证实了：
- DNA 是遗传物质。
  - RNA 转录自 DNA。
  - 某一给定的生物中的 A 和 T（或 G 与 C）的数量是一样的。
  - 双螺旋是通过碱基间的氢键维持的。
- 1.8 一个细胞的转录组的定义是：
- 一个细胞中所有 RNA 分子。
  - 一个细胞中编码蛋白质的 RNA 分子。
  - 一个细胞内核糖体 RNA 分子。
  - 一个细胞内转运 RNA 分子。
- 1.9\* 依赖于 DNA 的 RNA 聚合酶是如何进行 RNA 合成的？
- 它们使用 DNA 作为模板来使核苷酸多聚化。
  - 它们使用蛋白质作为模板来使核苷酸多聚化。
  - 它们使用 RNA 作为模板来使核苷酸多聚化。
  - 它们使核苷酸多聚化不需要模板。
- 1.10 哪种类型的功能性 RNA 是蛋白质合成所需结构的首要组成部分？
- 信使 RNA
  - 核糖体 RNA
  - 核小 RNA
  - 转运 RNA
- 1.11\* 一个细胞的蛋白质组被定义为：
- 一个细胞所能合成的所有蛋白质。
  - 细胞生命周期内所有出现的蛋白质。
  - 在某一时刻细胞内的所有蛋白质。
  - 在某一时刻细胞内活跃合成中的所有蛋白质。
- 1.12 蛋白质结构的哪一个水平描述了多亚基蛋白质的折叠构象？
- 一级结构
  - 二级结构
  - 三级结构
  - 四级结构
- 1.13\* 哪种类型的共价键在多肽链不同位置的半胱氨酸残基的连接中是重要的？
- 二硫键
  - 氢键
  - 肽键
  - 磷酸二酯键
- 1.14 一个细胞中大多数丰富蛋白质被认为是管家蛋白。它们的功能是什么？
- 它们负责各种细胞类型的特异功能。
  - 它们负责调控细胞内的基因组表达。
  - 它们负责从细胞中移除废物。



d. 它们负责细胞内发生的常规生物化学反应。

1.15\* 遗传密码子的简并性是指下面的哪一个?

a. 每一个密码子对应多个氨基酸。

b. 大多数氨基酸有多个密码子。

c. 有多个起始密码子。

d. 终止密码子也可以编码氨基酸。

1.16 下面哪一项不是蛋白质的生物学功能?

a. 生物催化

b. 调控细胞过程

c. 负载遗传信息

d. 在多细胞生物中转运分子

## 简答题

\* 奇数问题的答案见附录

1.1\* 给出一个发现 DNA、发现 DNA 是遗传物质、发现 DNA 的结构以及第一个基因组被描绘的发展时间线。

1.2 稳定双螺旋的是哪两种化学性相互作用?

1.3\* 为何 A 与 T, G 与 C 之间的特异性碱基配对为 DNA 复制提供了保真的基础?

1.4 RNA 和 DNA 之间两个重要的化学差异是什么?

1.5\* 为何信使 RNA 的半衰期相对其他 RNA 分子要短?

1.6 信使 RNA 的翻译是否与其从 DNA 模板上转录的形式一致?

1.7\* 细胞是否有时缺少转录组? 解释你答案的意义。

1.8 氢键、静电相互作用和疏水力是如何在蛋白质的二级、三级、四级结构中起到重要作用的?

1.9\* 仅由 20 种氨基酸合成的蛋白质是如何具有这么多结构与功能的多样性的?

1.10 除了 20 种氨基酸, 蛋白质还因为两个因素而具有其他化学多样性, 这两个因素是什么, 它们的重要性如何?

1.11\* 密码子 5'-UGA-3' 是如何既能作终止密码子又能作修饰后的氨基酸 (硒代半胱氨酸) 的密码子的?

1.12 基因组是如何指导一个细胞的生物学功能的?

## 论述题

\* 奇数问题的指导见附录

1.1\* 正文中讲到 Watson 和 Crick 在 1953 年 3 月 7 日发现 DNA 的双螺旋结构。请阐述。

1.2 讨论为何双螺旋立刻被广泛接受为 DNA 的正确结构?

1.3\* 什么实验使得遗传密码在 20 世纪 60 年代被阐明?

1.4 转录组和蛋白质组被分别视为基因组表达的一个中间产物和一个终末产物。请以我们对基因组表达的理解来评价上述用词的优缺点。

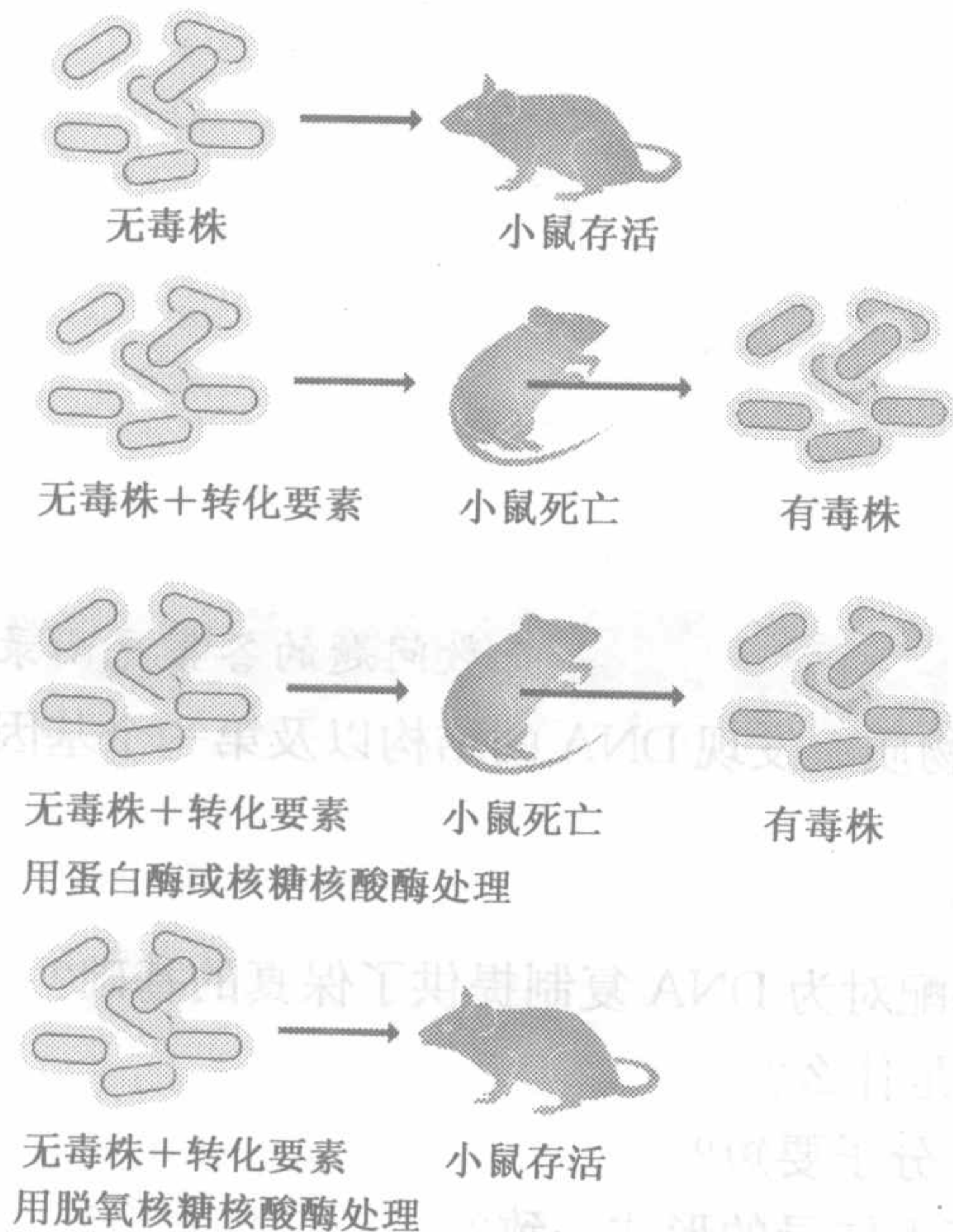


## 图形测试

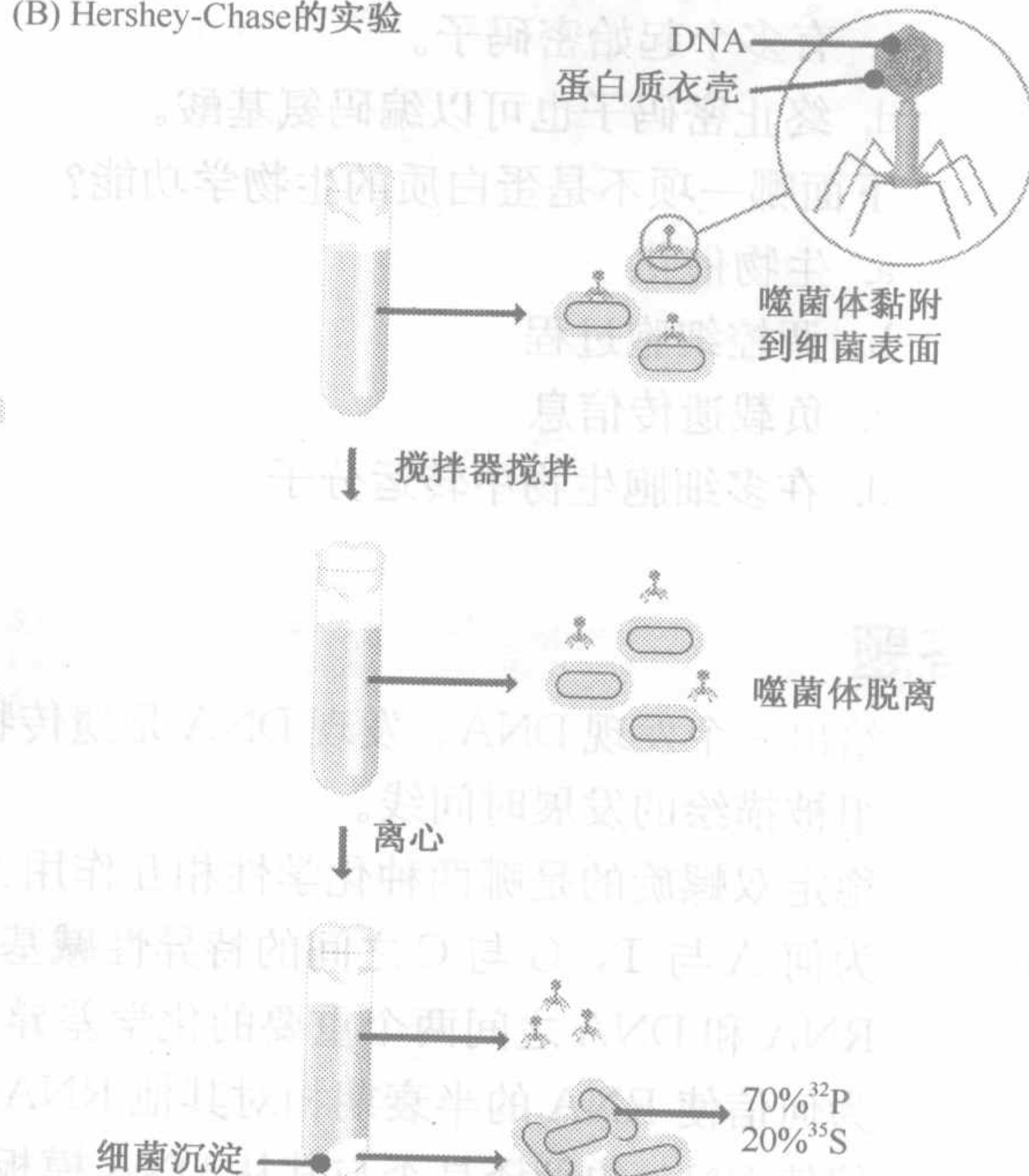
\* 奇数问题的答案见附录

1.1\* 讨论下述实验是如何有助于证明 DNA 而非蛋白质包含了遗传信息的。

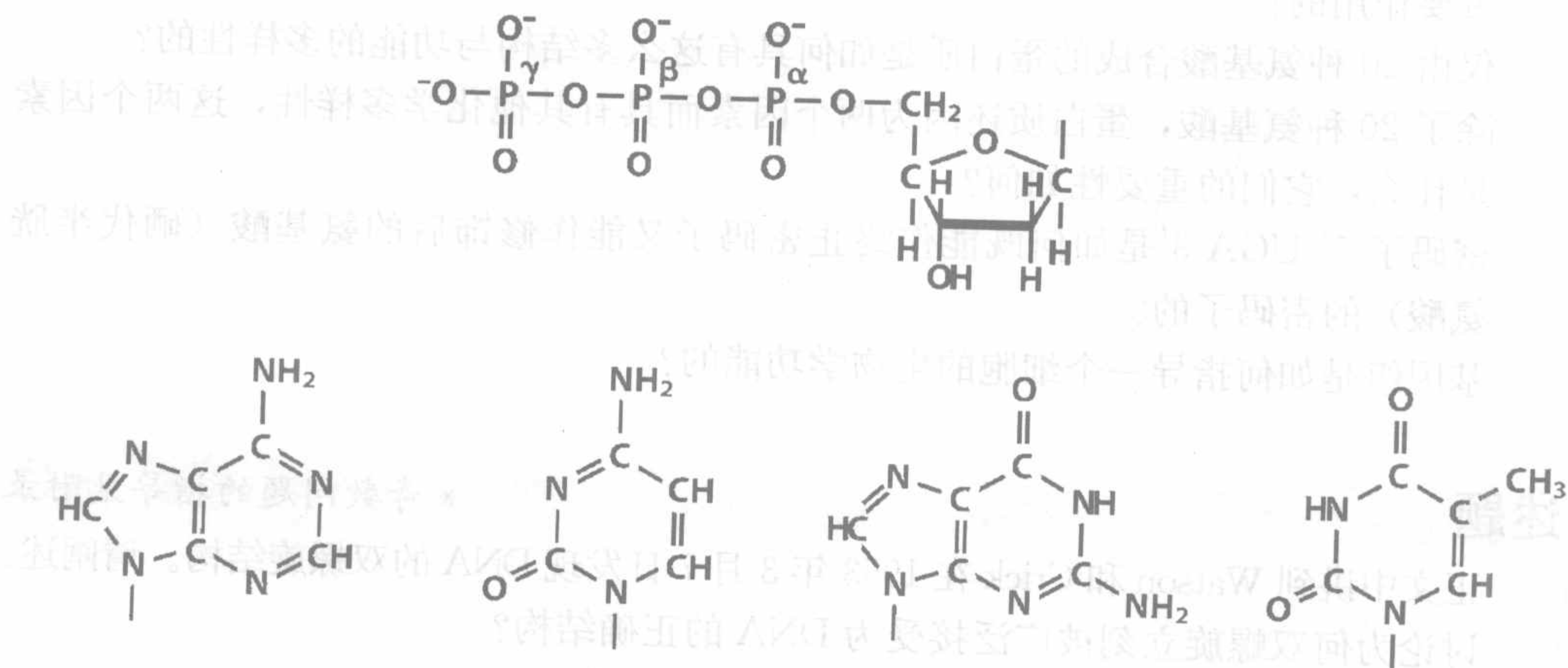
(A) 转化要素



(B) Hershey-Chase 的实验

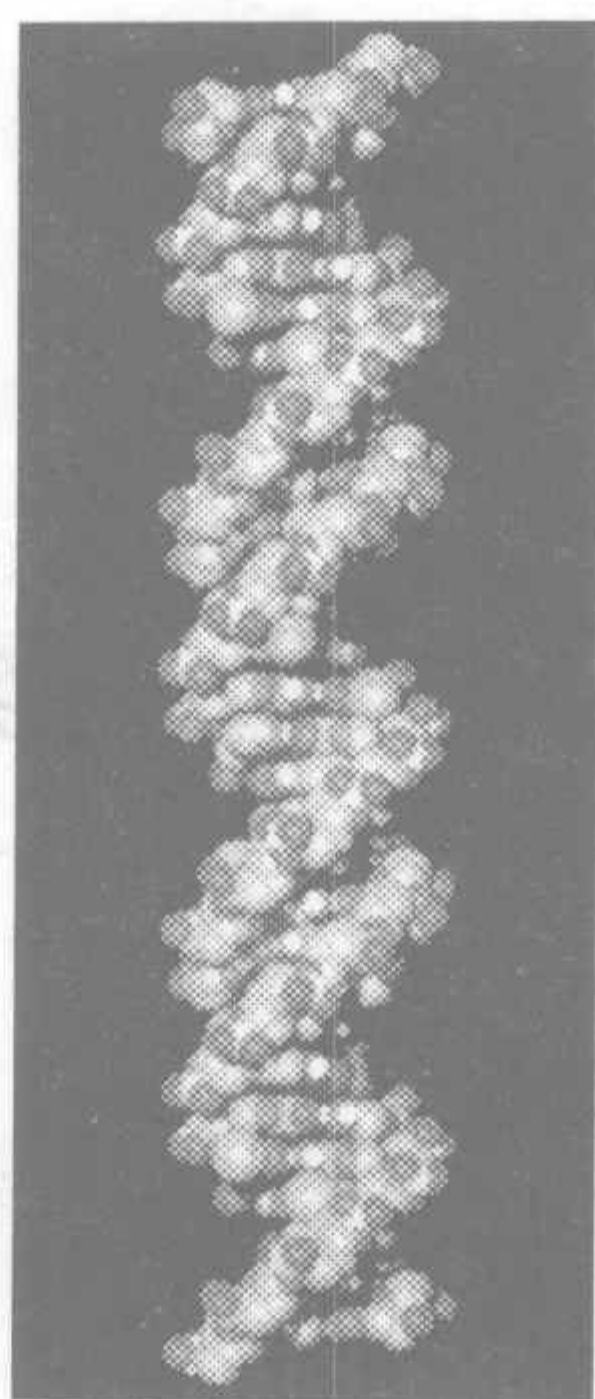


1.2 鉴别脱氧核糖、磷酸基团和不同的碱基。你是否能够确认脱氧核糖上的 1'-5' 的碳原子？

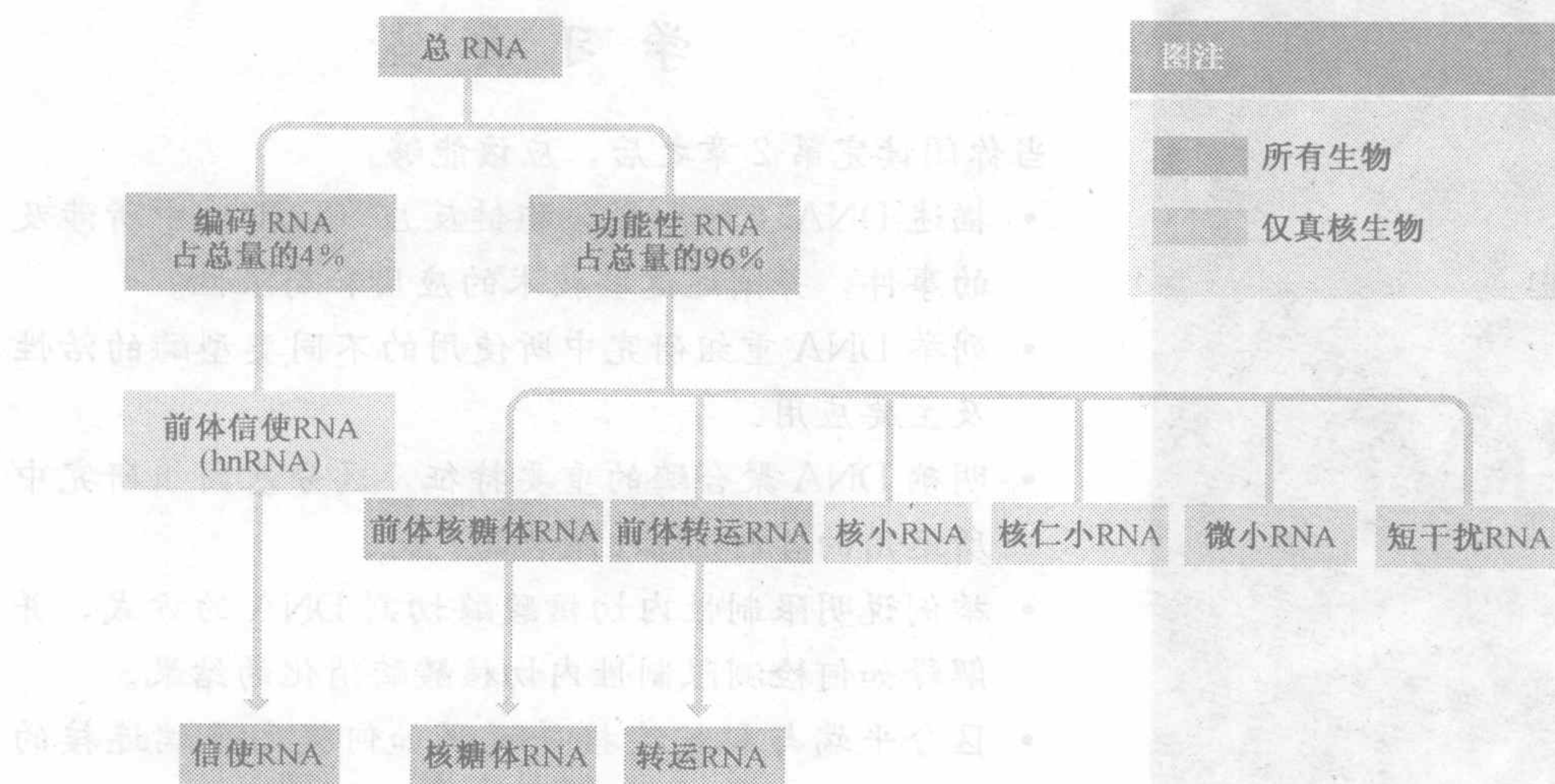


1.3\* 就下面的 B 型 DNA 的空间填充模型，请描述该分子的重要结构特点。





#### 1.4 解释原核与真核细胞中 RNA 的差异。



### 拓展阅读

关于双螺旋发现的书籍和文献以及 DNA 研究方面其他的重要里程碑

**Brock, T.D.** (1990) *The Emergence of Bacterial Genetics*. Cold Spring Harbor Laboratory Press, New York. A detailed history that puts into context the work on the transforming principle and the Hershey-Chase experiment.

**Judson, H.F.** (1979) *The Eighth Day of Creation*. Jonathan Cape, London. A highly readable account of the development of molecular biology up to the 1970s.

**Kay, L.E.** (1993) *The Molecular Vision of Life*. Oxford University Press, Oxford. Contains a particularly informative explanation of why genes were once thought to be made of protein.

**Lander, E.S. and Weinberg, R.A.** (2000) Genomics: journey to the center of biology. *Science* **287**: 1777-1782. A brief descrip-

tion of genetics and molecular biology from Mendel to the human genome sequence.

**Maddox, B.** (2002) *Rosalind Franklin: The Dark Lady of DNA*. HarperCollins, London.

**McCarty, M.** (1985) *The Transforming Principle: Discovering that Genes are Made of DNA*. Norton, London.

**Olby, R.** (1974) *The Path to the Double Helix*. Macmillan, London A scholarly account of the research that led to the discovery of the double helix.

**Watson, J.D.** (1968) *The Double Helix*. Atheneum, London. The most important discovery of twentieth century biology, written as a soap opera.



## 2.1 用于 DNA 操作的酶

## 2.2 DNA 克隆

## 2.3 聚合酶链反应 (PCR)

## 学 习 要 点

当你阅读完第 2 章之后, 应该能够:

- 描述 DNA 克隆和聚合酶链反应 (PCR) 中所涉及的事件, 并阐述这些技术的应用和局限性。
- 列举 DNA 重组研究中所使用的不同类型酶的活性及主要应用。
- 明确 DNA 聚合酶的重要特征, 区分基因组研究中所运用的各种 DNA 聚合酶。
- 举例说明限制性内切核酸酶切割 DNA 的方式, 并解释如何检测限制性内切核酸酶消化的结果。
- 区分平端与黏端连接并说明如何提高平端连接的效率。
- 详述质粒克隆载体的主要特征并描述如何将这些载体用于克隆实验中。
- 描述如何将  $\lambda$  噬菌体载体用于 DNA 克隆。
- 举出用于克隆长片段 DNA 的载体, 并评价每种载体的优点及局限性。
- 概括酵母、动物和植物中如何克隆 DNA。
- 描述如何进行 PCR 反应, 要特别注意引物和热循环中所用温度的重要性。



实际上，我们知道的有关基因组及基因组表达的所有事情都是通过科学研究发现的，理论研究在该领域或分子与细胞生物学的其他任何领域中作用甚微。虽然对如何得到那些事实了解不多也可以学习有关基因组的“事实”，但是为了真正了解这一学科，我们必须详细审视基因组研究中所运用的技术和科学方法。后续五章覆盖了这些研究方法。首先，我们探讨了用于研究 DNA 分子的以 DNA 克隆和聚合酶链反应为核心的技术。这些技术对小片段 DNA 包括单个基因很有效，从而使我们在该水平上获得丰富信息。第 3 章继续探讨为构建基因组图谱而发明的方法，并描述了大约一个世纪前首次发明的遗传作图技术是如何用互补的各种基因组物理作图法进行补充的。第 4 章在作图和测序之间建立联系，并表明虽然图谱对装配一个长的 DNA 序列可以提供有用帮助，但作图并不总是基因组测序的必要先决条件。在第 5 章中，我们着眼于用来理解基因组序列的不同方法，而在第 6 章中，我们检查了用于研究基因组表达的方法。当研读第 6 章时，你将开始体会到明白基因组如何指定活细胞的生物化学能力是当前生物学研究的主要挑战之一。

20 世纪 70 年代到 80 年代之间，分子生物学家用来研究 DNA 分子的技术被汇编成一套工具。在那之前，研究单个基因的唯一方法是经典遗传学，所用方法将在第 3 章描述。20 世纪 70 年代早期，生物化学研究为分子生物学家提供了可以在试管中操作 DNA 分子的酶，这一突破推动了更直接研究 DNA 方法的发展。这些酶在活细胞中天然存在，并在 DNA 复制、修复和重组过程中起作用，这些过程将在第 15、16 和 17 章中介绍。为了确定这些酶的功能，许多酶被纯化，并对它们催化的反应进行了研究。然后分子生物学家利用纯化的酶作为工具按预定的方式操作 DNA 分子，用它们进行 DNA 分子拷贝，将 DNA 分子切成小片段并重新将它们连接在一起形成自然界不存在的组合体（图 2.1）。这些操作形成了 **DNA 重组技术**（recombinant DNA technology）的基础，运用 DNA 重组技术可以将天然存在的染色体片段和质粒构建成为新的或“重组”的 DNA 分子。

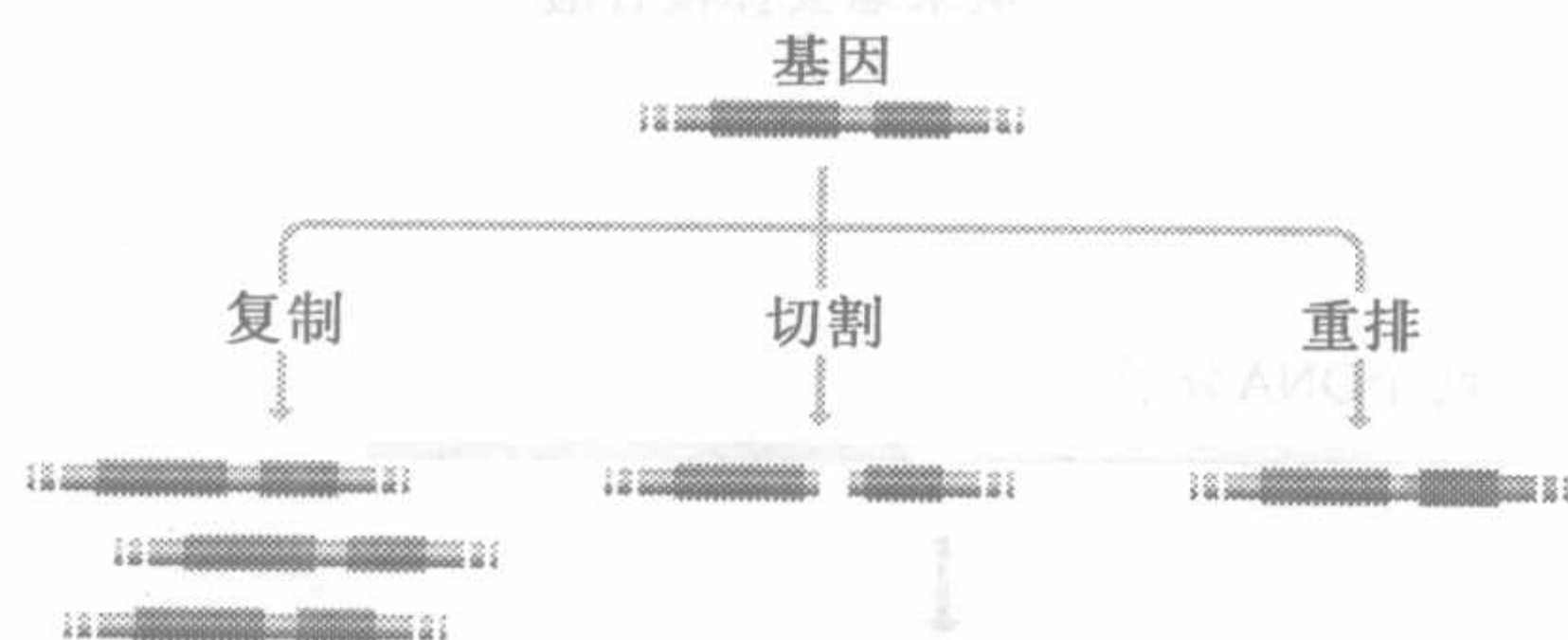


图 2.1 用 DNA 分子进行操作例子

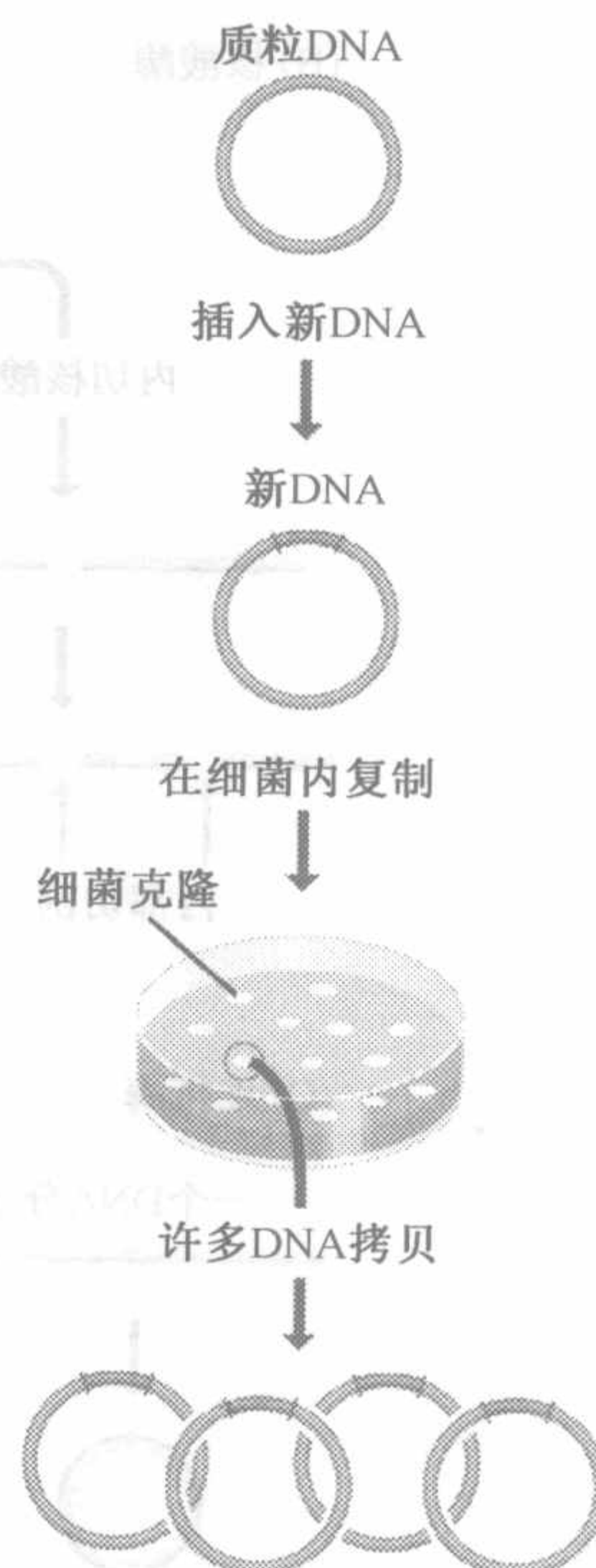


图 2.2 DNA 克隆

DNA 重组方法学导致 **DNA 克隆**（DNA cloning）或 **基因克隆**（gene cloning）的产生，通过此方法将小 DNA 片段（可能包含单个基因）插入到质粒或病毒染色体中并由此在细菌或真核类宿主中复制（图 2.2）。在此例中，要克隆的 DNA 片段被插入到质粒载体中，随后质粒就在细菌宿主内复制。



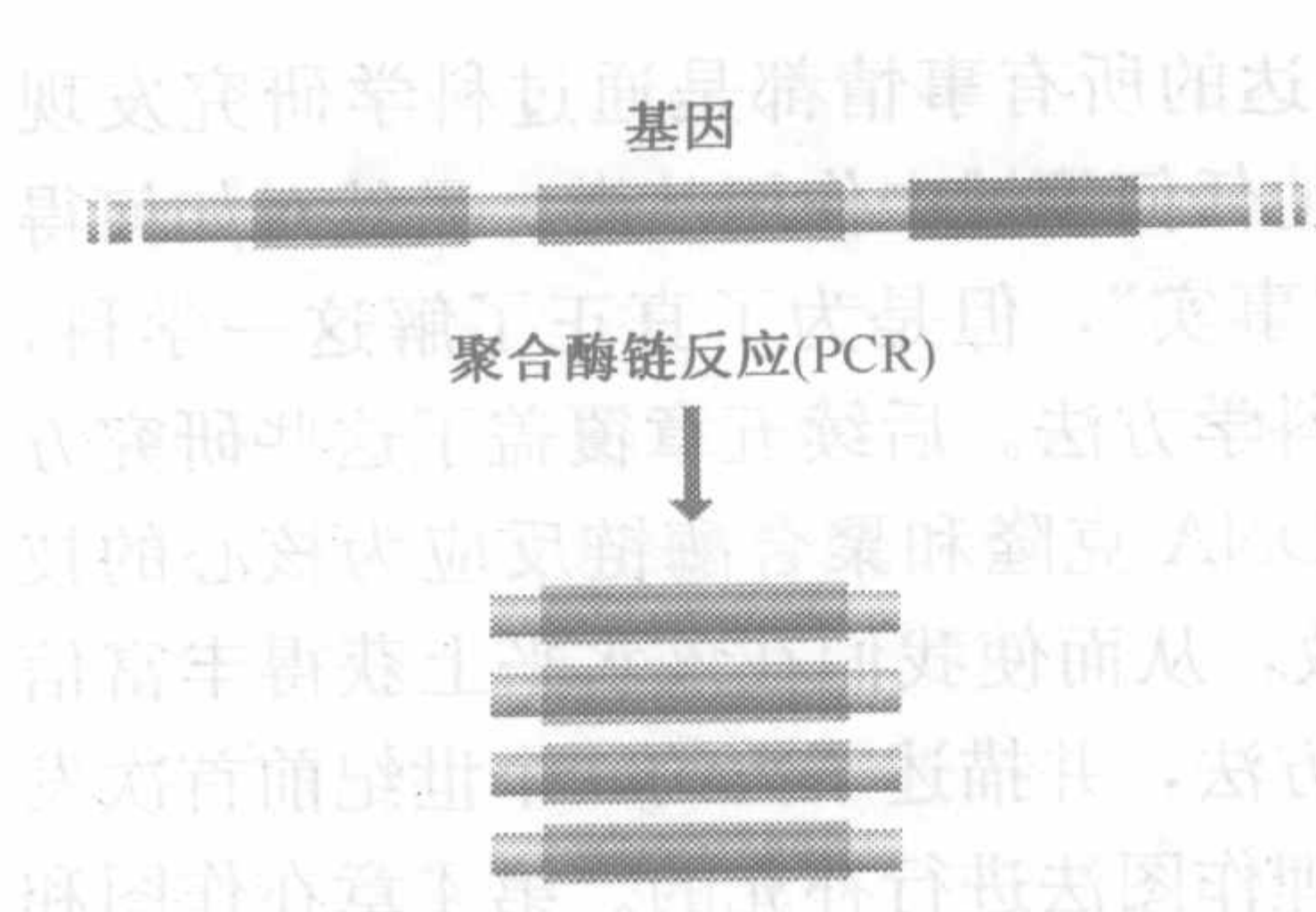


图 2.3 聚合酶链反应 (PCR) 用来复制 DNA 分子中所选择的片段  
在此例中, 单个基因被复制

2.2)。我们将在 2.2 节中准确描述如何进行基因克隆以及该技术引发分子生物学变革的原因。

到 20 世纪 70 年代末, 基因克隆很好地建立起来。下一个主要技术突破是 20 世纪 80 年代中期聚合酶链反应 (polymerase chain reaction, PCR) 的发明。PCR 不是一项复杂技术, 它所做的一切是 DNA 分子小片段的重复拷贝 (图 2.3), 但它在生物学的许多领域而不只是基因组研究中变得非常重要。PCR 在 2.3 节中详细介绍。

#### (A) DNA 聚合酶

DNA 模板

RNA 模板

DNA 拷贝

#### (B) 核酸酶

内切核酸酶

外切核酸酶

内部切割

从末端去掉核苷酸

#### (C) 连接酶

一个 DNA 分子

两个 DNA 分子

图 2.4 DNA 聚合酶 (A)、核酸酶 (B) 及连接酶 (C) 的活性  
(A) 中, DNA 依赖的 DNA 聚合酶活性如左边所示, RNA 依赖的 DNA 聚合酶如右边所示。(B) 中, 显示了内切核酸酶和外切核酸酶的活性。(C) 中, 浅色的 DNA 分子与其自身相连 (左边), 与另一个 DNA 分子相连 (右边)



## 2.1 用于 DNA 操作的酶

DNA 重组技术是 20 世纪 70~80 年代出现的促进基因表达相关领域快速发展的重要因素之一。DNA 重组技术的基础是能够在试管中操作 DNA 分子。这反过来依赖于能够提供纯化的酶，而且酶的活性已知并能被控制，这些酶才能被用于对被操作的 DNA 分子进行特异性改变。分子生物学家可用的酶分为四大类。

- **DNA 聚合酶** (DNA polymerase) (2.1.1 节)，是一类以现存 DNA 或 RNA 为模板合成新的互补多聚核苷酸的酶 [图 2.4 (A)]。
- **核酸酶** (nuclease) (2.1.2 节)，通过切断连接两个核苷酸之间的磷酸二酯键降解 DNA 分子 [图 2.4 (B)]。
- **连接酶** (ligase) (2.1.3 节)，在两个不同分子的末端核苷酸或单个分子的两末端核苷酸之间形成磷酸二酯键而将 DNA 分子连接在一起 [图 2.4 (C)]。
- **末端修饰酶** (end-modification enzyme) (2.1.4 节)，改变 DNA 分子末端，为连接实验的设计增加重要的可操作空间，并提供了一种用放射性及其他标记物标记 DNA 分子的方法 (技术注解 2.1)。

### 技术注解 2.1 DNA 标记

将放射性、荧光类或其他类型标记物连接到 DNA 分子上

DNA 标记是许多分子生物学实验的中心环节，包括 Southern 杂交 (2.1.2 节)、荧光原位杂交 (FISH; 3.3.2 节) 和 DNA 测序 (4.1 节)。它通过检测标记物发出的信号，能够在硝酸纤维素膜、尼龙膜、染色体或者凝胶上确定某个特定 DNA 分子的定位。在某些实验中也用到标记的 RNA 分子 (技术注解 5.1)。

放射性标记物经常用来标记 DNA 分子。在合成核苷酸时，可用<sup>32</sup>P 或<sup>33</sup>P 取代其中一个磷原子，用<sup>35</sup>S 取代磷酸基团中的一个硫原子或者用<sup>3</sup>H 取代一个或多个氢原子 (图 1.4)。放射性标记的核苷酸仍然作为 DNA 聚合酶的底物，因此能通过 DNA 聚合酶催化的任一链的合成反应而掺入到 DNA 分子中。标记的核苷酸或单个磷酸基团也可以通过 T4 多聚核苷酸激酶或末端脱氧核糖核苷酸转移酶的催化作用而连到 DNA 分子的一端或两端 (2.1.4 节)。放射性信号可以通过闪烁计数来检测，但在大多数分子生物学应用中需要定位信息，因此通过 X-光敏感的胶片 [放射自显影术 (autoradiography); 图 2.11 示例] 或放射性敏感的磷屏成像 (phosphorimaging) 进行检测。对不同放射性标记物的选择依赖于所用方法的要求。<sup>32</sup>P 这种同位素具有高的发射能量因此其敏感性高，但由于信号的发散其高敏感性同时伴随着低分辨率。像<sup>35</sup>S 或<sup>3</sup>H 之类的低发射能量同位素，敏感性低但分辨率高。

健康与环境问题表明近年来放射性标记物已不再受欢迎，在许多实验中它们正被非放射性替代物大量取代。其中最有用的是荧光标记物，它是诸如 FISH 技术 (3.3.2 节) 和 DNA 测序技术 (4.1.1 节) 的重要成分。不同发射波长的荧光标记物 (如不同颜色) 被掺入到核苷酸中或者直接与 DNA 分子连接，可以用合适的胶片检测及通过荧光显微镜检测或用荧光检测仪检测。其他类型的非放射性标记利用化学发光物质，但缺点是信号不是由标记物直接产生，而标记分子必须用化学试剂处理后才能发光。一种流行的方法是用碱性磷酸酶标记 DNA 分子，利用在酶的去磷酸化作用下产生化学发光的二氧环烷来检测。



接下来的两章讲述的是获得基因组序列的技术和策略。DNA 测序在这些技术中是至关重要的，但是测序有一个极大的局限性：即使是最精确的技术，在一个反应中也很难测出大于 750bp 的序列。这就意味着长的 DNA 分子不得不由一系列短的序列拼接而成，即需将大分子分解为片段，测出每一段的序列，再用计算机寻找重叠的部分，从而拼接成长的序列（图 3.1）。这种**鸟枪法**（shotgun method）是小的原核生物基因组测序的标准方法（4.2.1 节），但是对于较大的基因组来说这是相当困难的，

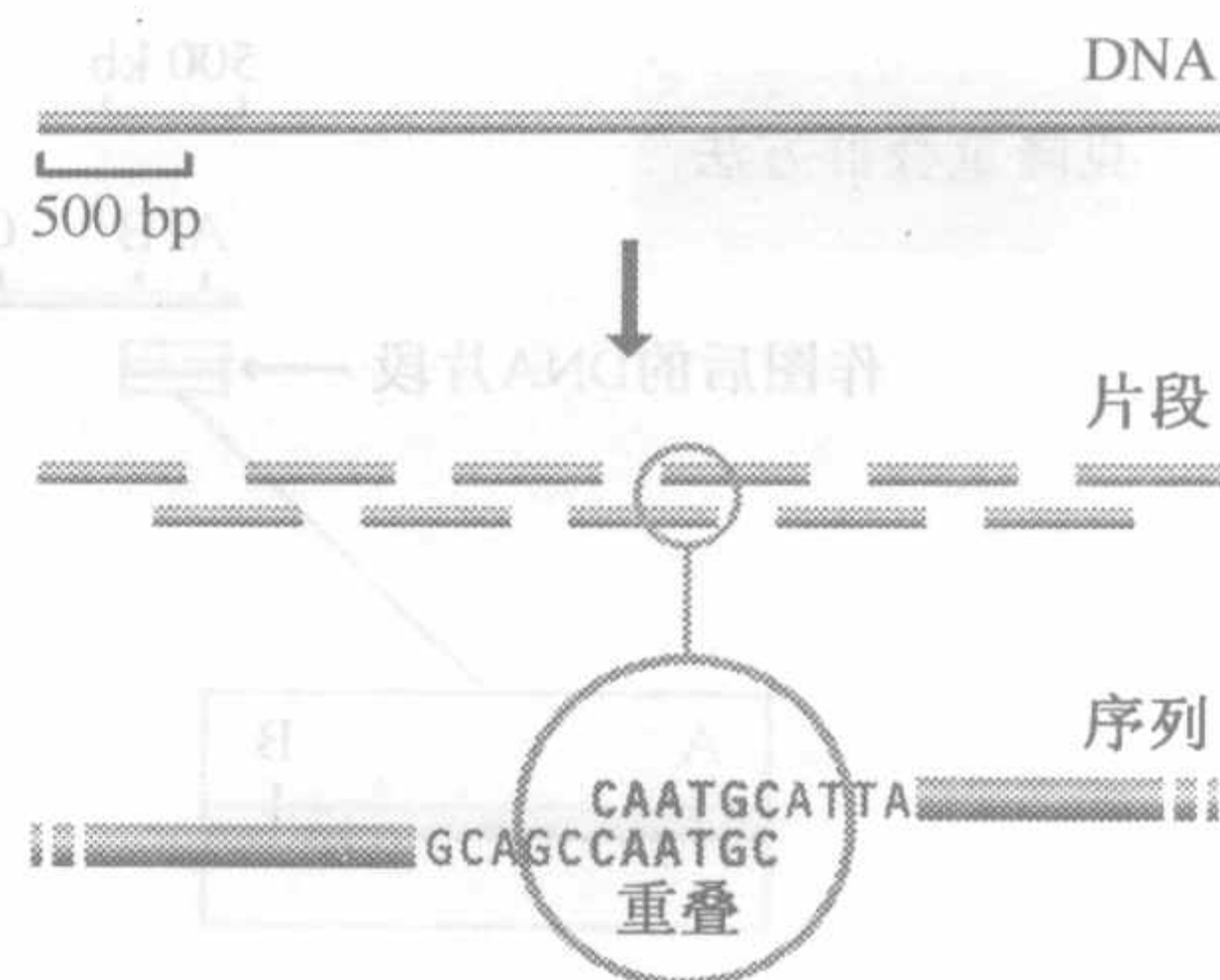
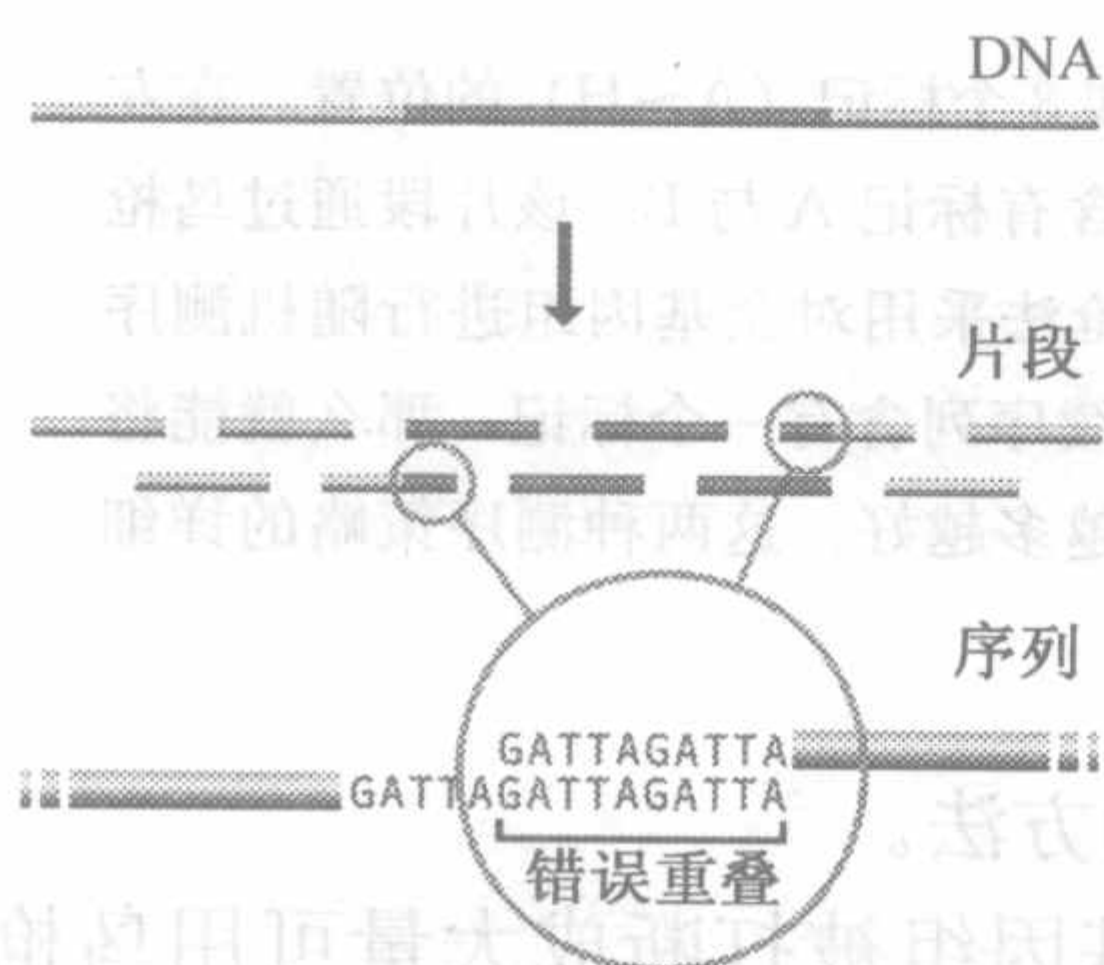


图 3.1 鸟枪法组装序列

将 DNA 分子打断为小片段，测定每一段的序列，通过查找每个片段序列的重叠部分再组装得到主体序列。实际上，需要几十个碱基对的重叠以确认两个序列相衔接

因为随着片段数的增加，所需要的数据分析会越来越复杂（ $n$  个片段可能的重叠数为  $2n^2 - 2n$ ）。鸟枪法的第二个问题是当分析基因组的重复区域时会发生错误。当一段重复序列被分解成片段后，许多片段将含有相同的或非常相似的序列基序。因此这些序列很容易被重组，导致部分重复区域被遗漏，或者将同一染色体或不同染色体的两个片段错误地连接在一起（图 3.2）。

(A) 串连重复DNA的问题



(B) 基因组范围内重复的问题

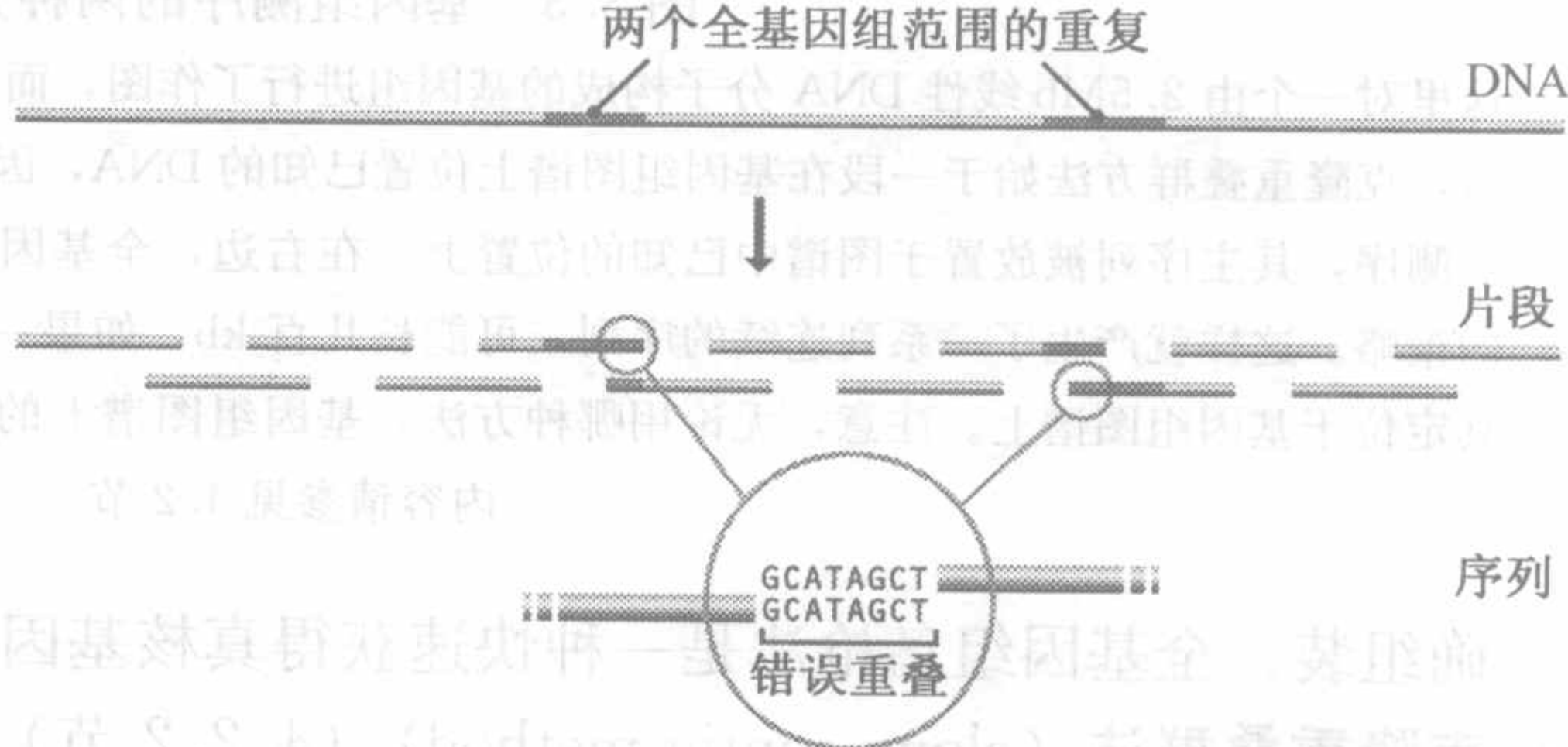


图 3.2 鸟枪法中所遇到的问题

(A) 该 DNA 分子含有一个由许多拷贝的 GATTAGATTA 序列组成的串联重复元件。分析序列时，来自于 DNA 不同部位的两个片段看起来是重叠的。如果未能发现这个错误，那么主体序列将遗漏这个 DNA 分子的中间片段。

(B) 在第二个例子中，该 DNA 分子含有基因组范围内的一个重复元件的两个拷贝。分析这些序列时，两个片段看上去是重叠的，不过一个片段含有重复元件的左半部分，而另一个片段则含有第二个重复元件的右半部分。在这种情况下，如果不能识别出这种错误，将会导致这两个重复元件之间 DNA 片段大量序列的丢失。如果这两个重复元件在不同的染色体上，则会将这些染色体的序列错误地组装在一起

由于在测定含重复序列的 DNA 大分子方面存在困难，因此鸟枪法本身不适合用于进行真核基因组的测序。相反，必须首先建立一个基因组的图谱（map），通过标明基因和其他显著特征的位置，为测序提供指导。一旦得到了基因组的图谱，基因组计划中的测序阶段可以采用下面两种方法之一进行测序（图 3.3）。

• **全基因组鸟枪法**（whole-genome shotgun method）（4.2.3 节） 采用与标准的鸟枪法相同的方法，只是使用基因组图谱上的显著特征作为界标，指引着将用鸟枪法获得的大量短序列拼接成主序列。参考基因组图谱也可以确保含有重复 DNA 的区段正



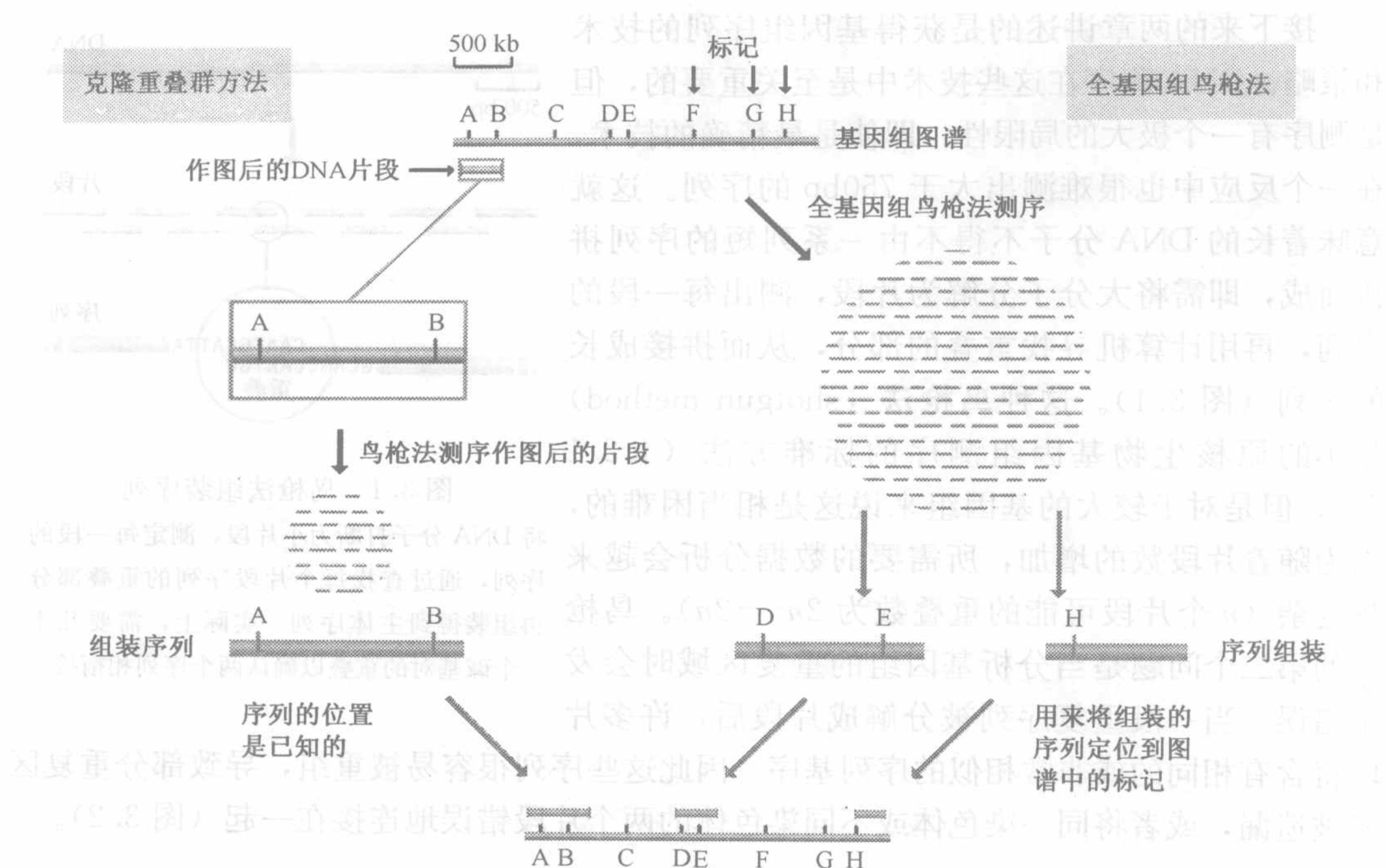


图 3.3 基因组测序的两种方法

这里对一个由 2.5Mb 线性 DNA 分子构成的基因组进行了作图，而且已知 8 个标记（A~H）的位置。在左边，克隆重叠群方法始于一在基因组图谱上位置已知的 DNA，因为它含有标记 A 与 B。该片段通过鸟枪法测序，其主序列被放置于图谱中已知的位置上。在右边，全基因组鸟枪法采用对全基因组进行随机测序的策略。这样就产生了一系列连续的序列，可能长几百 kb，如果一个连续序列含有一个标记，那么就能将其定位于基因组图谱上。注意，无论用哪种方法，基因组图谱上的标记越多越好。这两种测序策略的详细内容请参见 4.2 节

确组装。全基因组鸟枪法是一种快速获得真核基因组的方法。

- **克隆重叠群法** (clone contig method) (4.2.2 节) 基因组被打断成大量可用鸟枪法准确测序的片段，每个片段长数百 kb 或数个 Mb，一旦一个片段的序列测定完成，就被定位在基因组图谱的正确位置上。这种逐步测序的方法比全基因组鸟枪测序法要花更多的时间，但是它可以得到更准确无误的序列。

应用这两种方法，基因组图谱为实施人类基因组计划的测序工作提供了框架。如果基因组图谱上标明了基因的位置，那么就可以把克隆群的起始部分放在基因组内令人感兴趣的区域，从而可以尽快获得重要基因的序列。

### 3.1 遗传图谱和物理图谱

传统上将基因组作图方法分为两类。

- **遗传作图** (genetic mapping) 应用遗传学技术构建的能在基因组上显示基因和其他序列特征位置的图谱。遗传学技术包括杂交育种实验，对人类则是检查家族史（家谱 pedigree）。遗传学作图在 3.2 节介绍。
- **物理作图** (physical mapping) 应用分子生物学技术直接检测 DNA 分子，从而构



- d. 单核苷酸多态性被用于确认那些能导致明显可见表型差异的点突变。
- 3.3\* 下述哪一项不是生化表型通常被用于构建人类遗传图谱的理由:
- 人类没有用于遗传作图的可见特征。
  - 生化表型可以通过简单的血型分析来筛选。
  - 一些容易描述的生化表型由具有非常多等位基因的基因决定。
  - 进行控制的人类繁殖实验是不道德的。
- 3.4 真核基因组除了基因以外还使用 DNA 标记来作图是因为:
- DNA 标记不需要两个或更多的等位基因来作图。
  - 基因图谱可能不能覆盖基因组的大型区域。
  - 绝大多数基因包含多个能被轻易定位的等位基因。
  - DNA 标记相对于遗传标记变异更小。
- 3.5\* 微卫星比小卫星更常用于 DNA 标记的原因是:
- 基因组中小卫星所在的位点太多。
  - 限制性酶能够被用于微卫星但不能应用于小卫星。
  - 真核基因组中很少有微卫星, 所以它们很容易被确认和分析。
  - 微卫星遍布真核基因组并且容易用 PCR 扩增。
- 3.6 下面哪一个遗传学标记在人类基因组中的数目最多?
- RFLP
  - 小卫星
  - 微卫星
  - 单核苷酸多态性
- 3.7\* 遗传连锁的原则是:
- 某一个基因的不同等位基因定位于一个染色体的相同位置。
  - 发现某些特征(如果蝇眼睛的颜色)由多个基因负责。
  - 观察到某些基因如果定位于同一个染色体将会被一起遗传。
  - 观察到染色体的深染区不包含基因。
- 3.8 有丝分裂和减数分裂的差异是有丝分裂有以下特征:
- 产生的两个双倍体细胞与亲代细胞在遗传学方面是等同的。
  - 同源染色体之间 DNA 的交换(crossing-over)。
  - 产生的两个双倍体细胞与亲代细胞在遗传学方面是不同的。
  - 产生的四个双倍体细胞与亲代细胞在遗传学方面是等同的。
- 3.9\* 下面哪一项准确地描述了两个基因之间的重组频率?
- 两个基因靠得越近, 它们之间的重组频率就越高。
  - 两个基因离得越远, 它们之间的重组频率就越高。
  - 如果两个基因在同一个染色体上, 它们之间不可能发生重组。
  - 如果两个基因在不同的染色体上, 它们之间的重组频率就比较高。
- 3.10 在分析一个人类家系以确定两个基因的连锁性如何的过程中, 最好是:
- 总结出在子代中最常见的基因型是其亲代的基因型。
  - 总结出在子代中最常见的基因型是重组子。



- c. 进行测交试验来确定基因之间的连锁性。
- d. 确定祖父母的基因型。
- 3.11\* 下述哪一项不是限制人类及其他复杂真核生物遗传图谱准确性的因素?
- a. 对于许多真核生物而言不可能得到足够的后代。
- b. 重组热点可能干扰遗传作图。
- c. 遗传作图只使用基因而复杂真核生物没有足够的基因来为整个基因组作图。
- d. 距离上万个碱基对的基因或标记可能出现在遗传图谱的相同位置。
- 3.12 中期染色体最初被用于荧光原位杂交,但是结果多少有其局限性是因为:
- a. 染色体的许多区域是浓缩的且不能被探针杂交。
- b. 探针会优先与多个染色体上的重复序列杂交。
- c. 浓缩状态的染色体不稳定,当染色体变得舒松的时候杂交信号就变弥散了。
- d. 只可能作出低分辨率的图谱,因为染色体是浓缩的。
- 3.13\* 分裂间期染色体对于通过荧光原位杂交来精细作图是有用的,因为它们:
- a. 是浓缩最少的染色体类型。
- b. 彼此间很容易通过结构来区分。
- c. 具有该技术所需要的染色质转录激活区域。
- d. 允许基因组的物理图谱分辨率达到 1kb。
- 3.14 序列标签位点具有以下哪个特点?
- a. 它们在基因组中只出现一次并且具有一个 RFLP 位点。
- b. 它们在基因组中只出现一次并且已知序列。
- c. 它们的序列已知并且必定含有重复的 DNA 序列。
- d. 它们必定还有一个基因的序列,而没有重复 DNA 序列的存在。
- 3.15\* 下述哪一个序列不能被用作一个序列标记位点?
- a. 表达序列标签。
- b. 随机基因组序列。
- c. 简单序列长度多态性。
- d. 限制性片段长度多态性。
- 3.16 放射杂交群为物理作图提供了一个有用的机制,因为:
- a. 在任一给定的杂交细胞中只存在一部分的人类基因组。
- b. 宿主仓鼠细胞缺乏人类遗传标记的同源序列。
- c. 宿主仓鼠细胞对射线不敏感。
- d. 仓鼠基因组的完整物理图谱是已知的。

## 简答题

\* 奇数问题的答案见附录

- 3.1\* 为什么基因组测序需要图谱?如果没有一个基因组图谱,在获得基因组序列过程中的主要困难是什么?
- 3.2 清晰地解释基因组的遗传图谱与物理图谱之间的差别。
- 3.3\* PCR 技术是如何使得 RFLP 的分析既快又方便的?在使用 PCR 来作 RFLP 图谱之前需要什么?



ORF 扫描结果时出错的机会就会降低。

## 单纯的 ORF 扫描对高等真核生物 DNA 效果不佳

虽然 ORF 扫描对细菌基因组很有效，但在高等真核生物的 DNA 序列中定位基因的效果不佳。部分原因是真核生物基因组中真正基因之间的间隔很大（如人类基因组中接近 62% 的序列是基因间序列），发现假 ORF 的概率就会增加。但对人类基因组和高等真核生物基因组所普遍存在的主要问题是它们的基因经常被内含子隔开（1.2.3 节），因此在 DNA 序列中就不会出现连续的 ORF。许多外显子小于 100 个密码子，一些外显子包含的密码子小于 50 个，将可读框延伸到内含子中通常会引起一个看起来接近 ORF 的终止序列（图 5.4）。换句话说，高等真核生物基因不会以长 ORF 的形式出现在基因组序列中，单纯的 ORF 扫描不能定位它们。

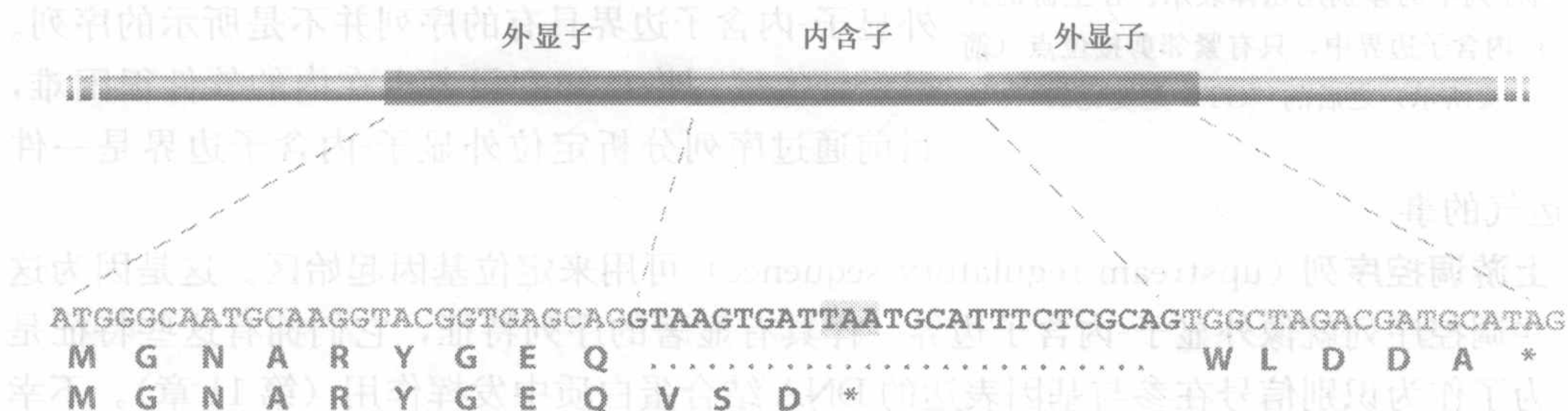


图 5.4 内含子使 ORF 扫描复杂化

图中显示出包含一个内含子的短基因的核苷酸序列。从基因翻译出的蛋白质中正确的氨基酸序列在紧邻核苷酸序列下面列出：在该序列中，内含子被删掉，因为它在 mRNA 翻译成蛋白质之前就从转录产物中被去除。在下面一行中，没有认识到内含子的存在就把序列翻译出来。这种错误的结果就使氨基酸序列在内含子内部出现了终止。氨基酸序列是用一个字母的缩写书写的（表 1.2）。星号表示终止密码子的位置。内含子会在 12.2.2 节中详细介绍

解决内含子产生的问题是生物信息学家为 ORF 定位而编写新软件程序所面临的主要挑战。对 ORF 扫描的基本程序已经进行了三项改良。

- **密码子偏倚**（codon bias）被考虑在内。“密码子偏倚”是指特定生物体的基因中并不是所有密码子的使用频率都是平等的。例如，在遗传密码中亮氨酸是由 6 个密码子编码的（TTA、TTG、CTT、CTC、CTA 和 CTG；图 1.20），但在人类基因中，亮氨酸大多是由 CTG 编码的，而且几乎不是由 TTA 或 CTA 编码。同样，在 4 个缬氨酸密码子中，人类基因利用 GTG 的频率比利用 GTA 的频率高 4 倍。密码子偏倚的生物学原因并不清楚，但所有生物体都有偏倚，不同种属中的偏倚也不同。预期真正的外显子会表现出密码子偏倚，而三联核苷酸的随机排列却不会。因此，被研究的生物体中密码子偏倚被写入 ORF 扫描软件中。

- **外显子-内含子边界**（exon-intron boundary）可因其具有特定的序列特征而被搜索到，尽管不幸的是，这些序列的特异性不会重要到使它们的定位变得简单。上游外显子-内含子边界的序列通常为：

5'-AG↓GTAAGT-3'

箭头表示准确的分界点。然而，只有紧邻箭头之后的“GT”是不变的：在序列的



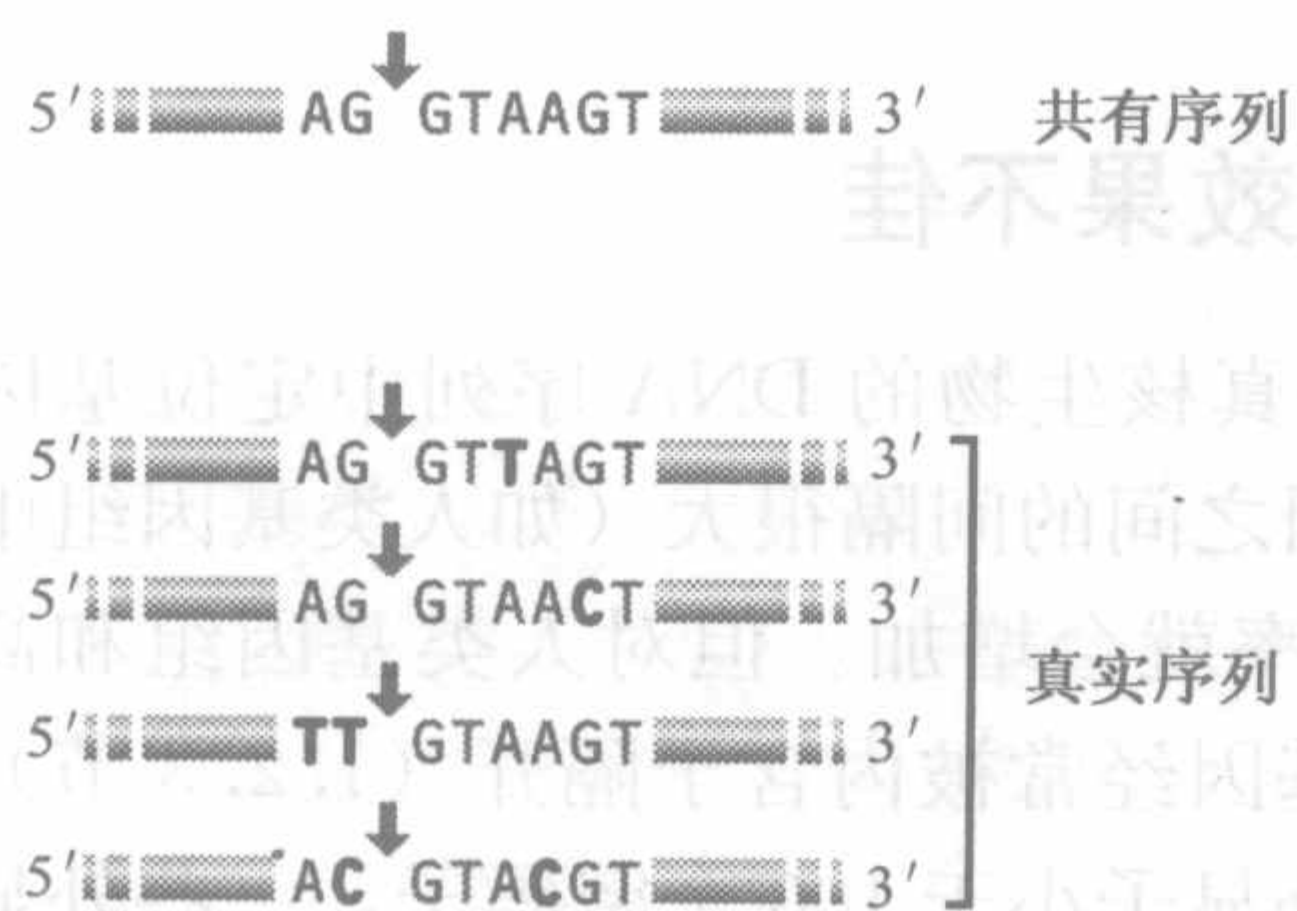


图 5.5 上游外显子-内含子边界的共有序列和在真正基因中发现的真实序列之间的关系

共有序列中的差别用粗体表示。在上游的外显子-内含子边界中，只有紧邻剪接位点（箭头所示）之后的“GT”是变化的

其他位置经常会发现图中没有出现的核苷酸。换句话说，该序列是共有的（consensus），我们的意思是该序列表示在所有已知上游外显子-内含子边界的每个位置上出现频率最高的核苷酸，但在任何特定的边界序列中一个或多个位置上可能具有不同的核苷酸（图 5.5）。下游内含子-外显子边界更不明显：



其中“Py”是指嘧啶核苷酸（T 或 C）的一种，“N”指任意核苷酸。只寻找这些共有序列将不会定位更多的外显子-内含子边界，因为大多数外显子-内含子边界具有的序列并不是所示的序列。已经证实编写将已知变量考虑在内的软件很困难，目前通过序列分析定位外显子-内含子边界是一件碰运气的事。

- **上游调控序列**（upstream regulatory sequence）可用来定位基因起始区。这是因为这些调控序列就像外显子-内含子边界一样具有显著的序列特征，它们拥有这些特征是为了作为识别信号在参与基因表达的 DNA 结合蛋白质中发挥作用（第 11 章）。不幸的是，对于外显子-内含子边界来说，调控序列是变化的，在真核生物中的变化比原核生物更大，并且真核生物中并非所有基因都拥有同样的调控序列。因此，运用这些来定位基因是有问题的。

这三种简单 ORF 扫描的衍生方法虽然有局限性，但普遍适用于所有高等真核生物的基因组。根据它们基因组的特征，另外的策略也可能适用于单个生物体。例如，脊椎动物基因组包含着许多基因上游都有的 **CpG 岛**（CpG island），这些序列大约有 1kb 长，其中的 GC 含量比整个基因组的平均含量要高。人类基因中有 40%~50% 的基因上游含有 CpG 岛。这些序列是有特色的，当一个这样序列位于脊椎动物 DNA 中时，就可以大胆推测在紧邻该序列的下游区存在一个基因。

## 为功能性 RNA 定位基因

ORF 扫描适用于蛋白质编码基因，但对于那些功能性 RNA 基因，如 rRNA 和 tRNA（1.2.2 节）又是如何定位呢？这些基因不包含可读框，因此不能通过上面描述的方法进行定位。然而，功能性 RNA 分子确实具有它们自己的特征，这些特征可以用来帮助在基因组序列中发现它们。这些特征中最重要的是能够折叠成二级结构，如 tRNA 分子所具有的三叶草（cloverleaf）结构 [图 5.6 (A)]。这些二级结构并不是像 DNA 双螺旋那样通过两条单独多聚核苷酸之间的碱基配对结合在一起，而是通过相同多聚核苷酸的不同部位间的碱基配对结合在一起，我们称之为**分子内碱基配对**（intramolecular base pairing）。为了使分子内形成碱基配对，该分子中两部分的核苷酸序列必须是互补的，为了形成复杂的结构，如三叶草结构，所有配对的互补序列在 RNA 序列内必须按照特定的顺序进行排列 [图 5.6 (B)]。这些特征提供了很丰富的信



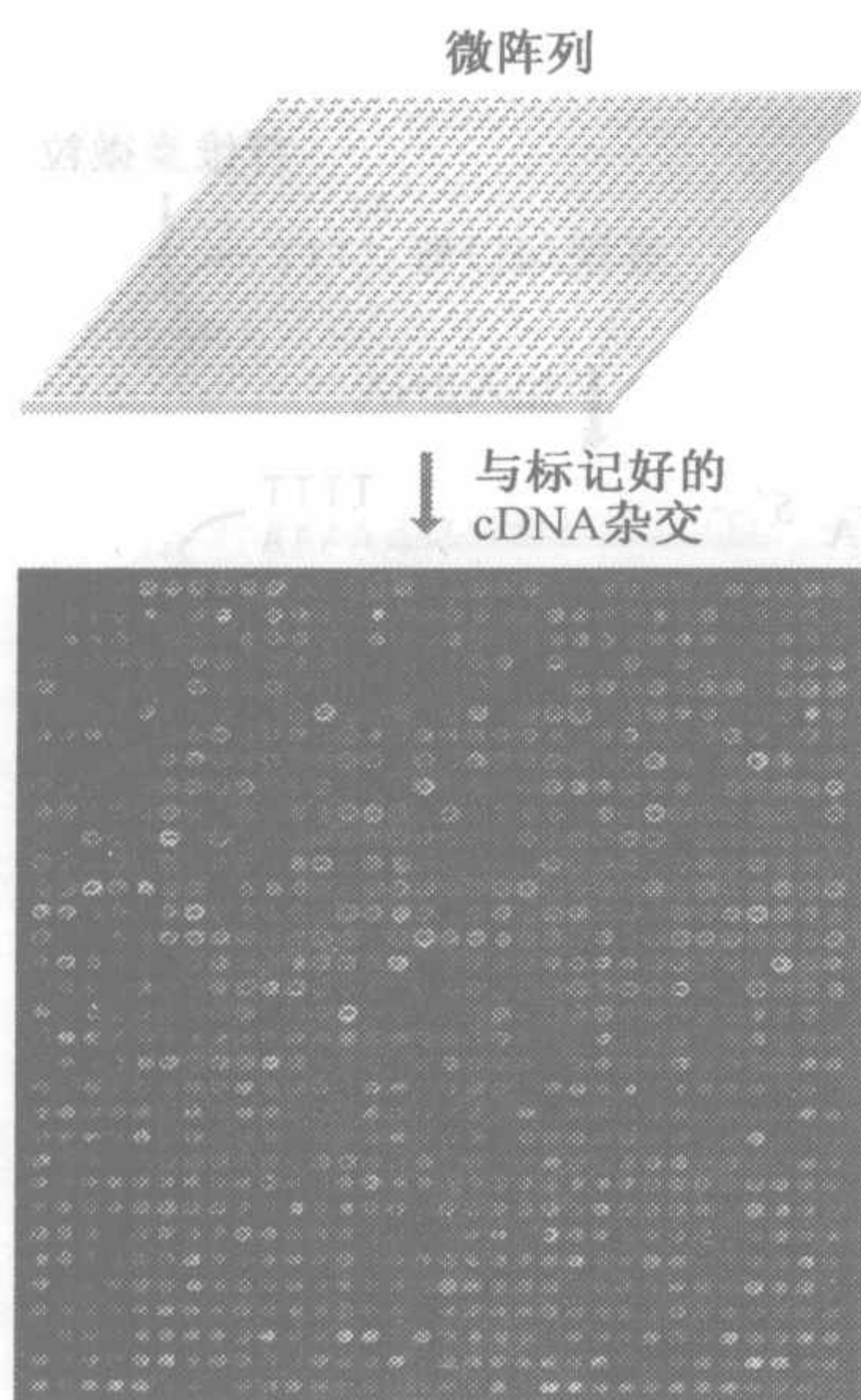


图 6.2 微阵列分析

一个 cDNA 样品经过荧光标签标记后杂交到微阵列上。标签通过激光共聚焦扫描检测后信号强度被转换成一个拟色光谱（红色代表最强的杂交，接下来依次是橙色、黄色、绿色、蓝色、紫蓝色和紫色，最后一个代表杂交的背景强度）。关于芯片制备和使用的更多信息参见技术注解 3.1。图由 Tom Strachan 惠赠，并经 *Nature* 杂志允许采用

有一个探针分子能够代表每一个相应的基因。这一点对于微阵列而言，是通过使用来自于感兴趣基因的 PCR 产物或 cDNA；对于 DNA 芯片而言，就是通过每一个杂交点合成一个可能包含多达 20 种不同的寡聚核苷酸链的混合物（它们的序列分别与相应基因的不同部位匹配）（图 6.3）。第二项要求（即确定不同 mRNA 在转录组中的相应数量）是可以满足的，因为微阵列或芯片中的每一个点都包含多达  $10^9$  拷贝数的探针分子，这比使用阵列检测的一个小量转录组中任何一个 mRNA 预期的拷贝数都高，也就是说任意一个杂交点都不会被饱和（饱和即杂交反应多到使每一个探针分子都被靶基因分子碱基互补配对结合了）。杂交反应的数量是不同的，每一个杂交点的信号强度依赖于转录组中每一个特定 mRNA 的数量（图 6.4）。

通过上面的阅读，微阵列和芯片分析看来似乎是简便的过程。但在实践中，一系列新的难题出现了。首先，对于几乎是最简单的转录组而言，杂交分析并不能产生充分的特异性来区分细胞内存在的所有 mRNA。这是因为，两个不同的 mRNA 可能具有相似的序列并可能交叉杂交到阵列中另一个特定的探针上。当两个或多个同源基因（5.2.1 节）在同一个组织中都有表达时这种情况会经常发生。某一个转录组中可能包含一簇相关的 mRNA，其中的每一个 mRNA 都可能会一定程度的杂交到该基因家族的不同成员。如此，要辨别每一个

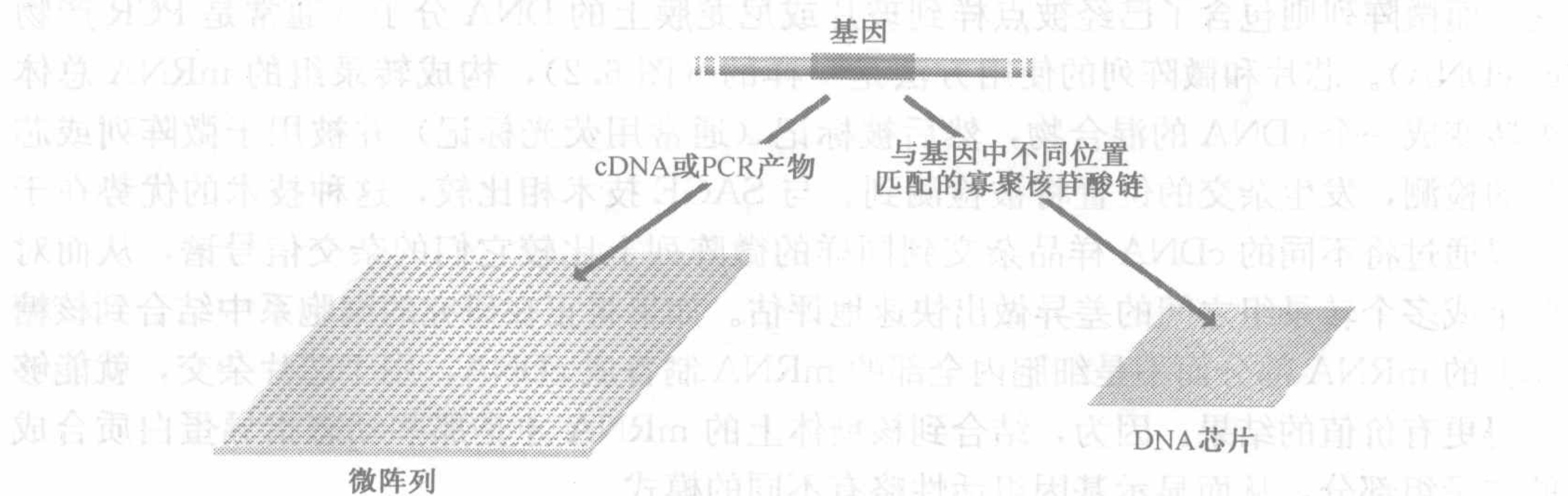


图 6.3 微阵列和 DNA 芯片

微阵列中的每一个点都包含来自感兴趣基因的 cDNA 或 PCR 产物，而 DNA 芯片上的每一个点包含了与相应基因不同区段匹配的寡聚核苷酸链的混合物

mRNA 的相对数量，甚或是确认哪一个 mRNA 存在与否都将是困难的。当两个或多个不同 mRNA 来自同一个基因时，一个类似的问题出现了。这种情况在脊椎动物中相对多见，其原因是选择性剪接（alternative splicing）——来自同一个前体 mRNA 的外显



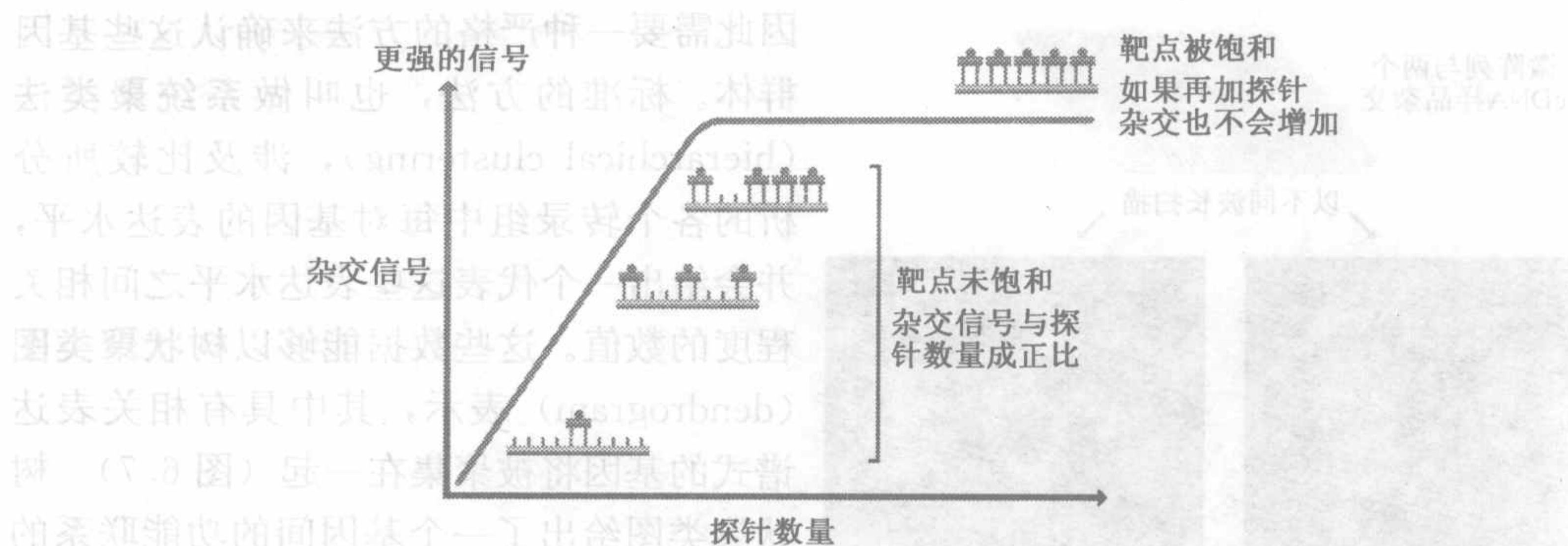


图 6.4 杂交的信号强度和探针数量的关系

子以不同的组合形式加工成一系列相关但又不同的成熟 mRNA (图 6.5)。如果想将上述变异都检测出来并做到精确的定量, 那么阵列就必须被仔细地设计。

如果研究的目的是为了比较两个或多个转录组的时候, 下面将讲述的新难题就进一步凸显了, 而且这是很常见的情景。为了使比较有效, 两个不同的微阵列或芯片上同一个基因之间的杂交强度的差异必须能够反映出 mRNA 数量的真实差异, 而不应该由于诸如阵列上靶 DNA 的数量、探针的标记效率或者杂交过程的效率等实验因素的不同造成的杂交强度差异。即使在单个实验室中, 这些因素也很难被绝对精确地控制好, 而要在不同实验室之间达到准确地重复则或多或少是不可能的事情。这就意味着数据的处理必须包括标准化过程, 以使来自不同阵列实验的结果能够被准确地比较。因此, 阵列中必须包含阴性对照, 这样就能确定每一次实验的背景信号强度。与此同时, 阳性信号应该每次都得到同样的信号强度。对于脊椎动物转录组, 肌动蛋白基因通常被作为阳性对照, 因为在不考虑发育阶段及疾病状态的情况下, 它在给定的组织中表达量相当的恒定。一个更加令人满意的选择是设计的实验使用同一芯片进行一次分析, 从而使得两个转录组能够直接进行比较。这可以通过使用不同的荧光探针来标记 cDNA 样品, 然后在合适的波长范围来扫描阵列, 以确定每个杂交点上两种荧光信号的相对强度, 由此来确定两个转录组的 mRNA 组成之间的差异 (图 6.6)。

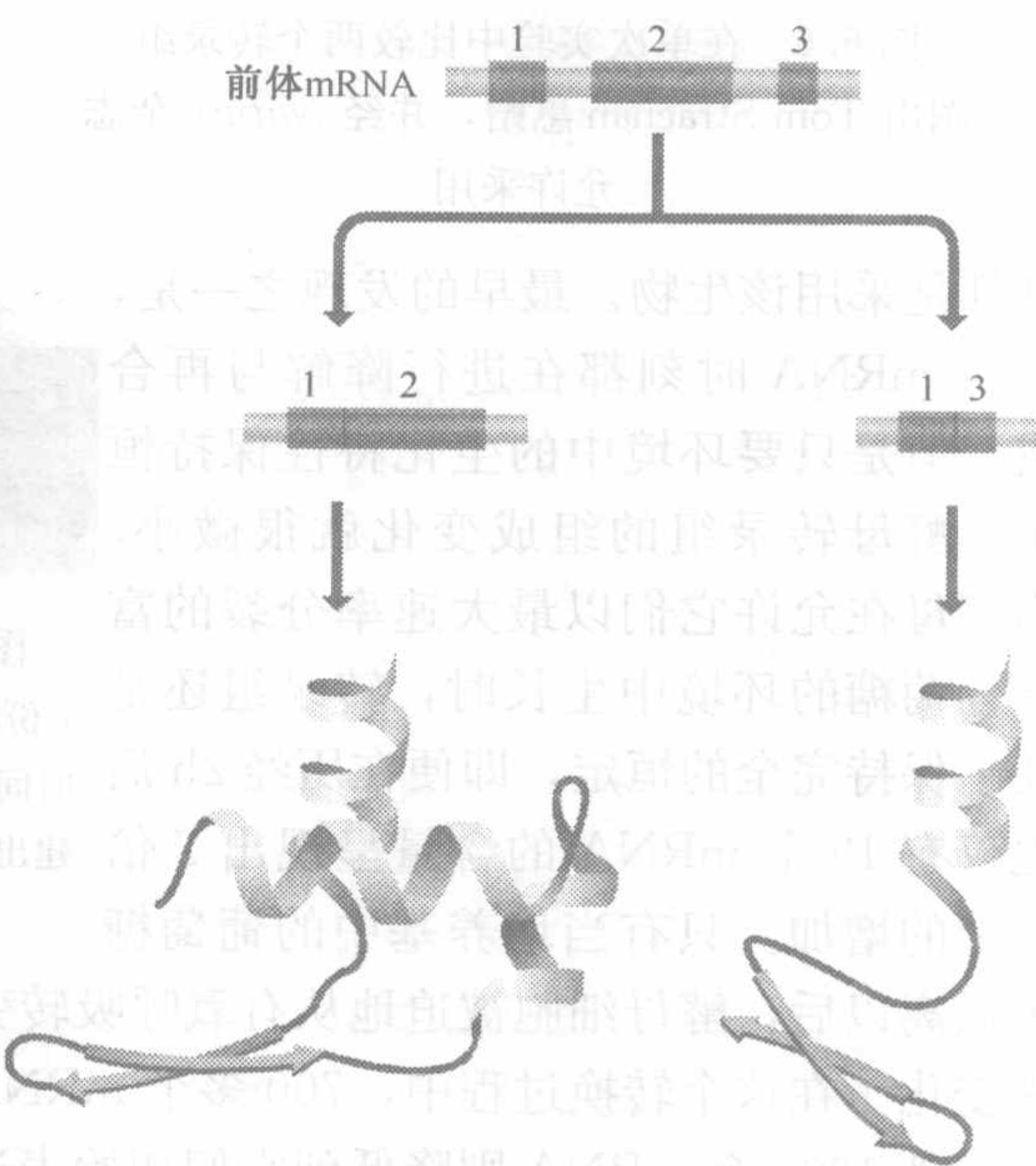


图 6.5 选择性剪接

选择性剪接产生不同的外显子连接组合, 导致由同一个前体 mRNA 合成出不同的蛋白质

假设两个或多个转录组之间能够进行准确的比较, 那么就可以区分出不同基因表达模式中极其复杂的差异。那些表现出相似表达谱的基因很可能是具有相关功能的基因,



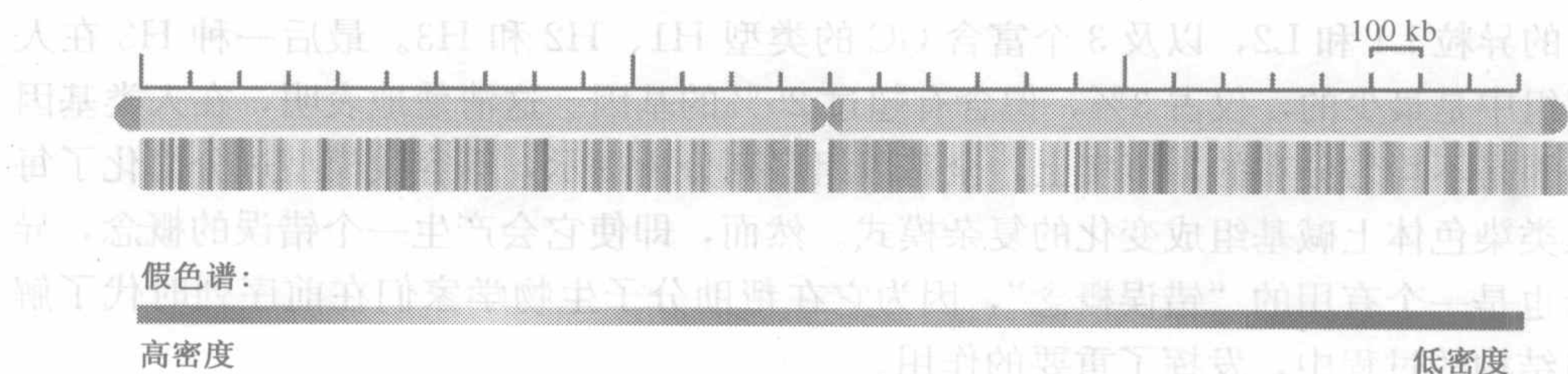


图 7.11 拟南芥 5 条染色体中最大染色体上基因的密度

1 号染色体长 29.1 Mb, 测序的部分用红色表示, 着丝粒和端粒用蓝色表示。染色体下方的基因图以伪彩色光谱给出了基因密度, 从深蓝色 (低密度) 到红色 (高密度)。密度为每 100 kb 有 1~38 个基因不等。经 AGI (The *Arabidopsis* Genome Initiative) 授权 (*Nature*, 408, 797~815. © 2000 Macmillan Magazines Limited)

在序列草图完成的前几年, 人们就提出过“人类染色体中基因不均匀分布”的这一猜测。有两个证据, 其中一个与染色体染色后出现的带型有关。染色时染料 (表 7.1) 与 DNA 分子结合, 但在大多数情况下对某些碱基对有倾向性。例如, 吉姆萨对富含 A 和 T 的 DNA 区域有较高亲和力。因此, 我们认为人类核型图中黑色的 G 带是基因组中富含 AT 的区域 (图 7.5)。A+T 占整个基因组的 59.7%, 所以 G 带中的 AT 含量肯定大于 60%。因此, 根据细胞遗传学的预测, 在 G 带中基因较少, 因为基因中的 AT 含量一般为 45%~50%。通过比较人类基因组序列草图与人类核型图, 人们确认了这种假设。

基因不均匀分布的第二个证据来自基因组结构的等容线 (isochore) 模型。根据这种模型, 脊椎动物和植物 (可能还包括其他真核生物) 的基因组是许多 DNA 片段的嵌合体, 每个片段至少长 300 kb, 并有统一的碱基组成但与邻近的片段之间有一些差别。下列实验也支持等容线模型: 将基因组 DNA 断裂成约 100 kb 的片段, 用特异性结合 AT 或 GC 富含区的染料处理, 再用密度梯度离心加以分离 (技术注解 7.1)。对于人类

### 技术注解 7.1 超速离心技术

#### 分离细胞成分和生物大分子的方法

20 世纪 20 年代高速离心的发展产生了从破裂细胞中分离细胞器和其他组分的技术。首先利用的是差速离心 (differential centrifugation) 技术, 即采用不同的转速, 收集细胞提取物中的较轻成分。例如, 完整细胞核相对较大, 可以经过 1000 g 离心 10 min 收集, 线粒体较轻, 要求 20 000 g 离心 20 min。设计好离心参数, 就可以得到较高纯度的不同细胞成分。

**密度梯度离心** (density gradient centrifugation) 首次使用于 1951 年。在这种方法中, 细胞成分不是在正常水溶液中离心。而是在离心管中加入蔗糖, 以形成密度梯度, 越往管底, 密度越大。细胞组分加在上方, 在非常高的转速下 (至少 500 000 g) 离心数小时。这种情况下, 细胞成分的迁移率有赖于其沉降系数, 而沉降系数又与分子的大小和形状有关。例如, 真核生物核糖体的沉降系数为 80S (S 指 Svedberg 单位, Svedberg 是瑞士科学家, 他首次将超速离心应用于生物学领域), 而较小的细菌核糖体的沉降系数为 70S。

第二种密度梯度离心法, 使用的是 8 mol/L 的氯化铯, 比用于测 S 值的蔗糖溶液密度高。这种溶液开始是均一的, 离心过程中才形成梯度。细胞成分沿离心管向下迁移, 但 DNA 和蛋白质等分子并不到达底部, 相反, 每种分子会停留在与其浮力密度 (buoyant density) 相同的介质中 (图 7.23)。这一技术在分子生物学中有很多应用, 可以分离碱基组成不同的 DNA 片段, 和不同构象的 DNA 分子 (如超螺旋、环状和线性的 DNA), 也可以分离正常 DNA 和重氮同位素标记的 DNA (15.1.1 节)。



基因组 DNA，可见 5 个组分，每个组分代表具有不同碱基组成的等容线模型：2 个富含 AT 的异粒 L1 和 L2，以及 3 个富含 GC 的类型 H1、H2 和 H3。最后一种 H3 在人类基因组中是最少的，仅占 3%，但含有超过 25% 的基因。这清楚地表明，在人类基因组中基因是不均匀分布的。事实上，全基因组序列分析提示，等容线模型过分简化了每一条人类染色体上碱基组成变化的复杂模式。然而，即使它会产生一个错误的概念，异粒理论也是一个有用的“错误概念”，因为它在帮助分子生物学家们在前序列时代了解基因组结构的过程中，发挥了重要的作用。

7.2.2 在核基因组中基因是如何组织的

因为真核生物染色体上的基因密度变化不一，这就意味着很难找到那些能代表全基因组的“典型”基因排列区域。尽管存在这样的困难，我们还是很明确一点：在不同真核生物中，基因分布总体上是很不一样的。但是，我们又需要了解这些差异，因为它们反映了这些基因组的遗传学特征和进化史上的重要差别。为了阐明这些差异，我们将在本章节中，深入了解人类基因组中的一小部分，并与其他物种基因组中的相应部分进行比较。在阅读这部分内容时，读者应该集中于所描述的明显特征，而单个染色体上发生的各种变化也提示我们：对于一个特定基因组中是否存在某种基因排布模式的问题，很难做出明确而又简洁的回答。

基因只构成了人类基因组中的一部分

在核基因组中基因是如何组织的？为回答这个问题，我们将学习一下第 12 号染色体上的一个 50 kb 的片段（图 7.12），这个片段中包含如下几个遗传学特征。

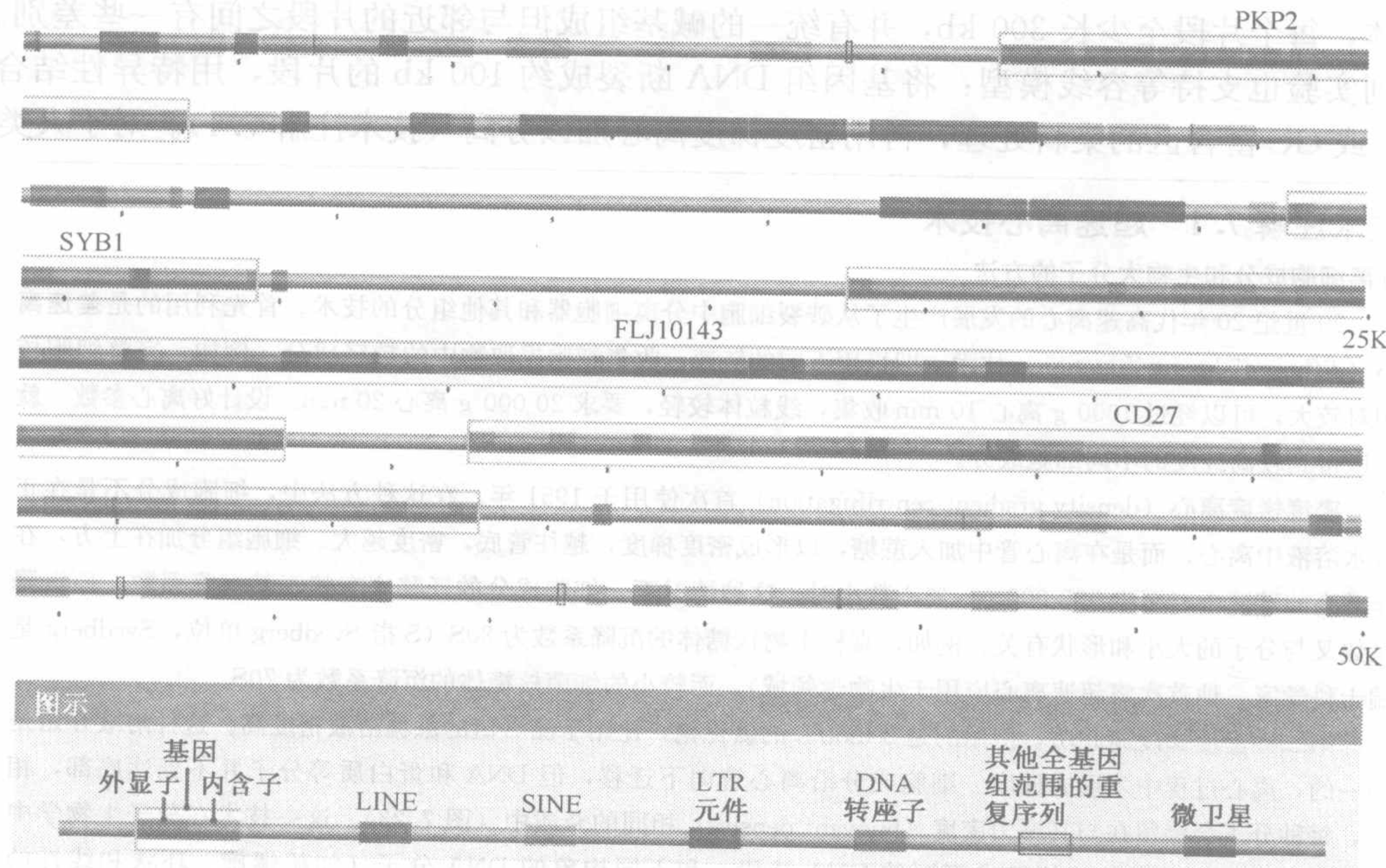


图 7.12 人基因组的一个片段

该图显示了人第 12 条染色体中的一个 50kb 片段中，基因、基因片段、全基因组重复序列和微卫星的定位







8.3.3 细胞器基因组的遗传组成

细胞器基因组比细胞核基因组小得多，因此我们推测其基因数目有限，事实也的确如此。线粒体基因组再次表现出很大的可变性，基因数目从恶性疟原虫的 5 个到原生动物异养鞭毛虫 (*Reclinomonas americana*) 的 92 个 (表 8.6)。所有线粒体基因组都含有非编码 rRNA 基因和至少一部分呼吸链组分蛋白的基因，而呼吸链正是线粒体的主要生化特征。基因含量更高的线粒体基因组还编码 tRNA、核糖体蛋白，参与转录和翻译的蛋白质以及将胞质蛋白运输进入线粒体的转运蛋白 (表 8.6)。大多数叶绿体基因组都含有大约 200 个基因，编码 rRNA、tRNA、核糖体蛋白以及参与光合作用的蛋白质 (图 8.13)。

表 8.6 线粒体基因组特点

特征	恶性疟原虫	莱茵衣藻	人	酿酒酵母	拟南芥	异养鞭毛虫
基因总数	5	12	37	35	52	92
基因类型						
蛋白质编码基因	3	7	13	8	27	62
呼吸复合物	3	7	13	7	17	24
核糖体蛋白	0	0	0	1	7	27
转运蛋白	0	0	0	0	3	6
RNA 聚合酶	0	0	0	0	0	4
翻译因子	0	0	0	0	0	1
功能性 RNA 基因	2	5	24	27	25	30
核糖体 RNA 基因	2	2	2	2	3	3
转运 RNA 基因	0	3	22	24	22	26
其他 RNA 基因	0	0	0	1	0	1
内含子数目	0	1	0	8	23	1
基因组大小/kb	6	16	17	75	367	69

从表 8.6 可以发现细胞器基因组的一般特征。这些基因组负责编码细胞器中的一些蛋白质，但并非全部。其他的蛋白质则由细胞核基因编码，在胞质合成，然后转运到细胞器中。如果细胞具有将蛋白质运输进入线粒体和叶绿体的机制，那为什么细胞器蛋白质不都由细胞核基因编码呢？对这个问题，我们还没找到令人信服的答案。但推测可能是因为一些细胞器基因组编码的蛋白质疏水性很强，不能通过线粒体和叶绿体膜，从而不能由胞质进入细胞器。细胞使这些蛋白质进入细胞器的唯一途径就是原地生产。

总结

原核生物包括两类截然不同的生物：细菌和古生菌。细菌基因组位于拟核——原核细胞中的轻微着色区。DNA 黏附在结合蛋白核心上，形成 40~50 个超螺旋环 DNA，



9 号和第 22 号染色体发生转位，产生一个异常产物，称之为费城染色体（Philadelphia chromosome），可以导致慢性髓性白血病（图 10.3）。如此同样的转位反复发生说明在核内这些染色体的交互地域通常是比较接近的。也有证据表明，至少在某些生物体内，某些染色体是比较倾向于占据核周位置的，该地域的基因组表达相对较少，也正是这些染色体包含活性基因较少，如鸡基因组的大染色体（7.1.2 节）。

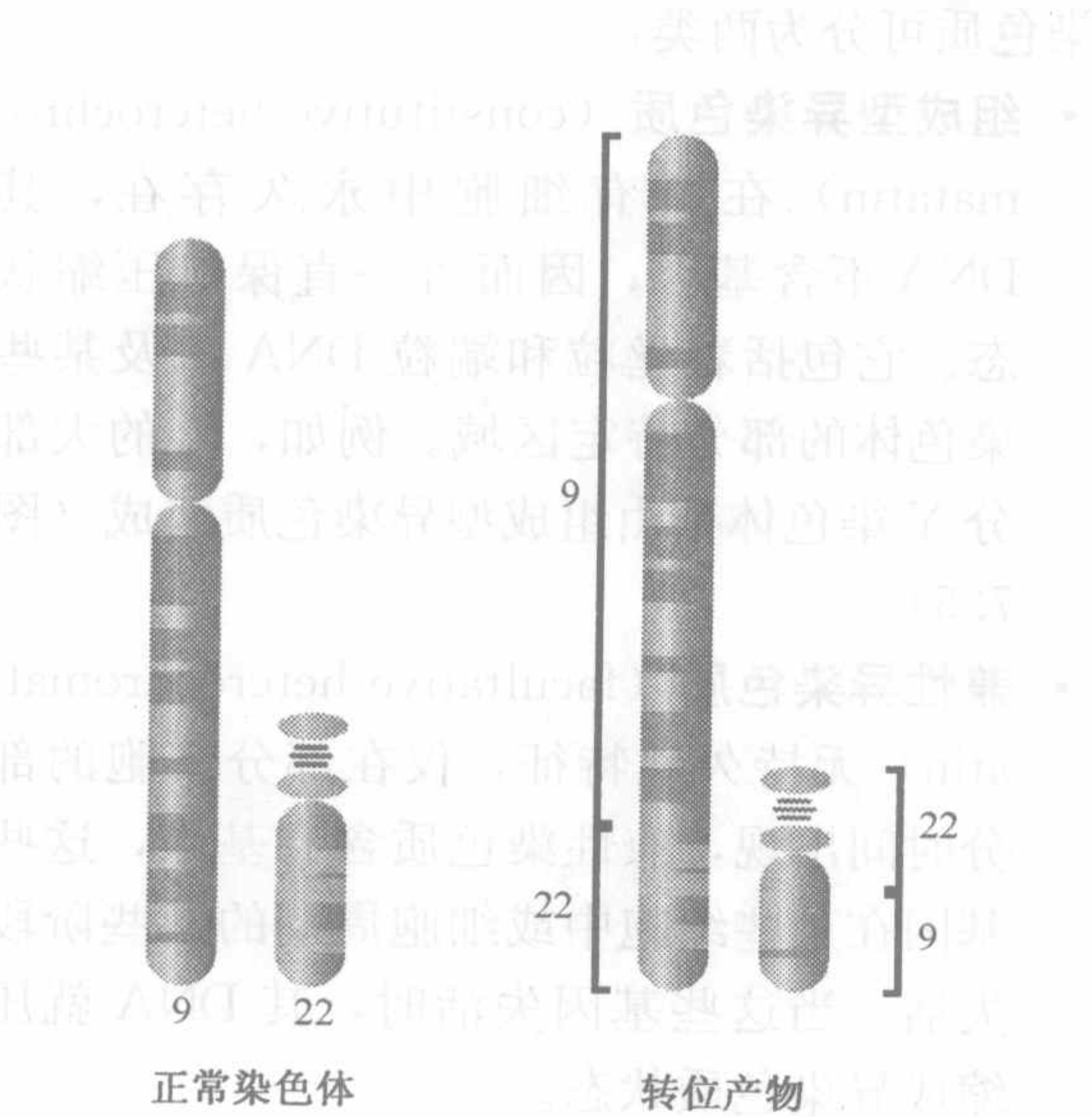


图 10.3 人第 9 号和第 22 号染色体转位产物  
左边显示的是人第 9 号和第 22 号染色体，右边是转位产物。费城染色体指的是转位产物中较小的那个，通常缺口会被正确修复，但偶尔的错误修复会在造成杂交产物。可以认为，费城染色体之所以相对高频率地发生，是因为人第 9 号和第 22 号染色体在细胞核内占据了相对接近的地域。人第 9 号染色体的缺口位于 ABL 基因，其产物参与了信号通路（14.1.2 节）。转位之后则在这个基因的起点加上了一个新的编码序列，从而生成了一个可以导致细胞转化并产生慢性髓性白血病的异常蛋白质

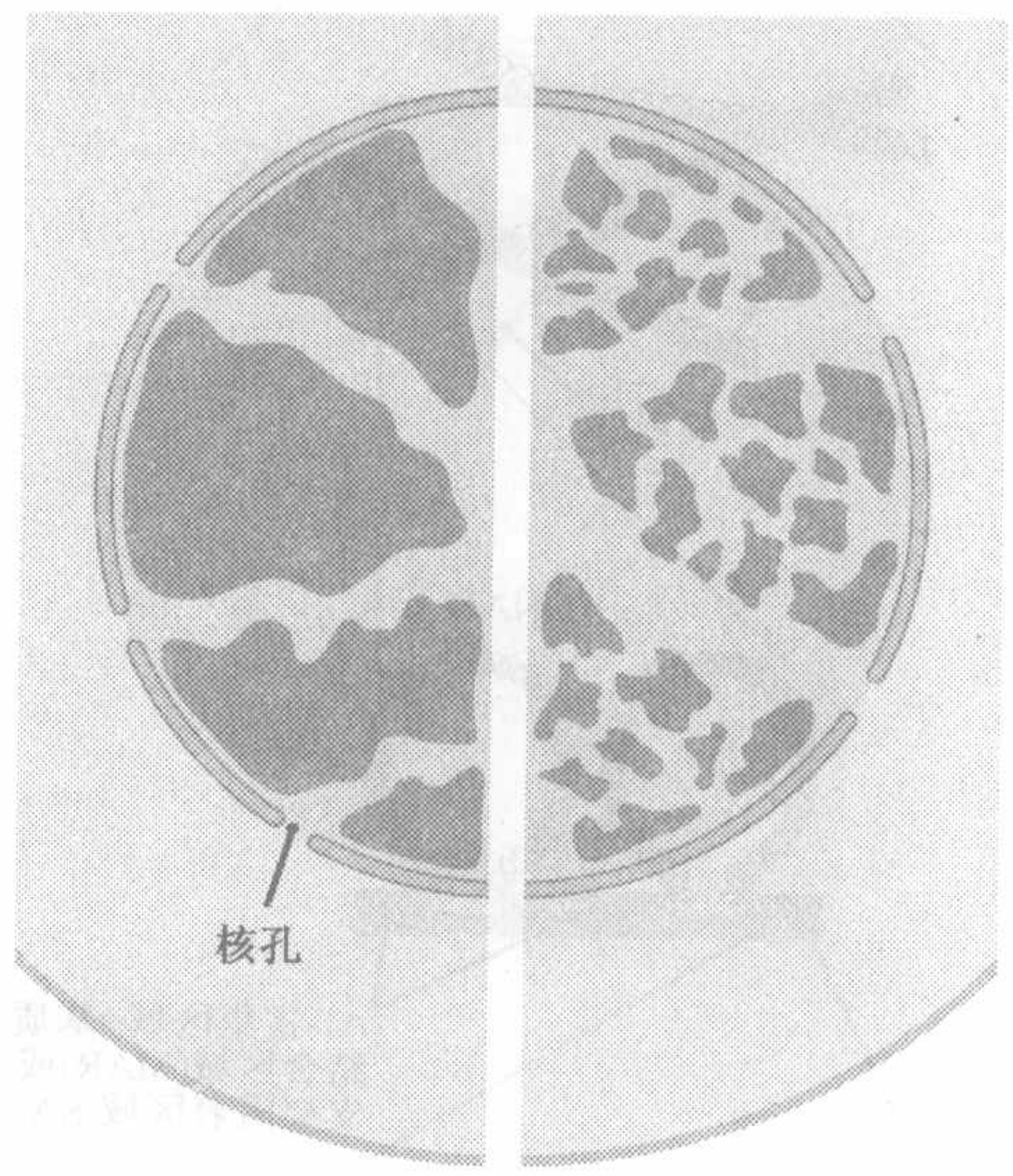


图 10.4 染色体地域  
左半图显示了每一个地域形成一个区的原始模型，提示活性基因位于各个地域的表面。右半图显示了通道贯通于地域之中的改进模型

另外一个富有争议的话题是活性基因在每个染色体地域内的定位。人们曾经一度认为，活性基因一般在各个地域的表面，接近非染色质区域，因此可以接近参与基因转录的酶和蛋白质（图 10.4）。现在，这种看法受到了质疑，因为实验表明转录出来的 RNA 既在各个地域的表面，又在其内部。更细致的显微观察发现在染色体各地域之间有隧道贯穿，从而连接了非染色质区域的不同部分，这样转录机器就可以进入这些地域的内部了。

### 10.1.2 染色质区域

在第 7.1.1 节中我们已经知道染色质是真核细胞核中基因组 DNA 和染色体蛋白质形成的复合体。染色质的结构也是多层次的，它存在着以核小体和 30 nm 染色质纤丝（图 7.2 和图 7.3）为主的两种最低包装形式，直至细胞有丝分裂过程中才可见的最高压缩形式即有丝分裂中期染色体。有丝分裂结束后，染色体逐渐松散，以至于不能分辨



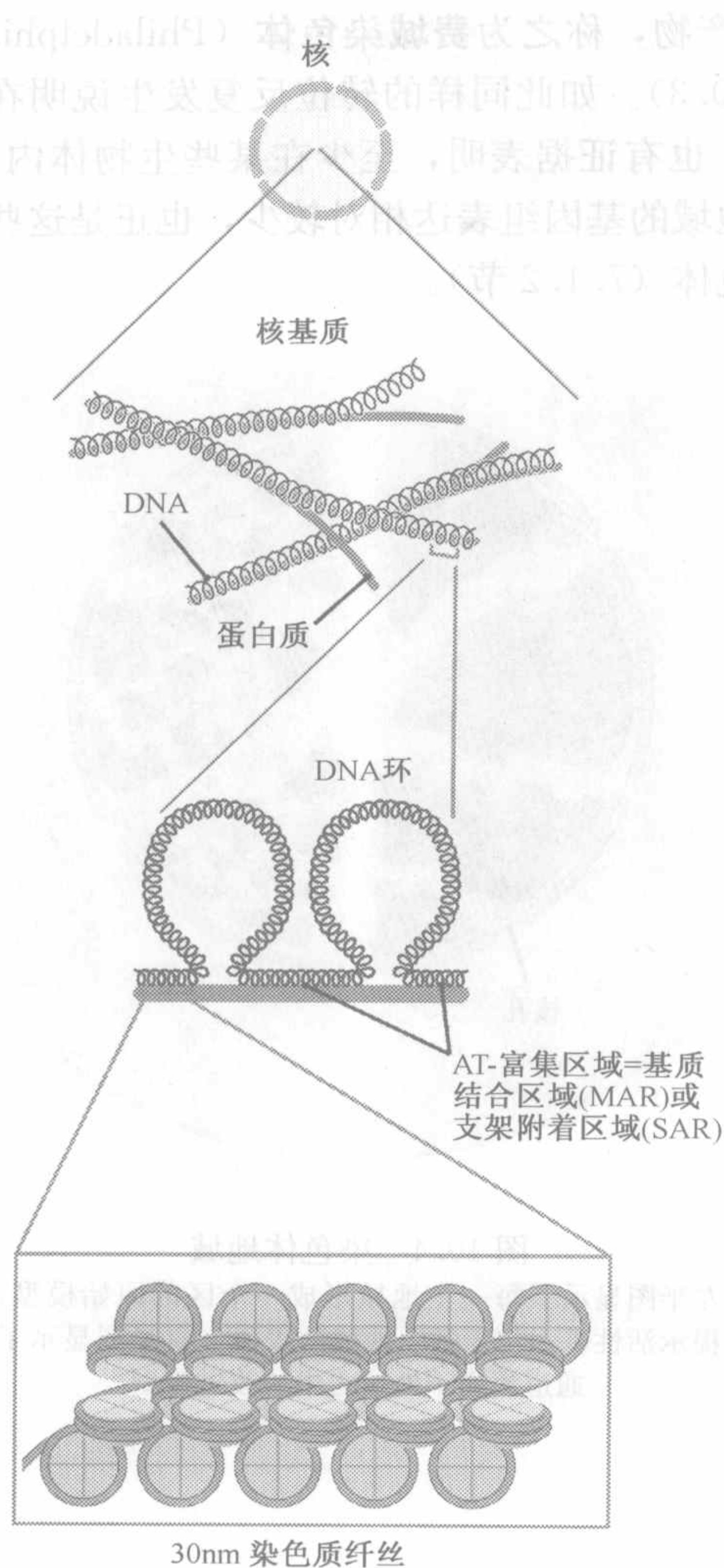


图 10.5 DNA 在细胞核中组织的一种方案  
核基质是基于蛋白质的纤维状结构，它的精细组成和在核中的排列还没有描述。一般认为，主要以 30 nm 染色质纤丝形式存在的常染色质（图 7.3）通过富含 AT 的称为基质结合区或支架附着区域（MAR 或 SAR）的序列附着在基质上（图 10.5）。

连接在核基质结合位点之间的 DNA 环称为**结构域**（structural domain）。表达的基因周围的 DNA 区域可被定义为**功能域**（functional domain）。二者的确切关系是个非常有趣的问题。用脱氧核糖核酸酶 I（DNaseI）处理一段纯化的染色质，可界定功能域的范围。DNaseI 是一种 DNA 结合蛋白，不能接近 DNA 的致密区（图 10.6）。对 DNaseI 敏感的区域会延伸到正在表达的一个或一套基因的两端，虽然我们还不清楚这种结构是 30 nm 纤丝还是串珠样的结构，但这表明该区域染色质形成了较开放的结构 [图 7.2

其各自的单体结构。光学显微镜观察非分裂状态的细胞核时，只能观察到核内是由着色深浅不同区域的混合体组成的。深色区域称为**异染色质**（heterochromatin），包含相对致密但不如中期染色体结构紧密的结构。异染色质可分为两类：

- **组成型异染色质**（constitutive heterochromatin）在所有细胞中永久存在，其 DNA 不含基因，因而可一直保持压缩状态。它包括着丝粒和端粒 DNA 以及某些染色体的部分特定区域。例如，人的大部分 Y 染色体都由组成型异染色质组成（图 7.5）。
- **兼性异染色质**（facultative heterochromatin）无持久性特征，仅在部分细胞的部分时间出现，兼性染色质含有基因，这些基因在某些细胞中或细胞周期的某些阶段失活。当这些基因失活时，其 DNA 就压缩成异染色质状态。

一般认为异染色质的结构是高度致密的，参与基因表达的蛋白质不能接近 DNA。相反，含有活性基因的其他染色体 DNA 区域则相对较疏松，可允许参与表达的蛋白质进入，这些区域叫做**常染色质**（euchromatin），它们在整个细胞核中分散存在。DNA 在常染色质中具体的结构方式还不清楚，但用电子显微镜可以看到许多环状 DNA，它们主要以 30 nm 染色质纤丝的形式存在，长度介于 40~100 kb。这些环通过富含 AT 的 DNA 片段即**基质结合区域**（matrix-associated region, MAR）或叫**支架附着区域**（scaffold attachment region, SAR）与核基质相连



细菌 RNA 聚合酶结合于启动子序列

细菌中 RNA 聚合酶结合的靶序列叫做启动子 (promoter)。这个术语最先在 1964 年被遗传学家用于描述乳糖操纵子中位于三个基因紧邻上游的区段的功能 (图 11.15)。当该位点突变失活后, 基因就不能表达: 因此这一位点似乎负责启动下游基因的表达。现在我们知道这是由于该位点是负责转录操纵子的 RNA 聚合酶的结合位点。

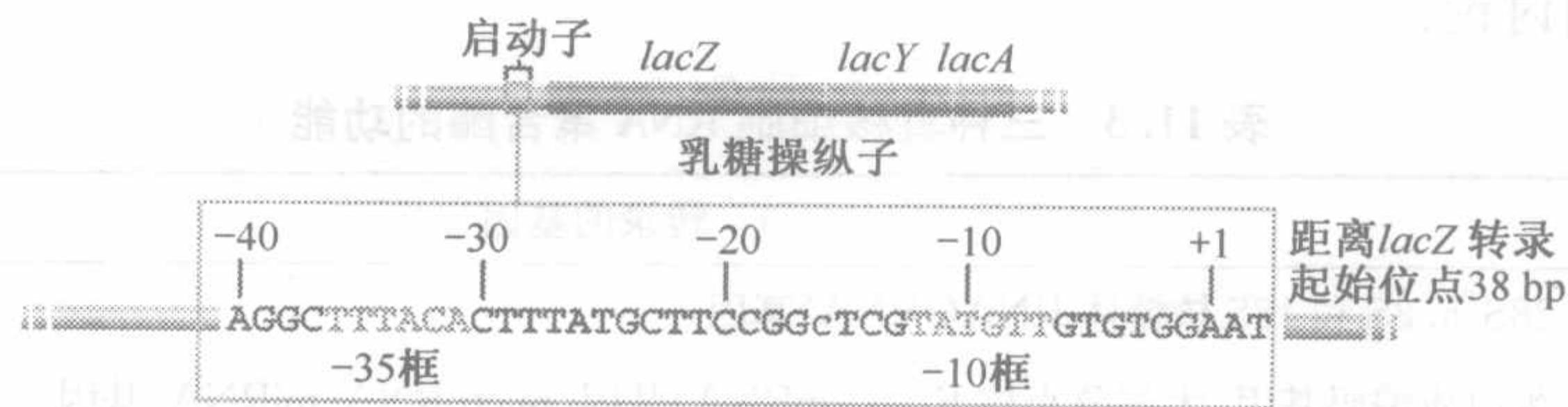


图 11.15 大肠杆菌乳糖操纵子的启动子

启动子位于操纵子中第一个基因 *lacZ* 的立即上游。DNA 序列显示了启动子的 2 个独特序列组分—35 和—10 框的位置。这些序列与一致序列的比较见正文, 关于乳糖操纵子的详细介绍见图 8.8 (A)

大肠杆菌的启动子序列最早是从对一百多个基因上游区段的比较中获得的。过去一般认为所有基因的启动子序列都是相似的, 所以可通过基因上游区段的比较识别出来。分析表明, 大肠杆菌启动子由两部分组成, 每个长 6 个核苷酸, 记作 (图 11.15):

—35 框                      5'-TTGACA-3'  
—10 框                      5'-TATAAT-3'

这些是共有序列, 因此可用于描述大肠杆菌中所有启动子序列的“平均”情况; 但对于每个特定基因可能会稍有不同 (表 11.4)。框的名字表明它们相对转录起始位点的位置, 转录起始点的核苷酸记作“+1”, 它位于基因编码区起点上游 20~600 个核苷酸。两个框之间的间隔距离非常重要, 因为必须确保两个基序位于双螺旋的同一侧, 方能促进它们与 RNA 聚合酶上 DNA 结合组分之间的相互作用 (11.2.3 节)。

表 11.4 大肠杆菌启动子序列

启动子	序列	
	—35 框	—10 框
共有序列	5'-TTGACA-3'	5'-TATAAT-3'
乳糖操纵子	5'-TTTACA-3'	5'-TATGTT-3'
色氨酸操纵子	5'-TTGACA-3'	5'-TTAACT-3'

真核细胞启动子更加复杂

在真核生物中, “启动子” 一词用于指所有对基因转录起始有重要作用的序列。对于某些基因, 这些序列可能有多个, 各具有不同功能, 不仅包括作为起始复合物组装位点的核心启动子 (core promoter) [有时也叫做基本启动子 (basal promoter)], 还包括一个或多个位于核心启动子上游的上游启动子元件 (upstream promoter element)。核心启动子上起始复合物的组装通常可以在没有上游元件的情况下发生, 但效率很低。这



表明结合于上游元件的蛋白质中至少有一些是可以“启动”基因表达的转录激活物，也证实了这些序列在“启动子”中的存在。

真核生物的三种 RNA 聚合酶分别识别不同的启动子序列；事实上，是不同启动子间的差异限定了哪些基因可以被哪类聚合酶转录。对于脊椎动物的启动子详述如下(图 11.16)：

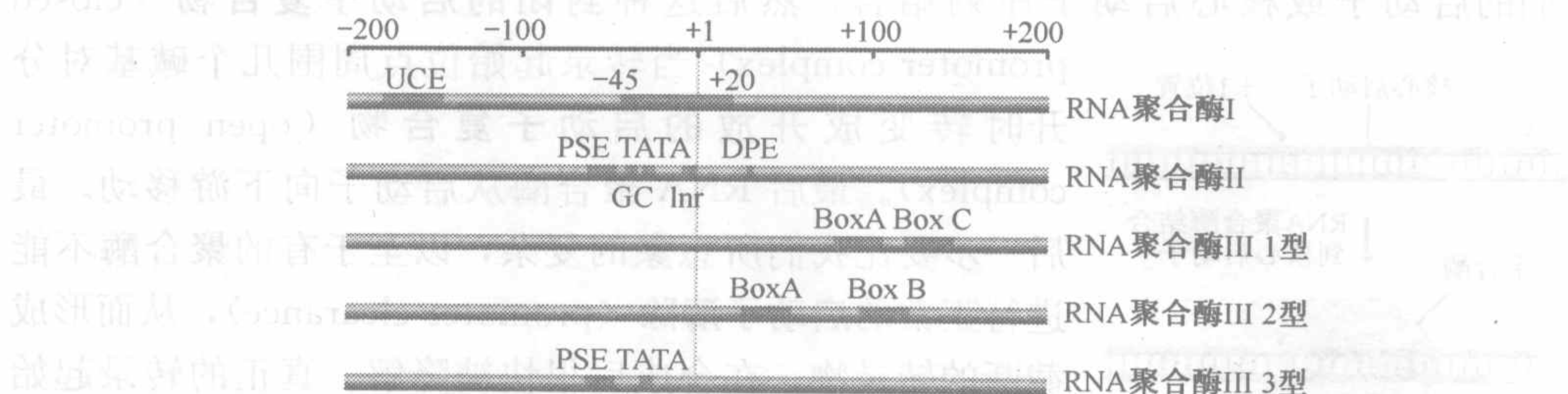


图 11.16 真核启动子的结构  
缩写见正文

- RNA 聚合酶 I 启动子由覆盖转录起始位点核苷酸 $-45 \sim +20$ 的核心启动子和上游大约 100 bp 的上游控制元件 (upstream control element) 组成。
- RNA 聚合酶 II 启动子变化较多，可延伸至转录起始位点上游几千个碱基。核心启动子由两部分组成， $-25$  或 TATA 框 (TATA box) (共有序列为  $5'$ -TATAWAWAR- $3'$ ，其中 W 代表 A 或 T，R 代表 A 或 G) 和位于  $+1$  位核苷酸附近的起始子序列 (initiator sequence, Inr) (共有序列为  $5'$ -YCANTYY- $3'$ ，其中 Y 代表 C 或 T、N 代表任意核苷酸)。RNA 聚合酶 II 转录的某些基因只含有这两种核心启动子组分中的一种，还有一些居然两种组分都不含有。后者称为“裸”基因，虽然它们的转录起始位点不如含有 TATA 和 (或) Inr 序列的基因固定，但它们仍然可以被转录。某些基因还有其他的序列，也被看作核心启动子，例如：
  - 下游启动子元件 (downstream promoter element, DPE, 位于  $+28 \sim +32$ )，序列不固定，因其能结合 TFIID 而被鉴定，TFIID 蛋白质复合体在转录起始中起着重要作用 (11.2.3 节)
  - TATA 框立即上游的 7 bp GC 富含基序，由 TFIIB 识别，TFIIB 是起始复合体的另一成分。
  - 近端序列元件 (proximal sequence element, PSE)，位于 RNA 聚合酶 II 转录的 snRNA 基因上游的  $-45 \sim -60$  处。
  - 与核引物相同，RNA 聚合酶 II 转录基因有不同的上游引物序列，其功能可见 11.3.2 节。
- RNA 聚合酶 III 启动子不固定，可以分为至少三类。其中两种不太一样，它们位于所转录基因的內部，通常这些序列覆盖  $50 \sim 100$  bp，两侧区域是由一个可变区分开的两个保守序列框组成。另外一类与 RNA 聚合酶 II 启动子非常相似，含有 TATA 框和多个上游启动子元件 (有时候包括上面提到的近端序列元件)。有趣的是，这种排列也见于小分子核 RNA 基因家族中的 U6 基因，但该家族的其他成员却由 RNA



苷酸尾，许多前 mRNA 还包含内含子，因而需进行剪接，另外少数 mRNA 要进行 RNA 编辑。“加帽”的功能已经被了解，但多聚腺苷酸化的原因仍然是个谜。对于剪接和编辑，已经清楚这些事件的成因（前者将阻止 mRNA 翻译的内含子除去，后者改变 mRNA 的编码特性），但这些机制进化的原因尚不清楚。为什么基因首先要有内含子？为什么要编辑 RNA 而不是在 DNA 中编码所需要的序列？

真核细胞 mRNA 边合成边加工，转录一开始即加帽，转录仍在进行时剪接和编辑就开始了，而多聚腺苷酸化是 RNA 聚合酶 II 终止机制固有的一部分。因为有太多不同的事件需要同时描述，所有这些过程放到一起处理易引起混乱，因此我们将把 RNA 编辑放到本章的最后介绍，因为它具有与 rRNA 和 tRNA 加工中的化学修饰相似的方式。我们将在研究了加帽、延伸和多聚化反应之后再讨论有关剪接的问题。

## 在转录起始之后，RNA 聚合酶 II 转录物的加帽反应立即发生

虽然，真核生物中 mRNA 编码基因的转录起始阶段最后一步是 RNA 聚合酶 II 的最大亚基 C 端结构域 (CTD) 发生磷酸化 (11.2.3 节)，但紧接其后的并非就是延伸的起始。在面对那些区分“启动子清除” (promoter clearance) 与“启动子脱离” (promoter escape) 的事件时，还存在一些理解上的灰色地带。启动子清除是指由前起始复合体向一个已经开始合成 RNA 的复合体过渡；启动子脱离是聚合酶脱离启动子区域开始产生转录物 (图 12.20)。正、负延伸因子相反的效应会影响聚合酶开始合成有效 RNA 的起始能力。如果负性因子占主导地位，则在聚合酶尚未移动到离转录起点 30 个核苷酸以外时，转录即中断。所以，启动子脱离会成为一个重要的调控点，但是尚不知道此阶段的调控是如何进行的。

成功的启动子脱离同加帽过程是相关联的，加帽会在转录物达到 30 个核苷酸长度以前完成。加帽的第一步是加一个附加的鸟苷酸到 RNA 的 5' 端，它不是通过正常的 RNA 聚合反应，而是通过末端核苷酸的 5' 三磷酸与 GTP 的三磷酸之间的反应完成。末端核苷酸的  $\gamma$ -磷酸 (最外面的磷酸) 和 GTP 的  $\beta$  和  $\gamma$  磷酸被去除，形成一个 5'-5' 键 (图 12.21)。该反应被鸟苷转移酶 (guanylyl transferase) 催化。第二步反应是在鸟苷酸甲基转移酶 (guanine methyltransferase) 的催化下将一个甲基加到鸟嘌呤环的 7 位 N 原子上，使鸟嘌呤变成 7-甲基鸟嘌呤。这两种酶与 CTD 有接触，它们可能都是启动子清除时 RNA 聚合酶 II 复合物的内在成分。

这种 7-甲基鸟嘌呤结构叫做 0 型帽子，是酵母中最常见的形式。另外在较高等的真核生物中会发生更多的修饰 (图 12.21)。

• 第二个甲基取代第二个核苷酸的 2'-OH 上的 H，称 1 型帽子 (type 1 cap)。

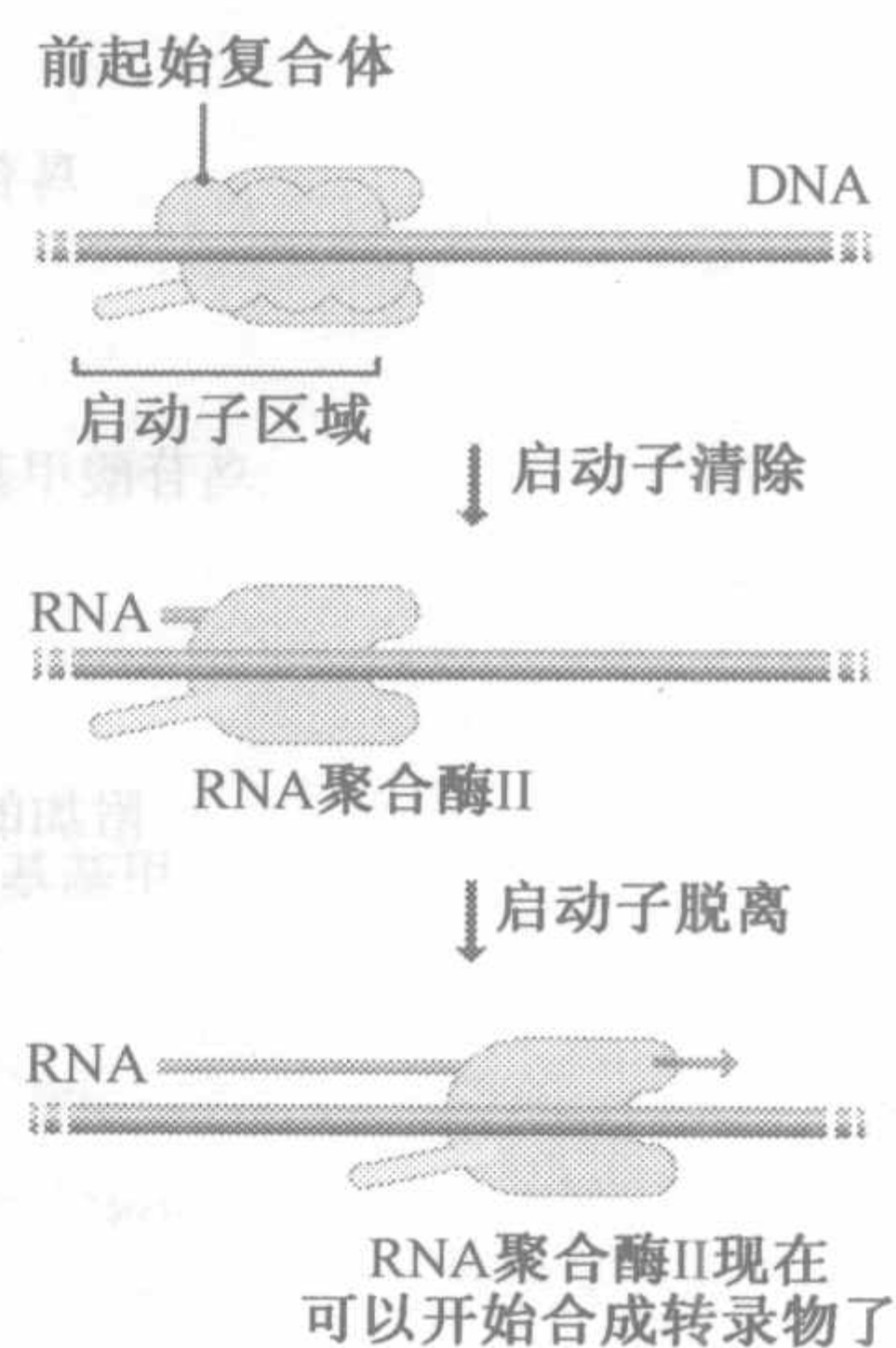


图 12.20 启动子清除与启动子脱离

启动子清除指由前起始复合体向一个已经开始合成 RNA 的复合体过渡；启动子脱离是聚合酶脱离启动子区域开始产生转录物。注意图中仅为示意，并不反映合成转录物的 RNA 聚合酶 II 复合体的形状或亚基组成



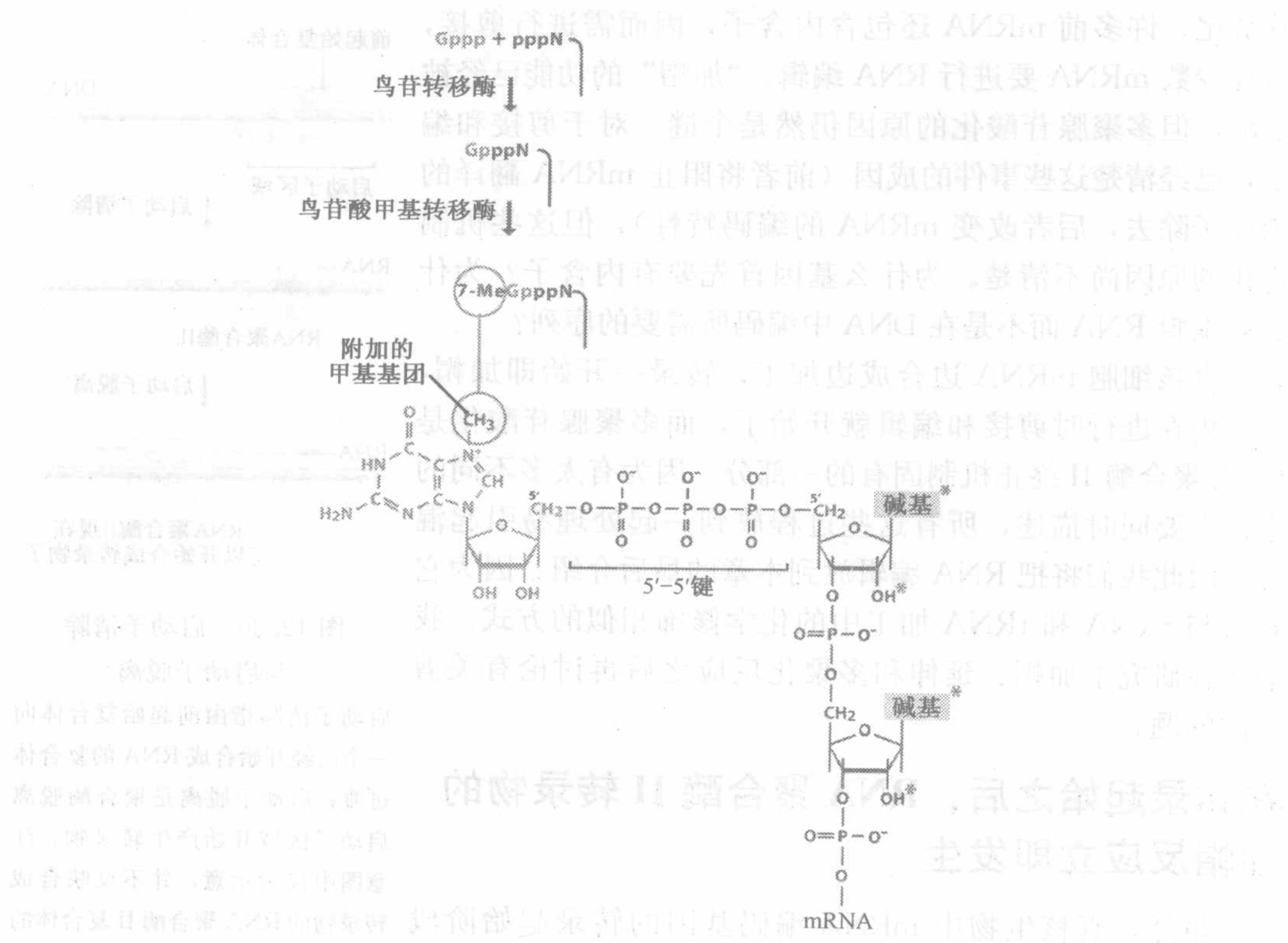


图 12.21 真核 mRNA 加帽

图中上半部分说明的是加帽反应。一个 GTP 分子 (Gppp) 与 mRNA 的 5' 端反应形成一个三磷酸键。然后末端 G 的第 7 位 N 原子被甲基化。下半部分显示的是 0 型帽子的化学结构, 星号显示的是在 1 型和 2 型帽子结构中可能发生甲基化的位置

- 如果第二个核苷酸是 A, 则甲基可能加到 A 的 6 位 N 原子上。
- 第三个核苷酸的 2'-OH 也可能加上甲基, 形成 2 型帽子 (type 2 cap)。

RNA 聚合酶 II 催化合成的所有 RNA 都以某种方式加帽。这意味着与 mRNA 相同, 由该酶转录的 snRNA 也被加帽 (表 11.3)。帽子结构对于 snRNA 和 mRNA 从核中运出可能也是重要的 (12.2.7 节), 但研究得最清楚的是它在 mRNA 翻译中的作用, 这一点我们将在 13.2.2 节中介绍。

## 真核细胞 mRNA 的延伸

如上所述, 细菌和真核细胞中转录延伸过程基本上是相同的。一个主要的区别是形成的转录物的长度。最长的细菌基因只有几个 kb, 在几分钟之内就能被转录, 细菌 RNA 聚合酶的转录速度是每分钟几百个核苷酸。相反, 尽管真核细胞 RNA 聚合酶 II 每分钟能合成 2000 个核苷酸, 但它需要数小时才能转录一个基因, 这是因为在许多真核细胞基因中存在多个内含子 (12.2.2 节), 因而被转录的 DNA 相当长。例如, 人 dystrophin 基因的前体 mRNA 长 2400 kb, 它的合成需 20 h。

由于真核细胞基因非常长, 因而需要转录复合物保持稳定。RNA 聚合酶 II 本身不能满足这一要求。当在体外应用纯化的酶进行催化反应时, 因为酶经常在模板上暂停或



基因组表达的最终结果是蛋白质组，即活细胞合成的功能性蛋白质的合集。蛋白质组中各种蛋白质类型及其相对丰度，代表了新蛋白质合成和已有蛋白质降解之间的平衡。化学修饰及其他加工作用也能改变蛋白质组的生化功能。合成、降解和修饰/加工共同作用使蛋白质组可以满足细胞的变化需求，并对外界刺激作出反应。

本章我们将学习蛋白质组成分的合成、加工和降解。为了理解蛋白质的合成，我们将首先学习 tRNA 在解读遗传密码中的作用，以及核糖体中发生的、最终使氨基酸聚合成多肽的事件。核糖体事件有时被认为是单个基因表达的最终阶段，但刚合成的多肽要在折叠以后才有活性，并可能必须经历切割和化学修饰才能发挥功能。我们将在 13.3 节学习这些加工事件。在本章的最后，我们将学习细胞如何降解不再需要的蛋白质。

## 13.1 tRNA 在蛋白质合成中的作用

tRNA 在翻译中起关键作用。1956 年 Francis Crick 预言了它们的存在，作为接头分子，它们连接 mRNA 和合成中的多肽。这既是一种物理连接，tRNA 结合到 mRNA

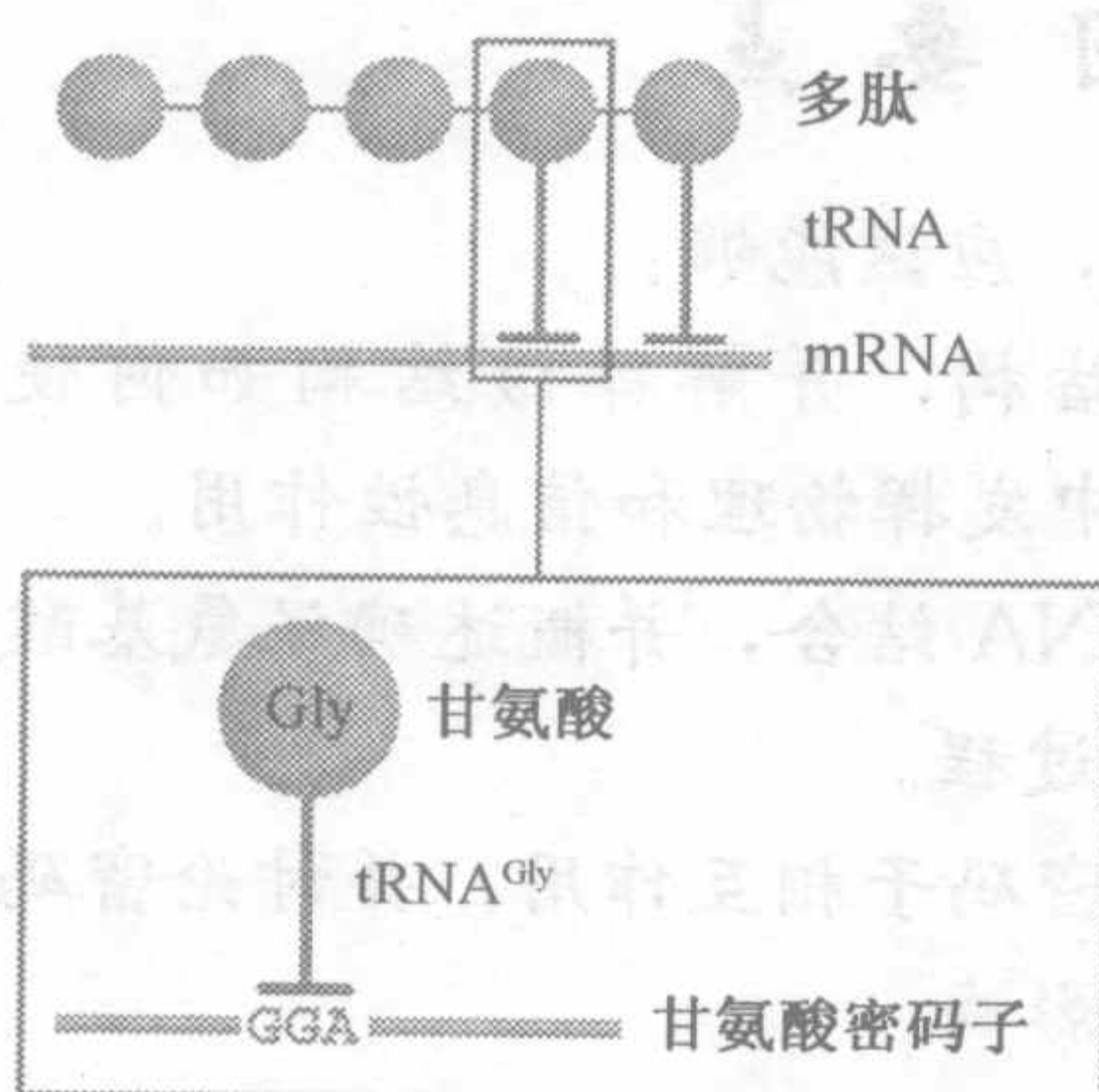


图 13.1 tRNA 在翻译中的接头作用

上图显示了 tRNA 的物理作用，即在多肽和 mRNA 之间形成连接。下图显示了信息连接，即 tRNA 携带由三联体密码子限定的氨基酸，并与密码子结合

和延伸中的多肽上；同时又是一种信息连接，通过 mRNA 中核苷酸序列的遗传密码，tRNA 确保合成中的多肽具有对应的氨基酸序列（图 13.1）。为了理解 tRNA 如何发挥这种双重作用，我们必须先研究氨酰化（aminoacylation），在此过程中正确的氨基酸与对应的 tRNA 相连接，以及密码子-反密码子识别（codon-anticodon recognition），即 tRNA 和 mRNA 之间的相互作用。

### 13.1.1 氨酰化：将氨基酸连接到 tRNA 上

细菌含有 30~45 种不同的 tRNA，真核生物则多达 50 种。因为遗传密码仅指定 20 种氨基酸，这意味着所有生物中至少存在部分同工 tRNA (isoaccepting tRNA)，即对应相同氨基酸的不同 tRNA。描述 tRNA 的术语中，用氨基酸特异性的上标再加数字后缀，如 1、2 等，表示，以区分不同的同工 tRNA，例如，两种对应甘氨酸的 tRNA 表示为 tRNA<sup>Gly1</sup> 和 tRNA<sup>Gly2</sup>。

### 所有 tRNA 都具有相似的结构

最小的 tRNA 只有 74 个核苷酸，而最大的也很少超过 90 个核苷酸。因为它们很小，且可以纯化出单一类型的 tRNA，所以它们是最早被测序的核酸之一，早在 1965 年纽约康奈尔大学 Robert Holley 及其同事就已对其进行了测序。测序结果发现了一些意料之外的特征：除标准 RNA 核苷酸（A、C、G 和 U）之外，tRNA 还含有一些被修饰的核苷酸，每种 tRNA 中含有 5~10 个，已发现共 50 种以上的修饰类型（12.1.3 节）。



第一个被测序的 tRNA 分子是酿酒酵母 (*Saccharomyces cerevisiae*) 的 tRNA<sup>Ala</sup>, 结果表明, 这个分子可形成多种碱基配对的二级结构。随着越来越多的 tRNA 被测序, 逐渐清楚所有 tRNA 都可以形成一种特定的结构, 即三叶草 (cloverleaf) 结构(图 13.2), 具有如下的特征:

- **受体臂** (acceptor arm), 由分子 5' 端和 3' 端之间的 7 个碱基对组成。氨基酸连接到 tRNA 3' 端, 即 tRNA 都具有的 CCA 末端序列的腺苷酸上 (12.1.3 节)。
- **D 臂** (D arm), 因其结构中始终存在修饰核苷二氢尿嘧啶而得名 (图 12.18)。
- **反密码子臂** (anticodon arm), 含有反密码子三联体核苷酸, 在翻译时与 mRNA 进行碱基配对。
- **V 环** (V loop), 在 I 类 tRNA 中含 3~5 个核苷酸, 在 II 类 tRNA 中含 13~21 个核苷酸。
- **T $\Psi$ C 臂** (T $\Psi$ C arm), 因其序列中总含有胸腺嘧啶-假尿嘧啶-胞嘧啶序列而得名。

几乎所有 tRNA 都可以形成三叶草结构, 主要的例外是脊椎动物线粒体所使用的 tRNA, 它由线粒体基因组编码, 有时缺失部分结构, 如人线粒体 tRNA<sup>Ser</sup> 就没有 D 臂。除了保守的二级结构以外, tRNA 某些部位的核苷酸完全恒定 (总是同一个核苷酸) 或半恒定 (总是嘌呤或总是嘧啶), 而且修饰核苷酸的位置也几乎总是相同。

很多恒定核苷酸的位置对于 tRNA 的三级结构很重要。X 射线晶体学研究显示, D 环和 T $\Psi$ C 环的核苷酸形成碱基配对, 使 tRNA 折叠成紧密的 L 型结构(图 13.3)。L 型结构的每个臂长约 7 nm, 直径约为 2 nm, 氨基酸结合位点位于一条臂末端, 反密码子位于另一条臂的末端。这种附加的碱基配对意味着碱基堆积力 (1.1.2 节) 从 tRNA 的一端到另一端基本上是连续的, 从而保证了结构的稳定性。

## 氨酰 tRNA 合成酶将氨基酸连到 tRNA 上

tRNA 的氨酰化 (分子生物学中称之为“负载”) 由称为氨酰-tRNA 合成酶 (aminoacyl-tRNA synthetase) 的一组酶催化完成。氨酰化反应分两步进行。氨基酸首先和 ATP 反应形成活化的氨基酸中间体, 然后这个氨基酸被转到 tRNA 的 3' 端, 在氨基酸的 -COOH 基团与 tRNA 最后一个核苷酸 (总是 A) 糖基 2' 或 3' 碳原子的 -OH 基团之间形成连接 (图 13.4)。

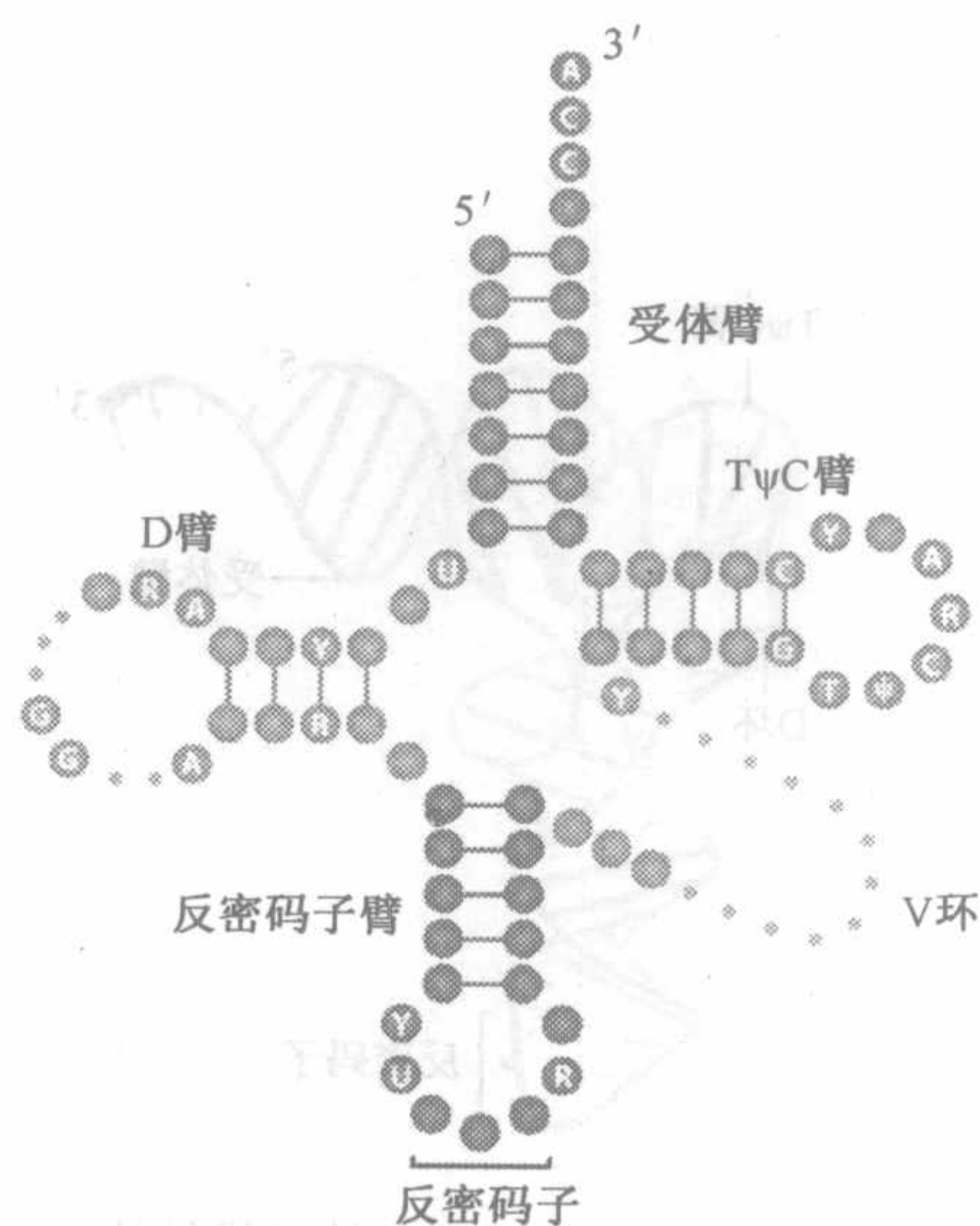


图 13.2 tRNA 的三叶草结构

tRNA 以传统的三叶草结构表示, 不同组分都作了标记, 并指出了恒定核苷酸 (A、C、G、T、U、 $\Psi$ , 其中  $\Psi$  是假尿嘧啶) 和半恒定核苷酸 (缩写: R, 嘌呤; Y, 嘧啶)。用较小点表示并非所有 tRNA 中都存在的可选核苷酸。标准编号体系将 5' 端定为 1 位, 3' 端定为 76 位; 其中包括部分而非全部的可选核苷酸。恒定和半恒定核苷酸位于第 8、11、14、15、18、19、21、24、32、33、37、48、53、54、55、56、57、58、60、61、74、75 和 76 位。反密码子核苷酸位于第 34、35 和 36 位





## 14.1 基因组活性的瞬时变化

## 14.2 基因组活性的永久性和半永久性变化

## 14.3 发育过程中基因组活性的调节

## 学习要点

当你阅读完第14章之后，应该能够：

- 区别分化与发育，概括在这两种过程中基因组的表达是如何调控的。
- 举例描述胞外信号化合物，如乳铁蛋白和类固醇类激素造成基因组活性瞬时变化的不同途径。
- 详细描述细菌的代谢抑制。
- 讨论信号从细胞表面受体传递到基因组的不同途径。
- 举例描述造成基因组活性永久性和半永久性变化的不同途径，明确涉及基因组重排、染色体结构变化与反馈环过程之间的区别。
- 讨论λ噬菌体的溶原性感染周期和枯草杆菌芽孢 (*Bacillus subtilis*) 形成方面的研究给我们提供了哪些分化和发育相关的基本信息。
- 解释为什么秀丽隐杆线虫 (*Caenorhabditis elegans*) 是一种有用的模式生物，描述在线虫的阴门发育的过程中细胞命运是如何确定的。
- 描述黑腹果蝇 (*Drosophila melanogaster*) 胚胎形成过程中的遗传学事件。
- 讨论黑腹果蝇、脊椎动物和植物中同源异形基因的功能。



我们已经阐述了基因组的表达决定蛋白质组内容的途径，反过来蛋白质组又决定了细胞的生化特征。没有一种生物的生化特征是完全恒定的。即使最简单的单细胞生物在环境变化时也能改变它们的蛋白质组，从而使它们的生化能力可以与其可获得的营养物质以及主要物理化学条件不断地保持协调。多细胞生物的细胞对胞外环境的变化也同样做出应答，二者唯一的区别在于多细胞生物的主要刺激物包括激素、生长因子和营养。由此引发的基因组活性的瞬时变化能够使蛋白质组不断地被重塑，以满足外部环境对细胞的要求（图 14.1）。基因组活性的其他变化是永久性的，或至少半永久性的，从而不可逆地改变细胞的生化特征。这些变化导致细胞的分化（differentiation），即细胞具有特定的生理功能。许多单细胞生物能发生分化，一个例子就是细菌，如杆菌（*Bacillus*）的芽孢生成细胞的产生，但我们更多地是将分化与多细胞生物联系起来。多细胞生物是由不同类型的特化细胞（人类有 250 多种）来组成组织和器官的，这些复杂的多细胞结构和生物体整体的装配需要不同细胞的基因组活性的协调。这种协调包括瞬时和永久的变化，而且在生物体的发育（development）过程中必须持续很长时间。

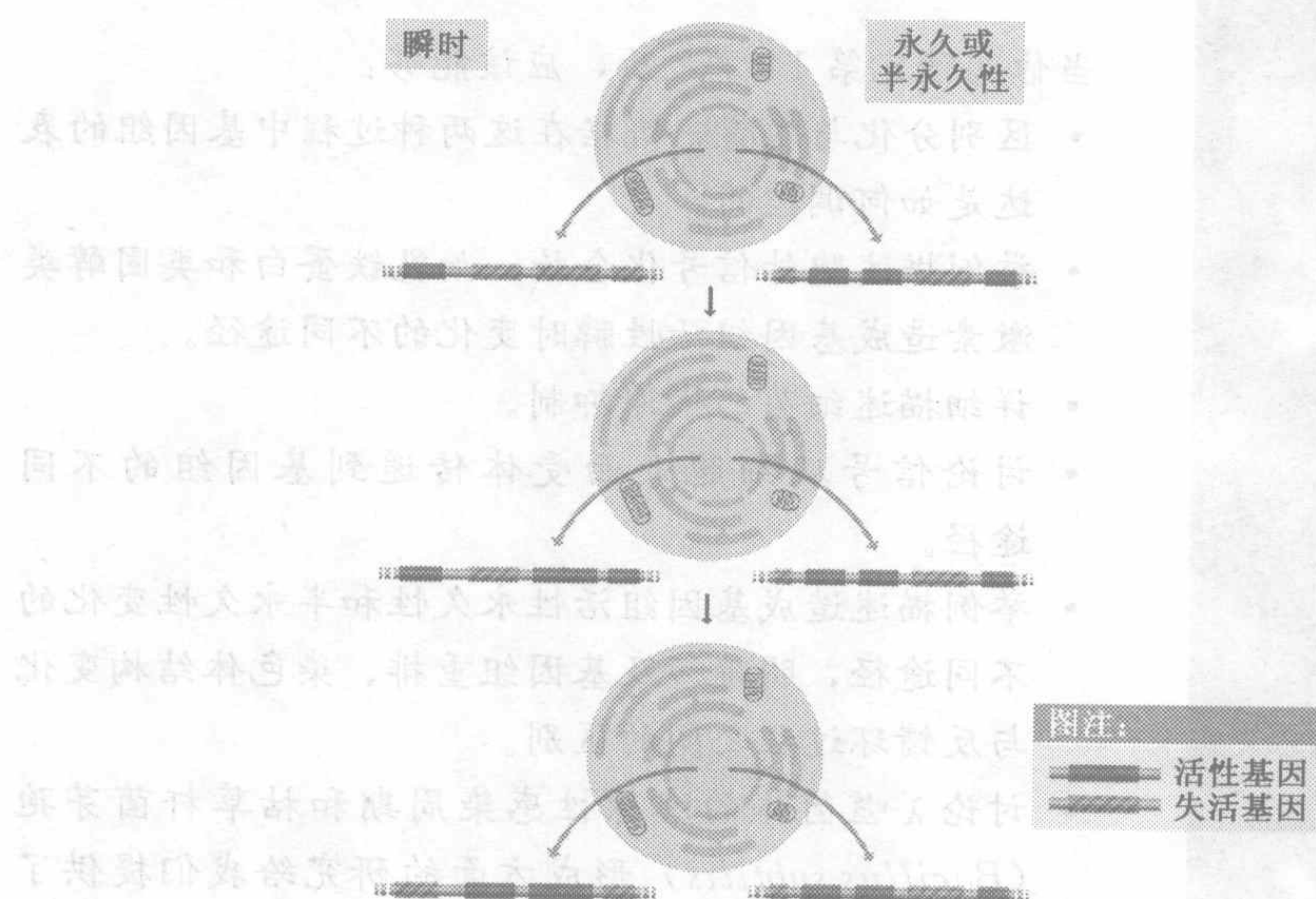


图 14.1 基因组活性的两种调节方式

左边的一套基因受瞬时调控，由胞外环境的变化决定是否表达。右边的基因发生了表达谱的永久或半永久性的变化，使同样的三个基因持续表达

在单个基因的表达途径中，有很多步骤可以进行调节（表 14.1），在第 10~13 章中列举了不同机制生物学功能的例子。这一章的目的不是重述这些基因特异的调控系统，而是解释基因组作为一个整体其活性是如何调节的。这样做时，我们必须记住生物界的多样性和个体基因组中基因数量的庞大性，因此，我们有理由认为任何可能进化形成的调控基因表达的机制都有可能已经发挥了作用。因此，我们可以在基因表达途径中的每一个环节举出调节的例子是很自然的。但是所有这些调控点在对基因组整体的活性进行调节时有同等的重要性吗？目前的观点是并非如此。我们的理解只是基于对几个生

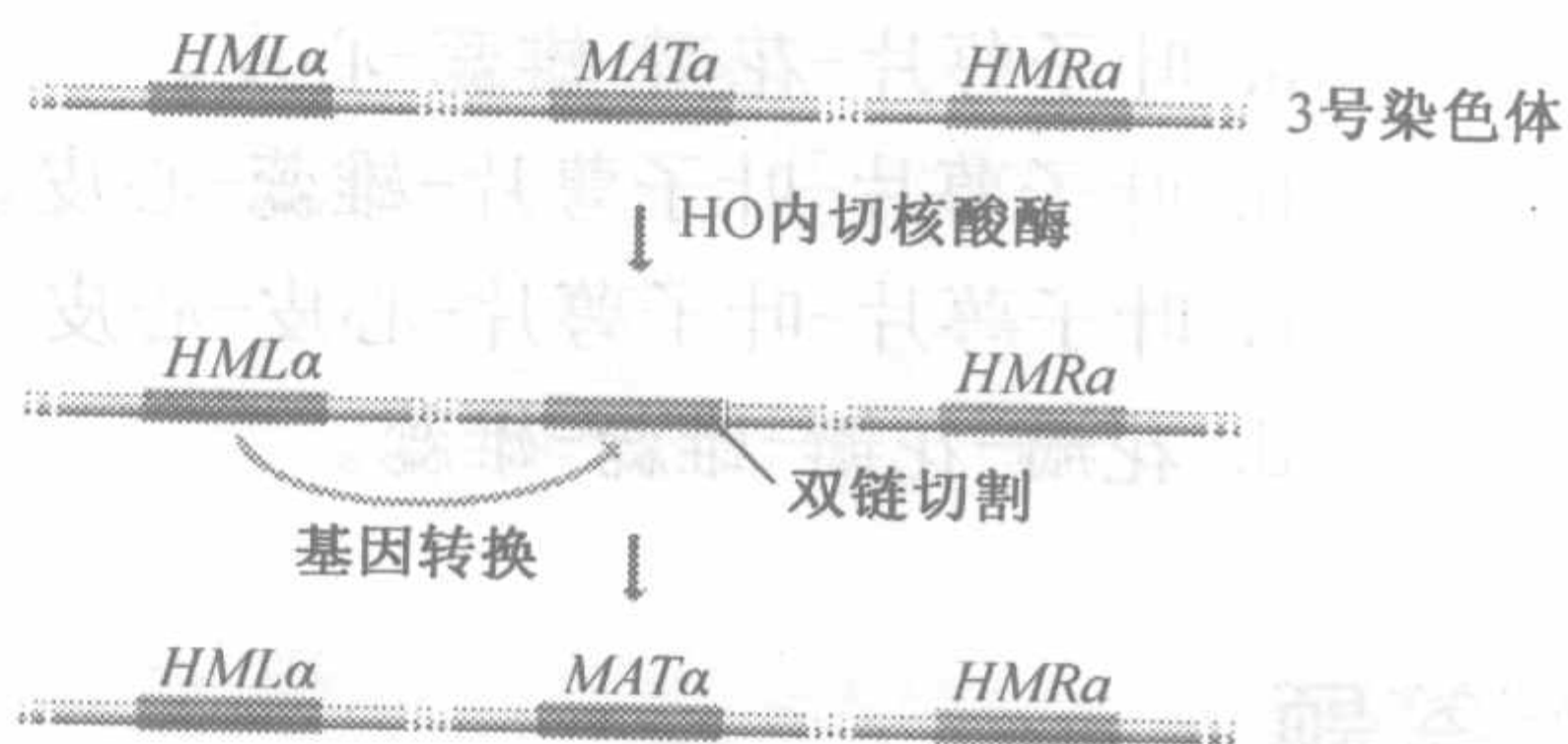
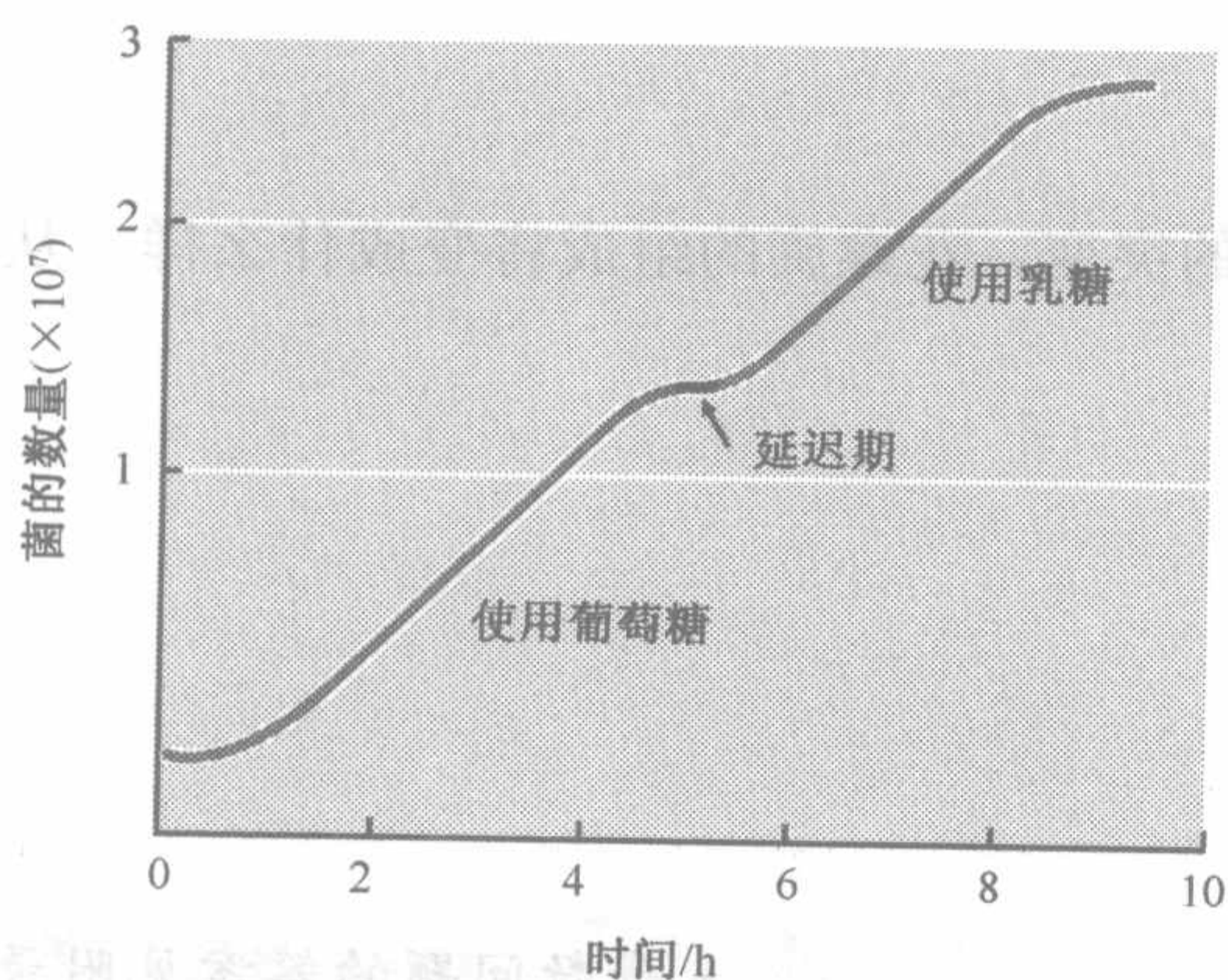


- 14.3\* 秀丽隐杆线虫和黑腹果蝇是高等真核生物的发育研究的理想模式生物吗？  
 14.4 用模式生物研究高等真核生物的发育有什么价值吗？  
 14.5\* 高等真核生物的发育理想模式生物的关键特点是什么？

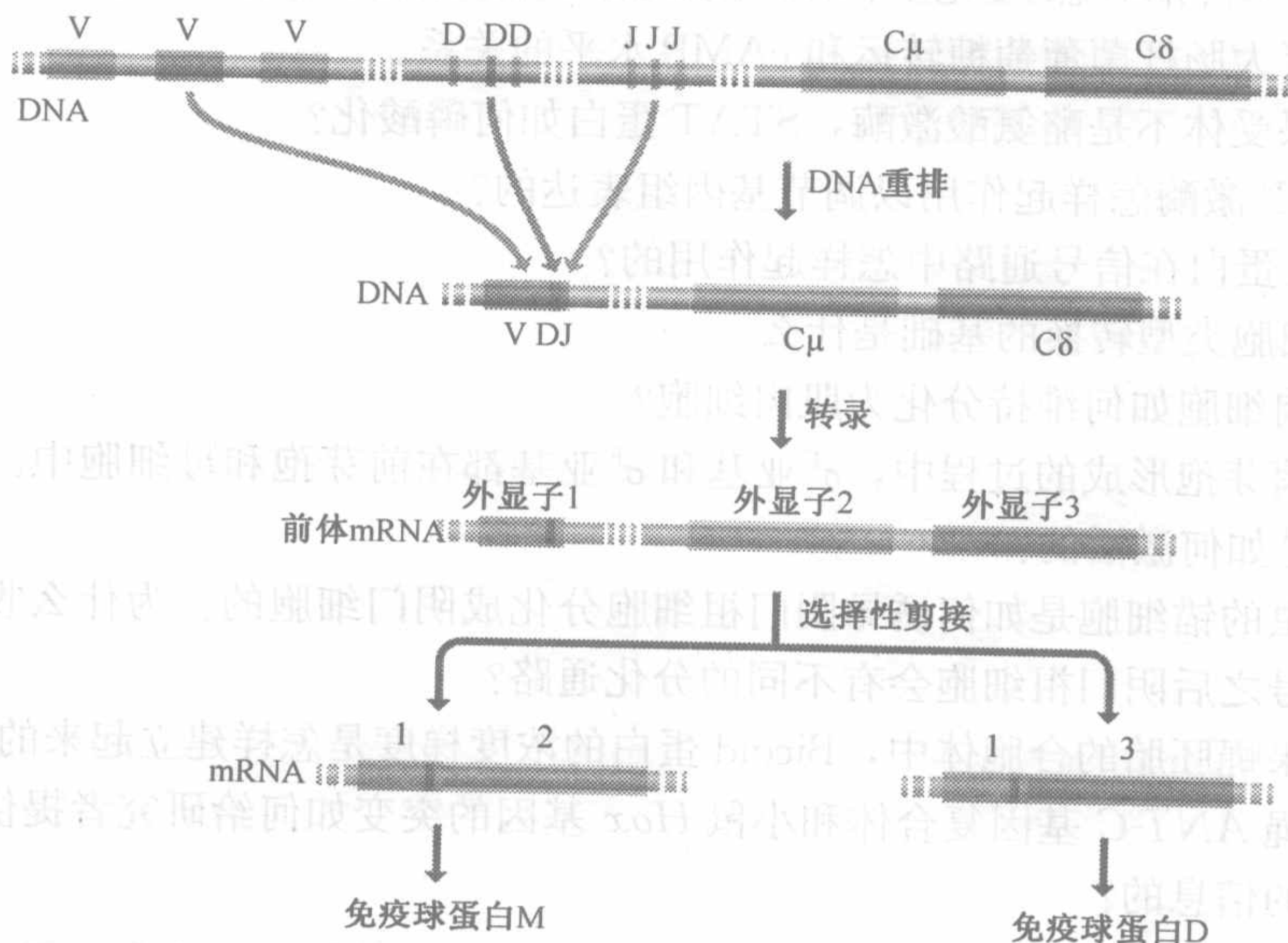
## 图形测试

奇数问题的答案见附录

- 14.1\* 如下图（左）显示的是当大肠杆菌生长在含有两种糖的培养基里，细胞在利用乳糖之前先利用葡萄糖，用什么术语描述这种现象？解释这个过程的机制。  
 14.2 如下图（右）描绘了 *MAT* 基因座从 *MAT $\alpha$*  基因型转变为 *MAT $\alpha$*  基因型，用什么术语描述这个转变？这个转变的机制是什么？



- 14.3\* 讨论 B 细胞产生免疫球蛋白多样性的过程。



- 14.4 如图显示果蝇胚胎发育过程中蛋白质的分布。这些蛋白质有哪些类型的基因编码？哪些蛋白质调节这些基因的表达？



头 : 胸 : 腹

Huckebein  
Tailless  
Orthodenticle  
Empty spiracles  
Buttonhead  
Giant  
Krüppel  
Knirps

## 拓展阅读

### 细胞外信号物质进入细胞

**He, J. and Furmanski, P.** (1995) Sequence specificity and transcriptional activation in the binding of lactoferrin to DNA. *Nature* **373**: 721–724.

**Tsai, M.-J. and O'Malley, B.W.** (1994) Molecular mechanisms of action of steroid/thyroid receptor superfamily members. *Annu. Rev. Biochem.* **63**: 451–486.

**Winge, D.R., Jensen, L.T. and Srinivasan, C.** (1998) Metal ion regulation of gene expression in yeast. *Curr. Opin. Chem. Biol.* **2**: 216–221.

### 通过细胞表面受体的信号转导

**Horvath, C.M.** (2000) STAT proteins and transcriptional responses to extracellular signals. *Trends Biochem. Sci.* **25**: 496–502.

**Karin, M. and Hunter, T.** (1995) Transcriptional control by protein phosphorylation: signal transmission from the cell surface to the nucleus. *Curr. Biol.* **5**: 747–757.

**Maruta, H. and Burgess, A.W.** (1994) Regulation of the Ras signaling network. *Bioessays* **16**: 489–496.

**Robinson, M.J. and Cobb, M.H.** (1997) Mitogen-activated kinase pathways. *Curr. Opin. Cell Biol.* **9**: 180–186.

**Schlessinger, J.** (1993) How receptor tyrosine kinases activate Ras. *Trends Biochem. Sci.* **18**: 273–275.

**Spiegel, S., Foster, D. and Kolesnick, R.** (1996) Signal transduction through lipid second messengers. *Curr. Opin. Cell Biol.* **8**: 159–167.

**Whitman, M.** (1998) Feedback from inhibitory SMADs. *Nature* **389**: 549–551.

### 基因组重排

**Alt, F.W., Blackwell, T.K. and Yancopoulos, G.D.** (1987) Development of the primary antibody repertoire. *Science* **238**: 1079–1087. *Generation of immunoglobulin diversity.*

**Nasmyth, K. and Shore, D.** (1987) Transcriptional regulation in the yeast life cycle. *Science* **237**: 1162–1170. *Yeast mating-type switching.*

### 多梳蛋白

**Chan, C.S., Rastelli, L. and Pirrotta, V.** (1994) A Polycomb

response element in the Ubx gene that determines an epigenetically inherited state of repression. *EMBO J.* **13**: 2553–2564.

**Déjardin, J., Rappailles, A., Cuvier, O., Grimaud, C., Decoville, M., Locker, D. and Cavalli, G.** (2005) Recruitment of *Drosophila* Polycomb group proteins to chromatin by DSP1. *Nature* **434**: 533–538.

### 反馈环

**Popperl, H., Bienz, M., Studer, M., Chan, S.K., Aparicio, S., Brenner, S., Mann, R.S. and Krumlauf, R.** (1995) Segmental expression of HoxB-1 is controlled by a highly conserved autoregulatory loop dependent upon exd/pbx. *Cell* **81**: 1031–1042.

**Regulski, M., Dessain, S., McGinnis, N. and McGinnis, W.** (1991) High affinity binding sites for the Deformed protein are required for the function of an autoregulatory enhancer of the *deformed* gene. *Genes Devel.* **5**: 278–286.

### Sporulation in *B. subtilis*

**Errington, J.** (1996) Determination of cell fate in *Bacillus subtilis*. *Trends Genet.* **12**: 31–34.

**Sonenshein, A.L.** (2000) Control of sporulation initiation in *Bacillus subtilis*. *Curr. Opin. Microbiol.* **3**: 561–566.

**Stragier, P. and Losick, R.** (1996) Molecular genetics of sporulation in *Bacillus subtilis*. *Annu. Rev. Genet.* **30**: 297–341.

### 线虫阴门发育

**Aroian, R.V., Koga, M., Mendel, J.E., Ohshima, Y. and Sternberg, P.W.** (1990) The *let-23* gene necessary for *Caenorhabditis elegans* vulval induction encodes a tyrosine kinase of the EGF receptor subfamily. *Nature* **348**: 693–699.

**Katz, W.S., Hill, R.J., Clandinin, T.R. and Sternberg, P.W.** (1995) Different levels of the *C. elegans* growth factor LIN-3 promote distinct vulval precursor fates. *Cell* **82**: 297–307.

**Kornfeld, K.** (1997) Vulval development in *Caenorhabditis elegans*. *Trends Genet.* **13**: 55–61.

**Labouesse, M. and Mango, S.E.** (1999) Patterning the *C. elegans* embryo: moving beyond the cell lineage. *Trends Genet.* **15**: 307–313. *Reviews the developmental pathways of C. elegans.*

**Sharma-Kishore, R., White, J.G., Southgate, E. and Podbilewicz, B.** (1999) Formation of the vulva in



- d. Toponemic
- 15.3\* 哪位科学家最先提出“断裂-再连接”模型用以解决 DNA 复制的拓扑学问题?
- a. Dibrück  
b. Kornberg  
c. Meselson 和 Stahl  
d. Watson 和 Crick
- 15.4 下面哪项是大肠杆菌 DNA 回旋酶的功能?
- a. 消除 DNA 复制过程中产生的基因组过度螺旋。  
b. 消除转录过程中产生的基因组过度螺旋。  
c. 向 DNA 分子中介导超螺旋。  
d. 以上所有。
- 15.5\* 哪一类型的 DNA 分子使用滚环复制方式进行复制?
- a. 细菌基因组  
b. 部分噬菌体基因组 (如  $\lambda$  噬菌体)  
c. 线粒体基因组  
d. 酵母基因组
- 15.6 DNA 复制起始的位点被称之为:
- a. 增强子  
b. 起始子  
c. 复制起始点  
d. 启动子
- 15.7\* DNA 合成中引物的作用是什么?
- a. 为下一个核苷酸的加入提供 5'-磷酸基团。  
b. 为 DNA 合成提供一个可以水解放能的 5'-磷酸基团。  
c. 为下一个核苷酸的加入提供一个 3'-羟基。  
d. 为 DNA 链合成提供核苷酸。
- 15.8 在 DNA 合成中下面哪种 DNA 聚合酶的活性被用来校对?
- a. 3'→5' 外切核酸酶  
b. 5'→3' 外切核酸酶  
c. 单链内切核酸酶  
d. 双链内切核酸酶
- 15.9\* 什么是冈崎片段?
- a. DNA 前导链上合成的短片段多聚核苷酸。  
b. DNA 滞后链上合成的短片段多聚核苷酸。  
c. 滞后链上合成用于 DNA 合成的引物。  
d. DNA 聚合酶的蛋白水解片段。
- 15.10 哪个蛋白质可阻止复制叉处单链 DNA 分子的降解或重结合?
- a. 解旋酶  
b. 引导酶



- c. 单链结合蛋白
- d. 拓扑异构酶
- 15.11\* 细菌中, 下列哪个酶可以去除在滞后链每条冈崎片段起始处的 RNA 引物?
- a. DNA 聚合酶 I
- b. DNA 聚合酶 III
- c. DNA 连接酶
- d. RNase H
- 15.12 大肠杆菌中, 哪个蛋白质在 DNA 复制结束后结合在一起的姐妹染色体的分离中发挥作用?
- a. DnaB
- b. DNA 聚合酶
- c. 拓扑异构酶 IV
- d. Tus
- 15.13\* 下面关于端粒酶的判断哪条是正确的?
- a. 端粒酶是一种 RNA 依赖的 DNA 聚合酶。
- b. 端粒酶是一种 RNA 依赖的 RNA 聚合酶。
- c. 端粒酶是一种 DNA 依赖的 DNA 聚合酶。
- d. 端粒酶是一种 DNA 依赖的 RNA 聚合酶。
- 15.14 果蝇中的端粒类似下列哪种序列?
- a. 着丝粒
- b. 微卫星
- c. 逆转座子
- d. DNA 转座子
- 15.15\* 在细胞周期的哪个期 DNA 复制进行?
- a. M
- b. G<sub>1</sub>
- c. S
- d. G<sub>2</sub>
- 15.16 哪项技术被用于研究基因组不同区域复制起始时间?
- a. 荧光原位杂交
- b. 质谱
- c. 微阵列分析
- d. PCR

## 简答题

- 15.1\* 在 Meselson-Stahl 实验前, 我们不知道 DNA 复制是“弥散性”、“半保留”或“全保留”的。描述经这几种不同方式复制后子代分子中 DNA 组分的区别。
- 15.2 描述 DNA 分子替代复制的机制。



- d. 丢失了磷酸二酯键
- 16.12 下面哪条描述的是核苷酸切除修复?
- 含有损伤核苷酸的双链 DNA 区域被去除,同时由新 DNA 替换。
  - 单一受损的核苷酸被切除并由新的核苷酸取代。
  - 单一受损的碱基被切除并由新的核苷酸取代。
  - 含有损伤核苷酸的单链 DNA 区域被去除,同时由新 DNA 替换。
- 16.13\* 下面哪项是错配修复的特点?
- 修饰后的核苷酸被识别。
  - 环丁二聚体被切除。
  - 亲代链和新合成的子代链被区分开。
  - 正确的读码框被鉴定。
- 16.14 大肠杆菌中亲代链和新合成的子代链之间是如何区分的?
- 子代链一合成好即被甲基化。
  - 子代链不被迅速甲基化。
  - 子代链不能迅速和核小体蛋白结合。
  - 子代链包含用于起始 DNA 合成的 RNA 引物中的核糖核酸。
- 16.15\* 大肠杆菌细胞如何在 SOS 应答中复制受损 DNA?
- DNA 的损伤区域从基因组中去除。
  - 在损伤位点核苷酸随机插入。
  - 在损伤修复前所有的 DNA 合成停止。
  - mRNA 通过重组插入到 DNA 的受损区域。

## 简答题

\* 奇数问题的答案见附录

- 基因组中突变产生的机制是什么?
- 复制前 DNA 修复和复制后 DNA 修复系统有什么区别?
- DNA 聚合酶在 DNA 合成过程中如何选择正确的核苷酸?
- 碱基类似物 2-氨基嘌呤如何在 DNA 中制造突变?
- 热如何影响 DNA 的结构? 热诱导的 DNA 损伤如何普遍存在? 这些损伤的影响是什么?
- 非编码 DNA 序列中的突变如何影响基因组的表达?
- 人类中,家族性高胆固醇血症是由于一个编码脂蛋白 B-100 的 LDL 受体结合结构域的基因发生了功能丢失突变,并进行显性遗传,为什么?
- 如何利用条件致死突变来鉴定或描述微生物中重要的基因产物?
- 人类免疫球蛋白基因 V 区高频突变的基础是什么?
- 碱基切除修复的步骤是什么?
- 非同源末端连接系统修复 DNA 双链断裂的过程如何?
- 大肠杆菌中 LexA 蛋白在 SOS 应答中的作用是什么?



## 论述题

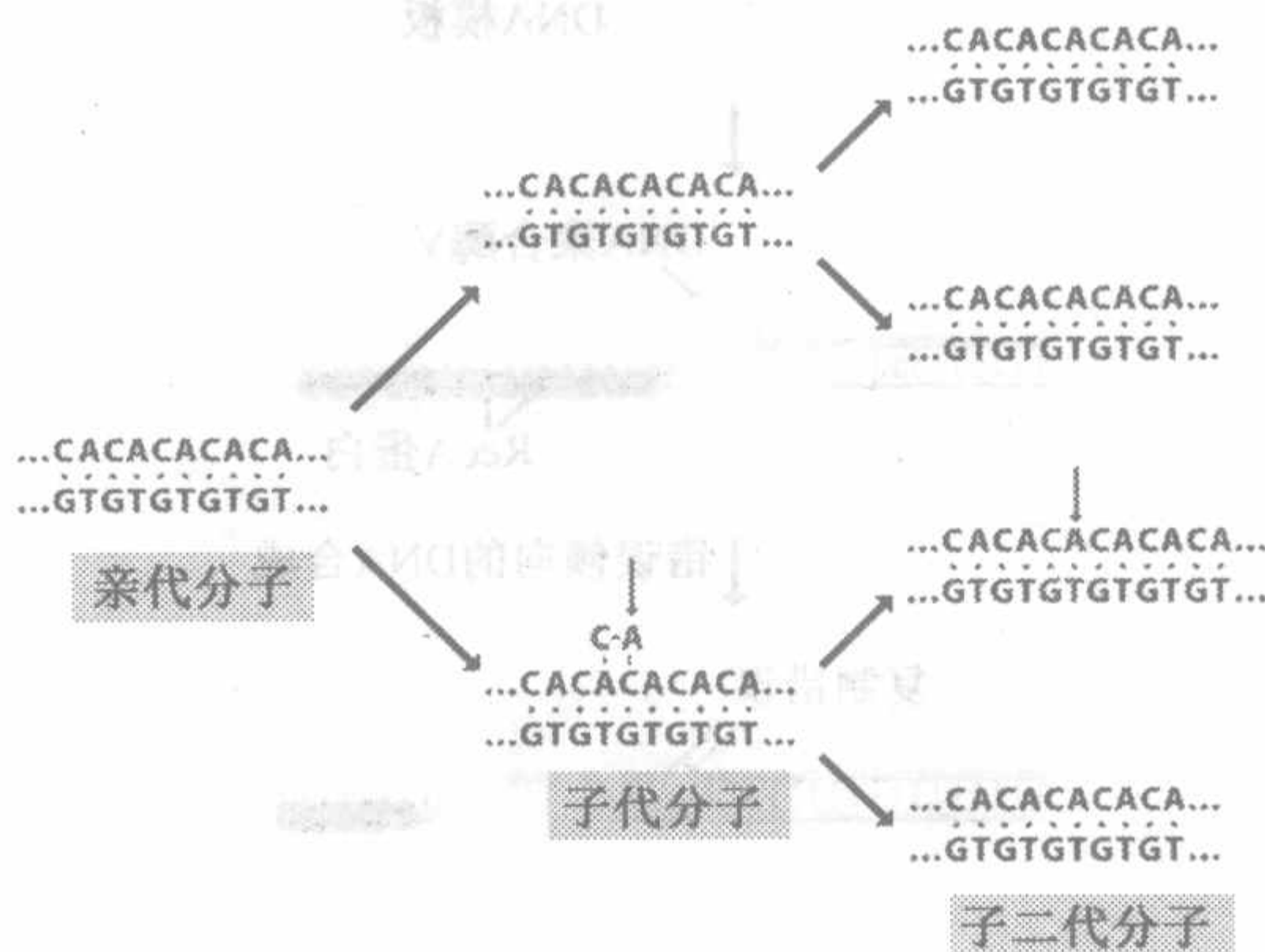
\* 奇数问题答案的指导见附录

- 16.1\* 解释为什么嘌呤到嘌呤或嘧啶到嘧啶的点突变叫做转换，而嘌呤到嘧啶或嘧啶到嘌呤的改变叫做颠换。
- 16.2 大量突变中转换和颠换的预计比例应该是多少？
- 16.3\* 了解更多关于三核苷酸扩增的疾病的知识，包括试图对三核苷酸扩增产生疾病机理进行解释的假说。
- 16.4 评价程序性突变的证据？
- 16.5\* 耐辐射奇球菌 (*Deinococcus radiodurans*) 对放射和其他物理和化学诱变剂有较高的抗性。讨论它的特殊特点如何在其基因组序列中体现出来。

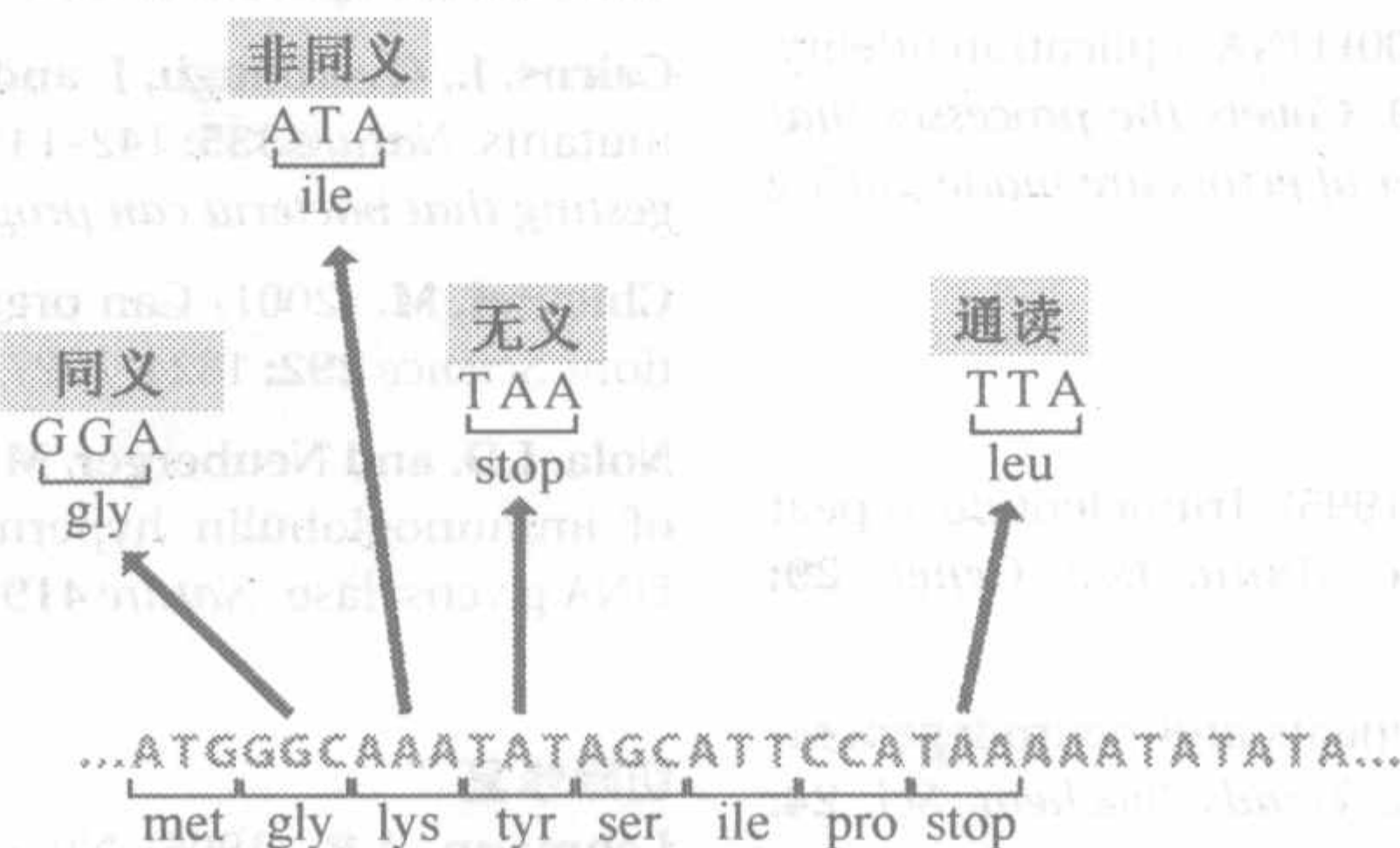
## 图形测试

\* 奇数问题的答案见附录

- 16.1\* 本图中是哪种突变（箭头所指处）？这种突变的原因是什么？



- 16.2 讨论本图中不同类型突变的潜在影响。



- 16.3\* 从 DNA 上将损伤碱基切除的酶是什么？碱基被切除后，哪条通路用来修复 DNA？



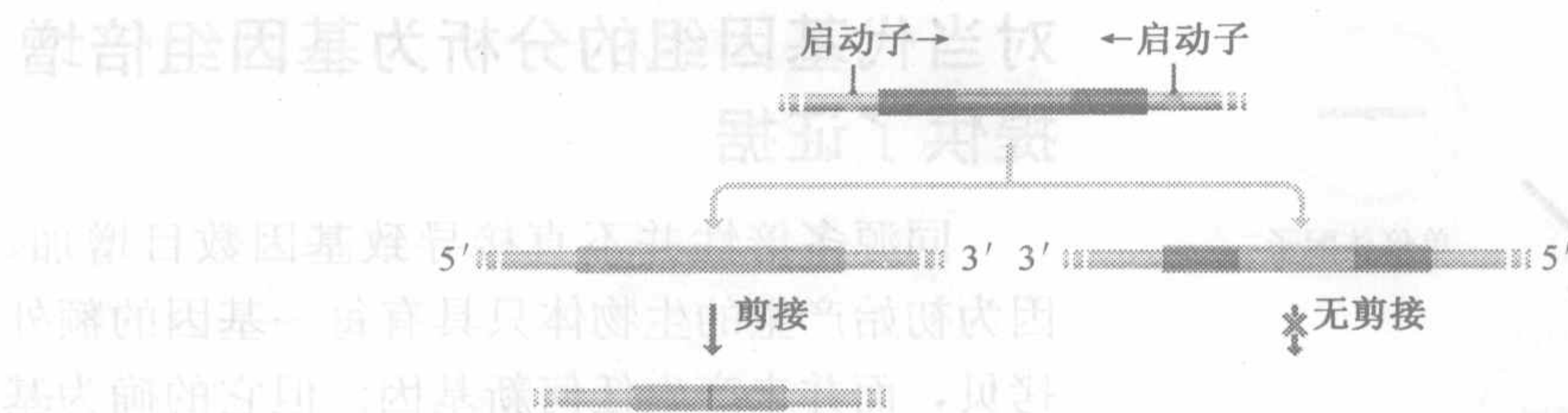


图 18.12 基因的反义 RNA 将包含基因的内含子序列

在图的左侧，基因以经典的方式被转录，产生前体 mRNA，前体 mRNA 的内含子随后被剪切去除。在图的右侧，基因的下游有另一个启动子，这个启动子起始的转录，会产生一条反义 RNA。因为反义 RNA 中的内含子区域不再含有内含子的剪接序列，所以不会被剪接体剪接，内含子会一直保留（12.2.2 节）

裂产生配子的过程中发生了错误，从而导致配子是二倍体而不是单倍体时，基因组倍增就有可能发生（图 18.13）。如果两个二倍体配子融合，将会产生同源多倍体（autopolyploid），此时四倍体细胞核中含有每条染色体的 4 个拷贝。同源多倍性像其他类型的多倍性一样（18.2.2 节），在植物中并不罕见。同源多倍体通常可以存活，因为每条染色体仍具有同源配对体，并能在减数分裂时形成二价体。这使得同源多倍体可以繁殖，但通常不能与其来源生物体进行变种间杂交。这是因为当二倍体和四倍体杂交时，产生的子代三倍体因为有一套染色体缺乏同源配对体而不能繁殖（图 18.14）。同源多倍性是物种发生的一种机制，一对物种通常定义为两种不能在变种间互交的生物体。确实有研究者观察到通过同源多倍性产生的新的植物物种，特别是 Mendel 实验的重新发现者之一 Hugo de Vries。在对晚樱草（*Oenothera lamarokicina*）的研究中，他发现了这种通常为二倍体的植物的四倍体变种，并命名为 *Oenothera gigas*。同源多倍性在动物中，特别是具有明显两性分化的动物中很少。这可能是因为细胞核中含有超过一对性染色体会引起种种问题。然而并非完全不可能，在一种哺乳动物阿根廷大鼠 red viscacha 中，确实存在着四倍体基因组。

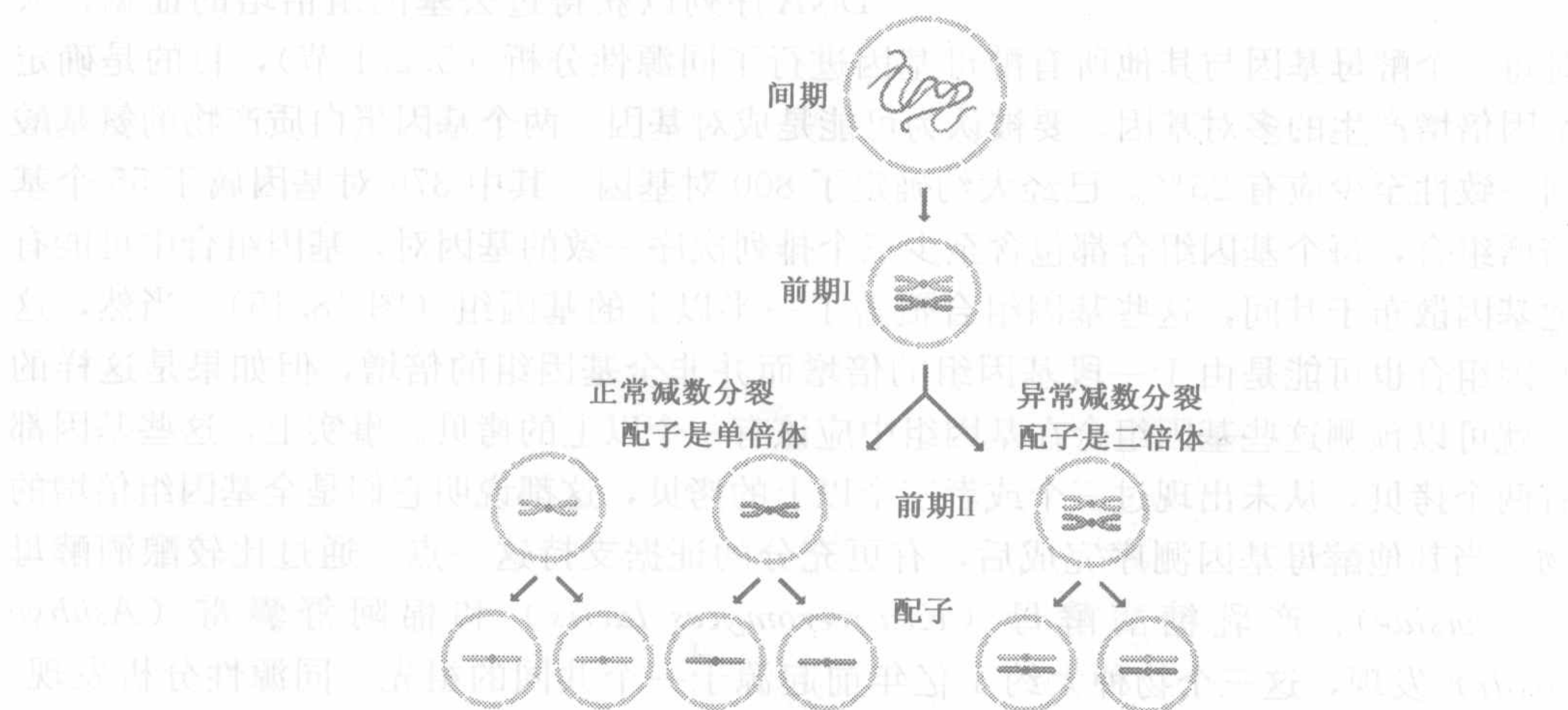


图 18.13 同源多倍体化的基础

在减数分裂中发生的正常事件如图左侧简易形式所示（与图 3.16 比较）。在图右侧前期 I 与前期 II 之间发生了异常，一对同源染色体没有分离至不同的细胞核中，产生的配子是二倍体而不是单倍体



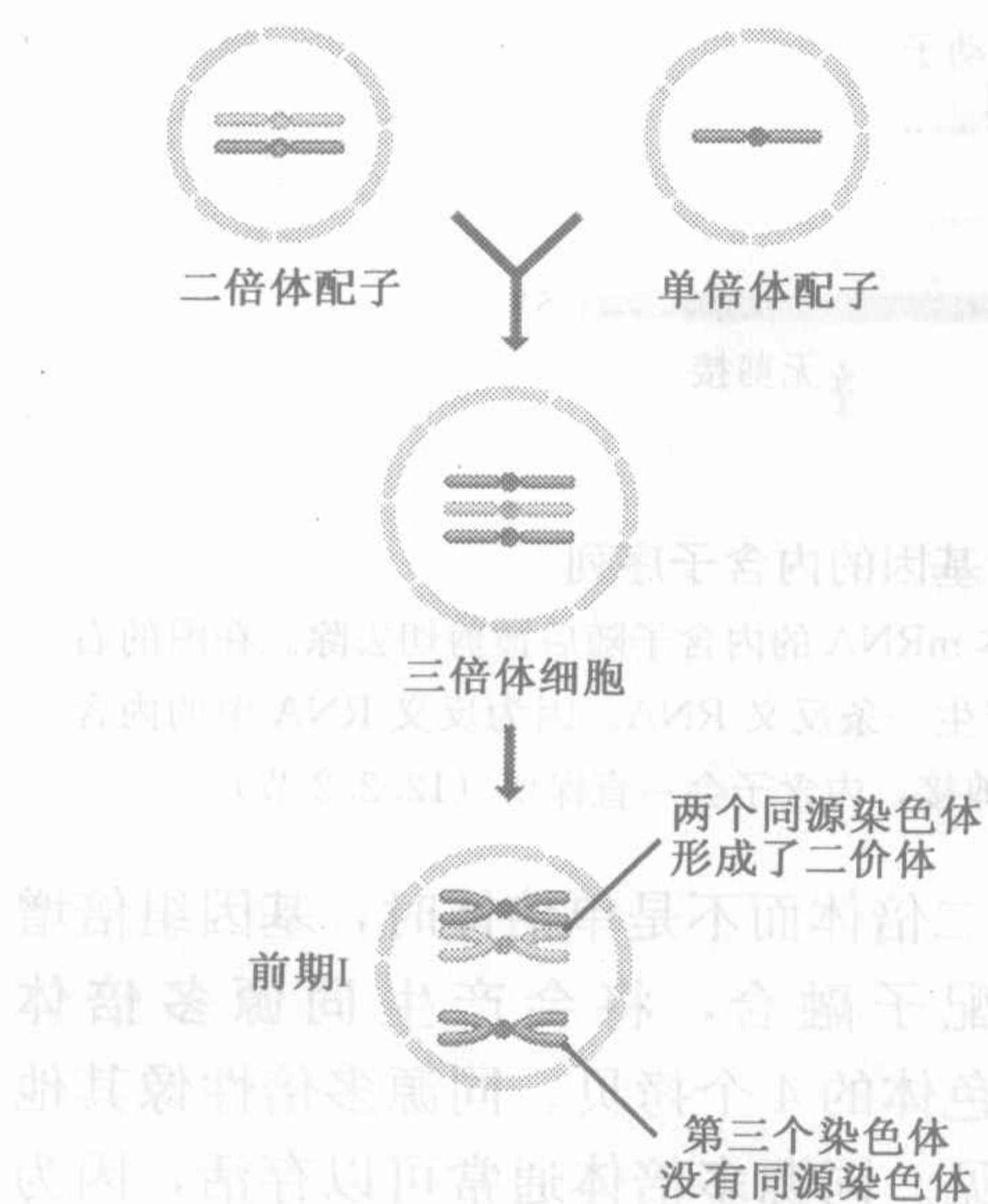


图 18.14 同源多倍体不能与其亲本间互交  
异常减数分裂 (图 18.13) 产生的二倍体配子与正常减数分裂产生的单倍体配子融合后, 产生三倍体细胞核。每条同源染色体具有三个拷贝。在下次减数分裂的前期 I 时, 两条同源染色体中可以形成二价体, 而第三条染色体则没有对应体。这会破坏后期染色体的分离 (图 3.16) 并阻止减数分裂圆满完成。这就意味着不能产生配子, 因而三倍体生物是不育的。注意: 在三条同源染色体中任何两条都可以形成二价体, 而不仅在图中所示的两个同源染色体之间

## 对当代基因组的分析为基因组倍增提供了证据

同源多倍性并不直接导致基因数目增加, 因为初始产生的生物体只具有每一基因的额外拷贝, 而并未产生任何新基因。但它的确为基因增加提供了可能性, 因为额外基因并不是细胞功能必需的, 它们可进行不损害生物体存活的突变。既然这样, 我们能不能从现存基因组中找到由于基因组倍增而获得大量基因的证据呢?

从我们前面对基因组改变方式的理解, 我们可以预见, 得到基因组片段倍增的证据将相当困难。可以想象来自基因组倍增的许多额外基因拷贝会退化, 在 DNA 序列中不再可见。那些因为倍增的功能对生物体有用或进化出新功能而保留下来的基因, 可以被鉴定出来。但不可能分辨出它们是来自全基因组倍增或仅是许多小片段的倍增。要标记出一次基因组倍增, 就必须发现大的多组基因倍增, 两个组合中的基因排布顺序相同。基因组中倍增基因组合仍然能被发现的可能性取决于将基因移动到新位置的重组事件发生的频率。

为查找酵母 (*Saccharomyces cerevisiae*) DNA 序列以获得过去基因组倍增的证据, 人们

对每一个酵母基因与其他所有酵母基因进行了同源性分析 (5.2.1 节), 目的是确定由基因倍增产生的多对基因。要被认为可能是成对基因, 两个基因蛋白质产物的氨基酸序列一致性至少应有 25%。已经大约确定了 800 对基因, 其中 376 对基因属于 55 个基因倍增组合, 每个基因组合都包含至少三个排列次序一致的基因对, 基因组合中可能有其他基因散布于其间, 这些基因组合覆盖了一半以上的基因组 (图 18.15)。当然, 这种基因组合也可能是由于一段基因组的倍增而并非全基因组的倍增, 但如果是这样的话, 就可以预测这些基因组合在基因组中应该有一个以上的拷贝。事实上, 这些基因都只有两个拷贝, 从未出现过三个或者三个以上的拷贝, 这都说明它们是全基因组倍增的产物。当其他酵母基因测序完成后, 有更充分的证据支持这一点。通过比较酿酒酵母 (*S. cerevisiae*)、产乳糖酶酵母 (*Kluyveromyces lactis*) 和棉阿舒囊霉 (*Ashbya gossipii*) 发现, 这三个物种大约 1 亿年前起源于一个共同的祖先, 同源性分析发现, 在这之后酿酒酵母基因组内发生了一次基因组倍增事件。因此, 应该存在着一种基因, 这种基因在酿酒酵母基因组中存在着两个拷贝, 而在产乳糖酶酵母和棉阿舒囊霉中只有一个拷贝。事实上也确实是这样, 这一新的分析发现, 大约有 10% 的酿酒酵母基因起



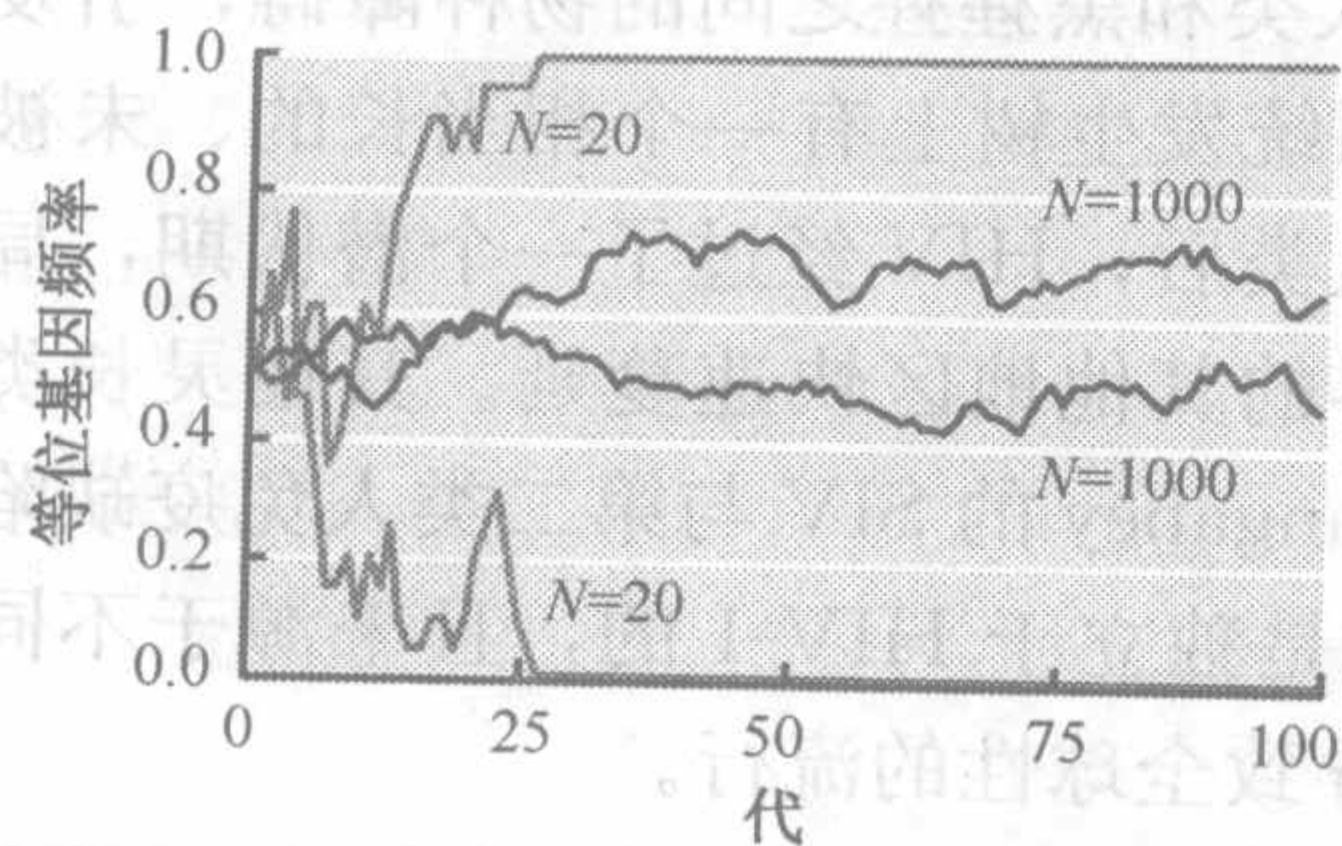


图 19.21 随机基因漂移对等位基因频率的影响

图中显示的是一次模拟实验的结果，实验模拟了一对等位基因在两个大小不同的群体中经历100代过程中的频率变化情况，一个群体包含20个个体，另一个包含100个个体。在两个群体中，一个等位基因的频率最终恒定，而另一个等位基因则最终消失，但是消失的时间和群体的大小有很大的关系。摘自 M. Jobling et al. *Human Evolutionary Genetics* 19.21)。

由于自然选择或者随机基因漂移的作用，一个等位基因开始在群体中占据主导地位，并最终达到100%的频率。自此以后，此等位基因就恒定不变(fixed)了。数学模型指出，经过一段时间后，不同的等位基因开始在一个群体中恒定，产生一系列的同一个基因取代物。如果一个物种分裂为两个并不广泛杂交的群体，那么两个群体中等位基因的频率会经历不同的变化，经过几十代以后这两个群体就会产生显示不同的遗传特征。最终，在不同的群体中会产生不同的基因取代(gene substitution)，但在这之前我们也可以根据等位基因的频率来区分不同的群体。这些不同可以用来确定种群分裂的时间，也可以用来确定两个群体或者其中的一个群体是否经历了一次生物瓶颈(bottleneck)，即在某一个时期内，群体的大小显著地减少了。

## 现代人的起源——是否出自非洲

我们有充分的理由相信，人类起源于非洲，因为所有最古老的前人类化石都是在那里发现的。化石学证据揭示，人科的成员首次迁移出非洲是在1百万年前，但这些不是现代人类，它们被称为直立人(*Homo erectus*，图19.22)的早期物种。这是人类首次地理扩散，最终扩散到旧大陆的所有地方。

对直立人类扩散后发生的事件尚存在争议。通过比较骨骼化石，古生物学家得出结论，定居于旧大陆不同地区的直立人产生了那些地方的现代人类，该过程称多区域进化

同个体相互关系的信息。单个生物体的生殖细胞发生突变会在群体中产生新的等位基因和单体群。每一种等位基因都有自己的等位基因频率(allele frequency)，随着时间在自然选择(natural selection)和随机基因漂移(random genetic drift)的作用下发生改变。自然选择就是通过不同等位基因对生物的适合度(fitness)的影响而起作用的，正如达尔文所说，自然选择会保留那些有利的突变，拒绝有害的突变。自然选择会减少降低生物体适合度的等位基因频率，增加增强生物体适合度的等位基因频率。在现实中，只有为数不多新的等位基因能够对生物体的适合度产生重大的影响，所以大多数新的等位基因不受自然选择的影响，但它们的发生频率仍受随机基因漂移的影响而发生改变，这些影响包括出生、死亡以及繁殖过程中的偶然性(图

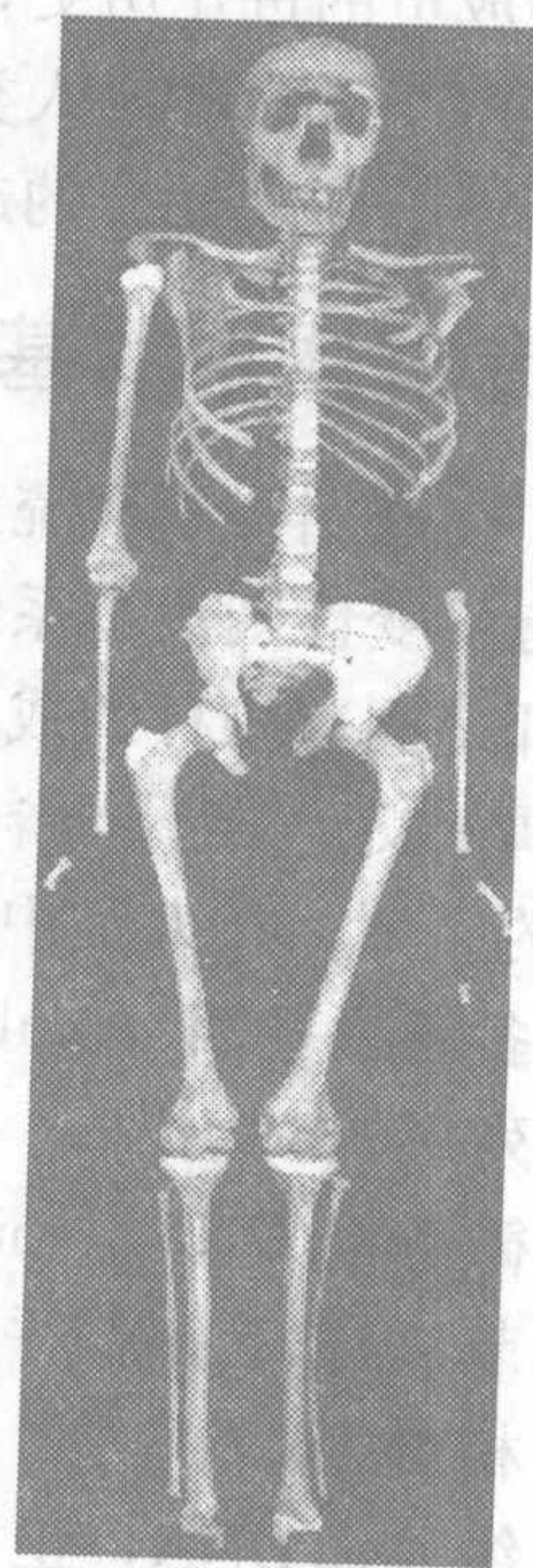


图 19.22 一副成年男性直立人的骨骼  
此标本又被称为“纳里奥科托姆男孩”，大约有160万年，在肯尼亚的特纳湖附近发现



(multiregional evolution) [图 19.23 (A)]。不同地理区域的人群可能有一定数量的通婚，但总的来说，不同人群在他们的进化史中保持独立。

对于多区域假设的疑问最初来自于对化石的重新解释，随之将争论推向高潮的是一篇文章，该文章发表于 1987 年，构建了基于世界不同地区 147 个有代表性人群的线粒体 DNA 的 RFLP 分析的系统发生树。该系统发生树（图 19.24）肯定了现代人的祖先居住在非洲，但推测 20 万年前他们还在那里。这个结论是将线粒体分子钟用于系统发生树而得出的，它显示，所有现代线粒体 DNA 的祖先 DNA 存在于 14 万~29 万年。系统发生树显示这个线粒体基因组定位于非洲，则具有这些基因组的人，即那位被称为线粒体夏娃 (mitochondrial Eve) 的女性（线粒体 DNA 仅仅是通过母系遗传），一定是非洲人。

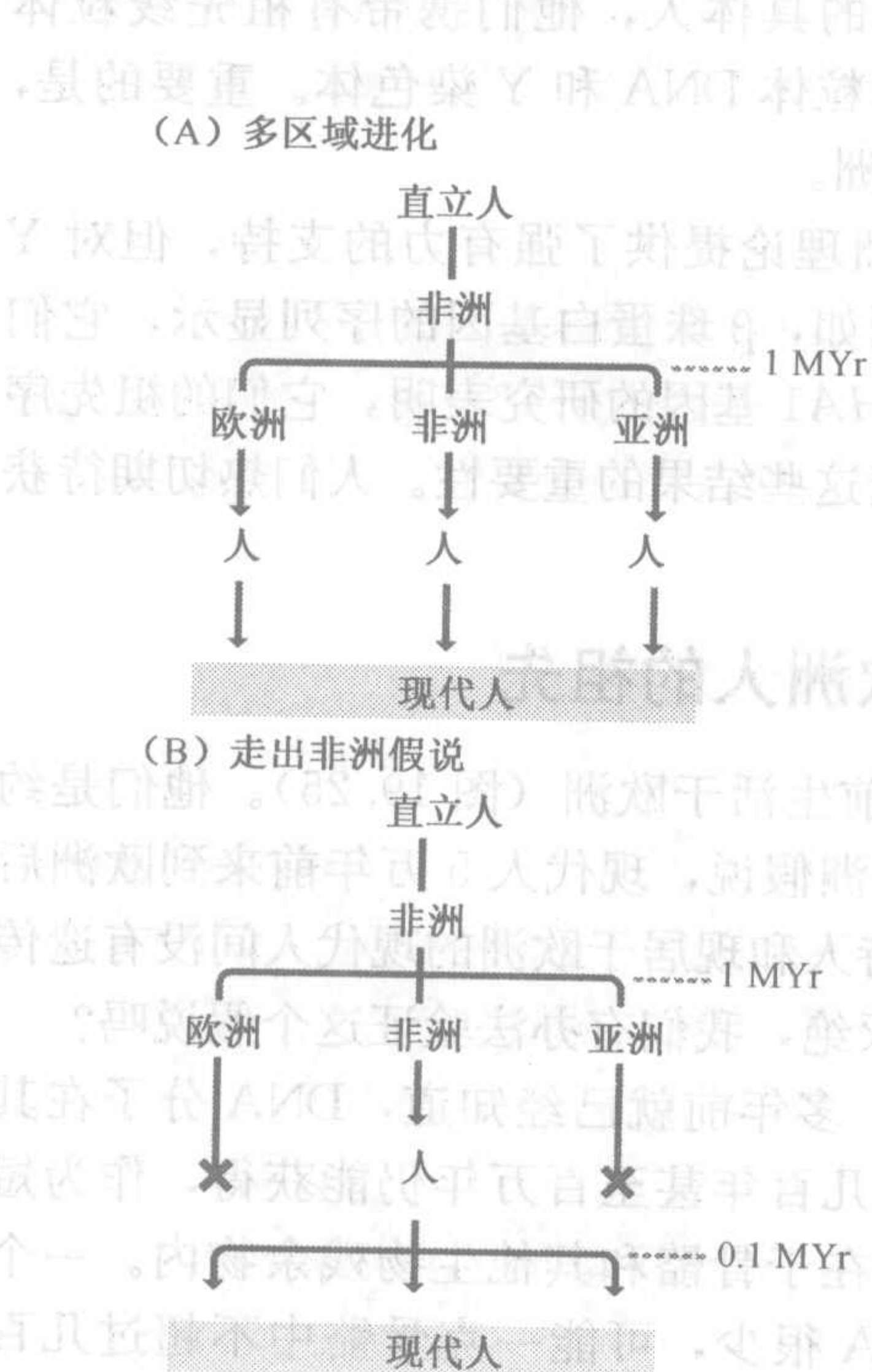


图 19.23 有关现代人类起源的两种对立假设

(A) 多区域假说认为，近代人类在 100 万年前离开非洲，在世界的不同地方进化成现代人。(B) 走出非洲假说认为，旧大陆的直立人群体是被随后迁出非洲大陆的现代人取代

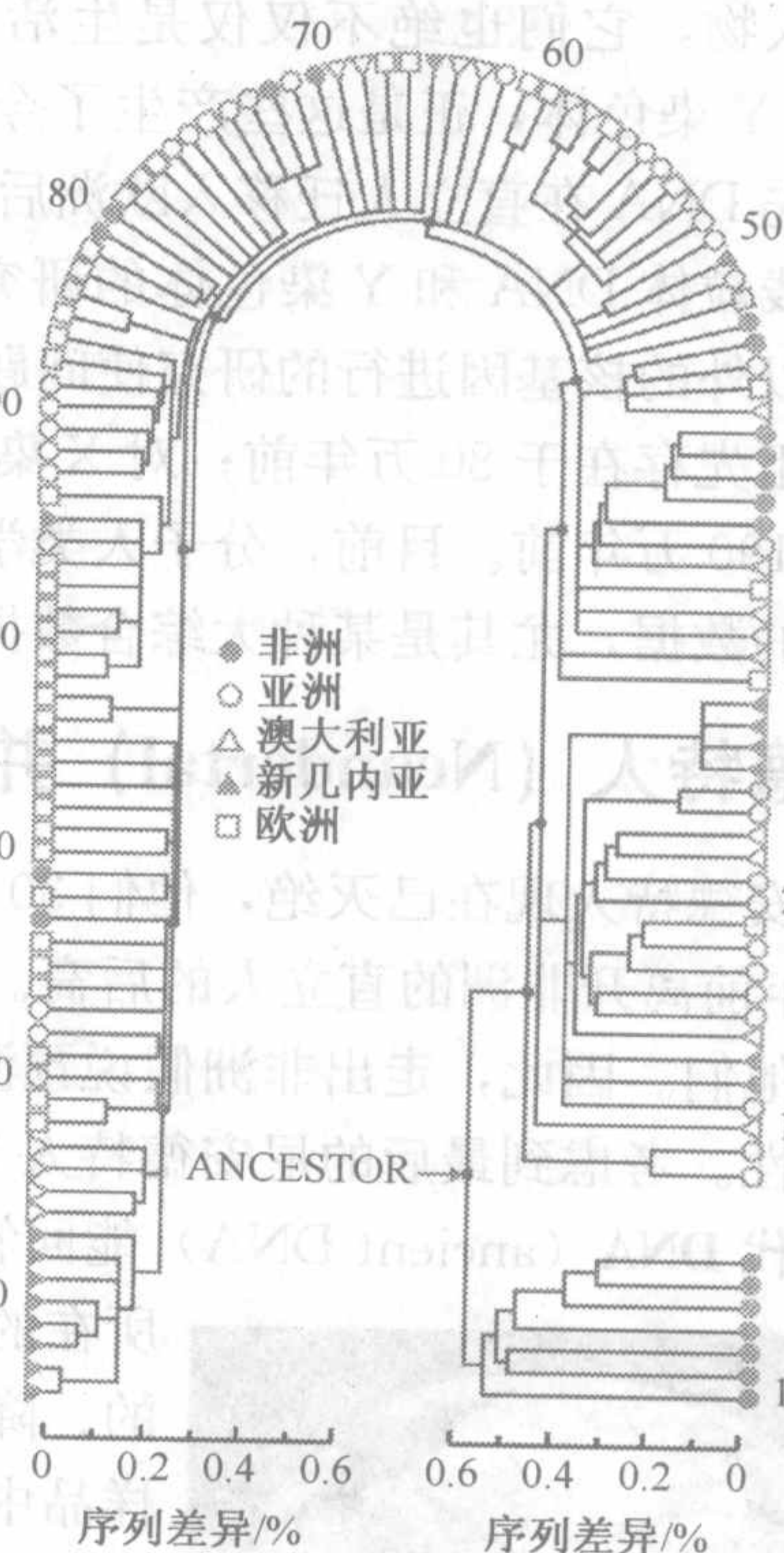


图 19.24 用 147 个现代人的线粒体 RFLP 数据构建的系统发生树

祖先线粒体 DNA 具推测存在于非洲，因为第一个分支位于 7 个现代非洲人的线粒体基因组与其他基因组之间，由于这个分支完全由非洲人组成，因此认为祖先也是非洲人。底部的标尺显示的是线粒体 DNA 的序列差异，它可以用来确定树上分支点的年代。根据分子钟推测，祖先序列存在于 14 万~29 万年。

(Cann et al., 1997)



Adenine	腺嘌呤	DNA 和 RNA 中发现的一种嘌呤碱基。
Adenosine deaminase acting on RNA(ADAR)	作用于 RNA 的腺嘌呤脱氨酶	一种能够对多种真核生物 mRNA 进行编辑的酶,它可以使腺嘌呤脱氨变成次黄嘌呤。
Adenylate cyclase	腺苷酸环化酶	将 ATP 转化为环 AMP 的酶。
A-DNA	A 型 DNA	DNA 双螺旋的一种构型,在细胞中存在但不常见。
Affinity chromatography	亲和层析	一种使用配体结合待纯化分子的柱层析方法。
Agarose gel electrophoresis	琼脂糖凝胶电泳	在琼脂糖凝胶中进行的电泳,用于分离长度为 100bp~50kb 的 DNA 分子。
Alarmone	信号素	是一种应急反应的激活因子,包括 ppGppp 和 pppGpp。
Alkaline phosphatase	碱性磷酸酶	能将磷酸基团从 DNA 分子的 5' 端去除的一种酶。
Alkylating agent	烷化剂	一类可以在核酸碱基上添加烷基的诱变剂。
Allele	等位基因	基因的两个或多个可替换型中的一种基因型。
Allele frequency	等位基因频率	一个等位基因在人群中出现的频率。
Allele-specific oligonucleotide (ASO) hybridization	等位基因特异性寡核苷酸杂交	利用一条寡聚核苷酸探针确定两条可变核苷酸序列哪条存在于某一 DNA 分子中。
Allopolyploid	异源多倍体	来自于不同物种的两个配子融合产生的多倍体。
Alphoid DNA	类 $\alpha$ -DNA	位于人染色体中着丝粒区域的串联重复核苷酸序列。
Alternative polyadenylation	可变多聚腺苷酸化	一个 mRNA 两个或两个以上不同位点多聚腺苷酸化。
Alternative promoter	选择性启动子	作用在同一个基因上的两个或多个不同启动子之一。
Alternative splicing	可变剪接	通过连接不同的外显子而从一个 mRNA 前体产生出两种或者多种成熟 mRNA 的现象。
Alu	Alu 序列	在人类及相关哺乳动物基因组中发现的一种短重复元件(SINE)。
Alu-PCR	Alu-PCR	运用 PCR 检测克隆的 DNA 片段中 Alu 序列相对位置的一种克隆指纹图谱技术。
Amino acid	氨基酸	蛋白质分子中的一个单体单位。
Aminoacyl or A site	氨酰或 A 位点	翻译过程中,氨酰-tRNA 占据的核糖体位点。
Aminoacylation	氨酰化	氨基酸结合到 tRNA 受体臂的过程。
Aminoacyl-tRNA synthetase	氨酰-tRNA 合成酶	催化一个或更多 tRNA 氨酰化的酶。
Amino terminus	氨基端	具有一个游离氨基的多肽末端。
Amplification refracton mutation system(ARMS test)	扩增不应突变系统	一种 SNP 分型技术,进行一对引物引导下的 PCR,其中的一个引物覆盖 SNP 位点。
Ancestral character state	原始特征态	一组生物的远古共同祖先所具有的特征态。
Ancient DNA	古代 DNA	古代生物样本中残留的 DNA。
Annealing	退火	寡聚核苷酸引物结合到 DNA 或 RNA 模板上的过程。
Anticodon	反密码子	tRNA 分子中第 34~36 位的三个核苷酸,与 mRNA 中的密码子发生碱基配对。
Anticodon arm	反密码子臂	tRNA 分子结构中的一部分。
Antigen	抗原	能激发免疫反应的物质。
Antitermination	抗终止	细菌中一种调节转录终止的机制。
Antiterminator protein	抗终止子蛋白	一种蛋白质,能与细菌的 DNA 结合,介导抗终止。
AP endonuclease	AP 核酸内切酶	碱基切除修复中的一种酶。
Apomorphic character state	变形特征态	所研究物种中的一个亚组所进化出的一组特征态,与原始特征态不同。
AP(apurinic/apyrimidinic) site	AP(无嘌呤/无嘧啶)位点	DNA 分子中核苷酸碱基丢失的位置。



Apoptosis	凋亡	程序性细胞死亡。
Archaea	古生菌	两大类原核生物之一,大多数都发现于极端环境中。
Artificial gene synthesis	人工基因合成	从一系列重叠的寡聚核苷酸中构建一个人造基因。
Ascospore	子囊孢子	如酿酒酵母这样的子囊菌有丝分裂的一个单倍体产物。
Ascus	子囊	酿酒酵母单次有丝分裂产生的包含四个子囊孢子的结构。
Attenuation	衰减作用	某些细菌依据细胞内氨基酸水平调节氨基酸生物合成操纵子表达的过程。
AU-AC intron	AU-AC 内含子	是真核生物核基因中的一种内含子,内含子的最初的两个核苷酸是 5'-AU-3',最后两个核苷酸是 5'-AC-3'。
Autonomously replicating sequence(ARS)	自主复制序列	特别是指在酵母中一段可以赋予无复制能力的质粒具有复制能力的 DNA 序列。
Autopolyploid	同源多倍体	来自于相同物种的两个配子融合产生的,这两个配子都不是单倍体。
Autoradiography	放射自显影	通过 X 光敏感性胶片的曝光来检测放射性标记的分子。
Autosome	常染色体	不是性染色体的染色体。
Auxotroph	营养缺陷型	一类只能在含有野生型微生物不需要的某营养成分的条件下生长的突变型微生物。
Backtracking	反向移动	RNA 聚合酶沿着 DNA 模板链向后移动一小段距离。
Bacteria	细菌	两大类原核生物之一。
Bacterial artificial chromosome(BAC)	细菌人工染色体	基于大肠杆菌 F 质粒的一种高容量克隆载体。
Bacteriophage	噬菌体	感染细菌的一种病毒。
Bacteriophage P1 vector	噬菌体 P1 载体	基于噬菌体 P1 的一种高容量克隆载体。
Barcode deletion strategy	条码缺失策略	一种用于在酿酒酵母中进行大规模筛选缺失突变的方法。
Barr body	巴氏小体	失活 X 染色体具有的高度浓缩的染色质结构。
Basal promoter	基本启动子	真核启动子中起始复合体组装的位置。
Basal promoter element	基本启动子元件	很多真核启动子中存在,开始基本水平的转录起始的序列模块。
Basal rate of transcription	基础转录速率	特定启动子单位时间内转录起始发生的个数。
Base analog	碱基类似物	一类具有同 DNA 分子中某一种碱基类似结构使其可作为诱变剂的化合物。
Base excision repair	碱基切除修复	通过切除并替换异常碱基的一种 DNA 修复过程。
Baseless site	无碱基位点	DNA 分子中核苷酸碱基丢失的位置。
Base pair	碱基对	由两个互补的核苷酸通过氢键形成的结构,当缩写成“bp”时,就是双链 DNA 分子长度的最小单位。
Base pairing	碱基配对	通过碱基对,一条多聚核苷酸与另一条多聚核苷酸结合,或者一条多聚核苷酸的一部分与该多聚核苷酸的另一部分结合。
Base ratio	碱基比例	在双螺旋 DNA 中的 A 与 T 或 G 与 C 的比例,Chargaff 发现碱基比例总是接近 1。
Base stacking	碱基堆积力	发生在双螺旋 DNA 分子中相邻碱基对之间的疏水性相互作用。
Basic domain	碱性结构域	一种 DNA 结合结构域。
B chromosome	B 染色体	群体中部分个体(不是全部)中含有的一条染色体。
B-DNA	B 型 DNA	活细胞中 DNA 双螺旋最普遍的构型。
Beads-on-a-string	串珠模型	由核小体小球连在 DNA 链上形成的一种未折叠的染色质结构。



- 代谢物活化蛋白 320, 424  
代谢物阻遏 423  
代谢组 188  
代谢组学 188  
带-螺旋-螺旋基序 304  
单倍体 4  
单倍体不足 506  
单核苷酸多态性 70  
单核苷酸重复单元 220  
单链结合蛋白 471  
单体群 591  
单一孤儿 161  
单源性 580  
蛋白电泳 177  
蛋白酶 7, 401  
蛋白酶体 409  
蛋白谱 177  
蛋白质-蛋白质交联 390  
蛋白质工程 157  
蛋白质水解切割 401  
蛋白质相互作用图谱 185  
蛋白质折叠 401  
蛋白质组 5, 17, 21, 382  
蛋白质组学 177  
等电点 177  
等电聚焦 177  
等容线 207  
等位基因 69  
等位基因频率 592  
等位基因随机分离 76  
等位基因特异性寡核苷酸杂交 496  
等位基因自由组合 76  
第二信使 427  
颠换 494  
点突变 73, 494  
电离辐射 502  
电泳 496  
电子显微镜技术 390  
凋亡途径 485  
调节型 507  
调节型控制 318  
定点诱变 157  
定量 PCR 595  
定位克隆 121  
定向进化同源基因 148  
定向诱变 157  
定型复合体 355  
端粒 53, 205  
端粒 DNA 478  
端粒酶 478  
短干扰 RNA 16, 370  
短散布核元件 209  
短散在重复元件 260  
短修补 514, 515  
钝端 38  
多半胱氨酸锌指 304  
多次击中 584  
多次取代 584  
多蛋白 405  
多等位基因 69, 591  
多基因家族 216  
多聚 C 尾 42  
多聚 G 尾 42  
多聚蛋白 255  
多聚嘧啶束 353  
多聚腺苷酸化编辑 368  
多聚腺苷酸化特异因子 349  
多聚腺苷酸结合蛋白 349  
多拷贝 155  
多区域进化 592  
多梳基因家族 434  
多梳应答元件 434  
多态性 591  
多维蛋白质鉴定技术 184  
多源性 580  
二倍体 4  
二度生长 424  
二级结构 17, 18  
二价体 79  
二硫键 18  
发夹结构 139  
发育 418  
发育调节因子模块 322  
翻译 5, 17  
翻译终止 400  
翻译起始 390  
翻译跳跃 399  
反馈环 435  
反密码子臂 383  
反式取代 286  
反向移动 336  
反转录基因 556  
泛素 409  
泛素化 285  
放射性标记 7  
放射杂合体组 99  
放射杂交体 98  
放射自显影 33, 41  
非编码 RNA 15, 290  
非标准密码子 23  
非同义 503  
非同源末端连接 510, 517  
非外显 506  
非遗传息肉结肠直肠癌 519  
非中期染色体 96  
费城染色体 277  
分化 418  
分支 579  
分支迁移 528  
分支系统学说 577  
分子伴侣 403  
分子伴侣素 404  
分子灯塔 74  
分子内碱基配对 138  
分子梳理 92  
分子钟 553, 582  
蜂毒肽 405  
弗里德赖希氏共济失调 499  
浮力密度 207  
辅激活蛋白 323  
辅助噬菌体 112  
辅阻抑物 320  
负超螺旋 228  
附加体转移 87  
复合型转座子 262  
复制蛋白 A 471, 473  
复制工厂 474  
复制后复合物 482  
复制后修复 533  
复制滑移 498, 555  
复制介导蛋白 471  
复制起点 464  
复制起始位点 45, 53  
复制前复合物 482  
复制体 472  
复制型 111  
复制型转座 258  
复制型转座的 DNA 转座子 537



引发前复合物 465	正交变电场凝胶电泳 92	转换 494
引发小体 470	支架附着区域 278	转基因小鼠 155
引物 34, 113, 467	直接修复 510	转录 5
隐蔽剪接位点 355	直立人 592	转录偶联修复 515
隐性 76	直系同源 581	转录泡 335
应答模块 322	指导 RNA 368	转录物 142
应急反应 340	质粒 44, 230	转录物特异性调节 395
荧光原位杂交 33, 89, 94, 276	质量/电荷比 178	转录组 4, 5, 13, 17, 170
营养缺陷型 87, 506	滞后链 467	转录组重构 175
优势对数值 86	中介子 317	转染 50
有根树 580	中期染色体 202	转位 397
有丝分裂 79	中枢 187	转运 RNA 16
有丝分裂期 482	中枢蛋白质 187	转座 221, 258, 527, 536
右臂 50	终止密码子 21	转座酶 261
右手螺旋 11	终止因子 324	转座噬菌体 262
诱变剂 494, 499	终止子序列 476	转座子 258
诱导 254	种系发生 576	转座子标记 161
诱导物 319	种属间印迹 143	转座子标记技术 153
原基因组 549	重叠基因 251	着色性干皮病 519
原聚体 250	重复 DNA 的 PCR 122	着丝点 205
原生质体 55	重复 DNA 指纹图谱 122	着丝粒 53, 202
原始特征态 577	重复性 DNA 219	紫外辐射 502
原噬菌体 254, 534	重复元件 123	自发错误 494
原养型 506	重排 140	自然选择 592
杂合 DNA 240	重取样分析 585	自私 DNA 408
杂合子 76	重取样值 586	自主复制序列 465
杂交 41	重组 46, 80, 81, 527	足迹法 306
杂交分析 58	重组酶 535	组成型 507
杂交探针 41	重组频率 82	组成型控制 318
杂种不育 263	重组热点 82	组成型异染色质 126, 278
早内含子 565	周期素 483	组蛋白 200
增强体 324	周期素依赖性激酶 483	组蛋白密码 285
增强子 321	珠蛋白 217	组蛋白去乙酰化酶 284
增殖细胞核抗原 468	主要成分分析 595	组蛋白修饰 285
折叠途径 403	专门方法 143	组蛋白乙酰化 282
真实树 580	专性寄生 250	组蛋白乙酰转移酶 283
整合酶 535	转导 87	最大简约法 585
整合子 231	转化 46, 87	左臂 50
整体调节 394	转化生长因子 $\beta$ 428	作图试剂 98
正超螺旋 228	转化要素 6	作用于 RNA 的腺嘌呤脱氨酶 367