普通高等教育“十一五”国家级规划教材

基因组学 第3版

杨金水 编著

高等教育出版社
HIGHER EDUCATION PRESS

普通高等教育“十一五”国家级规划教材

基因组学 第3版

JIYINZUXUE

杨金水 编著

高等教育出版社·北京
HIGHER EDUCATION PRESS BEIJING

内容提要

基因组学是当代生命科学发展最为迅速、关注度较高的学科之一。本书是国内第一本专门并全面介绍基因组学及其研究最新进展的学术著作，自2002年第1版、2007年第2版出版发行以来，在专业领域产生较大影响。本次修订突出系统性和简明实用性，重点阐述了基因组学的基本概念、基本理论，介绍了研究基因组的基本思路与技术手段，充分吸收了近几年国内外学科重要进展。

全书共分14章，分别是：基因组、遗传图绘制、物理图绘制、基因组测序与序列组装、基因组序列注释、基因组解剖、基因的转录调控、转录物组、蛋白质组、基因组表观遗传、基因组的复制、基因组进化的分子基础、基因组进化的模式、基因组与生物进化。

与第2版相比，第3版在如下方面做出了重大调整：第4章新增第三代测序；第5章增加基因本体及采用Gene Ontology注释基因方法；第8章由原“RNA修饰与加工”改为“转录物组”，第8章中涉及小RNA、非编码RNA及其在不同层次的基因调控内容进行更新；第9章由原来的“蛋白质合成与加工”改为“蛋白质组”，相关内容进行更新；第10章扩充了表观遗传学内容。书后新增索引，各个章节根据学科进展和教学需要，对有关数据、图表、思考题、参考文献等也进行了更新和调整。

本书可供综合性大学、高等师范院校、农林院校及医学院校本科生使用，也可供研究生及有关科研人员参考。

图书在版编目（CIP）数据

基因组学／杨金水编著. -- 3版. -- 北京：高等教育出版社，2013.1

ISBN 978-7-04-036836-9

Ⅰ.①基… Ⅱ.①杨… Ⅲ.①基因组－高等学校－教材 Ⅳ.①Q343.2

中国版本图书馆CIP数据核字（2013）第012830号

策划编辑 吴雪梅　　责任编辑 单冉东　　封面设计 张　楠　　责任印制 朱学忠

出版发行 高等教育出版社
社　　址 北京市西城区德外大街4号
邮政编码 100120
印　　刷 保定市中画美凯印刷有限公司
开　　本 850mm×1168mm　1/16
印　　张 27.25
字　　数 660千字
购书热线 010-58581118
咨询电话 400-810-0598

网　　址 http://www.hep.edu.cn
　　http://www.hep.com.cn
网上订购 http://www.landraco.com
　　http://www.landraco.com.cn
版　　次 2002年6月第1版
　　2007年7月第2版
　　2013年1月第3版
印　　次 2013年1月第1次印刷
定　　价 48.00元

物 料 号 36836-00

第3版前言

由于DNA测序技术的革新与发展，目前国内外基因组测序计划已经覆盖到几乎所有具有重要经济价值和理论研究意义的物种。同时，由于DNA测序技术的改进提高了DNA测序的效率，极大地降低了测序成本，促使了转录物组研究的蓬勃与深入发展，获取了许多此前无法探知的基因组表达谱信息。此外，生物信息技术的开发与应用，加快与深化了对结构基因组和功能基因组海量数据的分析、整理与归纳，使研究者得以从生命的整体视角，从生物体的不同结构层次和活性水平了解与认识基因组结构与功能、保守与进化的生物学意义。

《基因组学》第2版出版至今已经过去了5个年头。5年时间在科学研究的历史长河中只是一个瞬间，但对基因组学领域的研究而言，过去的5年取得了许多值得关注的进展。为了尽可能地反映基因组学领域的研究现状，为读者学习基因组学提供一些有益的参考，第3版沿袭原有的结构框架，在第2版的基础上，对原有的一些内容做了修改或重新撰写。在第4章中添加了第三代DNA测序，第5章中增加了基因本体以及采用Gene Ontology注释基因的方法。第8章的题目由原来的“RNA的修饰与加工”改为“转录物组”，第9章的题目由原来的“蛋白质的合成与加工”改为“蛋白质组”，相关内容进行更新。此外，近年来在基因组小RNA和非编码RNA研究领域取得了许多突破性进展，为在不同层次的调控水平解读基因功能提供了新的思路，为此在第8章中专门补充了许多相关的内容。

修订过程中也对各章节的参考文献进行了一次全面的梳理，尽可能地列出了相关领域的最新文献。限于篇幅，仍然有许多基因组学研究领域的重要进展未能在新版中体现，读者如有兴趣可以查阅相关的参考文献。

本书的再版编写工作得到了复旦大学生命科学院各方面的支持与帮助。作者实验室的王莹同学和刘爽同学在本书的文字校对、文献查阅以及编写内容安排等方面做了大量工作，并提出了一些好的建议，在此表示衷心感谢。

杨金水

2012年8月于复旦大学

第2版前言

截至2006年11月1日，国际上已完成测序的基因组为456个，其中古细菌29个。真细菌384个，真核生物44个。正在测序的古细菌56个，真细菌基因组995个，真核生物基因组632个，环境微生物基因组63个，已完成和正在进行的基因组测序计划总共为2 202个。由于DNA测序方法的改进。基因组测序计划的数目正在迅速地增加。与此同时，基因组注释与功能基因组的研究也呈现百花齐放的局面。基因组学的研究已深入到生命科学的各个领域，正在深刻地影响着生命科学未来的发展方向。

为了适应基因组学蓬勃发展的现状，及时反映基因组学领域近年来所取得的重要成果与重大进展，作者对已有的内容进行了扩充与调整。再版《基因组学》仍然保持原书的结构框架，仅将原第10章“染色质结构与基因表达”改为“基因组表观遗传”，并相应地增补了与之有关的内容。虽然各个章节都在原有内容的基础上作了更新，但在基因组注释和基因组进货方面增添了更多的内容。这也是近年来基因组学研究中最令人注目的领域。

再版《基因组学》仍将重点放在一些基本概念的阐述上。由于篇幅有限，许多研究进展只能简要提及，读者如何有兴趣可以查阅相关的参考文献。

本书的再版编写工作得到了复旦大学生命科学院各方面的支持与帮助。作者实验室的洪芳同学、王玉锋同学和查笑君同学在本书的文字校对、文献查阅以及编写内容安排等方面做了大量的工作，在此表示衷心感谢。

杨金水

2006年11月8日于复旦大学

第 1 版前言

在人类基因组计划的影响下，分子生物学的主要目标已经从传统的单个基因的研究转向对生物整个基因组结构与功能的研究。生命科学正从全新的视角研究与探讨生长与发育、遗传与变异、结构与功能以及健康与疾病等生物学与医学基本问题的分子机理，并形成了一门新的学科分支——基因组学。基因组学研究的对象涉及原核生物和真核生物不同的种属，其所研究的内容触及生命学科的各个领域，对生命科学的未来发展将产生重大影响。

为了向青年学生介绍基因组学的基本概貌，作者在已有教学的基础上，参考有关的书籍，收集与整理了近年来基因组学研究的最新资料，编著了《基因组学》一书。本书共分 14 章，第 1 章至第 5 章主要涉及基因组的结构，重点基因组遗传图与物理图绘制的原理与方法，这是基因组测序与序列组装的基础，同时对基因组序列诠释的依据与注解的方法进行了分析与探讨。第 6 章至第 10 章着重介绍基因组的功能，包括基因水平以及基因组水平的表达与调控。第 11 章至第 14 章讲述基因组的进化，内容涉及基因组进化的分子机制以及进行的模式与生物多样性的关系。

基因组学是一门年青的学科，并处在迅速发展之中。书中所介绍的许多方法甚至某些实验结果在读者看到本书时或许已有改变或修正，因此希望读者随时关心基因组学相关领域的最新进展，不必囿于已有的结论。此外书中不少章节还介绍了基因组学研究中一些尚未定论的观点以及某些目前还缺少有效方法进行研究的难题，目的是为了读者提供更多的思维空间，或许能从中找出自己现在或将来感兴趣的研究方向。

本书的编写工作得到赵寿元先生和乔守怡教授的全力支持与帮助。作者实验室的钱晓茵博士撰写了本书第 13 章“比较基因组学”一节，柯越海博士为作者提供了有关人类起源、进化与迁徙相关的基因组学的大量资料，左开井博士、王东和黄骥同学为本书绘制插图出力不少，在此一并表示衷心感谢。

杨金水

2002 年 2 月 19 日于复旦大学

目 录

第 1 章

基 因 组

基因组（genome）一词系由德国汉堡大学 H. 威克勒教授于 1920 年首创，用以表示真核生物从其亲代所继承的单套染色体，或称染色体组。更准确地说，基因组是指生物的整套染色体所含有的全部 DNA 序列。由于在真核细胞的线粒体和植物的叶绿体中也存在 DNA，因此又将线粒体或叶绿体所携带的 DNA 称为线粒体基因组或叶绿体基因组。原核生物基因组则包括细胞内的染色体和质粒 DNA。此外，非独立生命形态的病毒颗粒也携带遗传物质，称为病毒基因组。所有生命都具有指令其生长与发育，维持其结构与功能所必需的遗传信息，本书中将生物所具有的携带遗传信息的遗传物质总和称为基因组。

基因组学（genomics）一词系由 T. 罗德里克（T. Roderick）于 1986 年首创，用于概括涉及基因组作图、测序和整个基因组功能分析的遗传学学科分支，并已用来命名一个学术刊物 *Genomics*。基因组学是伴随人类基因组计划的实施而形成的一个全新的生命科学领域。

基因组学与传统遗传学其他学科的差别在于，基因组学是在全基因组范围研究基因的结构、组成、功能及其进化，因而涉及大范围高通量收集和分析有关基因组 DNA 的序列组成，染色体分子水平的结构特征，全基因组的基因数目、功能和分类，基因组水平的基因表达与调控以及不同物种之间基因组的进化关系。基因组学的研究方法、技术和路线有许多不同于传统遗传学的特点，各相关领域的研究仍处于迅速发展和不断完善的过程中。

1.1 遗传的分子基础

绝大多数生物，包括低等生物和高等生物的基因组都由脱氧核糖核酸（deoxyribonucleic acid，DNA）组成，少数病毒基因组则为核糖核酸（ribonucleic acid，RNA）。基因组所含有的遗传信息由 DNA 或 RNA 分子中核苷酸的排列顺序所决定，它们组成独立的结构单位——基因。基因所包含的信息可由特定功能的蛋白质解读，这类蛋白质附着在 DNA 或 RNA 分子

的一定位置，起始一系列的生化反应，合成基因的编码产物，这一过程称之为基因表达。基因表达由两个主要步骤组成：第一步以 DNA 分子为模板合成 RNA 拷贝，称为转录；第二步由 RNA 拷贝指令蛋白质的合成，称为翻译。DNA、RNA 和蛋白质这 3 种生物有机大分子的关系可以描述为生命的中心法则。如图 1.1 所示，DNA 处于中心法则的源头，一切生命的奥秘都储存在基因组中。

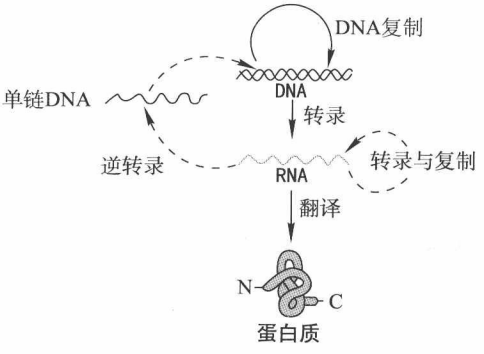

图 1.1　中心法则——遗传信息流

遗传信息流的总的方向是从 DNA 到 RNA，再从 RNA 到蛋白质。逆转录可将 RNA 的遗传信息转移到 DNA，但逆转录的遗传信息的表达仍然要经由 DNA 到 RNA，再从 RNA 到蛋白质的流程，并不改变遗传信息流的方向

遗传信息在上下代细胞之间和个体的世代之间的传递是通过 DNA 的复制和细胞的分裂完成的。复制使亲代 DNA 加倍，通过细胞分裂将两份相同的 DNA 拷贝分配到两个子代细胞中。DNA 在复制时偶尔会发生突变与重组，使其所携带的遗传信息改变，它们是生命进化与生物多样性的源泉。生命的所有现象都与 DNA、RNA 和蛋白质的结构与功能有关。

1.1.1　DNA 的化学与生物学

生命活动无论在何种水平，每一种特定的生物学功能总有特定的结构基础与之对应。DNA 几乎是所有生命形式都必须具备的遗传物质。DNA 的生物学功能主要是储存遗传信息，DNA 的物理结构和 DNA 的化学性质与其功能是相适应的。

核苷酸与多聚核苷酸

DNA 是一种长链多聚分子，由 4 种核苷酸组成，这 4 种核苷酸可以任何次序排列连接成可达数百上千万个核苷酸的长链分子。每个核苷酸分子都含有 3 个组分（图 1.2）。

2′－脱氧核糖（2′-deoxyribose）　这是一种五碳糖，即由 5 个碳原子组成的核糖。2′－脱氧核糖系指核糖的 2′－碳原子上连接的羟基（—OH）基团由氢原子取代。

含氮碱基（nitrogenous base）　共有 4 种，即 2 个嘧啶分子，分别为胞嘧啶（cytosine）和胸腺嘧啶（thymine）；2 个嘌呤分子，即腺嘌呤（adenine）和鸟嘌呤（guanine）。这些碱基通过 β-*N*-糖基键（β-*N*-glycosidic bond）分别与嘧啶环的 1 位氮原子和嘌呤环的 9 位氮原子共价连接。

磷酸基团　磷酸基团与核糖分子的 5′－碳原子相连。由糖分子与碱基组成的分子称为核苷（nucleoside）；加上磷酸后成为核苷酸。核苷酸含有的磷酸基团可分为 3 类，即单磷酸，双磷酸，三磷酸。

虽然细胞中均含有单磷酸、双磷酸和三磷酸基团的核苷酸，但只有核苷三磷酸才是合成 DNA 的底物。4 种核苷三磷酸的全称为：2′－脱氧腺嘌呤－5′－三磷酸，2′－脱氧胞嘧啶－5′－三磷酸，2′－脱氧鸟嘌呤－5′－三磷酸，2′－脱氧胸腺嘧啶－5′－三磷酸。这 4 种核苷酸的简称依次分别为 dATP，dCTP，dGTP 和 dTTP，当注明为 DNA 的序列时，可简写成 A，C，G 和 T。当未注明具体的脱氧核糖核苷酸，而是泛指 4 种核苷酸时，可以写成 dNTP，“N”代表 A，C，G 和 T 中的任意一种。

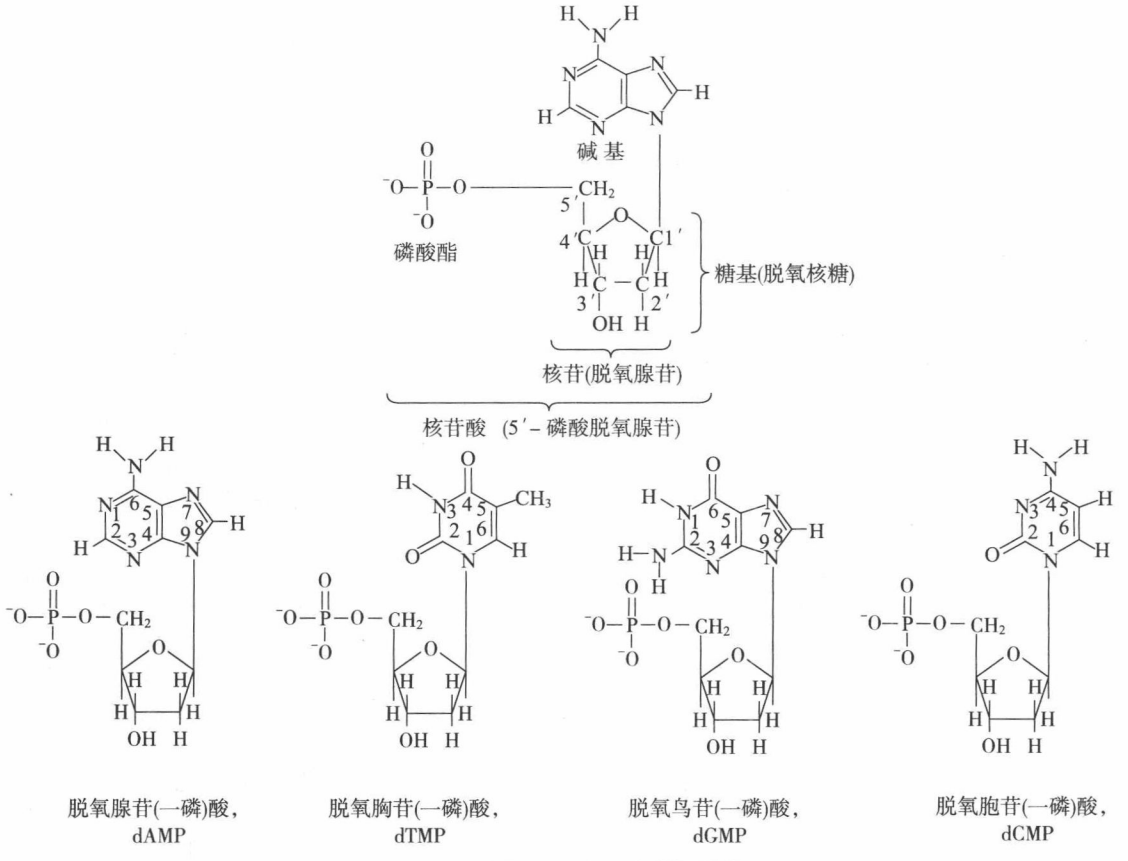

图 1.2 DNA 的组成单元为脱氧核苷酸

每个脱氧核苷酸均由一个 2′-脱氧核糖，一个磷酸基团和一个碱基组成。碱基有 4 种，即腺嘌呤、胞嘧啶、鸟嘌呤和胸腺嘧啶。磷酸基团与脱氧核糖的 5′-碳原子相连，碱基则连接在核糖的 1′-碳原子上

由于核苷酸单体中提供聚合反应所需能量的三磷酸基团直接与核糖的 5′碳原子相连，发生 DNA 聚合反应时，下一个核苷酸单体与前一个核苷酸单体的聚合总是由下一个核苷酸单体的 5′-三磷酸与前一个核苷酸单体的 3′-OH 缩合生成磷酸脂键。因此多聚核苷酸链的化学反应只能是 5′→3′方向，所有天然的 DNA 聚合酶都只执行 5′→3′的合成，同样的极性也表现在以 DNA 为模板合成 RNA 拷贝的反应中（图 1.3）。核苷酸聚合反应的方向性使双链 DNA 的复制复杂化。因为 DNA 的两条互补单链化学极性正好相反，在复制时必须采取不同的策略，即 3′链采取连续复制，而 5′链采取间断复制。

DNA 的双螺旋结构

2 条反向平行的 DNA 单链彼此相互缠绕组成双螺旋分子，有 2 种化学作用稳定双螺旋结构：

碱基配对（base-pairing） 位于 2 条 DNA 单链中的碱基可相互配对。配对只发生在腺嘌呤（A）与胸腺嘧啶（T）或鸟嘌呤（G）与胞嘧啶（C）之间，因为只有 A 与 T 和 G 与 C 配对才使 DNA 双螺旋成为最稳定的状态（图 1.4），A 与 T 或 G 与 C 称为互补碱基对。

碱基堆积（base-stacking） 碱基堆积系指与 DNA 双螺旋主轴垂直的相邻碱基对杂环之间的互作，可增加双螺旋的稳定性。

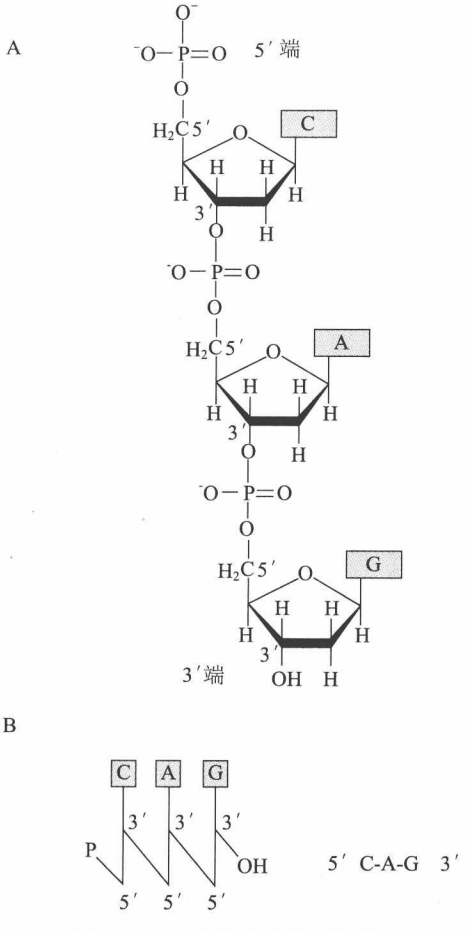

图 1.3　多聚核苷酸分子结构

A. 4 种脱氧核苷酸以磷酸二酯键连接，由此聚合成长链分子。B. DNA 分子的合成按 5′→3′方向进行，加入的每个核苷酸总是以 5′－磷酸与 DNA 单链的最后一个核苷酸 3′－OH 缩合形成磷酸二酯键

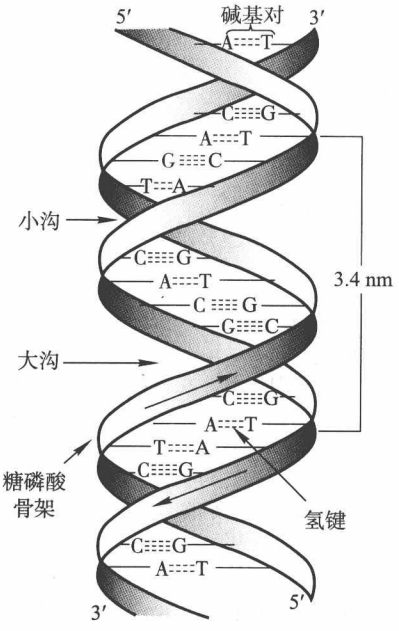

图 1.4　DNA 分子的双螺旋结构

每旋转一圈约 10.5 个核苷酸，两个大小凹槽沿 DNA 分子交替排列

碱基配对具有重要的生物学意义。在双链 DNA 复制时，每条单链均可作为模板以碱基互补的方式合成新链，它们的序列组成与亲链完全相同，使遗传信息以极其简单而准确的方式忠实地传递给下一代。

碱基配对还有另一特征，即 A/T 碱基对与 G/C 碱基对之间氢键数目的差异。A/T 碱基对含两对氢键，G/C 碱基对为 3 对氢键，因此 G/C 碱基对比 A/T 碱基对更加稳定。或者说，A/T 碱基对比 G/C 碱基对更易解链。正是这一特点，基因组 DNA 中的某些序列为适应特别的功能便含有更多的 A/T 或 G/C 序列。如 DNA 转录与复制的起始区要求迅速解链，该区段含有较多的 A/T 碱基对。

DNA 单链彼此缠绕时，沿着双螺旋的走向交替分布两个凹槽，一个较宽且较深的凹槽称为大沟（major groove），另一个较窄且较浅的凹槽称为小沟（minor groove）。DNA 双螺旋中两个交替分布的大小沟具有特征性的结构信息，在基因表达中起重要作用。DNA 结合蛋白的

特定功能域可伸入大小沟中，通过氨基酸侧链与双螺旋大小沟中碱基杂环上的基团互作，由此识别 DNA 序列所包含的组成与结构的信息。

除少数单链 DNA 病毒外，活体中的 DNA 分子均以双链形式存在。单链 DNA 病毒的复制也必须先转变为双链 DNA，然后以半保守方式复制，再合成单链 DNA 病毒。

DNA 双螺旋构象

Watson 和 Crick 描述的 DNA 双螺旋是天然 DNA 的构象之一，即 B 型 DNA。其他的 DNA 双螺旋构象还包括 B′-，C-，C′-，C″-，D-，E-和 T-DNA，所有这些构象都是右旋，此外还有左旋构象的 Z-DNA。

双螺旋 DNA 分子的不同构象影响到蛋白质接触双螺旋内部的程度，沟槽内表面化学基团的组成与位置提供了 DNA 结合蛋白可识别的空间信息。DNA 结合蛋白自身所具有的结构能使其阅读 B 型 DNA 大小沟槽中特定的碱基顺序，是基因调控的核心内容。一些特别的序列组成也能影响 DNA 构型，如果周期性即每隔 10 个碱基对重复出现连续的 A/T 碱基对，所在的 DNA 区段将发生明显的弯曲。原因是，A/T 碱基对形成的与 DNA 主轴垂直的上下两个平面有点倾斜。当倾斜连续发生在一个方向时，就会发生机械的弯曲。锥虫（trypanosome）动基体（kinetoplast）的线性 DNA 片段电泳时迁移率降低，因为 DNA 分子中出现周期性重复的 $(A/T)_{5\sim6}$与 $(G/C)_{4\sim6}$序列，使其构型偏离一般的线形 DNA，从而影响电泳行为。现已证明弯曲 DNA 在基因的表达调控中起重要作用。

DNA 拓扑学

DNA 有两种构型，即线性 DNA 与环状 DNA。细菌质粒、大肠杆菌染色体、线粒体、叶绿体以及哺乳类 DNA 病毒基因组均由共价连接的双链 DNA 组成。λ 噬菌体染色体在生活史中有线性 DNA 与环状 DNA 两种状态。

生物活体细胞中的 DNA 总是与蛋白质结合在一起。结合的蛋白质可使 DNA 的双螺旋轻微解旋，因此单位长度的与蛋白质结合的 DNA，其螺旋圈数要少于游离的 B 型 DNA。当除去结合的蛋白质后，DNA 分子又恢复正常的 B 型螺旋圈数。对线性 DNA 而言，因存在游离的末端，增加螺旋圈数所产生的张力可由游离的末端释放。双链闭环 DNA 的情况却与此不同，因为闭环 DNA 的螺旋圈数在拓扑学上是固定的。如果没有其他的力学补偿，闭环 DNA 的双螺旋圈数是不能改变的。因此当结合蛋白质离开天然环状 DNA 后，会同时发生两种事件：①DNA 的双螺旋圈数恢复到正常 B 型 DNA 的数值；②环状 DNA 发生扭曲，扭曲方向与右旋方向相反，但圈数相同，这种 DNA 分子称为负超螺旋。如果超螺旋 DNA 的一条单链中出现单个缺口，超螺旋构型即将消失，拓扑学的约束力也不再存在（图 1.5）。基因的转录，DNA 的复制、修复与重组均涉及 DNA 分子的拓扑学变化。

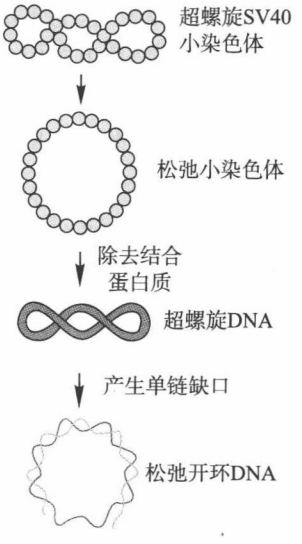

图 1.5 SV40 小染色体拓扑学结构

当除去与超螺旋小染色体结合的蛋白质后，为了补偿因回复 B 型 DNA 产生的张力，双链闭环 DNA 出现负超螺旋。当在超螺旋双链 DNA 的单链中产生缺口时，单链 DNA 释放张力转变为开环结构

DNA 甲基化

无论高等生物还是低等生物，其活体中的 DNA 都有不同程度的化学修饰，主要为甲基化。细菌中甲基化大多发生在腺嘌呤的 6 位氮原子与胞嘧啶 5 位碳原子。高等生物 DNA 的甲基化主要涉及胞嘧啶 5 位碳原子。

细菌中有专门的甲基化酶，它们可识别特异的 4 ~ 8 个碱基对序列并对其中特定碱基甲基化。哺乳动物中甲基化主要涉及双碱基 CpG，使其转变为甲基化^{m}CpG。高等植物中甲基化包括回文序列 CpG 和 CpNpG，N 为任意碱基。

基因由 DNA 组成

20 世纪 20 年代人们即已完成了细胞核的细胞化学研究，并证实染色体由 DNA 和蛋白质组成。由于 DNA 的组成过于简单，当时人们猜测只有蛋白质才有足够的多样性成为遗传物质，因此认为基因应由蛋白质组成。直到 20 世纪中期，2 项关键的实验报道才改变了人们关于基因分子基础的看法。

第一项实验由哥伦比亚大学 Avery OT、Macleod CM 和 McCarty M 于 1944 年完成。此前，一位英国人已经证实，肺炎链球菌（*Streptococcus pneumoniae*）死去的细胞中有一种未知的成分可使非致病肺炎球菌活细胞转变为致病品系。Avery OT 及其合作者完成了一系列实验，结果表明，与预期的想法相反，转化的物质不是蛋白质而是 DNA。这一重要发现当时并未引起人们的重视，许多人也不接受 DNA 是遗传物质的结论，部分原因是大多数微生物学家还不清楚转化是一种遗传现象还是一种生理现象。另一种批评意见认为，Avery 采用的脱氧核糖核酸的纯度值得怀疑，因为不能排除在纯化脱氧核糖核酸过程中存在痕量蛋白质污染的可能。

第二项实验是由美国纽约冷泉港实验室的 Hershey AD 和 Chase M 于 1952 年完成的，其结论无懈可击。他们的实验直接针对噬菌体生活史中感染的第一步，并清楚地证明只有 DNA 才具有遗传活性。在 Hershey-Chase 实验完成前一年，Watson J 和 Maaloe O 就已尝试用放射性标记追踪噬菌体感染中 DNA 的命运，可惜他们以及其他一些人的实验都未能给出结论性的证据，主要因为无法区分沾污在细菌细胞表面的噬菌体与感染中产生的新的噬菌体。Hershey AD 和 Chase M 采取同样的放射性标记方法，但在程序上作了重要修改。他们用放射性标记的噬菌体 T2 感染大肠杆菌，保温数分钟以使噬菌体吸附在细胞表面并将其内部的 DNA 注入到细胞中。然后将混合物在振荡器上振动，使空心的噬菌体从细胞表面脱落，再经离心收集含有噬菌体基因的细菌。结果证实有 70% 的噬菌体 DNA 和 20% 的噬菌体蛋白质保留在细菌组分中。在完成生活史循环后，新产生的噬菌体中仍含有 50% 原来的 DNA，而蛋白质降为 1%。这些结果表明，噬菌体转移到细菌中的基因与注入的 DNA 呈平行关系，DNA 是噬菌体的遗传物质并可遗传到下一代噬菌体中。

1.1.2　RNA 的化学与生物学

细胞中 RNA 的总量是 DNA 的 5 ~ 10 倍。RNA 的主要功能是传递遗传信息，参与基因的表达与调控。某些病毒，如逆转录病毒和许多动物、植物的单链与双链病毒基因组由 RNA 组成。

细胞中的 RNA 种类很多，其中含量较高的包括核糖体 RNA（ribosomal RNA，rRNA），转运 RNA（transfer RNA，tRNA）和信使 RNA（messenger RNA，mRNA）。此外，大多数细

胞中均含有其他一些胞质内小 RNA（small cytoplasmic RNA，scRNA）和核仁小 RNA（small nucleolar RNA，snoRNA），真核生物还含有核内小 RNA（small nuclear RNA，snRNA）。细胞中约 80% 的 RNA 由 rRNA 和 tRNA 组成，mRNA 约占 RNA 总量的 5%。

在真核生物中还发现一类在基因表达调控中起重要作用的小分子干扰 RNA，即微小 RNA（micro-interfering RNA，microRNA，miRNA）和小干扰 RNA（small interfering RNA，siRNA）（He L 等，2004）。这类非编码 RNA 分子的生物学功能多姿多彩，不仅涉及基因组的表达调控，而且参与基因组 DNA 的程序化和染色质的重建，是后基因组（post-genome）时代重要的研究内容之一。

RNA 的化学组成

RNA 与 DNA 在结构上有很大的相似性，也由 4 种核苷酸组成。由所含碱基代表分别称为腺嘌呤（A），鸟嘌呤（G），尿嘧啶（U）和胞嘧啶（C）（图 1.6）。RNA 多聚分子的合成方式与 DNA 相似，相邻核苷酸之间亦由 3′，5′-磷酸二酯键连接，具有相同的化学极性。不同之处在于，RNA 核苷酸中连接在核糖 2′-碳原子上的基团为羟基而非氢原子。这一看似微小的差别极其重要，因为 2′—OH 比 2′—H 的化学性质更为活泼，造成了 RNA 与 DNA 理化性质与生物学功能的不同（图 1.7）：

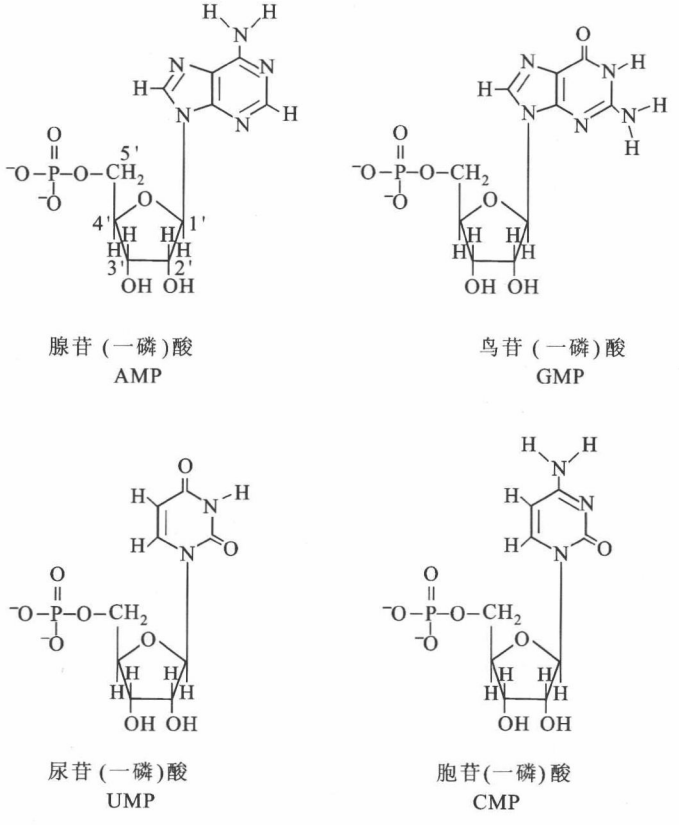

图 1.6 组成 RNA 分子的 4 种核苷酸