

杨金水 编著

基因组学

Genomics



高等教育出版社
HIGHER EDUCATION PRESS

100

基因组学

杨金水 编著



高等教育出版社
HIGHER EDUCATION PRESS

图书在版编目(CIP)数据

基因组学/杨金水编著. - 北京:高等教育出版社, 2002.6

ISBN 7-04-011087-3

I. 基... II. 杨... III. 基因组 - 高等学校 - 教材
IV. Q343.2

中国版本图书馆 CIP 数据核字(2002)第 038027 号

策 划 吴雪梅
编 辑 吕庆娟
封面设计 王凌波
责任排版 李 杰
责任印制 陈伟光

基因组学
杨金水 编著

出版发行	高等教育出版社	社 址	北京市东城区沙滩后街 55 号	邮政编码	100009
购书热线	010-64054588	传 真	010-64014048	网 址	http://www.hep.edu.cn
免费咨询	800-810-0598				http://www.hep.com.cn
经 销	新华书店北京发行所	印 刷	北京民族印刷厂	版 次	2002 年 6 月第 1 版
开 本	850×1168 1/16	印 次	2002 年 6 月第 1 次印刷	定 价	36.00 元
印 张	20.75				
字 数	500 000				

©高等教育出版社 2002

版权所有 侵权必究

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到生命学科各个领域,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,重点介绍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化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,目的是为读者提供更多的思维空间,或许能从中找到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
于复旦大学
2002年2月19日

目 录

1	基因组	1
1.1	遗传的分子基础	2
1.2	基因组顺序复杂性	12
1.3	基因与基因家族	15
1.4	染色体	17
1.5	基因组	19
2	遗传图绘制	21
2.1	遗传图与物理图	22
2.2	遗传作图标记	23
2.3	遗传作图的方法	25
2.4	人类遗传图	35
3	物理图绘制	37
3.1	限制性作图	38
3.2	基于克隆的基因组作图	42
3.3	荧光标记原位杂交	46
3.4	顺序标签位点(STS)作图	48
3.5	人类基因组物理图	52

4	基因组测序与序列组装	55
4.1	DNA 测序的方法学	56
4.2	DNA 顺序的组装	62
4.3	基因组测序的其他路线	68
4.4	人类基因组的测序与组装	69
5	基因组序列诠释	74
5.1	搜寻基因	75
5.2	基因功能预测	78
5.3	从基因组到细胞	88
6	基因组解剖	95
6.1	真核生物基因组解剖	96
6.2	原核生物基因组解剖	103
6.3	转座因子与分散重复顺序	106
6.4	串接重复顺序及其分布	111
6.5	人类基因组草图顺序中的编码基因	111
6.6	拟南芥基因组编码基因	114
7	基因的转录调控	118
7.1	原核生物基因的转录	119
7.2	真核生物基因的转录	124
7.3	古细菌基因的表达调控	136
7.4	DNA 结合蛋白在基因转录调控中的作用	136
8	RNA 的修饰与加工	145
8.1	细胞中的 RNA 组分	146

8.2	mRNA 的加工	147
8.3	非编码 RNA 的加工	160
8.4	RNA 前体内部的化学修饰与加工	163
8.5	RNA 干涉	167
9	蛋白质组	170
9.1	tRNA 在蛋白质合成中的作用	171
9.2	蛋白质合成中核糖体的作用	175
9.3	蛋白质翻译后加工	181
9.4	蛋白质周转	186
10	染色质结构与基因表达	191
10.1	核小体的结构与基因表达	192
10.2	位置效应	196
10.3	后生遗传与基因表达	200
10.4	基因表达与染色质重建	206
11	基因组的复制	214
11.1	DNA 复制的问题	215
11.2	原核生物基因组的复制	217
11.3	真核生物核基因组的复制	222
11.4	细胞器基因组的复制	231
11.5	基因组复制的调控	234
12	基因组进化的分子基础	240
12.1	突 变	241
12.2	重 组	253
12.3	转 座	260

13	基因组进化的模式	265
13.1	遗传系统的起源	266
13.2	新基因的产生	270
13.3	非编码顺序的扩张	280
13.4	比较基因组学	285
14	基因组与生物进化	291
14.1	分子系统发生学	292
14.2	分子系统发生学与生物进化	295
14.3	生物新特征的产生	305
	名词解释	311

1

基因组

- 1.1 遗传的分子基础
 - 1.1.1 DNA 的化学与生物学
 - 1.1.2 RNA 的化学与生物学
 - 1.1.3 蛋白质的结构
- 1.2 基因组顺序复杂性
 - 1.2.1 C 值与 C 值悖理
 - 1.2.2 顺序复杂性
 - 1.2.3 重复顺序
 - 1.2.4 单一顺序
 - 1.2.5 基因主要位于单一顺序
- 1.3 基因与基因家族
 - 1.3.1 编码 RNA 基因
 - 1.3.2 编码蛋白质基因
 - 1.3.3 基因家族
 - 1.3.4 异常结构基因
 - 1.3.5 假基因
- 1.4 染色体
 - 1.4.1 真核生物染色体
 - 1.4.2 原核生物染色体
- 1.5 基因组
 - 1.5.1 人类基因组
 - 1.5.2 其他生物基因组

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,生物所具有的携带遗传信息的遗传物质总和称为基因组。

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 2 个步骤组成,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如图 1.1。

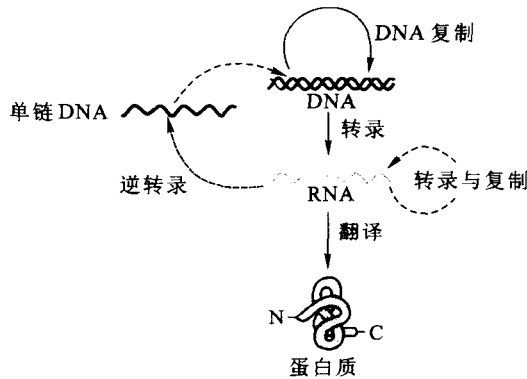


图 1.1 遗传信息流的方向

遗传信息在世代之间的传递是由 DNA 的复制完成的。复制使亲代 DNA 加倍,通过细胞分裂将 2 份相同的 DNA 拷贝分配到 2 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

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达数百万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,即腺嘌呤和鸟嘌呤(adenine, 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。

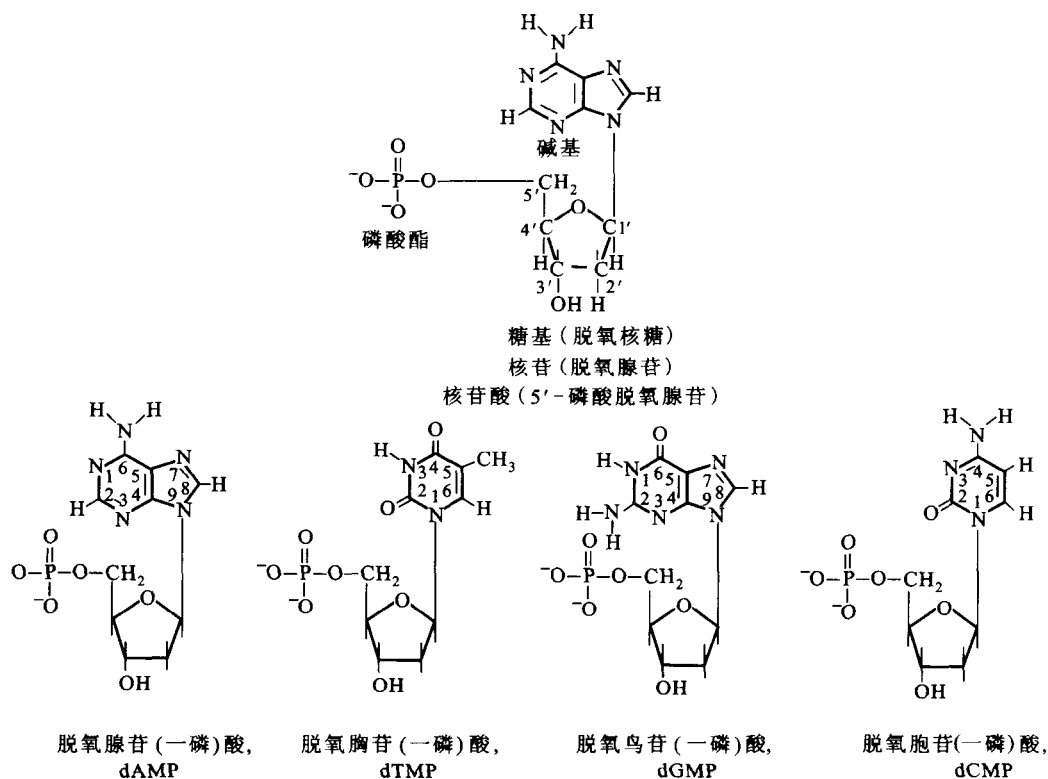


图 1.2 DNA 的组成单元为脱氧核苷酸

每个脱氧核苷酸均由 1 个 2'-脱氧核糖, 1 个磷酸基团和 1 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

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, 但只有三磷酸核苷酸才是合成 DNA 的底物。4 种三磷酸核苷酸的全称为: 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 的顺序时, 可简写成 A, C, G 和 T。

多聚核苷酸链的化学反应只能是 5'→3' 方向, 所有天然的 DNA 多聚酶都只执行 5'→3' 的合成 (图 1.3), 同样的极性也表现在以 DNA 为模板合成 RNA 拷贝的反应中。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 2 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。

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 配对才是最稳定的方式 (图 1.4), A 与 T 或 G 与 C 称为互补碱基对。

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:

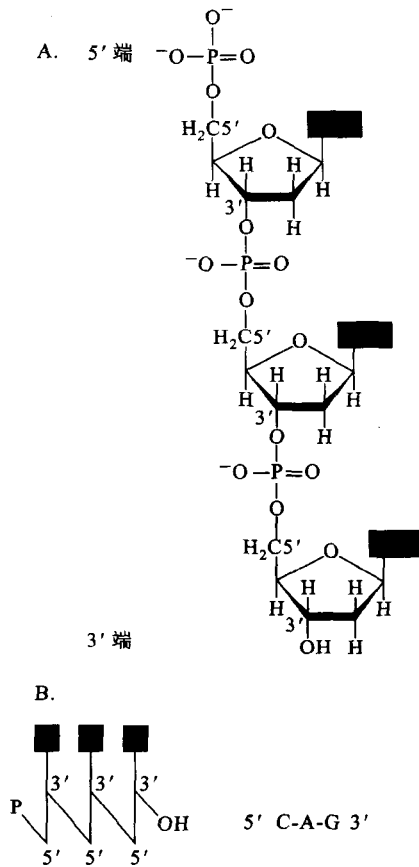


图 1.3 4 种脱氧核苷酸以磷酸酯键连接, 由此聚合成长链分子 DNA 分子的合成按 5'→3' 方向进行, 加入的每个核苷酸总是以 5'-磷酸与 DNA 单链的最后 1 个核苷酸 3'-OH 缩合形成磷酸酯键

碱基堆积 (base-stacking) 碱基堆积系指相邻碱基对杂环之间的互动, 可增加双螺旋的稳定性。

碱基配对具有重要的生物学意义, 它可以极其简单而准确的方式以已有的亲本链为模板合成新的 DNA, 使遗传信息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 碱基对之间氢键的差异。A/T 碱基对含 2 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。如 DNA 转录与复制的起始区要求迅速解链, 该区段有较多的 A/T 顺序。

DNA 单链彼此缠绕时, 沿着双螺旋的走向交替分布 2 个凹槽, 一个较宽与较深的凹槽称为大沟 (major groove), 另一个较窄及较浅的为小沟 (minor groove)。DNA 双螺旋中 2 个交替分布的大小沟具有特征性的遗传信息, 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, 识别

DNA 顺序所包含的信息。

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。

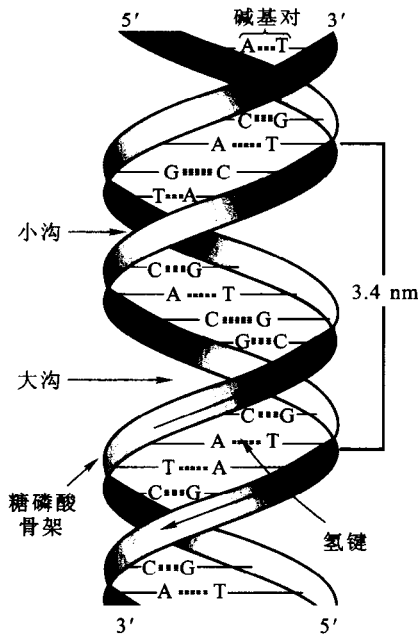


图 1.4 DNA 分子的双螺旋结构
每旋转 1 圈约 10.5 个核苷酸,2 个大小凹
槽沿 DNA 分子交替排列

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-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-DNA 中特定的核苷酸顺序。一些特别的顺序组成也能影响 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。锥虫(trypanosome)动基体(kinetoplast)的线性 DNA 片段电泳迁移率降低,因为 DNA 分子中出现周期性重复的(A/T)₅₋₆与(G/C)₄₋₆序列,使其构型偏离一般的线性 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 2 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。

活体中的 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-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 2 种事件: ① DNA 的双螺旋圈数恢复到正常 B-DNA 的数值; ② 环状 DNA 发生扭曲, 扭曲方向与右旋方向相反, 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。

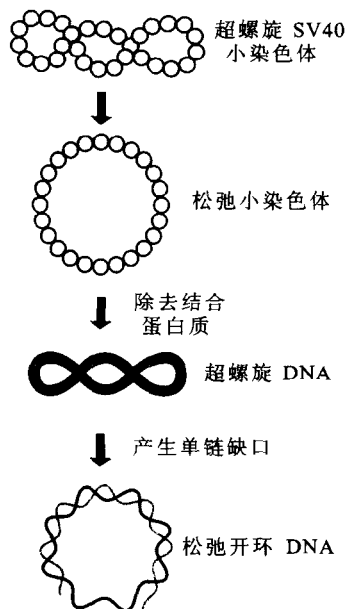


图 1.5 SV40 小染色体的拓扑学结构

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 型构型产生的张力, 双链闭环 DNA 出现负超螺旋构型。当在超螺旋双链 DNA 的单链中产生缺口时, 单链 DNA 释放张力变为开环结构

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。细菌中甲基化发生在腺嘌呤的 6 位氮原子与胞嘧啶的 5 位碳原子。高等生物 DNA 的甲基化主要涉及胞嘧啶的 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 bp 顺序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^mCpG。高等植物中甲基化包括回文顺序 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, 并证实染色体由等量的 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, Macleod 和 McCarty 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链球菌活细胞转变为致病品系。Avery 及其合作者完成了一系列细心的实验, 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。

第二项实验是由纽约冷泉港美国国家实验室的 Alfred Hershey 和 Martha Chase 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 和 Ole Maaloe 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染在细菌细胞表面的噬菌体与感染中产生的新的噬菌体。Hershey 和 Chase 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:噬菌体接种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、植物及昆虫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小分子细胞质 RNA (small cytoplasmic RNA, scRNA),真核生物还含有小分子细胞核 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%,其余为 scRNA 和 snRNA。

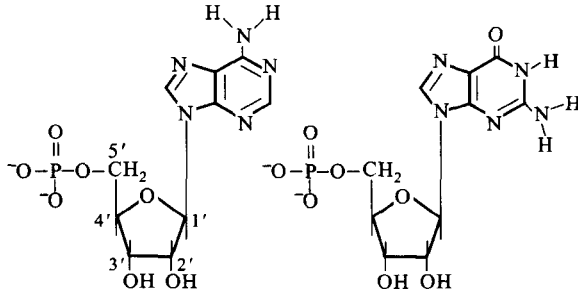
RNA 的化学组成

RNA 与 DNA 在结构上有很大的相似性,也由 4 种核苷酸组成,所含碱基分别为腺嘌呤(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 整个理化性质与生物学功能的不同:

① 2'-OH 非常靠近连接 2 个核苷酸的磷酸二酯键位置,使 RNA 的磷酸二酯键对碱性环境变得十分敏感,即使在正常细胞内略微偏碱的 pH 条件下也很容易降解。

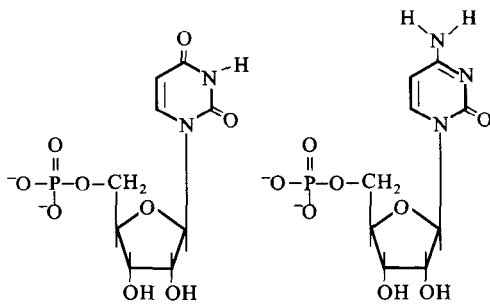
② 2'-OH 的化学作用使 RNA 构型的选择范围受到很大限制,通常 RNA 双螺旋区段均在数十碱基对以下。

③ 长链 RNA 增加了 2-OH 与磷酸酯键互作引起分子断裂的可能(图 1.7),限制了 RNA 长度的增加,这是 RNA 远比 DNA 相对分子质量小得多的主要原因。由于不能形成



腺苷(一磷)酸, AMP

鸟苷(一磷)酸, GMP



尿苷(一磷)酸, UMP

胞苷(一磷)酸, CMP

图 1.6 组成 RNA 分子的 4 种核苷酸,与脱氧核糖核苷酸的主要差别是连接核糖 2'-碳原子的基团为羟基(-OH)

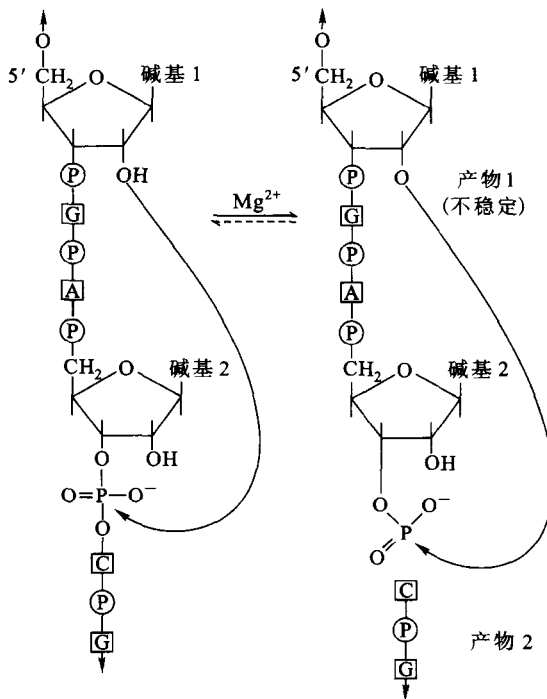


图 1.7 核苷酸 2'-OH 在碱性条件下极易与磷酸酯键互动引起 RNA 分子的断裂,这是 RNA 分子不稳定的主要原因

很长的多聚分子, RNA 只能储存有限的遗传信息。

④ 2-OH 可参与同磷酸或碱基的互作从而稳定 RNA 分子的折叠构型, 所以 RNA 分子比 DNA 分子更易形成稳定而紧凑的三级结构, 并使 RNA 分子获得了某些重要的催化功能。

RNA 与 DNA 的另一差别是, DNA 中胸腺嘧啶(T)的位置在 RNA 中由尿嘧啶(U)取代, 这一变化对 RNA 与 DNA 的生物学功能有深远影响。当 RNA 分子中的胞嘧啶甲基化脱氨基形成尿嘧啶后, 细胞无法区分突变的尿嘧啶与正常的尿嘧啶, 无疑会增加 RNA 的突变机率。可以想象, 如果 DNA 中胸腺嘧啶(T)也由尿嘧啶(U)取代, 将会面临同样的问题。天然的 DNA 中只存在胸腺嘧啶(T)而无尿嘧啶(U), 因此当 DNA 顺序中出现胞嘧啶甲基化脱氨基形成的尿嘧啶时, 可立即由细胞的修复系统识别并予以剔除与修复, 维持了遗传的稳定性。这是生命在进化过程中最终选择了 DNA 而不是 RNA 作为遗传信息主要载体的另一个重要原因。

RNA 的高级结构

除了少数动物基因组由类似 A-DNA 的双链 RNA 分子组成之外, 绝大多数天然 RNA 分子均以单链形式存在。几乎所有的单链 RNA 分子, 不管其大小如何都会形成或长或短的分子内双螺旋结构, 这是单链 RNA 分子的一个显著特征(图 1.8)。RNA 分子内碱基配对的规则与 DNA 的碱基配对类似, 即 A 与 U 配对, G 与 C 配对。G 也可与 U 配对, 但稳定性比 G/C 配对差, 因为前者只能形成 2 对氢键。

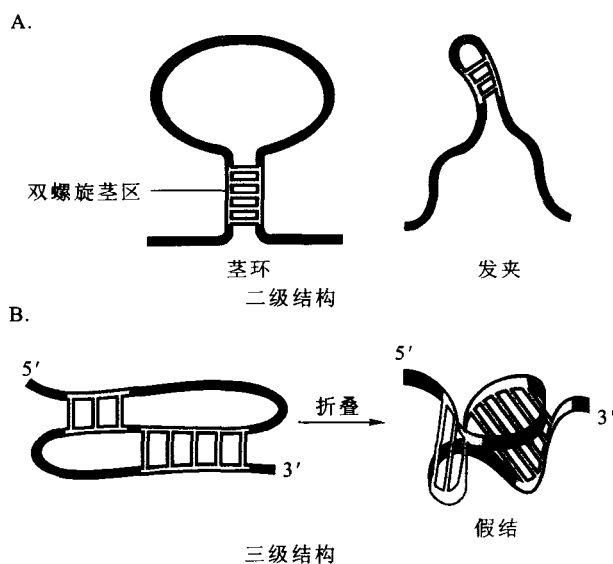


图 1.8 单链 RNA 的分子内碱基配对可形成不同构型的高级结构
A. 二级结构; B. 三级结构

RNA 的分子内双螺旋由单链区段回折形成, 常常呈现类似茎环的结构。以这种方式形成的双螺旋一般都很短, 约为 10 对左右核苷酸。由于单链回折形成的双螺旋很少有完全的碱基互补顺序, 因此 RNA 双螺旋中常常出现如 G/U 的不完全配对或配对间断。

RNA 的化学修饰

几乎所有细胞中的 RNA 在合成之后均会受到不同程度的化学修饰, 涉及 RNA 分子的