

医学生物学参考资料

(二)

《分子生物学引论》



中国人民解放军第四军医大学生物学教研室

1079
66
1

说 明

这本《分子生物学引论》是我们翻印复旦大学生物系的一本讲义。该书综合了大量近些年的有关研究资料，对于基因的本质以及细胞中的基本生命过程，特别是对于真核细胞基因组的特点和基因调控，都有较好的描述。对于学习《分子生物学》，特别是《基因工程》很有参考价值。

由于分子生物学进展迅速，已经渗透并影响到医学的许多学科，因此这本书对医学院校中的教师和医生以及研究生，都有参考意义。在我校科研处的支持下，由西北大学生物系胡楷同志提供样本，同意加以翻印。我们作为内部的《医学生物学参考资料》（二），期望能对读者有所帮助。

由王泰清、李永顺、刘素英同志负责校对，发现原样本中的少数遗误之处，也加以校正。附录的参考文献，加以集中，排在每章之后。印刷工作由教材处印刷所负责。

四 医 大 生 物 学 教 研 室

1980年3月

目 录

第一部分 基本概念

第一章 基因的物质本性.....	1
§ 1.1 核酸	1
§ 1.2 DNA分子结构的双螺旋模型.....	5
§ 1.3 遗传密码	10
§ 1.4 基因和基因组	12
§ 1.5 基因突变	15
§ 1.6 几种RNA病毒	18
第二章 细胞中的基本生命过程.....	21
§ 2.1 DNA复制的生物化学	21
§ 2.2 细胞内DNA复制的机制.....	24
§ 2.3 限制酶的作用	32
§ 2.4 细胞中的各种RNA	36
§ 2.5 转录	42
§ 2.6 遗传密码的实验研究	47
§ 2.7 细胞内蛋白质合成的机制	52
第三章 细菌中基因的调控.....	59
§ 3.1 大肠杆菌乳糖转录子的调控	59
§ 3.2 细菌中其他调控体系	65

第二部分 真核基因组

第四章 真核生物的基因.....	70
§ 4.1 染色体和DNA	70
§ 4.2 真核基因 I —rRNA基因和tRNA基因.....	75
§ 4.3 真核基因 II —为蛋白质记码的基因	85
§ 4.4 内部间隔区和隔裂基因	94
第五章 重复DNA	99
§ 5.1 真核DNA中的重复顺序和单一顺序	99
§ 5.2 异染色质	104

§ 5.3	卫星DNA	107
§ 5.4	Y染色体上的重复DNA	117
第六章	真核染色体	122
§ 6.1	组蛋白和非组蛋白	122
§ 6.2	染色体的基本结构单位——核体	128
§ 6.3	染色体的复制	135
§ 6.4	几种低等真核生物的染色体	138
第七章	真核的转录	141
§ 7.1	真核的三种RNA多聚酶	141
§ 7.2	真核mRNA	143
§ 7.3	真核转录的调控, (1)固醇类激素的作用	145
§ 7.4	真核转录的调控, (2)其他因素	150
§ 7.5	异质核内RNA	153

第一章 基因的物质本性

§ 1.1 核酸

在1953年以前，人们对基因的物质本性只能作一些模糊的猜测。经典遗传学早已证明基因在染色体上。把细胞磨碎，用离心方法提取染色体，进行化学分析，发现染色体主要由两种物质组成，一是蛋白质，一是核酸，两者在高等生物（真核生物）染色体中的含量大致各占一半，染色体中除了这两种物质之外几乎没有别的物质。因此当时人们认为基因的物质本性只有三个可能性，或是核酸，或是蛋白质，或是“核酸蛋白质”，即蛋白质与核酸结合所成的复合物。直到1953年，Watson和Crick提出脱氧核糖核酸的分子结构模型，一下子就使大多数生物学家承认基因是核酸。

核酸是在十九世纪就发现的，长期以来，人们不知道核酸在生物体内有什么作用，但对核酸的化学结构则逐渐分析清楚。核酸是一种高分子，这和蛋白质相同。但蛋白质的单体是氨基酸，而核酸的单体则是核苷酸(nucleotide)。每个核苷酸分子由一个五碳糖分子，一个磷酸分子，和一个杂环化合物(嘧啶或嘌呤)分子构成。核酸分两大类，即核糖核酸(RNA)和脱氧核糖核酸(DNA)，是按其核苷酸中五碳糖的不同而区分的。RNA中的五碳糖叫核糖(ribose, $C_5H_{10}O_5$)，DNA中的五碳糖叫脱氧核糖(deoxyribose, $C_5H_{10}O_4$)，后者比前者少一个氧原子。从图1.1.1可见，前者在2'地位的C原子上有个OH基，而后者在这地位是以一个H原子来代替这个OH基。图中注明的1', 2', 3', 4', 5'，是这两种五碳糖中五个C原子的习惯命名。

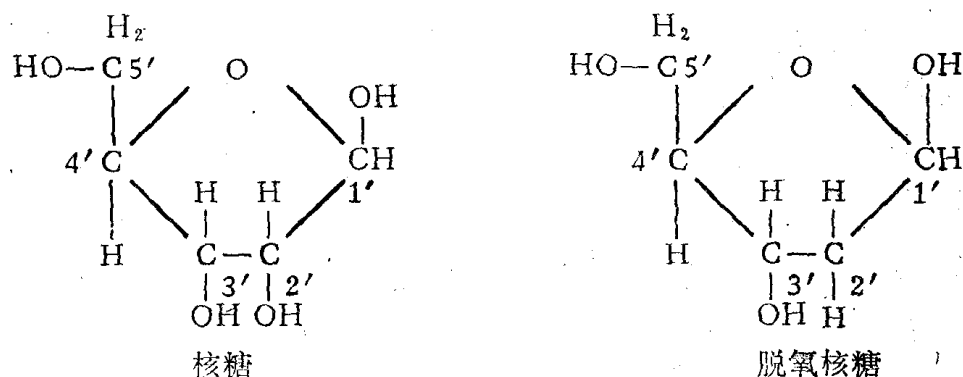


图1.1.1 核糖和脱氧核糖的结构

杂环化合物嘧啶或嘌呤总是结合在1'地位的C上的。这两类杂环化合物的一般结构和原子命名见图1.1.2。

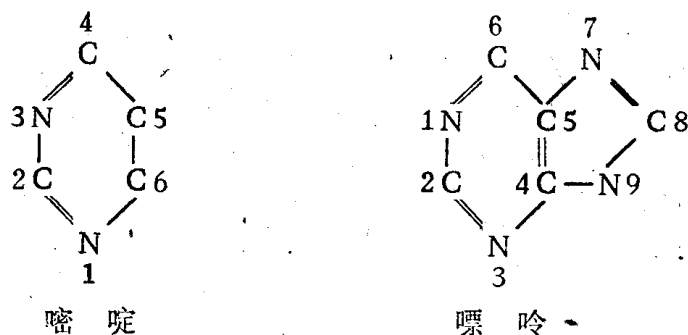


图1.1.2 嘧啶和嘌呤的一般结构

和五碳糖结合时，嘧啶类总是以3位的N和五碳糖中1'位C相接，嘌呤类总是以9位的N和五碳糖中1'位C相接。

DNA中最常见的杂环化合物是两种嘌呤(腺嘌呤, adenine, 鸟嘌呤guanine)和两种嘧啶(胞嘧啶Cytosine, 胸腺嘧啶thymine)。RNA中最常见的也是四种, 即腺嘌呤, 鸟嘌呤, 胞嘧啶, 和尿嘧啶(uracil)。这些嘌呤和嘧啶分子中因为有一 NH_2 和一 NH —基团, 所以带碱性, 因此一般常把他们简单称之为碱基(base)。从图1.1.3.可见, DNA中常见的胸腺嘧啶和RNA中常见的尿嘧啶是极为相似的, 只是前者在5位C上有个甲基, 而后者没有。

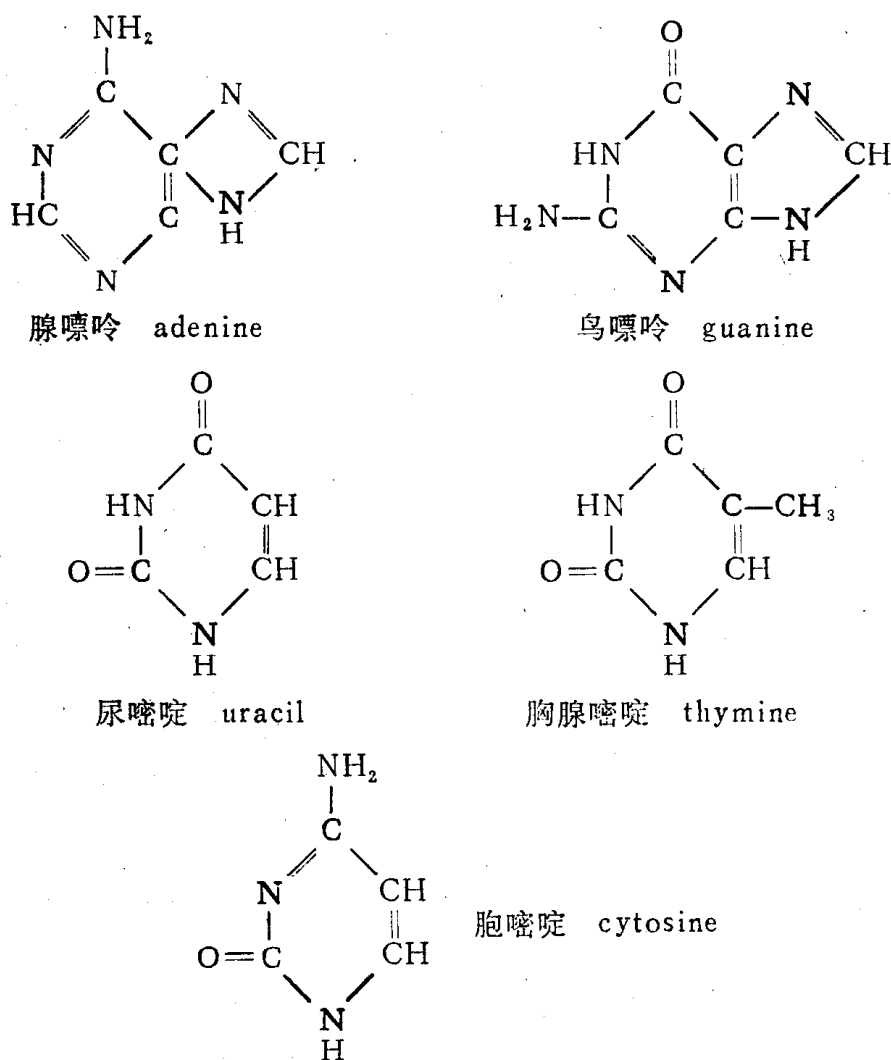
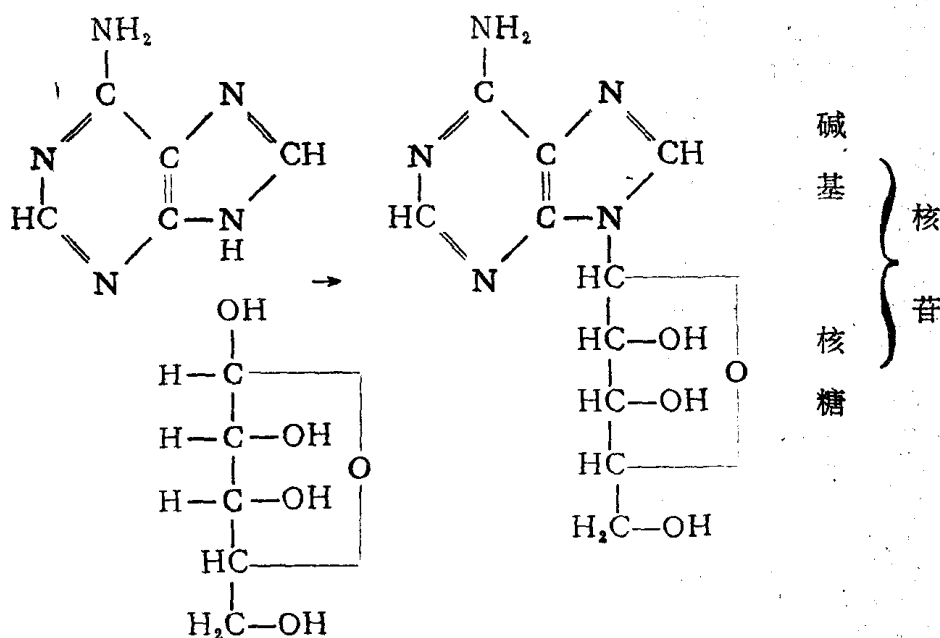


图1.1.3 核酸中常见的碱基

嘌呤的9位N或嘧啶的3位N可以和核糖(或脱氧核糖)的1'位C以N—C糖苷键结合, (脱去一分子H₂O):



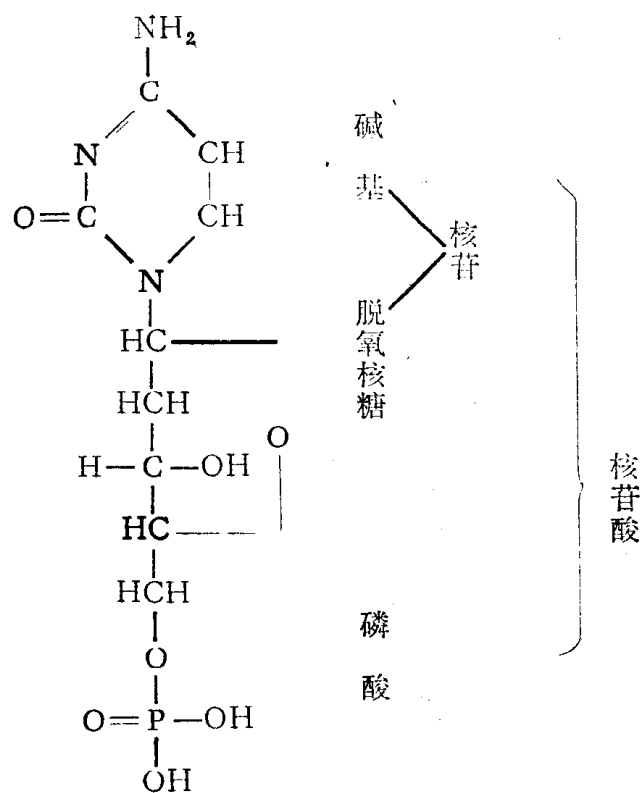
这样一个碱基和一个五碳糖结合, 成为一个核苷 (nucleoside)。DNA和RNA中常见的核苷各四种:

	核 糖	脱氧核糖
腺嘌呤	腺嘌呤核苷	腺嘌呤脱氧核苷
adenine	adenosine	deoxyadenosine
鸟嘌呤	鸟嘌呤核苷	鸟嘌呤脱氧核苷
guanine	guanosine	deoxyguanosine
胞嘧啶	胞嘧啶核苷	胞嘧啶脱氧核苷
cytosine	cytidine	deoxycytidine
尿嘧啶	尿嘧啶核苷	
uracil	uridine	
胸腺嘧啶		胸腺嘧啶脱氧核苷
thymine		deoxythymidine

DNA和RNA中常见的核苷酸。(这些核苷酸也常以“5'—磷酸腺嘌呤核苷”, “5'—磷酸腺嘌呤脱氧核苷”, 等等的方式命名。)

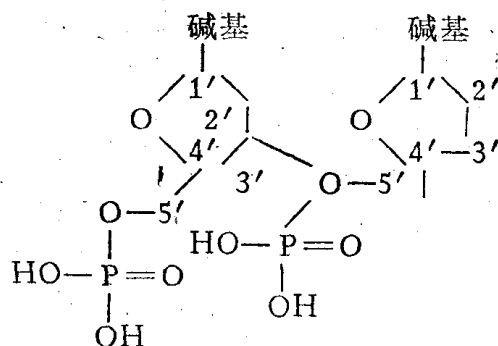
	核 糖	脱氧核糖
腺嘌呤 Aae	腺嘌呤核苷酸 adenylic acia	腺嘌呤脱氧核苷酸 deoxyadenylic acid
鸟嘌呤 Gua	鸟嘌呤核苷酸 guanylic acid	鸟嘌呤脱氧核苷酸 deoxyguanylic acia
胞嘧啶 Cyt	胞嘧啶核苷酸 cytidylic acid	胞嘧啶脱氧核苷酸 deoxycytidylic acid
尿嘧啶 Ura	尿嘧啶核苷酸 uridylic acid	
胸腺嘧啶 Thy		胸腺嘧啶脱氧核苷酸 deoxythymidylic acia

核苷再和一个磷酸结合，就成为核苷酸（nucleotide）。五碳糖上除了1'位 C 已和碱基结合之外，核糖上尚有2',3',5'三个 C 上都有 OH基，都可和磷酸结合，（脱去一分子水，成磷酸酯键）脱氧核糖上则有3'和5'两个 C 都可和磷酸结合，最常见的是磷酸接在5'位上。



如上图就是5'—磷酸胞嘧啶脱氧核苷，也可叫5'—胞嘧啶脱氧核苷酸。

DNA 和 RNA 都是由许多核苷酸聚合而成的“多核苷酸”（Polynucleotide）是链形的高分子。一般讲，可以把核酸看作为5'核苷酸的多聚物。两个5'核苷酸分子结合时，是一个核苷酸5'位上的磷酸和另一个核苷酸上空着的3'位结合，形成磷酸二酯键，称为3'—5'磷酸二酯键。



这样可以一直延长下去。所以一条多核苷酸链，总是一端在5'位有个磷酸，而另一端最后一个核苷酸则在3'位有个空闲的OH基；前者称为多核苷酸的5'端，后者称为3'端。这也就好像一条多肽链有个氨基端和一个羧基端那样。

核酸的分子量一般是很大的，尤其是DNA。例如大肠杆菌的染色体是一个DNA分子，分子量为 2.5×10^9 ，约含有 8×10^6 个核苷酸。 λ 噬菌体的DNA分子量为 3.1×10^7 ，约 9×10^4 个核苷酸，已经算是较小的噬菌体了。

§ 1.2 DNA分子结构的双螺旋模型

在真核细胞中，DNA主要在细胞核内的染色体上，而RNA则主要在细胞质内。在1953年以前，虽然已有一些实验说明基因是DNA，但由于人们对DNA分子结构还不清楚，不能设想DNA怎么能具有遗传物质的功能。J.D.Watson和F.H.C.Crick在1953年发表DNA分子结构的双螺旋模型，（即Watson-Crick模型），就解决了这个问题。他们的模型主要是根据那几年对DNA所作的X线衍射研究和化学成分分析和结构研究而提出的。

根据碳原子和氧原子的已知键角，人们早以知道象多核苷酸这样的长链在空间中必是螺旋形。在1953年以前，也有些人提出DNA分子结构的一些模型，也都以螺旋形长链为出发

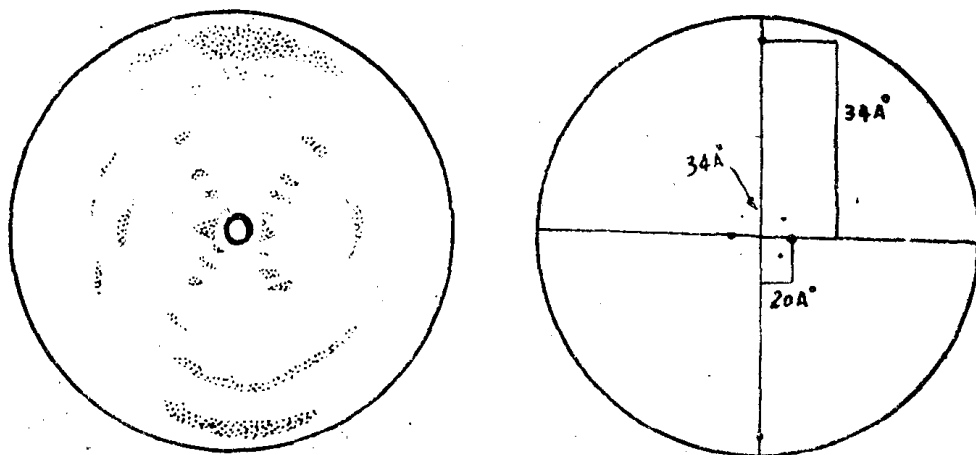


图1.2.1 左，DNA纤维衍射图 右，左图的解释。

（注1）分析X光衍射照片所用的Bragg公式是

$2d \sin \theta = n\lambda$ ， θ 是衍射角；衍射点至照片中心的距离/被测物体至照片的距离即为 $\tan \theta$ ， d 是物体中存在的重复间距， λ 是X线波长， n 是1,2,3...各个正整数。

在 $n\lambda$ 恒定时， d 越大则 θ 越小。所以图1.2.1中，相当于3.4 Å的点很远，而20 Å一点较近。

点。但这些模型和化学上的已知事实不很符合，不能令人满意。后来 M.H.F. Wilkins, Rosalind Franklin 等用小牛胸腺 DNA 作纤维衍射，得到较好的照片，（图 1.2.1），为 Watson 和 Crick 提出他们的模型打下了基础。

这些照片上，子午线（垂直线）上有个很强的衍射点，相当于 $3.4 \text{ \AA}$ 的重复间距（注1），人们早已设想这是两个邻近核苷酸之间的间距，（也就是邻近两个碱基之间的间距）。赤道线（水平线）上的衍射点相当于 $20 \text{ \AA}$ 间距，Watson 和 Crick 很自然地把这数值想象为螺旋的直径。

在 $20 \text{ \AA} \times 20 \text{ \AA} \times 3.4 \text{ \AA}$ 这样一个单位空间内，测得电子密度相当于两个核苷酸分子所具有的电子。这使他们设想 DNA 分子是由两条螺旋形多核苷酸链互绕而成。（图 1.2.2）。

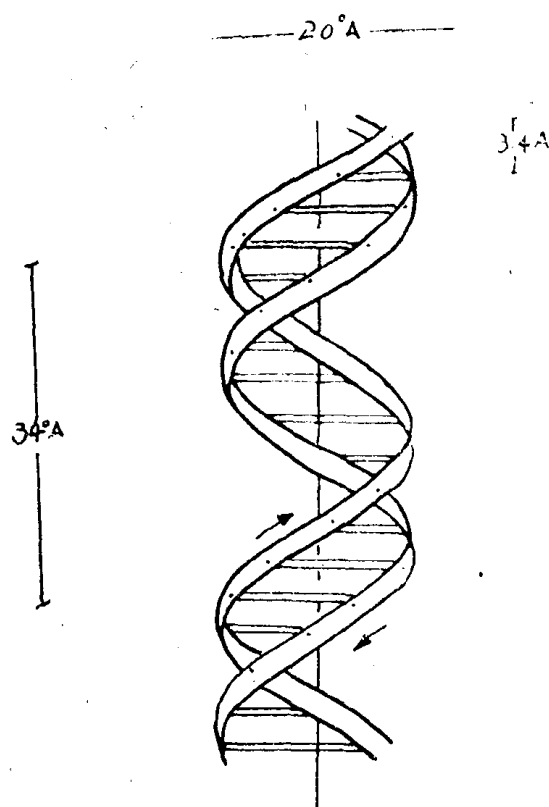


图 1.2.2 DNA 分子双螺旋模型
Watson-Crick 原图。

根据结构化学原理考虑，这两条螺旋在空间上以同一方向旋绕，（可以说都是右旋），但两条链的分子方向（即核苷酸单体结合方向）则相反，设以一链为 $5'$ 端到 $3'$ 端，则另一链为 $3'$ 端到 $5'$ 端。（图 1.2.3）。

为了解释衍射照片中那一系列相当于 $34 \text{ \AA}$ 的层线（图 1.2.1），他们设想同一链上两个邻近核苷酸之间的旋转角为 36° ，（这和结构化学理论很符合），因此连续十个核苷酸就旋转 360° ，即一个整圈，这是螺旋的旋绕周期，其长度为 $3.4 \text{ \AA} \times 10 = 34 \text{ \AA}$ 。这模型是设想多核苷酸链的磷酸——五碳糖基线是位于双螺旋的外周，而碱基是向内的，即向着螺旋的轴心。嘌呤和嘧啶本来都是平面的分子，现在就设想碱基平面都和双螺旋长轴垂直。

这模型最主要之点是设想两条链上的碱基两两成对，一个碱基和相对碱基之间靠氢键结合，由此保持整个结构的稳定性。（图 1.2.3）。

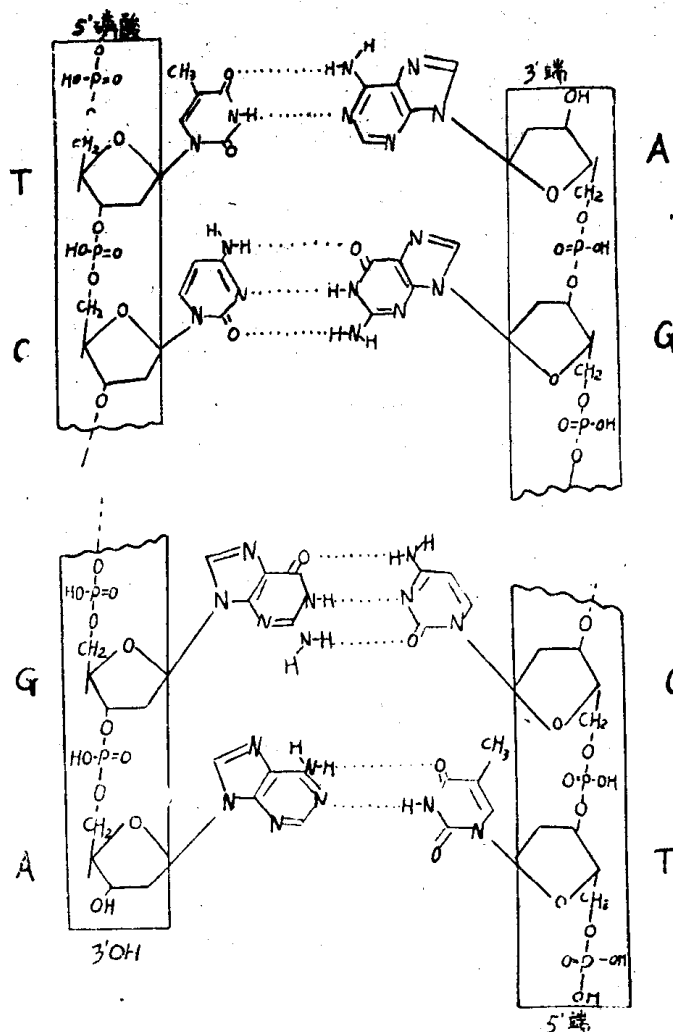


图1.2.3 DNA 双链的碱基配对。点线表示氢键。淡蓝色部分表示磷酸—脱氧核糖基线。

由于以上这些前提，就发现：并非随便哪两个碱基都能互相配对，碱基配对时有很大的限制。首先，嘌呤只能和嘧啶配，反之亦然。这是因为，双螺旋的直径是固定的，如果同一平面上两个碱基都是嘌呤，则地位不够；如果两个都是嘧啶，则彼此距离太远，超过氢键的作用距离，不能形成氢键。

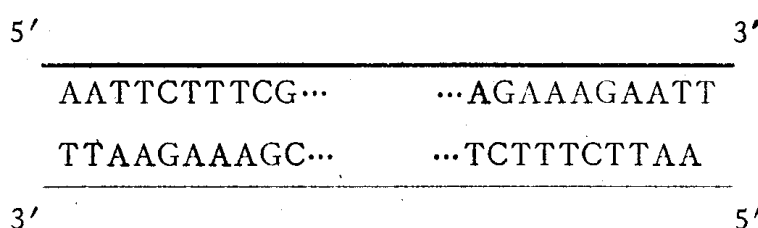
由于氢键结合的要求，发现腺嘌呤（简称为A）只能和胸腺嘧啶（T）配对，鸟嘌呤（G）只能和胞嘧啶（C）配对。碱基对平面上下相叠，象一叠玻璃板那样，彼此并无干扰和限制。所以以一条链来说，核苷酸前后顺序随便怎样都可以；但一条链上的顺序决定后，另一条链上的核苷酸顺序就自动决定。两条链是严格互补的。

这个模型立刻解释了DNA化学分析的历来结果。各种不同生物的DNA中，A的含量总是等于T，G总是等于C，但 $(A+T)/(G+C)$ 这比数则不同的生物不同；

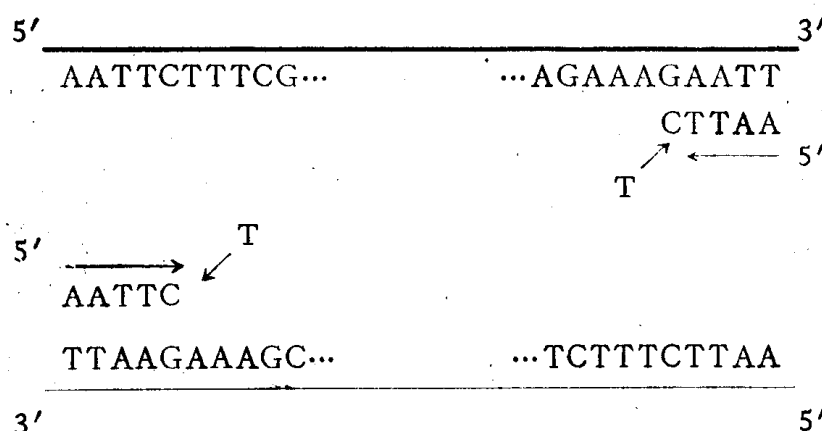
	A	G	C	T	$\frac{A+T}{G+C}$
大肠杆菌 <i>E. Coli</i>	26.0%	24.9%	25.2%	23.9%	1.00
粘质沙雷氏菌 <i>Serratia marcescens</i>	20.7	27.2	31.9	20.1	1.69
结核杆菌 <i>Mb. tuberculosis</i>	19.3	28.2	34.9	17.5	0.58
面包酵母菌 <i>Sacch cerevisiae</i>	31.7	18.3	17.4	32.6	0.80
海胆 <i>Psammechinus miliaris</i> (精虫)	32.6	17.8	17.8	31.9	1.81
海胆 <i>Arbacia lixula</i> (精虫)	32.6	19.1	19.2	30.5	1.61
蝗虫 <i>Locusta migratoria</i>	29.3	20.5	20.7	29.3	1.41
蛙(精虫)	29.7	20.8	20.4	29.1	1.43
槽白鱼(精虫)	27.9	19.5	21.5	28.2	1.28
鸽(红血球)	28.7	22.0	21.3	27.9	1.31
鸡(红血球)	28.8	20.5	21.5	29.2	1.38

所以 Watson-Crick 模型和过去各种实验结果都很符合。最使生物学家高兴的是碱基互补配对原理足以解释基因的复制；多核苷酸链上碱基前后顺序不受限制足以解释基因的极大多样性。基因复制可以这样设想：设原来一个基因，共 207 对核苷酸，如下：

(只写出两端各十对)：



在复制时先是双链拆开，然后附近可得的核苷酸单体一个个按碱基配对原则和原有两条链相配，这样就合成新的链。最后就成为两个基因：



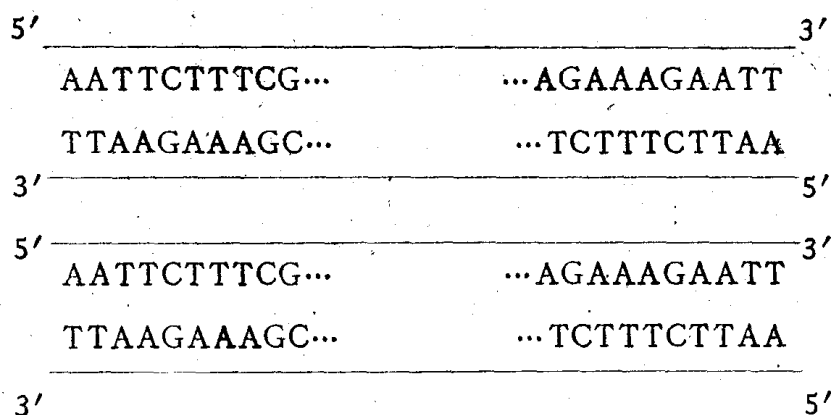


图1.2.4. DNA样板复制示意图

原来的链可说是新链合成的样板 (template)。

DNA 分子靠碱基互补配对形成双链，并能进行样板复制，这些原理为后来的许多实验所完全证实。现在讲一些较有兴趣的实验。

在1960年的时候，J.Marmur 等发现，DNA 在食盐溶液 (0.15—0.2M) 中水浴加温到接近100℃时，会突然发生变性 (denaturation)：260nm 紫外线吸收突然增加40%，溶液粘度突然降低。这是DNA双链拆开，成为单链。(在pH的碱性溶液中也会这样。) 发生变性的DNA，如果在水浴中很缓慢地降温，(从90℃降低到25℃约花两小时)，又会恢复原来性状，称为复性 (renaturation)。这样用缓慢降温的办法使拆开的链重新结合成双链，称为退火淬合 (annealing)。

由双链拆开所得的两条链，可用甲基白蛋白硅藻土 (MAK) 柱层析方法分开，分别回收。回收所得的两条链往往一条比重较高，称为重链 (H)，一条比重较低，称为轻链 (L)。分别分析重链和轻链中的碱基成分，完全符合 A—T，G—C 配对的模型。

双链DNA在电子显微镜下看起来时是条比较硬直的，直径20 Å的条状，单链DNA则往往很蜷曲，看不清楚。把两个来源不同，但遗传上有一定相似程度的DNA，拆成两条链，把甲的H和乙的L放在一起，进行淬合，也能成功，在一定部分的链段上形成双链。这个方法有个浪漫的名子，叫做“分子杂交”，(molecular hybridization)。图1.2.5是大肠杆菌的λ噬菌体的两个品系，其DNA进行分子杂交后在电子显微镜下所见的图象，可以看到

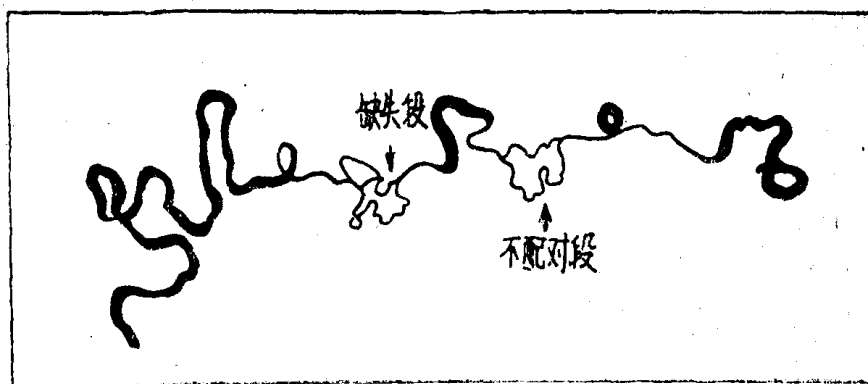


图1.2.5 λ噬菌体两个不同品系的DNA的“分子杂交”，电子显微镜图。引自Westmoweland等，1969。

两条链大部分配成双链，但有一个缺失段和一个不配对段，和遗传学分析完全符合。这很类似经典细胞遗传学上“相对染色体”配对的情况。（但相对染色体配对是双链 DNA 和双链 DNA 配对，而这里讲的都是单链和单链配成双链，两者不可混淆）。

后者 DNA 中核苷酸顺序的测定方法建立，碱基互补配对的原则就得到直接的证实。在 1974 年测得大肠杆菌中乳糖基因 (lac) 的开头部分（所谓操纵区 operator），共 21 碱基对如下：

5' AATTGTCAGCGGATAACAATT 3'
3' TTAACAGTCGCCTATTGTTAA 5'

§ 1.3 遗传密码

基因是个生物学概念。一般说，一个基因决定一个蛋白质。

基因决定蛋白质，其基本的规律叫做“氨基酸密码”（the amino-acids code），也有人称之为“遗传密码”（the genetic code）；其要点就是三个核苷酸决定一个氨基酸。核苷酸有 4 种，3 个核苷酸的排列共有 $4^3 = 64$ 种，这叫 64 种密码子（codon）。氨基酸密码在 1966 年已由 M. Nirenberg, H. G. Khorana 等用各种实验方法全部解出如下，（见 § 2.6 节）：

UUU 苯丙	UCU 丝	UAU 酪	UGU 半胱
UUC "	UCC "	UAC "	UGC "
UUA 亮	UCA "	UAA 	UGA 
UUG "	UCG "	UAG 	UGG 色
CUU 亮	CCU 脯	CAU 组	CGU 精
CUC "	CCC "	CAC "	CGC "
CUA "	CCA "	CAA 谷胺	CGA "
CUG "	CCG "	CAG "	CGG "
AUU 异亮	ACU 苏	AAU 门胺	AGU 丝
AUC "	ACC "	AAC "	AGC "
AUA "	ACA "	AAA 赖	AGA 精
AUG 甲硫	ACG "	AAG "	AGG "
GUU 缬	GCU 丙	GAU 门冬	GGU 甘
GUC "	GCC "	GAC "	GGC "
GUA "	GCA "	GAA 谷	GGA "
GUG "	GCG "	GAG "	GGG "

 是终止密码子。

遗传密码是整个生物界统一通行的。为便于使用，再把明密码对照列表如下：

ala 丙氨酸 : GCU, GCC, GCA, GCG
 arg 精氨酸 : CGU, CGC, CGA, CGG, AGA, AGG。
 asn 门冬酰胺: AAU, AAC。
 asp 门冬氨酸: GAU, GAC。
 cys 半胱氨酸: UGU, UGC。
 gln 谷氨酰胺: CAA, CAG。
 glu 谷氨酸 : GAA, GAG。
 gly 甘氨酸 : GGU, GGC, GGA, GGG。
 his 组氨酸 : CAU, CAC。
 ile 异亮氨酸: AUU, AUC, AUA。
 leu 亮氨酸 : UUA, UUG, CUU, CUC, CUA, CUG。
 lys 赖氨酸 : AAA, AAG。
 met 甲硫氨酸: AUG。
 phe 苯丙氨酸: UUU, UUC。
 pro 脯氨酸 : CCU, CCC, CCA, CCG。
 ser 丝氨酸 : UCU, UCC, UCA, UCG, AGU, AGC。
 thr 苏氨酸 : ACU, ACC, ACA, ACG。
 try 色氨酸 : UGG。
 tyr 酪氨酸 : UAU, UAC。
 val 缬氨酸 : GUU, GUC, GUA, GUG。
 (终止) : UAA, UAG, UGA。

氨基酸密码表一般用RNA“字母”，用U而不用T；RNA中的U相当于DNA中的T。这是因为直接决定蛋白质的是RNA而不是DNA。如果要把密码改写成DNA字母，只要把上面两表中的U改成T就行。

细胞中所有的RNA都是DNA决定的，（只有几种RNA病毒是例外，见§1.6节）。DNA按碱基互补配对原则而复制，也按碱基互补配对原则而决定RNA。这后一种过程，称为“转录”，（transcription），此时DNA中的A是和RNA中的U互补。所以转录过程和DNA复制过程有相似之处。只是RNA分子一般是单链的，双链DNA一般只有其中一链是RNA的样板，其另一链则也可说是RNA的“原稿”。转录时RNA在3'端不断合成，同时从5'端开始从样板链上脱下。

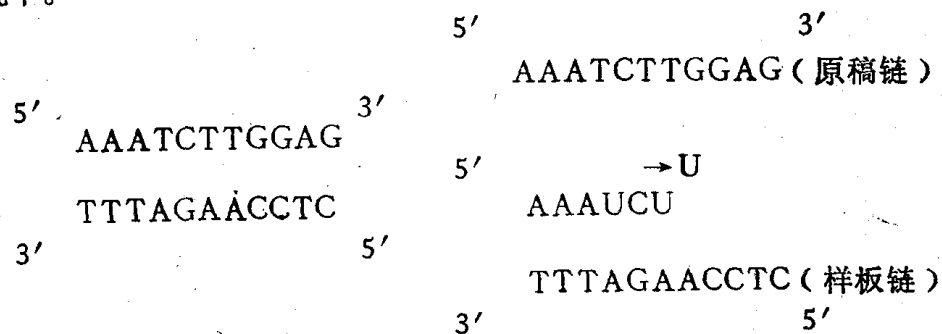


图1.3.1 转录。绿线表示DNA，红线表示RNA

所以分子生物学家在实验室内也可以把某种RNA和它原来的样板DNA做“分子杂交”。

蛋白质是由RNA决定的，但并非细胞中所有的RNA都有决定蛋白质的作用，有些RNA分子并不起决定蛋白质的作用。起决定蛋白质作用的RNA特称为信使RNA（messenger RNA，简写mRNA）。mRNA按遗传密码而决定蛋白质中的氨基酸顺序，这个过程叫做“翻译”，（translation），就是把mRNA中每个核苷酸“字母”翻译成蛋白质中一个氨基酸“字母”。

翻译过程的实现，要依靠另一种分子量很小的RNA，称为转运RNA（transfer RNA，简写tRNA）。每个tRNA分子只有几十个核苷酸。每种生物都各有几十种tRNA分子。每个tRNA分子在其3'端接一个一定的氨基酸，而在其分子的失约半当中的地方有三个核苷酸组成一个“反密码子”（anticodon），按碱基互补配对原则和mRNA中的一个密码子配对。

tRNA既然是一种RNA，当然也是从DNA转录而来的。但各种tRNA分子，其初级结构（即核苷酸顺序）中只有中间三个核苷酸组成“反密码子”，其他核苷酸并不起“氨基酸密码”的作用。换句话说，tRNA不起mRNA的作用，不是mRNA，不决定蛋白质；它们是翻译的工具而不是被翻译的“原文”。

密码表中有三个密码子是“终止密码子”（termination codon）。一般生物体内没有一种tRNA具有与这三个密码子配对的反密码子，因此翻译过程进行到这些终止密码子的地方就必然终止。

翻译过程可以在体外实验中进行，就是说用“无细胞体系”（cell-free system）来进行蛋白质的生物合成。由于遗传密码是整个生物界统一的，所以可以用一种生物的翻译机器来翻译另一种生物的mRNA。所谓“翻译机器”，除了那几十种tRNA之外，还要各种酶，以及一种叫“核糖体”（ribosome）的细胞器（见§ 2.4节）。取一些小麦胚，把小麦胚细胞磨碎，提取其核糖体和各种tRNA和各种酶，以及各种氨基酸，加入大鼠前胰岛素的mRNA，就可在试管中合成前胰岛素。

§ 1.4 基因和基因组

大肠杆菌的全部遗传物质是一个环形的双链DNA分子，其分子量为 2.5×10^9 dalton（注1），含 4×10^6 碱基对（base pairs，简写bp）。按Watson-Crick模型，碱基对间距为 $3.4 \text{ \AA}$ 计算，这个分子的全长为 $3.4 \text{ \AA} \times 4 \times 10^6 = 1.36 \text{ mm}$ ，但在细菌细胞中则围绕成直径不到 $1 \mu\text{m}$ 的一团。大肠杆菌的基因全都直线地排列在这环形DNA上。这个环形DNA分子就叫做大肠杆菌的“基因组”（genome）。基因组的定义是一个细胞内（或一个病毒颗粒内）的全部遗传物质，一般就是指DNA。

有些病毒的基因组很小，例如 $\phi \times 174$ 噬菌体，它是以大肠杆菌为宿主的。在成熟的、具有感染能力的噬菌体中，在蛋白质外壳内有个单链的环形DNA，共5,375个核苷酸，这就是它的基因组。在感染大肠杆菌时，这单链环形DNA进入细菌，很快就按碱基配对原则合成双链环形DNA。一般把噬菌体颗粒中的单链称为+链，进入细菌后合成的互补链为-链。转录是以一链为模板的，所以转录出来的RNA，其核苷酸顺序和+链相同。

（注1） $1\text{g} = 6.022 \times 10^{23} \text{ dalton}$ 。（ 6.022×10^{23} 是阿伏加德罗常数。）

$$\therefore 1 \text{ dalton} = (6.022 \times 10^{23})^{-1} = 1.66 \times 10^{-24} \text{ g}$$

如果分子量是 2.5×10^9 ，则克分子量是 $2.5 \times 10^9 \text{ g}$ ，一个分子就重 $2.5 \times 10^9 \text{ dalton}$ 。

原来根据遗传学研究, 已知 $\phi \times 174$ 有 9 个基因, 决定 9 种蛋白质, 基因的排列次序是 A—B—C—D—E—J—F—G—H, (图1.4.1) Sanger等(1977)把+链的5,375个核苷酸的顺序全部测定。8个基因的起迄点都已确切测定, 只有基因C的起迄点尚未正式测定, 但亦可作较合理的推测(见下表)。

基 因	起迄点(核苷酸序数)	全长(核苷酸数)	说 明
D	No. 390—No. 845	456	
E	568—840	273	包含在D内
J	848—961	114	
F	1001—2275	1,275	
G	2387—2911	525	
H	2923—3906	984	
A	3973—5375, 1—133	1,536	包含在A内未
B	5064—5375, 1—48	360	正式测定, 是
C	134—391	258	推测

由于链是环形的, 所以核苷酸的命序带有随意性。现在所采用的是以某种核酸酶(限制酶Pst I)对这个DNA的切断点开始, 作为 No.1, 按3', 5'磷酸二酯键的方向顺数, 即每个磷酸接在前一个核苷酸的3'和后一个核苷酸的5'上。也就是说, 如果在 No.5375与No.1之间切断, 则 No.1是5端, No.5375是3'端。

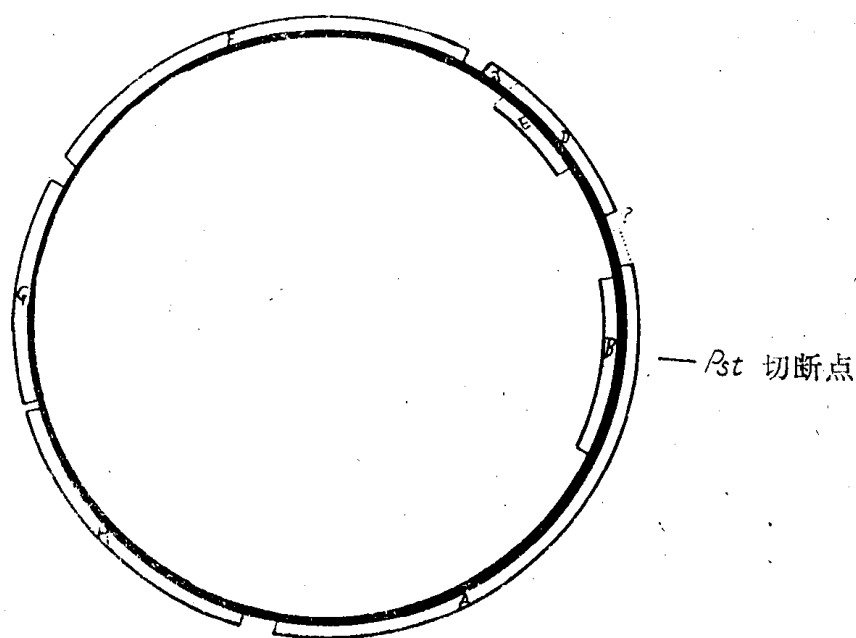


图1.4.1 $\phi \times 174$ 噬菌体的基因组。+链的3', 5'磷酸二酯键结合是逆时钟方向。

(自Sanger, F.等1977, Nature, 265, 687—695)

这个工作说明的问题是极多的, 现在只讲三点:

(1) 有两个基因是包含在其他基因之内的, 即基因B包含在基因A之内, 基因E包含在基因D之内。这种现象已知不仅为 $\phi \times 174$ 所特有。这种现象之所以可能, 是由于一种特殊机制, 称为“密码移框”, (将在§1.5节中说明)。