



# 第三讲 生物信息的传递

## (上) 从**DNA**到**RNA**

基因作为唯一能够自主复制、永久存在的单位，其生物学功能是以蛋白质的形式表达出来的。**DNA**序列是遗传信息的贮存者，它通过自主复制得到永存，并通过转录生成信使**RNA**，翻译生成蛋白质的过程来控制生命现象。



基因表达包括转录（**transcription**）和翻译（**translation**）两个阶段。

转录是指拷贝出一条与**DNA**链序列完全相同（除了**T→U**之外）的**RNA**单链的过程，是基因表达的核心步骤。

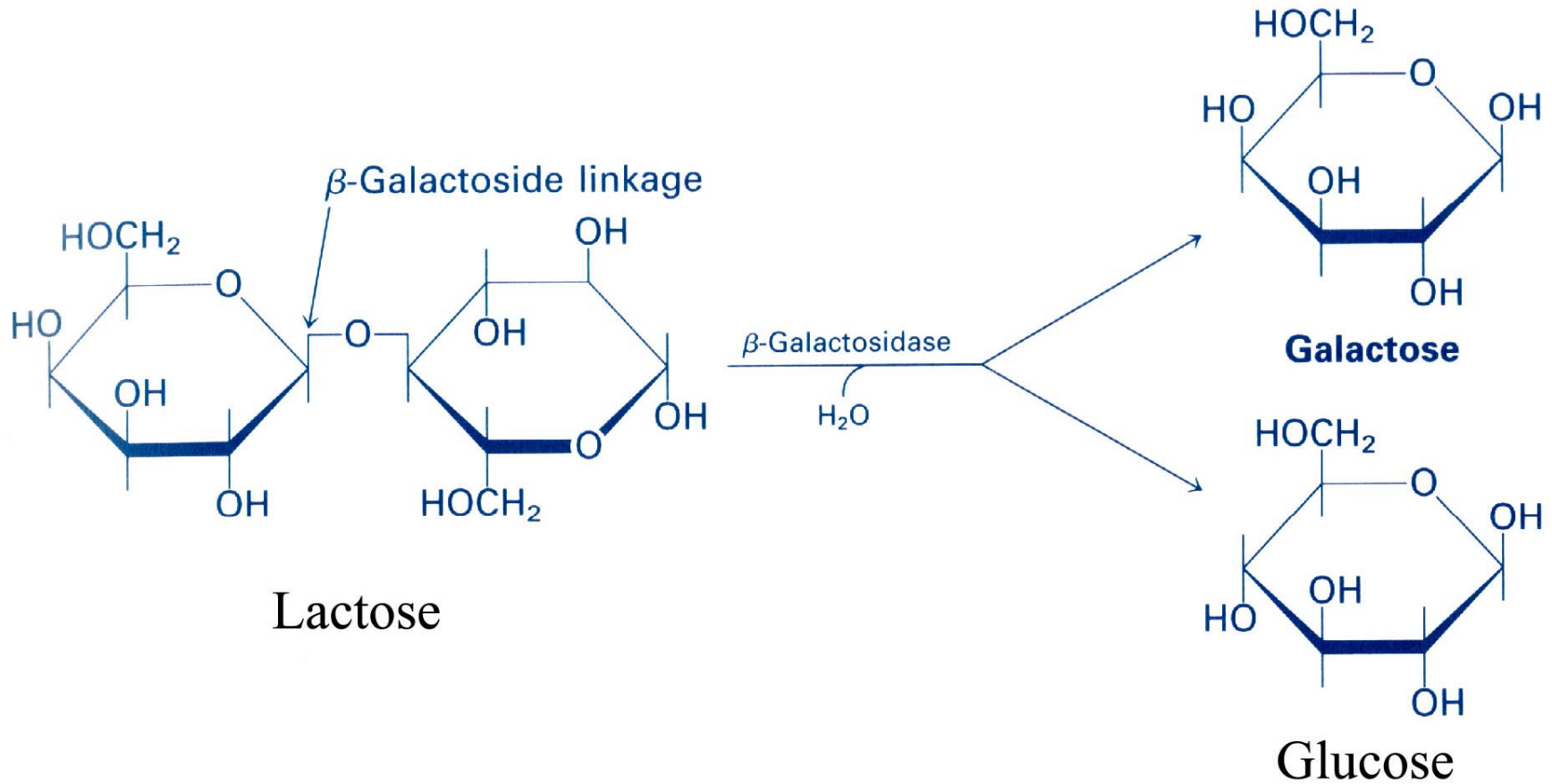
翻译是指以新生的**mRNA**为模板，把核苷酸三联遗传密码子翻译成氨基酸序列、合成多肽链的过程，是基因表达的最终目的。



**Francois Jacob (Left), Jacques Monod (Center) & Andre Lwoff (Right)**

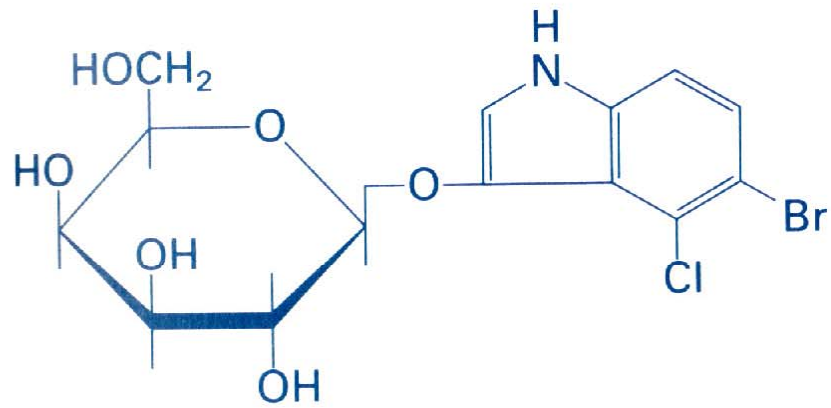


(a)



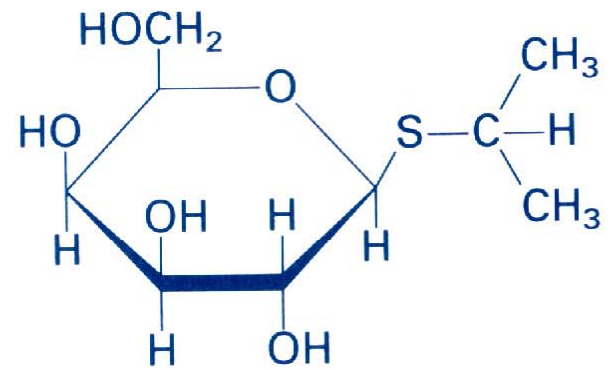


(b)



X-gal

(c)



Isopropyl- $\beta$ -D-thiogalactoside  
(IPTG)



**DNA**分子中的核苷酸排列顺序不但决定了胞内所有**RNA**及蛋白质的基本结构，还通过蛋白质（酶）的功能间接控制了细胞内全部有效成份的生产、运转和功能发挥。



与mRNA序列相同的那条DNA链是编码链（**coding strand**）或称有意义链（**sense strand**），另一条根据碱基互补原则指导mRNA合成的DNA链则被称为模板链（**template strand**）或称反义链（**antisense strand**）。

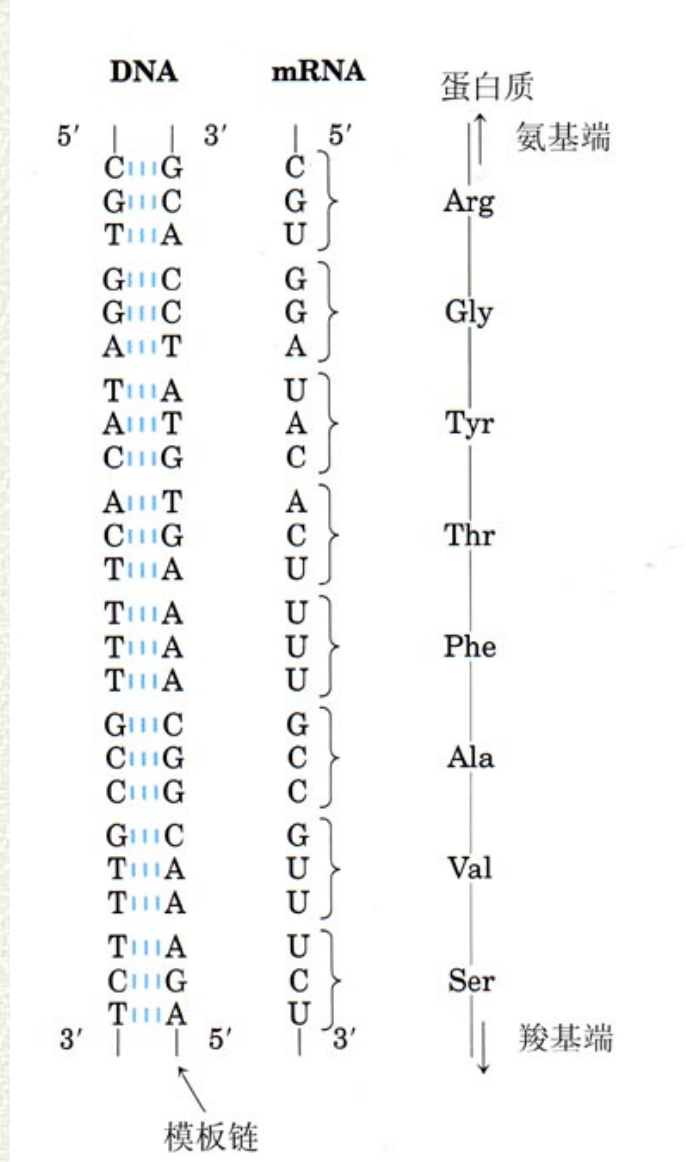


图3-1 DNA模板与mRNA分子及多肽链之间存在共线性关系。



**DNA**和**RNA**虽然很相似，只有**T**或**U**及核糖的第二位碳原子上有所不同，但它们的生物学活性却很不同。

**RNA**主要以单链形式存在于生物体内，其高级结构很复杂，它们既担负着贮藏及转移遗传信息的功能，又能作为核酶直接在细胞内发挥代谢功能。



因为只有**mRNA**所携带的遗传信息才被用来指导蛋白质生物合成，所以人们一般用**U、C、A、G**这4种核苷酸而不是**T、C、A、G**的组合来表示遗传性状。



生物体内拥有三类RNA，即：

编码特定蛋白质序列的mRNA；

能特异性解读mRNA 中的遗传信息并将其转化成相应氨基酸后加入多肽链中的tRNA；

直接参与核糖体中蛋白质合成的rRNA。



## 3.1 RNA的转录

### 3.1.1 转录的基本过程

无论是原核还是真核细胞，转录的基本过程都包括：

模板识别  
转录起始  
通过启动子  
转录的延伸和终止。

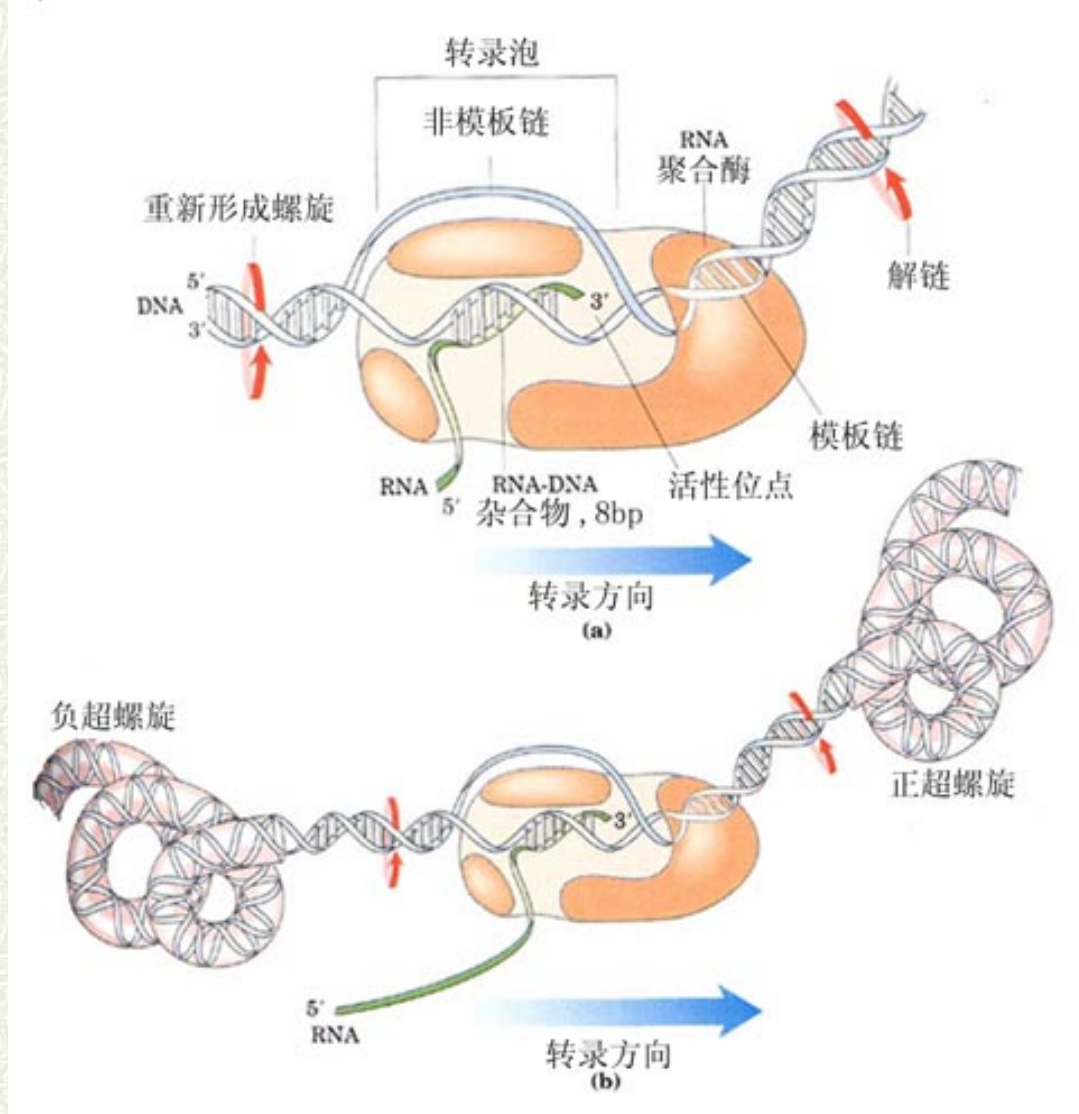


图3-2 大肠杆菌中依赖于DNA的RNA转录过程图示。转录泡中单链DNA的长度约在17bp左右，被聚合酶复合物所保护的DNA序列约为35bp左右。



模板识别阶段主要指**RNA**聚合酶与启动子**DNA**双链相互作用并与之相结合的过程。转录起始前，启动子附近的**DNA**双链分开形成转录泡以促使底物核糖核苷酸与模板**DNA**的碱基配对。

转录起始就是**RNA**链上第一个核苷酸键的产生。



转录起始后直到形成**9**个核苷酸短链是通过启动子阶段，此时**RNA**聚合酶一直处于启动子区，新生的**RNA**链与**DNA**模板链的结合不够牢固，很容易从**DNA**链上掉下来并导致转录重新开始。



一旦**RNA**聚合酶成功地合成9个以上核苷酸并离开启动子区，转录就进入正常的延伸阶段。所以，通过启动子的时间代表一个启动子的强弱。

一般说来，通过启动子的时间越短，该基因转录起始的频率也越高。



**RNA聚合酶离开启动子，沿DNA链移动并使新生RNA链不断伸长的过程就是转录的延伸。**

**大肠杆菌RNA聚合酶的活性一般为每秒50-90个核苷酸。随着RNA聚合酶的移动，DNA双螺旋持续解开，暴露出新的单链DNA模板，新生RNA链的3'末端不断延伸，在解链区形成RNA-DNA杂合物。**



真核生物**RNA**聚合酶不能直接识别基因的启动子区，需要一些被称为转录调控因子的辅助蛋白质按特定顺序结合于启动子上，**RNA**聚合酶才能与之相结合并形成复杂的前起始复合物（**preinitiation transcription complex, PIC**），以保证有效地起始转录。



**37°C时，转录生成mRNA的速度大约是每分钟2500个核苷酸，即每秒钟合成14个密码子，而蛋白质合成的速度大约是每秒钟15个氨基酸。正常情况下，从一个基因开始表达到细胞中出现其mRNA的间隔约为2.5分钟，而再过半分钟就能在细胞内测到相应的蛋白质。**



- 1、什么是DNA的半保留复制？
- 2、DNA复制的生物学意义是什么？
- 3、原核生物中主要有那几种DNA聚合酶？请说出它们的主要功能？
- 4、说出人类细胞中最大的和最小的染色体，它们各有多少bp。



- 5、说出DNA作为遗传物质的最基本特征。
- 6、有什么证据说明染色体可能是由核小体组成的？
- 7、大肠杆菌DNA完全伸展时有多长？
- 8、重叠基因是怎样发现的，是从哪种生物里首次发现的？



## 3. 1. 2 转录机器的主要成分

### 1. RNA聚合酶

**RNA聚合酶**是转录过程中最关键的酶，主要以双链**DNA**为模板，以4种核苷三磷酸作为活性前体，并以**Mg<sup>2+</sup>/Mn<sup>2+</sup>**为辅助因子，催化**RNA**链的起始、延伸和终止，它不需要任何引物，催化生成的产物是与**DNA**模板链相互补的**RNA**。



**RNA或RNA-DNA双链杂合体不能作为模板。**

原核和真核生物的**RNA**聚合酶虽然都能催化**RNA**的合成，但在其分子组成、种类和生化特性上各有特色。



大多数原核生物**RNA**聚合酶的组成是相同的，大肠杆菌**RNA**聚合酶由2个  $\alpha$  亚基、一个  $\beta$  亚基、一个  $\beta'$  亚基和一个  $\omega$  亚基组成，称为核心酶。加上一个  $\sigma$  亚基后则成为聚合酶全酶（**holoenzyme**），相对分子质量为  $4.65 \times 10^5$ 。

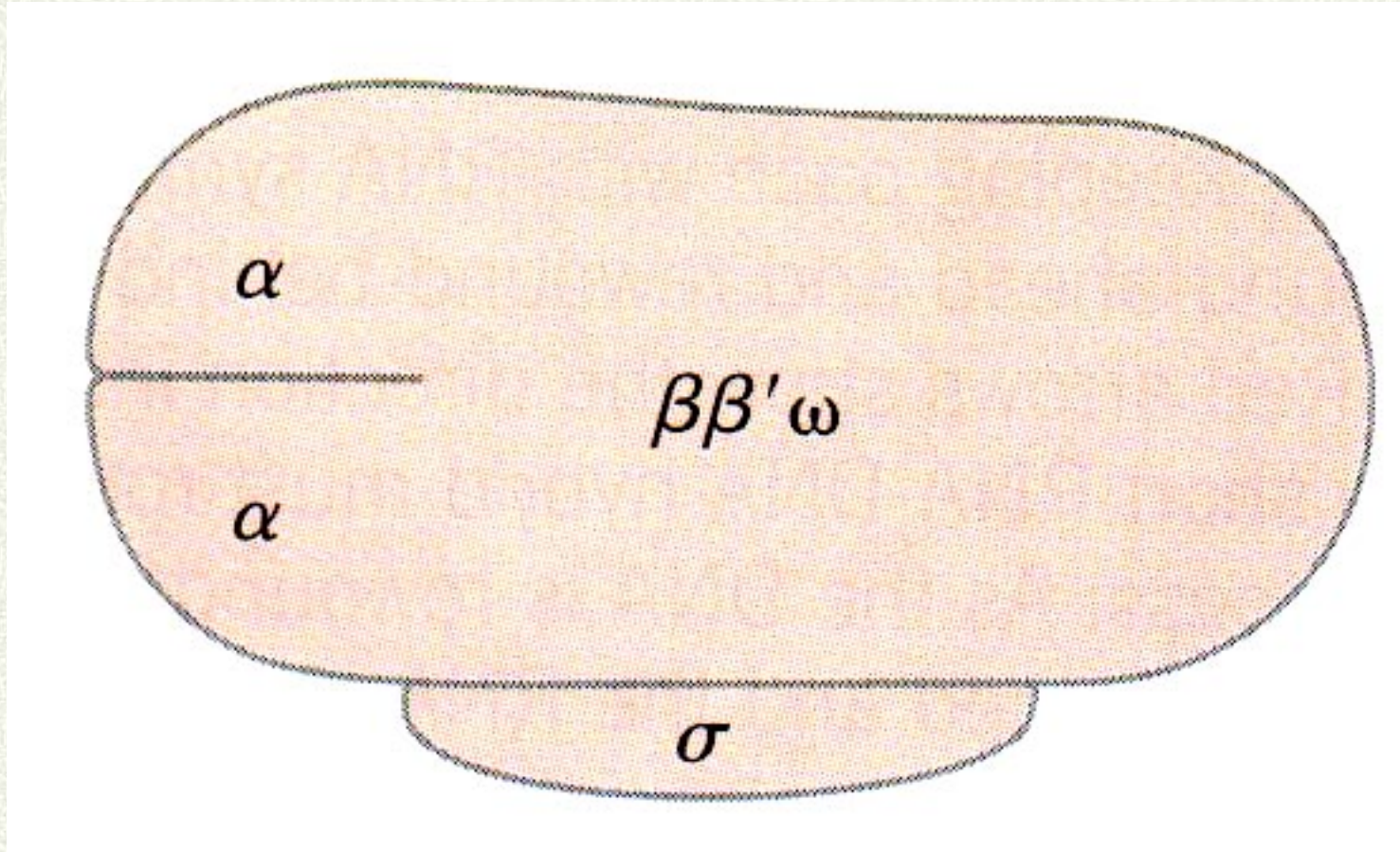
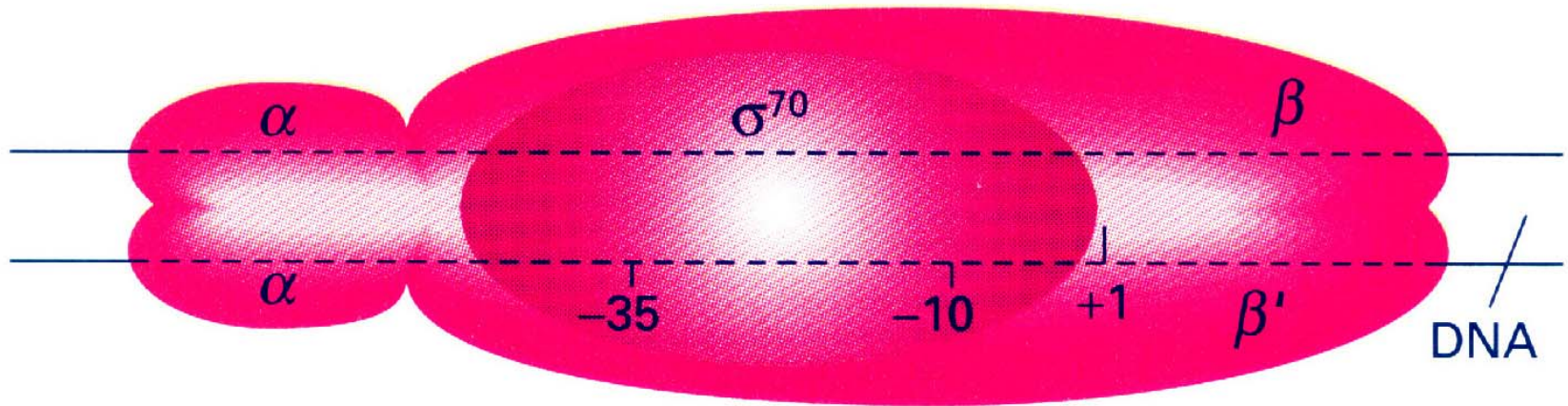


图3-3 大肠杆菌RNA聚合酶的主要成分与功能分析



← upstream, downstream →



Direction of transcription





### 表3-1 大肠杆菌RNA聚合酶的组成分析

亚基	基因	相对分子量	亚基数	组分	功能
$\alpha$	rpoA	$3.65 \times 10^4$	2	核心酶	核心酶组装，启动子识别。
$\beta$	rpoB	$1.51 \times 10^5$	1	核心酶	$\beta$ 和 $\beta'$ 共同形成RNA合成的活性中心。
$\beta'$	rpoC	$1.55 \times 10^5$	1	核心酶	
	?	$11 \times 10^4$	1	核心酶	未知
$\sigma$	rpoD	$7.0 \times 10^4$	1	$\sigma$ 因子	存在多种 $\sigma$ 因子，用于识别不同的启动子



$\alpha$  亚基可能与核心酶的组装及启动子识别有关，并参与**RNA**聚合酶和部分调节因子的相互作用。

**T4**噬菌体感染大肠杆菌后对  $\alpha$  亚基的一个精氨酸残基进行**ADP**糖基化修饰，造成**RNA**聚合酶全酶对启动子亲和力降低。



由  $\beta$  和  $\beta'$  亚基组成了聚合酶的催化中心，它们在序列上与真核生物RNA聚合酶的两个大亚基有同源性。 $\beta$  亚基能与模板DNA、新生RNA链及核苷酸底物相结合。



σ 因子的作用是负责模板链的选择和转录的起始，它是酶的别构效应物，使酶专一性识别模板上的启动子。核心酶在T7噬菌体DNA上约有1300个结合位点，平均结合常数为 $2 \times 10^{11}$ 。



σ 因子可以极大地提高**RNA**聚合酶对启动子区**DNA**序列的亲合力，酶底结合常数提高 $10^3$ 倍，酶底复合物的半衰期可达数小时甚至数十小时。

σ 因子还能使**RNA**聚合酶与模板**DNA**上非特异性位点的结合常数降低 $10^4$ 倍，使酶底复合物的半衰期小于1秒。



表3-2 大肠杆菌中的  $\sigma$  因子能识别并与启动子区的特异性序列相结合

因子	基因	功能	-35区	间隔 (bp)	-10区
$\sigma^{70}$	rpoD	广泛	TTGACA	16-18	TATAAT
$\sigma^{32}$	rpoH	热休克	TCTCNC CCTTGA A	13-15	CCCCAT NTA
$\sigma^{54}$	rpoN	氮代谢	CTGGNA	6	TTGCA



枯草杆菌中有6种不同相对分子质量的 $\sigma$ 因子，其中 $\sigma^{55}$ 是主要存在形式，出现在营养细胞中， $\sigma^{29}$ 则主要出现在孢子形成阶段，参与孢子形成期基因转录的调控。

在大肠杆菌中，最常见的调控因子是由**rpoD**基因所编码的 $\sigma^{70}$ 。



由**rpoH**编码的  $\sigma^{32}$ 是与热休克启动子所控制的基因转录密切相关。

由**rpoN**编码的  $\sigma^{54}$ 则参与细胞的氮代谢。由**T4**噬菌体所编码的  $\sigma^{55}$ 能与大肠杆菌**RNA**聚合酶的核心酶结合，启动**T4**晚期基因的转录。



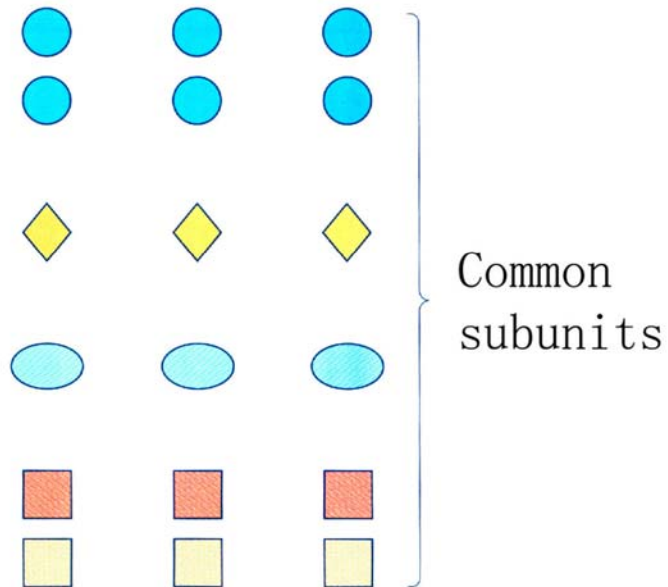
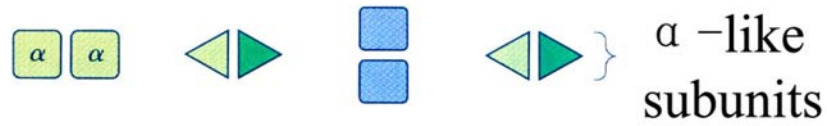
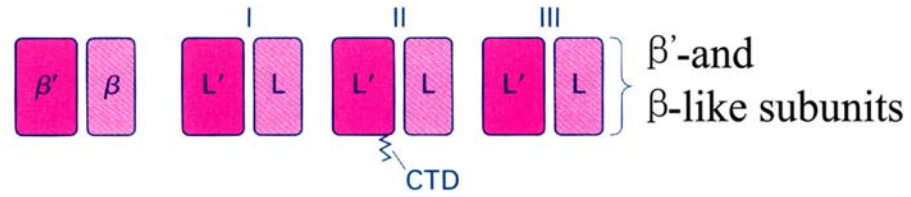
真核生物中有**3类RNA聚合酶**，其结构比大肠杆菌**RNA聚合酶**复杂，它们在细胞核中的位置不同，负责转录的基因不同，对 $\alpha$ -鹅膏蕈碱的敏感性也不同。

真核生物**RNA聚合酶**一般有**8-14个亚基**所组成，相对分子质量超过 **$5 \times 10^5$** 。



E.coli core  
RNA polymerase  
( $\alpha_2 \beta \beta'$ )

Eukaryotic  
RNA polymerases





真核生物RNA聚合酶的主要特征：

聚合酶中有两个相对分子质量超过 $1 \times 10^5$ 的大亚基；

同种生物3类聚合酶有“共享”小亚基的倾向，即有几个小亚基是其中3类或2类聚合酶所共有的。



表3-3 真核细胞中三类RNA聚合酶特性比较

酶	细胞内定位	转录产物	相对活性	对 $\alpha$ -鹅膏蕈碱的敏感程度
RNA聚合酶I	核仁	rRNA	50%-70%	不敏感
RNA聚合酶II	核质	hnRNA	20%-40%	敏感
RNA聚合酶III	核质	tRNA	约10%	存在物种特异性









## 2. 转录复合物

启动子选择阶段包括**RNA聚合酶全酶**对启动子的识别，聚合酶与启动子可逆性结合形成**封闭复合物**（**closed complex**，此时，**DNA**仍处于双链状态）。

然后，封闭复合物转变成开放复合物（**open complex**），聚合酶全酶所结合的**DNA**序列中有一小段双链被解开。对于强启动子来说，从封闭复合物到开放复合物的转变是不可逆的，是快反应。



# 图3-4 RNA合成的起始

反应成分	参数	状态
		全酶
<p>↑ ↓ 聚合酶与 DNA 结合</p> 	结合平衡常数 $K_R = 10^6 \sim 10^9$	二元闭合复合物
<p>↓ DNA 开始解链</p> 	反应速率 $k_2 = 10^{-3} \sim 10^{-1} \text{ base/s}$	二元开链复合物
<p>↓ 转录起始</p> 	反应速率 $k_1 \approx 10^{-3} \text{ base/s}$	三元复合物
<p>↓ <math>\sigma</math> 因子被释放</p>  	聚合酶 离开启动区时间 $> 1 \sim 2 \text{ s}$	RNA 合成 开始



开放复合物与最初的两个**NTP**相结合并在这两个核苷酸之间形成磷酸二酯键后即转变成包括**RNA**聚合酶、**DNA**和新生**RNA**的三元复合物。除**RNA**聚合酶之外，真核生物转录起始过程中至少还需要**7**种辅助因子参与。



# 表3-4 真核生物RNA聚合酶II所形成的转录起始复合物

蛋白质	亚基数	亚基的分子量 (ku)	功能
RNA聚合酶II	12	10-220	催化RNA的生物合成
TBP	1	38	与启动子上的TATA区相结合
TFIIA	3	12, 19, 35	使TBP及TFIIB与启动子的结合比较稳定
TFIIB	1	35	与TBP相结合, 吸引RNA聚合酶和TFIIF到启动区上
TFIID	12	15-250	与各种调控因子相互作用
TFIIE	2	34, 57	吸引TFIIH, 有ATP酶及解链酶活性
TFIIF	2	30, 74	结合RNA聚合酶II并在TFIIB帮助下阻止聚合酶与非特异性DNA序列相结合
TFIIH	12	35-89	在启动子区解开DNA双链, 使RNA聚合酶II磷酸化, 接纳核苷酸切除修复体系



一般情况下，转录起始复合物可以进入两条不同的反应途径：

合成并释放2-9个核苷酸的短**RNA**转录物，即所谓的流产式起始；

尽快释放  $\sigma$  亚基，转录起始复合物通过上游启动子区并生成由核心酶、**DNA**和新生**RNA**所组成的转录延伸复合物。



转录延伸复合物较为稳定，可长时间与**DNA**模板结合而不解离。只有在遇到转录终止信号时，**RNA**聚合酶才停止加入新的核苷酸，**RNA-DNA**杂合物解离，释放转录产物并导致聚合酶本身从模板**DNA**上脱落。



## 3.2 启动子与转录起始

大肠杆菌RNA聚合酶与启动子的相互作用主要包括启动子区的识别、酶与启动子的结合及 $\sigma$ 因子的结合与解离。



### 3. 2. 1 启动子区的基本结构

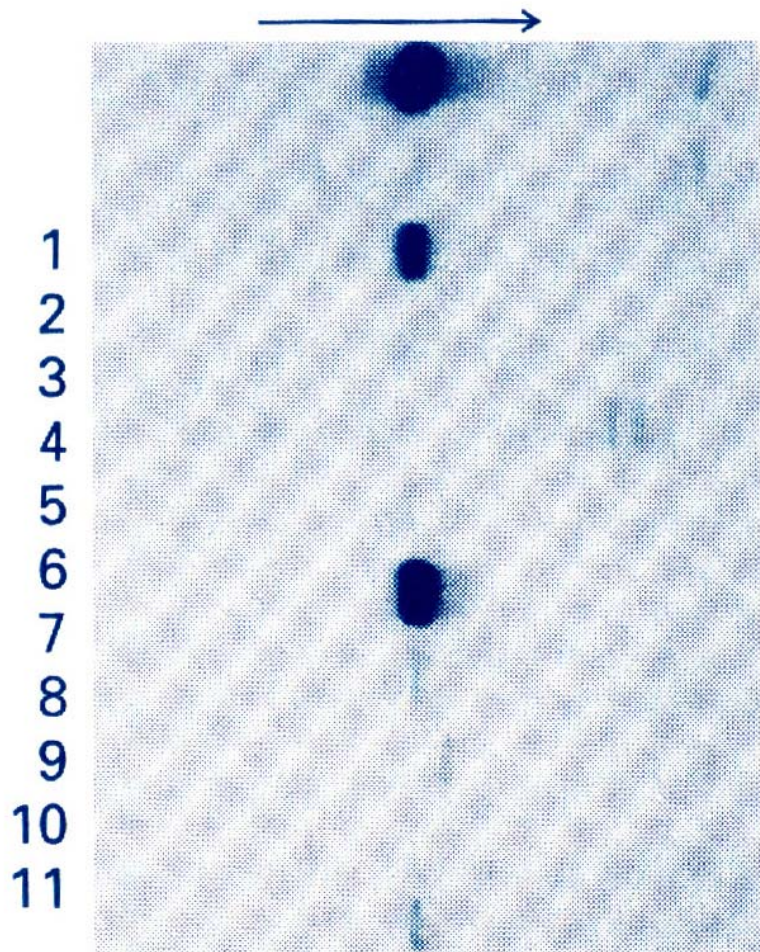
启动子是一段位于结构基因5'端上游区的DNA序列，能活化RNA聚合酶，使之与模板DNA准确地相结合并具有转录起始的特异性。



转录的起始是基因表达的关键阶段，而这一阶段的重要问题是RNA聚合酶与启动子的相互作用。启动子的结构影响了它与RNA聚合酶的亲和力，从而影响了基因表达的水平。



转录起点是指与新生**RNA**链第一个核苷酸相对应**DNA**链上的碱基，通常为嘌呤。把起点5'末端的序列称为上游（**upstream**），而把其3'末端的序列称为下游（**downstream**）。起点为+1，下游方向依次为+2、+3.....等，上游方向依次为-1、-2、-3.....等等。



Initiator sequence

+1  
 ●  
 CCCTCATTCT  
 CAGGCAGGGA  
 TAGGCAATCA  
 GTTACATGGA  
 GCCCAAGGG  
 ATGGCAACCG  
 TTGACAGACT  
 CCGCAACTA  
 CTTACAACCG  
 CCTGCATGCG  
 CCGCAAGTT

Consensus sequence:

+1  
 ●  
 PyPyAN<sup>T</sup>PyPy  
 A



转录单元 (transcription unit) 是一段从启动子开始至终止子 (terminator) 结束的DNA序列。



启动子区有什么结构特点呢？

**Pribnow**将**RNA聚合酶全酶**与**模板DNA**结合后，用**DNaseI**降解**DNA**，得到**41~44**个核苷酸对的**DNA**片段。

序列分析发现，在被保护区内有一个由**5**个核苷酸组成的保守序列，是聚合酶结合位点，称为**Pribnow**区，其中央大约位于起点上游**10bp**处，所以又称为**-10**区。



RNA 聚合酶结合到  
DNA 特定位点



加核酸酶

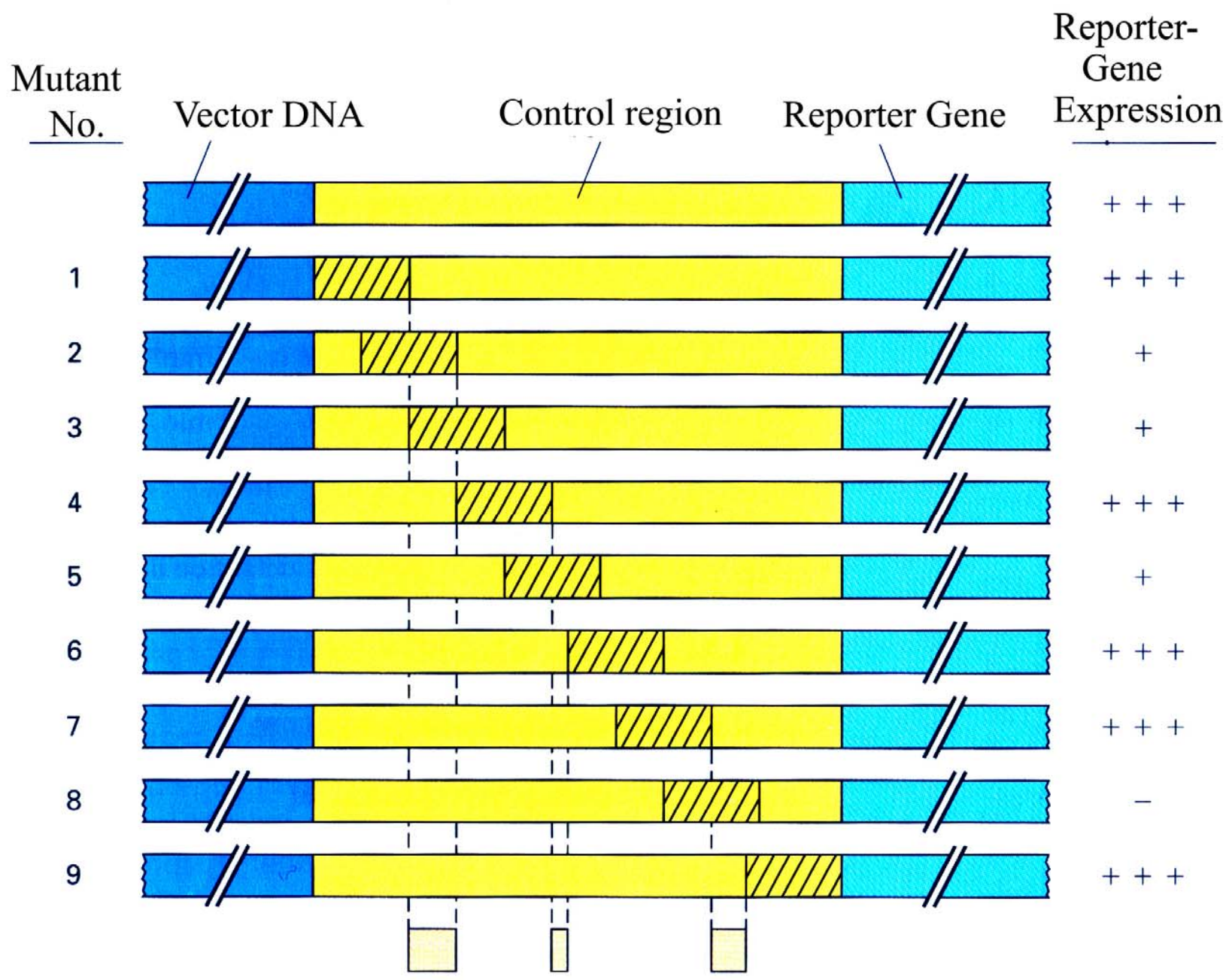


提取 DNA



确定被保护的 DNA 序列

图 3-5 分离转录启动子区 DNA 序列程序图示





科学家又从噬菌体的左、右启动子 $P_L$ 及 $P_R$ 和SV40启动子的-35 bp附近找到了另一段共同序列：**TTGACA**。

现已证实大部分启动子都存在这两段共同序列，即位于-10 bp处的**TATA**区和-35 bp处的**TTGACA**区，它们是**RNA**聚合酶与启动子的结合位点。

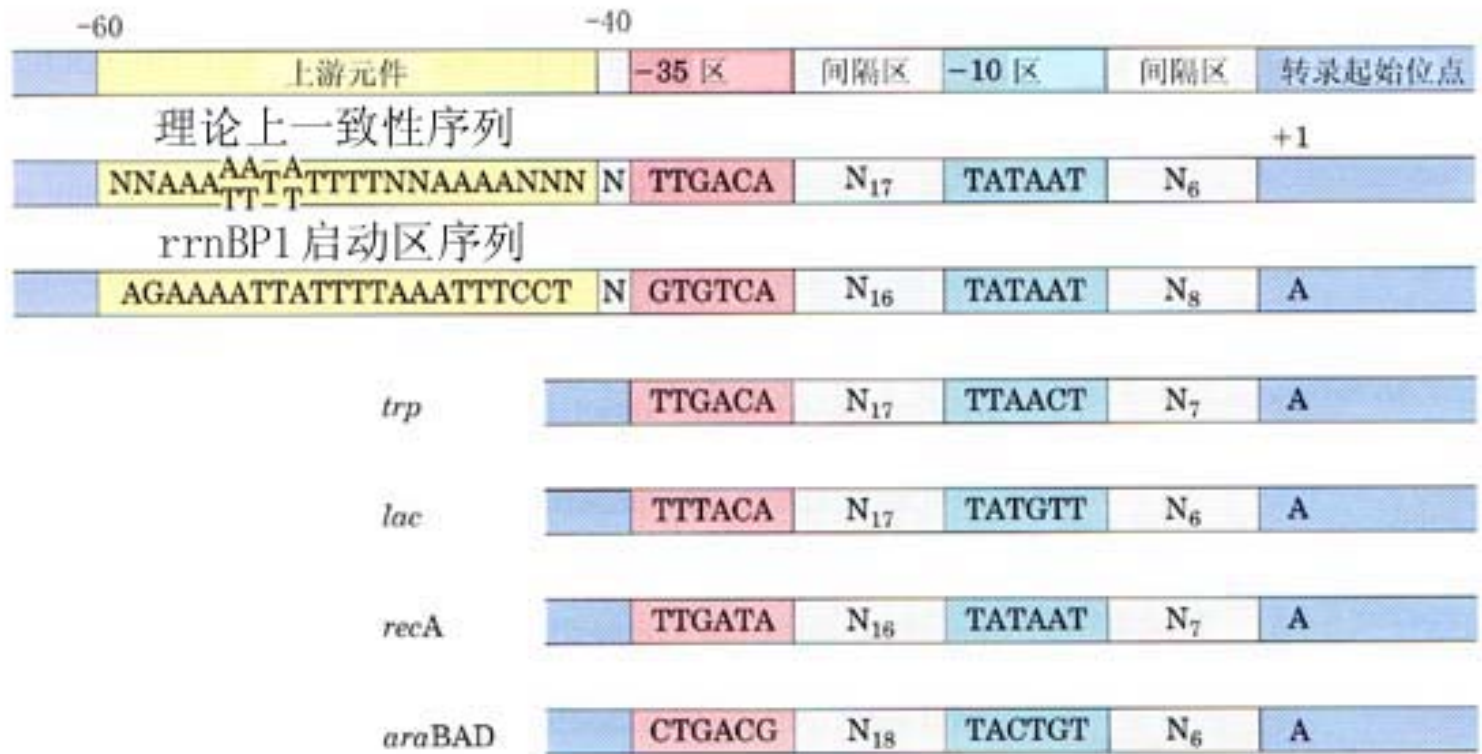


图3-6 大肠杆菌RNA聚合酶全酶所识别的启动子区

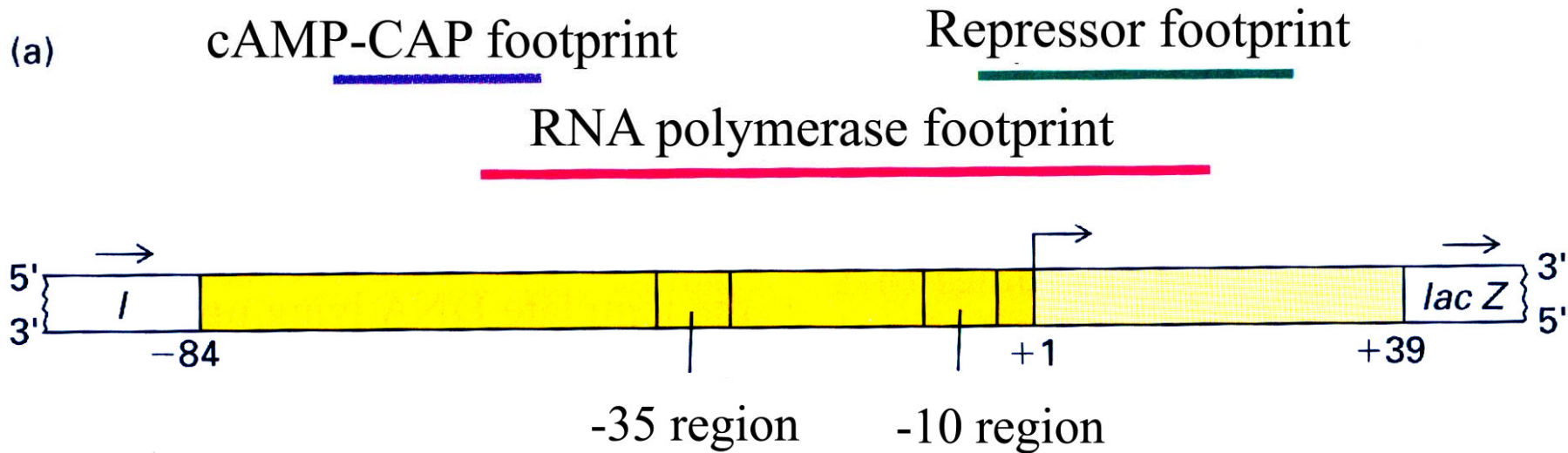


-35区

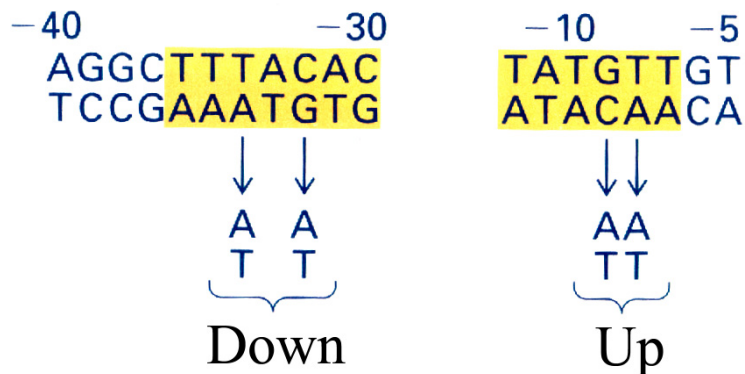
-10区

.....T<sub>85</sub>T<sub>83</sub>G<sub>81</sub>A<sub>61</sub>C<sub>69</sub>A<sub>52</sub>.....T<sub>89</sub>A<sub>89</sub>T<sub>50</sub>A<sub>65</sub>A<sub>100</sub>...  
...

在真核生物基因中，Hogness等发现类似Pribnow区的Hogness区，在转录起始点上游-25~-30 bp处，保守序列为TATAAA，也称TATA区。在起始位点上游-70~-78 bp处还有另一段共同序列CCAAT，称为CAAT区（CAAT box）。

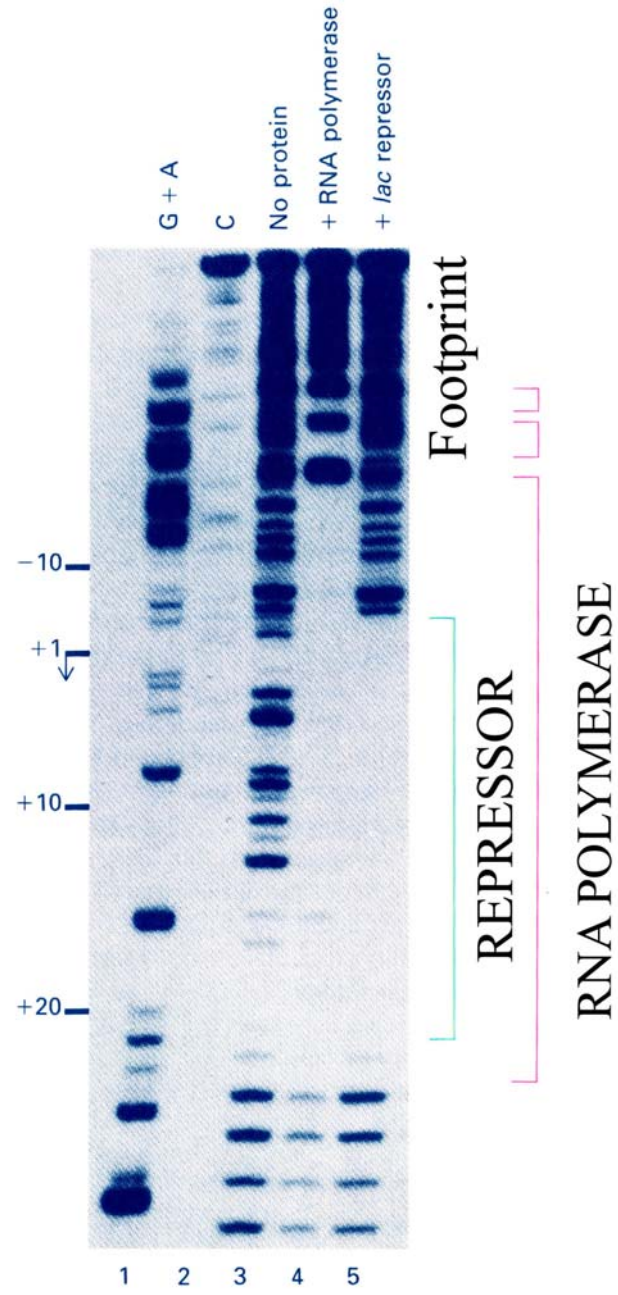


(b) Promoter mutations



(c)  $O^c$  mutations





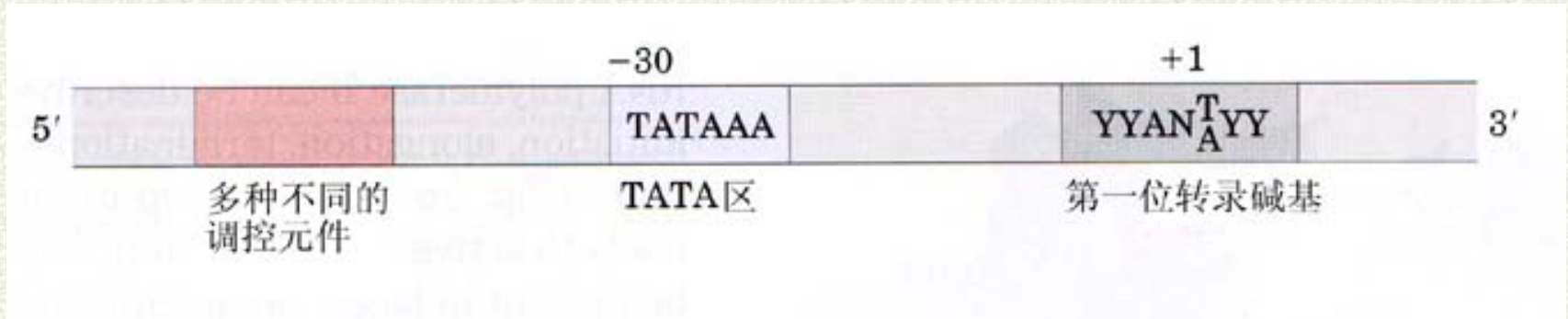


图3-7 真核生物RNA聚合酶II所识别的启动子区



真核基因的启动子在-25~-35区含有TATA序列，在-70~-80区含有CCAAT序列（CAAT box），在-80~-110含有GCCACACCC或GGGCGGG序列（GC box）。习惯上，将TATA区上游的保守序列称为上游启动子元件（upstream promoter element, UPE）或称上游激活序列（upstream activating sequence, UAS）。



如果把 **Pribnow** 区从 **TATAAT** 变成 **AATAAT** 就会使该启动子发生下降突变；  
如果增加 **Pribnow** 区的共同序列，将乳糖操纵子的启动子中的 **TATGTT** 变成 **TATATT**，就会提高启动子的效率，称为上升突变。



已在SV40的转录单元上发现其转录起始位点上游约200bp处有两段72bp长的重复序列，它们不是启动子的一部分，但能增强或促进转录的起始，因此，称这种能强化转录起始的序列为增强子或强化子（**enhancer**）。



**TATA区的主要作用是使转录精确地起始。如果除去TATA区或进行碱基突变，转录产物下降的相对值不如CAAT区或GC区突变后明显，但发现所获得的RNA产物起始点不固定（图3-10）。**

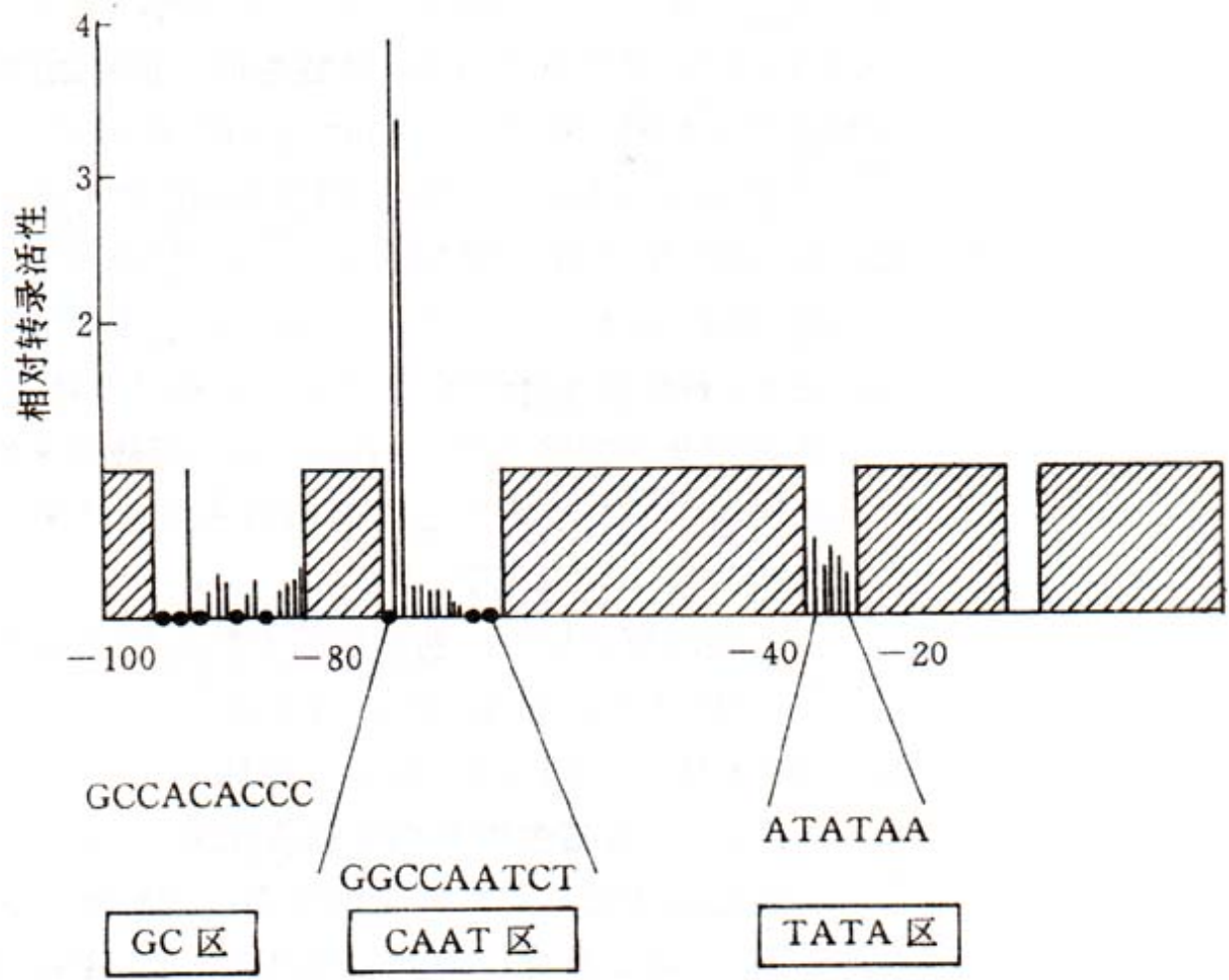


图3-10 启动子区主要顺式作用元件与β-珠蛋白基因转录活性



研究SV40晚期基因启动子发现，上游激活区的存在与否，对该启动子的生物活性有着根本性的影响。若将该基因5'上游-21~-47核苷酸序列切除，基因完全不表达（图3-11）。

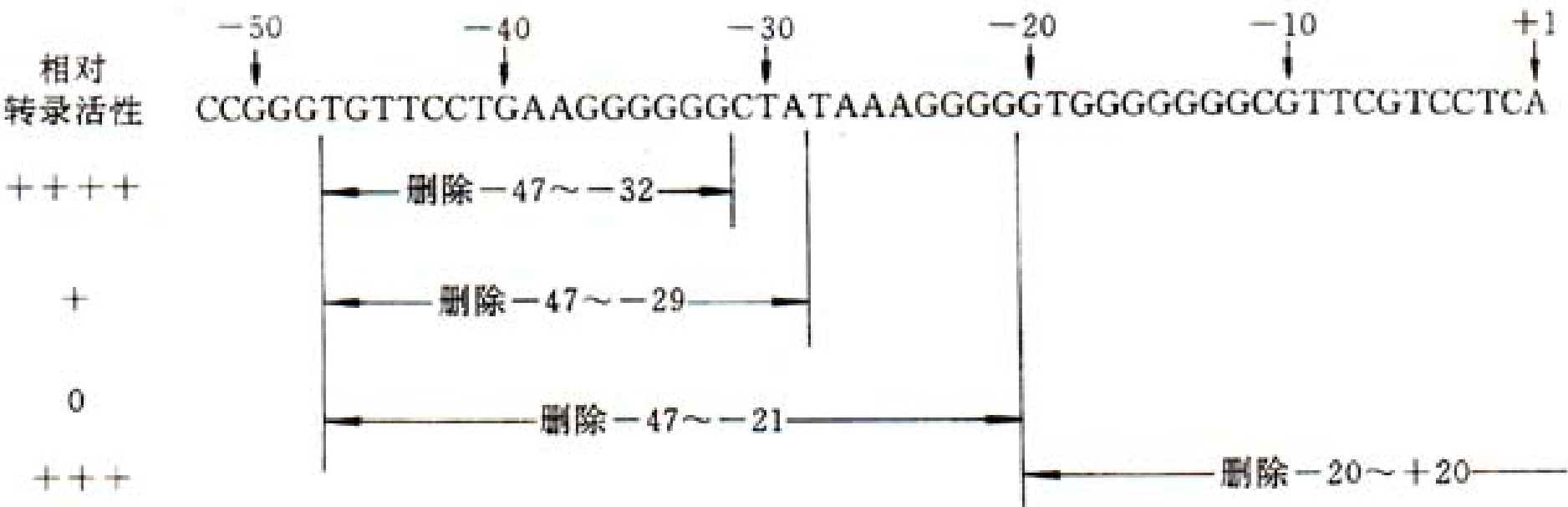


图3-11 SV40 基因启动子上 TATAAA 及邻近区域对基因转录活性的影响



**CAAT区和GC区主要控制转录起始频率，基本不参与起始位点的确定。**

**SV40的早期基因，缺少TATA和CAAT区，只有6个串联在上游-40~-110位点的GC区；而组蛋白H<sub>2</sub>B，不含GC区，但有两个CAAT区，一个TATA区。**

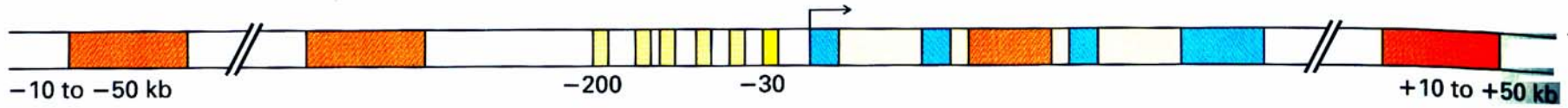


### 表3-5 真核细胞中的部分转录调控因子分析

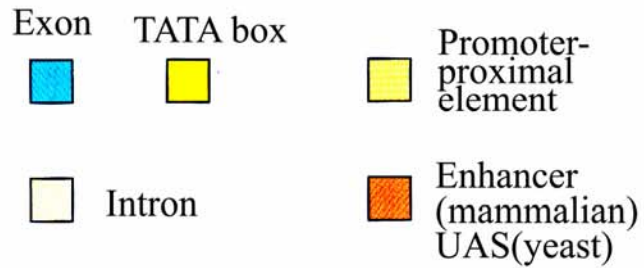
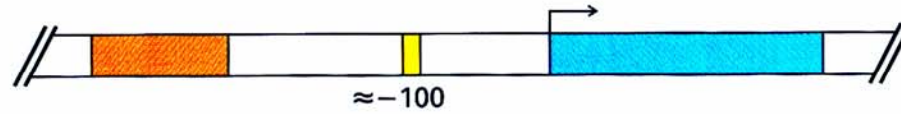
转录因子	DNA结合区的特征	DNA结合区序列	在生物体内的分布
组成型转录因子			
CTF/NF1	CAAT区	5'-GCCAATCT-3'	广泛
CP	CAAT区	5'-GCCAATCT-3'	广泛
C/EBP	CAAT区	5'-GCCAATCT-3'	广泛
Sp1	GC区	5'-GGGCGG-3'	广泛
Oct-1	八碱基区	5'-ATGCAAAT-3'	广泛
Oct-2	八碱基区	5'-ATGCAAAT-3'	B-淋巴细胞
应答蛋白			
热激因子	HSE	5'-CNNGAANNTCCNNG-3'	受热激以后
血清反应因子	SRE	5'-CCATATTAGG-3'	受血清生长因子刺激
细胞特异性因子			
GATA-1	-	5'-GATA-3'	成红细胞
Pit-1	-	5'-ATATTCAT-3'	腺体细胞
MyoD1	E区	5'-CANNG-3'	成肌细胞
NF-κB	κB位点	5'-GGGACTTCC-3'	淋巴细胞



(a) Mammalian Gene



(b) *S.cerevisiae* gene





### 3.3 原核与真核生物mRNA的特征比较

**mRNA**在大肠杆菌细胞内占总**RNA**的**2%**左右，**tRNA**占**16%**，**rRNA**则占**80%**以上。

真核细胞的**mRNA**往往以较大相对分子量的前体**RNA**出现在核内，只有成熟的、相对分子质量明显变小并经化学修饰的**mRNA**才能进入细胞质，参与蛋白质的合成。所以，真核细胞**mRNA**的合成和功能表达发生在不同的空间和时间范畴内。



### 3.3.1 原核生物mRNA的特征

1. 半衰期短。原核生物中，**mRNA**的转录和翻译是在同一个细胞空间里同步进行的，蛋白质合成往往在**mRNA**刚开始转录时就被引发了。



大多数细菌mRNA在转录开始1分钟后就开始降解。mRNA降解的速度大概只有转录或翻译速度的一半。科学家发现每过大约2分钟，体系中出现新生蛋白质的速度就下降50%。



**2.原核细胞的mRNA（包括病毒）有时可以编码几个多肽，而真核细胞的mRNA最多只能编码一个多肽。**

**我们把只编码一个蛋白质的mRNA称为单顺反子mRNA（monocistronic mRNA），把编码多个蛋白质的mRNA称为多顺反子mRNA（polycistronic mRNA）。**



几乎所有**mRNA**都可以被分成3部分：编码区和位于**AUG**之前的5'端上游非编码区以及位于终止密码子之后不翻译的3'端下游非编码区。编码区从起始密码子**AUG**开始经一连串编码氨基酸的密码子直至终止密码子。



第一个蛋白质合成终止以后，核糖体分解成大、小亚基，脱离**mRNA**模板，第二个蛋白的翻译必须等到新的小亚基和大亚基与该蛋白起始密码子相结合后才可能开始（图**3-12A**）。

前一个多肽翻译完成以后，核糖体大、小亚基分离，小亚基也可能不离开**mRNA**模板，而是迅速与游离的大亚基结合，启动第二个多肽的合成（图**3-12B**）。

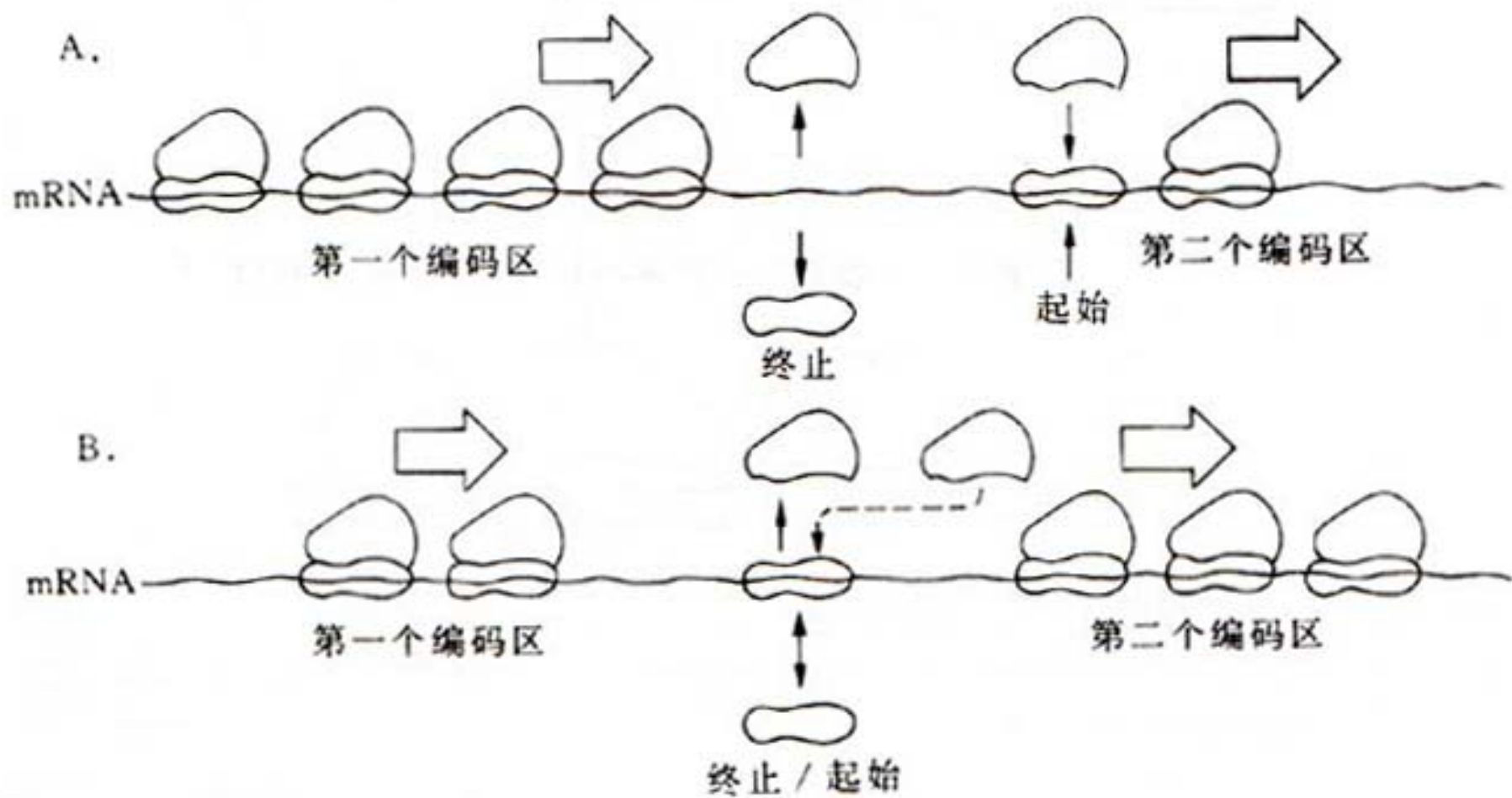
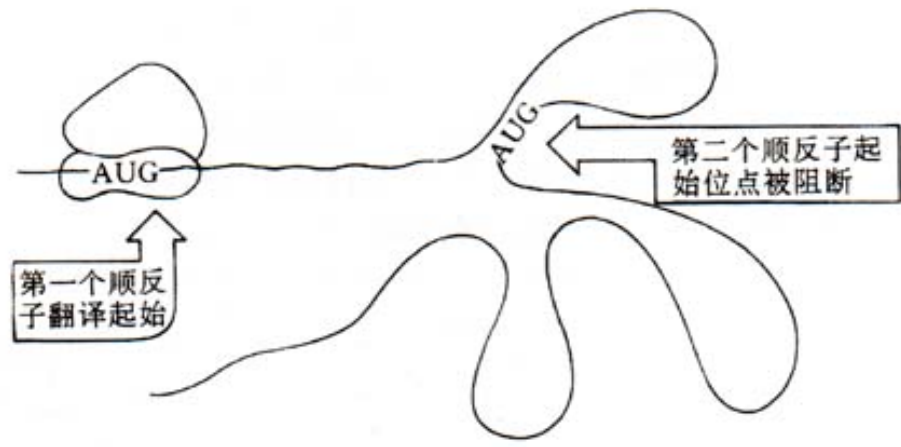


图 3-12 原核生物多顺反子基因中后续编码区翻译起始受顺反子之间距离的影响 A. 当两个顺反子之间距离较远时, 前一顺反子翻译的终止与后一顺反子翻译的起始是相互独立的; B. 当两个顺反子之间距离较近时, 前一顺反子翻译的终止与后一顺反子翻译的起始相衔接, 30 S 小亚基始终与 mRNA 结合, 只有 50 S 大亚基可能与 mRNA 分离



一个顺反子的翻译有时完全取决于它前面顺反子的翻译，因为只有第一个翻译起始位点是暴露的，在这个顺反子翻译产生多肽的过程中，核糖体的运动破坏了后续顺反子的二级结构，使起始位点较容易与核糖体相结合形成起复合物（图3-13）。



随着另一个顺反子的翻译,mRNA 的二级结构被破坏

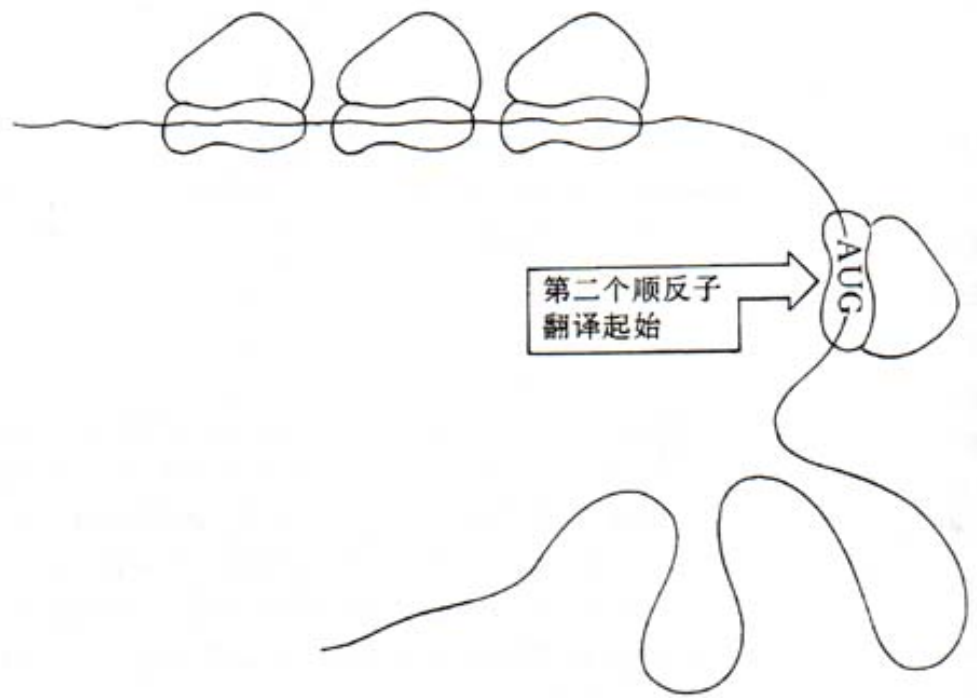


图 3-13 mRNA 的次级结构有可能控制翻译的起始



**3. 原核生物 mRNA 的 5' 端无帽子结构，3' 端没有或只有较短的多聚 (A) 结构。**

原核生物起始密码子 AUG 上游 7-12 个核苷酸处有一被称为 SD 序列 (Shine - Dalgarno sequence) 的保守区，因为该序列与 16S-rRNA 3' 端反向互补，所以被认为在核糖体-mRNA 的结合过程中起作用。



4.原核生物常以**AUG**（有时**GUG**，甚至**UUG**）作为起始密码子，而真核生物几乎永远以**AUG**作为起始密码子。



### 3.3.2 真核生物mRNA的特征

编码功能蛋白的真核基因都通过**RNA聚合酶II**进行转录，几乎都是单顺反子，其长度在几百到几千个核苷酸之间。

一个完整的基因，不但包括编码区（**coding region**），还包括**5'**和**3'**端长度不等的特异性序列，它们虽然不编码氨基酸，却在基因表达的过程中起着重要作用。

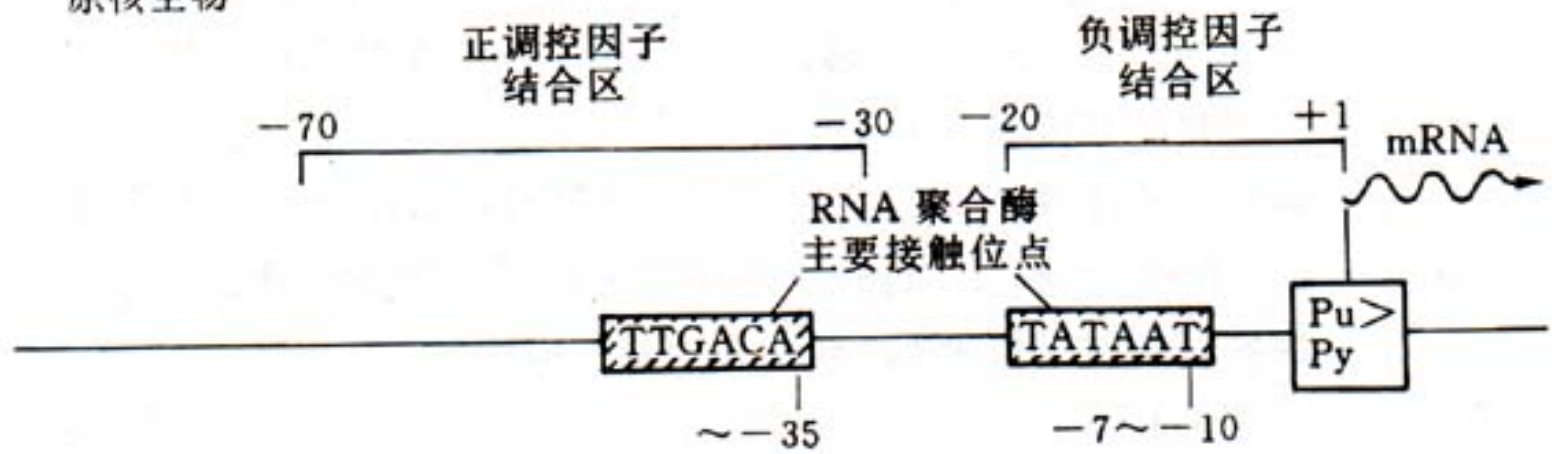


“基因”的分子生物学定义是：产生一条多肽链或功能RNA所必需的全部核苷酸序列！真核生物mRNA结构上的最大特征是5'端的帽子及3'的多聚(A) 结构。

**A gene can be defined as following:  
The entire nucleic acid sequence  
that is necessary for the synthesis of  
a functional polypeptide or RNA  
molecule.**



原核生物



真核生物

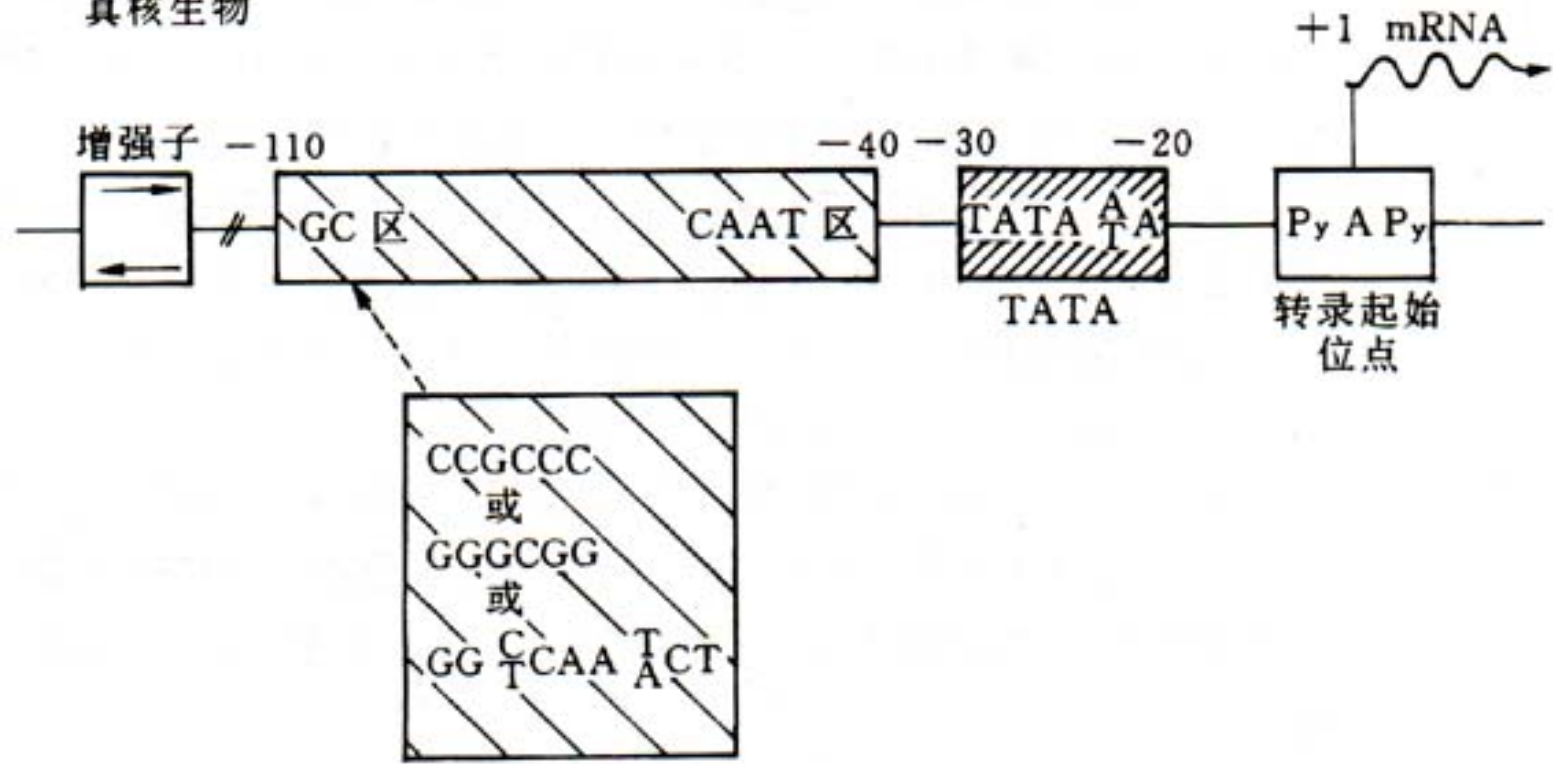


图 3-9 原核和真核生物基因转录起始位点上游区的结构比较



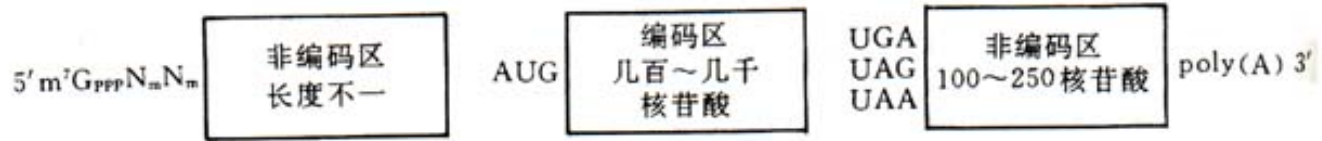
1. 真核生物mRNA的5'端存在“帽子”结构。真核生物基因转录一般从嘌呤（主要是A，也可能是G）起始，其5'端大都经过修饰，第一个核苷酸保留了5'端的三磷酸基团。



用核酸酶处理成熟**mRNA**，其**5'**端并不产生预期的核苷三磷酸，而产生以**5'-5'**三磷酸基团相连的二核苷酸，**5'**终端是一个在**mRNA**转录后加上去的甲基化鸟嘌呤。



### A. 真核生物 mRNA 的结构模式



### B. 真核生物 mRNA 的“帽子”结构

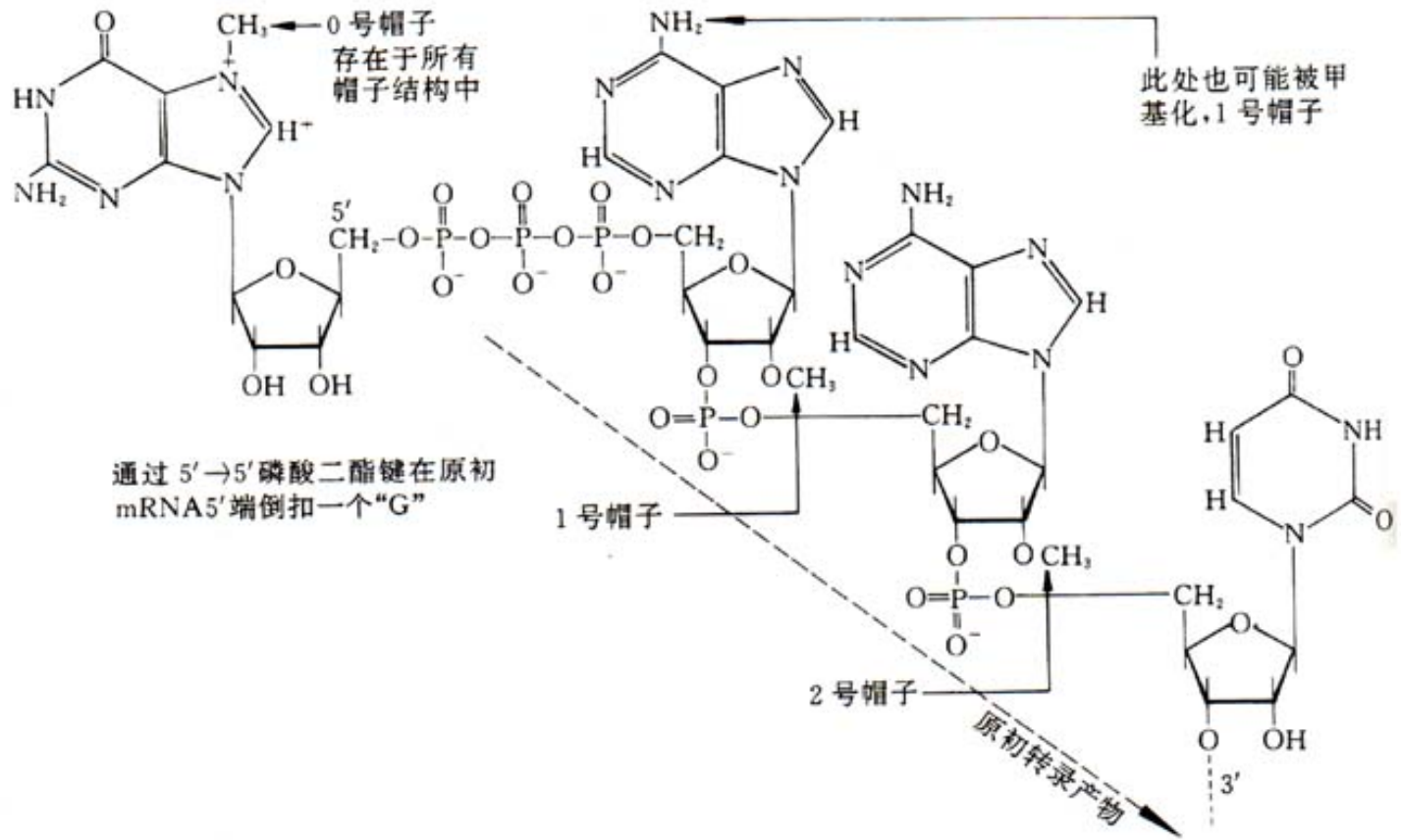


图 3-15 真核生物 mRNA 结构



**mRNA**的帽子结构常常被甲基化。第一个甲基出现在所有真核细胞的**mRNA**中（单细胞真核生物**mRNA**主要是这个结构），由尿苷酸-7甲基转移酶催化，称为零类帽子（**cap0**）。



如在第二个核苷酸(原mRNA 5'第一位)的2'-OH位上加另一个甲基，这步反应由2'-O-甲基转移酶完成。一般把有这两个甲基的结构称为1类帽子（cap1），真核生物中以这类帽子结构为主。

当mRNA原第二位核苷酸是腺嘌呤时，其N<sup>6</sup>位有时也被甲基化，这一反应只能在2'-OH被甲基化以后才能发生。



在某些生物细胞内，mRNA链上的第三个核苷酸的2'-OH位也可能被甲基化，因为这个反应只以带有1类帽子的mRNA为底物，所以被称为2类帽子（cap2）。有2类帽子的mRNA只占有帽mRNA总量的10%-15%以下。



2. 绝大多数真核生物**mRNA**具有多聚(A)尾巴。除组蛋白基因外，真核生物**mRNA**的3'末端都有多聚(A)序列，其长度因**mRNA**种类不同而变化，一般为**40-200**个左右。



真核基因的3' 末端转录终止位点上游 15 ~ 30bp 处的保守序列 **AAUAAA** 对于初级转录产物的准确切割及加多聚 (A) 是必需的 (图3-16)。

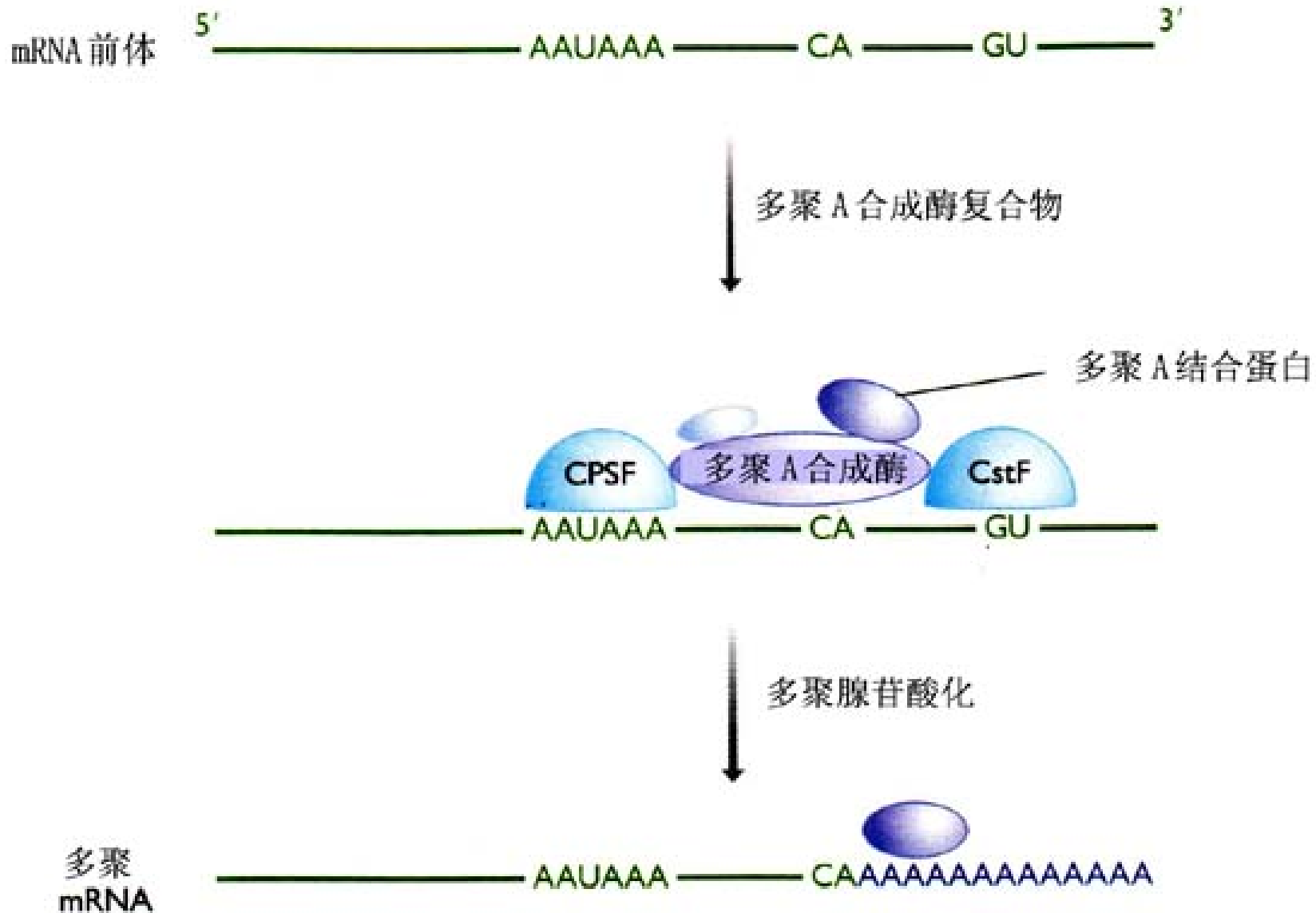


图3-16 真核生物mRNA中的加多聚A反应



**多聚(A)是mRNA由细胞核进入细胞质所必需的形式，它大大提高了mRNA在细胞质中的稳定性。mRNA刚从细胞核进入细胞质时，其多聚(A)尾巴一般比较长，随着mRNA在细胞质内逗留时间延长，多聚(A)逐渐变短消失，mRNA进入降解过程。**



### 3.4 终止和抗终止

RNA聚合酶启动基因转录后，就沿模板5'→3'方向不停地移动，合成RNA链直到碰上终止信号时才与模板DNA相脱离并释放新生RNA链。所有参与形成RNA-DNA杂合体的氢键都被破坏，模板DNA链与有意义链重新组合成DNA双链。



### 3.4.1 不依赖于 $\rho$ 因子的终止

终止位点上游一般有一个富含GC碱基的二重对称区，由这段DNA转录产生的RNA容易形成发卡式结构。在终止位点前面有一段由4-8个A组成的序列，所以转录产物的3'端为寡聚U，这种结构特征的存在决定了转录的终止。

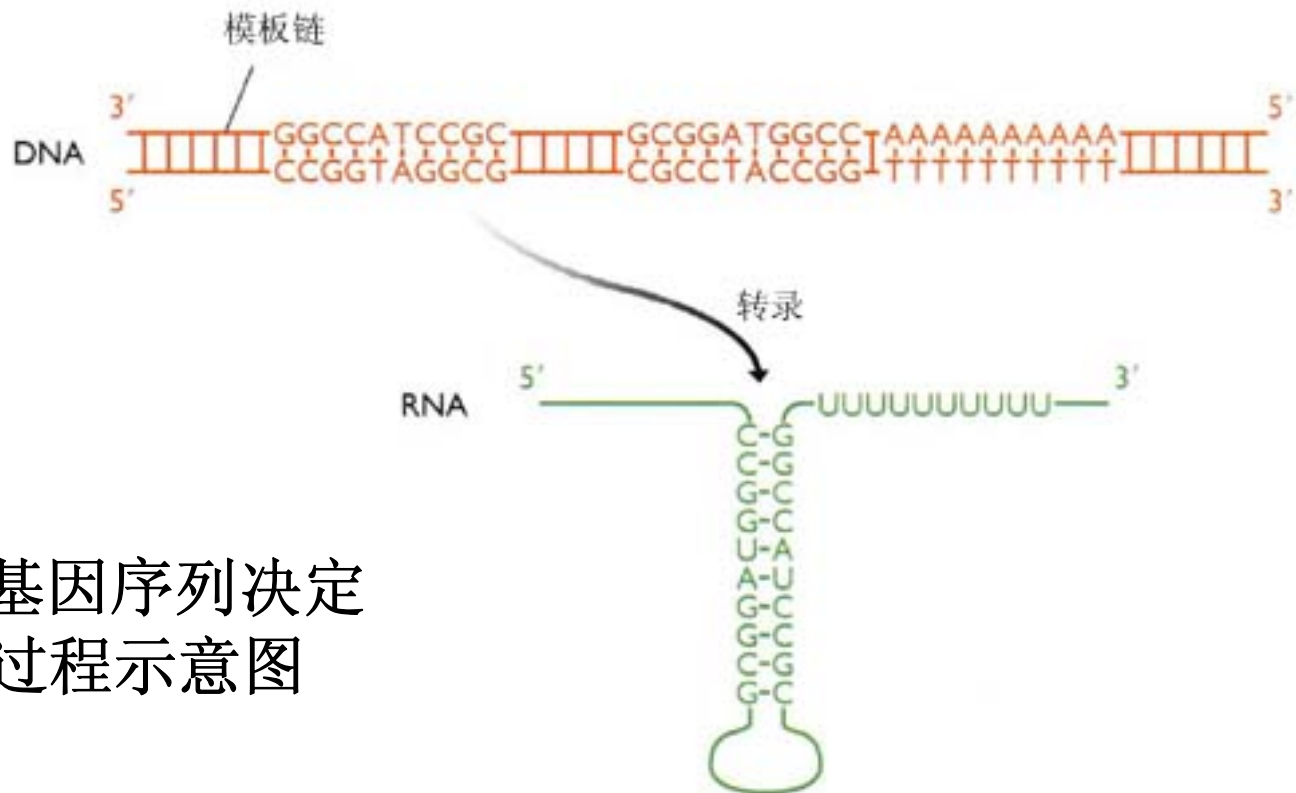
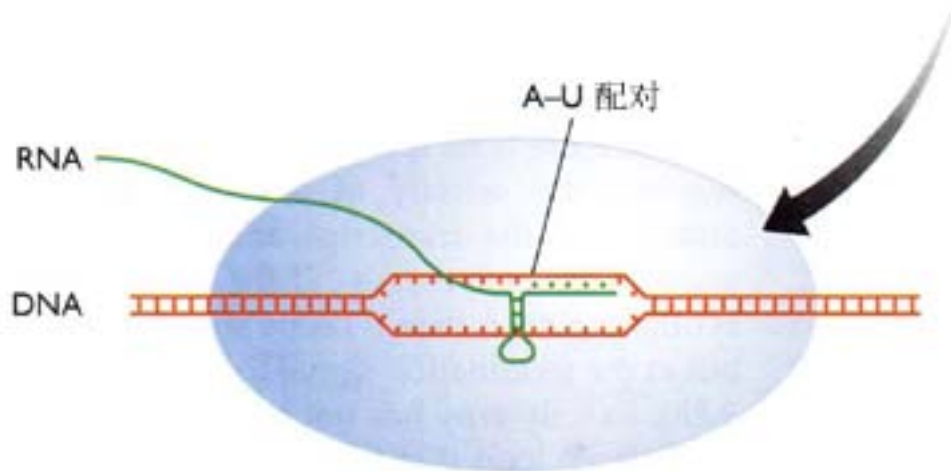


图3-17 由基因序列决定的转录终止过程示意图





新生RNA中出现发卡式结构会导致RNA聚合酶的暂停，破坏RNA-DNA杂合链5'端的正常结构。寡聚U的存在使杂合链的3'端部分出现不稳定的rU•dA区域，两者共同作用使RNA从三元复合物中解离出来。



终止效率与二重对称序列和寡聚U的长短有关，随着发卡式结构（至少6bp）和寡聚U序列（至少4个U）长度的增加，终止效率逐步提高。



### 3.4.2 依赖于 $\rho$ 因子的终止

提纯的RNA聚合酶并不能识别特异性的转录终止信号，而加入大肠杆菌  $\rho$  因子后该聚合酶就能在DNA模板上准确地终止转录。 $\rho$  因子是一个相对分子质量为  $2.0 \times 10^5$  的六聚体蛋白，它通过催化NTP的水解促使新生RNA链从三元转录复合物中解离。

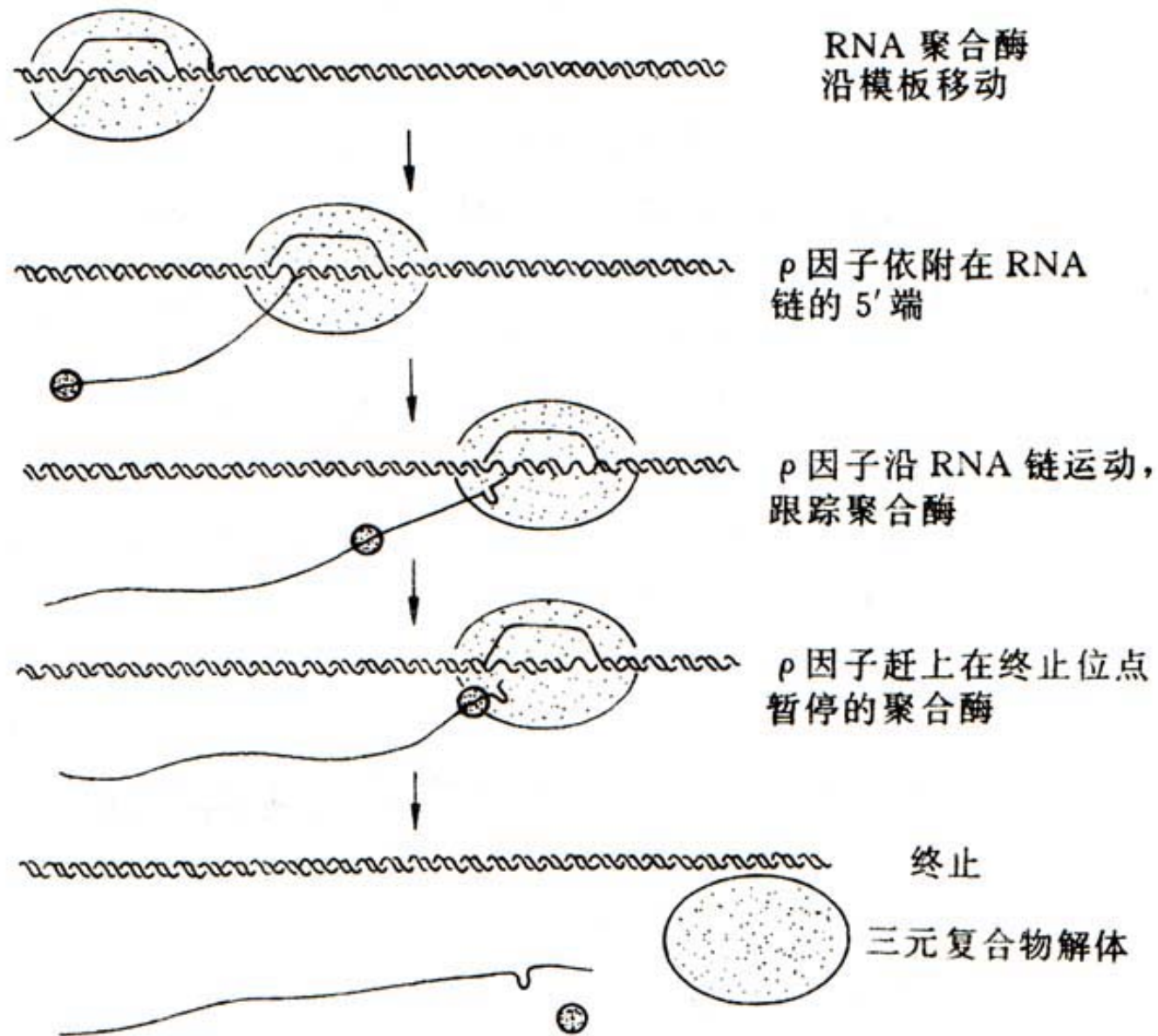
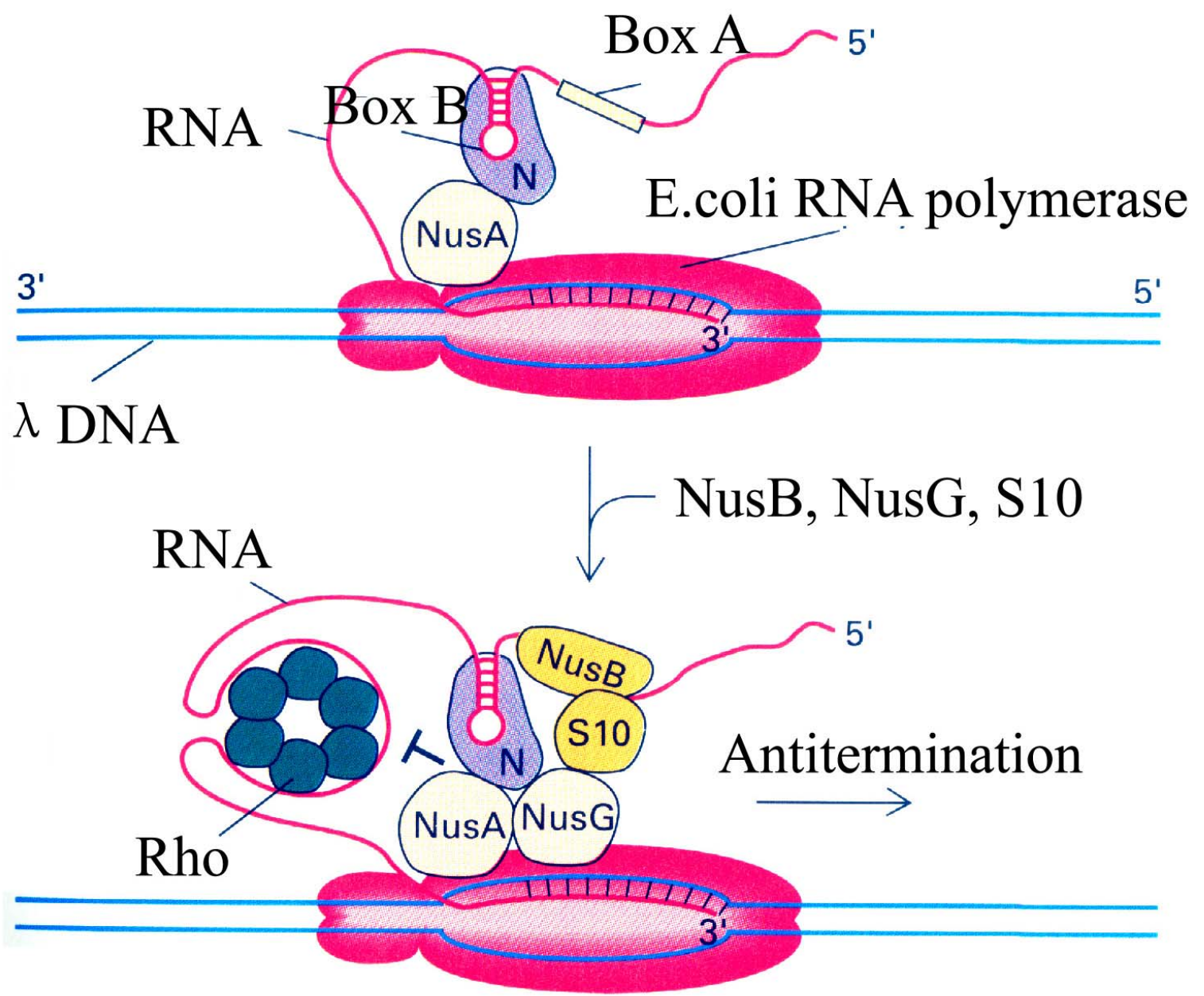


图 3-18  $\rho$  因子参与的 RNA 合成终止模式



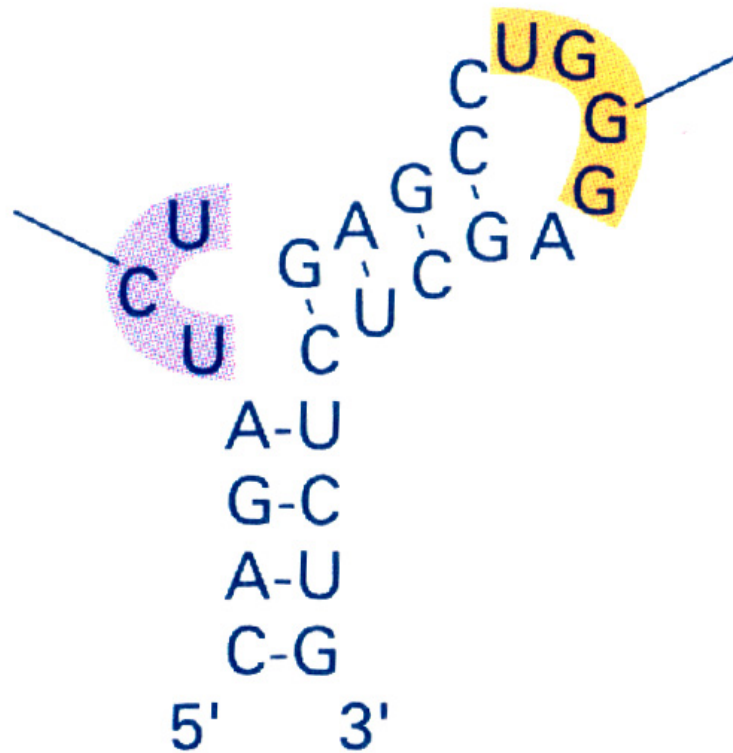
RNA合成起始以后， $\rho$  因子即附着在新生的RNA链上，靠ATP水解产生的能量，沿着5'→3'方向朝RNA聚合酶移动，到达RNA的3'-OH端后取代了暂停在终止位点上的RNA聚合酶，使之从模板DNA上释放mRNA，完成转录过程。



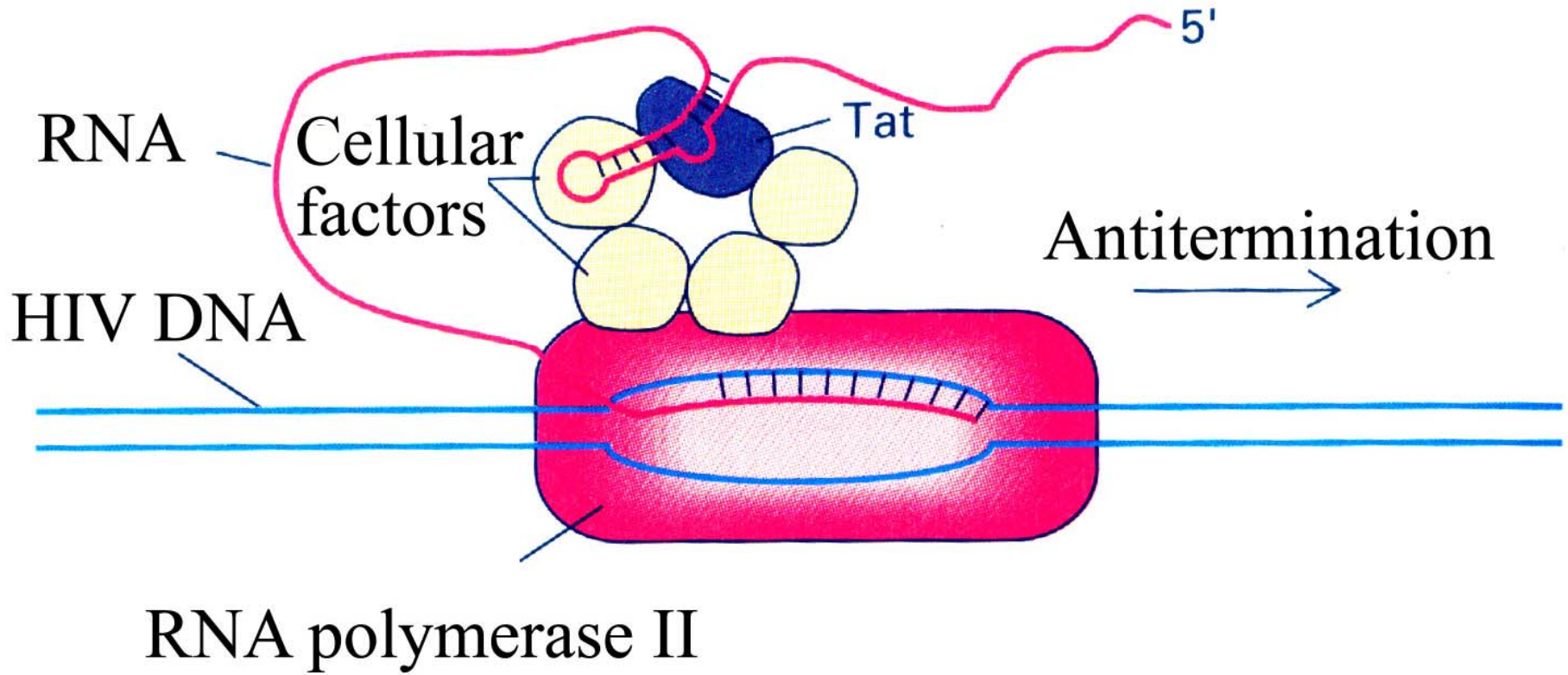


# (a) TAR sequence

Tat  
recognition  
site



Cellular factor  
recognition site



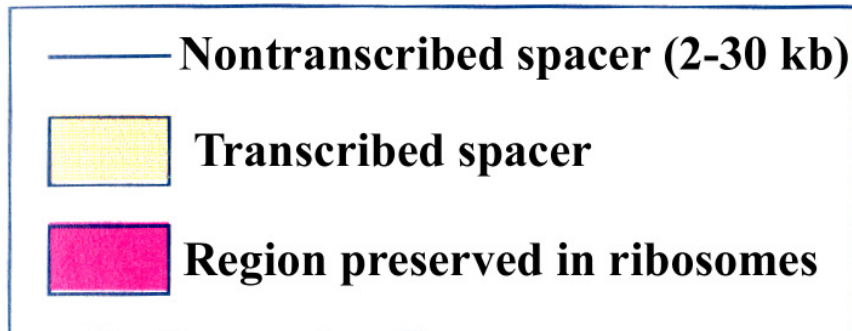
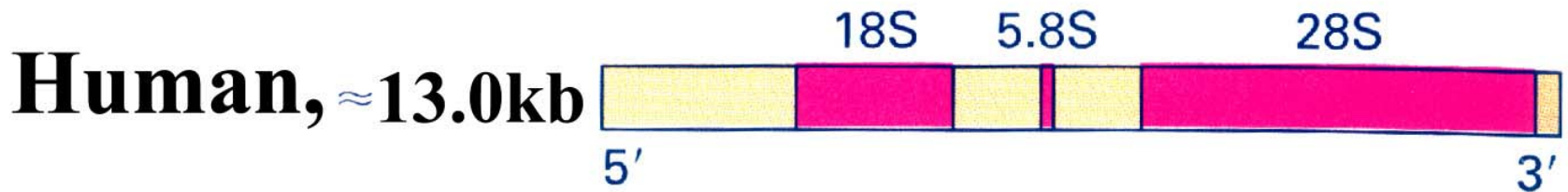


## 3.5 RNA Pol I和Pol III介导的基因转录

**RNA Pol I** 只用于合成**pre-rRNA**。有三大特征：1. **pre-rRNA**组成一个长的转录单元；

2. 真核生物基因组上rDNA区中成熟rRNA基因相对位置和方向永远相同（5'—> **18 S**，**5.8 S**和**28 S** —> 3'）；

3. 无论原核还是真核生物中，pre-rRNA转录单元都显著长于成熟rRNA的总和。





# (b) TANDEM ARRAY

Transcription unit

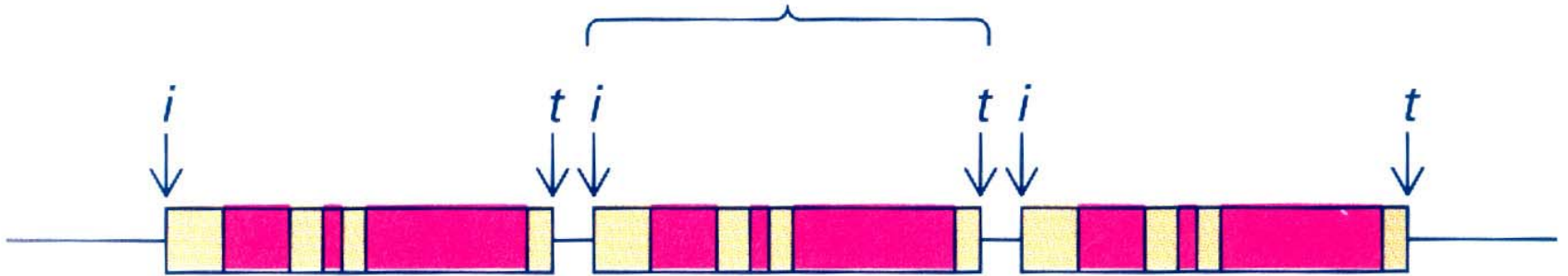


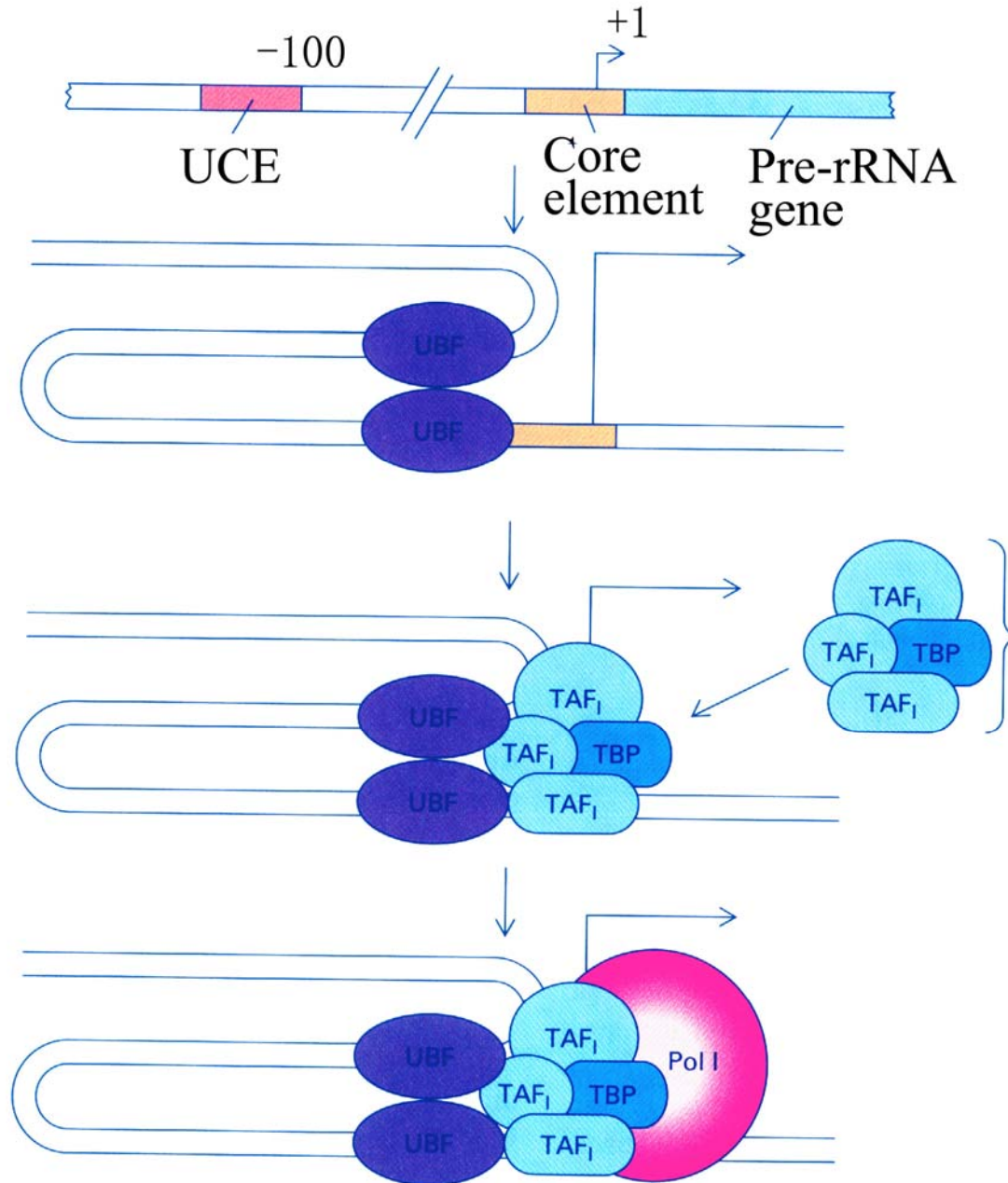


TABLE 11-6 Lengths of Primary rRNA Transcripts and Cytoplasmic rRNAs in Various Species

	Primary transcript*		Ribosomal RNA Length*		
	S Value	Length (Kb)	26S-28S (Kb)	16S-18S (Kb)	Percentage of Precursor Preserved
<i>Esherichia coli</i> (prokaryote)	30	6.0	3.0	1.5	75
<i>Saccharomyces cerevisiae</i> (yeast)	37	6.6	3.8	1.7	77
<i>Dictyostelium discoideum</i> (slime mold)	37	7.4	4.1	1.8	80
<i>Drosophila melanogaster</i> (fruit fly)	34	7.7	4.1	1.8	76
<i>Xenopus laevis</i> (frog)	40	7.9	4.5	1.9	81
<i>Gallus domesticus</i> (chicken)	45	11.2	4.6	1.8	57
<i>Mus musculus</i> (mouse)	45	13.7	5.1	1.9	51
<i>Humo sapiens</i> (human)	45	13.7	5.1	1.9	51

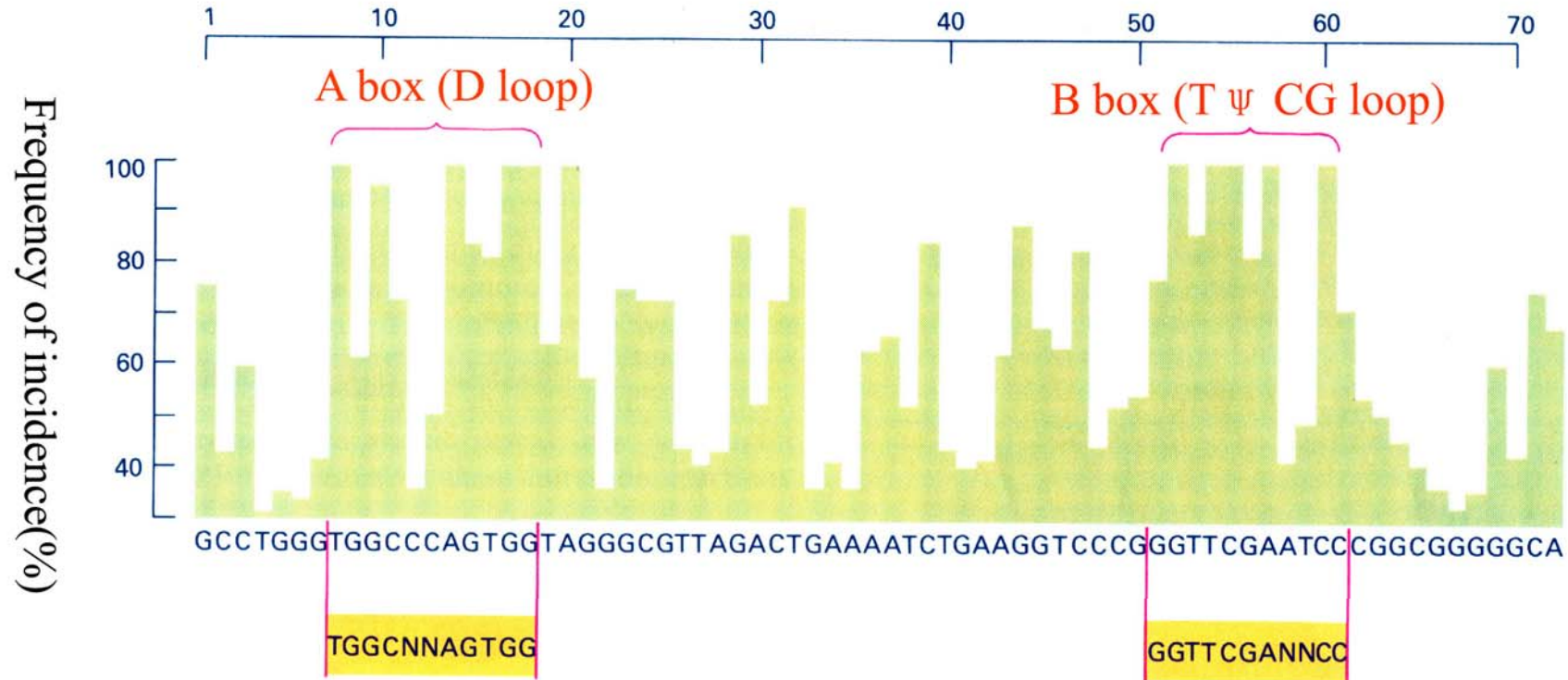
\*The lengths of the various RNA molecules are estimates based on gel electrophoresis and direct measurements of electron micrographs. The size is of the first major product with definite 5' and 3' end.

SOURCE:B. Lewin, 1980, Gene Expression, vol.3, Wiley, p.867





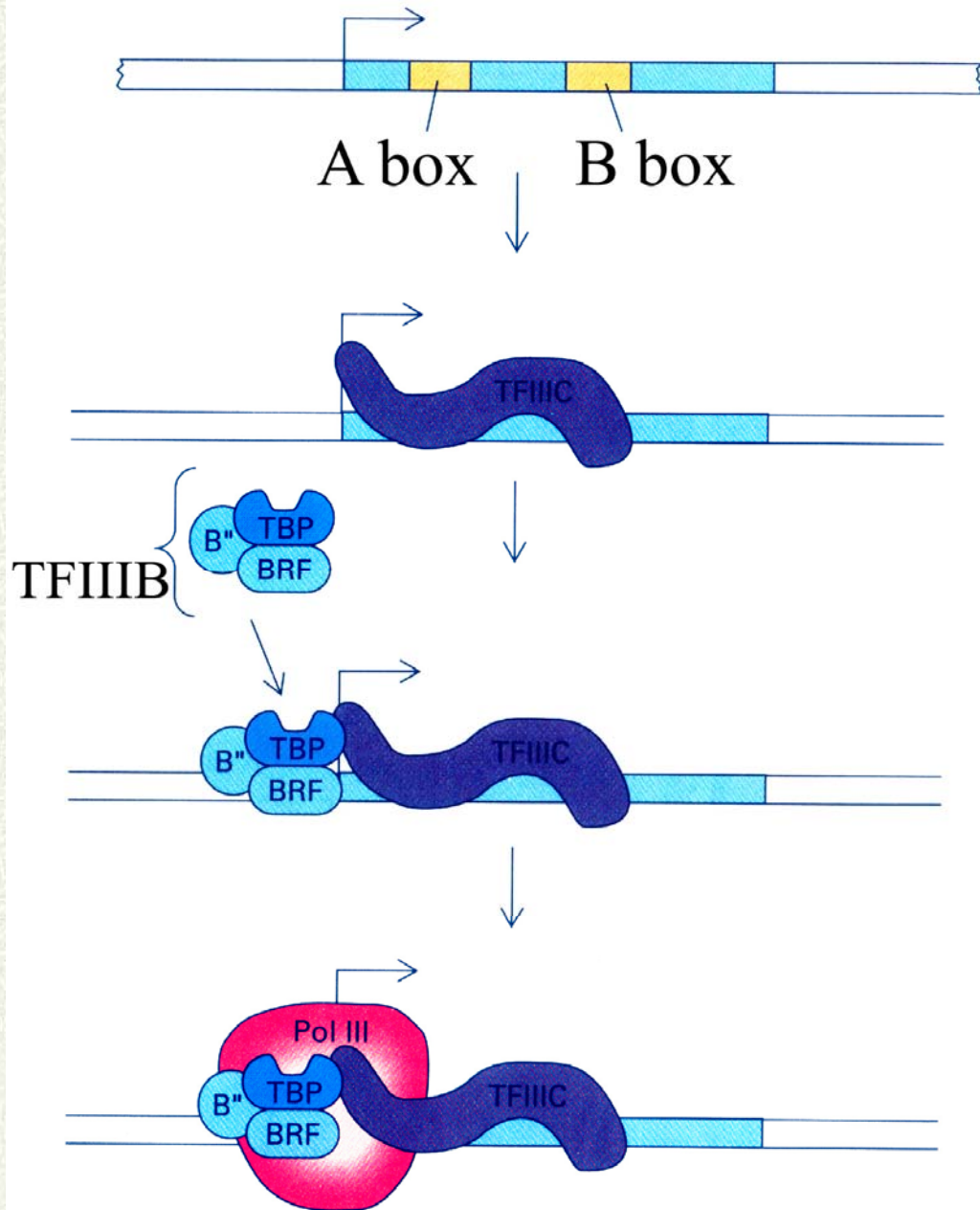
# tRNA基因的保守性分析



Consensus sequences

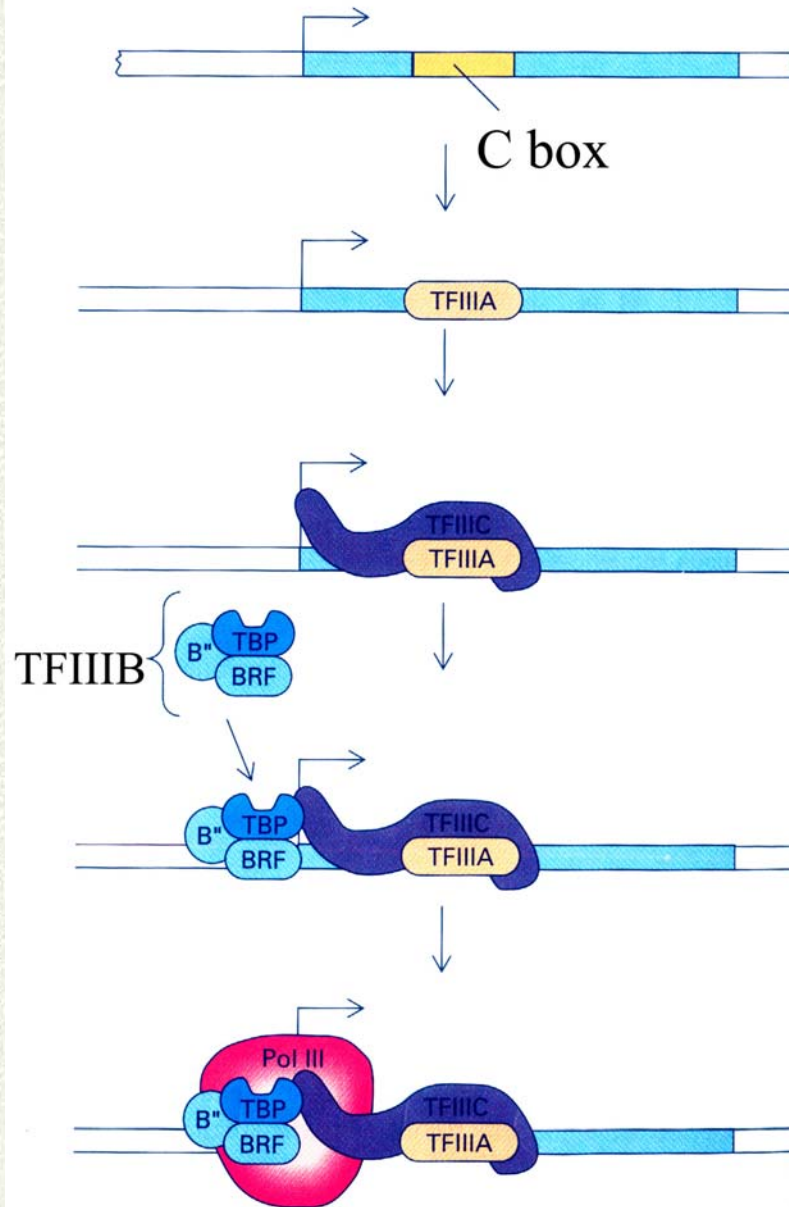


# (a) tRNA gene





## (b) 5S-rRNA gene





## 3.6 内含子的剪接、编辑及化学修饰

### 3.6.1 RNA中的内含子

因为真核基因表达往往伴随着**RNA**的剪接过程（**splicing**），从**mRNA**前体分子中切除被称为内含子（**intron**）的非编码区，并使基因中被称为外显子（**exon**）的编码区拼接形成成熟**mRNA**。

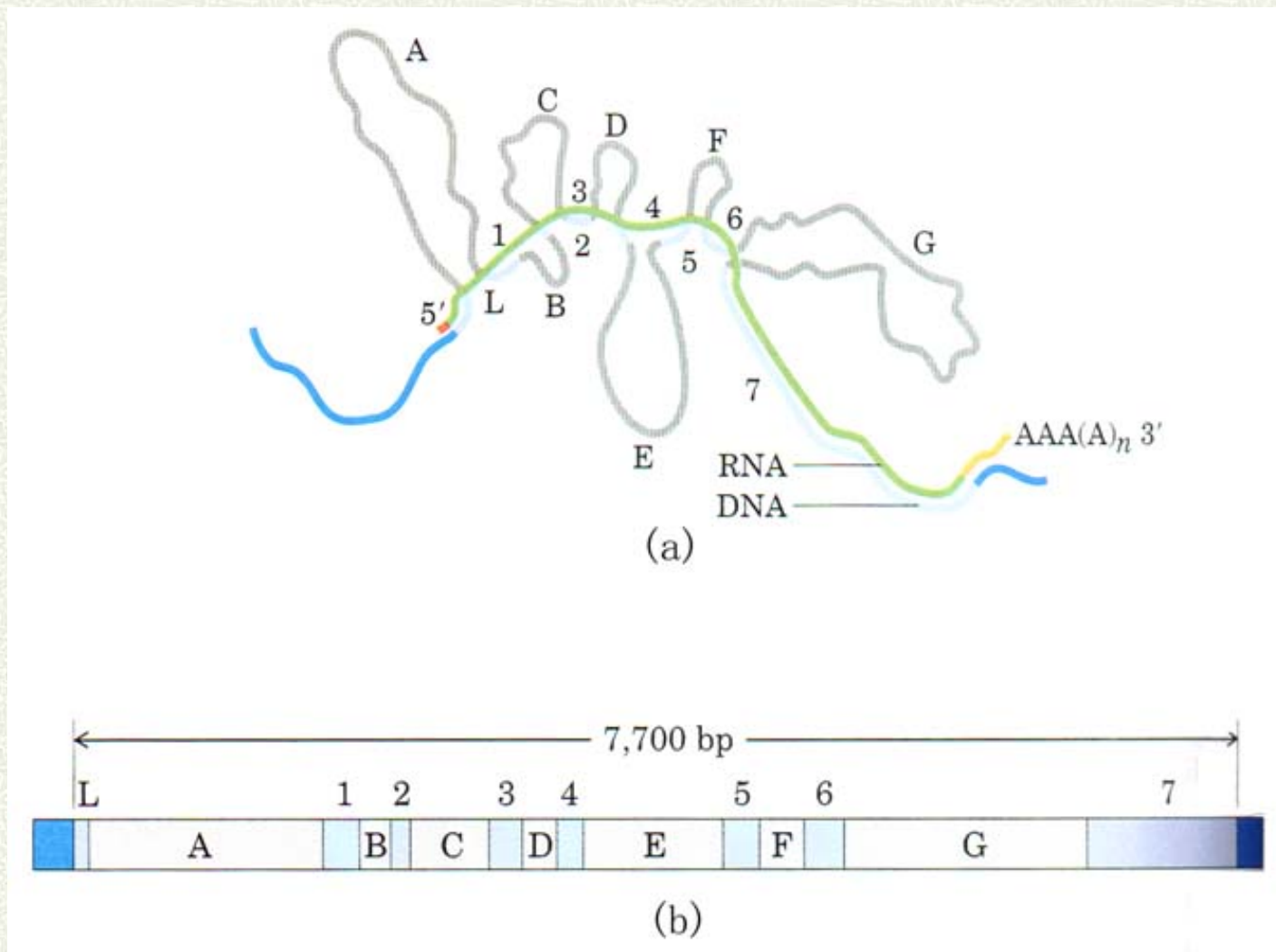


图3-20 用DNA-RNA杂交的方法验证鸡卵清蛋白基因中存在非编码的内含子区。 a. 成熟mRNA与带有该基因的互补链DNA序列杂交示意图； b. 鸡卵清蛋白的基因结构示意图。



真核基因大多是断裂的，也就是说，一个基因可由多个内含子和外显子间隔排列而成。研究表明，内含子在真核基因中所占的比例很高，甚至超过**99%**（表3-7）。



**表3-7 部分人类基因中内含子序列所占的比重分析**

基因	长度 (kb)	内含子数量	内含子所占比重 (%)
胰岛素	1.4	2	67
$\beta$ -球蛋白	1.4	2	69
血清蛋白	18	13	89
胶原蛋白组分VII	31	117	71
VIII因子	186	25	95
萎缩性肌强直因子	2400	78	>99



由**DNA**转录生成的原始转录产物——核不均一**RNA**（**hnRNA**, **heterogeneous nuclear RNA**），即**mRNA**的前体，经过5'加“帽”和3'酶切加多聚腺苷酸，再经过**RNA**的剪接，编码蛋白质的外显子部分就连接成为一个连续的可译框（**open reading frame, ORF**），通过核孔进入细胞质，作为蛋白质合成的模板。



表3-8 RNA加工过程及其生理功能

加工过程	推测的生理功能
加帽子反应	MRNA从细胞核向细胞质运转，翻译起始
加多聚A反应	转录终止，翻译起始和mRNA降解
RNA的剪接	从mRNA, tRNA和rRNA分子中切除内含子
RNA的切割	从前体RNA中释放成熟tRNA和rRNA分子

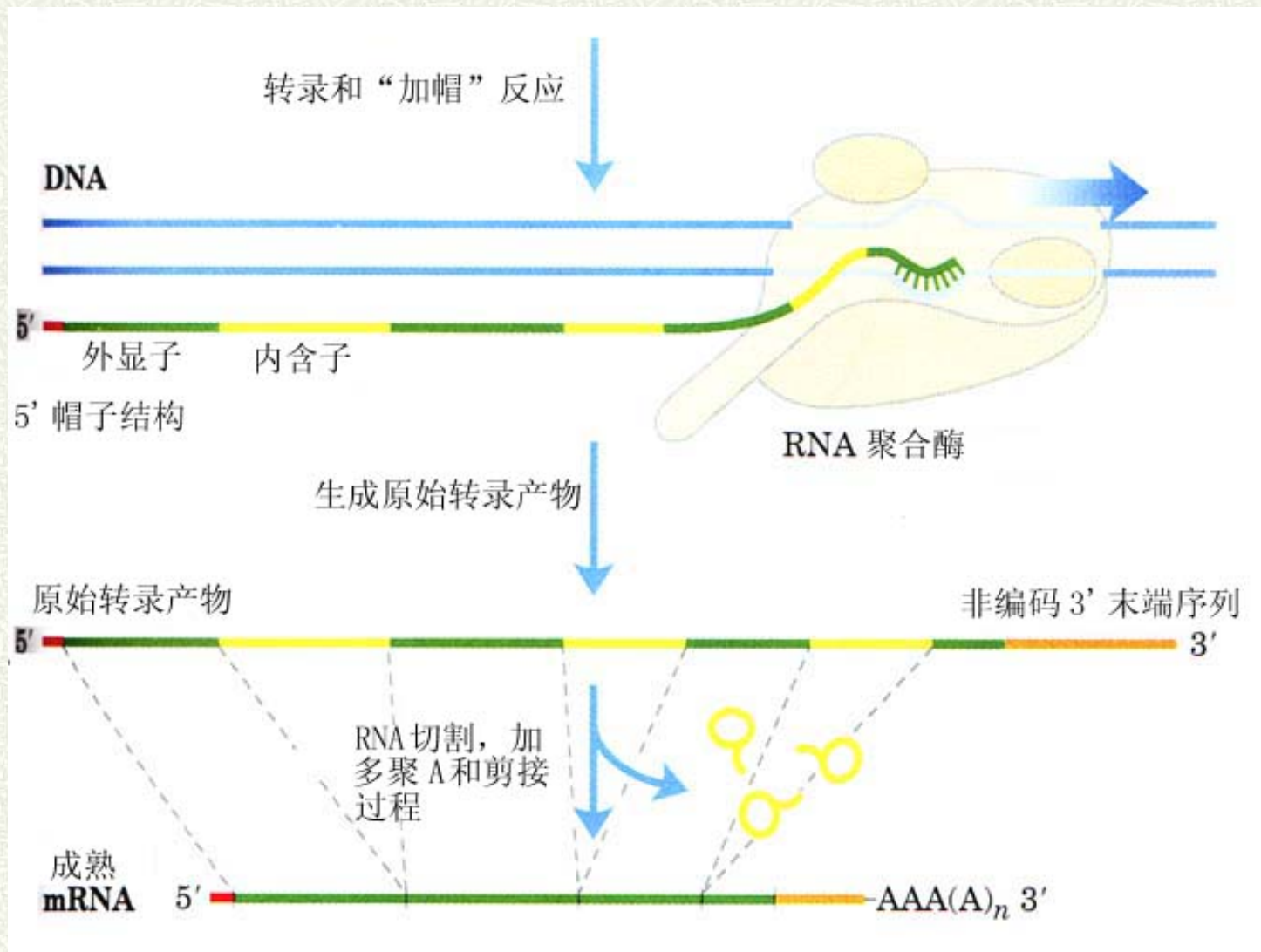


图3-21 原始转录产物的生成及其主要加工剪接过程图示



真核基因平均含有**8-10**个内含子，前体分子一般比成熟**mRNA**大**4-10**倍。

内含子的“功能”及其在生物进化中的地位是一个引人注目的问题。

许多人类疾病是内含子剪接异常引起的。地中海贫血病人的珠蛋白基因中，大约有**1/4**的核苷酸突变发生在内含子的**5'**或**3'**边界保守序列上，或者直接干扰了前体**mRNA**的正常剪接。



表3-9总结了存在于生物体内的各种内含子，其中**GU-AG**或**AU-AC**分别代表了不同内含子的5'和3'边界序列。除了边界序列之外，外显子与内含子交界处的序列，内含子内部的部分序列都有可能参与内含子的剪接（图3-22）。



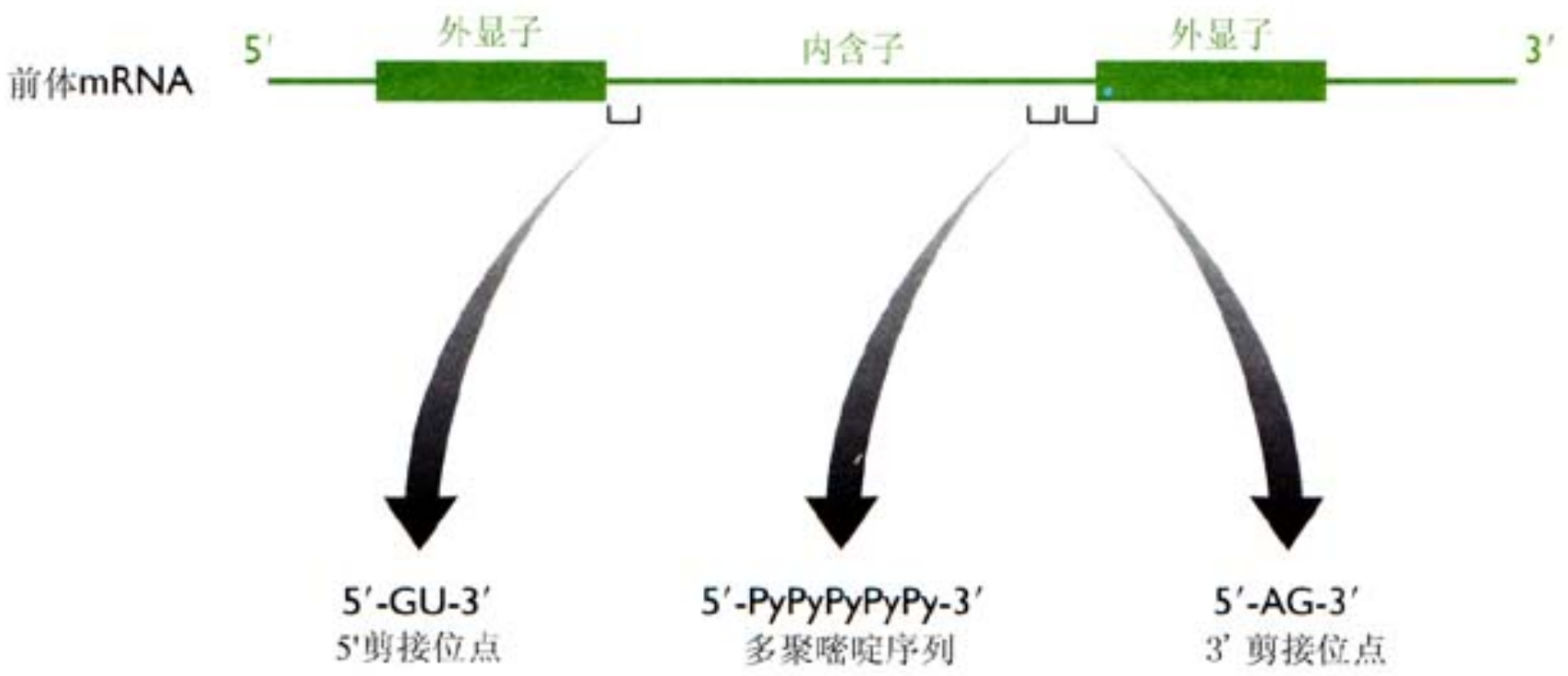
## 表3-9 生物体内的各种内含子

内含子类型	细胞内定位
GU-AG	细胞核，前体mRNA（真核）
AU-AC	细胞核，前体mRNA（真核）
I类内含子	细胞核，前体rRNA（真核）， 细胞器RNA，少数细菌RNA
II类内含子	细胞器RNA，部分细菌RNA
III类内含子	细胞器RNA
双内含子	细胞器RNA
tRNA前体中的内含子	细胞核，tRNA前体（真核）



## 3.6.2 RNA的剪接

转录产生的核内mRNA前体分子与蛋白质结合，形成RNA和蛋白质组成的RNP复合物（ribonucleo-protein particle）。随着RNA链的延伸，每个内含子5'和3'两端的复合物成对联结，产生60S的颗粒—剪接体（spliceosome），进行RNA前体分子的剪接。



**B.**  
 5' 剪接位点相邻的保守序列      5'-AG ↓ GUAAGU-3'  
 3' 剪接位点相邻的保守序列      5'-PyPyPyPyPyNCAG ↓ -3'

图3-22 脊椎动物前体mRNA中常见内含子剪接所必需的保守序列

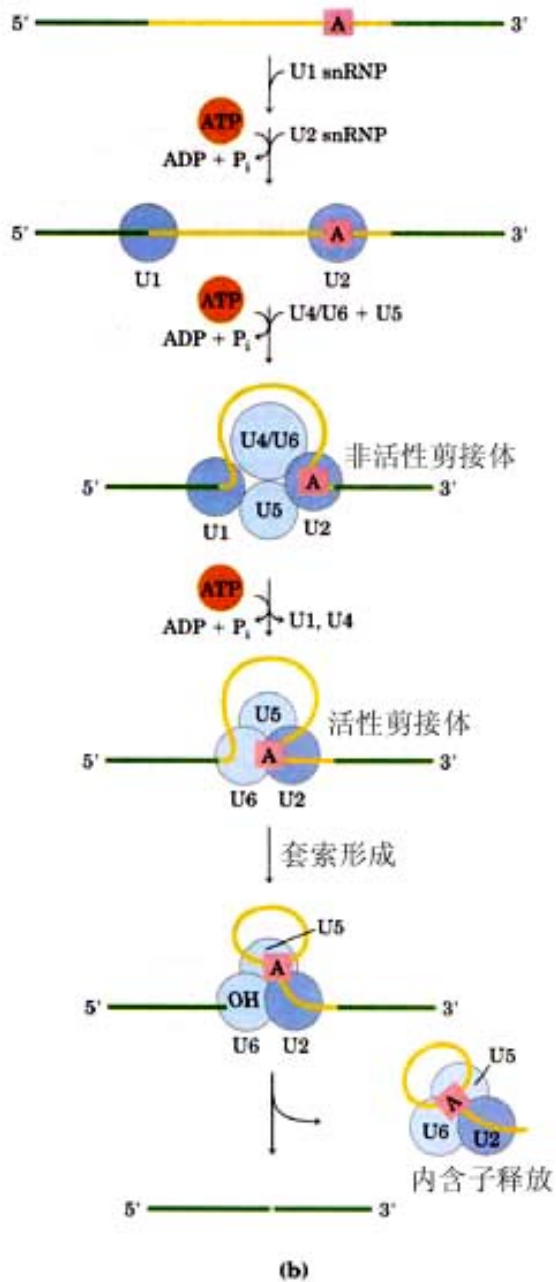


图3-23 真核生物 mRNA前体中内含子剪接过程示意图。



由U1 snoRNA以碱基互补的方式识别mRNA前体5'剪接点，由结合在3'剪接点上游富嘧啶区的U2AF（U2 auxiliary factor）识别3'剪接点并引导U2 snRNP与分支点相结合，形成剪接前体（pre-spliceosome）。

剪接前体进一步与U4、U5、U6 snRNP三聚体相结合，形成剪接体。



哺乳动物细胞中mRNA前体上的snRNP是从5'向下游“扫描”，选择在分支点富嘧啶区3'下游的第一个AG作为剪接的3'受点。

AG前一位核苷酸可以影响剪接效率，一般说来，  
**CAG=UAG>AAG>GAG。**

如果mRNA前体上同时存在几个AG，可能发生剪接竞争。



在高等真核生物个体发育或细胞分化过程中可以有选择性地越过某些外显子或某个剪接点进行变位剪接，产生出组织或发育阶段特异性mRNA，称为内含子的变位剪接。

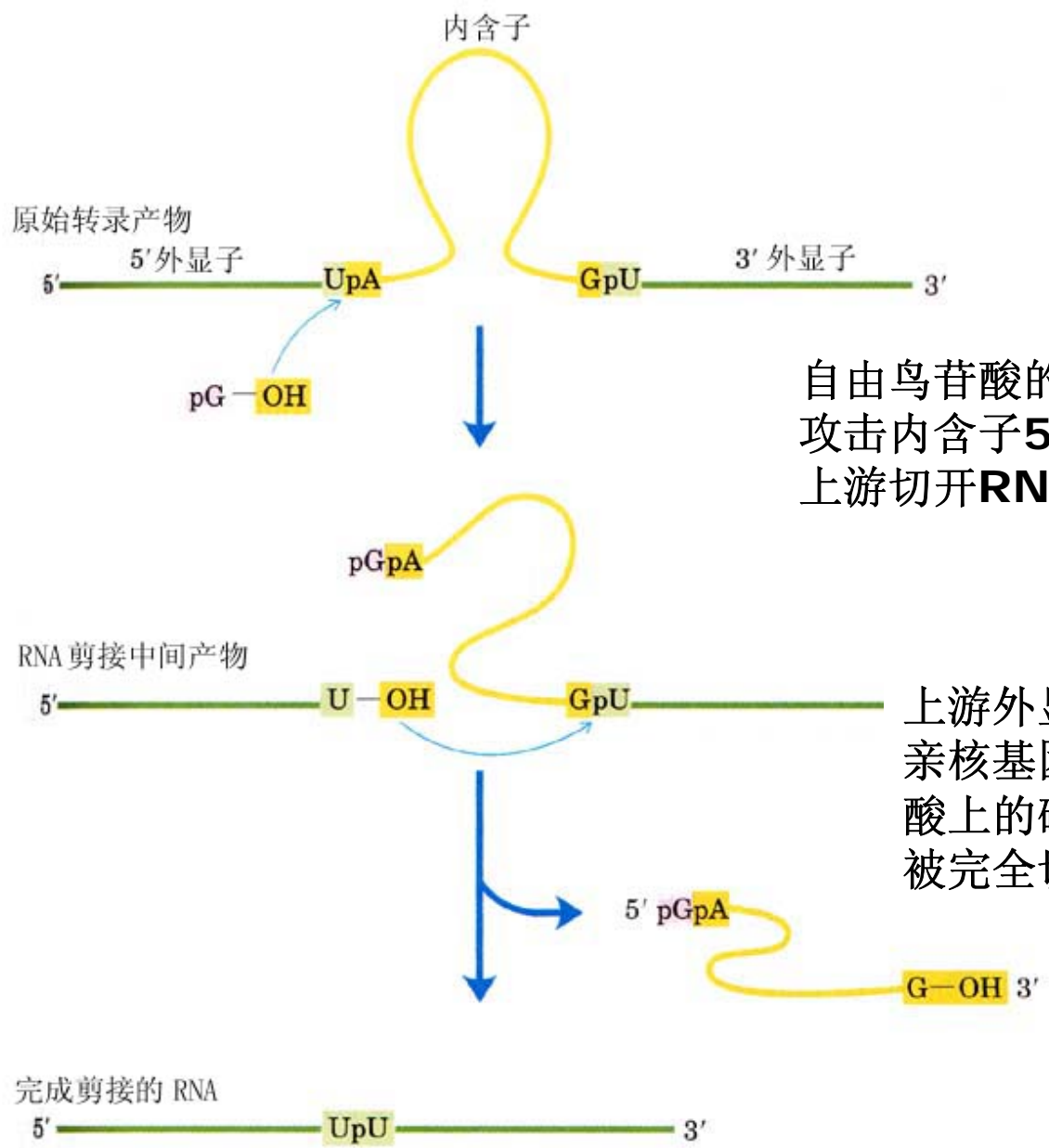
脊椎动物中大约有5%的基因能以这种方式进行剪接，保证各同源蛋白质之间既具有大致相同的结构或功能域，又具有特定的性质差异，进而拓展了基因所携带的遗传信息。



与mRNA前体中主要（GU-AG类）和次要（AU-AC类）内含子的剪接方式不同的是I、II类内含子，因为带有这些内含子的RNA本身具有催化活性，能进行内含子的自我剪接。



在I类内含子切除体系中，鸟苷或鸟苷酸的3'-OH作为亲核基团攻击内含子5'端的磷酸二酯键，从上游切开RNA链。再由上游外显子的自由3'-OH作为亲核基团攻击内含子3'位核苷酸上的磷酸二酯键，使内含子被完全切开，上下游两个外显子通过新的磷酸二酯键相连。



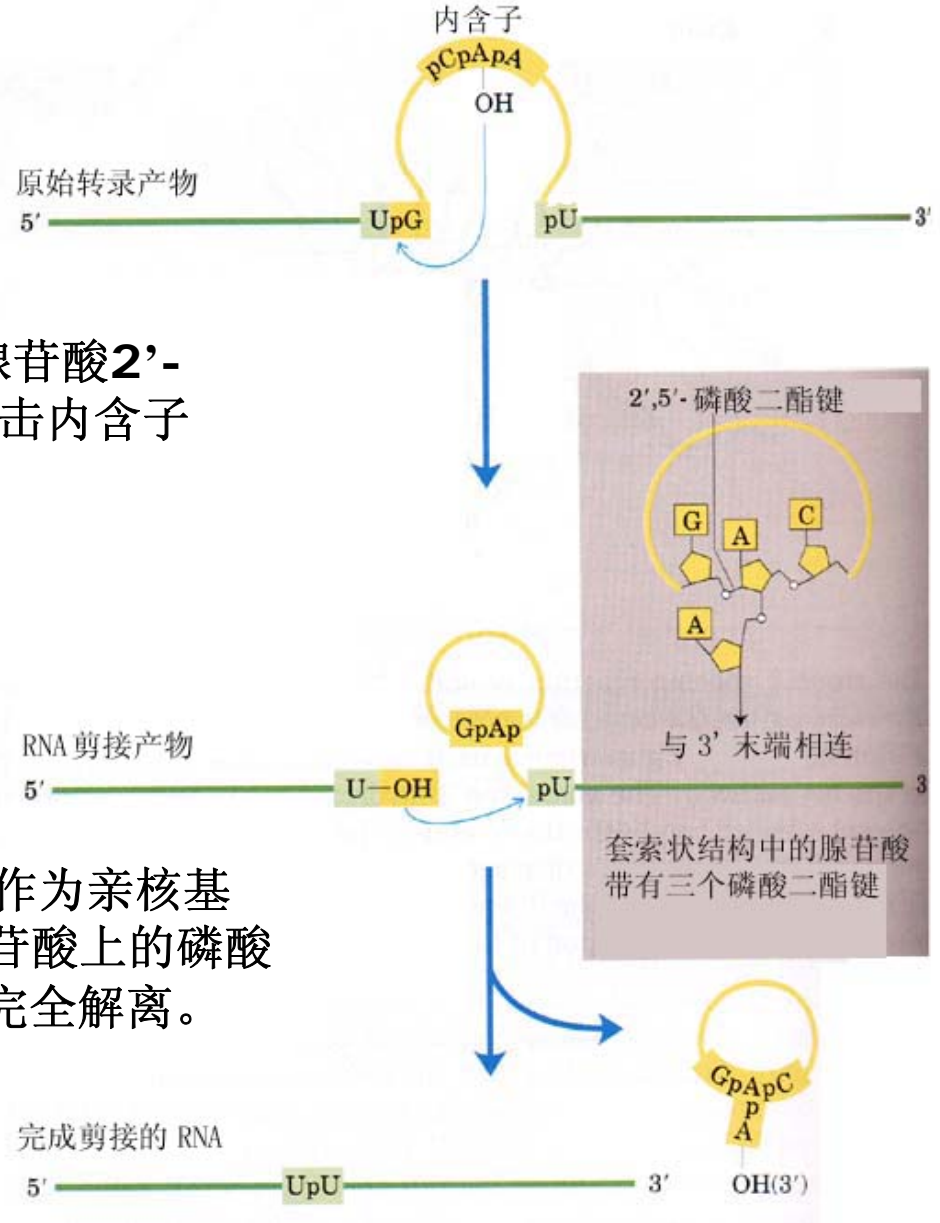
自由鸟苷酸的3'-OH作为亲核基团攻击内含子5'端的磷酸二酯键，从上游切开RNA链。

上游外显子的自由3'-OH作为亲核基团攻击内含子3'位核苷酸上的磷酸二酯键，使内含子被完全切开。

图3-24 I类内含子的自我剪接过程



在II类内含子切除体系中，内含子本身的某个腺苷酸2'-OH作为亲核基团攻击内含子5'端的磷酸二酯键，从上游切开RNA链后形成套索状结构。再由上游外显子的自由3'-OH作为亲核基团攻击内含子3'位核苷酸上的磷酸二酯键，使内含子被完全切开，上下游两个外显子通过新的磷酸二酯键相连。



内含子本身的某个腺苷酸2'-OH作为亲核基团攻击内含子5'端的磷酸二酯键。

上游外显子的3'-OH作为亲核基团攻击内含子3'位核苷酸上的磷酸二酯键，使套索结构完全解离。

图3-25 II类内含子的自我剪接过程

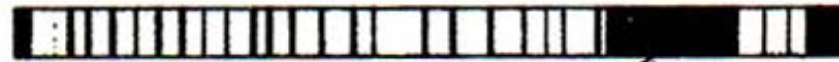


### 3.6.3 RNA的编辑和化学修饰

RNA的编辑（RNA editing）是某些RNA，特别是mRNA的一种加工方式，它导致了DNA所编码的遗传信息的改变，因为经过编辑的mRNA序列发生了不同于模板DNA的变化。



载脂蛋白B基因有29个外显子



CAA

第2153密码子编码谷氨酰胺



编辑

CAA

UAA



肝中剪接的mRNA编码  
含4563个残基的蛋白质



肠mRNA有UAA密码子在  
2153位密码子终止合成

图3-26 哺乳动物载脂蛋白基因转录产物的编辑



载脂蛋白基因编码区共有4563个密码子。在肝脏中，该基因转录产生完整的mRNA并被翻译成有4563个氨基酸的全长蛋白质，相对分子质量为 $5.1 \times 10^5$ 。在肠中合成的却是只包含有2153个密码子的mRNA，翻译产生相对分子质量为 $2.5 \times 10^5$ 的蛋白质。



研究发现，该蛋白其实是全长载脂蛋白的N端，它是由一个在序列上除了2153位密码子从CAA突变为UAA之外完全与肝脏mRNA相同的核酸分子所编码的，C→U突变使编码谷氨酰胺的密码子变成了终止密码子。



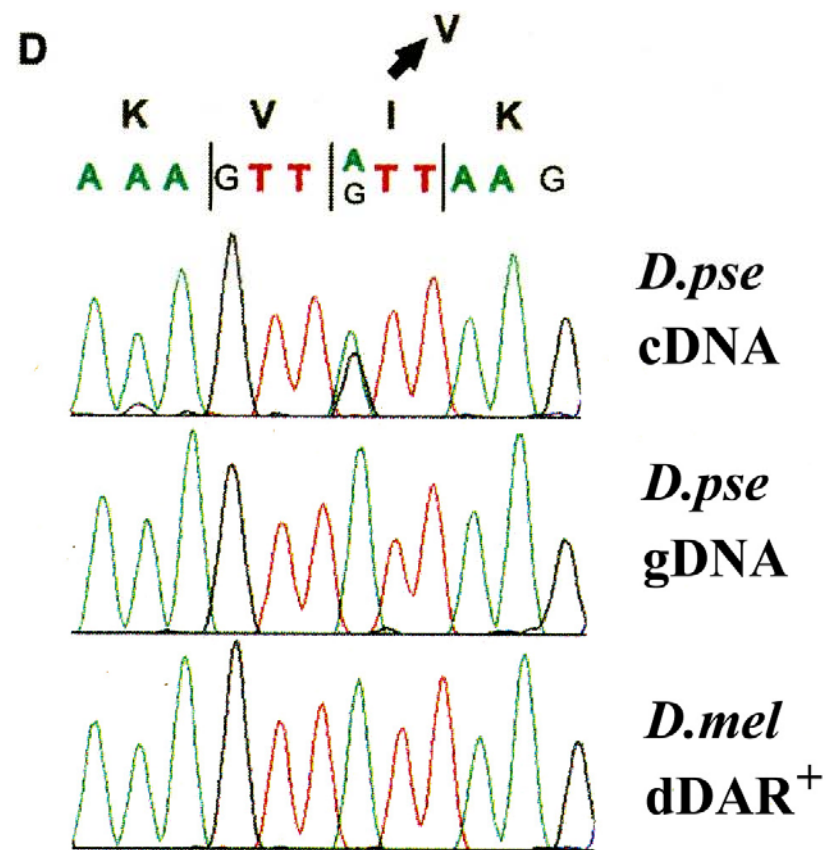
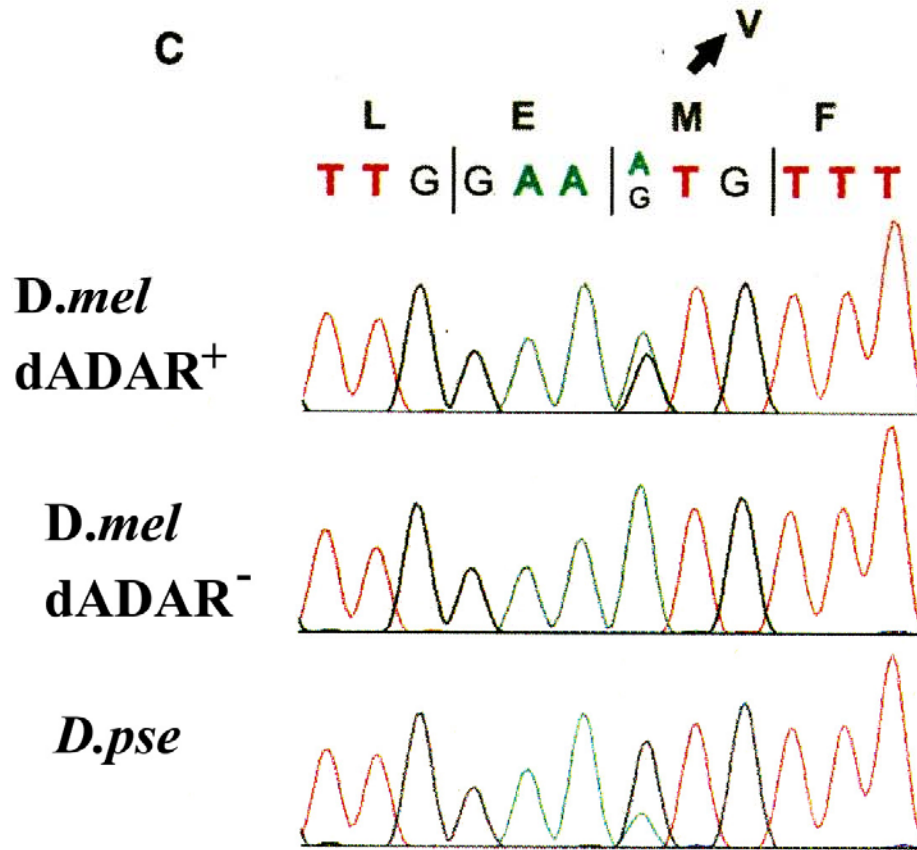
RNA的编辑虽然不是很普遍，但在真核生物中却时有发生（表3-10）。

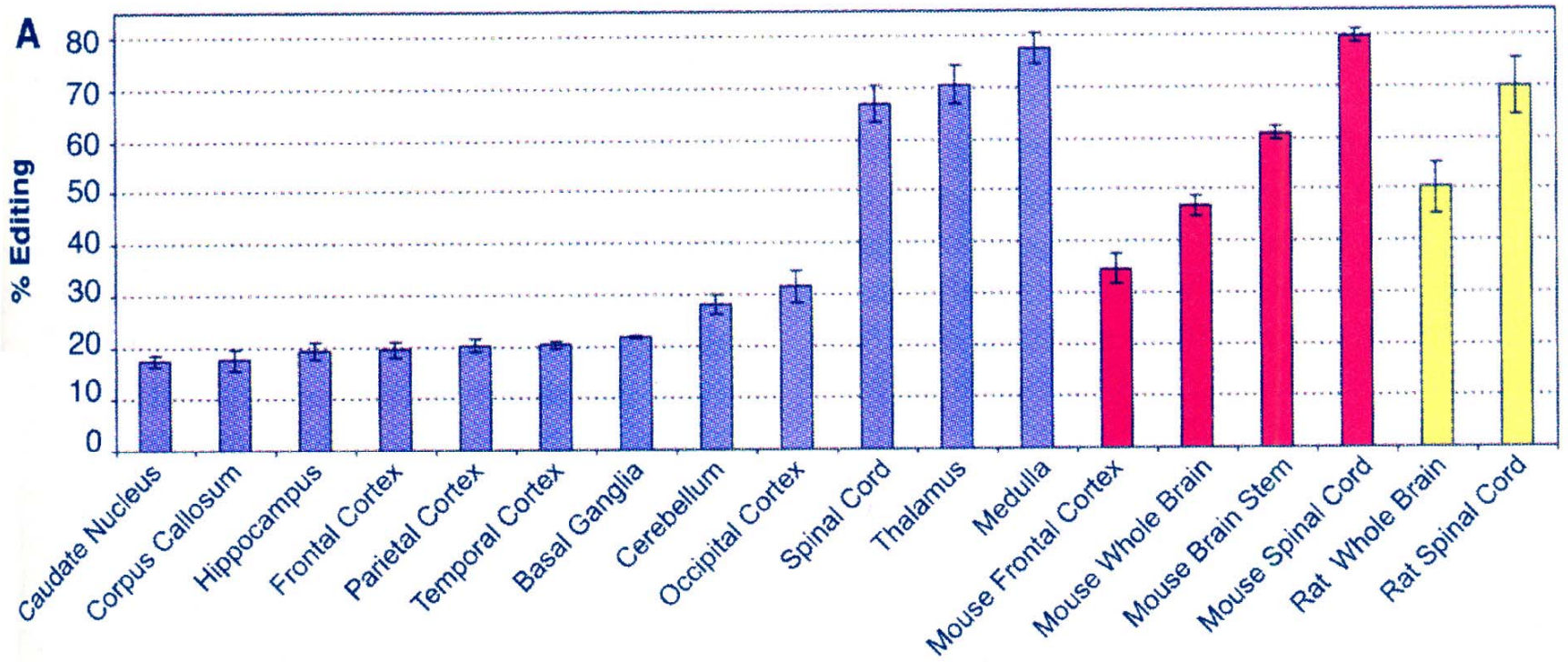
大鼠脑中谷氨酸受体蛋白mRNA经编辑后，多个编码谷氨酸的密码子变成了在控制通过神经递质的离子流过程中有重要影响的精氨酸，表明RNA的编辑可能是充分发挥生理功能所必须的。



# 表3-10 哺乳动物中RNA编辑的实例

组织	靶标RNA	所改变的碱基	结果
肝脏, 肠	载脂蛋白B	C→U	谷氨酰胺密码子→终止子
肌肉	半乳糖苷酶	U→A	苯丙氨酸密码子→酪氨酸
睾丸, 肿瘤等	Wilms肿瘤基因-1	U→C	亮氨酸密码子→脯氨酸
肿瘤	神经纤维瘤基因-1	C→U	精氨酸密码子→终止子
脑	谷氨酸受体蛋白	A→I	多个谷氨酸密码子→精氨酸







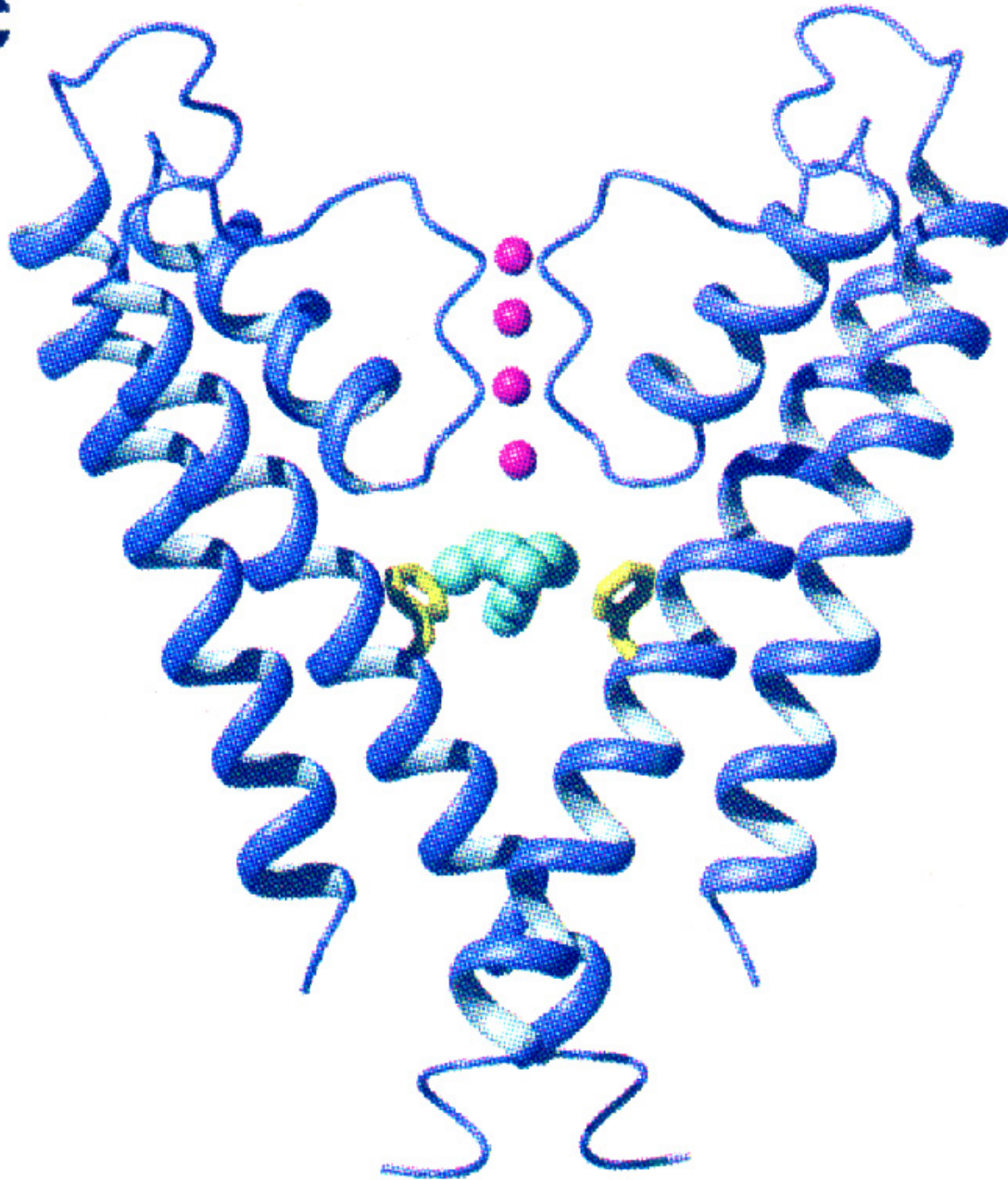
**B**

**hKCNA1**  
**Sqkv1a**  
**dShaker**  
**TuKv1**  
**ceShaw**  
**dShaw**  
**dShal**  
**dShab**  
**KcsA**

\*                      ▼  
**KIVGSLCAIAGVLTIALPVPVIVSNFN**  
**KLVGSLCAIAGVLTIALPVPVIVSNFN**  
**KIVGSLCVIAGVLTIALPVPVIVSNFN**  
**KVVGSLCAITGVLCIALPVPVIVSNFN**  
**RLVGSLCAVMGVLTIALPVPVIVSNFA**  
**MFVGALCALAGVLTIALPVPVIVSNFA**  
**KIVGGVCSLSGVLVIALPVPVIVSNFS**  
**KVIGTVCCICGVLVIALPIPIIVNNFA**  
**RLVAVVVMVAGITSFGLVTAALATWFFV**



C





RNA编辑的另一种形式是尿苷酸的缺失和添加。研究发现，利什曼原虫属细胞色素b mRNA中含有许多独立于核基因的尿嘧啶残基，而特异性插入这些残基的信息来自指导RNA (guide RNA)。



指导RNA与被编辑区及其周围部分核酸序列虽然有相当程度的互补性，但该RNA上存在一些未能配对的腺嘌呤，形成缺口，为插入尿嘧啶提供了模板（图3-27）。反应完成后，指导RNA从mRNA上解离下来，而mRNA则被用做翻译的模板。

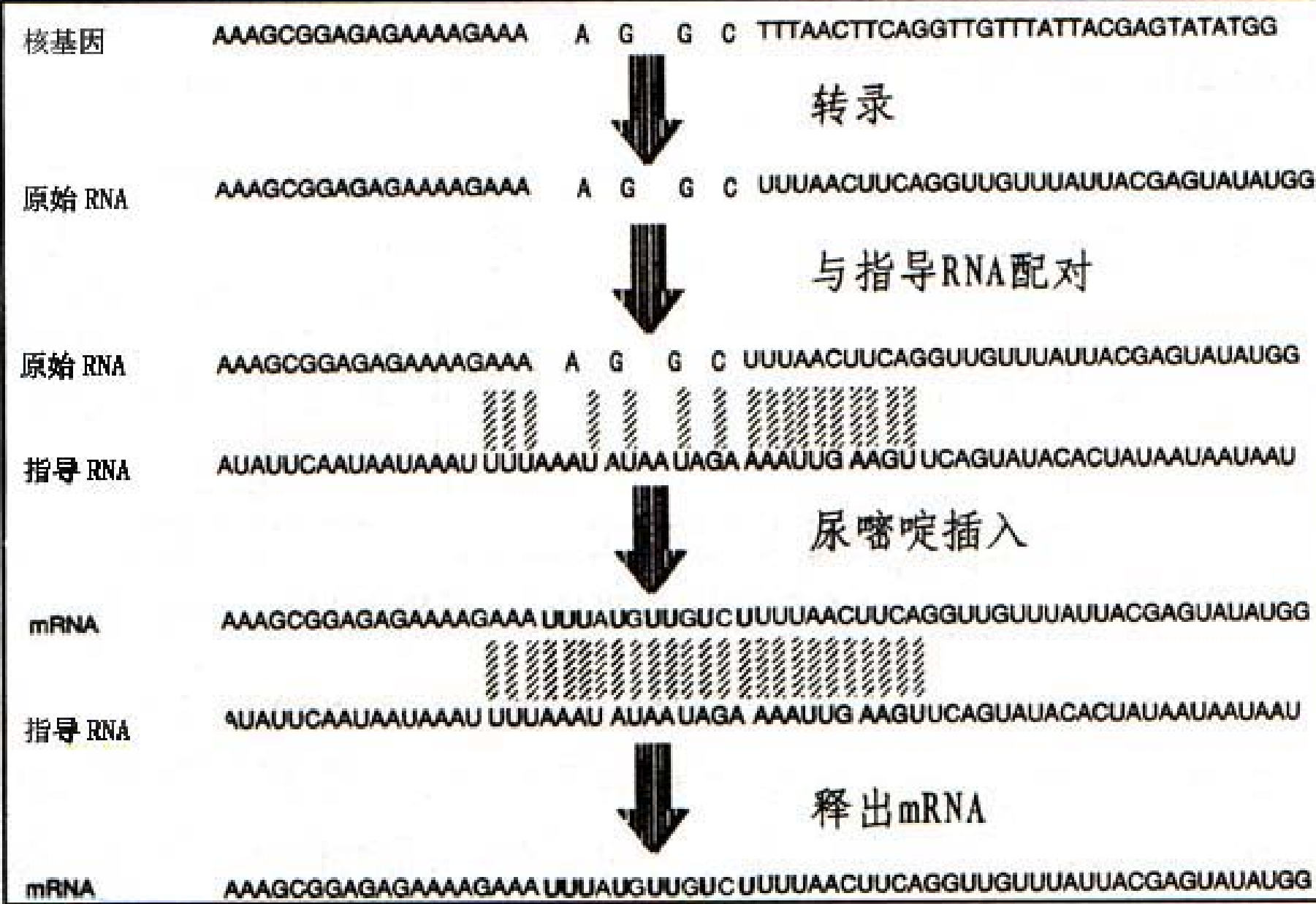


图3-27 指导RNA和RNA的编辑机制



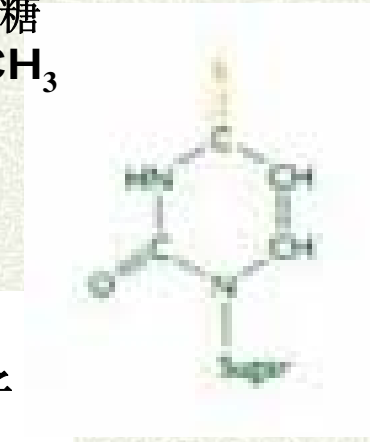
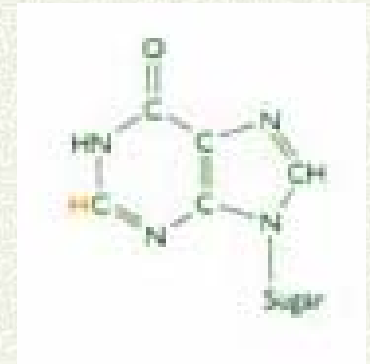
除了RNA的编辑之外，有些RNA，特别是前体rRNA和tRNA，还可能有特异性化学修饰。

图3-28是最常见的6大类化学修饰途径及其产物。研究证实，仅人细胞内rRNA分子上就存在106种甲基化和95种假尿嘧啶产物，虽然尚未发现特征性保守序列，但有实验证据表明，RNA的化学修饰可能具有位点特异性。



甲基化  
在核苷酸的碱基或核糖基上加一个或多个-CH<sub>3</sub>

去氨基化  
从碱基上去掉氨基

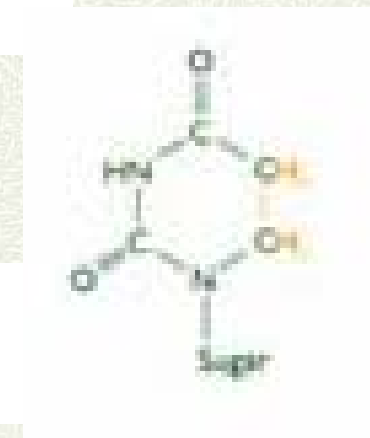


硫代  
用硫取代碱基分子上的氧

碱基同分异构化，碱基环结构上发生分子替代



二价键的饱和化  
使二价键饱和



核苷酸的替代  
用不常见核酸替换常见核苷酸

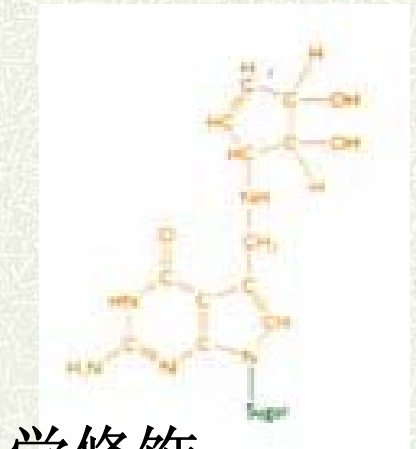


图3-28 rRNA和tRNA分子中的核苷酸化学修饰



有实验表明，分子量只有70-100个核苷酸的核仁RNA（snoRNAs）参与RNA的化学修饰，因为这些RNA能通过碱基配对的方式，把rRNA分子上需要修饰的位点找出来（图3-29）。

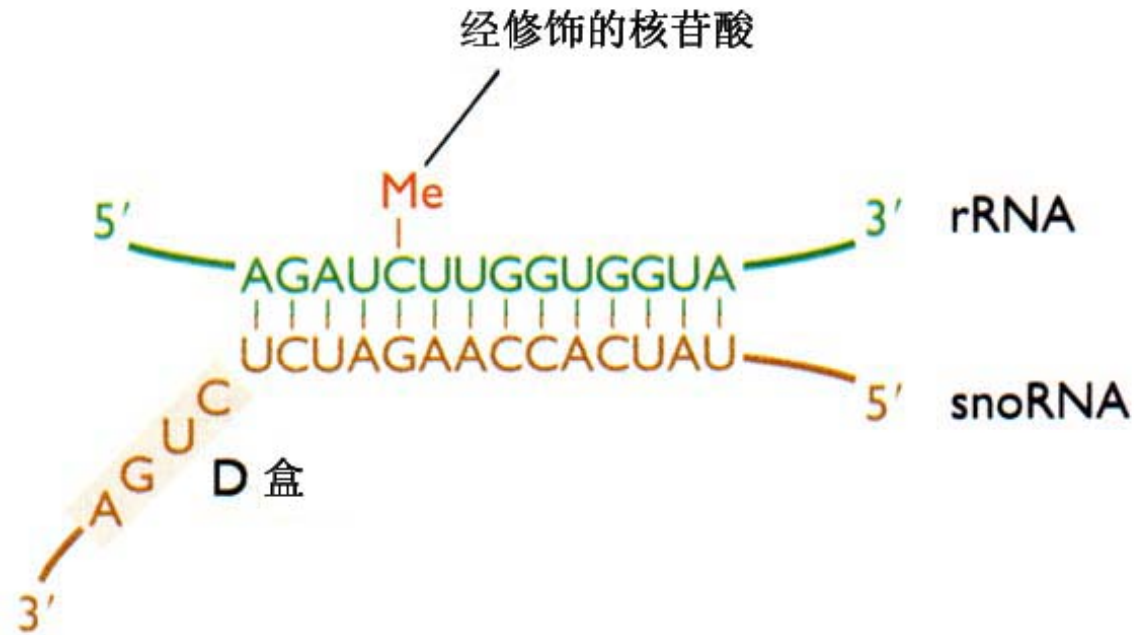


图3-29 酵母U24 snoRNA指导的甲基化



- 1、请说出生物体内**RNA**的种类和功能。
- 2、与**mRNA**序列相同的那条链是什么链？
- 3、什么是**DNA**模板与**mRNA**及蛋白质产物之间的共线性关系？
- 4、转录一般被分为哪几个步骤？



- 5、原核生物**RNA**聚合酶哪有几个亚基组成？
- 6、说说启动子的一般构造。
- 7、请说说转录终止子与翻译终止密码的结构特点。