



## 遗传物质的分子基础

什么是遗传物质？20世纪初，从1900年孟德尔论文的重新发现到1910年摩尔根的果蝇伴性遗传实验以来，生物学家一直在寻找答案。1928年 **Griffith** 等的肺炎链球菌的转化实验，1944年 **Avery** 等的单因子转化实验，1952年 **Hershey** 和 **Chase** 的噬菌体感染实验，1957年 **Fraenkel-Conrat** 等的烟草花叶病毒重建实验，先后证实 **DNA** 是遗传物质，在不具备 **DNA** 的病毒中，**RNA** 是遗传物质，即核酸是遗传物质，是遗传信息的载体。紧接着要回答的第二个问题是，遗传物质的分子基础是什么？1953年 **Watson** 和 **Crick** 的 **DNA** 双螺旋结构模型的提出，给出了圆满的答案。第三个问题是遗传信息如何传递？基因如何表达？这得益于“中心法则”的提出和遗传密码的破译以及“**DNA** 半保留半不连续复制模型”的证明。与此同时，开展了对真核生物染色体结构和功能的探讨。随着现代遗传学、分子生物学和基因组学研究的深入进行，我们对基因的表达调控、**RNA** 的转录与加工、蛋白质的翻译、基因的结构和功能等有了越来越深入的了解和认识。

## 3.1 核酸是遗传物质

### 3.1.1 肺炎链球菌的转化实验

DNA是遗传物质的概念源于1928年Griffith等进行的肺炎链球菌(*Streptococcus pneumoniae*旧称肺炎双球菌*Diplococcus pneumoniae*)的转化实验。肺炎链球菌有两种类型,一种为光滑型(S型),能在琼脂固体培养基上形成大而光滑闪亮的菌落,细胞外包有一层多糖荚膜,是一种致病细菌。人感染后引起肺炎,小鼠感染后产生败血症而死亡。另一种为粗糙型(R型),菌落粗糙,细胞外无荚膜,不能致病。在实验中他们发现了肺炎链球菌的转化作用。将加热灭活的SIII型菌液和活的RII型菌液分别注射小白鼠,都不致死;但当将加热灭活的SIII型菌液与活的RII型菌液混合后,注射小白鼠,结果小白鼠都死亡。此时,可以在死亡小白鼠的体内或血液中分离到大量活的SIII型肺炎链球菌。这说明灭活的SIII型菌中的遗传物质使RII型菌转化为SIII型菌,这种改变遗传性状的现象称为细菌的转化(transformation)。利用从人体内得到的R型和S型肺炎链球菌抽提液的混合物注射小白鼠也能观察到致死结果,表明在抽提液中含有一种转化因子。由于Griffith等没有做单因子转化实验,当时还不能说明何种物质引起转化作用。

直到1944年Avery等三位美国科学家不仅在体外成功重复了上述实验,而且用生物化学的方法对S型菌提取液的所有成分分离后,进行了单因子转化实验,证明转化因子是DNA,而不是多糖荚膜、蛋白质和RNA,而且转化频率随着DNA的纯度的提高而增加。Avery等的实验结果首次证明转化因子是DNA,取得了DNA是遗传物质的第一个和最重要的一个证据,明确了DNA是遗传信息的载体。

### 3.1.2 噬菌体感染实验

1952年Hershey和Chase用标记放射性同位素的方法,进行了噬菌体感染实验,为证明DNA是遗传物质提供了更直接的证据。

T2噬菌体是感染大肠杆菌(*E. coli*)的一种噬菌体,它由蛋白质外壳和DNA核心构成。在噬菌体中,其蛋白质是唯一含硫(S)的物质,而DNA是唯一含磷(P)的物质。利用这一特性,他们在放射性<sup>32</sup>P和<sup>35</sup>S存在的情况下,使噬菌体进行繁殖,再用这种带有放射性物质标记的噬菌体去感染无放射性的大肠杆菌,几分钟后离心除去未吸附的噬菌体,再利用捣碎机捣碎使噬菌体与大肠杆菌分开。对其进行离心,发现从大肠杆菌表面释放的噬菌体外壳在上清液中,它们由蛋白质组成,含有80%放射性标记<sup>35</sup>S,而大肠杆菌在沉淀中,含有80%放射性标记<sup>32</sup>P。实验表明在噬菌体感染过程中,只有DNA进入细菌细胞,而蛋白质外壳留在细菌体外。同时,感染产生的子代噬菌体颗粒含有亲代<sup>32</sup>P活性的30%,但仅含不到1%的亲代噬菌体蛋白质,说明侵入的DNA在短时间内便繁殖出同原来一样的具有完整蛋白质外壳的子代噬菌体。该实验证明具有遗传作用的是DNA而不是蛋白质。

### 3.1.3 烟草TMV的重建实验

能感染烟草的烟草花叶病毒(tobacco mosaic virus, TMV)由蛋白质外壳和RNA核心组成。从TMV病毒分别提取它的蛋白质和RNA,再将它们放在一起,可以得到具有感染力的病毒颗粒。1957年, Fraenkel-Conrat等人将两个不同的TMV株系(S株系和HR株系)的蛋白质和RNA分别提取出来,然后相互对换,将S株系的蛋白质和HR株系的RNA,或反过来将HR株系的蛋白质和S株系的

RNA放在一起,重建形成两种杂种病毒,去感染烟草叶片。结果是叶片上所产生的病斑跟所含 RNA密切相关,是什么样的 RNA就形成什么样的子代病毒粒子,即子代病毒粒子的外壳蛋白是由 RNA决定的,而不是由蛋白质决定的。这个实验令人信服地证明在不具有 DNA的病毒中, RNA是遗传物质。

## 3.2 核酸的分子结构

### 3.2.1 核酸的分子组成

核酸是遗传信息的载体,也是生物体的基本组成物质。生物界的核酸根据其化学组成可以分为两大类: DNA(*deoxyribonucleic acid*, 脱氧核糖核酸)和 RNA(*ribonucleic acid*, 核糖核酸)。DNA和RNA的基本结构单位都是核苷酸。核苷酸由碱基、戊糖及磷酸基3部分组成。

DNA和RNA都含有4种主要碱基,其中腺嘌呤(*adenine*, A)、鸟嘌呤(*guanine*, G)和胞嘧啶(*cytosine*, C)是两者共有的,而尿嘧啶(*uracil* U)为RNA特有,胸腺嘧啶(*thymine*, T)为DNA所特有。DNA的碱基组成具有很明显的特征,即  $A/T = G/C = 1$ ,而RNA的碱基组成一般没有这种特点。DNA和RNA都含有戊糖,其本质的区别在于它们含有的戊糖的类别不同: RNA含有D-核糖,而DNA含有2'-脱氧-D-核糖。

### 3.2.2 DNA的分子结构

DNA的一级结构是指核苷酸在DNA分子中的排列顺序。在DNA的一级结构中因为各种脱氧核苷酸的脱氧核糖和磷酸都是相同的,所以碱基顺序也就代表了核苷酸顺序。

DNA的二级结构是指DNA通过分子间相互作用形成的双链或双螺旋分子,即DNA双螺旋结构(*double helical structure*)。1953年,Watson和Crick根据Wilkins及Franklin对DNA纤维X射线衍射图研究证明DNA分子具有螺旋结构的信息,Chargaff等发现的DNA中碱基含量  $A = T, G = C, A + G = T + C$ 的定律,以及对4种碱基的物化数据的分析结果等,提出了著名的DNA双螺旋结构模型(*double helical model of DNA*)。不朽的DNA双螺旋结构模型的诞生具有划时代的意义,成为遗传学发展史上最重要的里程碑之一。

DNA的高级结构是指DNA的超螺旋(*supercoil*)结构和染色体DNA所具有的复杂折叠状态。超螺旋是DNA双螺旋的螺旋轴盘绕而形成的螺旋,是DNA三级结构的一种形式。DNA的三级结构是指双螺旋链的扭曲。例如在B-DNA双螺旋中,每10个核苷酸长度旋转一圈,这时双螺旋最稳定,处于能量最低态。

生物体中绝大多数DNA以超螺旋的形式存在。超螺旋是具有方向性的,有正超螺旋和负超螺旋之分。双螺旋DNA的松开导致形成负超螺旋,而DNA的旋紧则导致形成正超螺旋。正超螺旋和负超螺旋这两种形式在一定条件下可以相互转变。天然的DNA都呈负超螺旋,但在体外溴乙锭、放线菌素D等扁平分子(致癌物)可以嵌入DNA的碱基对之间,随着嵌入分子数量的增加,可以使负超螺旋变为松弛型,再变为正超螺旋。在遗传学研究中,基因转录、DNA复制、DNA损伤与修复、重组等过程均涉及超螺旋的变化。

### 3.2.3 RNA的分子结构

细胞内的RNA分子绝大多数以单链形式存在,碱基配对常常形成分子内的二级结构——双螺旋,这种结构是其行使功能的分子基础。RNA的分子内双螺旋由单链区段回折形成,呈现类似

茎环的结构,由于回折区段不完全的碱基互补,常常出现不完全配对或配对间断。在各种类型的 RNA 分子中由于不同的茎-环结构与单链区段的不同结合形成各种特定的 RNA 的高级结构形式。

### (1) tRNA 分子

自 1965 年 Holley 等首次测定酵母丙氨酸 tRNA 的一级结构以来,至今已经做出全序列分析的 tRNA 有 3 300 种以上。几乎所有的 tRNA 都是小分子,长 75~94 个核苷酸,5' 端有末端磷酸化基团,3' 端为 CCA-OH 序列,分子中稀有碱基含量较高,而且许多被甲基化修饰;分子内部有碱基互补区,都能形成三叶草形二级结构。该二级结构一般由氨基酸接受臂、二氢尿嘧啶环(D 环)、反密码子环和 T $\Psi$ C 环组成,有的还在反密码子环和 T $\Psi$ C 环之间存在额外环[图 3-1(a)]。tRNA 的三级结构都为倒 L 形[图 3-1(b)]。

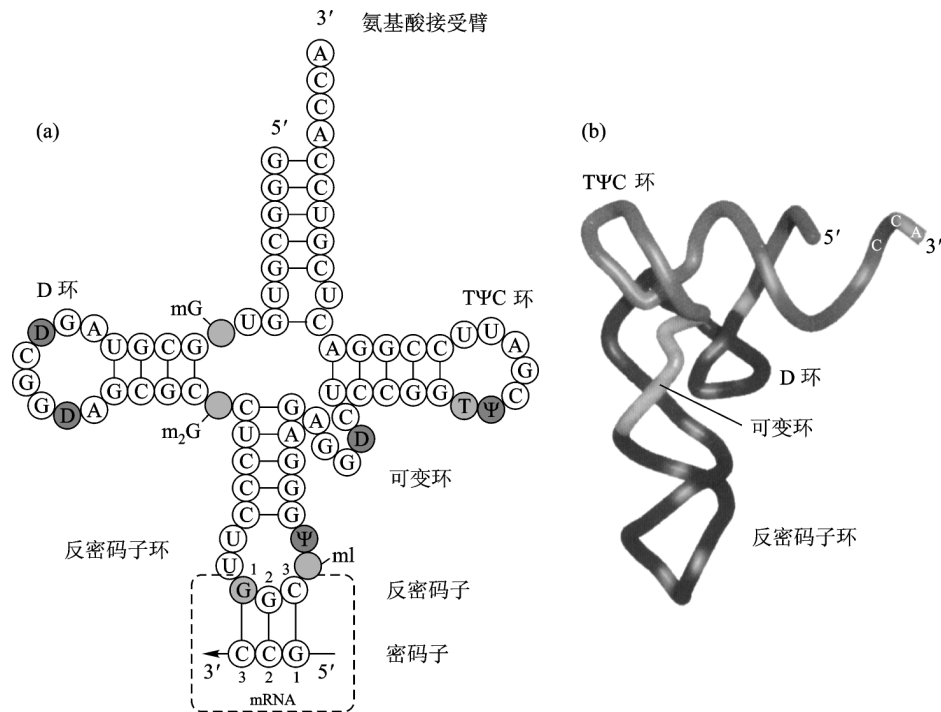


图 3-1 tRNA 分子结构(引自 Lodish, 2004)

(a) tRNA 三叶草形二级结构 (b) tRNA 倒 L 形的三级结构

### (2) rRNA 分子

大肠杆菌的 5S、16S 和 23S rRNA 的一、二级结构分别含有 120、1 542 和 2 904 个核苷酸。大肠杆菌的 5S rRNA 的二级结构模型表明,它由 120 个左右的核苷酸组成 5 个双螺旋区和其他单链区(图 3-2),含有两个功能域,一个功能域中含有 5' CGAAC 序列,这是与 tRNA 分子 T $\Psi$ C 环上的 GT $\Psi$ CG 序列相互作用的部位,使 tRNA 与核糖体结合。另一个功能域与 23S rRNA 中的一段序列互补,对核糖体大亚基与 rRNA 的相互作用有重要意义。

不同来源的 16S rRNA 具有相似的二级结构,大肠杆菌 16S rRNA 的二级结构主要分为 4 个功能域:5' 端功能域、中心功能域、3' 端大功能域和 3' 端小功能域。在这些功能域中有与 mRNA 互补的部位,有结合肽酰 tRNA 的部位,有些则在 30S 与 50S 亚基结合中起作用等,涉及多种识别与作用功能。

研究表明,RNA 的二级结构在基因的表达与调控过程中起着十分重要的作用,例如, rRNA 与 mRNA 间的碱基配对控制着蛋白质的起始; rRNA 与 mRNA 间的碱基配对促进了翻译过程; rRNA 发

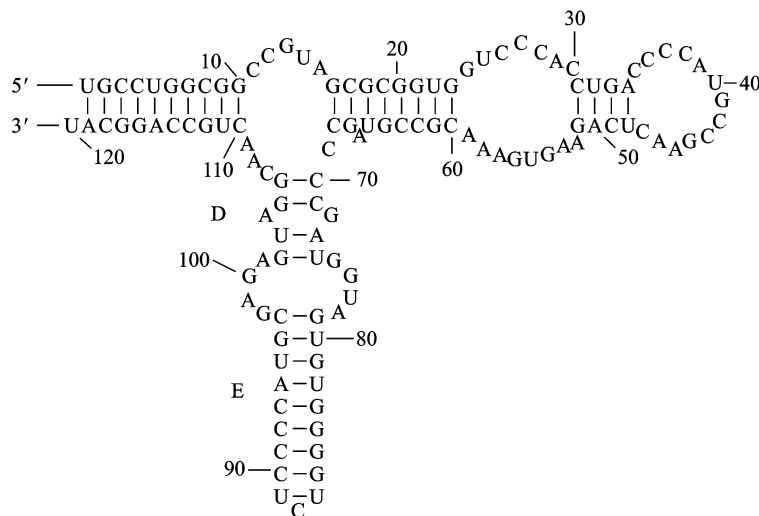


图 3-2 大肠杆菌 5S rRNA 的二级结构模型

夹结构及茎环结构控制了转录的终止、翻译的效率以及 mRNA 的稳定性；RNA-RNA 间的碱基配对在内含子的剪切过程中也起着重要的作用。特别是近年对小分子 RNA 的结构和功能的研究，揭示出它们在生物的发育、生长、繁殖和基因表达与调控中的多种重要功能(详见第 15 章)。

### 3.3 DNA 复制

#### 3.3.1 DNA 复制的基本规律

DNA 作为遗传物质应该具备两个基本功能，一是遗传信息在世代间的传递，从而保证物种的延续和遗传的稳定性；二是在细胞和个体的生长发育等过程中使遗传信息得以表达，从而表现与亲代相似的性状。必须准确进行 DNA 复制( replication)，即以原来的 DNA 分子为模板合成出相同的分子，通过亲代 DNA 分子的复制将遗传信息传递给子代。

1953 年 Watson 和 Crick 在提出 DNA 双螺旋结构模型后不久，接着又提出了 DNA 的复制机制——DNA 复制的半保留模型，其基本规则是：① 一般按半保留( semiconservative)半不连续方式进行。② 复制起始( initiation)在原点( origin)的特定序列上。③ 在复制的起点处控制复制。④ 复制叉( fork)的移动有单向或双向。⑤ 链的延伸方向只能是 5'端→3'端方向。⑥ 存在模板的条件下，DNA 聚合酶以短的 RNA 片段作为引物( primer)开始合成 DNA 的短片段。⑦ 存在各种 DNA 链的合成起始机制，除了 RNA 引发外，还存在其他的一些机制，包括 DNA 链与一个末端蛋白共价结合，以及缺口的共价延伸，或者亲本链已被环出的末端等。⑧ 终止也是在复制过程中的某个固定点。⑨ 复制的机制取决于基因组结构和构象来保持产生完整的染色体。⑩ 即使在同一个细胞内也可进行多种复制机制的操作。

#### 3.3.2 半保留半不连续复制

DNA 半保留复制模型的正确性，在 1958 年首先由 Meselson 和 Stahl 以大肠杆菌为材料，利用  $^{15}\text{N}$  及  $^{14}\text{N}$  同位素标记的实验所证实。此后，Cairns 1963 年对大肠杆菌 DNA 放射自显影实验以及 Taylor 在 1958 年进行的蚕豆染色体放射自显影实验，都说明真核生物与原核生物一样，DNA 的复制方式也

是按半保留方式进行的。

DNA 半保留半不连续复制 (DNA semiconservative and semidiscontinuous replication) 指的是亲代 DNA 分子的两条链, 在半保留复制过程中, 一条链按  $5' \rightarrow 3'$  端方向连续合成, 另一条链按  $5' \rightarrow 3'$  端方向不连续合成的方式, 合成一系列不连续的冈崎片段 (Okazaki fragment), 它是在蛋白质和酶促作用下进行的复杂的生化反应, 涉及 30 多种蛋白质的协同作用。参与 DNA 复制的关键酶有: DNA 聚合酶、引物(发)酶、DNA 连接酶、DNA 拓扑异构酶、DNA 解旋酶和单链 DNA 结合蛋白等(图 3-3)。无论真核生物还是原核生物, DNA 分子复制实质上是染色体的复制。

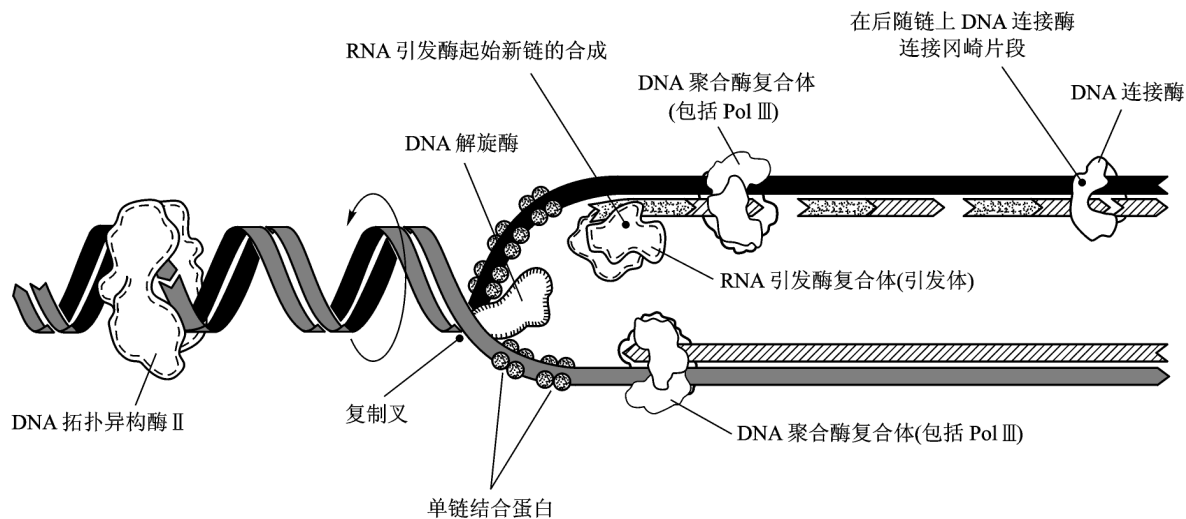


图 3-3 DNA 的半保留半不连续复制模型(引自 Hartl 1998)

### 3.3.3 环状双链 DNA 复制方式

#### (1) 滚环复制

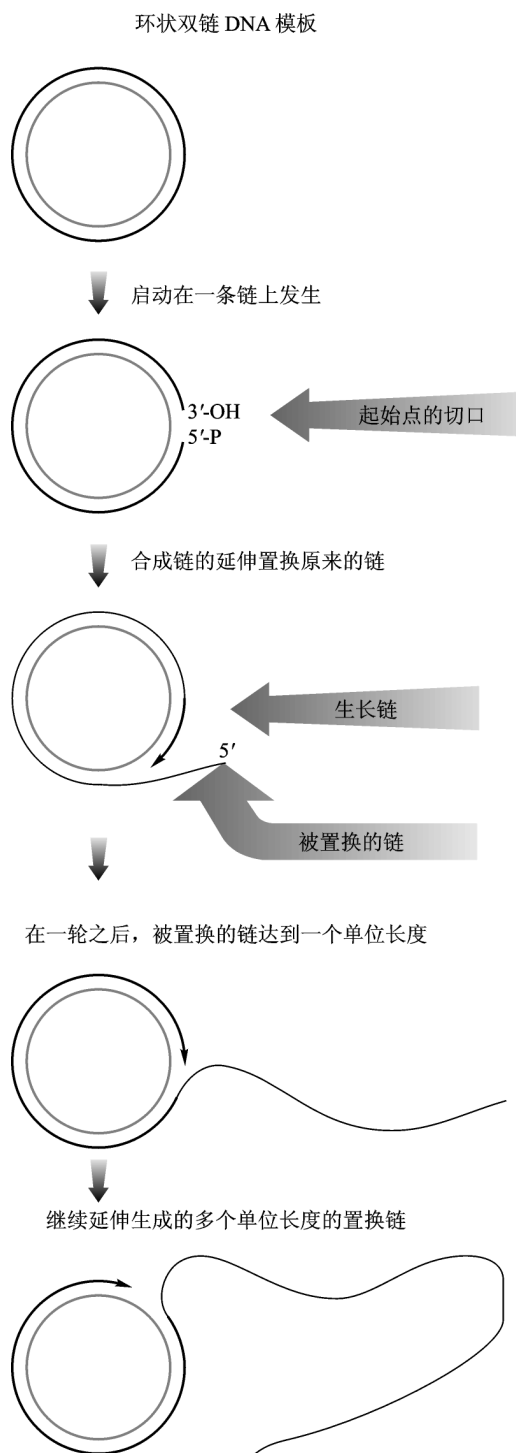
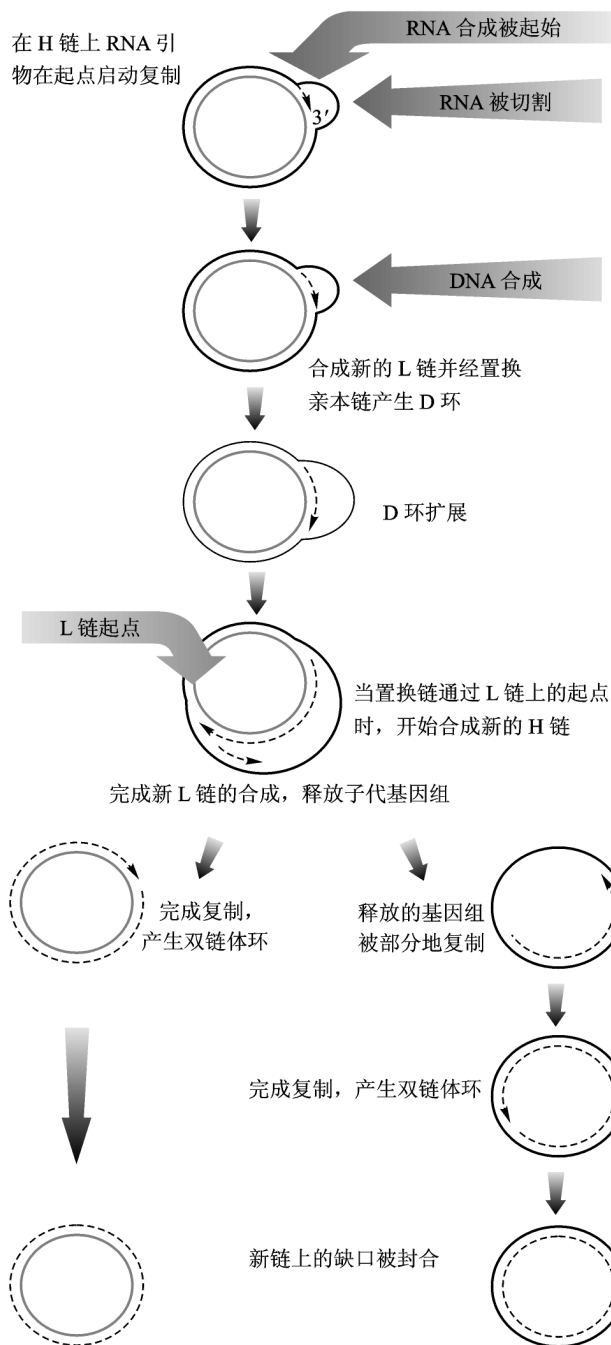
滚环复制又称  $\sigma$  复制, 该模型能解释双链及单链子代 DNA 是怎样从复制型产生的。滚环复制中, 在环状双链 DNA 的一条链上造成专一性缺口, 产生游离的  $3'-OH$  末端作为引物, 以另一条完整环状单链作为模板, 由 DNA 聚合酶催化延伸, 新合成的链沿着环状的模板链滚动; 这种环状结构称为滚环 (rolling circle)。随着它的滚动,  $3'$  端不断延长,  $5'$  端不断地被置换甩出而成为一条单链, 该单链可以延长出基因组的多个拷贝单位, 按照单位长度断裂, 就会产生多个单链线性拷贝(图 3-4)。线性状态可以以一条单链维持, 或通过合成互补链形成双链体。 $\lambda$  噬菌体的增殖、接合, 以及真核生物  $\text{mtDNA}$  的扩增都是以这种复制方式进行的。

#### (2) $\theta$ -型复制

大肠杆菌 DNA 环状双链分子在其 DNA 复制过程的中间产物, 在放射自显影观察时可形成一个  $\theta$  结构。这是由于从复制起点 ( $\text{oriO}$ ) 开始复制, 形成两个复制叉, 双向复制所产生的结果, 其外形酷似  $\theta$  字而得名。 $\theta$ -型复制需要 RNA 引物进行半保留半不连续复制。一条单链总是和模板链互补地结合在一起形成子链。

#### (3) 线粒体的 D 环复制

哺乳动物  $\text{mtDNA}$  复制是不对称的, 且双链不是同时复制。先复制双链中的一条链, 待该链复制到  $2/3$  的长度时, 另一条链才开始复制。这是由于重链和轻链上有各自的复制起点, 因而两条链的复制进度相差甚远。其复制起始的特点是, 在起始区形成 D 环 (displacement loop, D-loop) 结构, 即重链 (H 链) 和以重链为模板合成的新链组成的双链部分以及被新链替代出来的单链 (轻链, L 链) 部分形成

图 3-4 滚环复制模型(仿自 Lew *in*, 2004)图 3-5 线粒体 DNA 的 D 环复制模型(仿自 Lew *in*, 2004)

哺乳动物线粒体 DNA 的 D 环维持一种开放形式,  
每条单链的复制有各自的复制起点

的结构。D 环含有 500~600 bp, 它是不稳定的, 处于降解—合成状态, 这样来保持该区双链被打开, 即 D 环结构。其复制过程见图 3-5。图中可见线粒体 DNA(哺乳类)的 D 环维持开放的形式, 每条单链复制有各自的复制起点。

### 3.3.4 真核生物染色体端粒的复制

端粒(telomere)是真核生物染色体末端的一种特殊结构。端粒的主要功能是防止染色体末端受核酸酶的降解,维持染色体结构的稳定性和完整性,为线状染色体的末端复制提供基础。此外,端粒与染色体联会、细胞分裂和细胞衰老等也有密切关系。

端粒 DNA 的序列比较特殊,由一系列短的串联重复序列组成,可用  $G_n(A/T)_m$  的一般式来表示,其中  $n > 1$ ,  $m$  为  $1 \sim 4$ 。例如,四膜虫为 TTGGGG,线虫为 TTAGGC,哺乳类为 TTAGGG 等。端粒末端的这些特殊序列一般形成环状结构。研究表明当端粒(TTAGGG) $_n$  序列的 3' 单链末端序列折回,碱基配对取代其上游相同序列时,将形成一个类似 D 环的 t 环(telomere loop, t loop) 结构(图 3-6)。这是由于在该区域一系列被替换出的 TTAGGG 重复序列环出形成一个未配对的单链区,而折回的端粒尾端序列与该区域的同源链配对的结果。这种结构是由端粒结合蛋白(telomere binding protein, TBP) TRF  $\alpha$  与端粒的 TTAGGG 重复序列结合的一种蛋白)所催化形成的。该结构的形成提供了一个有序的高级结构,使凸出 3' 端单链埋藏在 DNA 分子内部以免与端粒酶接触,同时也保护了单链。

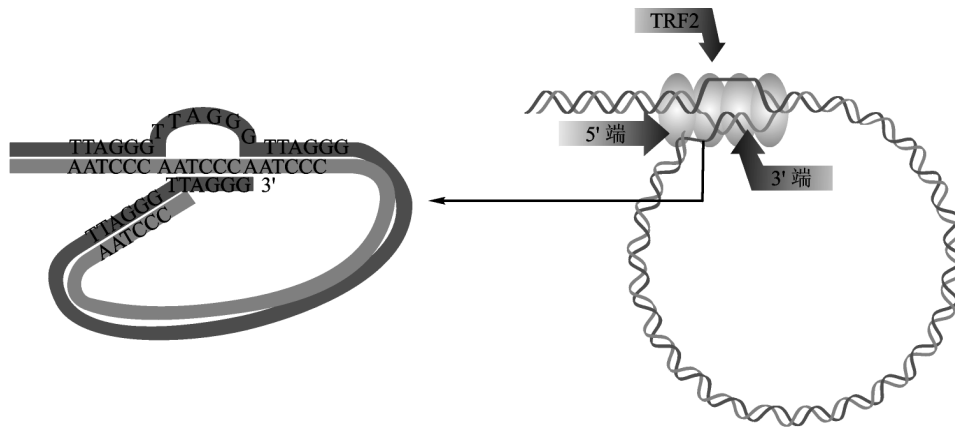


图 3-6 t 环结构(仿自 Lewin, 2004)

端粒的 3' 单链末端(TTAGGG)重复序列从双链体 DNA 中置换出同源序列,形成 t 环。该反应由 TRF2 催化

端粒的复制不是由 DNA 聚合酶完成的,而是由端粒酶(telomerase)催化合成的。端粒酶是一个大的核糖核蛋白分子,由多条多肽与一单个 RNA 分子组成。端粒酶 RNA 的核心序列是复制端粒 DNA 的模板,它是一富含 C 的  $15 \sim 22$  核苷酸重复序列。端粒酶是一种特殊的反转录酶,其活性只限于利用端粒酶特异的 RNA 作为模板。因此,端粒酶的作用是不连续的,模板 RNA 序列与 DNA 引物严格按照碱基互补定位结合,利用末端 G-T 发夹结构的 3'-OH 作为引物,反转录酶合成一段 DNA 序列后,端粒酶就要移位,进行下一段的催化合成。延伸后的染色体 3' 端可使复制子链长度保持与亲链相同(端粒的长度是由端粒结合蛋白调控的),由此解决冈崎片段 3' 端复制的问题(图 3-7)。



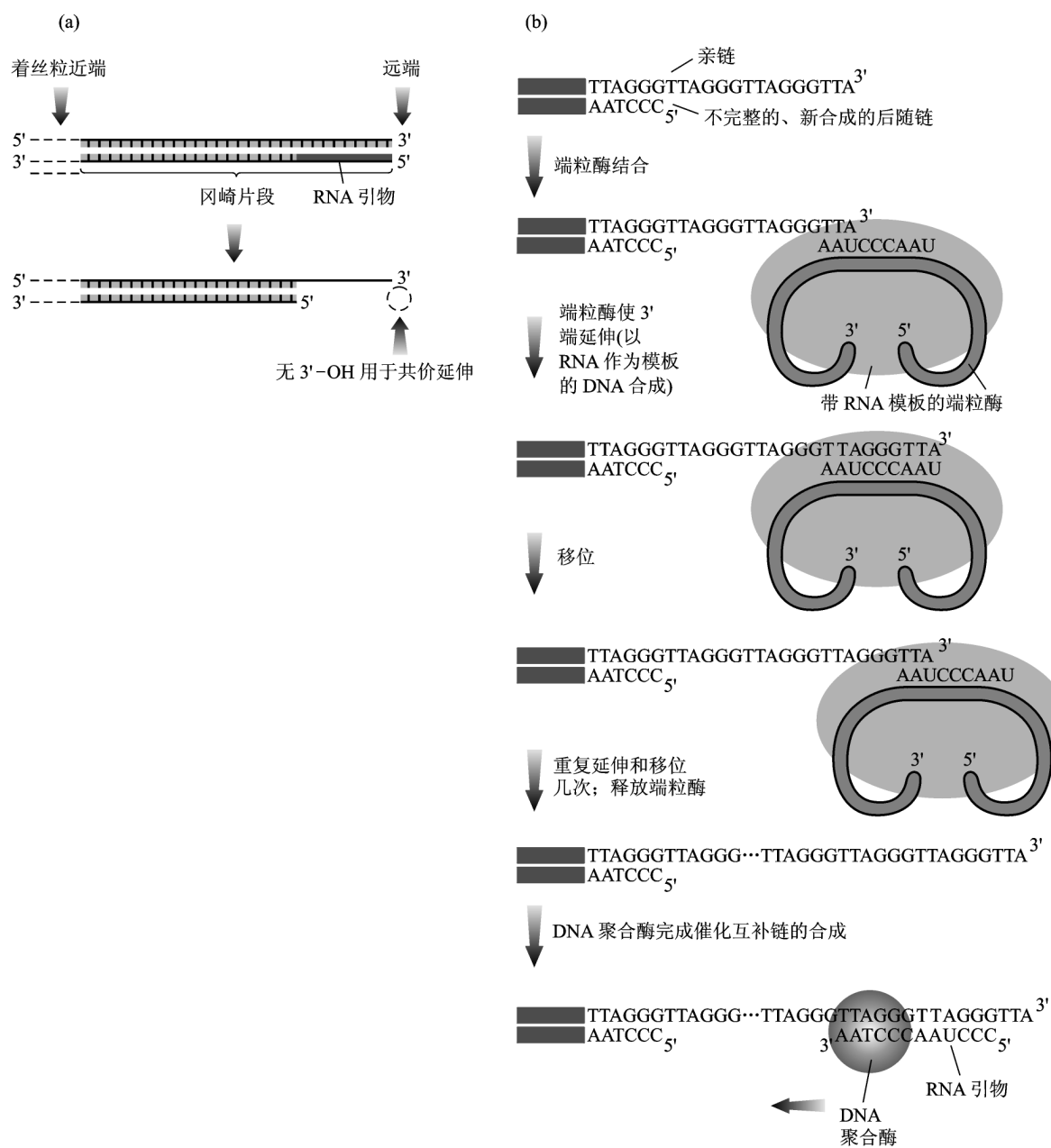


图 3-7 端粒 DNA 的合成(仿自 Sunstad, 2003)

(a) 端粒后随链引物 (b) 端粒酶解决了端粒 DNA 合成的引物问题

## 3.4 RNA 转录与加工

### 3.4.1 RNA 聚合酶

转录(transcription)是 DNA 的遗传信息被拷贝成 RNA 的遗传信息的过程。在该过程中以 DNA

为模板,在依赖于 DNA 的 RNA 聚合酶的催化下,以 4 种 NTP ATP、CTP、GTP 和 UTP 为原料,合成 RNA。在此过程中我们一般将 DNA 双链分子上带有遗传信息的链称为非模板链(nontemplate strand),或基因链(gene strand),或有义链(sence strand),或编码链(coding strand),它与 mRNA 序列一致,代表的是从遗传密码到蛋白质序列相联系的 DNA 序列。另一条与其互补的链称为模板链(template strand),或反基因链(antigene strand),或反义链(antisense strand),它是作为模板合成 mRNA 的 DNA 链。所有基因均以模板链为模板进行转录,因而转录产物的碱基顺序必定与基因链的遗传信息一致(图 3-8)。

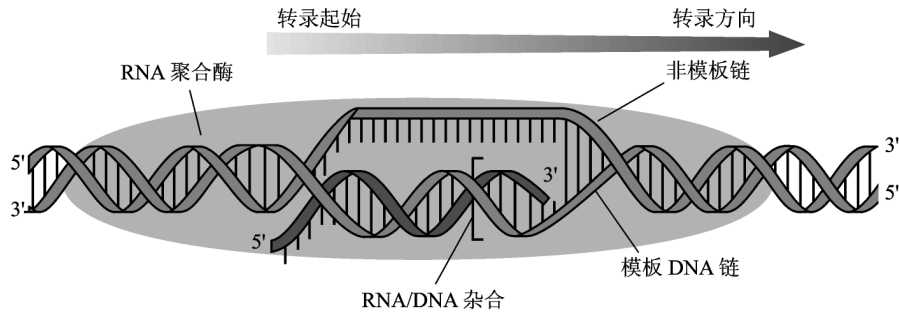


图 3-8 RNA 的转录过程(引自 Russell 1998)

RNA 聚合酶在 DNA 模板链上移动的方向为  $3' \rightarrow 5'$ , 转录产物为  $5' \rightarrow 3'$ 。在真核细胞中, rRNA 在核仁内转录, mRNA 与 tRNA 在核质中转录。转录过程可以分为模板识别(template recognition)、转录起始(transcription initiation)、延伸(elongation)和终止(termination) 4 个基本阶段。模板识别是指 RNA 聚合酶与启动子相互作用并结合到启动子上的过程;转录起始是 RNA 聚合酶结合到启动子上后开始合成 RNA;延伸是 RNA 链被延伸, RNA 合成主要在此阶段;终止是指延伸被终止,转录物(transcript)从模板上脱离下来。

#### (1) 大肠杆菌 RNA 聚合酶

大多数原核生物 RNA 聚合酶的组成是相同的,而且只有一种 RNA 聚合酶转录所有的基因。最典型的是大肠杆菌 RNA 聚合酶,它主要由 4 种亚基组成,包括 2 个  $\alpha$  亚基、一个  $\beta$  亚基、一个  $\beta'$  亚基和一个  $\sigma$  亚基。 $\alpha$  亚基的功能是在核心酶组装和启动子识别中起作用,还可与一些激活蛋白质结合对基因调控发挥作用。 $\beta$  和  $\beta'$  亚基组成聚合酶的催化中心,这两个亚基的蛋白质序列与真核生物 RNA 聚合酶的两个大亚基有同源性,表明 RNA 聚合酶结构与功能的相关性。 $\sigma$  亚基又称  $\sigma$  因子(sigma factor),相对分子质量为  $(32 \sim 90) \times 10^3$ ,在细菌细胞中存在多种不同的  $\sigma$  因子,用于识别不同的启动子。 $\alpha_2\beta\beta'$  组成大肠杆菌核心酶(core enzyme), $\sigma$  因子加入后形成全酶(holoenzyme)。只有全酶才能专一性起始转录。 $\sigma$  因子上的几种氨基酸能与 -10 启动子序列的编码链上的碱基专一结合,这便是  $\sigma$  因子决定启动子专一性的分子机制(图 3-9)。

一个  $\sigma$  因子(E. coli 中是  $\sigma^{70}$ )起始大部分基因的转录,而其他  $\sigma$  因子识别不同的启动子,可以诱导特殊基因的协同表达。原核生物一种 RNA 聚合酶转录所有的基因,其中  $\sigma$  因子控制启动子的识别。由  $\sigma$  因子替代引导核心酶去识别一组不同的启动子。

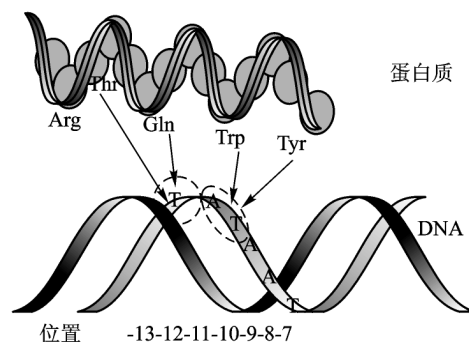


图 3-9  $\sigma$  因子决定启动子的专一性(引自 Lewis 2004)  
在  $\sigma^{70}$  上的氨基酸与 -10 启动子序列的编码链中特定碱基接触

### (2) 真核生物 RNA 聚合酶

真核生物与原核生物不同,细胞中有 3 类 RNA 聚合酶,最早是根据它们从亲和层析柱上洗脱的次序不同分别称为 RNA 聚合酶 I、II 和 III。但后来发现不同真核生物的 3 类 RNA 聚合酶的洗脱顺序并不完全相同,于是采用了它们对  $\alpha$ -鹅膏蕈碱( $\alpha$ -amanitin)的敏感程度不同来进行分类。3 类 RNA 聚合酶分别负责不同的基因转录(表 3-1)。

表 3-1 真核生物 RNA 聚合酶的特性比较

聚合酶类别	细胞核内定位	转录产物	对 $\alpha$ -鹅膏蕈碱的敏感程度
RNA 聚合酶 I	核仁	28S, 5.8S 和 18S rRNA	不敏感
RNA 聚合酶 II	核质	核前 mRNA( pre mRNA), 核内小分子 RNA( snRNA)	敏感
RNA 聚合酶 III	核质	rRNA, 5S rRNA, U6 snRNA, 小分子核仁 RNA( snoRNA), 小分子细胞质 RNA( scRNA)	中等敏感

从组成结构上看,真核生物 RNA 聚合酶一般由 12~15 个亚基组成,是一个很大的蛋白质复合物,其相对分子质量大于  $500 \times 10^3$ 。例如将酵母的 RNA 聚合酶与大肠杆菌的核心酶相比较发现,在组成上酵母 RNA 聚合酶 I、II 和 III 分别有两个大亚基(L'和 L)与大肠杆菌的  $\beta$  和  $\beta'$  亚基类似,另两个亚基与  $\alpha$  亚基类似。同时,酵母 RNA 聚合酶 I、II 和 III 彼此之间有 5 个小亚基( $10 \times 10^3 \sim 27 \times 10^3$ )是相同的。这说明 RNA 聚合酶在进化上高度保守。

由于没有细菌  $\sigma$  因子的类似物,因此真核生物 RNA 聚合酶不能直接识别基因的启动子区,这也是与原核 RNA 聚合酶的模板识别方式不同之处。真核生物 RNA 聚合酶需要一些转录起始因子(transcription initiation factor, TIF)或通用转录因子(general transcription factor, GTF)帮助其识别启动子。首先转录起始因子按特定顺序结合于启动子上形成复合物,帮助 RNA 聚合酶定位到 DNA 上的转录起始位点, RNA 聚合酶才能与之相结合,形成前起始复合物。由于多个转录起始因子的相互作用使 DNA 分子构象发生改变,从闭合状态转换成开放形式。转录起始因子与 RNA 聚合酶共同组成了转录起始的基本装置(basal apparatus)。

### 3.4.2 启动子与增强子

#### (1) 原核生物启动子

启动子(promoter)是决定 RNA 聚合酶转录起始位点的 DNA 序列。RNA 聚合酶特异与其结合,而启动转录。原核生物的启动子一般处在结构基因的上游。一个典型的原核生物启动子主要由 4 个区域组成:转录起始位点(transcription initiation site)、-10 序列(-10 sequence)、-35 序列(-35 sequence)以及 -10 序列和 -35 序列间的间隔碱基数目[图 3-10 a]。

转录起始位点是指在合成 RNA 时 DNA 链上第一个核苷酸掺入的位点,即与新生 RNA 链第一个核苷酸对应 DNA 链上的碱基,通常为一个嘌呤。常把起点前面或左边即 5' 端的序列称为上游(upstream),其后面或右边即 3' 端的序列称为下游(downstream)。在描述碱基位置时,一般用阿拉伯数字表示,转录起点的位置记为 +1,下游的核苷酸依次记为 +2, +3, +4, ..., 上游方向依次记为 -1, -2, -3, ...。注意在 -1 与 +1 之间没有 0 位置。

-10 序列也称 Pribnow 框/盒(Pribnow box),是由 Pribnow 和 Schaller 于 1975 年发现的。他们将原核生物 RNA 聚合酶全酶同原核基因的 DNA 结合后,用 DNase I 水解,得到与 RNA 聚合酶结合而未被水解的 DNA 片段(即被蛋白质保护的 DNA),经测序分析几乎都含有一个由 6 bp 组成的共有序列(consensus sequence): TATAATK T<sub>80</sub> A<sub>95</sub> T<sub>45</sub> A<sub>60</sub> A<sub>50</sub> T<sub>96</sub> 下标数字表示出现碱基最高频率的百分数[图 3-10 b]。该共有序列的中心通常位于转录起点上游 10 bp 处,故称 -10 序列。由于其富含 AT,有

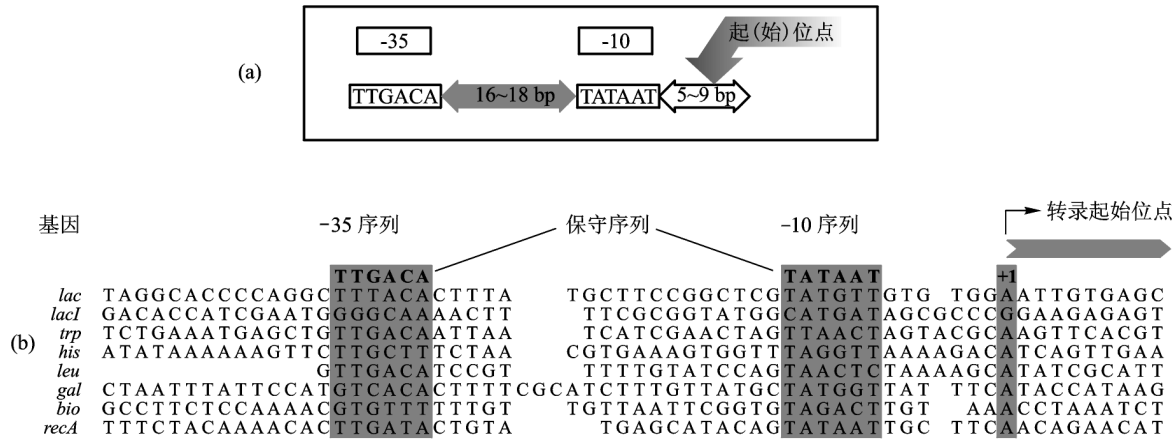


图 3-10 原核生物启动子结构(引自 Lewis 2004)

(a) 一个典型的原核生物启动子由 -35 序列、-10 序列、-10 序列和 -35 序列间的序列长度和转录起始位点这 4 部分组成 (b) 不同原核生物启动子序列的比较

利于解链。-10 序列的碱基组成对转录的效率影响很大,若发生 TATAAT→AATAAT 突变,产生下降突变(down mutation),转录效率下降;在 *lacO* (乳糖操纵子)的 -10 序列,若发生 TATGTT→TATAAT 突变,则发生上升突变(up mutation),转录效率会上升。这说明 -10 序列碱基的组成对启动子功能的重要性。推测 -10 序列的功能是与 RNA 聚合酶紧密结合,形成开放启动复合物,使 RNA 聚合酶定向转录。

-35 序列,其共有序列为 TTGACA (T<sub>82</sub> T<sub>84</sub> G<sub>78</sub> A<sub>65</sub> C<sub>54</sub> A<sub>45</sub>),该序列的中心位于转录起始位点上游 35 bp 处[图 3-10 (b)]。该序列是 RNA 聚合酶的识别位点。RNA 聚合酶的核心酶并不能识别 -35 序列,只有  $\sigma$  因子可以识别该序列并与之结合,使 RNA 聚合酶在启动子区域形成封闭复合物。因此,-35 序列的功能是为 RNA 聚合酶的识别提供信号,为转录选择模板链。

-10 序列和 -35 序列间的最佳距离在 90% 的原核生物启动子中为 16~18 bp,在该间距内的碱基序列的组成并不重要,但碱基的数目即该距离的长短却十分重要。碱基的数目的变化会影响基因转录活性的高低,强启动子一般为(17±1) bp,当间距小于 15 bp 或大于 20 bp 时都会降低启动子的活性。对 -10 序列和 -35 序列间碱基的长度关系,被解释为 -35 序列与 -10 序列间碱基数目的改变,会使它们产生的超螺旋关系发生改变(增减一个 bp 会使两者之间的夹角发生 36° 的变化),影响聚合酶与 DNA 的空间关系,若要使两者在这个区域内保持正确取向,就必须增加结合自由能,使两者弯曲。这说明适宜的距离可以为 RNA 聚合酶提供合适的空间结构,便于转录的起始。当然并不是所有的启动子都是以上所述的典型启动子结构,有些启动子缺少其中的某一结构组成,因此不能被 RNA 聚合酶所识别,这时需要其他辅助蛋白质参与,以弥补 RNA 聚合酶与启动子之间相互作用的缺陷。

## (2) 真核生物启动子

真核生物有 3 类 RNA 聚合酶,分别识别 3 种不同类型的启动子,它们在结构上各有特点。

① RNA 聚合酶 I 识别的启动子结构 RNA 聚合酶 I (RNA Pol I) 只转录 rRNA 一种基因,包括 5S、18S 和 28S rRNA。所有真核生物 rRNA 基因均为多拷贝,彼此首尾串联,簇集在染色体的特定区域,共同被转录在一个转录产物上,然后经加工成为 3 种 rRNA。

人类细胞的 RNA Pol I 识别的启动子研究得最清楚,它由两个分开的启动子成分构成,一是位于起始位点周围 -45 至 +20 区域内的核心启动子(core promoter),二是在起始位点 5' 上游从 -180 至 -107 区域的上游控制元件(upstream control element, UCE)。核心启动子本身就足以起始转录,而 UCE 的存在可以大大提高核心启动子的转录起始效率。两个区域内的碱基组成和一般的启动子结

构有所差异,均富含 G—C 对,两者有 85% 的同源性。

② RNA 聚合酶 III 识别的启动子结构 RNA Pol III 基因的产物为一些相对分子质量较小的细胞质 RNA (cytoplasmic RNA, scRNA)、rRNA、5S rRNA、7SL RNA (参与细胞内蛋白质转移)、U6 RNA (转录后加工)等。RNA Pol III 基因的启动子很特殊,既有上游启动子也有下游启动子。下游启动子位于它们转录的基因编码序列内,通常核心启动子在 +50 bp 至 +100 bp 之间,由 2 个分开的框序列组成, A 框和 C 框,或 A 框和 B 框。因此, RNA Pol III 基因的启动子有 3 种类型: I 型基因内启动子、II 型基因内启动子和 III 型基因外启动子。

I 型基因内启动子:是 5S rRNA 的基因所具有的启动子,不含 TATA 框,转录控制区(核心启动子)位于 +50 至 +80 的区段,由 A box 和 C box 组成。这类启动子需要 TF III A、TF III B 和 TF III C 的结合来定位 RNA 聚合酶。

II 型基因内启动子:是 rRNA 的基因启动子,也无 TATA 框,由 A box 和 B box 构成基因的不连续启动子。A box 长 11 bp,位于 +10 bp 至 +20 bp; B box 长 16 bp,位于 +50 bp 至 +65 bp。只需要 TF III B 和 TF III C 的结合来定位 RNA 聚合酶。

III 型基因外启动子:是某些 scRNA、U6 RNA 和 7SL RNA 的基因启动子,属上游启动子,它由 TATA 框、近端顺序元件(proximal sequence element, PSE)和 Oct 元件 3 个上游元件组成。TATA 框可能决定聚合酶类型的特异性, PSE 和 Oct 元件的存在可以大大增加转录效率。

③ RNA 聚合酶 II 识别的启动子结构 RNA 聚合酶 II (RNA Pol II) 主要负责蛋白质基因和部分核内小 RNA (small nuclear RNA, snRNA) 的转录,其识别的启动子位于转录起始位点的上游,结构最为复杂。RNA Pol II 识别的启动子由核心启动子(core promoter)或基本启动子(basal promoter)和上游启动子元件(upstream promoter element, UPE)组成。核心启动子是指在体外测定到的由 RNA Pol II 进行精确转录起始所要求的最低限度的一套 DNA 序列元件。一个典型的核心启动子与起始位点紧密靠近,长约 80 个核苷酸,从转录起点向上游(-40)或下游(+40)延伸。

RNA Pol II 核心启动子由 4 种元件组成,包括:① TATA 框(TATA box):在转录起始点上游 25 bp 至 30 bp 处有一段被称做 Hogness 序列的保守序列 TATA(A/D)A(A/D),通常简称 TATA 框(盒),其作用是选择正确的起始位点,保证精确起始并产生基础水平的转录,当某些基因缺少 TATA 框时可由 InR 来替代这一作用。TATA 框还影响转录的效率。没有 TATA 框的启动子称为无 TATA 框启动子(TATA-less promoter),它占 RNA Pol II 启动子的半数或更多。② TB II B 识别序列(TB II B recognition element, BRE):位于 -32 bp 至 -37 bp 之间,作用是介导 RNA Pol II 的结合,影响转录起始点的选择。③ 起始子(initiator, InR):一般由 P<sub>2</sub>CAP<sub>2</sub> 构成,位于 -2 至 +4 之间,可能提供 RNA Pol II 的识别。只包含起始子的启动子是能被 RNA Pol II 识别的最简单形式。④ 下游启动子元件(downstream promoter element, DPE):当一个启动子无 ATAT 框时,通常在位于 +28 至 +32 处具有一个下游启动子元件(DPE)。DPE 对于含起始密码子但不含 TATA 框的启动子的活性十分重要,但对含 TATA 框的启动子却无作用[图 3-11(a)]。

一般地,一个核心启动子可能包含以上 4 种元件中的 2 种或 3 种。因此,一个核心启动子既可由 TATA 框和 InR 组成,还可通过 InR 和 DPE 组成。这样,核心启动子的变化是不大的。但是,在 RNA Pol II 启动子中上游启动子元件的变化和组合却是很大的,这样与核心启动子再组合起来就形成了 RNA Pol II 启动子的复杂性和多样性。

RNA Pol III 启动子的上游启动子元件一般位于起始位点上游 -50 bp 至 -200 bp 区域,主要有 CAAT 框、GC 框、八聚体框(octamer box)等。CAAT 框是位于转录起始上游约 -75 bp 处的一个保守序列 GG(T/C)CAATCT。实验揭示 CAAT 框内的碱基突变对转录起始的影响很大,说明它决定了启动子起始转录的效率及频率,但是并不影响启动子的专一性。而且,该元件对启动子的影响在正方向和反方向排列时均能产生作用。GC 框位于转录起始上游约 -90 bp 附近,含有共有序列 GGGCGG,可以在启动子中正、反方向排列,一份或数份,其功能与 CAAT 框相似,主要是决定启动子起始转录的效率。八

聚体框一致序列为 ATTTGCAT,是转录因子 Oct 1和 Oct 2识别结合的位点[图 3-11b]。

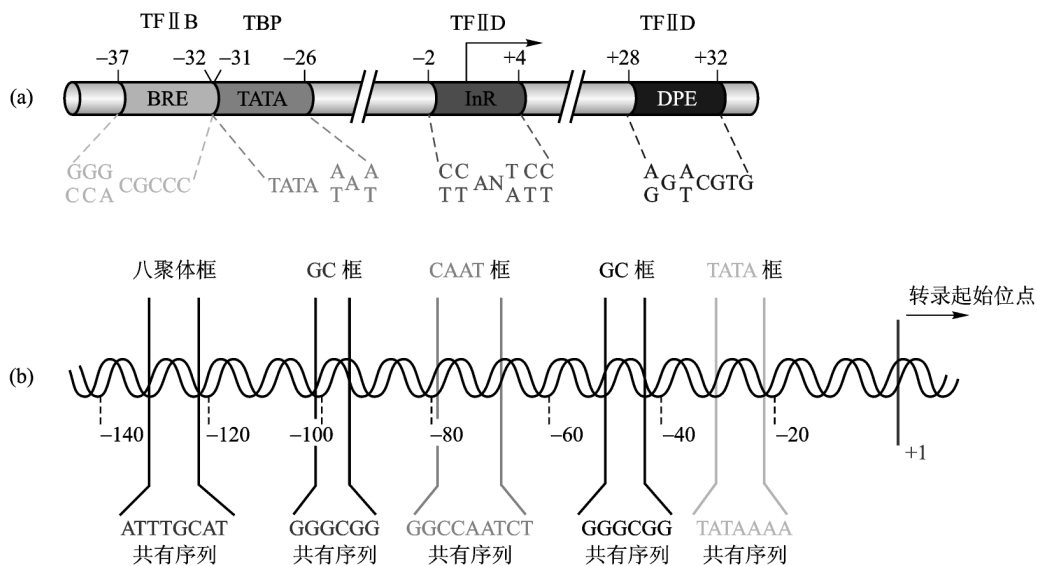


图 3-11 RNA 聚合酶 II 识别的启动子结构

(a) RNA Pol II 识别的核心启动子结构主要由 BRE、TATA 框、InR 和 DPE 等组成(引自 Watson 等, 2004)

(b) RNA Pol II 启动子的上游启动子元件有 CAAT 框、GC 框、八聚体框等(仿自 Snustad 等, 2003)

### (3) 增强子

增强子(enhancer)是增强真核基因转录的一类调节序列。它是通过启动子来增强转录的一种远端性控制元件。增强子可位于基因的 5'端、3'端或者基因的内含子中,增强的是同它连锁的基因的转录频率,一般它可使转录频率增加 10~200 倍。

首先被发现的是 SV40 增强子。1981 年, Benerj 等在 SV40 DNA 中发现一段 140 bp 的序列,位于 SV40 早期基因的上游 200 bp 处,是两个串联的 72 bp 的重复序列,其中核心序列为 GGTGTG-GAAAG。它们不是启动子的组成部分,但能增强或促进转录的起始。缺失实验显示两个重复缺失一个,基因转录水平正常,两个都缺失将会大大降低转录水平。有人发现,若将  $\beta$ -珠蛋白基因置于含有上述 72 bp 重复的 DNA 分子上,发现其在体内的转录水平提高了 200 倍以上,甚至将该 72 bp 序列放在离转录起点上游 1400 bp 或下游 3300 bp 处时,它都产生增强作用。后来发现增强子在真核基因中比较普遍存在,例如反转录病毒基因、免疫球蛋白基因、胰岛素基因、胰糜蛋白酶基因等。

增强子的主要作用特点是:① 具有远距离效应:常在上游 -200 bp 处,但可增强远处启动子的转录,即使相距十几 kb 也能发挥其作用。② 无方向性:在 DNA 双链中没有 5'端与 3'端固定的方向性,无论在靶基因的上游、下游或内部都可发挥增强转录的作用。③ 顺式调节:只调节位于同一染色体上的靶基因,而对其他染色体上的基因无作用。④ 无物种和基因的特异性,对同源或异源基因同样有效,可以接到异源基因上发挥作用,如将 SV40 的增强子接到兔  $\beta$ -珠蛋白基因前,引入 HeLa 细胞,此珠蛋白基因转录增强 200 倍。⑤ 具有组织特异性:SV40 的增强子在 3T3 细胞中比多瘤病毒的增强子要弱,但在 HeLa 细胞中 SV40 的增强子比多瘤病毒的要强 5 倍,抗体基因的增强子只有在 B 淋巴细胞中才起作用。增强子的效应需特定的蛋白质因子参与。⑥ 有相位性:增强子的活性与其在 DNA 双螺旋结构中的空间方向性有关(其作用和 DNA 的构象有关)。⑦ 有的增强子可以对外部信号产生反应:如热休克基因在高温下才表达。编码重金属蛋白的金属硫蛋白基因在镉和锌存在下才表达。某些增强子可以被固醇类激活。⑧ 增强子含有许多可与不同转录因子结合的基序,可以不同的组合方式调控基因的表达,在已研究过的绝大多数基因的表达调控模式中都发现组成增强子的不同基序或模块(module)之间可形成不同的组合,它们是基因差别表达(differential expression)的主要原因。

### 3.4.3 原核生物 tRNA 和 rRNA 的加工

多数转录的初始产物无生物活性,在生物体内进行加工处理后才具有生物活性。在这里加工是指转录后加工(post transcriptional processing),它是对初级转录物(各种前体 RNA 分子)进行剪接、加工使之成为成熟的 mRNA、rRNA 或 tRNA 的过程。

原核生物 rRNA 初始转录物多为串联在一起的多顺反子(polycistron),少数为单顺反子,另有由 rRNA 和 tRNA 串联组成。例如,在大肠杆菌 tyu 操纵子上有 4 个串联排列的 rRNA 基因( $rRNA_1^{Tyr}$ 、 $rRNA_2^{Tyr}$ 、 $rRNA_3^{Tyr}$  和  $rRNA_4^{Tyr}$ )和一个 EF Tu 蛋白基因。同时,大肠杆菌 rRNA 基因常常插在 tRNA 转录单位内。大肠杆菌共有 7 个 rnm 操纵子(A—G),编码 3 种不同的 rRNA 基因,在多顺反子中基因的排列顺序是相同的,即都是 16S—rRNA—23S—5S—rRNA 顺序。为了得到成熟的 rRNA 和 tRNA,细菌细胞利用了多种核酸酶对前体分子进行后加工。

### 3.4.4 真核生物 tRNA 和 rRNA 的加工

#### (1) 真核生物 tRNA 的加工

真核生物 rRNA 的基因与原核不同,表现在:① 真核 rRNA 的前体分子数目多、单顺反子、成簇排列、有基因间隔区等。② 真核的前体 rRNA 分子中含有内含子(intron),其加工过程中要剪接内含子,还都要加 3'-CCA。而且真核 rRNA 内含子的切除和其他内含子的切除是不同的,首先,内含子中没有交界序列,没有内部引导序列;其次切除依赖于蛋白质的 RNase。③ 不是转酯反应。④ 其剪接原则上是依赖于对 rRNA 共同的二级结构的识别。

目前对酵母前体 rRNA 分子结构及其加工过程了解得比较深入。在酿酒酵母(*Saccharomyces cerevisiae*)的 272 个核 rRNA 基因中有 59 个 rRNA 基因为断裂基因,均含有一个内含子,位于反密码子环的 3'端,与反密码子只隔一个核苷酸,内含子长度变化在 14~16 bp。对酵母 rRNA 前体内含子的剪接机制研究揭示,内含子的序列顺序和大小,对剪接并不产生影响,而是依赖 rRNA 的共同二级结构由内切核酸酶进行识别酶切。加工主要是切除内含子,连接外显子(图 3-12)。首先由酵母内切核酸酶识别 rRNA 前体共同结构,在特异部位切割产生带有 2',3'-环磷酸的 5'半分子和带有 5'羟基的 3'半分子,同时释放出完整的内含子。剩下的 rRNA 2 个半分子再经折叠,5'半分子在 2',3'-环磷酸单酯酶作用下转变成带 2'-磷酸的 5'半分子,3'半分子在 ATP 存在下由激酶催化生成带 5'-磷酸及一个 3'羟基裂口的,类 rRNA 结构半分子。RNA 连接酶再将这两个半分子连接成完整的 rRNA 分子,5'端多余的 2'-磷酸由特异的磷酸单酯酶除去。这种经过剪切和修饰的 rRNA 分子由于缺乏 3'-CCA—OH 是没有活性的,还必须在 rRNA 核苷酸转移酶的作用下进行 3'端 CAA 的加成。

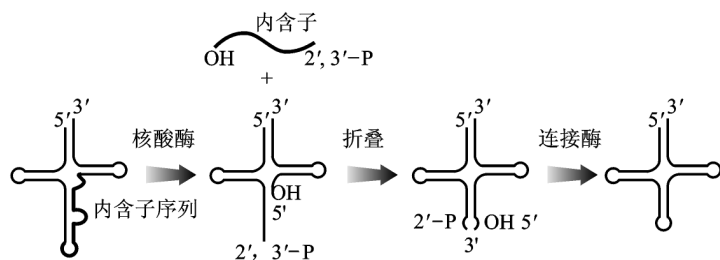


图 3-12 酵母 rRNA 内含子的剪接(引自 Lew in 2004)

#### (2) 真核生物 rRNA 加工

真核生物的 18S、5.8S 和 28S rRNA 基因串联在一起形成一个转录物,初级转录物为 45S 前体,5S rRNA 与它们分开转录,这和原核的 rRNA 基因不同。一些高等的真核 rRNA 基因中没有内含子,某些低等真核生物 rRNA 前体中含有 I 型(group I)内含子。例如,哺乳动物细胞在核仁中合成的 rRNA 初级转录物为 45S 前体,它的长度是 3 种成熟 rRNA 总长度的 2 倍,在核中经过一系列的核酸

内切酶和外切酶的作用,最终加工形成 18S、5.8S 和 28S rRNA(图 3-13)。共有 100 多个甲基化位点在转录过程或以后被甲基化。有人认为甲基化是 45S 前体上最终成为成熟 rRNA 区域的标志。在 rRNA 前体的加工过程中需要核仁小分子 RNA(small nuclear RNA, snoRNA) 参与, snoRNA 与 rRNA 前体分子的切割、修饰、甲基化及 3' 端的产生有关。

### 3.4.5 真核生物 mRNA 前体的加工

原核生物由于转录和翻译的偶联, mRNA 一般不经过加工,中间没有加工的时间。少数情况下多顺反子 mRNA 先被内切酶切成较小的单位再作为翻译的模板。真核生物则不同,转录在细胞核中,翻译在细胞质中,其中大多数基因为断裂基因,存在内含子与外显子。因此, mRNA 前体必须要经过加工过程。真核 mRNA 前体的加工一般要经过 4 个步骤: mRNA 的 5' 端加帽、mRNA 的 3' 端多聚腺苷酸化(加尾)、内含子的切除与外显子的连接以及对某些碱基的修饰等。

#### (1) mRNA 的 5' 端加帽

在转录延伸过程的早期,通过增加 7-甲基鸟嘌呤(7-methyl guanosine, 7-MG)帽结构修饰 pre-mRNA 的 5' 端。当生长的 RNA 链只有大约 30 个核苷酸长时,帽结构就被加到 5' 端。成熟的真核生物并没有游离的 5' 端,只有所谓的帽子结构,该结构是通过转录加工加上去的。mRNA 的 5' 端加帽是一个多步加工过程,首先是鸟苷转移酶(guanylyl transferase)将一个鸟苷酸加在 5' 端 RNA 的前端,产生 5'-5' 端对接的磷酸二酯键。然后由鸟嘌呤甲基转移酶(guanine methyl transferase)和 2-O-甲基转移酶将一个甲基基团加到嘌呤环的 7 位氮原子使 5' 端帽子鸟嘌呤转变为 7-甲基鸟嘌呤(图 3-14)。

真核生物的帽子结构有 3 类,即 Type 0 cap: m7GpppX; Type I cap: m7GpppXm; Type II cap: m7GpppXmYm。一般认为 mRNA 5' 加帽的功能主要表现在 4 个方面: ① 保护 mRNA 5' 端不被降解。细胞内的许多 RNA 酶(RNase),它们可攻击游离的 RNA 分子。RNA 酶的降解从 5' 端起始,当在 mRNA 的 5' 端加上 m7GpppG 帽子后,带有 3 个连接磷酸的 5' 帽可阻止存在细胞内的 RNase 切割。② 为核糖体识别 mRNA 提供信号,提高翻译效率。真核生物 mRNA 必须通过 5' 帽结合蛋白才能接触核糖体,起始翻译。缺少加帽的 mRNA 由于不能被 5' 帽结合蛋白识别,其翻译效率比加帽的 mRNA 低 20 倍。③ 作为进出细胞核的识别标记。凡由 Pol III 转录的 RNA 均在 5' 端加帽,包括 snRNA,这是 RNA 分子进出细胞核的识别标记。大多数 snRNA 转录后在细胞核中接收 5' 端单甲基 m7G 加帽,然后转移到细胞质与 snRNP 蛋白结合。在细胞质中 snRNA 5' 帽需再修饰成为三甲基带帽结构 m2,2,7G,随后重新返回细胞核参与 mRNA 的剪接加工。U6 snRNA 由 Pol III 转录,在其 5' 端保留的三磷酸基团无帽子结构,因而不能输出细胞核。某些突变型中被输送到细胞质中的 snRNA 由于不能合成三甲基带帽结构,不能返回细胞核。④ 提高 mRNA 的剪接效率。5' 帽结合蛋白涉及第一个内含子剪接复合物的形成,直接影响 mRNA 的剪接效率。

#### (2) mRNA 的 3' 端多聚腺苷酸化

由 RNA Pol III 转录合成的初级转录物的 mRNA 3' 端都被内切酶切除,然后加上多聚腺苷的尾巴。新合成的 mRNA 的 3' 端含有明显的加尾信号,位于 poly(A)上游 10~20 个核苷酸处,其一致顺

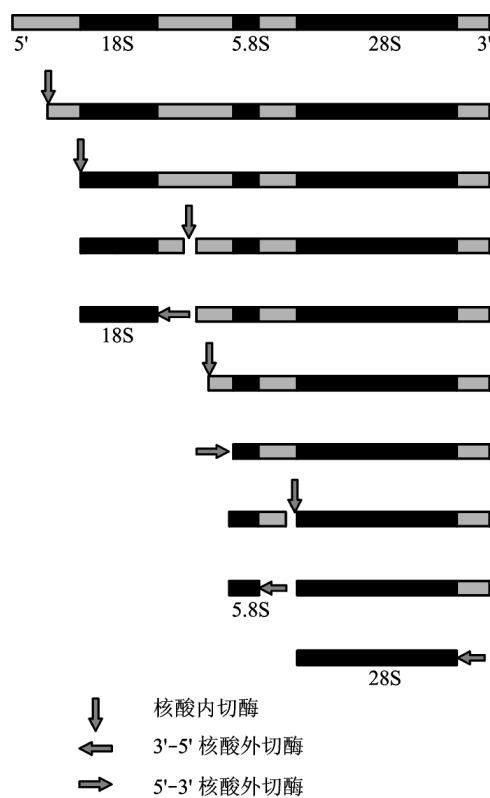


图 3-13 高等真核生物 rRNA 前体的加工  
(引自 Lewis, 2004)



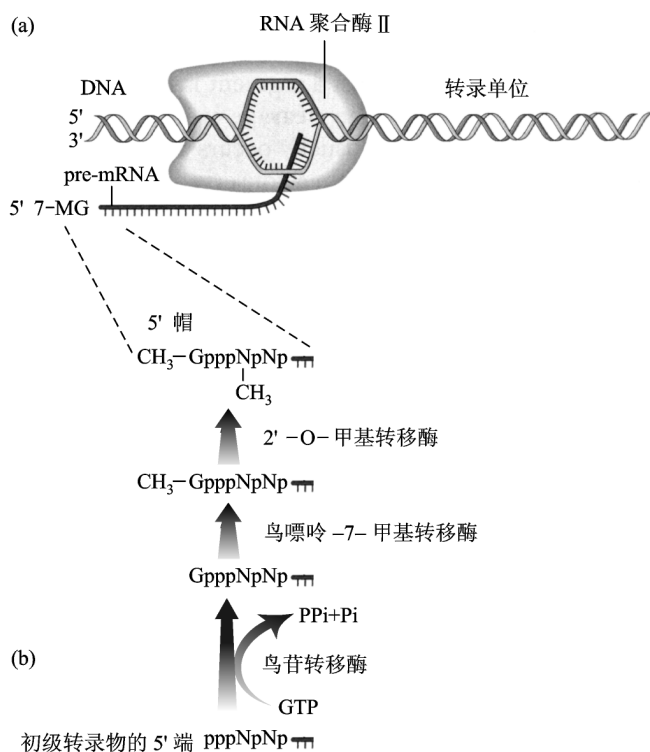


图 3-14 mRNA 前体的 5' 端加帽(7-MG)(引自 Snustad 等, 2003)

(a) 一个由 RNA 聚合酶 II 转录的基因的早期阶段 (b) 7-MG 帽结构的生化合成途径

序为 5'-AAUAAA-3'。首先必须在加尾的核苷酸位置切断 RNA，然后在游离的 3' 端由 poly(A) 聚合酶进行多聚腺苷酸化。一旦 mRNA 前体被切断之后，多聚腺苷酸化立即开始。多聚腺苷酸化可细分为两个阶段，首先依赖 5'-AAUAAA-3' 信号和与之结合的 CPSF 缓慢合成约 10 个核苷酸，然后依赖已合成的寡聚腺苷酸迅速在其末端加上 200 或更多个腺苷酸(图 3-15)。

mRNA 的 3' 端多聚腺苷酸化的作用可能是：① 增加 mRNA 的稳定性。② 提高 mRNA 翻译效率。③ poly(A) 可影响 mRNA 前体最后一个内含子的剪切，缺少 poly(A) 使剪接效率降低 5~10 倍。

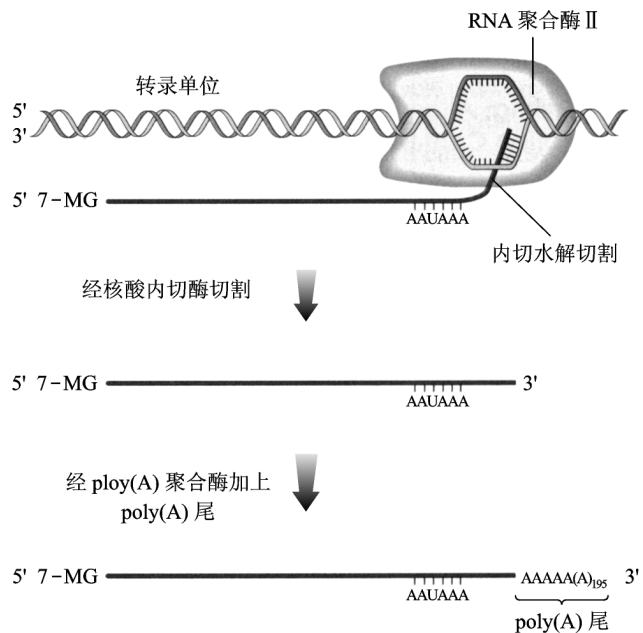


图 3-15 mRNA 3' 端的切割和加 poly(A) 尾(引自 Snustad 等, 2003)

### (3) RNA 的剪接

真核细胞基因 DNA 的间插序列( *intervening sequence* ),即不连续基因( *interrupted gene* )的间插序列称为内含子( *intron* )。被内含子隔开的基因序列,即出现在成熟 RNA 中的序列称为外显子( *exon* )。真核生物前体 mRNA 切除内含子,连接外显子形成成熟的 mRNA 的过程称为 RNA 剪接( *RNA splicing* )。

对内含子与外显子的序列分析表明,内含子的边界顺序由保守的基序组成。分析比较大量的真核生物 mRNA 内含子发现,它们的两侧边界均有一对保守顺序,即 5' 端为 GU, 3' 端为 AG,这类称为 GU-AG 的内含子均以相同方式剪切。只是在不同的真核生物中,内含子的一致顺序有所不同。

① I 类内含子的剪接 I 类内含子常分布于低等真核生物的细胞器中,如四膜虫的 rRNA,大部分真菌如酵母的 mtDNA,植物叶绿体基因的内含子等。I 类内含子属于自催化剪接( *autocatalytic splicing* )内含子。其结构特点是:边界序列为 5' U $\downarrow$ ...G $\downarrow$  3',内含子中有中部核心结构( *central core structure* )和内部引导序列( *internal guide sequence, IGS* )。内含子中可与外显子配对的序列称为内部引导序列,由 6 个核苷酸组成,GGAGGG(四膜虫)。最初认为 IGS 的作用是通过和两个外显子近侧区配对,使外显子并在一起,现在认为其作用是决定剪接的专一性。I 类内含子有一个共同的二级结构,共 9 个配对区( P1~P9),其中 P4 和 P7 是 I 类内含子中共有的保守序列。P3, P4, P6 和 P7 是核心结构,是可以执行催化的最小区域。其他的配对区在不同的内含子中是不同的。

I 类内含子的剪接机制属自剪接( *self splicing* )反应,通过 3 次连续的转酯反应完成(图 3-16)。

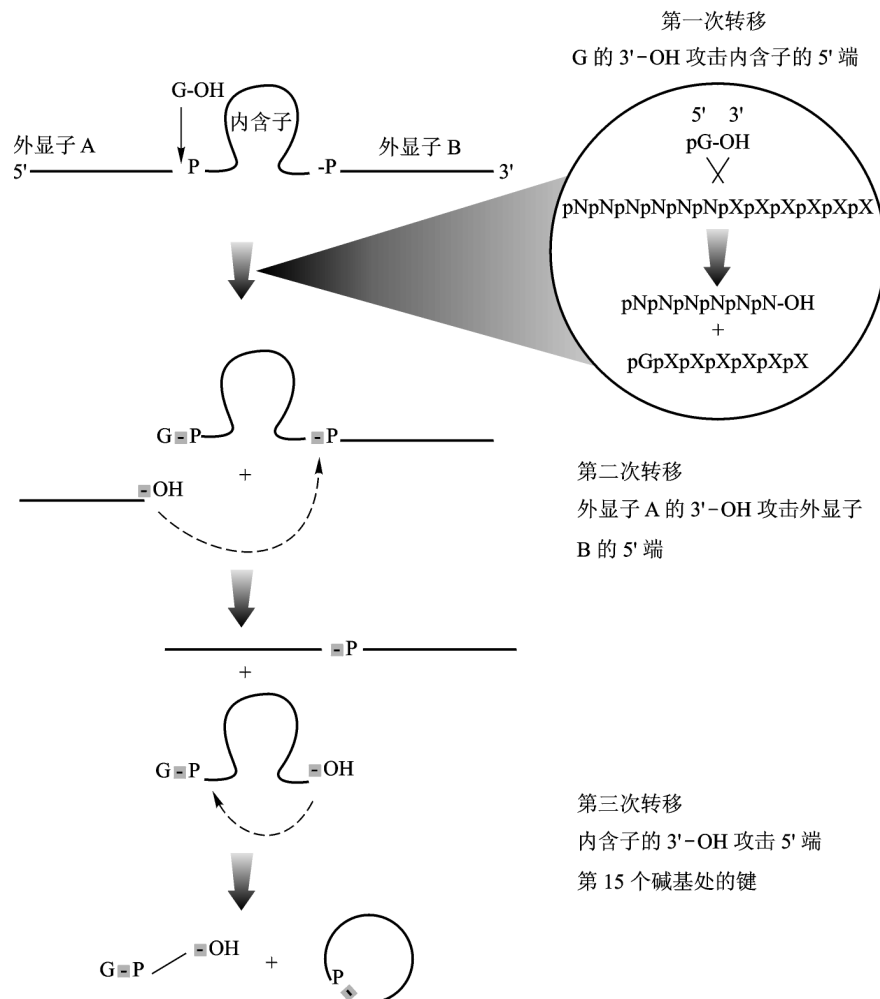


图 3-16 I 类内含子的剪接机制(引自 Lew 等, 2000)

首先利用游离 G 提供的 3' 端自由羟基, 连接内含子和上游外显子(外显子 A) 的磷酸二酯键, 在内含子的 5' 端加上鸟嘌呤使其延伸, 并在上游外显子 3' 端产生自由羟基。接着, 上游外显子 3' 端的自由羟基进攻内含子与下游外显子(外显子 B) 的磷酸二酯键, 使上游外显子与下游外显子连接并将 5' 端有鸟嘌呤的内含子以线性形式释放出来。最后, 通过第三个转酯反应, 内含子的 3' 端与鸟嘌呤连接形成环状结构。

② II 类内含子的剪接 II 类内含子分布在酵母 mRNA, 细胞色素氧化酶的 a 亚基、b 亚基, 玉米 mRNA、rRNA 以及真核 mRNA 前体中。II 类内含子也属自催化剪接内含子。其边界序列为 5' ↓ GUGCG... YnAG ↓ 3', 符合 GU-AG 法则。其二级结构形成 6 个茎环结构, 使两个并列的功能区靠近(图 3-17)。功能区 5 和功能区 6 被两个碱基隔开。功能区 6 含有 1 个不配对 A 残基, 其上带有 2'-OH, 发动第一次转酯反应。A 的 2'-OH 攻击 5' 端外显子-内含子边界的 G。第二次由 G 攻击 3' 端剪接点, 完成转酯反应, 内含子形成套索状结构。与核基因 mRNA 的剪接过程相比, II 类内含子的最大特点是不需要其他成分的参与, RNA 分子本身能形成剪接所需的空问结构和催化活性区域。

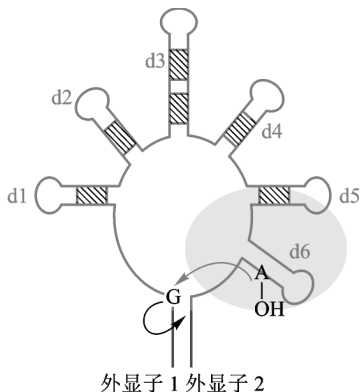


图 3-17 II 类内含子的二级结构与转酯反应位点(引自 Lewin, 2004)

③ 核 mRNA 前体内含子的剪接 真核生物基因组中的大多数内含子都不能自催化剪接, 称为核 pre-mRNA 内含子(nuclear pre-mRNA intron)。这些内含子的切除是由剪接体(spliceosome)完成的。剪接体是一个 40~60 S 的核糖核蛋白复合物, 由剪接装置(splicing apparatus)的各组分装配而成, 包括核内小分子 RNA 和 RNA 结合蛋白, 它们结合在一起以核内小分子核糖核蛋白(small nuclear ribonucleoproteins, snRNP) 的形式存在。

snRNP 是根据它所含有的 U-RNA(因为它含有修饰的尿嘧啶, 分为 U1、U2 RNA 等) 特殊种类而命名。参与剪接体的构成并在剪接过程中起作用的 snRNP 分别称为 U1、U2、U4、U5 和 U6。每个 snRNP 都包含 10~20 个蛋白。其他一些蛋白瞬时与剪接体结合, 但本身不是 snRNP 的组成成分, 称为剪接因子(splicing factor)。

剪接体的形成与 mRNA 前体的剪接过程紧密相连。首先在可变剪接因子(alternative splicing factor, ASF)/剪接因子 1(splicing factor 1, SF1) 的参与下由 U1 识别并直接与 5' 端剪接位点配对, 结合在 mRNA 前体 5' 端剪接点, U2 ASF 结合在 3' 端剪接位点富含嘧啶区, SF1/BBP 蛋白将两个剪接位点蛋白桥连起来, 形成早期前剪接复合体(early presplicing complex, E 复合体)(图 3-18)。U2 结合到分支点上形成 A 复合体。U5/U4/U6 三聚体与 A 复合体结合形成 B1 复合体, 完成剪接体的装配。各组分进行位置调整, U5 从外显子部位移向内含子位点, U6 结合在 5' 位点, mRNA 前体发生空间结构变化, 释放出 U1, 使 B1 复合体转变成 B2 复合体。在 ATP 供能的条件下, U4 被解离, U6/U2 催化转酯反应, U5 结合在外显子的 3' 剪接位点, 内含子 5' 端被切开, 形成套索(loop) 结构, 生成 C1 复合物。U2/U5/U6 保持与套索结构结合, 形成 C2 复合体, 进行第二次转酯反应, 内含子 3' 端被切开, 外显子连接起来(图 3-19)。

核 mRNA 的剪接过程和 II 类内含子十分相似, 根本区别是: 核 mRNA 前体本身不能形成二级结构, 必须依赖于 snRNA 和 snRNP(U1、U2、U4、U5 和 U6) 的帮助才能形成剪接体, 进行剪接, 是在剪接体的作用下通过转酯反应而完成, 形成一个套索结构和剪接了的外显子。

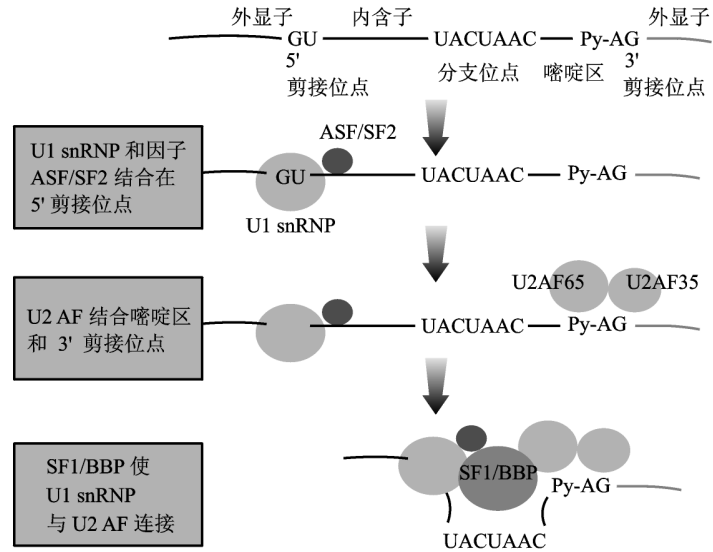


图 3-18 mRNA 前体剪接过程(引自 Lew 2004)

两端剪接位点的相互作用使 E 复合体形成

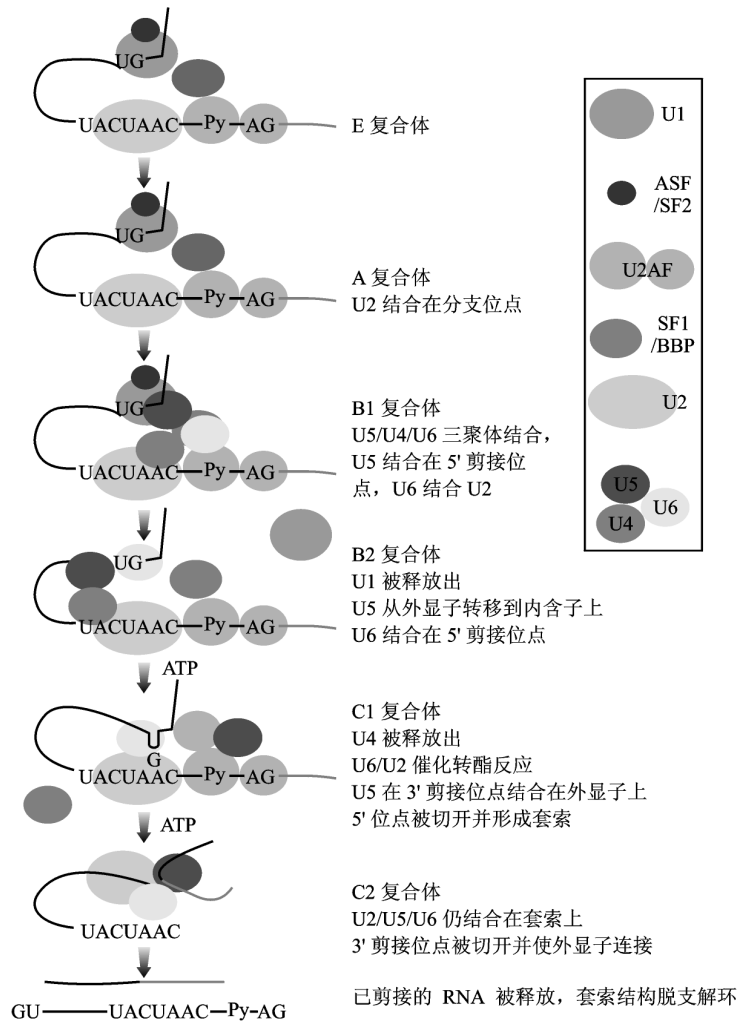


图 3-19 mRNA 前体剪接机制

剪接体通过几个中间复合体过程而形成,由剪接体完成 mRNA 前体的剪接

## 3.5 遗传密码与蛋白质合成

### 3.5.1 遗传密码的性质

遗传密码 (genetic code) 是联系 mRNA 的碱基序列和蛋白质氨基酸序列的桥梁, 具有以下性质:

① 遗传密码是三联体密码 (triplet code) 1 个密码子由 3 个核苷酸组成, 它特异编码多肽链中的 1 个氨基酸 (表 3-2)。

表 3-2 通用遗传密码表

		第二位(中间)核苷酸									
		U		C		A		G			
第一位 5'端 核苷酸	U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U C A G	
		UUC		UCC		UAC		UGC			
		UUA	Leu	UCA		UAA	终止	UGA	终止		
		UUG		UCG		UAG	终止	UGG	Trp		
	C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U C A G	
		CUC		CCC		CAC		CGC			
		CUA		CCA		CAA	CGA	Gly			
		CUG		CCG		CAG	CGG				
	A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U C A G	
		AUC		ACC		AAC		AGC			
		AUA		ACA		AAA	AGA	Arg			
		AUG	Met	ACG		AAG	Lys		AGG		
	G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U C A G	
		GUC		GCC		GAC		GGC			
		GUA		GCA		GAA	GGA	Gly			
		GUG		GCG		GAG	GGG				

② 遗传密码无逗号 密码与密码之间没有任何不编码的核苷酸, 阅读 mRNA 时是连续的。

③ 遗传密码不重叠 遗传密码不重叠是指一个蛋白质中为氨基酸编码的密码子上的 3 个核苷酸只参与编码 1 个氨基酸, 在多核苷酸链上任何两个相邻的密码子不共用任何核苷酸, 核苷酸本身不重叠使用。阅读 mRNA 时是以密码子为单位, 连续阅读。尽管密码子是不重叠的, 但是有的基因可能是重叠的, 这是因为同一段 DNA 序列通过不同的阅读框可以编码几种蛋白质, 例如噬菌体  $\phi$ X174 中的重叠基因。

④ 遗传密码具有通用性 在所有的生物中, 密码子字典几乎是通用的, 既适用于原核生物, 也适用于多数真核生物。

⑤ 遗传密码具有简并性 (degeneracy) 61 种有义密码子决定 20 种氨基酸, 必然同一个氨基酸有多个密码子。实际上除 AUG (Met) 和 UGG (Trp) 以外, 每个氨基酸都有一个以上的密码子, 这种现象

称为密码的简并。例如,有 9 种氨基酸( Phe, Tyr, Asp, Asn, Glu, Gln, Cys, His, Lys)各有 2 个密码子; 1 种氨基酸( Ile)有 3 个密码子; 5 种氨基酸( Pro, Thr, Ala, Gly, Val)各有 4 个密码子; 3 种氨基酸( Ser, Leu, Arg)各有 6 个密码子。同一氨基酸的不同密码子称为同义密码子( synonym)。一组密码子,为一种或两种氨基酸编码,仅在它们的 3'端碱基不同称为密码子家族( codon family, code word family)。密码子表中有 8 个密码子家族,每个家族有 4 种密码子,编码一种氨基酸,这时密码子的第三位碱基对密码子的性质没有影响。密码虽有简并性,但它们使用的频率并不相等,有的密码子使用的机会较多,有的几乎不用。

⑥ 有起始密码子和终止密码子 蛋白质合成的起始和终止信号含在密码子中,密码子 AUG 是与 N-甲酰甲硫氨酰-tRNA 结合的密码子,在原核生物中启动蛋白质的合成,所以称为起始密码子( initiation codon)。后来证明无论在真核生物还是在原核生物中 AUG(Met)都是用作起始密码子,但在少数情况下原核生物也用 GUG 作为起始密码子。另有 3 个密码子 UAA、UAG 和 UGA 不编码任何氨基酸,因此称为无义密码子( nonsense codon)或终止密码子( stop codon),或链终止密码子( chain terminating codon)。3 种无义密码子也被分别称为赭石型( ochre, UAA)、琥珀型( amber, UAG)、乳白型( opal, UGA)或赭色型( umber, UGA)密码子。终止密码子没有相应的 tRNA 存在,只供释放因子识别来实现翻译的终止,对一个序列来说是很重要的位点。

蛋白合成所用的阅读框( reading frame)决定于起始密码子和终止密码子的位置。在基因组序列分析中,潜在的基因往往是根据开放阅读框( open reading frame, ORF)而定义。

### 3.5.2 tRNA 与遗传密码

#### (1) 反密码子中的“摆动”(wobble)问题

在蛋白质的合成过程中, tRNA 的反密码子在核糖体内是通过碱基的反向配对与 mRNA 上的密码子相互作用的。反向配对的基本方式如下:



一个 tRNA 只有一个反密码子,而一个反密码子可识别两个以上的密码子,这种现象称为摆动假说( wobble hypothesis),即在密码子与反密码子的配对中,前两位碱基配对严格遵守碱基配对原则,第三位碱基可以“摆动”,因而使某些 tRNA 可以识别 1 个以上的密码子。根据摆动假说,可以推测 tRNA 识别密码子的数目由它的反密码子第一位碱基的性质来决定。第一位碱基为 C 或 A 时,可识别一种密码子;第一位碱基为 U 或 G 时,可以识别两种密码子;第一位碱基为 I 时,可以识别三种密码子。这合理地解释了密码子的简并性。tRNA 含有多种修饰碱基是造成密码子与反密码子发生摆动的重要原因之一。与其他 RNA 的简单甲基化修饰不同, tRNA 存在着从简单的甲基化到整个嘌呤环的重排等多种不同的修饰。当反密码子被修饰后,会产生特殊的配对方式。一般反密码子第一位( 5'端)上的 U 通常以修饰形式存在,改变配对性质。A 通常被变为 I 出现在反密码子第一位,能与 U、C 和 A 中的任何一种碱基配对。

#### (2) tRNA 的种类

遗传信息从 mRNA 分子转移到结构不相同的蛋白质分子,信息是以能被翻译成单个氨基酸的三联体密码子形式存在的,在这里起作用的是 tRNA 的解码机制。tRNA 在蛋白质合成中处于关键地位,它不但为每个三联体密码子翻译成氨基酸提供了接合体,还为准确无误地将所需氨基酸运送到核糖体上提供了运送载体。首先,氨基酸(AA)与 tRNA 结合生成有活性的 AA-tRNA,然后 tRNA 的反密码子识别 mRNA 的密码子。但是氨基酸本身不能识别密码子,只有结合到 tRNA 上生成 AA-tRNA,才能被带到 mRNA-核糖体复合物上,插入到正在合成的多肽链的适合位置。在这些过程中 tRNA 分工明确。

① 起始 tRNA 和延伸 tRNA 能特异识别 mRNA 模板上起始密码子的 tRNA 称为起始 tRNA, 其他 tRNA 统称为延伸 tRNA。原核生物的起始 tRNA 携带的是甲酰甲硫氨酸 (Met)。原核生物中 Met-tRNA 必须首先甲酰化生成 Met-tRNA 才能参与蛋白质的生成; 真核生物的起始 tRNA 携带的是甲硫氨酸 (Met)。

② 同工 tRNA (iso-tRNA) 由于遗传密码具有简并性, 一种氨基酸可能有多个密码子, 即多个 tRNA 代表同一种氨基酸, 在此将几个代表相同氨基酸的 tRNA 称为同工 tRNA。在一个同工 tRNA 组内, 所有 tRNA 均专一于相同的氨酰-tRNA 合成酶。同工 tRNA 有不同的反密码子, 能识别该氨基酸的各种同义密码, 又要有某种结构上的共同性, 而被 AA-tRNA 合成酶识别, 目前从一级结构上无法解释。有证据说明, tRNA 的二级和三级结构对它的专一性起着重要的作用。

③ 校正 tRNA 有些 tRNA 的突变体能抑制基因突变, 称为 tRNA 校正基因。在 tRNA 校正系统中, 第一次突变发生在 mRNA 上, 改变了某个密码子。在蛋白质的结构基因中, 一个核苷酸的改变可能使代表某种氨基酸的密码子变成终止密码子 (UAG、UGA、UAA), 使蛋白质合成提前终止, 产生无义突变。而第二次发生的抑制突变或校正突变是改变 tRNA 上的反密码子, 这样它仍然可识别取代了原来靶密码子的突变密码子, 插入一个氨基酸, 使蛋白质的功能得到恢复。这样, 无义突变的校正 tRNA 可通过改变反密码子区校正无义突变。错义突变是由于结构基因中某种核苷酸的变化使一种氨基酸的密码变成另一种氨基酸的密码。错义突变的校正 tRNA 通过反密码子区的改变将正确的氨基酸加到肽链上, 合成正常的蛋白质。有关生化和遗传学研究表明, 一种蛋白质的结构基因发生变异所产生的有害结果可被第二次变异消除。第二次变异就是对第一次变异的校正。所以基因一旦变异可以通过校正基因产生的校正 tRNA 进行校正。借此机制可以保证 mRNA 密码子翻译的准确性 (详见第 15 章)。

### 3.5.3 密码子的例外与特殊属性

#### (1) 线粒体密码子的偏离

最初人们认为生物中所有的遗传密码都是相同的, 即通用的。但是, 在线粒体基因组中线粒体密码子中存在不符合通用密码的偏离 (code deviation) 情况 (表 3-3)。在一些高等动物、果蝇、酵母以及真菌等生物线粒体基因组中线粒体基因的终止密码子 UGA 变为编码 Trp 的密码子, 另外一些编码核基因的密码子变为终止密码子, 其余的则是编码氨基酸的密码子相互改变。为什么线粒体的密码子会发生变化? 按摆动假说推算, 要识别 61 种密码子至少需要 31 种 tRNA, 而哺乳动物线粒体中只有 22 种 tRNA, 因此必须存在特殊的识别机制。其关键在于密码子与反密码子配对方式的简化, 如果一种 tRNA 可以识别一个密码子家族的 4 个密码子, 最少需要 23 种 tRNA, 而线粒体中 AGG 和 AGA 变为终止密码子 (表 3-3), 又可减少一种 tRNA, 正好是 22 种 tRNA。

#### (2) 密码子的使用情况

密码子虽然具有简并性, 但研究表明密码子的使用却具有选择、偏爱和倾向性。有的密码子使用的机会较多, 有的几乎不用。例如, 在大肠杆菌核糖体蛋白质基因的 1209 个密码子中, 编码苏氨酸的 4 个密码子中的 ACU 被使用 36 次, ACC 26 次, 而 ACA 只有 3 次, ACG 则完全没有使用。在动物密码子和其他生物的密码子中都存在类似情况。在简并密码子中, 不同生物往往偏向于使用其中一种。这种被经常使用的密码子称为偏爱密码子 (prefer codon)。同样, 每种氨基酸所具有的密码子数目并不与该氨基酸在蛋白质中出现的频率成正比。

#### (3) 稀有密码子

在原核生物和真核生物中编码硒代蛋白质 (selenoprotein) 的基因比较特殊。大多数蛋白质中不常用的氨基酸是翻译后加工的产物, 但硒代蛋白中的硒代半胱氨酸是翻译前由丝氨酸的修饰产生, 然后与其对应的 tRNA 结合。硒代半胱氨酸由通常为终止密码子的 UGA 编码 (表 3-3)。这种内部的 UGA 可被硒代-Cys-tRNA 识别, 在编码硒代蛋白的 mRNA 中, 称为硒代半胱氨酸插入序列的二级

结构使蛋白质合成装置翻译这种密码子。

表 3-3 偏离标准(通用)遗传密码的一些例子

生物	密码子	预期编码	实际编码
线粒体基因组			
哺乳动物(mammal)	UGA	终止密码 stop	Trp
	AGA, AGG	Arg	终止密码 stop
	AUA	Ile	Met
果蝇(Drosophila)	UGA	终止密码 stop	Trp
	AGA	Arg	Ser
	AUA	Ile	Met
酿酒酵母 (Saccharomyces cerevisiae)	UGA	终止密码 stop	Trp
	CUN	Leu	Thr
	AUA	Ile	Met
真菌(fung)	UGA	终止密码 stop	Trp
玉米(maize)	CGG	Arg	Trp
核基因组和原核生物基因组			
几种原生生物	UAA, UAG	终止密码 stop	Gln
Candida cylindracea	CUG	Leu	Ser
Micrococcus sp.	AGA	Arg	终止密码 stop
	AUA	Ile	终止密码 stop
Euphotis sp.	UGA	终止密码 stop	Cys
Mycoplasma sp.	UGA	终止密码 stop	Trp
	CGG	Arg	终止密码 stop
依赖顺序关系的密码子重排			
各种生物	UGA	终止密码 stop	硒代半胱氨酸

注: N 代表任何核苷酸。

### 3.5.4 核糖体的结构与功能

核糖体是合成蛋白质的细胞器,普遍存在于各类细胞以及真核生物线粒体和叶绿体中,是一种无膜包被的颗粒状结构,其唯一功能是按照 mRNA 的指令高效而精确地合成多肽链。

真核生物核糖体为 80S,其中 60S 大亚基由 49 种蛋白与 28S、5.8S 和 5S rRNA 组成,40S 小亚基由 33 种蛋白质和 18S rRNA 组成;原核生物核糖体为 70S,其中 50S 大亚基由 34 种蛋白质(编号 L1, L2, L3, ..., L34)与 5S rRNA 和 23S rRNA 组成,30S 小亚基由 21 种蛋白质(编号 S1, S2, S3, ..., S21)与 16S rRNA 组成。

核糖体的空间结构很复杂。rRNA 组成总分子量的 60%~65%对维持其结构完整与功能都是必需的,它们通过分子内或分子间碱基配对形成复杂的三级与四级构象。核糖体的几个结构域在蛋白质合成过程中具有特别的重要性。如大肠杆菌核糖体 30S 亚基上有 mRNA 的结合位点;16S rRNA 的 3'端有一段序列(CCUCC)与 mRNA 起始密码子 AUG 上游的 SD 序列(Shine-Dalgarno sequence)互补结合,促进起始复合物的形成;核糖体上有 3 个 rRNA 结合位点:A 位点(aminacyl rRNA site,氨基乙酰 rRNA 位点)、P 位点(peptidyl rRNA site,肽酰 rRNA 位点)和 E 位点(exit site,离开位点)。A 位点在延长过程中与进入的 rRNA 结合,而 P 位点与携带新生多肽链的 rRNA 结合。A 位点和 P 位点两者靠得很近,由大小两个亚基共同组成;E 位点,空载的 rRNA 从此位点排出,该位点位于大亚基上;大亚基还具有肽基转移酶结构域,可催化氨基酸间形成肽键,这是蛋白质合成中的关键反应,具有这种酶活性的不是蛋白质,而是位于大亚基中的 rRNA;GTP 酶结构域,为延伸因子 EF-G 的结合位点,对 GTP 具有活性,可催化肽酰-rRNA 从 A 位点转移到 P 位点,促使肽链的延伸。除此之外,还



有为多种蛋白质合成因子提供的结合位点。

### 3.5.5 蛋白质的合成

蛋白质的合成即翻译(translation),是由 rRNA 将相应的氨基酸带到核糖体-mRNA 复合物上,按 mRNA 上的遗传信息装配成特定肽链的过程。核糖体大小亚基在细胞内常游离存在于细胞基质中,在蛋白质的合成中,它们要与 mRNA 结合形成完整的核糖体后,才能行使翻译的功能。其基本过程包括翻译的起始、肽链的延伸、合成的终止 3 个步骤。请查阅生物化学教材的相关章节,本书不再赘述。

## 3.6 中心法则及其发展

### 3.6.1 中心法则与遗传信息流

Watson 和 Crick 于 1953 年提出不朽的 DNA 双螺旋模型后,接着提出 DNA 半保留的复制方式,揭示了遗传信息的贮存和复制的分子基础,解决了 DNA 的自我复制问题,巩固了 DNA 作为遗传物质和遗传信息载体的地位。在这里,遗传信息是指核酸中的碱基序列以及蛋白质中的氨基酸序列。Crick 于 1958 年又提出了中心法则(central dogma)来说明遗传信息的传递方向和途径,所表示的遗传信息流的传递如下:



认为遗传信息的传递方向是从核酸(DNA)到核酸(RNA)或从核酸(RNA)到蛋白质,而不可逆转。由此可以看出中心法则的主要内容是:DNA 是自身复制的模板,DNA 通过转录作用将遗传信息传递给 RNA,最后 RNA 通过翻译作用将遗传信息表达成蛋白质。这样,中心法则揭示了遗传信息的传递方向,反映了 DNA、RNA 和蛋白质之间的相互关系。随着分子遗传学和分子生物学研究的深入和发展,中心法则的内容和形式都得到了修正、补充和发展。

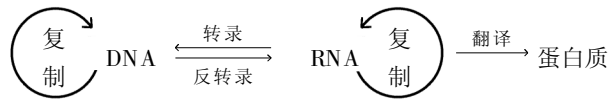
### 3.6.2 中心法则的修正与发展

#### (1) RNA 的复制

发现 RNA 依赖的 RNA 聚合酶催化以 RNA 为模板的 RNA 合成,即 RNA 复制。很多 RNA 病毒,如流感病毒、双链 RNA 噬菌体及多数单链 RNA 噬菌体,在感染宿主细胞后,它们的 RNA 在宿主细胞内进行复制,这种复制是以导入的 RNA 为模板,而不是通过 DNA。说明在某种情况下,RNA 像 DNA 一样是可以复制的,这是对中心法则的一种补充。

#### (2) RNA 反向合成 DNA

遗传信息从 DNA 向 RNA 的定向转移并不是绝对的。1970 年,Temin 等在某些引起肿瘤的单链 RNA 病毒(如 Rous 肉瘤病毒),即反转录病毒(retrovirus)中,发现一种反转录酶(reverse transcriptase),能以病毒 RNA 为模板,反向合成 DNA。在感染循环中,RNA 通过反转录(reverse transcription)成单链 DNA,单链 DNA 又形成双链 DNA,插入寄主基因组,成为细胞基因组的一部分,像其他基因一样遗传。这说明遗传信息流也可以反过来,从 RNA→DNA。因此,反转录使 RNA 序列作为遗传信息使用。这在遗传信息的传递问题上开辟了一条新的途径,说明遗传信息的形式可相互转化,这是中心法则的一项重要的新发展。



### (3) RNA 的自催化剪接

1981年, Cech等人在四膜虫中发现自催化剪接的 rRNA。1983年, Altman等发现大肠杆菌的核糖核酸酶的催化活性取决于 RNA 而不是蛋白质。这意味着 RNA 可以不通过蛋白质而直接表现出本身的某种遗传信息,而这种信息并不以核苷酸三联体来编码。这是对中心法则的又一次补充和发展。

### (4) DNA 水平的基因重排

基因重排 (gene rearrangement) 在原核生物和真核生物中都存在。例如,人的正常造血细胞向 B 或 T 淋巴系定向转化时,在重排信号的启动下发生特异性免疫球蛋白 Ig 和 TCRs (T cell receptors) 基因的重新组合,在结构上使分散排列在 DNA 上的 V、D、J 基因片段发生重排,形成了免疫球蛋白和 T 细胞受体的功能性基因。DNA 水平的基因重排是对中心法则的一大补充。这说明即使 DNA 作为遗传信息载体,为行使细胞的某种特殊功能也是可以发生改变的。

### (5) RNA 编辑

mRNA 因核苷酸的插入、缺失或置换,而改变了原 DNA 模板的遗传信息,翻译出不同于基因编码的氨基酸序列,称为 RNA 编辑 (RNA editing)。由于 RNA 编辑,使得基因产物的结构不能从基因组 DNA 序列中推导得到。编辑具有增加或改变遗传信息的作用。

### (6) 基因中内含子的切除和外显子的连接

1977年 Berger 首次报道腺病毒基因中存在内部间隔区,1978年 Gilbert 提出内含子、外显子概念,揭示了基因的不连续性。一个基因的外显子和内含子共同转录在一条转录产物中,然后将内含子去除而将外显子连接起来形成成熟的 RNA 分子。内含子和外显子的发现与 mRNA 的成熟加工是对中心法则的一个重要修正,它没有改变遗传信息流的方向,但 DNA 与成熟的 mRNA 并不是共线性,这是由于 RNA 加工所造成的结果。

### (7) 普里昂的感染与繁殖问题

普里昂又称朊粒 (prion) 或“感染性蛋白质粒子” (proteinaceous infectious particle), 是一种不含核酸分子只由蛋白质分子构成的病原体,能引起人及其他哺乳动物中枢神经系统疾病。如人类的库鲁症 (Kuru, 新几内亚震颤病)、克-雅氏综合征 (Creutzfeldt Jakob disease, CJD)、克-什氏综合征 (Gerstmann Straussler syndrome, GSS); 动物的羊瘙痒病 (scrapie)、疯牛病 (mad cow disease) 等。

普里昂完全由蛋白质组成。它既能作为蛋白质病原感染人类与其他哺乳动物,也能在被感染体内扩增繁殖。中心法则明确指出遗传信息流只能从核酸到核酸,从核酸到蛋白质,不能从蛋白质到核酸,也不能从蛋白质到蛋白质。蛋白质不是遗传物质,怎么能指导在被感染体内的扩增繁殖? 对朊粒相关蛋白 (prion related protein, PrP) 的研究揭示 PrP 是一种相对分子质量  $28 \times 10^3$  的疏水球蛋白。在人和其他哺乳动物的基因组内存在一个与朊粒相应的基因 PrP, 而且该基因是保守的,在正常脑内的正常表达时,其产物为 PrP<sup>C</sup>, 该蛋白可被蛋白酶完全水解;在被感染朊粒的脑细胞中该类蛋白是以 PrP<sup>Sc</sup> 的形式存在,对蛋白酶具有抗性。对 PrP<sup>C</sup> 与 PrP<sup>Sc</sup> 两者的比较发现它们在一级结构上是完全相同的,但在二级结构上却十分不同。PrP<sup>C</sup> 的  $\alpha$  螺旋的含量约为 40%, 几乎没有  $\beta$  片层结构,而 PrP<sup>Sc</sup> 则含有 50%  $\beta$  片层结构,只有 20% 的  $\alpha$  螺旋。这种结构上的差异可能导致了两者对蛋白酶抗性的不同以及其他特性的差异。

当将患有瘙痒病的羊组织接种到小鼠后,小鼠将在 75~150 天内发病。如果将小鼠的 PrP 基因敲除,小鼠则变成抗感染,不会发病。这说明内源性蛋白质对感染是必需的,推测可能是因为它提供转化为感染物的原始物质; PrP 基因敲除后,没有 PrP<sup>C</sup> 的小鼠正常存活,因此推测病因是 PrP<sup>Sc</sup> 作用的扩大,即朊粒的感染繁殖。

所谓朊粒的感染繁殖就是由正常的  $\text{PrP}^f$  蛋白质转变为  $\text{PrP}^{Sc}$  的过程。1991年 Prusiner等证明  $\text{PrP}^f$  转变为  $\text{PrP}^{Sc}$  是一个翻译后加工过程,即以  $\text{PrP}^{Sc}$  为模板,结合  $\text{PrP}^f$ ,使其转变成与  $\text{PrP}^{Sc}$  相同的构象。羊瘙痒病的发生伴随着脑细胞中的  $\text{PrP}^f$  转变为  $\text{PrP}^{Sc}$  的现象。是不是  $\text{PrP}^{Sc}$  会导致朊粒病呢?实验表明  $\text{PrP}^{Sc}$  本身不足以产生疾病,  $\text{PrP}^{Sc}$  只有产生于细胞内或者进入一个已经表达  $\text{PrP}^f$  的细胞时才有毒性,然后才能导致疾病的发生。缺少  $\text{PrP}$  基因的鼠不能被感染而发病,证明了  $\text{PrP}^f$  对疾病的发生是必不可少的,所有推测瘙痒病的发生是感染性的  $\text{PrP}^{Sc}$  与内源性的  $\text{PrP}^f$  共同作用的结果。显然,这是一种遗传信息未发生改变的表现遗传(epigenetic),因为正常的和患病的细胞有相同的  $\text{PrP}$  基因,但  $\text{PrP}^{Sc}$  形式是感染物,而  $\text{PrP}^f$  是正常型,对身体无害。

朊粒作为一种能够自我繁殖的病原体,对中心法则提出了新的挑战。众所周知,病毒的繁殖都是以病毒自身的核酸作为遗传物质进行核酸的复制,然后指导病毒外壳蛋白的合成,这是符合中心法则的规则的,而朊粒是不含核酸的蛋白质病原体,它的感染与繁殖对此提出了挑战。在细胞内,  $\text{PrP}^f$  必须与  $\text{PrP}^{Sc}$  结合,以  $\text{PrP}^{Sc}$  为模板使其转变成  $\text{PrP}^{Sc}$ 。这不同于一般的 DNA 复制和一般的蛋白质修饰,而是一种典型的表现遗传变异(epigenetic variation)。表现遗传变异是指在基因的核苷酸序列不发生改变的情况下,由于基因的修饰如甲基化、组蛋白的乙酰化等导致基因活性发生了改变,使基因决定的表型出现变化,且可传递少数世代,但这种变化是可逆的。在正常人体内,  $\text{PrP}$  基因的表达、 $\text{PrP}^f$  的产生是符合中心法则的。因此,在某种意义上看,朊粒的特殊的复制也是对中心法则的一种补充和发展。

### 3.7 基因的现代概念

在现代遗传学的分析中,某一个基因座在染色体上有一个特定的位置,一个基因在染色体上是一个编码的座位(coding locus),是一个携带从上一代到下一代遗传信息的物质单位和功能单位。按遗传术语讲,一个基因是合成一条有功能的多肽或 RNA 分子所必需的完整的 DNA 序列。除了编码区外,大多数基因也包含非编码的间插序列和转录控制区。一个结构基因包括启动子、RNA 编码区和终止区(图 3-20)。在真核生物的 RNA 编码区中一般还存在内含子和外显子的结构,在 RNA 的成熟加工中要切除内含子,将外显子连接起来。而启动子区域如前所述是很复杂的,真核生物与原核生物不同,真核生物不同的基因类型之间也不相同。在真核生物中由于启动子是按“混合相配”(mix and match)原理来构建的,不同的基因具有不同的启动子,因而实际上是没有一个固定不变的启动子结构模式的。

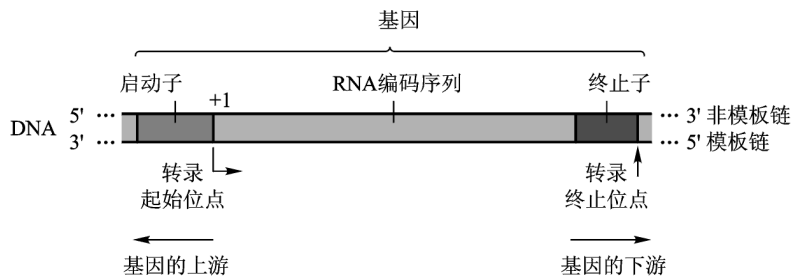


图 3-20 基因的结构(引自 Russell 1998)

一个基因应该由启动子、RNA 编码区和终止子等几部分组成

现代基因的概念将基因的结构和基因的功能联系起来,特别强调基因是合成一条有功能的多

肽或 RNA 分子所必需的完整的 DNA 序列。这里,完整的 DNA 序列提示我们不能忽略 RNA 编码区两端的转录调控区。在基因组研究中,往往利用被起始密码与终止密码所界定的一串密码子的可读框,作为鉴别和寻找编码蛋白质的基因的依据。严格地讲,根据现代基因的概念,这只是找到为蛋白质编码的 RNA 编码区,还不能认为找到了一个完整的基因。

当今,随着分子生物学技术和遗传学的发展以及各项基因组计划的实施,通过比较基因组、功能基因组的研究,人们会对基因的本质有更进一步的认识和了解,必将揭示和丰富基因的结构、功能的内容,基因概念也将会得到不断的更新和发展。

## ❓ 思考题

1. 如何得知 **DNA** 是遗传物质? 其实验设计的思路是什么? 你现在若要证明什么是遗传物质,将如何设计实验?
2. **DNA** 复制的基本规则有哪些? **DNA** 的复制模型有哪些?
3. 真核生物染色体端粒的复制有哪些特点?
4. 原核生物启动子的结构特点是什么?
5. 真核生物有 3 种不同的 **RNA** 聚合酶,识别不同的启动子类型,3 种不同的启动子的结构特点是什么?
6. **RNA Pol III** 启动子的上游启动子元件有哪些,各有哪些功能?
7. 何谓增强子? 增强子的主要作用特点有哪些?
8. **mRNA** 前体加工中加尾和加帽的生物学功能有哪些?
9. 何为 **RNA** 剪接,不同的 **RNA** 剪接方式有哪些特点?
10. 遗传密码的性质有哪些? 反密码子中的“摆动”会产生什么结果? 并请解释线粒体的密码子有哪些例外。
11. 何谓“中心法则”? 中心法则提出后有哪些修正? 目前又有哪些对中心法则的挑战? 请谈谈你的认识。