

第三章 基因表达调控

基因表达 (gene expression) 是基因所贮存的遗传信息表达为表型的过程, 它包括了基因的转录和翻译。基因表达调控 (regulation of gene expression or gene control) 体现在基因表达的全过程中, 主要发生在下面三个主要阶段: ①染色质激活。②转录起始、转录后加工修饰及转运。③翻译起始及翻译后加工修饰。

原核生物与真核生物的基因组结构及细胞结构有很大差异, 因此决定了它们的基因表达调控方式不同。原核生物中, 营养状况、环境因素对基因表达起着十分重要的作用; 而真核生物尤其是高等真核生物中, 激素水平、发育阶段等是基因表达调控的主要手段, 营养和环境因素的影响则为次要因素。

第一节 原核生物基因表达调控

1965年, 法国科学家 Jaques Monod 与 François Jacob 因为提出操纵子 (operon) 学说, 荣获诺贝尔生理学或医学奖, 开创了原核生物基因表达调控研究的新纪元。原核生物没有细胞核, 转录和翻译偶联在一起进行, 因此转录起始成为原核生物基因表达调控的主要环节。原核生物基因转录起始调控大多通过操纵子模型来实现。

一、操纵子是原核生物转录调控的基本单位

(一) 操纵子模型的特点

操纵子是基因表达的协调单位 (coordination unit), 通常由调控序列 (regulation sequence) 和结构基因在基因组中成簇串联组成。调控序列由启动子 (promoter, P)、操纵基因 (operator, O) 和调节基因 (regulator, R) 组成。一个操纵子的若干个结构基因排列在一起, 共用一个启动子及转录终止信号, 转录产生一条携带几个多肽链编码信息的 mRNA 长链, 称为多顺反子 mRNA (polycistronic mRNA) (图 3-1)。许多原核生物启动子序列非常相似, 如在转录起始点上游 -10 及 -35 区域存在的共有序列, Pribnow 框和 TATA 框。这些共有序列中的任一碱基突变或变异都会影响 RNA 聚合酶与启动子的结合及转录起始。因此, 共有序列决定启动子的转录活性大小。操纵基因是原核阻遏蛋白 (repressor protein) 的结合位点。当操纵基因结合阻遏蛋白时会阻碍 RNA 聚合酶与启动子的结合, 或使 RNA 聚合酶不能沿 DNA 向前移动, 阻遏转录, 介导负性调节 (negative regulation)。原核操纵子调控序列中还有一种特异 DNA 序列可结合激活蛋白 (activator protein), 增强 RNA 聚合酶的转录活性, 使转录激活, 介导正性调节 (positive regulation)。

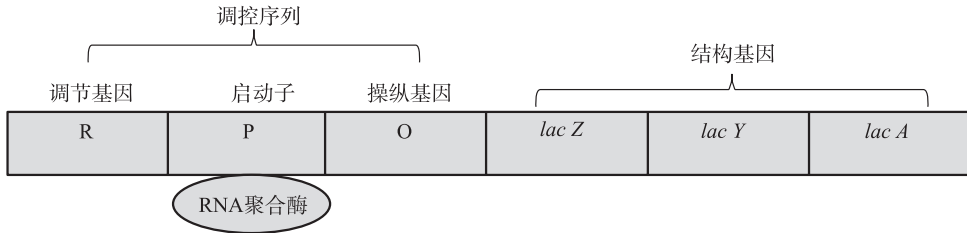


图 3-1 操纵子的基本结构

(二) 乳糖操纵子

很早就发现，在含有乳糖和葡萄糖的混合培养基中，*E. coli* 首先利用葡萄糖维持生长。1961 年，法国科学家 Jaques Monod 与 François Jacob 根据大量实验数据发表《蛋白质合成中的遗传调节机制》一文，揭示了这种现象的机制，提出乳糖操纵子 (*lac operon*) 学说，阐明乳糖操纵子是原核生物转录调控的基本模式。

1. 乳糖操纵子的结构 乳糖操纵子主要负责调控大肠杆菌及其他肠道菌科乳糖的运输及代谢。它包含了一个启动子 (P)、一个操纵基因 (O)、一个调节基因 (I)、三个相连的结构基因及一个终止子 (T)。I 基因具有独立的启动子 (PI)，编码阻遏蛋白，后者可与 O 基因结合，使操纵子受到阻遏而处于关闭状态。三个结构基因 Z、Y、A 分别编码 β - 半乳糖苷酶 (β -galactosidase)、 β - 半乳糖苷通透酶 (β -galactoside permease)、 β - 半乳糖苷乙酰基转移酶 (β -galactoside acetylase) (图 3-2)。除此之外，乳糖操纵子启动子上游还有一个分解代谢物基因活化蛋白 (catabolite gene activator protein, CAP) 结合位点，诱导乳糖操纵子的转录激活。



图 3-2 大肠杆菌乳糖操纵子的结构

2. 乳糖操纵子的作用机制 乳糖操纵子是典型的可诱导型操纵子，受两种机制的调控：一是对 RNA 聚合酶结合到启动子上的正调控；二是对操纵基因的负调控。

细菌通常优先利用葡萄糖为能源，即使加入乳糖、阿拉伯糖等其他糖类，只要有葡萄糖的存在，它只选择葡萄糖代谢，直到葡萄糖消耗完毕，代谢其他糖类的酶才会被诱导，这种现象叫做“葡萄糖效应”。这种现象可以用乳糖操纵子正、负反馈机制来解释。

(1) 乳糖操纵子的负调控：在没有乳糖存在时，乳糖操纵子处于阻遏状态。此时，I 基因在 PI 启动子作用下表达的乳糖阻遏蛋白与 O 基因结合，故阻断 RNA 聚合酶与 P 序列结合，抑制转录启动。阻遏蛋白的阻遏作用并非绝对，偶有阻遏蛋白与 O 序列解聚。因此，每个细胞中可能会有很少量的 β - 半乳糖苷酶、 β - 半乳糖苷通透酶生成。

当有乳糖存在时，乳糖操纵子即可被诱导。真正的诱导剂并非乳糖本身。乳糖经 β - 半乳糖苷通透酶催化、转运进入细胞，再经原先存在于细胞中的少数 β - 半乳糖苷酶催化，转变为半乳糖和葡萄糖 (图 3-3)。另外， β - 半乳糖苷酶可以转化乳糖以 β -1,6- 糖苷键方式结合，形成别乳糖 (乳糖的结合方式是 β -1,4- 糖苷键，图 3-4)。别乳糖则作为一种

诱导剂结合阻遏蛋白，使蛋白构象变化，导致阻遏蛋白与 O 序列解离、发生转录，使 β -半乳糖苷酶分子增加 1 000 倍。

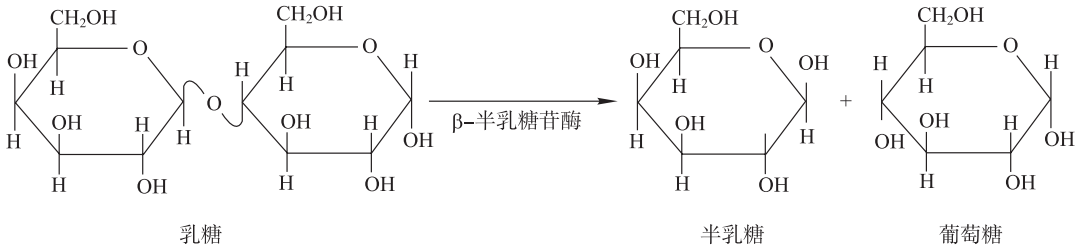


图 3-3 β -半乳糖苷酶的催化反应

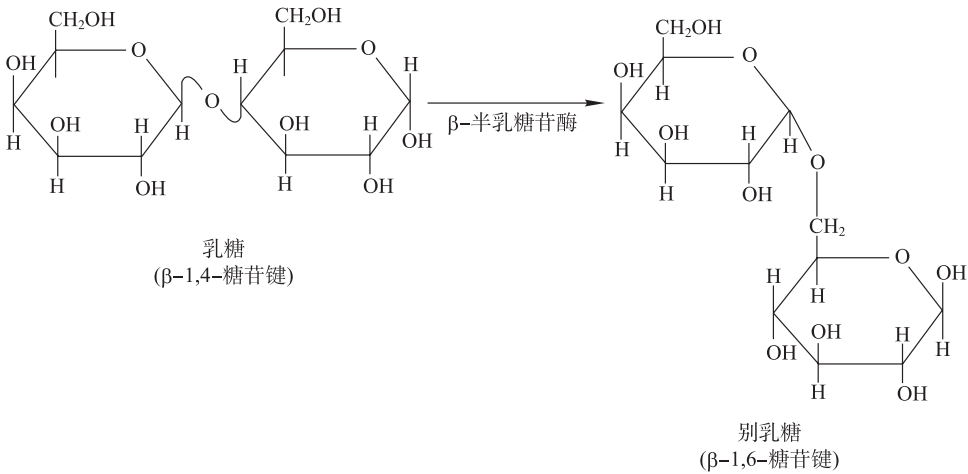


图 3-4 乳糖的转化

(2) 乳糖操纵子的正调控：CAP 是同二聚体，在其分子内有 DNA 结合区及 cAMP 结合位点。CAP 特异地结合在启动子附近的 CAP 位点时，能促进 RNA 聚合酶与启动子结合，促进转录。但游离的 CAP 不能与启动子结合，必须在细胞内有足够的 cAMP 时，CAP 首先与 cAMP 形成复合物，此复合物才能与启动子相结合。葡萄糖的降解产物能降低细胞内 cAMP 的含量，当向乳糖培养基中加入葡萄糖时，造成 cAMP 浓度降低，CAP 便不能结合在启动子上。此时即使有乳糖存在，RNA 聚合酶不能与启动子结合，此时虽已解除了对操纵基因的阻遏，也不能进行转录，所以仍不能利用乳糖。

(3) 乳糖操纵子的协同调节：对乳糖操纵子来说，CAP 是正性调节因素，阻遏蛋白是负性调节因素。两种调节机制使得细菌在葡萄糖和乳糖存在与否的情况下有 4 种不同的协同调节方式（图 3-5）：

1) 葡萄糖和乳糖都存在：这种状态基因是关闭的。虽然乳糖的存在可以解除阻遏蛋白对转录的抑制作用，但由于葡萄糖降低了细胞内 cAMP 的水平，降低 cAMP-CAP 复合物的水平，使 CAP 不能结合到 CAP 结合位点上，乳糖操纵子的正调控作用不能发挥，导致转录不能启动，因此维持基本转录水平。

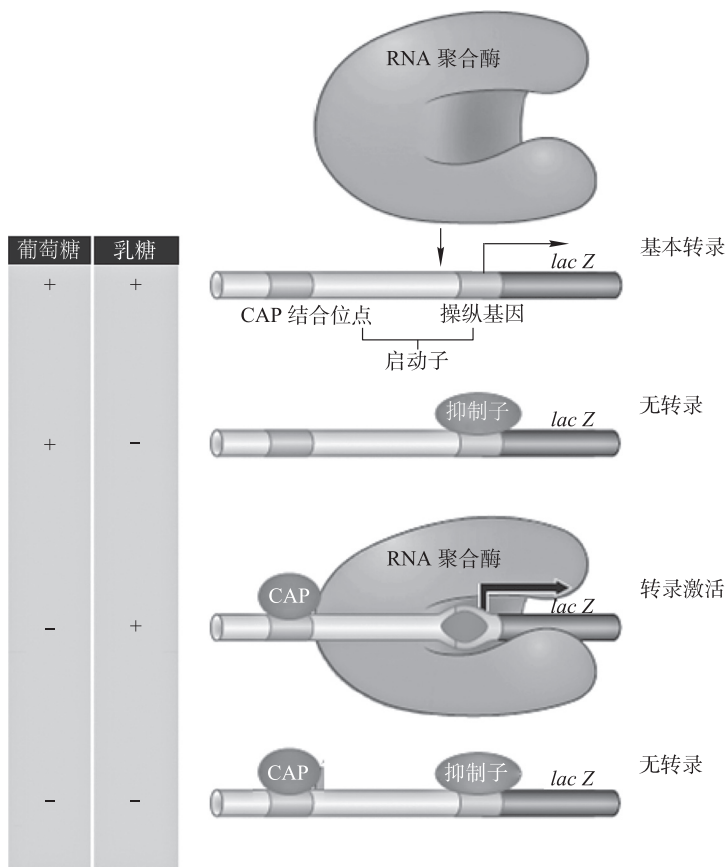


图 3-5 乳糖操纵子转录起始的调节机制

2) 葡萄糖存在、乳糖不存在：这种状态基因是关闭的。因为阻遏蛋白对转录的抑制作用无法解除，葡萄糖降低了 cAMP-CAP 复合物的水平，导致乳糖操纵子的正调控作用不能发挥，负调控作用不能解除，转录不能启动。

3) 葡萄糖不存在、乳糖存在：这种状态基因被打开。阻遏蛋白由于诱导剂的存在失去负调控作用，此时葡萄糖不存在使得 CAP 可以发挥正调控作用，基因被打开，启动转录。如此看来，只有协同调节，乳糖操纵子才能发挥作用。

4) 葡萄糖和乳糖都不存在：这种状态基因是关闭的。由于别乳糖不存在，无法解除阻遏蛋白对转录的抑制，即使 CAP 可以发挥正调控作用，但没有诱导剂，阻遏蛋白的负调控依然可以使基因处于关闭状态。

🔗 深入学习 3-1：阿拉伯糖操纵子

🔗 图 3-1 阿拉伯糖操纵子的结构

🔗 图 3-2 阿拉伯糖操纵子转录起始的调控机制

🔗 深入学习 3-2：色氨酸操纵子

🔗 图 3-3 色氨酸操纵子的结构及阻遏调控作用

🔗 图 3-4 色氨酸操纵子的衰减机制

二、原核生物转录调控的主要环节

原核生物的基因组较小，如大肠杆菌的基因组约为 4×10^6 bp，约含 4 000 个基因。通常只有 5% 左右的基因高表达，其余大多数基因不表达或表达水平极低。虽然原核基因的表达也受转录起始、转录终止、翻译调控及 RNA、蛋白质的稳定性等多级调控，但其表达开、关的关键机制主要发生在转录起始。

(一) 转录水平调控

其特点包括以下 3 方面：① σ 因子决定 RNA 聚合酶的识别特异性：原核生物只有一种 RNA 聚合酶，核心酶催化转录的延长， σ 亚基识别特异启动序列，即不同的 σ 因子协助启动不同基因的转录。② 操纵子模型的普遍性：除个别基因外，原核生物绝大多数基因按功能相关性成簇地连续排列在染色体上，共同组成一个转录单位即操纵子，如乳糖操纵子等。在同一个启动序列控制下，转录出多顺反子 mRNA。③ 阻遏蛋白与阻遏机制的普遍性：很多原核操纵子系统，特异的阻遏蛋白是控制转录的重要因素。当阻遏蛋白与操纵基因结合或解离时，结构基因的转录被阻遏或去阻遏。

(二) 翻译水平调控

原核生物中，翻译和转录是同步进行的。除了在转录水平的调控以外，在蛋白质水平也可以进行调节，保证了基因表达调控的有效性和精确性。

1. 影响翻译起始效率的因素 首先，在细菌 mRNA 起始密码子 AUG 上游 8~13 核苷酸的位置上有一个 SD 序列，能与细菌 16S rRNA 3' 端识别，有助于 mRNA 在核糖体的小亚基上准确定位。因此，SD 序列的结构特征会影响翻译起始效率。其次，蛋白质与自身 mRNA 结合可以调控翻译起始水平引起翻译阻遏。组成核糖体的蛋白质有 50 多种，它们的合成与 rRNA 保持相应的水平。当有过量的游离核糖体蛋白质存在时，可引起它自身以及有关蛋白质合成的阻遏，这种在翻译水平的阻遏作用称为翻译阻遏 (translational repression)。能够起阻遏作用的调节蛋白都可以与 rRNA 相结合而影响翻译。如组成核糖体小亚基的 S8 蛋白质，能与 16S rRNA 的茎-环结构结合；组成核糖体大亚基的 L5 蛋白质，它的 mRNA 的 5' 端能够形成一个类似于 16S rRNA 茎-环结构的结构，也能与 S8 蛋白质结合。当 16S rRNA 含量不足时，多余的 S8 与 L5 的 mRNA 结合，阻遏 L5 蛋白质的合成。而当 16S rRNA 含量丰富时，S8 蛋白质能全部与之结合，不影响 L5 蛋白质的合成。

2. 反义 RNA 抑制 mRNA 的翻译 细菌和病毒存在一类调节基因，能够转录产生反义 RNA (antisense RNA)。反义 RNA 能够与特定的 mRNA 序列互补结合，形成一个双链区，从而抑制 mRNA 的翻译。

3. 密码子使用频率影响翻译延伸速度 除了色氨酸和甲硫氨酸外，其余的氨基酸都有 2 个或 2 个以上的遗传密码子。常用密码子往往使用频率较高，稀有密码子往往使用频率较低。当基因中的密码子是使用频率较高的常用密码子时，mRNA 的翻译速度快，反之，mRNA 的翻译速度慢。

第二节 真核生物基因表达调控

真核生物的基因表达更加复杂，在多个环节上都受到严格的调节控制。与原核生物相比，真核生物基因表达调控具有一些显著的特点：①真核生物基因表达体现出多级调控。真核生物基因表达涉及多个环节，主要包括基因的活化、转录、转录后修饰、翻译和翻译后修饰等。遗传信息传递过程中的任何改变，均可能导致基因表达的变化。转录起始同样是真核生物基因表达调控的最关键点。②真核生物基因表达调控以正性调节为主。这一点与原核生物基因表达调控以负性调节为主恰恰相反。虽然已经发现一些真核生物基因含有负性顺式作用元件，但真核生物基因表达调控仍以正性调节为主。原核细胞的 RNA 聚合酶可以直接与启动子结合，多数基因只要解除了阻遏蛋白的抑制作用，就可以一定程度的激活基因转录。真核细胞中的基因转录则比较复杂，首先需要基因相应染色质区域结构变得疏松，组蛋白解离，然后先由基本转录因子如 TF II D 等辨认结合启动子区的 TATA 框等功能元件，再招募 RNA 聚合酶形成转录前起始复合物，基因转录的启动通常还需几种能够识别结合增强子的特异转录因子以及辅助激活因子的协同作用。

一、真核生物 RNA 聚合酶和启动子是真核生物转录调控的关键

(一) 真核生物 RNA 聚合酶的特点

RNA 聚合酶又称为 DNA 依赖的 RNA 聚合酶 (DNA dependent RNA polymerase)，通过识别并结合启动子，从而启动基因的转录。由于结构基因不同，启动子的碱基序列存在差异，RNA 聚合酶与不同启动子之间的亲和力也不同。

RNA 聚合酶在原核生物中只有一种，在真核生物中有 3 种，RNA 聚合酶 I 位于细胞核的核仁区，催化转录 rRNA 前体；RNA 聚合酶 II 位于细胞核基质中，转录生成核不均一 RNA (hnRNA) 及一些重要的基因表达调节作用的非编码 RNA，如长链非编码 RNA (lnc RNA)、微小 RNA (miRNA) 和 piRNA (与 Piwi 蛋白相互作用的 RNA)；RNA 聚合酶 III 存在于核基质中，转录 tRNA、5SrRNA 及 snRNA 基因。在 3 种真核生物 RNA 聚合酶中，最重要的是 RNA 聚合酶 II，共具有 12 个亚基，最大亚基的分子结构具有独特的羧基末端结构域 (carboxyl terminal domain, CTD)，该结构域由 7 个氨基酸残基 (Y-S-P-T-S-P-S) 的重复序列 (27~52 次) 所构成。RNA 聚合酶 I 和 III 中都没有 CTD 结构域。已知存在不同的 CTD 激酶 (kin28, Ctk1)，分别催化其 S2 或 S5 的磷酸化修饰。这些 CTD 激酶的活性改变均与细胞周期素浓度改变相关。真核 RNA 聚合酶 II 的 CTD 结构不仅是转录和转录加工复合物装配的平台，其不同位点的可逆磷酸化 / 去磷酸化修饰，对转录的起始、延伸、终止及转录加工均具有重要的调节作用。

(二) 真核生物启动子

真核生物基因的启动子是 RNA 聚合酶的基本转录因子识别并结合的转录调控区中一段特异的 DNA 序列，通常包含一组转录调控功能组件，每一个功能组件的 DNA 序列均由 7~20 个碱基组成。它和原核生物的启动子有很多不同：①有多种元件，包括 TATA 框、GC 框、CAAT 框、OCT 等 (图 3-6)；②结构不恒定，有的有多种框，如组蛋白 H2B，有的只有 TATA 框和 GC 框，如 SV40 早期转录蛋白；③它们的位置、序列、距离和方向都不完全相同；④有的有远距离的调控元件存在，如增强子；⑤这些元件常常起到控制转录效率和选择起始位点的作用；⑥不直接和 RNA 聚合酶结合，转录时先和其他转录激活因

子相结合，再和聚合酶结合。

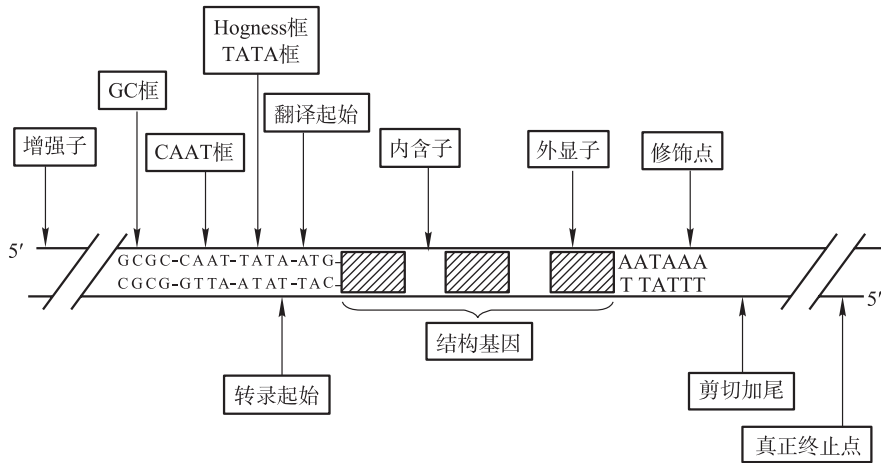


图 3-6 真核生物启动子结构

🔗 深入学习 3-3：真核生物的 3 种启动子

实验证实，启动子具有严格的方向性，需定位于转录起始点的上游。典型的启动子由 TATA 框及其上游的 GC 框与 CAAT 框所构成，拥有典型启动子的基因通常具有一个固定的转录起始位点并具有较高的转录活性。然而，有些基因的启动子可以缺失 GC 框和（或）CAAT 框，甚至可以缺失 TATA 框。不含 TATA 框的基因通常有多个转录起始位点，可转录出多种 5' 端不同的转录物。不含 TATA 框的启动子，有的含有起始子元件（initiator, Inr），有的富含 CpG 岛（CpG island），有的还含有数个 GC 框。这类启动子的活性通常较低，主要在胚胎发育和组织再生中发挥重要作用。而所谓 CpG 岛是指在真核生物基因组的某些区域中，CpG 序列密度很高，成为鸟嘌呤和胞嘧啶的富集区（CpG 岛将在第四章基因表达的表观遗传调控中重点介绍）。值得一提的是，许多真核基因的启动子区域可同时拥有 TATA 框和 Inr。

（三）增强子和沉默子

见第二章第二节深入学习 2-3，2-4。

（四）真核生物转录因子

真核生物的顺式作用元件都需要与特异的 DNA 结合蛋白相结合，才能发挥对转录以及其他相应过程的调节功能。这些能够直接或间接地识别结合顺式作用元件，并参与调控靶基因转录的蛋白质统称为转录调节蛋白（transcription regulatory protein）或转录因子（transcription factors, TF）。大多数转录因子以反式作用方式调节基因转录，因此也称反式作用因子（trans-acting factor）。但也有少数转录因子以顺式作用方式调节基因转录，可称为顺式作用蛋白（cis-acting protein）。顺式作用蛋白是指那些能够调节自身表达的蛋白质，可以看作是反式作用因子的特例。

二、真核生物转录调控的主要环节

（一）转录前调控

真核生物的重要特征之一是 DNA 与组蛋白质结合，形成以核小体（nucleosome）为基

本单位的染色质结构，赋予了真核生物基因转录前独特的调控机制。真核基因常需根据机体生长发育的需要，调节基因转录活性，因此染色质水平调控对基因表达有着重要的影响。常常涉及 CpG 岛甲基化修饰、组蛋白修饰和核小体重塑等当今表观遗传学领域的研究热点（详见第四章第一节表观遗传调控研究内容）。

1. 转录基因与核小体的结构 基因转录的模板是染色质 DNA，在转录时，RNA 聚合酶和大约 50 bp 的 DNA 结合，其中有 12 bp DNA 需发生解链。RNA 聚合酶（相对分子质量约 5.0×10^5 ）比一个核小体（相对分子质量约 2.6×10^5 ）大近似 2 倍。因此，早期的观点认为，只有在染色质结构被破坏、核小体解体、DNA 裸露的条件下，RNA 聚合酶才能进行转录。随着人们对染色质结构研究的日渐深入，近年来有不少实验表明 RNA 聚合酶进行基因转录时并不是在裸露的模板链进行的。如在 SV40 的染色体中，正在转录的基因仍然有核小体结构。用微球菌核酸酶处理染色质，然后进行凝胶电泳，发现转录的基因和不转录的基因都出现 200 bp 间隔的 DNA “梯”带。这也说明转录的基因和非转录的基因其核小体结构没有显著差异，转录基因对微球菌核酸酶更为敏感，降解更快，导致核小体数目略少。也可能是 RNA 聚合酶所经过的核小体部分中组蛋白暂时分离，待 RNA 聚合酶离去后立即恢复了核小体的结构。对核小体结构动态变化的研究表明，真核细胞染色质中核小体存在两种构象改变：①相邻的或较远的核小体聚集。②解聚为 H2A-H2B 二聚体和 (H3-H4)₂ 四聚体。组蛋白 (H3-H4)₂ 四聚体组装到核小体中可使核小体形成稳定构象，然而 H2A-H2B 二聚体结合到核小体极易破坏核小体的稳定构象，使核小体结构疏松。这种疏松结构的核小体可能正是 RNA 聚合酶有效转录的基础。通常，这两种聚合物处于动态平衡状态，但外界因素如连接区的组蛋白、装配蛋白或转录因子等可促使 H2A-H2B 二聚体和 (H3-H4)₂ 四聚体平衡的改变。

核小体的相位与基因转录也有着重要的联系。相位（phase positioning）指的是同类型细胞中，组蛋白八聚体在 DNA 序列上的特殊定位。由于染色质中核小体的相位改变，使涉及某个基因的启动子和增强子序列被暴露出来，这样的核小体相位就会对该基因的调控产生重要的影响。特异序列的核小体相位还可以促进远距离调控元件的联系，通过顺式作用元件彼此接近，形成一种特殊的三维结构。如在有关的体外模型中类似的结构可以使爪蟾卵黄蛋白基因转录活性提高 10 倍。

2. 活性染色质的特征 真核细胞的染色质可分为常染色质（euchromatin）和异染色质（heterochromatin）。对昆虫、线虫等低等生物染色体的研究表明，基因转录激活是在常染色质上进行的。转录发生之前，常染色质特定的区域被解螺旋或结构变得很疏松，使用 DNase I 对其进行处理，可将 DNA 降解为约 200 bp 或其倍数的片段。其基因表达处于活性状态，故称为活性染色质（active chromatin）。异染色质结构高度致密，处于凝聚状态；因此无论是 DNase I，还是基本转录因子或 RNA 聚合酶都无法识别结合，其基因表达处于阻遏状态。

通过对活性染色质部分的染色质蛋白分析表明：①组蛋白 H1 与活性染色质结合并不紧密；②尽管 4 种核内组蛋白在活性染色质中的含量正常，但乙酰化程度比非活性染色质中的高；③在活性染色质中，组蛋白 H2B 较少磷酸化；④变异较少的 H2A 在活性染色质中有很高程度的富集；⑤活性染色质中的核小体结合了两种相关的、相对分子质量相对小的高迁移率群蛋白质（high-mobility group protein, HMG protein）HMG14 和 HMG17，这些 HMG 蛋白是非组蛋白，全部在染色质中，每 10 个核小体结合 1 个，而且其氨基酸序列在

进化过程中是高度保守的，它们有很重要的功能。

3. 蛋白质修饰与基因活化调节 核小体组蛋白上某些特定的氨基酸常可被共价修饰。修饰作用包括乙酰基化、甲基化和磷酸化等。其中最主要的修饰方式是组蛋白赖氨酸残基的乙酰化，它促使相邻核小体的聚合受阻，还影响泛素（ubiquitin, Ub）与组蛋白 H2A 分子的结合，导致蛋白质的选择性降解。组蛋白 H3、H4 是蛋白酶修饰的主要靶点，它们的乙酰化使得核小体间的 DNA 因产生较多的负螺旋而易于从核小体上脱离，对核酸酶的敏感性增高。组蛋白 H1 对核小体起装配作用，确定核小体的方向性，并能进一步将核小体从 10 nm 纤维组装到 30 nm 纤维中，而活化的染色质对组蛋白 H1 的亲合力极低，因为组蛋白 H1 的丝氨酸位点在有丝分裂期发生磷酸化所致。构成核小体的组蛋白并不与 DNA 序列专一结合，而且组蛋白种类少，进化中很稳定。在不同物种、不同组织细胞的染色质中，功能完全类似；由此看来，组蛋白只是在染色质结构中起作用，组蛋白的修饰作用是降低组蛋白所携带的正电荷，短暂松弛它们与 DNA 的结合，使染色质活化，利于基因转录。组蛋白修饰作用往往具有时空性。

细胞核内除组蛋白外，还含有多达数百种的非组蛋白，它们在不同组织细胞里的种类和数量都不相同，具有组织特异性。非组蛋白的相对分子量一般比组蛋白大，性质也不像组蛋白那样稳定，在细胞中变化迅速。在代谢旺盛的组织中，当 RNA 大量合成时，非组蛋白的合成量也同时增加。非组蛋白与 DNA 结合有特异性。一般认为，在基因表达中，组蛋白是起抑制作用，而非组蛋白则是起调节作用。已知的非组蛋白中主要有多种是参与核酸代谢和修饰的酶类，其中许多非组蛋白则是具有重要调控作用的反式作用因子。非组蛋白对基因转录的调控可用非组蛋白的转位模型来解释。首先非组蛋白特异性地连接到一个 DNA 位置上，在磷酸化酶作用下，非组蛋白被磷酸化带有负电荷，与具有负电荷的 DNA 相斥，并强烈地与带正电荷的组蛋白结合；于是这种组蛋白-非组蛋白聚合物从 DNA 上转移位置，DNA 裸露出来不再受组蛋白的抑制而转录其特定的 DNA 区段。

高迁移率蛋白质群是一组较丰富而不均一、富含电荷的非组蛋白。它们的相对分子量不大，一般 $\leq 3.0 \times 10^4$ ，因在聚丙烯酰胺电泳中迁移率很高而被命名。它们与染色质其他组分的结合较松弛，具有转录活性的核小体常常缺乏 H1，却有相对分子量在 3.0×10^4 以下的没有组织特异性的非组蛋白 HMG14 和 HMG17 存在。如用 DNA 酶 I 处理鸡红细胞染色质时，优先释放出 HMG14 和 HMG17 这两种非组蛋白。新近的研究发现，在爪蟾卵提取物体系中，HMG17 掺入到 DNA 复制后新组装的染色质中，会促进 RNA 聚合酶 III 的基因转录；而 HMG14 则可直接参与 RNA 聚合酶 II 对染色质中基因的转录等等。可见这些 HMG 蛋白与核小体结合是产生活性染色质的重要细胞过程。

4. DNA 的甲基化和去甲基化与基因活性的关系 甲基化多发生在 CpG 二核苷酸对上。真核生物 DNA 中 20% ~ 70% 的胞嘧啶存在着甲基化修饰，以卫星序列的甲基化程度最高。使用某些 II 型限制性内切酶有助于研究 DNA 序列的甲基化状况。如限制酶 *Hpa* II 识别并切割未甲基化的 CCGG 序列，对甲基化的 CG 对则不起作用，而限制酶 *Msp* I 识别并切割所有的 CCGG 序列，因此，可以用 *Msp* I 来识别 CCGG 序列的存在与否，用 *Hpa* II 来鉴别这些 CCGG 序列是否甲基化。实验证明，DNA 甲基化抑制转录。如在特异性表达某些基因的组织中，基因激活比非表达组织中 5-mCCGG 明显降低 30% 左右；同时，哺乳动物细胞核内 80% 的 CG 甲基化存在于有组蛋白 H1 的紧密结构的核小体中。可见基因表达与 DNA 甲基化呈负相关。近年有关的研究还表明，DNA 甲基化对转录的抑制主要取决

于 CG 甲基化的密度和启动子的强度这两个因素。启动子附近高甲基化是阻遏作用的主要决定因素之一。弱启动子可以被散在的 CpG 甲基化阻遏，但是增强子可使启动子恢复转录；若是甲基化位点或者程度进一步增加，转录就会完全停止。说明转录的激活和阻遏之间的转换，决定于甲基化的密度和启动子强度间的平衡。

甲基化不是一种原始现象，而可能是出现于某一进化时期，并随着进化程度的提高而逐步增强。因此甲基化或去甲基化对基因表达调控的作用随生物的不同种类而异，即便同一种生物，甲基化对不同基因活性也有不同的效应。如 rDNA 能够在甲基化的情况下依然进行转录。对于其他大多数基因，则是通过甲基化或去甲基化的方式来调节转录。

5. 基因丢失、重排、扩增与基因活性的调节 分子生物学的研究已经发现，在个体发育过程中，作为 RNA 合成的 DNA 模板也会发生规律性的变化来控制基因表达和生物体的发育。这种变化往往是由于基因本身或其拷贝数发生了永久性的改变，包括基因丢失、基因扩增和基因重排等方式。显然，这些调控方式与转录和翻译水平的调控不同，是从根本上使一些细胞的基因组发生改变。

(二) 转录水平的调控

1. 转录起始调控 转录起始同样也是真核基因表达调控的最重要环节。以 RNA 聚合酶 II 为例来阐述真核基因转录水平的调控过程。在真核生物的转录起始过程中，首先由基本转录因子 TF II D 的亚基 TATA 框结合蛋白 (TATA box binding protein, TBP) 识别、结合启动子 TATA 框或起始子元件，再由 TF II D 的另一种亚基 TBP 相关因子 (TBP associated factor, TAF) 招募 RNA 聚合酶 II 和其他基本转录因子 TF II A、TF II B、TF II F 和 TF II H，组装形成转录前起始复合物 (preinitiation complex, PIC)。在此过程中，TBP 是唯一具有序列特异性的 DNA 结合因子，在转录前起始复合物的组装中起核心指导作用。TAF 具有细胞特异性，与特异转录因子一起决定基因的时空特异性表达。TF II H 具有解旋酶活性，可解开 DNA 双螺旋，便于转录的发生；TF II H 还具有激酶活性，能使 RNA pol II 的羧基末端结构域 (CTD) 磷酸化，这也是调节转录前起始复合物形成的重要环节。然而，许多基因的转录前起始复合物还需要特异转录因子识别并结合增强子才能有效启动基因的转录。在转录起始的过程中，所涉及的每一个因子都可能受到内外环境信号的调节，从而使转录起始的频率、效率产生差异，实现基因的时空特异性表达。

2. 转录后水平的调控 转录后水平的调控主要涉及修饰、剪接、编辑和转运等多个环节。

(1) mRNA 的修饰：除了组蛋白基因以外，其他真核 mRNA 在转录后都需在 5' 端加帽、3' 端加多聚腺苷酸 [poly (A)] 尾。帽子结构和 poly (A) 尾对于维持 mRNA 的稳定性有着重要作用。帽子及 poly (A) 尾形成的具体机制详见第六章第一节。蛋白质的合成速率往往与细胞质内其 mRNA 的水平和稳定性密切相关。真核生物管家基因的 mRNA 一般均较稳定，因为它们合成的蛋白质产物是维持细胞基本功能所必需的。但是，其他一些蛋白质，尤其是那些决定细胞周期的蛋白质，其合成必须按照严格的时间顺序开启和关闭。真核 mRNA 的降解主要有两种途径，究竟以何种途径降解，主要由 mRNA 的特异性所决定。第一种，mRNA poly (A) 尾的逐步短缩途径，是最常见的 mRNA 降解途径。当 mRNA 进入细胞质后，一方面可以作为模板指导蛋白质的合成，另一方面核酸外切酶会从 mRNA 3' 端开始切割，使 poly (A) 尾逐步缩短；当剩下约 30 个腺苷酸时，发生脱帽

现象, mRNA 随即被迅速降解。第二种, mRNA poly (A) 尾被特异核酸内切酶切割途径。这些核酸内切酶可以识别特定 mRNA 3' 非翻译区 (3'-untranslated region, 3'-UTR) 的特异序列, 将 mRNA 的 poly (A) 尾直接整个切除, 导致 mRNA 的迅速降解。

(2) 选择性剪接: 尽管多数真核生物的 hnRNA 只能加工形成一种 mRNA, 但也有许多真核基因的 hnRNA 可以通过选择性剪接 (alternative splicing) 的方式, 选择不同的加尾位点, 或去除不同的内含子而被加工形成不同的 mRNA。例如, 原肌球蛋白基因的 hnRNA 含有 11 个外显子, 通过选择性剪接, 可以产生 7 种 mRNA, 分别为脑、肝、骨骼肌、平滑肌和成纤维细胞编码 α -原肌球蛋白。就人类而言, 可能 95% 的基因存在着选择性剪接。

(3) mRNA 的编辑: mRNA 编辑 (mRNA editing) 也是 mRNA 转录后加工的一种调控方式 (第六章详述了 RNA 的加工与编辑)。

(4) mRNA 的转运: 估计只有约 1/5 的成熟 mRNA 能够通过核膜进入细胞质, 作为蛋白质翻译的模板。而留在细胞核内的 mRNA 大约在 1 h 内就会被降解。mRNA 通过核膜输出的过程是一个主动耗能的过程, 需要借助核输出受体 (nuclear export receptor) 的帮助。目前, 调控 mRNA 分子从细胞核转运至细胞质的机制尚不清楚, 仍待进一步研究。

(三) 翻译水平的调控

翻译水平的调控中, 最为重要的几个方面是: mRNA 自身的稳定性、翻译起始的调控及 mRNA 非编码区的调控等。

1. 翻译起始的调控 蛋白质翻译的速率很大程度上取决于翻译起始的速率。真核细胞内许多变化都可能激活胞内特异的蛋白激酶, 使一些翻译起始因子如真核生物翻译起始因子 -2 (eukaryotic initiation factor-2, eIF-2) 发生磷酸化, 从而影响翻译的速率。eIF-2 和 GTP 形成的复合物介导着起始甲硫氨酰 -tRNA 与核糖体小亚基之间的结合, 并识别结合 mRNA 上的起始密码子 AUG, 随后 eIF-2 发挥 GTPase 活性, 水解 GTP, 释放能量, eIF-2-GDP 解离, 大亚基结合小亚基, 形成完整的核糖体, 肽链翻译开始。eIF-2-GDP 的结合非常紧密, 需在鸟苷酸释放蛋白 (guanine nucleotide releasing protein, 即 eIF-2B) 的作用下, 才能释放 GDP, 重新结合 GTP, 再次参与到翻译的起始过程中 (图 3-7a)。但是当 eIF-2 被磷酸化后, eIF-2 不但不再释放 GDP, 而且 eIF-2 与 eIF-2B 之间的结合也会异常稳固, 使得 eIF-2 和 eIF-2B 的循环利用受到抑制 (图 3-7b), 蛋白质翻译速率大为降低。在哺乳动物细胞中, 细胞由生长期进入非生长期 (G_0 期) 的部分机制就涉及 eIF-2 的磷酸化调节。

2. mRNA 非编码区的调控 真核 mRNA 的非编码区包括 5'-UTR 和 3'-UTR。研究发现, 5'-UTR 的帽子结构、帽子结构与起始密码子 AUG 之间的距离、起始密码子 AUG 旁侧序列以及一些特异的二级结构 (如茎-环结构) 与特异蛋白的结合, 都会影响蛋白质的翻译效率。

大部分真核生物的蛋白质生物合成是按照翻译起始的扫描模型, 从靠近帽子结构的 AUG 开始。然而有些原癌基因和生长调节因子基因的 mRNA, 其翻译起始并不完全遵循上述规律。这是因为作为起始密码子的 AUG 与其旁侧序列关系密切, 当 5' 端第一个 AUG 附近序列不利于起始密码子扫描/识别时, 核糖体小亚基可能会无视第一个 AUG 而滑向下游第二个, 甚至第三个 AUG, 这种现象称为遗漏扫描 (leaky scanning)。Kozak 对 699 种脊椎动物 mRNA 的起始密码子两侧的核苷酸序列进行了分析, 并提出了相关的共有序列

gccRccAUGG，其中小写字母表示相对保守，而大写字母表示绝对保守，其中 AUG 代表起始密码子，上游的 R 代表嘌呤 A 或 G，但 A 出现的频率更高。后人将其称为 Kozak 序列。遗漏扫描现象的发生可能就是由于 5' 端 AUG 不在 Kozak 序列中。在一些特定情况下，细胞可能还会通过遗漏扫描机制调节相关蛋白质翻译的相对丰度。如某些细胞可以通过增加起始因子 eIF-4F 的表达，而偏爱使用距离帽子结构最近的 AUG 作为起始密码子。

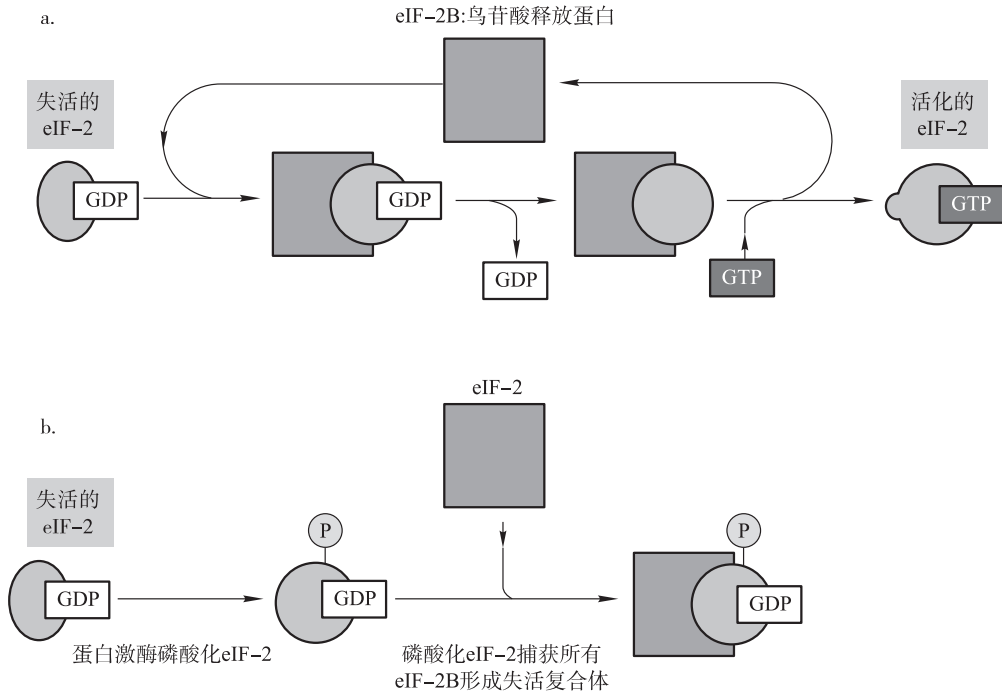


图 3-7 真核生物 eIF-2 的循环利用

另外一些 mRNA，在其正常的起始密码子 AUG 的上游有一个或多个 AUG 也能作为起始密码子，但这些上游开放阅读框 (upstream open reading frame, uORF) 翻译后会很快遇到终止密码子而释放出一个无功能的多肽。uORFs 存在的意义可能在于起调节作用，它们的 AUG 可与正常起始密码子 AUG 竞争，从而使正常的翻译产物维持在较低水平。

除此之外，还有一些特殊的蛋白质可以识别结合特定 mRNA 3'-UTR 序列，从而调节 mRNA 的降解速率。如转铁蛋白 (transferrin) 的 mRNA 5'-UTR 有一茎环结构——铁反应元件 (IRE)。当细胞质中可溶性铁离子的水平较低时，一种特异性翻译抑制蛋白——铁反应元件结合蛋白 (iron-response element binding protein, IREBP) 识别结合 IRE，抑制 mRNA 与核糖体小亚基的结合，从而阻断转铁蛋白的翻译。当细胞质中可溶性铁离子的水平较高时，铁离子结合 IREBP 使其变构，从 IRE 上脱落，蛋白质翻译抑制被解除，转铁蛋白的翻译效率显著增高。

(四) 翻译后水平的调控

新生肽链多数不具备活性，必须经过各种修饰、折叠成为成熟的蛋白质，才能发挥其特定功能。蛋白质的修饰有多种类型，如二硫键的形成、羟基氨基酸的磷酸化、精氨酸和赖氨酸的乙酰化和甲基化、脯氨酸和赖氨酸的羟基化以及丝氨酸、苏氨酸和天冬酰胺的糖

基化。这些修饰需要在特异蛋白酶的催化下完成，因此，对这些蛋白酶活性的调节也可以影响蛋白质的成熟过程，进而影响蛋白质的功能。

另外，虽然蛋白质折叠所需的信息蕴含在蛋白质的一级结构中，但蛋白质的顺利折叠需要依赖分子伴侣、二硫键异构酶、肽-脯氨酸顺反异构酶等蛋白质。因此，对这些酶（蛋白质）的活性和含量进行调节也属于蛋白质翻译后水平的调节范畴。

最后，真核细胞内各种蛋白质的数量，既取决于新生肽链的合成速率，也取决于它们在体内的半衰期（half-time）。细胞内蛋白质除可以通过溶酶体水解外，还可以通过精细的依赖 ATP 的泛素-蛋白酶体途径降解。因此，蛋白质的降解也是受到严格调控的。

（五）非编码 RNA 参与真核基因的表达调控

随着人类基因组测序的完成，发现其中蛋白质的编码序列仅占全序列的 2%，余下的大部分 RNA 并不能编码蛋白质，称为非编码 RNA。这些非编码 RNA 在真核基因表达过程中的多个环节发挥着重要的调控作用。

1. 短链非编码 RNA 指的是那些碱基组成小于 200 个核苷酸的非编码 RNA。短链非编码 RNA 的种类很多，目前发现的主要有 microRNA（miRNA）、小核 RNA（small nuclear RNA, snRNA）、核仁小 RNA（small nucleolar RNA）、snoRNA 等。

在短链非编码 RNA 的生成和执行功能的过程中，Argonaute（Ago）蛋白家族具有重要的作用。依据转录物的序列长度及其所依赖的 Ago 蛋白家族成员的不同，将短链非编码 RNA 分为三类：miRNA（约 22 nt）、内源性的小干扰 RNA（endo-siRNA，约 21 nt），和 Piwi-interacting RNA（piRNA，24~31 nt）。这些短链非编码 RNA 的发生机制、生理功能及其调控作用各不相同，以 miRNA 为例进行阐述。

miRNA 是一类 21~23 nt 的小 RNA，miRNA 的合成过程大致可以分成三个阶段：首先，miRNA 基因转录形成具有类似环形结构的初级产物 pri-miRNA，可长达几个 kb。哺乳动物典型的 pri-miRNA 具有 3 个特殊结构的区域：末端的环状结构、33 bp 左右的 dsRNA 主干和单链 RNA 片段区，它可以通过两条途径形成相应的 pre-miRNA。第一条途径需要一类特殊的 RNA 酶-Drosha 及其辅助蛋白（人类的是 DGCR8，果蝇的是 Pasha）。其中，DGCR8 能够识别并且结合到 pri-miRNA 的 33 bp RNA 主干和单链 RNA 片段区，协助 Drosha 切割转录产物，释放成熟的 pre-miRNA。此外，对于位于果蝇和哺乳动物基因组内含子中的有些 miRNA 基因，它们的 pre-miRNA 形成并不需要 Drosha 参与，如具有套索结构的内含子在剪接体（spliceosome）完成剪切后，经过脱支酶（debranching enzyme）的作用，最终成为具有发卡结构的 pre-miRNA。随后，pre-miRNA 被 Exportin 5 等运送到细胞质中。成熟 miRNA 的长度由 RNA 酶 III 家族的两个成员共同决定：位于核内的 Drosha 和位于细胞质中的 Dicer。Drosha 先切割出成熟 miRNA 的一端，Dicer 就像一把尺子，从已经切好的一端量出约 22 nt 进行切割，形成双链 RNA（图 3-8）。双链 RNA 被运送到以 Ago 蛋白为核心的 RNA 诱导沉默复合体（RNA-induced silencing complex, RISC）中，其中一条链作为成熟的 miRNA 留在 RISC 中，另一条链则被视为底物而降解。

有关 miRNA 调控机制的研究发现，miRNA 可以通过碱基互补配对的方式，与目标基因 mRNA 的 3'UTR 或编码区域结合，进而抑制 mRNA 的翻译或直接降解 mRNA，最终发挥抑制基因表达的作用。生物信息学预测发现，每个 miRNA 都有众多的靶基因，而每个基因的 mRNA 又有可能受到多个 miRNA 的调控，由此构成了复杂的调控网络。在哺乳动物基因组中，30% 以上基因的 mRNA 都受到 miRNA 的调节。此外，一些 miRNA 隶属于

同一个家族，它们具有相同的种子区域，通过碱基互补配对与目标 mRNA 的关键部位结合，即从 5' 端起 2~7 位具有相同的序列。如果短链非编码 RNA 的调控作用发生异常，就有可能导致机体生命活动的紊乱，甚至导致疾病的发生。

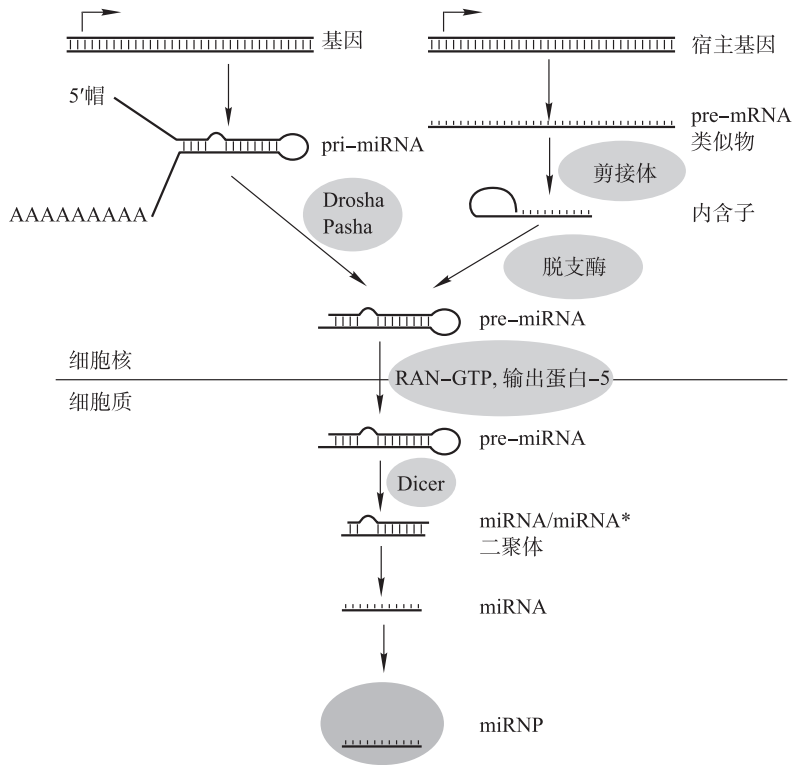


图 3-8 miRNA 的成熟加工

2. 长链非编码 RNA (long noncoding RNA, lncRNA) 指的是碱基组成大于 200 个核苷酸的非编码 RNA，占非编码 RNA 的 80% 左右，数量远超过短链非编码 RNA。长链非编码 RNA 的序列保守性较低，在不同物种间差异很大。早期认为长链非编码 RNA 只是转录过程的副产物，并不发挥任何生物功能，但是，近年来对长链非编码 RNA 的研究逐渐深入，发现其在基因表达过程中发挥着重要的调控作用。

lncRNA 因其种类、数量、功能都不明确，目前没有统一的分类标准，可以根据其生物来源、转录长度及功能等进行分类。根据其生物来源大致有 5 类：蛋白质编码基因框架结构断裂产生；染色体重组过程中，由多个未转录且彼此分开的基因序列重组产生；反转录过程中，由非编码基因复制产生；lncRNA 中相邻重复单元结构由串联复制产生；直接由转座因子插入产生。根据转录物长度可分为 5 类：lncRNA，基因间区长链非编码 RNA (long-integenic noncoding RNA, lincRNA)，位于基因间区的超长链非编码 RNA (very long-integenic noncoding RNA)，宏 RNA (macroRNA) 与启动子相关联的 lncRNA。根据其功能可分为 4 类：具有增强子作用的 lncRNA，作为 miRNA 初级转录物的 lncRNA，作为 piRNA 初级转录物的 lncRNA，竞争性内源 RNA。除此之外，根据与蛋白质编码 RNA 的相似程度可分为 mRNA 样 lncRNA (mRNA-like lncRNA, mlncRNA) 和 lincRNA。mlncRNA

多数由 RNA 聚合酶 II 合成，在其 5' 端有甲基鸟苷帽，在其 3' 端有 poly (A) 尾。除此之外，相对于蛋白质编码基因的位置，可将其分为 5 种类型，包括正义型、反义型、基因内型、基因间型和双向型 lncRNA。

lncRNA 在生物体内广泛参与 DNA 甲基化、组蛋白修饰、染色质重塑等生物学过程，能够直接与转录因子、功能性 RNA 分子、染色质重塑复合物作用，分别在转录水平、转录后水平以及表观遗传水平上对靶基因的表达进行调控。不仅如此，有些 lncRNA 还可作为某些功能性短链非编码 RNA (short non-coding RNA, sncRNA) 发挥作用。

在生物体内，lncRNA 主要通过 4 种方式行使其生物学功能：①信号分子，通过识别转录因子在信号通路中的联合作用来调控靶基因的表达；②诱饵分子，直接与转录因子、蛋白质分子等相关物质结合，阻断其对靶基因的作用，间接调控目标基因的转录；③引导分子，招募染色质修饰相关酶，并指导该蛋白复合物顺式或反式定位到调控位点；④支架分子，作为中央平台招募多种蛋白质分子并形成核糖核蛋白复合物，通过影响组蛋白修饰在表观遗传水平上对靶基因进行调控。例如由同源异型框基因 C (homeobox C, HOXC) 转录产生的同源异型框基因反义基因间 RNA (HOX antisense intergenic RNA, HOTAIR) 在多梳抑制复合物 2 (polycomb repressive complex 2, PRC2) 和赖氨酸特异性脱甲基酶 1 (lysine-specific demethylase 1, LSD1) / 转录因子沉默元件 1 辅阻遏物 (corepressor for repressor element-1 silencing transcription factor, CoREST) / 转录因子沉默元件 1 (repressor element-1 silencing transcription factor, REST) 复合体的招募过程中发挥着支架分子作用，而在指导 PRC2 靶向作用于 HOXD 相应基因位点时又充当引导分子的角色。长链非编码 RNA 在生物体的生长、发育、衰老、死亡等过程中都发挥着重要的调控作用，第四章我们将阐述它在表观遗传调控中的重要作用。

(孔璐)

网上学习



学习目标



重点内容简介



参考文献



思考题