

## 原核生物基因的表达

郭兴华

(中国科学院微生物研究所, 北京 100080)

生物在代谢活动和生长发育过程中, 将蕴藏在 DNA 中的遗传信息传递到 RNA 分子上, 然后再转移到蛋白质分子, 这就是基因表达。本文只就原核生物基因的表达加以讨论。

## 表达调控上的差别

原核基因组与真核基因组不论结构组织, 还是转录调控以及翻译后加工等都有很大的差别, 可以综合如表 1。

## (一) 原核基因组与真核基因组在结构和

## (二) 内原基因的表达与调控

表 1 原核基因组和真核基因组的差别<sup>[1-7]</sup>

基因组	原核生物	真核生物
大小及复制起点	小, 一般只有一个 DNA 复制起点	大, 具有许多 DNA 复制起点
染色体数目及形状	染色单体, 一般呈环状 DNA	一个基因组有若干个染色体, 一般不呈环状
DNA 和蛋白的关系	染色体 DNA 不和蛋白固定结合	染色体 DNA 和蛋白固定地结合
重复序列的多少	少	有大量重复序列
不编码序列	少	多
相关基因的集中程度	功能上密切相关的基因构成操纵子, 并常转录成多基因 mRNA	功能上相关的基因集中程度不如原核生物, 很少有关操纵子的报道
依赖于 DNA 的 RNA 多聚酶	一个细胞一种酶	一个细胞有多种酶, 作用各异
启动子区	-10 区: 5'-TATAAT-3' -35 区: 5'-TTGACA-3'	在 100bp 内有 TATA box 和 CCAAT 或 GGGCGG 顺序
调控转录的元素	除启动子无其它元件, 偶尔有原核增加子样序列	增强子, 上游激活序列和转录活性因子
转录后加工	转录形成成熟的 RNA	转录成前 RNA, 剪接后为成熟的 RNA
翻译起始信号	SD box	比较复杂
翻译后加工	不能糖基化, 切掉信号肽	糖基化, 切掉信号肽
转录终止信号	终止子 { Rho <sup>+</sup> 衰减子 Rho <sup>-</sup>	3'-端有 5'-AAUAAA-3' 可能是 polyA 的识别位点

1. 转录水平上的调控<sup>[1,7-9]</sup>: 遗传信息由 DNA 转录到 mRNA 是由 RNA 聚合酶进行的。原核生物的 RNA 聚合酶全酶 (holoenzyme), 主要是由 5 个亚基 ( $\alpha, \beta, \beta', \sigma$ ) 组成。核

心酶只能使已开始合成的 RNA 链延长, 但不具识别特异转录起点的作用。 $\sigma$  因子能识别 -35 区特异的 DNA 序列,  $\beta'$  因子能和 DNA 模板结合,  $\beta$  因子具有起始和催化的作用。

在原核生物中, 转录启动分三步进行(图1)。

(1) 识别: RNA 聚合酶与启动子上的  $R_{\sigma}$  和  $R_c$  部位作顺序专一性的接触, 形成复合物;

(2) 解链: 在起始转录位点的附近发生解链, 两个 DNA 单链局部分开;

(3) 起始: RNA 聚合酶使第一和第二个核苷酸形成磷酸二酯键, 接着 RNA 聚合酶向前推进, 形成新生的 RNA 链。

在大肠杆菌中, 至少有两种已知的  $\sigma$  因子。 $\sigma^{70}$  是第一个被发现的  $\sigma$  因子, 它识别大肠杆菌启动子的范围较广泛;  $\sigma^{32}$  识别的范围较窄, 它涉及对大肠杆菌热休克反应。

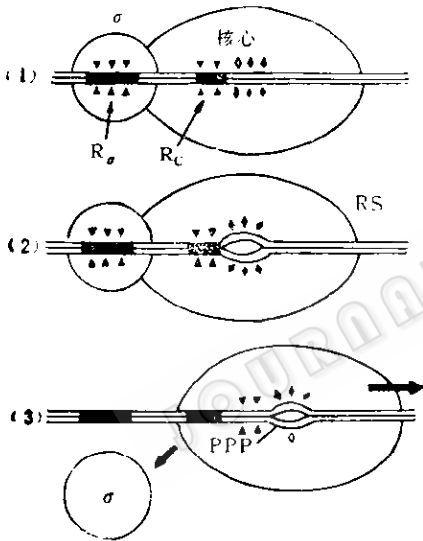


图1 转录启动模型 (Pribnow 1979)

枯草杆菌有一个较复杂的转录系统, 至少有 5 个以上的  $\sigma$  因子参与, 它们和核心酶结合形成了不同类型的全酶。由于每个  $\sigma$  因子识别一个特定的保守区, 所以不同的全酶在细胞生长的不同阶段或病毒感染过程中起着不同的作用。

在枯草杆菌中除了 RNA 聚合酶与大肠杆菌有所不同外, 很多基因是受两个或更多的启动子控制的, 这些启动子使得基因在营养生长和孢子形成的不同时期的基因得以正常表达。这些启动子可能是以重叠的形式存在, 或以串联的形式存在, 这些启动子提供了由生长变成

稳定期启动子开关的机制。这种机制特性, 暗示了在枯草杆菌中控制基因表达有多种多样的途径。

原核生物转录的衰减作用和阻抑作用: 衰减作用是细菌辅助阻抑作用的一种精细调控。这种作用多存在于合成代谢(尤其是氨基酸的合成代谢)的操纵子调控系统中。阻抑作用虽然也有终止转录的作用, 但和衰减作用的机制完全不同。在阻抑作用中, 阻抑物与操纵基因结合使操纵子转录不能开始。衰减作用是使操纵子的转录开始后还没进入第一个结构基因时便终止, 也就是说, 对于已经开始的转录, 则只能通过衰减作用使它中途停顿下来。不过这种终止作用并不使所有正在转录中的 mRNA 全部中途终止, 而是仅有部分中途停止转录, 在细菌细胞内这两种作用相辅相成, 体现着生物体内的精细的调控作用。

原核生物转录的终止作用。终止子是转录单位的终止区, 如果没有终止子, 将会发生通读(Read-Through), 转录就无法控制, 基因的产物就变得乱七八糟, 所以终止子是转录单位的最后一个调控元件, 它有特定结构, 这种结构不因环境条件的变化而改变。

2. 翻译水平上的调控<sup>[9]</sup>: 在比较和分析了大量数据之后, Shine 和 Dalgarno 得出, S/D 序列是 3—9 个碱基的长度, 它位于起始密码子 AUG 或 GUG 上游 3—11 个碱基的位置上。S/D 序列和起始密码子一起形成核糖体结合位点(Ribosomal Biding Sites, 简称 RBS)。S/D 序列和起始密码之间的距离及 mRNA 的二级结构, 在确定翻译起始效率中起着重要的作用。

在革兰氏阳性菌中, 已经证明起始密码附近的 mRNA 和核糖体 16SRNA 3'-末端有高度的互补性, 保守程度比大肠杆菌更强。S/D 序列和 16S 核糖体 RNA 之间相互反应的自由能值提高到 -50—-80KJ/mol 之间, 而相对应的大肠杆菌的能值是在 -17—-92KJ/mol 的大范围内, 这可说明, 在体外翻译系统中, 枯草杆菌能识别来自革兰氏阳性菌的 mRNA, 而不识别来自革兰氏阴性菌的 mRNA。

### (三) 外源基因的表达与调控<sup>[10-13]</sup>

外源基因在受体中的表达,是指外源基因能在受体中合成功能性的蛋白质。功能蛋白质的合成,决定于转录、翻译、翻译后加工和新生肽链的折叠,任何一个步骤出现问题,即导致表达的失败。外源基因需要有宿主细胞内 RNA 聚合酶的识别部位;有效的翻译需要 mRNA 上具有核糖体的结合位点;蛋白质的大量分泌要有信号序列和翻译后加工修饰以及信号肽的切除。这些都是外源基因表达所要涉及到的问题。

一般说,真核基因在原核细胞中的表达较难,而原核基因在原核细胞中的表达较易,然而同样是原核细胞,革兰氏阳性菌和革兰氏阴性菌也有差异。如革兰氏阳性菌的基因可在革兰氏阴性菌的大肠杆菌中表达,这是假定遗传元件是通用的,即大肠杆菌的转录体系和翻译体系对由其它细菌来的外源基因的识别,相对讲是不加选择的。但对革兰氏阳性的枯草杆菌来说则不然,它对外来的基因是严加选择的。因此,大肠杆菌的基因在枯草杆菌中一般不能表达。

1. 真核生物基因在原核生物中表达的困难与解决办法。

(1) 带有内含子的真核基因不能在原核细胞中表达,因为原核细胞缺乏真核细胞转录后加工体系, mRNA 中的内含子不能被切除,不能形成成熟的 mRNA,所以真核细胞的基因不能表达。克服的办法是,人工合成 DNA 或从真核细胞分离 mRNA 并反转录成 cDNA, cDNA 有完整的编码序列但无内含子,这样真核的基因就可在原核生物中表达。

(2) 真核基因表达调控的启动子不被原核 RNA 聚合酶识别,真核基因转录的 mRNA 缺乏 S/D 序列,不能与原核的核糖体结合。克服的办法是,将真核基因克隆到原核启动子和 S/D 序列的下游,使真核基因处于原核启动子控制之下,转录出的 mRNA 有原核的 S/D 序列,这样便可在原核细胞中正常表达。

(3) 真核基因的产物——蛋白质,在原核

细胞中不稳定,容易被细菌蛋白酶破坏。克服的办法是,在克隆时,将真核基因插在几个原核密码子之后,通过翻译,得到原核多肽和真核基因产物在一起的蛋白,细菌蛋白酶不易降解这种蛋白,经过分离纯化,切去原核多肽,即可得到真核基因的产物。

2. 不同类型基因产物的表达:由于分子克隆的载体是多种多样的,所以当外源基因插入不同载体的不同位置上时,可产生不同的产物,因此可将基因产物分为融合蛋白、天然蛋白和分泌蛋白等,现分述如下。

(1) 融合蛋白的表达:外源基因通常以融合蛋白的形式而表达。往往是把外源基因融合到  $\beta$ -内酰胺酶或  $\beta$ -半乳糖苷酶等肽链上,融合蛋白的 N 端由原核 DNA 序列编码, C 端由真核 DNA 的完整序列编码。这样的蛋白是由一条短的原核肽链和真核蛋白结合在一起的,故称融合蛋白。融合蛋白有一定的好处:它可以大量生产,因为转录和翻译是由正常的细菌遗传元件指导开始的;它比天然蛋白更稳定,因为细菌蛋白酶不易破坏它;它比细菌蛋白质更大,所以更容易在蛋白质电泳胶上鉴定,把它从胶上切下后冷冻磨碎,可用于抗原试验。为了得到正确的真核蛋白,在插入真核基因 DNA 时,其阅读框架要与原核 DNA 片段的密码阅读框架一致,翻译时才不致产生码组移动。

(2) 天然蛋白的表达:要得到天然完整的蛋白,关键是要把外源基因放在载体的适当位置上。载体 pWT571 由 Trp 启动子和其 S/D 序列组建而成,它有一个 EcoRI 起点,这个切点正好包括了起始密码子 (ATGAATTC),克隆的基因在正确的阅读框内将能表达 N-末端的 fmet-Asn 蛋白序列。假若用 EcoRI 切开,接着用 S1 酶除去 5'-AATT,变成平末端,把它连接于编码蛋白缺乏起始密码 N-乙酰甲硫氨酸的片段上,也可得到天然蛋白。

如果把带有起始密码 ATG 的真核基因插入到原核启动子和 S/D 序列的下游,组成一个杂合的核糖体结合区,经转录和翻译,也可得到天然蛋白。

(3) 分泌蛋白的表达<sup>[12,14,15]</sup>: 基因产物要分泌 (Secretion) 到细胞周质空间或外泌 (Excretion) 到细胞外培养基内, 必要的条件是把外源基因融合到分泌载体的信号序列 (Signal sequence) 上, 外源基因经过转录和翻译, 借助信号肽 (Signal peptide) 把蛋白前体进行穿膜转位于周质空间或细胞膜外, 再由信号肽酶切掉信号肽, 变成成熟蛋白。

3. 使外源基因高效表达的措施: 外源基因能否在宿主中高效表达, 涉及因素很多, 诸如:

(1) 选用强启动, 可调控的高表达载体: 基因表达的第一步是转录, 转录产物——mRNA 的多少, 决定于启动子的强弱, 所以选用强启动的高表达载体是外源基因表达的关键。就目前所知, 在大肠杆菌中使用的强启动子有: LacUV5、Trp、Tac、 $\lambda P_L$  和  $T_{10}$  启动子等, 现就 Tac 启动子叙述如下: Tac 是由色氨酸操纵子的启动子 Trp 和乳糖操纵子的启动子 Lac 拼接而成的一个嵌合启动子。Tac 启动子的转录效率比 LacUV5 大 11 倍, 比 Trp 大 3 倍。原核生物中 -10 处的共有序列为 5'TA-TAATG3', -35 处的共有序列为 5'TTGACA3'。取 Trp 中 -35 的共有序列和 Lac -10 处的共有序列, 构建了 Tac 启动子, 能被 RNA 多聚酶正确识别, 因此 Tac 比两者亲本转录效率都高。利用 Tac 启动子已构建了很多高表达载体, 如 pDR540 和 pKK 系列。Tac 启动子受 Lac 阻抑物调控, 因此在 Lac 受体中, 如 JM103、JM105 等菌株中, Tac 启动子被阻抑, 然而在培养菌的适当时间里, 补加异丙基- $\beta$ -D-硫代半乳糖苷 (IPTG) 时, 这种阻抑是可以被解除的。

(2) 调整 (S/D) 与起始密码之间的距离: 要使外源基因在细菌中高效表达, 除了启动子, 第二个重要的因素就是有效的 RBS, 也就是要调整原核 S/D 序列和外源基因 ATG 之间的距离。距离过长或过短都影响外源基因的表达。有时仅仅 2—3 对碱基的差别, 蛋白质的合成率相差 20 倍。调整距离的方法之一是: 当外源真核基因与载体连接后, 在距真核基因的 ATG 上游约 100bp 内, 要有一个限制性位点。用特

异的限制酶消化和外核酸酶修饰到不同长度, 然后插入带有 S/D 序列的原核携带式启动子, 便得到一套重组 DNA。它们每一个在原核 S/D 序列和真核 ATG 之间都有不同长度的距离, 选择那些距离适宜的阳性重组体, 即能高水平表达天然蛋白。由于这段距离是在 ATG 上游, 不参与表达, 不像密码阅读框架那样严格。

(3) 通过温度敏感突变体增加外源基因的拷贝数: 在细菌生长繁殖的过程中, 质粒拷贝数过多对宿主并没有好处。当宿主细胞迅速生长时, 抑制重组质粒的复制, 当细胞生物量积累到一定水平后, 再使细胞中质粒 DNA 复制, 增加质粒的拷贝数。拷贝数增加必将伴随外源基因表达水平的提高。质粒 pCZ101 是温度调控 DNA 复制的例子。用牛生长激素基因与 pCZ101 克隆, 转化细胞在 25°C 生长时, 质粒 DNA 仅有 10 个拷贝, 宿主细胞大量生长。转化细胞在 37°C 生长时, 重组质粒大量繁殖, 每个细胞的拷贝数增加到 1000 个, 外源基因得以高水平表达。

(4) 阻抑和解阻抑效应的利用:  $P_L$  启动子是被 cI 产物阻抑所调控的, 当它用于含温度敏感的 cI 阻抑物的细菌如  $\lambda$ cI857 时, 就可创造一个热诱导系统。在温度 (29—31°C) 时, cI 基因产物有活性, 阻抑效应发挥作用, 抑制了  $\lambda P_L$  启动子下游的基因产物的合成, 此时细菌大量生长繁殖。当温度升高到 42°C 时, cI 基因产物失活, 解阻抑效应发挥作用,  $P_L$  启动子下游的外源基因产物大量产生。利用这个原理, 曾使得  $\lambda$  噬菌体 cII 蛋白高达细胞总蛋白的 4%。

(5) 提高质粒的稳定性: 基因工程遇到的难题之一是质粒的不稳定性, 质粒若不稳定, 就很难使外源基因高效表达。质粒的不稳定性主要有两种类型: 一种是分离的不稳定性 (Segregational instability), 它使整个质粒由宿主中丢失, 这是由于细胞在分裂过程中不准确的分配造成的。另一种是结构的不稳定性 (Structural instability), 它造成缺失、重排或插入 DNA 新片段。重排可在分子内或分子间进行,

是由于宿主有很强的重组能力造成的。质粒的稳定性与质粒的本身、宿主和培养条件等有关,一种质粒在不同的宿主中稳定度不一样;反过来,一种宿主对不同质粒稳定性的保持程度也不一样。可以通过自然选择、人工诱变或质粒重新构建和宿主的改造等方法来提高质粒的稳定性。

(6) 使用蛋白酶缺陷型的宿主: 宿主的蛋白酶很容易降解外源蛋白, 尤其是天然的外源蛋白, 为要得到更多的基因产物, 就要对宿主进行改造, 使之成为蛋白酶缺陷型, 而细菌细胞中的蛋白酶又不只一种, 因此要对它改造成多种蛋白酶缺陷型的菌株, 这当然不是一件容易的事, 尽管如此, 目前有不少实验室已获得各种各样的蛋白酶缺陷株。实验证明, 外源基因的产物在蛋白酶缺陷型的宿主中, 要比在蛋白酶野生型的宿主中多的多。

### 参 考 文 献

4. 盛祖嘉、沈仁权: 分子遗传学, 5—90, 复旦大学出版社,

- 1988.
2. Serfling E et al.: *Trends Genet.* 1: 224, 1985.
  3. Vogt T F et al.: *Science*, 236: 301, 1987.
  4. Wight P A et al.: *J. Biol. Chem.* 262: 5959, 1987.
  5. Van der Veen R et al.: *EMBO J.*, 6: 1079, 1987.
  6. Soinick D: *Cell*, 24: 157, 1985.
  7. James A et al.: *Trends in Biochemical Science*, 16(11): 394, 1991.
  8. Ernst-L, Winnacker: From gene to clone, Introduction to gene technology, VCH, 210—223, 1987.
  9. Sydney Kustu et al.: *Trends in Biochemical Science*, 16(11): 397, 1991.
  10. 齐义鹏, 黄永秀: 基因工程原理和方法, 212, 四川大学出版社, 1988.
  11. Ernst-L Winnacker: From gene to clone. Introduction to gene technology, VCH, 224—313, 1987.
  12. 郭兴华、贾士芳、门大鹏: 微生物学通报, 12(5): 215, 1985.
  13. Old R W et al.: Principles of gene Manipulation, Third edition, 1985.
  14. Doi R H: Tibeck-September, 232, 1986.
  15. Thomas Eckhardt et al.: *J. Bacteriology*, 169(9): 4249, 1987.
  16. 郭兴华、贾士芳、许怡: 待发表。