

影响大肠杆菌中外源基因表达的因素

隋广超 胡美浩

(北京大学生物系, 北京 100871)

摘要 大肠杆菌已经被广泛地应用于表达各种外源基因, 但是, 不同的外源基因在表达效率上却有很大的差异. 文章综述了影响大肠杆菌中外源基因表达的因素, 这将有助于认识大肠杆菌中外源基因表达的规律, 以便采取有效的方法提高外源基因在大肠杆菌中的表达效率.

关键词 大肠杆菌, 外源基因, 载体, 基因表达

随着生物技术的发展, 基因工程已经开始与医药生产, 疾病的治疗有了紧密的结合. 利用大肠杆菌 (*E. coli*) 来合成某些生物分子, 具有低成本, 高效率, 短周期的特点, 已经越来越成为人们研究的热门领域. 但是, 在利用 *E. coli* 表达外源基因的过程中, 往往会遇到许多困难, 使表达效率难以得到最大限度的提高, 因而影响了产率.

影响外源基因在 *E. coli* 中表达效率的因素包括: 载体(启动子)和宿主菌的选择, 培养条件的控制, cDNA 密码子的选用, mRNA 的一、二级结构和表达产物的后加工处理, 等等. 本文就以上这些影响表达效率的因素以及提高外源基因在 *E. coli* 中表达效率的方法加以综述.

1 表达载体的选择

用来在 *E. coli* 中表达外源基因的载体有很多种. 对于载体的基本要求是: 重组质粒有较高的拷贝数, 并可在菌体内稳定存在.

构成一个表达载体的组成部分是^[1]: a. 强的启动子(在-35至-10区); b. 启动子下游的 polylinker 区; c. polylinker 区的后面, 有转录终止序列; d. 抗性筛选标记. 对于某些特殊的载体, 内部还要带有其它一些基因. 一般来说, 表达载体不宜过大, 以 2.5—5.0kb 为佳. 太大的载体, 既不利于自身复制, 也会在

表达较小的外源基因时, 难于进行重组体的筛选.

启动子是影响外源基因表达效率的关键因素. 由于外源基因的表达往往会影响细菌正常的生长和分裂, 某些外源蛋白甚至还对菌体有毒, 目前选用的启动子多为可控制表达 (controllable expression)^[2], 就是在菌体生长到一定时期进行诱导, 而诱导前的基础表达水平很低或没有. 常用的启动子有: P_L , P_R , P_{trp} , P_{lac} , $P_{\phi 10}$, P_{trc} 等等, 可以通过升高温度或加入异丙基- β -D-硫代半乳糖苷 (IPTG) 进行诱导的.

有些外源蛋白对启动子有一定的选择性. Surek 等人^[3]发现, 表达尿激酶原 (pro-UK) 时, 利用 P_{lac} 最好, P_{trp} 次之, P_L 最差.

增加启动子的数目, 可以提高表达效率. Shibui 等人^[4]就利用两个串联的 P_{lac} 启动子, 使人的 cardiodilatin 蛋白得到了高效表达.

对于某些小分子量的蛋白质, 有时要选用形成融合蛋白的表达载体. Gan 等人^[5]在表达蛇毒中 echistatin 蛋白 (49 个氨基酸) 时, 发现它不能单独表达. 利用融合蛋白表达, 可达菌体总蛋白量的 20% 以上. 由于 echistatin 是通过蛋氨酸与前面的蛋白相连, 必须将基因内部的 Met-28 突变为 Leu-28, 结果并未影响产物

的活性。

转录终止子 (terminator) 位于基因的下游附近, 使得被转录出的 mRNA 尽可能的短, 以减少不必要的能量消耗。

2 外源基因中密码子的使用

带有相应反密码子 (anti-codon) 的 tRNA 将氨基酸引导到 mRNA 上, 进行蛋白质的翻译合成。然而, 在不同种类的生物中, 各种 tRNA 的含量是有很大的区别的。哺乳动物的细胞中, 这种差别还不太明显。但在原核生物中, 却由于不同 tRNA 含量上的差异很大, 产生了对密码子的偏爱性。对应的 tRNA 丰富或稀少的密码子, 分别称为偏爱密码子 (biased codons) 或稀有密码子 (rare codons)^[6]。

通过统计在 *E. coli* 中含量丰富的一些蛋白质中密码子的使用情况, 发现 AGA, AGG (编码精氨酸) 等 8 种密码子是 *E. coli* 中的稀有密码子^[7]。

一般来说, 含有较高比例的稀有密码子的外源基因, 其表达效率往往不高。在这种情况下, 要达到高效表达, 必须采取一定的措施。

Makoff 等人^[8], 在 *E. coli* 中表达破伤风芽孢杆菌毒素的 C 片段时, 发现该基因中含有许多稀有密码子: AUA, AGA, AGG 等, 表达量仅占细胞总蛋白量的 3%—4%。通过碱基突变, 把几乎所有的稀有密码子改为 *E. coli* 偏爱的密码子后, 表达效率提高到占细胞总蛋白量的 11%—14%。

Brinkmann 等人^[9]对精氨酸的稀有密码子 (AGA, AGG) 影响外源基因的表达情况进行了深入研究。利用一种编码精氨酸的 tRNA_{AGA/AGG} 的基因——*dnaY* 基因, 与重组质粒进行共转染, 来提供这一稀有 tRNA, 从而使几种富含这些稀有密码子的外源基因在 *E. coli* 中的表达效率, 由原来的占细胞总蛋白 5% 以下, 提高到 30% 以上。他们还发现, 含有这些稀有密码子的外源基因, 在 *dnaY* 基因不存在时诱导, 宿主菌的生长受到抑制, 菌体内质粒的拷贝数也会降低。

Spanjaard 等人^[10]还发现, 若外源基因中出现 AGA-AGA 或 AGG-AGG 的连续稀有密码子, 翻译进行到这里时, 会由于相应 tRNA 的缺乏, 使大约 50% 的核糖体在 mRNA 上向上游移动一位, 这往往造成翻译过程的提前终止, 形成了副产物。将 *argU* (即 *dnaY*) 基因或 T4 tRNA_{Arg} 基因共转化到宿主细菌中, 可以抑制这种读码的移动。

因此, 密码子偏爱性对外源基因表达效率的影响, 主要是通过翻译过程中 tRNA 的浓度对蛋白质合成速度的限制来进行的。在肽链的延长过程中, tRNA 携带相应的氨基酸与核糖体上的 A 位相作用。如果上到 A 位的 tRNA 与该处的密码子相对应, 就可以进行肽键的合成及肽链的延伸。如果不是特异的 tRNA, 就会掉下, 再由其它 tRNA 去尝试。所以, 对那些偏爱密码子, 正确的氨基酸会很快被连接上; 但对稀有密码子, 要经过多次的这种互相辨认才能找到正确的 tRNA^[7]。这样, 外源蛋白的合成会在此发生停顿。在含有较多的稀有密码子, 甚至相同的稀有密码子连续出现时, 会产生明显停顿, 抑制蛋白质合成。有时, 这些地方会发生密码子的错配, 而某些密码子又有较明显的错配倾向^[11,12]。这些状况的存在, 会严重地干扰菌体的正常代谢, 影响细菌的生长和分裂。

在利用 *E. coli* 表达外源基因时, 小分子量蛋白质, 可选用 *E. coli* 偏爱的密码子来合成其 cDNA; 对高分子量的蛋白, 在克隆到 cDNA 后进行表达时, 要针对密码子的偏爱性采取措施, 如, 提高某种氨基酸的 tRNA 浓度, 进行适当的碱基突变, 等等。

3 mRNA 的一级和二级结构的作用

外源基因转录出的 mRNA 的一级和二级结构与外源蛋白的合成有很大关系, 也是影响表达效率的重要环节之一。目前, 虽然这方面的研究并不像转录水平的调控进行得深入, 但已有证据表明, mRNA 调节外源蛋白的合成主要是在翻译的起始过程, 而且, 起始密码子

AUG 附近的核苷酸序列 (长度约为 40 个核苷酸) 是很重要的. 它是核糖体的结合位点 (ribosome binding site, 即 RBS). RBS 序列的一级和二级结构, 决定了核糖体与 mRNA 的结合, 从而影响翻译的起始及外源蛋白的合成效率^[1].

3.1 mRNA 的一级结构

在 AUG 上游的 3—11 个核苷酸处, 有一段 SD 序列 (Shine-Dalgarno Sequence). 长度一般为 5 个核苷酸, 富含 G, A, 例如: GGAGG. 功能是与核糖体 16S rRNA 的 3' 端互补配对, 使核糖体结合到 mRNA 上, 以利于翻译的起始. 对于不同的启动子, 它们的 RBS 序列可以不同, 但是 SD 序列的构成大致相同^[1]. 外源基因的起始密码子与 SD 序列之间有合适的距离, 一般, 以 4—10 核苷酸为佳.

Makoff 等人^[13], 使用双顺反子的 RBS 序列. 即, 在外源基因及其相对较弱的 RBS 序列的上游附近, 插入一个很强的 RBS 序列, 这样, 上游的翻译起始可以促进下游的 RBS 的功能, 使外源基因的合成效率得到提高.

某些高分子量的外源蛋白, 尤其是含有较多蛋氨酸的情况下, 其 mRNA 序列中有多个 AUG. 这样, 如果某一个或几个 AUG 的上游有与 SD 序列相似的序列, 会发生翻译起始的竞争, 合成副产物, 影响表达效率. Sung 等人^[14]在表达甲状旁腺激素 (PTH) 时发现, 起始竞争的副产物竟占总表达产物的 30% 以上.

外源基因的高效表达还要求 mRNA 5' 端的序列与 *E. coli* 中的基因有一定的共同性. 否则, 即使带有偏爱密码子的基因表达效果也可能不好^[6].

3.2 mRNA 的二级结构

核糖体的 30S 亚基与 mRNA 的靠近, 要求 mRNA 的 5' 端有一定的空间结构. 决定这一空间结构的核苷酸不仅是载体上 SD 序列附近的区域, 还包括起始密码子下游的序列. 把起始密码子 AUG 中的 A 作为 +1 时, mRNA 与核糖体结合的部分是 -20 至 +14 的区域^[15]. 这些核苷酸的微小变化, 往往会影响

mRNA 的二级结构, 可导致表达效率上 2000 倍以上的巨大差异^[1]. Hall 等人^[16], 把 SD 序列上游附近的一个碱基突变后, 外源基因的表达效率降为原来的 4%. Wood 等人^[17], 在表达免疫球蛋白 μ 重链时, 突变了 AUG 下游附近的三个核苷酸, 提高表达效率近 100 倍.

有时, 改变距离较远的核苷酸也有很大作用. 如, Gross 等人^[15]发现, 在 IFN- β mRNA 的 AUG 下游, 第 11 个密码子中一个核苷酸的突变, 可以使表达效率有 28 倍的差异.

这些现象的产生, 是由于碱基突变, 改变了形成 mRNA 5' 端二级结构的自由能, 影响了核糖体 30S 亚基与 mRNA 的结合, 因而造成了蛋白合成效率上的差异.

因此, 要提高表达效率, 需对 AUG 附近的核苷酸进行分析, 得出可能的空间结构. 一般认为, 在 mRNA 起始区域无发夹结构或即使有类似的结构, 而 SD 序列和起始密码子 AUG 又不被包裹在这一结构内部时, 有利于翻译的起始^[15-17].

4 mRNA 稳定性的影响

对应外源蛋白的 mRNA 的稳定性, 也是影响表达效率的一个很重要的因素.

Gross 等人^[15]研究发现 IFN- β 的表达量与 mRNA 的半寿期成正比例关系. 最长的半寿期为 4.5min. Wood 等人^[18]在表达 β -半乳糖苷酶时发现, 诱导表达后, mRNA 的合成速度提高了 42 倍, 稳定存在的 mRNA 量仅提高了 4 倍. 可见, 大部分 mRNA 被迅速降解.

mRNA 的降解方式有外切和内切, 可在 5' 端和 3' 端. 某些 mRNA 的 SD 序列上游 20 多个核苷酸, 有一段富含 U 的区域, 易受核酸内切酶的作用, 但它对翻译起始是必不可少的. 蛋白合成时, 核糖体及起始因子的结合, 对这一区域起到保护作用, 避免被核酸内切酶降解. 而且, 核糖体结合与内切酶作用有相关性. 有充足的核糖体时, 内切酶的作用机率会减少^[15].

mRNA 的 3' 端结构也影响 mRNA 的稳定

性. 在 *E. coli* 中, 约有 1000 个拷贝的 REP (repetitive extragenic palindromic) 序列存在于染色体上, 它在 mRNA 的 3' 端出现时, 可以避免 3' 至 5' 核酸外切酶的作用^[19].

在 mRNA 有适宜的结构并含有 *E. coli* 偏爱的密码子时, 会减少核糖体在 mRNA 上的滞留, 在提高翻译效率的同时, 也加强了对 mRNA 的保护.

5 宿主的选择及培养条件的控制

宿主菌的生长情况与培养条件是密切相关的. 二者对外源基因的表达也有一定影响.

5.1 宿主的选择

在外源蛋白合成时, 会引起菌体内的一些变化. 不同的菌种, 这些变化又不尽相同.

细菌含高拷贝数的质粒时, 会出现蛋白合成因子减少, 热激蛋白 (heat shock protein) 成倍提高等现象^[20]. 这些不利于外源蛋白的合成和稳定存在.

Surek 等人^[3]在表达 pro-UK 时发现, 在 htpR 基因 (编码一种热激蛋白) 缺失的细菌中表达时, 产物能够稳定存在; 但在其它几种菌株中, 随诱导时间延长, 产物会被降解. 原因是, 诱导后产生的热激蛋白, 会激发蛋白水解酶, 降解表达产物. Sung 等人^[14]也发现, 利用缺失 lon 基因 (编码一种蛋白酶) 的菌株表达 PTH, 约为其它菌株内表达量的三倍. 可见, 选用与蛋白酶相关的基因缺失的菌株作为宿主, 可减少产物降解, 提高产率.

某些特殊的启动子, 需要特殊的宿主菌. 如, pET3a 载体^[21]是利用 $\phi 10$ 启动子, 需 T7 RNA 聚合酶特异地合成 mRNA, 这就要用带有溶源的 T7 RNA 聚合酶基因的细菌 [如, BL21 (DE3)] 作宿主. 诱导表达时, 先诱导合成 T7 RNA 聚合酶, 再表达外源基因. 表达量可占细菌总蛋白的 50% 以上. 还有几种 BL21 (DE3) 的衍生菌株, 适于表达对细菌有不同程度毒性的外源蛋白.

在工业化生产中, 为降低成本, 需少加或不加抗菌素. 这会使细菌在分裂时, 发生质粒

丢失现象. Hiraga 等人^[22], 利用一种带有 ccd 基因 (与细菌的分裂有关) 的菌株及相应的表达载体, 使得不带有表达质粒的细菌停止分裂而死亡, 带有质粒的细菌则可以存活.

5.2 培养条件的控制

培养条件包括培养液成分, 温度的选择, 诱导及培养时间的长短等等.

一般来说, 使用营养成分丰富的培养液, 易于细菌的生长和表达.

Jelenc 等人^[12]发现, 体外翻译系统中, 理想的条件是含有适量的 Mg^{2+} , K^+ , NH_4^+ , 较高浓度的 Ca^{2+} 和多胺类, 以及充足的 ATP 和 GTP. 在这种条件下, 可以显著降低蛋白质合成中的错读.

温度的影响有时也很重要. 例如, pro-UK 不适于用温度诱导表达, 而且, 30°C 时表达产物稳定性要明显高于在 37°C 时的表达^[3].

有时, 培养条件的改变, 会影响外源蛋白的表达方式. Blackwell 等人^[23]在 *E. coli* 中表达一种转移酶时, 培养液中加入甜菜碱和山梨醇来改变渗透压, 使表达产物由包含体 (inclusion body) 形式转变为活性状态. 在 25°C 诱导, 这一现象更为明显.

一般来说, 在 25°C 至 37°C 诱导, 外源蛋白易于以活性状态存在; 在 37°C 以上诱导, 则会形成包含体^[23].

要选择适当的诱导时间来诱导表达. 细菌在对数生长期, 即 $A_{600nm} = 0.6$ 附近时, 生长状态良好, 易于诱导而合成外源蛋白. 提前诱导菌体太少, 外源蛋白产量低; 诱导太迟, 细菌过老, 也不利于基因表达.

诱导后的培养时间, 随不同的表达载体或启动子而不同. 对启动子 P_{tac} , $P_{\phi 10}$, P_{lac} , 一般仅需 2—3h, P_R , P_L , P_{trp} , P_{trc} 需时间较长.

因此, 必须结合外源蛋白的性质及表达载体的特点, 选择出恰当的表达条件.

另外, 表达产物的后加工虽不涉及到外源基因的表达, 但是, 采取适当的方法从包含体得到活性蛋白, 可以提高产物的回收率^[24].

总之, 研究提高外源基因在 *E. coli* 中表达

效率的工作已进行的十分广泛, 被发现的影响表达效率的因素也有很多, 它们之间有很大的相关性. 单单改进其中一个方面往往不能取得明显的效果, 有时还存在一定的机遇性. 但是, 认识到各因素的产生和作用的方式, 就能够提出改进表达效率的方法. 这方面工作的深入研究, 有很深远的理论意义和实际应用价值.

参 考 文 献

- 1 Walker J M, Gingold E B. Molecular biology biotechnology, 2nd edition, London: Burlington House, The Royal Society of Chemistry, 1988; 53-66
- 2 Kobayashi M, Kurusu Y, Yukawa H. Appl Biochem Biotechnol, 1991; 27: 145
- 3 Surek B, Wilhelm M, Hillen W. Appl Microbiol Biotechnol, 1991; 34: 488
- 4 Shibui T, Uchida M, Nagahari K *et al.* Agric Biol Chem, 1988; 52: 1145
- 5 Gan Z R, Condra J H, Gould R J *et al.* Gene, 1989; 79: 159
- 6 Kurland C G. FEBS Lett, 1991; 285 (2): 165
- 7 Grosjean H, Fiers W. Gene, 1982; 18: 199
- 8 Makoff A J, Oxer M D, Romanos M A. Nucleic Acids Res, 1989; 17 (24): 10191
- 9 Brinkmann U, Mattes R E, Buckel P. Gene, 1989; 85: 109
- 10 Spanjaard R A, Chen K, Walker J R *et al.* Nucleic Acids Res, 1990; 18 (17): 5031
- 11 Scorer C A, Carrier M J, Rosenberger R F. Nucleic Acids Res, 1991; 19 (13): 3511
- 12 Jelenc P C, Kurland C G. Proc Natl Acad Sci USA, 1979; 76 (7): 3174
- 13 Makoff A J, Smallwood A E. Nucleic Acids Res, 1990; 18 (7): 1711
- 14 Sung W L, Zahab D M, Barbier J R *et al.* J Biol Chem, 1991; 266 (5): 2831
- 15 Gross G, Mielke C, Hollatz I *et al.* J Biol Chem, 1990; 265 (29): 17627
- 16 Hall M N, Gabay J, Debarbouille M *et al.* Nature, 1982; 295 (18): 616
- 17 Wood C R, Boss M A, Patel T P *et al.* Nucleic Acids Res, 1984; 12 (9): 3937
- 18 Wood T K, Peretti S W. Biotechnol Bioeng, 1991; 38: 397
- 19 Higgins C F, McLaren R S, Newbury S F. Gene, 1988; 72: 3
- 20 Brinbaum S, Bailey J E. Biotechnol Bioeng, 1991; 37: 736
- 21 Studier F W. J Mol Biol, 1991; 219 (1): 37
- 22 Hiraga S, Jaffe A, Ogura T *et al.* J Bacteriol, 1986; 166: 100
- 23 Blackwell J R, Horgan R. FEBS Lett, 1991; 295 (1-3): 10
- 24 Rudolph R. In: Tschesche H ed. Modern methods in protein and nucleic acid research; review articles. Berlin: W. de Gruyter, 1990: 147-171

反义核酸在肿瘤研究中的应用*

孙丛梅 周爱儒

(北京医科大学生生化教研室, 北京 100083)

摘要 反义核酸研究已活跃于肿瘤研究及基因治疗领域. 反义核酸通过碱基配对特异性地抑制基因表达, 因此为研究肿瘤中癌基因和生长因子的功能及癌基因突变检测提供了更为有效的手段, 并为肿瘤的基因治疗提供了可能途径. 文章综述了反义核酸在基因治疗中所面临的问题及部分解决办法.

关键词 反义核酸, 寡聚脱氧核苷酸, 肿瘤

近年来, 随着反义核酸研究的深入, 更多的注意力集中于针对肿瘤的研究. 反义核酸, 它是与目的基因或目的基因 mRNA 互补, 并

以碱基配对方式与目的序列结合的核酸. 它可

*国家教委博士点基金, 国家自然科学基金资助项目.
收稿日期: 1992-11-11, 修回日期: 1993-03-22

circular and linear DNA, and other RNA N-glycosidases also have this endonucleolytic activity. The same molecular mechanism may exist in the action of trichosanthin on 28S rRNA, supercoiled DNA and HIV-1 RNA.

Key words ribosome topography, ribosome-inactivating protein (RIP), RNA N-glycosidase, trichosanthin

Transcriptional Regulations of Genes in Eucaryotic Cells. Hu Meihao. (*Department of Biology, Peking University, Beijing 100871*). *Prog. Biochem. Biophys. (China)*, 1994; **21** (2): 117

This review summarizes the transcriptional regulations in the eucaryotic genes transcribed by three kinds of RNA polymerases. The regulatory strategies differ for higher eucaryotic cells with their huge DNA contents. First, much greater numbers of transcriptional factors are required. And second, these regulatory proteins simultaneously bind at the nearby specific sites on DNA with proper orders. This demonstrated that the control of transcription in eucaryotic cells involves the interaction of protein factors with specific DNA sequence elements and the interactions between protein factors.

Key words transcriptional regulations in eucaryotic genes, RNA polymerase, transcriptional factors

The Interaction of Nuclear Factors in the Regulation of Gene Expression. Shao Hongbo, Chu Liye. (*The Laboratory for Biotechnology, Siping Normal College, Siping 136000*). *Prog. Biochem. Biophys. (China)*, 1994; **21** (2): 122
Nuclear factors are increasingly important in playing part in regulating the expression of genes. Eukaryotic transcriptional initiation is

controlled by complicated interactions between *cis*-acting DNA motifs and *trans*-acting proteins, which consist of nuclear factors. Four sequences involved in DNA sequence recognition have been determined as follows: zinc fingers; leucine-zippers; helix-turn-helix and helix-loop-helix motifs. Research from viral and animal systems turn to plant gene expression systems. Evidence has shown that the interaction of nuclear factors is the basis and prerequisite for the regulation of gene expression.

Key words nuclear factors, gene expression, *cis*-acting sequences, *trans*-acting factors, protein domains, plant genes

Factors Influencing the Expression of Foreign Genes in *E. coli*. Sui Guangchao, Hu Meihao. (*Department of Biology, Peking University, Beijing 100871*). *Prog. Biochem. Biophys. (China)*, 1994; **21** (2): 128

E. coli (*Escherichia coli*) has been widely used in expressing foreign genes, but different foreign genes may exhibit very different expression efficiencies. This article is the review of factors that influence expression of foreign genes in *E. coli*, and it will be helpful to know the information in this field, in order to take effective measures to improve expression efficiencies of foreign genes in *E. coli*.

Key words *E. coli*, foreign gene, vector, gene expression

The Application of Antisense in Cancer Research. Sun Congmei, Zhou Airu. (*Department of Biochemistry, Beijing Medical University, Beijing 100083*). *Prog. Biochem. Biophys. (China)*, 1994; **21** (2): 132

Antisense can be used to control the expression of specific genes. When targeted to specific messenger RNAs or specific sequences of the