

密码子偏爱性与外源蛋白的表达

卢希彬¹ 郭玉函² 戴平^{1*}

(1 华东师范大学生命科学学院 上海 200062 2 上海水产大学 上海 200090)

摘要 在功能蛋白的异源表达中,由于不同宿主偏好的密码子不同,很多蛋白往往表达水平很低甚至不表达。因此,很多情况下,我们需要对基因重新设计与合成。就密码子的偏好性及密码子优化的一些策略简要地进行了综述。

关键词 外源蛋白 密码子优化 基因表达

中图分类号 Q781

现已知,外源蛋白的表达水平与多种因素有关,如启动子和终止子的强弱、质粒拷贝数、密码子的偏爱性、基因二级结构、蛋白稳定性、翻译效率、培养条件等等。选择合适表达载体和宿主系统是基因表达研究者考虑的首要问题,而基因本身结构则是一个十分重要因素^[1]。对于异源蛋白表达,稀有密码子过多存在会严重影响蛋白表达水平,无论原核系统和真核系统都是如此。尤其是串联稀有密码子出现在基因读码框的5'端时,会导致表达水平急剧下降,甚至耗竭胞内 tRNA 资源,引起外源 mRNA 翻译提前终止和移码突变。因此,从基因结构本身来进行蛋白表达优化,是重组蛋白表达的一个基本策略。本文就密码子偏爱性及密码子优化问题进行了综述。

1 密码子的偏爱性

遗传密码有 64 种。在组成蛋白质 20 种氨基酸中,只有甲硫氨酸和色氨酸仅对应唯一密码子(分别为 AUG 和 UGG),其余 18 种氨基酸均拥有 2 至 6 种不同密码子,这些编码相同氨基酸的不同密码子称为同义密码子。密码子的简并性使得一种蛋白质可以采用多种不同核苷酸序列编码,不同生物甚至同种生物不同的基因,对于同一氨基酸所对应的同义密码子,使用频率也不相同^[2]。那些利用率最高称为最佳密码子(optimal codons),利用率低或不被经常利用称为稀有密码子(rare codons),这种现象称作遗传密码偏爱性。大肠杆菌,酵母,哺乳动物细胞,昆虫细胞等表达系统都表现出某种程度密码子利用率的差异或偏爱。如 *E. coli* 和人之间密码子偏性具有明显不同,哺乳动物细胞通常使用 AGG 和 AGA 编码 Arg,但却很少在 *E. coli* 中使用(分别为 2.1% 和 2.4%)。许多研究表明,氨基酸密码子 AGA 和 AGG 的高频出现可能大大影响蛋白产量。当这些密码子出现在 N 端附近时候,这种影响将达到最大。如果一个基因存在过多稀有密码子,这种蛋白往

往是低表达甚至是不表达。特别是在蛋白 N 端附近出现多个稀有密码子的时候,情况更为严重。大肠杆菌基因有 8 个使用频率很低的密码子:编码精氨酸密码子 AGA, AGG, CGG, CGA;亮氨酸密码子 CUA;异亮氨酸密码子 AUA;色氨酸密码子 UCG 和脯氨酸密码子 CCC。表 1 所示是几种宿主稀有密码子及其编码氨基酸。具体密码子利用情况的数据库网址是:<http://www.kazusa.or.jp/codon/>。

表 1 罕用密码子*

Table 1 Rare codons

大肠杆菌	酿酒酵母	果蝇	人	氨基酸
AGG	AGG			Arg
AGA		AGA		Arg
AUA		AUA		Ile
CUA				Leu
CGA	CGA	CGA	CGA	Arg
CGG	CGG	CGG	CGG	Arg
CCC				Pro
UCG			UCG	Ser
	CGC		CGC	Arg
	CCG		CCG	Pro
	CUC			Leu
	GCG		GCG	Ala
	ACG		ACG	Thr
		UUA		Leu
		GGG		Gly
		AGU		Ser
		UGU		Cys
			CGU	Arg

收稿日期:2005-12-20 修回日期:2006-02-08

* 通讯作者,电子信箱:pdai@bio.ecnu.edu.cn

* 引自 Zhang S P, et al. Gene, 1991, 105: 61~72

无论是在原核细菌还是在真核生物体内,密码子使用频率与相应 tRNA 丰度呈正相关。通常表达量高的基因含有较少种类密码子,而且这些密码子又对应于含量高 tRNA 分子,使细胞能以更快的速度合成目标蛋白;而对于需求量少的蛋白质,其基因中含有较多与低丰度 tRNA 相对应密码子,用于控制该蛋白质合成速度。例如在 *E. coli* 中, tRNA^{Arg} 阅读 AGG 和 AGA 两个 Arg 密码子,而在细胞中 tRNA^{Arg} 只具有很低水平。如果一个哺乳动物基因含有较多 AGG 和 AGA 密码子,当其在 *E. coli* 中表达时,表达水平很可能会很低。在这种情况下,选择宿主菌所偏爱密码子对基因进行点突变或重新合成基因,是基因表达研究者常常采用的一种策略。

2 密码子的优化

很多密码子优化报道都是采用 *E. coli* 表达哺乳动物蛋白,在 *E. coli* 中表达哺乳动物蛋白时,由密码子优化带来表达量增加通常在 5~15 倍之间^[3,4],经常占 *E. coli* 可溶性蛋白的 5%。通常我们可以采用以下两种措施:替换稀有密码子和将稀有密码子对应 tRNA 基因和外源基因共表达。

2.1 密码子替换

2.1.1 起始密码子周边序列的优化 对基因编码区密码子优化通常能使外源蛋白表达量得到大幅提高,即使只对编码区 5' 端进行优化也能对蛋白表达量产生显著作用。如果一个基因起始的 25 个密码子中存在较多稀有密码子,则会严重影响蛋白表达^[5]。在构建表达载体时,应避免起始密码子附近形成发夹结构,而保证翻译有效进行,因为翻译效率通常是由翻译起始点两侧 mRNA 区域二级结构的稳定性决定的。Wu 等^[6]在表达甲烷细菌中一种谷氧还蛋白样蛋白(mtGrx)时,利用 *E. coli* 偏爱密码子合成基因,却检测不到蛋白表达,Northern 分析结果表明:蛋白表达失败原因不在转录而在翻译这一环节,5' 端融合 GST 标签做融合表达时,蛋白表达量则大大提高。可能融合这段基因产生了更有利于蛋白翻译的二级结构。已经知道:5' 非编码区(5' UTR)如果 GC 含量较高,则会使翻译速率下降。在哺乳动物细胞中表达人促红细胞生成素(EPO)时, Kim 等^[7]在构建表达载体时,5' 邻近区域选用酵母偏好密码子(第三位碱基偏好 A/T,使 GC 含量下降,剩余编码序列选用人所偏好的密码子(第三位碱基偏好 G/C),使蛋白表达量提高了 13.8 倍。因此,在构建表达载体时,应根据起始密码子周边序列特征加以修饰。

2.1.2 基因编码区的优化 异源蛋白表达,特别是哺乳动物蛋白在 *E. coli* 中表达时,由于表达机制差异,很多基因往往表达水平很低。例如,直到最近才实现了人血红蛋白过表达。为了达到血红蛋白良好表达水平,Alpha-球蛋白 cDNA 不得不用大肠杆菌偏爱密码子进行重新合成。Zhou 等^[8]在不改变蛋白氨基酸序列前提下,对一种疟疾疫苗(FALVAC-1)基因进行了改造,选用

E. coli 偏爱密码子,使蛋白表达量至少提高了 3 倍。利用大肠杆菌偏好密码子(氨酰密码子 CGT)将野生型牛血红蛋白基因中编码氨酰稀有密码子(AGA 或 AGG)替换,通过基因合成得到密码子优化重组基因,将其克隆入原核表达载体 pGEX2T,经 IPTG 诱导,同样条件下,野生型牛血红蛋白基因无法得到表达,密码子优化重组基因能够得到良好表达,表达产物约占细菌总蛋白的 18%^[9]。除了替换稀有密码子,编码区核苷酸碱基组成对外源蛋白表达也有重要作用。Sinclair 等^[10]在毕赤酵母中表达人葡萄糖苷酶(GBA)时,按毕赤酵母偏好密码子优化以后,蛋白表达量提高了 10.6 倍,密码子未经优化而只改变碱基中 GC 含量(从 59.8%降低到 41.3%),蛋白表达量也提高了 7.5 倍,推测可能是由于 mRNA 碱基组成变化形成了更有利于翻译的 mRNA 二级结构,增强了蛋白稳定性,从而提高了蛋白表达量。

2.1.3 终止密码子周边序列的优化 绝大多数生物都有偏爱围绕终止密码子序列框架。大肠杆菌偏爱终止密码子为 UAA,而哺乳动物和酵母偏爱终止密码子为 UGA。同时,终止密码子下游碱基对翻译有效终止有影响,大肠杆菌中 UAAN, N 影响力次序为:U>G>A,C;哺乳动物中 UAGN, N 影响力次序为:A,G>C,U。如果终止密码子附近序列没有最佳化,可能发生明显增加翻译通读,因此减少了蛋白表达。例如,在兔网状细胞无细胞翻译系统里,UGAC 翻译通读可以高达 10%,而第四个碱基如果为 A,G 或 C,翻译通读为<1%。在设计表达载体时,可以使用两个终止密码子,以防止造成翻译通读。

总的来说,翻译起始、翻译终止序列框架和密码子利用应该仔细选择,以利于蛋白最高水平表达。

3 修饰宿主提高表达

在细胞中,tRNA 种类和数量直接反应了其 mRNA 使用密码子种类和数量偏爱性。由于密码子偏爱性不同,外源基因在 *E. coli* 中表达时,会因为稀有密码子存在,缺乏某种或某几种 tRNA 而导致翻译终止或错误。tRNA 不足会造成翻译终止,移码突变和氨基酸掺入等问题。这可以借助表达编码稀有 tRNAs 基因来克服这一问题。对于 *E. coli*,argU 基因是促进表达人基因首要目标,该基因编码稀有 tRNA^{Arg}(阅读 AGG 和 AGA 密码子),tRNA^{Ile}(阅读 AUA),tRNA^{Leu}(阅读 CUA 和 CUG),tRNA^{Pro}(阅读 CCC 和 CCU)。当宿主细胞 tRNA 含量增加后,蛋白表达会得到大幅度提高^[11]。如 Novagen 公司的 Rosetta™ 菌株就是专门用于带有 *E. coli* 稀有密码子真核蛋白表达菌株。

但有一点要注意,改变细胞内 tRNA 浓度可能会对新陈代谢产生影响,产生完全功能 tRNA 可能还需要其它细胞组分,而在单独过表达这些 tRNA 时这些组分供应可能有限。而且,tRNA 过表达策略还不够灵活,与 *E. coli* 相比,难以对真菌或哺乳动物宿主细胞进行工程改造。对于某些应用,如 DNA 疫苗,宿主工程更是不可能,

只好通过修饰密码子来获得高表达^[12]。

总体来说,在基因工程^{重组蛋白}表达中,基因^{结构}本身对^{重组蛋白}表达起着至关重要的作用。黄阳滨等^[13]根据^{酿酒酵母}密码子偏爱性等原则设计^了适合在^{酵母细胞}中表达的^{rhGH}(人生长激素)编码序列,然后利用^{常规}方法进行全基因合成,或在通过^{PCR}方法获得^的cDNA上进行点突变,大幅度地提高了^{rhGH}表达水平,得到了表达量高(1000mg/L以上)、比活性高、提纯方便和性状稳定^的rhGH产品,大大^{降低}了生产成本,非常适合大规模生产,该项目已申请专利。密码子优化^的另一个非常成功^的应用是提高病毒基因在哺乳动物细胞系中^的表达,一般通过全部^{重组}合成基因实现。在DNA疫苗研究中通常对病毒基因^的密码子进行优化,通过提高^{目的}抗原^的表达来提高免疫原性。

当然,基因结构序列只是影响^{重组蛋白}表达^的因素之一,表达载体与宿主系统^的选择,转录启动子,核苷酸序列^的组成,^{重组蛋白}稳定性,培养条件等都会对^{重组蛋白}表达产生影响。对于^{重组蛋白}表达,是一个实际加理论^的过程,需要解决^的问题很多。在表达某一^{重组蛋白}时,我们应首先根据^该重组蛋白^的性质,确定产物^的表达形式:即可溶性表达还是包涵体表达。选择合适^的载体与宿主系统,对于含有稀有密码子^的目的^{基因},利用宿主偏爱^的密码子加以改造,优化密码子基因本身^的核苷酸序列,或者增加融合^的来^{提高}tRNA^的识别效率,从而使得难表达^的基因获得表达,提高^{重组蛋白}表达量。迄今,科学家们仍在思考^{为什么}进化压力导致了密码子^的使用偏性,但不^{知道}是什么原因导致了密码子^的偏性,已日益清楚^的是密码子偏性对异源^{重组蛋白}表达有着深远^的影响。相信在异源^{重组蛋白}表达中,密码子使用会越来越受到基因表达研究者^的重视。

参考文献

- [1] Makrides S C. Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol Rev*, 1996, 60 (3):512
 [2] 张惠展. 基因工程概论. 上海:华东理工大学出版社,2000. 206~207
 [3] Mechold U, Gilbert C, Ogryzko V. Codon optimization of the

BirA enzyme gene leads to higher expression and an improved efficiency of biotinylation of target proteins in mammalian cells. *Journal of Biotechnology*. 2005,116(3): 245~249

- [4] Srivastava P, Bhattacharaya P, Pandey G, et al. Overexpression and purification of recombinant human interferon alpha2b in *Escherichia coli*. *Protein Expression, and Purification*. 2005,41(2):313~322
 [5] Chen G F, Inouye M. Suppression of the negative effect of minor arginine codons on gene expression; preferential usage of minor codons within the first 25 codons of the *Escherichia coli* genes. *Nucleic Acids Res*, 1990,18(6): 1465~1473
 [6] Wu X, Jornvall H, Berndt KD, et al. Codon optimization reveals critical factors for high level expression of two rare codon genes in *Escherichia coli*; RNA stability and secondary structure but not tRNA abundance. *Biochemical and Biophysical Research communications*, 2004, 313(1):89~96
 [7] Kim C H, Oh Y, Lee T H. Codon optimization for high-level expression of human erythropoietin(EPO) in mammalian cells. *Gene*, 1997,199(1-2):293~301
 [8] Zhou Z, Schnake P, Xiao L, et al. Enhanced expression of a recombinant malaria candidate vaccine in *Escherichia coli* by codon optimization. *Protein Expression and Purification*, 2004, 34 (1):87~94
 [9] 张雅丽,杨国庆,郭英,等. 密码子优化^的牛^{痘病毒}基因在大肠杆菌中^的表达. *高技术通讯*,2002,4:42~46
 [10] Sinclair G, Choy F Y. Synonymous codon usage bias and the expression of human glucocerebrosidase in the methylotrophic yeast, *Pichia pastoris*. *Protein Expression and Purification*, 2002,26(1):96~105
 [11] Kane J F. Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Curr Opin Biotechnol*,1995, 6 (5):494~500
 [12] Gustafsson C, Govindarajan S, Minshull J. Codon bias and heterologous protein expression. *Trends in Biotechnology*, 2004, 22(7):346~353
 [13] 黄阳滨,叶建明,杜碧金,等. 重组人生长激素^的生产方法. *中国*,200310108760. 6,2005. 5. 25

Codon Bias and Heterologous Protein Expression

LU Xi-bin¹ GUO Yu-han² DAI Ping^{1*}

(1 East China Normal Univestity Shanghai 200062,China 2 Shanghai Fisheries University Shanghai 200090,China)

Abstract Differences in codon usage among organisms often lead to the expression problems concerning heterologous gene expression. Therefore, in many instances, the gene need to be redesigned and synthesized. This codon preference and some strategies for codon optimization were reviewed.

Key Wods Heterologous protein Codon optimization Gene expression