

基因设计对重组蛋白表达的影响研究进展

蔡海莺^{1,2}, 李杨^{1,2}, 张辉^{1,2}, 冯凤琴^{1,2}

1 浙江大学食品科学与营养系, 浙江 杭州 310058

2 浙江大学馥莉食品研究院, 浙江 杭州 310058

蔡海莺, 李杨, 张辉, 等. 基因设计对重组蛋白表达的影响研究进展. 生物工程学报, 2013, 29(9): 1201-1213.

Cai HY, Li Y, Zhang H, et al. Effects of gene design on recombinant protein expression: a review, 2013, 29(9): 1201-1213.

摘要: 利用异源表达系统生产重组蛋白已成为现代基因工程和生物工程研究热点和重点。但是研究者发现并非所有的基因都能在异源宿主中高效表达,除了宿主、分泌途径、启动子等因素外,基因自身的序列也蕴含了多种影响蛋白表达的因素,如密码子偏爱性,密码子对偏爱性,GC含量,mRNA二级结构,mRNA稳定性等。从基因设计的角度对影响蛋白表达的因素和方法进行了综述,尤其是对密码子优化和密码子对优化,详细讨论了与传统基因优化理念截然不同的密码子协调化及密码子对协调化等最新进展。

关键词: 重组表达, 基因设计, 基因优化, 密码子优化, 密码子协调化, 密码子对偏爱性

Effects of gene design on recombinant protein expression: a review

Haiying Cai^{1,2}, Yang Li^{1,2}, Hui Zhang^{1,2}, and Fengqin Feng^{1,2}

1 Department of Food Science and Nutrition, Zhejiang University, Hangzhou 310058, Zhejiang, China

2 Fuli Institute of Food Science, Zhejiang University, Hangzhou 310058, Zhejiang, China

Abstract: It has become a hotspot and keystone in gene engineering and bioengineering to produce recombinant proteins through heterologous expression systems. Unfortunately, not all the genes could be successfully and effectively expressed in heterologous hosts. The role of gene itself in regulating translation process through its intrinsic sequence characteristics such as codon bias, codon pair bias, GC content, mRNA secondary structure and mRNA stability, has been gradually elucidated.

Received: December 7, 2012; **Accepted:** May 14, 2013

Supported by: Key Science and Technology Programs in Zhejiang Province (No. 2012C12005-2), Key Innovation Team of Science and Technology in Zhejiang Province (No. 2010R50032).

Corresponding author: Fengqin Feng. Tel/Fax: +86-571-88982163; E-mail: fengfq@zju.edu.cn

浙江省重大科技专项 (No. 2012C12005-2), 浙江省重点科技创新团队一般项目 (No. 2010R50032) 资助。

网络出版时间: 2013-08-21

网络出版地址: <http://www.cnki.net/kcms/detail/11.1998.Q.20130821.1715.003.html>

Here we review these factors that influence the translation processes and their corresponding optimization methods in the process of gene design. We emphatically discussed codon bias and codon pair bias and their optimization methods. In particular, the latest theories of codon harmonization and codon pair harmonization were discussed and compared with the traditional codon and codon pair optimization strategies in gene design.

Keywords: recombinant expression, gene design, gene optimization, codon optimization, codon harmonization, codon pair bias

随着现代基因工程和生物工程理论和技术的快速发展,无论是基础研究领域,还是生物医药、农业、环保和食品等应用领域,利用异源表达系统生产更高水平的目的重组蛋白都已成为研究热点和重点。相对于原始基因在自身宿主的表达,利用异源宿主表达重组蛋白具有许多优点:首先,异源表达可能显著提高目的蛋白的表达量^[1];其次,异源表达中通常所选的宿主具有相对透彻的研究背景,以及成熟的发酵和分离纯化工艺,有利于提高生产效率、节约成本;另外,异源表达中通常所选的宿主具有生物安全性,在食品和医药方面,有利于避免使用病原宿主而引发的潜在风险^[2]。

1977年,Genentech公司的Itakura等^[3]首次完成基因的异源表达,将化学合成的人体蛋白生长抑素基因以和 β -半乳糖苷酶基因融合的形式在原核生物大肠杆菌*E. coli*中表达,并得到具有功能活性的生长抑素。随着生物技术不断发展,绝大多数的基因都能通过PCR (Polymerase chain reaction) 和 RT-PCR (Reverse transcription - Polymerase chain reaction) 等核酸扩增技术得到,但是基因合成并不能被完全替代。人们把PCR扩增技术获得的完整基因克隆到表达载体并转化到表达宿主后发现,克隆的基因经常不能表达出重组蛋白,或仅仅表达出很低水平的重组蛋白。研究人员通过各种手段一定程度地

缓解了这一问题,如提高宿主的分泌能力,使用蛋白酶缺陷的宿主,提高基因的拷贝数,采用强的启动子,融合表达,共表达分子伴侣等。但是这些手段都忽略一个潜在但又非常重要的问题,采用PCR扩增的编码蛋白基因在不同的生物体内对应的DNA序列可能差异巨大。而这些编码重组蛋白的核酸序列在其他生物体内表达时,可能在转录、翻译等不同表达水平受到各种因素的限制。因此,从不同角度对基因序列进行优化或者重新根据蛋白序列进行基因设计以提高蛋白重组表达的效率逐渐成为现代生物技术研究热点之一。但目前基因设计和优化理论尚不完善,各种优化效果参差不齐,本文根据现有报道,系统综述了基因设计对重组蛋白表达影响的研究进展。

1 密码子优化

1.1 密码子影响蛋白表达机制

不同物种的基因在密码子使用上存在着明显的偏好性,甚至同物种内不同功能的基因其密码子使用频率也存在较大的差异,密码子偏爱性对重组蛋白的异源表达具有深刻复杂的影响^[4]。在原核生物内,密码子使用频率已经被鉴定为影响翻译的基因序列元件中重要的要素^[5]。密码子偏爱性在真核生物内同样是影响蛋白表达水平的关键因素之一。目前,密码子偏爱性

影响蛋白表达的主要机制已基本确定,生物体偏爱的密码子与体内对应的 tRNA 的丰度基本呈正相关,而 tRNA 的丰度决定在蛋白翻译延伸的过程中可用的氨基酸数量,从而影响了蛋白合成的效率^[6]。

1.2 密码子优化方法

一方面异源表达系统具有自己的独特优势,另一方面,由于不同生物的密码子偏爱性存在差异,异源表达可能原始编码重组蛋白的基因从偏爱的密码子变成影响翻译的稀有密码子。如何降低这一风险已成为异源表达系统的重要步骤之一。

1.2.1 改造宿主

既然已知 tRNA 的缺乏可能导致对应的稀有密码子在蛋白表达过程中使翻译速率降低甚至终止,从而使蛋白水平显著降低,研究者为了避免蛋白表达在翻译环节受到限制,应用了一种在宿主体内(通常是细菌)通过质粒补充稀有密码子对应的 tRNA 的策略进行蛋白表达的优化^[7]。这种方法在 *E. coli* 中应用非常成功(例如 pRARE2),许多外源蛋白在 *E. coli* 中表达量都显著提高^[8]。同时,在异源宿主中共表达特殊的 tRNA 能在一定程度上修饰宿主的密码子偏爱性效应并提高目的蛋白的表达量的现象,也间接证明密码子偏爱性是通过控制可用的 tRNA 丰度影响蛋白的翻译过程。该方法已被广泛用作密码子优化提高蛋白表达的策略之一。

1.2.2 稀有密码子替换

为了使外源基因在异源表达系统内有效表达,通过优化基因的密码子是十分有效的方法。拥有所用宿主的稀有密码子越多的基因,越难在该宿主中表达出理想水平的重组蛋白,另外,

如果稀有密码子出现在靠近蛋白的 N-端部分,或者以成串形式出现,则蛋白的翻译速率将进一步降低^[9]。密码子优化策略有几种,一种是选用高频使用的密码子替代基因中存在的已选宿主的稀有密码子^[10]。可以通过定点突变的方法逐步把原始基因改成预期的基因序列,能有效地降低成本。另外,这种方法的优点还包括不需要对基因转录后的 mRNA 序列作出较大改动,假设原始基因的 mRNA 在进化过程中已经具有比较合理的结构,那么这种方法能降低由于 mRNA 大幅度改变带来的不良影响。另一种方法是在排除原始基因中的稀有密码子的基础上进行重新合成,同时使密码子组成频率更接近宿主^[11]。这种方法除了能对密码子进行修饰,还能根据研究者需要对基因序列进行其他方面的优化,包括限制性酶切位点改造、mRNA 稳定性优化等。

1.2.3 异源表达宿主的选择或同源表达

密码子的偏爱性与物种的亲缘关系有一定的相关性,亲缘关系近的物种具有相对类似的密码子偏爱性,为了降低密码子偏爱性对蛋白表达影响的风险,较为简单的方法是在与基因宿主有类似密码子偏爱性的异源表达宿主中表达这一基因。

利用同源表达(Homologous expression)技术发掘生物体自身的蛋白,尤其是酶资源,受到越来越多的关注。同源表达技术可以用来表达在其他异源表达系统中难以表达的蛋白,显著提高该蛋白的产量。Mayfield 等^[12]在黄孢原毛平革菌 *Phanerochaete chrysosporium* 中成功表达了启动子改造后的自身蛋白锰过氧化氢酶基因。Jia 等^[13]在洋葱伯克霍尔德氏菌 *Burkholderia*

cepacia 通过同源表达的方式过量表达了该菌株自身的脂肪酶基因。同源表达与异源表达相比,具有以下几方面的优点: 1) 基因来源于自身, 表达菌株并不是严格意义上的转基因, 在应用领域 (尤其是食品) 更容易推广; 2) 表达的蛋白与宿主有更好的相容性, 更有可能获得较高的表达水平; 3) 需要较少的基因修饰, 如密码子和密码子环境方面的修饰等。

1.3 密码子协调化

随着对蛋白表达研究的不断深入, 研究人员发现即使宿主中高表达的基因中也存在一定的稀有密码子。这引发了人们对密码子和蛋白表达关系的新思考。由于核糖体对 mRNA 不同区域的翻译速率存在差异, 并且蛋白质在核糖体能够进行翻译时折叠, Thanaraj 等提出假设, 翻译动力学可能影响新生肽链的折叠过程。他们进一步研究发现, 高级结构或结构域之间的连接区域和

蛋白域边界 (Protein domain boundary) 通常由 RNA 上翻译缓慢的区域编码 (图 1), 同时组成这部分区域的氨基酸也大多是能够通过与核糖体的新生肽链通道发生粘连作用而减缓翻译速率的^[14-15]。因此, 蛋白合成与蛋白的体内折叠并不是独立的两个过程, 而是有着非常密切的相关性。另外, Komar 等^[16]通过体外翻译实验证明同义密码子之间的更换并不是严格的沉默突变, 它能够影响蛋白质翻译延伸的速率, 从而导致蛋白质折叠和功能发生变化。这表明基因核苷酸序列也一定程度上决定了重组蛋白的结构和功能。上述研究表明, 密码子的优化必须与蛋白的高级结构联系起来, 如果通过一味地替换稀有密码子来提高重组蛋白的表达可能会起到适得其反的结果。目前仍有很多蛋白不能通过密码子优化来实现高效表达也可能是欠缺这方面的考虑。

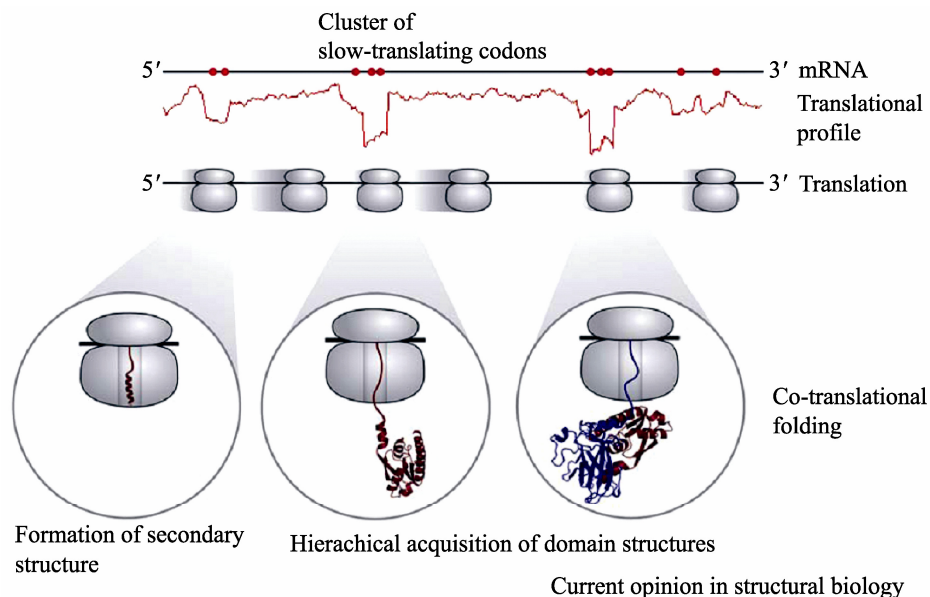


图 1 新生肽链翻译的不连续性与翻译时折叠相关^[15]

Fig. 1 Local discontinuous translation actively coordinates co-translational folding of the nascent chain^[15].

既然连接蛋白高级结构元件的域边界区能够偏好地使用一系列特定的氨基酸,同时采用稀有的密码子编码这些氨基酸来帮助蛋白的合理折叠,Angov 等^[17]提出了一种被称为密码子协调化(Codon harmonization)的密码子优化规则,将基因在原始宿主内通过编码蛋白域边界密码子的选择,间接调控蛋白折叠的机制复制到需要异源表达的宿主。密码子协调化首先需要密码子使用频率和蛋白的二级结构来确定 mRNA 的慢速翻译区域,即编码蛋白结构元件的连接或终止区域;然后通过同义密码子替换使结构基因在新的宿主中每个位置的密码子使用频率与原始宿主的尽量相符。

这种优化方法对于提高重组蛋白的异源表达已被多次证明十分有效。由于恶性疟原虫 *Plasmodium falciparum* 结构基因 AT 含量高达 80%^[18],因此利用 *E. coli* 异源系统高效表达这类基因编码的蛋白异常艰难。Darko 等^[19]利用密码子协调化的优化方法,分别获得了编码恶性疟原虫菌株 FVO 的主裂殖子表面蛋白 1 (Merozoite surface protein, MSP1) 的 C 端部分 42 kDa 的蛋白片段 MSP1₄₂ (FVO) 基因的两种优化基因 FMP003 和 LSA-NRC^H,分别对应在预测的蛋白域连接区域改变了一个密码子和对全序列的密码子协调化优化,两种基因都成功实现了恶性疟原虫 MSP1₄₂ (FVO) 蛋白片段的可溶性表达。Angov 等^[17]在 Darko 等研究的基础上,系统比较了 3 种恶性疟原虫疫苗候选蛋白 MSP142 (FVO)、MSP142 (3D7) 和 MSP142 (Camp) 的原始基因和密码子协调化优化基因在 *E. coli* 的表达情况,结果表明通过优化的基因序列都能显著提高蛋白的表达量和可溶性,对应的蛋白表

达水平提高了 4~1 000 倍不等。另外,他们还比较了对编码 MSP142 (FVO) 的基因分别采取不同的优化方法的优化效果,结果显示密码子协调化获得的序列 LSA-NRCH 在 *E. coli* 的表达量明显高于对应的替换所有稀有密码子的优化方法获得的序列 LSA-NRCE。Chowdhury 等^[20]通过密码子协调化的优化方法成功地在 *E. coli* 表达系统中高效表达了恶性疟原虫蛋白 Pfs48/45,而且获得的重组蛋白 CH-rPfs48/45 能够被对应的天然蛋白的抗原决定簇的单克隆抗体识别。

因此,密码子协调化是现有密码子优化策略的一个重要补充,有望帮助目前难以在异源表达系统中重组表达的蛋白实现高效表达,尤其是需要可溶性、功能性表达的蛋白。

2 密码子对 (Codon pair) 优化

2.1 密码子对与蛋白表达水平关系

除了密码子使用频率的非随机性外,越来越多的证据表明密码子环境 (Codon context) 也影响密码子与反义密码子之间的识别,从而影响翻译延伸的速率。最早的证据来源于对无义突变密码子的研究,研究者发现无义突变抑制子 (Suppressor) 对无义密码子的抑制效率明显地受到附近核苷酸的影响^[21]。同样的现象也发生在抑制子对错义突变包括移码突变的抑制^[22-23]。显然,机体抑制子对无义密码子的抑制活性受密码子环境影响表明终止密码子的翻译终止同样是环境依赖性的。另外,对大肠杆菌蛋白的编码序列的统计分析显示,密码子对 (Codon pair) 同样具有很高的偏爱性。实际统计得到的 3 721 (61²) 种密码子对与以 61 种密码子使用频率计算得到的对应密码子对的随机期望

值 (Random expectation) 相比, 部分密码子对的实际出现频率高于预测值, 称为被过度代表的 (Overrepresented); 部分密码子对的实际出现频率低于预测值, 称为代表不足的 (Underrepresented)^[24-27]。越来越多的实验表明位于翻译核糖体表面的 A 和 P 位点相邻的氨酰 tRNA 之间的相容性, 可能是引起翻译过程中密码子环境效应的原因^[28]。这种相容性可能通过改变翻译速率, 从而成为调节蛋白表达的重要因素之一。

进一步的统计分析证明了这个结论, 在大肠杆菌基因组内, 密码子对利用模式与基因的表达水平存在相关性, 相对于低水平表达的基因, 编码高水平表达的蛋白基因, 倾向于包含较多高频出现但代表不足的密码子对。其中, 高频的密码子对可能保证了基因使用的是高频的密码子, 从而有利于基因的翻译表达^[24-25]。和密码子偏爱性类似, 不同生物的密码子对使用频率之间也有明显差异, 具有各自的密码子对偏爱性 (Codon pair bias)^[27-29], 因此, 编码重组蛋白的目的基因需要在外源的表达系统中表达时, 密码子对偏爱性可能是影响蛋白重组表达的重要因素之一。

2.2 密码子对优化

如果需要在大肠杆菌中高效重组表达某一外源基因, 对密码子对进行优化可能是不可或缺的一环。Hatfield 等^[30]利用密码子对偏爱性, 发明了一种新型的基因设计优化方法 CODA (Computationally optimized DNA assembly)。在 *E. coli* 中对编码酿酒酵母 Ty3 反转座子 GAG 基因的衣壳部分片段进行优化表达的结果显示,

相对于原始基因在 *E. coli* 中的表达, 尽管优化该基因片段的密码子使用频率 (密码子优化) 的方法能显著提高目的蛋白的表达水平, 但是在密码子优化基础上进行密码子对偏爱性的优化, 既避免过度代表密码子对出现的优化方法, 还能使蛋白表达在密码子优化的水平上获得进一步的提高。

同样地, 利用密码子对偏爱性, 科学家还设计了有目的地使蛋白表达显著降低的应用系统。接种减毒活疫苗是用来预防病毒、细菌和原虫等病原引起的流行性疾病的常用策略之一。Coleman 等^[31]通过合成性减毒病毒工程 (Synthetic attenuated virus engineering, SAVE) 的方法, 重新设计合成了经过密码子对同义替换处理的编码脊髓灰质炎病毒衣壳蛋白的 DNA 大分子。在保证其他影响翻译的因素 (如密码子使用频率、RNA 二级结构等) 不变的情况下, 在接种的小鼠中提高代表不足的密码子对同义替换的基因设计引起对应病毒衣壳蛋白的翻译速率明显下降, 从而导致包含该氨基酸依赖变化的脊髓灰质炎病毒毒性减弱; 反之, 提高过度代表的密码子对的基因设计则提高了对应病毒衣壳蛋白的翻译水平。令人意外的是, 这种结果与前期研究者在 *E. coli* 系统中获得的结论, 即代表不足的密码子对有利于蛋白的翻译刚好相反。随后, Muller 等^[32]利用 SAVE 方法通过全基因组范围的密码子对替换, 对流感病毒株 A/PR/8/34 进行病毒毒性减弱的理性设计。与 Coleman 等的结果类似, 相对于野生型的流感病毒, 通过使用代表不足的密码子对替换设计的病毒在小鼠中的毒性显著减弱。上述结果可能暗示

在哺乳动物小鼠体内密码子偏爱性对蛋白表达的作用模式与原核生物 *E. coli* 的模式非常不同。这可能跟密码子环境保守性有关, 高等真核生物相比其他种类生物只有较低的密码子环境保守性^[27]。但是, 这并不能解释接下来的实验结果。Coleman 等^[33]还通过密码子对替换的基因设计对肺炎链球菌 *Streptococcus pneumoniae* 进行毒性减弱, 他们在保证氨基酸序列不变的基础上尽量使用代表不足的密码子对重新编码 *S. pneumoniae* 的血清 3 型 (SP3) 的溶血素 (Pneumolysin, *ply*) 基因。与野生型或者 *ply* 缺失的 Δ *ply* SP3 菌种相比, 密码子对改造菌种的溶血素蛋白表达量明显降低, 并且对小鼠的毒性以及在肺中引起的炎症反应都显著性下降。这表明同样属于原核生物的 *S. pneumoniae* 和 *E. coli*, 密码子对对于蛋白表达的作用模式也非常不同。

综上所述, 密码子对偏爱性与蛋白表达存在明显的相关性, 因此, 通过对密码子对偏爱性进行优化是提高异源表达重组蛋白的重要环节之一。然而, 不同生物体内, 由于密码子对偏爱性对蛋白翻译过程的作用模式可能不同, 相关的作用机理也比较模糊, 因此, 应针对不同的生物自身特点来进行密码子对偏爱性的优化。

2.3 密码子对协调化

与密码子协调化理论对应, 研究发现高级结构或结构域之间的连接区域和蛋白域边界往往由 RNA 上翻译缓慢的区域编码, 因此编码组成这部分区域的氨基酸序列也经常存在能够使翻译速率降低的密码子对。经统计计算分析, 编码人免疫缺陷病毒衣壳蛋白和酿酒酵母 Ty3 反转座子 GAG 的衣壳部分片段的基因序列在蛋白

边界区都不同程度地存在过度代表的密码子对^[30], 这些密码子对被认为在翻译过程中, 通过降低翻译速率而有意暂停, 来保证有足够的时间来完成新生肽链的高级结构和结构域的折叠。

Trinh 等^[34]在保证密码子使用频率在哺乳动物细胞中相对稳定的情况下, 通过对抗 Her2/Neu 单链抗体 (Single chain variable region fragment, ScFv) 中连接重链可变区 (V_H) 和轻链可变区 (V_L) 之间的接头, 以及单链抗体与人抗鼠铁传递蛋白受体 IgG3 重链 C_H3 的接头 (GGGS)₃ 两处编码序列进行密码子对优化, 并转染大鼠细胞进行表达, 结果显示, 优化序列中单个核苷酸的改变使得该融合蛋白表达提高 30 倍以上, 并且通过对 mRNA 定量分析证实蛋白表达量的提高完全由翻译效率提高实现。然而, 与实验预期相反的是, 密码子对的优化是将过度代表的密码子对换成代表过少的密码子对, 而前期的研究表明代表过少的密码子对在 *E. coli* 中是有利于翻译的, 但作为蛋白域边界的接头需要翻译有意的暂停来帮助蛋白高级结构的折叠。因此, Trinh 等认为对于该实验表达的融合蛋白可能在接头处不需要翻译暂停太长时间。然而, 根据 Coleman 等和 Muller 等的研究结果, 这种解释可能并不妥当, 而可能是由于不同生物中, 密码子对作用于蛋白翻译的模式并不相同, 与 *E. coli* 相反, 在哺乳动物中, 过度代表的密码子对有利于蛋白的翻译。如果这一结论成立的话, 则能与密码子协调化的结论相呼应, 即高级结构或结构域之间的连接区域和蛋白域边界需要一些降低翻译的因素, 如稀有密码子、密码子对等, 以便提供蛋白的高级

结构或结构域在翻译时充分正确地折叠。但是, 由于相关的研究还较少, 这一假设还需要进一步验证。

3 其他基因设计要素

3.1 基因 GC 含量

研究表明, 基因 GC 含量是基因设计和优化的重要指标^[35], 不同物种间基因组的 GC 含量有显著差异, 在宿主中表达某些其他来源的基因时很容易引起人们的注意。GC 含量通常间接对基因表达进行调控和影响。

3.2 mRNA 稳定性

蛋白的表达调控分为转录水平和翻译水平, 一般来说, 转录水平起关键作用, 但是同时, 翻译的效率与 mRNA 的降解直接相关, 因此, 也间接影响着转录后水平的调控。如原核生物中, 存在稀有密码子的 mRNA 由于翻译延伸过程受到影响, 使得 mRNA 得不到更多核糖体结合后的有效保护, 也将导致 mRNA 的降解从而使积累水平显著降低, 蛋白表达已被多次证明与 mRNA 稳定性相关^[36-37]。但是对于 mRNA 稳定性的优化比较困难, 缺乏比较成熟的理论体系, 相反, 过高的 mRNA 稳定性意味着过高的 GC 含量及稳定的二级结构等不利于蛋白翻译的因素。

3.3 mRNA 二级结构

研究显示核糖体仅仅能够与单链 RNA 结合并起始翻译^[38], 另外, 较大的折叠自由能可能减缓核糖体的延伸从而降低翻译效率^[39]。因此, mRNA 如果形成大而稳定的二级结构如发卡结构和茎环结构, 尤其是起始密码子附近的稳定

二级结构, 将会影响 mRNA 在翻译过程中核糖体的结合和延伸, 从而降低翻译的效率和最终的蛋白表达水平^[39-41]。

Kudla 等^[42]构建了包含 154 种不同随机密码子同义突变的 GFP 蛋白突变体库, 通过对其在大肠杆菌中的蛋白表达水平, 研究 mRNA 折叠自由能对翻译效率的影响。结果表明, 在排除密码子偏爱性因素后, GFP 蛋白表达丰度与转录的 mRNA 的前 40 个核苷酸片段的折叠自由能显著相关。Tuller 等^[39]通过对大肠杆菌和酿酒酵母转录组的研究发现, mRNA 的前 40 核苷酸片段的平均自由能明显小于下游长度为 40 的核苷酸片段的平均自由能, 表明生物在进化过程中可能要求更有效的翻译起始, 从而驱使 mRNA 起始密码子附近二级结构稳定性降低。这种进化选择性在其他物种中同样存在^[43]。另外, Tuller 等^[39]还发现 mRNA 的 41~80 核苷酸的平均自由能明显大于其他长度为 40 的核苷酸片段的平均自由能, 这可能是生物体通过这段核苷酸片段形成自身的稳定结构, 防止其与核糖体结合位点的核苷酸序列形成潜在的有害结构的一种保护翻译有效起始的策略。Desmit 等^[44]通过对噬菌体 MS2 的外壳蛋白基因的同义密码子突变体在大肠杆菌中的表达, 定量分析了翻译起始区的 mRNA 二级结构与翻译效率的关系, 结果表明翻译起始区 mRNA 折叠自由能每增大 1.4 kcal/mol, 对应的蛋白翻译起始率和表达量则降低 10 倍。但是, 必须指出的是, 只有当翻译的起始受到 mRNA 二级结构影响, 或者该 mRNA 与核糖体的亲和度较低的时候, 这种比例关系才会存在。

因此,在基因序列设计和优化的时候,合理地利用翻译起始区的 mRNA 二级结构的规律将有效提高目的蛋白的异源表达水平。上述推断已被国内外研究者多次应用并证明有效。Punginelli 等^[45]通过突变大肠杆菌甲酸脱氢酶 N 的亚基 G (FdnG) 信号肽的第一个精氨酸 (R5),降低了翻译起始区强稳定性茎环结构的形成,显著减小了 mRNA 该区域折叠自由能,使目的蛋白的表达水平最高提高了 60 倍。王珊珊等^[41]对近平滑假丝酵母的(R)-羧基还原酶基因的 mRNA 翻译起始区中+1~+78 区进行二级结构的优化,使酶蛋白在大肠杆菌的表达水平比优化前提高了 4~5 倍。通过 mRNA 二级结构优化,

尤其是翻译起始区的二级结构优化提高目的蛋白的异源表达在基因工程领域的应用已受到越来越多的重视。

3.4 其他因素

其他要素或优化方法还包括检查单核苷酸重复和密码子重复,核对核糖体结合位点,起始密码子环境,终止密码子及其环境,避免内含子、隐蔽剪接位点、AT 富含区、内部核糖体进入位点 (IRES)、重组位点等不利元件,选择合适的 UTR 序列和信号肽序列,合理设计酶切位点、接头、融合基因、检测和纯化标签等(如表 1),使基因进一步优化,检查无误后进行基因合成,以备后续的遗传转化和表达。

表 1 其他影响蛋白表达的基因设计因素

Table 1 Other gene design factors that influence protein expression

Gene design factors	Influences and mechanisms	References
GC content	Low GC content tends to terminate transcription	[35]
	High GC content may result in strong secondary structures of mRNA Related to codon bias	
mRNA stability	Influencing the accumulation of mRNA	[36–37]
mRNA secondary structures	Influencing the initiation and elongation efficiency of translation	[39–41]
Long mononucleotide repeats	Resulting in mutation through transcription and translation slippage Related to codon bias and codon pair bias	[46]
Codon repeats	Related to codon pair bias	[47]
SD sequence	Ribosome binding site in prokaryotes Related to mRNA stability	[48]
Kozak sequence	Ribosome recognition site in eukaryotes	[49]
Start codon context	Strong mRNA secondary structures nearby start codon may decrease the efficiency of translation initiation	[36, 50]
Stop codons	Different in termination efficiency of translation	[51]
Stop codon context	Related to translation termination (especially higher eukaryotes)	[52]
Intron	Resulting in shift mutation through improper splicing Enhancing translation efficiency	[53]
AT-rich region	Similar to transcription termination signal	[35, 54]
Internal ribosome entry site (IRES)	Influencing translation efficiency and integrality of protein	[55]
UTR region (Untranslated region)	Regulating translation	[56–57]
Signal peptide	Targeting and translocating of protein	[58–59]
Protease cleavage site	Separation of proteins from fusion form with proteins or tags	[60]

4 讨论

目前,越来越多的物种被发展成重组蛋白表达系统,研究比较深入并且应用相对广泛的表达系统包括大肠杆菌表达系统、枯草芽胞杆菌表达系统、链霉菌表达系统、毕赤酵母表达系统、酿酒酵母表达系统、昆虫表达系统、哺乳动物表达系统和植物表达系统等,另外,一些细胞的细胞器(如叶绿体和线粒体等)也被用于重组蛋白表达。利用细胞作为异源表达宿主生产重组蛋白被形象地称为细胞工厂,分子农业,细胞生物反应器等。基因设计有助于原始基因难以表达的蛋白在异源重组系统内表达,以及提高目的蛋白的表达量。同样,高效异源重组表达能使异源表达系统宿主作为生物反应器生产除蛋白多肽外的其他代谢中间产物。本实验室结合传统的微生物筛选及菌种诱变等方法,利用重组表达系统高效表达 sn-1,3 专一性脂肪酶蛋白,并对其进行酶制剂化,以用于功能油脂 1,3-二油酸-2-棕榈酸甘油三酯(OPO)的酶法制备。本实验室在 sn-1,3 专一性脂肪酶的微生物高效表达方面已取得一定进展,今后将在 sn-1,3 专一性脂肪酶在模式生物中的重组表达方面,结合蛋白和宿主自身特点,进一步利用基因设计和优化提高脂肪酶的表达。

尽管通过异源表达系统提高重组蛋白的表达量的报道已不胜枚举,但是绝大多数的蛋白还不能通过异源系统表达实现产业化之路。通过基因工程和生物工程的手段,不断改善和优化目的蛋白异源重组表达,还是今后发展的重要研究领域。另外,尽管人们在基因设计和优化方面已获得很多成功,对影响基因在转录和翻译表

达水平的各种影响因素及其作用机制的研究也取得长足进步,但是由于这些影响因素之间并非独立存在,它们往往相互作用,构成围绕蛋白表达调控的动态网络,而目前的基因设计分析和优化手段并不能达到从整体上完全掌控和调节这些因素,更多的是从单个因素或者少数几个较为重要的因素进行优化和设计。因此,现行的优化和设计方法都不同程度地存在局限性,目前还没有非常成熟基因优化理论。因此,利用基因设计优化策略提高异源宿主蛋白表达的方法应该被看作是必要条件,而不是充分条件,同时还应该考虑细胞、个体及环境等其他因素在基因转录、翻译等表达水平的影响。

REFERENCES

- [1] Valero F. Heterologous expression systems for lipases: a review. *Methods Mol Biol*, 2012, 861: 161-178.
- [2] Schuster E, Dunn-Coleman N, Frisvad JC, et al. On the safety of *Aspergillus niger*--a review. *Appl Microbiol Biotechnol*, 2002, 59(4/5): 426-435.
- [3] Itakura K, Hirose T, Crea R, et al. Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. *Science*, 1977, 198(4321): 1056-1063.
- [4] Fiers W, Grosjean H. Codon usage. *Nature*, 1979, 277(5694): 328-328.
- [5] Lithwick G, Margalit H. Hierarchy of sequence-dependent features associated with prokaryotic translation. *Genome Res*, 2003, 13(12): 2665-2673.
- [6] Karlin S, Mrazek J, Campbell A, et al. Characterizations of highly expressed genes of four fast-growing bacteria. *J Bacteriol*, 2001, 183(17): 5025-5040.
- [7] Brinkmann U, Mattes RE, Buckel P. High-level expression of recombinant genes in *Escherichia*

- coli* is dependent on the availability of the dnaY gene product. *Gene*, 1989, 85(1): 109–114.
- [8] Zdanovsky AG, Zdanovskaia MV. Simple and efficient method for heterologous expression of clostridial proteins. *Appl Environ Microbiol*, 2000, 66(8): 3166–3173.
- [9] Clarke TFt, Clark PL. Rare codons cluster. *PLoS ONE*, 2008, 3(10): e3412.
- [10] Komar AA, Guillemet E, Reiss C, et al. Enhanced expression of the yeast Ure2 protein in *Escherichia coli*: the effect of synonymous codon substitutions at a selected place in the gene. *Biol Chem*, 1998, 379(10): 1295–1300.
- [11] Mellitzer A, Weis R, Glieder A, et al. Expression of lignocellulolytic enzymes in *Pichia pastoris*. *Microb Cell Fact*, 2012, 11(1): 61.
- [12] Mayfield MB, Kishi K, Alic M, et al. Homologous expression of recombinant manganese peroxidase in *Phanerochaete chrysosporium*. *Appl Environ Microbiol*, 1994, 60(12): 4303–4309.
- [13] Jia B, Yang JK, Liu WS, et al. Homologous overexpression of a lipase from *Burkholderia cepacia* using the lambda Red recombinase system. *Biotechnol Lett*, 2010, 32(4): 521–526.
- [14] Thanaraj TA, Argos P. Ribosome-mediated translational pause and protein domain organization. *Protein Sci*, 1996, 5(8): 1594–1612.
- [15] Zhang G, Ignatova Z. Folding at the birth of the nascent chain: coordinating translation with co-translational folding. *Curr Opin Struct Biol*, 2011, 21(1): 25–31.
- [16] Komar AA, Lesnik T, Reiss C. Synonymous codon substitutions affect ribosome traffic and protein folding during *in vitro* translation. *FEBS Lett*, 1999, 462(3): 387–391.
- [17] Angov E, Hillier CJ, Kincaid RL, et al. Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. *PLoS ONE*, 2008, 3(5): e2189.
- [18] Weber JL. Analysis of sequences from the extremely A+T-rich genome of *Plasmodium falciparum*. *Gene*, 1987, 52(1): 103–109.
- [19] Darko CA, Angov E, Collins WE, et al. The clinical-grade 42-kilodalton fragment of merozoite surface protein 1 of *Plasmodium falciparum* strain FVO expressed in *Escherichia coli* protects *Aotus nancymai* against challenge with homologous erythrocytic-stage parasites. *Infect Immun*, 2005, 73(1): 287–297.
- [20] Chowdhury DR, Angov E, Kariuki T, et al. A potent malaria transmission blocking vaccine based on codon harmonized full length Pfs48/45 expressed in *Escherichia coli*. *PLoS ONE*, 2009, 4(7): e6352.
- [21] Bossi L, Ruth JR. The influence of codon context on genetic code translation. *Nature*, 1980, 286(5769): 123–127.
- [22] Gaber RF, Culbertson MR. Codon recognition during frameshift suppression in *Saccharomyces cerevisiae*. *Mol Cell Biol*, 1984, 4(10): 2052–2061.
- [23] Bossi L, Roth JR. Four-base codons ACCA, ACCU and ACCC are recognized by frameshift suppressor suffJ. *Cell*, 1981, 25(2): 489–496.
- [24] Boycheva S, Chkodrov G, Ivanov I. Codon pairs in the genome of *Escherichia coli*. *Bioinformatics*, 2003, 19(8): 987–998.
- [25] Gutman GA, Hatfield GW. Nonrandom utilization of codon pairs in *Escherichia coli*. *Proc Natl Acad Sci USA*, 1989, 86(10): 3699–3703.
- [26] Irwin B, Heck JD, Hatfield GW. Codon pair utilization biases influence translational elongation step times. *J Biol Chem*, 1995, 270(39): 22801–22806.
- [27] Moura GR, Pinheiro M, Freitas A, et al. Species-specific codon context rules unveil non-neutrality effects of synonymous mutations. *PLoS ONE*, 2011, 6(10): e26817.
- [28] Buchan JR, Aucott LS, Stansfield I. tRNA properties help shape codon pair preferences in open reading frames. *Nucleic Acids Res*, 2006, 34(3): 1015–1027.

- [29] Moura G, Pinheiro M, Arrais J, et al. Large scale comparative codon-pair context analysis unveils general rules that fine-tune evolution of mRNA primary structure. *PLoS ONE*, 2007, 2(9): e847.
- [30] Hatfield GW, Roth DA. Optimizing scaleup yield for protein production: Computationally Optimized DNA Assembly (CODA) and Translation Engineering. *Biotechnol Annu Rev*, 2007, 13: 27–42.
- [31] Coleman JR, Papamichail D, Skiena S, et al. Virus attenuation by genome-scale changes in codon pair bias. *Science*, 2008, 320(5884): 1784–1787.
- [32] Mueller S, Coleman JR, Papamichail D, et al. Live attenuated influenza virus vaccines by computer-aided rational design. *Nat Biotechnol*, 2010, 28(7): 723–726.
- [33] Coleman JR, Papamichail D, Yano M, et al. Designed reduction of *Streptococcus pneumoniae* pathogenicity via synthetic changes in virulence factor codon-pair bias. *J Infect Dis*, 2011, 203(9): 1264–1273.
- [34] Trinh R, Gurbaxani B, Morrison SL, et al. Optimization of codon pair use within the (GGGS)₃ linker sequence results in enhanced protein expression. *Mol Immunol*, 2004, 40(10): 717–722.
- [35] Baca AM, Hol WG. Overcoming codon bias: a method for high-level overexpression of Plasmodium and other AT-rich parasite genes in *Escherichia coli*. *Int J Parasitol*, 2000, 30(2): 113–118.
- [36] Mitarai N, Sneppen K, Pedersen S. Ribosome collisions and translation efficiency: optimization by codon usage and mRNA destabilization. *J Mol Biol*, 2008, 382(1): 236–245.
- [37] Tanaka M, Tokuoka M, Shintani T, et al. Transcripts of a heterologous gene encoding mite allergen Der f 7 are stabilized by codon optimization in *Aspergillus oryzae*. *Appl Microbiol Biotechnol*, 2012, 96(5): 1275–1282.
- [38] Desmit MH, Vanduin J. Control of prokaryotic translational initiation by messenger-RNA secondary structure. *Progr Nucl Acid Res Mol Biol*, 1990, 38: 1–35.
- [39] Tuller T, Waldman YY, Kupiec M, et al. Translation efficiency is determined by both codon bias and folding energy. *Proc Natl Acad Sci USA*, 2010, 107(8): 3645–3650.
- [40] Freischmidt A, Liss M, Wagner R, et al. RNA secondary structure and *in vitro* translation efficiency. *Protein Expr Purif*, 2012, 82(1): 26–31.
- [41] Wang S, Zhang R, Geng Y, et al. Improved expression and catalytic efficiency of (R)-carbonyl reductase in *Escherichia coli* by secondary structure optimization of mRNA translation initiation region. *Chin J Biotech*, 2009, 25(12): 1907–1913 (in Chinese).
王珊珊, 张荣珍, 耿亚维, 等. mRNA 翻译起始区二级结构优化提高(R)-羧基还原酶的表达及催化效率. *生物工程学报*, 2009, 25(12): 1907–1913.
- [42] Kudla G, Murray AW, Tollervey D, et al. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science*, 2009, 324(5924): 255–258.
- [43] Zhou T, Wilke CO. Reduced stability of mRNA secondary structure near the translation-initiation site in dsDNA viruses. *BMC Evol Biol*, 2011, 11.
- [44] Desmit M H, Vanduin J. Secondary structure of the ribosome binding-site determines translational efficiency - a quantitative-analysis. *Proc Natl Acad Sci USA*, 1990, 87(19): 7668–7672.
- [45] Punginelli C, Ize B, Stanley NR, et al. mRNA secondary structure modulates translation of Tat-dependent formate dehydrogenase N. *J Bacteriol*, 2004, 186(18): 6311–6315.
- [46] Gu T, Tan S, Gou X, et al. Avoidance of long mononucleotide repeats in codon pair usage. *Genetics*, 2010, 186(3): 1077–1084.
- [47] Tats A, Tenson T, Remm M. Preferred and avoided codon pairs in three domains of life. *BMC Genomics*, 2008, 9: 463.
- [48] Agaisse H, Lereclus D. STAB-SD: a

- Shine-Dalgarno sequence in the 5' untranslated region is a determinant of mRNA stability. *Mol Microbiol*, 1996, 20(3): 633–643.
- [49] Kozak M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*, 1986, 44(2): 283–292.
- [50] Gingold H, Pilpel Y. Determinants of translation efficiency and accuracy. *Mol Syst Biol*, 2011, 7: 481.
- [51] Sun J, Chen M, Xu J, et al. Relationships among stop codon usage bias, its context, isochores, and gene expression level in various eukaryotes. *J Mol Evol*, 2005, 61(4): 437–444.
- [52] Pacho F, Zambruno G, Calabresi V, et al. Efficiency of translation termination in humans is highly dependent upon nucleotides in the neighbourhood of a (premature) termination codon. *J Med Genet*, 2011, 48(9): 640–644.
- [53] Kamo K, Kim AY, Park SH, et al. The 5'UTR-intron of the *Gladiolus* polyubiquitin promoter GUBQ1 enhances translation efficiency in *Gladiolus* and *Arabidopsis*. *BMC Plant Biol*, 2012, 12: 79.
- [54] Nishikubo T, Nakagawa N, Kuramitsu S, et al. Improved heterologous gene expression in *Escherichia coli* by optimization of the AT-content of codons immediately downstream of the initiation codon. *J Biotechnol*, 2005, 120(4): 341–346.
- [55] Attal J, Theron MC, Puissant C, et al. Effect of intercistronic length on internal ribosome entry site (IRES) efficiency in bicistronic mRNA. *Gene Expr*, 1999, 8(5/6): 299–309.
- [56] Wang J, Lu C, Min D, et al. A mutation in the 5' untranslated region of the BRCA1 gene in sporadic breast cancer causes downregulation of translation efficiency. *J Int Med Res*, 2007, 35(4): 564–573.
- [57] Koda A, Minetoki T, Ozeki K, et al. Translation efficiency mediated by the 5' untranslated region greatly affects protein production in *Aspergillus oryzae*. *Appl Microbiol Biotechnol*, 2004, 66(3): 291–296.
- [58] Klatt S, Konthur Z. Secretory signal peptide modification for optimized antibody-fragment expression-secretion in *Leishmania tarentolae*. *Microb Cell Fact*, 2012, 11: 97.
- [59] Futatsumori-Sugai M, Tsumoto K. Signal peptide design for improving recombinant protein secretion in the baculovirus expression vector system. *Biochem Biophys Res Commun*, 2010, 391(1): 931–935.
- [60] Pozzuolo S, Breme U, Salis B, et al. Efficient bacterial expression of fusion proteins and their selective processing by a recombinant Kex-1 protease. *Protein Expr Purif*, 2008, 59(2): 334–341.

(本文责编 郝丽芳)