

• 生物工程与大健康 •

霍毅欣 北京理工大学生命学院教授, 博士生导师。1999 年在南开大学获得微生物学学士学位, 2005 年同时获得北京大学的生物化学与分子生物学专业博士学位和法国巴黎七大的分子微生物学及病毒学专业博士学位。2006-2011 年在加州大学洛杉矶分校做博士后研究, 2015 年 10 月起在北京理工大学任教。现任中国生物工程学会青年工作委员会副主任、中国化工学会生物化工青年学者工作委员会常务委员、中国生物工程学会氨基酸专业委员会委员。主要研究方向是代谢工程和合成生物学, 已在 *Science*、*Nature Biotechnology*、*Nature Communications*、*Current Opinion in Biotechnology*、*Nucleic Acids Research* 等刊物发表论文 30 余篇。



密码子优化策略在异源蛋白表达中的应用

杨云彭¹, 马晓焉¹, 霍毅欣^{1,2}

1 北京理工大学 生命学院 分子医学与生物诊疗工业和信息化部重点实验室, 北京 100081

2 苏州工业园区洛加大先进技术研究院 (苏州), 江苏 苏州 215123

杨云彭, 马晓焉, 霍毅欣. 密码子优化策略在异源蛋白表达中的应用. 生物工程学报, 2019, 35(12): 2227-2237.

Yang YP, Ma XY, Huo YX. Application of codon optimization strategy in heterologous protein expression. Chin J Biotech, 2019, 35(12): 2227-2237.

摘要: 酶在医疗和生物药物方面有着广泛的应用, 不仅可以用来治疗各种疾病, 还在临床诊断和人体健康等方面有着重要的影响。利用微生物来表达异源蛋白已经成为获取酶最简单快速的方法。为获得高浓度和高质量的异源蛋白, 常用的方法是对基因序列进行密码子优化。传统的密码子优化策略主要基于密码子偏好性和 GC 含量, 忽略了翻译动力学和代谢水平等复杂多样的变化因素。文中从基因水平、转录水平、翻译水平、翻译后水平以及代谢水平等多方面考虑出发, 提供了一个较为全面的密码子优化策略, 主要包括密码子偏好性、密码子协调性、密码子敏感性、调整基因序列结构以及一些其他影响因素。同时对每种策略的内容、理论支持以及应用范围等方面作了全面的总结, 并将各策略的优缺点进行了系统的比较, 为异源蛋白表达提供了全方位、多层次、多选择的优化策略, 也为酶工业和生物药物等方面提供参考。

关键词: 密码子优化策略, 异源蛋白表达, 密码子偏好性, 密码子协调性, 密码子敏感性

Received: June 26, 2019; **Accepted:** September 3, 2019

Supported by: National Key Research and Development Program of China (No. 2017YFD0201400), National Natural Science Foundation of China (No. 21676026).

Corresponding author: Yi-Xin Huo. Tel: +86-10-68918158; E-mail: huoyixin@bit.edu.cn

国家重点研发计划 (No. 2017YFD0201400), 国家自然科学基金 (No. 21676026) 资助。

网络出版时间: 2019-10-31

网络出版地址: <http://kns.cnki.net/kcms/detail/11.1998.Q.20191031.1130.001.html>

Application of codon optimization strategy in heterologous protein expression

Yunpeng Yang¹, Xiaoyan Ma¹, and Yi-Xin Huo^{1,2}

¹ Key Laboratory of Molecular Medicine and Biotherapy, School of Life Sciences, Beijing Institute of Technology, Beijing 100081, China

² SIP-UCLA Institute for Technology Advancement (Suzhou), Suzhou 215123, Jiangsu, China

Abstract: Enzymes are widely used in medical and biopharmaceuticals. They can be used not only for various disease treatments, but also clinical diagnosis. The use of microorganisms to express heterologous proteins has become the easiest and fastest way to obtain enzymes. In order to obtain high concentration and high-quality heterologous proteins, a common method is codon optimization of gene sequences. The traditional codon optimization strategy is mainly based on codon bias and GC content, ignoring complex and varied factors such as translational dynamics and metabolic levels. We provide here comprehensive codon optimization strategy based on gene level, transcriptional level, translational level, post-translational level and metabolic level, mainly including codon bias, codon harmonization, codon sensitivity, adjustment of gene sequence structure and some other influencing factors. We also summarize the aspects of strategy content, theoretical support and application. Besides, the advantages and disadvantages of each strategy are also systematically compared, providing an all-round, multi-level and multi-selection optimization strategy for heterogeneous protein expression, and also providing references for the enzyme industry and biopharmaceuticals.

Keywords: codon optimization strategies, heterologous protein expression, codon bias, codon harmonization, codon sensitivity

酶是各种反应的生物催化剂,在医疗和生物制药等方面有着广泛的用途,可用于疾病诊断治疗、临床检测和养生保健等方面^[1-3]。例如,在大肠杆菌中生产的1-天冬酰胺酶可用于治疗白血病^[4],在枯草芽孢杆菌生产的纳豆激酶在血栓药物治疗上有着巨大前景^[5]。微生物酶还可以用于治疗由于遗传问题、酶缺乏或消化紊乱引起的疾病,同时,还可以用于诊断行业的糖尿病检测等诊断程序^[6]。目前在背景清晰的模式菌株里表达异源蛋白已成为获取微生物酶的最简单方法。但在进行异源蛋白表达时,密码子优化策略还存在很大的局限性。现有的优化策略主要从宿主本身出发,意在消除稀有密码子、调整GC含量和mRNA的稳定性等,但该策略仅局限于tRNA浓度、基因结构和mRNA稳定性等,忽略了翻译动力学、翻译后折叠及代谢水平等这些复杂多样的变化因素,导致一些异源蛋白的表达量仍很低^[7-8]。此外,目前常用的密码子优化策略的另一个弊端是更多地强调蛋白的表达数量,对蛋白的质量没有严格

的要求,常常出现蛋白翻译或折叠错误等^[7]。随着生物技术的飞速发展,人们对于生物的作用机制更加了解,也为密码子优化提供了一些新的策略,包括密码子的协调性^[9]、密码子的敏感性^[10]以及一些其他转录和翻译水平影响因素(图1A)。密码子协调性不同于密码子偏好性,该策略更加注重蛋白的质量,通过增加基因的耐受性来减少蛋白翻译后折叠错误的出现,从而维持了蛋白表达数量和质量的协调性。密码子敏感性策略则考虑到了氨酰化tRNA在胁迫条件下的波动性,为在底物不足或氨基酸饥饿条件下提供了一条新的密码子优化策略。

本文综合考虑蛋白表达的各个影响因素,将密码子的优化策略分为5大类,即密码子偏好性、密码子协调性、密码子敏感性、调整基因序列和其他影响因素。同时,总结了各个策略的内容、理论支持、优缺点以及应用范围等方面,为异源蛋白的表达提供一个较为全面的密码子优化策略,也为酶工业、生物药物的制备和精准医疗等

方面提供参考。

1 密码子优化策略

1.1 密码子偏好性

在遗传密码中, mRNA 上 3 个相邻的碱基组成一个密码子, 一个密码子可编码一种氨基酸。在生物体中有 61 种密码子、20 种氨基酸^[11], 这是由于遗传密码具有简并性, 同一氨基酸对应的多个密码子被称为同义密码子^[12-13]。但这些同义密码子在翻译的过程中使用的频率并非一致, 这种现象被称为“密码子偏好性”^[14-15]。密码子偏好性对蛋白的表达有着直接的影响, 能通过 tRNA 在翻译水平上改变蛋白的翻译速度^[16]。在同一生物中, 携带同一种氨基酸的不同 tRNA 称为同功受体 tRNA, tRNA 的浓度通常与其读取的密码子的频率正相关^[14-15]。在高度表达的基因中, 编码同种氨基酸的密码子中往往有一个密码子在频率上占主导地位, 并且该密码子通常由最丰富的 tRNA 的同功受体读取^[17]。研究表明, 这样有偏好性使用的机制可加快蛋白质合成, 减少氨基酸取代错误^[18]。因此, 密码子偏好性和同功受体 tRNA 浓度可能共同进化^[19-21], 并且与低水平表达的基因相比, 高水平表达的基因的选择压力更为显著^[22]。

密码子偏好性优化策略是目前最常用的密码子优化策略, 主要是将供体密码子与宿主基因组中具有最高频率的同义密码子进行替换, 利用宿主中最丰富的密码子来编码优化序列中的氨基酸^[23-24]。该理论假定 tRNA 和蛋白翻译呈直接线性关系, 且忽略了蛋白翻译的动力学相关影响^[7], 当宿主细胞内的密码子频率较高时, 相应的 tRNA 水平较高, 翻译速率较快, 更利于蛋白含量的表达 (图 1B)。密码子使用偏好性的分析在生物医药方面具有重要应用, 通过改变密码子分配从而产生更多蛋白质的策略已经广泛用于蛋白质药物和

核酸疗法的生物生产中。例如, 根据密码子的偏好性在大肠杆菌中成功地表达出了抗生素、胰岛素和疫苗等^[25]。

利用密码子偏好性策略来提高异源蛋白的表达水平已得到人们的广泛认可, 但也有部分实验结果表明, 利用密码子偏好性的优化策略来进行蛋白表达时, 非但没有增加蛋白的表达含量, 反而使蛋白的含量降低^[7-8]。密码子偏好性策略主要基于高频密码子对应的 tRNA 浓度较高, 可直接加快蛋白的翻译速率。但随后人体内翻译速率实验已证明 tRNA 浓度与翻译水平并不呈直接相关^[26-28], 甚至有实验证明, 相同的 tRNA 可以以显著的差异速率解码不同的密码子^[29]。因此, 当将供体密码子替换成宿主细胞内高频密码子时, 相应密码子的 tRNA 浓度较高, 但实际用于蛋白质合成的氨酰化 tRNA 浓度并不一定较高, 相应的合成蛋白的速率就会受到很大影响。同时, 当采用该方法策略获得异源性蛋白时, 高频密码子的使用消除了稀有密码子使用时造成的核糖体暂停效应, 导致部分蛋白折叠错误^[7,18,30], 最终形成包涵体, 影响蛋白的质量, 不利于微生物酶工业化生产 (表 1)。

1.2 密码子协调性

有研究表明, 在 mRNA 翻译较慢的区域将使用频率较低的密码子改为使用频率较高的同义密码子可能对酶活性产生有害影响。在氯霉素乙酰转移酶 (CAT) 基因上将 16 个稀有密码子替换成高频密码子, 可影响 mRNA 突变区域的核糖体运输并消除特定的翻译暂停, 导致蛋白质错误折叠水平增加, 比活率下降了 20%^[31]。相反, 将稀有密码子引入含有高频密码子使用的区域则会改变底物特异性。P-糖蛋白 (P-gp) 是多药耐药基因 1 (MDR1) 的表达产物, 也是 ATP 驱动的外排泵, 与 P-gp 底物药物的药物代谢动力学和癌症的多

药耐药性相关。在 P-gp 表达的过程中,将高频密码子替换成稀有密码子,tRNA 则成为蛋白表达的限制因素,影响共翻译折叠和 P-gp 插入细胞膜的时间,从而改变药物和抑制剂相互作用位点的结构^[32]。如果引入密码子的频率不当,可能会使蛋白的天然结构和功能发生变化,增加重组蛋白药物的免疫原性,降低治疗效果。因此,选择合适频率的密码子是成功表达有活性蛋白的关键因素。为了最大限度地模拟宿主内供体的天然翻译动力学,顺利表达异源蛋白,我们可以采用密码子协调性的优化策略。该策略主要是将供体密码子与宿主中使用频率最为相似的密码子进行替换,利用频率相似的密码子来编码蛋白。密码子协调性优化策略对酶工业生产过程的经济性具有重大影响,可显著地降低生产成本。例如,在小牛血清凝乳酶生产的过程中,通过协调密码子的使用,使凝乳酶原的量增加 70%,为商业化生产带来了极大效益^[33]。此外还有文献报道,将 mRNA 缓慢翻译的区域段用小于或等于天然供体密码子的同义密码进行编码,最终使异源蛋白的表达含量提高了 4-1 000 倍,可用于疫苗生产^[34]。同时,还可以利用密码子的随机分布^[33]来维持密码子的协调,该方法主要是基于整个基因组中密码子的频率分布或高度表达基因的翻译表来将供体密码子分配到宿主的每个密码子上。在这种情况下,密码子被随机分配,其概率由宿主密码子频率表给出^[35-38]。随后,可通过计算“用于替换供体密码子的可能性 (Likelihoods for selection to replace the donor codon, LSR-值)”来寻找最佳的密码子(图 1C)^[9]。目前,已有大量文献报道根据密码子协调性优化策略成功地提高了异源蛋白的表达量,为微生物酶生产的迅速发展提供了支持^[7, 9, 34, 39-42]。

在进行异源蛋白表达时,如果仅引入最高频率的密码子,则有可能造成相应氨酰化 tRNA 的浓度快速消耗,使底物成为蛋白翻译的限制因

素^[43],造成翻译的过早终止或者错误的氨基酸掺入,从而影响蛋白的含量和质量。利用密码子协调性,将频率最相似的密码子引入到宿主中,则能在一定水平上增加基因的耐受性,减少和避免出现同功受体 tRNA 缺乏的情况出现。且该策略更加强调模拟宿主内供体的天然翻译动力学,防止由于强制过表达时造成分子伴侣系统被破坏,从而使蛋白能最大限度地宿主细胞内高质量表达。在翻译中,同功受体 tRNA 的种类和氨酰化的总浓度决定了密码子的翻译速度^[29]。在密码子协调性策略中,最佳密码子并非最高频密码子,因此,其 tRNA 需要更长的时间才能到达核糖体的 A 位点,导致核糖体在同源密码子上的转运延迟^[44]。这样的延迟调节了共翻译折叠和分子伴侣相互作用的时间范围,特别是在多结构域蛋白质中,增加了成功折叠事件的机会,避免了包涵体的形成(表 1)。

1.3 密码子敏感性

根据氨酰化 tRNA 的可利用性在不同的生长和胁迫条件下有显著不同^[45-46],提出了“密码子敏感性”这一概念^[10]。密码子敏感性指的是在细胞内氨基酸缺乏的条件下,tRNA 与氨基酸结合能力的强弱。结合能力随氨基酸浓度降低呈缓慢降低的 tRNA 的敏感性较低,其相应的密码子被称为低敏感性密码子;结合能力随氨基酸浓度降低呈快速降低的 tRNA 的敏感性较高,其相应的密码子被称为高敏感性密码子。随后的实验直接证实了该理论预测,在细胞缺乏相对应氨基酸时,丝氨酸、亮氨酸、苏氨酸和精氨酸家族中同功受体 tRNA 选择性氨酰化^[30, 45, 47]。并根据 tRNA 可用性估算了密码子翻译速率,展示了 tRNA 在整个大肠杆菌基因组翻译过程中敏感性的动态变化^[48],为微生物酶的工业化增加了新元素。

理论上,蛋白的合成速度与细胞内 tRNA 和氨基酸的浓度相关^[49-50],但是,大肠杆菌 tRNA

的浓度在所有代谢条件下基本恒定^[51], 因此, 在底物缺乏的条件下, 即当核糖体翻译机制对某种氨基酸的需求大于其生物合成速率时, 氨基酸的供应成为了蛋白合成的限制性因素。由于密码子的频率与敏感性并不呈对应的关系^[10], 频率较高的密码子, 其敏感性并不一定低。在饥饿条件下表达异源蛋白时, 如果宿主中频率较高的密码子正好敏感性也较高, 那么, 采用密码子偏好性来进行密码子优化不仅不会使蛋白的表达量提高, 反而会使蛋白的表达量降低。在这种情况下, 如果采用的是低敏感性的密码子, 即使氨基酸的浓度降低, 该低敏感性密码子对应的 tRNA 相较其他的同功受体 tRNA 也有较高的氨酰化水平, 也能维持蛋白的持续表达 (图 1D)。反之, 高敏感性的密码子所对应的 tRNA 的氨酰化水平会随着氨基酸的浓度降低呈现较快的下降趋势, 即使细胞内其他同功受体 tRNA 仍维持高水平的氨酰化状态也不能被细胞加以利用, 会造成蛋白表达的暂缓甚至停止。

密码子敏感性的优化策略可广泛地应用于极端条件下或对于培养基有严格要求的蛋白表达, 也可将其用于对蛋白质量要求较高的结构生物学、抗体或生物药物的研究中。目前, 利用密码子敏感性的优化策略还未得到推广, 只有文献报道了关于大肠杆菌的密码子敏感性数值^[10], 其他物种尚未见报道 (表 1)。

1.4 调整基因序列

在异源蛋白表达中, 研究者关注更多的是转录和翻译过程, 但基因序列本身对蛋白的表达也有一定影响。常见的对基因序列优化方法有调整 GC 含量^[52-54]、避免碱基重复^[55-56]、消除限制酶识别位点^[57]、Chi-site 延伸重组热点^[57-58]等 (图 1E, 表 1)。有研究表明, 不同物种的 GC 含量显著不同, 基因序列中的 GC 含量能影响基因的表

达和调控^[59]。GC 含量过高 (大于 70%) 会造成 RNA 二级结构稳定性增加, 减慢或暂停翻译; GC 含量过低 (小于 30%) 则会减慢转录延伸, 不利于蛋白的表达。GC 含量的调整可通过相关的软件进行分析。例如, 在大肠杆菌中, 通过调整 GC 含量实现了肽脱甲酰酶的表达, 成为了治疗癌症的潜在药物^[60]。此外, 密码子进行优化时还需避免与位于开放阅读框中可能干扰 mRNA 加工和翻译功能的重要 RNA 基序相似, 如 Shine-Dalgarno 序列^[28], 这可能会造成核糖体暂停^[61]。在真核生物中则需要考虑到 CpG 的含量^[62]和 TATA 盒^[63]等因素, 这对转录的启动有着重要的影响。

1.5 其他因素

影响异源蛋白表达的因素复杂多样, 除了上述因素外, 还包括一些其他的转录水平和翻译水平的影响因素。如 mRNA 二级结构的稳定性^[64]、消除串联稀有密码子^[65]、避免密码子重复^[66]、调整核糖体结合位点^[67]和注意起始终止密码的环境^[68-69]等 (图 1F, 表 1)。mRNA 的折叠能量对翻译效率具有显著影响, 特别是在起始密码子附近, 因为更稳定的 RNA 二级结构在翻译起始前需要更多的能量展开。在大肠杆菌中, 非最高频密码子与降低的折叠能量相关, 因为它们与其最高频密码子相比更倾向于富含 AT, 因此当非最高频密码子位于编码 DNA 序列的 5'末端附近时, 翻译效率提高^[70]。这表明密码子最优性必须与松散的 RNA 二级结构相平衡, 以便最大限度地提高翻译效率^[59, 71-72], 此外, 胞内直接翻译速率的测量^[73-74]和核糖体起始和延伸的计算机模拟^[75]证明了翻译起始通常是蛋白质合成中的限速步骤。在翻译起始过程中, 使用低频密码子不仅会减慢核糖体翻译速率, 甚至还可以通过降低核糖体从 mRNA 起始位点移出的速率来影响同一转录物上其他核

糖体的翻译启动^[76-77]。在细菌中核糖核酸酶 E 位点也可以影响 mRNA 结构的稳定性^[78]。而在真核

生物中还需考虑潜在的 polyA 位点，防止提前终止翻译^[79]。

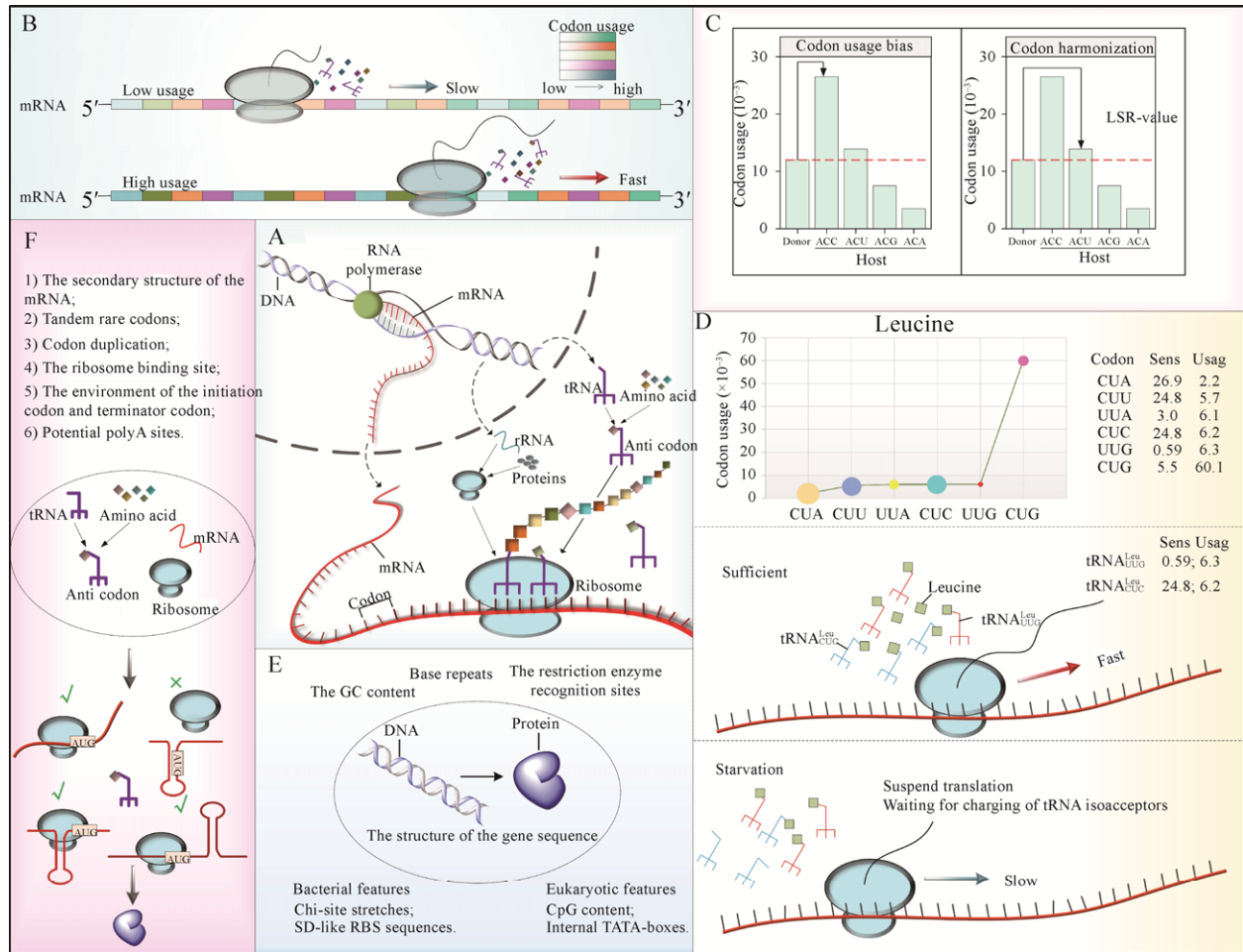


图 1 密码子优化策略示意图

Fig. 1 Schematic diagram of the codon optimization strategies. (A) Various factors affecting protein expression. Processes such as gene sequence, transcription, translation and post-translational processing affect protein expression. (B) Codon bias strategy. The frequency of codons is related to protein translation. Low-frequency codons have low tRNA concentration, and protein translation can be performed lower; in contrast, high-frequency codons have high tRNA concentration, and protein translation can be performed faster. (C) Comparison of codon bias strategy and codon harmonization strategy. Codon bias is the replacement of the codon in the donor with the most frequent synonymous codon in the host, while the codon harmonization is the replacement of the codon in the donor with the most similar synonym in the host. The best codon can be found by calculating the LSR value through the website. (D) Codon sensitivity strategy. The frequency and sensitivity of codons are not linear. In the case of leucine, the higher the codon frequency, the lower the sensitivity (bubble size represents sensitivity). Under starvation conditions, low-sensitivity codons are more "stable" and can continuously express proteins. (E) The effects of gene sequences on protein expression. The GC content, base repeats, the restriction enzyme recognition sites, Chi-site stretches, SD-like RBS sequences, CpG content and internal TATA-boxes can affect the expression of the protein. (F) Other factors affecting protein expression. Such as the secondary structure of the mRNA, rare codons, codon duplication, the ribosome binding site, the environment of the initiation codon or terminator codon and potential polyA sites, etc.

表 1 密码子优化策略比较

Table 1 Comparison of codon optimization strategies

Strategy	Codon bias	Codon harmonization	Codon sensitivity	Adjust the structure of the gene sequence	Other influencing factors
Content	Replace the codon in the donor with the highest frequency synonymous codon in the host	1) Replace the codon in the donor with the most similar frequency synonymous codon in the host 2) Append the donor codon to each host codon based on the codon frequency	Replace the codon in the donor with a low-sensitivity synonym in the host	1) Adjust the GC content 2) Avoid base repeats 3) Avoid the restriction enzyme recognition sites 4) Chi-site stretches 5) SD-like RBS sequences 6) CpG content 7) Internal TATA-boxes	1) The secondary structure of the mRNA 2) Avoid tandem rare codons 3) Avoid codon duplication 4) Adjusting the ribosome binding site 5) Note the environment of the initiation codon and terminator codon 6) Potential polyA sites
Theoretical support	The highest frequency codon has the highest tRNA concentration	1) Codons with similar frequency can maximize the natural translation dynamics 2) Add donor codons to each host codon to increase gene tolerance	Low-sensitivity codons maintain a high level of aminoacylation under conditions of amino acid starvation	Base sequence can affect protein expression and regulation	Transcriptional processes and translation processes affect protein expression and regulation
Advantage	High protein expression in quantity	1) Increase gene tolerance 2) Increase protein quality	Provides a new optimization strategy for protein expression under extreme conditions or starvation conditions	Ignore some factors affecting protein expression at the genetic level	Ignore some factors affecting protein expression at the transcriptional level and translational level
Disadvantage	Cause protein folding error	Not pay attention to protein expression in quantity	1) Only <i>E. coli</i> codon sensitivity is known; 2) This strategy has not been widely adopted	There are still many uncertain factors	Many mechanisms have not yet been found
Application	Suitable for most protein expression	Suitable for most protein expression	Stress conditions or other extreme conditions	Make appropriate adjustments based on specific gene sequences	Make appropriate adjustments based on specific gene sequences

2 总结与展望

密码子优化是实现异源蛋白高效表达的关键技术手段。密码子偏好性策略适用性较广,能用于大多数异源蛋白表达,但由于蛋白表达量过高易造成折叠错误,形成包涵体。此时,可采用密码子协调性优化策略,将供体中密码子与宿主中频率最相似的同义密码子进行替换或将供体密码子根据密码子频率分配到宿主每个密码子上,避免了因密码子频率过高而引起的翻译终止或折叠错误等现象出现。密码子协调性策略在异源蛋白表达中的使用可提高功能性蛋白表达的可靠性,对结构生物学和生物技术有着重要的影响。此外,该策略为大肠杆菌等模式菌株进行异源蛋白表达提供了极好的发展前景,也为许多潜在的疫苗和生物药物的开发和制备提供了大力支持。密码子敏感性优化策略则考虑到了氨酰化 tRNA 的可利用性,将高敏感性的密码子替换成低敏感性的密码子,使蛋白在体内氨基酸缺乏的条件下也能维持稳定表达。该策略在蛋白质工程、酶工程、代谢工程以及合成生物学中有着广泛应用,可用于在胁迫条件或其他极端条件下异源蛋白的表达,也可用于氨基酸衍生物的发酵生产。但该策略还未得到广泛应用,因其研究还未完善,目前只报道了关于大肠杆菌的密码子敏感性数值,其他生物密码子敏感性尚在研究中。密码子偏好性、密码子协调性和密码子敏感性优化策略主要是在不同培养条件下,调整细胞体内 mRNA、tRNA、氨酰化 tRNA、氨基酸含量以及蛋白表达的关系,最大限度地模拟天然蛋白的翻译过程。因此,上述 3 种策略不受物种限制,可广泛地应用于大肠杆菌和酵母表达系统。基因结构和其他转录翻译水平的影响因素复杂多样,原核生物和真核生物的转录翻译机制不同,在密码子优化中考虑的因素也不同,可根据宿主和异源蛋白的基因序列进行适当调整,从而提高蛋白的表达水平。

异源蛋白的表达已成为酶工业和疫苗制造过程的重要内容,密码子的优化策略则是其中的关键因素。随着科技的飞速发展,生物的生命活动解析会更为清晰,密码子的优化策略也将更加丰富多样,从而为酶工业生产和生物药物制备带来新策略,为人类生命健康带来新期望。

REFERENCES

- [1] Singh R, Kumar M, Mittal A, et al. Microbial enzymes: industrial progress in 21st century. *3 Biotech*, 2016, 6(2): 174.
- [2] Gurung N, Ray S, Bose S, et al. A broader view: microbial enzymes and their relevance in industries, medicine, and beyond. *Biomed Res Int*, 2013, 2013: 329121.
- [3] Jin C. Preface for special issue on enzyme engineering (2018). *Chin J Biotechnol*, 2018, 34(7): 1021–1023 (in Chinese).
金城. 2018 酶工程专刊序言. *生物工程学报*, 2018, 34(7): 1021–1023.
- [4] Jain R, Zaidi DK, Verma Y, et al. L-Asparaginase: A promising enzyme for treatment of acute lymphoblastic leukemia. *People's J Sci Res*, 2012, 5(1): 29–35.
- [5] Cho YH, Song JY, Kim KM, et al. Production of nattokinase by batch and fed-batch culture of *Bacillus subtilis*. *New Biotechnol*, 2010, 27(4): 341–346.
- [6] Vellard M. The enzyme as drug: application of enzymes as pharmaceuticals. *Curr Opin Biotechnol*, 2003, 14(4): 444–450.
- [7] Buhr F, Jha S, Thommen M, et al. Synonymous codons direct cotranslational folding toward different protein conformations. *Mol Cell*, 2016, 61(3): 341–351.
- [8] Zhu D, Cai G, Wu D, et al. Comparison of two codon optimization strategies enhancing recombinant *Sus scrofa* lysozyme production in *Pichia pastoris*. *Cell Mol Biol (Noisy-le-grand)*, 2015, 61(2): 43–49.
- [9] Rehbein P, Berz J, Kreisel P, et al. “CodonWizard”—An intuitive software tool with graphical user interface for customizable codon optimization in protein expression efforts. *Prot Expres Purif*, 2019, 160: 84–93.

- [10] Elf J, Nilsson D, Tenson T, et al. Selective charging of tRNA isoacceptors explains patterns of codon usage. *Science*, 2003, 300(5626): 1718–1722.
- [11] Crick FHC, Barnett L, Brenner S, et al. General nature of the genetic code for proteins. *Nature*, 1961, 192(4809): 1227–1232.
- [12] Lagerkvist U. “Two out of three”: an alternative method for codon reading. *Proc Natl Acad Sci USA*, 1978, 75(4): 1759–1762.
- [13] Ikemura T. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. *J Mol Biol*, 1981, 146(1): 1–21.
- [14] Sharp PM, Cowe E, Higgins DG, et al. Codon usage patterns in *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Homo sapiens*: a review of the considerable within-species diversity. *Nucleic Acids Res*, 1988, 16(17): 8207–8211.
- [15] Plotkin J B, Kudla G. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet*, 2011, 12(1): 32–42.
- [16] Sabi R, Tuller T. Modelling the efficiency of codon-tRNA interactions based on codon usage bias. *DNA Res*, 2014, 21(5): 511–526.
- [17] 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.
- [18] Hanson G, Collier J. Codon optimality, bias and usage in translation and mRNA decay. *Nat Rev Mol Cell Biol*, 2017, 19(1): 20–30.
- [19] Sharp PM, Emery LR, Zeng K. Forces that influence the evolution of codon bias. *Philos Trans Roy Soc B Biol Sci*, 2010, 365(1544): 1203–1212.
- [20] Yannai A, Katz S, Hershberg R. The codon usage of lowly expressed genes is subject to natural selection. *Genome Biol Evol*, 2018, 10(5): 1237–1246.
- [21] Hershberg R, Petrov DA. Selection on codon bias. *Ann Rev Genet*, 2008, 42(1): 287–299.
- [22] Ehrenberg M, Kurland CG. Costs of accuracy determined by a maximal growth rate constraint. *Quart Rev Biophys*, 1984, 17(1): 45–85.
- [23] Quax TEF, Claassens NJ, Söll D, et al. Codon bias as a means to fine-tune gene expression. *Mol Cell*, 2015, 59(2): 149–161.
- [24] Villalobos A, Ness JE, Gustafsson C, et al. Gene Designer: a synthetic biology tool for constructing artificial DNA segments. *BMC Bioinform*, 2006, 7: 285.
- [25] Kane JF. Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Curr Opin Biotechnol*, 1995, 6(5): 494–500.
- [26] Curran JF, Yarus M. Rates of aminoacyl-tRNA selection at 29 sense codons *in vivo*. *J Mol Biol*, 1989, 209(1): 65–77.
- [27] Bonekamp F, Dalbøge H, Christensen T, et al. Translation rates of individual codons are not correlated with tRNA abundances or with frequencies of utilization in *Escherichia coli*. *J Bacteriol*, 1989, 171(11): 5812–5816.
- [28] Li GW, Oh E, Weissman JS. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature*, 2012, 484(7395): 538–541.
- [29] Sørensen MA, Pedersen S. Absolute *in vivo* translation rates of individual codons in *Escherichia coli*: the two glutamic acid codons GAA and GAG are translated with a threefold difference in rate. *J Mol Biol*, 1991, 222(2): 265–280.
- [30] Gong M, Gong F, Yanofsky C. Overexpression of *tnaC* of *Escherichia coli* inhibits growth by depleting tRNA_{2^{Pro}} availability. *J Bacteriol*, 2006, 188(5): 1892–1898.
- [31] 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.
- [32] Kimchi-Sarfaty C, Oh JM, Kim IW, et al. A “Silent” polymorphism in the *MDR* gene changes substrate specificity. *Science*, 2007, 315(5811): 525–528.
- [33] Menzella HG. Comparison of two codon optimization strategies to enhance recombinant protein production in *Escherichia coli*. *Microb Cell Fact*, 2011, 10: 15.
- [34] 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.
- [35] Welch M, Govindarajan S, Ness JE, et al. Design parameters to control synthetic gene expression in

- Escherichia coli*. PLoS One, 2009, 4(9): e7002.
- [36] Kodumal SJ, Patel KG, Reid R, et al. Total synthesis of long DNA sequences: Synthesis of a contiguous 32-kb polyketide synthase gene cluster. Proc Natl Acad Sci USA, 2004, 101(44): 15573–15578.
- [37] Wang XX, Li XJ, Zhang ZL, et al. Codon optimization enhances secretory expression of *Pseudomonas aeruginosa* exotoxin A in *E. coli*. Prot Expres Purif, 2010, 72(1): 101–106.
- [38] Menzella HG, Reid R, Carney JR, et al. Combinatorial polyketide biosynthesis by *de novo* design and rearrangement of modular polyketide synthase genes. Nat Biotechnol, 2005, 23(9): 1171–1176.
- [39] 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.
- [40] Angov E, Legler PM, Mease RM. Adjustment of codon usage frequencies by codon harmonization improves protein expression and folding//Evans TC Jr, Xu MQ, Eds. Heterologous Gene Expression in *E. coli*: Methods and Protocols. Totowa, NJ: Humana Press, 2011: 1–13.
- [41] Mignon C, Mariano N, Stadthagen G, et al. Codon harmonization-going beyond the speed limit for protein expression. FEBS Lett, 2018, 592(9): 1554–1564.
- [42] Asam C, Roulias A, Parigianni MA, et al. Harmonization of the genetic code effectively enhances the recombinant production of the major birch pollen allergen bet v 1. Int Arch Allergy Immunol, 2018, 177(2): 116–122.
- [43] Fuglsang A. Codon optimizer: a freeware tool for codon optimization. Prot Expres Purif, 2003, 31(2): 247–249.
- [44] Chu D, Barnes DJ, von der Haar T. The role of tRNA and ribosome competition in coupling the expression of different mRNAs in *Saccharomyces cerevisiae*. Nucleic Acids Res, 2011, 39(15): 6705–6714.
- [45] Dittmar KA, Sørensen MA, Elf J, et al. Selective charging of tRNA isoacceptors induced by amino-acid starvation. EMBO Rep, 2005, 6(2): 151–157.
- [46] Dong HJ, Nilsson L, Kurland CG. Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. J Mol Biol, 1996, 260(5): 649–663.
- [47] Lindsley D, Bonthuis P, Gallant J, et al. Ribosome bypassing at serine codons as a test of the model of selective transfer RNA charging. EMBO Rep, 2005, 6(2): 147–150.
- [48] Wohlgemuth SE, Gorochofski TE, Roubos JA. Translational sensitivity of the *Escherichia coli* genome to fluctuating tRNA availability. Nucleic Acids Res, 2013, 41(17): 8021–8033.
- [49] Ingolia NT. Ribosome profiling: new views of translation, from single codons to genome scale. Nat Rev Genet, 2014, 15(3): 205–213.
- [50] Mazumder GA, Uddin A, Chakraborty S. Comparative analysis of codon usage pattern and its influencing factors in *Schistosoma japonicum* and *Ascaris suum*. Acta Parasitol, 2017, 62(4): 748–761.
- [51] Yamao F, Andachi Y, Muto A, et al. Levels of tRNAs in bacterial cells as affected by amino acid usage in proteins. Nucleic Acids Res, 1991, 19(22): 6119–6122.
- [52] Kiktev DA, Sheng ZW, Lobachev KS, et al. GC content elevates mutation and recombination rates in the yeast *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA, 2018, 115(30): E7109–E7118.
- [53] Newman ZR, Young JM, Ingolia NT, et al. Differences in codon bias and GC content contribute to the balanced expression of TLR7 and TLR9. Proc Natl Acad Sci USA, 2016, 113(10): E1362–E1371.
- [54] Kudla G, Lipinski L, Caffin F, et al. High guanine and cytosine content increases mRNA levels in mammalian cells. PLoS Biol, 2006, 4(6): e180.
- [55] Gustafsson C, Govindarajan S, Minshull J. Codon bias and heterologous protein expression. Trends Biotechnol, 2004, 22(7): 346–353.
- [56] Chamary JV, Parmley JL, Hurst LD. Hearing silence: non-neutral evolution at synonymous sites in mammals. Nat Rev Genet, 2006, 7(2): 98–108.
- [57] Parret AHA, Besir H, Meijers R. Critical reflections on synthetic gene design for recombinant protein expression. Curr Opin Struct Biol, 2016, 38: 155–162.
- [58] Taylor AF, Amundsen SK, Smith GR. Unexpected DNA context-dependence identifies a new

- determinant of Chi recombination hotspots. *Nucleic Acids Res*, 2016, 44(17): 8216–8228.
- [59] Kudla G, Murray AW, Tollervey D, et al. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science*, 2009, 324(5924): 255–258.
- [60] Han JH, Choi YS, Kim WJ, et al. Codon optimization enhances protein expression of human peptide deformylase in *E. coli*. *Prot Expres Purif*, 2010, 70(2): 224–230.
- [61] Mohammad F, Woolstenhulme CJ, Green R, et al. Clarifying the translational pausing landscape in bacteria by ribosome profiling. *Cell Rep*, 2016, 14(4): 686–694.
- [62] Bauer AP, Leikam D, Krinner S, et al. The impact of intragenic CpG content on gene expression. *Nucleic Acids Res*, 2010, 38(12): 3891–3908.
- [63] Jonkers I, Kwak H, Lis JT. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *eLife* 2014, 3: e02407.
- [64] Kertesz M, Wan Y, Mazor E, et al. Genome-wide measurement of RNA secondary structure in yeast. *Nature*, 2010, 467(7311): 103–107.
- [65] Clarke IV TF, Clark PL. Increased incidence of rare codon clusters at 5' and 3' gene termini: implications for function. *BMC Genom*, 2010, 11: 118.
- [66] Gustafsson C, Minshull J, Govindarajan S, et al. Engineering genes for predictable protein expression. *Prot Expres Purif*, 2012, 83(1): 37–46.
- [67] Mackie GA. RNase E: at the interface of bacterial RNA processing and decay. *Nat Rev Microbiol*, 2012, 11(1): 45–47.
- [68] Shell SS, Wang J, Lapierre P, et al. Leaderless transcripts and small proteins are common features of the mycobacterial translational landscape. *PLoS Genet*, 2015, 11(11): e1005641.
- [69] Srivastava A, Gogoi P, Deka B, et al. *In silico* analysis of 5'-UTRs highlights the prevalence of Shine–Dalgarno and leaderless-dependent mechanisms of translation initiation in bacteria and archaea, respectively. *J Theor Biol*, 2016, 402: 54–61.
- [70] Goodman DB, Church GM, Kosuri S. Causes and effects of N-terminal codon bias in bacterial genes. *Science*, 2013, 342(6157): 475–479.
- [71] Gu WJ, Zhou T, Wilke CO. A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. *PLoS Comput Biol*, 2010, 6(2): e1000664.
- [72] Tsao D, Shabalina SA, Gauthier J, et al. Disruptive mRNA folding increases translational efficiency of catechol-*O*-methyltransferase variant. *Nucleic Acids Res*, 2011, 39(14): 6201–6212.
- [73] Morisaki T, Lyon K, Deluca KF, et al. Real-time quantification of single RNA translation dynamics in living cells. *Science*, 2016, 352(6292): 1425–1429.
- [74] Wu B, Eliscovich C, Yoon YJ, et al. Translation dynamics of single mRNAs in live cells and neurons. *Science*, 2016, 352(6292): 1430–1435.
- [75] Shah P, Ding Y, Niemczyk M, et al. Rate-limiting steps in yeast protein translation. *Cell*, 2013, 153(7): 1589–1601.
- [76] 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.
- [77] Potapov I, Mäkelä J, Yli-Harja O, et al. Effects of codon sequence on the dynamics of genetic networks. *J Theor Biol*, 2012, 315: 17–25.
- [78] Del Campo C, Bartholomäus A, Fedyunin I, et al. Secondary structure across the bacterial transcriptome reveals versatile roles in mRNA regulation and function. *PLoS Genet*, 2015, 11(10): e1005613.
- [79] Proudfoot NJ. Ending the message: poly(A) signals then and now. *Genes Dev*, 2011, 25(17): 1770–1782.

(本文责编 郝丽芳)