



密码子中的密码: 密码子偏好性与基因表达的精细调控

任桂萍^{1,2}, 董璿莹^{1,2}, 党云琨^{1,2*}

1. 云南大学省部共建云南生物资源保护与利用国家重点实验室, 昆明 650091;

2. 云南大学生命科学学院, 生命科学研究中心, 昆明 650500

* 联系人, E-mail: ykdang@ynu.edu.cn

收稿日期: 2019-05-04; 接受日期: 2019-06-05; 网络版发表日期: 2019-07-08

国家自然科学基金(批准号: 31871254)、云南大学省部共建云南生物资源保护与利用国家重点实验室开放课题和华东理工大学生物反应器工程国家重点实验室开放课题资助

摘要 同义突变由于不改变编码蛋白质的氨基酸序列, 常被认为是“沉默”突变. 实际上, 同义密码子的选择在进化尺度上是受到限制的, 从而致使同义密码子的使用频率存在差异, 称为密码子偏好性. 密码子偏好性在转录、转录后加工、mRNA稳定性、翻译起始、延伸、蛋白折叠等方面都起着精细调节的作用. 因此, 同义突变在很多情况下可导致癌症等各类疾病的发生. 本综述在分子机制层面简述了近年来关于密码子偏好性对翻译和转录过程调节作用的进展, 以及对于基础研究及医学方面的意义.

关键词 密码子偏好性, 致病同义突变, 转录, 翻译

在组成蛋白质的20种基本氨基酸中, 有18种由2~6个同义密码子所编码. 根据经典的蛋白质折叠热力学理论, 蛋白质的高级结构由蛋白质一级结构(蛋白质的序列)决定^[1]. 由于同义密码子之间的切换不改变蛋白质一级序列, 因此过去常把同义突变认为是“沉默”突变. 然而生物体在演化和适应的过程中, 同义密码子的选择并非是“中性”的(即随机选择). 对于某个物种的基因组而言, 一些同义密码子会频繁出现, 称为优化密码子; 一些密码子则很少出现甚至不出现, 称为非优化密码子或稀有密码子, 这一现象称为密码子偏好性(codon usage bias)^[2]. 目前所有已知物种的基因组都存在密码子偏好性, 尽管这种偏好性各有不同^[3].

密码子偏好性作为基因和基因组的一个静态特征, 是一个生物在长期演化过程中适应和选择的结果. 早期研究发现, 密码子的出现频率与对应的tRNA基因数量呈现显著的正相关性^[4,5], 因此认为胞内各tRNA表达差异与蛋白翻译机器之间的协同进化是引起密码子偏好性的主要原因之一^[6,7], 并由此形成了两种不矛盾的假说. 翻译准确性假说认为在演化过程中, 选择优化密码子是由于其能够降低翻译过程中被非匹配tRNA识别的概率, 从而更好地避免错误蛋白的产生^[8]; 翻译效率性假说则认为, 选择优化密码子是其需要解码的时间更短, 从而增加核糖体的使用效率^[9]. 近年来的研究还发现, 转录、mRNA稳定性及细胞减数分裂

引用格式: 任桂萍, 董璿莹, 党云琨. 密码子中的密码: 密码子偏好性与基因表达的精细调控. 中国科学: 生命科学, 2019, 49: 839–847
Ren G P, Dong Y Y, Dang Y K. Codon codes: Codon usage bias influences many levels of gene expression (in Chinese). Sci Sin Vitae, 2019, 49: 839–847, doi: 10.1360/SSV-2019-0103

过程中的同源重组等过程也与密码子偏好性存在着协同进化的关系^[10-12]。

密码子偏好性预示着同义突变至少在某些情况下并不是“沉默”的。事实上, 同义突变经常会导致蛋白质产量和功能的改变, 因此在实验室和工业生产中, 通过修改异源基因的密码子偏好性来适配表达载体(如大肠杆菌)已成为提高异源蛋白表达的一种常见策略^[13,14]。此外, 在医学领域, 同义突变导致的疾病也早已被报道, 如儿茶酚-*o*-甲基转移酶(COMT)基因中的3个常见同义单碱基多态位点(synonymous SNP, sSNP)可以影响对疼痛敏感性的差异^[15], 而耐药基因MDR1(multidrug resistance 1)中的1个sSNP能够改变底物和抑制剂的结合活性^[16]。早在2011年, Sauna和Kimchi-Sarfaty^[17]就系统性地整理出大约50种由同义突变导致的各类疾病。近年来, 随着技术的进步, 尤其是全基因组关联性分析(genome-wide association study, GWAS)的广泛应用, 更多的sSNP被发现致病或与疾病紧密相关。截止2019年, 美国NCBI网站的dbSNP数据库中收集的致病同义突变大约有1700个(<https://www.ncbi.nlm.nih.gov/snp>)。尽管这些数据中大部分还需要进一步的实验验证, 但研究显示, 同义突变的致病概率与错义突变几乎相当^[18]。由此可见, 研究密码子偏好性无论在基础理论、生产实践和医疗卫生方面皆有重要意义。过去的30多年中, 人们已经从遗传学、分子生物学、生物化学、系统生物学和病理学等方面对密码子偏好性和同义突变开展了广泛研究, 篇幅有限, 本文仅对这些研究进展做一个简要的概况。

1 密码子偏好性对翻译过程的影响

近年来随着深度测序技术的普及, 测量组织甚至单细胞内的总体mRNA含量已不再是障碍。然而, 以mRNA水平来推测对应的蛋白丰度存在很大误差。在哺乳动物中, 以小鼠成纤维细胞为例, 41%~54%的蛋白量差异是由翻译过程造成的, 仅有约40%的差异由mRNA量决定, 而蛋白降解在其中的影响很微弱^[19,20]。在丝状真菌粗糙脉孢菌中, 尽管mRNA水平与蛋白质质量存在着很强的相关性($r=0.74$), 这个结果仍然意味着有约30%的蛋白表达变化是在翻译水平上受到调控的^[21]。因此, 加深在翻译水平上的认知是未来研究的重点。密码子偏好性由于与tRNA丰度存在强烈的相关

性, 预示着其在翻译过程的各个方面中都可能起重要的调控作用。

1.1 翻译起始

对于内源基因, 翻译起始通常是整个蛋白质翻译过程的限速步骤^[22]。在原核生物中, 核糖体通过识别mRNA上的Shine-Dalgarno序列来协助核糖体定位起始密码子。Kudla等人^[23]发现, 含不同程度稀有密码子的*gfp*基因最终产生的蛋白量之间存在约250倍差异, 而这种差异主要源于起始密码子AUG附近的mRNA二级结构的稳定性。Goodman等人^[24]进一步研究发现, mRNA 5'端核糖体结合区域(RBS)附近的密码子偏好性选择会改变RNA局部二级结构的稳定性, 从而影响核糖体识别mRNA的翻译起始位点(图1A)。

真核生物的翻译起始主要依赖翻译蛋白复合体识别mRNA的5'端帽子结构, 然后向3'端扫描来识别起始密码子^[25], 因此翻译起始点附近较强的二级结构可阻止蛋白质的翻译起始^[21,26,27]。另外, 翻译的初期延伸阶段对蛋白表达也有很大影响。在果蝇(*Drosophila melanogaster*)、线虫(*Caenorhabditis elegans*)、酵母(*Saccharomyces cerevisiae*)等物种的基因(尤其是高表达基因)中, 可读框(open reading frame, ORF)的N端常富含稀有密码子, 除了避免在起始密码子附近形成强二级结构外, 可能还为了避免延伸过快而出现多个核糖体在mRNA上阻塞, 从而引起蛋白异常折叠或翻译提前终止^[28]。总之, 较慢的初期延伸阶段可以促进翻译效率^[27,29,30]。在酵母中, 替换外源*luciferase*基因的起始密码子附近区域内两个稀有密码子为优化密码子, 可显著降低荧光素酶活性^[31]。

1.2 翻译延伸速度

氨酰-tRNA进入核糖体A位点与密码子准确匹配的过程是翻译延伸过程的限速步骤^[32], 由于密码子偏好性与tRNA丰度存在着紧密的相关性, 因此一个合理的假设是核糖体在解码稀有密码子时总体需要更长的时间。确实, 在大肠杆菌中提高三个匹配稀有密码子的tRNA表达水平可加快翻译延伸, 减少停滞^[33]。然而, 核糖体在mRNA上解码的过程还可能受到如mRNA二级结构、tRNA的局部重复利用、新生肽段与核糖体出口隧道之间相互作用等影响^[26,34,35]。叠加上述因素, 传统生物化学手段考察单个或几个基因的翻译延伸速度

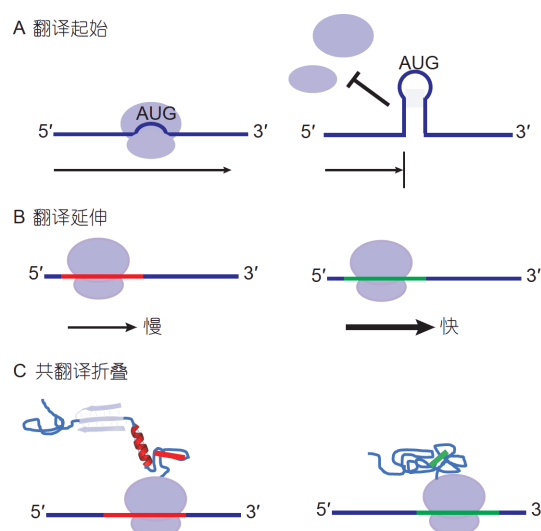


图 1 密码子偏好性调控翻译的作用机制。A: mRNA 二级结构对翻译起始的调控。翻译起始松散的 mRNA 二级结构有利于核糖体对 AUG 的识别, 翻译正常进行(左); 稳定的 mRNA 结构阻止核糖体继续扫描, 无法识别 AUG, 翻译不能进行(右); B: 密码子偏好性对翻译延伸速度的影响。核糖体在富含稀有密码子的区域上移动速度较慢(左); 在富含优化密码子的区域移动速度较快(右); C: 密码子偏好性改变蛋白质的结构和构象。稀有密码子有利于蛋白质正确折叠, 优化密码子不利于蛋白质正确折叠。在 B 和 C 中, 红色线条表示稀有密码子区域, 绿色线条表示优化密码子区域

Figure 1 Mechanism of codon usage bias in fine-tuning translation. A: Translation initiation is regulated by mRNA secondary structure. A loose secondary structure surrounding the start codon (AUG) facilitates ribosome anchorage to AUG and allows translation to commence (left panel). However, a strong secondary structure hinders ribosomal scanning and disrupts translation initiation (right panel). B: Translation elongation is regulated by codon usage bias. Ribosomes generally move slowly on mRNA regions rich in rare codons (left panel) and fast on regions rich in optimal codons (right panel). C: Local peptide folding and structure is regulated by codon usage bias. Regions rich in rare codons can facilitate local folding of the proteins, whereas regions rich in optimal codons may result in aberrant protein structures. In B and C, red and green lines indicate mRNA regions rich in rare and optimal codons, respectively

很难在总体上准确反映核糖体解码不同密码子的速率。

核糖体印迹测序(ribosome profiling)技术是目前唯一可以在全转录组水平上考量核糖体解码状态的方法。其原理是通过分析受核糖体保护的 mRNA 片段的序列, 从而推知核糖体 A 位点在 mRNA 对应密码子上的停留时间。然而早期的结果并没有观察到翻译速率与密码子偏好性之间的显著关联^[36]。后期的研究发现, 这是由于放线菌素(cycloheximide, 常用的翻译延伸抑制剂)渗入出芽酵母的核糖体 E 位却不能立刻冻结核糖体的移动, 从而造成实验的杂音干扰^[37~39]。本实验室

随后在粗糙脉孢菌中开展了核糖体印迹测序, 首次在转录组的水平上清晰地展示了核糖体解码速率与稀有密码子之间的显著负相关性, 更重要的是这种关联也存在于每一组同义密码子之间^[40]。另外, 在体外翻译实验中发现, 含全优化密码子的荧光素酶 mRNA 翻译速率可比野生型(即野生型)快 1.5 min, 进一步证实了长期以来对于稀有密码子解码效率低的假说(图 1B)^[40]。在哺乳动物细胞中, Yan 等人^[41]通过将 mRNA 固定于细胞膜表面构建了一个单分子实时监测系统。由 mRNA 编码区的 N 端携带一个荧光标签, 通过实时观测翻译出的荧光标签沿着 mRNA 运行的速度, 可以准确测量核糖体解码 C 端密码子的速率, 结果显示, 优化密码子的平均解码效率比稀有密码子快 58%。

1.3 翻译准确性

与 RNA 转录相比, 蛋白质的翻译过程不仅耗能而且错误率很高, 估计每翻译 1000~10000 个密码子即会出现一次错误^[42]。根据翻译准确论假说, 正在进行翻译的核糖体 A 位点上, 近似匹配或非匹配 tRNA 会与匹配的 tRNA 竞争结合密码子。优化密码子对应的匹配 tRNA 丰度高, 因此发生错误翻译的概率较小^[43]。当遇到稀有密码子时, 匹配 tRNA 的浓度较低, 因而导致核糖体延宕。如果核糖体最终不能顺利找到匹配的 tRNA, 则可发生三种情况: (i) 近似或非匹配 tRNA 进入 A 位点发生替代性翻译^[44], 在酵母中这种错误配对发生的概率约为 10^{-5} , 枯草杆菌中可达 10^{-2} ^[45,46]; (ii) 在 mRNA 的特定序列上发生 -1 或 +1 位移码, 如酵母中的 *Ty1* 基因^[47]; (iii) 翻译提前终止, 如酵母中遇到缺乏匹配 tRNA 的密码子(即稀有密码子)可导致激活 no-go 降解(no-go decay, NGD)途径提前终止翻译并降解 mRNA^[26]。

1.4 蛋白质折叠

稀有密码子与优化密码子在基因内通常不是混合在一起, 而是倾向于富集成簇且与蛋白结构功能存在关联。例如, 较长的蛋白质一般存在较多缓慢翻译的片段, 且这些缓慢翻译的片段簇主要位于保守结构域的下游^[48]; 形成功能域的部分倾向使用优化密码子, 而结构松散区域趋于使用稀有密码子^[49]。这些现象暗示, 稀有密码子可能有利于在翻译过程中调节结构域折叠的精确度。

最初的实验证据来自于人的多重耐药基因*MDR1*。位于*MDR1*中的3个sSNP改变了蛋白质的活性及生理功能, 暗示MDR1的折叠可能受到了影响^[16]。类似的现象随后也在与人类囊性纤维化病相关基因*CFTR*及大肠杆菌*SufI*基因中被报道^[50,51]。更直接的证据则来自对粗糙脉孢菌调控生物节律的核心基因*frequency(frq)*的研究。*frq*基因含有大量稀有密码子, 替换部分稀有密码子为优化密码子尽管在翻译水平上促进了FRQ蛋白量的上升, 但意外地引起生物节律紊乱。具体原因是源于优化*frq*基因的密码子改变了FRQ蛋白的构象及磷酸化修饰, 从而影响蛋白的功能和稳定性^[52]。同样的现象也可在调控果蝇生物钟的*period*基因和哺乳动物视觉晶体蛋白gamma-B crystallin的编码基因中观测到^[53,54]。

蛋白质的翻译速率是如何影响蛋白质折叠的? 实际上, 蛋白质的折叠除了依赖自身序列特征外, 也常常需要分子伴侣蛋白参与^[55]。本实验室结合体内和体外实验不同优化程度密码子编码的荧光素酶蛋白活性发现, 插入一部分富含稀有密码子区段显著增加了正确折叠蛋白的比例^[40], 其原因可能在于翻译速率的降低允许初生多肽链有充足的时间与分子伴侣蛋白互作(图1C)^[56,57]。总之, 含有较多稀有密码子的基因最终产生的蛋白质总体稳定性更高, 且这些基因往往参与维系恶性生理表型相关的信号通路^[58]。

2 密码子偏好性对转录过程的影响

由于核糖体是目前生物体内已知的唯一识别密码子的分子机器, 因此对密码子偏好性作用的认知往往局限于翻译过程。在哺乳动物中, tRNA丰度与密码子偏好性的对应关系比较弱, 且tRNA的含量在不同组织细胞中也呈现较大的差异^[59], 因此翻译效率并不能完全解释同义突变的效应。近年来的证据显示, 在真核生物中, 密码子偏好性可以在转录的多个层面对mRNA水平造成影响。

2.1 mRNA剪切

真核生物的绝大多数基因都含有内含子, 尽管这一特性使得同一个基因可以通过不同的剪切方式获得不同的成熟mRNA, 但移除内含子的过程必须被精确调控以获得正确的成熟mRNA^[60]。由sSNP导致的

mRNA异常剪切可以分为两种情况: (i) sSNP破坏了外显子边缘上的GT/AG序列致使相邻内含子保留, 或添加了新的外显子切割位点致使部分外显子丢失。这类sSNP比较少见, 如抑癌基因*MEN1*和黑色素瘤相关基因*CDKN2A*^[61,62]; (ii) sSNP破坏了位于mRNA前体上的外显子剪切增强子序列(exonic splicing enhancer, ESE)。在转录过程中, 剪切调控蛋白会识别并结合ESE来指导剪切体识别外显子的边界。因此, 发生在ESE上的sSNP可导致该外显子出现跳跃剪切(exon skipping, 图2A)。比较著名的例子为, *BRCA1/2*和*GATA2*基因上的sSNP导致相应位置的外显子跳跃剪切^[63-65]。

2.2 mRNA稳定性

不同mRNA的半衰期从几分钟到数小时不等。一般认为, mRNA的5'和3'UTR上的一些调控元件是决定mRNA稳定性的关键因素^[66], 其中也包含microRNA的靶点。microRNA是调控mRNA稳定性的重要因素, 约84%的microRNA是依赖介导mRNA降解来调控基因表达的^[67]。绝大多数microRNA的靶点位于3'UTR, 但也有部分靶向编码区内部。同义突变如果发生于靶点, 则会影响microRNA对靶点的识别, 从而改变靶向mRNA的稳定性并引起相应的生理变化(图2B, 左)^[68]。

除了microRNA, ORF内是否还存在其他影响RNA稳定性的因素? 早期在转录组水平上的分析并没有发现mRNA半衰期与密码子适应性指数(基因使用密码子偏好的程度)存在明确的关联性^[69]。但将部分酵母基因的优化密码子突变为稀有密码子会降低mRNA稳定性, 且这个过程需要核糖体与mRNA结合^[70,71]。为了解开这个谜题, Presnyak等人^[72]把每一个密码子对于RNA稳定性的贡献做了单独评估, 发现稀有密码子倾向于在半衰期短的mRNA中聚集, 而优化密码子则在半衰期长的mRNA富集。类似现象随后又在大肠杆菌、裂殖酵母、斑马鱼等物种中被相继报道^[73-75]。另外, 在出芽酵母中发现, 此过程依赖于含DEAD-box的RNA解旋酶Dhh1蛋白(图2B, 右), 在稀有密码子区域, Dhh1结合移动速度较慢的核糖体, 进一步激活mRNA脱去5'端帽子引起mRNA降解^[76]。

2.3 转录

最初有关真核生物中密码子调控转录的证据来源于对转录因子在基因编码区的识别。Sternachis等人^[10]

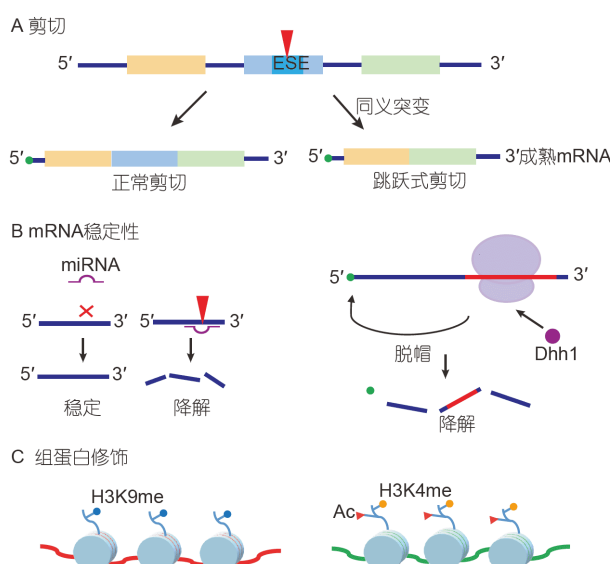


图 2 密码子偏好性改变对转录的影响. A: 同义突变破坏了外显子剪切增强子序列(ESE)引起跳跃式剪切; B: 密码子偏好性影响mRNA稳定性. 同义突变(红色三角形)产生microRNA识别位点, 导致mRNA降解(左); 核糖体在富含稀有密码子区域(红色线条)上移动速度较慢时, 被Dhh1蛋白识别, 进一步引起mRNA脱帽, 导致mRNA被降解(右); C: 密码子偏好性影响组蛋白修饰. 富含稀有密码子的基因可导致组蛋白H3K9me3修饰(左); 而富含优化密码子的基因可促进组蛋白H3K4me3和H3K9Ac修饰增加(右)

Figure 2 Mechanism of codon usage bias in fine-tuning transcription. A: Synonymous mutations cause exon skipping by disrupting sequences of the exonic splicing enhancer (ESE). B: Codon usage bias influences mRNA stability. Synonymous mutations (red triangle) may create recognition sites of microRNAs, resulting in mRNA degradation (left panel). On the other hand, upon stalling of mRNA due to rare codons (red lines), ribosomes can be recognized by Dhh1, resulting in mRNA decapping and degradation (right panel). C: Codon usage bias could influence histone modifications. In particular, a gene rich in rare codons (red line) could induce tri-methylation of histone H3 lysine 9 (H3K9me3) (left panel), whereas a gene rich in optimal codons (green line) could induce methylation of H3K4 and acetylation of H3K9 (right panel)

利用染色质上各个区域对于DNaseI的不同敏感性并结合深度测序技术, 发现大约14%的人类基因编码区可被转录因子识别和结合. 由于转录因子的识别位点偏好GC, 从而对结合区内密码子形成了偏向GC的选择压. 据此推理, 发生于这些区域的同义突变可能改变基因的转录状况. 更为直接的证据来自粗糙脉孢菌. 首先, 在全基因组水平上发现密码子偏好性与mRNA及蛋白的丰度之间存在着非常高的关联性, 预示着mRNA的水平可能会受到密码子偏好性的影响^[21]. 其次, 无论是内源还是外源基因, 将野生型基因的所有密码子同义替换为优化密码子可使蛋白表达量最多增

加120倍, 而这种增加主要来自于mRNA水平增加^[21]. 与酵母中不同的是, 这个效应并没有翻译过程参加, 不同优化程度的mRNA之间也没有显著的稳定性变化^[21]. 最后, 密码子偏好性可以影响组蛋白H3的修饰水平(图2C), 如稀有密码子可导致组蛋白H3K9甲基化修饰发生^[77]. 本实验室在哺乳动物中也发现, 改变致癌的KRAS基因部分稀有密码子为同义优化密码子可促进转录水平的升高和组蛋白修饰, 如H3K9乙酰化的形成^[77].

在真核生物中, 转录的终止依赖于位于3'UTR上的转录终止信号(AAUAAA及类似的富含AU的变体)及附近一段富含A/U的序列, 从而引导转录终止复合体切割新生RNA并在3'末端添加poly(A)尾^[78]. 利用poly(A)测序技术发现, 粗糙脉孢菌的很多基因ORF内部也有微弱的转录终止信号. 分析这些ORF内部的提前转录终止位点(premature transcriptional termination, PTT)附近的序列发现, 周围都富含稀有密码子^[11]. 同样的现象也存在于小鼠的细胞系中, 暗示至少在部分物种中转录终止与同义密码子选择形成了协同进化关系, 最终导致密码子偏好性形成^[11]. 对于经历自然选择却依然存在的这部分PTT是否具有生理意义, 是下一步值得探讨的课题.

3 总结与展望

受经典理论影响, 同义突变一直被认为是“沉默”突变. 近年来受益于技术进步, 尤其是高通量测序等组学技术的进步, 不仅颠覆了人们对过去的认知, 而且由于密码子偏好性便于定量, 也使得该领域进入了定量生物学的时代. 目前的结果表明, 密码子偏好性可以精细调节基因表达的各个层面, 一些致病同义突变的机理也得到初步揭示. 因此, 可以将基因中同义密码子的选择使用看作是二级遗传密码, 在各个层面精细调控基因表达.

由于技术或思维的限制, 密码子偏好性在近年来才逐渐引起人们的关注, 很多方面仍需要深入研究. (i) 国内外并未对致病同义突变开展系统性的遗传流行病学研究, 导致人们不了解哪些致病同义突变的发生概率更高、更加值得深入研究. 因此, 尽快开展相关方面的研究将有助于确定高发的致病同义突变, 从而能够在一些模式动物上建立疾病模型并开展深

入研究; (ii) 从分子机制上看, 尽管目前对密码子偏好性及同义突变在翻译水平上的研究有了较深入的了解, 但这些研究大多集中在简单模式生物上, 对于复杂动物模型, 如小鼠或人的研究仍然很缺乏, 而对于其对转录的影响更是处于非常早期的阶段, 即使在简单模式生物中也大多不明其分子机制; (iii) 一些关键的实验方法仍需要突破. 例如, 对翻译准确性的研究, 由于缺乏准确的tRNA丰度和修饰的测量方法, 以及缺乏在组学层面上对翻译错误的衡量方法, 至今仍然

没有明确的答案; 另外, 由于密码子偏好性的成因复杂, 迄今报道的一些始于20世纪80~90年代的评估密码子偏好性的方法还存在较多的局限性. 因此, 需要配合各方面的进展, 尤其是当前生物大数据的分析来改进或发展新的评估方法, 以期协助预测密码子偏好性及同义突变带来的影响. 总之, 近年来关于人类同义突变致病的报道和关注不断增加, 因此关于此方面的诊断、分子机制及相关治疗手段的研究, 将是未来的一个新兴领域.

参考文献

- 1 Anfinsen C B. Principles that govern the folding of protein chains. *Science*, 1973, 181: 223–230
- 2 Hershberg R, Petrov D A. Selection on codon bias. *Annu Rev Genet*, 2008, 42: 287–299
- 3 Athey J, Alexaki A, Osipova E, et al. A new and updated resource for codon usage tables. *BMC Bioinf*, 2017, 18: 391
- 4 Ikemura T. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol*, 1985, 2: 13–34
- 5 Ikemura T. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: A proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *J Mol Biol*, 1981, 151: 389–409
- 6 Plotkin J B, Kudla G. Synonymous but not the same: The causes and consequences of codon bias. *Nat Rev Genet*, 2011, 12: 32–42
- 7 Quax T E F, Claassens N J, Söll D, et al. Codon bias as a means to fine-tune gene expression. *Mol Cell*, 2015, 59: 149–161
- 8 Drummond D A, Wilke C O. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell*, 2008, 134: 341–352
- 9 Cannarozzi G, Cannarozzi G, Schraudolph N N, et al. A role for codon order in translation dynamics. *Cell*, 2010, 141: 355–367
- 10 Stergachis A B, Haugen E, Shafer A, et al. Exonic transcription factor binding directs codon choice and affects protein evolution. *Science*, 2013, 342: 1367–1372
- 11 Zhou Z, Dang Y, Zhou M, et al. Codon usage biases co-evolve with transcription termination machinery to suppress premature cleavage and polyadenylation. *eLife*, 2018, 7: e33569
- 12 Pouyet F, Mouchiroud D, Duret L, et al. Recombination, meiotic expression and human codon usage. *eLife*, 2017, 6: e27344
- 13 Duret L, Mouchiroud D. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc Natl Acad Sci USA*, 1999, 96: 4482–4487
- 14 Carlini D B, Stephan W. *In vivo* introduction of unpreferred synonymous codons into the *Drosophila Adh* gene results in reduced levels of ADH protein. *Genetics*, 2003, 163: 239–243
- 15 Nackley A G, Diatchenko L. Assessing potential functionality of catechol-*O*-methyltransferase (COMT) polymorphisms associated with pain sensitivity and temporomandibular joint disorders. *Methods Mol Biol*, 2010, 617: 375–393
- 16 Kimchi-Sarfaty C, Oh J M, Kim I W, et al. A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science*, 2007, 315: 525–528
- 17 Sauna Z E, Kimchi-Sarfaty C. Understanding the contribution of synonymous mutations to human disease. *Nat Rev Genet*, 2011, 12: 683–691
- 18 Chen R, Davydov E V, Sirota M, et al. Non-synonymous and synonymous coding SNPs show similar likelihood and effect size of human disease association. *PLoS ONE*, 2010, 5: e13574
- 19 Schwanhäusser B, Busse D, Li N, et al. Global quantification of mammalian gene expression control. *Nature*, 2011, 473: 337–342
- 20 Vogel C, de Sousa Abreu R, Ko D, et al. Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. *Mol Syst Biol*, 2010, 6: 400
- 21 Zhou Z, Dang Y, Zhou M, et al. Codon usage is an important determinant of gene expression levels largely through its effects on transcription. *Proc Natl Acad Sci USA*, 2016, 113: E6117–E6125
- 22 Bentele K, Saffert P, Rauscher R, et al. Efficient translation initiation dictates codon usage at gene start. *Mol Syst Biol*, 2013, 9: 675

- 23 Kudla G, Murray A W, Tollervey D, et al. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science*, 2009, 324: 255–258
- 24 Goodman D B, Church G M, Kosuri S. Causes and effects of N-terminal codon bias in bacterial genes. *Science*, 2013, 342: 475–479
- 25 Archer S K, Shirokikh N E, Beilharz T H, et al. Dynamics of ribosome scanning and recycling revealed by translation complex profiling. *Nature*, 2016, 535: 570–574
- 26 Doma M K, Parker R. Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature*, 2006, 440: 561–564
- 27 Tuller T, Carmi A, Vestsigian K, et al. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell*, 2010, 141: 344–354
- 28 Powers E T, Balch W E. Costly mistakes: Translational infidelity and protein homeostasis. *Cell*, 2008, 134: 204–206
- 29 Li G W, Oh E, Weissman J S. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature*, 2012, 484: 538–541
- 30 Pechmann S, Frydman J. Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat Struct Mol Biol*, 2013, 20: 237–243
- 31 Chu D, Kazana E, Bellanger N, et al. Translation elongation can control translation initiation on eukaryotic mRNAs. *EMBO J*, 2014, 33: 21–34
- 32 Koutmou K S, Radhakrishnan A, Green R. Synthesis at the speed of codons. *Trends Biochem Sci*, 2015, 40: 717–718
- 33 Fedyunin I, Lehnhardt L, Böhmer N, et al. tRNA concentration fine tunes protein solubility. *FEBS Lett*, 2012, 586: 3336–3340
- 34 Ramu H, Vázquez-Laslop N, Klepacki D, et al. Nascent peptide in the ribosome exit tunnel affects functional properties of the A-site of the peptidyl transferase center. *Mol Cell*, 2011, 41: 321–330
- 35 Kaminska M, Havrylenko S, Decottignies P, et al. Dynamic organization of aminoacyl-tRNA synthetase complexes in the cytoplasm of human cells. *J Biol Chem*, 2009, 284: 13746–13754
- 36 Ingolia N T, Ghaemmaghami S, Newman J R S, et al. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science*, 2009, 324: 218–223
- 37 Hussmann J A, Patchett S, Johnson A, et al. Understanding biases in ribosome profiling experiments reveals signatures of translation dynamics in yeast. *PLoS Genet*, 2015, 11: e1005732
- 38 Gerashchenko M V, Gladyshev V N. Translation inhibitors cause abnormalities in ribosome profiling experiments. *Nucleic Acids Res*, 2014, 42: e134
- 39 Schneider-Poetsch T, Ju J, Eyler D E, et al. Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. *Nat Chem Biol*, 2010, 6: 209–217
- 40 Yu C H, Dang Y, Zhou Z, et al. Codon usage influences the local rate of translation elongation to regulate co-translational protein folding. *Mol Cell*, 2015, 59: 744–754
- 41 Yan X, Hoek T A, Vale R D, et al. Dynamics of translation of single mRNA molecules *in vivo*. *Cell*, 2016, 165: 976–989
- 42 Ogle J M, Ramakrishnan V. Structural insights into translational fidelity. *Annu Rev Biochem*, 2005, 74: 129–177
- 43 Shah P, Gilchrist M A. Effect of correlated tRNA abundances on translation errors and evolution of codon usage bias. *PLoS Genet*, 2010, 6: e1001128
- 44 Kramer E B, Farabaugh P J. The frequency of translational misreading errors in *E. coli* is largely determined by tRNA competition. *RNA*, 2007, 13: 87–96
- 45 Meyerovich M, Mamou G, Ben-Yehuda S. Visualizing high error levels during gene expression in living bacterial cells. *Proc Natl Acad Sci USA*, 2010, 107: 11543–11548
- 46 Stansfield I, Jones K M, Herbert P, et al. Missense translation errors in *Saccharomyces cerevisiae*. *J Mol Biol*, 1998, 282: 13–24
- 47 Belcourt M F, Farabaugh P J. Ribosomal frameshifting in the yeast retrotransposon Ty: tRNAs induce slippage on a 7 nucleotide minimal site. *Cell*, 1990, 62: 339–352
- 48 Zhang G, Ignatova Z. Generic algorithm to predict the speed of translational elongation: Implications for protein biogenesis. *PLoS ONE*, 2009, 4: e5036
- 49 Zhou M, Wang T, Fu J, et al. Nonoptimal codon usage influences protein structure in intrinsically disordered regions. *Mol Microbiol*, 2015, 97: 974–987
- 50 Lazrak A, Fu L, Bali V, et al. The silent codon change I507-ATC→ATT contributes to the severity of the ΔF508 CFTR channel dysfunction. *FASEB J*, 2013, 27: 4630–4645

- 51 Zhang G, Hubalewska M, Ignatova Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat Struct Mol Biol*, 2009, 16: 274–280
- 52 Zhou M, Guo J, Cha J, et al. Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature*, 2013, 495: 111–115
- 53 Fu J, Murphy K A, Zhou M, et al. Codon usage affects the structure and function of the *Drosophila* circadian clock protein PERIOD. *Genes Dev*, 2016, 30: 1761–1775
- 54 Buhr F, Jha S, Thommen M, et al. Synonymous codons direct cotranslational folding toward different protein conformations. *Mol Cell*, 2016, 61: 341–351
- 55 Saibil H. Chaperone machines for protein folding, unfolding and disaggregation. *Nat Rev Mol Cell Biol*, 2013, 14: 630–642
- 56 Spencer P S, Siller E, Anderson J F, et al. Silent substitutions predictably alter translation elongation rates and protein folding efficiencies. *J Mol Biol*, 2012, 422: 328–335
- 57 Gloge F, Becker A H, Kramer G, et al. Co-translational mechanisms of protein maturation. *Curr Opin Struct Biol*, 2014, 24: 24–33
- 58 Lian X, Guo J, Gu W, et al. Genome-wide and experimental resolution of relative translation elongation speed at individual gene level in human cells. *PLoS Genet*, 2016, 12: e1005901
- 59 Dittmar K A, Goodenbour J M, Pan T. Tissue-specific differences in human transfer RNA expression. *PLoS Genet*, 2006, 2: e221
- 60 Wang G S, Cooper T A. Splicing in disease: Disruption of the splicing code and the decoding machinery. *Nat Rev Genet*, 2007, 8: 749–761
- 61 Petronzelli F, Sollima D, Coppola G, et al. CDKN2A germline splicing mutation affecting both P16INK4 and P14ARF RNA processing in a melanoma/neurofibroma kindred. *Gene Chromosome Canc*, 2001, 31: 398–401
- 62 Carrasco C A, González A A, Carvajal C A, et al. Novel intronic mutation of MEN1 gene causing familial isolated primary hyperparathyroidism. *J Clin Endocrinol Metab*, 2004, 89: 4124–4129
- 63 Hansen T V O, Steffensen A Y, Jønson L, et al. The silent mutation nucleotide 744 G→A, Lys172Lys, in exon 6 of BRCA2 results in exon skipping. *Breast Cancer Res Treat*, 2010, 119: 547–550
- 64 Raponi M, Kralovicova J, Copson E, et al. Prediction of single-nucleotide substitutions that result in exon skipping: identification of a splicing silencer in BRCA1 exon 6. *Hum Mutat*, 2011, 32: 436–444
- 65 Wehr C, Grotius K, Casadei S, et al. A novel disease-causing synonymous exonic mutation in *GATA2* affecting RNA splicing. *Blood*, 2018, 132: 1211–1215
- 66 Chen C Y A, Shyu A B. Emerging themes in regulation of global mRNA turnover in *cis*. *Trends Biochem Sci*, 2017, 42: 16–27
- 67 Guo H, Ingolia N T, Weissman J S, et al. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*, 2010, 466: 835–840
- 68 Brest P, Lapaquette P, Souidi M, et al. A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease. *Nat Genet*, 2011, 43: 242–245
- 69 Wang Y, Liu C L, Storey J D, et al. Precision and functional specificity in mRNA decay. *Proc Natl Acad Sci USA*, 2002, 99: 5860–5865
- 70 Caponigro G, Muhlrud D, Parker R. A small segment of the MAT alpha 1 transcript promotes mRNA decay in *Saccharomyces cerevisiae*: A stimulatory role for rare codons. *Mol Cell Biol*, 1993, 13: 5141–5148
- 71 Hu W, Sweet T J, Chamnongpol S, et al. Co-translational mRNA decay in *Saccharomyces cerevisiae*. *Nature*, 2009, 461: 225–229
- 72 Presnyak V, Alhusaini N, Chen Y H, et al. Codon optimality is a major determinant of mRNA stability. *Cell*, 2015, 160: 1111–1124
- 73 Boël G, Letso R, Neely H, et al. Codon influence on protein expression in *E. coli* correlates with mRNA levels. *Nature*, 2016, 529: 358–363
- 74 Mishima Y, Tomari Y. Codon usage and 3' UTR length determine maternal mRNA stability in *Zebrafish*. *Mol Cell*, 2016, 61: 874–885
- 75 Harigaya Y, Parker R. Analysis of the association between codon optimality and mRNA stability in *Schizosaccharomyces pombe*. *BMC Genomics*, 2016, 17: 895
- 76 Radhakrishnan A, Chen Y H, Martin S, et al. The DEAD-box protein Dhh1p couples mRNA decay and translation by monitoring codon optimality. *Cell*, 2016, 167: 122–132.e9
- 77 Fu J, Dang Y, Counter C, et al. Codon usage regulates human KRAS expression at both transcriptional and translational levels. *J Biol Chem*, 2018, 293: 17929–17940
- 78 Porrua O, Libri D. Transcription termination and the control of the transcriptome: Why, where and how to stop. *Nat Rev Mol Cell Biol*, 2015, 16: 190–202

Codon codes: Codon usage bias influences many levels of gene expression

REN GuiPing^{1,2}, DONG YingYing^{1,2} & DANG YunKun^{1,2}

1 State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Yunnan University, Kunming 650091, China;

2 Center for Life Sciences, School of Life Sciences, Yunnan University, Kunming 650500, China

Synonymous mutations, which do not change the primary sequences of encoded proteins, are often recognized as silent mutations. In particular, the selection of synonymous codons at an evolutionary scale is constrained and has resulted in uneven occurrences, a phenomenon called codon usage bias. In recent decades, the roles of codon usage bias in fine-tuning transcription, post-transcriptional processing, mRNA stability, translation initiation, elongation, and peptide folding have been revealed. Upon disrupting these biological processes, synonymous mutations have been reported to cause various diseases such as cancer. Herein, we briefly review the current understanding concerning the molecular mechanisms of transcription and translation processes regulated by codon usage bias as well as the implications for basic and translational research in medicine.

codon usage bias, pathogenic synonymous mutation, transcription, translation

doi: [10.1360/SSV-2019-0103](https://doi.org/10.1360/SSV-2019-0103)