

RNA 翻译的复杂性：不翻译、部分翻译、 从头翻译及过度翻译*

郝亚静 骆健俊 张宝 陈润生**

(中国科学院核酸生物学重点实验室, 中国科学院生物物理研究所, 北京 100101)

摘要 经典分子生物学的中心法则描述了遗传信息的传递方向。中心法则认为 RNA 只有通过翻译产生蛋白质才在生命活动中发挥功能。但是随着分子生物学的发展, 很多 RNA 其本身就可以承担生命学功能。而且 RNA 的形式也不仅仅只有线性这一种。本文总结了 RNA 转录后命运的 4 种形式: 不翻译、部分翻译、从头翻译和过度翻译。RNA 命运的多样性使得对于翻译的理解比经典的中心法则规定的内容将更加丰富、复杂。充分了解 RNA 转录后的命运, 对以后研究 RNA 的功能提出了更高的要求, 也为我们真正而全面地了解 RNA 的功能提供了可能。

关键词 RNA 翻译, 不翻译, 部分翻译, 从头翻译, 过度翻译
学科分类号 Q752

DOI: 10.16476/j.pibb.2017.0101

经典中心法则描述了遗传信息的传递方向, DNA 转录产生 mRNA, 而 mRNA 利用所携带的遗传信息指导蛋白质的合成。中心法则认为蛋白质是生命活动的主要承担者。然而随着科学技术的进步, 科学家们发现, 人类基因组中 70% 以上的 DNA 可以转录产生 RNA, 而只有 3% 左右的 DNA 能转录产生具编码蛋白质能力的 mRNA, 表明绝大多数转录产生的 RNA 分子都属于非编码 RNA。由于这些非编码 RNA 不能翻译产生蛋白质, 一度被认为是转录“噪音”、“暗物质”, 而对非编码 RNA 的研究直到最近十几年才开始受到广泛的重视, 并迅速成为现代生命科学研究的最热门领域之一。为更好地探索生命体中的这些“暗物质”信息, 涌现出了一批非编码 RNA 专家数据库, 其中比较著名的有 NONCODE^[1]与 LNCipedia^[2]等。随着越来越多非编码 RNA 数据的积累, 特别是针对许多非编码 RNA 的重要功能的解析, 进一步揭示了生命活动的复杂性, 同时也宣示新一轮基因功能研究革新时代的到来。

通常认为非编码 RNA (non-coding RNAs, ncRNA) 是不能翻译产生蛋白质的 RNA 转录本。而

最近的研究发现这类 RNA 有的可以部分翻译产生小肽, 这样的小肽(小蛋白)同样也能在生命活动中发挥重要的作用。例如: 来自于 ncRNA 的 TAL 小蛋白可以影响果蝇的早期发育^[3]; 又如 LINC00948 之前被定义为长链非编码 RNA, 随后发现其可以产生一段长度为 46 aa 的小蛋白(MLN), 并且这个小蛋白在人类中通过抑制关键钙离子泵的活性来调控肌肉收缩^[4]; 随后 Nelson 等^[5]又发现一个与 MLN 相反功能的同样来源于非编码 RNA 的小蛋白(DWOF), 这个小蛋白可以通过激活相同的钙离子泵来增强肌肉活性。上述的实验结果说明, 翻译过程比我们想象的要复杂, RNA 作为模板, 不仅仅是从起始密码子翻译到终止密码子, 也可能部分翻译。另外, 在生物体的细胞中除了线性 RNA, 也存在着很多环状 RNA(circular RNAs, circRNAs), 这些 circRNAs 表达量相对较低, 最开始一直被大

* 国家自然科学基金(31520103905, 91440116)资助。

** 通讯联系人。

Tel: 010-64888543, Fax: 010-64871293, E-mail: rschen@ibp.ac.cn

收稿日期: 2017-03-16, 接受日期: 2017-06-27

家所忽视, 近来, 科学家发现这类 circRNAs 在人类生物体中也是广泛存在的, 而且研究表明, 这些 circRNAs 同样可以翻译产生蛋白质, 但其产生蛋白质的过程不同于线性 RNA, 核糖体可以在 circRNAs 上转圈, 从而可以产生比其本身携带的遗传信息更长的蛋白质. RNA 世界的复杂, 导致我们不能只是简单地通过中心法则的翻译准则来解释这些 RNA 翻译现象的存在. 本文归纳整理了最新的翻译现象并总结出了 RNA 转录本的 4 种不同命运: 不翻译、部分翻译、从头翻译(经典翻译流程)和过度翻译.

1 不翻译

人类基因组 70% 以上的区域可以转录, 但是真正能够翻译产生蛋白质的区域仅有 3% 左右, 这就暗示着人类基因组存在着大量的 ncRNA. 这些 ncRNA 既包括一些保守的小 RNA 分子(tRNA, snRNA, miRNA 等)还包括近些年来发现的承担着

重要生命角色的长链非编码 RNA. 长链非编码 RNA 是一类长度大于 200 核苷酸, 但是不能编码产生蛋白质的一类核酸分子. 之前大量的数据挖掘证实了这类核酸分子的广泛存在. ENCODE(即“DNA 元件百科全书”计划, Encyclopedia of DNA Elements, 简称 ENCODE)的测序计划拼接组装产生的人类转录本中, 60% 都是非编码 RNA(图 1). HOTAIR 是第一个被发现具有反式转录调控作用的长链非编码 RNA^[6], 其表达与多种肿瘤的发生发展以及转移预后密切相关. XIST 是被发现的另一个明星的长链非编码 RNA, XIST 的表达对于 X 染色体失活至关重要^[7]. lncTCF7 通过激活 Wnt 信号通路来促进人类肝癌干细胞的自我更新^[8]. 以及随后发现的 LINC01186^[9]、ADINR^[10]、ASNR^[11]等等. 小的非编码 RNA 不能翻译产生有功能的蛋白质可能受其长度的限制, 那么长链非编码 RNA 不能够翻译产生蛋白质又是受怎样的调控呢?

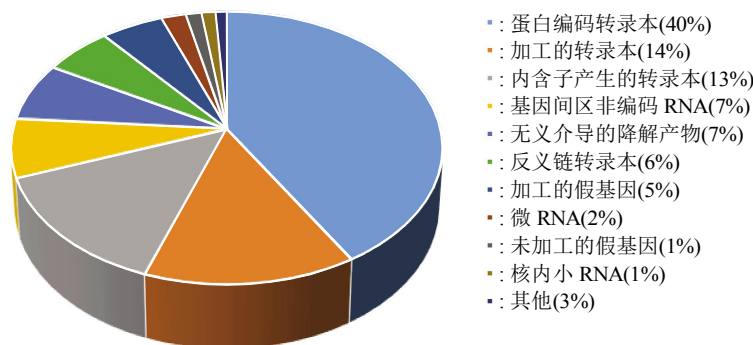


Fig. 1 Distribution of types of transcripts

图 1 转录本类型的分布

统计 ENCODE 组装产生的人类转录本中各转录本的比例. 数据来自于 GENCODE V24. 从图中可以看出编码基因的转录本占到全部转录本的 40% 左右, 其余 60% 为非编码转录本.

利用 ENCODE 数据库中 HeLa 细胞系的测序数据, 我们针对 GENCODE V24^[12]所有注释的长链非编码 RNA 的翻译情况进行了比较分析(图 2). 结果表明, 虽然在 HeLa 细胞中可以转录的长链非编码 RNA 有很多, 但是能从细胞核进入细胞质的仅占 11% 左右. 翻译的发生离不开核糖体, 而核糖体定位在细胞质中, 这就表明绝大多数非编码 RNA 不能够翻译产生蛋白质主要可能是因为无法被转运到细胞质, 从而无法接触到翻译机器. 这也说明大部分的非编码 RNA 其实是在核内实现生物

学功能的. MALAT1 就是位于细胞核并发挥重要作用的一个非编码 RNA, 其不仅可以通过调节 SR 蛋白的激活水平来调节可变剪切^[13], 还和肺癌的恶化和转移相关^[14]. 此外, PVT1 和 BCAR4 也主要定位在细胞核. PVT1 干扰 MYC 的磷酸化, 随后增加 MYC 的稳定性然后在癌症中导致 MYC 的积累^[15]. BCAR4 在乳腺癌中通过两个不同的区域分别结合转录因子 SNIP1 和 PNUTS, 随后影响下游基因的表达调控^[16].

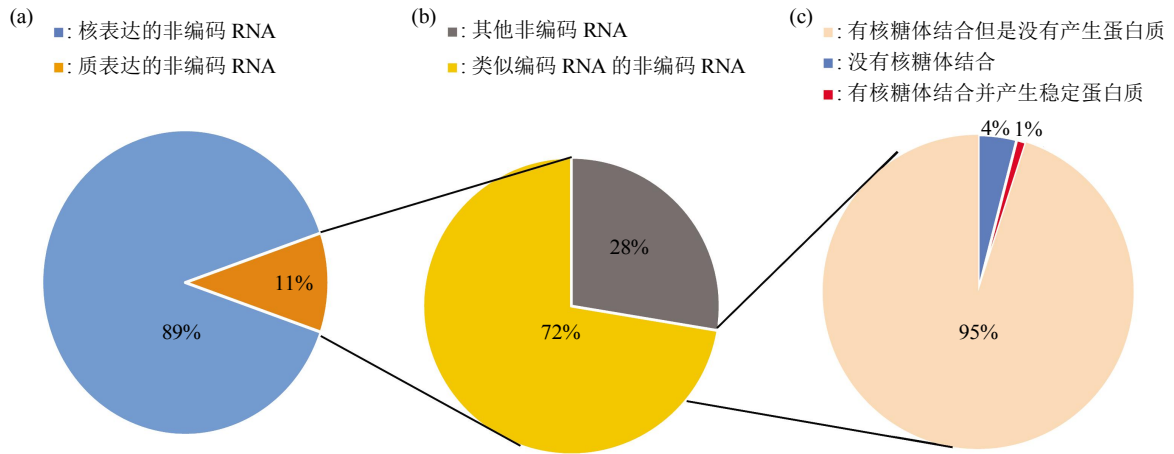


Fig. 2 Diversity of the fate of the non-coding RNAs from the example of HeLa cells

图 2 HeLa 细胞系为例探讨非编码 RNA 命运的多样性

利用 HeLa 相关测序数据分析发现: DNA 转录产生的长链非编码 RNA 有 89%位于细胞核, 只有 11%可以从细胞核转运进入细胞质(a). 可以转运进细胞质的非编码 RNA 72%具有类似于 mRNA 的结构, 既有 5'帽子结构又有 3'polyA 尾巴, 将这一类非编码 RNA 称为类似编码 RNA 的非编码 RNA(mRNA-like ncRNA)(b). mRNA-like ncRNA 96%可以接触到翻译机器, 只有 1%可以产生蛋白质(c).

常规翻译过程中, 核糖体通过识别 5'帽子结构结合 RNA 并启动扫描过程. 5'帽子结构和 3' polyA 尾巴的存在使得位于细胞质的 RNA 不被降解从而可能被翻译机器识别并启动翻译过程. 通过获取 HeLa 细胞系的 PolyA-seq、non-PolyA-seq 以及 CAGE-seq 的数据, 我们对于可以从细胞核进入细胞质的 11%左右的长链非编码 RNA, 进行了 RNA 结构的进一步分析. 研究发现, 能够出核的非编码 RNA 中, 72%都是既有 5'帽子结构又有 3' polyA 尾巴, 这一类非编码 RNA 通常被叫做 mRNA-like 的非编码 RNA. 对于剩余的 28%不具有完整结构信息的非编码 RNA, 虽然它们可以出核, 但是由于不具有完整的 RNA 帽子和尾巴, 从而不能导致翻译机器的结合和启动, 这样就无法产生稳定的翻译产物——蛋白质. 现在关于不具备完整结构也可以出核的非编码 RNA 还需要进一步的研究.

那对于这些既可以出核, 又有完整帽子与尾巴结构的非编码 RNA 为什么也不可以产生蛋白质? 难道蛋白质产生还需要其他信息的存在, 还是其产生了蛋白质却被快速降解呢? 为了解决这个问题. 我们获取了 HeLa 细胞的核糖体图谱(ribosome profiling)的数据, 发现的确如所预料, 96%类似编码 RNA 的非编码 RNA(mRNA-like ncRNA)都可以结合到核糖体这个翻译机器. 随后根据核糖体翻译蛋白每 3 个核苷酸翻译产生 1 个氨基酸, 从而真正

活跃转录的开放读码框 (ORF) ribosome profiling 产生的序列(reads)比对后会呈现三周期性的原则, 利用 RiboTaper^[17]算法来看这些结合到翻译机器的 RNA 是否可以产生蛋白质. 我们发现虽然 96%的 mRNA-like ncRNA 编码 RNA 可以结合翻译机器, 但是有可能产生蛋白质的只有 1%. 所以这些 RNA 之所以不能翻译, 应该还存在着其他的问题. 但是这类原因又是什么现在还不清楚. 为此, 我们对这 1%可能产生蛋白质的非编码 RNA, 进一步利用质谱实验进行了检测, 结果发现只有 1 个候选非编码 RNA 能被质谱检测到, 而其余都尚不能被质谱数据验证. 暗示着就算预测可以产生蛋白, 而这些蛋白质也有可能由于不能稳定存在也会是其没有翻译产物的原因.

概括而言, 非编码 RNA 不翻译的原因主要可能是由于存在以下 4 种情况(图 3):

- 定位于细胞核, 无法接触到翻译机器;
- 虽然被转运出细胞核, 但仍无法接触到翻译机器;
- 被转运出细胞核, 能结合翻译机器, 但是由于某些未知的因素, 仍无法翻译;
- 被转运出细胞核的, 能结合翻译机器, 并且可以正确启动翻译过程, 但是翻译产生的蛋白质不能稳定存在, 快速被降解, 导致无法检测到稳定产物.

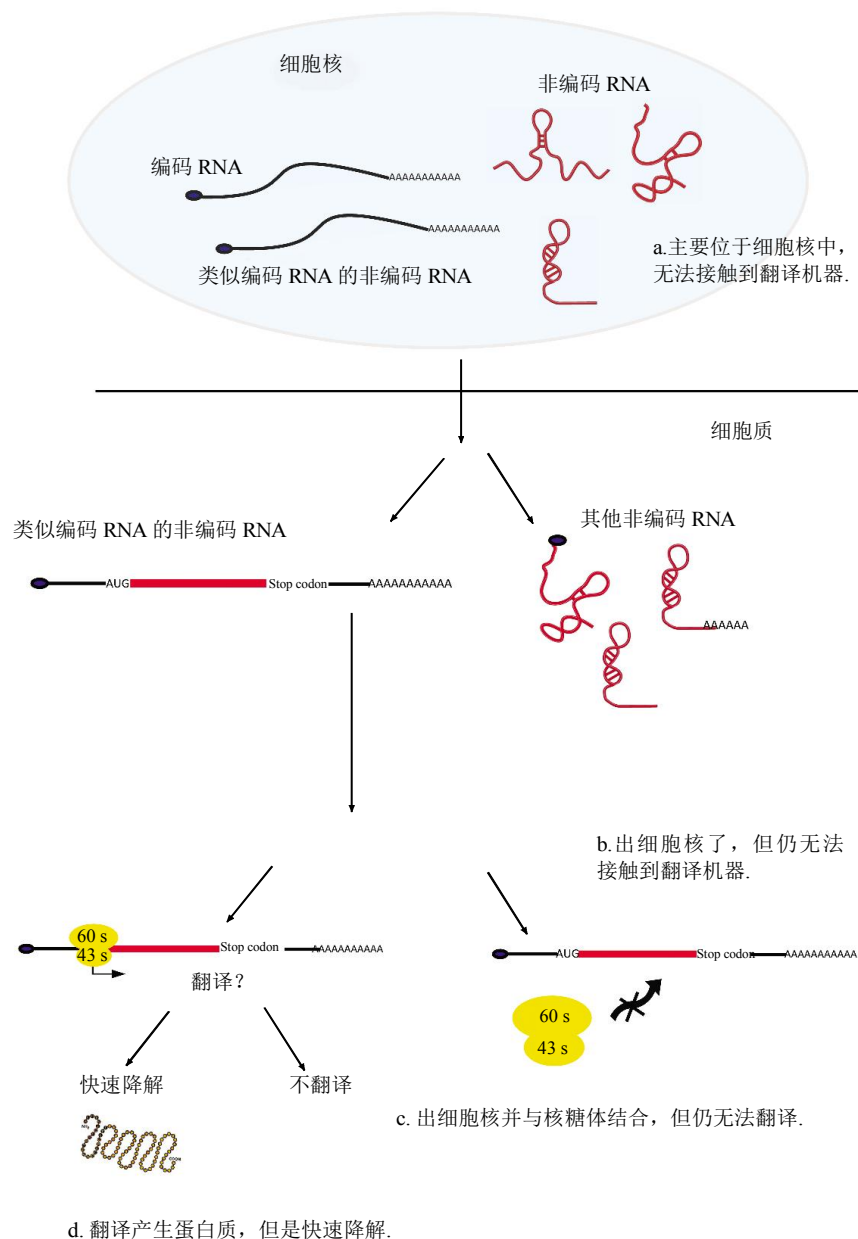


Fig. 3 On the causes of non-coding RNA's no-translation

图 3 非编码 RNA 不翻译的原因探讨

通过生物信息学分析, 发现非编码 RNA 不能产生有功能的蛋白质的原因主要可以分为以下 4 点: a. 定位于细胞核, 无法接触到翻译机器; b. 虽然被转运出细胞核, 但仍无法接触到翻译机器; c. 被转运出细胞核, 能结合翻译机器, 但是由于某些未知的因素, 仍无法翻译; d. 被转运出细胞核的, 能结合翻译机器, 并且可以正确启动翻译过程, 但是翻译产生的蛋白质不能稳定存在, 快速被降解, 导致无法检测到稳定产物. Stop codon: 终止密码子.

2 部分翻译

部分翻译是指 RNA 并没有把作为模板的全部信息翻译成肽链, 而只是翻译了一部分信息. 在经典翻译过程中, 核糖体结合到 mRNA 上, 从而开

始扫描直到遇到第一个符合条件的 AUG 并开始启动翻译. 然而近来研究表明存在着很多不符合经典翻译流程的例子. 这些事件中翻译的起始并不是起始于第一个扫描遇到的 AUG, 而是从中间某一个密码子来启动翻译, 这个密码子可以是典型的起始

密码子 AUG, 也可以是其他 60 个密码子中的任意一个. 当然 AUG 的起始还是占绝大多数. 其中一个著名的例子就是 LINC00948, 它从第 2 个起始密码子开始翻译产生 1 个长为 46 个氨基酸的名为 MLN 的小蛋白^[4]. 这个 MLN 小蛋白在钙离子吸收过程中的功能我们在前面已经描述过了. 另外, 也发现了以非 AUG 起始的翻译例子: 利用我们自主构建的数据库 NONCODE, 对其所包含的非编码 RNA 进行潜在的非编码 RNA 蛋白质库的构造, 随后将质谱数据和非编码 RNA 蛋白质库进行比对, 发现了 103 个来自于非编码的肽段. 对于这些肽段的结构和序列信息进行解析, 其中就存在着 2 个以非 AUG 起始的肽段. 同时, 本实验室专门构建了一个数据库 SmProt^[18], 该数据库收集了所有的小蛋白序列, 特别关注了来自于现在的非编码 RNA 的小蛋白. 在数据库中非 AUG 起始的例子也有很

多(SPROHSA029394、SPROHSA027116 等). 这种非 AUG 起始的现象以及非第一个 AUG 起始的翻译情况, 可以导致作为模板的长非编码 RNA 不能把全部信息翻译成多肽链, 从而造成了部分翻译(图 4). 这类翻译的起始可能是核糖体启动扫描然后在转录本中间位置进行翻译, 也可能是核糖体直接结合到相应的起始位置而非采取扫描的方法来起始翻译. 随后我们从文献调研了部分翻译的现象. 发现部分翻译如果采取扫描而非从第一个 AUG 起始也可能是上游开放阅读框(upstream open reading frame, uORF)调控的翻译^[19]. uORF 是一种位于 5' 非翻译区 (5'untranslated region, 5'UTR) 区域的开放阅读框. 如果是直接以非扫描方式启动翻译, 可能是通过 IRES(内部核糖体进入位点, Internal ribosome entry site)介导的翻译情况^[20].

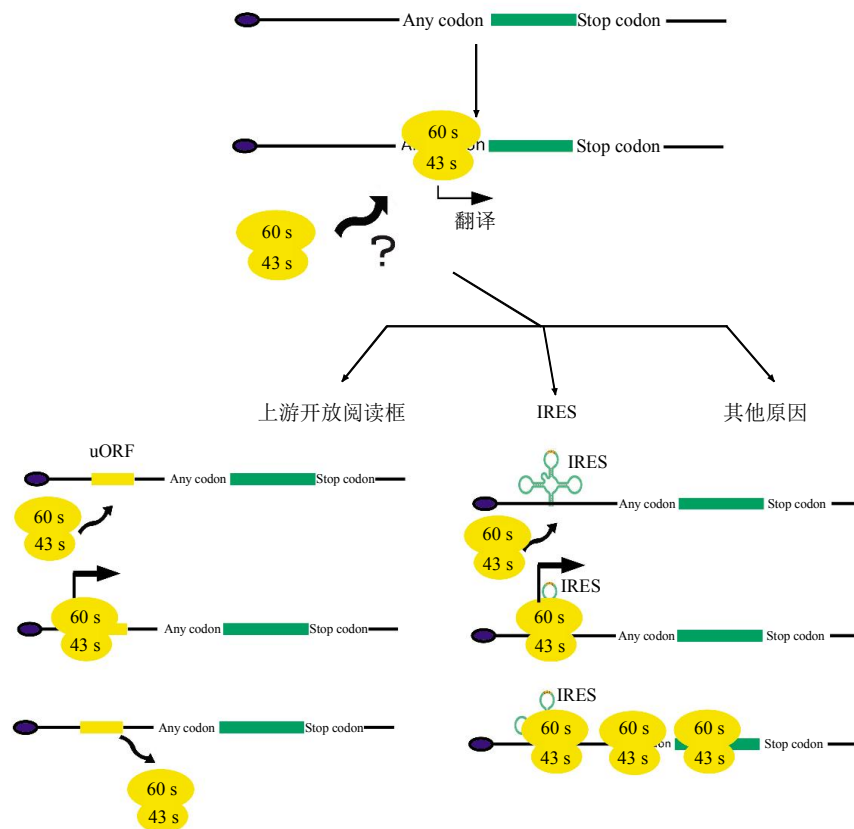


Fig. 4 Part-translation

图 4 部分翻译

部分翻译是指翻译的起始不是遵循传统扫描模型并从第一个 AUG 起始的翻译, 而是其从非第一个 AUG 或者任何一个非 AUG 密码子起始的翻译. 部分翻译的原因可能有很多, 但是现在有 2 个已知的可以解释这种现象, 首先是 uORF 的存在. uORF 是通常是第一个 AUG 出现的位置, 其有时候无法产生稳定的蛋白质, 导致我们观测到的蛋白质主要是来源于非第一个 AUG 的读框. 其次是 IRES, IRES 的存在使得翻译机器可以在 IRES 的位置结合到 RNA 上, 从而使得翻译的进行不依赖于 AUG 起始密码子. 当然还有其他的原因需要我们进一步探讨. Any codon: 任意密码子; Stop codon: 终止密码子. IRES: 核糖体进入位点.

uORF 是指位于 5'UTR 区域的开放阅读框, 以 AUG 起始和标准终止密码子结束. uORF 调控现象是指, 在真核生物转录本中 uORF 可以调控下游主要蛋白的翻译. 因此, 正常的第一个 AUG 无法起始翻译可能是由于 uORF 的存在. uORF 有时无法起始翻译, 有时会逃逸扫描从而产生小蛋白, 但这些小蛋白很容易被降解. uORF 现象可以解释一部分非第一个 AUG 起始的翻译事件. IRES 又称内部核糖体进入位点, 一般真核 mRNA 的翻译都需要 5'帽子来介导核糖体结合, 但真核生物和病毒中还存在着一些例外情况, 例如一些基因 5'端具有一段较短的 RNA 序列(约 150~250 碱基), 这类 RNA 序列能折叠成类似于起始 tRNA 的结构, 从而介导核糖体与 RNA 结合, 启动蛋白质翻译, 这段非翻译 RNA 被称为内部核糖体进入位点序列. IRES 最初是 1988 年在脊髓灰质炎病毒(poliovirus, PV)和脑脊髓炎病毒(encephalomyocarditis virus, EMCV) RNA 基因组中发现的^[21-22], 随后在哺乳动物、植物以及酵母中也均发现了 IRES 序列存在^[23-24]. IRES 主要是通过其折叠产生的二级结构或者更为复杂的三级结构来介导核糖体的结合. 虽然 IRES 的基本原理是通过其折叠形成的结构来完成的, 但是不同的 IRES 其介导翻译的原理又不同. IRES 的存在可以为非 AUG 起始的翻译提供一个解释.

部分翻译不仅是在起始位置有遗传信息的丢失, 其在终止位置可能也有遗传信息的丢失. 非典型终止密码子导致的终止现象现在研究不多, 但是在我们的数据库中也有发现蛋白质的终止端并非是典型的终止密码子的情况. 当然这类非正常终止密码子的终止现象还需要进一步的研究和探讨.

3 从头翻译 (经典翻译流程)

经典的中心法则提出, DNA 可以转录形成 RNA, 然后 RNA 通过核糖体翻译来产生蛋白质从而发挥生物学功能, 并且 DNA 本身也可以通过自我复制来扩增和作为遗传物质遗传到子代^[25]. 1970 年, 通过对 RNA 肿瘤病毒的研究, 科学家们发现病毒中作为遗传物质的 RNA 不但可以自我复制, 还可以通过逆转录形成 DNA 来影响宿主体内的生物学活动, 从而中心法则得以修订.

从头翻译指教科书中所定义的广泛发生的翻译现象. 翻译机器结合到 RNA 上, 扫描遇到第一个 AUG 然后起始翻译产生蛋白质的过程. 也就是中心法则中 RNA 翻译产生蛋白质的过程. 关于从头

翻译在教科书以及已发表的综述中都有较详细的介绍, 所以本文不再详细阐述, 只是简单介绍一下其发生的基本流程.

经典翻译的过程主要包括三部分: 翻译的起始、延伸和终止.

a. 翻译起始

在蛋白质合成的过程中需要核糖体大小亚基, 起始 tRNA 和几十种蛋白因子的参与, 在模板 mRNA 编码区 5'端形成核糖体-mRNA-起始 tRNA 复合物, 然后沿 mRNA 移动直到遇到起始密码子 AUG, 它能在 AUG 处停下来并起始翻译.

b. 肽链延伸

在起始复合物形成以后, 首先 eEF1 和 GTP 结合携带 tRNA 到核糖体上 A 位. 随后 GTP 水解, 释放 eEF1A, 然后 tRNA 和 A 位上的密码子进行碱基配对. 随后在肽酰转移酶的作用下, 新加入的氨基酸的氨基对之前肽段的羧基进行亲核攻击完成转肽.

c. 翻译终止

在肽链延伸过程中, 当终止密码子 UAA、UAG 或 UGA 出现在核糖体 A 位点时, 没有相应的氨基酰 tRNA 能与之结合, 而释放因子能识别这些密码子并与其结合, 水解 P 位点上多肽链与 tRNA 之间的二酯键. 接着, 新生的肽链和 tRNA 从核糖体上释放, 核糖体上大小亚基解体, 蛋白质合成结束.

4 过度翻译

CircRNAs 是一类由前体 RNA 经过可变剪接形成的转录本. circRNAs 的发现从病毒到真核, 从单个到大规模, 到目前 circRNAs 的发现已经数不胜数. 那么这么多的 circRNAs 到底有什么功能呢? 目前对 circRNAs 的功能研究则集中在阐释其在 RNA 水平上的功能, 比如与转录因子结合调控转录^[26]; 作为竞争性内源 RNA(competitive endogenous RNA, ceRNA)在转录后水平调控 mRNA 的降解^[27]; 作为 microRNA 海绵吸收大量的 microRNA^[28]. 而对于 circRNAs 是否可以指导翻译或者本身是否可以翻译, 则是另一个层面上关于 circRNAs 的功能. 第一个提出 circRNAs 可以翻译产生蛋白质是 1986 年的一篇报道, 该研究发现类病毒基因组为 circRNA, 其反转录出来的 cDNA 包含一个 122 个氨基酸的 ORF^[29]. 1995 年第一篇发现在真核生物中 circRNAs 可以翻译产生蛋白质的文章发表在

《科学》杂志(*Science*), 这篇文章同时证明 IRES 在真核生物起始阶段是必要的^[30]. 随后几年, 关于 circRNAs 翻译的研究却进入一个瓶颈期, 迟迟没有新的研究出现. 直到 2014 年一篇文章中的研究发现 circRNAs 翻译不同于真核生物线性 RNA 的翻译, 它可以产生比自己遗传信息还要多的蛋白质. 这篇文章发现一种叫做水稻黄斑病毒(rice yellow mottle virus, RYMV)的拟病毒, 该病毒可以导致水稻黄斑病变^[31]. 该拟病毒仅由 220 碱基的闭合环状 RNA 构成, 按照以往正常编码的经验, 三联体密码子计算其最多编码 73 个氨基酸, 但实际上这个病毒却编码了 16 ku 的蛋白, 这远远超过正常情况下编码蛋白质的量. 这就说明在翻译时核糖体第一次走完环状 RNA 全部序列之后, 在起始密码子前的一个碱基和起始密码子的“AU”继续作为 ORF 的密码子进行第二轮编码, 当再次到达起始位点的时候, 原先起始密码子的“A”与前面的 2 个碱基“UG”构成了第一终止密码子, 后面还有第 2 和第 3 终止密码子存在, 因此个别情况下还会表达出 18 ku 或 23 ku 大小的蛋白. 2017 年的三篇文章也加深了我们对于 circRNAs 翻译的了解. 分别是 2017 年 2 月发表在 *Cell Research* 杂志上的一篇文章, 文章中发现 m⁶A 修饰是起始 circRNAs 翻译的一个方式并且这种起始翻译的方式在 circRNAs 的翻译中是广泛存在的, 而且仅仅一个 m⁶A 修饰的存在就可以启动翻译进程^[32]. 以及 2017 年 4 月发表在 *Molecular Cell* 上的两篇文章, 都证实了 circRNAs 的翻译是广泛存在的, 并且都找到了一个 circRNA 翻译的例子 (Circ-ZNF609^[33]和 circMbl^[34]), 并进行了低通量实验的证实. CircRNAs 翻译的存在也拓展了对现有翻译的了解. 上述的例子中有的 circRNAs 翻译产生的蛋白质是翻译机器在 circRNAs 旋转不到一圈产生的, 也有的是翻译机器在 circRNAs 上旋转多圈产生的. 我们觉得类似于 RYMV 病毒的这一类 circRNAs 的翻译都可以归类为过度翻译, 即翻译机器可以在 circRNAs 旋转进而产生长度大于其本身遗传信息的一段蛋白质(图 5). 在最新一版的 circRNAs 数据库 circRNADb^[35]中收录了 32 914 个 circRNAs, 其中 46 个 circRNAs 可以产生有质谱数据支持的蛋白质. 这说明 circRNAs 产生蛋白质不是随机的或者不稳定的蛋白质. 另外 circScan 方法通过探究 circRNAs 和 RNA 结合蛋白之间的互作关系, 也发现大量 circRNAs 和独立于 5'帽子结构

的翻译起始相关蛋白存在互作, 这些现象暗示很多 circRNAs 可能通过这些因子来起始翻译产生功能蛋白^[34]. 当然这类蛋白质是不是 circRNAs 过度翻译的产物还需获得完整的蛋白质才可以知晓.

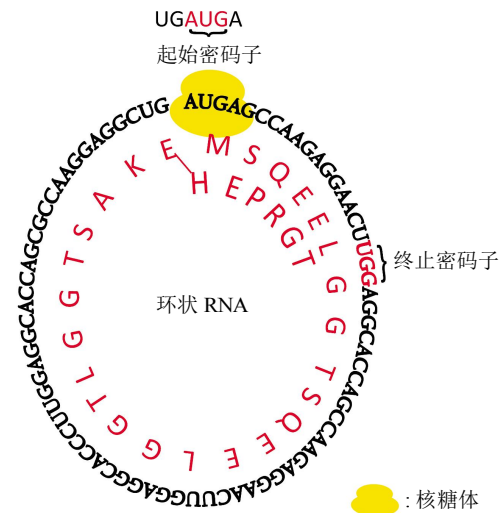


Fig. 5 Model of Over-translation

图 5 过度翻译模型

CircRNAs 的翻译可以产生比其本身携带的遗传信息更多的蛋白质, 这类翻译称为过度翻译. 如图所示 CircRNA 可以产生的氨基酸序列为: MSQEELGGTSAKEHEPRGT. 其中黑色的碱基代表 CircRNA 的核苷酸序列, 而红色的代表翻译产生的氨基酸序列. 黄色的代表核糖体.

5 展 望

RNA 是中心法则的核心分子, 既可以从 DNA 获取遗传信息, 又可以将遗传信息传递到蛋白质. 最初认为 RNA 只是遗传信息的传递者, 越来越多的研究表明 RNA 不仅可以传递遗传信息, 本身也可以以 RNA 分子的形式发挥功能. 前边列举的典型非编码 RNA 的功能就是一个很好的例证. DNA 转录产生 RNA, RNA 在翻译成蛋白质前还进行了选择性可变剪接、RNA 编辑、选择性多聚腺苷化等一系列加工修饰. 这些加工修饰增加了 RNA 分子的多样性, 也使得本身的遗传信息得以在 RNA 层面改变, 从而丰富了蛋白质产物. RNA 分子在转录后各种加工和修饰的存在也间接增加了 RNA 翻译的多样性. 本文的介绍说明了在转录加工修饰后 RNA 转录本命运多样性, 其既可以直接以 RNA 分子来承担生命活动中的重要角色, 也可以只翻译其中一部分来产生小蛋白从而发挥功能. 此

外还可以按照标准翻译的流程进行扫描翻译以及在特定 RNA 分子中存在的过度翻译. RNA 研究相对 DNA 来说是复杂的, 而又是非常有趣的. RNA 翻译的复杂性比经典的中心法则规定的内容将更加丰富、复杂.

最初科学家们专注于蛋白质功能的研究, 认为 RNA 只有翻译蛋白质才能在生命活动中发挥作用, 将大量的非编码 RNA 归类为“垃圾”. 但是随着功能非编码 RNA(XIST、HOTAIR 等)的出现, 才使得这些被埋藏的宝藏得以被发现. 现有的研究又发现这些非编码 RNA 可能编码产生小蛋白. 由于小蛋白的长度以及表达量低的特点在原有的技术中被大家所忽略. 这两年以来这些小蛋白(MLN、DWORF 等)的功能逐渐被大家认可. 所以着眼于 RNA 研究时, 不能再将其简单的分为非编码 RNA 和编码 RNA. 而是应该根据其所有的可能命运形式来研究其功能. 其即可能编码蛋白质发挥功能, 而其编码蛋白质的模板 RNA 本身也有很重要的功能. 充分了解 RNA 转录后的命运, 对以后研究 RNA 的功能提出了更高的要求, 也为我们真正的全面的了解 RNA 的功能提供了可能.

参 考 文 献

- [1] Zhao Y, Li H, Fang S, *et al.* NONCODE 2016: an informative and valuable data source of long non-coding RNAs. *Nucleic Acids Res*, 2016, **44**(D1): D203-208
- [2] Volders P J, Verheggen K, Menschaert G, *et al.* An update on LNCipedia: a database for annotated human lncRNA sequences. *Nucleic Acids Res*, 2015, **43**(Database issue): D174-180
- [3] Martinez Arias A, Galindo M I, Pueyo J I, *et al.* Peptides encoded by short ORFs control development and define a new eukaryotic gene family. *PLoS Biology*, 2007, **5**(5): e106
- [4] Anderson D M, Anderson K M, Chang C L, *et al.* A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell*, 2015, **160**(4): 595-606
- [5] Nelson B R, Makarewich C A, Anderson D M, *et al.* A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. *Science*, 2016, **351**(6270): 271-275
- [6] Gupta R A, Shah N, Wang K C, *et al.* Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*, 2010, **464**(7291): 1071-1076
- [7] Brown C J, Hendrich B D, Rupert J L, *et al.* The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell*, 1992, **71**(3): 527-542.
- [8] Wang Y, He L, Du Y, *et al.* The long noncoding RNA lncTCF7 promotes self-renewal of human liver cancer stem cells through activation of Wnt signaling. *Cell Stem Cell*, 2015, **16**(4): 413-425
- [9] Hao Y, Yang X, Zhang D, *et al.* Long noncoding RNA LINC01186, regulated by TGF-beta/SMAD3, inhibits migration and invasion through Epithelial-Mesenchymal-Transition in lung cancer. *Gene*, 2017, **608**: 1-12
- [10] Xiao T, Liu L, Li H, *et al.* Long noncoding RNA ADINR regulates adipogenesis by transcriptionally activating C/EBPalpha. *Stem Cell Reports*, 2015, **5**(5): 856-865
- [11] Chen J, Liu L, Wei G, *et al.* The long noncoding RNA ASNR regulates degradation of Bcl-2 mRNA through its interaction with AUF1. *Scientific Reports*, 2016, **6**: 32189
- [12] Harrow J, Frankish A, Gonzalez J M, *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res*, 2012, **22**(9): 1760-1774
- [13] Tripathi V, Ellis J D, Shen Z, *et al.* The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell*, 2010, **39**(6): 925-938
- [14] Gutschner T, Hammerle M, Eissmann M, *et al.* The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Research*, 2013, **73**(3): 1180-1189
- [15] Tseng Y Y, Moriarity B S, Gong W, *et al.* PVT1 dependence in cancer with MYC copy-number increase. *Nature*, 2014, **512**(7512): 82-86
- [16] Xing Z, Lin A, Li C, *et al.* lncRNA directs cooperative epigenetic regulation downstream of chemokine signals. *Cell*, 2014, **159**(5): 1110-1125
- [17] Calviello L, Mukherjee N, Wyler E, *et al.* Detecting actively translated open reading frames in ribosome profiling data. *Nat Methods*, 2016, **13**(2): 165-170
- [18] Hao Y, Zhang L, Niu Y, *et al.* SmProt: a database of small proteins encoded by annotated coding and non-coding RNA loci. *Briefings in Bioinformatics*, 2017(DOI: 10.1093/bib/bbx005)
- [19] Andrews S J, Rothnagel J A. Emerging evidence for functional peptides encoded by short open reading frames. *Nature Reviews Genetics*, 2014, **15**(3): 193-204
- [20] Jackson R J, Hellen C U, Pestova T V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol*, 2010, **11**(2): 113-127
- [21] Pelletier J, Sonenberg N. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature*, 1988, **334**(6180): 320-325
- [22] Jang S K, Krausslich H G, Nicklin M J, *et al.* A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during *in vitro* translation. *Journal of Virology*, 1988, **62**(8): 2636-2643
- [23] Mokrejs M, Vopalensky V, Kolenaty O, *et al.* IRESite: the database of experimentally verified IRES structures (www.iresite.org). *Nucleic Acids Res*, 2006, **34**(Database issue): D125-130
- [24] Hellen C U, Sarnow P. Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes & Development*, 2001, **15**(13): 1593-1612
- [25] Crick F. Central dogma of molecular biology. *Nature*, 1970,

- 227(5258): 561–563
- [26] Li Z, Huang C, Bao C, *et al.* Exon-intron circular RNAs regulate transcription in the nucleus. *Nat Struct Mol Biol*, 2015, **22**(3): 256–264
- [27] Memczak S, Jens M, Elefsinioti A, *et al.* Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*, 2013, **495**(7441): 333–338
- [28] Hansen T B, Jensen T I, Clausen B H, *et al.* Natural RNA circles function as efficient microRNA sponges. *Nature*, 2013, **495**(7441): 384–388
- [29] Kos A, Dijkema R, Arnberg A C, *et al.* The hepatitis delta (delta) virus possesses a circular RNA. *Nature*, 1986, **323**(6088): 558–560
- [30] Chen C Y, Sarnow P. Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs. *Science*, 1995, **268**(5209): 415–417
- [31] Abouhaidar M G, Venkataraman S, Golshani A, *et al.* Novel coding, translation, and gene expression of a replicating covalently closed circular RNA of 220 nt. *Proc Natl Acad Sci USA*, 2014, **111**(40): 14542–14547
- [32] Yang Y, Fan X, Mao M, *et al.* Extensive translation of circular RNAs driven by N6-methyladenosine. *Cell Research*, 2017, **27**(5): 626–641
- [33] Legnini I, Di Timoteo G, Rossi F, *et al.* Circ-ZNF609 is a circular RNA that can be translated and functions in myogenesis. *Mol Cell*, 2017, **66**(1): 22–37 e29
- [34] Pamudurti N R, Bartok O, Jens M, *et al.* Translation of CircRNAs. *Mol Cell*, 2017, **66**(1): 9–21 e27
- [35] Chen X, Han P, Zhou T, *et al.* circRNADb: A comprehensive database for human circular RNAs with protein-coding annotations. *Scientific Reports*, 2016, **6**: 34985

The Complexity of RNA Translation: Non-translation, Part-translation, De Novo-translation and Over-translation*

HAO Ya-Jing, LUO Jian-Jun, ZHANG Bao, CHEN Run-Sheng**

(Key Laboratory of RNA Biology, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China)

Abstract The central dogma of molecular biology describes the flow of genetic information within the biological system. According to the classical central dogma, RNA can only play its role in life through translation. With the development of molecular biology, rapidly growing evidence shows that RNAs perform a vast array of functions within organisms even without translation. Moreover, RNA can fold into different types aside from linear RNA. Here, we summarized the four forms of RNA for their post transcriptional fate: non-translation, part-translation, *de novo*-translation and over-translation. Non-translation means the RNAs can't be translated to proteins as many non-coding RNAs do. Part-translation describes a phenomenon that the RNA can't translate into a peptide chain using all the information as a template, but only a part of it. Whereas the *de novo*-translation refers to the classical translation process. Over-translation means the RNA can encode a protein using more than their own genetic information, which only occurs in circular RNAs. The diversity of RNA's fate extends largely our understanding of translation above the classical central dogma. Further understanding of the fate of RNA post transcription will put forward more challenging demand on the research for the functions of RNA, and meanwhile, will also present the potential solutions for comprehensive understanding of the functions of RNA.

Key words RNA translation, non-translation, part-translation, *de novo*-translation, over-translation

DOI: 10.16476/j.pibb.2017.0101

* This work was supported by a grant from The National Natural Science Foundation of China (31520103905, 91440116).

**Corresponding author.

Tel: 86-10-64888543, Fax: 86-10-64871293, E-mail: rschen@ibp.ac.cn

Received: March 16, 2017 Accepted: June 27, 2017