

RNA 编 接

陈 丁 丁

(中国药科大学药理研究室, 南京 210009)

提 要

本文概要介绍了锥虫基因表达与调控的一种普遍而独特的方式: RNA 编接 (RNA editing)。初级转录本经 RNA 编接系统加工后, 成熟 RNA 中插入了数目不等的、非基因组编码的尿嘧啶核苷酸残基 (U) 并切除了某些由基因组编码的 U 序列。编接插入的“额外”U 序列可占成熟 RNA 的 60% 左右。这一现象的发现, 向人们提出了这样的问题: 遗传信息都存在于基因组 DNA 序列中吗? 是否还存在于其它形式的遗传物质?

关键词 锥虫, RNA 编接, 转录与加工

自从 1981 年 Cech 发现 ribozyme 以来, RNA 已成为分子生物学的焦点之一。对分子生物学家来说, 锥虫 (trypanosome), 特别是在锥虫的 RNA 代谢中, 充满了各种使他们惊奇不已的分子事件。第一个发现是, 基本上所有锥虫 mRNA 都具有一段共同的前导序列或微小外显子 (mini-exon), 后者是通过一种不寻常的反式剪接 (trans splicing) 即不连续转录连接到 mRNA 的 5'-端上去的^[1,2]。这一反式剪接加工方式现已众所周知, 并继而在其它生物系统中如线虫和牛痘病毒中发现^[3,4]。

然而, 在锥虫中还有第二个更令人惊奇的现象, 这就是 RNA 的“editing”^[5,6]。根据其涵义, 以及与剪接 (splicing) 一词译名相呼应, 作者将 RNA editing 译作 RNA 编接。RNA 剪接是指在初级转录本的基础上, 进行选择性的剪切, 去除内含子部分后, 重新连接为成熟 RNA 的加工过程, 没有添加额外的新信息和新顺序。与电影制作时的蒙太奇技术有某些相似之处。而 RNA 编接则是指除来自初级转录本的序列部分外, 成熟 RNA 中还有额外添加的新信息和新顺序, 并切除了某些由基因组编

码的序列。既增加“新”信息, 又删除原有的“旧”内容, 与出版编辑工作形式有雷同之处。

RNA 编接是锥虫科线粒体进行转录时的一种共同特征。1986 年在布氏锥虫 (*T. brucei*) 和簇生型短膜虫 (*Crithidia fasciculata*) 中首次发现了 RNA 编接现象, Benne 等人的研究表明^[7], 在它们的线粒体细胞色素 c 氧化酶亚基 II (COII) 基因的转录物 5'-端附近, 含有 4 个非基因组编码的 U 残基。进一步的研究证实, 由于这些额外编加的 U 序列的出现, 恰好形成完整的开放阅读框 (ORF), 从而抑制了该基因分子内的移码, 产生有活性的细胞色素 c 氧化酶亚基 II。1987 年 Feagin 等指出^[8], 布氏锥虫细胞色素 b 蛋白 (CYb) mRNA 的 5'-端含有 34 个非基因组编码的 U 序列, 后者出现的结果是向 mRNA 中引入了 AUG 起始密码子。随后发现, 这些有机体的线粒体中, 有许多转录物都显示出进行了某种程度的编接, 在它们的 5'-端或接近 5'-端处, 编接插入了一定数量的额外 U。

RNA 编接系统的发现向人们提出了某些重大概念上的问题, 如遗传信息是否都存在于

基因组序列中? 指导 RNA 编接过程的遗传信息是什么? 和 RNA 编接系统的起源及其进化意义等。早先人们认为, 由于 RNA 编接插入的是数量有限且位置相对固定的 U 序列, 存在范围亦有限, 因而对这些问题并非不能作出解释^[5-7]。然而, 不久前 Feagin 等^[9]和 Shaw 等^[10]共同研究发现的事实使问题变得复杂起来。RNA 编接现象可能比原先认为的更为普遍和多样化。

虽然以前在某些锥虫中发现过细胞色素 c 氧化酶亚基 III(COIII) 基因, 但是一直未能在布氏锥虫细胞核或线粒体中找到该基因。Feagin 等^[9]用 *C. fasciculata* 和 *L. tarentolae* 的 COIII 基因保守区域的一组寡核苷酸探针, 终于探明布氏锥虫的 COIII 基因存在于线粒体基因组中。他们随后分离鉴定了该基因的 mRNA, 并通过 RNA 和 cDNA 的顺序测定, 确定了该 mRNA 约 65% 的序列。令人吃惊的是, 在被测 mRNA 序列中, 约 60% 核苷酸是由 RNA 编接系统添加的额外 U 序列, 这些大量的编接信息并不出现在基因组中。此外, 不象其它已描述过的 RNA 编接那样, COIII 转录物的编接发生于基因的 3'-端, 而且遍及被测区域的三分之二。由一个编接的非编码 U 残基参与形成了 UAA 顺序, 很可能就是该基因的终止密码子。因此, 很明显且不可思议的事实是, 为 COIII 编码的信息中有一半以上是而非线粒体 DNA 携带的, 也不出现在核 DNA 中。这些“额外”的编接序列似乎无源而至, 它们到底从何而来? 难道存在非核酸性的遗传物质吗?

由 Feagin 和 Shaw 等发现的第二个意想不到的结果是^[9,10], 在锥虫中的 RNA 编接不仅涉及插入, 而且还有将基因组编码的 U 从转录物中切除移去的过程。这已在 COIII 基因转录物和三种锥虫的另外两个基因转录物中被证实。这样, RNA 编接的本质是一有增也有减的过程, 与其名称 RNA editing 就完全一致了。

但是这些发现又产生了更多的问题。例

如, 编接是如何进行的并发生于生成 mRNA 的哪个阶段? 虽然编接的机制基本上还是个秘密, 但是许多证据表明^[6,8,9], RNA 编接发生于转录后, 而且是从 RNA 分子的 3'-端向 5'-端进行的。因此 Feagin 等根据与假定的编接中间物相互补的 cDNA 结构指出, 编接可能至少包括两步反应: 首先插入不同数量的 U, 和随后通过调整修剪 (trimming) U 区形成大小合适的产物。

RNA 编接发生于转录后的最有力证据是检测到未经编接的基因初级转录物。当用编接后 mRNA 的探针与线粒体 RNA 作 Northern 吸印杂交时, Feagin 和 Shaw 发现既有全长的 mRNA 也有一些较小杂交分子的印渍。他们推测后者就是经过部分编接的中间物。从对应于这些推测的中间物的 cDNA 顺序看, 它们似乎反映了不同长度的编接中间物, 并且揭示编接是从转录物的一端即 3'-端开始的。虽然还没有确凿的证据证明这些 cDNA 忠实地代表了编接中间物, 但是却由此提示编接是从 poly(A) 化的转录物 3'-端逐步向 5'-端进行的。

对编接中间物的分析也表明, 跟随向转录物插入 U 序列后的步骤, 可能是一调整修剪的过程。Feagin 等指出, 许多 CYb 的 cDNA 5'-端 T 区长度比成熟 mRNA 3'-端的 U 区长。因此可以推断, 这些 cDNA 代表了还未经过修剪的中间物。如果编接确实是一个两步加工过程, 那么根据修剪反应, 对上述从转录物中切除遗传编码的 U 序列现象, 就是可以理解的了。

关于决定编接过程高度专一性的机制问题基本上还是个谜。为什么某些转录物被编接而另外一些则否? 为什么编接具有多样性? 编接加工的信号是建立在初级转录物的顺序上, 还是具有拷贝预先存在已编接过的分子的机制? 对编接后的转录物进行序列分析, 无论对于插入还是切除 U 的过程, 都没有揭示任何规律。顺序专一性的唯一提示, 是观察到转录物被编接基因的有意义链强烈偏重于含 C 而不是 G 残基。这也许说明在编接位点对嘌呤的某种需

要。假如编接过程确实包括U残基的插入和修剪缺失这样的独特步骤,那么专一性作用对两个步骤都是必须的。因为如已指出的,U残基插入起始于转录物的3'-端。故核糖体不可能在这一过程中发挥什么作用,然而它们却可能与修剪过程有关。

关于RNA编接的作用意义和机制,由两方面的观察提供了某些潜在性线索。首先,全部已知转录物被编接的五个基因,除一个外,都缺少适当的完整起始密码子ATG。这可能表明这些不能正确起始翻译的转录物需要并可被编接系统利用,以生成成熟的mRNA。因此,编接过程是基因表达的一种转录后调控机制。第二个更引人注目的发现是,在布氏锥虫中的编接过程本身亦受到发育的调节控制。仅仅在布氏锥虫的昆虫发育阶段,氧化途径才是其全部的功能活动。进入哺乳动物血液发育阶段(bloodstream stage)后,由于缺乏线粒体,所有的ATP只能由酵解提供,Feagin和Stuart测定了在其血液发育阶段,布氏锥虫线粒体各种不同的mRNA发生时间和形式^[11]。有趣的是,他们发现某些mRNA(COII和CYb)是以血流形式(bloodstream form)但不是编接形式出现的,血流形式即所谓的树桩形式(stumpy form),它们一般是在血液发育阶段后期为随后的人侵昆虫阶段而预备产生的。他们测定的其它mRNA则两种形式都有。这种差别说明,编接本身是受到发育的调节控制的,编码编接系统的基因在锥虫的各个发育阶段都是开放表达的,而且编接加工很可能在基因的表达和调控以及细胞的生长发育过程中起着某种重要的调节作用。即编接过程原初可能是作为一种翻译调节机制而进化发展起来的,但随后又进一步被发展,用于克服累积突变对个体生长发育的不良影响。

类似于锥虫RNA编接的现象,在哺乳动物apolipoprotein-B基因和副粘液病毒SV5(paramyxovirus SV5)的P、V蛋白及麻疹病毒(measles virus)P蛋白基因的研究中已经发现。研究表明,人的apolipoprotein-B单拷

贝基因产生两种类型的mRNA转录物,其中一种在位点2153处用一U取代了DNA编码的C,结果向此mRNA中引入了一个终止密码子,导致了Apo-B48蛋白质的合成^[12,13],鼠肝细胞也是通过相同的机制产生Apo-B48蛋白质的,而且鼠肝中的这一过程受甾体激素的调控^[14,15]。副粘液病毒的P和V蛋白由独立的两种mRNA编码,它们之间的区别仅仅在于,在编码P蛋白的mRNA中插入了两个非基因组编码的鸟嘌呤核苷酸残基(G)^[16]。在麻疹病毒编码P蛋白的约50% mRNA中,由于转录后插入了一个G残基,结果产生的蛋白质,由P蛋白的氨基端区域与一富含半胱氨酸残基的片段连接而成,后者由存在于该mRNA中的一隐秘阅读框(cryptic reading frame)编码^[17]。因此,RNA编接现象可能比原先猜想的存在范围更加广泛和普遍。

参 考 文 献

- 1 Borst P. *Ann Rev Biochem*, 1986; 55: 701
- 2 陈丁丁. 生命的化学, 1988; 8(6): 25
- 3 Sharp P A. *Cell*, 1987; 50: 147
- 4 Krause M, Hirsh D. *Cell*, 1987; 49: 753
- 5 Eisen H. *Cell*, 1988; 53: 331
- 6 Benne R. *Biochim Biophys Acta*, 1989; 1007: 131
- 7 Benne R, Van Den Burg J, Brakenhoff J P J *et al. Cell*, 1986; 46: 819
- 8 Feagin J E, Jasmer D P, Stuart K. *Cell*, 1987; 49: 337
- 9 Feagin J E, Abraham J M, Stuart K. *Cell*, 1988; 53: 413
- 10 Shaw J M, Feagin J E, Stuart K *et al. Cell*, 1988; 53: 401
- 11 Feagin J E, Stuart K. *Mol Cell Biol*, 1988; 8: 1259
- 12 Powell L M, Walls S C, Pease R J *et al. Cell*, 1987; 50: 831
- 13 Chen S H, Habib G, Yang C Y *et al. Science*, 1987; 238: 363
- 14 Davison N O, Powell L M, Wallis S C *et al. J Biol Chem*, 1988; 263: 13482
- 15 Tennyson G E, Sabatos C A, Higuchi K *et al. Proc Natl Acad Sci USA*, 1989; 86: 500
- 16 Thomas S M, Lamb R A, Paterson R G. *Cell*, 1988; 54: 891
- 17 Cattaneo R, Kaelin K, Baczko K *et al. Cell*, 1989; 56: 759

【本文于1990年1月6日收到,9月10日修回】