

内含子与基因表达

唐红生* 洪剑明 邱泽生

(首都师范大学生物系 北京 100037)

基因是负载遗传信息的 DNA 片段。绝大多数真核生物的基因都是不连续基因,即可表达的基因序列(exon,外显子)被通常不可表达的介入序列(intervening sequence 即 intron 内含子)所介断,形成编码序列和不编码序列相间排列的情况。 n 个内含子排列于 $n+1$ 个外显子之中,故有内和外之称。

基因的表达是一个非常复杂的过程,包括转录和转译两个水平。内含子通常在转录后的加工过程中被去除。然而,现已证明内含子不仅在这两个水平上对基因表达具有影响,更为重要的是,含可读框架(open reading frame ORF)的内含子,其 ORF 序列可以表达为酶,而这些酶又对基因的表达具有重要的或潜在的影响。因此,内含子与基因表达的关系主要是指内含子 ORF 基因本身的表达和内含子对基因表达的影响两个方面。本文首次从这一角度对内含子的重要作用作一简述。

内含子根据保守序列组分、二级结构和拼接(splicing)机制可分为 I 类内含子(group I intron)和 II 类内含子(group II intron)^[1-3]。III 类内含子(group III intron)和双生内含子(twintrons 即 introns-within-introns)的概念也早已明确^[4,5]。此外,还有 rRNA 内含子和 tRNA 内含子^[2]。Scott O. 等首次报道半知菌类(deuteromycetes)的核 18 S rDNA 基因有 mRNA 型内含子^[6]。

植物和动物内含子是相似的,在其 5' 和 3' 拼接点均含有短而高度保守的序列纹区(motif)。然而动物内含子和植物内含子亦有许多重要的区别^[7]:动物内含子含有多聚嘧啶束(polypyrimidine tract)和松散的保守分支点序列;动物内含子长度范围从 80 nt 到大于

100 kb,平均为 1127 nt,而植物内含子较短,长度范围是 80 nt 到 2 kb,平均为 183 nt 且富含 AU 序列;较长的动物内含子一般分布于较短的外显子中,而较短的植物内含子通常由较长的外显子所包围。以前认为动物和植物内含子有效拼接的最小长度是 70—80 nt^[7],最近,美 Chris B. R. 等发现了已知最短只有 20—33 个标准核酸序列的草履虫(*P. tetraurelia*)的内含子^[8]。

一、内含子 ORF 的表达

早在 70 年代,两位欧洲学者报道了内含子中 4 种可编码逆转录酶的序列^[9]。以后将内含子中这些可以表达的序列亦称之为 ORF。现已知 ORF 编码的蛋白质有三方面的功能:(1) 逆转录酶活性,并在内含子移动(intron mobility)中起作用^[10]。有报道此酶对内含子和相邻外显子高度专一^[11]。(2) 成熟酶(maturases)活性,在所编码的内含子的拼接中起作用^[10]。(3) 专一位点核酸内切酶活性,其切点即为不含内含子基因组的内含子插入位点或此点附近^[12]。原核生物此酶的切点位于内含子插入位点的一定距离上^[13]。

Bernd S. 等首次报道酵母细胞色素 c 氧化酶亚基 I 基因的第一个内含子编码成熟酶并刺激在大肠杆菌中的同源重组;该作者还证明此内含子在杂交过程中是移动的,说明它也编码了核酸内切酶^[14]。据报道酵母 Cox I 基因内含子所编码的核酸内切酶 I-See III,其 20 个碱

本文曾与印莉萍副教授讨论并承赵微平教授审阅,特表谢意。

* 现工作单位: 中国农业科学院生物技术研究中心 北京 100081

基长度的靶识别位点高度专一, 因而此核酸内切酶可能成为大基因组结构分析中有用的工具酶; 通过带转变(band-shift)等实验证明此酶在内含子 RNA 拼接中也具有作用^[10]。John C. 等证明酵母 mtDNA Cox-1 基因中 II 类内含子 al 1 和 al 2 所编码的逆转录酶或以已切除的内含子 RNA 序列为模板, 或以 Cox-1 的 pre-mRNA 为模板起始 al 2 3' 末端 cDNA 的合成, 此酶还对内含子的移动、切除和 RNA 的拼接有作用, 另外其活性与此内含子所编码的成熟酶相关^[11]。这些证据说明了内含子编码蛋白质(酶)的复杂性。

内含子 ORF 的翻译一般与上游外显子同时进行, 或在起始密码子上进行。交互拼接(alternative splicing)是不与上游外显子相连续并缺少起始信号的 I 类内含子基因表达的方式^[10]。

Georg M. 等^[17]比较了 34 个 II 类内含子 ORF 所编码的蛋白质和 4 个非内含子 ORF 编码的蛋白质; 绝大多数蛋白质都有一个逆转录酶区, 此区后是一个 X 功能区(domain X)和一个类锌指区(Zn²⁺-finger-like-region)。一些蛋白质缺少逆转录酶区的关键氨基酸, 一些缺少锌指结构的部分区段, 因而推测普遍具有的 X 区为关键区。X 区在逆转录和 RNA 拼接中具有重要作用^[10]。有报道指出 X 区还与成熟酶的活性相关^[11]。

V. Knoop & A. Brennicke^[18]证明在大肠杆菌质粒内含子中, 高度保守的功能区 IV 和 V (domains IV and V) 是 II 类内含子的典型特征; 功能区 V 上游之 ORF 与 II 类内含子编码的成熟酶同源(homologous), 在所推断的 3' 拼接点下游的可读框架(ORF 104)与农杆菌(*Agrobacterium tumefaciens*)T-DNA 中心部分所编码的一个小多肽高度相似; 该文作者指出这一保守序列(ORF 104)及其在较大的范围内与移动 DNA 序列(mobile DNA sequences)相联系的保守位点都说明了其功能的重要性。

最近, P. du Jardin 等报道了马铃薯叶绿

体 trank 基因内含子 ORF 蛋白质在体内全系列的表达, 并证明此蛋白质具有成熟酶活性。这是首次内含子所编码的多肽在高等植物体内表达的报道^[19]。暗示着高等植物基因内含子, 通过其所编码的酶(成熟酶或核酸内切酶), 对高等植物的基因表达产生影响。

二、内含子对基因表达的影响

真核基因的表达在多水平多因素上受到调控和影响。大量事实表明内含子对基因的表达具有重要作用:

1. 极大地增加基因表达的量

早有报道玉米 Adh1 内含子 1 和 sh1 内含子可以使玉米培养细胞标记基因(marker gene)的表达增强 100 倍; 玉米 Bz1 和 phyA 内含子则可增强标记基因 5—6 倍的表达^[20]。相似的报告还有, 玉米 sh1 内含子刺激羊草和紫狼尾草培养细胞的表达; 玉米 Adh1 内含子 1 刺激水稻培养细胞 GUS 酶的表达; 以上作用不只局限于单子叶植物, 茄中的一个内含子刺激菜豆 CAT 嵌合基因 3 倍增长量的表达^[20]。相反的例证为 Suan R. 等将拟南芥(*Arabidopsis*)之多聚泛汰素(polyubiquitin)基因 UBQ3、UBQ10 和 UBQ11 内含子中的起动机序列去除, 代之以相应邻近 cDNA 序列(flanking cDNA sequence), 结果均稳定地引起标志酶(GUS、LUX)表达活性降低了 2.5—3 倍^[20]。

内含子对基因表达的增强作用, 对一些基因来讲, 是由于前体 mRNA 中的内含子增加了细胞质中的 mRNA, 从而在转录水平上增加基因的表达^[7]。如矮牵牛 rbc 内含子增加烟草细胞中稳定的 mRNA 的量达 5 倍之多^[20]。在另一些基因中, 内含子的增强作用源于增加了基因转录的起始反应, 因为增强子(或部分序列)即包含于内含子中^[7]。总之, 内含子由于含有增强子的组分, 或由于拼接体(spliceosome)及多聚腺苷酸化作用(polyadenylation)保护前体 mRNA 免受 RNase 作用^[7], 亦或由于目前还未知的机制, 增加了基因的表达。

2. 内含子序列中含有重要的调控基因表达的元件

内含子除含有增强子(或部分序列)以增加基因表达外,有证据^[21]表明鼠B细胞K链基因内含子的增强子序列通过诱导去甲基化和促进组织专一转录而调节B细胞的分化。Alexander G.等通过转基因鼠证明免疫球蛋白k基因内含子对调控体细胞超突变(somatic hypermutation)具有重要作用^[22]。

鼠hprt基因在胚胎干细胞中的表达需要内含子1中一个位置和方向(-230 bp)依赖性元件(长度420 bp),推测这一元件是hprt启动子整合的下游部分^[23]。

人apoB基因在转基因鼠中的高水平、组织专一的表达除需增强子外还必需其第二个内含子中的一些序列^[24]。

A. Graessmann等^[25]对单纯疱疹病毒(HSV)t k编码区和SV 40病毒T-抗原内含子进行体外单一位点甲基化,结果均导致基因的失活。该作者分析DNA甲基化影响了转录的延长而不是起始;转录衰减位点极有可能位于内含子上,从而抑制了转录的延长。逐渐多的证据表明,在哺乳动物细胞中转录的延长也是一个重要的调节步骤^[25,30]。

3. 内含子的拼接是转录后加工的重要内容,并直接在翻译水平影响基因的表达。

基因的线性表达被内含子所中断发生于三个层次:翻译前-RNA拼接,翻译时-mRNA跳跃,翻译后-蛋白质拼接^[26]。此方面文献较多,本文不作冗述。

4. 丝状真菌(*P. anserina*)的衰老过程与其线粒体内含子DNA的缺失相关。

Annie S.等^[27]利用PCR技术证明:(1)衰老和幼嫩培养种所含DNA分子都准确地缺失了内含子的序列。(2)衰老种比幼嫩的培养种缺失的程度大得多。(3)DNA内含子的缺失可能是一逆转录酶介导的结果。(4)内含子 α 编码的蛋白质可以通过增高线粒体逆转录酶活性调节这一过程。

5. 衣藻(*C. reinhardtii*)psbA基因缺失内含子的突变体,基因型虽无明显影响,但可能影响细胞适应环境改变的灵活性和其他一些生理过程,例如DNA重组的调节过程和psbA基因前体mRNAs的有效拼接过程^[28]。

6. 由于I类内含子在地衣真菌(*C. chlorophaea*)SSU rRNA基因高保守区的数目位置和序列的不同,引起其种群个体间和个体中小亚基rDNA变异^[29]。

7. 人EGFR基因染色质结构分析表明其内含子1序列在细胞致乳癌的过程中具有作用^[30]。

综上所述,以前认为含而不显的内含子不仅可以表达,而且对基因表达和相关的生理过程有重要影响。这也再次提醒我们由核酸序列到氨基酸序列的基因表达过程的复杂性。我们知道,任何碱基序列的改变都可能导致巨大的变异,而如此之多的内含子序列绝不可能只是被简单地去除,其内必然隐藏着许多我们所未知的作用。随着今后的研究进展,内含子的重要作用将会得到更为深刻的认识。

摘 要

以前认为含而不显的内含子,现已证明其内藏的可读框架(ORF)不仅可以表达为成熟酶、逆转录酶和核酸内切酶,而且这些酶又对基因的表达具有重要的或潜在的影响。内含子可以极大地增强基因的表达,在转录和翻译水平上影响基因的表达过程;内含子序列中还含有调控基因表达的元件;还有报道指出内含子还与衰老过程、rDNA群体和个体变异、对环境的适应性以及癌症等相应的生理过程有关。

参 考 文 献

- [1] Martin W. et al., 1993, *Current Genetics*, 24: 301—306.
- [2] 孙乃恩等, 1993, 分子遗传学, 217页.
- [3] Eiji O. et al., 1993, *Nucleic Acids Res.*, vol. 21, 5: 1297—1305.
- [4] Robert G. D. et al., 1993, *Nucleic Acids Res.*, vol. 21, 10: 2389—2394.

- [5] Donald W. C. et al., *Nucleic Acids Res.*, vol. 22, 6: 1029—1036.
- [6] Scott O. R. et al., 1993, *Current Genetic*, 23: 338—342.
- [7] Kenneth R. L. & Wirginia W. 1993, *Plant Mol. Biol.*, 24: 449—463.
- [8] Chris B. R. et al., 1994, *Nucleic Acids Res.*, vol. 22, 7: 1221—1225.
- [9] 程惟丽, 1987, 生命的化学, 7: 3 第32页.
- [10] Aeorg M. et al., *Nucleic Acids Res.*, 1993, vol. 21, 22: 4991—4997.
- [11] John C. K. et al., 1993, *Cell*, vol. 73, 133—146, April 9.
- [12] Bertrand S. et al., 1992, *Gene*, 113: 1—8.
- [13] Mary B. et al., 1993, *J. EMBO.*, vol. 12, 5: 2141—2149.
- [14] Bernd S. et al., 1994, *Current Genetic*, 25: 336—341.
- [15] M. Schapira, 1993, *Nucleic Acids Res.*, vol. 21, 16: 3683—3689.
- [16] Carole H. S. et al., 1994, *Nucleic Acids Res.*, vol. 22, 7: 1135—1137.
- [17] G. Mohr et al., 1993, *Nucleic Acids Res.*, vol. 21, 22: 4991—4997.
- [18] V. Knoop & A. Brennicke, 1994, *Nucleic Acids Res.*, vol. 22, 7: 1167.
- [19] P. du Jardin et al., 1994, *Current Genetic*, 25: 158—163.
- [20] Suan R. N. et al., 1993, *Plant Mol. Biol.*, 21: 895—906.
- [21] M. Lichtenstein et al., 1994, *Cell*, vol. 76, March 11: 913—923.
- [22] Alexander G. B. et al., 1994, *Cell*, vol. 77, April 22: 239—248.
- [23] Alan R. B. et al., 1994, *Mol. & cellu. Biol.*, April: 2234—2256.
- [24] Thomas M. M. 1992., *Gene*, 122: 289—296.
- [25] A. Graessmann et al., 1994, *Mol. & Cellu. Biol.*, March: 2004—2010.
- [26] 蔡宝立, 张富国, 1994, 生命的化学, 14(2): 21—23.
- [27] Annie S. C. et al., 1993, *J. Mol. Biol.*, 234: 1—7.
- [28] Udo J. & Silvia H. 1993, *Plant Mol. Biol.*, 22: 91—99.
- [29] Paula T. D. 1993, *Gene*, 134: 67—74.
- [30] S. A. Chrysogelos et al., 1993, *Necleic Acids Res.* vol. 21, 24: 5736—5741.

牛磺酸跨膜转运的意义及机制

宿燕岗 杨英珍 陈灏珠

(上海市心血管病研究所 上海医科大学附属中山医院 200032)

有关氨基酸跨膜转运早已研究很多, 根据是否依赖于 Na 及是否受膜电位、pH 或阴离子的影响, 已认可十余种独立的氨基酸转运系统, 包括 A 型、ABC 型、C 型和 G 型等。但对心肌膜上氨基酸跨膜转运却了解很少, 尤其对心肌细胞内最丰富游离氨基酸的牛磺酸 (taurine) 了解更少。研究牛磺酸与心脏, 中心问题是研究心肌细胞膜通过何种机制建立起细胞内外的浓度差及其对心脏的作用。近年来有关此方面的研究取得了很多进展, 本文就有关牛磺酸跨膜转运的意义及机制等方面的研究进展综述如下。

一、牛磺酸跨膜转运的意义

牛磺酸是正常存在于体内的含硫氨基酸, 有着广泛的生理和药理作用^[1]。近年来有关牛磺酸与心脏方面的研究愈来愈受到重视。牛磺酸占心肌游离氨基酸含量的 50% 以上, 它主要存在于细胞内。在不同的哺乳类动物, 心肌细胞内外牛磺酸含量比自 50—200:1 不等。虽然牛磺酸的基本生理功能尚不十分明确, 但根据已有的实验资料证实, 牛磺酸跨膜转运对渗透

感谢中国科学院上海细胞生物学研究所 顾全保副研究员在本文撰写中提出的宝贵意见。