

起始点的角色这个话题又重新提起。从前面介绍的实验结果看,小肽能形成 α 螺旋, β 折叠, β 转角,也能形成疏水簇。近程、中程作用力足以维持它们稳定存在,二硫键还原及许多变性条件都不能破坏它们的有序结构。所以,Creighton认为,尽管现在没有证据表明伸展态中的有序结构对卷曲有动力学上的重要性(如RNase A中N端稳定螺旋仍未显示动力学上的重要性),它们的作用是不能忽视的。就象BPTI中能偏向30-51二硫键生成的那个有序结构一样,一旦这些有序结构与其它的和卷曲起始相关的构象或特殊作用偶联起来,它们对于卷曲起始来说就会显得非常重要。

早些时候,Ptitsyn经过热力学及动力学的计算指出, α 螺旋可以预先存在于松散的肽链中作为结构胚胎启动蛋白质卷曲。但他认为 β 折叠与 β 转角由于没有预先存在于伸展态结构中而不参与启动卷曲。现在看来,这后一观念也被实验结果更新^[4]。

参 考 文 献

- 1 Jaenicke R. Folding and association of proteins. *Prog Biophys Mol Biol*, 1987; **49**: 117
- 2 Kim P S, Baldwin R L. Intermediate in the folding reactions of small proteins. *Annual Rev Biochem*, 1990; **59**: 631
- 3 Tsou C L. Folding of the nascent peptide chain into a biologically active protein. *Biochemistry*, 1988; **27**: 1809
- 4 Wright P E, Dyson H J, Lerner R A. Conformation of peptide fragments of proteins in aqueous solution; Implications for initiation of protein folding. *Biochemistry*, 1988; **27**: 7167
- 5 Hughson F M, Wright P E, Baldwin R L. Structural characterization of a partly folded apomyoglobin intermediate. *Science*, 1990; **249**: 1544
- 6 Oas T G, Kim P S. A peptide model of a protein folding. *Nature*, 1988; **336**: 42
- 7 Dill K A, Shortle D. Denatured states of proteins. *Annu Rev Biochem*, 1991; **60**: 795
- 8 Richards F M. The protein folding problem. *Scientific American*, 1991; **264** (1): 34

真核核基因 RNA 剪接研究的新进展

徐 军 陈润生

(中国科学院生物物理研究所, 北京 100101)

提 要

简述了真核生物核基因 RNA 剪接研究的一些新进展, 主要包括剪接体的结构组成, 剪接体的逐步组装, 剪接反应的影响因素等方面。

关键词 剪接体, snRNA, PRP, 预组装

现已发现的绝大多数真核蛋白质基因是不连续的, 编码蛋白的序列被不编码的序列所间隔。前者称作外显子(exon), 后者称作内含子(intron), 转录时先生成一条包括有内含子的前 mRNA(pre-mRNA), 然后经过剪接得到成熟的 mRNA, 可以作为模板翻译蛋白质。

剪接由剪接体完成。剪接体由特定的分子 RNA 和蛋白质构成。

剪接是基因表达过程中一个非常重要的环节, 对于它的逐步了解大大丰富了我们在分子这一层次上对生命活动的认识, 同时内含子的起源和存在的生物学意义是一个极其诱人的课题, 对它的研究将推进我们对于整个基因组结构规律的认识。

1 内含子的结构

真核生物核基因 mRNA 绝大多数内含子都遵循 GU-AG 规则：5' 剪接位点为 GU，3' 剪接位点为 AG。距 3' 端 15-40 核苷酸处为分枝点 (branch point) 序列 (图 1)。酵母中为 5' UACUAA* C3'，其中 A* 高度保守；哺乳动物中分枝点序列较离散。分枝点与 3' 剪接位点间有一段保守的嘧啶区，它对于哺乳动物来讲至关重要：一可能是帮助识别分枝点，二可能是帮助识别 3' 剪接位点。

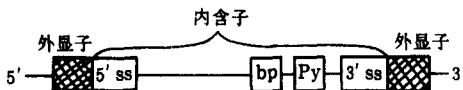


图 1 内含子的结构

bp: 分枝点; Py: 嘧啶区; 5' ss: 5' 端剪接位点; 3' ss: 3' 端剪接位点

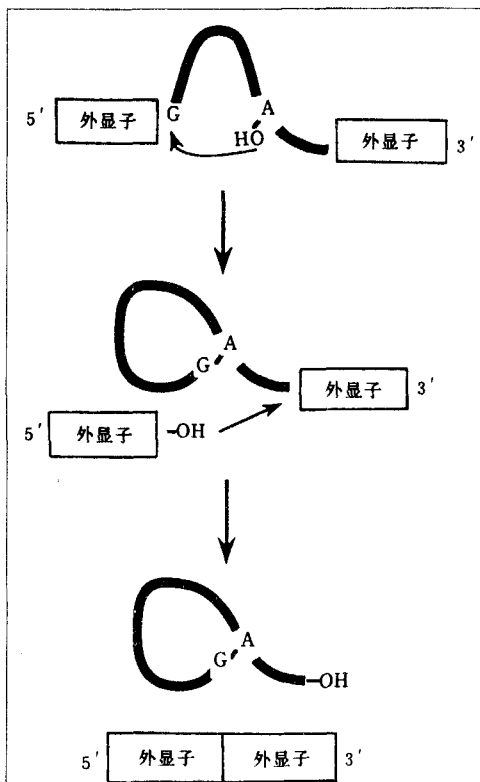


图 2 剪接反应由两步转酯反应构成

2 剪接反应

剪接反应分两步完成。第一步由分枝点 A

的 2' 羟基亲核攻击 5' 剪接位点 G，发生转酯反应 (图 2)：G 与第一个外显子的最后一个碱基间的磷酸二酯键断裂，与 A 形成新的磷酸二酯键，即形成“套索” (lariat) 内含子；第二步反应由第一个外显子 3' 端最后一个碱基的 3' 羟基亲核攻击内含子 3' 剪接位点，发生相似的转酯反应，于是两个外显子连起来，内含子则被剪下来，由核酸酶降解掉。

3 剪接体的组成

核基因前 mRNA 的剪接由剪接体完成。通过电子显微镜观察到剪接体呈现 40 到 60nm 大小的微粒并具有特定的形态，剪接体由三部分组成：前 mRNA，核小 RNA (small nuclear RNA, snRNA)，蛋白因子。

3.1 核小 RNA

真核生物细胞中有许多 100—300 核苷酸的小分子 RNA，每个细胞中大约有 10 万到 100 万个。它们有的由 RNA 聚合酶 II 转录，有的由 RNA 聚合酶 III 转录，有的还象 mRNA 一样戴上 5' 帽子。核中的小 RNA 即称作 snRNA。我们感兴趣的主要是 U1, U2, U4, U5, U6 这 5 种 snRNA。

U1, U2, U5 的进化保守程度较差。酵母的 U1, U2, U5 的长度分别是哺乳动物的 3.5, 6.3, 2 倍，它们也比其它真菌中各自的相应物长。U2 中相当一部分序列的缺失不影响功能，而且可以用人的 U2 替代比它长几倍的酵母的 U2 而不发生功能上的改变^[1]。但 U4, U6 则不同，它们的长度甚为保守。U4, U6 间有两段互补配对序列，其中茎 (stem) I 含 8 对碱基，茎 II 含 16 对碱基 (图 3)，U6 这两段序列中 80% 以上的碱基保守。这种能形成螺旋高级结构的序列的高度保守是不寻常的。与 U2 不同，U6 的缺失将阻止剪接的发生。除 U4, U6 外，最近发现 U2 与 U6 间也存在碱基互补配对。

有人从功能上推断 snRNA 可能起源于第二类自我剪接的内含子。

3.2 蛋白因子

蛋白因子分为外在蛋白和整合蛋白。外在

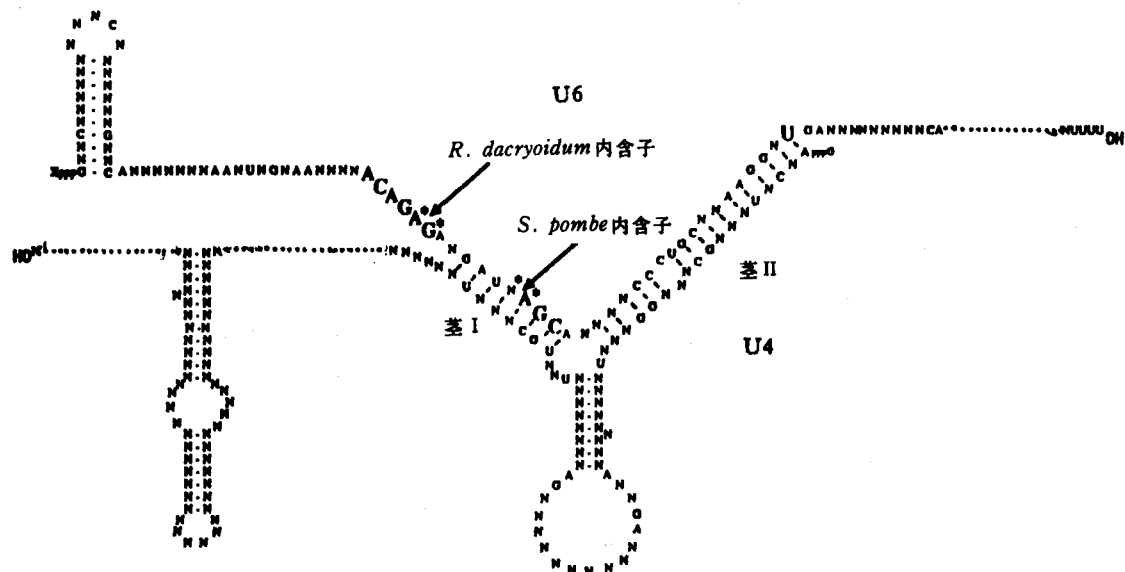


图3 U4与U6间通过碱基互补配对形成两段双螺旋：茎I和茎II

↓ 指示 mRNA 型内含子插入的位置

蛋白在剪接过程的某一时相中加入到剪接体上，完成功能后又释放出来；整合蛋白是指结合到 snRNA 不同部位上的蛋白。结合有整合蛋白的 snRNA 称 snRNP (small nuclear ribonucleoprotein)。若体系中不加入 ATP，仅有 U1 snRNP 能够结合到前 mRNA 底物 5' 剪接位点上去，但是提纯的 U1 snRNA 则无此功能，可见整合蛋白对于 snRNP 功能的实现是不可缺少的。整合蛋白又可分作核心蛋白 (core protein) 与特异蛋白，其中核心蛋白指同时出现在 5 种 snRNP 中的整合蛋白，特异蛋白指出现在不同的 snRNP 中的整合蛋白。哺乳动物中至少有 9 种核心蛋白。核心蛋白对于 snRNPs 高级结构的稳定来讲是必不可少的。特异蛋白 U1 有 3 个，U5 有 7 个甚至更多。特异蛋白的特征结构 (motif) 提示人们这些蛋白间或与 RNA 间的作用影响到 snRNAs 分子间或 snRNA 与内含子间的相互作用。特征结构是指蛋白中特定氨基酸区域，它们决定了该蛋白与其它蛋白或核酸间结合的方式。哺乳动物中发现 U5 特异蛋白 IBP (intron binding protein) 具有与 RNA 相结合的特征结构，它可结合于嘧啶

区和 3' 剪接位点，可能在 snRNA 与内含子间起桥梁作用。

酵母中与前 mRNA 剪接有关的蛋白又称作 PRP (precursor RNA processing) 蛋白。目前至少已分离了 30 多种 PRP 蛋白。但在酵母和人各自的剪接体系中，只发现酵母的 PRP8 与人的 220kD IBP 蛋白间有较高的同源性：前者的抗体可沉淀后者。它们都整合于 U5 snRNA。

PRP 蛋白有的是外在蛋白，有的是整合蛋白。比如 PRP3, PRP4, PRP6 是整合蛋白，整合于 U4-U6 snRNAs。外在蛋白各自在剪接体上出现的时相不尽相同。如 PRP11 只出现在活化状态的剪接体上。有的 PRP 蛋白具有特征结构。如 PRP4, PRP17 有 G 蛋白 β 亚基型特征结构。这一特征结构由一个相同的肽段重复 5 次而形成，可能与蛋白间相互作用及调控有关。PRP6, PRP9, PRP11 中有与锌指类似的特征结构。其中 PRP9 这一特征结构中的 His 或 Cys 发生突变将影响到剪接^[2]。各 PRP 蛋白的功能也不相同。如 PRP17, PRP18, PRP29 在剪接体上可能仅仅起到一种维持特定构象的功

能。

PRP5, PRP28 属于 DEAD Box 蛋白家族。两者都与第一步剪接反应有关, 而且若体系中缺少 PRP5, 剪接体则不能形成。DEAD 蛋白有 7 个保守性区域, 其中一个的氨基酸组成为 Asp-Glu-Ala-Asp, 用单字母符号表示即 DEAD。7 个保守性区域中有两个位于 ATP 结合部位中。真核生物翻译起始因子 eIF4A 是最先为人们广泛研究的 DEAD 蛋白, 它同时具备 ATP 水解活性和 RNA 拓扑异构酶活性。eIF4A 与其它起始因子共同作用, 用 ATP 水解产生的能量解开起始密码子上游的 mRNA 二级结构, 促使核糖体 40S 亚基结合上去。这种 RNA 二级结构的改变是可逆的, 取决于碱基配对状态, 而 RNA 拓扑异构酶能够改变配对状态。

PRP2, PRP16, PRP22 属 DEAH Box 蛋白家族。DEAH 蛋白与 DEAD 蛋白在组成和功能上有许多相似之处 (H 代表 His)。DEAH 蛋白 PRP16 具有 RNA 依赖的 ATP 水解活性。这里的 RNA 不可用单链 DNA 替代。核基因前 mRNA 剪接需要 ATP 提供能量, 但是反应的中间产物及最终产物中均未发现结合有 ATP 磷酸根, 而且反应中磷酸键的数目恒定, 即无须外界提供能量。因此看来 ATP 的作用可能在于辅助反式作用因子结合到底物上去以促使剪接的发生, 而剪接反应本身可能类似于第二类内含子的自我剪接是不需要 ATP 的。ATP 水解酶 PRP16 只结合于包含有套索内含子的剪接体上, 所以 PRP16 可能主要与第二步剪接反应相关, 而且第二步剪接反应的发生很可能要求 ATP 的水解。PRP16 从剪接体上释放是否需要 ATP 水解有待进一步探讨。另外两个 DEAH 蛋白 PRP2, PRP22 各自有两段序列与 PRP16 有较高的同源性: 一段长 320 个氨基酸, 包括尚未证实的 ATP 结合位点, 对于这一段氨基酸, PRP2, PRP22 与 PRP16 的同源性均达到 50%; 另一段是羧端的 280 个氨基酸, PRP2, PRP22 与 PRP16 同源性分别为 42% 和 35%。很可能三者均具有 RNA 依赖的 ATP 水

解活性, 与剪接体发生暂态的高亲和力的相互作用, 改变剪接因子的构象, 促使组装及剪接一步一步向前推进。

4 剪接体的组装

剪接体遵循一定的顺序组装起来 (图 4), 首先体系中必须有 U5 snRNP。在 U5 存在的前提下, U1 通过碱基互补配对与 5' 剪接位点处序列相结合。哺乳动物中配对的长度大约为 9 个碱基, 其中外显子 3' 端 3 个碱基, 内含子 5' 端包括 GU 在内 6 个碱基, 与 U1 5' 端第 3 位到第 11 位碱基配对 (图 5)。

实验发现 U1 结合于 5' 剪接位点要求分枝点-嘧啶区-3' 剪接位点区域序列的完整。如果分枝点处发生突变, 则 U1 的结合能力大为下降。因此推测可能有某一因子 (也可能是 U1 的亚基) 紧随 U1 的结合之后而结合于上述区域作为 U1 与该区域间的中介。哺乳动物中已发现至少有两种蛋白结合于这一区域, 包括前面述及的整合于 U5 的 220kD 的 IBP 蛋白。HeLa 细胞中也发现有 hnRNP 蛋白和 IBP 蛋白结合于 3' 剪接位点附近^[3]。

然后 U2 以与 U1 类似的方式结合于分枝点, 但与保守 A 之间并不互补配对 (图 5)。U2 结合于分枝点首先在酵母中发现, 随后在哺乳动物中也得到了证实。U2 对于分枝点的识别必须以结合于 5' 剪接位点或游离态的 U1 的存在为前提, 而且 U2 的结合要求 ATP 提供能量。DEAD 蛋白 PRP5 参与了这一结合过程, 可能与 ATP 的水解有关。U2 结合于分枝点还可以由突变实验得到佐证。在哺乳动物和酵母中均发现 U2 snRNP 的突变可以不同程度地补偿分枝点序列的突变。

U1 结合到 5' 剪接位点, U2 结合到分枝点, 使得这两段核苷酸在空间上相互接近 (图 4, B), 这是第一步剪接反应发生的必要条件。

U1, U2 分别结合到前 mRNA 上适当位点后, U4-U6, U5 再以三聚体的形式结合到剪接体上。这个三聚体的聚合及解聚与 ATP 相关, 而与前 mRNA 的存在与否无关。在不含前

mRNA 的抽提液中同样可以形成 U4-U5-U6 复合物. U4-U6 及 U4-U5-U6 三元复合物的稳定性与 PRP 蛋白有关. U6 整合蛋白 PRP24 的抗体可以沉淀 U6, U4 与 U6 间结合不稳定时则可同时沉淀 U4 和 U6^[4]. 这可能是由于 PRP24 上抗体结合的位点, 当 U4-U6 结合牢固时被屏蔽, 结合松弛时方显露出来. 故

PRP24 可能有促进 U4, U6 间结合的作用. PRP4 则可能与 U4-U5-U6 三元复合物的稳定性有关. PRP4 的抗体能抑制剪接体的组装^[5]. 在 U4 5' 端环区 (loop) 作缺失突变 (图 3), 虽然能形成 U4-U6 复合物, 但 PRP4 不能与之结合, 也不能形成 U4-U5-U6 复合物.

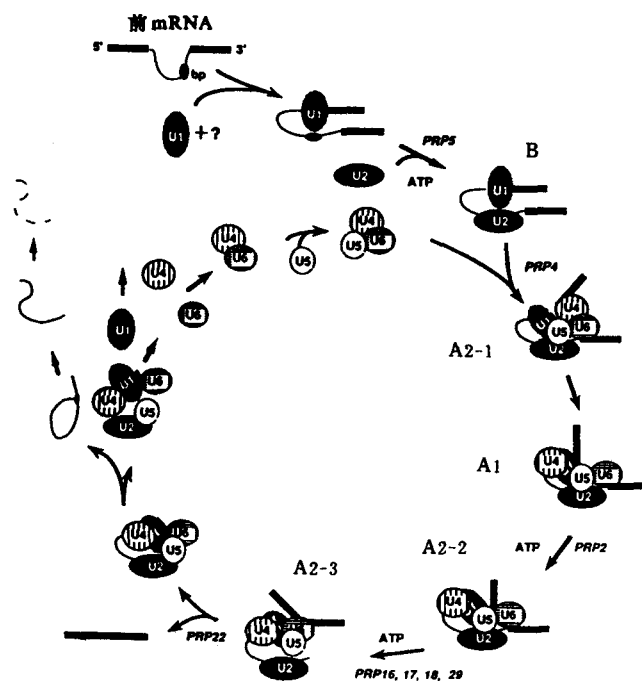


图 4 剪接体组装以及剪接过程

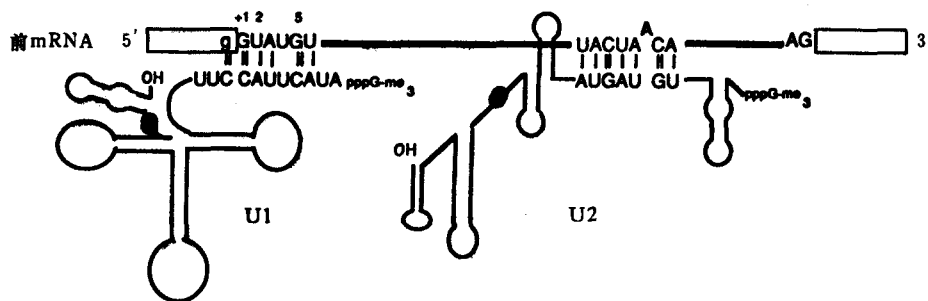


图 5 U1 与 5' 剪接位点间碱基配对, U2 与分枝点间碱基配对

5 剪接反应的影响因素

5.1 影响第一步反应

DEAH 蛋白 PRP2 可能具有 RNA 依赖的

ATP 水解活性. 它对于第一步剪接反应的发生不可或缺^[6] (图 4, A1—A2-2). 温度敏感突变体 prp2-处于热失活状态时, 虽然剪接体可以组装起来, 但不发生剪接反应. 免疫沉淀证明

PRP2 直接或通过某种中介结合于前 mRNA 上,但不出现在剪接产物上.推测 PRP2 在第一步反应前结合上来而在第二步反应前或反应中释放.

DEAD 蛋白 PRP5, PRP28 均与第一步剪接反应有关^[7].通过 PRP28 与 PRP24 双突变发现,PRP28 可能参与了 U4-U6 间解旋.解旋是剪接的需要. U1 与 5' 剪接位点, U2 与分枝点以及 U4, U6 间的碱基互补配对在剪接反应发生时都需要打开,这些局部的双螺旋可能是 ATP 酶/拓扑异构酶的作用区域. PRP28 有拓扑异构酶活性,它与 PRP24 相作用,而 PRP24 有稳定 U4-U6 间结合的作用.由于 PRP28 的解旋及其它因子的共同作用,在剪接体组装的最后一个阶段,第一步剪接反应发生之前, U4, U6 间结合力已大幅度减弱,于是 U6 激活.这里 U4 可看作反义负向调控,即与 U6 结合使其处于惰性状态,一旦时机成熟, U4, U6 间螺旋解开, U6 的催化活性被激活而发挥作用. U6 的这一催化活性可能与第二步剪接反应有关.

哺乳动物中, 5' 剪接位点的突变常常致使隐蔽的剪接位点被激活.有人认为隐蔽的剪接位点优先出现在与 U1 5' 端有最佳匹配的区域中,而且选择与 U1 5' 端 C8, C9 配对的碱基.比如兔珠蛋白 β 亚基 mRNA 中 5' 剪接位点 G1 突变为 U 或 C, 剪接位点向上游移动一个碱基而成为 GU 或 GC. 如果新的剪接位点内含子一侧为 G,如上例,则剪接反应可以进行彻底而得到产物,否则只能完成第一步剪接反应而造成中间产物积累,因为套索的 G-A 被改变了.

酵母中情形有些不同.将剪接位点处 G1 突变为 A, 剪接位点并没有移动,这时由于套索的 G-A 变为 A-A, 所以出现中间产物积累.可能 U5 与确定 5' 剪接位点有关.它包含一个由 9 个碱基组成的保守的环区.若在上述 G1→A1 突变的基础上,在 U5 的这一环区中同时发生一个碱基替换和一个碱基缺失,则新的剪接位点被激活,重又可形成 G-A 套索,于是反应可以进行彻底^[8]. 内含子第五位 G5 的突变将

导致剪接位点向上游或下游移动若干碱基,新的剪接位点与原剪接位点间无明显同源性,剪接将滞留在第二步反应前. U1 上作相应突变恢复两者间第五位上碱基互补配对,不能够将剪接位点复原到正常位置^[9] (这一点上 U1 不同于 U2, U2 的突变可以不同程度地抑制分枝点的突变).相反,在剪接位点 5' 上游外显子中特定碱基上作突变,这一突变不影响 U1 与前 mRNA 间的结合却能使剪接位点复原,剪接反应即可进行彻底.这说明 U1 结合于 5' 剪接位点仅仅决定剪接发生的区域,不能够决定断开这个区域中具体哪个键,后者可能与 U5 及外显子有关.

5.2 影响第二步剪接反应及剪接体解聚

DEAH 蛋白 PRP16 具有 ATP 水解活性.若将其免疫沉淀,则出现中间产物积累,再加入纯化的 PRP16 和 ATP, 反应又可以继续进行下去^[10]. 它的功能可能与分枝点的确认有关.若它发生突变,则分枝点突变的底物亦可被剪接,即对分枝点精度的要求下降.这一点下面还要谈到.

U4-U6 间有两段碱基配对.酵母 U6 的突变实验表明:茎 I 上有两个区域对于突变很敏感.一个序列为 CAGC, 另一个序列为 ACA-GAG. 在酵母 *S. pombe* 中, CAGC 中插入了一个 mRNA 型内含子,酵母 *R. dactyloides* 中则是 ACAGAG 中插入了一个 mRNA 型内含子(图 3).可能这两段碱基在空间上与 3' 剪接位点相邻,由于剪接中的“事故”而将内含子插进去, U6 上这两个敏感区域的突变将造成中间产物的积累.一种解释认为上述区域为催化部位,另一种解释认为由于突变改变了 RNA 立体构象,以至于起催化作用的蛋白不能够结合上来.

第二步中很关键的一点是对 3' 剪接位点的确认.有实验报道人 U5 snRNP 可能参与了 3' 剪接位点的识别.

剪接完成后,剪接体要解聚,以使剪接因子再次被利用而进行新一轮剪接.剪接因子可能是依次离去的而非“一哄而散”.例如酵母

中 PRP8 先于剪接后的内含子离开剪接体。

解聚也是由若干因子共同作用完成的。DEAD 蛋白 PRP22 参与了剪接体的解聚^[6]。若使 PRP22 热失活, 则成熟的 mRNA 将继续滞留在剪接体上。另外至少还有 PRP26, PRP27 也参与了剪接体的解聚。剪接过程示意在图 4 中。

5.3 对校正机制的影响

剪接与翻译之间有许多相似之处: RNP 有序地结合到 RNA 底物上去; ATP 或 GTP 水解促使蛋白因子的结合或释放, 导致核糖体或剪接体构象的改变, 使得反应逐步进行; 两个过程都需要拓扑异构酶, 以及具有锌指等特征结构的蛋白。更为重要的是两者在校正机制上的相似之处。

肽链延伸时, 首先是肽链延伸因子 EF-TU, GTP, 氨酰基-tRNA 三元复合物进入 A 位, 然后 GTP 水解, 同时核糖体对氨酰基-tRNA 正确与否作出判断。如果正确, 则留下氨酰基-tRNA 而释放 EF-TU, GDP; 否则将三者一同剔出 A 位。我们知道识别准确率与 NTP 水解速度成反比。如果 EF-TU 突变而改变了它和 GTP 间亲和力或 GTP 水解速度, 则作出判断所要求的时间将被缩短, 而识别准确率下降, 出现错误的氨基酸加入到肽链上的情形。

若分枝点发生 A→T 突变, 剪接将被阻止。PRP16 上 ATP 结合位点的突变可抑制该突变^[11]。这一方面说明了分枝点在剪接中的功能与 ATP 的结合或水解有关; 另一方面说明 PRP16 突变体降低了剪接体对分枝点的识别, 使一个突变的底物被剪接。在此 PRP16 的作用机制尚不清楚: 是类似于 EF-TU, 由于 NTP 的水解而发生构象的改变, 或是类似于 eIF4A, 具有 RNA 拓扑异构酶活性, 能利用 ATP 水解能量来打开 mRNA 二级结构。另外, U2 的突变也能部分地抑制分枝点的突变。

非 G-A 的套索将导致中间产物的积累。若 U5 保守环区发生多碱基的突变, 则可降低第二步反应对底物的要求而将反应进行下去。

真核生物核基因不采用第一类、第二类内含子的自我剪接方式, 而是将剪接的信息从内含子本身转移到外部因子, 这就可以给内含子提供以较大的自由度, 使其可以在长度与序列上发生变异, 同时又保证了剪接的精度。对于这一高度精确的行为的分子调控机制的了解, 将大大丰富我们对高等真核生物的认识。

作者感谢中国科学院遗传研究所童克忠先生的热忱帮助。

参 考 文 献

- 1 Shuster E O, Guthrie C. Human U2 snRNA can function in pre-mRNA splicing in yeast. *Nature*, 1990; **345**: 270
- 2 Legrain P, Chouluka A. The molecular characterization of PRP6 and PRP9 yeast genes reveals a new cysteine/histidine motif common to several splicing factors. *EMBO J*, 1990; **9**: 2775
- 3 Jacquier A. Self-splicing group I and nuclear pre-mRNA introns: How similar are they? *Trends Biochem Sci*, 1990; **15**: 351
- 4 Shannon K W, Guthrie G. Suppressors of a U4 snRNA mutation define a novel U6 snRNP protein with RNA-binding motif. *Genes Dev*, 1991; **5**: 773
- 5 Banroques J, Abelson J. PRP4, a protein of the yeast U4/U6 small nuclear ribonucleoprotein particle. *Mol Cell Biol*, 1989; **9**: 3710
- 6 Company M, Arenas J, Abelson J. Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes. *Nature*, 1991; **349**: 487
- 7 Strauss E, Guthrie C. A cold-sensitive mRNA splicing mutant is a member of the RNA helicase gene family. *Genes Dev*, 1991; **5**: 629
- 8 Newman A, Norman C. Mutations in yeast U5 snRNA Alter the specificity of 5' Splice-site cleavage. *Cell*, 1991; **65**: 115
- 9 Seraphin B, Rosbash M. Exon mutations uncouple 5' splice site selection from U1 snRNA pairing. *Cell*, 1990; **63**: 619
- 10 Schwer B, Guthrie C. PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome. *Nature*, 1991; **349**: 494
- 11 Burgess S, Couto J R, Guthrie C. A putative ATP binding protein influences the fidelity of branchpoint recognition in yeast splicing. *Cell*, 1990; **60**: 705