

## mRNA 加工 I: 剪接 (mRNA Processing I: Splicing)

本文介绍 mRNA 加工过程中的剪接机制，包括断裂基因的证据、RNA 剪接的过程、信号、机制以及调控，并简要提及了 RNA 自剪接现象。

### 14.1 断裂基因 (Genes in Pieces)

大多数真核基因被非编码区打断。RNA 聚合酶在转录时会转录整个序列，产生不均一核 RNA (hnRNA)。细胞必须通过**剪接 (splicing)** 过程去除这些非编码 RNA 序列。基因中的间隔序列 (intervening sequences, IVS) 或内含子 (introns) 是非编码区，而编码区则称为外显子 (exons)。有些基因，尤其在低等真核生物中，没有内含子。内含子存在于真核生物的 mRNA、tRNA 和 rRNA 基因中，以及叶绿体和线粒体基因中。内含子的数量和大小差异很大。例如，哺乳动物的二氢叶酸还原酶 (DHFR) 基因有 6 个外显子，总长约 2000 个碱基，但整个基因长达 31,000 个碱基。 $\alpha$ -胶原基因有 50 个外显子，大小从 45-249 个碱基不等，基因总长约 40,000 个碱基。同样大小的基因可能含有不同数量和大小的内含子。有些物种的特定基因含有内含子，而其他物种的同一基因可能没有。例如，植物线粒体细胞色素氧化酶 (cytochrome oxidase) 亚基 II 基因在不同植物物种中内含子的存在情况不同。

#### 14.1.1 断裂基因的证据：R-loop 技术 (Evidence for Split Genes, R-loop)

R-loop 技术是证明断裂基因存在的证据之一。其原理是 mRNA 与其 DNA 模板杂交形成双链杂合体，通过电子显微镜 (EM) 进行观察。

#### 14.1.2 RNA 剪接过程 (RNA Splicing Process)

RNA 剪接是将不成熟 RNA 中的内含子切除并将外显子连接起来形成最终产物的过程。原始转录物 (primary transcript) 或 pre-mRNA 经过剪接后成为成熟转录物 (mature transcript)。

#### 14.1.3 剪接信号 (Splicing Signals)

**剪接信号**是指导剪接发生的特异性序列。典型的剪接信号模式为：外显子/**GU**-内含子-**AG**/外显子。在哺乳动物中，5' 剪接位点的共有序列为 AG/GUAAGU，3' 剪接位点的共有序列为 YNCUR**AC**-**Yn**NAG/G。其中，/ 表示外显子-内含子边界，Y 为 U/C，Yn 为约 9 个嘧啶组成的序列，R 为 A/G，A 是参与分支的特殊腺苷酸，N 是任意碱基。3' 剪接位点由长度不等的多聚嘧啶区 (polypyrimidine tract, PPT) 界定，PPT 区域招募因子结合到 3' 剪接位点和分支点序列 (branch point sequence, BPS)。通过突变共有序列或研究有剪接问题的患者基因，证实了这些剪接信号对于正常剪接至关重要。

#### 14.1.4 剪接对基因表达的影响 (Effect of Splicing on Gene Expression)

剪接的存在实际上促进了基因表达。**内含子可以提高基因表达效率**，例如  $\beta$ -珠蛋白基因的高效表达依赖于内含子。内含子刺激高效的 mRNA 3'-末端形成，并提高翻译效率。mRNA 在细胞核中与多种蛋白质复合形成信使核糖核蛋白 (mRNP)，这些蛋白质随 mRNA 转运到细胞质。剪接过程中，一些蛋白质被添加到 mRNP 的外显子连接处，形成外显子接点复合物 (exon junction complex, EJC)。EJC 是内含子刺激基因表达所必需且充分的，可能通过促进 mRNA 与核糖体的结合来实现。因此，是剪接过程中添加的蛋白质而不是剪接本身引起了刺激作用。EJC 还能促使含有提前终止密码子的错误 mRNA 降解，从而通过去除无生产力地占据核糖体的受损 mRNA 来提高效率。

### 14.2 细胞核 mRNA 前体的剪接机制 (The Mechanism of Splicing of Nuclear mRNA Precursors)

#### 14.2.1 分支中间体 (A Branched Intermediate)

细胞核 mRNA 前体的剪接通过**套索机制**进行，中间体呈分支状，类似套索 (lariat)。剪接过程分两步：(1) 套索状中间体的形成；(2) 剪接过程完成。在第一步中，内含子中分支点的 A 的 2'-OH 攻击第一个外显子与内含子连接处的磷酸二酯键，产生游离外

显子 1 和套索状内含子-外显子 2 中间体。在第二步中，外显子 1 的游离 3'-OH 攻击内含子与外显子 2 的磷酸二酯键，连接两个外显子，并释放套索状内含子。

#### 14.2.2 分支点信号 (A Signal at the Branch)

分支点存在一个共有序列。在酵母中，分支点序列几乎不变，为 UACUAAC。在高等真核生物中，共有序列变化较大。在这两种情况下，分支的核苷酸都是序列中最后一个 A。酵母的分支点序列还能指导剪接机器选择下游合适的 AG 作为 3' 剪接位点。

#### 14.2.3 剪接体 (Spliceosomes)

细胞核 mRNA 前体的剪接发生在剪接体 (spliceosome) 上。剪接体由五种小核 RNA 蛋白 (small nuclear RNA proteins, **snRNPs**) 和多种非 snRNP 相关蛋白因子组成。

snRNPs (发音为 "snurps") 由偶联到蛋白质上的小核 RNA (small nuclear RNAs, snRNAs)

组成。五种 snRNPs 和/或 snRNAs 包括 **U1, U2, U4, U5, 和 U6**。它们具有三个作用：

(1) 识别 5' 剪接位点和分支点；(2) 将这些位点聚集在一起；(3) 催化 (或协助催化)

RNA 的切割。RNA-RNA、RNA-蛋白质和蛋白质-蛋白质相互作用在剪接过程中都非常重要。

snRNAs 和蛋白质剪接因子都识别剪接信号。U1 snRNP 首先识别 5'-剪接位点，然后被 U6 snRNP 取代。U2 snRNP 识别分支点，U2AF 蛋白 (U2-associated factor) 识别 3'-剪接位点。U5 snRNP 结合到 5'-和 3' 剪接位点。

U1 snRNA 具有与 5' 和 3' 剪接位点共有序列互补的区域，可能通过碱基配对将它们聚集在一起。U6 snRNA 与 5'-剪接位点碱基配对。U6 在剪接体中与内含子 5'-末端非常接近。酵母 U6 中不变的 ACA (nt 47-49) 与内含子的 UGU (nt 4-6) 碱基配对。U6 在剪接的初始步骤前后都与剪接底物结合。存在 U2-U6 复合物，**U2-U6 碱基配对有助于形成剪接体的活性位点**。U6 snRNP 通过 U6 snRNA 与内含子 5'-末端碱基配对结合。U6 与剪接底物的结合对于剪接过程至关重要。U6 在剪接过程中也与 U2 结合。

U2 snRNA 与分支点的保守序列碱基配对。这种碱基配对对于剪接至关重要。U2 也与 U6 形成重要的碱基对，形成**螺旋 I** (helix I)，帮助这些 snRNPs 定位进行剪接。U2 的 5'-末端与 U6 的 3'-末端相互作用形成螺旋 II (helix II)，这对于哺乳动物细胞的剪接很重要，但在酵母细胞中不是必需的。

U5 snRNP 与两个外显子结合，将它们定位以便进行第二步剪接。U5 snRNP 结合到第一个外显子的 3'-末端和第二个外显子的 5'-末端，以定位两个外显子进行剪接。

U4 snRNP 与 U6 结合，但在剪接中没有直接作用。剪接开始后，U4 从 U6 解离。U4 的作用是结合并隔离 U6，直到 U6 参与剪接。与 U4 碱基配对形成**茎 I** 的一些 U6 碱基也参与与 U2 的碱基配对。去除 U4 允许 U6 与 U2 碱基配对并帮助形成活性剪接体。U6 snRNA 参与催化。剪接的两个步骤都是转酯反应 (trans-esterification reaction)，其中一个磷酸二酯键断裂并形成另一个磷酸二酯键。

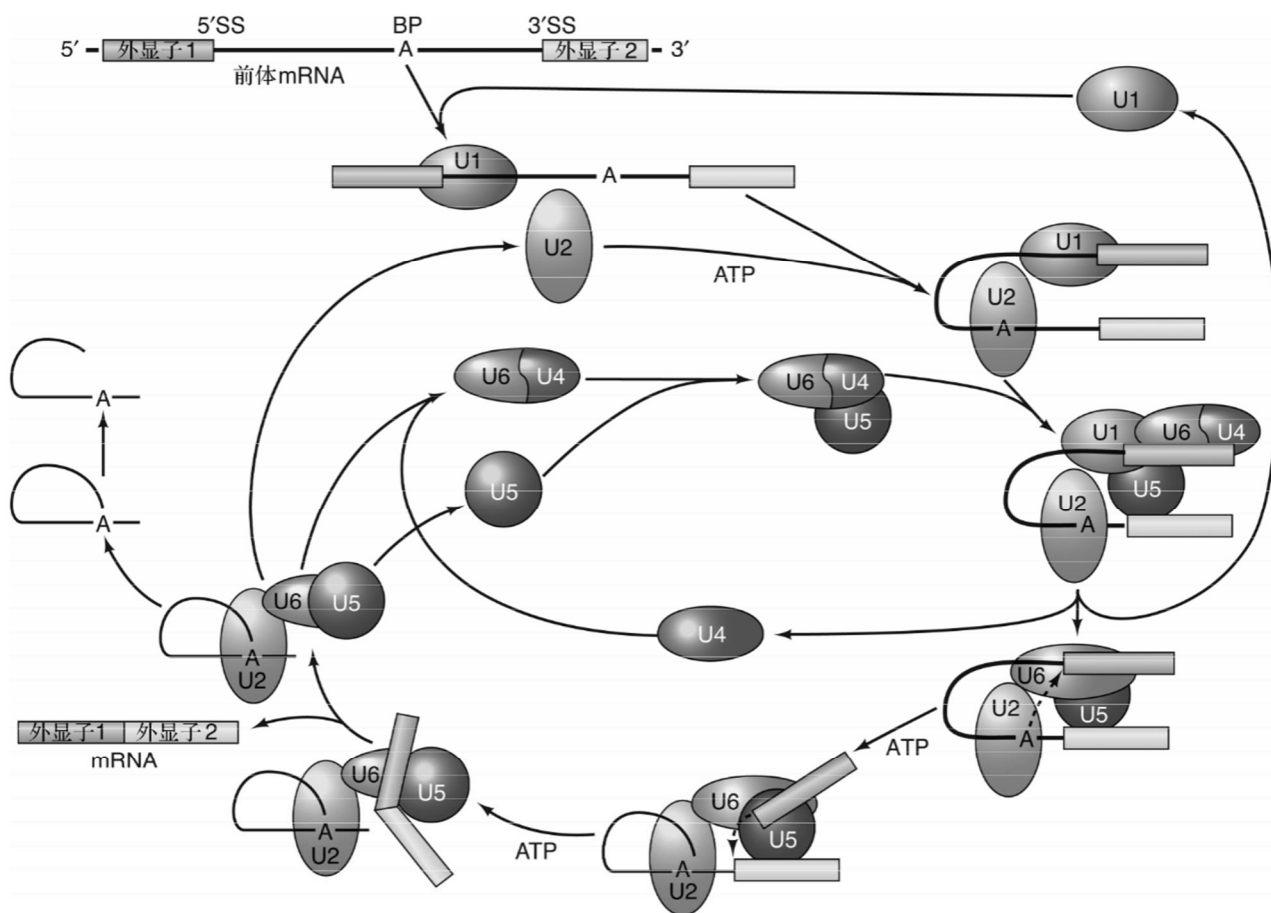


图 14.25 剪接体循环。循环过程中的各个事件已在正文中阐述。(Source: Adapted from Sharp, P. A. Split genes and RNA splicing, *Cell* 77: 811, 1994.)

#### 14.2.4 剪接体的组装及功能 (Spliceosome Assembly and Function)



**所有 snRNPs 都含有相同的一组七个 Sm 蛋白**。这些蛋白是系统性自身免疫疾病患者体内抗体的常见靶点。Sm 蛋白结合到 snRNAs 上的共同 Sm 位点。每个 snRNP 仍然有自己特异性的蛋白质组。Sm 蛋白形成一个中间有孔的**甜甜圈**状结构。

在后生动物 (metazoans) 中，存在一种罕见的变异内含子类型。它们的 5'-剪接位点和分支点序列与典型的不同。一组丰度较低的 snRNPs，包括 U11, U12, U4atac, 和 U6atac，与 U5 一起构成低丰度剪接体 (minor spliceosome)，用于剪接罕见的 U12-型 pre-mRNA 内含子。这些构成 U12 剪接体的 snRNPs 位于细胞质中，而主要剪接体位于细胞核中。后生动物细胞含有低丰度 snRNAs：U11 对应 U1；U12 对应 U2；U4atac 对应 U4；U6atac 对应 U6。

剪接体的组装是逐步进行的。剪接体的组装、功能和拆卸被称为**剪接体循环** (spliceosome cycle)。U1 是第一个结合到剪接前体的 snRNP。

酵母剪接体循环包括以下步骤：(1) 定向复合体 (commitment complex, CC) 形成：剪接底物 + U1 + 其他因子。CC 在 5'-位点对内含子剪接进行定向。(2) U2 在 ATP 帮助下加入，形成 A 复合体。(3) U4-U6 和 U5 加入，形成 **B1** 复合体。(4) U4 解离：U6 取代 U1；U1 和 U4 退出；**U6 与 U2 碱基配对**。形成活化的剪接体 (**B2 复合体**)。(5) 在 ATP 帮助下，在 **C1** 复合体中进行**第一步剪接**。(6) 在第二次 ATP 帮助下，在 **C2** 复合体中进行**第二步剪接**。(7) 剪接后的内含子结合到 I 复合体，然后从 snRNPs 解离，去分支并降解。snRNPs 被回收用于另一个剪接复合体。

#### 14.2.5 定位、剪接位点选择、可变剪接 (Commitment, Splice site selection, and Alternative Splicing)

**snRNPs 需要额外的剪接因子帮助结合外显子-内含子边界**。Slu7 识别正确的 AG。

U2AF 和 SR 蛋白参与特定位点剪接的定位。U2AF (U2-associated factor) 和 SR 蛋白 (含有富含丝氨酸/S 和精氨酸/R 的结构域)。一些剪接因子桥接内含子和外显子，从而定义这些 RNA 元件用于剪接。其他因子可以改变剪接位点的选择。

### 3' 剪接位点选择 (3'-Splice Site Selection)

3' 剪接位点 AG 通常位于分支点下游 18-40 nt 之间。靠近分支点的 AG 通常被跳过。

剪接因子 Slu7 对于选择正确的 AG 是必需的。**没有 Slu7，正确的 AG 不会被使用。**

U2AF (U2AF35 + U2AF65) 参与 3' 剪接位点识别。U2AF65 结合到多聚嘧啶区，U2AF35 结合到 3' 剪接位点的 AG。实验证明 Slu7 对于在 3' 剪接位点选择正确的 AG 是必需的。

### 定位 (Commitment)

剪接因子 SC35 是一种 RNA 结合蛋白 (RNA-binding protein)，属于 **SR 蛋白家族**。

SC35 导致定向复合体的形成。SC35 的定向活性是特异性的，不源于一般的 RNA 结合能力。SR 蛋白是剪接激活因子，包含两个结构域：(1) RNA 识别模体 (RNA-recognition motif, RRM)，用于 RNA 结合；(2) RS 结构域，介导剪接机器内部蛋白质之间的相互作用以促进剪接。

### 桥接蛋白与定位 (Bridging Proteins and Commitment) **内含子界定**

酵母中不存在 SR 蛋白，但酵母和哺乳动物的 CC 具有许多共同特征。酵母 Mud2p 不仅与内含子 5'-末端的 U1 snRNA 相互作用，还与分支点桥接蛋白 (branchpoint bridging protein, **BBP**) 和靠近内含子 5'-末端的 **yPrp40p** 相互作用。Prp40p 是 mRNA 前体加工蛋白，是 U1 snRNP 的组分之一。BBP 在内含子 5'-和 3'-末端之间形成桥梁。在酵母 CC 中，BBP 结合到内含子 5'-末端的 U1 snRNP 蛋白，并结合到内含子 3'-末端附近的 **Mud2p**。BBP 也结合到内含子 3'-末端附近的 RNA。因此，它桥接内含子，可能在剪接前定义内含子中发挥作用。哺乳动物的 BBP 对应物 SF1 (mBBP) 可能在哺乳动物 CC 中发挥相同的桥接功能。

### RNA 聚合酶 II CTD 的作用 (Role of the RNA Polymerase II CTD)

剪接、加帽和多聚腺苷酸化由 RNA 聚合酶 II 的 C-末端结构域 (CTD of Rpb1) 协调。

在外显子界定 (exon definition) 中，剪接因子识别外显子的末端并剪切内含子。**所有**

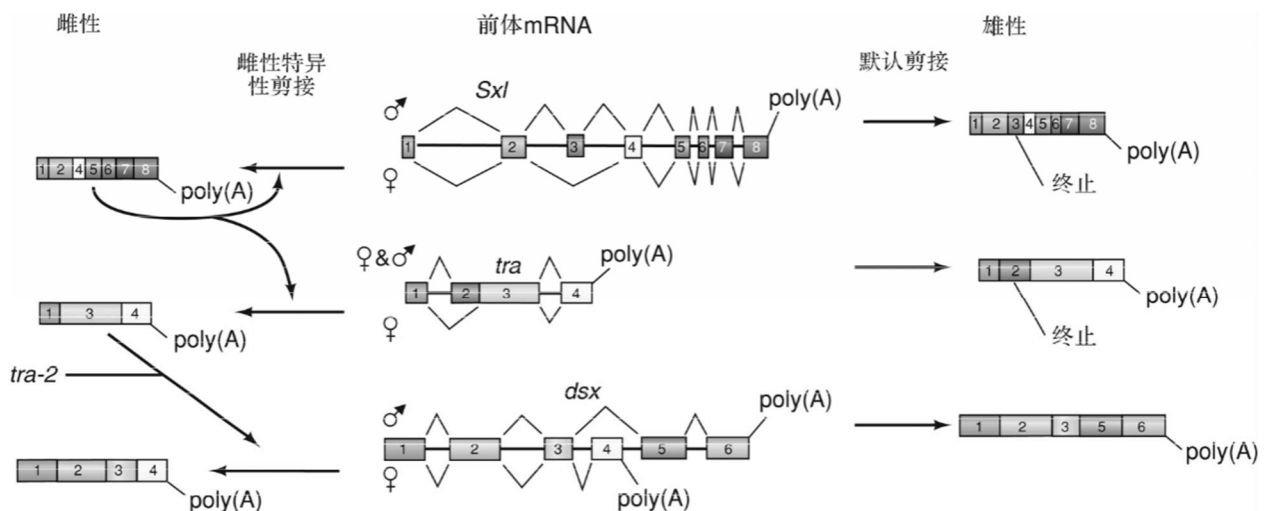
**外显子必须完整才能确定其为外显子**。在内含子界定 (intron definition) 中，识别的是内含子的末端。CTD 刺激使用外显子界定的底物剪接，但不刺激使用内含子界定的底物。高等真核生物主要以外显子界定为主，酵母以内含子界定为主。高等真核生物的外显子通常较小 (<300nt)，酵母的内含子通常较小，而外显子大小似乎没有限制。CTD 与剪接因子结合，将剪接因子组装到外显子的两端，促进剪接反应。实验表明 CTD 能够刺激剪接反应，并帮助剪接底物募集到活性剪接体。CTD-GST (CTD 融合蛋白) 不能刺激含有不完整外显子的剪接底物。CTD 与 snRNPs 和 SR 蛋白结合。RNA Pol II 转录出第一个外显子，CTD 介导剪接因子组装到 pre-mRNA 外显子的两端，从而界定该外显子。CTD 以同样方式介导第二个外显子，并拉近两个外显子，为剪接做准备。随着 RNAP 继续转录，两个外显子被剪接在一起。

### **选择性剪接/可变剪接 (Alternative splicing)**

选择性剪接是从一种剪接模式切换到另一种。大约 50% 的 pre-mRNA 可以以不止一种方式剪接，产生两种或多种可变 mRNA，从而产生不同的蛋白质。其意义在于：(1) 决定基因蛋白质产物的性质；(2) 控制基因表达。果蝇 DSCAM 基因可以以 38,000 种可变方式剪接。

果蝇的性别决定系统是选择性剪接的例子。Sex lethal (Sxl)、transformer (tra) 和 doublesex (dsx) 三个基因 pre-mRNA 的选择性剪接方式在雄性和雌性中不同。在雌性中，Sxl 转录物的雌性特异性剪接产生活性产物，该产物进一步增强 Sxl 转录物的雌性特异性剪接，同时也引起 tra 转录物的雌性特异性剪接，产生活性 tra 产物。活性 tra 产物与另一个基因 tra-2 的产物一起，引起 dsx 转录物的雌性特异性剪接，其产物使雄性特异性基因失活，导致雌性个体发育。相反，在雄性中，Sxl 转录物的雄性特异性剪接产生无活性产物，因为外显子 3 含有终止密码子。因此，tra 转录物也进行默认 (雄性特异性) 剪接，产生无活性产物。没有 tra 产物，细胞按照默认的雄性特异性剪接模式剪接 dsx 转录物，其产物使雌性特异性基因失活，导致雄性个体发育。

选择性剪接受到 RNA 剪接因子的控制。Sxl 和 Tra 的产物能够决定 tra 和 dsx 转录物的剪接位点。Sxl 和 Tra 可能是剪接因子，引起雌性特异性剪接模式的定位。Tra 蛋白和 Tra-2 蛋白通过结合在 dsx pre-mRNA 雌性特异性 3'-剪接位点下游的调控区而发挥作用。Tra 和 Tra-2 是 dsx pre-mRNA 进行雌性特异性剪接所必需的。Tra 和 Tra-2 单独不足以引起定位，外加任何一种有活性的 SR 蛋白就足以确保 dsx pre-mRNA 发生雌性特异性剪接。



**图 14.36 果蝇性别决定中选择性剪接的级联反应。** *Sxl*、*tra* 和 *dsx* 基因前体 mRNA 分子的结构在雌雄果蝇个体中是一致的，见图中部所示。雄性特异剪接模式和雌性特异剪接模式分别在结构图的上、下方标出。*Sxl* 前体 mRNA 的雌性特异性剪接涉及 1、2、4~8 等 7 个外显子的拼接，而 *Sxl* 前体 mRNA 的雄性特异性剪接涉及全部 8 个外显子的拼接，其中外显子 3 含有终止密码子，这意味着该转录物的雄性特异性剪接产生短的无活性的蛋白质。同样，*tra* 前体 mRNA 的雌性特异性剪接涉及 1、3、4 外显子的拼接，产生有活性的蛋白质，而 *tra* 前体 mRNA 的雄性特异性剪接涉及全部 4 个外显子的拼接，其中外显子 2 含有终止密码子，产生无活性蛋白质产物。最左侧的长箭头指示基因的蛋白质产物对剪接反应具有正控效应，即雌性的 Sxl 蛋白引发 *Sxl* 前体 mRNA 和 *tra* 前体 mRNA 的雌性特异性剪接反应。雌性 *tra* 基因产物与 tra-2 蛋白协同作用，引发 *dsx* 基因转录物的雌性特异性剪接反应。（Source: Adapted from Baker, B. S. Sex in flies: The spice of life. *Nature* 340: 523, 1989.）

补充知识：mRNA 的可变剪接极大地增加了蛋白质组的多样性和复杂性，使基因能够承担复杂功能并实现精细调节。59% 的人类基因具有可变剪接，每个基因平均有 2.6 种剪接型。多数可变剪接基因表达于神经系统和免疫系统。

#### 14.2.6 剪接的调控 (Control of Splicing)

选择性剪接模式的调控方式包括：(1) 转录物从不同的启动子处开始；(2) 有些外显子可被忽略；(3) 5'-剪接位点的选择；(4) 3'-剪接位点的选择；(5) 内含子保留 (retained intron)；(6) 多聚腺苷酸化引起 pre-mRNA 切割，使下游外显子缺失。细胞可精细调



控选择性剪接，对一段序列在不同情况下以外显子或内含子区别对待。决定对剪接信号特异性识别的因素包括：(1) 剪接因子刺激在特定剪接位点的定位；(2) 外显子含有促进剪接的**外显子剪接增强子** (exonic splicing enhancer, ESE) 和抑制剪接的**外显子剪接沉默子** (exonic splicing silencer, ESS)；(3) 内含子也有内含子剪接增强子和内含子剪接沉默子。这些序列与细胞类型特异性或发育阶段特异性蛋白质因子结合，或响应外界刺激，从而促进或抑制邻近位点的剪接。SR 蛋白结合 ESE，hnRNP (heterogeneous nuclear ribonucleoprotein) 蛋白结合 ESS。hnRNP 蛋白结合 hnRNA，是剪接阻遏物 (splicing repressors)。它们缺乏 RS 结构域，不能招募剪接机器，但能阻止其结合的特定位点的使用。具有 ESS 活性的 hnRNP 通常是 hnRNPA1。hnRNPA1 蛋白至少有 3 种作用机制抑制外显子剪接。

### 14.3 RNA 的自剪接 (Self-Splicing RNAs)

RNA 的自剪接是分子生物学的重要发现之一。首先在四膜虫 (Tetrahymena) 26S rRNA 基因中发现。自剪接内含子分为 Group I 内含子和 Group II 内含子。