

RNA 剪接因子结构与功能研究进展*

吴 瑛^{1 2)} 赫荣乔¹⁾**

(¹⁾中国科学院生物物理研究所, 视觉信息加工重点实验室, 北京 100101 ;

²⁾Washington University, School of Medicine, St. Louis, MO 63110, USA)

摘要 RNA 剪接是一个多步骤、形成多种中间状态复合物的复杂过程. 尽管在已经发现的一百多种 pre-mRNA 剪接相关因子中, 仅研究了约 8% 相关蛋白质的空间结构, 已充分显示对剪接相关因子三维晶体结构以及溶液结构的测定与研究, 在理解 RNA 剪接的复杂机理以及生物学特性中具有不可替代的重要意义.

关键词 RNA 剪接, snRNA 结合蛋白, hnRNA 结合蛋白, 剪接体, 晶体结构, 溶液结构, RNA 结合结构域

学科分类号 Q75

RNA 剪接 (RNA splicing) 是 tRNA、rRNA, 特别是 mRNA 加工与成熟的重要生物学过程, 也是蛋白质分子多样性产生的关键机制之一. RNA 剪接需要多种因子参与, 包括杂性核 RNA (heterogeneous nuclear RNA, hnRNA) 结合蛋白 hnRNP (heterogeneous nuclear ribonucleoprotein)、小型核 RNA (small nuclear RNA, snRNA) 结合蛋白 snRNP (small nuclear ribonucleoprotein) 等. 在真核细胞内, RNA 原初转录物的分子很大, 通过剪接产生成熟的 mRNA 分子. hnRNP 与 hnRNA 结合形成核酸蛋白质复合物 hnRNPP (hnRNP particles) 可穿梭于细胞核与胞质之间, 具有转运和剪接 RNA 的作用. 另一方面, snRNP 与细胞核内 snRNA (分子质量为 100~200 nt 的小 RNA) 紧密结合, 而构成核内小核酸蛋白质复合物 snRNPP (snRNP particles), 参与 RNA 剪接、多聚腺苷化以及转录拷贝 3' 端的成熟.

最近, Reed 等通过蛋白质组学方法证明, 与 pre-mRNA 剪接相关的蛋白质因子大约有 145 种, 它们参与 RNA 剪接过程中的不同环节^[1]. 托普霉素亲和层析^[2] (tobramycin affinity-selection) 以及麦芽糖亲和层析^[3] (maltose-binding affinity) 等方法能识别和分离 70 种以上的相关剪接蛋白质. 由于 RNA 剪接机制愈来愈受到国内外同行的关注, 在最近几年中, 部分 pre-mRNA 剪接因子, 如某些重要 snRNP、hnRNP 等的晶体及溶液结构获得了一些研究成果, 为阐释 RNA 剪接的复杂机制提供了至关重要的信息.

1 hnRNP

hnRNP-A1 (又名解链蛋白 1, unwinding

protein 1, A1) 是真核细胞核 hnRNPP 复合物中含量十分丰富的分子. A1 能与 pre-mRNA 结合, 形成 hnRNPP 复合物, 并通过对几种富含 Ser/Arg 剪接因子的拮抗, 以球状 (global) 调节因子的方式参与 RNA 交替剪接 (alternative splicing), 表现为选择性地跨过某些外显子, 实现 5' 端交替剪接. A1 蛋白不断地穿梭在细胞核和胞质之间, 能将 polyA⁺ mRNA 从细胞核运输到胞质, 该过程的调节与其 C 端富含 Gly 结构域的结构相关^[4]. 了解 hnRNP A1 与 RNA 相互作用中空间结构方面的信息, 将有助于揭示 RNA 穿过细胞核膜以及剪接过程中的分子细节.

RNA 结合结构域 (RNA binding domain, RBD; 又称 RNA 识别基序 RNA-recognition motifs, RRM) 是蛋白质与 RNA 相互作用的常见结构域之一^[5]. 在一定条件下, 单个 RBD 就能特异识别 RNA 并进行高亲和的相互作用, 但在某些情况下, 需要多个 RBD 之间的协同才能实现蛋白质与 RNA 的结合. 人类 hnRNP A1 的 N 端具有两个 RNA 结合域, 即 RBD1 和 RBD2. 晶体结构 (分辨率为 0.19 nm) 研究表明^[2], hnRNP A1 含有两个在空间上相对独立的 BRD, 彼此由一条 linker 连接. 在未结合 RNA 时, 连接肽在空间上具有较大的柔性, 这有利于 RBD 与 RNA 直接相互作用. 两个 RBD 在空间位置彼此靠近, 呈反平

* 国家自然科学基金委纳米生物学重大计划重点项目 (90206041) 和中国科学院创新方向性项目 (KSCX2-SW214).

** 通讯联系人.

Tel: 010-64889876, E-mail: herq@sun5.ibp.ac.cn

收稿日期: 2003-02-10, 接受日期: 2003-03-24

行排列, 构成 RNA 结合平台 (binding platform), 并且依靠两对 Arg-Asp 离子键来实现稳定. 反平行 RNA 结合平台的存在, 使得 RNA 能够结合并集聚于该平台, 形成浓缩状态, 从而有利于 RNA 分子的剪接和运输.

尽管 A1 具有两个 RBD (与其类似物 U1A 的 RBD 同源), 分别包含了 4 条反平行 β 折叠和两段 α 螺旋. 但是, 这两个 RBD 之间在结构上具有一定的差别, 与 RBD2 相比较, RBD1 的 N 端起始处存在一段 3_{10} - α 螺旋, 这可能是与 RNA 直接结合的结构基础. Krainer 和 Steitz 的研究工作很好地展示了两个 RBD 在空间上的折叠和相互靠近的完美空间几何特性, 及其与 RNA 结合时的相互协同机制^[4,5].

1995 年, Ishikawa 等分离纯化了一种能与 UUAG 片段特异结合的 hnRNP, 命名为 hnRNP D₀ (简称 D₀)^[6]. 该蛋白质分子的中部具有两个 RBD, 且在其 C 端区域包括一个 RGG-box, 富含 Gly 与 Arg 残基, 因而被称作 2 × RBD-Gly 结构. 在 hnRNP 家族中, 如 A1、A2/B1 和 D₀, 共同具有能与 RNA 特异结合的 2 × RBD-Gly 结构. RBD 则被认为是 RNA 剪接因子所共同拥有的特征结构. 因此, 在 RBD 蛋白家族中, 这些蛋白质被归类为 2 × RBD-Gly 亚家族. 在 D₀ 分子中, RBD 的 N 端区域有 19 个氨基酸残基插入序列, 调节 RBD 对 RNA 底物的选择性结合, 类似的插入序列在 hnRNP A2/B1 蛋白质中也能观察到.

尽管 D₀ 含有两个 RBD (靠近 N 端为 RBD1 ; C 端为 RBD2), 但其中任意一个都能与 UUAG 片段特异性结合^[7], 同时还能选择性地与人类端粒重复序列单链 d (TTAGGG)_n 特异性结合. RBD 的 N 端区域由两段 α 螺旋和 4 条反平行 β 折叠重叠形成, 这是 RNP-type-RBD 的共同结构特征. 与 RBD2 不同的是, RBD1 的两条 α 螺旋具有 N-帽状盒 (capping boxes), 第二条 β 折叠存在一个 β 凸起 (β -bulge), 一条附加的反平行 β 折叠与一个 β -转角的类似结构形成一个环区 (loop). 该环区的空间结构由氢键参与稳定. 从 RBD 整体上看, 与 RNA 相互作用的必需氨基酸残基在结合区的中央及边缘都有分布, 中央区的氨基酸残基与 RNA 之间为非特异性相互作用, 而边缘区的残基则与特异性识别相关.

2001 年, Katahira 等^[8]采用 NMR 技术比较了 D₀ 中两个 RBD 在溶液中的结构, 获得更为详细的

结果. RBD2 折叠成为 RNP-type RBD 的特征结构, 即 α 螺旋和 β 折叠共同参与的紧密折叠结构, 一条反平行 β 折叠与两段 α 螺旋面对面叠在一起. 除了与 RNP-type RBD 有相同的 4 条反平行 β 折叠外, 还在其上游发现了一条额外折叠, 被称为 $\beta 4 (-)$. 为此, RBD2 就含有 5 条 β 折叠, 这与 RBD1 有较大的不同. 化学位移微扰实验显示, 与 RNA 相互作用的必需氨基酸位于 RBD2 的 β 折叠边沿, 即位于 $\beta 4 (-)$ 与 4 个 β 折叠起始端之间 ($\beta 4-\beta$). 而 RBD1 的必需基团却位于对侧的 α 螺旋上. 进一步实验表明, RBD2 的 RNA 结合位点能以相同的方式与 DNA 相互作用. 当未结合 RNA 时, RBD2 的 $\beta 4-\beta$ 区域表现为相对缓慢 (毫秒至微秒级) 的构象交换 (conformational exchange). RBD 与 RNA 复合物的形成可以淬灭该构象交换. RBD2 在空间上的柔性, 可能使 D₀ 在与 RNA 识别过程中产生不同的构象, 从而发挥诱导契合的作用.

2 snRNP

与 pre-mRNA 剪接相关的 snRNPs, 如 U1、U2、U4/U6 及 U5, 通常含有 7 种 Sm 蛋白 (B/B'、D1、D2、D3、E、F 及 G). 这些蛋白质可以通过 Sm 结合位点在 4 种 snRNAs 周围聚集, 并且在其分子内部共享一段特征序列 (基序, motif), 分别存在于由一段非保守 linker 连接的两个片段 (Sm1 和 Sm2) 中. Sm-B'/B (分子质量为 27/26 ku) 二聚体可以被自身免疫性疾病系统红斑狼疮 (systemic lupus erythematosus, SLE) 的特异性抗体 anti-Sm 所识别, 并且呈现强阳性反应, 这是由于该二聚体的 N 端享有与 SLE 特异性抗原相同的一段多肽序列所致^[9]. 从一级结构来看, Sm-B'/B 富含 Pro (20%) 和 Gly (15%) 残基, 特别是分布在 C 端区域的 Pro 残基, 形成重复排列单位 (repetitive unit), 这一特点与其他 snRNP 的一级结构类似. 值得注意的是, 其中一个含 8 个 Pro 的保守序列可能与 RNA 或其他蛋白质的相互作用有关.

迄今为止, 在 snRNP 中三维空间结构研究得较为仔细的是 U1A. Nagai 等研究了 U1A 的 RBD 与 RNA (21 nt) 茎环-2 (stem loop-2) 所形成复合物的晶体结构 (分辨率为 0.192 nm)^[10]. 需要说明的是, RNA 分子容易形成发卡结构, 通常 stem loop-2 为蛋白质提供了结合位点. 剪接蛋白 U1A 的作用位点就分布于 RNA 茎环顶端的 loop^[11],

loop 上的 10 个核苷酸与 U1A 的 β 折叠相互作用形成一个开放结构 (open structure), 其中 heptamer (AUUGCAC) 识别保守性很强的 RNP1 和 RNP2 基序以及延长于 C 端的 RBD. Heptamer 与芳香族氨基酸的侧链基团直接或通过水分子形成氢键而间接相互作用. 这些结构信息揭示了 RBD 识别 RNA 不但具有序列特异性, 而且具有空间立体化学特异性.

Neuhaus 等^[12]通过制备人类 U1A 氨基端片段 U1A117 (含 2 ~ 117 氨基酸残基), 并采用 NMR 技术研究了该片段的溶液中结构. U1A 羧基端区域在空间结构上具有较高的有序性, 这一点较 U1A 与 RNA 茎环-2 形成复合物的晶体结构有较大差别. α 螺旋 (Asp90 ~ Lys98) 与另一条 β 折叠在空间上靠近, 通过螺旋上 Ile93、Ile94、Met97 与折叠层上 Leu44、Phe56、Ile58 残基相互接触, 防止 β 折叠的疏水氨基酸残基暴露而与溶剂分子相作用, 确保蛋白质空间结构的自身稳定. 当 U1A 与 RNA 结合时, C 段 α 螺旋发生空间位移, 并旋转 135° , 使得 Tyr13、Phe56、Gln54 重叠到 RNA 碱基上, 同时也使 Leu44 与 RNA 接触. 在 U1A-RNA 复合物形成过程中, C 段 α 螺旋的空间位移, 是通过氨基酸残基间的疏水作用 (Ile93、Ile94、Ile58、Leu41、Val62、His10) 以及氢键的形成 (Ser91 ~ Thr11) 而得以稳定. α 螺旋在空间位置上的移动, 改变了 U1A 在 Thr89、Asp90、Ser91 残基间肽链的扭矩, 宛如一个盖子遮住了蛋白质上的 RNA 结合位点.

U1A 与 RNA 的相互作用受到特别关注, 该分子机制可能成为研究蛋白质与核酸相互作用, 形成络合物的诱导契合模型. 通过分子动力学方法, Baranger 等^[13]模拟了 U1snRNA 的 stem loop-2 结构以及 U1A 蛋白的结构, 来预测非结合状态 U1A 的溶液结构, 以阐释在与 RNA 结合时蛋白质分子的诱导动力学过程. 根据 U1A 与 stem loop-2 形成复合物的晶体结构^[8]和溶液结构^[10], 他们的分子动力学模拟及计算结果显示, U1A 与 RNA 的相互作用具有明显的诱导契合机制性质.

Nagai 等^[14]研究了 U2B''/U2A' 与 U2 发夹环 IV 形成三重复合物的晶体结构 (分辨率为 0.24 nm). 与其类似物 U1A 不同的是, U2B'' 仅仅在 U2A' 存在的情况下, 才与其同类 (cognate) RNA 相结合. 在 U2A' 富含 Leu 重复序列区域内, 一条平行 β 折叠形成的凹面, 与 U2B' 表面的 RNP 结构域相互作

用. 该结构域在 U2B' 分子中的相对面, 为 RNA 结合结构域. 另外, U2A' 的碱性 C 端区域与 RNA 茎环相互作用. 晶体结构揭示了蛋白质-蛋白质相互作用是怎样影响和调节蛋白质对 RNA 的特异识别, 以及少数关键氨基酸残基在空间上处于什么样的位置, 才使得 U2B'' 和 U1A 具有识别其各自同类 RNA 茎环的能力, 从而形成交替相互作用的机制.

另一个重要的 snRNP 为 U5^[15], 在进化上具有很高的保守性. Reuter 等^[16]从人类 U5 中纯化并研究了一种 snRNP 特异蛋白质, U5-15 ku. X 射线晶体衍射 (分辨率为 0.14 nm) 的研究结果显示, U5-15 ku 具有与人类硫氧还蛋白 (thioredoxin) 相似的折叠结构. 与硫氧还蛋白相比较, U5-15 ku 的另外 37 个氨基酸残基导致该分子空间结构在某种程度上的改变, 从而为其他剪接体蛋白或 RNA 提供结合位点. 在硫氧还蛋白家族中, 蛋白质分子均含有一个标准二硫键 (canonical disulfide bond), 而 U5-15 ku 蛋白却代之以一个分子内二硫键 (intramolecular disulfide bond). 尽管 U5-15 ku 蛋白缺乏二硫键异构酶 (protein disulfide isomerase, PDI) 的活性, 但是酵母细胞直向同源基因 (orthologue gene) 损耗的体内实验表明, pre-mRNA 的剪接过程与该蛋白的功能密切相关.

U5 能与其他剪接因子相互作用形成复合物, 如 Slu7 (RNA 剪接过程中识别 3' 端 AG 位点的因子) 以及 Prp18 (pre-mRNA processing protein 18) 等. 剪接因子 Prp18 为 pre-mRNA 剪接的第二步所必需. Harbor 等通过 X 射线晶体衍射技术研究了 N 端缺失 79 个非保守氨基酸残基的啤酒酵母 Prp18- δ 79 (与天然状态剪接活性相同) 的空间结构^[17]. Prp18 具有一个以 5 段 α 螺旋折叠形成的球状核心, 被称作 X-bouddle. 该分子具有两个由保守氨基酸残基构成的表面, 保守性相对较低的表面与 Slu7 相互作用有关, 而在保守性较高的分子表面有一个氨基酸残基保守性最高的区域 (靠近一条由 19 个非保守氨基酸组成的肽段), 在两个 α 螺旋之间形成一个 loop 结构, 可能与 U5 的结合相关. 这些结果表明, Prp18 在 Slu7 和 U5 之间起到一个桥梁的连接作用. 另外, 在 pre-mRNA 剪接的不同阶段, 剪接因子所形成的复合物具有不同的性质, 如 Moore 等^[18]观察到剪接因子所形成的复合物 C (complex C) 包含了 U2、U5、U6、Prp7、Prp16、Prp17 以及 Prp22 等蛋白质因子. NMR 研究显示, 酵母 Prp40 氨基端的两个 WW 结构域 (WW-

domain) 是其功能上的关键结构^[19].

最近研究表明, U4/U6 中存在一种特殊的组分, Cyclophilin SnuCyp-20, 并与细胞核内 pre-mRNA 剪接相关. SnuCyp-20 与 U4/U6-60 ku、U4/U6-90 ku、人类 Prp4 和 Prp3 (orthologues of *Saccharomyces cerevisiae*) 等因子之间形成稳定的复合物. 通过分子置换法, Ficner 等观察了 SnuCyp-20 晶体结构 (分辨率为 0.20 nm), 其空间结构与人类 cyclophilin A (hCyp-A) 类似^[20]. 对于分子表面结构而言, 尽管两种蛋白质具有较大的差别, 但是 SnuCyp-20 与 hCyp-A 的催化中心在空间结构上几乎完全相同. SnuCyp-20 具有的脯氨酰顺反异构酶 (peptidyl-prolyl-*cis/trans*-isomerase) 活性也说明了这一点. 除了 N 端 7 个延长的氨基酸残基外, 还在 α 螺旋 1 与 β 折叠 3 以及 α 螺旋 2 与 β 折叠 8 形成的两个 loop 中还分别插入了 5 个和 1 个氨基酸残基, 使得这两个 loop 在空间结构上发生变化. 特别是 α 螺旋 1 与 β 折叠 3 组成的 loop 被扩展, 在局部形成了一个较为宽大的以疏水性质为主的裂隙. Ficner 等认为, 被扩展的 loop 可能是与 U4/U6-60 ku 蛋白的相互作用的结构基础.

Nagai 等观察到 7 种 Sm 蛋白 (B/B'、D1、D2、D3、E、F 及 G) 含有一段共同基序, 与 U1、U2、U5 以及 U4/U6 结合, 在 snRNPP 内形成一个球状核心结构域 (globular core domain)^[21]. D3B 和 D1D2 二聚体复合物的蛋白质晶体结构显示, 它

们具有一种共同的折叠结构, 其中包括 N 端一段 α 螺旋和与之相连的 5 条反平行 β 折叠. 二聚体之间的肽链彼此紧密重叠形成核心区域, 两个亚基的相互作用面也包含在该区域中. 蛋白质晶体结构显示, 这 7 种 Sm 蛋白可能形成一个闭环结构 (图 1), 而 snRNA 可能结合在闭环结构中央的正电荷区域^[22]. 根据该模型可以推测, 参与核心结构域组成的每一个 Sm 蛋白只能来自于单拷贝基因. 为了检验该模型正确与否, Nagai 等从化学计量方法着手, 研究了 Sm 蛋白在酵母剪接体的 snRNPP 中所占的比例. 他们构建了 7 株酵母菌, 分别都产生一种标记有钙调蛋白结合肽 (calmodulin-binding peptide, CBP) 的 Sm 蛋白 (CBP-Sm). 每一株菌用标记有 protein-A 的任意一种 Sm 蛋白 (protein-A-Sm) 转染. 当某种 Sm 以 CBP-Sm (D1) 蛋白形式由染色体 DNA 表达时, 可以观察到由质粒表达的 protein-A-Sm (D2) 与 CBP-Sm (D1) 之间 (通过 calmodulin 形成微珠状结构的相互作用信号进行观察) 具有很强的亲和性, 表明不同种类的 Sm 蛋白之间具有明显的相互作用. 但是, 同种 Sm 蛋白之间, 如 D1 与 D1、D2 与 D2 等, 却没有观察到相互作用. 这些研究结果提示, 每一种 Sm 蛋白都是以单拷贝的方式参与形成剪接体 snRNP 核心结构域, 该结构域为七聚体 (heptamer) 的环状复合体结构.

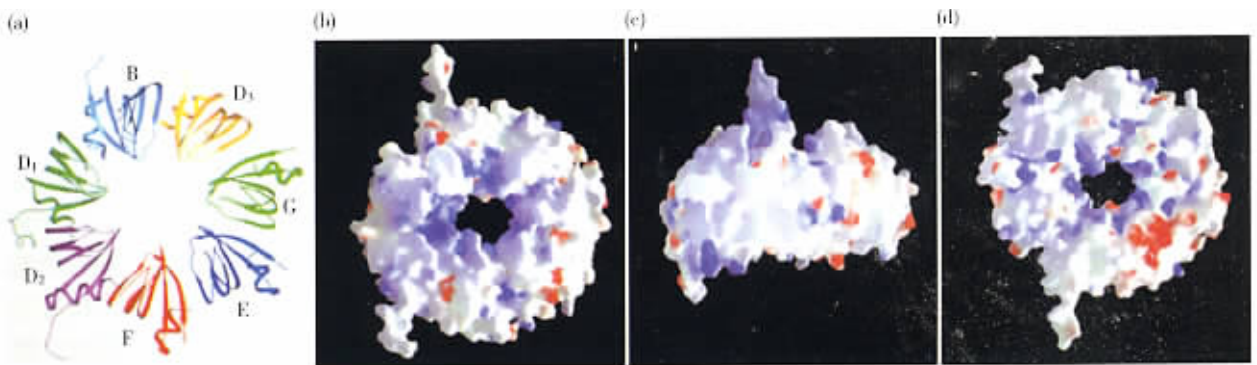


Fig. 1 A high order assembly of the human core snRNP proteins^[22]

(a) ribbon diagram of the heptamer model ; (b) surface representation of the heptameric ring with electronic potential (blue , positive ; red , negative) ; (c) side view ; (d) bottom view .

近几年的研究表明, snRNPP 的形成需要一种称之为 SMN (survival of motor neuron) 的蛋白质复合物 (含 Germin-2、-3、-4、-5 等蛋白质), 该基因的突变与遗传性脊柱肌肉萎缩症 (spinal

muscular atrophy , SMA) 的发病密切相关. SMN 复合物不但参与 pre-mRNA 剪接, 同时还参与转录体的形成. 组成复合物的 Germin 蛋白组分能直接与 7 种 Sm 蛋白相互作用, 通过富含尿嘧啶的 U

snRNAs 介导 snRNPs 的装配与形成. Sattler 等通过研究 SMN 的三维空间结构发现其 Tudor 结构域在 Sm 蛋白结合过程起关键作用, 该结构域通过一个保守的负电荷表面, 与 Sm 蛋白 C 端的 Arg 残基以及富含 Gly 残基的尾部相互作用^[23]. 最近, Dreyfuss 等^[24]观察到 SMN 能特异识别 U1 snRNA 的发卡结构 stem-loop 1 (SL1) 并与之结合. SL1 突变体的体外结合实验以及爪蟾胚胎的微量注射结果, 均证明了 SMN 与 SL1 相互作用对于 U1 snRNP 的装配至关重要.

3 KH-domain 剪接因子

在晶体及溶液结构方面研究得较为深入的 KH 结构域 (K-homologous domain) RNA 剪接因子是 FMR1 (fragile X mental retardation 1), 它仅仅与单链 RNA 结合. 研究表明, KH 结构域分子是一个蛋白质家族, 其中包括 vigilin (高密度脂蛋白的一种相结合蛋白, 含有 5 个 KH 结构域) 等. FMR1 的突变与缺失是一种人类遗传性智力低下, 即脆性 X 染色体综合症 (Fragile X syndrome) 发病的关键因素. Pastore 等应用 NMR 技术研究了 FMR1 蛋白 (含有两个 KH 结构域, KH1 和 KH2) 的溶液结构, 观察到 KH 结构域由 β 折叠和 α 螺旋交替排列而成, 即由 “ $-\beta_1-\alpha_1-\alpha_2-\beta_2-\beta_3-\alpha_3-$ ” 构成一个稳定的空间结构^[25, 26]. 进一步研究表明^[21], KH₁ 的 β_1 与 α_1 之间 (包括一个 -G-X-X-G-基序) 形成一个柔性很高的 loop, 能特异识别 RNA (poly-rG) 序列, 而 FMR1 分子 C 端含有 RGG-box 的区域与 RNA 的非特异性相互作用有关.

4 展 望

在参与 RNA 剪接的相关因子中, 仅有不到约 8% 蛋白质的三维空间结构进行了研究, 并且在已经研究的因子中, 只有部分 Sm 蛋白质, 如 B/B'、D1、D2、D3 等形成的复合物、hnRNP A1 以及 hnRNP D₀ 的结构得到了较为完整的结构. 尽管其他一些剪接相关蛋白, 如 DEA (D/H) -box RNA 螺旋酶家族的真核转译起始因子 eIF4A (eukaryotic translation initiation factor 4A), TIP-39 (tuberoindundibular peptide-39) 片段、hnRNP K、YB-1 (human Y-box protein 1) 等晶体结构也有报道, 但是大都是多肽片段与 RNA 复合物的三维结构. 近年来的研究表明, RNA 剪接不但与胚胎发育、性别确定、激素分泌和细胞衰老等密切相关,

还与人类某些重大疾病的病理过程相关, 如神经退行性疾病等^[27, 28]. 随着本领域研究的不断深入, RNA 剪接因子结构与功能的研究必将成为一个重要的国际前沿领域.

参 考 文 献

- 1 Zhou Z L, Lichlinder J J, Gygl S P, *et al.* Comprehensive proteomic analysis of the human spliceosome. *Nature*, 2002, **419** (12): 182 ~ 185
- 2 Hartmuth K, Urlaub H, Vornlocher H P, *et al.* Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. *Proc Natl Acad Sci USA*, 2002, **99** (26): 16719 ~ 16724
- 3 Zhou Z, Sim J, Griffith J, *et al.* Purification and electron microscopic visualization of functional human spliceosomes. *Proc Natl Acad Sci USA*, 2002, **99** (19): 12203 ~ 12207
- 4 Xu R M, Jokhan L, Cheng X, *et al.* Crystal structure of human UPI, the domain of hnRNP A1 that contains two RNA-recognition motifs. *Structure*, 1997, **5** (4): 559 ~ 570
- 5 Shamoo Y, Krueger U, Rice L M, *et al.* Crystal structure of the two RNA binding domains of human hnRNP A1 at 1.75 Å resolution. *Nat Struct Biol*, 1997, **4** (3): 215 ~ 222
- 6 Kajita Y, Nakayama J, Aizawa M, *et al.* The UAG-specific RNA binding protein, heterogeneous nuclear ribonucleoprotein D₀. Common modular structure and binding properties of the 2 × RBD-Gly family. *J Biol Chem*, 1995, **270** (38): 22167 ~ 22175
- 7 Nagata T, Kurihara Y, Matsuda G, *et al.* Structure and interactions with RNA of the N-terminal UAG-specific RNA-binding domain of hnRNP D₀. *J Mol Biol*, 1999, **287** (2): 221 ~ 237
- 8 Katahira M, Miyanoiri Y, Enokizono Y, *et al.* Structure of the C-terminal RNA-binding domain of hnRNP D₀ (AUF1), its interactions with RNA and DNA, and change in backbone dynamics upon complex formation with DNA. *J Mol Biol*, 2001, **311** (5): 973 ~ 988
- 9 Rokeach L A, Jannatipour M, Haselby J A, *et al.* Primary structure of a human small nuclear ribonucleoprotein polypeptide as deduced by cDNA analysis. *J Biol Chem*, 1989, **264** (9): 5024 ~ 5030
- 10 Oubridge C, Ito N, Evans P R, *et al.* Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*, 1994, **372** (6505): 432 ~ 438
- 11 Nagai K, Oubridge C, Ito N, *et al.* Crystal structure of the U1A spliceosomal protein complexed with its cognate RNA hairpin. *Nucleic Acids Symp Ser*, 1995, (34): 1 ~ 2
- 12 Avis J M, Allain F H, Howe P W, *et al.* Solution structure of the N-terminal RNP domain of U1A protein: the role of C-terminal residues in structure stability and RNA binding. *J Mol Biol*, 1996, **257** (2): 398 ~ 411
- 13 Pitici F, Beveridge D L, Baranger A M. Molecular dynamics simulation studies of induced fit and conformational capture in U1A-RNA binding: Do molecular substrates code for specificity? *Biopolymers*, 2002, **65** (6): 424 ~ 435
- 14 Price S R, Evans P R, Nagai K. Crystal structure of the spliceosomal U2B''-U2A protein complex bound to a fragment of U2 small nuclear RNA. *Nature*, 1998, **394** (6694): 645 ~ 650
- 15 Makarov E M, Makarova O V, Urlaub H, *et al.* Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science*, 2002, **298** (5601): 2205 ~ 2208
- 16 Reuter K, Nottrott S, Fabrizio P, *et al.* Identification, characterization and crystal structure analysis of the human

- spliceosomal U5 snRNP-specific 15 kD protein. *J Mol Biol*, 1999, **294** (2): 515 ~ 525
- 17 Jiang J, Horowitz D S, Xu R M. Crystal structure of the functional domain of the splicing factor Prp18. *Proc Natl Acad Sci USA*, 2000, **97** (7): 3022 ~ 3027
- 18 Jurica M S, Licklider L J, Gygi S R, *et al.* Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *RNA*, 2002, **8** (4): 426 ~ 439
- 19 Wiesner S, Stier G, Sattler M, *et al.* Solution structure and ligand recognition of the WW domain pair of the yeast splicing factor Prp40. *J Mol Biol*, 2002, **324** (4): 807 ~ 822
- 20 Reidt U, Reuter K, Achsel T, *et al.* Crystal structure of the human U4/U6 small nuclear ribonucleoprotein particle-specific SnuCyp-20, a nuclear cyclophilin. *J Biol Chem*, 2000, **275** (11): 7439 ~ 7442
- 21 Walke S, Bragado-Nilsson E, Seraphin B, *et al.* Stoichiometry of the Sm proteins in yeast spliceosomal snRNPs supports the heptamer ring model of the core domain. *J Mol Biol*, 2001, **308** (1): 49 ~ 58
- 22 Kambach C, Walke S, Young R, *et al.* Crystal structure of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell*, 1999, **96** (3): 375 ~ 387
- 23 Selenko P, Sprangers R, Stier G, *et al.* SMN tudor domain structure and its interaction with the Sm proteins. *Nat Struct Biol*, 2001, **8** (1): 27 ~ 31
- 24 Yong J, Pellizzoni L, Dreyfuss G. Sequence-specific interaction of U1 snRNA with the SMN complex. *EMBO J*, 2002, **21** (5): 1188 ~ 1196
- 25 Musco G, Stier G, Joseph C, *et al.* Three-dimensional structure and stability of the KH domain: molecular insights into the fragile X syndrome. *Cell*, 1996, **85** (2): 237 ~ 245
- 26 Musco G, Kharrat A, Stier G, *et al.* The solution structure of the first KH domain of FMR1, the protein responsible for the fragile X syndrome. *Nat Struct Biol*, 1997, **4** (9): 712 ~ 716
- 27 王俊峰, 廖祥儒, 付伟. 小型核酶的结构和催化机理. *生物化学与生物物理进展*, 2002, **29** (5): 674 ~ 677
- Wang J F, Liao X R, Fu W. *Prog Biochem Biophys*, 2002, **29** (5): 674 ~ 677
- 28 王付龙, 徐祥, 梁华平, 等. 转录因子圈套策略研究进展. *生物化学与生物物理进展*, 2001, **28** (6): 802 ~ 805
- Wang F L, Xu X, Liang H P, *et al.* *Prog Biochem Biophys*, 2001, **28** (6): 802 ~ 805

Structural Analyses of RNA Splicing Factors : Progress and Perspective *

WU Ying^{1 2)}, HE Rong-Qiao^{1)**}

(¹)Laboratory of Visual Information Processing, Institute of Biophysics, The Chinese Academy of Sciences, Beijing 100101, China;

²⁾Washington University, School of Medicine, St. Louis, MO 63110, USA)

Abstract Pre-mRNA splicing is a critical step in eukaryotic gene expression and regulation. The removal of introns occurs in spliceosome, a macromolecular machine containing small nuclear ribonucleoproteins (snRNP), heterogeneous nuclear ribonucleoproteins (hnRNP) and K-homologous domain (KH) proteins. Crystallographic and NMR studies have begun to provide insight into the architecture of snRNP and the structural basis for RNA-protein as well as protein-protein interactions involved in pre-mRNA splicing. Elucidating molecular mechanisms underlying pre-mRNA splicing awaits systematic structural analyses of splicing factors and complexes.

Key words RNA splicing, snRNA binding protein, hnRNA binding protein, spliceosome, structure, RNA binding domain

* This project is supported by grants from Nano-Bioscience of National Natural Foundation of Sciences (90206041) and Creative Foundation of The Chinese Academy of Sciences (KSCX2-SW214).

** Corresponding author. Tel : 86-10-64889876, E-mail : herq@sun5.ibp.ac.cn

Received : February 10, 2003 Accepted : March 24, 2003