

微型述评

RRM RNA 结合蛋白的结构与功能

杜光伟 周 严 袁建刚 强伯勤

(中国医学科学院基础医学研究所, 医学分子生物学国家重点实验室, 北京 100005)
(中国协和医科大学基础医学院)

摘要 RRM RNA 结合蛋白是一类含一个或数个 RRM 结构域及附属结构域的 RNA 结合蛋白, 参与 RNA 前体的剪接、RNA 的细胞定位、RNA 的稳定性等多种转录后调控过程。在 RRM 基序中含有许多保守的氨基酸以保证对 RNA 的结合活性, 但是这一家族的不同蛋白质却能特异地结合各种不同的 RNA 分子。RRM RNA 结合蛋白与某些人类遗传性疾病及肿瘤相关。

关键词 转录后调控, RNA 结合蛋白, RRM 基序

学科分类号 Q522

哺乳动物组织细胞的生长和分化过程中, 基因的表达在转录、转录后和翻译水平都受到严格的调控。转录调控是基因表达调控的第一步, 曾被认为是基因表达的主要调控机制。但是随着对转录后调控机制越来越多的了解, 发现转录后调控在基因表达调控中同样起着重要作用。基因从核内转录后, 新合成的 pre-mRNA 或核异质 RNA (hnRNA) 通过加帽、pre-mRNA 的剪接和合成多聚腺苷酸 (polyA) 尾等一系列的反应而成为成熟的 mRNA。成熟的 mRNA 被转运送至细胞质, 在细胞质内 mRNA 的定位、翻译和代谢等过程受到严格的调控。转录后的调控决定着基因表达最终产物的种类和表达量^[1]。

在转录后调控的过程中涉及到多种蛋白质, 其中最重要的是各种 RNA 结合蛋白的参与。随着越来越多的 cDNA 及其产物的鉴定, 发现了多种具有 RNA 结合基序的 RNA 结合蛋白。这些蛋白质参与 RNA 在转录后受调控的每一步骤, 其结构与功能的异常往往可导致疾病的发生。

已知的 RNA 结合蛋白中, 研究得最多的一类是具有 RRM (RNA recognition motif) 的 RNA 结合蛋白。RRM 也被称为 RNP 基序 (ribonucleoprotein motif), 或 CS-RBD (consensus sequence RNA-binding domain)^[1]。目前, 在动物、植物、真菌和细菌等不同物种中已经发现约有 300 个 RRM RNA 结合蛋白。本文主要介绍 RRM RNA 结合蛋白的结构、功能及与某些疾病的关系。

1 RRM RNA 结合蛋白的结构

RRM RNA 结合蛋白的明显特征是这类蛋白质都含有一个或数个明显的 RRM。在某些 RRM RNA 结合蛋白中还含有一个或数个位于蛋白质不同位置的附属结构域。RRM 通常在序列上相对保守, 而附属的结构域在不同的蛋白质中各不相同。

1.1 RRM 的结构

RRM 通常由 80~90 个氨基酸组成, 其中含有分别具 8 个和 6 个保守氨基酸的 RNP1 和 RNP2 基序, 在 RNP1 和 RNP2 的两侧是一些散布的疏水性氨基酸。RRM 的 $\beta\alpha\beta\alpha\beta$ 的二级结构形成 4 条反向平行 β 片层结构, 2 个 α 螺旋与 β 片层结构的方向相垂直。RNP1 和 RNP2 位于 β_3 和 β_1 片层结构的中间位置。从 RRM 的三维结构显示了 RRM 保守氨基酸的 2 个作用。第一, RNP1 和 RNP2 的带电和芳香族侧链在溶剂中暴露在外侧, 并可能通过氢键和环的堆积力直接与 RNA 结合。第二, 位于 RNP1 最后位置的芳香侧链指向折叠的结构域内部, 和两个 α 螺旋的其他高度保守的疏水氨基酸形成结构域的疏水核心^[2]。

1.2 RRM RNA 结合蛋白的附属结构域

RRM RNA 结合蛋白除了具有一至四个 RRM 结构域外, 通常还含有其他附属结构域。有趣的是, RRM RNA 结合蛋白附属结构域存在转录因子

激活结构域中常见的特征, 如hnRNP C具有富含酸性氨基酸的附属结构域. 从结构和功能相似性来看, 有两种 RRM RNA 结合蛋白的附属结构域已经比较清楚^[3]. 这两者的空间结构不一样, 但是都具有明显的 RNA 结合活性. 第一类是在大部分核异质核糖核蛋白, 核仁蛋白与核仁纤维蛋白等中找到的 Gly 富含结构域. 第二类附属结构域是 SR 结构域 (富含丝氨酸和精氨酸), 存在于参与剪接的 SR 家族中. SR 结构被认为是参与剪接的活性标志.

2 RRM RNA 结合蛋白的 RNA 结合活性

RNA 的结合实验表明可能与 RNA 相互作用的三个 RRM 的主要结构是: β 片层结构, 连接片层结构的环, RRM 的氨基和羧基两端的区域. 虽然 RNP1 和 RNP2 的高保守氨基酸对 RNA 的结合是必需的, 但是可能并不能区分不同的 RNA 分子. RNA 结合的特异性主要存在于 RRM 基序的可变区, 特别是在环和末端的位置. 例如, 将连接 snRNP U1A $\beta 2$ 和 $\beta 3$ 的氨基酸 (环 3) 用另一 snRNP 蛋白 U2B' 的相应氨基酸替代后, 杂合蛋白表现为 U2B' 的结合特异性^[4]. 对 U1 70K, U1A 和 hnRNP C 三种蛋白的突变分析发现紧接着 RRM 羧基端的氨基酸对 RNA 的结合特异性是高度必需^[3]. 因此虽然在 RRM 基序中含有许多保守的氨基酸以保证对 RNA 的结合活性, 但是这一家族的不同蛋白质却能特异地结合各种不同的 RNA 分子^[1].

RRM RNA 结合蛋白与 RNA 的结合已经被许多实验数据所证实. 含多个 RRM 的许多 RNA 结合蛋白 (如 hnRNPA1, PABP, U2AF⁶⁵ 等) 对 RNA 的结合特异性需要连续的 RRM. 具有多个 RRM 结构域的 RNA 结合蛋白的 RNA 结合活性并不仅仅是单个 RRM 活性的总和. 某些蛋白质的多个 RRM 能同时结合不同的 RNA 序列. 例如, U1A 的第一个 RRM 与 U1 SnRNA 结合, 第二个 RRM 与 mRNA 前体结合. 此外, 在某些情况下, 富含 Gly 附属结构域和 SR 附属结构域也参与 RNA 的加工, 但它们对 RNA 的结合都是非特异的. Biamonti 等^[3]提出了一种 RNA 结合的机制, RNA 结合蛋白的附属结构域非特异地将蛋白停靠在 RNA 上, 从而使相应的 RRM 结构域能特异地与 RNA 结合. RRM 的构象在结合 RNA 和未结合 RNA 时是基本一致的. 暴露的 β 片层 RNA 结合表

面与 RNA 衔接形成一个开放的平台, 而不是将 RNA 埋藏在结合的裂缝中. 结合的 RNA 保持相对暴露可以使它与其他 RNA 序列或 RNA 结合蛋白相互作用. 此外, 结合在 RNA 上的蛋白质为 RNA 分子在细胞不同位置的定位提供了信号.

在发育的过程中, 对 RNA 的调控将使细胞对刺激信号快速反应以适应不同的生理需要, 这一速度要比从转录开始到蛋白质表达要快得多. 外界信号并不是直接作用于基因的转录产物上, 而是可能通过控制参与 RNA 剪接的反式作用因子的磷酸化来调控的. 例如, 磷酸化后可能改变了蛋白质的 RNA 结合活性、亲和力或特异性, 激活 RNA 的翻译或通过 RNA 结合蛋白将 RNA 运送到细胞质内特定的位置.

3 RRM RNA 结合蛋白与疾病

RRM RNA 结合蛋白在细胞代谢过程中起着重要作用. 在人和多种其他生物中已经发现许多 RRM RNA 结合蛋白在表达量变化或突变后将引起发育异常和疾病.

Elav 类蛋白是一类密切相关的 RNA 结合蛋白. *Elav* (embryonic lethal abnormal visual phenotype) 基因最先在果蝇中克隆, 并发现其产物在神经系统的发育和功能维持中起重要作用. 最近已克隆了这一家族的多个成员^[5]. Elav 类蛋白含有 3 个高度保守的 RRM 结构域. 其中两个是串联重复的, 第三个 RRM 与前两个 RRM 由富含碱性氨基酸的片段分开. Elav 家族在人中有 4 个成员, HuD, HuC, He1N1 和 HuR. 正常情况下 HuD, HuC 和 He1N1 常表达于终末分化的神经元. HuR 则是一种在多种组织中都有表达的基因. 已经发现 Elav 类蛋白与某些神经系统的自身免疫疾病有关, Elav 改变会造成神经组织的破坏和神经自身免疫疾病的发生, 出现痴呆病、小脑退行性病变、脑干炎或脊髓炎等表现.

生育力的下降影响到世界上 2% ~ 12% 的夫妇. 大部分男性不育是由于不能产生足够的精子或根本不产生精子. 过去研究发现在 Y 染色体长臂远端共有 3 个非重叠的区域, 与正常精子的产生都有关, 这 3 个区域被称为 AZFa_c (azoospermia factor, AZF). 在 AZFb 和 AZFc 区域, 已克隆多个候选基因, 非常有趣的是, 这些基因都是 RRM RNA 结合蛋白. RBM (RNA binding motif protein) 是在 Y 染色体上首先克隆的男性不育相关的

候选基因, RBM 具有一个 RRM, 在人和鼠的雄性生殖细胞中特异表达. 最近又克隆了另外两个位于 AZFc 的候选基因, *DAZ* (deleted in azoospermia) 和 *SPGY* (spermatogenesis gene on the Y). RBM 与 *DAZ*、*SPGY* 的序列差异较大, 但是都具有相同的组件结构, 即具有单一的 RRM 和一段重复序列. 在临床检测中发现, 不育患者含 *DAZ/SPGY* 区域的高频率缺失 (不算点突变) 达到了 5%, 说明 *DAZ/SPGY* 基因在人精子发生过程的重要作用.

过去研究发现人染色体易位 t (12; 16) (q13; p11) 会导致人粘液样脂肪肉瘤 (liposarcoma). Crozat^[5] 克隆了含有一个 RRM 的 *TLS* (translocated in liposarcoma) 基因, 并发现 *TLS* 的 RRM 被 CHOP (转录因子 C/EBP 家族的一个成员) 的 DNA 结合和亮氨酸拉链结构域替代后会引致人粘液样脂肪肉瘤. *EWS* 是一种与 *TLS* 有着很高同源性的 RRM RNA 结合蛋白. 造成 Ewing's 肉瘤和其他原发的神经外胚层肿瘤的 t (11; 22) (q24; q12) 平衡易位涉及到 *EWS* 基因. 有趣的是, 在 Ewing's 肉瘤中, 也是 *EWS* 的 N 端与另一类转录因子家族成员 *FLI-1* 的 DNA 结合结构域相融合. 结构上高度相似的这两个蛋白质在 Ewing's 肉瘤和粘液样脂肪肉瘤存在类似的发病机制, 说明 RNA 结合蛋白的 RRM 和其他转录因子的 DNA 结合结构域的替换可能代表一种新的肿瘤发生的分子机制^[5]. 近几年, 又发现 *TLS* 和 *EWS* 与其他转录因子的融合而造成多种肿瘤的发生, 进一步证实了某些 RNA 结合蛋白在肿瘤发生中的重要作用.

4 结束语

目前对 RRM RNA 结合蛋白功能的了解还不够深入. 但是已经知道这一类蛋白质参与 RNA 前体的剪接、RNA 的细胞定位、RNA 维持稳定等多种转录后调控过程, 多种 RRM RNA 结合蛋白的突变将会造成疾病的发生. 由于这一类蛋白质具有重要的生理功能和分子功能, 有必要对它们进一步深入地研究. 今后的研究应该引入一些新的思路和方法, 重点研究与其他蛋白质的相互作用、在信号传导通路中的作用、对所结合 RNA 序列的选择方

式等内容. 此外, 对各种 RRM RNA 结合蛋白之间的关系虽然有了初步的分析, 但是还不够全面, 分类的标准也不统一. 有必要在收集所有已知 RRM RNA 结合蛋白, 结合现有的功能信息, 对 RRM RNA 结合蛋白作一个系统的分类. 这将为今后深入了解 RRM RNA 结合蛋白的功能提供重要依据.

参 考 文 献

- 1 Burd C G, Dreyfuss G. Conserved structure and diversity of functions of RNA-binding proteins. *Science*, 1994, **265** (5172): 615 ~ 621
- 2 Kenan D J, Query C C, Keene J D. RNA recognition: towards identifying determinants of specificity. *TIBS*, 1991, **16** (6): 214 ~ 220
- 3 Biamonti G, Riva S. New insights into the auxiliary domains of eukaryotic RNA binding proteins. *FEBS Letters*, 1994, **340** (1~2): 1~8
- 4 Crozat A, Aman P, Mandahl N, *et al.* Fusion of CHOP to a novel RNA-binding protein. *Nature*, 1993, **363** (6430): 640~644
- 5 Antic D, Keene J D. Insights from model systems: Embryonic lethal abnormal visual RNA-binding proteins involved in growth, differentiation, and posttranscriptional gene expression. *Am J Hum Genet*, 1997, **61** (2): 273~278

RRM RNA Binding Protein: Structure and Function. DU Guang-Wei, ZHOU Yan, YUAN Jiann-Gang, QIANG Bo-Qin (*Institute of Basic Medical Science, Chinese Academy of Medical Sciences & Peking Union Medical College, National Laboratory of Medical Molecular Biology, Beijing 100005, China*).

Abstract RRM RNA binding protein is a kind of RNA binding protein which contains one or more RNA recognition motifs and some affiliated motifs. It participates in genes' post-transcriptional regulations including splicing of pre-mRNA, cellular localization and stability maintenance of RNA. Conservative amino acids in the RRM motif are essential for RNA binding, but the specific binding depends on other factors. Some RRM RNA binding proteins are related to genetic diseases and tumors.

Key words posttranscriptional regulation RNA, binding protein, RRM motif