

Abstract The Wnt genes belong to an ever-expanding family of proto-oncogenes that express in species ranging from *Drosophila* to man. Wnt proteins, as well as other growth factors or signaling molecules, have the characteristics of secreted growth factors. As recently, a group of proteins in Frizzled (Fz) family have been identified as the putative receptors of Wnt family, several components are implicated in the downstream of Wnt/Fz signaling

pathway, such as Dishevelled, GSK3 (glycogen synthase kinase 3), APC (adenomatous polyposis coli) and β -catenin. This signal transduction pathway plays a pivotal role not only in the formation and organization of developing organs, but also in biological tumorigenesis.

Key words Wnt, Frizzled, biological development, oncogenesis

RNA 和蛋白质的相互作用

张庆硕 王恩多

(中国科学院上海生物化学研究所, 分子生物学国家重点实验室, 上海 200031)

摘要 RNA 与蛋白质的相互作用是许多基本的细胞生理过程得以实现的决定性因素。近年来, 随着技术的改进和新方法的建立, RNA 和蛋白质的相互作用研究取得了长足进步。目前科研人员已经鉴定了许多 RNA 上的蛋白质结合位点, 也发现了许多蛋白质中的 RNA 结合结构域, 并对它们的结构特征进行了比较详细的研究。这些都为最终阐明 RNA 和蛋白质相互作用的分子机制, 从而从本质上认识相关的细胞生理过程打下了坚实的基础。

关键词 识别, 结构域, RNA 结合蛋白

学科分类号 Q512, Q522

RNA 参与许多基本的细胞生理过程: 携带来自 DNA 的遗传信息, 参与形成核糖体、拼接体、端粒酶等许多核酸蛋白颗粒的结构, 有些 RNA 还具有酶活性。但是, 几乎所有的 RNA 生物功能的发挥都需要蛋白质可逆或不可逆地结合。而且, 在绝大多数情况下, 蛋白质与 RNA 的相互作用在上述那些生理过程中起着决定性的作用^[1]。

RNA 和蛋白质相互作用的研究过去一直远远落后于 DNA 和蛋白质相互作用的研究, 主要原因是很难大量制备序列确定的纯 RNA。近年来, 科研工作者建立了运用噬菌体 RNA 聚合酶进行体外转录的系统并改进了 RNA 化学合成的技术, 大大促进了 RNA 与蛋白质相互作用的研究^[2]。目前已经鉴定了许多 RNA 中的蛋白质结合位点, 也发现了为数不少的蛋白质中的 RNA 识别结构域。

1 蛋白质识别的 RNA 的二级结构单元

RNA 的结构非常复杂。RNA 链内的碱基配对产生各种各样的二级结构, 除螺旋 (helix) 外, 还有发夹 (hairpin)、膨泡 (bulge)、内环 (bubble)

和假结 (pseudoknot)。这些结构单元之间进一步的相互作用使 RNA 形成复杂的三级结构。另外, RNA 中还含有为数不少的非沃森-克里克碱基对如 G-U 和 C-U 等。这些不规则的碱基对使经典的 A 型螺旋扭曲变形, 大沟加宽。膨泡和内环也可以使核糖磷酸骨架扭曲而改变沟的宽度。

RNA 的复杂结构使 RNA-蛋白质的识别比 DNA-蛋白质的识别要复杂得多。DNA 具有规则的双螺旋结构, 差别主要在于碱基对 (一般是相对暴露的大沟中的碱基对) 所提供的氢键受体和供体的不同, 因而 DNA 和蛋白质的相互作用主要是蛋白质中的基团和 DNA 中的碱基对形成专一性的氢键及其与胸腺嘧啶的甲基间的疏水作用, 这种“直接读出” (direct readout) 机制已经证明基本上是正确的。尽管某些 DNA 的骨架构象也会显示出一些微妙的变化, 如在 *EcoR* I 核酸酶位点中发现有纽结 (kinks) 结构^[3], 在 TATA 结合蛋白与 DNA 形成的复合物中发现有螺旋的强烈弯曲和碱基对去堆积现象^[4], 但是蛋白质和 DNA 的这种特殊构象

间的相互识别即“间接读出”(indirect readout)机制只占非常小的比例。与此相反, RNA 不规则的二级结构和复杂的三级结构使 RNA 与蛋白质的识别以“间接读出”机制为主。所有的 RNA 结构, 包括发夹、膨泡、内环、假结和螺旋本身都可以作为蛋白质专一性识别的目标。而且, RNA 的独特的三级结构使蛋白质同 RNA 的核糖磷酸骨架之间也可以产生专一性的识别, 甚至于有的蛋白质识别相应的 RNA 可以仅仅通过专一性地识别该 RNA 的整体构象来实现, 这是“间接读出”机制的一种极端情况。

下面通过介绍一些研究得比较深入的 RNA 结

合蛋白来说明 RNA 是如何被蛋白质识别的。

1.1 HIV Rev ——内环结合蛋白

内环是指在一段 RNA 的双链区中间由于两条链上都有未配对的碱基而产生的一个环状结构^[5]。艾滋病病毒 (HIV) 的 Rev 蛋白结合的 Rev 应答元件 (Rev responsive element, RRE) 就是一个 RNA 内环 (图 1a)。该内环富含嘌呤碱基, 其中环的部分含有 GG 和 GUA 残基并形成 G-G、G-A 不规则碱基对^[6]。内环是 Rev 蛋白结合所必需的, 单独内环也足以产生与 Rev 高亲和力的结合。环区的 G-G 对可以用 A-A 对替代, 因为它们可以使结合位点的糖磷酸骨架产生相同的形变。

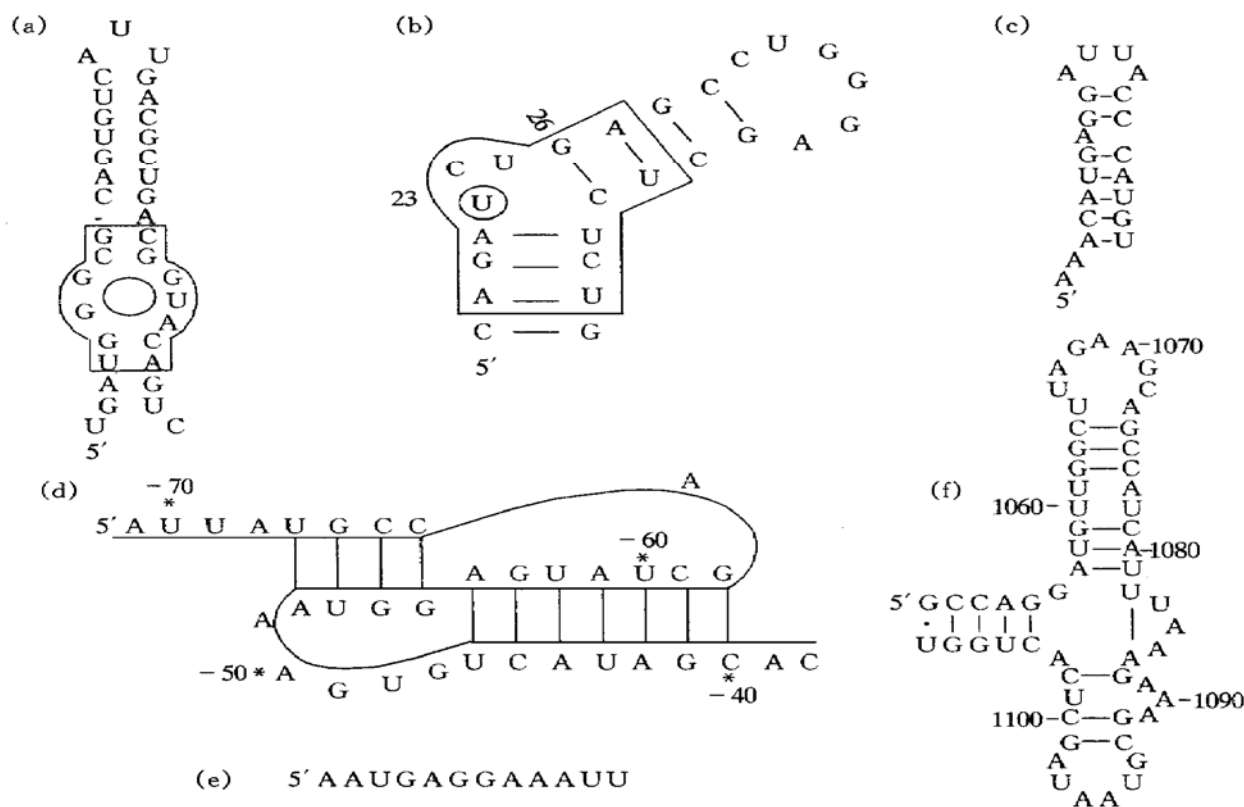


图 1 一些 RNA 结合蛋白识别的 RNA 的结构单元

(a) Rev 蛋白识别的一个内环结构, (b) Tat 蛋白识别的一个 RNA 膨泡, (c) R17 外壳蛋白识别的一个 RNA 发夹, (d) T4 基因 32 蛋白识别的一个 RNA 假结^[7], (e) regA 蛋白的识别位点——一段简单的 RNA 序列, (f) L11 蛋白识别的 RNA 的片段^[11]。

1.2 HIV Tat ——膨泡结合蛋白

膨泡是指在一段 RNA 的双链区只有一条链上含有未配对的碱基^[5]。HIV Tat 蛋白结合病毒 mRNA 的 Tat 应答区 (Tat responsive region, TAR) 内的一个 RNA 膨泡 (图 1b), 从而允许病毒转录物积聚。TAR 应答区形成一个具有长茎的发夹结构, 茎区含有一个 UCU 膨泡^[7]。该膨泡必须具备

的基本结构是: 环上的碱基至少要有两个, 而且含有 5' 端的 U (即 U23); G26/C39 和 A27/U38 碱基对不能改变; A22 和 U23 的 5' 磷酸被乙基保护。

1.3 R17 外壳蛋白 ——发夹结合蛋白

发夹结合蛋白很常见。位于环区的碱基比位于 A 型双螺旋区的碱基更加暴露, 所以发夹的环区是蛋白质结合的很好的位点。典型的例子包括核糖体

蛋白、snRNP (small nuclear ribonucleoprotein) U₁A、U₁70K 和 U₂B, 还有一些噬菌体外壳蛋白. 其中研究得最清楚的是 R17 噬菌体的外壳蛋白和一个发夹结构的相互作用 (图 1c)^[8]. R17 外壳蛋白的 RNA 识别位点具有以下三个基本特征^[1]: 第一, 发夹是必需的; 第二, 序列为 RNYA (R= 嘌呤, N= 任意碱基, Y= 嘧啶) 的一个四碱基的发夹环对结合最有利; 第三, 茎区特定位置上一个不配对的嘌呤碱对识别很重要.

1.4 假结结合蛋白

当一个 RNA 环和其紧邻的单链区形成碱基对时, 就可以形成假结结构. T4 噬菌体基因 32 蛋白结合在它自身的 mRNA 上, 对翻译进行自调控^[9]. 这个 mRNA 的 5' 端形成一个假结 (图 1d), 基因 32 蛋白和这个假结的结合能进一步增强其与 mRNA 的结合. 核糖体蛋白 S4 和 S15 也都结合假结结构.

以上详细介绍了一些被蛋白质识别的 RNA 结构单元. 与此相比, T4 噬菌体 regA 蛋白识别的 RNA 位点是一段与起始密码子重叠的短序列, 该位点不存在任何稳定的二级结构 (图 1e). 很可能这是最简单的一种情况, 即蛋白质识别 RNA 上特定的单链序列. 然而, 大多数蛋白质识别的 RNA 结构都象上面介绍的那样, 比较复杂, 任何一种 RNA 结构都可能是潜在的结合位点. 在另外的一些情况下, 如核糖体大亚基蛋白 L11 和一个长达 58 个碱基的 rRNA 结构域的结合, L11 很可能不是识别碱基, 而是在三维结构的水平上识别 RNA 的骨架 (图 1f).

2 蛋白质中的 RNA 结合结构域

为了确定识别 RNA 的蛋白质结构, 研究人员在众多的 RNA 结合蛋白中寻找氨基酸序列同源性. 目前已经发现了许多 RNA 结合结构域 (RNA binding domain, RBD, 表 1).

2.1 RNP 模体

这是研究得最多也是了解得最清楚的 RNA 结合结构域, 又名 RNA 结合蛋白的共有序列 (RNP consensus sequence, RNP-CS)、共有序列 RNA 结合结构域 (consensus sequence RNA-binding domain, CS-RBD) 或 RNA 识别模体 (RNA recognition motif, RRM). 目前已经在 200 多种 RNA 结合蛋白中发现这一结构域, 这些蛋白分别结合 RNA 前体、mRNA、rRNA 和 snRNA, 涉及

RNA 的加工、转运和代谢等多方面. RNP 模体的一致序列长为 80~90 个氨基酸残基, 含有两段短的强保守序列 RNP1 和 RNP2. RNP1 和 RNP2 长度分别为 8 个和 6 个氨基酸残基, 含有许多保守的疏水氨基酸, 特别是酪氨酸和苯丙氨酸, 二序列之间间隔约 30 个氨基酸残基^[10]. RNP 模体的二级结构模式为 $\beta\alpha\beta\beta\alpha\beta$, 其中四股反平行的 β 折叠片位于中间, α 螺旋位于两侧. RNP1 和 RNP2 位于中间的两股 β 折叠上, 在 RNA 的结合中起至关重要的作用. 保守的芳香族氨基酸中, 有两个突出到溶液中, 其余的则和另外的疏水氨基酸一起形成疏水内核.

RNA 结合研究表明, RNP 模体中有三个精细的结构元件可能与 RNA 相互作用: β 折叠、连接 β 折叠的环区、氨基和羧基端邻近区域^[10]. RNP1 和 RNP2 上高度保守的氨基酸尽管对 RNA 的结合至关重要, 但是可能并不参与识别不同的 RNA 序列. 与 RNA 结合专一性有关的主要识别子位于 RNP 模体上最可变的区域, 特别是末端和环区. 这表明: RNP 模体的 β 折叠, 含有许多高度保守的残基, 构成了一个结合表面; 高度可变的精细的识别子依托这一结构, 决定其 RNA 结合专一性. 上述结果也合理地解释了为什么这一家族的不同蛋白含有许多高度保守的结构元件却能结合功能如此之多的 RNA 分子.

2.2 富含精氨酸模体

富含精氨酸模体 (the arginine-rich motif, ARM) 发现于病毒和核糖体的一些蛋白中, 是很短 (10~20 个氨基酸) 的富含精氨酸的序列, 除了都富含精氨酸外, 几乎没有同源性. 上文提到的 Rev 和 Tat 都是 ARM 蛋白. Rev ARM 专一性结合 RRE 的螺旋, 至少有 6 个氨基酸 (包括 4 个精氨酸) 是专一性结合必需的.

RNA 上的 ARM 识别位点很复杂, 包括茎环 (N 蛋白)、内环 (Rev) 或膨泡 (Tat), 而且最主要的识别元件可能是其构象而非一级结构的特殊序列. Rev 和 RRE 的相互作用已明确地显示了这一点. 除了 RRE 和 Rev 之间可能的氢键外, RRE 的一个关键的结构组成是非沃森克里克碱基对代替, 这表

一性识别位点. RNA 结合和模建实验还表明, TAR 和 RRE 中的膨泡核苷酸使本来很深的大沟扭曲变形, 从而能形成氢键的原子得以与其接近.

2.3 双链 RNA 结合结构域

该结构域一致序列长为 65~ 68 个氨基酸残基, 同源性约为 30%, 发现于双链 RNA 结合蛋白中^[1]. 有一些双链 RNA 结合蛋白专一性结合双链 RNA 或 DNA-RNA 杂合双链, 大肠杆菌 RNase III 则对含有特定内环结构的螺旋结合最强. RNase III 和果蝇 staufer 蛋白的双链 RNA 结合结构域 (the double stranded RNA binding domain, dsRBD) 含有一个 α - β - β - α 拓扑结构, 三股反平行的 β 折叠在内, 外侧每边一段 α 螺旋. 环 2 (位于第一和第二 β 折叠之间) 和环 4 (位于第三股 β 折叠和第一个 α

螺旋之间) 上含有一些带正电荷的残基及一些氢键供体和受体, 可能参与 RNA 的结合^[11].

另外的 RNA 结合结构域还有 RGG Box (Arg-Gly-Gly 重复)^[12]、KH 结构域 (the K homology domain, the KH domain)^[13]、锌指结构 (zinc fingers)^[14]、冷休克结构域 (cold shock domain)^[15]、Smr-protein 模体^[16] 等. 其中需要说明的是, 现在认为锌指结构主要结合 DNA, 而非 RNA, 尽管它们最初是作为 RNA 结合蛋白被发现的, 但是它们在结合 DNA 的调控蛋白中更常见.

表 1 RNA 结合蛋白的 RNA 结合结构域

结构域名字	长度 (氨基酸)	例子	参考文献
RNP 模体	80	U snRNP 蛋白、hnRNP 蛋白、RNA 加工和翻译的调控因子	[10]
富含精氨酸模体	12~ 18	噬菌体 N 蛋白, HIV Rev, Tat	[10]
双链 RNA 结合结构域	65~ 68	双链 RNA 激活的蛋白激酶 DAI、果蝇 staufer 蛋白、人类细胞 TAR 结合蛋白	[11]
RGG box		hnRNP 蛋白、核蛋白	[12]
KH 结构域	50	hnRNP 蛋白、核糖体蛋白 S3、酵母拼接调节因子 MER1、NusA、GAP-associated p62、FMR-1	[13]
锌指结构	23	TF III, p43, XFG 5-1	[14]
冷休克结构域	70~ 80 15~ 25	DNA CCAAT-或 Y box 结合蛋白爪蟾 mRNP 蛋白 p54, p56	[15]
Smr-protein 模体		SnRNP 蛋白	[16]

3 结 论

尽管已经发现了不少蛋白质中的 RNA 识别结构域, 也鉴定了许多 RNA 中的蛋白质结合位点, 但是要想深入认识 RNA 与蛋白质的相互作用的分子机制, 必须解出大量的 RNA 与蛋白质复合物的精细结构. 这要借助于 X 光晶体衍射和核磁共振技术, 它们可以提供原子水平的结构信息. 然而, 目前只有三种氨酰 tRNA 合成酶与其相关 tRNA 的复合物和两种蛋白质与 RNA 发夹的复合物的晶体结构是已知的, 核磁共振也只在研究氨基酸或小肽和 RNA 的复合物的结构方面比较成功.

参 考 文 献

- 1 Draper D E. Protein-RNA recognition. Annu Rev Biochem, 1995, 64: 593~ 620
- 2 Milligan J F, Groebe D R, Witherell G W, *et al.* Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. Nucleic Acids Res, 1987, 15 (21): 8783~ 8789
- 3 McClarin J A, Frederick C A, Wang B C, *et al.* Structure of the DNA-EcoRI endonuclease recognition complex at 3 Å resolution. Science, 1986, 234 (4783): 1526~ 1541
- 4 Kim J L, Nikolov D B, Burley S K. Co-crystal structure of TBP recognizing the minor groove of a TATA element. Nature, 1993, 365 (6446): 520~ 527
- 5 Chastain M, Tinoco I. Structural elements in RNA. Prog Nucleic Acid Res Mol Biol, 1991, 41: 131~ 177
- 6 Heaphy S, Finch J T, Gait M J, *et al.* Human immunodeficiency virus type 1 regulator of virion expression, Rev, forms nucleoprotein filaments after binding to a purine rich bubble located within the Rev-responsive region of viral mRNAs. Proc Natl Acad Sci USA, 1991, 88 (16): 7366~ 7370
- 7 Nagai K. RNA-protein interactions. Curr Opin Struct Biol, 1992, 2 (2): 131~ 137
- 8 Witherell G W, Gott J M, Uhlenbeck O C. Specific interaction

- between RNA phage coat proteins and RNA. *Prog Nucleic Acid Res Mol Biol*, 1991, **40**: 185~ 220
- 9 Puglisi J D, Wyatt J R, Tinoco I. Conformation of an RNA pseudoknot. *J Mol Biol*, 1990, **214** (2): 437~ 453
- 10 Burd C G, Dreyfuss G. Conserved structures and diversity of functions of RNA-binding proteins. *Science*, 1994, **265** (5172): 615 ~ 621
- 11 Nagai K. RNA-protein complexes. *Curr Opin Struct Biol*, 1996, **6** (1): 53~ 61
- 12 Girard J P, Lehtonen H, Caizergues-Ferrer M, *et al.* GAR1 is an essential small nucleolar RNP protein required for pre-rRNA processing in yeast. *EMBO J*, 1992, **11** (2): 673~ 682
- 13 Siomi H, Matunis M J, Michael W M, *et al.* The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. *Nucleic Acids Res*, 1993, **21** (5): 1193~ 1198
- 14 Miller J, McLachlan A D, Klug A. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J*, 1985, **4** (6): 1609~ 1614
- 15 Murray M T, Schiller D L, Franke W W. The sequence analysis of cytoplasmic mRNA-binding proteins of *Xenopus* oocytes identifies a family of RNA-binding proteins. *Proc Natl Acad Sci USA*, 1992, **89** (1): 11~ 15
- 16 Cooper M, Johnston L H, Beggs J D. Identification and characterization of Uss1p (Sdb23p): a novel U6 snRNA-associated protein with significant similarity to core proteins of small nuclear ribonucleoproteins. *EMBO J*, 1995, **14** (9): 2066~ 2075

RNA-Protein Interactions. ZHANG Qing-Shuo, WANG Er-Duo (*State Key Laboratory of Molecular Biology, Shanghai Institute of Biochemistry, The Chinese Academy of Sciences, Shanghai 200031, China*)

Abstract RNA-protein interactions play crucial roles in many fundamental cellular processes. Recently studies on RNA-protein interactions are getting considerable progress with the improvement of techniques and the appearance of new methods. Many protein-binding sites of RNAs have been characterized, some RNA-binding domains of proteins have been found, and the structural features of them have been investigated in detail. The results from above work can provide important information to understand the molecular mechanism of RNA-protein interactions and the related cellular processes.

Key words recognition, domain, RNA-binding protein

白血病抑制因子促进胚泡植入的研究进展*

蔡理全 段恩奎 祝 诚

(中国科学院动物研究所, 计划生育生殖生物学国家重点实验室, 北京 100080)

摘要 白血病抑制因子是一多效性细胞因子, 能促进哺乳动物的早期胚胎发育和启动胚泡植入, 其基因表达和生物合成受母体因素 (如甾体激素和其他细胞因子) 的控制. gp130 是白血病抑制因子家族受体的亲和力转化亚基, 其同源/异源亚基的二聚体能够激活酪氨酸激酶, 通过不同途径调节靶基因的表达.

关键词 白血病抑制因子, 胚泡植入, gp130

学科分类号 Q492.1, Q516

白血病抑制因子 (leukemia inhibitory factor, LIF) 是子宫内膜腺体分泌、分子质量为 45~56 ku 的糖蛋白^[1]. 它是一种多效性细胞因子, 可作用于多种靶细胞, 而且, 在体内许多组织/器官都有其分布.

哺乳动物的胚泡植入是一复杂的生理过程, 它要求有侵入能力的胚泡和具接受性的子宫内膜相互作用^[2], 即双方建立起“对话”机制. 一些细胞因子, 如 LIF、白介素 (interleukine, IL)、表皮生长因子 (epidermal growth factor, EGF)、转化生长因子 (transforming growth factor, TGF) 等在子宫内

膜的“植入窗口”期具有重要的生理功能.

本文试图就 LIF 在围植入期的基因表达、生理功能及其受体信号转导方面的近期进展加以综述, 为探讨 LIF 对胚泡植入的调节机理提供思路.

1 子宫内膜的 LIF 基因表达

Bhatt 等^[2]研究表明, LIF 在妊娠第 4 天小鼠的子宫内膜腺体中表达最高. 当延缓植入时, LIF

* 中国科学院留学经费择优支持项目 (96063).

收稿日期: 1997-10-27, 修回日期: 1998-02-19