

文章编号: 0529-6579 (1999) 01-0083-88

核仁小分子 RNA 基因的起源初探^{*}

张尚宏, 屈良鹄

(中山大学生物工程研究中心, 广州 510275)

摘 要: 通过分析、比较含 Box C/D 和含 Box H/ACA 这两类 snoRNA 的一些代表与几种剪接体 snRNA 的序列和二级结构特征, 研究它们的进化关系, 并探讨 snoRNA 是起源于内含子还是后来才插入到内含子中这一与 snoRNA 的起源密切相关的问题. 结合对剪接体 snRNA 起源的有关研究, 提出两类 snoRNA 最初是起源于原始 II 类内含子中的不同部分的观点.

关键词: snoRNA, 剪接体 snRNA, 内含子, 起源

分类号: Q 349.55 **文献标识码:** A

核仁小分子 RNA (snoRNA) 主要可以分为含 Box C/D、含 Box H/ACA 和 MRP RNA 等 3 类^[1,2], 含 Box C/D 和含 Box H/ACA 这两类有很多种, 其中不少是由核 mRNA 基因中的内含子编码的^[3]. 这一特殊的基因结构形式为研究基因的起源和进化提供了新的材料.

对于 snoRNA 基因的起源问题, 目前已有的研究还不多. 只有 MRP RNA 与 RNase P 中的 RNA 的进化关系可以比较肯定^[4], 而其他两类 snoRNA, 特别是由内含子编码的 snoRNA 的起源, 则是仍有待深入探讨的问题. 由于这两类 snoRNA 基因有不少是作为某些现代基因中内含子的组成部分, 因此要研究它们的起源, 关键是要解决 snoRNA 基因是起源于内含子 (原来就是内含子的结构成分), 然后有一部分演化成独立转录的基因; 还是起源于独立的基因序列, 后来才有一部分 snoRNA 基因插入到内含子中? 这两种起源方式从基因进化的可能性来看都是许可的, 从而对此一直就有两种与之相应的观点^[3,5]. 不过, 由于还没有找到足够的事实依据, 这两类观点只能作为对 snoRNA 基因起源的初步推测.

本文拟通过分析 snoRNA 与剪接体核小分子 RNA (spliceosomal snRNA) 的进化关系来探讨 snoRNA 的起源问题, 以求在这方面取得某些突破和进展.

1 剪接体 snRNA 起源研究的新进展

剪接体 snRNA 主要有 U1, U2, U4, U5, U6 等 5 种, 它们在核 mRNA 前体的剪接加工中起着重要的作用. 由于核 mRNA 前体的剪接与含有 II 类内含子的 RNA 剪接的生物化学途径十分相似, 因此就有认为这两类剪接以及核 mRNA 内含子与 II 类内含子都具有进化关

^{*} 基金项目: 国家自然科学基金重点项目 (39730300)

收稿日期: 1998-06-15 第一作者简介: 张尚宏, 男, 36 岁, 副教授

系的观点^[6], 进一步还推测 5 种剪接体 snRNA 是源于原始 II 类内含子的一些片段^[7]. 对 II 类内含子、各种剪接体 snRNA 以及相应的 RNA 剪接过程的深入研究确实是获得了不少支持这些观点的事实依据^[8-9]. 不过, 也有人认为这些相似性只是趋同进化的结果, 是不同起源的体系为了达成一些相似的生化反应而形成的类似解决途径^[10].

随着有关研究的深入, 近年来获得了一些能进一步说明核 mRNA 前体剪接与 II 类(内含子)剪接的进化关系的新结果. 首先, 在剪接反应的立体化学过程方面, 两者有着惊人的相同之处^[9-11]. 如果不认为它们是起源于同一原始体系, 则难以解释这种在反应细节上的相似性. 其次, II 类内含子的某些部分(结构域)在剪接中的作用可以由某种剪接体 snRNA 代替. 如 II 类内含子中的 ID3 域缺失后, 受到影响的剪接过程可由外源的 U5 复原^[12]. II 类内含子中的 ID3 域是固定剪接下来的 5' 外显子所必需的, 这与核 mRNA 前体剪接中 U5 的作用相当. 这种可替代性说明 II 类内含子与剪接体中的 RNA 有着结构上的相应关系. 此外, II 类内含子在纤细裸藻(*Euglena gracilis*)等叶绿体基因组中的发现^[13], 为研究原始的 II 类内含子如何进化为核 mRNA 内含子提供了更进一步的资料. 从结构上看, II 类内含子很明显是由原始 II 类内含子进化而来的, 它们的大小较小且均一(约 100 碱基对), 剪接过程与 II 类内含子相似, 但不能自剪接, 必须要有其他因子的参与^[9]. 从而, II 类内含子在更多的方面与核 mRNA 内含子相似, 可以认为它们是以原始的 II 类内含子作为共同祖先, 沿不同的途径趋同进化的结果^[9-13].

因此, 虽然核 mRNA 前体剪接与 II 类剪接、核 mRNA 内含子与 II 类内含子的特定进化关系还不能完全确定, 但从已有的资料来看, 这是最合理的解释.

2 snoRNA 与剪接体 snRNA 的关系

snoRNA 与剪接体 snRNA 在结构上的一些相似性以及功能上的对应性是明显的^[14]. 在结构上, 两者的大小差不多, 5' 端的帽子结构(对于独立转录的 RNA 来说)也十分相似(大部分是由聚合酶 II 转录). 与剪接体 snRNA 在核 mRNA 前体加工中的作用相应, 一些 snoRNA 在 rRNA 前体加工中也具有类似的功能. 那么, 这两类小分子 RNA 是否还有更本质的、进化上的关系呢? 由于目前对剪接体 snRNA 的起源已经有了比较深入的认识, 因此, 如果它们与 snoRNA 有一定的进化关系的话, 那对 snoRNA 起源的研究是很有启示的. 另一方面, 在两者存在着进化关系的情况下, 也会由于分歧年代的久远和功能的不尽相同, 不容易从核苷酸序列上揭示出它们之间的内在联系. 因此, 本研究主要从分析、比较这些小分子 RNA 的二级结构入手.

5 种剪接体 snRNA 的二级结构目前已经比较清楚, 如人类的 U1, U2 和 U5 的二级结构^[15, 16](图 1a). 在不同的生物中, 同一种 snRNA 的核苷酸序列可有较大的变化, 但都有一些区域十分保守, 二级结构也基本上不变^[13]. 至于 snoRNA 的二级结构, 则可以分为几种情况(不包括 MRP RNA): ① 独立转录且含 Box C/D 的 snoRNA; ② 由内含子编码且含 Box C/D 的 snoRNA; ③ 独立转录且含 Box H/ACA 的 snoRNA; ④ 由内含子编码且含 Box H/ACA 的 snoRNA. 其中①的二级结构比较多样化, 目前已确定结构的有 U3, U8, U13 等^[16]. ②则一般有 3'-5' 末端互补序列形成的“茎”结构, 其他区域的序列再形成另外一些茎-环结构. 目前已确定完整二级结构的有脊椎动物的 U14 等^[17]. ③和④的二级结构较为相似, 一般都是 5' 部分的茎环结构与 3' 部分的茎环结构由一较短的单链 RNA(绞链

区) 相连, 形成明显分为两部分的二级结构. 不过, ③比④一般要大, 二级结构也就复杂一些. 已有多种 Box H/ACA snoRNA 的二级结构基本上被确定^[2].

由于独立转录与内含子编码的 snoRNA 的产生(熟化)方式有所不同, 它们在二级结构上就会有一些不同的特点. 而 5 种剪接体 snRNA 都是由独立基因编码的, 因此在二级结构上较适合与独立转录的 snoRNA 比较. 另外, 为了避免物种间差异的影响, 有关的比较还应利用同种物种中的数据或多个物种的共同序列.

先比较独立转录且含 Box C/D 的 snoRNA 与剪接体 snRNA 的二级结构. 人类的 U3, U8 和 U13 snoRNA 的二级结构^[16]如图 1b. 比较 U8 和 U2, U2(图 1a, b)除了在 5'端的部分外, 其余区域与 U8 十分相似. 在相应茎环的序列也有不少相同的地方. 另外, Box C 在 U8 上的位置对应于 Sm 蛋白结合位点在 U2 上的位置, 而且两者的基序也相差不多. U13 与 U2 之间也有类似的相似性, 这很可能是由于 U8 和 U13 snoRNA 本身有较密切的进化关

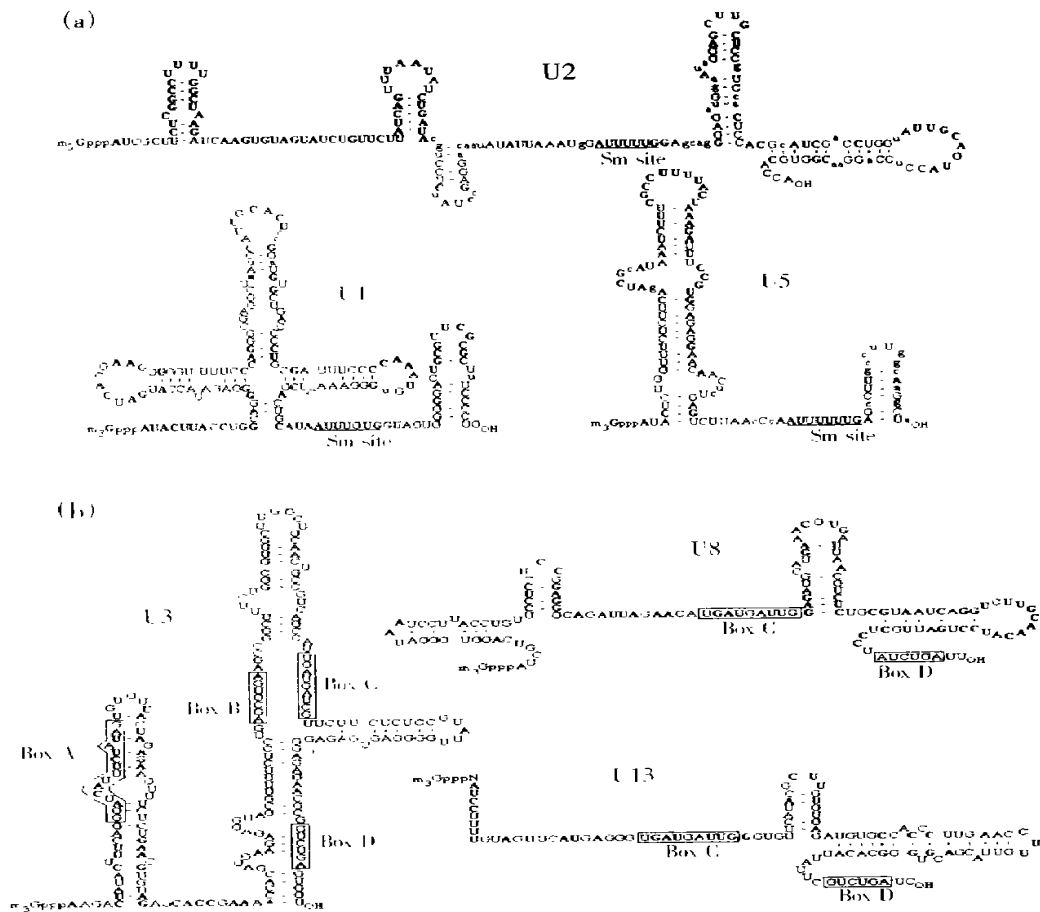


图 1 人类几种剪接体 snRNA 和 snoRNA 结构的比较

Fig 1 Comparison of the structure of several human spliceosomal snRNAs and snoRNAs

(a) U1, U2 和 U5 的二级结构. 大写字母表示该位置上的核苷酸在研究过的物种中有 1/2 以上是相同(黑体表示全部或绝大部分相同); 小写字母表示该位置上的核苷酸不保守, 小写黑体则表示在不同物种中基本上都是同类核苷酸

(b) U3, U8 和 U13 的二级结构. 黑体字母表示该核苷酸与有关剪接体 snRNA 相应位置上的核苷酸相

(C) 1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

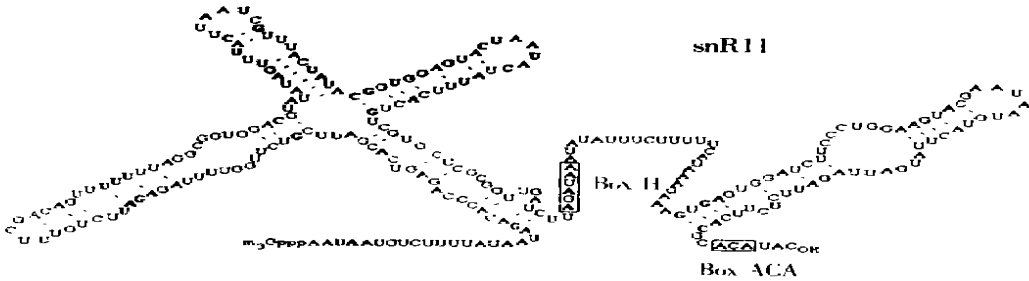


图 2 酵母 snR11 的二级结构

Fig 2 The secondary structure of yeast snR11

黑体字母表示该核苷酸与人类 U1 相应位置上的核苷酸相同

系. 至于 U3, 它与 U8 和 U13 在结构上差别较大. 从整体上看, U3 中含 Box C 和 D 的部分在 3'区域, 这一部分似乎与 U8 或 U13 有一定的同源关系, 而 U3 的 5'区域的二级结构则与 U5 这种剪接体 snRNA 5'部分的发夹结构较为相似.

含 Box H/ACA snoRNA 的二级结构与剪接体 snRNA 的也有一些相似之处. 这类 snoRNA 都有与 U5 或 U1 类似的特点, 即: 5'部分的发夹结构 (茎-环结构) 与 3'部分的发夹结构由一单链 RNA 的“绞链”相连^[2, 15]. 在 Box H/ACA snoRNA 的“绞链”中 (近其 5'端), 有 Box H (ANANNA 基序); 在 U5 和 U1 的“绞链”中, 则有 Sm 蛋白的结合位点, 而且在“绞链”的近 5'端处, 还可有类似 Box H 的基序 (往往在 Sm 蛋白结合位点的 5'上游区或包括部分 Sm 蛋白结合位点, 序列数据见文 [15]). 此外, 在酵母中还有一种独立转录的 snoRNA (snR11) 的二级结构^[1]与一般物种 U1 的^[15]十分相似 (图 1a, 图 2; 酵母 U1 的结构比较特殊^[15], 故只把 snR11 与可作为一般物种代表的人类 U1 比较).

综上所述, 独立转录且含 Box C/D 的 snoRNA 中的 U8 和 U13 与剪接体 snRNA U2 的关系较为密切, 而 U3 似乎与 U2 和 U5 有一定的关系. 至于含 Box H/ACA 的 snoRNA, 则是与 U5 和 U1 有关联的. 这种关系不但表现在二级结构上, 而且在部分区域的序列上也体现出来, 因此应该是一种进化上的关系. 另外, 由内含子编码且含 Box C/D 的 snoRNA 与剪接体 snRNA 的关系在二级结构分析中不明显, 但它们仍然可能有一些内在的联系.

3 snoRNA 的起源

在 snoRNA 与剪接体 snRNA 存在着进化关系以及同类的 snoRNA 都有共同祖先的前提下, 就不难对 snoRNA 的起源提出一些新的看法. 如前所述, 对于几种剪接体 snRNA 的起源, 越来越多的证据显示它们是源于原始 II 类内含子中的某些片段. 这样, 在起源上, snoRNA 很有可能也是来自原始 II 类内含子的片段, 特别是与 U2, U5, U1 相应的片段. 而且, 两类 snoRNA 的来源有所不同: 含 Box C/D 的源于与 U2 以及 U5 5'部分相应的片段; 含 Box H/ACA 的则是源于与 U5 和 U1 相应的片段. 在某些植物的同一个内含子中确是有同时存在含 Box C/D 和 Box H/ACA snoRNA 的情况 (待发表).

认为 snoRNA 直接起源于原始 II 类内含子的某些片段而不是源于已经从原始 II 类内含子演化而来的剪接体 snRNA 的祖先, 主要是因为仍然有很多现代的 snoRNA 基因存在于核 mRNA 内含子中. 按照这样一种模式, 这些由内含子编码的现代 snoRNA 理应与 II 类内含子中特定区域的片段的关系更为密切, 只是由于进化年代的久远、有关 RNA 分子 (片段)

的多变以及目前资料的不充分, 还未能找到一一对应的关系. 即使如此, 从现代 II 类内含子结构的复杂性及其各结构域中茎环的多样性^[9] 来看, 由原始 II 类内含子中的片段形成内含子中的 snoRNA 是完全有可能的.

在本文提出的 snoRNA 起源的模式中, snoRNA 是作为原始内含子的组成部分而起源的. 独立转录的 snoRNA 则是在其后的进化过程中才产生, 即由内含子中的 snoRNA 基因转变为独立的 snoRNA 基因, 这一过程无疑与内含子本身的进化有密切的关系. 这样, 在现代基因结构中, 核 mRNA 内含子编码 snoRNA 就是一种原始基因结构的遗迹. 核 mRNA 内含子的剪接显然要比其他内含子的自剪接进化, 但这并不意味着核 mRNA 内含子在各个方面都是进化的, 它们与外显子的连接区域就有可能是一种十分原始的进化遗迹^[18]. 因此, 一些核 mRNA 内含子具有编码 snoRNA 这一原始遗迹也就不足为奇了.

另一种观点认为内含子中的 snoRNA 是在进化后期才由独立转录的 snoRNA 插入到内含子而形成^[3, 9]. 因为同一种 snoRNA 基因在宿主基因内含子中的位置以及拷贝数并非恒定不变, 而是随着宿主基因或生物的不同会有变化; 而且, 在同一物种的两个同源基因中, 其中一个的内含子编码 snoRNA, 另一个的内含子却可不含 snoRNA; 此外, 有些 snoRNA 在某类生物中是独立编码, 在另外的生物中却是由内含子编码^[3]. 这些现象确实是可以由内含子和 snoRNA 的移动性来解释, 但这并没有限定 snoRNA 的最早起源, 从而也并不排除 snoRNA 最早是起源于内含子中的序列的可能性. 只要考虑 snoRNA 起源后的进化, 即它们与内含子一样, 都可能有在基因组中“移动”以及“扩增”或“丢失”的特性, 就可以解释现代基因组中各种 snoRNA 的存在状况.

4 结 语

snoRNA 在所有研究过的真核生物中都存在, 是 rRNA 的加工和修饰必不可少的^[3, 14], 从而它们在生命活动中起着十分重要的作用, 其起源也应可以追溯到生物进化的较早期. 从分子进化的共同祖先观点看, 很多现代差异较大的生物大分子(基因)都有可能来源于远古的一种生物大分子祖先. 本文已得出两类 snoRNA 如不同的剪接体 snRNA 那样, 是来源于原始 II 类内含子的不同片段的结论. 进一步, 剪接体 snRNA 和 snoRNA 的起源似乎还可以追溯到更远古的分子祖先, 因为原始 II 类内含子的不同片段之间还可能有着内在的或进化上的联系. 这首先可以从现代的剪接体 snRNA 看出: 除了 U6, 其余几种 snRNA 都含有与 Sm 蛋白结合的保守序列, 而且 U1 和 U5 的二级结构还有一定的相似性^[15]. 此外, Box C/D 和 Box H/ACA 这两类 snoRNA 也有一些很有意思的相似之处^[2]. 因此, 原始的 II 类内含子极有可能是由一些相似甚至相同的片段组成^[18], 继而才向多样化发展, 并从其中一些内含子产生各种剪接体 snRNA 及两类 snoRNA 的直接祖先.

参考文献:

- [1] BALAKIN A G, SMITH L, FOURNIER M J. The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions. *Cell*, 1996, 86: 823 ~ 834
- [2] GANOT P, CAIZERGUES-FERRER M, KISS T. The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation.

- Genes & Dev, 1997, 11: 941~956
- [3] MAXWELL E S, FOURNIER M J. The small nucleolar RNA. *Annu Rev Biochem*, 1995, 35: 897~934
- [4] MORRISSEY J P, TOLLERVEY D. Birth of the snoRNAs: the evolution of RNase MRP and the eukaryotic pre-rRNA-processing system. *Trends Biochem Sci*, 1995, 20: 78~82
- [5] SOLLNER-WEBB B. Novel intron-encoded small nucleolar RNAs. *Cell*, 1993, 75: 403~405
- [6] CECH T R. The generality of self-splicing RNA: relationship to nuclear mRNA splicing. *Cell*, 1986, 44: 207~210
- [7] SHARP P A. "Five easy pieces". *Science*, 1991, 254: 663
- [8] WISS J A. Guides to the heart of the spliceosome. *Science*, 1993, 262: 1978~1979
- [9] MICHEL F, FERAT J L. Structure and activities of group II introns. *Annu Rev Biochem*, 1995, 64: 435~461
- [10] WEINER A M. mRNA splicing and autocatalytic introns: distant cousins or the products of chemical determinism? *Cell*, 1993, 72: 161~164
- [11] PADGETT R A, PODAR M, BOULANGER S C, et al. The stereochemical course of group II intron self-splicing. *Science*, 1994, 266: 1685~1688
- [12] HETZER M, WURZER G, SCHWEYEN R J, et al. Trans-activation of group II intron splicing by nuclear U5 snRNA. *Nature*, 1997, 386: 417~420
- [13] COPRTYONO D W, HALLICK R B. Group II and group III introns of twintrons: potential relationships with nuclear pre-mRNA introns. *Trends Biochem Sci*, 1993, 18: 467~471
- [14] BACHELLERIE J P, MICHOT B, NICOLOSO M, et al. Antisense snoRNAs: a family of nucleolar RNAs with long complementarities to rRNA. *Trends Biochem Sci*, 1995, 20: 261~264
- [15] GUTHRIE G, PATTERSON B. Spliceosomal snRNAs. *Annu Rev Genet*, 1988, 22: 387~419
- [16] BASERGA S J, STEITZ J A. The diverse world of small ribonucleoproteins. In: GESTELAND R F, ATKINS J F, eds. *The RNA World*. Plainview, New York: Cold Spring Harbor Lab Press, 1993. 359~381
- [17] TERNS M P, GRIMM C, LUND E, et al. A common maturation pathway for small nucleolar RNAs. *EMBO J*, 1995, 14: 4860~4871
- [18] ZHANG S H. Origin and evolution of exon/intron junctions. *Speculat Sci Technol*, 1998, 21: 17~27

A Preliminary Study on the Origin of Small Nucleolar RNAs

ZHANG Shanghong^{*}, QU Lianghu

Abstract: The secondary structure and partial sequences of some box C/D and box H/ACA snoRNAs as well as several spliceosomal snRNAs were analyzed and compared, so as to reveal the evolutionary relationship between snoRNAs and spliceosomal snRNAs. For the origin of snoRNAs, it is discussed whether they first originated from introns or they inserted into introns after their origin. With the study on the origin of spliceosomal snRNAs, it is proposed that the two kinds of snoRNAs first originated respectively from different sections of primitive group II introns.

Keywords: snoRNA, spliceosomal snRNA, intron, origin

^{*} Biotechnology Research Center, Zhongshan University, Guangzhou 510275, China