

miRNA (microRNA) 家族的研究进展

叶 茂 陈跃磊 明镇寰*

(浙江大学生命科学院, 杭州 310029)

摘要 近年来, 科学家在许多真核生物中发现了一类能时序调控发育进程, 长度约为 21 个核苷酸的小分子 RNA, 并称其为 miRNA (microRNA). 最近研究表明, 它与早先在 RNAi (RNA interference) 中发现的 siRNA (small interfering RNA) 具有很大的相关性, 并在不同水平上参与了生物体内的遗传调控和基因重组等重要过程. miRNA 基因调控的特殊方式, 高度的保守性, 可能的加工机制以及它的发现所引发的讨论已引起人们的普遍关注.

关键词 miRNA, siRNA, 时序调控

学科分类号 Q522

1993 年, Lee 等^[1] 在秀丽新小杆线虫 (*Caenorhabditis elegans*) 中发现了第一个能时序调控胚胎后期发育的基因 *lin-4*. 时隔 7 年之后, Reinhart 等^[2, 3] 同样又在线虫 *C. elegans* 中发现了第二个异时性开关基因 *let-7*, 并将这类基因所编码的能时序调控发育进程, 长度约为 21 个核苷酸 (nt) 的小分子 RNA 称为 stRNA (small temporal RNA). 近一年来, 随着生物信息学的发展、分子克隆技术的改进和模式物种 cDNA 文库的建立, 美国 and 德国等科研人员又相继在线虫 (*C. elegans* 和 *C. briggsae*)、果蝇 (*Drosophila melanogaster*)、HeLa 细胞、斑马鱼 (*Danio rerio*)、拟南芥 (*Arabidopsis thaliana*) 和水稻 (*Oryza sativa*) 等许多真核模式生物和细胞中找到了上百个相类似的小分子 RNA, 并将其称为 miRNA (microRNA), 而基因 *lin-4* 和 *let-7* 所编码的 stRNA 则被认为是 miRNA 的代表^[4-9]. 尽管有很多方面仍不明了, 但目前还是普遍认为, miRNA 并不是 mRNA 和其他 RNA 简单的破碎产物, 而是很可能存在着一个庞大的 miRNA 家族, 在不同水平上发挥着发育控制、基因调控、RNA 切割和基因重组等方面的功能^[10-12]. 基于最新的研究成果, 本文对 miRNA 的特点, 可能的功能及其加工调控机制, *let-7* 基因的保守性和 miRNA 的研究前景作一简要综述.

1 miRNA 的特点

miRNA 有 3 个明显特征: a. 广泛存在于真核生物中, 是一组不编码蛋白质的短序列 RNA, 它本身不具有开放阅读框架 (ORF); b. 通常的长度

为 20~24 nt, 但在 3' 端可以有 1~2 个碱基的长度变化 (对 miRNA 的具体长度范围尚无统一标准, 在拟南芥和烟草中发现的 26 nt RNA, 和在四膜虫 (*Tetrahymena*) 中发现的能使大核部分 DNA 失活的 28 nt RNA 也被归于其中^[10]); c. 成熟的 miRNA, 5' 端有一磷酸基团, 3' 端为羟基, 这一特点使它与大多数寡核苷酸和功能 RNA 的降解片段区别开来^[4-7]. 除此之外, 多数 miRNA 还具有高度保守性、时序性和组织特异性.

miRNA 基因不是随机排列的, 其中有一些是成簇的 (cluster), 而且簇生排列的基因常常协同表达^[4-6]. 最典型的是一组高度相关的 miRNA 基因 (*mir35~mir41*) 集中簇生在 *C. elegans* 2 号染色体的 1 kb 片段上, 并从同一个前体上加工形成 7 个成熟的 miRNA^[6]. 这样的例子还有很多, 如 miR-6 的前体中簇生着 3 个仅有细微差异的重复序列^[5]. 关于簇生排列的原因尚未见深入研究.

细胞特异性和组织特异性是 miRNA 表达的主要特点. 由实验显示, 在组织培养的 S2 细胞 (即 Schneider-2 细胞, 从发育 20~24 h 的果蝇胚胎中提取得到) 可发现 miR-12 等. 却无法找到 miR-3~miR-6^[5]. 类似地, miR-171 (也有文献称之为 miR-39) 在拟南芥的花序和花组织中高水平表达, 而在茎、叶组织中却没有表达的迹象^[8]. 虽然被 miR-171 (miR-39) 调控的基因在其他组织中也都存在, 但仅在花序组织中大量表达, 这就使人怀疑

* 通讯联系人.

Tel: 0571-88273604, E-mail: zhming@mail.hz.zj.cn

收稿日期: 2002-11-20, 接受日期: 2002-12-28

miRNA 的表达和分布是否决定了组织和细胞的特异性. 而另一部分 miRNA 在不同细胞、组织和物种间的表达, 可能意味着它们在基因表达调控等方面具有更为广泛和普遍的作用.

目前, 大多数已发现的 miRNA 的表达都具有时序性. *mir3~ mir7* 基因只在果蝇早期胚胎形成时表达, 而 miR-1、miR-8 和 miR-12 的含量在果蝇幼虫阶段急剧上升并在成虫期维持在较高水平. 与此同时, 在所有阶段都存在的 miR-9 和 miR-11 的含量却急剧减少^[5]. 基因表达的严格或不严格时序性, 表达水平的显著变化, 以及所在组织和细胞的特异性, 都暗示着 miRNA 可能参与了深远而复杂的基因表达调控, 并决定发育和行为等的变化^[13].

2 miRNA 的功能及可能的加工机制

对 *lin-4* 和 *let-7* 的研究表明, miRNA 的主要功能很可能是进行转录后调控 (posttranscriptional regulation). 成熟的 miRNA 存在着与靶 mRNA 的 3' 端非翻译区 (UTR) 互补配对的位点, 两者识别结合, 使翻译无法进行, 从而抑制基因表达^[4~8]. 但也有研究表明, 有些 miRNA 能与 mRNA 中的开放阅读框 (ORF) 结合. 这种下调机制往往引起发育过程中某种特定蛋白质的数量急剧减少, 从而成为进入下一个发育阶段的标志^[11, 14]. 目前已知的 miRNA 的功能多与发育的时序调控有关, 但也不可能参与了空间发育、应激性、细胞周期和基因重组等过程^[13]. 虽然目前在一些真核模式生物中发现了 miRNA, 在原核生物中仍没有 miRNA 的相关报道, 但还是不能肯定 miRNA 的这种调控机制是真核生物所特有的. 相信随着研究的深入, 在这方面会有新的进展.

迄今的研究认为, miRNA 的形成需要 Dicer 酶的参与, Dicer 酶是一种多结构域的 RNase III 蛋白, 它具有两个纵列的 RNase III 结构域、一个 ATP 结合基序 (Dead 盒 RNA 解螺旋酶域, Dead-box RNA helicase domain), 以及一个 dsRNA (double-stranded RNA) 结合域. 其加工过程可能是: 首先在 Dicer 酶的参与下, 将一个约 70 nt、具有茎环结构的稳定的前体加工成为一个不稳定的双链 RNA 分子, 然后迅速降解成约 22 nt 的单链 RNA 分子, 之后再被 PPD (PAZ & Piwi domain) 蛋白质家族的成员识别结合, 形成一个 RNA-蛋白质复合体, 即 miRNA 核蛋白体 (miRNP), 进而形成成熟的

miRNA, 以反义 RNA 的形式与靶基因 3' 端 UTR 结合, 引起翻译阻遏 (图 1). 这种加工可以从前体两端的任意一端开始, 但就目前情况而言, 从前体 3' 端加工得到的 miRNA 要比 5' 端的多, 这也说明了前体的 3' 和 5' 位置并不是识别加工过程中的决定因素^[4~7]. 在整个加工过程中, Dicer 酶的识别和加工过程需要 ATP 供能, 而后来与 PPD 蛋白质家族的识别结合却不需要 ATP 的参与^[15]. 果蝇中发现的 Dicer 与线虫中的 Dcr-1, 人细胞中发现的 Helicase-MOI, 拟南芥中的 CARPEL FACTOR (CAF) 等同源物构成一个 Dicer 酶家族^[8, 13, 16, 17]. PPD 蛋白质家族的成员包括果蝇中的 Ago2、线虫 *C. elegans* 中的 Rde-1、拟南芥中的 Ago1、真菌中的 Qde-2 以及转录延伸因子 eIF2C2 等.

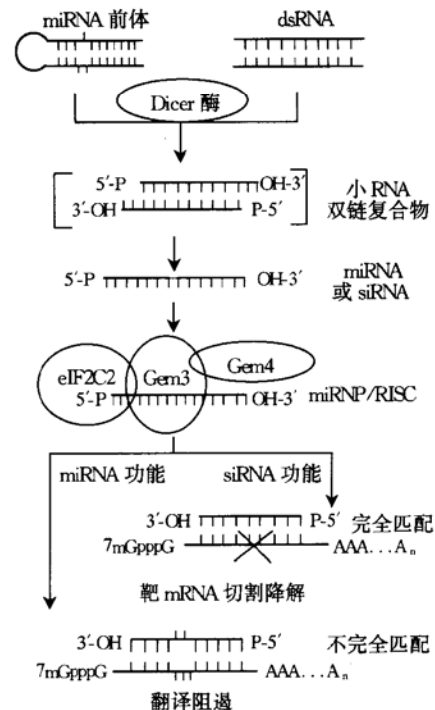


Fig. 1 Model for a common pathway in which miRNAs direct translational repression and siRNAs direct target RNA destruction (RNAi) ^[12]

图 1 miRNA 产生翻译阻遏和 siRNA 指导 RNAi 的共同途径的模式^[12]

小 RNA 双链复合物推测是短时存在的中介物.

上述加工机制与 siRNA (small interfering RNA) 的加工作用过程极其相似 (图 1). siRNA 也是一类长度为 21~25 nt 的双链 RNA, 同样在 Dicer 酶参与下, 由 dsRNA 加工而来. 它能与靶 mRNA 互补配对, 结合后导致 mRNA 降解, 使基因沉默, 这也就是 RNA 干涉 (RNA interference,

RNAi) 的机理. 早期研究认为 miRNA 与 siRNA 没有多大关联, 因为两者存在着明显的区别: miRNA 为单链, siRNA 为双链; miRNA 与 mRNA 的 3'-UTR 并不完全配对, 而 siRNA 却是完全互补的; 在正常生长的动物中一般不会有 RNAi 和 RNA 沉默的现象, 只有在外源性 RNA 导入的情况下才会发生. 但最新的研究表明这两者之间似乎具有很大的相关性, 如 Llave 等^[11] 在拟南芥中发现 miR-171 (miR-39), 它可与转录因子 *SCL* (*Scarecrow-like*) 家族 3 个成员的内部序列完全互补, 并能使它像与 siRNA 结合那样被降解. 而先前发现的 *let-7* 本身就参与形成一个 RNA 诱导沉默复合物 (RNA-induced silencing complex, RISC), 进入 RNAi 途径^[12]. 通过免疫沉淀法, 有人在哺乳动物细胞内发现了 miRNP, 它是一种 eIF2C2/GEMIN3/GEMIN4 复合物, 后两种蛋白质与运动神经元的存活 (survival of motor neurons, SMN) 有关^[18]. 关于 miRNP 的组分争议比较多, 因为采用不同方法在不同生物中提取得到的蛋白质复合物成分不一样. 最新研究认为 miRNP 就是 RISC. Hutvagner 和 Zamore 通过实验, 认为 miRNA 或 siRNA 功能的发挥取决于 miRNA 与靶结合位点匹配的程度, 而用于识别 miRNA 进入何种生化途径的 PPD 蛋白家族中的很多成员都是 RISC 的组分, 这使得 miRNA 与

siRNA 功能界限变得不清晰^[12, 16]. 因此从这个意义上说, 将 miRNA 和 siRNA 合为一个分子 RNA 家族就显得更为科学合理了.

3 miRNA 的保守性

许多 miRNA 的保守性非常显著^[3~8]. 在线虫 *C. elegans* 中所发现的 miRNA, 85% 都可以在 *C. briggsae* 的基因组中找到同源序列^[6], 而在拟南芥中发现, 16 个 miRNA 中的 8 个也可以在水稻基因组中找到完全一致的序列. 同样, miR-171 (miR-39) 的相似序列在拟南芥、水稻和烟草中可用杂交比对的方法检测到^[11]. 这种高度的保守性被认为与其功能的重要性有着密切的关系, 同时也为生物早期进化的同源性提供了某种证据.

最具有 miRNA 保守性的是 *let-7* RNA, 它广泛存在于两侧对称的生物体中 (图 2). 这种只有 21 nt 的 RNA 在脊椎动物、半索动物、软体动物、环节动物以及节肢动物等的生物中都可找到, 却不存在于某些刺胞动物和海绵动物中. *let-7* 序列的保守性的确令人吃惊, 线虫 *C. elegans* 中的 *let-7* miRNA 可以在果蝇基因组, 人的第 9、11 和 22 条染色体上各找到一个与之完全一致序列, 另外在人的第 9 条染色体上还有一个仅差 1 个核苷酸的序列^[3].



Fig. 2 Phylogenetic comparison of *let-7* RNA expression in animals^[3]

图 2 动物中 *let-7* RNA 表达情况的系统发生比较^[3]

“+”表示在该物种体内存在 *let-7* RNA 的表达, “-”表示在该物种体内无法探测到 *let-7* RNA, “Dev.”表示该物种发育早期没有 *let-7* RNA, 但成体后出现 *let-7* RNA 的表达.

为了更好地体现许多物种中 *let-7* 基因的高度保守性, 并反映物种间的亲缘关系, 我们查询了美国国家生物信息中心 (NCBI) 数据库 (<http://www.ncbi.nlm.nih.gov>) 和不久前建立的非编码 RNA (Non-coding RNA, ncRNA) 数据库 (<http://biobases.ibch.poznan.pl/ncRNA/micro.html>), 采用生物信息学的方法, 用 clustalw (1.4 版) 和 treeview (win32) (1.0.0.0 版) 软件对 2002 年 10 月之前发现的 21 个 *let-7* 基因及其同源序列进行多序列比对, 并做出相应的系统进化树图 (图 3)。

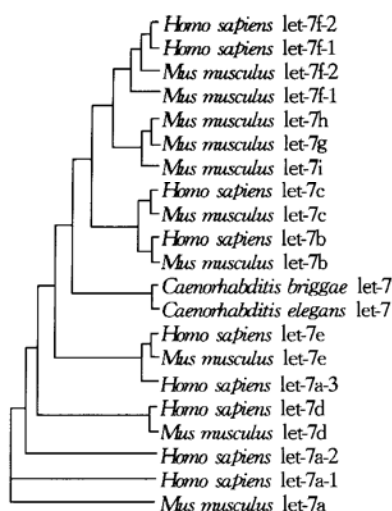


Fig. 3 An evolutionary and phylogenetic tree of *let-7* gene and its homologs in *Homo sapiens*, *Mus musculus*, *Caenorhabditis elegans* and *Caenorhabditis brigglae*
图 3 在人、小鼠、线虫 *C. elegans* 和 *C. brigglae* 中发现的 *let-7* 基因及其同源序列的系统进化树

从 *let-7* 基因的高度保守性我们可以猜测它出现在进化的早期。虽然在有些生物中未发现它的踪影, 但也可能是所测试的基因遗失或分化, 不能排除这些生物的共同先祖拥有 *let-7* 的可能性^[3]。如果 *let-7* 确是共同祖先早期就拥有的基因, 那么与它协同作用的 RNA-蛋白质的进化必然也已长期存在。这样就极有可能解释 *let-7* 的高度保守, 因为基因和其相互作用的 RNA-蛋白质复合物这两者中任何一方的改变都要另一方付出巨大的代价, 这就可能是它与目标基因协同进化的结果。

4 关于 miRNA 的讨论和展望

4.1 miRNA 为什么选择这样的长度

有人认为, 21 nt 可能是已知的 mRNA 表达过程中所需的最小热力学稳定长度, 而且将 70 nt 左

右的前体加工为 21 nt 左右的 miRNA, 其速度要比从 mRNA 转录到蛋白质翻译调控的全过程快得多, 因此采用这种方式就能更加迅速地调控基因表达^[4]。我们也认为大约 21 nt 的排列组合可以满足它所调控的基因数量, 而不会造成浪费, 并且转录的长度越短, 转录时的出错率也就越小, 这对调控过程可能更为有效。

4.2 miRNA 的数量

尽管迄今为止已发现了 100 多个 miRNA, 但可以肯定的是这个数目远未达到饱和, 因为许多脊椎动物的 miRNA 仅被发现了一次^[13], 这也说明 miRNA 的低水平表达。现在有很多 miRNA 是在混合期细胞中发现的, 如果能仔细筛选出细胞类型或特定时期, 就有可能发现比较罕见的 miRNA。长期以来人们也为高等和低等生物之间复杂性的基因根源困惑不已。人基因组中的编码基因约有 3.5 万个, 除去重组可能引起的功能复杂, 大量的非编码 DNA 的作用仍待确定, 也许它们在物种差别上起到了重要的作用。Mattick^[19]声称内含子和非编码 DNA 占据了人类基因组转录的 95% 以上, 而一个由 Gingeras 领导的小组^[20]更是在人的 21, 22 染色体上, 发现高于预期 10 倍以上的 DNA 序列被转录, 这些都使我们对新 miRNA 的发现充满希望。

4.3 miRNA 和 siRNA 的进化关系

siRNA 和 miRNA 的界限已越来越不明显了, 两者加工和表达机制上的相似性, 使人们推测它们都是早期 RNA 在进化上的遗留物, 并逐步被蛋白质调控所代替。现在争论较多的是这两种机制的从属关系和进化顺序。以前曾有人认为 siRNA 是 miRNA 的补充, 因为 miRNA 参与正常情况下生长发育的调控, 而 siRNA 不参与动物体的正常生长, 只有在外源性 dsRNA 诱导下才产生。到底哪个出现得更早些, 是 RNA 的彻底降解还是表达的抑制? 我们倾向于在早期的生命进化中, 类似于 siRNA 的调控机制出现得更早, 并且它可能是早期基因调控的主要方式。因为外源基因的入侵在早期的生物进化中就出现了, 采取这种方式就能比较有效地破坏外源基因并且防止转座子转移。但是这种基因的抑制方式过于彻底, 导致调节的不可逆, 而且费力费时, 所以在进化过程中可能逐步产生了 miRNA 来代替它, 从而使 miRNA 成为发挥调控作用的主导因子。

鉴于最近一年有关 miRNA 的研究获得了突破性

的进展,《Science》杂志将它和先前发现的 siRNA 所组成的小分子 RNA 评为 2002 年的年度突破. miRNA 对中心法则中 RNA 次要的中介角色的重要补充,将促使生物学家重新思考细胞遗传调控及其发育等方面的问题.我们相信,随着研究的深入,miRNA 将在生命起源和早期进化、基因复杂性、疾病原理等方面起到更为深远的作用.

致谢 miRNA 的发现者之一,美国麻省理工学院 Bartel 教授、Lau 硕士和美国麻省医科大学 Zamore 教授为本文的完成提供了相关咨询和资料,浙江大学生命科学院傅承新教授对本文的完成也给予悉心指导,在此一并表示衷心感谢!

参 考 文 献

- 1 Lee R C, Feinbaum R L, Ambros V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*, 1993, **75** (5): 843~854
- 2 Reinhart B J, Slack F J, Basson M, *et al.* The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature*, 2000, **403** (6772): 901~906
- 3 Pasquinelli A E, Reinhart B J, Slack F, *et al.* Conservation of the sequence and temporal expression of *let-7* heterochronic regulatory RNA. *Nature*, 2000, **408** (6808): 86~89
- 4 Ruvkun G. Molecular biology: glimpses of a tiny RNA world. *Science*, 2001, **294** (5543): 797~799
- 5 Lagos-Quintana M, Rauhut R, Lendeckel W, *et al.* Identification of novel genes coding for small expressed RNAs. *Science*, 2001, **294** (5543): 853~858
- 6 Lau N C, Lim L P, Weinstein E G, *et al.* An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science*, 2001, **294** (5543): 858~862
- 7 Lee R C, Ambros V. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science*, 2001, **294** (5543): 862~864
- 8 Reinhart B J, Weinstein E G, Rhoades M W, *et al.* MicroRNAs in plants. *Genes & Dev*, 2002, **16** (13): 1616~1626
- 9 Llave C, Kasschau K D, Rector M A, *et al.* Endogenous and silencing-associated small RNAs in plants. *Plant Cell*, 2002, **14** (7): 1605~1619
- 10 Baulcombe D. DNA events: an RNA microcosm. *Science*, 2002, **297** (5589): 2002~2003
- 11 Llave C, Xie Z X, Kasschau K D, *et al.* Cleavage of *Scarecrow-like* mRNA targets directed by a class of *Arabidopsis* miRNA. *Science*, 2002, **297** (5589): 2053~2056
- 12 Hutvagner G, Zamore P D. A microRNA in a multiple-turnover RNAi enzyme complex. *Science*, 2002, **297** (5589): 2056~2060
- 13 McManus M T, Sharp P A. Gene silencing in mammals by small interfering RNAs. *Nature Reviews Genetics*, 2002, **3** (10): 737~747
- 14 Dennis C. Gene regulation: The brave new world of RNA. *Nature*, 2002, **418** (6894): 122~124
- 15 Holen T, Amarzguioui M, Wigger M T. Positional effects of short interfering RNAs targeting the human coagulation trigger tissue factor. *Nucleic Acid Research*, 2002, **30** (8): 1757~1766
- 16 Ishizuka A, Siomi M C, Siomi H. A *Drosophila* fragile X protein interacts with components of RNAi and ribosomal proteins. *Genes & Dev*, 2002, **16** (19): 2497~2508
- 17 Hutvagner G, McLachlan J, Pasquinelli A E, *et al.* A cellular function for the RNA-interference enzyme dicer in the maturation of the *let-7* small temporal RNA. *Science*, 2001, **293** (5531): 834~838
- 18 Mourelatos Z, Dostie J, Paushkin S, *et al.* miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes & Dev*, 2002, **16** (6): 720~728
- 19 Mattick J S, Gagen M J. The evolution of controlled multitasked gene networks: The role of introns and other noncoding RNAs in the development of complex organisms. *Mol Biol Evol*, 2001, **18** (9): 1611~1630
- 20 Kapranov P, Cawley S E, Drenkow J, *et al.* Large-scale transcriptional activity in chromosomes 21 and 22. *Science*, 2002, **296** (5569): 916~919

Progress in The Research of miRNAs (microRNAs) Family

YE Mao, CHEN Yue-Lei, MING Zhen-Huan*

(College of Life Science, Zhejiang University, Hangzhou 310029, China)

Abstract Recently, a class of ~ 21 nucleotides (nt) small RNA have been discovered in many eukaryotes, termed miRNAs (microRNAs), which were first identified as key temporal regulators in development. So far, large quantities of studies have revealed that miRNAs have played important roles in genetic control at many different levels and rearrangement of genome. Besides, its association with siRNA (small interfering RNA) previously discovered in RNAi (RNA interference) in the further researches becomes much closer than it has ever been considered. Its pathway directing translational repression, the surprisingly high conservation of certain miRNA, the mechanism of process of mature miRNA and genetic regulation compared with that of siRNA were focused. Finally several discussions arising people's interests caused by the discovery of miRNA are made.

Key words miRNA, siRNA, temporal regulation

* Corresponding author. Tel: 86-571-88273604, E-mail: zhming@mail.hz.zj.cn

Received: November 20, 2002 Accepted: December 28, 2002