Cell, Vol. 107, 823–826, December 28, 2001, Copyright ©2001 by Cell Press

microRNAs: Tiny Regulators with Great Potential

Minireview

Victor Ambros[1]
Department of Genetics
Dartmouth Medical School
Hanover, New Hampshire 03755

Animal genomes contain an abundance of small genes that produce regulatory RNAs of about 22 nucleotides in length. These microRNAs are diverse in sequence and expression patterns, and are evolutionarily widespread, suggesting that they may participate in a wide range of genetic regulatory pathways.

Cells contain a variety of noncoding RNAs, including components of the machinery of gene expression, such as tRNAs and rRNAs, and regulatory RNAs that influence the expression of other genes (see Eddy, 2001 for a recent review, as well as references therein). Lately, it has become increasingly apparent that noncoding RNAs are impressively diverse, and that a significant fraction of the genes of all organisms do not encode proteins. One class of small noncoding RNAs—the microRNAs (or miRNAs)—has recently been recognized to be quite numerous and phylogenetically extensive (Lee and Ambros, 2001; Lau et al., 2001; Lagos-Quintana et al., 2001). microRNA genes produce tiny transcripts of about 22 nucleotides (~22 nt) in length that probably function as antisense regulators of other RNAs. As expected for a newly described large gene family, there are at the moment far more hypotheses and unanswered questions than there are concrete facts.

Genetic Identification of microRNAs in Caenorhabditis elegans: The lin-4 and let-7 Gene Products

The best-studied miRNA genes are *C. elegans lin-4* and *let-7*, which were identified from the genetic analysis of developmental timing in the nematode (see Rougvie, 2001 for a review, and references therein). *let-7* or *lin-4* mutant worms fail to execute certain developmental switches, resulting in the abnormal repetition of certain larval stages. The cloning of first *lin-4* (Lee et al., 1993) and later *let-7* (Reinhart et al., 2000) revealed these two genes to be particularly deviant—unusually small, encoding no protein products, and producing exceedingly short (~22 nt) transcripts from characteristic hairpin RNA precursors (Figure 1). The 22 nt *lin-4* and 21 nt *let-7* RNAs are antisense translational repressors of mRNAs that encode proteins of the heterochronic developmental timing pathway of the worm (Lee et al. 1993; Wightman et al., 1993; Moss et al., 1997; Slack et al., 2000). For example, *lin-4* RNA is complementary to sequences in the 3′ untranslated region (UTR) of *lin-14* and *lin-28* mRNAs. The synthesis of LIN-14 and LIN-28 proteins is repressed by *lin-4* during the early larval stages of *C. elegans* development to cause the proper sequence of stage-specific developmental events.

When first described, *lin-4* and *let-7* seemed to be unique, since no similar tiny regulatory RNA had been encountered in other organisms. Were the *lin-4* and *let-7* RNAs evolutionary freaks, peculiar to worm developmental timing, with little relevance to the world of biology at large? Two remarkable discoveries helped to dispel these doubts. The first of these was that *let-7* RNA is phylogenetically conserved—in size and nucleotide sequence—in essentially all the bilaterally symmetric animals (Pasquinelli et al., 2000). Moreover, *let-7* has a similar developmental profile in diverse taxa, suggesting the conservation of an ancient developmental timing pathway. Indeed, homologs of the worm *let-7* target, *lin-41*, can be found in insects and vertebrates with their *let-7* complementary sites intact. These findings indicated that the *lin-4* and *let-7* class of regulatory genes was not just a worm oddity, and likely represents a gene family that has evolved from an ancient ancestral small RNA gene.

A second discovery that supported the relative ubiquity of tiny RNAs like *lin-4* and *let-7* was the finding that small antisense RNAs of about 22 nt in length (called siRNAs) are central to RNA interference (RNAi) (see Sharp, 2001 and references therein). RNAi is an evolutionarily conserved genetic surveillance mechanism that can degrade an mRNA in response to the presence of double-stranded RNA corresponding to the targeted mRNA. *lin-4* and *let-7* are not siRNAs (they do not trigger degradation of their targets), but the ubiquity of siRNAs suggested that small RNAs of the size of *lin-4* and *let-7* have been part of the eukaryotic milieu for a very long time. Indeed, the *lin-4* and *let-7* ~22 nt RNAs are processed from their stem-loop precursor transcripts by the same enzyme, Dicer, that generates the ~22 nt siRNAs from a dsRNA trigger (Grishok et al., 2001; Hutvagner et al., 2001; Bernstein et al., 2001; Ketting et al., 2001). Since Dicer is phylogenetically widespread, genes like *lin-4* and *let-7* could also be commonplace.

Stealth Genes No Longer

If other tiny RNA genes like *lin-4* and *let-7* lurk undiscovered in animal genomes, how would they be unmasked? Because these genes would be noncoding, traditional computational gene finding methods tuned to protein coding potential would miss them, and they would not be represented in conventional cDNA libraries prepared from polyadenylated mRNA. However, they would be expected to produce hairpin precursor transcripts of about 70 nt, and mature RNAs of about 22 nt. Therefore, genes like *lin-4* and *let-7* should emerge from structure-based informatics searches of noncoding genomic sequences, and from cDNA cloning efforts tailored to very small transcripts.

Recently, methods were developed to clone cDNAs corresponding to the ~22 nt siRNAs produced during RNAi (Elbashir et al. 2001a). These methods were adapted to the preparation of cDNAs corresponding to size-selected (~22 nt) RNAs expressed in worms, flies, and human cells (Lee and Ambros, 2001; Lau et al., 2001; Lagos-Quintana et al., 2001). From these cDNA libraries, dozens of sequences corresponding to novel transcripts of about 22 nt were identified, and dubbed "microRNAs." The genes that produce these transcripts are located

[1]Correspondence: vra@dartmouth.edu

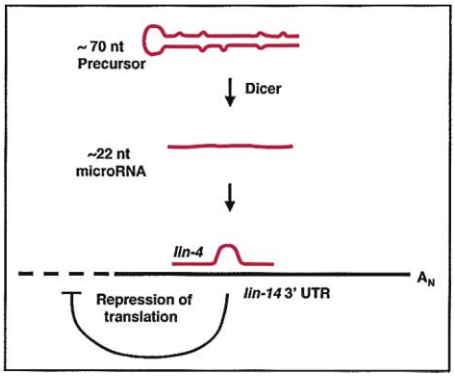

Figure 1. *lin-4* Base Pairs with *lin-14*

A canonical microRNA, *lin-4*, is processed from its precursor transcript by Dicer, and base pairs with a complementary sequence in the 3′UTR of a target, in this case *lin-14* mRNA, thereby triggering a *cis*-acting translational repressive activity.

within exons of, or between, protein coding genes. Longer (~70 nt) precursors were identified for many of these novel RNAs, and the precursors are predicted to form a hairpin reminiscent of the *lin-4* and *let-7* precursors.

The characteristic hairpin structure of miRNA precursors was used as the basis for an informatics screen conducted in parallel to the cDNA cloning. Noncoding genomic sequences that are conserved between *C. elegans* and a related nematode, *C. briggsae*, were tested computationally for predicted secondary structure (Lee and Ambros, 2001). Although this approach was less efficient than the cDNA cloning (owing in part to the limited amount of available *C. briggsae* sequence), some of the miRNAs that were identified by informatics were not represented in the cDNA sequences. This indicates that computational approaches are a valuable complement to cDNA cloning in efforts to exhaustively catalog the miRNA genes in sequenced genomes.

A significant fraction of the microRNA genes seems to be very well conserved phylogenetically. Of the 62 *C. elegans* microRNA genes described so far (including *lin-4* and *let-7*), nine are conserved in *Drosophila*, and seven are conserved in *Homo sapiens*. For these evolutionarily related miRNAs, the sequence of the ~22 nt mature miRNA shows the greatest conservation. Such tight conservation in the miRNA sequence presumably reflects complementarity to multiple conserved target sequences. In some cases, the miRNA and its antisense targets could be involved in similar pathways across diverse evolutionary distances, as seems to be the case for the temporal regulator *let-7*. However, until we learn more about the biology of miRNAs, particularly what range of molecular interactions they engage in, it is difficult to predict all the factors that could influence the evolutionary conservation or divergence of their sequences.

How many different microRNAs are there? It is early still, but indications are that miRNA genes will prove to be quite numerous. The 90-odd different miRNA genes described so far in worms, flies, and humans were from screens that were not pushed to saturation, so it is likely that many more will be identified in these and other organisms. As more organisms are analyzed, the tally of unique miRNAs will likely be extensive, and miRNA genes could prove to be one of the most phylogenetically numerous and diverse classes of noncoding RNA genes.

What Do miRNAs Do?

The new fly, worm, and human microRNAs exhibit varied expression patterns, from uniform during development, to relatively stage-specific and/or tissue-specific. This suggests a variety of roles for miRNA genes, including the regulation of developmental timing, spatial patterning of cell fates, or cellular and organismal physiology. An accurate picture of miRNA function will require detailed knowledge of their spatial patterns of expression, and the identification of loss-of-function and gain-of-function mutations. Currently only two microRNAs, *lin-4* and *let-7*, have been analyzed genetically and assigned to pathways and biological functions. It is curious that other miRNA genes were not discovered previously through the common routine of cloning genes by mutation. These little genes may present a relatively small target size for mutagenesis, and some of them may function redundantly with other miRNAs of similar sequence, which would cause them to be under-represented among genes identified by screens for visible phenotypes. Nevertheless, one suspects that at least some of the microRNA genes are among the uncloned genetic loci previously identified by mutation. Unaware of the abundance of miRNAs, investigators cloning genes by phenotype (including human disease genes) could have easily overlooked these tiny noncoding sequences. Armed with a new awareness, we can now be more attentive to miRNA candidate genes.

By what molecular mechanisms do miRNAs act? Do they all regulate gene expression? Are they all translational repressors, like *lin-4* and *let-7*? It is probably wise to keep an open mind to the potential diversity of regulatory activities for miRNAs. Based on the precedent of *lin-4* and *let-7*, it is reasonable to expect that other microRNAs could also base pair with particular mRNA 3′UTR regulatory elements and repress translation. In the case of *lin-4*, its target mRNAs (*lin-14* and *lin-28*) are translationally repressed yet remain associated with polyribosomes (Olsen and Ambros, 1999; Seggerson et al., 2002), suggesting a block at a step after translational initiation. However, other miRNAs might not work the same as *lin-4*; and some could even regulate mRNA localization or stability, rather than translation. In principle, a microRNA could have either a positive or negative regulatory impact on gene expression, depending on how miRNA base pairing influences the structure and composition of the mRNA ribonucleoprotein (RNP). miRNAs could bind to protein factors and thereby recruit them to specific mRNAs, or they could act as "gatekeepers" that alter mRNA secondary structure, and thereby indirectly control the binding of other regulatory factors (Figure 2). A recent study suggests that *lin-4* can work this way. Specifically, *lin-4* seems to repress *lin-28* indirectly, by enabling a separate and independent repressive activity (yet to be characterized molecularly) to access the *lin-28* 3′UTR (Seggerson et al., 2002).

More broadly speaking, we should perhaps recognize that miRNA targets may not be restricted just to messenger RNAs. Some miRNAs could base pair with and regu-

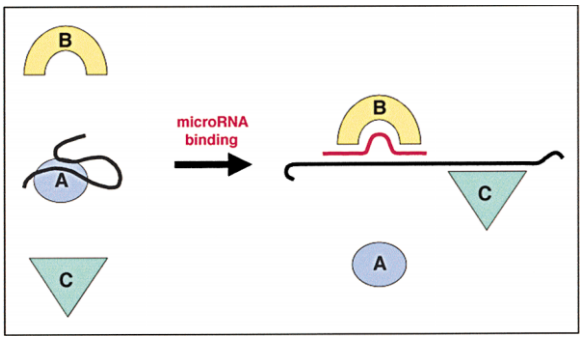

Figure 2. Regulation of RNP Structure and Function by a Generic microRNA

Base-pairing of the miRNA (red) to a target RNA (black) alters the RNA secondary structure, displacing proteins (A), and/or directly recruiting other proteins (B), and/or exposing binding sites for other proteins (C).

late noncoding RNAs, perhaps even other miRNAs. A miRNA might even function without base pairing to nucleic acids targets at all, for example by competing with another RNA for binding to protein. Perhaps we can think of miRNAs as free-range RNP structural elements, and in that sense they could potentially regulate in *trans* any biological process involving RNA::RNA or RNA::protein interactions (Figure 2).

The identification of the regulatory targets of each miRNA will be a challenge, but will be critical for placing them in genetic pathways and biological contexts. This task will probably require a combination of informatics, biochemical, and genetic approaches. One problem at the outset is that we do not yet know all the parameters defining a functional miRNA::target interaction, so the informatics approaches should be designed to cast a broad net, followed by validation using biochemical evidence of in vivo interactions, and genetic epistasis in support of functional interactions.

Our current thinking about how miRNAs interact with their mRNA targets is shaped by *lin-4* and *let-7*. For both *lin-4* and *let-7*, the base-pairing between miRNA and target is discontinuous, consisting of two short helices. This topology is probably critical to the specific outcome of the interaction, which in the case of *lin-4* and *let-7* is translational repression without mRNA destabilization. A continuous ~22 nt double-stranded helix seems to be the key to triggering RNAi and consequent target degradation (Elbashir et al., 2001b), and so miRNAs such as *lin-4* and *let-7* may form interrupted helices with their targets in order to bypass RNAi and engage in other modes of regulation. It remains to be seen whether any of the miRNAs actually do act as siRNAs—by forming a continuous 22 nt dsRNA hybrid with a target and thus triggering its degradation.

Why are the miRNA so tiny, and what predictions about their biology can be made based on their small size? The approximately 22 nt length is probably dictated by the enzymology of Dicer (Elbashir et al., 2001a). Although the unprocessed (~70 nt) form of a miRNA could be functional in some cases, currently we are assuming that most miRNAs function as 21 or 22 nt molecules. There are more than 10^{13} different theoretical 22 nt sequences, and so in terms of biological specificity through Watson/Crick base-pairing, each miRNA has enormous potential. Aside from the base-pairing specificity of a miRNA, how much additional biological information can be contained in ~22 nucleotides? This relatively short length seems insufficient to include complex structural elements signaling stability, intracellular transport, etc., in addition to antisense sequence elements. Therefore, we may find that many different miRNAs utilize a common protein (or RNP) complex for intracellular trafficking and protection from degradation.

How Are miRNAs Regulated?

lin-4 and *let-7* expression is developmentally regulated; *lin-4* RNA accumulates at the second larval stage, and *let-7* at the fourth stage. Several of the new miRNAs also display temporally regulated expression, indicating that miRNA genes can respond to a variety of developmental signals. Although some of the miRNAs that are contained in introns might be generated as a byproduct of pre-mRNA splicing, most miRNAs seem to come from intergenic regions and are probably transcribed from autonomous miRNA promoters, as is the case for *lin-4* and *let-7*. Key questions in this regard include: What RNA polymerase(s) transcribe these genes, how are their regulatory sequences structured, and what transcription factors couple their expression to developmental signals?

Some constitutively transcribed miRNAs could be regulated on the level of processing of the mature ~22 nt transcript from its precursor. Among the miRNAs identified in *C. elegans*, some appear to be processed rather inefficiently, leaving the ~70 nt precursor as the major species detected (Lee and Ambros, 2001). These may represent a class of miRNA that is processed only under specific conditions, which would imply that Dicer, and/or other components of miRNA processing be regulated.

Most of the miRNAs occur in isolation along the genome, but a significant fraction of them occur in clusters containing multiple miRNAs that are apparently expressed coordinately as a long precursor, and processed into individual miRNAs (Lagos-Quintana et al., 2001; Lau et al., 2001). Clustered miRNAs are sometimes similar in sequence, but can also be distinct. Some miRNA clusters could represent a multipronged attack on a common target, or the coordinate control of a set of independent targets.

Outlook for miRNAs

The history of miRNA research, beginning with the novelty of *lin-4* almost 10 years ago, and including this recent abundant harvest of new miRNAs, has been marked by surprises at almost every turn. Currently, there is an embarrassment of miRNA riches, accompanied by a corresponding measure of ignorance. The miRNA genes promise a wealth of regulatory molecules whose functions are still almost completely unknown. The questions about microRNAs discussed here, and other questions beyond the scope of this review, are just beginning to be addressed. The answers to these questions are likely to be complex, given the number and diversity of different miRNAs. *lin-4* and *let-7* provide welcome role models for how other miRNAs might work, but even for these emissaries, our knowledge of mechanism is still rather rudimentary. Moreover, one suspects that the diversity and abundance of miRNA genes re-

flects a broad spectrum of functions and mechanisms, requiring that we approach the study of them with a mindset open to surprise and delight.

Selected Reading

Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Nature *409*, 363–366.

Eddy, S.R. (2001). Nat. Rev. Genet. *2*, 919–921.

Elbashir, S.M., Lendeckel, W., and Tuschl, T. (2001a). Genes Dev. *15*, 188–200.

Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W., and Tuschl, T. (2001b). EMBO J. *23*, 6877–6888.

Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A., Ruvkun, G., and Mello, C.C. (2001). Cell *106*, 23–34.

Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T., and Zamore, P.D. (2001). Science *293*, 834–838.

Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J., and Plasterk, R.H. (2001). Genes Dev. *15*, 2654–2659.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Science *294*, 853–858.

Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). Science *294*, 858–862

Lee, R.C., and Ambros, V. (2001). Science *294*, 862–864.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). Cell *75*, 843–854.

Moss, E.G., Lee, R.C., and Ambros, V. (1997). Cell *88*, 637–646.

Olsen, P.H., and Ambros, V. (1999). Dev. Biol. *216*, 671–680.

Pasquinelli, A.E., Reinhart, B.J., Slack, F.J., Martindale, M.Q., Kuroda, M.I., Maller, B., Hayward, D.C., Ball, E.E., Degnan, B., Muller, P., et al. (2000). Nature *408*, 86–89.

Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., and Ruvkun, G. (2000). Nature *403*, 901–906.

Rougvie, A.E. (2001). Nat. Rev. Genet. *2*, 690–701.

Sharp, P.A. (2001). Genes Dev. *15*, 485–490.

Seggerson, K., Tang, L., and Moss, E.G. (2002). Dev. Biol., in press.

Slack, F.J., Basson, M., Liu, Z., Ambros, V., Horvitz, H.R., and Ruvkun, G. (2000). Mol. Cell *5*, 659–669.

Wightman, B., Ha, I., and Ruvkun, G. (1993). Cell *75*, 855–862.