Molecular Cell, Vol. 16, 861–865, December 22, 2004, Copyright ©2004 by Cell Press

Transcription and Processing of Human microRNA Precursors

Review

Bryan R. Cullen*
Howard Hughes Medical Institute and
Department of Molecular Genetics and Microbiology
Duke University Medical Center
Durham, North Carolina 27710

MicroRNAs have recently emerged as key posttranscriptional regulators of eukaryotic gene expression, yet our understanding of how microRNA expression is itself controlled has remained rudimentary. This review describes recent insights into the mechanisms governing microRNA transcription and processing in vertebrates and their implications for understanding the regulation of microRNA biogenesis.

First discovered in the nematode *C. elegans*, the ~22 nt long class of noncoding RNAs termed microRNAs (miRNAs) are now thought to be expressed in all metazoan eukaryotes (reviewed by Bartel [2004]). Over 200 different miRNAs have been identified in rodents, and almost all of these are also found in humans. Moreover, a number of human miRNAs are conserved in *C. elegans*, which is especially striking as this organism appears to encode somewhat fewer miRNAs (116 are currently known).

Whereas the function of the large majority of miRNAs remains unclear, specific miRNAs are known to play important roles in the regulation of apoptosis and cell proliferation in fruit flies, neuronal asymmetry in *C. elegans*, leaf and flower development in plants, and hematopoietic differentiation in humans (Bartel, 2004). Perhaps the best characterized miRNAs are lin-4 and let-7, which regulate the timing of larval development in *C. elegans* by downregulating the expression of specific target mRNAs after binding to complementary sequences present in the mRNA 3′ untranslated region (3′ UTR). The fact that most, and possibly all, vertebrate miRNAs are expressed in a developmentally regulated or tissue-specific manner (Lagos-Quintana et al., 2002) suggests that they also control aspects of human development. While the identification of the mRNA targets for specific miRNAs and a full description of their mechanism(s) of action therefore remain important issues, the focus of this review is on the regulation of miRNA expression and, more specifically, on the transcription and processing of miRNA precursors. Throughout, I will focus on what is known about these processes in vertebrate animals, although relevant observations obtained in invertebrates and plants will be discussed where appropriate. It is important to note, however, that certain aspects of the miRNA expression pathway may differ significantly between vertebrates on the one hand, and invertebrates and plants (in particular) on the other.

pri-miRNA Transcription and Structure

Animal miRNAs are initially transcribed as part of one arm of an ~80 nucleotide (nt) RNA stem-loop that in turn forms part of a several hundred nt long miRNA precursor termed a primary miRNA (pri-miRNA) (Lee et al., 2002). Whereas most human miRNAs are genomically isolated, several are found in miRNA clusters that are transcribed and expressed coordinately. Analyses of the pri-miRNA precursors of 15 human miRNAs, five isolated and the others in clusters, have shown that all derive from pri-miRNA precursors that bear a 5′ 7-methyl guanosine cap and a 3′ poly-A tail (Cai et al., 2004; Lee et al., 2004a).

Five full-length pri-miRNA precursors have now been cloned and characterized, three from humans (Figure 1A), one from *C. elegans*, and one from plants. All are capped, polyadenylated RNAs that appear to be noncoding (Aukerman and Sakai, 2003; Bracht et al., 2004; Cai et al., 2004; Lee et al., 2004a). The pri-miRNA precursor for the clustered human miRNAs miR-23a~miR-27a~miR-24-2 is an unspliced ~2.2 kilobase (kb) RNA, and the ~3.4 kb pri-miRNA for the isolated human miRNA miR-21 is also unspliced. In contrast, the pri-miRNA for human miR-155 contains two introns and is polyadenylated at two alternate poly-A sites to give spliced pri-miRNA precursors of ~0.6 and ~1.4 kb (Figure 1A).

The DNA sequences that flank the 5′ end of pri-miR-23a~miR-27a~miR-24-2 and pri-miR-21 are able to express artificially linked protein-coding genes in transfected human cells (Cai et al., 2004; Lee et al., 2004a). Moreover, in *C. elegans*, fusion of the predicted promoter elements for the lsy-6 and let-7 miRNAs to the green fluorescent protein (*GFP*) indicator gene gives a pattern of GFP expression in transgenic animals that accurately reiterates the miRNA expression pattern (Johnson et al., 2003; Johnston and Hobert, 2003). Together these data argue that most, and probably all, animal pri-miRNAs are transcribed by RNA polymerase II and indicate that pri-miRNAs are structurally analogous to mRNAs. This result is consistent with not only the regulated expression pattern of most miRNAs but also the large size of many pri-miRNAs.

A recent computational analysis of 161 vertebrate miRNAs (Rodriguez et al., 2004) concluded that 30 were located in an exon of a noncoding RNA, 27 in an intron in a noncoding RNA, and 90 in an intron in a protein coding mRNA. 14 miRNAs were found in both exonic and intronic locations, depending on the alternative splicing pattern of the flanking gene. The finding that many miRNAs are located within the introns of protein coding mRNAs (Figure 1A) suggests that these intronic miRNAs may be processed out of intron lariats, as previously also reported for some small nucleolar RNAs (snoRNAs) (Weinstein and Steitz, 1999), and that their expression is therefore coordinately regulated with the expression of the flanking mRNA. Indeed, a limited analysis of 5 intronic miRNAs showed a very strong correlation between the tissue-specific expression pattern of the flanking mRNA and the embedded miRNA in four cases (Rodriguez et al., 2004). However, as several intronic miRNAs are present in the antisense orientation, relative to the flanking mRNA, it remains possible that at least

*Correspondence: culle002@mc.duke.edu

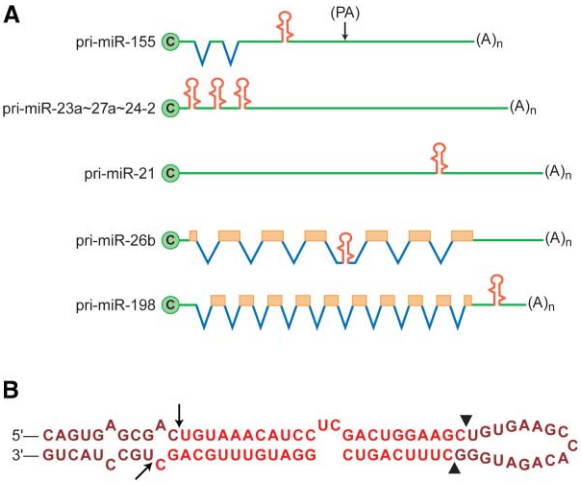

Figure 1. Structure of pri-miRNAs

(A) Schematic of the structure of five human pri-miRNAs. miRNA stems are shown in red, noncoding sequences in green, introns in blue and coding exons in beige. This figure is not to scale. PA, alternate poly-A-site.

(B) Structure of the human pri-miR-30a RNA hairpin. Drosha cleavage sites are shown by arrows and Dicer cleavage sites by triangles.

some intronic miRNAs are also transcribed by their own promoter.

Finally, I note that a very small number of human miRNAs are found within the 3′ UTR of coding mRNAs; e.g., miR-198 in the 3′ UTR of the mRNA for human follistatin related protein (Figure 1A). While the processing of this pri-miRNA would be predicted to inhibit expression of the linked open reading frame by removing the mRNA poly(A) tail, it remains unclear how efficiently individual pri-miRNA precursors are processed. Moreover, it is possible that expression of the linked open reading frame may actually be mediated, in at least some cases, by an alternatively spliced transcript bearing a different 3′ UTR (Rodriguez et al., 2004). Finally, recent data suggest that 3′ truncated pri-miRNA precursors may still be translated at some level of efficiency (Cai et al., 2004).

Nuclear Processing of pri-miRNAs by Drosha

RNase III enzymes are a family of double-stranded RNA (dsRNA) specific ribonucleases that are thought to be expressed in all living cells (Conrad and Rauhut, 2002). A characteristic of RNase III enzymes is that dsRNA cleavage introduces a 2 nt 3′ overhang at the cleavage site. In bacteria, RNase III is involved in rRNA maturation and mRNA degradation. In yeast, the RNase III enzyme Rnt1p participates in the nuclear processing of rRNAs, small nuclear RNAs, and snoRNAs and also plays a role in the degradation of unspliced pre-mRNAs and intron lariats (Danin-Kreiselman et al., 2003). Interestingly, this requires the recognition by Rnt1p of a conserved class of RNA tetraloop followed by cleavage 12 and 14 nt away in the flanking stem. Thus, Rnt1p appears to function as a "helical RNA ruler" (Chanfreau et al., 2000).

Human cells express three members of the RNase III enzyme family. One of these is restricted to mitochondria and is believed to function in mitochondrial rRNA processing. The other two RNase III enzymes, termed Drosha and Dicer, play essential but distinct roles in miRNA processing (Bartel, 2004). While Drosha is also thought to participate in the nuclear processing of human rRNAs (Wu et al., 2000), the cytoplasmic enzyme Dicer may function exclusively in the generation of miRNAs and the similar small interfering RNAs (siRNAs) that mediate RNA interference (RNAi). In the case of Dicer, and probably also Drosha, processing again involves binding to a specific site on an RNA stem or stem-loop followed by cleavage of the stem at a set distance from that site.

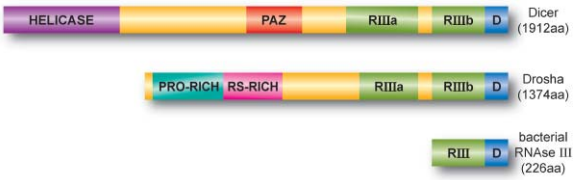

Figure 2. Domain Organization of Human Drosha and Dicer

RIII, RNase III catalytic domain; D, dsRNA binding domain; PRO-RICH, proline-rich domain; RS-RICH, arginine/serine rich domain.

Sequence analysis of bacterial RNase III identified two functional domains; i.e., a carboxy-terminal dsRNA binding domain and an amino-terminal catalytic domain (Figure 2) (Conrad and Rauhut, 2002). Bacterial RNase III enzymes form dimers, and this is critical for the function of the catalytic domain. Whereas these domains are conserved in the larger human enzymes, both Drosha and Dicer actually contain two RNase III catalytic domains (Figure 2) that have recently been shown to form an intramolecular dimer (Han et al., 2004; Zhang et al., 2004).

As noted above, miRNAs form part of one stem of an ~80 nt stem-loop that in turn forms part of the longer pri-miRNA. This stem-loop is cleaved by Drosha to liberate a shorter, ~60 nt hairpin, termed a pre-miRNA, that contains a 2 nt 3′ overhang (Figure 1B) (Lee et al., 2003). Importantly, one end of the pre-miRNA is coincident with one end of the mature miRNA (note that the miRNA may form part of either arm of the pre-miRNA).

Inspection of the RNA stem-loops that are predicted to flank miRNAs reveals no sequence similarity. However, these stem-loops are of comparable size and are all predicted to form simple but imperfect hairpins. A typical example, shown in Figure 1B, encodes the human miRNAs miR-30a-5p and miR-30a-3p (this precursor is unusual in that it gives rise to two mature miRNAs). Moreover, the RNA stem loop shown, together with short, ~6 nt single-stranded flanking sequences, represents the minimal sequence required for in vivo (Zeng and Cullen, 2003) and in vitro (Lee et al., 2003) processing by Drosha. Of note, this minimal sequence includes not only the 63 nt pre-miR-30a sequence, consisting of an ~22 bp stem and 15 nt terminal loop, but also a flanking ~9 bp stem. Other pre-miRNAs require up to 40 nt of flanking sequence for appropriate excision by Drosha. Mutational analysis of the extended miR-30a stem-loop has shown that disruption of the stem, either within or outside the pre-miR-30a sequence, blocks Drosha processing (Lee et al., 2003; Zeng and Cullen,