ABSTRACT

MicroRNAs (miRNAs) are generated from stem-loop-structured double-stranded RNA precursors by the consecutive action of the two RNase III-type endoribonucleases Drosha and Dicer. However, such structures are very common on cellular transcripts and specific features have evolved that guide and regulate processing of stem-loop-structured hairpins into mature and functional miRNAs. These features include sequence motifs and local RNA structures but also trans-acting factors such as RNA binding proteins. The menu of features required for miRNA biogenesis is summarized in this review.

INTRODUCTION

Small RNA-guided post-transcriptional gene silencing pathways are found in all eukaryotes and play important roles ranging from regulating endogenous physiological processes to genome stability and cellular defense strategies [1]. One particular pathway is guided by microRNAs (miRNAs) that repress specific target genes through RNA-RNA pairing [2]. MiRNAs are found in all eukaryotes except of the model yeast Saccharomyces cerevisiae, which specifically lost this pathway [3]. A hallmark of miRNAs is their generation from double-stranded (ds) RNA precursors through the action of specialized ribonucleases of the RNase III family [4].

miRNAs are single stranded and typically 18-24 nucleotides (nt) long. They associate with members of the Argonaute protein family and guide them to complementary target RNAs [5]. In mammals, such target sites are often located in the 3' untranslated region (3' UTR) of mRNAs and miRNAs are only partially complementary to their target sites [6]. Most importantly, nts 2-7 are fully complementary and serve as the ‘seed sequence’ important for target site recognition and binding [2]. The remaining part of the miRNA contributes to pairing with the target but is not essential. The bound Argonaute protein then induces deadenylation, decapping and mRNA decay through the interaction with additional proteins mainly members of the GW protein family (refer to as TNRC6 proteins in mammals) [7-10]. Some Argonaute proteins are endoribonucleases themselves and in case the bound miRNA is fully complementary to the target (rare in animals but very common in plants), the target RNA is endonucleolytically cleaved [11,12].

The targeting rules of miRNAs highlight the importance of the correct sequence of the miRNA. Since miRNAs are processed from dsRNA precursors, accurate and highly regulated processing is essential for proper targeting of the correct target sequences. For example, shifting the processing position within the precursor miRNA by only one nucleotide, the seed sequence would be different and an entire new spectrum of RNAs might be targeted while binding to regular targets might be lost. To achieve accuracy, miRNA precursors have evolved specific sequence and structural features that position the processing enzymes at the correct place. Our current knowledge about these features are summarized in this review.

PRIMARY miRNA FEATURES GUIDING PROCESSING BY THE MICROPROCESSOR COMPLEX

MiRNAs are transcribed from specific miRNA genes, which can be mono- or poly-cistronic (referred to as miRNA clusters) [2]. The majority of miRNAs are transcribed by RNA polymerase II, which leads to capping and polyadenylation since these processes are essential for transcription elongation and termination [13]. Thus, a nuclear miRNA precursor, immediately after transcription (also known as primary miRNA transcript or pri-miRNA), is characterized by
a stem-loop structure formed by a 70–90 nt long sequence containing an approximately 10 nt loop (Fig. 1A). In mammals, the stem is approximately 35 nt long and flanked by a single stranded 5′ extension carrying the cap and a single stranded 3′ part that ends with the poly(A) tail [14]. These extensions are removed by a first processing step by a protein complex composed of two molecules of DiGeorge critical region 8 (DGCR8) and one molecule of the RNase III-enzyme Drosha known as the microprocessor complex [14-18].

Such hairpins may form at many positions of the transcriptome but primary miRNAs are nevertheless highly specifically recognized. This is due to the nature of dsRNA primary miRNA molecules. They evolved a number of features that guide the microprocessor.

The 5′ flanking region ends with a short UG motif that is recognized by Drosha [19]. The stem is followed by a short highly complementary dsRNA to form the lower basal stem (Fig. 1A). The stem is followed by a bulged GHG motif, which is also important for Drosha cleavage. Drosha contains two RNase III domains and cleaves the stem on both strands (Fig. 1A, indicated as red arrow heads). The positioning of Drosha leaves two nt 3′ overhangs, which is a prerequisite for further downstream processing [20]. A rather important feature is found right above the two cleavage sites. The double stranded nature of this region determines cleavage efficiency and therefore mismatches or even weak interactions are often depleted from this particular region [21-23]. The upper end of the stem is also highly complementary and rarely mismatched. The stem is followed by a single stranded loop region of about 10 nt in length that contains a GUGU sequence motif that recruits and positions DGCR8 [24]. DGRC8 itself forms a dimer and helps to correctly position Drosha on the primary miRNA [25]. Downstream of the stem on the single stranded 3′ extension, a conserved sequence motif is found [19]. This CNNC motif serves as binding site for the SR protein SRSF3, which acts broadly and binds to most canonical primary miRNAs [22].

Not all primary miRNAs contain all features mentioned above. Nevertheless, the more of these sequence elements and dsRNA features are present, the better the hairpin is cleaved. Large scale mapping of cleavage events, however, also found various atypical cleavage events including 5′ and 3′ nicks of the double stranded stem or even inverse processing, i.e. binding of the microprocessor to the primary transcript in flipped orientation [26]. Such atypical processing might occur when specific features are missing. A recent study revealed a Drosha dsRNA recognition site (DRES) on various stem-loop-structured RNAs, which lack other features (Fig. 1A). DRES can guide the microprocessor to such sites leading to the generation of non-canonical miRNA processing [27].

The evolution of the features highlighted above demonstrates that the generation of mature canonical miRNAs is tightly regulated and pervasive processing of hairpin structures that spontaneously form on many transcripts, is prevented. In vitro processing and mapping experiments suggest that approximately 1000 human primary miRNAs are efficiently processed [26]. Some non-canonical processing is likely tolerated to allow for plasticity of miRNA-guided gene silencing during changing environmental conditions and the evolution of novel miRNA-guided regulatory circuits.

Many miRNAs are organized in clusters and transcribed together as long primary transcript [28]. Interestingly, several miRNAs are rather efficiently processed although their double stranded precursors are sub-optimal. This is particularly true for miR-451, which forms a rather short hairpin and is independent of Dicer processing [29-31]. It was found that neighboring optimal miRNA hairpins can support microprocessor cleavage in a process termed ‘cluster assisted miRNA processing’ [32,33]. This phenomenon is observed for a number of primary miRNAs including viral miRNAs [34,35]. Mechanistically, the two proteins SAFB2 and ERH interact with the microprocessor and bridge two clustered miRNAs [36,37]. This leads to efficient processing of a sub-optimal substrate within a miRNA cluster.

Taken together, these few examples highlight the importance of the double-stranded nature of the primary miRNA with its specific sequence features and genomic organization, which are prerequisites and determinants of efficient miRNA generation in many different organisms.

Figure 1. Structure and sequence features of double stranded miRNA precursors. (A) Primary miRNA transcripts (pri-miRNAs) contain various structural features as well as contact points for the microprocessor (Drosha, DGCR8) and the RNA binding protein SRSF3. In the absence of most indicated features a Drosha double-stranded RNA recognition site (DRES) can recruit the microprocessor for processing. Drosha cleavage sites are indicated by red arrowheads. (B) The microprocessor product miRNA precursors (pre-miRNA) serves as substrate for Dicer. Dicer requires the 2 nts 3′ overhang produced by Drosha but does not further rely on specific sequence or structure motifs. A GYM motif, however, enhances Dicer processing. Dicer cleavage sites are indicated with red arrowheads.
DOUBLE-STRANDED miRNA PRECURSORS IN THE CYTOPLASM: PROCESSING BY DICER

Cleavage by the microprocessor generates a specific basal end of the miRNA precursor (also referred to as pre-miRNA) characterized by a 5′ phosphate and a two nt 3′ overhang (Fig. 1B). This end is specifically recognized by the export receptor Exportin5 [38-40] and both features are required for the second cleavage step facilitated by the cytoplasmic RNase III enzyme Dicer [41,42]. Dicer is a highly conserved multi-domain enzyme. Similarly to Drosha, it contains two RNase III domains that cleave the double-stranded stem of the pre-miRNA hairpin. Therefore, while Drosha defines one end of the mature miRNA, Dicer produces the second one. Dicer anchors the 5′ and the 3′ ends of the hairpin in specific pockets located within specific domains referred to as the PAZ and platform domains [43]. In subsequent steps, the hairpin is loaded into the catalytic site and both strands are cleaved leaving a short double stranded RNA that contains the mature miRNA and the opposing strand initially referred to as the miRNA star strand. It is now commonly accepted that the two sequences are termed miRNA-5p and miRNA-3p reflecting the position of the mature miRNA on the stem of the pre-miRNA. Both strands can become the mature miRNA (Dicer mechanisms and structural features are reviewed in [43,44]).

While the recognition of primary transcripts by the microprocessor requires a number of different RNA structure and sequence features (see above), Dicer processes Drosha products without further selection. MiRNA definition and selection is therefore carried out by Drosha and channeling into the pathway does not require further selection and recognition at later stages. Nevertheless, a recent study identified a short sequence motif located at the apical cleavage site of the 3′ arm of the pre-miRNA stem. Based on the sequence context, it was termed GYM (a paired G, a paired pyrimidine and a mismatched C or A) motif (Fig. 1B) [45,46]. This motif directly interacts with Dicer and increases cleavage efficiency at this particular position.

Besides cleavage of the pre-miRNA hairpin at the two positions (indicated with red arrow heads in Fig. 1B), Dicer carries out a second very important function: it initiates strand selection and the transfer of the mature strand to an Argonaute protein, a process commonly referred to as RISC or Argonaute loading [47-50]. The two strands are not randomly selected and loaded. Instead, the strand with the less stably paired 5′ end becomes the mature miRNA and Dicer may contribute to this asymmetric strand selection process [51,52]. Furthermore, Dicer interacts with a dsRNA binding protein partner (e.g. Loqs or R2D2 in flies and TRBP in human). Particularly in Drosophila, a role during asymmetry sensing has been assigned to R2D2 [53]. The detailed molecular mechanisms however, still remain rather enigmatic.

REGULATION OF THE DOUBLE STRANDED miRNA PRECURSOR BY RBPs

Besides SRSF3, which binds the 3′ flanking region of most primary miRNAs, RBPs that sequence-specifically bind and regulate miRNA processing of individual primary miRNAs, have been reported (summarized in [54]). Large-scale biochemical as well and cellular screening approaches discovered many RBPs with affinity to primary miRNAs [54,55]. A well-characterized member of this group of RBPs is LIN28 [56-60]. In human cells, LIN28 expresses two protein variants – LIN28A and LIN28B. LIN28A is cytoplasmic and interacts with two motifs located at the apical loop and stem within miRNA precursors of the let-7 miRNA family. LIN28A is expressed in stem cells and recruits the terminal uridylytransferases TUT4 and TUT7, which oligo-uridylate the double stranded pre-let-7 miRNAs at their 3′ ends [61-63]. This modification prevents Dicer recruitment and instead attracts the 3′ to 5′ exoribonuclease DIS3L2, which rapidly degrades the precursor. LIN28B appears to be expressed in a broader panel of cell types and localizes to the nucleus, where it binds to primary let-7 transcripts. However, the proposed inhibitory mechanism appears to be different. LIN28B sequesters the dsRNA substrate into nucleoli and thus prevents export into the cytoplasm [64].

Some pre-miRNAs receive only a 1 nt 3′ overhang from the nuclear microprocessor due to structural constraints. Such pre-miRNAs are only poor Dicer substrates. They are nevertheless efficiently processed because TUT2, TUT4 and TUT7 can bind to these pre-miRNAs and add a single U to their 3′ ends generating a 2 nt 3′ overhang that can be efficiently processed by Dicer [65]. Moreover, such an addition can also lead to a miRNA arm switch and directly affects miRNA strand selection. This has been reported for pre-miR-324 and a role in glioblastoma pathology has been suggested [66].

RNA polymerase III (pol III) generates non-coding RNAs of medium length including pre-tRNAs, the 5S rRNA and many other functional RNAs [67]. All these ncRNAs are highly structured which is important for their specific functions. However, not all of these RNAs can fold spontaneously. Instead, they require help of RNA chaperones. The RBPs and Lupus autoantigen La is one of the first described RNA chaperones that associates with the 3′ end of pol III transcripts and helps folding them into the correct structure [68,69]. However, some pre-tRNAs can adopt alternative structures in the absence of La leading to extended hairpin structures. Interestingly, these alternative structures can be channeled into the miRNA pathway and are exported and efficiently processed by Dicer [70]. This highlights the importance of the correct structure not only of miRNA precursors but non-coding RNAs in general.

PERSPECTIVES

Not all mature miRNAs follow the path through microprocessor and Dicer maturation. A large number of non-canonical miRNA hairpins have been reported including microprocessor- or Dicer-independent miRNAs [71]. Prominent examples are mirtrons, which are intron-miRNA chimeras that are liberated by the splicing machinery and directly channeled into Dicer without prior Drosha cleavage [72]. Or, miR-451 has a very short double stranded stem and cannot be processed by Dicer. Instead it is cleaved by a catalytic member of the Argonaute protein family [30,31].
In addition, short structured RNAs can also form during transcriptional initiation and some of these RNAs are transported to the cytoplasm where they are processed by Dicer to mature miRNAs [73,74]. In addition, other sources of ds-RNAs can be processed by the miRNA processing machinery including small nucleolar RNAs (snoRNAs) or even tRNAs [75-77]. Nevertheless, many of these miRNAs do not reach high cellular concentrations and thus it remains unclear whether processing of such RNAs is a consequence of processing plasticity and tolerated noise or whether small RNAs have indeed evolved from such bi-functional dsRNA precursors. In some cases, both might be true and active evolution towards novel functionalities could be observed.

Highlighted are just some of the reported aspects how the double-stranded character of an RNA can affect downstream processing and fueling them into distinct regulatory pathways. RBPs but also RNA modifications [78] can directly change the biophysical properties of an RNA-RNA helix suggesting many more and so far unrecognized ways of influencing double-stranded RNA processing.

REFERENCES
ables the Processing of Suboptimal Stem-Loop Structures in Clustered Primary miRNA Transcripts. Mol Cell 78: 876-889 e876