Regulation of microRNA biogenesis and function

Thomas Treiber; Nora Treiber; Gunter Meister
Biochemistry Center Regensburg (BZR), Laboratory for RNA Biology, University of Regensburg, Regensburg, Germany

Summary

MicroRNAs (miRNAs) are considered as key regulators of literally all cellular functions. Therefore, miRNA biosynthesis and their individual cellular functions must be tightly regulated as well. MiRNAs are transcribed as primary transcripts, which are processed to mature miRNAs in two consecutive maturation steps. Finally, the mature miRNA is incorporated into a miRNA-protein complex, where it directly interacts with a member of the Argonaute (Ago) protein family. The miRNA guides such protein complexes to partial complementary target sites, which are typically located in the 3’ untranslated region (UTR) of mRNAs leading to inhibition of gene expression. MiRNA activity and abundance is regulated on various levels ranging from transcription and processing to target site binding and miRNA stability. Recent advances in our understanding of how miRNA activity is regulated in mammalian cells are summarised and discussed in this review article.

Keywords

microRNAs, Dicer, Drosha, Argonaute, gene silencing

Introduction

MicroRNAs (miRNAs) are short non-coding RNAs that are viewed as fundamental regulators of cell function (1, 2). MiRNAs are approximately 20–25 nucleotides (nt) long and are generated from double stranded RNA precursors (see below). One strand is selected and interacts with a member of the Ago protein family to form a miRNA-induced silencing complex (miRISC), also referred to as micro-ribonucleoprotein (miRNP). The miRNA guides such complexes to partially complementary target RNAs. Most of the target sites are located within the 3’ untranslated region (UTR) of mRNAs. However, functional miRNA binding sites in the 5’ UTR as well as in the open reading frame have also been reported (3, 4). Nucleotides 2–8 of the miRNA are particularly important for pairing with the target mRNA. This sequence motif is referred to as the miRNA seed sequence (1). Depending on the recognition site, binding of miRISC to the cognate target can have different outcomes (5). In case the target site is perfectly complementary to the miRNA, the miRNA functions like short interfering RNAs (siRNAs) and the target is sequence-specifically cleaved by miRISC. This is very rare in mammals but more frequent in plants. Binding to partially complementary target sites is the rule in mammals and leads to repression of translation or degradation of the target transcript (6). This degradation process, however, is different to RNAi-like mechanisms and involves the recruitment of deadenylase complexes such as the CCR4-NOT complex to the mRNA to remove or shorten the poly(A) tail. Poly(A) tail shortening induces the removal of the 5’ cap of the mRNA, a process referred to as de-capping. Uncapped mRNAs are rapidly removed from the cell by 5’ to 3’ exoribonucleases such as Xrn1 (7). Ribosome profiling experiments in conjunction with mRNA level measurements revealed that mRNA decay accounts for approximately 85%, while translational repression contributes only about 15% to miRNA-guided gene silencing effects in mammals (8).

MiRNA-guided gene silencing is mediated by Ago proteins, which are the direct binding partners of mature miRNAs and therefore central components of miRISC. In mammals, four different Ago proteins exist (Ago1-Ago4) and they are characterised by PAZ (PIWI-Argonaute-Zwille), MID and PIWI (P-element-induced wimpy testes) domains. The PAZ domain anchors the 3’ end of the miRNA, while the MID domain contains a highly specific binding pocket for the 5’ end of the miRNA (9). In addition, the MID domain discriminates between the four different bases at the 5’ terminal nucleotide and prefers binding to uridines and, with a slightly lower affinity, to adenosines explaining the U bias at the 5’ end of miRNAs (10). The PIWI domain is structurally similar to RNase H and indeed some PIWI domains possess endonuclease activity and mediate target RNA cleavage in RNAi (11, 12). In mammals, only Ago2 is endonucleolytically active (13, 14).

MiRNA-guided regulation of gene expression has been implicated in literally every cellular pathway. In addition, each cell type expresses a specific subset of miRNAs to ensure that cell type specific miRNA profiles are established and maintained. For example, expression of a neuron-specific miRNA in non-neuronal cells resulted in shifting the global gene expression program towards transcript profiles typically found in neurons (15). Another prominent example for the importance of miRNAs is the effect of a number of miRNAs on the reprogramming of somatic cells to induced pluripotent stem cells (iPS) (16).
miRNA biogenesis

MiRNAs are transcribed as parts of longer primary transcripts (pri-miRNAs) that are generated mainly by RNA polymerase II (22, 23). MiRNA genes are frequently located in transcripts of protein-coding genes, where they mainly reside in introns (24, 25). Other miRNAs are transcribed as part of long non-coding RNAs and are often arranged as clusters in the genome, which leads to one pri-miRNA being subsequently processed into several functional miRNAs (24). Like other Pol II transcripts, pri-miRNAs possess a 5’ cap and a 3’ poly-A-tail (23). Within the primary transcripts, miRNAs form stem-loop structures, which contain the mature miRNA as part of an imperfectly paired double stranded stem connected by a short terminal loop. In the canonical miRNA biogenesis pathway, these structures are recognised by the microprocessor complex, a multiprotein complex with the two core components Drosha and Di George Syndrome critical region gene 8 (DGCR8) (26–29). The double-stranded RNA binding protein DGCR8 binds to the base of the stem loop structure and thereby guides the positioning of the RNase III enzyme Drosha, which constitutes the catalytic center of the complex (30). Drosha cleaves the double-stranded stem about 11 bp from the base and generates a two nucleotide (nt) overhang at the 3’ end (27, 30). This cleavage reaction liberates a hairpin-shaped RNA molecule of 70–100 bp called pre-miRNA precursor or pre-miRNA. The mammalian microprocessor contains several additional protein factors, most of which have not been assigned a specific function in miRNA processing yet (27). Notably, the two DEAD-box RNA helicases DDX5 and DDX17 (also known as p68 and p72) have been identified as integral part of the microprocessor and are necessary for the efficient cleavage of a subset of pri-miRNAs (31).

The pre-miRNAs are specifically recognised by the nuclear export receptor Exportin 5 and exported in a Ran-GTP dependent manner (32–34). In the cytosol, the pre-miRNA is further processed by the RNase III enzyme Dicer (35). This enzyme binds the 3’ overhang of the pre-miRNA with its PAZ domain and thereby positions the substrate correctly for the cleavage by the two catalytic domains 22 nt upstream within the double stranded stem (36, 37). A recent study has reported that in metazoa, the 5’ end of the pre-miRNA is contacted by Dicer proteins as well. This interaction contributes to pre-miRNA recognition and correct processing (38). The result of Dicer cleavage is a double stranded RNA of 22 nt in length with 3’ overhangs of 2nt on both ends (1). One of the strands (the mature miRNA) is transferred into an Ago protein, whereas the other strand (the star (*)-strand) is degraded. In mammals, the strand selection and RISC assembly is accomplished by a complex that contains Dicer, Ago and the double stranded RNA binding protein TRBP (39–41). Statistical analyses have discovered that generally the strand with the less stable base pairing at the 5’ end is chosen as guide strand (42, 43), which requires an analysis of the thermodynamic properties of the double stranded miRNA after dicing. Recently, it has been proposed that the sensor for the thermodynamic asymmetry is the helicase domain of human Dicer itself. In this model, the double stranded Dicer cleavage product is repositioned on Dicer after the cleavage reaction. During this process, the helicase domain of Dicer senses the thermodynamic stability of the ends and positions the double stranded RNA in an orientation that allows for correct guide strand incorporation into RISC (►Fig. 1) (44).

Regulation of miRNA biogenesis

Given the tremendous impact of miRNA-guided gene regulation on almost all aspects of cellular biology, it is not surprising that miRNA levels are tightly controlled and that deregulation can lead to various diseases including cancer (6, 17, 20). The first and one of the most important layers governing miRNA abundance is the regulation of pri-miRNA transcription. The transcription of miRNAs encoded within introns of protein coding genes is mostly driven by the promoter of the host gene leading to a strong correlation of mRNA and miRNA expression. However, an analysis of chromatin signatures characteristic for promoters has identified additional promoters for about one third of the intronic miRNAs enabling a separate regulation from the host gene (45). Intergenic miRNAs have dedicated promoters, which show all features commonly associated with PolII-mediated transcription such as histone marks, CpG islands, transcription factor binding sites etc. (45). Clustered miRNAs share one promoter and are coregulated as part of long pri-miRNAs. Via the promoters, pri-miRNA transcription can be strongly regulated by binding of transcription factors. The tumour suppressor p53, for instance, has been shown to upregulate the transcription of miR-34 family members, which in turn repress important factors for cell proliferation and survival, such as Bcl2 and Cdk4 and 6 (46–49).

In addition to the transcriptional regulation, it was noted that also the processing of pri-miRNA transcripts can be the limiting
factor for miRNA expression and can be regulated at different steps leading to the accumulation of specific precursor forms (50, 51). This posttranscriptional regulation can affect a large number of miRNAs and is especially prominent in cancer samples and cell lines (50, 52). Since these discoveries, a small but increasing number of mechanisms that specifically regulate the processing of miRNA subsets have been reported and will be discussed below.

**Regulation of Drosha processing**

An emerging group of proteins that can modulate pri-miRNA processing by the microprocessor complex in response to diverse stimuli are transcription factors. For example, activation of Smad proteins by stimulation of cells with bone morphogenic protein (BMP) or tumor growth factor β (TGF-β) can stimulate the maturation of specific miRNAs. This activity is independent of Smad phosphorylation and binding of the Co-Smad4 and leads to increased recruitment of the pri-miRNAs to the microprocessor complex, enabling a more efficient cleavage by Drosha (53). Analysis of the about 20 miRNAs that are affected by this regulation revealed a consensus sequence that strongly resembles the Smad-binding element in DNA and is located in the double stranded stem of the pri-miRNAs. This motif is bound by the DNA-binding MH1 domain of Smad1, 3 and 5 and establishes the specificity of the regulated miRNAs (54). Another well-characterised Smad domain, the MH2 domain, binds to p68 (DDX5), which is part of the microprocessor complex and is necessary for the induction of miRNA processing by the Smad proteins (Fig. 1) (53).

Similarly, activation of the p53 pathway through DNA damage leads to an association of the p53 protein with the microprocessor complex via the p68 helicase (Fig. 1). This recruitment causes increased processing of a set of pri-miRNAs including mir-145, which exerts a tumour suppressive function via repression of c-myc (55). In contrast to the Smad-regulated miRNAs, no common sequence motif could be delineated from the induced miRNAs and it remains unclear how the target-miRNAs are selected.

An alternative mechanism is used by the estrogen receptor alpha (ERα), which associates with the microprocessor complex and inhibits the cleavage of specific miRNAs after ligand binding (Fig. 1). Again, the helicase subunits of the microprocessor p68 and p72 are crucial for the recruitment of ERα (56). Interestingly, several estrogen-regulated miRNAs target mRNAs that are transcriptionally upregulated during estrogen-response. Thus, the inhibition of miRNA maturation by ERα synergises with its transcriptional effects by stabilising the target mRNAs (56). However, how the specificity for the regulated miRNAs is achieved, needs to be elucidated.

A second group of factors regulating microprocessor cleavage of a subset of miRNAs are RNA binding proteins that also have a known role in RNA splicing (Fig. 1). The KH-type splicing regulatory protein (KHSRP) binds with high affinity to conserved terminal loop regions of a specific set of miRNAs. The binding motif is similar to binding sites in mRNA and consists of a single stranded GGG-triplet. KHSRP binds to pri-miRNAs including pri-let-7a and is necessary for microprocessor cleavage. Knockdown of KHSRP leads to a strong decrease in mature let-7a levels. KHSRP is present in most cells and constitutively promotes processing of so far 14 identified miRNA targets (57). Its activity can be augmented by phosphorylation of three serine residues by the ATM kinase, which increases the affinity of KHSRP for the pri-
miRNA-loops in response to DNA damage (58). The stimulatory role of KHSRP in miRNA processing is antagonised by hnRNPA1, which binds to the same site in pri-miRNAs and competes with KHSRP binding (59). In addition to its inhibitory role, hnRNPA1 can also activate microprocessor cleavage of pri-miR-18a by binding to the double stranded stem resulting in a more relaxed local helix structure more favourable for Drosha cleavage (60, 61). Interestingly, in spite of being part of the large miR-17–92 cluster, miR-18a is the only member that is regulated post-transcriptionally by hnRNPA1 (60, 61). This observation illustrates how regulation of miRNA processing is used to uncouple the expression levels of individual clustered miRNAs through posttranscriptional mechanisms.

Similar to the regulation of miR-18 by hnRNPA1, the SR-protein SF2/ASF can bind to the lower stem of pri-miR-7 and promote Drosha cleavage. As the translation of SF2/ASF mRNA is repressed by miR-7, this leads to the formation of a negative feedback loop that attenuates SF2 expression (62).

Adenosine deaminase acting on RNA (ADAR) enzymes bind double stranded RNA and deaminate adenosine to inosine. Apart from their function in mRNA editing they can also accept certain miRNA precursors as substrates resulting in an I-U wobble base pair in the modified RNA. For miR-142 it has been shown that Drosha processing of the edited pri-miRNA is substantially impaired. Instead, the pri-miRNA is recognised and degraded by the I-U specific nuclease Tudor-SN, resulting in a severe decrease in mature miRNA levels (63).

Regulation of pre-miRNA export

The pre-miRNA export receptor Exportin 5 is, like other importin-β/karyopherin family proteins, composed of HEAT repeats that form a solenoid-shaped molecule with a large binding groove for the cargo. In Exportin 5, this binding region is strongly basic and can perfectly accommodate the negatively charged RNA double helix of the pre-miRNA stem. An additional small pocket at the base of the groove binds the 2 nt 3’ overhang and ensures the specificity for Drosha cleavage products. The terminal loop is not bound and can vary in size and shape (64). This mode of binding might allow for the co-export of proteins bound to the terminal loop but not the double stranded pre-miRNA stem.

C-terminally truncated mutations of Exportin 5 have been identified in tumours with microsatellite instability and result in a global defect of pre-miRNA export. The resulting general decrease of mature miRNA levels is a hallmark of miRNA profiles of cancer cells (65). So far there is only indirect evidence for specific regulation of nuclear export of individual miRNA precursors. Several pri-miRNAs including mir-31 are processed by the microprocessor but retained in the nucleus in most cell lines (52). However, the detailed mechanisms as well as the functional consequences remain to be elucidated.

Regulation of Dicer processing

The best-studied regulator of pre-miR processing by Dicer is the protein Lin-28 (Fig. 1). In an evolutionary highly conserved mechanism, this protein binds to the terminal loop of almost all let-7-family miRNAs (66–68). This high affinity interaction is mediated by a cold shock domain and two zinc fingers that display a high homology to the NCp7 protein of HIV-1 (69–71). Lin-28 is highly expressed in embryonic stem cells and leads to a complete block in let-7 maturation thus ensuring that the expression of several let-7 targets (including c-myc) that are essential for the maintenance of the pluripotent state, remain unpressed. During differentiation, the expression of Lin-28 is lost, and mature let-7 miRNAs can be produced (66). Repression of Dicer processing involves the cytoplasmic polyuridylation of the pre-miRNA at its 3’ end (72), a reaction catalysed by terminal uridyltransferases (TUTases). Recent high-throughput sequencing data of miRNAs has shown that a considerable amount of pre-miRNAs carry at least a single U added to their 3’ end (73). Pre-let-7 bound Lin-28 recruits a specific TUTase, TUT4, and thereby acts as a processivity-factor mediating the specific polyuridylation of its bound pre-miRNA (74–76). In addition to the regulation of pre-miRNA processing by Dicer, it has been proposed that Lin-28 also plays a role in the nucleus by blocking Drosha cleavage. This has been shown in Caenorhabditis elegans where Lin-28 binds pri-let-7 co-transcriptionally and inhibits microprocessor activity (77). Very recently, it has been reported in mammals, that a homolog of Lin-28, Lin-28b, represses Drosha processing by sequestering pri-let-7 to nucleoli, where they are not accessible for the microprocessor complex (78). Apart from stem cells, Lin-28 and Lin-28b are expressed in specialised tissues and are frequently upregulated in cancer, where they promote tumour growth by repressing let-7 miRNAs.

Members of the let-7 family are not the only targets of repression by Lin-28 and TUT4. It has been shown that in cardiac tissue, pre-miR-1 can be uridylated and degraded in a Lin-28-dependent manner. Yet, under normal conditions, this process is prevented by the binding of the MBNL1-protein to the terminal loop of miR-1, thereby competing with Lin-28 for binding. In myotonic dystrophy, however, MBNL1 is sequestered by binding to expanded CUG or CCUG sequences that cause the disease. Thus, Lin-28 can bind pre-miR-1 and the expression of the mature miRNA is lost, which contributes to the cardiac manifestation of the dystrophy through dysregulation of membrane channels (79).

Regulation of miRISC stability

The final regulatory level of miRNA abundance is the stability of miRISC. More and more specific mechanisms that can destabilise miRNAs as well as miRISC protein components are uncovered. In stem cells, the tripartite motif protein Lin-41 acts as specific ubiquitin ligase for Ago proteins. As Ago proteins appear to be limiting for the activity of the miRNA pathway, this results in the global attenuation of miRNA-guided gene silencing (80). Upon differenti-
ation of stem cells, Lin-41 expression is downregulated, at least partially as result of repression by let-7 family members that are released from the maturation block imposed by Lin-28 and destabilise Lin-41 mRNA (80).

On the miRNA side, it is becoming evident that different miRNAs have distinct inherent half-lives that might even be encoded in their sequence. miR-382, for example, is relatively short lived and degraded by the exosome. A sequence in the 3’ end of the miRNA is crucial in this process (81). Moreover, several retinal miRNAs are rapidly turned over upon light stimulation (82). The detailed mechanisms of active miRNA turn over in mammals, however, has not been elucidated. In C. elegans, it has been demonstrated that miRNAs, which are not bound to target RNAs, are degraded by the 5’ to 3’ exoribonuclease Xrn2 (83). It is therefore very likely that more, so far unrecognised factors might influence miRNA stability and thereby regulate gene silencing.

Outlook

It is becoming more and more apparent that miRNA activity is heavily regulated at all steps of the miRNA pathway. It is likely that numerous, so far not identified RNA binding proteins are engaged in the regulation of miRNA function. It might even be possible that each miRNA family or miRNA cluster possesses its own repertoire of regulatory proteins. In addition, key protein components of the miRNA pathway are subject to phosphorylation (84). Currently, we do not understand how miRNA pathways are embedded into cellular signalling networks and it is possible that kinases and phosphatases are involved in the interplay of miRNAs and signalling molecules.

Very recently, pseudogenes have been implicated in the regulation of miRNA activity as well. Numerous conserved pseudogenes exist in the human genome and many of them contain 3’ UTRs with conserved miRNA binding sites. Interestingly, these binding sites are active and function as decoys or sponges sequestering miRNAs and preventing them from binding to their cognate mRNA targets. Such RNAs have been termed competing endogenous RNAs (ceRNAs) (85). These new findings highlight the complexity of regulatory miRNA networks. In order to understand the impact of miRNAs on disease and the development of miRNA-based therapeutics, it is essential to characterise such miRNA networks in molecular detail.

Acknowledgements

We apologise to those, whose work was not cited due to space constraints. Our research is supported in part by grants of the European Research Council (ERC, sRNAs to G.M.), the EU FP7 project ‘ONCOMIRs’, the Bavarian Genome Research Network (BayGene to G.M.) and the Deutsche Forschungsgemeinschaft (DFG).

Conflicts of interest

None declared.


