Review Article

MicroRNAs in Gene Regulation: When the Smallest Governs It All

Dominique L. Ouellet,1, 2 Marjorie P. Perron,1, 2 Lise-Andrée Gobeil,1, 2 Pierre Plante,1, 2 and Patrick Provost1, 2

1 Centre de Recherche en Rhumatologie et Immunologie, Centre de Recherche du CHUL, 2705 Boulevard Laurier, Ste-Foy, Quebec, Canada G1V 4G2
2 Faculté de Médecine, Université Laval, Quebec, Canada G1K 7P4

Received 30 January 2006; Accepted 17 April 2006

Encoded by the genome of most eukaryotes examined so far, microRNAs (miRNAs) are small ∼21-nucleotide (nt) noncoding RNAs (ncRNAs) derived from a biosynthetic cascade involving sequential processing steps executed by the ribonucleases (RNases) III Drosha and Dicer. Following their recent identification, miRNAs have rapidly taken the center stage as key regulators of gene expression. In this review, we will summarize our current knowledge of the miRNA biosynthetic pathway and its protein components, as well as the processes it regulates via miRNAs, which are known to exert a variety of biological functions in eukaryotes. Although the relative importance of miRNAs remains to be fully appreciated, deregulated protein expression resulting from either dysfunctional miRNA biogenesis or abnormal miRNA-based gene regulation may represent a key etiologic factor in several, as yet unidentified, diseases. Hence is our need to better understand the complexity of the basic mechanisms underlying miRNA biogenesis and function.

Copyright © 2006 Dominique L. Ouellet et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

In 1990, a group of plant biologists attempted to accentuate the purple pigmentation of petunia petals by increasing the synthesis of anthocyanin via overexpression of a transgene encoding chalcone synthase. Unexpectedly, this transgene induced the formation of white flowers, in association with a block in pigment synthesis and a 50-fold reduction in transgene mRNA levels; this intriguing phenomenon was termed cosuppression [1, 2].

Three years later, in the field of developmental biology, Lee et al [3] identified two lin-4 transcripts, with the smaller (∼21 nt) being complementary to seven repeated sequences in the 3’ nontranslated region (NTR) of the heterochronic gene lin-14 mRNA, identified previously by Wightman et al [4]. These findings suggested that lin-4 could regulate lin-14 translation via an antisense RNA:mRNA interaction and play an important role in developmental timing in the nematode Caenorhabditis elegans (C elegans) [3, 5].

These studies converged in 1998, when Fire et al [6] obtained evidences about the involvement of double-stranded (ds) RNA intermediates in a phenomenon termed RNA interference (RNAi). The authors noticed that dsRNA species induced a more potent genetic interference than either strands alone in C elegans. A year later, while investigating posttranscriptional gene silencing (PTGS) as a natural antiviral defense mechanism, Hamilton and Baulcombe [7] observed the presence of antisense viral RNA of ∼25 nt in virus-infected plants. The authors noted that these small RNAs were long enough to convey sequence specificity and suggested that they may be important specificity determinants of PTGS. Subsequent papers reporting that dsRNA-induced mRNA degradation is mediated by 21 to 23 nt RNAs [8, 9] prompted molecular biologists and geneticists to search for the endogenous source of small RNAs. In 2001, three independent groups defined miRNAs as a novel family of small (∼22 nt) endogenous RNAs that are diverse in sequence and expression patterns, evolutionarily widespread, and involved in sequence-specific, posttranscriptional regulatory mechanisms of gene expression [10–12].

We now know that miRNA genes are encoded in the genome of most eukaryotic organisms and transcribed by RNA polymerase (pol) II into primary miRNAs (pri-miRNAs). These structured RNAs are then processed by the nuclear RNase III Drosha, acting in concert with the DiGeorge syndrome critical region 8 (DGCR8) protein within a
Figure 1: mRNA regulation mediated by microRNAs (miRNAs). miRNA genes are transcribed by RNA polymerase II into primary miRNAs (pri-miRNAs), which are processed by Drosha, acting in concert with DiGeorge syndrome critical region 8 (DGCR8) protein within the microprocessor complex, into ∼60 to 70 nt miRNA precursors (pre-miRNAs). Following export via exportin-5, pre-miRNAs are cleaved by Dicer, acting in concert with transactivating response RNA-binding protein (TRBP) within the pre-miRNA processing complex, to generate an imperfect miRNA:miRNA$^*$ duplex of ∼21 to 23 nt. After a strand selection/separation process, the mature miRNA is loaded into an effector miRNA-containing ribonucleoprotein (miRNP) complex that will recognize and mediate repression or cleavage of specific mRNAs. Synthetic small interfering RNAs (siRNAs) can be introduced into cells and be incorporated into the endogenous miRNA-guided RNA silencing machinery to mediate cleavage of the targeted mRNA.

In humans, conservative predictions indicate that up to 30% of the genes may be regulated by such a mechanism! Thus, potentially all the cellular pathways may be governed by miRNAs, which may contribute to the fine tuning of gene expression on a global level. The importance of miRNAs in gene regulation will be better appreciated when their function or deregulation, or that of the cellular machinery mediating their biosynthesis and function, will be identified among the underlying causes of several genetic disorders. Indeed, it is easy to conceive that protein overexpression resulting from defective miRNA-based mRNA regulation may compromise normal cell function and cause genetic diseases. In turn, the responsible gene(s) may be responsive to RNAi-based inactivation, illustrating the transition from fundamental research to clinical applications of RNAi.

Today, the miRNA mimetics small interfering RNAs (siRNAs) are increasingly important molecular tools, as they often are the method of choice used by researchers that aim at elucidating the function of a gene. More importantly, perhaps, is the high potential of the approach for therapeutic applications, leading several biotechnology firms to develop and refine tools, and improve the design of new therapeutic strategies in order to take advantage of the natural RNA silencing machinery to silence the expression of disease-causing genes. This requires a better understanding of the miRNA-based RNA silencing machinery in human.
COMPONENTS OF THE RNAi MACHINERY

The endogenous miRNA-guided RNA silencing machinery is composed of several different proteins, protein complexes, and types of RNAs. How these elements integrate with each other to form this important functional cascade is the subject of intense investigations. We will first discuss the protein components, identified so far, that are governing miRNA biogenesis and function (see Figure 1). The subsequent sections will cover the identification of miRNAs and their targets, the biological roles of miRNAs, as well as their involvement in diseases.

RNA polymerase II

RNA pol II, which governs the transcription of protein-encoding messenger RNAs (mRNAs), has been identified as the major transcriptional unit for miRNA genes [17, 26] after some speculations about the potential implication of RNA pol III [25]. The pri-miRNA transcripts, which can be longer than 1000 nts [26] and up to several kilobases long, possess the signature of RNA pol II characterized by a 5′ guanylate (m7G) cap and a 3′ poly(A) tail [17]. Although miRNA genes can be found as clusters forming their own transcriptional units [19, 26], ~40% are transcribed from the intronic sequence of protein-encoding genes [27, 28]. A study by Rodriguez et al. [27] has shown that the expression of a large subset of mammalian miRNAs may be transcriptionally linked to the expression of other genes, coding for both proteins and ncRNAs. Although the majority of pri-miRNAs are noncoding RNAs, whose genomic regions do not correlate with known transcripts [29], some of them contain open reading frame (ORF) susceptible to be translated. However, analysis of both endogenous and overexpressed pri-miRNAs showed that very little full-length pri-miRNA transcripts reached the cytoplasm, probably because they were processed by Drosha before they could be exported from the nucleus [17].

Drosha

The RNase III Drosha is a class II endoribonuclease that was identified, cloned, and first implicated in preribosomal RNA (pre-rRNA) processing [30] (see Figure 2). Members of the class II RNAIII family are characterized by a duplication of the RNase III domain (RIIID), a C-terminal dsRNA binding domain (dsRBD), as well as a proline-rich region (PRR) and an arginine/serine(RS)-rich domain in the N-terminal region [18, 30]. Previously known as the human RNase III, Drosha was further identified as the enzyme mediating the first step in miRNA biogenesis through conversion of pri-miRNAs into pre-miRNAs [18] (see Table 1), confirming previous findings obtained with nuclear fractions of human cultured cells [19].

Drosha homologues are expressed in C elegans [13, 31], Drosophila melanogaster (D melanogaster) [13, 31], and Mus musculus [32], but not in Schizosaccharomyces pombe (S pombe) (http://www.sanger.ac.uk/Projects/S_pombe/) and Arabidopsis thaliana (A thaliana) [30]. The absence of Drosha in lower species reveals fundamental differences in the initiation steps of small regulatory RNA biosynthesis, which may have been evolved during the course of evolution.

Pri-miRNA processing by Drosha yields a pre-miRNA product with termini bearing the signature of RNases III, that is, a 5′ phosphate and 2 nt overhangs at the 3′ hydroxylated end [18, 19, 33]. From the junction of the loop and the adjacent stem, Drosha cleaves pri-miRNAs after approximately two helical turns into the stem to produce ~70 nt premiRNAs [18]. Lee et al. used deletional mutagenesis on miR-30a followed by in vitro processing to show that sequences covering ~20 nt upstream and ~25 nt downstream of the expected cleavage site were necessary and sufficient to support processing [18]. Beyond the pre-miRNA cleavage sites, approximately one helix turn of stem extension is also essential for efficient processing. While Drosha cleavage sites are determined largely by the distance from the terminal loop, variations in stem structure and sequence around the cleavage site can fine-tune the actual cleavage sites chosen [34]. A cleaving model was proposed in which the two RIIIDs of Drosha form an intramolecular dimer to create a catalytic site for substrate processing [15]. This model is analogous to that proposed for Dicer [35]. The two RIIIDs of human Drosha are distinct in their roles within the dimer: the RIIIDa cuts the 3′ strand, while the RIIIDb cleaves the 5′ strand, independently of each other [15]. Han et al. suggested that Drosha may reorientate itself after the recognition of the 3′ end of pri-miRNAs [15] and, as for human Dicer, places the processing center at <20 base pairs from the terminus [35].

Fractionation of HEK 293 cell nuclear extracts by gel filtration chromatography identified a pri-miRNA processing activity corresponding to a molecular mass of >700 kDa [15]. This activity peak shifted to ~650 kDa following treatment of the extract with RNase A, indicating that Drosha may function in a large complex of ~650 kDa. Analysis of Drosha immunoprecipitates by mass spectrometry revealed the presence of DGCR8 in that complex [13, 14]. A distinct, larger Drosha complex containing the DEAD box RNA helicase DDX17/P72, the heterogeneous nuclear ribonucleoprotein M4 (hnRNPM4), and the protein product of Ewing’s sarcoma gene (EWS) was reported [14]. As reviewed in Arvand and Denny [36], EWS belongs to a family of genes that encode proteins that may serve as adapters between the RNA pol II complex and RNA splicing factors. Because Drosha has also been previously shown to participate in pre-rRNA processing [30], this large Drosha complex has been suggested to mediate such pre-rRNA processing activities [14].

DGCR8/microprocessor

DGCR8 was identified in anti-Flag immunoprecipitates prepared from an HEK-293 cell line stably expressing Flag-Drosha [14]. This Drosha-DGCR8 complex, which has also been observed in other organisms [13, 16], has been termed microprocessor [14]. DGCR8 contains two dsRBDs and a WW domain that could interact with the N-terminal proline-rich region of Drosha [14].
An accessory protein required for optimal pri-miRNA processing; its absence results in a serious genetic disease [37,38].

Discovered in 2000, it was originally reported to be involved in pre-rRNA processing [30].

DGCR8

Drosha

AAA

Pri-miRNA

DGCR8

Drosha

AAA

Microprocessor

DGCR8

Drosha

AAA

Nucleus

Cytoplasm

Nuclear membrane

Exportin-5

Achilles’s heel of the miRNA pathway, it is targeted and saturated by some viral RNAs, such as VA1 from adenovirus [206].

The genome of A thaliana (DCL 1–4) and D melanogaster (DCR-1 and DCR-2) harbors four and two Dicer isoforms, respectively, in contrast to a single Dicer form in C elegans (Dicer-1), S pombe (DCR1), and H sapiens (Dicer).

Among the human Ago protein family, isoforms 1 to 4 bind to siRNAs and miRNAs, but only Ago2 is present in miRNA cleavage-competent RISC [93–95].

TRBP

TRBP may play a dual, yet opposing, role in HIV-1 pathogenesis and miRNA-guided RNA silencing [207].

Cell membrane

Pre-miRNA processing complex

Pre-miRNA

Dicer

TRBP

Ago2

TRBP

Exportin-5

Export of pre-miRNAs was greatly reduced by the inhibition of the Ran guanine nucleotide exchange factor, suggesting

The genome of A thaliana (DCL 1–4) and D melanogaster (DCR-1 and DCR-2) harbors four and two Dicer isoforms, respectively, in contrast to a single Dicer form in C elegans (Dicer-1), S pombe (DCR1), and H sapiens (Dicer).

Among the human Ago protein family, isoforms 1 to 4 bind to siRNAs and miRNAs, but only Ago2 is present in miRNA cleavage-competent RISC [93–95].

The DGCR8 gene is located in the q11.2 region of the human chromosome 22 that contains ∼30 genes and is a part of a common monoallelic deletion [37]. Patients carrying this heterozygous deletion and other chromosomal abnormalities in this region display clinical phenotypes defined as the Di-George syndrome, Conotruncal anomaly face syndrome, and Velocardiofacial syndrome [38]. Congenital heart defects, characteristic facial appearance, immunodeficiency, and behavioral problems are other manifestations of these genetic disorders [38].

Exportin-5

Subcellular localization studies previously showed that pri-miRNA and pre-miRNA processing is compartmentalized into the nucleus and cytoplasm, respectively [19], suggesting the existence of a pre-miRNA nuclear export step. Less than two years later, three independent groups reported the identification of Exportin-5, a member of the nuclear karyopherin β transporter family [21], as the nuclear pre-miRNA transporter [20, 22, 23].

Exportin-5-mediated pre-miRNA transport was either reduced upon downregulation of Exportin-5 by siRNAs [23] or enhanced upon overexpression in mammalian cells [39]. Export of pre-miRNAs was greatly reduced by the inhibition of the Ran guanine nucleotide exchange factor, suggesting
Table 1: Characteristics of the major protein components of the miRNA-guided RNA silencing pathway.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Role/function</th>
<th>Accession mRNA number</th>
<th>mRNA (bp)</th>
<th>Accession protein number</th>
<th>Amino acids</th>
<th>Molecular weight (kDa)*</th>
<th>Isoelectric point*</th>
<th>Intracellular localization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drosha</td>
<td>Processing of pri-miRNA into pre-miRNA</td>
<td>NM_013235</td>
<td>4764</td>
<td>NP_037367</td>
<td>1374</td>
<td>159</td>
<td>7.81</td>
<td>Nuclear</td>
</tr>
<tr>
<td>DGCR8</td>
<td>Assistance of Drosha function in the microprocessor</td>
<td>NM_022720</td>
<td>4461</td>
<td>NP_073557</td>
<td>773</td>
<td>86</td>
<td>5.94</td>
<td>Nuclear</td>
</tr>
<tr>
<td>Exportin-5</td>
<td>Nuclear export of pre-miRNA</td>
<td>NM_020750</td>
<td>5231</td>
<td>NP_065801</td>
<td>1204</td>
<td>136</td>
<td>5.79</td>
<td>Nuclear membrane</td>
</tr>
<tr>
<td>Dicer</td>
<td>Processing of pre-miRNA into miRNA:miRNA*</td>
<td>NM_177438</td>
<td>10 276</td>
<td>NP_803187</td>
<td>1922</td>
<td>219</td>
<td>5.68</td>
<td>Cytoplasmic, ER</td>
</tr>
<tr>
<td>TRBP</td>
<td>Assistance of Dicer function in pre-miRNA processing complex</td>
<td>M60801</td>
<td>1368</td>
<td>AAA36765</td>
<td>345</td>
<td>38</td>
<td>7.38</td>
<td>Cytoplasmic</td>
</tr>
<tr>
<td>Ago2</td>
<td>Component of miRNPs, repression of translation</td>
<td>NM_012154</td>
<td>3567</td>
<td>NP_036286</td>
<td>859</td>
<td>97</td>
<td>9.19</td>
<td>Cytoplasmic</td>
</tr>
<tr>
<td>FMRP</td>
<td>Component of miRNPs, repression of translation</td>
<td>NM_002024</td>
<td>4362</td>
<td>NP_002015</td>
<td>632</td>
<td>71</td>
<td>7.42</td>
<td>Cytoplasmic, nuclear</td>
</tr>
</tbody>
</table>

* Calculated from protein calculator v.3.2 (http://www.scripps.edu/~cdputnam/protcalc.html). ER, endoplasmic reticulum.

that it is catalyzed by RanGTP [22]. In fact, RanGTP was necessary for specific binding of pre-miRNAs by Exportin-5 [23]. Recognition of pre-miRNAs bearing a 2 nt 3′ overhang by Exportin-5 was superior than pre-miRNAs with 5′ or no overhangs [40]. As for the pre-miRNA terminal loop and stem, which should be more than 16 nt in length, their recognition is not sequence-specific [20, 40]. Recognition of a minihelix motif in the RNA allows transport by Exportin-5, as demonstrated by the efficient transport of VA1 RNA from adenovirus 5 [41]. The exact coordination links between Exportin-5 and the nuclear and cytoplasmic steps of miRNA biogenesis remain obscure and need further investigation.

**Dicer**

Dicer is a ribonuclease III that was first identified as an enzyme capable of generating ~21–23 nt RNA guide sequences from dsRNA to initiate RNAi in *Drosophila S2* cells [42]. Within a two-month period, three papers reported that null mutations in the Dicer gene altered developmental timing, in association with defective miRNA maturation and accumulation of pre-miRNAs, in *C elegans* [43, 44] and *Drosophila* [45]. Human Dicer cDNA, which had been identified two years before [46], was later cloned and the recombinant protein expressed, allowing the characterization of its RNA binding properties and RNase activity [47, 48]. Localized mainly in the cytoplasm [49] or the endoplasmic reticulum [47] of cultured cells, human Dicer is a large protein composed of several domains: an N-terminal putative ATPase/helicase domain containing a DECH box, a domain of unknown function (DUF283), a PIWI/Ago/Zwille (PAZ) domain, and a C-terminal RIID, composed of tandem RNase III motifs and a C-terminal dsRBD [35, 42, 47, 48].

Recently, data reported by Zhang et al [35] pointed towards the existence of a single catalytic center in human Dicer. The authors proposed a model in which Dicer would function through intramolecular dimerization of its two RIID, assisted by the flanking RNA binding domains, PAZ, and dsRBD. The PAZ domain of Dicer may participate in the recognition of the terminal 3′ overhangs of its pre-miRNA substrate [35]. In this model, each RIID cuts a single strand of the RNA duplex substrate after two turns of α-helices, at the end opposite to that cleaved by Drosha, to produce a new end bearing a hydroxylated 2 nt 3′ overhang and a phosphorylated 5′ end. The 2 nt overhang is measured by the alignment of the dimer rather than by the distance between active residues on one peptide chain, whereas the length of the product (~21 nt) is determined by the distance between the PAZ domain and the active site [35].

Genetic studies revealed that Dicer is essential for mammalian development, as Dicer-deficient mice die at the embryonic stage [50, 51]. However, the DCR-1 gene can be disrupted in mouse embryonic stem (ES) cells by conditional gene targeting. The generated Dicer-null ES cells are viable, despite being completely defective in the generation of miRNAs, and display severe defects in differentiation both in vitro and in vivo [52]. Similar conditional inactivation of the Dicer gene in ES cell lines compromised proliferation as well as miRNA maturation, possibly rationalizing the phenotype...
observed in Dicer-null animals [53]. Epigenetic silencing of centromeric repeat sequences [52, 53] and expression of homologous small dsRNAs [52] were also markedly reduced in Dicer-null ES cells. Re-expression of Dicer in knockout cells rescued these phenotypes [52]. These results suggest the involvement of Dicer in multiple fundamental biological processes in mammals, ranging from stem cell differentiation to maintenance of centromeric heterochromatin structure and centromeric silencing [52].

It is relevant to note that Dicer activity is potently stimulated by limited proteolysis induced by low concentrations of proteinase K in vitro [47, 48], indicating the presence of intrinsic regulatory domains of Dicer activity. As recently reported, cellular proteins interacting with Dicer such as Ago2 [54], fragile X mental retardation protein (FMRP) [55], TRBP [56, 57], and the protein kinase R (PKR)-activating enzyme (PACT) [58] may also represent key regulators of Dicer activity. In addition, Dicer was recently shown to be a part of an effector miRNP [24], thereby coupling the initiation and effector steps of miRNA-guided RNA silencing.

**TRBP**

TRBP was identified and characterized in 1991 as a cellular factor acting in synergy with the viral Tat protein in the transactivation of the long terminal repeat (LTR) of human immunodeficiency virus type 1 (HIV-1), leading to viral gene transcription [59]. TRBP exists mainly in two different isoforms: TRBP1 and TRBP2 [60], which possess three dsRBDS and a basic C-terminus, coexist in the cell and are encoded by two alternatively initiated isoforms of mRNA that differ at their 5′ ends. TRBP2 is 21 amino acids longer than TRBP1 [60–62]. TRBP has also been shown to bind Tax of human T-cell leukemia virus 1, although this interaction inhibits the transactivating activity of Tax [63]. Another function of TRBP is the inhibition of the interferon-induced dsRNA-regulated PKR [64].

Recently, TRBP has been reported to play a role in miRNA-guided RNA silencing. TRBP was identified by proteomic analysis of immunoprecipitates prepared from HEK 293-derived stable cell lines expressing a Flag-tagged Dicer [56]. Further analyses revealed the association of Dicer-TRBP with Ago2 and the requirement of TRBP for the recruitment of Ago2 to the siRNA bound by Dicer. TRBP was shown to facilitate the cleavage of pre-miRNAs in vitro and optimize RNA silencing mediated by siRNAs and endogenous miRNAs [57]. These results support a role for TRBP, the first Dicer-interacting protein identified since Ago2, in assisting Dicer function in a pre-miRNA processing complex and contributing to RISC assembly by the recruitment of Ago2 to the miRNA.

A Dicer interaction with the *Drosophila* homologue of human TRBP, Loquacious (Loqs), which share 34% identity at the amino acid level, was also observed by two independent groups [65, 66]. As for human TRBP, Loqs was required for normal processing of pre-miRNAs by Dicer-1 [65, 66] and for efficient miRNA-mediated silencing in various contexts [65]. Thus, every known fly RNase III is paired with a dsRBBD-containing protein that facilitates its function in small RNA biogenesis.

**R2D2**

The siRNA-generating complex purified from *Drosophila* S2 cells consists of two stoichiometric subunits: Dicer-2 and R2D2 [67]. R2D2, which was named so because it contains two dsRNA-binding domains (R2) and is associated with DCR-2 (D2) in *Drosophila* [67], is homologous to *C elegans* RDE-4 [68, 69]. The Dicer-2/R2D2 complex, but not Dicer-2 alone, binds to siRNA and enhances sequence-specific mRNA degradation mediated by the RISC. R2D2 has been shown to act as a biosensor for detecting thermodynamic differences of base pairing at the extremities of an siRNA [70]. Thus, in concert with Dicer-2, R2D2 determines which siRNA strand will be incorporated into the RISC and may also discriminate an impostor siRNA [71, 72]. These results indicate that R2D2 bridges the initiation and effector steps of the *Drosophila* RNAi pathway by facilitating siRNA passage from Dicer to RISC. Whether a similar mechanism is operating in humans remains to be investigated.

**RISC and miRNPs**

The miRNA or siRNA generated by Dicer is loaded into an effector miRNP or siRNP complex, respectively, and guides it for the recognition and regulation of the mRNA target. The miRNA specifically recognized by the RNP complexes will initially be either cleaved or translationally repressed, depending on whether the guide:mRNA pairing is perfect or not [25]. In humans, mRNA regulation by miRNAs is believed to consist mainly in translational repression, although a recent study reported that miRNAs downregulate a greater number of transcripts than previously thought [73]. Yekta et al [74] demonstrated that miR-196 shows perfect complementarity (presence of a single G:U wobble) with HOXB8 mRNA and directs its cleavage in mouse embryos. Genes for miR-196 map to homeobox (HOX) clusters, which encode transcription factors crucial for the developmental program in animals. Bagga et al [75] observed that the let-7 miRNA induces degradation of its target, lin-41, in *C elegans*. Furthermore, they observed that the level of the lin-4 miRNA targets, lin-14 and lin-28, is decreased in response to lin-4 expression. These observations suggest that miRNAs containing partial miRNA complementary sites may not only be subjected to translational repression, but also be targeted for degradation in vivo.

Initial studies on the RISC reported the existence of a large (≈ 150 kDa to ≈ 500 kDa) multiprotein RNP complex exhibiting sequence-specific nuclease activity [54, 76, 77]. Small RNAs of ≈ 21 to 25 nt were found to copurify with the RISC isolated from *Drosophila* S2 cells [76], a characteristic shared by RISC complexes from other species [78]. A study by Pham et al [78] provided the first glimpse of the mechanism involved in RISC assembly. The authors proposed a three-step model for RISC formation in *Drosophila*. Isolation of three distinct complexes, named R1, R2, and R3, by native gel electrophoresis showed that siRNA binding to Dicer-2 is
responsibility for R1 formation. R1 is probably the ~360 kDa complex described as the RISC [77]. R1 serves as a precursor to form both the R2 and R3 complexes. R3 is a large ATP-enhanced complex that contains unwound siRNAs, cofractionates with known RNAi factors, binds and cleaves targeted mRNAs in a cognate siRNA-dependent manner [78].

Recently, three studies published in the same issue of Cell [24, 79, 80] provided additional insights on the composition, assembly, and function of the RISC. Gregory et al. [24] showed that the human RISC is composed of at least three proteins: Dicer, TRBP, and Ago2. Recently, the dsRBD protein PACT was also found to be associated with Dicer, hAgo2, and TRBP in a ~500 kDa complex and to function as a component of the RISC [58]. At first, an ATP-dependent helicase was proposed to separate the two siRNA strands, one of which is thought to bind to Ago2. However, a recent consensus model suggests that Ago2 directly receives the double-stranded siRNA and cleaves the siRNA passenger strand instead, thereby liberating the ss guide for mediating cleavage or repression of the RNA target [24, 79, 80]. In contrast, passenger-strand cleavage is not important for the incorporation of miRNAs that are derived from mismatched duplexes, suggesting that this mechanism may not apply to endogenous miRNAs in humans.

In 2002, Mourelatos et al. [81] reported the identification and characterization of a miRNP complex showing high similarity with the RISC. The authors isolated a wide range of different miRNAs forming a complex with three major proteins: Gemin3, Gemin4, and EIF2C2 (hAgo2). Gemin3, a 105 kDa DEAD-box putative helicase, may be involved in unwinding the double-stranded miRNA and releasing the miRNA* strand for recognition of the target.

P/-GW-bodies

Where does the miRNP-mediated mRNA regulation or cleavage occur in the cell? Recent studies revealed the existence of specific cytoplasmic foci, referred to as processing (P-bodies) [82, 83] or GW182-containing bodies (GW-bodies) [84]. The GW bodies, which were named so because they contain the GW182 RNA-binding protein, are enriched in proteins that are involved in mRNA degradation [85]. Liu et al. [82] demonstrated the localization of Ago proteins into mammalian P-bodies. In fact, Ago proteins were found to interact with GW182 [86]. Silencing of GW182 or mutations that prevented Ago proteins from localizing in P/-GW-bodies impaired translational repression of mRNAs [86]. The presence of exogenous siRNAs was also detected in these bodies [87]. These studies support a functional link between cytoplasmic P/-GW-bodies and mRNA translation repression mediated by miRNAs.

These cytoplasmic P/-GW-bodies may not be the only sites of mRNA degradation in the cell. Two independent groups also detected a RISC-like activity in the nucleus of cultured mammalian cells [88, 89]. It is tempting to speculate that the nuclear effector complex mediating this activity may be closely related to the RNA-induced initiation of transcriptional gene silencing (RITS) complex found in S pombe [90].

Ago2

Ago2 is a member of the PAZ and Piwi domain (PPD) protein family, which is composed of highly basic proteins that are present in metazoans and fungi, but not in the budding yeast Saccharomyces cerevisiae [91, 92]. Eight members of the Ago family are expressed in humans [93], and the isoforms Ago1 to 4 are closely related. All four can bind siRNAs and miRNAs, but only Ago2 is present in an mRNA-cleavage competent RISC [94, 95]. Several paralogues of Ago proteins are found across the kingdoms and their number varies from 1 in S pombe [96] to more than 20 in C elegans [43, 97].

Structural studies have provided key insights into the mechanism of RNAi. Ago2 is composed of a central PAZ domain and a C-terminal PIWI domain. The nuclear magnetic resonance solution structure of the Drosophila Ago1 PAZ domain bound to RNA was resolved recently [98, 99]. The structure consists of a left-handed, six-stranded β-barrel capped at one end by two α-helices and wrapped on one side by a distinctive appendage, which comprises a long β-hairpin and a short α-helix. Combined structural and binding studies of the PAZ domain indicated that it provides a binding pocket for the 3’ protruding ends of siRNAs [98–101].

Structural studies revealed that the PIWI domain consists of 5-stranded β-sheets surrounded by three helices [102] and mediates binding of the ss RNA 5’ end [103–105]. The structure of Archaeoglobus fulgidus PIWI domain in complex with an siRNA-like duplex, which mimics the 5’ end of a guide RNA strand bound to an overhanging target mRNA, has been solved. This study revealed the presence of a highly conserved metal-binding site that anchors the 5’ nt of the guide RNA [105]. Structural studies also determined PIWI as the catalytic domain for the nuclease activity of Ago2, given its resemblance to RNase H [95, 102, 104], in terms of structure and activity; like RNase H, Ago2 activity is dependent on divalent cations such as Mg2+ or Mn2+ [106]. The PIWI domain and RNase H also share a DDE motif, similar to those present at the catalytic center of integrase proteins [107].

Structural information from Pyrococcus furiosus Ago [102], together with the demonstration that Ago2 is the core slicing machine of the human RISC [95, 108], provided strong evidences suggesting that the PIWI domain may be responsible for mediating this “slicer” activity. This possibility was further supported by the observed inhibition of target mRNA cleavage activity upon deletion of the DDE motif of hAgo2 [95]. In siRNA-guided RNA silencing, Ago2 cleaves, in an ATP-independent manner, the phosphodiester backbone of the target mRNA between nucleotides 10 and 11, as calculated from the RNA guide 5’ end [80, 94, 109].

The mechanism of translation repression mediated by Ago proteins is still unclear, although recent evidences suggest the possibility that some Ago-containing complexes may repress translation in P-bodies [110, 111].

FMRP

In human, loss-of-function mutations in the FMR1 (fragile mental retardation 1) gene product FMRP is the cause of the most common mental retardation, the fragile X syndrome.
An expansion of the CGG repeat in the 5′NTR of FMR1 is associated with DNA methylation problems of both the CpG island and the CGG repeat itself, resulting in an inhibition of transcription and translation [114, 115].

FMRP is a cytoplasmic RNA-binding protein found to be associated with polyribosomes as part of an mRNA ribonucleoprotein (mRNP) complex, suggesting a role for FMRP in mRNA translation regulation [112]. In fact, this protein of 632 amino acids, containing two K-homology (KH) domains and an RGG box, acts as a negative regulator of translation in vitro and in vivo [116–118]. A relationship between FMRP and the RNAi pathway was unveiled by the copurification of dFMR1 with the *Drosophila* RISC, which also contains Ago2 and the vasa intronic gene (VIG) [119]. Similarly, Ishizuka et al. [55] used a tandem affinity purification approach to isolate an RNP complex that contains dFMR1, Ago2, the RNA helicase Dmp68, and the ribosomal proteins L5/5S RNA and L111. Ishizuka’s group demonstrated that dFMR1 is a component of the RISC effector complex and is associated with Dicer and Ago2 [55]. Knockdown of dFMR1 by introduction of dFMR1 dsRNA had only mild effects on the efficiency of RNAi [55, 119].

The work of Jin et al. [120] suggested that FMRP could interact with miRNAs, Dicer, and Ago1 in mammalian cells in vivo, raising the possibility that FMRP could use miRNAs to regulate translation of specific mRNAs. Indeed, a recent study from our laboratory showed that human FMRP can act as an mRNA acceptor protein for Dicer and facilitate the assembly of miRNAs on specific target RNA sequences [121]. This activity appears to be mediated by the KH domains. In this study, the requirement of FMRP for efficient RNAi in vivo was unveiled by reporter gene silencing assays using various small RNA inducers, which also supported its involvement in an ss siRNP effector complex in mammalian cells. These results defined a possible role for FMRP in miRNA-guided RNA silencing and provided further insight into the molecular defects in patients with the fragile X syndrome.

**VIG**

The VIG protein has been shown to be associated with the *Drosophila* RISC [119]. An evolutionarily conserved protein expressed in *C elegans, A thaliana, S pombe*, and mammals, VIG has no recognizable protein domains other than an RGG box, a motif that is known to bind RNA. Although no function has been assigned to VIG, its human homologue, plasminogen activator inhibitor (PAI)-RBP-1, was originally identified as a protein having an affinity for AU-rich elements (ARE) located in the 3′NTR of PAI RNA and regulating its stability [122]. The authors also demonstrated the importance of Dicer and miR-16, a human miRNA containing a sequence complementary to ARE, in conferring instability to ARE-containing mRNAs. This suggests an interesting connection between the components of the miRNA-guided RNA silencing pathway and regulation of the stability of mRNAs containing AREs, which are known to act *in cis* to regulate rapid turnover of unstable mRNAs [123] in their 3′NTR. The exact role of VIG in that context remains to be investigated.

**Tudor-SN**

The staphylococcal nuclease Tudor (Tudor-SN) has been identified as a component of the RISC in *C elegans, D melanogaster*, and humans [124]. Tudor-SN contains five staphylococcal/micrococcal domains and a tudor domain. At first, Tudor-SN was suspected to be the nuclease responsible for the RISC-mediated mRNA target cleavage. However, studies demonstrating that the nuclease activity of the RISC is Mg<sup>2+</sup>-dependent [106] and produces 5′-phosphomonoester ends [125] did not support this hypothesis, as Tudor-SN is rather a Ca<sup>2+</sup>-dependant nuclease that generates 3′-phosphomono- and dinucleotides from DNA or RNA substrates [126].

Recently, a novel relationship was established between Tudor-SN and adenosine deaminases that act on RNA (ADARs). Members of the ADAR family exhibit affinity with dsRNAs and mediate an RNA editing reaction that substitutes adenosine (A) residues by inosines (I) in cellular mRNAs or other dsRNA targets [127]. Scadden [128] showed that Tudor-SN specifically interacts with and promotes the cleavage of model hyper-edited dsRNA substrates containing multiple IU and UI pairs. Yang et al. [129] have recently reported that the edition of pri-miR-142 resulted in the suppression of its processing by Drosha, and was instead degraded by Tudor-SN. Similarly, pre-miRNAs have also been shown to be edited by ADARs [130]. ADAR-induced modification of pri- and pre-miRNA sequences may also contribute to diversifying and influencing the genetic control mediated by miRNAs. For example, structural changes induced by A-to-I edition of pri- and pre-miRNAs may hamper their recognition and processing by the dsRNA-cleaving Drosha and Dicer RNases [131–133]. These studies reveal a new function for RNA editing in the control of miRNA biogenesis.

**RITS complex**

In the fission yeast *S pombe*, dsRNA arising from centromeric repeats targets the formation and maintenance of centromere function through RNAi-mediated histone H3 lysine-9 (K9) methylation [90]. This is accomplished by the effector complex RITS, which contains the proteins Ago1, Chp1, and Tas3, in addition to small RNAs [96] generated by Dicer [134]. Homologous to centromeric repeats [135], these small RNAs appear to guide the RITS components to heterochromatic regions, such as the centromeres, the mating-type region, and the telomeres [136, 137]. Upon centromeric binding, RITS promotes Clr4-mediated methylation of histone H3 K9, recruitment of Swi6 [138], and formation of heterochromatin [96]. Recently, a study concluded that Dicer and the RNAi machinery were involved in the formation of heterochromatin in higher vertebrate cells, as suggested previously [134]. The discovery of the effector RITS complex supports a nuclear function for small RNAs derived from Dicer.
IDENTIFICATION OF miRNAs AND THEIR TARGETS

As the major protein components of the miRNA-guided RNA silencing pathway are being identified and characterized, hundreds of new miRNAs are being discovered in several different species. The fact that the interaction between miRNAs and the mRNA targets they regulate is based mainly on partial, rather than perfect, complementarity renders target identification rather arduous. However, improvement of our understanding of the determinants governing mRNA recognition by miRNAs has allowed the development of several predictive bioinformatic tools. The growing number of miRNA targets and functions, as revealed by various experimental approaches, let us foresee the importance and complexity of the gene regulatory network utilizing miRNAs.

Identification of miRNAs

Almost 8 years after the discovery of the ncRNA lin-4, known for its crucial role in developmental timing in C. elegans [3, 5], three independent groups defined miRNAs as a novel family of small (~22 nt) regulatory RNAs that are diverse in sequence and expression patterns, and evolutionarily widespread [10–12]. The authors used different strategies to identify new miRNAs from various species. miRNAs showing features reminiscent of Dicer cleavage can be cloned by reverse transcription-polymerase chain reaction (RT-PCR) on size-fractionated RNA populations. If their sequences are known or predicted, and if they are abundant enough, miRNAs can be detected by Northern blot analysis.

Rapid and large-scale identification of miRNAs prompted experts in the field to establish guidelines for miRNA annotation and institute different criteria, based on expression and biogenesis, for an RNA to be considered as an miRNA [139]. First, a 22 nt RNA transcript must be detected by Northern blot analysis. Second, the RNA transcript must be detected in a cDNA library prepared from size-fractionated RNA samples. Third, bioinformatic analyses must predict a hairpin-loop structure encoded in the genome and the sequence has to be located on one arm of this structure with a lowest free energy. The hairpin should have small bulges and generally 60–80 nt in length. Phylogenetic conservation among species represents another important feature which, however, excludes miRNAs that have either disappeared, appeared, or evolved during the course of evolution.

Computational algorithms designed to identify hairpin-loop structures and sequence conservations across species are very useful, especially for less abundant or tissue-specific miRNAs. These small RNAs can be regrouped into families, based on the sequence of their 5′ region [139]. One computational algorithm that has been developed and tested with C. elegans, miRScan, uses different characteristics to identify miRNA genes. It has been designed to find conserved sequences upstream and downstream of the miRNA foldback, identify specific adjacent sequences that can be involved in miRNA transcription or processing, and determine the location of cotranscribed miRNAs in orthologous host genes [140].

Identification of miRNA targets

Identification of miRNA targets is a key step in understanding the biological function of miRNAs. However, the progress of this work is hampered by the mode of mRNA recognition by the regulatory miRNAs itself, which is based on imperfect sequence complementarity [25]. Characterization of a few experimentally validated miRNA:mRNA interactions allowed to establish a context in which this interaction is favored and helped to develop very useful bioinformatic approaches to identify them. Initial studies indicate that a given miRNA may regulate several different miRNAs and that, conversely, a specific mRNA can be regulated by more than one miRNAs.

Several algorithms currently available on the web, such as TargetScan (http://genes.mit.edu/targetscan/) [146], Miranda (http://www.microrna.org) [147], and DIANA-microT (http://www.diana.pcbi.upenn.edu/) [148], combine different parameters of the sequence requirements for miRNA:mRNA binding as predictive methods to identify targets. These computational tools are designed to scan the 3′ NTR of miRNA targets, to search for the miRNA seed and to determine the free energy of the interaction. They can also take into account the phylogenetic conservation and the presence of more than one miRNA binding site in a given 3′ NTR. Because each of these methods uses different miRNA:mRNA target predictive determinants, the results obtained may differ from one to another. Nevertheless, these bioinformatic tools are crucial in providing initial cues as to the possible miRNA targets regulated by specific miRNAs. They also offer a certain basis for initiating experimental validation on miRNA:mRNA target pair of interest. In turn, a better comprehension of the interaction between miRNAs and their targets will permit the improvement of these predictive methods.

Vella et al. [143, 149] studied the well-characterized let-7:lin-41 interaction to better understand the architecture and requirements of miRNA:miRNA target recognition. Although the lin-41 mRNA target bears six putative let-7 miRNA binding sites, only two of them appear to be necessary for lin-41 regulation (see Figure 3). These two sites are separated by a 27 nt sequence. Generally, the miRNA seed consists in a perfect pairing between miRNA nucleotides 2 to 8 with a
sequence located in the 3′NTR of its mRNA target. Although pairing of the 3′ region of an miRNA seems to be less important, it may compensate a weaker binding of the 5′ region. The authors also observed that lin-41 regulation by let-7 was lost upon substitution of the intervening 27 nt sequence by another [143]. This suggests that miRNA:mRNA interactions do not rely solely on the regions of complementarity and may be more complex than previously thought.

In spite of the difficulties to identify miRNA targets, several groups have found their way to assign a biological function to some miRNAs. Hematopoietic cell lineages derived from mouse bone marrow express specific miRNAs that regulate differentiation. Chen et al [150] analyzed three miRNAs, miR-181, miR-223, and miR-142 that were differentially or preferentially expressed in hematopoietic cells. They observed that overexpression of these miRNAs in undifferentiated progenitor cells derived from mouse bone marrow altered lineage differentiation. They further analyzed the effects of miR-181 in vivo by infecting mouse Lin− bone marrow cells with a viral vector expressing this miRNA and observed that mice expressing miR-181 had a substantial increase in B-lymphoid (CD19+) cells. Lim et al [73] used a microarray approach to identify miRNA targets after overexpression of known miRNAs. They found that 174 genes were downregulated following overexpression of miR-124, an miRNA preferentially expressed in the brain, in HeLa cells. Incidentally, the target genes were expressed at low levels in the brain. Thus, the expression of miR-124 in HeLa cells caused a shift in gene expression profile towards that of the brain. Using the same approach, expression in HeLa cells of miR-1, expressed in muscle, shifted the expression profile of HeLa cells towards that of the muscle.

**BIOLOGICAL FUNCTION OF miRNAs**

As experimental evidences are accumulating on how miRNAs recognize and regulate specific mRNA targets, we are beginning to understand the exact function of each miRNA as well as the cellular processes they are regulating. Information pertaining to the biological function of miRNAs in different species, which is the subject of this section, is summarized in Table 2.

**miRNAs and development**

Developmental studies of the nematode *C. elegans* led to the discovery of the first small noncoding regulatory RNA, lin-4. *C. elegans* proceeds through four larval development stages termed L1 to L4. Transition from one stage to the next is dictated by temporally regulated heterochronic genes, which are involved in developmental regulatory cascades. Wightman et al [4] first reported that short repetitive sequences in the 3′NTR of the heterochronic gene *lin-14* were negative regulatory elements of Lin-14 expression. The observed increase in Lin-14 protein synthesis associated with two gain-of-function mutations in the 3′NTR of the heterochronic gene *lin-14* were negative regulatory elements of Lin-14 expression. The observed increase in Lin-14 protein synthesis associated with two gain-of-function mutations in the 3′NTR of the heterochronic gene *lin-14* were negative regulatory elements of Lin-14 expression. The observed increase in Lin-14 protein synthesis associated with two gain-of-function mutations in the 3′NTR of the heterochronic gene *lin-14* were negative regulatory elements of Lin-14 expression. The observed increase in Lin-14 protein synthesis associated with two gain-of-function mutations in the 3′NTR of the heterochronic gene *lin-14* were negative regulatory elements of Lin-14 expression.

Recently, Boehm and Slack [169] found that lin-4 and lin-14 expression control life span through adulthood, since lin-4 loss-of-function mutant is associated with a shorter life span as compared to wild-type nematodes, whereas overexpression of lin-4 prolonged it. They also noted that animals carrying a temperature-sensitive loss-of-function mutation in *lin-14* had a 31% longer life span than wild-type, which is consistent with the phenotype observed with *lin-4* [169].

A second small ncRNA, let-7, was later identified and found also to regulate the transition from late larval L4 to adult stage through the regulation of heterochronic genes.
Table 2: Biological functions of miRNAs in different species.

<table>
<thead>
<tr>
<th>Species</th>
<th>Expression</th>
<th>miRNA target</th>
<th>Validation (Expt/Pred)</th>
<th>Mode of regulation</th>
<th>Process regulated</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizosaccharomyces pombe</td>
<td>—</td>
<td>12 different small RNAs</td>
<td>Expt</td>
<td>Translational repression/mRNA cleavage</td>
<td>Maintenance of heterochromatin</td>
<td>[135]</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>let-7 lin-41 lin-12 ras</td>
<td>Expt</td>
<td>Translational repression/mRNA cleavage</td>
<td>Control of the last developmental stage L4</td>
<td>[75, 149, 152–156]</td>
</tr>
<tr>
<td>Drosophila melanogaster</td>
<td>—</td>
<td>miR-48 miR-84 miR-241</td>
<td>Expt</td>
<td>Translational repression</td>
<td>Control of the L2 to L3 transition</td>
<td>[157]</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>bantam hid</td>
<td>Expt</td>
<td>Translational repression</td>
<td>Inhibition of apoptosis</td>
<td>[158]</td>
</tr>
<tr>
<td>Mus musculus</td>
<td>—</td>
<td>miR-2 miR-6 miR-11 miR-13 miR-308 miR-142 miR-181 miR-223</td>
<td>Expt</td>
<td>Translational repression</td>
<td>Inhibition of apoptosis</td>
<td>[159]</td>
</tr>
<tr>
<td></td>
<td>Hematopoietic cells</td>
<td>ND ND Expt ND</td>
<td>Hematopoietic cell differentiation</td>
<td></td>
<td></td>
<td>[150]</td>
</tr>
<tr>
<td>Mouse embryo &lt; day 7</td>
<td>miR-196 HOXB8</td>
<td>Expt</td>
<td>mRNA cleavage</td>
<td>Control of the developmental program</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homo sapiens</td>
<td>—</td>
<td>miR-124 ND</td>
<td>Expt</td>
<td>Brain-specific gene expression</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Brain</td>
<td>ND ND Expt ND</td>
<td>Brain-specific gene expression</td>
<td></td>
<td></td>
<td>[73]</td>
</tr>
<tr>
<td></td>
<td>Muscle</td>
<td>miR-1 ND Expt ND</td>
<td>Muscle-specific gene expression</td>
<td></td>
<td></td>
<td>[73]</td>
</tr>
<tr>
<td></td>
<td>Overexpression in brain tumor glioblastoma</td>
<td>miR-21 ND Expt ND</td>
<td>Inhibition of apoptosis</td>
<td></td>
<td></td>
<td>[160]</td>
</tr>
<tr>
<td></td>
<td>Downregulation in CLL</td>
<td>miR-15a miR-16-1 Bcl2 Expt</td>
<td>Translational repression</td>
<td>Induction of apoptosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Downregulation in numerous cancer cultured cells</td>
<td>miR-143 miR-145 ND Expt ND</td>
<td>Oncogenesis</td>
<td></td>
<td></td>
<td>[163]</td>
</tr>
<tr>
<td></td>
<td>Overexpression in BL patient</td>
<td>miR-155 ND Expt ND</td>
<td>Oncogenesis</td>
<td></td>
<td></td>
<td>[164]</td>
</tr>
<tr>
<td></td>
<td>Overexpression in B-cells lymphomas and lung cancer cell lines</td>
<td>miR-17-92 cluster ND Expt ND</td>
<td>Oncogenesis</td>
<td></td>
<td></td>
<td>[165]</td>
</tr>
</tbody>
</table>
in *C. elegans* [152]. Northern blot analyses revealed that the miRNA let-7 is expressed in a wide range of species, including worm, fly, and human, as opposed to lin-4, and seems to regulate late developmental transition in different species [153]. Three let-7 miRNA family members, miR-48, miR-84, and miR-241, were identified on the basis of sequence identity of 8 consecutive nucleotides in their 5′ region [11, 170]. let-7 regulates lin-41, hbl-1, and daf-12 [149, 152–155]. The other members of the let-7 family appear to regulate hbl-1 in the L2 to L3 transition [157]. let-7 also appears to be important for zebrafish embryo development, since injection of a synthetic let-7 miRNA duplex into zebrafish zygotes causes severe growth defect [171]. Embryos of a maternal-zygotic zebrafish Dicer mutant that were unable to process pre-miRNA into miRNA showed abnormal brain morphogenesis [172]. This brain defect was rescued by the injection of a preprocessed, mature miRNA, miR-430. The miRNA expression profile of zebrafish embryos is highly tissue-specific during segmentation and later stages, but not in early development, suggesting that miRNAs may play a more prominent role in differentiation or maintenance of tissue identity, rather than in directing tissue fate [173].

miRNAs have also a major role in developmental regulation in fly. This conclusion came from miRNA loss-of-function analyses using 2′ O-methyl (Me) antisense oligoribonucleotides in *Drosophila* embryos [159]. In these analyses, depletion of as many as 25 of 47 miRNAs expressed in early development caused a severe developmental phenotype. In situ hybridization analyses, using probes recognizing 38 different miRNAs in *Drosophila* embryos, indicated that the expression profile of most of them is comparable to their vertebrate counterparts [174], suggesting an evolutionarily conserved role for miRNAs in development.

Recently, two groups independently reported the cloning of the mouse and chicken homologues of *C. elegans* lin-41 [175, 176]. They found that mlin-41 and clin-41 are implicated in limb development. Bioinformatic analyses confirmed the presence of let-7 binding sites in the 3′ NTR of these two genes. In mice, targeted disruption of the *Dicer*1 gene was lethal in early development, indicating that Dicer function is essential for proper development in mammals [50]. Harfe et al [177] used an inducible inactivation system of *Dicer*1 to study its importance in late development in mice. In this model, depletion of Dicer led to a severe defect in limb formation.

### miRNAs and heterochromatin

The RNAi pathway was also reported to play a role in nuclear events such as genome rearrangement [178], gene inhibition [90, 134, 179], and chromosome segregation [134], supporting the idea that the genome integrity itself is preserved by small regulatory RNAs. In a model proposed by Noma et al [136], dsRNA transcripts are cleaved by Dicer to produce siRNAs, which are incorporated into the RITS complex and guide it to heterochromatic regions, probably through interactions with DNA or native RNA transcripts. Once localized at the siRNA homologous target sequence, the RITS complex recruits the Ctr4 methyltransferase that catalyzes methylation of histone H3 at lysine 9. This creates binding sites for the heterochromatin protein Swi6 which, in turn, leads to the recruitment of additional Ctr4 and further H3-Lys9

<table>
<thead>
<tr>
<th>ORF2 in the 3′ NTR of all remaining PFV-1 mRNAs</th>
<th>Restriction of PFV-1 accumulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>HeLa, 293T miR-32 Expt ND [166]</td>
<td></td>
</tr>
<tr>
<td>Liver-specific expression miR-122 5′ noncoding region of the HCV genome Expt ND</td>
<td>Facilitation of HCV replication [167]</td>
</tr>
<tr>
<td>T cells miR-29a nef Pred ND</td>
<td>Inhibition of HIV-1 replication [168]</td>
</tr>
<tr>
<td>T cells miR-29b vpr Pred ND</td>
<td>Control of the cell cycle arrest in G2 [168]</td>
</tr>
<tr>
<td>T cells miR-149 vpr Pred ND</td>
<td>Control of the virus assembly [168]</td>
</tr>
<tr>
<td>T cells miR-378 env Pred ND</td>
<td>Control of viral particle production [168]</td>
</tr>
<tr>
<td>T cells miR-324-5p vif Pred ND</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2: Continued.**

BL: Burkitt lymphoma; CLL: chronic lymphocytic leukemias; Expt: experimental; HCV: hepatitis C virus; HIV-1: human immunodeficiency virus type 1; ND: not determined; PFV-1: primate foamy virus type 1; Pred: predicted.
methylation of adjacent nucleosomes. These modifications allow the binding of RITS in a Dicer-independent manner via the chromodomain of Chp1 and the maintenance of gene repression at the transcriptional level. Although this process was first documented in *S. pombe* [96, 136, 180], recent reports indicate the occurrence of a similar transcriptional gene silencing phenomenon in the nucleus of human cells [181, 182]. Altogether, these studies reveal a key role for Dicer-derived small RNAs in guiding the RITS complex and regulating the transcriptional and posttranscriptional status of host gene expression.

**miRNAs in cell growth and apoptosis**

Cell growth and programmed cell death are important processes implicated in both development and differentiation. The *bantam* gene identified in *Drosophila* was first discovered on the basis of its effect on tissue growth: tissues were larger when *bantam* was overexpressed and smaller when *bantam* expression was suppressed. Although smaller, the flies were proportional and did not exhibit patterning defects [158]. Later, the same group determined that the *bantam* gene encoded for an miRNA, not for a protein, that controlled the proapoptotic gene *hid*. Thus, *bantam* promotes proliferation while inhibiting apoptosis [183]. Additional miRNAs involved in the regulation of pro-apoptotic genes in *Drosophila* were discovered in loss-of-function experiments. In that context, a family of miRNAs comprising miR-2, miR-6, miR-11, miR-13, and miR-308 has been found to be required for suppression of embryonic apoptosis [159].

Chang et al. [184] used a library of miRNA antisense oligonucleotides bearing 2'-O-Me modifications to inhibit miRNA function in HeLa cells. Monitoring changes in cell growth and apoptosis, they identified several important regulatory miRNAs. In the highly malignant human brain tumor glioblastoma, miR-21 was strongly overexpressed. When miR-21 was knocked down in glioblastoma cultured cells, caspases were activated, causing an increase in cell apoptosis [160]. This suggests a role for miR-21 as a suppressor of apoptosis in this malignant tumor [160]. In chronic lymphocytic leukemias (CLL), the antiapoptotic protein B-cell lymphoma 2 (Bcl2) is overexpressed [161]. Interestingly, frequent deletions and downregulation of the miR-15 and miR-16 genes at the chromosome locus 13q14 are observed in the majority of CLLs [162]. These findings suggest a role for miR-15a and miR16-1 as repressors of Bcl2 expression and possible inducers of apoptosis [161].

**RELATIONSHIP BETWEEN miRNAs AND DISEASES**

Given their recognized importance in gene regulation, a link between miRNAs and several major diseases is expected. For example, defects in miRNA-mediated regulation of mRNA translation may lead to overexpression of specific proteins, which accumulation may cause diseases. In fact, intriguing connections between miRNAs and diseases, such as cancer and viral infections, are emerging.

**miRNAs and cancer**

A recent study reported that human miRNA genes are frequently located at fragile sites and genomic regions involved in cancer [185]. Indeed, Calin et al. [162] observed frequent deletions and downregulation of miR-15 and miR-16 genes at 13q14 in CLL. These miRNAs have been shown to negatively regulate the antiapoptotic Bcl2 protein at the posttranscriptional level [161]. BCL2 repression by these miRNAs induced apoptosis in a leukemic cell line model [161], thereby providing a link between the absence of miR-15/miR-16 and leukemia.

Northern blot analyses showed that miR-143 and miR-145 expression is downregulated in various human cell lines derived from breast, prostate, cervical, lymphoid cancers, and, particularly, colorectal tumors [163]. Potential targets of these miRNAs have been previously implicated in oncogenesis [163].

A relationship between miRNAs and Burkitt lymphoma (BL) has been suggested. miR-155 is encoded within nucleotides 241–262 of the BIC gene, which is located on chromosome 21. Both the BIC and miR-155 genes are overexpressed in some BL patients, but not in all BL cases [164, 186]. Abnormal miRNA expression may thus contribute to the transformation of B cells [164].

Another miRNA cluster, miR-17-92, is often overexpressed in tumor samples from B-cell lymphomas when compared to normal cell lines [165]. This cluster is present in an amplified DNA region encoding for the ORF c13orf25. Alignment of this ORF between mouse and human indicates that the polycistron and its immediate flanking sequences only are conserved. The c13orf25 transcript contains seven pre-miRNAs encoding for miR-17-5p, miR-17-3p, miR-18, miR-19a, miR-20, miR-19b-1, and miR-92-1. Using a microarray analysis of 191 mature miRNAs, five miRNAs from the cluster were found to be highly expressed in B-cell lymphomas, in correlation with an increased expression of c13orf25 [165]. These studies revealed that the miR-17-92 cluster can act as a potential human oncogene, and was referred as oncomiR-1 by Hammond and colleagues [165].

Another group subsequently reported that the miR-17-92 cluster was overexpressed in human lung cancer cell lines [187]. When analyzing the subcellular localization of the c13orf25 transcript, they observed a nuclear localization, restricting its cytoplasmic localization for translation. This suggested that this transcript can act as a vehicle for the expression of the miR-17-92 cluster. Its predicted targets include the tumor suppressor genes PTEN and RB2 [146]. In a mouse B-cell lymphoma model, the overexpression of miR-17-92 accelerates tumor formation induced by the product of the MYC gene [165], which encodes an important transcription factor that regulates cell proliferation, growth, and apoptosis. Modification of MYC expression is a commonly observed deregulation leading to tumorigenesis. O’Donnell et al. showed that c-myc, through binding to sequences near the miR-17-92 cluster genomic locus, activates miR-17-92 expression [188]. Among the predicted targets of the miR-17-92 cluster is the transcription factor E2F1, which appears to be negatively regulated by miR-17-5p and miR-20a [146].
E2F1 is a cell cycle promoter induced upon c-myc expression. Conversely, c-myc expression is induced by E2F1 [189–191]. Therefore, a balance between the gene regulatory processes involving miRNAs and transcription factors may contribute to finely tune E2F1 expression and to generate a tightly controlled proliferative signal [188].

The let-7 miRNA also seems to be involved in cancer pathogenesis. Calin et al reported that let-7 genes are deleted in many cancers [185]. Moreover, a reduction in let-7 expression has been observed in samples of human lung cancers or cancer cell lines. Patients associated with a reduced let-7 expression had the worst prognosis after a potentially curative resection [192]. It is relevant to note that the overexpression of let-7 in lung cancer cells inhibits growth in vitro [192]. In C elegans, let-60, the ortholog of the human oncogene RAS, was found to contain eight putative let-7 binding sites in its 3′NTR [156]. As for the three human RAS genes, they also contain multiple let-7 binding sites, allowing let-7 to regulate RAS expression. Evidences of a downregulatory role for let-7 in RAS expression came from experiments using reporter genes fused to the 3′NTR of NRAS and KRAS [156]. Introduction of let-7a inhibitors relieved this repression [156]. In this study, miRNA microarray analyses of 21 different samples from lung cancer patients revealed that the expression of the let-7 gene family was reduced, in association with an increased expression of RAS protein. These data suggest a role for the let-7 miRNA family in the regulation of RAS during development of lung cancer [156].

Deregulated expression of protein components of the miRNA biosynthetic pathway may also be implicated in cancer formation. Karube et al [193] recently observed a diminution of Dicer expression in nonsmall cell lung cancer samples taken from 67 patients, as assessed by RT-PCR. As for the let-7 miRNA, this reduction was also associated with shorter postoperative survival [193]. Thus, the presence of Dicer, which mediates miRNA biogenesis, appears to be required for maintaining normal cell function.

miRNA expression profiling of the affected tissues may eventually be important for improving the diagnosis of diseases, such as cancer [194]. Using a new miRNA profiling method, Lu et al [195] analyzed mammalian miRNAs from 334 samples, including human cancers. They observed a characteristic general downregulation of miRNA expression in tumors, as compared to normal tissues [195]. Similarly, Jiang et al used an RT-PCR approach using primers specific to 222 pre-miRNAs to monitor their expression in human cancer cell lines [196]. Monitoring of global changes in miRNA expression profiles will be useful to establish possible links between miRNAs and diseases.

**miRNAs and viruses**

Several studies have reported a role for RNA silencing pathways in host defense mechanisms against viruses in plants [197], and reports suggest that they may also play a similar role in humans. The interaction between RNA silencing pathways and viruses, such as HIV-1, is complex and multifaceted [198]. Some viral RNAs exhibit secondary structures that are prone to Dicer processing, as evidenced by the discovery of miRNAs derived from Epstein-Barr virus (EBV), a virus belonging to the herpesvirus family, in infected Burkitt's lymphoma cells [199]. In this case, miRNAs originate from five precursors present in two different clusters of the genome of EBV [199]. The overall impact of viral miRNAs on cellular and viral gene expression remains to be fully appreciated. They may target and regulate specific human miRNAs, thereby ultimately influencing cell function and viral replication. Indeed, the potential host mRNA targets of these miRNAs, as predicted by bioinformatical analyses, are implicated in many biological processes, such as transcription, cell proliferation, apoptosis, B cell-specific chemokine and cytokine synthesis, and signal transduction [199]. These findings illustrate how a virus may exploit the RNA silencing machinery for its own purpose.

miRNAs derived from a virus may also be turned against some of its mRNAs, as exemplified by miR-BART2. This EBV miRNA has been shown to be perfectly complementary to the EBV gene BALF5, encoding for a DNA polymerase, and to target it for degradation [199, 200].

The pathogenic human Kaposi’s sarcoma-associated herpesvirus (KSHV) was recently shown to encode an array of 11 distinct miRNAs, all of which are expressed at readily detectable levels in latently KSHV infected cells [201]. Computer analysis of potential mRNA host targets for these viral miRNAs included several mRNAs previously shown to be downregulated in KSHV-infected cells, suggesting that KSHV miRNAs play a critical role in the establishment and/or maintenance of KSHV latent infection [201].

The genome of HIV-1 encodes a gene called nef, which is located in the 3′ region and is overlapping with the LTR. Omoto et al identified a nef-derived miRNA, called miR-N367, produced in cells persistently infected with HIV-1. This miRNA has been shown to downregulate the transcription of the HIV-1 genome in human T cells by targeting the nef-responsive element of its 5′ NTR U3 region and the nef sequence located in the 3′ NTR [202, 203]. HIV-1 was also found to generate a siRNA that can mediate nucleic-acid-based immunity and to encode a suppressor of RNA silencing in its Tat protein [204].

Additional evidences suggest that viruses have evolved to take advantage of RNA silencing pathways to enhance the probability of successful infection. For example, the simian virus 40 (SV40) genome was found to encode a pre-miRNA from which two miRNAs can be derived. Expressed at late times in infection, these miRNAs are perfectly complementary to the early viral mRNAs, and target those for degradation [205]. SV40-infected cells show a reduced expression of viral T antigens, are less sensitive to lysis by cytotoxic T cells, and trigger less cytokine production [205].

Cellular miRNAs may also play an important role in virus/host interactions. For example, miR-32 was found to restrict retrovirus primate foamy virus type 1 (PFV-1) accumulation in human cells [166]. However, PFV-1 may counteract this cellular restriction through expression of Tas, a protein inhibiting RNA silencing in mammalian cells [166]. Moreover, a study showed that the sequestration of miR-122, an mRNA highly and specifically expressed in the liver,
resulted in a marked loss of autonomously replicating hepatitis C virus (HCV) RNAs [167]. HCV replication thus appears to be facilitated by a genetic interaction between miR-122 and the 5′NTR of the HCV genome, making miR-122 a potential target for an anti-HCV intervention.

As for a possible regulation of HIV-1 replication by human miRNAs, computational predictions identified four possible HIV-1 targets: the nef gene targeted by miR-29a and miR-29b, the vpr gene targeted by miR-149, the vpu gene targeted by miR-378, and the vif gene targeted by miR-324-5p [168]. Microarray profiling confirmed the expression of these miRNAs in HIV-1 replication-competent human T lymphocytes [168]. Further investigation is required to determine the biological significance of these cellular miRNAs-HIV-1 interactions.

CONCLUSION

miRNAs are now recognized as key regulators of gene expression. Not surprisingly, causal links between deregulation of miRNA expression and some important genetic diseases are gradually emerging. A better characterization of the miRNA expression profiles observed in various clinical situations may ultimately be useful to physicians in providing signatures for specific tumors or infectious diseases. Further investigations that aim at elucidating and understanding the mechanisms involved in miRNA biosynthesis and function are crucial for the design and development of potentially important diagnostic tools and new therapeutic strategies.

ACKNOWLEDGMENTS

We are grateful to Gilles Chabot for graphic design. M. P. Perron and L.-A. Gobeil are supported by doctoral and master studentships from the Natural Sciences and Engineering Research Council of Canada (NSERC), respectively. P. Provost is a New Investigator of the Canadian Institutes of Health Research (CIHR) and The Arthritis Society. This work was supported by Discovery Grant 262938-03 from NSERC and Grant EOP-64706 from Health Canada/CIHR. D. L. Ouellet and M. P. Perron contributed equally to this work. P. Provost is the author for correspondence.

REFERENCES


[85] Song J-J, Liu J, Tolia NH, et al. The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in...


[138] Ekvall K, Nimmo ER, Lavenzar JP, et al. Mutations in the fission yeast silencing factors clr4+ and rik1+ disrupt the localization of the chromo domain protein Swi6p and impair cen-


