FEBS Letters 579 (2005) 5830–5840 FEBS 29898

Minireview

Identification and characterization of small RNAs involved in RNA silencing

Alexei Aravin, Thomas Tuschl*

Laboratory of RNA Molecular Biology, The Rockefeller University, New York, NY 10021, USA

Received 25 July 2005; accepted 11 August 2005

Available online 18 August 2005

Edited by Shou-Wei Ding

Abstract Double-stranded RNA (dsRNA) is a potent trigger of sequence-specific gene silencing mechanisms known as RNA silencing or RNA interference. The recognition of the target sequences is mediated by ribonucleoprotein complexes that contain 21- to 28-nucleotide (nt) guide RNAs derived from processing of the trigger dsRNA. Here, we review the experimental and bioinformatic approaches that were used to identify and characterize these small RNAs isolated from cells and tissues. The identification and characterization of small RNAs and their expression patterns is important for elucidating gene regulatory networks.

Keywords: MicroRNA; miRNA; Small interfering RNA; siRNA

1. Introduction

1.1. Discovery of small RNAs guiding RNA silencing processes

The first gene silencing process guided by small RNAs was discovered in 1993 in *Caenorhabditis elegans* [1,2]. It was shown that a 21-nucleotide (nt) processing product of the non-coding RNA transcript of the *lin-4* gene mediated repression of *lin-14* mRNA. The repression was dependent on partial sequence complementarity between the *lin-4* 21-nt RNA and the 3′ untranslated region (UTR) of *lin-14* mRNA. The *lin-4* mutant was initially discovered in 1981 in a screen for genes controlling the invariant cell lineages of *C. elegans* [3]. Several years later, another *C. elegans* 21-nt small RNA derived from the non-coding RNA transcript of *let-7* was shown to regulate *lin-41* mRNA [4]. lin-4 and let-7 small RNAs are unrelated in sequence, yet their primary transcripts share a common structural feature: the lin-4 and let-7 21-nt sequence is embedded in the stem of a 30-bp double-stranded RNA (dsRNA) hairpin. A double-stranded precursor suggested that lin-4 and let-7 feed into general gene silencing mechanism triggered by dsRNA in *C. elegans*. Because let-7 is evolutionary conserved in all bilateral symmetrical animals it was further suggested that this general silencing mechanism was also conserved [5]. More recent genetic screens have identified additional small RNA coding genes, such as *lsy-6* in *C. elegans* [6] and *bantam* in *Drosophila melanogaster* [7], controlling neuronal asymmetry and apoptosis, respectively. While genetic approaches provide not only insight into the nature of the gene product but also its function, other experimental and computational approaches have led to the identification of now hundreds of such small RNA coding genes, referred to as microRNAs (miRNAs) [8–10].

While small RNAs were studied in *C. elegans*, studies in plants on the mechanism of co-suppression, transcriptional and post-transcriptional gene silencing (TGS and PTGS, respectively), and virus-induced gene silencing were tracking down "aberrant" RNA as the trigger of these processes [11,12]. Filamentous fungi also showed similar mechanisms known as quelling, repeat-induced point mutation and meiotic silencing [13]. Related processes contributing to transposon and transgene silencing were also described in *D. melanogaster* [14,15]. Similar to miRNA-guided gene silencing, these silencing processes are evolutionary conserved and homology-dependent, i.e., require extensive sequence similarity between the trigger sequence (e.g., the transgene) and the targeted genes or gene products.

The molecular nature of the trigger of homology-dependent gene silencing was only revealed upon the discovery of the process of RNAi in *C. elegans* in 1998 [16]. Injected long dsRNA (hundreds of bps) induced degradation of mRNAs cognate to the dsRNA trigger that sometimes lasted for several generations. This method was shown to be effective in other invertebrate animals as well as in mouse oocytes and early embryos [17]. However, long dsRNA applied to differentiated somatic mammalian cells not only triggers RNAi but also a strong sequence-unspecific response causing global translational arrest and apoptosis [18]. The first report linking the production of small RNAs to homology-induced gene silencing in plants appeared in 1999 [19]. Small 25-nt sense and antisense RNAs were only found in transgenic plants that showed PTGS, suggesting that these small RNAs might function as specificity determinants of silencing. The initially reported 25-nt RNAs were later shown to represent two distinctly sized small RNA populations of 21- and 24-nt [20]. Analysis of the biochemical mechanism of RNAi in *D. melanogaster* embryo lysate revealed that long dsRNA is processed to 21-nt small RNAs [21]. After processing, the 21-nt small RNAs are base-paired with a 2-nt 3′ overhang and carry a

*Corresponding author.
E-mail addresses: aravina@rockefeller.edu (A. Aravin), ttuschl@mail.rockefeller.edu (T. Tuschl).

0014-5793/$30.00
doi:10.1016/j.febslet.2005.08.009

5′ phosphate and free 3′ hydroxyl [22]. Because such 21-nt base-paired processing intermediates were effectively targeting mRNAs for degradation, they were named small interfering RNAs (siRNAs). In order to characterize siRNAs that are processed from long dsRNA, the first small RNA cloning protocol was developed [22]. Cloning of the siRNAs obtained by processing of exogenously added dsRNA to *D. melanogaster* embryo lysate identified at the same time endogenous small RNAs, including novel miRNAs, and small RNAs cognate to transposable elements [8,22]. These initial observations triggered new experiments aimed at elucidating cellular sources of dsRNA-derived small RNAs and their role in gene regulation. Small RNAs of different origin and function were subsequently identified in different organisms raging from fission yeast to human (Table 1).

1.2. Different types of small RNAs

RNA silencing is an evolutionary ancient regulatory mechanism. Although its core protein machinery is widely conserved in eukaryotes, excluding the yeast *Saccharomyces cerevisiae*, different sources of dsRNA and functionally distinct RNA silencing effector complexes are encountered in different organisms. Effector complexes are composed of a single-stranded small RNA [23] directly bound to a member of the Argonaute (Ago) protein family [24–26]. Ago proteins contain an endoribonuclease domain [27]. However, only a subset of Ago proteins is catalytically active and responsible for mediating RNA cleavage [28,29]; cleavage-incompetent Ago members likely mediate other sequence-specific regulatory processes, such as transcriptional and translational silencing. The core effector complex may also be associated with auxiliary proteins [30].

1.2.1. miRNAs. miRNAs are the most abundant type of small RNAs in plants and animals (for recent review see [31,32]). miRNAs are excised from primary transcripts (pri-miRNAs) by two rounds of endoribonuclease III processing involving first Drosha, producing a hairpin-shaped pre-miRNA, and then Dicer (reviewed in [33]). After Dicer processing, miRNAs emerge as siRNA-duplex-like structures, but only one strand, the mature miRNA, is then predominantly incorporated into the Ago effector complexes. The discarded RNA strand is frequently referred to as miRNA* and is degraded. In plants, miRNAs guide mRNA cleavage and the highly complementary target sites are readily identified using bioinformatic tools [34]. In contrast, animal miRNAs preferentially target mRNAs at partially complementary yet evolutionary conserved sites, which are predominantly located within the 3′ UTR [35–37].

Table 1
Different types of endogenous small RNA

Class of small RNA	Size of mature form (nt)	Structure of precursor	Biogenesis	Mechanism of action	Organism
miRNA	20–23	Imperfect hairpin	Successive cleavage by Drosha and Dicer resulting in a mature form with defined sequence	Translational repression,[a] mRNA cleavage[b]	*C. elegans* [6,9,10,53,56,97] *D. melanogaster* [7,8,40,93] *X. laevis* [101] *D. rerio* [61,80,96] Mammals [8,37,59,61,94,95,100,102–109] Plants [110–116] Viruses [47,117,118]
rasiRNA	23–28	Long dsRNA	Processing of long dsRNA by Dicer resulting in multiple short RNAs[d]	Regulation of chromatin structure, transcriptional silencing[c]	*S. pombe* [38] *T. brucei* [44] *C. elegans* [58] *D. melanogaster* [15,40] *D. rerio* [96] *A. thaliana* [20,41]
Endogenous siRNA	20–23	Long dsRNA	Processing of long dsRNA by Dicer.[e] Biogenesis requires RdRP activity	mRNA cleavage	*C. elegans* [53] *A. thaliana* [51]
tncRNA	19–23	Unknown	Unknown. Mature forms have defined sequence[f]	Unknown	*C. elegans* [53]

[a]Majority of animal miRNAs.
[b]Majority of plant miRNAs.
[c]Shown in fission yeast and plants; indirect evidences for *D. melanogaster* and *C. elegans*.
[d]rasiRNA in *S. pombe* and plants processed from long dsRNA by Dicer. Size of rasiRNA in *D. melanogaster* (24-27 nt) does not correspond to the size of standard Dicer products (21-23 nt).
[e]In plants miRNA-directed cleavage of transcript produce register for subsequent Dicer processing resulting in defined sequences of siRNAs spaced 21 nt apart from each other.
[f]Processing of tiny non-coding RNA depends on Dicer but no hairpin precursor can be identified.