Distinct ribonucleoprotein reservoirs for microRNA and siRNA populations in C. elegans

SAM G. GU,1 JULIA PAK,2,3 SERGIO BARBERAN-SOLER,1 MUSTAPHA ALI,1 ANDREW FIRE,2,3 and ALAN M. ZAHLER1

1Department of Molecular Cell and Developmental Biology and Center for Molecular Biology of RNA, University of California Santa Cruz, Santa Cruz, California 95064, USA
2Department of Pathology and Genetics, Stanford University School of Medicine, Stanford, California 94305, USA
3Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA

ABSTRACT

MicroRNAs (miRNAs) are regulatory molecules that share both biosynthetic derivation (cleavage from short hairpin precursor RNAs) and functional roles (downregulation of specific mRNAs through targeted degradation and/or translational inhibition). A distinct family of small RNAs, termed siRNAs, have some common characteristics but exhibit distinct modes of biosynthesis and function. In this study, we report procedures for purification of a predominant species of miRNA-containing ribonucleoprotein complexes from Caenorhabditis elegans and demonstrate that this population is distinct from the predominant pool of siRNA-containing ribonucleoprotein complexes. An observed miRNP-associated RNA population consisting predominantly (>95%) of miRNAs supported the unique identity of miRNPs as biological effectors within the cell, provided clean material for analysis of changes in miRNA spectra during development, and provided strong evidence of miRNA character for a number of novel small RNAs. Likewise, the RNA spectrum derived from partial siRNP purification was useful in defining functional characteristics of this more diverse population of small RNAs.

Keywords: miRNA; siRNA; miRNP purification; microarray; small RNA cloning

INTRODUCTION

MicroRNAs (miRNAs) are a class of ~22 nucleotide (nt) RNA molecules that target specific messenger RNAs (mRNAs) for silencing at the post-transcriptional level (for reviews, see Ambros 2004; Bartel 2004). Since the initial discoveries of lin-4 and let-7, two miRNAs that regulate developmental timing in Caenorhabditis elegans (Lee et al. 1993; Wightman et al. 1993; Reinhart et al. 2000; Slack et al. 2000), hundreds of miRNA genes have been identified in animals and plants. miRNA expression is often dynamically regulated, consistent with the role of these genes in cell fate determination and developmental timing (for reviewed, see Plasterk 2006).

As a defining feature of the miRNA gene family, the mature miRNA sequences reside on either side of the stem–loop structures formed in primary miRNA transcripts (Lau et al. 2001; Lee and Ambros 2001). The current model of miRNA maturation involves two RNA-processing steps, each catalyzed by a different RNase III-like enzyme (Hutvagner et al. 2001; Ketting et al. 2001; Han et al. 2006). The primary miRNA transcripts are first processed by the DROSHA–nuclease complex, resulting in the release of stem–loop structures called precursor miRNAs (pre-miRNAs) (Han et al. 2006). Pre-miRNAs are then further processed by the DICER nuclease into ~22 nt duplexes that carry two nucleotide 3' overhangs (Grishok et al. 2001; Hutvagner et al. 2001; Ketting et al. 2001). Usually, only one strand is selected as the mature miRNA and assembled into effector complexes, referred to as miRNA-associated ribonucleoprotein complexes (miRNPs). Members of the argonaute family of proteins have been shown to be important for miRNA accumulation and function (Grishok et al. 2001). The exact mechanisms of gene silencing by miRNPs are still unclear. Previous studies have provided evidence supporting different mechanisms, including inhibition of mRNA translational initiation or post-initiation steps (Wightman et al. 1993; Olsen and Ambros 1999; Humphreys et al. 2005; Pillai et al. 2005; Petersen et al. 2006), as well as promoting mRNA deadenylation and...
degradation (Bagga et al. 2005; Lim et al. 2005; Liu et al. 2005; Giraldez et al. 2006; Wu et al. 2006). These different studies suggest that miRNAs in animals may mediate silencing through a variety of intertwined mechanisms.

In order to characterize the components and activities of miRNPs, we developed a chromatographic procedure to enrich for endogenous miRNPs from *C. elegans* to high purity. This procedure yielded a fraction of ribonucleoprotein complexes that uniquely house miRNAs. *C. elegans* has also been a valuable model organism for the studies of other small noncoding RNAs, including, most notably, the short interfering RNAs (siRNAs) (e.g., Pak and Fire 2007; Sijen et al. 2007). Although both populations require the DICER nuclease, miRNAs and siRNAs differ in some of their biogenesis pathways and in their gene-silencing mechanisms (Grishok et al. 2001; Duchaine et al. 2006). miRNA and siRNA complexes have recently been shown to be separable by size exclusion chromatography in *C. elegans* (Sijen et al. 2007). In this study, we demonstrate biochemical purification and analysis of the RNA components of the distinct complexes containing miRNA and siRNAs.

**RESULTS**

**A unique miRNP complex in *C. elegans***

We began our analysis of miRNA complexes in *C. elegans* with a candidate-independent approach. Crude extracts derived from mixed-stage populations were first subjected to a high salt extraction (500 mM KCl), which promotes dissociation of miRNPs from polysomes (Caudy et al. 2003). miRNAs remained in the supernatant after depletion of ribosomes by ultracentrifugation, as determined by Northern blots (data not shown). After dialysis, the extract was fractionated sequentially by anion exchange, cation exchange, and size-exclusion chromatography. After each purification step, all of the major miRNA-containing fractions, determined by Northern blotting with probes to miR-35, miR-71, miR-1, or lin-4, were pooled and further fractionated at the next step. miRNP complexes tracked by such methods could bind to both anion and cation exchange columns under mild salt conditions (Fig. 1A,B). Size-exclusion chromatography indicated that miRNPs have sizes ranging from \(~100\) kDa to \(~500\) kDa, with a peak in the 160–200 kDa fractions (Fig. 1C; fractions B2, B1, and C1).

We combined the sizing column fractions enriched in miRNAs, the B2, B1, and C1 fractions, referred to as miRNP-200, for further analysis (Fig. 1C). Fractions B2, B1, and C1 share common protein profiles by SDS polyacrylamide gel electrophoresis (SDS-PAGE) analysis (Supplemental Fig. S1). With this minimal number of proteins, individual protein bands for miRNP-200 were individually excised from the SDS-PAGE and analyzed by liquid chromatography coupled tandem mass spectrometry (LC-MS/MS) (see Supplemental Fig. S1 for a complete list; Supplemental Table S1). Selective purification of at least one major component of this complex, tudor staphylococcal nuclease homolog (TSN-1) (Caudy et al. 2003), in the miRNP200 fraction has been confirmed with anti-TSN-1 antisera (data not shown). Although signatures with potential matches to other proteins of interest (e.g., ALG-1 [Argonaute-Like Gene], and ALG-2), (Grishok et al. 2001) were also observed in the mass spectrometry data, further purification will be required to evaluate these assignments. Due to the nonquantitative nature of the LC-MS/MS analysis, we are unable to determine the relative abundance of proteins in miRNP-200 or whether one or several miRNP complex comprises the major species.

**miRNPs can assemble on specific substrate RNAs***

As a measure of the functionality of the purified miRNPs, we asked whether they could assemble onto partially complementary target RNA sequences while being held in a ribonucleoprotein complex, a predicted function of miRNPs. In order to do this, we immobilized an RNA containing two of the confirmed target binding sites for lin-4 from the 3′-UTR of lin-14 (Lee et al. 1993). In vitro-transcribed RNA was covalently linked by its 3′ end to adipic acid dihydrazide agarose beads (Caputi et al. 1999). Incubation of these RNA-coated beads with miRNP fractions shows that the *lin-4* miRNP specifically assembles onto these beads, and can be depleted from the extract, while *mir-35* does not appear to interact with the *lin-14* 3′-UTR substrate (Fig. 2A). We then tested the assembly...
of miRNPs onto a different substrate, one containing a binding site for mir-58 (58S) that we designed based on a prediction of potential miRNA–mRNA binding sites. We compared the binding of miRNPs to the 58S substrate versus an immobilized control RNA substrate (C) containing a run of 18 adenosines (Fig. 2B). mir-58 assembled onto this substrate with high efficiency, and showed depletion from the flow-through. Quantitation with a PhosphorImager revealed that mir-58 bound to the 58S substrate 9.2 times more efficiently than to the C substrate. In contrast, mir-35 and lin-4 showed no evident depletion from the column flow-through, and only showed a minor preference for binding to the longer 58S substrate over the C substrate: 1.6- and 1.9-fold, respectively. From these results we conclude that miRNAs in the purified miRNP complexes retain the ability to bind specifically to partially complementary substrate RNAs.

Characterization of miRNP-associated small RNAs

To characterize the RNA component of the purification product, we extracted total RNAs from the miRNP-200 fractions. Following alkaline phosphatase treatment, the RNAs were 5' end-labeled with \(^{32}\)P using T4 polynucleotide kinase (PNK) and then resolved by denaturing PAGE. Signals detected on the autoradiograph were exclusively in the 21–23-nt range, consistent with miRNA character (Fig. 1D).

We then constructed a cDNA library of the miRNP-200-associated small RNAs using a 5' ligation-dependent cDNA cloning protocol as described (Lau et al. 2001). A total of 1093 sequences were obtained from sequencing of these clones. Among the sequenced clones in the cDNA library, 79% are in the size range of 21–24 nt and 1040 (95%) match known miRNAs (Table 1; Supplemental Fig. S2). Since we collected fractions containing the majority of miRNAs at each of the chromatographic steps, the sequencing data indicate that the purification procedure efficiently enriches miRNPs as the predominant, if not exclusive, RNP complexes.

From the sequenced cDNA library, we detected 63 out of the 114 C. elegans miRNAs that are registered at miRBase (version 9.0) (Griffiths-Jones et al. 2006), a miRNA database. In addition, we also detected 5 of the 18 new miRNAs that were only recently discovered in an extensive pyrosequencing study of \(~400,000\) size-selected RNAs from C. elegans (Supplemental Fig. S2) (Ruby et al. 2006). Two different sequences in our collection, both 23-nt long, do not match any known miRNAs.
genomic sequences are extended at both ends, both can generate stem–loop RNA structures as predicted by M-fold (Zuker 2003), a characteristic feature of miRNA genes (Fig. 3A). These two small RNAs satisfy some of the criteria for classification as miRNA genes, including their detection in a small RNA cloning procedure and a hairpin RNA secondary structure for their putative precursor. The exclusiveness of miRNAs in miRNP-200 further increases the likelihood of the two being miRNAs. We were unable to confirm these two putative miRNAs by Northern blotting nor could we detect four of the five microRNAs that we detected which were recently reported by Ruby et al. (2006) (data not shown). Given that a known miRNA, lsy-6, is expressed in a small number of cells (Johnston and Hobert 2003) and its abundance is so low that it could not be detected by pyrosequencing of 400,000 small RNAs from C. elegans (Ruby et al. 2006), lack of a Northern blot signal should not of itself invalidate these as miRNAs. Since these two miRNAs were found in a sequencing effort from purified miRNPs and are found in predicted hairpin structures from the genome, these criteria may be enough to validate these as miRNAs. Therefore, we have assigned them as new C. elegans miRNAs, cel-mir-1021 and cel-mir-1022.

miRNP versus siRNP

Previous genetic analyses suggested that RNAi and miRNA pathways in C. elegans involve different silencing mechanisms and can assemble into distinct complexes through interactions with the DICER protein (Grishok et al. 2001; Duchaine et al. 2006). This implies that the endogenous miRNAs and siRNAs exist in distinct sets of functional complexes. Consistent with this model, there are only 12 sequences out of 1093 that match antisense exonic sequences and therefore represent possible endogenous siRNAs in the sequenced miRNP-200 cDNA library. We next set out to test whether miRNPs and siRNA complexes could be separated chromatographically. Crude extract, prepared under mild salt condition (50 mM KCl) from worms undergoing feeding RNAi against the sel-1 gene (Timmons et al. 2001; Duchaine et al. 2006; Pak and Fire 2007), was fractionated using either an anion exchange column or a sizing column. All the fractions were sequentially probed for siRNAs and miRNAs by Northern blotting analysis. In the anion exchange FPLC, siRNPs eluted in a distinct peak, at a lower salt concentration than miRNPs, suggesting that siRNPs are less negatively charged on their surface than miRNPs (Fig. 4A). The difference in their sizes is more prominent: siRNPs eluted as a complex with an apparent molecular weight of ~60 kDa in the size-exclusion chromatography, while miRNPs peaked in the 160–200 kDa fractions (Fig. 4B). These different FPLC profiles indicate that miRNAs and siRNAs are assembled into distinct reservoirs in the cell.

We then asked if siRNA-containing complexes could also be chromatographically separated. Similar to enriching for miRNAs, we fractionated the extract made from worms undergoing RNAi against sel-1 by anion exchange, cation exchange, and size-exclusion chromatography. After each purification step, all of the major sel-1 siRNA-containing fractions, determined by Northern blotting, were pooled and further fractionated at the next step (data not shown). We purified small RNAs from the final siRNA-enriched fractions and constructed cDNA libraries using two different methods. One is the 5’ ligation-dependent protocol that
is specifically capable of cloning small RNAs with 5’ monophosphate (used to clone miRNP-200 small RNAs; see above). The other is a 5’ ligation-independent cloning protocol that is capable of cloning siRNAs with 5’ di- or triphosphate, as well as miRNAs (Pak and Fire 2007; Sijen et al. 2007). When comparing results from different cDNA sources but with the same cloning method (5’ ligation-dependent), we noticed a much lower percentage of miRNA clones relative to siRNAs (antisense to mRNAs) identified in the siRNP fractions (7.5% miRNAs versus 43.8% siRNAs) than for the miRNP-200 fractions (95.2% miRNAs versus 1.1% siRNAs) (Table 2). This is consistent with miRNP-200 being highly enriched for miRNAs specifically and miRNPs being chromatographically distinct from the siRNPs.

We obtained a similar fraction of siRNA clones from the purified siRNPs using the two different cloning methods (Table 2). The cDNA library generated by using the 5’ ligation-independent method yielded more sequences that match the *C. elegans* genome (Table 2), and therefore was used for subsequent analysis. In examining the small RNA clones from the siRNP fractions, we first noticed that the most abundant class of RNAs was that of ribosomal RNAs (rRNAs) (42%) (Table 1). A priori, these clones may represent either small RNAs involved in regulation of the rRNA genes or may be the result of degradation of the highly abundant rRNAs. Two predictions follow from the first possibility (Ruby et al. 2006). First, the small RNAs should be oriented antisense to the rRNAs. When we aligned a subset of these sequences against the rRNA gene cluster on chromosome I, we found, in fact, that all clones corresponded to sense sequence. Interestingly, certain sequences appear overrepresented (Supplemental Fig. 3A). We speculate that aspects of rRNA secondary structure may differentially expose these sequences for cleavage. The second prediction is that the sizes of these small RNAs should reflect the precision observed for either Dicer-mediated cleavage or RdRP-mediated RNA synthesis. We observed that the rRNA fragments do not exhibit an obvious preference for any length (Supplemental Fig. 3B). We, therefore, conclude that these sequences are a result of rRNA degradation. In addition, their apparent greater abundance in the siRNP-containing fractions may be due to a general deenrichment in other types of small RNAs in these fractions as cloning methods can only distinguish relative amounts of different small RNA species rather than their absolute levels.

The second most abundant class we observed corresponded to sequence that is antisense to mRNAs (34%; Table 1). For convenience, we designate these as endogenous siRNAs (Lee et al. 2006). *sel-1* siRNAs comprised only 1% of the total sequences. Even when we disregard the rRNA sequences (given that their presence in the sequencing pool probably does not reflect that among the endogenous small RNA pool), *sel-1* siRNAs still only comprise 1.3% of small RNAs in the siRNPs, while the revised endogenous siRNA percentage is 43.3%. We do not observe a significant enrichment in these species over those observed from small RNA extracts (Pak and Fire 2007). It is possible that a small depletion of miRNAs and other small RNAs in siRNPs may not be noticeable in the relatively large number of siRNAs.

**TABLE 2.** Relative amounts of miRNAs and siRNAs identified in different cloning experiments from different sources

<table>
<thead>
<tr>
<th>Source</th>
<th>Unenriched small RNAs</th>
<th>miRNPs</th>
<th>siRNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloning method (5’ ligation-dependent or independent)</td>
<td>Dependent</td>
<td>Independent</td>
<td>Dependent</td>
</tr>
<tr>
<td>miRNAs (%)</td>
<td>64.6</td>
<td>8.2</td>
<td>95.2</td>
</tr>
<tr>
<td>Endogenous siRNAs (%)</td>
<td>10.6</td>
<td>37.6</td>
<td>1.1</td>
</tr>
</tbody>
</table>

Unenriched small RNAs are from Pak and Fire (2007). Total number of miRNP clones is 1093, total number of siRNP 5’ ligation-dependent clones is 92, and total number of siRNP 5’ ligation-independent clones is 340.
We have performed a cursory analysis of the endogenous siRNAs detected in the siRNPs (Supplemental Table S2). Although we obtained 14 examples of single genes which correspond to multiple siRNA clones (e.g., Y52B11A.gc10), the rich complexity of the endogenous siRNA pool precludes an analysis of their relative abundances with this small number of sequences. However, multiple siRNA clones corresponding to a single gene is consistent with the existence of a specific set of genes that are subject to endogenous RNAi (Ambros et al. 2003). An analysis of the 5’-most nucleotide of the siRNAs revealed a preference for guanosine (4% C, 13% A, 37% T, and 61% G) as noted previously (Ambros et al. 2003; Ruby et al. 2006). There seems to be no strong bias for any particular chromosome or chromosomal subregion as a particular source of these endogenous siRNAs (data not shown).

Microarray analysis of miRNA expression during development

To study the expression of miRNAs from *C. elegans* on a genome-wide level, we developed a microarray platform, which contains DNA oligonucleotide probes for all 114 *C. elegans* mature miRNAs that have been reported to miRBase (Release 9.0) (Griffiths-Jones et al. 2006). Trials using total small RNAs yielded high background and noisy signals for the hybridization (data not shown), possibly due to the crosshybridization contributed by non-miRNA small RNAs. We asked if using miRNAs purified from the miRNP-enriched fractions could improve the quality of the array signals. To test this in the developmental context, we enriched miRNPs from *C. elegans* hermaphrodites synchronized at six different developmental stages (embryos, L1–L4 larval stages, and early adult). The miRNP purification procedure was simplified for the array study by not including the final size-exclusion chromatography column. The miRNPs enriched by sequential anion and cation FPLC columns contained RNAs that are homogeneously in the size range of 21–23 nt (Supplemental Fig. S4). Purified miRNAs from each stage were labeled using Cy3 and hybridized onto the microarray together with Cy5-labeled reference RNA, which was a mixture of equal amounts of miRNAs from miRNPs isolated from all six developmental stages. A 3-nt difference between probe and target feature will lead to a loss of hybridization signal in this study, and greater than a 1-nt difference can lead to a significant reduction in signal according to control DNA oligos on the array (data not shown). This helps to ensure that miRNAs belonging to the same general family crosshybridize at most minimally to the features for each other on the array. Gratifyingly, our array measurements accurately recapitulated the temporal expression features of *lin-4* and *let-7*, two well-characterized miRNAs whose expression is initiated at the L1–L2 and L2–L3 transitions, respectively (Fig. 5A; Lee et al. 1993; Wightman et al. 1993; Reinhart et al. 2000).

For 81 miRNAs, we detected hybridization signals that were above the cutoff intensity as defined by the average intensity of the control spots plus two standard deviations (see Materials and Methods). We then compared the signal intensities of the array with the relative abundance as indicated by the large-scale miRNA sequencing (Ruby et al. 2006). We noticed that among the 33 miRNAs whose array intensities are below the cutoff value, 32 are included in the 48 low-abundance miRNAs, each represented by less than 0.088% of the total clones, or more than 10-fold less than expected if all miRNAs had been equally represented in the library (1/114) (Ruby et al. 2006). Therefore, the low intensities of these 33 miRNAs correlate strongly with their low abundance in vivo. For the other 16 low-abundance miRNAs, some of which were not even detected in the large-scale sequencing, their array signals above background may be due to crosshybridization with abundant miRNAs. Therefore, we only include the array data for the remaining 65 miRNAs (Fig. 5A) in the high-confidence data set for the rest of this work. We also validated the microarray data by probing Northern blots of the same miRNA preparations used to probe the arrays, with 12 different miRNA probes (Supplemental Fig. S5). Results obtained by the two methods of detection were generally consistent for the developmental time course, especially in regard to those miRNAs that the microarray predicted to show greater than fourfold changes during development (for example, *lin-4*, *mir-35*, *mir-40*, *mir-71*, *mir-80*, *mir-237*).

To survey the extent of expression changes during development, we examined the difference between the highest and lowest expression levels for each miRNA. Among the 65 miRNAs in this analysis, 40 show expression differences greater than fourfold across the different developmental stages. These include all the heterochronic *lin-4* and *let-7* family members. The expression of some miRNAs changes as much as 128-fold during development. These results indicate that, for the majority of miRNAs, expression is variable and dependent on developmental stage.

To identify any generalized expression patterns, developmental expression profiles of the 40 miRNAs whose expression levels vary greater than fourfold during development were analyzed using a hierarchical clustering method (Eisen et al. 1998). These miRNAs are clustered into two groups (Fig. 5B). In one cluster of 19 miRNAs, expression is initiated at one of the larval stages and continues stably until the adult stage (Fig. 5B). A closer examination indicated that the onset of expression for these miRNAs mostly occurs at the L2–L4 stages, with *mir-71* being the only one whose expression is strongly upregulated by L1. In the other group (21 miRNAs), expression is restricted to embryonic and L1 stages, and repressed at later larval stages (Fig. 5B). Some miRNAs in the second group show mild expression at the adult stage possibly due to the presence of some early embryos in the gonads of a fraction of the young adults. Intriguingly, we did not observe any
examples of miRNAs whose predominant expression was at the young adult stage.

Seventeen groups of miRNAs are encoded as clusters in the *C. elegans* genome (Ruby et al. 2006). Many miRNAs in a cluster share the same seed region and, therefore, could be capable of downregulating the same set of target genes (Bartel 2004). These miRNAs were suggested to be coexpressed (Lau et al. 2001). Our high-confidence data set provides quantitative developmental expression profiling for eight such clusters (Fig. 6A). miRNAs in each of seven clusters (miR-35/36/37/38/40/41, miR-42/43/44, miR-48/241, miR-51/53, miR-54/55/56, miR-64/65/66/229, and miR-73/74) have very similar expression profiles within the cluster (Fig. 6A). However, we observed distinct temporal expression for miR-2 and miR-71, which have different sequences in their seed regions and are encoded 7 kb away from each other in the genome (Fig. 6A).

miRNAs that have same seed region at the 5’ end can also be assigned to groups. Currently, 24 families are assigned for a total of 78 miRNAs in *C. elegans* (Ruby et al. 2006). Our data set covers 55 of these miRNAs in 13 families (Fig. 6B). For all of these families, we observed a strong tendency for coexpression during development for miRNAs in the same family. These data strongly suggest that transcriptional coregulation mechanisms exist to ensure that potentially redundant miRNAs have similar developmental expression profiles even when not cistronic. For example, miR-51 through miR-56 all share the same seed region and developmental expression profile, yet they are encoded in three different regions of the genome. There are a small number of exceptions to this phenomenon. For example miR-229,
miR-84, mir-72, and miR-79 have different developmental expression profiles, or at least do not demonstrate the same level of developmental temporal regulation, as other family members sharing the same seed region.

**DISCUSSION**

In this study, we purified miRNP complexes that are stable through an initial high salt (0.5 M KCl) extraction step and three different FPLC fractionation steps. miRNPs that were size-fractionated immediately after the crude extract was made using milder salt conditions showed similar behavior during column chromatography (Fig. 4), indicating that the initial salt extraction did not disrupt the miRNP structure other than helping to release the miRNPs from polysome fractions. High-level enrichment of miRNPs was achieved due to their shared bulk properties and affinity for both the anion and cation exchange FPLC columns.
At these two steps, the miRNPs were collected in continuous narrow-ranged fractions that were eluted under mild salt conditions. The miRNP sizes peaked in the 160–200-kDa range as determined by size-exclusion chromatography. Proteins present in these fractions (miRNP-200) include at least one previously reported miRNP component (TSN-1) (Caudy et al. 2003). Other protein bands were also observed in miRNP-200; thus, a definitive assignment of protein composition for the complex will await additional fractionation. It is likely that the miRNPs can be further enriched because the components identified together have a molecular weight much greater than the apparent molecular weight of the miRNPs. Although the miRNP complexes identified in this analysis could certainly function directly as effector miRNPs in trans-acting regulation, it is also possible that the purification product includes distinct miRNA-associated complexes that are involved in distinct miRNA-dependent activities (biogenesis, processing, assembly, and/or turnover). The labeling of RNAs associated with miRNP-200 (Fig. 1D) indicates that there are no other major classes of ribonucleoproteins present in these fractions besides miRNPs.

We found that RNAs associated with the miRNP-200 fractions were homogenous in size and consisted mostly of the known mature miRNAs from *C. elegans*. Using miRNPs as a source of miRNAs provided substantial enrichment when compared to size selection of small RNAs from total RNA. Depending on the experiment, non-miRNAs accounted for 16%–62% of ~22-nt size-selected RNA clones from total RNA (Pak and Fire 2007) compared to only <5% from miRNP-200-associated RNAs. We obtained very similar profiles from two sequencing experiments using the same miRNAs from the purified mixed-stage miRNPs (data not shown), indicating the scale of the sequencing experiments provides reproducible if not necessarily linear representations of the underlying RNA population. For the sequenced miRNP-200-associated miRNAs, the clone numbers of individual miRNA species ranged from zero for 64 different miRNAs, including the 19 miRNAs that were not detected in the large-scale sequencing study (Ruby et al. 2006), to 186 for miR-58 (19.6% of the total clones) (Supplemental Fig. S2; Fig. 3B, relative frequency of clones of the top 20 most prevalent miRNAs in miRNP-200). In general, the relative abundance of the miRNAs cloned from the miRNP-200 complex closely parallels the relative miRNA abundance indicated by the pyrosequencing of ~400,000 size-selected miRNAs (Ruby et al. 2006). Our miRNP-200 cloning results provide evidence that at least 68 known miRNAs are assembled into the miRNP-200 complexes, and with similar efficiency to their overall abundance.

In previous studies where size-selected total small RNAs were used for cloning and sequencing, sequences opposite the mature miRNAs on the hairpin precursors, the miRNA*’s, were sometimes detected, although at much lower frequencies than their corresponding mature miRNAs (Ruby et al. 2006). In two studies (one covered ~3400 miRNA-matching sequences (Lim et al. 2003) and ~310,000 for the other (Ruby et al. 2006), the average ratios of miRNA* to miRNA were both found to be ~1%. In our study, we only observed two alignments that matched any miRNA*, namely, cel-mir-2* and cel-mir-239b*, out of 1040 sequenced miRNA clones. The rarity of miRNA* sequences in miRNP-200 supports the idea that the miRNP assembly machinery is highly precise in determining that only one strand of the double-stranded dicer product is loaded into a miRNP complex.

In this study we provide biochemical evidence that the endogenous siRNAs and miRNAs exist in distinct complexes. This is consistent with previous genetic analyses suggesting that RNAi and miRNA pathways use different sets of argonaute proteins (Grishok et al. 2001; Duchaine et al. 2006). Sijen et al. (2007) recently demonstrated that *C. elegans* miRNAs and siRNAs are separable by gel exclusion chromatography of cell extracts. They report that siRNAs exist in RNP complexes of 100–150 kDa while let-7 miRNA was detected in a larger peak centered at 400 kDa. These results are consistent with the separation that we report here, although we detected siRNAs peaking at 67 kDa and miRNAs at 200 kDa. The discrepancy in the sizes of the miRNA and siRNA complexes that these two studies report may reflect differences in purification protocols; for example we use a 0.5 M salt extraction step in our purification that they do not, and they are loading a cruder extract on their columns. An alternative explanation for the size differences is that determination of the complexes is approximate, and it may be difficult to compare results using different standards on different columns. However, both studies conclude that miRNPs are distinct from and larger than siRNPs.

In previous studies using mammalian and fly systems, miRNAs were programmed into RNA-induced silencing complexes (RISCs) with RNAi activity in vitro by using either hairpin miRNA precursors or short RNA duplexes (Hutvagner et al. 2001). This raised the possibility that the specificity of the miRNP assembly pathway is less pronounced in the in vitro systems and that the mature miRNAs derived from the synthetic miRNA precursors were assembled into siRNPs as a result. Alternatively, it is possible that, unlike their counterparts in *C. elegans*, the miRNPs and siRNPs from mammals and flies overlap in regard to the assembly pathway or effector components. It is possible that miRNAs can be loaded into several different types of effector complexes that exist at different concentrations. From this perspective, it is intriguing that the worm siRNPs were fractionated in a narrow range by both the anion exchange and the sizing FPLC, suggesting that the siRNAs were predominantly assembled into a similar family of RNP complex. Furthermore, the presence of both endogenous and exogenously-triggered siRNAs in
the same fraction suggests similar complexes could be involved in both pathways. Some divergence in genetic requirements between these two pathways has been noted (Duchaine et al. 2006; Lee et al. 2006; Sijen et al. 2007). The sizing FPLC results indicate that the siRNPs have an apparent size of ~60 kDa, which limits the number of the core protein components of siRNP. We are currently in the process of purifying the endogenous worm siRNPs.

Gene expression profiling during animal development can provide important information for the analysis of gene function and coordinate regulation. A previous study used Northern blotting to probe the expression of many of the C. elegans miRNAs during different developmental stages (Lim et al. 2003). The investigators found that a large number of the miRNAs in this organism have their expression developmentally regulated. Based on the Northern blots of Lim et al., expression of miRNAs was only described as on or off at the different developmental stages. The chromatographically purified miRNPs provide somewhat more quantitative measurements for the changes in expression of miRNAs during C. elegans development. In this study, we were able to detect temporal expression for more than half of the miRNAs for which analysis was possible. Most of these developmentally regulated miRNAs clearly fall into two groups: embryo-enriched or larval-enriched. Among the 40 miRNAs with greater than fourfold changes during development, we found 21 enriched in embryos but not in later larval stages. Sixteen of the embryo-specific miRNAs belong to three families: the mir-35 family (eight miRNAs), the miR-42 family (two miRNAs), and the miR-51 family (six miRNAs) (Fig. 5B). These embryo-specific miRNAs may function to silence both zygotic transcripts and unwanted maternally loaded mRNAs, as has been demonstrated for miR-430 in zebrafish (Giraldez et al. 2006). For the other group of developmentally regulated miRNAs, expression is initiated at larval stages and continues to the adult stage. Nineteen miRNAs that have over fourfold changes during development belong to this group. Some miRNAs are certainly involved in temporal patterning during cell lineage specification, including the well-studied lin-4 and let-7 miRNAs, which causes reiteration of L1 or L4 stage developmental patterns, respectively (Lee et al. 1993; Wightman et al. 1993; Reinhart et al. 2000). Others may be tissue-specific, with their distribution following patterns of developmental events in the relevant tissue as a fraction or total tissue in the animal. Since the temporal features of many important model developmental events, such as neural, vulval, and germ line determination, have been well characterized with regard to larval development, and there is expression data for almost every mRNA during larval development (Jiang et al. 2001), this study can provide useful information for approaches to identify candidate target mRNAs for these temporally regulated miRNAs.

To summarize, our study has introduced a procedure to enrich miRNPs through a fractionation protocol that is dependent on following miRNAs, not any candidate protein. Our analysis of the resulting RNA population suggests that assembly into corresponding complexes is indicative of miRNA character. Given the complexity of the different small RNAs regarding their biogenesis, processing, and function, this work addresses the importance and feasibility of developing an in vitro system to study the mechanisms of miRNA-dependent gene silencing and for learning more about the nature of endogenous siRNA targets.

MATERIALS AND METHODS

miRNA purification

Large quantities of mixed-stage or synchronized C. elegans N2 strain were grown and harvested as described (McGhee 1999). Animals were ground into powder in a mortar and pestle under liquid nitrogen and stored at −80°C. To make the crude extract, 20 g of the powder were resuspended in 60 mL buffer A (50 mM HEPES-NaOH, pH 7.5, 50 mM KCl, 5 mM MgCl₂, 1 mM DTT, and 0.5 mM PMSF) and then homogenized in a Dounce homogenizer with 15 strokes. After insoluble material was removed by centrifugation in a JA-20 rotor (Beckman) at 27,000 g for 15 min, the crude extract was stirred in the cold room for 1 h, with slow addition over a 10-min period of 4 M KCl to a final concentration of 500 mM. The crude extract was then centrifuged in a Ti-60 rotor (Beckman) at 100,000 g for 1 h. The supernatant was dialyzed in buffer A and then fractionated through a Mono Q 10/100 GL column (Amersham Biosciences) with a gradient of 0.05–2 M KCl on the AKTA FPLC system (Amersham Biosciences). miRNA-containing fractions, as determined by Northern blotting, were pooled together and dialyzed against 20 mM HEPES-NaOH (pH 7.5)/1 mM DTT. The extract was further fractionated through a Mono S 5/50 GL (Amersham Biosciences) column with a gradient of 0–1 M KCl. miRNA-containing fractions were pooled together and fractionated using a Superdex 200 column (Amersham Biosciences). The peak miRNA-containing fractions (B2, B1, and C1) were pooled and dialyzed against buffer D (50 mM HEPES-NaOH [pH 7.5], 140 mM KCl, and 5 mM MgCl₂). After addition of glycerol to a final concentration of 20% v/v, the dialyzed samples were flash frozen in liquid N₂ and stored at −80°C.

Fractionation of siRNP complexes

Wild-type N2 animals were grown on large NGM plates containing 25 µg/mL carbenicillin and 232.5 µg/mL IPTG and seeded with HT115 bacteria containing the plP12.2+ plasmid (containing the sel-1 trigger sequence) as a food source. plP12.2+ is a derivative of the C. elegans dsRNA feeding vector L4440 carrying a 458-bp segment of the sel-1 coding sequence (TGCA TGGAGCGG ATCGGA AAATGGCGGATGCGCA). Following 7–10 d at 16°C, C. elegans populations were washed free of bacteria in 5% sucrose, 50 mM NaCl, 5 mM EDTA (pH 8), and frozen as pellets in liquid nitrogen. Animals were ground into powder in a mortar and pestle under liquid nitrogen and then resuspended in buffer A (50 mM HEPES-NaOH [pH 7.5], 50 mM KCl, 5 mM MgCl₂, 1 mM DTT, and 0.5 mM PMSF). Extract was
prepared as described for miRNPs purification but without the initial high salt extraction. siRNPs were fractionated across Mono Q, Mono S, and Superdex 200 columns prior to RNA extraction from the siRNA-containing fractions.

**Protein identification**

Data for protein identification was acquired at the Vincent Coates Foundation Mass Spectrometry Laboratory, Stanford University Mass Spectrometry. Proteins in miRNP-200, made from the mixed-stage animals, were resolved on a 10% SDS-PAGE. Protein bands were excised, followed by in-gel trypsin digestion. Trypsin-generated peptides were analyzed by liquid chromatography-coupled tandem mass spectrometry (LC-MS/MS). Proteins were identified by comparing the mass spectrometry data against the NCBInr protein database through the Mascot search engine (http://www.matrixscience.com/).

**miRNA and siRNA Northern blots**

Extracted RNAs were resolved on 10% polyacrylamide gel (Acrylamide:Bis-Acrylamide, 19:1) containing 7 M urea and electrophoretically transferred to Hybond N membranes (Amerham Biosciences). The membrane was probed with 32P-labeled cDNA for a specific miRNA in 20 mL of hybridization buffer containing 200 mM sodium phosphate, pH 7.0, and 5% SDS at 37°C overnight. Membranes were washed three times (15 min each) at 25°C with wash buffer containing 300 mM NaCl, 30 mM sodium citrate, and 0.1% SDS. Hybridization signal was detected by autoradiography using a Phosphorimage (Molecular Dynamics). For the siRNA Northern blot, sel-1 probes were 60-nt oligonucleotides corresponding to the sel-1 sequence contained in the trigger plasmid (pJP12.2+) overlapping by 20:

5'-GACTAACCGAGAAAAGATCCAAACTCATATGCAAC CAGGAAGTGTCACCACCTGGGAGA-3'
5'-AGGAATGGCCACACTGGAGGTTACATCTTTGAGGTATT AAGATGCTTTGGCCCCAAAGAGGG-3'
5'-AAGATGCTTGGCCACAAAGGAGTACATCCGCTCAATTG GACCTGGAGACAGATTTATTTA-3'
5'-GAATCAGAAGATTTATTTAGCTTGTCGGAGACGGAGGCCT AATCAAATTTCGCAACCTGGCCT-3'
5'-TCAAAAATTTTCGAAACTCCGCTCTCCGTACTTTGTTAGCTGC TGCTGAGTCAGGAAGTGCCGA-3'

**RNA affinity chromatography**

Bait RNA (8 nmol) was oxidized with sodium periodate and cross-linked to 200 mL of adipic acid dihydrazide agarose beads as previously described (Caputi et al. 1999). The RNA beads were incubated with miRNPs at room temperature for 40 min with gentle agitation. The beads were subsequently washed four times with 1 mL buffer D. miRNPs were eluted from the beads by boiling in SDS-PAGE sample buffer.

**Small RNA cloning**

The 5' ligation-dependent cloning procedure used for miRNP-200 is derived from Lau et al. (2001). Approximately 2 µg of RNA extracted from the mixed-stage miRNP were added to a ligation reaction mixture containing 20 µM adenylated linker (rAppCTG TAGGCAACATCAA/3ddC) and T4 RNA ligase. The product was gel fractionated and products between 36 nt and 44 nt were eluted. The 3'-linked products were then ligated to 10 µM of 5' linker (ATCGTArGrGrArGrArGrAr) and the 53–61-nt products were similarly gel purified. Reverse transcription was performed using the primer (ATTGATGTGCTGCTACAG) followed by 20 cycles of PCR, a 1:40 dilution, then 10 additional cycles of PCR using the primers (ATTGATGTGCTGCTACAG) and (ATCGTGACCCACCTGAAA). The PCR products were gel extracted, digested with Banl, and concatamered with T4 DNA ligase. The concatamerized product was cloned into the TOPO-TA vector (Invitrogen) and sequencing of plasmid inserts was performed following PCR amplification. 5' ligation-independent cloning for siRNPs was performed as described (Pak and Fire 2007).

**Microarray**

DNA oligonucleotide probes were designed for all C. elegans mature miRNAs reported to miRBase (version number 9.0) (Griffiths-Jones et al. 2006). Probes were resuspended in Pronto spotting solution (Corning) at a concentration of 50 µM and robotically spotted onto Epoxide slides (Corning) according to the manufacturer recommendations. Each spot was printed in quadruplicate. Purified miRNAs were labeled with the mirVana miRNA labeling kit (Ambion) and amine-reactive Cy3 and Cy5 dyes according to the manufacturer recommendations as described (Shingara et al. 2003). Three times miRNA hybridization buffer (Ambion) was added to the labeled miRNAs and the solution was heated at 95°C for 3 min. Slides were hybridized for 14–16 h at 42°C. Following hybridization, the slides were washed and dried prior to scanning with an Axon Instruments 4000 series scanners. Microarray scanning data were normalized and further processed using R and Bioconductor (Dudoit et al. 2003). Specifically, median normalization was used before differential expression analysis was done with lmFit (Smyth 2004). Hierarchical clustering was carried out using the Cluster package (Eisen et al. 1998). Microarray spot intensity signals, where the average intensity across all the experiments was above the average signal intensity plus two SD values of all “untargeted” spots, was labeled as “high confidence” for further analysis. “Untargeted” spots were oligonucleotides containing sequences not found in the C. elegans genome or containing only printing buffer.

**SUPPLEMENTAL MATERIALS**

Supplemental tables and figures can be downloaded at: http://bio.research.ucsc.edu/people/zahler/miRNAsupp.htm

**ACKNOWLEDGMENTS**

We thank R. Plasterk for anti-TSN-1 sera, V. Hunt for large-scale worm growth, M. Jurica and J. Little for assistance with FPLC fractionation, L. Hoang for assistance with extract preparation and RNA electrophoresis analysis, and L. Shiue and M. Ares for assistance with the UCSC microarray facility. We thank H. Noller, M. Jurica, M. Ares, Jr., Z. Ni, D. Haussler, M. Robertson, S. Strome, A. Chisholm, and all of the Zahler and Fire laboratory.
Distinct RNP complexes for miRNAs and siRNAs

members for discussion and support. This research is supported by NIH Grants R01-GM61646 to A.M.Z., R01-GM37706 to A. F., P41-HG02731 to the UCSC Center for Genomic Science, the W.M. Keck Foundation for support of the UCSC Center for Molecular Biology of RNA, the UC-MEXUS/CONACyT program for support of S.B.-S., and the UCSC MARC U*STAR program T34 GM007910 for support of M.A.

Received March 23, 2007; accepted May 24, 2007.

REFERENCES


Distinct ribonucleoprotein reservoirs for microRNA and siRNA populations in *C. elegans*

Sam G. Gu, Julia Pak, Sergio Barberan-Soler, et al.

*RNA* 2007 13: 1492-1504 originally published online July 24, 2007
Access the most recent version at doi:10.1261/rna.581907

**References**

This article cites 38 articles, 16 of which can be accessed free at:
http://rnajournal.cshlp.org/content/13/9/1492.full.html#ref-list-1

**Email Alerting Service**

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here.](http://rnajournal.cshlp.org/subscriptions)