Structural basis of microRNA length variety

Julia Starega-Roslan, Jacek Krol, Edyta Koscianska, Piotr Kozlowski, Wojciech J. Szlachcic, Krzysztof Sobczak and Wlodzimierz J. Krzyzosiak*

Laboratory of Cancer Genetics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14 Street, 61-704 Poznan, Poland

ABSTRACT

The biogenesis of human microRNAs (miRNAs) includes two RNA cleavage steps in which the activities of the RNases Drosha and Dicer are involved. miRNAs of diverse lengths are generated from different genes, and miRNAs that are heterogeneous in length are produced from a single miRNA gene. We determined the solution structures of many miRNA precursors and analysed the structural basis of miRNA length diversity using a new measure: the weighted average length of diced RNA (WALDI). We found that asymmetrical structural motifs present in precursor hairpins are primarily responsible for the length diversity of miRNAs generated by Dicer. High-resolution northern blots of miRNAs and their precursors revealed that both Dicer and Drosha cleavages of imperfect specificity contributed to the miRNA length heterogeneity. The relevance of these findings to the dynamics of the dicing complex, mRNA regulation by miRNA, RNA interference and miRNA technologies are discussed.

INTRODUCTION

In animal cells, the post-transcriptional regulation of gene expression by microRNAs (miRNAs) is initiated by the binding of the effector complex miRISC to partially complementary sequences in the 3′-UTR of mRNAs. Binding typically results in mRNA repression by translation inhibition or by deadenylation and degradation (1). The majority of human mRNAs have been predicted to be regulated by miRNAs (2), and the number of experimentally validated targets increases constantly (3). Proteomics studies (4,5) have shown the complexity of miRNA regulatory networks in a new light, and recent technological advances have allowed for the determination of miRNA–mRNA interaction maps *in vivo* (6).

In human cells, most miRNA genes are transcribed by RNA polymerase II (Pol II), and the main miRNA maturation pathway begins with the nuclear processing of primary precursors (pri-miRNAs) by a microprocessor complex (7,8). The microprocessor core contains the ribonuclease type III enzyme Drosha and the DGCR8 protein, which is an RNA-binding protein. DGCR8 binds to the junction of the double-stranded stem of miRNA-containing hairpins and its single-stranded flanks to form a platform to which Drosha binds (9). Pri-miRNA cleavage by Drosha occurs one helical turn away from the junction, towards the hairpin terminal loop. The product of Drosha cleavage, which is typically a ∼60-nt pre-miRNA hairpin, is exported from the nucleus by Exportin-5 and Ran-GTP (10). In the cytoplasm, the pre-miRNA is handed over to the RISC-loading complex (RLC), which contains another ribonuclease III family enzyme (Dicer), the Tar-RNA-binding protein (TRBP) and an argonaute family protein (AGO1-4) (11–13). Dicer anchors itself via its PAZ domain to the 3′-terminus of the pre-miRNA, and the pre-miRNA is cleaved about two helical turns away by a single RNA processing center formed by two RNase III domains (14). A typical product of the pre-miRNA cleavage is an imperfect duplex composed of miRNA and miRNA* strands that contain 5′-monophosphates and free OH groups at the 2-nt 3′-protruding ends.

Mature miRNAs generated from different miRNA genes may differ in length (miRNA length diversity) (15), and individual miRNA genes may give rise to several miRNA species that differ in length (miRNA length heterogeneity). The rapidly growing interest in miRNA heterogeneity is caused by the fact that it diversifies and enriches the miRNA universe and increases the regulatory potential of miRNAs (16), even if the miRNA fraction undergoes some 5′-end nucleotide biases at the AGO2 loading step (17). In addition, this type of miRNA variation has important practical implications for the construction of miRNA precursor-based expression cassettes, which are designed to release

siRNAs or artificial miRNAs with specific sequences. In spite of the importance of these issues, the question how the structure of miRNA precursors predetermines the length diversity and heterogeneity of miRNAs has not yet been formally addressed.

In this study, we focused on the structural aspects of miRNA biogenesis. We investigated the Dicer step in miRNA biogenesis, the structural diversity of pre-miRNAs and the relationship between the structure of precursors and the specificity of Dicer cleavage. We examined recombinant Dicer reactions with numerous synthetic pre-miRNAs and some of their mutants and identified the primary determinants of miRNA length diversity. Then, we addressed the following question: what are the roles of cleavages generated by endogenous Drosha and Dicer in generating non-uniform miRNA ends? We monitored the effects of these two miRNA precursor processing events in a cellular system that overexpresses various pri-miRNAs. We concluded that both processing steps involving Drosha and Dicer generate substantial miRNA length heterogeneity that can be reduced to some extent by AGO2 binding. Finally, we discuss the importance of these observations for the biological function of miRNAs and their implications for the RNAi and miRNA technologies.

Materials and Methods

Cell culture and RNA isolation

HeLa, HT29, MCF7 and HEK293T cells were obtained from the ATCC collection and cultured according to supplier's instructions. Total RNA was isolated from the cell lines using the Tri-Reagent method (MRC).

Northern blots of miRNA and pre-miRNA

RNA (30–40 µg) was resolved in a 12%-denaturing polyacrylamide gel in 0.5× TBE. The XC dye migrated 10 cm and 30 cm during miRNA and pre-miRNA detection, respectively. The marker lanes contained a mixture of radio-labelled RNA oligonucleotides (17, 19, 21, 23 and 25 nt) and the RNA Low Molecular Weight Marker (USB Corp.). RNA was transferred onto GeneScreen Plus membrane (Perkin Elmer) using semi-dry electroblotting (Sigma-Aldrich) and immobilized using UV irradiation (UVP). The membrane was probed with $[\gamma^{32}P]$ ATP (5000 Ci/mmol; Hartmann Analytics)-labelled oligonucleotides (Supplementary Table S1) that were complementary to either miRNA or miRNA*. The hybridization was performed overnight at 37°C in a buffer containing 1% SDS, 5× SSC and 1× Denhardt's Solution. The radioactive signals were quantified by PhosphorImager (Multi Gauge; Fujifilm).

Oligonucleotides

The DNA oligonucleotides used in this study were obtained from IBB Warsaw. Chemically synthesized RNAs were purchased from Curevac and Metabion.

Preparation of RNA substrates

DNA templates for the pre-miR-19b-1, pre-miR-24-1, pre-miR-27b, pre-miR-31, pre-miR-33a, pre-miR-124-2, pre-miR-187, pre-miR-208a, pre-miR-210, pre-miR-214 and pre-miR-496 transcripts were obtained by chemical synthesis and purified by polyacrylamide gel electrophoresis. Each oligomer contained the SP6 or T7 RNA polymerase promoter sequence at the 3′-end (Supplementary Table S2). The experimental strategy and protocols for the preparation of DNA templates and *in vitro* RNA transcripts that are free of detectable 5′-end heterogeneity have been described earlier (18,19). The RNA sequences of the chemically synthesized pre-miR-132, pre-miR-136, pre-miR-139, pre-miR-367, pre-miR-526b, pre-miR-549, pre-miR-591, pre-miR-637 and the mutant transcripts pre-miR-19b-1-Mut, pre-miR-33a-Mut, pre-miR-136-Mut and pre-miR-549-Mut are also shown (Supplementary Table S2). Transcripts were 5′-end-labelled with T4 polynucleotide kinase (Epicenter) and $[\gamma^{32}P]ATP$ (4500 Ci/mmol; ICN), gel-purified and stored at −80°C until use. Pre-miR-496-C and pre-miR-526b-C (chemically synthesised RNA, both lacking the 3′-terminal C-residue) were 3′-end-labelled using T4 RNA ligase (Fermentas) and $[\gamma^{32}P]pCp$ (3000 Ci/mmole; Hartmann Analytics) followed by dephosphorylation of the 5′- and 3′-ends with alkaline phosphatase (Pharmacia) and phosphorylation of the 5′-end with T4 polynucleotide kinase. The product was purified by polyacrylamide gel electrophoresis and stored at −80°C until use.

Nuclease digestions and metal ion-induced cleavages of RNA

Prior to probing the structures, the ^{32}P-labelled transcripts were mixed with an excess of homologous unlabelled RNA and then denatured and renatured by heating at 90°C for 1 min, which was followed by slow cooling to 37°C. The RNA was subjected to limited digestion at 37°C in a solution resembling the intracellular environment (10 mM Tris–HCl, pH 7.2, 40 mM NaCl and 1 mM MgCl₂) or in an optimal buffer for Dicer activity (20 mM Tris–HCl, pH 8.0, 150 mM NaCl and 2.5 mM MgCl₂; 0.5 mM ZnCl₂ was also present in the reactions with nuclease S1).

Structure determination was performed as described earlier (20). Briefly, 8 µl of the RNA prepared in the appropriate buffer described above (50 000 c.p.m., ~15 fmol) was mixed with 2 µl of a probe at different concentrations. The final concentrations of the probes in the reactions were as follows: S1 nuclease—0.3, 0.6, 1.2 U/µl; T1 ribonuclease—0.1, 0.2, 0.3 U/µl; T2 ribonuclease—0.03, 0.04, 0.05 U/µl; V1 ribonuclease—0.03, 0.04, 0.05 U/µl; Pb²⁺ ions—0.1, 0.2 mM. The reactions with the nucleases and lead ions were stopped after 10 min by the addition of an equal volume of a stop solution containing 7.5 M urea and 20 mM EDTA with dyes. The samples were electrophoresed in a 15%-polyacrylamide gel under denaturing conditions at 1500 V, followed by autoradiography at −80°C using an intensifying screen. The products of the structure-probing reactions were also visualized

and analysed by PhosphorImager (ImageQuant v5.1—Molecular Dynamics).

Electrophoresis in nondenaturing conditions

RNA structure homogeneity was analysed for each of the investigated transcripts by electrophoresing the radio-labelled samples in a 10%-nondenaturing polyacrylamide gel (150/140/1 mm; acrylamide/bisacrylamide—29:1) buffered with 10 mM Tris–HCl, pH 7.2, 40 mM NaCl and 1 mM $MgCl_2$ or 20 mM Tris–HCl, pH 8.0, 150 mM NaCl and 2.5 mM $MgCl_2$ at a fixed temperature of 37°C. Prior to gel electrophoresis, the ^{32}P-labelled transcripts (~5000 c.p.m.) were denatured and renatured as described in the preceding section and mixed with an equal volume of solution containing 7% sucrose and dyes. The electrophoresis was performed at 100 V with buffer circulation at 2 l/h and was followed by phosphorimager analysis.

RNA cleavage assay using recombinant Dicer

Prior to the reaction with Dicer (Ambion), the 5′-end-labelled RNAs were denatured and renatured by heating the sample at 80°C for 1 min and then slowly cooling to 37°C. The RNAs (50 000 c.p.m., ~5 fmol) were incubated with human Dicer (~0.5 pmol) at 37°C for different amounts of time, as specified in the figure legends. The reactions were stopped by the addition of an equal volume of gel loading buffer (7.5 M urea and 20 mM EDTA with dyes). The samples were analysed by electrophoresis in a 15%-polyacrylamide 7.5 M urea gel that was run in 1× TBE buffer along with the products of the alkaline hydrolysis and limited T1, S1 or P1 nuclease digestion of the same RNA molecule. The products were detected by autoradiography and then quantitatively analysed with a PhosphorImager (ImageQuant v5.1—Molecular Dynamics).

To visualize Dicer cleavage reactions by northern blot, ~4 pmol of unlabelled 5′-phosphorylated pre-miRNAs were subjected to the reaction with recombinant Dicer as described earlier, except that the time of the reaction was extended to 60 min. Small aliquots were separated in a 12%-polyacrylamide gel, transferred to nylon membranes and hybridization was performed with the miRNA or miRNA* specific probes (Supplementary Table S1) as described earlier.

Preparation of the ladders

The alkaline hydrolysis ladder was generated by incubating the labelled RNA in formamide buffer containing 0.5 mM $MgCl_2$ at 100°C for 10 min. The RNAs (~50 000 c.p.m.) were partially digested with T1 ribonuclease under semi-denaturing conditions (10 mM sodium citrate, pH 4.5, 0.5 mM EDTA and 3.5 M urea) for 10 min at 55°C using 0.2 U/μl of the enzyme. The S1 ladder was generated by incubating 5′-end-labelled RNA with 0.2 U/μl of S1 in a buffer containing 10 mM Tris–HCl, pH 7.2, 40 mM NaCl, 1 mM $ZnCl_2$ and 1 mM $MgCl_2$ for 1 min at 75°C. The P1 ladder was generated by incubating the labelled RNA with 1 U/μl of P1 ribonuclease in a buffer containing 6.5 M urea, 20 mM sodium phosphate, pH 9.0 and 1 mM $ZnCl_2$ for 20 min at 55°C as described earlier (21).

DNA transfection

HeLa and HEK293T cells were grown to 90% confluence and transfected, using Lipofectamine 2000 (Invitrogen), with 10 μg of the plasmid construct (System Biosciences, Open Biosystems, GeneCopoeia, Cell Biolabs) that encoded an appropriate miRNA precursor (Supplementary Table S3). The efficiency of the transfection was monitored using a fluorescent reporter gene. The cells were harvested 24 h after transfection, and the isolated RNA was used for northern blot and primer extension analyses.

Primer extension

Prior to primer extension, each primer (Supplementary Table S4) was gel-purified. For the primer extension assay, 10 μg of the total RNA and 300 fmol of the 5′-end-labelled [γ^{32}P]ATP (5000 Ci/mmol; Hartmann Analytics) primer were used. After the primer was annealed at 42°C for 10 min, it was extended for 30 min at 42°C using 2 U of AMV Reverse Transcriptase (Promega). The reaction was stopped by addition of an equal volume (10 μl) of formamide loading dye, and the products were separated by electrophoresis in a 15%-polyacrylamide gel with 7.5 M urea along with size marker. Radioactive signals were quantified using a PhosphorImager (Multi Gauge, Fujifilm).

Statistical analysis

In all of the tests where the *P*-value was calculated, $P \leq 0.05$ was considered significant. The Student's *t*-test with Welch's correlation for unequal variance was calculated using Statistica (StatSoft) or Prism v. 4.0 (GraphPad Software). End-heterogeneity (in the Dicer *in vitro* cleavage test), 5′-end-heterogeneity (in the primer extension experiment) and length-heterogeneity (using the northern blots) were calculated using the equation $H = 1 - F_{max}$, where H represents the heterogeneity and F_{max} is the fraction of the most abundant product (%).

To generate a bubble-chart graph (MS Excel), we analysed the lengths of the Dicer cleavage products, and the cleavage intensities were represented as bubble areas [as calculated from the densitometric analysis (ImageQuant v5.1—Molecular Dynamics)].

RESULTS

The biogenesis machinery of miRNAs typically generates heterogeneous products

miRNAs isolated from various organisms and tissues frequently show length heterogeneity. This can be seen in the northern blot analysis of several human miRNAs that are not derived from members of miRNA families (Figure 1). In the three analysed cell lines, HeLa, MCF7 and HT29, miR-191 was detected in three length variants, miR-16-1, miR-21 and miR-31 were detected in two length variants, and miR-25 was represented by a single band.
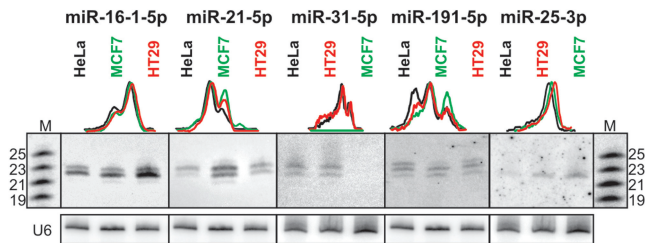
Figure 1. miRNA length heterogeneity is a frequent phenomenon. RNA (40 μg) isolated from different cell lines was analysed by northern blot using specific probes complementary to miRNAs (indicated above), as described in 'Materials and Methods' section. Quantitative representations of the miRNA length variants are shown using the peaks obtained from the phosphoimaging analysis. M denotes the 19–25-nt RNA size marker. U6 is the hybridization control.

The heterogeneity patterns observed in the different cell lines did not differ significantly. A length comparison of these different miRNAs showed that they varied from 22 nt for miR-25; 22 and 23 nt for miR-16-1, miR-21 and miR-31; and 22, 23 and 24 nt for miR-191, which covers the upper part of the miRNA length range (15). As the length diversity of miRNAs may be attributed to the different structures of the pre-miRNAs, we began our study with a detailed structure analysis of pre-miRNAs that were then subjected to a Dicer cleavage assay.

Structures of pre-miRNAs used for Dicer cleavage studies are highly diverse

Earlier analyses of the predicted structures of pre-miRNA hairpins revealed their great variety (22,23). In this study, we aimed to determine the relationship between the sites of Dicer-induced cleavage and the structural features of pre-miRNAs. For this purpose, we selected 19 pre-miRNAs of different lengths that contained different numbers, types and locations of RNA structure motifs. As some of the pre-miRNAs were synthesized by *in vitro* transcription, we used rigorous quality control measures (19,24) to ensure that the transcripts had homogeneous ends. Each transcript was analysed for sequence and structure homogeneity prior to the structure analysis. The secondary structures of the pre-miRNAs were determined using state-of-the-art biochemical structure probing methods (19,20), as shown for pre-miR-24-1, pre-miR-27b and pre-miR-33a in Supplementary Figures S1 and S2. The structures of all 19 pre-miRNAs are shown in Supplementary Figure S3.

Primary and secondary Dicer cleavages and their alternative detection methods

To examine how Dicer cleaves pre-miRNA substrates, we performed a time-course cleavage assay with recombinant Dicer for two precursors, pre-miR-496 and pre-miR-526b, which were labelled at either the 5′- or 3′-ends (Figure 2A and B). We selected these pre-miRNAs because they could attain identical nucleotide sequences after 5′- or 3′-end-labeling ('Materials and Methods' section). The positions of the Dicer cleavages were assigned using homologous ladders: P1 and S1 nuclease ladders that have the

same end-groups as the Dicer cleavage products and/or standard T1 RNase and alkaline hydrolysis ladders (Figure 2 and Supplementary Figure S4). The patterns of pre-miRNA cleavage by Dicer revealed two types of cut, which we defined as 'primary' and 'secondary', i.e. cleavages that occur in intact precursors and in single-nick intermediates, respectively. The local heterogeneity of the cuts generated by Dicer within each arm was not taken into account in this classification. Dicer cleavage sites that are observed at the 3′-side of pre-miRNA, which is labelled at the 5′-end, represent the primary cleavages that are generated by the RNase IIIa domain of Dicer. The fragments generated by Dicer from the 5′-side represent the cumulative effect of primary and secondary cleavages by the RNase IIIb domain. Using 3′-end-labelled pre-miRNA, we could distinguish between primary cleavages in the 5′-arm and combined primary and secondary cleavages in the 3′-arm. It may be noted from the comparisons of the cleavage intensities within the 5′- and 3′-arms of each pre-miRNA labelled at either the 5′- or 3′-end that the radioactive signals from shorter products were the major signals regardless of the labelled end, indicating that secondary cleavages make a large contribution to the observed cleavage patterns (Figure 2). A diagram of the observed cleavage types and fragments generated from the pre-miRNAs that were labelled at either the 5′- or 3′-ends is shown in Figure 2C. Notably, the cleavage patterns obtained for the 5′- and 3′-end-labelled pre-miR-496 and pre-miR-526b also showed that the primary and secondary cuts in each arm occurred mainly at the same sites. This means that the specificity of the primary and secondary cleavage sites within each arm well correspond to each other.

We have also used northern blot analysis with miRNA and miRNA* specific probes as an alternative method to visualize the intermediates and products of pre-miR-136 and pre-miR-139 cleavage by recombinant Dicer and compared results with those obtained using ^{32}P end-labelled pre-miRNA. It can be seen (Figure 2D and E) that the cleavage products detected by northern blot agree well with the products generated by Dicer from the 5′-end-labelled pre-miRNAs. To compare the cleavages generated by Dicer in all 19 pre-miRNAs, we used transcripts labelled at their 5′-ends.

The length diversity of miRNAs depends strongly on the pre-miRNA structure

The data shown in Figures 2, 3A and 4 and Supplementary Figure S5A is a rich source of information on the relative intensities of 5′- and 3′-arm cleavages (Supplementary Figure S5B) as well as on influence of precursor structure on the lengths of miRNAs generated from different pre-miRNAs. To analyse these results, we combined information on the cleavage sites within the 5′- and 3′-arms of the precursors with quantitative data on the cleavage intensities. The heterogeneity of the Dicer cleavage products was calculated for both arms and is shown in Figure 3B as the cumulative fraction of signals from products other than the dominant product. The results show that the length heterogeneity of the miRNAs that
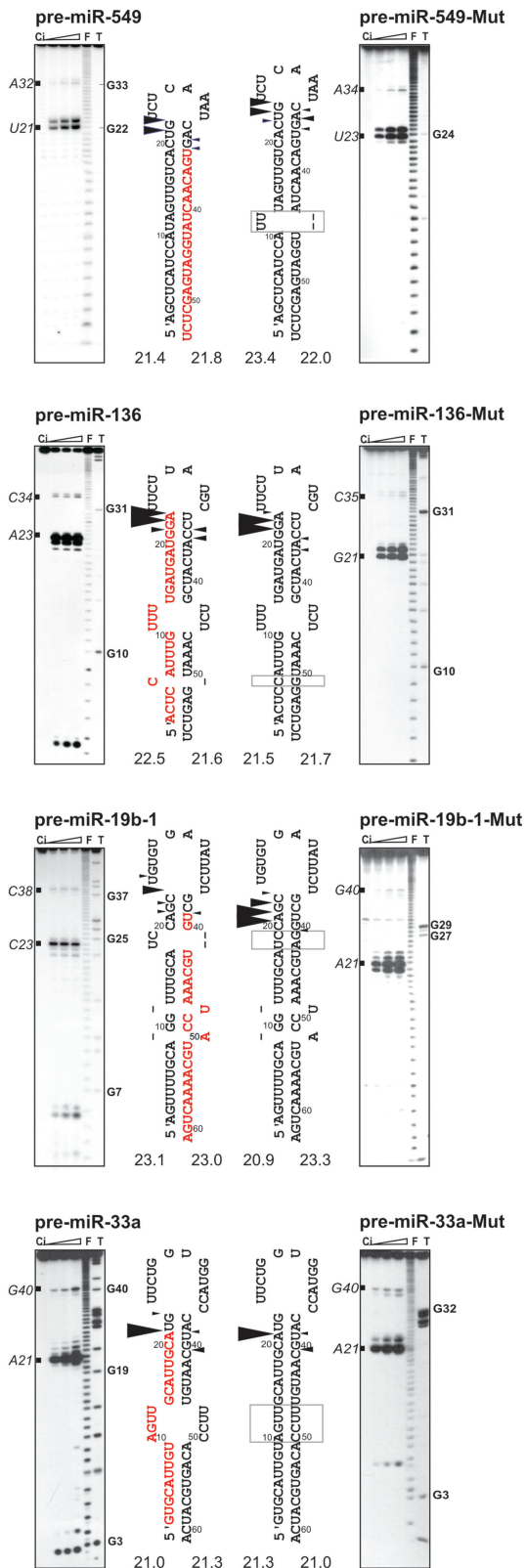
Figure 2. Different experimental systems used to analyse Dicer cleavage products. A comparative analyses of the cleavage patterns generated by recombinant Dicer (Ambion) in a time-dependent reaction: 1, 2 and 5 min for pre-miR-496 (**A**) and pre-miR-526b (**B**) labelled at the 5′- (left panel) or 3′-end (right panel). Lane Ci, incubation control without Dicer; lane S, S1 ladder; lane P, P1 ladder; lane T, guanine-specific ladder; lane F, formamide ladder. The positions of the selected cleavage sites (marked with black boxes) and G-residues were counted from the 5′-end of the precursors in all autoradiograms. In the middle, the structures of these precursors are shown. Asterisks represent the labeling site on the 5′-end (red) and the 3′-end (blue). The Dicer cleavage sites (marked with red arrowheads in the precursor labelled at the 5′-end or blue arrowheads in the 3′-end-labelled precursor, the size of which is proportional to the radioactive signal intensity: very strong, strong, medium and weak) are shown in the secondary structure models, where the reported miRNA sequences are marked in red. (**C**) The scheme of pre-miRNA processing by recombinant Dicer labelled at the 5′- (left, red) and 3′-end (right, blue). From left to right: unprocessed precursor (substrate); single-nick intermediate products; precursor that is cleaved in both arms (product). The expected electrophoretic band migrations of the Dicer cleavage products that were labelled at either 5′- or 3′-ends are shown schematically below. (**D** and **E**) The products of the Dicer cleavage assay of the unlabelled pre-miR-136 (D) and pre-miR-139 (E) were subjected to northern blot with specific 5′- or 3′-end probes detecting miRNA derived from 5′- or 3′-arm, respectively (lanes 1 and 2 for pre-miR-136 and pre-miR-139, respectively). For comparison, on the right, the structure of the pre-miRNA and the result of the Dicer *in vitro* cleavage assay of the labelled precursors are shown. The blue arrowheads represent the cleavages by Dicer in the 3′-arm being only primary cuts. The red arrowheads mark the cleavages in the 5′ precursor arm (the sum of primary and secondary cuts). The blue and red arrows indicate the corresponding cleavage sites. The migration of the miRNA, intermediates and pre-miRNA in the northern blot are represented with black lines on the right side. M1 denotes the Low Molecular Weight Marker (USB Corp.); M2 denotes the 17–25 nt RNA marker. Other designations are the same as in (A) and (B).

Figure 3. Dicer generates heterogeneous products of diverse lengths. (**A**) An analysis of Dicer cleavages in selected 5′-end-labelled pre-miRNAs that were incubated with recombinant Dicer (Ambion) for 1, 2 and 5 min. Below each precursor arm, the WALDI parameter is presented for the Dicer cleavage products generated from this arm. The designations are the same as in Figure 2A and B. (**B**) The heterogeneity of the products that were generated from the 5′-end-labelled precursors from the 5′ (the sum of primary and secondary cleavages) and 3′-arms (primary cleavages) was calculated based on the data presented in Figures 2, 3A, 4 and Supplementary Figure S5A as described in 'Materials and Methods' section. Dots and horizontal lines indicate the individual and average values of miRNA heterogeneity, respectively. Grey dots represent the heterogeneity of products generated from symmetrical precursors, while white dots represent those generated from asymmetrical precursors. Signals that accounted for <5% of all the cleavage products were not considered in the analysis. (**C**) A bubble-chart (MS Excel) depicting the fractions of cleavage products that were generated by Dicer in the 5′-arm (left-hand side) and 3′-arm (right-hand side) of the 5′-end-labelled precursors analysed (*y*-axis). Note that the left diagram represents the sum of the primary and secondary cleavage products, while the right diagram shows the primary cleavage products only. The *x*-axis represents the lengths of the products, and the bubble size (area) represents the fraction of the product that was generated from a given arm (as calculated from the densitometric analysis of the results presented in Figures 2, 3A, 4 and Supplementary Figure S5A). The values of WALDI parameter from the 5′- and 3′-arms are shown as red vertical lines. The precursors were sorted according to the increasing value of the WALDI parameter generated from the 5′-arm. White and grey bubbles indicate products that were diced from the 'asymmetrical' (i.e. including at least one asymmetrical motif in the stem; labelled ASYM) and 'symmetrical' precursors (labelled SYM), respectively. The insets (i.e. the bar charts) present average WALDI values for the 'symmetrical' and 'asymmetrical' precursors with error-bars representing the standard deviation (SD). These values were formally compared using a Student's *t*-test with Welch's correlation for unequal variance, and the *P*-values are shown in the insets. Signals that accounted for <5% of all the cleavage products were not considered in graph (C).

are generated from the precursor 5′-arm (combined primary and secondary cleavage products) was very similar to that of miRNAs generated from the 3′-arm (primary cleavage products) [0.31 and 0.35, respectively

(Figure 3B)]. We could not identify any correlation between pre-miRNA structure and miRNA heterogeneity if the presence of asymmetrical versus symmetrical motifs was considered. We hypothesize that more subtle features

Figure 4. *In vitro* Dicer cleavages in the pre-miRNAs and their mutants. Cleavages by recombinant Dicer of the 5′-end-labelled pre-miR-19b-1, pre-miR-33a, pre-miR-136, pre-miR-549 and the mutant variants. The sequences that were mutated in the precursors are marked with rectangles. Below each precursor arm, the WALDI parameter is presented for Dicer cleavage products generated from this arm. Other designations and symbols are the same as in Figure 2.

of pre-miRNA architecture may be involved in determining miRNA heterogeneity.

The pre-miRNAs that harbor either symmetrical RNA structural motifs (i.e. mismatches or symmetrical internal loops) or both symmetrical and asymmetrical motifs (i.e. bulges or asymmetrical internal loops) were ranked according to the weighted average length of diced RNA (WALDI), and the results are presented in Figure 3C (vertical red lines). Although the ranges and standard deviations of the WALDI parameter are lower for cleavage products within the 5′-arm than for the products within the 3′-arm (SD: 0.52 versus 0.85, respectively), the average lengths of these products are very similar (21.8 and 21.9 nt, respectively). When the cleavages that were generated in the pre-miRNAs that contained only symmetrical RNA structure motifs (grey bubbles in Figure 3C) were compared with those in pre-miRNAs that contained at least one asymmetrical motif (white bubbles), it became apparent that the latter gave rise to significantly longer products (average WALDI: 21.5 nt versus 22.2 nt, and 21.6 nt versus 22.2 nt, for 5′- and 3′-arms, respectively, see Figure 3C inset).

Longer miRNAs derive from pre-miRNA arms containing excessive nucleotides

To illustrate the general observations described in the previous paragraph using specific examples, we note that the structures of pre-miR-132 and pre-miR-136 (Figures 2D and 3A) differ in a single C-bulge present in the 5′-arm of the latter and that otherwise their hairpin stems have similar architectures. These precursors have nearly identical cleavage patterns, but the products of the Dicer cleavages in the 5′-arm of pre-miR-136 are 1-nt longer than those generated from pre-miR-132 (WALDI: 22.5 and 21.6 nt, respectively). This length difference may suggest that only the consecutive base pairs are counted by Dicer during cleavage site determination. Dicer cut in the 5′-arm of pre-miR-637 and generated predominantly a 22-nt fragment (WALDI: 21.9 nt) and cut in its 3′ arm to release a 24-nt fragment (Figure 3A). Again, the difference in lengths between the released fragments may be explained by the different number of bulged nucleotides along the hairpin stem in both precursor arms. Conversely, the same total number of bulged nucleotides in the stem portion that spans the Dicer anchoring site and the cleavage site explains the generation of the 23-nt fragment from each arm of pre-miR-19b-1 (Figure 4).

Pre-miRNA mutagenesis supports the conclusions from the analysis of natural precursors

To demonstrate the influence of asymmetrical structural motifs on the lengths of miRNAs generated by Dicer, we created specific pre-miRNA mutants. The asymmetrical motif was removed from pre-miR-19b-1 and pre-miR-136. In pre-miR-549, a 2-nt UU bulge was inserted, and in pre-miR-33a, the symmetrical internal loop AGUU/UUCC was replaced with a fully paired sequence (Figure 4). Natural precursors and their mutants were 5′-end-labelled and subjected to Dicer cleavage assay. In all cases, products derived from

pre-miRNA arms containing excessive nucleotides were longer than those released from arms not containing these nucleotides, which was reflected by the differences in the corresponding WALDI parameters. Conversely, no substantial change in the length of released fragments was observed between pre-miR-33a to pre-miR-33a-Mut, in which a symmetrical internal loop was eliminated (Figure 4). Taken together, the results of the mutational analysis provided additional evidence that the asymmetrical motifs had a much stronger effect on the lengths of miRNAs released by Dicer than symmetrical motifs. Our experimental observations were supported by the results of a comparative analysis of the predicted secondary structures of all human pre-miRNAs and the lengths of miRNAs deposited in miRBase version 14. The presence of excessive nucleotides in any pre-miRNA arm gave rise to longer miRNAs generated from this arm (Figure 5).

Drosha cleavages contribute to miRNA length heterogeneity

To gain insight into Drosha and Dicer cleavages in cells, we overexpressed a number of pri-miRNAs (Supplementary Table S3) in HEK293T cells and analysed their processing products using high-resolution northern blot analysis (Figure 6). Overexpression of the primary transcripts did not change the processing specificity characteristics of the endogenous miRNAs (Supplementary Figure S6). The processing efficiency of different primary transcripts and pre-miRNAs varied considerably, as shown by the different proportions of the pre-miRNA and miRNA fractions in the total northern blot signal. Both Drosha and Dicer cleavages were highly efficient in the case of pre-miR-214, for which >95% of the total signal was in the miRNA fraction.

The northern blot experiments confirmed the frequent occurrence of miRNA length heterogeneity (Figure 6A). Most of the analysed miRNAs were heterogeneous in length, and only miR-136 and miR-148a were

homogeneous. The mean value of the heterogeneity was 0.28 for the 14 miRNAs shown in Figure 6A and C. We also analysed the pre-miRNA fraction using a high-resolution northern blot from long-electrophoretic runs (Figure 6B). Of the 14 pre-miRNAs, six (pre-miR-25, pre-miR-31, pre-miR-136, pre-miR-139, pre-miR-141 and pre-miR-432) did not show length heterogeneity. In all other cases, the heterogeneity ranged from 0.08 to 0.49. Overall, the average pre-miRNA length heterogeneity was 0.18 (Figure 6B and 6C). We also determined the heterogeneity of the miRNAs generated from the six pre-miRNAs that were homogeneous in length. Dicer generated heterogeneous products from five of six pre-miRNAs (the average heterogeneity of these six miRNAs was 0.23) (Figure 6C). To confirm the northern blot results, the primer extension assay was performed (Figure 6D) on RNA isolated from HEK293T cells transfected with pri-miRNA encoding plasmids. We analysed the status of the 5′-ends of the miRNAs generated from 14 pre-miRNAs by either Drosha or Dicer. The analysis of the radioactive signal indicated that the 5′-ends generated by Drosha were less heterogeneous than those generated by Dicer (average heterogeneity: 0.09 versus 0.18, respectively). The case of homogeneous miRNA derived from heterogeneous pre-miRNA was noteworthy. The northern blot results showed that from heterogeneous pre-miR-148a (heterogeneity 0.42), only one dominant mature miRNA, or alternatively a population of miRNA variants of the same length, was generated from the pre-miRNA 5′ arm (Supplementary Figure S7). The primer extension results showed that the 5′-end of the mature miRNA was heterogeneous (heterogeneity 0.24). These results allowed us to conclude that (i) more than one length variant of pre-miR-148a is a substrate for Dicer; (ii) at least two mature miR-148a variants of the same length exists (shift in sequence by 1 nt); and (iii) the 3′-end of pre-miRNAs is more heterogeneous than the 5′-end (Supplementary Figure S7). Taken
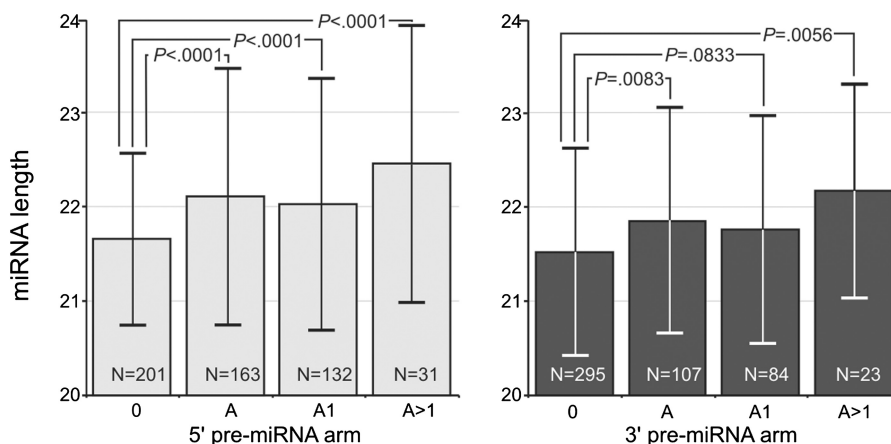

Figure 5. The correlation between miRNA lengths and the presence of excessive nucleotides in pre-miRNA arms. The lengths of miRNAs generated from the pre-miRNA arms that did not contain asymmetrical motifs (0) compared to the lengths of miRNA generated from pre-miRNA arms containing at least one asymmetrical motif (A), containing one asymmetrical motif (A1) and containing more than one asymmetrical motif (A > 1). The results are shown separately for the 5′- and 3′-arms. *N*, number of analysed miRNAs; the Mann–Whitney test *P*-values are indicated in the graphs. Standard deviations are shown as error bars. The results are based on a bioinformatics analysis of 721 human miRNA sequences and pre-miRNA structures.
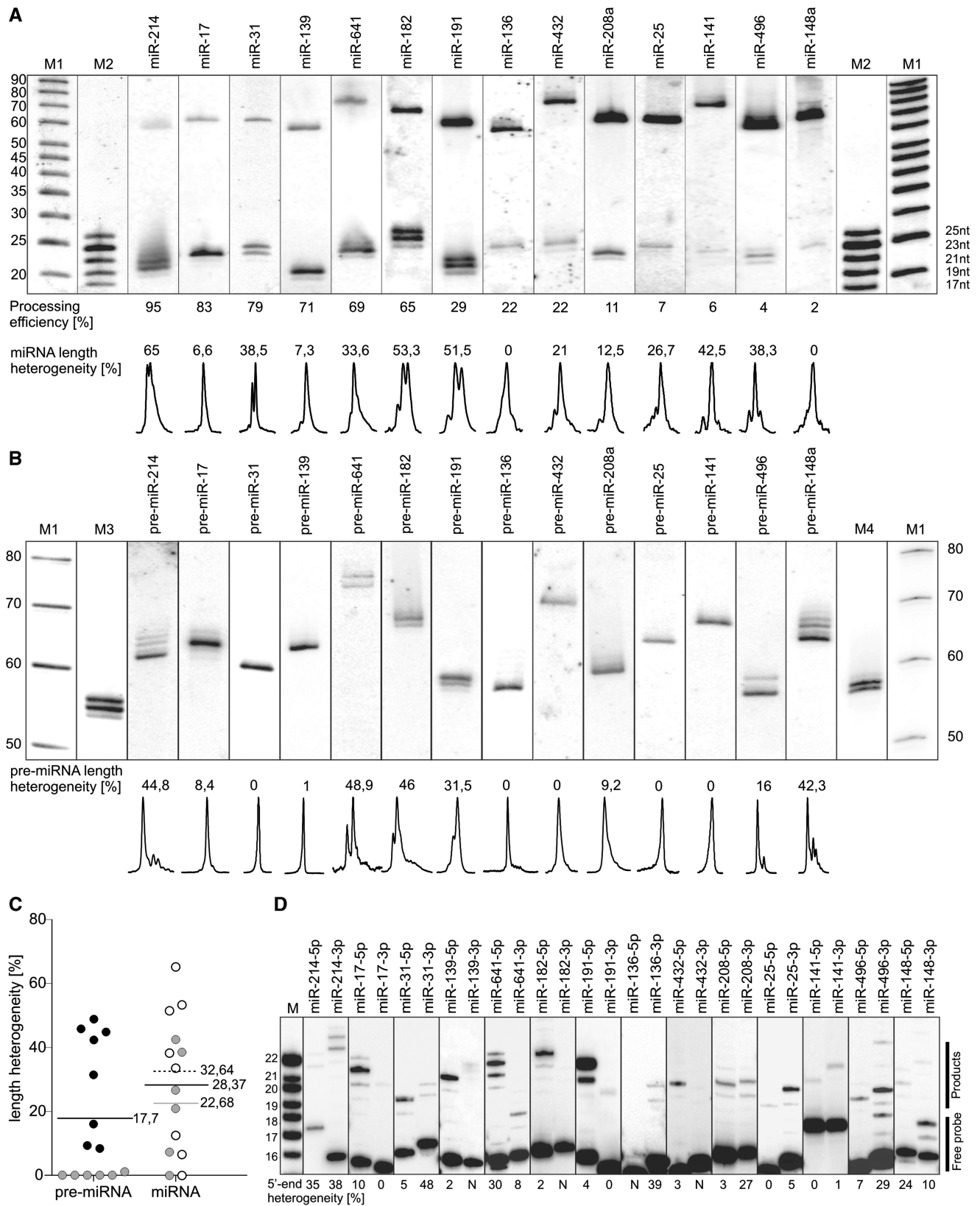
Figure 6. Analysis of the length heterogeneity of miRNAs and pre-miRNAs. (**A**) Northern blots showing the efficiency of precursor processing and the length variation of the miRNAs. RNA (30 μg) that was isolated from HEK293T cells after transfection with different plasmid-encoded pri-miRNAs (indicated above the gel) was analysed in the blot. We analysed the total miRNA fraction in the total northern blot hybridization signals (miRNA and pre-miRNA fraction) quantitatively, and the result is shown below the gel (calculated as the processing efficiency in percent). The fraction of the miRNA variants other than the most abundant miRNA (i.e. the miRNA length heterogeneity) was calculated as in the Figure 3B,

Continued

together, our data demonstrate that both processing steps contribute to miRNA length heterogeneity and that the role of Dicer cleavage inaccuracy may be somewhat greater.

DISCUSSION

In this study, the specificity of cleavages induced in miRNA precursors by both Drosha and Dicer was analysed by high-resolution northern blot and primer extension analyses. In most cases, the considerable heterogeneity of the miRNAs was observed, which most likely resulted from imprecise cleavages by Drosha and Dicer and could be further biased by AGO2 binding (17). The inaccurate generation of the miRNA 5′-end by Drosha or Dicer, which has also been observed in miRNA discovery efforts by deep sequencing (25–27), has important functional implications, even if the 5′-end variability is reduced by AGO2 binding (17). The miRNAs that have shifted 5′-ends have different seed sequences, and they may regulate different sets of targets (16,27,28). We have predicted the target pools for the 5′-end variants of miR-214, miR-17, miR-191 and miR-25 (Supplementary Figure S8). Our *in silico* predictions together with both bioinformatics (16) and experimental (29) results obtained by others support the notion that not one but more miRNA 5′-end variants should be annotated because they all may contribute to miRNA function.

The imprecise processing of miRNA precursors also has important implications for the rapidly developing RNAi and miRNA technologies. The vector-based siRNAs or artificial miRNAs that are expressed from pri-miRNA shuttles are thought to have the same sequence and exert the same effects as the effective siRNAs or miRNA mimics that are obtained by chemical synthesis. However, the relaxed specificity of the Drosha and Dicer cleavages will result in only a fraction of the silencing reagent with the desired sequence; the rest will have shifted sequences. The release of the right allele-specific siRNA from the pri-miRNA type vector is still more challenging. The results of the analysis of the processing products of 14 different pri-miRNAs (Figure 6) could serve as a guide for the selection of suitable shuttles that express artificial miRNAs or siRNAs in cells.

The role of the structure of the pre-miRNA hairpins as an intrinsic factor in the regulation of mammalian miRNA biogenesis has not been previously analysed in detail. In this study, we determined the secondary structures of 19 pre-miRNAs and demonstrated that their diverse structures strongly influence the specificity of Dicer cleavage. In particular, the presence of bulges and asymmetrical internal loops within the approximately two helical turns that span the Dicer anchoring site and cleavage site were the major source of miRNA length diversity (Figure 7A). Cleavage by Dicer alone has also been shown to generate the substantial length heterogeneity of its products (Figure 7B). The model suggests that RNA structure plays a substantial role in determining miRNA length diversity and that Dicer flexibility plays the prevalent role in determining miRNA heterogeneity. In cells, both RNA and Dicer are likely to exhibit structural flexibility during active complex formation. The WALDI parameter introduced in this study to characterize the products of Dicer, which acts as a ruler, allowed us to evaluate the specificity of Dicer cleavage in a quantitative manner and may be useful for further studies of this kind.

Neither the crystal structure of human Dicer nor that of the Dicer complex with its accessory proteins (TRBP and AGO) and/or pre-miRNAs has been reported thus far. Most of the relevant structural information comes from biochemical studies (14) and from the crystal structures of Dicer from *Giardia intestinalis* (30,31) and bacterial RNase III (32). Recent electron microscopy imaging of the human Dicer–TRBP complex (33) and the human RLC complex (34) provided low-resolution structural information about the molecular architecture of these complexes and suggested how pre-miRNA docking might occur. The results of our study show that pre-miRNA structure may substantially contribute to the dynamics of the dicing complex. The stems of human pre-miRNA hairpins are typically mosaics of base pairs and internal loops of various types and sizes and have, on average, 2.7 structure-disturbing motifs per precursor (22). We propose that the unmatched nucleotides of pre-miRNAs that are accommodated within the substrate channel are often not counted by Dicer when it measures the distance to its cleavage site. Thus, the accumulation of structural imperfections in pre-miRNA

Figure 6. Continued
and the results are presented below the gel. A quantitative representation of the miRNA length variants is shown using the peaks obtained from the phosphoimaging analysis; M1 denotes the Low Molecular Weight Marker (USB Corp.); M2 denotes the 17–25 nt RNA marker. (**B**) Northern blots presenting the length variation of the pre-miRNAs. The experiments were performed as in (A), but longer electrophoretic runs were applied. The fractions of the pre-miRNA variants other than the most abundant pre-miRNA were determined as in Figure 3B and are shown below the gel. The quantitative representation of the pre-miRNA length variants is shown using the peaks obtained from the phosphoimaging analysis. M1 denotes the Low Molecular Weight Marker (USB Corp.) M3 and M4 denote markers of the same sequence but differ in length by 1 nt (53 and 54 nt, and 58 and 59 nt, respectively). (**C**) Dots and black horizontal lines indicate the individual and average values of the length heterogeneity of all pre-miRNAs and miRNAs, respectively. The numbers represent the average length heterogeneity. The black dots represent pre-miRNAs that are heterogeneous in length. The grey dots represent either precursors that were homogeneous in length or miRNAs that were generated from these precursors. The grey line represents the average values of the miRNA length heterogeneity for those miRNAs that were generated from the homogeneous in length pre-miRNAs. The white dots show miRNAs that were generated from heterogeneous precursors. The dotted line represents the average value of the length heterogeneity of the miRNAs that were generated from heterogeneous precursors. (**D**) Primer extension analyses of RNA isolated from cells that had been transfected with the same plasmid-encoded pri-miRNAs as in Figure 6A. Oligodeoxynucleotide primers (Supplementary Table S4) complementary to either the miRNA or miRNA* sequences (which detect miRNAs derived from the 5′-arm-5p or the 3′-arm-3p) were annealed to the RNA (10 μg). M denotes the 16–22 nt DNA marker. The fraction of miRNA 5′-end variants other than the most abundant primer extension product was calculated, and the result is shown below the gel. *N*, products not detected or detected at level too low to analyse. The migration of the free probe and a fraction of the extended products are represented by black lines on the right side.
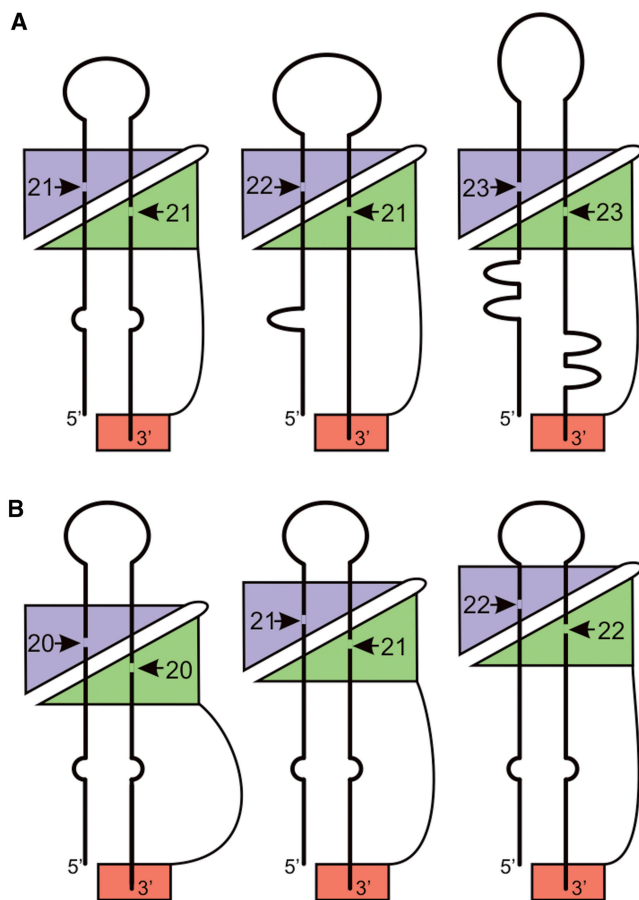
Figure 7. Schematic representations of the different phenomena involved in creating miRNA length variation. (**A**) miRNA length diversity—miRNAs of diverse lengths are generated from different precursors. Symmetrical or asymmetrical structural motifs in pre-miRNA precursor are represented by single base mismatch or bulges, respectively. Dicer RNaseIII domain A (RIIIA) is indicated using a green triangle, and domain B (RIIIB) is indicated using a violet triangle. The PAZ domain is shown as a red rectangle. The arrows and numbers represent the Dicer cleavage sites and the lengths of the Dicer products, respectively. (**B**) miRNA length heterogeneity—miRNAs of heterogeneous lengths are generated from individual precursor. All representations and symbols are the same as in (A). For simplicity, the pre-miRNA is shown as flexible in generating length diversity (A), and Dicer is shown as flexible in generating miRNA length heterogeneity (B).

hairpins may result in the higher plasticity of the precursor structures, and this may be considered another successful strategy for the enrichment of miRNA diversity and increased complexity of miRNA regulatory networks.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

REFERENCES

1. Filipowicz,W., Bhattacharyya,S.N. and Sonenberg,N. (2008) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat. Rev. Genet.*, **9**, 102–114.
2. Friedman,R.C., Farh,K.K., Burge,C.B. and Bartel,D.P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.*, **19**, 92–105.
3. Papadopoulos,G.L., Reczko,M., Simossis,V.A., Sethupathy,P. and Hatzigeorgiou,A.G. (2009) The database of experimentally supported targets: a functional update of TarBase. *Nucleic Acids Res.*, **37**, D155–D158.
4. Baek,D., Villen,J., Shin,C., Camargo,F.D., Gygi,S.P. and Bartel,D.P. (2008) The impact of microRNAs on protein output. *Nature*, **455**, 64–71.
5. Selbach,M., Schwanhausser,B., Thierfelder,N., Fang,Z., Khanin,R. and Rajewsky,N. (2008) Widespread changes in protein synthesis induced by microRNAs. *Nature*, **455**, 58–63.
6. Chi,S.W., Zang,J.B., Mele,A. and Darnell,R.B. (2009) Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature*, **460**, 479–486.
7. Denli,A.M., Tops,B.B., Plasterk,R.H., Ketting,R.F. and Hannon,G.J. (2004) Processing of primary microRNAs by the Microprocessor complex. *Nature*, **432**, 231–235.
8. Gregory,R.I., Yan,K.P., Amuthan,G., Chendrimada,T., Doratotaj,B., Cooch,N. and Shiekhattar,R. (2004) The Microprocessor complex mediates the genesis of microRNAs. *Nature*, **432**, 235–240.
9. Han,J., Lee,Y., Yeom,K.H., Nam,J.W., Heo,I., Rhee,J.K., Sohn,S.Y., Cho,Y., Zhang,B.T. and Kim,V.N. (2006) Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell*, **125**, 887–901.
10. Lund,E., Guttinger,S., Calado,A., Dahlberg,J.E. and Kutay,U. (2004) Nuclear export of microRNA precursors. *Science*, **303**, 95–98.
11. Chendrimada,T.P., Gregory,R.I., Kumaraswamy,E., Norman,J., Cooch,N., Nishikura,K. and Shiekhattar,R. (2005) TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature*, **436**, 740–744.
12. Haase,A.D., Jaskiewicz,L., Zhang,H., Laine,S., Sack,R., Gatignol,A. and Filipowicz,W. (2005) TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. *EMBO Rep.*, **6**, 961–967.
13. Maniataki,E. and Mourelatos,Z. (2005) A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. *Genes Dev.*, **19**, 2979–2990.
14. Zhang,H., Kolb,F.A., Jaskiewicz,L., Westhof,E. and Filipowicz,W. (2004) Single processing center models for human Dicer and bacterial RNase III. *Cell*, **118**, 57–68.

15. Griffiths-Jones,S., Saini,H.K., van Dongen,S. and Enright,A.J. (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Res.*, **36**, D154–D158.
16. Chiang,H.R., Schoenfeld,L.W., Ruby,J.G., Auyeung,V.C., Spies,N., Baek,D., Johnston,W.K., Russ,C., Luo,S., Babiarz,J.E. *et al.* (2010) Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. *Genes Dev.*, **24**, 992–1009.
17. Frank,F., Sonenberg,N. and Nagar,B. (2010) Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature*, **465**, 818–822.
18. Sobczak,K., de Mezer,M., Michlewski,G., Krol,J. and Krzyzosiak,W.J. (2003) RNA structure of trinucleotide repeats associated with human neurological diseases. *Nucleic Acids Res.*, **31**, 5469–5482.
19. Krol,J. and Krzyzosiak,W.J. (2006) Structure analysis of microRNA precursors. *Methods Mol. Biol.*, **342**, 19–32.
20. Krol,J., Sobczak,K., Wilczynska,U., Drath,M., Jasinska,A., Kaczynska,D. and Krzyzosiak,W.J. (2004) Structural features of microRNA (miRNA) precursors and their relevance to miRNA biogenesis and small interfering RNA/short hairpin RNA design. *J. Biol. Chem.*, **279**, 42230–42239.
21. Cruz-Reyes,J., Piller,K.J., Rusche,L.N., Mukherjee,M. and Sollner-Webb,B. (1998) Unexpected electrophoretic migration of RNA with different 3' termini causes a RNA sizing ambiguity that can be resolved using nuclease P1-generated sequencing ladders. *Biochemistry*, **37**, 6059–6064.
22. Kozlowski,P., Starega-Roslan,J., Legacz,M., Magnus,M. and Krzyzosiak,W.J. (2008) In Ying,S.-Y. (ed.), *Current Perspectives in microRNAs (miRNA)*. Springer, Netherlands, pp. 1–16.
23. Krol,J., Starega-Roslan,J., Milanowska,K., Nowak,D., Kubiaczyk,E., Nowak,M., Majorek,K., Kaminska,K. and Krzyzosiak,W.J. (2006) In Clarke,N. and Sanseau,P. (eds), *microRNA: Biology, Function & Expression*. DNA Press, Eagleville, PA, pp. 95–110.
24. Krzyzosiak,W.J., Napierala,M. and Drozdz,M. (1999) In Barciszewski,J. and Clark,B.F.C. (eds), *RNA Biochemistry and Biotechnology*. Kluwer Academic Publishers, Dordrecht, pp. 303–314.
25. Morin,R.D., O'Connor,M.D., Griffith,M., Kuchenbauer,F., Delaney,A., Prabhu,A.L., Zhao,Y., McDonald,H., Zeng,T., Hirst,M. *et al.* (2008) Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res.*, **18**, 610–621.
26. Calabrese,J.M., Seila,A.C., Yeo,G.W. and Sharp,P.A. (2007) RNA sequence analysis defines Dicer's role in mouse embryonic stem cells. *Proc. Natl Acad. Sci. USA*, **104**, 18097–18102.
27. Wu,H., Neilson,J.R., Kumar,P., Manocha,M., Shankar,P., Sharp,P.A. and Manjunath,N. (2007) miRNA profiling of naive, effector and memory CD8 T cells. *PLoS One*, **2**, e1020.
28. Seitz,H., Ghildiyal,M. and Zamore,P.D. (2008) Argonaute loading improves the 5' precision of both MicroRNAs and their miRNA* strands in flies. *Curr. Biol.*, **18**, 147–151.
29. Azuma-Mukai,A., Oguri,H., Mituyama,T., Qian,Z.R., Asai,K., Siomi,H. and Siomi,M.C. (2008) Characterization of endogenous human Argonautes and their miRNA partners in RNA silencing. *Proc. Natl Acad. Sci. USA*, **105**, 7964–7969.
30. MacRae,I.J., Li,F., Zhou,K., Cande,W.Z. and Doudna,J.A. (2006) Structure of Dicer and mechanistic implications for RNAi. *Cold Spring Harb. Symp. Quant. Biol.*, **71**, 73–80.
31. MacRae,I.J., Zhou,K., Li,F., Repic,A., Brooks,A.N., Cande,W.Z., Adams,P.D. and Doudna,J.A. (2006) Structural basis for double-stranded RNA processing by Dicer. *Science*, **311**, 195–198.
32. Gan,J., Tropea,J.E., Austin,B.P., Court,D.L., Waugh,D.S. and Ji,X. (2006) Structural insight into the mechanism of double-stranded RNA processing by ribonuclease III. *Cell*, **124**, 355–366.
33. Lau,P.W., Potter,C.S., Carragher,B. and MacRae,I.J. (2009) Structure of the human Dicer-TRBP complex by electron microscopy. *Structure*, **17**, 1326–1332.
34. Wang,H.W., Noland,C., Siridechadilok,B., Taylor,D.W., Ma,E., Felderer,K., Doudna,J.A. and Nogales,E. (2009) Structural insights into RNA processing by the human RISC-loading complex. *Nat. Struct. Mol. Biol.*, **16**, 1148–1153.