Efficient processing of primary microRNA hairpins by Drosha requires flanking non-structured RNA sequences

Yan Zeng‡ and Bryan R. Cullen§
Center for Virology and Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, NC 27710, USA

Running Title: Processing human primary microRNAs

‡Present address: Department of Radiation Oncology, Virginia Commonwealth University, 401 College St., Richmond, VA 23298
§To whom correspondence should be addressed: Dept. of Molecular Genetics and Microbiology, Duke University Medical Center, Box 3025, Durham, NC 27710. Tel.: 919-684-3369; Fax: 919-681-8979; E-mail: culle002@mc.duke.edu.

Drosha is a member of the ribonuclease (RNase) III family that selectively process RNAs with prominent double-stranded features. Drosha plays a key role in the generation of precursor microRNAs from primary microRNA (pri-miRNA) transcripts in animal cells, yet how Drosha recognizes its RNA substrates remains incompletely understood. Previous studies have indicated that, within the context of a larger pri-miRNA, an approximately 80-nucleotide-long RNA hairpin structure is necessary for processing by Drosha. Here, by performing in vitro Drosha processing reactions with RNA substrates of various sizes and structures, we show that Drosha function also requires single-stranded RNA extensions located outside the pri-miRNA hairpin. The sequence of these RNA extensions was largely unimportant, but a strong secondary structure within the extension, or a blunt-ended pri-miRNA hairpin, blocked Drosha cleavage. The requirement for single-stranded extensions on the pri-miRNA hairpin substrate for Drosha processing is currently unique among the RNase III enzymes.

Ribonuclease (RNase) III family enzymes are expressed in both prokaryotes and eukaryotes and are involved in the processing, maturation, and degradation of a wide variety of RNAs, including ribosomal RNAs, transfer RNAs, small nuclear RNAs, small nucleolar RNAs, microRNAs (miRNAs), and small interfering RNAs (siRNAs) (1, 2). These RNases generate RNA products that feature an imperfect or perfect duplex with an ~2 nucleotide (nt) 3' overhang at the site of cleavage. This characteristic staggered 3' end structure results from the independent cleavage of the two RNA strands by the two catalytic sites located within a single double-stranded RNA (dsRNA) processing center formed by two RNase III domains. These RNase III domains may derive from two proteins, as seen with bacterial RNase III, or from a single protein, as seen with Dicer and Drosha (3, 4).

The RNase III family can be divided into four subclasses (1). Class 1 consists of bacterial enzymes with a minimal RNase III domain and a single dsRNA binding domain (dsRBD). Class 2 consists of fungal enzymes, such as Rnt1p in Saccharomyces cerevisiae and Pac I in...
Schizosaccharomyces pombe, which contain an extra N-terminal region with no recognizable motifs. Class 3 consists of the Drosha orthologs found in animals. These proteins have two RNase III domains and one dsRBD in the C-terminal half, and a Proline-rich domain and an Arginine-rich (R-rich) domain in the N-terminal half of the protein. Class 4 RNase III enzymes consist of the Dicer homologs expressed in Schizosaccharomyces pombe, plants, and animals. Their C-terminal half appears similar to Drosha, but the N-terminal half features different domain structures.

How these RNase III enzymes select and cleave their RNA substrates has been the subject of several studies. Escherichia coli RNase III, a class 1 enzyme, targets a broad spectrum of RNAs, apparently regulated only by RNA antideterminants, i.e., certain basepaired sequences at defined positions along a helical substrate are disfavored (5). Human Dicer, a class 4 enzyme, preferentially recognizes a 2 nt 3' overhang on a dsRNA, then cuts ~20 nt away to generate a short RNA duplex (3, 6, 7). Rnt1p, a class 2 enzyme, shows the highest specificity, as it selects for a NGNN tetraloop and cleaves 14-16 bp into the stem of the flanking RNA hairpin (8, 9). Another class 2 enzyme, Pac I, however, does not have such stringent requirements (10). Of note, while the enzymes mentioned above show clearly distinct substrate specificities, they are all capable of processing a blunt-ended dsRNA substrate effectively in vitro.

The class 3 RNase III Drosha, forms a complex with a protein partner, termed DGCR8 in humans and Pasha in flies and worms, that catalyzes the cleavage of long primary miRNA transcripts (pri-miRNAs) to produce the ~ 60 nt hairpin RNAs termed precursor miRNAs (pre-miRNAs) (4, 11-13). Pre-miRNAs are further processed by Dicer to yield mature, ~22 nt long miRNAs. miRNAs are abundant, endogenous, non-coding RNAs that post-transcriptionally regulate gene expression in multicellular organisms (14). Because Drosha produces pre-miRNAs that then serve as substrates for Dicer, and because Dicer primarily uses the terminal structure of the pre-miRNA hairpin created by Drosha cleavage to determine where it will subsequently cut, pri-miRNA cleavage by Drosha imparts much of the specificity of miRNA processing in animal cells.

It has been demonstrated that, for a pri-miRNA to be efficiently processed by Drosha, the targeted hairpin must consist of a large terminal loop of ≥ 10 nt and a stem region somewhat longer than the one present in the final pre-miRNA (11, 15, 16). In all previously reported experiments, which analyzed Drosha activities in vitro and miRNA expression in transfected cells, the miRNA-containing hairpin was always embedded within a longer transcript and thus surrounded by extra RNA sequences derived from either its endogenous flanking genomic sequence or from the expression vector used (4, 11, 15-18). One of these studies showed that, when transcribed from an RNA polymerase III promoter in transfected cells, at least 40 nt of additional sequence on each side of a pre-miRNA structure was required for efficient miRNA production (18). Some of the flanking nucleotides likely formed the short stem extension beyond the pre-miRNA stem that is known to be essential for pri-miRNA processing (11, 15, 16), but what role the rest of these extra sequences play remains unclear. Do they provide specific RNA sequences or structures that enhance processing or does this simply represent a requirement for flanking single-stranded RNA (ssRNA)? Furthermore, little is known about how Drosha interacts with the
various structural elements required for efficient pri-miRNA processing. To address these questions, we have analyzed how various flanking RNA sequences affect pri-miRNA hairpin recognition and cleavage by Drosha.

MATERIALS AND METHODS

Plasmid construction. pGEX-4T-1-Drosha RBD encodes part of the Drosha protein extending from leucine 1254 to its C-terminus and was made by amplifying the relevant DNA fragment with primers: 5'-CTGGAATTCTAGTTGAATCAGGATTGGAAT-3' and 5'-GCGCTCGAGTTATTTCTTGATGTCTTCAGT-3', digesting the PCR product with EcoRI and XhoI, and inserting it into the EcoRI and XhoI sites of pGEX-4T-1. pGEX-4T-1-Drosha R-rich encodes the sequence from proline 216 to leucine 333 of Drosha and was similarly made by PCR subcloning using primers: 5'-GCGAATTCCCAATGAGAGAAGGTCC-3' and 5'-GCCTCGAGCTAATTCTGGTGTGCATC-3'. pGEX-4T-1-DGCR8 2XRBD encodes a DGCR8 protein fragment extending from glutamic acid 502 to the C-terminus, and was made by amplifying a DNA fragment from a FLAG-DGCR8 expression plasmid (a gift of Dr. R. Shiekhattar) with primers: 5'-GCAGAAATCCCAGTGAGAGAAGGTCC-3' and 5'-GCCTCGAGCTAATTCTGGTGTGCATC-3'. pGEX-4T-1-DGCR8 2XRBD encodes a DGCR8 protein fragment extending from glutamic acid 502 to the C-terminus, and was made by amplifying a DNA fragment from a FLAG-DGCR8 expression plasmid (a gift of Dr. R. Shiekhattar) with primers: 5'-GCAGAAATCCCAGTGAGAGAAGGTCC-3' and 5'-GCCTCGAGCTAATTCTGGTGTGCATC-3'. pGEX-4T-1-DGCR8 2XRBD encodes a DGCR8 protein fragment extending from glutamic acid 502 to the C-terminus, and was made by amplifying a DNA fragment from a FLAG-DGCR8 expression plasmid (a gift of Dr. R. Shiekhattar) with primers: 5'-GCAGAAATCCCAGTGAGAGAAGGTCC-3' and 5'-GCCTCGAGCTAATTCTGGTGTGCATC-3'. pGEX-4T-1-DGCR8 2XRBD encodes a DGCR8 protein fragment extending from glutamic acid 502 to the C-terminus, and was made by amplifying a DNA fragment from a FLAG-DGCR8 expression plasmid (a gift of Dr. R. Shiekhattar) with primers: 5'-GCAGAAATCCCAGTGAGAGAAGGTCC-3' and 5'-GCCTCGAGCTAATTCTGGTGTGCATC-3'. pGEX-4T-1-DGCR8 2XRBD encodes a DGCR8 protein fragment extending from glutamic acid 502 to the C-terminus, and was made by amplifying a DNA fragment from a FLAG-DGCR8 expression plasmid (a gift of Dr. R. Shiekhattar) with primers: 5'-GCAGAAATCCCAGTGAGAGAAGGTCC-3' and 5'-GCCTCGAGCTAATTCTGGTGTGCATC-3'. pGEX-4T-1-DGCR8 2XRBD encodes a DGCR8 protein fragment extending from glutamic acid 502 to the C-terminus, and was made by amplifying a DNA fragment from a FLAG-DGCR8 expression plasmid (a gift of Dr. R. Shiekhattar) with primers: 5'-GCAGAAATCCCAGTGAGAGAAGGTCC-3' and 5'-GCCTCGAGCTAATTCTGGTGTGCATC-3'. pGEX-4T-1-DGCR8 2XRBD encodes a DGCR8 protein fragment extending from glutamic acid 502 to the C-terminus, and was made by amplifying a DNA fragment from a FLAG-DGCR8 expression plasmid (a gift of Dr. R. Shiekhattar) with primers: 5'-GCAGAAATCCCAGTGAGAGAAGGTCC-3' and 5'-GCCTCGAGCTAATTCTGGTGTGCATC-3'. pGEX-4T-1-DGCR8 2XRBD encodes a DGCR8 protein fragment extending from glutamic acid 502 to the C-terminus, and was made by amplifying a DNA fragment from a FLAG-DGCR8 expression plasmid (a gift of Dr. R. Shiekhattar) with primers: 5'-GCAGAAATCCCAGTGAGAGAAGGTCC-3' and 5'-GCCTCGAGCTAATTCTGGTGTGCATC-3'. pGEX-4T-1-DGCR8 2XRBD encodes a DGCR8 protein fragment extending from glutamic acid 502 to the C-terminus, and was made by amplifying a DNA fragment from a FLAG-DGCR8 expression plasmid (a gift of Dr. R. Shiekhattar) with primers: 5'-GCAGAAATCCCAGTGAGAGAAGGTCC-3' and 5'-GCCTCGAGCTAATTCTGGTGTGCATC-3'.

Preparation of recombinant Drosha. Human 293T cells were transfected with pCK-Drosha-FLAG, which expresses C-terminally FLAG-tagged human Drosha (11). Two days later, cell extracts were prepared in lysis buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, and 0.4 % NP-40) and mixed with anti-FLAG-agarose beads (Sigma) at 4°C for ~ 1 hour. Beads were washed four times with the same lysis buffer and once with reaction buffer (20 mM HEPES-KOH pH 7.6, 100 mM KCl, 0.2 mM EDTA, and 5% glycerol). Proteins were then eluted by incubating the beads with reaction buffer containing 150 ng/μl of 3x FLAG peptide (Sigma) at 4°C for ~ 30 minutes. Aliquoted supernatants were stored at -80°C.

Enzymatic assays. Drosha processing experiments were performed in the reaction buffer mentioned above supplemented with 2 mM DTT, 7 mM MgCl2, and ~0.5 unit/μl RNasin (Promega). Purified Drosha was mixed with ~105 cpm (~0.1 ng) of a 32P labeled RNA substrate and incubated at 37°C for 60-90 minutes. Reactions were terminated by adding an equal volume of 2x loading buffer (98% formamide, 20 mM EDTA, and 0.1% bromophenol blue), heated at 95°C for ~10 minutes, resolved on a 10% denaturing gel, and analyzed by autoradiography and/or on a PhosphorImager (Amersham). 5’-32P labelled φX174 HinfI DNA markers (Promega) were used as size standards. All RNA substrates were tested at least twice.

DNA templates for RNA probes were prepared by PCR using T7 promoter-added primers. RNA probes were in vitro transcribed (Promega) in the presence of [α-32P]CTP. For some of the probes, full length RNAs were gel isolated prior to use.

To prepare the circular pri-miR-31 substrate, the linear, labeled transcript was first synthesized as described above, dephosphorylated by alkaline phosphatase (Roche), purified, phosphorylated by T4 polynucleotide kinase (New England Biolabs) in the presence of ATP, and then treated with T4 RNA Ligase (New England BioLabs).
Biolabs). Products were separated on a 10% denaturing gel, and RNAs isolated from gel slices. Partial alkaline hydrolysis of RNA was performed by incubating the RNA in 0.1M NaHCO3 pH 9.0, along with 20 μg of yeast tRNA, at 95°C for 6 minutes, the RNAs were then precipitated with ethanol.

miRNA expression from a plasmid in transfected cells. DNA spanning the pre-miR-31 sequence was amplified from human genomic DNA (Clontech), digested with HindIII and XhoI, and then cloned into a modified pSuper vector (15). The constructs were then co-transfected with a plasmid expressing a control short hairpin RNA specific for green fluorescent protein (GFP) into 293T cells, RNA isolated two days later, and analyzed by Northern blotting as described (15, 16).

RNA binding to glutathione-S-transferase (GST) proteins. A DNA template for ssRNA substrate transcription was prepared by PCR amplification of a 43 bp multiple cloning site fragment derived from pCMV (17). The DNA template for pre-miR-30 transcription was prepared by annealing and extending the two oligonucleotides: 5’-TGTAATACGACTCTATAGGTAAAC ATCCTCGACTGGAAGCT-3’ and 5’-ACGGCAAAACATCCGACTGAAAGCCCATCATCG AGG-3’ (the putative loop region is underlined). The DNA template for pre-miR-30 L5 transcription was prepared by annealing and extending the two oligonucleotides: 5’-TGTAATACGACTCTATAGGTAAAC ATCCTCGACTGGAAGCT-3’ and 5’-ACGGCAAAACATCCGACTGAAAGCCCATCATCG AGG-3’ (the putative loop region is underlined, and the 10 nt 3’ extension is italicized). Other DNA templates and RNA probes were synthesized as described above.

For protein purification, DH5α cells transformed with pGEX-4T-1 or one of its derivatives were induced, lysed, and GST-proteins bound to glutathione beads (Amersham) as previously described (19). The beads were washed three times with 10 mM Tris pH 7.6, 0.5 M LiCl, and 0.1% TritonX-100, followed by one wash with binding buffer (20 mM Tris pH 7.6, 0.1 M KCl, 0.1% Tween-20, 0.1% TritonX-100), and then incubated with ~90 μl of binding buffer containing 40 units of RNasin, 10 mg/ml of poly(dI-dC) (Sigma), and various 32P-labeled RNA substrates (~10^4 cpm of each) at 4°C for ~25 minutes. Beads were afterwards washed four times with binding buffer. RNA was then eluted with 100 μl of 1% SDS, 0.15 M NaCl, and 30 μg of yeast tRNA, extracted with phenol/chloroform, and precipitated with ethanol. Bound RNA was analyzed by electrophoresis and autoradiography. In a parallel experiment, proteins bound to glutathione beads were directly analyzed by electrophoresis followed by Coomassie staining to confirm that the relevant proteins were indeed purified. All the experiments were performed at least two times with identical results.

RESULTS

Drosha-mediated cleavage of pri-miRNAs in vitro requires flanking ssRNA - We prepared FLAG epitope tagged Drosha enzyme by transfecting human 293T cells
with the plasmid pCK-Drosha-FLAG (11), purifying proteins with FLAG-agarose, and then eluting the immunoprecipitate (IP) with 3x FLAG peptide. The IP contained Drosha-FLAG and presumably also endogenous DGCR8 (4). For RNA substrates, we prepared transcripts encoding three human miRNAs, miR-31, miR-223, and miR-30a (miR-30 for short).  

32P-labelled RNA substrates, including various flanking sequences beyond the pre-miRNA hairpin, were mixed with the Drosha IP and tested for the generation of the ~60 nt pre-miRNA intermediate (marked by asterisks in Figs. 1-7). Figs. 1-3 list all the natural or near natural pri-miRNA variants tested along with their predicted secondary structures, and also show some of the representative autoradiographs. RNA substrates were named according to the numbers of extra nucleotides 5' and 3' to the predicted pre-miRNA cleavage product, e.g., miR-31(13+16) denotes a miR-31 variant with 13 nt flanking the 5' end, and 16 nt flanking the 3' end, of the predicted pre-miR-31 hairpin (Fig. 1). The RNA located outside of a pre-miRNA hairpin can be tentatively demarcated into two distinct components: one is the essential stem extension located immediately adjacent to the pre-miRNA hairpin, and the other is the extra RNA at the ends that presumably forms ssRNA extensions. Importantly, we found that these predicted ssRNA flanking sequences greatly facilitated Drosha cleavage.

Fig. 1 presents the results using pri-miR-31 variants. The miR-31(13+16) substrate was processed efficiently by Drosha, generating largely pre-miR-31 (lane 2) and very few processing intermediates (indicated by arrowheads). Judged from their characteristic sizes, these intermediates were RNAs cut by Drosha at the authentic 5' or 3' cleavage site, but not at both. Such products have been reported previously (4). These singly-cut RNAs likely represent dead-end products in vitro, since they lack an essential feature required for de novo processing (i.e. a stem extension beyond the Drosha cleavage sites, see below). When isolated from gels and treated with Drosha again, they were indeed totally resistant to cleavage (data not shown). Eliminating part of the 5' flanking ssRNA, as in miR-31(7+16), led to accumulation of singly-cut intermediates (Fig. 1, lane 6). Further deletion of part of the 3' ssRNA extension, as in miR-31(7+10) or miR-31(7+7), led to a further reduction in pre-miR-31 production (Fig. 1, lanes 7 and 8). When deletions were made even closer to the pre-miR-31 region, e.g. in miR-31(4+10) or (4+7), Drosha processing became undetectable (Fig. 1, lanes 9 and 10).

Very similar results were also obtained for miR-223 (Fig. 2) and miR-30 (Fig. 3). For example, compared with miR-223(29+21), the 3' ssRNA-shortened substrates miR-223(29+15) and miR-223(29+12) were much less efficiently cleaved by Drosha in vitro. Blunt-ended RNA hairpins were cleaved very poorly or not at all, such as miR-31(7+5) (Fig. 1), miR-223(16+15) (Fig. 2, and Fig. 5, see below) and miR-30(11+9) (Fig. 3, lane 2). In general, the longer the native flanking sequence the miRNA retained on both sides, the better substrate the RNA was for Drosha processing in vitro. However, flanking ssRNA at the 3' end did appear to be more critical than the one at the 5' end. Thus, miR-223(16+21) was a better Drosha substrate than miR-223(24+15), although neither was as good as miR-223(24+21) (Fig. 2). Finally, and consistent with previous reports (11, 15, 16), a stem extension beyond the pre-miRNA hairpin, within the longer pri-miRNA, was always essential, although the exact length requirement varied among the three miRNAs tested. miR-223 apparently requires the longest extension, as miR-223(13+30) and miR-223(13+21) were...
hardly processed by Drosha (Fig. 2). In contrast, miR-31(7+16) and miR-31(7+10) were still reasonable Drosha substrates (Fig. 1, lanes 6 and 7), while further shortening the stem extension to make miR-31(4+10) or miR-31(4+7) abolished cleavage (Fig. 1, lanes 9 and 10).

**Flanking ssRNA sequences function in a largely sequence-independent manner in vitro** - Figs. 1-3 show secondary structure predictions for miRNA hairpins based on MFOLD. The actual RNA folding details at the top of the hairpin and at the base of the stem, and the conformations of the flanking RNAs, might be dynamic and/or different from the RNA structures proposed here. For example, some residues from the 5' ssRNA extension could potentially form hydrogen bonds with those from the 3' side. Nevertheless, we hypothesized that Drosha preferred flanking RNA sequences that did not fold into a helical conformation. We performed several experiments to test this idea, and these results are shown in Figs. 4-6. In Fig. 4A, when the predicted 6 nt 5' ssRNA extension and 11 nt 3' extension in the natural but truncated miR-31(13+16) substrate were replaced by arbitrary sequences, the new RNA, called (6)+7+5+(11), was still effectively cleaved by Drosha in vitro (compare lane 2 and lane 6). miR-223 is another example (Fig. 4B). Here, the artificial (8)+16+15+(6) variant was cleaved at an efficiency close to that of the natural miR-223(24+21) substrate.

To examine flanking sequence requirements in more detail, we turned to miR-223(16+21) (Fig. 5). This RNA is predicted to form a simple structure containing a 6 nt 3' ssRNA overhang, and in vitro processing by Drosha was weaker than seen with miR-223(24+21), but still readily detectable (compare lane 2 with lane 12, Fig. 5). Deleting the 6 nt 3’ overhang eliminated detectable Drosha processing (lane 4), which was rescued by adding back 6 nt (lane 6) or 9 nt (lane 8), but not 3 nt (lane 10) of an arbitrary ssRNA sequence. This rescue was not simply due to larger RNA size, since introducing a 6 nt 5’ extension that is predicted to form a 6 bp stem with the arbitrary 6 nt 3’ extension abolished processing (data not shown). Furthermore, when an artificial hairpin was appended to the 6 nt 3’ extension to make 16+15+(6+D), the new RNA was processed less efficiently than the parental 16+15+(6) RNA (Fig. 6, compare lane 4 with lane 2), while an identically sized RNA substrate lacking the predicted hairpin structure, 16+15+(6+S), was processed more efficiently (Fig. 6, lane 6). From these results, we concluded that flanking ssRNA sequences strongly enhance Drosha cleavage of pri-miRNA hairpins but that the particular sequence of the ssRNA extensions was not critical as long as the flanking ssRNA sequences were of sufficient length (> 3nt) and adopted a largely single-stranded structure.

**Free RNA ends are not required for Drosha processing** - If Drosha first recognizes an unpaired 5’ end and/or 3’ end, and then scans along the RNA for a suitable stem-loop structure, such a mechanism could explain why a ssRNA extension is required. To test this scenario, we chose a pri-miR-31 substrate, which was larger than the ones examined above and contained an ~85 nt sequence flanking each side of the mature pre-miR-31. The linear RNA was a good substrate for Drosha processing in vitro (Fig. 7B, lane 2) and was properly processed when transcribed in transfected cells (15). We prepared its closed, circular version as described in Materials and Methods (Fig. 7A, lane 2). The identity of the circular RNA was confirmed by partial alkaline hydrolysis, as this treatment collapsed it to a position corresponding to the linear RNA and yielded a smear below it.
Drosha cleaved this circular RNA efficiently to generate pre-miR-31 and the predicted single byproduct (lane 5, Fig. 7B). Thus, free RNA ends are not necessary for Drosha processing of pri-miRNA hairpins in vitro.

miRNA expression in cells requires ssRNA sequences flanking the pri-miRNA hairpin. We next asked if RNA variants that were good substrates for Drosha cleavage in vitro were also good substrates for processing to a mature miRNA in cells. We inserted the corresponding DNAs behind an RNA polymerase III dependent promoter, the H1 promoter, transfected the resultant expression plasmids into human 293T cells, and then examined miRNA expression by Northern blotting. The vector we used (15) contributed ~8 nt at each side of the cloned RNAs (Fig. 8). We found that even longer sequences were required for, or at least enhanced, miRNA maturation in this setting. As shown in Figure 8, miR-31(51+51), which had 51 nt of natural RNA sequence flanking each side of pre-miR-31 (substrate a), yielded a high level of mature miR-31, miR-31(51+16) and miR-31(13+51) had reduced levels of miR-31 production (substrates b and c), while miR-31(13+16) gave only small amounts of mature miR-31 (substrate d). We then added artificial sequences to this shorter pri-miRNA transcript to bring it back to the size of miR-31(51+51). These sequences were designed so that they were clearly different from the natural ones and also so that they would not form a strong secondary structure. Adding an arbitrary 38 nt sequence at the 5’ side of miR-31(13+51) to make miR-31(38+13+51) largely restored miR-31 expression (compare substrates e and f). However, making miR-31(51+16+35) and miR-31(38+13+16+35) only partially restored miR-31 expression (substrates e and g). The failure to completely rescue miR-31 expression by adding back artificial sequences in this latter instance could be due to the loss of a positive contribution from the natural flanking sequences or due to the introduction of negative effects by the new flanking ssRNAs.

The R-rich region of the Drosha subunit preferentially binds ssRNA – We have previously shown that Drosha prefers to process pri-miRNA hairpins bearing a large, ssRNA terminal loop (15), and here, we have further demonstrated that ssRNA extensions are required for Drosha cleavage in vitro. To identify which part(s) of the Drosha:DGCR8 complex interacts with ssRNA, we expressed and purified individual domains of the Drosha and DGCR8 subunits as GST-fusion proteins in E. coli and tested their interaction with various RNA substrates. Fig. 9 presents the results obtained using the dsRBDs of Drosha and DGCR8 and the R-rich domain of Drosha. Fig. 9A shows the domain structures of the proteins, and Fig. 9B lists the different RNA substrates used in the binding experiments. Substrate a is a 43 nt RNA derived from a vector sequence and is used here as a representative of ssRNA. Mfold predicts that it contains no consecutive helical RNA region longer than 5 bp. Substrate b is the native pre-miRNA for miR-30, which is a 63 nt hairpin bearing a 2 nt 3’ overhang (11). Substrate d, the L5 variant, is similar to substrate b, but, due to substitution (15), it contains a small 5 nt terminal loop instead of the predicted 15 nt loop (see Materials and Methods). Compared with wild-type pri-miR-30, a pri-miRNA bearing the L5 mutant is processed much less efficiently by Drosha (15), underscoring the importance of a large terminal loop. Substrate e, L5+10, has 10 nt of arbitrary ssRNA sequence appended to the original 3’ overhang of substrate d.
Substrates c and f are the same pri-miR-223 RNA substrates listed in Fig. 2.

All the Drosha and DGCR8 protein fragments were expressed as GST-fusions in bacteria (Fig. 9D), with the Drosha RBD and the DGCR8 2XRBDs being expressed at a higher level than the R-rich region of Drosha. GST-DGCR8 2XRBDs bound avidly to all RNAs tested (Fig. 9C, lanes 7 and 14). Single, individual dsRBDs of DGCR8 bound RNA as well as the 2XRBDs (data not shown). The GST-Drosha R-rich domain fusion, however, pulled down little or no substrate d (Fig. 9C, lane 13, L5 RNA). The GST-Drosha R-rich domain fusion did, however, interact very well with all other substrates tested, such as the putative ssRNA (substrate a) and hairpin RNAs with a wildtype, presumably large and flexible, terminal loop (substrates b, c, and f) or a 3’ ssRNA extension, as in substrates e and f. GST alone or GST-Drosha RBD did not bind any RNA under these conditions, and other Drosha truncations were expressed too poorly in E. coli for us to test their RNA binding potential. Curiously, if poly(dI-dC) was omitted in the binding reactions, Drosha RBD could then exhibit RNA binding activity (data not shown).

**DISCUSSION**

The principal finding of our current study is that efficient Drosha processing of a pri-miRNA substrate in vitro needs a substantial ssRNA flanking sequence attached to an extensive, ~80 nt pri-miRNA stem-loop structure. Such a requirement for flanking ssRNA sequences has not been observed for other RNase III type enzymes. E. coli RNase III, yeast Rnt1p and PacI can all cleave dsRNAs with basepaired 5’ and 3’ ends (10, 20). While Dicer clearly prefers a 2 nt 3’ overhang, human Dicer nevertheless cleaves dsRNAs containing a blunt end with only a small drop in efficiency in vitro (3, 7). Even allowing for some uncertainty at the terminal structure, such as breathing, our data indicate that Drosha requires much longer ssRNA sequences flanking its dsRNA substrate than do other RNase III enzymes.

Drosha does not cleave a fully helical RNA (21), thus suggesting that ssRNA is involved in mediating this protein:RNA interaction. We previously showed that Drosha strongly prefers a large, unstructured terminal loop on its pri-miRNA substrates (15), which together with the essential ssRNA overhangs identified in this study may thus satisfy the predicted requirement for single-stranded RNA for Drosha cleavage. We believe we have now largely defined the RNA elements that are essential for Drosha cleavage in vitro, and it is apparent that Drosha actually engages a very large RNA surface. Unlike other RNase III enzymes, which can function alone and recognize smaller and simpler RNA structures, at least in vitro, pri-miRNA processing is actually mediated by a protein complex that minimally consists of the catalytic subunit Drosha and a protein partner called DGCR8 in humans or Pasha in invertebrates (4, 12, 13). DGCR8 greatly enhances, and is likely indeed necessary, for Drosha activity. There is little information as to how these proteins interact with their RNA substrates. As the first step towards achieving such an understanding, we show here that, under our assay conditions, the dsRBDs of DGCR8 were capable of binding to both RNA with a largely single-stranded conformation and to RNA with a mostly helical structure (Fig. 9). While the dsRBD of Drosha showed a very low affinity for RNA, interestingly we found that the R-rich region of the Drosha subunit had a preference for ssRNA. For the R-rich region, the ssRNA can be either at the top of the stem, i.e., in the terminal loop, or flanking.
the base of the RNA hairpin (Fig. 9). A 2 nt 3’ overhang together with a small terminal loop, as seen in the pre-miR-30 L5 variant, is insufficient to support binding. The Drosha R-rich region is not the sole determinant on the enzyme that requires ssRNA, as it does not distinguish between miR-223(24+21), a good substrate for Drosha cleavage \textit{in vitro}, and miR-223(16+15), whose cleavage by Drosha was never observed. Contributions to recognition of ssRNA regions within the pri-miRNA substrate may thus also come from other, as yet undefined, parts of the Drosha/DGCR8.

A recent paper (4) indicated that a Drosha mutant devoid of the R-rich region analyzed here could still function as an active pri-miRNA processing enzyme when transiently expressed in cells. The Drosha deletion mutant was expressed at a much higher level than the full length protein (4), so it will be interesting to see if it indeed retains the same specific cleavage activity and substrate specificity as the full length enzyme. Since Drosha can self-associate, it is formally possible that endogenous Drosha protein could form a complex with the overexpressed Drosha deletion mutant and then exhibit cleavage activity \textit{in vitro}. Our data do not address the question of what role the R-rich domain plays in an intact Drosha protein in cells, and the mechanisms governing Drosha-RNA interactions certainly need to be investigated further.

For many pri-miRNAs, RNA folding algorithms predict that sequences at the 5’ side and the 3’ side, beyond the pre-miRNA hairpin, can anneal to form a very long, imperfect stem. A modest stem extension adjacent to the pre-miRNA intermediate is indeed essential for the excision of the pre-miRNA intermediate from a pri-miRNA substrate, but a longer dsRNA conformation is not beneficial and can be inhibitory (11, 15, 16; and this study). The exact sequence of the ssRNA extension is apparently not critical, but a strong RNA secondary structure within the flanking sequence, or formed between the 5’ and 3’ extensions, is distinctively disfavored (Figs. 4-6). Although it is currently unclear why ssRNA extensions are needed, or how they regulate Drosha function, we favor the hypothesis that the flanking ssRNA sequences form part of the Drosha:RNA interface, e.g., Drosha may simultaneously bind to the stem-loop structure as well as to the overhang(s). Alternatively, Drosha may be intrinsically unable to bind or correctly position itself directly onto a hairpin structure. It is conceivable that the extra flanking sequences may be needed initially to tether or recruit the Drosha:DGCR8 complex to RNA. In the absence of a suitable stem-loop structure nearby, however, the enzyme may rapidly dissociate from ssRNA binding sites. As Drosha is capable of cleaving a circular substrate (Fig. 7), we can exclude the possibility that the RNA overhang contributes a free 5’ or 3’ end necessary for Drosha function. Considering that many miRNAs are encoded within the introns of their host genes (22), our data are consistent with the prediction that Drosha can operate directly on lariat RNAs. It is also, of course, possible that Drosha cleavage \textit{in vivo} might be facilitated by other, as yet unknown, proteins.

We found that the minimal RNA element required for \textit{in vitro} Drosha cleavage, identified here as an ~80 nt RNA hairpin structure plus ~10 nt ssRNA overhang(s), was ineffective in mature miRNA production when transcribed from the H1 promoter in transfected cells (Fig. 8). This is consistent with an earlier report that at least 40 nt of extra sequences on each side of the pre-miRNA hairpin are required for
efficient miRNA production in cells (18). Part of these 40 nt extra RNA would form the stem and the ssRNA extensions essential for Drosha recognition and cleavage in vitro (11, 15, 16), but how the additional nucleotides contribute to pri-miRNA processing remains unknown. Drosha might need an even larger RNA structure for cleavage in vivo. Alternately, the extra RNA sequences might affect transcription, RNA folding and/or RNA stability. Most miRNAs are transcribed from RNA polymerase II promoters in vivo, so it is also possible that transcription from the H1 promoter, an RNA polymerase III promoter, can inhibit Drosha function. A more detailed analysis of sequence requirements for miRNA processing in vivo will be required to fully address these questions.

Acknowledgments - We thank N. Kim and R. Shiekhattar for providing FLAG-tagged Drosha and DGCR8 expression plasmids, respectively.

REFERENCES


\textbf{FOOTNOTES}

\*This work was supported by NIH grant GM071408 to Bryan Cullen.

\textsuperscript{1}The abbreviations used are: RNase, ribonuclease; miRNA, microRNA; siRNA, small interfering RNA; nt, nucleotide(s); dsRNA, double-stranded RNA; dsRBD, dsRNA binding domain; R-rich, Arginine-rich; pri-miRNA, primary miRNA transcript; pre-miRNA, precursor miRNA; ssRNA, single-stranded RNA; DTT, dithiothreitol; GFP, green fluorescent protein; GST: glutathione S-transferase.

\textbf{FIGURE LEGENDS}

FIG. 1. \textbf{Schematic of miR-31 transcripts and their susceptibility to Drosha cleavage \textit{in vitro}.} Arrows indicate the putative Drosha cleavage sites that liberate the pre-miR-31 RNA hairpin. The hairpin symbol below represents pre-miR-31. RNA variants are named based on the numbers of extra nucleotides 5’ and 3’ to the pre-miRNA. For example, 13+16 means the variant has 13 nt extra on the 5’ side and 16 nt on the 3’ side, and 13+10 means the variant has 13 nt extra on the 5’ side and 10 nt on the 3’ side. A residue different from the endogenous transcript is underlined. A series of + and – signs are used to denote the efficiencies with which Drosha cleaved its substrates. “-”: no cleavage at all; “++++”: the highest level of processing; reduced levels were judged based on reduced pre-miRNA production and increased intermediate accumulation. RNA variants with their original \textit{in vitro} processing data shown in the bottom panel are indicated as substrates a, b, c, d, e, f, and g. The size markers to the left of the autoradiograph were DNAs. Asterisks indicate the position of pre-miR-31, and arrowheads point to processing intermediates.

FIG. 2. \textbf{Drosha processing of miR-223 transcripts \textit{in vitro}.} See Fig. 1 legend for the meaning of symbols and labels.

FIG. 3. \textbf{Drosha processing of miR-30 transcripts \textit{in vitro}.} See Fig. 1 legend for explanation of symbols and labels. Underlined letters at the overhangs represent nucleotides different from the endogenous pri-miR-30, and they were changed to fit the transcription start site used by T7 RNA polymerase or for maintaining secondary structures.

FIG. 4. \textbf{Sequence of the ssRNA extension can be replaced without significantly affecting cleavage.} A, the 6 nt 5’ overhang and 11 nt 3’ overhang of pri-miR-31(13+16) were substituted with arbitrary sequences (underlined) to make the (6)+7+5+(11) mutant. The hairpin symbol represents pre-miRNA. Three RNA substrates, i.e. 13+16, 7+5, and (6)+7+5+(11) (substrates a, b, and c) were subjected to \textit{in vitro} processing by Drosha, as shown in the right panel. Size of...
DNA markers is shown to the left of the autoradiograph, and the asterisk indicates the pre-miRNA band. B, The native pri-miR-223(24+21) overhangs were likewise replaced to make (8)+16+15+(6). Labels are the same as in A. Cleavage percentage was calculated as the ratio of the intensity of the pre-miRNA band (marked by an asterisk) divided by that of the remaining full length substrate, and corrected for cytosine contents.

FIG. 5. Size requirement for the 3’ overhang of the pri-miR-223 transcript. Six RNA substrates (a, b, c, d, e, and f) containing a 6 nt natural 3’ overhang, no overhang, 6 nt, 9 nt or 3 nt of an arbitrary 3’ ssRNA overhang (underlined), or extensions at both the 5’ and 3’ sides, were examined for Drosha processing in vitro. Labeling is the same as in Fig. 4.

FIG. 6. Drosha cleavage of miR-223 substrates with different 3’ extensions. As shown, substrate b, 16+15+(6+D), has a predicted 8 bp stem and a 4 nt loop, flanking the pri-miRNA hairpin, while substrate c, 16+15+(6+S), is predicted to have a single-stranded 3’ extension. Cleavage efficiencies were calculated and labeled is the same as in Fig. 4B.

FIG. 7. Processing of linear and circular substrates. A, a linear pri-miR-31 transcript and a self-ligated derivative were fractionated on a 10% denaturing gel. Bands a and b, indicated by arrows, were excised from the gel and the RNA eluted. B, RNAs from bands a and b in panel A were subjected to Drosha cleavage in vitro (lanes 2 and 5) or to partial alkaline hydrolysis (lanes 3 and 6). Different predicted RNA species are diagrammed at the right. Pre-miR-31 is represented by a light colored bar, and the flanking sequences by darker colored bars.

FIG. 8. Mature miR-31 expression in transfected cells. The pri-miR-31 substrates a-g were transcribed from plasmids and contain different sequences flanking the predicted pre-miR-31 intermediate. Underlined are the elements common in all substrates. Sequences at the 5’ and 3’ ends are from the vector. N1 and N2 represent natural sequences flanking miR-31(13+16) (Fig. 1). A1 and A2 represent arbitrary sequences. The bottom panels show the results of a Northern analyses for miR-31, with a GFP siRNA used as control, using RNA derived from transfected 293T cells. Sizes of DNA markers are shown at the left. Relative expression was calculated as the mature miR-31 signal divided by that of the GFP siRNA signal, with cleavage of substrate (a) set as 100%.

FIG. 9. RNA binding to individual domains of Drosha and DGCR8. A, domain structure of the Drosha and DGCR8 subunits. The R-rich region and the dsRBD domain of Drosha were separately expressed as GST-fusions in bacteria. Part of the DGCR8 from glutamic acid 502 to the end of the protein, which includes the two dsRBDs, was also expressed as a GST-fusion in bacteria. B, list of the RNA substrates (a, b, c, d, e, and f) that were tested for protein binding in C. For predicted RNA structure details, see Materials and Methods and Fig. 2. C, RNA binding to GST-proteins immobilized on glutathione beads. Lanes 1, 2, 3, 8, 9, and 10 show ~ 1/3 of the total input RNA. Lanes 4-7 and 11-14 show RNAs bound by recombinant proteins in vitro. For lanes 4-7, RNA substrates a, b, and c were mixed in a single solution and incubated with the proteins. For lanes 11-14, RNA substrates d, e, and f were also mixed and incubated with the proteins. Similar results were obtained when only a single RNA was used for binding. D, Coomassie staining of a protein gel to show that the expected GST-fusion proteins were indeed made.
miR-31

5' - CUGU UCGGA GGAGAG CCAAA AUGGCAUUGGC GUUU C
3' - GACG AGUCU CCGUUC CUGUU UCAC ACGGCUG UCC CAA U

a. 13+16
   G C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C

b. 13+10
   G
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C

13+7
   G
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C

c. 7+16
   G
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C

7+10
   G
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C

d. 7+10
   G
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C

e. 7+7
   G
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C

7+5
   G
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C

f. 4+10
   G
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C

g. 4+7
   G
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C
   C

Substrate | Drosha  
---|---
| a | b | c | d | e | f | g |
100 nt | - | + | - | + | + | + |
82 nt  |  Δ |  Δ |  Δ |  Δ |  Δ |  Δ |
66 nt  |  * |  * |  * |  * |  * |  * |

Fig. 1
miR-30

5' - GGUAAUA CU A A → UC A
UU G UU GA CAUG UGC CUGUAAACAUCC GACUGAAGCUC GACUGCUGAGG

3' - GG UU G CC C C

27+22
GGUAAUA CU A A
UU G UU GA CAUG UGC C

27+15
GA CAGUG GCG C
CU GUCAU GU 
GG CC C C

22+15
GU GUGCUUU A A
GA CAGUG GCG C
CU GUCAU GU 
GG CC C C

22+15
GU GUGCUUU A A
GA CAGUG GCG C
CU GUCAU GU 
GG CC C C

22+15
GU GUGCUUU A A
GA CAGUG GCG C
CU GUCAU GU 
GG CC C C

22+15
GU GUGCUUU A A
GA CAGUG GCG C
CU GUCAU GU 
GG CC C C

22+15
GU GUGCUUU A A
GA CAGUG GCG C
CU GUCAU GU 
GG CC C C

Substrate | a | b | c | d
---|---|---|---|---
Drosha | + | + | + | +

100 nt
82 nt
66 nt

Fig. 3
Fig. 4
Fig. 5
Fig. 6
Fig. 7
a. 51+51: 5'-AAGCUU-N1-miR-31(13+16)-N2-CUGCAGUU-3'
b. 51+16: 5'-AAGCUU-N1-miR-31(13+16)-CUGCAGUU-3'
c. 13+51: 5'-AAGCUU-miR-31(13+16)-N2-CUGCAGUU-3'
d. 13+16: 5'-AAGCUU-miR-31(13+16)-CUGCAGUU-3'
e. 51+16+35: 5'-AAGCUU-N1-miR-31(13+16)-A1-CUGCAGUU-3'
f. 38+13+51: 5'-AAGCUUA2-miR-31(13+16)-N2-CUGCAGUU-3'
g. 38+13+16+35: 5'-AAGCUUA2-miR-31(13+16)-A1-CUGCAGUU-3'

N1: CAUAACAAACGAAGGAGGUAUGUCCUCGUUCAUCUC
N2: AGCUUGGCUCUCCCUCCUCCUCCUCCGCUU
A1: CCGAUGCAGAAACAC UCUGUACACUUCUGAUG
A2: GUACAAGGAGAAGACUAAGCAGAUGCACAUUCAACU

**Substrate**

<table>
<thead>
<tr>
<th>Substrate</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
<th>g</th>
<th>mock</th>
</tr>
</thead>
<tbody>
<tr>
<td>118 nt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>66 nt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22 nt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Relative expression (%) | 100 | 38 | 32 | 13 | 48 | 84 | 21 |

miR-31
Northern

GFP siRNA
Northern

**Fig. 8**
**Fig. 9**

### Table A

<table>
<thead>
<tr>
<th>Drosha</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-rich</td>
</tr>
<tr>
<td>R-rich</td>
</tr>
<tr>
<td>Rnaselli</td>
</tr>
<tr>
<td>Rnaselli</td>
</tr>
</tbody>
</table>

### Table B

- a: ssRNA
- b: pre-miR-30
- c: miR-223(16+15)
- d: pre-miR-30 L5
- e: pre-miR-30 L5+10
- f: miR-223(24+21)

### Table C

**RNA**

<table>
<thead>
<tr>
<th>a</th>
<th>b</th>
<th>c</th>
</tr>
</thead>
</table>

**GST pulldown**

**RNA**

<table>
<thead>
<tr>
<th>d</th>
<th>e</th>
<th>f</th>
</tr>
</thead>
</table>

**GST pulldown**

<table>
<thead>
<tr>
<th>1/3 of input</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

### Table D

**Protein standards (kD)**

- 70
- 60
- 50
- 40
- 30

**Coomassie stain**
Efficient processing of primary microRNA hairpins by Drosha requires flanking non-structured RNA sequences
Yan Zeng and Bryan R. Cullen

J. Biol. Chem. published online June 1, 2005

Access the most updated version of this article at doi: 10.1074/jbc.M504714200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/early/2005/06/01/jbc.M504714200.citation.full.html#ref-list-1