Extensive terminal and asymmetric processing of small RNAs from rRNAs, snoRNAs, snRNAs, and tRNAs

Zhihua Li1,2, Christine Ender3, Gunter Meister3, Patrick S. Moore2, Yuan Chang2 and Bino John1,2,*

1School of Medicine, University of Pittsburgh, Pittsburgh, PA 15260, 2Cancer Virology Program, Hillman Cancer Center, 5117 Centre Avenue, Pittsburgh, PA 15213, USA and 3Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany

Received September 2, 2011; Accepted March 22, 2012

ABSTRACT

Deep sequencing studies frequently identify small RNA fragments of abundant RNAs. These fragments are thought to represent degradation products of their precursors. Using sequencing, computational analysis, and sensitive northern blot assays, we show that constitutively expressed non-coding RNAs such as tRNAs, snoRNAs, rRNAs and snRNAs preferentially produce small 5' and 3' end fragments. Similar to that of microRNA processing, these terminal fragments are generated in an asymmetric manner that predominantly favors either the 5' or 3' end. Terminal-specific and asymmetric processing of these small RNAs occurs in both mouse and human cells. In addition to the known processing of some 3' terminal tRNA-derived fragments (tRFs) by the RNase III endonuclease Dicer, we show that several RNase family members can produce tRFs, including Angiogenin that cleaves the TΨC loop to generate 3' tRFs. The 3' terminal tRFs but not the 5' tRFs are highly complementary to human endogenous retroviral sequences in the genome. Despite their independence from Dicer processing, these tRFs associate with Ago2 and are capable of down regulating target genes by transcript cleavage in vitro. We suggest that endogenous 3' tRFs have a role in regulating the unwarranted expression of endogenous viruses through the RNA interference pathway.

INTRODUCTION

Recent applications of deep sequencing methods have led to the identification of a surprising diversity of non-protein-coding RNAs (ncRNAs), including degradation-like small RNA fragments derived from miRNAs (1,2), snoRNAs (3–5) and tRNAs (6–12). Evidence is accumulating that these small RNA fragments are precisely processed to participate in diverse biological processes and are conserved in distantly related species (2,3,5–7,13). The first report of a small snoRNA-derived RNA and its function as a miRNA was reported 3 years ago (3). Since then, ~10 additional snoRNA-derived small (~18 nt) RNAs that can function as miRNAs have been described (5). We have also reported a group of human and viral unusually small (~15 nt) RNAs (usRNAs) that are derived from miRNAs as well as other non-coding regions, and can regulate gene (e.g., RAD21) expression (2). Production of these usRNAs are evolutionarily conserved (13,14) and are associated with hippocampal functions in mice (13).

It is common to find small RNA fragments that match long RNA transcripts such as mRNAs in small RNA sequencing data (8). The presence of such small RNA fragments is generally attributed to the extreme sensitivity of deep sequencers in detecting low-abundance RNAs including degradation products. However, even very low-abundance RNAs, as low as four copies per cell, originating from the Cyclin D1 (CCND1) promoter region are reported to have regulatory functions (15). The presence of such small RNA fragments is generally attributed to the extreme sensitivity of deep sequencers in detecting low-abundance RNAs including degradation products. However, even very low-abundance RNAs, as low as four copies per cell, originating from the Cyclin D1 (CCND1) promoter region are reported to have regulatory functions (15). While many of the RNA fragments in deep-sequencing data frequently correspond to low abundance, potentially non-functional RNA products, RNA fragments found at even greater abundance
than most miRNAs are also frequently observed in small RNA sequencing data (2,7). Common sources for such RNA fragments are tRNAs and rRNAs. Since tRNAs and rRNAs constitute most of the cellular RNA, it is reasonable to assume that these RNAs generate much more degradation products than other RNAs, resulting in the preferential detection of their degradation products in deep-sequencing. While the notion of degradation products is appealing, it was recently shown that tRNAs and rRNAs undergo stress-induced cleavage to produce stable RNA products, and this mechanism is conserved from yeast to human cells (11). In human cells, tRNA fragments have been reported to induce transient translational arrest (12). Various tRNA fragments are produced in cells under stress; notably during starvation in Thermophila (9) and serum deprivation in Giardia lamblia (10), or during development in Aspergillus fumigatus (16). Stress-induced tRNA cleavage is mediated by Rny1p in yeast (17) and angiogenin (ANG) in humans (12,18). These stress-induced tRNA cleavage products are 30–50 nt in length and seem to form related RNA classes, termed sitsRNA (10), tRNA (12,19) and tRNA halves (11,16,18). An additional class of small RNAs that are ~20–30 nt long, and are derived from 5’ or 3’ ends of tRNAs or from the genomic region following the 3’ end of tRNAs, broadly termed tRNA fragments (tRFs) were also reported (7). Subsequently, another class of seemingly related tRNA-derived small RNAs (tsRNAs) was found to function similarly to miRNAs. These tsRNAs are processed by either Dicer or RNase Z, depending on the tRF locations within both the mature tRNAs and their precursors (6). In summary, the widespread occurrence of stable small RNAs and their emerging roles in cellular processes indicate that we cannot ignore such molecules as randomly generated degradation products. Systematic analysis of all sequence reads from deep sequencing data may prove useful in defining new biological features and mechanisms relating to both canonical and non-canonical forms of small RNAs.

In this study, we found that the majority of the ~20 nt long fragments deriving from the mature sequences of the widely expressed tRNAs, rRNAs, snoRNAs and snRNAs, but not mRNAs, are produced in a specific cleavage pattern from the 5’ or 3’ ends. We note that this pattern was previously observed for snoRNAs (3–5), and tRNAs (6–8). Similar to the processing of miRNAs, these terminal RNAs accumulate in the cell in an asymmetric manner that favors the expression of either the 5’ or the 3’ end fragments. The terminal fragments, are largely independent of canonical RNA interference (RNAi) processing machineries, which include Dicer that processes precursor miRNAs (pre-miRNAs), and DGCR8, which partners with Drosha to form the microprocessor complex that processes primary mRNA transcripts to pre-miRNAs. The biogenesis of tRFs is likely to depend on multiple RNase family members, including ANG that cleaves tRNAs within the TψC loop in a sequence-specific manner. Notably, antisense sequences of 3’-terminal tRFs are highly enriched in genomic regions harboring retroviral sequences, and 3’ tRFs are found to cleave target RNA through endogenous association with Ago2.

### Materials and Methods

#### Small RNA sequence analysis

Small RNAs reads from the B lymphoma BCP1 cell line were processed and mapped to the transcriptome as reported earlier (2). Small RNA sequences from mouse embryonic stem cells and HEK293 cells were retrieved from NCBI (GSE 12521 and GSE 16579). A total of 522 tRNA genes were curated from gtrNAdb (20), excluding 109 predicted pseudo tRNA genes. Perfect matches between reads and tRNAs were imposed to obtain the most reliable list of tRNA-derived RNAs. To eliminate any bias in our analysis due to terminal CCA motifs that are known to occur in tRNAs, the terminal CCA of tRNAs were masked. For comparison of the abundance of miRNA and different classes of terminal RNA in Dicer or DGCR8 knockout libraries, reads from different libraries were normalized using the total number of mRNA fragments in each library. For normalization, the total number of mRNA fragments (≥15 nt) in each library was used to first divide the read counts of each distinct RNA in that library, and then scaled by multiplying by the total number of mRNA reads in the reference (wild-type) library. To focus on the most reproducible set of tRFs, a threshold of 10 reads/million was used for the analysis of tRFs (Figures 3A, 4B, C and 8B; Supplementary Figure S5).

To detect putative tRNA orthologs, we identified 137 human and mouse tRNA pairs with significantly high similarity (>80%) over >90% of the tRNA sequences. For analysis of retrotransposons, coordinates for human LTR, LINE and SINE elements were downloaded from UCSC genome database (21). To be considered as a potential tRF binding site, we required perfect match of the antisense sequence to the LTR, LINE and SINE elements. To eliminate artifacts due to the different sizes of LTR, LINE and SINE regions, the number of putative binding sites was normalized with respect to the total length of each of those elements. Since the length of the regions spanned by LINE and SINE elements were larger than LTR elements, the appropriate normalization factors of 0.42 and 0.67 were used to multiply the number of binding sites in LINE and SINE elements, respectively. All tRF sequences identified in BCP1 and were at least 10 nt long were used for the analysis. Since tRNAs can share similar terminal sequences, the terminal sequences that match to the most anti-LTR tRFs were selected as representative regions for illustrations.

#### Northern blot analysis

Northern blots were based on our recently published protocol (22). Briefly, total RNA (10–50 µg) was separated on 18% denaturing polyacrylamide gels and electrotransferred to positively charged nylon membranes from Roche (Roche Applied Science, Indianapolis, IN, USA). LNA–DNA mixed oligonucleotide probes were synthesized by IDT (Integrated DNA Technologies, Inc. Coralville, IA, USA). Probes were labeled with the small molecule Digoxigenin (DIG) using Oligo End Tailing Kit (Roche Applied Science, Indianapolis, IN, USA). Hybridization
was carried out overnight in Ultrahyb buffer (Ambion, Austin, TX, USA) at 42°C and washing was performed in two rounds of 0.1 × SSC/0.1% SDS at 60°C. DIG signals were detected with alkaline phosphatase-conjugated anti-DIG antibody (Cat. 11093274910, Roche). Probes used for northern blotting corresponds to (LNAs in bold): 5′ end tRF of tRNA LysTTT (5′-TGGCCGCCCAACCGAGA-3′); 5′ end tRF of tRNA LysTTT (5′-CTGAGCCTACCGGGAAA-3′); 5′ end tRF of tRNA PheGAA (5′-GGACCGGAAACCCG-3′); 5′ end tRF of tRNA PheGAA (5′-CTCCACAAGCTACCGGGA-3′); miR-17 control RNA (5′-CTACCTGACTGTAAGCCTTGG-3′), 5.8S rRNA (5′-TGATCCACGCTAAGGT-3′); U6 snRNA (5′-ATATGGGACGCTTACGGAATT-3′).

For Ago2-assocciations studies, immunoprecipitated RNA and 5μg total RNA isolated from HEK293 cells using Trisfast (Peqlab) was separated by 12% denaturing RNA PAGE and transferred to a nylon membrane (GE Healthcare) for northern blotting corresponds to (LNAs in bold):

Anti-DIG antibody (Cat. 11093274910, Roche). Probes used for northern blotting corresponds to (LNAs in bold): 5′ end tRF of tRNA LysTTT (5′-TGGCCGCCCAACCGAGA-3′); 5′ end tRF of tRNA LysTTT (5′-CTGAGCCTACCGGGAAA-3′); 5′ end tRF of tRNA PheGAA (5′-GGACCGGAAACCCG-3′); 5′ end tRF of tRNA PheGAA (5′-CTCCACAAGCTACCGGGA-3′); miR-17 control RNA (5′-CTACCTGACTGTAAGCCTTGG-3′), 5.8S rRNA (5′-TGATCCACGCTAAGGT-3′); U6 snRNA (5′-ATATGGGACGCTTACGGAATT-3′).

**In vitro cleavage assay**

Recombinant human ANG (Sigma-Aldrich) was dissolved and diluted to a final concentration of 10μg/ml in 30 mM HEPES, 30 mM NaCl and 0.01% BSA. The working buffer in absence of ANG was used as the control buffer. RNase A was purchased from Qiagen (Valencia, CA, USA). RNase I and RNase T1 were purchased from Ambion (Carlsbad, CA, USA). Incubations of total RNA were carried out at 37°C for 0, 1, 2 and 4 h. Incubations of in vitro transcribed RNA were carried out at 37°C for 0.5 h. The incubation was stopped by adding Trizol (Invitrogen) into the reaction mixture from which RNA was recovered by ethanol precipitation. Ago complex-containing anti-FLAG beads were incubated in a reaction containing 5μg target RNA, 1μM ATP, 0.2 mM GTP, 10μU/ml RNasin (Promega), 1μM 5′-adenosyl-Met (500μM), 1μM DTT (1M), 1μl Guanylintransferase and RNA from in vitro transcription reaction was incubated for 3 h at 37°C. RNA was purified by denaturing RNA–PAGE and recovered by ethanol precipitation. Ago complex-containing anti-FLAG beads were incubated in a reaction containing 5μl target RNA, 1μM ATP, 0.2 mM GTP, 10μU/ml RNasin (Promega), 100 mM KCl, 1.5 mM MgCl2 and 0.5 mM DTT for 1.5 h at 30°C. The reaction was stopped by proteinase K treatment and RNA was isolated by phenol/chloroform extraction and ethanol precipitation. Cleavage products were analysed by denaturing RNA–PAGE followed by autoradiography using BioMax MS film (Kodak) and an intensifying screen (Kodak).

**Production of wild-type and mutated tRNA molecules**

To ensure minimum structural disruptions in mutated tRNA (LysTTT) sequences, the RNAs were folded by Vienna RNA Package (24), and then manually inspected using VARNA (25). Wild-type and mutated tRNA sequences were constructed using mirVana miRNA Probe Construction kit (Ambion, Austin, TX, USA). Briefly, wild-type or mutated tRNA DNA oligos were synthesized (IDT, IA, USA) and annealed to a shorter DNA oligo harboring T7 promoter sequence. The partial double-stranded DNA was filled in by Klenow polymerase and then subjected to in vitro transcription by T7 RNA polymerase. Synthetic RNA was purified by Trizol (Invitrogen), before performing in vitro cleavage assays.

**RESULTS**

**Constitutive ncRNAs but not mRNAs generate termini-specific RNAs**

We previously reported on ~15 nt long unusually small RNAs (usRNAs) and their sub-groups (2), consisting of 3′ end-specific motifs in human and viral genomes. Two of these usRNA subgroups share a CCA element in their 3′ end. To further study RNAs with the terminal CCA motif, we reanalyzed 307,269 RNA sequence reads (10–30 nt) that were identified in the human KSHV-infected primary-effusion lymphoma cell line, BCPI (26). The CCA motif occurs precisely at the 3′ end in 24,601 reads (4,482 distinct sequences, Figure 1A). Terminal, 3′-specific CCA motifs are known to occur in almost all mature tRNAs (~70–100 nt), which act as sites for amino acid residue attachment and as stabilizers for tRNA–ribosome interactions (27). Since the majority (85.6%) of reads containing terminal CCA motifs consisted of usRNAs or even smaller RNAs that have not been the focus of previous studies (6,7,10–12,16,18,19), we sought to fully characterize the spectra of ~10–30
base long terminal CCA-containing RNAs. To reduce artifacts from short length reads, the analysis was limited to those reads that perfectly matched tRNA ends. More than half (13,579 of 24,601 reads) of the terminal CCA-containing small RNAs matched tRNAs. We next tested whether tRNAs produce other 10–30 nt long RNAs at other positions in tRNAs in an unbiased manner as expected for a degradation process involving random exo/endo-nuclease cleavage. The majority of tRFs precisely match 5' or 3' ends of tRNA (Figure 1B and C) and these 5'–3' terminal tRFs matches to 76% (447/522) of human tRNAs. We find a significantly higher proportion of tRFs reads (76%) than previously reported (7) for tRFs (~55%), likely because gel-based isolation of RNAs >10 nt seem to be necessary to retain the full spectrum of usRNAs (2). Interestingly, 5' and 3' tRFs display different size distributions (Figure 1D). The size range of 5' terminal tRFs peaks at 14 and 15 nt, which corresponds to a smaller size estimate than the previous report of 18–19 nt (7); this difference may be due to our inclusion of a broader range (10–30 nt) of RNA fragments, or is a result of differences between the cell lines in the two studies.

We next examined if other classes of RNAs also generate termini-specific small RNAs. The phenomenon of 5'–3' specific processing is observed across all major classes of ncRNAs but not mRNAs (Figure 2). Terminal tRFs are the most abundant fragments found in deep sequencing, followed by rRNA-derived RNAs (Supplementary Figure S1; Figure 1B and C). All ncRNAs generate both 5' and 3' products except snRNAs, which produce small RNAs exclusively from their 3' termini. We note that 5' modifications on mRNAs and snRNAs (28) hinder the cloning of 5' fragments, which might be the most plausible reason for the absence of 5' snRNA-derived reads in deep sequencing data. However, despite the large number of mRNAs, very few mRNA-derived fragments are detected at 5'–3' ends and the proportion of mRNA-derived fragments across all mRNAs is negligible (Figure 2) suggesting that the processing of mRNAs into stable fragments is not a common phenomenon. In comparison to tRNA-derived tRFs with a read number density of ~1.0
read per base, mRNAs yield a noticeably low read density of 0.0005. Furthermore, there were no reads with purely poly(A) tails of at least 15 nt long sequences, suggesting that end processing of mRNAs is a rare event. Although termini-specific stable fragments of mRNAs are not detected in our analysis, it is important to note that very low abundance (1 copy/10 cells) short RNAs also occur near transcription start sites of mRNAs (29,30). These transcription start site-associated (29) RNAs (TSSa-RNAs) or transcription initiation (30) RNAs (tiRNAs) are distinct from termini-specific RNA fragments, because they map to a wide range of up/down-stream regions near transcription start sites. Thus, small RNAs from ncRNA termini are likely to be produced by a different mechanism than that used for TSSa-RNA and tiRNA generation, as well as mRNA degradation.

**Asymmetric processing of 5’–3’ terminal fragments**

During miRNA biogenesis, each miRNA precursor is processed to generate a small (~22 nt) duplex, from which one of the strands (miRNA) is preferentially incorporated into Argonaute (AGO) and thus stabilized, while the other strand, termed the passenger strand is often degraded (31,32). Since the 5’ and 3’ ends of both tRNAs (33) and many snoRNAs (28) form RNA duplexes, we speculated that the terminal RNAs may be generated asymmetrically from the 5’ and 3’ ends of the duplex. To explore this possibility, for each tRNA, we compared the number of 5’ terminal RNAs to that of 3’ terminal RNAs. Among the 447 tRNA genes that match terminal tRFs, 335 (75%) tRFs match both 5’ and 3’ terminal tRFs. Remarkably, 93% of these (313/335) tRFs manifest considerable (2 x) difference between their 5’ and 3’ tRF levels, while nearly half of them (161) yield a difference of 10 x. To ensure that these biases are not due to tRNA's that are similar in their sequences, we reanalyzed our data using a non-redundant set of 99 tRNAs that shared <90% sequence identity to reconfirm the observed biases (Figure 3A). To further confirm that the biases are not limited to our BCP1 library and to rule out protocol-based artifacts, we also analyzed the results from an independent library from HEK293 cells (34). We found similar patterns that were additionally confirmed using northern blots for four tRFs derived from two different tRNAs (Supplementary Figure S2). These results support the notion that tRNAs are generally processed into two terminal tRFs but only one fragment is preferentially maintained in cells, reminiscent of miRNA maturation. The bias is stronger for snoRNAs and snRNAs; the 95% (66/69) snoRNAs generate terminal RNAs primarily from either the 5’ or 3’ ends (Figure 3B), while all terminal snRNA fragments are exclusively derived from 3’ ends (Figure 2). Terminal RNAs of rRNAs also seem to manifest a general preference toward 3’ terminal (Figure 3C). Since 5’ and 3’ ends of rRNAs likely do not form intra-molecular base pairs to form a double-stranded structure within the molecule (35), the asymmetry in ncRNA terminal RNAs could be due to stabilization of the 5’ or 3’ fragment that is cleaved off by the processing machinery.

We next sought to find out whether the terminal RNAs and their asymmetric processing bias are characteristics that are evolutionarily conserved. Analysis of small RNA libraries (36) from mouse embryonic stem cells (mESCs) revealed that mESCs contain very abundant terminal RNAs derived precisely from the 5’ and 3’ ends of tRNAs (Figure 4A) that also manifest asymmetry in processing (Figure 4B). Similar to human terminal tRFs, termini-specific processing (Supplementary Figure S3) and asymmetric stabilization (Supplementary Figure S4) are also observed for mouse rRNAs, snoRNAs and snRNAs. Furthermore, comparison of the asymmetric processing bias of individual human tRNAs and their putative mouse orthologs reveal a modest conservation in their processing bias in generating terminal tRFs (Supplementary Figure S5). Intriguingly, the processing biases of 20–30 nt long tRFs are more correlated (r = 0.67) between human and mice than that of
the 15- to 19-nt-long tRFs. Taken together, these data suggest that the mechanism to generate terminal RNAs from various ncRNAs is a conserved process between human and mice.

Terminal small RNAs are primarily independent of Dicer and DGCR8

Since a few Dicer-dependent RNA products from tRNAs (8) and snoRNAs (3) are known, we investigated whether terminal RNAs are commonly processed through the canonical miRNA pathway. Unlike other studies (4,36) that normalized the data sets using RNAs derived from tRNAs and snoRNAs, we used mRNA-derived fragments for normalization. We used a normalization strategy (see ‘Materials and Methods’ section) based on mRNA fragments because of our finding that mRNAs but not tRNAs and snoRNAs, produce fragments that largely resemble random degradation. Since the dependence of shorter tRFs (~15 nt) on Dicer/DGCR8 has not been investigated yet, we separately analyzed the terminal RNAs of the two different size groups (20–30 and 15–19 nt). As a class, each terminal small RNA group manifest statistically insignificant ($P > 0.05$) changes after either dicer 1 or dger 8 knockouts (Supplementary Figure S6A). However, it is important to note that specific RNAs within tRNAs, snoRNAs and snRNAs do manifest considerable (>2-fold) difference in both Dicer and DGCR8 knockdowns (Supplementary Figure S6B). We next tested whether the observed asymmetric processing of 5’ and 3’ tRFs is maintained during Dicer/DGCR8 deletions. To compare asymmetric processing across different samples, the processing bias was evaluated in each library and for the 20–30 and 15–19 nt size groups (Figure 4C). Consistent with the notion that these RNAs are largely independent of the RNAi pathway, the asymmetry profiles within each group are constant across wild-type and dicer 1/dgcr8 knockout mES cells, yielding Pearson correlation coefficients ($r$) in the range of 0.89–0.97.

Presence of endogenous Ago2-associated 3’ tRFs that are capable of cleaving target RNAs

To further investigate processing and function of tRF, we selected two sequences (HisGTG, LeuCAG) for further...
in vitro studies (Figure 5). We first analyzed possible Dicer requirements for their biogenesis (Figure 5A). Total RNA from wild-type (Dicer+/+) and Dicer knock out (Dicer−/−) mouse embryonic fibroblasts (MEFs) was used for Northern blotting using probes against the HisGTG (Figure 5A, upper panel) and the LeuCAG fragment (lower panel). Indeed, small RNA production from these tRNAs is readily detectable in Dicer-deficient cells, indicating that these tRFs are independent of Dicer processing. To probe into the possibility that some terminal RNAs might use the RNAi pathway to inhibit target transcripts, we further examined whether tRFs associate with Ago2. Endogenous Ago2 was immunoprecipitated using anti-Ago2 antibodies and co-immunoprecipitated RNA was analyzed by Northern blotting using probes specific to HisGTG and LeuCAG (Figure 5B). Both tRFs were detected in anti-Ago2 immunoprecipitates demonstrating that tRFs can be specifically loaded into Ago protein complexes. Although previous studies have indicated that Ago2 co-immunoprecipitates with tRF-like fragments (6,8),
It has not been shown yet whether tRFs can cleave target RNAs. Indeed, based on the selection of high abundance of reads in Ago2-associated data, northern blotting assays confirmed that 3′ tRFs from two tRNAs are associated with endogenous Ago2 (Figure 5B). Analysis using Flag-HA-Ago2 complexes also reconfirmed the association of these 3′ tRFs with Ago2 (Figure 5C). Since Ago2, but not Ago1 cleaves target mRNA (37,38), we immunoprecipitated Flag-HA-Ago1 and Flag-HA-Ago2 were incubated with a 32P-cap-labeled target RNA (~100 nt long), which contained a perfect complementary sequence to the 3′ termini of tRNAs (HisGTG and LeuCAG). Lanes indicated with T1 represent RNaseT1 digestions of the RNA substrates as ladders. The RNA sequence complementary to tRNA HIS-3′ and tRNA LEU-3′ is indicated by a black bar on the left.

**Figure 5.** Presence of Ago2-associated 3′ tRFs that can cleave target RNA. (A) Total RNA from Dicer knockout mouse cells (Dicer−/−) or wild-type cells (Dicer+/+) were analyzed by northern blotting using probes specific for the 3′ of tRNAs (HisGTG and LeuCAG), and miR-125 b. (B) Endogenous Ago2 was immunoprecipitated from human HEK293 lysates, and the co-immunoprecipitated RNAs were extracted and analyzed by northern blots using probes against the 3′ tRFs (HisGTG and LeuCAG). GST antibody is used as control (C) Immunoprecipitated Flag-HA-Ago1 and Flag-HA-Ago2 were incubated with a 32P-cap-labeled target RNA (~100 nt long), which contained a perfect complementary sequence to the 3′ termini of tRNAs (HisGTG and LeuCAG). Lanes indicated with T1 represent RNaseT1 digestions of the RNA substrates as ladders. The RNA sequence complementary to tRNA HIS-3′ and tRNA LEU-3′ is indicated by a black bar on the left.

...ANG and other RNases can cleave tRNAs to produce terminal tRFs...
in a time-dependent manner, suggesting ANG is involved in generating a variety of tRNA fragments, the sizes of which correspond to previously reported 30–50 nt sitRNAs (12), as well as the ~20 nt 5’ (Supplementary Figure S7) and 3’ (Figure 6A) terminal RNAs reported in our study. Processing of rRNA and snRNA, by ANG seems to reflect random hydrolysis, resulting in a series of RNA fragments (Supplementary Figure S8). Since ANG belongs to the super family of RNase A (41), we examined if RNase A produces terminal tRFs at low concentrations, and tested the cleavage pattern of additional RNases. At a much (~30×) lower concentration (0.3 μg/ml) of RNase A than that was used for ANG, RNase A was able to generate both the ~20 and ~25 nt tRF bands specifically after 1-h digestion (Figure 6B). Intriguingly, RNase I, a bacterial RNase that can non-specifically digest all di-nucleotide bonds (42), also generated similar tRFs at low concentrations. In contrast, RNase T1, which cleaves only after unpaired G residues (43), generates fragments that are different than ANG, and RNase A/I (Figure 6B). These observations indicate that multiple RNases expressed in a given cell-type are likely responsible for the diverse range of abundant terminal tRFs with varying lengths.

To further analyze the roles of different tRNA regions on tRF production, we designed a series of mutations to disrupt specific functional sub-structures of the tRNA and monitored the tRNA cleavage patterns generated by ANG (Figure 7 and Supplementary Figure S9). The mutations were designed by an in silico screening of base substitutions to yield candidates that preserve secondary structures outside the mutated domains. ANG cleavage of wild-type tRNAs produced ~20 nt long 3’ terminal tRFs (Figure 7A). Mutations that individually disrupt the acceptor stem, the D stem, the anti-codon stem and the TcC stem were first examined (Supplementary Figure S9A–S9D). None of these mutations interfered with the production of the terminal tRNA fragments by ANG after in vitro cleavage of 0.5 h. Another mutation that completely altered the tRNA clover-leaf structure also did not visibly affect tRF production (Supplementary Figure S9E). However, three different mutants, each involving point mutations around the ANG cleavage site on the TcC loop, resulted in complete disappearance of the terminal RNA bands (Figure 7C). Since the in vitro transcribed tRNAs do not have modified nucleosides, these results also indicate that post-transcriptional modifications of tRNA are not a requirement for terminal RNA processing. In summary, the major determinant in processing of terminal small tRNAs by ANG and possibly related RNases is the TcC loop within tRNA.

Figure 6. Multiple RNases generate terminal tRFs from total RNA. (A) ANG (10 μg/ml) generates ~20 nt long-terminal tRF as well as longer sitRNAs from total RNA extracts. In comparison an equal concentration of RNase A completely degrades the RNA within an hour, while a control lane with solution buffer (Buffer) generate terminal tRFs at much lower levels. (B) in vitro cleavage assay (t ~ 1 h) using total RNA from HEK293 cells (10 μg/lane) and different RNases at a series of concentrations indicate that at low concentrations, both RNase A and RNase I are able to produce similar terminal tRFs as that of ANG (arrows). In contrast, RNase T1 generates a different set of 3’ tRFs.
Endogenous retroviral replication loci are preferentially antisense to 3' terminal tRFs

The binding of tRNAs to retroviral primer binding sites (PBS) facilitates the initiation of retroviral genome replication (44,45), such as that for the human immunodeficiency virus (HIV). Since endogenous human retroviral sequences comprise ~7% of the human genome (46), it is possible that cells have evolved mechanisms to regulate replication of these retroviral elements, particularly for preventing accidental transcription of deleterious retroviral sequences. We therefore investigated whether the terminal tRFs have the potential to bind to human endogenous retroviruses (HERVs). The majority of HERVs exist in the human genome as long terminal repeat (LTR) retrotransposons (47) which, together with non-viral retrotransposons LINEs (long interspersed nucleotide elements) and SINEs (small interspersed nucleotide elements), form roughly 40% of the mammalian genome (48). Terminal 3' tRFs are remarkably more complementary to HERV LTRs than to LINEs and SINEs (Figure 8A). Furthermore, even though the number of distinct 3' and 5' BCP1 tRFs differ only by a factor of two (1002 versus 507), 3' tRF sequences are 26-fold (2614 versus 100) more frequently complementary to retrotransposon elements than the 5' tRFs; corresponding to a normalized enrichment of 13-fold. Similarly, 3' tRFs are 4- to 6-fold more enriched (normalized) to complementary sites in LINEs and SINEs than 5' tRFs. Analysis of the major LTR elements that are complementary to BCP1 tRFs (Figure 8B), indicate that the top candidate tRF (LysTTT tRNA, 947 reads) is nearly identical to a known antisense DNA oligonucleotide that targets the PBS region of HIV and can inhibit HIV replication (49). Moreover, the two 3' tRFs (LeuCAG and HisGTG tRNAs) that we found to associate with Ago2 are also capable of degrading their target RNAs, and are complementary to the ERVL-E and HERVH LTR elements (Figure 5C).

DISCUSSION

We present here an in-depth study of terminal-specific small RNAs from tRNAs, snRNAs, snoRNAs, tRNAs and mRNAs that are generally missed in deep sequencing studies because they are commonly filtered out as potential degradation products during bioinformatics analysis. While such strict filtering greatly helps in focusing on the most abundant and likely functional small RNA molecules, emerging studies continue to provide evidence that degradation-like products can have functional impact and are abundant. Most importantly, whether these small RNA products are functional or cellular noise, their study continue to open up novel cellular mechanisms and interesting cellular attributes (1–4,6–13,16,18,19,30,50–53).

Our own detailed analyses of small RNAs derived from various classes of RNAs reveal novel, biologically important features for these small RNAs and support the notion that terminal RNAs are not due to artifacts of cloning methods used in deep sequencing methods. For clarity, we note that the tRFs in this study are derived from mature RNAs and do not include tRFs that could be also processed from mature RNA precursors such as that of pre-tRNAs (6,7). Comparisons across all novel and other previously reported classes of tRNA/ snoRNA-derived RNAs, (3,4,6–8), indicate that the production of terminal- and 5’–3’ end specific small RNAs, are common features of constitutively expressed ncRNAs. The distinct size distribution of terminal RNAs, as shown by sequencing and northern blots, seem atypical for RNA turnover products generated by common nuclear proof-reading pathways, where mutant or defective RNAs are degraded by nuclear exosomes (50,54), or by 5’–3’ exo-nucleases Rat1 and Xrn1 (52). The extensive identification of both 5' and 3' terminal tRFs with similar size distributions in independent deep-sequencing studies that include Ago bound small RNAs, suggests that tRFs exist due to a yet unknown terminal-specific degradation mechanism or by a processing mechanism that yields functional RNA fragments. The phenomena of terminal RNAs is also
present in mice indicative of a conserved terminal RNA processing/degradation mechanism. While our in vitro cleavage assay results suggest that various RNase families including ANG are likely the factors that drive tRF production, the association of these RNAs to AGO as well as the dearth of mRNA-derived tRFs further seem to support the notion that these RNAs could have a biological function, perhaps in a coordinated manner (55).

Why do ncRNAs but not mRNAs manifest terminal-specific processing of small RNAs? Since mRNAs are comprised of many constitutively expressed RNAs, terminal RNA processing is unlikely to be a simple phenomenon for all constitutively expressed RNAs, and provides additional support to the notion that the phenomenon is independent of a random degradation process. We also did not detect a preference for terminal small RNAs among other well-known ncRNAs such as Y-RNAs (56). One possibility is that terminal RNA processing might be a hallmark of RNAs involved in fundamental and ancient core processes such as translation. Consistent with the notion of terminal RNAs as a process that evolved early, we postulate that 3' tRFs may have evolved as RNAi-based modulators that block the replication of endogenous retroviral sequences (Figure 8C). Indeed, the RNAi pathway evolved as an immune defense mechanism against viruses in basal organisms that do not have protein-based adaptive immune systems (57). While these concepts support the view that terminal RNAs might be limited to core ncRNAs as a pervasive phenomenon, we cannot yet rule out that other classes of RNAs under various environmental conditions (e.g. stress) may also preferentially generate such terminal small RNAs.

Figure 8. 3' terminal tRFs might inhibit endogenous retroviral replication. (A) LTRs are highly enriched in complementary sequences of 3' terminal tRFs. (B) The major tRNA terminal regions that produce abundant tRFs (reads > 10, column 4) and their complementary LTR elements. (C) A potential model for the 3' tRF pathway. The binding of 3' tRF to the transcribed viral RNA could recruit double-stranded RNA specific endonucleases such as the highly efficient Ago2, enabling the rapid cleavage of the transcribed endogenous viral RNA.
Indeed, we found that tRFs are endogenously associated with Ago2 and are able to guide Ago2 to cleave the target RNA. Detailed experimental testing of this hypothesis could take several years of research not only because of the complexity of the regulation but also the difficulty in working with endogenous viral products. It also remains difficult to study these RNAs because common technologies such as RT-PCR are either not reliable or laborious (e.g. northern blot) for detecting these RNAs. Our own studies would have been considerably more difficult than if we did not use recently developed northern blot protocols (22) to routinely assay for the expression of these small RNAs. It is likely that emerging third generation sequencing technologies and other multiplexing technologies, combined with multiplexing, will provide methods for cheaper and more robust detection of these difficult-to-study small RNAs that are typically smaller than even miRNAs (59).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures 1–9.

ACKNOWLEDGEMENTS

We thank members of the laboratories of Y.C. and P.S.M., B.J. for helpful discussions and Mr Boles for computer systems administration. C.E thanks Marie Curie Actions for the Intra-European Fellowship.

FUNDING

American Cancer Society Research Professorships [RP0099401 to Y.C., RP0099601 to P.S.M.]; American Cancer Society [RSG0905401 to B.J.]; Bavarian Genome Research Network [BayGene to G.M.]; Deutsche Forschungsgemeinschaft [SFB 960 and FOR855 to G.M.]; European Research Council [sRNAs to G.M.]; National Institutes of Health (NIH) [GM079756 to B.J., CA136363 and CA136806 to Y.C./P.S.M.]. Funding for open access charge: NIH.

Conflict of interest statement. None declared.

REFERENCES


