Research Article

Base Composition Characteristics of Mammalian miRNAs

Bin Wang

Department of Chemistry, Robert C. Byrd Biotechnology Science Center, Marshall University, Huntington, WV 25755, USA

Correspondence should be addressed to Bin Wang; wangb@marshall.edu

Received 8 November 2012; Revised 15 January 2013; Accepted 29 January 2013

Academic Editor: Satoshi Obika

MicroRNAs (miRNAs) are single-stranded, short (15–27 nucleotides) RNA sequences that repress protein synthesis via base pairing to some portion of a messenger RNA (mRNA), such as the 3' untranslated region (3' UTR), the 5' untranslated region (5' UTR), or the coding region [1, 2]. Endogenous miRNAs have been found in various organisms, including animals, plants, and viruses. Mammalian miRNAs are evolutionarily conserved, are scattered throughout chromosomes, and play an important role in the immune response and the onset of cancer. For this study, the author explored the base composition characteristics of miRNA genes from the six mammalian species that contain the largest number of known miRNAs. It was found that mammalian miRNAs are evolutionarily conserved and GU-rich, where U residues are more frequent than other bases. Adenosine residues are generally the most frequent occupants of positions 2 and 3 of the 5' end of miRNAs. Unlike G and U residues that may pair with C/U and A/G, respectively, A residues can only pair with U residues of target mRNAs, which may augment the recognition specificity of the 5' seed region.

1. Introduction

MicroRNAs (miRNAs) are single-stranded, short (15–27 nucleotides) RNA sequences that repress protein synthesis via base pairing to some portion of a messenger RNA (mRNA), such as the 3' untranslated region (3' UTR), the 5' untranslated region (5' UTR), or the coding region [1, 2]. Endogenous miRNAs have been found in various organisms, including animals, plants, and viruses. In mammals, precursor miRNAs (pre-miRNAs) 60–100 nucleotides in length form a hairpin stem-loop structure and are processed by Dicer, a highly conserved RNase III family endonuclease that is found in almost all eukaryotic organisms [3, 4]. This processing yields a miRNA : miRNA duplex that is approximately 21 nucleotides long. One or both of the duplex’s short RNA strands is incorporated into RNA-induced silencing complexes (RISCs, which are composed of Dicer, Argonaute, and other nonspecified proteins) and functions as a mature miRNA that can base pair with mRNA targets, inducing either the degradation or translational repression of the mRNA [3, 4].

The regulation of miRNA is essential for organisms because miRNA plays critical roles in numerous biological processes, including the proliferation and differentiation of cells, as well as apoptosis. It is predicted that mammalian miRNAs can regulate approximately 30% of protein-coding genes [4]. The dysfunctional posttranscriptional modulation of gene expression caused by miRNAs has been related to diseases such as cancer and neurodegenerative disorders [5–8]. The investigation of miRNAs and their gene targets has thus attracted much attention in recent years and has become a major focus of research in medicine and molecular biology [7, 9–13].

For this study, the author explored the base composition of mature miRNAs from the six mammalian species that contain the largest number of known miRNAs. Position-specific base dominance in certain miRNAs has been investigated by other researchers; for example, Gkirtzou et al. discussed the sequence composition on and around mature miRNAs for humans and mice [14], and Wang et al. described position-specific base dominance in human miRNAs [15]. However, this paper provides more complete, detailed, and updated information regarding the base composition characteristics of mature miRNAs in four primates and two rodents, thus providing additional knowledge regarding mammalian miRNAs. We found that the mammalian miRNAs are evolutionarily conserved and GU-rich, where U residues are more frequent than other bases. Adenosine residues are generally
miRNAs were obtained from miRBase [15], a database that contains the sequences of mature miRNAs, for each species studied. In Excel, the miRNA sequences were arranged such that each Excel column contained only the nucleotides at a specific position of the miRNAs studied (e.g., column B contained nucleotides at position 1 of the 5' end of all human miRNAs; column C contained nucleotides at position 2 of the 5' end of all human miRNAs studied). The nucleotides in each column were copied and pasted into an online RNA base composition calculator (http://www.currentprotocols.com/WileyCDA/CurPro3Tool/toolid-7.html) in order to determine the number and percentage of A, G, U, and C at every position of each miRNA of each species studied.

### 2. Materials and Methods

The sequences of mature miRNAs were obtained from miRBase [http://www.mirbase.org/], a database that contains experimentally supported base sequences for various miRNAs [16–19]. The miRNA data for Homo sapiens, Pongo pygmaeus, Pan troglodytes, Macaca mulatta, Mus musculus, and Rattus norvegicus were exported from the miRBase database and then transferred into Microsoft Excel. In Excel, each miRNA sequence was pasted into a single column, and the frequency of each nucleotide at each position was calculated using an online RNA base composition calculator. The results were then formatted into a table for easy comparison.

| Species          | No. of miRNA sequences investigated | No. and % of A, G, U, and C in miRNAs investigated | Base composition at position 1 of the 5' end of miRNAs | Base composition at position 1 of the 3' end of miRNAs |
|------------------|-----------------------------------|---------------------------------------------------|------------------------------------------------------|------------------------------------------------------|
| Homo sapiens     | 1897                              |                                                   |                                                      |                                                      |
|                  | A 9510 (23%)                      | A 578 (30%)                                       | A 409 (22%)                                         |                                                      |
|                  | G 11383 (28%)                     | G 263 (14%)                                       | G 463 (24%)                                         |                                                      |
|                  | U 10755 (26%)                     | U 673 (36%)                                       | U 630 (33%)                                         |                                                      |
|                  | C 9236 (23%)                      | C 383 (20%)                                       | C 395 (21%)                                         |                                                      |
| Pongo pygmaeus   | 600                               |                                                   |                                                      |                                                      |
|                  | A 2945 (23%)                      | A 178 (30%)                                       | A 110 (18%)                                         |                                                      |
|                  | G 3438 (26%)                      | G 69 (11%)                                        | G 141 (24%)                                         |                                                      |
|                  | U 3731 (29%)                      | U 264 (44%)                                       | U 236 (39%)                                         |                                                      |
|                  | C 2928 (22%)                      | C 89 (15%)                                        | C 113 (19%)                                         |                                                      |
| Pan troglodytes  | 525                               |                                                   |                                                      |                                                      |
|                  | A 2653 (23%)                      | A 166 (32%)                                       | A 103 (20%)                                         |                                                      |
|                  | G 2961 (26%)                      | G 64 (12%)                                        | G 118 (22%)                                         |                                                      |
|                  | U 3265 (29%)                      | U 229 (44%)                                       | U 198 (38%)                                         |                                                      |
|                  | C 2527 (22%)                      | C 66 (12%)                                        | C 106 (20%)                                         |                                                      |
| Macaca mulatta   | 488                               |                                                   |                                                      |                                                      |
|                  | A 2479 (23%)                      | A 144 (29%)                                       | A 89 (18%)                                          |                                                      |
|                  | G 2741 (26%)                      | G 59 (12%)                                        | G 117 (24%)                                         |                                                      |
|                  | U 3042 (29%)                      | U 213 (44%)                                       | U 179 (37%)                                         |                                                      |
|                  | C 2392 (22%)                      | C 72 (15%)                                        | C 103 (21%)                                         |                                                      |
| Mus musculus     | 1157                              |                                                   |                                                      |                                                      |
|                  | A 5687 (23%)                      | A 336 (29%)                                       | A 262 (23%)                                         |                                                      |
|                  | G 6634 (26%)                      | G 161 (14%)                                       | G 266 (23%)                                         |                                                      |
|                  | U 6901 (27%)                      | U 417 (36%)                                       | U 379 (33%)                                         |                                                      |
|                  | C 5926 (24%)                      | C 243 (21%)                                       | C 250 (21%)                                         |                                                      |
| Rattus norvegicus| 680                               |                                                   |                                                      |                                                      |
|                  | A 3496 (24%)                      | A 203 (30%)                                       | A 162 (24%)                                         |                                                      |
|                  | G 3639 (25%)                      | G 86 (12%)                                        | G 150 (22%)                                         |                                                      |
|                  | U 4149 (28%)                      | U 243 (36%)                                       | U 233 (34%)                                         |                                                      |
|                  | C 3472 (23%)                      | C 148 (22%)                                       | C 135 (20%)                                         |                                                      |

The six largest miRNA pools were exported from the miRBase database, including Homo sapiens (1921 miRNAs), Mus musculus (1157 miRNAs), Rattus norvegicus (680 miRNAs), Pongo pygmaeus (600 miRNAs), Pan troglodytes (525 miRNAs), and Macaca mulatta (488 miRNAs). A small number of human miRNAs were found to have slightly different names but identical sequences. The names of miRNAs in miRBase are first assigned to their hairpin precursors and then inherited by the mature miRNAs. The naming problem is caused by the occasional existence of multiple miRNA members in the same family, where the mature sequence from one stem of the hairpin precursor is slightly different from the same sequence in a miRNA family member, while the sequences from the other stem of the hairpin precursor are identical. After deleting these redundant sequences, 1897 out of 1921 of
Table 2: Percentages (%) of A, G, U, and C in every mature miRNA of each species studied. This table contains base composition information for the second-through-eighth nucleotides from both ends of the miRNAs. Bases shown in red are the most favored residues at positions 2 and 3 of the 5' end; bases in magenta are the most favored residues at positions 4–6 of the 5' end; bases in orange are the most favored residues at position 2 of the 3' end; bases in green are the most favored residues at position 3 of the 3' end; bases in blue are the most favored residues at position 4 of the 3' end; and bases in purple are the most favored residues at positions 5 and 7 of the 3' end of the miRNAs examined.

| Species          | Location of base residue, as indicated by end of miRNA (5' or 3') and position number |
|------------------|--------------------------------------------------------------------------------------|
|                   | 5', 2  | 5', 3 | 5', 4 | 5', 5 | 5', 6 | 5', 7 | 3', 2 | 3', 3 | 3', 4 | 3', 5 | 3', 6 | 3', 7 | 3', 8 |
| Homo sapiens     |        |       |       |       |       |       |       |       |       |       |       |       |       |
| A29              |        |       |       |       |       |       |       |       |       |       |       |       |       |
| G28              |        |       |       |       |       |       |       |       |       |       |       |       |       |
| U22              |        |       |       |       |       |       |       |       |       |       |       |       |       |
| C21              |        |       |       |       |       |       |       |       |       |       |       |       |       |
| G28              |        |       |       |       |       |       |       |       |       |       |       |       |       |
| U22              |        |       |       |       |       |       |       |       |       |       |       |       |       |
| C21              |        |       |       |       |       |       |       |       |       |       |       |       |       |
| Homo troglodytes |        |       |       |       |       |       |       |       |       |       |       |       |       |
| A35              |        |       |       |       |       |       |       |       |       |       |       |       |       |
| G26              |        |       |       |       |       |       |       |       |       |       |       |       |       |
| U20              |        |       |       |       |       |       |       |       |       |       |       |       |       |
| C19              |        |       |       |       |       |       |       |       |       |       |       |       |       |
| Homo troglodytes |        |       |       |       |       |       |       |       |       |       |       |       |       |
| A34              |        |       |       |       |       |       |       |       |       |       |       |       |       |
| G27              |        |       |       |       |       |       |       |       |       |       |       |       |       |
| U23              |        |       |       |       |       |       |       |       |       |       |       |       |       |
| C16              |        |       |       |       |       |       |       |       |       |       |       |       |       |
| Macaca mulatta   |        |       |       |       |       |       |       |       |       |       |       |       |       |
| A30              |        |       |       |       |       |       |       |       |       |       |       |       |       |
| G29              |        |       |       |       |       |       |       |       |       |       |       |       |       |
| U21              |        |       |       |       |       |       |       |       |       |       |       |       |       |
| C20              |        |       |       |       |       |       |       |       |       |       |       |       |       |
| Mus musculus     |        |       |       |       |       |       |       |       |       |       |       |       |       |
| A31              |        |       |       |       |       |       |       |       |       |       |       |       |       |
| G28              |        |       |       |       |       |       |       |       |       |       |       |       |       |
| U20              |        |       |       |       |       |       |       |       |       |       |       |       |       |
| C21              |        |       |       |       |       |       |       |       |       |       |       |       |       |

human miRNAs were included for investigation in this study. No identical sequences were observed within any of the other five species.

Column 3 in Table 1 provides the total base composition, including total numbers and percentages of A, G, U, and C in all mature miRNAs of each species studied. Overall, G and U residues are more abundant than A and C residues in the miRNA sequences investigated. More Us thanGs are present in five of the six species, with human miRNAs being the exception. Based on canonical Watson-Crick base pairing, this bias in miRNA composition indicates the abundance of A residues in the target mRNA regions.

David Bartel’s research team discovered the high frequency of U and low occurrence of G at the 5’ end of Caenorhabditis elegans miRNAs [20]. Data shown in Table 1, column 4, confirmed this tendency in mammalian species. At position 1 of all experimentally confirmed miRNAs in each species studied, the frequency of U residues ranges from 36% to 44% (shown in red in Table 1), whereas G residues range from 11% to 14% (shown in green in Table 1). In addition, C residues are less frequent at the 5’ end of mammalian miRNAs, indicating a preference for base pairs with lower thermodynamic stability (i.e., favoring AU rather than GC base pairs) at position 1.

Data presented in Table 1, column 5, provide the base composition at the 3’ end of the miRNAs studied. Interestingly, the 3’ end demonstrates the same abundance of U residues as found in the 5’ end of miRNAs, indicating an enrichment of A residues within the 5’ end of mRNA targets. G and C residues are less frequent in the 3’ end of miRNAs as well, however, to a lesser degree than in the 5’ end.

It has been reported that the presence of an A residue in the mRNA target, across from position 1 of the 5’ end of a miRNA, improves binding between the miRNA and its target. This occurs even though actual base pairing between the A residue and the nucleotide at position 1 of the 5’ end of the miRNA is not necessary [3, 4, 21–24]. Fewer studies have focused on the 3’ end of miRNAs [14]. The strong bias toward U residues at position 1 of both ends of miRNAs in the six mammalian species studied in this paper implies that both ends, rather than just the 5’ end, may participate in target recognition. One study demonstrated that miRNA can
Position 4 at the 5′ end of these miRNAs is more frequently G in primates and U in rodents (shown in purple in Table 2). Position 5 at the 3′ end of these miRNAs is most frequently U (shown in purple in Table 2); U is also the most frequent occupant of position 7 (shown in purple in Table 2), except in human miRNAs. The base preferences at positions 6 and 8 of the 3′ end of miRNAs are not consistent across the six species studied.

The specificity of miRNA targeting in mammals is not necessarily restricted to the 5′ and 3′ ends of the seed pairing regions, and previous reports suggest that seed matches are not always sufficient for the repression of targets in mammals [3]. Therefore, base composition beyond the eighth nucleotide from the 5′ end of miRNAs was also investigated in this study. Base composition information for the ninth-through-sixteenth nucleotides from the 5′ end of the miRNAs is presented in Table 3. Positions 9, 13, and 16 demonstrate a strong preference for G at position 2 (shown in orange in Table 2), followed by a preference for U at position 3 (shown in green in Table 2).
end of a miRNA, improves binding between the miRNA and its target. This occurs even though the A or U does not need to base pair with the nucleotide at position 9 of the 5′ end of the miRNA [3, 4, 22]. Our study demonstrates that at position 9 of the miRNAs, there is a strong preference for U (ranging from 29% to 35% in the six species investigated), which is able to base pair with the A residue in the mRNA target. This finding implies that the nucleotide at position 9 of the 5′ end of a miRNA plays an important role in improving both the recognition of specific targets and the functional efficiency of the miRNA.

It has been reported that residues at positions 13–16 of the 5′ end of the miRNAs are also important for stabilizing miRNA-mRNA interactions, especially when base pairing in the 5′ seed region is suboptimal [3, 4, 22]. The lengths of the mature miRNAs investigated in this study range from 15 to 27 nucleotides. For some short miRNAs, residues at positions 13–16 are located in the 3′ half of the miRNAs; however, for some long miRNAs, these residues are situated in the central region of the miRNAs. Therefore, whether nucleotide pairing at positions 13–16 of the 5′ end of a miRNA is important may, at least partially, depend on their exact locations in the miRNA investigated.

Overall, mammalian miRNAs are evolutionarily conserved and GU-rich, where U residues are more frequent than other bases. This finding implies that their target mRNA sites should be A-rich. Another interesting finding is that, in the miRNA sequences investigated, A residues are clearly the most frequent occupants of positions 2 and 3 of the 5′ end of miRNAs. Unlike G and U residues that may pair with C/U and A/G, respectively, A residues can only pair with U residues of target mRNAs, which may augment the recognition specificity of the 5′ seed region.

Conflict of Interests
The author declares that she has no competing financial interests.

Acknowledgments
This work was supported by National Science Foundation Grant EPS-1003907, NASA WVSCC Research Initiation Grant (to B. Wang), and NASA WV EPSCoR Research Seed Grant (to B. Wang).

References
[1] J. S. Ross, J. A. Carlson, and G. Brock, “miRNA: the new gene silencer,” American Journal of Clinical Pathology, vol. 128, no. 5, pp. 830–836, 2007.
[2] C. K. Sen and S. Roy, “miRNA: licensed to kill the messenger,” DNA and Cell Biology, vol. 26, no. 4, pp. 193–194, 2007.
[3] A. Grimson, K. K. H. Farh, W. K. Johnston, P. Garrett-Engele, L. P. Lim, and D. P. Bartel, “MicroRNA targeting specificity in mammals: determinants beyond seed pairing,” Molecular Cell, vol. 27, no. 1, pp. 91–105, 2007.
[4] W. Filipowicz, S. N. Bhattacharyya, and N. Sonenberg, “Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight?” Nature Reviews Genetics, vol. 9, no. 2, pp. 102–114, 2008.
[5] Y. Tanaka and T. Kozu, “miRNA in disease and therapeutic development,” Tanpakushitsu Kakusan Koso, vol. 51, no. 16, pp. 2509–2514, 2006.
[6] J. Weiler, J. Hunziker, and J. Hall, “Anti-miRNA oligonucleotides (AMOs): ammunition to target miRNAs implicated in human disease?” Gene Therapy, vol. 13, no. 6, pp. 496–502, 2006.
[7] C. Jay, J. Nemunaitis, P. Chen, P. Fulgham, and A. W. Tong, “miRNA profiling for diagnosis and prognosis of human cancer,” DNA and Cell Biology, vol. 26, no. 5, pp. 293–300, 2007.
[8] T. H. Thai, P. A. Christiansen, and G. C. Tsokos, “Is there a link between dysregulated miRNA expression and disease?” Discovery Medicine, vol. 10, no. 52, pp. 184–194, 2010.
[9] F. Ahmed, H. R. Ansari, and G. P. S. Raghava, “Prediction of guide strand of microRNAs from its sequence and secondary structure,” BMC Bioinformatics, vol. 10, article 105, 2009.
[10] Z. Hu, “Insight into microRNA regulation by analyzing the characteristics of their targets in humans,” BMC Genomics, vol. 10, article 594, 2009.
[11] M. Brameier, “Genome-wide comparative analysis of microRNAs in three non-human primates,” BMC Research Notes, vol. 3, article 64, 2010.
[12] G. M. Borchert, N. W. Holton, J. D. Williams et al., “Comprehensive analysis of microRNA genomic loci identifies pervasive repetitive-element origins,” Mobile Genetic Elements, vol. 1, pp. 8–17, 2011.
[13] M. Schnall-Levin, O. S. Rissland, W. K. Johnston, N. Perrimon, D. P. Bartel, and B. Berger, “Unusually effective microRNA targeting within repeat-rich coding regions of mammalian mRNAs,” Genome Research, vol. 21, no. 9, pp. 1395–1403, 2011.
[14] K. Gkirtzou, I. Tsamardinos, P. Tsakalides, and P. Poirazi, “MatureBayes: a probabilistic algorithm for identifying the mature miRNA within novel precursors,” PLoS ONE, vol. 5, no. 8, Article ID e1843, 2010.
[15] Z. Z. Wang, B. S. Gong, H. K. Wang et al., “MicroRNA regulation constrains the organization of target genes on mammalian chromosomes,” FEBS Letters, vol. 585, no. 12, pp. 1897–1904, 2011.
[16] S. Griffiths-Jones, “miRBase: the microRNA sequence database,” Methods in Molecular Biology, vol. 342, pp. 129–138, 2006.
[17] S. Griffiths-Jones, R. J. Grocock, S. van Dongen, A. Bateman, and A. J. Enright, “miRBase: microRNA sequences, targets and gene nomenclature,” Nucleic Acids Research, vol. 34, pp. D140–D144, 2006.
[18] S. Griffiths-Jones, H. K. Saini, S. Van Dongen, and A. J. Enright, “miRBase: tools for microRNA genomics,” Nucleic Acids Research, vol. 36, no. 1, pp. D154–D158, 2008.
[19] A. Kozomara and S. Griffiths-Jones, “MiRBase: integrating microRNA annotation and deep-sequencing data,” Nucleic Acids Research, vol. 39, no. 1, pp. D152–D157, 2011.
[20] N. C. Lau, L. P. Lim, E. G. Weinstein, and D. P. Bartel, “An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans,” Science, vol. 294, no. 5543, pp. 858–862, 2001.
[21] J. G. Doench and P. A. Sharp, “Specificity of microRNA target selection in translational repression,” Genes and Development, vol. 18, no. 5, pp. 504–511, 2004.
[22] J. Brennecke, A. Stark, R. B. Russell, and S. M. Cohen, “Principles of microRNA-target recognition,” PLoS Biology, vol. 3, no. 3, p. e85, 2005.
[23] B. P. Lewis, C. B. Burge, and D. P. Bartel, “Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets,” *Cell*, vol. 120, no. 1, pp. 15–20, 2005.

[24] C. B. Nielsen, N. Shomron, R. Sandberg, E. Hornstein, J. Kitzman, and C. B. Burge, “Determinants of targeting by endogenous and exogenous microRNAs and siRNAs,” *RNA*, vol. 13, no. 11, pp. 1894–1910, 2007.

[25] I. Lee, S. S. Ajay, J. I. Yook et al., “New class of microRNA targets containing simultaneous 5′-UTR and 3′-UTR interaction sites,” *Genome Research*, vol. 19, no. 7, pp. 1175–1183, 2009.

[26] E. Borenstein and E. Ruppin, “Direct evolution of genetic robustness in microRNA,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 17, pp. 6593–6598, 2006.