



Investigation of the sequence constitution of miRNA and siRNA

Lucia PAPPALARDO*

Department of Biology and Chemistry, American University of Sharjah, P.O. B. 26666 Sharjah,
United Arab Emirates

*Corresponding author; E-mail: lpappalarDO@aus.edu, Tel: +971 6 5152425, Fax: +971 6 5152450.

ABSTRACT

MicroRNAs (miRNAs) and small-interfering RNAs (siRNAs) are small non-coding RNAs that play important regulatory roles in animals and plants. The primary sequences of 9164 miRNAs and 14238 siRNAs were analyzed to determine the occurrence of each nucleotide in specific positions of the sequences. The results show that there are positions in which the composition is not completely random. In general the nucleotide cytosine is underrepresented in both miRNA and siRNA sequences, while the nucleotides uracil and adenine are overrepresented in miRNAs and siRNAs, respectively. Possible implications between these findings and the biological functions of small non-coding RNAs are discussed.

© 2011 International Formulae Group. All rights reserved.

Keywords: microRNA, RNA primary sequence, uracil

INTRODUCTION

MicroRNAs (miRNAs) and small-interfering RNAs (siRNAs) are small non-coding RNAs, about 21- to 23-nucleotide long, that play important regulatory roles in animals and plants by targeting mRNAs for cleavage or translational repression. Many studies have been done to unravel the structure and the mechanism by which siRNAs and miRNAs function, but the overall mechanism is not yet fully understood (Lee et al., 2004; Zeng, 2006; Rana, 2007; Bartel, 2009; Yang and Yuan, 2009; Suzuki and Miyazono, 2011). Both siRNAs and miRNAs in the last stage of their mechanism assemble into an RNA-induced silencing complex (RISC), known respectively as siRISC and miRISC. siRISC silences the gene expression

by recognizing, binding and cleaving the target RNA. Usually the cleavage occurs when the small RNA is a perfect match to the target RNA sequence. miRISC instead, recognizes and binds the target mRNA in an imperfect match. Translation is repressed but the target RNA is not cleaved.

In this work we perform a statistical analysis of the residue composition of all the miRNA and siRNA sequences publically available through online databases.

MATERIALS AND METHODS

9164 miRNA and 14238 siRNA sequences were downloaded online respectively from miRBase (<http://www.mirbase.org>), (Griffiths-Jones 2004; Griffiths-Jones et al., 2006; Griffiths-

Jones et al. 2008) and siRecords database (<http://sirecords.biolead.org>), (Gong et al. 2006; Rene et al., 2006; Gong et al., 2008; Rene et al., 2009). Only none identical sequences were retained. Occurrence of each residue A, C, G and U, at each position 1 to 21 was determined.

RESULTS AND DISCUSSION

The composition of miRNA and siRNA primary sequences, available through online databases, was analyzed. The occurrence of each residue, A, C, G and U, at each position 1 to 21 of the known miRNA and siRNA sequences was determined. The results are reported in figures 1 and 2, with a threshold at 25% to better visualize deviations from random occurrence. The graphs show that there are positions in which the composition is not completely random. It is already known that the U residue is overrepresented at 5' end (position 1) of mature miRNA sequences (Zeng, 2006; Bartel, 2009). We find indeed a 58% probability for U residues to be represented in this position, while G residue is underrepresented with only 8% probability to be found at this position. Our analysis also shows that the U residue is overrepresented in position 9 with a 36% probability to be found at this position. The U residue is also slightly overrepresented from positions 11-21. The C residue is in general underrepresented with occurrence of less than 20% in several positions. The A residue is overrepresented at positions 2, 3, 5 and 7. Positions 6, 8 and 10 present a random composition. In general miRNA sequences have an overrepresentation of U residues, while the C residue is underrepresented.

For siRNAs the composition at 5' end (position 1) is reversed compared to miRNA; indeed we find the G residue is overrepresented in position 1 with a 40% probability, while U is underrepresented with 11% probability. The A residue is

overrepresented in position 2, with 40% probability and slightly overrepresented in several positions. The U residue is underrepresented at the first four positions. Positions in the middle of the sequences present random compositions. In comparison with miRNA sequences we have an overrepresentation of A residues instead of U, while C residues are underrepresented in both siRNA and miRNA. These tendencies are more accentuated in the sequences of siRNAs with high activities (Figure 2b). siRNAs with high activities are defined in <http://sirecords.biolead.org>.

To validate our statistical analysis we computer-generated 3200 random 22-nucleotide RNA sequences and we applied to them the same statistical analysis as for the miRNA and siRNA sequences. The results are reported in Figure 3. As expected, Figure 3 shows about 25% probability to find each residue at each position.

It is known that the first nucleotide of small interfering RNA determines which strand is selected as the guide strand, while residues 2-8 are required to match the target mRNA to form a duplex. It is also known that siRNA requires A-form helical geometry at the cleavage site which is at position 10-11. In miRNA it is likely that bulges or other non-helical elements are present at this site (Lee et al., 2004; Rana, 2007; Khvorova et al., 2003).

The NMR solution structure of a model mimicking the interaction of let-7 miRNA with its complementary site (LCS 2) in the 3' untranslated region (3'-UTR) of the lin-41 mRNA has been determined by Cevic et al. (2008). The two stems regions of the complex are separated by an asymmetric internal loop that adopts a well-defined structure in which three uracils form a base triple. One of the uracils involved in the internal loop corresponds to the position 9 in miRNA sequence, where U is overrepresented (Figure 1).

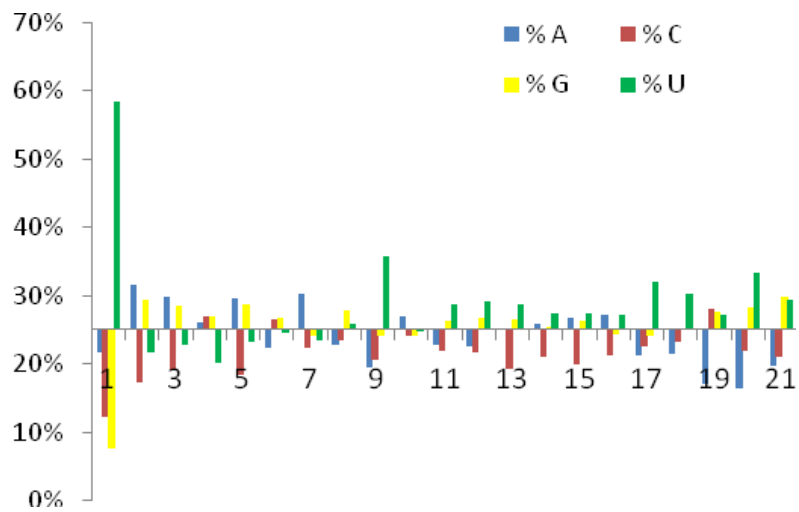


Figure 1: Occurrence of A, C, G and U residues at positions 1 to 21 in miRNA.

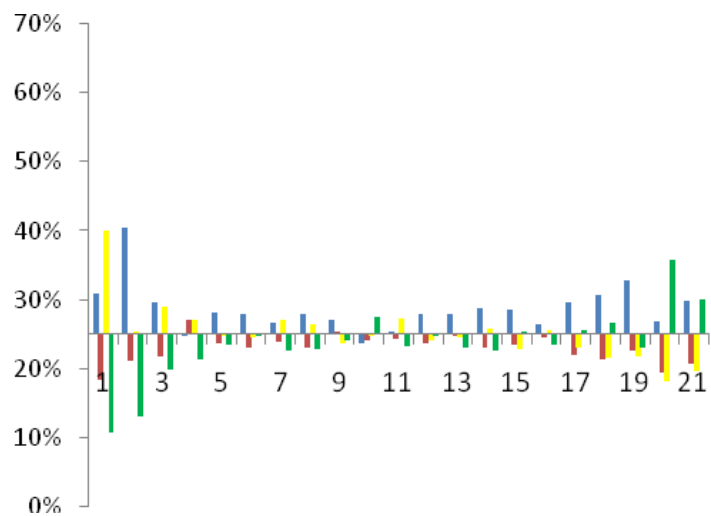


Figure 2A: Occurrence of A, C, G and U residues at positions 1 to 21 in siRNA.

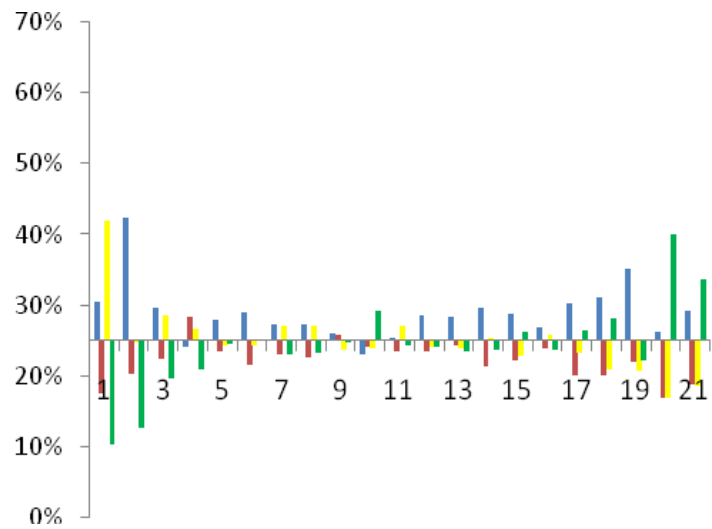


Figure 2B: Occurrence of A, C, G and U residues at positions 1 to 21 in siRNA with high activities.

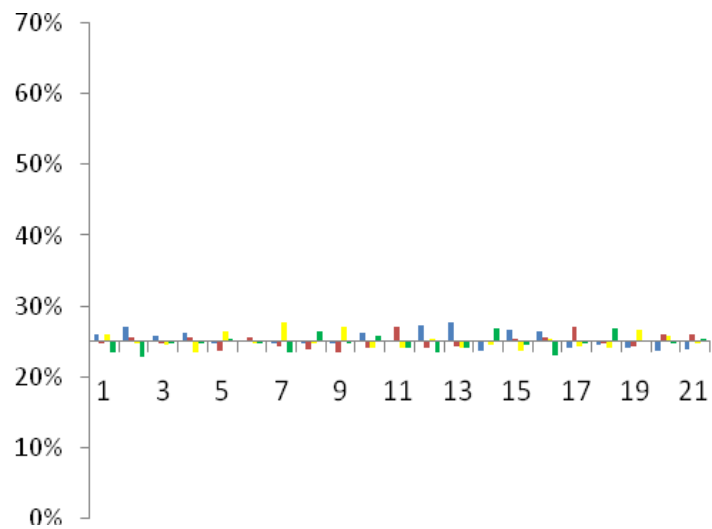


Figure 3: Occurrence of A, C, G and U residues at positions 1 to 21 in randomly generated 21 long RNA.

We speculate that the higher presence of uracil in miRNA sequence, in particular at position 9, may be necessary to favor the formation of U-rich internal loop or bulges that abolish stacking in miRNA/complementary site duplex, due to the fact that U is the nucleotide base with the lowest polarisability, and induce a bending of the stem around that position. The bending of the stem may be a structural requirement for one or more stages in miRNA gene-regulatory mechanism.

ACKNOWLEDGMENTS

The author would like to thank Dr. M. Petersen for preliminary discussion on this work and is immensely grateful to Dr. Prof. P.N. Borer for his comments and proof-reading the manuscript.

REFERENCES

- Bartel DP. 2009. MicroRNAs: target recognition and regulatory functions. *Cell*, **136**: 215-233.
- Cevc M, Thibaudeau C, Plavec J. 2008. Solution structure of a let-7 miRNA: lin-41 mRNA complex from *C. Elegans*. *Nucleic Acids Res.*, **36**: 2330-2337.
- Gong W, Ren Y, Xu Q, Wang Y, Lin D, Zhou H, Li T. 2006. Integrated siRNA design based on surveying of features associated with high RNAi effectiveness. *BMC Bioinformatics*, **7**: 516.
- Gong W, Ren Y, Zhou H, Wang Y, Kang S, Li T. 2008. siDRM: an effective and generally applicable online siRNA design tool. *Bioinformatics*, **24**: 2405-2406.
- Griffiths-Jones S. 2004. The microRNA Registry. *Nucleic Acids Res.*, **32**: (Database issue): D109-111.
- Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. 2006. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*, **34**: (Database issue): D140-144.
- Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. 2008. miRBase: tools for microRNA genomics. *Nucleic Acids Res.*, **36**: 154-158.
- Khvorova A, Reynolds A, Jayasena SD. 2003. Functional siRNAs and miRNAs exhibit strand bias. *Cell*, **115**: 209-16.
- Lee YS, Nakahara K, Pham JW, Kim K, He Z, Sontheimer EJ, Carthew RW. 2004. Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell*, **117**: 69-81.
- Rana TM. 2007. Illuminating the silence: understanding the structure and function of small RNAs. *Nat Rev Mol Cell Biol.*, **8**: 23-36.
- Ren Y, Gong W, Xu Q, Zheng X, Lin D, Wang Y, Li T. 2006. siRecords: an extensive database of mammalian siRNAs with efficacy ratings. *Bioinformatics*, **22**: 1027-1028.
- Ren Y, Gong W, Zhou H, Wang Y, Xiao F, Li T. 2009. siRecords: a database of mammalian RNAi experiments and efficacies. *Nucleic Acids Res.*, **37**: (Database issue): D146-149.
- Suzuki HI, Miyazono K. 2011. Emerging complexity of microRNA generation cascades. *J. Biochem.*, **149**(1): 15-25.
- Yang J, Yuan YA. 2009. A structural perspective of the protein-RNA interactions involved in virus-induced RNA silencing and its suppression. *Biochim. Biophys. Acta*, **1789**: 642-652.
- Zeng Y. 2006. Principles of micro-RNA production and maturation. *Oncogene*, **25**: 6156-6162.