



Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences

Eric Bonnet¹, Jan Wuyts¹, Pierre Rouzé² and Yves Van de Peer^{1,*}

¹Department of Plant Systems Biology, Flanders Interuniversity Institute for Biotechnology (VIB), Ghent University, Technologiepark 927, B-9052 Ghent, Belgium and ²Laboratoire associé de l'INRA, France

Received on January 21, 2004; revised on June 2, 2004; accepted on June 17, 2004
Advance Access publication June 23, 2004

ABSTRACT

Motivation: Most non-coding RNAs are characterized by a specific secondary and tertiary structure that determines their function. Here, we investigate the folding energy of the secondary structure of non-coding RNA sequences, such as microRNA precursors, transfer RNAs and ribosomal RNAs in several eukaryotic taxa. Statistical biases are assessed by a randomization test, in which the predicted minimum free energy of folding is compared with values obtained for structures inferred from randomly shuffling the original sequences.

Results: In contrast with transfer RNAs and ribosomal RNAs, the majority of the microRNA sequences clearly exhibit a folding free energy that is considerably lower than that for shuffled sequences, indicating a high tendency in the sequence towards a stable secondary structure. A possible usage of this statistical test in the framework of the detection of genuine miRNA sequences is discussed.

Availability: The dataset, software and additional data files are freely available as supplementary information on our Website.

Contact: yves.vandeppeer@psb.ugent.be

Supplementary information: <http://www.psb.ugent.be/bioinformatics/>

INTRODUCTION

It is generally accepted that the functional activity of many RNA molecules is determined by their specific secondary (and tertiary) structure, which is, in turn, determined by Watson–Crick, wobble and other non-canonical base pairings. As a consequence, the stability of an RNA with a specific secondary structure and involved in functional processes in the cell is expected to be higher than that of RNAs which do not have such a structure, or for which the secondary structure is of less functional importance. Starting from this hypothesis, Maizel and co-workers (Le *et al.*, 1988, 1989; Chen

et al., 1990) proposed assessing thermodynamic stability and statistical significance of secondary structure features using a method that combines minimum free energy (MFE) values and Monte Carlo simulations. Following this pioneer work, several authors have tried to find significant secondary structures in messenger RNAs (Seffens and Digby, 1999; Workman and Krogh, 1999; Katz and Burge, 2003). For other classes of RNA sequences, Rivas and Eddy (2000) demonstrated that, for most of the non-coding RNAs examined, the stability of the secondary structure was not sufficiently different from the predicted stability of a random sequence to be useful as a general gene-finding approach.

MicroRNAs (miRNAs) are a class of small regulatory non-coding RNAs that have generated much excitement recently (Eddy, 2001; Bartel and Bartel, 2003; Carrington and Ambros, 2003). Analysis of miRNAs is leading to new paradigms for control of gene expression during development in plants and animals. MiRNAs arise from larger precursor molecules that can fold into a stable stem–loop structure (Lee *et al.*, 1993; Lagos-Quintana *et al.*, 2001; Lau *et al.*, 2001; Llave *et al.*, 2002; Reinhart *et al.*, 2002). Those structures are processed by ribonuclease III-like nuclease (Dicer in animals and Dicer-like in plants (Hutvagner and Zamore, 2002; Schauer *et al.*, 2002) and all have a typical stem–loop shape (Lee *et al.*, 1993; Reinhart *et al.*, 2000, 2002; Lagos-Quintana *et al.*, 2001; Lau *et al.*, 2001; Llave *et al.*, 2002).

Here, we present a study on the MFE values of secondary structures in miRNAs and other non-coding RNAs. Original MFE values of the secondary structure are compared with values obtained by randomly shuffling the original sequences.

MATERIALS AND METHODS

Non-coding RNA sequences

A total of 506 miRNA precursor sequences were downloaded from the RFAM database (Ambros *et al.*, 2003; Griffiths-Jones *et al.*, 2003; Griffiths-Jones, 2004). The available sequences

*To whom correspondence should be addressed.

Table 1. Basic statistic values for the sequences used in the study

	%A	%T	%G	%C	Count	Length	%GC	% Paired bases	MFE
tRNA									
<i>A.thaliana</i>	20 ± 3	24 ± 3	32 ± 3	24 ± 3	125	75 ± 5	55 ± 4	30 ± 3	-27.5 ± 3.7
<i>C.elegans</i>	18 ± 3	23 ± 3	33 ± 3	25 ± 3	125	75 ± 7	58 ± 4	30 ± 3	-27.5 ± 5
<i>D.melanogaster</i>	19 ± 3	23 ± 3	33 ± 3	25 ± 3	125	75 ± 8	58 ± 4	31 ± 3	-29 ± 5
<i>H.sapiens</i>	20 ± 4	24 ± 4	31 ± 4	24 ± 4	125	75 ± 5	55 ± 6	30 ± 3	-26 ± 5
miRNA									
<i>C.briggsae</i>	25 ± 4	29 ± 4	23 ± 3	22 ± 4	51	92 ± 7	45 ± 5	36 ± 2	-37 ± 6
<i>C.elegans</i>	26 ± 5	29 ± 5	23 ± 4	21 ± 5	107	97 ± 5	44 ± 7	34 ± 4	-36 ± 9
<i>D.melanogaster</i>	26 ± 4	32 ± 4	22 ± 3	20 ± 3	78	88 ± 13	42 ± 5	35 ± 3	-32 ± 6
<i>H.sapiens</i>	23 ± 5	28 ± 5	26 ± 4	23 ± 6	158	89 ± 14	49 ± 8	36 ± 3	-40 ± 9
<i>M.musculus</i>	23 ± 5	29 ± 5	26 ± 5	22 ± 5	69	69 ± 5	48 ± 9	37 ± 3	-32 ± 5
<i>A.thaliana</i>	26 ± 4	31 ± 3	22 ± 3	21 ± 4	43	135 ± 49	43 ± 5	37 ± 2	-57 ± 17
rRNA	21 ± 7	26 ± 4	30 ± 4	22 ± 6	581	88 ± 25	52 ± 10	32 ± 4	-33 ± 12

Values are expressed as the mean ± SD. Base pairs represent the number of base pairs in the secondary structure relative to the total number of nucleotides in the sequence (the maximum value is thus 50%). The results for rRNAs are showed globally, since the number of different species selected was very high.

cover five animal genomes and one plant genome: *Homo sapiens*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *C.briggsae*, *Mus musculus* and *Arabidopsis thaliana*. Most of the sequences were recently isolated using direct cloning techniques, but some were predicted using computational tools. However, this set should not be considered as complete and is the subject of very active research (Carrington and Ambros, 2003; Bartel, 2004).

Transfer RNAs were downloaded from the genomic tRNA database (Lowe and Eddy, 1997). In order to have a comparable number of sequences, 500 sequences were selected at random from four eukaryotic genomes, namely *H.sapiens*, *A.thaliana*, *C.elegans* and *D.melanogaster*.

Ribosomal RNA sequences are much longer (1500–3000 nt) and have a much more complex secondary and tertiary structure than the other types of non-coding RNAs. A selection of regions that are similar in length compared with miRNAs and tRNAs was made. From the European ribosomal RNA database, 120 randomly chosen eukaryotic rRNA sequences from the small ribosomal subunit were selected (Wuyts et al., 2004). From each of these sequences, five areas were extracted that fold into a secondary structure pattern that does not include any significant base pairing with other parts of the molecule. These areas correspond to helices 12–13, 23/e1–23/e3, 24, 42–44 and 49 (Wuyts et al., 2001).

Some basic statistics about nucleotide composition, sequences length and MFE values are displayed in Table 1.

RNA secondary structure

Methods to predict the secondary structure from the primary sequence information are based on minimizing the free energy of the molecule by maximizing the number of favorable base pairing interactions (Zuker and Stiegler, 1981; Schuster et al., 1994). MFE is estimated by considering the minimum energy values obtained by complementary base pairs

decreased by the stacking energy of successive base pairs or increased by the destabilizing energy associated with non-complementary bases (Zuker and Stiegler, 1981; Schuster et al., 1994). For this reason, the dinucleotide bias in RNA sequences is considered to be important for the predicted MFE and several authors have pointed out the importance of taking into account the dinucleotide distributions when randomizing sequences (Workman and Krogh, 1999; Rivas and Eddy, 2000).

In our study, only the results of the optimal folding were used. The biologically correct structure for tRNAs and rRNAs is often not the calculated optimal structure but one that is within a percent of the optimal structure (suboptimal structure) (Zuker and Stiegler, 1981; Durbin et al., 1998). Therefore, the results should be very close for the biologically active structure. Unfortunately, the real *in vivo* secondary structure of miRNA is not yet known.

Mononucleotide shuffling

This type of shuffling is trivial and is done simply by permuting the nucleotides of the sequence at random. The dinucleotide frequencies are completely distorted using this method.

Dinucleotide shuffling

The goal here is to shuffle a sequence while keeping the dinucleotide distribution (or frequencies) constant. An implementation of the algorithm as described by Workman and Krogh (1999) was used. The dinucleotide and mononucleotide frequencies are exactly preserved.

First-order Markov model

We also used the DiShuffle program (Katz and Burge, 2003). This program derives a first-order Markov model from the conditional probabilities found in the initial sequence.

A random nucleotide is chosen as a seed for a new sequence after which nucleotides will be added according to the conditional probabilities. The process stops when the sequence has the same length as the original. This method produces shuffled sequences with dinucleotide frequencies close to the original sequences but that do not have exactly the same values. Mononucleotide frequencies are not preserved.

Randomization test

A variant of the Z-score procedure (Le *et al.*, 1988, 1989; Workman and Krogh, 1999; Katz and Burge, 2003) was used to determine if the MFE value is significantly different from that of random sequences. The MFE value of the sequence is compared with the distribution of MFE values obtained by randomly reshuffling the order of the nucleotides in the native sequence. This procedure is known as Monte Carlo and randomization tests, and is used in many fields of biology (Efron, 1979; Eddington, 1995; Manly, 1997). A high number of permutations (minimum 1000) are required to have an accurate probability (Edwards, 1985; Jockel, 1986; Eddington, 1995), but as the length of the sequences used here is rather small, this is not a computational issue. Under this model, no assumption is made upon the nature of the distribution and the probability is straightforward to compute, as detailed below.

The randomization test can be summarized as follows:

- (1) Compute MFE of the secondary structure inferred from the original sequence.
- (2) Randomize the order of the nucleotides in the original sequence and compute MFE for the inferred structure based on the shuffled sequence.
- (3) Repeat step 2 a great number of times (1000) in order to build the distribution of MFE values.
- (4) If N is the number of iterations and R the number of randomized sequences that have a MFE value less or equal to the original value, then p is defined here as:

$$p = \frac{R}{N + 1}.$$

To fold sequences into their secondary structure, we used the ViennaRNA fold package (Hofacker *et al.*, 1994) that implements Zucker's energy minimization algorithm. The software that implements the randomization test of secondary structure MFE, called RANDFOLD, is available under the terms of the General Public License (see supplementary data).

RESULTS AND DISCUSSION

Two examples of randomized distributions are given in Figure 1. Figure 1a shows the distribution of MFE values for secondary structures inferred from randomized sequences of the let7 miRNA precursor. Let-7 and Lin-4 genes are the founding members of the miRNAs and control developmental timing in *C.elegans* (Lee *et al.*, 1993; Reinhart *et al.*,

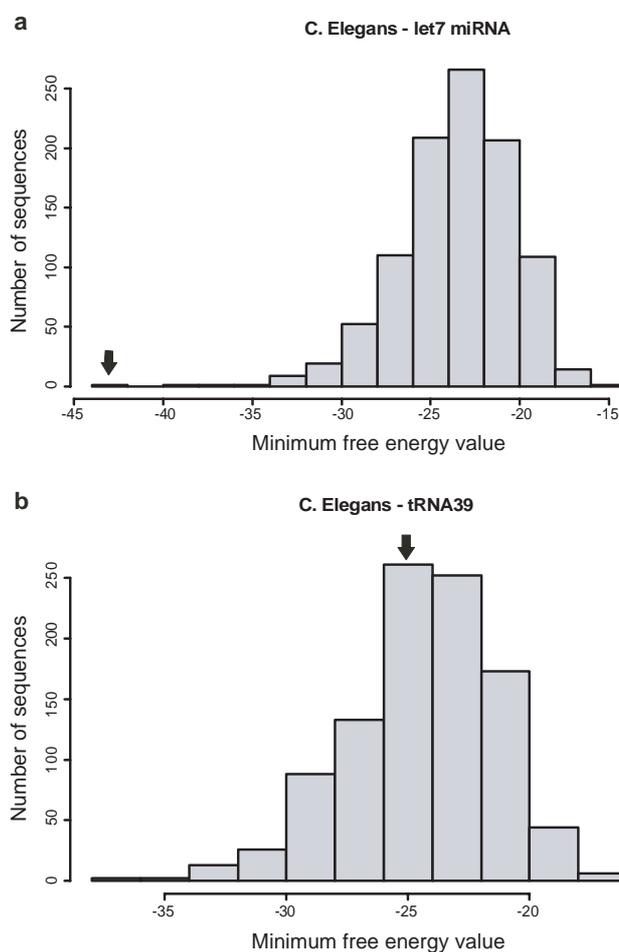


Fig. 1. Distribution of MFE values for randomized sequences (dinucleotide shuffling) for miRNA gene let-7 (a) and for a tRNA gene (b). The black arrow indicates the MFE for the original structure.

2000). The MFE value for the original structure of let7 lies clearly in the extreme tail of the distribution, which corresponds with a low P -value of 0.001. This implies that the MFE value of the original structure of the let7 miRNA gene is significantly different from that of a random sequence.

Figure 1b shows the distribution of MFE values for structures inferred from randomized sequences of a tRNA sequence, Ce-chr4.tRNA39-ArgACG (Lowe and Eddy, 1997) of *C.elegans*. In this case, the MFE value for the original tRNA structure falls in the middle of the distribution giving a P -value of 0.345. Therefore, the null hypothesis cannot be rejected and the MFE value for the original structure is not significantly different from that of a random sequence.

A graph showing the distribution of P -values for all the sequences used in this study is available as supplementary data.

Figure 2 shows the results of the randomization tests for miRNAs, rRNAs and tRNAs sequences (the complete list of

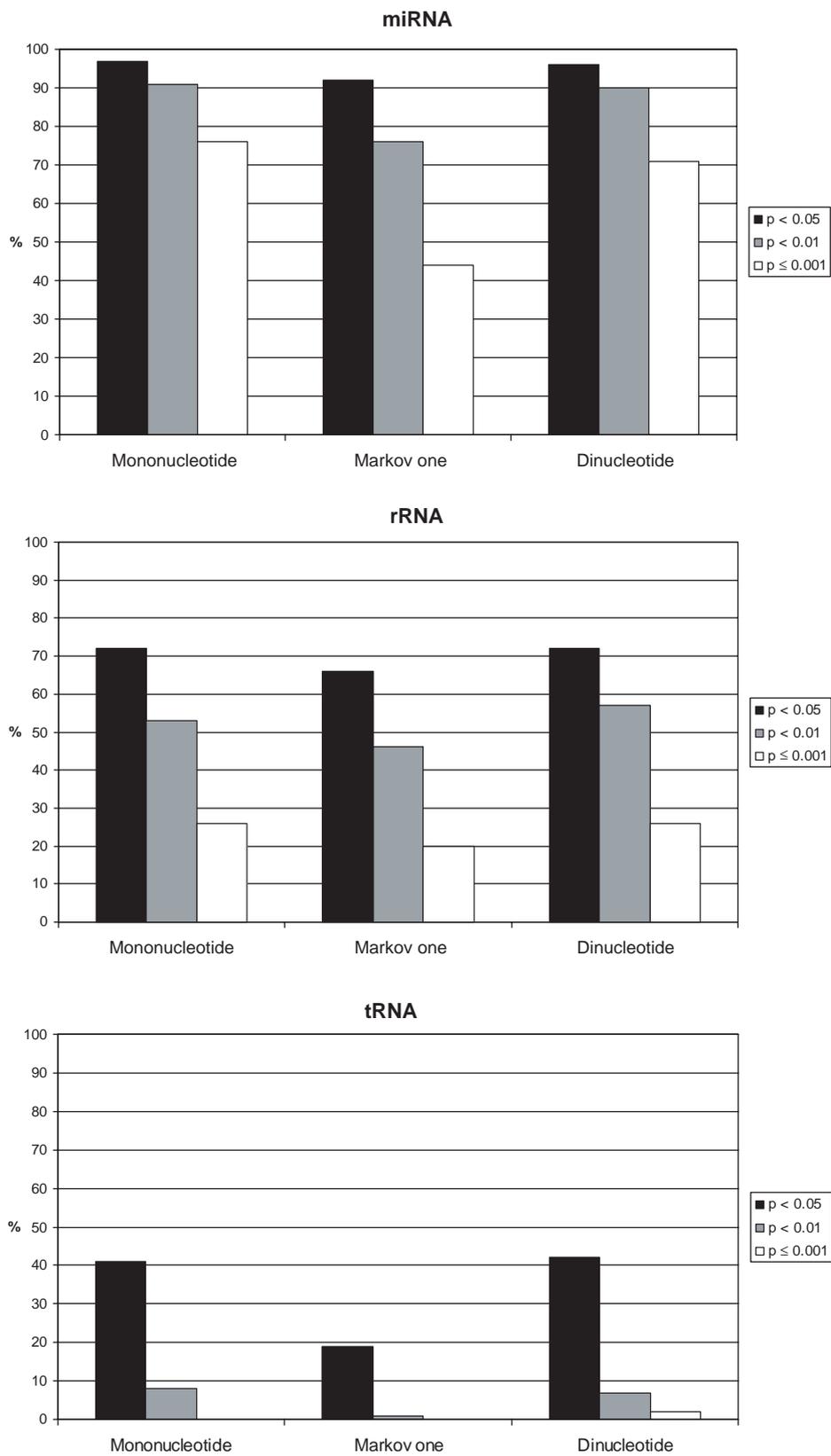


Fig. 2. Proportion of sequences (given in %) that have a *P*-value below 0.05, 0.01 and 0.001, for each type of shuffling method used and for each of the three types of non-coding RNAs used in this study.

MFE values and their associated probabilities is available as supplementary data). The proportion of miRNA sequences that have a structure with a lower MFE than structures based on randomized miRNA sequences is very high for the three levels of significance defined. For example, more than 90% of the miRNA sequences have a P -value <0.01 with the mononucleotide and dinucleotide shuffling. There is also a high proportion of sequences with a P -value of 0.001, based on the three types of randomization methods ($\geq 70\%$ for the mononucleotide and dinucleotide shuffling and $>40\%$ with the first-order Markov).

Ribosomal RNA sequences show a less high proportion of sequences for each P -value level compared to miRNA. For example, for $P < 0.01$, only 50% of the sequences show an MFE value that is lower than MFE values for structures inferred from randomized sequences. For tRNA sequences, $<8\%$ of the sequences have a P -value <0.01 . This result is in agreement with previous findings (Rivas and Eddy, 2000) where 26 tRNA sequences out of 1415 (1.8%) were found to have a P -value of 0.01.

The values obtained with the first-order Markov type of shuffling are always slightly smaller than the values obtained with the other types of shuffling. This is due to fluctuations in the energies for the Markov sequences. Those fluctuations arose due to higher variations in GC content while shuffling sequences maintain a constant GC:AT ratio (Workman and Krogh, 1999). This could artificially increase the number of low values for the MFE of randomized sequences.

Thermodynamic stability of the secondary structure does not seem to be of major importance for tRNAs and rRNAs. For rRNAs, it is well known that the stability is enhanced by interactions with protein structures in the ribosome (Xing and Draper, 1995). For tRNAs, Workman and Krogh (1999) already suggested that the tertiary structure, which is not taken into account in secondary structure predictions, might be of major importance for this class of molecules. For miRNAs, precursor sequences do not have tight interactions with proteins before their cleavage by Dicer (or Dicer-like) enzymes (Lee *et al.*, 2002). Therefore, a stable secondary structure might be needed for this class of non-coding RNAs to avoid early degradation. Indeed, the great majority of the miRNA precursors have an MFE value significantly smaller than MFE values found for structures inferred from randomized sequences. We believe that this observation can be used for the more successful annotation of plant miRNA genes. Usually, in order to identify miRNA genes, the MFE of the sequence is calculated. Once this MFE is lower than a specified cutoff value the precursor sequence is accepted. Some authors also suggested to use a threshold based on the MFE normalized by the length of the sequence (Pervouchine *et al.*, 2003). This kind of approach has been proven to be very successful, especially in animal genomes, where all known miRNA genes are of comparable length and contain few unpaired bases (Lai *et al.*, 2003; Lim *et al.*, 2003a,b). However,

in plant genomes the miRNA precursor sequences are often much longer (Llave *et al.*, 2002; Reinhart *et al.*, 2002). It is poorly understood how this property would affect the cutoff value that is to be used in miRNA annotation. In our approach the MFE value is compared to that of random sequences with the same (di-)nucleotide composition, leading to a less error-prone annotation of miRNAs in plants (Bonnet *et al.*, 2004).

The miRNAs used in our study have heterogeneous origins, and were obtained both by direct cloning methods and by pure computational prediction. In order to avoid a possible bias introduced by considering predicted miRNA precursors, we analyzed a subset of miRNA precursor sequences that were validated by northern blot experiments (Lagos-Quintana *et al.*, 2001; Lau *et al.*, 2001; Lee and Ambros, 2001; Calin *et al.*, 2002; Lee *et al.*, 2002; Schmittgen *et al.*, 2004). The set is composed of 47 sequences covering three different animal taxa, namely *C.elegans*, *D.melanogaster* and *H.sapiens*. Unfortunately, no plant miRNA precursor sequence has been verified experimentally up to now (Reinhart *et al.*, 2002). Table 2 shows the MFE values and their associated P -values computed using 1000 repeats and a dinucleotide randomization method. The results show clearly that most of the P -values are very low, namely 45 sequences out of 47 ($\sim 96\%$) have a P -value <0.07 . Only two sequences (*C.elegans* mir-52 and mir-69) have very high P -values (0.113 and 0.201). The fact that both those sequences are U rich and that mir-69 have an unusual high MFE value of -18 kcal/mol might explain why they have such high P -values. Based on these results, the level of sensitivity of the test appears to be high. However, the fact that a relative high number of rRNA sequences that also have a low P -value clearly suggest that the MFE randomization test alone cannot be used as a valid miRNA sequence detection tool. It should be used in association with other type of evidence such as the absolute MFE value, comparative genomics or other secondary structure characteristics in order to reliably detect genuine miRNA sequences (Bonnet *et al.*, 2004). In this respect, for the integration of the MFE randomization test for validating potential miRNA precursor sequences, a cutoff level must be defined. Preferably, this value should be evaluated on a set of precursor sequences that have been characterized experimentally. The cutoff value will be a trade-off between discarding 'true' sequences (false negatives) and accepting 'false' ones (false positives). For example, if we consider the results of Table 2 for *H.sapiens*, a cutoff value set to $P = 0.05$ would discard only two sequences from this set.

Our results also strengthen what is known from the biogenesis of miRNA precursors: a stable secondary structure with specific properties is probably needed for a correct processing of those molecules by specific enzymes. Mature miRNAs are processed from their stem-loop precursors by double-stranded specific nucleases of the so-called Dicer

Table 2. MFE and computed *P*-values for experimentally validated miRNA precursor sequences (see text for details)

	miRNA	MFE	<i>P</i>	
<i>C.elegans</i>	cel-mir-35	-54	0.001	
	cel-mir-36	-50	0.001	
	cel-mir-37	-45	0.001	
	cel-mir-38	-50	0.001	
	cel-mir-39	-45	0.001	
	cel-mir-40	-46	0.001	
	cel-mir-1	-41	0.001	
	cel-mir-42	-42	0.001	
	cel-mir-43	-47	0.001	
	cel-mir-44	-42	0.001	
	cel-mir-45	-41	0.001	
	cel-mir-80	-31	0.001	
	cel-mir-60	-35	0.001	
	cel-mir-34	-37	0.002	
	cel-mir-84	-23	0.003	
	cel-mir-52	-26	0.113	
	cel-mir-69	-18	0.201	
	<i>D.melanogaster</i>	dme-mir-1	-36	0.001
		dme-mir-2b-2	-37	0.001
dme-mir-6-2		-27	0.001	
dme-mir-6-3		-32	0.001	
dme-mir-2b-1		-28	0.002	
dme-mir-6-1		-26	0.006	
<i>H.sapiens</i>	hsa-let-7d	-43	0.001	
	hsa-mir-16	-38	0.001	
	hsa-mir-20	-31	0.001	
	hsa-mir-21	-36	0.001	
	hsa-mir-28	-51	0.001	
	hsa-mir-29	-25	0.001	
	hsa-mir-30a	-37	0.001	
	hsa-mir-33	-37	0.001	
	hsa-mir-92-1	-36	0.001	
	hsa-mir-93-1	-45	0.001	
	hsa-mir-101	-36	0.001	
	hsa-mir-105-1	-43	0.001	
	hsa-mir-224	-37	0.001	
	hsa-mir-30d	-28	0.001	
	hsa-let-7d	-43	0.001	
	hsa-mir-219	-55	0.002	
	hsa-mir-27a	-38	0.002	
	hsa-mir-15a	-31	0.003	
	hsa-mir-24-1	-26	0.003	
	hsa-mir-216	-41	0.005	
	hsa-mir-220	-43	0.011	
	hsa-mir-107	-30	0.016	
	hsa-mir-18	-22	0.055	
	hsa-mir-24-2	-27	0.069	

family (Hutvagner and Zamore, 2002; Schauer *et al.*, 2002). Recently, another RNase protein named Drosha has been identified as the core nuclease that executes the initiation step of miRNA processing in the nucleus (Lee *et al.*, 2003). The question remains how these RNases can recognize their targets in a very specific manner whereas the primary sequences do not show any conserved elements. In order to answer

this question, it would be important to precisely determine the structures of miRNA precursors and RNase III proteins. According to the results of our study, secondary structure might play a crucial role here.

ACKNOWLEDGEMENTS

We would like to thank Sven Degroeve and Steven Maere for helpful discussions. J.W. is indebted to the 'Vlaams Instituut voor de Bevordering van het Wetenschappelijk-Technologisch Onderzoek in de Industrie' for a postdoctoral fellowship.

REFERENCES

- Ambros, V., Bartel, B., Bartel, D.P., Burge, C.B., Carrington, J.C., Chen, X., Dreyfuss, G., Eddy, S.R., Griffiths-Jones, S., Marshall, M. *et al.* (2003) A uniform system for microRNA annotation. *RNA*, **9**, 277–279.
- Bartel, B. and Bartel, D.P. (2003) MicroRNAs: at the root of plant development? *Plant Physiol.*, **132**, 709–717.
- Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.
- Bonnet, E., Wuyts, J., Rouzé, P. and Van de Peer, Y. (2004) Detection of 71 potential conserved plant microRNAs in *Arabidopsis thaliana* and *Oryza sativa* identifies important target genes. *Proc. Natl Acad. Sci., USA*, (in press).
- Calin, G.A., Dumitru, C.D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., Aldler, H., Rattan, S., Keating, M., Rai, K. *et al.* (2002) Frequent deletions and down-regulation of microRNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc. Natl Acad. Sci., USA*, **99**, 15524–15529.
- Carrington, J.C. and Ambros, V. (2003) Role of microRNAs in plant and animal development. *Science*, **301**, 336–338.
- Chen, J.H., Le, S.Y., Shapiro, B., Currey, K.M. and Maizel, J.V. (1990) A computational procedure for assessing the significance of RNA secondary structure. *Comput. Appl. Biosci.*, **6**, 7–18.
- Durbin, R., Eddy, S.R., Krogh, A. and Mitchison, G. (1998) *Biological Sequence Analysis*. Cambridge University Press, Cambridge.
- Eddington, E.S. (1995) *Randomization Tests*. Marcel Dekker, New York.
- Eddy, S.R. (2001) Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.*, **2**, 919–929.
- Edwards, D. (1985) Exact simulation based inference: a survey, with additions. *J. Stat. Comput. Sim.*, **22**, 307–326.
- Efron, B. (1979) Computers and the theory of statistics: thinking the unthinkable. *Soc. Indus. Appl. Math.*, **21**, 460–480.
- Griffiths-Jones, S. (2004) The microRNA Registry. *Nucleic Acids Res.*, **32**, D109–D111.
- Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A. and Eddy, S.R. (2003) Rfam: an RNA family database. *Nucleic Acids Res.*, **31**, 439–441.
- Hofacker, I.L., Fontana, W., Stadler, P.F., Bonhoeffer, L.S., Tacker, M. and Schuster, P. (1994) Fast folding and comparison of RNA secondary structures. *Monatsh Chem.*, **125**, 167–188.
- Hutvagner, G. and Zamore, P.D. (2002) RNAi: nature abhors a double-strand. *Curr. Opin. Genet. Dev.*, **12**, 225–232.
- Jockel, K.H. (1986) Finite sample properties and asymptotic efficiency of Monte Carlo tests. *J. Stat. Comput. Sim.*, **14**, 336–347.

- Katz,L. and Burge,C.B. (2003) Widespread selection for local RNA secondary structure in coding regions of bacterial genes. *Genome Res.*, **13**, 2042–2051.
- Lagos-Quintana,M., Rauhut,R., Lendeckel,W. and Tuschl,T. (2001) Identification of novel genes coding for small expressed RNAs. *Science*, **294**, 853–858.
- Lai,E.C., Tomancak,P., Williams,R.W. and Rubin,G.M. (2003) Computational identification of *Drosophila* microRNA genes. *Genome Biol.*, **4**, R42.
- Lau,N.C., Lim,L.P., Weinstein,E.G. and Bartel,D.P. (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science*, **294**, 858–862.
- Le,S.V., Chen,J.H., Currey,K.M. and Maizel,J.V.,Jr (1988) A program for predicting significant RNA secondary structures. *Comput. Appl. Biosci.*, **4**, 153–159.
- Le,S.Y., Chen,J.H. and Maizel,J.V. (1989) Thermodynamic stability and statistical significance of potential stem–loop structures situated at the frameshift sites of retroviruses. *Nucleic Acids Res.*, **17**, 6143–6152.
- Lee,R.C. and Ambros,V. (2001) An extensive class of small RNAs in *Caenorhabditis elegans*. *Science*, **294**, 862–864.
- Lee,R.C., Feinbaum,R.L. and Ambros,V. (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*, **75**, 843–854.
- Lee,Y., Ahn,C., Han,J., Choi,H., Kim,J., Yim,J., Lee,J., Provost,P., Radmark,O., Kim,S. and Kim,V.N. (2003) The nuclear RNase III Drosha initiates microRNA processing. *Nature*, **425**, 415–419.
- Lee,Y., Jeon,K., Lee,J.T., Kim,S. and Kim,V.N. (2002) MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J.*, **21**, 4663–4670.
- Lim,L.P., Glasner,M.E., Yekta,S., Burge,C.B. and Bartel,D.P. (2003a) Vertebrate microRNA genes. *Science*, **299**, 1540.
- Lim,L.P., Lau,N.C., Weinstein,E.G., Abdelhakim,A., Yekta,S., Rhoades,M.W., Burge,C.B. and Bartel,D.P. (2003b) The microRNAs of *Caenorhabditis elegans*. *Genes Dev.*, **17**, 991–1008.
- Llave,C., Kasschau,K.D., Rector,M.A. and Carrington,J.C. (2002) Endogenous and silencing-associated small RNAs in plants. *Plant Cell*, **14**, 1605–1619.
- Lowe,T.M. and Eddy,S.R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.*, **25**, 955–964.
- Manly,B.F.J. (1997) *Randomization, Bootstrap and Monte Carlo Methods in Biology*. Chapman Hall, London.
- Pervouchine,D.D., Graber,J.H. and Kasif,S. (2003) On the normalization of RNA equilibrium free energy to the length of the sequence. *Nucleic Acids Res.*, **31**, e49.
- Reinhart,B.J., Slack,F.J., Basson,M., Pasquinelli,A.E., Bettinger,J.C., Rougvié,A.E., Horvitz,H.R. and Ruvkun,G. (2000) The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature*, **403**, 901–906.
- Reinhart,B.J., Weinstein,E.G., Rhoades,M.W., Bartel,B. and Bartel,D.P. (2002) MicroRNAs in plants. *Genes Dev.*, **16**, 1616–1626.
- Rivas,E. and Eddy,S.R. (2000) Secondary structure alone is generally not statistically significant for the detection of noncoding RNAs. *Bioinformatics*, **16**, 583–605.
- Schauer,S.E., Jacobsen,S.E., Meinke,D.W. and Ray,A. (2002) DICER-LIKE1: blind men and elephants in *Arabidopsis* development. *Trends Plant Sci.*, **7**, 487–491.
- Schmittgen,T.D., Jiang,J., Liu,Q. and Yang,L. (2004) A high-throughput method to monitor the expression of microRNA precursors. *Nucleic Acids Res.*, **32**, e43.
- Schuster,P., Fontana,W., Stadler,P.F. and Hofacker,I.L. (1994) From sequences to shapes and back: a case study in RNA secondary structures. *Proc. R. Soc. Lond. B Biol. Sci.*, **255**, 279–284.
- Seffens,W. and Digby,D. (1999) mRNAs have greater negative folding free energies than shuffled or codon choice randomized sequences. *Nucleic Acids Res.*, **27**, 1578–1584.
- Workman,C. and Krogh,A. (1999) No evidence that mRNAs have lower folding free energies than random sequences with the same dinucleotide distribution. *Nucleic Acids Res.*, **27**, 4816–4822.
- Wuyts,J., Perrière,G. and Van de Peer,Y. (2004) The European database on small subunit ribosomal RNA. *Nucleic Acids Res.*, **32**, D101–D103.
- Wuyts,J., Van de Peer,Y. and De Wachter,R. (2001) Distribution of substitution rates and location of insertion sites in the tertiary structure of ribosomal RNA. *Nucleic Acids Res.*, **29**, 5017–5028.
- Xing,Y. and Draper,D.E. (1995) Stabilization of a ribosomal RNA tertiary structure by ribosomal protein L11. *J. Mol. Biol.*, **249**, 319–331.
- Zuker,M. and Stiegler,P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.*, **9**, 133–148.