Statistical Inference for Well-ordered Structures in Nucleotide Sequences

Shu-Yun Le
Laboratory of Experimental
and Computational Biology
NCI Center for Cancer Research
National Cancer Institute, NIH
Bldg. 469, Room 151, Frederick
Maryland 21702, USA
shuyun@ncifcrf.gov

Jih-H. Chen
Advanced Biomedical
Computing Center
SAIC-NCI/FCRDC
Frederick, MD 21702, USA
chen@ncifcrf.gov

Jacob V. Maizel, Jr.
Laboratory of Experimental
and Computational Biology
NCI Center for Cancer Research
National Cancer Institute, NIH
Bldg. 469, Room 151, Frederick
Maryland 21702, USA
jmaizel@ncifcrf.gov

Abstract

Distinct, local structures are frequently correlated with functional RNA elements involved in post-transcriptional regulation of gene expression. Discovery of microRNAs (miRNAs) suggests that there are a large class of small non-coding RNAs in eukaryotic genomes. These miRNAs have the potential to form distinct fold-back stem-loop structures. The prediction of those well-ordered folding sequences (WFS) in genomic sequences is very helpful for our understanding of RNA-based gene regulation and the determination of local RNA elements with structure-dependent functions. In this study, we describe a novel method for discovering the local WFS in a nucleotide sequence by Monte Carlo simulation and RNA folding. In the approach the quality of a local WFS is assessed by the energy difference (E_{diff}) between the optimal structure folded in the local segment and its corresponding optimal, restrained structure where all the previous base pairings formed in the optimal structure are prohibited. Distinct WFS can be discovered by scanning successive segments along a sequence for evaluating the difference between E_{diff} of the natural sequence and those computed from randomly shuffled sequences. Our results indicate that the statistically significant WFS detected in the genomic sequences of Caenorhabditis elegans (C.elegans) F49E12, T07C5, T07D1, T10H9, Y56A3A and Y71G12B are coincident with known fold-back stem-loops found in miRNA precursors. The potential and implications of our method in searching for miRNAs in genomes is discussed.

1 Introduction

Complete genomic sequence data are accumulating at an unprecedented pace. The sequence data consist of an alphabet of four nucleotides (nt) with implicit rules for transcription, translation and cell organization. Analyses of massive sequence data in the database provide a great opportunity for us to explore the unrevealed biological properties. Currently, the computational tools for predicting novel protein coding genes in the complete genome are quite efficient and advanced. However, efficient methods for predicting novel non-coding RNAs (ncRNAs) and other RNA functional elements in whole-genome sequence data fall short.

Recent advances [1] in RNA studies indicate that RNA has functions in various biological processes beyond carrying the message for protein synthesis and ribosome structure. Included among these functions are catalysis and regulation of gene expression mediated by self-splicing ribozymes [1], small microRNAs (miRNAs) [2, 3], translational frame-shifting [4], translational repression elements [5], internal ribosome entry sequences [6], iron-response elements [7], constitutive transport elements and Rev response elements [8], and mRNA localization [9] etc.. Most of the functional RNA elements involve higher order structure rather than simple, linear sequence motif [1]. It has been estimated that 98% of the transcriptional output of human genome is ncRNAs [10]. The number of miRNAs in eukaryotic genomes was estimated from hundreds to thousands per genome [11]. These small RNA precursors of miRNAs often range from 60 to 80 nt and are able to form a conserved fold-back stem-loop structure in which the $\sim$ 22-25 nt miRNA sequence is within one arm of the stem-loop [12]. Knowledge discovery of such functional RNA sequences in genomic sequences by computational method is highly desirable.

Here, we describe SigED (Significant scan of Energy Difference), a computational method for detecting statistically significant well-ordered folding sequences (WFS) in a genomic sequence. SigED is designed to identify the local structural features by computing standardized z-scores, $SigZscr_e$ (see Methods section) and statistical evaluation in a random sample. The detected well-ordered structures are expected to be both thermodynamically stable and uniquely folded. Also, SigED is capable of finding unusually unstable folding regions and/or distinct loop structures. Using SigED, we search for the distinct fold-back stem-loops of miRNA precursors in the sequences of *Caenorhabditis elegans* (*C.elegans*). Our results show that the statistically significant WFS structures correlate with functional RNA elements. In conjunction with the previously developed EDscan [13] it is used in searching for potential, structured functional elements in genomic sequences.

2 Methods

Functional RNA molecules in modern organisms are presumably evolved to have distinct structure. It has been suggested that structured functional RNAs possess well-ordered conformations that are both thermodynamically stable and uniquely folded [13-16]. Schultes *et al* [16] recently proposed three measures to define the stability and uniqueness of RNA secondary structures. However, these measures did not directly consider the morphology of the folded structure. We [13] previously proposed a quantitative measure, E_{diff}, to characterize the thermodynamic stability and well-ordered conformation of a local RNA secondary structure in EDscan. The measure $E_{diff}(S_i)$ of a given RNA segment (S_i) is defined as the difference of free energies between the folded global minimal energy structure $(E(S_i))$ and its corresponding optimal restrained structure (ORS) in which all the previous base pairings in the lowest free energy structure are forbidden $(E_f(S_i))$. We have

$$E_{diff}(S_i) = E_f(S_i) - E(S_i) \quad and$$

$$Zscr_e(S_i) = \frac{E_{diff}(S_i) - E_{diff}(w)}{std(w)}$$

where $E_{diff}(w)$ and $std(w)$ are the mean and standard deviation, respectively, of the E_{diff} scores computed by sliding a fixed-length window in steps of a few nt from 5' to 3' along the sequence. The greater the $E_{diff}(S_i)$ (or $Zscr_e$) of the segment is, the more well-ordered the folded RNA structure is expected to be. The measure E_{diff} and $Zscr_e$ are obviously dependent on the structural feature of a local segment. The scores $Zscr_e$ of local segments in a sequence can be computed by the previously developed computer program EDscan [13].

What is the typical behavior of E_{diff} in a random sample that is related to the local segment? We adapt Monte Carlo simulations as the means of estimating the uncertainty of E_{diff} in a random sample. In Monte Carlo simulations, we first compute $E_{diff}(S_i)$ for a given segment S_i in the sequence, We then generate a large number of randomly shuffled sequences ($RS_{i,1}$, ... ,$RS_{i,m}$) for the local segment S_i, where the number m is determined by the length (LS_i) of the segment S_i. In the current version of SigED, we set $m = 100$ if $LS_i \leq 60$, $m = 200$ if $60 < LS_i \leq 150$ and $m = 300$ if $LS_i > 150$. In the random shuffling, nucleotides at all sites of the local segment are sequentially swapped with a randomly chosen site elsewhere in the segment [17]. As a result, randomly shuffled sequences have the same length and same base compositions as the natural segment S_i. Similarly, we can compute $E_{diff}(RS_{i,1}), \ldots, E_{diff}(RS_{i,m})$ for each random sequence and calculate their sample mean $E_{diff}(RS_i)$ and sample standard deviation $std(RS_i)$.

To facilitate statistical inference, we then define a standard z-score, $SigZscr_e(S_i)$ for the given segment.

$$SigZscr_e(S_i) = \frac{E_{diff}(S_i) - E_{diff}(RS_i)}{std(RS_i)}$$

In the random sample, the distribution of the random variable $SigZscr_e(RS_{i,j})$ $(1 \leq j \leq m)$ is expected to follow a Normal distribution. The statistical significance of the measure $E_{diff}(S_i)$ of the segment S_i can be easily estimated by means of the classical Normal distribution.

We can also divide E_{diff} into two parts, $Estem_{diff}$ and $Eloop_{diff}$, to characterize the structural features of the the base-pairing regions and loops, respectively. Where $Estem_{diff}$ is defined as the energy difference contributed by base-pairing stacking only between the lowest free energy structure and its corresponding ORS. $Eloop_{diff}$ is defined as the energy difference contributed by loops only between the two structures as mentioned above. Similarly, we can define the other two z-scores, $SigStem_e(S_i)$ and $SigLoop_e(S_i)$ for the given segment S_i. The measure $SigLoop_e$ and $SigStem_e$ can help us to evaluate a significantly unstable folding region that, for example, may be a potential target for hybridization of antisense agents.

The computer program SigED is designed to compute the three z-scores of $SigZscr_e$, $SigStem_e$ and $SigLoop_e$ by scanning successive segments along a nucleotide sequence. In the computation of the lowest free energy for a folded segment, SigED employs the dynamic programming algorithm and Turner energy rules [18-19]. SigED is implemented in Fortran 90 running on Unix, such as SGI/Octane and SGI/Onyx platform with IRIX 6.5. SigED may take considerable heavy computation for a long sequence. The program SigED is often used in conjunction with EDscan and works in a particular region that was previously determined to be of interest by EDscan [13].

3 Results and Discussion

3.1 The fold-back stem-loop of *let-7* RNA precursor is coincident with statistically significant WFS

It is known that *let-7* precursor can fold into a conserved fold-back stem-loop in *C. elegans*, *D. melanogaster* and human. The 21-nt *let-7* RNA of *C. elegans* is located at the 5' arm of the distinct stem-loop and 20 out of 21-nt are base paired. We searched for WFS in the *C. elegans let-7* gene sequence (Accession No. AF274345) by EDscan and SigED, respectively.

All scores are computed by moving the 75-nt window in steps of 3 nt from 5' to 3' along the sequence. Our results are displayed in Fig. 1. The global maxima of $Zscr_e$ and $SigZscr_e$ are 7.74 and 8.26, respectively, and are identified in the segment 1756-1830. It is clear that both EDscan and SigED can detect the WFS exclusively. The WFS 1756-1830 found in the gene sequence is statistically significant and the well-ordered structure is not expected by chance. Except for the distinct WFS, we also find the three other interesting regions, segment 1828-1902 ($Sigstem_e = -2.94$), 2038-2112

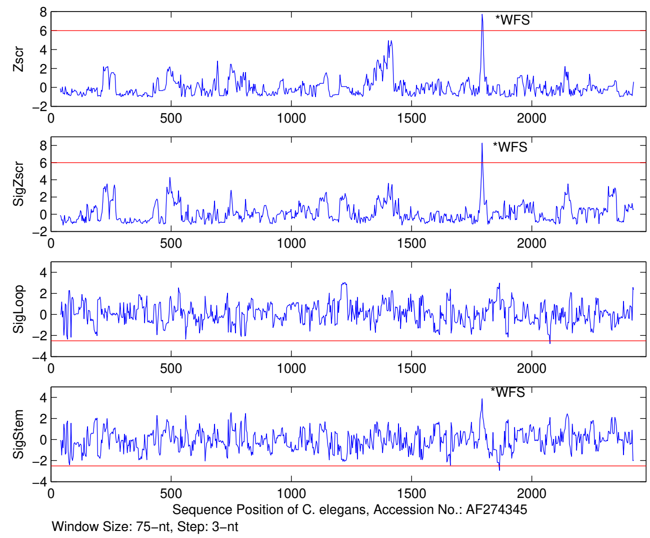

Figure 1. Distributions of the scores, $Zscr_e$ from EDscan and $SigZscr_e$, $SigLoop_e$ and $SigStem_e$ from SigED computed in *C. elegans let-7* RNA gene. In the computation, the 75-nt window was moved in steps of 3-nt from 5' to 3' along the sequence. The plot was produced by plotting the scores of each segment against the position of the middle nt in the segment. The maxima of $Zscr_e$, $SigZscr_e$ and $SigStem_e$ are found at the same segment 1756-1830. The detected WFS is coincidence with the *C. elegans let-7* RNA. The significantly unstable folding regions are detected in the segment 2038-2112 by the minimal $SigLoop_e$ (-2.79) and segment 1828-1902 by the minimal $SigStem_e$ (-2.94). The distinct loop structure is found in the segment 1186-1260 by the maximal $SigLoop_e$ (3.03).

($SigLoop_e = -2.79$) and 1186-1260 ($SigLoop_e = 3.03$) that contain significantly unstable regions or distinct loop structure, respectively (see Table 1). The capability of SigED is clear in that it indicates the statistical significance of WFS explicitly by means of the measure $SigZscr_e$. In contrast to the more rapid EDscan, SigED also identifies the significantly unstable folding region by the minimal $SigLoop_e$ and $SigStem_e$, and the distinct loop structure by the maximal $SigLoop_e$ in scanning the sequence.

3.2 Distinct structures of miRNA precursors and the corresponding WFS identified in *C. elegans* genomes

The 55 miRNAs encoded in *C. elegans* have been recently reported [20]. The 22-nt miRNA 81 (mir-81) was identified to be encoded in cosmid T07D1 sequence (Accession No. U41531). Figure 2 graphically displays the observed distributions of the scores $SigZscr_e$,

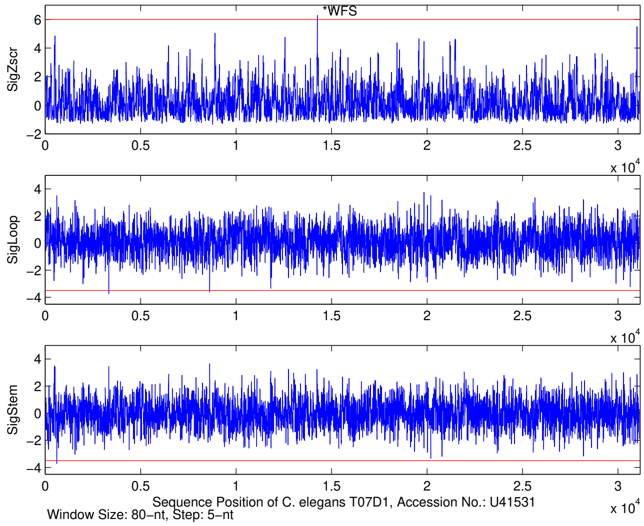

Figure 2. Distributions of the scores of $SigZscr_e$, $SigLoop_e$ and $SigStem_e$ computed in *C. elegans* T07D1 sequence. In the computation, the 80-nt window was moved in steps of 5-nt from 5' to 3' along the sequence. The maxima of $SigZscr_e$ (6.29) is identified in the WFS 14211-14290 that includes a miRNA mir-81 (14261-14282). Two significantly unstable folding regions are found in the segment 3291-3370 by the minimal $SigLoop_e$ (-3.73) and segment 566-645 by the minimal $SigStem_e$ (-3.72). For further details see the caption to Figure 1.

$SigLoop_e$, and $SigStem_e$ computed in the T07D1 sequence. The scores were computed by scanning the 80-nt window in steps of 5-nt from 5' to 3' along the T07D1 sequence (1-31150). The maximal score of $SigZscr_e$ was 6.29 and found in the segment 14211-14290. The 22-nt mir-81 (14261-14282) was located at the right arm of the fold-back stem-loop (14211-14288) and 18 nt out of 22-nt were in the base-pairing region. It is clear that the distinct WFS can be explicitly identified by $SigZscr_e$ from the $\sim$ 31100 observations. We also found two significantly unstable folding regions, segment 3291-3370 with $SigLoop_e = -3.73$ and 566-645 with $Sigstem_e = -3.72$.

We also searched for miRNAs, mir-50, mir-62, mir-70, mir-85 and mir-86 in the *C. elegans* genomic sequences. Among them, mir-70 and mir-86 were located at the genes T10H9.5 (32586-29037, Accession No. AF067949) and Y56A3A.7 (113210-100367, Accession No. AL132860) that were encoded in the reverse complementary strains. The mir-50 of 24-nt is located at the region 98223-98246 (17054-17077, where the start position of gene Y71G12B.11 was numbered

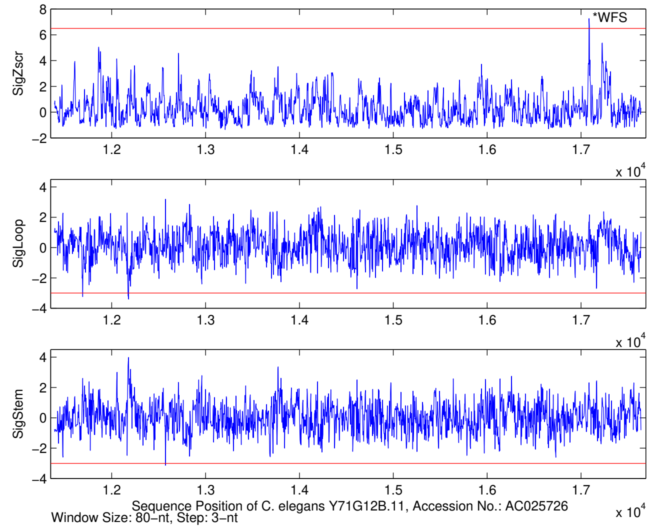

Figure 3. Distributions of $SigZscr_e$, $SigLoop_e$ and $SigStem_e$ computed in *C. elegans* Y71G12B.11 gene that is located at the region 81170-98845 of Y71G12B sequence. In the plot, the start position of Y71G12B.11 gene is numbered to position 1. The WFS 17044-17123 with $SigZscr_e = 7.27$ is identified by sliding the 80-nt window in a steps of 3-nt along the sequence. In fact, the WFS is located at the region 98213-98292 of Y71G12B sequence and encloses the 24-nt mir-50 (17054-17077). The two significantly unstable folding regions are found in the segment 12139-12218 with $SigLoop_e = -3.40$ and 12532-12611 with $SigStem_e = -3.13$. For further details see the caption to Figure 1.

by position 1) in the gene Y71G12B.11 (Accession No. AC025726). The detected WFS 98213-98292 (17044-17123) had $SigZscr_e = 7.27$ and was easily detected as shown in Figure 3. Mir-50 was located at the left arm of the fold-back stem-loop of the WFS and the 21 nt out of the 24-nt miRNA was in base-pairing (Table 1). Mir-62 of 22-nt was located at the region 11866-11887 of gene T07C5 sequence (Accession No. Z50006). The detected WFS 11829-11898 was coincident with the mir-62 (Fig. 4). The miRNA of 22-nt was in the right arm of the fold-back stem-loop of the WFS 11829-11898 (Table 1). Among them, 18 bases out of the 22-nt mir-62 were in the base-pairing region of the well-ordered structure. Similarly, we identified the fold-back structures of WFS in the genomic sequences of F49E12 (Accession No. Z66520), T10H9, and Y56A3A. The WFS 16090-16169 was detected by scanning a 80-nt window in steps of 3-nt along the F49E12 (40450-nt) sequence and its $SigZscr_e$ was 5.82, the maximal score in the region (Fig. 5). This is statistically very signifi-

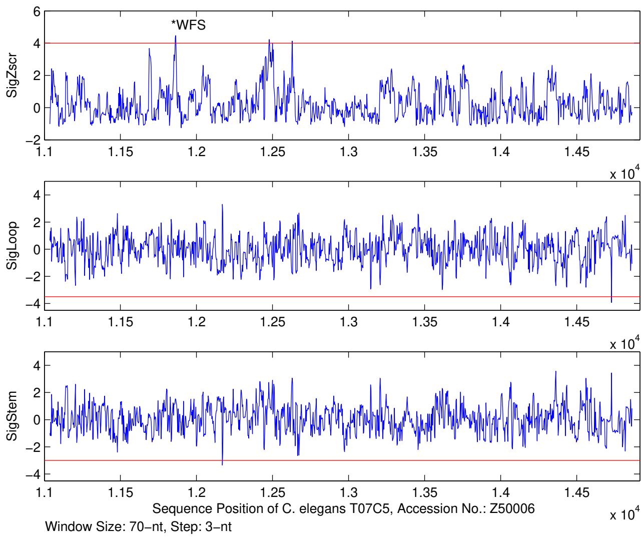

Figure 4. Distributions of $SigZscr_e$, $SigLoop_e$ and $SigStem_e$ computed in *C. elegans* T07C5 sequence. The WFS 11829-11898 with $SigZscr_e = 4.46$ is identified by sliding the 70-nt window in a steps of 3-nt along the sequence. The 22-nt mir-62 is located at 11866-11887. The unstable folding regions are detected in the segment 14694-14763 with $SigLoop_e = -3.93$ and segment 12135-12204 with $SigStem_e = -3.34$. For further details see the caption to Figure 1.

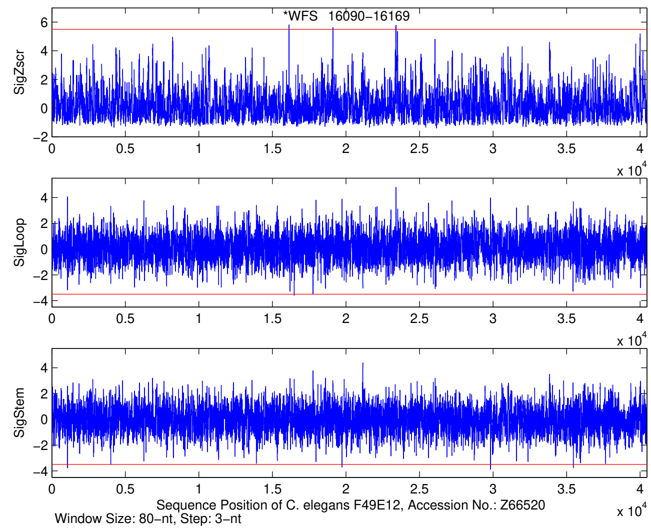

Figure 5. Distributions of $SigZscr_e$, $SigLoop_e$ and $SigStem_e$ computed in *C. elegans* F49E12 sequence. The plot was produced by plotting the three scores of a 80-nt segment against the position of the middle base in the segment as it was moved successively by 3-nt each time from 5' to 3' along the sequence. The maximal $SigZscr_e$ is 5.82 and found in the segment 16090-16169 that encloses a 24-nt mir-85 (16140-16163). The another interesting WFS 23353-23432 contains a distinct loop structure with the maximal $SigLoop_e$ of 4.80. Several unstable folding regions are also found in the computation. The two of them are region 16441-16520 with the minimal $SigLoop_e$ (-3.60) and region 29791-29870 with the minimal $SigStem_e$ (-3.87).

cant and not be expected to occur by chance. The 24-nt mir-85 was located at the right arm of the fold-back stem-loop (16090-16169 in F49E12) and its 20 nt of the 24-nt miRNA are in base-pairing region.

The two miRNAs mir-70 and mir-86 were detected in the reverse complementary sequence (RCS) of their corresponding loci. The distinct fold-back structure of mir-70 precursor was coincident with the WFS 32400-32321 (RCS) that was identified by scanning a 80-nt window along the gene T10H9.5 sequence (32586-29037, RCS). Mir-86 precursor was coincident with the WFS 102922-102858 (RCS) that was detected by scanning a 65-nt window along the gene Y56A3A.7 (113210-100367, RCS). The $SigZscr_e$ values of the two detected WFS were 3.48 and 3.50, respectively (Table 1). The likelihood of the two WFS in the random sample are less than 0.005.

Draper [15] has suggested that RNA functional elements whose folded structure conformation plays a crucial role are known to possess specific structural features that are both thermodynamically stable and uniquely folded. In the method SigED, the conformational property in the RNA secondary structure is measured based on the consideration of both thermodynamic stability and uniqueness of the minimal free energy structures folded in both the natural and randomly shuffled sequences. Our computational results show that WFS with high $SigZscr_e$ are coincident with the reported miRNAs in our tested examples. Statistical extremes with high $SigZscr_e$ determined by SigED represent such significant folding segments where predicted structures are expected to be well-ordered.

Recently, we reported a computational method EDscan [13] to discover WFS in nucleotide sequences. Both the EDscan and SigED provide a computational tool for discovery of WFS. Those detected 60-80 nt WFS can be used to explore those undiscovered miRNAs in eukaryotic genomes with the help of structure and sequence comparison. But the method SigED puts our emphasis on the statistical inference for the uniqueness of the morphology of the folded RNA secondary structure. Our computational experiments indicate that the new method can also offer a means of searching for unusually unstable folding regions and distinct loop structures in a genomic sequence. The new features added in SigED should be useful in the theoretical design of antisense RNA structures in antisense therapeutics. The previously developed EDscan however failed of success in characterizing the unstable structural features from the bulk distribution of $Zscr_e$ [13].

With the combination of the two methods EDscan and SigED we can do the computational experiment in a large scale. In general, we can find those potential interesting regions with high $Zscr_e$ in a genome sequence by the more rapid EDscan and sliding various fixed-length windows with a range of sizes. Then we can evaluate the random probabilities of those WFS by SigED and localize the critical sequence for detailed analysis of its secondary and/or tertiary structure. SigED provides us a new computational tool for the determination of potential functional elements with structure dependent functions that may be exploited for possible regulatory or therapeutic purposes.

Acknowledgments

The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. The source code of SigED is available at the web http://protein3d.ncifcrf.gov/shuyun/siged.html.

References

[1] Simons, R.W. and M. Grunberg-Manago, eds. *RNA Structure and Function* Cold Spring Harbor Lab. Press, New York, 1998.

[2] G. Storz, An expanding universe of noncoding RNAs. *Science* **296,** 2002, pp. 1260-1263.

[3] S.R. Eddy, Non-coding RNA genes and the modern RNA world.. *Nature Rev Genet* **2,** 2001, pp. 919-929.

[4] I. Brierley, P. Digard, and S.C. Inglis, Characterization of an efficient coronavirus ribosomal frameshifting signal requirement for an RNA pseudoknot. *Cell* **57,** 1989, pp. 537-547.

[5] J. Dubnau, and G. Struhl, RNA recognition and translational regulation by a homeodomain protein. *Nature* **379,** 1996, pp. 694-699.

[6] C.U.T. Hellen, and P. Sarnow, Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes & Development* **15,** 2001, pp. 1593-1612.

[7] M.W. Hentze, S.W. Caughman, J.L. Casey, D.M. Koeller, T.A. Rouault, J.B. Harford, and R.D. Klausner, A model for the structure and functions of iron-responsive elements. *Gene* **72,** 1988, pp. 201-208.

[8] M.H. Malim, J. Hauber, S.-Y. Le, J.V. Maizel Jr., and B.R. Cullen, The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA, *Nature* **338,** 1989, pp. 254-257.

[9] P.M. Macdonald, and C.A. Smibert, Translational regulation of maternal mRNAs. *Curr Opin Genet Dev* **6,** 1996, pp. 403-407.

[10] J.S. Mattick, Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep.* **2,** 2001, pp. 986-991.

[11] G. Riddihough, The other RNA world. *Science* **296,** 2002, pp. 1259.

[12] V. Ambros, *et al.* A uniform system for microRNA annotation. *RNA* **9,** 2003, pp. 277-279.

[13] S.Y. Le, J.H. Chen, D. Konings, and J.V. Maizel, Jr. Discovering well-ordered folding patterns in nucleotide sequences. *Bioinformatics* **19,** 2003, pp. 354-461.

[14] S.Y. Le, K. Zhong, and J.V. Maizel, Jr. RNA molecules with structure dependent functions are uniquely folded. *Nucl. Acids Res* **30,** 2002, pp. 3574-3582.

[15] D.E. Draper, Strategies for RNA folding. *Trends Biochem Sci* **21,** 1996, pp. 145-149.

[16] E.A. Schultes, P.T. Hraber, and T.H. LaBean, Estimating the contributions of selection and self-organization in RNA secondary structure. *J. Mol. Evol.* **49,** 1999, pp. 76-83.

[17] Knuth, D. The art of computer programming. Vol. 3. Addison-Wesley, Reading: MA, 1973.

[18] M. Zuker, and P. Steigler, Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.* **9**, 1981, pp. 133-149.

[19] D.H. Mathews, J. Sabina, M. Zuker, and D.H. Turner, Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* **288,** 1999, pp. 911-940.

[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* **294,** 2001, pp. 858-862.

```
Table 1.  WFS and its fold-back stem-loop structures computed in C. elegans

1. let-7 gene of C. elegans
  folding region: 1756 ~ 1830; energy = -37.7, SigZscr: 8.26, SigStem: 3.88
gaTCCGGTGAGGTAGtAGGTTGTATAGTTTGGaatattaccaCCggtGAACTATGCAATTTtCTACCTTACCGGA
       --------------------- (let-7)

  folding region:  2038 ~ 2112; energy = -8.5, SigLoop: -2.79
GGTAaTGTATctggAGAgaataaTCTaatcGTATGTACTGTtgAgtaATGtaTCcatgGAgcCGTTtgACatttc

  folding region:  1186 ~ 1260; energy = -11.8, SigLoop: 3.03
catctttcGTGTACATTTGcaacattttctggatcatcaatCAAGTGTGCACtgaccactctctctctcctccta

  folding region:  1828 ~ 1902; energy = -9.7, SigStem: -2.94
GGAGacAGAActcttCGAAGctGCGtcgtctTGCtctcacaactttctTTTCGttTTCTtcCTCCtcttactttc
  The significant regions were computed by SigED and moving the 75-nt
window (W75) in steps of 3 nt from 5’ to 3’ along the sequence.

2. T07C5.1, W70, WFS: 11829-11898, SigZscr = 4.46
  folding region: 11829 ~ 11898; energy = -25.5
gGTGAGTTAGATctCATATCcTTCCGcaaaaTGGAAatGATATGtaATCTAGCTTACagGTTgttcAACa
      11866-11887 (22-nt),  mir-62   ----------------------

3. T07D1.2, W80, WFS: 14211-14290, SigZscr = 6.29
  folding region: 14211 ~ 14290; energy = -35.9
CCAACAGtcGGTTTTCACcGTGATCTgAGAGCAatccaaaaaTGCtttTCTgAGATCATcGTGAAAGCTagTTGTTGGct
                   14261-14282 (22-nt),  mir-81   ----------------------

4. Y71G12B.11a, W80, WFS: 98213-98292(17044-17123), SigZscr = 7.27
  folding region: 98213 ~ 98292; energy = -42.3, Position 81170 was numbered by 1.
cgCCGGCCGcTGATATGTCTGGTATtctTGGGTTTgAACttccagcGTTGAACCCGcATATTAGACGTATCGaCGGCCGG
          ------------------------  mir-50, 98223-98246 (17054-17077) (24-nt)

5. F49E12.8, W80, WFS: 16090-16169, SigZscr = 5.82
  folding region: 16090 ~ 16169; energy = -32.8
gtCGGAGCcCGATTTTTCAATAgTTTGaAACCAGTGtacaCATaaaTGGTTaCAAAgTAtTTGAAAAGTCGtGCTCTGaa
                   16140-16163 (24-nt),  mir-85   ------------------------

6. T10H9.5, W80, WFS: 32400-32321(187-266), SigZscr = 3.48
  folding region: 187 ~ 266; energy = -25.2
atgaAAACTATcGAAATACTAtCGACGaATaACAcTTatgaagAAaTGTaATaCGTCGtTGGTGTTTCcATAGTTTgaat
        Reverse complementary sequence (RCS)    -----------------------
        32352-32330 (235-257) (23-nt),  mir-70
Position 32586 was taken as position 1, and position 29037 was taken as
3550. The gene T10H9.5 is at reverse complementary strain (32586-29037).

7. Y56A3A.7, W65, WFS: 102922-102858 (10289-10353), SigZscr = 3.50
  folding region: 102922(10289) ~ 102858(10353); energy = -22.3
tcTAAGTGAATgCTttGCCaCAGtCTTCgATgttctgaaATGAAGcCTGGGCtcAGATTCGCTTA
  -----------------------  mir-86, RCS 102920-102898 (10291-10353) (23-nt)
Position 113210 was taken as position 1, and position 100367 as 12844.
The gene Y56A3A.7 is at reverse complementary strain (113210-100367)

All structures were computed by mfold [19] and the base-pairing regions
in the structures were indicated by capital letters.
```