# Discovery of Structural Feature in Nucleic Acid Sequences by Computational Methods

- 1. Development of New Algorithms in Analyses of RNA Structure.

- **Chen, J.-H.,** ABCC, SAIC; **Liu, W.-M.**, Indiana Univ.
- **Zhang, K.**, Dept. of Comp. Sci., Univ. of Western Ontario, Canada.

- 2. Data Mining of RNA Functional Elements

- **Lab of Cullen, B.R.** , Duke Univ. Medical Center
- **Lab of Elroy-Stein, O.**, Tel-Aviv Univ.
- **Lab of Groner, Y.** , Weizmann Institute of Sciences

# 1. Development of New Algorithms in Analyses of RNA/DNA Structures

- A. **Development of the computer program ed_scan in searching for statistically significant 'well-determined' folding patterns in RNA/DNA sequences.**

- B. **Development of the computer program st_comp that is used to determine if the predicted RNA structure is well-ordered structure.**

- **The well-determined or well-ordered structures are both thermodynamically stable and distinctly folded in RNA sequences.**

- C. **Development of empirical energy rules for guiding RNA folding based on RNA structural database and folding principle of current thermodynamic experimental data using genetic algorithm**

# D. Modification and improvement of two computer programs

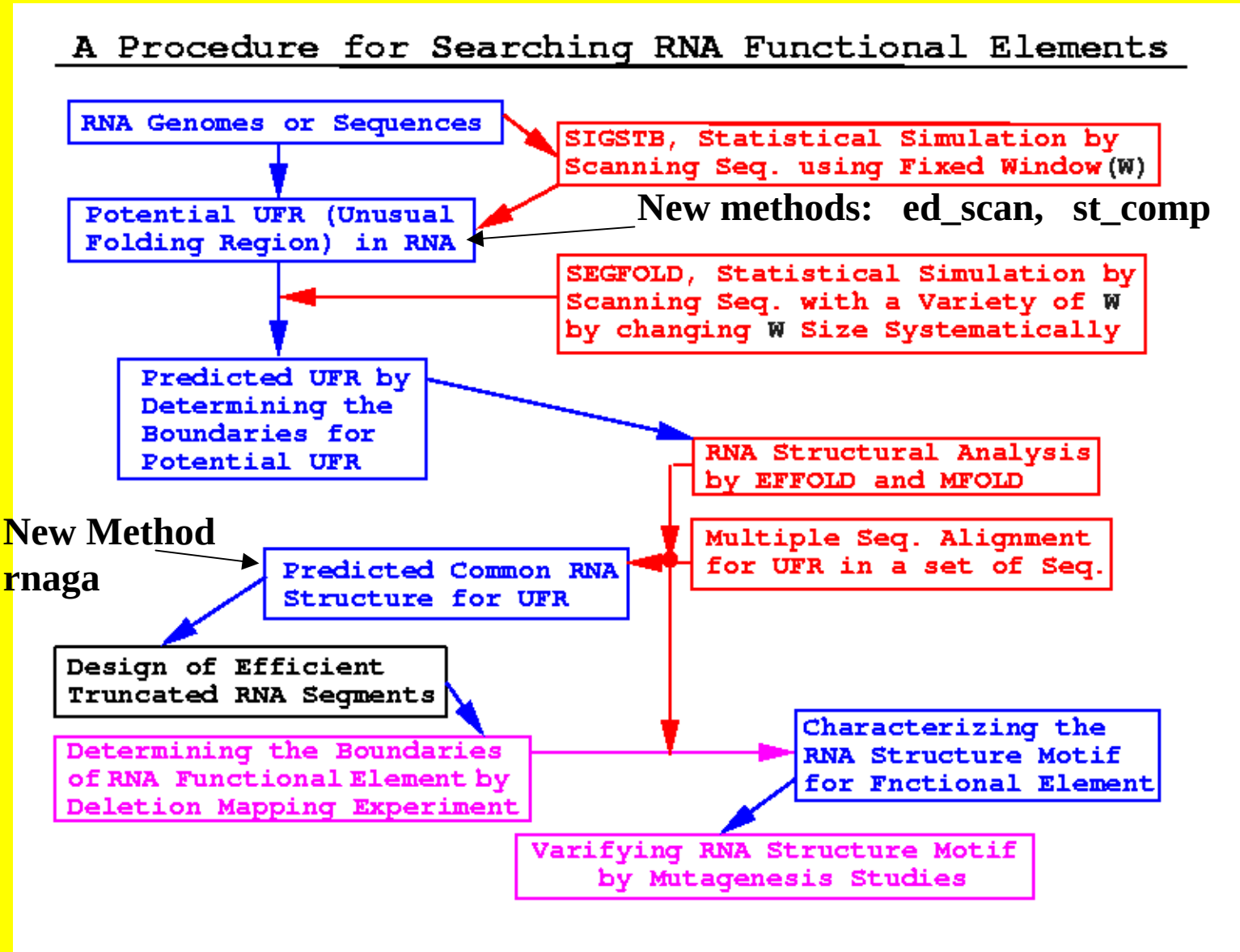
1. **RNA_match computes a similarity measure, maximal matching score, between two RNA structures.**

2. **RNAGA employs a genetic algorithm (GA) to search for a common secondary structure without the need for pre-aligned homologous RNA sequences. The predicted common structures are automatically optimized by not only the thermodynamic stability but also the structural similarity among homologous sequences.**

# 2. Data Mining of RNA Functional Elements

- **A. Data mining of functional elements in 24 microbial genomes.**
- **B. Searching for significant unusual folding regions (UFRs) in 1196 orthologous mouse and human full-length mRNAs**
- **C. Searching for significant UFRs in 5' and 3' UTR sequences.**
- **D. Y-shape structure motif in the 5'UTR of onco-protein mRNAs.**
- **E. Common RNA structure of the RRE elements of 151 HIV-1, 18 HIV-2, and 30 SIV mRNAs**
- **F. Structure comparison for Rnase P RNA structures in the database.**

# Our Strategy and Tactics :



# Searching for Well-determined Folding Patterns in RNA/DNA (program ed_scan)

Hypotheses:

A. **Evolutionary constraints imposed by structural property of functional RNA elements or RNA molecules are to have a well-determined structure that is both thermodynamically stable and well-ordered folded.**

B. **Local folding in DNA and RNA is closely associated with its biological functions.**

C. **The distinct, well-determined structure folded by a local segment is closely correlated with a functional element whose structure property play a crucial role in the gene expression.**

Method:

1. **Defining the Uniqueness of RNA/DNA Secondary Structures**.

A. **The lowest free energy (E) of the global minimal structure in a local segment**

B. **The lowest free energy ($E_f$) of the folded, restrained structure in which all base-pairs in the global minimal structure are forbidden to form.**

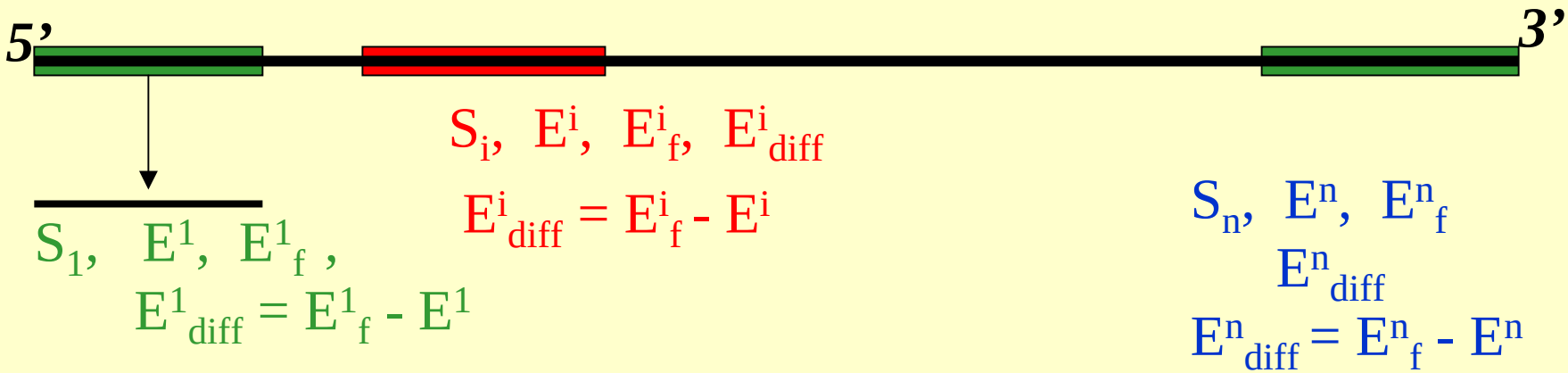
C. **The quality of the well-determined structure of a local segment is quantitatively measured by the energy difference ($E_{diff}$) and its z-score($Zscr_e$)**

$E_{diff} = E_f - E$ and $Zscr_e = (E_{diff} - E_{diff}(w))/std_w$

**$E_{diff}(w)$ and $std_w$ are the sample mean and standard deviation of the $E_{diff}$ scores computed by sliding a fixed window stepped a few nucleotides each time from 5' to 3' along the test sequence**.

2. How ed_scan finds statistically well-determined folding patterns in a sequence.

A. The principle: choose successive, overlapping segments




**$E_{diff}(w)$ and $std_w$ are a sample mean of $E^1_{diff}$,…,$E^i_{diff}$, and $E^n_{diff}$**

$Z^1scr_e = (E^1_{diff} - E_{diff}(w))/std_w$; $Z^iscr_e = (E^i_{diff} - E_{diff}(w))/std_w$

## B. General Procedure:

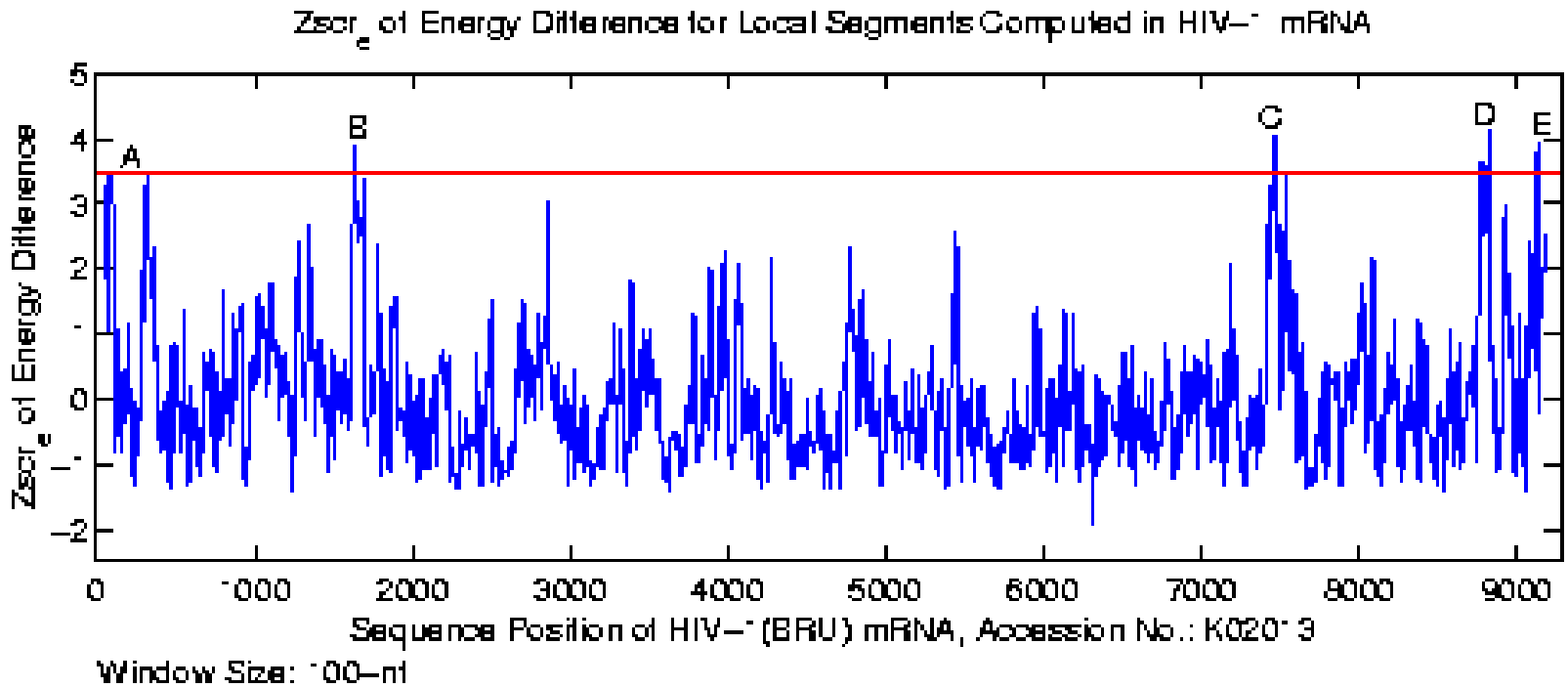
1. **$Zscr_e$ values are computed in the sequence by sliding a fixed window with a proper size (for example 100-nt).**

2. **Detection of the potential interesting regions that have high $Zscr_e$ values based on the profile of $Zscr_e$ in the step 1.**

3. **The precise locations of those potential targets in which the folded structure is highly well-determined are inferred by an extended search in the regions determined from the step 2.**
   **For example, window size can be changed from 80 to 300-nt.**

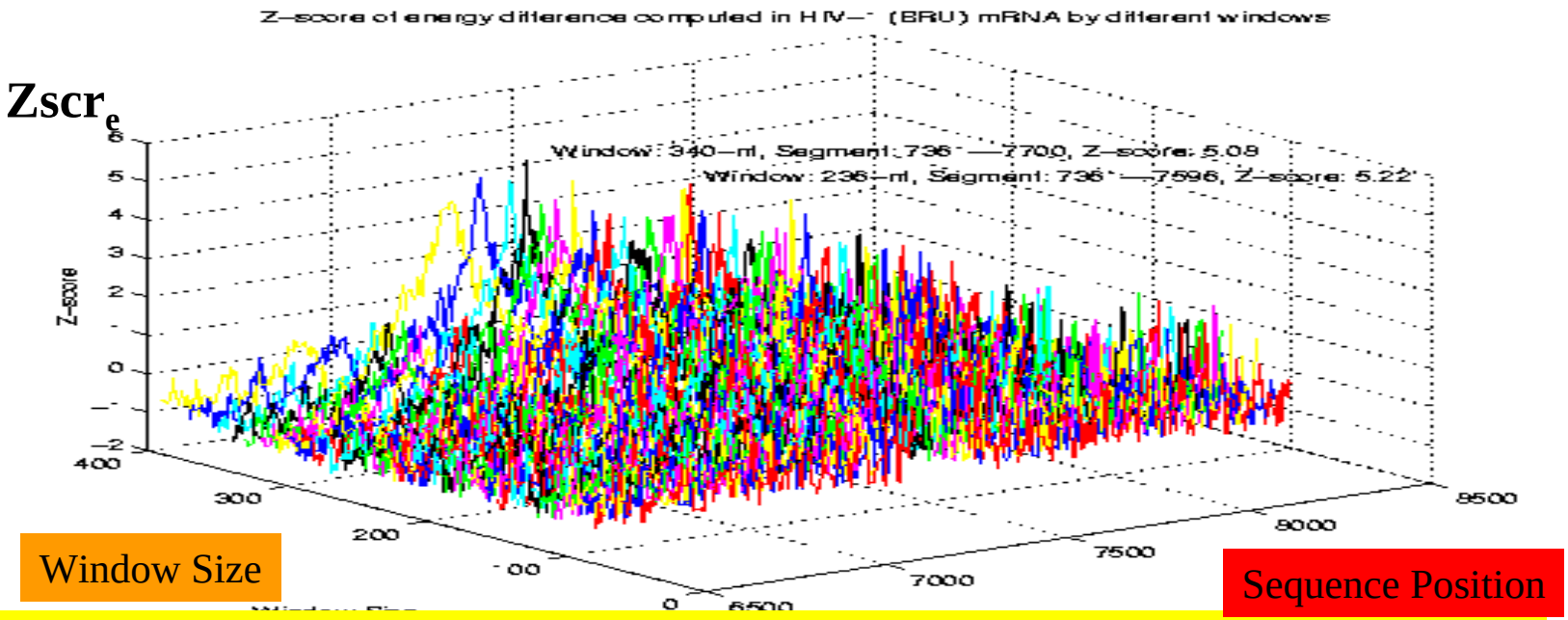
4. **The optimized well-determined folding regions in the sequence are determined by a 3D plot ($Zscr_e$, window size, and the segment position)**

# $Zscr_e$ of the free energy difference between the optimized and its corresponding restrained structure of a local segment in HIV-1 (isolate BRU) mRNA sequence




Window Size: 100-nt; Stepped by 5-nt each time

# A 3-D plot of $Zscr_e$, the start position of a local segment and the window size in the searching domain (6501-8500) of HIV-1 (BRU) mRNA

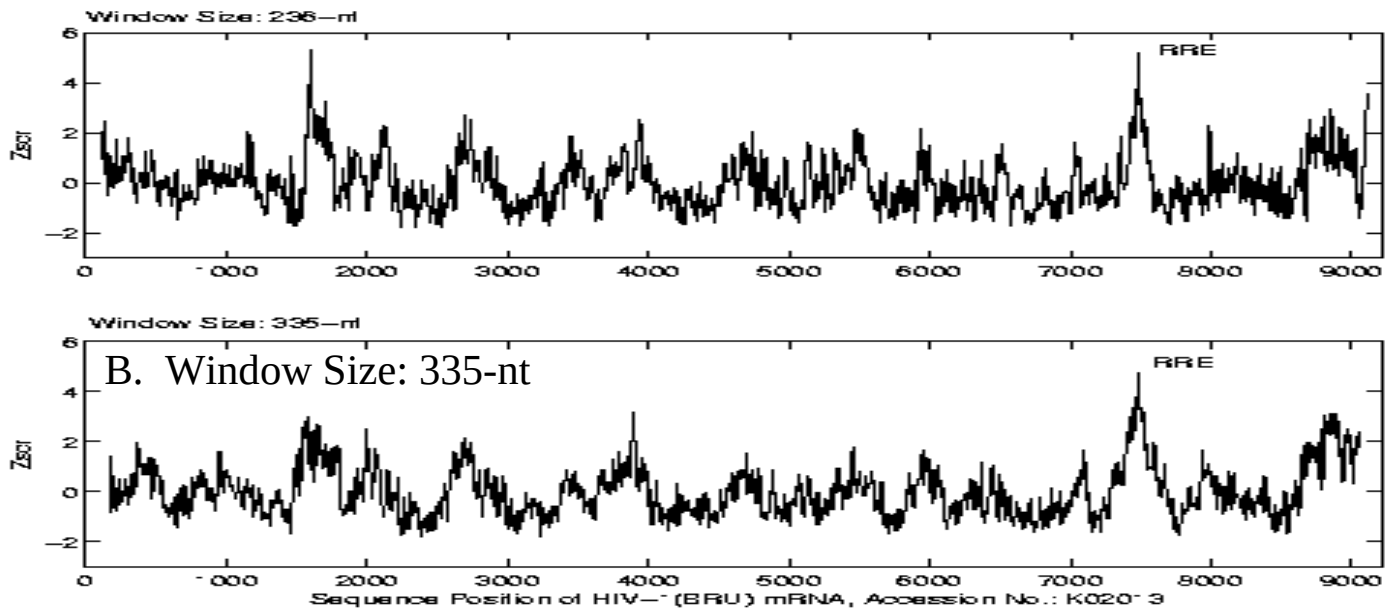



Red: $Zscr_e$ (= 5.22) was computed by window of 236-nt, corresponding segment 7361-7596

Black: $Zscr_e$ (=5.01) was computed by window of 340-nt, corresponding segment 7311-7650

**The window size was systematically changed from 80 to 350-nt and plotted by magenta, green, black, cyanic red, blue and yellow, respectively. In addition to these data, $Zscr_e$ computed by 236, 370 and 390-nt were plotted by red, blue and yellow**

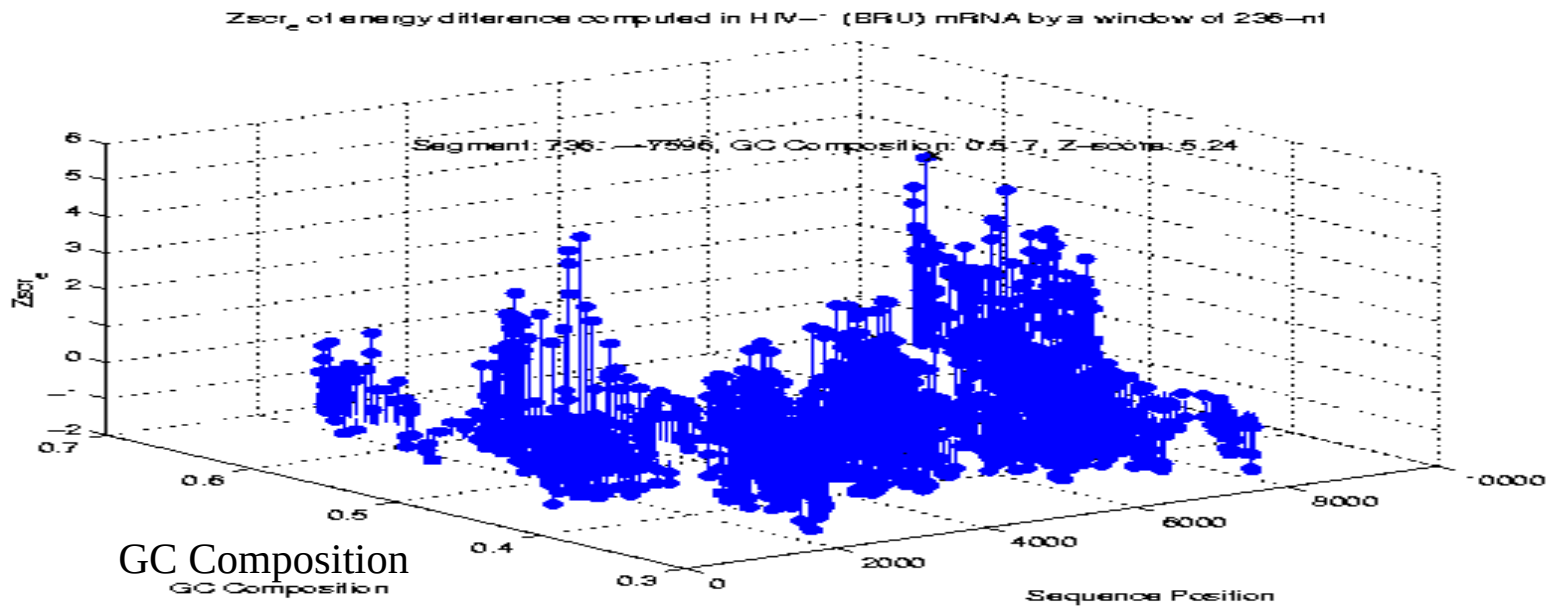
$Zscr_e$ of the free energy difference between the optimized and its corresponding restrained structure of a local segment in HIV-1 (isolate BRU) mRNA sequenc




Sequence position

The plot was produced by plotting $Zscr_e$ of a local segment against the position of the middle base in the segment and stepped by 5-nt each time from 5' to 3' of the sequence

A 3-D stem plot of $Zscr_e$, GC composition and the start position of a local segment computed in HIV-1 (BRU) mRNA by a fixed window of 236-nt

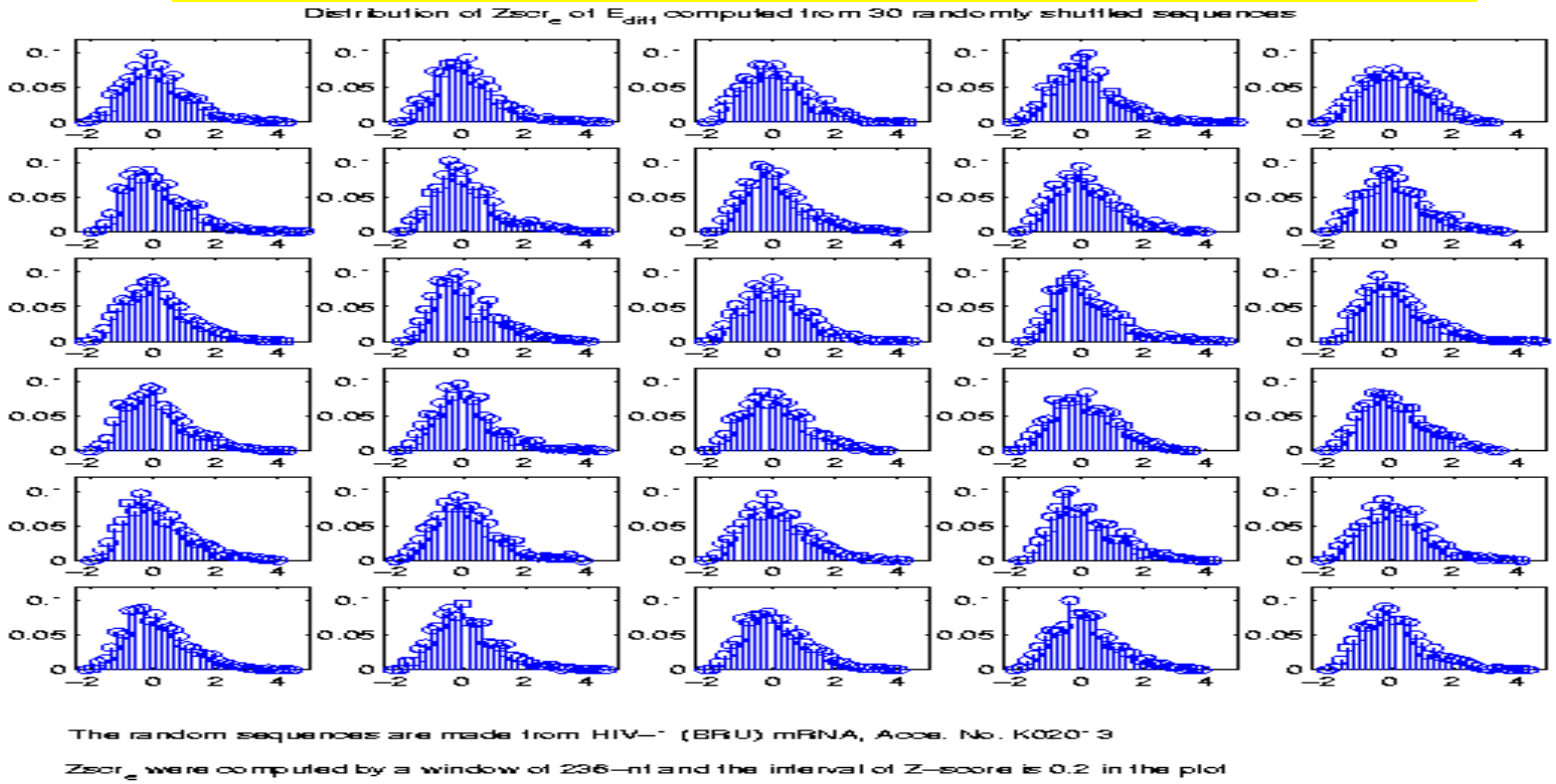



Sequence Position

The maximum $Zscr_e$ (= 5.24) corresponds to the well-determined folding segment 7361-7596 whose GC composition is 0.517.

**The GC composition of the 1799 observations ranged from 0.31 to 0.58**

**Distributions of $Zscr_e$ of the free energy difference computed from 30 randomly shuffled sequences of HIV-1 (BRU) mRNA.**



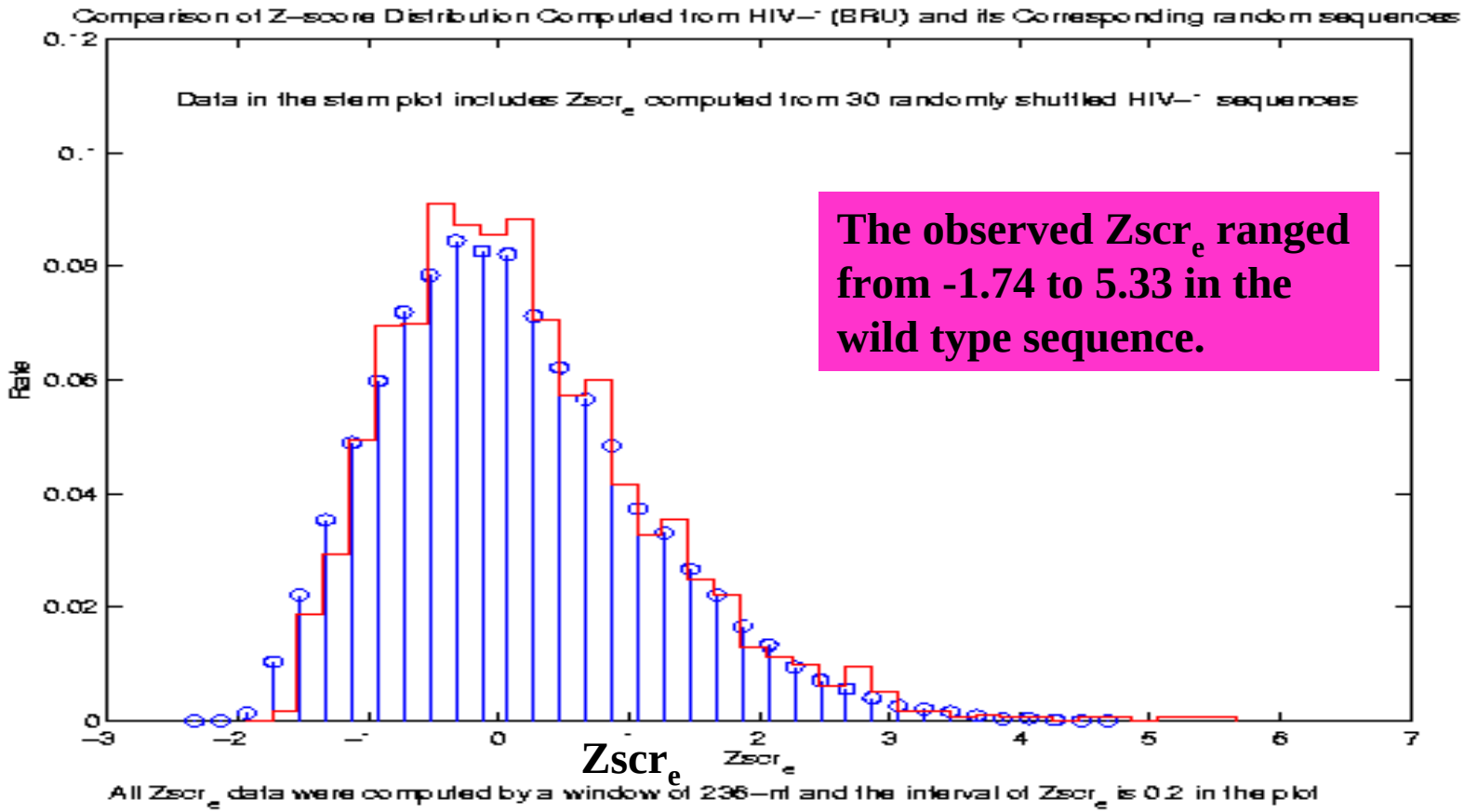

**$Zscr_e$ were computed by a window of 236-nt and the interval of $Zscr_e$ is 0.2 in the plots.**

**Only 4 out of total 53,970 observations of $Zscr_e$ were greater than 4.5, and the scores ranged from -2.13 to 4.723.**

# Comparisons of $Zscr_e$ distributions computed from HIV-1(BRU) mRNA and its corresponding 30 randomly shuffled sequences

Red: data are from the wild type HIV-1(BRU)
Blue: data are from the 30 randomly shuffled sequences

**Fraction**

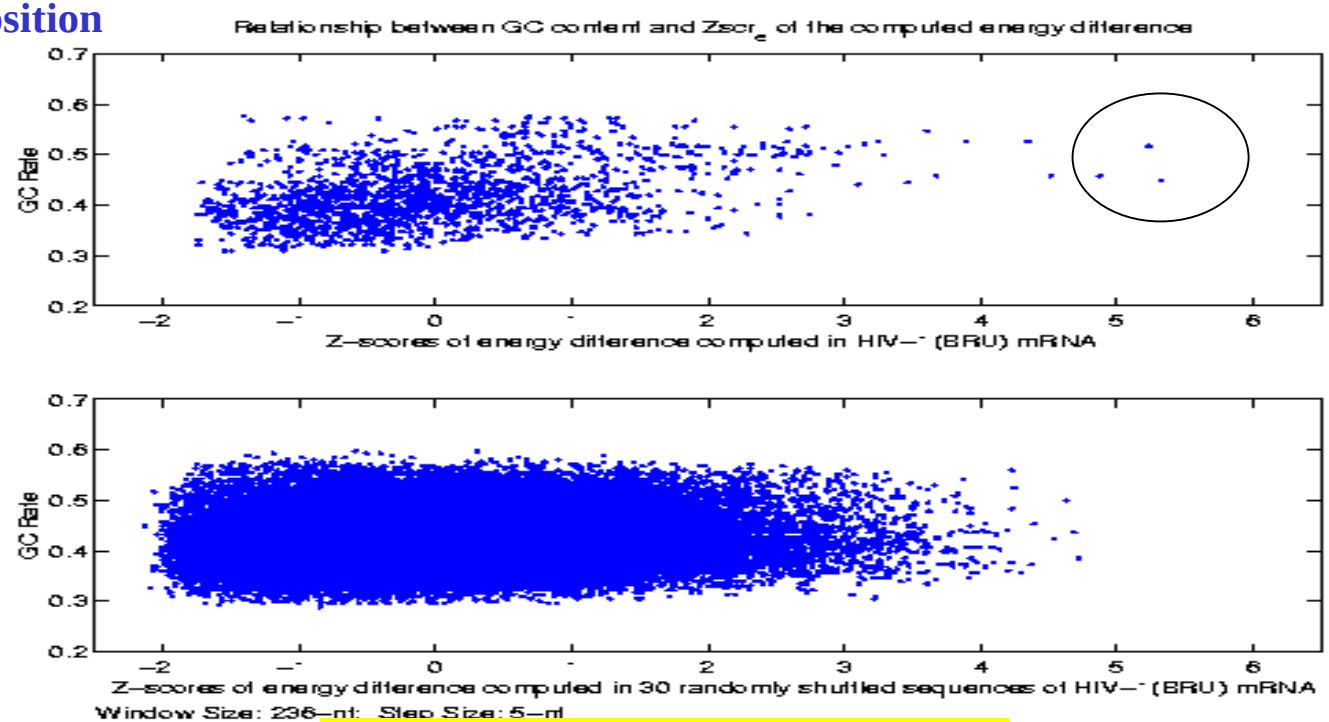



**The observed $Zscr_e$ ranged from -1.74 to 5.33 in the wild type sequence.**

**All $Zscr_e$ data were computed by window of 236-nt, and the interval of $Zscr_e$ in the plot is 0.2**

**Relationships between GC composition and $Zscr_e$ of the local segment of 236-nt computed in the wild type sequence of HIV-1 (top) and 30 randomly shuffled sequences (bottom) of the wild type sequence.**

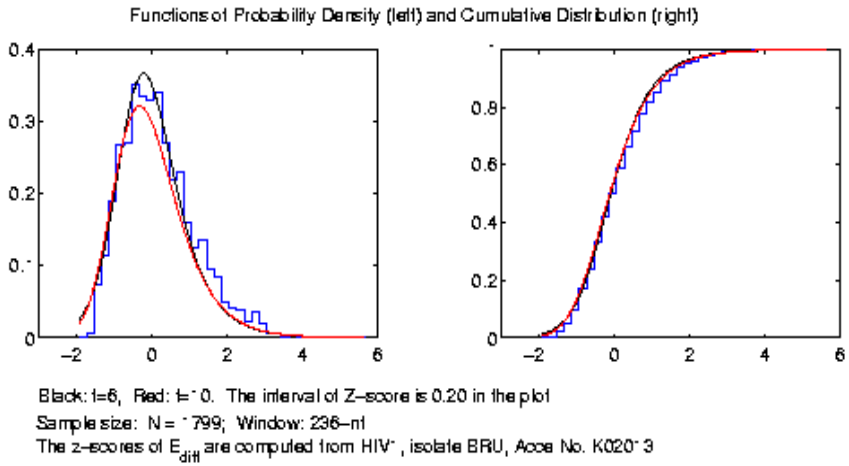
**GC Composition**




**$Zscr_e$ of free energy difference**

**The significant well-determined folding patterns are apparently separated from the bulk distribution in the wild type sequence.**

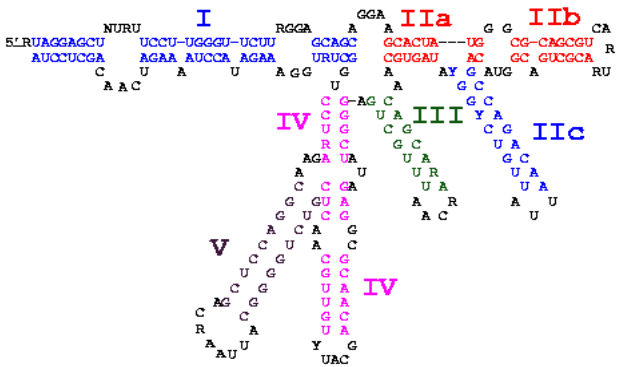
**Empirical probability density function (left) and empirical distribution function (right) plotted together with linear transformed, theoretical probability density functions (left) and cumulative distribution functions (right) of noncentral *t* distribution.**

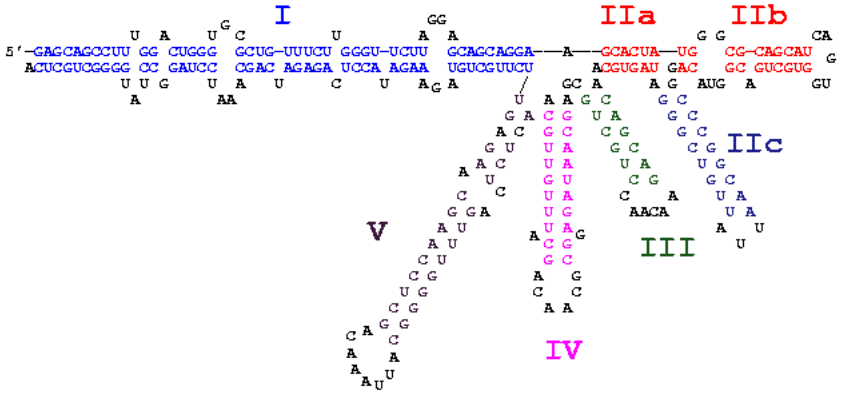



**The plot is for Zscr$_e$ data (1799 observations) computed from the wild type HIV-1 (BRU) mRNA by a fixed window of 236-nt.**

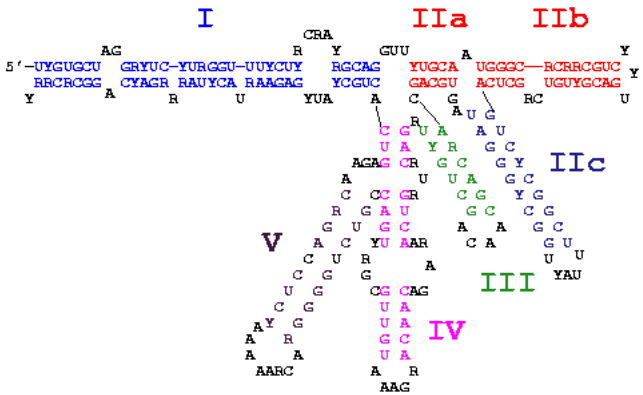
**The theoretical curves with the degree of freedom *f* = 6 are draw in black and those with *f = 10 are* drawn in red. The empirical step functions in the plot are plotted with step size 0.2.**

Common Structure of HIV-1 RRE

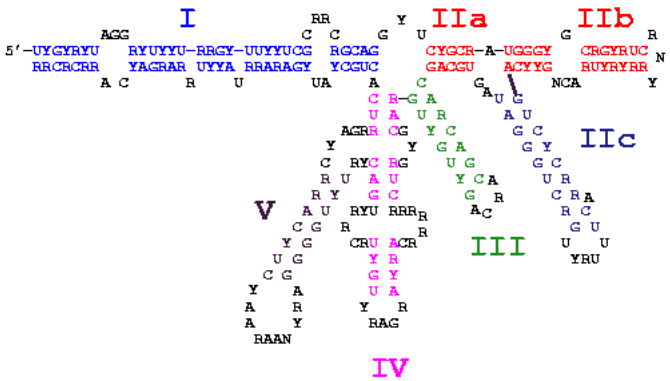



Common Structure of SIVcpz RRE




Common Structure of HIV2 RRE




Common Structure of SIV RRE

# *let-7 RNA*

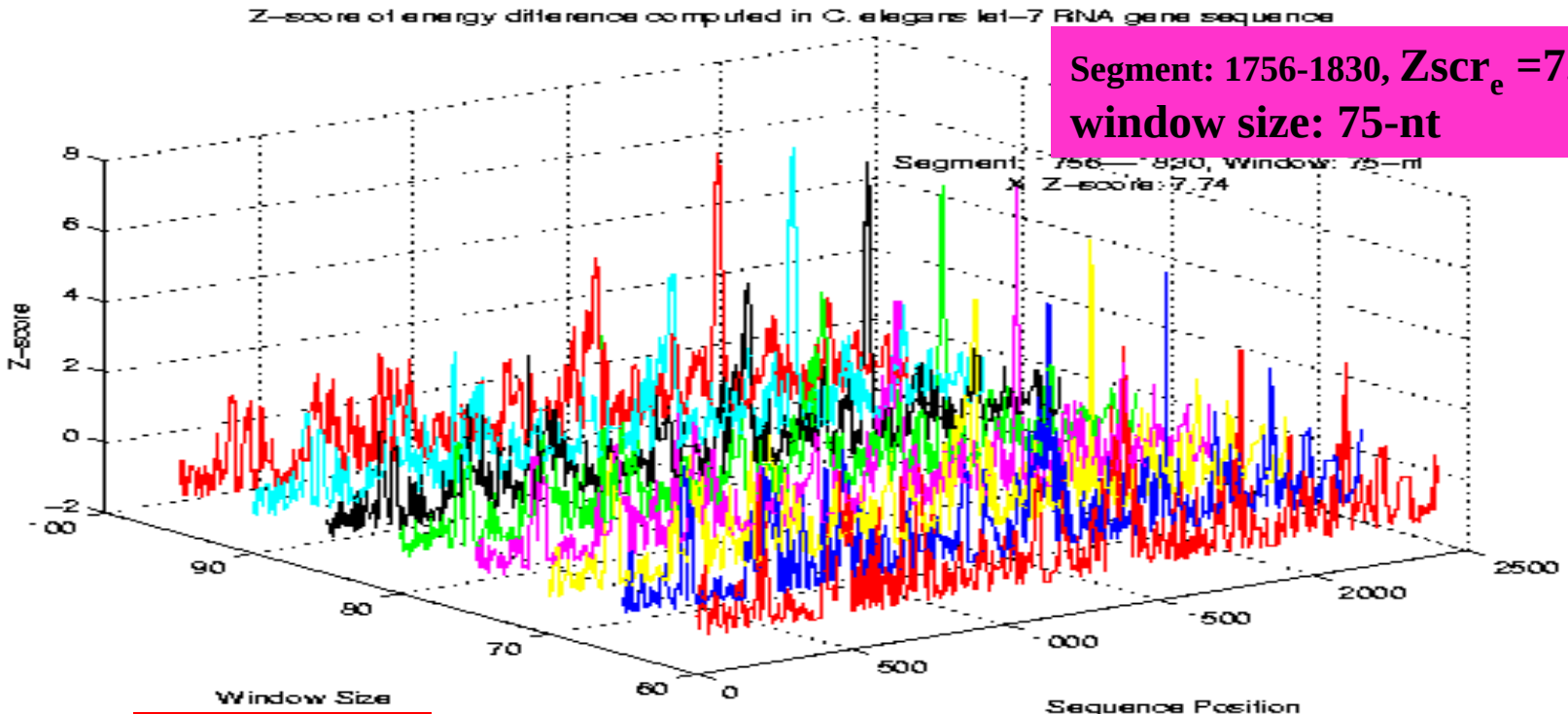
***The let-7 RNA* is a small RNA having about 21 nucleotides. The sequence and function of *let-7 RNA* is conserved in the *nematode Caenorhabaditis elegans* and *C. briggsae*. It may control late temporal transitions during development across animal phylogeny.**

**The small *let-7 RNA* regulates the timing of *C. elegans* development.**

**Its sequence is complementary to the sequence in the 3' untranslated region of a set of protein-coding target genes that are normally negatively regulated by the RNAs**

**The 21-nt *let-7 RNA* can be folded to a stable stem-loop structure (~73 nt) with nearby sequence.**

**A 3-D plot of $Zscr_e$ , the start position of a local segment and the window size computed in *C. elegans let-7* RNA gene sequence (Acce. No. AF274345)**




**Segment: 1756-1830, $Zscr_e$ =7.74**
**window size: 75-nt**

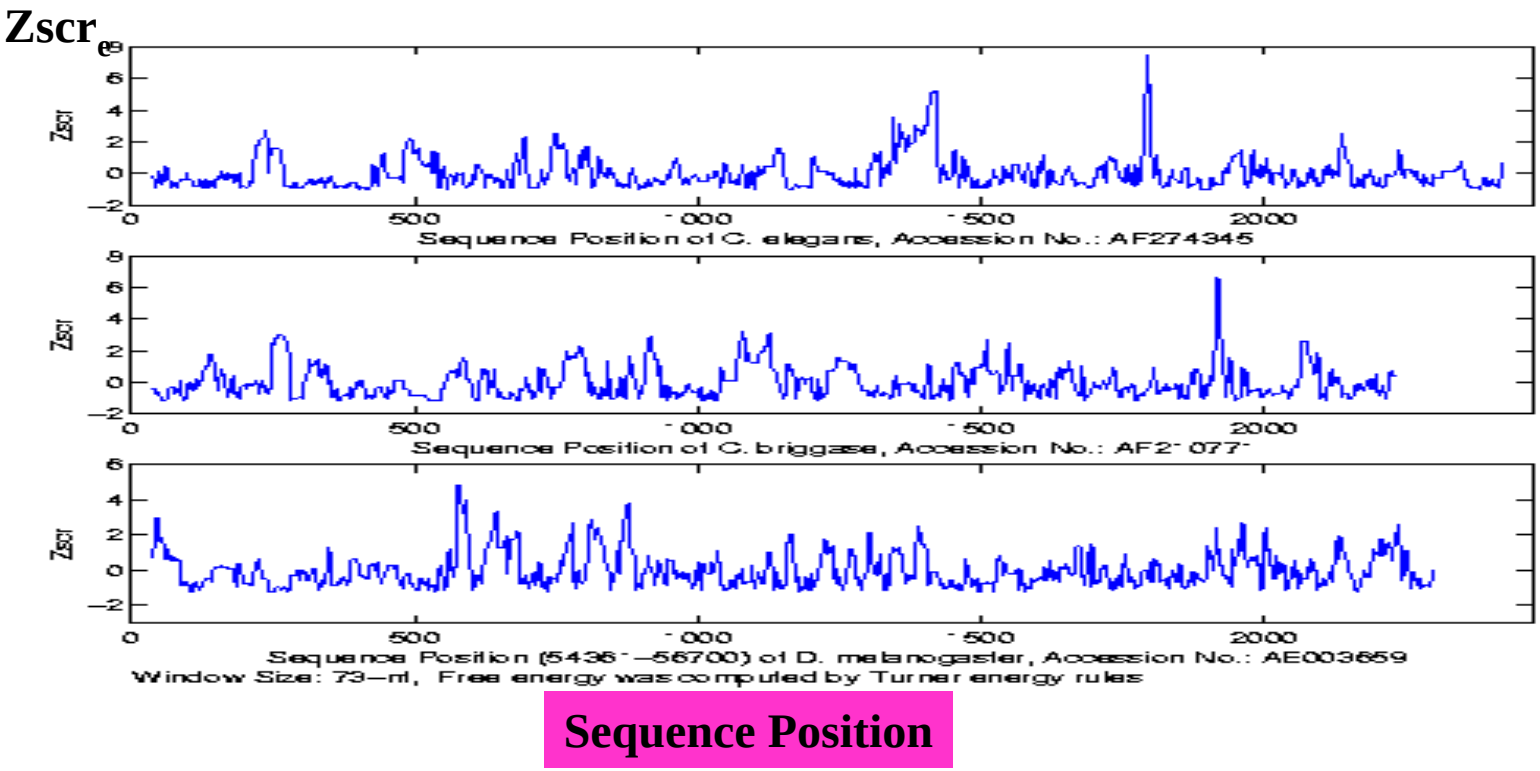
**Window Size**

**Sequence Position**

**The $Zscr_e$ values were computed by moving a fixed window stepped 3-nt each time from 5' to 3'.**

**The window size was systematically changed from 60 to 95-nt by a step of 5-nt and the corresponding curve was systematically plotted by red, blue, yellow, magenta, green, black and cyanic, respectively.**

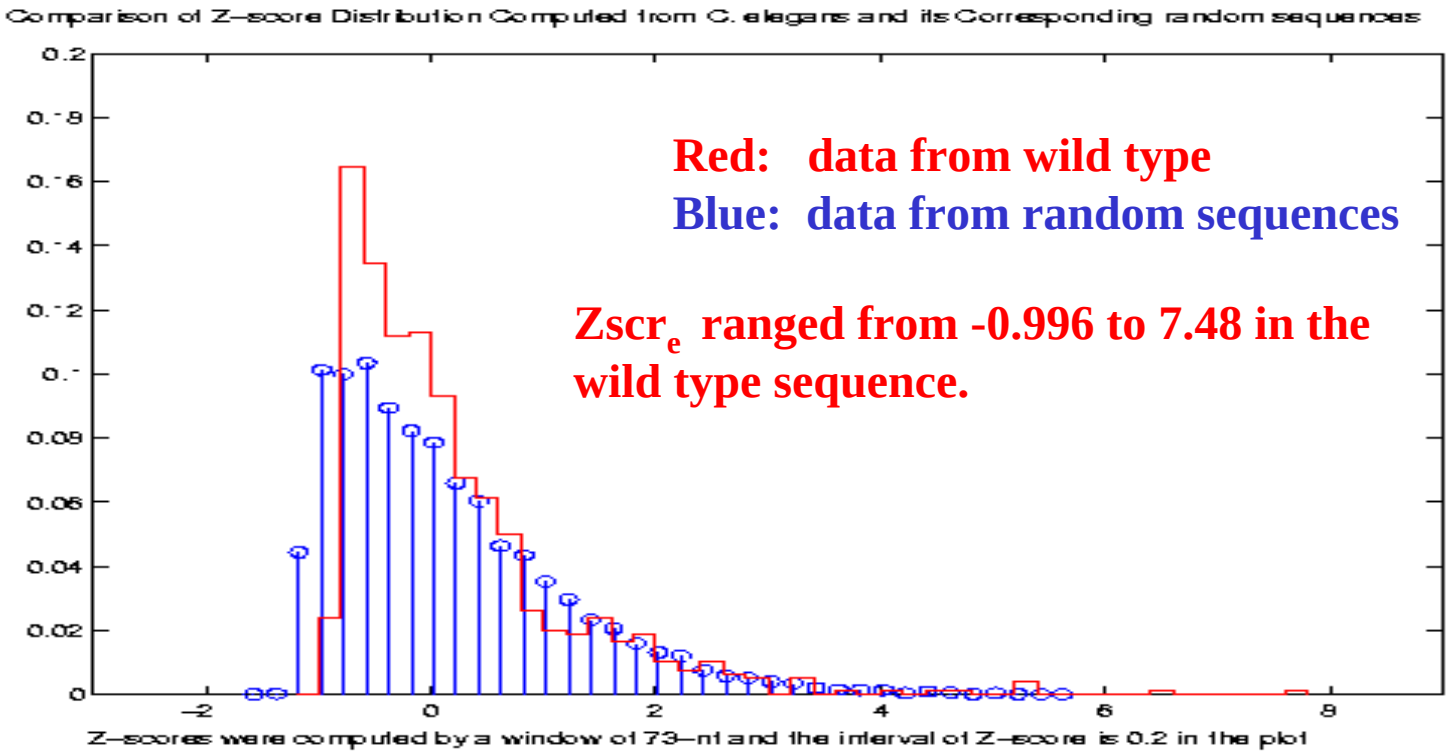
# $Zscr_e$ Plots of the three genomic sequences of *C.elegans* (top), *C. briggase* (middle) and *D. melanogaster* (bottom)




**The plot was produced by plotting $Zscr_e$ against the position of the middle base in the window of 73-nt. In the plot of *D. melanogaster* , the sequence position 54,361 is numbered as position 1.**

# Comparisons of $Zscr_e$ distributions computed from *let-7* RNA gene sequence of *C. elegans* and its corresponding 30 randomly shuffled sequences

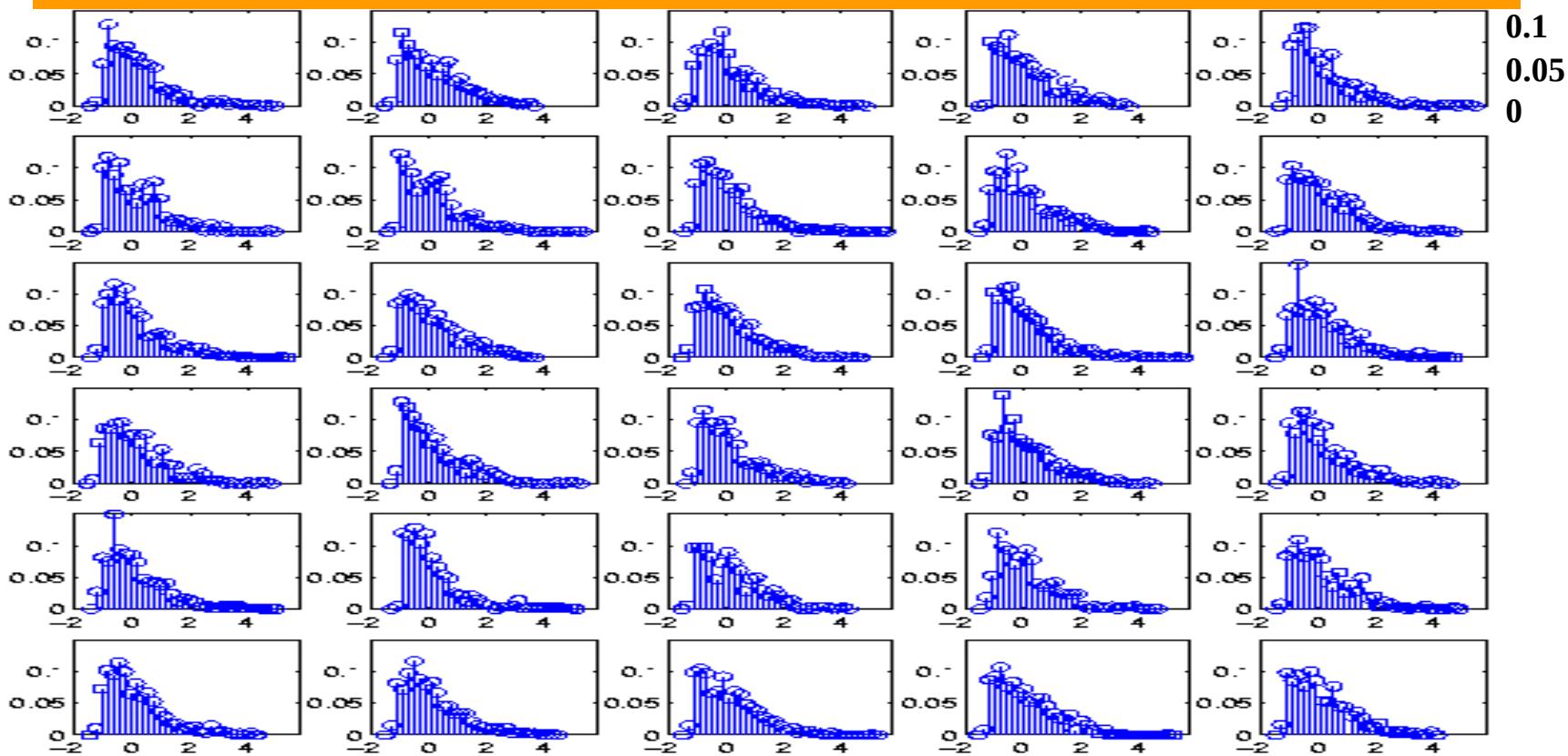
**Fraction**




**$Zscr_e$ of Free Energy Difference**

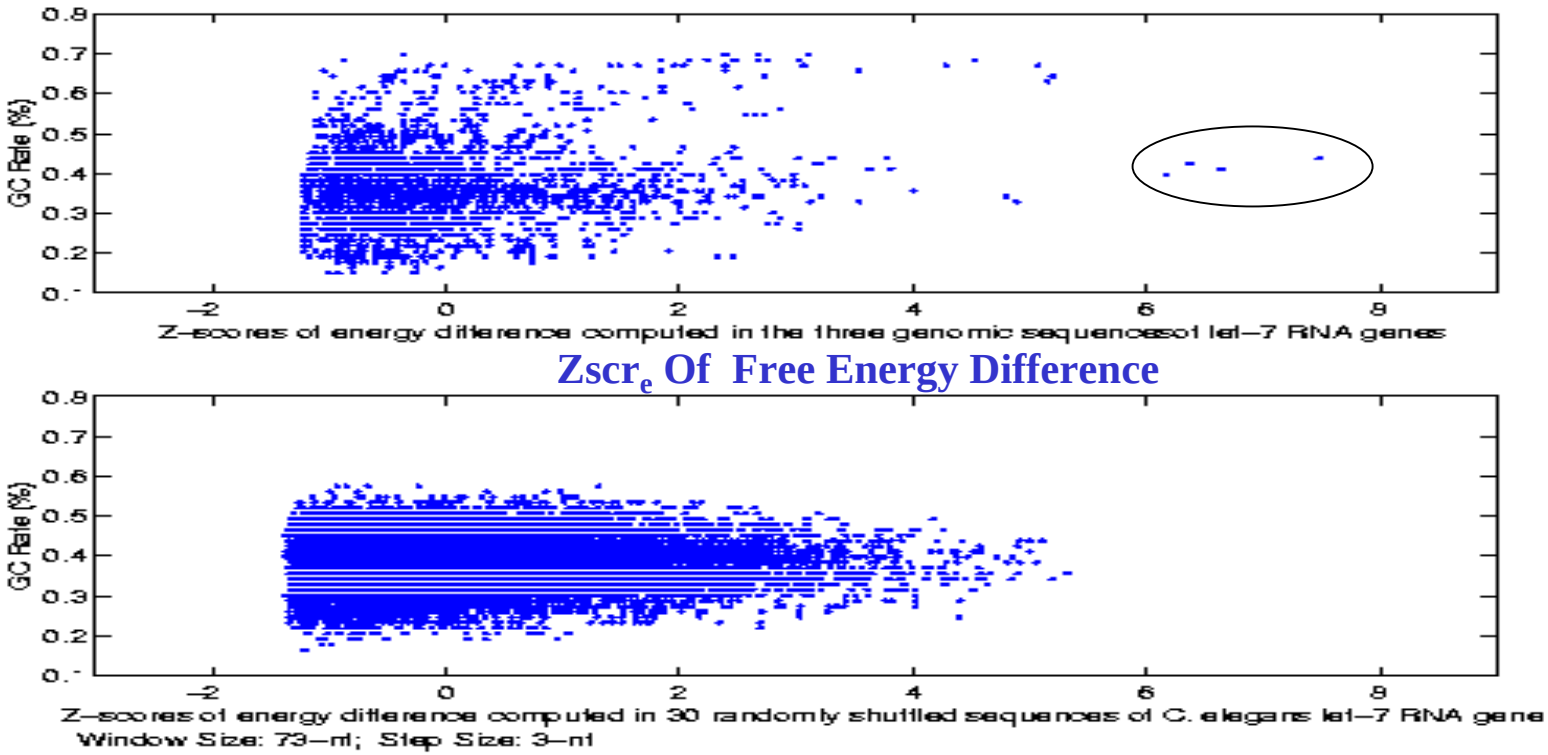
**All data were computed by moving the fixed window of 73-nt stepped 3-nt from 5' to 3' along the sequence. None of 23,880 $Zscr_e$ scores is greater than 5.32 in 30 randomly shuffled sequences.**

# Distributions of $Zscr_e$ of the free energy difference computed from 30 randomly shuffled sequences of *C.elegans let-7* RNA gene




**$Zscr_e$ scores are ranged from -1.37 to 5.32 in the 30 randomly shuffled sequences. Only 10 out of 23,880 observations of scores are greater than 4.89.**

**Relationships between GC composition and $Zscr_e$ of the local segment of 73-nt computed in the wild type sequence of *C.elegans let-7 RNA gene* (top) and 30 randomly shuffled sequences (bottom) of the wild type sequence**



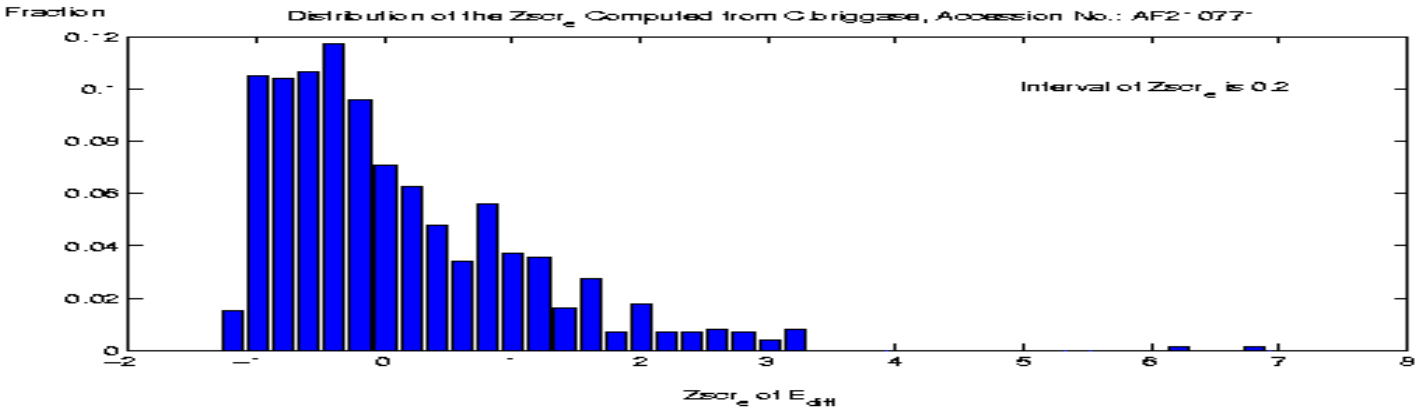

**$Zscr_e$ Of Free Energy Difference**

**$Zscr_e$ values are ranged from -1.00 to 7.48 and GC compositions are from 0.151 to 0.699 in the wild type sequence**

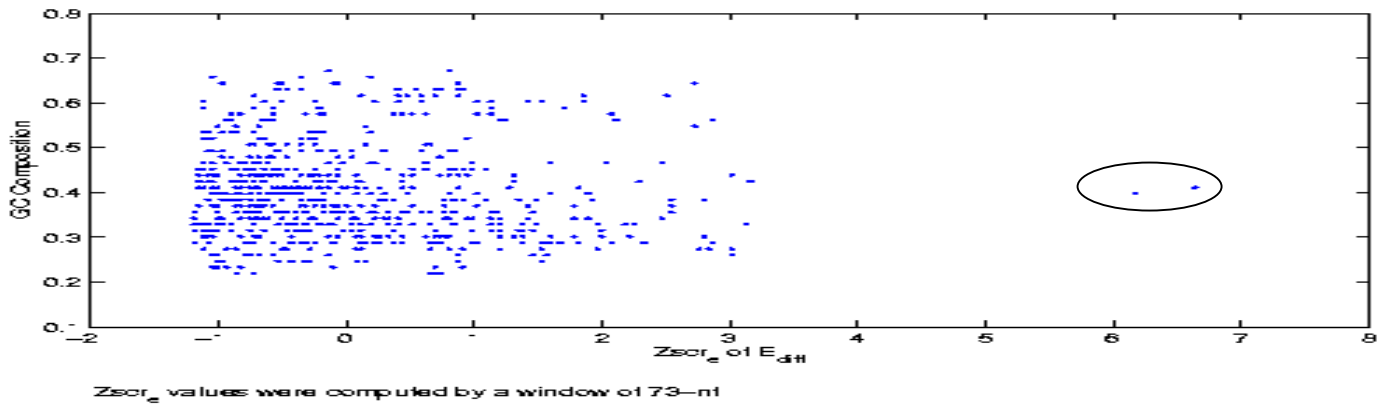
**$Zscr_e$ values are ranged from -1.37 to 4.89 and GC compositions are from 0.164 to 0.575 in 30 random sequences.**

# Distributions of $Zscr_e$ of the free energy difference computed from *C.briggase let-7* RNA gene (Accession No. AF210771)

## *Fraction*



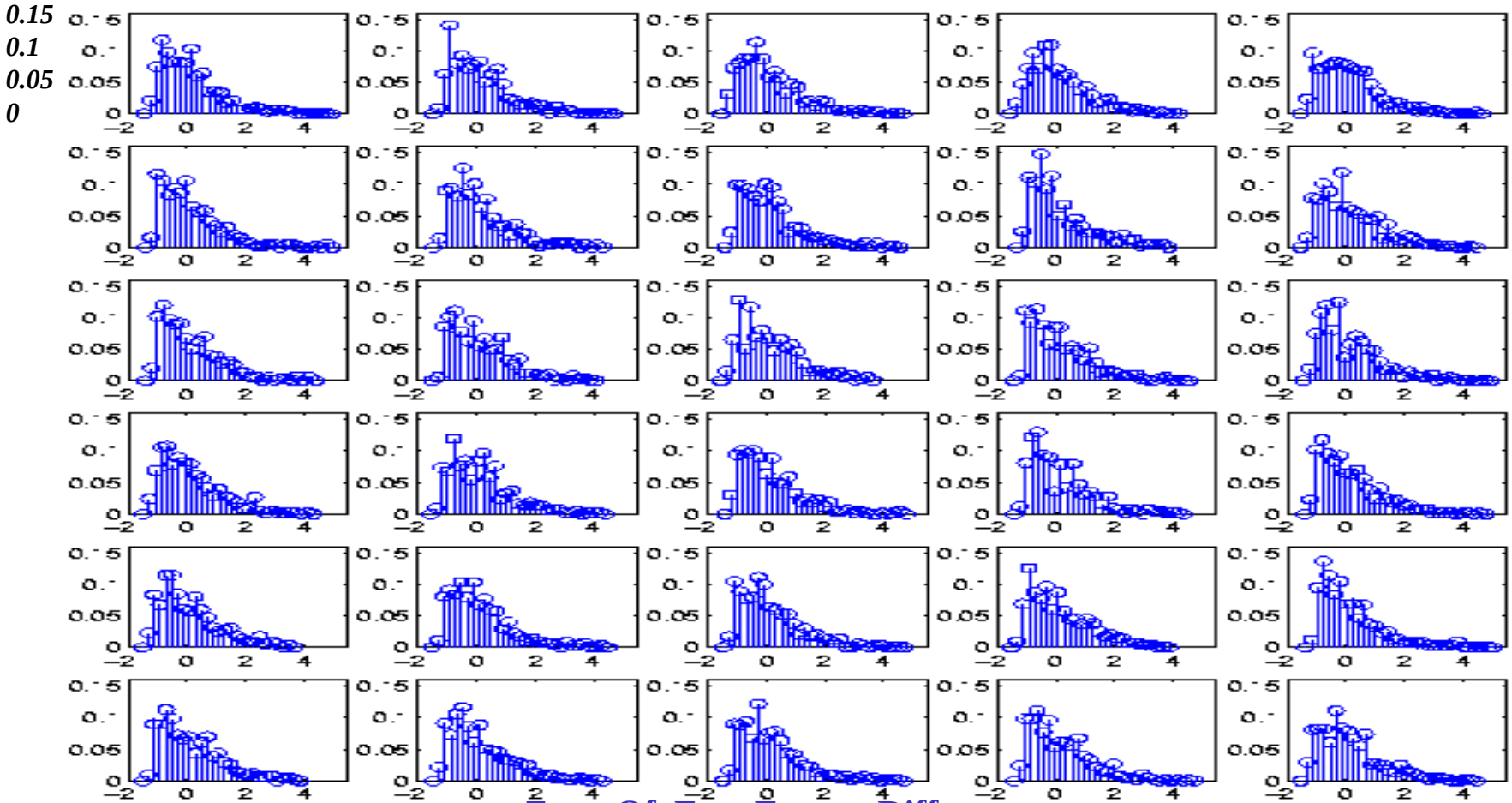

## *GC Composition*




**$Zscr_e$ values are ranged from -1.19 to 6.64 and GC compositions are from 0.219 to 0 .671 in the wild type sequence**

**$Zscr_e$ values are computed by moving the fixed window of 73-nt stepped successively in 3 nt from 5' to 3'.**

# Distributions of $Zscr_e$ of the free energy difference computed from 30 randomly shuffled sequences of *D.melanogaster* *let-7* RNA gene



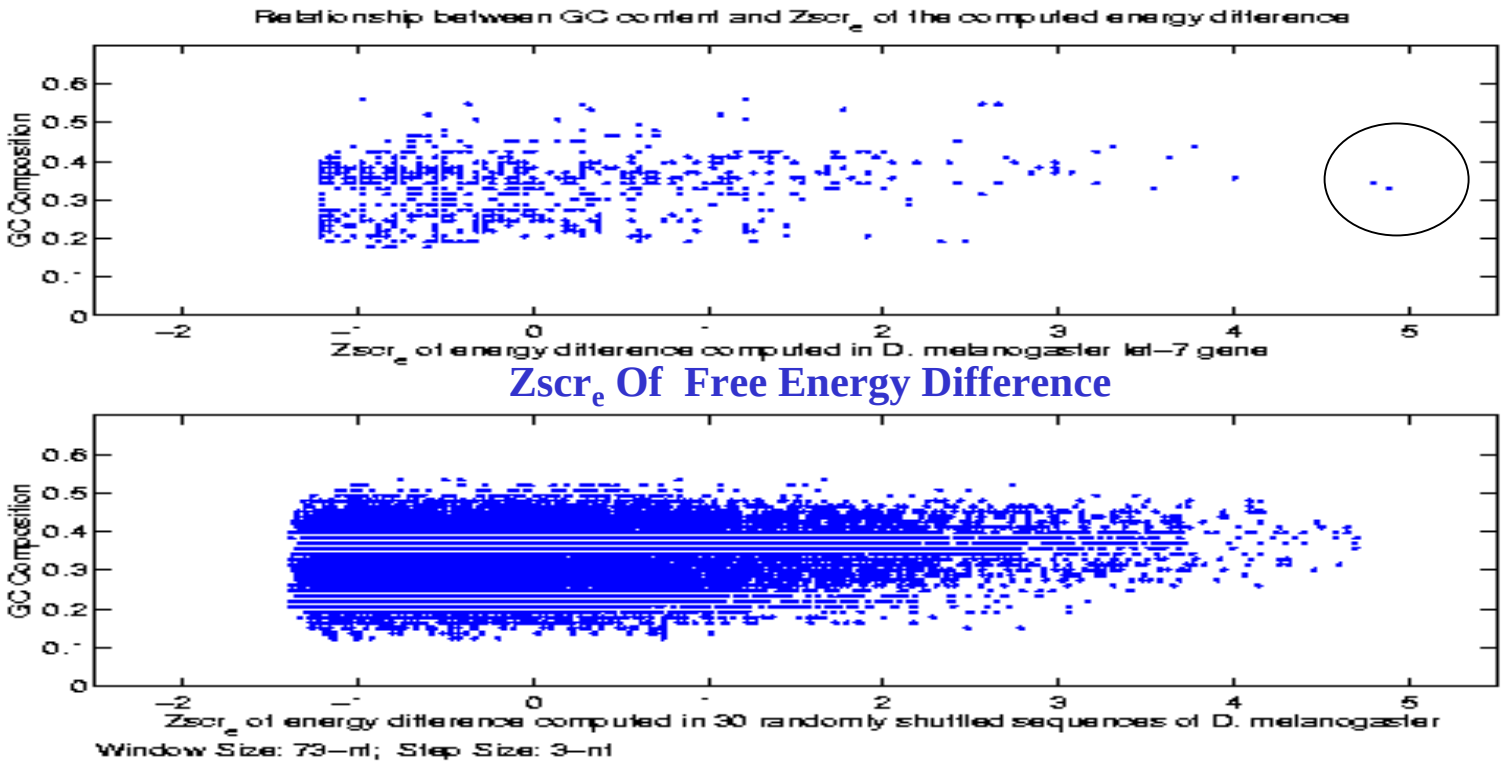

$Zscr_e$ Of Free Energy Difference

The random sequences are made from D. melanogaster let−7 RNA gene, Acce. No.AE003659

$Zscr_e$ were computed by a window of 73−nt and the interval of Z−score is 0.2 in the plot

**All data were computed by moving the fixed window of 73-nt stepped 3-nt from 5'to 3' along the sequence. None of 22,680 $Zscr_e$ scores is greater than 4.70 in 30 randomly shuffled sequences.**

**Relationships between GC composition and $Zscr_e$ of the local segment of 73-nt computed in the wild type sequence of *D.melanogaster let-7 RNA gene* (top) and 30 randomly shuffled sequences (bottom) of the wild type sequence**

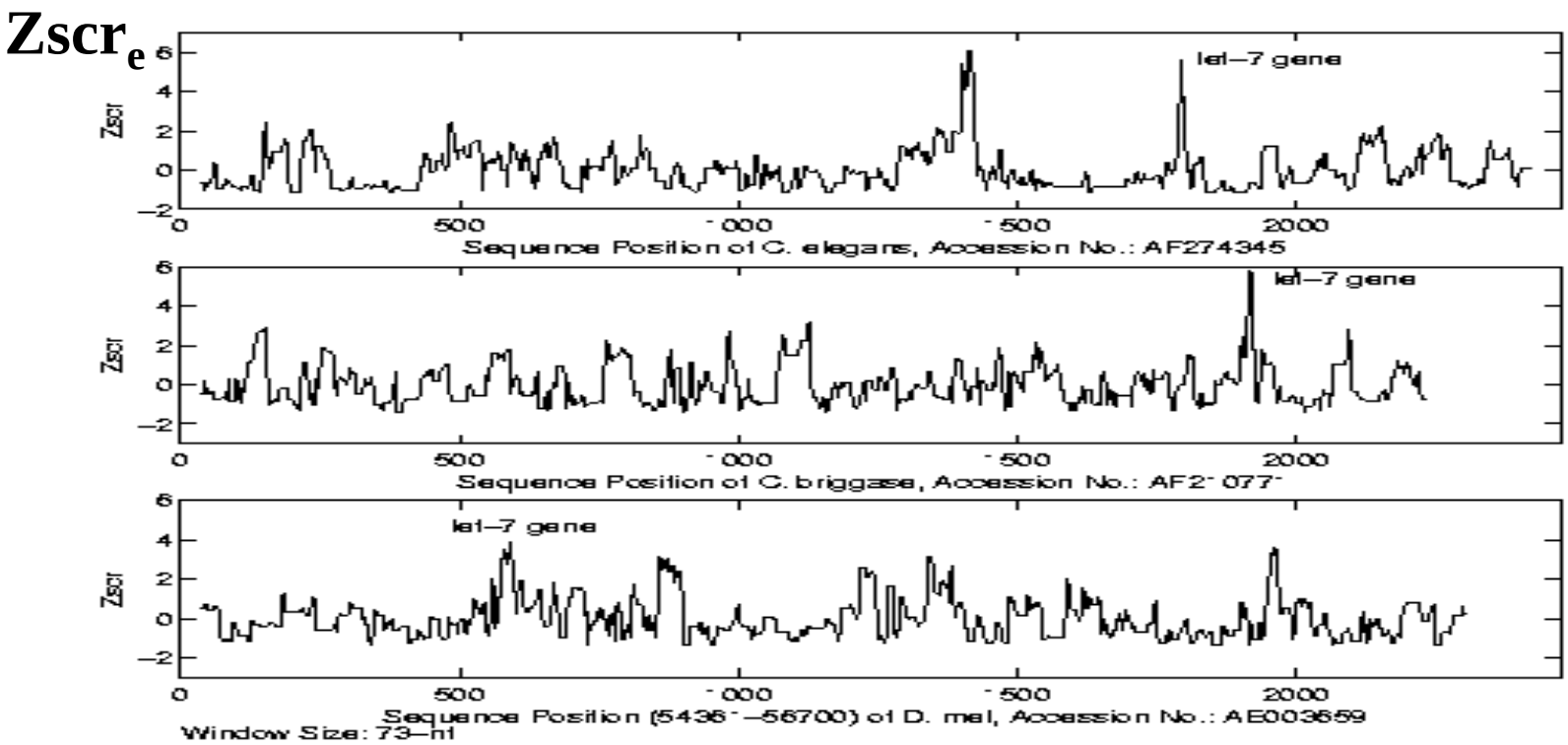



**$Zscr_e$ Of Free Energy Difference**

**$Zscr_e$ values are ranged from -1.21 to 4.89 and GC compositions are from 0.178 to 0.562 in the wild type sequence**

**$Zscr_e$ values are ranged from -1.38 to 4.70 and GC compositions are from 0.123 to 0.534 in 30 random sequences.**

# $Zscr_e$ Plots of the three genomic sequences of *C.elegans* (top), *C. briggase* (middle) and *D. melanogaster* (bottom)




**Sequence Position**

**The plot was produced by plotting $Zscr_e$ against the position of the middle base in the window of 73-nt. In the plot of *D. melanogaster* , the sequence position 54,361 is numbered as position 1. The free energies of the folded segments were computed by energy parameters derived from nearest-neighbor thermodynamics of single-stranded DNA sequences.**

# Conclusion

- **$Zscr_e$ of the folded structure is a good measure to define quantitatively both the stability and uniqueness of functional elements in sequences.**

- **The statistically significant well-determined folding patterns can be apparently discriminated from the bulk distribution of folded segments. The distinct well-ordered conformations in our examples can not be expected to be found in large number of randomly shuffled sequences.**

- **$Zscr_e$ of the energy difference in a sequence does not follow a normal distribution, however it can be well described by a linearly transformed non-central *t* distribution (LTNSTD).**

- **Statistical extremes of the well-determined folding patterns can be estimated by the derived LTNSTD based on the sample mean, sample standard deviation, and coefficient of skewness of $Zscr_e$ data computed in a nucleic acid sequence.**

# Evaluation of Well-ordered RNA structures that are both thermodynamically stable and distinctly folded by computational method (st_comp)

- **Hypotheses**

- **Evolution of functional RNA elements (molecules) from random sequences is constrained by the intrinsic structural property.**

- **Increase the thermodynamically stable of the folded structure by adding more base-pairs in stem helical regions and reduce irregularities in stem regions.**

- **Reduce the possibility of alternative stable structures folded in the sequence in a dynamic RNA folding.**

- **The unique conformation can not be expected by chance.**

# Quantitative Measure for the Uniqueness of RNA Secondary Structure

- **Maximal Matching Score (MMS): represents the maximal structure similarity between two RNA secondary structures**

- **MMSNVR: The measure is the sample mean of all MMS observations computed from the structural comparison between the structure of a wild type RNA and predicted optimized structures from a set of randomly shuffled sequences (e.g. 300 random sequences) of the wild type RNA.**
- **$MMSNVR_E$ : if the structure of a wild type RNA is predicted by the lowest free energy structure (mfold).**
- **MMSNVRp : if the structure of a wild type RNA is predicted by the phylogenetic comparison method.**

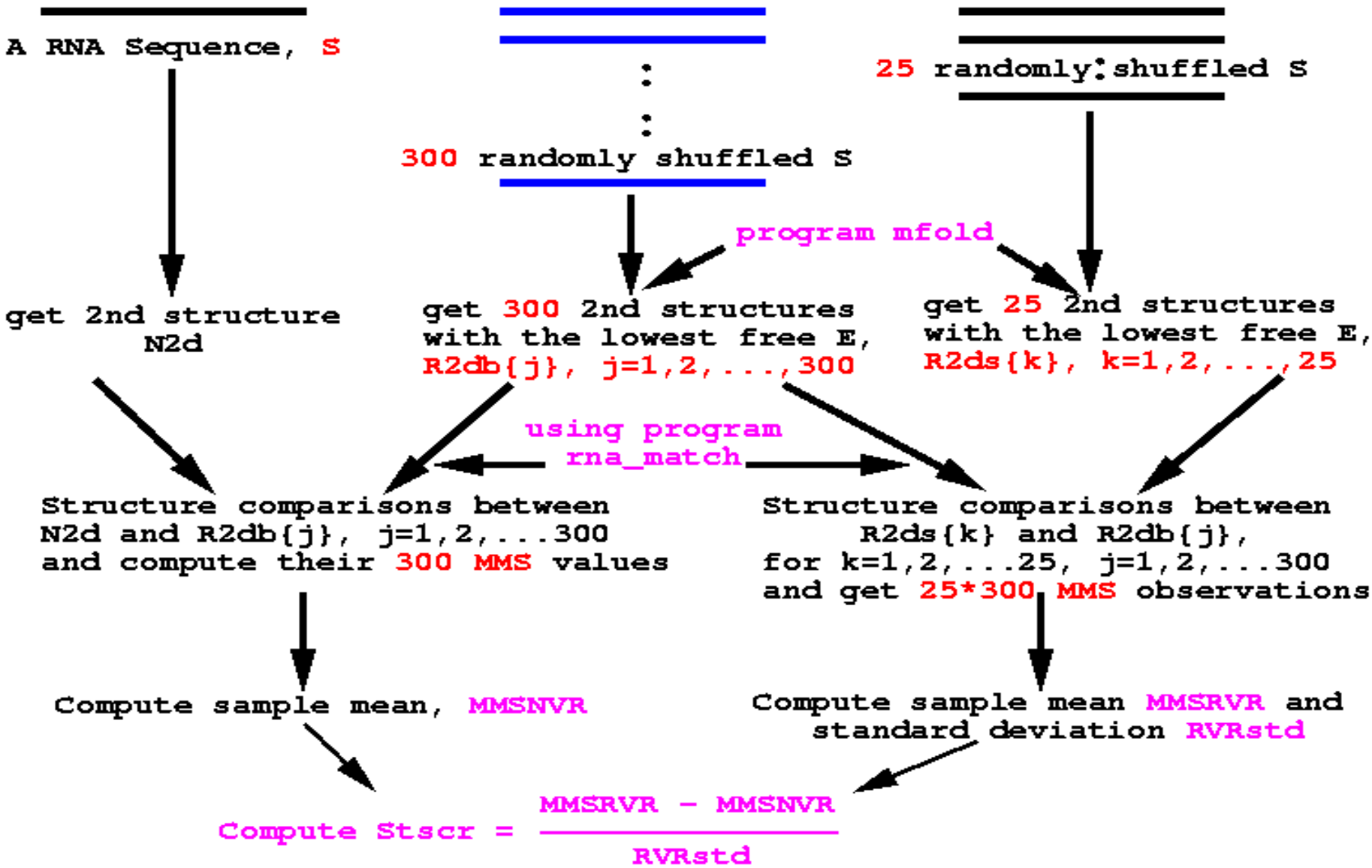
- **MMSRVR: The measure is the sample mean of all MMS observations computed from the structural comparison between the structures of 25 other randomly shuffled sequences and the randomly shuffled sequences (e.g. 300 random sequences) used in the structural comparison between the wild type and random structures.**
- **RVRstd: is the sample standard deviation of the maximal matching scores between the other 25 random structures and the 300 random structures.**

# A standard score, Stscr for evaluating the uniqueness of RNA secondary structure

- **Definition of Stscr: It is asignificance score of structural uniqueness.**
- **Stscr = (MMSRVR - MMSNVR) / RVRstd**

- **The standard score, Stscr is used to estimate how different a real structure (either predicted or phylogenetically determined) from a large number of random structures. The larger the score Stscr of the real structure the more well-ordered the real structure. The uniqueness of the well-ordered conformations in the real structure implies that its distinct structure feature can not be derived from a random folding. The folded structure of RNA functional elements is expected to be significantly more ordered than that of random RNA sequences results from constraints imposed by intrinsic structural properties.**

# Method of St_comp



# The procedure of estimating the uniqueness of an RNA secondary structure

- **1. Generating a set of 300 randomly shuffled sequences for the natural RNA sequence and predicting their optimized structures with the lowest free energy.**

- **2. Performing structure comparisons between the structure from an RNA sequence and each of 300 random structures derived by step 1. The structure comparison is made by using program rna_match. The average of maximal matching scores, MMSNVR from the comparisons is computed.**

- **3. Generating a set of other 25 randomly shuffled sequences for the natural RNA sequence and predicting their optimized structures with the lowest free energy.**

- **4. Computing the average of maximal matching scores, MMSRVR and the sample standard deviation, RVRstd from the structure comparisons of the 25 random sequences and the 300 random sequences used in step 1.**

- **5 Computing the significance score of structural uniqueness, Stscr**

# Table 1. RNA secondary structural comparisons among100 tRNA and corresponding randomly shuffled sequences

| tRNA | $Stscr_E$ | $MMSNVR_E$ | $Stscr_P$ | $MMSNVR_P$ | MMSRVR | RVRstd |
|---|---|---|---|---|---|---|
| DA2480 | 2.93 | 45.54 | 2.93 | 45.54 | 78.34 | 11.20 |
| DA5280 | 1.16 | 48.63 | 2.12 | 29.13 | 72.12 | 20.28 |
| DA7680 | 3.16 | 34.39 | 3.16 | 34.39 | 76.51 | 13.28 |
| DC0380 | 2.40 | 36.90 | 2.40 | 36.90 | 79.37 | 17.73 |
| DC5020 | −0.79 | 74.77 | 3.60 | 21.45 | 65.14 | 12.15 |
| DD2680 | −0.53 | 82.66 | 2.48 | 29.36 | 73.27 | 17.72 |
| DD2920 | −0.73 | 89.99 | 2.11 | 28.93 | 74.35 | 21.55 |
| DD4000 | 1.99 | 49.97 | 3.49 | 29.62 | 77.14 | 13.63 |
| DD4080 | 0.55 | 62.33 | 2.19 | 28.05 | 73.73 | 20.90 |
| DD5280 | 2.05 | 41.99 | 2.05 | 41.99 | 66.98 | 12.19 |
| DD5320 | 2.02 | 41.21 | 2.71 | 32.61 | 66.43 | 12.46 |
| DE1230 | 0.56 | 57.67 | 1.82 | 38.99 | 65.99 | 14.84 |
| DE1660 | −0.68 | 84.87 | 2.88 | 31.90 | 74.78 | 14.87 |
| DE6160 | 2.62 | 19.01 | 2.62 | 19.01 | 64.43 | 17.31 |
| DF1180 | −1.06 | 93.18 | 2.65 | 32.62 | 75.85 | 16.31 |
| DF5220 | 2.83 | 36.39 | 2.73 | 37.10 | 56.61 | 7.15 |
| DF5900 | −0.27 | 56.81 | 2.14 | 29.19 | 53.76 | 11.50 |
| DF5930 | 2.46 | 31.91 | 2.46 | 31.91 | 49.60 | 7.18 |
| DF9160 | 1.49 | 44.80 | 2.55 | 28.39 | 67.91 | 15.48 |
| DG2440 | 0.58 | 57.35 | 2.19 | 27.84 | 68.01 | 18.36 |
| DG2921 | 2.31 | 47.67 | 2.61 | 44.61 | 70.84 | 10.04 |
| DG4070 | 0.29 | 54.35 | 2.09 | 34.16 | 57.64 | 11.25 |
| DG5040 | 1.61 | 41.94 | 2.20 | 35.47 | 59.67 | 11.01 |
| DG7740 | −0.53 | 94.42 | 3.19 | 27.12 | 84.85 | 18.09 |
| DG8100 | 0.37 | 66.84 | 1.98 | 45.08 | 71.88 | 13.53 |
| DH1700 | 1.97 | 36.97 | 1.97 | 36.97 | 75.11 | 19.39 |
| DH2880 | −0.10 | 74.57 | 2.01 | 46.81 | 73.19 | 13.11 |
| DH2960 | −0.13 | 76.77 | 3.00 | 41.13 | 75.34 | 11.41 |
| DH4360 | 1.07 | 55.19 | 2.24 | 34.52 | 74.12 | 17.68 |
| DH5120 | 0.09 | 49.09 | 2.55 | −12.75 | 51.43 | 25.17 |
| DI2220 | 2.79 | 20.09 | 2.79 | 20.09 | 70.09 | 17.89 |
| DI2701 | 2.12 | 40.13 | 2.12 | 40.13 | 75.39 | 16.63 |
| DI2400 | 4.42 | 24.17 | 4.42 | 24.17 | 78.99 | 12.41 |
| DK1230 | 0.01 | 70.44 | 2.56 | 28.69 | 70.57 | 16.29 |
| DK4340 | −0.87 | 86.20 | 2.69 | 38.57 | 74.56 | 13.37 |
| DK5320 | −0.86 | 63.11 | 0.78 | 45.56 | 53.91 | 10.73 |
| DK6280 | 2.02 | 7.78 | 2.02 | 7.78 | 64.70 | 28.22 |
| DL0980 | −0.85 | 102.11 | 2.62 | 27.52 | 83.88 | 21.47 |
| DL1141 | −1.17 | 90.75 | 5.06 | 1.99 | 74.07 | 14.26 |
| DL1200 | 3.36 | 9.31 | 3.36 | 9.31 | 75.68 | 19.76 |
| DL1231 | −1.15 | 94.87 | 3.09 | 13.43 | 72.83 | 19.24 |
| DL1543 | 0.49 | 72.80 | 4.49 | 10.71 | 80.48 | 15.55 |
| DL1662 | 0.99 | 54.73 | 3.49 | 17.14 | 67.93 | 13.40 |
| DL1750 | −0.21 | 84.23 | 4.99 | 7.79 | 81.14 | 14.70 |
| DL2522 | 0.45 | 67.17 | 4.36 | −11.87 | 76.20 | 20.18 |
| DL3200 | −0.81 | 87.07 | 2.75 | 27.07 | 73.43 | 16.83 |
| DL4070 | −0.12 | 75.41 | 3.88 | 9.78 | 73.39 | 16.37 |
| DL4700 | 1.48 | 42.86 | 2.03 | 35.87 | 61.41 | 12.56 |
| DL4760 | 2.15 | 46.11 | 2.15 | 46.11 | 77.91 | 14.79 |

# Table 1. RNA secondary structural comparisons among100 tRNA and corresponding randomly shuffled sequences

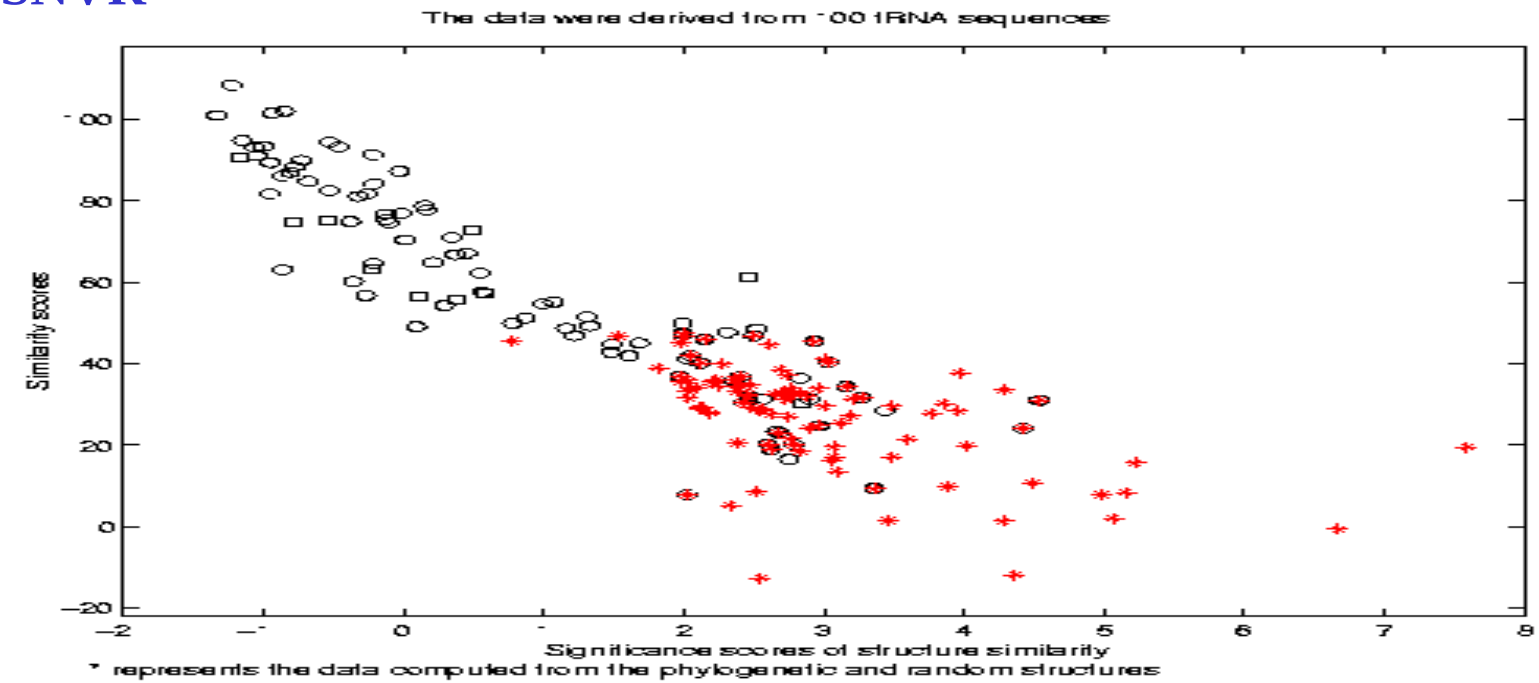
| tRNA | $Stscr_E$ | $MMSNVR_E$ | $Stscr_P$ | $MMSNVR_P$ | MMSRVR | RVRstd |
|---|---|---|---|---|---|---|
| DL5880 | −0.36 | 60.18 | 2.76 | 33.89 | 57.13 | 8.43 |
| DL7740 | −1.23 | 108.47 | 3.07 | 19.67 | 83.01 | 20.62 |
| DL7920 | 1.33 | 49.31 | 4.28 | 1.31 | 70.97 | 16.26 |
| DM4000 | −0.96 | 81.83 | 2.39 | 33.29 | 67.92 | 14.52 |
| DN4620 | 2.46 | 61.31 | 7.59 | 19.45 | 81.42 | 8.16 |
| DP0680 | 2.90 | 31.38 | 3.00 | 29.53 | 83.15 | 17.85 |
| DP1360 | 3.03 | 40.33 | 3.03 | 40.33 | 85.85 | 15.03 |
| DQ1340 | −0.78 | 88.39 | 2.05 | 34.14 | 73.45 | 19.16 |
| DQ3220 | −0.03 | 87.34 | 3.97 | 37.81 | 86.92 | 12.36 |
| DQ4880 | 0.15 | 78.86 | 3.05 | 16.20 | 82.14 | 21.63 |
| DQ5080 | 0.78 | 50.00 | 2.03 | 31.56 | 61.66 | 14.86 |
| DQ6160 | 2.77 | 32.13 | 2.77 | 32.13 | 74.34 | 15.25 |
| DR1660 | −0.46 | 93.23 | 3.85 | 30.14 | 86.52 | 14.65 |
| DR1663 | −0.95 | 89.47 | 2.47 | 34.79 | 74.29 | 15.97 |
| DR3320 | 0.21 | 64.89 | 2.38 | 20.55 | 69.20 | 20.44 |
| DR5080 | 0.38 | 55.69 | 2.03 | 33.09 | 60.83 | 13.44 |
| DR6051 | −0.26 | 81.69 | 3.20 | 31.31 | 77.91 | 14.54 |
| DS0261 | 0.34 | 71.02 | 5.22 | 15.59 | 74.89 | 11.37 |
| DS1230 | 2.75 | 16.52 | 3.46 | 1.47 | 74.72 | 21.14 |
| DS2480 | 2.84 | 30.34 | 3.12 | 25.43 | 80.30 | 17.56 |
| DS2520 | 1.22 | 46.93 | 3.07 | 17.12 | 66.59 | 16.11 |
| DS2922 | −1.00 | 93.46 | 3.76 | 27.80 | 79.68 | 13.78 |
| DS6745 | −0.22 | 91.43 | 6.67 | −0.70 | 88.51 | 13.38 |
| DT0661 | 2.50 | 46.88 | 2.50 | 46.88 | 81.20 | 13.75 |
| DT1542 | 2.60 | 20.21 | 2.60 | 20.21 | 78.35 | 22.33 |
| DT2600 | 2.43 | 30.45 | 2.43 | 30.45 | 63.63 | 13.66 |
| DT3880 | 1.99 | 47.43 | 1.99 | 47.43 | 68.07 | 10.36 |
| DT4700 | −0.22 | 64.68 | 2.23 | 36.23 | 62.15 | 11.63 |
| DT4880 | 0.16 | 77.81 | 4.28 | 33.62 | 79.57 | 10.74 |
| DT6160 | 1.68 | 45.08 | 1.54 | 46.78 | 66.47 | 12.75 |
| DT7740 | 1.31 | 51.53 | 2.90 | 24.27 | 73.93 | 17.14 |
| DT9991 | 2.68 | 22.98 | 2.68 | 22.98 | 71.46 | 18.06 |
| DV2600 | 3.27 | 31.72 | 3.27 | 31.72 | 66.29 | 10.58 |
| DV3200 | 4.54 | 31.02 | 4.54 | 31.02 | 78.20 | 10.40 |
| DV3960 | −0.23 | 63.28 | 2.83 | 18.51 | 59.86 | 14.61 |
| DV4000 | 2.36 | 35.92 | 2.36 | 35.92 | 68.66 | 13.90 |
| DW4360 | 3.43 | 28.49 | 2.96 | 33.93 | 68.27 | 11.60 |
| DW5080 | 0.11 | 56.55 | 2.72 | 31.09 | 57.64 | 9.77 |
| DX0260 | −1.05 | 91.21 | 4.02 | 19.88 | 76.49 | 14.08 |
| DX4320 | 2.96 | 24.73 | 2.96 | 24.73 | 71.59 | 15.84 |
| DX4440 | −0.54 | 75.21 | 2.83 | 32.86 | 68.43 | 12.55 |
| DX4960 | 2.57 | 31.43 | 2.46 | 32.57 | 60.06 | 11.16 |
| DX5160 | 0.87 | 51.31 | 2.28 | 39.93 | 58.27 | 8.03 |
| DY0660 | 2.74 | 32.90 | 2.74 | 32.90 | 77.85 | 16.41 |
| DY2920 | −0.94 | 101.54 | 2.33 | 5.11 | 73.90 | 29.54 |
| DY3280 | −1.33 | 101.07 | 5.15 | 8.35 | 82.03 | 14.32 |
| DY3770 | −0.01 | 77.07 | 3.96 | 28.36 | 76.94 | 12.26 |
| DY4880 | −0.38 | 74.93 | 1.98 | 35.84 | 68.61 | 16.53 |
| DY5040 | 2.52 | 48.54 | 2.52 | 8.54 | 66.73 | 7.22 |
| DY5220 | −0.32 | 81.15 | 2.38 | 34.55 | 75.65 | 17.25 |
| DY6743 | 2.66 | 23.39 | 2.76 | 21.54 | 76.21 | 19.84 |

# Table 1. RNA secondary structural comparisons among 14 ribonuclease P RNAs and corresponding randomly shuffled sequences

| RNAse P RNA | $Stscr_E$ | $MMSNVR_E$ | $Stscr_P$ | $MMSNVR_P$ | MMSRVR | RVRstd |
|---|---|---|---|---|---|---|
| A. nid (Anacystis nidulans) | 1.59 | 147.30 | 5.35 | 56.17 | 185.71 | 24.20 |
| A. tum (Agobacterium tumefaciens) | 1.24 | 139.95 | 2.65 | 90.63 | 183.59 | 35.12 |
| B. bre (Bacillus brevis) | 0.28 | 174.17 | 2.45 | 50.37 | 190.1 | 57.07 |
| B. sub (Bacillus subtilis) | 5.09 | 24.81 | 5.43 | 14.67 | 175.61 | 29.64 |
| B. bur (Borrelia burgdorferi) | −0.08 | 159.27 | 2.83 | 49.23 | 156.10 | 37.75 |
| B. the (Bacteroides thetaiotaomicron) | 1.48 | 126.50 | 3.55 | 55.26 | 177.51 | 34.42 |
| C. lim (Chlorobium limicola) | 2.43 | 82.81 | 3.12 | 60.08 | 163.61 | 33.21 |
| C. par (Cyanophora paradoxa cyanelle) | 1.07 | 138.98 | 3.80 | 50.06 | 174.02 | 32.63 |
| D. des (Desulfovibrio desulfuricans) | −0.26 | 184.76 | 5.43 | 25.52 | 177.57 | 27.99 |
| E. coli (Escherichia coli) | −0.42 | 202.93 | 2.15 | 118.67 | 189.15 | 32.70 |
| P. pur (Porphyra purpurea chloroplast) | 1.42 | 133.10 | 3.29 | 71.39 | 180.06 | 33.01 |
| S. bik (Streptomyces bikiniensis) | 0.65 | 172.59 | 2.85 | 69.70 | 202.89 | 46.80 |
| T. aqu (Thermus aquaticus) | 4.23 | 95.71 | 3.46 | 116.28 | 208.41 | 26.64 |
| T. nea (Thermotoga neapolitana) | 1.21 | 173.12 | 5.84 | 41.55 | 207.39 | 28.42 |

# Significance Score of Structure Uniqueness Computed from 100 tRNA sequences

**MMSNVR**



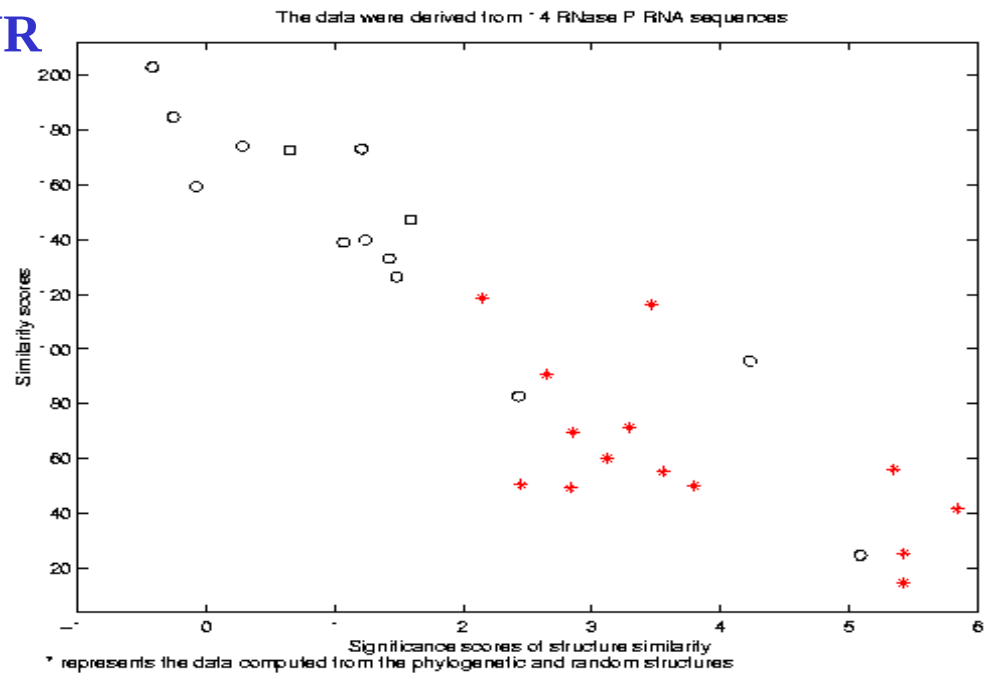

## Significance Score of Structure Uniqueness

**Red: represents the score data computed from phylogenetic structures of tRNA molecules.**

**Black: represents the score data computed from the predicted optimized structures by mfold.**

# Significance Scores of Structure Uniqueness Computed from 14 Ribonuclease P RNAs

**MMSNVR**




## Significance Score of Structure Uniqueness

**Red: represents the score data computed from phylogenetic structures of Rnase P molecules.**

**Black: represents the score data computed from the predicted optimized structures by mfold.**