

Computational Discovery of Distinct RNA Elements with Functional Structures in Genomic Sequences

Shu-Yun. Le

Laboratory of Experimental and
Computational Biology, CCR, NCI, NIH

RNAs are conformational molecules.

RNAs perform a wide range of functions in biological systems

Regulated controls of replication and transcription of virus.

Post-transcriptional regulation of gene expression

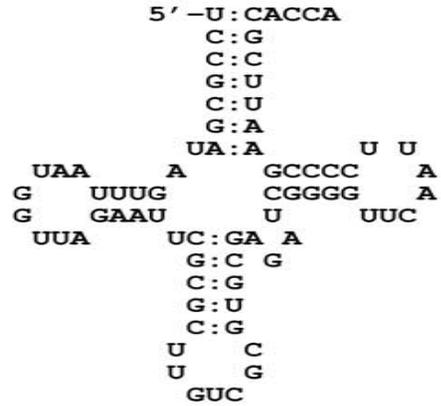
mRNA processing, localization and metabolism

Regulated controls of mRNA translation

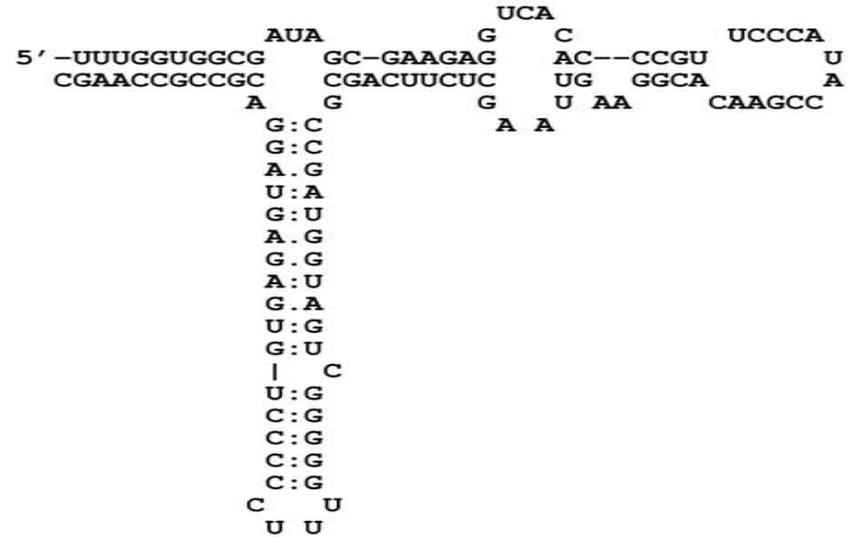
Interactions between RNA/RNA and RNA/Protein play a crucial role in the regulatory mechanism.

Micro-RNAs having the potential to form fold-back stem-loops are involved in RNA-based gene regulation.

Yeast Aspartic acid tRNA



5S rRNA of Bacillus subtilis

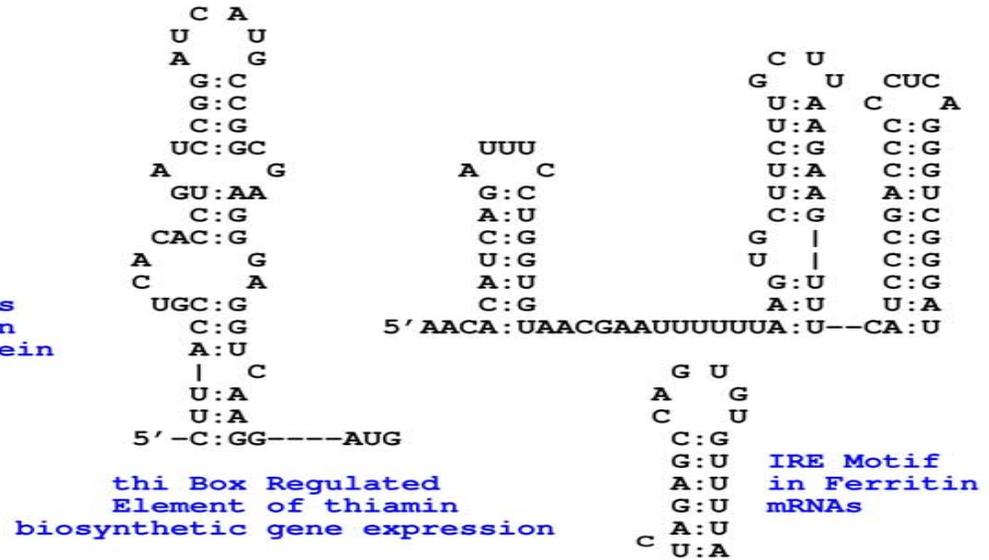


HIV-1 TAR Element

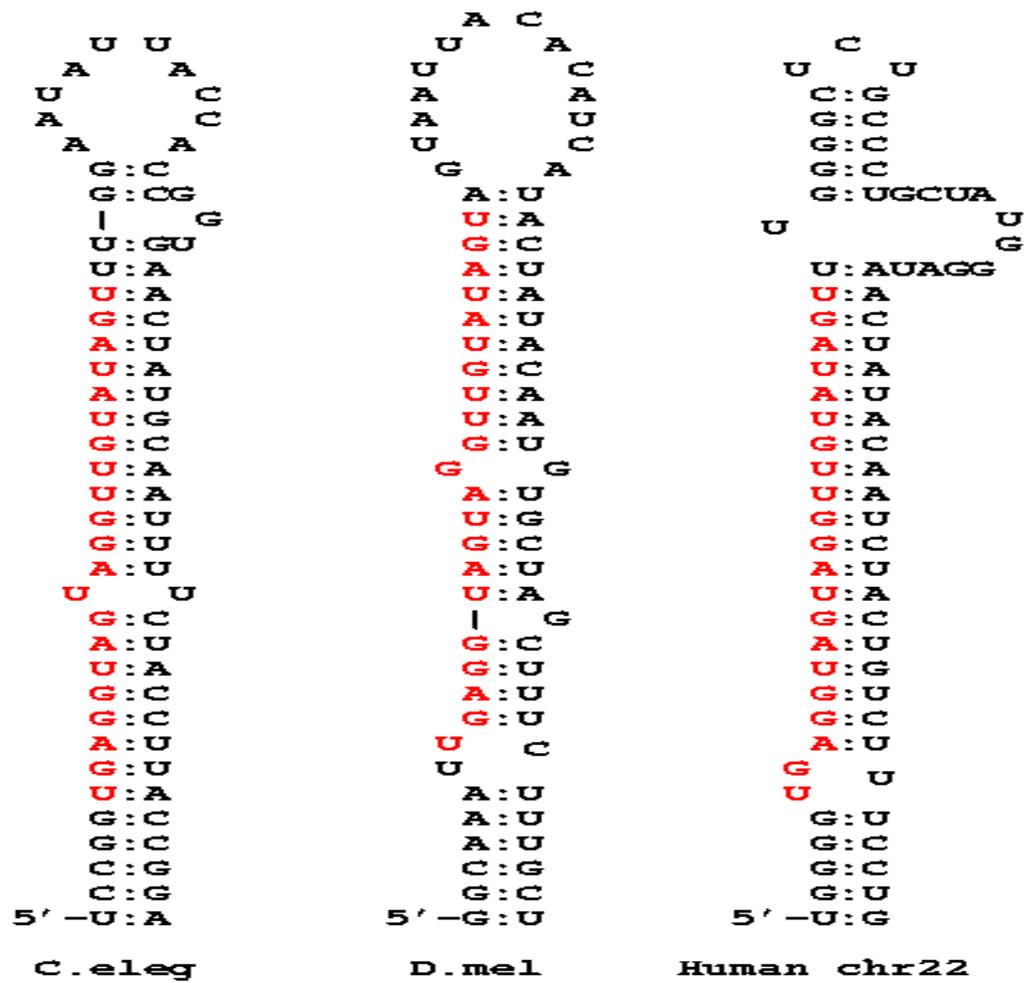


SECIS
elements
in human
Selenoprotein
M mRNA

DsrA small noncoding RNA of E.coli



thi Box Regulated
Element of thiamin
biosynthetic gene expression



RNA Secondary Structures of *let-7* microRNAs

FSRs are uniquely folded

Functional RNA molecules often contain important structural motifs including distinct loop sequences and specific combinations of base pairings.

Some FSRs, such as tRNA and RNase P RNA are not well represented by the most thermodynamically stable structure.

Studies of RNA folding energy landscapes suggested that functional structured RNAs (FSRs) were often thermodynamically more stable than would be anticipated for equivalent random sequences.

An Example for RNA Secondary Structure Prediction

RNA Sequence:

5'-UACGCCACAAAGUGGCCCC-3'

RNA Structure:

	A	E (kcal/mol) computed by Turner's Rule
	A A	4.5 (for a hairpin loop of 3 bases)
	C--G	
	A--U	-2.1 (for a base-pairs stacking of C:G and A:U)
	C--G	-1.8
	C--G	-2.9
5'-UACG--CCC-3'		-3.4
		-1.1 (for a terminal mismatching stacking)
		-6.8 Total Energy

Calculation of the lowest free energy:

Dynamic Programming Algorithm

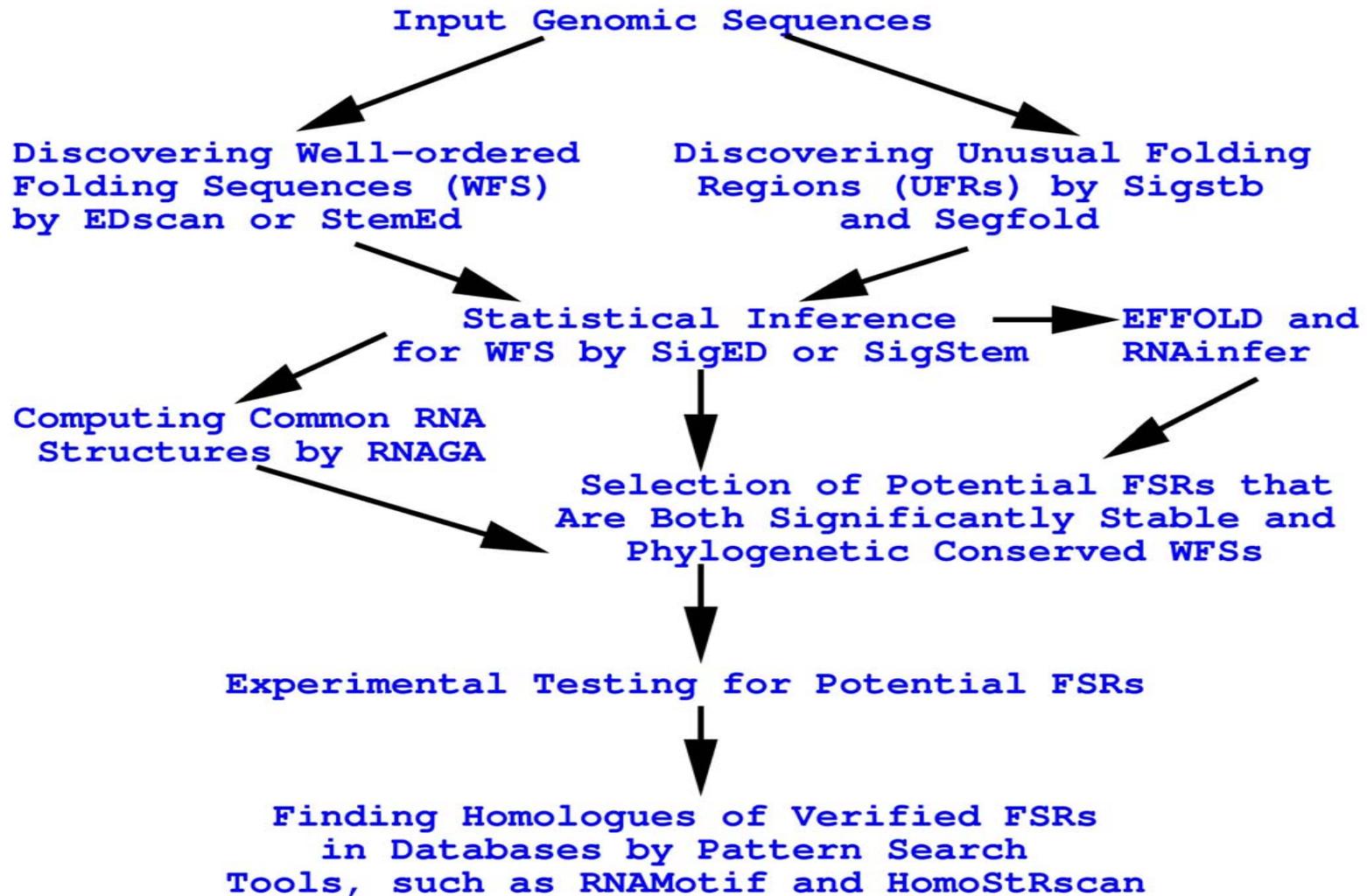
$V(i,j)$ minimal energy in a fragment, $S(i,j)$, $i:j$

$W(i,j)$ minimal energy in a fragment, $S(i,j)$, regardless i and j formed a base-pair or not

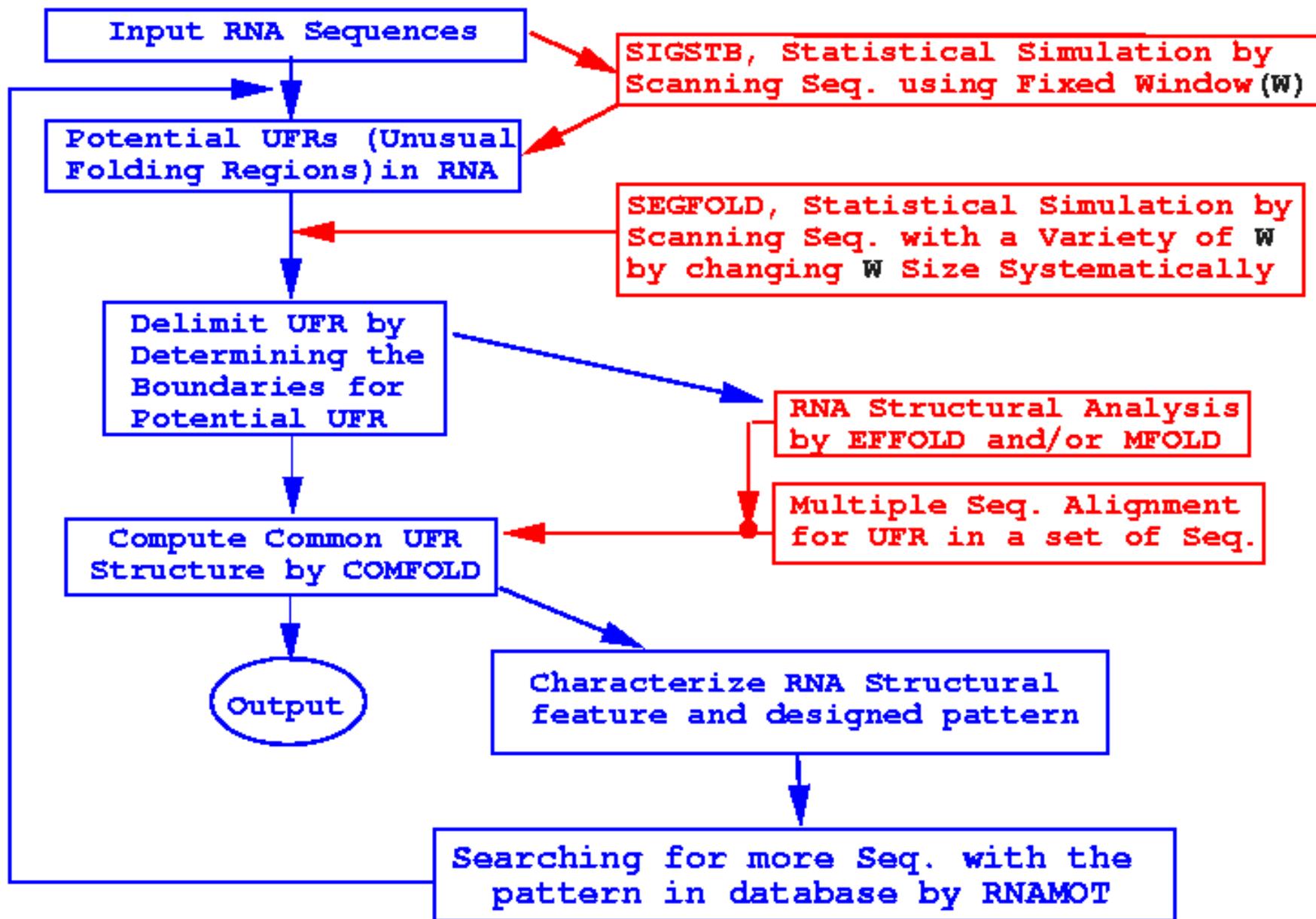
$$V(i,j) = \min \{ \begin{array}{l} Eh(i,j) \quad \text{minimal } E \text{ of the structure if } i:j \text{ closes a hairpin loop} \\ V(i+1,j-1) + Es(i,j) \quad \text{a stacking } E \text{ if } i:j \text{ stack over } (i+1):(j-1) \\ EIL(i,j), \quad EBL(i,j) \quad \text{minimal } E \text{ if } I:j \text{ closes a interior or bulge loop} \\ \min \{ V(i,k) + V(k+1,j) \} \quad i \leq k < j \end{array} \}$$

$$W(i,j) = \min \{ \begin{array}{l} V(i,j) \\ V(i+1, j) + d(i) \quad \text{dangling contribution of base } i \\ V(i, j-1) + d(j) \\ V(i+1, j-1) + d(i) + d(j) \\ \min \{ W(i,k) + W(k+1, j) \}, \quad i \leq k < j \end{array} \}$$

Procedure of Discovering FSRs in Genomic Sequences



Flowchart of Data Mining of Structure Features in RNAs



SIGSTB and SEGFOLD:

SIGSTB and SEGFOLD are variations of Zuker's Mfold.

Mfold is Zuker's implementation of a dynamic programming algorithm to predict the maximally stable secondary structure from an RNA sequence.

Lowest free energy (E) is computed by the Turner energy rules.

SIGSTB

It is used to search for potential unusual folding regions (UFRs) by scanning a fixed window along the RNA sequence. In the scanning, we compute two quantitative measures Sigscr and Stbscr.

SEGFOLD

It employs a systematic search with automatically varying window size to define the limits of the local UFRs in an RNA sequence.

Using Significance score (Sigscr) and Stability score (Stbscr) we evaluate:

What is the E difference between a real folding and random foldings?

What difference is the folding E between a selected segment and other segments in the sequence?

UFR: A region that has unusual lower and higher scores of Sigscr and Stbscr.

How SIGSTB and SEGFOLD Find Statistically Unusual Features

choose successive, overlapping segments by sliding a fixed window



Fold the local segment, S_i and compute the lowest free energy (E_i) from the natural sequence.

Shuffle the local segment and get a set of random segments; fold these random segments and compute the sample mean (E_r) and the sample standard deviation (std_r) of those E values.

$$Sigscr = (E_i - E_r) / std_r$$

For all E values ($E_1, \dots, E_i, \dots, E_n$) computed from the real RNA we also calculate the sample mean (E_w) and sample standard deviation (std_w)

$$Stbscr = (E_i - E_w) / std_w$$

Discovering FSRs in the 5'UTR

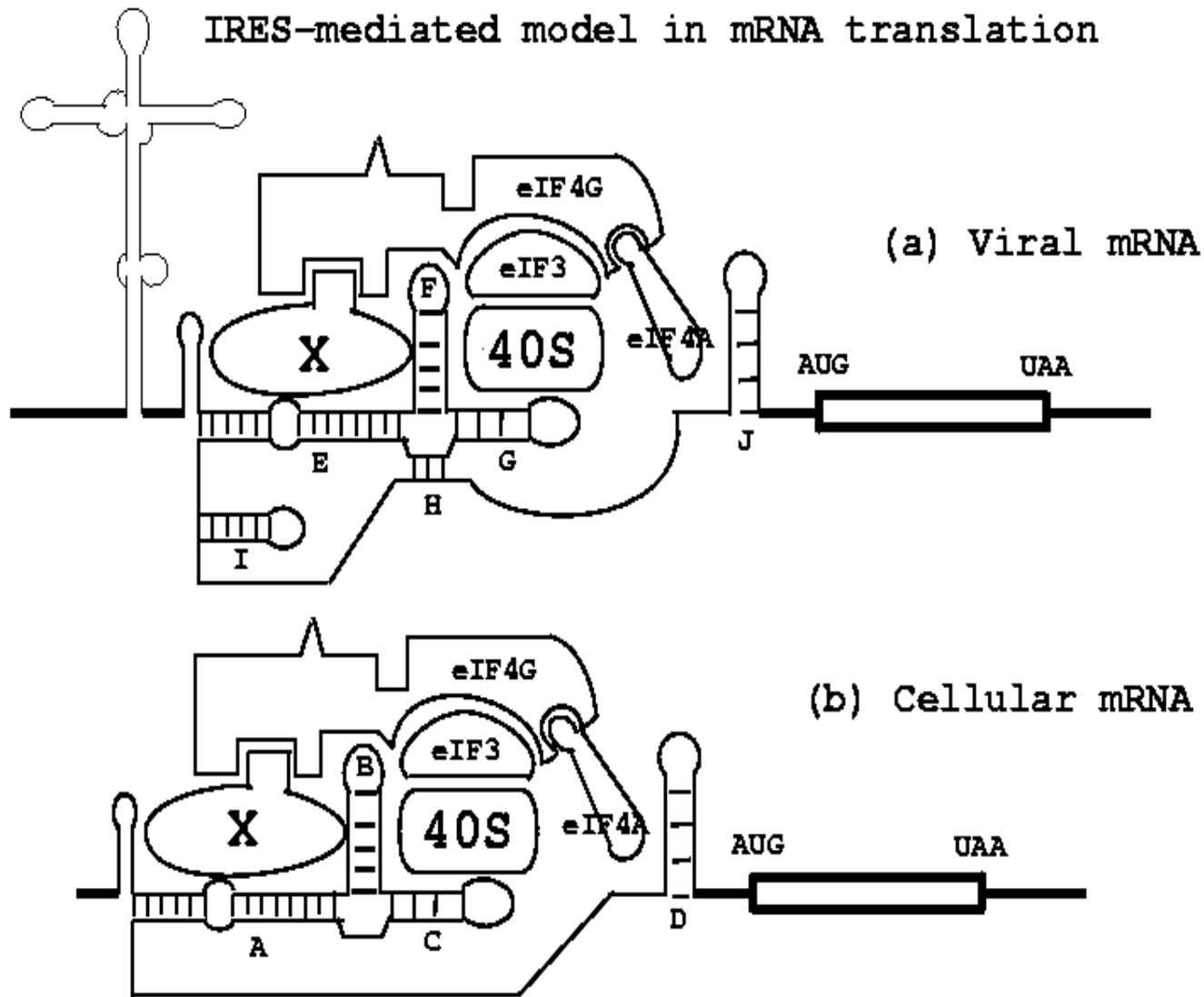
IRES: Internal ribosome entry segment in the 5'UTR

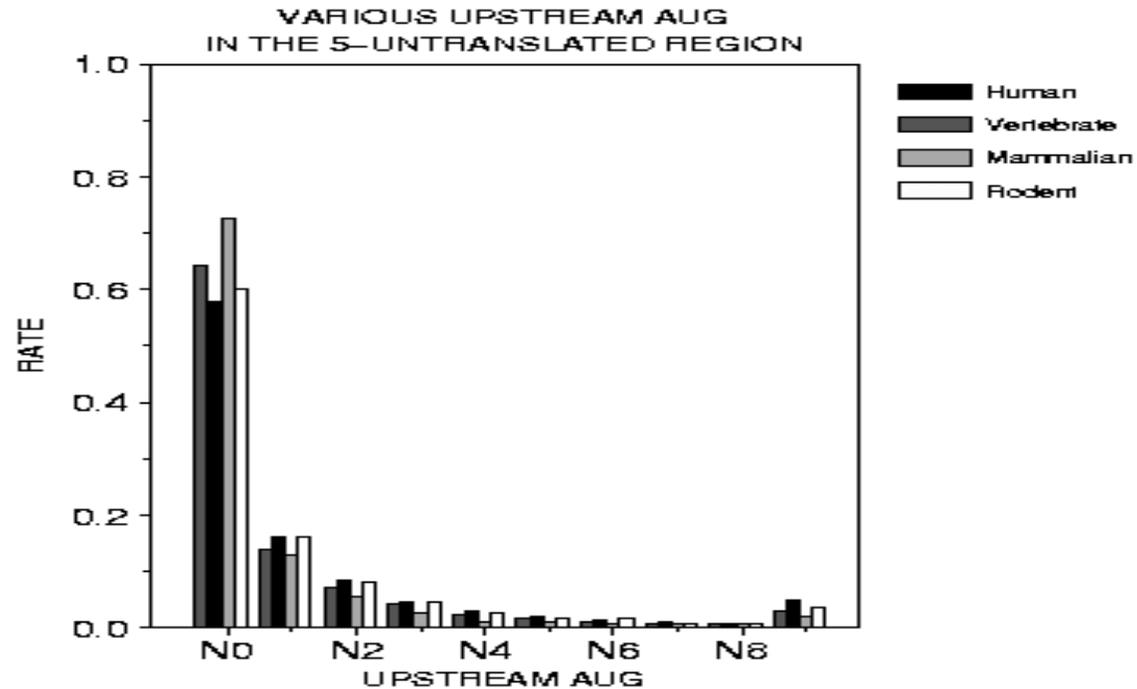
Function: Allow the translational machinery to skip over upstream AUGs

Detected IRES elements in cellular mRNAs are quite divergent and their sizes range from less than 100 nt to ~ 630 nt.

- A. mRNA is considered as the immediate source of information for translation from sequences in the four bases, A, C, G, U to the twenty amino acids of proteins.**
- B. A mRNA has the coding region, 5' untranslated region (5'UTR) before the initiator AUG and 3'UTR following the coding region.**
- C. Most mRNAs encoding oncoproteins and cell factors possess a long, GC-rich 5'UTR that contains one or more upstream AUG (uAUG) triplets and a complex RNA secondary structure.**
- D. Experimental studies revealed that there are cis-acting internal ribosome entry sites (IRES) in the 5'UTR which allowed the translational machinery to skip over the uAUG and initiate translation in the start codon.**

IRES-mediated model in mRNA translation





5'UTR	Total No.	Mean of Length
Human	6669	207-nt
Other mammal	2106	138-nt
Rodent	7056	192-nt
Vertebrate	2846	152-nt
Invertebrate	4054	205-nt

Cellular IRES and 5' UTR

1580 nt, 15 uAUGs

_____ AUG
human AML1/RUNX1

1022 nt, 3 uAUGs

_____ AUG
human PDGF2/c-sis (platelet-derived growth factor B)

483 nt, 3 uAUGs

_____ AUG
human FGF-2 (fibroblast growth factor 2)

368 nt, 4 uAUGs

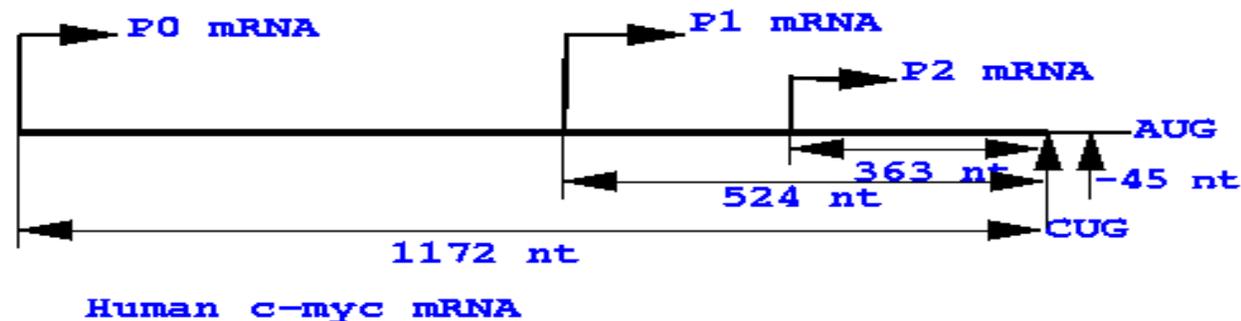
_____ AUG
human eIF4G (eukaryotic initiation factor 4 gamma)

221 nt

_____ AUG
human BiP (immunoglobulin heavy chain binding prote

1038 nt, 1 uAUGs

_____ AUG
human VEGF (Vascular Endothelial Growth Factor)



Computation of Common RNA Secondary Structure

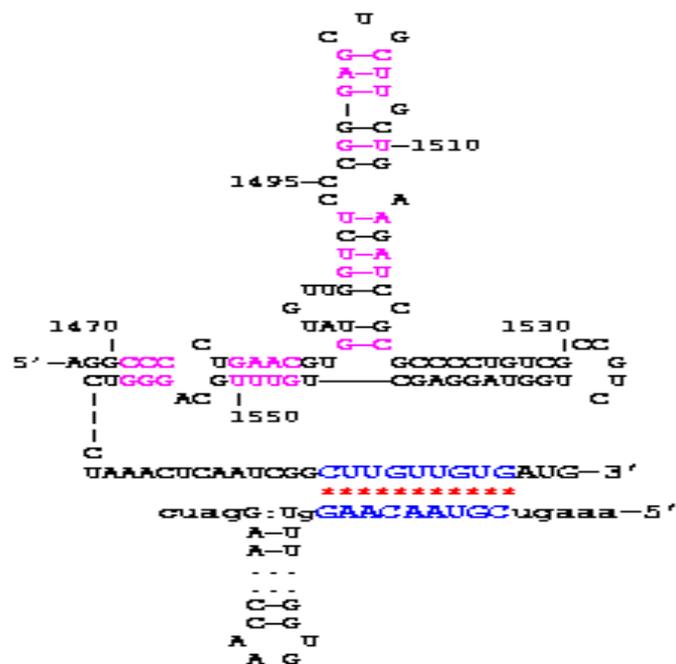
COMFOLD and RNAGA

COMFOLD follows the steps of phylogenetic method and includes four steps:

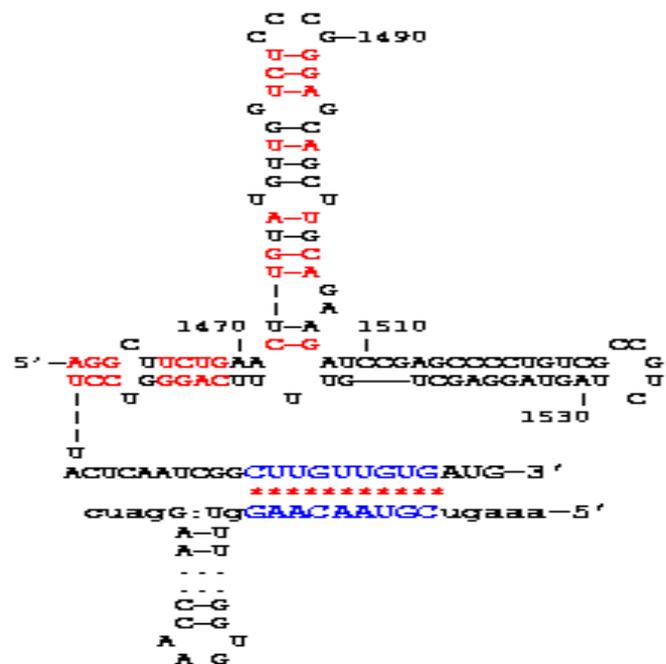
- 1. Align a set of homologous sequences by multiple alignment.**
- 2. Generate a list of possible thermodynamically favored stems folded in the sequence by program EFFOLD considering the uncertainties of energy parameters for the formation of RNA duplexes and loops in Turner's energy rules. The primary approach is to simulate the energy rules by perturbing the parameters within the range of the experimental errors under a predetermined normal distribution. Using the "simulated energy rules" and dynamic programming algorithm of mfold, the most stable structures are computed, and all thermodynamically favorable stems in the simulations are compiled.**
- 3. Make a conserved stem list by inspecting equivalent base pairings for each computed stem by program MATCH and multiple sequence alignment.**
- 4. Build the optimal structure with the maximum score from the conserved stem list by program BUILD.**

**Common Structural motif Predicted
in the 5'UTR Upstream of the Start
Codon AUG of AML mRNAs**

**a. Human AML1-a mRNA
(1468-1583)**



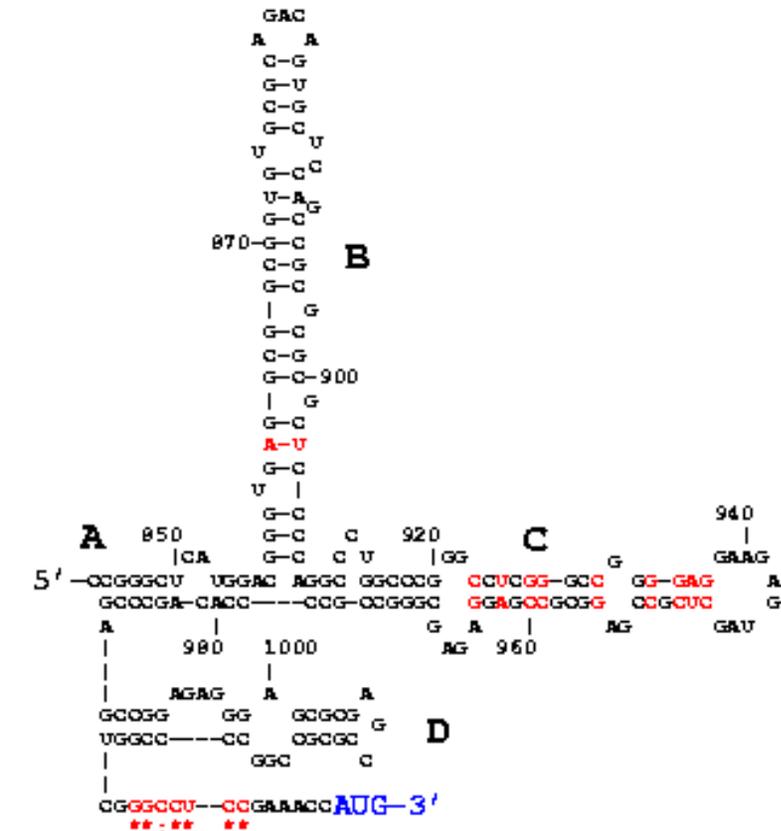
**b. Mouse PEBP2ab2 mRNA
(1461-1576)**



Common RNA Structures Predicted in the 3' Portion of VEGF 5' UTR

(a) Human (844-1041)

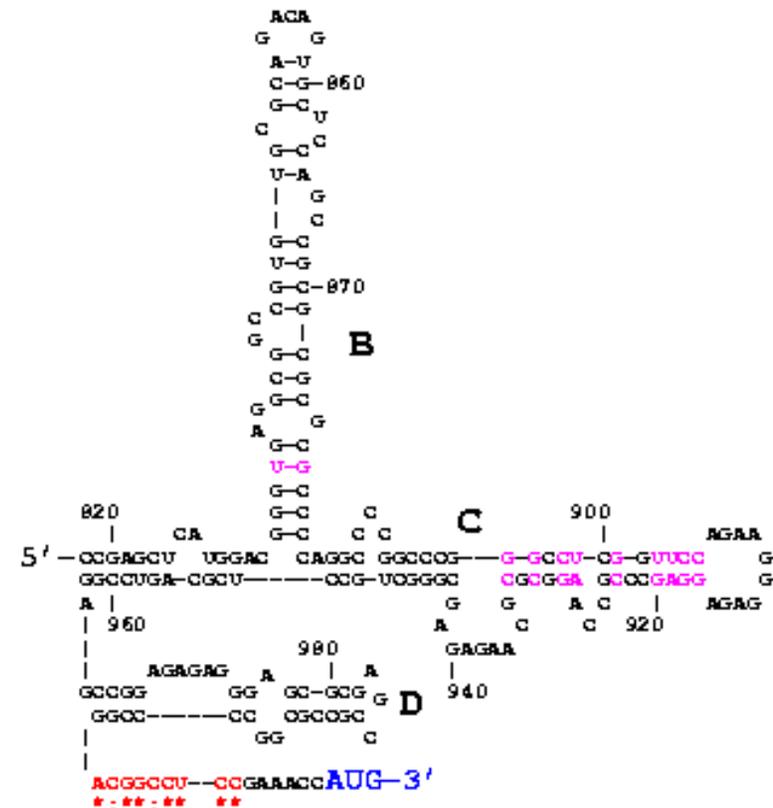
(b) Mouse (818-1017)



3' - AUUACUAGG UGGAACAAUVCUGAAA-5'

A-U
A-U
...
...
C-G
C-G
A U
A G

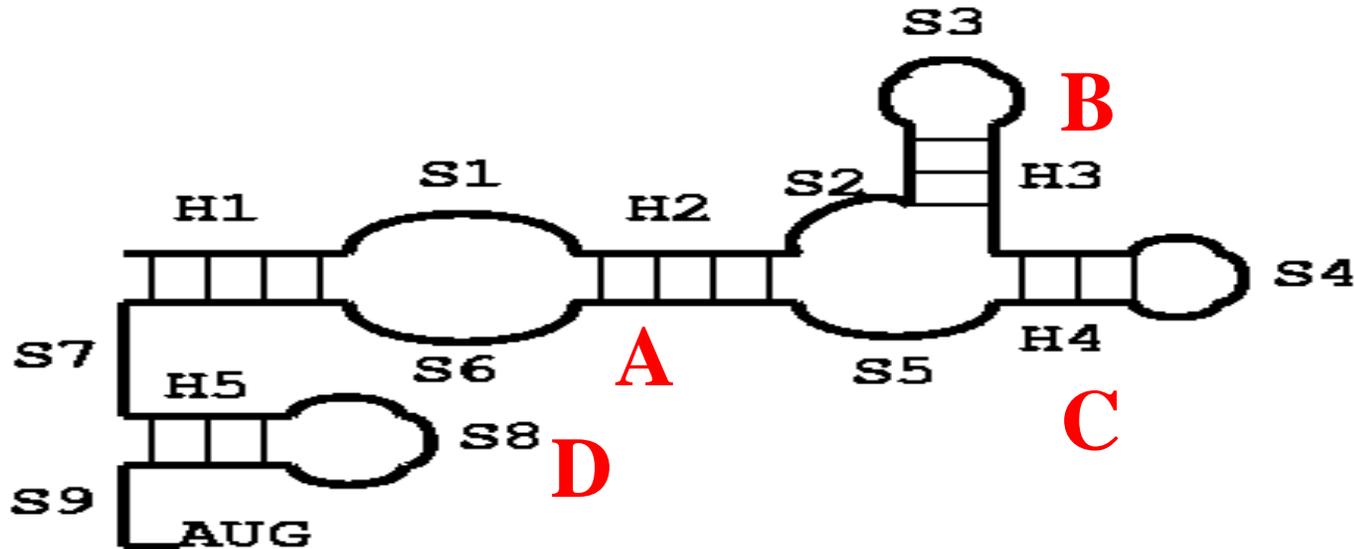
Human 18S rRNA 3'-end (1823-1869)



3' - AUUACUAGG UGGAACAAUVCUGAAA-5'

A-U
A-U
...
...
C-G
C-G
A U
A G

Structural Motif in the 5'NTR



A description file for pattern search by the program RNAMOT:

H1 S1 H2 S2 H3 S3 H3 H4 S4 H4 S5 H2 S6 H1 S7 H5 S8 H5 S9

H1 4:6 0; H2 4:9 0; H3 4:6 0; H4 4:6 0; H5 4:6 0

S1 0:7; S2 0:1; S3 3:17; S4 3:17; S5 0:5; S6 0:6

S7 2:9; S8 3:17; S9 3:11 AUG

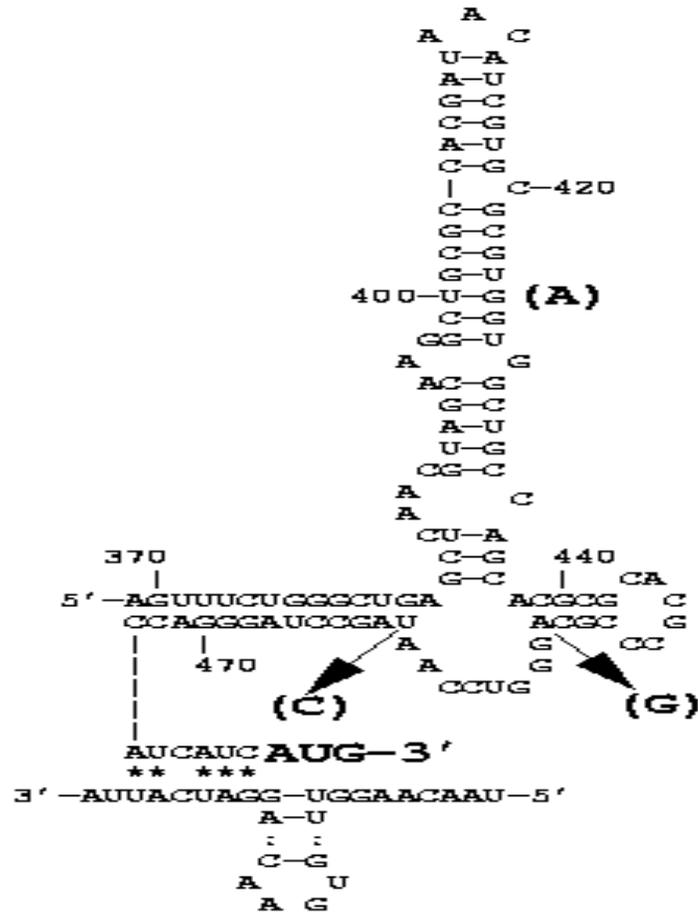
R H3 H4 H2; M 0; W 9

5'UTR	Size (nt)	No of uAUG	Region of Y Motif	Stem D	Complementary Sequence
human					
abl(M14753)	340	6	225-321	N	GGU--ACC-UAUUAUuACUUU-AUG
bcr(X02596)	488	2	373-455	N	GGCGG--CGC(9 nt)CGGC-(6 nt)AUG
erb(Y00479)	333	3	117-249	Y	GGcAUCC(9 nt)UUGaaGUGA-AUG
erbA(X55005)	466	4	258-382	Y	GGcAUCC(9 nt)UUGaaGUGA-AUG
IL15(X91233)	316	10	221-290	N	UAAgGAUUUACCGUGGCUUU-(5 nt)AUG
Int2(X14445)	491	3	396-456	Y	GAUGCC-(3 nt)AUG
mas (M13150)	267	3	117-249	N	CCaACCU-GaGGCcU-(4 nt)AUG
mos (J00119)	479	1	369-473	N	AUcAUC-(0 nt)AUG
mouse					
abl2(U13835)	144	3	55-125	N	UCCaGCCUcCGAC-(0 nt)AUG
abl3(X07539)	219	0	90-166	N	GGUCCuugugaGCCacgUUGUGGU..AUG
abl4(X07541)	1168	11	1059-1145	N	CACCUaUUAUuGCUUU-AUG
Int2(Y00848)	864	3	776-833	Y	GAUGCC-(3 nt)AUG
mas (U96273)	341	5	86-212	N	CACCg-(0 nt)AUG
VEGF(U41383)	1014	1	818-962	Y	AcGGcCU-CC(6 nt)AUG
rat					
BiP(M14866)	206	0	114-169	N	CCGCUgagcgACuGACU-(19 nt)AUG
FGF2(M22427)	532	0	310-375	N	GUCCgGCU-(8 nt)CUG
			438-510	N	GUCCcgggGCCGCGG-(7 nt)AUG
c-myc(Y00396)	413	0	71-222	Y	UUAUU-UGA(3 nt)CUG
mos(X52952)	482	2	369-477	N	UAAUc-(0 nt)AUG

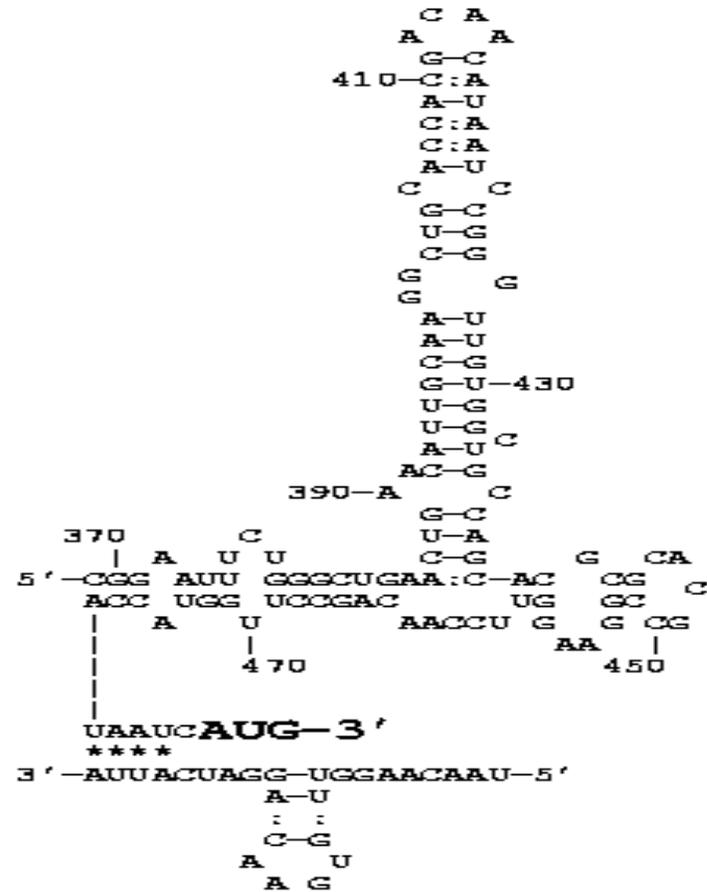
RNA higher order structure folded in the 5' leader of oncoprotein mos mRNA

(A) Human and African Green Monkey c-mos mRNA (369-473)

(B) Rat c-mos (369-477)



Human 18S rRNA 3'-end (1830-1869)

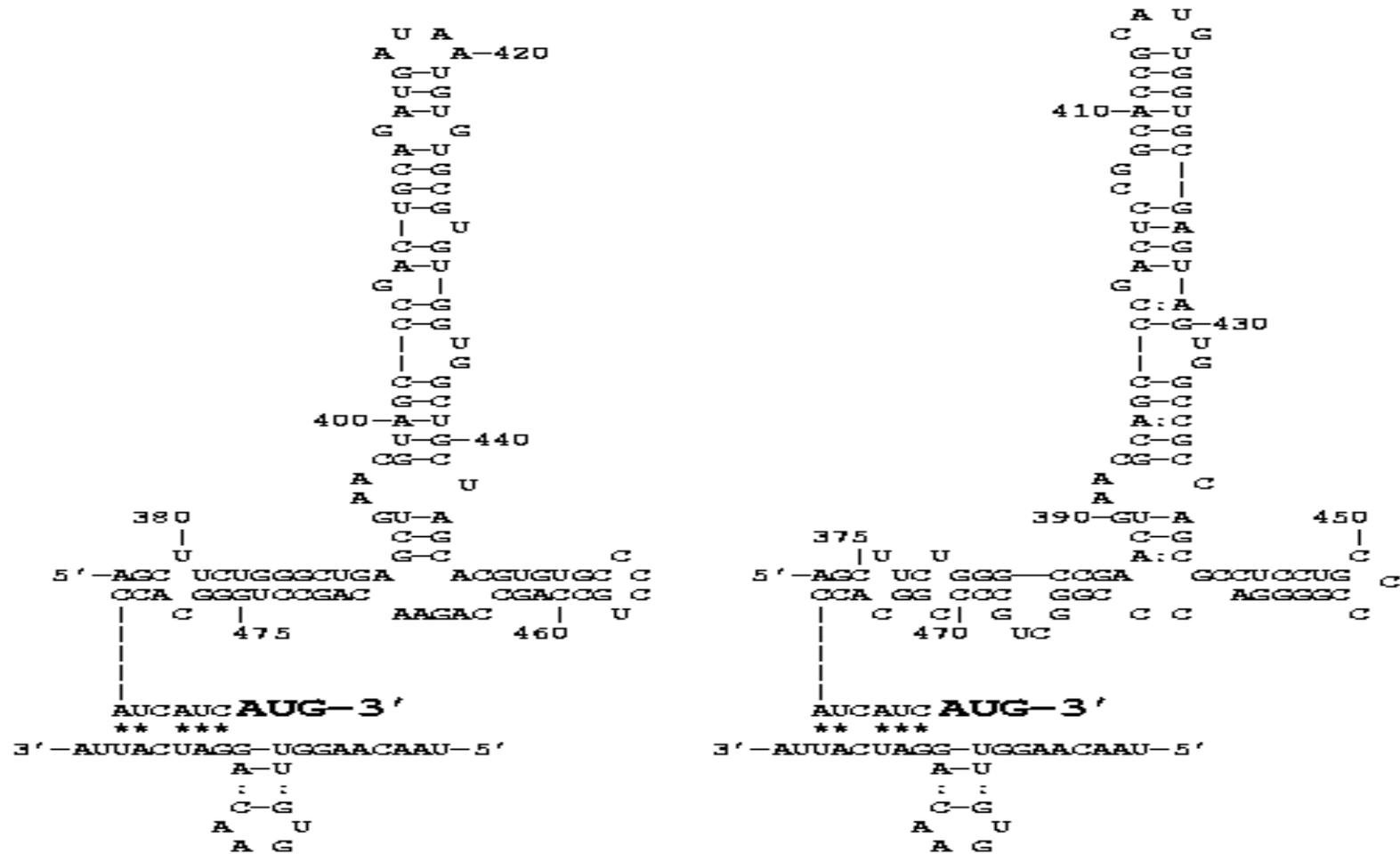


(The base mutation in the African Green Monkey c-mos mRNA is indicated in parentheses.)

**RNA higher order structure folded in
the 5' leader of oncoprotein mos mRNA**

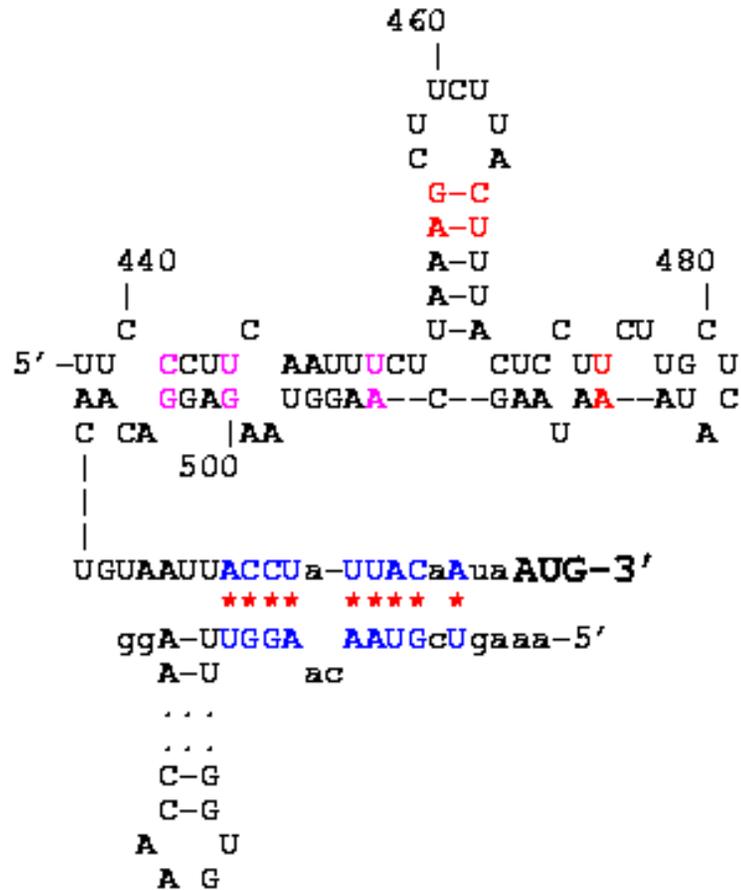
(C) Chicken c-mos (377-481)

(D) Xenopus c-mos (373-477)

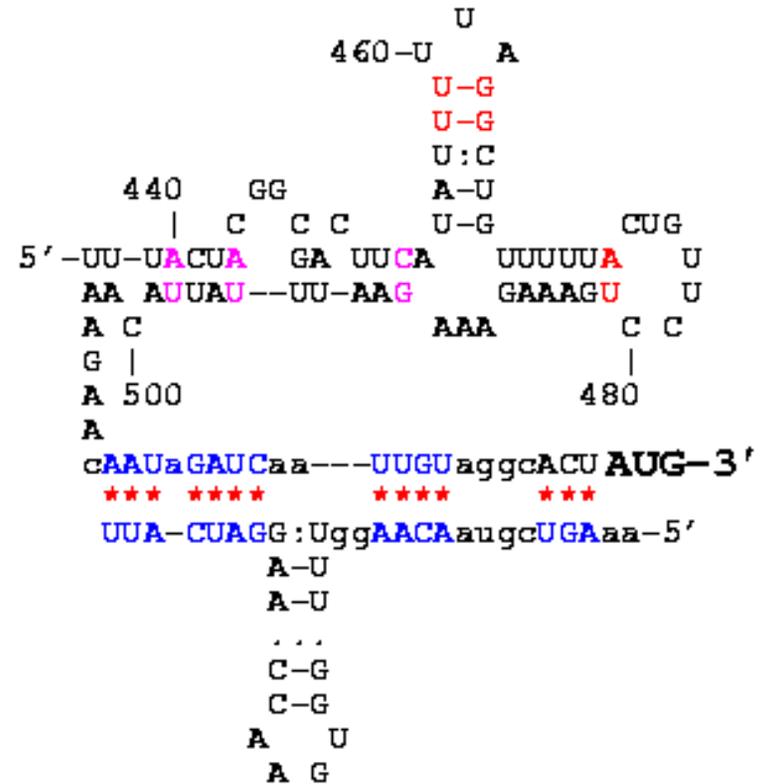


RNA Structures folded in the 5'UTR of Yeast mRNAs
for Eukaryotic Initiation Factor 4 Gamma (eIF4G)

(A) TIF4631 (438--531)



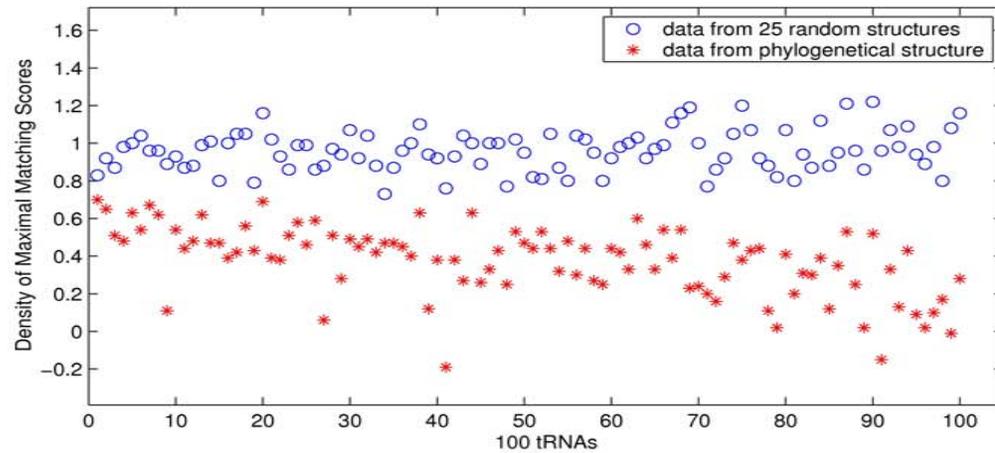
(B) TIF4632 (437--531)



5'UTR	Size (nt)	No of uAUG	Region of Y Motif	Stem D	Complementary Sequence
hamster					
BiP(M17169)	150	0	64-125	N	GGCCACagcGCcGGC-(3 nt)AUG
green monkey					
mos(X12449)	479	1	369-473	N	AUcAUC-(0 nt)AUG
bovine					
VEGF(M32976)	533	0	351-494	Y	A-GGcCU-CC(6 nt)AUG
chicken					
mos(M19412)	487	4	377-481	N	AUcAUC-(0 nt)AUG
Xenopus					
mos(X13311)	483	2	373-477	N	AUcAUC-(0 nt)AUG
Yeast					
TIF4631					
eIF4F(L16923)	528	11	438-507	N	ACCUaUUAC(4 nt)AUG
TIF4632					
eIF4F(L16924)	528	4	437-502	N	AAUaGAUCaaUUGU-AgGcACU-AUG

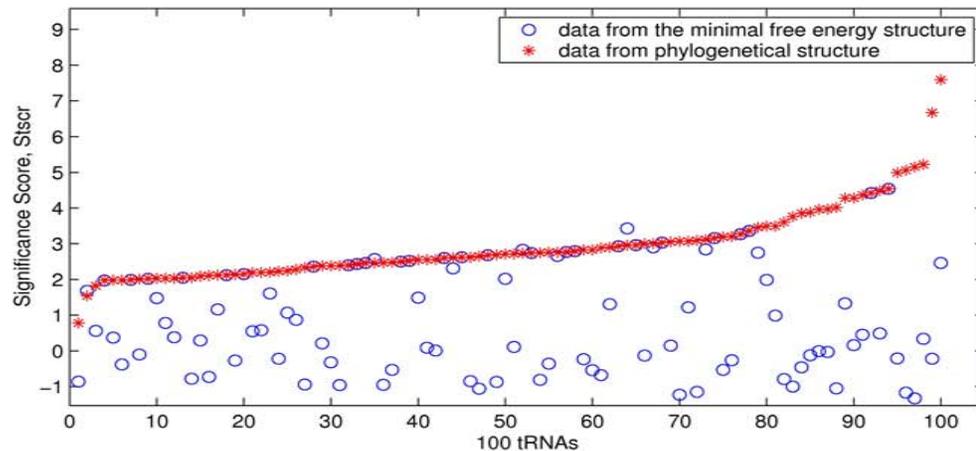
Evaluation of Structural Uniqueness

1. A quantitative measure, MSS (maximal similarity score)
2. Uniqueness of a RNA structure can be estimated by evaluating
Difference between the average MSS computed from a natural RNA and
Its related, randomly shuffled sequences and those MSS computed from
Random versus random sequences.
3. Computational experiment on 100 tRNAs indicates:
Structural conformations folded by tRNAs are significantly different from
Those of corresponding random structures, and the uniqueness of the
Common cloverleaf structure is statistically significant.



**RR: sample mean of MSS
(maximal similarity scores)
computed from 7500 randomly
shuffled sequences for each tRNA.**

**NR: sample mean of MSS
computed from the natural tRNA
versus 300 corresponding
randomly shuffled sequences**



$$Stscr = (RR - NR) / std$$

*The greater Stscr, the statistically
more significant is the well-ordered
structure of the natural RNA.*

The well-ordered structures of FSRs are both thermodynamically stable and uniquely folded.

The measurement of thermodynamic stability (folded free energy) alone is not sufficient enough for us to characterize the structural features folded in FSRs.

FSRs consist of a well-ordered folding sequence (WFS)

Searching for Well-ordered Folding Sequences in RNA/DNA (programs StemEd and *SigStem*)

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 folding pattern.**
- B. **Local folding in DNA and RNA is closely associated with its biological functions.**
- C. **The distinct, well-ordered 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:

Defining the Uniqueness of RNA/DNA Secondary Structures.

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

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

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

$E(S_i)$: lowest free energy of the optimal structure in a local segment S_i

$E_f(S_i)$: lowest free energy of the optimal restrained structure in S_i , where all the previous base pairings formed in the optimal structure are prohibited.

$E_{diff}(w)$ and $std(w)$ are the sample mean and standard deviation of the E_{diff} of all overlapped fixed-length segments computed by sliding a window stepped a few nucleotides each time from 5' to 3' along the sequence.

Calculation of the lowest free energy:

Dynamic Programming Algorithm

$V(i,j)$ minimal energy in a fragment, $S(i,j)$, $i:j$

$W(i,j)$ minimal energy in a fragment, $S(i,j)$, regardless i and j formed a base-pair or not

$$V(i,j) = \min \{ \begin{array}{l} Eh(i,j) \quad \text{minimal } E \text{ of the structure if } i:j \text{ closes a hairpin loop} \\ V(i+1,j-1) + Es(i,j) \quad \text{a stacking } E \text{ if } i:j \text{ stack over } (i+1):(j-1) \\ EIL(i,j), \quad EBL(i,j) \quad \text{minimal } E \text{ if } I:j \text{ closes a interior or bulge loop} \\ \min \{ V(i,k) + V(k+1,j) \} \quad i \leq k < j \end{array} \}$$

$$W(i,j) = \min \{ \begin{array}{l} V(i,j) \\ V(i+1, j) + d(i) \quad \text{dangling contribution of base } i \\ V(i, j-1) + d(j) \\ V(i+1, j-1) + d(i) + d(j) \\ \min \{ W(i,k) + W(k+1, j) \}, \quad i \leq k < j \end{array} \}$$

The minimal energy of an optimal structure folded by a RNA, $S(l,n)$ is $W(l,n)$ that is computed from the recurrent computation of $W(i,j)$.

Computational time complexity is $O(n^3)$

To improve the speed of the DP algorithm, the new approach *StemEd* is restricted to search for all possible stem-loops in excluding the multi-branch and opening bifurcate stem-loops.

The computational complexity of $W(i,j)$ is reduced to $O(n^2)$.

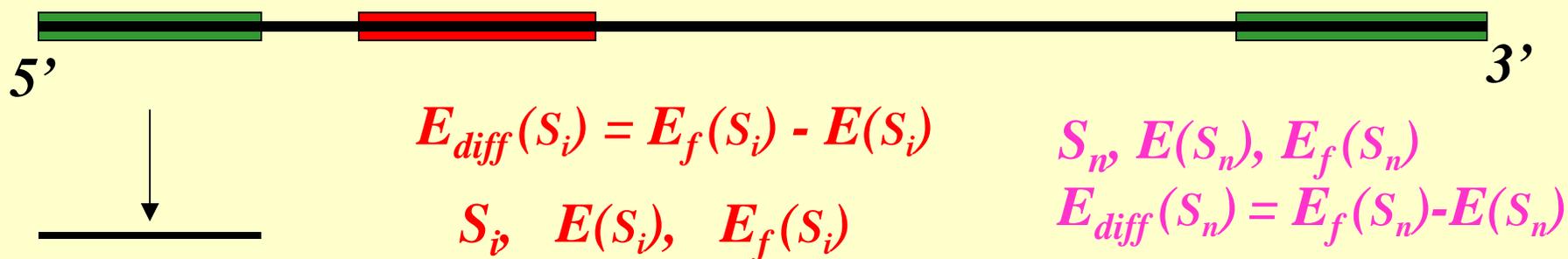
$$V(i, j) = \min \{ \begin{array}{l} Eh(i, j) \\ Es(i, j) + V(i+1, j-1) \\ EIL(i, j), \quad EBL(i, j) \end{array} \}$$

$$W(i, j) = \min \{ \begin{array}{l} V(i, j) \\ V(i+1, j) + d(i) \\ V(i, j-1) + d(j) \\ V(i+1, j-1) + d(i) + d(j) \end{array} \}$$

All energies of folded structures are computed by the DP algorithm and Turner Energy Rules (*Mathews et al., 1999, J. Mol. Biol., 288: 911-940*).

How Does *StemEd* /*EDscan* find statistically well-ordered folding segments in a sequence?

A. The principle: choose successive, overlapping segments



S_1 , computing $E(S_1)$ and the optimized RNA 2D structure, T .

prohibiting all base pairings in the predicted optimized $St. T$.

Re-computing the minimal $E_f(S_1)$ by DP when the previous base pairings are forbidden.

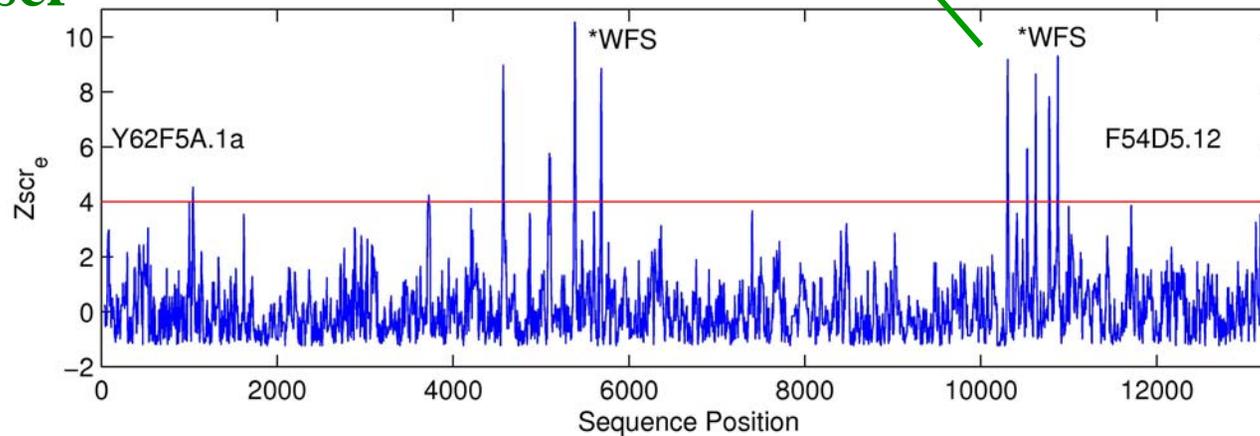
Computing Energy difference, $E_{diff}(S_1) = E_f(S_1) - E(S_1)$

Computing the normalized $Zscr_e(S_j) = (E_{diff}(S_j) - E_{diff}(w)) / std(w)$

$E_{diff}(w)$ and $std(w)$ are computed from $E_{diff}(S_1), \dots, E_{diff}(S_i), \dots, E_{diff}(S_n)$

mir-35, mir-37, mir-38, mir-39 and mir-40

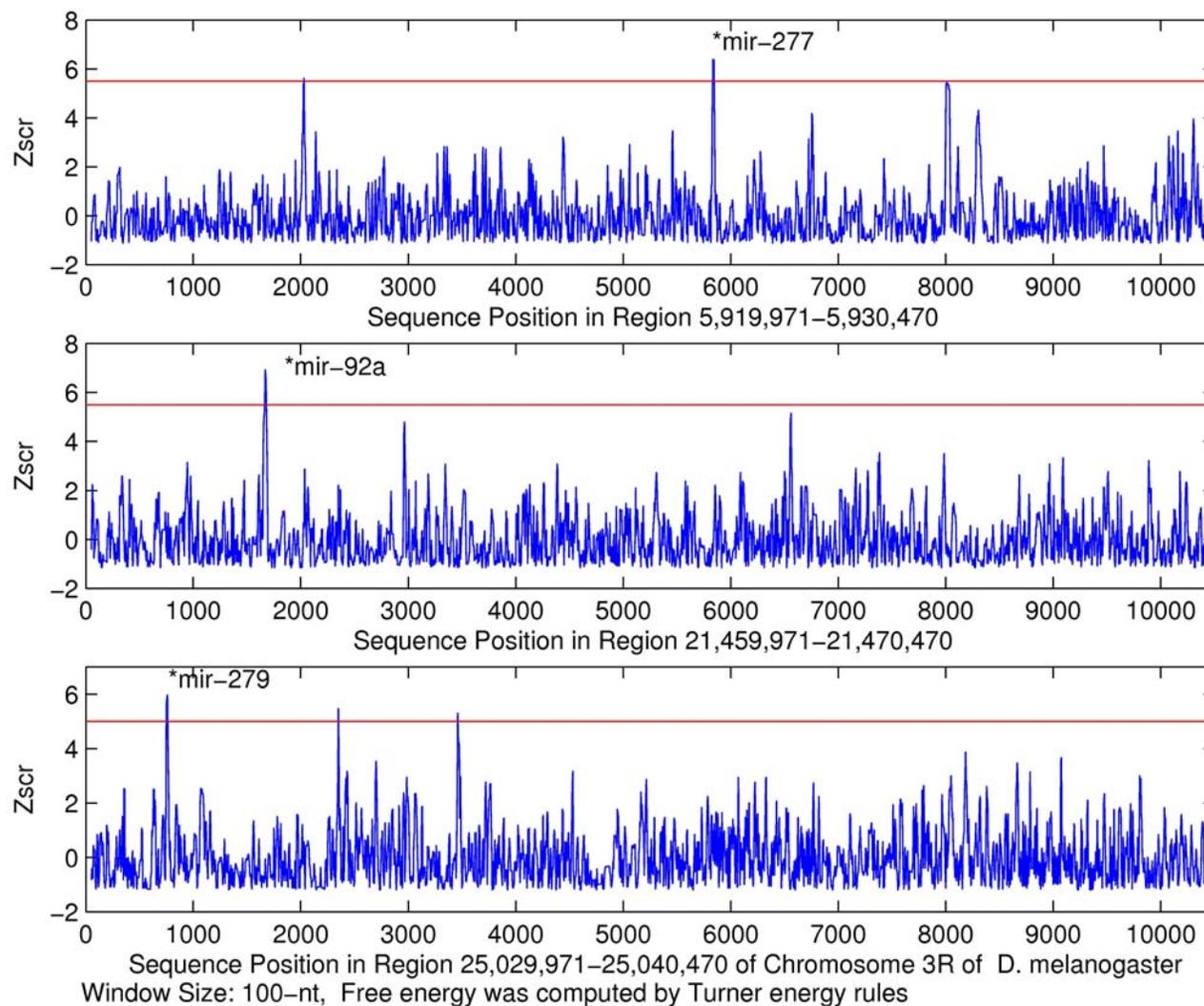
Zscr_e Zscr_e of Local Segments Computed in the Region from Genes F62F5A.1a to F54D5.12 of *C. elegans*



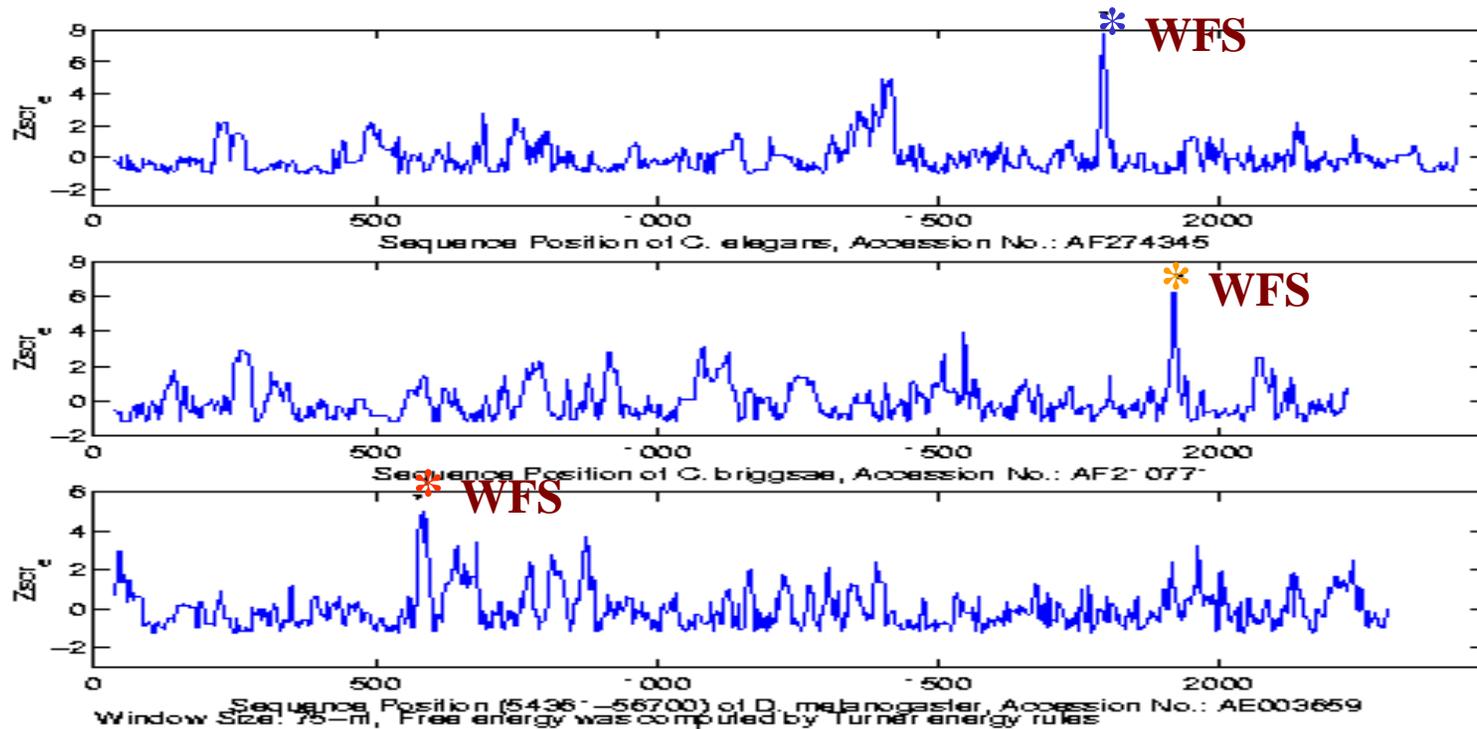
Window Size: 80-nt

**Sequence Region from Genes F62F5A.1a to F54D5.12
of *C. elegans* Chromosome II**

Chromosome 3R of *D. melanogaster*

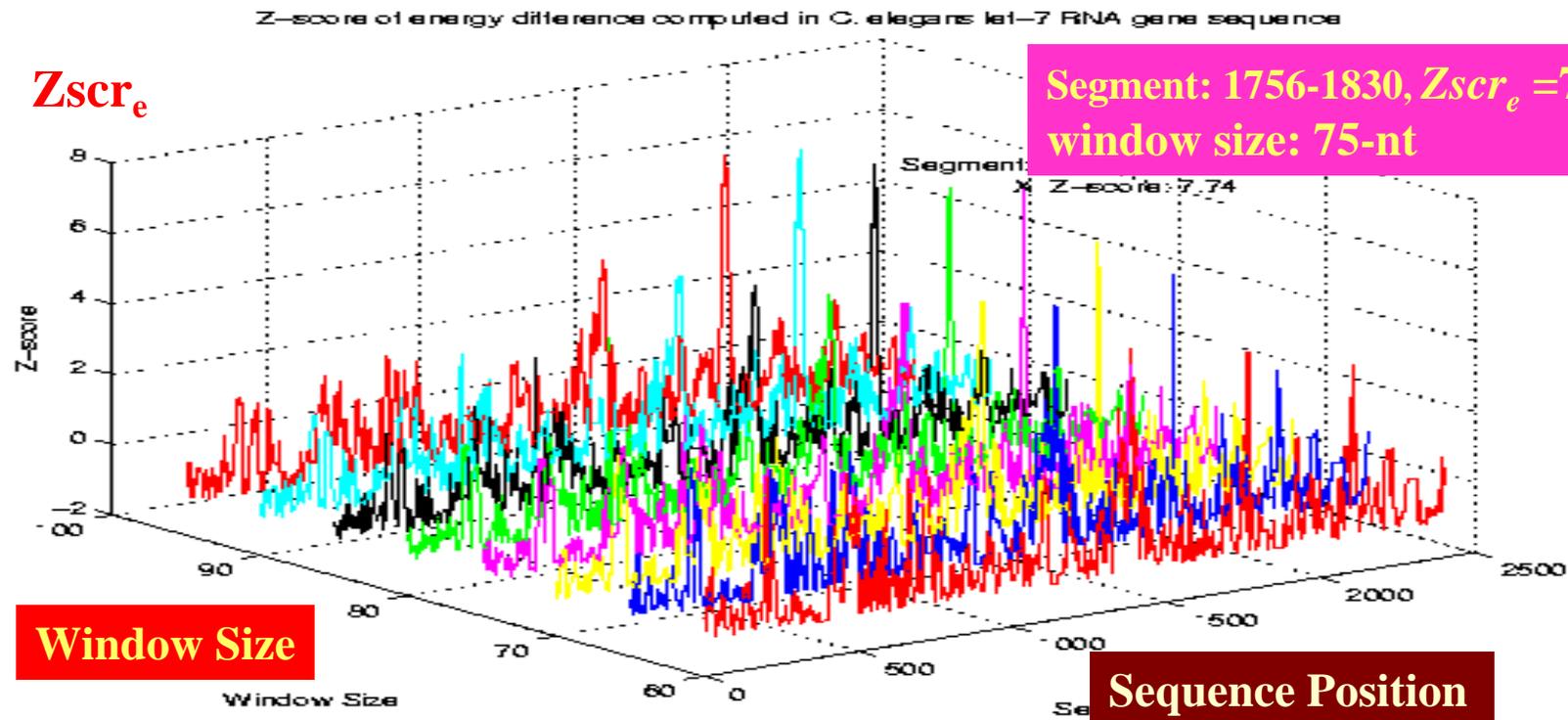


$Zscr_e$ Plots of the three genomic sequences of *C.elegans* (top), *C. briggsae* (middle) and *D. melanogaster* (bottom)



The plot was produced by plotting $Zscr_e$ against the position of the middle base in the window of 75-nt. In the plot of *D. melanogaster*, the sequence position 54,361 is numbered as position 1.

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)

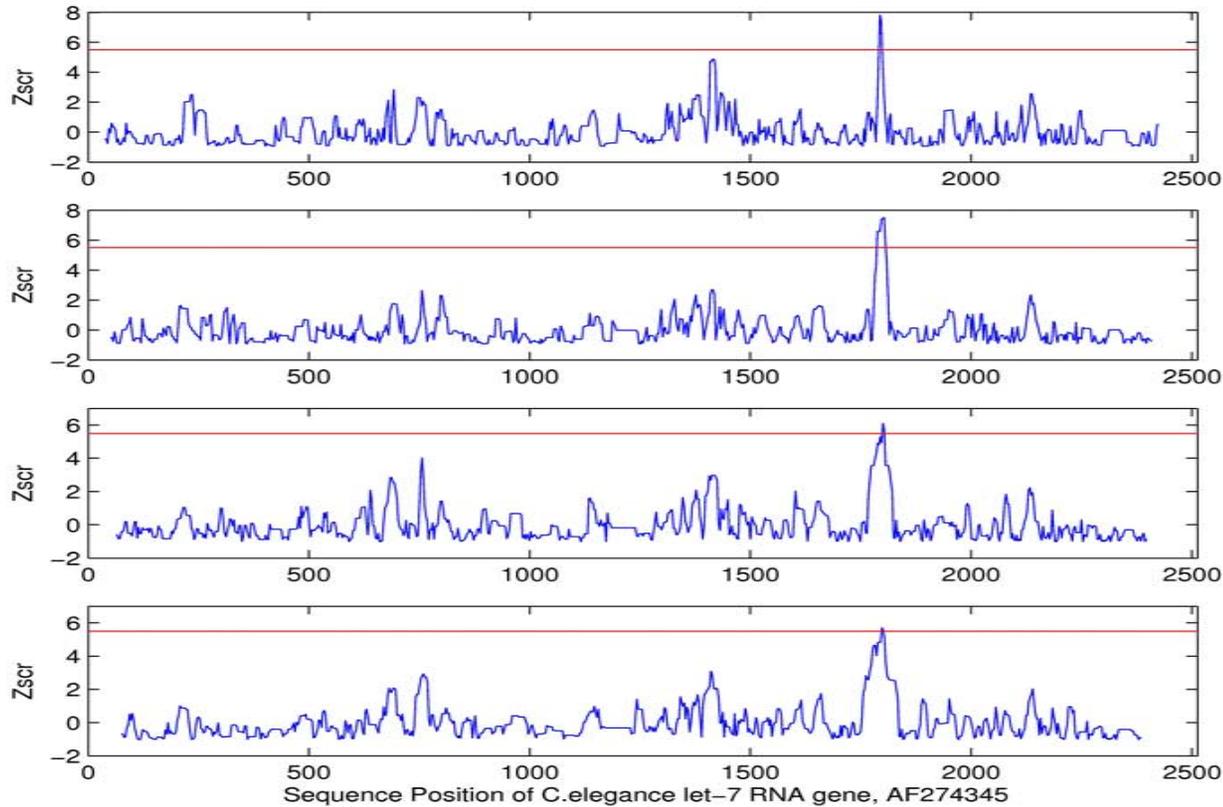


The $Zscr_e$ values were computed by moving a fixed window stepped 3-nt each time from 5' to 3'.

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.

Stability of the statistical inference for detecting WFS by *StemEd*

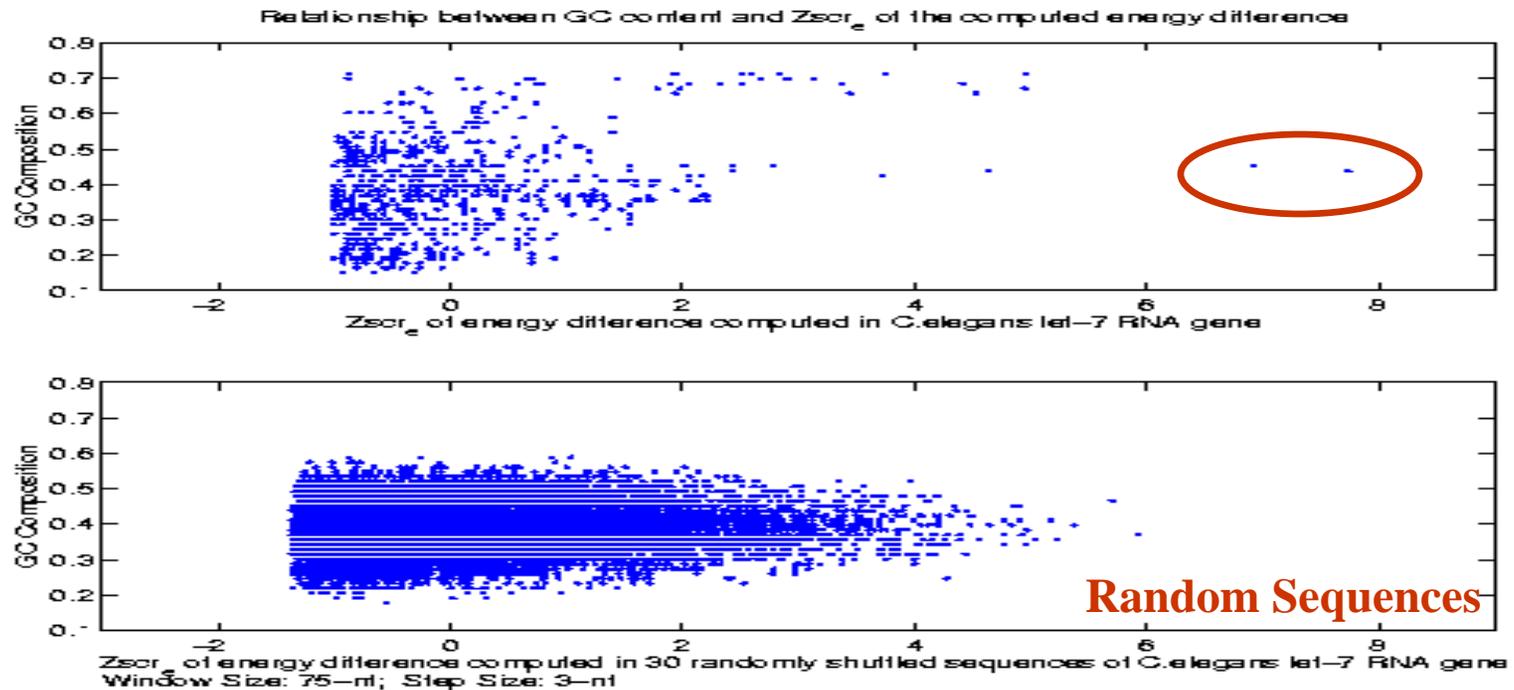
Window Size



Zscr computed by sliding a 75, 100, 125 and 150-nt window are shown in Figs. a-d, respectively

Let-7 miRNA gene sequence of C.elegans, AF274345

Relationships between GC composition and $Zscr_e$ of the local segment of 75-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



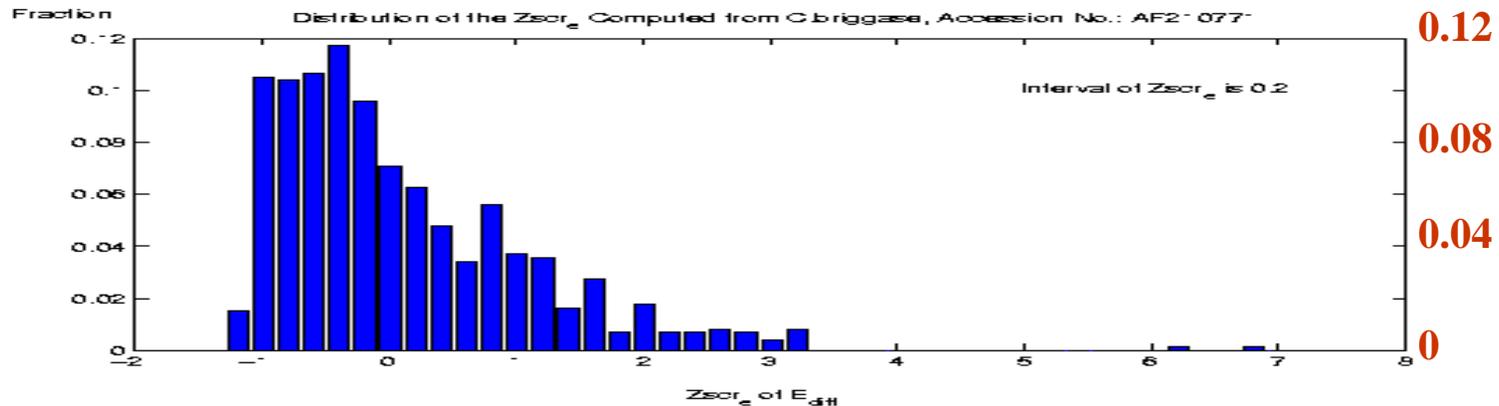
$Zscr_e$ Of Free Energy Difference

$Zscr_e$ ranged from -1.0 to 7.74 in the natural *let-7* gene, and the GC composition ranged from 0.15 to 0.70

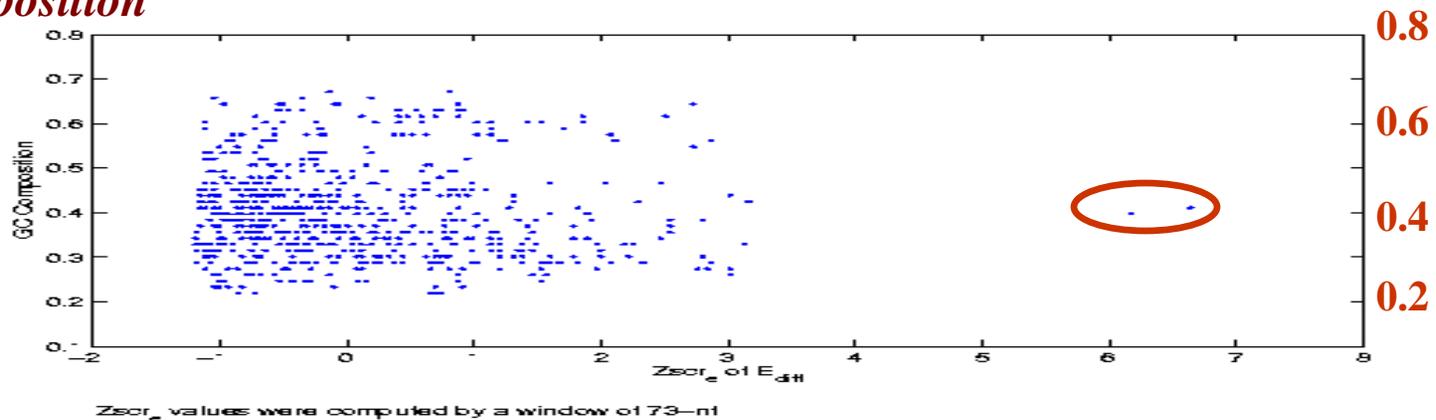
In the random sequences $Zscr_e$ ranged from -1.37 to 5.93 and GC composition ranged from 0.16 to 0.58

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 75-nt stepped successively in 3 nt from 5' to 3'.

Conclusion

- E_{diff} of the folded structure is a good measure to define quantitatively both the stability and uniqueness of the folding segment in a sequence.
- The statistically significant well-ordered folding structure 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.
- The general behavior of E_{diff} in a random sample that is related to the local segment is not represented in the measure, $Zscr_e$ of the local segment.
- To well estimate the statistical significance of $Zscr$ for a given segment we have to know what is the general behavior of E_{diff} of a set of random sequences that are made by randomly shuffling the local segment rather than the complete sequence.

Computational method, *SigStem* or *SigED* is designed to identify the local Structural features by computing energy difference E_{diff} and statistical evaluation in a random sample.

A. Monte Carlo simulations are adapted to estimate the uncertainty of E_{diff} in a random sample.

- 1. Computing $E_{diff}(S_i)$ for a given segment S_i in the sequence.**
- 2. Generating a large number of randomly shuffled sequences, $RS_{i,1}, RS_{i,2}, \dots, RS_{i,m}$ for the segment S_i , where the number m is determined by the length of the segment.
($m = 200$ if $60 < \text{length} \leq 150$)**
- 3. Computing $E_{diff}(RS_{i,k}), 1 \leq k \leq m$, for m random sequences.**
- 4. Computing the sample mean, $E_{diff}(RS_i)$ and sample standard deviation, $std(RS_i)$ for those energy difference, $E_{diff}(RS_{i,k})$.**

B. To characterize the structural features of the base-pairing regions and loops, respectively, E_{diff} is divided into two parts,

$$E_{diff} = E_{stem_{diff}} + E_{loop_{diff}}$$

C. Computing the standard z-scores, $SigZscr_e(S_i)$, $SigStem_e(S_i)$ and $SigLoop_e(S_i)$ for the given segment of the biological sequence.

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

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

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

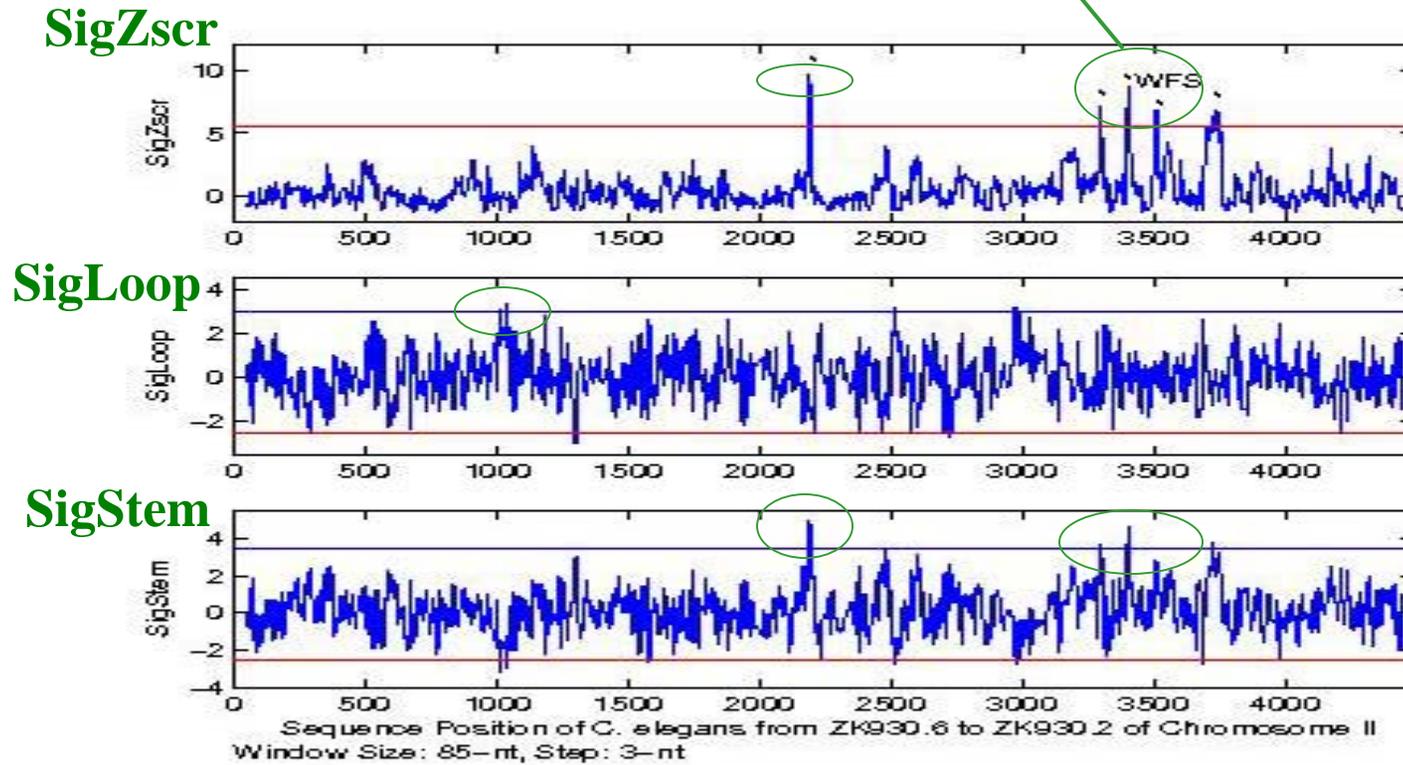
D. *SigED* is used to compute the three z-scores of $SigZscr_e(S_i)$, $SigZscr_e(S_j)$, and $SigZscr_e(S_k)$ by scanning successive segments along a nucleotide sequence similar to that *EDscan* did.

(<http://protein3d.ncifcrf.gov/shuyun/rna2d.html>)

E. *SigED* may take considerable heavy computation for a long sequence. It is often used in conjunction with *EDscan* and works in a particular region that was previously determined to be of interest by *EDscan*.

WFS with high statistical significance

mir-42, mir-43 and mir-44

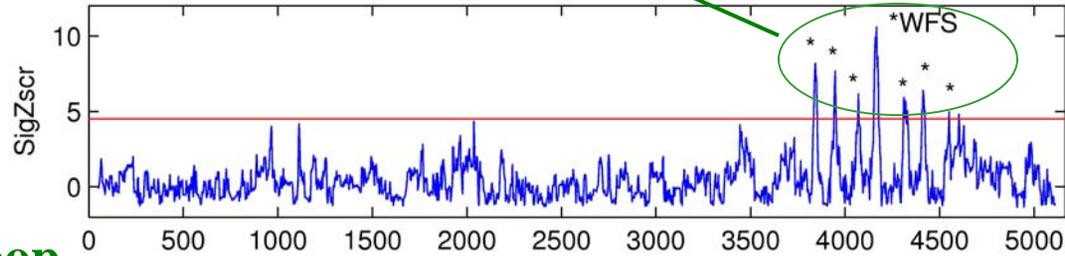


Noncoding region between ZK930.6 and ZK930.2 of chromosome II

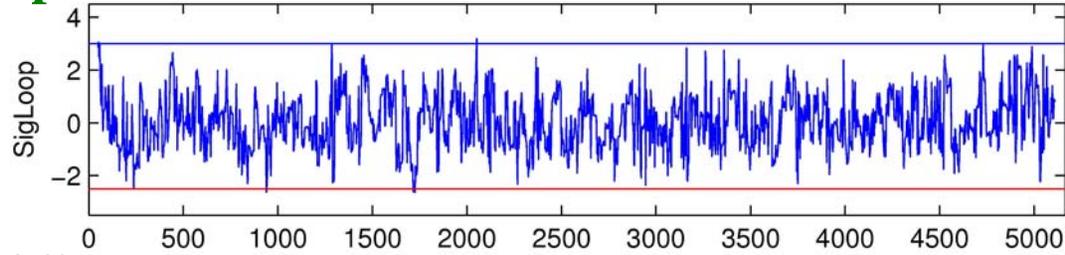
mir-35, mir-36, mir-37, mir-38, mir-39, mir-40 and mir-41

WFS with high statistical significance

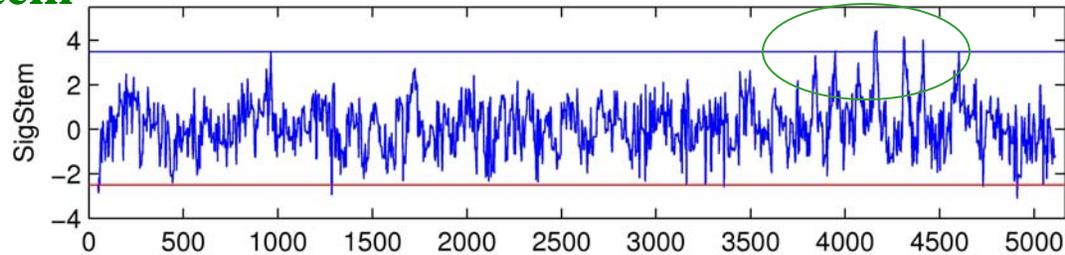
SigZscr



SigLoop



SigStem



Sequence Position of *C. elegans* from Y62F5A.1a to F54D5.12 of Chromosome II
Window Size: 95-nt, Step: 3-nt

Statistical analysis of computed E_{diff} from known miRNA precursors and their corresponding randomly shuffled sequences

<i>Genomes</i>	<i>N</i>	<i>Computed from natural seq.</i>		<i>From randomly shuffled seq.</i>	
		<i>SigZscr (std)</i>	E_{diff} (<i>std</i>)	E_{diff} (<i>std</i>)	<i>SigZscr (std)</i>
<i>Human</i>	<i>207</i>	<i>6.11 (2.25)</i>	<i>20.52 (5.89)</i>	<i>3.50 (0.55)</i>	<i>0.0 (1.0)</i>
<i>Mouse</i>	<i>208</i>	<i>5.76 (2.34)</i>	<i>19.14 (6.19)</i>	<i>3.43 (0.52)</i>	<i>0.0 (1.0)</i>
<i>Rat</i>	<i>187</i>	<i>5.93 (2.15)</i>	<i>20.40 (6.13)</i>	<i>3.54 (0.46)</i>	<i>0.0 (1.0)</i>
<i>G. Gal</i>	<i>121</i>	<i>6.13 (1.77)</i>	<i>19.81 (5.02)</i>	<i>3.30 (0.36)</i>	<i>0.0 (1.0)</i>
<i>Fly</i>	<i>78</i>	<i>5.80 (1.73)</i>	<i>18.22 (4.68)</i>	<i>3.14 (0.41)</i>	<i>0.0 (1.0)</i>
<i>C.eleg</i>	<i>116</i>	<i>6.09 (2.40)</i>	<i>20.15 (7.03)</i>	<i>3.36 (0.41)</i>	<i>0.0 (1.0)</i>
<i>C.brig</i>	<i>50</i>	<i>6.77 (2.30)</i>	<i>21.84 (6.03)</i>	<i>3.36 (0.34)</i>	<i>0.0 (1.0)</i>
<i>A. tha</i>	<i>92</i>	<i>8.48 (2.17)</i>	<i>30.54 (8.22)</i>	<i>3.99 (0.80)</i>	<i>0.0 (1.0)</i>
<i>O. sat</i>	<i>122</i>	<i>7.72 (2.61)</i>	<i>30.35 (9.89)</i>	<i>4.35 (0.66)</i>	<i>0.0 (1.0)</i>

Unit of E_{diff} : kcal/mol

The Reported miRNAs and the Corresponding WFS in *C.elegans*

Gene	Corresponding Well-ordered Folding Sequences (WFS)	SigZscr
lin-4	U UCCUGAGACCUC AAGUGUGAGUGUACUAUUGAUGCUUCACACCUGGGCUCUC	3.34
let-7	UGUGGAUCCGG UGAGGUAGUAGGUUGUAUAGUU UGGAAUUAUACCACCGGUAACUAUGCAAUUUUCUACCUUACCGGAG	8.35
mir-1	GUGACCGUACCGAGCUGCAUACUUCUUAUCAUGCCCAUACUAUAUCAUAAAUGGAUA UGGAAUGUAAAAGAAGUAUGUA GAACGGGGUGGU	7.01
mir-2	CAUCAAGCGGUGGUUGAUGUGUUGCAAUUUAUGACUUUCA UAUCAACGCCAGCUUUGAUG	3.85
mir-34	AG AGGCAGUGUGGUUAGCUGGUUG CAUAUUUCUUGACAACCGGUACCUUACUGCCACCCGAACAUGUCGUCC	5.11
mir-35	UCAGAUCGAGCCAUUGCUGGUUUUCUCCAGUGGUACUUUCCAUAUAGAACUA UCACCGGGUGGAAAACUAGCAGU GGCUCGAUCUUUUC	9.49
mir-36	GUCGGGGAACCGCGCCAAUUUUCGCUUCAGUGCUAGACCAUCAAAGUGUCUA UCACCGGGUGAAAACUUCGCAUG GGUCCCGAC	7.36
mir-37	CCCUUGGACCAGUGUGGGUGUCCGUUGCGGUGCUACAUCUCUAAUCUGUA UCACCGGGUGAACACUUGCAGU GGUCCUC	5.55
mir-38	AGGUCUUGUCCGGUUUUUCCGUGGGUUAACGCAUCCAAAAGUCUCUA UCACCGGGAGAAAACUGGAGU AGGACCG	10.01
mir-39	GAGAGCCCAGCUGAUUUCGUCUUGGUAUAAGCUCGUCUAUUGAGAUUA UCACCGGGUGUAAAUCAGCUUG GCUCUGGUGU	6.97
mir-40	CCGCACCUCAGUGGAUGUAUGCCAUGAUGAUAAGAUUAUCAGAAAUCUA UCACCGGGUGUACAUCAGCUAA GGUGCGGGU	6.85
mir-41	UCCAGAGACCUCUGGUGUUUUCUGCAGUGAUAGAUACUUCUAACAUCUCGCUA UCACCGGGUGAAAACUACCUA GGUCUGGAGCC	3.45
mir-42	GGACUUUGUGGGUGUUUGCUUUUUCGGUGAUGUUCUCCAGUUCUUCUU CACCGGGUAAAACUACAG AGGUCCAAAAGGGG	7.80
mir-43	GCCCGUGACAUCAAGAAACUAGUGAUUAUGCCAAACACAGGGACA UAUCACAGUUUACUUGCUGUCGC GGGCGG	10.39
mir-44	GGCCAAUCUGGAUGUGCUCGUUGGUAUAGACGUAACACGAAACUGUUAUA UGACUAGAGACACAUCAGCU UGGCCUG	6.76
mir-45	GUGCCACGUGGAUGUCUCGUUAGUAUAUAUGUCCCAAGCAAGGACUA UGACUAGAGACACAUCAGCU UGGCC	9.80
mir-46	GCUGAAGAGAGCCGUCUAUUGACAGUUAAGACCACGAGUCGUUGUGUG UGUCAUGGAGUCGCUCUCUUCAGAU	5.55
mir-47	AAACUGAAGAGAGCAGUCUAUUGACAGUCGGUUAUCGAAAUCUUU UGUCAUGGAGGGCCUCUCUUCAGUA	7.88
mir-48	aactctgggaatgagagctaggtggtggtgatgagatccggttcaat TGCGATCTACTGAGCCTACCTCA agttcccgggagtt (antisense)	8.17
mir-49	AAAAGACCACCGUCCGAGUUUGUUGAUGUGUCCCAAGCAAAUCAUGAGUCU UAGCACCACGAGAGCUGCAGA UGGAGGUUC	3.86
mir-50	UGCCCGCCGGCCG CUGAUUAGUCUGGUUUUCUUGGGUU UGAACUUCAGCGUUGAACCCGCAUUAUAGACGUAUCGACGGCCGGCGGGG	10.32
mir-51	CGUC UACCCGUAGCUCCUAUCCAUGUUA ACUGGUCAAAAAGUGAAACUUGGAAGCAGGUACA	3.84
mir-52	UCCAACUCUAACAGUC CACCCGUACAUAUGUUUCCGUGCU UGACAGCGAAGCUCAAUCACGUAUCAAUGAAAGGGUAGCCGGUUAUUGAAGUUG	3.24
mir-53	ACCCGUACAUAUUUGUUUCCGUGCU UGACUUCAAAAGCUCAAUCACGGCACAAUUAUUGGGUC	3.55
mir-54	CGCUCUGACUAGGAUAUGAGACGACGAGAACAUGUCUUUUUAAAAGACUUG UACCCGUAAUCUUAUAAUCCGAGUC AGGGCUAGCUGA	5.49
mir-55	GGGACUCGGCAGAAACCUAUCGGUUAUACUUUUUGGAUAUGCUA UACCCGUAAUAGUUUCUGCUGAG CCCCUUAU	7.95
mir-56	CUGUUCU UGGCGGAUCCAUUUUGGGUUGUA CCUCAUCCUAAAUUUGACGGUACCCGUAAUGUUUCCGUGAGAACCGACU	7.51
mir-57	UACCCUGUAGAUCCGAGCUGUGU UUGAAACAAUCAUACAGGAGCUAGACUACAAGGUGCACGAACAACCGAA	4.39
mir-58	CAUAUCCAUUGCCUACUCUUCGCAUCUCAUCACUUCGUCCAAUACCAUAGGGA UGAGAUCGUUCAGUACGGCAA UGGAC	5.52
mir-59	UAUGACAUCGUCCUGAAAACGAAACGGAACAAAAGUUAAGAUUAUGAU UUGCAAUCGUUUUACAGGAUGAUG	5.63
mir-60	UCUUGAACUGGAAGAGUGCCAUAUAAUUAUGACAAAGUACGUGA UAUUUAGCACAUUUUUCUAGUUCAGACUUGA	10.43
mir-61	UAUCGCUGAACCCUCGAGAUGGGUUAUCGGGGCUUAGUCCUUCUCCGUAUGGCAA UGACUAGAACCUGUACUCAUCUC GAGGUUUCGGUGA	8.11
mir-62	GGUGAGUAGAUUCUAUACCUUCCGCAAAAUGGAAU UGAUUAGUAAUCUAGCUUACAGG	5.18
mir-63	GACACAAUUUCUAACUCGUCGUGAGUCAUCGUUCUAGCUGAAAAGGACAC UAUGACUAGAAAGCGAGUUGGAAA UAGUGGUUCUA	8.24
mir-64	CGCCGAA UAUGACACUGAAGCGUUACCGAA CCGUUUUCCACACCUGGAUUCGGUGCAACGAUCAGUGGCAUGCUCGGCU	5.70
mir-65	AUGGAGCCUUCGCCGAU UAUGACACUGAAGCGUAACCGAA CACCAUUAUUUUGAGAUUCUG.. (25 nt)..GUUGCCUAAUAAA	4.09
mir-66	CCACAAAAGGCCAU CAUGACACUGAUUAGGGAUGGUA UGAAUGUUAAGAUCGCCGAUCA.. (20 nt)..AUGGCGUAGUGGUU	7.20
mir-67	GUCGAUCCGCUCAUUCUGCCGGUUGUUAUGCUAUAUACAGAUUAAGCA UCACAACCUCUAGAAAAGAGUAGA UCGAUUU	9.96
mir-68	UUUUGAAAUAUUAUUUUCUGAAUUUACACUUCAGUUAGUUAUUAACGUUUUUAUUAAGGAUGGUUAU UCGAAGACUCAAAAGUGUAGAC	3.06
mir-69	UUAAUUAAAUUUUUUUUAAUUUUUACACGGGUUAUUAAGUAUAU UGAAAUAUAAAAGUGUAGACAU	3.33
mir-70	UCAAUUAAA.. (25 nt)..CGACGAAUAACACUUAUGAAGAAAUG UAUACGUCGUUGGUUUUCCAU AGUUUGAAUUGUUUAU	4.15
mir-71	CUGCUCUGAACGA UGAAAGACAUUGGUAGUGA GACGUCGGAGCCUCGUCGUAUCACUAUUCUGUUUUUCGCCGUCGGGAU	4.57
mir-72	GGUCCCGUCAGACU AGGCAAGAUUGGCAUAGCUG AAUAGUACGCUAUAACAACUAUCAGCUUCGCCACAUUCGCCACGACUGAUGU	3.69
mir-73	CACACACGACUGGACUCCAUUCGAGCCACAGCUAUCACGAAUUG UGGCAAGAAUGGCAUCUAC CGUUUUUCAUGU	2.77
mir-74	AAAUGGUUCA.. (20 nt)..CUCUUCCAGCCUACAUCUCAACCUGGGC UGGCAAGAAUGGCAUCUAC CGUUUUUCAACCA	6.46
mir-75	UUGCUUUGAAGAAUUGCAGUCGGUUGCAAGCUUAAAUA CAAAUCCGAAUUGUUAUUAAGCUACCAACCGGCUCA AGUCUGAAAAGACCA	4.71

T07D1.2 gene of *C. elegans* (Accessoin No. U41531), *miRNA -81*

WFS: 14211-14290, energy = -35.90, SigZscr = 6.29

CCAACAGtcGGTTTTCAcGTGATCTgAGAGCAatccaaaaTGctttTCTgAGATCATcGTGAAAGCTagTTGTTGGet
 1 14211 14288 7 (mir-81) ----- (22 nt)
 2 14220 14279 9 UC C G AUCC
 3 14230 14269 7 CCAACAG GGUUUUCAC GUGAUCU AGA---GCA A
 4 14238 14261 3 GGUUGGU UCGAAAGUG UACUAGA UCU CGU
 5 14241 14255 3 GA C G UUU AAAA

Region: 566-645, energy = -17.20, Sigstem = -3.72

GGTccCGTTTGcCATgTGAATgtTGGAAgCACATcgaATGtTGCaTTCattGTTcGgATcGTttCAAATGcataACCcat
 1 566 642 3
 2 571 635 6
 3 577 627 2
 4 579 624 2
 5 582 621 5
 6 589 614 4
 7 594 609 3
 8 597 605 3

Region: 3291-3370, energy = -11.00, SigLoop = -3.73

tAGGGTggACTatthttgaTGTGaaTGctggagGTAAaTGCAGaaAAGGAGATcggATTTCTTTaggggacaGGTaACTCT
 1 3292 3370 5
 2 3299 3364 3
 3 3309 3331 4
 4 3315 3325 3
 5 3335 3353 8

Region: 8576-8655, energy = -13.90, Sigstem = 3.66

acataaGTTCCcTCTaTTTTacaacAAGAcAGAcaaaaaagaaGTTTcaCGCGctgctGCGaCGacAAACcaGGAACatt
 1 8582 8652 5
 2 8588 8608 3
 3 8592 8604 4
 4 8619 8645 4
 5 8625 8639 2
 6 8627 8636 3

Y71G12B.11 gene of *C. elegans* (Searched sequence is region 81170-98845
in Y71G12B, Accession No. AC025726), **miRNA-50**

WFS: 17044-17123, energy = -42.30, SigZscr = 7.27
cgCCGGCCGcTGATATGTCTGGTATTctTGGGTTgAACTtccagcGTTGAACCCGcATATTAGACGTATCGaCGGCCGG
----- (mir-50, 24 nt)

1	17046	17123	7	
2	17054	17115	15	CCGGCCG ^C UGAU AUGUCUGGUAU ^{UCU} UGGGUUU ^G AAC ^{UUC}
3	17072	17099	7	GGCCGGC ^A GCUAUGCAGAUUAUA ^C GCCCAAG-UUG ^{CGA}
4	17080	17092	3	

Region: 12139-12218, energy = -13.10, SigLoop = -3.40, SigStem = 3.99
gacGAGtagggCTCcTCCAGgctataattTTGCGCTAGAAcAATTtaaAATTTTCTtttttacGTTGCAACTGGAtataacc

1	12142	12152	3
2	12154	12212	5
3	12168	12207	7
4	12175	12194	4
5	12180	12190	4

Region: 12532-12611, energy = -10.50, SigLoop = 3.19, SigStem = -3.13
tGCGCTtcAGCTGTAGAAattCCCTttaaatctcacttttcataacaaccacAGGGctcatTTTTAtTAGTTttccAGTGC

1	12533	12611	5
2	12540	12602	5
3	12545	12596	5
4	12552	12586	4

Conclusion:

WFSs with high $SigZscr_e$ are coincident with the reported *miRNAs* in *C. elegans*, *C.brig*, human and other genomes.

With the combination of the two methods *EDscan/StemEd* and *SigED/SigStem* we can do the computational experiment in a large scale.

SigED can also offer a means of searching for unusual unstable folding regions and distinct loop structures in RNAs.

Statistical extremes with high $Zscr_e$ and $SigZscr_e$ represent significant folding segments where predicted structures are expected to be well-ordered. *EDscan/StemEd* and *SigED/SigStem* provide us a new tool for the discovery of potential functional elements with structure dependent functions.

With the additional information, such as miRNA phylogenetic conservation, the method can be used to search for ncRNAs in genome.

Detecting ncRNAs of known structured RNAs in genomes

HomoStRscan: A novel algorithm to search for homologous structured RNAs by scanning a genomic sequence.

It uses information of both the primary sequence and the secondary structural constraints to the query RNA in detail.

It provides two types of scoring matrices:

- a. Base matching score for 4 types of bases, A, C, G, U
- b. Base-pairing matching score for 16 types of base-pairings, with including non-canonical base pairs, A:G, G:G,

It computes MSS between the query RNA and the computed matching structure inferred in a flexible window.

Homologous structured RNAs are detected by robust statistical inference from the MSS distribution computed by moving the window along genome.

The two profiles of 5Srna+ and 5Srna- of 5S rRNA queries used in HomoStRscan

a. The profile 5Srna+

5Srna+_B.sub_matured

```
TTTGGTGGCG auaGCGAAGA GgtcacACCC GTtcccatac
cgaacACGGa aGTtaagCTC TTCaGCgCCG ATGGTAGTcG
GGGGtttCCC CctGTGAGAG TAGGaCGCCG CCAAGc
```

>

>

(1)	1	115	10
(2)	14	66	2
(3)	16	63	6
(4)	27	53	2
(5)	29	49	4
(6)	68	104	11
(7)	80	92	5

>

b. The profile 5Srna-

5Srna-_B.sub_matured_RCS

```
gCTTGGCGGC GtCCTACTCT CACaGGGGGa aaCCCCCgAC
TACCATCGGc GctGAAGAGc ttaACTtCCG Tgttcggtat
gggaACGGGT gtgacCTCTT CGCtatCGCC ACCAAA
```

>

>

(1)	2	116	10
(2)	13	49	11
(3)	25	37	5
(4)	51	103	2
(5)	54	101	6
(6)	64	90	2
(7)	68	88	4

>

Table 3. The predicted 5S rRNAs in the genome of E.coli k12

rRNAs Location	Product	Predicted	MSS	
			+	-
		MSS > mean+6*std	500	500
228756..228875	+ 5S rRNA	228756..228875	555	
2724089..2724208	- 5S rRNA	2724089..2724208		555
3421059..3421179	- 5S rRNA	3421059..3421179		550
3421305..3421424	- 5S rRNA	3421305..3421424		555
3944324..3944443	+ 5S rRNA	3944324..3944443	555	
4038097..4038216	+ 5S rRNA	4038097..4038216	553	
4169216..4169335	+ 5S rRNA	4169216..4169335	555	
4210619..4210738	+ 5S rRNA	4210619..4210738	555	
Total number of observations with MSS > 500			5	
Total number of observations with MSS > 500				3

Table 4. The predicted 5S rRNAs in the genome of *Staphylococcus aureus* subsp. *aureus* MW2

rRNAs Location	Product	Prodicated	MSS	
			+	-
		MSS > mean+6*std	508	509
496386..496500	+ 5S rRNA	496386..496500	546	
534799..534913	+ 5S rRNA	534799..534913	546	
540641..540755	+ 5S rRNA	540641..540755	546	
545852..545966	+ 5S rRNA	545852..545966	546	
1959247..1959361	- 5S rRNA	1959247..1959361		546
2137179..2137293	- 5S rRNA	2137179..2137293		546
2250748..2250862	- 5S rRNA	2250748..2250862		546
Total number of observations with MSS > 508			4	
Total number of observations with MSS > 509				3

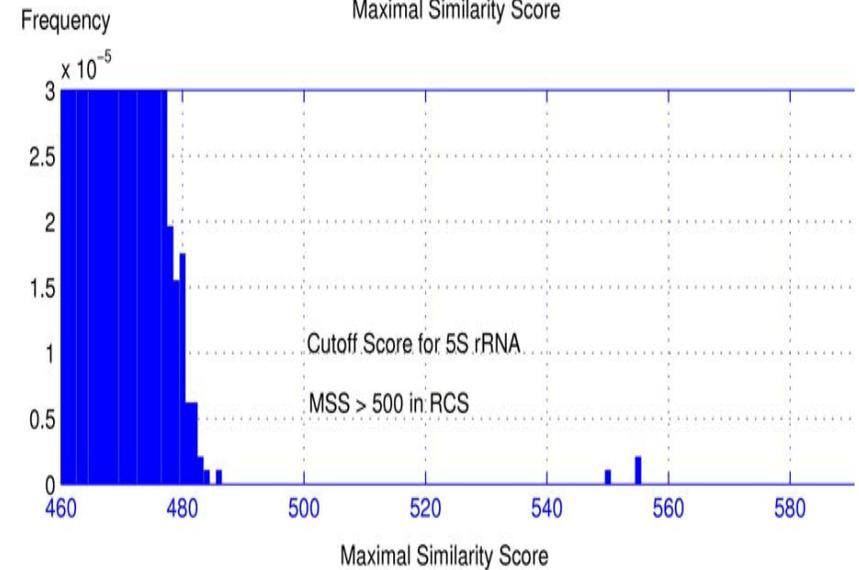
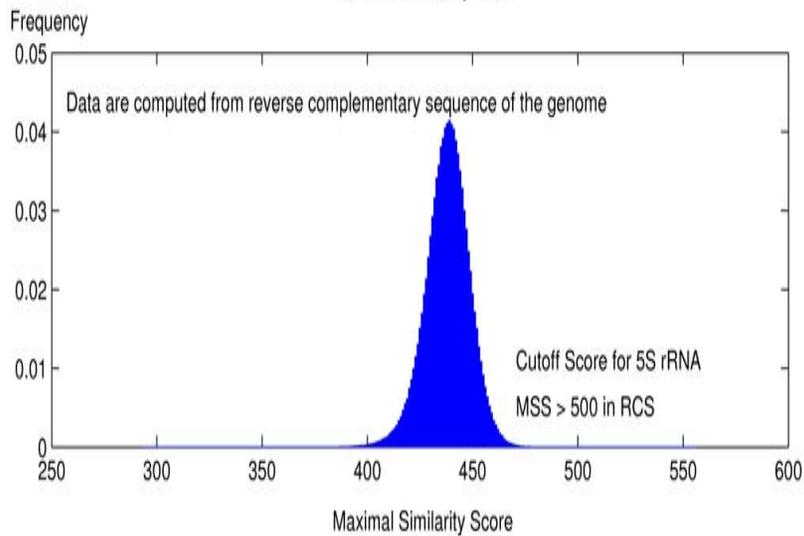
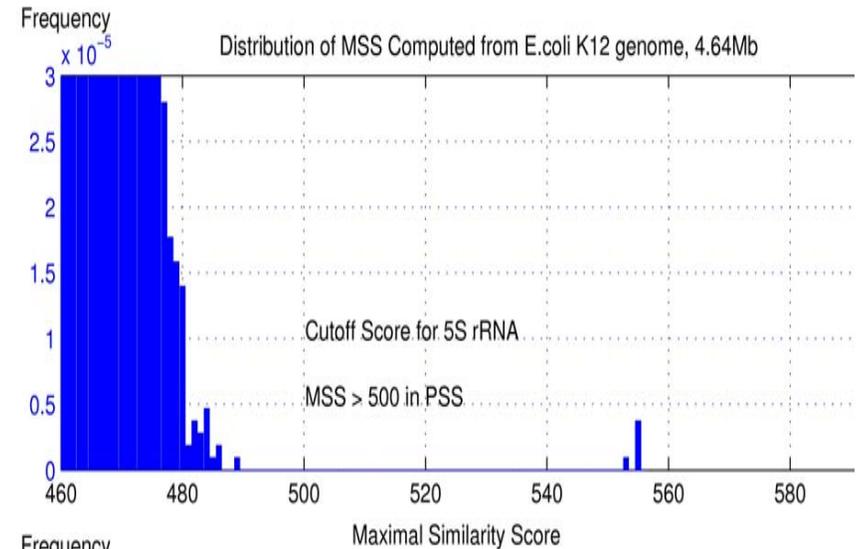
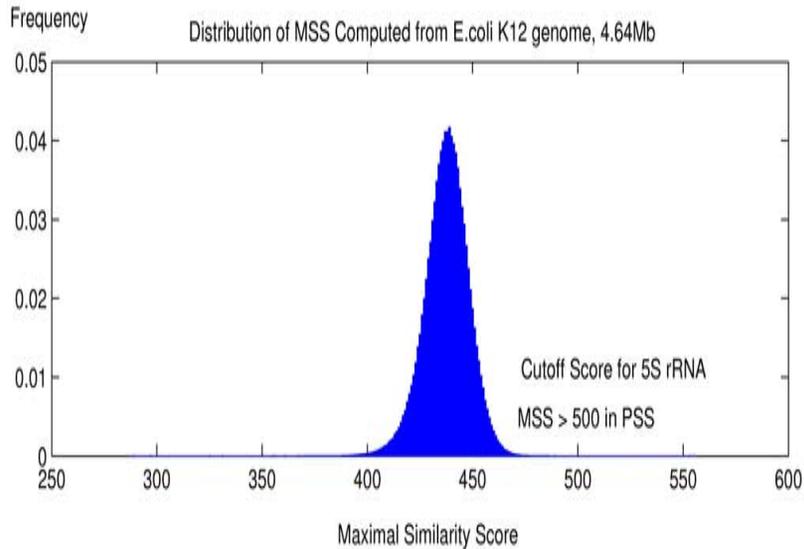
Table 5. The predicted 5S rRNAs in the genome of *Streptococcus agalactiae* 2603V/R by HomoStRscan

rRNAs Location	Product	MSS > mean+6*std	505 +	505 -
16411..17917	+ 16S rRNA			
18234..21136	+ 23S rRNA			
Predicted	+ 5S rRNA	21211..21326	539	
22242..23748	+ 16S rRNA			
24065..26967	+ 23S rRNA			
Predicted	+ 5S rRNA	27042..27157	539	
91219..92725	+ 16S rRNA			
93042..95944	+ 23S rRNA			
Predicted	+ 5S rRNA	96019..96134	539	
165248..166754	+ 16S rRNA			
167071..169973	+ 23S rRNA			
170027..170188	+ 5S rRNA	170048..170163	539	
250375..251881	+ 16S rRNA			
252198..255100	+ 23S rRNA			
Predicted	+ 5S rRNA	255175..255290	539	
348582..350088	+ 16S rRNA			
350405..353307	+ 23S rRNA			
353361..353522	+ 5S rRNA	353382..353497	539	
417726..419232	+ 16S rRNA			
419549..422451	+ 23S rRNA			
Predicted	+ 5S rRNA	422526..422641	539	
Total number of observations with MSS > 505			7	
Total number of observations with MSS > 505				0

5S rRNA MSS distribution computed from *E. coli* K12

MSS computed from positive stranded sequence are shown in the top

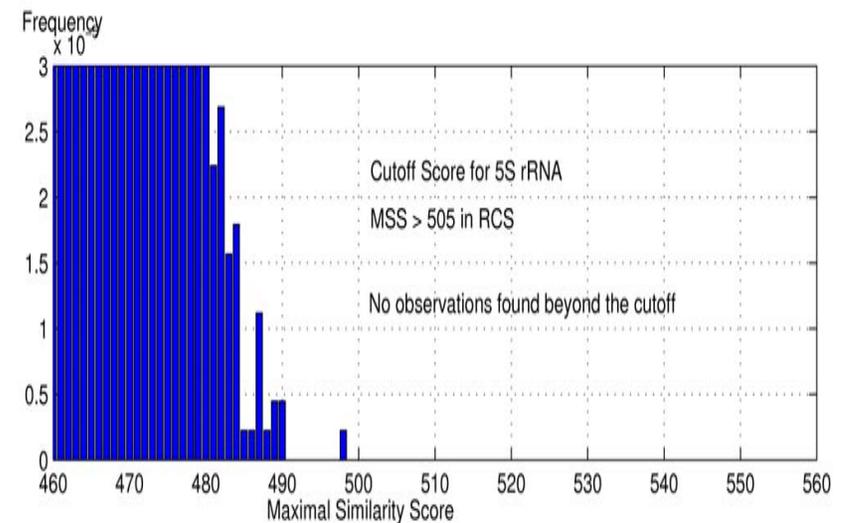
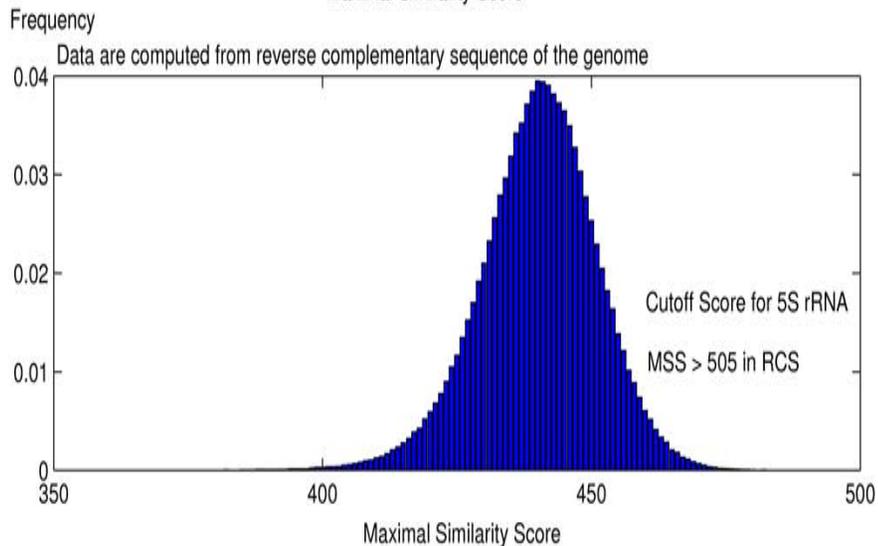
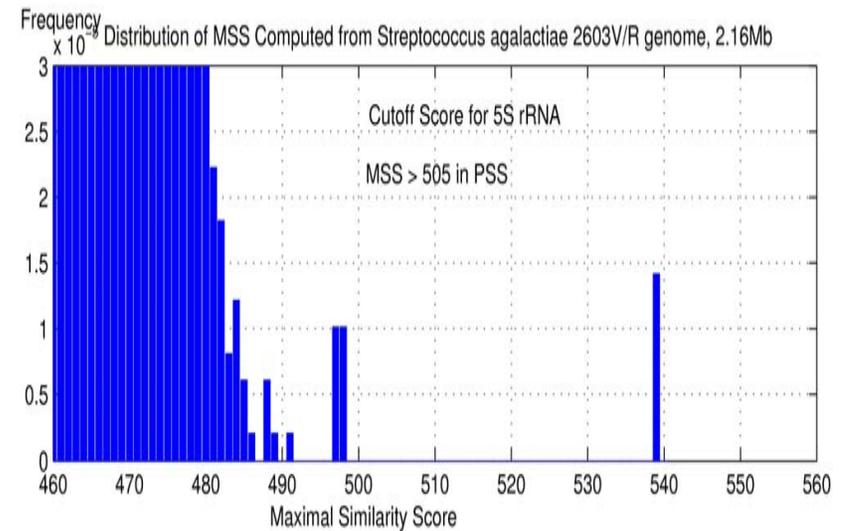
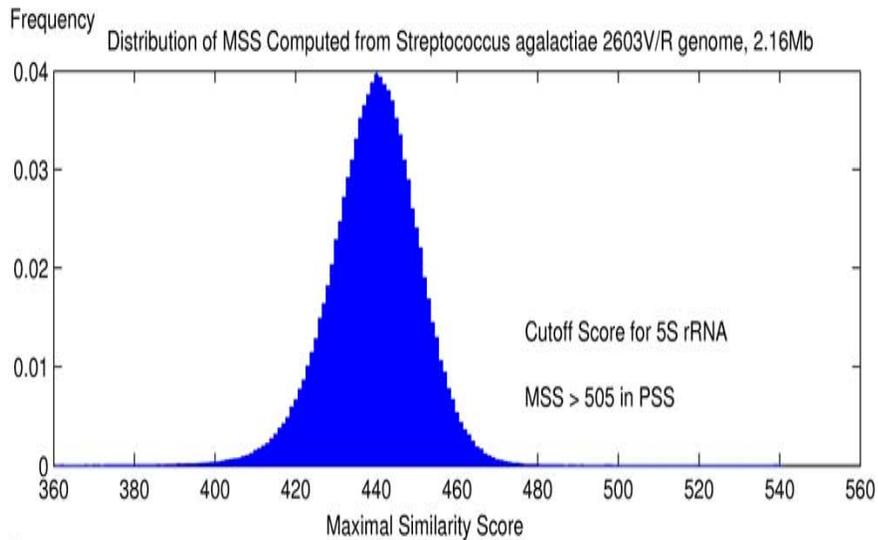
MSS computed from reverse complementary sequence are shown in the bottom.



5S rRNA MSS distributions computed from *S. agalactiae* (strain 2603 V/R)

MSS computed from positive stranded sequence are shown in the top

MSS computed from reverse complementary sequence are shown in the bottom.



E.coli (K12) NC_000913 (Original, U00096), 4.64 Mbp							
		mean + 5*std =		300	319	300	319
		MSS >		301	322	301	322
Transfer RNAs	Location	Product		72+	85+	72-	85-
225381..225454	+	tRNA-Ile		315			
225500..225572	+	tRNA-Asx		324			
228928..229001	+	tRNA-Val		323			
236931..237004	+	tRNA-Sec		323			
262095..262167	+	tRNA-Thr		312			
563946..564019	+	tRNA-Arg		330			
695656..695727	-	tRNA-Gln				344	
695768..695839	-	tRNA-Gln				344	
695890..695963	-	tRNA-Met				314	
695982..696053	-	tRNA-Gln				345	
696091..696162	-	tRNA-Gln				345	
696189..696270	-	tRNA-Leu				316	330
696283..696356	-	tRNA-Met				314	
779777..779849	+	tRNA-Lys	331				
779988..780060	+	tRNA-Val	322				
780066..780138	+	tRNA-Lys	331				
780291..780363	+	tRNA-Val	322				
780370..780442	+	tRNA-Lys	331				
780592..780664	+	tRNA-Lys	331				
780800..780872	+	tRNA-Lys	331				
925110..925194	-	tRNA-Ser				312	358
1030851..1030935	-	tRNA-Ser				316	356
1096791..1096875	-	tRNA-Ser				312	358
1286470..1286551	-	tRNA-Tyr				318	343
1286764..1286845	-	tRNA-Tyr				318	343
1744459..1744532	+	tRNA-Val	322				
1744540..1744613	+	tRNA-Val	322				
1989841..1989924	-	tRNA-Leu				308	344
1989940..1990010	-	tRNA-Cys				325	
1990068..1990140	-	tRNA-Gly				324	
2041493..2041579	-	tRNA-Ser				312	359
2042571..2042643	+	tRNA-Asn	327				
2056052..2056124	-	tRNA-Asn				327	
2057873..2057945	+	tRNA-Asn	327				
2060282..2060354	+	tRNA-Asn	327				
2284231..2284304	+	tRNA-Pro	315				
2464329..2464400	+	tRNA-Arg	319				
2516064..2516136	-	tRNA-Ala				320	
2516179..2516251	-	tRNA-Ala				320	
2518951..2519023	+	tRNA-Val	322				
2519071..2519143	+	tRNA-Val	322				
2519193..2519265	+	tRNA-Val	322				
2519273..2519345	+	tRNA-Lys	331				
2727392..2727464	-	tRNA-Sec				326	
2783785..2783857	-	tRNA-Ile				326	
2815809..2815882	-	tRNA-Arg				330	

Transfer RNAs	Location	Product	72+	85+	72-	85-
2816084..2816157	-	tRNA-Arg			330	
2816223..2816296	-	tRNA-Arg			330	
2816498..2816571	-	tRNA-Arg			330	
2816578..2816667	-	tRNA-Ser			309	349
2945409..2945482	+	tRNA-Met	314			
2945519..2945592	+	tRNA-Met	316			
2945629..2945702	+	tRNA-Met	316			
2997009..2997079	-	tRNA-Gly			325	
3108383..3108455	+	tRNA-Phe	324			
3213239..3213311	+	tRNA-Ile	326			
3315857..3315930	-	tRNA-Met			314	
3319716..3319799	-	tRNA-Leu			302	333
3421220..3421292	-	tRNA-Thr			323	
3424598..3424670	-	tRNA-Asx			324	
3424716..3424789	-	tRNA-Ile			315	
3706248..3706321	-	tRNA-Pro			326	
3833850..3833939	+	tRNA-OTHER		328		
3941057..3941129	+	tRNA-Val	326			
3944496..3944569	+	tRNA-Sec	323			
3944581..3944653	+	tRNA-Trp	322			
3979988..3980061	+	tRNA-Arg	320			
3980122..3980195	+	tRNA-His	327			
3980219..3980302	+	tRNA-Leu		327		
3980348..3980421	+	tRNA-Pro	328			
4034730..4034803	+	tRNA-Ile	315			
4034849..4034921	+	tRNA-Asx	324			
4165951..4166023	+	tRNA-Val	326			
4172967..4173039	+	tRNA-Thr	318			
4173051..4173132	+	tRNA-Tyr	318	342		
4173252..4173323	+	tRNA-Gly	315			
4173333..4173405	+	tRNA-Thr	326			
4207352..4207424	+	tRNA-Sec	326			
4360132..4360204	-	tRNA-Phe			324	
4389938..4390010	+	tRNA-Gly	324			
4390050..4390122	+	tRNA-Gly	324			
4390161..4390233	+	tRNA-Gly	324			
4493973..4494054	+	tRNA-Leu	316	339		
4603651..4603734	-	tRNA-Leu				327
4603772..4603855	-	tRNA-Leu				327
4603887..4603970	-	tRNA-Leu				327
Predicted new candidates for tRNA genes						
		MSS >	301	322	301	322
344558..344635	+	tRNA-Trp	304			
649910..649984	-	tRNA-Ser			304	
1785332..1785407	-	tRNA-Glu			305	
3264993..3265080	-	tRNA-Ile				323

Acknowledgement

Jacob V. Maizel, Jr.

CCR, NCI, NIH

Jih-H. Chen :

**Advanced Biomedical Computing
Center, SAIC-NCI/FCRDC**

Kaizhong Zhang:

Univ. of Western Ontario

Orna Elroy-Stein

Tel Aviv Univ., Israel

Yoram Groner

Weizmann Institute of Science

Ben-Zion Levi

**Dept. of Food Engineering and
Biotechnology, Technion, Israel**