Biologia Molecular Computacional
IBI5035/QBQ2507 - 2023

Estrutura e função de RNAs

Eduardo Moraes Rego Reis
Instituto de Química - USP

Os níveis de organização da estrutura do RNA

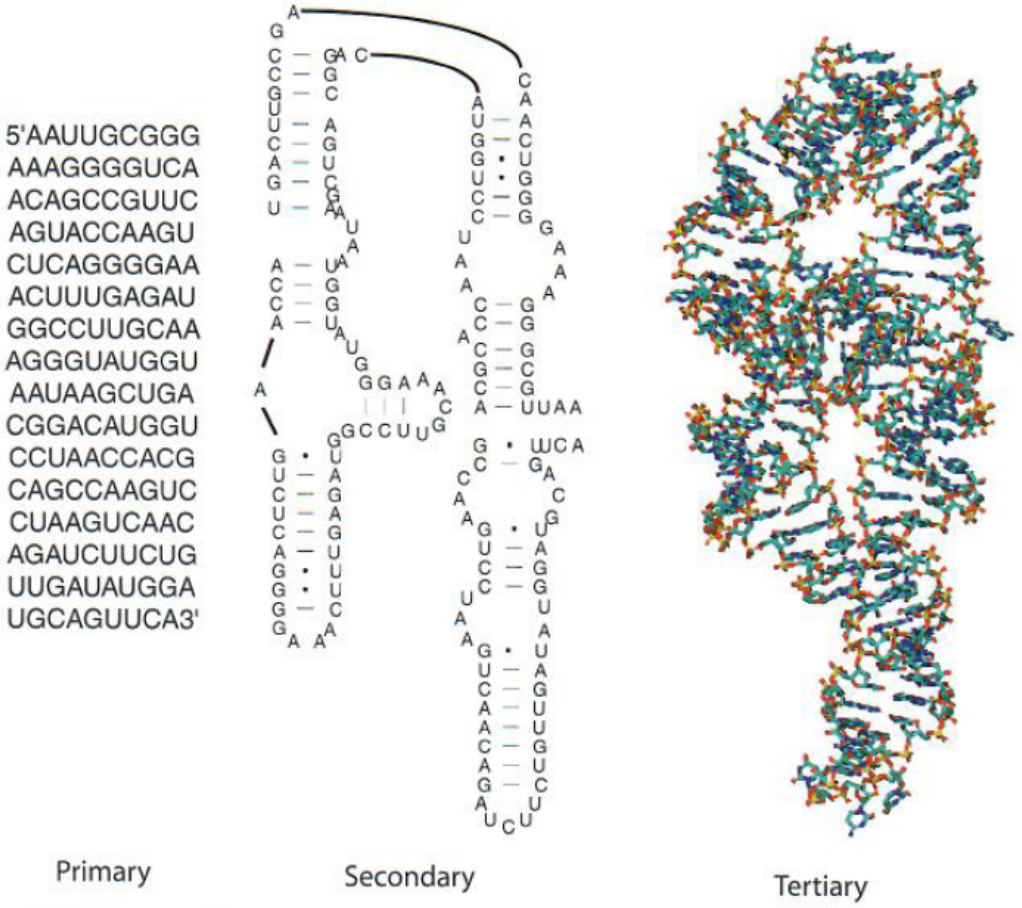
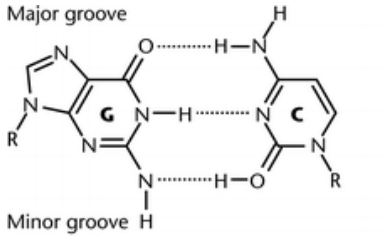
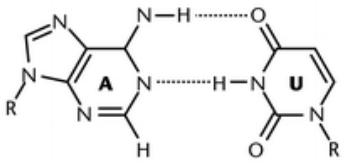


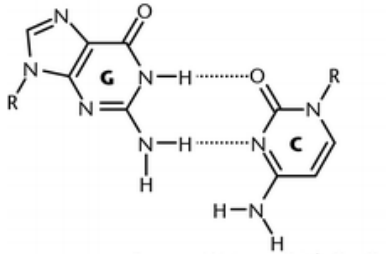
FIGURE 6.1 **The three levels of organization of RNA structure. From left to right are the primary sequence, the secondary structure (Cannone et al., 2002), and the tertiary structure (Cate et al., 1996) of a domain of the group I intron from *Tetrahymena*. The secondary structure illustrates the canonical base pairs, and the tertiary structure is the actual three-dimensional arrangement of atoms.**

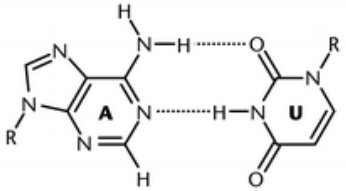
As bases nas **moléculas de RNA** podem fazem **pareamentos canônicos e não canônicos**

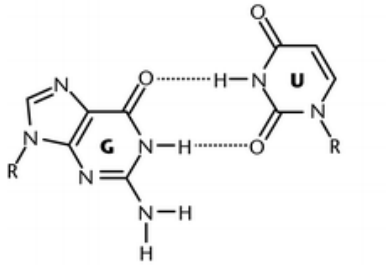


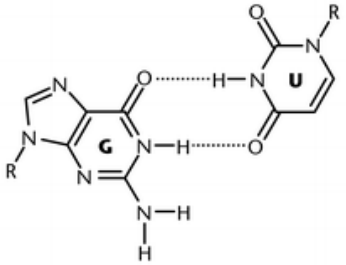
Watson–Crick G • C

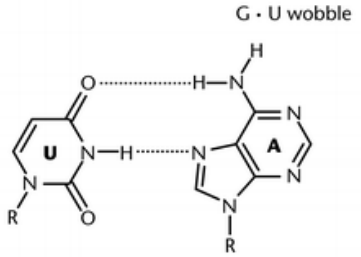

Watson–Crick A • U

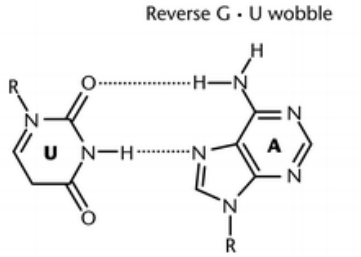

Reverse Watson–Crick G • C

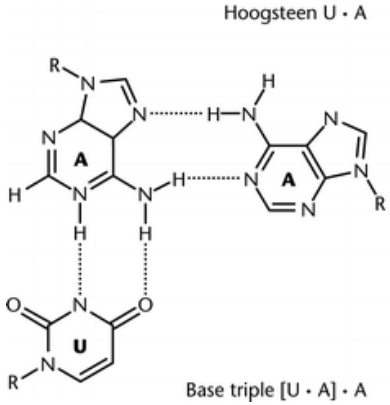

Reverse Watson–Crick A • U

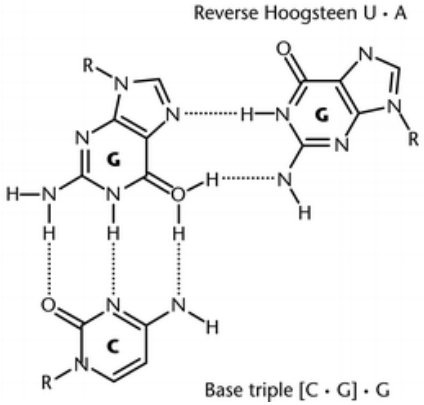

G • U wobble


Reverse G • U wobble


G • U wobble


Reverse G • U wobble

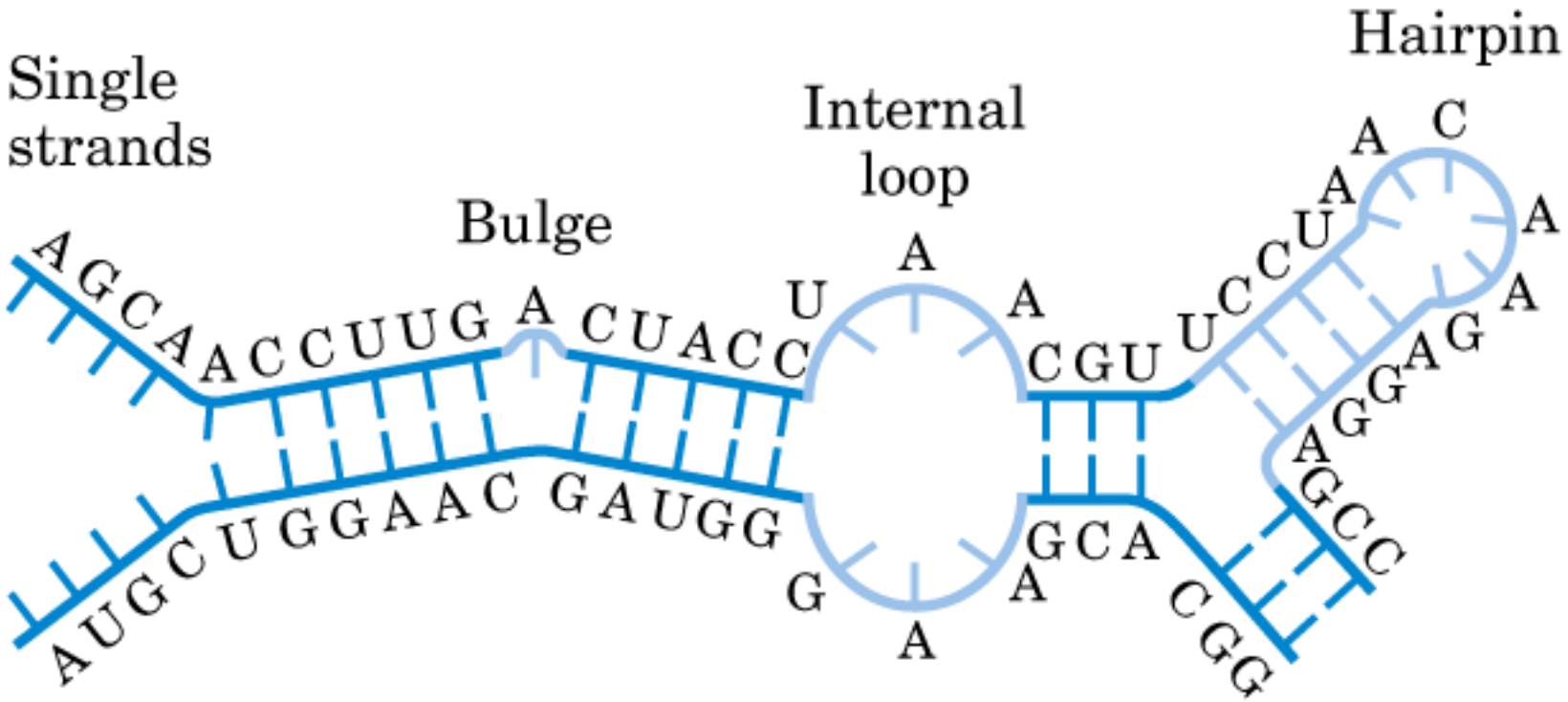

Hoogsteen U • A


Reverse Hoogsteen U • A

Base triple [U • A] • A

Base triple [C • G] • G

Ligações de hidrogênio intramoleculares dão origem a trechos de dupla fita e estruturas secundárias típicas



Estruturas secundárias de RNAs podem fazer interações intra e intermoleculares

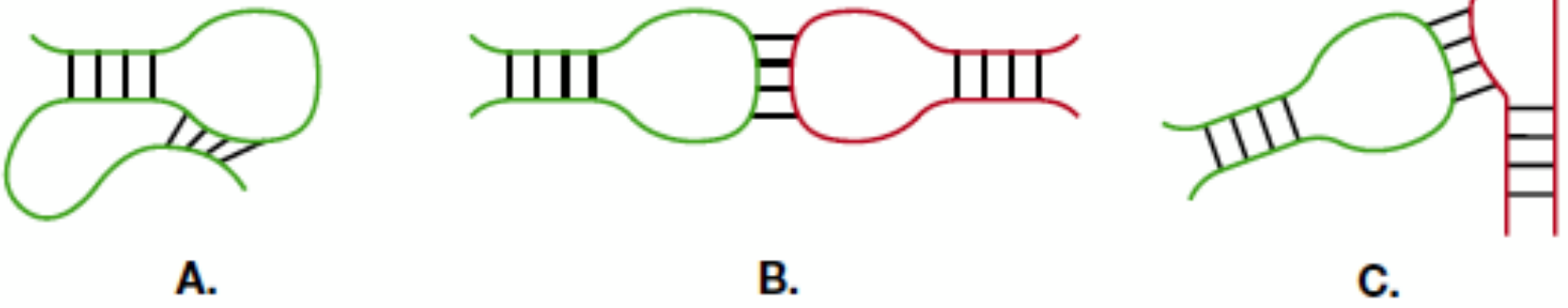


Figure 5.3. Examples of known interactions of RNA secondary structural elements. (*A*) Pseudo-knot. (*B*) Kissing hairpins. (*C*) Hairpin-bulge contact. (Adapted from Burkhard et al. 1999b.)

Qual a relevância de se estudar a estrutura de RNAs?

Estruturas secundárias e terciárias são essenciais para diferentes funções exercidas pelos RNAs

Estrutura secundária do 3'UTR do retrotransposon R2.

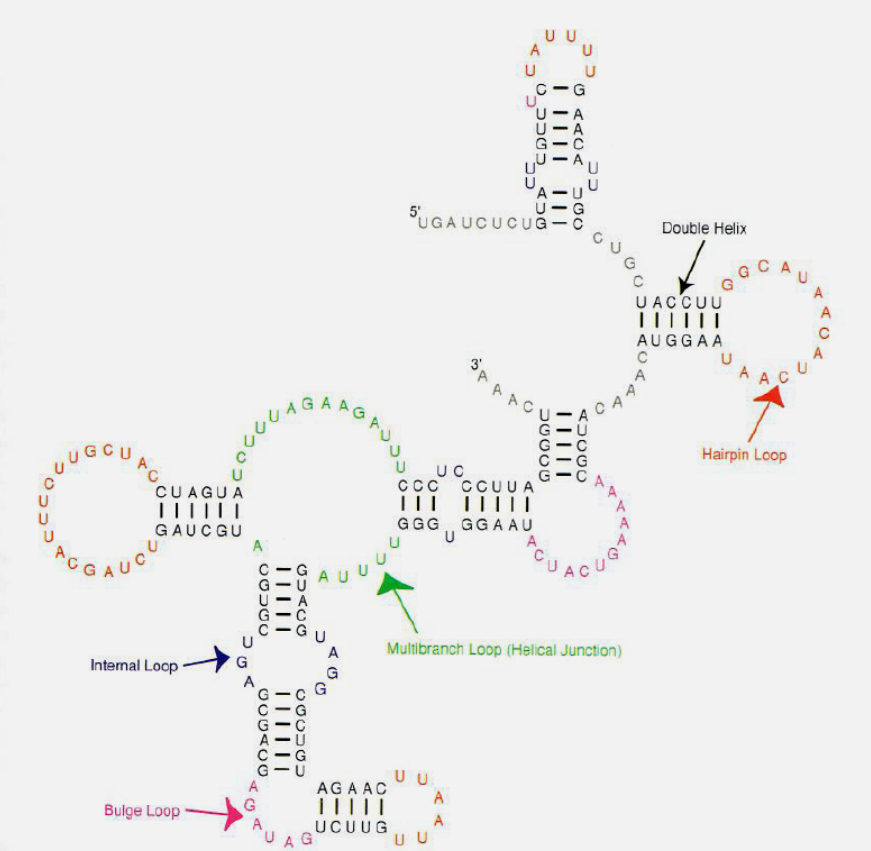
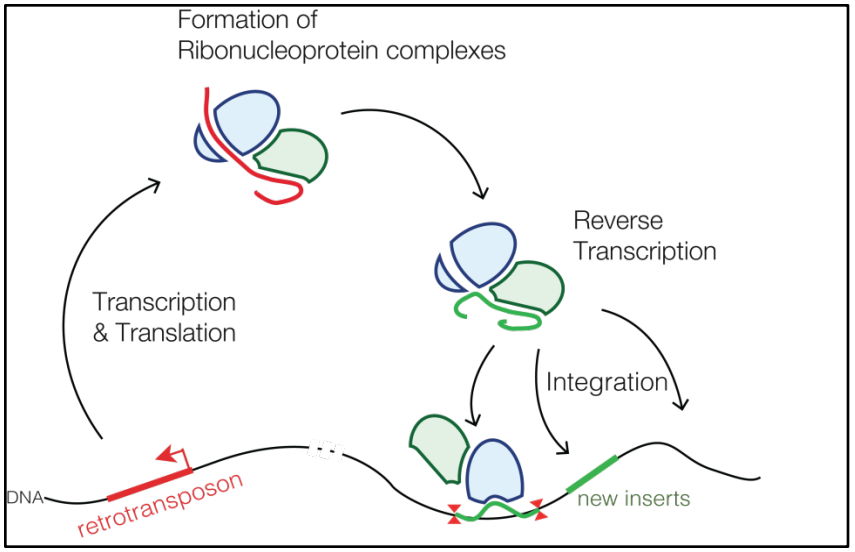


FIGURE 6.2 The RNA secondary structure of the 3′ UTR from the *D. sucinea* R2 element (Lathe & Eickbush, 1997; Mathews et al., 1997). Base pairs in nonhelical regions, known as loops, are colored by type of loop.



A estrutura secundária estável tem função de atuar como primer para transcrição reversa do retroelemento

Estrutura de RNAs e o controle da expressão gênica

Table 1 | **Examples of the diverse roles of RNA structure in gene expression**

RNA type	Examples	Example roles of RNA structures	Refs
Transcription			
Long and short ncRNAs	*Xist*, *HOTAIR*, *ANRIL*, promoter associated RNAs	Double stem–loop and other structural motifs recruit Polycomb complex for gene silencing (mammals)	111–115
Mitochondrial RNA		G-quadruplex structures cause transcription termination (mammals)	116
Riboswitch	Adenine, guanine, lysine, glycine, T box, TPP, SAM, pre-Q1	Structure change upon ligand binding results in either transcription termination or activation (bacteria)	5,117–122
Splicing			
mRNAs	Tau, cardiac troponin	Protein binding to stem–loop regulates alternative splicing (mammals)	123,124
	CD59, *XBP1*	IRE1α recognizes stem–loop for splicing (mammals)	125
	14-3-3ξ	Inter-intronic RNA pairing results in mutually exclusive splicing (*Drosophila melanogaster*)	126
Riboswitch	Group I ribozyme, TPP	Binding to metabolites alters splicing (bacteria, fungi, plants)	127,128

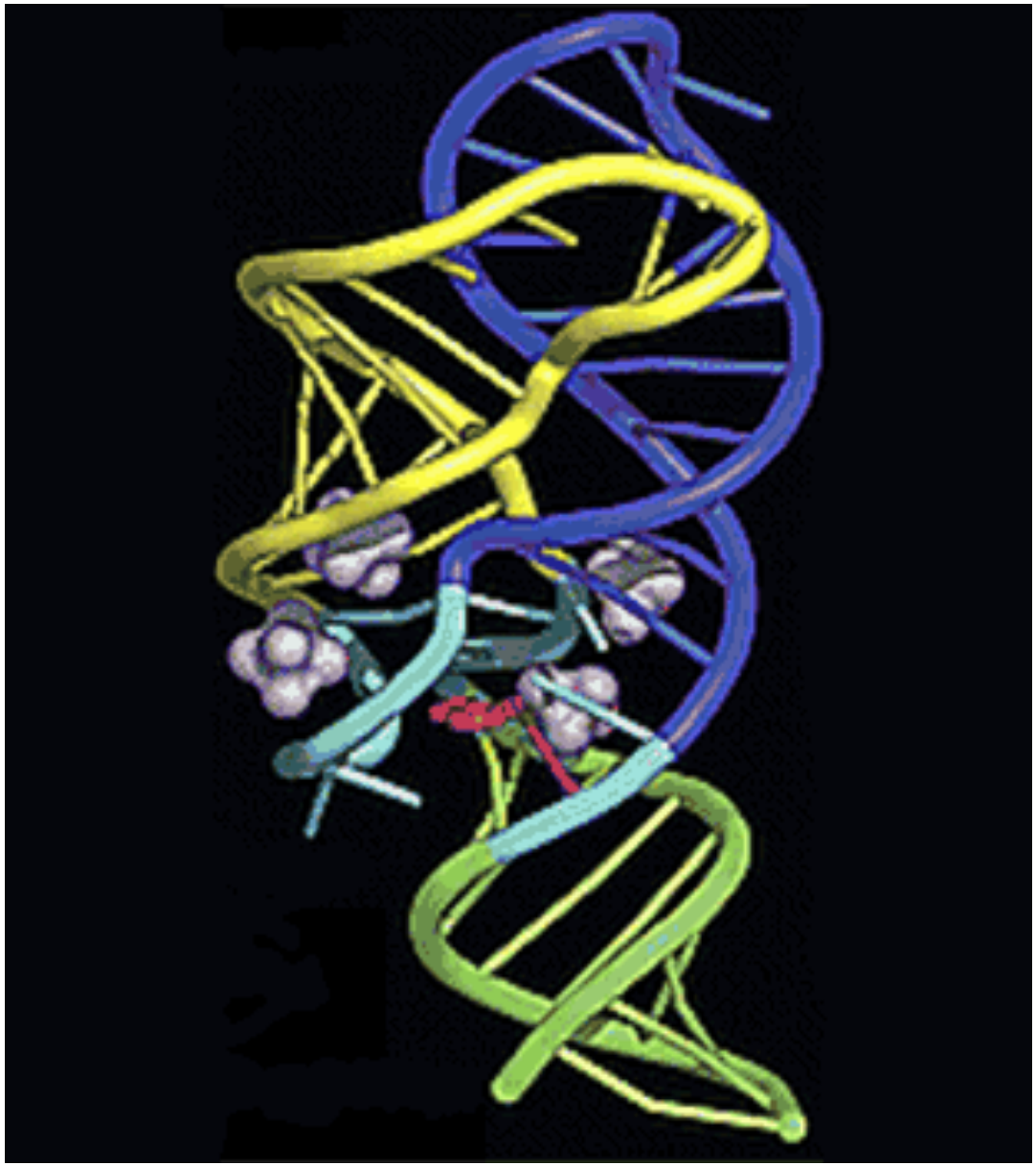
Wan et al., 2011 Nature Review Genetics **12**:641

Table 1 | **Examples of the diverse roles of RNA structure in gene expression**

RNA localization			
mRNAs	*Hac1*	Localization to yeast endoplasmic reticulum membrane	129
	ATP2, *ATM1*	Localization to yeast mitochondria	130
	fs(1)K10	A-form helix causes localization to anterior of *Drosophila* oocyte	131
	PSD95/CaMKIIa	G-quadruplex in 3′UTR targeting to neuritis (mammals)	132
	β-actin	Localizes to the leading edge of fibroblasts or neurons (mammals)	133,134
ncRNA	Promoter RNA	Stem–loop results in nucleoli localization (mammals)	135
Translation			
mRNAs	*p27*, *VEGFA*	Protein binding causes structural changes (mammals)	20,22
	Collagen genes, amyloid precursor protein, ferritin	Stem–loop at 5′UTR (mammals)	136–138
	BCL2, *ERA*, *TRF2*	G-quadruplex in 5′UTR affects translation (mammals)	139–141
	URE2	Stem–loop as internal ribosomal entry site (yeast)	142
ncRNA	rRNA	Binding of Z-DNA-binding domain to rRNA structures block translation (bacteria and mammals)	143
Riboswitch	FourU, ROSE element, CSPA, TPP, SAM	Structure change on ligand binding and temperature variance changes accessibility of Shine–Dalgarno sequence for ribosomal recognition (bacteria)	9,102, 144–146

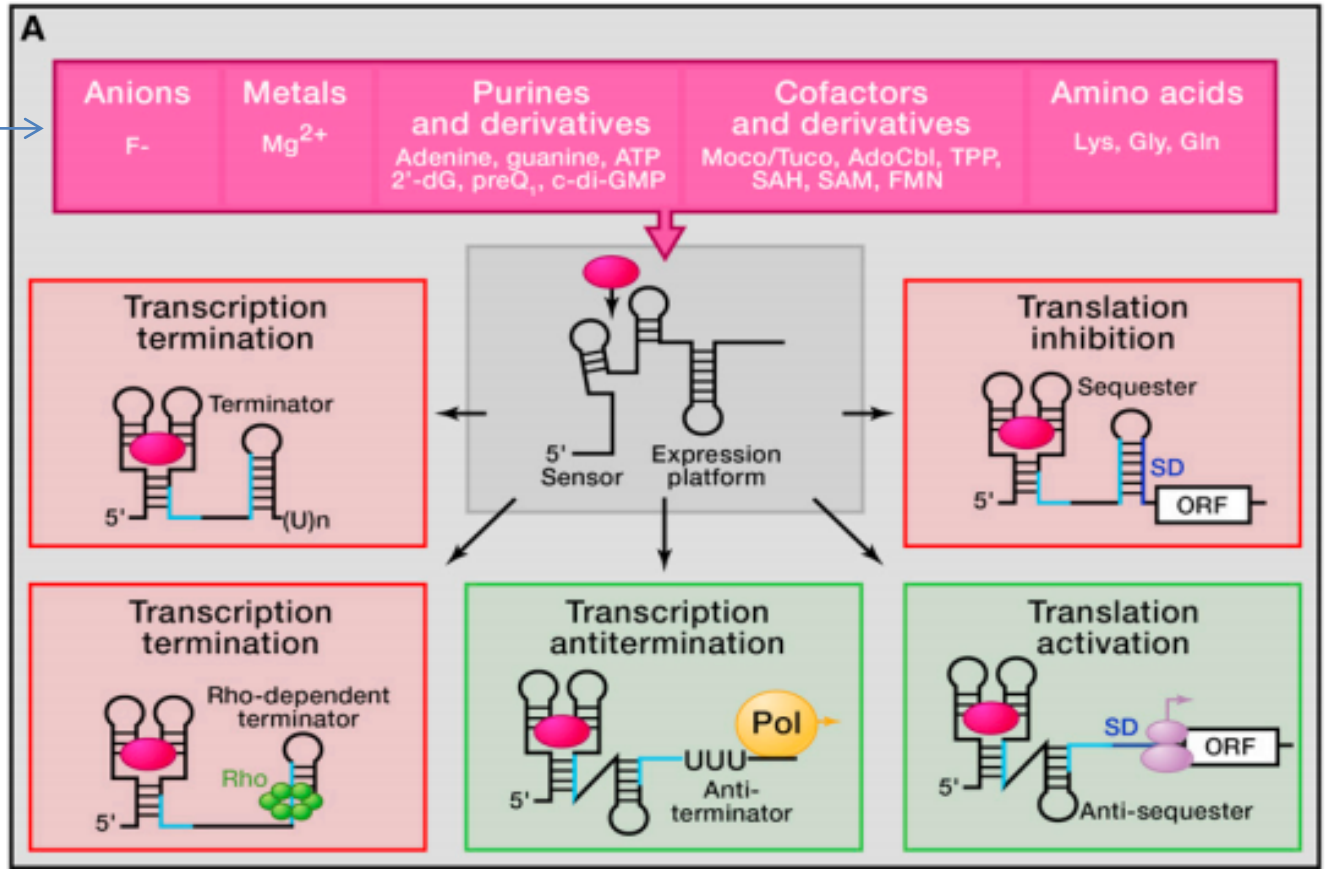
Estruturas secundárias e terciárias determinam papéis regulatórios de RNAs

Exemplo:
Estrutura terciária de um RNA mensageiro de bactéria que liga adenina ("**Riboswitch**") e estabiliza a tradução do mRNA



Bactérias dispõe de diversos mecanismos que ativam ou reprimem a expressão gênica baseados em ***riboswitches*** *(regulação em cis)*

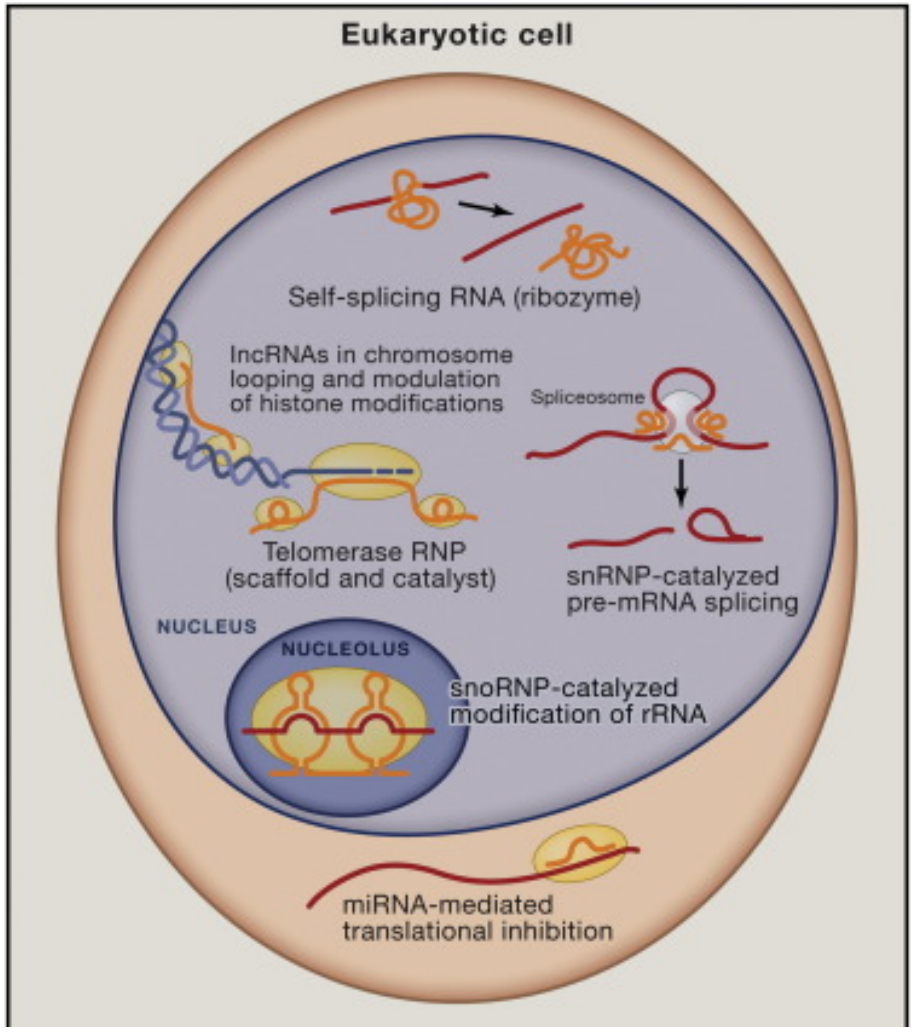
Exemplos de ligantes

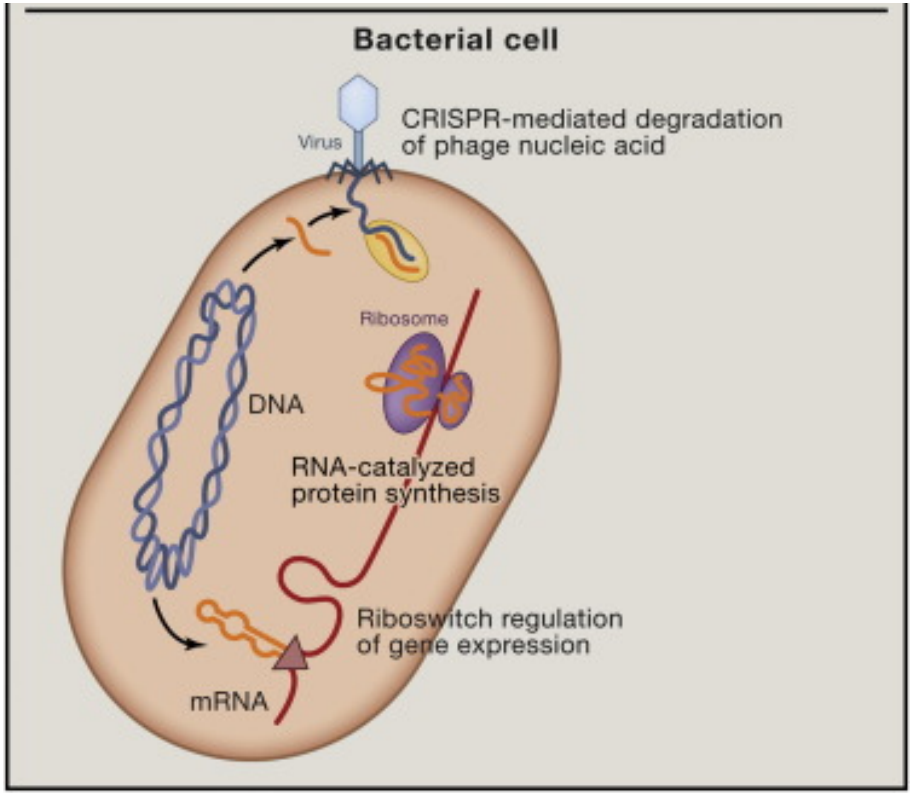


Ligação do ligante ao "riboswitch" induz mudanças conformacionais que **afetam a transcrição** do DNA (através da formação de grampos de terminação/antiterminação), a **tradução** (através do sequestro ou exposição do RBS ou fatores proteicos), ou a **estabilidade** do RNA mensageiro

Serganov & Nudler, 2013 Cell 152:17–24

RNAs não codificadores (ncRNAs) estão presentes em todos os domínios da vida, regulando a expressão genética e contribuindo para a organização e estabilidade do genoma



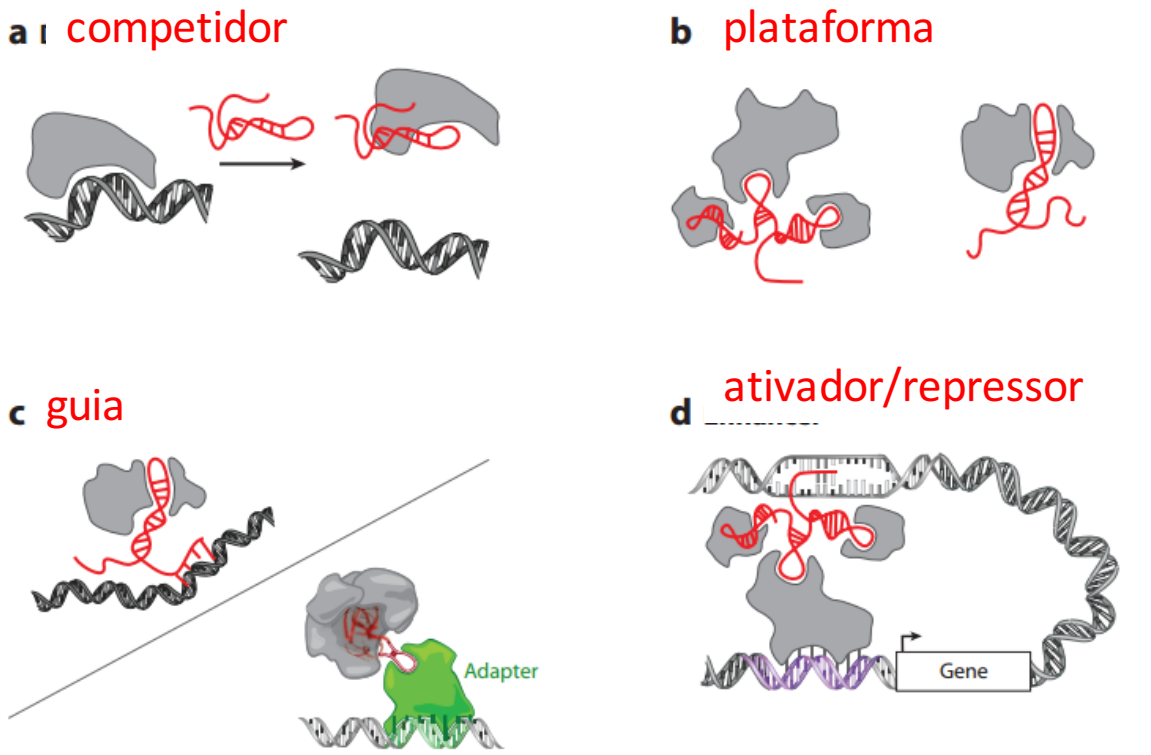


Álguns processos:

- transcrição
- splicing
- tradução

Cech TR, Steitz JA. The noncoding RNA revolution-trashing old rules to forge new ones. Cell. 2014 Mar 27;157(1):77-94

Os mecanismos de ação de **RNAs nãocodificadores longos (lncRNAs)** com papel na regulação epigenética e transcricional em eucariotos dependem da sua estrutura secundária e terciária



- Ativação/repressão da expressão gênica
- Organização de complexos ribonucleoproteicos

Figure 4

Models of long noncoding RNA (lncRNA) mechanisms of action. (*a*) The lncRNAs can act as decoys that titrate away DNA-binding proteins, such as transcription factors. (*b*) These lncRNAs may act as scaffolds to bring two or more proteins into a complex or spatial proximity and (*c*) may also act as guides to recruit proteins, such as chromatin modification enzymes, to DNA; this may occur through RNA-DNA interactions or through RNA interaction with a DNA-binding protein. (*d*) Such lncRNA guidance can also be exerted through chromosome looping in an enhancer-like model, where looping defines the *cis* nature and spread of the lncRNA effect.

Rinn & Chang, 2012 Annu. Rev. Biochem. 81:145–66

Ferramentas para predição de estrutura de RNAs

https://en.wikipedia.org/wiki/List_of_RNA_structure_prediction_software

List of RNA structure prediction software

From Wikipedia, the free encyclopedia

This **list of RNA structure prediction software** is a compilation of software tools and web portals used for RNA structure prediction.

Contents [hide]

Métodos de predição de estrutura secundária de RNAs

- **critério termodinâmico**

Buscam identificar **regiões de complementaridade intra-intermolecular** com formação de trechos de dupla-fita que resultam em estruturas secundárias estáveis utilizando métodos de **minimização de energia**.

- **critério evolutivo**

Buscam identificar **covariações de bases** na sequencia primária que mantenham estruturas secundárias conservadas em diferentes espécies.

Determinação de estruturas secundárias de RNA com menor energia livre (ΔG)

A **variação da energia livre** (ΔG) que contribui para a estabilização de estruturas secundáriasno RNA é calculada pela soma da contribuição individual de **interações locais** que estabilizam ou desestabilizam a sua formação:

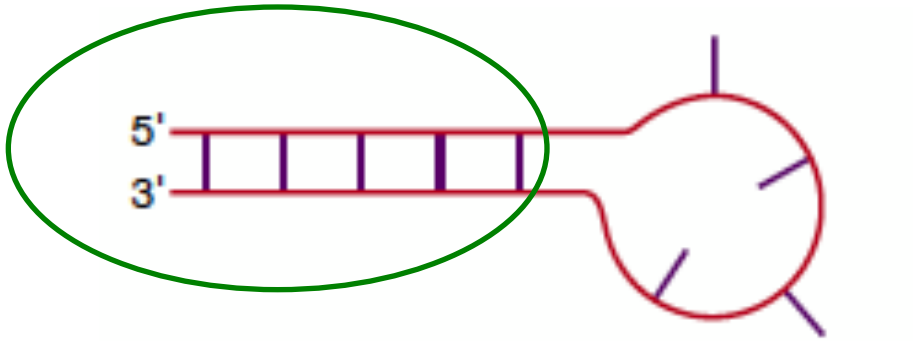
- Empillhamento de pares de bases adjacentes nos elementos de estrutura secundária (estabiliza estrutura secundária, **ΔG < 0**)

- Voltas (desestabiliza estrutura secundária, **ΔG > 0**)

Empillhamento de pares de bases adjacentes (ΔG < 0)

Quanto maior o número de ligações de hidrogênio maior a estabilidade:
C-G: 3 ligações hidrogênio
A-U: 2 ligações hidrogênio

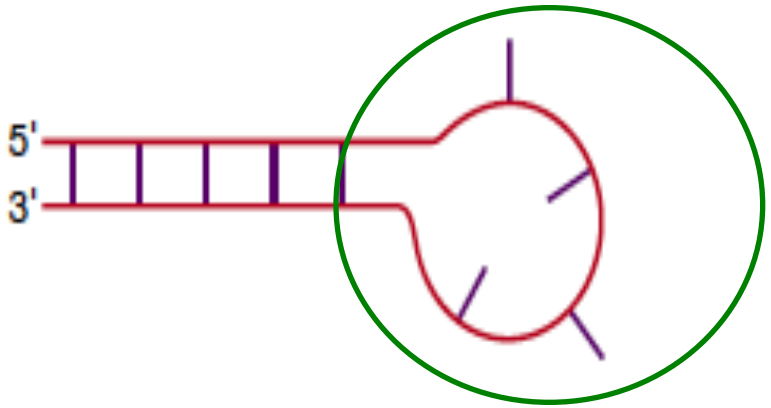
A vizinhança da base influencia na estabilidade



	Stacking Energies for base pairs					
	A/U	C/G	G/C	U/A	G/U	U/G
A/U	-0.9	-1.8	-2.3	-1.1	-1.1	-0.8
C/G	-1.7	-2.9	-3.4	-2.3	-2.1	-1.4
G/C	-2.1	-2.0	-2.9	-1.8	-1.9	-1.2
U/A	-0.9	-1.7	-2.1	-0.9	-1.0	-0.5
G/U	-0.5	-1.2	-1.4	-0.8	-0.4	-0.2
U/G	-1.0	-1.9	-2.1	-1.1	-1.5	-0.4

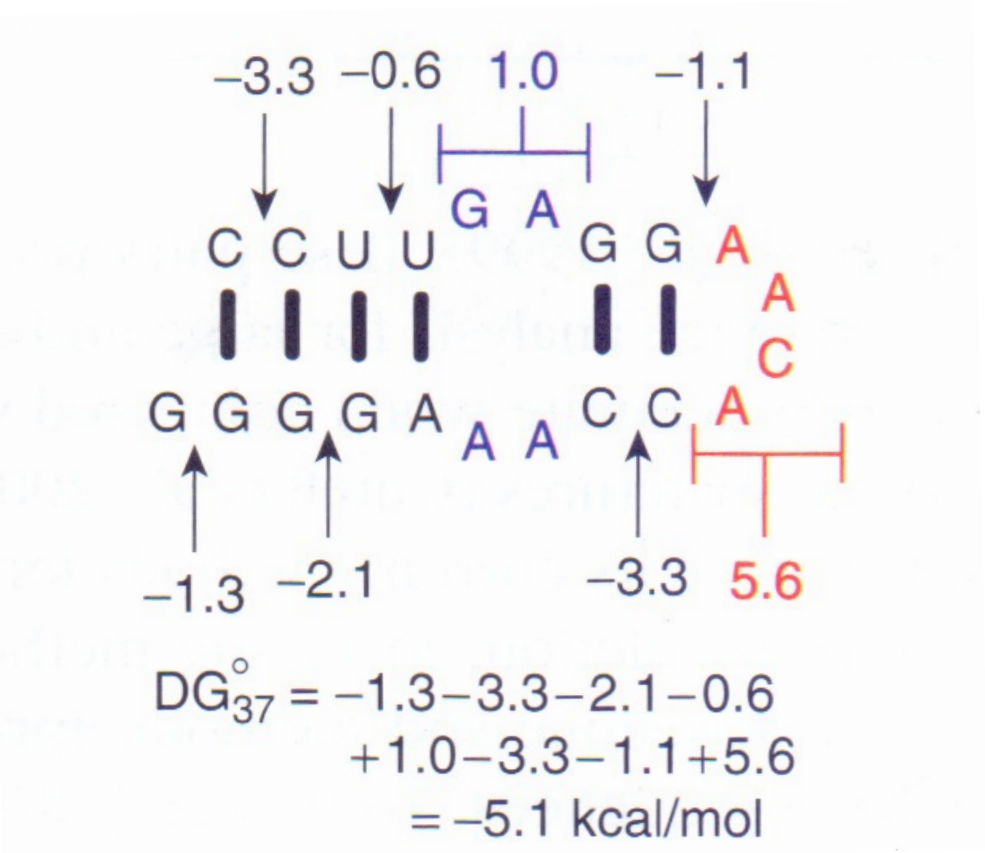
Voltas (“loops”) (ΔG > 0)

Quanto maior o trecho de bases desemparelhadas menor a estabilidade da estrutura



	Destabilizing Energies for Loops				
Number of Bases	1	5	10	20	30
Internal	--	5.3	6.6	7.0	7.4
Bulge	3.9	4.8	5.5	6.3	6.7
Hairpin	--	4.4	5.3	6.1	6.5

Predição de estruturas estáveis a partir da determinação de estruturas secundárias de RNA com **menor energia livre** (ΔG min)



rCCUUGAGGAACACCAAAGGGG

Relação entre ΔG e equilíbrio químico definem a estabilidade de estruturas secundárias

RNA sem conformação "random coil" $\rightleftharpoons$ (K_{eq}) RNA estruturado conformação "A"

$$K_{eq} = \frac{[\text{conformação "A"}]}{[\text{"random coil"}]} = e^{-\Delta G°/RT}$$

Energia livre de Gibbs

Exemplo prático: Estabilidade predita da conformação A = -5.1 Kcal/mol
Portanto:
K_{eq} = 3900 , ou seja, existem 3900 moléculas na conformação A para cada molécula sem conformação (random coil)

“Dot Matrix”

Método para identificação de regiões de complementaridade

- útil para vizualizar regiões de auto-complementaridade na sequencia primária do RNA

- auxiliam a predição da existência de estruturas secundárias

• Dot Matrix de RNA do viroide de tubérculo de batata (< 350 nt)

• diagonais indicam diferentes trechos da molécula que apresentam complementaridade intramolecular e portanto podem formar estruturas secundárias estáveis

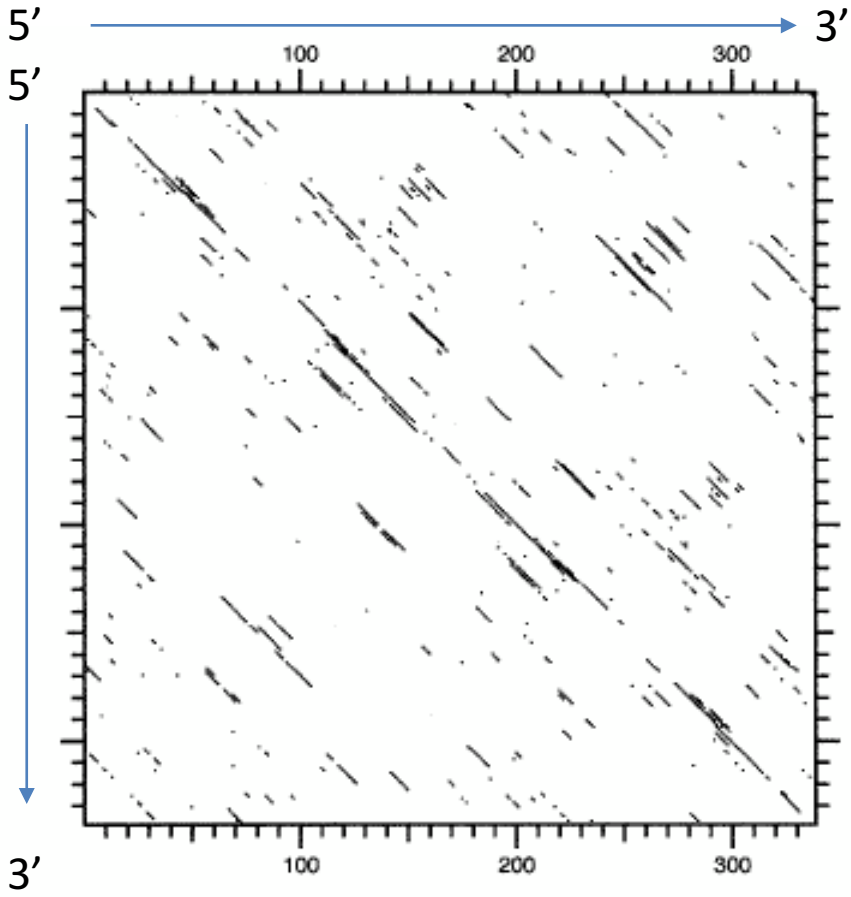
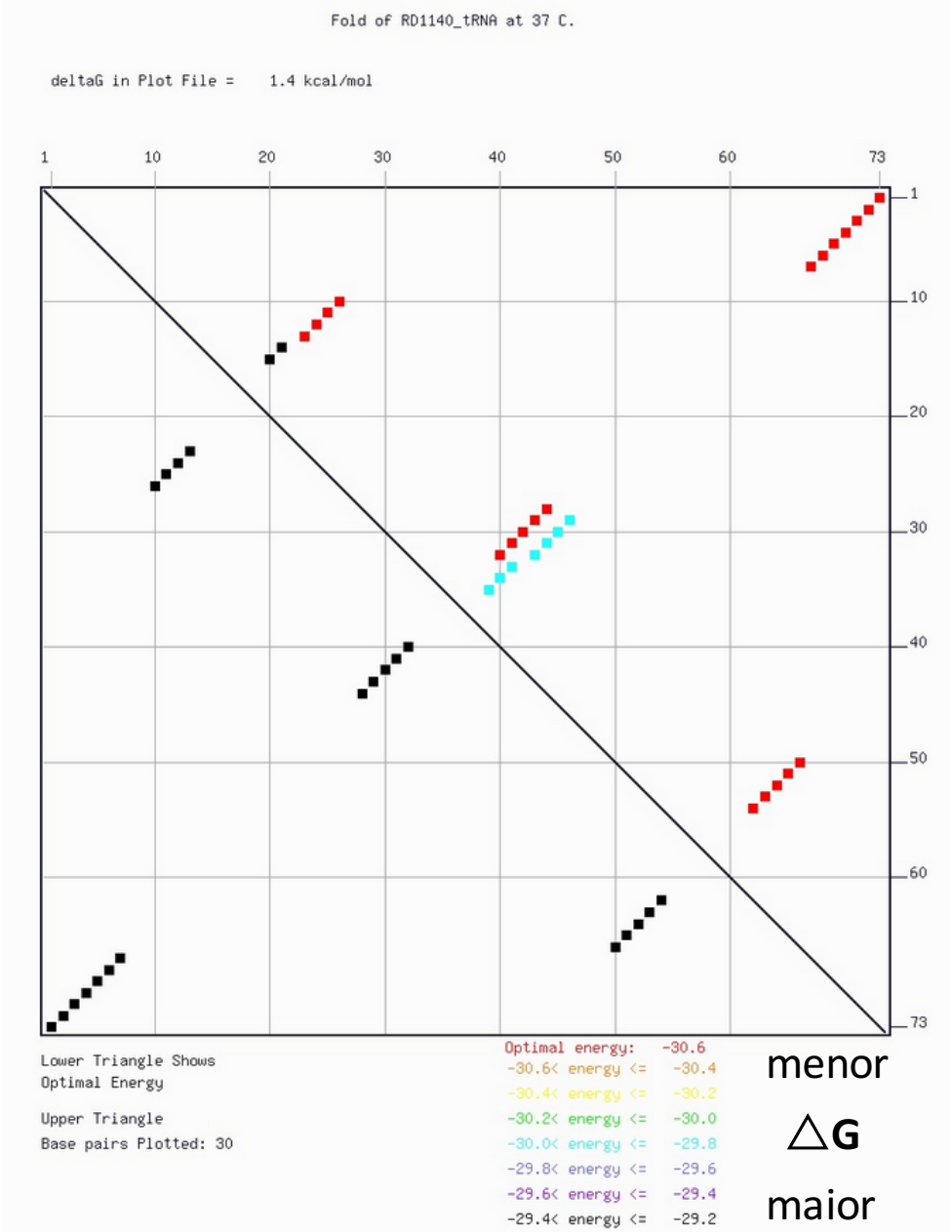
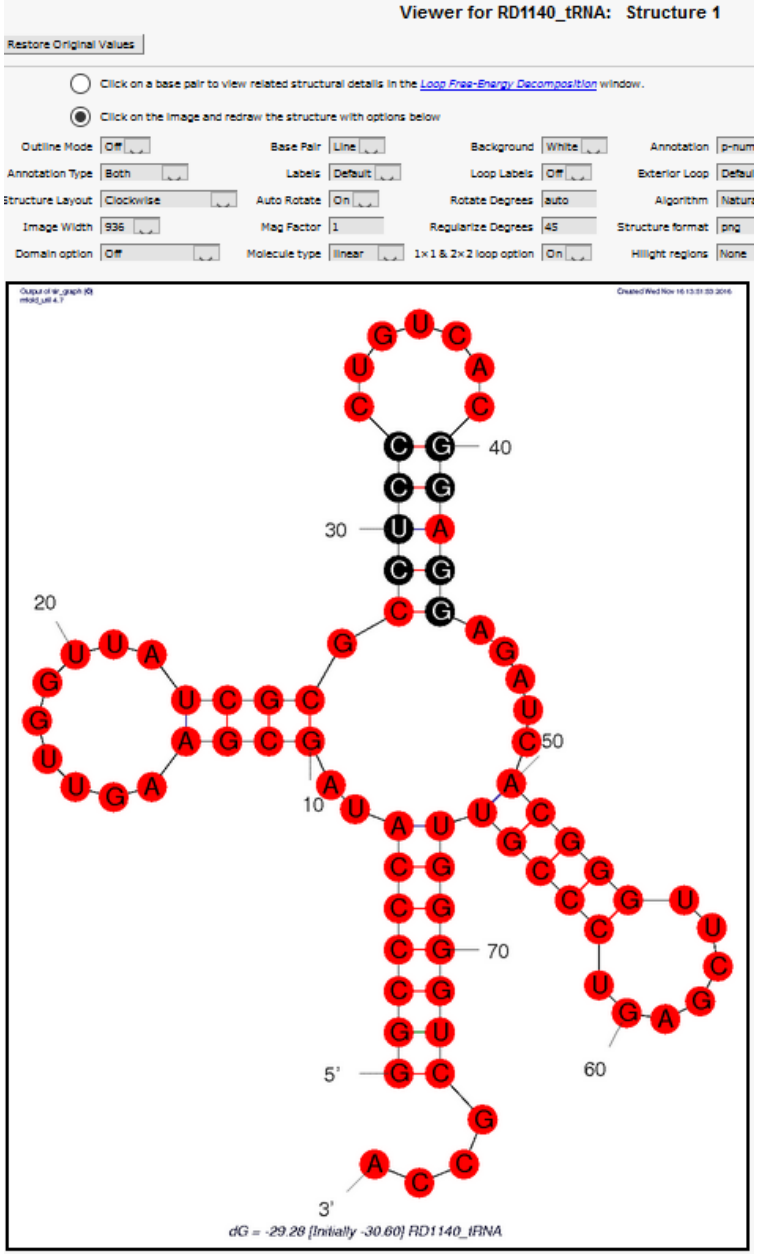


Figure 5.5. Dot matrix analysis of the potato tuber spindle viroid for RNA secondary structu
using the MATRIX function of DNA Strider v. 1.2 on a Macintosh computer.

Predição da estrutura do RNA transportador (77 nt)

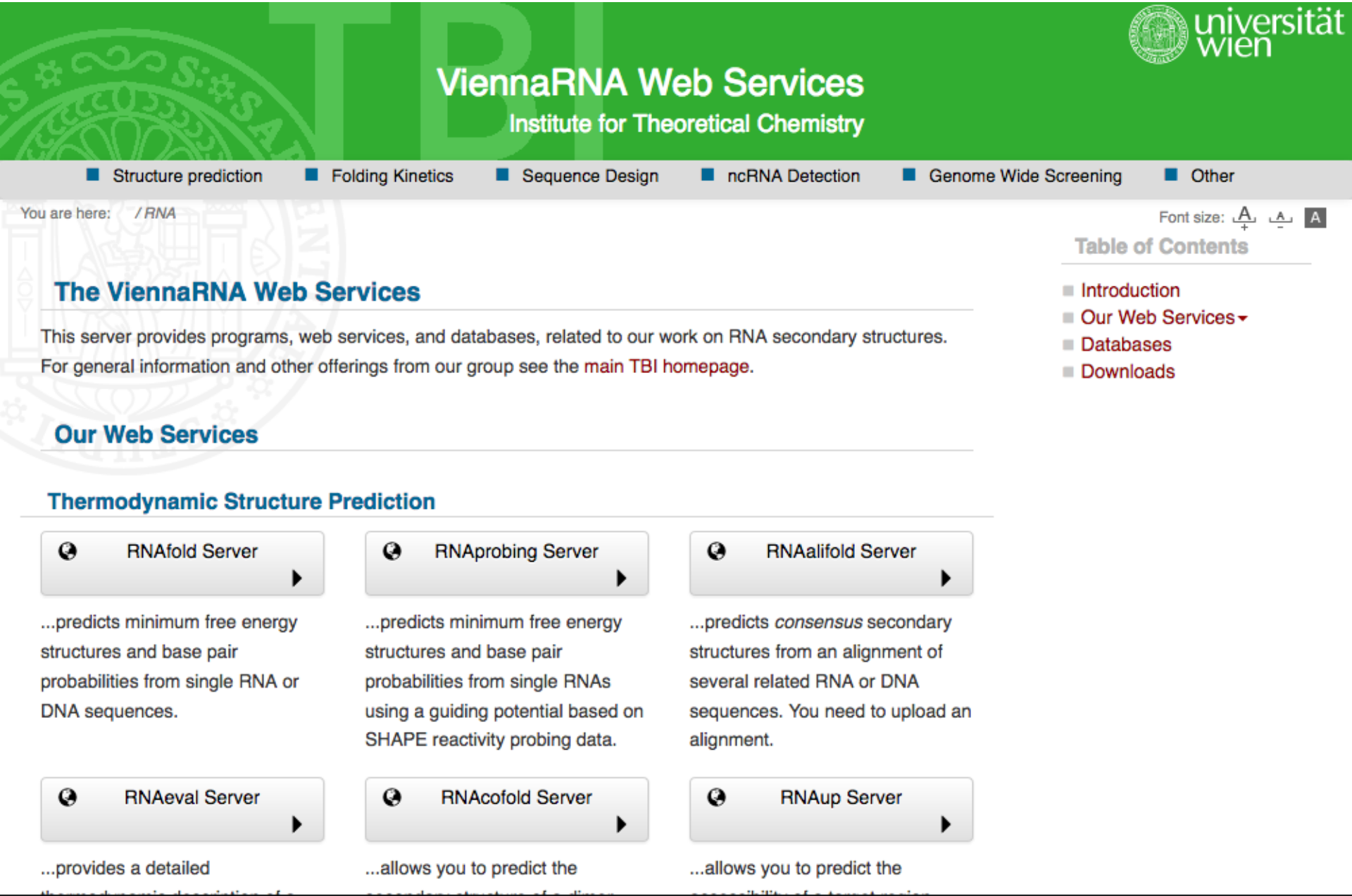




Loop Free-Energy Decomposition Structure 1
RD1140_tRNA
ΔG = -30.60

Structural element	δG	Information
External loop	-1.70	4 ss bases & 1 closing helices.
Stack	-1.50	External closing pair is G_1-C_{73}
Stack	-2.50	External closing pair is G_2-U_{72}
Stack	-3.30	External closing pair is C_3-G_{71}
Stack	-3.30	External closing pair is C_4-G_{70}
Stack	-3.30	External closing pair is C_5-G_{69}
Stack	-2.10	External closing pair is C_6-G_{68}
Helix	-16.00	7 base pairs.
Multi-loop	2.10	External closing pair is A_7-U_{67} 8 ss bases & 4 closing helices.
Stack	-2.20	External closing pair is A_{50}-U_{66}
Stack	-2.40	External closing pair is C_{51}-G_{65}
Stack	-3.30	External closing pair is G_{52}-C_{64}
Stack	-3.30	External closing pair is G_{53}-C_{63}
Helix	-11.20	5 base pairs.
Hairpin loop	4.40	Closing pair is G_{54}-C_{62}
Stack	-3.30	External closing pair is C_{28}-G_{44}
Stack	-2.10	External closing pair is C_{29}-G_{43}
Stack	-2.40	External closing pair is U_{30}-A_{42}
Stack	-3.30	External closing pair is C_{31}-G_{41}
Helix	-11.10	5 base pairs.
Hairpin loop	5.00	Closing pair is C_{32}-G_{40}
Stack	-3.40	External closing pair is G_{10}-C_{26}
Stack	-2.40	External closing pair is C_{11}-G_{25}
Stack	-2.40	External closing pair is G_{12}-C_{24}
Helix	-8.20	4 base pairs.
Hairpin loop	6.10	Closing pair is A_{13}-U_{23}

http://rna.tbi.univie.ac.at/



RNAfold WebServer

Results for minimum free energy prediction

The optimal secondary structure in dot-bracket notation with a minimum free energy of **-30.20** kcal/mol is given below.
[color by base-pairing probability | color by positional entropy | no coloring]

```
1      GGCCCCAUAGCGAAGUUGGUUAUCGCGCCUCCCUGUCACGGAGGAGAUCACGGGUUCGAGUCCCGUUGGGGUCGCCA
1      (((((((..((((.........)))).(((((.......)))))).....(((((.......)))))))))))....
```

You can download the minimum free energy (MFE) structure in [Vienna Format| Ct Format]. You can get thermodynamic details on this structure by submitting to our RNAeval web server.

Results for thermodynamic ensemble prediction

The free energy of the thermodynamic ensemble is **-30.90** kcal/mol.
The frequency of the MFE structure in the ensemble is **31.97** %.
The ensemble diversity is **8.24** .
You may look at the **dot plot** containing the base pair probabilities [EPS|PDF|IMAGE CONVERTER].

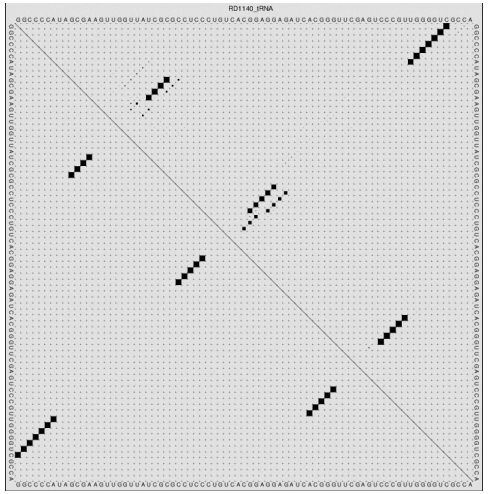
The centroid secondary structure in dot-bracket notation with a minimum free energy of **-30.20** kcal/mol is given below.
[color by base-pairing probability | color by positional entropy | no coloring]

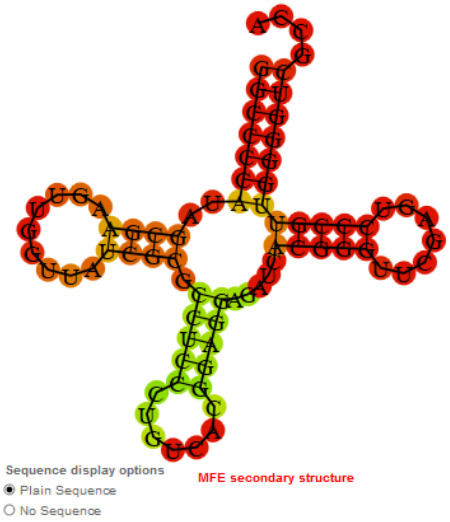
```
1      GGCCCCAUAGCGAAGUUGGUUAUCGCGCCUCCCUGUCACGGAGGAGAUCACGGGUUCGAGUCCCGUUGGGGUCGCCA
1      (((((((..((((.........)))).(((((.......)))))).....(((((.......)))))))))))....
```

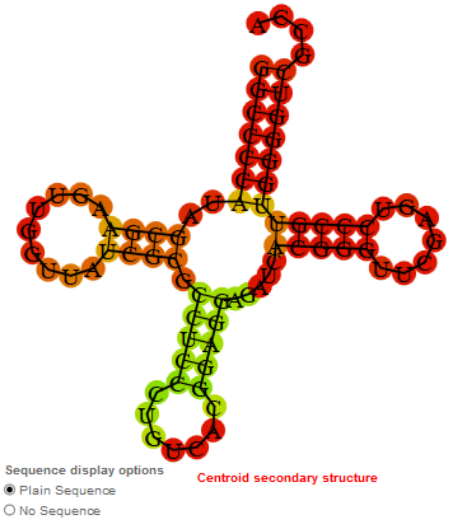
You can download the minimum free energy (MFE) structure in [Vienna Format| Ct Format]. You can get thermodynamic details on this structure by submitting to our RNAeval web server.

Graphical output

You may look at the interactive drawing of the MFE structure below. If you do not see the interactive drawing and you are using Internet Explorer, please install the Adobe SVG plugin. **A** regions the color denotes the probability of being unpaired.



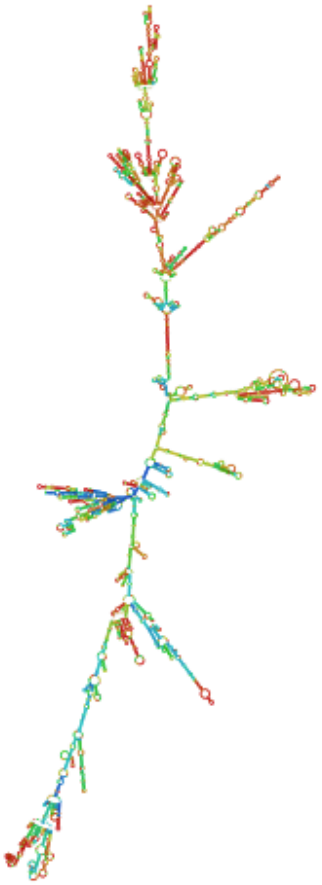




Predição termodinâmica de estruturas secundárias

A estrutura secundária centróide de uma sequência de RNA é a estrutura secundária com distância mínima de pares de bases para todas as outras estruturas preditas no modelo

minimum free energy of **-1957.30** kcal/mol



Sequence display options

◉ Plain Sequence

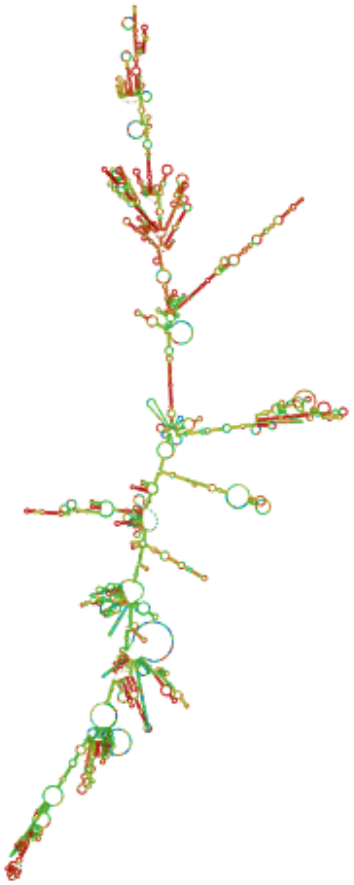
○ No Sequence

MFE secondary structure

The free energy of the thermodynamic ensemble is **-2082.59** kcal/mol.

The frequency of the MFE structure in the ensemble is **0.00** %.

The ensemble diversity is **1812.57**.



Sequence display options

◉ Plain Sequence

○ No Sequence

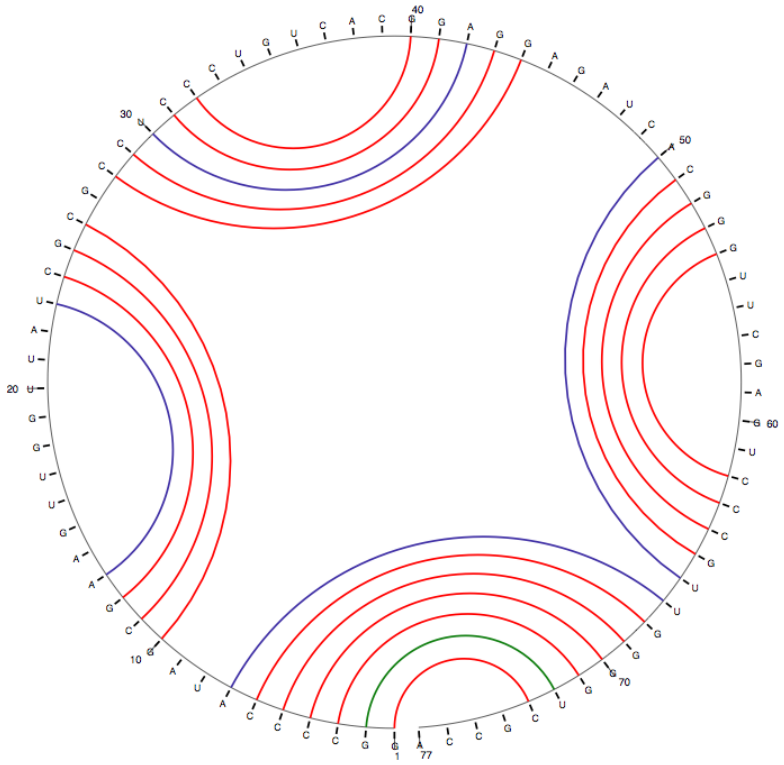
Centroid secondary structure

Outros formatos para representação de estrutura 2D de RNAs

“dot-bracket”

```
1       GGCCCCAUAGCGAAGUUGGUUAUCGCGCCUCCCUGUCACGGAGGAGAUCACGGGUUCGAGUCCCGUUGGGGUCGCCA
1       (((((((..((((.........)))).(((((.......)))))).....(((((.......)))))))))))....
```

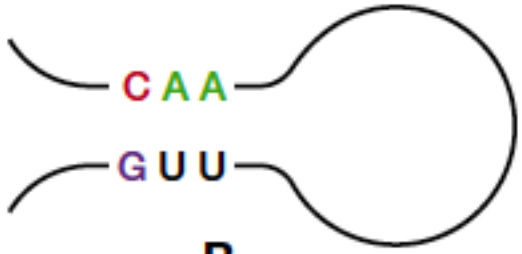
gráficos circulares



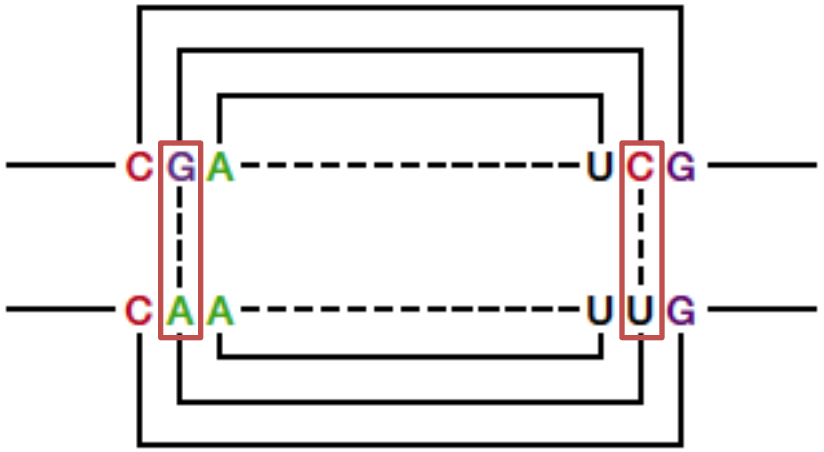
Métodos de predição de estrutura secundária de RNAs baseados em critérios evolutivos

RNAs com sequencias diferentes podem assumir a mesma estrutura secundária

A.



B.



covariação de bases:

G→A e C→U

Figure 5.1. Complementary sequences in RNA molecules maintain RNA secondary structure. Shown is a simple stem-and-loop structure formed by the RNA strand folding back on itself. Molecule *A* depends on the presence of two complementary sequences CGA and UCG that are base-paired in the structure. In *B*, two sequence changes, G → A and C → U, which maintain the same structure, are present. Aligning RNA sequences required locating such regions of sequence covariation that are capable of maintaining base-pairing in the corresponding structure.

Predição de estruturas secundárias através da **identificação de covariância de bases** na sequência de RNAs conservados (homologos) em diferentes espécies

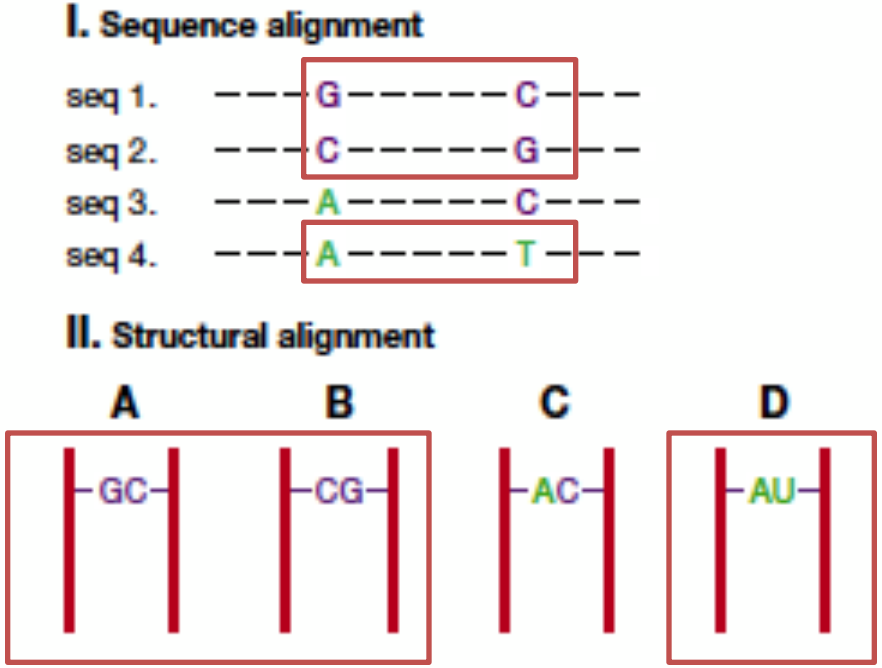


Figure 5.11. Conservation of base pairs in homologous RNA molecules influences structure prediction. The predicted structure takes into account sequence covariation found at aligned sequence positions, and may also use information about conserved positions in components of a phylogenetic tree. In the example shown, sequence covariations in A, B, and D found in sequences 1, 2, and 4, respectively, permit Watson–Crick base and G-U base-pairing in the corresponding structure, but variation C found in sequence 3 is not compatible. Sometimes correlations will be found that suggest other types of base interactions, or the occurrence of a common gap in a multiple sequence alignment may be considered a match. Positions with greater covariation are given greater weight in structure prediction. Molecules with only one of the two sequence changes necessary for conservation of the base-paired position may be functionally deleterious.

Predição de estrutura de tRNA a partir de análise de covariação de bases

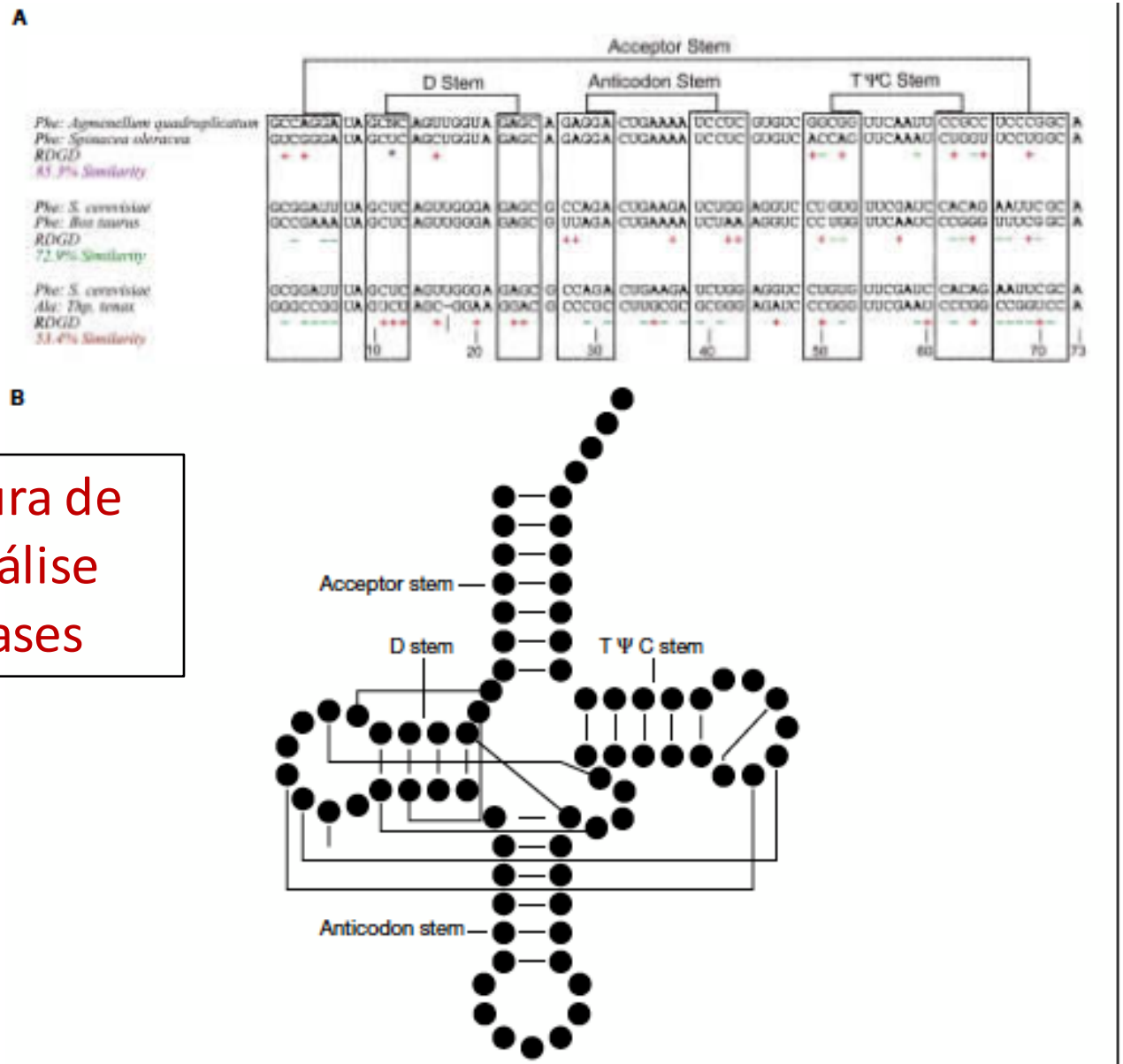


Figure 5.12. Covariation found in tRNA sequences reveals base interactions in tRNA secondary and tertiary structure. (*A*) Alignment of tRNA sequences showing regions of interacting base pairs. (+) Transition; (−) transversions; (|) deletion; (ˣ) ambiguous nucleotide. (*B*) Diagram of tRNA structure illustrating base–base interactions revealed by a covariance analysis. Adapted from the Web site of R. Gutell at http://www.rna.icmb.utexas.edu.

transição

A → G

G → A

C → U

U → C

Transversão

A → C

A → U

C → A

U → A

G → C

G → U

C → G

U → G

Methods of Covariation Analysis in RNA Sequences

Secondary and tertiary features of RNA structure may be determined by analyzing a group of related sequences for covariation. Two sequence positions that covary in a manner that frequently maintains base-pairing between them provides evidence that the bases interact in the structure. Combinations of the following methods have been used to locate such covarying sites in RNA sequences (see R. Gutell for additional details and at http://www.rna.icmb.utexas.edu/METHODS/menu.html).

1. Optimally align pairs of sequence to locate conserved primary sequence, mark transitions and transversions from a reference sequence, and then visually examine these changes to identify complementary patterns that represent potential secondary structure.
2. Perform a multiple sequence alignment, highlight differences using one of the sequences as a reference, and visually examine for complementary patterns.
3. Mark variable columns in the multiple sequence alignment by numbers that mark changes (e.g., transitions or transversions) from a reference sequence; examine marked columns for a similar or identical number pattern that can represent potential secondary structure.
4. Perform a statistical analysis (Chi-square test) of the number of observations of a particular base pair in columns *i* and *j* of the multiple sequence alignment, compared to the expected number based on the frequencies of the two bases.
5. Calculate the mutual information score (mi*xy*) for each pair of columns in the alignment, as described in the text and illustrated in Figure 5.13.
6. Score the number of changes in each pair of columns in the alignment divided by the total number of changes (the ec score), examine the phylogenetic context of these changes to determine the number of times the changes have occurred during evolution, and choose the highest scores that are representative of multiple changes.
7. Measure the covariance of each pair of positions in the alignment by counting the numbers of all 16 possible base-pair combinations and dividing by the expected number of each combination (number of sequence × frequency of base in first position × frequency of base in second position), choose the most prevalent pair, and examine remaining combinations for additional covariation; then sum frequency of all independently covarying sites to obtain covary score.

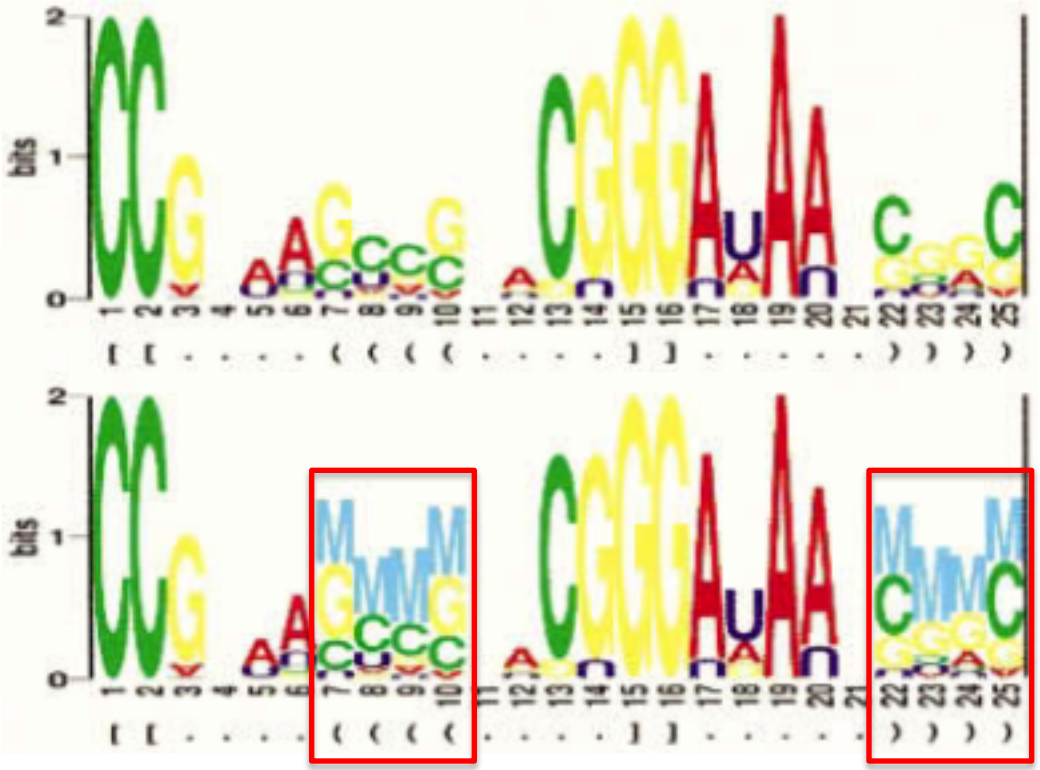
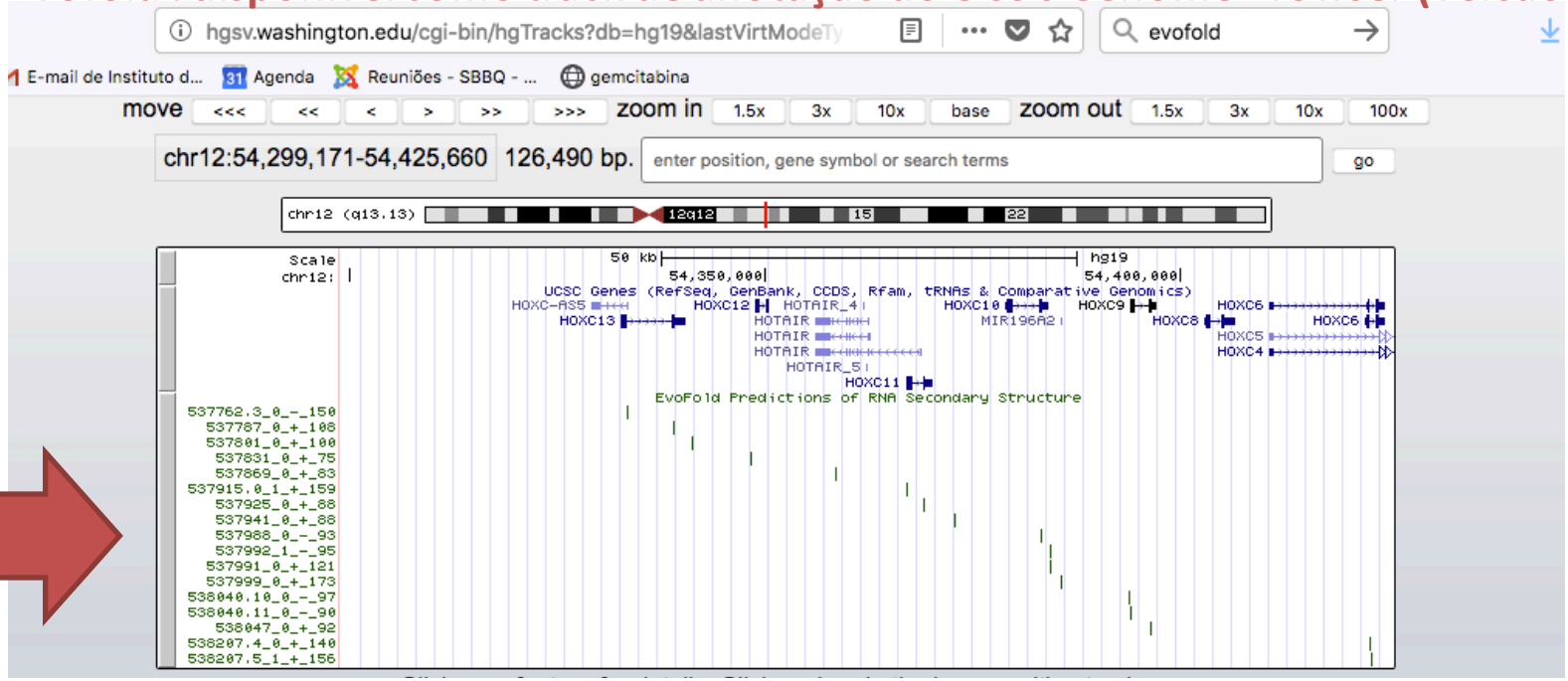
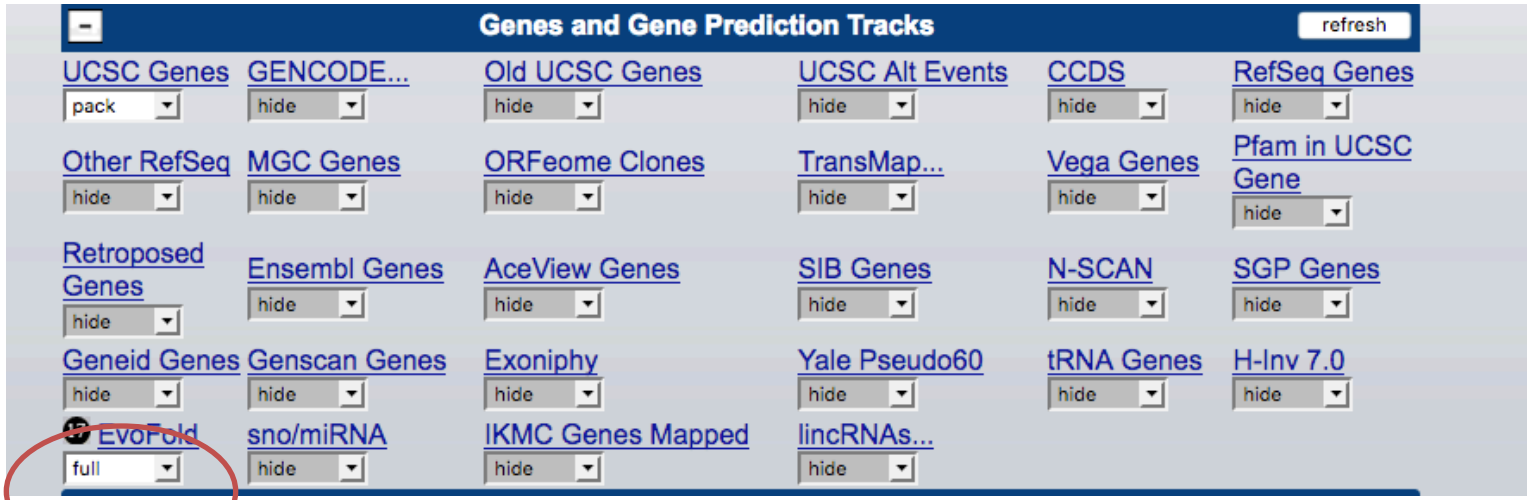


Figure 5.13. RNA structure logo. The top panel is the normal sequence logo showing the size of each base in proportion to the contribution of that base to the amount of information in that column of the multiple sequence alignment. The relative entropy method is used in which the frequency of bases in each column is compared to the background frequency of each base. Inverted sequence characters indicate a less than background frequency (see Chapter 4, page 196). The bottom panel includes the same information plus the mutual information content in pairs of columns. The amount of information is indicated by the letter M, and the matching columns are shown by nested sets of brackets and parentheses. All sequences have a C in column 1 and a matching G in column 16. Similar columns 2 and 15 can form a second base pair stacked upon the first. Columns 7–10 and 25–22 also can form G/C base pairs most of the time. Sequences with a G in column 7 frequently have a C in column 25, and those with a C in column 7 may have a G in column 25. Thus, there is mutual information in these two columns (Gorodkin et al. 1997 [using data of Tuerk and Gold 1990]).

Predição global de regiões estruturadas em RNAs em genomas a partir de modelos filogenéticos

Evofold : disponível como track de anotação do UCSC Genome Browser (versão hg19)





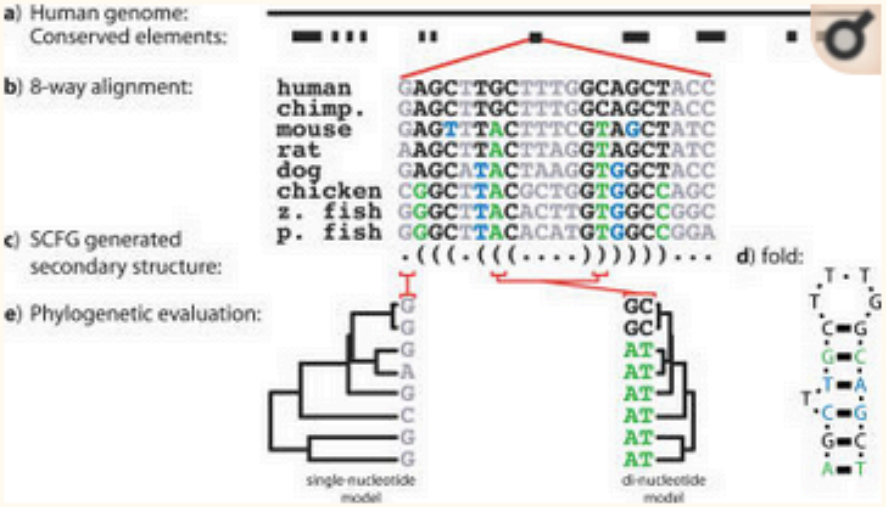


Figure 1

Outline of EvoFold Prediction Method

(A) Schematic representation of human genome and conserved elements. The conserved elements define the input alignments.

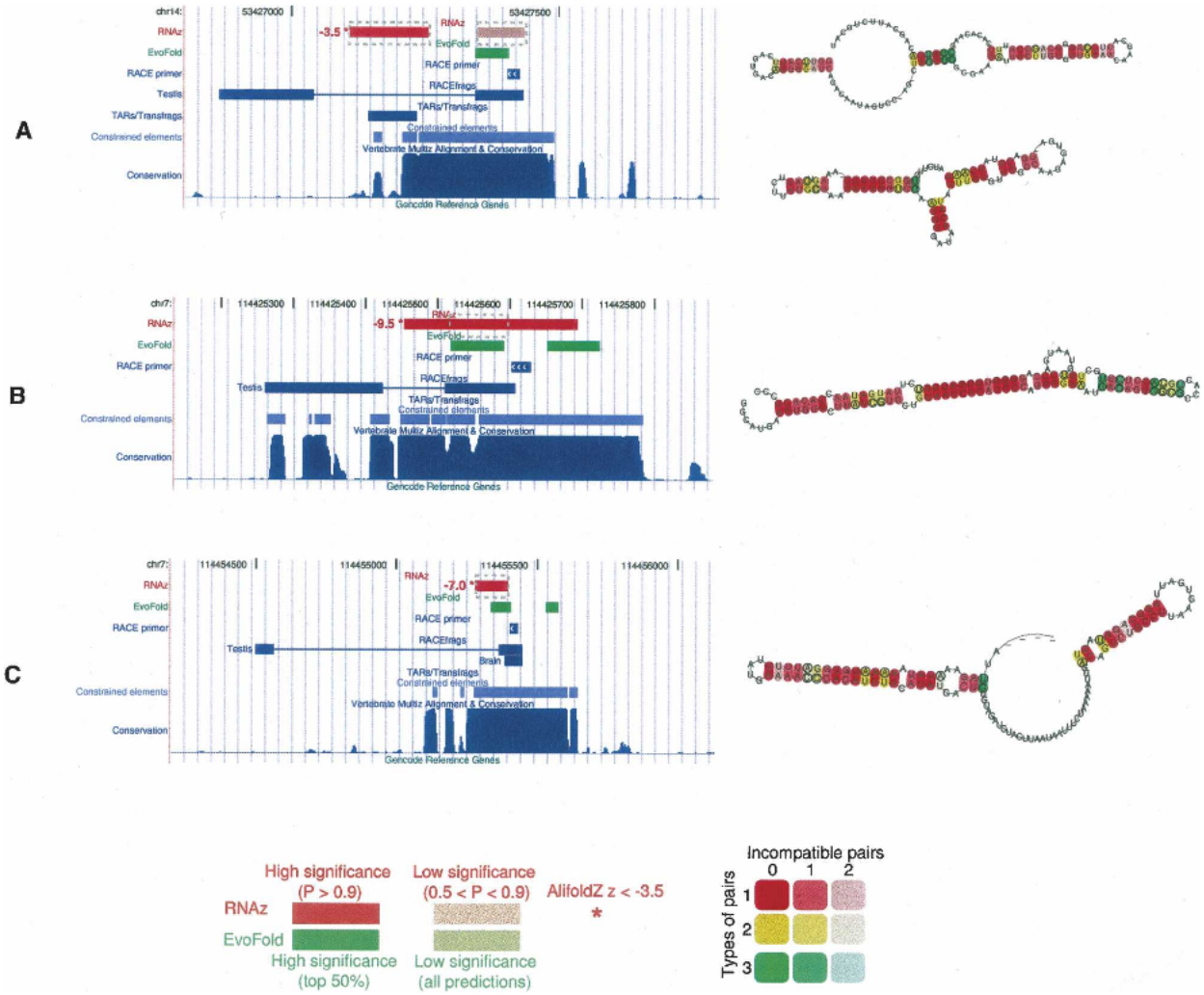
(B) Segment of eight-way genomic alignment.

(C) The SCFG of the fRNA model defines a distribution over all possible secondary-structure annotations. One of the many possible secondary structures is shown in parenthesis format. Substitutions in pairing regions of the alignment are color-coded relative to human: compensatory double substitutions are green, and compatible single substitutions are blue.

(D) Color-coded fold corresponding to the secondary-structure annotation of the alignment.

(E) Two phylogenetic models are used to evaluate the possible secondary-structure annotations: unpaired columns are evaluated using a single-nucleotide phylogenetic model. Paired columns are combined and evaluated using a di-nucleotide phylogenetic model. Horizontal branch lengths reflect the expected number of substitutions.

Conservação de sequencia indica regiões estruturadas de RNAs

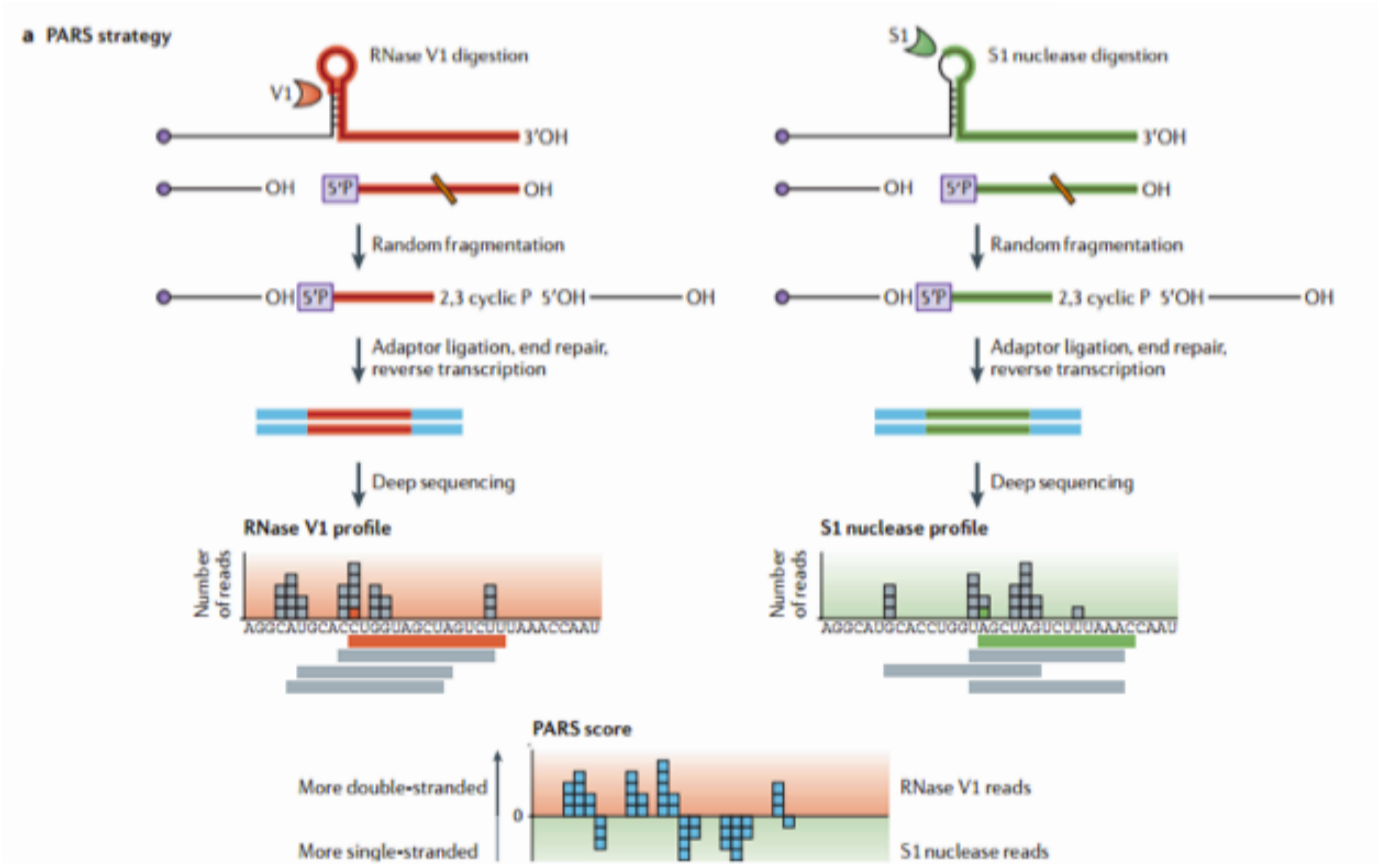


Washietl et al. 2007 Genome Research 17:852–864

Determinação experimental da estrutura de RNAs

Digestão com endonucleases seguida de sequenciamento permite distinguir regiões inacessíveis (contem estruturas secundárias) e acessíveis (fita simples) do RNA

PARS: Parallel Analysis of RNA Structure

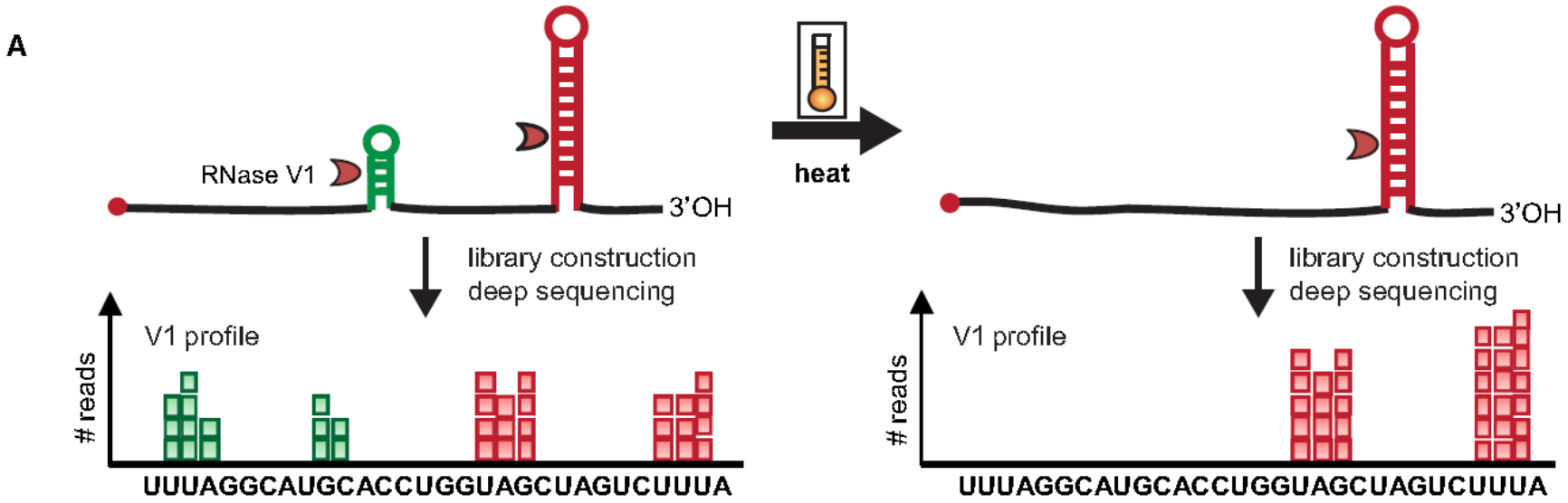


RNAse V1: digere RNA dupla fita

S1 nuclease: digere RNA fita simples

Wan et al., 2011 Nature Review Genetics **12**:641

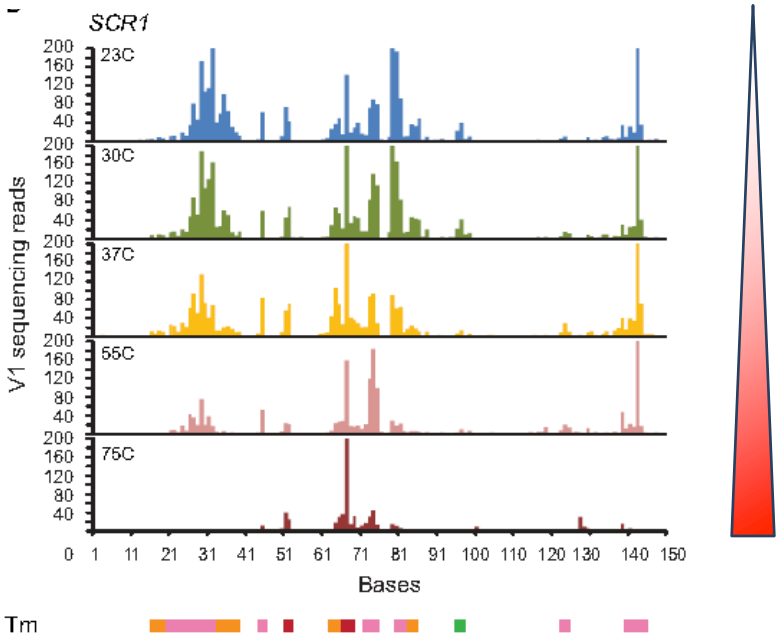
PARTE:
Parallel Analysis of RNA Structures with Temperature Elevation



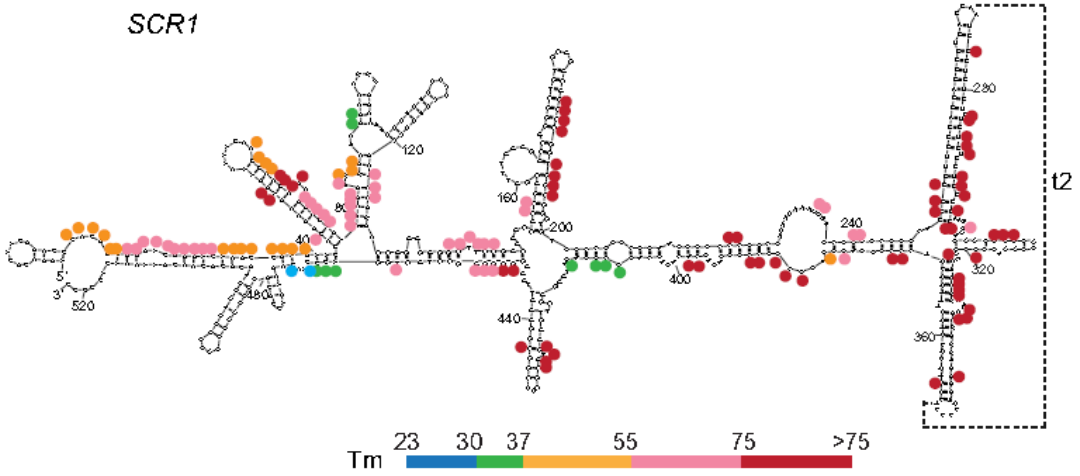
Wan et al., 2011 Nature Review Genetics **12**:641

Exemplo: PARTE no estudo da dinâmica estrutural do RNA que compõe a partícula de reconhecimento de sinal SCR1

- Estabilidade



- Permite refinar a informação sobre a estabilidade das estruturas secundárias presentes na molécula



https://rmdb.stanford.edu/

RMDB Browse Deposit Analyze Tools About MyRMDB Login

RMDB has upgraded to version 2.2 BETA for online preview. More changes and updates are in progress.
Please report bugs and make suggestions. We really appreciate your feedback.

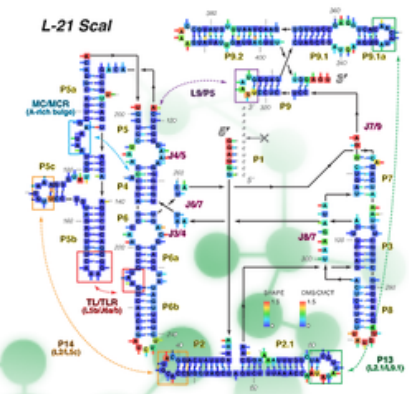
Welcome
Browse
Deposit
Download
Analyze
Help

A Repository of RNA Structure Probing

The RNA Mapping Database is an archive that contains results of diverse structural mapping experiments performed on ribonucleic acids. Results `ted to the repository are manually curated and annotated, ensuring reliability and quality of reported data. With a total of 13,286,924 datapoints , RMDB currently houses 426 entries, describing more than 745 experiemnts of 123,127 RNA constructs in several solution conditions and has been growing rapidly .

In addition, RMDB contributes greatly to the **Eterna** videogame/synthetic biology project, and the **RNA Puzzles** blind RNA structure prediction challenges. It will continue to benefit data sharing for RNA structue mapping community.

The ultimate goal of the RMDB is to provide a centralized, curated, data-sharing platform with visualization and structure prediction tools for rapid analysis and meta-analysis of these data.



Tutoriais sobre estrutura de RNAs na página da disciplina

Bibliografia sugerida

- Baxevanis & Ouellette. Bioinformatics: a practical guide to the analysis of genes and proteins. 3rd edition.
- Mount. D. Bioinformatics – Sequence and Genome analysis – 2nd Edition.
- Genome-wide Measurement of RNA Folding Energies. Wan et al.,2012 Mol Cell48:169-81