

Variations on RNA folding: Locally stable structures and RNA hybridization

Defense Talk

Stephan Bernhart

Institute for Theoretical Chemistry
University of Vienna

Wien, 9/21/2007

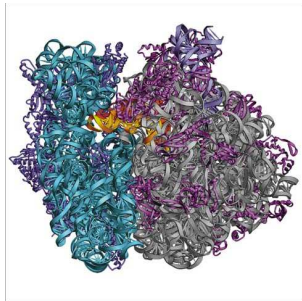
Outline

- 1 Introduction
 - RNA structure
 - RNA structure prediction
- 2 Bimolecular Secondary structures of RNA molecules
 - Introduction
 - Implementation
 - Application
- 3 Local partition function
 - RNAPfold
 - Results
 - Accessibility
 - Results

RNA biology

Biological functions of RNA

- tRNA, mRNA, rRNA
- maturation: RNase P, snoRNAs
- guide RNAs for editing
- spliceosomal RNAs
- functional motifs in mRNA
- Signal recognition particle
- miRNA, siRNA, piRNA
- T-Box RNA
- vault RNAs
- telomerase RNAs
- RNase MRP

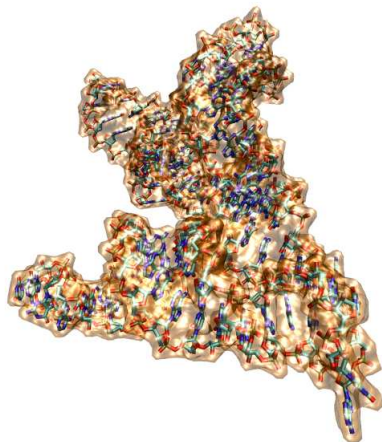
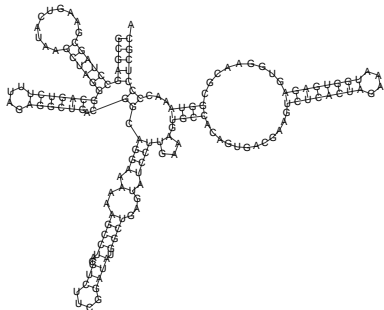


Ribosome

Levels of RNA structure description

Primary, secondary and tertiary structure

CGCAGCCUAGCGAAGUCAUAAGCUAAGGCCAGUCUUUAGAGGCCUGACGGCCAGGAAAAAAGCCUA CGUCUUCGGAAUAGGCCUGA GUAUCCUUGAAAAGUGCCACAGUGACGAAGUCUCACUAGAAAUGGUGAGAGUGGAACCGCGUAAAACCCUCGCA

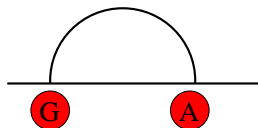


RNA secondary structure

Definition

List of base pairs i, j such that:

- only GC, AU or GU base pair
- only one base-pair per base
- minimum distance 3
- base pairs do not cross

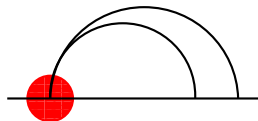


RNA secondary structure

Definition

List of base pairs i, j such that:

- only GC, AU or GU base pair
- only one base-pair per base
- minimum distance 3
- base pairs do not cross

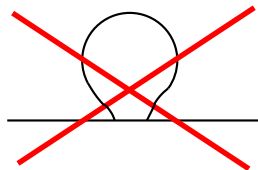


RNA secondary structure

Definition

List of base pairs i, j such that:

- only GC, AU or GU base pair
- only one base-pair per base
- minimum distance 3
- base pairs do not cross

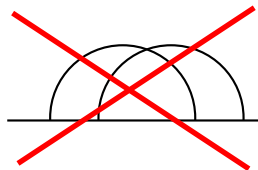


RNA secondary structure

Definition

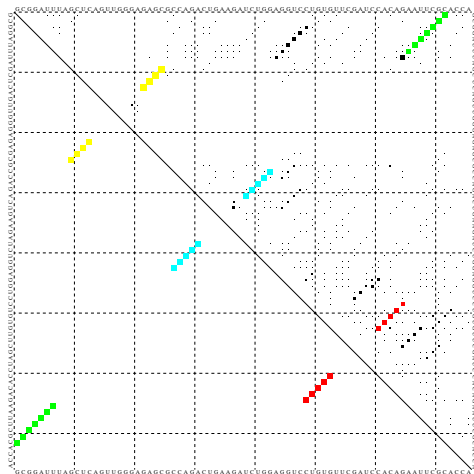
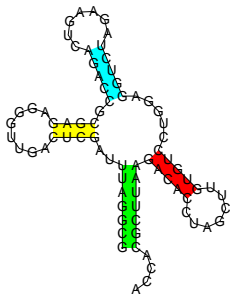
List of base pairs i, j such that:

- only GC, AU or GU base pair
- only one base-pair per base
- minimum distance 3
- base pairs do not cross



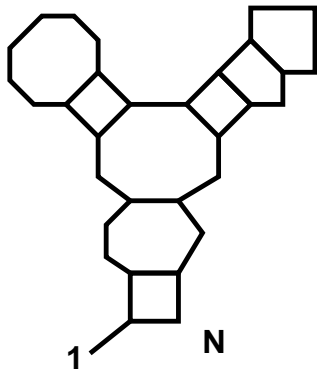
RNA secondary structure

Visualization



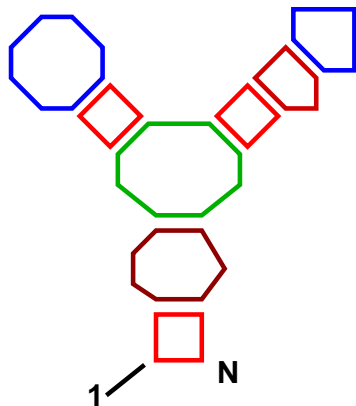
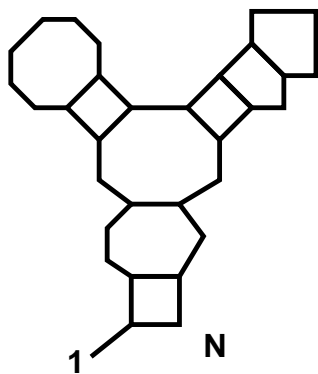
RNA secondary structure

Energy - Loop decomposition



RNA secondary structure

Energy - Loop decomposition



Hairpin-, Interior- and Multiloops

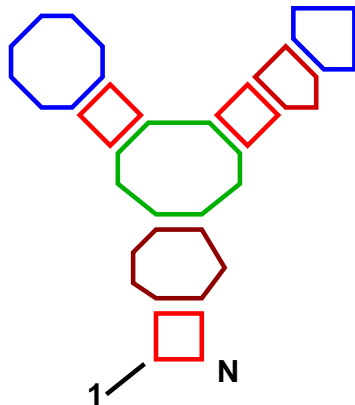
RNA secondary structure

Energy - Loop decomposition

Every Loop type is assigned an energy

- experimentally derived
- Energies are relative to open chain
- Energies are additive, independent

Hairpin-, Interior- and Multiloops



RNA secondary structure

Mfe structure prediction

- Structure of minimal energy for given sequence
- Loop energies are independent, additive
- Base pairs divide structure into inner and outer part
- Dynamic Programming

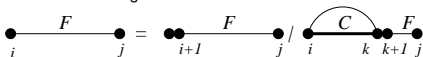
RNA secondary structure

Dynamic Programming Procedures

Dissect a problem into small, easy to solve sub-problems, solutions of sub-problems are tabulated

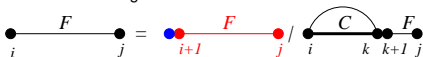
- Start with small sub-sequence
- Add bases one by one
- each base can either be unpaired or paired

Basic RNA Folding



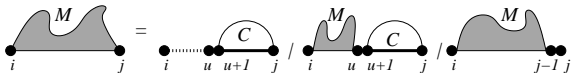
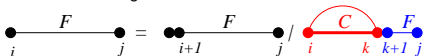
$$F_{ij} = \min\{F_{i+1,j}, \min_k C_{i,k} + F_{k+1,j}\}$$

Basic RNA Folding



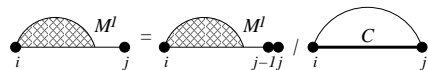
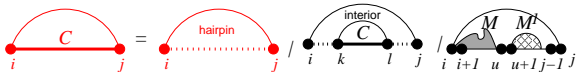
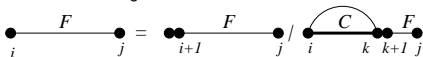
$$F_{ij} = \min\{F_{i+1,j}, \min_k C_{i,k} + F_{k+1,j}\}$$

Basic RNA Folding



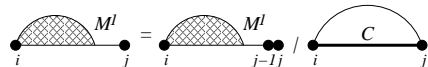
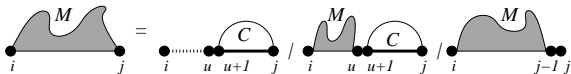
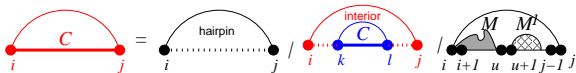
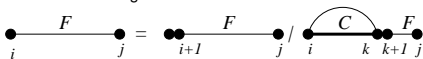
$$F_{ij} = \min\{F_{i+1,j}, \min_k C_{i,k} + F_{k+1,j}\}$$

Basic RNA Folding



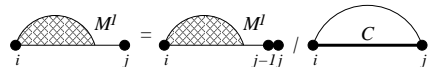
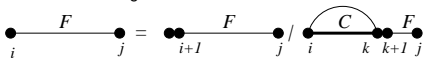
$$C_{i,j} = \min\{\mathcal{H}(i,j), \min_{k,l} \mathcal{I}(ij, kl) C_{k,l}, \min_u \mathcal{M}(ij) M_{i+1,u} M_{u+1,j-1}^1\}$$

Basic RNA Folding



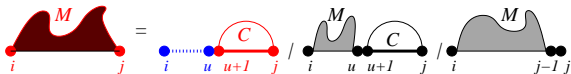
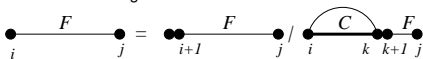
$$C_{i,j} = \min\{\mathcal{H}(i, j), \min_{k,l} \mathcal{I}(ij, kl) C_{k,l}, \min_u \mathcal{M}(ij) M_{i+1,u} M_{u+1,j-1}^1\}$$

Basic RNA Folding



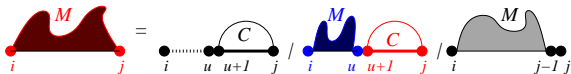
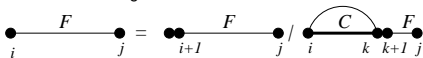
$$C_{i,j} = \min\{\mathcal{H}(i, j), \min_{k,l} \mathcal{I}(ij, kl) C_{k,l}, \min_u \mathcal{M}(ij) M_{i+1,u} M_{u+1,j-1}^1\}$$

Basic RNA Folding



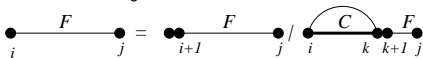
$$M_{i,j} = \min\{\min_u(u - i + 1)a + C_{u+1,j}, \min_u M_{i,u} + C_{u+1,j}, M_{i,j-1}a\}$$

Basic RNA Folding



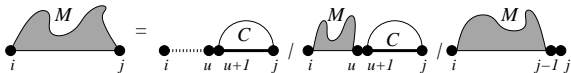
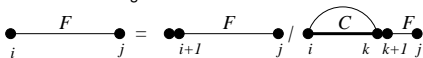
$$M_{i,j} = \min\{\min_u (u-i+1)a + C_{u+1,j}, \min_u M_{i,u} + C_{u+1,j}, M_{i,j-1} + a\}$$

Basic RNA Folding



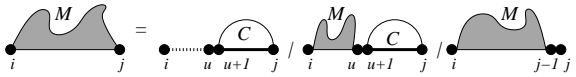
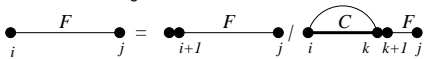
$$M_{i,j} = \min\{\min_u (u-i+1)a + C_{u+1,j}, \min_u M_{i,u} + C_{u+1,j}, M_{i,j-1} + a\}$$

Basic RNA Folding



$$M^l(i, j) = \min M^l(i, j-1) + C(i, j)$$

Basic RNA Folding



$$M^1(i, j) = \min M^1(i, j - 1) + C(i, j)$$

RNA secondary structure

Partition function

- Stacking energies in same range as thermal energy
- Probability of structure $s \propto e^{-\frac{E_s}{RT}}$
- $Q = \sum_s e^{-\frac{E_s}{RT}}$
- Probability of structure $s = e^{-\frac{E_s}{RT}} / Q$
- Any probability of a structural feature can be computed
- E.g. probability to get certain base pairs
- Use mfe decomposition, $\min \rightarrow \sum$, $+$ $\rightarrow *$

RNA secondary structure

Partition function

$$Q(i, j) = Q(i + 1, j) + \sum_{i < k \leq j} Q^B(i, k) Q(k + 1, j)$$

$$Q^B(i, j) = \mathcal{H}(i, j) + \sum_{i < k < l < j} \mathcal{I}(ij; kl) Q^B(k, l) + \\ + \sum_k M(i + 1, k) M^1(k + 1, j - 1)$$

$$M(i, j) = M(i + 1, j) + \sum_{i < k < j} Q^B(k, j) + \sum_{i < k < j} M(i, k - 1) Q^B(k, j)$$

$$M^1(i, j) = M^1(i, j - 1) + Q^B(i, j)$$

Outline

- 1 Introduction
 - RNA structure
 - RNA structure prediction
- 2 Bimolecular Secondary structures of RNA molecules
 - Introduction
 - Implementation
 - Application
- 3 Local partition function
 - RNAPfold
 - Results
 - Accessibility
 - Results

Bimolecular secondary structures of RNA Molecules

Biology

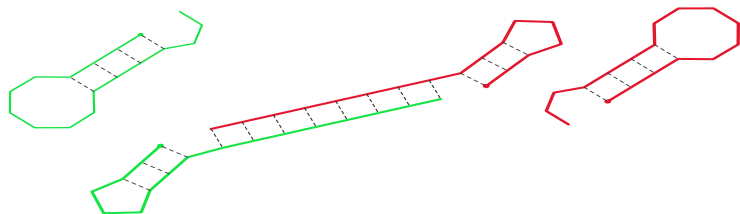
RNA RNA interaction as highly selective targeting

Examples:

- miRNA/siRNA – mRNA interaction
- snoRNA/rRNA interaction
- Hfq dependent regulation in prokaryotes, e.g. RyhB, OxyS
- Many ncRNAs of unknown function

Bimolecular secondary structures of RNA Molecules

Recursions

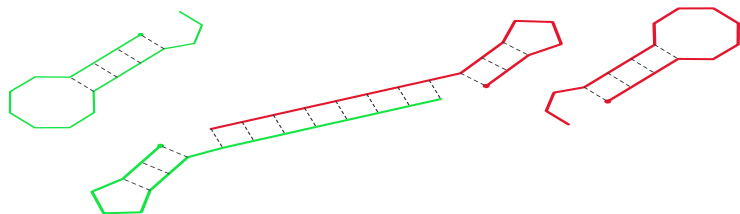


Computation of joint partition function:

- Concatenate Sequences
- Keep track of concatenation point (c_p)
- Treat loops containing c_p as exterior loops
- Add duplex initiation penalty
- Penalty added in post-processing step

Bimolecular secondary structures of RNA Molecules

Recursions



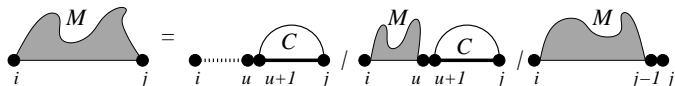
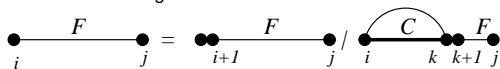
Computation of joint partition function:

- Concatenate Sequences
- Keep track of concatenation point (c_p)
- Treat loops containing c_p as exterior loops
- Add duplex initiation penalty
- Penalty added in post-processing step

Bimolecular secondary structures of RNA Molecules

Recursions

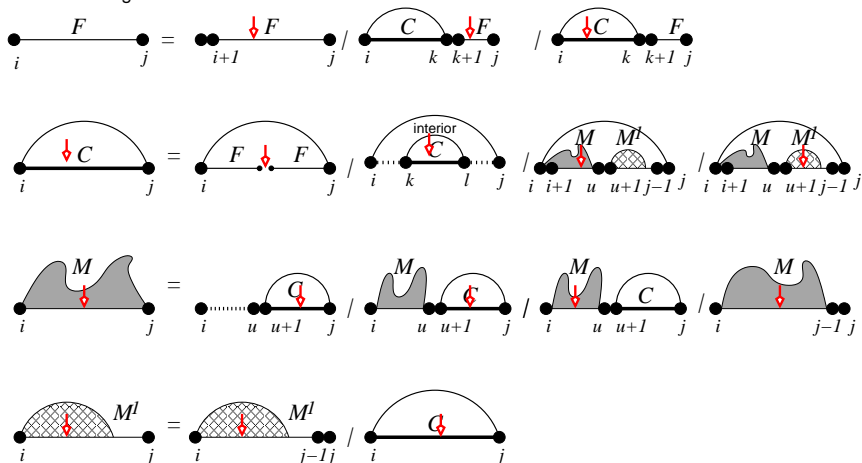
Basic RNA Folding



Bimolecular secondary structures of RNA Molecules

Recursions

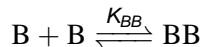
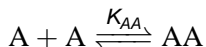
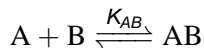
RNA Cofolding



Bimolecular secondary structures of RNA Molecules

Concentration Dependency

Two RNA molecules A , B with concentration $[A]_0$ and $[B]_0$, resp. give rise to 5 species:



Equilibrium constants computed out of partition functions:

$$K_{AB} = \frac{Q_{AB}}{Q_A Q_B} \quad K_{AA} = \frac{Q_{AA}}{Q_A Q_A} \quad K_{BB} = \frac{Q_{BB}}{Q_B Q_B}$$

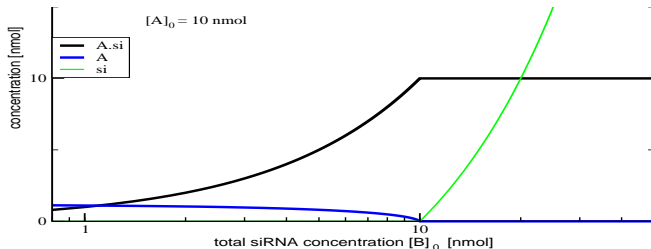
Bimolecular secondary structures of RNA Molecules

Concentration Dependency

- Use Mass conservation and equilibrium constants
- Generate two quadratic equations in two variables
- Use Newton's iteration method to solve system

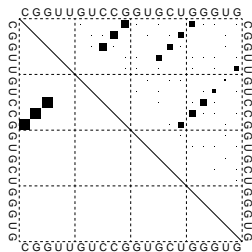
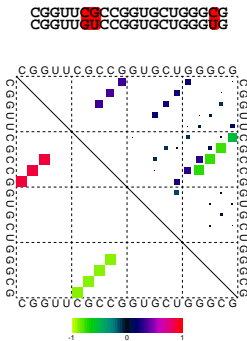
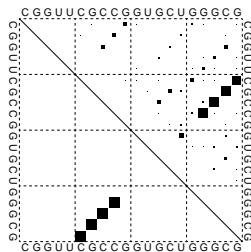
$$0 = f([A], [B]) := [A] + K_{AB}[A][B] + 2K_{AA}[A][A] - [A]_0$$

$$0 = g([A], [B]) := [B] + K_{AB}[A][B] + 2K_{BB}[B][B] - [B]_0$$



Case study: role of GU base pairs and miR efficacy

Difference dot plots



Case study: Role of GU base pairs and miR efficacy

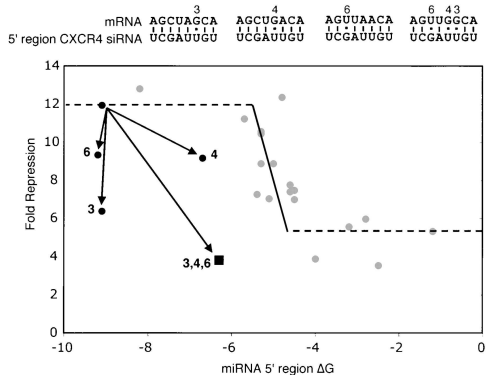
Doench and Sharp

- Investigated effects of mutations on miRNA efficacy
- Introducing GU base pairs reduced efficacy
- Loss of miRNA function not due to difference in binding energy
- Explained effect as due to GU base pairs unfavorable for siRNA function

Doench JG and Sharp PA; (2004) Genes Devel., 18:504-511

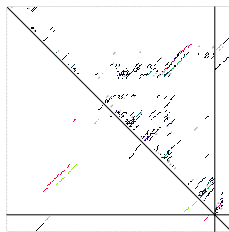
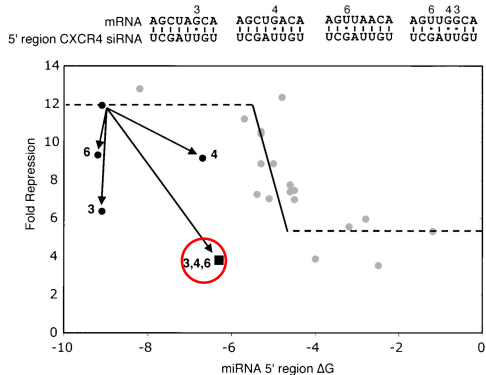
Case study: Role of GU base pairs and miR efficacy

Doench and Sharp



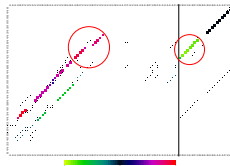
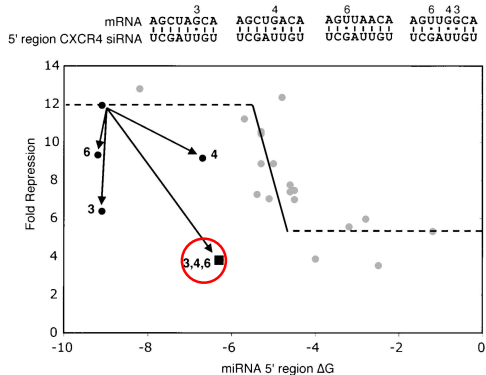
Case study: Role of GU base pairs and miR efficacy

Doench and Sharp



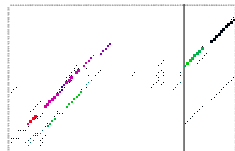
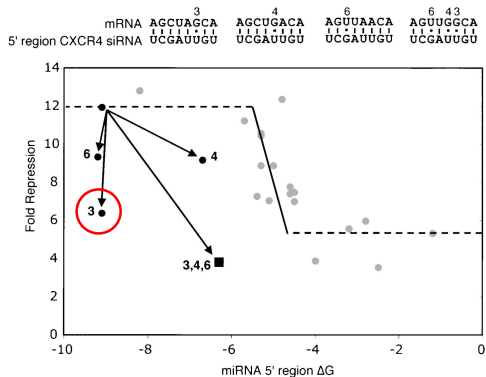
Case study: Role of GU base pairs and miR efficacy

Doench and Sharp



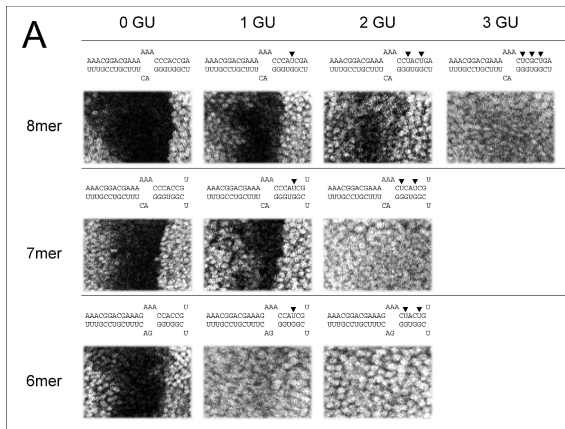
Case study: Role of GU base pairs and miR efficacy

Doench and Sharp



Case study: Role of GU base pairs and miR efficacy

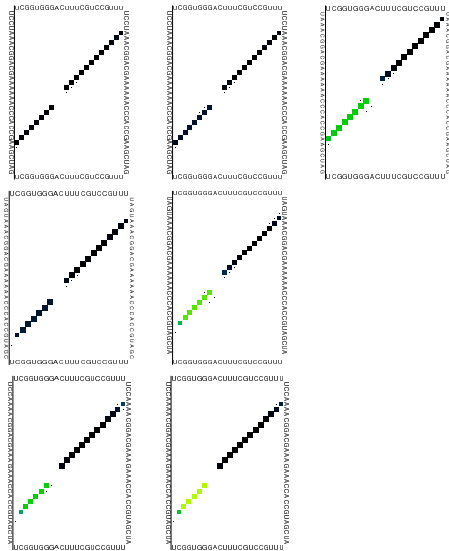
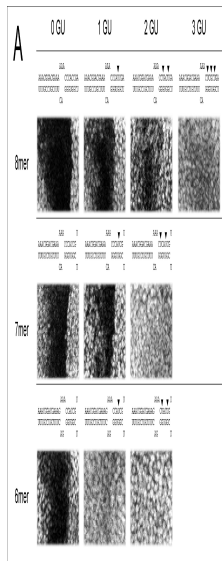
Brennecke et. al.



Brennecke J, Stark A, Russell RB, Cohen SM; (2005) PLoS Biol, 3(3):e85

Brennecke et. al.

Difference dot plot



Bimolecular secondary structures of RNA Molecules

Conclusion

- Sophisticated thermodynamical analysis in agreement with experimental data
- Can not explain reduced efficacy of all mutated binding sites
- No need to treat GU base pairs differently
- mRNA structure and therefore target site accessibility is important

Outline

- 1 Introduction
 - RNA structure
 - RNA structure prediction
- 2 Bimolecular Secondary structures of RNA molecules
 - Introduction
 - Implementation
 - Application
- 3 Local partition function
 - RNAPfold
 - Results
 - Accessibility
 - Results

Local partition function

Motivation

- RNA folding is $\mathcal{O}(n^3)$
- Huge amount of data
- Faster analysis tools
- Boundaries of transcripts often unknown
- Substructures of large Molecules (IRES, SECIS,...)
- Prediction of long range base pairs bad anyway

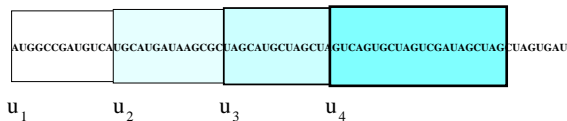
Local partition function

Motivation

- RNA folding is $\mathcal{O}(n^3)$
- Huge amount of data
- Faster analysis tools
- Boundaries of transcripts often unknown
- Substructures of large Molecules (IRES, SECIS,...)
- Prediction of long range base pairs bad anyway

Local partition function

sliding window approach



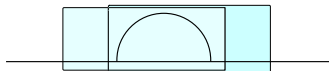
- Compute Q and pair probabilities for a stretch of length L , starting at u
- Increment u , compute partition function, continue, until end of molecule
- We compute all possible windows (i.e. increment 1)
- Reusing entries computed earlier, complexity is $\mathcal{O}(nL^2)$ instead of $\mathcal{O}(nL^3)$

Local partition function

Pair probabilities

What is the pair probability?

Local Pair probability is the mean probability over all possible windows a pair can be in:



$$\pi^L(i, j) = \frac{1}{l - (j - i) + 1} \sum_{u=j-L}^i p^{u, L}(i, j)$$

Computing all $p^{u, L}(i, j)$ is $\mathcal{O}(nL^3)$, but we can derive a recursion for the averages directly:

Local partition function

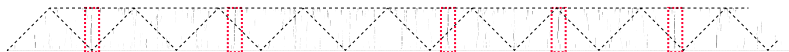
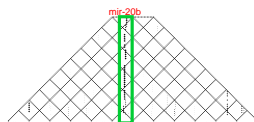
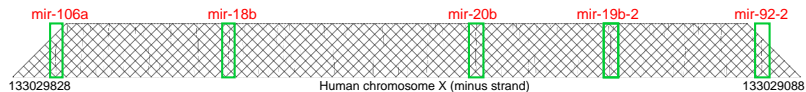
Pair probabilities

- i, j not enclosed by a base pair
- Within interior loop
- Within multi loop

$$\begin{aligned}\pi^L(i, j) &= \sum_{u=j-L}^i \frac{Q(u, i-1)Q^B(i, j)Q(j+1, u+L)}{Q^{u,L}(u, u+L)} \\ &+ \sum_{u \leq k < i; j < l \leq u+L} \pi^L(k, l) \frac{\mathcal{I}(kl, ij)}{Q^B(k, l)} \\ &+ \sum_{u \leq k < i; j < l \leq u+L} \pi^L(k, l) \frac{M(k+1, i-1) + M(j+1, l-1)}{Q^B(k, l)} \\ &+ \sum_{u \leq k < i; j < l \leq u+L} \pi^L(k, l) \frac{M(k+1, i-1)M(j+1, l-1)}{Q^B(k, l)}\end{aligned}$$

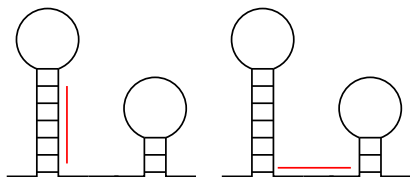
Results

miR cluster



green: annotated in Rfam (human), red : unannotated in homologue (dog)

Predicting Accessibility

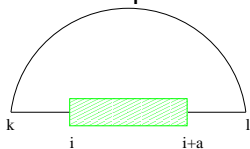


- Compute probability of a stretch of bases to be unpaired
- Equivalent to compute energy necessary to open a binding site
- Target site accessibility
- Local Version of RNAUp

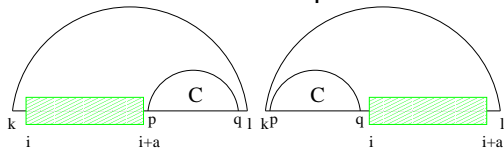
Predicting Accessibility



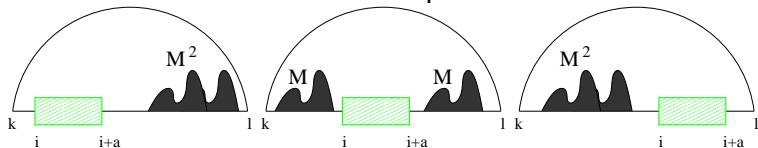
Hairpin:



Interior loop:

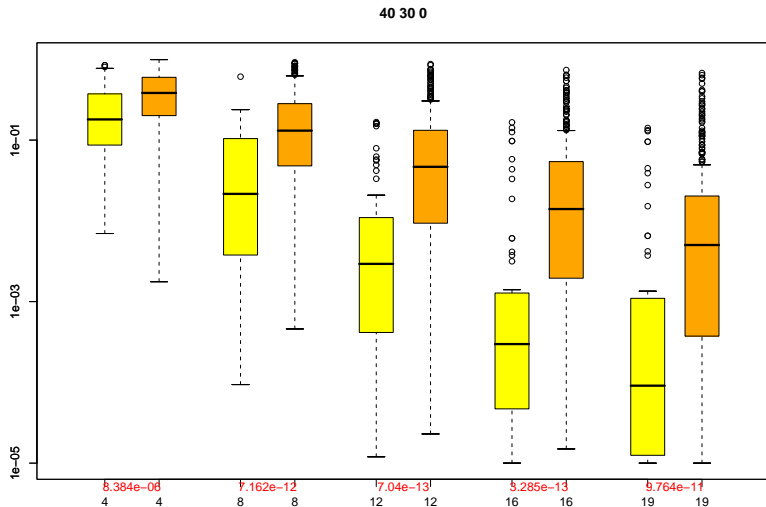


Multi loop:



Accessibility

Performance in separating non-working from working siRNAs



Over

I want to thank

Peter Schuster, Ivo Hofacker, Peter Stadler, Christoph Flamm,

Stefan Washietl, Dilimulati Yusufujiangaili, M T Wolfinger,
Hakim Tafer, Ulli Mückstein, Judith Invansits, Richard Neuböck,
W.A. Svrcek-Seiler, Andrea Tanzer, Rainer Machne, Caro
Thurner, Lukas Ender, Andreas Gruber, Jana Hertel, Stefanie
Widder, Roman Stocsits, Claudia Stocsits, Stefan Müller . . .

. . . and you for your attention

Consensus structures

- Often conservation on structural but not on sequence level
- Mutations can retain base pairs (consistent, compensatory)
- Predict structure of an alignment of RNA molecules
- Compute mean energy, add conservation score
- RNAalifold

Modifications of RNAalifold

- **Result can be biased**
- Include sequence weighting
- Gaps are scored like bases
- Use energy evaluation dependent on sequences

Modifications of RNAalifold

- Result can be biased
- Include sequence weighting
- Gaps are scored like bases
- Use energy evaluation dependent on sequences

Modifications of RNAalifold

- Result can be biased
- Include sequence weighting
- Gaps are scored like bases
- Use energy evaluation dependent on sequences

Modifications of RNAalifold

- Result can be biased
- Include sequence weighting
- Gaps are scored like bases
- Use energy evaluation dependent on sequences

Implementation

$$Q(i, j) = \sum_{s \in \text{Sequences}} Q(i, j, s)$$

Energy evaluation:

Implementation

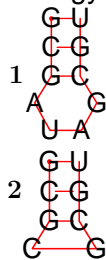
$$Q(i, j) = \sum_{s \in \text{Sequences}} w(s) Q(i, j, s)$$

Energy evaluation:

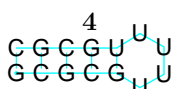
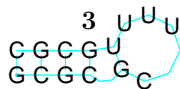
Implementation

$$Q(i, j) = \sum_{s \in \text{Sequences}} Q(i, j, s)$$

Energy evaluation:



	*****	*	*****	
sequence_2	AGCGUUCUUGCGC	--GUGUUUUUGCGCUUGCU		30
sequence_3	AGCGUUCUUGCGC	--GU--UUUUUGCGCUUGCU		28
sequence_1	AGCGUUCUUGCGAUA	AGCGUUUUUGCGCUUGCU		32
old	((((((.....(((.....)))).....))))).			-5.95
new	((((((.....((((.....)))).....))))).			-5.83



Implementation

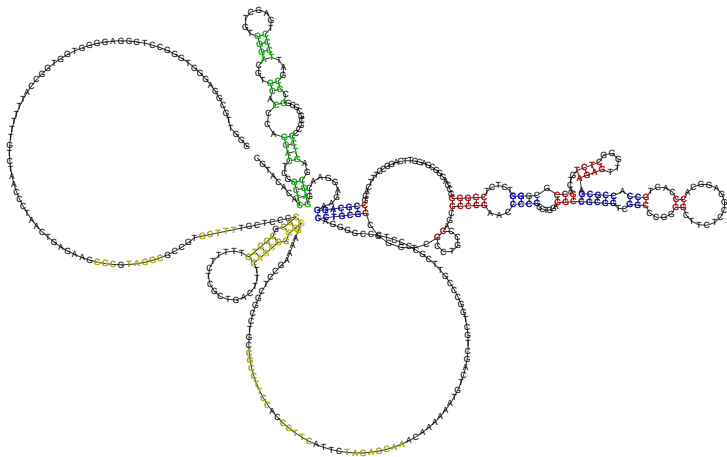
$$Q(i, j) = \sum_{s \in \text{Sequences}} Q(i, j, s)$$

Energy evaluation:

- Use length in sequence, not in alignment
- Use next position in sequence, not in alignment (for gaps)
- Translational arrays, arrays with 3', 5' neighbors

Results

Telomerase



blue,yellow: both wrong, red: both right, green: new right

Over

I want to thank Peter Schuster, Ivo Hofacker, Peter Stadler, Christoph Flamm, Stefan Washielt, Dilimulati Yusufjuangaili, M T Wolfinger, Hakim Tafer, Ulli Mückstein, Judith Invansits, Richard Neuböck, W.A. Svrcek-Seiler, Andrea Tanzer, Rainer Machne, Caro Thurner, Lukas Endler, Andreas Gruber, Jana Hertel, Dirk Stermann. . .
. . . and you for your attention