Sequence-Structure Relations of Single RNA Molecules and Cofolded RNA Complexes

Ulrike Mückstein

Institute for Theoretical Chemistry
University of Vienna

Vienna, 2005

Outline

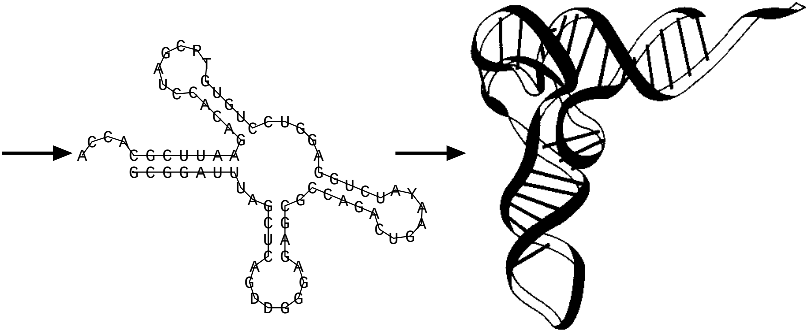
Motivation
- Folding RNA
- Overview

Results
- RNA sequence to structure mapping
- Co-folding two RNA molecules

Discussion

RNA primary, secondary and tertiary structure

GCGGGAAUAGCUCAGUUGGUAGAGCACGACCUUGCCAAGGUCGGGGUCGCGAGUUCGAGUCUCGUUUCCCGCUCCA



Outline

- Motivation
    - Folding RNA
    - Overview
- Results
    - RNA sequence to structure mapping
    - Co-folding two RNA molecules
- Discussion

Relation between RNA sequences and their secondary structures

- RNA sequence to structure mapping
- Evolutionary dynamics of an RNA population in a flow reactor
- RNA-RNA interaction by co-folding two RNA molecules

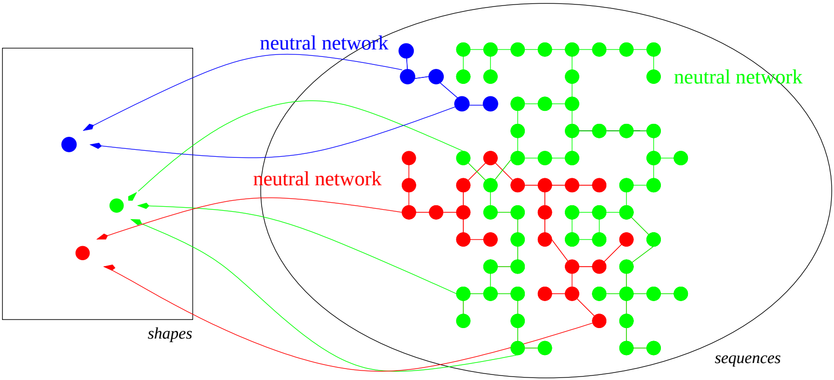
Relation between RNA sequences and their secondary structures

- RNA sequence to structure mapping
- Evolutionary dynamics of an RNA population in a flow reactor
- RNA-RNA interaction by co-folding two RNA molecules

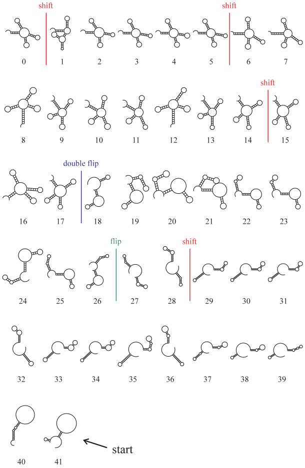
Relation between RNA sequences and their secondary structures

- RNA sequence to structure mapping
- Evolutionary dynamics of an RNA population in a flow reactor
- RNA-RNA interaction by co-folding two RNA molecules

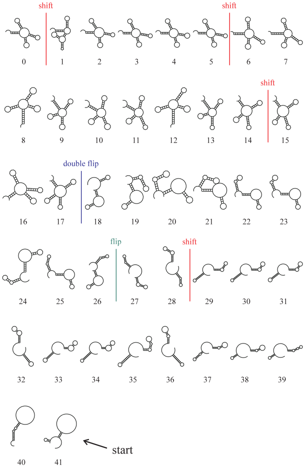
RNA sequence to structure mapping



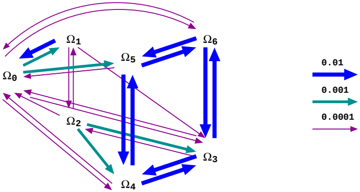
Evolutionary dynamics in a flow reactor



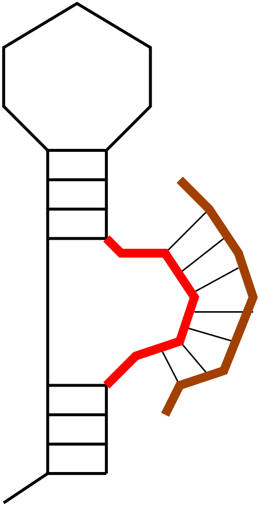
Evolutionary dynamics in a flow reactor



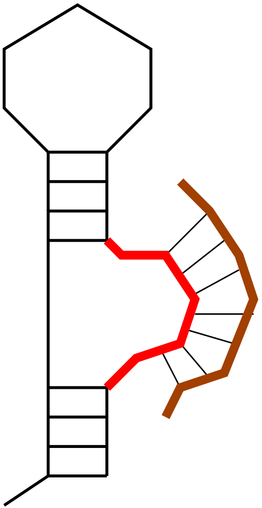
kind of tansition 🟩 continous 🟨 discontinous		distance ham
`(((((...((((........)))).(((((.......))))).....(((((.......))))).))))))....`	Ω_0	0
`((.(((...((((........)))).(((((.......))))).....(((((.......))))).))).))....`	Ω_1	2
`((.(((...((((........)))).(((((.......))))).....(((((.......)))))..)))))....`	Ω_2	2
`.(((((...((((........)))).(((((.......))))).....(((((.......)))))..))))).…`	Ω_3	2
`((((((...((((........)))).(((((.......))))).....(((((.......)))))..))))))...`	Ω_4	2
`(((((((..((((........)))).(((((.......))))).....(((((.......))))).)))))))...`	Ω_5	2
`.((((((..((((........)))).(((((.......))))).....(((((.......))))).))))))....`	Ω_6	2

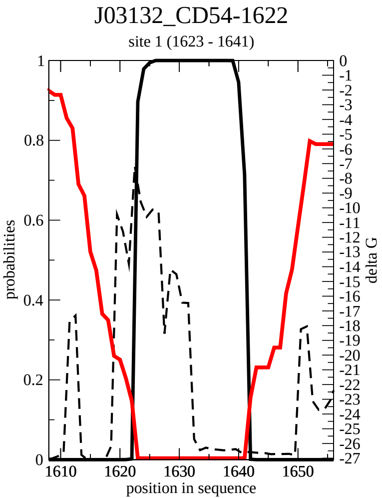


Co-folding two RNA molecules



Co-folding two RNA molecules





J03132_CD54-1622

site 1 (1623 - 1641)

RNAup

Outline

Motivation
  Folding RNA
  Overview

Results
  RNA sequence to structure mapping
  Co-folding two RNA molecules

Discussion

Generic properties of RNA folding

- More sequences than structures
- Few common and many rare structures
- Common structures form extended neutral networks
- Shape space covering

Generic properties of RNA folding

- More sequences than structures
- Few common and many rare structures
- Common structures form extended neutral networks
- Shape space covering

Generic properties of RNA folding

- More sequences than structures

- Few common and many rare structures

- Common structures form extended neutral networks

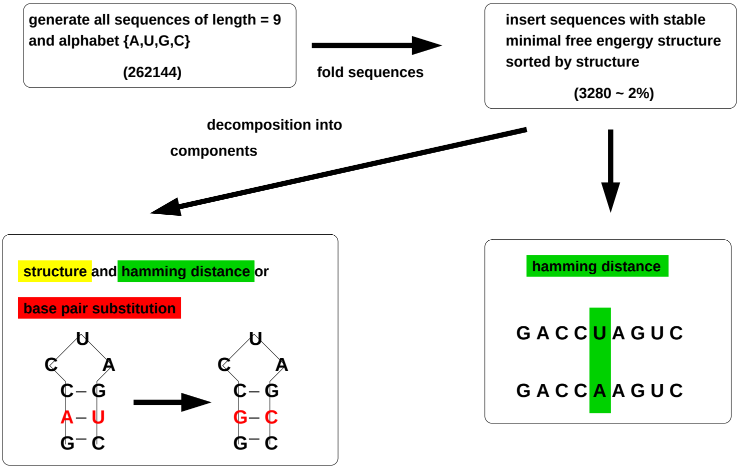
- Shape space covering

Generic properties of RNA folding

- More sequences than structures
- Few common and many rare structures
- Common structures form extended neutral networks
- Shape space covering

Exhaustive folding and enumeration of the sequence spaces $\mathcal{I}_{AUGC}^{(\ell=9)}$ and $\mathcal{I}_{AUGC}^{(\ell=10)}$

PostgreSQL Database



generate all sequences of length = 9 and alphabet {A,U,G,C}

(262144)

fold sequences

insert sequences with stable minimal free engergy structure sorted by structure

(3280 ~ 2%)

decomposition into components

structure and hamming distance or

base pair substitution

hamming distance

GACCUAGUC

GACCAAGUC

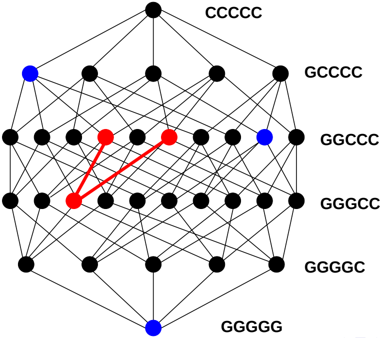
Stored Procedures **Vigl Graph Library** **Vienna RNA package**

More sequences than structures

$$3^\ell \text{ structures} < 4^\ell \text{ sequences}$$

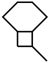
`(((...)))`   `GCGAUGCGC`

- $\mathcal{I}^{(\ell=9)}_{AUGC}$
  - stable mfe structure $3239 \sim 2.3\%$
  - 4 different structures
- $\mathcal{I}^{(\ell=10)}_{AUGC}$
  - stable mfe structure $40345 \sim 3.8\%$
  - 9 different structures



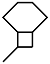
Few common and many rare structures

$\mathcal{I}^{(\ell=9)}_{AUGC}$



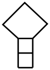
`((....)).`

73.2%



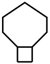
`.((....))`

15.5%



`(((...)))`

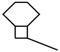
9.9%



`((.....))`

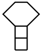
1.4%

$\mathcal{I}^{(\ell=10)}_{AUGC}$



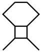
`((....))..`

22.8%



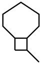
`(((....)))`

20.6%



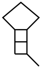
`.((....)).`

19.7%



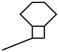
`((.....)).`

16.7%



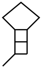
`(((...))).`

7.4%



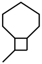
`..((....))`

4.7%



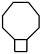
`.(((...)))`

4.1%



`.((.....))`

2.0%

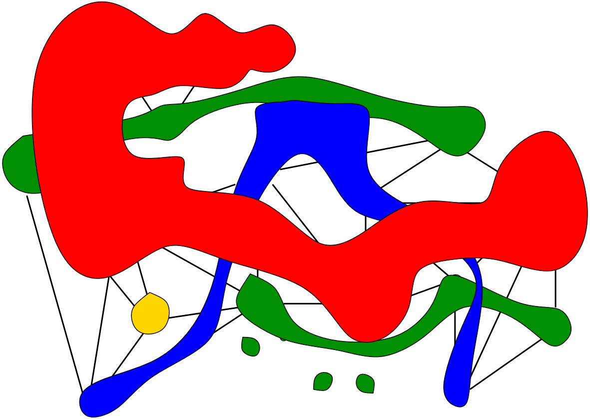


`((......))`

1.9%

Common structures form extended neutral networks

((....)).	**73.2%**
.((....))	**15.5%**
(((...)))	**9.9%**
((.....))	**1.4%**



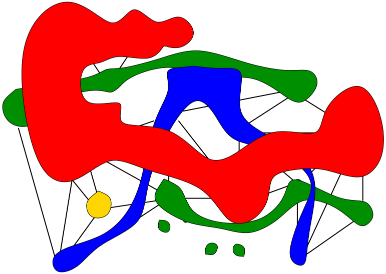
Shape space covering

((....)).

.((....))

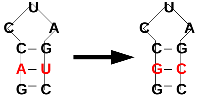
(((...)))

((.....))



structure and hamming distance or

base pair substitution



hamming distance

G A C C U A G U C

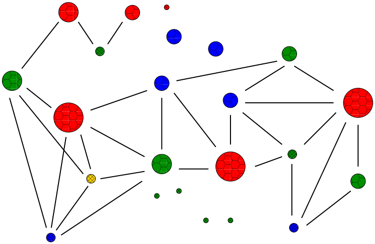
G A C C A A G U C

((....)).

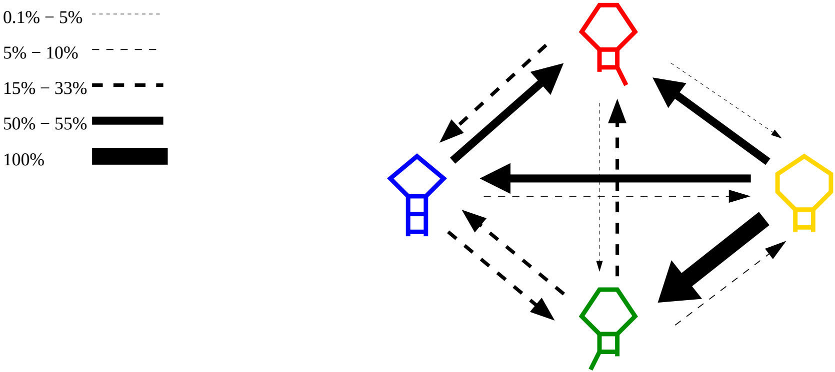
.((....))

(((...)))

((.....))



Shape space topology induced by the sequence-structure map



Discussion

- study topology of RNA sequence to structure mapping by exhaustive folding and enumeration of sequence spaces
- extended neutral networks allow a population of RNA molecules to explore sequence space by neutral drift
- asymmetric accessibility relations support evolutionary innovation

Discussion

- study topology of RNA sequence to structure mapping by exhaustive folding and enumeration of sequence spaces
- extended neutral networks allow a population of RNA molecules to explore sequence space by neutral drift
- asymmetric accessibility relations support evolutionary innovation

Discussion

- study topology of RNA sequence to structure mapping by exhaustive folding and enumeration of sequence spaces
- extended neutral networks allow a population of RNA molecules to explore sequence space by neutral drift
- asymmetric accessibility relations support evolutionary innovation

Outlook

- functional RNAs work by interacting with other RNAs
- Camille Stephan-Otto Attolini et al. studied the cofolding map of interacting RNA sequences:
  - the cofolding map admits large neutral networks and long neutral paths similar to the folding map of single RNA molecules
- tool to compute RNA-RNA interaction by co-folding two RNA molecules

Outlook

- functional RNAs work by interacting with other RNAs
- Camille Stephan-Otto Attolini et al. studied the cofolding map of interacting RNA sequences:
  - the cofolding map admits large neutral networks and long neutral paths similar to the folding map of single RNA molecules
- tool to compute RNA-RNA interaction by co-folding two RNA molecules

Outlook

- functional RNAs work by interacting with other RNAs
- Camille Stephan-Otto Attolini et al. studied the cofolding map of interacting RNA sequences:
    - the cofolding map admits large neutral networks and long neutral paths similar to the folding map of single RNA molecules
- tool to compute RNA-RNA interaction by co-folding two RNA molecules

Outlook

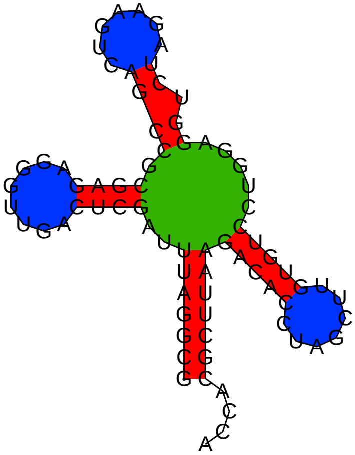
- functional RNAs work by interacting with other RNAs
- Camille Stephan-Otto Attolini et al. studied the cofolding map of interacting RNA sequences:
  - the cofolding map admits large neutral networks and long neutral paths similar to the folding map of single RNA molecules
- tool to compute RNA-RNA interaction by co-folding two RNA molecules

Outline

- Motivation
  - Folding RNA
  - Overview
- Results
  - RNA sequence to structure mapping
  - Co-folding two RNA molecules
- Discussion

Outline

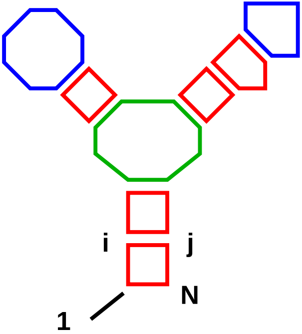
RNA Secondary Structure



- Hairpin Loops
- Interior Loops
- Multiloops
- Exterior Loop

Loop decomposition of RNA Secondary Structure

- $F(S) = \sum_{L \in S} F_L$.



Basic Algorithms

- McCaskill: Dynamic programming algorithm for calculation of the full equilibrium partition function Z for RNA secondary structure in $O(N^3)$

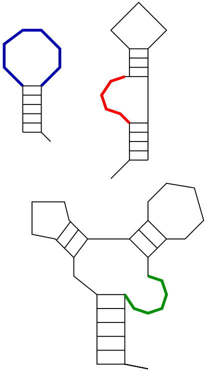
- Vienna RNA Package

Basic Algorithms

- McCaskill: Dynamic programming algorithm for calculation of the full equilibrium partition function Z for RNA secondary structure in $O(N^3)$
- Vienna RNA Package

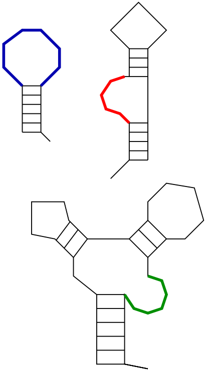
Complete ensemble of secondary structures

Probability of an unstructured region

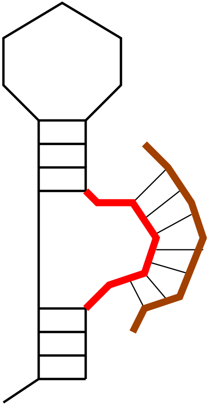


Complete ensemble of secondary structures

Probability of an unstructured region



Probability of interaction in unstructured regions



Information provided by our algorithm

- structural context of a binding site
  - $P_u[i,j]$ region $[i,j]$ contains no secondary structure
- location of possible binding sites
  - $P^*[i,j]$ probability of a regional interaction
- energetics of RNA-RNA interaction
  - $\Delta G = \Delta G_u + \Delta G_h$

Information provided by our algorithm

- structural context of a binding site
  - $P_u[i,j]$ region $[i,j]$ contains no secondary structure
- location of possible binding sites
  - $P^*[i,j]$ probability of a regional interaction
- energetics of RNA-RNA interaction
  - $\Delta G = \Delta G_u + \Delta G_h$

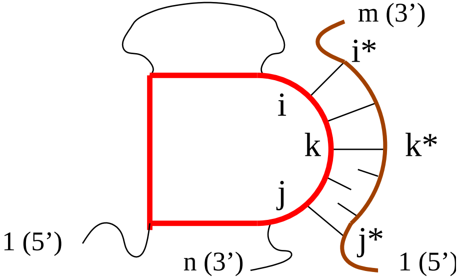
Information provided by our algorithm

- structural context of a binding site
  - $P_u[i,j]$ region $[i,j]$ contains no secondary structure
- location of possible binding sites
  - $P^*[i,j]$ probability of a regional interaction
- energetics of RNA-RNA interaction
  - $\Delta G = \Delta G_u + \Delta G_h$

$P_u[i,j]$ region $[i,j]$ contains no secondary structure

$$P_u[i,j] = \underbrace{\frac{Z[1,i-1]\times 1\times Z[j+1,N]}{Z[1,N]}}_{\text{exterior}} + \underbrace{\sum_{p<i\le j<q} P_{pq}\frac{Z_{pq}[i,j]}{Z^b[p,q]}}_{\text{enclosed}}.$$

Z^I Interaction of region $[i, j]$ with region $[i^*, j^*]$



$$Z^I[i, j, i^*, j^*] = \sum_{\substack{i<k<j \\ i^*>k^*>j^*}} Z^I[i, k, i^*, k^*] e^{-\frac{1}{kT} I(k, k^*; j, j^*)}.$$

probability and optimal free energy of an interaction

- conditional probability given that two molecules bind at all

$$P^*[i,j] = \frac{Z^*[i,j]}{\sum_{k<l} Z^*[k,l]}. \qquad Z^*[i,j] = P_u[i,j] \sum_{i^*>j^*} Z^I[i,j,i^*,j^*].$$

- optimal free energy of a regional interaction

$$\Delta G_i = min_{k \le i \le l}\{\Delta G[k,l]\}. \qquad \Delta G[i,j] = (-kT)\ln Z^*[i,j].$$

probability and optimal free energy of an interaction

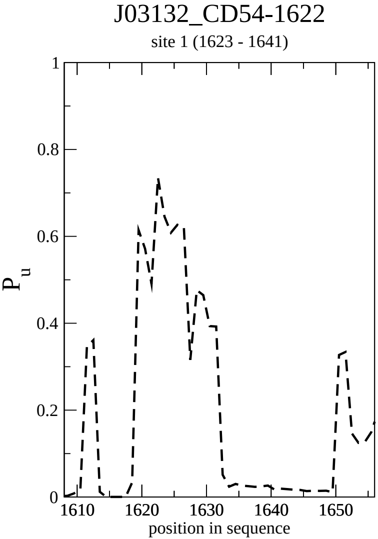
- conditional probability given that two molecules bind at all

$$P^*[i,j] = \frac{Z^*[i,j]}{\sum_{k<l} Z^*[k,l]}. \qquad Z^*[i,j] = P_u[i,j] \sum_{i^*>j^*} Z^I[i,j,i^*,j^*].$$

- optimal free energy of a regional interaction

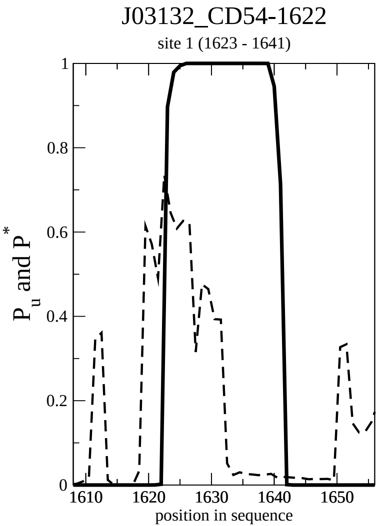
$$\Delta G_i = min_{k \leq i \leq l}\{\Delta G[k,l]\}. \qquad \Delta G[i,j] = (-kT) \ln Z^*[i,j].$$

Interaction of CD54 mRNA with siRNA si1622



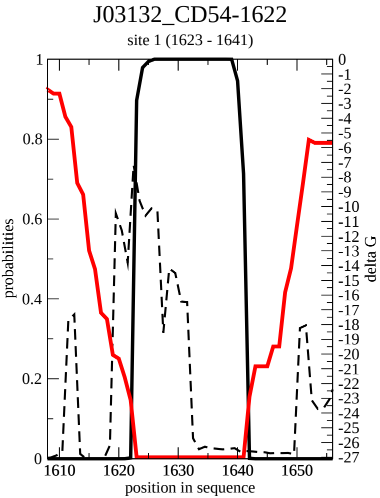
Vickers et. al. J.Biol.Chem., 278(9):7108-118,2003

Interaction of CD54 mRNA with siRNA si1622



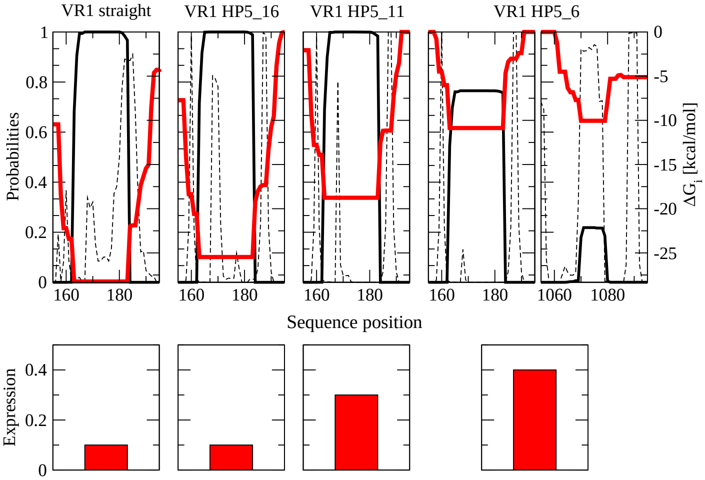
Vickers et. al. J.Biol.Chem., 278(9):7108-118,2003

Interaction of CD54 mRNA with siRNA si1622



Vickers et. al. J.Biol.Chem., 278(9):7108-118,2003

Local RNA target structure influences siRNA efficacy



Schubert et.al. J.Mol.Biol.,348(4):883-93,2005

Discussion

- Realistic description of interaction between small RNAs and their targets
- Structural context of a possible binding site
- Location of possible binding sites
- Energetics of RNA-RNA interaction
- Time complexity is $O(N^3)$, memory requirement $O(N^2)$

Discussion

- Realistic description of interaction between small RNAs and their targets
- Structural context of a possible binding site
- Location of possible binding sites
- Energetics of RNA-RNA interaction
- Time complexity is $O(N^3)$, memory requirement $O(N^2)$

Discussion

- Realistic description of interaction between small RNAs and their targets
- Structural context of a possible binding site
- Location of possible binding sites
- Energetics of RNA-RNA interaction
- Time complexity is $O(N^3)$, memory requirement $O(N^2)$

Discussion

- Realistic description of interaction between small RNAs and their targets
- Structural context of a possible binding site
- Location of possible binding sites
- Energetics of RNA-RNA interaction
- Time complexity is $O(N^3)$, memory requirement $O(N^2)$

Discussion

- Realistic description of interaction between small RNAs and their targets
- Structural context of a possible binding site
- Location of possible binding sites
- Energetics of RNA-RNA interaction
- Time complexity is $O(N^3)$, memory requirement $O(N^2)$

Thanks to

Peter Schuster
Peter Stadler
Ivo Hofacker
Christoph Flamm
Stephan Bernhart
Hakim Tafer
Andrea Tanzer
Camille Attolini