

Prediction of RNA-RNA interaction

Ivo Hofacker

Institute for Theoretical Chemistry, University of Vienna

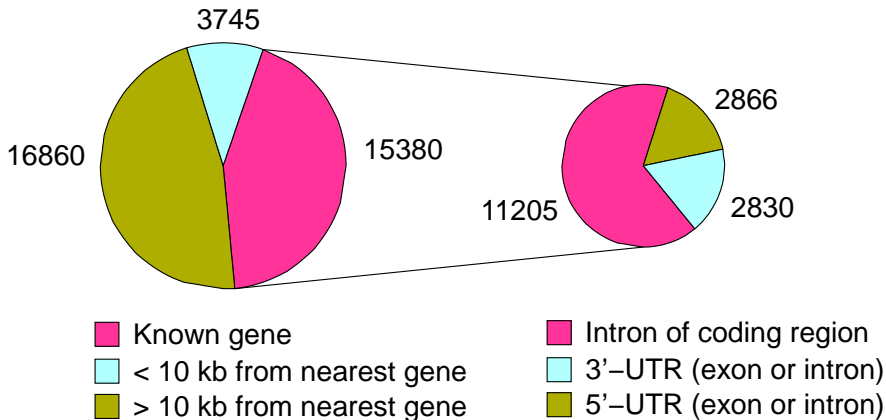
<http://www.tbi.univie.ac.at/~ivo/>

IMA, Minneapolis, October 2007

Motivation

- Experimental and bioinformatical methods find novel ncRNAs *en masse*
- Give no hint as to the function of these novel ncRNAs
- Functional characterization of ncRNAs is difficult and slow
- Most ncRNAs function through interaction with other RNAs
- Identification of interaction partners is the easiest approach to learn about possible functions
- Most obvious in the case of miRNA target prediction

Example: RNAz predicted ncRNAs in the Human Genome



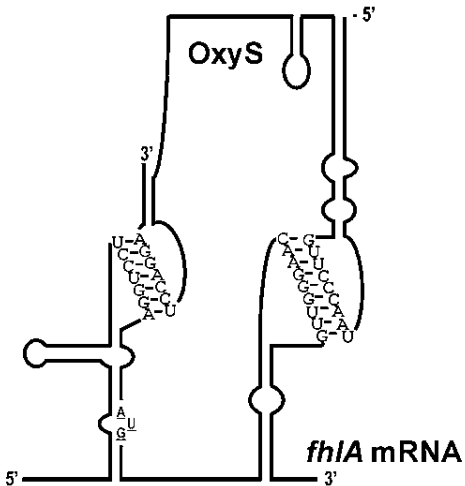
Tens of thousands predicted ncRNAs. Less than 2% of predictions are homologous to known ncRNAs.

Well known Examples of RNA-RNA Interaction

- micro RNAs regulate mRNA translation
- snoRNAs guide methylation and pseudouridylation of rRNA
- some well studied bacterial examples
 - RyhB is transcribed under low Fe, binds several mRNA of Fe binding proteins (sdh, sodB) and leads to mRNA degradation
 - GadY interacts with the 3' UTR of GadX and inhibits its degradation
 - DsrA is expressed at low temperatures and stimulates the translation of RpoS a translational regulator
 - OxyS is expressed under oxidative stress and inhibits translation of its targets RpoS and flhA
 - T-box motifs bind uncharged tRNAs to control transcription of aminoacyl synthetases

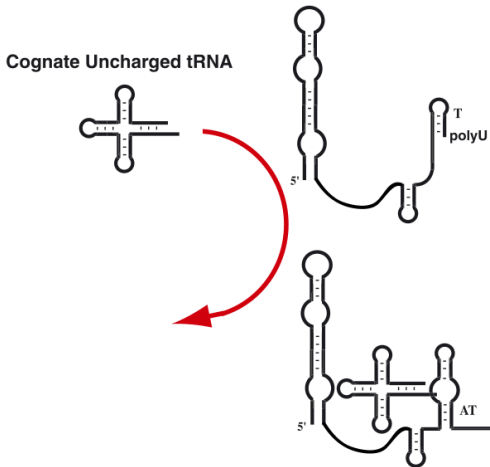
Interaction of OxyS and *fhIA*

Binding of OxyS to *fhIA* mRNA makes the ribosome binding site (start codon) inaccessible



Transcriptional control by T-box Motifs

Concentration of un-charged tRNAs controls transcription of its aminoacyl synthetase



Challenges

- Few well-studied examples
- Energetics of many interaction motifs are unknown
- Length of the interacting region is often quite small
- Binding is a concentration dependent process
- Folding kinetics rather than thermodynamics may play a role
- A single small RNA may have many targets
- RNA chaperones such as Hfq may be required for binding
- ncRNAs often act within RNPs, what's the influence of the protein?

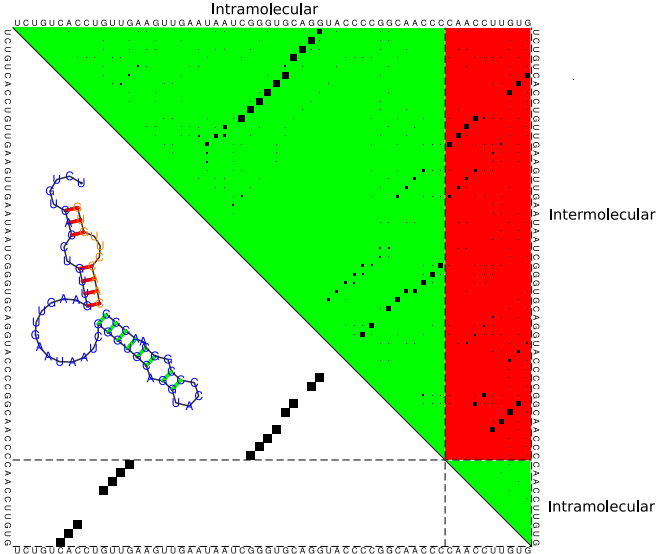
Overview of Prediction Strategies

- Co-folding by concatenation of two sequences, e.g. RNAcofold, pairfold, DINAMELT, Nupack
- Co-folding with pseudoknot-like structures, IRIS
- Using only inter-molecular interaction, i.e. assume that both molecules are unstructured by themselves. RNAhybrid, RNAduplex, codeRNAplex
- Combine interaction search with accessibility calculations. RNAup, RNApIfold + RNAplex, oligowalk

Simple Co-folding of two RNAs

- Poor man's approach to cofolding:
 - Concatenate two RNAs using a short linker
 - Use conventional folding programs such as mfold
- Proper way:
 - Use modified folding algorithm that keeps track of the break between the strands
 - Any loop containing the break point is treated specially.
 - Implemented in the RNAcifold program of the Vienna RNA package
- Limited to structures that are pseudo-knot free for concatenated sequences.

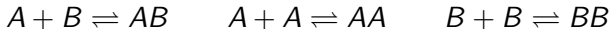
Pair Probabilities from RNAcofold



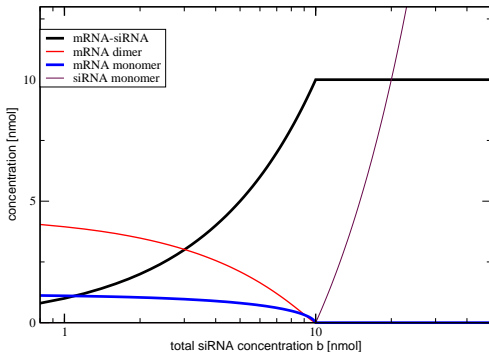
Concentration Dependence of RNA-RNA interactions

Binding processes are always concentration dependent

For two RNAs we have three reactions in equilibrium:

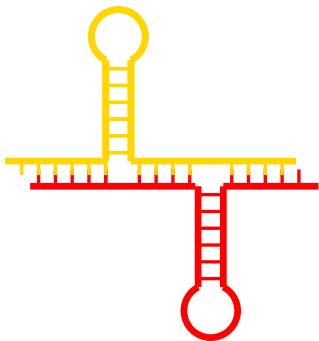


Compute concentrations of all five monomers and dimers.

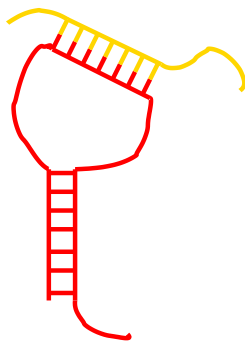


Structures (not) Predicted by RNAcofold

knot-free



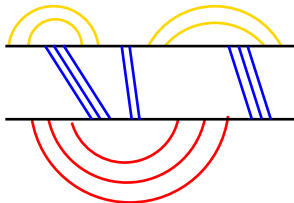
pseudo-knotted



Predicting more complex Structures

Without restricting allowed structure motif RNA-RNA interaction is NP-complete

- The most general algorithms (Alkan 2006, Pervouchine 2004) allow structures where
 - Intra-molecular pairs form pseudo-knot free structures
 - Inter-molecular pairs are not allowed to cross
- Run time is too slow for most purposes ($\mathcal{O}(n^3 \cdot m^3)$)



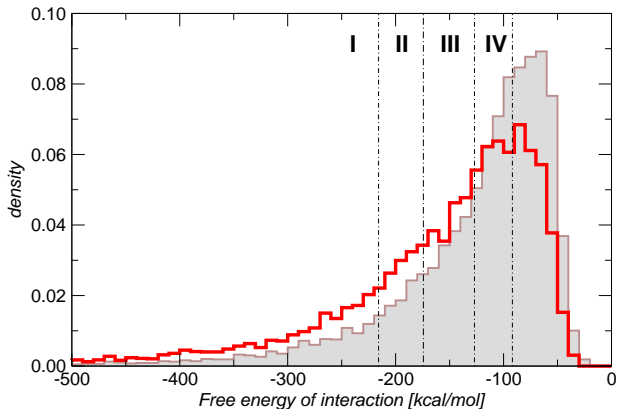
Fast Interaction Search

Methods for fast interaction search

- Search for sequence complementarity by BLAST
- Better: Interaction search using thermodynamics
- Simplified folding algorithm without intra-molecular pairs.
- Runs in $\mathcal{O}(n \cdot m)$ time.
- Used in RNAhybrid (miRNA target prediction), RNAduplex, RNAplex

What's the effect of neglecting intra-molecular structure?

Frequency of ncRNA - mRNA Interactions



RNA-mRNA interaction interaction energies (from RNAduplex)
red: ncRNA candidates from RNAz, grey: shuffled sequences.
Enrichments relative to randomly chosen conserved regions:
I: 2.3, II: 1.9, III: 1.4, IV: 1.1

Combining Interaction and Accessibility

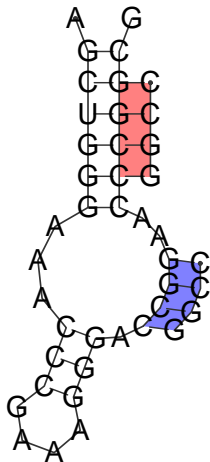
Two ingredients for efficient hybridization

- Complementarity
- Accessibility

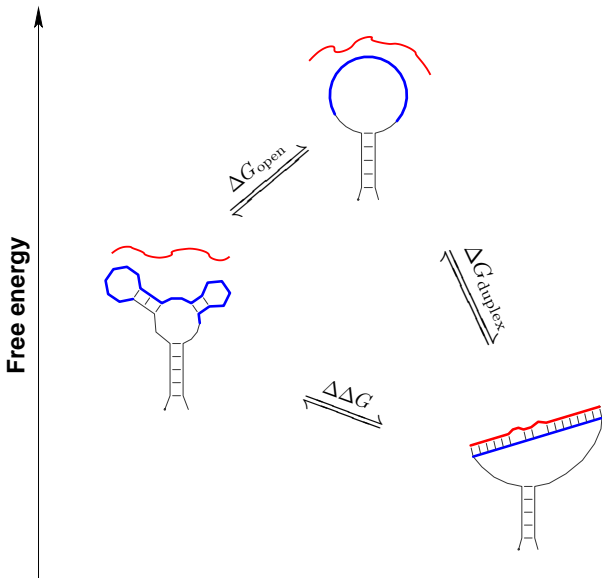
How to quantify these?

Complementarity \rightarrow interaction energy

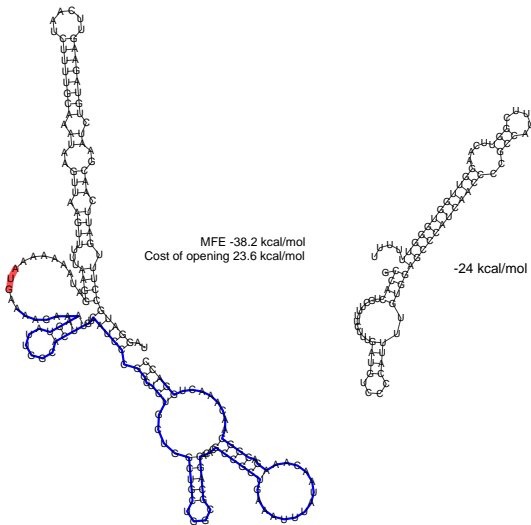
Accessibility \rightarrow probability to be unpaired



RNA Hybridization as a two Step Process

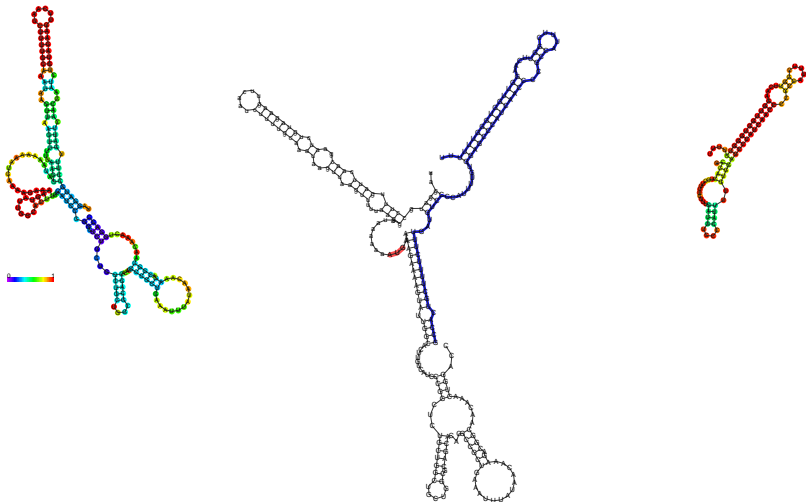


Example: ompN and RybB



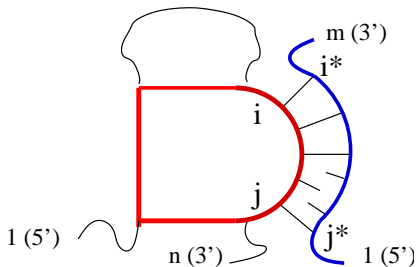
```
GCCAC-----TGCTTTTCTTTGATGTCCCCATTTT-GTGGA-----GC-CCATCAACCCGCCATTTTCGGTT---CAAG-GTTGGTGGGTTTTT  
|||      |||  |||||  |||  |||||  |||||      |||  |||  ||  ||  ||  |||||      ||||  ||  |||  |||||  -40.30  
AGGTCAAACAACGGC-AGAAACAATATT--TAAAGTCGCCGCACACGACGGGTCGTCGGT-CGTCTCGGCCCTACTGTTACGGTTATGAAAAGAAACC-3'
```

Example: ompN and RybB



$$\Delta G_{\text{open}} = 1.6 + 3.9 \text{ kcal/mol}, \quad \Delta\Delta G = -16 \text{ kcal/mol}$$

The RNAup Approach



- Compute probability that a site at $[i..j]$ is unpaired (equivalent to the energy ΔG_{open} needed to force it open).
- Consider all possible ways of binding to the region $[i..j]$ to compute the interaction energy $\Delta G_{\text{interact}}$
- Total binding energy is the sum of these contributions:
$$\Delta\Delta G = \Delta G_{\text{open}} + \Delta G_{\text{interact}}$$
- Currently, restrict interactions to a single region

Computing Accessibility

ΔG_{open} is equivalent to the probability that the region $[i..j]$ is unpaired in equilibrium $\Delta G_{open} = -RT \ln P^u[i,j]$

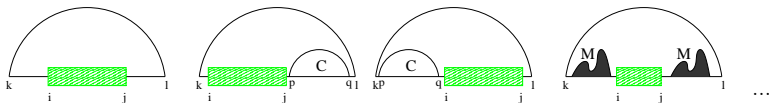
- Constrained folding $\Delta G_{open} = \Delta G^{constr} - \Delta G^{free}$
- Boltzmann sampling, works for short regions only
- Direct computation by modified folding algorithm

Computing Accessibility

ΔG_{open} is equivalent to the probability that the region $[i..j]$ is unpaired in equilibrium $\Delta G_{open} = -RT \ln P^u[i, j]$

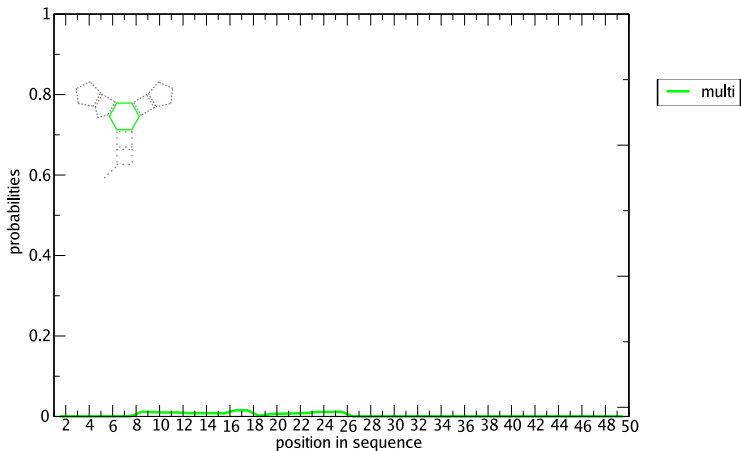
- Constrained folding $\Delta G_{open} = \Delta G^{constr} - \Delta G^{free}$
- Boltzmann sampling, works for short regions only
- Direct computation by modified folding algorithm

$$P^u[i, j] = \frac{Z_{1, i-1} Z_{j+1, n}}{Z_n} + \sum_{h < i, j < l} p_{h, l} \cdot \text{Prob}([i, j] | (k, l))$$



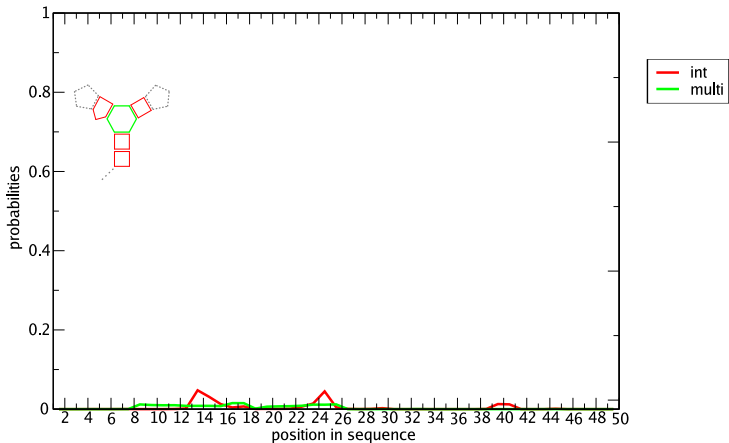
RNAup

Structural Information



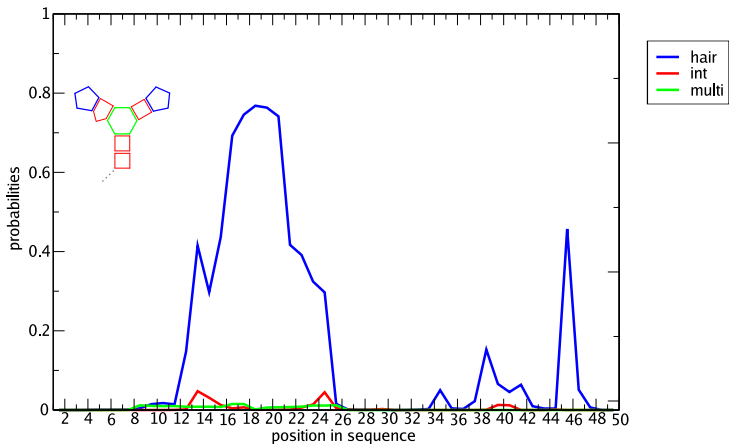
RNAup

Structural Information



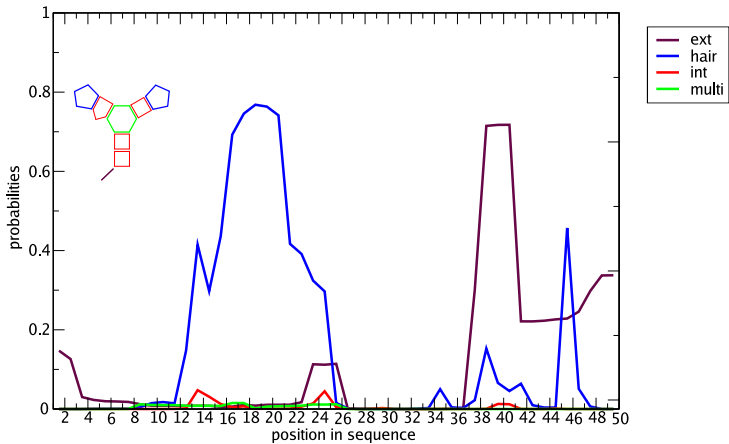
RNAup

Structural Information



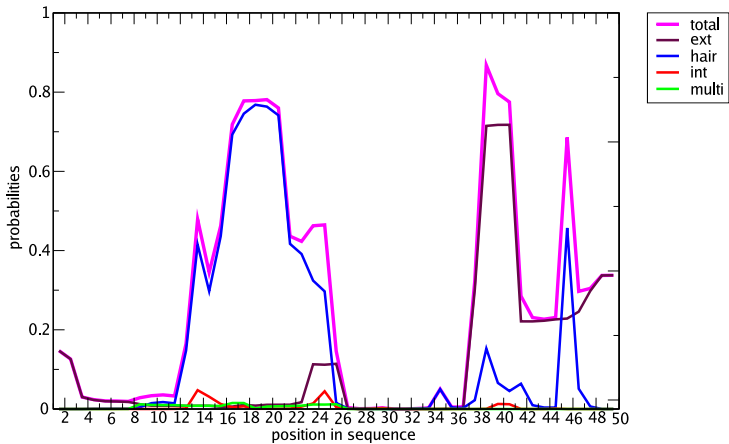
RNAUp

Structural Information

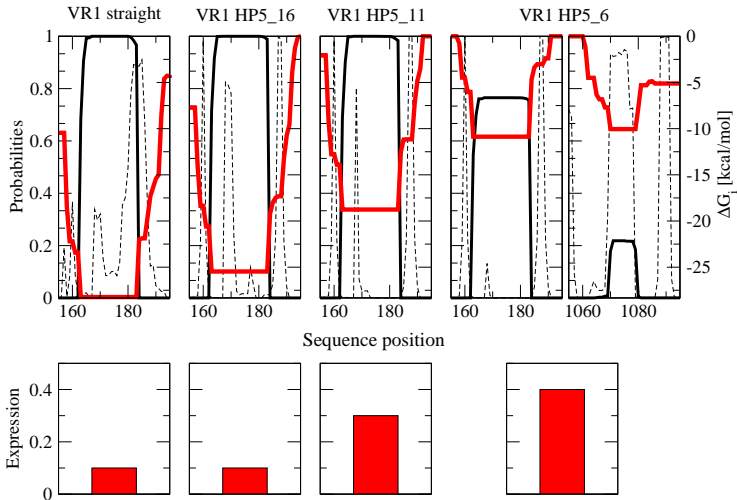


RNAup

Structural Information



Example: siRNA Binding

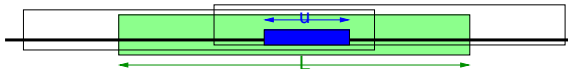


Data taken from Schubert et al 2006

A scanning Version of RNAup

Can we adapt this method for fast searching in large databases?

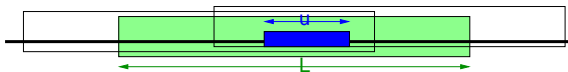
- *Local* folding algorithms can scan very large sequences by restricting the size of local structures to some maximum L .
- RNAplfold computes the probability that regions of length u are unpaired by averaging over all windows of length L
- Runtime is linear in the length of the database $\mathcal{O}(n \cdot L^2)$



A scanning Version of RNAup

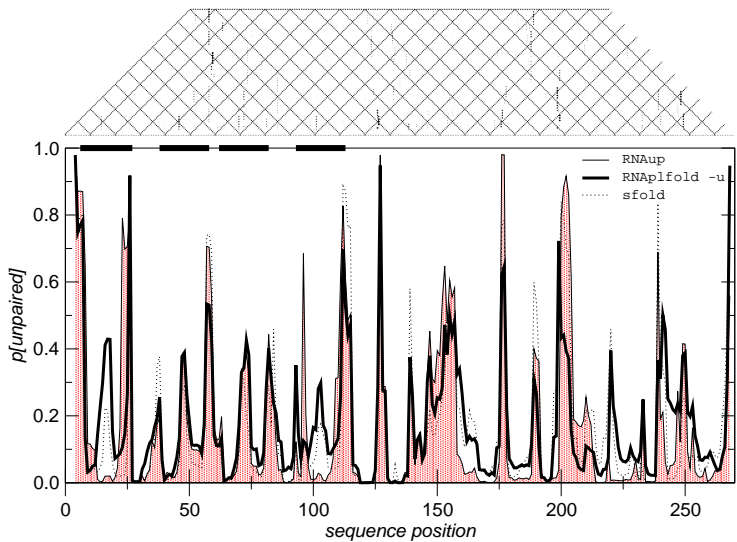
Can we adapt this method for fast searching in large databases?

- *Local* folding algorithms can scan very large sequences by restricting the size of local structures to some maximum L .
- RNAplfold computes the probability that regions of length u are unpaired by averaging over all windows of length L
- Runtime is linear in the length of the database $\mathcal{O}(n \cdot L^2)$



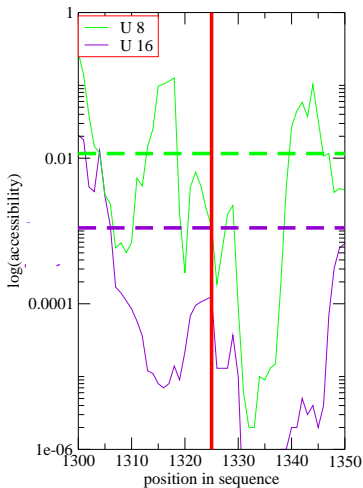
Computes average over all windows containing the region

$$\pi^L[i, j] = \frac{1}{L - (j - i) + 1} \sum_{u=j-L}^i P^{u, L}[i, j]$$

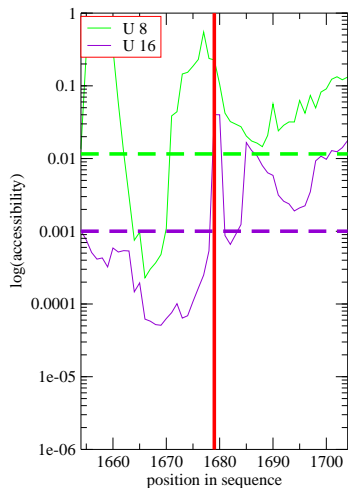


Accessibility of miRNA targets

NON WORKING -36.5 kcal/mol

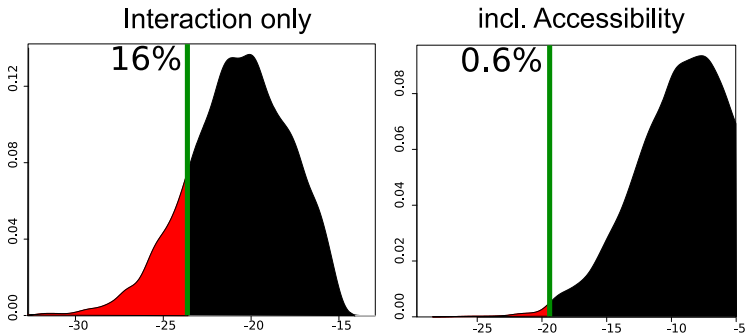


WORKING -28.3 kcal/mol

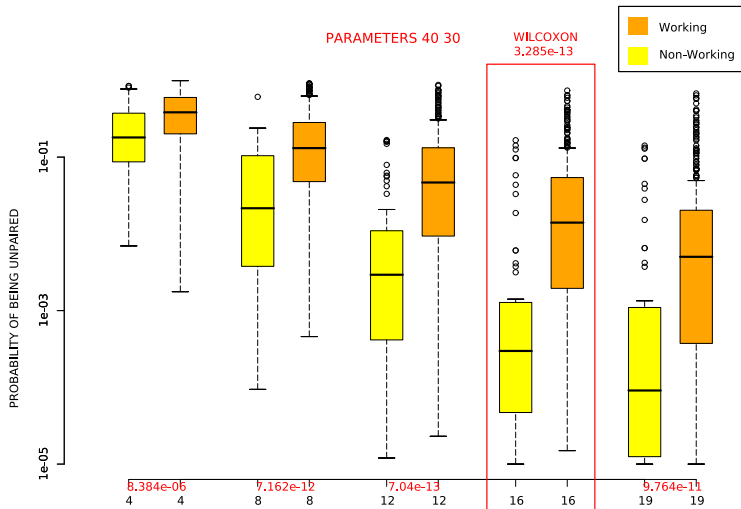


Accessibility and miRNA targets

miRNA 26



Accessibility predicts siRNA efficiency



Summary

Several approaches to RNA co-folding of varying complexity

- Including complex structures results in very slow algorithms
- RNAcofold and RNAup consider a limited set of interactions but run as fast as conventional folding algorithms
- Search for potential interactions and local accessibility can be performed by fast scanning algorithms, e.g. RNAplfold, RNAplex
- Strong interaction and accessibility are equally important for miRNA and siRNA binding
- Be aware of the limitations of each approach

Acknowledgments

- Ulli Mückstein
- Hakim Tafer
- Stephan Bernhart
- Gregor Obernosterer
- Peter Stadler
- Jörg Hackermüller