Prediction of RNA-RNA Interaction
slides by Mathias Möhl and Rolf Backofen
RNA-RNA interaction

(Waters and Storz, Cell 2009)
RNA-RNA interaction

(Waters and Storz, Cell 2009)
How is a mRNA-target recognized?

- idea 1: only hybridization energy counts
  
  \[
  \text{opt} \begin{pmatrix}
  \text{A} & \text{C} & \text{A} & \text{C} & \text{A} & \text{U} & \text{C} & \text{G} \\
  \text{G} & \text{G} & \text{A} & \text{G} & \text{C} & \text{A} & \text{A} & \text{U} & \text{C} & \text{A} & \text{U} & \text{G} & \text{U} \\
  \text{A} & \text{C} & \text{A} & \text{U} & \text{C} & \text{G} & \text{A} & \text{A} & \text{U} & \text{C} & \text{A} & \text{U} & \text{G} & \text{U} \\
  \text{A} & \text{C} & \text{U} & \text{C} & \text{U} & \text{U} & \text{G} & \text{U} \\
  \text{U} & \text{A} & \text{C} & \text{G} & \text{A}
  \end{pmatrix}
  \]

- many approaches build on that: RNAhybrid, RNAplex, TargetRNA etc.

- problem: structure

- hence: additional information in **accessibility** of site

  remark: accessibility = single-strandedness
Approaches to Target Detection

**approach 1**: maximize duplex energy (RNAhybrid, RNAplex, etc.)

\[
\begin{bmatrix}
\text{U-G} \\
\text{C-C-U-C-G-U-U-A} \\
\text{A-U-C-U-C-U-U-G-U} \\
\end{bmatrix}
\]

problem:

\[
\begin{bmatrix}
\text{G-G-U-G-G-G-A-U-U-G-C} \\
\text{C-U-U-A-C-A-U-C-G-A} \\
\text{G-G-U-G-G-G-A-U-U-G-C} \\
\end{bmatrix}
\]

**approach 2**: common structure by concatenation: (RNAcofold, PairFold)

- given: two sequences
- concatenate them
- use Zuker’s algorithm

\[
\begin{bmatrix}
\text{G-G-U-G-G-G-A-U-U-G-C} \\
\text{C-U-U-A-C-A-U-C-G-A} \\
\text{G-G-U-G-G-G-A-U-U-G-C} \\
\end{bmatrix}
\]

problem:
approach 3: RNAup/IntaRNA:
- determine probability for region i..j being unpaired
- calculate ensemble energy from probability
- hybridize unpaired region with second RNA

Comparison:
- approach 3: just one interaction possible
- approach 2: more than one interaction possible, but only external interactions (no pseudoknots in concatenated structure)
The Idea of IntaRNA

**IntaRNA** = **Interacting RNA**
similar to RNAup, but much faster
(optimized for scanning genomes)

\[ E = E_{hybrid} + ED_{i,i'}^{mRNA} + ED_{k,k'}^{ncRNA} \]

\[ \text{as in } RNAhybrid \quad RNAplfold \quad RNAplfold \]
Efficient Unpaired Probabilities (RNAplfold)

Given RNA sequence $S[1..n]$, compute probability $Pr[x..y|\text{unpaired}|S]$ that positions $x..y$ of $S$ are unpaired.

Recall matrices of McCaskill: $Q, Q^b, Q^m, Q^{m1}$, and introduce “outside” matrices $\hat{Q}, \hat{Q}^b$

$$Pr[x..y \text{ unpaired}|S] = Q^u(x, y)/Q(1, n)$$

Cases

I $x..y$ external, $O(1)$

II $x..y$ in hairpin, closed by $(i, j)$, naive $O(n^2)$

III $x..y$ in internal loop, closed by $(i, j)$, 5’ or 3’ of inner base pair $(k, l)$, $O(1)$ (internal loop size restricted)

IV $x..y$ in multiloop, closed by $(i, j)$, 5’, 3’, or between inner base pairs, naive $O(n^3)$

RNA Accessibility in cubic time Stephan H. Bernhart, Ulrike Mückstein, Ivo L. Hofacker. AMB 2011
Two Parts of One Problem?

- however: there are more complex structures
  - double kissing hairpins
  - ...
Example

- more than one internal interaction site
- example: OxyS-fhIA interaction

predicted complex [Alkan et al: JCB 2006]

(Argamana and Altuviaa: JMB 2000)
**Generalized Problem**

**approach 4**: predict joint mfe structure of two sequences

- like Zuker on two sequences at the same time, including loops between the sequences
- no pseudoknots, no crossing interaction
- proven to be NP-complete
- NP-completeness because of ZIG-ZAG structure
- without ZIG-ZAGs polynomial algorithm
approach 4: predict joint mfe structure of two sequences
- like Zuker on two sequences at the same time, including loops between the sequences
- no pseudoknots, no crossing interaction
- proven to be NP-complete
- NP-completeness because of ZIG-ZAG structure
- without ZIG-ZAGs polynomial algorithm
Given sequences $R$ and $S$, compute the maximal number of intramolecular and intermolecular base pairs

Fragments/subproblems $R[i..j], S[k..l]$

Decomposition cases

1. unpaired base at either end of one sequence
2. closed structure at either end of one sequence
3. base pair enclosing either sequence
4. interaction between left or right ends of sequences
5. decomposition at $i \leq i' \leq j, k \leq k' \leq l$ into two subproblems $R[i..i'], S[k..k']$ and $R[i' + 1..j], S[k' + 1..l]$

Complexity $O(n^6)$ time / $O(n^4)$ space. Why no zig-zags?
RNA-RNA interaction: literature 1/2

approach 1:

approach 2:
RNA-RNA interaction: literature 2/2

approach 3:


Anke Busch, Andreas S. Richter, Rolf Backofen, IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions

approach 4:

Hamidreza Chitsaz, Raheleh Salari, S. Cenk Sahinalp, Rolf Backofen, A partition function algorithm for interacting nucleic acid strands, Bioinformatics 2009

Raheleh Salari, Mathias Möhl, Sebastian Will, S. Cenk Sahinalp, Rolf Backofen, Time and space efficient RNA-RNA interaction prediction via sparse folding, RECOMB 2010