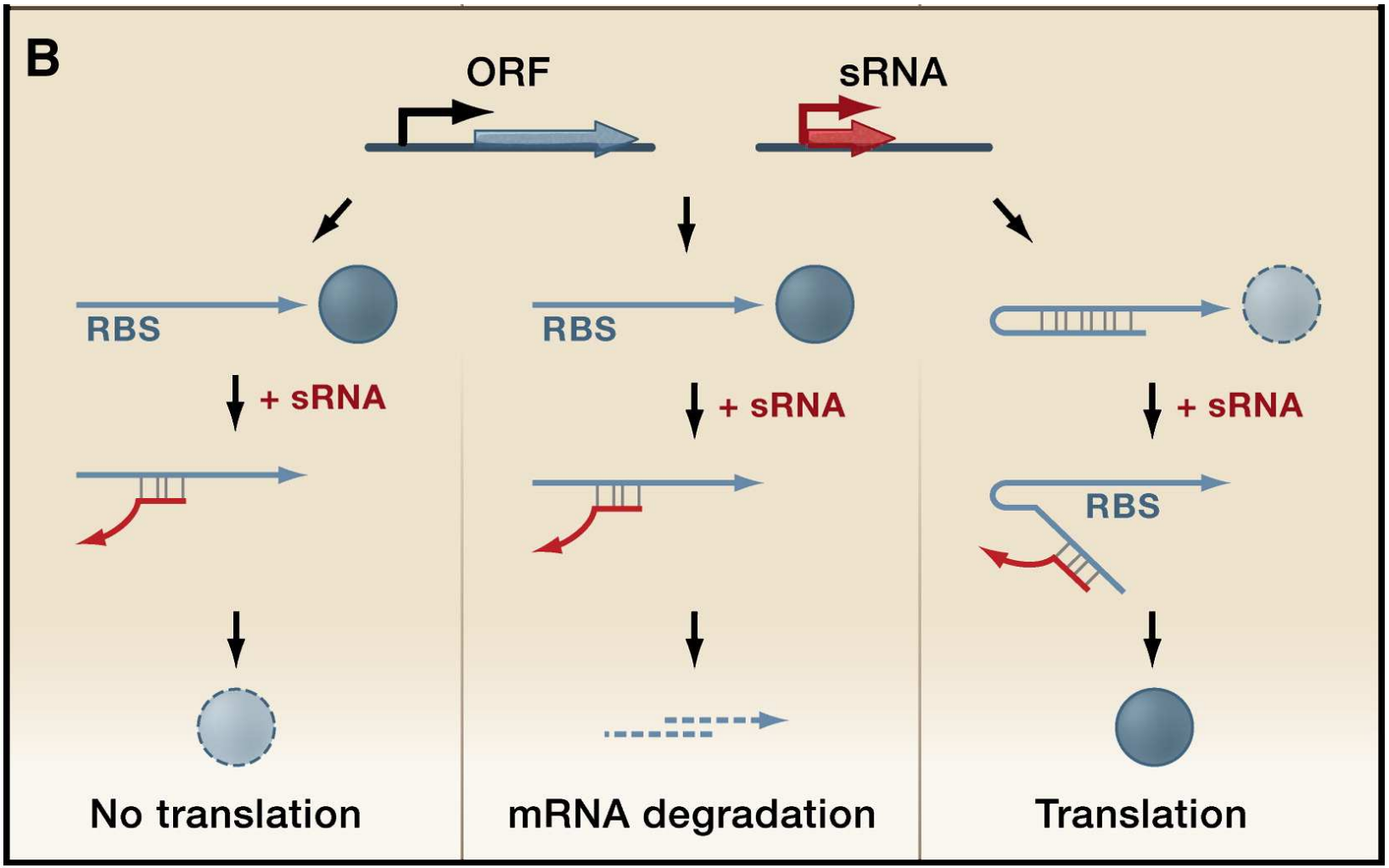
Prediction of RNA-RNA Interaction

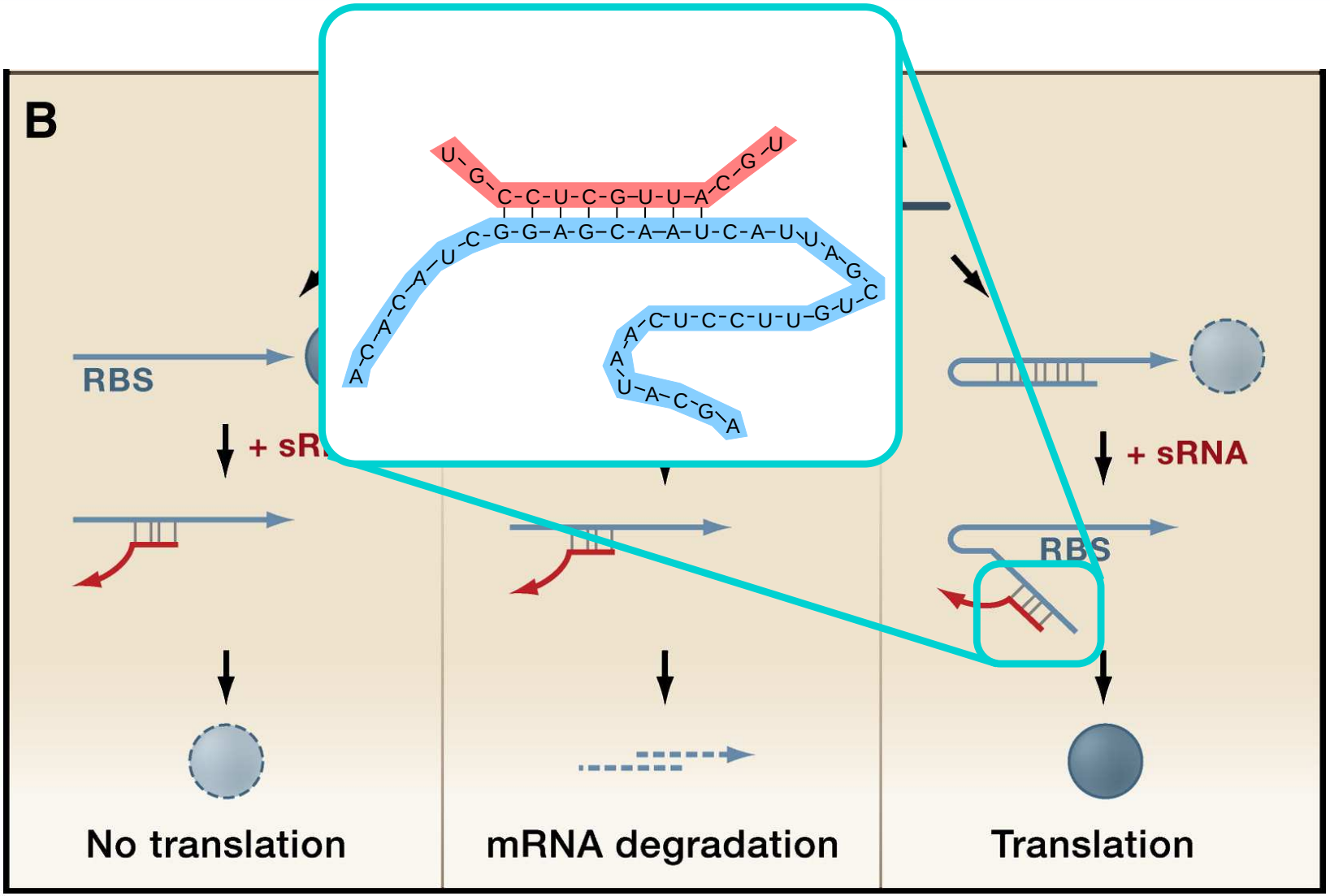
slides by Mathias Möhl and Rolf Backofen

RNA-RNA interaction



(Waters and Storz, Cell 2009)

RNA-RNA interaction

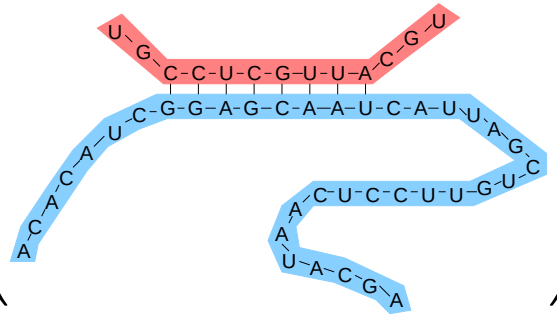


(Waters and Storz, Cell 2009)

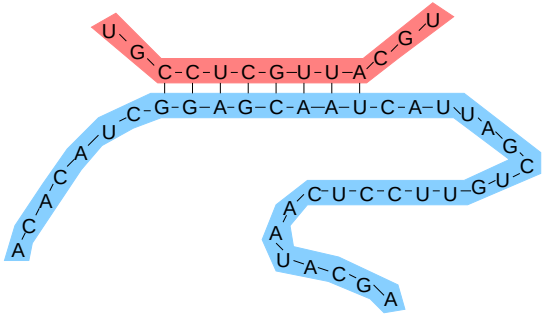
How is a mRNA-target recognized?

- idea 1: only hybridization energy counts

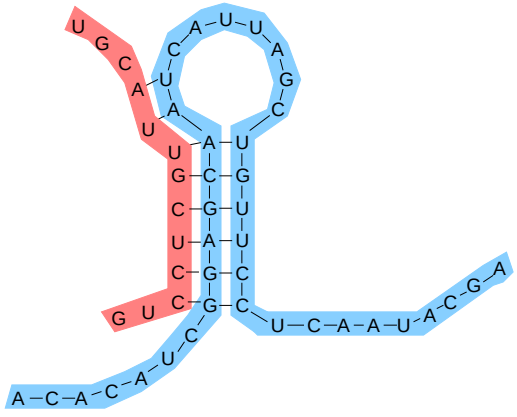
$$opt \left(\quad \right)$$



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



versus

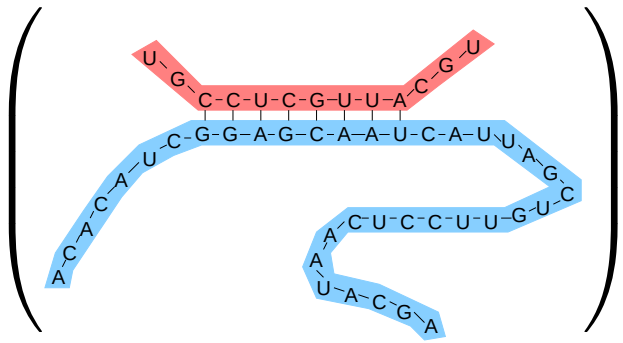


- hence: additional information in **accessibility** of site

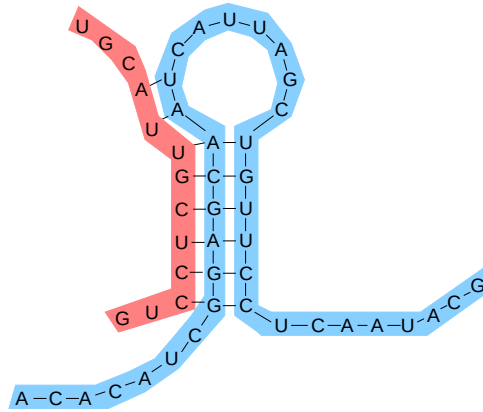
  remark: accessibility = single-strandedness

Approaches to Target Detection

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

opt 

problem:

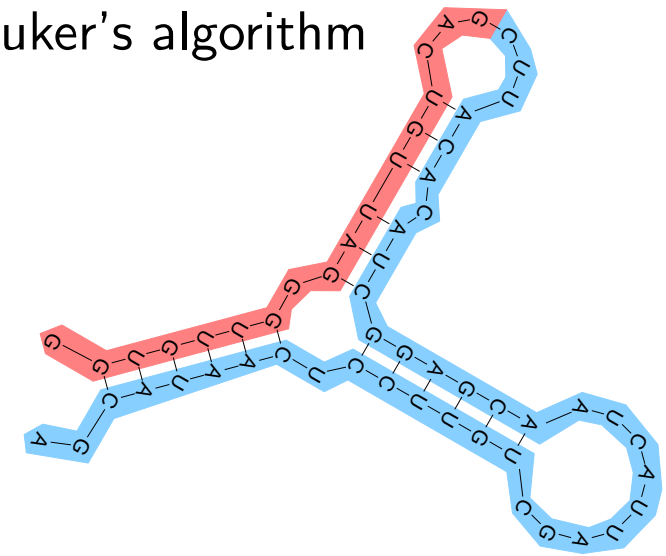


approach 2: common structure by concatenation:
(RNAcofold, PairFold)

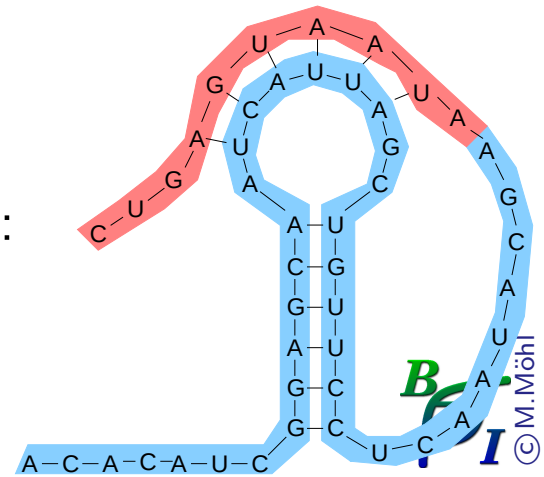
- given: two sequences
- concatenate them
- use Zuker's algorithm

GGUGUUGGGAUUGUCAG
CUUACACAUCGGAGCAAUCAUUAGCUGUUCCUCAAUACGA

GGUGUUGGGAUUGUCAG CUUACACAUCGGAGCAAUCAUUAGCUGUUCCUCAAUACGA



problem:



Approaches to Target Detection

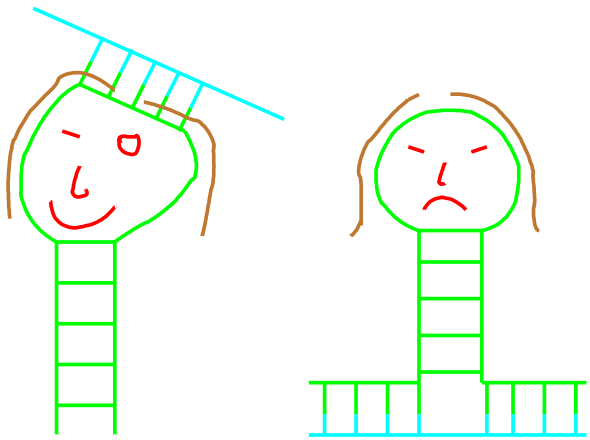
approach 3: RNAup/IntaRNA:

- determine probability for region i..j being unpaired
- calculate ensemble energy from probability
- hybridize unpaired region with second RNA

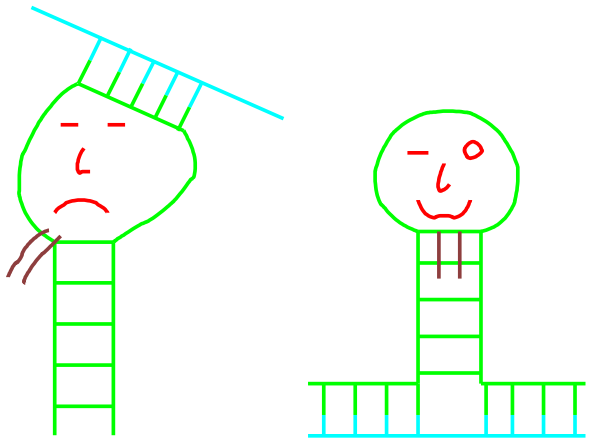
Comparison:

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

RNAup

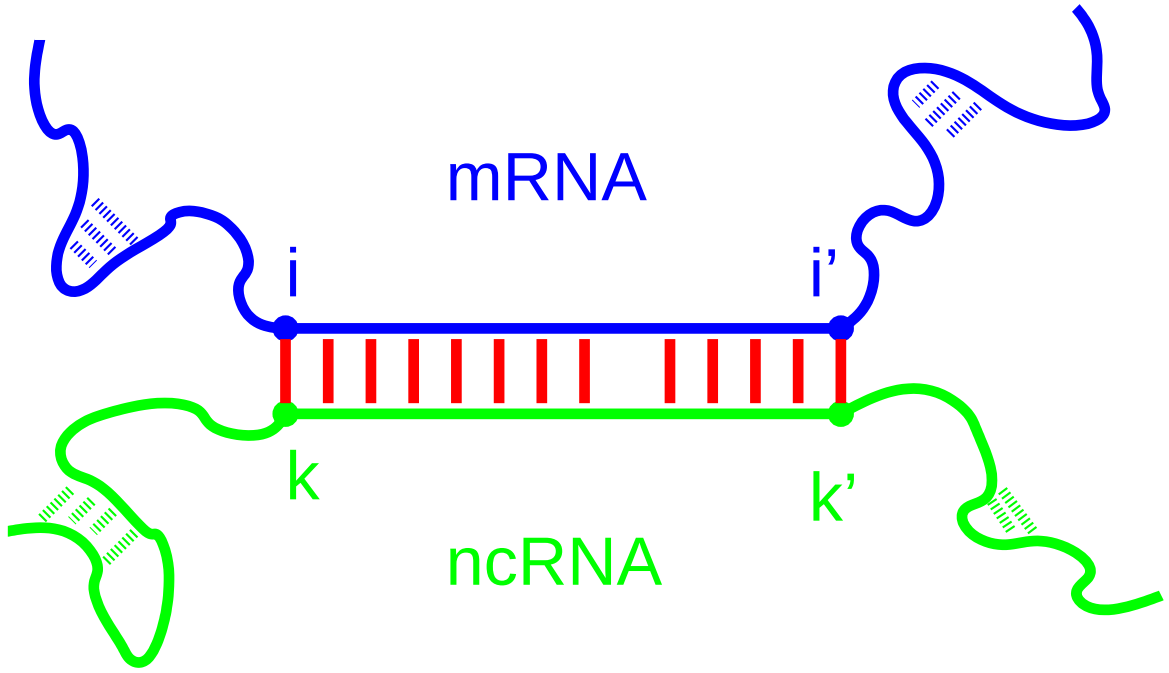


RNAcofold



The Idea of IntaRNA

IntaRNA = **Int**era**c**ting **RNA**
similar to RNAup, but much faster
(optimized for scanning genomes)



$$E = E^{hybrid} + ED^{mRNA}_{i,i'} + ED^{ncRNA}_{k,k'}$$

E^{hybrid} ↓ as in *RNAhybrid*

$ED^{mRNA}_{i,i'}$ ↓ *RNAplfold*

$ED^{ncRNA}_{k,k'}$ ↓ *RNAplfold*

Effificient Unpaired Probabilities (RNAplfold)

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

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

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

Cases

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

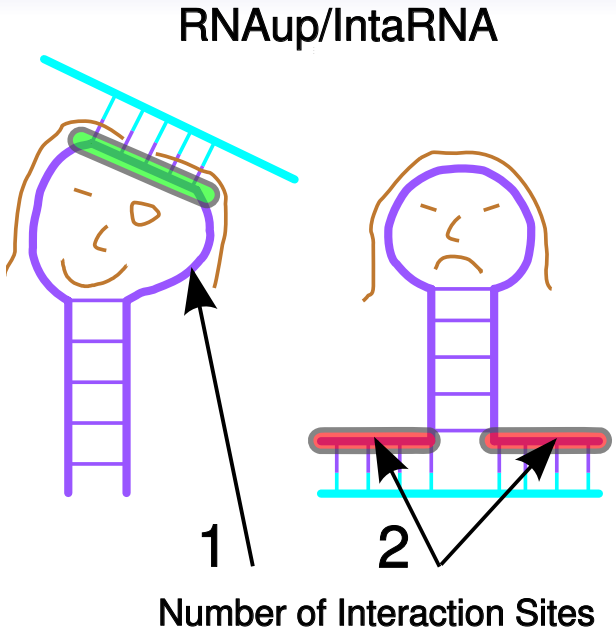
II $x..y$ in hairpin, closed by (i, j), naive $O(n^2)$

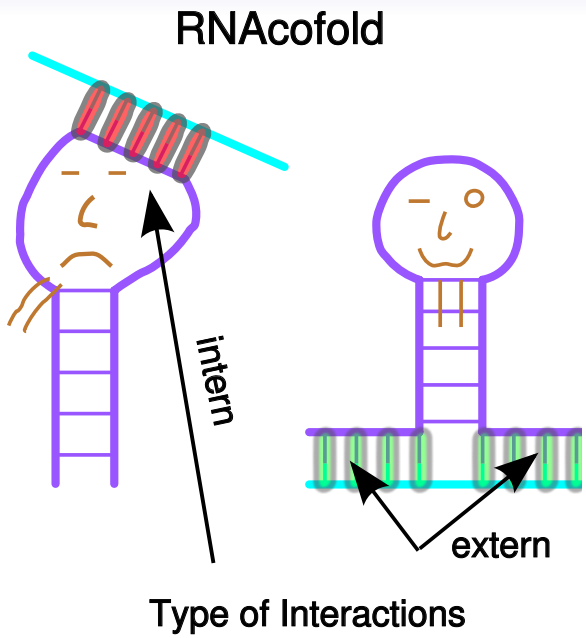
III $x..y$ in internal loop, closed by (i, j), 5' or 3' of inner base pair (k,l), $O(1)$ *(internal loop size restricted)*

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

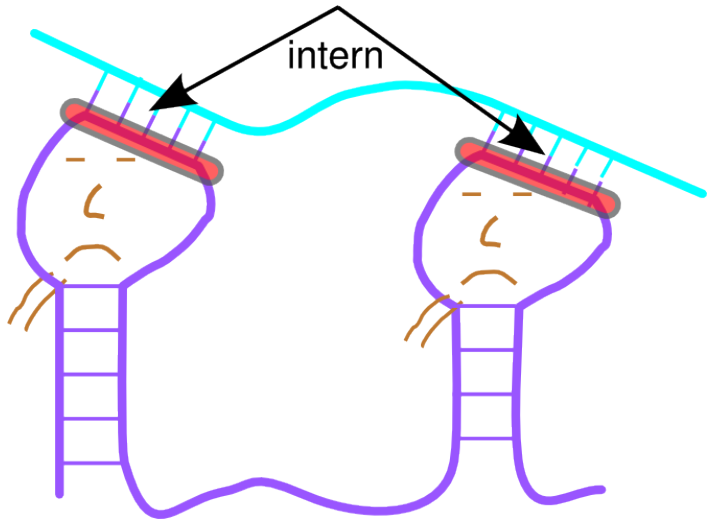
RNA Accessibility in cubic time Stephan H. Bernhart, Ulrike Mückstein, Ivo L. Hofacker. AMB 2011

Two Parts of One Problem ?



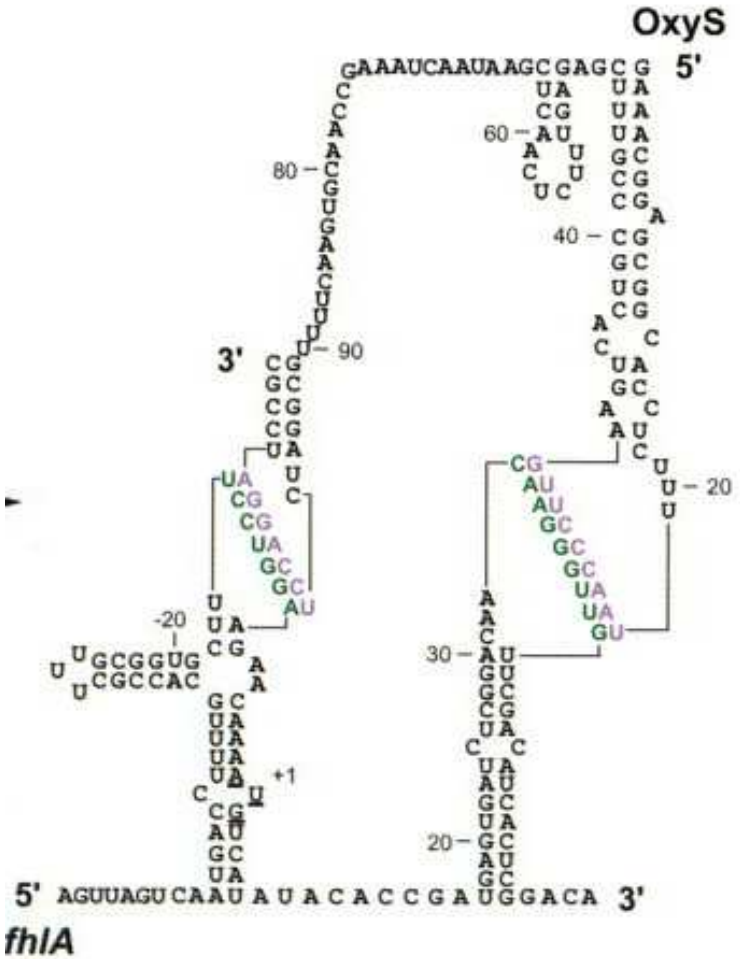


- however: there are more complex structures
  - double kissing hairpins
  - ...

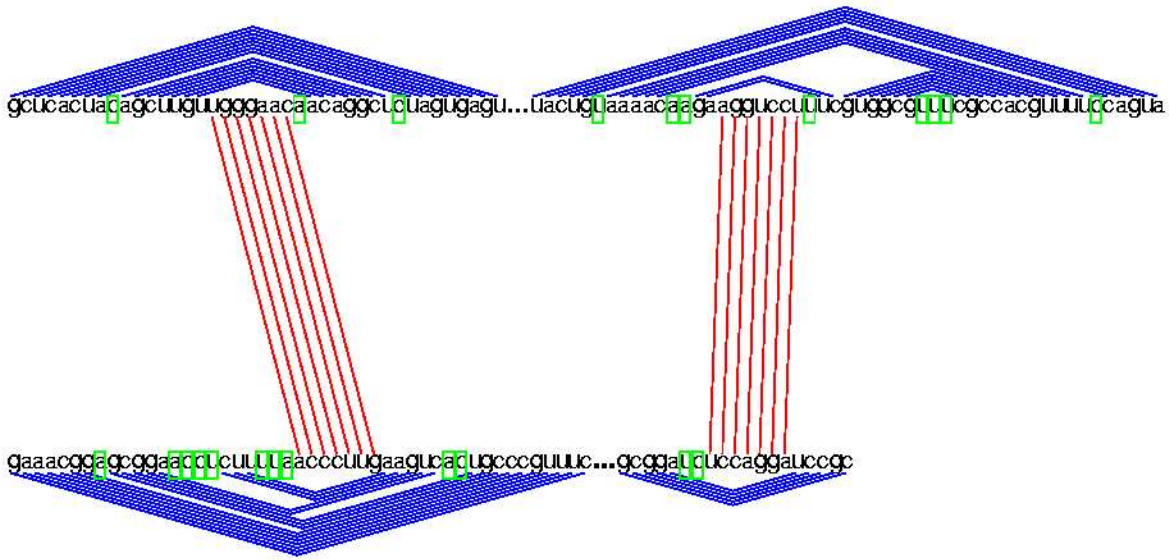


Example

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



(Argamana and Altuviaa: JMB 2000)

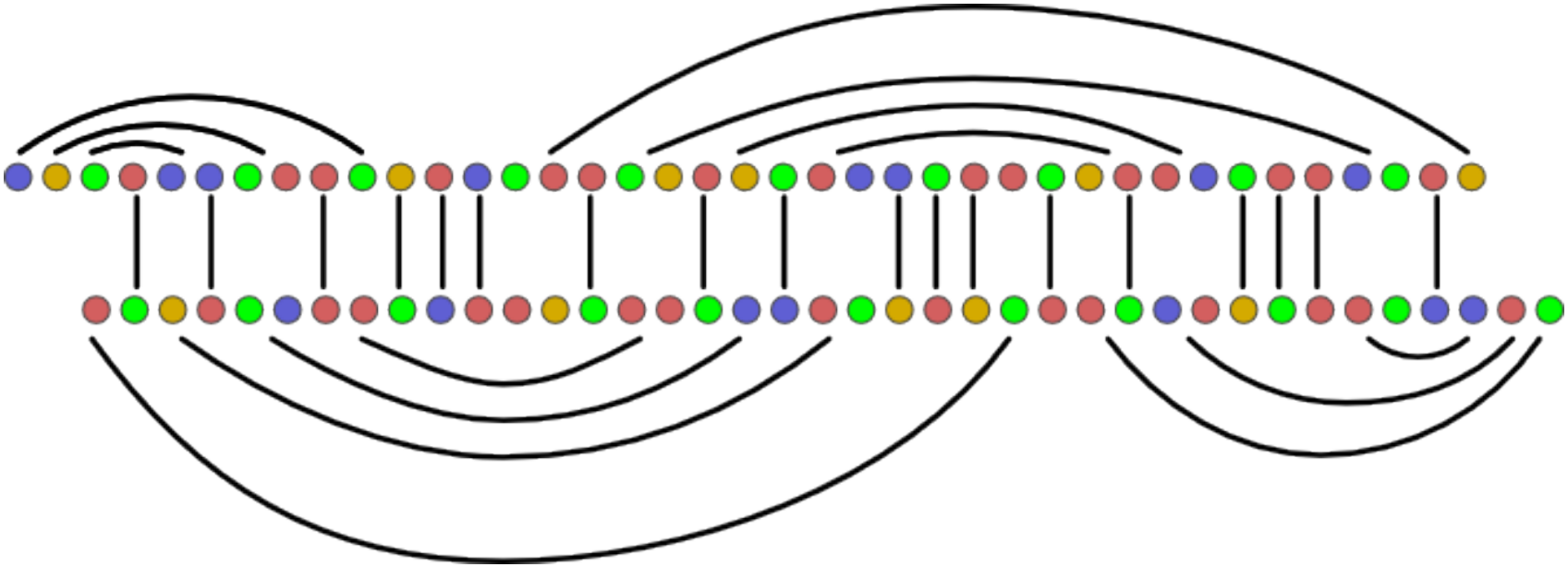


predicted complex [Alkan et al: JCB 2006]

Generalized Problem

approach 4: predict joint mfe structure of two sequences

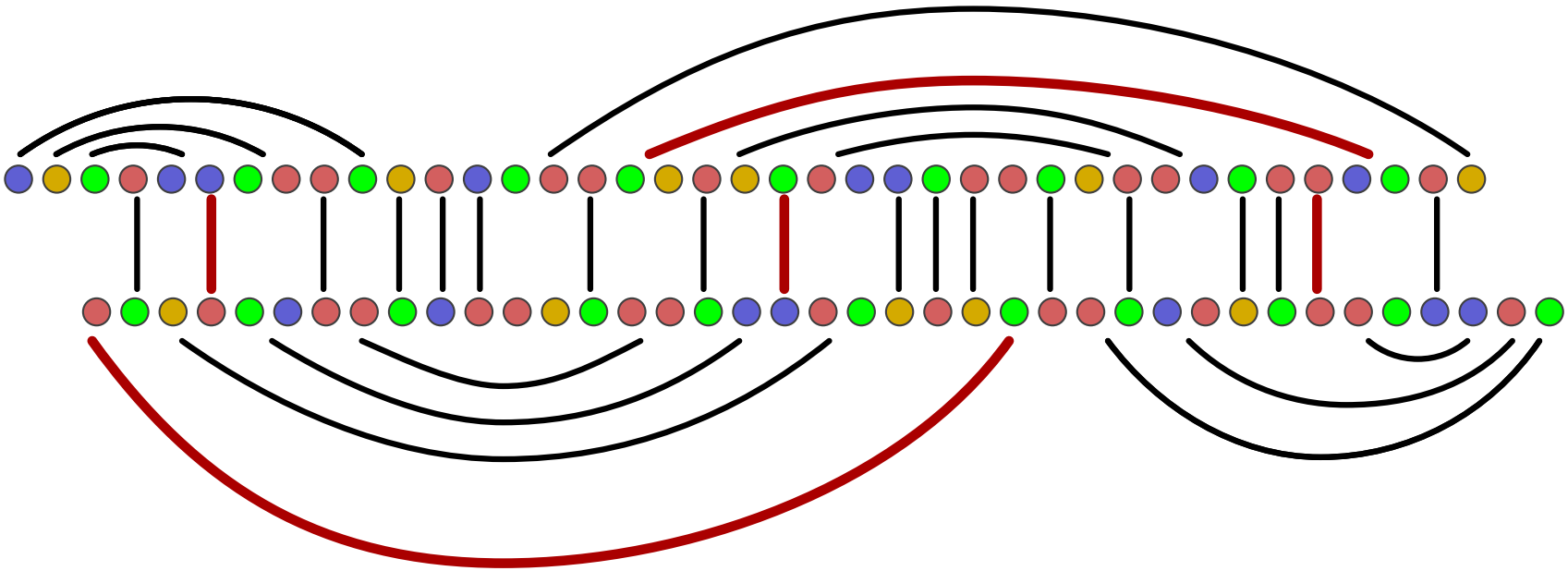
- like Zuker on two sequences at the same time, including loops between the sequences
- no pseudoknots, no crossing interaction
- proven to be NP-complete
- NP-completeness because of ZIG-ZAG structure
- without ZIG-ZAGs polynomial algorithm



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



Efficient Base Pair Maximization without Zig-Zag

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

I unpaired base at either end of one sequence

II closed structure at either end of one sequence

III base pair enclosing either sequence

IV interaction between left or right ends of sequences

V 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:

Hakim Tafer, Ivo L. Hofacker, RNAplex: a fast tool for RNA-RNA interaction search, Bioinformatics 2008

approach 2:

Stephan H. Bernhart, Hakim Tafer, Ulrike Mückstein, Christoph Flamm, Peter F. Stadler, Ivo L. Hofacker, Partition function and base pairing probabilities of RNA heterodimers, Algorithms Mol Biol 2006

RNA-RNA interaction: literature 2/2

approach 3:

Ulrike Mückstein, Hakim Tafer, Jorg Hackermuller, Stephan H. Bernhart, Peter F. Stadler, Ivo L. Hofacker, Thermodynamics of RNA-RNA binding, Bioinformatics 2006

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