

Some more Fun Things to Compute from the Partition Function of the RNA-RNA Interaction Model

Peter F. Stadler

Bioinformatics Group, Dept. of Computer Science & Interdisciplinary Center for
Bioinformatics, **University of Leipzig**

Max-Planck-Institute for Mathematics, Leipzig
Institute for Theoretical Chemistry, Univ. of Vienna (external faculty)
The Santa Fe Institute (external faculty)

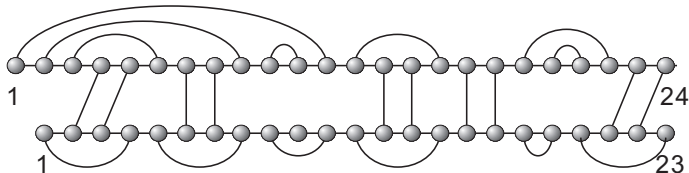
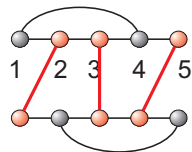
Joint Work with:
Fenix W.D. Huang, Jing Qin & Christian M. Reidys

Benasque, Jul 27 2012

Alkan's RIP Model

Two arbitrary secondary structures and non-crossing intermolecular base-pairs

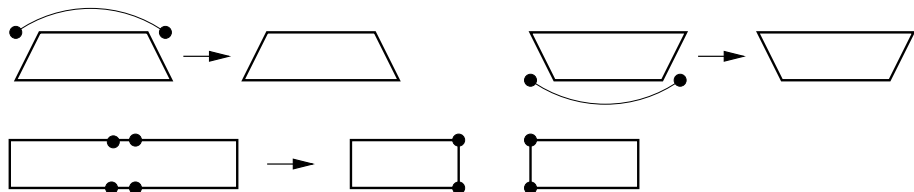
Forbidden configuration: the "zigzag"



Solvable by dynamic programming in the absence of "zigzags":
previous work by several groups:

Alkan, Pervouchine, Mneimneh, Backofen & Sahinalp

RIPing it apart



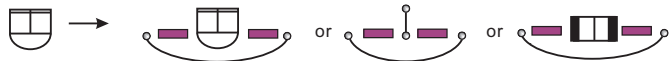
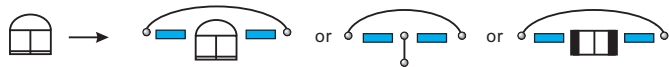
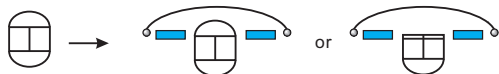
- 1 one of the partners is enclosed by a base pair:
→ “remove” this pair to reduce to a smaller problem.
- 2 neither of the partners is enclosed by a base pair:
Then there are breakpoints p and q in the two sequences such that no pairs connect the block structure $x[1, p] : y[q + 1, n]$ with $x[p + 1, n] : y[1, q]$.
→ cut at p and q and treat the two blocks separately.

Our unambiguous grammar

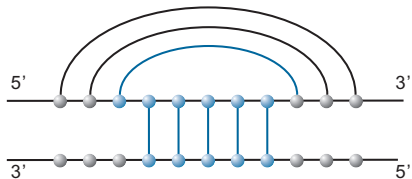
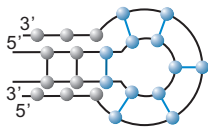
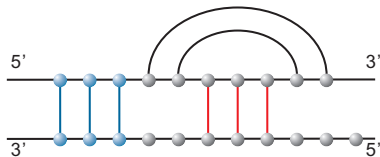
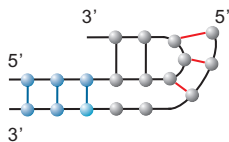
Procedure (a)



Procedure (b)

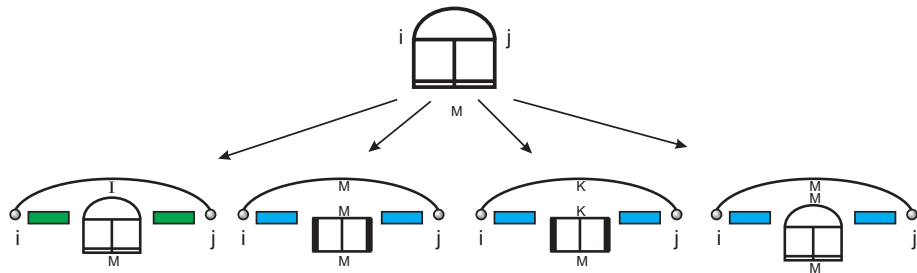


Full Energy model

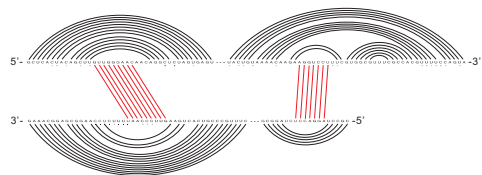


additional structural elements that need to be scored
multiloop-like model

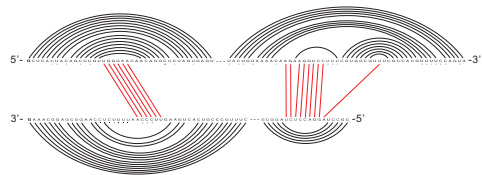
Full Energy model



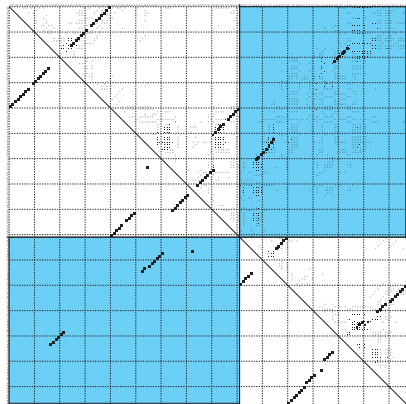
- Ugly but doable:
 - Hamidreza's talk just before
- $\mathcal{O}(n^6)$ time and $\mathcal{O}(n^4)$ memory
- Most of the arrays are used to store information for backtracing: $16 + 24 + 18 + 15 = 73$ four-dimensional arrays
- Improved version:
 $15 + 20 + 20 = 65$ four-dimensional arrays with a stochastic backtracing

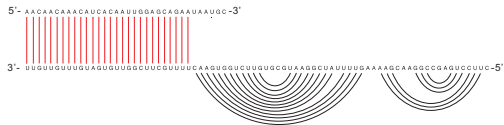


(A)

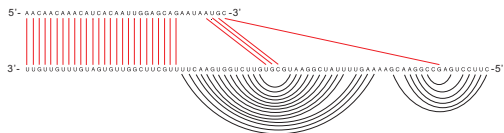


(B)

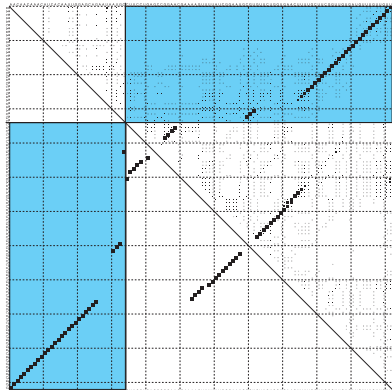




(A)

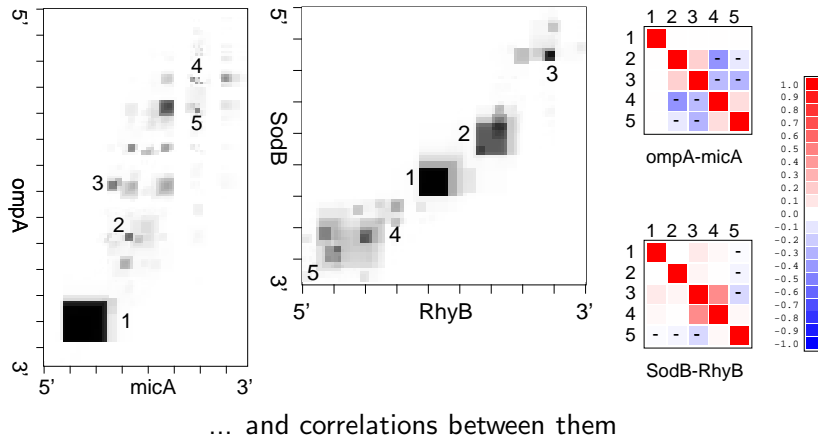


(B)



Interaction Regions

Probability $\pi_{i,j}$ that the basepair i,j is contained in an interacting region



RIP for Multiple sequence alignments

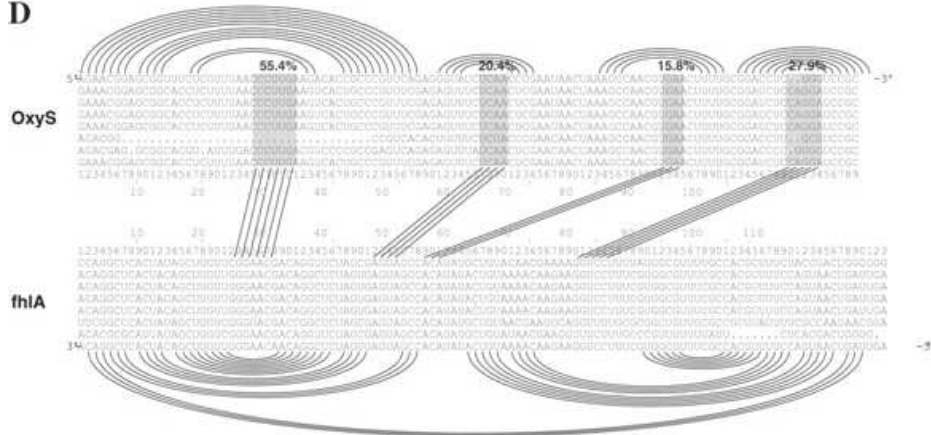
Andrew X. Li, Manja Marz, Jing Qin, Christian M. Reidys

- RNAalifold-like energy model: average of the energies of the individual aligned sequences
- small bonus energies for sequence covariations

RIP for Multiple sequence alignments

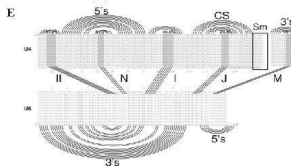
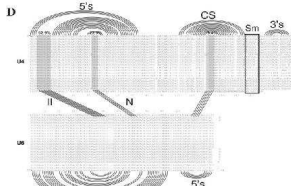
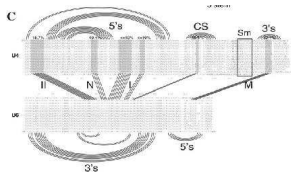
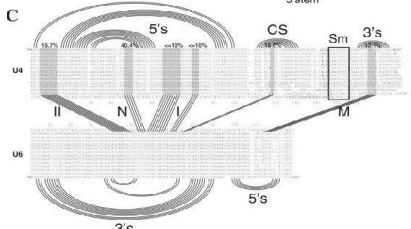
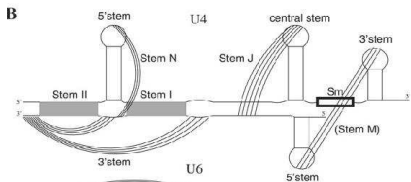
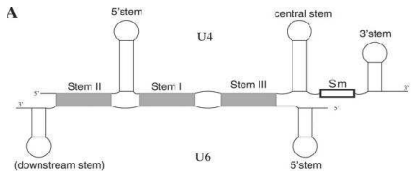
Andrew X. Li, Manja Marz, Jing Qin, Christian M. Reidys

D



RIP for Multiple sequence alignments

Andrew X. Li, Manja Marz, Jing Qin, Christian M. Reidys



U4/U6

top:

protostomia

middle:

insects

bottom:

deuterostomia