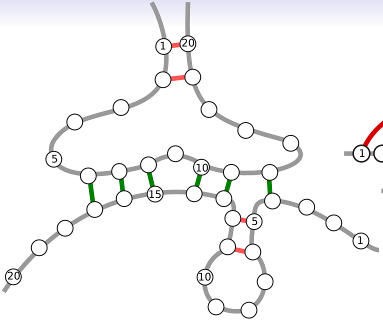
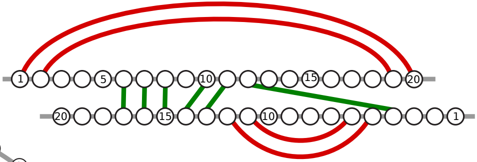
Prediction of RNA-RNA-Interaction

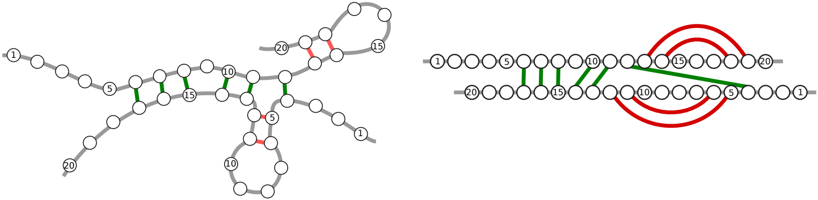




Can Alkan, Emre Karakoc, Joseph H. Nadeau, S. Cenk Sahinalp, Kaizhong Zhang. RNA-RNA interaction prediction and antisense RNA target search. JCB 2006

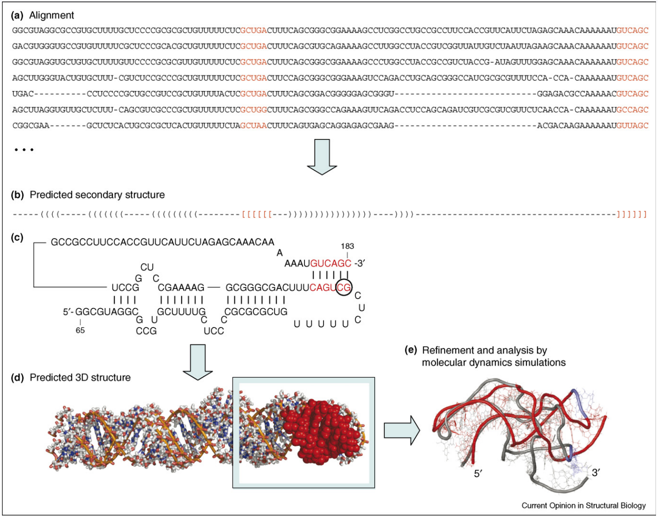
- define problem RIP (with and without PKs)
- prove NP-completeness even without PK for Base pair-energy model and more complex models
  (reduction from "longest common subsequence of multiple binary strings", mLCP)

Relation between PK-Prediction and RIP



- RNAcofold: concatenate RNAs A and B, predict PK-free structure
- specific restrictions on the structure of the interaction complex
- Can we apply pseudoknot-prediction to concatenation? Difference to Alkan-algorithm?

Semiautomatic RNA 3D Structure Modeling



(a) Alignment

```
GGCGUAGGCGCCGUGCUUUUGCUCCCCGCGCGCUGUUUUUCUCGCUGACUUUCAGCGGGCGGAAAAGCCUCGGCCUGCCGCCUUCCACCGUUCAUUCUAGAGCAAACAAAAAAUGUCAGC
GACGUGGGUGCCGUGUUUUUCGCUCCCGCACGCUGUUUUUCUCGCUGACUUUCAGCGUGCAGAAAAGCCUUGGCCUACCGUCGGUUAUUGUCUAAUUAGAAGCAAACAAAAAAUGUCAGC
GGCGUAGGUGCUGUGCUUUUGUUCCCCGCGCGUUGUUUUUCUCGCUGACUUUCAGCGGGCGGAAAAGCCCUGGCCUACCGCCGUCUACCG-AUAGUUUGGAGCAAACAAAAAAUGUCAGC
AGCUUGGGUACUGUGCUUU-CGUCUCCGCCCGCUGUUUUUCUCGCUGACUUCCAGCGGGCGGGAAAGUCCAGACCUGCAGCGGGCCAUCGCGCGUUUUCCA-CCA-CAAAAAAUGUCAGC
UGAC----------CCUCCCCGCUGCCGUCCGCUGUUUUACUCGCUGACUUUCAGCGGACGGGGGAGCGGGU---------------------------GGAGACGCCAAAAACGUCAGC
AGCUUAGGUGUUGCUCUUU-CAGCGUCGCCCGCUGUUUUUCUCGCUGGCUUUCAGCGGGCCAGAAAGUUCAGACCUCCAGCAGAUCGUCGCGUCGUUCUCAACCA-CAAAAAAUGCCAGC
CGGCGAA-------GCUCUCACUGCGCGCUCACUGUUUUUCUAGCUAACUUUCAGUGAGCAGGAGAGCGAAG---------------------------ACGACAAGAAAAAAUGUUAGC
```

• • •

(b) Predicted secondary structure

```
-----((((-----(((((((-----(((((((((--------[[[[[[---))))))))))))))))----))))--------------------------------------]]]]]]
```

(c)

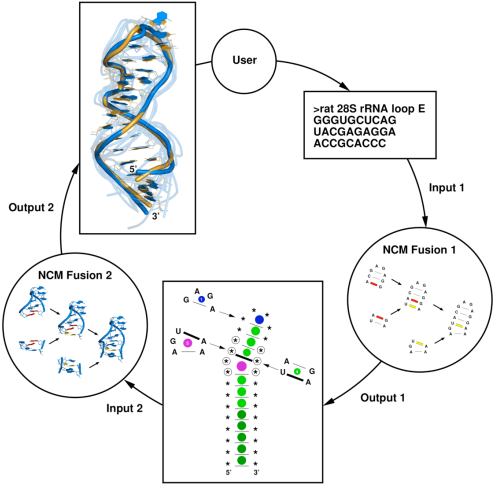
(d) Predicted 3D structure

(e) Refinement and analysis by molecular dynamics simulations

Current Opinion in Structural Biology

Bruce A Shapiro, Yaroslava G Yingling, Wojciech Kasprzak and Eckart Bindewald. Bridging the gap in RNA structure prediction Current Opinion in Structural Biology. 2007

An automated pipeline: MC-Fold/MC-Sym

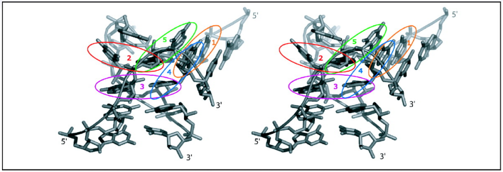


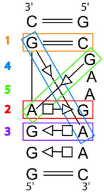
Marc Parisien & Francois Major. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. Nature 2008.

Potential obstacles

- Reliability of secondary structure prediction
  →prediction from alignments, covariance
- Pseudoknots
  → pseudoknot prediction
  → covariance analysis of large multiple alignments
- Non-canonical base pairs
  → experimental loop energies? learn from 3D-structures!
- 3D-motifs (due to non-canonical base pairs)
  → learn from 3D-structures, isostericity

Non-canonical Base Pairs, 3D-Motifs and Isostericity





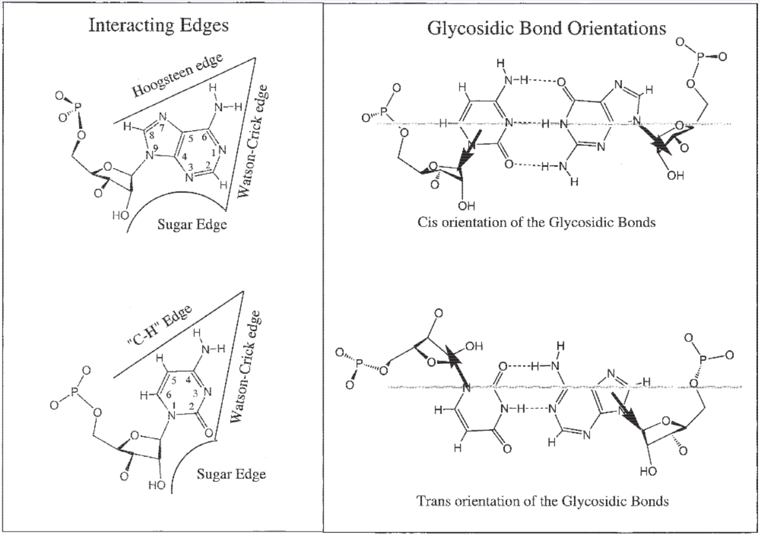
1		Watson-Crick			
	cis	A	C	G	U
Watson-Crick	A	I4	I2	I3	I1
	C	I2	I6	I1	I5
	G	I3	I1		I2
	U	I1	I5	I2	I6

2&3		Sugar-Edge			
	trans	A	C	G	U
Hoogsteen	A	I1	I1	I1	I1
	C	I1	I1		I1
	G			I2	
	U	I2		I2	

4&5		Sugar-Edge			
	trans	A	C	G	U
Sugar-Edge	A	I1		(I2)	
	C	I1		I2	
	G	I1		I2	
	U	I1		I2	

Recurrent structural RNA motifs, Isostericity Matrices and sequence alignments. Aurélie Lescoute, Neocles B. Leontis, Christian Massire and Eric Westhof. NAR 2005.

Non-Canonical Base Pairs



Interacting Edges

Glycosidic Bond Orientations

Cis orientation of the Glycosidic Bonds

Trans orientation of the Glycosidic Bonds

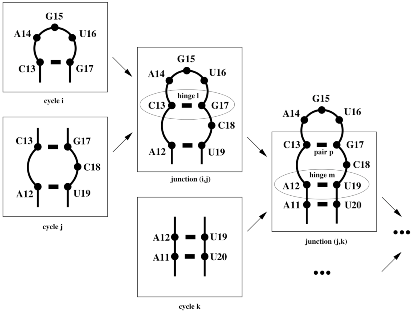
Leontis, N.B. and Westhof, E. Geometric nomenclature and classification of RNA base pairs. RNA 2001

Back to MC-Fold/MC-sym

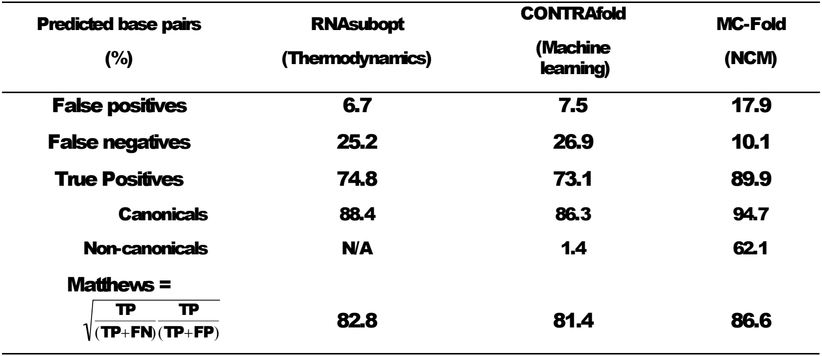
- NCMs: Nucleotide Cyclic Motifs from PDB (531 structures)
- MC-fold predicts secondary structure *including* non-canonical base pairs by merging NCMs
- Probability-based scoring

$$\begin{aligned}Pr[structure|seq] &= Pr[NCMs|seq] \times Pr[junctions|NCMs] \\ &\quad \times Pr[hinges|junctions] \times Pr[pairs|hinges]\end{aligned}$$

- predict sub-optimals



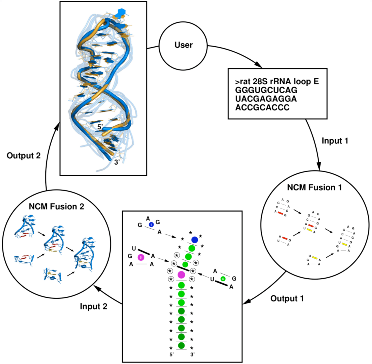
Prediction Performance of MC-Fold



Predicted base pairs (%)	**RNAsubopt (Thermodynamics)**	**CONTRAfold (Machine learning)**	**MC-Fold (NCM)**
False positives	**6.7**	**7.5**	**17.9**
False negatives	**25.2**	**26.9**	**10.1**
True Positives	**74.8**	**73.1**	**89.9**
Canonicals	**88.4**	**86.3**	**94.7**
Non-canonicals	**N/A**	**1.4**	**62.1**
Matthews = $\sqrt{\frac{TP}{(TP+FN)}\frac{TP}{(TP+FP)}}$	**82.8**	**81.4**	**86.6**

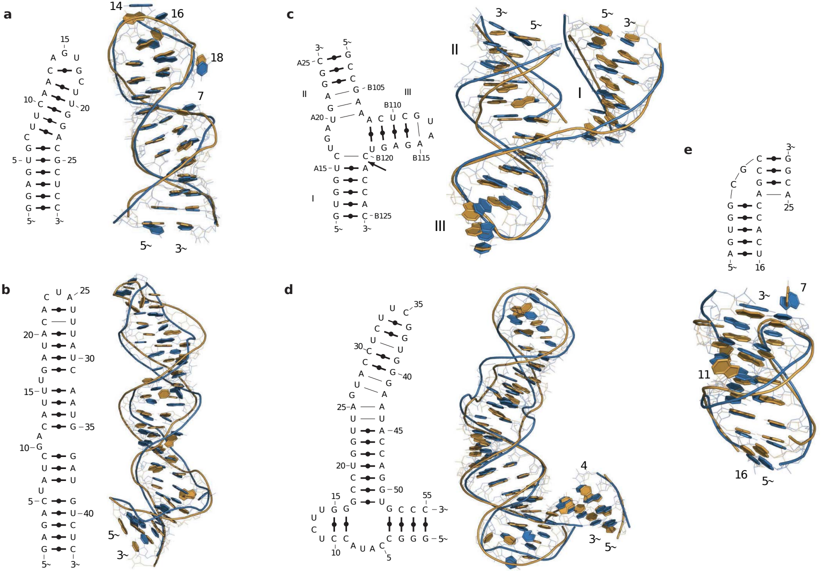
1968 base pairs (1665 Watson-Crick) in 264 hairpins from 182 different PDB structures

MC-Sym



- libraries of 3D-fragments for each NCM
- solve combinatorial puzzle, satisfy steric/RMSD constraints
- Las-vegas algorithm (no exhaustive enumeration, could fail to produce solution)
- run-time in pipeline 24h

Example Predictions of MC-Fold/MC-Sym



[Parisien&Major, Nature 2007]

Rfam / Infernal

- Infernal: scan genomic data for RNA family members

  Inference of RNA alignments

- important tool for Rfam

  Rfam 10.1 (June 2011, 1973 families)

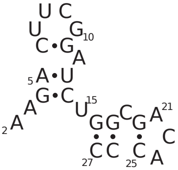
  `http://rfam.sanger.ac.uk/`

- in Rfam: 'hand-curated' seed alignments ⇒ full alignments
- use Stochastic Context Free Grammars to model RNA families
- model of a family: *Consensus Model (CM)*

input multiple alignment:

```
[structure]  . : : < < < _ _ _ _ > - > > : < < - < . _ _ _ . > > > .
      human  . A A G A C U U C G G A U C U G G C G . A C A . C C C .
      mouse  a U A C A C U U C G G A U G - C A C C . A A A . G U G a
        orc  . A G G U C U U C - G C A C G G G C A g C C A c U U C .
             1       5         10        15        20        25   28
```

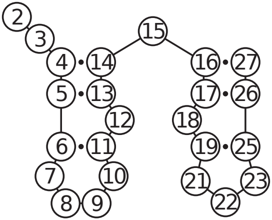
example structure:



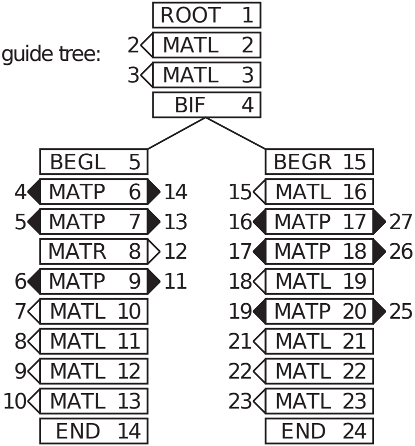
Infernal

Construct grammatical description

consensus structure:



guide tree:



Infernal

- Construct CM from guide tree
- Expand nodes of guide tree:

  Add match, insertion, and deletion states
- learn transition and output probabilities from alignment
- CM comparable to profile HMM for protein families (Pfam)