New features in the Vienna RNA Package

Ivo Hofacker

Institute for Theoretical Chemistry, University of Vienna

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

Bled, Februar 2006

The Vienna Package at Age 12

- Version 1.6 released in January 2006
- several 1.5α, β in 2002 -2005
- Version 1.4 May 2003
  Matthews '99 energy parameters
- Version 1.3 April 1999
  RNAsubopt, Perl module
- Version 1.2.x 1997
  Walter '94 parameters
- Version 1.1 1995
  dangling ends for MFE folding
- Version 1.0 1994

New in Version 1.6

- Additional sub-packages
  - Kinfold – RNA folding kinetics
  - RNAforester – structure alignment
- New and updated programs
  - RNAfold (with circular folding)
  - RNAcofold (partition function version)
  - RNAduplex
  - RNAsubopt (stochastic backtracking)
  - RNALfold, RNAplfold
  - RNApaln
  - RNAalifold
- Updated Perl Module (type: perldoc RNA)
- More Perl utilities

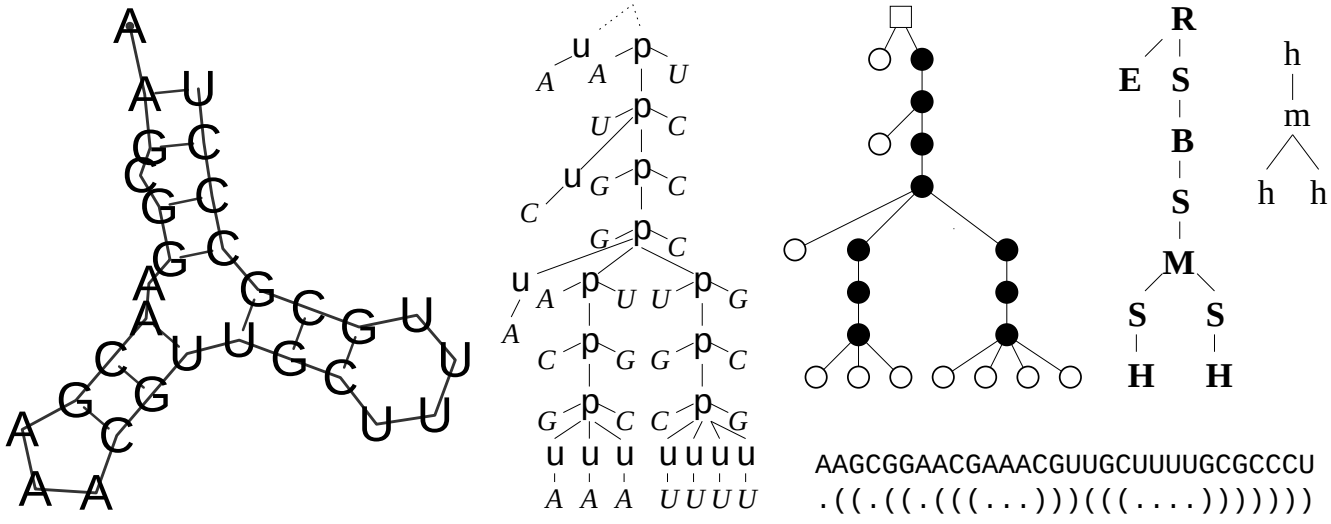
RNAforester

- written by Matthias Hoechsmann
- computes tree *alignments*
- global and local similarity alignments
- multiple alignments
- visualize pairwise structure alignments

Tree Representations of RNAs

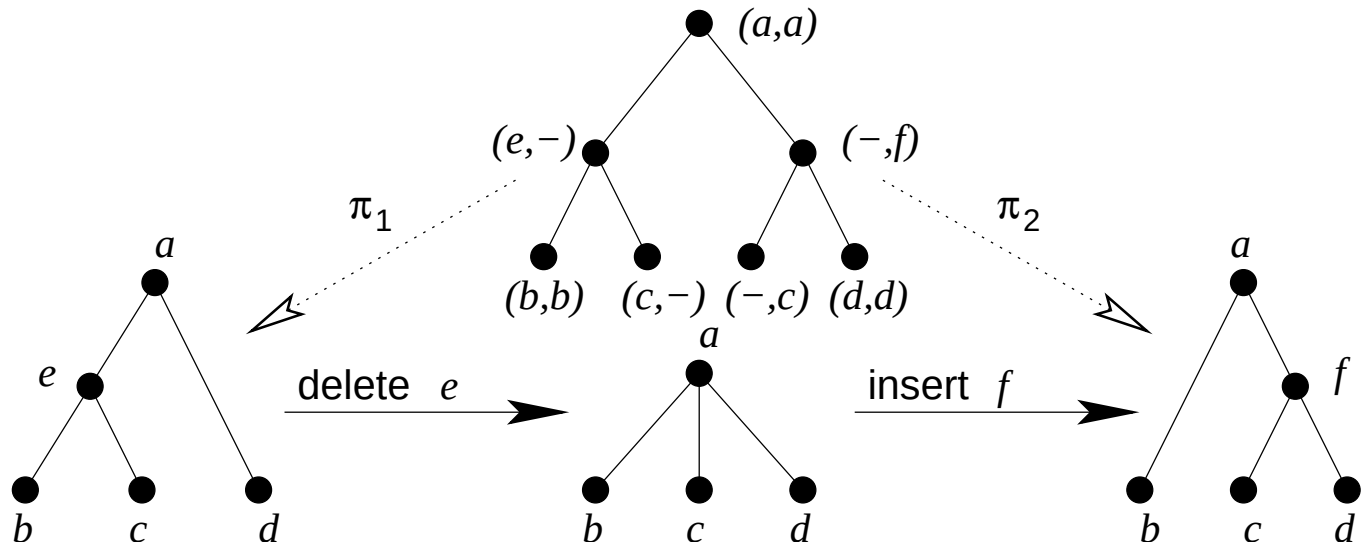
Knot-free secondary structures are trees

Basis for many comparison methods (see RNAdistance)



Tree Alignments

- analog of alignment to trees
- different from tree editing
- allows for progressive multiple alignments

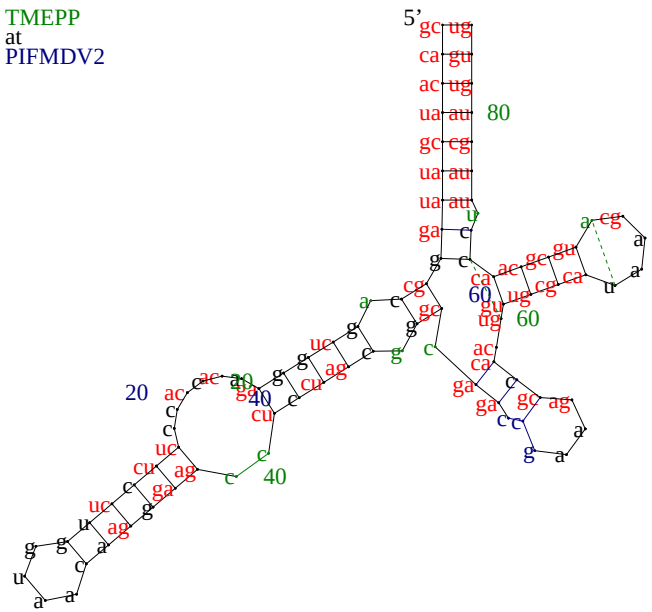


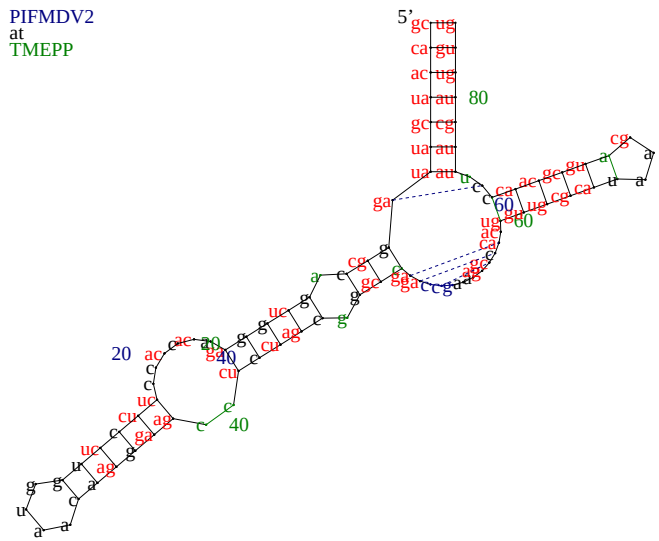
RNAforester Example

```
> RNAfold < ires.seq | RNAforester -l -2d
local optimal score: 117
starting at positions: 0,0

PIFMDV2    gcauguuggcc-gugggaacaccuccuugguaacaagga--cccac-gg-ggccgaaagccauguccuaac-ggaccc-aacaugu
TMEPP      cacacaaaggcagcggaacccccccuccugguaacaggagccucugcggccaa---aagccacguggauaagauccaccuuugugug
   * * * ** * * **  * ******** *     *  * *   **  *       ***      **

PIFMDV2    ((((((((((-(((((......((((((....))))))--)))))-))-((.(....))).(((((....-))))))-)))))))
TMEPP      (((((((.(((.(((((......((((((....))))))..)))))).)))..---.......((((((...))))))..)))))))
   ******* *** **************************  ***** ***      **   * **** *** *****  *******
```





Stochastic Backtracking

- Produce a Boltzman weighted sample of structures $p(S) = \exp(E(S)/RT)$
- Natural analog to backtracking in the partition function algorithm
- Any thermodynamic average can be computed approximately by sampling

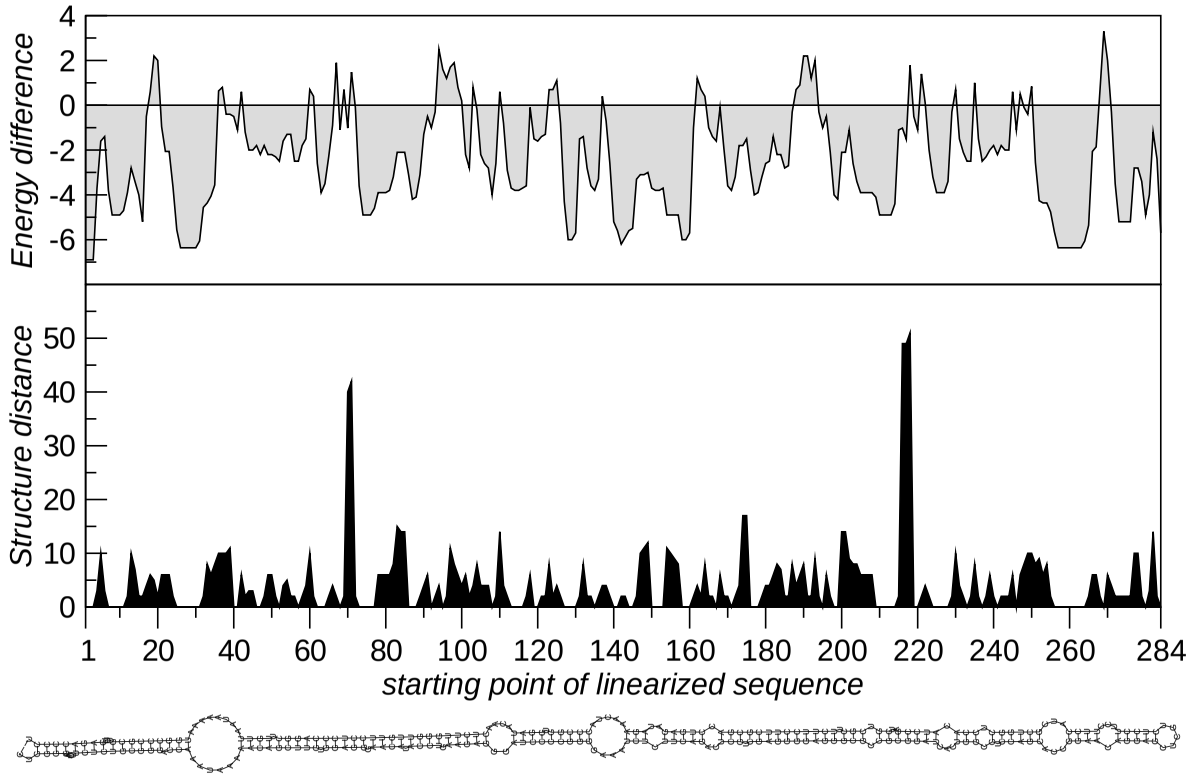
```
$ RNAsubopt -p 6 < tRNA.seq
GCGGAUUUAGCUCAGUUGGGAGAGCGCCAGACUGAAGAUCUGGAGGUCCUGUGUUCGAUCCACAGAAUUCGCACCA
(((((((..((((........)))).(((((....).)))))....((((.......))))))))))....
(((((....(((((((.((......)).)))))..)).((((..((((........))))..))))))))....
(((((....(((((((((.......))))..))))..).((((..((((........))))..))))))))....
(((((((...(((((..((......))..)))))).....((..((((........))))..))))))))....
(((((((.((((........)))).(((((....).))))).(((........))))....)))))))....
(((((((.((((........)))).(((((....).))))).(((........)).))...)))))))....
```

Folding circular RNAs

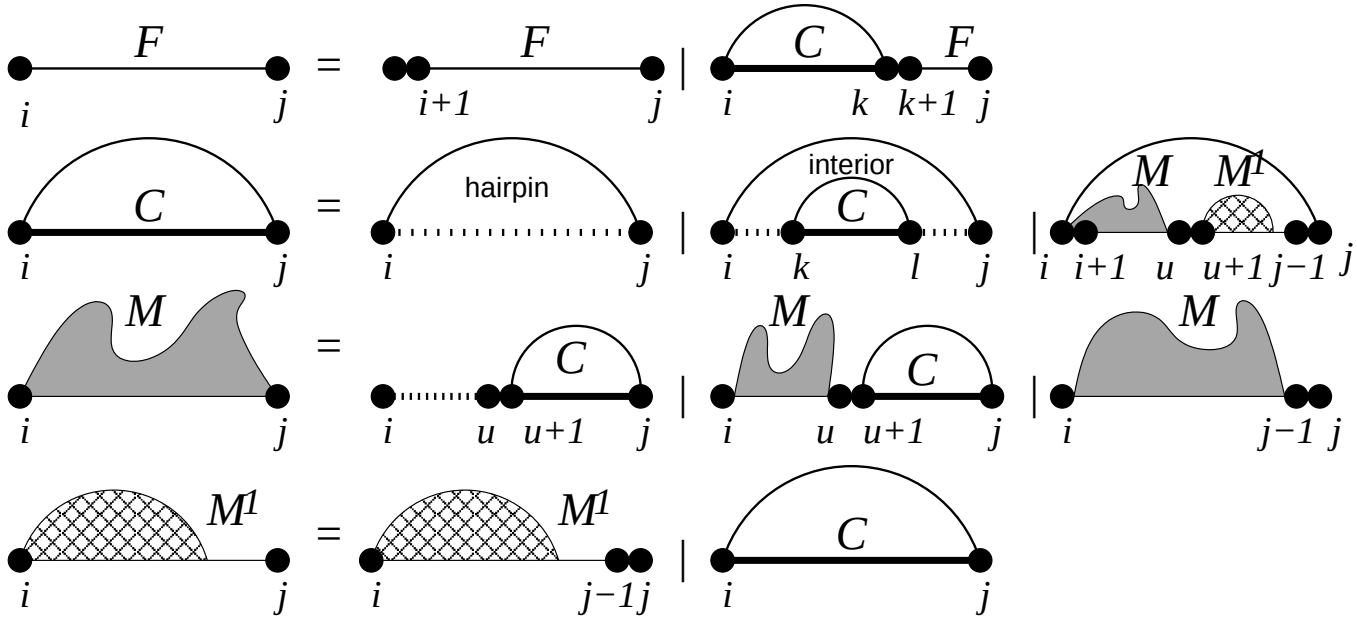
- Some RNAs, e.g. viroid genomes, are circular
- The "opposite" of co-folding
- Easy to implement as yet-another folding variant
- Usual solution doubles memory requirements
- Our solution: post-processing step to linear folding
- implemented for the MFE case in "RNAfold -circ"

Folding circular RNAs

Poor man's solution: just cut the sequence somewhere
Predicted structure depends on the cut-point



Reminder: Folding linear RNAs



F_{ij} best free energy on the subsequence from i to j.

C_{ij} best free energy of a substructure closed by the pair (i, j)

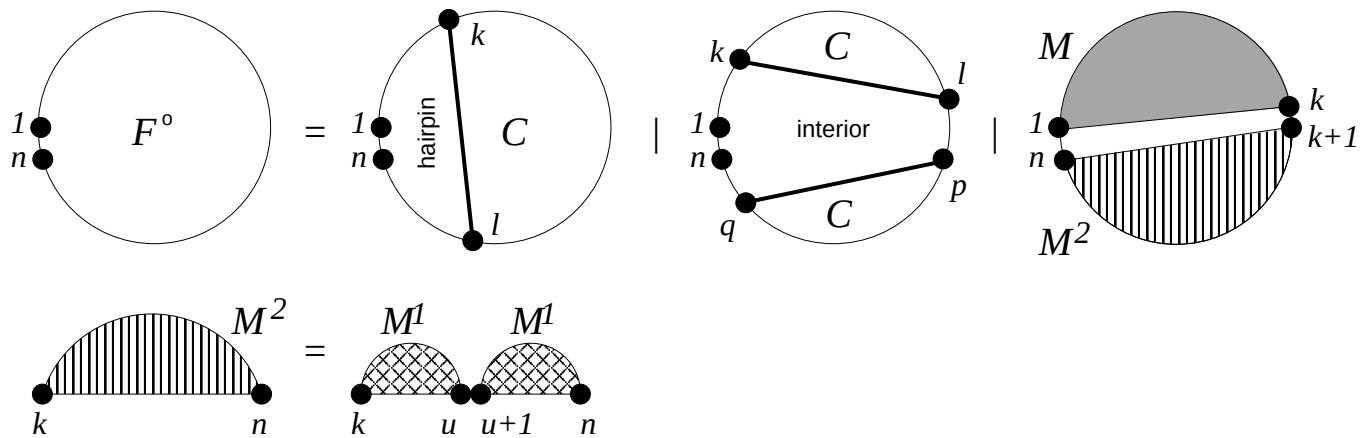
M_{ij} best free energy given that $i..j$ is part of a multi-loop and contains at least on base pair.

M^1_{ij} as M_{ij} but contains exactly **one** pair i, h

Implementing Circular Folding

Relative to linear folding only the loop containing the cut has to be re-evaluated.
Three cases: cut in Hairpin, Interior-, or Multi-loop

$$F^{\circ} = \min\{F^{\circ}_H, F^{\circ}_I, F^{\circ}_M\}$$



Circfold recursion

(1) $F_H^\circ = \min_{p<q} \{C_{pq} + \mathcal{H}(q,p)\}$

(2) $F_I^\circ = \min_{k<l<p<q} \{C_{pq} + C_{kl} + \mathcal{I}(q,p,l,k)\}$

(3) $F_M^\circ = \min_{1<k<n} \left\{ M_{1,k} M_{k+1,n}^2 + a \right\}$

(4) $M_{kn}^2 = \min_{k<u<n} \left(M_{ku}^1 + M_{u+1,n}^1 \right)$

Cofolding two RNAs

Three approaches to predicting hybridization structures

- RNAduplex
  - compute best structure without intra-molecular pairs
  - $\mathcal{O}(n \cdot m)$ time
- RNAcofold
  - Folding by concatenation
  - MFE & partition function
  - computes equilibrium concentrations from free energies
- RNAup
  - computes binding energies as sum of interaction and unfolding $\Delta G_{\mathrm{tot}} = \Delta G_u + \Delta G_i$
  - still a separate download

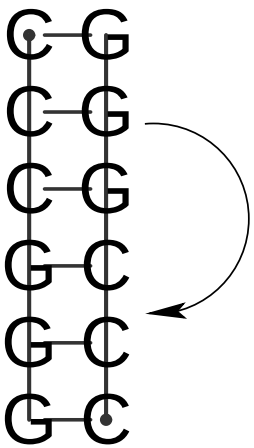
Cofolders Beware

Hybridization is concentration dependent. This means:

- The dot plot produced by "RNAcofold -p" is a weird mixture of monomer and dimer states. Use "RNAcofold -a" to get conditional pair probabilities.
- Previous versions of RNAeval and RNAcofold did not include the "duplex initiation penalty", i.e. all energies are now shifted.
- Symmetric hybrid structures incur a symmetry penalty which is currently ignored.

Cofolding and Symmetry

A homo-dimer with symmetric structure has only half as many conformations – Entropic penalty of $R \ln 2$ (≈ 0.4 kcal/mol).
How can include this correction?



- Partition function case is easy, just subtract $RT \ln 2$ from the free energy: All asymmetric structures are actually counted twice
- MFE case can be difficult:

case a) predicted cofold is asymmetric: no problem.
case b) predicted cofold is symmetric: subtract symmetry penalty *and* make sure all asymmetric structures are still worse in energy.

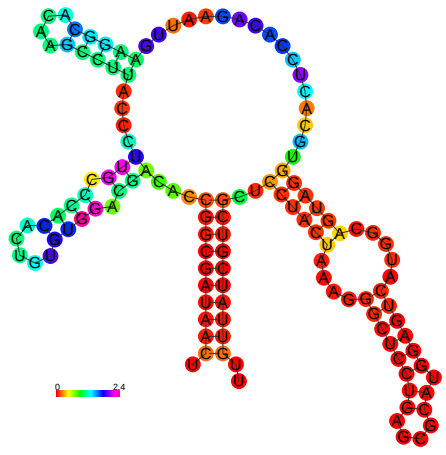
Proper solution requires an algorithm that finds best assymetric structure

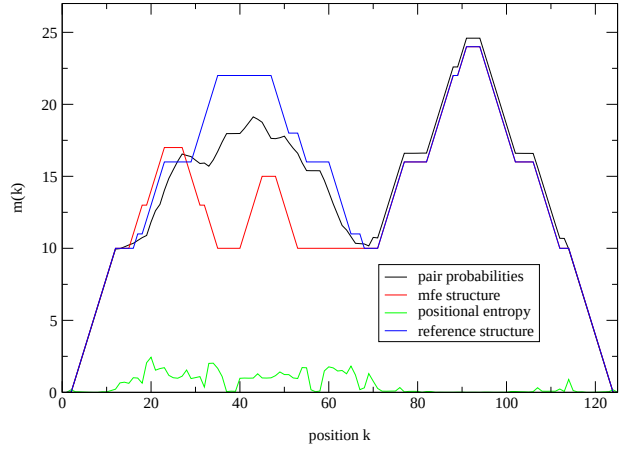
Perl Utilities

Several new Perl utilities help produce cool colorful postscript plots

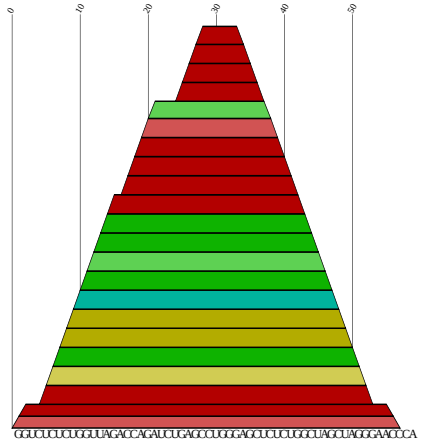
- RNAplot –pre 'some postcript code'
  adds annotation macros and your PS code to a structure plot
- rotate_ss.pl
  rotate or mirror an RNA structure plot
- relplot.pl
  colors an RNA structure by pair probability or reliability
- colorrna.pl and coloraln.pl
  color consensus structure and alignments to show co-variations

Examples






```
                ((..((((((((((.(((((...((((......))))))))))))))))))..)).
HIVMAL          GGUCUCUCUUGUUAGACCAGGUC-GAGCCCGGGAGCUCUCUGGCUAGCAAGG-AACCCA
HIVU455         GGUCUCUCUUGUUAGACCAGAUC-GAGCCUGGGAGCUCUCUGGCUAGCGAGGGAACCCA
HIVCAM1         GGUCUCUCUGGUUAGACCAGAUCUGAGCCUGGGAGCUCUCUGGCUGACUAGGGAACCCA
HIVD31          GGUCUCUCUGGUUGGACCAGAUCUGAGCCUGGGAGCUCUCUGGCUAGCUAGGGAACCCA
HIVLAI          GGUCUCUCUGGUUAGACCAGAUUUGAGCCUGGGAGCUCUCUGGCUAACUAGGGAACCCA
HIVOYI          GGUCUCUCUAGCUAGACCAGAUCUGAGCCCGGGAGCUCUCUGGCUAACUAGGGAACCCA
HIVBCSG3C       GGUCUCUCUGGUUAGACCAGAUCUGAGCCUGGGAGCUCUCUGGCUAGCUAGGGAACCCA
HIVRF           -GUCUCUCUGGUUAGACCAGAUCUGAGCUUGGGAGCUCUCUGGCUAGCUAGGGAACCCA
HIVELI          GGUCUCUCUGGUUAGACCAGAUUUGAGCCUGGGAGCUCUCUGGCUAGCUAGGGAACCCA
HIVZ2Z6         GGUCUCUCUGGUUAGACCAGAUUUGAGCCUGAGAGCUCUCUGGCUAGCUAGGGAACCCA
HIVNDK          GGUCUCUCUGGUUAGACCAGAUUUGAGCCUGGGAGCUC--UGGCUAAUUAGGGAACCCA
HIVANT70        GGUCUCUCGGUUAGAGGACCAGGUCUGAGCCCGGGAGCUCCCUGGCCUCUAGCUGAACCCG
HIVMVP5180      GGUCUUAGUUAGAGGACCAGGUCUGAGCCCGGGAGCUCCCUGGCCUCUAGCUGAACCCG
```



What's missing

- Soon to be added:
  - RNAup
  - local and cofold versions of RNAalifold
  - Z-score calculations in RNALfold
  - Perl bindings for some of the newer algorithms
- Latest Turner energy parameters (including DNA)
- Dependency on ionic conditions
- More structural alignment tools
- More frequent releases

Quote from `http://mjg59.livejournal.com/` Matthew Garrets journal:

Friday, December 16th, 2005
This [`http://www.tbi.univie.ac.at/~ivo/RNA/`] *is excellent. Every time I need to write a new piece of software, I find that it's just been added to Vienna. I'm beginning to get a little suspicious.*