de.NBI and its Galaxy interface for RNA folding

Jörg Fallmann, Jan Engelhardt

Institute for Bioinformatics
University of Leipzig

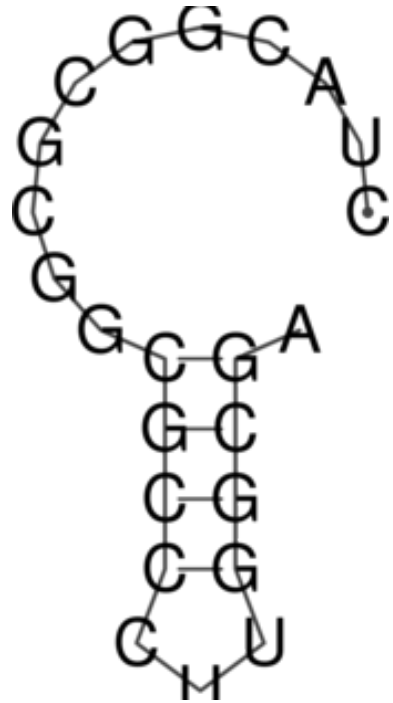
September 29, 2017

You can download the pdfs you will need today here
http://www.bioinf.uni-leipzig.de/~fall/RNA_folding_workshop-presentation.pdf
http://www.bioinf.uni-leipzig.de/~fall/Exercises.pdf

Goal: Use `RNAfold` to do a simple structure prediction.

- Upload the file `rna.fa` into your Galaxy session.
- Start `RNAfold` with standard parameters
- Look into the output

- CUACGGCGCGGCGCCCUUGGCGA
- .....((((...)))). (-5.00)



Goal: Use `RNAfold` to do a structure prediction using the partition function

- Start `RNAfold` using --partfunc
- Look into the output

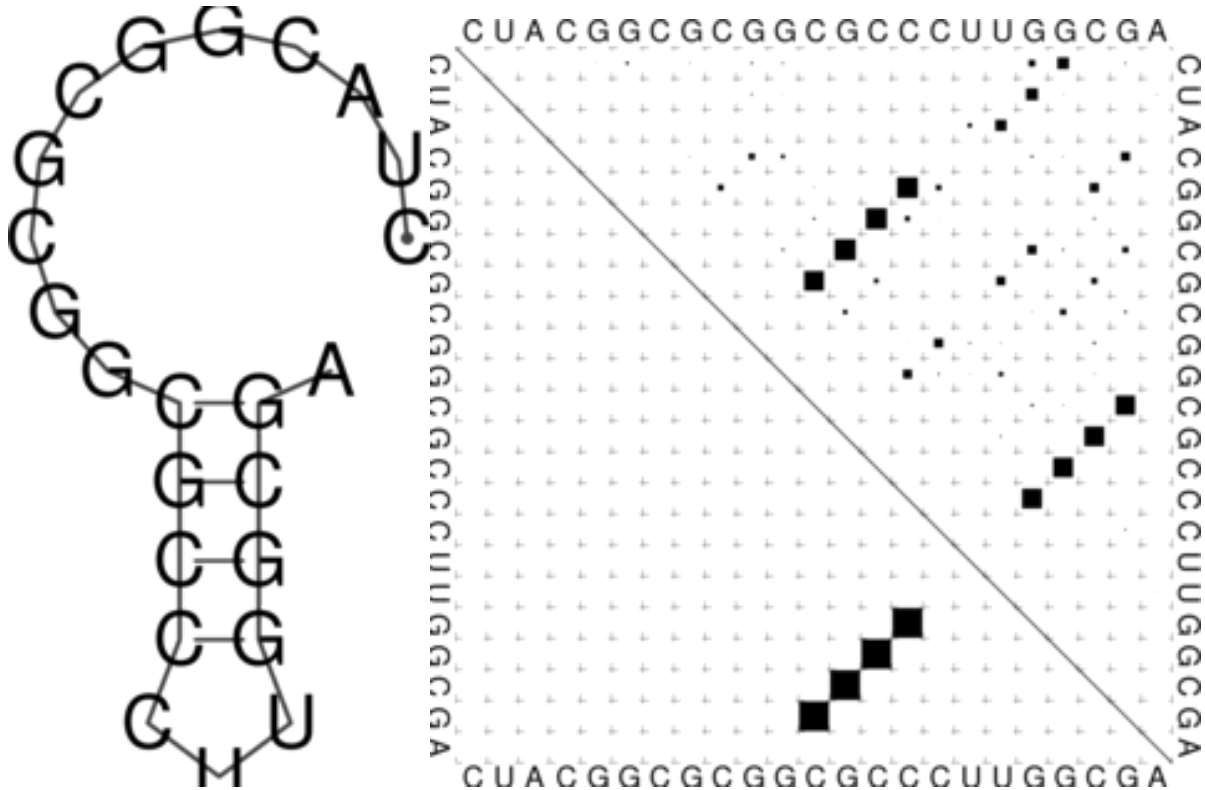
- CUACGGCGCGGCGCCCUUGGCGA
- MFE: .....((((...)))). (-5.00)
- PF: ....{,{{...||||...)}}}. [-5.72]

The partition function is a rough measure for the well-definedness of the MFE structure. The third line shows a condensed representation of the pair probabilities of each nucleotide, similar to the dot-bracket notation, followed by the ensemble free energy ($-kT * ln(Z)$) in kcal/mol

The partition function allows us to calculate the proportion of a certain structure in the ensemble.
Use RNAfold -p to get the ensemble free energy, which is related to the partition function via **F = -RT*ln(Q)**, for the unconstrained (Fu) and the constrained case (Fc), (use option -C), and evaluate **pc = exp((Fu - Fc)RT)** to get the desired probability.

- CUACGGCGCGGCGCCCUUGGCGA
- MFE: ..........((((...)))). (-5.00)
- PF: ....{,{{...||||...)}}}. [-5.72]
- CS: ..... { 0.00 d=4.66 }
- MEA: ......((...))((...))... { 2.90 MEA=14.79 }
- frequency of mfe structure in ensemble 0.311796; ensemble diversity 6.36

Pseudo bracket notation:
Here, the usual '(', ')', '.', represent bases with a strong preference (more than 2/3) to pair upstream (with a partner further 3'), pair down-stream, or not pair, respectively. '', '', and ',' are just weaker version of the above and '|' represents a base that is mostly paired but has pairing partners both upstream and downstream. In this case open and closed brackets need not match up.


C U A C G G C G C G G C G C C C U U G G C G A
C U A C G G C G C G G C G C C C U U G G C G A
C U A C G G C G C G G C G C C C U U G G C G A
C U A C G G C G C G G C G C C C U U G G C G A

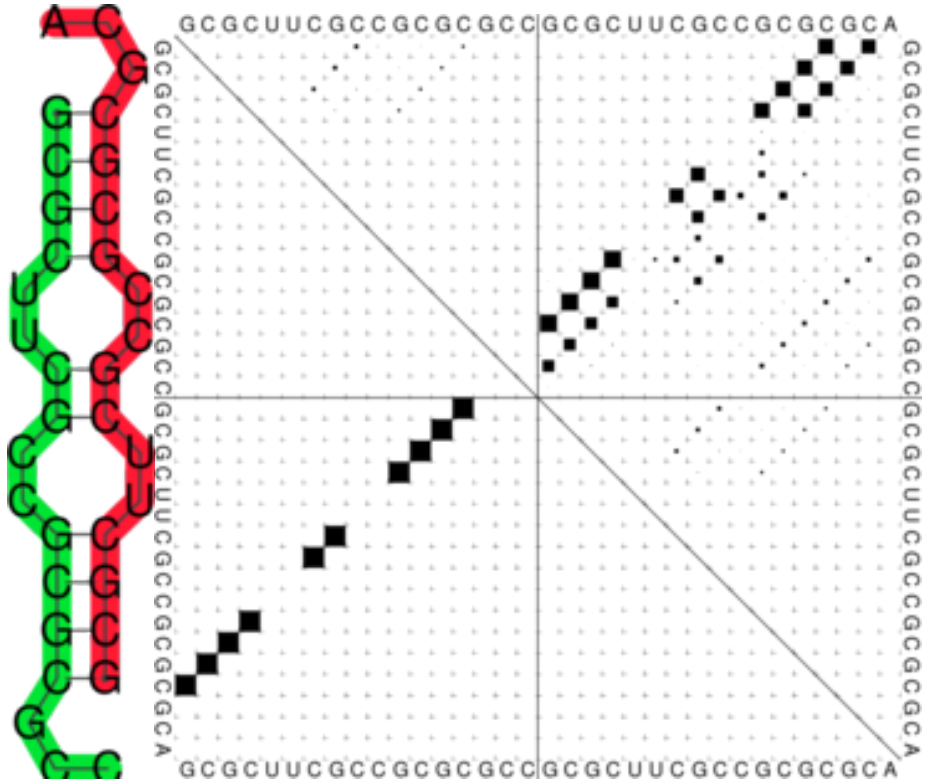
Goal: Use SHAPE-directed `RNAfold` to do a structure prediction

- Upload the file `rna.simple.shape` into your Galaxy session.
- Start RNAfold using the shape file and -shapeMethod=D
- Look into the output

Goal: Use `RNAcofold` to predict the cofolding of two sequences.

- Upload the file `cofold.txt` into your Galaxy session. (Look at it)
- Start RNAcofold using cofold.txt with the --partfunc option
- Look at the output.

- GCGCUUCGCCGCGCGCC&GCGCUUCGCCGCGCGCA
- ((((..((..((((...&))))..))..))))... (-17.70)
- ((((..(,.((((,,.&))))..),.))))),,. [-18.26]
- frequency of mfe structure in ensemble 0.401754 , delta G binding= -3.95



Cofold can use concentrations of molecules for duplex prediction, but this is slow for longer sequences.

Goal: Use `RNAduplex` to predict *only* intermolecular base pairs of two sequences.

- Upload the file `duplex.txt` into your Galaxy session. (Look at it)
- Start RNAduplex using duplex.txt with standard parameter
- Look at the output.

RNAduplex does not use concentrations and neglects intramolecular interactions, faster but less reliable, good prefilter.

Goal: Use `RNAup` to test the `RNAduplex` result.

- Start RNAup using duplex.txt with --include_both
- Look at the output and compare it with the RNAduplex result.

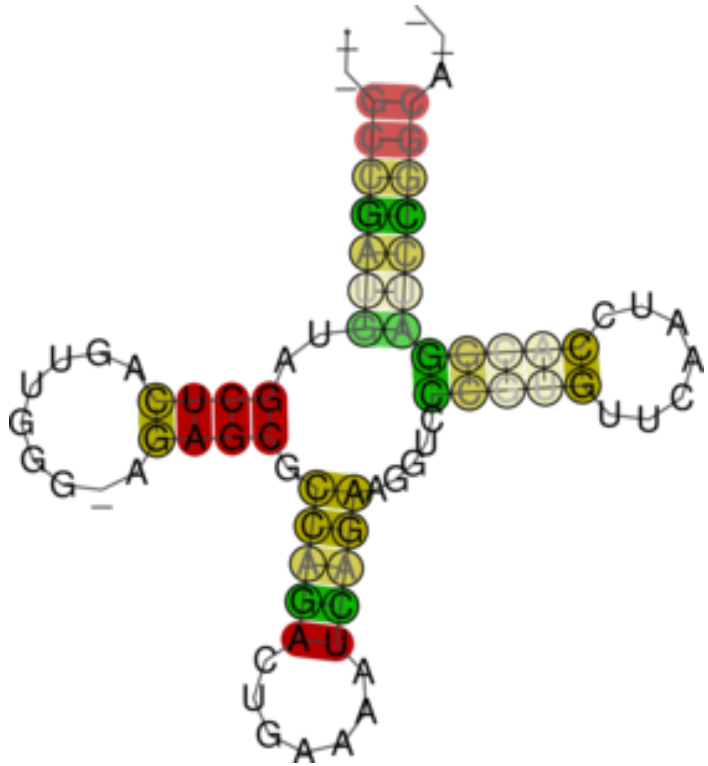
RNAup is also taking intramolecular interactions into account.

Goal: Use `RNAalifold` to predict the consensus structure

- Upload the clustal file `alifold.aln` into your Galaxy session.
- Edit the data type of alifold.aln to 'clustal'
- Use RNAalifold with the alifold.aln and --partfunc (Calculate partition function: 1)
- (Download the output) and look at it
- `Bonus`: Fold the sequences (alifold.fa) individually (RNAfold) and compare the results.

Goal: Use `RNAalifold` to predict *and visualize* the consensus structure

- Use RNAalifold with alifold.aln and --color and --aln
- (Download the output) and look at it



RNAalifold uses covariance information from sequence alignment to predict a consensus structure.

Goal: Use `RNAcode` to predict coding sequences in a MAF alignment.

- Upload the file `oskar.27way.rnacode.maf` into your Galaxy session.
- (Change its data type to maf)
- Use RNAcode with the maf file, --cutoff 0.05, --best_region, --best_hit, with GTF output
- (Download the output) and look at it

Goal: Use `RNAz`

- Upload the file `oskar.27way.rnaz.maf` into your Galaxy session.
- Use RNAz with the maf file
- (Download the output) and look at it