

# Supplementary Material

## Design of RNAs: Comparing Programs for inverse RNA folding

Alexander Churkin, Matan Drory Retwitzer, Vladimir Reinharz,  
Yann Ponty, Jérôme Waldispühl, Danny Barash

### 1 Command Line Interfaces

#### 1.1 `RNAinverse`

The command line interface of `RNAinverse` allows subtle optimizations while the main parameters are inserted upon software request. More advanced options exist for custom alphabet, energy parameters and base pairing. Those would not be discussed here as they are a very rare usecase. For the average user, the following are the ones that will be most used.

- T Rescale energy parameters for a given temperature.
- F Select the minimization algorithm. `m` for energy minimization or `p` for partition function.
- R The number of output sequence to output for the same structure. Negative number will force the software to continue until a perfect match.
- noGU Do not allow GU pairs.
- noClosingGU Do not allow GU pairs at the end of helices.

Therefore, to find a maximum of 50 solutions, using both partition and energy minimization algorithms, for the structure

`(((((...(((....))))...(((....))))...))))`





### 1.3 RNAiFold

The command line interface of `RNAiFold` has over 50 options allowing for an extremely fine tuning of the desired output. For the average user, the following are the ones that will be most used. There is two way to enable those options, or through a file, where the option name is on a line preceded by a `#` instead of a `-`, followed on the next line by the desired option. Usually, the option can be simply given as argument on the command line.

- `-RNAscdstr` The target structure. Multiple target can be set, they must be on the same line separated by the pipe `|` symbol. The structures must have the same length.
- `-RNAseqcon` The admissible sequences, in IUPAC format. It must be one string the same length as the structure.
- `-maxGCcont` The maximal GC content admissible in the sequences.
- `-minGCcont` The minimal GC content admissible in the sequences.
- `-TimeLimit` The amount of time allowed to run (default 600 seconds).
- `-MAXsol` The maximum number of solutions to be reached under the time limit

Therefore, to find a maximum of 50 solutions, under an hour, for the structure

```
(((...(((...)))...(((...))))...)))
```

allowing any sequence with a mandatory GC base between the first and last nucleotide, and with a GC content between 60% and 70%, the command would be as follows:

```
./RNAiFold -RNAscdstr "(((...(((...)))...(((...))))...)))" \  
-RNAseqcon GNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNC \  
-maxGCcont 70 -minGCcont 60 -MAXsol 50 -TimeLimit 3600
```

### 1.4 antaRNA

The distribution of `antaRNA` as a Python2.7 executable and all options are given as arguments. A similar ensemble of constraints exists.

- Cstr** The target structure in the dot bracket notation. A fuzzy notation can be used to define blocks allowed to base pair together using any lowercase and uppercase letter.
- Cseq** The admissible sequences, in IUPAC format. It must be one string the same length as the structure.
- tGC** Target GC content, in  $[0, 1]$ , which also serves as a minimum.
- tGCmax** Maximal GC content admissible in the sequences.
- tGCvar** Variance ( $\sigma^2$ ) in the case of normal distribution, -**tGC** serves as the expected value  $\mu$ .
- t** The amount of time allowed to run (default 600 seconds).
- n** Number of solutions to be produced.

Therefore, to find a maximum of 50 solutions, under an hour, for the structure

((((...(((.....)))...(((.....))))...))))

allowing any sequence with a mandatory GC base between the first and last nucleotide, and with a GC content between 60% and 70%, the command would be as follows:

```
python antaRNA.py -Cstr "((((...(((.....)))...(((.....))))...))))" \
-Cseq GNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNC \
-tGC 0.6 -tGCmax 0.7 -n 50 -t 3600
```

Pseudoknotted structures can be considered with the -p parameter if pKiss\_mfe or HotKnots or IPKnot is installed,.

In addition, all parameters of the ant colony search algorithms can be directly modified through the command line, from the random seed to initiate the search -s, the number of ants exploring (-aps, default 10), the pheromone evaporation rate(-er, default 0.2), and a wealth of others.

