CondAlt 2.1 Manual

Amirhossein Manzourolajdad and John L. Spouge

Introduction

The *CondAlt 2.1* Software Package predicts alternative structures of an RNA sequence based on conditional base-pair probabilities, as follows. First, the computed MFE structure S_1^* predicts the dominant structure. Then, the *UNAFold* Software [2, 4-7] computes McCaskill base-pair probabilities conditioned on excluding base pairs close to (i.e., with endpoints near) base pairs in S_1^*. A calculation of McCaskill conditional base-pair probabilities yields a longest bulge-containing stem seed L^*, in which the conditional probability of every base-pair is higher than 0.5. Finally, we predict an alternative structure S_2^* as the lowest-energy structure containing L^*. the Materials and Methods Section of "Structural Prediction of RNA Switches using Conditional Base-Pair Probabilities" article gives algorithmic details.

Folding-energy Parameters

Calculation of base-pair probabilities as well as conditional base-pair probabilities are computed via version 3.0 energy parameters of *UNAfold* (version 4.0.0) [8].

Installation

UNAfold [2] has to be already installed and available in the main path. The downloadable zip file contains the stand-alone main program in Perl 'altmfe.pl' and accompanying java programs required for the main program. All input and output files are to be in the same folder as 'altmfe.pl'.

Input Parameters

1. -i <fasta file>: Input RNA sequence. The name of file MUST have a ".seq" extension. The input must be in fasta format and start with '>' in its first line. There must be only one RNA sequence in the following line and it must contain unambiguous nucleotides; A, C, G, and U. The end of file must contain at most one newline. The length of the sequence must be in the range of [1,500].
2. -d <distance threshold>: Dissimilarity Threshold τ. This parameter is used to define the size of the meta-stable structure to be eliminated in conditional probabilities calculations; i.e., $\mathfrak{S}$. Higher τ leads to higher number of eliminated base pairs. Default is set to 5.
3. -t <temperature>: Folding temperature T. S_2^* can be predicted at different temperatures. For a given T, S_1^* is still predicted at the default temperature 37°C, but quantities $\mathcal{P}$, L^*, and S_2^* are all calculated at T. See Materials and Methods.

Output Files

Both the alternative structures and the seed locations are available to user. After running the Perl file with the above input parameters, two output files: [input.seq].out and [input.seq]MFE1.aux.

1. [input.seq].out contains for the dominant structure S_1^* and alternative structure S_2^* (in that order): (1) the free energy; (2) the sequence; and (3) the resulting structures in Stockholm format.
2. [input.seq]MFE1.aux contains the base-pairs in the stem seed L^*.

A Simple Example:

Consider the following input RNA sequence and parameters:

<fasta file>=input.seq

```
>seq
CGUACGUAGCUAGUCGUACGUAGCUACGUACUGA
```

<distance threshold> = 5

<temperature> = 37

Because of the defaults, the following commands are equivalent:

```
perl -i altmfe.pl input.seq -d 5 -t 37

perl -i altmfe.pl input.seq -d 5

perl -i altmfe.pl input.seq -t 37

perl -i altmfe.pl input.seq
```

They yield:

input.seq.out:

```
>-17.5
CGUACGUAGCUAGUCGUACGUAGCUACGUACUGA
.((((((((((.......))))))))))...
>-6.4
CGUACGUAGCUAGUCGUACGUAGCUACGUACUGA
..............((((((...))))))...
```

And input.fastaMFE1.aux:

```
F 16 31 1
F 17 30 1
F 18 29 1
F 19 28 1
F 20 27 1
F 21 26 1
```

References

1. Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, Schuster P. Fast folding and comparison of RNA secondary structures (the Vienna RNA package). Monatshefte für Chemie / Chemical Monthly. 1994;125(2):167-88.
2. Markham NR, Zuker M. UNAFold: software for nucleic acid folding and hybridization. Methods in molecular biology. 2008;453:3-31.
3. McCaskill JS. The Equilibrium Partition Function and Base Pair Binding Probabilities for RNA Secondary Structure. Biopolymers. 1990;29.
4. Jaeger JA, Turner DH, Zuker M. Predicting optimal and suboptimal secondary structure for RNA. Methods in enzymology. 1990;183:281-306.
5. Zuker M. On finding all suboptimal foldings of an RNA molecule. Science. 1989;244(4900):48-52.
6. Zuker M. Prediction of RNA secondary structure by energy minimization. Methods Mol Biol. 1994;25:267-94.
7. Zuker M, Mathews DH, Turner DH. Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide. In: Barciszewski J, Clark BFC, editors. RNA Biochemistry and Biotechnology. Dordrecht: Springer Netherlands; 1999. p. 11-43.
8. Mathews DH, Sabina J, Zuker M, Turner DH. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. Journal of molecular biology. 1999;288(5):911-40.