

## A Simple Protocol for the Inference of RNA Global Pairwise Alignments

### Recap

Similar secondary structures are usually similar in functionality.

### SPS

Is called the Sum of Pairs Score, Similarity is:

$1 - (\text{edit distance} / \text{unaligned length of shorter sequence})$

### Sequence Identity

The identity is the number of identical nucleotides divided by the shorter sequence length, so a measure for how many parts of the sequences are 'equal' or so to say.

### needle

**needle** is an implementation of the sequence based Needleman-Wunsch algorithm, which is basically only the classical edit-distance (Levenstein-distance) algorithm.

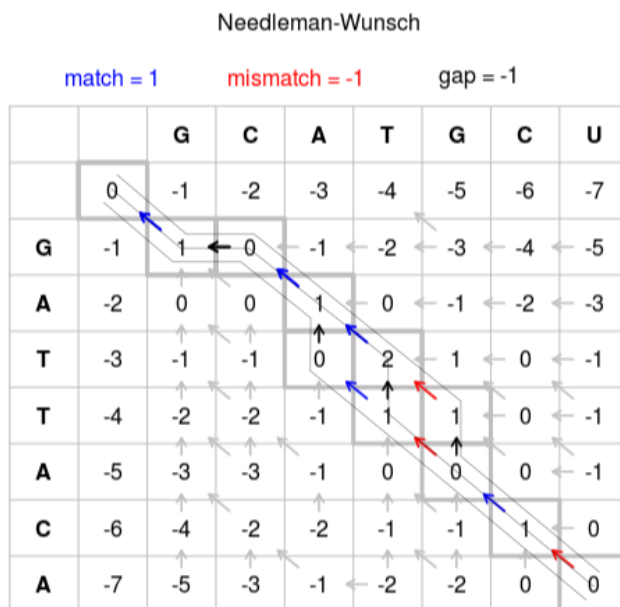
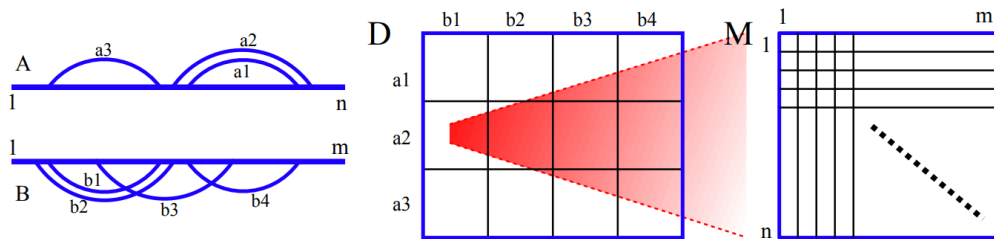
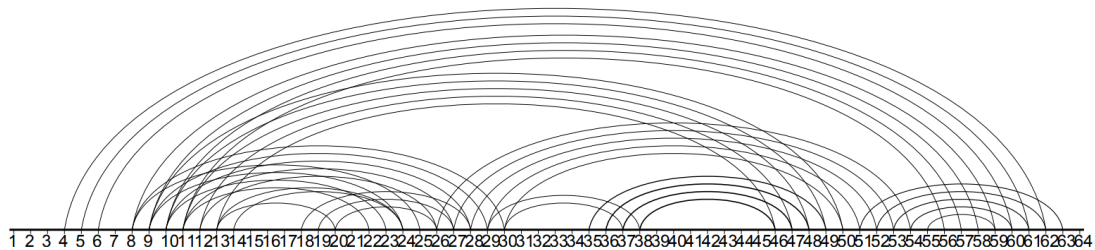


Diagram illustrating the classification of RNA secondary structure changes. The top part shows a sequence with arcs: `---.---.(( (....)))...-.-..`. Below it, two sequences are aligned: `---A---CGUUGACUGACAAC---AC` and `ACGAUCACGU--ACUAGC--CUGAC`. The bottom sequence has arcs: `....(( ( (---....)) ---.))`. Labels with arrows indicate: **arc altering** (top), **arc removing** (top), **deletion** (left), **base match** (bottom left), **arc mismatch** (bottom middle), and **arc match** (bottom right).

## LocARNA

Is a folding and aligning sequence algorithm. First a base pair probability matrix is being created using RNAfold, this is then used to prune possible arcs. Only these with realistic chances are further considered.

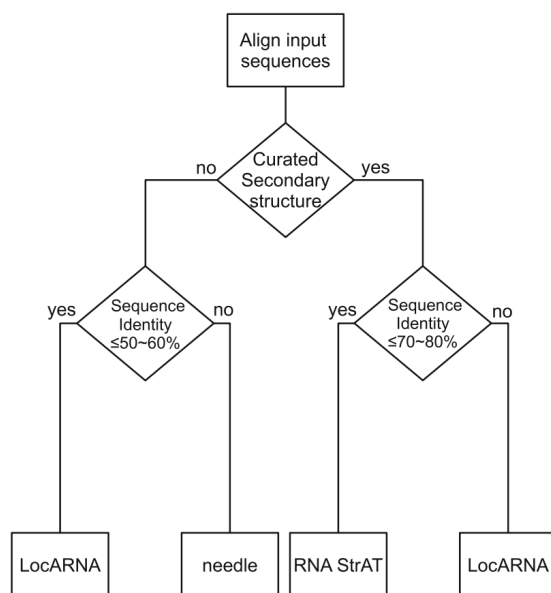
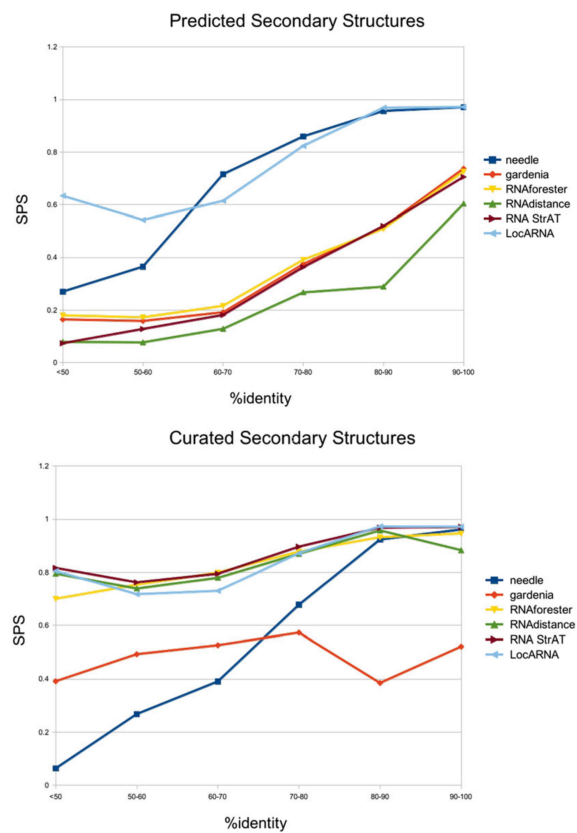
$$p_{\text{cutoff}} = 0.1$$



Then, Alignment scores for all of these realistic arcs are computed, caching smaller one's first. Meanwhile, the full path matrix is being filled out with values.



## Comparison



<https://github.com/fkarg/things-to-talk-about/tree/master/bioinfII>