

Algorithms in Structural Bioinformatics

I. Emiris and E. Chrysina, Assignment 1

Announced 15/03/21, Deadline: 28/03/21

Assignments must be submitted on **e-class: Assignments** as a single file whose name includes your last name. For each question, include a PDF with your answers; further files (e.g. code) may be included.

1. RNA folding

Find all optimal secondary structures of the RNA sequence *AAUACUCCGUUGCAGCAU* with the following simplified Zuker minimization algorithm. Starting from the slides' algorithm, use the following initialisation:

$$j + 5 > i \implies V(i, j) = W(i, j) = \infty, \quad i > j,$$

hairpin energy $h(i, j) = 2(i - j + 5)$, and stem energy $s(i, j) = -4, 0, 4$, for Watson-Crick bonds, *GU*, and all other possible pairs respectively. Ignore multiloops and buldges to simplify V so that it only has the first two cases.

Implement your algorithm in Matlab, R, Python or other convenient system; submit your code. Print the filled-in tables W, V . Draw (by hand) an optimal fold, show its bonds, and the corresponding backtrack path.

2. c-RMSD and d-RMSD

Given are 80 conformations of a specific molecule in file "80_conformations.txt" on **e-class: Documents/Assignments** with $n = 369$ atoms on the backbone (hence in correspondence). The file starts with 2 lines containing 80 and n . The rest of the file uses tabs to define 3 columns containnig n triplets $x \ y \ z$ of each conformation hence $2 + 80n$ rows:

```
80
369
2.816 -11.005 10.087
4.43 -10.545 10.011
...
```

Implement c-RMSD and d-RMSD in Matlab, Octave, Maple or other system offering linear algebra (SVD); submit your code. If your system provides either of these functions, it is OK to just use it.

1. Use c-RMSD to compute the optimal translation and rotation minimizing c-RMSD between the first 2 conformations.
2. Compute the c-RMSD distances between all $\binom{80}{2}$ pairs of conformations. Compute the mean and median c-RMSD distance and draw a histogram of all distances in 10 classes.
3. Repeat (2) for d-RMSD using all $k = \binom{n}{2}$ distances within each conformation, or a random subset of $k = 3n$ distances.