

## Challenge 2: Quest for Stable RNA Components

RNA tectonics is an approach that utilizes known RNA structural components to build complex RNA devices {Grabow:2011fc, Shu:2004dv, Afonin:2010ip, Delebecque:2011fc}. We have extracted over 2,500 unique RNA structural components from known X-ray and NMR structures and are challenging you to design loops that stabilize the known secondary structure. This concept is illustrated below.

[SORRY NOT AVAILABLE YET]

You are given the following datasets to work with.

```
/challenges/data/components/  
    /comps  
        comps_[id]_[component_num].ent  
            FORMAT  
                PDB file format for the RNA component  
    /seqs  
        comps_[id]_[component_num].seq  
            FORMAT  
                (file_path) (sequence with breaks)  
    /ss  
        ss_comps.txt (text file containing the component id and a  
            secondary structure  
            comps_[id]_[component_num] ((((...))))
```

Your goal is to identify sequences that replace the [BREAKS] in the sequence files, such that the target secondary structure is still maintained. The following is a manual example.

### Step 1: Get the Inputs

For each structure in the /ss/ss\_comps.txt file get the sequence from the corresponding comps\_[id]\_[num].seq file

```
/seqs  
comps_2aar_54.seq  
2aar_comps.txt,AAGCCGAAGUGGC[BREAK]GCUACACCUCAGAAGGUGAGAGUCC  
UGUAGGCGA
```

```
/ss  
ss_comps.txt  
2aar_54 ..(((...((((()))))(((...).(((.....))))))...)).. 35.955
```

### Step 2: Align the target structure with the sequence

```

..(((...((((          )))))(((...).(((.....))))...)))..
AAGCCGAAGUGGC [ BREAK ] GCUACACCUCAGAAGGUGAGAGUCCUGUAGGCGA

```

**Step 3: Replace [BREAK] with a sequence such that when the sequence is folded with any secondary structure prediction program, the minimal energy fold contains the target secondary structure.**

```

..(((...((((.....))))))(((...).(((.....))))...)))..
AAGCCGAAGUGGCGAACGCUACACCUCAGAAGGUGAGAGUCCUGUAGGCGA

```

```

>RNAfold
AAGCCGAAGUGGCGAACGCUACACCUCAGAAGGUGAGAGUCCUGUAGGCGA
..(((...((.(.....((.(.....))))))..))))..
minimum free energy = -16.60 kcal/mol

```

FAIL! My guess didn't produce the correct secondary structure.

**Step 4: Save data for each design in the following format**

```
2aar_54,[GAAC],ensemble defect,MFE defect,...
```

where the first column contains the component\_id, the second column contains sequences replacing [BREAK]. If multiple [BREAK] exist use the format [SEQ1,SEQ2,SEQ3...]. The remaining columns are metrics that define how far from the target secondary structure the given component with the designed sequence is such as ensemble, MFE, or probability defect {Zadeh:2011ee}.

