

# AST Übung 8 - RNA

## Einführung

Ziel dieser Übung ist es sich kurz mit der Arbeit mit RNA Strukturen vertraut zu machen. Es soll die Sekundärstruktur eines Moleküls ermittelt und mit Hilfe von Web-basierten Tools vorhergesagt werden. Anschließend wird mit Sequenzdesign experimentiert.

## Klammernotation

Öffnen Sie das Modell 2jxv in *Chimera*. Hierbei handelt es sich um eine mit NMR Spektroskopie ermittelte Struktur. Schließen Sie zur Übersicht daher alle bis auf eines der Modelle (zunächst im Model Panel die Gruppierung der Modelle mit group/ungroup aufheben).

Betrachten Sie das Molekül und notieren Sie sich dessen Sekundärstruktur in der Klammernotation (*dot-bracket, vienna*). Falls Ihnen in der ribbon Darstellung die Basenpaare nicht deutlich genug erscheinen, können Sie zu einer Atomdarstellung wechseln und mögliche Wasserstoffbrückenbindungen anzeigen lassen (Tools -> Structure Analysis -> FindHBond). Nehmen Sie für ein Basenpaar mindestens zwei Wasserstoffbrückenbindungen an. Zur Übersicht kann es hilfreich sein die Rückgratatome zu verstecken. Diese können Sie mit Select -> Structure -> backbone -> full auswählen.

## Sekundärstrukturvorhersage

Nutzen Sie im Folgenden den *ViennaRNA* Web Services Server, welcher verschiedene Programme zum Arbeiten mit RNA Strukturen über ein Webinterface bereitstellt:

<http://rna.tbi.univie.ac.at/>

Mit dem RNAfold Tool können Sie für eine RNA Sequenz die Sekundärstruktur mittels Energiefunktionen unter verschiedenen Parametern vorhersagen. Kopieren Sie die Sequenz der RNA (Tools -> Sequence -> Sequence und anschließend Edit -> Copy Sequence) und führen Sie eine Sekundärstrukturvorhersage mit den Standardparametern durch. Betrachten Sie die Ergebnisseite und vergleichen Sie die vorhergesagte Struktur mit der tatsächlichen.

## Sequenzdesign mit Simulated Annealing

Kopieren Sie sich die folgende Datei in ihr Homeverzeichnis:

```
cp /home/olzhabaev/Public/rna_design_annealing.py .
```

Hierbei handelt es sich um ein Python Skript mit dem "*Simulated Annealing*" Algorithmus aus der vorherigen Übung. Dieses wurde für RNA Sequenzdesign angepasst. Die Zielfunktion (Z. 72) ruft hierbei ein externes Programm auf, welches für eine Sequenz und eine Sekundärstruktur die Wahrscheinlichkeit schätzt, mit der sich die Sequenz in die Struktur faltet. Dieser Wert wird zurückgegeben und im Algorithmus maximiert. Initial wird eine zufällige Sequenz erzeugt. Testschritte erzeugen zufällige Mutationen in der Sequenz, wobei Basenpaare entsprechend angepasst werden.

Fügen Sie die Sekundärstruktur des Moleküls in die Anführungszeichen in Zeile 89 ein und speichern Sie die Datei. Starten Sie eine Konsole und führen Sie das Skript aus:

```
python rna_design_annealing.py
```

Pro Iteration wird die aktuelle Sequenz und entsprechende Falt-Wahrscheinlichkeit ausgegeben. Passen Sie ggf. die Parameter des Algorithmus (Z. 98-100) an um bessere Ergebnisse zu erhalten und lassen Sie das Skript mehrmals Laufen. Versuchen Sie eine Sequenz zu mit einer möglichst hohen Falt-Wahrscheinlichkeit zu erzeugen.

Kopieren Sie die beste gefundene Sequenz und Öffnen Sie zwei Instanzen des Vienna RNAeval Tools. Hiermit können für eine gegebene Sequenz und Sekundärstruktur die Energie und Energiebeiträge berechnet werden. Führen Sie dies für die Sequenz von 2jxv und die durch den Algorithmus gefundene Sequenz durch, und vergleichen Sie die Ergebnisse. Welche Sequenz wird sich eher in die Struktur falten und welche Teile der Sequenz / Struktur tragen dazu bei?

## Fragen

- i. Nennen Sie Vorteile der Klammernotation zur Darstellung von RNA Sekundärstrukturen.
- ii. Welches Problem würde bei der Darstellung einer RNA Sekundärstruktur mit Pseudoknoten mittels der Klammernotation auftreten? Wie könnte es sich lösen lassen?
- iii. Warum haben G/C Basenpaare in der Regel bessere Energiebeiträge?