

11. Übungsblatt “Einführung in die Bioinformatik” SS 2008

Prof. Dr. Daniel Huson und Juliane D. Klein

27. Juni 2008

1 Nussinov Algorithmus (7 Punkte)

Bitte implementieren Sie den Nussinov Algorithmus einschließlich des Tracebacks. Als Eingabe soll eine RNA-Sequenz im Fasta-Format eingelesen werden. Die Ausgabe soll die DP-Matrix, den Score und die berechnete Struktur beinhalten. Die Struktur wird als eine Liste von Basenpaarungen angegeben. Diese Liste wird in einen String geschrieben, den man mit dem Aufruf der Funktion `getSecondaryStructure()` erhält.

Ein Beispiel für eine Eingabe mit zugehöriger Ausgabe finden Sie in `example.fasta` und `example.txt`.

Implementieren sie für diese Aufgabe die Klasse `Nussinov` und verwenden sie die Funktionen der Basis-klasse `NussinovBase`. Kompilieren Sie das Programm wieder zusammen mit den Java-Dateien der Klassen `FastA` und `FastABase`.

2 Zuker Algorithmus (2 Punkte)

Bitte laden Sie sich die Sequenz des *persimmon viroids* (AB.366022) von der Seite des NCBI's herunter. Bestimmen Sie nun für die RNA-Sequenz mögliche Sekundärstrukturen mit Hilfe des Zuker Algorithmus. Wählen Sie dafür unterschiedlichen Temperaturen (37, 60, 75 Grad). Nutzen sie dafür das Tool unter <http://frontend.bioinfo.rpi.edu/applications/mfold/cgi-bin/rna-form1-2.3.cgi>.

Speichern Sie jeweils eine geeignete Struktur im pdf Format ab und schicken Sie diese an Ihren Tutor. Welche der gefundenen Strukturen ist am wahrscheinlichsten und warum?

Es gibt zwei große Virenfamilien. Machen Sie sich kurz vertraut mit ihnen und ihrer typischen Struktur. Zu welcher Familie könnte dieser Virus gehören? Welche andere Familie von Viren gibt es und wie sieht deren RNA-Struktur aus? Wählen Sie selber einen Vertreter dieser anderen Familie und wiederholen sie die Strukturvorhersage bei einer geeigneten Temperatur. Schicken Sie auch dieses Bild bitte an ihren Tutor.

3 Charakterisierung von Proteinsekundärstrukturen (1 Punkt)

Gehen Sie bitte auf die Seite www.rcsb.org und suchen Sie den Eintrag für das *thyroxin-binding globulin* (2CEO). Lassen Sie sich die beiden verschiedenen Möglichkeiten für die Sekundärstruktur anzeigen. Wieviele Helices, Beta-Sheets und wieviele Turns werden jeweils vorhergesagt?

Laden Sie die Darstellung “display FASTA with Secondary Structure” herunter und schicken Sie diese an Ihren Tutor.

Allgemeiner Hinweis: Alle Fragen, die in den Aufgaben gestellt werden, gehen in die Vergabe der Punkte mit ein.