

# Angewandte Bioinformatik I - Übung

## Strukturbioinformatik: RNA

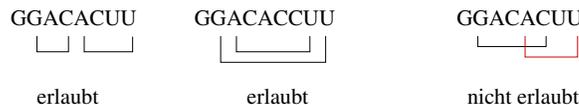
Alle Daten, die Sie zum Lösen der folgenden Aufgaben benötigen, können Sie sich von der folgenden Seite herunterladen. Dort finden Sie auch die Links zu allen hier beschriebenen Werkzeugen.

<http://gobics.de/lectures/appliedBI/rna/>

### Strukturvorhersage durch Maximierung von Basenpaarungen

Einer der ersten Ansätze zur Vorhersage von Sekundärstrukturen bei RNAs, basierte auf der Maximierung der Anzahl von konsistenter Basenpaare. Im Folgenden sollen Sie sich mit Sekundärstrukturen und verschiedenen Darstellungsformen vertraut machen.

**Aufgabe 1** Bestimmen Sie für die gegebene Sequenz: AGUAAACAGC die Sekundärstruktur mit der maximalen Anzahl an Basenpaaren, wobei zwischen zwei paarenden Basen mindestens eine Ungepaarte stehen soll. Überlegen Sie sich dazu, welche Basen überhaupt miteinander Paaren können und wählen Sie dann die Paare aus, die sich nicht widersprechen, d.h. sich nicht überkreuzen:



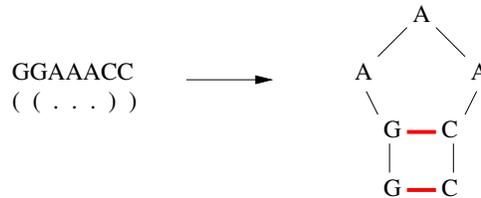
Geben Sie die endgültige Struktur in der Vienna-Darstellung an. Hier steht ein Punkt für eine nichtgepaarte Base und ein komplementäres Klammernpaar für eine Paarung. Für die Sequenz GGAAACC, in der nur G und C paaren können, wird die Struktur z.B. wie folgt dargestellt:

```
GGAAACC
((...))
```

**Aufgabe 2** Stellen Sie die Sequenz und ihre Struktur graphisch dar:

AGGAAGGCAAUGCCCUGACCAUCUCCA  
 .(((.(((...))..((...))))).

Bsp:



### Mfold-Webserver zur Vorhersage von RNA Sekundärstrukturen

Die Weiterentwicklung des Algorithmus zur Maximierung der Anzahl von Basenpaaren bei der Strukturvorhersage resultierte in einem Ansatz, der die bei der Faltung freiwerdende Energie berücksichtigt. Die Theorie besagt: Je mehr Energie bei einer chemischen Reaktion frei wird, umso stabiler ist die entstandene Bindung und umso wahrscheinlicher, dass sich genau diese Bindungen real ausbilden. Mit der neuen Strategie soll deshalb die Struktur mit der minimalen freien Energie (MFE), auch optimale Struktur genannt, bestimmt werden. Sie werden im Folgenden das Programm Mfold kennen lernen, das nach diesem Prinzip arbeitet.

Das M in Mfold steht für „multiple“ und deutet bereits an, dass neben der MFE Struktur noch weitere Strukturen für eine Sequenz berechnet werden können. Es ist möglich bestimmte Restriktionen für die Faltung vorzugehen, so kann z.B. festgelegt werden, dass bestimmte Bereiche einer Sequenz ungepaart bleiben sollen.

**Aufgabe 3** Laden Sie sich die Sequenz der DsrA RNA von Ecoli auf Ihren Desktop herunter. Die DsrA RNA ist eine nichtkodierende RNA, welche sowohl Transkription als auch Translation reguliert. Unerlässlich für ihre regulatorische Funktion ist die Struktur. Im Folgenden sollen Sie sich die Sekundärstruktur dieser RNA mit Mfold vorhersagen lassen.

Gehen Sie auf die Seite von Mfold auf dem „Rensselaer bioinformatics web server“. Kopieren Sie die Sequenz der DsrA RNA in das dafür vorgesehene Fenster und tragen Sie einen Namen für die Sequenz ein. Mit dem Knopf  starten Sie die Faltung.

- Die graphische Darstellung der Struktur finden Sie unter dem Punkt *Individual Structures*. Warum gibt es hier mehr als nur eine Struktur? Um die Frage beantworten zu können, schauen Sie sich die Beschreibung der Parameter: *percent suboptimality* und *window* an. Probieren Sie ihren Einfluss auf die Faltung aus.
- Vergleichen Sie die gegebenen Strukturen miteinander. Dazu wählen Sie auf der Ergebnisseite unten die drei Strukturen aus und klicken auf den Knopf . Was bedeuten die farbigen Hervorhebungen?
- Was stellt der „Energy Dot Plot“ dar? Wie unterscheidet er sich von dem Dotplot beim Strukturvergleich mit allen Strukturen?

Wissen Sie welche Bereiche einer RNA miteinander paaren müssen bzw. welche Abschnitte keine Bindung eingehen dürfen, dann können Sie diese Informationen in die Faltung einfließen lassen.

**Aufgabe 4** Gehen Sie auf die Startseite von Mfold. Schauen Sie sich die Beschreibung zu *constraint information* an.

- a) Setzen Sie die Parameter so, dass die ersten 3 Basen der Sequenz gepaart werden. Lassen Sie sich eine neue Struktur mit diesen Vorgaben berechnen und vergleichen Sie die neuen Strukturen mit der ursprünglich optimalen Struktur. Hat Ihr Vorhaben geklappt?
- b) Setzen Sie die Parameter so, dass bei der Faltung der DsrA RNA die Basen 4-10 ungepaart bleiben.
- c) Wie können Sie die Faltung mit Hilfe der *constraint information* noch beeinflussen?

Weitere Informationen zur Benutzung des Mfold-Servers gibt es in dem Artikel „Mfold web server for nucleic acid folding and hybridization prediction“ von M. Zuker. Es gibt einen Link auf diesen Artikel auf der Seite des Mfold-Servers.

## **RNAalifold : Vorhersage einer gemeinsamen Sekundärstruktur für eine Schar von Sequenzen**

Sie haben einen Datensatz von mehreren RNA Sequenzen gegeben und interessieren sich für ihre strukturellen Gemeinsamkeiten. Eine mögliche Strategie für das Problem sieht vor, die Sequenzen mit einem Sequenzalignmentprogramm zu aligniert und anschließend die Struktur zu den alignierten Sequenzen zu bestimmt.

An dieser Stelle soll die gemeinsame Struktur für einen Satz von vier Sequenzen, die zur Familie der *srpB* RNA gehören, bestimmt werden. *srpB* RNA kommt bei verschiedenen Enterobakterien vor, doch ihre Funktion ist bis heute nicht ganz geklärt. Dennoch geht man davon aus, dass die Struktur darin eine wichtige Rolle spielt.

**Aufgabe 5** Laden Sie sich die Fasta-Datei der *srpB* RNA Sequenzen herunter. Alignieren Sie diese Sequenzen mit dem Sequenzalignmentprogramm ClustalW, und speichern Sie das Alignment lokal auf Ihrem Desktop. Gehen Sie zum Webinterface von RNAalifold und lassen Sie sich Ihr Alignment falten. Dazu wählen Sie mit  das Alignment der *srpB* RNAs, das jetzt auf Ihrem Desktop liegen müsste und klicken auf .

- a) Vergleichen Sie die ausgegebene Sequenz und Struktur mit dem ursprünglichen Alignment.

## **Links zu Werkzeugen:**

RNAalifold : <http://rna.tbi.univie.ac.at/cgi-bin/alifold.cgi>

ClustalW : <http://www.ebi.ac.uk/Tools/clustalw/index.html>

Mfold: <http://frontend.bioinfo.rpi.edu/applications/mfold/>