

Aufgabenblatt 10

Needleman-Wunsch und Gap-scores

Abgabe: **18.01.2005** bis **17 Uhr** über Goya

Maximal: 20 Punkte

Namen nicht vergessen!

1. Implementiert den Algorithmus zum Befüllen der Tabelle so wie er in der Vorlesung (Foliensatz 12, Folie 24) vorgestellt wurde. Darüber hinaus sollen die Kosten für ein Insert/Delete, ein Replace und ein Match frei wählbar sein.

Der Aufruf der Java-Klasse soll durch

```
java GruppeXAufgabe10 -A datei1 -B datei2 -R  $c_r$  -I  $c_{i/d}$  -M  $c_m$ 
```

erfolgen.

Die Ausgabe des Programms soll folgendes beinhalten:

- (a) die erstellte Matrix des globalen Aligments **3 P**
(b) alle besten Alignments, die durch das Traceback gefunden wurden. **4 P**

Gebt sowohl die erstellte Matrix als auch die besten Alignments für die beiden Beispielstrings mit den Kosten für ein

- Insert/Delete, $c_{i/d} = 1$
- Replace, $c_r = 2,5$
- Match, $c_m = 0$

auf der Lösung mit an.

2. Erklärt den Unterschied zwischen End-free Alignment und Lokalem Alignment. Gebt jeweils ein Beispiel an wo jede dieser beiden Methoden verwendet wird. **3 P**
3. Gegeben sind die beiden String **ATGCCTGCT** und **ATTCT**. Gebt z.B. mit Hilfe des implementierten Programms das / ein beste(s) Alignment und den dazugehörigen Edit-Abstand für lineare Gap-Kosten von $c_{i/d} = c_r = 1$ an. Erstellt nun zum Vergleich die Matrix mit konstanten Gap-Kosten, $w = 2$ und gebt das / alle beste(n) Alignments an. Auch hier ist $c_r = 1$. **4 P**
4. Welche Alignment-Methode mit welchem Gap-Score-Modell ist für die folgenden Probleme die beste? Gebt außerdem an, welche Optimierungsmöglichkeiten bezüglich Speicherplatz und Laufzeit-Verhalten möglich sind und gebt die unteren Grenzen für beides an.
- (a) Wie ähnlich sind sich 2 Proteinsequenzen? **2 P**
(b) Finden einer mRNA bzw. cDNA-Sequenz im publizierten humanen Genom. **2 P**
(c) Alignment von 2 Proteinsequenzen aus der gleichen Proteinfamilie? **2 P**