

Prof. Dr. Knut Reinert  
Enrico Siragusa  
Sascha Meiers  
Christoph Hartmann

Institut für Informatik  
AG Algorithmische Bioinformatik

# Algorithmen und Datenstrukturen in der Bioinformatik

## Viertes Übungsblatt WS 11/12

Abgabe Montag, 14.11., 15:00 Uhr

Name: \_\_\_\_\_ Übungsgruppe: A  B  C

Matrikelnummer: \_\_\_\_\_

Niveau I

---

### Aufgabe 1: Online-Übung

Berechnen Sie mit Hilfe eines Needleman Wunsch Web Applets die paarweise Ähnlichkeit der menschlichen Haemoglobin- $\alpha$ - und Haemoglobin- $\beta$ -Untereinheit und der Haemoglobin- $\beta$ -Untereinheit der Maus. Die Sequenzen finden sich bei UniProt unter folgenden IDs:

HBA\_HUMAN HBB\_HUMAN HBA\_MOUSE

Erklären Sie kurz Ihre Ergebnisse.

---

### Aufgabe 2: Overlap alignment

In der Vorlesung wurden Algorithmen vorgestellt, die ein optimales *lokales* und *globales* Alignment zweier Sequenzen berechnen. Entwickeln Sie auf Basis dieser Algorithmen einen Algorithmus, der ein optimales *Overlap*-Alignment berechnet. (Ein Overlap-Alignment sucht die maximale Übereinstimmung des Endes einer Sequenz mit dem Anfang der anderen Sequenz, siehe Skript)

Berechnen Sie mit ihrem Algorithmus das optimale Overlap-Alignment der Sequenzen  $S_1 = \text{CTAC}$  und  $S_2 = \text{TCCG}$ . Ein Match trägt 3 zum Score des Alignments bei, ein Mismatch  $-3$  und ein InDel  $-2$ .


---

## Niveau II

---

### Aufgabe 3: Dynamische Programmierung II

Die *längste gemeinsame Subsequenz* zweier Sequenzen sei definiert als die längste Buchstabenfolge, die beiden Sequenzen gemein ist. Buchstaben der Subsequenz müssen in derselben Reihenfolge in beiden Sequenzen auftauchen, aber nicht notwendigerweise direkt aufeinanderfolgen.

Beispiel: Die längste gemeinsame Subsequenz von PAMPELMUSE und HASELNUSS ist AELUS.

- Geben sie einen Algorithmus in *Pseudocode* an, der die *Länge* der längsten gemeinsamen Subsequenz von zwei Sequenzen  $S_1, S_2$  bestimmt.
- Wie läßt sich nicht nur die Länge, sondern auch die Subsequenz ermitteln?
- Überlegen sie sich kurz, wie sich der Algorithmus von 2 auf  $k$  Sequenzen erweitern läßt. Wie ist die Laufzeit?