

## Algorithmen und Datenstrukturen in der Bioinformatik

### Sechstes Übungsblatt WS 09/10

Abgabe Montag 30.11.2009 12:00

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

Matrikelnummer: \_\_\_\_\_ Ich kann Aufgabe \_\_\_\_\_ nicht vorrechnen.

---

#### Exercise 14: Alignment in linear space

Give the pseudo-code of an algorithm that computes the optimal global alignment score of two sequences using only linear spaces. **Hint:** use what you know about the Needleman-Wunsch algorithm. Which values in the matrix are still needed? Which can be safely dropped?

---

In the following exercises, you will perform all steps of the FastA algorithm for two sequences  $S_1 = \text{LICLLI}$  and  $S_2 = \text{LICLIWLLWCLLI}$ .

#### Exercise 15: FastA – Hot-spots and diagonal runs

- Start by drawing the dot plot for these sequences.
  - Now identify all hot-spots of length  $k = 2$ , by
    - creating a hash table with positions for sequence  $S_1$  (draw the hash table like on the slides),
    - filling the list of diagonal runs by walking over sequence  $S_2$ .
  - Re-score the diagonal runs using as score 1 for matches and  $-1$  for mismatches. What changes?
  - Identify the highest-scoring diagonal run INIT1.
- 

#### Exercise 16: FastA – Chaining

- Construct the topological ordering from the diagonal runs. As scores, use 2 for all letters and  $-1$  for all other columns. Diagonal “shortcuts” are allowed.
- Apply the modifications to this graph to make it amenable for the DAG\_shortest\_paths algorithm. Which are these?
- Use the algorithm to find the best chaining of diagonal runs in the dot plot.