Advanced Algorithms in Bioinformatics (P4) Sequence and Structure Analysis

Freie Universität Berlin, Institut für Informatik David Weese, Sandro Andreotti Sommersemester 2011

> 7. Exercise sheet, 8. June 2011 Discussion: 15. June 2011

Exercise 1.

BWT

• For the text **tacaacaatacaagag\$** construct the BWT and the arrays *C* and *Occ*. Use them to search for the pattern **aca**.

Exercise 2.

BWT - compressing *L*

• Let R be the MTF encoding of L and Y the corresponding list of characters. Give an algorithm in pseudocode to decode R into L

Exercise 3.

BWT - compressing *pos*

Present an example that proves the following assumption stated in the script:
If we mark every η-th row in the matrix *M* the worst case time of a *pos* query is O(^{η-1}/_ηn)

Exercise 4.

Chaining

• Prove the lemmata used for the Manhattan distance and the sum-of-pairs distance in the chaining problem, as discussed in the lecture (Lemma 1 and Lemma 3).