



---

Silke Trissl

Email: [trissl@informatik.hu-berlin.de](mailto:trissl@informatik.hu-berlin.de)

---

## Lectures on Bioinformatics: Assignment 4

More information about the exercises can be found at  
[http://zope.informatik.hu-berlin.de/forschung/gebiete/wbi/teaching/archive/sose09/v1\\_biophysik](http://zope.informatik.hu-berlin.de/forschung/gebiete/wbi/teaching/archive/sose09/v1_biophysik)

### Exercise 1: Global alignment using dynamic programming.

- Write a program to implement the algorithm based on dynamic programming for the global alignment of the two DNA sequences. The program should read the DNA sequences from a file in fasta format. The output of the program should be similarity between both sequences. Use the following costs for the calculation: Insertion -1, Deletion -1, Replacement -1 and Match +1. Choice of the programming language is yours. Some exemplary DNA sequences can be downloaded from the web site.
- Modify your program in such a way that it also outputs all optimal alignments between the two sequences as text. An alignment should look as follows

```
ACCG--TGAA-TAGGCA
| | |   | | | | | |
ACCTTGTC AAGTAGGTA
```

- Use your program to calculate all optimal alignments between the following pairs of sequences. How many optimal alignments exist in each case?
  - ACCGTTGACCACACACAG
  - CACA
  - TTTTTTTTTTT
  - TTTTTTTTT
  - TTTTTTTTTTTTT
  - TTTTTTTT

## Exercise 2:

Derive a formula which calculates how many optimal global alignments are there between a string of length  $n$  and a string of length  $m$ , if both strings are defined over the same one-element alphabet?

## Exercise 3: Database search and alignment

Titin is an important and very long muscle protein. Amino acid sequences are known for human and *Drosophila* as well as multiple partial DNA sequences.

- i. Search both titin proteins in Uniprot database and describe their functional importance.
- ii. Download the sequences for human and *drosophila* using the following links and calculate their similarity with your program (without expenditure of the Alignments!). How long does it take?

### **Drosophila:**

[http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?-id+626991V6atU+\[embl-AccNumber:AJ271740\]+-e](http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?-id+626991V6atU+[embl-AccNumber:AJ271740]+-e)

### **Human:**

[http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?-id+626991V6atU+\[embl-AccNumber:AJ277892\]+-e](http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?-id+626991V6atU+[embl-AccNumber:AJ277892]+-e)

**Note: Results of all exercises and the code should be zipped into one file including a README file describing the commands for running your code. Your code must be able to perform all task asked above.**