

# Introduction to Bioinformatics - project 1

27.10.2022

## 1 Task

- Implement (in Python  $\geq 3.5$ ) Needleman-Wunsch algorithm (global sequence alignment)
- Download the sequences (in FASTA format) of homologous genes (differing by around 10%) from the NCBI database. Test the operation of the algorithms. Get results with two different scoring functions.
- Download the human and hamster insulin protein sequences and make comparisons as above.
- Write a short report (max three pages, but structured! - problem description, methods, results, discussion)
- Task specification:
  1. Task should be submitted either in Jupyter notebook (with all required files to be run) or as a Python script that can be run from command line (ex. `./nw.py -a fasta1.fa -b fasta2.fa -o output.txt`)
  2. Input: Two sequences in fasta format and output file.
  3. Output: Nicely formatted score matrix, optimal alignment and score. Optimal alignment should also be saved in given output text file.

## 2 Scoring and deadlines

- Deadline: 9.11.2022 23:59
- Exceeding the deadline has a chance to send your project for the next three days, but this will cost you 1.5 points from the maximum mark for each day of delay.
- Punktacja projektu (10p + 2p bonus points\*):
  - Needleman-Wunsch implementation - 4p
  - Homologous genes alignment - 2p
  - Insulin sequence alignment - 2p
  - Report - 2p (Report is obligatory. Projects without the report will not be graded.
  - \*Retrieving multiple optimal alignments - +2pkt
- The works will be checked with anti-plagiarism. The consequence of downloading is 0 points for the entire project (no possibility of improvement)
- The project is individual.
- Remember about clean code and leaving comments in your code.