

Choosing a project for your Code-Review

You can choose one of the options below or pick your own idea.

The aim of the code review is to learn how to write good and readable code and also to have fun implementing interesting tasks and ideas.

The deadline for the first working version of your code will be announced separately. The “first working” version means your teacher is able to execute your code without getting any errors. Your code may be not fully functional but it must work with no errors, so you are recommended not to wait until the deadline to send your code for review so you could elucidate any troubles while you still have time to solve them.

Remember: if a program runs perfectly on your computer it still can fail to run on another computer.

Steps to pass the project:

- 1) Create an empty private repository on github.com;
- 2) Make an initial commit to the main branch;
- 3) Add your teacher as a collaborator of the repository;
- 4) Create a “dev” branch from main and work in dev from now on;
- 5) Write the first working version of your project and push the code into dev;
- 5) Create a pull request from dev into master. In the description write instructions - what your code can do and how to execute it;
- 6) Your teacher has **one week** to review your code and send remarks;
- 7) You have **one week** to fix your mistakes and/or to improve functionality;
- 8) Repeat steps 5)-7) until your project is accepted.

Table of content

Sequence Analyzer	3
Alignment editor	4
3D structure analyzer	5
Sequence alignment	6
Motif search	7

Sequence Analyzer

Your script takes a file with a sequence (RNA/DNA/Protein) as input and outputs its features.

Basic functionality

- *Output the molecule type - RNA/DNA/Protein*
- *Output the nucleotide/amino acid statistics*
- *For DNA output GC content*
- *For RNA/DNA output reverse complement sequence*
- *For RNA/DNA output ORFs and translated peptides*

Extra

- *Parsing different formats (FASTA, etc.)*
- *GUI*

All paths can be absolute or relative. The program should check the ability to perform operations (the existence and non-existence of files and folders), and report errors if they occur.

Alignment editor

Your script reads a FASTA file with aligned sequences and allows the user to edit the alignment.

Basic functionality

- *Parse FASTA format*
- *Add insertions or deletions to a given column for a selected set of rows*
- *Store default match/mismatch/gap scores*
- *Calculate the alignment score*
- *Save the alignment in FASTA format*

Extra

- *Parsing/saving different alignment formats (Stockholm, CLUSTAL, etc.)*
- *GUI*
- *Allowing different scoring functions*
- *Affine gap scores*

All paths can be absolute or relative. The program should check the ability to perform operations (the existence and non-existence of files and folders), and report errors if they occur.

3D structure analyzer

Your script takes a coordinate file with a 3D structure (RNA/DNA/Protein) as input and outputs its features.

Basic functionality

- *Parse PDB format*
- *Output the number of DNA/RNA/Protein chains*
- *Output ligands*
- *Output nucleotide / amino acid statistics*
- *Output the number of missing (unresolved) residues*
- *Save RNA-only/DNA-only/Protein-only atoms in PDB format*

Extra

- *Parsing/saving mmCIF format*
- *GUI*
- *Find all chain breaks*
- *Find all atom-atom contacts below a specified distance threshold*

All paths can be absolute or relative. The program should check the ability to perform operations (the existence and non-existence of files and folders), and report errors if they occur.

Sequence alignment

Your script reads a FASTA file with two unaligned sequences and aligns them.

Basic functionality

- *Parse FASTA format*
- *Perform the Needleman–Wunsch algorithm for global alignment*
- *Allow different scoring functions*
- *Implement affine gaps mode*
- *Save the aligned sequences in FASTA format*

Extra

- *Saving in different alignment formats (Stockholm, CLUSTAL, etc.)*
- *GUI*
- *Implementation of the Smith-Waterman algorithm for local alignment*
- *Fitting/Overlap alignment modes*
- *Multiple alignment*

All paths can be absolute or relative. The program should check the ability to perform operations (the existence and non-existence of files and folders), and report errors if they occur.

Motif search

Your script takes a file with a sequence (RNA/DNA/Protein) and a motif (shorter sequence) as input and performs the motif search.

Basic functionality

- *Parse FASTA format with a sequence and a motif*
- *Output all the positions of the motif within the sequence*
- *Output all the positions of the motif or its reverse complement within the sequence*
- *Output all the positions of the motif or its reverse complement within the sequence allowing up to K mismatches*
- *Allow input of a nucleotide sequence and an amino acid motif to search within all possible translated peptides*

Extra

- *GUI*
- *Search for the input motif within a set of sequences*
- *Search for A shared motif within a set of sequences (no motif in input)*

All paths can be absolute or relative. The program should check the ability to perform operations (the existence and non-existence of files and folders), and report errors if they occur.