Homework 1: Implementing the sequence alignment algorithms

Objective: Implement the global sequence alignment algorithm (with banding) and local sequence alignment algorithm. For both programs please implement the affine gap penalty strategy.

You are NOT allowed to use existing packages or libraries. You can use any programming language; a 2-point bonus will be given the 3 correct submissions that run the fastest.

A: Input format:

The sequence is provided in the FASTA format. In the FASTA format, the first row is the tag of the sequence with a leading character '>'. The following rows are the actual sequence. For example, it may look like:

```
>dog protein X
EEEEE
KKKKK
AAAAA
FFF
```

It represents a dog protein sequence "EEEEEKKKKKAAAAAFFF". Notice the length of each row is variable. The program should accept two FASTA files for the alignment. Please use the BLOSUM62 (http://www.ncbi.nlm.nih.gov/Class/FieldGuide/BLOSUM62.txt) scoring matrix for match and mismatch; use -11 as the gap opening penalty and -1 as gap extension penalty.

B: Output format:

The output is the resulting alignment score and an alignment between the two input sequences. For amino acids (characters) whose matches have positive scores in BLOSUM62, use '|' to indicate a positive match; otherwise use '*'. For alignment of amino acid and gap, no symbol should be placed. For example:

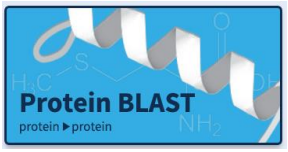
```
Score: 12345
EEEEEKKKKK
|||||
EEEEE-----

AAAAAFFF
*****|||
BBBBBFFF
```

Each row should contain 50-80 alignment columns such that the print out will not be messed up.

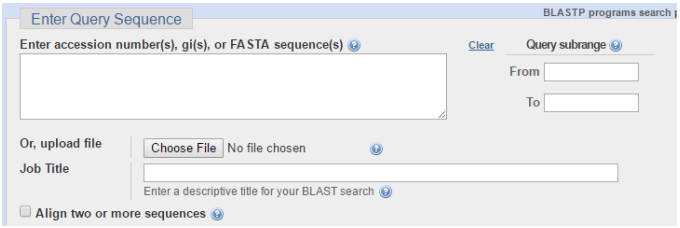
C: Test case

You can download two protein sequences (human hemoglobin alpha unit and mouse hemoglobin alpha unit) from our course website. Go to BLAST online server at http://blast.ncbi.nlm.nih.gov/Blast.cgi.

Click on “Protein BLAST”

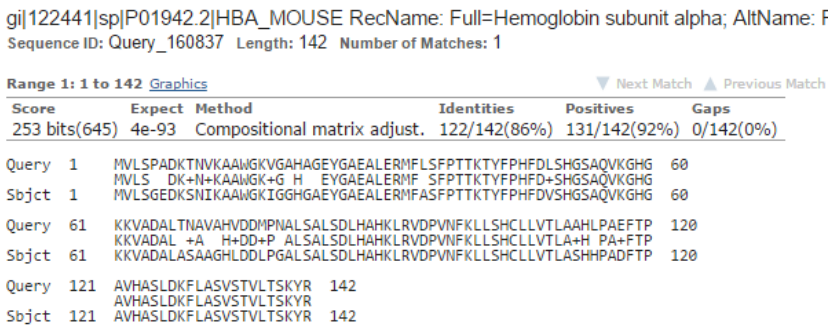


Check the box that says “Align two or more sequences” (at the bottom of the figure).



Copy the sequences into the boxes.

Scroll down and view the alignment generated by BLAST.



D: Submission

- Send your source code to my email (cczhong at KU dot edu) with title "EECS 730 HW1 submission" by **11:59 PM, Sep 30th**.
- Include a README file describing how to run your program (platform to compile, command line to run, *etc*).