

HW 3

Write a Perl script that takes two filenames as input. The first file contains two unaligned protein sequences in FASTA format and the second contains the same two proteins aligned in FASTA format. Your program should compute alignments (with BLOSUM 62 scoring matrix) using gap penalties -1, -3, -5, -7, -9, and -11 and output the gap penalty which produces the highest accuracy.

Test your program with the following PREFAB datasets and report the optimal gap penalties found.

1. 1f2dA_1oasA
2. 1cl2A_1qgnH
3. 1gca_1tlfA
4. 1efuB_1tfe
5. 1a04A_1ibjA

Submit your program as FirstnameLastnameTrain.pl by 1pm Wednesday March 21st via email to usman@cs.njit.edu. Include in your email the optimal gap penalties for each of the five datasets.

HINT:

Your program would have the following format:

```
@gaps=(-1, -3, -5, -7, -9, -11)

for(i=0;i<scalar(@gaps);i++){
    $g = $gaps[$i];
    system "perl computealignment.pl $unaligned $g";
    $acc[$i] = `perl score.pl $true alignment.txt`;
}
```

output gap penalty with highest accuracy by sorting \$acc.