

Part 3. Local sequence alignment with affine gap scores

Due: Tuesday October 30th, 2012

Overall weight: 30% of total project score

Specification: Extend the program you wrote in Part 2 to achieve the following:

- Perform local alignment (find best matching substring)
- Accept alignment scores formatted as a BLOSUM matrix (detail below)
- Gap penalties are "affine" - the combination of a gap opening penalty (paid once per group of gaps) plus a gap extension penalty (proportional to number of gaps within a group).

Output format: The output of your program must follow the format shown below - note the coordinates within the two strings of the aligned region, and that the '|' characters are replaced by the actual letter that matches between the two strings. Just as before, if the strings are too long to be printed on one line (assume one line is 80 characters) you will need to wrap the alignment around on multiple lines.

Score = 188, Identities = 49/152 (32%), Gaps = 6/152 (3%)

```
seq1  23  LPKTRTKALLTALTAAAAAALADVEFRHAL---DDSALDLSPIKGEEITDAVKSFR  79
      P      A      A  AL  FRH  D      S  G  T  AV  F
seq2   3   MPFNRSIAISATLAVGLLAPVVALGQEVFRHTVTGEDLKIMETSQPSGRD-TEAVRNFL  61
```

A sample output file is provided at: <http://www.cbcb.umd.edu/confcour/Fall2012/Output.txt>

Interface: follow the same rules as for Part 2, except that you must now accept two new values - gap opening and gap extension penalties - and a new file - the BLOSUM matrix specifying the substitution/match scores.

BLOSUM matrices: A sample matrix is provided at <http://www.cbcb.umd.edu/confcour/Fall2012/BLOSUM80.txt>

Note: most of the Bio* libraries contain utilities for reading in BLOSUM matrices.

Additional details: Any questions about this assignment should be sent to both myself and the TA. You can assume that the two FASTA files contain exactly one sequence (or if they contain more than one just use the first sequence in each file).

You can now use any of the Bio* libraries to read the FASTA files (but not to perform the alignments).