MAC5710 – Estruturas de Dados e sua Manipulação
IBI5038 – Introdução a Estruturas de Dados
IME – Primeiro Semestre de 2017

Quarto Exercício-Programa (EP4)
Professor: André Fujita

Data de entrega: até 23:55 do dia 25 de junho de 2017.

Alinhamento de sequências

Neste exercício-programa, sua tarefa consiste em, dadas duas sequências de aminoácidos, identificar todos o(s) melhor(es) alinhamentos.

A entrada consiste num arquivo contendo duas sequências de aminoácidos no formato FASTA. O FASTA consiste em diversas linhas, sendo que cada uma não ultrapasse 80 caracteres. A primeira linha consiste num símbolo de maior ">" seguido de um breve comentário sobre a sequência. As demais linhas são os nucleotídeos.

Exemplo de arquivo com duas sequências de aminoácidos em formato FASTA:

```
> Exemplo 1
ARNDARNDNDARNDARNDNDARNDARNDNDARNDARNDNDARNDARNDNDARNDARNDNDARNDARNDND
ARNDARNDNDARNDARNDNDARNDARNDNDARNDARNDNDARNDARNDNDARNDARNDNDARNDARNDND
ARNDARNDNDARNDARNDNDARNDARNDNDARNDARNDNDARNDARNDNDARNDARNDNDARNDARNDND
ARNDARNDNDARNDARNDNDARNDARNDNDARNDARNDNDARNDARNDNDARNDARNDNDARNDARNDND

> Exemplo 2
ARNDARNDNDARNDARNDNDARNDARNDNDARNDARNDNDARNDARNDNDARNDARNDNDARNDARNDND
ARNDARNDNDARNDARNDNDARNDARNDNDARNDARNDNDARNDARNDNDARNDARNDNDARNDARNDND
ARNDARNDNDARNDARNDNDARNDARNDNDARNDARNDNDARNDARNDNDARNDARNDNDARNDARNDND
ARNDARNDNDARNDARNDNDARNDARNDNDARNDARNDNDARNDARNDNDARNDARNDNDARNDARNDND
```

Para obter o(s) melhor(es) alinhamentos, seu programa deve usar o algoritmo Needleman-Wunsch (como visto em aula).

No algoritmo Needleman-Wunsch é necessário definir uma matriz de pesos. Utilize a matriz BLOSUM62 (https://en.wikipedia.org/wiki/BLOSUM#/media/File:BLOSUM62.gif).

Também é necessário definir a penalidade para o "gap". Permita que o usuário defina este valor.

A saída do programa consiste num arquivo que deve conter o "score" e TODOS os melhores alinhamentos. O "gap" deve ser representado pelo caractere "-" (hífen). Onde houver "match" entre as sequências, coloque uma barra vertical "|". Onde houver um "mismatch", coloque um asterisco. Por exemplo:

```
        *
-RNRNNCC
||||  | |
NRNR-NAC
```

Algumas fontes bibliográficas interessantes são:

[1] Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*. **22**: 4673 – 4680, 1994.

[2] Setubal J & Meidanis J. Introduction to Computational Molecular Biology. Boston. PWS Publishing Company. 296 páginas.

Você deve entregar o código fonte (escrito na linguagem C) e um *makefile*.