CCM0128 – Computação II
Bacharelado em Ciências Moleculares
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).

Observações:

- O cabeçalho do EP4 deve ser:

```
/* ************************************************************************ */
/* Nome: [coloque aqui o seu nome]                                          */
/* Numero USP: [coloque aqui seu numero USP]                                */
/*                                                                          */
/* Nome: [coloque aqui seu nome]                                            */
/* Numero USP: [coloque aqui seu numero USP]                                */
/*                                                                          */
/* Exercício-programa 4                                                     */
/* ************************************************************************ */
```

- EPs que não compilam receberão nota ZERO. O comando a ser usado na compilação do monitor será:
  gcc –Wall –ansi –pedantic –O2 –o ep4 ep4.c
  Certifique-se que seu EP compila no sistema operacional Linux com o comando acima. Mensagens de "warnings" serão penalizados na nota.
- Não serão aceitos EPs atrasados. Será considerado como EP não entregue.
- Você deve entregar somente o arquivo contendo o código fonte: *.c
  Outros arquivos que não sejam .c entregues "por engano" receberão nota ZERO.
- Seu programa NÃO precisa checar consistência de dados.
- O EP deve ser feito em até duas pessoas. Você pode conversar e discutir a solução com seus colegas, mas em hipótese alguma você deve mostrar e/ou ver o código dos outros. Qualquer problema com o código do EP deve ser tratado com o monitor da disciplina.
- EPs copiados parcialmente ou totalmente da internet ou de qualquer outra fonte será considerado plágio. EPs que tentem "mascarar" a cópia também serão considerados plágio.
- EPs com plágio receberão nota ZERO, o aluno será REPROVADO e seu nome será encaminhado a Comissão de Graduação.