

Universidade Federal do Rio Grande do Sul  
Instituto de Informática  
Departamento de Informática Teórica  
INF05018 – Biologia Computacional (2017/2)  
LISTA DE EXERCÍCIOS 02

---

Instruções:

- A resolução do exercício deve ser feita **individualmente**. Cópias evidentes entre trabalhos não serão aceitas.
  - Para cada uma das tarefas, deve-se entregar o código-fonte em linguagem Python. **Utilize somente a Python Standard Library** ([Python 2](#) / [Python 3](#)). O nome do arquivo deve identificar a tarefa (ex.: "e2.py").
  - Além do código-fonte deve-se entregar um único arquivo PDF com o nome "e2.pdf", apresentando o pseudocódigo do algoritmo desenvolvido e os resultados encontrados. As respostas devem **obrigatoriamente** ser apresentadas nos slides.
  - Os arquivos devem ser enviados em um arquivo compactado com nome no formato "IDA\_e2", sendo IDA as iniciais do nome do aluno.
  - Arquivos corrompidos serão desconsiderados.
  - A presença na aula de apresentação contabiliza 10% da nota.
  - Data de entrega: 03/10/2017 (terça-feira) até as 12h via Moodle, **exclusivamente**.
  - **Não serão aceitos trabalhos atrasados.**
- 

A hemoglobina dos vertebrados é responsável pelo transporte de oxigênio nas hemácias contidas no sangue. Em *Homo sapiens*, a hemoglobina A (HbA) constitui 97% da hemoglobina total em adultos saudáveis e sua estrutura é composta por quatro cadeias polipeptídicas: duas  $\alpha$ -globinas (em *H. sapiens*, codificadas pelos genes *HBA1* e *HBA2*) e duas  $\beta$ -globinas (*HBB*). Abaixo estão listadas as sequências aminoacídicas da  $\alpha$ -globina de sete espécies diferentes em formato FASTA. Nota-se que, embora existam variações de uma sequência quando comparada a outra, todas exercem a mesma função independente da espécie. São, portanto, ortólogas e possuem origem evolutiva comum.

```
>2DN3:A|Human (Homo sapiens) [TaxId: 9606]
MVLSPADKTNVKAAGKVGAGHAGEYGAEALERMFSLFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNA
VAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSK
YR
>2D5X:A|Horse (Equus caballus) [TaxId: 9796]
MVLSAADKTNVKAAWSKVGGHAGEYGAEALERMFSLGFPTTKTYFPHFDLSHGSAQVKAHGKKVGDALTTLA
VGHLDDLPGALSNSDLHAHKLRVDPVNFKLLSHCLLSTLAVHLPNDFTPAVHASLDKFLSSVSTVLTSK
YR
>1HDS:A|Deer (Odocoileus virginianus) [TaxId: 9874]
VLSAABKSBVKAAGKVGGNAAAPYGAZALZRMFLSFPTTKTYFPHFBLSHGSAZVKAHGZKVABALTKAV
GHLBBLPGTSLBLSBLHAHKLRVBPVBFKLLSHSLVTLATHLPBBFTPAVHASLBKFLABVSTVLTSKY
R
>2QSS:A|Cow (Bos taurus) [TaxId: 9913]
MVLSAADKGNVKAAGKVGGHAAEYGAEALERMFSLFPTTKTYFPHFDLSHGSAQVKGHGAKVAAALTKA
VEHLDDLPGALSELSDLHAHKLRVDPVNFKLLSHSLVTLASHLPSPDFTPAVHASLDKFLANVSTVLTSK
YR
>1QPW:A|Pig (Sus scrofa) [TaxId: 9823]
```

```

VLSAADKANVKAAGKVGQAGAHGAEALERMF LGFPTTKTYFPHFNLSHGSDQVKAHGQKVADALTKAV
GHLDDLPGALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHHPDDFNPSVHASLDKFLANVSTVLT SKY
R
>1FHJ:A|Maned wolf (Chrysocyon brachyurus) [TaxId: 68728]
VLSPADKTNIKSTWDKIGGHAGDYGGEALDRTFQSFPTTKTYFPHFDLSPGSAQVKAHGKKVADALT TAV
AHLDDLPGALSALSDLHAYKLRVDPVNFKLLSHCLLVTLACHHPTEFTPAVHASLDKFF TAVSTVLT SKY
R
>1HBR:A|Chicken (Gallus gallus) [TaxId: 9031]
MLTAEDKKLIQQAWEKAASHQEEFGAEALTRMFTTYPQTKTYFPHFDLSPGSDQVRGHGKKVLGALGNAV
KNVDNLSQAMAELSNLHAYNLRVDPVNFKLLSQCIQVVLAVHMGKDYTPEVHAAFDKFLSAVSAVLA EKY
R
>1OUT:A|Trout (Oncorhynchus mykiss) [TaxId: 8022]
SLTAKDKSVVKAFWGKISGKADVVGAEALGRDKMLTAYPQTKTYFSHWADLSPGSGPVKKHGGIIMGAIG
KAVGLMDDLVGMSALSDLHAFKLRVDPGNFKILSHNILLVTLAIHFPSDFTPEVHIAVDKFLAAVSAALA
DKYR

```

Implemente um conjunto de rotinas computacionais para realizar as tarefas que seguem:

1. Desenvolva um algoritmo que leia duas sequências  $n$  e  $m$ ;
2. Implemente o algoritmo de Needleman-Wunsch para alinhar estas sequências;
3. O programa deve imprimir na tela a tabela final de alinhamento, as duas cadeias alinhadas, o *score* e a identidade do alinhamento;
4. Regras:
  - Valor para *gap*: -4
  - Valor para *match*: +5
  - Valor para *mismatch*: -3
  - $$S_{i,j} = \max \begin{cases} S_{i-1,j-1} + s(a_i, b_j); & \text{match/mismatch} \\ S_{i,j-1} + w; & \text{gap seq1} \\ S_{i-1,j} + w; & \text{gap seq2} \end{cases}$$
5. Com base na identidade e no *score* do alinhamento, informe qual das espécies acima apresenta a maior semelhança em termos de sequência com *H. sapiens*.

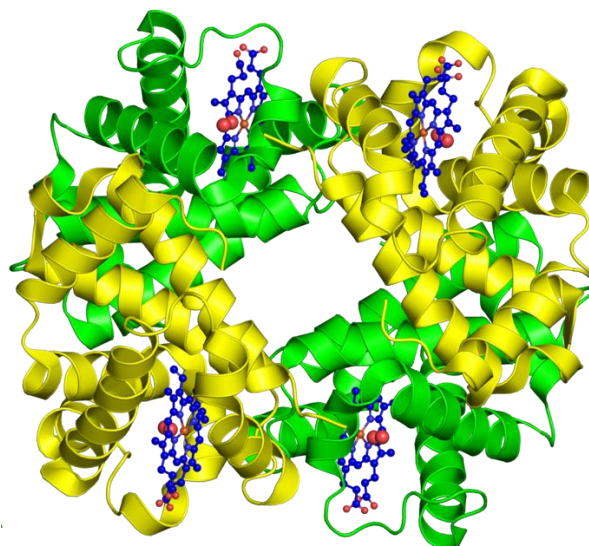
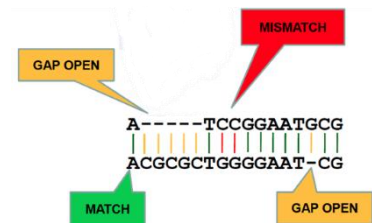


Figura 1: Hemoglobina humana (*H. sapiens*) colorida por subunidades distintas em visão frontal. PDB ID [1GZX](#).