

Universidade Federal do Rio Grande do Sul
Instituto de Informática
Departamento de Informática Teórica
INF05018 – Biologia Computacional (2017/2)
LISTA DE EXERCÍCIOS 03

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.: "e3.py").
 - Além do código-fonte deve-se entregar um único arquivo PDF com o nome "e3.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_e3", 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: 28/09/2017 (terça-feira) até as 12h via Moodle, **exclusivamente**.
 - **Não serão aceitos trabalhos atrasados.**
-

Ureases são enzimas níquel-dependentes que catalisam a reação de hidrólise da uréia em amônia e carbamato que, por sua vez, sofre hidrólise espontânea, formando ácido carbâmico e uma segunda molécula de amônia. Em plantas, essas enzimas tomam parte no metabolismo de compostos nitrogenados, principalmente disponibilizando a uréia como fonte de nitrogênio, assumindo papel vital no seu desenvolvimento. Adicionalmente, a urease é sintetizada em fungos, algas, bactérias e invertebrados, além de existir no solo. Dada sua ampla distribuição entre os reinos da vida, as ureases assumem um grande papel no metabolismo de nitrogênio na natureza. Considere as sequências abaixo:

```
>sp|P07374|UREA_CANEN Urease OS=Canavalia ensiformis PE=1 SV=3
MKLSPREVEKLGHLNAGYLAQKRLARGVRLNYTEAVALIASQIMEYARDGEKTVAQLMCL
GQHLLGRRQVLPVAVPHLLNAVQVEATFPDGTKLVTVHDPISRENGELQEALFGSLLPVPS
LDKFAETKEDNRIPGEILCEDECLTLNIGRKAVILKVTSGDRPIQVGSHYHFIEVNPYL
TFDRRKAYGMRLNIAAGTAVRFEPGDCKSVTLVSIENKVIKRGNAIADGPVNETNLEAA
MHAVERSKGFGHEEEKDASEGFTKEDPNC PFNTFIHRKEYANKYGPTTGDKIRLGD TNLLA
EIEKDYALYGDECVFGGK VIRDGMGQSCGHPPAISLDTVITNAVIIDYTGI IKADIGIK
DGLIASIGKAGNPDIMNGVFSNMI IGANTEVIAGEGLIVTAGAIDCHVHYICPQLVYEAI
SSGITTLVGGGTGPAAGTRATCTPSPQTQMRMLQSTDDLPLNFGFTGKGSSSKPDELHE
IIKAGAMGLKLHEDWGSTPAAIDNCLTIAEHHDIIQINIHTDTLNEAGFVEHSIAAFKGR
IHTYHSEGAGGGHAPDIIKVCGIKKNVLPSSSTNPTRPLTSNTIDEHLDMLMVCHHLDREIP
EDLAFASHSRIRKKTIAAEDVLNDIGAISIISDSQAMGRVGEVISRTWQTADKMKAGTGP
LKCDSSDNDNFRIIRYIAKYTINPAIANGFSQYVGSVEVGKLADLVMWKPSFFGTPKPEMV
IKGGMVAWADIGDPNASIPTPEPVKMRPMYGTGKAGGALSIAFVSKAALDQVRNVLYGL
NKRVEAVSNVRKLT KLDMKLN DALPEITVD PESYTVKADGKLLCVSEATTVPLSRNYFLF
```

```
>sp|P18314|URE1_KLEAE Urease subunit alpha OS=Klebsiella aerogenes
GN=ureC PE=1 SV=1
MSNISRQAYADMFGPTVGDKVRLADTELWIEVEDDLTTYGEEVKFGGGKVIRDGMGQGM
```

```
LAADCVDLVLTNALIVDHWGIVKADIGVKDGRIFAIGKAGNPDIQPNVTIPIGAATEVIA
AEGKIVTAGGIDTHIHWICPQQAEELVSGVTMVGGGTGPAAGTHATTCTPGPWYISRM
LQAADSLFPVNIIGLLGKGNVSQPDALREQVAAGVIGLKIHEDWGATPAAIDCALTVADEM
IQVALHSDTLNESGFVEDTLAAIGGRTIHTFHTEGAGGGHAPDII TACAHPNILPSSSTNP
TLPYTLNTIDEHLDMMLVCHHLDPDIAEDVAFaesRIRRETIAAEDVLHDLGAFSLTSSD
SQAMGRVGEVILRTWQVAHRMKVQRGALAEETGDNDNFRVKRYIAKYTINPALTHGIAHE
VGSIEVGKLADLVVWSPAFFGVKPATVIKGGMIAIAPMGDINASIPTPQPVHYRPMFGAL
GSARHHCRLTFLSQAAAANGVAERLNLRSIAIAVVKGCRTVQKADMVHNSLQPNITVDAQT
YEVVRVDGELITSEPADVLPMAQRYFLF
```

```
>sp|P18315|URE2_KLEAE Urease subunit beta OS=Klebsiella aerogenes
GN=ureB PE=1 SV=1
MIPGEYHVKPGQIALNTGRATCRVVVENHGDRIQVGSYHFAEVPNALKFDRQQAAGYR
LNIPAGTAVRFEFGQKREVELVAFAGHRAVFGFRGEVMGPLEVNDE
```

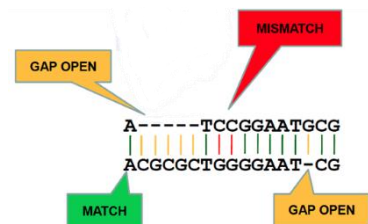
```
>sp|P18316|URE3_KLEAE Urease subunit gamma OS=Klebsiella aerogenes
GN=ureA PE=1 SV=1
MELTPREKDKLLLF TAALVAERRLARGLKLNYPESVALISAFIMEGARDGKSVASLMEEG
RHVLTREQVMGVPPEMIPDIQVEATFPDGSKLVTVHNPII
```

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 Smith-Waterman para alinhar estas sequências;
3. O programa deve imprimir na tela a tabela final de alinhamento, as duas cadeias alinhadas e a identidade do alinhamento;
4. Regras:

- Valor para *gap*: -2
- Valor para *match*: +1
- Valor para *mismatch*: -1

$$S_{i,j} = \max \begin{cases} S_{i-1,j-1} + s(a_i, b_j); & \text{match/mismatch} \\ S_{i,j-1} + w; & \text{gap seq1} \\ S_{i-1,j} + w; & \text{gap seq2} \end{cases}$$



Canavalia ensiformis

α

840 aa

Figura 1: Esquerda: Feijão-de-porco ou *jack bean* (*Canavalia ensiformis*). Direita: Urease eucariótica típica. PDB ID [3LA4](#).