



# Universidade Federal do Rio de Janeiro

## Centro de Ciências da Saúde

### Instituto de Biofísica Carlos Chagas Filho

**Disciplina:** CFB017 - Programação para Biociências

**Professor:** Dr. Vitor Lima Coelho

## TAC 3 - Biopython

1 - Escreva um programa Python que peça ao usuário uma sequência de DNA e imprima a sequência de mRNA e a sequência de proteína correspondentes.

Código:

```
# importação de biblioteca:
from Bio.Seq import Seq
# recebendo a sequência de DNA como input
DNA = Seq(str(input('Insira sua sequência de DNA:')))
# a função .transcribe realiza a transcrição
mRNA = DNA.transcribe()
# a função .translate() realiza a tradução A PARTIR de um RNA
ptn = mRNA.translate()
# imprimir cada sequência abaixo de um identificador
print('mRNA: ', mRNA, 'ptn: ', ptn, sep='\n')
```





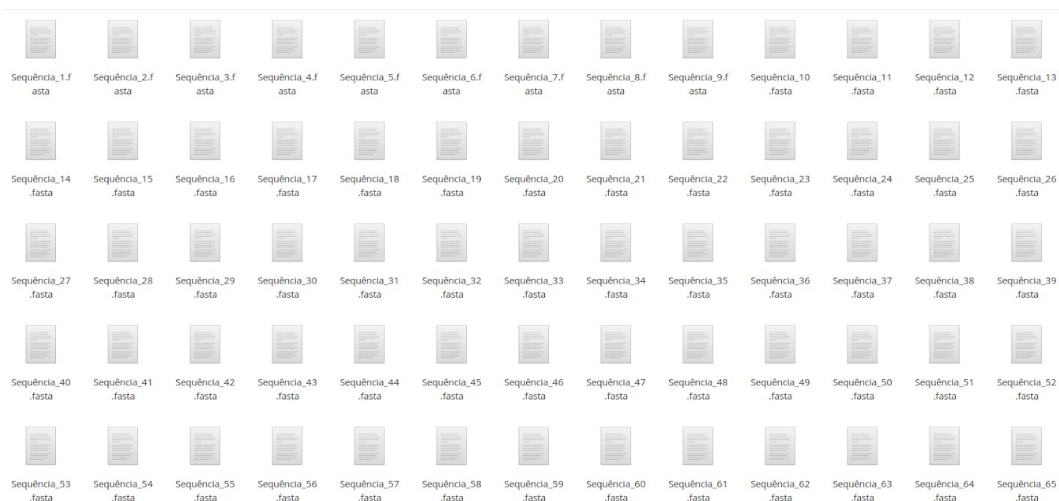
2 - Escreva um programa Python que pegue o arquivo sequencias.fasta e escreva N arquivos FASTA contendo em cada arquivo apenas uma sequência de sequencias.fasta, chamados de sequência\_i.fasta (onde i varia de 1 a N).

Código:

```
# importação de bibliotecas:
from Bio.Seq import Seq
from Bio import SeqIO
# para escrever um arquivo fasta
from Bio.SeqRecord import SeqRecord
# abrindo, lendo e identificando os itens do arquivo fasta:
multifasta = SeqIO.parse(open("/home/carol/Documents/CFB017/sequencias.fasta", 'r'), "fasta")
# contador para os arquivos:
num = 0
# leitura das linhas:
for line in multifasta:
    num = num + 1
# o SeqRecord precisa ser gerado nessa ordem (id primeiro e seq depois dá bug)
record = SeqRecord(line.seq, line.id)
# escrevendo o fasta inserindo a variável num no nome do arquivo
SeqIO.write(record, 'Sequência_%d.fasta' % num, "fasta")
```

Output:

+	×	Sequência_1.fasta	TAC3_questão2.py
1	>	Q99697 <unknown description>	
2		METNCRKLVSAQVQLGVQPAAVECLFSKDSEIKKVEFTDSPESRKEAASSKFFPRQHPGA	
3		NEKDKSQGKNEVDGAEDPSKKRQRRQTHFTSQQLQLEATFORNRYPDMSREEIAV	
4		WTNLTEARVRVWFKNRRRAKWRKRERNQQAELCKNGFGPQFNGLMQPYDDMPGYSYNNWA	
5		AKGLTSASLSTKSFPFFNSMNVNPLSSQSMFSPNSISSMSMSSMVPSAVTGVPGSSLN	
6		SLNNLNNLSSPSLNSAVPTPACPYAPPTPPYVYRDTCNSSLASRLKAKQHSSFGYASVQ	
7		NPASNLSACQYAVDRPV	





3 - Considerando o arquivo FASTA sequenciaDesconhecida.fasta, faça um código em Python que:

- Lê do usuário o caminho para o arquivo FASTA sequenciaDesconhecida.fasta;
- Lê do usuário o caminho para o arquivo multi-fasta de proteínas de *Trypanosoma cruzi* cepa CL Brener Esmeraldo-like TriTrypDB-47\_TcruziCLBrenerEsmeraldo-like\_AnnotatedProteins.fasta;
- Executa BLAST da sequência desconhecida de DNA contra as sequências de aminoácidos usando formato de saída nº 6 e e-valor igual à 0,05
- Imprima **somente o hit com o maior score**.

Código:

```
# importação de bibliotecas:
import pandas
import pandas as pd
from Bio.Seq import Seq
from Bio import SeqIO
from Bio.Blast.Applications import NcbiblastxCommandline
# blastx path:
blastx = "/home/carol/anaconda3/bin/blastx"
# Query: "/home/carol/Documents/CFB017/sequenciaDesconhecida.fasta"
seq_hipot = input(str('Sequencia desconhecida:'))
# Subject: "/home/carol/Documents/CFB017/TriTrypDB-47_Tcruzi_AnnotatedProteins.fasta"
Tcruzi_data = input(str('Tcruzi Data: '))
# Resultado:
Blast_TAC3 = r"/home/carol/Documents/CFB017/Blast_TAC3.txt"
# blast x
meu_blast = NcbiblastxCommandline(cmd = blastx ,query = seq_hipot,
subject = Tcruzi_data, evalue = 0.05, outfmt = 6, out = Blast_TAC3)
# redirecionando resultados:
stdout, stdeer = meu_blast()
# adicionando cabeçalho à tabela de resultados do blastx:
result = pd.read_csv("/home/carol/Documents/CFB017/Blast_TAC3.txt",
sep='\t',
names=["qseqid", "sseqid", "pident", "length", "mismatch", "gapopen", "qstart", "qend", "sstart", "send", "evalue", "bitscore"])
# ordenando resultado BLASTX por ordem crescente de bitscore:
max_score = result.sort_values('bitscore')
# printando a última linha (maior score):
print(max_score.iloc[[-1]])
```



Output:

```
Applications dom, out 18 12:12 PM
x Spyder (Python 3.7)
Arquivo Editar Pesquisar Código Executar Depurar Consoles Projetos Ferramentas Ver Ajuda
/home/carol/Documents/CFB017/TAC 3/TAC3_questão3.py
Console 1/A
In [24]: runfile('/home/carol/Documents/CFB017/TAC 3/TAC3_questão3.py', wdir='/home/carol/Documents/CFB017/TAC 3')
0 sequenciaDesconhecida TcCLB.503447.20:mRNA-p1 ... 0.0 1429.0
[1 rows x 12 columns]
```