

Thema 1, script 5

Verplichte onderdelen:

- Schrijf een programma dat, gegeven de naam van een mRNA FASTA-bestand, de frequentie van alle mogelijke codons in de betreffende sequentie bepaalt. Vorm een python dictionary die per triplet bijhoudt hoe vaak dit voorkomt.
- Reken de uitkomsten om naar percentages hoe vaak een triplet voor zijn aminozuur codeert (vergeleken met andere tripletten voor datzelfde aminozuur). Laat het programma voor elk aminozuur bepalen welk triplet het meeste voor dat aminozuur codeert, en geef deze uitkomsten weer op het scherm.
- Download van NCBI een flink aantal willekeurige menselijke mRNA FASTA-bestanden. Pas je programma aan zodat het een verzameling FASTA-bestanden kan inlezen en op grond van al deze bestanden tezamen één overzicht genereert met voor elk aminozuur het meest voorkomende daarvoor coderende triplet.
- Breidt het programma uit zodat het vervolgens een FASTA-bestand met een aminozuur-sequentie inleest, de sequentie “terug” vertaalt naar nucleotiden op grond van de meest waarschijnlijke tripletten, en deze nucleotide-sequentie op het scherm toont.
- Wijzig je programma zodat het een overeenkomstig mRNA FASTA-bestand naar de schijf wegschrijft.

Bonus-onderdelen:

- Laat het programma zelf een geschikte naam voor het uitvoerbestand vormen. Bv., een bestand genaamd *protein.fasta* zou kunnen worden omgezet in een nucleotide-sequentie die wordt opgeslagen in een bestand genaamd *protein_reverse-mrna.fasta*.
- Pas het programma aan zodat de gebruiker een willekeurig aantal FASTA-files met aminozuur-sequenties kan aangeven die allemaal onder een eigen aangepaste naam naar een bijbehorend reverse-mRNA FASTA bestand dienen te worden omgezet.
- Breid je programma uit zodat het een waarschuwing weergeeft op het scherm als de gebruiker een FASTA-bestand heeft gekozen dat alleen de aminozuren Alanine, Threonine, Glycine, en/of Cysteine lijkt te bevatten. (Hint: deze aminozuren worden immers gerepresenteerd door dezelfde letters als de vier nucleotiden.)
- Pas je programma aan dat de gebruiker één lijst met FASTA-bestanden kan invoeren. Je programma dient telkens automatisch te bepalen of het om nucleotide- of aminozuur-sequenties gaat. De bestanden met nucleotide-sequenties dienen te worden gebruikt om de meest voorkomende tripletten te bepalen; de bestanden met aminozuur-sequenties dienen te worden omgezet in reverse-mRNA FASTA-bestanden.

Lever je programma in via BlackBoard in de vorm van een python-bestand genaamd **script5.py**.

Let hierbij op dat je code:

- ✓ op tijd wordt ingeleverd;
- ✓ zonder foutmeldingen uitvoerbaar is;
- ✓ de (verplichte) onderdelen van de opdracht correct en volledig verricht;
- ✓ liefst zo efficiënt mogelijk geïmplementeerd is;
- ✓ en, een netjes leesbare programmeerstijl gebruikt.