

Praktische Opdracht Informatica Course 2a

Thematoets Python II - Course 2a 2018/2019

Bio-informatica 1: woensdag 23 januari t/m woensdag 30 januari 2019

Casus

1 ATP/GTP-binding site motief A (P-loop)

Het binden van ATP is een essentieel proces binnen de cel. ATP wordt door allerlei eiwitten gebruikt als katalysator voor het in gang zetten van celprocessen. Door middel van sequentie en crystallografische analyse is duidelijk geworden dat een groot deel van de eiwitten die binden met ATP (of GTP) een in meer of mindere mate gelijk motief delen. Het best geconserveerde motief is het glycine rijke gebied, welke een flexibele loop tussen een α -helix en een β -strand vormt. Deze loop gaat interacties aan met een van de fosfaatgroepen van een nucleotide. Het motief wordt over het algemeen het 'A' consensus patroon of de 'P-loop' genoemd. Zoals gezegd zijn er veel eiwitten die dit patroon bevatten, verdeeld over verschillende organismen. Het consensus patroon van de 'P-loop' is:

[AG]-x(4)-G-K-[ST]

2 Taak

Je loopt stage op een onderzoeksafdeling waar men zich bezighoudt met ATP binding sites, en dus ook met deze specifieke ATP binding site, de 'P-loop'. Jouw stagebegeleider geeft aan dat hij graag wil weten welke soorten hetzelfde patroon als de mens hebben. Hij is namelijk op zoek naar een modeldier voor een experiment. Om dit uit te zoeken heeft hij een fasta bestand gevonden waarin alle eiwitten uit UniProt die aan het motief voldoen. Deze voldoen echter aan het motief, niet aan de specifieke vorm waarin dit motief in de mens voorkomt. Een voorbeeld van een eiwit dat in het bestand voorkomt is:

```
>sp|Q8N139|ABCA6HUMAN (1617 aa )|ATP-binding . . . member 6 . [Homo sapiens  
(Human)]  
MNMKQKSVYQQTKALLCKNFLKKWRMKRESLLEWGLSILLGLCIALFSSSMRNVQ  
FPGMAPQNLGRVDKFNSSSLMVVYTPISNLTQQIMNKTALAPLLKGTSVIGAPNK  
THMDEILLENLPYAMGIIFNFTFSYKLIFQGYNSPLWKEDFSAHCWDGYGEFSC  
TLTKYWNRGFVALQTAINTAIEITTNHPVMEELMSVTAITMKTLFPITKNLLHN  
EMFILFLLHFSPLVYF . . .  
514 - 521: GhsgaGKS  
1320 - 1327: GpngaGKS
```

Jouw opdracht is om het motief van de mens te vinden in ieder eiwit in het bestand, om te zoeken in welke organismen deze sequentie allemaal voorkomt. Als resultaat wil jouw stagebegeleider dat je een GUI met een staafdiagram maakt met het aantal counts per organisme. Hoe vaak komt motief voor in een bepaald organisme? Figure 1 geeft een voorbeeld van zo'n grafiek. Let op: dit is geen echte data. In deze GUI is het de bedoeling dat je aan kan geven elke organismen je allemaal wilt zien, als een checkbox.

Er zijn meerdere verschillende P-loop motieven in de mens aanwezig. Je mag kiezen of je degene pakt die het meest voorkomt of je mag ze allemaal kiezen. Zolang je documenteert waarom en wat je gekozen hebt.

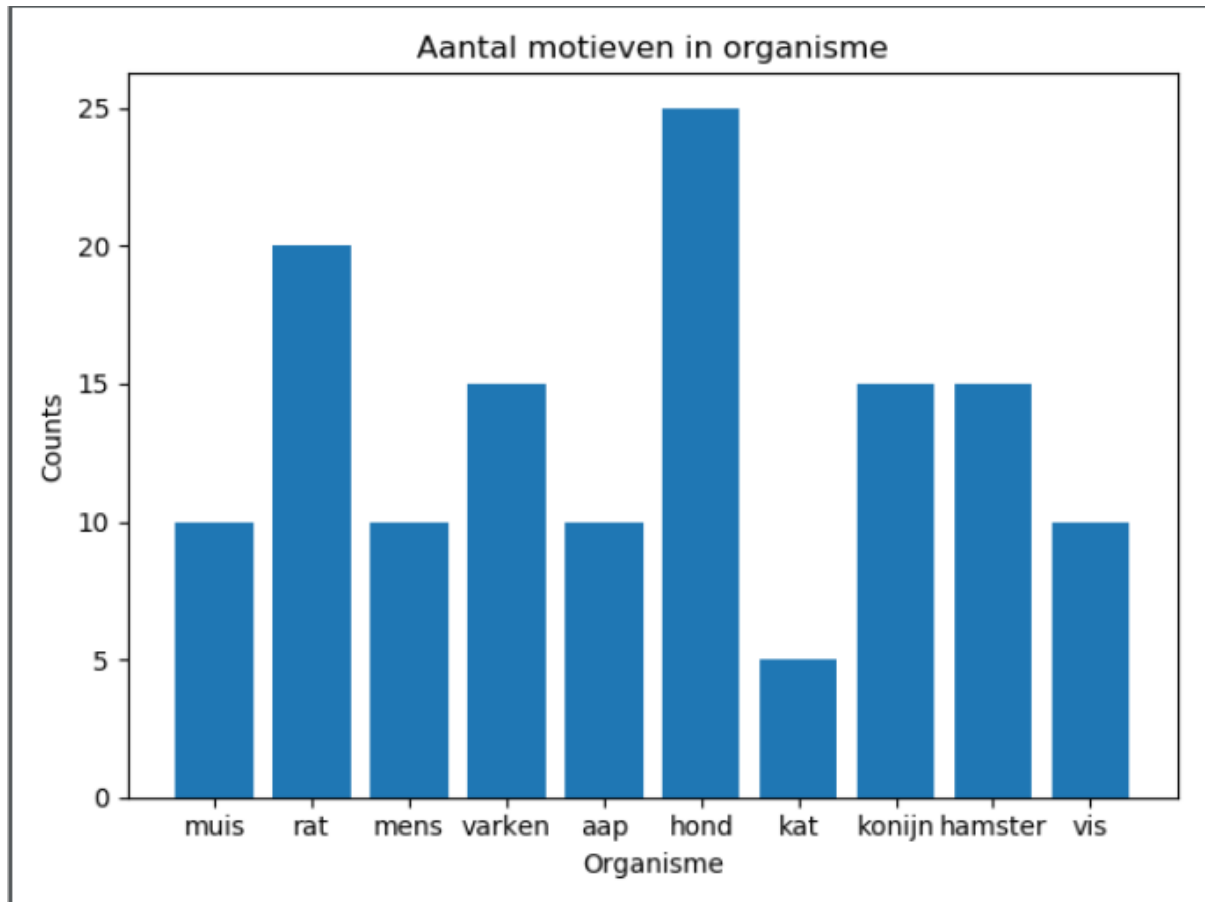


Figure 1: Voorbeeld counts van het aantal motieven gevonden in ieder organisme

Opdracht

Ga uit van het bestand HBI-OWE3a ploop.fa. Het op te leveren script dient aan de volgende functionele eisen te voldoen:

1. Lees het FASTA bestand ploop.fa
2. Indien het FASTA bestand niet bestaat wordt de exception afgevangen en krijgt de gebruiker een mededeling dat het bestand niet gevonden is.
3. Het programma zoekt in het bestand naar alle proteïnes die voldoen aan het kenmerk van de 'P-loop' van de mens (Homo sapiens). Let op: dit mag je niet hardcoden!
4. Het programma maakt een grafiek met het aantal counts per organisme waarin dit motief gevonden is.
5. Deze grafiek wordt weergegeven in een GUI met checkbox.

Het op te leveren script dient aan de volgende niet functionele (technische eisen) te voldoen.

1. Je zorgt dat je code modulair is. Dit houdt in:
 - (a) Je code bestaat uit functies
 - (b) Je maakt correct gebruik van parameters
 - (c) Je maakt correct gebruik van return statements
 - (d) Je zorgt voor goede docstrings en commentaar
 - (e) Je kiest voor duidelijke beschrijvende namen voor je variabelen
 - (f) Je code voldoet aan de regels van PEP8
2. Alle te verwachten exceptions worden afgevangen
3. Je maakt correct gebruik van datastructuren (lijsten en/of dictionaries)
4. De snelheid van je script is van geen belang
5. Het script is geschreven in python
6. Je mag gebruik maken van de volgende python packages:
 - (a) regex
 - (b) matplotlib
 - (c) tkinter

Tip: Aan de hand van het beoordelingsformulier kun je bepalen wat belangrijke aspecten zijn om op te focussen.