# Casus voor Assessment Python – Programmeren in Python 2

De centrale opdracht zal gebruik maken van een UniprotKB text file. Dit is een bestand waarin proteïne sequenties met bijbehorende features opgeslagen staan. In deze casus gaan wij niet kijken naar een enkel organisme, maar zijn we juist van plan om te kijken naar de conservering van een specifiek domein onder verschillende organismen. Wij kijken naar de conservering van het pectinesterase domein. Dit domein katalyseert de hydrolyse van pectine in pectaat en methanol. Het speelt een andere rol in verschillende organismen. In planten wordt het eiwit onder andere gebruikt om te rijpen, en plant pathogene organismen gebruiken het om de plant weefsel sneller af te breken en het rot proces dus in gang te brengen.

Afbeelding met tekst, Lettertype, lijn, schermopname

Automatisch gegenereerde beschrijving

Figuur : Pectinesterase consensus patroon

# Data beschrijving

Er wordt onderzoek gedaan naar conservering van domeinen op eiwitsequenties. Op dit moment heb je als Bio-Informaticus één bestand tot je beschikking. Dit bestand is een UniprotKB text file. Dit bestand biedt informatie over een collectie eiwit sequenties waar ons geconserveerde domein in terug komt. In het bestand staan verschillende line codes, deze line codes komen overeen met verschillende typen content. Neem de volgende link over in je browser: <https://web.expasy.org/docs/userman.html>. Via deze link kan je meer informatie vinden over hoe het bestand is opgebouwd. In figuur 2 is ook een overzicht van de verschillende line codes te vinden.

Afbeelding met tekst, schermopname, nummer, menu

Automatisch gegenereerde beschrijving

Figuur 2: UniprotKB bijbehorende line code en content tabel.

# De taken zijn als volgt:

1. Maak een GUI waarmee je het invoer bestand generiek kunt inladen.
2. Lees het bijgeleverde bestand op een juiste manier uit en parse de data.
3. Zoek welke vormen van de consensus sequenties gevonden worden tussen de sequenties. Consensus sequentie patroon: [GSTNP]-x(6)-[FYVHR]-[IVN]-[KEP]-x-G-[STIVKRQ]-Y-[DNQKRMV]-[EP]-x(3)-[LIMVA] (link: [PROSITE (expasy.org)](https://prosite.expasy.org/PS00800))
4. Maak een keuze uit minstens 1 van de volgende opties:
   1. Laat zien per organisme hoeveel eiwitten de bijbehorende consensus sequentie hebben.
   2. Laat zien per organisme welke consensus sequenties er zijn gevonden.
   3. Toon per accessiecode de bijbehorende sequentielengte
5. Maak een juiste plot voor minstens 1 van de volgende opties:
   1. Plot de top 10 van het aantal eiwitsequenties per organisme.
   2. Wat is de verhouding van het gebruik van het eerste aminozuur in de consensus sequentie (zie patroon bij punt 3 en figuur 1)
6. Werk met objecten.
7. Bouw je programma generiek op zodat je functies en methoden van objecten op verschillende manieren kunt aanroepen.
8. Voor info over de manier van beoordeling zie: [Assessment Beoordelingsformulier](https://han.onderwijsonline.nl/manage/information/file/MNZP6xD6/eyJpdiI6IkNua1k2SEhIcGFyb0VIcGYrcGNhc2c9PSIsInZhbHVlIjoiTFJtRWNkMHBxczQ0RFZHUzliWVd5Tk1VN1o4RDFuZ1V3Mm9mZnk4V0sxbi9ET08wS0d1cWw4YzFqMTNWS3pBZWY2bjNpZTNmS3JpNWwwNktseGN3bXc9PSIsIm1hYyI6ImNiZTIxYWVlNjRkN2Q3MTAyNGE2OWU5NmMxMmI4YTQ4MDljMmRkNzJhMTE2NDQ2NDg1NTU3NGMwYjE2NjBjMWIiLCJ0YWciOiIifQ==)
9. Mocht je extra relevante data willen laten zien via GUI of plot, staan wij hier zeker voor open!

## Richtlijnen

Probeer het programma flexibel op te stellen zodat je aan de gebruiker tekstueel en via een grafische weergave (plot en GUI) informatie kan laten zien.

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

(g) Niets is hard gecodeerd!

2. Alle te verwachten exceptions worden afgevangen   
3. Je maakt correct gebruik van verschillende datastructuren   
4. De snelheid van je script is van geen belang   
5. Het script is geschreven in Python v 3.\*   
6. Je mag alleen gebruik maken van de volgende python packages, indien je gebruikt wilt maken van een andere package, neem contact op met de docent:

(a) regex   
 (b) matplotlib   
 (c) tkinter

(d) random

(e) pickle

(f) os

(g) time

(h) numpy

## Assessment

Om deel te nemen aan het assessment moet je je GIT link inleveren via Handin. Wanneer je begint met de opdracht lever je hier meteen je Gitea link op in (dan kan je dit niet meer vergeten). Let dus goed op, de code en versie beheer gedurende de opdracht moet gedaan worden via onze GIT server. Lever je GIT link in voor: 11-01-2024 11:59, let op dit is midden op de dag! Te laat ingeleverd betekend geen deelname aan het assessment.

Gedurende het assessment zullen wij je vragen stellen over je code, werking van je code, waarom je bepaalde keuzes hebt gemaakt en wat je eventueel anders had kunnen doen. Dus bereid dit goed voor.