

Thema 1, eindopdracht

Maak bij deze opdracht zoveel mogelijk gebruik van functies

Verplichte onderdelen:

- Download vanaf de NCBI-website het Protein Data Bank-bestand behorende bij het eiwit betrokken bij het ontstaan van taaislijmziekte in de mens (heeft ID: 5UAK).
- Schrijf een programma dat het PDB-bestand met de structuur van het eiwit inleest en bepaal hieruit de aminozuur-sequentie. (Hint: zoek de regels met tag "SEQRES".) Vertaal de 3-letter aminozuur codes uit het PDB bestand naar de bijbehorende 1-letter codes, en sla het resultaat op in een string. (Hint: maak in deze eindopdracht zoveel mogelijk gebruik van functies en documenteer bij elke functie wat deze doet.)
- Download eveneens de bijbehorende aminozuur-sequentie van het taaislijmziekte-eiwit van NCBI (heeft ID: NP_000483.3). Lees ook deze aminozuur-sequentie in en print beide sequenties naar scherm. Vergelijk de aminozuur sequenties MET DE HAND/OOG uit het FASTA bestand met die uit het PDB-bestand. Is er een verschil?
- Bepaal de vervolgens middels het PDB-bestand waar zich alpha-helices en beta-sheets bevinden. (Hint: de relevante regels worden met de labels "HELIX" en "SHEET" aangegeven; zoek in de PDB-documentatie op welke delen je moet hebben om het begin en eind van een helix of sheet uit te knippen.)
HELIX documentatie: http://www.bmsc.washington.edu/CrystaLinks/man/pdb/part_42.html
SHEET documentatie: http://www.bmsc.washington.edu/CrystaLinks/man/pdb/part_44.html
- Selecteer de aminozuren die zich in een helix en sheet bevinden en plak die aan elkaar tot één helix- en één sheet-sequentie (gebruik hiervoor de sequentie uit het PDB-bestand). Schrijf beide sequenties (helix en sheet) naar één bestand in multi-fasta formaat. Bedenk daarbij per sequentie een gepaste header.
- Bepaal van de aminozuur-sequentie als geheel (dus niet van de helix/sheet specifiek) de compositie en geef die in de terminal weer middels een histogram (m.b.v: '#'). Doe hetzelfde met de helix- en sheet-sequenties. Vergelijk de resultaten voor de helices en sheets met die van het eiwit als geheel (MET HET OOG). Wat valt je op als je met het oog de histogrammen vergelijkt?

BONUS OPDRACHT

- Maak een viewer die de zogenaamde *hydrofobiciteit* van je eiwit in een grafiek weergeeft. De hydrofobiciteit of hydropathie geeft aan hoeveel een aminozuur "van water houdt", oftewel erin oplost. Dit heeft consequenties voor het eiwit in z'n omgeving in de cel. Plot per positie in je eiwitsequentie de hydrophobicity-score die daarbij hoort in een lijngrafiek. Schrijf je functie met behulp van Turtle graphics. (Hint: op http://en.wikipedia.org/wiki/Hydrophilicity_plot vind je een tabel met de waarden per aminozuur, en voorbeelden van grafieken; let op: het nulpunt van de y-as bevindt zich in het midden van je grafiek, de nul van de x-as begint wel helemaal links.)
- Zoek de positie van de helices en sheets op die je hiervoor hebt bepaald. Wat valt je op aan de grafiek op die plekken? (Hint: teken op die plekken de lijn van de grafiek in een andere kleur.)

Lever je programma in via BlackBoard in de vorm van een python-bestand genaamd `Groep_nummer_eindopdracht.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.