

## Thema 1, script 3

### Verplichte onderdelen:

- Schrijf een programma dat een FASTA-bestand met een door de gebruiker gegeven naam inleest en dat de inhoud identiek wegschrijft naar een FASTA-bestand met een andere gespecificeerde naam. Controleer dat het geproduceerde bestand inderdaad een kopie is van het origineel door de inhoud en eigenschappen van het bestand te vergelijken. (Hint: vergeet niet af te sluiten met `outputfile.close()`!)
- Zorg dat de nucleotiden/aminozuren als hoofdletters worden weggeschreven, zelfs als het invoerbestand in kleine letters is. De header-regel dient echter intact te blijven!
- Voeg aan je code een (integer) variabele toe die bepaalt hoeveel nucleotiden/aminozuren per regel in de uitvoer zullen worden weggeschreven. Het programma moet bijvoorbeeld in staat zijn om een FASTA-bestand met 70 tekens per regel in te lezen, en het weg te schrijven met -zeg- 50 tekens per regel. De header-regel dient opnieuw intact te blijven! (Hint: bewaar de header en de aaneengeplakte sequentie elk in een string-variabele.)
- Breid je programma uit zodat het meerdere door de gebruiker gespecificeerde FASTA-bestanden als invoer kan inlezen, maar slechts één enkel “multi-FASTA” bestand produceert waar alle headers met sequenties onder elkaar in worden vermeld. (Hint: lees [https://en.wikipedia.org/wiki/FASTA\\_format](https://en.wikipedia.org/wiki/FASTA_format) omtrent het [multi-]FASTA formaat.)

### Bonus-onderdelen:

- Breid je programma uit zodat het programma een waarschuwing op het scherm weergeeft wanneer er karakters voorkomen in de sequentie(s) die niet toegestaan zijn volgens het FASTA-formaat. (Hint: de toegestane nucleotide- en aminozuur-afkortingen vind je eveneens op [https://en.wikipedia.org/wiki/FASTA\\_format](https://en.wikipedia.org/wiki/FASTA_format).)
- Pas je programma aan zodat de nucleotiden/aminozuren worden weggeschreven in groepjes van 10 met een spatie ertussen. Bijvoorbeeld, voor 50 tekens per regel:  
    > Sequence header  
    MTEITAAMVK ELRESTGAGM MDCKNALSET NGDFDKAVQL LREKGLGKAA  
    KKADRLAAEG LVSVKVSDDF TIAAMRPSYL SYEDLDMTFV ENEYKALVAE  
    ...
- Pas je programma aan zodat het ook correct om kan gaan met invoer die zelf reeds bestaat uit multi-FASTA bestanden. Dat wil zeggen, alle sequenties in alle invoerbestanden dienen in één uitvoerbestand te worden weggeschreven.
- Laat je programma op het scherm weergeven hoeveel sequenties er in totaal in het uitvoerbestand zijn weggeschreven.

Lever je programma in via BlackBoard in de vorm van een python-bestand genaamd **script3.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.