

Projet de programmation Python pour la protéomique

Université de Rouen Normandie

M1 BIMS – M2 CCB4 2022–2023

Recherche dans une base d’empreintes peptidiques

À l’aide des exercices effectués en TP vous devez, dans un premier temps, fournir un script qui à partir d’un nom de fichier fasta contenant une protéine dite « requête » et d’un fichier multifasta contenant des protéines constituant une « base de données » permet de digérer *in silico* avec la trypsine les protéines « requête » et de la « base de données » et de calculer les masses des peptides pour les stocker dans des fichiers au format **mgf** (ceux correspondants aux peptides de la « base de données » seront regroupés dans un répertoire).

Puis vous devez écrire un programme Python qui prend en entrée le nom du fichier de masses peptidiques au format **mgf** de la « requête » et le nom du répertoire contenant les fichiers de masses peptidiques au format **mgf** de la « base de données » et qui renvoie la liste des masses contenue dans le fichier de la « base de données » ayant le plus de *hits* avec le fichier « requête ».

L’instruction :

```
for fileName in os.listdir("dir") :
```

permet de considérer les fichiers du répertoire `dir`. Il faut, bien sûr, importer le module `os`.