**Projet tutoré – Goran Bich**

**Exploration de nouvelles approches pour définir et évaluer une famille de régions répétées par alignement en utilisant l’annotation UniProt et les prédiction AlphaFold**

Les repeats WD forment une des plus grandes familles de repliements retrouvées chez l’homme avec plus de 2 000 instances réparties sur 300 protéines différentes. Ces régions répétées s’associent pour former une structure tridimensionnelle globulaire nommée bêta-propeller, possédant 4 à 8 pales ou blades, selon les familles de repeats.

Ces repeats sont difficile à caractériser et classifier de par leur grande variabilité au niveau de la séquence. Cette complexité est source de nombreuses erreurs d’annotations visibles dans la banque de données UniProt.

Là où le projet 3 a pour objectif de redéfinir les limites des domaines bêta-propellers formés par les repeats WD pour tenter de redéfinir les limites desdits repeats, ce projet propose une approche complémentaire, utilisant directement les repeats en tant qu’objets d’études. On s’intéressera principalement à l’alignement de séquence et structural de repeats tels qu’ils sont annotés sur UniProt. On pourra ainsi identifier des paramètres critiques au sein de la famille des WD et de filtrer et grouper en sous-groupes au besoin. On utilisera ces informations pour juger de la qualité de l’annotation actuellement disponible sur UniProt.

L’objectif final est de proposer plusieurs définitions différentes des repeats WD et de trouver des critères objectifs pour jauger de la validité de chacune.

Dans le cadre de ce projet vous :

* Interrogerez la banque de donnée UniProt pour récupérer l’annotation des repeats WD
* Alignerez, filtrerez et grouperez les séquences des repeats puis identifierez les résidus critiques
* Récupérerez les structures prédites par AlphaFold (et/ou les structures PDB) des repeats et les superposerez, les filtrerez et les grouperez
* Identifierez les résidus critiques dans la structure et comparerez avec les résultats obtenus par alignements de séquence
* Comparerez vos résultats/conclusions aux annotations existantes et tenterez d’identifier les cas ou l’annotation est la plus susceptible d’être fausse
* A partir des conclusions tirées des deux approches précédentes, vous identifierez différentes manières de définir les repeats WD au sein d’un bêta-propeller et trouverez des moyens de quantifier la pertinence de chacune des méthodes (il sera éventuellement possible d’échanger avec le groupe du projet 3 pour trouver d’autres pistes)