

Réévaluation des annotations UniProt d'une famille de régions répétées par alignement sur les structures prédites par 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 classer 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.

Leur propriété à former des structures globulaires et l'accès aux prédictions de cet ensemble de protéines grâce à AlphaFold permettrait via alignement structural de redéfinir les limites des bêta-propellers formées par les repeats WD. Ces nouvelles limites pourraient servir de base pour de nouveaux alignements de séquence permettant de recaractériser les limites des repeats WD et pourraient servir à corriger les annotations existantes sur cette famille de repeats.

De manière plus générale, la même approche pourrait être appliquée aux autres familles de repeats formant des bêta-propellers.

Dans le cadre de ce projet vous :

- Interrogez la banque de donnée UniProt et/ou InterPro pour récupérer les annotations de repeats disponibles
- Identifiez le nombre de repeats WD nécessaire pour un repliement complet en bêta-propeller
- Récupérez les structures prédites par AlphaFold des protéines contenant un repeat WD et découpez ces protéines en fragments grâce au désordre prédit ou pLDDT
- Utilisez une ou plusieurs structures de bêta-propeller de référence que vous superposez grâce à PyMol sur tous les fragments produits
- Identifiez la région superposée utilisant les infos contenues dans les fichiers PDB issues des superpositions pour proposer les nouvelles limites des bêta-propellers
- Alignez les séquences avec ces nouvelles limites pour tenter de mieux redéfinir les repeats WD
- Proposez de nouvelles annotations prenant en compte l'alignement de séquence et de structure