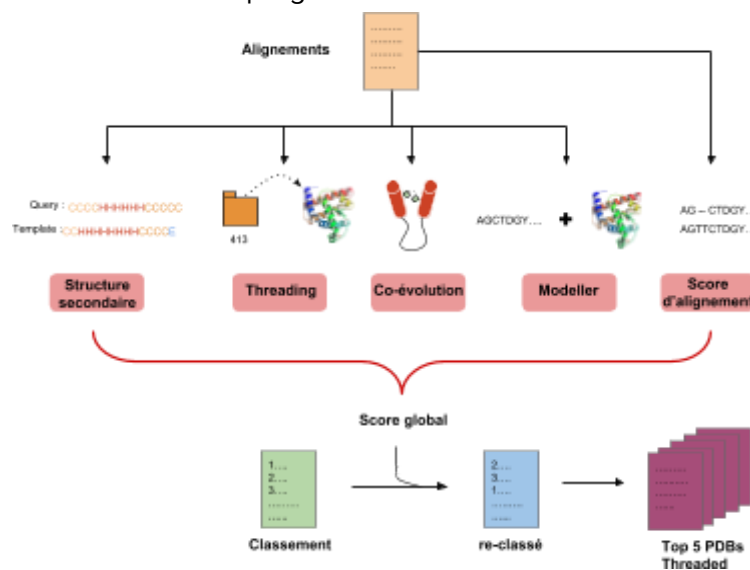


Avec le récent développement des nouvelles technologies de séquençage, le nombre de séquences en acides aminés de protéines a augmenté beaucoup plus vite que le nombre de structures 3D. Le défi est aujourd'hui de mieux comprendre la conformation des protéines, qui est étroitement reliée à leur fonction, en utilisant l'information des séquences en acides aminés.

Ce programme effectue une reconnaissance de motifs de repliement structuraux à partir de la séquence en acides aminés d'une protéine, et de la base de données de structures protéiques 3D HOMSTRAD. Un alignement profil-profil est effectué entre la séquence requête et les séquences supports de la base de données (équipe amont) qui permet un premier classement. A partir de ces alignements par paire, 5 scores sont calculés, estimant la similarité de structure entre les protéines requête et support (score d'adéquation des structures secondaires, score de threading, score de co-évolution, score issue de modélisation par homologie avec Modeller, et score d'alignement de l'équipe amont afin de prendre en compte le premier classement généré). Un score global est calculé comme combinaison linéaire de ces 5 scores, avec des poids ajustés par étalonnage sur un jeu de protéines test. Enfin, les protéines supports sont alors re-classées par probabilité de similarité structurale avec la protéine requête, et les 5 structures les plus vraisemblables sont générées par threading.

Par évaluation sur un jeu de séquences test, les scores calculés se sont avérés individuellement assez peu sensibles pour la reconnaissance de motifs de repliement. Le score le plus informatif est le score obtenu par threading, tandis que le moins performant est le score de co-évolution. Cependant, la complémentarité des méthodes permet d'obtenir un score combiné présentant de bonnes performances dans la reconnaissance de motifs de repliement protéique et permet entre autre de faire remonter des protéines du même *fold* dans le classement, comme l'Agglutinin qui passe de la 195<sup>e</sup> position à la première position lorsque la query est l'Agglutinin.

Concernant les séquences expérimentales, au moment de la rédaction seules 3 autres équipes ont posté leurs résultats. Il est possible de remarquer que certaines séquences comme trQ6HKV et recN ont des modèles issues de la même classe de métafold a, tandis que pour les autres les résultats sont cohérents avec une autre équipe mais pas avec les deux autres. Il est alors difficile d'être certains de la fiabilité du programme.



**Figure 1 : Schéma du principe général de la partie aval.** Un fichier foldrec contenant les alignements par paire entre la séquence étudiée et la base de données HOMSTRAD, ainsi qu'un score permettant de classer les différents modèles de repliement est récupéré. Suite à ça, plusieurs types de scores structuraux sont implémentés et combinés en un score global, servant à re-classer les modèles, afin de déterminer le motif de repliement auquel la séquence requête a le plus de chances d'appartenir.