

Sujet de thèse

Encadrants : Adrien NICOLAÏ (ICB) et Fabrice NEIERS (CSGA)

Titre : Construction *de novo* d'une enzyme synthétique permettant d'éliminer des polluants alimentaires et non alimentaires

Les protéines, faites de chaînes d'acides aminés, sont des macromolécules essentielles au monde du vivant et la résolution de leur structure tridimensionnelle (3-D) est un atout majeur à la compréhension de leur fonction biologique. Depuis plusieurs décennies, des efforts considérables ont été entrepris pour résoudre expérimentalement les structures atomiques des protéines, que ce soit par diffraction des rayons X, résonance magnétique nucléaire ou par cryo-microscopie électronique. Ainsi, près de 200 000 structures 3-D de protéines ont été résolues et déposées dans la banque de données en ligne Protein Data Bank. Cependant, cela représente une infime fraction des centaines de milliards de séquences de protéines référencées dans la base de données Universal Protein Resource (UniProt). Cette limitation est essentiellement due aux difficultés rencontrées expérimentalement pour résoudre la structure atomique d'une molécule biologique unique. Ainsi, prédire la structure 3-D adoptée par une protéine à partir de sa séquence d'acides aminés, le problème du repliement des protéines, est un challenge important, et ce depuis plus de 50 ans. Ces deux dernières années, la prédiction de la structure 3-D des protéines a connu une avancée majeure à la suite du développement du programme **AlphaFold** par DeepMind [1], la compagnie d'intelligence artificielle (IA) de Google. Ce programme révolutionnaire dans le monde de la biologie et de la bio-informatique est basé sur un réseau de neurones qui prédit la structure 3-D des protéines à partir de leur séquence d'acides aminés, avec une précision comparable à celle obtenue par diffraction des rayons X de cristaux protéiques. Cette avancée majeure permet aujourd'hui d'envisager l'utilisation de cet outil à différentes fins, que ce soit au niveau fondamental pour une meilleure compréhension des phénomènes biologiques à partir de prédictions de structures 3-D de protéines mais également pour le développement d'applications dans le domaine médical ou environnemental.

En utilisant le programme **AlphaFold**, notre projet s'attache à concevoir de nouvelles enzymes de la superfamille des Glutathione Transferase (GST) à partir de séquences d'acides aminés "artificielles" en vue de créer des protéines avec des propriétés catalytiques nouvelles et/ou optimisées. Ces enzymes GST, multifonctionnelles, sont notamment impliquées dans les processus de détoxification de l'ensemble des organismes vivants sur Terre incluant l'homme. En pratique, ces enzymes catalysent la biotransformation des composants toxiques en greffant une molécule de glutathion les rendant ainsi plus hydrophiles et donc plus facilement éliminables. De plus, ces enzymes sont robustes, tolérantes vis-à-vis des mutations et relativement simples à exprimer et à purifier, qui sont des propriétés essentielles pour le design de protéines synthétiques. Cela justifie donc fortement leur exploitation en biotechnologie. Le choix des séquences d'acides aminés sera établi pour imiter l'évolution [2]. Ces enzymes existent depuis l'ancêtre commun aux espèces vivant actuellement sur Terre et évoluent par adjonction de domaines et mutations des poches de reconnaissances des substrats. En effet, les structures des GST partagent une structure commune de type thioredoxin et au cours de l'évolution, des domaines structuraux se sont additionnés et/ou

insérés dans la séquence d'acides aminés et des mutations sont apparues afin de réaliser des fonctions biologiques très diverses et ce dans différents endroits de la cellule.

La conformation des structures 3-D déterminées par IA sera affinée et validée en utilisant des modèles physiques des forces d'interaction par simulations de dynamique moléculaire (MD). Cette technique permet d'aller au-delà des modèles statiques prédit par l'IA et de simuler la structure dans différentes conditions thermodynamiques de pression et de température. L'affinité du glutathion et de différents substrats pour les structures générées par IA sera calculée par MD dans différentes conditions. Les structures des GST les plus prometteuses prédites par la combinaison de l'IA et de la MD seront produites à la suite de la construction des gènes synthétiques correspondants, puis purifiées afin de mesurer leurs propriétés catalytiques et de sélectivité vis-à-vis de substrats comme des polluants alimentaires et non alimentaires. Les structures seront résolues expérimentalement par diffraction des rayons X. La dynamique prédite par MD pourra ainsi être comparée aux mesures par DRX des B-facteurs qui décrivent la flexibilité locale et globale de la protéine et aux données allostériques (dynamique entre sites catalytiques). Cette partie du projet s'appuie notamment sur les expertises complémentaires de l'équipe «Physique appliquée aux Protéines » de l'ICB (MD) et de l'équipe « Flaveur : de la molécule au comportement » du CSGA (GST, expériences) qui ont une longue expérience collaborative avec plusieurs publications co-signées [3,4,5] et une thèse co-dirigée [6].

Ces nouvelles enzymes synthétiques pourraient avoir un impact dans les domaines de la toxicologie alimentaire, de la décontamination des milieux pollués (bioremédiation) et en biotechnologie. Cette nouvelle procédure, pluridisciplinaire, permettra donc d'explorer théoriquement une multitude d'enzymes "artificielles" avec de nouvelles propriétés et de guider la sélection des plus intéressantes à synthétiser et pourrait conduire à une valorisation sous forme de brevets.

[1] Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596, 583-589 (2021).

[2] Frova, C. Glutathione transferases in the genomics era: new insights and perspectives. *Biomolecular Engineering*, 23, 149-169 (2006).

[3] Gonzalez D. et al. Characterization of a drosophila glutathione transferase involved in isothiocyanate detoxification. *Insect Biochemistry and Molecular Biology*, 95, 33-43 (2018).

[4] Grassein, P. et al. Curvature and torsion of protein main chain as local order parameters of protein unfolding. *The Journal of Physical Chemistry B*, 124(22), 4391-4398.

[5] Nicolai, A. et al. Free-energy landscape analysis of protein-ligand binding: the case of human glutathione transferase A1. *Applied Sciences*, 12(16), 8196 (2022).

[6] Grassein, P. Simulation of receptor-ligand recognition mechanisms of human Glutathione Transferases by free energy landscape calculation: applications to the science of taste and cancer. Thèse de doctorat (2019).