

Le dossier – Prix Nobel

La structure des protéines

→ **F. DIÉVART**
DUNKERQUE.

Le prix Nobel de chimie 2024 a couronné des découvertes permises par l'utilisation de l'intelligence artificielle (IA) afin de mieux comprendre la structure des protéines.

Ce prix a été décerné conjointement à un Britannique, Demis Hassabis, et un Américain, John M. Jumper, chercheurs de Google DeepMind et à un autre Américain, David Baker, de l'Université de Washington. Ce prix Nobel récompense deux types de travaux complémentaires concernant la structure des protéines, respectivement la prédiction de leurs structures et le dessin de nouvelles protéines. Ainsi, ce prix Nobel consacre un modèle dénommé AlphaFold développé par les chercheurs de GoogleMind, qui est capable de prédire les structures complexes des protéines et David Baker a été couronné, car ces travaux reposant sur l'IA ont permis de concevoir des protéines d'un genre entièrement nouveau.

Les lignes qui suivent sont principalement issues ou adaptées de l'article "Prix Nobel de chimie 2024 : design des protéines et prédiction de leurs structures, deux facettes d'une même médaille", publié le 10 octobre 2024 par Alessandra Carbone (professeur d'informatique et membre senior de l'Institut universitaire de France, Sorbonne Université) sur le site TheConversation.fr. Son préambule illustre l'apport des travaux récompensés : "... reconnaissance de deux avancées spectaculaires en sciences des protéines... comprendre comment les protéines interagissent est essentiel pour dévoiler les mécanismes qui gouvernent la vie, et permettre d'éclairer à

l'avenir la susceptibilité aux maladies et leur gravité, tout en facilitant le déchiffrement des liens entre gènes, maladies et médicaments."

Prédire la structure des protéines

L'ADN permet, via l'ARN de coder des chaînes d'acides aminés se repliant sur elles-mêmes et qui vont constituer les protéines. En se repliant, les protéines adoptent une forme tridimensionnelle qui est elle-même codée dans la séquence des acides aminés. Mais si le principe paraît simple, il était cependant difficile de prédire la structure tridimensionnelle d'une protéine à partir de la séquence connue de ses acides aminés et, inversement, lorsque la structure tridimensionnelle est connue, il était difficile de connaître la séquence d'acides aminés qui l'a produite. Les travaux de Demis Hassabis et John Jumper ont permis une avancée majeure dans ce domaine. Leur objectif initial était de reconstruire une structure protéique à partir de sa séquence d'acides aminés. Ces travaux ont été couronnés de succès puisqu'ils ont réussi à décrypter le code des protéines, rendant possible la prédiction de la structure de presque toutes les protéines naturelles non désordonnées.

Pour ce faire, en 2018, ils ont développé un logiciel d'IA dénommé AlphaFold, fournissant une prédiction de la structure des protéines à partir de leur séquence d'acides aminés, puis, une deuxième version de ce logiciel a été disponible en 2020 et dénommée AlphaFold2. Ce logiciel permet de prédire la structure de centaines de milliers de protéines. Il a été mis en accès libre à partir de 2021. John Jumper indiquait en 2020 dans la revue

Science qu'AlphaFold associe l'apprentissage profond à un "algorithme d'attention" qui, en quelque sorte, copie la manière dont un humain assemble les pièces d'un puzzle quand il connecte entre eux de petits amas de pièces préalablement réalisés, un amas représentant ici un petit assemblage d'acides aminés. Le système AlphaFold utilise 128 processeurs d'apprentissage automatique qui ont entraîné cet algorithme sur environ 170 000 structures protéiques déjà connues. Selon Wikipédia, l'IA peut ensuite modéliser des protéines constituées de dizaines à centaines d'acides aminés en tenant compte de ce qu'elle a appris sur la manière dont l'ordre de ces acides aminés dicte la façon très complexe dont la myriade de poussées et de tractions entre eux donne naissance aux formes 3D complexes des protéines, qui, à leur tour, déterminent leur fonctionnement. La connaissance de ces formes aide les chercheurs à concevoir des médicaments qui peuvent se loger dans les poches et les crevasses des protéines. Il devient alors possible de synthétiser des protéines avec la structure souhaitée.

Et, comme l'écrit Alessandra Carbone : "En 2020, ils ont présenté AlphaFold2, qui a atteint une précision d'environ 90 % par rapport aux structures cristallographiques connues, en accord avec les erreurs expérimentales lors de la détermination des structures. Cette avancée spectaculaire a bouleversé le monde de la biologie : du jour au lendemain, la "bibliothèque de protéines" accessibles aux chercheurs est passée des 200 000 structures cristallographiques à 200 millions de modèles AlphaFold2, quasiment identiques aux structures des cristaux ! Désormais, un biologiste, virologue ou médecin peut soumettre une séquence d'acides aminés à AlphaFold2,

et obtenir instantanément sa structure tridimensionnelle. Cette visualisation minutieuse de la position des atomes permet d'analyser les points vulnérables d'une protéine lors de ses interactions, qui peuvent être exploitées pour concevoir des médicaments bloquant ces interactions. Elle permet également d'identifier les résidus sensibles aux mutations futures, ainsi que ceux impliqués dans les interactions avec des partenaires moléculaires potentiels."

Comme le spécifie la phrase indiquant que la précision est de 90 %, les prédictions d'AlphaFold ne sont pas exemptes d'erreurs. Toujours selon Wikipédia, une étude publiée dans Nature Methods a comparé les structures protéiques déterminées expérimentalement via des techniques comme la cristallographie aux rayons X avec les prédictions générées par AlphaFold. Les résultats ont montré que, bien que globalement très précises, les prédictions de l'IA présentent de légères déviations par rapport aux structures obtenues expérimentalement. Martin Weigt, professeur en biologie computationnelle à Sorbonne Université, explique que les prédictions d'AlphaFold sont *"systématiquement un peu plus éloignées des structures expérimentales, même si ces écarts restent minimes"*. Ces petites différences soulignent que la détermination expérimentale des structures reste importante, notamment pour valider ou affiner les prédictions générées par AlphaFold.

Enfin, Google DeepMind a annoncé, le 8 mai 2024, la mise au point d'AlphaFold 3 codéveloppé avec Isomorphic Labs, elle aussi filiale d'Alphabet. AlphaFold 3 ne se limite plus aux protéines et peut également prédire la structure et les interactions avec l'ADN, l'ARN et certains ligands ou ions.

Concevoir de nouvelles protéines

En parallèle aux travaux de Demis Hassabis et John M. Jumper, David Baker travaillait d'une façon différente sur la structure des protéines : il cherchait, en exploitant des séquences d'acides aminés, à concevoir des protéines nouvelles dans l'objectif d'exploiter la nature à nos propres fins. En 2003, il a résolu le problème grâce à son logiciel dénommé Rosetta, lui permettant de concevoir une séquence d'acides aminés correspondant à une structure protéique donnée : *"En combinant optimisation combinatoire, chimie et physique, Rosetta permet aujourd'hui aux scientifiques de créer des protéines aux formes et fonctions spectaculaires, jusque-là inconnues."*

L'extension au domaine de la médecine est permise par le logiciel Rosetta@home, projet de calcul distribué ouvert le 22 septembre 2005 et mené par le laboratoire du professeur David Baker de l'université de Washington. En juillet 2017, le projet avait une puissance de calcul de 156 téraflops. Son objectif est de déterminer la structure de protéines, afin d'élaborer des traitements contre les principales maladies humaines.

Les nouvelles questions

Pour Alessandra Carbone, ces deux avancées sont majeures en biologie et elle identifie quatre prochaines étapes dans ce domaine de recherche :

- décrypter la fonction des protéines, c'est-à-dire comprendre ce qu'elles font lorsqu'elles interagissent ;
- explorer leur dynamique, c'est-à-dire la manière dont elles changent de forme pendant ces interactions ;

– étudier comment les protéines non structurées interagissent avec d'autres protéines en sachant qu'AlphaFold2 ne peut donc pas en prédire la structure alors qu'elles représentent environ la moitié des protéines humaines ;

– concevoir des protéines dotées de nouvelles fonctions, en lien avec les avancées que la biologie synthétique apportera.

En synthèse

Le prix Nobel de chimie 2024 récompense des chercheurs ayant contribué à mieux connaître la structure des protéines à partir de leurs séquences d'acides aminés. Il met en lumière la puissance de la biologie computationnelle et les immenses possibilités qu'elle ouvre pour les sciences du vivant.

Ces travaux illustrent l'apport majeur de l'informatique et notamment de l'IA à la recherche fondamentale et consacrent les investissements de Google dans la recherche.

Ce croisement entre la recherche et l'apport de l'informatique est désormais productif du fait de l'importance des bases de données, de la puissance des puces informatiques et du développement de l'IA, ce dernier étant à l'origine du prix Nobel de physique de 2024.

L'auteur a déclaré les liens d'intérêts suivants : honoraires pour conférences ou conseils ou défraiements pour congrès pour et par les laboratoires Alliance BMS-Pfizer, Amgen, Astra-Zeneca, Bayer, BMS, Boehringer-Ingelheim, Menarini, Novartis, Novo-Nordisk, Pfizer, Sanofi France, Servier.