

► Le Prix Nobel de chimie 2024 a été décerné à Demis Hassabis et John M. Jumper (société *Google DeepMind*) pour leurs travaux sur la prédiction des structures protéiques guidée par l'intelligence artificielle (IA), et à David Baker (Université de Washington, Seattle, États-Unis) pour les siens sur la conception computationnelle des protéines. Cet évènement marque une révolution dans le domaine de la biologie structurale et a conduit, entre autres, à la génération d'une bibliothèque de près de 200 millions de structures prédites de protéines destinée à accélérer la recherche. Cette révolution de l'IA a également permis la conception de plusieurs protéines artificielles d'intérêt médical présentées dans cette revue. ◀

L'essor de l'intelligence artificielle (IA) est un nouveau catalyseur de changements majeurs, à l'instar de la révolution industrielle du XIX<sup>e</sup> siècle, qui a profondément transformé notre société en la faisant basculer de l'agriculture vers le commerce et l'industrie. Cette tendance est mise en exergue par de nombreux articles et revues, qui soulignent l'impact transformateur de l'IA sur la recherche scientifique [1]. Le prix Nobel de physique 2024 a ainsi été décerné à John Hopfield et Geoffrey Hinton [40] (→), pour leur travail pionnier dans le développement des réseaux de neurones et des méthodes d'apprentissage automatique. Ces concepts sont au cœur de l'IA et permettent d'établir des modèles et des relations à partir de masses de données [2]. Les méthodes de modélisation de structures tridimensionnelles de macromolécules biologiques ont bénéficié de manière spectaculaire de ces avancées, comme en témoigne le prix Nobel de chimie 2024 décerné à Demis Hassabis et John M. Jumper, créateurs du programme disruptif *AlphaFold* guidé par l'IA [3], et à David Baker, créateur de la suite *Rosetta* qui est plus

## Prix Nobel de chimie 2024

David Baker, Demis Hassabis et John M. Jumper

### La révolution de l'intelligence artificielle en biologie structurale

Stéphanie Aguéro-Pizzolo<sup>1</sup>, Emmanuel Bettler<sup>1</sup>, Patrice Gouet<sup>2</sup>



<sup>1</sup> Laboratoire de biologie tissulaire et ingénierie thérapeutique, CNRS UMR 5305, Université Lyon-1, Lyon, France.

<sup>2</sup> Laboratoire de microbiologie moléculaire et de biochimie structurale, CNRS UMR 5086, Université Lyon-1, Lyon, France.

[stephanie.aguero-pizzolo@univ-lyon1.fr](mailto:stephanie.aguero-pizzolo@univ-lyon1.fr)

[emmanuel.bettler@univ-lyon1.fr](mailto:emmanuel.bettler@univ-lyon1.fr)

[patrice.gouet@ibcp.fr](mailto:patrice.gouet@ibcp.fr)

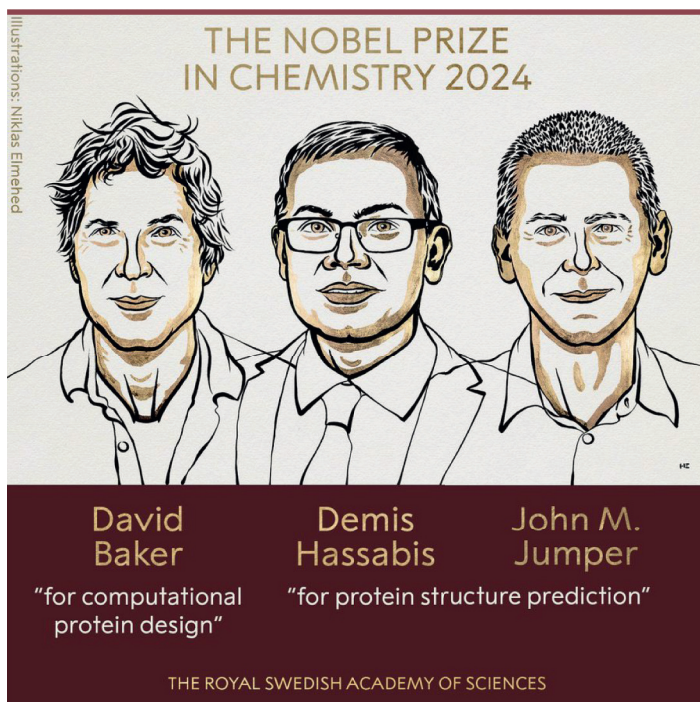
axée sur la conception de nouvelles protéines [4] (Figure 1).

Cette revue retrace, dans un premier temps, le chemin pavé de prix Nobel qui a mené la biologie structurale vers l'IA. Elle aborde ensuite, les spécificités des approches prédictives *AlphaFold* et *Rosetta*, et leurs apports à la recherche scientifique fondamentale et appliquée.

### L'évolution de la biologie structurale avant l'IA

La connaissance de la structure atomique à l'échelle de l'angström ( $10^{-10}$  m) des macromolécules biologiques telles que les protéines, les acides nucléiques (ARN, ADN), les lipides et les sucres complexes présente un intérêt majeur dans les domaines de la biologie, de la médecine, et des biotechnologies. Elle permet, par exemple, de mieux comprendre le mécanisme d'action de protéines d'intérêt thérapeutique [5] pour élaborer de manière rationnelle des molécules actives ou inhibitrices (antiviraux, antibiotiques, anticancéreux, etc.). L'acte fondateur de la biologie structurale remonte au 25 avril 1953 avec la parution, dans la revue *Nature*, de l'article de James Watson et Francis Crick décrivant la structure en double hélice de l'ADN [6]. Cette découverte résultait de la combinaison d'une approche expérimentale, basée sur la biophysique, et d'une approche déductive basée sur la minimisation d'énergie. En effet, les clichés de diffraction aux rayons X obtenus par Maurice Wilkins et Rosalind Franklin ne suffisaient pas pour établir la structure atomique de l'ADN. Ils étaient par contre essentiels pour comprendre son architecture hélicoïdale, dans laquelle viennent s'ajuster les paires de bases complémentaires





**Figure 1.** David Baker, biochimiste, est né en 1962 à Seattle, États-Unis. Il est professeur à l'Université de Washington de Seattle et directeur de l'*Institute for protein design*. Demis Hassabis, développeur en intelligence artificielle, est né en 1976 à Londres, Grande-Bretagne. Il détient un doctorat en neurosciences cognitives et est président-directeur général de la société *Google Deepmind* qu'il a co-fondé en 2010. John Michael Jumper, bioinformaticien, est né en 1985 à Little Rock, États-Unis. Il est titulaire d'un doctorat en chimie théorique et est président de la société *Google Deepmind*.

stabilisées par des liaisons hydrogène, imaginées par Watson et Crick. Cette découverte fut couronnée par l'attribution du prix Nobel de physiologie ou médecine en 1962 à Crick, Watson et Wilkins<sup>1</sup>. L'étape suivante, récompensée par le prix Nobel de chimie en 1962 à John Kendrew et Max Perutz pour l'observation des premières structures cristallines de protéines globulaires, respecte cette association symbiotique entre mesures expérimentales et minimisation d'énergie auxquelles s'est ajouté le calcul numérique grâce aux progrès de l'informatique [7]. En effet, en biocristallographie, le nombre de données expérimentales mesurées par la méthode de diffraction aux rayons X est généralement limité par rapport au nombre de degrés de liberté de la structure étudiée. Aussi, le processus mathématique itératif qui conduit à la détermination des coordonnées atomiques nécessite l'introduction de contraintes supplémentaires, qui tiennent compte des spécificités géométriques et énergétiques de l'assemblage (telles que la chiralité des résidus, les distances idéales de liaisons, etc.) [8]. L'attribution du prix Nobel de chimie en 2002 à Kurt Wüthrich pour le développement de la résonance magnétique nucléaire (RMN) [9] (→), et en 2017 à Jacques

Dubochet, Joachim Frank et Richard Henderson pour la cryo-EM [10] (→), a récompensé l'apport de nouvelles techniques complémentaires de biologie structurale. La RMN est particulièrement adaptée pour observer la dynamique des macromolécules en solution, tandis que la cryo-microscopie électronique (cryo-EM) est idéale pour aborder les assemblages moléculaires de grande taille [11]. Toutefois, ces nouvelles techniques n'ont pas supprimé la nécessité d'intégrer des contraintes énergétiques dans le processus de détermination expérimentale. De son côté, la cristallographie aux rayons X permet aujourd'hui de résoudre la structure d'assemblages aussi grands que le ribosome bactérien (dont la masse moléculaire est de 4,2 MDa), prouesse reconnue par l'attribution du prix Nobel de chimie à Venkatraman Ramakrishnan et Ada Yonath en 2009 [12] (→). (→) Voir *m/s* n° 12, 2017, page 1111. (→) Voir *m/s* n° 11, 2009, page 977.

### L'étude structurale des protéines par modélisation

Il est actuellement estimé que les structures protéiques comptent « seulement » 2 000 repliements différents ; l'accumulation de structures expérimentales dans la base de données *Protein Data Bank* (PDB) [13], qui compte plus de 200 000 entrées pour les protéines seules, aurait dû rapidement conduire à l'élaboration d'approches prédictives fiables basées uniquement sur la séquence.

En effet, dans la majorité des cas, les protéines se replient spontanément en fonction de leur séquence en acides aminés, pour adopter une structure stable formée d'un cœur hydrophobe entouré de résidus majoritairement hydrophiles en contact avec le solvant [14]. Les hypothèses actuelles suggèrent que ce processus d'effondrement hydrophobe, qui peut ne prendre que quelques microsecondes, est complexe et qu'il suit différentes voies et intermédiaires le long d'un entonnoir énergétique menant à une structure unique [15]. La méthode de prédiction de structure *ab initio* épouse ce concept, en associant des calculs de trajectoires par dynamique moléculaire à des calculs de minimisation d'énergie [16]. Elle est cependant limitée à l'étude des peptides et des petites protéines (de séquences inférieures à 100 résidus) en raison des temps de calcul nécessaires pour explorer l'espace conformationnel afin de trouver la structure de plus faible énergie [17]. La méthode d'enfilage (*threading*) utilise l'information spatiale de peptides de structure connue le long de la séquence étudiée et permet d'aborder des protéines de plus grande taille [18]. La modélisation par homo-

<sup>1</sup> Comme justement souligné plus tard, l'apport fondamental de Rosalind Franklin fut volontairement ignoré.



logie utilise, quant à elle, l'information spatiale de protéines de séquence significativement similaire (avec généralement plus de 30 % d'identité) et de structure connue [19]. Cette approche globale est rapide, mais elle est limitée par la nécessité de trouver au moins une protéine guide dans la PDB, où certaines classes sont sous-représentées comme les protéines membranaires (dues à des problèmes de solubilité qui rendent difficile la résolution de leur structure). Cette approche de modélisation a été largement employée avec des programmes comme *Swiss-Model* [20] et *Modeller* [21] jusqu'en 2018, date à laquelle est paru le programme *AlphaFold* développé par Demis Hassabis et John M. Jumper de la société *Google Deepmind* [3]. Ce programme basé sur l'IA [22] (→), a en effet obtenu des scores de confiance supérieurs aux autres programmes et algorithmes lors de la 13<sup>e</sup> rencontre biennale d'évaluation critique CASP (*Critical Assessment of Structure Prediction*), qui permet de confronter différents programmes de prédiction de structure [23].

### Demis Hassabis et John M. Jumper : la révolution de la prédiction de structures guidée par l'IA avec *AlphaFold*

La version suivante, *AlphaFold-2* [3], a causé une véritable révolution lors de la rencontre CASP14 de 2020 en surpassant, cette fois de loin, tous les programmes concurrents dans la catégorie « Modélisation libre ». Pour la première fois, une méthode purement informatique était en mesure de rivaliser avec la précision expérimentale pour déterminer la structure d'un seul domaine protéique.

Le code source d'*AlphaFold-2* a été rendu public en 2021 et quelques mois plus tard est apparue la version *AlphaFold-Multimer*, qui avait été spécifiquement entraînée pour prédire le cas difficile des complexes protéine-protéine [24]. Dans la foulée, une base de données des structures prédites nommée *AlphaFold DB*<sup>2</sup> a été créée par *Google DeepMind* et l'EMBL-EBI (*EMBL's european bioinformatics institute*), afin de couvrir l'ensemble des protéines de séquence connue et ainsi accélérer la recherche. Cette bibliothèque contient actuellement près de 200 millions d'entrées. L'amélioration de la précision des structures prédites bénéficie, par rebond, aux méthodes expérimentales telles que la biocristallographie. Cette dernière peut en effet utiliser des modèles générés par *AlphaFold* dans la méthode du

remplacement moléculaire<sup>3</sup>, afin de résoudre avec plus de succès le problème des phases des données de diffraction mesurées [8].

*AlphaFold-2* s'appuie sur l'utilisation de réseaux neuronaux profonds, dont le module central nommé *Evoformer* est à même d'analyser les données d'entrée (telles que la séquence de la protéine à modéliser, l'alignement de séquences multiples, parfois une empreinte structurale) pour prédire les distances et les angles entre paires d'acides aminés. La qualité de la modélisation est évaluée par différents scores, qui sont : 1) le pLDDT (*predicted local distance difference test*), un score de confiance local par résidu ; 2) le PAE (*predicted aligned error*), une métrique par paire indiquant l'erreur attendue dans la distance entre deux résidus lorsqu'ils sont alignés sur la structure prédite ; 3) le DockQ [25], une mesure globale de qualité pour les modèles multimériques protéine-protéine.

*AlphaFold-3* [26] est paru en 2024 et son code source a été divulgué peu après. Son architecture a été considérablement refondue, lui permettant de prédire les structures de protéines seules ou en complexe avec des acides nucléiques, des petites molécules, des ions, et d'inclure des résidus modifiés [27] (→).

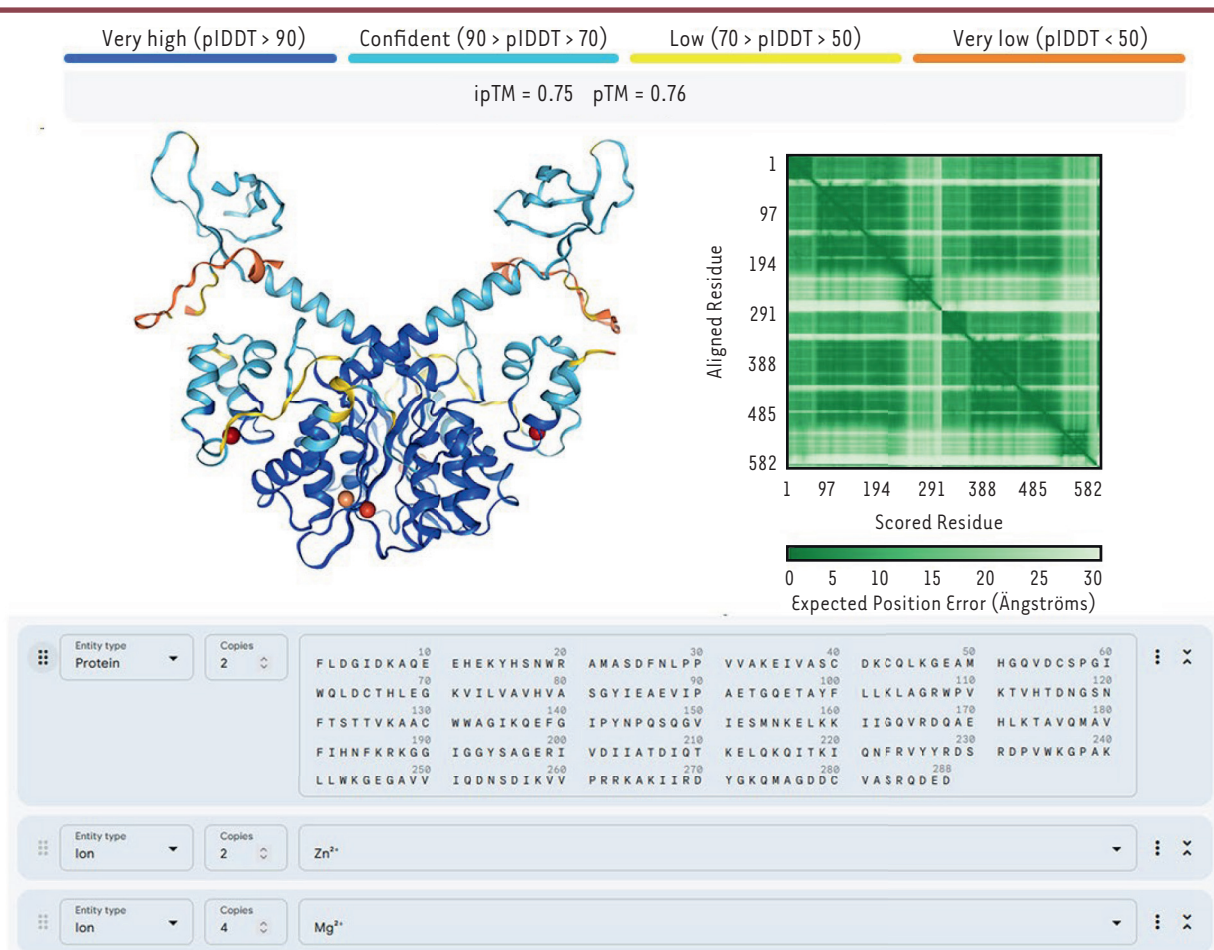
Le module *Pairformer* a remplacé l'*Evoformer*, permettant une meilleure modélisation des complexes biomoléculaires. Le module *Diffusion* utilise, quant à lui, les prédictions du *Pairformer* pour progresser itérativement vers une représentation tridimensionnelle de la structure moléculaire, en partant d'un nuage d'atomes. Il est à noter que ce nouveau module fait preuve d'une précision nettement améliorée pour les interactions protéine-protéine, notamment anticorps-antigène. *AlphaFold-3* conserve les scores de confiance pLDDT et PAE, et introduit de nouvelles métriques, à savoir : 1) le pTM (*predicted template modeling score*), une mesure de confiance globale de l'organisation structurale ; 2) l'ipTM (*interface pTM-score*) qui est spécifique aux interactions entre protéines et permet d'évaluer la fiabilité des interfaces dans un complexe.

*AlphaFold-3*, qui peut être utilisé via le web<sup>4</sup>, permet d'obtenir des résultats de prédictions de structures en quelques minutes et offre un accès démocratisé à la modélisation (Figure 2). Cependant, cette nouvelle évolution d'*AlphaFold* présente encore des limites [26], notamment un manque de fiabilité de prédiction pour les complexes protéine-ARN ou le risque de produire des « hallucinations », en structurant des zones qui sont intrinsèquement désordonnées. Par ailleurs, *AlphaFold-3* ne prend pas en compte les aspects dynamiques, ce qui pose problème, par exemple, lors de l'étude des protéines multifonctionnelles qui peuvent adopter différentes conformations. De même, il reste difficile de prédire l'effet sur la structure d'une simple mutation. Toutefois, la création par *DeepMind* de l'outil *AlphaMissense*, qui associe l'IA d'*AlphaFold* à des bases de données médicales [28] (→), permet d'améliorer l'esti-

<sup>2</sup> <https://alphafold.ebi.ac.uk>

<sup>3</sup> Les détecteurs de rayons X ne peuvent mesurer que l'intensité des ondes diffractées. L'information de phase est perdue. Grâce à la méthode du remplacement moléculaire, la connaissance d'une structure proche permet de calculer des phases qui approchent les phases expérimentales manquantes. La connaissance des intensités mesurées et des phases calculées permet de résoudre la structure de la macromolécule biologique cristallisée.

<sup>4</sup> <https://alphafoldserver.com>



**Figure 2. Structure homodimérique de l'intégrase (IN) du virus VIH-1 (en haut à gauche) prédite par AlphaFold-3 à partir de sa séquence et d'ions liés (panneau bleu en bas).** La protéine est représentée en mode ruban et est colorée en fonction de son score de confiance local pIDDT. Cette protéine modulaire de 288 résidus est constituée de trois domaines : un domaine amino-terminal (NTD, 1-47) un domaine central catalytique (CCD, 59-202) et un domaine carboxy-terminal (CTD, 220-270) [39]. Le NTD lie un ion Zn<sup>2+</sup> et le CCD deux ions Mg<sup>2+</sup>. Les structures des fragments NTD-CCD et CCD-CTD ont été résolues expérimentalement sous forme d'homodimère pour l'IN seule. La structure prédite de l'IN homodimérique entière montre de bonnes statistiques globales avec un pTM (*predicted template modeling score*) de 0.76 (>0.5 attendu) et un ipTM (*interface pTM-score*) de 0.75 (>0.6 attendu). L'analyse du PAE (*predicted aligned error*) (en haut à droite) montre une bonne rigidité au sein de chaque domaine (zones vert sombre) ainsi qu'entre les domaines CCD des deux monomères. La position du CTD par rapport au CCD et au NTD est légèrement flexible (zone vert clair) au sein de chaque monomère. La boucle reliant le NTD au CCD et la queue carboxy-terminale (271-288) sont plus flexibles (zones quasi-blanche). Ceci est confirmé par les couleurs jaune et orange de ces deux segments sur la représentation de la protéine en mode ruban (score de confiance pIDTT localement bas à très bas).

mation de l'effet pathogène de mutations chez l'humain (classées comme bénignes, incertaines ou pathogènes) sans pour autant fournir la structure du mutant étudié [29].

### David Baker : le développement visionnaire de la conception de protéines synthétiques, guidée par Rosetta

En 2003, David Baker a marqué une avancée majeure en concevant artificiellement, avec l'aide du programme de prédiction de structure *ab initio* Rosetta, développé dans son équipe, la première protéine synthétique nommée TOP7 [30]. Cette protéine, conçue pour avoir des

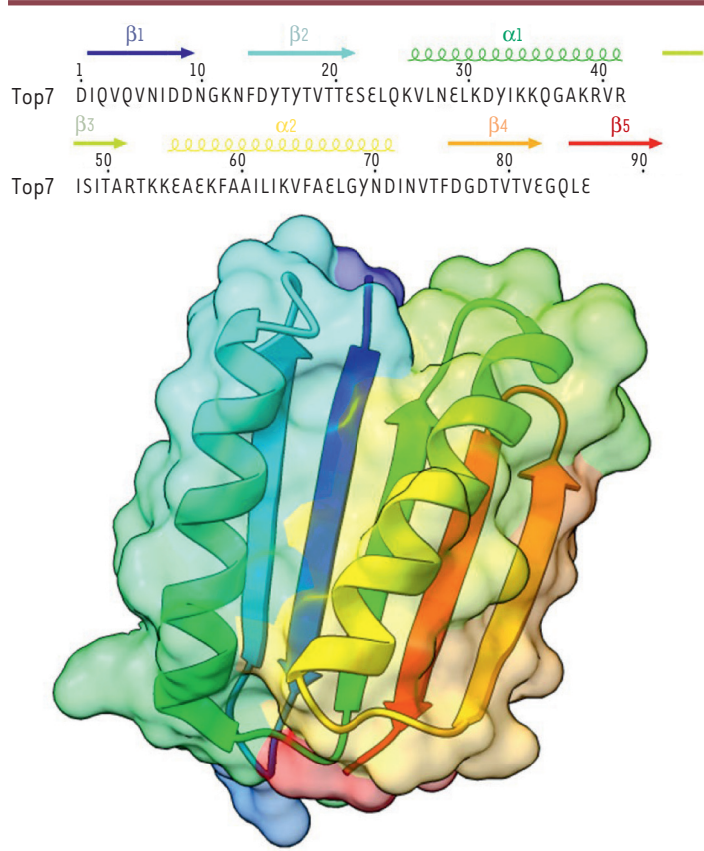
propriétés structurales et fonctionnelles spécifiques et uniques, se distingue par une séquence sans homologue connu et est capable de se replier de manière unique et stable (Figure 3). La détermination par cristallographie aux rayons X de la structure de cette protéine a montré que le modèle prédit était très proche de la structure expérimentale, validant ainsi la prédiction de Rosetta. En quoi cette avancée est-elle majeure ? Les protéines naturelles sont le fruit de millions d'années d'évolution convergeant vers leur adaptation à des besoins biologiques variés, définis par les contraintes exercées



par leur environnement. Ce processus, qui peut s'étaler sur des années, n'est pas compatible avec les défis contemporains tels que les pandémies, les maladies dégénératives ou la dégradation de l'environnement, qui exigent des solutions rapides et inédites. C'est dans ce contexte que les protéines conçues sur mesure trouvent tout leur intérêt. L'idée repose sur la modélisation du processus d'évolution naturel, mais orientée vers un besoin défini *a priori*. La conception computationnelle de protéines émerge ainsi comme un domaine de recherche innovant, avec le logiciel *Rosetta* comme outil emblématique [31]. Ce dernier s'appuie sur des algorithmes avancés et des données expérimentales, afin de prédire les structures possibles d'une protéine artificielle et d'ouvrir la voie à la création de nouvelles macromolécules adaptées à des applications précises. Au fil des années, *Rosetta* a évolué en intégrant des approches toujours plus performantes. En 2020, l'outil est passé de méthodes basées sur la minimisation d'énergie et les simulations de Monte Carlo<sup>5</sup> à l'apprentissage automatique et aux réseaux neuronaux [4]. Ces avancées lui permettent aujourd'hui d'explorer un espace de conception plus vaste et d'optimiser la stabilité ainsi que la fonctionnalité de ces protéines artificielles, accélérant de ce fait leur développement en vue de leur utilisation à des fins thérapeutiques ou industrielles.

Parmi les applications les plus remarquables se trouve le vaccin SKYCovione contre la COVID-19 [32]. Il est basé sur une nanoparticule protéique optimisée par conception computationnelle, qui est capable de présenter 60 copies du domaine RBD (*receptor binding domain*) de la protéine *Spike* du SARS-CoV-2 afin d'amplifier la réponse immunitaire. Approuvé en 2022, ce vaccin représente une première mondiale en validant l'utilisation pharmaceutique d'une protéine artificielle ; son développement met en lumière la capacité de cette technologie à répondre rapidement aux crises sanitaires globales.

Dans un autre domaine médical, des protéines conçues pour neutraliser les neurotoxines et cytotoxines du venin de cobras *Naja* [33] ont été développées en partenariat avec le *Jenkins Lab* (Université du Michigan, États-Unis). Ces protéines artificielles se lient à ces toxines pour les empêcher d'interagir avec leurs cibles après une morsure. Elles pourraient permettre de mieux



**Figure 3.** Séquence de la protéine TOP7 conçue *de novo* [30] et de ses éléments de structures secondaires avec une coloration arc-en-ciel de l'extrémité amino-terminale à l'extrémité carboxy-terminale (au-dessus de la séquence). La structure est représentée en-dessous en mode ruban et surface. Elle présente un repliement singulier de type  $\alpha/\beta$ , avec deux hélices  $\alpha$  adossées à un feuillet  $\beta$  composé de cinq brins antiparallèles.

traiter les populations rurales et le bétail des régions tropicales et subtropicales d'Afrique et d'Asie contre ce venin mortel.

Une autre avancée significative concerne une protéine ciblant les récepteurs de facteur de nécrose tumorale (*tumor necrosis factor*, TNF) [34], qui a démontré sa capacité à réduire l'inflammation de manière bien plus efficace que les traitements actuels tels que l'Enbrel®. Pour les pandémies futures, des anticorps entièrement conçus *de novo* ont été développés pour cibler les variants grippaux [35], offrant une réponse immunitaire robuste contre des souches mutantes. Cette avancée pourrait révolutionner les stratégies de lutte contre les virus émergents. En matière de biotechnologie, des nanopores conducteurs, c'est-à-dire soumis à un champ électrique, [36] conçus grâce à cette méthode permettent de détecter des molécules chimiques avec une précision exceptionnelle, ouvrant la voie à des applications avancées en diagnostic médical et en instrumentation industrielle. Enfin, des protéines artificielles capables de lier des chlorophylles [37] ont été développées pour améliorer l'efficacité de la capture de l'énergie lumineuse, une étape prometteuse vers des technologies de photosynthèse artificielle pour des énergies renouvelables.

<sup>5</sup> La méthode de Monte-Carlo ou simulation de probabilités multiples, est une technique mathématique utilisée pour estimer les résultats possibles d'un événement incertain.

La conception computationnelle de protéines, portée par les avancées en IA et en biologie synthétique, ouvre donc la voie à des développements inédits. L'amélioration continue des algorithmes IA et la baisse des coûts de synthèse de l'ADN promettent d'accélérer les avancées dans plusieurs domaines dont les thérapies géniques, la biologie synthétique et les technologies durables. Cependant, la capacité à créer des protéines inédites soulève des questions cruciales de biosécurité et d'éthique et la possibilité de synthétiser des biomolécules dangereuses, par exemple des toxines, intentionnellement ou par accident, requiert une vigilance accrue. Les experts recommandent ainsi de tracer toutes les séquences génétiques synthétisées dans des bases de données sécurisées. Une collaboration internationale, soutenue par des initiatives comme l'*international gene synthesis consortium*, est essentielle pour établir des normes et prévenir les abus tout en préservant l'innovation scientifique.

## Conclusion

L'IA a bouleversé la biologie structurale et permet aujourd'hui de prédire des structures de macromolécules biologiques avec un degré de confiance sans précédent. Sa facilité d'utilisation rend ce domaine de la biologie accessible à une communauté scientifique élargie. Les logiciels actuels (*AlphaFold-3* [26], *RoseTTAFold-2* [4], *Boltz-1* [38]...) restent toutefois encore tributaires de l'augmentation de la quantité de données expérimentales pour progresser, tout en suivant l'amélioration des méthodes d'apprentissage profond.

La conception assistée par ordinateur des protéines synthétiques requiert, quant à elle, une réflexion éthique et une régulation rigoureuse afin d'en garantir une utilisation responsable. Ce domaine est susceptible de transformer profondément la médecine et les biotechnologies à travers la création de nouvelles biomolécules et stratégies thérapeutiques, et offre d'importantes perspectives dans le domaine biomédical. ♦

## SUMMARY

### Nobel Prize in chemistry 2024: the revolution of artificial intelligence in structural biology

The 2024 Nobel Prize in chemistry has been awarded to Demis Hassabis and John M. Jumper (*Google DeepMind*) for the development of artificial intelligence-guided protein structure prediction and to David Baker (University of Washington, Seattle, USA) for the development of computational protein design. This event marks a revolution in the field of structural biology that has led, among other things, to the generation of a library of nearly 200 million predicted protein structures designed to speed up research. This revolution in AI has also led to the design of several artificial proteins of medical interest presented in this review. ♦

## REMERCIEMENTS

Nous remercions Xavier Robert, laboratoire MMSB, pour sa relecture critique de l'article.

## LIENS D'INTÉRÊT

Les auteurs déclarent n'avoir aucun lien d'intérêt concernant les données publiées dans cet article.

## RÉFÉRENCES

1. Wang H, Fu T, Du Y, et al. Scientific discovery in the age of artificial intelligence. *Nature* 2023 ; 620 : 47-60.
2. Li B, Gilbert S. Artificial Intelligence awarded two Nobel Prizes for innovations that will shape the future of medicine. *NPJ Digit Med* 2024 ; 7 : 336.
3. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021 ; 596 : 583-9.
4. Baek M, DiMaio F, Anishchenko I, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 2021 ; 373 : 871-6.
5. Schulz GE, Schirmer RH. *Principles of Protein Structure*. Springer-Verlag Berlin and Heidelberg GmbH & Co. K, 2<sup>e</sup> édition, 2013 : 334 p.
6. Watson JD, Crick FH. Molecular structure of nucleic acids ; a structure for deoxyribose nucleic acid. *Nature* 1953 ; 171 : 737-8.
7. Kendrew JC, Bodo G, Dintzis HM, et al. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* 1958 ; 181 : 662-6.
8. Le Du M-H, Legrand P, Sirigu S, Ravy S. *Introduction à la cristallographie biologique*. Les Ulis : EDP sciences, 2021 : 180 p.
9. Tabet JC, Rebuffat S. Prix Nobel de chimie 2002 pour la spectrométrie de masse et la résonance magnétique nucléaire. *Med Sci (Paris)* 2003 ; 19 : 865-72.
10. Neumann E, Farias Estrozi L, Effantin G, et al. La révolution de la résolution en cryo-microscopie électronique. *Med Sci (Paris)* 2017 ; 33 : 1111-7.
11. Schuller JM, Falk S, Fromm L, et al. Structure of the nuclear exosome captured on a maturing preribosome. *Science* 2018 ; 360 : 219-22.
12. Romby P, Marzi S, Westhof E. La structure atomique du ribosome en pleine lumière. *Med Sci (Paris)* 2009 ; 25 : 977-81.
13. Burley SK, Berman HM, Kleywegt GJ, et al. Protein Data Bank (PDB): The Single Global Macromolecular Structure Archive. *Methods Mol Biol* 2017 ; 1607 : 627-41.
14. Dobson CM. Protein folding and misfolding. *Nature* 2003 ; 426 : 884-90.
15. Bryngelson JD, Onuchic JN, Succi ND, Wolynes PG. Funnel, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* 1995 ; 21 : 167-95.
16. Lee J, Freddolino P, Zhang Y. Ab initio protein structure prediction. In : Rigden DJ, ed. *From Protein Structure to Function with Bioinformatics* : Springer-London, 2017 : 3-35.
17. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE. How fast-folding proteins fold. *Science* 2011 ; 334 : 517-20.
18. Bowie JU, Luthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science* 1991 ; 253 : 164-70.
19. Krieger E, Nabuurs SB, Vriend G. Homology modeling. *Methods Biochem Anal* 2003 ; 44 : 509-23.
20. Waterhouse A, Bertoni M, Bienert S, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* 2018 ; 46 : W296-W303.
21. Fiser A, Sali A. Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol* 2003 ; 374 : 461-91.
22. Jordan B. AlphaFold : un pas essentiel vers la fonction des protéines. *Med Sci (Paris)* 2021 ; 37 : 197-200.
23. Kryshtafovich A, Schwede T, Topf M, et al. Critical assessment of methods of protein structure prediction (CASP) - Round XIII. *Proteins* 2019 ; 87 : 1011-20.
24. Evans R, O'Neill M, Pritzel A. Protein complex prediction with AlphaFold-Multimer. *bioRxiv* 2021.
25. Basu S, Wallner B. DockQ: A Quality Measure for Protein-Protein Docking Models. *PLoS One* 2016 ; 11 : e0161879.
26. Abramson J, Adler J, Dunger J, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 2024 ; 630 : 493-500.
27. Taly A, Verger A. Prédiction de structures biomoléculaires complexes par AlphaFold 3. *Med Sci (Paris)* 2024 ; 40 : 725-7.
28. Jordan B. Les secrets des variants. *Med Sci (Paris)* 2023 ; 39 : 981-3.
29. Cheng J, Novati G, Pan J, et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science* 2023 ; 381 : eadg7492.
30. Kuhlman B, Dantas G, Ireton GC, et al. Design of a novel globular protein fold with atomic-level accuracy. *Science* 2003 ; 302 : 1364-8.
31. Leaver-Fay A, Tyka M, Lewis SM, et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol* 2011 ; 487 : 545-74.
32. Arunachalam PS, Feng Y, Ashraf U, et al. Durable protection against the SARS-CoV-2 Omicron variant is induced by an adjuvanted subunit vaccine. *Sci Transl Med* 2022 ; 14 : eabq4130.

33. Torres SV, Valle MB, Mackessy SP, *et al.* De novo designed proteins neutralize lethal snake venom toxins. *Res Sq* 2024.
34. Glogl M, Krishnakumar A, Ragotte RJ, *et al.* Target-conditioned diffusion generates potent TNFR superfamily antagonists and agonists. *Science* 2024 ; 386 : 1154-61.
35. Bennett NR, Watson JL, Ragotte RJ, *et al.* Atomically accurate de novo design of single-domain antibodies. *bioRxiv* 2024.
36. Vorobieva AA, White P, Liang B, *et al.* De novo design of transmembrane beta barrels. *Science* 2021 ; 371.
37. Ennist NM, Wang S, Kennedy MA, *et al.* De novo design of proteins housing excitonically coupled chlorophyll special pairs. *Nat Chem Biol* 2024 ; 20 : 906-15.
38. Wohlwend J, Corso G, Passaro S, *et al.* Boltz-1 Democratizing Biomolecular Interaction Modeling. *bioRxiv* 2024.
39. Rocchi C, Gouet P, Parissi V, Fiorini F. The C-Terminal Domain of HIV-1 Integrase: A Swiss Army Knife for the Virus? *Viruses* 2022 ; 14 : 1397.
40. Benani A, Messas E. John J. Hopfield et Geoffrey E. Hinton : de Hopfield et Hinton à AlphaFold : le Prix Nobel 2024 récompense les pionniers de l'apprentissage profond. *Med Sci (Paris)* 2025 ; 41 : 277-80.

## TIRÉS À PART

P. Gouet

REPÈRES



NOBEL 2024

**m/s**  
médecine/sciences

**Avec m/s, vivez en direct  
les progrès et débats  
de la biologie et de la médecine**

CHAQUE MOIS / AVEC LES ARTICLES DE RÉFÉRENCE DE M/S  
CHAQUE JOUR / SUR [WWW.MEDECINESCIENCES.ORG](http://WWW.MEDECINESCIENCES.ORG)

Abonnez-vous sur  
**[www.medecinesciences.org](http://www.medecinesciences.org)**