

accessibles au logiciel, mais pas leur orientation les uns par rapport aux autres.

Les segments de protéines qui sont naturellement désordonnés dans leur état, qui correspondraient à plus d'un tiers de tous les acides aminés du protéome humain, constituent un défi encore plus grand. Aucun algorithme ne peut actuellement prédire comment ils se replient, mais John Jumper note que la méthode des pLDDT peut aider à délimiter ces segments dans une structure.

Une caractéristique inattendue d'AlphaFold2 et de RoseTTAFold est leur capacité à prédire les structures des homodimères, ces complexes formés de deux protéines identiques, et des hétérodimères, quand les protéines assemblées diffèrent. Arne Elofsson et son équipe ont réussi à modéliser jusqu'à 59% des complexes à deux protéines qu'ils ont analysés avec AlphaFold2. Cependant, ce processus est plus difficile sur le plan informatique lorsque l'on tente d'identifier des complexes probables à partir de zéro, plutôt qu'en modélisant des paires d'interactions connues. Mais en appliquant plusieurs algorithmes d'apprentissage profond en parallèle, David Baker et son équipe ont identifié et modélisé des centaines de complexes multiprotéiques à partir de millions de paires d'interactions possibles dans le protéome de la levure *Saccharomyces cerevisiae*. Ici, pour filtrer les paires d'acides aminés qui interagissent le plus probablement, RoseTTAFold était environ cent fois plus rapide qu'AlphaFold2, celui-ci ayant ensuite été utilisé sur le sous-ensemble obtenu.

Sentant l'enthousiasme pour ce domaine, DeepMind a publié en octobre 2021 AlphaFold-Multimer, spécifiquement entraîné pour s'attaquer aux complexes protéiques. Le programme a prédit des interactions avec une haute précision pour 34% des complexes homodimériques testés, et pour 23% des complexes hétérodimériques.

FRONTIÈRES FONCTIONNELLES

De nombreuses questions restent encore hors de portée, note Debora Marks. «Si votre technologie copie la cristallographie, c'est formidable, mais de tels instantanés structuraux statiques ne disent rien du comportement dynamique d'une protéine donnée.» Ainsi, AlphaFold2 produit généralement une seule réponse «correcte» pour chaque séquence. Toutefois, de nombreuses protéines ont plusieurs conformations, toutes pertinentes pour la fonction, par exemple l'état activé ou inhibé d'une enzyme. Pour résoudre ces problèmes, des

techniques expérimentales qui montrent la structure de la protéine réelle dans de multiples états, comme la cryo-EM, sont encore nécessaires.

AlphaFold2 n'est pas non plus adapté pour prédire comment les changements d'acides aminés individuels modifient la structure des protéines, un facteur pourtant crucial pour comprendre l'effet des mutations. Une raison en est que l'algorithme converge vers une solution correcte à partir de nombreuses séquences légèrement différentes, explique Amelie Stein. Néanmoins, son équipe a obtenu des prédictions plus précises en couplant l'apprentissage profond à d'autres algorithmes d'analyse des mutations.

Les biologistes structuraux ne sont pas près de perdre leur emploi. En fait, ils sont même désormais en mesure de consacrer plus de temps à d'autres questions fondamentales. Randy Read, de l'université de Cambridge, au Royaume-Uni, note, par exemple, que les prédictions d'AlphaFold2 aident déjà beaucoup les cristallographes dans l'interprétation de leurs données lorsqu'elles sont incomplètes. Les concepteurs de protéines *de novo* tireront également profit d'AlphaFold2 en simplifiant la phase de calcul des modèles qui précède les tests en laboratoire.

Pour Mohammed AlQuraishi, la biologie structurale est à l'aube d'une nouvelle ère qui privilégiera la fonction des protéines plutôt que leur forme. Cette dernière ne sera plus une fin en soi, et la discipline va mériter la composante «biologie» de son nom.

— L'auteur —

> **Michael Eisenstein**
est journaliste scientifique,
basé à Philadelphie.

Cet article a paru sous le titre
«Artificial intelligence powers
protein-folding predictions»
sur le site Nature.com,
le 23 novembre 2021.

— À lire —

> **J. Jumper et al.**, Highly accurate protein structure prediction with AlphaFold, *Nature*, 2021.

> **K. Tunyasuvunakool et al.**, Highly accurate protein structure prediction for the human proteome, *Nature*, 2021.

> **W. Zheng et al.**, Protein structure prediction using deep learning distance and hydrogen-bonding restraints in CASP14, *Proteins*, 2021.

> **I. Humphreys et al.**, Computed structures of core eukaryotic protein complexes, *Science*, 2021.

Les agents jouent collectif

Simuler sur ordinateur la nage d'un banc de poissons, et sa réaction à une attaque de prédateurs, est désormais possible en le traitant comme un système multi-agents.

34

© Paul Cowel/shutterstock.com

Programmer un système multi-agents est
comme créer une fourmilière : chaque entité possède
une petite intelligence, mais leur interaction et leur
coordination en engendre une nouvelle forme.
Elle est étonnante et parfois d'une efficacité déconcertante.

Philippe Mathieu, Sébastien Picault et Jean-Christophe Routier



Au terme des huit saisons et soixante-treize épisodes de *Game of Thrones*, le spectateur, éberlué, pourrait se livrer à un rapide décompte : combien de cavaliers, de fantassins, de spectres et autres Marcheurs blancs a-t-il vu défiler, combattre, périr ? Heureusement pour la production, depuis la trilogie du *Seigneur des anneaux*, récompensée aux Oscars, les scènes de groupe au cinéma ou dans les séries sont de plus en plus confiées à l'ordinateur, qui dirige des acteurs artificiels grâce à des algorithmes.

En fait, l'attribution de règles de comportements et d'objectifs à une multitude d'entités informatiques autonomes s'applique à de nombreuses situations. Avec les mêmes principes, on mime des phénomènes naturels, on anime des mondes artificiels à des fins ludiques, on teste l'aménagement du territoire, on étudie des comportements sociaux. Toutes ces applications sont fondées sur un schéma commun : une modélisation centrée sur les individus au moyen de la simulation par agents.

Rappelons que pour étudier un phénomène, les scientifiques aiment à en construire une représentation simplifiée, qu'ils appellent un « modèle ». Les orbites elliptiques des planètes de Kepler sont par exemple un modèle réaliste de leurs trajectoires autour du Soleil. Cependant, les modèles que l'on peut soumettre directement à l'expérience deviennent de plus en plus rares. On a alors recours à des simulations qui testent le modèle sur un substitut du réel. Aujourd'hui, la plupart des simulations consistent à résoudre de façon exacte ou approchée des équations qui sous-tendent les modèles physiques. Cette approche mathématique permet, par exemple, de modéliser l'évolution

probable du climat à l'échelle planétaire ou d'une épidémie à l'échelle d'une population suffisamment homogène, de même que la régulation génétique au sein d'une cellule.

Mais une nouvelle façon d'envisager la simulation a émergé, au carrefour des recherches sur la vie artificielle (quand on laisse des ensembles d'entités virtuelles se développer et s'organiser) et des travaux sur la simulation des interactions d'individus. Le concept d'agents a repoussé les frontières de la modélisation au-delà des limites inhérentes à la seule mise en équations, notamment en fournissant des outils pour étudier les comportements en tant que tels. C'est ce qu'on appelle l'« intelligence collective ».

SE RECENTRER SUR L'INDIVIDU

En effet, l'une des faiblesses de l'approche mathématique de la simulation est son absence de portée explicative. On constate un comportement, telle une convergence vers un état, mais on ignore le détail des mécanismes en jeu. Les équations qui régissent le modèle portent sur le comportement moyen des individus et négligent les variations de comportement d'une entité à une autre. Le devenir d'un individu particulier et sa contribution à l'ensemble du phénomène restent inconnus.

À l'inverse, les modélisations dites « centrées sur des individus » sont focalisées sur les moyens plutôt que sur les fins. Elles posent le problème à l'échelle de l'individu : étant donné les entités qui interviennent dans un phénomène, quel doit être leur comportement pour que ce phénomène se manifeste ? Ce type de modélisation, qui permet de passer d'une simulation macroscopique à une