

2. Le jeu *Foldit*



Un joueur dispose à chaque instant d'une information sur le score (c'est-à-dire l'énergie qu'il faut minimiser) de la configuration moléculaire affichée. Pour améliorer ce score, le joueur a le choix entre plusieurs types d'actions sur la molécule à replier : saisir un élément de la molécule et le déplacer ; secouer (le squelette, les chaînes, ou tout) pour voir si cela améliore le score ; disposer quelques élastiques ou en enlever, ce qui aide à mieux contrôler la structure, etc. Un cahier de « recettes » (à gauche sur le côté) est aussi à sa disposition : chaque recette est une succession d'actions prédéfinies dont l'exécution a été reconnue efficace.

Étonnamment, les joueurs et les équipes de joueurs réussissent souvent à mieux replier les molécules que les meilleurs algorithmes spécialisés.

Foldit wiki

dans une fonction donnée. En janvier 2012, une enzyme qui catalyse la réaction dite de Diels-Alder a ainsi été perfectionnée, augmentant sensiblement son efficacité. Selon Seth Cooper, coordinateur de ce travail, « les gens possèdent des facultés naturelles de raisonnement spatial que les ordinateurs n'ont pas encore ».

Découvrir aussi des algorithmes

Récemment, les mécanismes de jeu du programme *Foldit* ont été complétés pour que les joueurs ou groupes de joueurs puissent définir des méthodes de repliement générales, autrement dit des algorithmes, et puissent les échanger, les copier et les faire évoluer. Ces « recettes » évolutives et en compétition ont amené de fécondes interactions et conduit à 5 400 propositions résultant de combinaisons de recettes élémentaires et de modifications variées des meilleures recettes trouvées précédemment. Deux recettes eurent un succès particulier auprès des joueurs. On remarqua alors des similitudes entre les deux méthodes et un algorithme récemment mis au point par des spécialistes de la programmation du repliement des protéines. Cet algorithme dépasse les meilleurs algorithmes publiés.

Les auteurs de l'article où est décrite cette découverte concluent que « les jeux scientifiques en ligne, non seulement résolvent des problèmes difficiles, mais

découvrent et formalisent des algorithmes nouveaux et efficaces ».

Ainsi, des joueurs utilisant leurs facultés naturelles de manipulations d'objets en trois dimensions et leur bon sens battent les meilleurs algorithmes et en conçoivent de nouveaux d'une complexité et d'une efficacité équivalentes à ceux que les spécialistes imaginent. Certes, les joueurs qui ont participé à ces expériences sont nombreux et, pris dans l'excitation de la compétition, certains ont consacré beaucoup de temps, d'énergie et d'ingéniosité pour arriver à ces exploits.

On ne peut s'empêcher de penser que les milliards d'heures perdues dans le monde à jouer à toutes sortes de jeux informatiques deviendraient utiles si plus de jeux scientifiques de cette nouvelle catégorie étaient pratiqués... Zoran Popovic, le directeur du *Center for Game Science*, énonce un objectif audacieux : « Notre but ultime est que toute personne puisse jouer et soit donc en mesure de gagner un prix Nobel. »

Un autre problème difficile de biologie computationnelle est celui de l'alignement multiple des séquences de molécules biologiques de type ADN, ARN ou protéines. On cherche à disposer ces séquences les unes sous les autres, de manière à ce qu'apparaissent les régions similaires et que ces zones de concordance soient, autant que possible, les unes sous les autres. On utilise des programmes qui maximisent le nombre de coïncidences entre les séquences. L'introduction de trous (c'est-à-dire de ruptures

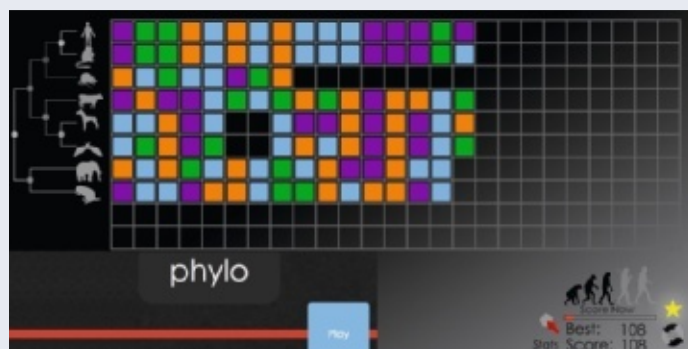
dans les séquences) est permise. Le résultat final est représenté par un tableau noté de la manière suivante : [a] deux séquences ayant un élément en commun au même endroit confèrent un point ; [b] si, en un endroit donné, deux séquences diffèrent, cela retire un point à la note du tableau ; [c] l'introduction d'un trou coûte cinq points et l'agrandissement d'un trou déjà existant en fait perdre un.

Les règles étant fixées, le calcul du meilleur alignement ne serait qu'une routine : en essayant tous les alignements possibles (en nombre fini) et en les notant tous, on déterminerait celui obtenant la meilleure note. Malheureusement, ce problème simple à énoncer est « complexe » à résoudre et aucun moyen algorithmique efficace n'est connu, car il appartient à la classe des problèmes NP-complets, classe de problèmes que personne ne sait traiter en temps raisonnable quand la taille des données augmente.

L'alignement des séquences est important pour déterminer la fonction des protéines et la phylogénie. L'alignement sert à identifier les zones fonctionnelles des protéines qui sont bien conservées lors de l'évolution et correspondent donc aux parties bien alignées des tableaux. Lorsqu'on détecte l'homologie d'une protéine A avec une protéine B de fonction connue, l'alignement suggère les fonctions de la protéine A.

Une seconde utilité des alignements multiples est liée à la recherche des parentés

3. Le jeu *Phylo*



Le programme *Phylo* transforme en jeu le problème de l'alignement de séquences génétiques. Le joueur déplace des cases colorées pour obtenir le plus grand nombre de couleurs identiques en colonnes. Des grands alignements sont examinés par le programme qui y repère des zones susceptibles d'être mieux alignées. Ces zones sont traduites en puzzles colorés.

On présente au participant l'un de ces problèmes en réserve. S'il trouve un meilleur alignement, celui-ci est inséré dans le grand alignement dont a été extraite l'énigme. De nombreux alignements ont ainsi été perfectionnés grâce aux joueurs qui, comme dans le cas de *Foldit*, obtiennent par leurs capacités visuelles et leur intelligence des résultats supérieurs à ceux des algorithmes.

entre espèces : quelle espèce a un ancêtre commun proche avec telle autre, et comment les espèces ancestrales se disposent-elles en arbre [équivalent aux arbres généalogiques] ? Un alignement multiple entre éléments d'une famille de protéines aide à calculer l'arbre phylogénétique le plus probable entre les protéines, qui est aussi un arbre reliant les organismes porteurs des protéines alignées. Les phylogénies ainsi reconstituées sont d'autant plus exactes que les alignements sont précis et optimaux. Toutes sortes d'algorithmes sont utilisés, complétés, perfectionnés et mis en commun par les chercheurs pour calculer des alignements multiples et en déduire des arbres phylogénétiques.

Alexander Kawrykow et ses collègues du Centre de bio-informatique de l'Université McGill à Montréal, inspirés par *Foldit*, ont mis au point un programme interactif nommé *Phylo* (<http://phylo.cs.mcgill.ca>) qui transforme les problèmes d'alignements multiples en jeux et propose à toute personne, même si elle n'a aucune connaissance de biologie, de rechercher de bonnes configurations. Des tableaux colorés sont affichés à l'écran et les joueurs les manipulent pour rechercher des alignements. Le programme indique à chaque instant la note du tableau affiché qu'il s'agit de faire progresser en déplaçant de petits carrés, dont il est inutile de savoir ce qu'ils représentent.

Encore plus facile à utiliser que *Foldit*, le programme *Phylo* s'est révélé d'une grande utilité. Les problèmes d'alignements soumis

aux utilisateurs concernent des séquences de taille limitée de façon à ce que l'œil humain s'y repère bien. Si un problème soumis à un joueur est résolu et obtient une note meilleure que celle de l'alignement connu au préalable (qui a été calculé par un algorithme rapide, produisant des alignements de qualité moyenne), alors l'alignement du joueur remplace celui qui était jusqu'alors enregistré dans la base de données.

Les organisateurs du jeu affirment que 70 pour cent des alignements soumis aux joueurs ont été améliorés et ont donc été remplacés par les résultats des joueurs humains. Après six mois de mise en place, le jeu avait attiré plus de 12 000 internautes qui avaient produit plus de 300 000 alignements variés. Comme le signalent les chercheurs de l'Université McGill, cela prouve à nouveau que le talent visuel des joueurs en ligne dépasse celui des algorithmes connus.

Une méthode d'avenir ?

Le programme *Phylo* et son usage sont gratuits. Un biologiste qui souhaite en bénéficier afin d'améliorer un alignement important pour ses travaux l'utilisera très facilement. Il suffit pour cela qu'il envoie l'alignement à améliorer à une adresse électronique donnée en respectant un format convenu ; le tableau sera transformé en jeu graphique et si les joueurs améliorent l'alignement initial, la solution graphique trouvée sera retraduite dans les notations

des biologistes, puis insérée là où il faut dans le grand alignement soumis, en même temps qu'il sera envoyé au chercheur à l'origine de la demande (voir la figure 3). Les méthodes de *Foldit* et de *Phylo* n'ont-elles d'intérêt que temporairement, ou permettront-elles longtemps à l'homme de battre la machine ?

Il paraît raisonnable de soutenir que dès qu'on aura mieux compris les problèmes, ou analysé comment travaille le cerveau humain et qu'on l'aura copié, ou encore dès qu'on disposera d'une puissance de calcul supérieure, alors le talent humain cessera de faire le poids face à la machine électronique, comme cela s'est produit pour le jeu d'échecs. *Foldit* et *Phylo* seront alors sans intérêt.

À l'inverse, on peut argumenter que les circuits neuronaux présents dans les aires visuelles du cortex sont le résultat d'une optimisation très fine après des dizaines ou centaines de millions d'années d'évolution ; en conséquence, ces circuits fonctionnent d'une manière extraordinairement subtile qui restera longtemps inaccessible à notre compréhension. Autrement dit, nous ne saurons pas programmer leurs mécanismes dans le silicium avant des lustres. En allant plus loin, on peut même imaginer qu'il sera impossible de comprendre dans le détail ce que font notre cerveau ou certaines de ses parties car, pour les comprendre, il faudrait y pénétrer et donc les détruire... Il est ainsi impossible de conclure aujourd'hui sur l'avenir de *Foldit* et de *Phylo*.

Transformer en jeu un problème scientifique réel et difficile est une belle idée mise en œuvre depuis peu, mais une idée proche l'a anticipée. Un groupe de chercheurs peut demander de l'aide à l'extérieur pour réaliser un travail qu'il n'arrive pas à effectuer seul, car cela exige trop de temps ou d'énergie. C'est ce qui s'est passé avec le projet *Galaxy Zoo*.

Des millions de galaxies

Pour comprendre la formation et l'évolution des galaxies, il est nécessaire de les classer selon leur forme et de disposer de nombreux exemples. Une classification sommaire distingue les galaxies spirales et les galaxies elliptiques. Un spécialiste disposant d'une image de galaxie indiquera assez rapidement son type et en classera quelques centaines sans se donner beaucoup de mal. La difficulté provient du nombre de galaxies que les chercheurs veulent prendre en compte (plusieurs millions), car il est important pour eux de

disposer de classifications aussi exhaustives que possible. Une première version du projet *Galaxy Zoo* (<http://www.galaxyzoo.org>) a permis, grâce aux internautes volontaires, de classer un million de galaxies. Les images provenaient d'une caméra numérique installée à l'Observatoire d'Apache-Point, au Nouveau-Mexique (États-Unis). Sans la mobilisation des 80 000 bénévoles, la caractérisation des 60 millions de galaxies repérées par le télescope spatial *Hubble* aurait été impossible.

Par ailleurs, lors de ces travaux d'observation et de classification, un objet étrange a été identifié par une internaute, Hanny van Arkel. L'objet, nommé *Hanny's Voorwerp*, est mystérieux et a suscité discussions et conjectures chez les spécialistes, prouvant à nouveau le bien-fondé d'un examen aussi exhaustif que possible, par des humains, des données astronomiques aujourd'hui disponibles en quantité... astronomique.

Si l'intelligence humaine réussit souvent mieux à calculer ou à classer que la machine, c'est que le problème a été bien traduit et mis en images pour que la recherche de sa solution devienne amusante ou très simple (dans le cas de *Galaxy Zoo*) grâce à un logiciel interactif mis en ligne et conçu dans ce but. Cela est possible depuis que les performances graphiques des ordinateurs de bureaux ont atteint les extraordinaires niveaux actuels.

Une autre raison du succès de ces méthodes est l'existence des réseaux informatiques et de la communication instantanée qu'ils permettent entre les organisateurs du travail et une multitude de joueurs potentiels dispersés dans le monde. Grâce à ces progrès technologiques, une masse importante de joueurs tente sa chance et fournit un travail irremplaçable. L'idée est de confier à une « foule » une tâche de calcul qu'on ne sait pas obtenir des machines et qu'un humain seul a peu de chance de mener à bien. L'intelligence humaine bat l'ordinateur dans ces exemples parce qu'elle s'est massivement mobilisée et coordonnée.

Cette idée de s'adresser au grand nombre pour obtenir ce qu'on ne réussit pas seul ou avec ses machines n'est pas

4. Karpov et Kasparov battent la foule

L'idée de confier à la foule une tâche trop difficile pour un être humain, et dont les exemples présentés dans l'article montrent qu'elle fonctionne souvent à merveille, a pu donner l'illusion que le *crowdsourcing* était apte à tout faire.

Il n'en est rien : dans un match sur Internet en 1996, Anatoli Karpov écrasa son adversaire, « Le reste du monde ». Toutefois, les coups étaient décidés à la majorité et certains joueurs

faibles votaient pour les mauvais coups.

Le match qui a été organisé en 1999 entre Gary Kasparov et « 50 000 bons joueurs » était plus difficile et a prouvé à nouveau que la foule, dans un tel cas, se fait battre par un individu seul. Alors que 50 000 personnes provenant de 75 pays différents et conseillées par quatre joueurs professionnels, dont Étienne Bacrot, collaboraient pour jouer et battre G. Kasparov en une fantastique

partie d'échecs, ce fut lui qui, en définitive, l'emporta après des échanges étendus sur 62 coups (voir www.chessgames.com/perl/chessgame?gid=1252350). Puisque l'ordinateur *Deep Blue* abattu G. Kasparov aux échecs, et que G. Kasparov a battu la foule, doit-on en conclure que, nécessairement, *Deep Blue* ou l'un de ses successeurs battrait lui aussi la foule ?

A. Karpov rejoua en juillet 2012 une partie majoritaire et gagna en 37 coups (www.europe-echecs.com/art/parte-majoritaire-contre-karpov-fin-de-la-partie-majoritaire-de-karpov-4421.html).

Ni les échecs ni la science ne sont démocratiques : ce n'est pas la majorité qui gagne. En science, le novateur a raison contre les habitudes de pensée de tous.



akora.com