

Département de biochimie de la même université. Ce programme, que l'on télécharge facilement (<http://fold.it/portal/>), propose des énigmes de repliement. Une molécule dessinée en trois dimensions est affichée à l'écran, et le joueur peut alors la faire tourner, en examiner les parties de près et en manipuler les diverses composantes. Il est donc en mesure de la replier en essayant de faire au mieux. Le joueur est guidé par des informations diverses, dont bien sûr le niveau d'énergie qu'il doit minimiser. Il peut revenir en arrière, agiter aléatoirement la molécule, en figer certaines parties, etc. Il n'est pas nécessaire de connaître la biochimie pour tenter sa chance, car ce qui est

important sera appris par des exercices d'introduction et de découverte. Le but d'un joueur est d'améliorer l'énergie du meilleur repliement connu.

Le programme *Foldit*, qui fonctionne depuis 2008, a aujourd'hui été utilisé par 240 000 joueurs enregistrés. Les meilleures structures produites par les joueurs sont mémorisées, examinées et testées par les experts du domaine. Pour obtenir de bons résultats, les joueurs sont incités à se grouper en équipes. Des tableaux de meilleurs scores sont maintenus à jour, et diverses formes de compétitions entre joueurs et équipes accroissent les motivations des participants.

En 2011, une équipe de joueurs dénommée *Contenders* a donné la forme exacte de la protéase d'un rétrovirus responsable d'une des formes du sida chez le macaque rhésus. Il s'agissait d'un problème non résolu depuis une quinzaine d'années. La solution de l'équipe *Contenders* a été publiée dans la revue *Nature Structural and Molecular Biology*. Le nom de l'équipe apparaît comme auteur de la publication parmi les autres chercheurs ayant travaillé sur le projet *Foldit*. Le repliement découvert aidera peut-être à la mise au point de nouveaux antiviraux s'appliquant au sida humain.

Le jeu a aussi été utilisé pour modifier des protéines et améliorer leur efficacité

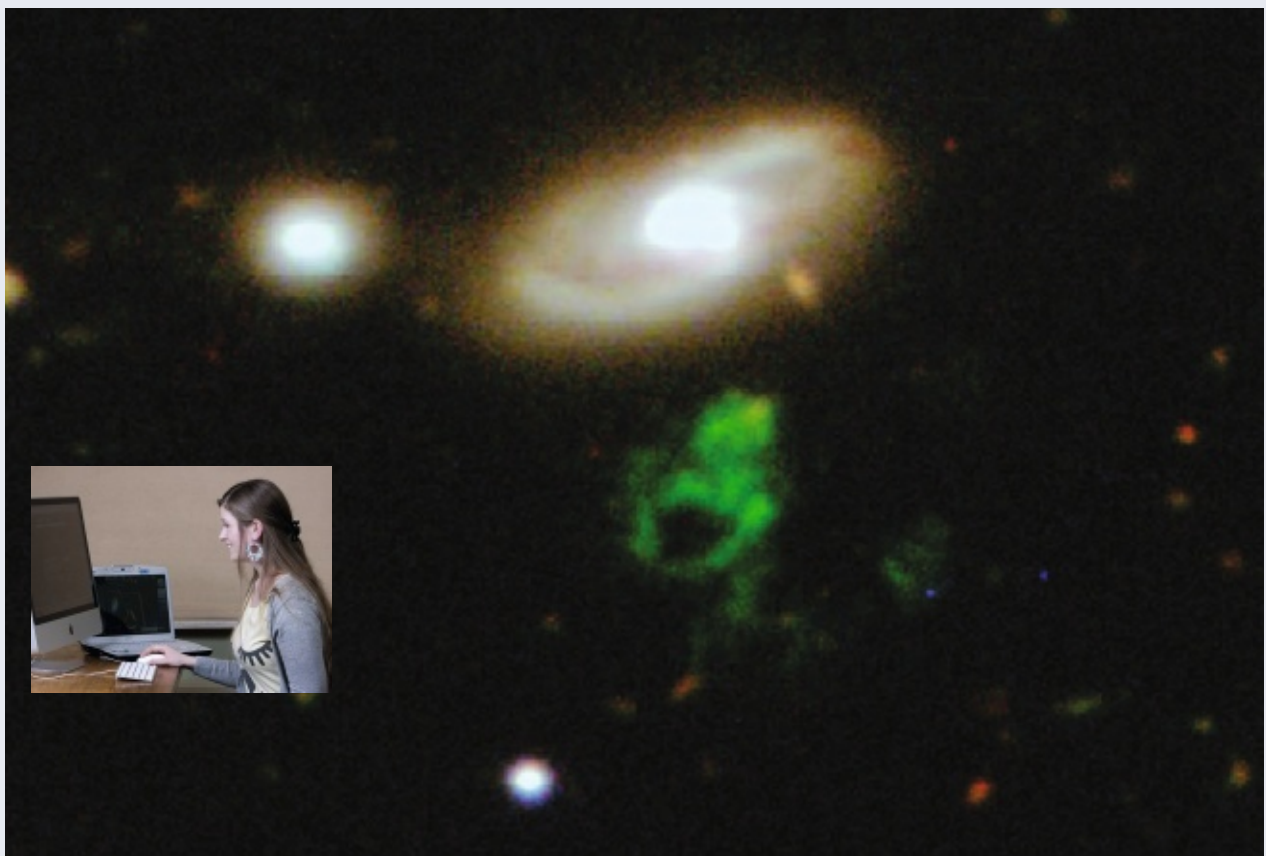
1. La grenouille de *Galaxy Zoo*

Les données recueillies par le télescope spatial *Hubble* sont trop nombreuses pour être traitées par les astronomes et trop complexes pour que la classification des galaxies soit

faite par un algorithme. Seul le recours à l'intelligence des bénévoles permet de mener le travail de classification nécessaire et rend disponibles aux chercheurs des données

scientifiques importantes (projet *Galaxy Zoo*). Durant ce travail de fouille des données portant sur des millions de galaxies, un étrange objet a été repéré par une internaute néerlandaise

prénommée Hanny (à gauche). L'objet, ici représenté en fausse couleur verte, a été nommé *Hanny's Voorwerp*. Il suscite conjectures et spéculations chez les astronomes.



Nasa et Université d'Arizona, *Galaxy Zoo Team*

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