

HUGH WOODIN ET LES GRANDS CARDINAUX

Pour Hugh Woodin (*ci-contre*), la hiérarchie des grands cardinaux (de grands ensembles infinis dont l'existence est posée par des axiomes complémentaires aux axiomes ZFC usuels de la théorie des ensembles) est aussi intrinsèque et absolue que celle des nombres entiers eux-mêmes. C'est selon lui une découverte majeure. Il est convaincu que les axiomes des grands cardinaux ne conduiront jamais à une contradiction, ce qui a pour conséquence que les énoncés qu'ils permettent de prouver sur les nombres entiers sont vrais, et que ces axiomes ont donc des conséquences concrètes, physiques. Il affirmait ainsi (dans « *The transfinite universe* », chapitre du livre collectif Kurt Gödel and the Foundations of Mathematics, paru en 2011) : « Dans les dix mille ans à venir, on ne trouvera aucune contradiction au sein de ces théories [affirmant l'existence de grands cardinaux]. Il s'agit d'une prédiction spécifique et non ambiguë sur l'univers physique. Elle ne se réduit à aucune vérité connue antérieurement. Cette prédiction découle du développement de la théorie des ensembles au cours des cinquante dernières années [...]. Elle se fonde sur ma conviction que la conception de l'univers transfini est pleinement sensée. Je fais cette prédiction indépendamment de toute spéculation sur les dispositifs informatiques susceptibles d'être développés au cours des dix mille prochaines années et qui pourraient avoir des effets sur l'efficacité de la recherche en mathématiques. »



➤ remis en cause, mais qu'elle est aussi le résultat d'une discussion et d'un affinement de notre conception de l'infini.

Cette discussion se nourrit à la fois de théorèmes et de points de vue qui renforcent notre intuition. Certaines options s'offrant aux mathématiciens sont ainsi éliminées non pas parce qu'on démontre qu'elles sont logiquement absurdes, mais parce qu'elles ne sont pas intellectuellement satisfaisantes – comme adopter brutalement l'axiome $V=L$. L'infini mathématique se montre intelligible et, comme pour d'autres sciences, nos connaissances dans ce domaine s'enrichissent d'année en année.

DES TRAVAUX ABSTRAITS AUX CONSÉQUENCES CONCRÈTES

Par ailleurs, considérant ces avancées de la théorie des ensembles depuis un demi-siècle, Hugh Woodin défend l'idée que la théorie des grands cardinaux et les travaux sur l'axiomatisation complémentaire de ZFC ne sont pas un travail purement abstrait, hors du monde. La compréhension nouvelle qu'ils donnent de l'infini a, selon lui, des « conséquences physiques ». Expliquons son point de vue.

Les axiomes de grands cardinaux se classent les uns par rapport aux autres et quand un axiome X est plus fort que l'axiome Y , alors X permet de prouver l'énoncé affirmant « le système $ZFC+Y$ est non contradictoire ». Or un tel énoncé s'exprime dans la théorie par une suite de symboles, et savoir s'il se

démontre n'est qu'une question combinatoire, que l'on sait toujours traduire en question portant sur les nombres entiers. Même si c'est étonnant, les axiomes de grands cardinaux ont donc des conséquences sur ce qu'on sait démontrer concernant les nombres entiers : l'infini joue sur le fini et plus on a d'infinis, plus on en sait sur le fini.

Ces effets sur les énoncés arithmétiques sont, d'une certaine façon, des effets physiquement testables. Penser qu'un axiome de grand cardinal X est acceptable – et Hugh Woodin clame que c'est le cas de beaucoup d'entre eux – signifie qu'en faisant fonctionner la théorie $ZFC+X$, on est certain qu'on ne tombera jamais sur une contradiction.

Or faire fonctionner une théorie, c'est dérouler les conséquences de ses axiomes, tâche que l'on peut confier à une machine. Croire en l'existence d'un grand cardinal, c'est donc croire que les machines utilisant le système ZFC complété par l'axiome correspondant à ce grand cardinal ne produiront jamais de résultats contradictoires tels que $0=1$. C'est donc un énoncé qui porte sur le monde physique. Il provient uniquement de notre compréhension de l'infini. En d'autres termes, les progrès de la théorie des ensembles enrichissent ce que nous pouvons affirmer de l'infini et donc du monde physique. Ainsi, ces avancées des mathématiciens et des logiciens tissent un fil qui relie l'infini le plus grand à la réalité matérielle, allant des objets les plus abstraits aux objets les plus concrets. ■

BIBLIOGRAPHIE

H. Woodin, **In search of ultimate-L**, *Bulletin of Symbolic Logic*, vol. 23(1), pp. 1-109, 2017.

J. Cavitt, **Set-theoretic geology, the ultimate inner model, and new axioms**, thèse soutenue à l'université Harvard, 2017.

C. Rittberg, **How Woodin changed his mind : New thoughts on the Continuum Hypothesis**, *Archive for History of Exact Sciences*, vol. 69(2), pp. 125-151, 2015.

Conférence filmée de Hugh Woodin, 2015 :
<https://www.youtube.com/watch?v=nVF4N1Ix5WI>

M. Heller et H. Woodin, **Infinity : New Research Frontiers**, Cambridge University Press, 2011.

H. Woodin, **Strong axioms of infinity and the search for V**, *Proceedings of the International Congress of Mathematicians 2010*, World Scientific, pp. 504-528, 2011.

DES SUJETS D'ACTUALITÉ, DES LECTURES CAPTIVANTES



Michio KAKU

Physicien et futurologue, cofondateur de la théorie des cordes.

1^{re} éd. 2019 • 448 p. • 9782807322332 • **24 €**

Les dernières avancées et perspectives sur le cerveau humain.



Jamie BARTLETT

Journaliste et expert britannique en médias sociaux et technologie.

1^{re} éd. 2019 • 224 p. • 9782807322301 • **19,50 €**

Un essai passionnant sur les enjeux de la révolution numérique



Christian de PERTHUIS

Créateur de la Chaire d'économie du climat à l'université Paris Dauphine.

1^{re} éd. 2019 • 256 p. • 9782807322097 • **19,50 €**

Une course contre la montre pour le climat



William NORDHAUS

Prix Nobel d'économie 2018 pour ses travaux sur l'économie du climat.

1^{re} éd. 2019 • 352 p. • 9782807325296 • **24,90 €**

Risques, incertitudes et solutions économiques face à un monde en réchauffement.



L'AUTEUR



JAMES E. DAHLGREN
maître de conférences dans le département
Wallace H. Coulter d'ingénierie biomédicale,
à l'institut de technologie de Géorgie et
à l'université Emory, à Atlanta, aux États-Unis

L'ESSENTIEL

> Les propriétés uniques de l'ADN le rendent idéal pour le stockage d'informations de toutes sortes, comme des livres ou des films.

> Dans l'ingénierie biomédicale, on l'utilise par ailleurs comme un code-barres permettant de réaliser une multitude d'expériences simultanément, ce qui permet de générer des données à des vitesses sans précédent.

> Cette technique offre la possibilité de tester plus rapidement et à moindre coût des pistes thérapeutiques, comme des traitements contre le cancer ou l'efficacité de vecteurs de médicaments.

Une vidéo stockée dans l'ADN

Un cheval au galop enregistré dans l'ADN d'une bactérie, c'est possible !
Dispositif de stockage numérique, outil de gestion de données lors d'expériences, code-barres... l'ADN endosse des rôles de plus en plus variés.

Des milliards d'années avant que les humains inventent le premier disque dur, l'information la plus précieuse – le patrimoine génétique – avait déjà son moyen de stockage : l'ADN. Au fil de l'évolution, toute forme connue de vie sur Terre s'est retrouvée dotée de cette molécule. Aujourd'hui, grâce aux récentes percées techniques qui nous permettent de lire et d'écrire l'ADN, des scientifiques voient dans cette vieille molécule un moyen de stocker d'autres types d'informations : celles produites à vitesse exponentielle à l'ère du *big data*.

L'idée de recourir à l'ADN pour stocker des informations autres que les gènes n'est pas neuve. Après tout, les 0 et les 1 du code informatique se heurtent aux limites de la physique. Un événement récent l'illustre : Myspace, un réseau social qui fut un jour l'un des plus populaires, a perdu les données d'une décennie au cours d'une migration de serveur. Assurer la protection des données à long terme, comme celles d'un site web qui redémarre après une longue période d'arrêt, est difficile quand les techniques employées sont vulnérables. En outre, les quantités d'énergie requises sont considérables.

Or l'ADN a le potentiel de contourner ces problèmes. Sa structure en double hélice est parfaitement adaptée au stockage d'information,

pour plusieurs raisons. D'abord, l'ADN est doté d'une très grande stabilité dans le temps : l'intégrité et la précision des informations sont donc garanties à long terme. Conservé dans un environnement froid et sec, l'ADN résiste plusieurs dizaines de milliers d'années, comme l'ont montré, en 2010, Svante Pääbo, à l'institut Max-Planck pour l'anthropologie évolutionniste, à Leipzig, et ses collègues en analysant le génome d'un spécimen de Néandertal à partir d'un os. Mais l'aspect le plus fascinant de la double hélice réside dans son exceptionnelle capacité de densification. Chaque cellule du corps humain contient un noyau de diamètre 10 micromètres. Pourtant, si l'on déroulait l'ADN qu'il renferme, il atteindrait 2 mètres. Mises bout à bout, les molécules d'ADN de toutes les cellules d'un individu mesureraient quelque 100 000 milliards de mètres. En 2014, les scientifiques ont calculé que l'archivage de 455 exaoctets (455×10^{18} octets) de données tiendrait, en théorie, dans un gramme d'ADN. Cette densité de stockage est 1 million de fois plus importante que celle des disques durs.

Malgré ces avantages, des obstacles scientifiques, économiques et éthiques doivent être levés avant que l'ADN ne remplace les traditionnels disques durs. Néanmoins, il commence déjà à faire ses preuves en tant que technologie de l'information. Ainsi, en 2016, >

> l'entreprise de postproduction hollywoodienne Technicolor est parvenue à archiver 1 million de copies du vieux film *Le Voyage dans la Lune*, de Georges Méliès, sur des brins d'ADN de synthèse, au lieu des fragiles microfilms habituels. Plus récemment, des chercheurs ont utilisé l'ADN comme outil pour concevoir des thérapies géniques, accélérer le développement de médicaments anticancer et même produire le premier «enregistrement en direct» de l'expression des gènes d'une cellule. Dans ce domaine en constante évolution, l'ADN est au cœur de recherches non seulement pour ses capacités de stockage, mais aussi pour ses possibilités de production de données à des vitesses sans précédent.

DES CODES-BARRES ADN

Au cours des dernières années, de plus en plus de scientifiques ont utilisé l'ADN comme «enregistreur moléculaire» pour suivre les résultats de leurs expériences. Le procédé implique souvent l'emploi de «codes-barres ADN»: afin de suivre les résultats d'une expérience, les scientifiques ont recours à une séquence d'ADN connue qui endossera le rôle d'étiquette moléculaire. Ils associent par exemple un résultat expérimental à la séquence ACTATC de nucléotides (les constituants de l'ADN), un autre à la séquence TCTGAT, et ainsi de suite.

Le principe du code-barres ADN est né au début des années 1990, lorsque Richard Lerner et Sydney Brenner, alors tous deux en poste à l'institut de recherche Scripps, en Californie, ont eu l'idée de l'utiliser comme un moyen de suivre des réactions chimiques. Leur concept était extrêmement innovant, mais en avance sur leur temps: les techniques permettant de séquencer facilement l'ADN à des coûts réduits n'existaient pas encore. Le potentiel de cette méthode a donc été exploité plus tard, quand le séquençage de nouvelle génération est apparu.

En 2005, une étape majeure a été franchie: des chercheurs ont réussi à analyser 25 millions de nucléotides en seulement 4 heures. Les progrès se sont poursuivis rapidement: on sait désormais lire des millions de séquences d'ADN en même temps, ce qui signifie qu'il est techniquement possible de réaliser des milliers d'expériences simultanément. L'analyse des expériences à codes-barres ADN relève donc à présent de la gestion de données. Au lieu de réaliser leurs tests un à un, les chercheurs peuvent émettre 20000 hypothèses, puis les tester toutes en même temps afin de déterminer lesquelles sont correctes.

Les biologistes ont été les premiers à recourir aux méthodes de codes-barres ADN. Puis, à mesure que le procédé est devenu plus accessible, des chercheurs issus d'autres domaines, dont l'ingénierie chimique et la science des matériaux, ont commencé à user de ces

Avec les codes-barres ADN, on peut tester 20000 hypothèses en même temps

techniques pour suivre leurs expériences à des échelles inédites. Par exemple, dans mon laboratoire de l'institut de technologie de Géorgie, aux États-Unis, des ingénieurs utilisent des codes-barres ADN pour améliorer la conception et le fonctionnement de nanoparticules. Ces nanoparticules sont destinées, *in fine*, à agir comme vecteurs, c'est-à-dire comme moyens de transport d'un médicament jusqu'à des cellules malades. Les nanotechnologies, qui reposent principalement sur l'ingénierie physique et chimique, semblent *a priori* sans rapport avec l'ADN. Mais si on considère l'ADN comme un outil de suivi et d'archivage de n'importe quelle donnée, son intérêt devient évident.

TESTER 500 NANOPARTICULES EN MÊME TEMPS

Pour les chercheurs en nanomédecine, il est plus aisé de concevoir des expériences destinées à trouver de nouvelles thérapies que de les réaliser et d'analyser les résultats. En effet, de nombreux paramètres comme la taille, la forme, la charge et la composition chimique des nanoparticules influent sur l'efficacité avec laquelle ces dernières transporteront un médicament jusqu'à des cellules malades. En outre, ces facteurs interagissent les uns avec les autres, ce qui complique la tâche des chercheurs. Une solution évidente serait de tester chaque nanoparticule individuellement. Mais, selon des données issues d'entreprises pharmaceutiques qui ont développé de telles nanoparticules, ce type de tests coûte plusieurs centaines de millions de dollars.

C'est là que les capacités de stockage de l'ADN présentent un réel intérêt. Pour augmenter le nombre de nanoparticules testées, on peut concevoir des milliers de nanoparticules aux structures chimiques variées (de grandes sphères chargées positivement ou de petits triangles à charge neutre, par exemple) et assigner un code-barres ADN à chacune d'elles. La nanoparticule 1, de structure chimique 1, porte le code-barres 1. La nanoparticule 2, de

structure chimique 2, le code-barres 2, etc. On peut ensuite administrer des centaines de ces nanoparticules à des cellules malades. Afin d'identifier les nanoparticules les mieux adaptées pour livrer un médicament, il suffit de séquencer l'ADN de ces cellules pour y quantifier les codes-barres présents. L'échelle de ces expériences est complètement nouvelle en nanomédecine. Dans mon domaine, à l'issue d'une expérience traditionnelle – le test d'une seule condition expérimentale –, on recueille en général une à cinq données. Mais d'ici à la fin de l'année, à l'aide de tels codes-barres ADN, mon laboratoire espère tester l'efficacité de 500 nanoparticules comme vecteurs dans 40 types de cellules différentes. Cela revient à pratiquer 20 000 expériences en même temps.

Pour ce faire, nous avons aussi besoin d'un outil capable de gérer ces données et de nous aider à les analyser. Nous avons donc d'abord testé la fiabilité de notre jeu de données : à quel

point les résultats issus d'une expérience reproduite plusieurs fois nous permettaient-ils de prédire l'issue d'une autre expérience ? Puis nous avons fait appel aux statistiques pour savoir si certaines caractéristiques des nanoparticules, comme leur taille, influent sur l'introduction du médicament dans les tissus cibles. Nous avons ainsi découvert que la chimie des nanoparticules, et non leur taille, affectait leur efficacité. Avec cette technique, nous espérons découvrir de nouvelles thérapies géniques plus rapidement, en utilisant moins de ressources. L'un de nos objectifs est d'identifier une nanoparticule capable de cibler spécifiquement les tumeurs. Cela réduirait les effets secondaires des traitements habituels, telles les nausées et la perte de cheveux.

Nous avons déjà eu quelques succès. En 2018, en utilisant cette approche, nous avons rapidement identifié des nanoparticules capables de livrer des médicaments aux cellules qui tapissent les vaisseaux sanguins (les cellules endothéliales), ainsi qu'à de nombreux types de cellules immunitaires. Or, avec les traitements existants, nombre de protéines des cellules immunitaires sont souvent hors de portée des agents thérapeutiques – de petites molécules ou des anticorps. Grâce à nos découvertes, les atteindre et modifier leur activité devient possible. Ces travaux ont intéressé de nombreux chercheurs en thérapie génique. Nous avons donc décidé de fonder GuideRx, une entreprise de codes-barres ADN destinée au développement de nouvelles thérapies géniques.

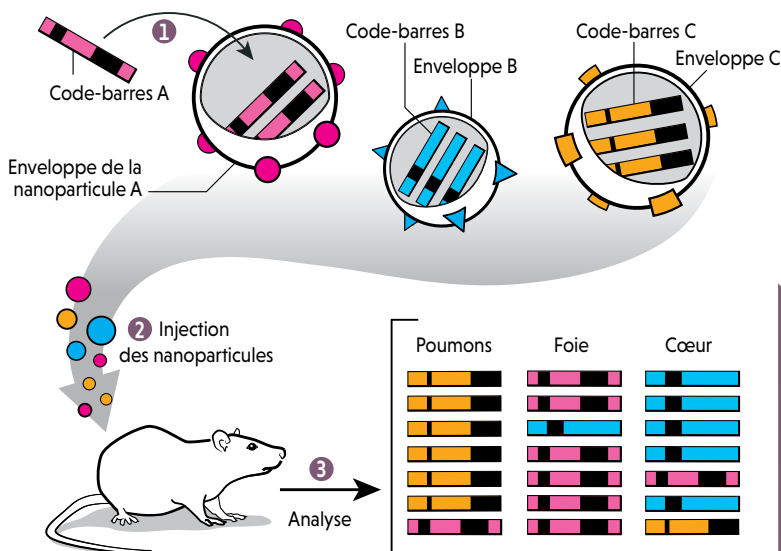
ÉTUDIER LES RÉSISTANCES DES CELLULES CANCÉREUSES

La technique des codes-barres ADN est devenue tellement commune que même au sein d'un seul domaine, elle est employée de diverses façons. C'est le cas, par exemple, en biologie du cancer, où l'on examine comment des mutations génétiques conduisent à des cancers et de quelle façon contrer les mécanismes en jeu. La résistance aux médicaments que les cellules cancéreuses développent chez les patients au fil du traitement est un défi majeur dans ce domaine.

En 2016, dans le laboratoire de Todd Golub, à l'université Harvard, des scientifiques ont eu recours aux codes-barres ADN pour étudier ces résistances. À l'aide d'un virus, ils ont inséré de façon permanente un code-barres ADN directement dans le génome de cellules cancéreuses. Une cellule cancéreuse de type A a reçu une séquence code-barres A, une cellule cancéreuse de type B, un code-barres B, etc. Les scientifiques ont ensuite mélangé les différentes cellules et les ont traitées avec un médicament contre le cancer. Une cellule sensible au traitement mourait ou ne se divisait plus. En revanche, si une cellule développait une

TRACER LES NANOPARTICULES AVEC DES CODES-BARRES ADN

Les thérapies géniques consistent à injecter du matériel génétique dans les cellules d'un organisme grâce à un vecteur – un moyen de transport capable de véhiculer le matériel génétique jusqu'au noyau des cellules. Si les vecteurs les plus courants sont dérivés de virus, on utilise aussi parfois des nanoparticules. Auparavant, tester l'efficacité de nanoparticules comme vecteurs était long et laborieux. Mais grâce aux codes-barres ADN, ces recherches sont beaucoup plus rapides : il est possible de réaliser des expériences avec des centaines de particules différentes en même temps. Pendant la phase de test (*ci-dessous*), on place un code-barres ADN différent à l'intérieur de l'enveloppe de chaque type de nanoparticule (1). Puis on injecte simultanément de nombreuses nanoparticules à un animal de laboratoire (2). Ensuite, on séquence l'ADN des cellules de l'animal pour déterminer quel code-barres elles contiennent : on sait ainsi quelle nanoparticule a atteint quel tissu (3). Grâce à cette méthode, les chercheurs identifient le modèle de nanoparticule qui convient le mieux selon l'objectif thérapeutique. Et le ciblage, plus spécifique, minimise les effets secondaires.



UN CHEVAL AU GALOP DANS UNE BACTÉRIE

Dans la plupart des recherches visant à coder des informations (images, textes, sons ou films) sur de l'ADN, les données sont archivées sur des brins d'ADN de synthèse. L'ADN est ensuite encapsulé dans des nanobilles de verre ou séché à l'air, ce qui permet de le conserver à très long terme. En 2017, Seth Shipman, alors à l'université Harvard, aux États-Unis, et ses

collègues ont testé une autre approche : stocker des images animées dans l'ADN de bactéries vivantes.

Leur image de départ était une photographie de main de 56 x 56 pixels de quatre couleurs de pixels différentes (quatre niveaux de gris, voir ci-dessous). En langage binaire, chaque couleur de pixel peut être codée par un duo de 0 et/ou de 1. Pour encoder l'image dans de l'ADN, les chercheurs ont associé chaque couleur de pixel à un nucléotide. Puis ils ont scindé les 3 136 pixels de l'image en 112 ensembles de 28 pixels, disséminés dans toute l'image. Chaque ensemble de pixels était ainsi codé par une succession de 28 nucléotides, à laquelle était ajouté un code-barres de 4 nucléotides permettant de reconnaître la zone et de la situer dans l'image. Les biologistes ont ensuite synthétisé les 112 fragments d'ADN obtenus en de nombreux exemplaires.

Pour intégrer ces fragments au sein du génome de bactéries *Escherichia coli*, ils ont détourné l'arsenal CRISPR de ces organismes – une sorte de système immunitaire qui enregistre, dans l'ADN bactérien, une trace des infections virales subies. Par électroporation, une technique qui consiste à appliquer un champ électrique sur les membranes cellulaires afin de les perméabiliser, les chercheurs ont introduit les fragments d'ADN dans les bactéries. Grâce à la machinerie CRISPR, dont les biologistes avaient forcé l'activation, les fragments d'ADN ont été copiés au sein du génome des bactéries. Le jour suivant, les chercheurs ont séquencé l'ADN de la population bactérienne,

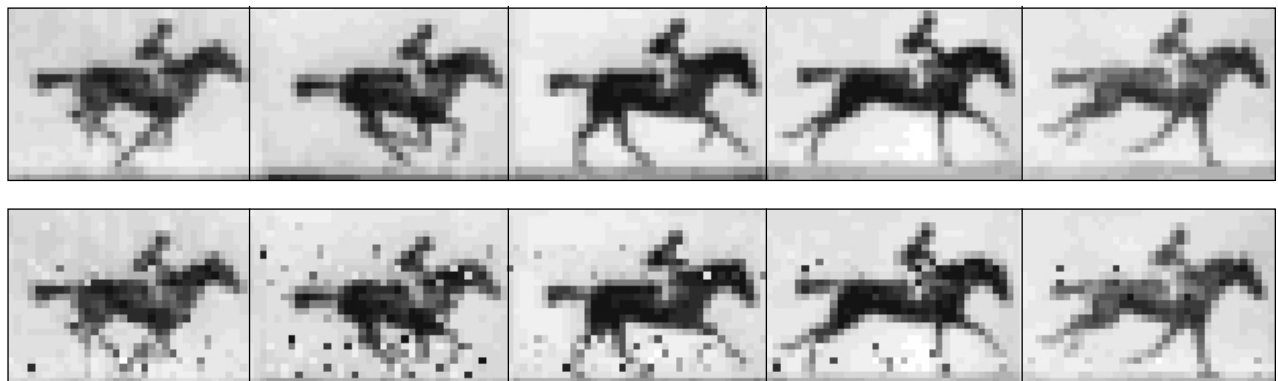
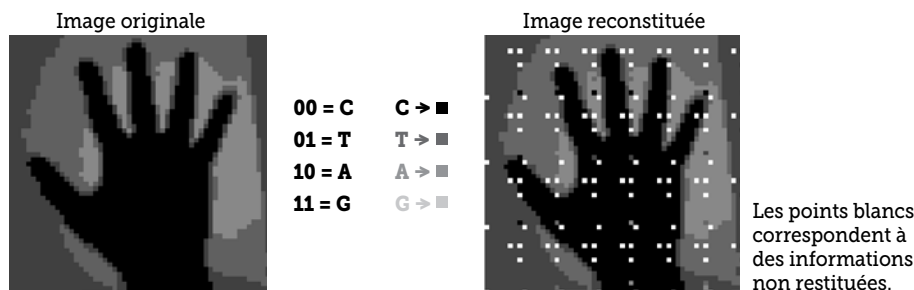
puis, grâce à un programme, ils ont converti les séquences nucléotidiques en langage binaire et ont ouvert l'image : 88 % des ensembles de pixels étaient correctement restitués. Les chercheurs ont ensuite réalisé la même expérience avec une image de main de 33 x 33 pixels comportant cette fois 21 couleurs de pixels. Les couleurs de pixels étaient codées non plus sur un nucléotide, mais sur un triplet. Avec cette méthode, 96 % des ensembles de pixels étaient fidèlement restitués 2 heures 40 après l'électroporation. Ensuite, le pourcentage de reconstitution de l'image diminuait légèrement dans les 24 heures, puis restait stable pendant les 6 jours suivants, au cours desquels 48 générations de bactéries se succèdent.

Les scientifiques ne se sont pas arrêtés à ce premier succès : ils ont souhaité enregistrer une séquence de cinq images issues du premier film d'Eadweard Muybridge, mettant en scène un cheval au galop et sa cavalière. Pour cela, ils ont exploité une caractéristique de l'arsenal CRISPR : lorsque les bactéries copient de l'ADN étranger au sein de leur propre ADN, l'ADN est copié dans un ordre précis selon la chronologie d'apparition. Ainsi, sur le même principe que pour l'image de main, ils ont d'abord introduit les brins d'ADN constituant la première image. Puis, le lendemain, ils ont refait de même avec ceux de la deuxième image, et ainsi de suite pendant cinq jours, réussissant de cette façon à coder une image animée dans de l'ADN bactérien.

En réalisant ces travaux, les chercheurs imaginaient-ils un futur où nos films favoris seraient stockés dans le génome de populations bactériennes ? Non, car les brins d'ADN de synthèse sont plus appropriés à cette fin. « Mais ce qui est vraiment intéressant dans cette étude, commente Christophe Antoniewski, directeur de la plateforme de bio-informatique ARTbio de l'institut de biologie Paris-Seine, c'est la perspective d'utiliser l'ADN comme enregistreur moléculaire *in vivo* de l'activité des cellules, afin d'étudier, par exemple, leur développement, ou leur dérèglement dans les cancers. »

HÉLÈNE GÉLOT
Journaliste scientifique

AVANT / APRÈS



La séquence originale (*en haut*) et celle reconstituée à partir des bactéries (*en bas*) ne diffèrent que de quelques pixels.

> résistance, elle se divisait rapidement. Par conséquent, au fil du temps, la quantité de codes-barres A, par exemple, augmentait si la cellule de type A devenait résistante au médicament ou diminuait si elle était neutralisée. En séquençant tous les codes-barres des cellules survivantes au cours du temps, le laboratoire a quantifié les réponses de chaque type cellulaire au médicament.

Quelques mois plus tard, ce fut le tour de l'équipe de Monte Winslow, de l'université de Stanford, de mener des recherches à l'aide de codes-barres ADN, cette fois pour trouver des inhibiteurs de métastases dans le cancer du pancréas. Les biologistes ont d'abord produit des lignées de cellules cancéreuses à partir de

cardiovasculaires. Dans les exemples cités plus haut, remplacez le mot « cancer » par une autre maladie et le mot « résistance » par n'importe quelle réponse à un médicament et vous percevrez l'étendue de leur potentiel.

ENCODER UN FILM DANS L'ADN D'UNE BACTÉRIE

Ce n'est pas tout. Les approches utilisant des codes-barres ADN reposent sur la lecture de séquences connues d'ADN. Mais cela suppose de préalablement « écrire » ces séquences. Or, jusqu'à récemment, la synthèse de séquences d'ADN restait compliquée dès qu'elles étaient un peu trop longues. Il était donc inenvisageable de convertir d'autres formes d'informations – images, films, états biologiques... – en séquences de nucléotides que l'on pourrait stocker et lire plus tard. Mais de nouvelles techniques d'écriture offrent désormais cette possibilité. Elles reposent sur des outils d'édition du génome dérivés du système CRISPR.

Ce système est un moyen de défense des bactéries contre les attaques de virus. Lorsqu'un virus infecte une bactérie, il se fixe à sa surface et introduit son ADN (ou ARN) à l'intérieur. Mais la bactérie garde alors une trace de l'attaque en insérant des extraits de cet ADN (ou ARN) étranger dans une région de son propre génome – une sorte de bibliothèque nommée CRISPR. En cas de nouvelle attaque du même virus, la bactérie le reconnaît alors grâce à sa bibliothèque CRISPR et détruit son matériel génétique avant qu'il n'ait le temps de se répliquer.

En 2017, dans le laboratoire du généticien George Church à l'université Harvard, Seth Shipman, aujourd'hui à l'université de Californie à San Francisco, a exploité ce mécanisme de défense pour enregistrer une image de main dans le génome d'une population de bactéries *Escherichia coli* (voir l'encadré page ci-contre). Dans de précédents travaux, l'équipe avait montré qu'il suffisait d'induire l'expression de deux protéines de ces bactéries – Cas1 et Cas2 –, associées au fonctionnement du système de défense CRISPR, pour forcer les cellules à insérer de courtes séquences d'ADN de synthèse dans leur bibliothèque CRISPR. Seth Shipman et ses collègues ont donc recommencé, cette fois avec des fragments d'ADN qui encodaient les pixels de l'image.

En séquençant l'ADN d'*E. coli*, ils ont retrouvé l'image initiale avec une précision de plus de 90%. Ils ont ensuite reproduit l'expérience, mais avec une difficulté supplémentaire : ils ont ajouté les séquences d'ADN à des temps différents et mis au point une méthode pour enregistrer la position de ces séquences les unes par rapport aux autres. En retrouvant l'ordre d'ajout des séquences au sein du >

Ils ont encodé une série d'images dans l'ADN et l'ont restituée

tumeurs primaires et de métastases provenant de modèles murins de la maladie. Parmi elles, ils n'ont gardé que celle qui, injectée chez des souris, formait le plus facilement des métastases. Puis, à l'aide d'un virus, ils ont introduit une étiquette d'ADN dans 96 lignées cellulaires issues de la lignée sélectionnée, avant de disposer chacune d'elles dans un puits distinct. Ils ont ensuite ajouté un inhibiteur potentiel dans chaque puits. Ainsi, chaque molécule était associée à un code-barres différent. Les scientifiques ont alors immédiatement injecté les cellules dans les vaisseaux sanguins de souris et suivi, grâce aux codes-barres, quelles cellules se sont développées dans les poumons. En identifiant les codes-barres présents ou absents des poumons, ils ont ainsi mis en évidence plusieurs molécules qui favorisaient ou empêchaient les métastases. Cette méthode leur a permis de tester plus de 700 inhibiteurs d'enzymes sur 20 fois moins de souris.

Dans ces approches, l'ADN est à la fois un enregistreur moléculaire – une étiquette qui permet de suivre des expériences menées simultanément et les repérer les unes par rapport aux autres – et un générateur de données, puisqu'il offre la possibilité de réaliser de nombreuses expériences en même temps. Les implications sont stupéfiantes : les mêmes techniques peuvent être appliquées contre les maladies auto-immunes, neurologiques ou

> génome de la bactérie, ils sont parvenus à encoder une série d'images dans l'ADN et à la restituer. Les chercheurs ont ainsi enregistré un extrait de la première séquence d'animation créée en 1878 par le photographe britannique Eadweard Muybridge, qui met en scène un cheval au galop.

ENREGISTRER EN TEMPS RÉEL L'EXPRESSION DES GÈNES

Plus récemment encore, en 2018, des chercheurs du laboratoire de Randall Platt, à l'École polytechnique fédérale de Zurich, en Suisse, ont mené ces approches encore plus loin en ciblant non plus l'ADN, mais l'ARN messager. Cette molécule est un cousin moléculaire clé de l'ADN, intermédiaire indispensable entre les gènes portés par l'ADN et la synthèse des protéines qu'ils codent. Au lieu de stocker des images dans des séquences d'ADN de synthèse, les chercheurs ont utilisé le système CRISPR d'une bactérie pour enregistrer en temps réel les ARN messagers qu'elle produisait, et donc l'expression de ses gènes. L'ensemble de tous les ARN messagers présents dans une cellule à un instant donné permet de savoir quelles protéines sont produites, et ainsi de mieux comprendre le fonctionnement de la cellule étudiée.

Afin d'enregistrer dans le génome d'une cellule les séquences des ARN messagers qu'elle a produits à divers moments, l'équipe de Randall Platt a tout d'abord testé différentes protéines Cas (associées au système CRISPR) issues de nombreuses souches bactériennes distinctes. Elle a ainsi découvert que les protéines Cas1 et Cas2 de la bactérie *Fusicatenibacter saccharivorans* étaient capables de convertir l'ARN messager en ADN et de l'encoder dans sa bibliothèque CRISPR. Les chercheurs ont utilisé ce procédé afin d'enregistrer le comportement d'une cellule à la suite de différents événements: un stress oxydatif, des conditions acides et l'exposition à un herbicide.

Ces résultats sont particulièrement intéressants: ils fournissent la preuve qu'il est possible d'enregistrer dans le génome d'une cellule, à un instant donné, les gènes qu'elle exprime, en vue d'une analyse ultérieure. Le laboratoire de Randall Platt continue d'améliorer cette technique. L'enregistrement en temps réel de l'expression des gènes pourrait s'étendre à d'autres domaines et permettre, par exemple, de suivre les événements qui rendent cancéreuse une cellule ou encore d'observer comment une cellule répond à une infection au cours du temps.

Alors que l'ADN est utilisé pour produire, suivre et stocker de l'information dans des domaines de plus en plus nombreux, on peut se demander s'il a la capacité de rivaliser avec les dispositifs électroniques usuels de stockage et de prendre en charge l'ensemble des

Conservé au sec et au frais, l'ADN résiste des dizaines de milliers d'années

données générées par les humains. Pour l'instant, la réponse est non: disques durs et cartes mémoires sont bien plus adaptés pour conserver des informations que le plus avancé des systèmes génétiques.

En même temps, comme toute technologie, les appareils électroniques conventionnels ont leurs limites. Ils prennent de la place et nécessitent des conditions environnementales particulières. Et même les dispositifs les plus robustes tiennent rarement plus de quelques décennies. À cause de ces obstacles, il deviendra de plus en plus difficile de conserver toutes les données que nous produisons.

Au contraire, l'ADN résiste des dizaines de milliers d'années s'il est conservé dans des environnements frais et secs. Il est aujourd'hui couramment stocké à -20°C , voire à -80°C dans les laboratoires, mais pourrait aussi être conservé sous des chaleurs extrêmes que les appareils électroniques ne supportent pas. En 2015 en effet, Robert Grass et Wendelin Stark, tous deux à l'École polytechnique fédérale de Zurich, ont montré que des molécules d'ADN encapsulées dans des billes de verre résistaient à une température de 70°C pendant une semaine sans qu'aucune erreur n'apparaisse. Et si les disques durs permettent de stocker 19 gigaoctets par centimètre carré, des estimations récentes suggèrent que l'ensemble des informations générées dans le monde tiendraient, en théorie, dans moins de 1 kilogramme d'ADN.

UNE RECHERCHE À ENCADRER

Il reste du chemin à parcourir avant que le stockage sur ADN devienne courant. Le premier obstacle réside dans l'extraction des informations: si elle est presque instantanée à partir d'un disque dur, récupérer des données encodées dans l'ADN requiert du séquençage, ce qui peut prendre de quelques minutes à une journée entière. Et même si les séquenceurs ont connu des avancées considérables ces dernières années, ils restent imposants et chers par rapport aux disques durs.

BIBLIOGRAPHIE

C. D. Sago et al., **High-throughput in vivo screen of functional mRNA delivery identifies nanoparticles for endothelial cell gene editing**, *PNAS*, vol. 115(42), pp. E9944-E9952, 2018.

F. Schmidt et al., **Transcriptional recording by CRISPR spacer acquisition from RNA**, *Nature*, vol. 562, pp. 380-385, 2018.

S. L. Shipman et al., **CRISPR-Cas encoding of a digital movie into the genomes of a population of living bacteria**, *Nature*, vol. 547, pp. 345-349, 2017.

G. M. Church et al., **Next-generation digital information storage in DNA**, *Science*, vol. 337, p. 1628, 2012.

Ces limites ne sont pas les seules considérations à prendre en compte. La banalisation du séquençage aura des répercussions sur le quotidien: tracer des personnes deviendra plus facile et de nouvelles vulnérabilités quant à la sécurité des données feront leur apparition. Les exemples d'atteinte à la vie privée sont déjà légion. Aux États-Unis, le séquençage de l'ADN est utilisé par les départements de police avec peu de vigilance. Lorsqu'elle enregistre l'ADN de personnes en état d'arrestation, même pour des crimes mineurs, la police enrichit de vastes banques de données génétiques. Cette procédure est considérée par certains comme l'équivalent moderne des empreintes digitales. Pourtant, il y a une différence majeure: les empreintes digitales n'identifient qu'un seul individu. En revanche, si une personne fournit son ADN, elle fournit une information qui peut permettre d'identifier n'importe quel membre de sa famille. En Chine, sous le couvert d'un programme de santé, des agents ont rassemblé des informations génétiques concernant près de 36 millions de personnes sans que l'on ne connaisse l'usage que le gouvernement en fera. Or cette population compte notamment de nombreux Ouïgours, les membres d'un groupe ethnique musulman victime de discriminations...

Jusqu'à aujourd'hui, les principales préoccupations sur le stockage d'informations génétiques concernaient le patrimoine individuel et la protection de l'identité. Mais dans le futur, si des informations comme des données de santé, des contrats légaux et d'autres histoires individuelles sont stockées dans de l'ADN, cela soulèvera de nombreuses questions en termes de sécurité physique des données et de cybersécurité. Des quantités si importantes d'informations rentreront dans un espace si petit: comment distribuer les données pour éviter qu'elles soient concentrées en un seul endroit? Et même si l'extraction devient plus simple, comment accèdera-t-on puis stockera-t-on à nouveau les données sans les exposer à des piratages malveillants ou à des pertes accidentelles?

L'ampleur du travail scientifique et éthique à accomplir me décourage parfois. Je pense alors aux frères Wright. Leur premier vol a duré 12 secondes, durant lesquelles ils ont parcouru 37 mètres. Soixante-six ans plus tard, les humains ont posé le pied sur la Lune. Ces exploits me laissent espérer que nous saurons exploiter le potentiel de l'ADN au cours des prochaines décennies et mettre tout en œuvre pour que ce potentiel soit mis au service du bien commun. ■

RENCONTRES

Samedi 5 octobre – 15h

PoulpeFiction – Chronique d'une intelligence hors du commun

Dans le cadre de l'exposition *Océan, une plongée insolite* à la Grande Galerie de l'Évolution

Avec : **Laure Bonnaud-Ponticelli**, Biologiste, enseignante-chercheuse, spécialiste des céphalopodes, Muséum et **Aline Richard Zivohlava**, Journaliste scientifique pour le site The Conversation France, auteure de *Dans la peau des bêtes* paru aux éditions Plon (2019).

Dès 8 ans - sur inscription : jardindesplantesdeparis.fr

Rotonde de la Ménagerie

Ménagerie, zoo du Jardin des Plantes - Paris 5^e

Lundi 7 octobre – 19h

L'agroécologie en pratique : trajectoires et expériences d'agriculteurs.trices

Modératrice : **Margaux Alarcon**, doctorante, Muséum

Avec : **Marie Françoise Bizard-Pasquet**, agricultrice gérante et membre de l'Académie d'Agriculture, **Kevin Morel**, ingénieur agronome, AgroParisTech Paris, **Véronique Lucas**, sociologue et agroécologue, INRA Montpellier

Lundi 14 octobre – 19h

De l'océan à l'assiette

Modérateur : **Christophe Lavelle**, biophysicien, commissaire de l'exposition *Je mange donc je suis* du Musée de l'Homme

Avec : **Guy Duhamel**, ichtyologue, Muséum, **Laurent Nicolle**, armateur, Groupe Le Garrec, **Elisabeth Vallet**, directrice Éthique Océan, **Julien Dumas**, chef du Lucas Carton, Paris

Auditorium de la Grande Galerie de l'Évolution

36 rue Geoffroy St-Hilaire, Paris 5^e

LES { Partagez les savoirs }
RENDEZ-VOUS DU MUSÉUM

Entrée gratuite

Au Jardin des Plantes

Détails sur jardindesplantesdeparis.fr

POUR LA
SCIENCE



L'ESSENTIEL

> Ces dernières années, un grand nombre d'études de psychologie ont échoué à répliquer les résultats de recherches précédentes.

> Une crise en a découlé, mais ces échecs sont aussi le signe que la psychologie s'est résolument engagée dans la voie de l'autocorrection.

> Loin d'en être affaiblie, elle en sort donc renforcée. D'autant plus que les effets classiques, corroborés par tout un faisceau d'études, tiennent bon.

L'AUTEUR



FRANÇOIS MAQUESTIAUX
professeur de psychologie cognitive
à l'université de Franche-Comté,
à Besançon, et membre de l'institut
universitaire de France

Le grand ménage de la psychologie

Non, vous ne vivrez pas plus longtemps si vous souriez sur les photographies.

Non, tenir un crayon entre les dents ne modifie pas votre jugement...

Nombre de conclusions d'études menées en psychologie doivent être soumises à un examen plus approfondi. Depuis quelques années, les chercheurs de cette discipline s'y attendent.

«**L**a plupart des résultats scientifiques publiés sont faux» : tel est le titre provocateur d'un article publié en 2005 par John Ioannidis, professeur de médecine à l'université Stanford, aux États-Unis. De fait, plusieurs cas de fraude seront découverts par la suite : quelques années après le cri d'alarme de John Ioannidis, le psychologue néerlandais Diederik Stapel sera par exemple radié de l'université de Tilburg, à l'âge de 45 ans, pour avoir inventé ou trafiqué les résultats de 55 de ses 137 publications scientifiques. Mais le problème dépasse largement ces tricheries manifestes. Il porte sur la nature et l'organisation même de la recherche dans plusieurs disciplines, qui entraînent la publication de nombreux « faux positifs » – des travaux qui semblent démontrer un phénomène, mais de façon illusoire (voir l'encadré page 48).

Si la question concerne toutes les disciplines expérimentales, les projecteurs se sont braqués sur la psychologie en 2015, avec la parution dans la revue *Science* d'une étude de

grande ampleur, associant 270 chercheurs et visant à reproduire 100 expériences. Les résultats étaient alarmants : seulement 36% des répliques menées ont confirmé les résultats originaux. Dès 1935, le philosophe viennois Karl Popper soulignait que la reproductibilité d'une découverte est le gage de sa qualité scientifique ; cette étude a donc déclenché un tsunami : la psychologie serait-elle une pseudoscience ? Les deux tiers de ses résultats seraient-ils à jeter à la poubelle ?

UNE CURE DE RÉPLICATION EST LANCÉE

En réalité, ce qui pouvait passer pour une catastrophe est une excellente nouvelle, puisque la psychologie est l'une des rares disciplines à s'être engagée dans la voie de l'autocorrection. Depuis un peu moins d'une dizaine d'années, une véritable cure de réplique a été lancée, afin de vérifier les effets allégués par un certain nombre de travaux. Il s'agit de reproduire les études originales à l'identique (même matériel, mêmes tâches, même environnement) ou avec de légères adaptations. ➤