

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