

# Les atouts de l'ADN synthétique pour stocker les données numériques

Hanna Siemiatycki

**Une technique de codage d'informations dans les molécules d'ADN a été mise au point en laboratoire. L'application à grande échelle se heurte encore à des difficultés d'écriture et de lecture rapide.**

Face à la production exponentielle de données numériques chaque année, et alors que les dispositifs de stockage actuels (disques durs, bandes magnétiques) ne parviennent plus à suivre le rythme, la piste du stockage moléculaire dans l'ADN synthétique semble prometteuse. Du moins, en ce qui concerne l'archivage de « données froides », ces informations très rarement relues mais pour lesquelles il est essentiel de conserver une copie (actes notariés, documents juridiques, médicaux ou audiovisuels, etc.).

Pour écrire un code binaire dans l'ADN, les chercheurs s'appuient sur sa structure moléculaire composée d'un enchaînement de quatre nucléotides, pouvant coder deux bits chacun : l'adénine (A), la cytosine (C), la guanine (G) et la thymine (T). Cet alphabet quaternaire est traduit en suite de 0 et de 1 en utilisant, par exemple, la convention 00 → A ; 01 → C ; 10 → G et 11 → T. 0010110111 devient ainsi AGTCT. Le séquençage de l'ADN permet ensuite de déchiffrer le message codé.

« Le stockage sur l'ADN a pour avantage de s'appuyer sur les méthodes très puissantes de la biotech, avec des séquenceurs qui n'ont pas d'équivalent pour d'autres molécules », explique Yannick Rondelez, directeur de recherche CNRS à l'Ecole supérieure de physique et de chimie industrielles (ESPCI). D'autant que l'ADN, déjà optimisé par la nature et extrêmement étudié, est devenu familier d'une large communauté de techniciens.

## Qualités intrinsèques

Les prémices de cette science, au croisement de la biologie, de la chimie et de l'informatique, remontent aux années 1960. Mais c'est en 2012 que deux preuves majeures – l'une aux Etats-Unis (George Church, Harvard) et l'autre au Royaume-Uni (Nick Goldman, Institut européen de la bio-informatique) – ont ancré ce concept dans la réalité. Parvenant à coder des kilooctets de données dans l'ADN, ils ont ouvert la voie à des start-up et à des industriels. La DNA Data Storage Alliance, un consortium public-privé composé d'une cinquantaine de membres du monde entier, a ainsi été créée en 2020 par quatre entreprises américaines, dont les géants Microsoft et Western Digital.

« Il y a trois pôles : les Etats-Unis, la Chine et l'Europe. La France est très active, on possède tout le spectre de compétences », indique Marc Antonini, directeur de recherche CNRS (laboratoire I3S, université Côte d'Azur et CNRS), qui dirige le Programme et équipements prioritaires de recherche (PEPR) exploratoire lancé en 2022 par le gouvernement dans ce domaine (20 millions d'euros sur sept ans).

Candidat tout trouvé du fait de sa longévité (des milliers, voire des millions d'années) et de sa frugalité énergétique, c'est surtout la compacité de l'ADN qui intéresse : « L'intégralité des données actuelles pourrait être stockée dans une boîte à chaussures d'ADN », fait valoir M. Rondelez. Néanmoins, des limites technologiques freinent son déploiement commercial : des mécanismes d'écriture (synthèse) et de lecture (séquençage) chronophages, coûteux et encore loin de l'automatisation.

## Plus vite et moins cher

« Ces futurs systèmes devront être évolutifs et aussi robustes que les autres solutions commerciales, abonde Karin Strauss, directrice de la recherche chez Microsoft. Notre travail en a jeté les bases, mais il reste encore beaucoup d'ingénierie à faire. » En partenariat depuis 2015 avec le laboratoire des systèmes d'information moléculaire (MISL) de l'université de Washington, les chercheurs de Microsoft sont parvenus à stocker plus d'un gigaoctet dans l'ADN. Et à développer l'accès sélectif à certaines informations dans une base de

données, une fonction primordiale alors que les utilisateurs ne consultent en général que 1 % de leurs archives. Plus important encore, ils ont construit le premier système de stockage dans l'ADN (seulement cinq octets) de bout en bout entièrement automatisé, publié dans Scientific Reports en 2019.

Au MISL, ils travaillent désormais à « développer de nouvelles technologies de lecture et à utiliser la synthèse enzymatique (non chimique, employant des enzymes) pour fabriquer l'ADN », précise Jeff Nivala, codirecteur de recherche. Tout comme M. Rondelez, copilote d'un volet du PEPR français, qui emploie cette voie de synthèse moins polluante pour « massifier la synthèse d'ADN, en parallélisant et en accélérant l'écriture des brins ». Aller cent fois plus vite, pour être cent fois moins cher, tel est l'objectif.

Depuis l'université de Tokyo, Anthony Genot, qui dirige le volet consacré aux applications, codéveloppe un démonstrateur basé sur la microfluidique (des réactions entre des liquides à l'échelle du microlitre) capable d'encoder et de décoder un gigaoctet dans l'ADN. « Pour convaincre les partenaires du projet – la BNF, l'INA et le Parlement européen – que ça marche », dit-il.

### Horizon 2030-2035

« Lorsqu'on va accélérer les synthèses, on va générer du bruit (au codage et au décodage) et probablement des dégradations de stockage qu'il faudra prendre en compte dans nos futurs codes correcteurs d'erreur », fait remarquer M. Antonini, qui codirige le volet consacré au traitement numérique de l'information. « On vise des systèmes automatisés et compétitifs avec les disques durs, à horizon 2030-2035 », ajoute-t-il.

Cofondateur, en 2022, de la start-up Pearcode, il travaille à des systèmes d'écriture et de lecture automatisés plus futuristes. Telles des boîtes dotées d'une connectique USB que l'utilisateur « pluggerait » sur son ordinateur ou qui s'installeraient en rack dans des data centers.

A Paris, Erfane Arwani, PDG et cofondateur de Biomemory, créée en 2021, indique qu'il disposera en 2025 d'un « prototype de "super-imprimante" à carte ADN, basé sur la biologie de synthèse. Elle fabriquera des cartes ADN d'un gigaoctet ». Une étape intermédiaire qui devrait mener à la conception, d'ici à 2028, d'une armoire de data center intégrant une imprimante à carte ADN d'un pétaoctet, où les cartes seront stockées les unes à côté des autres.

« Il reste peu probable qu'une mémoire principale d'ordinateur soit, un jour, constituée d'ADN », tient à nuancer Luis Ceze, architecte informatique et professeur de recherche au MISL. Quoi qu'il en soit, archiver dans l'ADN pourrait rendre bien des services comme contourner un rançongiciel, en cas de cyberattaque, ou servir de backup lors d'une panne généralisée d'ordinateurs, telle celle dont a été victime Microsoft en juillet.

Mais l'avènement de cette technologie dépendra avant tout des financements publics encore nécessaires pour changer d'échelle.

Mis à jour le 10 septembre à 14h : A été précisé que Yannick Rondelez est directeur de recherche CNRS.

[Cet article est paru dans Le Monde \(site web\) \(https://www.lemonde.fr/sciences/article/2024/09/03/les-atouts-de-l-adn-synthetique-pour-stocker-les-donnees-numeriques\\_6302640\\_1650684.html\)](https://www.lemonde.fr/sciences/article/2024/09/03/les-atouts-de-l-adn-synthetique-pour-stocker-les-donnees-numeriques_6302640_1650684.html)

### Note(s) :

Mis à jour : 2024-09-10 14:00 UTC +0200

© 2024 SA Le Monde. Tous droits réservés.

Le présent document est protégé par les lois et conventions internationales sur le droit d'auteur et son utilisation est régie par ces lois et conventions.

**Publi©** Certificat émis le **9 janvier 2025** à Réseau-Canopé à des fins de visualisation personnelle et temporaire.

**news.20240903.LMF.6302640\_1650684**