

Utiliser l'ADN pour stocker des données



Les technologies de stockage de données ont progressé extraordinairement depuis le premier disque dur développé par IBM en 1956. Cependant, alors que les nouvelles technologies ont permis d'augmenter considérablement les capacités de stockage dans des outils miniaturisés, elles restent fragiles, énergivores et peu adaptées pour conserver des données d'intérêt sur le long terme. La nature offre une solution : l'ADN.

L'acide désoxyribonucléique, ou ADN, est une macromolécule biologique utilisée dans presque toutes les cellules ainsi que chez de nombreux virus comme vecteur de leur information génétique. Alors que la succession de 0 et 1 constitue le langage binaire informatique, l'enchaînement des bases adénine (A), cytosine (C), guanine (G) et thymine (T) de l'ADN code l'information génétique. Les développements récents des techniques de synthèse et séquençage de l'ADN rendent envisageable l'utilisation de cette macromolécule pour le stockage à long terme de grandes quantités d'informations. L'ADN contenu dans un corps humain comprend l'équivalent en données de 150 000 des plus grands centres de stockage de données numériques, soit 150 milliards de Terabytes. Ainsi, l'ensemble des films produits depuis l'invention du cinématographe pourrait être conservé sous forme d'ADN dans un volume inférieur à un « morceau de sucre ». L'ADN, très stable, peut par ailleurs être conservé pendant plusieurs milliers d'années.

En 2020, l'équipe de recherche dirigée par Nicholas Guise au *Georgia Institute of Technology* (Atlanta, Géorgie) a reçu 25 millions de dollars de financement IARPA (*Intelligence Advanced Research Projects Activity*) pour développer de nouvelles

technologies de synthèse et séquençage d'ADN. La technologie de synthèse, développée en collaboration avec l'entreprise de biotechnologie [Twist Biosciences](#), repose sur un support en silicone creusé de micro-puits de quelques centaines de nanomètres de profondeur pouvant accueillir la [synthèse](#) chimique de molécules uniques d'ADN base par base, et permettant ainsi la production en parallèle de millions de séquences uniques sur une seule micro-puce. La micro-puce intègre les technologies de micro-fluidique, pour l'approvisionnement précis des réactifs aux sites de synthèse, et de semi-conducteurs électroniques permettant l'activation d'électrodes uniques, chaque puits constituant ainsi un mini bio-réacteur électrochimique. L'activation d'une électrode induit la création d'un environnement acide propice à l'ajout d'une base à la molécule d'ADN synthétisée, et la régulation de l'activité des électrodes par semi-conducteurs permet un contrôle précis de la séquence produite dans chaque puits. Le développement de cette nouvelle micro-puce permettrait une production d'ADN cent fois plus efficace en comparaison avec les techniques utilisées jusqu'à présent. L'objectif affiché est de pouvoir encoder plusieurs Terabytes de données par jour dans chaque micro-puce.

L'équipe travaille aussi en partenariat avec [Roswell Biotechnologies](#) sur des modèles de séquenceurs ENDseq (*Electronic Nano-Detection Sequencing*) qui consistent en des senseurs électroniques moléculaires uniques regroupés sur un semi-conducteur. En bref, des enzymes ADN polymérases utilisent le brin d'ADN d'intérêt pour produire un ADN complémentaire par ajout séquentiel de base. L'ajout d'une base induit une modification locale du courant électrique spécifique à chaque base, la mesure des variations du courant permet ainsi de déterminer la séquence de la molécule d'ADN d'intérêt. Enfin, l'équipe de Nicholas Guise développe des algorithmes de décodages plus fiables afin de minimiser le taux d'erreur du cycle synthèse-séquençage, qui est encore trop élevé pour garantir l'utilisation à grande échelle de cette technologie.

En complément des technologies électroniques et magnétiques permettant le stockage et l'accessibilité des données rapidement, le stockage d'information grâce à l'ADN pourrait donc être favorisé pour archiver des données sur le long terme. L'utilisation de l'ADN permet une réduction de l'empreinte physique et énergétique et du coût du stockage d'information. De manière importante, les connaissances permettant la synthèse et la lecture de l'information génétique ne deviendront pas

obsolètes puisqu'elles resteront essentielles à la compréhension de la nature et à la santé humaine.

Rédacteur - Benjamin Boumard, Attaché adjoint pour la Science et la Technologie, Atlanta,

deputy-univ@ambascience-usa.org

RÉFÉRENCES

<https://rh.gatech.edu/news/653264/data-dna-0>

<https://www.gtri.gatech.edu/newsroom/data-dna>

<https://gtri.gatech.edu/newsroom/25-million-project-will-advance-dna-based-archival-data-storage>

<https://www.researchgate.net/profile/Nicholas-Guise>

<https://www.youtube.com/watch?v=wxStlzunxCw>

<https://www.bbc.com/news/science-environment-59489560>

<https://www.youtube.com/watch?v=yAQHMkzlCLk>

<https://www.roswellbiotech.com/applications/genome-sequencing>

<https://www.twistbioscience.com/products/storage>

<https://www.youtube.com/watch?v=KUm173PZJBQ>