

Le potentiel de la mémoire à ADN

La quantité totale d'informations numériques sur Terre est aujourd'hui de l'ordre de 10^{20} (1 Zo = 1 zettaoctet = 10^{21} caractères), soit 10^{22} caractères, ou 5 milliards de milliards de pages imprimées de 2 000 caractères.

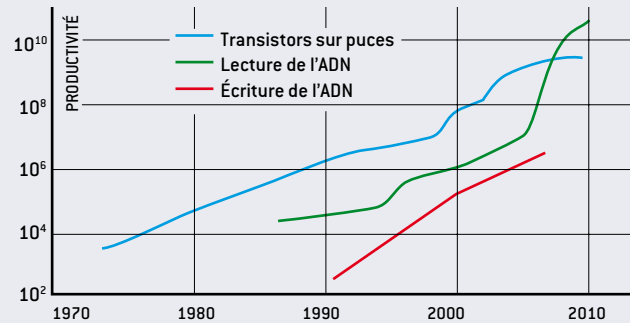
Si l'augmentation se poursuit au rythme des dernières décennies, nous serons incapables, par les moyens utilisés aujourd'hui, de répondre à la demande d'archivage de ces données.

Aussi étudie-t-on des moyens de stockage autres que magnétiques ou optiques. Les méthodes exploitant l'ADN pourraient être la solution pour le stockage massif à long terme, d'autant que la densité d'information envisageable avec de l'ADN est bien supérieure à celle des meilleures mémoires magnétiques ou optiques. Ce n'est peut-être pas étonnant : la mémorisation par l'ADN se fait à l'échelle de la molécule, ce qui n'est pas le cas en électronique ou en optique.

Les trois courbes calculées par Rob Carlson (un investisseur américain dans

la bioéconomie) montrent que le rythme de croissance de la puissance informatique, qui suit la loi de Moore (doublement tous les 18 mois à peu près depuis plus de quarante ans), reste plus faible que celui de l'efficacité des dispositifs d'écriture et de lecture de l'ADN. Cependant, si un moyen de stockage permet d'écrire et de lire vite, alors il ne conserve pas l'information longtemps.

Comparons les performances des quatre techniques que sont les mémoires



flash (clés USB, disques SSD), les disques durs des ordinateurs, les bandes magnétiques et l'ADN. Les temps d'accès se mesurent respectivement en millisecondes, en dizaines de millisecondes, en minutes et en dizaines d'heures.

Les durées de conservation sont respectivement de l'ordre de 5 ans, 5 ans, 15-30 ans et plusieurs siècles. Le stockage sur de l'ADN, bien que lent, apparaît donc beaucoup plus durable.

aléatoires selon un codage [expliqué dans l'encadré page 79], et ce jusqu'à la taille n , le nombre de nœuds du graphe.

- On filtre les séquences selon leur masse pour ne garder que les chemins de longueur n .

- On filtre en plusieurs étapes les séquences restantes en utilisant les mécanismes moléculaires connus, pour ne garder dans la soupe que les chemins passant une fois et une seule par chaque nœud.

- Les séquences qui restent indiquent des chemins hamiltoniens.

Cette méthode exploite le parallélisme des réactions chimiques qui, d'un endroit à l'autre de la soupe moléculaire, « calculent » simultanément, c'est-à-dire essaient des chemins différents en éliminant les chemins insatisfaisants. En théorie, le problème est résolu en un nombre d'étapes de réactions chimiques qui croît linéairement en fonction du nombre de nœuds du graphe. On disposerait donc d'une méthode de résolution en temps polynomial d'un problème NP-complet, et, en conséquence, d'une méthode polynomiale pour résoudre

n'importe quel problème de la classe NP, ce qui serait une avancée considérable.

Il faut cependant tempérer notre enthousiasme, car la quantité d'ADN nécessaire augmente exponentiellement en fonction de la taille du problème, à cause de la première étape de l'algorithme.

Plus que la masse de la Terre !

Une machine à ADN qui résoudrait le problème pour un graphe de 200 sommets pèserait environ 10^{25} kilogrammes, soit plus que la masse de la Terre ! L'algorithme impose la mise à disposition d'un très grand nombre de types distincts de la molécule d'ADN qui participeront à la réaction, car il faut que chaque chemin possible de longueur n soit présent au départ. Or les préparer n'est pas simple et prend du temps.

De plus, avec les méthodes utilisées, il existe une probabilité non nulle que certains chemins ne soient pas représentés et donc que l'on passe à côté d'une solution du problème. Ces obstacles n'ont pas permis de

tirer du procédé de Leonard Adleman une méthode plus efficace que celles des machines électroniques pour le problème des chemins hamiltoniens.

Depuis la création de ce premier algorithme, tout un domaine scientifique nouveau s'est développé. Ses chercheurs se retrouvent périodiquement au colloque *DNA computing and molecular programming*, dont la 22^e édition s'est réunie en septembre 2016 à Munich. Toutes sortes d'algorithmes à ADN sont aujourd'hui maîtrisés et ont donné lieu à des mises en œuvre pratiques à petites échelles.

On a en particulier proposé des systèmes assimilables à des ordinateurs où l'on entre des données sous la forme de soupes moléculaires et où, après le calcul, on lit le résultat à nouveau en analysant les molécules fabriquées pendant les phases de réaction et les longues étapes de filtrage [voir l'article de Y. Benenson et al. dans la bibliographie]. On a aussi conçu un système qui joue au jeu tic-tac-toe, une sorte de morpion sur une grille 3×3 [voir « Des assemblages d'ADN rompus au jeu et au travail », *Pour la Science*, janvier 2009].

Aujourd'hui pour certains problèmes combinatoires particuliers tels que celui de l'affectation des tâches, les chercheurs, notamment Jian-Jun Shu et ses collègues à Singapour, espèrent que les ordinateurs à ADN seront plus efficaces que les ordinateurs classiques. Dans ce problème, on se donne une série de travaux à exécuter et d'agents aux compétences variables selon les tâches auxquels on les affecte (confier la tâche T à l'agent A à un coût C) et l'on cherche à savoir comment assigner une tâche à chaque agent en minimisant le coût total.

Plus surprenant, des algorithmes de multiplication rapide fondés sur des méthodes moléculaires ont aussi été décrits, notamment par Aran Nayebe, de l'université Stanford.

Théorie... et pratique ?

En théorie, la puissance des machines à ADN est énorme. Avec un litre de liquide contenant 5 grammes d'ADN, on disposerait d'une mémoire de plus de 10^{21} octets et d'une puissance de calcul évaluée à environ 10^{17} opérations par seconde, soit 100 pétaflops, ce qui est supérieur aux performances du plus puissant ordinateur actuel, le chinois Tianhe-2, qui ne réalise que 33 pétaflops.

Ce potentiel est cependant délicat à exploiter et aujourd'hui, malgré tous les efforts faits, les ordinateurs à base d'électronique sont bien plus faciles à programmer, calculent mieux et plus vite. On pourra peut-être les surpasser en utilisant des soupes à ADN et il faut continuer à explorer cette voie. Mais ce n'est pas facile, car du côté électronique les progrès sont réguliers et donc les systèmes à ADN luttent contre un concurrent de plus en plus difficile à rattraper !

Notons aussi qu'il n'est pas absurde de penser que, dans le cadre de la physique classique, un modèle de calcul qui résoudrait en temps polynomial les problèmes NP-complets ne peut exister matériellement. Les gains envisageables avec les calculateurs à ADN risquent donc d'être réservés à des problèmes particuliers et ne conduire qu'à des progrès limités, contrairement aux espoirs initiaux. Même les ordinateurs quantiques, d'après le spécialiste Scott Aaronson,

ne pourront pas résoudre en temps polynomial les problèmes NP-complets, qui semblent condamnés à rester hors de portée de tout ce que nous pouvons concevoir comme calculateurs. Ne rêvons pas trop !

Mémoires massives et durables

Récemment, en perfectionnant les méthodes de stockage d'informations par l'ADN, on a cependant réalisé des exploits qui montrent que les mécanismes de stockage de données mis au point par l'évolution du vivant pourraient être utiles et que la manipulation de l'ADN nous donnera sans doute les clés permettant la mémorisation, sur le très long terme, de quantités massives de données.

En effet, lorsqu'on souhaite garder pour des siècles des quantités importantes d'informations, on sait que ni les disques optiques (*Blu-ray* ou autres) ni les mémoires magnétiques (bandes, clés USB à mémoire flash, disques souples ou durs à plateau rotatif, ou disque SSD – *Solid-state drive*) ne sont vraiment fiables. À l'opposé, des expériences réalisées avec de l'information stockée sous forme de séquences d'ADN indiquent que l'on peut s'y fier.

En 2015, Robert Grass et son équipe, de l'École fédérale polytechnique de Zurich, ont inscrit des informations sur de l'ADN placé dans des matériaux inorganiques. Ils ont utilisé des codes correcteurs d'erreurs pour introduire de la redondance dans l'information inscrite et corriger à la lecture les petites dégradations susceptibles de se produire.

Leur expérience a consisté précisément à traduire 83 kilooctets (83 000 caractères) d'informations sur 4 991 segments d'ADN, chacun long de 158 nucléotides, qui ont été encapsulés dans de la silice. Des traitements de vieillissement accéléré ont été appliqués pour mesurer la vitesse de dégradation de l'ADN. L'information originale a été lue sans erreur, même après avoir chauffé la silice à 70° C pendant une semaine, ce qui est thermiquement équivalent au stockage d'informations sur l'ADN pendant 2 000 ans à température ambiante. On en a déduit que des données peuvent être archivées dans l'ADN pendant des millénaires.

D'autres expériences ont été effectuées et l'on s'est amusé à écrire dans de l'ADN des sonnets de Shakespeare, un passage de l'enregistrement du discours « I have a dream » de Martin Luther King et l'article scientifique célèbre de James Watson et Francis Crick sur la structure de l'ADN.

Le coût estimé de ces enregistrements est aujourd'hui de l'ordre de 10 000 euros par mégaoctet (1 million de caractères) pour l'écriture, et de 200 euros par mégaoctet pour la lecture. C'est assez cher, mais cela baisse rapidement avec les progrès des méthodes de synthèse et de séquençage. Nous avons donc de bonnes raisons de croire qu'à court terme, dans une dizaine d'années, nous disposerons de méthodes de stockage de ce type, compétitives et adaptées aux longues durées, qui exploiteront l'ADN ou peut-être d'autres molécules biologiques. L'équipe de Zurich imagine ainsi bientôt stocker le contenu de l'encyclopédie Wikipédia dans une archive à base d'ADN.

Des codes-barres moléculaires

On envisage aussi d'utiliser ces méthodes pour réaliser des codes-barres moléculaires marquant des objets ou des solutions, ce qui aiderait à protéger de la contrefaçon des parfums ou des molécules pharmaceutiques. On mélangerait de très petites quantités de molécules de synthèse où une sorte de signature aurait été inscrite. Un appareillage de lecture donnerait rapidement l'information qu'un produit est authentique ou, à l'inverse, qu'il est contrefait. On pourrait également tracer ainsi la provenance d'armes à feu et pister les trafiquants.

Tout récemment, en juillet 2016, James Bornholt, de l'université de Washington, et des chercheurs de Microsoft ont présenté une méthode d'archivage massif fondée sur l'ADN permettant des accès en lecture relativement rapide pour l'information qui serait ainsi stockée. Une méthode qui atteint un record de 200 mégaoctets d'informations enregistrées.

Les calculs de la capacité de ces technologies montrent qu'elles ont un potentiel supérieur, à moyen terme, à celui des technologies à base de silicium.