

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 2000 caractères.

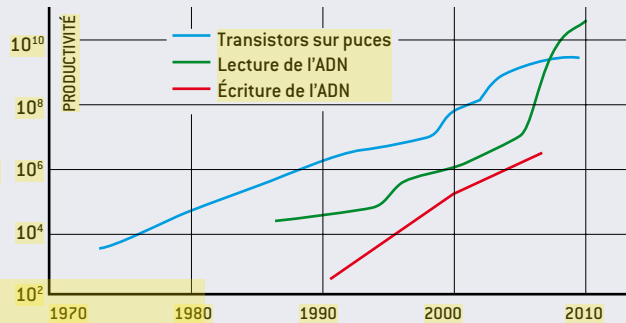
Si l'augmentation se poursuit au rythme des dernières décennies, nous serons incapables, par les moyens utilisés aujourd'hui, de répondre à la demande d'archivage de ces données.

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).