

SCIENTES • CHIMIE

Stockage des données numériques : les futurs data centers seront à l'échelle moléculaire

Des polymères synthétiques et de l'ADN pourraient permettre de conserver des données numériques, une alternative aux dispositifs actuels. En France, un programme du gouvernement vise à soutenir ces recherches innovantes, qui prendront encore plusieurs années.

Par Hanna Siemiatycki

Publié aujourd'hui à 05h00 • Lecture 10 min.

Article réservé aux abonnés



Dans le laboratoire de DNA Script, start-up française qui produit de l'ADN synthétique, alors abritée par l'hôpital Cochin, à Paris, le 6 décembre 2017. NICOLA LO CALZO POUR

« LE MONDE »

Chaque année, la quantité de données numériques produite par l'humanité croît de façon exponentielle. De 64 zettaoctets (64 000 milliards de gigaoctets) en 2020, on devrait monter à plusieurs milliers en 2035 ! Or, la multiplication des data centers – des infrastructures de stockage gigantesques – qui en découle pose de nombreux problèmes : emprise au sol et consommation énergétique faramineuses, obsolescence à environ cinq ans des supports de stockage (disques durs, bandes magnétiques, etc.) qu'il faut renouveler en permanence. La recherche d'alternatives compactes, pérennes et non polluantes s'impose.

Et si l'avenir du stockage des données numériques se jouait à l'échelle moléculaire ? Il ne s'agit pas là de science-fiction, mais d'un défi que se sont lancé des chercheurs français, il y a une dizaine d'années : celui de coder des données binaires sur des polymères synthétiques, comme ceux qui constituent nos plastiques ou autres objets du quotidien. Ces « polymères informationnels », comme ils les ont nommés, se situent au croisement de la chimie, de la biologie, de l'informatique et de l'électronique.

LA SUITE APRÈS CETTE PUBLICITÉ

Concrètement, les chercheurs utilisent les monomères, ces molécules le plus souvent organiques constitutives des chaînes polymères, comme un « langage » pour coder l'information binaire. Arbitrairement, ils définissent un monomère codant le bit-0, et un autre, dont la structure diffère, codant le bit-1. Le contrôle de leur agencement au moment de la synthèse (création de la chaîne polymère) aboutit à l'écriture de la suite de 0 et de 1 souhaitée. Pour déchiffrer ce message, le séquençage de la chaîne polymère permet de déterminer l'ordre des monomères.

« Trois fois mieux que l'ADN » synthétique

Jean-François Lutz, chimiste des polymères et directeur de l'Institut de science et d'ingénierie supramoléculaires (ISIS) à Strasbourg, où il vient d'installer son équipe, cherchait à ouvrir une voie différente de celle « *de biochimistes américains qui avaient commencé à stocker avec succès des données sur de l'ADN synthétique. Une "école de pensée" plus ancienne que la nôtre* », précise le pionnier des polymères informationnels.

Aussi a-t-il fallu d'abord maîtriser l'enchaînement des monomères lors de la synthèse, condition sine qua non à l'écriture du message binaire. Entre 2014 et 2016, les Français se sont ainsi appliqués à développer des approches combinatoires et itératives, donnant naissance à de nouvelles chimies. Jusqu'alors, les voies de synthèse traditionnelles n'assuraient qu'un contrôle très sommaire de la chaîne polymère. En parallèle de leurs recherches a progressivement émergé une concurrence désormais internationale, en Europe, aux Etats-Unis et en Asie.

Théoriquement, ces macromolécules sont prometteuses. Elles affichent une capacité de stockage largement supérieure aux disques durs, bandes magnétiques et mémoires flash. Car si 1 gramme d'ADN peut potentiellement stocker l'équivalent de plusieurs data centers, les polymères informationnels, eux, « *pourraient faire trois fois mieux* », selon leur inventeur. De plus, le défaut des plastiques qui mettent des millions d'années à se dégrader est ici un atout de taille. Ils ouvriraient la voie au stockage à très, très long terme ! Pour couronner le tout, leur conservation ne nécessite ni énergie ni contrôle de la température et de l'humidité.

Le Monde Application**La Matinale du Monde**

Chaque matin, retrouvez notre sélection de 20 articles à ne pas manquer

[Télécharger l'application](#)

Pour coder des quantités de l'ordre du kilooctet (des mots, une phrase, une petite image), la plupart des concepts de base – écrire, stocker, lire, effacer les données – ont été validés, rapporte M. Lutz. Mais pour passer du laboratoire aux applications, les verrous sont « énormes », dit-il : les vitesses d'écriture (synthèse) et de lecture (séquençage) sont bien trop lentes et coûteuses ; les données encodées sont difficiles à manipuler et à modifier – notamment à réécrire et à copier –, et lire de manière sélective certaines informations seulement est encore très ardu.

Cibler les « données froides »

Ces limites, son équipe tente de les surmonter dans le cadre d'un programme et équipement prioritaire de recherche (PEPR) lancé par le gouvernement en mai 2022. Doté d'un budget de 20 millions d'euros répartis équitablement en quatre volets sur sept ans, ce PEPR ambitionne de développer des dispositifs de stockage à long terme ciblant les « données froides », ces informations numériques très peu utilisées et rarement consultées, comme les archives audiovisuelles ou juridiques. D'où un partenariat avec la Bibliothèque nationale de France, l'Institut national de l'audiovisuel et le Parlement européen. Le volet que M. Lutz pilote se concentre entièrement sur les polymères informationnels – les trois autres volets se focalisant sur le stockage sur l'ADN synthétique, technologiquement plus avancé.

« Notre objectif est d'aller des preuves de concept bien établies vers la gestion de mégadonnées », annonce-t-il. Au programme : développer de nouvelles approches de synthèse, accélérer et paralléliser l'écriture et la lecture, et concevoir des supports de stockage bien réels. « En pratique, on repart de zéro. Financés par le PEPR, deux doctorants travaillent sur des synthèses de nouvelle génération », explique-t-il.

Pour eux, c'est l'engagement environnemental qui prime. « Je voulais absolument trouver un domaine applicatif. Ici, j'ai l'impression de réellement participer à réduire l'empreinte carbone des data centers », souligne avec entrain Benoit Pousse, 24 ans, chimiste organicien. Pour Hugo Guilmain, 26 ans, chimiste des macromolécules, c'est aussi « la partie importante, et vraiment motivante » de son travail.

Lire aussi | [Derrière l'IA, la déferlante des data centers](#)

A leurs côtés, Marc-André Delsuc œuvre aux confins de la chimie et de l'informatique pour élaborer des protocoles efficaces de codage et de décodage de l'information. « Je suis une sorte de "chimio-informaticien". Concrètement, j'écris des simulations : il faut introduire des erreurs à l'écriture puis vérifier notre aptitude à retrouver le message codé au moyen d'algorithmes de reconstruction », détaille le chimiste analytique.

A l'avenir, les dispositifs de stockage de polymères informationnels pourraient ressembler à de petites capsules métalliques, comme celles qui sont développées par la société française Imagene pour conserver de l'ADN synthétique. Mais le PEPR veut aussi réfléchir à des supports nanométriques plus complexes, « capables, au-delà du stockage passif, d'effacer, de manipuler, de lire localement »,

décrit M. Lutz. En ce qui concerne les « données chaudes », qui font appel à des dispositifs de mémoire dynamique et requièrent d'être accessibles instantanément, il se montre plus prudent : « *Nos polymères ne s'y prêteront peut-être pas.* »

Malléabilité structurelle

Pionniers incontestés de cette discipline, les Français mènent la danse au niveau international. Mais leurs concurrents, qui marchent dans leurs pas, tentent de les rattraper. Pour augmenter la capacité de stockage, des équipes se sont ainsi exercées à synthétiser le polymère codé le plus long possible. Depuis 2021, le groupe français détenait le record mondial : 440 bits restituant le portrait pixellisé d'Antoine Lavoisier, publié dans les *Comptes rendus chimie*. Mais, en février, des chercheurs sud-coréens les ont détrônés, en stockant 512 bits dans une seule macromolécule. L'article a été publié, avant évaluation par les pairs, dans *Research Square*. « *En nombre d'unités répétitives, c'est le record du monde. Mais, en matière de densité de stockage, nous sommes un peu en retard par rapport au P^r Lutz, confie Kyoung Taek Kim, professeur de chimie des polymères qui supervise ces recherches à l'université de Séoul. Nous voulions surtout changer la méthode de synthèse, pour limiter l'accumulation d'erreurs et minimiser le nombre d'étapes.* »

La densité de stockage est en effet l'un des avantages des polymères synthétiques vis-à-vis de l'ADN. Grâce à la flexibilité et à la grande diversité structurelle offertes par la chimie de synthèse, un monomère peut coder deux, trois, quatre bits, lorsque l'ADN est limité à son alphabet biologique, ne codant que deux bits par nucléotide. Cette malléabilité structurelle a aussi permis aux Français de faire état, en 2019, dans *Nature Communications*, de polymères informationnels « *photo-modifiables* » sous l'effet d'un stimulus lumineux. Des polymères lisibles ont pu être effacés : la lumière a modifié leur structure, induisant une suite de monomères identiques inapte à traduire un code binaire. Au contraire, d'autres polymères illisibles ont été révélés : un changement structurel sur des monomères a permis de différencier le bit-0 du bit-1. Mais, surtout, la lumière a permis de « réécrire » certains polymères, en transformant des monomères codant le bit-1 en bit-0. Des fonctionnalités potentiellement utiles pour protéger des données sensibles.

Pour stocker davantage d'informations, les chercheurs construisent aussi des bibliothèques de polymères. En 2018, l'équipe française a ainsi décrit dans la revue *Angewandte Chemie* une bibliothèque de seize polymères de quatre-vingts bits chacun, qui, superposés couche par couche, permettaient de coder une phrase de 1 280 bits (160 octets). Dans la même veine, une équipe chinoise publiait en 2019 dans *Nature Communications* sa faculté de coder un code-barres atteignant 3 672 octets grâce à une bibliothèque de dendrimères, des molécules à la structure arborescente, elles-mêmes conçues pour optimiser le stockage.

Après avoir évoqué l'écriture, place au défi de la lecture. A l'heure actuelle, la spectrométrie de masse en tandem (MS/MS) – méthode la plus mature pour déchiffrer les polymères informationnels en retrouvant l'ordre des molécules de la chaîne – fait face à de vives critiques : des instruments trop encombrants et coûteux (environ 500 000 euros), des temps de lecture trop lents et, surtout, une technique destructive qui nécessite de « casser » la molécule pour la lire.

Méthodes de lecture

Selon Roza Szweda, ancienne postdoc polonaise de l'équipe de M. Lutz, « *la plus grande faiblesse est de ne toujours pas disposer de méthodes efficaces pour lire les données codées* ». Rentrée en Pologne depuis quatre ans, avec « *l'ambition de constituer un groupe scientifique grand et fort* », elle a créé et dirige, depuis janvier, un laboratoire en partie consacré aux polymères informationnels, au sein de l'université Adam-Mickiewicz de Poznan. A la faveur d'un financement européen d'environ

2,3 millions d'euros, elle compte s'affranchir du spectromètre de masse : « *Nous travaillons à coder de l'information dans la lumière à l'aide de polymères fluorescents. En plus d'être réutilisables, ils sont lisibles instantanément, au moyen d'un fluorimètre, moins coûteux et beaucoup plus simple d'utilisation qu'un spectromètre de masse : plus compact, il peut être portable* », avance la chercheuse.

Pour Laurence Charles, spécialiste de la spectrométrie de masse appliquée aux polymères à l'université d'Aix-Marseille, pas question de s'en détourner. Partenaire du PEPR, elle collabore avec l'équipe strasbourgeoise pour la développer et l'optimiser. « *On met au point des molécules capables d'absorber la lumière UV directement dans l'appareil, afin de supprimer une étape d'irradiation très chronophage en amont de la lecture. Pour passer des minutes, voire des heures, aux millisecondes* », détaille-t-elle.

Autre piste, cette fois non destructive, explorée par les Français en collaboration avec l'Ecole polytechnique fédérale de Lausanne (EPFL) : le « séquençage nanopore ». Cette technique utilisée pour la lecture de l'ADN dispose déjà de dispositifs miniatures et portables. Sous l'application d'un courant, le passage du polymère informationnel à travers un nanopore (un trou à l'échelle du nanomètre) fournit un signal électrique permettant de reconstituer le code binaire. En 2020, les équipes ont montré que « *c'était possible, jusqu'à présent, sur des polymères très courts, de seulement quatre bits* », indique Matteo Dal Peraro, biophysicien et enseignant à l'EPFL.

« *C'est très compliqué, en termes de chimie. Mais en fournissant un gros effort de recherche, on pourrait remplacer la spectrométrie* », espère M. Lutz. Laurence Charles se montre plus nuancée : « *Aujourd'hui, ça fonctionne pour l'ADN parce que sa structure est figée. Il faudrait concevoir un nanopore unique pour chaque type de polymère. Mais peut-être que l'effort mérite d'être fait, car les nanopores permettraient de lire plusieurs chaînes en parallèle, ce qui n'est pas le cas de la spectrométrie de masse.* »

A Séoul, M. Kim s'essaie de son côté au séquençage par résonance magnétique nucléaire (RMN). En 2022, il a réussi à lire une image pixellisée de 192 bits. Une preuve de concept publiée dans *JACS Au*, « *qui ne fonctionne que sur des petites molécules* », rapporte-t-il. Laurence Charles est sceptique. « *Pour un appareil RMN, il faut compter trois fois le prix d'un spectromètre de masse, dit-elle. Et plusieurs heures pour obtenir un spectre bien plus compliqué à interpréter.* »

Modèle hybride

Quant à lever le verrou de la réécriture des données, la solution se trouverait-elle à l'intersection des polymères et de l'ADN ? A la tête d'une chaire de professeur junior créée à l'ISIS grâce au PEPR, Maria Nerantzaki dirige un groupe de recherche qui se concentre depuis deux ans sur le développement de macromolécules biohybrides. Et les résultats ne se sont pas fait attendre. En janvier 2024, son équipe a publié, dans *JACS*, une stratégie inédite de synthèse de polymères hybrides, composés d'une partie polymère codant l'information, liée chimiquement à un simple brin d'ADN.

En s'appuyant sur la propriété des simples brins d'ADN de s'autoassembler spontanément en double brin, les données peuvent être réécrites (remplacement d'un brin par un autre) ou effacées (retrait du brin). « *C'est la première étude de ce type jamais publiée dans le domaine* », se réjouit la chercheuse. Néanmoins, elle précise qu'« *il faudra encore beaucoup d'efforts de recherche pour arriver aux applications* ».

« *Ce papier a beaucoup de valeur. C'est une très bonne stratégie*, juge Erfane Arwani, PDG de Biomemory, une start-up française spécialisée dans le stockage de données sur l'ADN synthétique. *A l'heure actuelle, nous travaillons exclusivement sur l'ADN, mais de manière théorique, je pense qu'on ira plus loin avec du polymère non-ADN. Il serait prématuré d'essayer d'en commercialiser dès aujourd'hui*

pour le stockage massif de données. D'où le fait de démarrer avec l'ADN, un polymère mieux maîtrisé, pour ensuite se tourner vers un modèle hybride, avant de rejoindre du non-ADN complet. »

Tout le monde ne partage pas cet optimisme. Outre-Rhin, Michael Meier, professeur de chimie appliquée à l'Institut de technologie de Karlsruhe, doute de la capacité des polymères synthétiques à archiver les données : « *D'ici à dix ans, je ne pense pas que ce soit compétitif. Personnellement, je les vois plutôt performer dans des applications très coûteuses, là où le prix n'a pas d'importance, car il est avant tout question de sécurité. Pour l'étiquetage des billets de banque, par exemple. En ce sens, nous collaborons avec une entreprise dans le domaine de la lutte contre la contrefaçon. »*

La France bien lotie

Du côté des chercheurs strasbourgeois, c'est déjà chose faite. Depuis deux ans, une famille de leurs polymères informationnels – des polyuréthanes brevetés – est commercialisée par la société allemande Polysecure. Tels de véritables passeports numériques directement ancrés dans les matériaux, ils assurent la traçabilité des produits finis de différents secteurs. Car, ici, pas besoin d'une capacité de stockage trop importante pour incorporer un numéro de société, une date de péremption, une origine, ou encore une composition.

Au vu de leurs récentes publications, des chercheurs belges de l'université de Gand pourraient peut-être leur emboîter le pas : « *On perfectionne certaines chimies, pour ensuite entreprendre leur mise en œuvre dans plusieurs secteurs industriels* », indique le P^F Filip Du Prez, directeur de ces recherches.

Si la majorité des équipes internationales a obtenu des financements gouvernementaux, la France reste à ce jour un des pays les mieux lotis. « *Mais davantage d'investissements de la part du secteur privé seront nécessaires pour une réelle mise en œuvre* », commente Roza Szweda. De son côté, M. Kim, qui n'a reçu aucun soutien de l'Etat sud-coréen, ne désespère pas : « *Si le programme français connaît un véritable succès, nos responsables gouvernementaux pourraient probablement s'intéresser au domaine, et lancer un programme de ce type. Je l'espère. »*

Hanna Siemiatycki

Jeux

Découvrir

Mots croisés mini

Profitez tout l'été de grilles 5x5 inédites et ludiques, niveau débutant

Mots croisés

Chaque jour une nouvelle grille de Philippe Dupu