

Le séquençage des polymères numériques

Les polymères codés sont une nouvelle classe de macromolécules synthétiques dans lesquelles les unités monomères sont agencées en une séquence précise. Les avantages et applications potentielles de ces nouveaux polymères ont été décrits dans *L'Actualité Chimique* [1] ainsi que dans un article de vulgarisation récent [2]. Ces polymères sont synthétisés par des approches multi-étapes qui permettent de contrôler précisément l'enchaînement des monomères dans les chaînes formées. Les monomères peuvent donc être utilisés comme un alphabet pour écrire un message moléculaire. L'ADN est par exemple un polymère naturel codé basé sur un alphabet de quatre lettres : A, T, G et C. Mais le chimiste peut aussi développer des alphabets abiotiques pour préparer des polymères codés non naturels. Ainsi, en utilisant deux monomères différents, il est possible d'écrire un message binaire sur une macromolécule. On parle alors de polymères numériques [2]. Ces nouveaux polymères ouvrent des perspectives technologiques intéressantes car les unités élémentaires d'information qu'ils contiennent occupent un volume bien plus petit que les bits d'un disque dur traditionnel. De surcroît, ce sont souvent des solides ou des huiles qui sont stables sur de longues périodes à température ambiante. Ces nouveaux matériaux sont donc prometteurs pour des applications dans des domaines tels que le stockage de données, la lutte anti-contrefaçon ou la traçabilité de produits [2]. Toutefois, tous ces travaux sont encore au stade du laboratoire et beaucoup de points restent à améliorer. En particulier, la lecture de l'information inscrite sur une chaîne de polymère était, jusqu'à il y a peu, une opération chronophage et limitée à des séquences assez courtes. Cependant, des progrès significatifs ont été réalisés dans ce domaine au cours des deux dernières années. Ce sont ces avancées qui sont résumées ici.

Pour lire une macromolécule codée, on utilise une méthode dite de séquençage. Il s'agit d'une méthode analytique permettant de décrypter la séquence des monomères dans une chaîne de polymère. Par exemple, le séquençage est couramment utilisé en protéomique et en génomique pour déchiffrer la structure primaire des protéines et de l'ADN. Toutefois, les techniques de séquençage élaborées pour l'analyse des biopolymères ne sont pas toutes utilisables pour l'analyse des polymères synthétiques. Le séquençage par spectrométrie de masse en tandem et le séquençage par nanopores sont les méthodes les plus prometteuses pour la lecture de polymères numériques non naturels.

Séquençage par analyse nanopore

Le séquençage par nanopore est basé sur un principe d'électrophorèse et consiste à faire passer une chaîne unique de polymère codé, solubilisée dans un liquide conducteur, au travers d'un pore de diamètre contrôlé. Le passage de la macromolécule dans le pore génère des fluctuations de courant qui peuvent être mesurées et corrélées à la séquence de monomères. Cette mesure peut être effectuée au travers d'une protéine membranaire telle que l'hémolysine ou l'aérolysine, ou bien au travers d'un pore « solide », qui peut être un trou formé dans une surface de graphène ou dans un autre matériau inorganique plan. Le séquençage par nanopore a été d'abord conçu pour la lecture de l'ADN, mais il a fallu plusieurs décennies de recherche pour arriver à des systèmes commerciaux viables. L'extension de cette technique à l'analyse de

polymères numériques est beaucoup plus récente et de ce fait, un séquençage efficace n'a pas encore été démontré. Toutefois, la structure moléculaire des polymères codés peut être spécifiquement adaptée pour favoriser (ou défavoriser) les interactions de la chaîne avec le pore et ainsi faciliter sa lecture. Les premiers résultats allant dans ce sens ont été publiés très récemment [3].

Séquençage par spectrométrie de masse

La spectrométrie de masse en tandem (MS/MS) est à ce jour la méthode la plus efficace pour le séquençage de polymères numériques. Après injection d'une solution du polymère dans une source d'ionisation par électrobulbion (ESI) qui préserve l'intégrité des chaînes, une petite quantité de matière (quelques nanogrammes) est ionisée (mesure MS) puis soumise à des collisions (mesure MS/MS) qui engendrent des ruptures de chaînes. L'analyse des différents fragments permet de déduire la séquence codée du polymère étudié. La *figure 1* montre par exemple le séquençage d'un oligo(alcoxyamine amide) codé contenant la séquence modèle 10101. La masse de l'unité 1 étant intentionnellement légèrement plus lourde que celle de l'unité 0, il est très facile de lire la séquence en mesurant la différence de masse entre fragments. Nous avons montré que ce concept simple est applicable à un grand nombre de polymères synthétiques codés [4-5]. Nous avons aussi remarqué que les spectres MS/MS des polymères codés abiotiques sont souvent plus simples à analyser que ceux des biopolymères. Dans le cas de l'oligomère de la *figure 1*, les liaisons fragiles de type alcoxyamine N-O-C se fragmentent préférentiellement et conduisent à des spectres facilement déchiffrables.

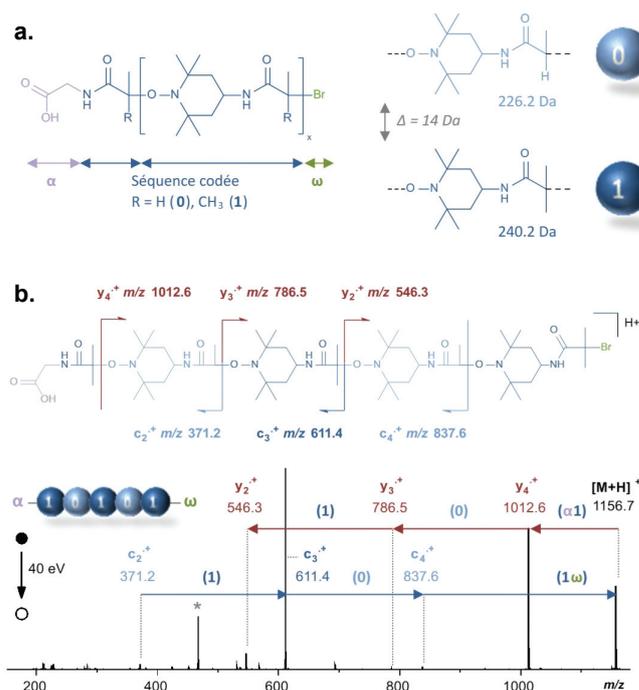


Figure 1 - Exemple de séquençage MS/MS d'un poly(alcoxyamine amide) : (a) structure générale du polymère codé ; (b) schéma de fragmentation et spectre ESI-MS/MS. * : fragment interne.

En jouant sur ces fragmentations sélectives, nous avons pu concevoir des macromolécules facilement « lisibles », ce qui est plus difficile à réaliser avec des biopolymères dont la structure chimique est imposée par la biologie. Ainsi, nous avons récemment décrit le séquençage de longues chaînes contenant jusqu'à 64 monomères codés [6]. Il est important de souligner que le séquençage MS/MS de polymères à séquences contrôlées contenant plus de vingt unités est en général ardu car les spectres obtenus sont très compliqués. En protéomique par exemple, le séquençage MS/MS de longues protéines n'est possible que si elles sont préalablement découpées en de petits peptides par hydrolyse enzymatique. Dans le cas des polymères codés non naturels, nous venons de montrer que de longues chaînes peuvent être directement séquencées dans un spectromètre de routine. Pour ce faire, des polymères numériques de type poly(phosphodiester) ont été conçus pour se fragmenter en deux temps (figure 2). Dans ces macromolécules, une liaison N-O-C a été incorporée entre chaque octet d'information (un octet est constitué de huit monomères codés). Dans des conditions MS/MS, ces liaisons fragiles se brisent prioritairement et la chaîne se fragmente donc en une série d'octets. Ces octets pouvant potentiellement avoir la même masse, chacun est marqué par un label spécifique (en bleu, figure 2) qui permet son identification. Chaque octet est ensuite séquencé individuellement (analyse MS³) et la séquence complète du polymère peut être reconstruite. Ces fragmentations en série sont toutes effectuées à l'intérieur du spectromètre en quelques minutes. Ainsi, en partant d'une macromolécule intacte, on peut lire sa séquence de manière complète (analyse *de novo*) en une seule mesure.

Automatisation du séquençage

Parmi les autres avancées récentes, nous avons développé, en collaboration avec l'équipe de Christine Carapito à l'Institut Pluridisciplinaire Hubert Curien de Strasbourg, un logiciel de lecture nommé MS-DECODER qui permet d'analyser automatiquement les spectres MS/MS et ainsi de déchiffrer l'information binaire stockée dans un polymère en quelques millisecondes [7]. MS-DECODER est un logiciel en libre accès qui peut être téléchargé par tout utilisateur potentiel [8]. Nous avons montré que ce logiciel permet de lire sans erreur les séquences binaires de nombreux polymères synthétiques et nous travaillons désormais à une extension qui permettra

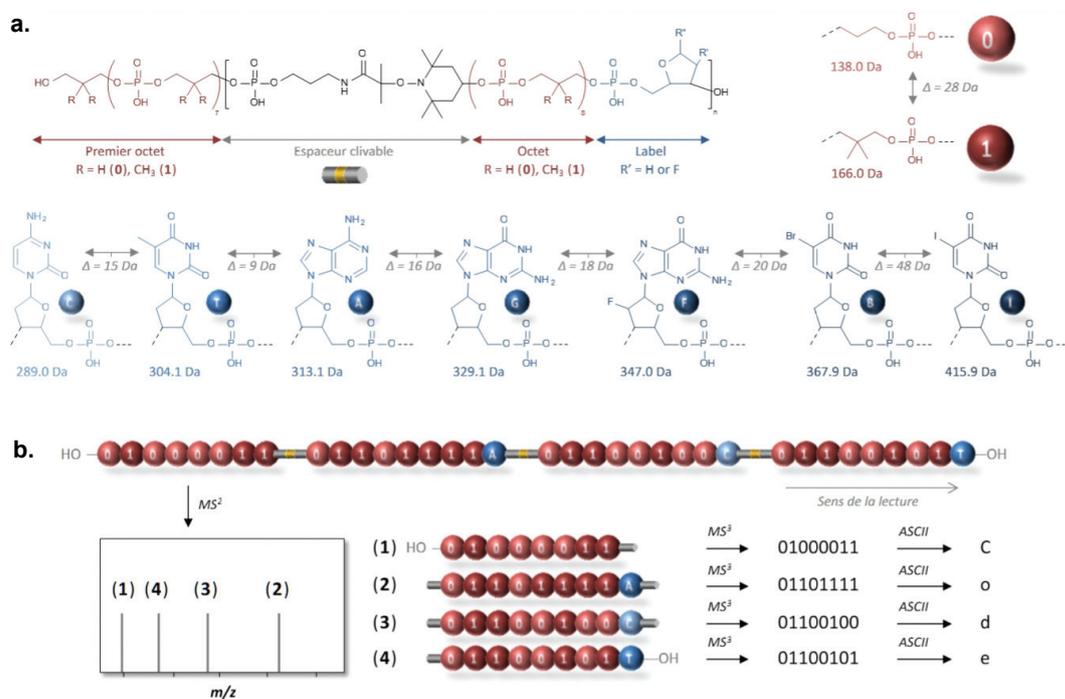


Figure 2 - Design macromoléculaire permettant la lecture de longues chaînes codées dans un spectromètre de masse de routine : (a) conception du polymère numérique ; (b) principe de la lecture par fragmentations successives [6].

de décrypter les longues chaînes codées décrites dans le paragraphe précédent.

En résumé, le développement de polymères codés est un domaine qui progresse très rapidement. Il est désormais non seulement possible de synthétiser de longues chaînes codées mais aussi de les lire facilement. Le prochain challenge consistera à passer du stade de la chaîne unique à des bibliothèques de chaînes, organisées en deux ou trois dimensions, permettant de stocker et de gérer la matière codée. Dans ce contexte, le prochain objectif pratique pour le séquençage sera de développer des méthodes permettant de lire directement un polymère incorporé dans un matériau solide sans préparation préalable de l'échantillon. Cette approche d'analyse ambiante et actuellement explorée pour ouvrir de nouvelles possibilités applicatives.

- [1] Lutz J.-F., Les polymères codés : une nouvelle propriété de la matière synthétique, *L'Act. Chim.*, **2016**, 404, p. 16.
- [2] Lutz J.-F., Plastiques binaires : l'avenir de la mémoire numérique?, *La Recherche*, mai **2017**, 523, p. 60.
- [3] Boukhet M. *et al.*, Translocation of precision polymers through biological nanopores, *Macromol. Rapid Commun.*, **2017**, 38, art. 1700680.
- [4] Roy R.K. *et al.*, Design and synthesis of digitally encoded polymers that can be decoded and erased, *Nat. Commun.*, **2015**, 6, art. 7237.
- [5] Gunay U.S. *et al.*, Chemoselective synthesis of uniform sequence-coded polyurethanes and their use as molecular tags, *Chem*, **2016**, 1, p. 114.
- [6] Al Ouahabi A., *et al.*, Mass spectrometry sequencing of long digital polymers facilitated by programmed inter-byte fragmentation, *Nat. Commun.*, **2017**, 8, art. 967.
- [7] Burel A., *et al.*, MS-DECODER: milliseconds sequencing of coded polymers, *Macromolecules*, **2017**, 50, p. 8290.
- [8] <https://msda.unistra.fr/public/MS-DECODER-1.4.2.zip>

Cette fiche a été réalisée par **Laurence Charles** (professeure, directrice adjointe de l'Institut de Chimie Radicalaire, Marseille, laurence.charles@univ-amu.fr) et **Jean-François Lutz** (directeur de recherche au CNRS, responsable de l'équipe Chimie Macromoléculaire de Précision et directeur adjoint de l'Institut Charles Sadron, Strasbourg, jflutz@unistra.fr).

Les fiches « Un point sur » sont coordonnées par un comité éditorial mené par Jean-Pierre Foulon et Séverine Bléneau-Serdel (contact : bleneau@lactualitechimique.org). Elles sont regroupées et en téléchargement libre sur www.lactualitechimique.org/spip.php?rubrique11.