



À chaque spectre de masse son hashtag

24 novembre 2016 | Andres Jordi
Catégories: Polluants | Eau potable

Des spécialistes des bases de données japonais, américains et européens ont créé un code qui simplifie l'accès à l'information de spectrométrie de masse en unifiant l'identification des spectres. En leur apposant une sorte de hashtag, le code appelé SPLASH, acronyme de « spectral-hash », facilite la recherche des spectres sur internet et permet de collecter et de comparer les informations disponibles sur un spectre donné dans toutes les bases de données. L'invention du code SPLASH vient maintenant de faire d'objet d'une publication dans la revue Nature Biotechnology.

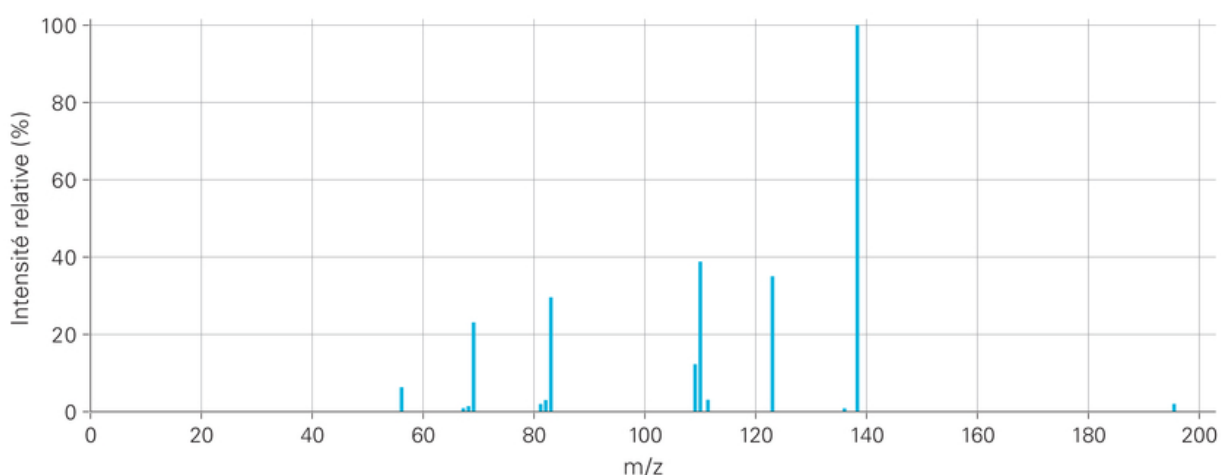
La spectrométrie de masse est une technique ultrasensible qui permet de détecter les composés chimiques en très faible quantité, qu'ils soient isolés ou en mélange. Sa puissance d'analyse est telle qu'elle pourrait détecter un morceau de sucre dans une piscine. En plus de mettre en évidence les substances déjà connues, la spectrométrie de masse permet aussi de déterminer la structure chimique de composés nouveaux. Depuis le développement des premiers spectromètres commercialisables dans les années 1950, les appareils et les méthodes ont été continuellement optimisés, de sorte que la spectrométrie de masse est devenue un instrument de recherche incontournable dans les domaines de la chimie, de la biologie, de l'environnement, du climat, de la médecine et de la médecine légale.

Contre la prolifération des données

Chaque jour, les experts du monde entier produisent des gigaoctets de données de spectrométrie de masse. Des millions de spectres sont actuellement stockés dans une vingtaine de bases de données, ce qui représente un volume de plusieurs millions de gigaoctets. Parmi ces spectres se trouvent plusieurs milliers de spectres de référence correspondant à des substances connues qui sont utilisés pour identifier les composés détectés au laboratoire par simple comparaison. Au-delà, les bases de données sont également alimentées en spectres de substances encore inconnues comme celles qui

sont aujourd'hui de plus en plus souvent détectées dans les végétaux, les champignons et les organismes marins. Les spectres sont toujours sauvegardés dans un format spécifique à chaque base de données, de sorte qu'il est par exemple impossible de savoir si une nouvelle substance inconnue X a déjà été décrite par ailleurs et si son spectre a déjà été enregistré. Cette situation handicape fortement les échanges d'informations entre scientifiques lorsqu'ils souhaitent, par exemple, discuter des propriétés de la substance X. C'est pour juguler cette prolifération des données de spectrométrie de masse que le système du code SPLASH a maintenant été imaginé.

Les programmes développés par les scientifiques du consortium international SPLASH sont en mesure de générer pour chaque spectre un code spécifique qui fonctionne comme un hashtag. Ce « marquage » rend les spectres repérables sur internet et permet de rassembler toutes les informations disponibles sur la substance dans les différentes bases de données. Avec leur code SPLASH, les spectres des substances encore inconnues reçoivent également un premier nom, ce qui facilite grandement les échanges à leur sujet.



Spectre de masse typique de la caféine. Le code SPLASH correspondant est :
splash10-000i-3900000000-73043667076aaf483c6e.
<http://mona.fiehnlab.ucdavis.edu/spectra/display/EA030313>

Du caractère indispensable des codes

Dans l'histoire des sciences, les chimistes se sont toujours heurtés à des problèmes de communication étant donné qu'une même substance était souvent connue sous différents noms, selon son lieu de découverte ou d'étude. La caféine, par exemple, a tout d'abord été nommée en référence au caféier, *Coffea arabica*, à partir duquel elle a été isolée pour la première fois. Cette substance est également connue sous d'autres noms tels que la 1,3,7-triméthylxanthine, la méthylthéobromine ou la théine. Dès le début du XXe siècle, l'UICPA (Union Internationale de Chimie Pure et Appliquée) entreprit d'unifier les principes de nomenclature, élaborant des règles, symboles et terminologies encore utilisés aujourd'hui dans le monde entier. D'après cette norme, le nom officiel et internationalement reconnu de la caféine est : 1,3,7-triméthyl-3,7-dihydro-1H-purine-2,6-dione. Ces règles standardisées de nomenclature sont particulièrement utiles pour la désignation des substances encore inconnues, même si elles reçoivent généralement des noms plus simples pour le quotidien de

la recherche.

Le code de l'UICPA est universel ; il est compris par les chimistes du monde entier. Il a cependant l'inconvénient d'être très long pour les molécules complexes, ce qui empêche quasiment de se faire une image de l'arrangement spatial des atomes. Les chimistes préfèrent donc la représentation graphique ou formule développée des molécules étant donné qu'elle livre des informations sur leur structure. Toutefois, si la représentation graphique est bien comprise des humains, elle n'est pas adaptée aux ordinateurs. Pour rendre les formules développées visibles dans les systèmes informatiques et les rendre repérables sur internet, l'UICPA a initié le développement de deux codes différents qui transcrivent les informations sur la structure des composés chimiques en chaînes de caractères déchiffrables par les machines. Ces codes, l'InChI-String et l'InChI-Key (InChI = International Chemical Identifier), fonctionnent à la manière de hashtags grâce auxquels les substances peuvent être retrouvées sur le Net. Les deux codes peuvent être générés pour toutes les molécules existantes avec un logiciel en accès libre. Les portails des substances chimiques et les bases de données en accès libre comme PubChem et ChemSpider ont, tout comme Wikipedia, déjà complété leurs informations du code InChI. Si un utilisateur entre par exemple le code de la caféine, ou même seulement une partie de ce code, dans un moteur de recherche, il se voit indiquer toutes les pages sur la caféine et a accès à la formule développée et à de nombreuses informations pertinentes pour les scientifiques.

Étant donné que chaque substance dispose non seulement d'une formule développée unique mais également d'un spectre de masse absolument spécifique, le développement du code SPLASH suit logiquement celui de l'InChI pour faire face à la prolifération des données et à la multiplicité des formats dans lesquels elles sont enregistrées.

Ce texte est basé sur un communiqué de presse de Sylvia Pieplow de l'Institut Leibniz de biochimie végétale (IPB) de Halle.

Links

Article original

Contact

Steffen Neumann
Institut Leibniz de biochimie végétale (IPB)
sneumann@ipb-halle.de

<https://www.eawag.ch/fr/portail/dinfo/actualites/news-archives/detail-de-larchive/a-chaque-spectre-de-masse-son-hashtag>