



## Jedem Massenspektrum sein Hashtag

24. November 2016 | Andres Jordi  
Themen: Trinkwasser | Schadstoffe

**Datenbankexperten aus Japan, Amerika und Europa haben einen Code entwickelt, mit dem es möglich ist, die Informationen von Massenspektren zu vereinheitlichen. Der Spectral-Hash oder Splash genannte Code erleichtert die Suche nach Spektren im Internet. Alle verfügbaren Informationen zu einem bestimmten Spektrum können mit diesem Spektren-Hashtag gezielt aus allen Datenbanken zusammengetragen und miteinander verglichen werden. Die Erfindung des Splash-Codes wurde jetzt in der Fachzeitschrift «Nature Biotechnology» veröffentlicht.**

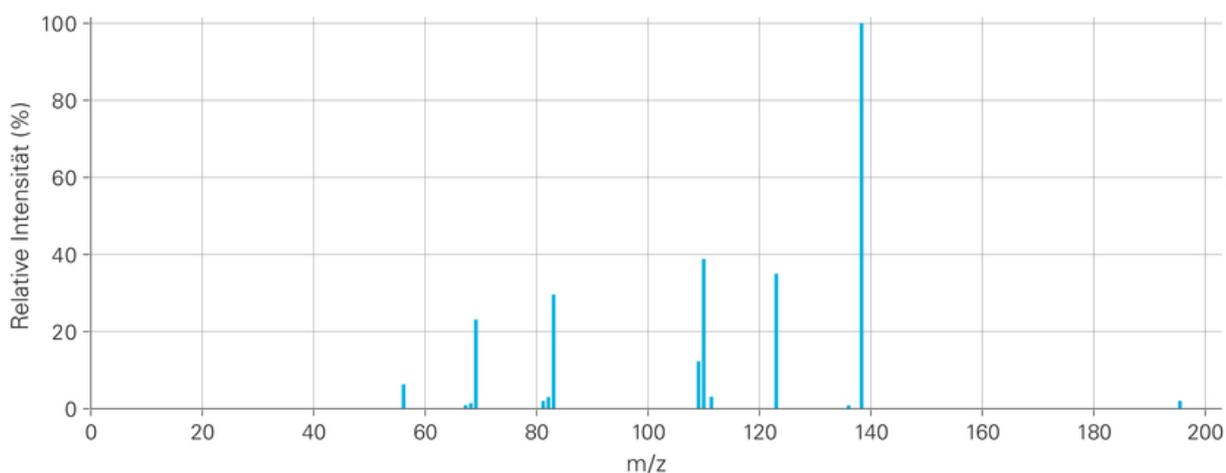
Die Massenspektrometrie ist ein hochsensibles Analyseverfahren, mit dem es möglich ist, geringste Substanzmengen – auch in Stoffgemischen – nachzuweisen. Die Technologie ist so leistungsfähig, dass man mit ihr ein Stück Würfelzucker in einem Schwimmbecken detektieren kann. Neben dem Nachweis bereits bekannter Substanzen, wird die Massenspektrometrie auch zur Strukturaufklärung neu entdeckter Verbindungen genutzt. Seit der Entwicklung der ersten kommerziellen Massenspektrometer in den 1950er-Jahren, wurden die Analysegeräte und Methoden ständig optimiert, sodass die Massenspektrometrie zu einem unentbehrlichen Werkzeug für die chemisch-biologische Grundlagenforschung, für Umwelt- und Klimaforschung, Medizin und Forensik geworden ist.

### Gegen den Datenwildwuchs

Weltweit tragen die Experten täglich Gigabytes an Massendaten zusammen. Millionen Spektren sind zurzeit in zirka 20 grösseren Datenbanken gespeichert – das entspricht einer Datenmenge von mehreren Millionen Gigabyte. Unter diesen Spektren sind mehrere tausend Referenzspektren bekannter Substanzen, auf die man bei Bedarf zum Vergleich der eigenen Messergebnisse zugreifen kann. Darüber hinaus werden die Datenbanken jedoch auch mit den Spektren noch unbekannter Substanzen gespeist, die man in jüngster Zeit vermehrt aus Pflanzen, Pilzen und marinen Organismen gewinnt. Die Speicherung der Spektren erfolgt dabei immer in dem jeweils datenbankspezifischen

Format, sodass zum Beispiel bei einer unbekanntem, noch namenlosen Substanz X nicht festgestellt werden kann, ob sie nicht schon an anderer Stelle beschrieben und als Spektrum gespeichert worden ist. Ein Informationsaustausch unter Wissenschaftlern, beispielsweise über wichtige Eigenschaften der Substanz X, wird dadurch erschwert. Diesem historisch gewachsenen Wildwuchs an Massendaten will man mit dem Splash-Code jetzt entgegenwirken.

Die von den Wissenschaftlerinnen und Wissenschaftler des internationalen Splash-Konsortiums entwickelten Programme können zu jedem vorhandenen Spektrum einen Code generieren, der ebenso wie ein Hashtag funktioniert. Dadurch werden Spektren im Internet nicht nur auffindbar, man kann zudem alle verfügbaren Substanzinformationen aus verschiedenen Datenbanken zusammentragen. Spektren von noch unbekanntem Substanzen erhalten mit dem Splash-Code ihren ersten Namen, was die Kommunikation über diese Stoffe extrem erleichtert.



Ein typisches Massenspektrum von Koffein. Der zugehörige Splash-Code lautet:  
*splash10-000i-3900000000-73043667076aaf483c6e.*

<http://mona.fiehnlab.ucdavis.edu/spectra/display/EA030313>

### Warum die Entwicklung von Codes unerlässlich ist

In der Geschichte der Wissenschaft standen Chemiker immer wieder vor Kommunikationsproblemen, denn gleiche Substanzen waren – je nach Entdeckungs- oder Erforschungsort – unter verschiedenen Namen bekannt. Koffein beispielsweise erhielt seinen Namen zunächst aus der Kaffeepflanze, *Coffea arabica*, aus der die Substanz zuerst isoliert wurde. Darüber hinaus ist Koffein unter einigen weiteren Namen wie 1,3,7-Trimethylxanthin, Methyltheobromin oder Thein bekannt.

Bereits im frühen 20. Jahrhundert sorgte die International Union of Pure and Applied Chemistry (IUPAC) für einheitliche Regeln für die Erstellung chemischer Nomenklaturen, Symbole und Terminologien, die bis heute weltweit angewendet werden. Demnach ist die offizielle international gültige chemische Bezeichnung von Koffein:

1,3,7-Trimethyl-3,7-dihydro-1H-purin-2,6-dion. Besonders bei der Benennung noch unbekannter Substanzen ist diese einheitliche Namensgebung hilfreich, obgleich sich immer auch zusätzliche Trivialnamen für den täglichen Umgang im Forschungsalltag etablieren.

Der IUPAC -Code ist allgemeingültig und wird von Chemikern weltweit verstanden; er hat aber den Nachteil, dass er besonders bei komplexen Verbindungen zu lang ist, um sich ein Bild über die räumliche Ausrichtung der Atome im Molekül zu machen. Chemiker bevorzugen daher die grafische Darstellung von Molekülen, die Strukturformel, da diese wichtige Informationen zum Aufbau der Verbindung enthält. Diese grafische Darstellung wird vom Menschen gut verstanden, von Computern hingegen nur bedingt. Um Strukturformeln mit dem Computer sichtbar und im Internet suchbar zu machen, wurden auf Initiative der IUPAC zwei verschiedene Codes entwickelt, die Strukturinformationen von chemischen Verbindungen in maschinenlesbare Zeichenketten umwandeln. Diese Codes, der InChI-String und der InChI-Key (von International Chemical Identifier) funktionieren wie Hashtags, mit denen die jeweilige Substanz im Internet wiederauffindbar ist. Beide Codes können für alle existierenden Verbindungen mit einer frei verfügbaren Software generiert werden. Öffentliche Datenbanken und Chemieportale wie Pubchem oder Chemspider, aber auch Wikipedia haben ihre Substanzinformationen um den InChI/InChI-Key erweitert. Gibt man den Code für Koffein oder Teile davon in die Suchmaschinen ein, so findet man alle relevanten Seiten zu Koffein, inklusive Strukturformel und vielen weiteren für Wissenschaftler interessanten Informationen.

Da jede Substanz nicht nur eine eindeutige Strukturformel, sondern auch ihr ganz spezifisches Massenspektrum aufweist, ist die Entwicklung des Splash-Codes die logische Fortführung des InChIs, die sich in der Konsequenz von wachsenden Massendatenmengen in verschiedenen Formaten ergibt.

*Der Text basiert auf der Pressemitteilung von Sylvia Pieplow vom Leibniz-Institut für Pflanzenbiochemie in Halle.*

## Links

Originalpublikation

## Kontakt

Steffen Neumann  
Leibniz-Institut für Pflanzenbiochemie  
[sneumann@ipb-halle.de](mailto:sneumann@ipb-halle.de)

<https://www.eawag.ch/de/info/portal/aktuelles/newsarchiv/archiv-detail/jedem-massenspektrum-sein-hashtag>