

Réflexions

Guy Ourisson

Institut de chimie des substances naturelles

L'INFORMATION MOLÉCULAIRE

CONSIDÉRATIONS SUR LA DOCUMENTATION EN CHIMIE

PRÈS de 500 volumes, chacun valant environ 1 000 à 2 000 F. C'est le *Gmelin*, l'un de ces *Handbücher* allemands si mal nommés et si utiles. Au moins, si l'on est assez riche, doit-on avoir, grâce à un bon investissement, tout ce qui est nécessaire pour une bonne documentation chimique ? Non bien sûr : il ne s'agit là que de chimie « inorganique » et cela exclut par conséquent le contenu d'un autre *Handbuch*, le *Beilstein*, qui, lui, ne contient que les informations relatives aux composés organiques du carbone, et ne comprend que 250 volumes (encore plus chers)... Il est vrai que plus de 5 millions de ces composés sont actuellement connus, et que, chaque semaine, s'y ajoutent une dizaine de milliers de nouvelles substances que les chimistes trouvent apparemment justifié de répertorier, de publier, de rendre accessibles par ordinateur, de vendre et de payer des fortunes. Y a-t-il un autre domaine de la science où une telle masse de documentation existe et soit indispensable ? Il ne semble pas.

L'archive essentielle

Il faut, avant tout, bien noter que ce ne sont pas *tous* les domaines de la documentation chimique qui présentent des caractéristiques « anormales » par rapport aux autres branches de la science, mais surtout les publications d'archivage de données.

Les publications primaires, aussi bien en écologie qu'en physique, en mathématique ou en linguistique, sont aussi importantes qu'en chimie : nombreuses revues nationales, en anglais, en volapük ou même en français, innombrables revues spécialisées, bien sûr en anglais, pérennes ou fugitives, toutes « indispensables » à qui veut, ou voudrait, se tenir au courant de ces domaines. Les livres, de même, ne doivent être en chimie ni plus abondants ni plus chers (certainement pas plus chers que les livres de médecine), ni plus indispensables statistiquement que dans les autres domaines : je connais bon nombre d'excellentes bibliothèques de chimie où le budget consacré aux livres n'atteint pas 10 % du total des dépenses.

Les publications de vulgarisation ou de pédagogie ne sont pas non plus anormalement abondantes ; par exemple, nos voisins d'outre-Rhin publient, pour les ensei-

gnants, l'excellent *Chemie in unserer Zeit*, qui n'est pas plus épais que les *Biologie in unserer Zeit*, *Physik, etc. in unserer Zeit* de la même série.

Les « organes de liaison », mensuels ou publications plus fréquentes publiées par les sociétés savantes à l'usage de leurs membres, ne sont, elles non plus, ni plus abondantes ni plus épaisses ni même plus intéressantes à conserver en chimie que dans d'autres disciplines.

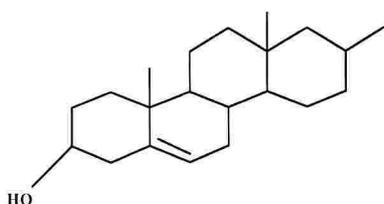
Ce qui caractérise la documentation en chimie, c'est essentiellement la valeur attachée à l'archivage, et à son corollaire, la recherche documentaire.

Qu'y a-t-il donc de si particulier à la chimie qui la rende si totalement dépendante de la documentation, des publications antérieures, de la capacité que nous avons — ou non — de retrouver une trace antérieure d'une recherche similaire, de résultats convergents ?

Bien entendu, comme dans toutes les sciences, il y a en chimie un poids extraordinaire de la notion de priorité. Il ne suffit pas d'avoir raison, il faut encore avoir eu raison le premier. Pourquoi s'en étonner ? Dans toutes les activités humaines, de la traversée de l'Atlantique à la défloraison, de la découverte de l'Amérique à l'invention d'une nouvelle façon d'apprêter le gigot de mouton, la notion de priorité est acceptée comme un des moteurs de notre ambition, de notre recherche de la gloire, quand bien même...

Création chimique

En chimie, il y a évidemment une dimension supplémentaire : c'est que « *la chimie crée son objet* ». Une journée de travail dans un laboratoire peut, sans que cela soit un exploit, conduire à la création d'une espèce moléculaire nouvelle, d'une substance qui n'avait jusqu'alors jamais existé dans l'univers, et qui pourra, selon les cas, rester une curiosité de laboratoire ou devenir un médicament contre une maladie jusqu'alors résistante, un parfum nouveau et jusqu'alors jamais humé ou un colorant d'une nuance jamais vue... Deux, dix, cent ans plus tard, il convient que la même sorte de matière, obtenue par un autre, puisse être identifiée : il faut un archivage de substances (d'où l'importance des collections qui existent dans cha-



que laboratoire), et surtout un archivage des modes d'obtention, de la description de toutes les propriétés caractéristiques, un archivage de la fiche d'identité de toutes les substances — des millions de substances déjà obtenues. Cet archivage doit permettre de retrouver cette publication princeps en partant de la seule propriété définitive et univoque : la structure moléculaire, ou, s'il s'agit d'un solide, la structure de son réseau. D'où l'importance des archivages dont l'accès est un code structural.

Il y a davantage : cette substance nouvelle, aux propriétés particulières, peut devenir objet de marché et être brevetable : son mode d'obtention, ses utilisations, selon les pays, peuvent être source d'argent, de création d'industries. L'archivage « accessible » devient source de profit — ou de pertes — et nécessité économique.

Comme nous sommes loin de la publication géniale de mathématique pure, ou de la description d'une espèce nouvelle de Flacourtiacée du sud de Madagascar...

Bref : la chimie produit des documents ponctuels, dont chacun doit devenir sans le moindre délai immédiatement accessible. D'où l'importance de ces immenses compilations que sont, chacun dans son domaine, le *Gmelin*, le *Beilstein*, le *Houben-Weyl*, le *Kirk-Othmer* et bien d'autres encyclopédies, mais aussi les *Chemical abstracts*, notre joie, notre fierté, et notre douleur.

connaît pas déjà. Mais n'importe quel chimiste, même s'il ignore que c'est la formule du cholestérol, peut déduire en l'apercevant du coin de l'œil que c'est sans doute une substance solide, qu'elle dérive probablement du métabolisme de polyterpènes, qu'elle doit réagir facilement avec l'anhydride acétique, que sa solubilité dans l'eau doit être quasi nulle, qu'elle doit donner un produit contenant un atome d'oxygène de plus si on la traite par certains oxydants, etc. Sa couleur ? aucun doute : elle doit être incolore. Son goût ? évidemment elle est insipide, etc. En fait, un tel « idéogramme » est immédiatement traduisible en termes de forme et de propriétés de molécule, et n'importe quel chimiste à peu près convenablement formé vous dira, en lisant la formule, quelles dimensions en auront les molécules et quelles en seront les

propriétés essentielles. On peut sans peine imaginer une discussion entre deux chimistes de nationalité différente, qui se comprennent parfaitement en écrivant des formules étonnantes au tableau, alors que chacun ignore la langue de l'autre.

La chimie s'écrit en idéogrammes. D'où une nécessité : indexer des idéogrammes. Et ceci conduit à tous les processus de recherches par le système DARC, par sous-structures, par arithmétique booléenne appliquée à des éléments structuraux. Nous nous émerveillons des Chinois, qui savent lire 6000 caractères. Mais le chimiste, lui, sait en lire plusieurs millions ! Et la difficulté des index des *Chemical abstracts* ou des systèmes documentaires en ligne vient de ce qu'il faut pouvoir « retrouver » chacune d'entre ces millions de structures.

Une difficulté supplémentaire vient d'une autre caractéristique d'une très grande partie de la chimie, qui ne décrit pas seulement des substances, mais aussi leurs transformations, leurs réactions. Nous tombons là dans une nouvelle dimension de la documentation chimique, très à la mode en 1987, mais pour laquelle nous ne disposons pas encore de système établi, accepté. Il faut

Formes bavardes

Si les publications chimiques n'étaient que cela, elles ne se distingueraient guère des autres publications scientifiques, si ce n'est, nous venons de le dire, parce qu'elles créent leur objet et ne se contentent pas de décrire ce que le monde leur impose. Mais, en fait, elles ont une autre particularité, sur laquelle Etienne, en son temps, avait insisté, et qui me semble d'une importance extrême.

La chimie, comme la Chine, s'écrit en idéogrammes. Une formule comme la suivante est tout aussi indéchiffrable par l'ignare qu'un idéogramme chinois.

Pas plus que ce dernier, elle ne peut être prononcée par qui ne la

R

réaliser une description univoque, l'archivage, puis la recherche des réactions chimiques : nature des transformations moléculaires qu'elles permettent, nature des réactifs, conditions opératoires, tout est important, tout doit être décrit, tout est brevetable si c'est nouveau, tout peut être vendu si c'est utile. Des livres comme le *Theilheimer (Synthetische Methoden)*, ou maintenant des programmes comme REACS, PAS-COP, ORAC, etc. sont placés sur ce créneau.

Un avenir non relié

Le plus difficile avec les prévisions, c'est de les faire à l'avance... Je mesure donc parfaitement le risque encouru à vouloir faire le profil de ce que va devenir la réalité de la publication en chimie à la fin du siècle. On peut cependant sans grand risque faire quelques prédictions.

D'abord, on peut être certain qu'en cette fin des années 80 la

multiplication des micro-ordinateurs changera complètement la nature physique des publications en rendant général le recours à des textes reproduits par photographie directe. Ceci exigera de l'auteur un changement d'habitudes : finies les corrections sur épreuves, les rectifications de dernière minute. Ce ne sera pas seulement « *Wysiwyg* » (« *what you see is what you get* »), mais aussi « *Wygiwyg* » (« *What you get is what you give* »)... Et passons sur les changements techniques évidents, puisqu'ils sont déjà partiellement réalisés : manuscrits fournis sur disquette, corrections et bons à tirer donnés par télécopie, etc.

Seconde prédiction, à peine plus risquée, concernant un autre aspect de la nature des publications : on pourrait diviser les publications scientifiques en deux catégories, celles que l'on lit au lit (ou dans l'avion, ou en regardant distraitemment « la Une »), et les autres. Les autres ne sont jamais feuilletées, on ne les lit pas, on n'en admire pas la beauté ou l'originalité, on y recherche des

renseignements; elles comprennent (outre les annuaires ferroviaires ou téléphoniques : serait-ce la raison du succès du Minitel ?) les *Beilstein* et autres *Chemical abstracts*. Les premières ne peuvent, pour l'instant du moins, être remplacées par un micro-ordinateur et ne le seront peut-être jamais. Les secondes le seront.

Ceci semble être une évidence, mais elle ne conduit malheureusement pas encore à des conclusions opérationnelles : il reste à savoir quel sera le matériel qui sera utilisé. Personnellement, je suis persuadé que l'avenir est, dans ce domaine, non aux banques de données centralisées, mais au disque optique numérique, consultable sur place. Je le crois, mais ne chercherai pas à vous en persuader : on peut fort bien penser au contraire que le chargement de données dans une banque centrale sera plus facile et que l'avenir ira plutôt aux formes multiples de télédéchargement. En tout cas, je ne peux pas imaginer que l'avenir soit aux séries de 500 volumes reliés.