

Le Centre d'Information Spectroscopique (C.I.S.)

par Mme D. Sandino et G. Emschwiller

Introduction

Le G A M S, association sans but lucratif, réunit les utilisateurs des méthodes spectroscopiques et physico-chimiques d'analyse, des constructeurs d'appareils scientifiques, des chercheurs, des ingénieurs de divers établissements ou laboratoires d'État, universitaires ou industriels.

Une des missions du G A M S est de fournir à ses adhérents de la documentation, principalement en spectroscopie. Celle-ci revêt plusieurs formes :

- publications périodiques scientifiques
- ouvrages spécialisés
- nouveautés en matière d'appareils scientifiques
- mais, surtout, dans certains domaines où la documentation joue un rôle essentiel pour l'analyse, le G A M S rassemble, depuis 1955, des collections de spectres d'origines diverses. L'ampleur de ces collections fait qu'elles constituent une sorte de banque de données spectrales dénommée : Centre d'Information Spectroscopique et Physico-chimique d'Analyse, par abréviation C.I.S.

I. Description de la banque de données

I. 1. Collections de spectres

Une collection de spectres est un outil indispensable pour de nombreux laboratoires mais sa constitution est devenue onéreuse. Dès 1955, le G A M S entreprit de rassembler de la documentation infrarouge à la disposition et avec le concours de ses adhérents (1). Mais on ne put guère recueillir qu'un millier de spectres par an, ce qui était notoirement insuffisant. On en vint donc nécessairement à faire appel à des éditeurs de données spectrales :

- Sadtler Research Laboratories
- Coblenz Society
- American Petroleum Institute (API)
- Documentation of Molecular Spectroscopy, d'origine allemande (DMS)
- Infrared Data Committee of Japan, etc...

Il faut noter que la publication de certaines collections a été arrêtée, ainsi DMS en 1974 et prochainement IR Data Committee of Japan ; de la sorte, la diversité des origines tend à faire place à une centralisation des données par les laboratoires Sadtler qui englobent dès à présent plus de la moitié de la documentation totale.

Le C.I.S. rassemble et diffuse, auprès de ses adhérents de l'information spectrale qui ne se limite pas à la spectroscopie infrarouge, mais s'étend à l'ultraviolet, au visible, à la spectroscopie Raman, à la Résonance Magnétique Nucléaire et à la spectrométrie de masse. Cependant la majeure partie de la collection concerne l'infrarouge.

Un bref inventaire de nos collections permettra de donner une idée de leur importance : nous possédons 200 000 spectres infrarouges. Il s'agit pour la plupart de spectres de corps purs, mais 31 000 d'entre-eux sont ceux de produits commerciaux de caté-

Groupement pour l'Avancement des Méthodes Spectroscopiques et physico-chimiques d'analyse, 88 Boulevard Malesherbes, 75008 Paris.

gories aussi diverses que les produits agricoles, biochimiques, tensio-actifs, alimentaires, monomères et polymères, etc. Dans la majorité des cas, le spectre est enregistré entre 2,5 et 15 μm domaine de l'infrarouge normal, mais nous possédons également 20 000 spectres dans le proche infrarouge (1 à 3 μm) et des spectres dans l'infrarouge lointain jusqu'à 50 μm .

Les premiers spectres édités étaient réalisés avec des appareils à prisme, mais actuellement 50 % des spectres ont été enregistrés avec des appareils à réseaux. Ces spectres ainsi obtenus sont mieux résolus et couvrent tous un domaine spectral plus étendu, allant de 3 μm à 40 μm .

Par ailleurs, nous possédons également :

- 28 500 spectres UV
 - 37 000 spectres RMN
 - 14 500 spectres de masse
- soit un total d'environ 300 000 spectres.

Les collections ainsi réunies s'accroissent régulièrement d'environ 10 000 spectres par an.

Le C.I.S., par l'ampleur de ses collections, particulièrement en infrarouge, semble unique en France et peut-être même en Europe (2). L'Institut für Spektrochemie und Angewandte Spektroskopie de Dortmund ne fait rien de comparable, bien que co-éditeur de la collection DMS comprenant environ 25 000 spectres infrarouges ; il poursuit la publication de catalogues où les spectres infrarouges sont associés aux spectres Raman, mais ne rassemble pas de documentation extérieure à son organisme. C'est surtout un centre de recherche.

I. 2. Catalogues et répertoires

Chaque collection de spectres possède ses propres catalogues alphabétiques, numériques et par formules moléculaires. Les catalogues Sadtler sont particulièrement complets puisqu'on y trouve réunies les références non seulement des spectres infrarouges Sadtler, mais aussi celles de la collection Coblenz, ainsi que les références des spectres ultraviolets et RMN de Sadtler.

Toutes les collections de spectres infrarouges, d'origines diverses ont été regroupées et classées par la Société ASTM (American Society for Testing and Materials). Des catalogues alphabétiques, numériques et par formules moléculaires regroupent 150 000 spectres : à chaque spectre est attribuée une référence qui comprend un numéro d'ordre suivi de deux (ou trois) lettres suivant les cas, décrivant son origine. La première lettre correspond à la collection du spectre par exemple : C pour Sadtler, A pour API, F pour DMS. La seconde décrit le domaine spectral couvert : A pour l'infrarouge normal et G pour l'infrarouge lointain. Par exemple, 180 CA est le numéro d'un spectre infrarouge standard dans la collection Sadtler. Pour les séries commerciales de Sadtler, une troisième lettre est introduite dans la référence, elle permet d'indiquer à quelle série commerciale appartient le spectre.

Enfin, un préfixe A est attribué aux spectres de la collection DMS correspondant à des produits inorganiques (AFA).

Seuls les laboratoires Sadtler, la Société Coblenz et API poursuivent des publications régulières, ce qui a amené la Société ASTM à envisager l'arrêt de ses compilations de données. Les laboratoires Sadtler seraient chargés de reprendre ce travail et surtout de continuer le classement des données spectrales relevées dans des publications scientifiques.

I. 3. Fichiers

Indépendamment de ces catalogues, ASTM a porté sur des bandes magnétiques tout un ensemble de données spectrales pour les 150 000 spectres répertoriés constituant ce qu'il est convenu d'appeler un fichier.

On y a porté les positions des bandes d'absorption infrarouges, codées avec une résolution de 0,1 μm dans un intervalle de 2 à 16 μm . Pour le lointain infrarouge, la résolution est de 1 μm .

Ce fichier contient également les renseignements sur les propriétés physico-chimiques des corps : point d'ébullition ou de fusion, composition chimique et les groupes fonctionnels, et enfin le numéro de référence du spectre.

Pour les 92 000 premiers spectres, le fichier ASTM a été transcrit sur des cartes perforées Wyandotte (3).

II. Critique des données

Outre ses fonctions de réunir, conserver et exploiter la documentation, le CIS s'est attaché, dans le passé, à corriger les erreurs trouvées dans celle-ci en revoyant avec minutie toutes les indications portées sur les spectres et sur les cartes perforées. Les erreurs étaient signalées à l'éditeur qui en assurait la correction.

Actuellement, cette vérification systématique n'est plus possible. En effet, le nombre sans cesse croissant des données spectrales de produits nouveaux et de plus en plus complexes ne permet qu'une vérification ponctuelle et sommaire.

Certaines collections, comme DMS et même IR Data Committee of Japan éditent les spectres infrarouges de manière peu lisible, le dessin étant beaucoup trop réduit. Heureusement les auteurs indiquent en parallèle les sommets des bandes d'absorption en nombre d'onde. Il y a lieu de leur préférer une autre collection lorsque les spectres figurent dans plusieurs. Dans certains cas, seule une partie du domaine spectral est couverte, ce qui rend, maintenant, le spectre à peu près inutilisable.

Enfin les conditions opératoires : la concentration du produit, le support utilisé, peuvent jouer un rôle important en modifiant la forme et l'intensité de certaines bandes. Ainsi certains composés réagissent avec le support, par exemple la pastille de bromure de potassium, et font apparaître des bandes supplémentaires qui seraient absentes avec d'autres supports, tel le nujol qui introduit lui-même certaines bandes entre 6,9 et 7,2 μm .

III. Utilisation des données

III. 1. Jusqu'en 1975, nous avons utilisé les cartes perforées pour procéder aux identifications de spectres infrarouges par tri mécanographique. Le procédé était long car le tri était réalisé bande par bande et demandait en moyenne 3 heures pour une recherche.

L'accroissement du nombre des données et donc du nombre des cartes perforées a fini par poser des problèmes de stockage en même temps qu'il entraînait une augmentation sensible de la durée du tri.

Enfin l'arrêt de la transcription sur cartes par ASTM nous a conduits à abandonner ce système et à recourir à l'utilisation de l'informatique.

III. 2. Plusieurs programmes de consultation des bandes magnétiques ASTM existent actuellement et sont commercialisés. Ils sont difficilement comparables car certains peuvent être adaptés sur un ordinateur maison, d'autres ne sont accessibles qu'en temps partagé, d'autres enfin uniquement par voie postale (4).

Dans tous les cas, le programme de recherche met à profit la structure spéciale du fichier qui se prête à l'utilisation d'instructions logiques du type «et» et «ou». Un spectre est représenté, par exemple, dans le fichier par un bloc de n bits, chaque bit représente une zone de 0,1 μm sur le spectre, des bits correspondant à un sommet de pic seront mis à 1, les autres à zéro. L'introduction d'une tolérance de 0,1 μm pourra être faite par glisse-

ment du spectre à droite puis à gauche et application de la loi du «ou».

Pour des conditions obligatoires, le programme pourra construire un masque constitué par un bloc de n bits tous nuls sauf dans une certaine zone correspondant à ces conditions obligatoires. Le spectre du fichier sera rejeté si le résultat de la loi «et» avec le masque est 0, sinon il sera conservé.

Un score est calculé pour chaque spectre satisfaisant aux conditions obligatoires, son calcul est en général complexe ; il tient compte des bandes signalées par l'utilisateur comme importantes, de la présence ou de l'absence de bandes supplémentaires et, lorsqu'une tolérance a été introduite sous forme de fourchette, il tient compte de l'écart entre la fréquence introduite et les fréquences du fichier.

Le numéro du spectre et le score correspondant sont rangés en mémoire pour permettre l'impression des résultats classés par ordre de score décroissant en fin de consultation.

III. 3. Citons parmi les principaux systèmes :

le système Eastman Kodak : il est le seul à utiliser toutes les données du fichier de base. De ce fait son bordereau de résultats est très complet, en particulier le nom des composés est inscrit en entier. Ce système de consultation, relativement lent, n'est disponible que par voie postale.

le système IRIS de Sadtler fonctionne en temps partagé aux USA.

Les résultats sont fournis sous forme de 3 tableaux :

- la liste des 20 meilleurs spectres avec leurs scores
- les propriétés physiques et chimiques codées pour ces 20 spectres
- les propriétés spectrales codées pour les 20 spectres.

le système SIRCH III de ASTM (fourni sous forme binaire translatable pour un IBM 1130 ou 360) ; il se caractérise par une grande vitesse de consultation. Les résultats comprennent les numéros d'identification et les propriétés physiques et chimiques correspondantes, sans indication du score.

le système FIRST I de DNA : Ce programme a été écrit par le même auteur que Sirch III et représente un progrès sensible. Les données peuvent être pondérées ou non, le programme en tient compte pour le calcul du score.

le système IRGO-SINGER (5) qui fonctionne en temps partagé aux USA et que nous utilisons par l'intermédiaire d'un terminal à la CEGOS. L'utilisateur doit répondre à un certain nombre de questions :

- l'introduction des données se fera-t-elle sous forme de fréquence ou de nombre d'onde ?
- l'indication des zones ou des bandes ne sont pas mesurées, pour ces zones le programme ne tient pas compte des données contenues dans le fichier
- zones d'absorption totale ; le programme ne conservera que les spectres ayant au moins une bande dans cette zone
- les pics obligatoires, le programme ne conservera que les spectres comportant les pics correspondants
- les pics principaux et les pics secondaires.

Il est possible à l'utilisateur de fixer le nombre maximal de résultats à lister sur le bordereau final, de préciser dans quelle subdivision du fichier la recherche doit être réalisée, de modifier la tolérance qui est normalement de 0,1 μm , de répondre à un certain

nombre de questions sur les propriétés physico-chimiques du composé.

le système INFRA 6 que nous utilisons également, mis aimablement à notre disposition par une importante société industrielle.

Les données fournies à l'ordinateur (bandes présentes et zones de non absorption) sont pondérées et créditées dans chaque groupe d'un certain nombre de possibilités d'erreurs.

Les recherches se font sur la totalité du fichier et le bordereau final fait état de 30 meilleurs spectres avec leurs scores respectifs.

III. 4. La principale critique à apporter à ces systèmes de recherche automatique vient du codage même des bandes d'absorption des spectres infrarouges. La Société ASTM n'a introduit aucune notion précise sur l'intensité relative des différentes bandes. Il peut en résulter des réponses erronées : les bandes d'absorption sont bien dans la position indiquée mais les intensités relatives ne sont pas respectées. A notre connaissance aucun système, sauf privé, ne tient compte de ce très important paramètre. Les programmes écrits par Fox (6) et Tanabe (7) sont encore à l'état expérimental et sont limités, pour le moment, à un nombre de composés trop restreint.

Un travail de dépouillement du bordereau fourni par l'ordinateur est donc nécessaire afin de visualiser les réponses et effectuer une comparaison effective entre le spectre inconnu et les spectres de référence sélectionnés par le programme de recherche automatique. Cela permet d'éliminer ceux dont la forme et l'intensité des bandes d'absorption sont trop éloignées du spectre étudié et d'aboutir à son identification.

Les programmes décrits ne sont pas prévus pour traiter des mélanges. Encore que certains auteurs R.W. Sebesta et G.G. Johnson (8) aient rédigé un programme capable d'identifier les constituants d'un mélange d'anthracène et d'acide benzoïque. Ce programme MIRET en est encore au stade expérimental.

Conclusion

En conclusion, comme on peut le voir, le CIS n'est peut-être pas vraiment une banque de données à part entière puisqu'il n'effectue le travail ni de saisie, ni de codage des données spectrales. Cependant, l'originalité du CIS réside principalement dans les possibilités qu'il offre à ses adhérents, à savoir :

- l'accès à une documentation spectrale très fournie, et ceci de manière ponctuelle ;
- l'identification de composés inconnus par leur spectre infrarouge à l'aide des techniques les plus modernes.

Bibliographie

- (1) E. Lœuille, Organismes d'études spectrographiques, Techniques de l'Ingénieur, 1962.
- (2) G. Emschwiller, Organismes d'études spectroscopiques, Techniques de l'Ingénieur, 1974.
- (3) ASTM, Codes and Instructions for Wyandotte ASTM, Punched Cards Indexing Spectral Absorption Data, May 1964.
- (4) Lachères, Champy et Normand, Compilation automatique des fichiers de spectres infrarouges par ordinateur, 31^e Congrès du GAMS, 5 au 9 juin 1972, Paris.
- (5) IRGO, Infrared Spectral Search by Computer Sharing. Instruction Manual.
- (6) R.C. Fox, Computer Searching of Infrared Spectra using peak location and intensity data. *Anal. Chem.* 1976, **48**, 717.
- (7) K. Tanabe, S. Saeki, Computer Retrieval of Infrared Spectra by a correlation coefficient method. *Anal. Chem.* 1975, **47**, 118.
- (8) R.W. Sebesta, G.G. Johnson, New Computerized Infrared Substance. Identification System. *Anal. Chem.* 1972, **44**, 260.