

cases the type of reaction, e.g., oxidation, reduction, halogenation, etc., is indexed per Alpha Number. But it should be mentioned that there is no syntax between the various partners of a reaction.

#### ACKNOWLEDGMENT

Thanks are due to the German Ministry of Research and Technology for supporting developmental work regarding the application of the system of journal literature.

#### REFERENCES AND NOTES

- (1) "Gebietsbeschreibung Anorganika", 2nd ed., IDC Internationale Dokumentationsgesellschaft für Chemie mbH Fachinformationszentrum Chemie, 6000 Frankfurt 90, Germany, March 1979, p. 3.
- (2) M. Isenberg and H. Kaltenhäuser, "Systematische Speicherung anorganischer Stoffe auf dem GREMAS-Ergänzungsband", Unveröffentlichter Bericht der IDC Internationale Dokumentationsgesellschaft für Chemie mbH, 6000 Frankfurt 90, West Germany, 1969.
- (3) W. Simmler, "Wissenschaftlich-technische Dokumentation in der chemischen Industrie—ein Beitrag zum Umweltschutz", In Deutscher Dokumentartag, 1976 (publiziert 1977), Verlag Dokumentation München, pp 164-77.
- (4) "Information über die anorganisch-chemische Literatur", *Nachr. Chem. Tech.*, **23**, 524-5 (1975).
- (5) H. Grünwald, "Information und Dokumentation auf dem Gebiet der anorganischen Chemie", *Nachr. Dok.*, **28**, 19-25 (1977).
- (6) "Magnetbandsystem Anorganische Chemie", *Nachr. Chem. Tech.*, **19**, 182-3 (1971); *Nachr. Dok.*, **22**, 172-3 (1971).
- (7) F. Ehrhardt, "Utilization of Chemical Abstracts Service (CAS) Data Bases. Application for the IDC-Inorganica-Dokumentationsystem", Bundesministerium für Forschung und Technologie, Report BMFT-FB ID 79-03, 1979, p. 53.
- (8) C. Weiske, "Das Klassifikationssystem des Chemischen Informationsdienstes", *DK-Mitt.*, **21**, 11-5 (1977).
- (9) C. Weiske, "Chemical Information Services", *Angew. Chem., Int. Ed. Engl.*, **9**, 550-5 (1970).
- (10) E. Meyer, R. Jansen, and E. Sens, "Das IDC-Thesaurus-System", *Nachr. Dok.*, **23**, 203-11 (1972).

## Interactive Simulation of Infrared, Mass, and $^{13}\text{C}$ NMR Spectra

M. RAZINGER,\* J. ZUPAN, M. PENCA, and B. BARLIČ

Boris Kidrič Institute of Chemistry, 61001 Ljubljana, P.O.B. 380, Yugoslavia

Received October 24, 1979

As a special part of a combined chemical information system,<sup>3</sup> the update of spectral data files is presented and discussed. The amending of infrared spectra in general is much more complicated than that of the MS or  $^{13}\text{C}$  NMR ones. The problems of developing an interactive simulation (updating) of IR spectra to make them as similar to the experimental ones as possible are elaborated and solutions suggested. The organization of records in spectral data files and data flow during the update procedure are presented in detail.

#### INTRODUCTION

Chemical information systems are quite numerous today.<sup>1</sup> They differ in the number of spectrometries upon which they are based, in the extent and quality of the spectral data bases, in the possibility for interactive work, etc. However, one thing is common to all: the quality of answers delivered depends strongly upon the quality of the spectral data on which they are based.<sup>2</sup> The flexibility of manipulation, update, and transformation of the data bases is quite a significant factor in the quality of the whole system.

As the main features of our combined chemical information system are presented elsewhere,<sup>3,4</sup> in this paper mainly the update and manipulation of spectra in the data files and also some aspects of graphical representation of spectral data are discussed. One of the main tasks in constructing the updating procedure was to make possible the manipulation of spectral data with simple instructions. Thus the user can easily modify the spectra already in the data banks, add new ones, or form new files.

Only by constant checking and improving of the quality of the data in the spectral files can we obtain better graphical representation of spectra and better performance of the information system.

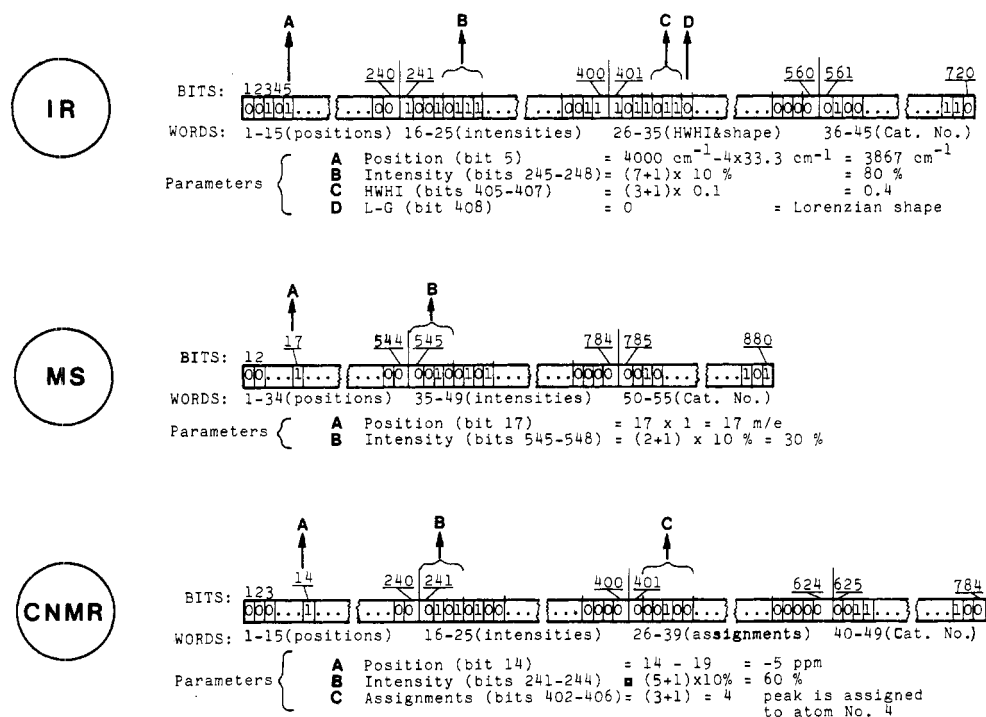
#### GRAPHICAL REPRESENTATION OF SPECTRA

Although the most vital information for the successful operation of an automatic spectral search system and for the graphical representation of spectral data files is the position of peaks in the spectrum, it is by no means the only important

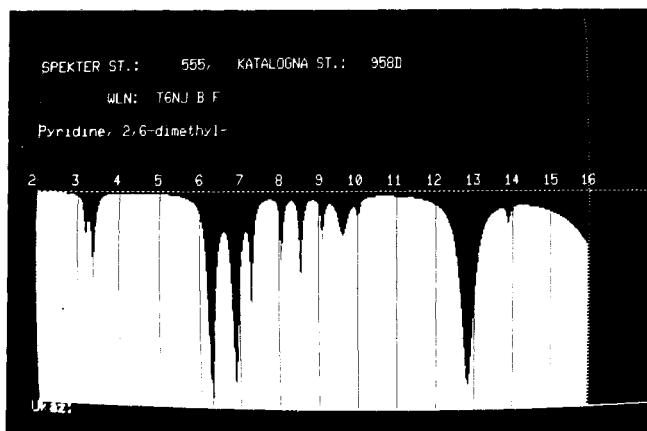
information. For the mass and most  $^{13}\text{C}$  NMR spectra the other important information needed besides peak position is peak intensity. Spectra of this type can be represented as a set of discrete lines of known positions and lengths which are adequate for obtaining quite useful approximations of the real spectrum. In infrared spectrometry the situation is quite different. Taking a look at an IR spectrum, one can hardly say that the peak positions and intensities completely define it. Band shapes are helpful in recognizing particular group motions, and the judgment of an experienced spectroscopist relies strongly on them.

The best way of solving the problem of adequate representation of real IR spectra is to digitize them.<sup>2,5</sup> This process requires direct connection of the spectrograph to the computer, i.e., online data acquisition, digitalization of spectrograms, or cataloged spectra with a digitizer. Either way requires specialized instruments which are not commonly found in many laboratories, while minicomputers themselves are increasingly used.

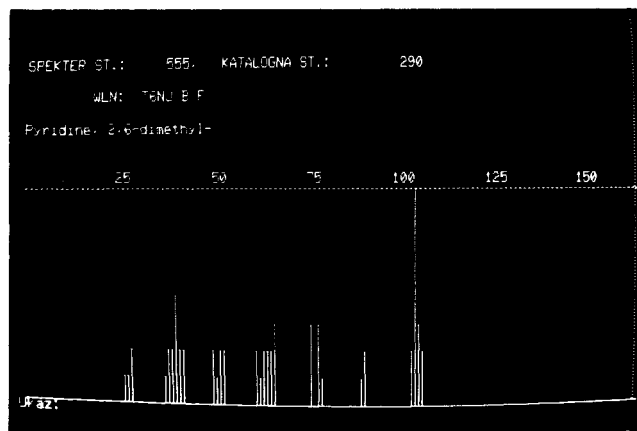
Instead of storing the digitized values of the whole spectrographic curve in the spectral data files, only some main parameters of each peak were stored for our system of data files. Peak positions and intensities are stored for all three spectrometries. In the case of  $^{13}\text{C}$  NMR spectra, carbon atom assignment number is associated with each peak as the additional parameter. For IR spectra, two additional parameters are stored. The first one is the half-width at half-intensity of the peak (HWHI), and the second parameter determines the shape of the peak in question. The value of this second parameter can be only 0 or 1, defining the function for calculation



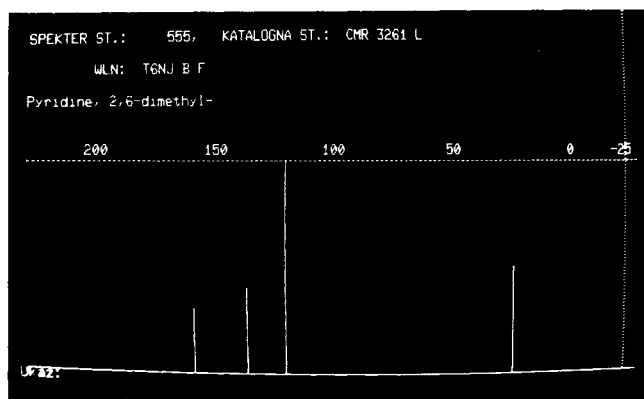
**Figure 1.** The organization of binary records in spectral data files. All actual parameters of one peak from each shown record are given explicitly. The bits from which these parameters are calculated are marked with arrows.



**Figure 2.** IR spectrum of 2,6-dimethylpyridine included in the spectral data bank as retrieved and displayed by the system.



**Figure 4.**  $^{13}\text{C}$  NMR spectrum of 2,6-dimethylpyridine included in the spectral data banks as retrieved and displayed by the system.



**Figure 3.** Mass spectrum of 2,6-dimethylpyridine included in the spectral data banks as retrieved and displayed by the system.

of the peak curve as Lorentzian or Gaussian, respectively.

The organization of the storage of these parameters within the records for the different spectral data files is shown in Figure 1. For mass and  $^{13}\text{C}$  NMR spectrometry, the spectrum

**Table 1.** List of All the Commands in the UPDATE Option with Their Respective Functions

command	function
spec	selection of the spectrometry; valid abbreviations are IR, MS, and CNMR
n	display of numerical parameters of the $n$ th spectrum in the previously specified file
n, m	display of numerical parameters of the $n$ th spectrum in the $m$ th file
ADD	initialization of the input of a new spectrum
MODIFY	modification of a spectrum
PEAK, n	modification of the $n$ th peak only
END	end of modification or addition of parameters
STORE	writing new or modified spectra on the physical device
ERASE, n, m	clearing the $n$ th record in the $m$ th file
IMAGE	display of the simulated spectrum on the screen
NAME	update of the chemical name
WLN	update of the WLN
OUT	exit from the option UPDATE
SOS	extended comment on how to use UPDATE

is simulated directly by plotting the input parameters. For IR spectrometry, the curve of the spectrum is first calculated

```

a Option?
UPDATE
Are you working on a graphic terminal?
NO
Instruction:
IR
Instruction:
4.14
      No. of Peak   Position   Intensity   Halfwidth   Shape
      1             770         50          1           1
      2             720         50          1           1
      3             670         50          1           1
Instruction:
MODIFY
Which peak are you changing (or generating)?
PEAK,1
This is peak no. 1: 770.0 50.0 1.0 1.0 Change: 70,3
Which peak are you changing (or generating)?
PEAK,3
This is peak no. 3: 670.0 50.0 1.0 1.0 Change: 660,100,2
Which peak are you changing (or generating)?
END
      No. of Peak   Position   Intensity   Halfwidth   Shape
      1             770         70          3           1
      2             720         50          1           1
      3             660        100          1           2
Instruction:
STORE,14
Spectrum no. 4 is stored in file no. 14
Instruction:
OUT
TTO -- STOP

b Option?
UPDATE
Are you working on a graphic terminal?
NO
Instruction:
CNMR
Instruction:
ADD
Spectrum is not yet defined. Use the instruction MODIFY!
Instruction:
MODIFY
Which peak are you changing (or generating)?
1
This is peak no. 1: 0.0 0.0 0.0 Change: 15,70,9
This is peak no. 2: 0.0 0.0 0.0 Change: 30,6,90,6
This is peak no. 3: 0.0 0.0 0.0 Change: 34,6,90,7
This is peak no. 4: 0.0 0.0 0.0 Change: 56,70,4
This is peak no. 5: 0.0 0.0 0.0 Change: 183,30,2
This is peak no. 6: 0.0 0.0 0.0 Change: END
      No. of Peak   Position   Intensity   Assignment
      1             15          70           9
      2             31          90           6
      3             35          90           7
      4             56          70           4
      5            183          30           2
Instruction:
NAME
Enter chemical name: Methionine
Instruction:
WLN
Enter WLN: QVYZ2S1
Instruction:
STORE,3
Spectrum no.23 is stored in file no. 3
Instruction:
OUT
TTO -- STOP

```

**Figure 5.** A computer printout of actual update procedures (a) MODIFY of an IR spectrum and (b) ADD of a CNMR spectrum. In this example, IR spectrum number 4 is read from the IR data file (file number 14 in the present organization) and all the parameters of all peaks are printed out for inspection. After using the command MODIFY the program asks for the number of peak to be changed (all the operator's entries are underlined). The new parameters are entered and the image of the modified spectrum can be displayed and inspected.

according to the input parameters as the sum of the analytical functions representing individual peaks and then displayed on a graphic videoterminal (DEC, VT55). Particularly for IR spectra, only the image simulated in this way makes possible visual comparison with the experimental spectrum. Figures 2, 3, and 4 are examples of IR, mass, and  $^{13}\text{C}$  NMR spectra (of the same compound) displayed on the videoterminal.

#### UPDATE AND MANIPULATION OF SPECTRAL DATA

Of the many options in our information system,<sup>3,4</sup> only the option UPDATE<sup>6</sup> will be discussed in detail. Table I gives the complete set of the commands which can be issued within this option. The commands are mainly self-explanatory, but when in doubt about the use of any of them, the user can always type the SOS command to get a more precise explanation of the functions of each command. For the updating procedure three commands besides the others, namely,

MODIFY, ADD, and STORE, are the most important.

MODIFY serves for the modification of the spectra already included in the data bank, i.e., for the alteration of various parameters of a chosen spectrum, for the inclusion and/or deletion of new peaks, etc. An example of the use of the command MODIFY is given in Figure 5a. The process in this figure can be repeated until the simulated spectrum looks as similar to the experimental one as possible. At the end of the modifying procedure all spectral parameters are displayed again for the final check. Using the command STORE and the number of the same file from which the initial spectrum was read, the modified record is written over the old one.

With the commands ADD and STORE, the parameters of the peaks of a spectrum under consideration can be written into the first encountered empty place of the predefined data file. Otherwise, the procedure is identical with that explained for the command MODIFY. An example of adding a new  $^{13}\text{C}$  NMR spectrum to the data bank is given in Figure 5b.

Chemical names of the compounds and their Wiswesser linear notations<sup>7</sup> are stored in separate files and can be updated with appropriate commands (see Table I). The data flow in the updating procedure is presented schematically in the Figure 6.

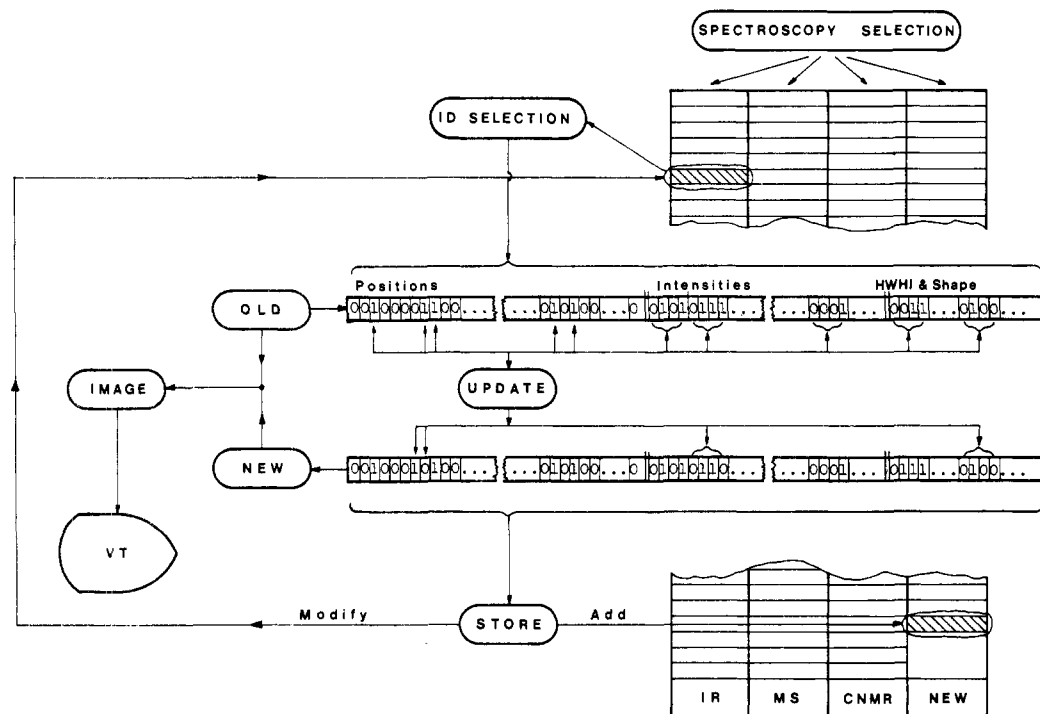
Figures 7, 8, and 9 illustrate a practical application of the modification procedure in a real search. Let us take the identification of an unknown compound using its IR spectrum as an example. With the help of the PEAK SEARCH option<sup>4</sup> one or more spectra with requested peak positions can easily be retrieved. It is evident that no positive identification by visual comparison is possible if the spectra from the retrieved list are represented by their peak positions only (Figure 7). If the spectrum in question is modified at its first appearance on the screen using the procedure shown in Figure 5a, it is made very similar to the cataloged spectrum (Figure 9). Then all further identifications will be very easy, comparing experimental spectrum with the result of modification shown in Figure 8.

#### ABOUT IMPLEMENTATION

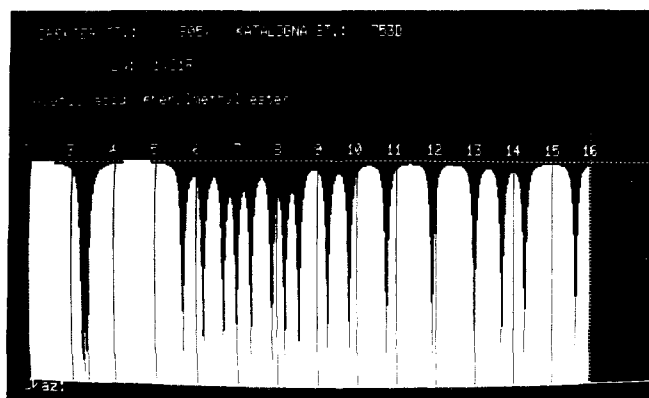
The entire information system including UPDATE routines is implemented on a PDP-11/34 minicomputer and requires 30K words of memory. The data bank (inverted and master files) for 1016 compounds needs about 1 Mbyte of mass storage on disk. The programs are written completely in the interactive DEC Fortran IV language and are packed into a three-level overlay structure to achieve better response time vs. space performance. The system is multiuser oriented.

In the present state, three master files of mass,<sup>8</sup> infrared,<sup>9,10</sup> and carbon nuclear magnetic resonance<sup>11</sup> spectra for each of the 1016 compounds are constantly used by the system. The chemical name, structure, and mass, infrared, and  $^{13}\text{C}$  NMR spectra of any compound are connected via the same identification number and can be accessed from any option in the system.

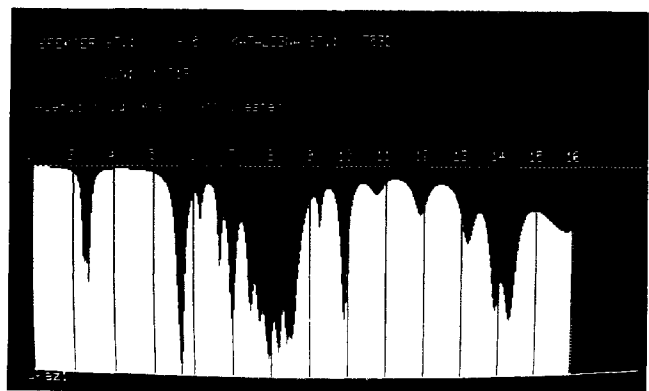
The instruction set used by the system consists of a list of simple statements, e.g., IMAGE, MODIFY, etc. In the operative version of the system, all instructions can be either in Slovenian or in English, and they could be changed easily to another language. The instructions are used on a two-level basis. The upper level instructions determine the option within the system, e.g., UPDATE, PEAK SEARCH, STRUCTURE GENERATION, etc. The lower level instructions are actual commands for some specific operation within an option, e.g., IMAGE, ADD, etc. The functions of some commands are dependent on the option from which the particular command is issued; e.g., the command IMAGE issued from the option STRUCTURE GENERATION causes a structure image (consisting of characters, dots, hyphens, etc.) to be printed or



**Figure 6.** The data flow in UPDATE option. The organization and update of an IR record which has more complicated structure than the MS and  $^{13}\text{C}$  NMR ones are shown.



**Figure 7.** Image of an IR spectrum on the screen. Of the four possible parameters, only the positions are entered while the other three parameters have default values. Slight differences in the intensities are due to peak overlapping.



**Figure 8.** Image of the spectrum in Figure 7 after the update.

displayed on the screen, while the same command typed in UPDATE option displays the spectrum on the screen of a (graphic) terminal.

In the present form, the records of the spectral data files are 45, 49, and 55 (16-bit) words long for IR,  $^{13}\text{C}$  NMR, and MS files, respectively. Those lengths permit storing parameters for 40 peaks in IR and  $^{13}\text{C}$  NMR records and for 60 peaks in the MS record. The limits can be extended easily by simply changing the parameters in the data file parameter table. All the UPDATE option programs consist of about 660 Fortran IV statements.

## CONCLUSION

Our aim to enable the user to simulate images of the spectra (especially the infrared ones) practically identical with the experimental spectra from a few parameters for each peak was achieved at least for not too complicated specimens. An advantage of this approach over the alternative possibility of the digitalization of spectra is a drastic reduction of data file size. In our solution the size of an infrared record is 45 words in contrast to the digitized data files of at least 320 tight-packed words needed for each IR record. Another advantage of our approach is that each user can generate easily his own spectral data bank according to his interests and needs without having to have access to specialized instrumentation and to be a skilled programmer.

An illustration of the ability of the UPDATE option is the fact that recently more than 10 000 intensities in over 700 IR spectra were checked and changed by two operators in five days, using this option.

## ACKNOWLEDGMENT

The authors thank Miljana Stevanović for help and gratefully acknowledge the financial support of ILIRIJA-VED-

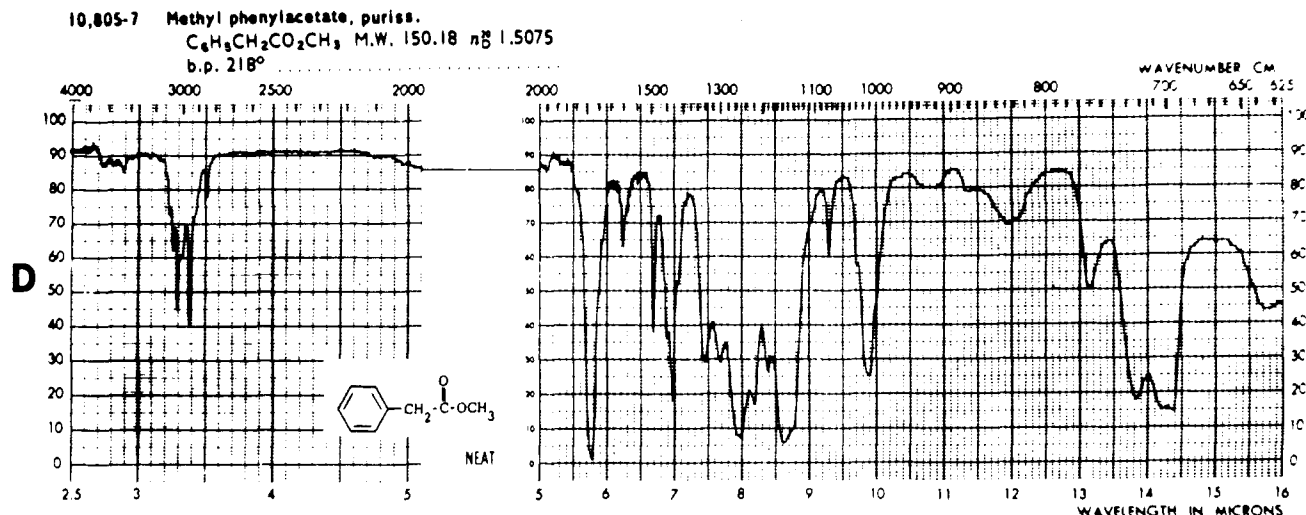


Figure 9. The spectrum of the compound in Figures 7 and 8 as given in the Aldrich catalog.<sup>9</sup>

ROG, Chemical and Cosmological works, Ljubljana, and the Research Community of Slovenia.

#### REFERENCES AND NOTES

- (1) Zupan, J. *Anal. Chim. Acta* **1978**, *103*, 273.
- (2) Clerc, J. T.; Zupan, J. *Pure Appl. Chem.* **1977**, *49*, 1827.
- (3) Zupan, J.; Penca, M.; Razinger, M.; Barlič, B.; Hadži, D. *Anal. Lett.* **1979**, *12*, 109.
- (4) Zupan, J.; Hadži, D.; Penca, M.; Razinger, M.; Barlič, B. *Anal. Chim. Acta* **1980**, *122*, 103.
- (5) Penca, M.; Zupan, J.; Hadži, D. *Kem. Ind.* **1978**, *27*, 137.
- (6) Razinger, M.; Zupan, J.; Penca, M.; Barlič, B.; Hadži, D. *Proceedings of the World Spectroscopy Conference*, Lisbon, 1979, in press.
- (7) Smith, E. G.; Baker, P. A. "The Wiswesser Line-Formula Chemical Notation", 3rd ed.; CIMI: Cherry Hill, N.J., 1975.
- (8) Aldermaston Mass Spectrometry Data Collection, Reading, U.K.
- (9) The Aldrich Library of Infrared Spectra, Pouchert, C. J., Ed.; Aldrich Chemical Inc.: Milwaukee, Wis., 1970.
- (10) Wyandotte-ASTM Infrared Data File; ASTM: Philadelphia, Pa., 1964.
- (11) CNMR Data Base, NIC, Delft, P.O.B. 36, 2600AA Delft, The Netherlands.

## Parent Compound Handbook—Successor to The Ring Index<sup>†</sup>

J. E. BLAKE,\* S. M. BROWN, T. EBE, A. L. GOODSON, J. H. SKEVINGTON, and C. E. WATSON

Chemical Abstracts Service, P.O. Box 3012, Columbus, Ohio 43210

Received March 10, 1980

This paper outlines the development of the PCH, reviews its growth in the three years since its introduction, briefly discusses its production, and illustrates some of the ways it can be useful to the chemical community.

#### INTRODUCTION

Chemical Abstracts Service (CAS) first published the "Parent Compound Handbook" (PCH) in March 1977 as the successor to "The Ring Index".<sup>1</sup> Since that time it has been updated regularly. The PCH is both a current-awareness service and a major reference work on ring systems and natural products for those who use *Chemical Abstracts* (CA) and its associated indexes. At present it contains more than 50 000 entries. These include chemical structure diagrams, CA index names, Wiswesser Line Notations (WLN),<sup>2,3</sup> CAS Registry Numbers, molecular formulas, systematic names for natural products, and ring data for cyclic parent compounds. The more recent entries also include CA references indicating the first occurrence of the ring systems in CAS processing. Access to the information that the PCH contains is provided by means

of a set of six indexes, which allow searching by ring analysis, ring substructure, name, WLN, molecular formula, and CAS Registry Number.

#### BACKGROUND AND HISTORY

From the initial publication of *Chemical Abstracts* in 1907, chemical structure diagrams have been used in CA issues to illustrate specific chemical substances. However, they were not used for the first ten years in the annual indexes because the alphabetical index entries for related substances (such as bromophenanthrene, chlorophenanthrene, and nitrophenanthrene) were widely scattered, depending on the names of the substituents. Only with the adoption of inverted index nomenclature in Volume 10 and in the First Decennial Index in 1916 did the inclusion of chemical structure diagrams in the indexes become feasible. Index entries for closely related substances were no longer scattered throughout the index at several headings depending on the substituent name, but were

<sup>†</sup> Presented in part to the Division of Chemical Information, 173rd National Meeting of the American Chemical Society, New Orleans, La., March 22, 1977.