

# Expert system for the interpretation of infrared spectra

G.N. Andreev, O.K. Argirov and P.N. Penchev

*Department of Chemistry, University of Plovdiv, 4000-Plovdiv (Bulgaria)*

(Received 7th December 1992; revised manuscript received 30th March 1993)

## Abstract

An expert system for the interpretation of infrared spectra EXPIRS was created. The main features of EXPIRS are: hierarchical organization of the characteristic groups, realized by frames; registration of the multiple use of spectral bands; taking into account the solvent absorption and the chemical inconsistencies; documenting the interpretation course and providing explanations on request. The ten most important heuristics used by an expert for interpretation of infrared spectra were formulated and some of them were tested with EXPIRS.

**Keywords:** Infrared spectrometry; Expert systems; Frames; Heuristics; Organic compounds

Computer-assisted interpretation of infrared (IR) spectra has drawn the attention of scientists for more than a decade. Several different approaches have been applied, including the utilization of correlation tables [1–5], symbolic [6–8] and fuzzy [9] logic, expert systems based on rules [10–14 and refs. cited therein] and table-driven procedure [15], frames [16] and, recently, neural networks [17–19]. Most of the authors have formulated their results in the terms of probability predictions for the characteristic groups in the studied compound. These systems produce results in the form of tables: functional groups vs. probability. The final decision for the presence of a given substructure was left to the user. In the general case, however, the user is not a specialist in IR spectroscopy.

An alternative approach is to base the interpretation rules on classical logic, leading to decisions clearly formulated by an expert. This can be

achieved using the heuristic knowledge of a human expert in conjunction with the positive characteristics of computers.

The present work deals with the formulation of the principle heuristics used by an expert to interpret IR spectra and the implementation of some of them in an expert system.

## BASIC CONCEPTS OF THE EXPERT

Until now, we have not found in the literature clear formulated heuristics used by an expert for the interpretation of vibrational spectra. Our experience in structural elucidation of organic compounds by IR spectroscopy and the practice with EXPert in InfraRed Spectroscopy (EXPIRS) have led to the following ten most important heuristics:

- (1) Taking into consideration the preliminary information about the studied sample.
- (2) Correction of the spectral band intensities, when some of the bands are “obviously too intensive”.

*Correspondence to:* G.N. Andreev, Department of Chemistry, University of Plovdiv, 4000-Plovdiv (Bulgaria).

(3) Comparison between the parameters of the spectrum bands (position, intensity, width) and characteristic group data.

(4) Discussion of the alternatives for the functional group combinations, possibly existing in the compound at hand, satisfying the spectral data.

(5) Discussion of the alternatives for the explanation of spectrum bands' origin: normal vibrations, overtones, combinations, Fermi resonances.

(6) Excluding from consideration the bands used during the interpretation of every alternative, i.e. single use of each band by discussion of every alternative.

(7) Taking into consideration the band shapes as well as the whole spectrum or any of its sectors.

(8) Taking into consideration the absorption of moisture and carbon dioxide.

(9) Taking into consideration spectrum registration conditions: physical condition (influence upon characteristic group intervals and band intensities), solute–solvent interactions, blocking of spectrum intervals caused by solvent absorption, sample concentration (hydrogen bonds), sample thickness, pressure, etc.

(10) Taking into consideration non-spectral reasons: chemical inconsistencies of the functional groups simultaneously predicted in a given alternative; chemical interaction between the solvent used and the predicted functional groups.

It should be noted that this enumeration is not in order of importance because the neglect of any heuristic can lead to wrong conclusions. Furthermore, such an order should not be accepted as an algorithm, because some of the heuristics must be repeatedly applied in the course of interpretation, depending on the nature of the studied molecule.

The third heuristic is the only heuristic used in all expert systems developed for the interpretation of IR spectra. Most of these also use some of the other heuristics described above. However, we have not found any utilization of heuristics 5, 7, 8 and 10 whereas 2 and 6 have been used but not in the same manner as an expert would use them.

The band of each characteristic group has a

different relative intensity depending on the other groups included in the same molecule. For example,  $\gamma(\text{CH})$  modes are very strong in the IR spectra of aromatic hydrocarbons but they often appear medium to strong in the spectra of aromatic carbonyl compounds. The common solution for this problem of the system for automated interpretation is enlargement of the intensity interval in the knowledge base. However, this automatically brings one to the so-called "hyperprediction". On the other hand, the expert detecting the presence of the "very strong"  $\nu(\text{CO})$  modes "expands" the intensity of the other spectrum bands instead of extending the intensity intervals in his mind. The application of this principle in the expert system will avoid the hyperprediction. Heuristic 2 takes into account the latter approach.

The results obtained from EXPIRS testing pointed out that the most significant reduction of the hyperprediction can be achieved by the application of heuristic 6.

## EQUIPMENT AND MATERIALS

EXPIRS was developed on an IBM-PC compatible computer with 640 kbyte RAM and the program was written in PASCAL (the system is available on request). 200 spectra from the Sadtler Fourier transform (FT)-IR library were used to test the system as well as spectra registered in our laboratory on a Perkin Elmer 1750 FT-IR spectrometer with a resolution of  $2\text{ cm}^{-1}$ .

More than 70 characteristic groups containing the elements C, H, O and N were incorporated in the system and the appropriate rules for their interpretation were programmed.

### *Program description*

The expert system is based on the concept of "characteristic group intervals". [20]. There are three levels in EXPIRS (Fig. 1). The first two levels involve the process of interpretation; the third one (documentation) is designed to register the interpretation course and the number of spectrum peaks used during the work of the program.

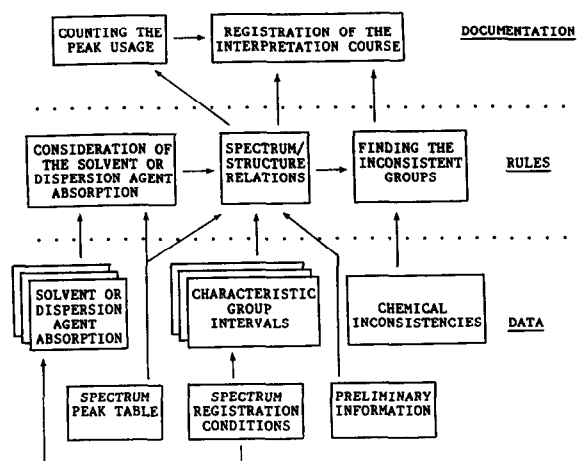


Fig. 1. Flow chart of EXPIRS.

The data for the characteristic intervals, solvent absorption and chemical inconsistency were realized as arrays of the appropriate variables.

Analysis of the literature, combined with our experience, reveals that the groups are best organized hierarchically. Such an approach avoids repeating the interpretation of the spectral intervals. For example, the primary  $-\text{CH}_2-\text{OH}$ , secondary  $>\text{CH}-\text{OH}$  and tertiary  $>\text{C}-\text{OH}$  alcohols have a common interval at  $3600-3200\text{ cm}^{-1}$  due to the stretching vibration of the hydroxyl group  $\nu(\text{OH})$ . The triple verification of this interval can be avoided if the characteristic OH group, which determines this interval, is used as a "parent group". In this case, the rules for the different alcohols (primary, secondary and tertiary) must take into account not only the entire spectrum, but also the status of the parent group OH. Such an approach was utilized in the description of other characteristic groups, including alkanes, alkenes, amines, aldehydes, etc. This type of group organization corresponds to the chemist's knowledge that the primary, secondary and tertiary alcohols are special cases of the concept "alcohol". In other words such an organization reflects the different levels of abstraction of the chemical structure in the chemist's mind.

We have used frames to realize the hierarchical organization of the characteristic groups in our expert system (Fig. 2). The principal frame in

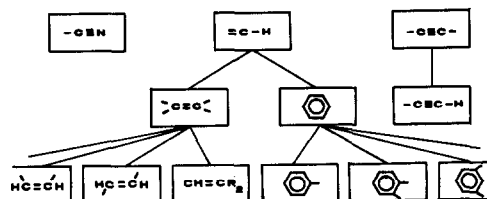


Fig. 2. Hierarchical organization of the frames in EXPIRS.

the program is the frame "Characteristic group"; two specifications of this frame are given in Fig. 3 for the groups  $\text{sp}^3\text{C}-\text{H}$  and  $\text{CH}_2$ .

When the system checks for the presence of the methylene group  $\text{CH}_2$  it needs data for its parent group  $\text{sp}^3\text{C}-\text{H}$ . If the status of the parent group has yet to be discovered, the procedure starts with its frame, etc. There are no limitations to the depth of such parent group checks. This makes the system independent on the group order in the data base and on the user's range of demand, which differentiate EXPIRS from the approaches utilized elsewhere [10,16,21].

One can readily see (Fig. 3) that the parent group and its "generics" belong to the same frame and the hierarchy mentioned above is de-

<u>Characteristic group:</u>	
Name:	$\text{sp}^3\text{CH}$
String for the screen:	$\text{sp}^3\text{C}-\text{H}$
Conclusion for presence:	Received by a procedure based on data for the characteristic intervals, spectrum, solvent absorption.
Parent group:	
Name:	Absent
Conclusion for presence:	Absent
Characteristic intervals:	
Interval No 1:	$3000-2800\text{ cm}^{-1}$ weak to strong
<u>Characteristic group:</u>	
Name:	$\text{CH}_2$
String for the screen:	$\text{CH}_2$
Conclusion for presence:	Received by a procedure based on data for the parent group, characteristic intervals, spectrum, solvent absorption.
Parent group:	
Name:	$\text{sp}^3\text{CH}$
Conclusion for presence:	Received
Characteristic intervals:	
Interval No 1:	$1480-1430\text{ cm}^{-1}$ weak to strong

Fig. 3. Two entities from the frame "characteristic group":  $\text{sp}^3\text{C}-\text{H}$  and  $\text{CH}_2$ .

terminated not as frame hierarchy, but by means of the connection "parent group" between the different specifications of the same frame "characteristic group".

#### Operation of the system

The interaction between the user and the system is determined by the appropriate menus. There are three main options: INTERPRETATION, CONCLUSIONS, SAVING THE RESULTS.

In the option INTERPRETATION the user:

- puts in the name of the spectrum file; interaction mode can be selected;
- defines the intensity margins for strong, medium and weak peaks;
- puts in the solvent or the disturbing media for the measured sample;
- gives preliminary information (based on analysis or the origin of the sample) for the absence of some groups as well as the kind of the groups of interest.

The option CONCLUSIONS provides information on the groups whose presence in the studied molecule is positive, negative or uncertain. The user can request explanation about the arguments on which the conclusion is based. Additional information on positively predicted groups which are not chemically coincident is available.

In the option SAVING THE RESULTS the user can save on disk or print the results of the interpretation.

#### RESULTS AND DISCUSSION

We have obtained excellent to satisfactory agreement between the predicted and existing characteristic groups by testing the developed expert system with the IR spectra, measured in our laboratory or included in the Sadtler FT-IR library.

The following examples of EXPIRS' work illustrate the importance of some heuristics described above.

The interpretation of the spectrum of DL-2-methylbutanoic acid [ $\text{CH}_3\text{CH}_2\text{CH}(\text{CH}_3)\text{COOH}$ ] gives a report for the presence of the following characteristic groups:  $sp^3\text{C-H}$ ,  $\text{CH}_2$ ,  $\text{CH}_3$ ,  $\text{COOH}$ , alkyl-COOH. As seen EXPIRS identifies all three functional groups of the studied substance. It further specifies that the carboxylic group is aliphatic. The presence of C-H with  $sp^3$ -hybridized carbon atom is also mentioned.

Along with the correct predictions we also received some incorrect answers by the spectral interpretation. The latter can be divided into two different types: (1) negative statement for a characteristic group existing in the studied molecule; (2) positive statement for a characteristic group that is absent in the studied compound.

We found that one could eliminate the first kind of incorrect answers by improving the data base. The errors of the second type, which appear more frequently (hyperpredictions), could not be avoided by the same manner. The latter are connected with the multiple use of the same bands in the course of interpretation. Our program was supplied with a counter in order to register the multiplicity of band usage.

The hyperprediction will be illustrated by the interpretation of the spectrum of isopropylbenzene. EXPIRS reveals the existence of the following characteristic groups:  $sp^3\text{C-H}$ ,  $\text{CH}_2$ ,  $\text{CH}_3$ , *i*-Pr, =C-H, C=C, *cis*-CH=CH, aryl, Ph-, *o*-aryl.

This interpretation is based on the spectrum bands given in the following format: wavenumber(bandwidth) - multiplicity of band's usage: 3083(m) - 1, 3064(m) - 1, 3029(s) - 1, 2963(s) - 1, 2929(s) - 1, 2889(s) - 1, 2873(s) - 1, 2802(m) - 1, 1949(m) - 0, 1872(w) - 0, 1804(m) - 1, 1744(w) - 0, 1665(w) - 1, 1605(m) - 2, 1533(w) - 0, 1496(m) - 0, 1465(s) - 1, 1451(s) - 1, 1386(m) - 3, 1364(m) - 3, 1320(m) - 1, 1300(m) - 1, 1279(m) - 1, 1208(w) - 0, 1102(m) - 0, 1079(m) - 1, 1048(m) - 2, 1027(m) - 2, 921(m) - 0, 905(m) - 0, 777(w) - 0, 761(s) - 4, 697(s) - 4, 531(m) - 1, where s = strong, m = medium and w = weak.

The multiple use of the 1605, 761 and 697  $\text{cm}^{-1}$  bands is the reason for the C=C, *cis*-CH=CH and *o*-aryl groups hyperprediction. Obviously, one could eliminate such incorrect predictions by tak-

ing into account the multiple use of each spectral band.

The ability of EXPIRS to take into consideration the influence of the solvent or dispersion agent used will be illustrated by the interpretation of the spectrum of D-glucose measured in nujol mull (wavenumber/relative intensity bandwidth): 3408/0.810, 3306/0.833br, 2925/0.905, 2855/0.828, 1459/0.720, 1376/0.662, 1340/0.630, 1296/0.568, 1224/0.568, 1203/0.570, 1149/0.695, 1111/0.751, 1078/0.611, 1050/0.753, 1024/0.824, 996/0.798, 916/0.597, 838/0.592, 775/0.559, 723/0.531, 614/0.709.

The following message appears after the interpretation:

(1) There are spectral data for PRESENCE of the following groups: OH, RCH<sub>2</sub>-OH, R<sub>2</sub>CH-OH, R<sub>3</sub>C-OH, Ar-OH, R-O-R, R-O-Ar, >N-H;

(2) The presence of the following groups is UNCERTAIN: sp<sup>3</sup>C-H, CH<sub>2</sub>, (CH<sub>2</sub>)<sub>4</sub>, CH<sub>3</sub>, i-Pr, t-Bu, Ar-NH.

The second report deals with the groups whose characteristic intervals overlap with those of the dispersion media (or solvent used).

An important feature of EXPIRS is that the user can follow the logic of the interpretation upon request. If the user wants to provide an explanation for any of the above results, he can use the option EXPLANATION. He goes to the option CONCLUSIONS and designate the corresponding group of interest. For example, the following messages will appear for the groups R<sub>2</sub>CH-OH and CH<sub>2</sub>:

R<sub>2</sub>CH-OH . . . . .

REPORT: The conclusion for presence of OH is POSITIVE

Is there a strong band between 1120 and 1080?

YES: 1111 (s)

There are spectral evidences for presence of R<sub>2</sub>CH-OH

Press any key to continue . . .

CH<sub>2</sub> . . . . .

REPORT: The conclusion for presence of sp<sup>3</sup>C-H is UNCERTAIN

## TOTAL OVERLAPPING WITH THE SOLVENT BANDS

The presence of the CH<sub>2</sub> is UNCERTAIN

Press any key to continue . . .

This possibility makes the expert system appropriate for educational purposes as well.

The authors wish to thank the Bulgarian National Fund of Research at the Ministry of Education and Science for partial financial support through Grant No. X-124/91.

## REFERENCES

- 1 T. Visser and J.H. van der Maas, Anal. Chim. Acta, 122 (1980) 363; Anal. Chim. Acta, 133 (1981) 451.
- 2 M. Farkas, J. Markos, P. Szepesvary, I. Bartha, G. Szalontai and Z. Simon, Anal. Chim. Acta, 133 (1981) 19.
- 3 G. Szalontai, Z. Simon, Z. Csapo, M. Farkas and Gy. Pfeifer, Anal. Chim. Acta, 133 (1981) 31.
- 4 B. Debska, J. Duliban, B. Guzowska-Swider and Z. Hippe, Anal. Chim. Acta, 133 (1981) 303.
- 5 W.-R. Leupold, C. Domingo, W. Niggemann and B. Schrader, Fresenius' Z. Anal. Chem. 303 (1980) 337.
- 6 L.A. Gribov and M.E. Elyashberg, J. Mol. Struct., 5 (1970) 179; J. Mol. Struct., 50 (1978) 351.
- 7 L.A. Gribov, M.E. Elyashberg and L.A. Moscovkina, J. Mol. Struct., 9 (1971) 357.
- 8 K. Funatsu, Y. Susuta and S. Sasaki, Anal. Chim. Acta, 220 (1989) 155.
- 9 T. Blaffert, Anal. Chim. Acta, 161 (1984) 135.
- 10 H.B. Woodruff and G.M. Smith, Anal. Chem., 52 (1980) 2321.
- 11 S.A. Tomellini, D.D. Saperstein, J.M. Stevenson, G.M. Smith, H.B. Woodruff and P.F. Seelig, Anal. Chem., 53 (1981) 2367.
- 12 S.A. Tomellini, R.A. Hartwick and H.B. Woodruff, Appl. Spectrosc., 39 (1985) 330.
- 13 B.J. Wythoff, C.F. Buck and S.A. Tomellini, Anal. Chim. Acta, 217 (1989) 203.
- 14 H.J. Luinge, Vib. Spectrosc., 1 (1990) 3.
- 15 M.O. Trulson and M.E. Munk, Anal. Chem., 55 (1983) 2137.
- 16 P. Edwards and P.B. Ayscough, Chemom. Intell. Lab. Syst., 5 (1988) 81.
- 17 E.W. Robb and M.E. Munk, Mikrochim. Acta (Wien) I, (1990) 131.
- 18 M.E. Munk, M.S. Madison and E.W. Robb, Mikrochim. Acta (Wien) II, (1991) 505.
- 19 K.J. Fessenden and L. Gyorgyi, J. Chem. Soc. Perkin Trans 2, (1991) 1755.

20 N.B. Colthup, L.H. Daly and S.E. Wiberlay, *Introduction to Infrared and Raman Spectroscopy*, Academic Press, New York, 1975; L.J. Bellamy, *The Infrared Spectra of Complex Molecules*, Methuen, London, 1964; K. Nakan-

ishi, *Infrared Absorption Spectroscopy*, Holden-Day, San Francisco, CA, and Nankodo, Tokyo, 1962.

21 H.J. Luinge and J.H. van der Maas, *Anal. Chim. Acta*, 223 (1989) 135.