P.N. Penchev, N.T. Kochev and G.N. Andreev; IRSS: A Programme System for Infrared Library Search. Comptes Rendus de l'Academie Bulgare des Sciences, 51(1-2), 67-70 (1998).

IRSS: A PROGRAM SYSTEM FOR INFRARED LIBRARY SEARCH[*]

P. N. Penchev, N. T. Kochev, G. N. Andreev

Introduction. The development of automated systems for structure elucidation and identification of organic compounds from their infrared (IR) spectra continues to attract the attention of the spectroscopists. One of the most widely used techniques is searching in spectral libraries [1,2].

In this article we describe the program system *IRSS* for searching in IR spectral databases. It is a user friendly menu driven program working in Microsoft Windows environment. The program can perform the following operations:

- loading an unknown or a library spectrum in one of the three available buffers. The JCAMP-DX file format can also be used for import of spectra.
- viewing of full or zoomed spectra in absorbance or transmittance units, peak tables, and the corresponding compounds' structures.
- editing header information of an unknown or a library spectrum, as well as the structure of a library's compound.
- peak search using three algorithms: forward, reverse, and scalar product.
- full spectral search applying four algorithms: sum of least squares, sum of absolute value differences, scalar product, and correlation coefficient.
- search for a given compound's chemical name.

[*] This study was supported by the National Science Fund of the Bulgarian Ministry of Education, Science and Technology under grant X 447/1995.

- interactive analysis of spectra of mixtures with the aid of multilinear regression, and subsequent graphic representation of the results.
- creating user-generated libraries, as well as deleting and merging of libraries.
- adding or removing of spectra to/from a library.
- peak-picking from a spectral curve of an unknown spectrum, as well as creating of peak table file from a spectral library one.
- printing of an active spectrum or search results.

System description. Measurements and processing of spectra. The program concept was tested with three different spectral libraries: Plovdiv Uni Library (611 spectra recorded in our laboratory), Sadtler Demo Library [3] (200 spectra), and Work Library (1000 spectra purchased from Chemical Concepts [4]).

In order to build up the Plovdiv Uni library, 611 spectra were registered on a Perkin-Elmer 1750 FT-IR Spectrometer from 4000 cm^{-1} to 450 cm^{-1} at resolution 4 cm^{-1} with 25 scans [5]. The Sadtler library spectra are taken as they are. The Work library was created from spectra delivered as JCAMP-DX files. These spectra were subjected only to smoothing [6].

Programming, hardware requirements, and availability. The program code (about 15000 lines) was written and compiled in Borland Pascal. The program was fully tested on 100% compatible IBM 286, 386, 486, and Pentium PC's. The last two platforms are strongly recommended in order to speed up search sessions and regression calculations. The operating environments Windows 3.1x and Windows 95 are fully tested.

University scientists and users of non-profit organizations can get a free copy of the executive files and a demonstrative library containing 100 spectra randomly taken from Plovdiv Uni Library on request: email to plamen@ulcc.uni-plovdiv.bg.

Data representation. The program can maintain unlimited number of spectral libraries. Each entry in a library corresponds to a given compound and is represented with the following data: chemical name, Wiswesser line-formula chemical notation, molecular formula, molecular mass, technique, boiling and melting points, field for comments, peak table, structure, and spectral data. The latter are full spectral curve in absorption values at 4 cm^{-1} data intervals in the 3700 - 500 cm^{-1} range. They are normalized in the range of 0-1 absorbance units (a.u.). To speed up the scalar product and the correlation coefficient search algorithms the mean, dispersion, and norm of each spectrum are kept in the spectral files. The corresponding peak tables' file can be derived from these spectral data.

Spectral Similarity Measures. Three types of hit quality indices (HQIs) are used in peak search. The forward and reverse HQIs are based on counting peaks' matches [3]. We proposed another one, scalar product HQI which is calculated according to the formula:

$$HQI = \frac{\sum_k A_k^{Unk} A_k^{Lib}}{\left\| A^{Unk} \right\| \cdot \left\| A^{Lib} \right\|}; \qquad (1)$$

the sum in the nominator being taken only for matched peaks.

This HQI turned out to be very useful in searching a mixture spectrum against the library's ones.

Four different HQIs were used for full spectral search; they compare the entire spectral curves. These measures are based on the following relations [1,5]: sum of least squares, sum of absolute value differences, scalar product, and correlation coefficient.

Interactive analysis of spectra of mixtures. In principle, the composition of a mixture can be determined directly from its spectrum. Assuming that the intermolecular interactions are negligible, the mixture spectrum represents a linear combination of the components' spectra. In this case the pseudo-concentrations of the components in a mixture (C) can be calculated from the spectra in the hit list (matrix S_{hl} obtained by the search of the mixture spectrum S_{mix}) according to the equation:

$$C = S_{mix} \cdot S_{hl}^{T} \cdot (S_{hl \cdot} S_{hl}^{T})^{-1} \;;$$

"T" and "-1" in the superscript denote a transposed and an inverse matrix, respectively. Matrix C does not represent the exact concentrations because of the normalization of the library spectra in the range 0-1 a.u. and the differences in samples preparation: all library spectra were registered with arbitrary path length (for liquids) or arbitrary amount of compound in the KBr pellet (for solids).

The calculations are performed with increasing size of matrix S_{hl}, and the program presents graphs *C = f(number of the hit list's compounds)*. The user can decide which compounds are in the mixture by comparison the relative stability of the corresponding curves.

Example of application. As an example we will consider the analysis of a mixture of 3-methyl-1-butanol and 1-pentanol taken in 9:1 volume proportion. Our previous tests showed that the best parameters for peak search of mixture spectra are: the library's peak-table file prepared by peak-picking using 0.03 a.u. threshold, peak table of mixture spectrum created with 0.01 a.u. threshold, and the tolerances for peaks' matches - ±7 cm^{-1} and ±1.0 a.u. The recorded mixture spectrum was searched against Plovdiv Uni library using all seven implemented search algorithms.

As could be seen from the search results (Table 1), all methods gave the exact components' concentrations. This coincidence is accidental because of the reasons stated above. More important characteristic for a qualitative analysis of mixtures is the relative stability of the results in respect of the number of hit list's spectra. An estimation of this stability is a relative standard deviation (RSD) of the components' concentrations. It is calculated only for the points on the curve whose number is greater than or equal to the hit list position of the compound. The user can obtain the RSD value just clicking with a mouse on the corresponding curve.

Table 1. Search results for the mixture of 3-methyl-1-butanol and 1-pentanol: positions in the hit lists, concentrations, and RSD of concentration for the mixture components.

		peak search methods			full spectral search methods			
compound		forward	reverse	scalar prod.	least sqs.	abs. value	correl. coeff.	scalar prod.
3-Methyl-1-butanol	position	1	2	2	1	1	1	1
	C	0.92	0.96	0.94	0.91	0.92	0.91	0.90
	RSD %	4.5	4.0	1.8	3.6	3.6	3.6	3.7
1-Pentanol	position	8	9	3	2	2	2	2
	C	0.10	0.11	0.10	0.11	0.11	0.10	0.10
	RSD %	4.1	6.7	5.5	4.7	5.0	5.1	5.1

The results in Table 1 show that the HQI calculated according to Eq. 1 gives the best hit list for multivariable regression calculations. Although the components are in the second and third positions in the hit list, the corresponding graphs point stable concentrations (Fig. 1) which do not depend on the number of hit list's spectra considered. The corresponding RSD for the other compounds in the hit list are greater than 40%.

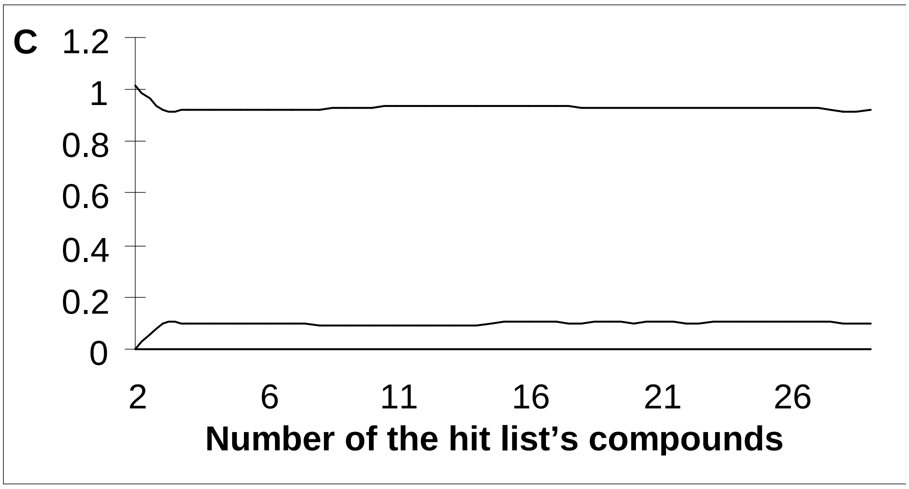

Fig. 1. Calculated concentrations (C) of 3-methyl-1-butanol (upper line) and 1-pentanol with the hit list obtained by scalar product peak search.

REFERENCES

[1] Luinge H.J. Vib. Spectrosc., **1**, 1990, 3-18. [2] Zupan J. In: Computer-supported Spectroscopic Data Bases, Chichester, Ellis Horwood, 1986. [3] Sadtler IR Search 2.25 University, Sadtler Research Laboratories, Division of Bio-Rad Laboratories, Inc. [4] Chemical Concepts, P.O. Box 10 02 02, D-69442 Weinheim, Fed. Rep. of Germany. [5] Penchev P. N., A. N. Sohou, G. N. Andreev. Spectrosc. Lett. **29**, 1996, No 8, 1513-1522. [6] Savitzky A., M.J.E. Golay. Anal. Chem., **36,** 1964, 1627-1639.

Department of Chemistry, University of Plovdiv,
24 Tsar Assen Str., 4000 Plovdiv, Bulgaria