

## ATR spectra database of organic compounds

S. Tsoneva, S. Nachkova and P. Penchev

**Abstract;** A spectral database of 100 attenuated total reflectance Fourier transform infrared spectra is created. The software IRSS previously developed in our lab is used to perform spectral search in the database. Despite the slight difference between the absorption and ATR FT-IR spectra, the comprehensive test search sessions of ATR spectra in IR libraries and vice versa (absorption spectra searched in the ATR library) show that an identification of the compound is a feasible task.

**Key words:** Library search; ATR spectral database.

### INTRODUCTION

The use of computer or printed spectral databases in NMR, MS and IR is now a routine practice in scientific and industrial labs [1]. The most applied is a search in spectral library of low resolution electron impact mass spectra that follows a chromatographic separation. Searching in IR spectral databases is also very popular because the IR spectra are easy, fast and cheap to produce, as well as, they reflect in great extent the compound's structure thus making them suitable for identification purpose [2].

The library search procedure consists of comparison of an IR spectrum of the unknown with a collection of reference IR spectra of compounds with known structure. The collection of spectra together with a chemical information and structure of the compounds is called a *spectral library*. The result obtained by a search, called a *hitlist*, is a list of spectra that are most similar to the query spectrum. The hits are sorted in the hitlist according to a real number called *hit quality index* (HQI) which reflects the spectral similarity between unknown and reference spectrum. If the unknown is among the library entries, then the correct answer often appears among the first several hits and a visual inspection of them makes it possible to identify the unknown: this is called *identity search* [3]. However, if the unknown compound is not contained in the spectral library, a more sophisticated interpretation of the hitlist is necessary. Assuming that similar spectra indicate similar structures, the hitlist structures characterize the unknown structure: this is the quintessence of the *similarity search*.

For both types of searches important aspects are the representation of the spectra, the spectral similarity measures used and the search algorithm applied [2]; for the similarity search another key characteristic is the method of analysis of the hitlist structures. Other characteristics of a library search system include the speed and versatility of the implemented search algorithms, the size, contents, and reliability of the data base, the options for an update of spectral library, the availability of modules for analysis of the hitlist entries and the possibility to derive spectrum-structure correlations [2].

This paper reports on a preparation of database of attenuated total reflectance (ATR) Fourier transform infrared (FT-IR) spectra. ATR is one of the more popular sampling techniques used by FT-IR spectroscopists because it is quick, non-destructive and requires virtually no sample preparation [4]. An ATR spectrum is slightly different from an IR one of the same compound measured in KBr pellet or as a thin layer: ATR spectral bands are narrower and band's relative intensities are also different from those of the absorption IR spectrum. That poses a challenge when IR spectra are searched in ATR library and vice-versa. That is why, the search capabilities of previously created software are also explored and reported here.

### EXPERIMENTAL

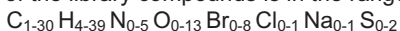
The method for library search is implemented into a Windows-based user-friendly program, called IRSS [5-6]. Seven different algorithms for the comparison of IR spectra are implemented: three methods for matching peaks [7] and four methods for comparing

full spectral curves [6]. Furthermore, IRSS contains software tools for an import of IR spectra in JCAMP-DX format, peak picking, and an interactive analysis of IR spectra of mixtures based on multiple linear regression techniques.

102 ATR FT-IR spectra are recorded with a VERTEX 70 FT-IR spectrometer (Bruker Optics); the spectrum is measured from 4500  $\text{cm}^{-1}$  to 600  $\text{cm}^{-1}$  at resolution 2  $\text{cm}^{-1}$  with 16 scans. The used ATR accessory is MIRacle™ with a one-reflection ZnSe element (Pike). For solid samples the stirred crystals of compound are pressed by an anvil to the reflection element; for liquid samples a drop is placed directly upon the crystal plate of the accessory. The spectra are saved as ASCII JCAMP-DX files (v. 4.24) and imported to compose the spectral database by use of the software IRSS [3-4]. The structures are prepared with ISISDraw, v. 2.4 (MDL, Inc.): MOL files are exported by this software and read by IRSS to create the chemical structure files. The software IRSS and the ATR spectral library is available from one of the authors (P.P.).

## RESULTS AND DISCUSSION

The molecular formula of the library compounds is in the range



and the mean molecular formula is



Forty out of 102 library compounds are present in the six FT-IR spectral libraries that have been generated earlier in our lab. The spectra (911 altogether) in these libraries are absorption IR spectra measured on two different FT-IR instruments. This duplication of spectra of the same compounds is not a flaw because "spectral search is greatly improved when using databases collected using the same sampling mode – especially when the sampling mode is ATR" [8]. These 40 ATR spectra are searched for identification in the six FT-IR spectral libraries: the used spectral similarity measure,  $HQI_4$ , is correlation coefficient of spectral curves, Eq. 1 [6].

$$HQI_4 = 999(S_4 + 1)/2; \quad S_4 = \frac{\sum_k (A_k^U - \overline{A^U})(A_k^R - \overline{A^R})}{\sqrt{\sum_k (A_k^U - \overline{A^U})^2 * \sum_k (A_k^R - \overline{A^R})^2}}, \quad (\text{Eq. 1})$$

where  $A_k$  are the absorbance values in the spectra (801 such values); U = unknown spectrum, R = reference one; the bar over a variable indicates its average value.

The search results are listed in Table 1. The detailed examination of the worse results (hit position of identical compound greater than 1) indicates that the absorption and the ATR FT-IR spectra of the same compound are substantially different. In some cases not only the relative intensity of the bands varies but the band width is different (5 cases). When spectra of compounds very structurally similar to the unknown are present in the library, the identification is hampered because the difference in the relative intensities puts other spectra at the top the hitlist (5 cases).

The intensity of the ATR bands increases with the increase of wave number, the fact explained with the larger depth of penetration at longer wave length. The same appears to be valid for the background (if present in the spectrum). There are 4 such cases when the background increases in ATR spectra with the decrease of wave number and this worsens the identification.

Hit #	Frequency
1	28
2	3
4	3
5	1

Table 1. Hit number of the correct identification and its frequency.

7	1
8	1
10	1
12	1
21	1
All	40

The ATR spectra are known to be with a very low background and it is the case for most of these 102 spectra. Of course, the background correction can solve this type of problem as it has been done with all absorption FT-IR spectra in the six libraries used in this study.

The search of the corresponding 40 absorption FT-IR in the ATR library gives very satisfactory results: two cases of second hit position, one case in third and one in ninth; the remaining 36 searches identify the unknown as a first hit in the hitlist. That can be explained with the smaller number of ATR spectra in the library.

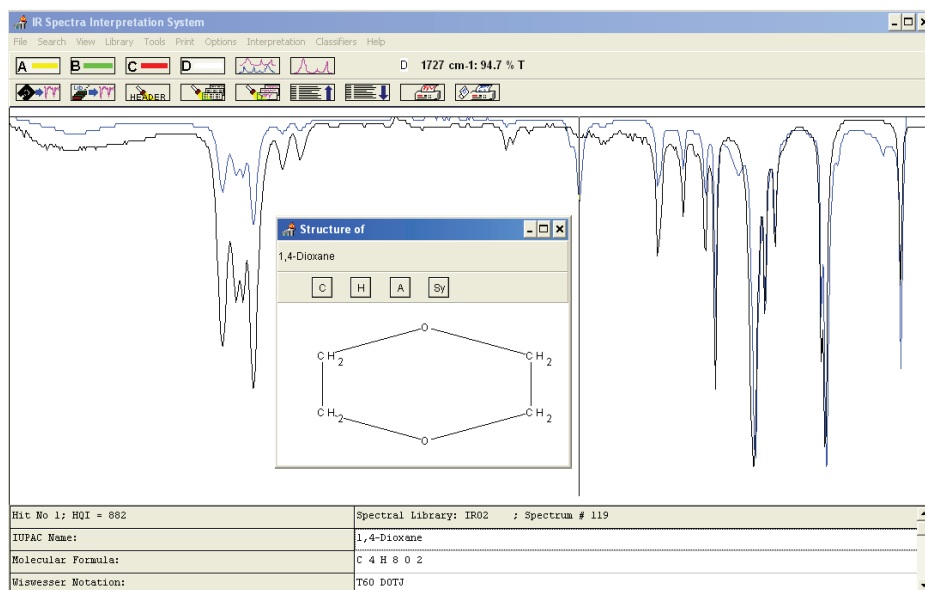


Figure 1. The main window of the program.

In order to illustrate the difference between both type of spectra, the ATR spectrum of 1,4-dioxane is searched in the six spectral libraries (911 spectra altogether) that have been created in our lab in the period 1994-2012. Figure 1 presents the main window of the program with the background in reverse color (to save printing ink). The query spectrum and the first hit (1,4-dioxane, too) are overlaid on the screen. The compound structure is given and the cursor (right to it) points to the ATR spectrum band at  $1724\text{ cm}^{-1}$  which shows a possible partial degradation of the sample. It can also be seen that the ATR spectrum is cut at  $600\text{ cm}^{-1}$  (i.e. zeroed between  $600$  and  $500\text{ cm}^{-1}$ ) because the ZnSe is absorbing IR radiation below  $600\text{ cm}^{-1}$ . That difference between both spectra plays an insignificant role by spectral comparison, giving HQI of 882 (999 is the highest HQI which is achieved between identical spectra).

## CONCLUSIONS

The created spectral database of 100 ATR FT-IR spectra is a useful and helpful addition to the already created databases of absorption IR spectra. The software IRSS previously developed in our lab can be used to perform search of ATR spectrum in the database of absorption IR spectra and vice versa an absorption IR one in ATR database. Despite the slight difference between the both types of spectra, an identification of the compound is a feasible task even when comparing different type of IR spectra.

## REFERENCES

- [1] Zupan J. ed. Computer-supported Spectroscopic Data Bases, Chichester: Ellis Horwood, Inc, 1986.
- [2] Luinge, H. Automated Interpretation of Vibrational Spectra. *Vib. Spectrosc.* 1990, **1**, 3-18.
- [3] Clerc, J.T.; 1987. Automated spectra interpretation and library search systems. In: Meuzelaar, H.L.C., Isenhour, T.L. (Eds.), *Computer-Enhanced Analytical Spectroscopy*. Plenum, New York, pp. 145–162.
- [4] Larkin P.; 2011. *Infrared and Raman Spectroscopy. Principles and Spectral Interpretation*, Elsevier.
- [5] P. Penchev, N. Kochev and G. Andreev; IRSS: A Programme System for Infrared Library Search. *Comptes Rendus de l'Academie Bulgare des Sciences*, 1998, **51**, 67-70.
- [6] Varmuza, K.; P. Penchev; H. Scsibrany. Maximum Common Substructures of Organic compounds Exhibiting Similar Infrared Spectra. 1998, *J. Chem. Inf. Comp. Sci.*, **38**, 420-427.
- [7] Penchev, P.; V. Miteva; A. Sohoy; N. Kochev; G. Andreev; Implementation and Testing of Routine Procedure for Mixture Analysis by Search in Infrared Spectral Library. 2008, *Bulg. Chem. Commun.*, **40**, 556-560.
- [8] ATR Spectral Databases – Optimize Search Results for ATR Spectral Data. Product Data Sheet, Pike Technologies.

## Acknowledgement

This work has been supported by the Bulgarian National Science Fund, Contract DDWU02/37 and Plovdiv University Project NI13ChF006.

## About the authors:

Correspondence author: Slava Tsoneva, Ph.D. student, Chemical Faculty, University of Plovdiv, E-mail: savava82@gmail.com.

Stefka Nachkova, Assist. Prof., Chemical Faculty, University of Plovdiv.

Plamen Penchev, Assoc. Prof., PhD, Faculty of Chemistry, University of Plovdiv.

**This paper has been reviewed**