Vol. 32 No. 14 KEXUE TONGBAO July 1987

AN AUTOMATIC INTERPRETATION SYSTEM OF INFRARED SPECTRA*

Song Guonan (宋果男), Che Xun (车 迅), Huang Weidong (黄威东),
Sun Faxiao (孙法校) and Wang Chuan (王 川)
(Dalian Institute of Chemical Physics, Academia Sinica)

Received January 11, 1984; revised April 21, 1986.

Infrared spectroscopy is an effective method for qualitative analysis and structure elucidation, and it can be widely applied to analysing gas, liquid and solid samples. However, the interpretation of infrared spectra is usually a difficult task for all but experienced spectroscopists. With the development and popularization of the computer, it is desirable to use the computer for interpreting the spectra. In recent years, some advances have been made in the studies of computer-aided interpretation of infrared spectra, such as the PAIRS program[1,2] which has been used on NICOLET FT-IR system and the SEARCH-2 program[3] of PERKIN-ELMER 3600 Data Station. But their interpreted results are still not very satisfactory. In this note, we will describe our automatic infrared interpretation system (IRIS).

I. IRIS Interpretation System

This system is performed on a NICOLET 7199B FT-IR system or a PERKIN-

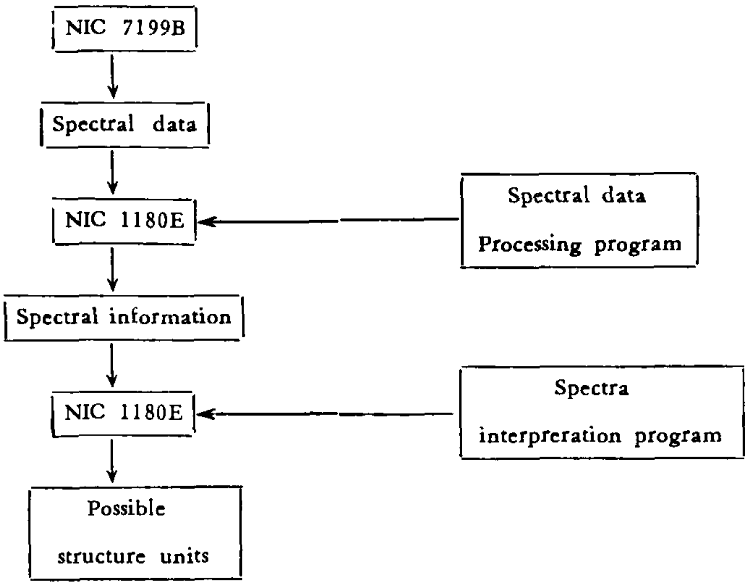

Fig. 1. The diagram of the interpretation system using a NIC 7199B FT-IR system.

---

* Project supported by the Science Fund of Academia Sinica.

ELMER 683 infrared spectrophotometer with a 3600 Data Station, and includes a data processing program and a spectra interpretation program in Basic or Fortran language. Fig. 1 is the diagram of the system. A digitized spectrum from infrared spectrophotometer is normalized by spectral data processing program, and then the useful information is extracted and transferred into the spectra interpretation program, from which the possible structure units are obtained.

II. THREE BASIC FUNCTIONS OF SPECTRAL DATA PROCESSING PROGRAM

(1) To correct the base-lines of the spectra in order to decrease the influences of the scattering, reflection and background absorption on spectral data.

(2) To normalize the spectral intensities. The normalization makes use of the most intensive peak located in the range of 600—2000 cm^{-1} and this peak is set to $T = 5\%$, and the others are adjusted respectively at the same absorbance proportion. In such a manner, most of the characteristic spectral peaks were of modest intensities, so that the interpretation results of the substances with long hydrocarbon chain can also be improved.

(3) To extract the useful information from the normalized spectrum, such as position, intensity and half-width of peak, and transfer them into the interpretation program.

III. SPECTRAL INTERPRETATION PROGRAM

The program is the key to the system. Its full capability depends on the compiler's ability to interpret spectra, that is, how many reliable rules of spectra he has mastered, and how much he can use ingeniously and correctly. Therefore, great attention should be paid to the following three aspects in compiling IRIS interpretation program.

(1) There are a large number of published rules of correlations between infrared spectra and molecular structures. However, some of them are not perfect because they were summarized from a few experimental data or the data observed on early low-performance instruments. In addition, some group frequencies are useless for the judgement of the existence of group and they can only be used to assign absorption bands of known compounds because they are in the regions of high interference. For those reasons, we have checked the published rules one by one, and selected the reliable and useful ones.

(2) To find new spectral rules based on the modern experimental data and theories. For example, we presented some new spectral rules after studying 386 spectra of amine[4], so that the unit of tertiary amine can be detected in IRIS.

(3) The overlapping of the regions of group frequency makes it difficult to distinguish the groups definitely using these frequencies. The strict criteria will increase the reliability of output, but will miss more detections. On the other hand, the over loose criteria can reduce the number of detections skipped, but they would bring about a lot of false detections. In compiling the program, we used reliable criteria as many as possible and defined the regions of the group frequency carefully in order to get the best interpretation results. The interpretation program of this system was compiled with more than 200 spectral rules, and can detect 110 common structure units.

IV. Examination of IRIS Interpretation System

The infrared spectra of more than 300 known compounds with various types of structural unit were interpreted automatically by IRIS system. Table 1 lists the interpretation results of the spectra of tert-butylcyclohexanol and mesitylene. All of these results are true, and they have given out the definite structural information. All of the interpretation results of the above compounds from PAIRS and SEARCH-2 are also listed in the same table. For tert-butylcyclohexanol, Nos. 1, 2, 4 and 6 in 14 possible structural units given by PAIRS are true, and for mesitylene, only Nos. 1, 2 and 6 in 13 possible structural units are true. The SEARCH-2 gave out more reliable results and only No. 7 for tert-butylcyclohexanol and No. 2 for mesitylene are false, but it cannot give out definite structural information, and is less useful in structure elucidation.

Table 1

Comparison of the Interpretation Results

	IRIS			PAIRS			SEARCH-2	
	No.	PSU	p	No.	PSU	p	No.	PSU
p-Tert-butylcyclohexanol	1	tert. butyl	0.9	1	methyl	0.70	1	alkyl
	2	sat. CH	0.9	2	alcohol	0.68	2	alkyl-methyl subst.
	3	CH$_3$	0.85	3	tert. alcohol	0.51		
	4	prim. OH/cyclosec.OH	0.8	4	sec. alcohol	0.51	3	alkyl-branched chain with multiple methyl subst.
				5	thiocarbonyl	0.50		
	5	CH$_2$	0.8	6	methylene	0.50		
	6	OH	0.7	7	amine	0.45		
				8	sec. amine	0.45	4	alkyl-hydroxy or amino subst.
				9	tert. amine	0.40		
				10	unsat. ether	0.36	5	aliphatic alcohol-prim. or sec.
				11	ether	0.36		
				12	gemdi methyl	0.35	6	aliphatic alcohol-prim. or sec. or cyclic
				13	sulfoxide	0.25		
				14	ether-epoxide	0.16		
							7	aliphatic cyclic ether
Mesitylene	1	sat. CH	0.9	1	aromatic	0.99	1	alkyl
	2	CH$_3$	0.85	2	benzene (1, 3, 5)	0.99	2	alkyl-methyl subst.
	3	unsat. CH	0.85	3	benzene (1, 4)	0.80		
	4	1,3,5-subst. benz.	0.85	4	heteroaromatic	0.50	3	aromatic hydrocarbon
				5	pyridine	0.50		
	5	conj. $C{=}C$	0.8	6	olefin	0.50	4	aromatic hydrocarbon-multiple subst.
	6	CH$_2$/Ar-CH$_3$	0.75	7	CHR=CHR, *cis*	0.50		
				8	sulfoxide	0.45		
				9	methyl	0.40		
				10	methylene	0.38		
				11	ketone	0.30		
				12	chelate	0.30		
				13	nitramine	0.25		

Note: PSU, possible structural unit; P, probability index.

In order to examine comprehensively the effectiveness of IRIS system, an additional program is set up to check each given result, to count the number of true detection (T), false detection (F) and missing detection (M) automatically and finally to present the ratio of successful detection in existing structural units, D_T, and the reliability of detection results, R_T. D_T and R_T are defined as

$$D_T = \sum T / (\sum T + \sum M),$$
$$R_T = \sum T / (\sum T + \sum F).$$

In this examination, D_T and R_T are 86% and 85% respectively. It means that 86% of all structural units of samples analyzed have been detected successfully by IRIS, and 85% of results given by IRIS are true. There are no comparable data published. However, it is estimated from Table 1 that R_T of PAIRS is much lower than that of IRIS.

REFERENCES

[1] Woodruff, H. B. et al., *Anal. Chem.*, **52**(1980), 2321.
[2] Tomellini, S. A. et al., *ibid.*, **53**(1981), 2367.
[3] Perkin-Elmer, "*Search-2 Applications Program*" *Rev.*, 1983.
[4] 宋果男，化学学报，**43**(1985), 184.