

SYSTEMATIC COMPUTER-AIDED INTERPRETATION OF INFRARED AND RAMAN VIBRATIONAL SPECTRA BASED ON THE CRISE PROGRAM

T. VISSER and J. H. van der MAAS*

Laboratory for Analytical Chemistry, University of Utrecht, Croesestraat 77a, 3522 AD Utrecht (The Netherlands)

(Received 7th November 1979)

SUMMARY

The CRISE computer program is used to correlate wavenumber regions and 6 structural elements containing carbon, hydrogen and oxygen on the basis of 2 standard files with 549 infrared and Raman spectra. The degree of correlation, including score percentages and interfering percentages, is established for different types of intervals in relation to various intensity thresholds. Specific regions (score 100%, interference 0%) proved to be rare, whereas pseudo-specific regions (score < 100%, interference 0%) are normally present. The usefulness of selective regions (score 100%, interference > 0%) is doubtful. The infrared and Raman results for a structural element can differ appreciably, yet neither technique is clearly superior for interpretative purposes.

Interpretation systems for some vibrational spectra can be compared objectively; the necessary parameters can be obtained from a file of coded spectra with the CRISE computer program in limited cases [1]. Files of both the infrared (i.r.) and Raman spectra of over 500 organic compounds are available here, and so it has been possible not only to study the interpretative value of each technique separately, but also to compare their particular utility. It should be noted that the comparison is not perfect as the scanning conditions for the i.r. and Raman spectra were slightly different.

EXPERIMENTAL

Two files, i.r. and Raman, were prepared; both contained the spectra of the same 549 liquid organic compounds. The samples contained the atoms CH (161), CHO (333), CHN (51) and CHNO (4). They were either commercial products or were obtained from the Laboratory of Organic Chemistry of this University. The purity of all compounds was $\geq 98\%$ (checked by g.c.).

The i.r. spectra were recorded on a Perkin-Elmer 180 spectrometer; the accuracy was 2 cm^{-1} in the region $4000\text{--}2000\text{ cm}^{-1}$ and 1 cm^{-1} in the region $2000\text{--}600\text{ cm}^{-1}$. The baseline was adjusted between 100 and 95% transmittance, the transmittance of the most intense band being 3–7% transmittance.

The Raman spectra were recorded on a Spectra-Physics 700 spectrometer with a Spectra-Physics 165 argon ion laser as light source. Spectra were recorded in the ranges $4000\text{--}2000\text{ cm}^{-1}$ and $2000\text{--}200\text{ cm}^{-1}$ with a precision of 2 cm^{-1} . The sensitivity was adjusted in such a way that the most intense band in the spectral region gave a reading between 85 and 95 scale divisions. For both techniques the peak height (intensity) is defined as the distance between the top of that peak and the baseline on both sides of the peak. A peak is considered to be present if its height is ≥ 3 scale divisions.

The program CRISE is written in FORTRAN EXTENDED IV. Data calculation performed on a CYBER 73-28 computer required a memory of 52K words and, depending on the structural element investigated, 15–30 CPU seconds per structural element. The different wavenumber precisions have been incorporated in the results of CRISE.

RESULTS

The basic files contain the atoms C, H, O and N. Only structural elements of C, H and O have been studied, as the number of compounds with nitrogen was considered to be too small for reliable results. The number of main elements is therefore eight, viz. C–H, C–C, C=C, C≡C, O–H, C–O, C=O and O–O. Of these, O–O, C–H and C–C have not been investigated: for obvious reasons, O–O is not present in the basic files; C–H and C–C would not produce useful intervals in i.r. or in Raman spectroscopy, because C–H is present in every compound and the C–C vibration is usually strongly coupled. The element C–H can be split into three semi-elements: –C–H, =C–H and ≡C–H, which form part of nine sub-elements (Table 1). Four of these are already included in the corresponding main elements, and as neither C–C nor the nitrogen-containing elements are included (see above), this leaves only one extra element, viz. –C–H (alkyl).

As reported previously [1], the CRISE program produces pseudo-specific and selective intervals in addition to specific intervals. Here, the number of non-specific intervals is kept within reasonable limits by setting a score percentage (*SP*) of at least 4% and a maximum interval width of 300 cm^{-1} . The number of combined regions or combined questions (*mQ*'s) has been restricted to five.

TABLE 1

The different types of structural elements

Semi-element	Sub-element	Main element	Semi-element	Sub-element	Main element
H–C–	H–C–C	C–C	H–C=	H–C=C	C=C
	H–C–H			H–C=O	C=O
	H–C–O	C–O		H–C=N	C=N
	H–C–N	C–N		H–C≡C	C≡C
				H–C≡N	C≡N

By means of CRISE, the presence of regions is investigated in relation to the intensity threshold. The results for each of the six primary elements are discussed below and a selection of the regions with the highest *SP*'s or the lowest *IP*'s (interfering percentages) are summarized in separate tables.

The main element —C—H (see Table 2)

Specific regions are not present. Although —C—H vibrations are always active, in Raman spectroscopy some vibrations have insufficient intensity. In i.r. spectroscopy, potentially specific regions are split up by several interfering peaks.

There are several pseudo-specific regions. The highest *SP*'s appear, in both i.r. as well as in Raman spectroscopy, in the —C—H stretching area around 2900 cm^{-1} . This is due both to the generally high intensity of these bands and to the greater chance of interfering bands in other regions. As appears from Fig. 1A, the slope of *SP* as a function of intensity threshold is about the same for both techniques, and there is also no great difference in the maxima.

Selective regions are few, and appear in i.r. spectroscopy only. All regions show *IP*'s over 50%. It should be noted that only 10 compounds of the investigated 549 do not contain an alkyl group.

There are several combined regions. As the starting *SP* is already 90%, the profit gained from these regions is very small.

It can be concluded that alkyl groups can be detected by Raman as satisfactorily as by i.r. spectroscopy.

The main element C=C (see Table 3)

Specific regions are not present. For obvious reasons, a specific interval would have to be found in the C=C stretching area around 1600 cm^{-1} . As C=C=C compounds do not show a peak in this region but at 2000 cm^{-1} instead, a specific region is not found in the Raman or in i.r. spectra. Moreover,

TABLE 2

Selection of the (combined) intervals obtained for the main element —C—H (*IT*, intensity threshold; *SP*, score percentage; *IP*, interfering percentage; *mQ*, number of combined intervals.)

Infrared					Raman								
<i>IT</i>	(Pseudo-specific)	<i>SP</i>	Selective	<i>IP</i>	<i>IT</i>	Combination	<i>SP</i>	<i>IT</i>	(Pseudo-specific)	<i>SP</i>	<i>IT</i>	Combination	<i>SP</i>
03	2949—2926	57	3007—2887	50	10	<i>mQ</i> = 2	95	03	2878—2847	71	20	<i>mQ</i> = 2	97
10	2965—2923	88	1473—1383	70	10	<i>mQ</i> = 3	96	10	2937—2884	93	10	<i>mQ</i> = 3	98
20	2965—2923	87			10	<i>mQ</i> = 4	98	20	2997—2884	96	10	<i>mQ</i> = 4	99
30	2967—2633	91			10	<i>mQ</i> = 5	98	30	2997—2840	91	10	<i>mQ</i> = 5	99
40	3007—2675	90						40	2997—2835	89			
50	2988—2707	83						50	2997—2827	84			

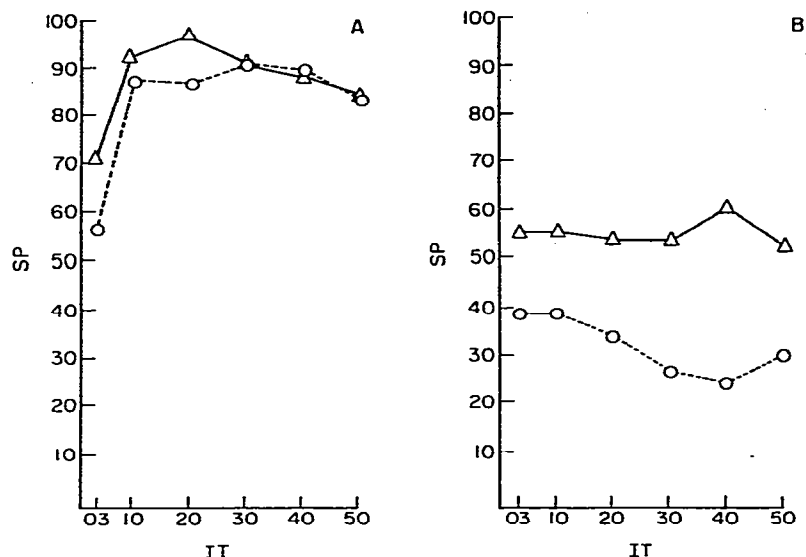


Fig. 1. The score percentage, SP , as a function of the intensity threshold, IT , for the main elements (A) $-C-H$, and (B) $C=C$. (---) Infrared; (—) Raman.

TABLE 3

Selection of the (combined) intervals obtained for the main element $C=C$ (Abbreviations as in Table 2.)

Infrared					Raman						
IT	(Pseudo-) specific	SP	IT	Combination	SP	IT	(Pseudo-) specific	SP	IT	Combination	SP
03	3042—3023	39	03	$mQ = 2$	54	03	1612—1569	54	40	$mQ = 2$	87
10	3042—3023	39	10	$mQ = 3$	65	10	1622—1562	54	40	$mQ = 3$	88
20	3042—3023	34	30	$mQ = 4$	69	20	1649—1500	53	40	$mQ = 4$	88
30	3042—3023	27	30	$mQ = 5$	73	30	1655—1500	53	40	$mQ = 5$	88
40	1508—1490	24				40	1669—1500	60			
50	1519—1490	30				50	1669—1500	52			

some double bonds are (pseudo-)symmetric and these vibrations do not absorb in the i.r. Also, in both i.r. and in Raman spectroscopy, there is interference from other structural elements ($C=O$ and $C=N$) in the $C=C$ region.

There are several pseudo-specific regions in both types of spectra. As can be seen from Fig. 1B, SP is only slightly influenced by the increase in the intensity threshold, particularly in Raman spectroscopy; although a raised threshold prevents weak peaks from scoring, it also involves less interference, so that the initial loss of SP is compensated by a broadening of the pseudo-specific region (see Table 3). It appears that the regions with the highest SP 's are due

to the C=C stretching vibration in Raman spectroscopy and to either the C=C (at an intensity threshold exceeding 30%) or the =C-H stretching around 3030 cm^{-1} in the i.r. spectra. This supports the rough-and-ready rule that the C=C band is more characteristic in Raman than in i.r. spectroscopy.

Selective regions are not present, for the above-mentioned reasons. There are several combined regions. Figure 2 shows that for Raman spectroscopy an *SP* of 87% is already attained with a combination of two *Q*'s, whereas for i.r. a 75% *SP* is not reached even with five. The gradual increase of *SP* in i.r. spectroscopy indicates that the regions are only slightly correlated; they may be related to some of the sub-elements of C=C.

It can be concluded that Raman is more powerful for the detection of C=C than i.r. spectroscopy.

The main element C=C (see Table 4)

There is one specific region, for Raman only. The region is clearly correlated to the C=C stretching vibration which is always Raman-active. No other bands are found in this region. The reason for the absence of an interval for i.r. spectroscopy is similar to that for C=C: a (pseudo-)symmetric C=C is infrared-inactive.

Pseudo-specific regions are few for i.r., and there is only one for Raman spectroscopy. In the i.r. spectra, the regions are related to either the $\equiv\text{C-H}$ or the C=C stretching vibration, the latter showing the highest *SP* (see Fig. 3A). For Raman spectroscopy, the specific region becomes pseudo-specific at intensity thresholds exceeding 30%.

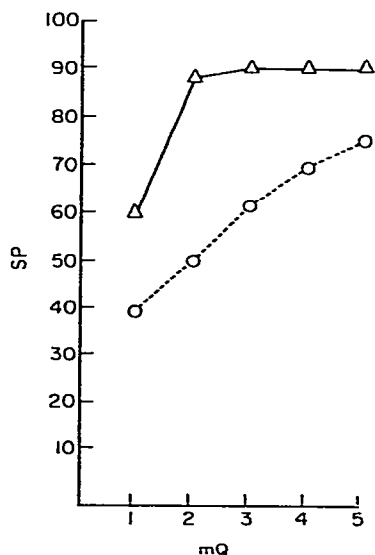


Fig. 2. The score percentage, *SP*, as a function of the number of combined intervals, *mQ*, for the main element C=C. (---) Infrared; (—) Raman.

TABLE 4

Selection of the (combined) intervals obtained for the main element $C\equiv C$
(Abbreviations as in Table 2.)

Infrared						Raman			
<i>IT</i>	(Pseudo-) specific	<i>SP</i>	<i>IT</i>	Combination	<i>SP</i>	(Pseudo-) specific	<i>SP</i>	Selective	<i>IP</i>
03	2112–2110	11	10	$mQ = 2$	58	2276–2098	100	2276–2098	0
10	2273–2118	36	10	$mQ = 3$	58	2320–2065	100	2320–2065	0
20	2273–2092	44				2320–2065	100	2320–2065	0
30	2273–2092	30				2303–2065	98		
40	2273–2092	23				2298–2098	98		
50	2273–2092	21				2276–2098	96		

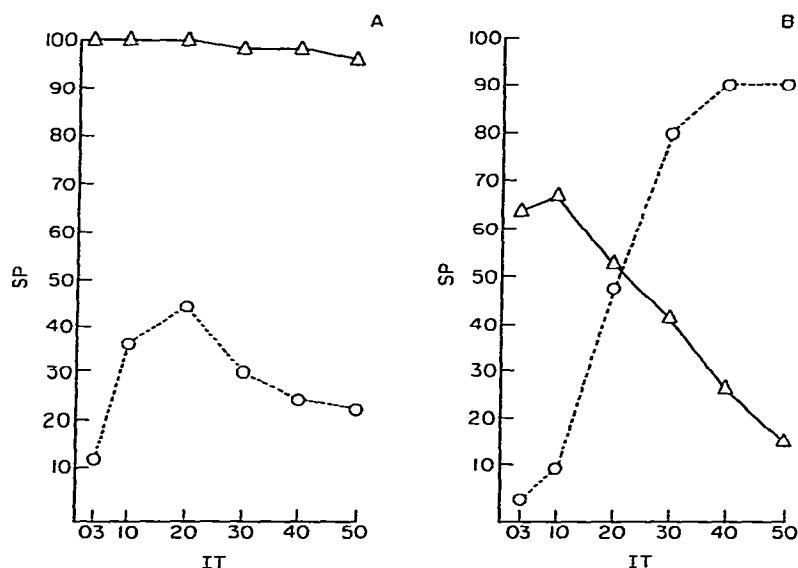


Fig. 3. The score percentage, *SP*, as a function of the intensity threshold, *IT*, for the main elements (A) $C\equiv C$, and (B) $C=O$. (---) Infrared; (—) Raman.

Selective regions are not present in i.r. spectra for the reasons set out above. There are a few combined regions. The maximum *SP* reached is 58% by combining two regions. Addition of an interval caused by the $\equiv C-H$ linkage does not increase the *SP*; the interval appears to be completely correlated with the former combination.

It can be concluded that Raman spectroscopy is eminently suited for the detection of this element.

The main element O—H (see Table 5)

Specific regions are not present. Although the O—H stretching vibration is strongly infrared-active, the presence of a specific region is prohibited by the

TABLE 5

Selection of the (combined) intervals obtained for the main element O—H
(Abbreviations as in Table 2.)

Infrared					
<i>IT</i>	(Pseudo-) specific	<i>SP</i>	<i>IT</i>	Combination	<i>SP</i>
03	3625—3602	4	50	$mQ = 2$	54
10	3350—3340	22	50	$mQ = 3$	57
20	3350—3323	35	50	$mQ = 4$	57
30	3350—3323	35			
40	3350—3323	35			
50	3350—3323	35			

carboxylic acids which do not show a band around 3300 cm^{-1} . Moreover, the $\equiv\text{C—H}$ and N—H stretching vibrations give interfering bands. For Raman spectroscopy, the lack of a specific region was expected, as the O—H is only slightly active.

There are several pseudo-specific regions in the i.r. spectra; a maximum *SP* of 35% is found. For Raman spectroscopy, no region reaches the minimum *SP* of 4%. Selective regions are absent for the reasons mentioned above. There are several combined regions for i.r. spectroscopy; a maximum *SP* is reached with five *Q*'s at an intensity threshold of 20%.

In conclusion, the presence of this element cannot be established by Raman and the score is rather small by i.r. spectroscopy.

The main element C—O (see Table 6)

Specific regions are absent, although the C—O stretching vibration is active in both i.r. and Raman spectroscopy. The absence of specific regions can be attributed to (i) the very low intensity ($< 3\%$) of some C—O bands, which in

TABLE 6

Selection of the (combined) intervals obtained for the main element C—O
(Abbreviations as in Table 2.)

Infrared							
<i>IT</i>	(Pseudo-) specific	<i>SP</i>	Selective	<i>IP</i>	<i>IT</i>	Combination	<i>SP</i>
03			1224—1026	97	50	$mQ = 2$	11
10			1228—1051	83	50	$mQ = 3$	13
20			1285—985	77	50	$mQ = 4$	15
30					50	$mQ = 5$	17
40	1189—1186	4					
50	1022—1018	7					

combination with the scanning conditions leads to the absence of a C—O band; and (ii) the strong coupling of C—O vibrations with skeletal modes; the bands are thus found over a wide region and interferences occur because many skeletal C—C vibrations give rise to peaks in the same region.

There are some pseudo-specific regions in the i.r. spectra, but the highest *SP* is only 9%. There are two selective regions in the i.r. spectra (*IP* > 80%) and several combined regions; for the latter regions, five *Q*'s yield a maximum *SP* of 17%. There are no analogous regions in Raman spectra.

It must be concluded that neither i.r. nor Raman spectroscopy is the obvious technique for the detection of the main element C—O.

The main element C=O (see Table 7)

Specific regions are absent. In the Raman spectra some C=O bands do not reach the minimum intensity of 3%; in both types of spectra, C=C bands interfere in the C=O stretching region.

There are two pseudo-specific regions in each type of spectra. As can be seen from Fig. 3B, the slope of *SP* as a function of intensity threshold for i.r. spectra is opposite to that for Raman spectroscopy. For i.r. spectra, this illustrates that for strongly active vibrations, and thus very intense bands, the influence of interfering bands decreases with an increase in the intensity threshold. For Raman spectra, *SP* decreases, as the intensity of the C=O bands is significantly smaller. Also, it can be seen that there are more interfering peaks at intensity thresholds below 30% in the i.r. spectra, as the *SP* in Raman is higher.

There is one selective region, for the i.r. spectra only. In contrast to all other selective regions, the *IP* found is small (8%). As the C=O band is always very intense in i.r. spectra, a high intensity threshold does not influence *SP* but it does decrease *IP* appreciably. There is only one combined region for each technique. The combination yields a maximum *SP* of 97% for i.r. and 69% for Raman spectroscopy.

It can be concluded that the presence as well as the absence of C=O can be established quite well, especially with the i.r. method.

TABLE 7

Selection of the (combined) intervals obtained for the main element C=O
(Abbreviations as in Table 2.)

Infrared							Raman					
<i>IT</i>	(Pseudo-specific)	<i>SP</i>	Selective	<i>IP</i>	<i>IT</i>	Combination	<i>SP</i>	(Pseudo-specific)	<i>SP</i>	<i>IT</i>	Combination	<i>SP</i>
03	3425—3423	7	1777—1631	53	50	<i>mQ</i> = 2	98	1793—1711	63	10	<i>mQ</i> = 2	69
10	1737—1728	10	1777—1631	28				1793—1708	66			
20	1737—1710	49	1777—1631	18				1765—1708	52			
30	1745—1684	81	1777—1631	15				1765—1708	41			
40	1795—1684	91	1777—1631	11				1763—1708	26			
50	1795—1684	91	1777—1631	8				1763—1708	15			

DISCUSSION

Some general remarks can be added to the detailed conclusions outlined above for the different structural elements.

Correlation

In agreement with earlier conclusions [2, 3], the number of specific regions is very small. As the main reason is the presence of interfering peaks, the number will be even smaller for larger basic files.

Pseudo-specific regions were always present in the i.r. spectra whereas in Raman those for C—O and O—H were absent. The *SP* values differ clearly for the structural elements investigated. Selective regions, except for C=O, show interfering percentages over 50%; as this type of region can be used only to establish the absence of an S_a , its usefulness seems very limited.

The combination of intervals (*Q*'s) sometimes produces a considerable increase of *SP*, but when the separate regions involved are mutually correlated, the gain is very small. Combinations of different techniques and of regions with different intensity thresholds have not been studied so far, because of difficulties in programming. These types of combinations may yield higher *SP*'s. A comparison of the *SP* maxima for i.r. and Raman spectroscopy (Table 8) shows that neither of the two techniques is significantly better than the other.

Intensity threshold

A uniform intensity threshold for all structural elements reduces the intensity data to a binary value: a peak is either present or absent. A high threshold reduces the number of peaks significantly, which is advantageous with respect to programming and to storage and processing of the spectral data. However, as can be seen from Table 8, the *SP* maxima of the elements investigated appear to be reached at different intensity thresholds, and it might be beneficial, therefore, to use various thresholds. For each S_a investigated, the maximum *SP* for a threshold of 10% is greater than or equal to the

TABLE 8

The maxima score percentage, *SP*, reached with one interval (*IT*, intensity threshold)

	Infrared		Raman	
	<i>IT</i>	<i>SP</i>	<i>IT</i>	<i>SP</i>
—C—H	30	91	20	96
C=C	10	39	40	60
C=C	20	44	20	100
O—H	50	35	—	—
C—O	50	7	—	—
C=O	50	91	10	66

SP for a threshold of 3%, which implies that peaks with intensities under 10% can be neglected. This supports a conclusion of van Eijk and van der Maas [4] who reported that loss of information starts at intensity thresholds exceeding 15%. It should be emphasized, however, that this phenomenon is valid for this particular file and the scanning conditions set. As Dupuis et al. [5] reported, there is a relatively small variation in information content of binary-coded spectra for intensity thresholds between 3 and 10%, except for files of saturated carbon-hydrogen compounds. The maximum intensity threshold was limited to 50%, because experiments indicated that the *SP* decreased for higher thresholds.

The study described above is a preliminary one, hence the number of structural elements investigated was limited to the six main ones. The procedure of establishing correlations between molecular structures and spectral data in relation to several variables is independent of the choice of the element and the basic file. The overall results of this study indicate that this approach to the systematic interpretation of vibrational spectra is promising. The CRISE program appears to offer the parameters necessary for the development of an interpretation system and for establishing the validity of the system.

REFERENCES

- 1 T. Visser and J. H. van der Maas, *Anal. Chim. Acta*, 122 (1980) 357.
- 2 T. Visser and J. H. van der Maas, *J. Raman Spectrosc.*, 7 (1978) 125.
- 3 T. Visser and J. H. van der Maas, *J. Raman Spectrosc.*, 7 (1978) 278.
- 4 C. G. A. van Eijk and J. H. van der Maas, *Fresenius Z. Anal. Chem.*, 291 (1978) 308.
- 5 P. F. Dupuis, A. Dijkstra and J. H. van der Maas, *Fresenius Z. Anal. Chem.*, 291 (1978) 27.