

the disadvantage of this method, it will be difficult to employ this sensor for the estimation BOD of the waste waters which contain organic compounds which can not be degraded or which can be assimilated very quickly by the microorganisms.

References

1. P. J. LeBlank, *J. Water Poll. Control Fed.*, **46**, 2202 (1974).
2. I. Karube, T. Matsunaga, S. Mitsuda, and S. Suzuki, *Biotechnol. Bioeng.*, **19**, 1535 (1977).
3. M. Hikuma, H. Suzuki, T. Yasuda, I. Karube, and S. Suzuki, *European J. Appl. Microbiol. Biotechnol.*, **8**, 289 (1979).
4. K. Riedel, R. Renneberg, M. Kühn, and F. Scheller, *Appl. Microbiol. Biotechnol.*, **28**, 316 (1988).
5. J. Kulys and K. Kadziauskiene, *Biotechnol. Bioeng.*, **22**, 221 (1980).
6. X. Zang, Z. Wang, and H. Jian, *Huanjing Kexue Xuebao*, **6**, 184 (1986); cited in *Chem. Abstr.*, **105**, 158286 (1986).
7. N. Kawabata and N. Nakamura, Abstracts of the International Symposium on New Sensors and Methods for Environmental Characterization, Kyoto, Japan (1986).
8. I. Karube, K. Yokoyama, K. Sode, and E. Tamiya, *Anal. Lett.*, **22**, 791 (1989).
9. Japanese Industrial Standard Committee, Testing Methods for Industrial Waste Water, JIS K 0102, Tokyo, Japan (1986), p. 47.
10. R. Rangarajan and G. A. Rechnitz, *Anal. Chem.* **47**, 324 (1975).
11. J. Mertens, P. V. den Winkle, and D. L. Mussart, *ibid.*, **48**, 272 (1976).
12. *Pure Appl. Chem.* **48**, 127 (1976).
13. IUPAC *Int. Bull.*, No. 1, 69 (1978).
14. I. Uemasu and Y. Umezawa, *Anal. Chem.*, **54**, 1199 (1982).
15. E. Pungor and Y. Umezawa, *ibid.*, **55**, 1432 (1983).

Expert System Approach for Vapor-Phase Infrared Spectra of Aromatic Compounds

Dongjin Pyo*, Hoon Hwang, and Junyeoun Lee

Department of Chemistry, Kangweon National University, Chuncheon 200-701

Received October 16, 1991

Computerized interpretation of vapor phase infrared spectra using a novel expert system approach for spectra/structure correlation for vapor phase spectra is introduced. Rapid identification of aromatic functional groups of components in gaseous mixture can be achieved using this expert system.

Introduction

The need for the identification of compounds present in complex gaseous mixture or after separation by gas chromatography (GC) using vapor-phase infrared (VPIR) spectrometry is becoming increasingly important. This is primarily due to advances in Fourier transform infrared (FT-IR) spectrometry allowing the acquisition of complete spectra in a few tenths of a second with detection limits of less than 1 ppb in long path infrared gas cells and below 10 nanograms in the GC/FT-IR interface. With a GC/FT-IR system hundreds of spectra can be generated per chromatogram. Clearly, the limiting processes are the spectral searching and in cases where the searching results are ambiguous, spectral interpretation steps. A powerful digital computer is needed to control the interferometer and perform the Fourier transformation from the time domain to the frequency domain in order to obtain the spectrum. This makes the incorporation of automatic spectral interpretation in the same computer with an obvious and sensible capability. It should be kept in mind that the ultimate goal of the chemist is not simply to produce a collection of spectra but rather to determine what chemical compounds are present in the unknown sam-

ples and what their significance is to the problem at hand.

The laborious task of spectral identification has been greatly simplified by the use of computer based library searching routines. Two general classes of algorithms have been developed: procedures that seek to make an exact identification of an unknown by direct comparison with spectra of known compounds (which require the spectrum of the unknown to be in the reference database), and more general approaches striving to identify all the functional groups of the unknown. A wide range of encoding schemes has been proposed¹ to represent spectra in a digital form suitable for library searching and interpretation. The ideal storage format retains the minimum amount of information necessary for correct identification of unknown compounds in a form permitting rapid numerical comparison. The optimum instrumental parameters for acquiring spectra for spectral searching and identification are a function of the sample, so that practical considerations require some degree of standardization in sampling and measurement. In addition, specific computer characteristics such as word size, amounts of memory, and types of mass storage devices also be taken into account.

Buechi *et al.*² have discussed the compilation of spectral libraries in terms of five operations: selection, digitization,

completion, formatting, and verification. For infrared spectral reference data bases, the problem of digitization has proven to be critical. Ideally the encoding scheme should be equally applicable to condensed-phase and vapor-phase spectra. Historically, however, the compilation of large libraries of condensed-phase spectra was begun over two decades before the commercial introduction of the first FT-IR spectrometer, so that most of these spectra were only available as hard copy of analog data. The problem of converting large databases of hard-copy spectra to a computer-compatible format has been approached by the use of graphics digitizers³. On the other hand, most vapor-phase reference spectra were measured using FT-IR spectrometers in the last ten years. Thus not only is their quality higher from an instrumental stand-point, but also they do not suffer from many of the other disadvantages of condensed-phase infrared spectra, such as shifting and relative intensity variation of bands depending on the method of sample preparation, baseline changes due to scattering, etc. Due to the inherent complexity of infrared spectra, definitive identification of unknown samples usually requires direct comparison with previously collected spectra. Most spectral searching algorithms are designed to scan through entire libraries in a sequential manner. These "brute force" methods have become only in recent years with mass storage devices. Much of the work of developing library-based spectral identification programs arose from the need to interpret the large number of spectra generated by GC/FT-IR experiments.

Although sequential searches can provide excellent results⁴, it is inefficient to search an entire spectral database to identify an unknown compound.

The situation is exacerbated when a good match is not found because of the limited size of the reference library. Investigation of a subset of the database, in the form of a class of known compounds, greatly reduces the amount of time required for the identification. In addition, the chemical functionality of molecules not present in the database can be ascertained. As an example of the success of this approach, Lytle⁶ was able to reduce the search time to less than 1% of a sequential search for the case in which seven characteristics were specified by reformatting the ASTM database of infrared spectra in an inverted format, to permit selective searching of library spectra based on the strongest absorption bands in the unknown spectrum.

Although exact compound identification is obviously most desirable in spectral searching algorithms, their success is always limited by the libraries being searched; that is, an unknown compound not in the library cannot be identified⁷. Consequently, a number of researchers have developed algorithms designed to interpret spectra rather than provide exact identification. Although these procedures rarely, if ever, provide an exact structural formula, they are very useful when a spectral library is not at hand or when the spectrum of the analyte is not contained in library. Functional group identification algorithms typically require many calculations and large mass storage devices that must be accessed very quickly. Complex decision trees have been developed⁸ that contain detailed information about known group frequencies. The operation of such algorithms typically requires that the IR spectrum be reduced to a list of peaks and then checked.

Procedures that attempt to analyze for the presence of specific functional groups either directly from the interferogram⁹ or after computing the Fourier transform of the absorbance spectra of the unknowns¹⁰ have also been developed, with the latter method showing considerable promise. By taking the Fourier transform of a spectrum containing peaks characteristic of a particular class of compounds, representative "synthetic interferograms", can be computed. Comparison of the unknown and group-specific interferograms is then done by calculating their dot product. This technique has been most successful in identifying functionalities that contain characteristic peaks well isolated from other classes of compounds.

This paper describes a new method for the interpretation of infrared spectra which, in contrast to earlier work, avoids the manual generation of rules for the interpretation of spectra. Our work was focused on the development of automatic characterization of spectral features of various functional groups for vapor-phase infrared spectra. The use of artificial intelligence to include spectral information for compound identification is a topic receiving increased attention. For example, the DENDRAL system¹¹ offered an important approaching method to modelling the work of a human operator in mass spectrometry. A computer program for an expert system type of analysis of infrared spectra was reported by Woodruff and Smith¹².

Although the goal of the automatic interpretation is the same as that of artificial intelligence approaches, the method of generation of rules are different.

In the artificial intelligence approaches, the generation of rules for the interpretation of spectra depends on more or less the knowledge of an experienced spectroscopist. In our work, rules are generated through the strict statistical treatment of infrared data from the database to identify patterns of absorbance that characterize certain functional groups. We have attempted to develop rules for which the position, intensity and width of each band in the spectrum are considered as important parameters to give structural information.

The preliminary results described in this paper indicate that our approach has better predicting power between sub-functionalities than any contemporary system, but also, once the spectra are correctly formatted, the speed of generating these rules is faster three or four times than other systems.

Experimental Section

The computer programs described were written in FORTRAN and run on an IBM personal computer. They are available from the authors. An IR database was used because it provided a large number of spectra, representing a variety of compound types. The database chosen was the EPA Vapor Phase Collection of 3300 spectra, available from James de Haseth at the University of Georgia. The spectra were stored in digital format on magnetic tapes. A general information "header" record, including compound name, formula, molecular weight, Chemical Abstracts Service (CAS) registry number, melting point, boiling point, Wiswesser line notation (WLN), etc., was followed by the digital spectra, measured at 2 cm⁻¹ resolution from 4000 to 450 cm⁻¹. The format of the records has appeared in the literature¹⁷.

For the application of most expert system software to find

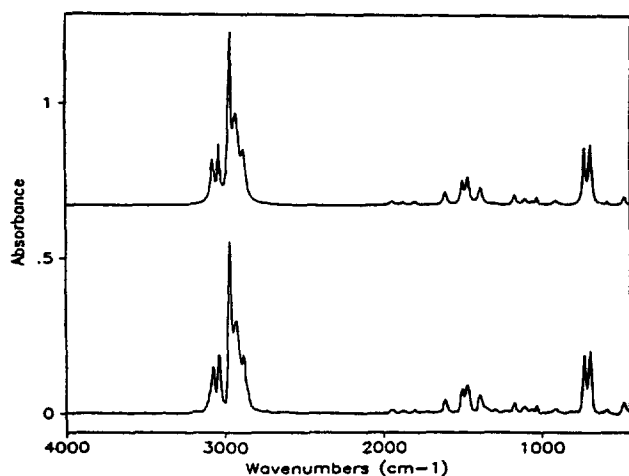


Figure 1. (a) Spectrum of isobutyl benzene, (b) Synthesized spectrum of 23 Lorentzian peaks from Table 1.

spectra/structure correlations, spectra must be changed from their usual format of (absorbance, wavenumber) pairs. We developed techniques for accurately calculating the center wavenumber, peak absorbance and full-width at half-height (FWHH), of all bands in the spectrum, including those bands that are partially or completely unresolved from a neighboring spectral feature. The technique that is being used involves first narrowing each band by Fourier self-deconvolution (FSD) and then fitting the deconvolved spectrum (in which all significant bands have been at least partially resolved) by a series of Lorentzian-Gaussian bands.

An algorithm was developed to find the optimum parameters for vapor-phase spectra. The widths of each band in a given spectrum are progressively reduced by an amount, Δ' . The series of spectra calculated in this manner is curve fit, and the sum of the squares of the difference (SSD) between each synthetic spectrum and the corresponding deconvolved spectrum is calculated. The conditions for calculating the most accurate band parameters can be found plots of SSD against Δ'^{13} . The band parameters calculated in this way are used as the input to the rule generation program.

Results and Discussion

In infrared spectra, it is usually impossible to determine the structure of any compound on the basis of an absorption band in one spectral region. To determine the structural units present in a compound by the use of group frequencies, it is therefore necessary to examine the entire infrared spectrum from 4000 to 450 cm^{-1} . Finding the group frequencies for each functional group is the most important step for structural elucidation using infrared spectra.

Obviously, the more accurate are the group frequencies, the better the structural elucidation for any compound. Once the group frequencies have been determined, band intensities and widths for those group frequencies become very useful parameters. We have therefore employed an algorithm which has generated very accurate band parameters (peak position, peak intensity, peak width) and utilized all these band parameters for spectra/structure correlation for infrared spectra.

Table 1. The Peak Table Spectrum of Isobutyl Benzene

Center	Height (Absorbance)	Width (FWHH)
3076.40	0.1367	27.00
3035.40	0.1701	15.43
2966.50	0.5188	21.22
2926.00	0.2617	42.43
2879.70	0.1321	30.86
2735.00	0.0029	13.50
1944.20	0.0128	30.86
1874.80	0.0088	28.93
1797.60	0.0119	30.86
1602.80	0.0434	25.07
1496.70	0.0691	17.36
1462.00	0.0892	28.93
1381.00	0.0581	30.86
1288.40	0.0049	38.58
1166.90	0.0342	23.15
1103.20	0.0216	30.86
1058.90	0.0113	17.36
1029.90	0.0262	15.43
921.30	0.0146	36.65
734.90	0.1811	19.29
696.30	0.1953	21.22
592.10	0.0130	11.57
486.10	0.0293	25.07

For the first experiments, band parameters for the vapor-phase spectra of thirty aromatic compounds were obtained. As an example, a spectrum (Figure 1a) of Isobutyl benzene (EPA database serial number 537) was converted to peak table spectrum (Table 1) using combined Fourier self-deconvolution and curve-fitting. Figure 1b is the synthesized spectrum of 23 Lorentzian peaks using peak parameters in Table 1. Excellent matches between these two spectra were shown. Band parameters for the vapor phase spectra of the thirty aromatic compounds listed in Table 2 were obtained in the same manner.

A computer routine was written to look for characteristic spectral features which are common to a certain compound class. In this routine, the peak table spectra are reformatted as binary points along the wavenumber axis, with all points one half width at half height (HWHH) on either side of the band center being represented as ones. All other points are represented as zeros (For more specific spectral representation, a region for each band that is bounded by a fraction of the HWHH can obviously be selected). As an example, the high wavenumber region of the spectrum of Isobutyl benzene show below.

```

000....001111.....111100000....000001111...11110000 etc.
|         |           |           |           |           |
| 4000 cm | 3104.....3020 |           | 2988...2848 |
|         |           |           |           |           |
(start)   (Aromatic C-H Stretch)   (Alkane C-H Stretch)

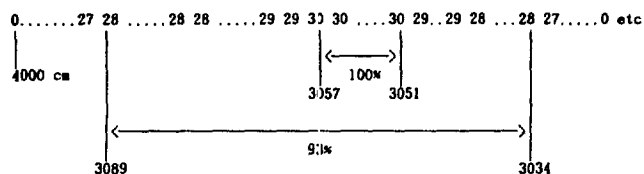
```

By overlapping spectra that have been formatted in this way, characteristic regions for a given functional group can be selected by finding the regions common to that class of spectra, *i.e.*, the regions where each spectrum contains ones.

Table 2. The List of 30 Aromatic Compounds as a Training Set (EPA Library)

0011	BENZENE, ETHOXY-,
0026	BENZENE, METHYL-,
0033	TOLUENE, 3,4-DICHLORO-,
0034	TOLUENE, 2,4-DICHLORO-,
0035	TOLUENE, P-CHLORO-,
0040	BENZENE, TERT-BUTYL-,
0075	BENZENE, 1,2,4-TRIMETHYL-,
0090	BENZENE, M-DIETHYL-,
0102	PHENOL, O-ETHOXY-,
0146	BENZENE, /1-CHLOROETHYL/-,
0187	DODECANE, 1-PHENYL-,
0200	PHENOL, O-ALLYL-,
0258	PHENOL, M-BUTYL-,
0394	ANILINE, M-FLUORO-,
0305	ANILINE, P-FLUORO-,
0311	BENZENE, 4-BROMO-1-ETHYL-,
0366	PHENOL, M-BROMO-,
0537	BENZENE, ISOPROPYL-,
0538	BENZENE, ETHYL-,
0539	BENZENE, SEC-BUTYL-,
0572	BENZENE, O-DIMETHOXY-,
0644	PHENOL, O-ISOPROPYL-,
0665	BENZENE, 1,4-DIFLUORO-,
0676	BENZENE, 1-BROMO-2-FLUORO-,
0677	BENZYLAMINE, P-ISOPROPYL-,
0689	BENZYL ALCOHOL, M-METHOXY-,
0799	BENZENE, 1,3,5-TRIMETHYL-,
0902	BENZENE, BUTYL-,
0981	BENZENE, ISOBUTYL-,
2525	BENZONITRILE, 2,4-DIMETHYL-,

An occurrence value for every wavenumber can be calculated simply by adding the binary spectra, so that if an occurrence value equal to the number of spectra is computed in a given region, that region would be a candidate for the spectra/structure correlation table. Using the binary spectral data for thirty aromatic compounds whose reference spectra were available in the EPA library of vapor-phase infrared spectra, occurrence values for every wavenumber were calculated by overlapping and summing these thirty binary spectra. Results



(Table 3) were obtained with threshold occurrence values of 100% (*i.e.*, all of the aromatic compounds in this region) and (*i.e.*, 28 of the 30 aromatic compounds absorbed in this region). As would be expected, broader regions were calculated with a threshold occurrence value of 93% than with a threshold occurrence value of 100% and additional spectral regions, 1064-1062 cm^{-1} , 1053-1051 cm^{-1} corresponding to the in-plane bending vibrational modes of the C-H moiety,

Table 3. Characteristic Regions Search Results for Aromatic Compounds with Thresholds 100% and 93%

100%	93%
1. 3057-3051 cm^{-1}	1. 3089-3034 cm^{-1}
2. 1610-1598 cm^{-1}	2. 1620-1581 cm^{-1}
3. 1500-1494 cm^{-1}	3. 1504-1479 cm^{-1}
	4. 1064-1062 cm^{-1}
	5. 1053-1051 cm^{-1}

Table 4. Statistical Calculations for Each Characteristic Region of Aromatic Compounds with 100% Threshold

Characteristic region	Average intensity	Intensity deviation	Ratio	Average width	Width deviation
3057-3051	0.17	0.16	0.96	36.00	17.25
1610-1598	0.24	0.26	1.07	26.42	6.79
1500-1494	0.47	0.44	0.92	20.52	5.78

were added.

In most previous work (*eg.*, 15), detailed absolute or relative peak intensity information has rarely been used in the interpretation of infrared spectra.

For the Program for the Automated Interpretation of Infrared Spectra (PAIRS) relative band intensity on a scale from 1 to 10, has been used (16). Band width information has rarely been used in the computerized interpretation of infrared data, although the width of any band can be input to PAIRS in a semiquantitative manner by defining it as sharp, medium, or broad.

In an attempt to get less ambiguous structural information than can be found using only wavenumber-based spectra/structure correlations, we have attempted to incorporate intensities and widths in the input to the program developed in our laboratory. To characterize the band intensity in a given spectral region, the absorbances of all spectra were averaged between wavenumbers located between one HWHH either side of the center wavenumber; the average relative peak intensity and standard deviation of relative intensities was then calculated in each region with a 100% threshold value (Table 4). Relative standard deviation values were used because absolute standard deviations tend to increase as intensities increase, and some weak bands are useful predictions of molecular structure. The lower are the values of relative standard deviation for the parameters of any band, the more characteristic is that band. Similar calculations have been made for peak widths. Average width (FWHH) and standard deviation of widths for each characteristic regions were calculated.

To predict the presence or absence of various functionalities, it is necessary to find probability values of correlations between structures and unknown infrared spectra. This probability value is called the confidence value in this paper. The steps taken to calculate the confidence value are as follows: the program looks at each spectrum only in the characteristic regions for the particular functionality (*eg.*, aromatic compounds). Each region is weighted differently according to the intensity deviation and width deviation value found

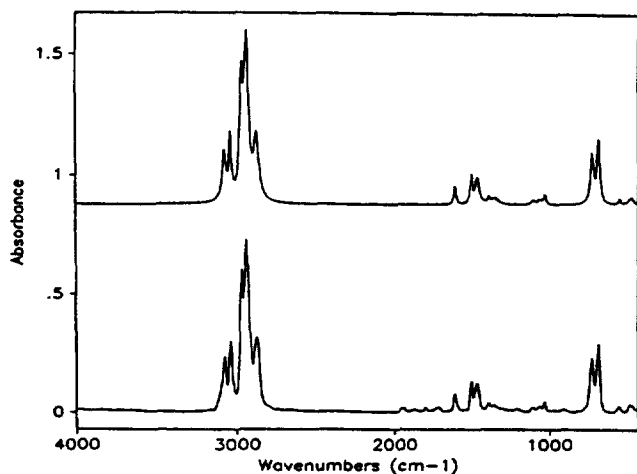


Figure 2. (a) Spectrum of butyl benzene, (b) Synthesized spectrum of 17 Lorentzian peaks from Table 5.

during the rule generation step, with the regions with lower deviation values being weighted more strongly than other regions. Standard deviation values are divided by the intensity values and then this values are inversed. The inversed values are used as weighting factors for each region. If the test spectrum has peaks in those regions, scores are given to it with different weighting factors. The decision-making routine looks for that region first, if no band is present in that region, the routine proceeds to the next functional group automatically.

For each region, relative peak intensity and peak width of the test spectra are tested to determine if those values are in the 99% confidence levels of intensities and widths. If the peak intensity of width values fall in the 99% confidence level limit ($x-2.586 < \mu < x+2.586$), scores are added for that compound. Scores are then converted to percentage confidence value of the particular class.

For the aromatic compounds, the lowest relative standard deviations both for intensities and widths were found two bands, namely the C-H stretching and C=C stretching band. Since all these compounds were aromatic compounds, these results give rise to a certain amount of optimism that relative band intensities will prove to be more useful in structural elucidation than has been recognized previously.

It is most important for the infrared spectroscopist to make decisions about functionalities. Many expert systems investigate more than 100 different classes, divided into major classes (aromatics, acids, esters, alcohols, etc.) and subclasses (monosubstituted benzenes, acetates, unsaturated acids, etc.).

To this point, we have only investigated an aromatic functionality; expanding the system to a much larger number of functionalities is routine but time-consuming in view of the time required for combined FSD and curve-fitting of all spectra in the database.

A confidence value (from 0% to 100%) was generated for each functional group investigated. The spectrum of butyl benzene (database serial number, 902) is used to illustrate the implementation of this approach. The original spectrum is shown in Figure 2a. The peak parameters generated by combined FSD and curve-fitting are shown in Table 5 and

Table 5. The Peak Table Spectrum of Butyl Benzene

Center	Height (Absorbance)	Width (FWHH)
3074.50	0.2111	28.93
3035.90	0.2640	17.36
2968.40	0.4550	23.15
2937.50	0.6766	32.79
2875.80	0.2689	36.65
1604.70	0.0804	17.43
1498.60	0.1214	15.43
1460.10	0.1206	32.79
1384.80	0.0285	17.36
1344.30	0.0280	54.01
1103.20	0.0213	28.93
1060.80	0.0204	28.93
1029.90	0.0455	17.36
736.80	0.2172	25.07
696.30	0.2746	21.22
565.10	0.0270	15.43
491.80	0.0345	38.58

Table 6. The PAIRS Table Spectrum of Butyl Benzene

Position (cm ⁻¹)	Relative intensity	Width
491	1	2
565	1	1
696	5	2
736	4	2
1029	1	2
1060	1	2
1103	1	2
1344	1	3
1384	1	2
1460	2	2
1498	2	2
1604	2	2
2875	4	2
2937	10	2
2968	8	2
3035	4	2
3074	4	2

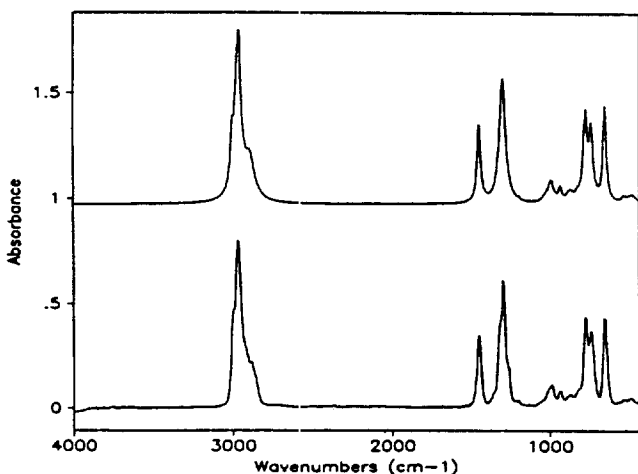
the synthetic spectrum reconstructed from these parameters is shown in Figure 2b. The spectrum contains peaks in all three characteristic regions for aromatics, and the confidence value that this compound is a aromatic compound was calculated to be 99.50%.

To compare the result obtained using our approach with the result obtained using PAIRS, the same spectrum was converted to a PAIRS peak table (Table 6). Using this data, the result obtained by PAIRS is shown in Table 7. The functional group with the highest expectation value was the thiophene group and aromatic group showed a 65% expectation value.

To illustrate the result of applying the rules generated for aromatic compound to compounds not containing a aro-

Table 7. The Results of PAIRS for Butyl Benzene

Group name	Expectation value
Thiophene	0.80
Aromatic	0.65
Olefin-(Non-Arom)	0.60
Olefin-CHR=CHR(C)	0.60
Methyl	0.50
Olefin-CHR=CR2	0.20
Olefin-CH2=CR2	0.20
Olefin-CHR=CHR(T)	0.20
Amine	0.20
Amine-1-A-Branch	0.20
Methylene	0.20
Arom-1.3-Subst	0.20
Arom-1.2-Subst	0.20
Alcohol-Primary	0.15
Alcohol-Secondary	0.15
Alcohol-Tertiary	0.15
Amine-2-Ali	0.15
Alcohol	0.15
Arom, -Monosubst	0.13
Olefin-CHR=CH2	0.13
Lactam	Unknown Group

**Figure 3.** (a) Spectrum of 1,4-dichloro benzene, (b) Synthesized spectrum of 16 Lorentzian peaks using parameters in Table 8.

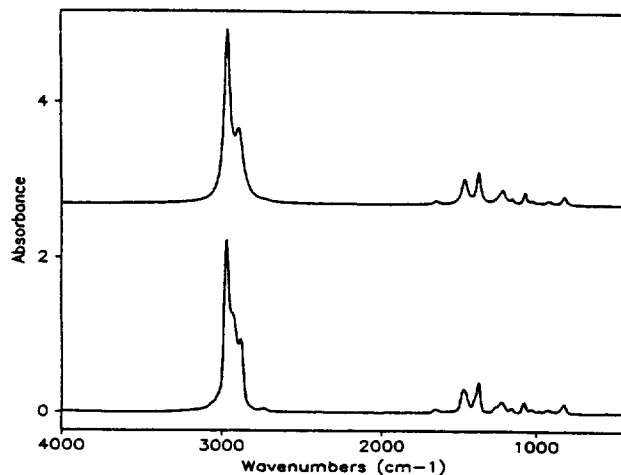
matic group, spectra of several other compounds were analyzed.

The first of these was 1,4-Dichloro butane (database serial number, 29). Figure 3a and 3b show the original spectrum and the synthesized spectrum respectively, and Table 8 shows the peak table spectrum for this compound. This compound yielded a 0% confidence value that it fell in the aromatic class because of the complete absence of all peaks in the aromatic stretching region, 3035-3051 cm^{-1} , 1610-1598 cm^{-1} , 1500-1494 cm^{-1} .

The second example was 2,4,4-trimethyl 2-pentene (serial number, 18) of which the original spectrum, synthesized spectrum and peak table spectrum are found in Figure 4a,

Table 8. The Peak Table Spectrum of 1,4-Dichloro Butane

Center	Height (Absorbance)	Width (FWHH)
3005.00	0.1881	19.29
2966.50	0.7646	46.29
2893.20	0.1815	69.44
1450.40	0.3615	27.00
1300.00	0.5839	46.29
1197.70	0.0079	11.57
1029.90	0.0219	52.08
993.30	0.0984	42.43
935.50	0.0622	25.07
875.70	0.0474	46.29
829.40	0.0284	32.79
779.20	0.3927	30.86
742.60	0.3122	28.93
659.70	0.4450	25.07
545.80	0.0213	30.86
495.70	0.0417	73.29

**Figure 4.** (a) Spectrum of 2,4,4-trimethyl 2-pentene, (b) Synthesized spectrum of 14 Lorentzian peaks using parameters in Table 9.

4b and Table 9.

Our system predicted a confidence value of 33.71% for the aromatic class. This is due to the fact that the spectrum of 2,4,4-trimethyl 2-pentene has a middle alkane bending band near 1500-1494 cm^{-1} region that overlaps with the C=C stretching region. However, it does not have any spectral feature near the C-H stretching region, 3057-3051 cm^{-1} , and also does not match with 1610-1598 cm^{-1} characteristic band of aromatic compounds.

To evaluate our system, a lot of spectra of compounds containing different functional groups were tested. Excellent results (Table 10) were obtained.

After creating a percentage confidence value for a functional group, the program should go through other different functionalities. To predict the presence or absence of various functionalities of an unknown compound, it is necessary to find representative spectral characteristics for various functionalities. The method which is developed only for one func-

Molecular Structure of Bicyclo[4.2.2]decapentaene

Oh Seuk Lee*, Yi Hwa Lee, and Eiji Ōsawa†

Department of Chemistry, Andong National University, Kyoungbuk 760-749

*Department of Knowledge-Based Information Engineering, Toyohashi University of Technology,

Toyohashi 441, Japan. Received October 19, 1991

Extensive search over the energy surface of bicyclo[4.2.2]decapentaene with MMP2 molecular mechanics method and AM1 semiempirical MO method revealed only one, deep energy minimum structure, which corresponds to **1**. The alternative structure **2** could not be identified as a stationary point. Although the deviation of benzenoid ring from planarity is large in the energy minimum structure ($\phi = 26^\circ$ (MMP2), 37° (AM1)), the bond lengths show no severe alternation.

Introduction

Geometrically distorted unsaturated compounds have attracted considerable attention because of their remarkable properties and chemical reactivities.¹⁻⁷ Among those compounds are cyclophanes of which structural features provide crucial informations on ring strain and aromaticity. The X-ray crystal structure analyses of [6]paracyclophanes⁸ and [5]metacyclophane⁹ showed that alternating bond lengths were not observed although the aromatic rings were greatly distorted from planarity. The available ¹H-NMR and ¹³C-NMR data for the most of the small [n]cyclophanes except for the very small [4]cyclophanes suggest intact aromaticity even for the smallest members of the family.¹⁰ Due to their unstability [4]cyclophanes can only be detected as a transient species so that calculational results become important.

The generation of bicyclo[4.2.2]decapentaene as a transient species which can be considered to be more strained than [4]paracyclophane was recently reported by Tsuji and Nishida.¹¹ The structure of bicyclo[4.2.2]decapentaene can be envisaged as at least two unique alternatives, namely [4]paracyclophadiene (**1**) and 1,6-etheno bridged *cis,trans,cis,trans*-cyclooctatetraene (**2**). Tsuji concluded that bicyclo[4.2.2]decapentaene is best represented by the structure **1** but that a possibility of structure **2** as a local minimum on the potential surface cannot be ruled out. We present in this article the structure of bicyclo[4.2.2]decapentaene as calculated by MMP2 molecular mechanics method¹² and AM1 semiempirical MO method.¹³

Results and Discussion

The two-parametric torsional energy surface of bicyclo[4.2.2]decapentaene was investigated in order to elucidate the relative stability of **1** and **2**. Whereas the dimensions of mole-

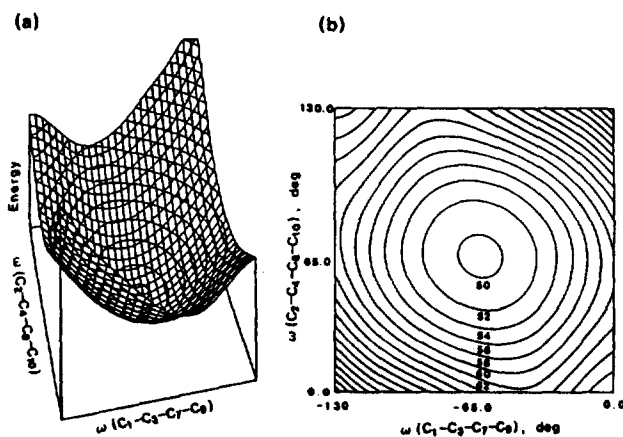
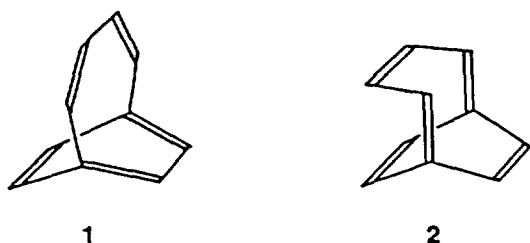


Figure 1. MMP2 torsional energy surface of bicyclo[4.2.2]decapentaene with dihedral angles at $C_1-C_3-C_7-C_9$ and $C_2-C_4-C_8-C_{10}$ as two variables. (a) Three-dimensional perspective. (b) Contour map. Energies in kcal/mol. See Figure 2 for numbering of atoms.

Table 1. MMP2-Calculated Bond Distances of Bicyclo[4.2.2]decapentaene in Å

C(1)-C(2)	1.385	C(5)-C(6)	1.385
C(1)-C(3)	1.406	C(3)-C(5)	1.407
C(2)-C(4)	1.406	C(4)-C(6)	1.407
C(3)-C(7)	1.516	C(4)-C(8)	1.516
C(7)-C(9)	1.353	C(8)-C(10)	1.353
C(9)-C(10)	1.476	C(2)-H(12)	1.104
C(1)-H(11)	1.104	C(16)-H(14)	1.104
C(5)-H(13)	1.104	C(8)-H(16)	1.103
C(7)-H(15)	1.103	C(10)-H(18)	1.104
C(9)-H(17)	1.104		

cular energy surface are generally much higher than two, there will be cases in conformational processes of small molecules where the energy surface can be more or less adequately described by taking only two dihedral angles as variables.¹⁴ Two dihedral angles involving bridgehead carbon atoms were chosen to characterize the conformational energy surface.

As shown in Figure 1 only one deep energy minimum

Table 9. The Peak Table Spectrum of 2,4,4-Trimethyl 2-Pentene

Center	Height (Absorbance)	Width (FWHH)
2964.50	2.0997	44.36
2891.20	0.7908	65.58
2727.30	0.0152	54.01
1652.90	0.0445	38.58
1465.80	0.3160	46.29
1373.30	0.3925	32.79
1251.80	0.0210	67.51
1219.00	0.1713	48.22
1155.30	0.0562	17.36
1074.30	0.1472	23.15
1028.00	0.0433	25.07
983.70	0.0164	25.07
923.90	0.0468	46.29
827.50	0.1127	30.86

Table 10. The Results of Calculating Confidence Values for Aromatic Functional Group

Name	Confidence value (%)
1-Octanol	27.12
Methyl cyclohexane	0.00
2,4,4-Trimethyl 2-pentene	27.12
Butric acid, ethyl ester	25.69
Cyclohexane	0.00
1,4-Dichloro butane	0.00
3-Methyl pentane	14.73
2,2,5-Trimethyl hexane	14.73
Dichloro methane	0.00
2-Methyl 2-pentane	25.69
N,N-Diethyl aniline	99.50
Phenyl ether	99.50
Cyclohexyl benzene	99.50
N,N-Diethyl benzylamine	99.50
Pentyl benzene	99.50
Benzoic acid methyl ester	99.50
M-Bromo aniline	99.50
M-Iodo aniline	85.32
Benzoic acid isobutyl ester	97.71
M-Nitro toluene	72.45

tional group can be used to select spectra which warrant further interpretation, whether that interpretation be by use of library searching, artificial intelligence, or manual spectral identification. Although the exact interpretation of infrared

data is usually a difficult and takes, we will continue to investigate other functional groups to complete the whole expert system.

In conclusion, this new approach of interpretation of infrared data based on inputting accurate band parameters to the expert system has potential for rapid identification of the components of the complex mixture samples. The methods can also be used in generating vapor-phase infrared spectra/structure correlation rules to obtain a better understanding of the correlation.

Acknowledgement. This research was supported by the Korea Research Foundation (1990) and Sunil Development Engineering.

References

1. J. T. Clerc and J. Zupan, *Pure Appl. Chem.*, **49**, 1827 (1977).
2. R. Buechi, J. T. Clerc, and C. Jost, *Anal. Chim. Acta*, **103**, 21 (1978).
3. M. F. Delaney and P. C. Uden, *Anal. Chem.*, **50**, 2156 (1978).
4. M. D. Erickson, *Appl. Spectrosc.*, **35**, 181 (1981).
5. P. R. Griffiths and J. A. de Haseth, "Fourier Transform Infrared spectrometry", Wiley-Interscience, New York, pp. 607-608 (1986).
6. F. E. Lytle, *Anal. Chem.*, **42**, 355 (1970).
7. J. Kwiatkowski and W. Riepe, *Anal. Chim. Acta*, **135**, 285 (1982).
8. H. B. Woodruff and G. M. Smith, *Anal. Chem.*, **52**, 2321 (1980).
9. R. C. Wiebolt, B. A. Hohne, and T. L. Isenhour, *Appl. Spectrosc.*, **34**, 7 (1980).
10. J. A. de Haseth and L. V. Azarraga, *Anal. Chem.*, **53**, 2292 (1981).
11. J. Lederberg, in "Biochemical Applications of Mass Spectrometry" (Georgy R. Weller, ed.), Wiley-Interscience, New York (1972).
12. H. B. Woodruff and G. M. Smith, *J. Chem. Inf. Comp. Sci.*, **24**, 33 (1984).
13. J. A. Pierce, P. R. Griffiths, and H. Gao, in "Computer-Enhanced Analytical Spectroscopy" (H. L. C. Meuzelaar and T. L. Isenhour, eds.), Plenum Publishing Co., New York, pp. 29-54 (1987).
14. D. R. Preuss and P. C. Jurs, *Anal. Chem.*, **46**, 520 (1974).
15. D. S. Frankel, *Anal. Chem.*, **56**, 1011 (1984).
16. S. A. Tomellini, J. M. Stevenson, and H. B. Woodruff, *Anal. Chem.*, **56**, 67 (1984).
17. P. R. Griffiths, L. V. Azarraga, J. A. de Haseth, R. W. Hannah, R. J. Jakobson, and M. M. Ennis, *Appl. Spectrosc.*, **33**, 543 (1979).