Prediction of chemical structure of aromatic hydrocarbons by pattern recognition of spectral lines obtained by fluorescence spectrometry in a supersonic jet

Cheng-Huang Lin, Totaro Imasaka *

Department of Chemical Science and Technology, Kyushu University, Hakozaki, Fukuoka 812, Japan

Abstract

A technique based on pattern recognition of data obtained by supersonic jet spectrometry is employed for the prediction of chemical structure. The degree of similarity is evaluated quantitatively by calculating a cross correlation factor between sample and reference molecules. A probability density function is determined by fitting the data to a specified equation. The functional group and its position are also predicted by a similar technique. The pattern recognition provides a method for prediction of the chemical structure and is applicable to samples that have not been examined by supersonic jet spectrometry.

1. Introduction

Supersonic jet spectrometry provides sharp line structure in ultraviolet and visible spectra [1-4]. This analytical method, therefore, shows selectivity and allows spectrometric differentiation between molecules with similar chemical structures. The application of mass spectrometry after multiphoton ionization gives information relating to the molecular weight of the sample. This information alone, however, is not always sufficient for differentiating between isomers. Although the advantage of supersonic jet spectrometry lies in its well-resolved spectral features, the mass resolution of this technique could be improved by expanding the gas and decreasing its velocity distribution, since a better mass resolution can be obtained by other techniques such as ion cyclotron resonance. It is the sharp spectral features given by supersonic jet spectrometry that are important and useful in the assignment of chemical species and structure.

When a complete database is available, a sample molecule can be accurately identified using the spectral data accumulated for a series of authentic compounds. Recently, a database constructed by accumulating the wavelengths of pure electronic (0-0) transitions was used for the assignment of aromatic hydrocarbons [5-7]. This approach is reliable, because of the sharp line structure of the spectrum. However, there is no straightforward relationship between spectrum and chemical structure. It is, therefore, difficult to assign the chemical species when no standard spectrum is available in the database for that species. It would be useful, therefore, to develop a new method for data analysis, which can be applied to samples whose spectra are not available, i.e. which have not been reported.

One possible approach to this problem involves theoretical calculations of spectral parameters. The energy for the 0-0 transition is estimated by calculation of the molecular orbital by the Pariser–Parr–Pople (PPP) method and by the complete neglect of differential overlap method for spectroscopy (CNDO/S) [8]. These methods are useful for

* Corresponding author. Fax: (81) 92-651-5606.

reducing the number of candidates in the assignment. However, the error in the estimation of the wavelength for the 0–0 transition is 1–3%, which is much larger than the error of the peak position obtained experimentally (<0.01%). For more accurate discussion, it is necessary to use a spectral pattern for assignment. In physical chemistry, the spectral profile, i.e. transition energy and intensity, is calculated from the wave function and the Franck–Condon factor by solving a Schrödinger equation. This may give additional information for assignment of the chemical species. However, this approach is not altogether accurate, and the error in the calculation of the vibrational energy sometimes exceeds 10%. In addition, such an approach is complicated and time-consuming even for a single molecule, and as a result is not suitable for applications to analytical spectroscopy. What is ideally needed is a method which will provide sufficient information for the prediction of chemical structure without a standard spectrum of the sample or a complicated theoretical calculation of spectral parameters.

This study describes an approach involving the recognition of patterns of spectral lines which are produced as a result of supersonic jet spectrometry. In order to correlate spectral feature with chemical structure, we calculate a cross correlation factor (CCF) between the spectra obtained for the sample and reference molecules. The line shape in the supersonic jet spectrum is so narrow that almost all the CCF values become zero when they are calculated for different compounds. The data, therefore, were optimized by expanding the spectral line to a specified function, e.g. a gaussian function, thus giving useful information. In order to evaluate the degree of spectral similarity quantitatively, the probability density is plotted against the CCF value for a series of aromatic hydrocarbons. In addition, the spectral data of the 0–0 transition are used for prediction of the chemical structure. The wavelength of the 0–0 transition hardly coincides with that of a different compound, but is correlated quantitatively with a probability density function, providing useful information for structure prediction. The probability densities obtained independently by the above two methods are then multiplied to improve the accuracy of prediction. The advantage of this pattern recognition is demonstrated by differentiation of the anthracene derivatives against non-anthracene compounds. The functional group and its position can also be predicted relatively. The present approach requires neither a standard spectrum of the sample nor a knowledge of complicated quantum chemistry, and it works quite well for assignment of the chemical species in supersonic jet spectrometry.

2. Experimental

2.1. Calculation

Fig. 1 shows a bar graph of the fluorescence excitation spectra for anthracene and its derivatives obtained by supersonic jet spectrometry. Similar vibrational structure is observed in the spectra of the compounds with similar chemical structure. This is consistent with the fact that vibrational structure is strongly associated with skeletal vibration of the anthracene ring. These similarities can be visually observed, but for quantitative purposes the degree of similarity should be calculated mathematically. It is possible to use a CCF value for evaluation of the similarity between the standard and reference molecules [9]. However, the spectral line is so narrow (<0.01 nm) that almost all the CCF values between different compounds become zero. Therefore, optimization is necessary to evaluate the degree of similarity.

In this study, the signal peak in the supersonic jet spectrum is assumed to have a gaussian distribution:

$$I_k(v) = \exp[-(v - v_k)^2/r^2] \qquad (1)$$

where $I_k(v)$ is the fluorescence intensity at the excitation frequency of v for peak k, and r is the linewidth assumed. The spectra of the chemical species $X(v)$ and $Y(v)$ are represented by

$$X(v) = \sum_1^n X_k(v) \qquad (2)$$

$$Y(v) = \sum_1^n Y_k(v) \qquad (3)$$

where n is the total number of spectral lines. In order to compare the vibrational structure in the excitation spectrum, it is necessary to subtract the frequency of the 0–0 transition from each peak position. The CCF value, C, is calculated between the standard and reference spectra by

$$C = \frac{\left[\displaystyle\int X(v - v_{0X}) \cdot Y(v - v_{0Y}) \, dv \right]}{\left[\displaystyle\int X^2(v - v_{0X}) \, dv \cdot \int Y^2(v - v_{0Y}) \, dv \right]^{1/2}}$$

(4)

where v_{0X} and v_{0Y} are the frequencies of the 0–0 transition for compounds X and Y, respectively. The denominator is a normalization factor. The CCF value is close to unity when the spectral feature of the standard is similar to that of the reference; CCF = 1 when CCF is calculated between the same compounds. In contrast, the CCF value is close to zero when the spectral features are completely different from each other. The CCF value also depends on the parameter r. In order to differentiate between the compounds giving similar spectra, this value should be small. However, if it is too small, all the CCF values become zero. Thus, appropriate fuzziness must be introduced by optimizing this parameter.

The CCF value was calculated using a notebook-type personal computer (NEC, PC-9801n). The program was written in BASIC. All the spectral data were cited from the references reported by other authors and from the data measured by ourselves. After digitizing the spectral data (frequency and intensity), the excitation spectrum was calculated using the assumed value for r. The CCF value was calculated between the spectra for the standard and reference molecules. The time period of calculation was typically 10 min.

3. Results and discussion

3.1. Preliminary study

The CCF values are calculated for various anthracene derivatives using 2-ethylanthracene as a standard sample. These results are shown in Table 1. The linewidth, r, is changed from 0.1 to 10 cm^{-1}. When r is assumed to be 0.1 cm^{-1}, it is difficult to evaluate the degree of similarity properly. In the present study, the optimum value of r appears to be between 1 and 10 cm^{-1}. The CCF values calculated for alkylanthracenes are significantly larger than those for chloroanthracenes, suggesting that the standard sample contains an alkyl group rather than a chlorine atom. The CCF values for anthracene derivatives substituted at position 2 are larger than those substituted at
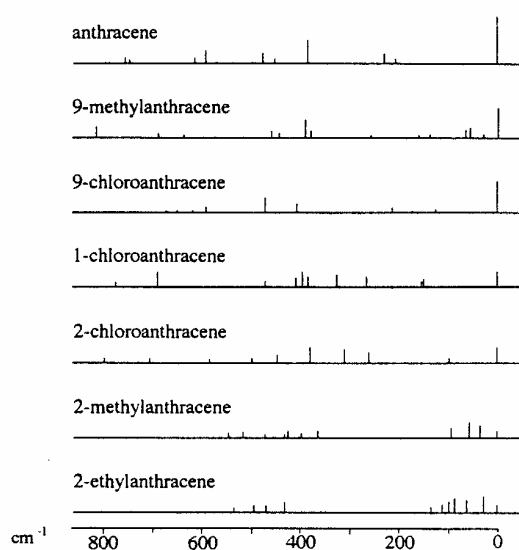

Fig. 1. Fluorescence excitation spectra for anthracene derivatives.

position 9, indicating a substituent group at position 2. The CCF value is not unity for 2-methylanthracene, indicating that the sample is not 2-methylanthracene. However, the large CCF value for 2-methylanthracene implies that the chemical structure is closely related, structurally, to 2-methylanthracene, and is possibly 2-ethylanthracene. This success in the preliminary study strongly encouraged us to perform further systematic investigations.

3.2. Skeletal structure

The CCF values were calculated for various aromatic hydrocarbons with different skeletons using 9-methylanthracene as a standard sample. A similar study was carried out using 1-methylnaphthalene. The results are shown in Table 2. The CCF values are the largest for the compounds with the same skeletons. Thus the skeletal structure can readily be predicted by this method.

Table 1
CCF values calculated for various aromatic hydrocarbons using 2-ethylanthracene as a standard sample

Compound	$r = 0.1$	$r = 1$	$r = 10$	Ref.
2-Ethylanthracene	1	1	1	[6]
2-Methylanthracene	0.243	0.301	0.495	[6]
9-Methylanthracene	0.001	0.060	0.236	[6]
2-Chloroanthracene	10^{-4}	0.008	0.048	[6]
9-Chloroanthracene	10^{-12}	0.004	0.040	[6]
Anthracene	10^{-19}	10^{-4}	0.002	[6]
1-Chloroanthracene	$<10^{-20}$	$<10^{-20}$	0.001	[6]

r: cm^{-1}.

C.-H. Lin, T. Imasaka / Talanta 42 (1995) 1111–1119

Table 2
CCF values calculated for various aromatic hydrocarbons with different skeletons using 9-methylanthracene and 1-methylnaphthalene as standard samples

Compound	9-Methylanthracene		1-Methylnaphthalene		Ref.
	$r = 1$	$r = 10$	$r = 1$	$r = 10$	
Benzene	0.002	0.017	10^{-15}	0.004	[12]
Indole	10^{-15}	0.001	$<10^{-20}$	10^{-5}	[13]
Naphthalene	10^{-6}	0.012	0.044	0.058	[14]
Fluorene	$<10^{-20}$	10^{-5}	10^{-6}	0.036	[15]
Anthracene	0.080	0.605	10^{-10}	0.001	[6]
Pyrene	0.002	0.026	10^{-5}	0.003	[16]

r: cm^{-1}.

A similar comparison was also carried out using toluene (methylbenzene) as a standard sample. The CCF values calculated for the same aromatic hydrocarbons are shown in lines 1–6 (benzene–pyrene) of Table 3. Surprisingly, the value for benzene is small; the values for pyrene and naphthalene are much larger at $r = 1$ cm^{-1} and those of indole and naphthalene at $r = 10$ cm^{-1}. As a result, the CCF values were also calculated for naphthalene derivatives (7–11, Table 3). However, the CCF values are not so large for naphthalene derivatives. The values for benzene derivatives were also calculated as trials. The results are shown in lines 12–26 of Table 3. Clearly, large CCF values are obtained for benzene derivatives, especially for alkylbenzenes. It should be noted that benzene has a high molecular symmetry (D_{6h}) and, as a result, the electronic transition ($^1A_{1g}$–$^1B_{2u}$) is essentially forbidden. Consequently, the spectrum is observed only when the electronic transition is coupled with vibrations breaking D_{6h} symmetry. This situation is completely different from other benzene derivatives. Thus benzene is a special case and is not a representative molecule for benzene derivatives. The compound with an OH or NH$_2$ substituent, e.g. p-cresol or p-toluidine, gives a small CCF value. Such compounds could be easily differentiated from benzene derivatives which contain only alkyl groups. The effect of substituent groups is further discussed in later sections.

3.3. Probability density function

In order to differentiate anthracene derivatives from non-anthracene compounds, the degree of similarity was evaluated by calculating the CCF values for various aromatic hydrocar-

bons using anthracene as a standard sample. The result is shown in column 2 of Table 4. It is apparent that compounds with an anthracene ring have larger CCF values. Fig. 2 shows the probability density of anthracene derivatives plotted against the CCF value. The solid curve was obtained by fitting the data to the following equation:

Table 3
CCF values calculated for various aromatic hydrocarbons using toluene as a standard sample

Compound	$r = 1$	$r = 10$	Ref.
Skeletal molecules			
Benzene	0.002	0.043	[12]
Indole	10^{-4}	0.158	[13]
Naphthalene	0.010	0.133	[14]
Fluorene	10^{-6}	0.026	[15]
Anthracene	0.002	0.024	[6]
Pyrene	0.015	0.040	[16]
Naphthalene derivatives			
2-Methylnaphthalene	0.021	0.034	[14]
Naphthalene	0.010	0.133	[14]
1-Methylnaphthalene	10^{-4}	0.008	[14]
2-Ethylnaphthalene	10^{-4}	0.062	[14]
1-Naphthol	10^{-6}	0.023	[17]
Benzene derivatives			
Benzene	0.002	0.043	[12]
Ethylbenzene	0.721	0.806	[18]
n-Butylbenzene	0.561	0.711	[18]
n-Propylbenzene	0.544	0.640	[18]
n-Hexylbenzene	0.516	0.737	[18]
tert-Butylbenzene	0.474	0.601	[18]
n-Pentylbenzene	0.423	0.482	[18]
Isopropylbenzene	0.246	0.447	[18]
1-Phenyl-1-butyne	0.124	0.265	[20]
p-Aminophenol	0.040	0.100	[19]
p-Difluorobenzene	0.041	0.012	[20]
1-Phenylpropyne	0.013	0.324	[20]
p-Cresol	10^{-6}	0.000	[19]
p-Toluidine	10^{-7}	0.067	[19]
p-Fluorophenol	10^{-15}	0.001	[19]

r: cm^{-1}.

Table 4
CCF values and similarities calculated for various aromatic hydrocarbons using anthracene as a standard sample.

Compound	CCF	Similarity (%)	Ref.
9-Hexylanthracene	0.606	100.0	[21]
9-Methylanthracene	0.600	100.0	[21]
2-Chloroanthracene	0.406	99.9	[6]
2-Phenylanthracene	0.270	96.0	[22]
9-Cyanoanthracene	0.267	95.7	[23]
9-Bromoanthracene	0.220	88.3	[23]
1-Chloroanthracene	0.181	76.7	[6]
2-Methylnaphthalene	0.174	73.9	[14]
9,10-Dibromoanthracene	0.122	48.3	[23]
9-Phenylanthracene	0.095	33.0	[22]
trans-2-Naphthol	0.089	17.9	[17]
2-Methylanthracene	0.062	15.7	[6]
9-Methoxyanthracene	0.051	10.9	[24]
n-Propylbenzene	0.050	10.5	[18]
9-Chloroanthracene	0.042	7.1	[6]
Tryptamine	0.038	6.2	[24]
Tryptophan	0.035	5.2	[25]
7-Azaindole	0.032	4.4	[26]
3-Methylindole	0.032	4.4	[13]
n-Hexylbenzene	0.029	3.6	[18]
9,10-Dichloroanthracene	0.025	2.7	[23]
Toluene	0.024	2.5	[18]
1-Naphthol	0.023	2.3	[17]
cis-2-Naphthol	0.023	2.3	[17]
3-Indole acetic acid	0.010	0.4	[27]
2-Ethylanthracene	0.009	0.3	[6]
1-Phenylpropyne	0.005	0.1	[20]
1-Phenyl-1-butyne	0.002	0.0	[20]
1-Methylnaphthalene	0.002	0.00	[14]
Phenylethyne	0.001	0.0	[20]
2-Indole propionic acid	0.000	0.0	[27]
1-phenyl-1-hexyne	0.000	0.0	[20]

The linewidth, r, is assumed to be $10 \, \text{cm}^{-1}$.

$$P = 1 - \exp(-C^2/k^2) \qquad (5)$$

where P is the probability of an anthracene derivative, C is the CCF value, and k is the constant obtained by fitting the data to the above equation. This curve is useful for quantifying the "similarity" of the spectrum by calculating the probability from the CCF value using Eq. (5). This "similarity" is listed in the third column of Table 4. The value for 9-hexylanthracene or 9-methylanthracene is 100%, indicating a high spectral similarity, and implying a high probability to be assigned to an anthracene derivative. It should be noted that the curve in Fig. 2 is saturated at high CCF values. In order to differentiate between compounds with closely related structures, the linewidth can be reduced, e.g. to $\approx 1 \, \text{cm}^{-1}$. It is interesting to note that the first seven compounds are anthracene derivatives in Table 4, but the similarity of other anthracene deriva-

tives such as 9,10-dichloroanthracene and 2-ethylanthracene are much lower than those for non-anthracene derivatives. Thus the reliability given by the present approach is not necessarily perfect.

In order to improve the accuracy of prediction, it is necessary to use an additional, simultaneous parameter, different from the CCF value, in order to discriminate against compounds having different spectral properties.

3.4. Wavelength of the 0–0 transition

In the previous section, only the vibrational structure of the spectrum is used for prediction of the chemical structure after subtraction of the frequency of the 0–0 transition. However, this frequency contains important information about the extension of π electrons in a molecule, which is related to the ring size. The second column of Table 5 shows the wavelengths of the 0–0 transition for compounds whose excitation spectra have been previously reported. The 14 anthracene derivatives are located between 364 and 387 nm, although two other compounds (pyrene and pyrazine) are also located in this spectral region. This implies that the wavelength of the 0–0 transition can be used to advantage for differentiation of anthracene derivatives. The probability density of the anthracene derivative is plotted against the wavelength of the 0–0 transition in Fig. 3. The probability density increases rapidly at shorter wavelengths and then decays gradually
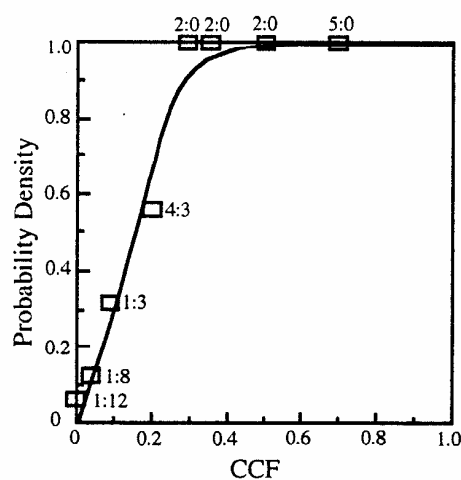

Fig. 2. Probability density of anthracene derivatives versus CCF value. The number in the Figure indicates the number of anthracene (left) and non-anthracene (right) compounds. The solid curve is calculated from Eq. (5) by fitting the experimental data to $k = 0.2$.

Table 5
Wavelengths of the 0–0 transitions for various aromatic hydrocarbons and probabilities of anthracene derivatives

Compound	0–0 transition (nm)	Probability (%)	Ref.
n-Propylbenzene	266.08	0.0	[18]
n-Hexylbenzene	266.15	0.0	[18]
Toluene	266.83	0.0	[18]
Phenylethyne	274.94	0.0	[20]
1-Phenylpropyne	275.96	0.0	[20]
1-Phenyl-1-butyne	276.49	0.0	[20]
1-Phenyl-1-hexyne	276.54	0.0	[20]
3-Indole acetic acid	285.40	0.0	[27]
3-Indole propionic acid	285.88	0.0	[27]
Tryptamine	286.39	0.0	[24]
3-Methylindole	286.68	0.0	[13]
Tryptophan	286.75	0.0	[25]
7-Azaindole	288.73	0.0	[26]
Naphthalene	312.30	0.0	[14]
1-Methylnaphthalene	314.73	0.0	[14]
2-Methylnaphthalene	315.40	0.0	[14]
1-Naphthol	317.90	0.0	[17]
trans-2-Naphthol	323.57	0.0	[17]
cis-2-Naphthol	326.94	0.0	[17]
2-Ethylanthracene	364.54	61.0	[6]
2-Methylanthracene	365.07	60.5	[6]
2-Chloroanthracene	367.09	59.5	[6]
1-Chloroanthracene	367.31	59.0	[6]
Pyrene	367.44	58.8	[16]
2-Phenylanthracene	370.80	52.0	[22]
9-Phenylanthracene	371.14	51.2	[22]
9-Methylanthracene	371.15	51.0	[21]
Pyrazine	372.80	49.5	[28]
9-Hexylanthracene	373.03	48.2	[21]
9-Chloroanthracene	373.25	47.2	[6]
9-Methoxyanthracene	373.40	47.0	[24]
9-Bromoanthracene	374.20	46.5	[23]
9-Cyanoanthracene	382.10	32.4	[23]
9,10-Dichloroanthracene	385.35	30.5	[23]
9,10-Dibromoanthracene	386.45	30.0	[23]
Benzo[a]pyrene	395.90	18.5	[29]
Fluoranthene	396.56	18.0	[30]
Perylene	415.45	7.5	[31]
Tetracene	446.37	1.0	[32]
Ovalene	466.22	0.0	[33]
Pentacene	536.90	0.0	[34]

at longer wavelengths. These data were fitted to the chi-square distribution function:

$$F_m(x) = [(1/2)\Gamma(m/2)](x/2)^{m/2-1} \exp(-x/2) \quad (6)$$

where $\Gamma(m/2)$ is the gamma function, m is the degree of freedom, and x is the wavelength. The probability to be assigned to an anthracene derivative is listed in column 3 of Table 5. The probabilities for pyrene and pyrazine are rather high, indicating the limitation of this method alone.

3.5. Product of probabilities

In order to improve the accuracy of prediction, the probabilities obtained independently (third columns of Tables 4 and 5) may be multiplied. The result is shown in Table 6. It is noteworthy that the first 14 compounds are anthracene derivatives and the probabilities for non-anthracene derivatives are completely zero. Thus, the discrimination of anthracene derivatives from other compounds is perfect for all

the compounds whose spectra are so far available in the reference literature. In this study, anthracene was used simply as a representative molecule of anthracene derivtives in calculation of the CCF values. More accurately, the calculation must be performed using the "averaged spectrum" of anthracene derivatives, which is obtained by accumulating all the spectra for anthracene derivatives. This procedure may further improve the accuracy of prediction.

3.6. Functional group

For complete knowledge of the chemical structure, the functional group and its position must also be predicted properly. Table 7 shows the CCF values calculated for various benzene derivatives using toluene as a standard sample. As expected, alkylbenzenes, especially n-alkylbenzenes, give large CCF values. Fig. 4 shows the probability of alkylbenzene derivatives plotted against the CCF value. Fig. 5 shows the probability density for alkylbenzenes calculated from the frequency shifts of the 0–0 transitions. The typical shifts by substitution with an alkyl group are $500-1000 \text{ cm}^{-1}$. The product of the probabilities is shown for benzene derivatives in column 6 of Table 7. Apparently, ethylbenzene gives the largest value among these compounds, indicating the distinct merit of the present pattern recognition.

A similar comparison was also performed using 9-methylanthracene as a standard sam-

Table 6
Products of probabilities of anthracene derivatives obtained from CCF values and wavelengths of the 0–0 transition

| Compound	Product of probabilities (%)
2-Chloroanthracene	59.4
9-Mthylanthracene	51.0
2-Phenylanthracene	49.9
9-Hexylanthracene	48.2
1-Chloroanthracene	45.3
9-Bromoanthracene	41.0
9-Cyanoanthracene	31.0
9-Phenylanthracene	16.8
2-Methylanthracene	9.5
9-Methoxyanthracene	5.1
9-Chloroanthracene	1.4
9,10-Dichloroanthracene	0.9
9,10-Dibromoanthracene	0.8
2-Ethylanthracene	0.2
2-Methylnaphthalene	0.0
n-Propylbenzene	0.0
$trans$-2-Naphthol	0.0
Tryptamine	0.0
Tryptophan	0.0
7-Azaindole	0.0
3-Methylindole	0.0
n-Hexylbenzene	0.0
Toluene	0.0
1-Naphthol	0.0
cis-2-Naphthol	0.0
3-Indole acetic acid	0.0
1-Phenylpropyne	0.0
1-Phenyl-1-butyne	0.0
1-Methylnaphthalene	0.0
Phenylethyne	0.0
3-Indole propionic acid	0.0
1-Phenyl-1-hexyne	0.0
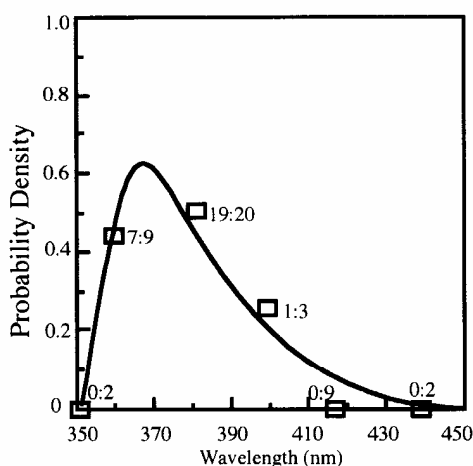
Fig. 3. Probability density of anthracene derivative against the wavelength of the 0–0 transition. The number in the figure indicates the number of anthracene (left) and non-anthracene (right) compounds. The solid curve is calculated from Eq. (6) by fitting the experimental data to $m = 4$.

ple. As expected, 9-alkylanthracene, i.e. 9-hexylanthracene, gives the largest CCF value, implying the presence of an alkyl group at the 9-position. It is interesting to note that the CCF values for anthracene derivatives substituted at the 2-position with an alkyl group are rather small. This is ascribed to vibrational and rotational motions of the methyl group, providing many split lines due to a shallow potential curve.

Owing to a lack of spectral data, a more accurate discussion is difficult. However, it is generally true that the functional group and its position can be predicted fairly well as the probability is always high when the CCF value is calculated using the standard sample with a closely related chemical structure.

Table 7

Probabilities calculated from CCF values using toluene as a standard sample and from frequency shifts of the 0–0 transition from benzene

Benzene derivative	CCF	PC (%)	FS (cm^{-1})	PFS (%)	P (%)
Ethylbenzene	0.806	98.3	501.1	73.3	72.0
n-Hexylbenzene	0.737	96.6	516.6	73.5	71.0
n-butylbenzene	0.711	95.8	512.4	73.4	70.3
n-Propylbenzene	0.640	92.3	506.8	73.3	67.6
tert-Butylbenzene	0.601	89.5	393.4	72.2	64.6
n-Pentylbenzene	0.482	76.6	515.2	73.5	56.3
Isopropylbenzene	0.447	71.3	430.3	72.8	51.9
p-Aminophenol	0.100	6.1	6690.3	0.0	0.0
p-Toluidine	0.67	2.8	4993.3	0.0	0.0
1-Naphthol	0.023	0.3	6633.0	0.0	0.0
trans-2-Naphthol	0.018	0.2	7184.2	0.0	0.0
cis-2-Naphthol	0.010	0.1	1722.9	21.7	0.0
p-Fluorophenol	0.001	0.0	2963.5	3.1	0.0

The linewidth, r, is assumed to be 10 cm^{-1}.

CCF, cross correlation factor; PC, probability from CCF (%); FS, frequency shift (cm^{-1}); PFS, probability from frequency shift (%); P, product of probabilities (%).
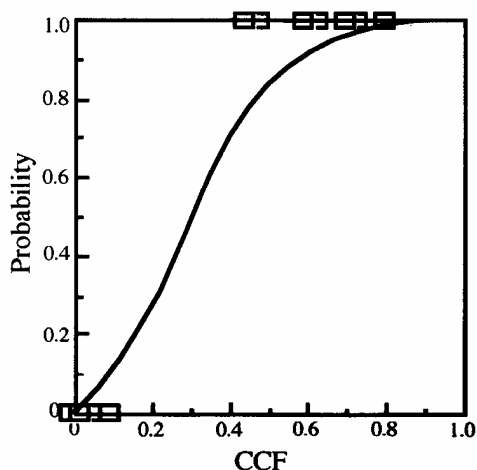

Fig. 4. Probability of alkylbenzenes against CCF value. Owing to the small number of compounds available, the data (0 or 1) are directly plotted in the Figure. The solid curve is calculated from Eq. (5) by fitting the experimental data to $k = 0.4$.
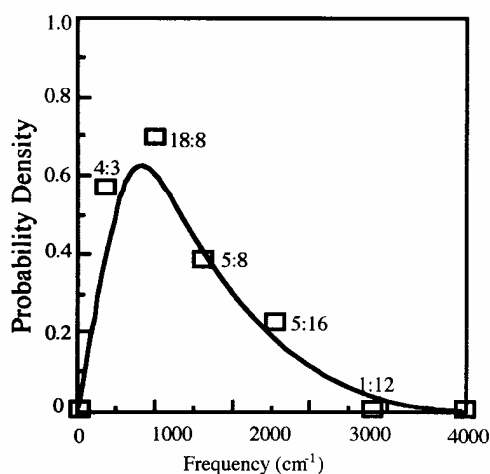

Fig. 5. Probability density of alkylbenzene against the frequency shift of the 0–0 transition from that of benzene. The solid curve is calculated from Eq. (6) by fitting the experimental data to $m = 4$.

4. Conclusion

A supersonic jet spectrum contains information concerned with skeletal vibrations, so that the skeletal structure can be well-predicted by pattern recognition. The wavelength of the 0–0 transition also contains information on the chemical structure. Thus the accuracy in prediction is improved by multiplying the above two probabilities obtained independently. Additional data concerned with fluorescence and multiphoton ionization spectra might also be useful for this purpose. Pattern recognition

may be applied to analysis of the mass fragment spectrum for more accurate prediction of the chemical structure.

In this study, the determination of functional group position must be regarded as preliminary, since the number of supersonic jet spectra reported is small at the present time. By increasing the number of accumulated data, it will become possible to discuss the effect of other substitutional groups, such as OH or NH_2. Their CCF values are quite small when an alkylbenzene is used as a standard sample, suggesting that it may be possible to discrimi-

nate such compounds from other aromatic hydrocarbons, especially alkylbenzene.

For more accurate prediction of chemical structure, it is necessary to accumulate a large number of spectra in the database, as described. For this purpose, it is necessary to develop an instrument which allows measurements of the supersonic jet spectrum in a short time period. For example, the sample should be replaced immediately without changing the conditions of the nozzle and the vacuum system [10]. Further, a widely-tunable laser source may be essential, because the tunable range of the dye laser currently used is very limited and is not suitable for application to unknown samples. A titanium-doped sapphire laser or an optical parameteric oscillator coupled with a frequency conversion system is quite attractive. A multi-frequency laser may be also used to reduce the time for recording a spectrum [11]. Such hardware developments are essential not only for construction of the database, but also for practical use of supersonic jet spectrometry in routine work.

References

[1] J.M. Hayes and G.J. Small, Anal. Chem., 55 (1983) 565 A.
[2] M.V. Johnston, Trends Anal. Chem., 3 (1984) 58.
[3] D.M. Lubman, Anal. Chem., 59 (1987) 31A.
[4] T. Imasaka and N. Ishibashi, Prog. Quantum Electron., 14 (1990) 131.
[5] T. Imasaka and N. Ishibashi, Spectrochim. Acta., Part B, 43 (1988) 661.
[6] C.H. Lin, H. Fukii, T. Imasaka and N. Ishibashi, Anal. Chem., 63 (1991) 1433.
[7] T. Imasaka and N. Ishibashi, Report For Grant-in-Aid for Scientific Research from the Ministry of Education of Japan, Project No. 01470065, 1991.
[8] T. Imasaka, K. Sakaki and N. Ishibashi, Chemom. Intell. Lab. Syst., 6 (1989) 281.
[9] T. Imasaka, K. Sakaki and N. Ishibashi, Spectrochim. Acta, Part B, 43 (1988) 703.
[10] T. Imasaka, T. Okamura and N. Ishibashi, Anal. Chem., 58 (1986) 2152.
[11] M. Nishida, C.H. Lin and T. Imasaka, Anal. Chem., 65 (1993) 3326.
[12] T.A. Stephenson, P.L. Radloff and S.A. Rice, J. Chem. Phys., 81 (1984) 1060.
[13] R. Bersohn, U. Even and J. Jorther, J. Chem. Phys., 80 (1984) 1050.
[14] S.M. Beck, D.E. Powers, J.B. Hopkins and R.E. Smalley, J. Chem. Phys., 73 (1980) 2019.
[15] A. Amirav, U. Even and J. Jortner, J. Chem. Phys., 67 (1982) 1.
[16] N. Ohta and H. Baba, Chem. Phys. Lett., 133 (1987) 222.
[17] A. Oikawa, H. Abe, N. Mikami and M. Ito, J. Phys. Chem., 88 (1984) 5180.
[18] J.B. Hopkins, D.K. Powers and R.E. Smalley, J. Chem. Phys., 72 (1980) 5039.
[19] R. Tembruell, T.M. Dunn and D.M, Lubman, Spectrochim. Acta, Part A, 42 (1986) 899.
[20] A.E.W. Knight and S.H. Kable, J. Chem. Phys., 89 (1988) 7139.
[21] J.A. Syage, P.M. Felker, D.H. Semmes, F.A. Adel and A.H. Zewail, J. Chem. Phys., 82 (1985) 2896.
[22] D.W. Werst, A.M. Brearley, W.R. Gentry and P.F. Parbara, J. Am. Chem. Soc., 109 (1987) 32.
[23] A. Amirav, C. Horwitz and J. Jortner, J. Chem. Phys., 88 (1988) 3092.
[24] S. Hirayama and F. Tanaka, Chem. Phys. Lett., 153 (1988) 112.
[25] T.R. Rizzo, Y.D. Park, L.A. Peteanu and D.H. Levy, J. Chem. Phys., 84 (1986) 2534.
[26] K. Fuke, H. Yoshiuchi and K. Kaya, J. Phys. Chem., 88 (1984) 5840.
[27] Y.D. Park, T.R. Rizzo, L.A. Peteanu and D.H. Levy, J. Chem. Phys., 84 (1986) 6539.
[28] J.L. Tomer, K.W. Holtzclaw and D.W. Pratt, J. Chem. Phys., 88 (1988) 1528.
[29] G.D. Greenblatt, E. Nissani, E. Zaroua and H. Yehuda, J. Phys. Chem., 91 (1987) 570.
[30] I.Y. Chan and M. Dantus, J. Chem. Phys., 82 (1985) 4771.
[31] M. Sonnenschein, A. Amirav and J. Jortner, J. Phys. Chem., 88 (1984) 4214.
[32] A. Amirav, U. Even and J. Jortner, J. Chem. Phys., 75 (1981) 3770.
[33] A. Amirav, U. Even and J. Jortner, Chem. Phys. Lett., 69 (1980) 14.
[34] A. Amirav, C. Horwitz and J. Jortner, Chem. Phys. Lett., 72 (1980) 21.