相関係数法による赤外スペクトル検索の研究

田村禎夫[®]， 田辺和俊， 平石次郎， 佐伯慎之助*

(1979 年 9 月 17 日受理)

相関係数法を用いた赤外スペクトルの検索について、455 種の液体・固体試料のスペクトルをモデルとして研究した。試料調製の方法の相関係数への影響及び適切な波数領域について検討した。検索に要する時間を短縮することを目的として、最強吸収帯の波数、1 次から 4 次までのスペクトルのモーメントをふるいとして用いたところ、時間を短縮できるとともに、検索の効率も向上し、赤外スペクトルのみでは、本質的に識別できないような化合物を別とすれば、すべての化合物を識別しうる可能性のあることが分かった。

1 緒 言

多くの物質の赤外スペクトルデータを電子計算機に格納し、それを用いて未知試料の同定を行うという、いわゆる赤外スペクトル検索については、これまでに多くの研究が行われている¹⁾。最も広く行われているのは ASTM データファイル²⁾を利用した検索システムであり、著者らも最近、未知試料の吸収帯の相対強度を考慮に入れた検索システムを作成した³⁾。ASTM データファイルは、基本的には、各吸収帯の位置（波長）を入力したものである。それを利用した検索システムでは、検索されて出てきたスペクトル番号をキーとして、そのスペクトルをスペクトルチャート集から探し出し、未知試料のスペクトルと比較して確認するという操作を行う必要がある。もし、スペクトル全体が数値化されて電子計算機に格納されているならば、この最終的な確認が容易であるとともに、ASTM ファイルを利用する場合と異なった論理で検索を行い、より良い検索方法が見いだされる可能性がある。このようなスペクトルのパターンを用いた検索方法として、田辺、佐伯は、相関係数法が有効であることを 110 種の液体試料のスペクトルデータを用いて示した⁴⁾。著者らは、相関係数法に先立って、スペクトルパターンの特徴を表すものをふるいとして用い

ることにより、検索に要する時間を短縮し、更に性能を向上させることができると考え、そのふるいとして最強吸収帯の振動数、及び 1 次から 4 次までのスペクトルのモーメントについて研究してきた⁵⁾⁶⁾。

今回、著者らは 455 種の液体、固体試料のデータを用いて、相関係数法に関する問題点を再検討し、更に上記のふるいを相関係数とともに用いて検索のテストを行った結果、赤外スペクトルでは、本質的に識別することができないような場合（例えば、メチレン鎖の長さだけが異なる化合物）は別として、すべてを識別できる可能性のあることが明らかとなったので、それについて報告する。

2 スペクトルデータの集積

モデルとして用いたスペクトルデータファイルは、既報⁶⁾において、スペクトルのモーメントについての研究を行った際に用いたものと同じであり、IRDC カードの作成に用いたスペクトルチャートから 455 枚を選び出し、チャートリーダーによって、それらのスペクトルを数値化したものである。化合物の選び方及び、チャートリーダーで数値化した場合の問題点については既報⁶⁾で述べたので省略する。

3 相関係数による検索

化合物 A, B のスペクトル間の相関係数 R は

* 工業技術院化学技術研究所：茨城県筑波郡谷田部町 東 1-1

$$R = \frac{\sum_{i=1}^n I_A(\nu_i) I_B(\nu_i) - (1/n) \{ \sum_{i=1}^n I_A(\nu_i) \} \{ \sum_{i=1}^n I_B(\nu_i) \}}{[\{ \sum_{i=1}^n I_A(\nu_i)^2 - (1/n) (\sum_{i=1}^n I_A(\nu_i))^2 \} \{ \sum_{i=1}^n I_B(\nu_i)^2 - (1/n) (\sum_{i=1}^n I_B(\nu_i))^2 \}]^{1/2}}$$

で与えられる。ここで $I_A(\nu_i)$ は化合物Aのスペクトルの波数 ν_i における吸光度である。上の式から明らかなように、すべての点で $I_A(\nu_i)/a + I_B(\nu_i)/b = 1$ の関係のあるとき (a, b は定数), $ab < 0$ ならば $R=1$, $ab > 0$ ならば $R=-1$ となる。AとBのスペクトルに類似性が少ないときには R は 0 に近い値となる。相関係数を計算する際のサンプリングの間隔は 10 cm^{-1} 程度までは、 R の値に影響のないことが液体試料で分かっている⁴⁾。本報告で扱っている試料には、固体試料が含まれており、固体試料の場合には液体試料の場合よりも吸収帯の半値幅が小さい傾向があるので⁵⁾, 5 cm^{-1} 間隔でサンプリングを行った。実際に幾つかの試料について波数間隔を変えて計算し、 5 cm^{-1} 間隔で十分であることを確認した。

3.1 試料調製の相関係数への影響

相関係数は、定義から明らかなように、試料の濃度に依存しない量である。しかしながら固体試料の場合には、試料調製の仕方によっては同一試料のスペクトル間の相関係数が1とはならない可能性がある。この試料調製の影響を調べるために、安息香酸、アセトアニリド及びイミダゾールについて、打錠の条件や濃度を変えてそれぞれ3種、合計9種の錠剤を作製し^{*}, YHP 2112A ミニコンピュータにオンラインで接続した⁷⁾ パーキンエルマー180型赤外分光光度計で測定した。

($1800 \sim 600 \text{ cm}^{-1}$) の領域について、同一試料間及び異種試料間の相関係数を計算した結果を Table 1 に示す。

Table 1 The value of the correlation coefficient

Between same substance		
Benzoic acid	I	II
II	0.967	
III	0.992	0.945
Acetanilide	IV	V
V	0.980	
VI	0.972	0.984
Imidazole	VII	VIII
VIII	0.969	
IX	0.948	0.953
Between different substances		
	III	IV
IV	0.064	
IX	-0.072	0.199

* 既報⁶⁾で、モーメントへの影響を検討した際に用いたものと同一である。

この結果から、相関係数を用いて検索するときには、およそ 0.95 以上の値を示すものは正解の可能性があると考えなければ

ならないことが分かる。液体試料の場合の同一試料間の相関係数は既報に示したように 0.997 以上となったが⁴⁾, これが固体試料の場合に 0.95 となることは、臭化カリウム錠剤の作製の仕方の影響が大きいことを示している。

3.2 波数領域と相関係数

上記の 455 種のスペクトル間の相関係数 $455 \times 454/2 = 103285$ 個を計算し、相関係数が 0.95 あるいは 0.90 以上となる化合物の組み合わせの数を求めた結果を Table 2 に示す。波数領域としては ($1800 \sim 600 \text{ cm}^{-1}$) と ($1400 \sim 800 \text{ cm}^{-1}$) の2種類を選んだ。その結果、狭いほうの ($1400 \sim 800 \text{ cm}^{-1}$) の波数領域のほうがむしろ適切であることが分かった。この波数領域で R が 0.95 以上となる組み合わせの数は 51 個であり、異なった物質間の R が 0.95 以上となる確率は $51/103285 \approx 5 \times 10^{-4}$ である。455 個の化合物の一つ一つを未知試料と考えて検索した場合に、自分自身以外にノイズとして出てくる化合物の数を調べてみると Table 3 のようになる。これは 422 個の化合物はノイズとなるものが無いが、その他の 33 個の化合物は (1~5) 個のノイズとなる化合物があることを示している。波数領域として指紋領域全体を取るよりも狭くしたほうが良いという結果は、これまでの研究の結果⁴⁾ と一致しているが、最適の波数領域を決めるには、化合物の数を更に多くして検討しなければならない。Table 2 には参考のために相関係数の値が 0.90 以

Table 2 Screening efficiency of correlation coefficient R , strongest band frequency ν_A , moments of spectrum $M_1 \sim M_4$, and their combinations. Table shows the number of hit couples†. Two searches have been tried setting the lower limit of R allowable for hit couples at 0.90 and 0.95, respectively

Level of R		Screening criteria			
		R	$\nu_{\Delta} \& R$	$M_1 \sim M_4 \& R$	$\nu_{\Delta}, M_1 \sim M_4 \& R$
(a)	$R > 0.90$	234	194	65	52
	$R > 0.95$	79	73	26	23
(b)	$R > 0.90$	156	121	38	31
	$R > 0.95$	51	43	16	12

Frequency range : (a) ($1800 \sim 600 \text{ cm}^{-1}$), (b) ($1400 \sim 800 \text{ cm}^{-1}$);

† The search has been made for the combination between 455 compounds. For example, 234 means that 234 combinations among $455 \times 454/2 \approx 10^5$ give R greater than 0.90.

Table 3 Distribution of 51 noises in Table 2

Noise	Number of compounds
0	422
1	22
2	7
3	2
4	1
5	1

上となる組み合わせの数も示した。

3.3 最強吸収帯, モーメントの効果

相関係数 R とともに最強吸収帯の波数 ν_A と 1 次から 4 次までのスペクトルのモーメント (M_1, M_2, M_3, M_4) を用いて検索した結果を同じく Table 2 に示す。当然のことながら, $\nu_A, M_1 \sim M_4$ を併用することによってノイズの数は減少するが, ν_A の効果は, $M_1 \sim M_4$ に比べて明らかに小さい。このことは R と ν_A との相関が, R と $M_1 \sim M_4$ とのそれよりも大きいことを示している。(1400~800) cm^{-1} の領域で R が 0.95 以上となる組み合わせの数は ν_A と $M_1 \sim M_4$ を併用すると 12 個である。

4 考 察

4.1 最強吸収帯の分布と波数領域

本研究で用いた 455 種の化合物の最強吸収帯の波数 ν_A の分布を Fig. 1 に示す。 ν_A の選択の方法については, 既に報告したので説明を省略する⁵⁾。この分布を見ると 700 cm^{-1} , 1460 cm^{-1} , 1700 cm^{-1} 付近に大きな山があることが分かる。相関係数 R と ν_A との間にある程度の相関があることが Table 2 から分かるが, これは強い吸収帯のある部分の相関が, 全体の R に大きく寄与することを考えれば当然の結果である。相関係数法によって検索する場合に波数領域が (1800~600) cm^{-1} よりも (1400~800) cm^{-1} のほうが効率が良いということは, 後者の場合には上記の ν_A の分布の三つの大きな山を含んでいない領域であることを考えると妥当な結果であると考えられる。

Fig. 1 Distribution of the strongest absorption band, ν_A , of 455 spectra treated in this work

4.2 ノイズとなる化合物

最強吸収帯の波数 ν_A , 1 次から 4 次までの四つのモーメント ($M_1 \sim M_4$) でふるいをかけ, その後に相関係数が 0.95 以上となる化合物の組み合わせが約 10 万のうち 12 個残るが, これはどのような化合物あるいはスペクトルの組み合わせであろうか。Table 4 にその組み合わせを示した。これをみると, その大部分はメチレン基の数のみが異なる化合物の組み合わせである。実際にスペクトルを比較しても非常に良く似ていることが分かり, これらの化合物は, 本質的に赤外スペクトルだけでは区別できないものであると考えられる。Table 4 (10) の化合物の組は, メチレン鎖の数の違いのみとは言

Table 4 The combination of substances, the spectra of which are highly correlated with each other

R	IRDC No.	Molecular structure
(1) 0.992	17440	$\text{NH-CO-(CH}_2)_4\text{-CH}_3$ $\text{CH}_3\text{-(CH}_2)_3\text{-CH-CO-O-CH}_2\text{-CH-CH}_2\text{OH}$ OH
	17448	$\text{NH-CO-(CH}_2)_{10}\text{-CH}_3$ $\text{CH}_3\text{-(CH}_2)_3\text{-CH-CO-O-CH}_2\text{-CH-CH}_2\text{OH}$ OH
(2) 0.987	16855	$\text{CH}_3\text{-(CH}_2)_3\text{-O-}$ -CHO
	17184	$\text{CH}_3\text{-(CH}_2)_6\text{-O-}$ -CHO
(3) 0.985	16857	$\text{CH}_3\text{-(CH}_2)_4\text{-O-}$ -CHO
	17184	$\text{CH}_3\text{-(CH}_2)_6\text{-O-}$ -CHO
(4) 0.984	17443	$\text{OHCH}_2\text{-CH}_2\text{-O-CO-(CH}_2)_5\text{-NH-CO-(CH}_2)_6\text{-CH}_3$
	17447	$\text{OHCH}_2\text{-CH}_2\text{-O-CO-(CH}_2)_5\text{-NH-CO-(CH}_2)_{10}\text{-CH}_3$
(5) 0.981	17266	$\text{CH}_3\text{-(CH}_2)_8\text{-CH}_2\text{-SO}_3\text{Na}$
	17372	$\text{CH}_3\text{-(CH}_2)_8\text{-SO}_3\text{Na}$
(6) 0.976	17182	$\text{CH}_3\text{-(CH}_2)_5\text{-O-}$ -CHO
	17184	$\text{CH}_3\text{-(CH}_2)_6\text{-O-}$ -CHO
(7) 0.975	16855	$\text{CH}_3\text{-(CH}_2)_3\text{-O-}$ -CHO
	16857	$\text{CH}_3\text{-(CH}_2)_4\text{-O-}$ -CHO
(8) 0.975	16855	$\text{CH}_3\text{-(CH}_2)_3\text{-O-}$ -CHO
	17182	$\text{CH}_3\text{-(CH}_2)_5\text{-O-}$ -CHO
(9) 0.966	16125	$\text{CH}_3\text{-(CH}_2)_9\text{-}$ -
	16238	$\text{CH}_3\text{-(CH}_2)_6\text{-}$ -
(10) 0.963	17024	$\text{HO-(CH}_2)_2\text{-O-(CH}_2)_2\text{-OC}_2\text{H}_5$
	17370	$\text{HOH}_2\text{C-(CH}_2\text{-O-CH}_2)_3\text{-CH}_2\text{OH}$
(11) 0.958	16125	$\text{CH}_3\text{-(CH}_2)_9\text{-}$ -
	16240	$\text{CH}_3\text{-(CH}_2)_7\text{-}$ -
(12) 0.957	16857	$\text{CH}_3\text{-(CH}_2)_4\text{-O-}$ -CHO
	17182	$\text{CH}_3\text{-(CH}_2)_5\text{-O-}$ -CHO

えないが構造的に類似性が高く、スペクトル的にも似ている (Fig. 2)。(1300~1400) cm^{-1} の範囲に差が見られるが、他の領域では大きな違いは見られない。(1400~800) cm^{-1} を 100 cm^{-1} 間隔で区切り、それぞれの領域でのこの二つのスペクトルの R を計算してみると Fig. 2 に示したようになり、(1300~1400) cm^{-1} の領域のみ R が小さくなる。このことは R をある波数領域ごとに計算し、1箇所でも小さい R を示すところがあれば、ふるい落とすという方法により、更に検索の効率を上げることができる可能性を示している。

Fig. 2 Spectra and correlation coefficients, R' , in every 100 cm^{-1} range (By the courtesy of the Infrared Data Committee of Japan)

I : 2-(2-Ethoxyethoxy)ethanol, $\text{HO}-(\text{CH}_2)_2-\text{O}-(\text{CH}_2)_2-\text{O}-\text{C}_2\text{H}_5$; II : 3,6,9-Trioxaundecane-1,11-diol, $(\text{CH}_2-\text{O}-\text{CH}_2)_3-\text{CH}_2\text{OH}$, $\text{CH}_2-\text{O}-\text{CH}_2-\text{CH}_2\text{OH}$

そこで、同一化合物のスペクトル間の相関係数を 100 cm^{-1} 間隔ごとに計算してみた。その結果が Table 5 である。ここでスペクトルの番号 I~IX は Table 1 の場合と同じである。この結果をみるとアセトアニリドの (1200~1100) cm^{-1} 及び (900~800) cm^{-1} の領域を除いては、高い相関を示すことが分かる。これはアセトアニリドではこの 2 箇所の波数領域には他の領域に比べて非常に弱い吸収帯があるのみで、試料調製の違いによる差が大きく R に寄与しているためであると考えられる。

Fig. 2 及び Table 5 の結果を考えると、未知試料のスペクトルと高い相関を示すスペクトルがあった場合には、更に一定波数領域ごとに区切って、それぞれの領域の R を計算し、次のような方法によって区別できることを示している。すなわち、すべての領域で高い相関を示す場合には残し、ある領域で相関が小さくなった場合には、その領域に「ある程度以上の強度の吸収帯」があるときには捨て、無いときには残すという論理が有効である可能性がある。この点については、Fig. 2 と同様な場

Table 5 Values of correlation coefficients in every 100 cm^{-1} range

Benzoic acid		Acetanilide		Imidazole	
I, II		IV, V		VII, VIII	
1400~1300 cm^{-1}	0.991		0.996		0.990
1300~1200 "	0.975		0.996		0.996
1200~1100 "	0.991		0.881		0.978
1100~1000 "	0.984		0.960		0.999
1000~900 "	0.988		0.972		0.998
900~800 "	0.946		0.651		0.992
I, III		IV, VI		VII, IX	
1400~1300 cm^{-1}	0.995		0.995		0.977
1300~1200 "	0.995		0.991		0.996
1200~1100 "	0.987		0.643		0.976
1100~1000 "	0.963		0.993		0.989
1000~900 "	0.996		0.964		0.986
900~800 "	0.985		0.862		0.976
II, III		V, VI		VIII, IX	
1400~1300 cm^{-1}	0.987		0.999		0.941
1300~1200 "	0.954		0.994		0.987
1200~1100 "	0.967		0.750		0.919
1100~1000 "	0.934		0.947		0.985
1000~900 "	0.973		0.951		0.977
900~800 "	0.880		0.660		0.951

合について更に検討を加える必要があろう。

5 結 論

455 種の化合物の赤外スペクトルをモデルとして、相関係数、最強吸収帯、スペクトルのモーメントを用いた検索の方法について検討した結果、赤外スペクトルでは本質的に区別しえないような場合を除いては、良い検索結果を与えることが分かった。Fig. 1 の最強吸収帯の分布をみると、先に報告した 1333 種の化合物の最強吸収帯の分布⁵⁾ とほぼ同様であるが、1460 cm^{-1} 付近に山があることだけが顕著な差になっている。これは、Table 4 のようにメチレン鎖を含む化合物が多いことと対応していると考えられ、その点ではスペクトル的に類似している化合物が割合に多いスペクトル群を取り扱っていると考えられる。従って、更に多くの化合物の集団に対してこの検索方法を適用しても、本研究の結果と少なくとも同程度の結果が得られることが期待される。

本研究に際し、スペクトルチャートの使用と、その一部の掲載を許可していただいた日本赤外データ委員会に感謝の意を表する。

文 献

- 1) 田辺和俊：化学の領域，**28**, 33 (1974).
- 2) "Codes and Instruction for Wyandott-ASTM Punched Cards", ASTM (1964).
- 3) K. Tanabe, T. Tamura, J. Hiraishi, S. Saeki : *Anal. Chim. Acta*, in press.
- 4) K. Tanabe, S. Saeki : *Anal. Chem.*, **47**, 118

- (1975).
5) 田村禎夫, 平石次郎, 田辺和俊, 佐伯慎之助: 分
化, **28**, 31 (1979).
6) 田村禎夫, 田辺和俊, 平石次郎, 佐伯慎之助: 分
化, **28**, 591 (1979).
7) 平石次郎, 松村 知: 分光研究, **28**, 22 (1979).

☆

**Study on computer retrieval of infrared spectra
by a correlation coefficient method.** Tadao TAMURA,
Kazutoshi TANABE, Jiro HIRASHI and Shinnosuke
SAEKI (National Chemical Laboratory for Industry,
1-1, Higashi, Yatabe-machi, Tsukuba-gun, Ibaraki)

A correlation coefficient method is applied to the
computer retrieval of infrared spectra by using 455
digitized spectra of solid and liquid samples. The
effect of the sampling technique for solid samples
are studied for determining the tolerance of the corre-

lation coefficient, R , and it is concluded that the
correct spectrum may give the value of R greater than
0.95. The frequency range adequate for the method
is also discussed. The frequency of the strongest
absorption band and the 1st~4th moments of spectra
are used in order to save the computation time before
calculating R . It is suggested that the correlation co-
efficient method can identify all compounds except for
the compounds which are not distinguishable essentially
by only infrared spectrum.

(Received Sept. 17, 1979)

Keywords

Computer retrieval

Correlation coefficient

Infrared spectra

6-ブロモ-2-ベンゾチアゾリルアゾ-2-ナフトールを含むゲルカラム による微量カドミウムイオンの簡易分析法*

李 龍根**, 上野 景平***①

(1979 年 9 月 1 日受理)

架橋ポリスチレンビーズ {ジビニルベンゼン 2%, 乾燥時粒径 (70~100) mesh} を, 6-ブロモ-2-ベン
ゾチアゾリルアゾ-2-ナフトール (BTAN) の 0.02% クロロベンゼン溶液に 24 時間浸せきすると,
ポリスチレン粒子は膨潤ゲル化する. この紅色ゲル粒子を内径 2.5 mm, 長さ 120 mm のガラス管に
100 mm の長さに充てんして分析カラムを調製する. このカラムに pH 10 (Clark-Lubs 緩衝液) に
した試料溶液を 0.2 ml/min の流速で流すと試料中のカドミウムの量に応じてカラムの流入端から赤
紫色に変化する. 変色帯の長さはカラムに流入したカドミウムの全量に比例するので, 一定量の試料溶
液を流した後, 変色帯の長さを測定し, あらかじめ作成した検量線からカドミウムの濃度を求める. こ
の方法で (0.05~0.1) ppm では $\pm 10\%$, (0.2~0.5) ppm では $\pm 5\%$ の相対誤差でカドミウムを定
量することができる. アルミニウム, コバルト(II), 水銀(II), ニッケル(II), 鉄(III), 銀, クロム
(VI) などは 10 倍ないし数十倍まで, 銅(II) は 5 倍まで, 又亜鉛はカドミウムと同量まで, ジンコ
ン, タイロン, チオ尿素, フッ化ナトリウムなどでマスクングすることができる. 各種環境水, 排水などに
本法を適用して好結果を得た.

1 緒 言

著者らはジチゾンのクロロベンゼン溶液で架橋ポリス

チレン粒子を膨潤ゲル化させ, このようなゲル粒子を充
てんしたマイクロカラムを用い, ppm レベル以下の水銀及
び亜鉛イオンを半定量的に簡易分析する方法を先に報告
した¹⁾²⁾. 本簡易分析法は, 簡単な操作で環境水中の微量
の有害重金属イオンを半定量的に分析でき, しかもその
感度は試験紙法はもちろん, 従来の検知管方式より優れ
ている点で特長がある. 又, この分析法は原理的には有
機試薬を用いる金属イオンの溶媒抽出比色法に外ならな
いので, 有機試薬と抽出比色条件を適当に選べば水銀,

* 試薬ゲルカラムを用いる微量金属イオンの簡易半定
量分析法(第3報). 前報は Y. K. Lee, K. J. Whang,
K. Ueno: *Talanta*, **23**, 244 (1976)

** 延世大学校理化学学科: 大韓民国 Seoul 120

*** 九州大学工学部合成化学教室: 福岡県福岡市東区箱
崎 6-10-1