

赤外吸収, ^{13}C -核磁気共鳴, 質量スペクトルの 複合検索における二, 三の考察

田辺 和俊[®], 平石 次郎, 田村 禎夫,
山本 修, 柳沢 勝, 和佐田宣英*

(1983 年 12 月 22 日受理)

赤外吸収, ^{13}C -核磁気共鳴, 質量スペクトルの 3 種のスペクトルデータを同時に用いる複合検索において, 検索結果 (検索雑音の割合) に対し, 入力ピークの本数, 検索ファイル作成時の収録ピークの強度下限などが及ぼす影響について考察した. その結果, 3 種のスペクトルで各 2 本, 計 6 本のピークのデータを入力することにより, 化合物 1 万程度の大容量のファイル検索が可能であることが分かった.

1 緒 言

既報¹⁾においては赤外吸収 (IR), ^{13}C -核磁気共鳴 (^{13}C -NMR), 質量スペクトル (MS) の 3 種のデータを同時に用いる複合検索が 1 種のみデータを用いる単純検索よりどの程度有効であるかを数値的に明らかにするために研究した. 本報では複合検索の有効性を実証するために, 更に種々の点について考察を行った.

2 方 法

検索用ファイルの作成方法及び検索論理についてはほぼ既報に記したものと同一である. ただし既報では検索用ファイルを作成する際に収録すべきピークの強度下限に関しては最強ピークの 1/100 としたが, 本報ではこの強度下限の違いが検索結果に及ぼす影響を調べることにした.

又検索用入力データについても既報では最強ピーク 1 本のみとしたが, 本報ではこの入力ピーク本数の違いが与える影響も調べることにした.

3 結果と考察

3.1 収録ピークの強度下限の影響

既報と同様, 入力ピークを各 1 本 (最強ピークのみ) とした場合における収録ピーク強度下限の及ぼす影響を調べた結果を Table 1 に示す. Table 1 の数値はノイ

Table 1 Average distributions (%) of error answers to the search noise as function of the intensity level of peaks to be compiled in the search file

Search method	Intensity level					
	1	10	20	30	40	50
IR only	4.30	3.93	3.55	3.25	3.04	2.93
NMR only	4.25	4.22	4.16	4.07	3.98	3.98
MS only	1.60	1.48	1.47	1.53	1.58	1.67
IR+NMR	0.94	0.81	0.70	0.61	0.54	0.47
IR+MS	0.47	0.38	0.33	0.29	0.26	0.24
NMR+MS	0.39	0.28	0.24	0.22	0.21	0.19
All	0.17	0.12	0.10	0.08	0.07	0.06

ズ (この定義については既報参照) の分布割合であるが, 全体的に見て収録ピークの強度下限が高いほどノイズの割合が減少していることが分かる. このことは本研究における検索論理を用いる限りではあまり強度の低いピークまで収録することは検索結果においてノイズを増加させる原因となるものであり, 一方, ファイル容量節約, 検索速度向上の点からも最強ピークの 50% 程度以上の強いピークのみを収録して検索用ファイルを作成すればよいことが分かる.

収録ピークの強度下限を高くすれば収録ピークの本数は当然少なくなるが, その様子を Table 2 に示す. NMR では本数の減少は顕著ではないが, IR と MS では顕著である. このようなスペクトルによるピークの分布の違いは既報においても考察したとおりである.

* 工業技術院化学技術研究所 : 305 茨城県筑波郡谷田部町東 1-1

Table 2 Average numbers of peaks compiled in the search file as function of the intensity level

Intensity level	IR	NMR	MS
0	30.47	5.73	45.50
10	21.18	5.63	9.42
20	14.42	5.39	5.45
30	10.40	5.00	3.77
40	7.65	4.55	2.86
50	5.66	3.84	2.27

3.2 入力ピークの本数の影響

次に入力ピークの本数の違いが検索結果（ノイズの割合）に及ぼす影響を調べるために、前記の結果よりピーク収録強度下限を 50 とし、ピーク 1～3 本の場合について検討した。その結果を Table 3 に示す。

Table 3 Average distribution (%) of error answers to the search noise as function of the number of peaks entered as the input data

	Number of input peaks		
	1	2	3
IR only	2.928	0.728	0.317
NMR only	3.982	1.434	0.936
MS only	1.669	0.708	0.678
IR+NMR	0.466	0.071	0.035
IR+MS	0.236	0.047	0.025
NMR+MS	0.190	0.045	0.027
All	0.059	0.011	0.004

ピークの選び方としては強いほうから 1～3 本を選んだが、当然のことながら入力ピークの本数が多いほど検索条件が厳しくなるのでノイズが減少することは明らかである。

3.3 複合検索における大規模容量のファイル検索の可能性の推定

IR, NMR, MS の 3 種のスペクトルデータを用いる場合にどの程度の大きさまでのファイル検索が可能であるか、又そのときに必要最低限の入力データ個数はどの程度になるかを推定することが本研究の目的である。そこで Table 3 の結果からこの推定を行ってみると、例えばファイル容量 1 万件に対する検索を行った場合のノイズの数は入力ピーク本数が各 1 本のときに 5.9 個、各 2 本のときに 1.1 個、各 3 本のときに 0.4 個になると考えられる。

この結果から IR, NMR, MS の 3 種のスペクトルデータを用いる複合検索によれば化合物 1 万程度のファ

イル容量の検索も十分可能であり、IR, NMR, MS の 3 種のスペクトルに対し各 2 本ずつ、合計 6 本程度の少数のピークのデータを入力するだけでもノイズ数の平均が 1.1 個程度で検索できることが分かる。

3.4 複合検索の有効性

以上のように 3 種のスペクトルデータを組み合わせて検索に用いるとノイズが極めて少なくなることが分かったが、ピーク収録強度下限 50 及び入力ピーク各 2 本の条件の場合に、933 回のテスト検索におけるノイズ数を集計した結果が Table 4 である。全体の約 94% の場合においてノイズが零、すなわち正答の評点が最高になり、約 99.7% の場合においてノイズが 3 個以内、すなわち正答が第 4 位以上に出力されてくることが分かる。

Table 4 Numbers of search noises in the IR+NMR+MS combined search using input data of two peaks and 50% intensity levels

Number of search noises	Number of searches
0	878
1	33
2	11
3	8
4	2
5	0
6	0
7	1

4 検索実例

複合検索のプログラムを作成し検索を実行した例を

```

CS
*** WELCOME TO COMBINED SEARCH PROGRAM ***
CS>S IR 2928 10
      21 HITS (SET # 1)
CS>L
*NO SCORE(IR CMR MS) SDBS COMPOUND-NAME
  1 76 76 - - 1250 HEPTADECYLBENZENE
  2 73 73 - - 1365 1,2,3,4,5,6,7,8,9,
  3 72 72 - - 893 CYCLOHEXANE
  4 69 69 - - 1304 TRANS-DECALIN
(OMITTED)
  21 51 51 - - 280 CIS-2-OCTENE
CS>A CMR 27 100
      3 HITS (SET # 2)
CS>L
*NO SCORE(IR CMR MS) SDBS COMPOUND-NAME
  1 76 69 84 - 1304 TRANS-DECALIN
  2 61 72 50 - 897 CYCLOHEXANE
  3 58 59 58 - 647 2-METHYLPYPERIDINE
CS>A MS 56 100
      1 HITS (SET # 3)
CS>L
*NO SCORE(IR CMR MS) SDBS COMPOUND-NAME
  1 64 72 50 70 897 CYCLOHEXANE
CS>E
*** BYE ***

```

Fig. 1 Example of the combined search

Sample is cyclohexane. For detailed explanation, see text. Underlined is input.

Fig. 1 に示す。試料はシクロヘキサンである。まず CS コマンドで複合検索プログラムを起動し、次に IR の入力データ 2928(波数)と 10(透過率)を入力して検索すると合格したスペクトルが 21 件あることを回答してくる。そこで L コマンドでその内容を見ると正答のシクロヘキサンは第 3 位に出力されている。そこで次に A コマンド(追加データによる検索)により ^{13}C -NMR のデータ 27(横軸 ppm)と 100(強度)を入力すると、21 件の内で合格のものが 3 件になったことを回答してくる。正答は第 2 位になっている。更に MS のデータ 56 と 100 を入力して検索すると正答のシクロヘキサンだけに絞られたことが分かる。

文 献

- 1) 田辺和俊, 平石次郎, 田村禎夫, 山本 修, 柳沢勝, 和佐田宣英: 分化, **33**, 95 (1984).

☆

Some considerations on the combined search by infrared, ^{13}C -nuclear magnetic resonance, and mass spectral data. Kazutoshi TANABE, Jiro HIRA-

ISHI, Tadao TAMURA, Osamu YAMAMOTO, Masaru YANAGISAWA, and Norihide WASADA (National Chemical Laboratory for Industry, 1-1, Higashi, Yatabe-cho, Tsukuba-gun, Ibaraki, 305)

As a continuation of the preceding paper, several points have been considered on the combined search by infrared, ^{13}C -nuclear magnetic resonance, and mass spectral data. The effect on the noise ratio in the search result of the threshold value of peak intensity in the coding has at first been examined, and it has been found that the appropriate threshold value is 50. Next, the effect of the number of input peaks has been considered, and it has been found that if data of two peaks per spectrum are entered and the threshold value of 50 is used, the noise ratio in the combined search is 0.01 %. It has been concluded from these facts that the combined search is effective for the identification of the correct answer from 10000 compounds.

(Received December 22, 1983)

Keyword phrases

combined search by infrared, ^{13}C -nuclear magnetic resonance, and mass spectra; combined search by spectral data; combined search by computer.
