フロッピーディスクを用いたパーソナル質量スペクトルデータベース

I. BASIC によるデータベースの構築とその利用

Personal Mass Spectral Data Base in Floppy Disk System

I. Construction and Utilization of the Data Base

by BASIC Language

小 山 保 二*・鈴 木 功*・前 田 浩 五 郎*

Yasuji Koyama, Isao H. Suzuki, and Kogoro Maeda

(昭和56年12月8日受理)

Mass spectral data of hydrocarbons from C_1 to C_{50} were extracted for personal use from EPA/NIH Mass Spectral Data Base into floppy disks of a small computer system.

The extracted data are represented in six terms, compound name, molecular weight, chemical form, ionization method and mass numbers of molecular ion, if present, and its fragment ions and further CAS number. The mass numbers are taken for only maximum 10 or 13 peaks higher than 10 % to the base peak in the order of their peak heights. No exact peak height information is included.

Construction programs of the database and retrieval system are all described with extended BASIC language. Long period experience has proved that the system is convenient for data addition, amendment of the database, and also that the retrieves of the stored data can be made effectively without the peak height values. The most abundant base peak was m/z 43, and the number of molecules having the m/z 43 peaks as the base peak is 151. Even with this large number, only several compound names could be extracted as the candidates for the unknown samples having the m/z 43 base peaks from about the 1800 hydrocarbon data by keying 5 to 10 m/z values as the search peaks.

1. はじめに

天然物、合成物を問わず有機化合物の同定や構造決定に於ける質量分析法の有用性が認められ、Organic Mass Spectrometry という新分野がひらけてすでに

久しい。とくに試料量が極めて微量の場合でもガスクロマトグラフィーとの結合によるいわゆる GC/MS が威力を発揮し、その広範な使用実績例をみても質量分析法の有用性はもはや多言を要しない。しかしこれが実際に可能になるにはいくつかの技術的な諸問題が解

*電子技術総合研究所（茨城県新治郡桜村梅園 1-1-4）

Electrotechnical Laboratory (1-1-4 Umezono, Sakura-mura, Niihari-gun, Ibaraki, Japan)

決されている必要がある。その中でとくに得られた質量スペクトルから最終的に同定や構造決定を行なうには既知化合物の質量スペクトルの集積が多くの場合必要で、構造推定に有効な手掛かりを与える。その効用が認められ、早くから広範な集積が行なわれ、その結果に基づいて近年質量スペクトルデータベースが各方面で活用されるようになってきた¹⁾ とくにコンピューターを用いた検索システムや公衆回線によるTSS利用のデータ検索が普及しだしている。しかしこれらを十分活用し得るにはそれ相当のコンピューターシステムが必要であるほか、TSSを利用する場合は一般に必ずしも安価でない使用料等の問題もあり、もう一つ普及が十分でなく、実際には印刷物によるデータブックを利用した手検索ですまざるを得ない研究者の数が極めて多い²⁾

研究室で、データブックを調べる程度の手軽さで、実験をしながらその進行の状況に応じてデータベースが使用可能となれば、その効用は極めて大きいと思われる。こうした言わばパーソナルなデータベースが今後の方向と思われ、それらをどう作り、どう利用するかが今後の課題であろう。そこで今回はこれらのパーソナルデータベースを作り、利用するという課題に取り組むことにし、『安価なシステムで誰にでも作れる』という観点からシステムの検討を行なった。そうした観点からまず使用するコンピューターとしてはLSI技術の急速な進歩に伴ない高性能化、低価格化の著しいマイクロコンピューター（一部のパーソナルコンピューターも含む）システムをを検討の対象にしたが、今回設定の観点を満たすのに十分と思われる。この種のハードウェアシステムでは今のところ BASIC 言語がもっとも広く普及しており、学習も容易で今回の目的に種々の利点を持っていて適合性が高い。それで今回は言語としては BASIC を使用することにした。

参照したデータベースは Heller³⁾によって編集された質量スペクトルデータベースであり、今回はこのデータ集から炭化水素のみを取り出し、1800余りの炭化水素データベースを完成させることができた。これらのデータを実際に構築する際の種々の技術上の問題点等について検討し、さらに収録されたデータを用いて実際に検索等を行なってみた。これらの例などの結果から今回検討のシステムについて、パーソナルデータベースとしての有用性を確認できたと思われる。それらの詳細について以下に述べる。

2. 基本的な考え方

今回の本システム設計を行なうにあたっては以下に述べる諸点を基本的な考え方とした。

1. パーソナルなレベルのシステムを使う。

最近の LSI 技術の進歩により論理回路の集積度が年々高まり、高性能で安価なマイクロコンピューターが量産されるようになってきて、かつてのミニコンレベルのものが極めて低価格で、しかも机上型の CRT、キーボードを含んだ小型のシステムが実用化しだしている。こうした状況では超大型のコンピューターに依存した集中型の方向のみでなく、これに知能度の高い端末をそなえた分散型の方向か、その端末の独立したパーソナルな方向に進むことはまず間違いない。手元に置いて個人個人が自分のものを使うという時代になるであろう。今回はまずその方向をしっかりとふまえることを第一とした。

2. 情報は本来極めてパーソナルなものである。

前項で述べたことを相補する形で次のことが言える。何がほしく、何を必要としているかは全く個人の問題でパーソナルなものである。ましてそれをどう利用するかは個人かまたは一つの組織の問題で他人は関与できない。ある人にとって極めて貴重な情報も他人には全く無価値ということが常に起る。情報のもつこの側面はもっと強調されてよい。今回はこの考え方を本稿の中心的な考え方にすえたい。

3. 自分のものは自分で作る。

従って標題のテーマをかかげたい。現在のところ多用なニーズにこたえるプログラム群も含めた便利なシステムが安価にという状況ではない。共通な使用目的等で検索するにはあまりにも周囲の経験が乏しくそれに関する情報が不足している。こうした状況では研究に意欲的で情報をほんとに必要とする研究者には在来のシステムでは多くの場合十分でなく、結局自分で作って行くことになる。前項2で述べたことから自分の目的に一番適しているものは結局自分で作って行くしかない。

自分のものは自分でつくり、そして使う。これが基本であろう。

4. 扱う対象は非常に小数。

幸いなことに生産や研究の現場では毎回対象としている化合物の種類はそう多いものではない。多くの場合およそ素性がわかっていて、その中で何であるかを知りたいとか、出発物質がわかっていて反応物にも

予想がつき得る場合が多い。何10万というデータを常に扱うことはあり得てもそう多くはない。今回はこうした一般的な場合を検討の対象にする。

5. 言語としては BASIC を使う。

前項までの状況で自分で容易に取り組むのに今のところ一番適合していると判断されるものは BASIC 言語を使うことであると考え、今世界の各方面で最も広く使われている高級言語はおそらく COBOL であろう。しかし科学技術の世界に限って言えばそれは言うまでもなく FORTRAN である。この共通言語の普及がプログラムの共有化にはたした役割は極めて大きい。しかし前項迄の諸条件を勘案すればそれは BASIC ということになる。マイクロコンピュータの普及とともに爆発的に広がりだし、いまや minimal BASIC から Extended BASIC に移行しつつあり、ANSI や ISO を中心とする国際標準化が進んでいて、今小型の分野でもっとも広く使われている基礎の固まった言語である。^{*}

以上のような特長は本質的に文字で表現されているデータベースの取り扱いにはむしろ適していると思われる。初心者でも取りかかれ、かつ互換性もあるという点で前項迄の目的に合うものと思われる。しかもデータの検索のみでなく、データベースの構築にも BASIC を使い前項3の要求にこたえるようにする。以上の理由で今回は BASIC 言語を使用する。

6. Two Drive Floppy System を使う。

前項4で述べたように取り扱いの範囲が限られているので、それぞれの化合物系ごとか目的系統ごとにフロッピーディスクに必要なデータを貯え、目的に応じて取り扱える人間の関与する方式をとることにする。今回の経験でも標準型(8インチ、単密度、片面)の一枚のディスクに約2500化合物のデータが蓄積可能であるので十分であろう。このように一枚のディスクは

データ収納専用ディスクとして使い、もう一枚のディスクはFDOSやデータ構築や検索用プログラムの蓄積用として使いわけて使用すると便利である。しかも前項1で述べた小型のシステムは殆んどTwo Drive Systemが標準になっているので今回は少し先の状況を勘案してこのシステムとする。

最近ではすべてシステムはOS(Operating System)の管理下におかれ、コマンドによる入出力可能な形が普通なのでプログラム、データ情報などのファイルの結合、切り離しが容易で、従って実用的なファイル管理機能がそなわっているものが多い。今回はこの機能を持つFDOS(Floppy Disk Operating System)のもとで動作するTwo Drive Floppy Systemとする。CP/Mが現在比較的低価格で汎用な形で普及しているので、今回の選択は決して特殊なものではない。

7. データは m/z のみにする。

以上の諸項目の考え方の上に立って実際に取り扱われる質量スペクトルデータの形式として基本的なことはピークの m/z のみを入れ、強度に関する情報は入れない。たかだか順序でその大小を表現するに止める。データを集録すると一般に同一化合物のスペクトルの再現性が問題になる。⁴⁾ 今回のデータベースはパーソナルで少しく検索の予備的な性格を強調する意味で m/z のみとする。この妥当性については以下の諸節にみられる通りである。

以上の項目を基本的な考え方とした。

3. 質量スペクトルデータの構築

3.1. Format についての考え方

今回参照したEPA/NIH質量スペクトルデータベースは印刷物として出版されている形式のものである。このデータ集では一つの化合物を質量数、化合物名、化学式、構造式、質量スペクトル、CAS登録番号の

^{*}以下に一般的なExtended BASICの言語体系の特長を列記しよう。

- i) 極めて簡単で自然な言語体系で、習得が容易である。
- ii) 一般的な数学関数も取り扱える内容を持っている。
- iii) 文字を常数または変数として取り扱え、常数は" "とすれば任意に設定でき、プログラム中で使える。
- iv) 文字変数は、AS, B\$ の形で定義でき、DIM AS(I), B\$(J) と宣言しておく。IF AS=B\$ GOTO 行番号 とすることで任意の文字列が、I, Jの範囲なら同一のものかどうか判別できる。しかもAS(I, J)により宣言されたDimensionの範囲ならAS中の任意の連続文字列を指定して比較できる。
- v) LEN(AS) で定義される関数で文字数の取り扱いができる。これは文字の取り扱いに極めて有効である。
- vi) 多くはインタプリタなのでプログラムを随時修正可能で、デバッグがそのままの形で可能となり、プログラムの修正、更新に都合よい。但し少しく実行時間がおそい。

6項目によって記述されている。これらの項目について実際に質量スペクトルデータベースを構築する上で今回用いたようなキーボードからデータを入力する方式では、鎖状の炭化水素などの一部の化合物を除いて構造式をそのままの形で入力することは困難であるので、今回は構造式については入力しないこととした。構造式は除くとしても化合物を表現する項目要素としてはその他の5項目はデータベースとして絶対に必要な項目である。そしてまた現在集積されている質量スペクトルデータベースは約23種程あり²⁾、例えばMSDCなどを見ても重要な項目はこの5項目である。このほかにはイオン化の方式などの実験条件の項目も重要と思われるので、この項目を加えることとした。以上の6項目で一化合物を表現することを基本としたが、将来分子構造に関する情報とか質量スペクトル以外のNMRとかIR等のデータを追加入力可能なようにSpare Columnを付け加えた。このことは今後のデータベースを拡張、充実する上で重要であると思われる。今回はこの項目を含めて合計7項目で化合物を表現することとした。

次にこれらの7項目を効率よくディスクファイルに書き込むためにそのStorage Formatについて検討した。1章で述べたように小型のマイクロコンピュータシステムによる端末処理を考慮して一ラインは80文字に限定する形式とした。この80文字×3ラインの中に先に述べた7項目を配置したFormatをFig. 1に示す。

1ライン目には分子量1から最大9999まで、化合物名は75文字まで表現できるものとした。

2ライン目には化学式として6～7構成原子までを考慮して20文字分、イオン化の方式は電子衝撃をEIなどと略記することとして10文字分、質量スペクトルについては基準ピークを筆頭にイオン強度の大きい順に最大10～13個のピークの m/z を収納できるように40文字分、これにCAS登録番号10文字分として都合80文字として構成した。

3ライン目は80文字分をすべてSpare Columnとして割り当てておくこととした。なお分子量の項目では分子量としての数字の先頭に英文字Mを付加した。こ

のことは1ライン目の最初の文字は必ず「M」となることを示しており、化合物を表現するいろいろな項目全体のヘッダーとしてMを使用できるもので、検索等の立場からはきわめて有用な方式であった。

3.2. データ構築の実際⁵⁾

実際にデータベースを構築して行く手順について以下に述べる。データベースの構築には通常データベースマネジメントシステム(DBMS)が用いられる。しかしここではあくまでも第2章3項目に準拠する立場なので、BASIC言語を用いて構築のプログラムも作った。

このプログラムのフローチャートをFig. 2(a)に示す。なお、このフローチャートは3.3で述べるデータ修正、追加のプログラム(b)に一部修正の上含まれるよう工夫した。従って、データ構築用プログラム(a)は(b)の左側の部分がほぼそれに該当するものとなる。詳細は後述する。ステップ数は約100である。

プログラミングはデータの書き込みを必ず再確認して書き込む方針とし、“人間はとかく間違い易い”ことを前提とした。従って3.1で述べたように一化合物あたり3ラインで構成されているが、一ラインごとにデータをディスクファイルに書き込む方式とした。

データ構築プログラムをRUNするとまずデータを格納するためのファイル(ユーザーが任意のファイルネームをつける)を開いてCRT上には次の形式

STORAGE FORM OF MASS SPECTRAL DATA

Minimum Items Required

Molecular Weight		Compound Name	
Chemical Form	Ionization	Mass Peaks	CAS Number
SPARE COLUMN			

M1~M9999

5		75	
M160	4a,8a-Ethanonaphthalene,1,4,5,8-tetrahydro-		
C12H14		104 129 130 91 117 115	20295-17-4
20		40	
10		10	

Seven Elements Ten Peaks in the order Registration
EI or CI of Magnitude Number

Fig. 1. Storage form of mass spectral data.

でキーイン待ちとなる。

MOLECULAR WEIGHT :

COMPOUND NAME ** :

オペレーターはそれぞれの要求に応じてキー操作するとキーインされた文字が一ラインに編成され、直ちにCRT上に現われてくる。そして今キーインした分子

量および化合物名に誤りがあるか否かを聞いてくるのでもし正しい内容でない場合には再び分子量からキーインし直すことにする。このように一ラインずつ正しく入力されているか否かを確認しながらデータを順次構築して行くという方式とした。一ライン目が正しい内容としてディスクファイルに書き込まれると続いて2ライン目がキーイン待ちとなる。

Fig. 2. Flow charts for Data Base construction (a) and for appending and amendment (b).

CHEMICAL FORM **: ✓
 IONIZATION ***: ✓
 MASS PEAKS ***: ✓
 CAS NUMBER ***: ✓

とそれぞれの項目をキーインすると化学式、イオン化方式、ピークの m/z および CAS 登録番号が一ラインに編成されて CRT 上に出力される。オペレーターはデータの正誤を確認してから一ライン目と同様な操作によって正しいデータをファイルに書き込むことでデータの構築を進めて行くことができる。なお今回参照したデータはイオン化の方式がすべて電子衝撃によるものであったので、データ入力の煩雑さを軽減するために空欄とした。このことは電子衝撃以外のイオン化方式によるデータの場合はそれを明確に入力データとして書き込むことになるので、今回のこの方式が支障となることはない。

つづいて三ライン目は

SPARE COLUMN ***: ✓

となるが現在データは何も入れていない。以上のように三ラインすべてがファイルに書き込み終るとデータの構築を終了するか引きつづき構築を行なうか質問してくるのでこれにこたえて適宜オペレーターの意向で決めることができる。実際にデータを構築していく場合には数日から数カ月は要すると思われるので、データはあくまで、ファイルの後部に順次追加していくデータ構築プログラムが大変便利である。

さて一般的には化合物名は 1-Propane, 2-methyl- の例のようにカンマ (,) が用いられる場合が少なくない。ところで今回実際に使用したコンピューターシステムは Fig. 3 に示すようないわゆるミニコンピューターシステムである。しかしその性能と構成は現在までのコンピューター技術の進歩により事実上マイクロコンピューターシステムにほとんど対応するものとなっており、このシステムを使っても実際は第 2 章の 1 項の要請に対応するものとなっている。このシステムで使用している Extended BASIC はカンマや Carriage Return は特定の働きをするように設計されているので、化合物名の中のカンマはそのままキーインすることは許されない。そこで今回はキーインするにはカンマの代りとしてセミコロン (;) を代用することにした。しかし実際にファイルに書き込む時はセミコロン

Fig. 3. Computer system for the present study.

をカンマに替えて書き込むことができるので収納されたデータは標準的な化合物名のままとされている。それ故 5 章で述べるようにピーク検索を行なってその結果を表示する、またはプリントアウトするといった実用上の点からは何ら支障ない。今回のこの処置は他の BASIC プログラムを使うときにも参考になるものと思われる。このほか化合物名として用いられているギリシャ文字とか特異な表現などについてはそれぞれもっともふさわしいように変換して表現することとした。これらを Table 1 に示す。

Table 1. Table of Symbols Used in This Study

EPA/NIH Mass Spectral DataBase	Present Personal Mass Spectral Data Base
,	;
$3^{3,8}$	3 (3, 8)
R^*	R^*
$\pm$	+-
α	A
β	B
γ	G
φ	P
ξ	X

3.3. 構築したデータの修正及び追加⁶⁾

3.2 で一化合物あたり 3 ラインで構成し、かつデータをファイルに書き込む時は一ラインづつキーインされ

た文字列が正しいと確認したものを書き込むという方式であることを述べた。しかしそれでもなお入手を介したキーボードからのデータ入力的方式では今後の経験からは最初に構築されたデータが正しいデータベースとはなり得ないのが実情であった。現実の問題としてデータをキーインする際に誤り易い項目として化合物名があげられる。この化合物名は大小英文字と数字やギリシャ文字で表現されているので、キーインする大小英文字の使い分けや長い化合物名である場合など誤って入力してしまった例が多かった。そのほかピークの m/z や CAS の登録番号の数値そのものを誤入力してしまった例も多く、不馴れた経験からの例では約 1800 のデータを取り扱ってそのうちほぼ 2~3% 程度の誤り率であった。このようにキーボードからのデータ入力では多少の誤入力は避けられないことであり、さらにオリジナルなデータからスペクトルの情報を取り出す際にも誤りが入ってしまうことが予想される。このような状況のもとでデータベースを完成させる過程ではデータを修正したり、入力もれのデータを追加するという操作がどうしても必要となる。これらのデータ修正、追加のプログラムのフローチャートは既に Fig. 2(b) に示した。以下にデータの修正および追加の手順について述べる。

まず誤って入力されたデータの一つ前のデータの CAS 登録番号を検索して次の修正すべきデータの入っているラインを CRT 上に表示する。そしてこのラインの内容の正誤を再確認して誤っている場合はそれぞれの項目を再度入力して正しいラインとして再構築する。

二ライン目と三ライン目も同様に誤ったデータの時のみ再入力することとし、誤っていないラインは同じデータのまゝとする。このように一ラインごとに修正、追加を行なう方式とすることによりデータ構築プログラムを一部変更するだけでこの修正、追加のプログラムの中で使用できることになる。Fig. 2(b) に示したようにデータを修正するかそれとも追加するかを指定することでそれぞれのルーチンを実行するプログラムとなっているので、目的に応じてオペレーターが選択することができる。なおここで追加するというのはデータをファイルに書き込む際に誤って入力洩れしたデータを後でしかるべき順序の箇所に挿入することであり、データを単に追加入力することなら 3.2 で述べた構築プログラムを用いてデータ構築を行

なえばそれでよい。

なお実際にはファイルの構成は使用している BASIC プログラムの制約をうけざるを得ず、ファイルアクセスの実態は次章に述べるごとくである。

4. データのアクセスとディスク内配置

4.1. データとファイルアクセス⁶⁾

データを構築する際にキーインしたデータをディスクファイルにアクセスする方式には今回使用した Extended BASIC には二通りの方式がある。

① 一つは今迄述べてきたデータ構築プログラムで採用しているシーケンシャルアクセス法である。この方式はレコードを前方より後方へ順番に処理する機能が中心で、ファイルの中からデータを読み出す、またはデータを書き込むかいつれかの一方を選択し、それを専用する方式になっている。そしてデータを構築する過程で随時データをファイルに書き込む場合にはデータ群に付けられたファイル名を指定してそれに順次書き加えていくという単純な手順でデータをファイルに収納できる方式となる。またデータ検索を行なう場合にも特にレコードを気にせずに順次検索項目とファイルデータとが比較でき、検索が簡単に不自由なく行なえる。しかしデータの修正とか追加を行なう場合に、同じファイル名の中で修正すべきデータを新しく正しいデータに直接書き換えるということはできない。データの修正を行なう場合には今までのファイル名とは別の名前の付いた新しいファイルにすべてのデータを書き換えるという手順をとらねばならない。すなわち、まず修正箇所の一つ前までのデータを別のファイルに書き移し、次に問題のレコードを修正してその新しいファイルにつけ加える。そのファイルにもとのファイルの修正箇所以後の部分を順次加えて新しいファイルを完成させる。このようにして修正後の正しいデータを収納したファイルが新しいファイル名で登録されることになる。データ検索を行なう立場からは検索対象とするファイル名は修正前のファイル名となっているので、前のファイル名を削除して修正後のデータのファイル名を最初に用いたファイル名に Rename しておくという方式をとらねばならない。

② ファイルアクセスのもう一方の方式としてランダムアクセス法がある。この方式ではファイルのレコード番号を指定することにより、ファイルの任意の部分のレコードの読み書きが可能である。それ故シーケ

ンシャルアクセス法ではデータの書き換えが不自由であったがこのアクセス法では同一ファイル名の中で自由に書き換えが可能である。すなわち修正すべきデータが収納されているレコードを読み出し (READ) て CRT 上に表示し、修正の必要を確認してから該当レコードに正しいデータを書き込む (WRITE) ことで目的が達せられる。しかしこのようなランダムアクセス法の場合は各データと対応するレコード番号を絶えず正確に把握しておかねばならない。またこの方式ではデータ修正の場合に限らず将来スペア

カラムに必要なデータを入力する際にも予め指定したレコード長の範囲なら何ら支障なくデータベースの拡張が行なえる。しかしこのアクセス法の場合でも連続して収納されているデータとデータの間に新しいデータを挿入することはできない。これを実現するには①で述べたように新しいファイルにデータを書き換えながら途中に必要なデータを書き込むという方式をとらざるをえない。さらにランダムアクセスにより構築されたデータの占める領域はシーケンシャルアクセスでのそれと比べて一般に大きくなり (今回の例では約2.2倍)、この方式では Floppy Disk あたり約1100個程度の化合物しか収納できない。従って今回は①の方式によってデータベースが構築、修正された。参考までにこれらファイルアクセスの機能を図示したものを Fig. 4 に示す。

4.2. データとプログラムの配置³⁾

以上前節で BASIC プログラムによって質量スペクトルデータを構築して行く実際の過程を述べ、さらにデータ修正プログラムの必要性についてもふれた。しかし勿論それで充分ではなく、さらに種々の検索プログラム等が作られる必要がある。今回使用したプログラムをデータ等すべて機能別に大別すると ① FDOS, ② BASIC のシステムプログラム、③データ構築、修正、検索等のユーザープログラム、④質量スペクトルデータ の4

つに分けられる。現在使用の Floppy Disk 一枚の記憶容量は約 300 KBytes であり、比較的高機能なためもあって FDOS と BASIC を含めたシステムソフト (①と②) で約60%の領域を専有してしまう。このことはデータベースを構築する立場からはデータ領域が十分でなく好ましくないことを意味する。しかし優れたシステムプログラムを常時使用できる利点も無視できない。そこで第2章6項で述べたようにこれら双方を満足する方法として Two Drive Floppy Disk を活用

Fig. 4. Comparison of sequential access and random access. Symbol X means the append can not be made there.

Fig. 5. Storage form of programs and data in floppy disks.

し、①～③のプログラム群を一枚のディスクに格納し、これをシステム起動を行なう主ファイルとした。一方データ群はデータ専用ファイルとして用いることとした。この結果データ領域が十分大きくとれることになり実用上都合がよい。このようなプログラム配置の様子を Fig. 5 に示す。なお将来これらのデータ群が著しく増加した場合には新しいディスクファイルにデータを収納すればよく、複数枚のディスクにデータを分散させることでデータ領域の問題は解決される。このようにデータが複数枚のディスクに分散されることは検索等の立場からプログラム実行の過程でデータファイルの交換等の手間がかかり、余分な作業が入ることになる。しかし一般的には対象とする化合物に対してはある程度の情報（組成元素の限定とか化合物群の組み分けなど）を基に検索等を活用するという場合が多いと思われる。従ってデータはファイルごとにまたはファイルの中をさらに複数のブロックに分割したファイル名で登録しておき、必要なファイルのみを選択して使用する方が有効ではないと思われる。また今回の経験から今まで述べてきたデータの分散、さらにこれらを適当に分割した方式が实际的と思われる。

以上述べてきた方式により構築されたデータを実際に使用してみた。それらを以下に述べる。

5. 炭化水素データの内容

5.1. 炭素数と水素数の相関⁷⁾

今回ディスクにより収録された炭化水素は C_1H_4 から $C_{50}H_{102}$ 迄総数1813個であった。まずはじめにこれらの総数について炭素数が1から50までについてそれぞれの化学式 (C_nH_m) がどのように分布しているかを調べた。その結果を Fig. 6 に示す。 C_n について C_1 から C_{10} までは順次増加しているが、 C_{10} をピークとしてこれを境にランダムに減少していくことが認められ、再び C_{20} で小さなピークを持つという分布となった。この結果は実際にこれらの分布が本当に自然界に存在する炭化水素全体を反映しているかどうかなお検討を要すると思われるが、炭素の数が奇数か偶数かによって C_{10} 以上ではその分布に差が認められる。これは対称、非対称などの構造上の問題が関係していると思われるが詳細は検討していない。

次に炭素数 n に対する水素数 m の種類の数の関係について調べた。その結果を Fig. 7 に示す。これは C_1 に対しては C_1H_4 の1種類、 C_2 に対しては C_2H_2 、 C_2H_4 、 C_2H_6 の3種類、 C_3 に対しては C_3H_4 、 C_3H_6 、 C_3H_8 の3種類というように $C_7 \sim C_{18}$ までは大よそ順次増加していることがわかる。しかし炭素数がこれ以上になってくると n の増加に対して水素数のそれはピー

Fig. 6. Abundance distribution of hydrocarbons C_nH_m from C_1H_4 to $C_{50}H_{102}$.

Fig. 7. Number of C_nH_m for a given carbon number n in C_nH_m .

クを持つような分布となった。さらに炭素数が C_{18} から C_{32} の間では奇数と偶数とでは明らかに偶数炭素が多く分布しているという結果となった。 C_{23} , C_{24} で極小値を持つような分布となっているが、この結果は収納されているデータの質によるものなのか、それとも化合物の構造に起因するものかは現在の結果だけでは判断できない。

そこで炭化水素の構造形態を鎖状か環状を含むかという単純な選別を行ない、これと炭素数との関係を調べた。その結果を Fig. 8 に示す。一般に炭化水素では炭素数が増大すると環状化して存在するものが増えると思われ、1800 余りの化合物についてそれらを調べてみると C_3 から C_{11} までは n の増加とともに環状化する割合が次第に大きくなり、その増加率はほぼ比例している。しかし炭素数 n が 12 以上になると環状と鎖状との割合が大よそ一定範囲 (約 77%~93%) 内に収まり、 n の数に無関係となることがわかった。しかし C_{23} についてはこの一定範囲からややはずれる結果となり、前に述べた結果などから判断して今回の収納データの質の問題である可能性を示して

いるものと思われるが、明確な結論は得ていない。なおここで言う環状形とは炭素結合が閉じた形態を一つ以上含むものをさし、Pentacosane, 13-phenyl (Me $(CH_2)_{11}CHPh(CH_2)_{11}Me$) のようにフェニル基と長い側鎖を持つ化合物も環状のグループとして取り扱っ

Fig. 8. Rate of ring-shaped compound number to the total hydrocarbon number for a given carbon number n in C_nH_m .

た。

C₃₂ 以上では化合物の数そのものが少ないのでここには記載していない。

5.2. ピーク検索プログラム⁵⁾

3章で述べた構成で質量スペクトルデータベースができていますが、データ検索を行なう場合どのような検索が有用かつ実用的であるかはデータベースの利用者によって多岐にわたると思われる。しかし多くの場合質量スペクトルの中から得られるピークに関する情報をいかに効率よく有効に利用し得るかということが多いようである。そこで実際に得られた質量スペクトルから指定されたピークの存在の有無であるとかピークの m/z の値の順序の関係などを調べ、さらにこれらを総合的に検討を加えることで少なくともある種の化合物であろうという推測が可能となるものと思われる。このようなピーク検索を考慮して以下の検索プログラムをつくってみた。

① ピーク高さの順序を考慮した検索 (in order search)

このプログラムをスタートさせると最初に
MASS PEAKS?

と聞いてくるので、基準ピークをはじめとしてイオン強度の大きい順に質量数を任意の数だけキーインする。データ構築プログラムの時述べたように誤った m/z の値をキーインして無駄な検索を防ぐために、キーインされたピーク群を一端 CRT 上に表示し、検索すべきピークの m/z であるか、ピーク順序が正しいかを確認してから検索の実行に移るという方式とした。ピークの m/z をキーインするとデータファイルの中から最初のレコードの内容が読み出される。そこでこの1レコード目の先頭の文字が「M」であれば無条件に次の2レコード目が読み出され、31文字以降に格納されているピークの m/z とその順序が完全に一致するかどうか比較される。この判定は前方一致の方式のため対象とするピークの数が少ないと多くの化合物が検索結果として出力されることになる。なお比較の結果一致しなかった場合には次のレコードを読み出して再び「M」をサーチして同様な過程を経ながら検索を続行する。検索対象のピーク群が存在した時は三ラインをプリントして再び検索を繰り返すことになる。データファイルがすべて参照し終ると検索された化合物の数をプリントし、検索を終了する。必要に応じて今と

違ったピーク順序による検索とか、指定するピークの数を変更するとか繰り返し実行可能なようにプログラムを構成した。

② ピーク高さの順序を無視した検索 (not in order search)

①で述べた例ではピークのイオン強度が基準ピーク、第2位、第3位……というようにその順位が明確に区別される場合には十分適応できるプログラムと言えよう。

Fig. 9. Flow chart for searching the candidate compound names by giving m/z values.

しかし実際に得られるスペクトルでは実験条件などによって相対的なイオン強度の値が異なる場合も起り得るものと思われる。このような場合にはピークの順序そのものにはこだわらず、相対的に大きいピークの存在の有無により比較検討を行なうことが有用と思われる。これに適応するピーク検索プログラムについて検討した。そのフローチャートを Fig. 9 に示す。データ構築の際に最大10~13個のピークを収録したのでこのプログラムでは検索の対象となるピーク数を基準ピークから第5位までの5個の特徴的なピークとすることとした。すなわち上位5個のピークはピーク高さの順序と無関係に存在するか否かを検索できるものである。このプログラムのフローチャートは①のプログラムとほぼ同様な構成となっているが、これらに追加した点としてはキーインされたピークの数を超えて基準ピーク以下その個数分だけ検索の対象とすることとその個数は最大5個までとすることである。実際に作った検索プログラムでは①と②は1本のプログラムにまとめられており、プログラムをスタートした後に①を選ぶか②を選ぶか目的に応じて選択できるように構成されている。今回作ったプログラムは検索結果を得ることを主な目的としたので、実行効率など十分と言えないが所期の目的を達することはできた。ステップ数は約200であった。

5.3. ピーク検索の例

今回のようなイオン強度の値が入力されていないパーソナルなデータベースを使用して特徴的なピークだけを選び、1800余りの化合物の中からどの程度の割合で検索結果が得られるか、何個までしぼれるか実用的な観点からも重要な問題である。そこでピーク検索の一応の目安を得るために基準ピークの分布を調べ、その中の分布のもっとも多い基準ピークを持つ分子種について先に述べたプログラムで検索を行なった。なおこの基準ピーク頻度分布も BASIC プログラムを用いて得られたもので、このような統計的な処理も比較的簡単に行なえ、目的の結果を手軽に得ることができた。

(1) 今回採録された炭化水素の全スペクトルの中から基準ピークの頻度分布を調べてみると m/z 43 ($C_3H_7^+$) が最も多く (151個)、以下 m/z 57, m/z 55……という順序で頻度が多いという結果となった。⁷⁾ これらの基準ピークの頻度の多い上位10位までを Table II

に示す。基準ピークの種類は m/z 16 ($C_2H_4^+$) から最大 m/z 610 ($C_{48}H_{34}^+$) までの196種であった。

Table II. Base Peak Distribution in the Order of the Abundance Number

BASE PEAK DISTRIBUTION				
m/z			M^+ Peak	
1	43	$C_3H_7^+$	151	12
2	57	$C_4H_9^+$	122	8
3	55	$C_4H_7^+$	112	58
4	41	$C_3H_5^+$	97	45
5	91	$C_7H_7^+$	85	45
6	67	$C_5H_7^+$	72	49
7	81	$C_6H_9^+$	53	30
8	83	$C_6H_{11}^+$	52	21
9	105	$C_8H_9^+$	45	26
10	56	$C_4H_8^+$	44	22
⋮	16	$C_2H_4^+$	1	1
⋮	610	$C_{48}H_{34}^+$	1	1
			1813	1094

m/z 43を基準ピークに持つ151個の分子種についてさらに2番目に大きいピークの分布を調べてみると m/z 29, 41をはじめ m/z 221 まで19種類であった。その中で m/z 57 ($C_4H_9^+$) が77個と一番多く約50%を占めることがわかった。さらに今度は m/z 43, 57 というピーク順序を与える77個の分子種の中から第3番目のピークの分布を調べると m/z 71 ($C_5H_{11}^+$) が41個と最も多かった。すなわち基準ピーク m/z 43 を持つ分子種151種の中からピーク強度の大きい3個のピークを取り出し、ピーク高さの順序を考慮した場合 m/z 43, 57, 71 となる化合物がもっとも多く存在している可能性が極めて高いといえてよいであろう。以下同様に上位5位までのピークについて調べてみると m/z 43, 57, 71, 41, 85 となる分子種は総数1813のうちたかだか20個であるという結果となった。この20種の化合物はすべて完全飽和型の鎖状炭化水素であり、元の分子は炭素数が13 ($C_{13}H_{28}^+$) より大きい化合物であった。

つぎに基準ピークの頻度の多い m/z 57 ($C_4H_9^+$) 122個の分子種についても m/z 43の例と同様に上位5個のピーク分布を調べてみると m/z 57, 43, 71, 85, 41 となる分子種が一番多く、15個であった。これらの結果を Table III に示す。基準ピークが m/z 43, 57 のいずれの場合も上位5位までのピークを pick up する

Table III. Results of the Searching Corresponding to Fig. 10

MASS PEAKS		IN ORDER	NOT IN ORDER
m/z 43		151	151
43 57		77	153
43 57 71		41	107
43 57 71 41		29	72
43 57 71 41 85		20	71
m/z 57		122	122
57 43		76	153
57 43 71		49	107
57 43 71 85		27	44
57 43 71 85 41		15	71

Table IV. Results of the Searching Corresponding to Fig. 12

MASS PEAKS		IN ORDER	NOT IN ORDER
m/z 55		112	112
55 83		24	51
55 83 41		21	50
55 83 41 97		13	24
55 83 41 97 69		6	14
m/z 41		97	97
41 69		28	43
41 69 55		9	30
41 69 55 27		2	6
m/z 91		85	85
91 92		16	31
91 92 39		5	6
91 92 39 65		3	5

ことで炭化水素全体の中から大よそ1%程度にしばれる可能性を示したものと考えられる。さらに対象とするピークを8個までとすると双方の場合とも6~7個になることもわかった。

(2) 以上ピーク高さの順序を考慮した検索結果についての例を述べたが、これら順序を考慮しない場合にはどの程度の割合で検索できるかについて検討した。基準ピークが m/z 43 (151個), m/z 57 (122個) であるいつれの化合物の場合も(1)で述べた通り上位5位までの頻度の多いピークを取り上げてみると、それらは m/z 41, 43, 57, 71, 85 の5個であり、ピーク高さの順序に差はあるがピークの m/z はすべて同じ上記の5個が共通に在り、しかもそれ以外にはないという結果となった。これらの5個のピークを順序を無視

して検索すると71個の分子種がこれに該当することがわかった。全化合物の約4%に相当する数である。これにひきかえ次に頻度の多い基準ピーク m/z 55, 41, 91などについて同様なピーク検索を行なった結果では上位5個ないし4個のピーク検索で14個ないし5~6個にしばれることがわかった。それらの結果を Table IV に示す。このことはピーク順序を無視した検索の場合でも基準ピーク m/z 43, 57 の例を除けば4~5個の特徴的なピークを pick up することだけで1800余りの全化合物の中から10~数個程度にしばれる可能性が極めて大きいことを示したものと言える。

実際のピーク検索の結果の例を(1)の場合 (in order search) を Fig. 10, Fig. 11, Fig. 12 に、(2)の場合 (not in order search) を Fig. 13 に示す。

6. Personal Data Base の評価と問題点

パーソナルな質量スペクトルデータベースに係わる技術的な諸問題について検討し、構築されたデータを用いたピーク検索の実例についても述べた。これらの結果から実用的にも十分使用可能であると思われる。さらに次の2点も言及出来よう。

(1) 今回構築されたデータベースでは3.1で述べたように質量ピークのイオン強度の値そのものは入力されていない。相対的なイオン強度の順序のみを入力しただけのものである。しかし5章で述べたように完全飽和型のもを除いたそのほかの例では1ケタにしばれることがわかり、ピーク強度の値そのものを比較しなくとも検索が十分行なえるということが確認されたものとしてよいであろう。実際に研究室や実験室で得られる質量スペクトルのイオン強度の安定性については測定系によって種々異なると言われているが、これらの点からもイオン強度の値そのものはさほど重要視しなくてよいものと思われる。これはなお今後の検討課題である。

(2) C_1H_4 から $C_{50}H_{102}$ までの総数1813個の炭化水素を1枚のFloppy Disk に収納できた。このデータはディスクファイルの中ではさらに C_1 から C_{10} , C_{11} から C_{17} , C_{18} から C_{50} の3つのブロックに分かれて収納してある。このようにデータをいくつかのグループにすることでデータの修正や追加などには取り扱いが容易となる利点があるほかピーク検索等で必要とするデータのグループのみを参照することも可能となるので効率のよい検索が行なえるなど優位な点が多い。

以上データベースの有用性について述べたが、実際の取り扱いに当たっての問題点について以下にふれる。

(3) 3.1 でデータの Storage Format について詳しく検討したが、その中で化合物の文字領域は75文字までとすることを述べた。しかし今回収納された1813の化合物のうち31個について特に長い化合物名を有しているため指定された75文字以内には収まらず76文字以降は入力されていない。このことは本来のデータベースの機能としては不十分とは思われるが、31種の化合物名を比較してみると75文字まですべて同一の化合物

名となるものは今回構築したデータベースには一つも存在しなかった。すなわち一つの化合物名とそれに対応する質量スペクトルは正確に1対1対応していることを示すもので、データ検索を行なう実用的な立場からはほとんど支障とはならないことを示すものと思われる。

(4) 3.2 で化合物名の中でカンマの代りにセミコロンを使用してデータを構築し、ピーク検索等に支障のないことを述べた。これは化合物名以外の項目を用いて検索を行なう場合は問題にならないが、もし化合物

**** IN ORDER MASS PEAKS ****				
MDB 1 CH4 TO C10H22				
MASS PEAKS ? 43 57 71 41 85				
SEARCH DATA : 0				
MDB 2 C11H10 TO C17H36				
M184 Nonane,5-butyl-				
C13H28	43	57	71	41 85 126 55 127 39 56 42 70 17312-63-9
M226 Tridecane,7-propyl-				
C16H34	43	57	71	41 85 29 55 140 56 42 69 99 55045-09-5
M226 Tridecane,6-propyl-				
C16H34	43	57	71	41 85 55 42 56 126 69 99 55045-10-8
SEARCH DATA : 3				
MDB 3 C18H10 TO C50H102				
M282 Eicosane				
C20H42	43	57	71	41 85 55 56 42 69 70 112-95-8
M282 Hexadecane,5-butyl-				
C20H42	43	57	71	41 85 55 56 125 69 42 225 6912-07-8
M296 Heneicosane				
C21H44	43	57	71	41 85 55 56 42 69 629-94-7
M296 Pentadecane,8-hexyl-				
C21H44	43	57	71	41 85 55 69 56 196 197 13475-75-7
M310 Docosane				
C22H46	43	57	71	41 85 55 56 69 42 629-97-0
M324 Tricosane				
C23H48	43	57	71	41 85 55 56 69 42 638-67-5
M324 Heptadecane,9-hexyl-				
C23H48	43	57	71	41 85 55 69 56 210 55124-79-3
M352 Pentacosane				
C25H52	43	57	71	41 85 55 56 69 42 629-99-2
M366 Docosane,11-butyl-				
C26H54	43	57	71	41 85 55 69 56 83 42 210 13475-76-8
M366 Heneicosane,11-pentyl-				
C26H54	43	57	71	41 85 55 69 224 56 225 83 42 14739-72-1
M366 Docosane,9-butyl-				
C26H54	43	57	71	41 85 55 69 56 83 182 42 55282-14-9
M366 Docosane,7-butyl-				
C26H54	43	57	71	41 85 55 69 56 83 154 55282-15-0
M366 Docosane,5-butyl-				
C26H54	43	57	71	41 85 55 126 69 56 83 309 70 55282-16-1
M380 Heptacosane				
C27H56	43	57	71	41 85 55 69 56 83 593-49-7
M380 Docosane,10-nonyl-10-propyl-				
C27H56	43	57	71	41 85 55 69 211 70 99 83 253 55282-31-0
M394 Docosane,7-hexyl-				
C28H58	43	57	71	41 85 55 69 182 56 83 42 55373-86-9
M408 Nonacosane				
C29H60	43	57	71	41 85 55 69 56 83 42 630-03-5
SEARCH DATA : 17				

Fig. 10. An example of the results of search by giving 5 peaks 43, 57, 71, 41, and 85 in this order.

フロッピーディスクを用いたパーソナル質量スペクトルデータベース

*** IN ORDER MASS PEAKS ***		
MDB 1 CH4 TO C10H22		
MASS PEAKS ?	57 43 71 85 41	
SEARCH DATA :	0	
MDB 2 C11H10 TO C17H36		
M212 Tridecane,2,5-dimethyl-	57 43 71 85 41 56 99 29 55 70	56292-66-1
M226 Dodecane,2-methyl-6-propyl-		
C16H34	57 43 71 85 41 55 56 99 140	55045-08-4
M240 Hexadecane,2-methyl-		
C17H36	57 43 71 85 41 55 29 56 99 42 69 113	1560-92-5
SEARCH DATA :	3	
MDB 3 C18H10 TO C50H102		
M254 Heptadecane,8-methyl-		
C18H38	57 43 71 85 41 126 55 154 56 127 155	13287-23-5
M268 Octadecane,2-methyl-		
C19H40	57 43 71 85 41 55 56 29 225 99 42 69	1560-88-9
M296 Eicosane,2-methyl-		
C21H44	57 43 71 85 41 55 56 29 69 99 253 42	1560-84-5
M296 Eicosane,10-methyl-		
C21H44	57 43 71 85 41 55 56 99 69	54833-23-7
M338 Tetracosane		
C24H50	57 43 71 85 41 55 29 69 83 56 99 70	646-31-1
M352 Heptadecane,9-octyl-		
C25H52	57 43 71 85 41 55 69 99 56	7225-64-1
M366 Hexacosane		
C26H54	57 43 71 85 41 55 56	630-01-3
M366 Eicosane,7-hexyl-		
C26H54	57 43 71 85 41 55 69 56 99	55333-99-8
M380 Heptadecane,8,8-dipentyl-		
C27H56	57 43 71 85 41 55 69 309 56 253 281 99	55282-28-5
M380 Hexadecane,8-hexyl-8-pentyl-		
C27H56	57 43 71 85 41 55 69 281 56 295 267 99	55282-29-6
M422 Tetracosane,2,6,10,15,19,23-hexamethyl-		
C30H62	57 43 71 85 41 55 56 69 113 99	111-01-3
M450 Docosane,11-decyl-		
C32H66	57 43 71 85 41 55 69 83 99 97	55401-55-3
SEARCH DATA :	12	

Fig. 11. Second example of the results by giving 5 peaks 57, 43, 71, 85, and 41 in this order.

*** IN ORDER MASS PEAKS ***		
MDB 1 CH4 TO C10H22		
MASS PEAKS ?	55 83 41 97 69	
M112 3-Hexene,2,2-dimethyl-, (E)-		
C8H16	55 83 41 97 69 39 27 29 112 57 13 43 70 690-93-7	
SEARCH DATA :	1	
MDB 2 C11H10 TO C17H36		
SEARCH DATA :	0	
MDB 3 C18H10 TO C50H102		
M250 Cyclohexane,1,1'-(2-propyl-1,3-propanediyl)bis-		
C18H34	55 83 41 97 69 111 207 153 43 67 82 96	55030-21-2
M346 Cyclohexane,1,1'-(3-(3-cyclopentylpropyl)-1,5-pentanediyllbis-		
C25H46	55 83 41 97 69 235 234 67 111 96 81 82	55401-70-2
M346 Cyclohexane,[6-cyclopentyl-3-(3-cyclopentylpropyl)hexyl]-		
C25H46	55 83 41 97 69 67 111 82 81 95 96 139	55401-72-4
M374 Cyclohexane,1,1'-(4-(2-cyclohexylethyl)-1,7-heptanediyl)bis-		
C27H50	55 83 41 97 69 67 234 111 235 82 81 95	55429-36-2
M388 Cyclohexane,1,1'-(4-(3-cyclohexylpropyl)-1,7-heptanediyl)bis-		
C28H52	55 83 41 97 69 263 96 67 111 262 81 82	55334-73-1
SEARCH DATA :	3	

Fig. 12. Third example of the results by giving 5 peaks 55, 83, 41, 97 and 69 in this order.

*** NO1 IN ORDER MASS PEAKS ***									
MDB 1 CH4 TO C10H22									
MASS PEAKS ? SU 83 41 97 69									
M112 3-Hexone, 2,2-dimethyl-, (E)-									
C8H16	55	83	41	97	69	39	27	29	112 57 15 43 70 690-93-7
SEARCH DATA : 1									
MDB 2 C11H10 TO C17H36									
M222 Cyclohexane, 1,1'-(2-methyl-1,3-propanediyl)bis-									
C16H30	83	55	41	97	69	207	67	111	139 82 39 43 2883-08-1
M222 Cyclohexane, 1-(cyclohexylmethyl)-4-(1-methylethyl)-									
C16H30	83	97	55	41	69	43	179	67	81 178 82 96 54965-61-6
M236 Cyclohexane, 1,1'-(2-ethyl-1,3-propanediyl)bis-									
C17H32	83	55	41	97	69	207	67	111	139 82 39 43 54833-34-0
SEARCH DATA : 3									
MDB 3 C18H10 TO C50H102									
M250 Cyclohexane, 1,1'-(2-propyl-1,3-propanediyl)bis-									
C18H34	55	83	41	97	69	111	207	153	43 67 82 96 55030-21-2
M304 Cyclopentane, 1,1'-(3-(2-cyclopentylethyl)-1,5-pentanedivyl)bis-									
C22H40	41	69	83	55	97	111	67	82	125 95 81 206 55255-85-1
M346 Cyclohexane, 1,1'-(3-(3-cyclopentylpropyl)-1,3-pentanedivyl)bis-									
C25H46	55	83	41	97	69	235	234	67	111 96 81 82 55401-70-2
M348 Cyclohexane, [6-cyclopentyl-3-(3-cyclopentylpropyl)hexyl]-									
C25H46	55	83	41	97	69	67	111	82	81 95 96 139 55401-72-4
M346 Cyclopentane, 1,1'-(4-(3-cyclopentylpropyl)-1,7-heptanedivyl)bis-									
C25H46	83	41	55	69	97	67	82	111	95 43 235 81 55429-35-1
M348 Dodecane, 1-cyclopentyl-4-(3-cyclopentylpropyl)-									
C25H48	83	41	55	69	97	43	57	67	111 82 71 125 7225-68-5
M348 Undecane, 1-cyclohexyl-3-(2-cyclohexylethyl)-									
C25H48	83	55	97	41	69	43	96	57	67 111 236 237 7225-69-6
M360 Cyclohexane, 1,4-bis(4-cyclohexylbutyl)-									
C26H48	83	55	97	41	69	82	67	81	95 111 125 221 55334-02-6
M374 Cyclohexane, 1,1'-(4-(2-cyclohexylethyl)-1,7-heptanedivyl)bis-									
C27H50	55	83	41	97	69	67	234	111	235 82 81 95 55429-36-2
M388 Cyclohexane, 1,1'-(4-(3-cyclohexylpropyl)-1,7-heptanedivyl)bis-									
C28H52	55	83	41	97	69	263	96	67	111 262 81 82 55334-73-1
SEARCH DATA : 10									

Fig. 13. An example of the results by giving 5 peaks not in the order of the peak heights.

名を検索対象としてキーインする時はやはりカンマの代りにセミコロンを用い、プログラムの中でセミコロンをカンマに直してファイルのデータと比較されることになる。このように化合物名の取り扱いには若干煩わしい手順を経なければならないが、データベースの構築者自身がこれを使用することを前提とするといふ述べたような規則はさほど障害とはならないものと思われる。

(5) 今回用いた BASIC はインタプリタであるので、プログラミングの初歩の段階では試行錯誤が簡単に行なえて大変都合がよい。しかしその反面実行速度が遅いという点がでてくる。このような欠点をカバーするために BASIC コンパイラの開発が行なわれており、この種のコンパイラを使用すれば実行速度は飛躍的に短縮されるが、まだ一般に広く普及するには至っていない。今回作ったピーク検索プログラムを使用した例では1800余りの化合物の中からピーク高さの順序を考

慮した場合には約5分、5個のピークで高さの順序を無視した場合では約15分という結果であった。これは決して満足すべき時間ではないが、パーソナルデータベースの利用に際し信頼性の高い検索データを得ることを主目的に考えるということでプログラム実行速度の遅さは第二義的な問題として扱うこととした。

7. おわりに

今回は印刷された質量スペクトルデータベースから炭化水素のみを取り出し、化合物を表現するそれぞれの項目をキーボードを介してディスクファイルに書き込むという手順でデータ構築を行なった。そして実際にデータを構築する過程における諸問題について詳しく述べた。これらの一連のデータ構築を行なった経験から

(1) 使用した言語は BASIC であり、覚え易い言語体系であるので種々の目的のプログラムも比較的容易

に作ることができ、所期の目的を達成できる。ここで使用するプログラムはあくまで目的を達するための手段であるという考え方を貫き、プログラミングの優秀さは追求しないこととした。いわゆる全天候型の完成を目指すよりもむしろ目的ごとのプログラムをつくり、正確な結果を得るという方がパーソナルデータベースの場合には得策と思われる。

(2) データベースを構築し利用する上で重要なことはユーザーが作ったプログラムやデータの収納されているファイルは必ず Back up 用として同じ内容のファイルを保管しておくことである。このようにコピーファイルを作っておくことにより万一データファイルがエラーを起こしてしまったとかオペレーションミスでデータを破壊してしまったという場合でも最小限の労力で今までの貴重な蓄積を活用できることになる。

現在前述の方式によって収録された炭化水素データを利用してピーク情報と化合物の構造との対応について調べ、いくつかの知見が得られている。これらの結果の詳細については別の機会に報告したい。

以上述べてきたことがこの種のパーソナルデータベースを構築される際の参考となれば幸いである。

最後に、今回使用した EPA/NIH Mass Spectral Data Base については S. R. Hepper 博士より多くの厚意を戴いたし、豊橋技術科学大学教授佐々木慎一博士、三菱油化中央研究所酒井稔夫博士、化学技術研究

所和佐田宣英氏よりは有益な御助言を戴き、成蹊大学教授飯田芳男博士には文献等について種々御援助を戴いた。また館 真利氏には原稿作成にあたって有益な御助言を戴いた。以上記して謝意としたい。

文 献

- 1) B. S. Middleditch, "Practical Mass Spectrometry, A Contemporary Introduction," Plenum Press, New York (1979).
- 2) 日本質量分析学会データ集積部会：物性データ調査研究報告（科学技術庁委託）質量分析データ，Vol. 1 (1979), Vol. 2 (1980).
- 3) S. R. Heller and G. W. A. Milne, EPA/NIH Mass Spectral Data Base Vol. 1 ~ 4, NBS USA (1978).
- 4) J. W. Elchelberger, L. E. Harris, and W. L. Budde, *Anal. Chem.*, **47**, 995 (1975).
- 5) 小山保二・鈴木 功・前田浩五郎，日本質量分析学会第28回年会講演要旨集，42 (1980).
- 6) 小山保二・前田浩五郎，第3回情報化学討論会論文集，86 (1980).
- 7) 小山保二・鈴木 功・前田浩五郎，日本質量分析学会第29回年会講演要旨集，155 (1981).

Keywords

Personal data base
Mass spectral data base
Floppy disk system
BASIC language
Hydrocarbon