

報 文

有機化合物の構造解析におけるコンピューターの応用

——スペクトルデータと化学的データの併用——

大島時生, 石田嘉明[Ⓐ], 斉藤啓治*, 佐々木慎一**

(1980 年 6 月 21 日受理)

未知有機化合物の分子式と IR, ¹H-NMR, ¹³C-NMR を入力し, それら物性データに対応する構造を出力させるためのコンピュータープログラムにおける化学的データの利用法とその入力法について検討した. スペクトルデータを効率的に解析し, 構造式の出力数を少なく制限する方法を見いだすために, 各種化合物の構造決定に重要な役割を持つ化学的データの種類と内容を調べ, 続いて化学的データと部分構造との対応関係を調べた. 数値化の困難な化学的データをコンピューターへ入力する方法として, タイム・シェアリング・システム (TSS) によるコンピューターの対話機能を応用し, コンピューターとその利用者が質問応答を繰り返すことによって構造推定を行う方式を用いた. その結果, 多くの試料化合物において構造式の出力数は大幅に減少することが分かった.

1 緒 言

有機化合物のスペクトルデータをコンピューターに入力し, 解析するプログラムシステムについては, 既に幾つかの方法が提案されている^{1)~5)}. この場合, 正しい構造式を決して落とさないこと, 考えられる構造異性体をすべて出力することが重要である. 佐々木らはスペクトルによりできるだけ異性体を限定して出力する方法として, 複数のスペクトルを相補的に解析する方法の有効なことを示した¹⁾. 佐々木らの自動構造解析システム (CHEMICS) は, CHO から成る化合物の分子式, IR, ¹H-NMR, ¹³C-NMR データを入力し, 各データと対応する部分構造を決定した後, それを組み合わせることによって, 試料の構造として考えられるすべての平面構造式を出力する. このシステムの中で定義している部分構造又は原子団をコンポーネントと呼ぶ¹⁾²⁾ ので, 以後この用語を用いる.

他方, 試料化合物の分子式とその不飽和数が大きくなるに従い, 解析の結果出力される構造異性体数は急激に

増加する. そこで推定構造式の数を更に少なく制限する方法として, データ検索を併用する方法 (CHEMICS-F), 化学者の高度な解析で得られた部分構造を入力する方法 (CHEMICS-II) が有効であることを示した³⁾⁴⁾. これに対し, Cheer らは定性反応試験の応用が有効なことを示している⁴⁾. 又, Shelley らはコンピューターの対話機能を応用するシステムの有効性を示している⁵⁾.

CHEMICS の実用性を更に高めるためにコンピューターによるスペクトル解析は従前どおり行わせることとし, この外にその未知物質の構造に関する知見 (化学的データと呼ぶ) を対話方式でそう入する方法を検討した. その結果, 構造式の出力数を少なく制限でき, 十分な実用性を発揮させることが分かったので報告する. なお, この実用化システムを CHEMICS-UBE と呼ぶことにする.

2 化学的データの種類と範囲

化学的データの内容を一般的に定義することは困難であるが, ここでは CHEMICS へ入力するスペクトルデータ以外のもので, 試料の構造に関する情報のうちコンポーネントと結び付けることのできるものを化学的データということにする. 例えば, 各種の定性試験, 誘導体確認法などから導かれる官能基, 合成反応に関して多数

* 宇部興産(株)中央研究所: 山口県宇部市大字小串 1978-5

** 豊橋技術科学大学: 愛知県豊橋市天伯町字雲雀ヶ丘 1-1

の類似の反応から推論される骨格構造、置換基などは化学的データと言いうる。更に便宜上、MS, UV などを実験者が解析して得られる部分構造も広義の化学的データの中に含むことにする。

化学的データは数値化が困難な場合が多いため、コンピュータプログラムで扱える形にするには、得られた知見を整理して部分構造の形に表現し、それを適当な表記法でコード化する必要がある^{3)~5)}。プログラムはそのコード記号を受け取り、コンポーネントとの関係を解説する⁶⁾。しかし、化学的データの中には部分構造の形に表現できないものも多く、又実際の試料に対して、どの程度の化学的データを加えれば効率的な解析ができるか不明なことも多い。従ってスペクトル解析において、どのような構造情報が必要であるかを調べるのが重要となる。そこで各種化合物について炭素数、分岐鎖数、不飽和数などの項目と出力数との関係を調査した。

2.1 調査方法

調査した化合物はできるだけ同系統の化合物を集め、飽和炭化水素 11, モノオレフィン類 13, 飽和アルコール 24, 飽和ケトン 14, 飽和モノカルボン酸 7, 飽和エーテル, アルデヒド 7, ベンゼン置換体 12, シクロヘキサン誘導体 25, モノテルペン類 15 の合計 128 検体である。各化合物のスペクトルを UNIVAC-1100/41 大型コンピュータによる CHEMICS で解析した。この際、コンピュータの利用可能な範囲として CPU 時間で 60 秒までとし、これを越えた場合には実行を中断させ、その時点までの出力構造式を調べた。

2.2 結果と考察

Table 1 に 12 種類の飽和アルコールの試験結果を示

Table 1 The numbers of generated structures for saturated alcohols (> : greater than)

Compounds	Mol. form.	
	¹ H-NMR	¹ H-NMR ¹³ C-NMR
2, 2-Dimethyl-1-pentanol	19	2
2-Ethyl-1-hexanol	50	34
2, 2, 4-Trimethyl-1-pentanol	46	2
2, 6-Dimethyl-4-heptanol	38	12
5-Ethyl-2-heptanol	118	37
3, 5, 5-Trimethyl-1-hexanol	134	4
6-Ethyl-3-octanol	288	>31
2-Propyl-1-heptanol	288	>30
1, 1, 2-Triethyl-1-butanol	51	>36
Undecanol	317	262
2-Butyl-1-octanol	>1163	>22
6-Ethyl-3-decanol	>555	>23

した。これを見ると ¹H-NMR のみの場合には炭素数の増加に伴い出力数が増えること、¹³C-NMR を加えると側鎖がエチル基以上の長さを持つもので多数出力することが分かる。又、直鎖でも出力の中には分岐鎖を持つ異性体が現れる。この傾向はシクロヘキサン誘導体の場合にも見られ、その理由として ¹H-NMR ではメチレンの集まった幅の広い強いシグナルと、メチルとの区別が困難になること、¹³C-NMR ではメチレン炭素同士の化学シフトが近接し、帰属が困難となること、の二つが考えられる。他の脂肪族化合物及びアルキルベンゼンなどにおいても、アルキル鎖の部分については同様の傾向を示す。

ベンゼン置換体の例として、Fig. 1 にベンジルアルコールのスペクトルデータを入力した場合に出力される構造式を示した。これを見ると、IR と ¹H-NMR からベンジル基の -CH₂-以外に CH₂=, =CH-, >CH(OH)-, >CH(C=C)- のメチレンやメチンも可能性を否定できぬと判断されるため、架橋環 [9] や縮合環 [11] が出力されてくる。又、アレン (=C=), ケテン (=C=O) などのプロトンを持たないコンポーネントの有無は決定できないため [6]~[8] の構造式も出力される。他方、¹³C-NMR を加えても [3]~[5] の構造式を出力し、¹H と ¹³C-NMR の両者を用いても芳香環と炭素-炭素間二重結合とを区別できないことを示している。これらの傾向は、調査に用いたすべてのベンゼン置換体で見られた。

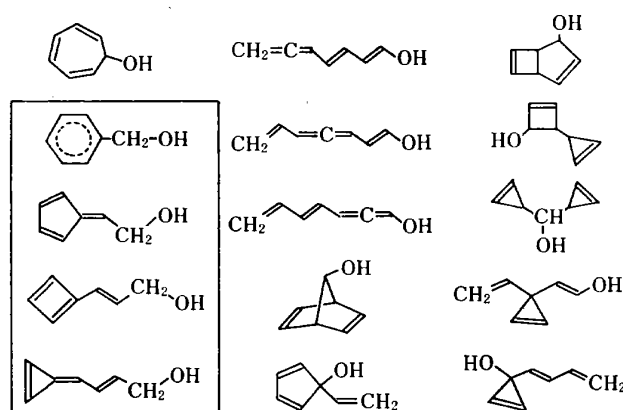


Fig. 1 Structures given for the IR, ¹H-NMR and ¹³C-NMR (enclosed) of benzylalcohol by CHEMICS

以上の調査の結果から、分析者が次のような限定条件を加えてコンピュータによるスペクトル解析を助けていけば、不要の出力数を大幅に制限できることが分かった。

(1) アルキル鎖に対し, 第三ブチル基, イソプロピル基, *gem*-ジメチル基, エチル基などの有無を決定し限定すること。

(2) 環状構造に対し, 環員数の範囲を限定すること。

(3) 芳香族化合物に対し, 単なる炭素-炭素二重結合の存在を否定, もしくは芳香環の存在を肯定すること。

(4) 水酸基に対し, アルコール性, フェノール性を決定し限定すること。

(5) カルボニル基に対し, カルボン酸, アルデヒド, エステル, ラクトンなどを決定し限定すること。

(6) エーテル結合型酸素に対し, ペルオキシ, メチレンジオキシ, アセタール又はオルトエステルなどの有無を決定し限定すること。

(7) 特に ^{13}C -NMR データがない場合, トロポロン, シクロプロペノン, ケテン, アレンなどの比較的特殊な構造に対し, あらかじめその有無を決定し限定すること。

そこで, これらの限定条件を Table 2 に示す質問メッセージの形に具体化することにした。

3 化学的データの処理法

CHEMICS-UBE では化学的データの入力方法に, コード表記法を用いず, コンピューターと実験者との簡単な質問応答を通してコンポーネントの有無を指定する方法を用いた。このために TSS を用いて, その対話機能を応用した。すなわち, コンピューター側から実験者に向かってスペクトルのみではその有無を決定できないコンポーネントの, 有無を尋ねる質問メッセージを出力し, これに対し実験者が, その存在を肯定する場合“YES”, 否定の場合“NO”, 不明の場合“UNKNOWN”のいずれかを返答データとして TSS 端末機から打ち込む。この場合, どのような質問メッセージを出力し, システムの中で返答データをどのように処理するかが重要となる。

3.1 質問項目の設定

プログラムで用意する質問メッセージの種類は, 入力しようとする化学的データと, それに対応するコンポーネントによって決定される。そこで, CHEMICS で設定している 189 種のコンポーネント¹⁾²⁾を, (I) 不飽和結合に関するもの, (II) 水酸基及びカルボニル基に関するもの, (III) アルキル鎖及びエーテル結合型酸素に関するもの, の三つに大分類し, 2.2 で述べた (1) から (7) の限定条件との対応を考えて, 27 項目の質問メッ

Table 2 The correlation of messages and “COMPONENT”s

Group	Message	“COMPONENT” number†
I	TROPOLONE	112
	METHYLENEPIOXO (CONNECTED WITH AROM. RING)	111
	CYCLOPROPENONE	115, 141
	FURAN	146
	KETENE	117, 143, 144, 175
	ALLENE	142, 145
	AROMATIC RING	114, 177~182
	C=C DOUBLE BOND	110, 118, 143, 189
	TRIPLE BOND	116, 176
	CYCLIC STRUCTURE	—
	PHENOLIC OH	113
II	ALCOHOLIC OH	126, 127
	ACETYL	35~40
	ACETOXYL	35
	FORMATE	119~124
	CARBOXYLIC ACID	128~133
	FORMYL	134~140
	ESTER OR LACTONE	147~152, 154~159
	TERT-BUTYL	1~6
III	ISO-PROPYL	19~25
	GEM.-DIMETHYL	7~12
	METHOXYL	26~31
	ETHYL	41~46
	R-O-O-R'	29, 119, 125, 148
	R-O-CH(R')-O-R''	47, 48, 50, 64, 66, 74~77
	R-O-CH(O-R')-O-R''	62, 63, 65, 68
	R-O-CH ₂ -O-R'	89

Group: (I) multiple bond, (II) hydroxyl and carbonyl, (III) alkyl and -O-; † See ref. 1), 2)

セージを設定した。

Table 2 に質問項目とコンポーネント番号¹⁾²⁾との対応関係を示した。

3.2 対話方式によるコンポーネントの処理

CHEMICS のスペクトル解析部では, 189 種のコンポーネントに対し, 分子式, IR, ^1H -NMR, ^{13}C -NMR の順に各データで否定されるものを取り除き, 最後まで残ったコンポーネント群を解析結果として構造式組み立て部へ渡す。解析結果は個々のコンポーネント番号とその個数範囲 {存在可能な最大 (MAX), 最小 (MIN) 値} とから成るリストとして示される。Fig. 2 に示すように, リストは分子式, IR, ^1H -NMR の解析後にも中間的なリスト (I)~(III) が作成されている。そこで, これらのリスト (I)~(III) に化学的データの入力用プログラムを接続し, 各リストの中に Table 2 で示したコンポーネントが残っている場合には, その質問メッセージを出力する。続いて返答データを受け取り, それが“NO”の場合には MAX を零とすることによってリストから取り除き, “UNKNOWN”の場合にはそのまま残す。又“YES”の場合には MIN を 1 とすることによって, そ

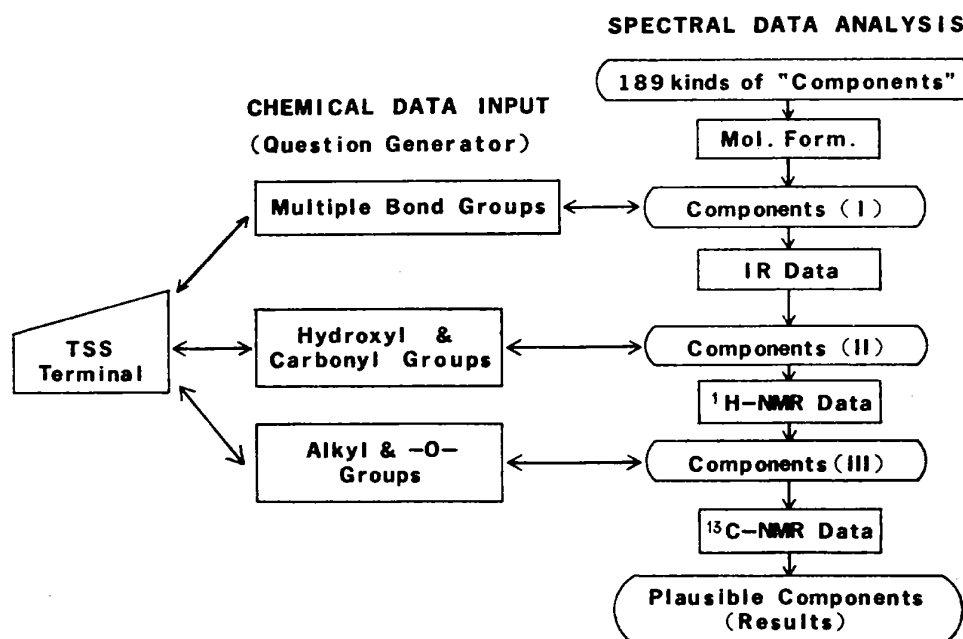


Fig. 2 Mode of interactive programs in CHEMICS-UBE

のコンポーネントが構造式の中に必ず含まれるように構造式組み立て部に指示する。ただし、化学的データよりスペクトルデータを優先させるために、その質問メッセージより後に続くスペクトル解析で否定された場合には“YES”は無効となる。なお、一度入力した返答データを変更したい場合には、リスト (I) にもどり、入力し直すことができる。

Table 2 に示したように質問メッセージは全体で 27 項目であるが、スペクトル解析で否定されたものは除くので、通常 (10~13) 項目のメッセージを出力する。Fig. 3 に質問応答の例を示した。

3.3 CHEMICS-UBE による構造解析例

微量のパラジウム触媒と硝酸ブチルの含まれている 1-ブタノールを 80 °C に長時間保った場合に生成する構造不明の化合物について、本システムを利用して構造解析を行った例を示す。この化合物の各スペクトルデータは Table 3 に示した。分子式 $C_8H_{16}O_2$ と 3 種のスペクトルによる解析では、最終的に 26 種のコンポーネントが残り 149 個の構造式を出力した (Fig. 4 左側)。他方、CHEMICS-UBE では、C=C、-OH 及び -O-O- の三つの質問に対し、“NO”を入力してそれらの存在を否定した (Fig. 4 右側)。これにより分子式で選ばれた 69 種のコンポーネントは C=C が無いことから 53, IR によって 29, -OH が無いことで 28, 1H -NMR によって 17, -O-O- 結合が無いことにより 12, 更に ^{13}C -NMR によって 10 種まで絞られ、これらを材料として

```

*** DATA ANALYSIS ***
DO YOU HAVE ANY INFORMATION IN NEXT SUB-STRUCTURE
FOR THE COMPOUND?
CYCLOPROPENONE ?
UNKNOWN
FURAN ?
UNKNOWN
KETENE ?
NO
ALLENE ?
NO
AROMATIC RING ?
UNKNOWN
C=C DOUBLE BOND ?
UNKNOWN
TRIPLE BOND ?
NO
CYCLIC STRUCTURE ?
YES
TYPE IN THE MAX. AND MIN. NUMBER OF
RING SIZE.
MAX MIN
8 5

```

Fig. 3 Example of the interactive mode

構造を組み立てさせると Fig. 5 に示す 13 個の構造式が出力された。この 13 個はこの段階では等しい確実度を持つものであるが、更に答を絞るとすれば MS のフラグメンテーション解析を考慮することにより、2-ブトキシテトラヒドロフラン (Fig. 5 の * 印) のみを最終結果として残すことができる。

その他の各種化合物について、スペクトルのみの場合と化学的データを併用した場合の出力構造式数を調査した結果を Table 4 に示す。いずれも出力数が大幅に減

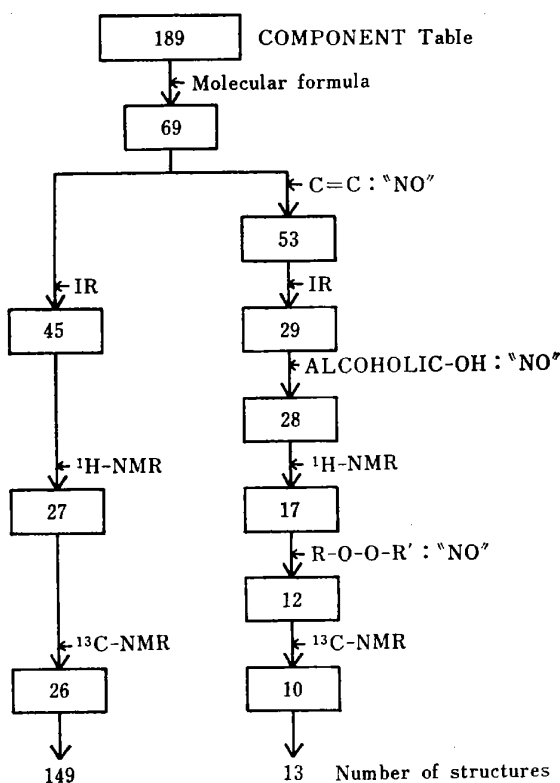


Fig. 4 Computation processes for an unknown compound derived from 1-butanol by CHEMICS (left) and CHEMICS-UBE (right). The number of components are expressed in enclosure

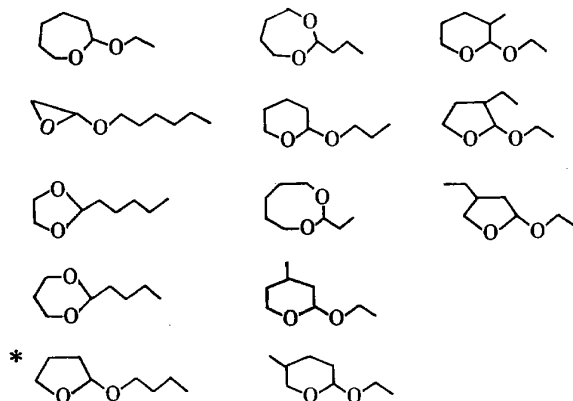


Fig. 5 Generated structures for an unknown compound derived from 1-butanol by CHEMICS-UBE

少しており, 本システムの方法の有効性を示しているものと言える。

3.4 対話方式の特徴

本システムで化学的データの入力法として採用した対話方式に対し, 化学的データをはじめから入力データに

Table 3 Spectral data of an unknown compound derived from 1-butanol

¹ H-NMR									
ppm	5.19	3.92	3.65	3.47	3.36	2.42	2.36	2.29	1.95
area	65	151	72	83	63	15	26	13	112
height	62	72	35	37	36	11	17	13	132
ppm	1.90	1.59	1.51	1.48	1.43	1.38	1.03	0.95	0.86
area	98	61	73	89	71	66	89	132	58
height	118	69	71	79	90	75	138	223	76
IR									
cm ⁻¹	2960	2935	2875	1705	1635	1465	1380	1275	1135
Intensity	81	76	69	25	28	43	34	27	60
¹³ C-NMR									
ppm	9.19	13.9	19.5	26.5	32.0	65.2	104.3		
Intensity	2824	7202	6150	4202	6197	6617	3073		

Table 4 The numbers of structures generated by CHEMICS-UBE for various compounds (> : greater than)

Compounds	S	S+C	Compounds	S	S+C
<i>o</i> -Cresol	>45	3	<i>l</i> -Menthol	>226	22
Guaiacol	>46	3	Borneol	>95	10
2,4-Xylenol	>231	6	Menthone	58	14
<i>o</i> -Allyloxyphenol	>12	6	α -Limonene	>24	5
<i>o</i> -Allyloxytoluene	>20	6	β -Pinene	>58	22
Cyclohexylpropionate	>37	18	Δ -3-Carene	>214	46
Cyclohexylbenzene	>91	1	Camphene	101	7
4-Cyclohexylphenol	>232	3	Camphor	>129	5
2-Cyclohexylcyclohexanone	147	3	<i>d</i> -Carvone	>13	9
4-Cyclohexylcyclohexanol	>761	7	Hinokitiol	>32	3

S: Spectral data only; S+C: Adding chemical data

加える一括処理方式がある。前者は TSS を利用し, 後者はバッチ処理形態で利用されるが, 本システムでは, コンピューターの演算時間やプログラム容量などの面で両方式間に大きな差は認められなかった。一括処理方式を対話方式に変更した場合, CPU 時間の増加は 1 秒以下であり, プログラムメモリーの増加は約 1 Kw 程度である (本システムは全体で 78 Kw に達する)。

他方, 対話方式はコンピューターの利用の面において, 下記のような利点がある。特に有機化合物の構造解析のような試行錯誤を要する問題に対し, コンピューターを活用する場合に有利である。

(1) TSS はコンピューターの利用者自身がプログラムを任意の時点に呼び出し, 結果を即座に端末機に出力できる (即時性)。

(2) 本システムはプログラムの実行や停止を自由に行え, 繰り返し実行も容易である (操作性)。

(3) 利用者はスペクトル解析部から出力するコンポーネントを直接追跡しながら, その中で不必要なものを

即座に削除することができる (解析経過のモニターリング)。

(4) 利用者はコンピューターから質問メッセージの出力された時点で、どんな化学的データが必要かを知ることができる (不足している情報の把握)。

上記 (3), (4) の特徴は一括処理方式では得られない利点であり、これらの特徴によって、構造解析の不得手な利用者にも容易に利用されるなど、実用性の増加することが分かった。

4 結 論

本研究の結果、有機化合物の構造解析において、構造異性体の範囲を制限するにはスペクトルのみの解析では必ずしも十分とは言えず、そこに不足しがちな構造情報の種類が明らかとなった。更にその構造情報を補う方法として化学的データに基づく幾つかの限定条件を加える方法の有効性を示すことができた。

又、コンピューターへの化学的データの入力方法として、コンピューターと実験者との対話方式を用いる方法の有効であることが明らかとなった。

文 献

- 1) S. Sasaki, H. Abe, Y. Kudo, T. Yamasaki : "CHEMICS; A Computer Program System for Structure Elucidation of Organic Compounds" in "ACS Symposium Series, No. 54, Computer-Assisted Structure Elucidation", D. H. Smith Ed., p. 108 (1977), (American Chemical Society, Washington D. C.).
- 2) S. Sasaki, H. Abe, Y. Hirota, Y. Ishida, Y. Kudo, S. Ochiai, K. Saito, T. Yamasaki : *J. Chem. Inf. Comp. Sci.*, **18**, 211 (1978).
- 3) S. Sasaki, T. Yamasaki : *Anal. Chim. Acta*, 印刷中.
- 4) C. J. Cheer, D. H. Smith, C. Djerassi, B. Tursch, J. C. Braekman, D. Doloze : *Tetrahedron*, **32**, 1807 (1976).
- 5) C. A. Shelly, H. B. Woodruff, C. R. Snelling,

R. E. Munk : "Interactive Structure Elucidation" in "ACS Symposium Series, No. 54, Computer-Assisted Structure Elucidation", D. H. Smith Ed., p. 108 (1977).

☆

Applications of a computer program for structure elucidation of organic compounds. Tokio OOSHIMA, Yoshiaki ISHIDA, Keiji SATO* and Shin-ichi SASAKI** (*Ube Industries Ltd., Central Research Laboratory, 1978-5, Kogushi, Ube-shi, Yamaguchi; **Toyohashi University of Technology, 1-1, Hibari-gaoka, Tenpaku-cho, Toyohashi-shi, Aichi)

A question-and-answer program has been added to the CHEMICS program system for structure elucidation of organic compounds containing C, H, and O. In the original CHEMICS the molecular formula, IR, ^1H -NMR, and C-13 NMR data are successively compared with the 189 components previously stored in the computer. It was found from the analyses of 128 compounds that the components with multiple bond group, hydroxyl, carbonyl, and/or ethereal oxygen group are sometimes difficult to be singled out by the input spectral data alone. To introduce a facility for input of appropriate information at the discretion of the user, the modified system has a software that can be used to put user's constraints on the results of automated analyses of spectral data. The user's judgement for the presence or absence of any component can be given to the computer through the TSS terminal. The usefulness of this dialogue program was proved by the computational analyses of an unknown compound derived from 1-butanol. A comparison is made between the number of structures generated from the new system and these from the original system for 20 organic compounds. As a result, it was found that the candidate structures could be reduced to only a small numbers in the new system.

(Received June 21, 1980)

Keywordphrases

computer program for structure elucidation; automated analysis of IR, ^1H -NMR, ^{13}C -NMR; man-machine interactive system for chemical information; determination of derivative of 1-butanol.