

# ライトペンによる化学構造式入力システムの試作

江崎 望\*・金岡 泰保\*\*・富田 真吾\*\*・岡田 敏彦\*\*

## Chemical Structure Input System by Computer Graphics

Nozomu EZAKI, Taiho KANAOKA, Shingo TOMITA, and Toshihiko OKADA

### Abstract

In this paper, we realize and examine an original chemical structure input system which make us perform input of the synthetic objective, "target molecule", which is the first step in computer-assisted synthetic analysis.

We use a color graphics display (NEC computer terminal, N6940 color graphics display) to realize this system, and input of the chemical structure is performed by the light-pen. Therefore, it is possible to treat quickly and easily for a person who is unaccustomed to use the teletype of computer.

### 1. まえがき

近年, 多数の専門的な知識や経験則が組織的に蓄えられたコンピュータと対話しながら専門家の発想を刺激しようとする, いわゆる知識ベースシステムが新しいコンピュータ利用法として注目されている。

その一つの応用分野として, 豊かな経験と模大な知識が要求される有機合成デザインがある。化学者 E. J. Corey はコンピュータを積極的に導入して, 知識ベースシステムの手法を基に, プロスタグラジンの合成経路探索に画期的な成果を収めている<sup>1-6)</sup>。

既に Corey 等は, コンピュータを用いる有機合成デザイン支援システムとして OCSS (Organic Chem-

ical Simulation of Synthesis)<sup>2)</sup> や LHASA (Logic and Heuristics Applied to Synthetic Analysis)<sup>3)</sup> を作成し, その概略を発表している (Fig. 1 参照) が, 詳細は公表されていない。

本論文は, 独自の有機合成デザイン支援システムの開発を目的として作成した構造式入力システムについて述べている。

本システムへの入力, ACOS-6TSS 端末を介したカラーグラフィックディスプレイ (NEC N6940) 上にライトペンで構造式を修正しながら描くことによって容易に行なうことができる。

### 2. 分子構造式入力システム

化学者が, コンピュータを用いて合成操作を行なうためには, まず, 化学者がコンピュータのノウハウを詳しく学ばなくてもよいように, 日頃から使い慣れた表現を使えるようにすることが望ましい。このため化学者にとって最もなじみ深い言語である化学構造式を用いることによってコンピュータとの間の情報伝達を行なわせ, しかも, コンピュータの操作法を容易にすることが本システム作成の目的である。

本システムは山口大学 計算機センターの ACOS-6 の制御のもとで, N-6940 カラーグラフィック端末を用いることにより動作するが, 本システムの大きな特徴は, プログラムへのすべての入力が, 端末付属のラ

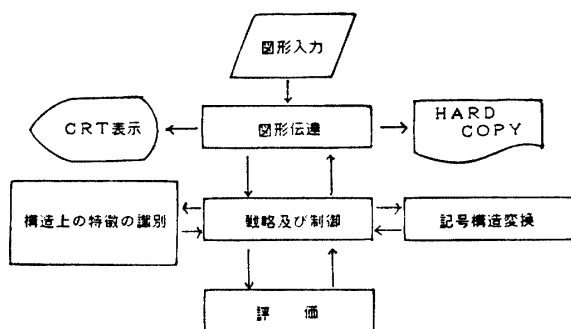


Fig. 1 Functional subdivisions of LHASA

\* 電子工学科 (現, 大学院電子工学専攻)

\*\* 電子工学科

イトペンによって行なわれることである。

Fig. 2 に本システム全体の流れを示す。

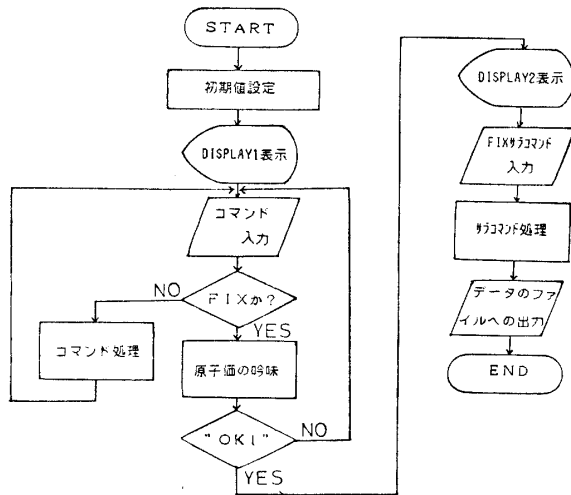


Fig. 2 Flow chart of Input system

## 2.1 構造式表示記号体系

本システムにおいて化学構造式を描写する際に用いる記号としては、原子を表示する“C”，“H”，“N”，“O”，“S”，“P”及びハロゲン元素を表わす“X”がある。また電荷を表示する“+”及び“-”，ラジカル（不対電子）を表示するために“/”という記号を用いた。

化学結合としては単結合，二重結合，三重結合及び点線（立体化学的配置を示す）がある。

構造式はグラフとして表示され，節は原子を，また節と節を結ぶ枝は化学結合を示す。通常無表示の節は炭素原子を表し，その他の原子は適当な記号を節上に書くことにより表される。水素原子は通常描く必要はない。

## 2.2 構造式入力方法の概略

本システムを起動した際，まず Fig. 3 に示したような表示が CRT 上に現われる。(DISPLAY 1)

ここで中央の四角形の領域を“BOX”，また各コマンド (DRAW, ERASE, etc.) にそれぞれ付けられた小さな四角形を“BUTTON”と呼ぶことにする。入力される化学構造式はこの“BOX”内において描かれることになる。

また構造式を入力するためにはコマンドを用いなければならないが，コマンドを作動するには，それに付けられた“BUTTON”にライトペンのペン先をあてればよい。ペン先を“BUTTON”に押し当てることに

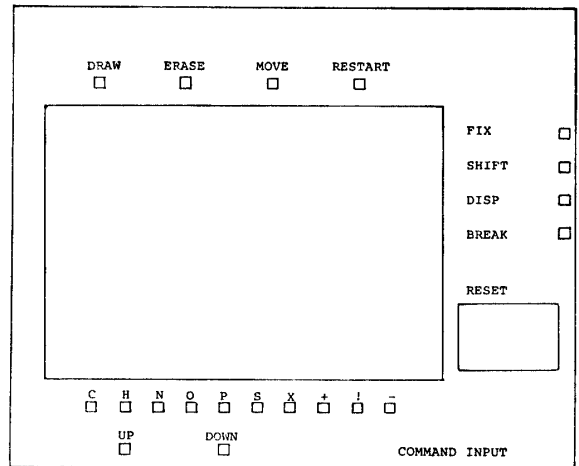


Fig. 3 DISPLAY 1

より，CRT 上の“BUTTON”の色は黄から赤に変化し，そのコマンドの制御下におかれたことを示す。但しこのコマンドの選択を行なえるのは，Fig. 3に見られる“COMMAND INPUT”というメッセージが現われている時のみに限られ，ある“BUTTON”の色が赤になっているときは“COMMAND INPUT”の表示は消されている。そこでコマンド入力待ちの状態（“COMMAND INPUT”表示時）に戻りたい時には Fig. 3に見られる“RESET”下の枠内（“RESET BUTTON”）にペン先を当てる必要がある。この操作を行なうことにより，今まで制御を受けついたコマンドの“BUTTON”の色は赤から黄に戻り，“COMMAND INPUT”の表示が再び現われ入力待ちとなる。

## 2.3 システムコマンドの解説

Fig. 3 に表されているように DISPLAY 1 では，構造式描写の際用いるコマンドとして DRAW, ERASE, MOVE, RESTART, の4つと，他に SHIFT, DISP, FIX, UP, DOWN, BREAK, RESET の7つの合計11のコマンドを設定した。

**2.3.1 DRAW** このコマンドは化学構造式を描写する際に必ず用いるもので，化学構造式の結合を引いたり，節上に特定の原子記号を置いたりする操作を行なう。

まず，結合を表示するには，“BOX”内に描こうとする結合の二つの節に相当する位置にラインペン押し当てればよい。そうすることによって二点間に実線が引かれ結合が存在していることを示す。多重結合を表したい場合には，すでに結合が描かれている二つの節上にもう一度ペンを当てることによって二重，三重

結合が示されるが、三重結合の上に更に結合を加えるとその結合は単結合に戻るようにした。また結合を描く際、一方、あるいは両方の節上に特定の原子記号や、電荷などを置きたいという場合には、“BOX”内をペンで指す前に、描こうとする原子記号を“BOX”下の原子記号群の中から選び、それにつけられた“BUTTON”をペンで指し、続いて“BOX”内の節の位置をペンで指せばよい。

**2.3.2 ERASE** このコマンドは誤まって描かれた結合や原子を消そうとするときに用いるもので、任意の結合を消したい場合には、その結合の両節をペンで指せばよい。しかし、多重結合の場合消される結合は一本だけである。

また原子を消したい場合には、その原子上をペンで2度指せば、その原子自体とそれに関わっているすべての結合が消される。これらの操作の結果、孤立した原子が残る場合が生じるが、そのときにはその原子も同時に消される。

**2.3.3 MOVE** このコマンドは、構造式の形を整えたりするときに用いるもので、“BOX”内に描かれている図形上の一つの節をライトペンで指し、次に“BOX”内の任意の点を指すと、最初ペンを当てる節上の原子はその点に移され、その原子が関わっているすべての結合も移される。

**2.3.4 RESTART** 入力していた構造式を他のものに代えたいときや、図形がひどく乱れたため最初から入力し直そうとするときに用いるもので、このコマンドの“BUTTON”をペンで指すと CRT 上のすべての表示が消され、Fig. 3 の DISPLAY 1 が再び現われる。そしてコンピュータ内部の情報もすべて消され、完全に初期化される。

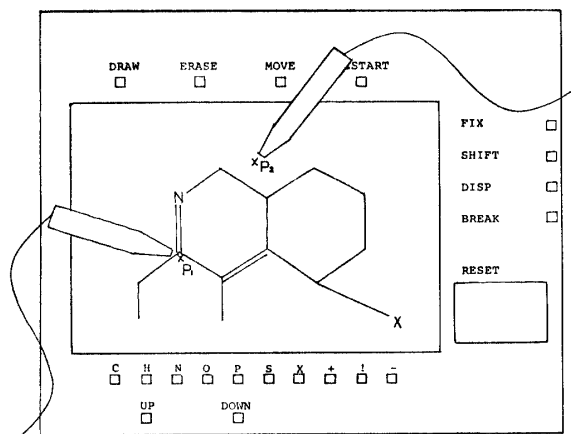


Fig. 4(a) Example of an operation at “SHIFT”

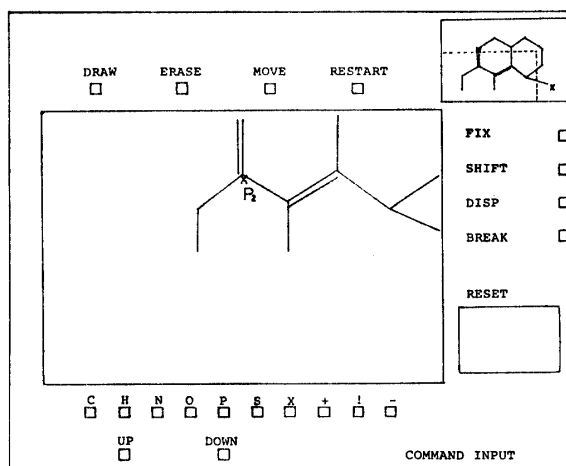


Fig. 4(b) Display after the figure was shifted

**2.3.5 SHIFT** 構造式を入力中、“BOX”内が構造式で満たされてしまい、これ以上“BOX”内に入力出来なくなってしまった時などに用いるコマンドで、図形を、それ自体の形状は保持したまま、任意の位置に移すものである。その例を Fig. 4 に示す。

Fig. 4(a) に示すように“BOX”内のある一つの節の位置  $P_1$  をペンで指し、次に“BOX”内の任意の点  $P_2$  をペンで指すと、同図(b)のように  $P_1$  にあった節は  $P_2$  に移り、他の節や結合も平行移動する。

このとき今まで描かれていた図形が“BOX”の隅などに移されると図形が“BOX”からはみ出してしまう場合が生じる。このため常に入力図形のすべてが見れるように、CRT 内の右上に小さな四角形の領域“Small BOX”を設け、その中に入力図形を表示するようにした。そして現在“Small BOX”内のどの部分が“BOX”内に表示されているかを示すために、その領域を点線で囲むようにしている。(Fig. 4(b) 参照)

この操作によって入力分子の図形は“BOX”内の大きさに限らず、無限に大きく描けるため、理論的にはどのような高分子の構造式でも入力できるようになった。(現在は、記憶領域の関係上原子数500の分子まで入力可能である。)

**2.3.6 DISP** このコマンドは SHIFT の操作で見えなくなった部分を含めて、今まで入力してある分子構造式を“BOX”内に表示しようとするものである。このコマンドは単に入力図形全体を眺めるためのものであって、このコマンドによって表示した全入力図形に DRAW, ERASE などの操作は加えられない。それ故“RESET BUTTON”を指すとこの DISP 画面以前の表示にもどるようになっている。

**2.3.7 DOWN** このシステムでは有機化合物を、構造式という二次元的な表現で入力するために、立体化学的構造を示すには特別な措置が必要となる。このコマンドはそのために設けられたもので、立体化学的構造を表示するためには、その構造式中の一つの結合の両端の節をペンで指せば、その結合は実線から点線となり立体構造をなしていることを示す。但し、この時、節のどちらか一方は炭素原子で、その原子を不斉原子と考えている。またその結合は単結合でなければならない。

**2.3.8 UP** このコマンドは前述の DOWN コマンドで点線表示となった結合をもとの実線に戻すもので、立体構造であるというメモリー上の情報も同時に消されるようになっている。

**2.3.9 BREAK** このコマンドを用いることによって、プログラムの行程は中断され続行不可能となり、状態は ACOS-6 FORTRAN のビルド・モードとなる。このため再開するためには、もう一度プログラムを起動し直す必要がある。

**2.3.10 FIX** これについては2.4で詳しく述べる。

**2.3.11 RESET** 前述のコマンド群の制御下から脱するには一旦この“RESET BUTTON”をペンで指す必要がある。このことによってシステムはコマンド入力待ちの状態となり、管面上に“COMMAND INPUT”の表示が出される。

**2.3.12 ATOMS “BOX”** 下の原子記号群は、2.3.1で述べたように DRAW の最中で用いる他に、それ自体を単独で扱うこともできる。すなわち“COMMAND INPUT”表示時に、特定の原子記号の“BUTTON”をペンで指し、続いて“BOX”内の任意の位置をペンで当てると、その位置に原子記号が表示され、再び“COMMAND INPUT”の表示が現れる。また電荷を表示する“+”、“-”、“/”なども同様に用いることが出来るが、これらは、“BOX”内にすでに構造式中の一つの原子として描かれている節上にしか描くことが出来ないようになっている。

## 2.4 FIX 及びサブコマンドの解説

入力図形をすべて描き終えたならば、そのデータをコンピュータで計算し易い形に変換するため、一旦ファイル上に出力する必要がある。そこでこのコマンドを用いるわけであるが、“BUTTON”をライトペンで指すと、最初に、入力されている化学構造式の各原子の原子価の妥当性がチェックされる。すなわち、もしある原子の原子価が規定値を超過すると、CRT 上の“VALENCE CHECK”の表示に続いて“EXCEED-

ED”というメッセージとその超過した原子記号が表示される。そしてどの原子が超過しているかを示すために“Small BOX”内の構造式中の原子記号の色が青から白に変化する。ここで原子価の規定値はそれぞれ C=4, H=1, N=3, O=2, P=3, S=2, X=1 に設定している。この例を Fig. 5 に示すが、この場合では N が規定値を超過して結合していることを示している。

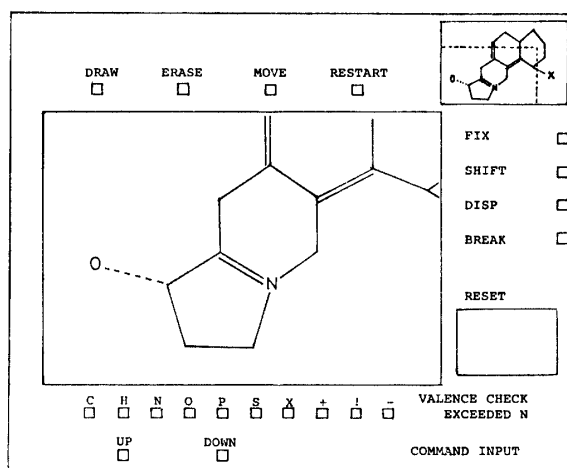


Fig. 5 “VALENCE CHECK”

このメッセージが出力されたのち、システムはコマンド入力待ちの状態になるので、ここで超過している原子の修正を行なわなければならない。修正が完了し、もう一度“FIX”の“BUTTON”を指し、すべての原子の原子価の妥当性が認められると、CRT 上は、“OK!”のメッセージが現われ、DISPLAY 2 に画面表示が代わる。(Fig. 6)

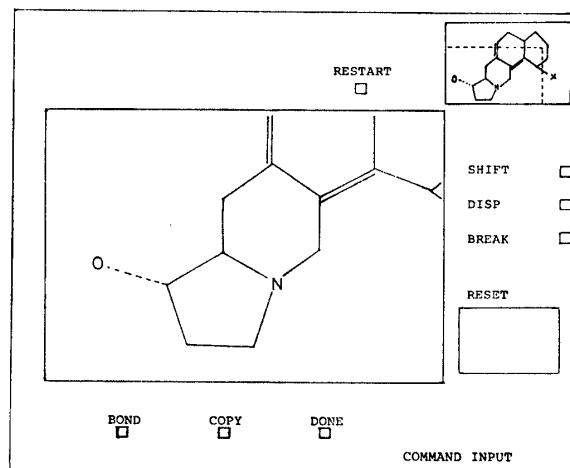


Fig. 6 DISPLAY 2

ここで新たに BOND, COPY, DONE などの文字が表示されているが、これらを FIX サブコマンドと呼ぶ。

**2.4.1 BOND** 化学者が標的分子から次の中間生成物を得ようとするとき、化学反応箇所(戦略箇所)を指定する際に用いるもので、構造式中で反応させたい結合部分の両端の節をペンで指すことにより、指定がなされたという情報を蓄えることができる。このとき、その指定部分は赤から白に色が変化する。

もし一度指定した戦略箇所の取り消しを行ないたい場合には、もう一度その両端の節をペンで指せばよく、その部分の色は元に戻る。

**2.4.2 COPY** 最終的な入力分子のハードコピーを取るために用いる。ハードコピー装置としては、計算機センターの N6943-01 を用いているが、この装置で DISPLAY 2 自体のハードコピーを取ると、“BOX”内が緑色で塗りつぶされているため、このままでは“BOX”内の構造式はハードコピーに現われてこない。そのため新たにコピー用の画面として DISPLAY 3 を表示させ、その画面のハードコピーを取るようにした。そのハードコピー図を Fig. 7 に示す。

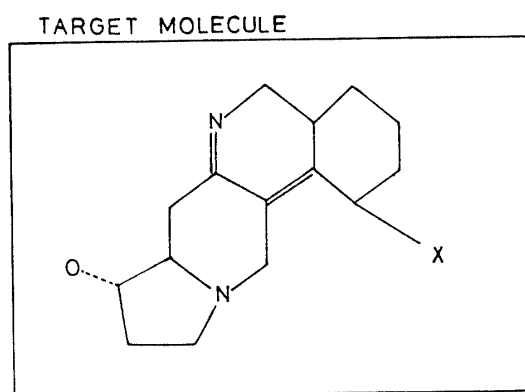


Fig. 7 DISPLAY 3

この COPY コマンドは次の DONE コマンドの働きも含むため、このコマンドを用いることによっても構造式の入力操作を完了することができる。

**2.4.3 DONE** このコマンドを用いることにより構造式の入力操作はすべて完了する。

“BOX”内に描かれた構造式は、そのデータをコンピュータ内部で計算しやすい形に変換する必要があるため、このコマンドによってすべてファイル上に転送する。

**2.4.4 RESTART** このコマンドは今まで入力された分子構造式を他のものに代えたい時などに用いる

もので、この“BUTTON”を指すと画面は DISPLAY 2 から DISPLAY 1 に変わり、すべての情報は消去され、初期化し直される。

### 3. 入力分子データのファイルへの転送

入力分子はグラフで表わされているため、「グラフ理論」を用いれば、例えば見かけは異なっている二つの図形が同じ構造であるかどうかの判定や、ある構造を部分として含むような(親)構造の探索、環の認定、環数の決定、官能基の認定などなどの処理が可能となる。そのためグラフをコンピュータで処理し、上述のような処理を行なわせるには、まずグラフをコンピュータに入力できる表現法に直さなければならない。

そのひとつの形式として考え出されたものに結合表<sup>7)</sup>と呼ばれものがある。

この結合表を作成するためには、入力されたデータをひとまず、ファイルに転送しなければならない。本システムでは COPY 及び DONE コマンドによってファイルへの出力を行なうが、その際一つの構造式に対して5種類のデータの出力を行なう。それらは以下のようなものである。

データ 1…それぞれの節がどの節と結合関係をもつかを示す。

データ 2…各節の原子記号、荷電状態など

データ 3…各結合の多重度

データ 4…各節の表示上の座標データ

データ 5…立体配置をなす結合及び戦略箇所情報

### 5. 分子構造式出力システム

前述のような操作でデータをファイルに出力し、結合表を得たならば、その結合表は構造上の特徴の識別や戦略などに活用され、その分子を表わす一意の表現としてコンピュータ内部に蓄えられる。ここでは化学者が現在コンピュータ内部に蓄えられている分子がどのようなものかを一見してわかるようにするために、この結合表を利用することにより、その分子を CRT 上に化学構造式として表わすことのできるシステムの作成に関して述べる。

このシステムは、まず入力データとして、4つのファイルを設定する必要がある。その4つのファイルからそれぞれ表示に必要なデータを得るわけであるが、その内容は、まず第1のファイルからは、結合をなす節番号とその結合の多重度、第2のファイルからは各節の原子記号、荷電状態などである。第3のファイル

からは各節の表示上の座標、第4のファイルからは立体化学的構造と戦略箇所のデータを得ることになる。

これらのデータから構造式を描くわけであるが、構造式の記号体系は入力のものであり、無表示の節には炭素原子が、また立体化学的構造は点線で示され、戦略箇所は白色表示となる。この例を Fig. 8 に示すが、本システムを起動すると“SAVED MOLECULE”の文字と構造式が CRT 上に表示され、更に構造式の各節に番号が付けられる。この番号は任意的なものではなく、コンピュータが一定の規則に従って計算した「絶対的」な、すなわち構造式と結合表とが厳密に1対1の対応関係を示すようになる番号である。

この出力システムの表示画面を DISPLAY 4 と呼ぶことにする。

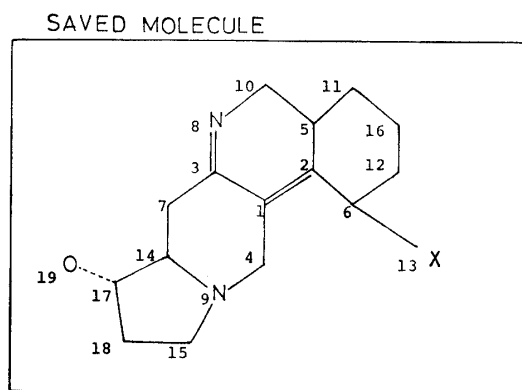


Fig. 8 DISPLAY 4

## 6. む す び

今回作成した入力システムによって、化学者が合成操作を開始するにあたって最初に行なう合成目標分子（標的分子）のデータの入力を非常に容易に、しかも素早く行なうことが出来るようになったと思われる。例えば原子数約30の分子の構造式を“BOX”内に描こうとするとき通常5～10分の時間ですべて操作を終えることが可能となる。このことはライトペンを用いていることに大きく起因しており、コンピュータ端末のキーボード操作に不慣れた場合でも、図形をコンピュータの CRT 上に描くという作業が容易に行なえる。

更に本システムの大きな特徴として、理論的には、どのような高分子の構造式でも入力可能であるということが挙げられる。これは、入力操作が CRT 上の“BOX”内だけに限らず、“SHIFT”の操作によって自由に拡大することが出来るからである。通常

Vol. 35 No. 1 (1984)

“BOX”内だけならば、せいぜい描くとしても原子数50ぐらいの構造式が限界であろう。それを本システムでは表示領域は“BOX”内に過ぎないが、入力可能領域では無限の拡がりを見ることが出来、時間さえかければどのような高分子でも入力可能である。

以上述べたように本システムにより、標的分子の入力はかなり容易になったと思われるが、分子構造の表現に不十分と思われる点がある。その一つが異性体に関する表現である。異性体の一つである幾何異性体 (cis-trans 異性体) の記述に関しては構造式の描写によって表すのみで、特に異性体情報は明記するようになっていない。また立体化学に関する表現法においても十分に満足のいくものとは言い難い。本システムでは不斉炭素原子を中心とする立体的配置の有無を表記するに止まり、鏡像異性体、ジアステレオ異性体などの立体異性体情報を蓄えることは出来ない。

その他の問題点としては2.4で述べているように、原子の原子価を一意に決めてしまい、特別な原子価をもつ場合を考慮していない点である。このことはコンピュータで処理を行なうには非常に規則性に乏しいものであり、解決が困難なものの一つといえる。

なお、化学反応データベース作成にも有用であると思われる、本システムを利用した構造式の固有表現化システムの作成も行っており、これについては別途述べる予定である。

謝 辞 日頃ご指導ご助言頂く本学高浪五男教授、井上克司助教授ならびに有益なるご意見を頂いた福永公寿博士に感謝する。又、本研究に際しご指導ご助言頂いた東京都臨床医学総合研究所神沼二真博士に感謝の意を表する。

## 参 考 文 献

- 1) 湊宏：“有機合成デザイン”講談社 (1977) p.1
- 2) 大野雅二，他：“コンピュータを利用する複雑な有機化合物の合成デザイン”南江堂 26 No.7~9 (1972)
- 3) E.J. Corey, W. Todd Wipke, Richard D. Cramer III, and W. Jeffrey Howe: “Computer-Assisted Synthetic Analysis. Facile Man-Machine Communication of Chemical Structure by Interactive Computer Graphics” *Journal of the American Chemical Society* 94, p.421 (1972)
- 4) E.J. Corey, W. Todd Wipke, Richard D. Cramer III, and W. Jeffrey Howe: “Tech-

- niques for Perception by a Computer of Synthetically Significant Structural Features in Complex Molecules”  
Journal of the American Chemical Society  
94, 431 (1972)
- 5) E.J. Corey, Richard D. Cramer III, and W. Jeffrey Howe : “Computer-Assisted Synthetic Analysis for Complex Molecules. Methods and Procedures for Machine Generation of Synthetic Intermediates”  
Journal of the American Chemical Society  
94, 440 (1972)
- 6) E.J. Corey, H.W. Orf, and David A. Pensak: “Computer-Assisted Synthetic Analysis. The Identification and Protection of Interfering Synthetic Intermediates”  
Journal of the American Chemical Society  
98, 210 (1976)

(昭和59年4月14日 受理)