

## 分子設計支援システムにおけるエディタと3次元表示

高橋一重・酒井真理子<sup>\*</sup>・可児利佳子・恒川尚

(株)東芝 総合研究所

近年計算機の発展により分子データベース、多方面の化学のアプリケーションソフトがつくられ、計算機は化学者の研究・開発に不可欠のものとなりつつある。

しかし分子データの入力には化学者にとっては気苦勞の多いものである。一方分子構造式から手軽に3次元モデルが生成でき、しかも3次元モデルをグラフィックディスプレイ上に表現できれば、化学者のイメージーションをおおいに刺激することができ

る。  
本報告ではコンピュータによる分子設計支援システムのうちとくに分子構造式の計算機への入力をスキャナから自動的に行うことを考慮したエディタと、分子の属性の抽出・推定をプリミティブの接続関係から判定すること、および3次元モデルを分子構造式から対話的に作成して分子構造式との対応づけをしながら3次元分子モデルの編集を行う3次元構造エディタについて述べる。

### On Editor and 3-dimensional Model of Molecular Design Support Systems

Kazushige TAKAHASHI, Mariko MAEDA, Rikako KANI, Shou TSUNEKAWA

Research and Development Center, TOSHIBA Corporation  
1, Komukaitoshibacho, Saiwai, Kawasaki, Kanagawa, 210, Japan

Recently molecular designer uses a computer system for the development of a molecular data base and a lot of useful chemical application software.

It is, however, time consuming to input chemical structure data into a computer system and convert them into 3-dimensional molecular models.

In this report, we propose a new chemical structure editor for an automatic reading and an interactive construction method for constructing 3-dimensional molecular model using primitives.

\* 88年3月まで在職

## 1 はじめに

特定の性質をもつ化合物の研究・開発をする際には、汎用の分子データベースを使用する以外に、個人専用の分子データベースを作成する必要がある。ここで時間を費やすのはデータ入力である。この労力を削減するために、図面読み取り技術を用いてスキャナで紙の上に書かれた分子構造式等を入力するという方法がある（文献1参考）。しかし現在では100%認識することは不可能であるので、自動読み取りの場合にも編集用のエディタが必要である。そこで自動読み取りのためにも使用することを考慮した分子構造式エディタを作成した。

また、普通の分子構造式はデータベースの中では分子のテーブル表現（分子を構成する原子の接続情報）をしているものが大部分である。一方市販のアプリケーションソフトの入力データとしては分子のテーブル表現データの他に、部分構造・化学的特徴等が必要になる。このため、分子のテーブル表現データを入力としてこれらの情報を生成する属性推定・抽出を行う部分を試作した。

ここでは、3次元データをただ表示するのではなく、分子の組み立てをする部分や分子の内部構造を活かした表示をするために、入力する分子構造式データを部分構造に分解し、その結合を対話的に決定することにより、分子の3次元モデルを生成する。その際、分子構造式の原子と3次元モデル分子の原子との対応づけを行う。この対応づけによって3次元表示の指定法が非常に使いやすいものになると期待される。

## 2 分子構造式入力エディタ

自動読み取りへの移行を考慮した分子構造式入力エディタの設計方針について述べる。

自動読み取りでは分子構造式のデータは図形的な意味しかもっておらず、化学的な意味は新たに解釈せねばならない。一方、エディタ入力の場合には、入力情報にある程度の化学の意味を持たせることは可能であるが、全意味を持たせた入力はきわめて煩雑になる。むしろ入力時には、化学的な意味をもたせず、操作性の良いエディタの入力部を作成し、化学的な解釈をするための変換部と解析部を付加したエディタとした。

データベース作成ツールとしては操作性に重点をおき、例えば同種類の分子をたくさん入力する場合の効率向上に注意した。またアプリケーション用の入力ツールとしては変換部・解析部を通して3次元座標の自動生成にもつながら、後述する3次元構造エディタとともに用いることができる。

入力部は作図・編集をする部分であり、ワークステーション上の図形編集機能をツールとしてもちいる。変換部では作図した分子データを原子テーブルとボンドテーブルからなる分子のテーブル表現データに変換する。解析部はテーブルデータをグラフィ的に解析することによって、図形的あるいは化学的な特徴を抽出して特徴データを求める。

エディタの概念図は図1にあるように、セグメントと呼ぶ単位ごとに編集する分子データを持つことが出来る。それぞれのセグメントごとにグラフィックディスプレイ上でコピー、移動等を行うことによって入力編集作業を行う。また、自動読み取りによりセグメントに分子データを入力することもできる。

以下入力・変換・解析の3つの部分について説明する。

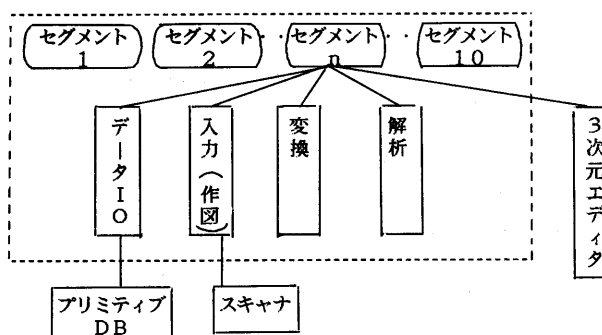


図1 エディタ概念図

## 2. 1 入力部

操作性が重要な入力部は、マウスでポップアップメニューを選択する形式にした。作図用プリミティブ（一部を図2に示す）をセグメントとして組み立てることにより、目的の分子構造式を構成する。似た分子構造式を多数入力する場合はセグメント間のコピーにより効率よく入力することができる。このプリミティブの組立は後に変換部が処理するので、プリミティブ同士をもっともらしい場所に配置するだけでよい。したがって、入力部で使用する作図用プリミティブはシンボル（官能基または原子の記号）と線分（ボンド）の位置座標データのみで構成される。また、化学的意味をもつプリミティブを使用することにより、データベースを構築する際コードを発生させやすく、作業能率をよくするという利点がある。

図3に入力エディタとしての基本的な機能をあげる。この図における消書・領域操作・チェックは変換と解析の結果を用いることにより行う。たとえば領域操作により、化学的にひとまとまりのもの（ベンゼン環等）をまとめて処理することができる。

またこの入力部におけるエディタは3次元構造エディタと並行して使用することができる。したがって入力部で作成した構造式を表示し、さらにその3次元座標を生成して3次元表示した両者を見ながら構造式または3次元構造の修正を行える。

|        |                                                                                                                                  |
|--------|----------------------------------------------------------------------------------------------------------------------------------|
| 原子     | C, H, O, N, S, . . .                                                                                                             |
| ボンド    | —    =    ≡                                                                                                                      |
| プリミティブ | COOH, OH, NH <sub>2</sub> , CH <sub>3</sub><br> |

図2 作図用プリミティブ

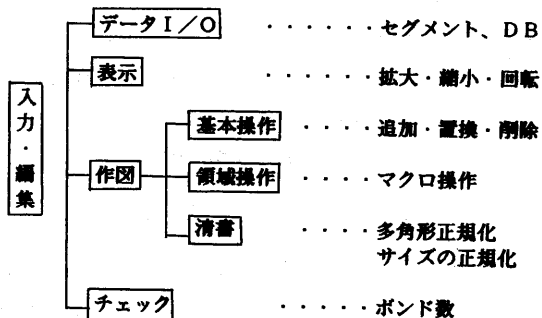


図3 エディタの機能

## 2. 2 変換部

変換部では入力部で生成したデータを、原子テーブルとボンドテーブルから構成される分子のテーブル表現データに変換する。この入力を自動読み取りで行う場合は、図4に示す手順で実行される。スキャナから入力した場合には、入力された画像から文字を構成する線を判別し、文字部分を抽出、認識し、残りの線を結合線としてベクトル化する。変換アルゴリズムの概要を図5に示す。原理的にはシンボルと線分の位置関係のみにより接続関係を決定する。たとえば2本の線分の両端点が近い場所にあるならば、

この2本の線分は2重結合を表していると解釈できる。また線分の端点の近くに他の線分の端点やシンボルがない場合にはそこに、炭素原子があるものと解釈する。

官能基によっては（例えば  $-\text{CONH}-$ ）2種類以上のボンドが同じ官能基の別々の原子と結合する場合もある。このため官能基のボンドとの結合情報からあらかじめシンボルデータベースを作成し、それを参照しながらテーブルの展開を行う。処理の途中ではシンボルと線分の位置座標を用いるが、最終的な分子のテーブル表現データは位置情報をもたない。

しかし図6のように、特別な意味をもつプリミティブは、テーブル形式のデータ構造では十分に表現できず、さらに階層的な分子のデータ構造が必要である。しかし、この単純データ構造に基づくアルゴリズムだけでも、ほとんどの分子を認識することができる。

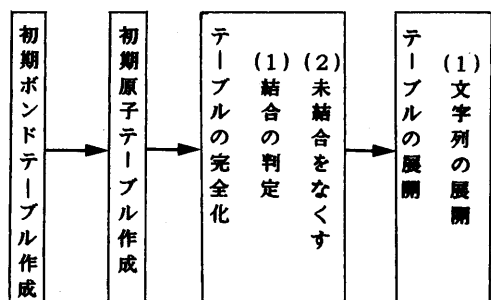


図5 変換アルゴリズム

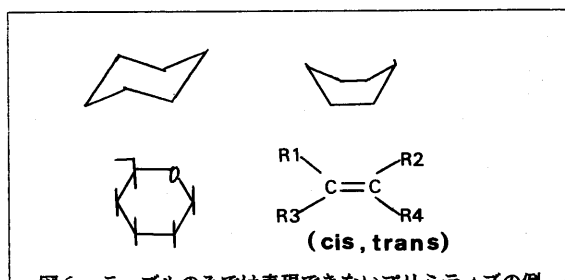


図6 テーブルのみでは表現できないプリミティブの例

### 2.3 解析部

解析部では、変換部で作成した分子のテーブル表現データから以下に述べる特徴データを求める。この際、入力部と解析部との独立性を保つために、入力部でのシンボルデータ（官能基）を利用しない。

この特徴データは、分子の部分構造の情報とその結合情報と位相的な情報が含まれる基本特徴データと、この基本特徴データをもとにしてさらに化学的な知識から属性推定をしてつくる属性データから構成される。

#### 2.3.1 基本特徴データの抽出

基本的な処理としては次の順序で行う。

- (1) 始めにテーブルデータのグラフ的な探索により、位相的な解析をする。このことにより、分岐情報とループ情報（ベンゼン環等）が得られる。
- (2) WLN記号による化合物の命名は計算機処理に適した方法である。しかし、無駄な部分が多いためこのWLN記述の一部を用いてアレンジし、部分構造記述テーブルを作成する。
- (3) 基本的な官能基を含むかどうかをチェックする。
- (4) 部分構造のはりあわせを行い分子構造の大まかな分類を行う。

(4) の出力データの概念図は図7に示したように基本的なプリミティブの接続情報である。また(2)の部分構造記号テーブルは 分子構造式の特徴をよく表しているの(3)の官能基のマッチングによる官能基の判定にもちいる。

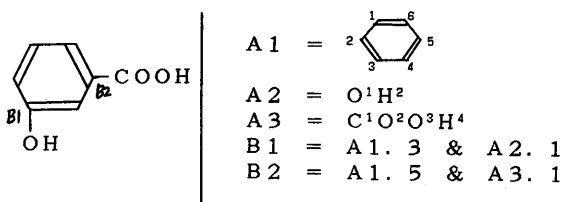


図7 解析部の出力の例

#### 2.3.2 属性データの推定

今回の基本的な試作においては化合物全般の知識に基づいた化合物の属性データの推定を行った。

入力データは解析部の前半で作成した基本特徴データで、プリミティブの接続情報データと部分構造記号テーブルのみである。内部処理は手続きにおこなった。

方法は官能基の接続関係に基づいて属性を抽出するものである。例えば同じ水酸基でも、ベンゼン環に付く場合とアルキルに付く場合とは異なる。このような手続きを数十種類いれ、出力時にプリミティブの属性情報の該当する部分にフラグをたてることにより属性データを作成した。

### 3 分子設計における3次元構造エディターの利用

3次元構造エディターを使用する目的には、化合物の立体情報を得る、分子の活性部位の検討、分子のモディファイ、エネルギー解析等が挙げられる。近年、分子設計支援ツールの3次元グラフィックスの開発は目ざましく、いろいろな方式があるが、我々は、次のような特徴を持ったソフトを提案する。

- (1) 3次元座標を持った部品の組合せにより、分子全体の3次元座標を生成する。
- (2) 2次元エディターとの併用により、目的に応じて化学構造式(2次元データ)からも3次元構造(3次元データ)からも対象を指示できる。

全体の流れを図8に示す。化学構造式を入力データとして、3次元データベースから部品の3次元構造情報を取り出し、必要に応じて分子の結合自由度を指定して初期3次元座標を生成する。初期座標値に対して計算による構造の最適化を行い、3次元構造を表示する。表示結果を利用して、対話的に分子をモディファイし、得られた3次元座標値をデータベース化することができる。

以下、特徴となる3次元座標生成、3次元構造表示について説明する。

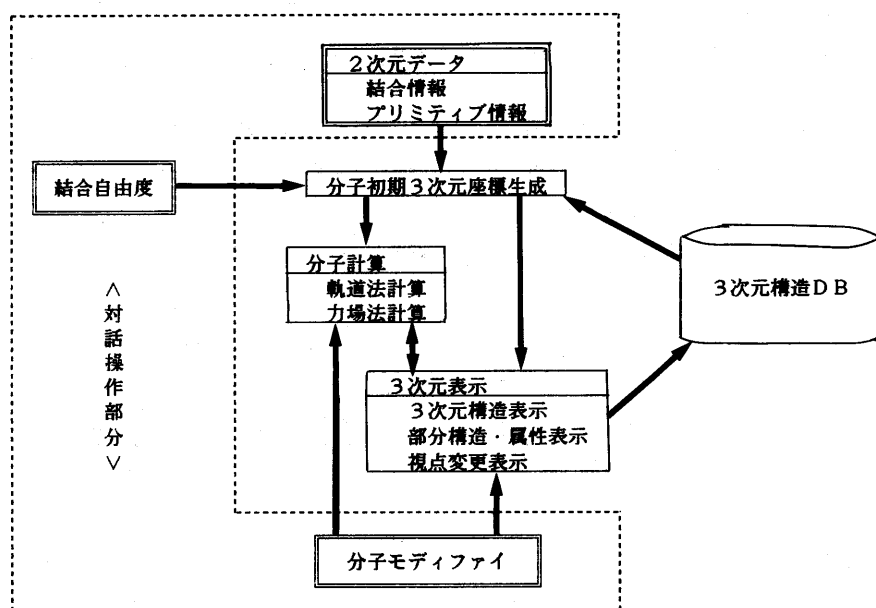


図8 全体の流れ

#### 3・1 3次元座標生成

2次元データから3次元データを生成する方法として、まずCPKモデルによって概略の3次元構造を得て、それを初期値とした配座計算を実行することによって、化学物質のより厳密な3次元構造を求める。このとき化学者側は、次の3つの機能を重視している。

- (1) 入力インターフェースが容易
- (2) ユーザとの対話的な座標生成
- (3) 出力データの再利用を図る

この方法に応えるために、既存の方法を使って、定められた結合長や結合角に従って順次端から組み立てていく方法による簡単なプログラムを試作した。これにより3次元座標生成を行い(必要に応じて構造データベースと照合)、初期値とする方法を考えた。しかし、次のような問題点が大きなネックとなった。

- ・上記のような単純な座標生成法のため、一つ一つの原子間での誤差が生じ値の信頼性が低い。
- ・ループの検出、一部の構造生成にバグが存在する。
- ・入力データとして、入力しやすい化学構造式ではなく、分子内の原子に番号を付けてその結合関係を示すデータであり、規制が多い。
- ・立体情報の表現機能が充実していない。

そこで、3次元座標値の既知な基本構造（以下基本プリミティブと呼ぶ）のデータを組み合わせ、目的分子の3次元座標値を発生する方法を新たに開発した。内部処理としては次のようになる。

- 1) 図9に示すような基本プリミティブの結合情報（分子全体の結合表、基本プリミティブの結合表、分子全体の原子と基本プリミティブの原子との対応表、基本プリミティブ同士の結合形式）を入力データとする。
- 2) 基準とするプリミティブを選定する。
- 3) 既に座標値の決定した基本プリミティブに結合する基本プリミティブを、変換行列によって平行移動する。これを全プリミティブについて実行する。
- 4) 結合自由度入力をユーザの必要に応じて行い、指定した基本プリミティブに回転の変換を施し、結合先プリミティブの最終的な3次元座標値を決定していき、目的分子の座標値を含む結合表を完成させる。

この方法によると、

- ・データベースに登録されている精度の高い既知座標値を組み立てているため、生成座標値の信頼性が高い。
- ・各プリミティブ同士の結合自由度をユーザが決定でき、対話的な分子設計支援が実現されている。また初期座標生成された構造を表示させながら、結合自由度の修正もできる。
- ・出力データの得られた3次元分子の部分構造をデータベースに登録することができるので、ユーザ独自の分子データの再利用ができる。

| table 1. 入力データ |          |   |    |   |    |   |   |   |   |
|----------------|----------|---|----|---|----|---|---|---|---|
| A              | 11       |   |    |   |    |   |   |   |   |
|                | 644      | 2 | 6  | 7 | 8  | 0 | 0 | 1 | 1 |
|                | 644      | 1 | 3  | 9 | 10 | 0 | 0 | 1 | 1 |
|                | 643      | 2 | 4  | 5 | 0  | 0 | 0 | 1 | 1 |
|                | 822      | 3 | 11 | 0 | 0  | 0 | 0 | 1 | 1 |
|                | 821      | 3 | 0  | 0 | 0  | 0 | 0 | 2 | 0 |
|                | 111      | 1 | 0  | 0 | 0  | 0 | 0 | 1 | 0 |
|                | 111      | 1 | 0  | 0 | 0  | 0 | 0 | 1 | 0 |
|                | 111      | 1 | 0  | 0 | 0  | 0 | 0 | 1 | 0 |
|                | 111      | 2 | 0  | 0 | 0  | 0 | 0 | 1 | 0 |
| B              | 11       | 2 | 0  | 0 | 0  | 0 | 0 | 1 | 0 |
|                | 11       | 4 | 0  | 0 | 0  | 0 | 0 | 1 | 0 |
|                | 2        |   |    |   |    |   |   |   |   |
|                | 5        |   |    |   |    |   |   |   |   |
|                | ETHYLENE |   |    |   |    |   |   |   |   |
|                | 22       |   |    |   |    |   |   |   |   |
|                | 2        |   |    |   |    |   |   |   |   |
|                | 8        |   |    |   |    |   |   |   |   |
|                | 644      | 2 | 3  | 4 | 5  | 0 | 0 | 1 | 1 |
|                | 644      | 1 | 6  | 7 | 8  | 0 | 0 | 1 | 1 |
| C              | 11       | 1 | 0  | 0 | 0  | 0 | 0 | 1 | 0 |
|                | 11       | 1 | 0  | 0 | 0  | 0 | 0 | 1 | 0 |
|                | 11       | 1 | 0  | 0 | 0  | 0 | 0 | 1 | 0 |
|                | 11       | 1 | 0  | 0 | 0  | 0 | 0 | 1 | 0 |
|                | 11       | 2 | 0  | 0 | 0  | 0 | 0 | 1 | 0 |
|                | 11       | 2 | 0  | 0 | 0  | 0 | 0 | 1 | 0 |
|                | 11       | 2 | 0  | 0 | 0  | 0 | 0 | 1 | 0 |
|                | 7        |   |    |   |    |   |   |   |   |
|                | COOH     |   |    |   |    |   |   |   |   |
|                | 22       |   |    |   |    |   |   |   |   |
| D              | 6        |   |    |   |    |   |   |   |   |
|                | 643      | 2 | 3  | 5 | 0  | 0 | 0 | 1 | 1 |
|                | 822      | 1 | 4  | 0 | 0  | 0 | 0 | 1 | 0 |
|                | 821      | 1 | 0  | 0 | 0  | 0 | 0 | 2 | 0 |
|                | 11       | 2 | 0  | 0 | 0  | 0 | 0 | 1 | 0 |
|                | 11       | 1 | 0  | 0 | 0  | 0 | 0 | 1 | 0 |
|                | 1        |   |    |   |    |   |   |   |   |
|                | 1        |   |    |   |    |   |   |   |   |
|                | 1        |   |    |   |    |   |   |   |   |
|                | 1        |   |    |   |    |   |   |   |   |

| table 2. 出力データ |   |    |   |    |   |   |   |   |   |
|----------------|---|----|---|----|---|---|---|---|---|
| tra_id=0       |   |    |   |    |   |   |   |   |   |
| tra_ske=0      |   |    |   |    |   |   |   |   |   |
| tra_all=11     |   |    |   |    |   |   |   |   |   |
| C44            | 2 | 6  | 7 | 8  | 0 | 0 | 1 | 1 | 1 |
| C44            | 1 | 3  | 9 | 10 | 0 | 0 | 1 | 1 | 1 |
| C43            | 2 | 4  | 5 | 0  | 0 | 0 | 1 | 1 | 0 |
| O22            | 3 | 11 | 0 | 0  | 0 | 0 | 1 | 1 | 0 |
| O21            | 3 | 0  | 0 | 0  | 0 | 0 | 2 | 0 | 1 |
| H11            | 1 | 0  | 0 | 0  | 0 | 0 | 1 | 0 | 0 |
| H11            | 1 | 0  | 0 | 0  | 0 | 0 | 1 | 0 | 0 |
| H11            | 1 | 0  | 0 | 0  | 0 | 0 | 1 | 0 | 0 |
| H11            | 1 | 0  | 0 | 0  | 0 | 0 | 1 | 0 | 0 |
| H11            | 2 | 0  | 0 | 0  | 0 | 0 | 1 | 0 | 0 |
| H11            | 2 | 0  | 0 | 0  | 0 | 0 | 1 | 0 | 0 |
| H11            | 4 | 0  | 0 | 0  | 0 | 0 | 1 | 0 | 0 |

- A: 分子全体の結合表  
 B: 基本プリミティブの結合表  
 C: 分子全体の原子と基本プリミティブの原子との対応表  
 D: 基本プリミティブ同士の結合形式

図9 3次元座標生成の入力データ

### 3・2 化学構造式から参照可能な3次元構造表示

分子の3次元表示は、分子モデルの視覚化を図ることだけを目的とするのではなく、化学情報を最大限に活用するため2次元構造式や属性データ等の他のデータとの関連付けられている必要がある。このため次のような機能を用意した。全体の機能構成を図10に示す。

#### (1) 対話型操作

メニュー選択による処理指定、マウス使用による容易な化合物の構造上の部分指定、キー入力による数値指定により、得たい情報をグラフィック表示したり、分子の見る視点を変えたり、分子データを自在に修正することを可能とする。例えば、指定された場所への移動、指定部分の計測情報の表示、化学計算結果の表示等である。また簡単な操作で、表示された初期3次元座標の構造を容易にモディファイする機能は効果的である。

#### (2) 2次元-3次元データの対応付け

2次元と3次元のデータを対応付けるために、2次元データと3次元座標値を一つのテーブルに記述する。これにより2次元データの構造情報を3次元構造に利用できる。また2次元エディタとの併用により、2次元データと3次元データの両方を対象に指定操作ができる。例えば、3次元構造の原子や部分を指定すると2次元構造の対応部分を色づけ表示したり、両データの同一部分抽出をする等、化学者の理解を深めることができる。また3次元構造で見えない部分の情報を知りたい時に2次元エディタで指定するなど、3次元構造上では指定できなかった操作も可能になる。

#### (3) 他のデータとの関連重視

他のデータ（名称データ、属性データ等）と対応をさせることにより、様々な化学情報をグラフィック表示できるので、化合物に対する認識を深めるだけでなく新しい発想のよい手がかりとなり得る。例えば、疎水性基の同色表示、温度別表示のように、分子の表面に化学性質を表現することができる。また化学計算の結果を利用して、エネルギーマップを同時表示したり、電子密度分布を分子上に表示することもできる。

この様に、分子を頭にイメージした通りの形に画面上に表現でき、知りたい情報が即座に分かりやすい表現で得られる3次元構造エディタを開発した。

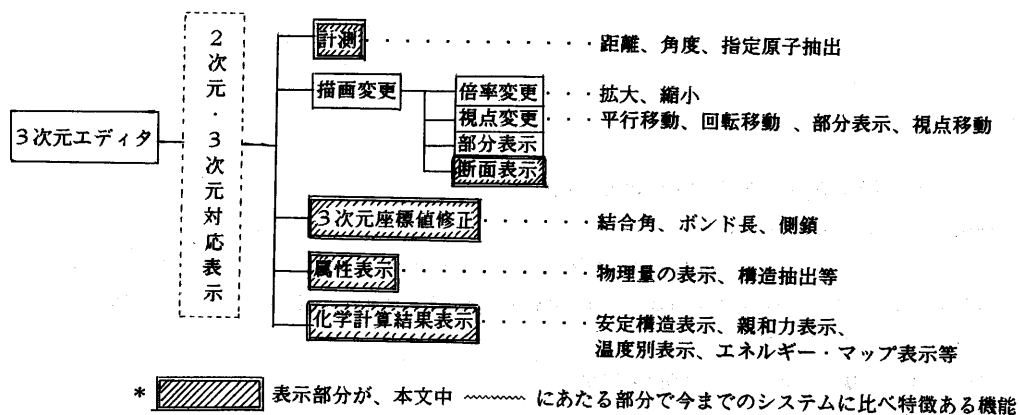
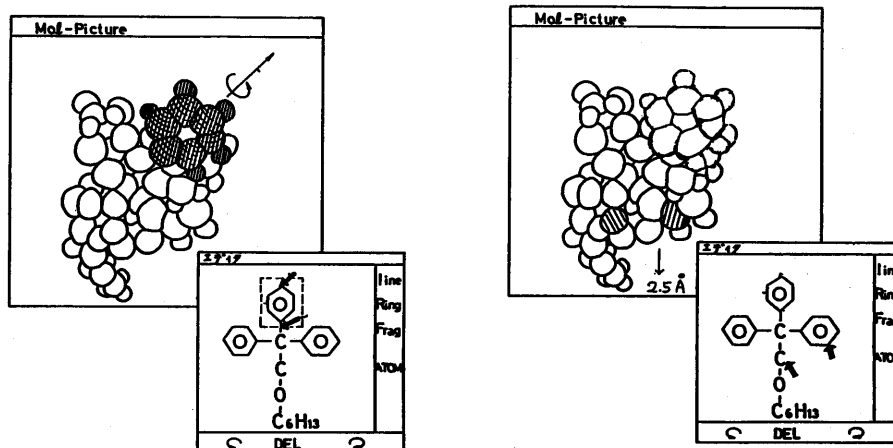


図10 エディタ機能概要



a) 指定部分の対応 (回転軸指定)

b) 距離の計測

図 11 3次元エディタの操作例

#### 4 おわりに

分子設計支援システムの基本的な部分である、分子構造式の入力と属性の推定・抽出、3次元モデル生成・表示を試作した。

分子構造式エディタは自動読み取りを使うことを考慮して、対話的に入力された分子構造式を解析する機能をつけた。このことで、解析結果の情報を編集時にも使えるため、対話的な分子構造式エディタとしてもさらに使いやすいものになった。

プリミティブから分子の3次元構造を対話的に組み立て、分子構造式中の原子と3次元構造中の原子との対応をとることにより3次元表示の指定法のマンマシーンインターフェースを向上させることができた。

しかし属性推定・抽出の部分は不完全であり、知識ベースを用い、どのように知識を蓄え推論するかが今後の課題としてのこされた。

#### 【参考文献】

(1) 岡崎・辻 : "構造規則に基づいた化学構造図式の入力方法"

第32回情報処理学会全国大会 (前期) p 1381

(2) "A Versatile, Efficient, and Interactive Program To Build Molecular Structure for Theoretical Calculations and Chemical Information Systems": JAMES KAO, CHARLES EYERMAN, LORAIN WATT, ROBERT MAHER, and DIANE LEISTER

J. Chem. Inf. Comput. Sci. 1985, 25 p400

(3) "化学の分野における知識ベースシステムに関する調査"

工業技術院化学技術研究所 1987

(4) R. Flettericks, and M. Zoller (eds) : "Computer Graphics and Molecular Modeling, Cold Spring Harbor Laboratory 1986

(5) 神沼伸英・鈴木勇 : "分子を描く" 啓学出版 1988

(6) 佐々木慎一他 : "化学者のためのパターン認識序説" 東京化学同人 1984