

分子内原子座標値を考慮した化学構造入力CAD

中嶋 正之 佐藤 眞一 安居院 猛

東京工業大学 像情報工学研究施設

分子模型のCG表示を作成するためには分子を構成する原子の三次元座標値が必要である。原子の座標値として結晶学的な測定結果を使用することも多いが、さほど正確さが要求されない場合などには、ユーザが入力した目的分子の化学構造をもとに、近似的な原子座標値を計算することが考えられる。

筆者らは、化学式から分子模型のCG表示を作成するシステムにおける化学構造の入力解析手法、ならびに計算機上での化学構造の表現法について検討を行っている。本報告では、オブジェクト指向の考え方をういた化学構造の記述手法と、それに基づいた入力解析動作、目的分子を構成する原子座標の生成などについてその概要を述べる。

CHEMICAL STRUCTURE INPUT CAD WITH COORDINATE APPROXIMATION IN A MOLECULE

Masayuki NAKAJIMA Shinichi SATOH Takeshi AGUI

Imaging Science and Engineering Laboratory, Tokyo Institute of Technology.

4259, Nagatada, Midori, Yokohama, Kanagawa, Japan

To make a molecular model by means of computer graphics, the coordinates of the atoms in a molecule are required. These coordinates are often obtained by making a reference to results of crystallographic measurements. For creating a somewhat rough model, however, approximate coordinates of atoms are calculated by considering the chemical structure of a molecule.

The authors have been interested in an input method and a representative scheme of chemical structure. The present paper describes object-oriented chemical structure representation and generation of the coordinates of the atoms in a molecule.

1. はじめに

分子模型のCG表示を作成するためには、分子を構成する原子の三次元座標値が必要である。精密で正確な表示が要求される場合には、タンパク質データベース⁽¹⁾のように結晶学的な測定結果として得られた原子のデータを使用することが多い。しかし、きほど正確さが要求されない場合や目的とする分子についての測定値が得られていないような場合には、目的分子の化学構造をもとに近似的な原子座標値を計算することが考えられる。

筆者らはそのような場合に用いられるシステムとして、図1に示すように、入力として与えた化学構造から原子の座標値の近似値を計算し、CG表示を作成するシステムを開発している。本システムでは主として有機化合物分子を取り扱いの対象としており、入力部では、多くの有機化合物分子の構造に共通して含まれる比較的単純な部分的構造を組み合わせることで目的の分子の構造を組み立てる。したがって、効率のよい化学構造の入力のためには、入力方法そのものにもまして、個々の部分的な構造をどのように記述・表現するかが問題となり、それに基づいた入力解析処理が必要と考えられる。本報告では、化学構造入力の問題点について検討し、化学構造を系統的に記述する表現法および実際の解析例について述べる。

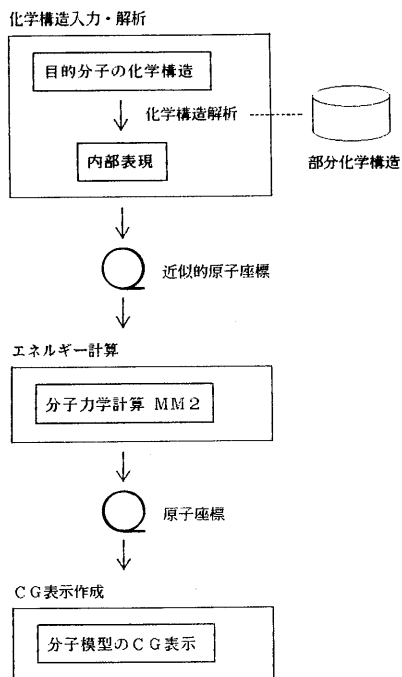


図1. システムの概要

2. 化学構造入力の問題点

2. 1. キーボードを用いた化学構造の入力

計算機上で化学構造を取り扱う場合、原子間の結合を表現する結合表を直接用いる方法や図形情報として原子の位置関係を逐一入力する方法⁽²⁾などがある。しかし、これらの方法は確実ではあるが、目的の化学構造を指定するための入力は煩雑なものになる。したがって、ユーザが希望の化学構造を入力するために用いる適切な化学構造の表現法が必要になる。

筆者らの検討しているシステムでは、基本的な化学構造の入力には主としてキーボードを用いるようにしている。化学構造をキーボードから入力するためには、英数字、記号などキーボードから打鍵可能な文字記号を用いて化学構造を表現しなければならない。筆者らのシステムでは当初、入力が容易であるという理由から、入力形式として通常の化学式を用いる方法⁽³⁾を採用した。この手法では、文字列としての化学式を生成する文法規則を作成しておき、それを用いて入力された化学式を構文解析し化学構造を得た。しかしこの入力方式は簡便ではあるが、環構造など図形的な構造の取り扱いがきわめて難しいなどの欠点があった。さらに入力の単位が一つ一つの原子であったため、ある集まりとして原子の集合を考慮することがなく、実用上希望する化学構造を入力するためには、より組織的かつ系統的な化学構造の表現法と入力法が必要になった。

化学式には分子式、示性式、構造式などいくつかの種類があり、用途に応じて使い分けられているが、有機化合物の表記のためには通常示性式、あるいは構造式が用いられることが多い。これらは原子記号、数字、原子間の結合を明記するための価標、環状構造の図形的な表現などが必要に応じて適宜用いられる。そのため、示性式あるいは構造式を計算機上で取り扱うためにはなんらかの工夫が必要になる。このためには、化学構造式のある規則を持った文字列として表現する各種の線形表記法⁽⁴⁾を利用することができる。

2. 2. WISWESSER 線形表記法

化学構造の線形表記法には各種の種類があるが、その一つである WISWESSER 線形表記法⁽⁵⁾ (以下 WLN と記す) は表記に使用される記号がすべて通常の計算機の文字セットに含まれ、表記が計算機上で自然に扱えるため、化合物検索などの手段として用いられている。WLN 表記では通常の化学式を記述するのとは多少異なった英記号の使い方をするが、ある一定の規則によって通常の化学構造式を WLN 表記にすることも、その逆も可能である。図2に WLN 表記を用いた化学構造の表記の例を示す。この表記法を用いれば、表記法自体に慣れさえすれば比較的簡単にキーボードから

化学構造を入力することが可能である。

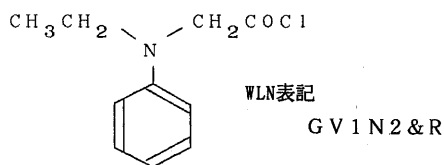


図2. 仮想化合物

入力形式として WLN 表記を採用することで、ある程度自由に化学構造を入力できるようになるが、入力された表記から複雑な化学構造をどのようにして生成するか、という問題が依然として残った。

3. オブジェクト指向を用いた化学構造の表現

3. 1. オブジェクト指向の導入

そこで筆者らは、化学構造を組み立てる際に基本となる官能基などのパーツを一種のオブジェクトとみなし、それらをクラスとして記述し、クラスから生成されたインスタンスによって化学構造を表現する方法を開発した⁽⁶⁾。この手法では、例えば置換基の有無の照会などパーツに関する操作の記述をそれぞれのパーツのクラスの記述の中に埋め込むことで、パーツ自体に関する情報と、メソッドと呼ばれるパーツに対する操作機能の両者を一体として記述していた。

しかしこの方法では、クラスに埋め込まれた機能は、化学構造を解析する場合に補助的に使用されるだけであり、オブジェクト指向でのメッセージ伝達の機構が十分生かされていないとはいえなかった。今回報告する入力手法はこの方式を化学構造入力に対しても適用したもので、入力を解析して化学構造を組み立てる際に必要となる各種の機能あるいは知識をオブジェクト自身に持たせ、それらを具体的に入力処理と組み合わせて化学構造を入力しようとしたものである。

以下ではまずクラス、クラスから生成されるインスタンスについて簡単に説明し、入力時に呼び出される機能・知識の記述などについて述べる。

3. 2. クラス

化学式を書く場合には、いくつかの基本となる構造を意識し、それらを相互に接続するのが普通である。例えば、図3に示すようにメタノールを表現する場合には、単に $\text{C}, \text{H}, 3, \text{O}, \text{H}$ と元素記号と数字を羅列するのではなく、メチル基 CH_3 に水酸基 OH を接続すると考えるのが一般的である。

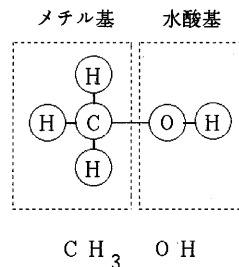


図3. メタノールの構造

この例におけるメチル基や水酸基のように、クラスは、各種の官能基など化学構造の基本となるパーツの記述である。クラスには、そのパーツを構成する原子、それらの原子間の結合様式などが記述され、部分的な化学構造を生成する際の単位となる。クラスはスーパークラスすなわち自分より上位のクラスを参照することができるので、複数のクラスに共通する情報はスーパークラスにまとめて記述しておけば、下位クラスはスーパークラスにある共通な情報の参照を行うことができる。

実際のクラスの記述の例としてベンゼン環のクラス benzene の概要を図4に示す。ベンゼンの水素原子を各種の原子団によって置換することによって、ベンゼン環を中心とした構造には多様なバリエーションが考えられる。そこで水素原子には置換可能であることを明示的に示すマーク(^)をつけておく。ここで水素原子の番号は有機化合物命名法における位置番号⁽⁷⁾に対応するもので、置換の位置を指定する際に用いられる。

```
class benzene
```

atom	C1, C2, C3, C4, C5, C6 H1^, H2^, H3^, H4^, H5^, H6^
各原子の相対的な座標値	

bond	C1-H1, C2-H2, ... C1-C2, C2-C3, ...
------	--

```
method
```

インスタンスの生成 インスタンスを表現する構造体を生成するために必要な諸元
--

水素原子の置換 水素原子の選択と置換手順

⋮

図4. クラス benzene

3. 3. インスタンス

クラスは必要に応じてインスタンスを生成する機能を持つ。クラスは部分的な化学構造の鋳型であり、インスタンスはその鋳型から作り出された部分的な化学構造の実体である。

インスタンスの例として、図5にベンゼンから作られたフェニレン(phenylene)基のインスタンスを示す。インスタンスには、各種の属性の他に、それを生成したクラスへのポイントが保持されており、インスタンスに対してなんらかの操作の要求が起こり、インスタンスに対してメソッドが発行された場合にはそれを生成したクラスが参照される。

3. 4. クラスのもつ機能

クラスには、メソッドと呼ばれるクラスのもつ機能、操作なども記述される。クラスを持つ機能は、そのクラスの生成したインスタンスに対して必要なメソッドを呼び出すことで利用できる。ここでは、それら機能のうちいくつかを説明する。

3. 4. 1. インスタンスの生成

クラスはインスタンス生成の要求に応じて、部分化学構造の実体であるインスタンスを生成する。目的とする分子の構造は、このようにして生成された普通複数個のインスタンスを接続したものととして表現される。

3. 4. 2. 相対的な座標値の保持・発生

官能基に含まれる原子の3次元座標値は、その官能基内のある原子を仮の原点とした相対的な座標値としてクラスの記述の中に保持されている。原点とされる原子は、官能基の種類によっていろいろと異なり、置換操作を行う際に都合がよいように決めてよい。例えばベンゼン環の場合では、図6に示すように、位置番号1の炭素原子を原点とし、位置番号1と4の炭素原子をx軸方向とし、ベンゼン環のなす平面がxy平面に一致するように座標軸を設定する。

目的分子の構造のうち、ある部分的構造をなす原子の座標値を得るには、その部分のインスタンスに対して、本機能の呼び出しを行えばよい。実際には本機能は次に述べる置換操作から内部的に呼び出され、分子を構成する原子の座標を、部分的な化学構造の単位を接続することで順次決定するために用いられる。

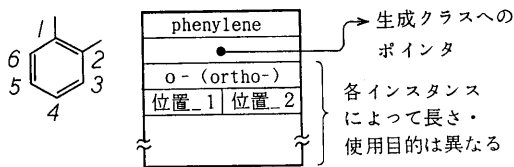


図5. フェニレン基

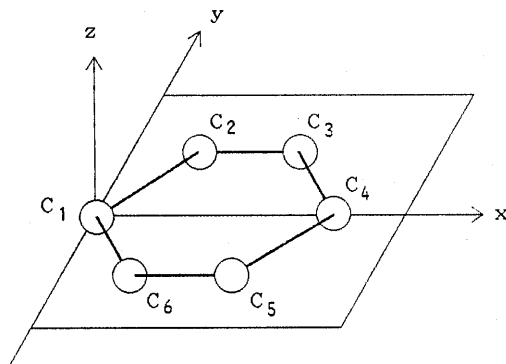


図6. ベンゼン環についての座標軸

3. 4. 3. 置換操作

ある原子団のある原子を別の原子あるいは原子団で置換するという操作は、部分的な構造から目的とする分子の構造を組み立てる際に基本となる操作である。

例えば図7のように、骨格B1, B2, B3のB2に結合している原子Aを、Gという新しく3つの原子からなる原子団S1, S2, S3で置換することを考える。この場合には、置換される原子Aが、置換する原子団の原子S1と置き代えられるわけだが、この際には元の結合B2Aの方向と、原子団G内の原子の相対座標を定めたときのx軸の方向とが一致するように原子団Gが接続される。また、置換前の結合距離B2AとB2S1は異なるので、原子B2と原子S1間の標準的な原子間距離を用いて結合長B2S1が調整される。

結合が炭素-炭素間の一重結合のように結合回りの自由な回転を許すものであると、図7に示すねじれ角 θ には自由度があり、接続される原子団Gの空間的な位置を一意に定めることができない。そのような場合にはデフォルトのねじれ角が使われることになる。例えば、直鎖のアルキル基で炭素数を一つ増やす場合には、図8(a)に示す末端の水素原子H1をメチル基CH3で置換するが、この場合の中央の結合C2-C3に関する結合C1-C2, C3-Cのデフォルト

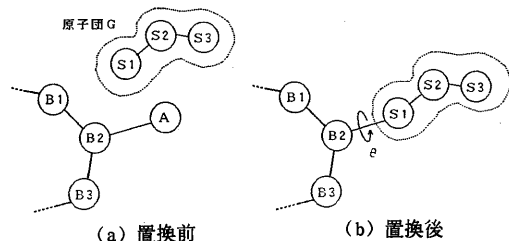
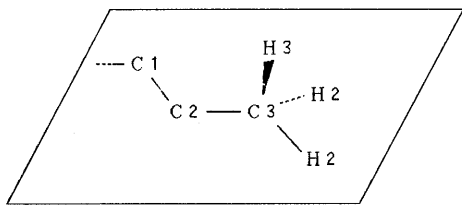
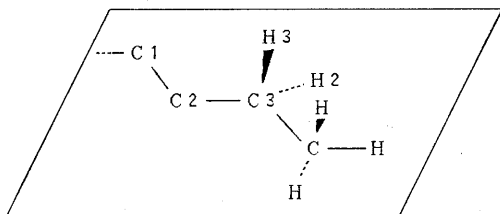


図7. 原子団Gによる原子Aの置換



(a) 置換前



(b) 置換後

図8. 直鎖アルキル基

トねじれ角は 180 度としてあり、図8(b)のように全トランス配座の構造になる。

3. 4. 4. 衝突のチェック

分子内の化学構造を組み立てたとき、それが妥当な空間配座すなわち原子どうしの位置関係になっているかどうかは大きな問題である。しかし、図1に示すように、入力された化学構造から生成される原子の近似的な三次元座標値は、分子力学計算⁽⁸⁾を行ないエネルギー的により安定な構造を計算するモジュールへの入力として用いられるので、あまり精密に原子の位置関係を考慮することにはさほど意味がない。ただし、原子団どうしが衝突しているような場合は満足な計算結果が得られないため、衝突している原子団を分離してから分子力学計算を開始する必要がある。したがって、官能基など原子団の衝突のチェックが必要である。

原子団相互の衝突の判定は、相互の原子間の距離、原子間の結合の交差などを調べることによって行っているが、衝突していることがわかった場合に、衝突を避けるために原子団の位置を調整するなどの機能は持っていない。自由回転を許す結合に関するねじれ角を変化させることで、そのような調整を自動的に行うことが可能である場合も多いが、現在のシステムでは衝突が生じた場合には対話的にねじれ角を調整するようにしている。

3. 5. インプリメント

システムの開発には、プログラミング言語としてC言語を用いたため、これまでに述べてきたようなオブジェクト

指向での基本要素の実現には、手続き型言語であるC言語の仕様の制限を受けることになる。

3. 5. 1. クラスの記述

一つのクラスの記述は、C言語での一つのモジュールに対応している。したがって、クラスの独立性はモジュールの独立性によるものになる。

3. 5. 2. インスタンス

クラスから生成されたインスタンスはメモリ上に動的に確保される構造体となる。クラスは、必要に応じてインスタンス生成に必要なメモリを動的に確保しようとするが、クラスによって生成するインスタンスは異なるため、どのような大きさ、メンバを持つ構造体を確保するかはクラス定義のモジュールの中に記述されている。

3. 5. 3. メソッド

メソッドは、C言語では関数となるが、原則としてC言語では局所的な関数を持つことができないため、クラス記述の中にメソッドの定義を直接埋め込むことはできない。そのため、メソッドは別モジュールとして独立させ、クラスにはそのメソッドの関数へのポインタの配列を持たせ、関数引数として間接的に関数を呼び出すことにした。しかし、逆にこのようにすることで、呼び出し方さえ統一しておけば、メソッドの独立性が確保されるため、共通な機能を共有するなど記述の面で有利であり、さらに保守性を向上させることができた。

3. 5. 4. 全体像

いままで述べてきたそれぞれの要素間の関係の概要を図示すると図9のようになる。

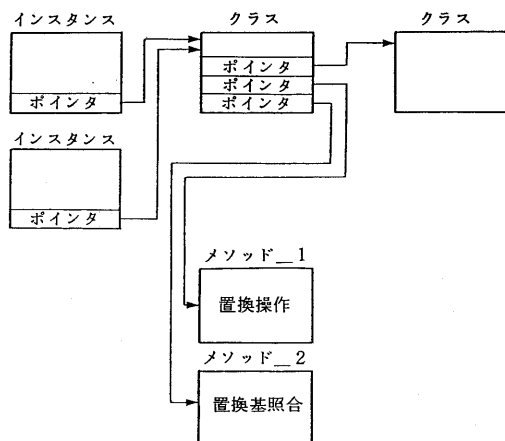


図9. クラス、インスタンス、メソッド

4. 化学構造入力と座標値生成

4. 1. 入力解析過程の概要

入力はユーザの入力した WLN 表記を読み込むことで開始され、WLN 表記に対応してインスタンスが順次生成される。生成されたインスタンスが、インスタンスオブジェクト自身のメソッド、すなわち、3. 4 で述べた各種の機能を呼び出し、部分的な化学構造の生成、相互の接続などが自動的に行われて化学構造が構成される。

4. 2. 原子の座標値の生成

目的とする分子は、各部分のインスタンスが接続された構造となっている。すでに述べたように、個々のインスタンスは分子内の部分的な構造に対応しており、その部分構造を成すそれぞれの原子について局所的な座標値を生成する機能を持つ。したがって、分子に含まれるおのおのの原子の実際の座標値は、座標軸を移動することでそれらの座標値を順次つなぎあわせることによって得られる。

このようにして生成された三次元座標値をもとに直ちに CG 表示を作り出してもよいが、ここで得られた座標値が十分に妥当なものであるとは必ずしもいえない。そこで、生成された原子の三次元座標値は、エネルギー的により安定な構造を計算するモジュールへの入力として用いられ、エネルギー計算の結果が CG 表示のための座標値として用いられる。

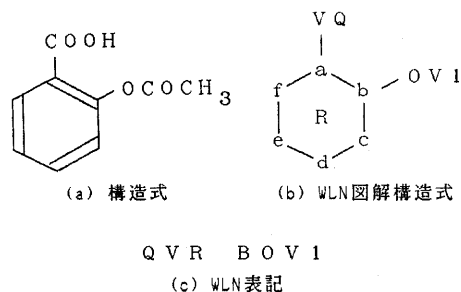
4. 3. 解析例

図 10 (a) に示すアセチルサリチル酸 $C_9H_8O_4$ の化学構造を入力する場合の解析過程の概略を述べる。

図 10 (a) の構造を書き直せば、WLN での図解構造型 (graphic formula) 図 10 (b) を得る。したがって入力した WLN 表記は図 10 (c) のように、QVR_BOV1 となる。_ は空白である。この表記では、Q はヒドロキシル原子団、V はカルボニル結合、R は非縮合のベンゼン環、O は水素の結合していない酸素原子、1 はメチル基をそれぞれ表しており、空白の後の B は環の位置を示すロカント (Locant) である⁽⁵⁾。

“QV” までの入力では、ヒドロキシル原子団とカルボニル結合よりカルボキシル基 $-COOH$ が生成される。これは、入力の過程で強制的に二つの部分構造をまとめ上げたのではなく、それぞれのインスタンスが連結した結果として生成されたものである。

“RB” については次のようになる。R についてベンゼン環に対応するインスタンスを生成するが、環ロカント B により、ortho- の属性を持ったフェニレン基のインスタンスが生成される。このインスタンスには生成時に、ベンゼンの位置番号 1, 2 の位置の水素原子を他の原子または原子団で置換した形をしているということが記述されているので、このインスタンスは、自動的に位置番号 1 に対応する



ここで、 Q: ヒドロキシル原子団
 V: カルボニル結合
 R: 非縮合のベンゼン環
 O: 水素の結合していない酸素原子
 1: メチル基

図 10. アセチルサリチル酸

環ロカント a および位置番号 2 に対応する b に適当な置換操作を行うことになる。環ロカント a には WLN 表記の約束により先に生成されていたカルボキシル基を表現するインスタンスが接続されるが、b に接続されるインスタンスはまだ決っておらず、この位置には、次の表記の読み込みの結果生成されたインスタンスが接続されることになる。

“OV1” については $-O-CO-CH_3$ の構造に対応するインスタンスの集合が生成され、これによって環ロカント b に關する置換操作が満たされる。生成されたインスタンスは最終的には、全体として図 11 のようになる。

このように正しい表記が過不足なく与えられた場合には問題が起きないが、不適當な入力のためなどで、矛盾のある化学構造が指定されることも考えられる。たとえば上の例で、環ロカント b に接続するインスタンスが別の場所の置換に用いられ、意図しない構造になることも考えられる。そのような場合には、フェニレン基のインスタンスに置換基の照会のメソッドが呼び出された時点で、異常を検出することができるため、入力構造異常の警告をすることが可能である。

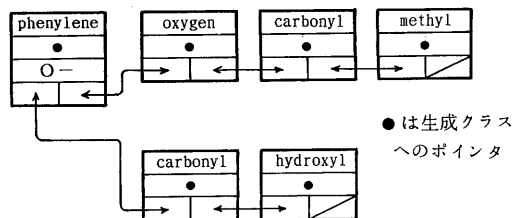


図 11. アセチルサリチル酸のインスタンスによる表現

なお、このようにして得られたアセチルサリチル酸の原子の三次元座標値をもとに、分子模型のCG表示を作成した例を図12、図13に示す。

5. あとがき

化学構造を組織的に入力し、内部処理することができるようオブジェクト指向を用いた化学構造表現手法を開発し、それに基づいた入力解析処理手順について述べた。

現段階のシステムでは、化学構造式の表記法としてWLN表記を用いているため、システムに対して化学構造を入力するためには、まずWLN表記自体になれる必要がある。WLN表示法は複雑な化学構造を文字列として表現できるという大きな利点があるが、自在に使いこなせるようになるためには修得の手間がかかる。筆者らは現在のこのシステムを核として、さらにユーザ・インタフェースを向上させたより対話性のよいシステムを開発中であり、ユーザが実際にシステムを利用する環境での評価を十分に行うことが重要な課題である。

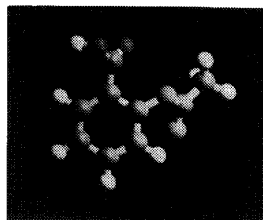


図12. アセチルサリチル酸の球棒モデル

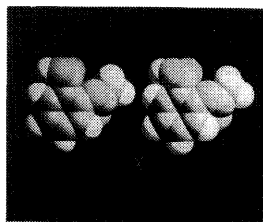


図13. アセチルサリチル酸の空間充填モデル
(ステレオペア)

参考文献

- (1) F. C. Bernstein et. al: "The Protein Data Bank: A Computer-based Archival File for Macromolecular Structures", Journal of Molecular Biology, 112, pp. 535-542 (1977)
- (2) 工藤: "化学構造表現法", 化学総説 No. 18, pp. 63-90, 日本化学会 (1978)
- (3) 安居院, 佐藤, 中嶋: "構文解析的手法を用いた化学構造入力方式", 昭和63年度信学全大, D-533 (1988)
- (4) 平山: "化学構造の線形表記法", 化学総説 No. 18, pp. 9-26, 日本化学会 (1978)
- (5) 平山, 佐々木: "WLN-化学構造式の線形表記法", 南江堂 (1975)
- (6) 安居院, 佐藤, 中嶋: "オブジェクト指向を用いた化学構造解析の一手法", 昭和63年度信学全大, D-214 (1988)
- (7) 廖: "全有機化合物名称のつけ方", pp. 17-42, 三共出版株式会社 (1986)
- (8) 大澤, 田辺, 杉江, 水野: "計算化学ガイドブック", 丸善株式会社 (1988)