

1. “計算機化学(Computer Chemistry)”の展開

化学の分野に計算機が導入されたのは比較的早いですが、当初は実験データの整理等を目的としたものであった。この過程での計算機の役割はあくまでも付随的なもので、“計算機化学(Computer Chemistry)”として一つの研究対象となるようなものではなかった。

その後の急速な計算機の発展に伴い、単なる計算補助的な仕事から、計算以外の機能の利用を目的とした利用が試みられるようになってきた。例えば、化合物データベースといった利用では、計算機の記憶能力と高速な検索能力が計算機に期待されている能力である。しかし、この化合物データベースが実用化されるまでには化学特有の問題に起因する様々なハードルを乗り越えることが必要であった。この計算機と化学とのギャップを埋めるのが“計算機化学”という学問の分野と言える。

この“計算機化学”に関する確たる定義は存在しない。化学分野で計算機を用いようとする時、化学と計算機との間に存在する様々なギャップを解決し、計算機を化学の分野の仕事に役立てることができるようになるのが“計算機化学”であると考えても問題はないであろう。

1. 1 “計算機化学”と“計算化学”

“計算機化学”という言葉の他に、最近では“計算化学(Computational Chemistry)”という言葉も使われるようになってきた。この2つの言葉は混同されがちであるがそれぞれ意味する所は異なるのでここで明確にしておく。

2つの言葉は計算機を用いるという点では同じ基盤に立つものである。しかし、計算機化学は計算機の持つ多彩な能力を化学の分野に役立てようとするもので、計算化学は計算機の強力な計算能力を十分に活用(活用せざるを得ない分野)することを目的としたものである。計算化学は主として、分子力学計算、分子軌道法計算、そして分子動力学計算といった分野の研究で利用されている。この計算化学においても、計算機と化学の接点としての技術は計算機化学の分野となる。

計算機化学は計算機と化学の接点における幅の広い分野をカバーするもので、計算化学は計算能力に強く依存する化学の分野を特に意識するものである。この点で計算化学は計算機化学に組み込まれるものであり、計算化学における分子の表示や計算結果のデータベース化といった仕事は計算機化学の仕事と言える。

2. 計算機と化学との接点

計算機は文字/数字データしか扱えないため、化学分野におけるデータベースは化合物の構造式を直接扱う必要のない文献や物性データ等の文字/数字主体のデータベースとして発展してきた。その後、化学構造式を取り扱う為の技術が開発されるにつれて、化合物構造式を検索対象とする化合物データベースが展開されるようになった。この化合物構造式を直接計算機で利用する技術が開発されたことで、その後の“計算機化学”の進歩は急速に展開されてきた。

このデータベースの内容やマンマシンインタフェース等の展開はこの化学特有の問題の解決と強くリンクしている。従って、これらの問題点を見直すことで化学における検索出力の展開を良く理解できるようになる。

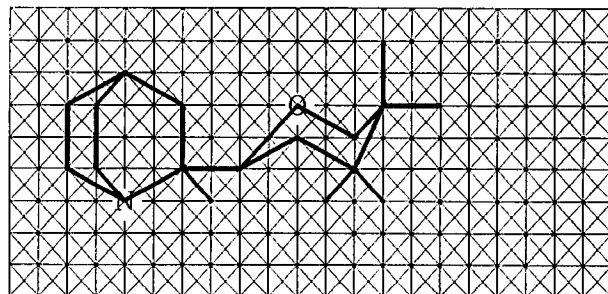
ここでのべている様々な化学に由来する問題点は計算機のハードに起因するもの、および化学特有の問題に起因するものとに分類することが可能である。ここでは化合物を計算機で扱う時に問題となる様々な要因をまとめだし、その問題点がいかようにして解決されていったかを簡単にまとめる。

2. 1 化合物構造式と0/1データとの連繫

2. 1. 1 化合物構造式の数値データによる表現方法
2次元構造式/3次元構造式

2. 1. 1. 1 構造式のシステムへの入力/出力

- ・化学タイピング
- ・化学タブレット



- ・グラフィックディスプレイの利用

線画/ソリッド/2, 3次元

2. 1. 2 化合物構造式と数値データの1体1対応

2. 1. 3 検索手法とその解決

- ・グラフィックディスプレイの利用
- ・検索システムのマンマシンインタフェースについて

● 検索システムの総合的形態

- ・検索システムの種類
 - 化合物検索
 - 反応検索
 - データ(物性等)検索
 - 3次元検索

- ・インハウス/パブリック

- ・検索システムの検索機能

○ システム全体構成に関する考察

- ・段階的構築(数次にわたる検索の実行)
 - プレスクリーニング
 - 本格検索
 - 確認検索
- ・検索結果の操作
 - AND/OR/その他の組み合わせ検索

○ 検索速度/精度向上の為の様々な工夫

第1次検索関連技術

- ・分子式による原子種／原子数による大まかなフィルタリング

第2次検索関連技術

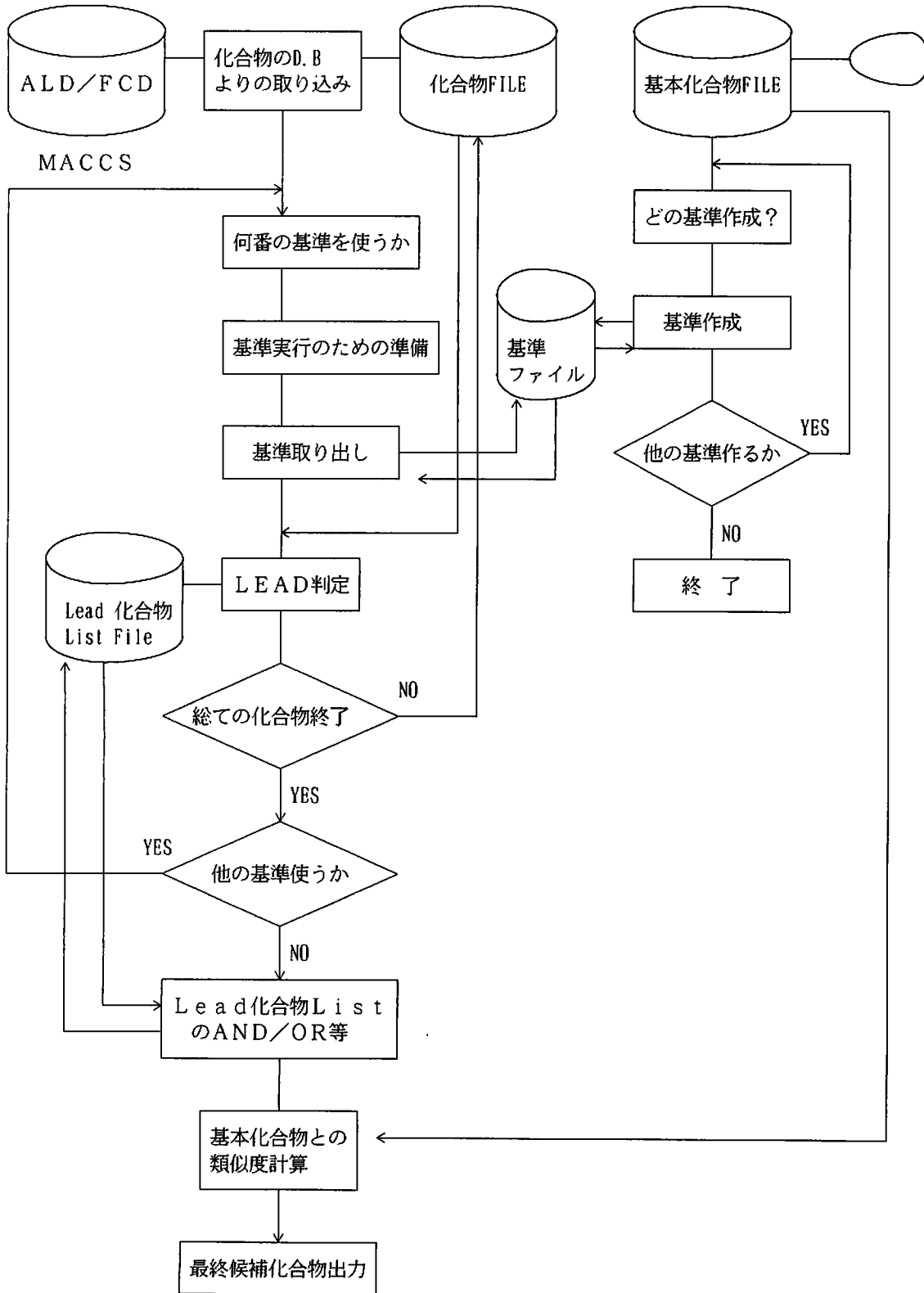
- ・フラグメント検索等の利用による、より詳細なフィルタリング

第3次検索関連技術

- ・アトムバイアトム検索による確認および最終選択過程

様々な検索様式に対応する為の工夫

CAS/CIS/ISI/etc



パブリックデータベース

現在世界的レベルで化合物関連のデータベースが普及している。これらのデータベース中最大のものはCAS (Chemical Abstracts Service)である。

このデータベースは一般に開かれたデータベースであり、料金を払いさえすればどんな人でも世界中何処からでも検索可能である。

インハウスデータベース

これは各企業や個人ベースでデータベースが構築されているもので、従って関係者以外はそのデータベースにアクセスすることはできない。この種のデータベースでは、データベース自体は市販のシステムを用い、そのデータベースの中のデータだけを個人や企業が作り上げるというアプローチが取られることが多い。このインハウスデータベースに用いられるシステムとしては米国MDL社が販売しているISISが最も有名である。

□ 化合物構造式の座標創出

化合物は基本的には3次元構造式を持つ。紙上に書かれる構造式は2次元座標を持っている。もうひとつの表現方法として計算機ではしばしば化合物を0次元で扱う事がある。この0次元では化合物を目でみる形で表現することは不可能であるが、化合物の原子種、結合種、原子同志の結合関係等の情報は持っている。

グラフィックディスプレイの発展していなかった時代、計算機による化合物の扱いは化合物構造式の表示は不可能であり、また構造式表示の必要ない作業に限定されていた。その後、グラフィックディスプレイの発展にともない化合物構造式の表示が容易となってきた。この結果、化合物表示のための2次元及び3次元座標データが必要となってきた。

① 0次元 ⇨ 2次元

一昔前のシステムを扱った人ならばわかると思うが、当時のシステムで利用された化合物構造式は例外をのぞけば、座標データを持たない単純な結合表で表現されていたことが多かった。この座標データを持たない結合表を用いて化合物構造表示をしようとするならば2次元座標の創出が必要となる。

この2次元座標の創出にはこれと定まった手法は存在しないが、よりアルゴリズムックなアプローチが幾つかとられているのでその概要をのべる。

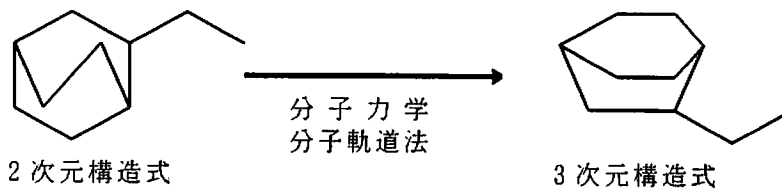
- ① 結合表から化合物の特徴をとらえる。
 - ・ 環及び側鎖の判別
 - ・ 環の状態の判別 (員数、縮合/孤立、スピロ、他)
 - ・ 側鎖の状態 (官能基、チェーン、他)
- ② 環及び縮合環のうちで最も大きい単位を取り出す
- ③ ②で取り出された単位を中心に据え置く
- ④ すでに座標がFixされた基本部分に側鎖部分を順に結合してゆく
- ⑤ 座標や結合の重なり状態をチェックする

* 主要な環 (5、6、7、8員環等) や官能基等ははじめから正しい2次元構造式の座標データをシステムでストアしておき、該当するものが存在するときはこのストアされているパーツを呼び出して表示すると、内部であらたに座標を発生させる必要がない。

② 2次元 ⇨ 3次元

最近展開されるシステムでは化合物の3次元座標を扱うことが常識となっている。この3次元座標を扱うことで、化合物の3次元表示のみならず、種々分子軌道法への入力構造式として用いる、通常では不可能な化合物の重ね合わせを行う等さまざまな目的に適用可能となる。

2次元座標データを3次元座標データにもどすのにも幾つかのアプローチがとられている。これらのアプローチに利用される技術としては、分子力学や分子軌道法が利用されることが多い (図)。



この分子力学や分子軌道法による3次元座標創出でも、その入力データとなる最初の2次元構造式の形があまりにもデフォルメされていると (紙面上の書き方では正しい書き方であっても) 分子力学の計算が不能になったり、分子軌道法による計算結果が全く異なる化合物になったりするので注意が必要である (図)。

通常は化合物の各原子に関する情報を読み取り、各原子が持つ3次元的な配向と結合距離を考慮し、簡単な分子模型をつくる程度に修正された3次元座標データを最初に生成し、その後分子力学や分子軌道法に持ってゆくことが多い。

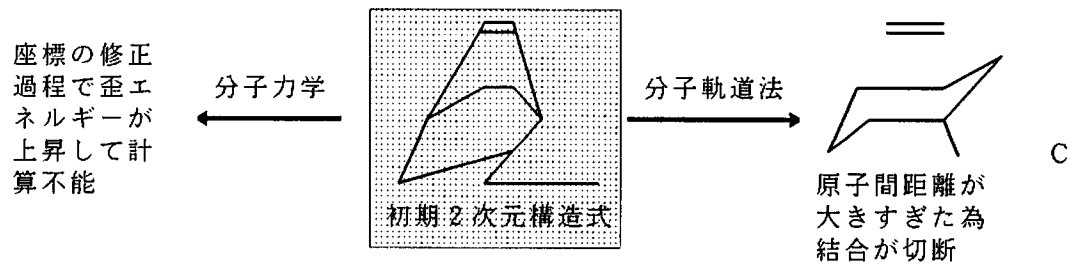


図 . 大きくデフォルメされた2次元構造式を用いた時

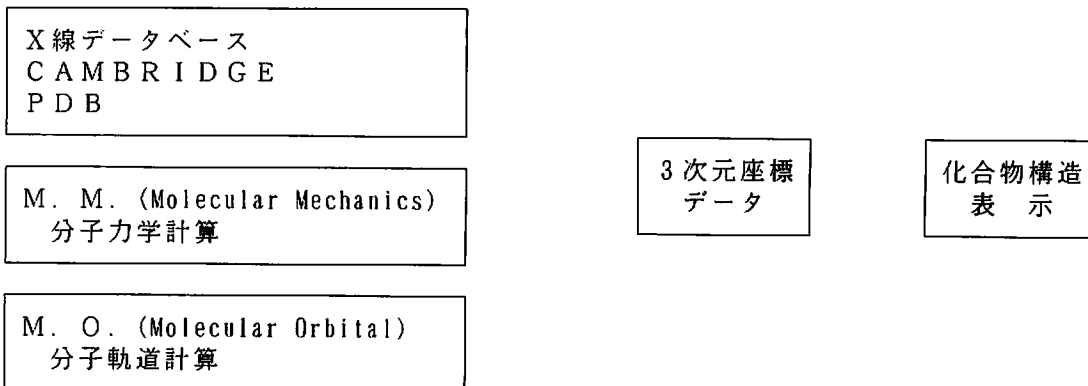
□ 化合物構造式のグラフィックディスプレイ上での表示

- ① スティック図 (スケルトン図)
- ② ボール&スティック図
- ③ 空間充填図
 - ・メッシュ表示
 - ・ソリッド表示
- ④ ドット図
- ⑤ その他の化合物表示手法
- ⑥ 関連計算機技術
 - ・グラフィックライブラリ
 - ・表示の為の技術
 - シェーディング
 - 透視変換
 - ステレオ図

□ 化合物表示法 (分子グラフィックス)

計算機による化合物構造表示は化合物の3次元座標が用意に得られるようになった事、及びグラフィックディスプレイの進歩により最近急速に発展してきた分野の一つである。化合物の構造式を立体的に表示させるには、各原子についての3次元座標が必要となる。このデータは古くはX線回折のデータより得られ、又簡単なものは分子模型を用いて化合物の3次元モデルを作成した。

現在、X線データベースはCAMBRIDGEのX線結晶データベースやPDB等のデータベースが充実され、容易に化合物の3次元モデルが得られるようになっている。また、計算時間はかかるが、分子軌道法でも3次元座標をもとめる事は可能である。



化合物表示の必要性

様々な表示手法

■ 化合物を対象とした表示手法

- ・スケルトン（骨格）図
- ・ボール&スティック（豆球）図
- ・スペースフィリング（空間充填）図
- ・ドット図
- ・ワイヤメッシュ図
- ・ソリッド図

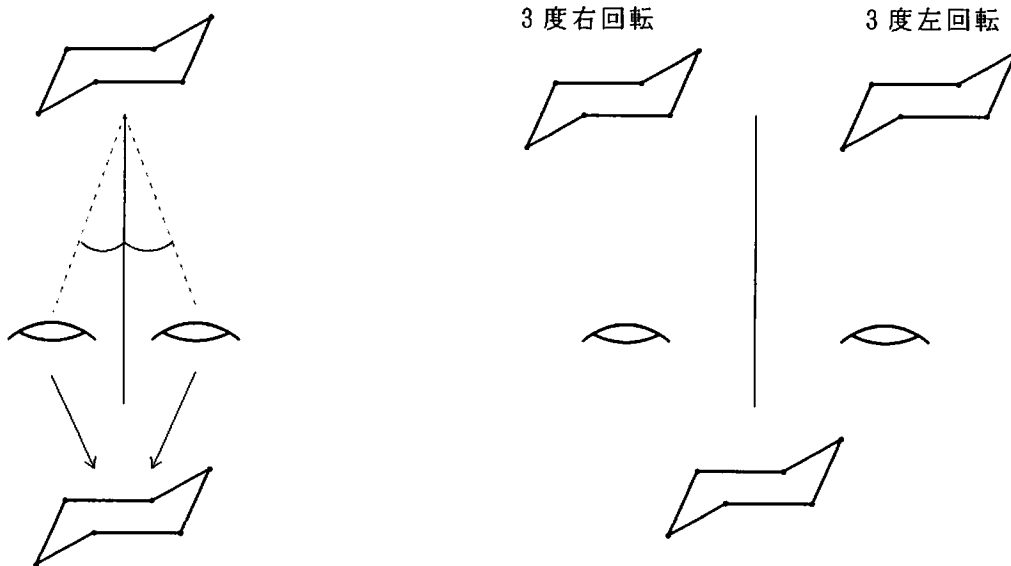
■ 蛋白/DNAを対象とした表示図

- ・リボン図
- ・ワイヤ図

■ 立体視を目的とした表示手法

- ・ステレオ図

昔から行われている手法で、左右少しずつ角度の異なる図を描き、それを左右の眼でみる事により立体像を得るものである。



ディスプレイ上に3度ずつ回転させた2画面を左右それぞれの眼でみて頭の中で2つの画面を合成しひとつの立体画像とするものである。この2画面を1つの画像にまとめる事は多少の訓練を必要とする。現在では特殊な眼鏡をとおしてディスプレイをみる事で、左右の画面を別々にみせる事で簡単に立体画面をえる事ができる。

- ・3次元ディスプレイの利用

■ より実感を出す為の表示手法

- ・レイトレーシング
- ・ラジオシティ法

表示の為のグラフィックディスプレイ

■ グラフィックディスプレイの種類

- ・ストレージ（リフレッシュ）型
- ・ランダムスキャン型
- ・ラスタースキャン型

■ 2次元及び3次元ディスプレイ

- ・2次元ディスプレイ
- ・3次元ディスプレイ

絵を描く為の表示技術

- ・透視変換
- ・シェイディング
- ・デップスキューイング
- ・陰線／陰面消却
- ・半透明表示

物を描く時の2タイプ

- ・ワイヤモデル
- ・ソリッドモデル

物体を操作する為の技術

- ・回転／移動／拡大／縮小
- ・アフィン変換

3次元ディスプレイ上での画面操作の技術

- ・ズームング
- ・パンニング
- ・クリッピング
- ・ピッキング

図形表示の為のソフト

- ・表示プログラムリスト
- ・
- ・

□ 化合物構造特徴の認識

概論： 化学の分野において何らかの解析作業を行う時には化合物の構造的な特徴を解析作業の目的に応じて的確に認識する事が大切な作業となる。 この作業は計算機を用いた解析作業でも全く同じである。 化合物構造式中に潜む様々な構造的な特徴を計算機を用いて取り出す技術は種々解析プログラムを作成する時に必要となる技術である。

ここでは化合物中に潜む様々な構造的な特徴を計算機を用いて取り出す技術についてまとめてみる。 但しスペクトル解析のようにスペクトルを基本として化合物情報を取り出す（計算機によるスペクトル解析）ものでなく、あくまでも化合物の結合表を基本とし、その結合表内部に潜む化合物情報の取り出しという点についてのべる。

化合物の構造的な特徴としてさまざまなものが存在する。 例えば、官能基、環及び側鎖、芳香族性、その他の特性が存在し、化学の問題はこれらの情報を整理し、利用しながら行われることになる。 ここではこれらのさまざまな特性のうち頻りに利用される、官能基の取り出し、環及び側鎖部分の判定、芳香族性の判定、立体及び幾何異性の判定について簡単にまとめる。

1. 官能基の認識

2. 環の認識

化合物構造は環部分と側鎖部分とから構成される。化学においては化合物中の環と側鎖を識別する事が一つの重要な作業となる。例えば、合成においても環と側鎖の合成は異なるアプローチをとり、スペクトルでも環と側鎖部分とでは異なるデータが得られることが多い。

計算機システムにおいても環の認識問題は重要なステップとなる。この環状部分の認識は大きく2種類の作業に分類される。

- ① 環と側鎖部分の認識
- ② 最小の環とより大きな環の認識

①は化合物中の環と側鎖部分を区別する作業である。この作業により環の存在と環の大きさ(孤立環のみ)が認識出来る。さらに、環が縮合している時には環の数や大きさは様々な形で算出可能となり、最終決定を行うことは困難である。このために縮合環を構成する環の数やその員数を決定するには環そのものにたいする定義が必要となる。この定義の一つとして、後にのべるSSSR (Smallest Set of Smallest Ring)がある。このSSSRは縮合環を構成する環の最小単位となる環を意味している。化学的には常にこのSSSRを対象としており、複数の環で構成されるより大きな環は考慮しないのが常である。このSSSRを決定すれば縮合環を構成する環数とその員数とを決定することが出来る。

以下ではこの①と②について順に説明する。

2. 1. 環と側鎖部分の認識

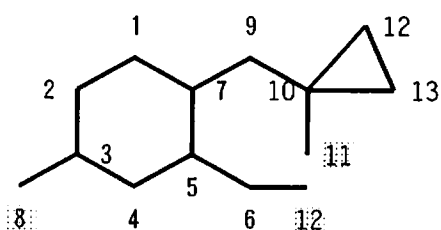
化合物構造式中の環状部分と側鎖部分を識別する事は化学の分野においては重要な作業である。この作業を行う方法として様々なアプローチがあるが、ここではそれらの中で最も基本的な手法である逐次法(アルゴリズム的なアプローチ)とマトリクスを用いた手法を紹介する。

① 逐次法による環と側鎖部分の認識

この手法は化合物中の側鎖部分の末端原子から順に原子を取り除く事と、環状部分を取り出して環の存在を確認した上で、環上の結合を切断し、再び末端原子を取り除く作業を繰り返しながら環状部分と側鎖部分の同時チェックを行う手法である。

イメージ的にわかり易いものであるが、プログラ的にはロジカルなものとなり、IF~THENの判断を多く含む複雑なプログラム構造となり、且つ比較的長い実行時間を必要とする。

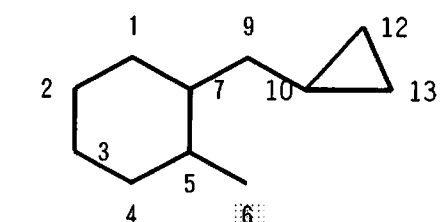
手順1 化合物中の末端原子の消却



左図において末端原子は8、11、12である。最初に、これらの原子につながる結合を切断する。

環 = 0

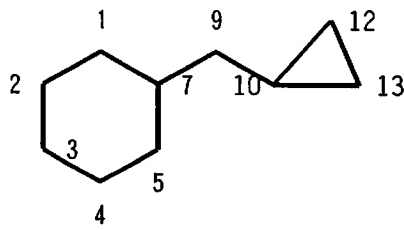
手順2 末端原子の削除(継続)



切断後の構造式は左図のようになる。この構造式中で末端原子となっている部分を同様に切断する。この手続きを、これ以上末端原子が存在しなくなるまでつづける。

環 = 0

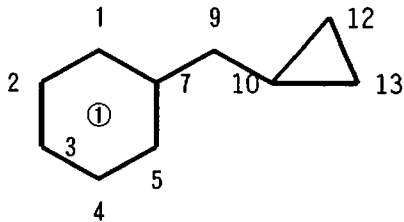
手順3 末端原子が存在しなくなった状態



この段階での最終構造式は左の図式のようなになる。末端原子から取り除くPRUNE手法では図からわかるように環と環に挟まれた原子(結合)はPRUNE不可能となる。

$$\text{環} = 0$$

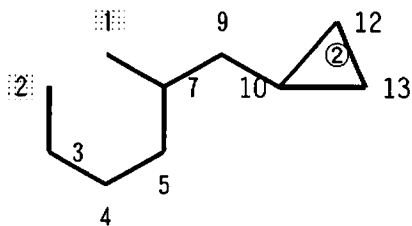
手順4 環構造の確認(発見)



PRUNE不可能な時点で、一旦環構造の認識を行う。この時、一個でも環が発見されればその環上の結合を切断する。ここでは6員環①が発見されたとする。

$$\text{環} = 1 \quad (\text{6員環})$$

手順5 発見された環構造部分の結合切断、及び作業の繰り返し



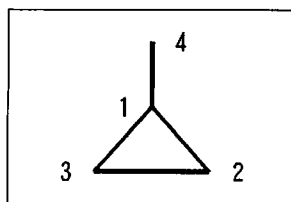
結合1-2を切断する事で左図のような構造式が得られる。この構造式に対し、(1)から(4)迄の手続きを繰り返し、総ての原子がチェックされる迄つづける。この手続きで②の3員環が発見される。

$$\text{環} = 2 \quad (\text{6員環}) \quad (\text{3員環})$$

② Burghard Schmidt アルゴリズムによる環/側鎖部分判定

* Burghard Schmidt & Jorg Fleischhauer, "A Fortran IV Program for Finding the Smallest Set of Smallest Rings of a Graph", J.C.I.C.S., 18, 204-206(1978).

この手法は化合物の結合表を基本としていくつかのマトリクス(配向隣接マトリクス及びパスマトリクス)を構築し、そのマトリクスから化合物内の環状部分と側鎖部分を一気に識別する手法である。

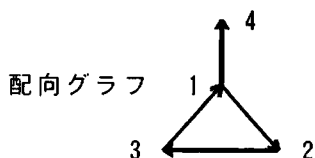


基本構造式

手順1 先ず対照とする化合物の結合表を作成する

	1	2	3	4
結合表	0	1	1	1
	1	0	1	0
	1	1	0	0
	1	0	0	0

手順 2 配向グラフ (ORIENTED GRAPH) の作成



まず化合物中の 1 原子を選び出し、出発原子とする。この原子を出発原子とし、隣接する原子に次々と連続的にベクトルを引いてゆく。分岐点がある時はマークしておき、グラフが一旦終点にたっした時、そのマークされた分岐点から再びベクトルを引く。この手続きを総ての原子/結合について繰り返す。

手順 3 配向隣接マトリクス (ORIENTED ADJACENCY MATRIX) の作成

配向グラフの作成と同様の順番で結合表中のデータを変換し、配向隣接マトリクスを作成する。この時出発原子を I とし、隣接原子の番号を J とすると、I と J 環の結合は結合表中で A_{IJ} と A_{JI} は 1 となっている。この中で A_{JI} を 0 とする。続いて、J 原子を起点として J に隣接する原子 L を探し、I と L の時と同様に A_{JL} を 1 に、 A_{LJ} を 0 とする。この手続きを総ての原子と結合について繰り返す事で配向隣接マトリクスがえられる。

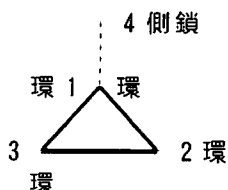
	1	2	3	4		1	2	3	4
1	0	1	1	1	→	1	0	1	0
2	1	0	1	0	→	2	0	0	1
3	1	1	0	0	→	3	1	0	0
4	1	0	0	0		4	0	0	0
結 合 表						配向隣接マトリクス			

手順 4. パスマトリクス (PATH MATRIX) の作成

前記配向隣接マトリクスを基本としてパスマトリクス (PATH MATRIX) を作成する。このパスマトリクスは原子を起点として配向グラフをたどった時、到達しうる原子を 1、到達しえない原子を 0 としたマトリクスである。

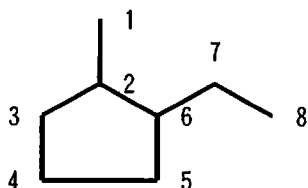
	1	2	3	4		1	2	3	4
1	0	1	0	1	→	1	1	1	1
2	0	0	1	0	→	2	1	1	1
3	1	0	0	0	→	3	1	1	1
4	0	0	0	0		4	0	0	0
配向隣接マトリクス						パスマトリクス			

手順 5 パスマトリクスより環及び側鎖原子の決定



パスマトリクス中、対角項の値が 0 の時には側鎖原子であり、0 でない時には環上原子と判定する。従って、先のパスマトリクスについては環上原子が 1、2、3 であり、側鎖原子は 4 となる。

QUIZ

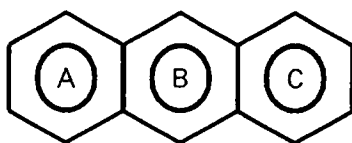


左図構造式の側鎖と環部分とをBurghard Schmidtアルゴリズムにより識別せよ。

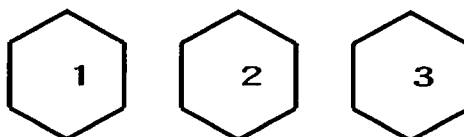
2. 2. 最小構成環 (SSSR: Smallest Set of Smallest Ring)の認識

SSSRとは化合物を構成する環のうち、これ以上小さくする事の出来ない最小の環を意味する。CoreyとWipke*1は環をReal RingとPseudo Ringの2種類に分類している。Real RingとはSSSRと同じであり、Pseudo Ringとは複数のReal Ringから構成されるものである。

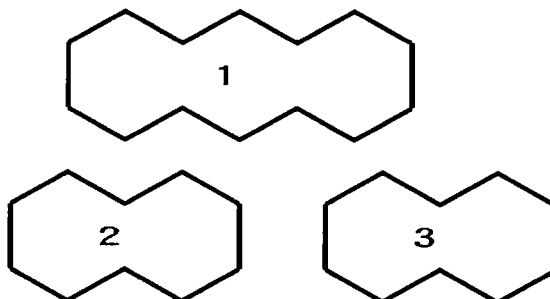
多環縮合化合物



SSSR
Real Ring



Pseudo Ring



* 1. E.J.Corey and W.T.Wipke, "Computer Assisted Design of Complex Organic Synthesis", Science, 166, 178-192(1969)

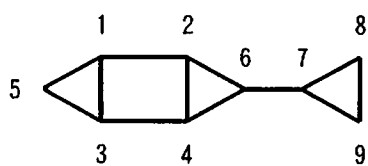
化学の分野において環構造を認識する時、これら2種類の環を識別する事が必要となる。REAL及びPSEUDOのどちらの環構造が重要かについては目的により異なる。例えば化合物の合成について簡単な例を考える。前記縮合3環性化合物の合成を考える時、B環をDIELS-ALDER法により合成する事を考えるならば、B環は独立して考える事が必要である。一方、アンスラセンのA環に置換基を導入する事考える時はアンスラセン全体を考慮した反応性について考慮する事が必要となる。

① Burghard Schmidt アルゴリズムによる SSSR 判定

この手法は 2. 1. の環と測鎖の判定の中でのべた②の方法を提唱したグループが発表したものである。つまり一連の手続きとしてはある化合物が与えられた時、最初に環と測鎖部分の判定により環構造部分を取り出し、つづいてその取り出された環構造部分について以下にのべる手法により SSSR を識別する。

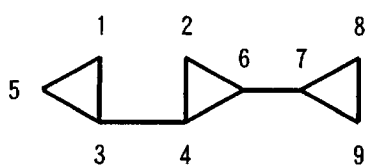
基本的には各原子について総ての隣接原子に関する原子番号情報を取り出して一つのマトリクスを形成し、そのマトリクスに関して一連の手続きを行うことで SSR を順次取り出す手法である。

手順 1



隣接原子	1	2	3	4	6	7
1	2	1	1	2	2	6
2	3	4	4	3	4	8
3	5	6	5	6	7	9
4	0	0	0	0	0	0

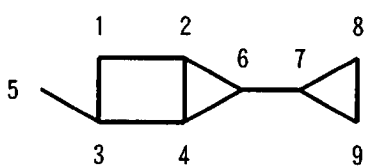
手順 2



隣接原子	1	2	3	4	6	7
1	0	0	0	0	0	6
2	0	0	0	0	0	8
3	5	6	5	6	7	9
4	0	0	0	0	0	0

RING-INDICES 1 - 2 - 4 - 3 - 1

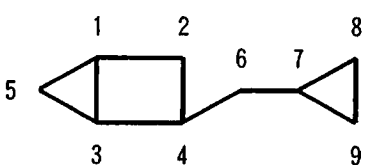
手順 3



隣接原子	1	2	3	4	6	7
1	0	0	0	0	0	6
2	0	0	0	0	0	8
3	0	6	0	6	7	9
4	0	0	0	0	0	0

RING-INDICES 1 - 3 - 5 - 1

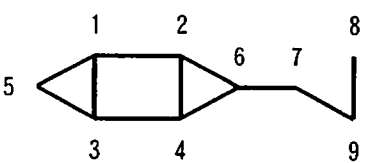
手順 4



隣接原子	1	2	3	4	6	7
1	0	0	0	0	0	6
2	0	0	0	0	0	8
3	0	0	0	0	7	9
4	0	0	0	0	0	0

RING-INDICES 2 - 4 - 6 - 2

手順 5



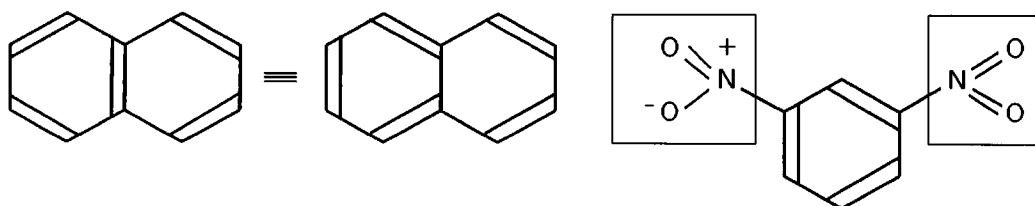
隣接原子	1	2	3	4	6	7
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	0

RING-INDICES 7 - 8 - 9 - 7

③ 芳香族性の認識

化合物内の芳香族性の認識は合成あるいは化合物データベース等でしばしば必要となる機能である。

合成的な立場から考えるならば、芳香族性（あるいは共役）があるか否かで合成アプローチがおおきく変化してくる。また、化合物データベースの立場で考えるならば化合物表示の特殊性からくる同一化合物でありながら入力構造式がことなる等の問題がある。



この検索の問題は必ずしも芳香族性に関係するわけではないが、化合物同義語の検索という点で重要となる。シンプルなシステムではこれら化合物や置換基はそれぞれ異なった化合物と判定してしまう。化学に親切的なシステムであるならばこのような化合物群や置換基群は同一のものと自動的に判定することが必要である。

④ 立体情報の認識

- ・ 幾何異性体
- ・ 立体異性体

⑤ 関連計算機技術

| | : ユーザーイニシャルの最初と最後のスペル
 P P_i : プログラム名
 MM : 月
 DD : 日
 YY : 年
 HH : 時間
 mm : 分
 dd : 2次元情報(2D)および3次元情報(3D)のコード情報
 XX : 未使用
 n n_{i0} : 未使用
 E E_{i2} : エネルギー値(分子モデリング等でエネルギー値が求められている時)
 R R_i : 内部登録番号
 | | | | : ユーザーイニシャルの最初から4番めまで

② Counts line

aaabbb|lffccssssxxxrrrpppiiimmmvvvvv -----> Counts line
 この部分の情報はモルファイルが表現している化合物の全体的な特徴をまとめてある部分である。化合物の原子数、結合数、キラリテイの有無、モルファイルのサイズ、モルファイルのバージョン等が含まれている。このライン中、|sxpriで表現される部分は他のファイルに関係する部分である。

a a a : 原子数(最大255)
 b b b : 結合数(最大255)
 | | | : 原子リストの数 化合物の構造式では利用しない
 f f f : 未使用
 c c c : キラリテイフラグ 0=アキラル、1=キラル
 s s s : Sテキストエントリーの数
 x x x : 反応化合物数+1
 r r r : 試薬の数 } 化合物の構造式では利用しない
 p p p : 成績体の数
 i i i : 中間体の数
 m m m : プロパティブロック(Properties block)のライン数
 v v v v v v : 現結合表のバージョン(現バージョンはV2000である)

③ Atom block

xxxxx.xxxxxxyyyzzzz.zzzz aaaddcccssshhbbbv -----> Atom block

x x x x x . x x x x y y y y y . y y y z z z z z . z z z z
 : 化合物のXYZ座標(実数情報としてそれぞれ10桁利用される)
 a a a : 原子シンボル
 d d : 同位体情報(-3~+4)
 c c c : チャージ 0;チャージ無し 1;+3, 2;+2, 3;+1,
 4;doublet(^), 5;-1, 6;-2, 7;-3
 s s s : 原子の立体パリティ 0;not stereo, 1;odd, 2;even, 3;either or unmark
 h h h : 水素原子数 1;H0(no H atoms), 2;H1, 3;H2, 4;H3, 5;H4
 b b b : 検索用立体情報(検索時にのみ利用)
 v v v : 原子価 0;制限なし、1~14;原子価、 15; 0価

但し、ddとcccの情報はモルファイルのV2000版ではProperties blockに組み込まれている。ddはM ISO、cccはM CHGのヘッダーのもとに書き込まれている。従って、V2000以前のモルファイルとV2000以降のモルファイルとでは情報の書き込まれている場所がことなるので注意が必要である。

④ Bond block

111222tttssxxxrrrccc -----> Bond block

1 1 1 : 第1原子番号
 2 2 2 : 第2原子番号
 t t t : 結合の種類 1; 単結合 2; 2重結合 3; 3重結合 4; 芳香族結合
 5; 単結合または2重結合 6; 単結合または芳香族結合

7; 2重結合または芳香族結合 8; 総ての結合
 但し、4~8は化合物検索時のみに用いられる
 s s s : 結合の立体情報 単結合の時 0; 非立体 1; UP 4; Either 6; Down
 2重結合の時 0; cis/transの決定にx y z座標必要
 3; cis or trans (either)
 x x x : 未使用
 r r r : 結合トポロジー 0; either 1; Ring 2; Chain
 但し、化合物検索時のみに用いられる
 c c c : 反応中心情報 (反応情報用のコード)

⑤ Properties block

```
M xxx
M .....
M END
```

} Properties block

このパターンはM xxx でしめされる。最初にあるM xxxパターンのうち、xxxの記号により、続く情報の内容が規定される。この規定された内容に従って、つづくカラムのパターンが変化する。

この情報の終了点はM ENDで示される。以下にこのxxxの組み合わせの一例をしめす。

M CHGnnn aaa vvv .. (aaa vvv) ... 原子のチャージに関する情報記述
 n n n : 1~8の数字
 a a a : 原子の番号
 v v v : 直前のa a aで指定された原子のチャージの値 (-1.5~+1.5)

M ISOnnn aaa vvv .. (aaa vvv) ... 同異体に関する情報記述
 n n n : 1~8の数字
 a a a : 原子の番号
 v v v : 基本原子との原子量の差 (例C¹³はC¹²との差が1なので1がはいる)

現在存在するコマースナル版のプログラムは大部分このモルファイル形式での化合物取扱が可能である。従って、自分でプログラムを作成する時、化合物の表現をこのモルファイル形式にしておくならば、既存のシステムとのデータ連携は簡単に行える。

□ 化合物構造式 (Zマトリクスによる構造式変換)

化合物の3次元構造情報を入力する手段として分子軌道法関連分野で伝統的に利用されてきた方法としてZマトリクスがある。このZマトリクスは3次元座標値入力を主目的とした手法である。この手法は主として量子化学計算プログラムの化合物構造式入力手法として用いられてきたが、最近では化合物の3次元座標値を分子力学等の手法により比較的簡単に求める事が可能となって来た為、前節で述べた結合表形式をとることが多くなって来ている。

Zマトリクスは1つの原子を中心とし、その原子に対する相対的な位置関係を結合長、結合角及び2面体角で表記してゆくものである。その入力方法は以下にしめすようになる。

- (1) 1つの原子をX, Y, Z座標の原点におく
- (2) この原子に繋がっている2番目の原子をX軸上におき、この時の1及び2番目の原子間距離を記す。
- (3) 2番目の原子に繋がる3番目の原子をXY平面上におき、1及び2番目との原子より形成される結合軸 (この場合はX軸) との角度および2番目の原子との結合距離を明記する。
- (4) 4番目の原子は、1, 2, 3番目の原子により形成される平面との角度 (2面体角) 及び2, 3, 4番目の原子の結合角及び3, 4番目の原子との結合距離として示される。
- (5) 5番目以降の原子は(4)の繰り返しで表現される。

番号	原子種	原子間距離	結合角	2面体角
1	C			
2	C	d_{12}	α_{123}	
3	N	d_{23}	α_{234}	β_{1234}
4	C	d_{34}	α_{345}	β_{2345}
5	C	d_{45}	α_{456}	β_{3456}
.....
.....

■ Zマトリクスの特徴

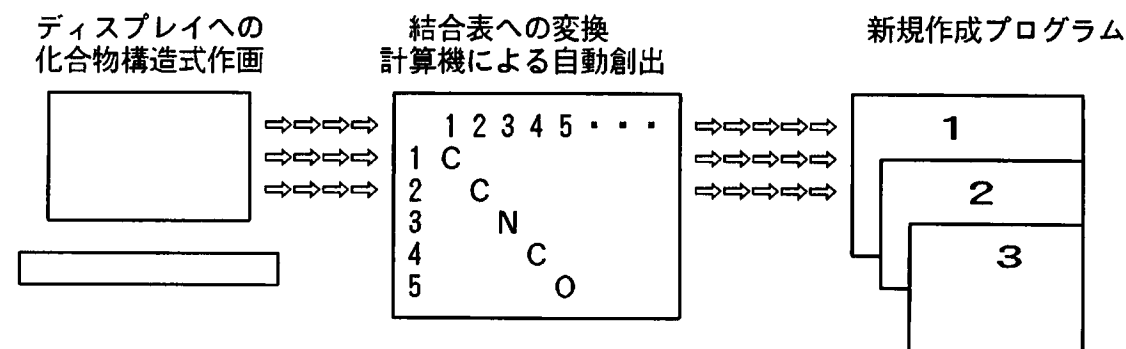
化合物の3次元情報をシステムに入力出来るという点がZマトリクスの特徴である。しかしながらこの方法では化合物の入力の為のデータ作りが面倒であり、一般的には分子模型を見ながら時間をかけてじっくりとデータを作成する事が必要となる。又この入力形式では環状化合物の入力は極めて困難であり（環を形成する時に最初の原子と最後の原子とで位置にズレが生じ易い）、又化合物の結合（1, 2, 3重結合、芳香属結合等）に関する情報も入っていない（データの持ち方で可能）事などにより、一般的には量子化学計算プログラムの入力として使われてイルニスギナイ。又、最近ではグラフィックを用いて予め3次元構造式として入力されている部分構造単位のパーツを組み立てて構造式を構築するのが一般的となりつつあり、またX, Y, Zのユークリッド座標系を用いる事が多くなっている。

□ グラフィックディスプレイを用いた化合物構造式の入力

化合物構造式のシステムへの入力には必要充分なだけの情報が簡単且つ性格に入力出来るという事が大切である。これまでに示した化合物入力手法は初期の目的をとげていると言える。しかしながら、計算機に不慣れな人が多い化学者にとり化合物構造式を簡単に入力出来るということは極めて重要な要素の一つとなる。この化学者に大きな負担をかけない入力うゆほうの解決策として、グラフィックディスプレイを用いた構造式の直接入力がある。

この手法によれば、化学者は普段利用している化合物構造式をそのままグラフィックディスプレイ上に描くだけで良い。システムへの入力に必要となる 構造式⇒数値データへの変換はシステムが自動的に行う。

ディスプレイからの化合物構造式入力の流れ



以下にディスプレイを用いた典型的な化合物構造式入力の例を示す。このシステムは米国MDL社の化合物検索システムであるMACCS (Molecular Access System)を例にとってある。

- ①画面上の一点にカーソルを置く、マウスのボタンを押すと画面上のその点が最初の原子の位置となる。
 - ②最初の原子とつながる原子の位置にカーソルを置き、再びボタンを強く押すとそこが2番目に指定された原子の位置となり、①で指定した原子との間に線が引かれる。
 - ③以下、①と②とを繰り返す事により化合物の構造式が画面上に描かれ、同時にシステムへの化合物構造式情報の入力が完了する。
- * 原子の種類について、炭素はしてはいる必要はない。炭素以外の原子はキーボードまたはディスプレイ上からの指定が必要である。

分子及び数値データの入力

以上示した手法は構造式に関する入力について示したが、この他に化学の分野で用いられる特有の入力データとしては分子式がある。分子式は原子主を示す記号と原子の数が添字として示されるが計算機への入力には添字を用いる事が出来ないので普通サイズの文字を用いる。又原子の種類順番については炭素原子を先頭とし、続いて水素原子を、3番目以降はアルファベット順に入力される。一般にこの形式での入力順番を“HILL ORDER”と言う。

分子式	C ₁₂ F ₂ H ₁₄ I ₂ N ₂ O ₆ S
HILL ORDER	C12H14F2I2O6S

数値データの入力には数値を書くべき場所及びその書き方が厳密に指定されたFORMATTED形式(書式付)の入力と、その制限がないFREE FORMAT形式(書式無)の入力とがある。

FORMATTED形式の入力では種々の制限があるので使用者はデータを入力するのが面倒であるが、プログラムそのものは簡単に作成できる。

FREE FORMAT形式では、連続して入力された数値データは一個のデータとみなされ、数値データの区別はスペースの存在により確認される。この形式であればユーザはデータの書く場所、入力形式等にとらわれる必要がないため数値データの入りは簡単となる。一方、プログラムの形式は複雑となり、作成は困難である。最近では、このFREE FORMAT形式による数値入力をサポートするサービスルーチンを備えるか、またはサブルーチンとして提供していることが多い。

FORMATTED 入力

```
1 2 3. 7 4 1 6 5 9 3. 6 1 E+5 -8 4. 4 4  
(プログラム)  
WRITE (6, 99) NO1, NO2, NO3, NO4  
FORMAT (F7. 3, 3X, I4, 5X, E7. 6, 2X, F6. 2)
```

FREE FORMAT入力

```
1 2 3. 7 4 1 6 5 9 3. 6 1 E+5 -8 4. 4 4
```

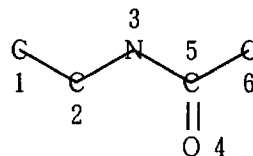
□ 化合物構造式の結合表 (CONNECTION TABLE) による表現

前記線型表記法は構造式が線型の文字データとして入力される為、構造式から文字データへの変換、および文字データから構造式への変換が大変な作業であった。このような欠陥をもちながらも利用されてきたのは、計算機のハード上での制約 (主としてメモリ) の比重が大きかった為である。計算機の進歩とともにこのような計算機に由来する制限は減少し、よりユーザフレンドリーな形式及び扱いやすさを主体とした構造式の変換様式が採用されるようになった。

当初採用された線型表記法にかわるものとして採用されてきたものが結合マトリクス表 (CONNECTION MATRIX) であり、それをより簡略な形にしたものが結合表 (CONNECTION TABLE) と呼ばれている。この様式では、計算機上でのメモリは多く必要となるが、特別な知識がなくとも構造式から結合表への変換、及び結合表から構造式への変換が容易という特徴を持つ。このため、この結合マトリクスおよび結合表様式は現在における化合物変換方式の基本となっている。

	1	2	3	4	5	6
1	C	1	0	0	0	0
2	1	C	1	0	0	0
3	0	1	N	0	1	0
4	0	0	0	O	2	0
5	0	0	1	2	C	1
6	0	0	0	0	1	C

結合マトリクス



結合マトリクスでの表現は、各原子に番号を付け、行列の番号とし、各項には結合の情報を入れ、又対称項には (X=Y) には原子の種類に関する情報を入力する。この時、水素原子は省略されることが多い。

このマトリクスを用いれば、立体及び幾何異性体の一部を除き全ての化合物が変換可能となる。変換ルールが簡単なので時間さえ気にしなければ誰にでも構造式→マトリクス→構造式の相互変換は可能である。

しかし、この表記法では構造式を表現するのに多くのメモリを必要 (N原子の時 N² 個) とし、特に原子数の多い化合物を扱う時には不経済である。従って、このマトリクスの良さを生かし、且つコンパクトに見易くしたのが結合表 (CONNECTION TABLE) である。これは前記マトリクスの情報に加え、化合物の3次元座標値、立体及び幾何異性体、チャージ、その他に関する情報等を表形式で表現するようにしたもので、以下にその一例を示す。

現在多くのシステムがこの結合表形式を採用している。これは計算機の進歩によりメモリが安くなり、比較的多くのメモリを必要としてもあまり問題とならなく成った事、線型表記法と比べた時に化合物の検索で不利とされていた幾つかの技術上における問題点 (化合物のユニークナンバリング、検索/データベース項参照) が解消された事、更には線型表記法では困難であった部分構造検索等が可能となった事などにより、この結合表を用いる事が可能となった為である。更に、この結合表には線型表記法を用いた時には不可能であった各原子及び結合間の相互関係、結合情報、3次元座標データその他の情報表、結合表による化合物の表現

原子番号	原子の種類	結合相手の原子番号及び結合の種類	立体不整の情報	UP/DOWN	幾何異性の情報	イオン等の情報	3次元座標データ
------	-------	------------------	---------	---------	---------	---------	----------

を簡単に利用出来る等の特徴がある為、化合物検索以外の目的/プログラム等にも幅広く用いることが可能である。

□ 結合表情報の統一

現在の化学システムの多くは前記結合表形式で化合物に関する情報を扱っていることを述べた。現在の環境で考えるならば一人の人が唯一のシステムだけを用いる事は少なく、

様々なシステムを相互に利用するということが多い。この事実は化合物情報を扱う結合表の形式を統一することの必要性を意味している。しかも、ローカルな範囲ではなく世界的なレベルで統一することが望ましい。

現在国際レベルの統一された機関により定義された結合表形式は存在していないが、現実的な問題として既に多くのシステムや機関で利用されているものとして「モルファイル」と呼ばれるものが存在する。

このモルファイルは計算機化学分野で老舗的な地位を持ち、化合物データベースという分野で圧倒的なシェアを有する米国MDL社が(MOLECULAR DESIGN LTD.)用いている化合物の表記形式である。この表記形式は当初一般には公表されていなかったが、1992年に正式に公表されて自由に利用することが可能となった。現在では特別な理由が無い限りこの「モルファイル」を用いるのが安全である。