

myPresto/VCOL 方式設計書
(Virtual Combinatorial Library 構築システム編)

第 1 版

目次 -

1. はじめに	2
1.1 システムの概要	2
1.2 システムの構成	3
2. 母核・側鎖構造登録部	4
2.1 入力情報	4
2.2 処理フロー	5
3. 仮想分子構造作成部	7
3.1 処理フロー	7
3.2 母核・側鎖の Z-matrix の結合方法	8
3.3 母核・側鎖の BOND 情報の結合方法	10
4. 使用するサンプルデータ	12

1. はじめに

本設計書では、母核分子に側鎖を自動的に付加することで、コンビナトリアルに分子を生成する「Virtual Combinatorial Library 構築システム」の方式について説明する。

1.1 システムの概要

母核構造および側鎖構造のデータファイル (mol2 形式) から、新規の仮想分子構造のデータファイル (mol2 形式) を生成するシステムである。本システムは、次の 2 つの処理部分からなる。

(1) 母核・側鎖構造登録部

母核構造および側鎖構造のデータファイル (mol2 形式) から、その構造情報を取り出してデータ化 (Z-matrix、BOND 情報) し、母核・側鎖構造データベースを作成する。

(2) 仮想分子構造作成部

母核・側鎖構造データベースから、母核構造とその置換基およびカテゴリー別に分けた側鎖構造セットを選択することで、指定置換基部分をセット内の各々の側鎖構造で置換した仮想分子を生成し、データファイル (mol2 形式) として出力する。

NOTE

母核・側鎖構造データベースは、XML 形式のフラットファイルとし、DBMS は使用しない。

1.2 システムの構成

本システムは、母核・側鎖構造登録部、仮想分子構造作成部の、2つの処理部分で構成する。以下に本システムの構成を示す。

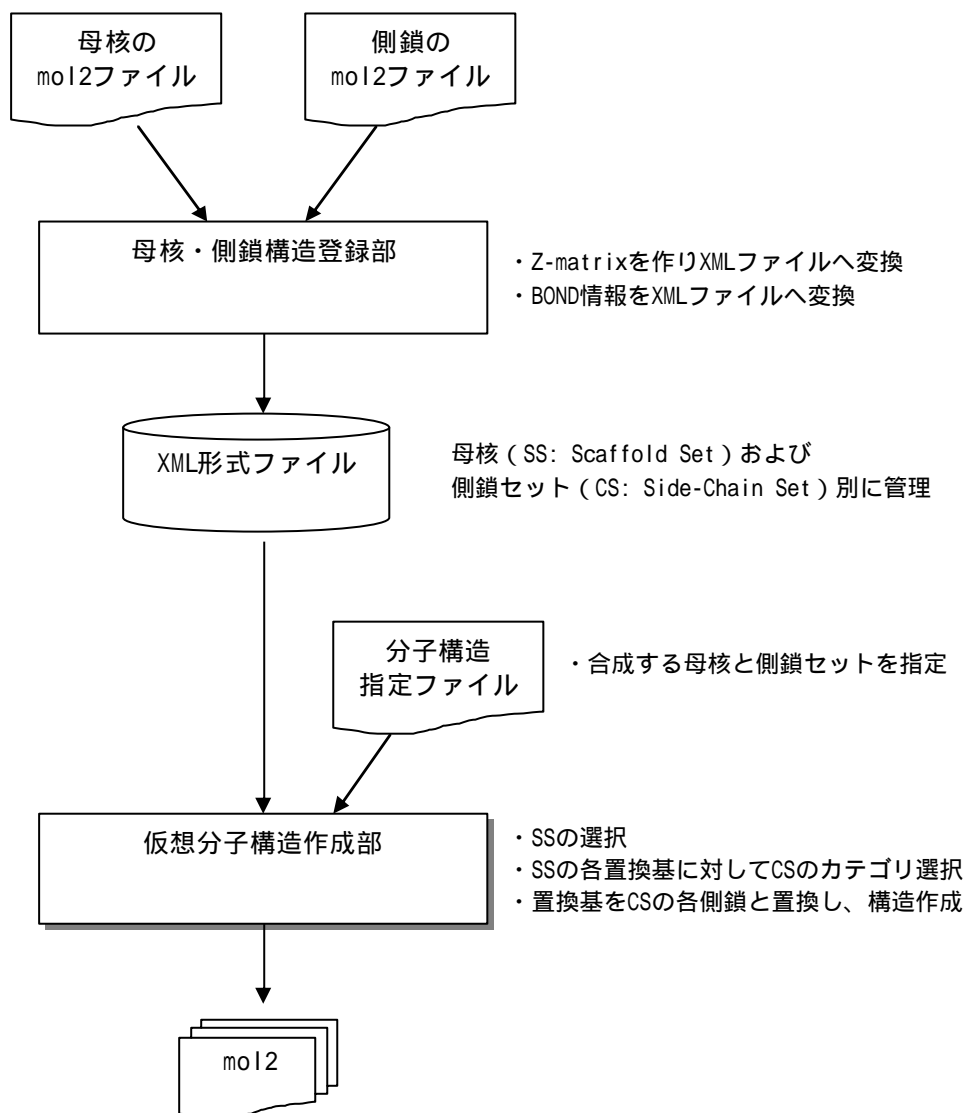


図 1 . システム構成

2. 母核・側鎖構造登録部

2.1 入力情報

母核および側鎖の登録時のファイルは以下の形式で準備することとする。

(1) 母核構造

母核構造の入力ファイルは、置換基が結合しうる部分をダミー原子「R」として作成したmol2ファイル形式とする。その際、原子タイプはR1, R2, ..., Rnを使用する。

複数の置換基（ダミー原子：R）を設定した場合、置換基は出現順序に従い、仮想分子合成プログラムの入力である側鎖セットの順序で置換される。

(2) 側鎖構造

側鎖構造の入力ファイルは、母核が結合しうる部分をダミー原子「X」として作成したmol2ファイル形式とする。その際、原子タイプはX1を使用する。

複数のダミー原子「X」は使用できないこととする。

2.2 処理フロー

母核構造および側鎖構造のデータファイル (mol2 形式) から、その構造情報を取り出してデータ化 (Z-matrix、BOND 情報) し、母核・側鎖構造データベースを作成する。

入力ファイル名、部分構造情報 (構造名、側鎖名) を入力する。

入力したmol2ファイルから、confgeneによりZ-matrixとbond情報のファイルを作成する。confgeneが出力したZ-matrixとbond情報、および、部分構造情報を含むXMLファイルを作成し、出力する。その際、カタログ番号を発行しXMLファイルに出力する。

作成されたXMLファイルを指定ディレクトリに格納する。

ログとして登録されたファイル名と、カタログ番号を出力する。

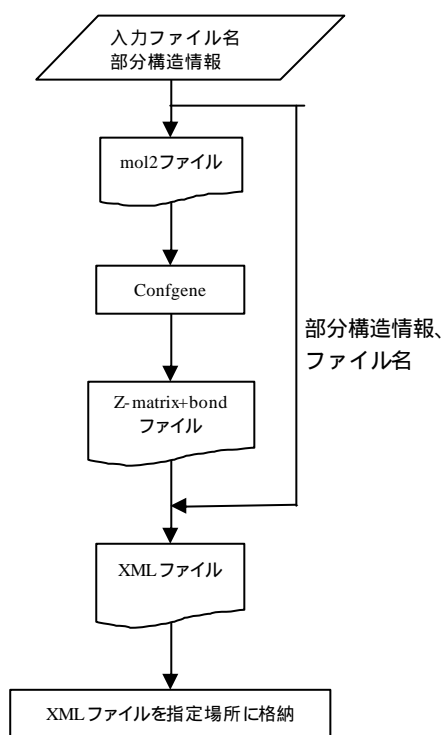


図 2 . 処理フロー

```
<vcl>
<id>
    4
</id>
<cs_id>
cs_002_01
</cs_id>
<set_id>
cs_002
</set_id>
<file>
    I\gzmat/CS/cs_002_01.gzmat
</file>
<gzmat>
X1      0  0.000000  0  0.000000  0  0.000000
C       1  1.7548858  0  0.000000  0  0.000000
H       2  1.1222162  1 109.7693253  0  0.000000
H       2  1.0809097  1 112.0950546  3 120.3639145
H       2  1.1065120  1 110.2582626  3 -116.9284515
#
    1      1      2      1
    2      2      5      1
    3      2      4      1
    4      2      3      1
</gzmat>
</vcl>
```

図3 .XML 形式ファイル例

3. 仮想分子構造作成部

3.1 処理フロー

母核・側鎖構造データベースから、母核構造とその置換基および側鎖セットを選択することで、指定置換基部分をセット内の各々の側鎖構造で置換した仮想分子を生成し、データファイル (mol2 形式) として出力する。

母核とその置換基および置換基に結合する側鎖セット情報 (SS_id, CS_id) を入力する。

入力の SS_id、CS_id から対象となる XML ファイルを検索する。

対象となる SS_id の XML ファイルより置換基の数を取得し、入力ファイルの CS_id の数 (同一 ID の場合もあり) と比較し、置換基より側鎖セット数が少ない場合、エラーを出力し、処理を中止する。

入力ファイルの指定どおりの組合せに従い、XML ファイルから母核、側鎖の Z-matrix を取得し、連結処理を行う (下記 3.2 節を参照)。

XML ファイルから各母核・側鎖の BOND 情報を取得し、連結処理を行う (下記 3.3 節を参照)。

連結した Z-matrix ファイルから原子座標を設定し、BOND 情報を加えた形での mol2 ファイルを出力する。

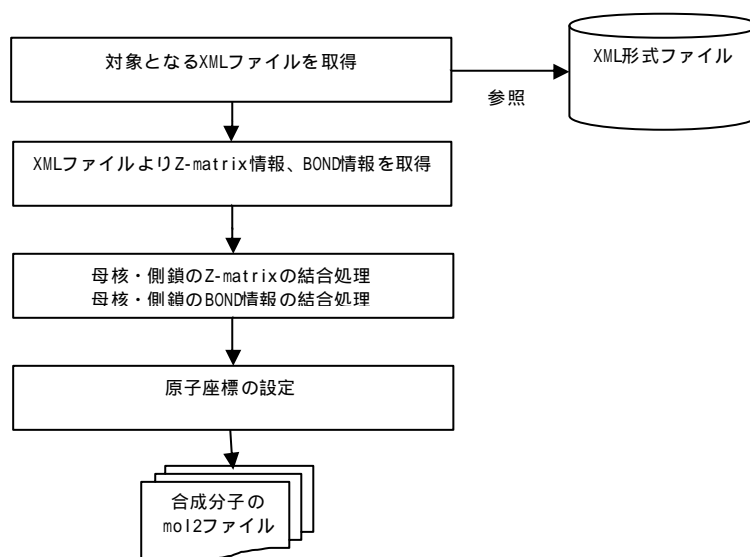


図 4 . 処理フロー

3.2 母核・側鎖の Z-matrix の結合方法

以下の方法に基づき、母核および側鎖の Z-matrix から、母核の置換基を側鎖の構造に置き換えた Z-matrix を作成する（図 4 (1) ~ (3) 参照）。

母核の元素数から 2 引いた数を、側鎖の各元素 ID=1,2 以外の ID に加える。

母核 Z-matrix より R1 の原子 ID (A) および、R1 が結合する原子 ID (C) を取得する。

側鎖 Z-matrix の原子 ID=1 を上記 で取得した (C) の原子 ID に置き換え、原子 ID=2 を (A) の原子 ID に置き換える。

R1 の結合角 (D)、二面角 (E) を成す原子の ID を取得する。

側鎖 Z-matrix の 2 行目の (D)、(E) の原子 ID を上記 の値に置き換える。

側鎖 Z-matrix の 3 行目の (E) の原子 ID を上記 の (D) の原子 ID の値に置き換える。

母核の R1 の行を、側鎖の 2 行目で置き換え、母核の最終行以降に側鎖の 3 行目以降を追加する。

さらに R_n ($n=2,3, \dots$) があれば、結合させる側鎖 Z-matrix との組合せで、上記 ~ の処理を行う。

(R_n がなくなるまで繰り返す)

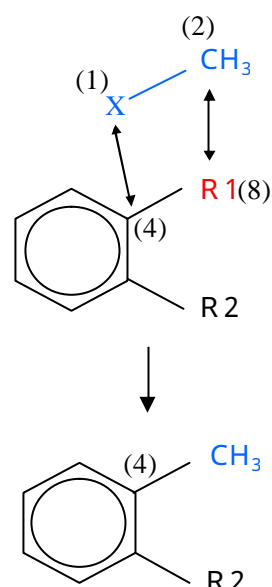
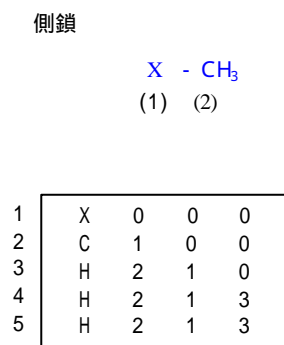
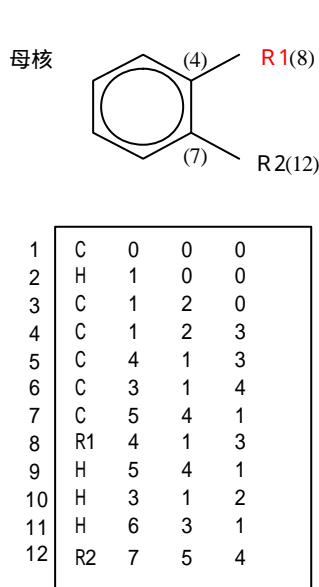


図 5 (1) 母核、側鎖の Z-matrix の例

図 5 (2) 母核、側鎖の連結

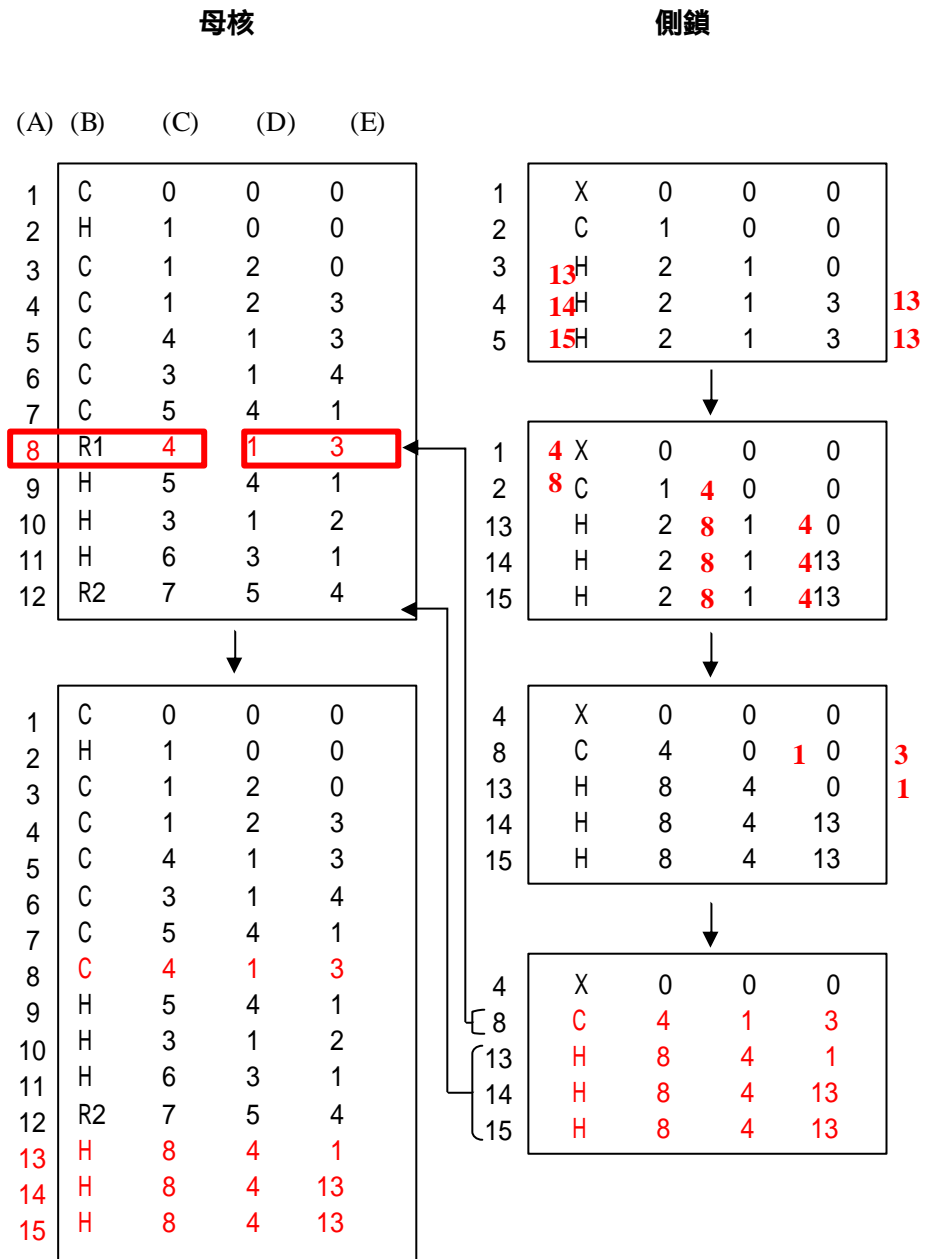


図5(3) 母核、側鎖のZ-matrixの結合アルゴリズム

(図中の ~ は、上記3.2節の本文の ~ に対応)

3.3 母核・側鎖の BOND 情報の結合方法

以下の方法に基づき、母核および側鎖の BOND 情報から、母核の置換基を側鎖の構造に置き換えた BOND 情報を作成する（図 5 (1) ~ (2) 参照）。

母核の元素数から 2 引いた数を、側鎖の各元素 ID=1,2 以外の ID に加える。

母核 Z-matrix より R1 の原子 ID および、R1 が結合する原子 ID を取得する（上記 3.2 節の の処理結果を利用）。

側鎖 Z-matrix の原子 ID=1 を上記 で取得した (C) の原子 ID に置き換え、原子 ID=2 を (A) の原子 ID に置き換える。

母核の最終行以降に側鎖の 2 行目を追加し、原子 ID でソートを掛ける。

さらに R_n ($n=2,3, \dots$) があれば、結合させる側鎖の BOND 情報との組合せで、上記 ~ の処理を行う。

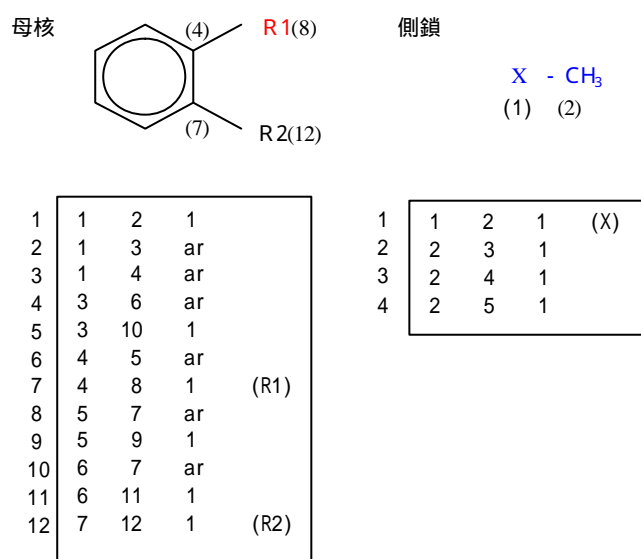


図 6 (1) 母核、側鎖の BOND の例

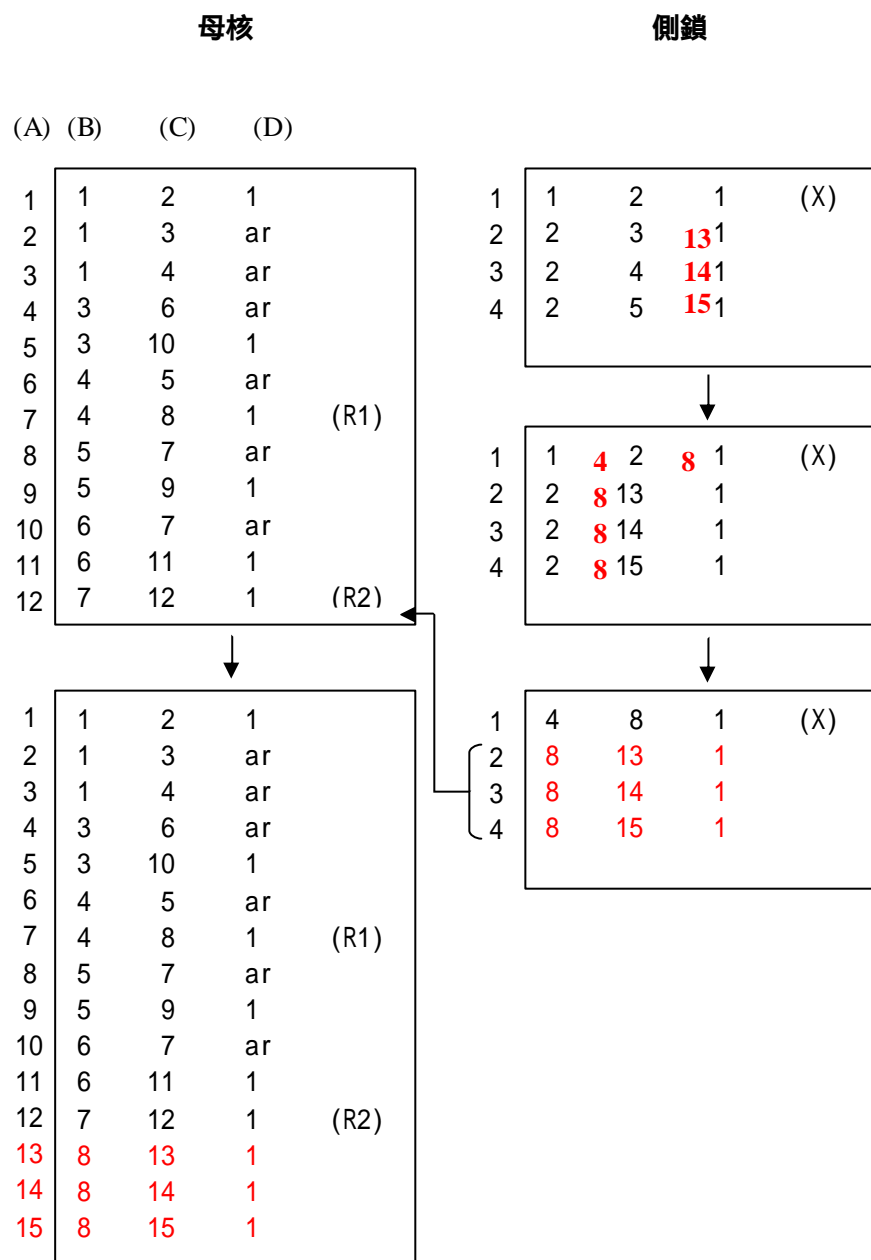


図 6 (2) 母核、側鎖の BOND 情報の結合アルゴリズム

(図中の ~ は、上記 3.3 節の本文の ~ に対応)

4. 使用するサンプルデータ

(1) 母核構造

資料 1 に示す母核セット (64 種) の mol2 ファイルを使用する。

(2) 側鎖構造

資料 2 に示す構造の 9 種の側鎖セット (合計 95 種) の mol2 ファイルを使用する。