myPresto 4.2

- VCOL -

USER MANUAL

Version 1.0

About This Document

This document accompanies the “myPresto 4.2 USER MANUAL”. Copyrights, program licensing conditions, the authors, and references are as stated in the “myPresto 4.2 USER MANUAL”.

Acknowledgements

This work was supported by grants from New Energy and Industrial Technology Development Organization (NEDO) and Ministry of Economy, Trade and Industry (METI), JAPAN.

This software was developed as part of a research project advanced by the late Dr. Yoshimasa Kyogoku.

Table of Contents

(Blank)

1 Introduction

1.1 Overview of system

Using data files (in mol2 format) of scaffold structures and side chain structures, the Virtual Combinatorial Library building system (VCOL) generates data files (in mol2 format) for new virtual molecular structures. The system consists of the following two processing components:

(1) Scaffold and side chain structure registration component

This component extracts structural information from data files (in mol2 format) of scaffold structures and side chain structures, converts the information to data (Z-matrix, BOND information), and creates a scaffold and side chain structure database.

(2) Virtual molecular structure building component

From the scaffold and side chain structure database, this component selects scaffold structures, substituents in those structures, and sets of side chain structures grouped by category. It then generates virtual molecules in which the specified substituent parts are replaced by the side chain structures in the set, and outputs these as a data file (in mol2 format).

■NOTE

The scaffold and side chain structure database is a flat file in XML format; DBMS is not used.

1.2 System structure

The system consists of two processing components: the scaffold and side chain structure registration component and the virtual molecular structure building component. The structure of the system is shown below.

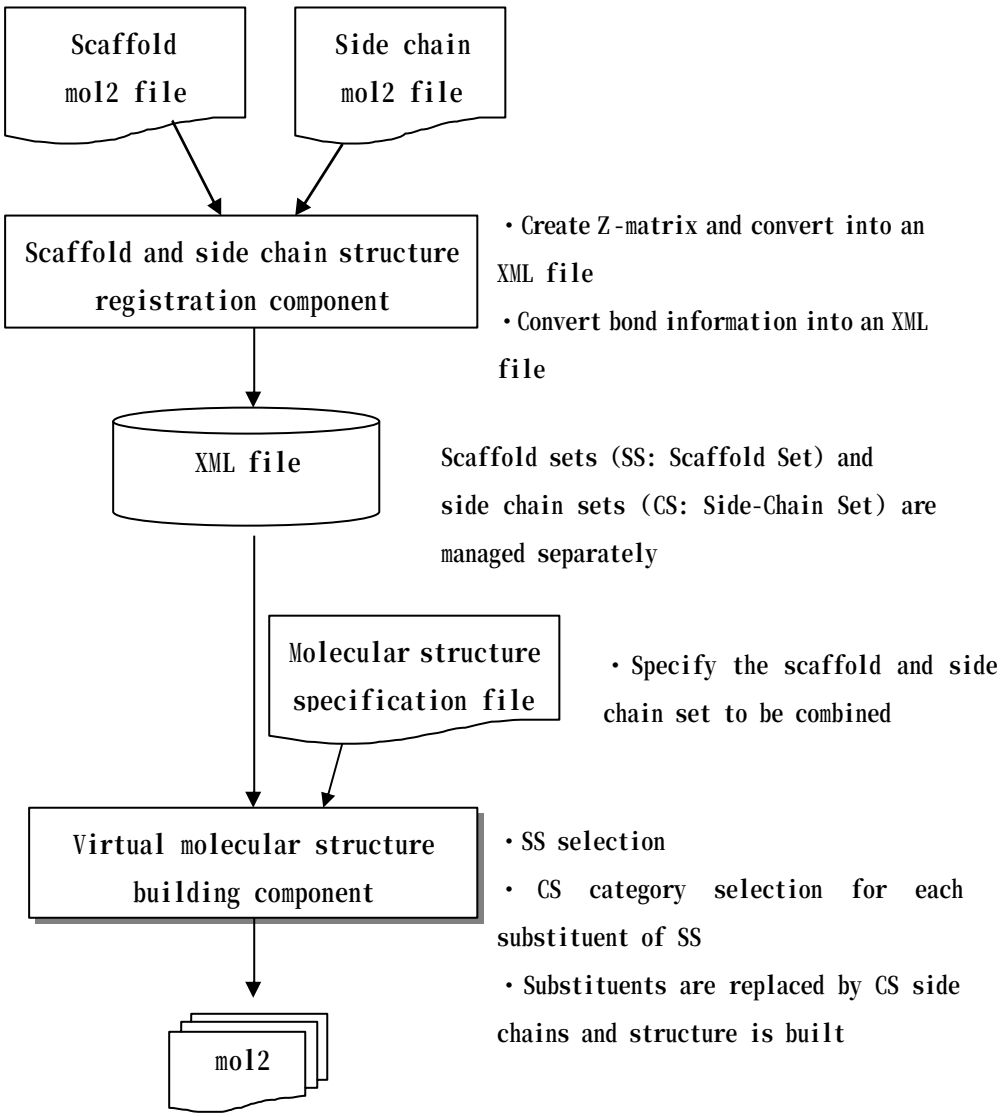


Fig. 1. System structure

2. Scaffold and side chain structure registration component

2.1 Input information

To register the scaffold and side chains, prepare a file in the format below.

（1）Scaffold structure

The scaffold structure input file should be in mol2 format with a dummy atom “R” used for the parts with which the substituent can bond. R1, R2, …, Rn are used as the atom types.

※If multiple substituents (dummy atom: R) are established, the substituents are replaced in the order of their appearance using the order of the side chain set that is the input of the virtual molecule combinatorial program.

（2）Side chain structure

The side chain structure input file should be in mol2 format with a dummy atom “X” used for the part with which the scaffold can bond. X1 is used for the atom type.

※Multiple dummy atoms “X” cannot be used.

2.2 Processing flow

Structural information is extracted from the scaffold structure and side chain structure data files (mol2 format), the information is converted to data (Z-matrix, bond information), and a scaffold and side chain structure database is created.

①Input the input file name and the part structure information (structure name, side chain name).
②Create a Z-matrix and bond information file from the input mol2 file using confgene.
③Create an XML file that includes the Z-matrix and bond information output by confgene and the part structure information, and output the file. At this time, issue a catalogue number and output it to the XML file.
④Store the created XML file in the specified directory.
⑤Output the registered file name and catalogue number as a log.

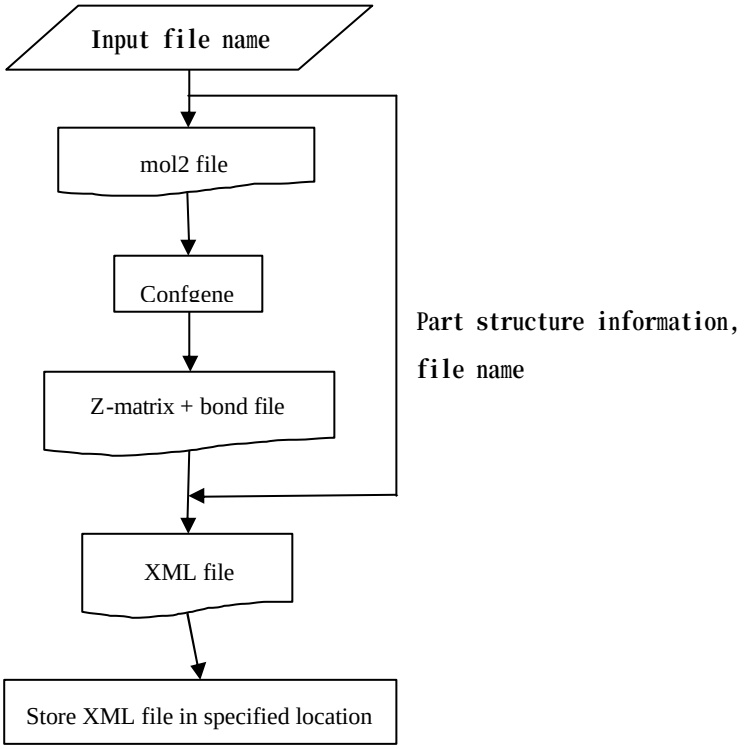


Fig. 2. Processing flow

```
<vcl>
<id>
        4
</id>
<cs_id>
cs_002_01
</cs_id>
<set_id>
cs_002
</set_id>
<file>
        INgzmat/CS/cs_002_01.gzmat
</file>
<gzmat>
X1        0  0.0000000   0   0.0000000   0    0.0000000
 C        1  1.7548858   0   0.0000000   0    0.0000000
 H        2  1.1222162   1 109.7693253   0    0.0000000
 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>
```

Fig. 3. Example of XML file

3. Virtual molecular structure building component

3.1 Processing flow

Scaffold structures, substituents in those structures, and sets of side chain structures are selected from the scaffold and side chain structure database to generate virtual molecules in which the specified substituent parts are replaced by the side chain structures in the set, and these are output as a data file (in mol2 format).

①Input the scaffold, the scaffold substituents, and the side chain set to be bonded to the substituents (SS_id, CS_id).
②Search for the target XML file in the input SS_id and CS_id.
③Obtain the number of substituents from the target SS_id XML file and compare this to the number of CS_id in the input file (some side chains may have the same ID). If the number of side chain sets is less than the number of substituents, output an error and stop processing.
④Obtain the scaffold and side chain Z-matrices from the XML file according to the combinations specified in the input file, and perform bonding processing (see section 3.2 below).
⑤Obtain the bond information of each scaffold and side chain from the XML file, and perform bonding processing (see section 3.3 below) .
⑥Set the atom coordinates from the bonded Z-matrix file and output a mol2 file with the added bond information.

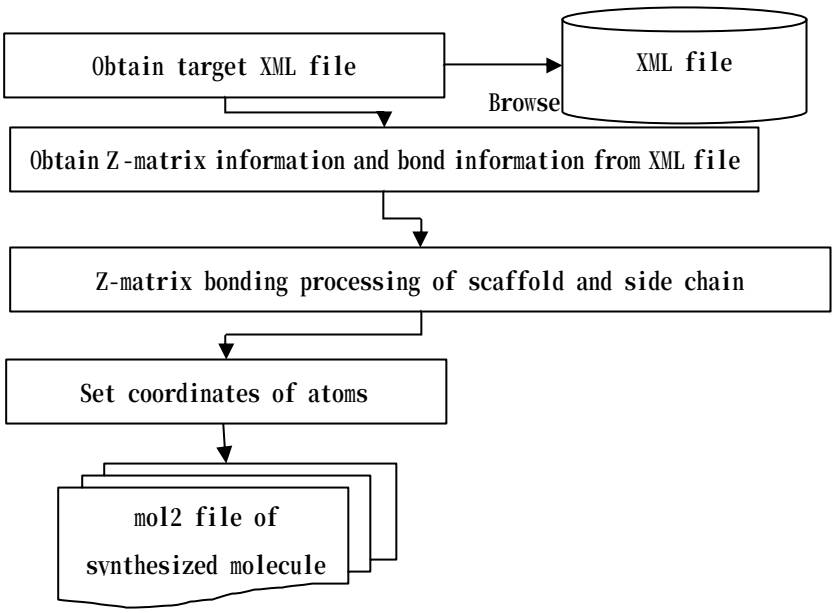
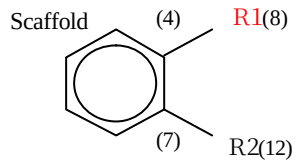


Fig. 4. Processing flow

3.2 Procedure for bonding the scaffold and side chain Z-matrix

Using the scaffold and side chain Z-matrices, follow the procedure below to create a Z-matrix that replaces the substituents of the scaffold with the side chain structures (see Fig. 4 (1) to (3)).

①Subtract 2 from the number of elements of the scaffold and add this number to IDs other than ID=1, 2 of the elements of the side chain.
②From the scaffold Z-matrix, obtain atom ID (A) of R1 and atom ID (C) to which R1 bonds.
③Replace ID=1 atoms in the side chain Z-matrix with the atom ID of (C) obtained above in ②, and replace ID=2 atoms with the atom ID of (A).
④Obtain the ID of the atom that forms the bonding angle (D) of R1 and the ID of the atom that forms the dihedral angle (E).
⑤Replace the atom IDs of (D) and (E) on the 2nd line of the side chain Z-matrix with the values of ④ above.
⑥Replace the atom ID of (E) on the 3rd line of the side chain Z-matrix with the value of the atom ID of (D) in ④ above.
⑦Replace the R1 line of the scaffold with the 2nd line of the side chain, and add the 3rd and following lines of the side chain to the scaffold after the last line.
⑧If there are more Rn (n=2,3, ・・・) , repeat ① to ⑦ above using each combination of Rn and the Z-matrix of the side chain to be bonded.
(Repeat until there are no more Rn.)



Side chain

X －CH_3

(1) (2)

1	C	0	0	0
2	H	1	0	0
3	C	1	2	0
4	C	1	2	3
5	C	4	1	3
6	C	3	1	4
7	C	5	4	1

1	X	0	0	0
2	C	1	0	0
3	H	2	1	0

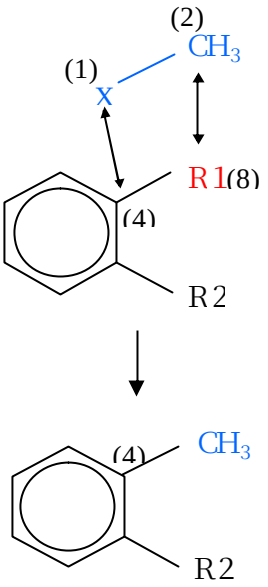


Fig. 5 (1) Example of scaffold and side chain Z-matrix

Fig. 5 (2) Linking of scaffold and side chain

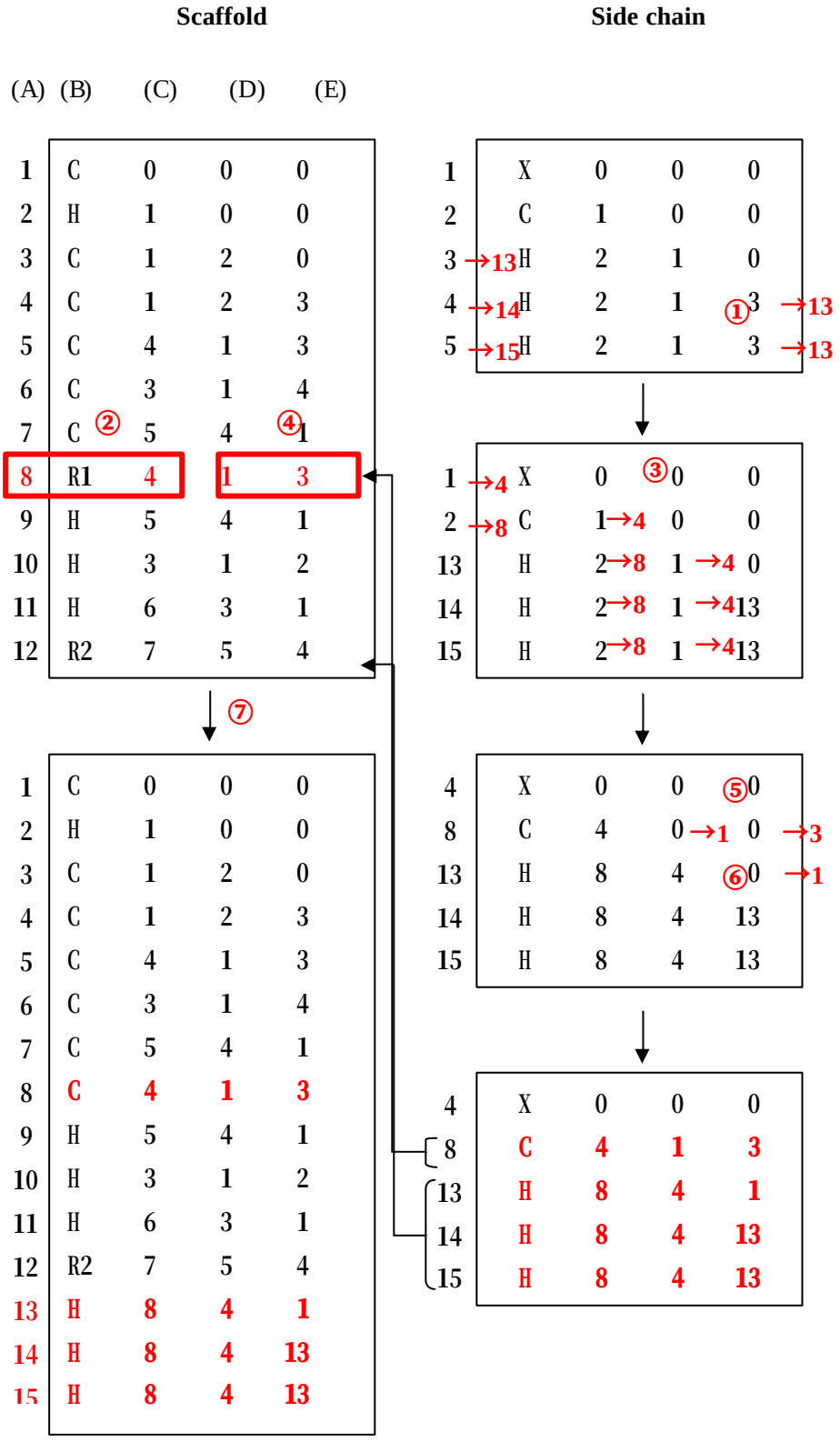


Fig. 5(3) Bonding algorithm of scaffold and side chain Z-matrices

(① to ⑦ above correspond to ① to ⑦ in the body text of section 3.2)

3.3 Procedure for bonding scaffold and side chain bond information

Using the scaffold and side chain bond information, follow the procedure below to create bond information that replaces the scaffold substituents with side chain structures (see Fig. 5 (1) and (2)).

①Subtract 2 from the number of elements of the scaffold and add this number to IDs other than ID=1, 2 of the elements of the side chain.
②Obtain the atom ID of R1 and the ID of the atom to which R1 bonds from the scaffold Z-matrix.
(Use the results of ② in section 3.2.)
③Replace ID=1 atoms in the side chain Z-matrix with the atom ID of (C) obtained above in ②, and replace ID=2 atoms with the atom ID of (A).
④Add the 2nd and following lines of the side chain to the scaffold after the last line, and sort by atom ID.
⑤If there are more Rn (n=2,3, ・・・) , repeat ① to ④ above using each combination of Rn and the bond information of the side chain to be bonded.

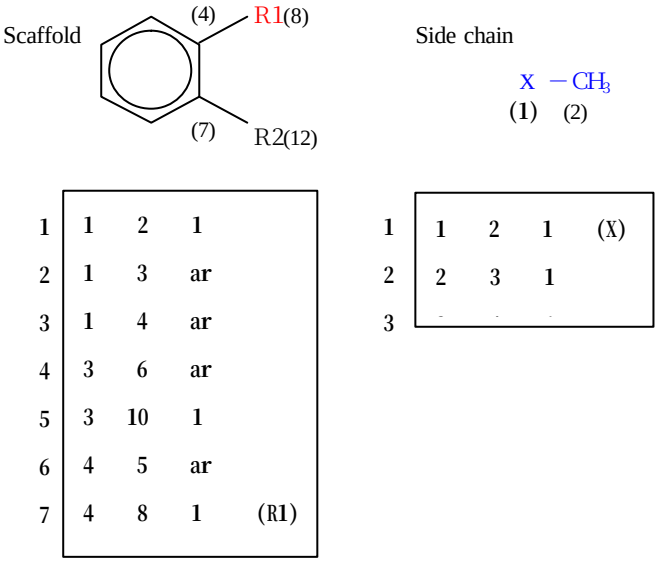


Fig. 6 (1) Example of scaffold and side chain bond

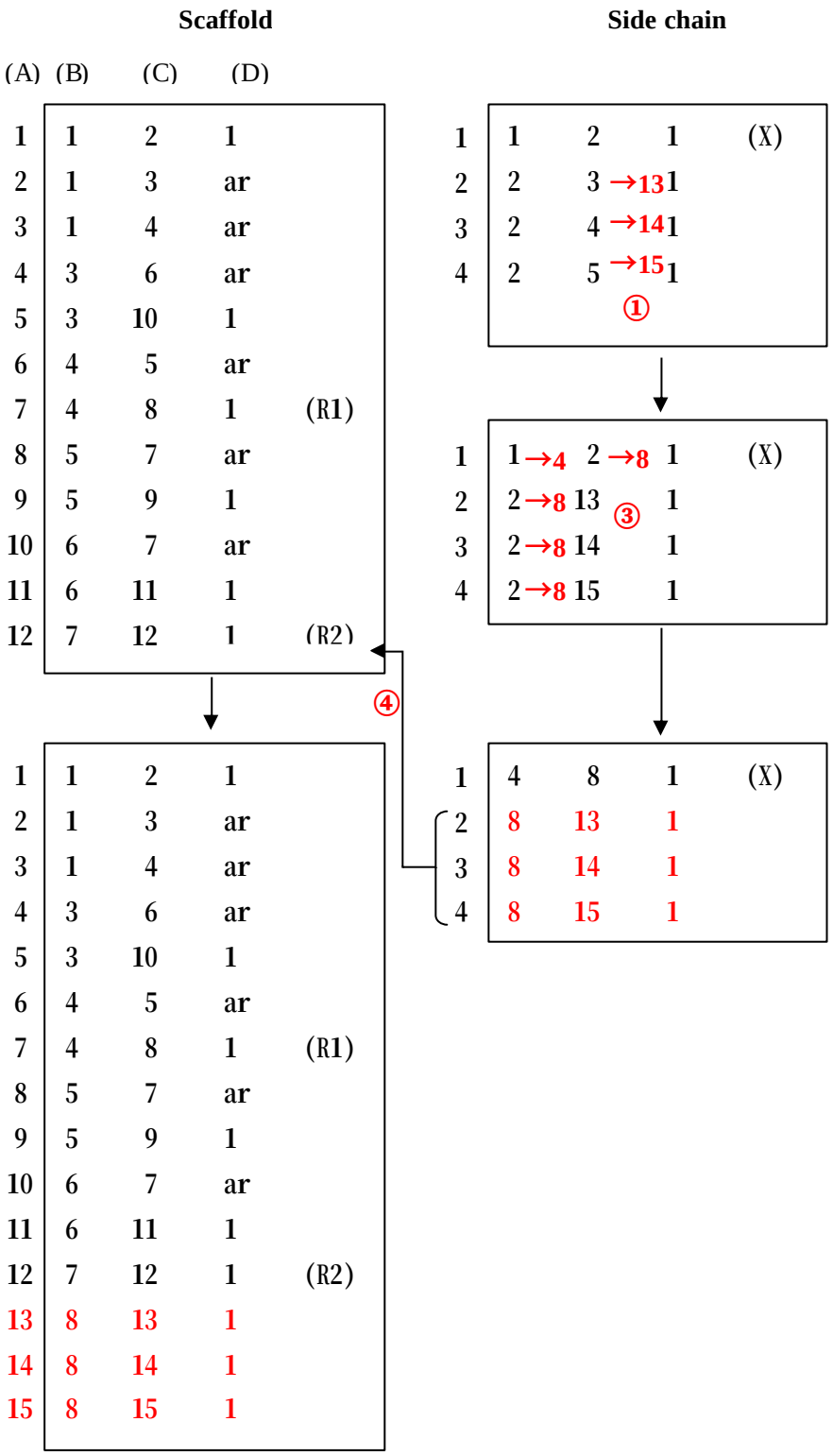


Fig. 6 (2) Bonding algorithm of scaffold and side chain bond information

(① to ④ above correspond to ① to ④ in the body text of section 3.3.)

4 Procedures for Using VCOL

4.1 Conformation generation tool (confgene)

This tool reads in a mol2 file and creates a Z-matrix format file that expresses the coordinates of the atoms that compose a molecule using the three parameters of bond distance, angle, and torsion.

When multiple conformations are specified, rotatable torsions are found and a Z-matrix file is generated for the conformations obtained by rotating these torsions.

4.1.1 Input

(1)Control file

The control file consists of four lines with the following items on each line:

(1-1) mol2 file name

(1-2) Number of generated conformations (when this is 1, the torsion is not rotated)

(1-3) Number of torsion rotation candidate angles

(1-4) Value of total electrical charge (when this is "a", the value is calculated automatically)

Example of control file)

```
LIG.mol2          ; Input mol2 file name
1                 ; Number of generated conformations
3                 ; Torsion rotation candidate angles(each rotation is360 ° /3 = 120)
a                 ; Automatic calculation of total charge
```

(2) mol2 file

Example of "CH_3" mol2 file)

```
@<TRIPOS>MOLECULE
m1
     4     3     1
SMALL
USER_CHARGES
@<TRIPOS>ATOM
     1 C1        5.2414    -3.0575    -0.0645 C.3       1  RES     0.2900
     2 H11       5.5684    -3.0437    -1.1215 H         1  RES     0.0000
     3 H12       5.6242    -3.9411     0.4265 H         1  RES     0.0000
     4 H13       5.7021    -2.1572     0.4219 H         1  RES     0.0000
@<TRIPOS>BOND
     1     1     2 1
     2     1     3 1
     3     1     4 1
@<TRIPOS>SUBSTRUCTURE
     1 RES         1
```

4.1.2 Output

The file with the file name "zmat.dat" is output in the format below.
“n” in the table is the number of atoms, and "m” is the number of bonds.

Item no.	Line	Column	Description	Type	Remarks
#1	1 to n	1	Atom name	Text	-
#2		2	Line of bond partner atom	Integer	First atom is 0
#3		3	Bond distance	Real number	First atom is 0.0
		4	Line of angle partner atom	Integer	Atom nos. 1 and 2 are 0
		5	Angle	Real number	Atom nos. 1 and 2 are 0.0
		6	Line of torsion partner atom	Integer	Atom nos. 1 to 3 are 0
		7	Torsion angle	Real number	Atom nos. 1 to 3 are 0.0
	n+1	1	Bond start indicator line	"#"	-
	n+2 to n+m+1	1	Bond consecutive number	Integer	-
		2	Line of atom 1 of bond	Integer	-
#4		3	Line of atom 2 of bond	Integer	-
		4	Bond order	Real number	-

Example of "CH_3" Z-matrix file)

```
 C       1  1.7548858   0   0.0000000   0    0.0000000
 H       2  1.1222162   1 109.7693253   0    0.0000000
 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        1        3    1
       3        1        4    1
```

4.2 Virtual molecule building tool

This tool reads in the Z-matrix files of the scaffold and side chain structure and outputs a mol2 file of the virtual molecule that synthesizes the two.

When multiple substituents exist in the scaffold structure, multiple side chain structures are synthesized.

4.2.1 Input

(1) Control file

The control file has the following number of lines: one line each for (1-1) and (1-2); (1-3) requires a number of lines equal to the number of side chains to be substituted.

(1-1) mol2 file output directory

The output file name is generated as a combination of the scaffold and side chain names.

(1-2) Scaffold Z-matrix file name

Specify “R” for the name of the atom in the scaffold to be replaced.

(1-3) Side chain Z-matrix file name

Specify "Xn" for the atom name of the side chain dummy atom (where ”n” is any text string) .

Control file example)

```
output/mol2              ; Output directory name
scaford.zmat             ; Scaffold Z-matrix file name
side1.zmat               ; Side chain 1 Z-matrix file name
side2.zmat               ; Side chain 2 Z-matrix file name
```

4.3.2 Output

The output mol2 file of the synthesized molecule will have the following name:

Scaffold name + "_" + side chain 1 name +...+ "_" + side chain n name

Example of mol2 file of synthesized molecule)

ss_01_1_cs_001_01.mol2 file that synthesizes ss_01_1.zmat and cs_001_01.zmat

```
@<TRIPOS>MOLECULE
ss_01_1_cs_001_01
    17    18     2
SMALL
NO_CHARGES
@<TRIPOS>ATOM
     1     H     0.0000   0.0000   0.0000     H    1 SS     0.0000
     2     C     1.0736   0.0000   0.0000     C    1 SS     0.0000
     3     C     1.7789   1.2162   0.0000     C    1 SS     0.0000
     4     H     1.2509   2.1590   0.0412     H    1 SS     0.0000
     5     C     3.1795   1.1458  -0.0346     C    1 SS     0.0000
     6     C     4.1586   2.2592  -0.0426     C    1 SS     0.0000
     7     H     4.0404   2.8543   0.9024     H    1 SS     0.0000
     8     H     4.0063   2.9218  -0.8996     H    1 SS     0.0000
     9     C     5.4272   1.5305  -0.0832     C    1 SS     0.0000
    10     F     6.7350   2.3229  -0.0762     F    2 C01    0.0000
    11     C     5.2833   0.2174  -0.1184     C    1 SS     0.0000
    12     H     6.0502  -0.5698  -0.1651     H    1 SS     0.0000
    13     C     3.8517  -0.0497  -0.1037     C    1 SS     0.0000
    14     C     3.1274  -1.2559  -0.0963     C    1 SS     0.0000
    15     H     3.6540  -2.2315  -0.1411     H    1 SS     0.0000
    16     C     1.7315  -1.2217  -0.0819     C    1 SS     0.0000
    17     H     1.1885  -2.1543  -0.0988     H    1 SS     0.0000
@<TRIPOS>BOND
     1     1     2  1
     2     2    16  2
     3     2     3  1
     4    16    17  1
     5    14    16  1
     6    14    15  1
     7    13    14  2
     8     5    13  1
     9    11    13  1
    10     3     5  2
    11     5     6  1
    12     3     4  1
    13     6     8  1
    14     6     7  1
    15     6     9  1
    16     9    10  1
    17     9    11  2
    18    11    12  1
@<TRIPOS>SUBSTRUCTURE
     1  SS       1
     2  C01     10
```

***myPresto* 4.2**

- VCOL -