

MiCMoS

Milano Chemistry Molecular Simulation

Description and User Manual

Tutorials

T3. Using the *Retcor* module as a molecule builder

PURPOSE: In this tutorial, it is described how to build chemically sound molecular models for systems whose atomic coordinates are unknown for subsequent use in Monte Carlo and Molecular Dynamics calculations.

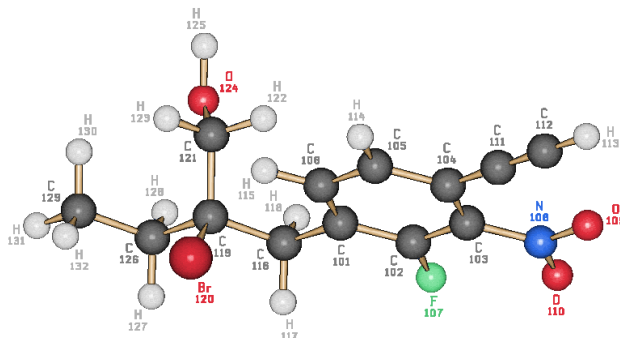
REFERENCE to main manual: Sections 1.2 (*Retcor*); 1.4.3 (.oih file); Figure 1.2 (chemical group reference); Appendix, Section A2 (full *Retcor* algorithm and reference Table A2.1).

FILES: You can find all the files you need to run this tutorial here:

https://sites.unimi.it/xtal_chem_group/images/MiCMoS_package/T3/large.oih

The starting point is the large.oih structure file, which must be prepared by hand from known explicit coordinates of three atoms. You will also need a macro to run the module *Retcor*. Place the .oih file and pertinent (Windows/Unix) macros into your working directory.

Consider the following problem. You want to prepare a .oeh file with explicit Cartesian coordinates for the large molecule shown in the picture below, for example as you are interested in performing a Molecular Dynamics simulation. Unfortunately, no entries are available in the Cambridge Structural Database and you don't know explicit atomic coordinates.



You might tackle the problem with a modern molecular building software, such as Avogadro (<https://avogadro.cc/>) or GaussView (<https://gaussian.com/gaussview6/>). However, you should be aware that MiCMoS routine *Retcor* can be used as a molecular builder on its own. The advantage of using *Retcor* is that it provides a correctly formatted .oeh structure file, ready for subsequent elaboration, at a cost of minimum intervention by the user.

The first step is to prepare an input .oih file. Open an empty large.oih file.

```
vi large.oih
```

Then, following the manual (Section 1.4.3), start typing the title line.

CAUTION: The first line of the *.oih* file is formatted: 1x,10a4,f8.3,f8.1,3x,f5.0,1x,f6.1. All other lines are free. The first character is skipped (1x), so fill it with "#". Then, you have 40 characters to type in the title - any symbol allowed. The next entries concern experimental crystal density, temperature, year of the corresponding Literature reference and crystallographic agreement factor, so you can safely skip them. In practice, **the only relevant entry is title.**

For example, a valid 25 character-long title line is:

```
1234567890123456789012345678901234567890
```

```
#LARGER99 large molecule
```

From now on, all lines are in free format. The second one is a dummy (type 0), and the third one contains the metric system. As you have an isolated molecule, type 1.0 1.0 1.0 90.0 90.0 90.0 (the basic Cartesian frame). The fourth line should contain the experimental sublimation enthalpy of the crystal; again, type 0.0. Finally, your file should look like:

```
#LARGER99 large molecule
0
1.0000 1.0000 1.0000 90.00 90.00 90.00
0.0
```

The next line is the number of atoms with explicit x, y, z coordinates. You must add coordinates of atoms 1-3, which are the minimal basis for construction of all other atomic positions. For each line, type in the sequential number of the *i*th atom, the Cartesian coordinates in Å, fragment id (=1, as you have only one molecular fragment), atom specie indicator according to Table 1.1 (manual, Section 1.4.2) and atomic charge (=0.0, as charges are unknown).

```
3
1 0.00000 1.39000 0.00000 1 12 0.0000
2 1.20378 0.69500 0.00000 1 12 0.0000
3 1.20378 -0.69500 0.00000 1 12 0.0000
```

The phenyl ring in the "large" compound is a good choice to start, but you could have started as well from any other point of the molecule. The coordinates above written correspond to three equispaced (1.390 Å) C atoms, with C1-C2-C3 angle of 120°: they are good to define any kind of flat 6-membered rings. The following configuration can be used for 3 equispaced (1.525 Å) sp³ carbons with C1-C2-C3 angle of 109.47° instead:

```
1 0.00000 0.00000 0.00000 1 13 0.
2 0.00000 1.52500 0.00000 1 13 0.
3 1.41396 2.09628 0.00000 1 13 0.
```

CAUTION: Avoid using 3 atoms lying exactly on the same line.

Now you have to specify the number of atoms/groups that must be defined by *Retcor*. It is difficult for non-experienced users to know in advance what this number should be. Type "#" for the moment and keep on preparing the *.oih* file. You must remember to update this parameter later.

From now on, you have to compile several lines specifying various structure-defining parameters. Each line will contain (i) 6 identification integer codes ("n1-n6") according to Figure 1.2 (main manual) and Figure A2.1 (Appendix); (ii) molecular fragment id (1 in this case); (iii) atomic specie indicator ("ispen") according to Table 1.1 (manual, Section 1.4.2); (iii) atom charge ("qrg", 0.0 for now); (iv)-(vi) Distance ("R"), torsion angle ("tors") and bond angle ("alph") parameters. These three latter numbers might be needed or not, depending on the sequence of 6 identification code numbers written before (see Tables 1.2 and A2.1).

CAUTION: The user must make sure that all atoms n3-n6 that are listed in the identification code have been generated before asking for the generation of atom n1, irrespective of sequential number.

The benzene ring atoms 4-6 are built in chain from explicit atoms 1-3 using the **"Z-matrix" mode**. This means that $n2 = -1$ and $n3 = 0$; $n4$, $n5$ and $n6$ correspond to the sequence that allows to define torsion and bond angles. We want that all C-C bonds have the same length ($R=1.39 \text{ \AA}$) and angles ($n1-n4-n5 = 120^\circ$), being part of a flat ring ($n1-n4-n5-n6$ torsion= 0°).

n1	n2	n3	n4	n5	n6	id	ispen	qrg	R	tors	alph
#											
4	-1	0	3	2	1	1	12	0.0000	1.390	0.00	120.0
5	-1	0	4	3	2	1	12	0.0000	1.390	0.00	120.0
6	-1	0	5	4	3	1	12	0.0000	1.390	0.00	120.0

Fluorine 7 and nitrogen 8 are built on the bisector of the CCC benzene angle, F7 attached to C2 and N8 to C3. In this case, torsion and alpha angles are dummies (**"bisector" mode**, $n2=n3=0$):

n1	n2	n3	n4	n5	n6	id	ispen	qrg	R	tors	alph
7	0	0	2	1	3	1	41	0.	1.300	0.0	0.0
8	0	0	3	2	4	1	19	0.	1.400	0.0	0.0

Nitro oxygens 9 and 10 are built attached to N8, at a distance of 1.22 \AA , one with torsion 0° and one with 180° , and trigonal 120° alpha angles (**"Z-matrix mode"**):

n1	n2	n3	n4	n5	n6	id	ispen	qrg	R	tors	alph
9	-1	0	8	3	2	1	30	0.0000	1.220	180.00	120.000
10	-1	0	8	3	2	1	30	0.0000	1.220	0.00	120.000

The acetylenic C11 is built on the bisector of the C3-C4-C5 angle (**"bisector mode"**). The second acetylenic carbon, C12, at a distance of 1.20 \AA from C11, along the C11-C4 direction (**"extension" mode**, $n2,n3,n6=0$), and the acetylenic hydrogen H13 in the same way, along the C12-C11 direction; tors and alpha are dummies:

n1	n2	n3	n4	n5	n6	mlc	ispn	qrg	R	tors	alph
11	0	0	4	3	5	1	11	0.0000	1.300	0.00	0.000
12	0	0	11	4	0	1	11	0.0000	1.200	0.00	0.000
13	0	0	12	11	0	1	1	0.0000	1.080	0.00	0.000

Let's apply once again the **bisector mode** preparation for H14, H15 and C16:

n1	n2	n3	n4	n5	n6	id	ispn	qrg	R	tors	alph
14	0	0	5	4	6	1	2	0.0000	1.080	0.00	0.000
15	0	0	6	5	1	1	2	0.0000	1.080	0.00	0.000
16	0	0	1	2	6	1	13	0.0000	1.450	0.00	0.000

Atom C19 (Z-matrix) is coplanar to the ring, and must be generated before H17 and H18 with the **Z-matrix mode**:

n1	n2	n3	n4	n5	n6	id	ispn	qrg	R	tors	alph
19	-1	0	16	1	2	1	13	0.0000	1.500	180.00	109.470

Note that n2=0 or -1 up to now. Thus, each implicit group contains one atom only. With the next instruction, we exploit the **"methylene" mode** (n3=0), to define at once the two hydrogens H17 and H18 of the CH₂ group.

n1	n2	n3	n4	n5	n6	id	ispn	qrg	R	tors	alph
17	18	0	16	1	19	1	3	0.0000	1.080	0.00	109.470

Note that hydrogens are bonded to C16. C1 and C19 serve to define connectivity, as H17 and H18 are in the plane bisecting the C1-C16-C19 angle. The torsion angle is a dummy, and *alph* is the HCH angle.

Atom C26 must be generated before C21 with the **Z-matrix mode**:

n1	n2	n3	n4	n5	n6	id	ispn	qrg	R	tors	alph
26	-1	0	19	16	1	1	13	0.0000	1.500	180.00	109.470

As C26 is now defined, the atom C21 can be generated as the fourth member of a pyramidal group, in which the three angles at the atom at the top of the pyramid are equal (**"pyramidal" mode**, n2 = -2, n3=0); tors and alph are dummies:

n1	n2	n3	n4	n5	n6	id	ispn	qrg	R	tors	alph
21	-2	0	19	16	26	1	13	0.0000	1.500	0.00	0.000

Atom Br20 is generated from atom C19, so that the three angles C16-C19-Br20, C21-C19-Br20 and C26-C19-Br20 are equal (**"methyne" mode**, n2=0); tors and alph are dummies. In practice, the group is tetrahedral:

n1	n2	n3	n4	n5	n6	id	ispn	qrg	R	tors	alph
20	0	19	16	21	26	1	43	0.000	1.700	0.00	0.0

The hydroxy oxygen O24 can be generated using one **Z-matrix** instruction:

It is time to close the *.oih* file with the space group block instructions. As the molecule is isolated, neither point nor translation symmetries are present. In the next line after the implicit atom input block type "1": this is the number of symmetry operations (the identity alone). In the next line, type in the identity matrix by row (1.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 1.0). Add in the next line the 0.000 0.000 0.000 translation vector. Then, add another line and type "0": there is no need to change the reference system for the isolated molecule. To conclude, add another line and type "0" again, as you have not non-library 6-12 parameters.

```
1
1.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 1.0
0.0000 0.0000 0.0000
0
0
```

The final *.oih* file should look like as follows:

```
#LARGER99 large molecule
0
1.0000 1.0000 1.0000 90.00 90.00 90.00
0.0
3
1 0.00000 1.39000 0.00000 1 12 0.0000
2 1.20378 0.69500 0.00000 1 12 0.0000
3 1.20378 -0.69500 0.00000 1 12 0.0000
27
4 -1 0 3 2 1 1 12 0.0000 1.390 0.00 120.0 [Z-matrix]
5 -1 0 4 3 2 1 12 0.0000 1.390 0.00 120.0 [Z-matrix]
6 -1 0 5 4 3 1 12 0.0000 1.390 0.00 120.0 [Z-matrix]
7 0 0 2 1 3 1 41 0.0000 1.300 0.0 0.0 [bisector]
8 0 0 3 2 4 1 19 0.0000 1.400 0.0 0.0 [bisector]
9 -1 0 8 3 2 1 30 0.0000 1.220 180.00 120.000 [Z-matrix]
10 -1 0 8 3 2 1 30 0.0000 1.220 0.00 120.000 [Z-matrix]
11 0 0 4 3 5 1 11 0.0000 1.300 0.00 0.000 [bisector]
12 0 0 11 4 0 1 11 0.0000 1.200 0.00 0.000 [extension]
13 0 0 12 11 0 1 1 0.0000 1.080 0.00 0.000 [extension]
14 0 0 5 4 6 1 2 0.0000 1.080 0.00 0.000 [bisector]
15 0 0 6 5 1 1 2 0.0000 1.080 0.00 0.000 [bisector]
16 0 0 1 2 6 1 13 0.0000 1.450 0.00 0.000 [bisector]
19 -1 0 16 1 2 1 13 0.0000 1.500 180.00 109.470 [Z-matrix]
17 18 0 16 1 19 1 3 0.0000 1.080 0.00 109.470 [methylene]
26 -1 0 19 16 1 1 13 0.0000 1.500 180.00 109.470 [Z-matrix]
21 -2 0 19 16 26 1 13 0.0000 1.500 0.00 0.000 [pyramidal]
20 0 19 16 21 26 1 43 0.000 1.700 0.00 0.0 [methine]
24 -1 0 21 19 16 1 29 0.0000 1.400 80.00 109.470 [Z-matrix]
22 23 0 21 19 24 1 3 0.0000 1.080 0.00 109.470 [methylene]
25 -1 0 24 21 19 1 5 0.0 0.800 180.00 106.000 [Z-matrix]
29 -1 0 26 19 16 1 13 0.0 1.500 180.00 109.470 [Z-matrix]
27 28 0 26 19 29 1 3 0.0 1.080 0.0 0.0 [methylene]
30 -1 0 29 26 19 1 3 0.0 1.080 60.00 109.47 [Z-matrix]
31 -1 0 29 26 19 1 3 0.0 1.080 180.00 109.47 [Z-matrix]
32 -1 0 29 26 19 1 3 0.0 1.080 -60.00 109.47 [Z-matrix]
25 0 0 24 0 0 1 5 0.0000 1.000 0.00 0.000 [reset]
1
1.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 1.0
0.0000 0.0000 0.0000
0
0
```

Finally, you have to execute **Retcor**.

```
./run.retcor large
```

It produces the *large.oe* and *large.dat* files. These files have no cell parameters as this is an isolated molecule. Atom coordinates are in Å. You may plot this *.dat* file with Schakal by E. Keller, or convert it into a *.xyz* file by hand to plot the molecule with any other graphic software of your choice. The result is the molecule shown below and corresponds to the following explicit coordinate list.

File large.dat (Schakal-readable format)

```
TITL  LARGER99  large mole
CELL  1.0000    1.0000    1.0000    90.0000    90.0000    90.0000
ATOM  C101     0.0000    1.3900    0.0000
ATOM  C102     1.2038    0.6950    0.0000
ATOM  C103     1.2038   -0.6950    0.0000
ATOM  C104     0.0000   -1.3900   -0.0000
ATOM  C105    -1.2038   -0.6950    0.0000
ATOM  C106    -1.2038    0.6950    0.0000
ATOM  F107     2.3296    1.3450    0.0000
ATOM  N108     2.4162   -1.3950   -0.0000
ATOM  O109     2.4162   -2.6150   -0.0000
ATOM  O110     3.4728   -0.7850    0.0000
ATOM  C111     0.0000   -2.6900   -0.0000
ATOM  C112     0.0000   -3.8900   -0.0000
ATOM  H113     0.0000   -4.9700   -0.0000
ATOM  H114    -2.1391   -1.2350    0.0000
ATOM  H115    -2.1391    1.2350    0.0000
ATOM  C116     -0.0000    2.8400    0.0000
ATOM  H117     0.5091    3.2000   -0.8818
ATOM  H118     0.5091    3.2000    0.8818
ATOM  C119    -1.4142    3.3400    0.0000
ATOM  Br120    -2.2156    2.7733   -1.3880
ATOM  C121    -2.1213    2.8400    1.2248
ATOM  H122    -1.8299    1.8179    1.4167
ATOM  H123    -3.1889    2.8866    1.0684
ATOM  O124    -1.7750    3.6378    2.3219
ATOM  H125    -2.2714    3.2614    3.1041
ATOM  C126    -1.4142    4.8400   -0.0000
ATOM  H127    -0.8959    5.2065   -0.8737
ATOM  H128    -0.8959    5.2065    0.8737
ATOM  C129    -2.8285    5.3399    0.0000
ATOM  H130    -3.3375    4.9799    0.8818
ATOM  H131    -2.8285    6.4199    0.0000
ATOM  H132    -3.3375    4.9799   -0.8818
SYMM  X, Y, Z
END
```

