MiCMoS
Milano Chemistry Molecular Simulation
Description and User Manual

Tutorials

T1. Generation of structural data files

PURPOSE: In this tutorial, it is described how a *.cif* crystallographic information file is interpreted by the **Retcif**, **Retcor** and **Retcha** sequence to produce a *.oeh* structural file, ready for subsequent calculations (static lattice, MC or MD).

REFERENCE to main manual: Section 1.1, Appendix Section A1 (**Retcif**); Section 1.2, Appendix Section A2 (**Retcor**); Section 1.3 (**Retcha**).

FILES: On https://sites.unimi.it/xtal_chem_group/images/MiCMoS_package/T1/sample.cif you can download the materials for this exercise. You also need *runretcif*, *runretcha* and *runretcor* macros available in the batch directory of the package. File sample.*cif* is an example of a minimal *.cif* file readable by the **Retcif** module (this is a subset of the CSD entry ACSALA01, the $P2_1/c$ structure of acetylsalicylic acid). Place sample.*cif* and pertinent (Windows/Unix) macros into your working directory.

CAUTION: Any user-prepared cif file should be reduced **exactly** to the format of sample.*cif*. There is some freedom only for atomic coordinates, which for example may or may not have (sigmas).

In the command prompt, type the running command:

```
./run.retcif sample
```

Then answer dialog mode:

```
This is the Retcif module version 3.0 nov 2015
type 0 to normalize Hs, 1 to keep csd Hs
```

Zero is the normal option for hydrogen atom generation (typing 1 here the program will retrieve hydrogen atom coordinates as they are on the *cif* file). Type 0: having renormalized hydrogen atom positions is mandatory due to the formulation of the force field. Using uncorrected X-ray H-atom positions will possibly lead to unreliable results. Use the option 1 only when you have neutron diffraction structures.

```
type 0 for normal, 1 for extended output
```

Type 0. Typing 1 will display all bond connectivities.

```
type 0 for H connect. checks, 1 retrieve anyway
```

If zero is given (normal option), the **Retcif** module will supply codes for the renormalized location of hydrogen atoms, and issue an error message if the hydrogen count is wrong with respect to what is in the .*cif* file. Typing 1 forces **Retcif** to retrieve what it has anyway, for subsequent corrections if necessary. Type 0 for the moment.

The program will print on screen:

```
structures retrieved      1 written on tape      1
```

This means that **Retcif** has found 1 structure in the .*cif* file, and it has been correctly interpreted and written into a sample .*oih* file. A .*pri* output file is also written, with a summary of the main information retrieved from .*cif*. Have a look at it:

Retcif *output*

```
 refcode is    SAMPLExx
sumformula sum  C9 H8 O4
species H  ispen and n.of atoms    1     8
sumformula moiety C9 H8 O4
SAMPLExx  space group name is  'P 21/c'
summary of atom counts
read from cif: total, non-H, H   21    13     8
n. of Hs in sumformula   8 fragments in as.unit    1
number of chemical units   1
after H assignment: atoms total, non-H,H   21    13     8
number of hydrogen placement cards in oih file    8
retrieved   4 sym.opn, Z on cif file    4
```

In case of problem in reconstructing the connectivity, you can verify that symmetry operations, space group and number of atoms are those expected. See now sample.*oih* file (a print is available in the next page); compare it with Section 1.4 in the manual to interpret the format.

Explicit coordinates are written for atoms 1-13 (the hydrocarbon backbone) and 21 (the acidic proton, whose coordinates **cannot** be guessed by the program). Note that atomic charges (last column) are all 0.0 and the appropriate atomic specie label (Manual, Table 1.1) has been assigned in the penultimate column. Then, a list of NHYD = 8 (Section 1.4.3) implicit atoms follows, whose coordinates are to be computed by **Retcor**. The first seven atoms are aromatic and aliphatic H; even though their coordinates were present in the .*cif*, **Retcif** has ignored them and has prepared instead a proper list of instructions as strings of n_1-n_6 identification codes. These will be used by **Retcor** to compute their correct explicit coordinates (Figure 1.2 in the Manual and Table A2.1 in the Appendix). The 8th atom is the -OH hydrogen: even though its coordinates are explicitly present in the atom list, we want that **Retcor** renormalize its bond distance to 1.0 Å. The last instructions concern symmetry operations and a flag IL=3 to signal that, if a PIXEL calculation is wanted, coordinates must be translated into the inertial reference frame (see also Section 3.2.5 in the main text).

Example of an .oih file, with some implicit atoms and atomic charges set to 0

```
#SAMPLExx 'P 21/c'                                      0.0
   0  0.
  11.4300    6.5910   11.3950    90.000    95.680    90.000
    0
   14
    1     0.1534     0.5640     0.0674    1  12   0.0000
    2     0.2462     0.4877     0.0095    1  12   0.0000
    3     0.2983     0.3047     0.0416    1  12   0.0000
    4     0.2604     0.1961     0.1338    1  12   0.0000
    5     0.1696     0.2676     0.1931    1  12   0.0000
    6     0.1165     0.4496     0.1598    1  12   0.0000
    7     0.0903     0.7573     0.0376    1  10   0.0000
    8     0.3659     0.7384    -0.0637    1  10   0.0000
    9     0.3974     0.8378    -0.1728    1  13   0.0000
   10     0.1203     0.8603    -0.0508    1  28   0.0000
   11     0.0104     0.8118     0.0965    1  27   0.0000
   12     0.2857     0.5878    -0.0880    1  23   0.0000
   13     0.4036     0.7801     0.0345    1  27   0.0000
   21     0.0737     0.9863    -0.0626    1   6   0.0000
    8
   14    0    0    3    2    4    1    2   0.0000   1.080      0.00      0.000
   15    0    0    4    3    5    1    2   0.0000   1.080      0.00      0.000
   16    0    0    5    4    6    1    2   0.0000   1.080      0.00      0.000
   17    0    0    6    1    5    1    2   0.0000   1.080      0.00      0.000
   18   -1    0    9    8   12    1    3   0.0000   1.080     57.57    109.471
   19   -1    0    9    8   12    1    3   0.0000   1.080    177.57    109.471
   20   -1    0    9    8   12    1    3   0.0000   1.080    297.57    109.471
   21    0    0   10    0    0    1    6   0.0000   1.000      0.00      0.000
    4     4
   1.0    0.0    0.0    0.0    1.0    0.0    0.0    0.0    1.0
   0.000000   0.000000   0.000000
  -1.0    0.0    0.0    0.0    1.0    0.0    0.0    0.0   -1.0
   0.000000   0.500000   0.500000
  -1.0    0.0    0.0    0.0   -1.0    0.0    0.0    0.0   -1.0
   0.000000   0.000000   0.000000
   1.0    0.0    0.0    0.0   -1.0    0.0    0.0    0.0    1.0
   0.000000  -0.500000  -0.500000
    3
    0
```

The last input block contains the 4 symmetry operations of $P2_1/c$, expressed as pairs of proper or improper rotation matrices S (9 line elements, corresponding to the sequence $S(1,1)$, $S(1,2)$, $S(1,3)$, $S(2,1)$, $S(2,2)$, $S(2,3)$, $S(3,1)$, $S(3,2)$, $S(3,3)$) and translation vectors. For example, the last first operation is the identity (+x, +y, +z) and the last one corresponds to +x, -1/2-y, -1/2+z.

Retcor will read the *.oih* file and produce *.oeh* and *dat* files. It is time to call it. Type in the command prompt:

```
                        ./run.retcor sample
```

The program should produce the following output on screen:

```
This is the Retcor module version 3.0 nov 2015
end data stream      1 read      1 written      0 errors
```

which means that the calculation ended regularly. **Retcor** will issue error messages when encountering singularities in its atom construction routines (an unlikely occurrence). For crystal

data, **Retcor** will also screen out the redundant space-group symmetry operations when the input molecule has point-group symmetry (an indispensable feature for running the subsequent energy calculation modules). A richer amount of information is printed in the samplecor.*pri* file. Note that the program recognizes, for each implicit entry in the input *.oih* file, what is the reference connectivity to be used to define new coordinates and prints the corresponding flag ("trigonal atom", "Z-matrix", "reset distance"; see Table A2.1 in the Appendix of the manual). You may want to verify here that the atom id flags that define the connectivity of your group are correct.

Output of **Retcor**

```
#SAMPLExx 'P 21/c'
trigonal atom    14 codes    3    2    4
trigonal atom    15 codes    4    3    5
trigonal atom    16 codes    5    4    6
trigonal atom    17 codes    6    1    5
Z-matrix tors-dist-angle    18       9    8    12
Z-matrix tors-dist-angle    19       9    8    12
Z-matrix tors-dist-angle    20       9    8    12
resetting distance between atoms    21    10
number of eq.pos,indep.,point group    4    4    0
end data stream      1 read      1 written      0 errors
```

Some new files have now appeared in the working directory. File sample.*dat* contains your molecular structure in crystallographic coordinates, ready to be plotted with SchaKal by E. Keller. File sampleort.*oeh* is the isolated molecule, after having get rid of unit cell and space group. Atoms are now explicitly listed, all expressed in Cartesian coordinates. This is the starting point, for example, to build a simulation box for a liquid. Finally, file sample.*oeh* contains the molecule in crystallographic coordinates, with all the atoms explicitly listed, with information on unit cell and symmetry operations. This file must be used to simulate crystals.

Both sampleort.oeh and sample.oeh do not have charges yet. To assign them, you should run **Retcha**, which will produce CLP-usable point-charge parameters. Command is:

```
./run.retcha sample
```

If the program issues the following message, the Extended Hückel calculation (see Section 1.3 in the Manual) ended correctly:

```
structures read and written    1    1
```

In *Retcha*, an error message is issued if the electron count reveals that the molecule is not a closed shell or has degenerate frontier orbitals, or has atoms for which no Extended Hückel parameters are available. File samplenoq.*oeh* is the original *.oeh* file saved without charges.

In the end, file sample.*oeh* will be ready to enter static lattice energy calculations, or MC/MD simulations. If you want to add charges to the isolated molecule, use the file sampleort.*oeh* produced by **Retcor** as an input to **Retcha**:

```
./run.retcha sampleort
```

In any case, a sample(ort)cha.*pri* is printed, which also contains the retrieved charges.

Example of an .oeh file with crystallographic coordinates and Extended Hückel charges.

```
#SAMPLExx 'P 21/c'                                0.000      0.0       0.    0.00
  0          0.0
  11.4300    6.5910   11.3950    90.000    95.680    90.000
    0.0
  21
   1    0.15340    0.56400    0.06740   1  12   -0.0532
   2    0.24620    0.48770    0.00950   1  12    0.5755
   3    0.29830    0.30470    0.04160   1  12   -0.3439
   4    0.26040    0.19610    0.13380   1  12   -0.2316
   5    0.16960    0.26760    0.19310   1  12   -0.3232
   6    0.11650    0.44960    0.15980   1  12   -0.2388
   7    0.09030    0.75730    0.03760   1  10    1.3326
   8    0.36590    0.73840   -0.06370   1  10    1.3401
   9    0.39740    0.83780   -0.17280   1  13   -0.8859
  10    0.12030    0.86030   -0.05080   1  28   -0.9080
  11    0.01040    0.81180    0.09650   1  27   -1.2101
  12    0.28570    0.58780   -0.08800   1  23   -0.7525
  13    0.40360    0.78010    0.03450   1  27   -1.1311
  14    0.36868    0.24670   -0.00548   1   2    0.2969
  15    0.30217    0.05387    0.15979   1   2    0.2843
  16    0.14005    0.18175    0.26531   1   2    0.2872
  17    0.04492    0.50446    0.20628   1   2    0.2857
  18    0.31897    0.89723   -0.22149   1   3    0.3477
  19    0.45815    0.96037   -0.15004   1   3    0.3089
  20    0.43805    0.72830   -0.22643   1   3    0.3516
  21    0.07316    0.98776   -0.06274   1   6    0.6678
   0
   4
  1.0    0.0    0.0    0.0    1.0    0.0    0.0    0.0    1.0
 0.000000  0.000000   0.000000
 -1.0    0.0    0.0    0.0    1.0    0.0    0.0    0.0   -1.0
 0.000000  0.500000   0.500000
 -1.0    0.0    0.0    0.0   -1.0    0.0    0.0    0.0   -1.0
 0.000000  0.000000   0.000000
  1.0    0.0    0.0    0.0   -1.0    0.0    0.0    0.0    1.0
 0.000000 -0.500000  -0.500000
   3
   0
```

CAUTION: The macro *run.cry* runs in sequence the ***Retcif***, ***Retcor***, ***Retcha*** and ***Crysaa*** modules deleting all unnecessary print files. This can be used when there are no special features and goes directly from the *.cif* file to the *.oeh* file with CLP potential atomic charges. The atom-atom lattice energy is also immediately provided (see Tutorial T4).

As file *.oih* above, the structure file is closed by the full list of symmetry operations and translation vectors.