Program: GRANADA

Version 4.5

Date: *Oct. 29, 2020*

RANDOM DISTRIBUTION OF MOLECULES AROUND A "SOLUTE KIND" POLYATOMIC SYSTEM

Luis A. Montero Cabrera

Universidad de La Habana, Facultad de Química, La Habana 10600, Cuba

Johns Hopkins University, Dep. of Chemistry, Baltimore, MD 21218, USA

granada4.5 is an executable program that is designed to read Cartesian coordinate files in XYZ format that contain atomic data of molecules in order to generate cells where environmental molecules are randomly surrounding it. Environmental molecules are randomly placed in cages as solvents, or over, or between surfaces. The placement of surrounding molecules avoids mutual overlapping by rejecting those configurations where any pair of atoms becomes nearer to the sum of their respective van der Waals radii scaled by a default factor of 0.75. More or less packed configurations can be selected as entering such factor by input options.

The produced files can be the same kind of Cartesian coordinates in the XYZ for further processing. It can also produce Cartesian coordinate input files for "Gaussian" system of programs, internal coordinate MOPAC files and "Orca" program system like input files. This program *does not* use any potential that can influence random distribution of molecules other than disallowing overlap farther than the van der Waals radii of component atoms.

Originally written by Luis A. Montero, Faculty of Chemistry, University of Havana 10400, Cuba at the Technical University of Dresden, in 1995, and completed at the University of Granada, Molecular Design and Modeling Group, Granada 18071, Spain, in June, 1996. Latest granada4 versions were produced at the Dep. of Chemistry, Johns Hopkins University, Baltimore, MD 21210, USA during March – October, 2020.

THE INPUT:

The main input is given by the command line according the following format:

`<program name> solute.xyz [solvent1.xyz [solvent2.xyz]…]`

Where `<program name>` could be either `granada4-3.exe` in the case of the Windows version or `granada4.4.x` in the case of the Linux version. `solute.xyz` is the name of the file with the Cartesian coordinates for the geometry of the central molecule or cluster. `solvent`*`i`*`.xyz` are the corresponding for eventual surrounding or solvent molecules. There could be up to 5 different solvent molecules. Default solvent is water, which coordinates are internally stored by the program. Their XYZ files can be avoided in the command line if being the only solvent.

A file named **`input.mmh`** must be present in the predefined directory and contains the general parameters for the calculation. Numerical input formats are loose and can be entered as numerical

data separated by commas.

First line:

```
nlim, dimen, nconf, nfsolv, (fracm(i),i=1,nfsolv)
```

nlim is the number of desired environmental or solvent molecules. In all cases, the origin of the Cartesian coordinates of the solute is reallocated to the center (middle) of its absolute coordinates.

dimen is the half-dimension (in Ångstroms) on all the Cartesian axes of a virtual cube wherein the environmental molecules will be randomly placed when a solution liquid like solution environment is being simulated. The volume of the cube will be `(2*dimen)**3` and no coordinate of solute molecules will have a greater value than `dimen`. Options SOLID1 and SOLID2 give a different meaning to this parameter (see below).

nconf is the number of random geometries to be generated. It means that the output `xxxx_solvated.xyz` file will hold `nconf` sequences of cell geometries with different random distributions of environmental molecules. The **default value** of `nconf` is 1. Cartesian coordinate files must follow the standard XYZ format for each cell, with an initial line containing the number of atoms, followed by a line of titles, and then a line for each atom with the symbol or atomic number followed by the coordinates X, Y and Z in angstroms, separated by spaces.

nfsolv is the number of environmental molecules other than water to be taken into account (the maximum is 5, and the implicit 0 for water).

fracm(i) is the molar fraction of the *i* environmental molecule. You can write up to `nfsolv` molar fraction values. They only refer to environmental molecules, NOT TO THE SOLUTE OR CENTRAL MOLECULE.

Second line:

It requests a single line with keywords and parameters for each run, separated by commas or blank spaces.

INCLUDE: Environmental molecules will be allowed to occupy the entire random space within the limits of the dimensions of the predetermined box, including cavities within the solute molecule. Superposition of molecules in clusters will be allowed up to 0.5 the van der Waals radii distances of component elements.

EXCLUDE: If positive, the solvent molecules will externally surround an ellipsoid with dimensions proportional to the central cluster molecule that is formed in each step. This is the **default option.** Superposition of molecules in clusters will be allowed up to 0.75 the van der Waals radii distances of component elements.

SOLID1(*dd*): Environmental molecules are placed where all solvent coordinates remain below a maximum distance of `dimen` Ångstroms perpendicular to the positive side of the XY plane of the

solute molecule. *dd* means the minimum distance in Ångstroms where all *x* and *y* coordinates of the solute will be placed with respect to the maximum and minimal coordinate values of the solute limits. This option could also be entered for fixed values by as **SOLID1_dd**, where *dd* could be 0.5, 1.0 or 1.5.

SOLID2(*dd*): Environmental molecules are placed between two surfaces of equidistant identical solutes. Solvent coordinates remain at `dimen` Ångstroms perpendicular equidistance's to the the XY plane of mirroring solute molecule. *dd* means the minimum distance in Ångstroms where all *x* and *y* coordinates of the solute will be placed with respect to the maximum and minimal coordinate values of the solute limits. This option can also be entered for fixed values as **SOLID2_dd**, where *dd* could be 0.5, 1.0 or 1.5.

FVDW(*f*) is entered when a different to default factors is desired to scale the sum of van der Waals radii of atoms. *f* is de desired scale factor different to the default value of 0.75.

MOPAC: means that there will be an additional *`solute`*`_solvated.mop` output file with random configurations in the form of internal coordinate entries for MOPAC. A text of **MOPAC** options limited to 80 columns must be included as text lines following this one of the main **granada4** input file. It will be used as the input data for most MOPAC program versions for semi-empirical calculations to corresponding expected outputs.

GAUSSIAN(*n*): means that there will be a *`solute`*`_solvated.gjf` output of Cartesian coordinates in the Gaussian input format. In this case, it is assumed that the molecules are neutral and singlet and any modification must be made by editing the output file so that it serves as a correct Gaussian input format. *n* value means the number of Gaussian command line to be written for each output configuration. For n=1,6 it could also entered as GAUSSIAN*n*. This desired Gaussian command lines must be entered as text lines following this one of the main **granada4** input file.

ORCA(*n*): means that there will be a *`solute`*`_solvated.inp` output of Cartesian coordinates in the Orca input format. In this case, it is assumed that the molecules are neutral and singlet and any modification must be made by editing the output file so that it serves as a correct Orca input format. *n* value means the number of Orca command line to be written for each output configuration. For n=1,6 it could also entered as ORCA*n*. This desired Orca command lines must be entered as text lines following this one of the main **granada4** input file.

MOPAC, **ORCA** and **GAUSSIAN** keywords are exclusive to each other.

Ending lines:

Ending lines are those expected if MOPAC, GAUSSIAN or ORCA input files are to be created. See above.

THE OUTPUT:

The output of Granada consists of several text files whose common name is "SOLVATED", with the corresponding extensions. A *SOLUTE*_SOLVATED.LOG file is also produced with details of the calculation.

AN EXAMPLE:

COMMAND LINE: `granada4-5.exe acrolein.xyz ethanol.xyz acetone.xyz`

`INPUT.MMH` file:

8,6.5,50,2,0.25,0.75
Include MOPAC fvdw(0.3)
pm3 specify geo-ok ef

In this case:

nlim = 8, dimen = 6.5, nconf = 50, nfsolv = 2, fracm (1) = 0.25, fracm (2) = 0.75

The scale factor to avoid atom – atom overlapping is set to 0.3.

It means that 0.25*8 = 2 ethanol molecules together with 0.75*8 = 6 acetone molecules will be surrounding an acrolein molecule. The run will create 50 different cell geometries where solvents are placed randomly around the central solute molecule.

After the run there will be two fixed output files. The one named *solute*_solvated.log contains a record of the generation of cells including cases where a given random geometry is discarded because some superposition occur. It could be originated when a newly generated molecule occupies the van del Waals room of a previously existing one. The other is a *solute*_solvated.xyz file with all the generated cell configurations.

Because the keyword INCLUDE is provided environmental or solvent molecules will be allowed to occupy the entire random space within the limits of the dimensions of the predetermined box, including cavities within the solute molecule.

As the keyword MOPAC is used, another *solute*_solvated.mop file will be created with the randomly generated geometries written by the internal coordinate MOPAC file format. The command line of each input molecule in the *solute*_solvated.mop file will be: `pm3 specify geo-ok ef`

.XYZ FILES

The .XYZ files have the following format:

Line 1:
Free format: N # Number of atoms in the entry.
Line 2:
80 cols .: # Comments
Line 3 to 2 + N:
For each atom I :
Free format: NAT (I) or SYMBOL # atomic number or chemical symbol of the atom
Free format: X (I) # X coordinate in Ångstroms
Free format: Y (I) # Y coordinate in Ångstroms

Free format: Z (I) # Z coordinate in Ångstroms