SOLVATE

USER MANUAL: QUICK GUIDE

SOLVATE version 2024.02

User Manual: Quick Guide

OVERVIEW

This is a manual of *Solvate Suite*. The modules of the main program interfaces with various widely used software in the scientific community (such as *ORCA*, *xTB*, *GROMACS*, *PackMol*, and *Q-Force*), enabling the sequential treatment of all stages of molecular modeling involving explicit interaction of the species of interest with the any solvent using molecular simulation procedures and electronic structure calculations in a hybrid approach (explicit/implicit solvent).

(1) The first problem that the program aims to solve is the creation of the simulation box, considering several user-specified criteria (such as the number of solvent molecules, shape and density of the simulation box, solute/solvent concentration, or proportion between solvent mixtures), as well as the construction of the system's topology and force field parameters (in the case of simulation methods based on classical dynamics).

(2) The second problem is submitting the simulation with a minimal adjusted set of parameters (some of which can be altered by the user during runtime; alternatively, the user can perform this step with their own simulation parameters).

(3) The third is the analysis of simulated properties (pressure, temperature, density, short-range interaction, total potential energy, and radial distribution function), including statistical analysis (such as stochastic average values, standard error, and standard deviation), graphical analysis, and extraction of properties by different criteria (such as block analysis and pressure filter).

(4) The fourth involves selecting a user-determined number of hydrogen-bonded structures or microsolvation clusters and solvent molecules per cluster identified from the simulation trajectory file based on the maximum connectivity criterion (hydrogen bonds) or minimum energy criterion (hydrogen bonds and microsolvation clusters).

(5) The fifth involves the treatment (by semi-empirical Quantum Chemistry methods) of microsolvation clusters for the calculation of free energy, considering three criteria: in the fixed geometry extracted from the simulation, in a completely optimized cluster structure, or in a relaxed solvent cavity and completely optimized solute geometry. In any case, the procedure removes all imaginary frequencies (which have a significant weight in the composition of free energy), and uses the quasi-harmonic correction (to handle low-frequency intermolecular modes), followed by the use of extrapolation methods for free energy to the theory level specified by the user (which also allows extrapolation of the basis set restriction in the implicit solvent methods when considering microsolvation clusters with at least the first solvation layer).

(6) The sixth and final step involves managing the significant number of files generated throughout the modeling process.

Step 1: Packing

➢ **solvate <*solute*> <*solvent*> -nm N [-it -bk]** [N = integer]

➔ **OBJECTIVE: Generation of the packaging box for the simulation.**

✔ ***Files***: The files `<solute>.ext` e `<solvent>.ext` contains the geometries of the solute and solvent, respectively. The *Solvate* program identifies geometries based on the order in which the files are passed into the command line, with the file with the solute structure being the first. It is possible to specify more than one solvent, informing the program of the number or proportion of each species in the packaging. By default, the program generates a simulation box with a single solute molecule in its center.[1-2]

✔ ***Nmol*`[-nm]`**: Specifies the number of solvent molecules to be included in the packaging process. If this parameter is specified, the program will automatically determine the dimensions of the simulation box that satisfy the criterion of minimum distance between molecules in the simulation box.

✔ ***Npck*`[-np]`**: Specifies the number of "packing layers" (which are defined so that there are at least 1 solvent molecules surrounding the solute in any direction). If this parameter is specified, the program will automatically determine the number of solvent molecules (`nmol` parameter) and the dimensions of the simulation box that simultaneously satisfy the criteria of the number of packing layers and minimum distance between the molecules in the simulation box (the *Solvate's* current default is 2.0 Å, there is no need to enter it on the command line; this parameter can be adjusted with the `tolerate` command line option).

✔ ***Shape*`[-sp]`**: Specify simulation box format. The available shapes are `sphere` (spherical, standard for *ORCA*), `box` (orthorhombic box) and `cube` (cubic, standard for *GROMACS*).

✔ ***Density*`[-ds]`**: Specifies the density of the simulation box. The requested density is reached by automatically adjusting the minimum distance criterion (`tolerance` parameter), keeping the other criteria (such as number of packing layers or number of solvent molecules) unchanged.

---

1 If the initial structures have not been previously minimized, use the *runCREST* module of the application suite to perform conformational analysis of the input structures for packaging.

2 Preferably use entries in the format .com (*Gaussian*) or .inp (*ORCA*) due to the load information and the multiplicity of the system. The other accepted formats are .xyz, .pdb, .log (*Gaussian*) and .out (*ORCA*).

Step 2: Simulation

➢ **solvate <*solute*> <*solvent*> -nm N -mdrun <*additional options*>**

➔ **OBJECTIVE: Submitting simulation using the box generated in the previous step.**

✔ ***MDRun***: In the case of boxes in the *inp* format, request a *Born-Openheimer molecular dynamics* (BOMD) simulation with the *ORCA* program at the semi-empirical level GFN2-xTB; If no additional parameters are provided, a first submission will be performed in an NVT ensemble at a temperature of 298 K, with 1000 steps of 1.0 fs.[3] In the case of boxes in the *gro* format, request a classical dynamics simulation with the *GROMACS* program; it is necessary to specify the type of similarity (minimization, balancing or production.

✔ ***Auto***: Start MD simulation with automatic settings.

✔ ***Init***: Execute an energy minimization with simulation box optimization procedure.

✔ ***Equi***: Execute MD equilibration procedure (with auto restart).

✔ ***Prod***: Execute MD production procedure (with auto restart).

✔ ***-prs***: Pressure for the MD simulation (in bar).

✔ ***-bar***: Type of barostat pressure coupling.

```
BRDS = Berendsen | CRES = C-rescale | PARR = Parrinello | MTTK
```

✔ ***-tmp***: Temperature for MD simulation (in K).

✔ ***-ter***: Type of thermostat temperature coupling.

```
BRDS = Berendsen | VRES = V-rescale | NOSE = Nose-Hoover | ADSN = Andersen
```

✔ ***-dts***: Simulation step size (in fs).

✔ ***-tot***: Total simulation time (in ps).

✔ ***-nrs***: Total number of simulation steps.

Notes: In the same folder with multiple files, it is possible to carry out different packaging with a variable number of solvent molecules, so in this step it is necessary to inform the number of solvent molecules with the nmol parameter, which allows the program to correctly identify the files generated. Although it is possible to perform the calculation directly with the simulation program, submission via Solvate guarantees that the files will have the appropriate format to be read in the remaining steps. Upon completion of the simulation, the trajectory and data files are obtained, whose data can be extracted with the DATAS module for simulation analysis.

3 In the NVT ensemble, the volume is maintained by imposing a force defined by a spring constant on the simulation box which, by default, is 10.0 $kJ \cdot mol^{-1} Å^{-2}$. This criterion can be adjusted in the input file by modifying the *cell/spring* parameter. For more information and adjustment options, consult the *ORCA* manual.