MCPB.py

(v. 1.0 Beta2, Pre-release date: June 26 2015)

```
Usage: -i input_file -s/--step step_number
       [--logf Gaussian/GAMESS-US output logfile]
       [--fchk Gaussian fchk file]
Options:
  -h, --help              Show this help message and exit
  -i INPUTFILE            Input file name
  -s STEP, --step=STEP    Step number
  --logf                  Gaussian/GAMESS-US output logfile
  --fchk                  Gaussian fchk file
```

The following is an introduction of the variables in the input_file:

(*Reminder:* there should be no blank lines in the input_file. The values or parameters should follow the variables separated by a space.)

Required variables:

original_pdb This is the file name of the original PDB file, which should have only one chain. The PDB file should have hydrogen atoms and metal ions in it. Users are advised to use an application like pdb4amber to clean up the PDB file first. They are also advised to add the hydrogen atoms by using a webserver such as H++ before performing the modeling in MCPB.py.

ion_ids The PDB atom ID of the complex's central metal ion. It should be an integer value, depending on how many metal ions are included in the metal complex.

ion_mol2files The name(s) of the ion(s) in the mol2 file(s) contained in the metal center. This can be one or several name(s), depending on how many kinds of ions are included in the metal center. The user can use antechamber to transfer the single ion PDB file to a mol2 file and then manually modify the atom type and the atomic charge of the metal ion in the mol2 file.

ion_info This variable is only required for the nonbonded model without re-fitting the residue charges (step number 4n2). In all, there are four data points required for each metal ion: 1) the residue name of the metal ion in the PDB file; 2) the atom name of the metal ion in the PDB file; 3) the element symbol of the metal ion; 4) the charge (or oxidation state, which needs to be an integer) of the metal ion. For example: ZN ZN Zn 2 (the first two are the residue and atom name of the Zn2+ ion in the PDB file, the third is its element symbol and the last one is its charge).

Optional variables:

group_name The group name the user has specified. The group name is the prefix for different kinds of modeling files e.g. PDB, fingerprint and Gaussian input files for different models. [The default is MOL.]

cut_off The cutoff value is used to indicate there is a bond between the metal ion and the surrounding atoms. The unit is Angstroms. [The default is 2.8.]

chgfix_resids Specify the residues' IDs whose charges are going to be fixed during the charge fitting. The fixed charge values are referenced from the mol2 files used during the modeling. [The default value of this variable is the null list.]

scmodel_chg Specify the total charge of the sidechain model. [The default value of the charge will be assigned automatically by the program, which is not guaranteed to be right. Careful check is suggested from runnning the Gaussian/GAMESS-US program. If it is not right, you can add this option with right charge into the input file and regenerate the modeling files.]

lgmodel_chg Specify the total charge of the large model. [The default value of the charge will be assigned automatically by the program, which is not guaranteed to be right. Careful check is suggested from runnning the Gaussian/GAMESS-US program. If it is not right, you can add this option with right charge into the input file and regenerate the modeling files.]

naa_mol2files The variable used to indicate non-amino acid mol2 file(s) in the metal complex if there are any nonstandard residue(s) in the metal complex. Examples of nonstandard residues include hydroxyl group and ligand molecules. For these residues, the user can use antechamber to generate the mol2 file(s) by first doing an AM1-BCC or HF/6-31G* RESP charge fit and then assigning an AMBER atom type (recommended for water or hydroxyl group) or a GAFF atom type (recommended for ligand). [The default value of this variable is the null list.]

force_field The user-designated name of the force field. The current version supports ff94, ff99, ff99SB, ff03, ff03.r1, ff10, ff12SB and ff14SB. [The default is ff14SB.]

gaff A variable used to indicate the use of a GAFF force field during the modeling. 0 means no, 1 means yes. [The default is 1.]

frcmod_files The variable used to indicate the parameter modification file(s) for the nonstandard residue(s) (e.g. frcmod file generated by parmchk for a ligand molecule) in the metal complex. It can be one name or several names seperated by space. [The default value of this variable is the null list.]

large_opt A variable used to indicate whether to do an geometry optimization in the Gaussian input file. Three options are aviable: 0, 1, or 2. 0 means no optimization, 1 means only optimizing the hydrogen positions, 2 means full geometry optimization. [The default is 0.]

sqm_opt A variable used to indicate the use of SQM in AmberTools to do a simulation of the sidechain and/or large model before using Gaussian to perform the calculation. *Please note:* if 1, 2 or 3 are chosen, the first step of the modeling process will take additional time (minutes for the sidechain model and hours for the large model). [The default is 0.]

- 0 – means no use of SQM.
- 1 – means the optimization is done only for the sidechain model.
- 2 – means the optimization is done only for the large model.
- 3 – means the optimization is done for both the sidechain and large models.

water_model The user-designated water model to be used in the molecular modeling. Options are TIP3P, SPCE and TIP4PEW. [The default is TIP3P.]

ion_paraset The user-designated ion parameter set to be used in the nonbonded model. (This option has no influence on the metal ion VDW parameters in the bonded model, in which the author has chosen certain VDW parameter sets for different ions.) There are four options for this variable: HFE, CM, IOD and 12_6_4 (reminder: there are underlines between the numbers). If you use the 12-6 Lennard-Jones nonbonded model, the recommended settings are the HFE set for the +1 and -1 ions, the CM set for the +2 ions, and the IOD set for the +3 and +4 ions. They are also the default settings for these metal ions.

software_version The version of software the user used to perform the QM calculations. Three options are available, g03 (which represents Gaussian03), g09 (which represents Gaussian09) and gms (which represents GAMESS-US). [The default is g03.]

The following is an explanation of the step number variables:

Here are the options for the step_number:

For step1 there are three options: 1a (default, same as specifying 1), 1m and 1n.

For step2 there are three options: 2e, 2s (default, same as specifying 2) and 2z.

For step3 there are four options: 3a, 3b (default, same as specifying 3), 3c and 3d.

For step4 there are three options: 4b (default, same as specifying 4), 4n1 and 4n2.

The following is the detailed explanation of the steps used in the modeling procedure:

Step1. Used to generate the modeling files (e.g. PDB, fingerprint and Gaussian input files) for different models (e.g. sidechain, standard and large models). Three options are available and their explanation is shown below. Default is 1a.

- 1a – Used to automatically rename the atom types of the center metal ions and the surrounding bonded atoms in the standard fingerprint file.
- 1m – Used to automatically rename only the atom type(s) of the center metal ion(s) to the AMBER atomic ion atom type style in the standard fingerprint file.
- 1n – Used to generate the standard fingerprint file without renaming the atom types. Users can rename the atom type of the metal ion(s) and its ligating atoms manually in the standard fingerprint file.

Please note: Between using Step1 and Step2, the Gaussian calculations (if needed), should be done for the sidechain model (to calculate the force constants) and the large model (to do the RESP charge calculation) using Gaussian input files. Prior to the calculation, users can change the parameters (such as the calculation method, basis set, etc.) in the Gaussian input files according to their own preferences. After finishing this procedure, the user can move on to Step2.

Step2. Used to generate the frcmod file for the modeling. In this step, a frcmod file (with pre.frcmod name at the end of the file name), will be pre-generated. This file includes all the parameters, except the bond and angle parameters related to the metal ions. Later, the final frcmod file will be generated which will include all the parameters. There are three methods to choose from: Empirical, Seminario and Z-matrix. Each of these methods generates the metal ion-related bond and angle parameters. Default is the 2s (Seminario method).

- 2e – The Empirical method,[1]can generate the metal ion-related bond and angle parameters efficiently without doing Gaussian calculations. It only supports Zn2+ ion in the current version.
- 2s – The Seminario method[3] generates the force field parameters based on sub-matrices of the Cartesian Hessian matrix obtained from quantum calculations. This method requires a Gaussian fchk file (which can be generated from a chk file by using the formchk command in Gaussian).*Reminder:* both the geometry optimization and force constant calculation procedures are needed to generate the final chk file and subsequent fchk file for the force constant calculations done by the Seminario method.
- 2z – The Z-matrix method generates the force field parameters by using the Cartesian Hessian matrix obtained from the quantum calculations. This method requires the force constant Gaussian output file (usually named as a log file) after the geometry optimization and force constant calculations.

Step3. Used to perform the RESP charge fitting and to generate the mol2 files for the residues within the metal ion complex. There are several fitting schemes available in this step. The four options are shown below. The default is 3b since Seminario/ChgModB was identified as the best combination in the work of Peters et al.[2]*Reminder:* the chgfix_resids variable is effective in this procedure, if the variable is specified, the charge restriction will be used as well as one of the following choices.

- 3a – Allows all the charges of the atoms in the ligating residues to change without any restrictions.
- 3b – Restrains the charges of the heavy backbone atoms in the ligating residues according to the user-chosen force field.
- 3c – Restrains the charges of the backbone atoms (both heavy and hydrogen atoms) in the ligating residues according to the user-chosenforce field.
- 3d – Restrains the charges of the backbone atoms (both heavy and hydrogen atoms) and C beta atoms in the ligating residues according to the user-chosen force field.

Step4. Generates the leap input file. The default is 4b.

- 4b – Generates the leap input file for the bonded model.
- 4n1 – Generates the leap input file for the nonbonded model and refits the charge of the ligating residues.
- 4n2 – Generates the leap input file for the nonbonded model without refitting the charge of the ligating residues.

Here are some suggestions for the parameterization procedure:

1) For the modeling of the bonded model, the following steps are usually needed (4 steps): 1a/1n→2e/2s/2z→3a/3b/3c/3d→4b

2) For the modeling of a non-bonded model with a refitted charge, users can follow the workflow (3 steps): 1m→3a/3b/3c/3d→4n1

3) For modeling with a normal nonbonded model (without fitting any charges), users usually only need one step to perform the modeling (1 step): 4n2.

The following is an explanation of the logf and fchk variables:

These variables are optional. If provided, they are only active in step2 and/or step3. The default log file name is group_name + '_sidechain_fc.log' for step2 and group_name + '_large_mk.log' for step3. The default fchk file name is group_name + '_sidechain_opt.fchk' and it is only active for step2 when using Gaussian software and Seminario method to obtain the force constant parameters.

If you are using Gaussian software and Seminario method to generate force constant parameters, it uses the fchk file of sidechain model to store the Cartesian Hessian matrix. If you are using Gaussian software and Z-matrix method

to generate force constant parameters, it uses the log file of sidechain model to store the force constant parameters. While if you are using GAMESS-US software and Seminario method to generate the force constants, it uses the log file of sidechain model to store the Cartesian Hessian matrix. In current version the software doesn't support GAMESS-US with Z-matrix method to generate the force constants.

Both the Gaussian and GAMESS-US software use the log file of large model to store the ESP charges.

References

[1] P. Li and K.M. Merz, Jr. Development and Validation of an Empirical Method to Generate the Bond and Angle Parameters for Metal Ion Containing Complexes in Protein System . *Manuscript in preparation*, 2015.

[2] M.B. Peters, Y. Yang, B. Wang, L. Fusti-Molnar, M.N. Weaver, and K.M. Merz, Jr. Structural Survey of Zinc-Containing Proteins and Development of the Zinc AMBER Force Field (ZAFF). *J. Chem. Theor. Comput.*, 6:2935–2947, 2010.

[3] J.M. Seminario. Calculation of Intramolecular Force Fields from Second-Derivative Tensors. *Int. J. Quantum Chem.*, 30:1271–1277, 1996.