POLCHAT
A polarisation-consistent charge-fitting tool

Stefano Caprasecca[1]

Version

This documentation refers to version 4.1.2.

Please note that the use of versions <4.0.0 is deprecated.

Presentation

PolChat is a Molecolab tool[2]. It performs a polarisable ESP fitting of a given molecule, according to one of the polarisable MM models available, based on the induced dipole formulation.

The next sections describe how to use the tool and the theory behind it.

Usage

Generation of input files

Four files are needed:

MOL.gesp a gesp file used to get the QM potential to be fit (generated e.g. by Gaussian);

MOL.mol2 a mol2 file used to get connectivity and QM dipole information (generated by antechamber);

MOL.pol a file with the values of atomic isotropic polarisabilities, in a.u.[3];

MOL.cns a file with constraints;

The paragraphs below illustrate how to obtain each of these files. Please follow the instructions in the correct order.

---

[1]stefano.caprasecca@for.unipi.it
[2]molecolab.dcci.unipi.it/tools — For help or to require a copy of the program, please write to benedetta.mennucci@unipi.it.

Generating the `gesp` **file**   This file contains the QM electrostatic potential on a grid of points around the molecule. It is produced by Gaussian from the following input file `MOL.com`:

```
#p b3lyp/6-311+G(d,p) pop=ESP IOp(6/50=1)

Calculate ES potential and print it to output file MOL.gesp

0 1
 O  0.0000 0.0000 0.0000
 H  1.0000 0.0000 0.0000
 H  0.0000 1.0000 0.0000

MOL.gesp
```

Additional options one may wish to use are the following:

`IOp(6/41=`n`)` use n layers in the ESP fit (default is 4)

`IOp(6/42=`n`)` density of points per unit area in the ESP fit (default is 1)

File `MOL.gesp` is then produced. It contains information on the atom positions, the QM dipole and the ES potential on the grid points.

Generating the `mol2` **file**   This file is generated by calling antechamber, which reads the gesp file produced in the previous step.
Call antechamber:

```
$> antechamber -i MOL.gesp -fi gesp -o MOL.mol2 -fo mol2

   -c resp -pf yes
```

The `mol2` file generated contains, among other things, the connectivity information. The structure of the connectivity is the following:

```
@<TRIPOS>MOLECULE
MOL
   60     59     1     0     0
[...]

@<TRIPOS>BOND
     1    1    2 1
     2    1   21 1
```

2

```
           3     1     22 1
           4     1     60 1
           5     2      3 2


     [...]
```

The first piece on information, at the beginning of the file, and starting with `@<TRIPOS>MOLECULE`, lists the number of atoms (60) and the number of bonds (59). Other bits of information follow.

In the second piece of information, starting with `@<TRIPOS>BOND`, lists all (59) bonds; in the example, the 1st bond is between atoms 1 and 2 and is single; the 2nd bond is between atoms 1 and 21 and is single; and so on.

In some cases you may wish to change the connectivity information. To do so, modify the list of bonds as you wish, and also update the number of bonds in the first part, as appropriate.

No further changes are needed.

Polarisability file The file `MOL.pol` must be edited by listing all the values of atomic polarisabilities (in a.u.[3]), in the same order the atoms appear in the `MOL.com` input file. The values can be taken from reference papers.

At the moment, one model is implemented: the AL model of Wang et al., *J. Phys. Chem. B* **115**, 3091 (2011). Work on other models is ongoing. If you need to use one specific model, try contacting the authors.

The said model can be used in two variants, as detailed in the Polarisation Model section.

Constraint file The `MOL.cns` file contains the constraints you wish to apply to your fit. Note that the program performs both a standard, non-polarisable ESP fit, and a polarisable one. Both fits are constrained as requested. For this reason, the resulting ESP charges may differ from the Gaussian-fitted ones (here, GESP), since the latter are only constrained to the total molecular charge.

The `MOL.cns` file is structured in the following way.[3]

total charge Required — Use only once
> The total charge constraint is *not optional* and must be entered only once. Use the keywords `chg` or `charge`, followed by the total molecular charge, as in the example:

```
    chg 1.00
```

charge of a fragment Optional — Use any number of times
> It is possible to specify that a certain fragment of the molecule must sum to a certain charge. Use the keywords `frg` or `fragment`, followed by the

---

[3]Please note that from version 4.0.0 it is no more possible to impose a constraint on the dipole moment.

fragment charge and the list of the atoms that constitute the fragment, as in the example below.
The keyword can be used any number of times.

```
frg  0.00 5-6,11,18-20
frg -0.50 7-10,25,27-30
```

This requires that the first fragment, made up by the atoms 5, 6, 11, 18, 19 and 20, must sum up to no charge, while the second fragment, made up by atoms 7, 8, 9, 10, 25, 27, 28, 29 and 30 must sum up to charge -0.5.

equivalence Optional — Use any number of times
The equivalence between atoms is specified using the keywords eqv or `equivalence`, followed by the list of the equivalent atoms, as in the example below.
The keyword can be used any number of times.

```
eqv 10-12
eqv 1,3
```

This requires that the three atoms 10, 11 and 12 must have the same charge, as well as atoms 1 and 3.

restraint Optional — Use only once
The presence of a restraint is specified using the keywords res or `restraint`, followed by the restraint parameter α, the type or restraint, and the list of atoms to be restrained, as in the example below.
The keyword can be used at most once.

```
res 0.005 1 41-45,56-60
```

This requires that a type-1 quadratic restraint of strength 0.005 is applied to the atoms from 41 to 45 and from 56 to 60. At the moment, only type 1 is allowed. It follows Model 2 of Bayly *et al.*, *J. Phys. Chem,* **97**, 10269 (1993). A typical value for the parameter is 0.005. Larger values might provide non-physical results. The use of more than one restraint command, although possible in principle, is not allowed as it could lead to non-physical results if used improperly.

There is no limit on how many constraints are imposed, as long as the conditions are linearly independent. The order the conditions are written in the file is irrelevant. See an example of a constraint file below:

```
     chg 0.00
     eqv 1-3
     frg 1.0 5,7
     eqv 15-16
     eqv 18,20
     res 0.005 1 1,3,5,9,11-15,19
```

Running the program

Compilation   The tool may either be distributed already compiled for a 64-bit machine, or in its source code. The availability of the source code depends on the Molecolab policy which may change. Please contact the authors should you need help.

    If you need to compile, after untarring the code, just run the command:

```
     $> make
```

Running   Once the program is compiled and the four input files are present, the program can be executed directly. The input files must be specified when calling the executable, using the following options:

Required arguments
- `-g --gesp` Followed by the gesp file name;
- `-m --mol2` Followed by the mol2 file name;
- `-p --pol` Followed by the polarisability file name;
- `-c --constr` Followed by the constraints file name;

Optional arguments
- `-x --screen` Activate charge–dipole screening as in the original article by Wang. The default is not to include it;
- `-db --database` Print database; followed by database file name;
- `-gi --gaussian` Print Gaussian-style input in log file; the default is to not print;
- `-cs --connshift` Shift connectivity by fixed amount in Gaussian-style input; followed by integer;
- `-r --resnum` Specify residue ID in Gaussian-style input; followed by integer; the default is 1;
- `-s --silent` Run in silent mode (minimum printout);
- `-v --verbose` Run in verbose mode (extra printout);
- `-d --debug` Run in debug mode (huge printout);
- `-h --help` Print out a help message and quit.

    For instance:

```
$> ./polchat.exe -g MOL.gesp -m MOL.mol2 -p MOL.pol
      -c MOL.cns -x -v -db MOL.db -gi -cs 25 -r 140
```

The database file produced is compatible with other Molecolab tools, including QMIP.[4] Its structure is as follows:

```
WAT    O   -0.8340   -0.6824   10.0600   8
WAT    H1   0.4170    0.3412    0.0000   1
WAT    H2   0.4170    0.3412    0.0000   1
[...]
```

The first two columns specify the residue name and the atom name. The other columns specify the MM charge, the MMPol charge (i.e. the charge that is used when the atom is polarisable), the polarisability in atomic units, and the atomic number.

Polarisable MM Model

The program is based on the induced dipole model, following the parametrisation of the AL model of Wang et al., *J. Phys. Chem. B* **115**, 3091 (2011).

In such model, the charge–dipole interaction is screened. However, please remember that, by default, such interaction is *not* screened. If you wish to activate the screening (consistently with the original model), please use the `-x / --screen` option when executing.

Theory

The fitting is based on the minimisation of the error function, which is the sum of squared differences between the calculated QM potential (read from the gesp file) and the potential resulting from the charges. Here we will use the indices i, j, ... to refer to gridpoints, and m, n, ... to refer to atoms. The QM potential V^{QM} is available on a grid.

$$J = \sum_i^{N_{\mathrm{G}}} \left[V_i^{\mathrm{QM}} - V_i^{\mathrm{chg}} - V_i^{\mathrm{pol}} \right]^2 + J_{\mathrm{c}} = \sum_i^{N_{\mathrm{G}}} \left[V_i^{\mathrm{QM}} - \sum_m^{N_{\mathrm{A}}} \frac{q_m}{r_{im}} - \sum_m^{N_{\mathrm{A}}} \frac{\boldsymbol{\mu}_m \cdot \mathbf{r}_{im}}{r_{im}^3} \right]^2 + J_{\mathrm{c}}$$

(1)

where $\mathbf{r}_{im} \equiv \mathbf{r}_i - \mathbf{r}_m$ and J_{c} is an additional term eventually including the constraints (see later).

---

[4]QMIP: QM/MM input preparation. L Cupellini, S Jurinovich & B Mennucci. Molecolab Tools. 2015-2017 `molecolab.dcci.unipi.it/tools`.

Main function

The minimum of this function is obtained by setting to zero the derivative with respect to the charges:

$$\frac{\partial J}{\partial q_x} = -2\sum_i^{N_G} \left[V_i^{QM} - V_i^{chg} - V_i^{pol} \right] \left[\frac{\partial}{\partial q_x}\left(\sum_m^{N_A} \frac{q_m}{r_{im}} \right) + \frac{\partial}{\partial q_x}\left(\sum_j^{N_A} \frac{\boldsymbol{\mu}_m \cdot \mathbf{r}_{im}}{r_{im}^3} \right) \right]$$

(2)

Some definitions:

$$\mathbf{D}_{mn} \text{ is the MMPol matrix tensor} \qquad (3N_A, 3N_A) \qquad (3)$$

$$A_{im} = \frac{1}{r_{im}} \qquad (N_G, N_A) \qquad (4)$$

$$\mathbf{B}_{im}^G = \frac{\mathbf{r}_{im}}{r_{im}^3} \qquad (N_G, 3N_A) \qquad (5)$$

$$\mathbf{B}_{mn}^A = \frac{\mathbf{r}_{mn}}{r_{mn}^3} \qquad (3N_A, N_A) \qquad (6)$$

$$\mathbf{C} = \mathbf{D}^{-1}\mathbf{B}^A \qquad (3N_A, N_A) \qquad (7)$$

$$\mathbf{E} = \mathbf{B}^G\mathbf{C} \qquad (N_G, N_A) \qquad (8)$$

$$\mathbf{F} = \mathbf{A} + \mathbf{E} \qquad (N_G, N_A) \qquad (9)$$

$$\mathbf{G} = \mathbf{F}^\dagger\mathbf{F} \qquad (N_A, N_A) \qquad (10)$$

$$\mathbf{H} = \mathbf{F}^\dagger\mathbf{V}^{QM} \qquad N_A \qquad (11)$$

Therefore, the potential and electric field at a point $\mathbf{r}_\eta$ induced by the distribution of charge on the atoms are:

$$V_\eta = \sum_m^{N_A} \frac{q_m}{r_{\eta m}} = \sum_m^{N_A} A_{\eta m} q_m = (\mathbf{A}\mathbf{q})_\eta \quad (\eta \text{ is a gridpoint}) \qquad (12)$$

$$\boldsymbol{\Phi}_\eta = \begin{cases} \sum_m^{N_A} \frac{q_m \mathbf{r}_{\eta m}}{r_{\eta m}^3} = \sum_l^n \mathbf{B}_{\eta m}^G q_m = (\mathbf{B}^G\mathbf{q})_\eta & \text{if } \eta \text{ is a gridpoint} \\ \sum_m^{N_A} \frac{q_m \mathbf{r}_{\eta m}}{r_{\eta m}^3} = \sum_l^n \mathbf{B}_{\eta m}^A q_m = (\mathbf{B}^A\mathbf{q})_\eta & \text{if } \eta \text{ is an atom} \end{cases} \qquad (13)$$

The dipole induced at a site m by the surrounding charges therefore is:

$$\boldsymbol{\mu}_m = \sum_n^{N_A} \mathbf{D}_{mn}^{-1}\boldsymbol{\Phi}_n = \sum_{n,n'}^{N_A} \mathbf{D}_{mn}^{-1}\mathbf{B}_{nn'}^A q_{n'} = \sum_{n'}^{N_A} \mathbf{C}_{mn'} q_{n'} = (\mathbf{C}\mathbf{q})_m \qquad (14)$$

and the electrostatic potential induced by this on a gridpoint i is:

$$\sum_m^{N_A} \frac{\boldsymbol{\mu}_m \cdot \mathbf{r}_{im}}{r_{im}^3} = \sum_m^{N_A} \mathbf{B}_{im}^G \boldsymbol{\mu}_m = (\mathbf{B}^G\boldsymbol{\mu})_i = (\mathbf{B}^G\mathbf{C}\mathbf{q})_i = (\mathbf{E}\mathbf{q})_i \qquad (15)$$

The terms in equations (12) and (15) represent the electrostatic potential due to the distribution of charge, and by the distribution of dipoles induced by it. They

must be differentiated with respect to the charges, as in equation (2):

$$\frac{\partial}{\partial q_x} \left(\sum_m^{N_A} \frac{q_m}{r_{im}} \right) = \frac{\partial}{\partial q_x} \left(\sum_m^{N_A} A_{im} q_m \right) = A_{ix} \qquad (16)$$

$$\frac{\partial}{\partial q_x} \left(\sum_m^{N_A} \frac{\boldsymbol{\mu}_m \cdot \mathbf{r}_{im}}{r_{im}^3} \right) = \frac{\partial}{\partial q_x} \left(\sum_m^{N_A} E_{im} q_m \right) = E_{ix} \qquad (17)$$

Therefore equation 2 becomes:

$$\frac{\partial J}{\partial q_x} = -2 \sum_i^{N_G} \left[V_i^{QM} - \sum_m^{N_A} A_{im} q_m - \sum_m^{N_A} E_{im} q_m \right] [A_{ix} + E_{ix}] =$$

$$= -2 \sum_i^{N_G} \left[V_i^{QM} - \sum_m^{N_A} F_{im} q_m \right] F_{ix} \qquad (18)$$

Setting this to zero to obtain the minimising charges gives:

$$2 \sum_i^{N_G} \sum_m^{N_A} F_{xi}^\dagger F_{im} q_m = 2 \sum_i^{N_G} F_{xi}^\dagger V_i^{QM} \qquad (19)$$

$$2 \sum_m^{N_A} G_{xm} q_m = 2 H_x \qquad (20)$$

This can be rewritten in matrix form as:

$$2\mathbf{G}\mathbf{q} = 2\mathbf{H} \qquad (21)$$

and the system is solved by:

$$\mathbf{q} = \mathbf{G}^{-1}\mathbf{H} \qquad (22)$$

Constraints

Constraints are of three different kinds: total charge of a fragment, equivalence between n atoms and restraint. The conservation of the total molecular charge is a particular case of the first kind, where the fragment includes all atoms.

In general, all constraints considered here are linear, $i.e.$, they can be written as:

$$\sum_m^{N_A} c_m q_m = d \qquad (23)$$

or, in matrix form:

$$\mathbf{c}^\dagger \mathbf{q} = d \qquad (24)$$

The constraints are introduced by adding a constraint term J_c to the error function, as in equation (1):

$$J_c = \sum_\alpha^{N_c} \lambda_\alpha \left(\sum_m^{N_A} C_{m\alpha} q_m - d_\alpha \right) \qquad (25)$$

where index α labels the N_c constraints, and matrix $\mathbf{C}$ contains the coefficients for each constraint, $\mathbf{c}_\alpha$, on the rows:

$$
\mathbf{C} = \begin{bmatrix} \vdots & \vdots & & \vdots \\ \mathbf{c}^1 & \mathbf{c}^2 & \cdots & \mathbf{c}^{N_c} \\ \vdots & \vdots & & \vdots \end{bmatrix}
\tag{26}
$$

J is now function of the N_A charges and of the N_c constraints: $J = J(\mathbf{q}, \boldsymbol{\lambda})$. Its derivative with respect to the charges is as in equation (19) plus a term coming from J_c:

$$
\frac{\partial J_c}{\partial q_x} = \sum_\alpha C_{x\alpha} \lambda_\alpha
\tag{27}
$$

so that:

$$
\frac{\partial J}{\partial q_x} = 2 \sum_m^{N_A} G_{xm} q_m + \sum_\alpha C_{x\alpha} \lambda_\alpha - 2H_x
\tag{28}
$$

Setting this to zero one obtains a modified version of equation (21):

$$
2\mathbf{G}\mathbf{q} + \mathbf{C}\boldsymbol{\lambda} = 2\mathbf{H}
\tag{29}
$$

Since the constraints introduced new variables $\boldsymbol{\lambda}$, it is also needed to differentiate with respect to them:

$$
\frac{\partial J}{\partial \lambda_x} = \frac{\partial J_c}{\partial \lambda_x} = \sum_m^{N_A} C_{mx} q_m - d_x
\tag{30}
$$

setting which to zero one obtains:

$$
\mathbf{C}^\dagger \mathbf{q} = \mathbf{d}
\tag{31}
$$

Equations (30) and (31) can be satisfied at once by solving:

$$
\begin{bmatrix} 2\mathbf{G} & \mathbf{C} \\ \mathbf{C}^\dagger & 0 \end{bmatrix} \begin{bmatrix} \mathbf{q} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} 2\mathbf{H} \\ \mathbf{d} \end{bmatrix} \quad \Rightarrow \quad \begin{bmatrix} \mathbf{q} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} 2\mathbf{G} & \mathbf{C} \\ \mathbf{C}^\dagger & 0 \end{bmatrix}^{-1} \begin{bmatrix} 2\mathbf{H} \\ \mathbf{d} \end{bmatrix}
\tag{32}
$$

The form of matrices $\mathbf{C}$ and $\mathbf{d}$ for each case of constraint is detailed below.

Charge of a fragment   The total charge of a fragment F is set to q^{frg}. Therefore the constraint function is:

$$
J_c = \lambda \left(\sum_{m \in F} q_m - q^{\text{frg}} \right)
\tag{33}
$$

Therefore:

$$
c_i = \begin{cases} 1, & i \in F \\ 0, & i \notin F \end{cases}
\tag{34}
$$

and $d = q^{\text{frg}}$.

Equivalence of atoms   If condition $q_r = q_s = q_t \dots$ holds, this can be split into a system of equations:

$$J_{c1} = \lambda_1 \left(q_r - q_s \right)$$
$$J_{c2} = \lambda_2 \left(q_s - q_t \right)$$

and so on. Therefore $C_{i1} = \delta_{ir} - \delta_{is}$, $C_{i2} = \delta_{is} - \delta_{it}$, and $\mathbf{d} = \mathbf{0}$.

Restraint   If the charges are restrained using a restraint parameter, an extra function is added so that large absolute values of charges are discouraged. The extra function is:

$$J_c = \alpha \sum_{m \in R} q_m^2 \tag{35}$$

where set R includes the atoms to be restrained. This function corresponds to that used in Model 2 of Bayly *et al.*, *J. Phys. Chem.* **97**, 10269 (1993). Note that a hyperbolic penalty function provided better results than the parabolic one, but the latter has linear derivatives in the charges and does not therefore require an iterative solution.

The derivative of the constrained function is linear in $\mathbf{q}$:

$$\frac{\partial}{\partial q_x} J_c = \begin{cases} 2\alpha q_x & x \in R \\ 0 & x \notin R \end{cases} \tag{36}$$

A diagonal term must be added to $\mathbf{G}$ of Equation(20):

$$\mathbf{G_r} = \mathbf{G} + \alpha \mathbf{\Delta}^r \tag{37}$$

where

$$\mathbf{\Delta}^r_{mn} = \begin{cases} 1 & m = n \in R \\ 0 & \text{otherwise} \end{cases} \tag{38}$$

Dimension of the constrained problem   If the problem has constraints (excluding restraints), the dimension of the problem changes from N_A to $N_A + N_c$, where N_c is the total number of constraints. Note that each constraint on the charge of the molecule or of some fragment counts as one, while each constraint on the equivalence between n atoms counts as $n - 1$.

Download

The binary code can be downloaded free of charge from the Molecolab website: `molecolab.dcci.unipi.it/tools`. You may need to send an email to require download.

Please contact the authors for comments or questions.

Citation

Please cite this tool as:

PolChat: A polarisation-consistent charge fitting tool.
S. Caprasecca, C. Curutchet & B. Mennucci.
Molecolab Tools. 2014-2017 `molecolab.dcci.unipi.it/tools`