PolChat
A polarisation-consistent charge-fitting tool

Stefano Caprasecca[1], *Carles Curutchet and Sandro Jurinovich*

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:

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

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

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

`xxx.cns` a file with constraints;

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

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 `xxx.com`:

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

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

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

xxx.gesp
```

[1] `stefano.caprasecca@for.unipi.it`

[2] `www.dcci.unipi.it/molecolab` — For help please write to `molecolab.pisa@gmail.com` or to the reference author.

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 xxx.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 xxx.gesp -fi gesp -o xxx.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
xxx
   60    59     1     0     0
[...]

@<TRIPOS>BOND
     1     1    2 1
     2     1   21 1
     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 `xxx.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 `xxx.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 `xxx.cns` file contains the constraints you wish to apply to your fit. Note that the program perform both a standard, non-polarisable ESP fit, and a polarisable one. Both fits are constrained as requested, including the constraint on the total dipole moment.

The `xxx.cns` file *must* contain, *on the first line*, the total charge of the molecule. Additional constraints are optional.

total charge (Required) The total charge constraint is *not optional* and must be entered on the *first line* of the file.

fragment charge (Optional) It is possible to specify that a certain fragment of the molecule must sum to a certain charge $\overline{c}$. Use the keyword `fragm`, then on the next line specify the number of atoms belonging to the fragment and the total charge, and on the following line the list of the atoms of the fragment, for instance:

```
fragm
 4  1.0
 9 15 17 19
```

This requires that the fragment made up by the 4 atoms 9,15,17 and 19 sums up to charge $+1.0$.

equivalence (Optional) The equivalence between atoms is specified using the keyword `equiv`, followed on the next line by the number of equivalent atoms and on the next line by the list of such atoms, as in the example:

```
equiv
 3
 9 10 11
```

This requires that the three atoms 9, 10 and 11 have the same charge.

total dipole (Optional) Constraining the total dipole (i.e. the sum of the fixed molecular dipole given by the charge distribution, plus all the induced dipoles) to a certain value can be done by using the keyword `dipole`. Subkeywords are:

`qm` use the QM molecular dipole as read from Gaussian calculation;

`esp` use the dipole computed from the ESP charges as from the `xxx.gesp` file;

`read` read the dipole moment, in a.u., from the `xxx.cns` file (on the next line). See examples below.

See the examples below:

```
dipole
qm
```

```
dipole
esp
```

```
dipole
read
0.250  -0.375   0.000
```

There is no limit on the number of constraints imposed, as long as the conditions are linearly independent. Only one dipole constraint may be imposed. The order the conditions are written in the file is irrelevant, except for the total charge, specified on the first line. See an example of a constraint file below:

```
0.00
dipole
read
0.250   -0.375   0.000
equiv
 3
 1  2  3
fragm
 2  1.0
 5  6
equiv
 2
15 16
equiv
 2
18 19
```

Running the program

Compilation The tool is available already compiled for a 64-bit machine. Should you need other solutions, the authors will be happy to help.

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:

-g (required) Followed by the gesp file name

-m (required) Followed by the mol2 file name

-p (required) Followed by the polarisability file name

-c (required) Followed by the constraints file name

-x (optional) Also include Wang Charge–Polarisability screening. The default is not to include it. Only use if the polarisable MM model includes such screening.

-h (optional) Print out a help message and quit

-s (optional) Run in silent mode (minimum printout)

-d (optional) Run in debug mode (extra printout)

if neither -s or -d are used, the program will run in normal mode (standard printout)

For instance:

```
    $> ./chpol.exe -g xxx.gesp -m xxx.mol2 -p xxx.pol

       -c xxx.cns -d
```

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` option when executing.

Theory

The function to minimize is the squared difference between the QM potential $\overline{V}$ calculated on a grid and that generated by a set of charges and dipoles induced by those charges:

$$J = \sum_i^m \left[\overline{V}_i - V^{\mathrm{chg}} - V^{\mathrm{pol}}\right]^2 + J_{\mathrm{c}} = \sum_i^m \left[\overline{V}_i - \sum_j^n \frac{q_j}{r_{ij}} - \sum_j^n \frac{\boldsymbol{\mu}_j \cdot \mathbf{r}_{ij}}{r_{ij}^3}\right]^2 + J_{\mathrm{c}} \quad (1)$$

where i runs over the m grid points, j over the n MM atoms, $\mathbf{r}_{ij} \equiv \mathbf{r}_i - \mathbf{r}_j$ and J_{c} is an additional term eventually including the constraints (see later).

Main function

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

$$\frac{\partial J}{\partial q_x} = -\sum_i^m 2\left[\overline{V}_i - V^{\mathrm{chg}} - V^{\mathrm{pol}}\right]\left[\frac{\partial}{\partial q_x}\left(\sum_j^n \frac{q_j}{r_{ij}}\right) + \frac{\partial}{\partial q_x}\left(\sum_j^n \frac{\boldsymbol{\mu}_j \cdot \mathbf{r}_{ij}}{r_{ij}^3}\right)\right] \tag{2}$$

Some definitions:

$$A_{ij} \equiv \frac{1}{r_{ij}} \tag{3}$$

$$\mathbf{B}_{ij} \equiv \frac{\mathbf{r}_{ij}}{r_{ij}^3} \tag{4}$$

$$\mathbf{C}_{ij} \equiv \sum_k^n D_{ik}^{-1}\mathbf{B}_{kj} \tag{5}$$

$$E_{ij} \equiv \sum_j^n \mathbf{B}_{ik}\mathbf{C}_{kj} \tag{6}$$

$$F_{ij} \equiv A_{ij} + E_{ij} \tag{7}$$

$$G_{ij} \equiv \sum_k^m F_{ik}^{\dagger}F_{kj} \tag{8}$$

$$H_i = \sum_j^m F_{ij}^{\dagger}\overline{V}_j \tag{9}$$

So:

$$\sum_j^n \frac{q_j}{r_{ij}} = \sum_j^n A_{ij}q_j \tag{10}$$

$$\boldsymbol{\Phi}_k = \sum_l^n \frac{q_l \mathbf{r}_{kl}}{r_{kl}^3} = \sum_l^n \mathbf{B}_{kl}q_l \tag{11}$$

$$\boldsymbol{\mu}_j = \sum_k^n \mathbf{D}_{jk}^{-1}\boldsymbol{\Phi}_k = \sum_k^n\sum_l^n \mathbf{D}_{jk}^{-1}\mathbf{B}_{kl}q_l = \sum_l^n \mathbf{C}_{jl}q_l \tag{12}$$

$$\sum_j^n \frac{\boldsymbol{\mu}_j \cdot \mathbf{r}_{ij}}{r_{ij}^3} = \sum_j^n \mathbf{B}_{ij}\boldsymbol{\mu}_j = \sum_j^n\sum_l^n \mathbf{B}_{ij}\mathbf{C}_{jl}q_l = \sum_l^n E_{il}q_l \tag{13}$$

and the derivatives become:

$$\frac{\partial}{\partial q_x}\left(\sum_j^n \frac{q_j}{r_{ij}}\right) = \frac{\partial}{\partial q_x}\left(\sum_j^n A_{ij}q_j\right) = A_{ix} \tag{14}$$

$$\frac{\partial}{\partial q_x}\left(\sum_j^n \frac{\boldsymbol{\mu}_j \cdot \mathbf{r}_{ij}}{r_{ij}^3}\right) = \frac{\partial}{\partial q_x}\left(\sum_j^n E_{ij}q_j\right) = E_{ix} \tag{15}$$

Therefore equation 2 becomes:

$$\begin{aligned}\frac{\partial J}{\partial q_x} &= -2\sum_i^m \left[\overline{V}_i - \sum_j^n A_{ij}q_j - \sum_j^n E_{ij}q_j\right][A_{ix}+E_{ix}] = \\ &= -2\sum_i^m \left[\overline{V}_i - \sum_j^n F_{ij}q_j\right] F_{ix} \qquad (16)\end{aligned}$$

Setting this to zero to obtain the minimizing charges gives:

$$\sum_i^m \left[\overline{V}_i - \sum_j^n F_{ij}q_j\right] F_{ix} = 0 \qquad (17)$$

$$\sum_i^m \sum_j^n F^{\dagger}_{xi}F_{ij}q_j = \sum_i^m F_{ix}\overline{V}_i \qquad (18)$$

$$\sum_j^n G_{xj}q_j = H_x \qquad (19)$$

This can be rewritten in matrix form as:

$$\mathbf{Gq} = \mathbf{H} \qquad (20)$$

and the system is solved by:

$$\mathbf{q} = \mathbf{G}^{-1}\mathbf{H} \qquad (21)$$

Constraints

Constraints are introduced in function J_c. These are of three different kinds: total charge of a fragment, equivalence between n atoms, and conservation of the total dipole moment. The conservation of the total molecular charge is a particular case of the first kind.

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

$$J_c = \lambda\left(\sum_{j\in F} q_j - \overline{q}\right) \qquad (22)$$

It follows that:

$$\frac{\partial}{\partial q_x}J_c = \lambda q_x \qquad (23)$$

$$\frac{\partial}{\partial \lambda}J_c = \sum_{j\in F} q_j - \overline{q} \qquad (24)$$

If the fragment comprises atoms from r to s, the new matrices are:

$$\mathbf{G}_c = \begin{bmatrix} G_{1,1} & G_{1,2} & \cdots & G_{1,r} & \cdots & G_{1,s} & \cdots & G_{1,n} & 0 \\ G_{2,1} & G_{2,2} & \cdots & G_{2,r} & \cdots & G_{2,s} & \cdots & G_{2,n} & 0 \\ \vdots & & & & & & & & 0 \\ G_{r,1} & G_{r,2} & \cdots & G_{r,r} & \cdots & G_{r,s} & \cdots & G_{r,n} & 1 \\ \vdots & & & & & & & & 1 \\ G_{s,1} & G_{s,2} & \cdots & G_{s,r} & \cdots & G_{s,s} & \cdots & G_{s,n} & 1 \\ \vdots & & & & & & & & 0 \\ G_{n,1} & G_{n,2} & \cdots & G_{n,r} & \cdots & G_{n,s} & \cdots & G_{n,n} & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} \quad \mathbf{H}_c = \begin{bmatrix} H_1 \\ H_2 \\ \vdots \\ H_r \\ \vdots \\ H_s \\ \vdots \\ H_n \\ \overline{q} \end{bmatrix} \tag{25}$$

The same holds even in the general case when the fragment comprises non-sequential atoms.

Equivalence of atoms If condition $q_i = q_j = q_k \ldots$ holds, this can be split into a system of equations:

$$J_{c1} = \lambda_1 (q_i - q_j)$$
$$J_{c2} = \lambda_2 (q_j - q_k)$$

and so on. For each condition:

$$\frac{\partial}{\partial q_x} J_{c1} = \lambda(\delta_{i,x} - \delta_{j,x}) \tag{26}$$

$$\frac{\partial}{\partial \lambda_1} J_{c1} = q_i - q_j \tag{27}$$

and the new matrices are:

$$\mathbf{G}_c = \begin{bmatrix} G_{1,1} & G_{1,2} & \cdots & G_{1,i} & \cdots & G_{1,j} & \cdots & G_{1,n} & 0 \\ G_{2,1} & G_{2,2} & \cdots & G_{2,i} & \cdots & G_{2,j} & \cdots & G_{2,n} & 0 \\ \vdots & & & & & & & & 0 \\ G_{i,1} & G_{i,2} & \cdots & G_{i,i} & \cdots & G_{i,j} & \cdots & G_{i,n} & 1 \\ \vdots & & & & & & & & 0 \\ G_{j,1} & G_{j,2} & \cdots & G_{j,i} & \cdots & G_{j,j} & \cdots & G_{j,n} & -1 \\ \vdots & & & & & & & & 0 \\ G_{n,1} & G_{n,2} & \cdots & G_{n,i} & \cdots & G_{n,j} & \cdots & G_{n,n} & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \end{bmatrix} \quad \mathbf{H}_c = \begin{bmatrix} H_1 \\ H_2 \\ \vdots \\ H_i \\ \vdots \\ H_j \\ \vdots \\ H_n \\ 0 \end{bmatrix} \tag{28}$$

Conservation of the total dipole moment If the total dipole moment must be equal to $\overline{\boldsymbol{\mu}}$, this corresponds to three conditions:

$$J_c = \lambda \left(\sum_j \boldsymbol{\mu}_j + \boldsymbol{\mu}^{\text{chg}} - \overline{\boldsymbol{\mu}} \right) \tag{29}$$

where $\boldsymbol{\mu}_j$ are the induced atomic dipole moments and $\boldsymbol{\mu}^{\text{chg}}$ is the fixed dipole given by the charges. Note that when non-polarisable ESP charges are being calculated, the induced dipole moments are assumed to be zero. In such case, the equations below still hold, with $\mathbf{C}_{jk} = 0$ and $\tilde{\mathbf{C}}_k = 0$, and therefore with $\mathbf{K}_k = \mathbf{r}_k$. Equation 29 can be rewritten as:

$$\begin{aligned} J_{\text{c}} =& \lambda \left(\sum_j \left(\sum_k \mathbf{C}_{jk} q_k + \mathbf{r}_j q_j \right) - \overline{\boldsymbol{\mu}} \right) = \\ =& \lambda \left(\sum_k \left(\tilde{\mathbf{C}}_k q_k + \mathbf{r}_k q_k \right) - \overline{\boldsymbol{\mu}} \right) = \lambda \left(\sum_k \mathbf{K}_k q_k - \overline{\boldsymbol{\mu}} \right) \end{aligned} \tag{30}$$

where $\tilde{\mathbf{C}}_k \equiv \sum_j \mathbf{C}_{jk}$ and $\mathbf{K}_k \equiv \tilde{\mathbf{C}}_k + \mathbf{r}_k$. The derivatives are:

$$\frac{\partial}{\partial q_x} J_{\text{c}} = \lambda \mathbf{K}_x \tag{31}$$

$$\frac{\partial}{\partial \lambda} J_{\text{c}} = \sum_k \mathbf{K}_k q_k - \overline{\boldsymbol{\mu}} \tag{32}$$

The new matrices are:

$$\mathbf{G}_c = \begin{bmatrix} G_{1,1} & G_{1,2} & \cdots & G_{1,n} & K_1^{(x)} & K_1^{(y)} & K_1^{(z)} \\ G_{2,1} & G_{2,2} & \cdots & G_{2,n} & K_2^{(x)} & K_2^{(y)} & K_2^{(z)} \\ \vdots & & & & & & \\ G_{n,1} & G_{n,2} & \cdots & G_{n,n} & K_n^{(x)} & K_n^{(y)} & K_n^{(z)} \\ K_1^{(x)} & K_2^{(x)} & \cdots & K_n^{(x)} & 0 & 0 & 0 \\ K_1^{(y)} & K_2^{(y)} & \cdots & K_n^{(y)} & 0 & 0 & 0 \\ K_1^{(z)} & K_2^{(z)} & \cdots & K_n^{(z)} & 0 & 0 & 0 \end{bmatrix} \quad \mathbf{H}_{\text{c}} = \begin{bmatrix} H_1 \\ H_2 \\ \vdots \\ H_n \\ \overline{\boldsymbol{\mu}}^{(x)} \\ \overline{\boldsymbol{\mu}}^{(x)} \\ \overline{\boldsymbol{\mu}}^{(x)} \end{bmatrix} \tag{33}$$

Download

The binary code can be downloaded free of charge from the Molecolab website: `www.dcci.unipi.it/molecolab/tools`.

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, S. Jurinovich.
Molecolab Tools. 2014 `www.dcci.unipi.it/molecolab/tools`