



readme.txt

<https://upjv.q4md-forcefieldtools.org/REDServer-Development/Documentation/readme.txt>

Documentation of February 2015

Last update of this documentation:

December 10th, 2024

Always reload this file, when reading this file with a web browser to be sure to read the latest version!

The Tripos Mol2 file format is not a correct input file format for PyRED!

The P2N file format is not a correct input file format for PyRED!

(Chemical equivalencing used in charge derivation is automatically carried out by PyRED: no need to execute Ante_RED.pl)

Use instead the PDB file format as input file format:

To be recognized a PDB input file must have a particular filename and the '.pdb' extension (see below)!

DO NOT FORGET TO ADD THE HYDROGEN ATOMS IN THE PDB INPUT FILES IF NEEDED!

All the chemical elements of the periodic table (but Francium & Radium) are handled by R.E.D.Server Development/PyRED

Molecules with even and odd numbers of electrons are both handled by R.E.D. Server Development/PyRED

Large molecules such as macromolecules, polymers or supramolecular systems are handled by R.E.D. Server Development/PyRED: such large molecular systems are not directly provided as inputs to R.E.D. Server Development/PyRED, but they are reconstructed from molecular fragments, which are generated from elementary building blocks or small input models with capping groups (with a number of atoms below 250)

I. THE PDB FILE FORMAT USED BY PyRED

Adaptation of <https://www.wwpdb.org/documentation/file-format-content/format33/sect9.html>

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"ATOM" or "HETATM"	
		(lines starting with "ANISOU" are excluded)	
7 - 11	Integer	serial	Atom serial number
13 - 16	Atom	name	Atom name
18 - 20	Residue name	resName	Residue name
23 - 26	Integer	resSeq	Residue sequence number
31 - 38	Real(8.3)	x	Orth. coordinates for X in Angstroms
39 - 46	Real(8.3)	y	Orth. coordinates for Y in Angstroms
47 - 54	Real(8.3)	z	Orth. coordinates for Z in Angstroms
77 - 78	LString(2)	element	Element symbol, right-justified
79 - 80	LString(2)	charge	Charge on the atom

Example of PDB input file composed of two conformations - requirements:

- A conformation is introduced by the "MODEL" keyword with a conformation index, which is incremented: MODEL 1, MODEL 2, ... MODEL n.
(the ENDMDL and TER keywords may be present, but are not read/used)
- The different conformations must have the SAME atom order.

- Two atoms belonging to the same residue cannot share the same name.

Example: conformations 'anti' and 'gauche' of ethanol:

```

      1      2      3      4      5      6      7      8
123456789012345678901234567890123456789012345678901234567890
-----|----|---|--|---|-----|-----|-----|
MODEL 1
ATOM    1  C1  ETO    1      1.1645 -0.4192  0.0000      C
ATOM    2  H11 ETO    1      2.1088  0.1164  0.0000      H
ATOM    3  H12 ETO    1      1.1258 -1.0533  0.8786      H
ATOM    4  H13 ETO    1      1.1258 -1.0533 -0.8786      H
ATOM    5  C2  ETO    1      0.0000  0.5509  0.0000      C
ATOM    6  H21 ETO    1      0.0488  1.1929  0.8777      H
ATOM    7  H22 ETO    1      0.0488  1.1929 -0.8777      H
ATOM    8  O3  ETO    1     -1.1890 -0.1968  0.0000      O
ATOM    9  H3  ETO    1     -1.9331  0.3883  0.0000      H
TER     10      ETO    1
ENDMDL
MODEL 2
ATOM    1  C1  ETO    1      1.2090 -0.2365 -0.0192      C
ATOM    2  H11 ETO    1      2.0673  0.4240  0.0606      H
ATOM    3  H12 ETO    1      1.2684 -0.9569  0.7932      H
ATOM    4  H13 ETO    1      1.2755 -0.7790 -0.9558      H
ATOM    5  C2  ETO    1     -0.0899  0.5530  0.0449      C
ATOM    6  H21 ETO    1     -0.1457  1.1129  0.9768      H
ATOM    7  H22 ETO    1     -0.1315  1.2706 -0.7639      H
ATOM    8  O3  ETO    1     -1.2219 -0.2625 -0.1100      O
ATOM    9  H3  ETO    1     -1.2727 -0.8706  0.6144      H
ENDMDL
```

For a molecule, columns 79-80 may be used to provide charge(s) on atom(s) (two characters; example: 1-, 1+, 2+ ...): by summing these charges PyRED calculates the molecule total charge value, which is required in QM geometry optimization/MEP computation. The molecule total charge value can also be provided by using the "MOLECULE'n'-TOTCHARGE" keyword ('n' = molecule index) in the 'Project.config' file. If different total charge values are determined from the PDB input file and from the 'Project.config' file, the value defined in the 'Project.config' file is the one, that is used (update Sept. 2024).

Example: ethanolate:

```

      1      2      3      4      5      6      7      8
123456789012345678901234567890123456789012345678901234567890
-----|----|---|--|---|-----|-----|-----|
HETATM  1  C1  ETO    1      1.2090 -0.2365 -0.0192      C
HETATM  2  H11 ETO    1      2.0673  0.4240  0.0606      H
HETATM  3  H12 ETO    1      1.2684 -0.9569  0.7932      H
HETATM  4  H13 ETO    1      1.2755 -0.7790 -0.9558      H
HETATM  5  C2  ETO    1     -0.0899  0.5530  0.0449      C
HETATM  6  H21 ETO    1     -0.1457  1.1129  0.9768      H
HETATM  7  H22 ETO    1     -0.1315  1.2706 -0.7639      H
HETATM  8  O3  ETO    1     -1.2219 -0.2625 -0.1100      O1-
Total charge value of ethanolate = -1
```

Example: Zwitterion of Glycine at pH = 7:

```

      1      2      3      4      5      6      7      8
123456789012345678901234567890123456789012345678901234567890
-----|----|---|--|---|-----|-----|-----|
ATOM    1  N   GLY    1     -1.920  0.218  -0.068  1.00  0.00      N1+
ATOM    2  HN1 GLY    1     -2.792 -0.255  0.051  1.00  0.00      H
ATOM    3  HN2 GLY    1     -1.850  0.559 -1.006  1.00  0.00      H
ATOM    4  HN3 GLY    1     -1.867  0.986  0.571  1.00  0.00      H
ATOM    5  CA  GLY    1     -0.817 -0.716  0.196  1.00  0.00      C
ATOM    6  HA1 GLY    1     -0.892 -1.087  1.218  1.00  0.00      H
ATOM    7  HA2 GLY    1     -0.874 -1.553 -0.500  1.00  0.00      H
ATOM    8  C   GLY    1      0.499 -0.003  0.017  1.00  0.00      C
```

```

ATOM      9  O  GLY      1      1.652 -0.663  0.208  1.00  0.00      0
ATOM     10  OXT GLY      1      0.516  1.164 -0.299  1.00  0.00     01-
Total charge value of the zwitterion of Glycine = 0

```

The chemical element columns (right justified columns 77-78; 1 or 2 characters) allows differentiating atoms with ambiguous atom names:

i.e. Ca (Calcium), Cd (Cadmium), Ce (Cerium), Cl (Chlorine) versus C (carbon) or Na (sodium), Nb (Niobium), Nd (Neodymium), Ni (Nickel) versus N (Nitrogen) or Si (Silicium), Sc (Scandium), Se (Selenium), Sn (Tin) versus S (Sulfur) or Pd (Palladium), Pt (Platinum), Pa (Protactinium) versus (Phosphorus) etc...

See examples below:

```

      1      2      3      4      5      6      7      8
123456789012345678901234567890123456789012345678901234567890
-----|-----| ---| --| ---|      -----|-----|-----|
ATOM      1  CA  GLY      1      1.218 -0.218 -0.000      C
-> This case corresponds to an alpha-carbon atom with the CA atom name

```

```

ATOM      1  CA1 UNK      1      1.218 -0.218 -0.000      Ca2+
ATOM      1  CA2 UNK      1      3.218 -0.218 -0.000      2+
-> These cases correspond to two calcium atoms with the CA1 and CA2 atom names
    in a single residue

```

```

ATOM      1  NA  MOL      1      1.218 -0.218 -0.000      N
-> This is a nitrogen atom with the NA atom name within the MOL residue

```

```

ATOM      1  NA  NA1      1      1.218 -0.218 -0.000      Na1+
ATOM      1  NA  NA1      2      3.218 -0.218 -0.000      1+
-> These correspond to two sodium atoms with the NA atom name in two different
    residues

```

The set of "CONNECT" keywords is used ONLY if the "MOLECULE'n'-CALCONNECT = OFF" keyword (which is defined in the 'Project.config' file; 'n' = molecule index) is used. Below is an example of PDB input file for Ethanol, which shares a single set of "CONNECT" keywords for the two reported conformations with identical atom orders (the elements and charges on atoms are not provided because optional, and Cartesian coordinates with 4 digits after the decimal point are used).

```

MODEL 1
HETATM    1  C1  ETH      1      1.1645 -0.4192  0.0000
HETATM    2  H11 ETH      1      2.1088  0.1164  0.0000
HETATM    3  H12 ETH      1      1.1258 -1.0533  0.8786
HETATM    4  H13 ETH      1      1.1258 -1.0533 -0.8786
HETATM    5  C2  ETH      1      0.0000  0.5509  0.0000
HETATM    6  H21 ETH      1      0.0488  1.1929  0.8777
HETATM    7  H22 ETH      1      0.0488  1.1929 -0.8777
HETATM    8  O3  ETH      1     -1.1890 -0.1968  0.0000
HETATM    9  H3  ETH      1     -1.9331  0.3883  0.0000
MODEL 2
HETATM    1  C1  ETH      1      1.2090 -0.2365 -0.0192
HETATM    2  H11 ETH      1      2.0673  0.4240  0.0606
HETATM    3  H12 ETH      1      1.2684 -0.9569  0.7932
HETATM    4  H13 ETH      1      1.2755 -0.7790 -0.9558
HETATM    5  C2  ETH      1     -0.0899  0.5530  0.0449
HETATM    6  H21 ETH      1     -0.1457  1.1129  0.9768
HETATM    7  H22 ETH      1     -0.1315  1.2706 -0.7639
HETATM    8  O3  ETH      1     -1.2219 -0.2625 -0.1100
HETATM    9  H3  ETH      1     -1.2727 -0.8706  0.6144
CONNECT   1    4    5    2    3
CONNECT   2    1
CONNECT   3    1
CONNECT   4    1
CONNECT   5    7    8    1    6
CONNECT   6    5

```

```

CONNECT      7      5
CONNECT      8      5      9
CONNECT      9      8

```

Remarks:

-1- The chemical elements are not provided in the example above because they are unambiguously deduced from the atom names. No charge on atoms is also provided.

-2- Cartesian coordinates with 4 digits after the decimal point are used for a better accuracy.

-3- For organic molecules it is recommended NOT to include the 'CONNECT' keywords in the PDB input files. Indeed if "MOLECULE'n'-CALCONNECT = ON" is used, PyRED computes the atom connectivities. When studying a bioinorganic complex, providing the 'CONNECT' keywords in the PDB input file may be suitable ("MOLECULE'n'-CALCONNECT = OFF") to force the use of 'long' bonds between the metal center and its organic ligands in the force field library (mol2 output file). Another approach is to modify the value of the "'METAL'-RAD4TOP" keyword ('METAL' = element in capital letters in the 'Project.config' file).
See <https://upjv.q4md-forcefieldtools.org/REDServer-Development/Documentation/Project.config>
<https://upjv.q4md-forcefieldtools.org/Tutorial/Tutorial-4.php#II4>

-4- The chemical elements of the molecule(s) involved in force field generation can be checked by the owner of a job after submission once the job runs by looking at the temporary PyRED journal log file available from the qstat interface (click on the Job ID at <https://cluster.q4md-forcefieldtools.org/qstat/qstat.php>)
If a chemical element is found not correct the owner of a job can kill her/his job at <https://upjv.q4md-forcefieldtools.org/REDServer-Development/delete-log.php> and resubmit her/his job after having corrected the wrong element(s) in the PDB input file(s).

-5- We do work on the automatic correction of erroneous PDB input files: corrections are achieved by the PBS script (which is regularly updated):
Please, compare the Mol_red'n'.pdb and the Mol_red'n'.pdb-ArchiveFile files to study the adaptations, which are automatically done.

II. THE DIFFERENT MODES OF EXECUTION OF PyRED

II.1 If OPT_Calc = ON, MEPCHR_Calc = ON (and Re_Fit = OFF):

Just provide a list of PDB input files with the following names

- one PDB input file per molecule (a PDB file may contain several conformations):

Mol_red'n'.pdb ('n' is the molecule index, which is incremented
'n' starts at 1 and not at 0)

Please note the uppercase of the first letter of the filename and the underscore '_' character (on Linux the case of the letters matters)

See examples below:

Example 1: a single PDB input file is provided:
This corresponds to a single molecule job
Mol_red1.pdb

Example 2: five consecutive PDB input files are provided: this job involves five molecules
Mol_red1.pdb
Mol_red2.pdb
Mol_red3.pdb
Mol_red4.pdb
Mol_red5.pdb

Example 3: This job involves only three molecules: three consecutive PDB input files
Mol_red1.pdb

Mol_red2.pdb
Mol_red3.pdb
Mol_red5.pdb (is not used as input by PyRED because
Mol_red4.pdb is not found)

II.2 If OPT_Calc = OFF, MEPCHR_Calc = ON (and Re_Fit = OFF):

For each PDB input file at least one QM geometry optimization file (obtained by using a QM program such as Gaussian/GAMESS/Firefly) has to be given: this output file is used as input by PyRED

Thus the different conformations of a molecule are provided in:

- a single PDB input file, and in
- different QM geometry optimization output files used as inputs by PyRED

Please note the uppercase of the first letter of each filename, as well as the underscore '_' and the hyphen-minus '-' characters

See examples below:

Example 1: This corresponds to a single conformation - single molecule job:

Mol_red1.pdb Mol_red1-1.log

Example 2: This job involves five molecules

Mol_red1.pdb	Mol_red1-1.log	(1 conformation)		
Mol_red2.pdb	Mol_red2-1.log	Mol_red2-2.log	(2 conformations)	
Mol_red3.pdb	Mol_red3-1.log	(1 conformation)		
Mol_red4.pdb	Mol_red4-1.log	(1 conformation)		
Mol_red5.pdb	Mol_red5-1.log	Mol_red5-2.log	Mol_red5-3.log	(3 conformations)

Each PDB input file and its corresponding QM geometry optimization output(s)/inputs for PyRED must have the SAME atom order

The following file formats for the Gaussian/GAMESS/Firefly geometry optimization output files are expected by PyRED:

The Gaussian geometry optimization output:

```
Entering Gaussian System
[...]
Stationary point found
[...]
Input orientation 'or' Standard orientation
[...]
Normal termination
```

The GAMESS or Firefly/PC-GAMESS geometry optimization output:

```
[...]
GAMESS VERSION = 'or' Firefly Project homepage
[...]
EQUILIBRIUM GEOMETRY LOCATED
[...]
COORDINATES OF ALL ATOMS ARE
[...]
TERMINATED NORMALLY
```

Remark:

Up to 3 different QM output files can be concatenated in a Mol_red'n'-'i'.log file ('n' and 'i' are the molecule and conformation indexes):

- 1- The geometry optimization file, which provides the required Cartesian coordinates involved in MEP computation
- 2- A frequency job, which provides the frequencies that can be analyzed with the Jmol applets
- 3- A single point energy computation using a high theory level, which can be involved in dE(QM) versus dE(MM) energy comparison (between conformations)

II.3 If Re_Fit = ON:

If Re_Fit = ON (OPT_Calc = OFF and MEPCHR_Calc = OFF are forced)

A previous job has to present in the working directory:
Data of this previous job are located in the directory defined by the DIR variable (see the 'System.config' file).

III. THE System.config FILE

The System.config file only contains plain text (no rich text file format)
It has to be prepared using a text editor such as vi, gedit, nedit, geany or the notepad

Please note the uppercase of the first letter of the filename

The presence of the 'System.config' file in the working directory is not mandatory

Keywords provided in the 'System.config' file are related to the tasks performed by the PyRED program.

Adding the 'System.config' file in the working directory allows overwriting the default tasks (OPT_Calc = ON, MEPCHR_Calc = ON, Re_Fit = OFF, etc...) carried out by PyRED

In the absence of the 'System.config' file PyRED can still be executed
However, in this case only default tasks are carried out by PyRED
See <https://upjv.q4md-forcefieldtools.org/REDServer-Development/Documentation/System.config>

Combining the 'REMARK SYSTEMCFG' keywords and keywords, which are defined in the System.config file, in the Mol_red'n'.pdb input files, allows bypassing the use of a System.config file in the archive file. Thus, if a single PDB input file is needed for a PyRED job, a user can directly submit this PDB input file as an archive file without the need of the Project.config/
System.config
files.

Example:

```
REMARK SYSTEMCFG OPT_Calc           = ON2
REMARK SYSTEMCFG MEPCHR_Calc        = ON
REMARK SYSTEMCFG Freq_Calc          = ON
REMARK SYSTEMCFG CHR_TYP             = RESP-X1
REMARK SYSTEMCFG METHOD_OPTCALC      = B3LYP
REMARK SYSTEMCFG BASSET_OPTCALC     = 6-31+G(d,p)
REMARK SYSTEMCFG METHOD_MEPCALC      = Default
REMARK SYSTEMCFG BASSET_MEPCALC     = Default
[ atom indexes and names, residue name(s) and number(s), X Y Z Cart. coordinates, and chemical
elements to the PDB file format ]
```

IV. THE Project.config FILE

The Project.config file only contains plain text (no rich text file format)
It has to be prepared using a text editor such as vi, gedit, nedit, geany or the notepad

Please note the uppercase of the first letter of the filename

The presence of the 'Project.config' file in the working directory is not mandatory

Keywords provided in the 'Project.config' file are related to the molecules involved in the PyRED job

Adding the 'Project.config' file in the working directory allows modifying the default information/keyword (molecule total charge, spin multiplicity, title, etc...) about the molecules involved in the PyRED job

In the absence of the 'Project.config' file PyRED can still be executed
However, in this case only default information about the molecules involved in the PyRED job are used
See <https://upjv.q4md-forcefieldtools.org/REDServer-Development/Documentation/Project.config>

Combining the 'REMARK PROJECTCFG' keywords and keywords, which are defined in the Project.config file, in the Mol_red'n'.pdb input files, allows bypassing the use of a Project.config file in the archive file. Thus, if a single PDB input file is needed for a PyRED job, a user can directly submit this PDB input file as an archive file without the need of the Project.config/System.config files.

Example:

```
REMARK
REMARK SYSTEMCFG OPT_Calc           = ON
REMARK SYSTEMCFG MEPCHR_Calc        = ON
REMARK SYSTEMCFG Freq_Calc          = ON
REMARK SYSTEMCFG MOD_GAUSSIAN_JOB   = Complex
REMARK SYSTEMCFG CHR_TYP            = RESP-X1
REMARK SYSTEMCFG METHOD_OPTCALC      = B3LYP
REMARK SYSTEMCFG BASSET_OPTCALC     = 6-31+G(d,p)
REMARK SYSTEMCFG METHOD_MEPCALC      = B3LYP
REMARK SYSTEMCFG BASSET_MEPCALC     = cc-pVTZ SCRF(IEFPCM,Solvent=Ether)
REMARK
REMARK PROJECTCFG MOLECULE1-TITLE   = [Heme(4-)-(Histidine)2_Fe(III)_complex]1-
REMARK PROJECTCFG MOLECULE1-TOTCHARGE = -1
REMARK PROJECTCFG MOLECULE1-SPINMULT = 6
REMARK PROJECTCFG MOLECULE1-ATMREORDR = OFF
REMARK PROJECTCFG MOLECULE1-INTRA-MCC1 = 0.0 | 1 2 3 4 5 6 | Remove
REMARK PROJECTCFG MOLECULE1-INTRA-MCC1 = 0.0 | 28 29 30 31 32 33 | Remove
REMARK PROJECTCFG FE-RAD4TOP        = 2.0
[ atom indexes and names, residue name(s) and number(s), X Y Z Cart. coordinates, and chemical
elements to the PDB file format ]
```

V. THE frcmod.user FILE

The frcmod.user file only contains plain text (no rich text file format)
It has to be prepared using a text editor such as vi, gedit, nedit, geany or the notepad

Please note the absence of uppercase for the filename letters

The frcmod file format belongs to AMBER
See the AMBER documentation: <https://ambermd.org/FileFormats.php>

The presence of the 'frcmod.user' file in the working directory is not mandatory

The user can provide missing or mandatory force field parameters (i.e. these reported in the 'frcmod.unknown' file generated in a previous job) within the 'frcmod.user' file