How to efficiently use R.E.D. Server Development/PyRED?

PyRED efficient use – Efficient PyRED workflow

Save CPU time by splitting a complex job into smaller ones, and progressing step-by-step to fix issues.

Questions?
Contact_at_q4md-forcefieldtools.org

Step 1: get the QM optimized geometries for each conformation/molecule

Use: `OPT_Calc = ON, Freq_Calc = ON, MEPCHR_Calc = OFF`

Step 2: get the QM MEP, atom types and charges for the multiple molecule job

Use: `OPT_Calc = OFF, Freq_Calc = OFF, MEPCHR_Calc = ON`

Step 3: check and correct the multiple molecule job

Use: `Re_Fit = ON` (other keywords do not matter)

Step 4: provide unknown FF parameters

Use: `Re_Fit = ON` (other keywords do not matter)

A 2 molecule job involving intra/inter-molecular charge constraints is demonstrated in this tutorial:
• Molecule 1 contains 2 conformations
• Molecule 2 contains a single conformation

Abbreviations : FF = force field; QM = quantum mechanics; MEP = molecular electrostatic potential.

Step 1: get the QM optimized geometries for each conformation/molecule

• QM geometry optimization is a bottleneck in FF generation, as it requires long computation times.
• Optimized geometries can also be selected after a conformational search…
• **Prepare PDB input files with chemical elements**: as PyRED handles all the elements of the periodic table, the chemical element columns (right justified columns 77-78; 1 or 2 characters) are used to differentiate atoms with ambiguous atom names.
See the '`readme.txt`' file:
https://upjv.q4md-forcefieldtools.org/REDServer-Development/Documentation/readme.txt

Prepare 3 jobs and run 3 single conformation single molecule jobs

• Use `OPT_Calc = ON`, `Freq_Calc = ON`, `MEPCHR_Calc = OFF` in the '`System.config`' file for each job (use `OPT_Calc = ON2` if the molecule is large).
• Define an informative title for each molecule, and specify the total charge value and spin multiplicity values (if different from 0 and 1, respectively) in the '`Project.config`' file.
• Use `MOLECULE1-ATMREORDR = ON` in the '`Project.config`' file to get a friendly atom order for atom indexing and RESP input files for each job.

Create the 3 archive files

```
# Conformation 1 molecule 1
tar -jcf job1-1.tbz job1-1/Mol_red1.pdb job1-1/*.config
# Conformation 2 molecule 1
tar -jcf job1-2.tbz job1-2/Mol_red1.pdb job1-2/*.config
# Conformation 1 molecule 2
tar -jcf job1-3.tbz job1-3/Mol_red1.pdb job1-3/*.config
```

Remark: Each '`System.config`' and '`Project.config`' keyword can also be added to the PDB input file allowing the direct upload of this file to R.E.D.Server Development using the '`REMARK SYSTEMCFG`' and '`REMARK PROJECTCFG`' keywords, respectively (see the '`readme.txt`' file).

Upload the 3 archive files separately

• Run the '`Px1`', '`Px2`' and '`Px3`' jobs, download the corresponding data on a local machine and study the results.
• Redo the jobs until all the geometry optimizations are successful and the generated conformations are those needed.

Results

The atoms of each QM optimized conformation/molecule are reordered, duplicated atom names are corrected, the total charge and spin multiplicity are validated, and the conformation(s) with imaginary frequency(ies) can be identified and/or discarded.

Step 2: get the QM MEP, atom types and charges for the multiple molecule job

Prepare the multiple molecule job

The job involves the 2 conformations of molecule 1 and the single conformation of molecule 2:

```
mkdir job1-2mols
cp  Px1/Data-R.E.D.Server/Mol_m1/Mol_m1-c1-qmra.pdb(*)  job1-2mols/Mol_red1.pdb
cp  Px1/Data-R.E.D.Server/Mol_m1/JOB1-gau_m1-1.out(*,#)  job1-2mols/Mol_red1-1.log

cat Px2/Data-R.E.D.Server/Mol_m1/Mol_m1-c1-qmra.pdb(*)  >> job1-2mols/Mol_red1.pdb
cp  Px2/Data-R.E.D.Server/Mol_m1/JOB1-gau_m1-1.out(*,#)  job1-2mols/Mol_red1-2.log
```

Warning: the atom orders in the 2 conformations of molecule 1 have to be identical in the PDB and QM files...

```
cp  Px3/Data-R.E.D.Server/Mol_m1/Mol_m1-c1-qmra.pdb(*)  job1-2mols/Mol_red2.pdb
cp  Px3/Data-R.E.D.Server/Mol_m1/JOB1-gau_m1-1.out(*,#)  job1-2mols/Mol_red2-1.log
```

(*) Atoms are reordered in PyRED outputs.
(#) Use '`JOB1-gau_m1-1.out`' for Gaussian, and '`JOB1-gam_m1-1.log`' for GAMESS/FIREFLY.

In the '`job1-2mols/System.config`' file

• Redefine '`OPT_Calc = OFF`', '`Freq_Calc = OFF`', '`MEPCHR_Calc = ON`'.

In the '`job1-2mols/Project.config`' file

• Redefine each molecule title, total charge and spin multiplicity values, incrementing the keywords as a function of the molecule indexes
• Add the intra/inter-molecular charge equivalencing (atom indexes are easily identified with the new atom order using the available JSmol applets with '`Atom labels = Index`')

Create the new archive file

```
tar -jcf job1-2mols.tbz job1-2mols/*
```

Upload the archive file

• Run the '`Px4`' job, download the corresponding data on a local machine and study the results.

Results

QM data (optimized geometries and MEP) are generated in a multiple molecule job.

Step 3: check and correct the multiple molecule job

Prepare the data to correct the 'Px4' job

```
mkdir job2-2mols
cd px4
cp Mol_red*                                          ../job2-2mols
cp --parents Data-R.E.D.Server/Mol_m*/JOB*       ../job2-2mols
rm Data-R.E.D.Server/Mol_m*/JOB*.inp       (GAMESS/Firefly input files), or
rm Data-R.E.D.Server/Mol_m*/JOB*.gjf       (Gaussian input files)
cp *.config                                          ../job2-2mols
```

Is charge equivalencing correctly carried out in the multiple molecule fit?

Indeed intra/inter-mcc may affect charge equivalencing in some cases
• Look at the force field libraries in the 'Px4/Data-R.E.D.Server/Mol_MM' directory
• If needed and to fine tune the multiple molecule charge fit call charge equivalencing rules used the single molecule fits (keywords MOLECULE1/2-CHEMEQSM) defined in the file:
Px4/Data-R.E.D.Server/Data-Default-Proj/Project-out.config
and force their use in the multiple molecule fit:

```
cd Data-R.E.D.Server/Data-Default-Proj/
grep CHEMEQSM ./Project-out.config  >> ../../../job2-2mols/Project.config
sed -i '/CHEMEQSM/CHEMEQMM/'           ../../../job2-2mols/Project.config
```

Are the generated atom types correct and does a new atom type has to be created?

PyRED may show inaccuracies with the 'CD' and 'CM' atom types
• Look at the force field libraries in the 'Px4/Data-R.E.D.Server/Mol_MM' directory
• Start from the atom type list (keywords MOLECULE1/2-ATMTYPE) defined in the file:
Px4/Data-R.E.D.Server/Data-Default-Proj/Project-out.config

```
grep -- "-ATMTYPE" ./Project-out.config  >> ../../../job2-2mols/Project.config
```

• Then adapt the ATMTYPE list of molecules 1/2 to the choices made in the 'job2-2mols/Project.config' file

Use the Re_Fit mode to avoid re-running QM MEP computation

Add 'Re_Fit = On' in the 'System.config' file

```
echo 'Re_Fit = On'  >> ../../../job2-2mols/System.config
```

Create a new archive file

It contains the updated 'System.config' and 'Project.config' files and the QM data from the previous job:

```
cd ../../..
tar -jcf job2-2mols.tbz job2-2mols/*
```

Need a better compression ratio for the archive file? *(maximum size for the uploaded archive file: 90 MB!)*

```
tar -cf job2-2mols.tar job2-2mols/*
7za a -mmt=on -mx=9 job2-2mols.tar.7z job2-2mols.tar
```

Upload the archive file

Run the 'Px5' job, download the corresponding data on a local machine and study the results.
('Px5' should last a few minutes on the cluster)

Results

The multiple molecule job contains corrected charge values and atom types.

Step 4: add unknown FF parameters

Is there any unknown/missing FF parameters in job 'P5x'?

• Look at the FF parameters in the '`Px5/Data-R.E.D.Server/Data-Default-Proj/`' directory.
• FF parameters are listed in 3 files:
'`frcmod.known`', '`frcmod.unknown`' and '`frcmod.correspondence`'.
• These files are based on the atom types defined in the previous step.
• Known FF parameters are listed in the '`frcmod.known`' and '`frcmod.correspondence`' files.
• Unknown FF parameters are listed in the '`frcmod.unknown`' file.

Prepare a last job if unknown FF parameters are found

```
mkdir job3-2mols
cd Px5
cp Mol_red*                                          ../job3-2mols
cp --parents Data-R.E.D.Server/Mol_m*/JOB*       ../job3-2mols
rm Data-R.E.D.Server/Mol_m*/JOB*.inp       (GAMESS/Firefly inputs), or
rm Data-R.E.D.Server/Mol_m*/JOB*.gjf       (Gaussian inputs)
cp *.config                                          ../job3-2mols
```

Create a 'frcmod.user' input file in the '`job3-2mols`' directory

```
cd ..
```

The '`frcmod.user`' input file must contain the unknown FF parameters listed in the '`frcmod.unknown`' file from the previous job; for more information see:
https://upjv.q4md-forcefieldtools.org/REDServer-Development/Documentation/ForceFieldParameters-LEaPScript.html
https://ambermd.org/FileFormats.php

Keep '`Re_Fit = On`' in the '`System.config`' file to avoid re-running QM MEP computation

Create a new archive file

```
tar -jcf job3-2mols.tbz job3-2mols/*
```

Need a better compression ratio for the archive file? *(maximum size for the uploaded archive file: 90 MB!)*

```
tar -cf job2-2mols.tar job2-2mols/*
7za a -mmt=on -mx=9 job2-2mols.tar.7z job2-2mols.tar
```

Upload the archive file

Run the '`Px6`' job, download the corresponding data on a local machine and study the results.
('`Px6`' should last a few minutes on the cluster)

Results

The multiple molecule job contain all the FF parameters based on the atom types defined in the previous step.

One is ready the edit/load the '`leaprc.q4mdfft`' script available in the '`Px6/Data-R.E.D.Server/Data-Default-Proj/`' directory in the LEaP program from AmberTools…