

R.E.D. version 2.0

User's Manual and Tutorial

Authors:

F.-Y. Dupradeau ^(1,3)

A. Pigache ⁽¹⁾

T. Zaffran ^(1,3)

P. Cieplak ⁽²⁾

⁽¹⁾DMAG EA 3901 & Faculté de Pharmacie, Amiens, France

⁽²⁾ The Burnham Institute, La Jolla, CA, USA

⁽³⁾ The Scripps Research Institute, La Jolla, CA, USA

Table of Contents

-I- WHY R.E.D. & X R.E.D.?	3
-II- WHAT IS NEEDED TO EXECUTE R.E.D. & X R.E.D.?	5
-III- HOW TO USE R.E.D. & X R.E.D.?	9
* General variables available in R.E.D.....	9
* Execution of R.E.D.....	11
* Execution of X R.E.D.....	11
* Execution of GAMESS & Gaussian in parallel within R.E.D.....	11
* Miscellaneous.....	13
-IV- INPUTS AUTOMATICALLY GENERATED BY R.E.D.	14
* Geometry optimization input for GAMESS.....	14
* Geometry optimization input for Gaussian.....	14
* MEP input for GAMESS & Gaussian.....	15
* Inputs for RESP.....	17
-V- THE 'INITIAL' PDB FILE	18
* PDB atom names.....	18
* Molecular orientation of the optimized geometry.....	19
* Atom connectivity.....	22
-VI- UPDATED TUTORIAL OF R.E.D. I FOR R.E.D. II	24
* Exemple of QMRA procedure.....	24
* Exemple of RBRA procedure.....	25
* Exemple of Multi-RBRA procedure.....	26
* Applications.....	28
-VII- NEW TUTORIAL FOR R.E.D. II	29
* Three conformation-two orientation RESP fit.....	29
-VIII- HOW TO REFERENCE R.E.D.?	32
-IX- REFERENCES	32

-I- WHY R.E.D. & X R.E.D.?

Derivation of RESP,^[1,2] and ESP^[3] charges for a new structure is an important step in molecular mechanics simulations using AMBER and other force fields. To get such atom-centered charges one proceeds in three steps:

First, the molecule of interest is optimized to determine a stable conformation [using a Quantum Mechanical (QM) software]. Then, the minimized structure is used to compute a Molecular Electrostatic Potential (MEP) on a three-dimensional grid (using again a QM software). Finally, the grid containing MEP values is exported into the "RESP" program, <http://amber.scripps.edu/Questions/resp.html>, which is used to fit the atomic charges to the MEP.

This protocol can also be applied to derive charges for several molecular conformations at once, and as such, it has been named as a multi-conformation RESP fit.^[1,2,4] This allows making atomic charges more 'general' and effective, and is useful in molecular dynamics simulations where the whole conformational space is going to be explored.

Although this method is routinely used nowadays, it still suffers from a number of limitation number:

- The whole procedure requires several steps involving different programs and various data format conversions between them. Consequently, the procedure is tedious, time-consuming, and numerous errors can be introduced without having a real way to check them. This is particularly true when the RESP fit is performed for large molecules.

- Although, in principle, any quantum programs could be used to optimize the initial structure and to compute the MEP, the "AMBER" community mainly uses the "Gaussian" program (<http://www.gaussian.com>),^[5a,5b] which is a quite expensive proprietary software. The "GAMESS" academic program (<http://www.msg.ameslab.gov/GAMESS/GAMESS.html>),^[6] which is provided at no cost, and has similar functionality for RESP and ESP charge derivation as "Gaussian", is not commonly used to develop RESP or ESP charges. Indeed, it is known that partial charges obtained using "GAMESS", are 'different' from those calculated using "Gaussian".

- Finally, starting from different sets of Cartesian coordinates for a given molecule, the RESP or ESP partial charges are, in some cases, not reproducible even using the "Gaussian" program. This makes potential errors in the protocol difficult to detect.

Thus, we developed the **R.E.D.** program (version 1.0, RESP ESP charge Derive, <http://www.u-picardie.fr/labo/lbpd/RED/>) to automatically derive RESP or ESP charges starting from an unoptimized PDB structure. **R.E.D. I** sequentially executes (i) either the "GAMESS" or the "Gaussian" program to optimize the molecular structure and to compute the corresponding MEP, and then (ii) the "RESP" program to fit the atom-centered charges to the grid determined in the previous step (see [Table 1](#) & [Figure 1](#)). Format conversions needed during the procedure, and the inputs for the "GAMESS", "Gaussian" and "RESP" programs are automatically generated.

The role of QM optimization thresholds on the charge values has been studied, and a new RESP fitting procedure based on single- or multi-reorientation(s) has been developed. This approach allows getting highly reproducible RESP or ESP charges that are independent of the QM software and the choice of the initial Cartesian coordinates. The charge reproducibility reaches the level of 0.0001 e. Although such an accuracy is not needed in molecular mechanics simulation, this allows defining atomic charges that can be considered as a reference set useful for reproducing published data or error checking. This set is obtained for a certain molecular orientation and a given QM theory level.

In **R.E.D. II** (version 2.0), we implemented the multi-conformation RESP and ESP fit that can be automatically carried out for a well-defined set of molecular conformations. Thus, 'multi-conformation' and 'multi-orientation' RESP fit can be performed together or independently according to the user choice. 'Standard' but also 'non-standard' RESP inputs can now be generated. The output from geometry optimization generated by one of the two QM programs can be used as input for MEP computation using the second QM program. Finally, RESP and ESP charges can be derived for chemical elements up to $Z = 35$ (Z is the total number of electrons).

X R.E.D. is a graphical interface to the **R.E.D.** program that can be used to modify **R.E.D.** variables.

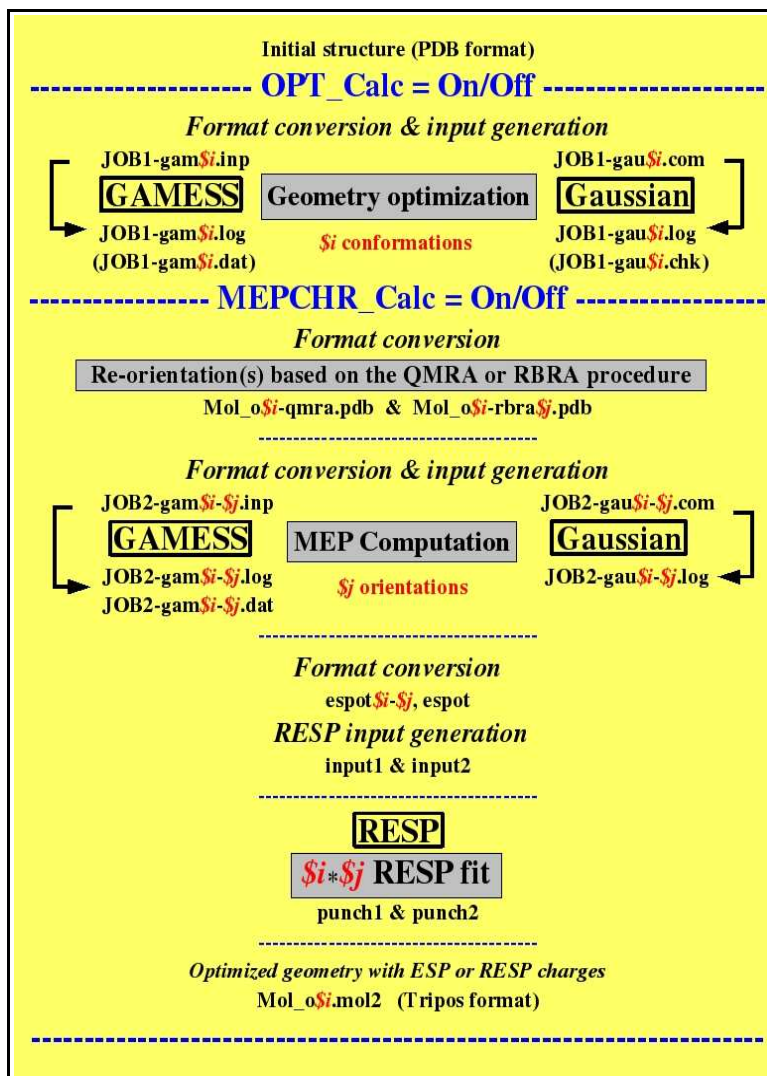


Figure 1: Execution of R.E.D. II

R.E.D. and **X R.E.D.** are available at no cost (but are copyrighted) for academic users on the Internet (see <http://www.u-picardie.fr/labo/lbpd/RED/>) after signing a license. On the contrary, a 1500 € fee is demanded to industrial/commercial users for the use of **R.E.D.** and **X R.E.D.** Thus, commercial users must contact us to get a license, and cannot use the license available on the Internet.

-II- WHAT IS NEEDED TO EXECUTE R.E.D. & X R.E.D.?

The **R.E.D. II** program (around 1550 line code) is written with the "perl" programming language (<http://www.perl.com>), which presents numerous advantages:

- It is an interpreted language meaning that the source code does not need to be compiled.
- It is well adapted to extract and format text files.
- It follows the "Open Source" philosophy, and is "freely" available on the Internet.
- It is available on the UNIX operating systems allowing the portability of **R.E.D.** on numerous machines (PC-LINUX, SGI-IRIX, HP-UX, IBM-AIX, SUN-SOLARIS etc...).
- "perl" functionality can also be easily increased using flexible modules.

R.E.D. II uses the "**FileHandle**", "**Math::Trig**" and "**warnings**" modules, which have to be compiled under the current operating system. However, these three modules were installed by default with "perl" on our systems (SGI-IRIX and PC-LINUX). If one of these modules is missing, it can be obtained from the Comprehensive PERL Archive Network Internet site (see <http://www.cpan.org>).

To check if these modules are installed on the system, use the following commands:

```
find / \( -name FileHandle.pm -o -name Trig.pm \) -print          and
find / -name warnings.pm -print
```

The **X R.E.D.** program is written with the "tcl/tk" programming language (also an interpreted language and "Open Source/Free" software, see <http://www.tcl.tk>) [8.3 (or newer) "tcl/tk" version has to be installed]. With **X R.E.D.**, the user executes **R.E.D.** in a graphical user-friendly interface and modifies **R.E.D.** variables using this interface (see [Figure 2](#)).

Below is the procedure that we applied to install "tcl/tk" (version 8.3) under SGI workstations (an older "tcl/tk" version is installed by default under IRIX 6.5):

- REQUIREMENTS: "perl" 5.x ; "tcl" 8.3.5 & "tk" 8.3.5 (or newer).
- HOW TO INSTALL "tcl/tk": Go to <http://www.tcl.tk> and download "tcl" and "tk" sources.

```
tar zxvf tcl8.3.5-src.tar.gz
tar zxvf tk8.3.5-src.tar.gz

su                                     password of root
mkdir /usr/local/tcltk
mv tcl8.3.5 tk8.3.5 /usr/local/tcltk
cd /usr/local/tcltk/tcl8.3.5/unix/
./configure --prefix=/usr/local/tcltk/tcl835
./make
./make install

cd /usr/local/tcltk/tk8.3.5/unix/
./configure --prefix=/usr/local/tcltk/tcl835
./make
./make install
```

Add "/usr/local/tcltk/tcl835/bin" to the "\$PATH" variable and use "wish" to execute "tcl/tk"

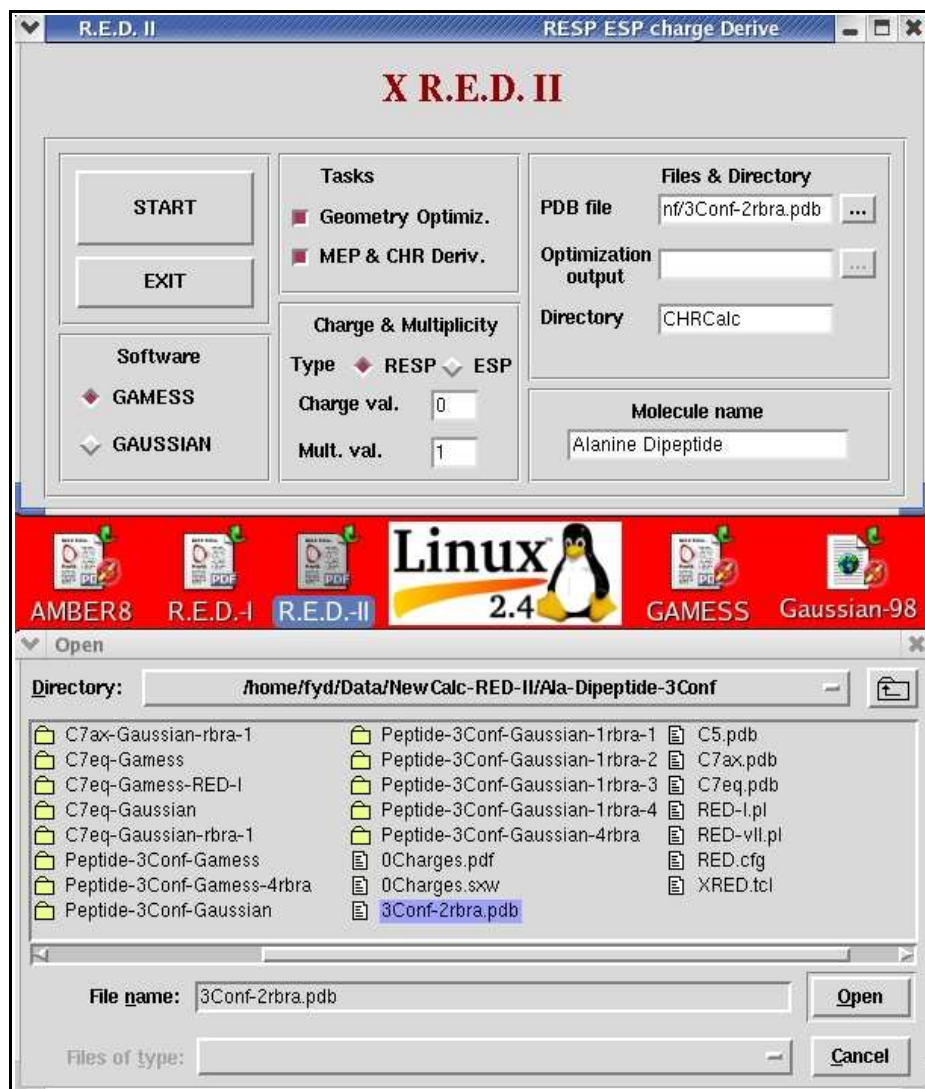


Figure 2: X R.E.D. graphical user-friendly interface

We tested the **R.E.D.** and **X R.E.D.** programs on SGI-IRIX and PC-LINUX ("Red-Hat 7.3", "Red-Hat 9.0" and "Debian 3.0" distributions) workstations ("perl" versions 5.4-5.6 and 5.8, and "tcl/tk" version 8.3.5), but both programs should work on all other UNIX operating systems. Although "perl" and "tcl/tk" are also available on Windows and Macintosh computers, we did not test **R.E.D.** and **X R.E.D.** on such operating systems. Indeed, the "GAMESS", "Gaussian" and "RESP" programs have to be installed on these operating systems separately (see [Table 1](#) below).

Under the UNIX operating systems, **R.E.D.** looks for the "GAMESS", "Gaussian" and "RESP" binaries and scripts using the "**which**" UNIX command. Thus, one should update the "**\$PATH**" environment variable (to install these software, see their corresponding documentations). One has also to test these programs in a 'standalone' mode before executing them with **R.E.D.** Below, is an example of ".cshrc" file (for the "csh" or "tcsh" shell) needed to execute the "GAMESS", "Gaussian" and "RESP" programs:

```
cat .cshrc

setenv GAUSS_SCRDIR /QM_Disk/QM_SCR
setenv g98root /usr/local
setenv SOFT /usr/local
source /usr/local/g98/bsd/g98.login

set path = ($path $SOFT/g98 $SOFT/gamess $SOFT/resp)
```

Derivation of RESP & ESP charges	Use	Type of software
UNIX OS Linux, IRIX, AIX, HP-UX, OSF1... Neither MacOS nor Win32 OS	Operating System	Proprietary or "Free"...
Gaussian ^[5a,5b]	QM software (geometry optimization & MEP computation)	Proprietary software Copyrighted
and/or		
GAMESS ^[6]	QM software (geometry optimization & MEP computation)	Academic software Copyrighted
RESP (part of AMBER) ^[7,8]	Fitting software for atom-centered charges	Academic software Copyrighted
R.E.D.	Automatic, reproducible & effective charge derivation	Academic program Copyrighted
"perl" (at least version 5.4) (For R.E.D. execution)	Interpreted programming language	Open Source software "Free" software
X R.E.D. (optional)	Graphical user interface for R.E.D.	Academic program Copyrighted
"tcl/tk" (at least version 8.3) (For X R.E.D. execution)	Interpreted programming language	Open Source software "Free" software

Table 1: The programs needed to derive RESP & ESP charges

Remarks regarding the use of "GAMESS" with R.E.D.

- The "GAMESS" execution script, named "**rungms**", has been modified so that "GAMESS" saves its MEP file(s) ("dat" or "punch" file) in its scratch directory ("**\$SCR**"). Thus, **R.E.D.** finds, and then converts this/these file(s) into the "RESP" format ["**espot**" file(s)]. See "**rungms**", around [lines 33](#) for the definition of the "GAMESS" scratch directory, and around [lines 73-80](#) for the modification that has to be carried out:

```
set SCR=/QM_Disk/QM_SCR          #"/QM_Disk/QM_SCR" is an example of scratch directory (~ line 33)
                                # (A lot of disk space has to remain available for temporary QM files...)
                                # This is also the scratch directory used by "Gaussian"
                                #
                                # (See the ".cshrc" file provided just above)

# file assignments.             # ~ lines 73-80
# all binary files should be put on a node's local disk ($SCR directory),
# both ASCII punch files might be written by NFS to user's permanent disk
set echo
setenv EXTBAS /dev/null
setenv IRCDATA $SCR/$JOB.irc
setenv INPUT $SCR/$JOB.F05
setenv PUNCH $SCR/$JOB.dat      # important modification !!!
setenv AOINTS $SCR/$JOB.F08
```

- On some UNIX machines (SGI ORIGIN, others ?), the "**ddikick.x**" program seems to not be used in the execution of "GAMESS". As **R.E.D.** automatically checks if it is installed, this will lead to the following error message:

```

                * Software checking *
rungms                [ OK ]
gameSS.01.x          [ OK ]
ddikick.x            [ NOT FOUND ]
resp                 [ OK ]

ERROR: Program(s) can NOT be executed

Press Enter to exit.
```

In this case, lines 219, 221-223 in the **R.E.D.** source code have to be commented:

```

#219#  chomp($ddikick=`which ddikick.x`);          #--- ddikick.x ---
open(STDOUT,">&OLDSTDOUT");
#221#  if(($ddikick =~ m/Command not found/ig)||($ddikick =~ m/not
in/ig)||($ddikick =~ m/no ddikick.x in/ig)||($ddikick eq "")){
#222#  print"\t  ddikick.x \t\t\t\t\t\t[ NOT FOUND ]\n"; $verif = 0;
#223#  }else{  print"\t  ddikick.x \t\t\t\t\t\t[ OK ]\n";    }
```


-III- HOW TO USE R.E.D. & X R.E.D.?

* *General variables available in R.E.D.*

The "GAMESS" program (<http://www.msg.ameslab.gov/GAMESS/GAMESS.html>)⁶ or the "Gaussian" program (versions 2003, 1998 or 1994) (<http://www.gaussian.com>)^{5a,5b} has to be installed. The user has to set the **R.E.D.** variable "\$QM_SOFT" = "GAMESS" or = "GAUSSIAN" (See line 1510 in the **R.E.D. II MAIN PROGRAM** section) according to the QM software installed on the system. For any version of the "GAMESS" program, the same binary names *i. e.* "rungms", "ddikick.x" and "gamess.\$i.x" ("i" is a number between "01" and "09"; "gamess.\$i.x" is recognized and executed by **R.E.D.**) are used. In the case of running various versions of the "Gaussian" program, different binary names are used ("g94", "g98" and "g03" for "Gaussian" 1994, 1998 and 2003, respectively). Thus, **R.E.D.** automatically checks which binary is installed, and runs the one that is found (Actually, the latest "Gaussian" version detected is executed by **R.E.D.**). We only tested **R.E.D.** with "GAMESS" and "Gaussian" 1998 and 2003. However, "Gaussian" 1994 can also be executed by **R.E.D.**

The QM software must be tested before using it with **R.E.D.**

The "RESP" program (<http://amber.scripps.edu/Questions/resp.html>) must also be installed on the operating system. It can be downloaded at no cost from the official AMBER Internet site (<http://amber.scripps.edu>).^[7,8] **R.E.D.** gives the opportunity to calculate RESP charges (used in Cornell *et al.* force field),^[9,10,11,12] and ESP charges (used in the 'old' Weiner *et al.* force field).^[13,14] The "\$CHR_TYP" variable has to be set = "RESP" or = "ESP" (See line 1515 in the **R.E.D. II MAIN PROGRAM** section) according to the user choice.

The "RESP" program should be tested before using it with **R.E.D.**

Two types of input can be provided within **R.E.D.**:

-1- The user ONLY provides an initial (unoptimized) PDB file (read the section -V- of this manual related to this file) and sets the **R.E.D.** variables "\$OPT_Calc" = "ON" and "\$MEPCHR_Calc" = "ON" (See lines 1512 and 1513 in the **R.E.D. II MAIN PROGRAM** section). In this case, **R.E.D.** first executes either the "GAMESS" program or the "Gaussian" program to optimize the initial PDB structure and to compute the MEP, and then the "RESP" program to fit the atom-centered charges to the MEP.

With **R.E.D. II**, the atomic charge derivation using a defined number of conformations (in principle, it could be an infinite number of conformations) has been developed (see Figure 1). This means that each conformation has to be optimized before computing its MEP and deriving the atomic charges for the whole molecular system. In this case, the initial (unoptimized) PDB file must contain the Cartesian coordinates of each conformation (read the section -V- of this manual related to this file). For each of them, a minimization input is automatically generated and a geometry optimization is carried out. Once the \$i conformations have been minimized, the \$i corresponding outputs (*i. e.* "JOB1-gam\$i.log" for "GAMESS" and "JOB1-gau\$i.log" for "Gaussian") are checked whether the calculations are finished correctly. Finally, these \$i files are concatenated into a single file (*i. e.* "JOB1-gam.log" for "GAMESS" and "JOB1-gau.log" for "Gaussian"). The latter file is used in the MEP computation and the derivation of RESP or ESP charges.

-2- The user may have already optimized the initial structure. For example, the geometry optimization may be performed on another computer, while the MEP computation and RESP fit may be executed on the user's workstation. In this case, the minimization output from the "GAMESS" or "Gaussian" program (modifying the "\$JOB_OPT" **R.E.D.** variable; see [line 1522](#) in the **R.E.D. II** MAIN PROGRAM section) can be provided as input by setting the **R.E.D.** variables "\$OPT_Calc" = "OFF", leaving "\$MEPCHR_Calc" = "ON". **R.E.D.** automatically extracts the minimized Cartesian coordinates from the corresponding output (*i. e.* the optimized Cartesian coordinates found after the words "EQUILIBRIUM GEOMETRY LOCATED" in the "GAMESS" output, or the geometry called "Standard orientation" found after the words "Stationary point found" in the "Gaussian" output are extracted).

As mentioned above, **R.E.D. II** can handle several conformations in the process of RESP or ESP charge derivation. However, the more conformations considered, the higher is the chance that geometry optimization will not converge for some of them. This leads to the "[FAILED]" **R.E.D.** execution. Thus, if the studied molecular system presents several important conformations, the best might be to run the optimization jobs (see the section -IV- of this manual, *i. e.* the two subsections named: "GAMESS" and "Gaussian" inputs for geometry optimization, describing the building of the corresponding QM inputs) for each conformation in a 'standalone mode' (*i. e.* without **R.E.D.**). Once all the minima have been located, the minimization outputs have to be concatenated into a single file, which can be loaded into **R.E.D.** using the two variables "\$OPT_Calc" = "OFF", "\$MEPCHR_Calc" = "ON", providing the full path to locate this file using the "\$JOB_OPT" variable.

Remark:

The atom order within the initial PDB file and in the minimization output(s) must be the same and more generally, the atom order in each conformation must be the same.

=> Read the section -V- of this manual related to the initial PDB file.

In **R.E.D. II.**, it is now also possible to use the geometry optimization output [containing \$i conformation(s)] generated by one of the two QM programs as input in **R.E.D.** for MEP computation using the other QM program. This is also achieved using the **R.E.D.** variables "\$OPT_Calc" = "OFF" and "\$MEPCHR_Calc" = "ON", providing once again the path to locate the concatenated minimization outputs using the "\$JOB_OPT" variable. Actually, **R.E.D. II** automatically detects the format of the optimization output (based on the QM software author names, warns the user which type of minimization output is found (*i. e.* "Gaussian" or "GAMESS")) and if the file is recognized [*i. e.* in agreement with the initial PDB file (same atom order), minimum found on the potential energy surface and normal termination of the geometry optimization job] or considered corrupted.

See the "Tutorials" for more information.

* Execution of R.E.D.

If **X R.E.D.** is not executed, the "**XRED**" variable in **R.E.D.** must remain set to "**OFF**" in order to execute **R.E.D.** ("**XRED**" = "**OFF**" by default, see [line 1508](#) in the **R.E.D. II MAIN PROGRAM** section). After having selected the correct variables for the **R.E.D.** execution, one of the following commands can be used to launch **R.E.D.**:

```
ln -s RED-vII.pl RED.pl          (only for R.E.D. II)
perl RED.pl                      or
perl RED.pl > RED.log           (To save the RED ".log" file; see Figure 3)

nohup perl RED.pl > RED.log &   (Typical use: distant workstation/computer cluster)
```

* Execution of X R.E.D.

As **X R.E.D.** executes **R.E.D.**, both programs have to be present in the working directory. If **X R.E.D.** is executed, the "**XRED**" variable in **R.E.D.** must be set to "**ON**" in the **R.E.D.** source code (see [line 1508](#) in the **R.E.D. II MAIN PROGRAM** section) otherwise the variables selected in the **X R.E.D.** graphical interface will not be taken into account!

```
wish XRED.tcl                    (Typical use: local workstation; see Figure 2)
wish8.3 XRED.tcl
("wish8.3" has to be used instead of "wish" if two versions of tcl/tk binary are installed on the machine)
```

If **X R.E.D.** is executed, the appropriate variables can be selected from the **X R.E.D.** graphic interface. In this case, the variables available in **R.E.D.** (See [lines 1509-1522](#) in the **R.E.D. II MAIN PROGRAM** section) are no longer used. When **X R.E.D.** is executed, a "**RED.cfg**" file is generated with the variables needed by **R.E.D.** for its execution.

R.E.D. I. is needed to run **X R.E.D. I.** while **X R.E.D. II** executes **R.E.D. II.** (see [line 668](#) in **X R.E.D. II**).

* Execution of "GAMESS" & "Gaussian" in parallel within R.E.D.

The "GAMESS" and "Gaussian" programs can also be launched 'in parallel mode' using **R.E.D. II** [*i. e.* using several processors on the same motherboard (Shared Memory Based or SMP systems) or using several 'nodes' having a single processor]. Indeed, "GAMESS" uses 'replicated' and 'distributed' memory while "Gaussian" uses 'virtual shared' memory (*i. e.* using Linda, <http://www.lindaspace.com/about/index.html>) or 'shared' memory (*i. e.* using the 'shmem' UNIX library). In the **R.E.D.** source code, see [lines 526 & 968](#) and [lines 552 & 1021](#) for the definition of the "perl" commands, which execute "GAMESS" and "Gaussian", respectively. Thus, one has only to modify the "**\$NP**" variable ("**\$NP**" being the number of 'SMP' processors) to execute "GAMESS" and "Gaussian" using several processors. Moreover, the optimization and MEP inputs in **R.E.D.** may also have to be slightly updated [by modifying the "**MWORDS**" and "**MEMDDI**" values in "GAMESS" inputs, by changing the number of processors "**N**" ("**%NProc**" = "**N**" or "**%NProcLinda**" = "**N**") and by increasing the "**%Mem**" value in "Gaussian" inputs]. See the respective "GAMESS" and "Gaussian" documentations for their execution in parallel.

See the "*Tutorials*" for more information.

```

-----
*   Welcome to R.E.D II   *
  RESP ESP charge Derive

MOLECULE = Dimethylphosphate
CHARGE   = RESP
-----

* Selected QM Software *
  GAMESS

* 3 Conformation(s) selected *

* Selected three atom based re-orientation(s) *
   5     6     7
   7     6     5

* Software checking *
rungms                [ OK ]
gamess.00.x           [ OK ]
ddikick.x             [ OK ]
resp                  [ OK ]

The structure/conformation(s) is/are being optimized ...

Conformation 1 ... [ OK ]
See the file "JOB1-gam1.log"

Conformation 2 ... [ OK ]
See the file "JOB1-gam2.log"

Conformation 3 ... [ OK ]
See the file "JOB1-gam3.log"

The MEP is/are being computed ...

Conformation 1 ... [ OK ]
See the file(s) "JOB2-gam1-(X).log"

Conformation 2 ... [ OK ]
See the file(s) "JOB2-gam2-(X).log"

Conformation 3 ... [ OK ]
See the file(s) "JOB2-gam3-(X).log"

The RESP charges are being derived ... [ OK ]
See the "punch2" file

The following Tripos Mol2 file(s) has/have been created.
Mol_o1.mol2
Mol_o2.mol2
Mol_o3.mol2

Execution time: 4 h 48 m 14 s

*****
R.E.D. I was developed by:
A. Pigache (1), P. Cieplak (2) & F.-Y. Dupradeau (1)

R.E.D. II was developed in Prof. D.A. Case's laboratory by:
T. Zaffran (1,3), P. Cieplak (2) & F.-Y. Dupradeau (1,3)

(1) DMAG EA 3901 & Faculte de Pharmacie, Amiens, France
(2) The Burnham Institute, La Jolla, CA, USA
(3) The Scripps Research Institute, La Jolla, CA, USA
*****

```

Figure 3: Typical ".log" file generated by R.E.D.

*** Miscellaneous**

The total (or formal) charge of the molecule ("**\$CHR_VAL**" variable) has to be set to the correct value ("**\$CHR_VAL**" = "0" by default). In **R.E.D. II**, a new variable has been created allowing to handle molecular spin multiplicity higher than "1" ("**\$MLT_VAL**" = "1" by default). It is particularly useful in RESP or ESP charge derivation for metal complexes.

A title for the studied molecular system can also be provided using the "**\$TITLE**" variable ("**\$TITLE**" = "**MOL**" by default), which is used in the "GAMESS", "Gaussian" and "RESP" inputs each time it is allowed.

Once the charge derivation procedure is completed, the files generated can be stored in a directory defined by the "**\$DIR**" variable ("**\$DIR**" = "**CHRCalc**" by default). This directory is incremented automatically if a previous one with the same name already exists. If this variable is kept empty (*i. e.* "**\$DIR**" = "" or = " "), the data stay in the working directory. If a given calculation does not correctly finish, **R.E.D.** detects the problem, the word "**[FAILED]**" is printed, and the charge derivation process is stopped. In this case, the generated data remain also in the working directory.

See the "*Tutorials*" for more information.

-IV- INPUTS AUTOMATICALLY GENERATED BY R.E.D.

The **R.E.D.** program automatically generates all the required input files when it sequentially executes the "GAMESS" or "Gaussian" program and the "RESP" program. In most of the cases, the default keywords set up for those inputs by **R.E.D.** are correct. However, the user has still the opportunity to modify some inputs generated by **R.E.D.** (See the "GAMESS", "Gaussian" and "RESP" documentations in this case).

The minimization accuracy is important for deriving reproducible RESP and ESP charges that are independent of the QM software and initial (unoptimized) structure. See information below regarding optimization thresholds selected in the "GAMESS" and "Gaussian" programs. The following keywords have been chosen based on a large study of 17 structures and more than 100 orientations (See the section related to the initial PDB file for the importance of the molecular orientation in the process of charge derivation).

* *Geometry optimization input for "GAMESS"*

To modify this input, see [lines 504-521](#) in **R.E.D.** Below is a description of some specific keywords used in **R.E.D.**, which can be modified:

- For geometry optimization, **R.E.D.** uses the "6-31G*" basis set by default.
- "**QMTTOL**" = "1.0E-08", "**ITOL**" = "30", "**ICUT**" = "20", "**INTTYP**" = "**HONDO**" (just to increase the optimization accuracy, see the "GAMESS" documentation for more information).
- A SCF convergence criterion of 10^{-8} ("**CONV**" = "1.0E-08") is used (as in "Gaussian").
- The optimization threshold criterion "**OPTTOL**" = "1.0E-06" is used to set up the optimization accuracy. *Be aware that decreasing this value might slightly change charge values in some cases (Use even "**OPTTOL**" = "1.0E-07", if you can afford it!)*

* *Geometry optimization input for "Gaussian"*

To modify this input, see [lines 542-547](#) in **R.E.D.** Here are some particular keywords used in **R.E.D.**, which can be modified:

- For geometry optimization, **R.E.D.** uses the "6-31G*" basis set by default.
- A SCF convergence criterion of 10^{-8} "**SCF(Conver=8)**" is used (default in Gaussian).
- The optimization threshold criterion "**Tight**" is used to set up the optimization accuracy. *Be aware that decreasing this value might slightly change charge values in some cases (Use even "**VTight**", if you can afford it!)*

Remark:

Getting optimized structures with these thresholds might represent a real challenge in some cases. For small molecules such as DMSO or MeOH for instance, this should be straightforward with "GAMESS" or "Gaussian" inputs generated by **R.E.D.** However, achieving such convergence criteria for larger structures might be more difficult, particularly if several conformations are used in the atomic charge derivation. In this case, the best choice would be to run the geometry optimization(s) using the QM software without **R.E.D.** (*i. e.* in a 'standalone' mode), and modifying the optimization input(s), manually if the convergence fails. Once the minimization output(s) are obtained (with the optimized Cartesian coordinate set at the demanded threshold) and concatenated into a single file, the latter can be used as input in **R.E.D.** specifying the following **R.E.D.** variables: "**\$OPT_Calc**" = "OFF" and "**\$MEPCHR_Calc**" = "ON" (See the tutorials for more information).

Applying the "Tight" (or even "VTight") minimization convergence criterion in "Gaussian" may lead to a failed convergence in some cases. In this case, it is advised to use the "SCF (Conver=8)" keyword [decreasing the SCF threshold to "SCF(Conver=6)" makes the optimization even more difficult], and modify the inputs if the optimization fails [*i. e.* adding the calculation of the Hessians, and selecting the 'Standard orientation' (available in the Gaussian output) having the smallest optimization thresholds] as follows:

```
#P HF/6-31G* OPT=(Tight,CalcFC) SCF(Conver=8) Test or
#P HF/6-31G* OPT=(Tight,CalcAll) SCF(Conver=8) Test
```

In the case of geometry optimization using "GAMESS", we have not encountered any problem with getting "OPTTOL" = "1.0E-06" optimized structures for any molecule size. One has to use a huge number of optimization steps in "GAMESS", and to be patient. However, it might be better to calculate the Hessians every ten steps (for instance) using the following keywords:

```
$STATPT      HESS=CALC IHREP=10          $END
$FORCE       METHOD=ANALYTIC VIBANL=.F.  $END
```

Remark:

A reason for this difference between the two QM software might be, that the "Gaussian" program checks four thresholds (maximum and RMS forces, maximum and RMS displacements) while "GAMESS" checks only two of them (maximum and RMS gradients).

The optimization input names automatically generated by **R.E.D. II** for "GAMESS" and "Gaussian" are "JOB1-gam\$i.inp" and "JOB1-gau\$i.com" ("*i*" is the number of molecular conformations; "gam" and "gau" are used to differentiate the "GAMESS" and "Gaussian" files, respectively). The corresponding outputs differ by their names ("JOB1-gam\$i.log", "JOB1-gam\$i.dat" for "GAMESS", and "JOB1-gau\$i.log", "JOB1-gau\$i.chk" for "Gaussian"). When, "*i*" optimization outputs are produced, the "JOB1-gam\$i.log" or "JOB1-gau\$i.log" files are concatenated into a single file named "JOB1-gam.log" or "JOB1-gau.log". The latter is the file, which is used in the molecular reorientation process and MEP input generation.

*** MEP input for "GAMESS" & "Gaussian"**

The Cornell *et al.* force field uses RESP charges,^[9,10,11,12] which are derived based on MEP computed using the "6-31G*" basis set. The latter is important to generate implicit polarization, and results in dipolar moments, which are 20 % higher than those calculated in the gas phase. On the contrary, to derive ESP charges for the Weiner *et al.* force field,^[13,14] MEP have to be computed using the "STO-3G" basis set. See lines 940-958 & 995-1017 in the **R.E.D.** source code. **Thus, these lines should not be modified for classical RESP or ESP charge derivation!**

The name of the MEP inputs automatically generated by **R.E.D. II** for "GAMESS" and "Gaussian" are "JOB2-gam\$i-\$j.inp" and "JOB2-gau\$i-\$j.com" ("*i*" and "*j*" are the number of molecular conformations and orientations, respectively). The corresponding output files differ by their names ("JOB2-gam\$i-\$j.log", "JOB2-gam\$i-\$j.dat" for "GAMESS" and "JOB2-gau\$i-\$j.log" for "Gaussian"). The "JOB2-gam\$i-\$j.dat" or "JOB2-gau\$i-\$j.log" file(s) is/are automatically converted into *i* * *j* "espot*i-\$j*" file(s) [MEP file(s) in the "RESP" format] by a **R.E.D.** routine. This/these "espot*i-\$j*" file(s) is/are concatenated into a single "espot" file, which is the file called by the RESP fitting procedure.

Remarks about the geometry optimization and MEP inputs generated by R.E.D. for the automatic execution of the "Gaussian" and "GAMESS" programs:

The keywords used in these inputs have been selected to yield reproducible RESP or ESP charges independently of the QM software choice. This allows the utilization of the optimization outputs generated by one program for the MEP computation using the other one, as described above. However, there are several cases where it is simply impossible to get identical RESP or ESP charges using the two programs without modifying the basis set used or the source code of one of them:

- Indeed, the STO-3G basis set used for the second and third row chemical elements of the periodic table, the 6-31G basis set for the silicon atom and the polarization exponent used in the 6-31G* basis set for the K, Ca, Ga-Br atoms display some differences in the "GAMESS" and "Gaussian" programs. Calculations performed by those two programs for the systems involving above cases yield structural and MEP differences, and consequently different atomic charges. The "GFinput" and "GFprint" keywords have been added in the "Gaussian" inputs to display the basis set used, and to underline such differences. Another example is the utilization of the B3LYP method in geometry optimization. Different mathematical equations for this method (based on the VWN5 or VWN3 correlation functional) are implemented in each program. If this method is used to optimize geometries (*i. e.* the R.E.D. code has to be modified), it leads to slightly different geometries and RESP or ESP charges.

- The use of Cartesian or spherical harmonic functions for the basis set functions is another example of differences between the algorithms of the "GAMESS" and "Gaussian" programs. Particularly, such differences are observed for the third row elements of the periodic table. With R.E.D. II, the derivation of RESP or ESP charges for chemical elements having a total number of electrons up to $Z = 35$ is now possible. Thus, to allow the derivation of highly reproducible RESP charges for complexes containing metal atoms (Sc-Zn), spherical harmonic functions removing contaminant Cartesian functions are used within the two QM programs. Thus, the "ISPHER = 1" (for "GAMESS") and "5D" (for "Gaussian") keywords are used in the geometry optimization and MEP inputs if a third row element is detected in the studied molecule. In this case, the default algorithm available in these QM programs is not used (See lines 510, 511, 945, 946 & 545, 546, 999, 1000 in the R.E.D. source code to modify this feature).

- The atom radii used in the MEP computation is another source of difference between the "GAMESS" and "Gaussian" programs. The atom radii for the two first row elements of the periodic table are identical in the two QM software, and both QM programs yields identical RESP and ESP charges. The atom radii used by "GAMESS" for the third row chemical elements are automatically set to "1.8". On the contrary, the atom radii for this third row elements are unknown for "Gaussian", and must be provided by the user ("ReadRadii" keyword), except for the bromine atom, which has an atom radius equals to "2.3". To get highly reproducible RESP charges for metal complexes independently of the QM software, the strategy adopted by "GAMESS" has been applied in the "Gaussian" MEP input, *i. e.* atom radii equal to "1.8" is automatically used for the third row elements (See lines 1009-1017 in the R.E.D. source code to modify this feature).

To get more information about the usage of those QM programs, the users need to read the "GAMESS" (<http://www.msg.ameslab.gov/GAMESS/doc.menu.html>) and "Gaussian" (http://www.gaussian.com/g_ur/g03mantop.htm) documentations.

* Inputs for "RESP"

Inputs for the "RESP" program used in RESP or ESP charge derivation are automatically generated by **R.E.D.** based on the PDB atom names (read the section -V- of this manual related to the initial PDB file). In RESP charge derivation, a two stage fit is necessary, and the atomic charge equivalencing for chemically equivalent atoms is performed during the fit.^[1,2] Consequently, two different inputs are generated by **R.E.D.** (the "input1" and "input2" files), and the final RESP charges are available in the "punch2" file, which is generated by the "RESP" program. On the contrary, since a unique stage is required in ESP charge derivation,^[3] a single input is created by **R.E.D.** (the "input1" file), and the atomic charge equivalencing for chemically equivalent atoms is performed *a posteriori* to the fit. The ESP charges obtained from the fit (*i. e.* taken from the "punch1" "RESP" output) are made equivalent with a "perl" routine, and are saved in the "punch2" file (by analogy to the RESP charge derivation).

Remarks:

- If an initial PDB file with connectivity information is read by **R.E.D.**, then the final RESP or ESP charges available in the "punch2" file are also saved in the Tripos "mol2" file (read the section -V- of this manual related to the initial PDB file).

- With **R.E.D. II**, a "\$i" * "\$j" RESP fit, where "\$i" stands for orientations and "\$j" enumerates conformations, can be automatically performed. Consequently, the number of centers quickly increases when a multiorientation-multiconformation RESP fit is applied. For a molecular system, the center numbers is limited to 500 in the "RESP" program by default. Thus, if more than 500 centers are calculated in the RESP fit, one has to modify the "RESP" source code and recompile the program. The number of centers is defined by:

$$\text{\$Center-Nb} = \text{\$Atom-Nb} * \text{\$i} * \text{\$j}$$

Where "Center-Nb", "Atom-Nb", "\$i" and "\$j" are the number of centers, atoms, conformations and orientations, respectively.

It is strongly advised to always check the "output1" and "output2" "RESP" output files. For example, if too many constraints are used in the fit (originating for instance from multiorientation RESP fit), the "RESP" program might not be able to compute the charges, and error messages similar to the following one might be written in the "output1" or "output2" file:

```
qchng = NaN
chgopt: LU decomp gave almost-singular U
```

- **R.E.D. II**, does not handle intra-molecular charge restraints yet. Thus, if one needs to perform a RESP fit with such restraints one has to proceed in two steps. First, the RESP or ESP charges have to be calculated without restraints with **R.E.D.** Then, the "RESP" inputs have to be modified to incorporate appropriate charge restraints, and using the "espot" file, the "RESP" program has to be executed in a 'standalone mode', as follows:

For RESP charge derivation, a two stage fit is required:

```
resp -O -i input1 -e espot -o output1 -p punch1 -q qout \
-t qout1 -w qwts -s esout
resp -O -i input2 -e espot -o output2 -p punch2 -q qout1 \
-t qout2 -w qwts -s esout
```

For ESP charge derivation, a single stage fit is required:

```
resp -O -i input1 -e espot -o output1 -p punch1 -q qout \
-t qout1 -w qwts -s esout
```

-V- THE 'INITIAL' PDB FILE

This file is used each time **R.E.D.** (or **X R.E.D.**) is executed, and must be carefully prepared before executing **R.E.D.** With **R.E.D. II**, RESP and ESP charges can be derived using several conformations in the fitting process. Thus, the Cartesian coordinates of each conformation must be present in this initial PDB file. Each set of Cartesian coordinates representing a conformation must be separated by the "**TER**" keyword (the latter has to be present at the beginning of a line; See the new tutorial for **R.E.D. II**). As previously said, the atom order must be the same in each conformation.

* PDB atom names

Generalities

- In ESP charge derivation,^[3] the atomic charge values are fitted to reproduce the MEP, and the charge equivalencing of chemically equivalent atoms is performed *a posteriori* to the fit.

- In RESP charge derivation,^[1,2] the atomic charges are fitted to reproduce the MEP in a two stage fit. The atomic charge values are affected by the use of hyperbolic restraints, and the charge value equivalencing of chemically equivalent atoms is carried out during the fit.

* In 'standard' RESP charge derivation (*i. e.* the published way by the Kollman's group)^[1,2] a weak restraint (**qwt = 0.0005**) is used in the charge derivation for all the heavy atoms during the first stage, and a stronger restraint (**qwt = 0.001**) is applied only to the methyl and methylene (CH₃ and CH₂) carbons during the second stage.

* In 'non-standard' RESP charge derivation, the use of the "**qwt = 0.001**" restraint in the second stage has been extended to all types of carbon (in C=O, CH, C=C for instance), nitrogen, oxygen, silicon, phosphate and sulfur atoms. However, the use of such 'non-standard' RESP inputs is only recommended for expert users.

R.E.D. automatically generates the inputs used in ESP, 'standard' and 'non-standard' RESP charge derivation for the "RESP" program based on the atom names found in the initial PDB file. Thus, the user must edit this PDB file, and modify the atom names before running **R.E.D.** according to the three following rules:

-1- The atom names in the initial PDB file, whose the atomic charges have to be re-optimized in the second RESP stage (*i. e.* using the "**qwt = 0.001**" restraint) must present the "**T**" letter after their chemical symbol. With **R.E.D. I**, only 'standard' RESP inputs could be produced. Consequently, only the methyl and methylene carbons had to present a "**T**" letter in their atom name. The other atom names had only to use the letter of their chemical symbol [*i. e.* "**C**" for the other carbons (in C=O, CH, C=C for instance), "**O**" for oxygen, "**H**" for hydrogen etc...] in the definition of their atom name. In **R.E.D. II**, this rule has been extended to five other atoms (N, O, Si, P and S) and is used for the generation of 'non-standard' RESP inputs.

-2- The same number (positive integer) must be added to this/these letter(s) for equivalent atoms (whatever the equivalencing procedure is, *i. e.* during the fit or *a posteriori* to the fit). Consequently, non-equivalent atoms must have different atom names.

-3- Each hydrogen linked to an heavy atom must have the same number as this atom.

PDB atom names: Chemical symbol + "**T**" (if needed) + positive integer

Remarks:

- "HO1", "1H2" or "O4" (often found in PDB files) are not valid names. Using the three rules reported above, the correct way could be "H1", "H2" or "O4", respectively. "I1" or "Ag1" will also be rejected because their total number of electrons are above $Z = 35$.

- Equivalent atoms must have the same atom names [same letter(s) and number(s)] in the initial PDB file. The "InsightII" program (<http://www.accelrys.com>) renames automatically the atoms, which have the same names to differentiate them. On the contrary, the "VMD" program (<http://www.ks.uiuc.edu/Research/vmd/>) displays the atom name labels without modifying them, and thus, is very convenient to use for checking the selected atom names written in the PDB file. These 'user-modified' atom names are only necessary in the initial PDB file. Thus, the "Tripos" and PDB files generated after geometry optimization by **R.E.D.** have more classical atom names [the atom name is then composed of the chemical symbol (even for the methyl and methylene carbons) and a number which is automatically incremented].

- When several conformations are available in the initial PDB file, only the atom names (defined with the three rules defined previously) belonging to the first conformation are used in the RESP input generation, and are applied to all other conformations. More generally, the residue names, residue numbers, connectivity and re-orientation information based on three atoms (See the new tutorial for **R.E.D. II**) of the first conformation are used in the definition of other conformations when needed.

See the "*Tutorials*" for more information.

*** Molecular orientation of the optimized geometry**

It is well known that the number of MEP points depends on the molecular orientation in space.^[15,16] Consequently, the values of derived RESP and ESP atomic charges also depend on the molecular orientation of the optimized geometry.

The molecular orientation can be partially controlled in the "GAMESS" program using the molecular principal axes ("COORD" = "CART" keyword) and in the "Gaussian" program by placing the center of nuclear charge at the origin ("Symmetry" keyword). Thus, since "GAMESS" and "Gaussian" do not apply the same internal algorithm, the molecular orientation of the minimized structure generated by both programs will be different. Moreover, each QM program can generate different molecular orientations for a target minimum using its internal re-orientation algorithm. To address these problems, a new reorientation algorithm has been introduced in the **R.E.D.** program which can be applied to the optimized molecular system before calculating the MEP. It allows a full control of the molecular orientation of the optimized geometry independently of the QM program or initial structure selected. Consequently, highly reproducible RESP or ESP charges can be derived. This procedure applied to the optimized geometry uses a Rigid-Body Re-orientation Algorithm based on the choice of three atoms. The first two atoms define the (O , X) axis, the third one defines the (O , X , Y) plane; the Z -axis being defined as the $X \times Y$ cross-product (For more information, see <http://www.srl.gatech.edu/education/ME6175/notes/xforms.html>).

Three different procedures are available in **R.E.D.** to control the molecular orientation of the optimized geometry before computing the MEP. We strongly recommend to use the last one in RESP or ESP charge derivation:

-1- If **re-orientation** information based on three atoms (see case -2-, below) is not provided in the initial PDB file, then the internal Re-orientation Algorithm available in "GAMESS" or "Gaussian" is used to reorient the optimized Cartesian coordinates (We called this the QMRA procedure).

In this case, the molecular orientation of the optimized geometry is partially controlled, the charges derived using the two QM programs differ from each other, and are not reproducible.

It is important to underline that the "Gaussian" program reorients by default (using its own internal algorithm by placing the center of nuclear charge at the origin) a Cartesian coordinate set each time a molecular energy calculation is performed. This leads to the well known set of coordinates called 'Standard orientation'. Thus, at the end of a geometry optimization the Cartesian coordinates generated by "Gaussian" are always re-oriented with respect to the initial geometry (if the "**Symmetry**" keyword is used in the input). On the contrary, the "GAMESS" program only reorients the molecular Cartesian coordinates once at the beginning of a job (if the "**COORD**" = "**CART**" keyword is used). Thus, at the end of a geometry optimization the Cartesian coordinates generated by "GAMESS" are not perfectly re-oriented (according to the "GAMESS" internal algorithm). In order to get a well re-oriented optimized geometry using "GAMESS", one has to use the optimized Cartesian coordinates in a new single point energy run. In the **R.E.D. I** version, this was performed during the MEP computation (if the Rigid-Body Re-orientation Algorithm was not used). This has been changed in the **R.E.D. II** version, where a new routine has been incorporated (which uses the Jacobi matrix diagonalization method) to re-orient the "GAMESS" optimized geometry. This is a similar algorithm to the one used internally by "GAMESS", and presents the advantage that the "GAMESS" type molecular orientation can be generated without the need of executing the "GAMESS" program itself. Finally, such a 'standalone' re-orientation algorithm is not needed if the "Gaussian" program is used since the molecular orientation available at the end of the geometry optimization output is well re-oriented (according to the "Gaussian" internal algorithm).

-2- If re-orientation information based on three atom numbers (not atom names!) is provided in the initial PDB file, the minimized structure is automatically re-oriented using the Rigid-Body Re-orientation Algorithm (We called this the RBRA procedure).

Thus, in this case the molecular orientation of the optimized geometry is fully controlled and highly reproducible RESP and ESP charges can be derived independently of the QM software or initial structure choice. The charges are reproducible since the three atoms used in the RBRA procedure are known.

In order to apply this approach, one has to provide the following format in the initial PDB file before the set of Cartesian coordinates of the first conformation:

After the "**REMARK**" PDB keyword (which has to be present at the beginning of a line), one needs to add the "**REORIENT**" keyword (which has to be separated by only one space character from the "**REMARK**" keyword) *i. e.*

REMARK REORIENT atm_nb\$A atm_nb\$B atm_nb\$C (in the same line)

- Where "**\$A**", "**\$B**" and "**\$C**" stand for the atom numbers

- The keywords and the atom numbers must be written in the same line

-3- If "\$j" sets of three atom numbers ["\$j" is a positive integer representing the number of molecular orientation(s); an infinite number of orientations can be in principle used] are given, the optimized Cartesian coordinates obtained from the QM program output are reoriented "\$j" times using the Rigid-Body Re-orientation Algorithm (We called this the multi-RBRA procedure). The reoriented sets of Cartesian coordinates are then used to compute "\$j" MEPs. A "\$j" orientation fit (or multi-orientation RESP fit) is then applied using the "RESP" program.

This allows averaging out the charge value differences observed for a particular orientation over several different orientations. The derived atomic charges are also highly reproducible in this case.

Here is an example in which "\$j" = 4:

```
REMARK REORIENT atm_nb$A atm_nb$B atm_nb$C | atm_nb$C atm_nb$B atm_nb$A |
atm_nb$D atm_nb$C atm_nb$A | atm_nb$A atm_nb$C atm_nb$B
```

- The keywords and the atom numbers must be written in the same line
- The pipe character "|" is used as separator between two different orientations
- "\$A", "\$B", "\$C", and "\$D" are the atom numbers

For larger number of molecular reorientations, the corresponding format can be used, each line starting by the "REMARK REORIENT" keywords *i. e.* using several lines:

```
REMARK REORIENT atm_nb$A atm_nb$B atm_nb$C | atm_nb$C atm_nb$B atm_nb$A
REMARK REORIENT atm_nb$D atm_nb$B atm_nb$A | atm_nb$A atm_nb$B atm_nb$D
REMARK REORIENT atm_nb$E atm_nb$D atm_nb$A | atm_nb$A atm_nb$D atm_nb$E
```

As mentioned above, the **R.E.D. II** program can be used to derive RESP or ESP charges using several conformations (*i. e.* "\$i" conformations). The three procedures applied previously to control molecular orientation can be used to perform charge fitting for the "\$i" conformations provided in the initial PDB file:

- If no Rigid Body Re-orientation Algorithm information is provided in the initial PDB file, a "\$i" conformation RESP fit is performed using the molecular orientation calculated by the internal Re-orientation Algorithm of the QM program (QMRA procedure).

- If "\$j" orientations ("j" = 1 or higher positive integer, denoting the molecular orientation based on the Rigid-Body Re-orientation Algorithm) are requested, a "\$i" conformation * "\$j" orientation RESP fit is performed (RBRA or multi-RBRA procedure).

Remark:

In the latter case, the re-orientation information based on the three atom names is only provided for the first conformation (before the first set of Cartesian coordinates; See the new tutorial for **R.E.D. II**) and is used for all the "\$i" conformations.

The **R.E.D. II** program allows using the geometry optimization output (single conformation optimization output or concatenated file created after "\$i" conformation optimizations) generated by the "GAMESS" program as input for MEP computation by the "Gaussian" program (and *vice-versa*). If the Rigid-Body Re-orientation Algorithm is not selected, the molecular orientation used to

compute the MEP is based on the QM output and is not based on the QM software. This means that if a "Gaussian" geometry optimization output is used to compute MEP using the "GAMESS" program (for instance), then the molecular orientation selected is based on the "Gaussian" internal re-orientation algorithm. However, this particularity is obviously avoided if the Rigid-Body Re-orientation Algorithm is used.

R.E.D. II also automatically generates PDB files to check and visualize the molecular orientation of the optimized geometries in a graphic interface:

- The optimized structures, which have been re-oriented based on the internal Re-orientation Algorithm of the QM software (QMRA procedure) are named Mol_o*\$i*-qmra.pdb, where "*\$i*" is the conformation number. These files are generated if the QMRA or RBRA procedures are selected. However, the Cartesian coordinates available in the Mol_o*\$i*-qmra.pdb files are those used in the MEP computation only if the QMRA procedure is selected. If the RBRA procedure is selected, then the Mol_o*\$i*-qmra.pdb files are also generated. However, they are generated only for information (*i. e.* to simply check what would have been the molecular orientation if the QMRA procedure would have been selected).

- The optimized structures, which have been re-oriented based on the Rigid-Body Re-orientation Algorithm (RBRA procedure) are named Mol_o*\$i*-rbra*\$j*.pdb ("*\$i*" and "*\$j*" are the number of molecular conformation(s) and orientation(s), respectively). These files are generated only if the RBRA procedure is selected. In this case, the Cartesian coordinates available in such Mol_o*\$i*-rbra*\$j*.pdb files are those used in the MEP computation.

See the "*Tutorials*" for more information.

* *Atom connectivity*

If the atom connectivity information is available in the initial PDB file (*i. e.* the PDB file contains the lines with the keyword "**CONNECT**"), a Tripos file (*i. e.* "mol2" format <http://www.tripos.com/custResources/mol2Files/>) with optimized Cartesian coordinates and RESP or ESP charges is generated by **R.E.D.** This format is very powerful since the Cartesian coordinates, force field atom types, atomic charges, atom names and residue names are available in this file. In "AMBER8" (not implemented in "AMBER7"), such a Tripos format is automatically recognized by "LEaP" and allows the automatic generation of "LeaP" OFF libraries for instance (See the "AMBER8/LEaP" documentation).

Remarks:

- With **R.E.D. II**, "*\$i*" Tripos files are generated if "*\$i*" conformations are used in the atomic charge derivation. The charges available in the "*\$i*" Tripos files are identical since the derived charges are the average of those obtained for the different conformations. Thus, the "*\$i*" Tripos files differ only in their Cartesian coordinates, which belongs to the different conformations obtained after geometry optimization.

- The molecular orientation available in these Tripos files is not important since this file is supposed to be used in molecular mechanics simulation, where only the atomic charge values matter. The Cartesian coordinates present in such a file correspond also to the optimized geometry.

- We also use the Tripos format within "InsightII/Discover" (<http://www.accelrys.com>) to run molecular mechanics calculations with AMBER force fields converted by AmberFFC (<http://www.u-picardie.fr/labo/lbpd/AmberFFC/>),^[17] and to automatically calculate force field parameters using FFParmDev (<http://www.u-picardie.fr/labo/lbpd/FFParmDev/>, program still under development).

Restrictions:

- **R.E.D. II** generates only the atomic chemical symbol in the force field atom type column in the Tripos "mol2" format. Thus, force field atom types of the AMBER force fields have to be manually added in this column. Then, the Tripos file can be loaded within "AMBER8/LEaP" to create the Cartesian coordinates and topology files in the AMBER format (the "**prmcrd**" and "**prmtop**" files, respectively) and/or an "LEaP" OFF library.

- Since **R.E.D.** does NOT verify atom connectivity, the user must check carefully the connectivity provided in the initial PDB file. A simple way to generate safely and automatically a initial PDB structure with correct connectivity is to execute the Babel program (<http://ccl.net/cqa/software/UNIX/babel/index.shtml> or <http://smog.com/chem/babel>, last version: 1.6, this software is not developed anymore) or the Openbabel program (<http://openbabel.sourceforge.net>, which is the official version of Babel under development).

To convert a PDB file without connectivity information into a PDB file with connectivity included, use the following commands:

```
babel -ipdb Mol-noconnect.pdb -opdb Mol-connect.pdb
ln -s Mol-connect.pdb Mol.pdb if "$MOL_START" = "Mol.pdb" in R.E.D.
```

- Since only single bond connectivity are available in the PDB format (For more information, see http://www.rcsb.org/pdb/docs/format/pdbguide2.2/guide2.2_frame.html), only this type of bond is printed in the Tripos format by **R.E.D.** Thus, if the user wants to display double or triple bonds in a graphic user interface such as "InsightII" or "Sybyl" (<http://www.tripos.com>), the last column of the "**@<TRIPOS>BOND**" section must be manually modified (<http://www.tripos.com/custResources/mol2Files/>). However, since only single bonds are used in molecular mechanics, missing double or triple bonds in the Tripos format generated by **R.E.D.** do not cause any problem for simulations within "AMBER8" (or "InsightII").

With **R.E.D. II**, it is now possible to derive RESP or ESP charges using "\$i" conformations. In this case, the connectivity information are only provided for the first conformation (after the first set of Cartesian coordinates; See the new tutorial for **R.E.D. II**), which are used for all the "\$i" conformations. Since the atom order in each conformation is the same, the connectivity information is required only for the first conformation.

If no atom **connectivity** is provided in the initial PDB file, the RESP and ESP charges are only available in the RESP "punch2" file (which is normally created by the RESP program) and no Tripos file is generated.

See the "*Tutorials*" for more information.

-VI- UPDATED TUTORIAL OF R.E.D. I FOR R.E.D. II

* Example of QMRA procedure (using DMSO)

- Below is an initial PDB file (*i. e.* not optimized) needed for the execution of **R.E.D.**:

```
REMARK File name = dms0-noconnect.pdb; Directory name = /usr/people/fyd/DMSO
ATOM      1  CT1  DMSO      1      3.8750  0.6780 -8.4170  1.00  0.00
ATOM      2  H1   DMSO      1      3.8000  1.6900 -8.0760  1.00  0.00
ATOM      3  H1   DMSO      1      3.4060  0.0260 -7.7110  1.00  0.00
ATOM      4  H1   DMSO      1      3.3890  0.5830 -9.3660  1.00  0.00
ATOM      5  S2   DMSO      1      5.3590  0.2930 -8.5590  1.00  0.00
ATOM      6  O3   DMSO      1      5.4600 -1.0590 -9.0140  1.00  0.00
ATOM      7  CT1  DMSO      1      6.0550  0.4300 -7.1930  1.00  0.00
ATOM      8  H1   DMSO      1      7.0870  0.1630 -7.2900  1.00  0.00
ATOM      9  H1   DMSO      1      5.9800  1.4410 -6.8530  1.00  0.00
ATOM     10  H1   DMSO      1      5.5820 -0.2190 -6.4850  1.00  0.00
```

- => The two carbons are equivalent and belong to two methyl groups.
 - => The "CT1" atom name is used.
- => The six hydrogens are equivalent and are linked to a "CT1" carbon atom.
 - => Same number as the carbon atoms (*i. e.* "H1").
- => Sulfur and oxygen atom names: just have a number different from "1".

- No connectivity and no re-orientation information based on three atoms are provided in this initial PDB file.

=> No Tripos "mol2" file is generated and the RESP type charges are only available in the "punch2" file.

=> The molecular orientation of the optimized geometry is based on the internal algorithm used by the QM software.

=> Two warnings will be printed in the **R.E.D.** ".log" file:

```
WARNING: No three atom based re-orientation found in the PDB file
Re-orientation will be done according to the GAMESS Algorithm
```

```
WARNING: No connectivity information found in the PDB file
The Tripos Mol2 file(s) will NOT be generated !
```

- Below are examples of **R.E.D. II** variables (see section "MAIN PROGRAM" in **R.E.D.**)

```
$XRED = "Off";           # XRED will not be used
$NP    = "1";           # Number of processor used = 1 (in QM calculations)
$QM_SOFT = "GAMESS";   # GAMESS will be used in the QM calculations

$SOPT_Calc = "on";     # Geometry optimization will be performed
$MEPCHR_Calc = "on";  # The MEP & charges will be calculated

$CHR_TYP = "RESP";    # RESP charges will be derived
$CHR_VAL = "0";       # Total charge of DMSO = 0
$MLT_VAL = "1";       # Spin multiplicity of DMSO = 1

$TITLE = "DMSO";      # Title for the molecule/project
$DIR    = "CHRCalc";  # Directory name where the data will be stored
$MOL_START = "./Mol.pdb"; # Initial PDB File (ln -s dms0-noconnect.pdb Mol.pdb)
$JOB_OPT = "./JOB1.log"; # Minimization OUTPUT used only if $SOPT_Calc = "OFF"
```

Now, the **R.E.D.** program can be executed :

```
perl RED-vII.pl > RED.log &
```

The DMSO molecule is first minimized. Then, the optimized Cartesian coordinates are reoriented using the internal algorithm available in the "GAMESS" program (QMRA procedure). Then, the MEP and RESP charges are calculated (the ".log" file is saved in "RED.log").

*** Example of RBRA procedure (using EtOH)**

- Below is an initial PDB file needed for the execution of **R.E.D.**:

```
REMARK REORIENT 1 5 8
REMARK File name = EtOH.pdb; Directory name = /usr/people/fyd/EtOH
ATOM 1 CT1 EtO 1 3.1080 0.6530 -8.5260 1.00 0.00
ATOM 2 H1 EtO 1 2.8150 -0.3490 -8.7610 1.00 0.00
ATOM 3 H1 EtO 1 2.5170 1.0150 -7.7110 1.00 0.00
ATOM 4 H1 EtO 1 2.9560 1.2780 -9.3810 1.00 0.00
ATOM 5 CT2 EtO 1 4.5970 0.6740 -8.1320 1.00 0.00
ATOM 6 H2 EtO 1 4.7480 0.0490 -7.2770 1.00 0.00
ATOM 7 H2 EtO 1 5.1870 0.3120 -8.9470 1.00 0.00
ATOM 8 O3 EtO 1 4.9880 2.0120 -7.8180 1.00 0.00
ATOM 9 H3 EtO 1 5.9160 2.0250 -7.5720 1.00 0.00
CONNECT 1 2 3 4 5
CONNECT 2 1
CONNECT 3 1
CONNECT 4 1
CONNECT 5 1 6 7 8
CONNECT 6 5
CONNECT 7 5
CONNECT 8 5 9
CONNECT 9 8
END
```

=> The two carbons are not equivalent and belong to a methyl and methylene groups.

=> Different atom names "CT1" and "CT2" are used

=> The three methyl hydrogens have the same number as the methyl carbon, *i. e.* "H1" and the two methylene hydrogens bears the same number as the methylene carbon, *i. e.* "H2".

- Connectivity information is provided in the initial PDB file.

=> The charges are available in the "punch2" and Tripos mol2 files.

- A single re-orientation based on three atoms is requested ("REMARK REORIENT 1 5 8" in the initial PDB file).

=> The minimized Cartesian coordinates are reoriented and used in MEP computation.

- Below are examples of **R.E.D. II** variables (see the "MAIN PROGRAM" section):

```
$XRED = "Off"; # XRED will not be used
$NP = "1"; # Number of processor used = 1 (in QM calculations)
$QM_SOFT = "GAUSSIAN"; # GAUSSIAN will be used in the QM calculations

$OPT_Calc = "on"; # Geometry optimization will be performed
$MEPCHR_Calc = "on"; # The MEP & charges will be calculated

$CHR_TYP = "RESP"; # RESP charges will be derived
$CHR_VAL = "0"; # Total charge of EtOH = 0
$MLT_VAL = "1"; # Spin multiplicity of EtOH = 1

$TITLE = "Ethanol"; # Title for the molecule/project
$DIR = " "; # No directory name is provided
# The data will be stored in the working directory
$MOL_START = "./Mol.pdb"; # Initial PDB File (ln -s EtOH.pdb Mol.pdb)
$JOB_OPT = "./JOB1.log"; # NOT used since $OPT_Calc = "on"
```

Now, **R.E.D.** can be executed :

```
perl RED-vII.pl
```

(The ".log" file at this time is simply printed in the terminal; see [Figure 3](#))

The ethanol molecule is first minimized. Then, the set of minimized Cartesian coordinates is reoriented once according to the three atom numbers provided (RBRA procedure). Finally, MEP and charges are calculated using the reoriented Cartesian coordinates. A Tripos "mol2" file is generated containing the optimized Cartesian coordinates and RESP charges.

*** Example of Multi-RBRA procedure (using trans NMA)**

-1- Multi-orientation 'RESP' fit

- Below is an initial PDB file needed for the execution of **R.E.D.**:

```
REMARK REORIENT 1 5 7 | 7 5 1 | 5 6 7 | 7 6 5
REMARK REORIENT 1 7 9 | 9 7 1
REMARK File name = Mol.pdb; Directory name = /usr/people/fyd/NMeAcetamide
ATOM      1  CT1  NMA      1      2.0000  1.0000  0.0000  1.00  0.00
ATOM      2  H1   NMA      1      1.6360 -0.0270  0.0000  1.00  0.00
ATOM      3  H1   NMA      1      1.6360  1.5140  0.8900  1.00  0.00
ATOM      4  H1   NMA      1      1.6360  1.5140 -0.8900  1.00  0.00
ATOM      5  C2   NMA      1      3.5220  1.0000  0.0000  1.00  0.00
ATOM      6  O3   NMA      1      4.1440 -0.0600  0.0000  1.00  0.00
ATOM      7  N4   NMA      1      4.1200  2.1940  0.0000  1.00  0.00
ATOM      8  H4   NMA      1      3.5610  3.0350  0.0000  1.00  0.00
ATOM      9  CT5  NMA      1      5.5630  2.3280  0.0000  1.00  0.00
ATOM     10  H5   NMA      1      6.0200  1.3380  0.0000  1.00  0.00
ATOM     11  H5   NMA      1      5.8770  2.8730 -0.8900  1.00  0.00
ATOM     12  H5   NMA      1      5.8770  2.8730  0.8900  1.00  0.00
CONNECT   1    2    3    4    5
CONNECT   2    1
CONNECT   3    1
CONNECT   4    1
CONNECT   5    1    6    7
CONNECT   6    5
CONNECT   7    5    8    9
CONNECT   8    7
CONNECT   9    7   10   11   12
CONNECT  10    9
CONNECT  11    9
CONNECT  12    9
END
```

=> The carbon name of the carbonyl group: Atom name is NOT "CT2" but "C2".

=> The three hydrogens in the two methyl groups have the same number as the methyl carbon to whom they are bound, *i. e.* "H1" and "H5" (the two methyl groups are not equivalent).

- Connectivity information is provided in the initial PDB file (a Tripos file is generated).

- Six different orientations based on three atoms are requested (RBRA procedure): The minimized Cartesian coordinates are reoriented six times and for each orientation a MEP is calculated. The RESP type charges are derived using a six-orientation RESP fit.

- Below are examples of **R.E.D. II** variables:

```
$XRED    = "Off";           # If XRED = ON, R.E.D. has to be executed with XRED.tcl
$NP      = "1";            # Number of processor used for parallel run of QM calculation(s)
$QM_SOFT = "GAMESS";      # GAMESS will be used in geometry optimization & MEP computation

$OPT_Calc = "on";         # Geometry optimization will be done if $OPT_Calc = ON
$MEPCHR_Calc = "on ";    # MEP & charges will be calculated if $MEPCHR_Calc = ON

$CHR_TYP = "RESP";       # RESP charges will be derived
$CHR_VAL = "0";          # Total charge of NMeAcetamide = 0
$MLT_VAL = "1";          # Spin multiplicity of NMeAcetamide = 1

$title   = "MOL";         # Title for the molecule/project
$DIR     = "TEST";        # Directory name where the data will be stored
$MOL_START = "./Mol.pdb"; # Initial PDB File (unoptimized structure)
$JOB_OPT = "./JOB1.log";  # Minimization OUTPUT used if $OPT_Calc = "OFF"
```

Now, **R.E.D.** can be executed: `perl RED-vII.pl > RED.log &`

-2- Multi-orientation 'ESP' fit

- Use the same PDB file as for the example described above in "-1- Multi-orientation 'RESP' fit"

Remarks:

- The PDB file used has two atom carbon names, which have a "T" letter ("CT1" & "CT5"). This naming convention is used for the generation of the input of the second RESP stage. Since a single stage is used in the ESP fit, these "T" letters are not necessary for the input creation and are simply ignored if found.

- On the other hand, the atoms which will be equivalenced *a posteriori* to the fit must have identical names (for example the "H1" and "H5" atoms).

- Below are **R.E.D. II** variables to be used:

```
$XRED = "Off";           # XRED will not be used
$NP   = "1";           # Number of processor used = 1
$QMSOFT = "GAUSSIAN";  # GAUSSIAN will be used in the QM calculations

$OPT_Calc = "off";     # Geometry optimization was ALREADY performed
# See "-1- Multi-orientation 'RESP' fit"
# The MEP & charges will be calculated

$CHR_TYP = "ESP";     # ESP charges will be derived
# (Only useful for the Weiner et al. FF)
$CHR_VAL = "0";       # Total charge of NMeAcetamide = 0
$MLT_VAL = "1";       # Spin multiplicity of NMeAcetamide = 1

$TITLE = "Ethanol";   # Title for the molecule/project
$DIR   = "TEST";      # Directory name (where the data are saved)
# The directory name is automatically incremented
# i. e. TEST, TEST-1, TEST-2, etc...

$MOL_START = "./Mol.pdb"; # Initial PDB File
# Must be provided each time R.E.D. is executed
$JOB_OPT = "TEST/JOB1-gam.log"; # Obtained from "-1- Multi-orientation 'RESP' fit"
```

Now, **R.E.D.** can be executed : `perl RED-vII.pl > RED.log`

Six orientations are requested (RBRA procedure): The minimized Cartesian coordinates are reoriented six times and for each orientation a MEP is calculated. The 'ESP' type charges are calculated using a six-orientation fit.

Remark:

When the calculations are finished for this tutorial, the "*Multi-orientation 'RESP' fit*" data are stored in the "**TEST-1**" directory and the "*Multi-orientation 'ESP' fit*" data are available in the "**TEST**" directory!

*** Applications***a) "Multi-RED" run*

In this example, we want to run calculations for four different molecules. For each molecule a specific **R.E.D.** script with adequate variables needs to be prepared. For example, here we have four different "RED-vII.pl" scripts called: "RED-DMSO.pl", "RED-ETOH.pl", "RED-NMA-RESP.pl" and "RED-NMA-ESP.pl". Finally, a "Multi-RED" script (see below) has to be created that will execute these four scripts sequentially.

```

cd                                # Go in the $HOME directory
touch Multi-RED                   # Create an empty file
chmod +x Multi-RED                # Make it executable
vi Multi-RED                       # Edit

i                                  (Insertion mode)
cd DMSO
echo "DMSO job"
perl RED-DMSO.pl > RED-DMSO.log

cd ../EtOH
echo "EtOH job"
perl RED-EtOH.pl > RED-EtOH.log

cd ../NMeAcetamide
echo "2 NMA jobs"
perl RED-NMA-RESP.pl > RED-NMA-RESP.log
perl RED-NMA-ESP.pl > RED-NMA-ESP.log

cd ..
echo "All Done!"
:wq!                               (Save & quit "vi")

```

The four "RED-molecule.pl" scripts are executed sequentially using the following command:

```
./Multi-RED                        # Execution
```

b) Execution of R.E.D. with "XRED.tcl"

- The "**\$XRED**" **R.E.D.** variable has to be set to "**ON**". Consequently, the others are no longer used.

```
$XRED = "On";                      # XRED is, this time, used to execute RED
```

- Now, program **X R.E.D.** can be executed: `wish XRED.tcl`

- Use the graphical interface (see [Figure 2](#)) to select the variables needed, the initial PDB file, etc...

-VII- NEW TUTORIAL FOR R.E.D. II

Three conformation-two orientation RESP fit (Alanine dipeptide)

In the following tutorial we present the RESP charge fitting procedure involving three different conformations of the alanine dipeptide (C5, C7ax and C7eq conformations). Additionally, for each conformation two molecular orientations will be used, that are defined using three atoms (*i. e.* using the RBRA procedure). This corresponds to a "\$i" * "\$j" RESP fit (where "\$i" and "\$j" are the number of molecular conformations and orientations, respectively).

Initially, we create three different PDB files representing the three target conformations of the alanine dipeptide using for example a graphical program. The atom order in each PDB file must be the same. Next, those PDB files are used to prepare three Gaussian input files using the Babel program:

```
babel -ipdb C5.pdb -oxyz C5.xyz
babel -ipdb C7ax.pdb -oxyz C7ax.xyz
babel -ipdb C7eq.pdb -oxyz C7eq.xyz
```

These three ".xyz" files are transformed into ".com" Gaussian inputs by adding Gaussian keywords to allow the geometry optimization of each conformation. The corresponding jobs are executed using the following commands, modifying the inputs if the optimization(s) fail (*i. e.* adding the calculation of the Hessians, and selecting the 'Standard orientation' having the smallest maximum and RMS forces and the smallest maximum and RMS displacements):

```
g98 < C5.com > C5.log
g98 < C7ax.com > C7ax.log
g98 < C7eq.com > C7eq.log
```

The "Gaussian" outputs are concatenated into a single file called "JOB1-gau.log":

```
cp C5.log JOB1-gau.log
cat C7ax.log >> JOB1-gau.log
cat C7eq.log >> JOB1-gau.log
```

- This "JOB1-gau.log" file is then loaded as input into **R.E.D.** using the three input variables: \$OPT_Calc = "off", \$MEPCHR_Calc = "on" and \$JOB_OPT = **"/JOB1-gau.log"**

- The initial PDB file corresponding to the three conformations of the alanine dipeptide is built from the three PDB files representing each alanine dipeptide conformation. This file is presented on the next page with numerous comments. The atom names are detailed and allows the generation of 'standard' RESP inputs.

- Below are **R.E.D. II** variables to be used (see MAIN PROGRAM section):

```
$XRED = "Off"; # XRED will not be used
$NP = "4"; # Number of processors used = 4 (computer cluster)
$QM_SOFT = "GAUSSIAN"; # GAUSSIAN program will be used

$OPT_Calc = "off"; # Geometry optimization was ALREADY performed
$MEPCHR_Calc = "on"; # The MEP and charges will be calculated

$CHR_TYP = "RESP"; # RESP charges will be derived
$CHR_VAL = "0"; # Total charge of the Alanine dipeptide = 0
$MLT_VAL = "1"; # Spin multiplicity of the Alanine dipeptide = 1

$TITLE = "Alanine-dipeptide"; # Title for the Alanine dipeptide biomolecule
$DIR = "Alanine"; # Directory name (where the data will be stored)
$MOL_START = "./3Conf-Alanine.pdb"; # Initial PDB File.
$JOB_OPT = "./JOB1-gau.log"; # Obtained from 'manual' execution/concatenation
```

```

REMARK REORIENT 12 15 18 | 5 13 18
REMARK Conformation C5
ATOM 1 CT1 ACE 1 3.3280 -1.1550 0.0270 1.00 0.00
ATOM 2 H1 ACE 1 2.9220 -2.0690 -0.3900 1.00 0.00
ATOM 3 H1 ACE 1 4.1110 -0.7830 -0.6220 1.00 0.00
ATOM 4 H1 ACE 1 3.7740 -1.3680 0.9910 1.00 0.00
ATOM 5 C3 ACE 1 2.2960 -0.0640 0.2080 1.00 0.00
ATOM 6 O4 ACE 1 2.5910 1.0080 0.6700 1.00 0.00
ATOM 7 N5 ALA 2 1.0400 -0.3680 -0.1750 1.00 0.00
ATOM 8 C6 ALA 2 -0.0490 0.5660 -0.0320 1.00 0.00
ATOM 9 H5 ALA 2 0.8070 -1.2660 -0.5320 1.00 0.00
ATOM 10 H6 ALA 2 0.0610 1.0770 0.9170 1.00 0.00
ATOM 11 C8 ALA 2 -1.3450 -0.2400 -0.0340 1.00 0.00
ATOM 12 O9 ALA 2 -1.4020 -1.3370 -0.5290 1.00 0.00
ATOM 13 CT7 ALA 2 -0.0680 1.6080 -1.1590 1.00 0.00
ATOM 14 H7 ALA 2 0.8700 2.1480 -1.1570 1.00 0.00
ATOM 15 H7 ALA 2 -0.1930 1.1200 -2.1200 1.00 0.00
ATOM 16 H7 ALA 2 -0.8750 2.3210 -1.0270 1.00 0.00
ATOM 17 N10 NME 3 -2.4050 0.3660 0.5290 1.00 0.00
ATOM 18 CT2 NME 3 -3.7280 -0.2230 0.5080 1.00 0.00
ATOM 19 H10 NME 3 -2.2960 1.2820 0.8940 1.00 0.00
ATOM 20 H2 NME 3 -4.2340 -0.0470 -0.4360 1.00 0.00
ATOM 21 H2 NME 3 -3.6490 -1.2890 0.6610 1.00 0.00
ATOM 22 H2 NME 3 -4.3170 0.2070 1.3080 1.00 0.00
CONNECT 1 2 3 4 5
CONNECT 2 1
CONNECT 3 1
CONNECT 4 1
CONNECT 5 1 6 7
CONNECT 6 5
CONNECT 7 5 8 9
CONNECT 8 7 10 11 13
CONNECT 9 7
CONNECT 10 8
CONNECT 11 8 12 17
CONNECT 12 11
CONNECT 13 8 14 15 16
CONNECT 14 13
CONNECT 15 13
CONNECT 16 13
CONNECT 17 11 18 19
CONNECT 18 17 20 21 22
CONNECT 19 17
CONNECT 20 18
CONNECT 21 18
CONNECT 22 18
MASTER 0 0 0 0 0 0 0 0 22 0 22 0
TER
REMARK Conformation C7ax
ATOM 1 C ACE 1 3.1260 0.6910 -0.7590 1.00 0.00
ATOM 2 H ACE 1 3.3240 0.0780 -1.6300 1.00 0.00
ATOM 3 H ACE 1 3.0920 1.7320 -1.0540 1.00 0.00
ATOM 4 H ACE 1 3.9370 0.5760 -0.0510 1.00 0.00
ATOM 5 C ACE 1 1.8320 0.3520 -0.0540 1.00 0.00
ATOM 6 O ACE 1 1.4970 0.9560 0.9370 1.00 0.00
ATOM 7 N ALA 2 1.1030 -0.6410 -0.5990 1.00 0.00
ATOM 8 C ALA 2 -0.1370 -1.2130 -0.0750 1.00 0.00
ATOM 9 H ALA 2 1.4350 -1.0340 -1.4480 1.00 0.00
ATOM 10 H ALA 2 -0.3390 -2.0710 -0.7010 1.00 0.00
ATOM 11 C ALA 2 -1.3720 -0.3250 -0.2810 1.00 0.00
ATOM 12 O ALA 2 -2.3240 -0.7640 -0.8740 1.00 0.00
ATOM 13 C ALA 2 -0.0160 -1.6950 1.3730 1.00 0.00
ATOM 14 H ALA 2 0.7820 -2.4260 1.4490 1.00 0.00
ATOM 15 H ALA 2 0.1940 -0.8830 2.0520 1.00 0.00
ATOM 16 H ALA 2 -0.9450 -2.1710 1.6690 1.00 0.00
ATOM 17 N NME 3 -1.3420 0.9030 0.2550 1.00 0.00
ATOM 18 C NME 3 -2.4730 1.7950 0.1310 1.00 0.00
ATOM 19 H NME 3 -0.5160 1.2050 0.7220 1.00 0.00
ATOM 20 H NME 3 -2.6740 2.0340 -0.9070 1.00 0.00
ATOM 21 H NME 3 -3.3660 1.3510 0.5520 1.00 0.00
ATOM 22 H NME 3 -2.2490 2.7100 0.6640 1.00 0.00
TER
REMARK Conformation C7eq
ATOM 1 C ACE 1 -3.1140 0.4980 -0.8470 1.00 0.00
ATOM 2 H ACE 1 -3.1900 -0.2380 -1.6380 1.00 0.00
ATOM 3 H ACE 1 -3.9870 0.4290 -0.2100 1.00 0.00
ATOM 4 H ACE 1 -3.1010 1.4900 -1.2810 1.00 0.00
ATOM 5 C ACE 1 -1.8840 0.3350 0.0150 1.00 0.00
ATOM 6 O ACE 1 -1.6550 1.0930 0.9260 1.00 0.00
ATOM 7 N ALA 2 -1.0750 -0.7010 -0.2840 1.00 0.00
ATOM 8 C ALA 2 0.1230 -1.0190 0.4830 1.00 0.00
ATOM 9 H ALA 2 -1.2650 -1.2470 -1.0920 1.00 0.00
ATOM 10 H ALA 2 -0.0730 -0.7180 1.5030 1.00 0.00
ATOM 11 C ALA 2 1.3150 -0.2090 -0.0470 1.00 0.00
ATOM 12 O ALA 2 2.1520 -0.7020 -0.7570 1.00 0.00
ATOM 13 C ALA 2 0.4000 -2.5140 0.4230 1.00 0.00
ATOM 14 H ALA 2 -0.4510 -3.0640 0.8090 1.00 0.00
ATOM 15 H ALA 2 0.6040 -2.8290 -0.5930 1.00 0.00
ATOM 16 H ALA 2 1.2720 -2.7580 1.0180 1.00 0.00
ATOM 17 N NME 3 1.3370 1.0770 0.3430 1.00 0.00
ATOM 18 C NME 3 2.3350 1.9940 -0.1620 1.00 0.00
ATOM 19 H NME 3 0.5350 1.4360 0.8130 1.00 0.00
ATOM 20 H NME 3 3.3260 1.5880 -0.0070 1.00 0.00
ATOM 21 H NME 3 2.2110 2.1860 -1.2230 1.00 0.00
ATOM 22 H NME 3 2.2500 2.9290 0.3760 1.00 0.00
END

```

Reorientation information based on three atoms

- Optional, but recommended
- Present at the beginning of the line
- Sets of three atom numbers are provided
- Defined for the first conformation but used for all

PDB format (10 columns):

- 1: The ATOM or HETATM keyword
 - Present at the beginning of the line
- 2: Atom numbers
 - Used in the mol. reorientation based on three atoms
- 3: Atom names
 - Used in the RESP input generation
- 4: Residue names
 - Defined for the first conformation but used for all
 - Used in the optimized PDB & Tripos file generation
- 5: Residue numbers
 - Defined for the first conformation but used for all
 - Used in the optimized PDB & Tripos file generation
- 6-8: X, Y, Z Cartesian coordinates
 - Obviously different in each conformation
- 9-10: Not used by R.E.D.

Example of atom names used in the generation of 'standard' RESP inputs

- Defined for the first conformation but used for all

Connections generated by the Babel program (Optional, but recommended)

- Present at the beginning of the line
- Present just after the Cartesian coordinates of the first conformation
- Defined for the first conformation but used for all

This implies that the atom order must be the same in the conformations used

MASTER keyword not used by R.E.D.

TER keyword

- Needed and present at the beginning of the line
- Used to identify two sets of Cartesian coordinates (i. e. two conformations)

END keyword not used by R.E.D.

- Two files are used as inputs within **R.E.D.:**

3Conf-Alanine.pdb	Initial PDB file (previous page)
JOB-gau.log	Concatenated file originating from the geometry optimization of the C5, C7ax and C7eq conformations of the alanine dipeptide

- Here is the listing of the 40 files generated when using **R.E.D.:**

esout	One of the RESP outputs
espot	Contains the \$i * \$j MEP files concatenated into a single file \$i conformations * \$j orientations <i>i. e.</i> 6 MEPs in the RESP format Used as input by RESP
espot\$i-\$j	MEP files for each conformation/orientation in the RESP format (6 files) generated by R.E.D., but not used by RESP
input1	RESP input for the first RESP stage (generated by R.E.D. based on the atom names defined for the first conformation found in the initial PDB file)
input2	RESP input for the second RESP stage
JOB2-gau\$i-\$j.com	Gaussian MEP inputs (6 files), conformation \$i, orientation \$j
JOB2-gau\$i-\$j.log	Gaussian MEP outputs (6 files), conformation \$i, orientation \$j
Mol_o\$i.mol2	Optimized Cartesian coordinates for each \$i conformation with RESP charges in the Tripos format (3 files), <u>to be loaded in AMBER8/LEaP</u>
Mol_o\$i-rbra\$j.pdb	Optimized Cartesian coordinates for each \$i conformation, \$j orientation representing the molecular orientation calculated based on three atoms (RBRA procedure) in the PDB format (6 files)
Mol_o\$i-qmra.pdb	Optimized Cartesian coordinates for each \$i conformation representing the orientation generated by the Gaussian internal algorithm (QMRA procedure) in the PDB format (3 files). <u>These files are generated only for information</u> in order to inspect what would have been the molecular orientation used in the MEP computation if the QMRA procedure would have been selected instead of the RBRA one.
output1	RESP output generated for the first RESP stage which contains information about the RESP fit (restraints, charge values, RESP keywords...)
output2	RESP output generated at the second RESP stage
punch1	RESP output. It contains the atomic charges calculated during the first RESP stage
punch2	RESP output. It contains the <u>final atomic charges</u> calculated during the second RESP stage (used by R.E.D. to add the charge values in the Mol_o\$i.mol2 files)
qout1	Charges obtained in the first stage of RESP calculations
qout2	Charges obtained in the second stage of RESP calculations
File4REDDb.pdb	Optimized conformation(s) summarized in a unique file for R.E.D.D.B.

Remarks:

- It is highly recommended to use an initial PDB file that has hydrogen atoms put immediately after the heavy atom they are attached to. This allows to easily locate the methyl and methylene chemical groups and to quickly find potential errors in the PDB atom names and consequently in the RESP inputs. Unfortunately, many graphic interfaces generate PDB files in which all the hydrogens are located after all the heavy atoms. Such an approach is inconvenient for preparing RESP inputs, and should be avoided.

- If no information based on three atoms is provided, the QMRA procedure is selected (instead of the the RBBA one) in order to partially control the molecular orientation of the optimized geometries. Similar files are generated with the following differences:

A single molecular orientation is used for each conformation. Thus, only the files "espot\$i-\$j", "JOB2-gau\$i-\$j.com" and "JOB2-gau\$i-\$j.log", with "\$j" = 1, are generated.

The "Mol_o\$i-rbra\$j.pdb" files are not generated. The "Mol_o\$i-qmra.pdb" files are created, which have the same Cartesian coordinates as those that can be found in the MEP inputs.

-VIII- HOW TO REFERENCE R.E.D.?

- R.E.D. I

A. Pigache,⁽¹⁾ P. Cieplak⁽²⁾ & F.-Y. Dupradeau⁽¹⁾ Automatic and highly reproducible RESP and ESP charge derivation: Application to the development of programs RED and X RED, 227th ACS National Meeting, Anaheim, CA, March 28 - April 1, **2004**.

- R.E.D. II

F.-Y. Dupradeau,^(1,3) A. Pigache,⁽¹⁾ T. Zaffran^(1,3) & P. Cieplak⁽⁴⁾ Automatic, highly reproducible and effective RESP and ESP charge Derivation: Implementation of multiconformation and/or multiorientation RESP fit in the R.E.D. program, *manuscript submitted*.

(1) DMAG EA 3901 &, Faculté de Pharmacie, Amiens, France.

(2) Accelrys Inc., San Diego, CA, USA.

(3) The Scripps Research Institute, Dpt of Mol. Biology, La Jolla, CA, USA.

(4) The Burnham Institute, La Jolla, CA, USA.

-IX- REFERENCES

[1] WD Cornell, P Cieplak, CI Bayly & PA Kollman, *J. Am. Chem. Soc.* **1993**, 115, 9620-9631.

[2] CI Bayly, P Cieplak, WD Cornell & PA Kollman, *J. Chem. Phys.* **1993**, 97, 10269-10280.

[3] UC Singh & PA Kollman, *J. Comput. Chem.* **1984**, 5, 129-145, and references cited herein.

[4] CA Reynolds, JW Essex & WG Richards, *J. Am. Chem. Soc.* **1992**, 114, 9075-9079.

[5a] Gaussian 98, Revision A.11, MJ Frisch, GW Trucks, HB Schlegel, GE Scuseria, MA Robb, JR Cheeseman, VG Zakrzewski, JA Montgomery, Jr., RE Stratmann, JC Burant, S Dapprich, JM Millam, AD Daniels, KN Kudin, MC Strain, O Farkas, J Tomasi, V Barone, M Cossi, R Cammi, B Mennucci, C Pomelli, C Adamo, S Clifford, J Ochterski, GA Petersson, P Y Ayala, Q Cui, K Morokuma, DK Malick, AD Rabuck, K Raghavachari, JB Foresman, J Cioslowski, JV Ortiz, AG Baboul, BB Stefanov, G Liu, A Liashenko, P Piskorz, I Komaromi, R Gomperts, RL Martin, DJ Fox, T Keith, MA Al-Laham, CY Peng, A Nanayakkara, C Gonzalez, M Challacombe, PMW Gill, B Johnson, W Chen, MW Wong, JL Andres, C Gonzalez, M Head-Gordon, ES Replogle, & JA Pople, 1998 Gaussian, Inc., Pittsburgh PA.

[5b] Gaussian 03, Revision A.1, Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery Jr JA, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA,

Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA 2003 Gaussian, Inc., Pittsburgh PA.

[6] MW Schmidt, KK Baldrige, JA Boatz, ST Elbert, MS Gordon, JH Jensen, S Koseki, N Matsunaga, KA Nguyen, S Su, TL Windus, M Dupuis & JA Montgomery, *J. Comput. Chem.* **1993**, *14*, 1347-1363.

[7] DA Pearlman, DA Case, JW Caldwell, WS Ross, TE Cheatham, III, S DeBolt, D Ferguson, G Seibel, & P. Kollman, *Comp. Phys. Commun.* **1995**, *91*, 1-41.

[8] DA Case, TA Darden, TE Cheatham III, CL Simmerling, J. Wang, RE Duke, R Luo, KM Merz, B Wang, DA Pearlman, M Crowley, S Brozell, V Tsui, H Gohlke, J Morgan, V Hornak, G Cui, P Beroza, C Schafmeister, JW Caldwell, WS Ross & PA Kollman, **2004**, AMBER 8, University of California, San Francisco.

[9] WD Cornell, P Cieplak, CI Bayly, IR Gould, KM Mertz Jr, DM Ferguson, DC Spellmeyer, T Fox, JW Caldwell & PA Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.

[10] PA Kollman, R Dixon, W Cornell, T Fox, C Chipot & A Pohorille, The development/application of a 'minimalist' organic/biochemical molecular mechanics force field using a combination of *ab initio* calculations and experimental data. In *Computer simulation of biomolecular systems*, A Wilkinson, P Weiner, W van Gunsteren (eds), Kluwer, Dordrecht, **1997**, *3*, 83-96.

[11] TE Cheatham III, P Cieplak & PA Kollman, *J. Biomol. Struct. Dyn.* **1999**, *16*, 845-862.

[12] J Wang, P Cieplak & PA Kollman, *J. Comput Chem.* **2000**, *21*, 1049-1074.

[13] SJ Weiner, PA Kollman, DA Case, UC Singh, C Ghio, G Alagona, S Profeta Jr & PA Weiner, *J. Am. Chem. Soc.* **1984**, *106*, 765-784.

[14] SJ Weiner, PA Kollman, DT Nguyen & DA Case, *J. Comput. Chem.* **1986**, *7*, 230-252.

[15] RJ Woods, M Khalil, W Pell, S H Moffat & V H Smith Jr, *J. Comput. Chem.* **1990**, *11*, 297-310.

[16] C M Breneman & K B Wiberg, *J. Comput. Chem.* **1990**, *11*, 361-373.

[17] A Dejoux, P Cieplak, N Hannick, G Moyna & F-Y Dupradeau, *J. Mol. Model.* **2001**, *7*, 422-432.