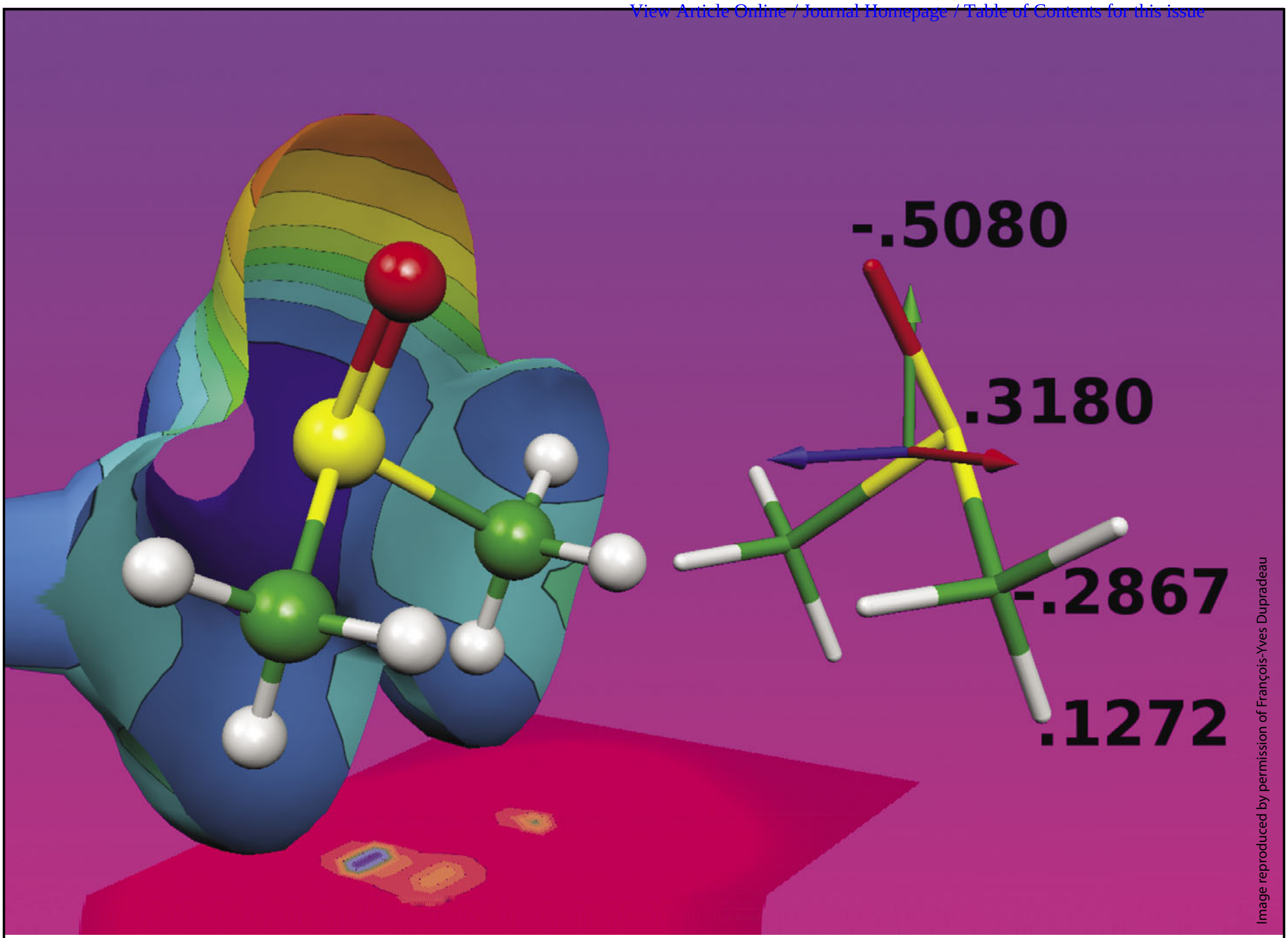

Image reproduced by permission of François-Yves Dupradeau

## Showcasing research from q4md-forcefieldtools.org (a collaboration between François-Yves Dupradeau and Piotr Cieplak)

**Title: The R.E.D. Tools: Advances in RESP and ESP charge derivation and force field library building**

Charge derivation and force field library building for new molecules useful in Amber, CHARMM, GLYCAM and OPLS force field based simulations.

Methods and computational tools useful for charge derivation and force field topology database building are presented. They give researchers the means to derive rigorously QM MEP-based charges embedded in force field libraries that are ready to be used in force field development, charge validation and/or MD simulations.

**As featured in:**

See F.-Y. Dupradeau, A. Pigache, T. Zaffran, C. Savineau, R. Lelong, N. Grivel, D. Lelong, W. Rosanski and P. Cieplak, *Phys. Chem. Chem. Phys.*, **12**, *7821*.

**www.rsc.org/pccp**
Registered Charity Number 207890

# The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building†

**François-Yves Dupradeau,*[a] Adrien Pigache,[a] Thomas Zaffran,[a] Corentin Savineau,[a] Rodolphe Lelong,[a] Nicolas Grivel,[a] Dimitri Lelong,[a] Wilfried Rosanski[a] and Piotr Cieplak*[b]**

***Received 2nd April 2010, Accepted 7th June 2010***
***First published as an Advance Article on the web 23rd June 2010***
**DOI: 10.1039/c0cp00111b**

Deriving atomic charges and building a force field library for a new molecule are key steps when developing a force field required for conducting structural and energy-based analysis using molecular mechanics. Derivation of popular RESP charges for a set of residues is a complex and error prone procedure because it depends on numerous input parameters. To overcome these problems, the R.E.D. Tools (RESP and ESP charge Derive, http://q4md-forcefieldtools.org/RED/) have been developed to perform charge derivation in an automatic and straightforward way. The R.E.D. program handles chemical elements up to bromine in the periodic table. It interfaces different quantum mechanical programs employed for geometry optimization and computing molecular electrostatic potential(s), and performs charge fitting using the RESP program. By defining tight optimization criteria and by controlling the molecular orientation of each optimized geometry, charge values are reproduced at any computer platform with an accuracy of 0.0001 e. The charges can be fitted using multiple conformations, making them suitable for molecular dynamics simulations. R.E.D. allows also for defining charge constraints during multiple molecule charge fitting, which are used to derive charges for molecular fragments. Finally, R.E.D. incorporates charges into a force field library, readily usable in molecular dynamics computer packages. For complex cases, such as a set of homologous molecules belonging to a common family, an entire force field topology database is generated. Currently, the atomic charges and force field libraries have been developed for more than fifty model systems and stored in the RESP ESP charge DDataBase. Selected results related to non-polarizable charge models are presented and discussed.

## Introduction

The atomic charge or monopole approximation is an important concept in chemistry. It plays a significant role in molecular simulations that use empirical force fields. Difficulties in developing such charges arise from the fact that they are not observables, as is an electron density. Moreover, no real criterion has been established to rigorously validate the quality of atomic charges. It has been required that these charges should be independent of the computational method, basis set, molecular orientation or conformation, or should be able to reproduce experimental multipole moments, or should be transferable and conform to the atom electronegativities.[1] Consequently, many different approaches have been proposed for the derivation of atomic charges: the Mulliken and Löwdin population analysis,[2,3] the atom in a molecule theory,[4] empirical approaches to reproduce some crystallographic[5] or liquid data,[6] the electrostatic potential (ESP) derived charges using semi-empirical[7] or *ab initio* methods,[8–12] and the AM1-BCC approach.[13,14] Unfortunately, no charge model has proved to be the best in all respects.

ESP atomic-centered charges are fitted to reproduce the molecular electrostatic potential (MEP), which is a molecular property directly derived from the self-consistent field (SCF) calculation.[8–12] The MEP is calculated at a large number of points defined on three-dimensional surfaces around the molecule of interest. For instance the Amber force fields use atomic charges which are fitted to the MEP calculated using the Connolly surface,[10,15] while the GLYCAM force field employs the CHELPG algorithm.[16] CHARMM force field developers have indiscriminately used both approaches, while Connolly surface-based OPLS-like charges have been reported.[17–20] ESP charges are known to optimally handle inter-molecular properties, which are essential for condensed phase simulations where the solute–solvent and solvent–solvent interactions have to be well represented and balanced. However, they may be less suited to reproduce intra-molecular properties and molecular

[a] *CNRS UMR 6219 & UFR Pharmacie, Université de Picardie-Jules Verne, 1, rue des Louvels, F-80037 Amiens cedex 1, France. E-mail: fyd@q4md-forcefieldtools.org; Fax: +33-(0)3-2282-7469; Tel: +33-(0)3-2282-7498*
[b] *Sanford|Burnham Institute for Medical Research, La Jolla, California, 92037, USA. E-mail: pcieplak@burnham.org; Fax: +1-858-713-9949; Tel: +1-858-646-3100*

† Electronic supplementary information (ESI) available: Examples of P2N input files for the R.E.D. program, a screen snapshot of the X R.E.D. graphical interface as well as the summary of the molecular systems studied in this work and submitted to R.E.DD.B. See DOI: 10.1039/c0cp00111b

conformations.[21,22] Moreover, they are not easily transferable between common groups of homologous molecules and depend on molecular orientation and conformation.[21–26] The origin of these problems comes from the observation that MEP points must lie outside the molecular van der Waals (vdW) surface. Thus, buried atoms such as carbon atoms in hydrocarbon chemical groups are not represented by a significant number of MEP points. Due to the statistical nature of the fitting process, these poorly defined centers lead to a number of artifacts in conformational energetic, large charge values and substantial orientation and conformation variability.[21–25] Several improvements have been successively introduced to limit these problems and implemented in various force fields. Reynolds *et al.* proposed to derive general ESP atomic charge values over a range of conformations.[27] Kollman's group used weak hyperbolic restrains to hold down ESP charges with a minor impact on the fit leading to the restrained ESP (or RESP) charge model.[28,29] The association of multiple conformation in charge derivation with the so-called non-polarizable "RESP" charge model lead to charge values widely used in Amber force field development, and recognized as particularly well-suited for condensed phase molecular dynamics (MD) simulations. Following a similar approach, Woods and coworkers developed a specific RESP model for carbohydrates.[30] Kollman and MacKerell's groups also extended the restrained fitting approach to polarizable force fields based on concentric Connolly surfaces.[31,32]

The selection of the *ab initio* basis set to compute MEP is a key aspect in RESP and ESP charge derivation. ESP charge values demonstrate important fluctuations depending on the basis set if a low level of theory is used. On the contrary, after reaching the 6-31G* basis set,[33,34] ESP charges tend to converge with respect to the size of the basis set used.[35] The Hartree–Fock (HF) method and the STO-3G basis set[34,36,37] were used to calculate ESP charges for the Weiner *et al.* force field,[38,39] while the HF/6-31G* theory level was applied to derivation of RESP charges for the Cornell *et al.* force field,[40] and its successive modifications.[41–44] The General Amber Force Field (GAFF) was designed based on the AM1-BCC charge model,[13,14] which was parametrized to match MEP computed at the HF/6-31G* theory level.[45] The GLYCAM force field uses HF/6-31G* MEP-based charge values for condensed phase simulations, while HF/cc-pVTZ is involved in computing MEP for gas phase calculations.[34,46,47] Indeed, the HF/6-31G* theory level yields dipole moments, which are approximately ten percent larger than those observed in the gas phase.[28,29] This effect is exploited in condensed phase simulations for taking into account the implicit polarization of the solute in an aqueous solution in the additive force field model.[40,43] In contrast, the B3LYP type exchange and correlation functionals[48] and the cc-pVTZ, cc-pVDZ and aug-cc-pVDZ basis sets[34,49] were applied to compute gas phase charge values for polarizable force fields.[31,32] In a different approach, Duan *et al.* used the B3LYP/cc-pVTZ theory level and implicit solvent model for constructing a third generation Amber force field for proteins.[50]

The elementary steps required for deriving RESP and ESP charges for a new molecule are as follows. First, the geometry of the molecule of interest is optimized, and then the MEP around the optimized geometry is computed. Both steps are carried out using quantum mechanical approaches. Finally, the charge values are fitted to reproduce the MEP computed in the previous step. The RESP and FITCHARGE programs have been developed for this purpose,[28,32] and are used in the development of the Amber, GLYCAM and CHARMM force fields. Although this method is routinely used to derive atomic charges for many force fields, it has, in our opinion, several limitations. Applied to a large set of molecules and/or conformations, the approach is tedious, time-consuming, error-prone and lacks a rigorous way to verify the calculated charges. In the process, different programs and scripts have to be sequentially used. Although in principle any *ab initio* program could be employed for quantum mechanics (QM) calculations, the Gaussian program,[51] which is a proprietary software, is mainly used by the Amber community.[52,53] The academic programs such as GAMESS-US (General Atomic and Molecular Electronic Structure System),[54] PC-GAMESS/Firefly,[55] and NWChem[56] which have similar functionalities to the Gaussian program with respect to MEP computation, are not widely used to derive RESP or ESP charges. Indeed, it is known that RESP or ESP atomic charges derived using the GAMESS-US or NWChem program are "different" from these calculated by Gaussian. Moreover, even using the Gaussian program, the RESP or ESP charges for a structure of interest are not easily reproducible and noticeable discrepancies between authors are observed. Finally, no program is available allowing automatic RESP or ESP charge derivation and force field library building for a new organic or bio-inorganic molecular fragment compatible with existing ones. The Antechamber program, introduced in the AmberTools, clearly solves some of the problems previously reported, but it is only capable of deriving charges for organic molecules, and neither addresses the problem of charge reproducibility nor the derivation of atomic charges for molecular fragments.[57] To the best of our knowledge no program exists that combines together the multiple conformation and multiple molecule fitting approach with the generation of a set of force field libraries compatible with a bio-molecular force field.

Here, we report on new approaches implemented in the RESP ESP charge Derive (R.E.D.) Tools that can be applied to derive non-polarizable RESP or ESP atomic charges for an ensemble of molecules, and to build a set of force field libraries for molecules and molecular fragments in the Tripos MOL2 file format.[58,59] When a large family of molecules is involved in the procedure, an entire force field topology database (FFTopDB) is generated. An extension of the RESP or ESP charge fitting method employing multiple molecular orientation feature is presented, and is applied to a preselected number of conformations and molecules. This procedure, which couples multiple orientation, multiple conformation and multiple molecule RESP or ESP fit enables automatic derivation of RESP and ESP charges for any complex bio-molecular system. Atomic charges derived in this way are independent of the QM program or initial Cartesian coordinate choice, and are reproduced with an accuracy of 0.0001 e. In the R.E.D. program various *ab initio* theory levels, surface algorithms for MEP computation, and different fitting approaches can be selected making it a versatile program capable of creating a large number of fixed-charge models and allowing for efficient comparisons between the charge sets generated. More than fifty molecular systems involved in multiple orientation, multiple conformation and/or multiple molecule RESP or ESP fit have been studied and are reported below. Charge values, optimized Cartesian coordinates, computational conditions and force field libraries generated for these molecular systems have been deposited in the RESP ESP charge DDataBase (R.E.DD.B.), and are freely available for downloading.[60] More generally, the goal of this work is not to provide a methodology for developing atomic charges suitable for any particular MD condition, but to give the researchers the means to derive rigorously QM MEP-based charges embedded in force field libraries ready to be used in force field development, charge validation and/or MD simulations.

## The R.E.D. Tools: the Ante_R.E.D., R.E.D. and X R.E.D. programs

The R.E.D. Tools consists of the Ante_R.E.D., R.E.D. and X R.E.D. programs. The Ante_R.E.D. and R.E.D. programs have been written using the practical extraction and report language (or Perl),[61] while the X R.E.D. program has been developed using the tool command language/graphical user interface toolkit or tcl/tk programming languages.[62] Perl and tcl/tk are interpreted and dynamic languages, and consequently the source code of these programs does not need to be compiled before being executed. Perl and tcl/tk follow the open source community philosophy, and are therefore freely available on the Internet for numerous platforms. Ante_R.E.D., R.E.D. and X R.E.D. are highly portable to UNIX, Macintosh and Windows operating systems, and are available for downloading from the q4md-forcefieldtools.org web site at http://q4md-forcefieldtools.org/RED/. The source code of the R.E.D. Tools is provided within the distribution.[63] A tutorial has been written demonstrating the use of the R.E.D. Tools.[64]

### The Ante_R.E.D. program & the P2N file format

Ante_R.E.D. is a preparatory program used for generating all necessary input files for the R.E.D. program. For each molecule, the input of Ante_R.E.D. has to be provided in the PDB file format.[65] Ante_R.E.D. automatically re-orders the atoms in a structure in such a way that the hydrogen atoms are located after the heavy atoms they are bound to, and generates a series of output files necessary for the execution of the R.E.D. program. Atom reordering is helpful for identifying the methyl and methylene groups necessary for preparing inputs for the RESP charge fitting program.[28] The Ante_R.E.D. program has to be executed separately for each molecule or molecular conformation. It produces a set of files, which are: (i) input files for the Gaussian, GAMESS-US and PC-GAMESS/Firefly programs containing the correct keywords required for the geometry optimization step, (ii) a text file summarizing the atom connectivities, and (iii) new PDB and P2N files with atom connectivities. The P2N file format has the .p2n extension, and is the format exclusively recognized by the R.E.D. program. It contains two columns of atom names: the first column, which is mandatory, is used for the automatic generation of the input file(s) required for the charge fitting step, while the second one, which is optional, is involved in preserving the PDB international atom name conventions found in force field libraries.[66] The P2N file format also contains specific keywords used by the R.E.D. program. These keywords are used for defining the IUPAC name [needed for identifying the molecule(s) in the process of charge derivation as well as for referencing the molecule(s) in a R.E.DD.B. project],[60,67] the total charge and spin multiplicity (required in QM computations) and the atom connectivities (required for establishing the topology) of each molecular system. Additional keywords are used for setting the rules determining the molecular re-orientation procedure, which is applied before the MEP computation step, as well as for defining charge constraints (intra-, inter-molecular charge constraints and intermolecular charge equivalencing) used during the charge fitting step. Detailed descriptions of the Ante_R.E.D. program and the definition of the new P2N file format associated with the R.E.D. Tools are available in a specific tutorial.[64] Examples of such a P2N file format for the ethanol, dimethylphosphate and *N*-acetyl-*O*-methyl-L-tyrosine-*N*′-methylamide molecules are presented in the Fig. S1 of the supplementary material.†

### The R.E.D. program & the Tripos MOL2 file format

Among the R.E.D. Tools, R.E.D. is the key program responsible for charge derivation and force field library building for a molecular system. R.E.D. automatically derives non-polarizable RESP and ESP charges for a set of $n$ molecules represented by $i$ conformations, where each of these conformations can be additionally represented by $j$ molecular orientations in space ($n$, $i$ and $j$ are positive integers greater or equal to one), and then generates the corresponding force field libraries in the Tripos MOL2 file format.[58] It handles different charge models depending on the method used in MEP computation and charge fitting. So far, the Connolly surface and CHELPG algorithms in MEP computation as well as one stage ESP, one stage RESP and two stage RESP fitting strategies have been implemented. The R.E.D. program sequentially performs the following steps: it (i) optimizes the geometry of the different conformations of a molecule, and computes the corresponding MEP on surfaces around each structure using the Gaussian, GAMESS-US or PC-GAMESS/Firefly program, and (ii) runs the RESP program in order to fit the atomic charges to the MEP determined in the previous step. The R.E.D. program automatically generates input files for the Gaussian, GAMESS-US, PC-GAMESS/Firefly and RESP programs and performs necessary format conversions. Attention has been paid to the preparation of the RESP inputs containing a minimum number of charge constraints. The flexibility of the P2N file format allows following different approaches for charge equivalencing for chemically equivalent atoms or for atoms considered chemically equivalent. This has particular importance when one needs to derive charge values for a whole molecule or a molecular fragment.[59,64] As previously defined by Williams,[12,68] the key here is to perform rigorous charge fitting leading to low RRMS values (relative root mean square between the MEP computed by QM and the one calculated using the fitted charge values) independently of the complexity of the assumed charge derivation protocol.

The molecular orientation of each optimized geometry is controlled in R.E.D. by either the re-orientation algorithm available in the QM programs,[69,70] or by a rigid-body re-orientation algorithm incorporated into the R.E.D. program itself.[71,72] These two re-orientation procedures have been named QMRA (for Quantum Mechanics Re-orientation Algorithm) and RBRA (for Rigid-Body Re-orientation Algorithm), respectively. Consequently, two different charge fitting methods depending on the re-orientation method are available in R.E.D. In the first approach, the internal Gaussian or GAMESS-US re-orientation scheme is applied to re-orient the optimized geometry (the way of controlling molecular orientation in GAMESS-US and PC-GAMESS/Firefly is identical). In the second one, the optimized geometry can be re-oriented *j* times using the RBRA approach, and charge fitting is carried out for the structure reoriented *j* times in space. By including several molecular orientations for an optimized geometry the charge uncertainty observed in charge derivation is substantially decreased. The procedure can be applied to every *i*th conformation.[72] Thus, starting from *n* P2N files characterizing *n* molecules, the geometry optimization, MEP computation and charge fitting steps are repeated *n* times, and yield *n* independent charge fits each of them involving *i* * *j* MEP computations. As a result, the force field libraries are generated for each conformation independently. For each molecule, the introduction of intra-molecular charge constraint(s) in a P2N file leads to an additional charge fit and to the generation of a force field library for the appropriate molecular fragment (Fig. 1A).[59] In the last step, the *n* molecules are combined together and a *n* molecule charge fit is carried out. At this stage intra-molecular charge constraints within a molecule, as well as inter-molecular charge constraints, and/or inter-molecular charge equivalencing between molecules can be setup during this *n* molecule fit. The procedure generates various types of molecular fragments such as those for amino acids, mono-saccharides and/or nucleotides.[47,59] All-atom as well as united-carbon force field library models can be generated.[38,40,59,73] In complex cases involving an ensemble of molecules, a large set of force field libraries can be built and stored in a FFTopDB (Fig. 1B). In principle, there is no limit to the number of the P2N files in conducting calculations by the R.E.D. program, and so far we have not observed any limitation in the fitting step using the RESP program. The comparison of the charge values obtained for every single molecule fit with these obtained in the multiple molecule charge fit provides an efficient way of verifying the charges derived in complex approaches.

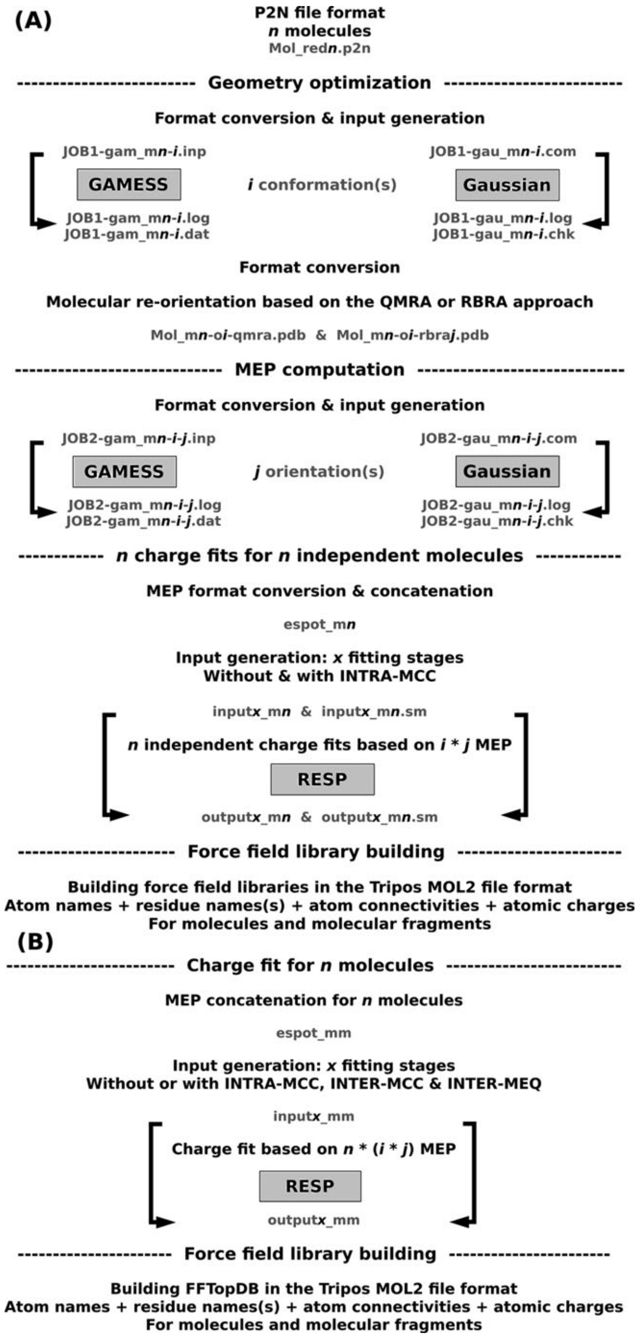
(A)
P2N file format
n molecules
Mol_redn.p2n
Geometry optimization
Format conversion & input generation
JOB1-gam_mn-i.inp
JOB1-gau_mn-i.com
GAMESS
i conformation(s)
Gaussian
JOB1-gam_mn-i.log
JOB1-gam_mn-i.dat
JOB1-gau_mn-i.log
JOB1-gau_mn-i.chk
Format conversion
Molecular re-orientation based on the QMRA or RBRA approach
Mol_mn-oi-qmra.pdb & Mol_mn-oi-rbraj.pdb
MEP computation
Format conversion & input generation
JOB2-gam_mn-i-j.inp
JOB2-gau_mn-i-j.com
GAMESS
j orientation(s)
Gaussian
JOB2-gam_mn-i-j.log
JOB2-gam_mn-i-j.dat
JOB2-gau_mn-i-j.log
JOB2-gau_mn-i-j.chk
n charge fits for n independent molecules
MEP format conversion & concatenation
espot_mn
Input generation: x fitting stages
Without & with INTRA-MCC
inputx_mn & inputx_mn.sm
n independent charge fits based on i * j MEP
RESP
outputx_mn & outputx_mn.sm
Force field library building
Building force field libraries in the Tripos MOL2 file format
Atom names + residue names(s) + atom connectivities + atomic charges
For molecules and molecular fragments
(B)
Charge fit for n molecules
MEP concatenation for n molecules
espot_mm
Input generation: x fitting stages
Without or with INTRA-MCC, INTER-MCC & INTER-MEQ
inputx_mm
Charge fit based on n * (i * j) MEP
RESP
outputx_mm
Force field library building
Building FFTopDB in the Tripos MOL2 file format
Atom names + residue names(s) + atom connectivities + atomic charges
For molecules and molecular fragments

**Fig. 1** Description of the tasks sequentially executed by the R.E.D. program initiated from a set of fully characterized molecules. (A) The geometry optimization, MEP computation and charge fitting steps are repeated *n* times for the *n* molecules involved in charge derivation. (i) Geometry optimization for *i* conformations, (ii) MEP computation for *j* orientations; both steps can be carried out using either the Gaussian or GAMESS (*i.e.* GAMESS-US or PC-GAMESS/Firefly) program; different approaches for controlling the molecular orientation of each optimized geometry are available, and (iii) atomic charge fitting using the RESP program involving *i* * *j* MEP; different charge fitting procedures depending on the selected molecular re-orientation scheme are implemented. *n*, *i* and *j* are positive integers greater or equal to one. (B) A multiple orientation, multiple conformation and multiple molecule charge fitting is carried out in a last step leading to the generation of a FFTopDB. $x$ = number of stage(s) in the fitting process ($x$ = 1 or 2, so far). Definitions of specific constraints used during the charge fitting steps: INTRA-MCC: intra-molecular charge constraint within a molecule; INTER-MCC: inter-molecular charge constraint between two different molecules; INTER-MEQ: inter-molecular charge equivalencing between atomic charges belonging to different molecules.[59] FFTopDB: force field topology database representing an entire set of force field libraries for molecules and molecular fragments generated in the Tripos MOL2 file format. Definition of the molecular re-orientation procedure: "QMRA": Re-orientation Algorithm based on the QM program; "RBRA": Re-orientation Algorithm based on a Rigid-Body transformation.[72]

The Tripos MOL2 file format employed here offers many advantages and is very attractive.[58] It contains information about the Cartesian coordinates, the atom and residue names, the force field atom types, the atomic charges and connectivities. Many molecular modeling programs are compatible with this file format, and it is recognized by the majority of the graphical interfaces. Consequently, this is the file format chosen for the force field libraries generated by the R.E.D. program. Furthermore, it has to be emphasized that R.E.D. builds force field libraries, which are force field atom type independent. This means that the force field atom types of a structure are not determined by R.E.D., but are replaced by the corresponding chemical elements. Indeed, RESP and ESP charge values derived using R.E.D. are fully compatible with Amber and GLYCAM force fields,[40–47] and can be used in CHARMM and OPLS force field based simulations as well.[17–20] Consequently, a dedicated program can be used to assign specific force field atom types, accordingly. For instance, the Antechamber program could be used to add GAFF atom types for organic molecules,[57] while Cornell *et al.* or GLYCAM force field atom types could be added for biopolymers using a script based approach and the LEaP program.[74] The well known Openbabel program has a capability of performing atom typing as well.[75] Such features are demonstrated in the tutorial describing the R.E.D. Tools.[64]

### The X R.E.D. program

X R.E.D. is the graphical user interface (or GUI) of the R.E.D. program. It makes easy access to R.E.D. variables, and provides a simple and efficient way to execute the R.E.D. program *via* the graphical environment. The screen snapshot of the X R.E.D. interface is presented in Fig. S2 of the supplementary material.†

## Methods

Geometry optimization and MEP computation were performed using the Gaussian (versions 98 and 03),[76,77] GAMESS-US [versions 24 Mar 2007 (R3) and January 2009 (R1)],[54] and the PC-GAMESS/Firefly (versions 7.1) programs,[55] on a 1.67 GHz SGI Altix running the SUSE Linux Enterprise Server 10 operating system, an IBM RS6000 based cluster (AIX 5.2), R5000 and R12000 SGI workstations (IRIX 6.5.22), and/or PC Linux based workstations (Fedora 6.0, 8.0 and CentOS 5.2). RESP and ESP charge fitting was carried out using the RESP program.[28] The latter program was modified and recompiled to slightly increase the charge accuracy as well as the maximum number of charges, Lagrange constraints and molecules allowed during the fitting step (the convergence criteria "qtol", the maximum number of charge values "maxq", the maximum number of Lagrange constraints "maxlgr" and the maximum number of molecules "maxmol" were adjusted to $1.0 \times 10^{-5}$, 5000, 500 and 200, respectively). The HF method and the 6-31G* basis set were used to optimize molecular geometries.[33–35] MEP were computed based on two different approaches: using either (i) the HF/6-31G* theory level in the gas phase,[28,29] or (ii) the density functional theory (DFT) method, the B3LYP exchange and correlation functionals, the IEFPCM continuum solvent model ($\varepsilon = 4$) to mimic organic solvent environment, and the cc-pVTZ basis set.[48–50] The HF/STO-3G theory level[34,36,37] was also tested to calculate MEP since it was used in ESP charge derivation for the Weiner *et al.* force field.[38,39] Both the CHELPG and Connolly surface algorithms used in MEP calculation were considered in this work.[10,15,16] Charge derivation and building force field library reported here were carried out by the R.E.D. Tools. Initial structures were constructed using the LEaP or InsightII program.[74,78] The corresponding optimized geometries and charge values were displayed using the LEaP or VMD program.[74,79]

More than fifty molecular systems have been considered in this work in order to demonstrate the different capabilities of the R.E.D. Tools. Considering the large amount of data generated only few characteristic results will be presented below. The entire set of data is summarized in the Table S3 of the supplementary material,† and is available in R.E.DD.B. It includes well-studied structures for which atomic charge values are known allowing for comparisons with published data and creating a benchmark. Several new molecular systems are also reported. The first group of studied structures includes organic molecules such as ethanol (*anti* and *gauche*+ conformations),[29,43,47,80,81] dimethylsulfoxide,[81–83] dimethylphosphate (*gauche*+, *gauche*+ conformation),[40,59] trifluoroethanol (*anti* and *gauche*+ conformations),[84–86] methoxyethane (*anti* and *gauche*+ conformations),[40,43,47,80,87] *N*-methylacetamide (*cis* and *trans* conformations),[28–29,40,43,80,87,88] 1-4-dioxane (*chair* and *twist*-boat conformations),[43,89,90] ethane-1,2-diol (*anti anti anti*, *anti gauche*+ *anti*, *gauche*+ *anti gauche*−, *gauche*+ *gauche*− *gauche*+ and *gauche*+ *gauche*+ *gauche*+ conformations),[29,47,80] methanol,[25,28,29,40,47,80] propanone, ethanoic acid,[43,80] acetonitrile,[25] formamide,[25,87] methanal,[87] furan,[87] pyrrole, benzene,[40,80] toluene,[80] chloroform,[81] cyclohexane (*chair* and *twist*-boat conformations).[43,80,90] These molecules were involved in explicit solvent MD simulations and/or force field development in the past. The second group of structures studied consists of bio-molecules such as alanine dipeptide (*C5*, *C7ax* and *C7eq* conformations),[21,25,40,80,91] as well as standard deoxyribonucleosides (*i.e.* deoxyadenosine, deoxycytidine, deoxyguanosine and thymidine in the *C2′-endo* and *C3′-endo* conformations) and ribonucleosides (*i.e.* adenosine, cytidine, guanosine and uridine in the *C3′-endo* conformation).[40,42] Finally, following the strategy proposed by Cieplak *et al.*,[59] charge derivation and force field library building were carried out for various molecular fragments of unusual amino acids as well as for standard nucleic acid nucleotides. The central, H3N(+)-terminal, COO(−)-terminal molecular fragments (as well as terminal neutral fragments) of α-aminoisobutyric acid[92,93] and *O*-methyl-L-tyrosine residues[94] were generated using the corresponding *N*-acetyl *N′*-methylamide amino acid (with $\varphi$, $\phi$ dihedral angles characteristic for the α-helix and/or β-sheet secondary structures), methylammonium and acetate. The central, *5′*-terminal and *3′*-terminal fragments of standard nucleic acid nucleotides for DNA and RNA were obtained using dimethylphosphate (*gauche*+, *gauche*+ conformation), the four deoxyribonucleosides (*C2′-endo* and *C3′-endo* conformations) and/or four ribonucleosides (*C3′-endo* conformation).

In the following section of the article we will discuss reproducibility of the RESP or ESP charge models. We will first compare the charge values of single conformation molecular systems determined using a given QM program, *i.e.* using either Gaussian or GAMESS-US. For every molecule, geometry optimization was performed using four different sets of initial Cartesian coordinates selected randomly. Computation of MEP and derivation of charge values were carried out using each optimized Cartesian coordinate set. Then, we will compare results obtained using the Gaussian and GAMESS-US programs. In this context, the role of *ab initio* threshold criteria during geometry optimization, and the impact of different optimized geometry re-orientation procedures, available in both programs, on the charge values will be addressed. Finally, we will discuss a rigid-body re-orientation algorithm based on the selection of a set of three atoms, which has been implemented in the R.E.D. source code to provide a general method for reorienting optimized geometries before MEP computation. This approach is independent of the QM program used. According to this strategy, the first selected atom is translated to the origin of axes, the first two atoms define the (O, X) axis while the third one is used to define the (O, X, Y) plane. The (O, Z) axis is automatically set as the cross-product between the (O, X) and (O, Y) axes.[71,72] This approach can be used for every optimized molecular geometry, and is the basis for multiple orientation charge fitting.

In the last section, we will demonstrate how multiple orientations and multiple conformations can be combined together during charge derivation. The R.E.D. program provides easy setup for handling MEP computation (using either the Connolly surface or the CHELPG algorithm), single stage ESP, single stage RESP as well as two-stage RESP fitting, which makes it an efficient tool for comparing various charge models. In addition, the introduction of intra-molecular charge constraint(s) during charge fit extends the number of charge models and allows building force field libraries of molecular fragments in a similar way as it is done for the central fragment of an amino acid employed for building polypeptide chains.[59] Examples of charge derivation involving multiple orientations, multiple conformations and multiple molecules will be then described. Including more than one molecule in charge derivation and introducing inter-molecular charge constraints and inter-molecular charge equivalencing during the charge fitting allows the determination of atomic charges for a large variety of molecular fragments.[47,59] Thus, inter-molecular charge constraints can be used for defining molecular fusion between two molecules by eliminating groups of atoms with zero sum of charge. This approach is applied in the process of an automatic generation of the force field libraries of molecular fragments, from which larger systems can be built, and is a standard strategy for creating libraries of the central and terminal amino acid and nucleotide fragments. By analogy, this method can be directly extended to other biomolecular systems such as oligosaccharides, glycoconjugates as well as bio-inorganic complexes. In addition to the above features, R.E.D. is capable of generating all-atom or united-carbon atom charge models, and create appropriate force field libraries, which can be readily used for validation in MD simulations.[38,40,59,73] The simultaneous formation of an ensemble of force field libraries for a family of structures or FFTopDB is then presented and discussed using standard nucleic acids as an example.

## Results and discussion

### Charge reproducibility

**General considerations.** A number of studies involving non-polarizable RESP and ESP charge derivation have been reported.[59,80,81,87,89] When attempting to reproduce these published values, we regularly observed differences between our results and the published ones. This effect was more pronounced for some molecules than for the others. This raises the following questions: what are the factors that affect the reproducibility of RESP and ESP charges used in an additive force field, is it possible to quantify the corresponding uncertainty in charge values, and how can this inaccuracy be fixed? From the late seventies until the early nineties, several algorithms related to MEP computation and ESP charge derivation were developed, and reviews covering this topic are available.[12,95] It has been shown that these different algorithms as well as molecular orientation and molecular conformation affect RESP and ESP charge values. Woods *et al.* demonstrated that after implementing in the GAMESS-US program the CHELP algorithm,[11] which was originally developed for the Gaussian 82 program, they were not able to reproduce the published charge values for model systems.[23] Breneman and Wiberg described an improved method for deriving ESP charges based on CHELP,[11] denoted as CHELPG, which uses a different method for selecting points at which MEP is calculated and applies a higher density of points.[16] The authors indicated that this new approach led to charge values considerably less dependent upon molecular orientation. They reported the standard deviations up to 0.018, 0.044 and 0.019 for charge values of formamide nitrogen, carbon and oxygen atoms, respectively, when applying the CHELP algorithm and the HF/6-31G**//HF/6-31G* theory level. Merz, Jr. also addressed the charge value dependence on molecular orientation,[24] using the MNDO method and the Connolly surface algorithm implemented by Singh and Kollman.[10] Finally, Spackman and Sigfridsson *et al.* compared the CHELP, CHELPG and Connolly surface algorithms for deriving atomic charges.[25,96] They demonstrated that atomic charges strongly depend on the approach chosen for selecting MEP points. As a result the Geodesic, CHELP-BOW and CHELMO methods were proposed as alternatives to the original algorithms, but they were neither extensively used in empirical force field development nor for constructing force field topology database, so far.

The reports described above reveal that controlling the orientation of the molecule with respect to the grid on which the MEP is calculated is a crucial point for obtaining reproducible charge values. *Ab initio* calculations are usually performed for subjectively selected input molecular geometries. Every molecular orientation can potentially yield different charge values. Consequently, a structure has to be re-oriented to lead to reproducible RESP or ESP charges, and the method used for reorientation needs to be reported. By default (*i.e.* setup associated with the "Symmetry" keyword), the Gaussian series of QM programs re-orient an input structure by placing the center of nuclear charge at the origin, which is called "standard orientation".[69] Gaussian applies its re-orientation algorithm whenever the energy is calculated. On the contrary, the GAMESS-US and PC-GAMESS/Firefly programs can re-orient a molecule based on its principal axis ("COORD = CART" keyword).[70] However, this re-orientation is not carried out by default, and is calculated only once at the beginning of a calculation. Since the internal re-orientation algorithm available in each QM program is different, the molecular orientation of the final optimized geometry generated by them, and consequently the corresponding atomic charges, will be different. Moreover, because of rounding-off errors in procedures executed during geometry optimization different molecular orientations can be generated for a given minimum energy configuration when using a given QM program (see Fig. 2).[69,70,97]

A second factor which affects the atomic charge value is the accuracy of the optimized Cartesian coordinate set used to compute the MEP. It is directly controlled by the geometry optimization threshold criteria setup in the QM programs. The Gaussian and GAMESS-US/PC-GAMESS/Firefly programs use tight SCF convergence criterion in geometry optimization, but each of them calls different algorithms to optimize a structure. During geometry optimization, the Gaussian program uses four different threshold criteria, which are the maximum force, RMS force, maximum displacement and RMS displacement.[69] On the contrary, in the GAMESS-US and PC-GAMESS/Firefly programs the geometry optimization convergence process is controlled only by two criteria, which are the maximum and RMS gradient.[70] As a result of this difference the stationary point accuracy is highly related to the shape of the potential energy surface. Particularly, if the minimum well depth is flat, the default minimization threshold criteria available in GAMESS-US and PC-GAMESS/Firefly might not give a precise representation of this stationary point. To avoid this problem, we modified default software keywords in order to obtain more accurate minimum energy geometries required for a better charge reproducibility. To optimize a structure using the Gaussian program, the default value of the input "Opt" keyword was modified to reach a RMS force convergence tolerance of $1.10 \times 10^{-5}$ ("Opt = Tight" keyword, the other threshold criteria are set up automatically in relation to this one). For geometry optimization by the GAMESS-US and PC-GAMESS/Firefly programs, the default gradient convergence tolerance was more strongly decreased in order to take into account the absence of the displacement criteria ("OPTTOL = $1.10 \times 10^{-6}$" keyword). Eventually, the HONDO Rys polynomial code ("INTTYP = HONDO" keyword) could be selected for computing all molecular integrals to produce slightly more accurate integral values.[98,99] We determined that these keywords used in the three interfaced QM programs represent a good compromise between the charge reproducibility and the calculation time required for geometry optimization.

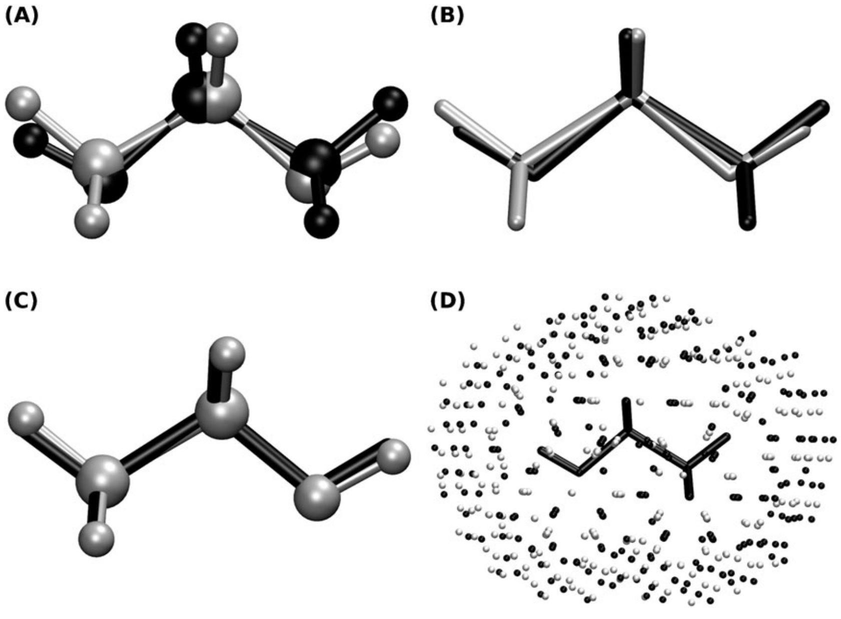

**Fig. 2** Representation of the molecular orientations of the optimized geometry of ethanol (*anti* conformation) generated using the Gaussian 98 and GAMESS-US program re-orientation algorithms. Molecular orientations and MEP points were displayed in the VMD program.[79] Only, the conformation of ethanol representing the lowest energy minimum was considered. (A) The two molecular orientations calculated using Gaussian 98 are shown ("ball and stick" representation). Orientation "-I-": gray color; orientation "-II-": black color (see Table 1 for the corresponding RESP charge values derived using each of these molecular orientations). (B) The two molecular orientations calculated using GAMESS-US are shown ("stick" representation). Orientation "-I′-": gray color; orientation "-II′-": black color (see Table 1 for the corresponding RESP charge values derived using each of these molecular orientations). (C) Comparison of the two molecular orientations "-I-" and "-I′-" generated using Gaussian and GAMESS-US (the molecular orientation "-I′-" is represented in black color for better characterizing the small differences between the two orientations). (D) The two molecular orientations of ethanol characterized in Fig. 2A were superimposed and displayed using the stick representation, while the corresponding MEP points are displayed using small spheres in black and gray.

RESP and ESP charge derivation strongly depends on the selected algorithm for MEP computation. When using HF and DFT methods for organic molecular systems, we observed that a criterion of $1.10^{-6}$ for single point energy convergence in the SCF is sufficient to achieve satisfactory results, and a tighter value such as $1.10^{-8}$ is not required to get reproducible charge values. This former value substantially decreases computation time required for calculating MEP using a large basis set in a multiple molecule approach. The atom radii as well as the algorithm defining the points at which the MEP is computed are other fundamental aspects of charge derivation.[10,15,16] The resultant charge values strongly depend on the input parameters defining the way the MEP is computed. For instance, the Connolly surface algorithm is affected by the scaling factor defining the vdW surface, the number of additional surfaces away from the first one, the distance between these surfaces as well as the density of points on them. Default values of 1.4 angstroms, 4 layers, 0.2 angstrom and 0.28 points per square au, respectively, are used in the originally developed algorithm that is incorporated in the Gaussian, GAMESS-US, and PC-GAMESS/Firefly programs.[69,70] These values are recommended by Kollman and co-workers, and were used in building the Amber force field topology database for proteins and nucleic acids.[59] To limit the dependency of charge values upon molecular orientation, the increase of the density of points in MEP computing has been proposed.[16] Unfortunately, this approach is not fully effective and dependence of charge values on molecular orientation is still observed.[25] Thus, Spackman implemented the geodesic algorithm in GAMESS-US, which produces a symmetric grid of points for a symmetric molecule, making the approach less rotationally dependent.[25] However, keeping the original approach defined for organic molecules without modifying the point selection scheme used for more than twenty years in Amber force fields ensures a rigorous compatibility of previous results with future developments. Thus, to alleviate the problem of charge reproducibility, the RBRA method described above is proposed. This new development raises a new question: which atoms should be used in the RBRA definition since the number of possible molecular re-orientations based on three atom selection exponentially increases with the molecule size? In principle, any set of atoms can be arbitrary selected in the RBRA procedure, and the actual version of the R.E.D. Tools does not provide any mode of selection for the atoms to be involved in this approach. To limit the number of possibilities heavy atoms might be chosen. These atoms could be either picked up randomly, selected from a set of atoms common to a family of molecules, or deduced from the symmetry of the molecules considered. The information related to chiral and prochiral atomic centers must also be provided since it describes the chemical group arrangement in the space, and thus can also affect atomic charge values.[67,100] If the molecular orientation(s) based on the specific choice of three atoms is correctly documented, then other researchers will be able to reproduce the atomic charges reported. Obviously, if another set of molecular orientations is used to compute the MEP a new set of charge values is derived. To limit the problem of charge value uncertainty observed for one molecular orientation, a multiple orientation RESP or ESP fit by analogy to the multiple conformation fit can be applied.[27,72] However, the main idea here is not to derive a converged set of charge values independently of the selected molecular orientations, but to generate reproducible charge values for the orientation(s) chosen by its authors.

Lastly, the fitting algorithm and restraints used in charge derivation as well as the process of equivalencing atomic charges carried out during or *a posteriori* to the fitting step affects charge values. This includes: equivalencing charges for chemically equivalent atoms or for atoms considered equivalent when using intra- and inter-molecular charge constraints as well as inter-molecular charge equivalencing.[28,29,59] A low value for the RRMS of the fit between the MEP computed by QM and the one produced by the fitted charges indicates the accuracy of the fit.[12,68] The R.E.D. program incorporates fitting methods, which minimize the number of charge constraints and differentiate charge equivalencing when one targets charge derivation for whole molecules and molecular fragments. Such a procedure ensures efficient charge fitting, yielding low RRMS values independently of the complexity of the case considered. Charge values generated using the R.E.D. Tools are reproduced with a maximal charge uncertainty of 0.0001 e independently of the initial structure or the QM program choice. Additionally, we did not observe any machine dependence of the derived atomic charges, at least at the level of accuracy applied here. Characteristic examples studied by the R.E.D. Tools are presented below.

**Impact of molecular orientation on RESP charges for organic structures.** The RESP atomic charge values obtained for three simple organic molecules (ethanol, dimethylsulfoxide and dimethylphosphate) using different re-orientation schemes and a single conformation are presented in Tables 1–3. The charges were derived using the internal re-orientation algorithm available in the Gaussian or GAMESS-US program or using the rigid body re-orientation algorithm implemented in the R.E.D. program.

Table 1 presents the RESP charge values for the ethanol molecule calculated in this work and compared with published values.[81] Only the lowest energy minimum (*anti* conformer) was considered. In this study, two different molecular orientations were generated for the final optimized geometry using each program internal re-orientation algorithm. These molecular orientations are represented and compared in Fig. 2A, 2B and C. The Maximum Charge Value Difference (next abbreviated MCVD in the text) obtained between the two molecular orientations generated by Gaussian or GAMESS-US were 0.022 e for the methyl carbon and 0.008 e for the methylene carbon. Using the two QM programs, the MCVD of 0.027 e for the methyl carbon are observed between the four molecular orientations. Finally, the MCVD between the values calculated in this work and those published by Fox and Kollman was more significant (0.045 e for the methyl carbon).[81] On the contrary, using the RBRA approach implemented into the R.E.D. program, highly reproducible charges are derived no matter which QM program or initial geometry representing the target minimum is chosen. A two-orientation RESP fit (orientation "-A-": methyl carbon, methylene carbon and oxygen atoms; orientation "-B-": oxygen, methylene carbon

**Table 1** RESP charge derivation and force field library building for the ethanol molecule (*anti* conformation) automatically carried out using the R.E.D. Tools. Charge values were derived from charge fitting and not arithmetic mean. The Connolly surface algorithm in MEP computation and the two-stage RESP fitting approach (hyperbolic restraints = $5.0 \times 10^{-4}$ and $1.0 \times 10^{-3}$) were considered as defined in the Amber force fields[28,29]

| QM Soft.[a]<br>Orient.nb | Gaussian<br>Ref.[b] | Gaussian<br>-I-[c] | Gaussian<br>-II-[c] | GAMESS<br>-I′-[c] | GAMESS<br>-II′-[c] | Both<br>RBRA-2[d] |
|---|---|---|---|---|---|---|
| Methyl-C | −.0990 | −.1009 | −.0788 | −.0767 | −.0737 | −.0859 |
| Methyl-H | .0345 | .0317 | .0261 | .0257 | .0244 | .0245 |
| Methylene-C | .3118 | .3533 | .3480 | .3483 | .3564 | .4132 |
| Methylene-H | −.0294 | −.0416 | −.0412 | −.0423 | −.0443 | −.0606 |
| O | −.6718 | −.6798 | −.6809 | −.6801 | −.6820 | −.6951 |
| H | .4143 | .4154 | .4157 | .4160 | .4148 | .4154 |
| MEP.pts[e] | [b] | 536 | 525 | 544 | 527 | 524, 529 |
| RRMS[f] | [b] | .141 | .143 | .140 | .142 | .145 |
| R.E.DD.B.[g] | na | W-7 | W-8 | W-6 | W-5 | W-9 |

[a] QM program used in geometry optimization and MEP computation: "Gaussian": results obtained using Gaussian 98. "GAMESS": results obtained using GAMESS-US. "Both": results obtained using Gaussian 98 or GAMESS-US lead to a single set of charge values when using the RBRA approach. [b] Results published in ref. 81: the molecular orientation, the RRMS value of the second stage RESP as well as the number of MEP points were not reported. [c] (i) Accurate geometry optimization, (ii) molecular orientation (named "-I-", "-II-", "-I′-" and "-II′-"; see Fig. 2A–C) of the optimized geometry controlled by the re-orientation algorithm implemented in Gaussian or GAMESS-US (or QMRA approach). [d] "RBRA-2": (i) Accurate geometry optimization, (ii) molecular orientation of the optimized geometry controlled by the RBRA approach implemented in the R.E.D. program; charge fitting step based on two molecular orientations defined using the following sets of three atoms: orientation "-A-": methyl carbon, methylene carbon and oxygen atoms; orientation "-B-": oxygen, methylene carbon and methyl carbon atoms. Identical atomic charge values (charge reproducibility ±0.0001 e) were obtained independently of the QM program used. [e] Number of MEP points generated for each molecular orientation. [f] The RRMS values calculated in the second stage RESP were taken directly from the outputs of the RESP program. [g] Data are available in R.E.DD.B. (the corresponding R.E.DD.B. code is provided); na = not available.

and methyl carbon atoms) can be performed to compute charges over these orientations. In order to characterize the charge dependence upon molecular orientation, the MEP generated for the different molecular orientations described in this example were further studied and compared. The number of MEP points generated for each molecular orientation are collected in Table 1. It shows that the number of MEP points slightly differ for each molecular orientation. Two molecular orientations (Fig. 2A) were superimposed and displayed in Fig. 2D for demonstrating the relative positions of the corresponding MEP points around the molecule. Fig. 2D demonstrates that MEP points for each orientation are located at different positions with respect to molecule atoms leading to different set of charges.

To demonstrate the generality of the conclusions reported for ethanol, other small organic structures were studied following the same approach. Table 2 compares the RESP charges calculated in this work with these published for dimethylsulfoxide.[81] In this new example, three different molecular orientations were generated using the Gaussian and GAMESS-US re-orientation algorithms for the same optimized geometry. Using Gaussian and GAMESS-US a single and two molecular orientations were generated, respectively. However, only two different sets of charge values were obtained for the dimethylsulfoxide molecule, since two different orientations yielded identical RESP charges. Using both QM programs, the MCVD for the dimethylsulfoxide carbon atom was only 0.007 e. More surprisingly, the difference in charge values was more substantial when compared with published results: the MCVD of 0.042 e was found between results obtained in this study and these reported by Fox and Kollman.[81] When a molecular re-orientation scheme is applied using the RBRA approach [the three atoms used for defining molecular orientation are as follows: carbon (*pro-S* configuration, *i.e.* "S" prochirality),[100] sulfur and oxygen atoms], the same set of atomic charges was derived no matter which of the QM programs is used.

Charge equivalencing of the atoms belonging to the methoxy groups is an important aspect in charge derivation for the *gauche*+, *gauche*+ conformer of dimethylphosphate. Two different strategies may be applied. In the first strategy, the two methoxy groups can be considered equivalent. Thus, when deriving charges for the Amber force fields the charges of the two methoxy oxygen atoms would be equivalenced in the first stage RESP while the charges of the two carbon and of the six hydrogen atoms would be equivalenced in the second stage of the fit. For GLYCAM charge equivalencing would be carried out in the single stage RESP. Such a scheme associated with a multiple conformational fit is suitable for studying the behavior of the molecule itself in MD simulations.[27] In the second approach the two methoxy groups are assumed to be not equivalent in order to take into account differences in molecular electronic environments. Cieplak *et al.* applied the latter approach to derive RESP charges for the Cornell *et al.* FFTopDB for DNA and RNA nucleotide fragments.[59] The first methoxy group in dimethylphosphate was used in the charge calculation of the deoxyribose and ribose O5′ atoms, while the second one was involved in the derivation of the sugar O3′ atomic charge, reflecting the different electronic environments of the O3′ and O5′ atoms in a nucleoside. However, in this approach, one faces the problem with defining which of the two methoxy groups needs to be used in the derivation of charges for the O3′ or O5′ atoms. The resulting ambiguity is demonstrated in Table 3, where the RESP charges of dimethylphosphate were derived using the internal re-orientation algorithm implemented in the QM programs as described above. Similar to the ethanol molecule case, two different orientations were generated by each QM program (molecular orientations are not displayed but are available in the corresponding R.E.DD.B. projects at the

**Table 2** RESP charge derivation and force field library building for the dimethylsulfoxide molecule automatically carried out using the R.E.D. Tools. Charge values were derived from charge fitting and not arithmetic mean. The Connolly surface algorithm in MEP computation and the two-stage RESP fitting approach (hyperbolic restraints = $5.0 \times 10^{-4}$ and $1.0 \times 10^{-3}$) were applied as defined in the Amber force fields[28,29]

| QM Soft[a]<br>Orient.nb | Gaussian<br>Ref.[b] | Gaussian<br>-I-[c] | GAMESS<br>-I′- and -II′-[d] | Both<br>RBRA-1[e] | Both<br>RBRA-2[f] |
|---|---|---|---|---|---|
| C | −.3244 | −.2826 | −.2897 | −.2867 | −.2808 |
| H | .1423 | .1263 | .1284 | .1272 | .1255 |
| S | .3155 | .3161 | .3177 | .3180 | .3163 |
| O | −.5205 | −.5089 | −.5085 | −.5080 | −.5078 |
| MEP.pts[g] | [b] | 624 | 617 | 627 | 627, 611 |
| RRMS[h] | [b] | .169 | .169 | .165 | .166 |
| R.E.DD.B.[i] | na | W-2 | W-1 | W-3 | W-4 |

[a] QM program used in geometry optimization and MEP computation: "Gaussian": results obtained using Gaussian 98. "GAMESS": results obtained using GAMESS-US. "Both": results obtained using Gaussian 98 or GAMESS-US lead to a single set of charge values when using the RBRA approach. [b] Results published in ref. 81: the molecular orientation, the RRMS value of the second stage RESP as well as the number of MEP points were not reported. [c] (i) Accurate geometry optimization, (ii) molecular orientation of the optimized geometry controlled by the re-orientation algorithm implemented in Gaussian (or QMRA approach): a single molecular orientation, named "-I-", was observed. [d] Similar procedure as in *c*, but using the re-orientation algorithm implemented in GAMESS-US. In this case, two molecular orientations, named "-I′-" and "-II′-", were observed: for each of them, a MEP with an identical number of MEP points and identical atomic charge values were obtained. [e] "RBRA-1": (i) Accurate geometry optimization, (ii) molecular orientation of the optimized geometry controlled by the RBRA approach implemented in the R.E.D. program was used, and charge derivation has been performed for the optimized geometry for which the molecular orientation was defined by the following set of three atoms: carbon (*pro-S* configuration), sulfur and oxygen atoms. Identical charge values (charge reproducibility ±0.0001 e) were obtained independently of the QM program used. [f] "RBRA-2": Similar procedure as in *e*, but charge fitting step based on two molecular orientations defined using the following set of three atoms: orientation "-A-": carbon (*pro-S* configuration), sulfur and oxygen atoms; orientation "-B-": oxygen, sulfur and carbon (*pro-S* configuration) atoms. Identical charge values (charge reproducibility ±0.0001 e) were obtained independently of the QM program used. [g] Number of MEP points generated for each molecular orientation. [h] The RRMS values calculated in the second stage RESP were taken directly from the outputs of the RESP program. [i] Data are available in R.E.DD.B. (the corresponding R.E.DD.B. code is provided); na = not available.

**Table 3** RESP charge derivation and force field library building for the dimethylphosphate molecule (*gauche*+, *gauche*+ conformation) automatically carried out using the R.E.D. Tools. Charge values were derived from charge fitting and not arithmetic mean. The Connolly surface algorithm in MEP computation and the two-stage RESP fitting approach (hyperbolic restraints = $5.0 \times 10^{-4}$ and $1.0 \times 10^{-3}$) were considered as defined in the Amber force fields[28,29]

| QM Soft[a]<br>Orient.nb | Gaussian<br>-I-[b] | Gaussian<br>-II-[b] | GAMESS<br>-I′-[b] | GAMESS<br>-II′-[b] | Both<br>RBRA-2[c] |
|---|---|---|---|---|---|
| CM1[d] | .0852 | .0691 | .0849 | .0697 | .1067 |
| H11[e] | .0260 | .0300 | .0260 | .0298 | .0198 |
| H12[e] | =H11 | =H11 | =H11 | =H11 | =H11 |
| H13[e] | =H11 | =H11 | =H11 | =H11 | =H11 |
| OM1[f] | −.4774 | −.4746 | −.4769 | −.4744 | −.4788 |
| P | 1.2200 | 1.2200 | 1.2205 | 1.2205 | 1.2174 |
| O1=O2[g] | −.7952 | −.7952 | −.7956 | −.7956 | −.7961 |
| OM2[f] | −.4746 | −.4774 | −.4744 | −.4769 | =OM1 |
| CM2[d] | .0691 | .0852 | .0697 | .0849 | =CM1 |
| H21[e] | .0300 | .0260 | .0298 | .0260 | =H11 |
| H22[e] | =H21 | =H21 | =H21 | =H21 | =H11 |
| H23[e] | =H21 | =H21 | =H21 | =H21 | =H11 |
| MEP.pts[h] | 759 | 759 | 758 | 758 | 756, 756 |
| RRMS[i] | .017 | .017 | .017 | .017 | .017 |
| R.E.DD.B.[j] | W-14 | W-13 | W-12 | W-11 | ns |

[a] QM program used in geometry optimization and MEP computation: "Gaussian": results obtained using Gaussian 98. "GAMESS": results obtained using GAMESS-US. "Both": results obtained using Gaussian 98 or GAMESS-US lead to a single set of charge values when using the RBRA approach. [b] (i) Accurate geometry optimization, (ii) molecular orientation of the optimized geometry controlled by the re-orientation algorithm implemented in Gaussian or GAMESS-US (each QM program generates two different molecular orientations, named "-I-", "-II-", "-I′-" and "-II′-", or QMRA approach). [c] "RBRA-2": (i) Accurate geometry optimization, (ii) molecular orientation of the optimized geometry controlled by the RBRA approach implemented within the R.E.D. program was used. Charge fitting step based on two molecular orientations defined using the following set of three atoms: orientation "-A-": methyl carbon one defined from the input atom order, phosphorus and methyl carbon two atoms; orientation "-B-": methyl carbon two, phosphorus and methyl carbon one atoms. [d] The charges of the two carbon atoms were not equivalenced in the second stage RESP. [e] The charges of the hydrogen atoms in each methyl group were made equivalent in the second stage RESP. [f] The charges of the two oxygen atoms in each methoxy group were not equivalenced in the first stage RESP and their charges were frozen during the second RESP stage. [g] The charges of the two phosphoryl oxygens were equivalenced in the first stage RESP only when using the QMRA approach, and the corresponding charges were frozen during the second RESP stage. [h] Number of MEP points generated for each molecular orientation. [i] The RRMS values calculated in the two RESP stages were taken directly from the outputs of the RESP program. [j] Data are available in R.E.DD.B. (the corresponding R.E.DD.B. code is provided); ns = not submitted.

q4md-forcefieldtools.org website). Interestingly, the carbon charge value of the first methoxy group in the first orientation corresponds exactly to the carbon charge value of the second methoxy group in the second orientation independently of the QM program used. Similar conclusions are observed concerning the two methoxy oxygen atoms. The MCVD calculated for the two dimethylphosphate carbon atoms were 0.016 e and 0.015 e between each molecular orientation generated by the Gaussian and GAMESS-US programs, respectively. The charge uncertainty caused by the different orientations of the two methoxy groups accounts for approximately 20% of the carbon charge value and closely corresponds to the difference in the carbon charge values obtained in the two stages RESP (data not shown). Although a MCVD of 0.016 e may not affect MD simulations, it is important to underline that it is rigorously explained as a rotational effect of the two methoxy groups in the *gauche*+, *gauche*+ conformation. To further substantiate this conclusion, the RBRA approach was applied to derive RESP charge for dimethylphosphate using two well-chosen orientations. The latter are characterized by the two following sets of three atoms: orientation "-A-": methyl carbon one defined from the input atom order, phosphorus and methyl carbon two atoms; orientation "-B-": methyl carbon two, phosphorus and methyl carbon one atoms. Considering that dimethylphosphate (*gauche*+, *gauche*+ conformation) presents a *C2* axis of symmetry, key features are that the central phosphorus atom and two symmetric atoms are selected to define the two considered orientations. Here, the two corresponding sets of three atoms are defined as follows: two symmetric atoms define the first and third atoms, while the central atom always specify the second one. Involving the corresponding MEP in two-orientation charge derivation leads to identical charge values for symmetric atoms (*i.e.* for the two methyl carbons, the methoxy oxygens and the phosphoryl oxygens) without the need of introducing charge equivalencing during the fitting step (see Table 3). In this example, the effect induced by the first orientation is cancelled out by the second one. This demonstrates the need for taking into account the molecular symmetry in the RBRA approach, or the necessity of using a symmetric grid of points in MEP computation as implemented by Spackman.[25]

**Example of charge derivation for a dipeptide structure.** In this subsection, a study of the effect of molecular orientation on the atomic charge values for a dipeptide is described. This effect is studied using the RBRA approach, and is applied to determine the charge values for the *C5* conformation of the alanine dipeptide or *N*-acetyl-L-alanine-*N*′-methylamide (constituted of the ACE-ALA-NME residues). Twelve different molecular orientations were generated for this conformation, and for each molecular orientation a set of RESP charges has been derived. Among the data generated, only the charge values of the four orientations presenting the highest charge variability are reported. The latter are compared with the charge values generated in the corresponding four molecular orientation RESP charge fit, and are presented in Table 4. The atomic charges were obtained using the Gaussian and the GAMESS-US programs. The MCVD for each heavy atom as well as the number of MEP points and RRMS value of each fit are also reported. As an example, the MCVD of 0.040 e (*i.e.* 7% of the charge value) and of 0.052 e (*i.e.* 10% of the charge value) are observed for the carbonyl carbon and nitrogen atoms of the alanine residue, respectively. This demonstrates that charges for all heavy atoms can exhibit quite a substantial dependence on molecular orientation, and this problem is not only restricted to buried centers such as methylene or methyl carbons.[28,29] This is also well illustrated by the MCVD of the α-carbon in the alanine residue (0.038 e) which is larger compared to the charge value itself (0.0287 e) observed for one of the orientations. Finally, no correlation between the differences of MEP point number and the differences of charge values for the molecular orientations considered were observed. Indeed, the largest MCVD (0.070 e, *i.e.* 25% of the charge value) is found for the methyl carbon of the ACE blocking group calculated between two molecular orientations which differ only by a single MEP point. This makes the charge orientation dependence unpredictable. To alleviate this problem and to make charges reproducible the molecular orientations used for the optimized geometry and applied in the charge derivation have to be reported. For example, the following sets of three atoms have been used for defining the four selected molecular orientations involved in charge derivation: (i) orientation “-A-”: ALA carbonyl-oxygen, NME nitrogen and NME carbon atoms; (ii) orientation “-B-”: ACE carbonyl-carbon, ALA β-carbon and NME carbon atoms; (iii) orientation “-C-”: ACE carbonyl-carbon, ACE carbonyl-oxygen and ALA α-carbon atoms; (iv) orientation “-D-”: ACE methyl-carbon, ACE carbonyl-carbon and ACE carbonyl-oxygen atoms. The atomic charges were derived using four different orientations to take into account dependence of charges on molecular orientation.

**Table 4** RESP charge derivation and force field library building for the *C5* conformation of the alanine dipeptide automatically carried out using the R.E.D. Tools. Charge values were derived from charge fitting and not arithmetic mean. MCDV: maximum charge value differences are not automatically calculated by the actual version of the R.E.D. Tools. The Connolly surface algorithm in MEP computation and the two-stage RESP fitting approach (hyperbolic restraints = $5.0 \times 10^{-4}$ and $1.0 \times 10^{-3}$) were considered as defined in the Amber force fields[28,29]

| Orient.nb | -A-[a] | -B-[a] | -C-[a] | -D-[a] | RBRA-4[b] | MCVD |
|---|---|---|---|---|---|---|
| ACE-C1[c] | −.3223 | −.2822 | −.3526 | −.3448 | −.3261 | .070 |
| ACE-H11[c] | .0952 | .0838 | .1027 | .0995 | .0954 | — |
| ACE-H12[c] | =H11 | =H11 | =H11 | =H11 | =H11 | — |
| ACE-H13[c] | =H11 | =H11 | =H11 | =H11 | =H11 | — |
| ACE-C[d] | .6831 | .6949 | .6991 | .7213 | .7005 | .038 |
| ACE-O[d] | −.5950 | −.5987 | −.5990 | −.6040 | −.5994 | .009 |
| ALA-N[d] | −.5007 | −.5343 | −.5099 | −.5523 | −.5254 | .052 |
| ALA-H[d] | .2872 | .3001 | .2860 | .3017 | .2941 | — |
| ALA-CA[d] | .0506 | .0287 | .0629 | .0665 | .0518 | .038 |
| ALA-HA[d] | .0694 | .0769 | .0644 | .0659 | .0692 | — |
| ALA-CB[c] | −.1445 | −.1286 | −.1198 | −.1199 | −.1272 | .025 |
| ALA-HB1[c] | .0561 | .0524 | .0484 | .0482 | .0510 | — |
| ALA-HB2[c] | =HB1 | =HB1 | =HB1 | =HB1 | =HB1 | — |
| ALA-HB3[c] | =HB1 | =HB1 | =HB1 | =HB1 | =HB1 | — |
| ALA-C[d] | .5394 | .5775 | .5379 | .5667 | .5571 | .040 |
| ALA-O[d] | −.5290 | −.5433 | −.5293 | −.5398 | −.5360 | .014 |
| NME-N[d] | −.4103 | −.4093 | −.4221 | −.4277 | −.4184 | .018 |
| NME-H[d] | .3143 | .3090 | .3143 | .3146 | .3128 | — |
| NME-C2[c] | −.3571 | −.3476 | −.3040 | −.3388 | −.3341 | .053 |
| NME-H21[c] | .1536 | .1495 | .1396 | .1492 | .1472 | — |
| NME-H22[c] | =H21 | =H21 | =H21 | =H21 | =H21 | — |
| NME-H23[c] | =H21 | =H21 | =H21 | =H21 | =H21 | — |
| MEP.pts[e] | 975 | 991 | 992 | 1017 | 4 values[f] | — |
| RRMS[g] | .104 | .111 | .109 | .108 | .108 | — |
| R.E.DD.B.[h] | W-50 | W-51 | W-52 | W-53 | W-54 | — |

[a] QM program used in geometry optimization and MEP computation: both the Gaussian 98 and GAMESS-US programs were employed. (i) Accurate geometry optimization, (ii) molecular orientation of the optimized geometry controlled by the RBRA approach implemented in the R.E.D. program. Among the twelve molecular orientations studied, the four ones (named “-A-”, “-B-”, “-C-”, and “-D-”), which demonstrate the largest charge value discrepancies were selected. Orientation “-A-” is defined by: ALA carbonyl-oxygen, NME nitrogen and NME carbon atoms; orientation “-B-”: ACE carbonyl-carbon, ALA β-carbon and NME carbon atoms; orientation “-C-”: ACE carbonyl-carbon, ACE carbonyl-oxygen and ALA α-carbon atoms; orientation “-D-”: ACE methyl-carbon, ACE carbonyl-carbon and ACE carbonyl-oxygen atoms. Each of these orientations was used in a single molecular orientation RESP fit. [b] “RBRA-4”: Four molecular orientation RESP fit involving the orientations defined in [a]. The charges reported were obtained using either the Gaussian 98 or the GAMESS-US program for both geometry optimization and MEP computation (charge reproducibility ±0.0001 e). [c] The charge values of the methyl carbon were re-optimized and the methyl hydrogen charges were made equivalent in the second stage RESP. [d] The charges of these atoms were optimized in the first stage RESP and their charges were frozen in the second stage RESP. [e] Number of MEP point generated for the given molecular orientations. [f] 975, 991, 992 and 1017. [g] The RRMS values calculated in the second stage RESP were taken directly from the outputs of the RESP program. [h] Data are available in R.E.DD.B. (the corresponding R.E.DD.B. code is provided).

Similar conclusions were obtained for α-aminoisobutyric acid in the R.E.D. Tools tutorial.[64,92,93]

As demonstrated in these examples, RESP charge MCDV of 0.07 e for buried carbons and 0.05 e for heteroatoms remain substantial values and cannot be ignored. Using *N*-acetyl-L-Alanine-*N′*-methylamide, non-negligible ESP charge MCDV were observed as well: the largest MCDV of around 0.10 or 0.03 e for the α-carbon was calculated when using the Connolly surface or CHELPG algorithm in MEP computation, respectively (Table S4 of the supplementary material†).[10,15,16] As reported by Spackman,[25] when using an algorithm such as CHELPG, which features high density of points, the charge dependence on molecular orientation is not rigorously solved. The charge uncertainty for an atom charge is not predictable and its impact remains undefined. This work shows that both RESP and ESP charges exhibit charge dependency on molecular orientation. Charge uncertainty is explained by a molecular orientation effect and is not related to errors caused by the finite precision of computations involving floating-point or integer values.[101] These examples clearly demonstrate the necessity of reporting information about the molecular orientation used in RESP or ESP charge derivation to rigorously control atomic charge values. Single or multiple re-orientation charge fit can be applied to fully control molecular orientation and to generate highly reproducible RESP and ESP charges independently of the algorithm involved in MEP computation and the density of points used.[72] Finally, it is important to emphasize that the use of multiple-orientation charge fitting presents as a single objective the reproducibility of charge values and not the generation of a converged set of charge values for an assumed molecular conformation.

**Limitations of the strategy presented.** Although a charge reproducibility of 0.0001 e has been attained for the bio-organic molecular systems reported in this work, the approach described presents several limitations. First, it is expected that if two different QM programs are used in geometry optimization and MEP calculations, they should use the same standard basis set to yield identical results. This condition is not always satisfied. For example, the GAMESS-US program uses different scaling factors for the Gaussian exponents in the STO-3G basis set for the second row elements compared to the Gaussian program.[102] Thus, different ESP charges can be derived for chloroform and methanethiol (data not shown) when using this basis set.[10,35,36] A similar problem can also be encountered, if the DFT B3LYP theory is used for geometry optimization. In GAMESS-US, the B3LYP method is based on the VWN5 correlation functional, while B3LYP in Gaussian uses the VWN3.[103,104] PC-GAMESS/Firefly has both methods implemented.

Finally, when a well-defined molecular orientation is used to derive RESP or ESP charges, the reproducibility of the charges directly depends on the accuracy of the calculated energy minimum. As mentioned above, to alleviate this problem, default Gaussian and GAMESS-US minimization threshold criteria have been modified in the R.E.D. source code. However, even using the keywords reported in the "Methods" section, a difference by one MEP point can be found between two identical orientations of a structure in some rare occasions. A single MEP point difference can slightly affect atomic charge values for a given structure and the MCVD of up to 0.0005 e might be observed in this case. In order to strictly attain a charge reproducibility of 0.0001 e and to limit rounding-off errors,[97] the "Opt = VTight" keyword in Gaussian and the "OPTTOL = $1.10 \times 10^{-7}$" keyword in GAMESS-US must be used.[69,70] However, the use of such keywords implies substantial increase in required computer CPU time for the sake of minor gain in charge reproducibility.

## Multiple orientation, multiple conformation, multiple molecule charge derivation and FFTopDB building

**General considerations.** For a given molecule the atomic charges derived from MEP strongly depend on the selected conformation.[21,22,26] Various approaches have been proposed to minimize this dependence. In the Amber additive force field model the RESP procedure, in which charges are restrained by a hyperbolic function and multiple conformation charge fitting approach have been extensively employed for generating a fixed and non-polarizable charge set suitable for MD simulations.[27–29] In this approach, selection of the conformations included in charge derivation is a key aspect in constructing an accurate force field. Usually the conformations corresponding to low energy minima are frequently selected. However, the problem of conformation selection is strongly related to the size of the molecular system under study. For large molecules, characterizing representative conformations might become a complex task, and using a systematic conformational search might not lead to satisfactory results. Among common encountered pitfalls are low energy minima containing internal hydrogen bonds. Such geometries are generally excluded from charge derivation since they lead to over-polarization effect.[59,105,106] In order to avoid the formation of artificial hydrogen bonds resulting from the computational conditions or from the chosen model itself, the use of geometrical constraints during the geometry optimization step is widely employed. Being limited by the size of the molecular system in QM computation, large structures are built from smaller fragments or residues, for which conformational properties are fully characterized. The different elementary building blocks constitute a FFTopDB which can be used for constructing any type of biopolymers such as proteins, nucleic acids and glycoconjugates. As an example, the charges of the Amber FFTopDB for proteins were derived based on capped amino acids represented by two conformations, whose $\varphi$, $\phi$ dihedral angles correspond to the values observed in the α-helix and β-sheet secondary structures.[50,59] Elementary nucleosides characterized by the canonical *C2′-endo* and *C3′-endo* conformations for the deoxyribose and ribose were used for constructing the Amber FFTopDB for standard nucleic acids, respectively.[59] Following a different approach, charges available in the GLYCAM force field were calculated over a large set of snapshots observed during MD simulations.[47,107]

The Amber, CHARMM and GLYCAM FFTopDB are collections of individual force field libraries, each containing appropriate atom names, atomic charges, force field atom types and the topology of a small molecule or a molecular fragment.[47,59,108] Several basic procedures are used for deriving

RESP charges for molecular fragments. One of them is the application of charge constraint(s) within a molecule or between two molecules to force a group of atoms to attain a specific charge value. Another one is to group several atoms together and constrain the sum of their charges to a zero value. This is usually applied to a group of atoms that are removed after the fit in order to create a molecular fragment from which a larger system can be built. In this case the total charge of a molecular fragment or the sum of the total charges of two complementary fragments generally take an integer value. Such charge constraint approach ensures a strict compatibility between the different force field libraries constituting a FFTopDB and eliminates subjective manual charge manipulation carried out *a posteriori* to the fit. The value of the assumed charge constraint and the constitution of the group of atoms to be constrained are strongly related. Consequently, a "well-adapted" value has to be used for a charge constraint. The closer is the constraint value for a considered group of atoms to the total charge value of this chemical group without constraint, the smaller is the impact of the constraint on the RRMS value of the fitting step, and the better this constraint can be considered. Direct application of this recipe is charge derivation and force field library building for the central and terminal fragments of an amino acid or a nucleotide belonging to the Amber FFTopDB for proteins or nucleic acids, respectively.[59] Similar strategies can also be applied to any type of biopolymers.[109] Finally, addition of extra points and lone pairs during the fitting step,[87] and exclusion of hydrogens during or *a posteriori* to the fitting step are other features used in charge derivation.[73] These approaches can be used for generating all-atom and united-carbon force field libraries.[38,40,59,73] Charge derivation and force field library building for amino acid and nucleotide fragments as well as the construction of all-atom and united-carbon FFTopDB are discussed below.

**A new dipeptide and its central, *N*-terminal and *C*-terminal fragments.** The capabilities of the R.E.D. Tools are illustrated by deriving charge values and building force field libraries for the central, H3N(+)-terminal and COO(−)-terminal molecular fragments of the unusual *O*-methyl-L-tyrosine amino acid. Appropriate calculations are performed using the *N*-acetyl-*O*-methyl-L-tyrosine-*N′*-methylamide dipeptide (constituted of the ACE-TYM-NME residues). *O*-methyl-L-tyrosine is a new residue added to the genetic code of *E. coli.* by Schultz and coworkers.[94] Charge derivation and force field library building are summarized in Fig. 3. In the absence of any information about the conformational preferences of this amino acid the two geometries corresponding to α-helix and/or β-sheet were selected for computing MEP. This approach was first proposed by Cieplak *et al.* and then used more recently by Duan *et al.*[50,59] Heavy atoms of the backbone were used to define dipeptide multiple orientations in space to assure highly reproducible charge values. The central fragment of *O*-methyl-L-tyrosine (the corresponding fragment or residue name is defined as "TYM") is derived by imposing two intra-molecular charge constraints for zeroing sum of charges on the *N*-acetyl and *N′*-methylamide groups. The *N*-terminal fragment of *O*-methyl-L-tyrosine ("NTYM" fragment name) is constructed by combining the methylammonium and *N*-acetyl-*O*-methyl-L-tyrosine-*N′*-methylamide molecules in charge derivation. Fragment NTYM is obtained by setting two charge constraints to a value of zero during the fitting step, which are: (i) an inter-molecular charge constraint between the methylammonium methyl group and the dipeptide NH-acetyl group, and (ii) an intra-molecular charge constraint for the dipeptide *N′*-methylamide group. The *C*-terminal fragment of *O*-methyl-L-tyrosine ("CTYM" fragment name) is built following a similar approach using acetate and *N*-acetyl-*O*-methyl-L-tyrosine-*N′*-methylamide. Fragment CTYM is obtained by constraining the total charge of two groups of atoms to zero

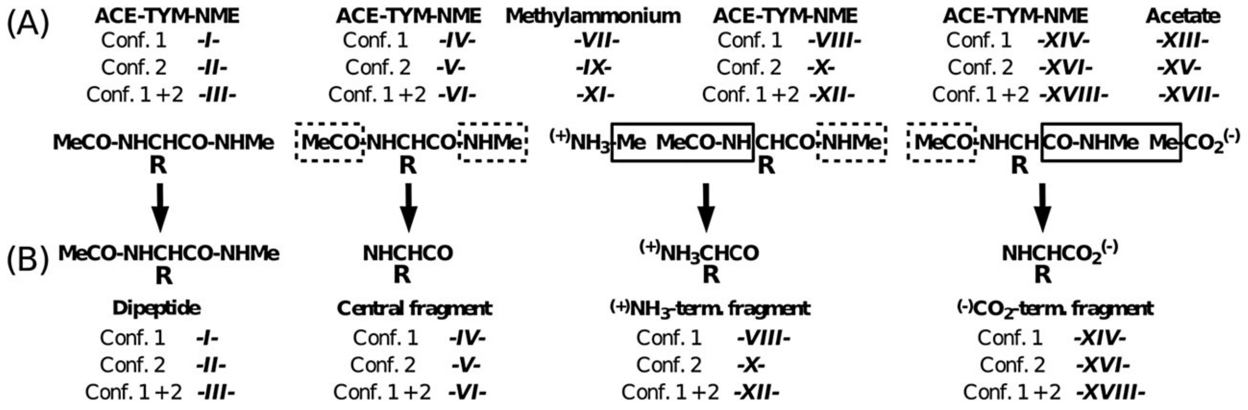

**Fig. 3** Multiple orientation, multiple conformation and multiple molecule charge derivation and force field library building for the *N*-acetyl-*O*-methyl-L-tyrosine-*N′*-methylamide dipeptide and the central, H3N(+)-terminal and COO(−)-terminal molecular fragments of the *O*-methyl-L-tyrosine amino acid automatically carried out using the R.E.D. Tools (see Table 5).[94] (A) Description of the molecules and conformations involved in the procedure (bold and italic roman numeral). Conf. 1: α-helix conformation ($\varphi$ = −72.4, $\phi$ = −34.8) of *N*-acetyl-*O*-methyl-L-tyrosine-*N′*-methylamide involved in a single conformation charge fit; Conf. 2: β-sheet conformation ($\varphi$ = −119.2, $\phi$ = 138.7) of *N*-acetyl-*O*-methyl-L-tyrosine-*N′*-methylamide involved in a single conformation charge fit; Conf. 1 + 2: α-helix and β-sheet conformations of *N*-acetyl-*O*-methyl-L-tyrosine-*N′*-methylamide involved in a two conformation charge fit. R group: side chain of *O*-methyl-L-tyrosine: CH2–C6H4–OCH3. A dash box defines an intra-molecular charge constraint within the dipeptide; a plain box defines an inter-molecular charge constraint between the dipeptide and methylammonium or acetate. (B) Selection of force field libraries generated in the Tripos MOL2 file format.

**Table 5** Multiple orientation, multiple conformation and multiple molecule charge derivation and force field library building for the *N*-acetyl-*O*-methyl-L-tyrosine-*N*′-methylamide dipeptide and the central, H3N(+)-terminal and COO(−)-terminal molecular fragments of the *O*-methyl-L-tyrosine amino acid automatically carried out using the R.E.D. Tools (see Fig. 3).[94] Charge values were derived from charge fitting and not arithmetic mean. (i) Accurate geometry optimization, (ii) molecular orientation of each optimized geometry controlled by the RBRA approach implemented in the R.E.D. program. When not specified MEP computation was carried using the HF/6-31G* theory level, and all-atom force field libraries were generated

| R.E.DD.B.[a] | | F-73 | F-74 | F-75 | F-76 | F-77 | F-78 | F-79 | F-80 |
|---|---|---|---|---|---|---|---|---|---|
| Surf[b] | | CS | CS | CS | CS | CS | CS | CS | CG |
| Fit[c] | | 1 E st | 2 R st | 2 R st | 2 R st | 2 R st | 2 R st | 2 R st | 2 R st |
| RRMS[d] | (i) | 0.044 | 0.049 | 0.048 | na | na | 0.048 | 0.048 | 0.077 |
| | (ii) | 0.041 | 0.046 | 0.040 | na | na | 0.048 | 0.042 | 0.074 |

[a] Data are available in R.E.DD.B. (the corresponding R.E.DD.B. code is provided). "F-78" and "F-79": MEP computations were carried out using the DFT method, the B3LYP exchange and correlation functionals, the IEFPCM continuum solvent model ($e = 4$) to mimic organic solvent environment, and the cc-pVTZ basis set as reported in the Duan *et al.* force field.[50] When an united force field library model is chosen, the selected hydrogen(s) are removed from each force field library. "F-75" and "F-79": united-methylene and methyl carbon force field libraries were generated by setting to a value of zero the charge value of each selected hydrogen atom during the charge fitting step; "F-76": united-carbon force field libraries were generated by summing the charge values of each hydrogen within the charge value of the carbon they are bound to *a posteriori* to the charge fitting step; "F-77": united-methylene and methyl carbon force field libraries were generated by summing the charge values of each hydrogen within the charge value of the carbon they are bound to *a posteriori* to the charge fitting step. [b] Surface algorithm used in MEP computation: CS = Connolly surface algorithm; CG = CHELPG algorithm. [c] Charge fitting step: 1 E st = one stage ESP; 2 R st = two-stage RESP (hyperbolic restraints = $5.0 \times 10^{-4}$ and $1.0 \times 10^{-3}$) as defined in the Amber force fields.[28,29] [d] RRMS values calculated in the single charge fitting step (1 E st) or in the second charge fitting step (2 R st) for each multiple orientation, multiple conformation and multiple molecule charge fit (number of structures "nmol" involved = 76). Two RRMS values are reported (i) in the presence and (ii) in the absence of intra- and inter-molecular charge constraints required for building the amino acid fragments. Values were taken directly from the outputs of the RESP program. na = not available: the RRMS values cannot be determined when the charge values of the hydrogen atoms are summed within the carbon charges in the united-atom model.

value. One group involves the acetate methyl group and the dipeptide CO-*N*′-methylamide group, and the other one the dipeptide *N*-acetyl group. As a result, the total charges of the central, *N*- and *C*-terminal fragments of *O*-methyl-L-tyrosine take integer values, and are compatible with other molecular fragments belonging to the Amber FFTopDB for proteins. Force field libraries for the TYM, NTYM, and CTYM molecular fragments are built by removing the atoms involved in the charge constraints from the molecules involved in charge derivation, and by adding a new atom connectivity between the methylammonium nitrogen atom and the dipeptide α-carbon for NTYM, and between the acetate carboxylate carbon and the dipeptide α-carbon for CTYM. In the present example, by combining eighteen P2N input files described in Fig. 3, the atomic charges required for force field libraries are derived for the dipeptide molecule and its different molecular fragments in a single R.E.D. execution. This approach, involving multiple orientations, multiple conformations and multiple molecules create a computational platform in which various factors affecting charge and RRMS values can be studied. These include: (i) the differences observed between single conformation charge fit *versus* multiple conformation charge fit for the dipeptide itself and its different fragments, (ii) the impact of the use of intra-molecular charge constraints when building the TYM central fragment, and (iii) the effect of intra- and inter-molecular charge constraints when building the NTYM and CTYM terminal fragments. In addition, it is possible to compare different charge models and/or force field library types, such as all-atoms or united-carbons obtained from the common set of the P2N input files. Charge values are not reported here due to the large amount of data generated, but are available in R.E.DD.B. for downloading. This database is constructed in the form of an ensemble of projects, each of them containing input and output files with detailed descriptions of the computational conditions. As an example, projects related to the *O*-methyl-L-tyrosine amino acid are listed in Table S3 of the supplementary material.† Global entities characterizing the charge fit, such as the RRMS values calculated for each multiple molecule fit in the presence and in the absence of intra- and inter-molecular charge constraints, are reported in Table 5. The small RRMS values observed and the relative weak impact of the charge constraints on these RRMS values are strong arguments demonstrating the effectiveness of the approaches presented. The latter are directly applicable to any type of dipeptide or set of dipeptides. The R.E.D. Tools tutorial describes charge derivation and force field library building for the molecular fragments of alpha-aminoisobutyric acid.[64,92,93] This second example is also available in R.E.DD.B. for downloading (the corresponding R.E.DD.B. codes are listed in the Table S4 of the supplementary material†).

**Examples of nucleic acid FFTopDB.** In the Amber force fields, the central, 5′-terminal and 3′-terminal fragments of a nucleotide are simultaneously generated in a single charge derivation procedure.[59] The strategy for building such nucleotide fragments is summarized in Fig. 4A. It involves application of two inter-molecular charge constraints between the methyl groups of dimethylphosphate and the HO5′ and HO3′ hydroxyl groups of the nucleoside of interest during the fitting step. As a result the total charge value of the central fragment of a nucleotide and the sum of the total charges of the two terminal fragments always equals the total charge of dimethylphosphate. Inter-molecular charge equivalencing between the charge values of the deoxyribose or ribose atoms (excluding the C1′ and H1′ atoms) during the fitting step ensures identical total charge values for the central, 5′- and 3′-terminal fragments of the nucleosides involved in the charge derivation, and a

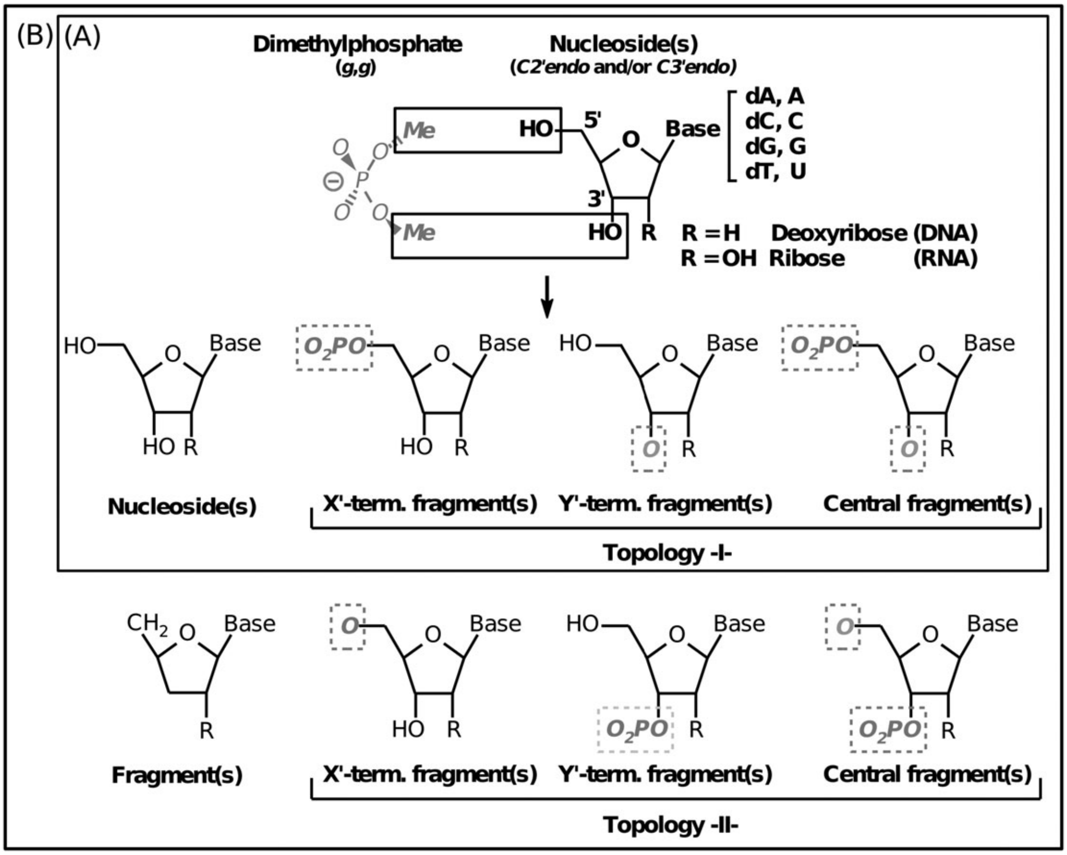

**Fig. 4** Building of the FFTopDB for standard nucleic acids based on multiple orientation, multiple conformation and multiple molecule charge derivation automatically carried out using the R.E.D. Tools (see Table 6). (A) Force field libraries available in the current Amber FFTopDB corresponding only to the topology -I-:[59] charge derivation involving dimethylphosphate (*gauche*+, *gauche*+ conformation) and the four DNA nucleosides (dA, dC, dG and dT; *C2′-endo* conformation) or the four RNA (A, C, G and U; *C3′-endo* conformation) nucleosides was carried out in two independent approaches. A plain box defines an inter-molecular charge constraint between the methyl groups of dimethylphosphate and the HO5′ and HO3′ hydroxyl groups of a selected nucleoside used during the fitting step. Additional inter-molecular charge equivalencing between the charge values of the deoxyribose or ribose atoms (but the C1′ and H1′ atoms) of the different nucleosides were also used during the fitting step. Topology -I-: the phosphate group taken from dimethylphosphate (gray color within a dash box) is arbitrarily connected the 5′ nucleoside carbon, and X′ = 3′ and Y′ = 5′ in the Amber FFTopDB. (B) Construction of a new FFTopDB corresponding to the topologies -I- and -II- using the R.E.D. Tools: charge derivation involving dimethylphosphate (*gauche*+, *gauche*+ conformation) and the eight DNA (*C2′-endo* and *C3′-endo* conformations) and RNA (*C3′-endo* conformation) nucleosides were carried out in a single R.E.D. execution. Topology -II-: the phosphate (gray color within a dash box) is arbitrarily placed at the 3′ position. Topologies -I- and -II- are both generated in a single R.E.D. execution.

compatibility between the different molecular fragments constituting the FFTopDB. This compatibility is assured by combining all the considered nucleosides in a single multiple molecule charge derivation procedure. The complexity of the approach and the manual construction of correct RESP inputs are significant limitations in this type of work. In this context, Amber FFTopDB for standard DNA and RNA nucleic acids were constructed in two independent procedures, one handling the deoxyribonucleosides and the other one the ribonucleosides. The two resulting FFTopDB were made compatible using subjective manual adjustment of charges performed *a posteriori* to the fit, and assuming that the phosphate group of dimethylphosphate is connected to the 5′-carbon in the nucleotide fragments.[59] The strategy based on multiple orientations, multiple conformations and multiple molecules developed in the R.E.D. Tools is particularly well suited for building the complex FFTopDB for nucleic acids without any need for manual charge value adjustments. The approach developed here is presented in Fig. 4B. The specific features of this methodology include: (i) no limitation for the number of nucleosides involved in charge derivation, (ii) generating two possible topologies (named as topologies -I- and -II-) where the phosphate group is connected either to the 5′- or 3′-nucleotide carbon, and (iii) a more general Y′ and X′ terminology is used for terminal fragments in order to build standard as well as non-standard or chemically engineered nucleic acids with different terminal groups instead of the regular H05′ and HO3′ ones. To illustrate the use of the R.E.D. Tools the construction of different FFTopDB for the standard nucleic acids has been performed. Multiple orientation, multiple conformation and multiple molecule charge derivation and force field library buildings have been carried out using dimethylphosphate (*gauche*+, *gauche*+ conformation) and various sets of standard nucleosides (four deoxyribonucleosides and/or four ribonucleosides in the *C2′-endo* and/or *C3′-endo* conformations). We performed calculations using the Connolly surface or CHELPG algorithm and various approaches for the fitting step. These include: one stage ESP fit, one stage RESP fit and two stage RESP fit. Only all-atom FFTopDB were arbitrarily considered here. As in the previous example, charge values are not reported in this article, but are available in the corresponding R.E.DD.B. projects. The nucleic acid FFTopDB generated in this work are listed in Table S3 of the supplementary material.† Only the RRMS values of the charge fitting step for the different FFTopDB are reported in Table 6. The relatively small values observed in all cases demonstrate the effectiveness of the approaches. Charge values are highly reproducible because of the use of multiple orientation feature, and the description of the atoms involved in the RBRA procedure, *i.e.* heavy atoms common for standard nucleoside pentoses were involved in the multiple molecular re-orientation procedure.

We would like to stress that the goal of this report is to demonstrate the capabilities of the R.E.D. Tools and not to validate the reported charge values in molecular mechanical conformational analyses or MD simulations. Potential applications include the impact of the conformation(s) involved in charge derivation on the stability of nucleic acid structures observed during MD simulations and the generation of FFTopDB for chemically engineered nucleic acids such as the glycerol and threose nucleic acids.[110,111] These developments are currently in progress, and the corresponding FFTopDB will be released in R.E.DD.B. in the near future.

**Table 6** Building of the FFTopDB for standard nucleic acids based on multiple orientation, multiple conformation and multiple molecule charge derivation automatically carried out using the R.E.D. Tools (see Fig. 4). Charge values were derived from charge fitting and not arithmetic mean.[59] (i) Accurate geometry optimization, (ii) molecular orientation of each optimized geometry controlled by the RBRA approach implemented in the R.E.D. program. MEP computation was carried using the HF/6-31G* theory level, and only all-atom force field libraries were generated

| R.E.DD.B.[a] | | F-45 | F-46 | F-47 | F-48 | F-49 | F-50 |
|---|---|---|---|---|---|---|---|
| Surf[b] | | CS | CG | CS | CG | CG | CS |
| Fit[c] | | 2 R st | 2 R st | 1 E st | 1 E st | 1 R st | 2 R st |
| RRMS[d] | (i) | 0.055 | 0.076 | 0.051 | 0.074 | 0.079 | F-45 |
| | (ii) | 0.055 | 0.075 | 0.050 | 0.074 | 0.078 | F-45 |

| R.E.DD.B.[a] | | F-51 | F-52 | F-53 | F-54 | F-55 | F-56 | F-60 |
|---|---|---|---|---|---|---|---|---|
| Surf[b] | | CS | CG | CS | CG | CG | CS | CS |
| Fit[c] | | 2 R st | 2 R st | 1 E st | 1 E st | 1 R st | 2 R st | 2 R st |
| RRMS[d] | (i) | 0.039 | 0.055 | 0.034 | 0.055 | 0.059 | F-51 | 0.059 |
| | (ii) | 0.038 | 0.054 | 0.033 | 0.054 | 0.059 | F-51 | 0.058 |

[a] Data are available in R.E.DD.B. (the corresponding R.E.DD.B. code is provided). FFTopDB for DNA: F-45 up to F-50. FFTopDB for RNA: F-51 up to F-56. FFTopDB for DNA and RNA: F-60. When not specified the phosphate group is located at the position 5′ of the nucleotide fragments. F-50 and F-56: the phosphate group is located at the position 3′ of the nucleotide fragments. [b] Surface algorithm used in MEP computation: CS = Connolly surface algorithm; CG = CHELPG algorithm. [c] Charge fitting step: 1 E st = one stage ESP; 2 R st = two-stage RESP (hyperbolic restraints = 5.0 × $10^{-4}$ and 1.0 × $10^{-3}$) as defined in Amber force fields;[28,29] 1 R st = one stage RESP (hyperbolic restraint = 1.0 × $10^{-2}$) as defined in the GLYCAM force field.[30] [d] RRMS values calculated in the single charge fitting step (1 E st and 1 R st) or in the second charge fitting step (2 R st) for each multiple orientation, multiple conformation and multiple molecule charge fit [number of structures "nmol" involved = 52 (for DNA FFTopDB), 28 (for RNA FFTopDB) and 76 (for DNA and RNA FFTopDB)]. Two RRMS values are reported: (i) in the presence, and (ii) in the absence of inter-molecular charge constraints and inter-molecular charge equivalencing required for building of nucleotide fragments: values were taken directly from the outputs of the RESP program.

**Other examples: FFTopDB for glycoconjugates and any other bio-polymers.** Deriving charges and building force field libraries involving multiple orientations, multiple conformations and multiple molecules for glycoconjugates can be obtained following highly analogous strategies to these previously presented. For example, FFTopDB for organo-glycoconjugates and glycolipids have been constructed using the R.E.D. Tools, and have been deposited in R.E.DD.B.[112,113] More generally, building FFTopDB for any type of bio-molecular systems can be generated using the R.E.D. Tools. An important extension of R.E.DD.B. will be released in the near future, and will contain examples of more complex approaches.[109]

## Conclusion

The R.E.D. Tools have been designed to automatically derive non-polarizable RESP and ESP atomic charges and create appropriate force field libraries for new bio-molecules and molecular fragments. New key features of these programs are: a method for controlling molecular orientation of each optimized geometry and capabilities of performing multiple orientation charge fitting whatever the density of points involved in MEP computation is. The resultant RESP and ESP charges are highly reproducible, with the errors of the order of 0.0001 e, which are independent of the QM program or the choice of initial structure. Although this level of accuracy is not required to achieve satisfactory MD simulations, it may become important for defining and reporting initial conditions of such simulations. This can have a particular importance in MD simulations and docking studies, where there is a need for establishing benchmarks, for facilitating tracking of errors and to have the ability to reproduce published data.[114–117] Furthermore, reproducible RESP and ESP charges might provide a rigorous starting point useful for the charge validation step in the process of developing a new force field. To take into account the charge dependence on molecular conformation, the multiple orientation and multiple conformation approaches have been combined together. In addition, the proper handling of charge constraints during charge fitting makes parameterizing molecular fragments a straightforward process. The procedure can be applied to a large ensemble of molecules in a single execution of the R.E.D. program in which a complex set of force field libraries (FFTopDB) can be built for any type of biopolymer such as nucleic acids, proteins, glycoconjugates or coenzymes.[109] Chemical elements up to bromine in the periodic table are fully handled by the R.E.D. Tools. Thus, RESP and ESP charge derivation and force field library building for bio-inorganic complexes containing fourth row transition metals are now directly accessible.[109] Moreover, both all-atom and united-carbon force field libraries can be generated. Finally, by modifying few keywords in the R.E.D. source code, a user can create a large number of new charge models and force field libraries that can be readily validated in MD simulations. This feature can be especially important if new methodologies leading to new charge models are developed.

Currently, version III.3 of the R.E.D. Tools is available for download from the q4md-forcefieldtools.org website. Recently, we also released a new web service named R.E.D. Server,[118] which provides the software and hardware required for RESP and ESP charge derivation and force field library building. R.E.D. Server employs the last version of the R.E.D. program (version R.E.D. IV β, June 2009), and the last versions of the Gaussian, GAMESS-US and PC-GAMESS/Firefly programs. New features are currently under development. They include: (i) the extension of RESP and ESP charge derivation to the use of Spackman's algorithm,[25] (ii) the handling of extra points or lone pairs in charge derivation,[87] (iii) the statistical analysis of the impact of any charge constraint used during a fitting step, (iv) the extension of the procedures reported to all chemical elements belonging to the periodic table,[109] and (v) the development of new charge models specific to bio-inorganic complexes.[109] The interface to the new i_RESP program by R.E.D. will provide a polarized force field dimension to the FFTopDB generated by the R.E.D. Tools and R.E.D. Server.[119] Finally, the next release of the R.E.D. Tools will be licensed and distributed under the GNU General Public License. We hope this will encourage further use, development and distribution of the R.E.D. Tools in the scientific community.

## Acknowledgements

We would like to acknowledge Mr A. Dubois, Mr O. Panis and Mr V. Grossu (Université de Picardie-Jules Verne), as well as Mrs J. Mills and Mrs M. Dunbar (Sanford|Burnham Institute for Medical Research) for their help regarding software licensing and copyright issues. We would like also to thank Dr D. Fox (Technical support, Gaussian, Inc.), Dr M. Schmidt (Iowa State University) and Dr A. Granovsky (Laboratory of Chemical Cybernetics at Moscow State University) for helpful discussions about the Gaussian, GAMESS-US and PC-GAMESS/Firefly programs, respectively. We are grateful to Mr T. A. D. Patko (California State University, Long Beach, CA, USA) for his contribution in porting the R.E.D. program on the Macintosh and Windows platforms. F.-Y. Dupradeau is grateful for computer time at the Institut du Développement et des Ressources en Informatique Scientifique (IDRIS, Orsay, France), the Centre Informatique National de l'Enseignement Supérieur (CINES, Montpellier, France) and the Plateforme de Modélisation et Calcul Scientifique (MeCS, Université de Picardie-Jules Verne). P. Cieplak is supported by the National Institutes of Health (grant No. GM079383).

## References

1 S. M. Bachrach, Population Analysis and Electron Densities from Quantum Mechanics, in *Reviews in Computational Chemistry*, ed. K. B. Lipkowitz and D. B. Boyd, VCH Publishers, New York, 1994, vol. 5, pp. 171–227.
2 R. S. Mulliken, *J. Chem. Phys.*, 1955, **23**, 1833–1840.
3 P. O. Löwdin, *J. Chem. Phys.*, 1950, **18**, 365–375.
4 R. W. F. Bader and T. T. Nguyen-Dang, *Adv. Quantum Chem.*, 1981, **14**, 63–124.
5 A. T. Hagler, E. Huler and S. Lifson, *J. Am. Chem. Soc.*, 1974, **96**, 5319–5327.
6 W. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1988, **110**, 1657–1666.
7 B. H. Besler, K. M. Merz Jr. and P. A. Kollman, *J. Comput. Chem.*, 1990, **11**, 431–439.
8 F. A. Momany, *J. Phys. Chem.*, 1978, **82**, 592–601.
9 S. R. Cox and D. E. Williams, *J. Comput. Chem.*, 1981, **2**, 304–323.
10 U. C. Singh and P. A. Kollman, *J. Comput. Chem.*, 1984, **5**, 129–145.
11 L. E. Chirlian and M. M. Francl, *J. Comput. Chem.*, 1987, **8**, 894–905.
12 D. E. Williams, Net Atomic Charge and Multipole Models for the Ab initio Molecular Electric Potential, in *Reviews in Computational Chemistry*, ed. K. B. Lipkowitz and D. B. Boyd, VCH Publishers, New York, 1991, vol. 2, pp. 219–271.
13 A. Jakalian, B. L. Bush, D. B. Jack and C. I. Bayly, *J. Comput. Chem.*, 2000, **21**, 132–146.
14 A. Jakalian, D. B. Jack and C. I. Bayly, *J. Comput. Chem.*, 2002, **23**, 1623–1641.
15 M. L. Connolly, *J. Appl. Crystallogr.*, 1983, **16**, 548–558.
16 C. M. Breneman and K. B. Wiberg, *J. Comput. Chem.*, 1990, **11**, 361–373.
17 C. Jenson and W. L. Jorgensen, *J. Am. Chem. Soc.*, 1997, **119**, 10846–10854.
18 R. H. Henchman and J. W. Essex, *J. Comput. Chem.*, 1999, **20**, 483–498.
19 E. Mayaan, A. Moser, A. D. MacKerell Jr. and D. M. York, *J. Comput. Chem.*, 2007, **28**, 495–507.
20 J. L. Knight and C. L. Brooks III, *J. Chem. Theory Comput.*, 2009, **5**, 1680–1691.
21 D. E. Williams, *Biopolymers*, 1990, **29**, 1367–1386.
22 T. R. Stouch and D. E. Williams, *J. Comput. Chem.*, 1992, **13**, 622–632.
23 R. J. Woods, M. Khalil, W. Pell, S. H. Moffat and V. H. Smith Jr., *J. Comput. Chem.*, 1990, **11**, 297–310.
24 K. M. Merz Jr., *J. Comput. Chem.*, 1992, **13**, 749–767.
25 M. A. Spackman, *J. Comput. Chem.*, 1996, **17**, 1–18.
26 P. Söderhjelm and U. Ryde, *J. Comput. Chem.*, 2009, **30**, 750–760.
27 C. A. Reynolds, J. W. Essex and W. G. Richards, *J. Am. Chem. Soc.*, 1992, **114**, 9075–9079.
28 C. I. Bayly, P. Cieplak, W. D. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, **97**, 10269–10280.
29 W. D. Cornell, P. Cieplak, C. I. Bayly and P. A. Kollman, *J. Am. Chem. Soc.*, 1993, **115**, 9620–9631.
30 R. J. Woods and R. Chappelle, *THEOCHEM*, 2000, **527**, 149–156.
31 P. Cieplak, J. W. Caldwell and P. A. Kollman, *J. Comput. Chem.*, 2001, **22**, 1048–1057.
32 V. M. Anisimov, G. Lamoureux, I. V. Vorobyov, N. Huang, B. Roux and A. D. MacKerell Jr., *J. Chem. Theory Comput.*, 2005, **1**, 153–168.
33 W. J. Hehre, R. Ditchfield and J. A. Pople, *J. Chem. Phys.*, 1972, **56**, 2257–2261.
34 F. Jensen, in *Introduction to Computational Chemistry*, John Wiley & Sons, Chichester, 1999.
35 P. C. Hariharan and J. A. Pople, *Chem. Phys. Lett.*, 1972, **16**, 217–219.
36 W. J. Hehre, R. F. Stewart and J. A. Pople, *J. Chem. Phys.*, 1969, **51**, 2657–2664.
37 W. J. Hehre, R. Ditchfield, R. F. Stewart and J. A. Pople, *J. Chem. Phys.*, 1970, **52**, 2769–2773.
38 S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta Jr. and P. Weiner, *J. Am. Chem. Soc.*, 1984, **106**, 765–784.
39 S. J. Weiner, P. A. Kollman, D. T. Nguyen and D. A. Case, *J. Comput. Chem.*, 1986, **7**, 230–252.
40 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.
41 P. A. Kollman, R. Dixon, W. D. Cornell, T. Fox, C. Chipot and A. Pohorille, The development/application of a 'minimalist' organic/biochemical molecular mechanic force field using a combination of *ab initio* calculations and experimental data, in

*Computer Simulation of Biomolecular Systems*, ed. A. Wilkinson, P. Weiner and W. F. van Gunsteren, Elsevier, Escom, The Netherlands, 1997, vol. 3, pp. 83–96.

42 T. E. Cheatham III, P. Cieplak and P. A. Kollman, *J. Biomol. Struct. Dyn.*, 1999, **16**, 845–862.

43 J. Wang, P. Cieplak and P. A. Kollman, *J. Comput. Chem.*, 2000, **21**, 1049–1074.

44 V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins: Struct., Funct., Bioinf.*, 2006, **65**, 712–725.

45 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.

46 R. J. Woods, R. A. Dwek, C. J. Edge and B. Fraser-Reid, *J. Phys. Chem.*, 1995, **99**, 3832–3846.

47 K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. Lachele Foley and R. J. Woods, *J. Comput. Chem.*, 2008, **29**, 622–655.

48 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.

49 T. Dunning Jr, *J. Chem. Phys.*, 1989, **90**, 1007.

50 Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. W. Caldwell, J. Wang and P. A. Kollman, *J. Comput. Chem.*, 2003, **24**, 1999–2012.

51 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *GAUSSIAN 09 (Revision A.2)*, Gaussian, Inc., Wallingford, CT, 2009.

52 D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham III, S. DeBolt, D. Ferguson, G. Seibel and P. A. Kollman, *Comput. Phys. Commun.*, 1995, **91**, 1–41.

53 D. A. Case, T. E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K. M. Merz Jr., A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *J. Comput. Chem.*, 2005, **26**, 1668–1688.

54 M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis and J. A. Montgomery, *J. Comput. Chem.*, 1993, **14**, 1347–1363.

55 A. V. Nemukhin, B. L. Grigorenko and A. A. Granovsky, *Moscow Univ. Chem. Bull.*, 2004, **45**, 75–102.

56 R. A. Kendall, E. Apra, D. E. Bernholdt, E. J. Bylaska, M. Dupuis, G. I. Fann, R. J. Harrison, J. Ju, J. A. Nichols, J. Nieplocha, T. P. Straatsma, T. L. Windus and A. T. Wong, *Comput. Phys. Commun.*, 2000, **128**, 260–283.

57 J. Wang, W. Wang, P. A. Kollman and D. A. Case, *J. Mol. Graphics Modell.*, 2006, **25**, 247–260.

58 The Tripos MOL2 file format, http://www.tripos.com/custResources/mol2Files/.

59 P. Cieplak, W. D. Cornell, C. I. Bayly and P. A. Kollman, *J. Comput. Chem.*, 1995, **16**, 1357–1377.

60 F.-Y. Dupradeau, C. Cézard, R. Lelong, E. Stanislawiak, J. Pêcher, J. C. Delepine and P. Cieplak, *Nucleic Acids Res.*, 2007, **36**, D360–D367.

61 S. Spainhour, E. Siever and N. Patwardhan, *Perl in a Nutshell*, O'Reilly, Sebastopol, CA, 2nd edn, 2002.

62 P. Raines and J. Tranter, *TCK/TK in a Nutshell*, O'Reilly, Sebastopol, CA, 1999.

63 F.-Y. Dupradeau and J. Rochette, *J. Mol. Model.*, 2003, **9**, 271–272.

64 F.-Y. Dupradeau and P. Cieplak, *RESP charge derivation using the Ante_R.E.D.-1.x & R.E.D. III.x programs*, Université de Picardie-Jules Verne, Sanford|Burnham Institute for Medical Research, 2007, http://q4md-forcefieldtools.org/Tutorial/Tutorial-1.php.

65 The PDB file format, http://http://www.wwpdb.org/docs.html.

66 J. L. Markley, A. Bax, Y. Arata, C. W. Hilbers, R. Kaptein, B. D. Sykes, P. E. Wright and K. Wüthrich, *J. Biomol. NMR*, 1998, **12**, 1–23.

67 International Union of Pure and Applied Chemistry, http://www.iupac.org/.

68 D. E. Williams, *J. Comput. Chem.*, 1994, **15**, 719–732.

69 A. Frisch, M. J. Frisch and G. W. Trucks, in *Gaussian 03 User's Reference*, Gaussian Inc., Wallingford, CT, USA, 2nd edn, Manual version 7.1, 2005.

70 *GAMESS User's Guide*, Section 2 Input Description, Department of Chemistry, Iowa State University, Ames, 2009.

71 D. Rosen, *Rigid Body Transformations*, Georgia Institute of Technology, http://www.srl.gatech.edu/education/ME6104/notes/xforms.html.

72 A. Pigache, P. Cieplak and F.-Y. Dupradeau, Automatic and highly reproducible RESP and ESP charge derivation: Application to the development of programs R.E.D. and X R.E.D., *227th ACS National Meeting*, Anaheim, CA, USA, March 28–April 1, 2004.

73 L. Yang, C.-h. Tan, M.-J. Hsieh, J. Wang, Y. Duan, P. Cieplak, J. W. Caldwell, P. A. Kollman and R. Luo, *J. Phys. Chem. B*, 2006, **110**, 13166–13176.

74 C. E. A. F. Schafmeister, W. S. Ross and V. Romanovski, *LEaP*, University of California, San Francisco, CA, 1995.

75 The Open Babel Package, version 2.2.3, http://openbabel.sourceforge.net/.

76 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, P. Salvador, J. J. Dannenberg, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle and J. A. Pople, *GAUSSIAN 98 (Revision A.11)*, Gaussian, Inc., Pittsburgh, PA, 2001.

77 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, *GAUSSIAN 03 (Revision E.01)*, Gaussian, Inc., Wallingford, CT, 2004.

78 *Insight II*, *User Guide*, Molecular Modeling System, Molecular Simulations Inc., San Diego, CA, 2000.

79 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.

80 A. St-Amant, W. D. Cornell, P. A. Kollman and T. A. Halgren, *J. Comput. Chem.*, 1995, **16**, 1483–1506.

81 T. Fox and P. A. Kollman, *J. Phys. Chem. B*, 1998, **102**, 8070–8079.

82 M. L. Strader and S. E. Feller, *J. Phys. Chem. A*, 2002, **106**, 1074–1080.

83 A. Vishnyakov, A. P. Lyubartsev and A. Laaksonen, *J. Phys. Chem. A*, 2001, **105**, 1702–1710.
84 M. Fioroni, K. Bruger, A. E. Mark and D. Roccatero, *J. Phys. Chem. B*, 2000, **104**, 12347–12354.
85 D. Roccateno, G. Colombo, M. Fioroni and A. E. Mark, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 12179–12184.
86 S. Jalili and M. Akhavan, *J. Comput. Chem.*, 2009, **31**, 286–294.
87 R. W. Dixon and P. A. Kollman, *J. Comput. Chem.*, 1997, **18**, 1632–1646.
88 P. Cieplak and P. A. Kollman, *J. Comput. Chem.*, 1991, **12**, 1232–1236.
89 A. E. Howard, P. Cieplak and P. A. Kollman, *J. Comput. Chem.*, 1995, **16**, 243–261.
90 J. Wang and P. A. Kollman, *J. Comput. Chem.*, 2001, **22**, 1219–1228.
91 M. D. Beachy, D. Chasman, R. B. Murphy, T. A. Halgren and R. A. Friesner, *J. Am. Chem. Soc.*, 1997, **119**, 5908–5920.
92 A. W. Burgess and S. J. Leach, *Biopolymers*, 1973, **12**, 2599–2605.
93 I. L. Karle, *Biopolymers*, 2001, **60**, 351–365.
94 L. Wang, A. Brock, B. Herberich and P. G. Schultz, *Science*, 2001, **292**, 498–500.
95 M. M. Francl and L. E. Chirlian, The Pluses and Minuses of Mapping Atomic Charges to Electrostatic Potentials, in *Reviews in Computational Chemistry*, ed. K. B. Lipkowitz and D. B. Boyd, VCH Publishers, New York, 2000, vol. 14, pp. 1–31.
96 E. Sigfridsson and U. Ryde, *J. Comput. Chem.*, 1998, **19**, 377–395.
97 J. H. Wilkinson, *Rounding Errors in Algebraic Processes*, Dover, New York, 1994.
98 H. F. King and M. Dupuis, *J. Comput. Phys.*, 1976, **21**, 144–165.
99 M. Dupuis, J. Rys and H. F. King, *J. Chem. Phys.*, 1976, **65**, 111–116.
100 M. North, *Principles and Applications of Stereochemistry*, Stanley Thornes, Cheltenham, 1998.
101 J. Demmel, *Trading Off Parallelism and Numerical Stability*, Computer Science Division Tech Report UCB//CSD-92-702, University of California, Berkeley, 1992.
102 M. S. Gordon, M. D. Bjorke, F. J. Marsh and M. S. Korth, *J. Am. Chem. Soc.*, 1978, **100**, 2670–2678.
103 P. J. Stephens, F. J. Devlin, C. F. Chablowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623–11627.
104 R. H. Hertwig and W. Koch, *Chem. Phys. Lett.*, 1997, **268**, 345–351.
105 N. Homeyer, A. H. C. Horn, H. Lanig and H. Sticht, *J. Mol. Model.*, 2006, **12**, 281–289.
106 V. M. Anisimov, I. V. Vorobyov, B. Roux and A. D. MacKerell Jr., *J. Chem. Theory Comput.*, 2007, **3**, 1927–1946.
107 M. Basma, S. Sundara, D. Calgan, T. Varnali and R. J. Woods, *J. Comput. Chem.*, 2001, **22**, 1125–1137.
108 B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187–217.
109 C. Cézard, E. Vanquelef, J. Pêcher, P. Sonnet, P. Cieplak, E. Derat and F.-Y. Dupradeau, RESP charge derivation and force field topology database generation for complex bio-molecular systems and analogs, *236th ACS National Meeting*, Philadelphia PA, USA, August 17–August 21, 2008.
110 Z. Lilu, P. Adam and E. Meggers, *J. Am. Chem. Soc.*, 2005, **127**, 4174–4175.
111 K.-U. Schöning, P. Scholz, S. Guntha, X. Wu, R. Krishnamurthy and A. Eschenmoser, *Science*, 2000, **290**, 1347–1351.
112 S. G. Gouin, E. Vanquelef, J. M. Garcia Fernand, C. Ortiz Mellet, F.-Y. Dupradeau and J. Kovensky, *J. Org. Chem.*, 2007, **72**, 9032–9045, and R.E.DD.B. code "F-71".
113 S. Abel, F.-Y. Dupradeau and M. Marchi, Etudes par simulations de dynamique moléculaire explicites de micelles de dodecyl maltoside: influence de la conformation de la tête polaire maltose et du champ de forces sur la structure des agrégats, *Journées Modélisation de l'ENS-ENSCP*, 15–16juin, 2009, and R.E.DD.B. code "F-72".
114 Reproducibility, http://www.gromacs.org/WIKI-import/Reproducibility.
115 S. E. Murdock, K. Tai, M. H. Ng, S. Johnston, B. Wu, H. Fangohr, C. A. Laughton, J. W. Essex and M. S. P. Sansom, *J. Chem. Theory Comput.*, 2006, **2**, 1477–1481.
116 C. I. Williams and M. Feher, *J. Comput.-Aided Mol. Des.*, 2008, **22**, 39–51.
117 M. Feher and C. I. Williams, *J. Chem. Inf. Model.*, 2009, **49**, 1704–1714.
118 E. Vanquelef, S. Simon, G. Marquant, J. C. Delepine, P. Cieplak and F.-Y. Dupradeau, R.E.D. Server: a web service designed to automatically derive RESP and ESP charges and to generate force field libraries for new molecules and new molecular fragments, Université de Picardie Jules Verne - Sanford|Burnham Institute for Medical Research, 2009.
119 P. Cieplak, F.-Y. Dupradeau, Y. Duan and J. Wang, *J. Phys.: Condens. Matter*, 2009, **21**, 333102.