

R.E.DD.B.: A database for RESP and ESP atomic charges, and force field libraries

François-Yves Dupradeau^{1,*}, Christine Cézard¹, Rodolphe Lelong¹,
Élodie Stanislawiak¹, Julien Pêcher¹, Jean Charles Delepine² and Piotr Cieplak³

¹Université de Picardie Jules Verne - Faculté de Pharmacie & UMR CNRS 6219, ²Université de Picardie Jules Verne, Amiens, France and ³Burnham Institute for Medical Research, La Jolla, CA, USA

Received September 1, 2007; Revised October 1, 2007; Accepted October 2, 2007

ABSTRACT

The web-based RESP ESP charge DataBase (R.E.DD.B., <http://q4md-forcefieldtools.org/REDDB>) is a free and new source of RESP and ESP atomic charge values and force field libraries for model systems and/or small molecules. R.E.DD.B. stores highly effective and reproducible charge values and molecular structures in the Tripos mol2 file format, information about the charge derivation procedure, scripts to integrate the charges and molecular topology in the most common molecular dynamics packages. Moreover, R.E.DD.B. allows users to freely store and distribute RESP or ESP charges and force field libraries to the scientific community, via a web interface. The first version of R.E.DD.B., released in January 2006, contains force field libraries for molecules as well as molecular fragments for standard residues and their analogs (amino acids, monosaccharides, nucleotides and ligands), hence covering a vast area of relevant biological applications.

INTRODUCTION

Classical molecular dynamics simulations are nowadays the theoretical tools of choice to study biomolecular systems such as oligosaccharides and glycoconjugates (1,2), proteins (3,4), DNA/RNA and their complexes (5–7). At the basis of those tools is the force field, which partitions the total energy into classical terms. Most of the parameters needed to describe these energetic contributions can be found in a number of available force fields, such as AMBER (8–11)/GAFF (12), CHARMM (13–15), GLYCAM (16–18), GROMOS (19,20) and OPLS (21,22), except for the atomic charges required for evaluating the

electrostatic energy. Each force field provides its own set of atomic charges for the standard residues in its specific format. However, for any non-standard residue or molecule the charges need to be derived anew, which is a tedious and limiting step for computational biologists involved in structural studies. Several empirical and semi-empirical methods have been developed for deriving atomic charges. Among them are approaches in which the charges are varied to fit the properties of crystals (23) or liquids (21). Another group of methods is based on quantum mechanical calculations. Appropriate atomic charges are derived either from Mulliken (24) and Lowdin (25) population analysis or by applying the theory of atoms in molecule (26), or fitting them to the molecular electrostatic potential (MEP) [reviewed in (27) and (28)]. The latter approach has proved to be highly accurate in simulation of condensed-phase properties, and to handle inter-molecular properties. Hence, nowadays ESP and RESP methods (29–34), belonging to the last group of methods, are widely used to derive atomic charge values for biomolecules.

We developed the RESP ESP charge Derive (R.E.D.) program (35,36) to automatically derive RESP or ESP charges. We designed also, available online, the RESP ESP charge DataBase (R.E.DD.B., <http://q4md-forcefieldtools.org/REDDB>). R.E.DD.B. stores force field libraries containing charge values, molecular structures and topologies, as well as information about the charge derivation procedure and scripts to integrate the data in the most common molecular dynamics packages. Such a strategy ensures computational biologists and chemists a rigorous reproducibility and description of the published results. Up to date, to the best of our knowledge, no system exists that rigorously defines the charge derivation conditions, stores, compares, improves charge models and allows the free distribution of the corresponding charge values in the form of force field libraries for biomolecular systems.

*To whom correspondence should be addressed. Tel: +33 3 2282 7498; Fax: +33 3 2282 7469; Email: fyd@q4md-forcefieldtools.org
Correspondence may also be addressed to Piotr Cieplak. Tel: +1 858 646 3100; Fax: +1 858 713 9949; Email: pcieplak@q4md-forcefieldtools.org

The AMBER parameter database is maintained at the school of Pharmacy and Pharmaceutical Sciences (University of Manchester, UK) (37) and a database for modified RNA nucleotides is available in the SantaLucia's group (38). However, those tools exhibit number of limitations such as the absence of description of the computational conditions used to derive the available data or the restriction of these data to a particular class of biomolecules. Furthermore, several attempts of publishing force field libraries have been made by the authors themselves as supplementary material of a given publication and/or on their own website [see for instance recent publications such as (39) and (40)]. However, so far, no universal database exists centralizing all these efforts from various groups, and providing computational conditions used to reproduce the published data.

METHODOLOGY

In ESP and RESP methods (27–34), the atomic charges are fitted to the MEP(s) around a molecule or a set of molecules. The strategy to derive such charges is described in detail elsewhere (29–34), thus we will only report here the main steps of the adopted methodology, which is independent of the surface/grid algorithm used [Connolly (41,29), CHELP (42) and CHELPG (43)]:

- The geometry of the molecule is optimized using an *ab initio* method (in the case of *i* conformations this step is performed for each of them).
- The MEP(s) around each of the *i*-th conformation of the target molecule is computed using the geometry obtained in step (a) using a suitable *ab initio* method. For each of the *i*-th conformation, the number of *j* molecular orientations might be generated and then for each of them an appropriate MEP is computed. This ensures that charges have reproducible values (35).
- The atom-centered charges are fitted to reproduce the *i***j* MEPs. This can be done using any of appropriate fitting programs, such as the RESP module of the AMBER program suite (30,44), the FITCHARGE module of the CHARMM package (34,45) or the Pdm97 program (46).
- Finally, in the case of complex charge derivation procedure involving *n* different molecules, the steps (a), (b) and (c) are repeated *n* times. Then the corresponding MEPs are concatenated, and a multiple molecule RESP or ESP fit is carried out. This step can be performed with or without additional charge constraint for atom(s) and/or group(s) of atoms (32,36).

This procedure is routinely used to derive atomic charges for any force field; nevertheless, it suffers from several limitations. The Gaussian (47) and GAMESS (48) packages are most often used for carrying out *ab initio* calculations in steps (a) and (b). The charges generated in such a way are not easily reproducible, and can exhibit some noticeable discrepancies between programs and authors (35). Application of this methodology to large molecules and to different conformations is tedious,

time-consuming, error-prone and lacks a rigorous way to check the calculated charges. Consequently, many atomic charge values published in the literature are not reproducible, and/or of poor quality.

The R.E.D. program (35,36) allows the user to derive RESP or ESP charges starting from unoptimized structures, with an accuracy in charge reproducibility of the order of $\pm 0.0001 e$. Indeed, complex charge fitting procedures involving multiple molecules, multiple conformations and multiple orientations, can be automatically carried out leading to highly reproducible atomic charges independently of the quantum mechanical software or the initial Cartesian coordinate set. Those charges are also recognized as highly effective or suitable for molecular dynamics simulation (9–11). Special set of charge constraints has been implemented in order to derive charges for any type of molecular fragments or set of molecules. These constraints control the charge values on specific atom(s) and/or group(s) of atoms in a molecule or in a set of molecules. The outputs of the R.E.D. program are in the Tripos mol2 file format. They can be used in any molecular dynamics program, and can be directly stored in R.E.DD.B.

DATABASE DEVELOPMENT

R.E.DD.B. is developed using HTML, PHP MySQL, Javascript and Java. The database is implemented in MySQL (version 12.22) and consists of eight interconnected tables, while all web-based forms and queries to the database are written in PHP (version 4.3.10). The database is hosted on a Linux server located at the Université de Picardie Jules Verne in Amiens.

Moreover, since the R.E.DD.B. website uses the Jmol program (50) to visualize the charge values and the structures stored in R.E.DD.B., the J2SE Runtime Environment (51) has to be installed on the user front-end machine. R.E.DD.B. molecule name entries and R.E.DD.B. project titles are cross-linked to Wikipedia (52) providing general information about the molecules and projects available in R.E.DD.B. When no matching article exists in Wikipedia, the user could initiate the writing of a new entry about the molecules/project described in R.E.DD.B.

The R.E.DD.B. website as well as the database contents can be easily browsed using any popular web browser under any operating system. The first version of R.E.DD.B. has been released in January 2006. Since this date, the homepage has been accessed nearly 3000 times, and more than 50 official users from throughout the world are registered. Furthermore, the R.E.DD.B. homepage also contains numerous links towards helpful tools and programs, bibliography and tutorial involving atomic charge derivation (53).

DATABASE CONTENTS

R.E.DD.B. presently stores force field topologies for around 150 molecular systems, covering a vast area of relevant biological applications, divided into two

sub-entities, namely 'Whole molecule' and 'Molecule fragment'. An overview of the R.E.D.D.B. home page is presented in Figure 1.

- Whole molecule project (denoted by 'W-*i*', where *i* is the number of the whole molecule projects available at the time the project was submitted to R.E.D.D.B. by its author): it corresponds to an intact [not broken into fragment(s)] molecule. Specific examples are presented below:

(i) Projects involving a single molecule. Projects W-20 up to W-23 involve *N*-methylacetamide (NMA). These projects contain topologies and RESP atomic charges for NMA in *cis* and *trans* conformations. Two or four molecular orientations and one or two conformations were used during the charge derivation procedure. Such data are needed in force field development to model the peptide bond (9).

(ii) More complex projects involving different small organic or inorganic molecules. Projects W-46 up to W-49. These projects contain topologies and RESP or ESP charges based on the Connolly surface (41) or CHELPG (43) algorithm for 10 common solvent molecules (Dimethylsulfoxide, Ethanol, Trifluoroethanol, Methanol, Acetone, Acetic acid, Acetonitrile, Benzene, Toluene and Chloroform) fitted together using the R.E.D.-III program. Two or four molecular orientations and a single conformation were used for each solvent molecule. Such data are required for condensed-phase simulation, and for building boxes of solvent.

- Molecule fragment project (denoted by 'F-*i*', where *i* is the number of molecule fragment projects available at the time the project was submitted to R.E.D.D.B. by its author): corresponds to a fragment of an organic or inorganic macro-molecule. This means that in the process of charge fitting some atoms were removed from the structure(s). Specific examples are presented below:

(i) Projects F-23 up to F-28: present the central fragment of *O*-methyl-L-tyrosine (TYM), the 21st amino acid synthesized by the Schultz's group (54). These projects contain topologies and RESP or ESP charges obtained using the Connolly surface (41) or CHELPG (43) algorithm for the NHCHCH₂PhOMeCO fragment. The charges were derived using *N*-Acetyl-*O*-methyl-L-tyrosine-*N'*-methylamide (capped amino-acid ACE-TYM-NME) and using intra-molecular charge constraints set to zero for the ACE and NME residues during the fitting step. One or two conformations (close to an α -helix: $\varphi = -72.41$, $\psi = -34.82$, $\chi = -171.96$, or a β -sheet: $\varphi = -119.15$, $\psi = 138.71$, $\chi = -60.94$) and two molecular orientations for each conformation were used during the charge derivation procedure.


(ii) Project F-58: unusual nucleotides (55). This project contains topologies and RESP charges for eight fragments of two 2-(2'-Hydroxyphenyl)benzoxazole (or HBO) *C*-aryl nucleotides (HBO enol-imine

oriented in the DNA major or minor groove). Three structures were used in the charge derivation, i.e. Dimethylphosphate (conformation *gauche*, *gauche*; single molecular orientation), HBO *C5*- and *C3*-aryl-nucleosides (conformation *C2'endo* for both nucleosides; single molecular orientation). The procedure was automatically carried out using the R.E.D.-III program.

(iii) Project F-60: this project contains RESP atomic charges for the thirty-two components of a DNA/RNA force field topology database. Seventy-six structures were used during the charge derivation procedure, i.e. Dimethylphosphate (represented by a *gauche*, *gauche* conformation and four molecular orientations), the four DNA nucleosides (represented by *C2'endo* and *C3'endo* conformations and six molecular orientations for each conformation) and the four RNA nucleosides (represented by *C3'endo* conformation and six molecular orientations). The nucleotide fragments are then generated by the fusion between Dimethylphosphate and the corresponding nucleoside. Although these charges are similar to those calculated in the Cornell *et al.* (9) AMBER force field, their advantage is to be highly reproducible. Moreover, no manual adaptation of the total charge of the terminal DNA and RNA fragments is required, since the DNA and RNA nucleosides are fitted in a single R.E.D.-III run.

Further details on fragment charge derivation can be found in the tutorial describing how to use the R.E.D.-III program (53). It has to be noted that these fragments are and must be compatible with previously existing ones available in force field topology databases such as the Cornell *et al.* (9) AMBER force field or the Duan *et al.* (11) force field.

Each entry in the database is identified by a R.E.D.D.B. code and includes the following mandatory elements: (i) RESP or ESP charge values and molecular topologies for the whole molecules or molecule fragments in the Tripos mol2 file format (49); (ii) Files in the Protein Data Bank (PDB) format (56) containing the molecular orientation(s) and the Cartesian coordinates optimized using an ab initio method for each conformation of the molecule(s) used in the charge derivation process and (iii) inputs files for the used fitting program. Five extra force field related files might also be stored in the database if the author of the R.E.D.D.B. project decides to make them available. Those include scripts to convert the Tripos mol2 files into AMBER or CHARMM force field libraries, additional force field parameters compatible with the libraries provided and any information files the author wishes to provide. Moreover, given the computational complexity of some projects, all the computational conditions and details used during the different steps of the charge derivation procedure are always fully and thoroughly documented. Hence, this will help a future user to reproduce submitted charges for an educational or reviewing purpose.



Home Register Submit project Update project List projects Download projects

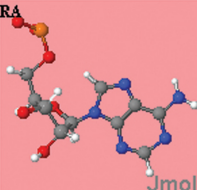
RESP ESP charge DDataBase

Home Page

R.E.D.D.B. is a database of RESP and ESP charges of high quality and reproducibility, and force field libraries for small molecules and molecular fragments: A database for computational biologists involved in AMBER/GAFF, CHARMM, GLYCAM & OPLS force field based biological studies

Characteristic examples available in R.E.D.D.B.

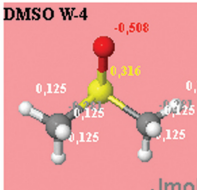
New DB for DNA & RNA
From a 76 struct. RESP fit
(4 orient + 4*6 orient + 4*2 conf*6 orient)
Connolly surface & HF/6-31G*



off 20% 100%
 spin

NEW DNA-RNA fragments NEW
R.E.D.D.B. code: [F-60](#)

20 examples...
From different RESP fits
Mol., conf. & orient. fully defined...
Connolly surface & HF/6-31G*



off 20% 100%
 spin

Selected examples
Selection of R.E.D.D.B. codes

USERS = 52 WHOLE MOLECULES = 79 MOLECULE FRAGMENTS = 71

Page visited **002970** times since January, 16th, 2006.

Home Register Submit project Update project List projects Download projects

Last update of this R.E.D.D.B. page: October, 1st, 2007
Web site design by E. Stanislawiak

UNIVERSITÉ de Picardie Jules Verne
Universit  de Picardie Jules Verne - Burnham Institute for Medical Research
Internet page   2005-2008. All rights reserved.
R.E.D.D.B. projects free.

BURNHAM INSTITUTE for MEDICAL RESEARCH
From Research, the Power to Cure

Figure 1. General overview of the R.E.D.D.B. home page.

We decided to distribute the force field libraries in the Tripos mol2 file format, since it has some advantages over the PDB format (49,57). Indeed, not only does the Tripos mol2 file format contain information present in a PDB file, i.e. Cartesian coordinates, atom names and numbers, residue names and numbers, but it can add the following features, such as:

- force field atom types and atom partial charges,
- atoms connectivities defining the topology of the molecule (bonds and bond types), required in molecular simulation,
- any arbitrary number of digits after the decimal point for the Cartesian coordinates.

Our choice was also motivated by the fact that the Tripos mol2 file format is force field independent, and can be easily converted into a library specific to AMBER (LEaP OFF file) or CHARMM (RTF or PSF file). This file format became more and more popular over the years, and is now recognized by most of the molecular modeling related packages: Babel (58) and Openbabel (59), LEaP version of AMBER 8/AMBER 9 (60). Moreover, a Tripos mol2 file can easily be converted into a CML file (61) to display the molecular structure and atomic charge values using the Jmol (50) program within any browser as implemented in R.E.DD.B.

Finally, R.E.DD.B. can also be seen as a database of quantum mechanically optimized molecules that can be directly downloaded and used in the PDB file format.

INTERACTING WITH THE DATABASE

Submission of new data

To submit a new entry in R.E.DD.B., a registration is mandatory. Once registered, the user has to follow four steps to fully upload her/his project. The submission forms contain extensive text to help the user throughout the four uploading stages.

Step 1: Registration and/or login

Step 2: General Information and charge derivation procedure

The user has to provide the database with general information about the project: name, description and abstract, up to five keywords related to the system under study, reference to publication if available and indicate if additional files (scripts to convert the mol2 files into proprietary libraries, additional force field parameters and extra files in a free format. . .) are included in the project. At this stage, the user can choose whether or not to disclose the project, since it can be temporary protected or hidden to the scientific community for a period of 3, 6 or 12 months. Then, the user has to submit general information on the charge derivation procedure by supplying the following information: whether the R.E.D. program was used to derive the charges or not, the type of project ('Whole molecule' or 'Molecule fragment'), the number of molecules used to derive the charges, and the number of Tripos mol2 files to be uploaded.

Step 3: Information regarding quantum mechanical calculations and the fitting step

For every molecules of the project, the author has to provide its name, the number of molecular conformations, and the number of molecular orientations for each conformation used in the charge derivation process with the corresponding re-orientation procedure. Next, the information must be provided for the following calculation steps: details about (i) the geometry optimization step [which quantum mechanics program was used, basis set, theory level, and specific program keyword(s)]; (ii) the MEP computation step [which quantum mechanics program was used, basis set, theory level, surface or grid algorithm and specific program keyword(s)] and (iii) the fitting step [program, and number of stage(s) used in the fit].

Step 4: Upload section

Based on the information given in the previous steps, a 'Summary' is generated, and the user is asked to check the data and to modify/correct them, if necessary. Below the summary, is the interface to upload all the required files to the server, i.e. the mandatory Tripos mol2, the PDB, the fitting program input files, and the five optional force field-related files (Figure 2).

Finally, a R.E.DD.B. code, or project code, is automatically attributed to a R.E.DD.B. project once it is successfully recorded in R.E.DD.B. This code should be used by its author(s) to reference a R.E.DD.B. project in the corresponding publication, thus helping readers to study and re-use the data described in the manuscript and ensuring the reproducibility of the published results. It has to be noted that every project can be accessed anytime by an author to update her/his project(s), such as attributing an actual publication reference to the project, and/or to upload non mandatory additional force field related files.

Searching the database

Any user can search the database. Downloading projects stored in R.E.DD.B. does not require a registration. The user can search the database by listing all the projects ('List projects'), or by using the 'Download projects' link. In the latter, four different search fields are available: 'project code', 'molecule keyword', 'molecule name' and 'author last name'. The information returned to the user consists of a link pointing towards the project 'Summary of information' describing the computational conditions, a list of all the files in the project, individually downloadable, as well as a compressed archive file containing the whole project data.

Miscellaneous tools

Tools to correct and delete projects have also been developed in R.E.DD.B. but are only available to the R.E.DD.B. system administrator. In order to correct a project, or withdraw it from R.E.DD.B., the user must send an email to the administrator. A Golden Book is

UPLOAD SECTION

Maximum size for each uploaded file: ~ 2 Mo

TRIPOS FILES			
Name of the Tripos mol2 file (number = 1)	Fragment HBA3	File to upload...	<input type="text"/> <input type="button" value="Parcourir..."/>
Name of the Tripos mol2 file (number = 2)	Fragment HBA5	File to upload...	<input type="text"/> <input type="button" value="Parcourir..."/>
Name of the Tripos mol2 file (number = 3)	Fragment HBA	File to upload...	<input type="text"/> <input type="button" value="Parcourir..."/>
Name of the Tripos mol2 file (number = 4)	Fragment HBAN	File to upload...	<input type="text"/> <input type="button" value="Parcourir..."/>
Name of the Tripos mol2 file (number = 5)	Fragment HBI3	File to upload...	<input type="text"/> <input type="button" value="Parcourir..."/>
Name of the Tripos mol2 file (number = 6)	Fragment HBI5	File to upload...	<input type="text"/> <input type="button" value="Parcourir..."/>
Name of the Tripos mol2 file (number = 7)	Fragment HBI	File to upload...	<input type="text"/> <input type="button" value="Parcourir..."/>
Name of the Tripos mol2 file (number = 8)	Fragment HBIN	File to upload...	<input type="text"/> <input type="button" value="Parcourir..."/>

PDB FILES	
PDB file to upload (name = 'Dimethylphosphate', number = 1)...	<input type="text"/> <input type="button" value="Parcourir..."/>
PDB file to upload (name = '2-(2'-Hydroxyphenyl)benzoxazole C5-aryl-nucleoside', number = 2)...	<input type="text"/> <input type="button" value="Parcourir..."/>
PDB file to upload (name = '2-(2'-Hydroxyphenyl)benzoxazole C3-aryl-nucleoside', number = 3)...	<input type="text"/> <input type="button" value="Parcourir..."/>

RESP INPUT FILES	
RESP input1 to upload...	<input type="text"/> <input type="button" value="Parcourir..."/>
RESP input2 to upload...	<input type="text"/> <input type="button" value="Parcourir..."/>

FILE 1	
Force field related file you wish to upload...	<input type="text"/> <input type="button" value="Parcourir..."/>

Figure 2. Interface to upload the files constituting a R.E.DD.B. Project. The files uploaded for the project F-58 (55) are presented as an example. Eight molecular fragments in the Tripos mol2 file format, three molecules in the PDB file format, two RESP input files and a script for converting the Tripos mol2 files into LEaP OFF libraries are uploaded.

also available, where users can write messages, comments and suggestions concerning the database and web site.

CONCLUSION AND FUTURE PERSPECTIVES

R.E.DD.B. is a publicly accessible repository offering to computational biologists and chemists a molecular force field topology database as well as a new source of highly effective and reproducible atomic charge values. The Tripos mol2 files contained in a given R.E.DD.B. project are force field independent and are for immediate use for any force field-based molecular modeling package, since they contain most of the 'external' information needed to build the topology file: Cartesian coordinates, atom and residue names, atom connectivities and atomic charge values. The atom types specific to the force field chosen by the user and the connecting atoms between molecular fragments are added using a script-based approach or any dedicated program.

The systems collected so far in R.E.DD.B. database comprise: amino acids, monosaccharides, nucleotides, ligands, solvents and cover standard as well as non-standard systems and/or residues. Thus, the content of the database (RESP or ESP atomic charge values, topologies, as well as the information about procedures applied for deriving charges) is helpful for any computational biologist and chemist for research and educational purposes.

We also highly recommend future R.E.DD.B. users to execute the R.E.D. program to derive charges before submitting them in the database, since all the files required for submission to R.E.DD.B. are automatically generated by the R.E.D. program. Moreover, with the next release of the R.E.D. program, version IV, which can handle force field topology databases for complex systems, we expect that resulting projects will be submitted to R.E.DD.B. Such projects, in order to be validated, will have to go through a 'peer-review' system, reinforcing the quality and the reproducibility of the submitted data. Number of

computational biologists and chemists, from around the world, working in the field of molecular dynamics and force field parameterization have already agreed to review projects submitted to R.E.DD.B.

We also would like R.E.DD.B. to become a unique database and unified platform for collecting published atomic charges and topologies. For this purpose, we plan to gather most of the online available force field libraries, set of charges, as well as new data found in the literature, and include them in R.E.DD.B. after approval by their authors. New R.E.DD.B. tools are currently under development to improve the capabilities of the database and web interface. As an increase in the project submissions is expected in the near future, those tools will facilitate the treatment and analysis of high volume of data, and consequently, queries to the database will be efficiently processed.

ACKNOWLEDGEMENTS

The authors of R.E.DD.B. are grateful to Professor M.-A. Fliniaux (former vice-President of Université de Picardie Jules Verne), Professor J. Rochette (Director of the UPRES EA 3901 team), to Professor P. Sonnet (Faculté de Pharmacie & UMR CNRS 6219) and to Professor D. A. Case (Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA, USA) for supporting this project. We would like also to thank Mr A. Dubois, Mrs S. Lemaire (Université de Picardie Jules Verne), Mrs J. Mills and Mrs M. Dunbar (Burnham Institute for Medical Research) regarding their assistance on copyright issues. F.-Y.D. is grateful to acknowledge computer time at the Institut du Développement et des Ressources en Informatique Scientifique (IDRIS, Orsay, France) and the Centre Informatique Nationale de l'Enseignement Supérieur (CINES, Montpellier, France). P.C. was supported by the National Institutes of Health (grant No. GM079383). Funding to pay the Open Access publication charges for this article was provided by funds obtained from the development of the R.E.D. program.

Conflict of interest statement. None declared.

REFERENCES

1. Woods,R.J. (1998) Computational carbohydrate chemistry: what theoretical methods can tell us. *Glycoconjugate J.*, **15**, 209–216.
2. Lipkowitz,K.B. (1998) Applications of computational chemistry to the study of cyclodextrins. *Chem. Rev.*, **98**, 1829–1873.
3. Adcock,S.A. and McCammon,J.A. (2006) Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem. Rev.*, **106**, 1589–1615.
4. Daggett,V. (2006) Protein folding simulation. *Chem. Rev.*, **106**, 1898–1916.
5. Cheatham,T.E. III and Case,D.A. (2006) Using AMBER to simulate DNA and RNA. In Spomer,J. and Lankas,F. (eds), *Computational Studies of RNA and DNA*, Dordrecht, Springer, Vol. 2, pp. 45–72.
6. MacKerell,A.D. Jr and Nilsson,L. (2006) Theoretical studies of nucleic acids and nucleic acid-protein complexes using CHARMM. In Spomer,J. and Lankas,F. (eds), *Computational Studies of RNA and DNA*, Dordrecht, Springer, Vol. 2, pp. 73–94.
7. Spackova,N., Cheatham,T.E. III and Spomer,J. (2006) Molecular dynamics simulations of nucleic acids. In Spomer,J. and Lankas,F. (eds), *Computational Studies of RNA and DNA*, Dordrecht, Springer, Vol. 2, pp. 301–326.
8. Weiner,S.J., Kollman,P.A., Case,D.A., Singh,U.C., Ghio,C., Alagona,G., Profeta,S. and Weiner,P. (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.*, **106**, 765–784.
9. Cornell,W.D., Cieplak,P., Bayly,C.I., Gould,I.R., Merz,K.M., Ferguson,D.M., Spellmeyer,D.C., Fox,T., Caldwell,J.W. *et al.* (1995) A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.
10. Wang,J., Cieplak,P. and Kollman,P.A. (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.*, **21**, 1049–1074.
11. Duan,Y., Wu,C., Chowdhury,S., Lee,M.C., Xiong,G., Zhang,W., Yang,R., Cieplak,P., Luo,R. *et al.* (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.*, **24**, 1999–2012.
12. Wang,J., Wolf,R.M., Caldwell,J.W., Kollman,P.A. and Case,D.A. (2004) Development and testing of a general amber force field. *J. Comput. Chem.*, **25**, 1157–1174.
13. MacKerell,A.D. Jr, Wiórkiewicz-Kuczera,J. and Karplus,M. (1995) An all-atom empirical energy function for the simulation of nucleic acids. *J. Am. Chem. Soc.*, **117**, 11946–11975.
14. MacKerell,A.D. Jr, Bashford,D., Bellott,M., Dunbrack,R.L., Evanseck,J.D., Field,M.J., Fischer,S., Gao,J., Guo,H. *et al.* (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, **102**, 3586–3616.
15. Foloppe,N. and MacKerell,A.D. Jr (2000) All-atom empirical force field for nucleic Acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.*, **21**, 86–104.
16. Kirschner,K.N. and Woods,R.J. (2001) Solvent interactions determine carbohydrate conformation. *Proc. Natl Acad. Sci. USA*, **98**, 10541–10545.
17. Basma,M., Sundara,S., Calgan,D., Vernali,T. and Woods,R.J. (2001) Solvated ensemble averaging in the calculation of partial atomic charges. *J. Comput. Chem.*, **22**, 1125–1137.
18. Kirschner,K.N. and Woods,R.J. (2001) Quantum mechanical study of the nonbonded forces in water-methanol complexes. *J. Phys. Chem. A*, **105**, 4150–4155.
19. Oostenbrink,C., Villa,A., Mark,A.E. and van Gunsteren,W.F. (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameters sets 53A5 and 53A6. *J. Comput. Chem.*, **25**, 1656–1676.
20. Soares,T.A., Hünenberger,P.H., Kastenzholz,M.A., Kräutler,V., Lenz,T., Lins,R.D., Oostenbrink,C. and van Gunsteren,W.F. (2005) An improved nucleic acid parameter set for the GROMOS force field. *J. Comput. Chem.*, **26**, 725–737.
21. Jorgensen,W.L. and Tirado-Rives,J. (1988) The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.*, **110**, 1657–1666.
22. Jorgensen,W.L., Maxwell,D.S. and Tirado-Rives,J. (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, **118**, 11225–11236.
23. Hagler,A.T., Huler,E. and Lifson,S. (1974) Energy functions for peptides and proteins. I. Derivation of a consistent force field including the hydrogen bond from amide crystals. *J. Am. Chem. Soc.*, **96**, 5319–5327.
24. Mulliken,R.S. (1955) Electronic population analysis on LCAO-MO molecular wave functions. I. *J. Chem. Phys.*, **23**, 1833–1840.
25. Lowdin,P.O. (1950) On the non-orthogonality problem connected with the use of atomic wave functions in the theory of molecules and crystals. *J. Chem. Phys.*, **18**, 365–375.
26. Bader,R.W.F. and Nguyen-Dang,T.T. (1981) Quantum theory of atoms in molecules-Dalton revisited. *Adv. Quantum Chem.*, **14**, 63–124.
27. Williams,D.E. (1991) Net atomic charge and multipole models for the ab initio molecular electricpotential. In Lipkowitz,K.B. and

- Boyd, D.B. (eds), *Reviews in Computational Chemistry*, Wiley VCH, New York, Vol. 2, pp. 219–271.
28. Francl, M.M. and Chirlian, L.E. (2000) The pluses and minuses of mapping atomic charges to electrostatic potentials. In Lipkowitz, K.B. and Boyd, D.B. (eds), *Reviews in Computational Chemistry*, Wiley VCH, New York, Vol. 14., pp. 1–31.
 29. Singh, U.C. and Kollman, P.A. (1984) An approach to computing electrostatic charges for molecules. *J. Comput. Chem.*, **5**, 129–145.
 30. Bayly, C.I., Cieplak, P., Cornell, W.D. and Kollman, P.A. (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges the RESP model. *J. Phys. Chem.*, **97**, 10269–10280.
 31. Cornell, W.D., Cieplak, P., Bayly, C.I. and Kollman, P.A. (1993) Application of RESP charges to calculate conformational energies, hydrogen-bond energies, and free-energies of solvation. *J. Am. Chem. Soc.*, **115**, 9620–9631.
 32. Cieplak, P., Cornell, W.D., Bayly, C. and Kollman, P.A. (1995) Application of the multimolecule and multiconformational RESP methodology to biopolymers: charge derivation for DNA, RNA, and proteins. *J. Comput. Chem.*, **16**, 1357–1377.
 33. Woods, R.J. and Chappelle, R. (2000) Restrained electrostatic potential atomic partial charges for condensed-phase simulations of carbohydrates. *J. Mol. Struct. (THEOCHEM)*, **527**, 149–156.
 34. Anisimov, V.M., Lamoureux, G., Vorobyov, I.V., Huang, N., Roux, B. and MacKerell, A.D. Jr (2005) Determination of electrostatic parameters for a polarizable force field based on the classical drude oscillator. *J. Chem. Theory and Comput.*, **1**, 153–168.
 35. Pigache, A., Cieplak, P. and Dupradeau, F.-Y. (2004) Automatic and highly reproducible RESP and ESP charge derivation: application to the development of programs RED and X RED. *227th ACS National Meeting* Anaheim, CA, USA.
 36. Dupradeau, F.-Y., Pigache, A., Zaffran, T., Grivel, N., Savineau, C., Lelong, R., Lelong, D. and Cieplak, P. (2007) Automatic, highly effective and reproducible RESP and ESP charge derivation, and force field library building: implementation of multi-orientation, multi-conformation and multi-molecule RESP and ESP charge fitting, and Tripos file format generation in the R.E.D. tools. Manuscript in preparation.
 37. Bryce, R. AMBER Parameter Database. The School of Pharmacy & Pharmaceutical Sciences, University of Manchester, Oxford Road, Manchester, M13 9PL, UK. <http://pharmacy.man.ac.uk/amber/> (15 October 2007, date last accessed).
 38. Saro, P. and SantaLucia, J. Jr, RNA Modified Parameters DataBase. <http://ozone3.chem.wayne.edu:8080/Modifieds/> (15 October 2007, date last accessed).
 39. Pérez, A., Marchán, I., Svozil, D., Cheatham, T.E. III, Laughton, C.A. and Orozco, M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. *Biophys. J.*, **92**, 3817–3829.
 40. Song, K., Hornak, V., De Los Santos, C., Grollman, A.P. and Simmerling, C. (2007) Molecular mechanics parameters for the FapydG DNA Lesion. *J. Comput. Chem.*, doi 10.1002/jcc.20625 [Epub ahead of print].
 41. Connolly, M.L. (1983) Analytical molecular surface calculation. *J. Appl. Cryst.*, **16**, 548–558.
 42. Chirlian, L.E. and Francl, M.M. (1987) Atomic charges derived from electrostatic potentials: a detailed study. *J. Comput. Chem.*, **8**, 894–905.
 43. Breneman, C.M. and Wiberg, K.B. (1990) Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J. Comput. Chem.*, **11**, 361–373.
 44. Case, D.A., Darden, T.A., Cheatham, T.E. III, Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Merz, K.M., Pearlman, D.A. *et al.* (2006), AMBER 9. University of California, San Francisco, CA.
 45. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M. (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, **4**, 187–217.
 46. Williams, D.E. (1997) pdm97 <http://www.netsci.org/Resources/Software/Modeling/QM/pdm97.html> (15 October 2007, date last accessed).
 47. Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Montgomery, J.A. Jr, Vreven, T., Kudin, K.N., *et al.* (2004) Gaussian 03, Revision C.02. Gaussian, Inc., Wallingford, CT.
 48. Schmidt, M.W., Baldridge, K.K., Boatz, J.A., Elbert, S.T., Gordon, M.S., Jensen, J.H., Koseki, S., Matsunaga, N., Nguyen, K.A. *et al.* (1993) General atomic and molecular electronic structure system. *J. Comput. Chem.*, **14**, 1347–1363.
 49. Tripos mol2 file format. http://www.tripos.com/index.php?family=modules,SimplePage,,&page=sup_mol2&s=0 (15 October 2007, date last accessed).
 50. Jmol: an open-source Java viewer for chemical structures in 3D, <http://jmol.sourceforge.net/> (15 October 2007, date last accessed).
 51. J2SE Runtime Environment. <http://java.sun.com/j2se/1.5.0/download.html> (15 October 2007, date last accessed).
 52. Wikipedia. http://en.wikipedia.org/wiki/Main_Page/ (15 October 2007, date last accessed).
 53. Dupradeau, F.-Y. and Cieplak, P. Tutorials describing the Ante_R.E.D. and R.E.D. III programs and the R.E.DD.B. Database. <http://q4md-forcefieldtools.org/Tutorial/> (15 October 2007, date last accessed).
 54. Wang, L., Brock, A., Herberich, B. and Schultz, P.G. (2001) Expanding the genetic code of *Escherichia coli*. *Science*, **292**, 498–500.
 55. Dupradeau, F.-Y., Case, D.A., Yu, C., Jimenez, R. and Romesberg, F.E. (2005) Differential solvation and tautomer stability of a model base pair within the minor and major grooves of DNA. *J. Am. Chem. Soc.*, **127**, 15612–15617.
 56. PDB format. http://www.pdb.org/pdb/static.do?p=file_formats/pdb/index.html (15 October 2007, date last accessed).
 57. Dupradeau, F.-Y. and Cieplak, P. The Tripos mol2 file format. <http://q4md-forcefieldtools.org/Tutorial/Tutorial-1.php#15> (15 October 2007, date last accessed).
 58. Walters, P. and Stahl, M. Babel – A Molecular Structure Information Interchange Hub. <http://smog.com/chem/babel> (15 October 2007, date last accessed).
 59. Open Babel: The Open Source Chemistry Toolbox. http://openbabel.sourceforge.net/wiki/Main_Page (15 October 2007, date last accessed).
 60. Schafmeister, C.E.A.F., Ross, W.S. and Romanovski, V. (1995) LEaP University of California, San Francisco, CA.
 61. Murray-Rust, P. and Rzepa, H.S. (2003) Chemical Markup, XML, and the World-Wide Web. 4. CML Schema. *J. Chem. Inf. Comput. Sci.*, **43**, 757–772.