BIOINFORMATICS APPLICATIONS NOTE Vol. 28 no. 5 2012, pages 752–753
doi:10.1093/bioinformatics/bts020

Databases and ontologies

Advance Access publication January 11, 2012

GROMACS molecule & liquid database

David van der Spoel[1,*], Paul J. van Maaren[1] and Carl Caleman[2,*]

[1]Department of Cell and Molecular Biology, Uppsala University, Husargatan 3, Box 596, SE-75124 Uppsala, Sweden and [2]Coherent Imaging Division, Center for Free-Electron Laser Science, Deutsches Elektronen-Synchrotron Notkestrasse 85, DE-22607 Hamburg, Germany

Associate Editor: Jonathan Wren

ABSTRACT

Motivation: The molecular dynamics simulation package GROMACS is a widely used tool used in a broad range of different applications within physics, chemistry and biology. It is freely available, user friendly and extremely efficient. The GROMACS software is force field agnostic, and compatible with many molecular dynamics force fields; coarse-grained, unified atom, all atom as well as polarizable models based on the charge on a spring concept. To validate simulations, it is necessary to compare results from the simulations to experimental data. To ease the process of setting up topologies and structures for simulations, as well as providing pre-calculated physical properties along with experimental values for the same we provide a web-based database, containing 145 organic molecules at present.

Results: Liquid properties of 145 organic molecules have been simulated using two different force fields, OPLS all atom and Generalized Amber Force Field. So far, eight properties have been calculated (the density, enthalpy of vaporization, surface tension, heat capacity at constant volume and pressure, isothermal compressibility, volumetric expansion coefficient and the static dielectric constant). The results, together with experimental values are available through the database, along with liquid structures and topologies for the 145 molecules, in the two force fields.

Availability: The database is freely available under http://virtualchemistry.org.

Contact: spoel@xray.bmc.uu.se; carl.caleman@cfel.de

Received on December 7, 2011; revised on January 2, 2012; accepted on January 5, 2010

1 INTRODUCTION

Molecular dynamics simulations can be a very useful tool in understanding fundamental interactions in simple molecular liquids and crystals as well as complex biomolecular systems. The accuracy of predictions based on molecular dynamics simulations are limited by the applied force field. Force fields are normally developed to reproduce certain specific experimental properties, such as density or energies, or to reproduces known molecular structures. A model that captures the basic physics underlying the atomic interactions can in principle be used to predict outside the realm of parameterization, and this is in fact often done. In order to validate the predictions based on force fields, we have recently performed force field benchmarks of proteins (Lange *et al.*, 2010) and liquids (Caleman *et al.*, 2012).

*To whom correspondence should be addressed.

The latest incarnation of the GROMACS software was described by Hess *et al.* (2008) while a recent overview paper (van der Spoel and Hess, 2011) describes future prospects for the software suite. Here we present a database holding simulation resources, namely liquid structures and topology files. The construction of molecular topologies—describing connectivity, atom types and partial charges—is unfortunately still cumbersome. There is a large user base for the GROMACS software—over 2000 people that subscribe to a mailing list—and therefore we envision that there will be a significant demand for the kind of service the molecule & liquid database provides, since questions about topology building are among the most common ones on the mailing list.

2 METHODS

The database contains GROMACS topology files (.top/.itp) for all the included molecules, for both the Generalized Amber Force Field (GAFF) (Wang *et al.*, 2004) and OPLS/AA (Jorgensen and Tirado-Rives, 2005) force fields. The molecular models were built using Molden (Schaftenaar and Noordik, 2000) and PRODRG (Schuettelkopf and van Aalten, 2004), and optimized using quantum calculations with Gaussian 03 (Frisch *et al.*, 2004) at the Hartree-Fock level employing the 6-311G** basis set (Blaudeau *et al.*, 1997; Krishnan *et al.*, 1980), as provided by the Basis Set Exchange website (Schuchardt *et al.*, 2007).

The GAFF topologies were generated using the Antechamber software (Wang *et al.*, 2004, 2005) using partial charges calculated with the restrained electrostatic potential method (Bayly *et al.*, 1993). The GAFF topologies were converted into GROMACS format using the amb2gmx.pl script (Mobley *et al.*, 2006) (`http://ffamber.cnsm.csulb.edu/`)—note that at present the ACPYPI tool (http://code.google.com/p/acpype/) is the recommended tool for converting from AMBER to GROMACS. OPLS/AA topologies were built using GROMACS tools and verfied manually.

Gas-phase structures as well as liquid structures (after equilibration) are available for all molecules in pdb format. Finally, the database contains simulated and, where available, experimental physical properties (fully referenced) for up to eight different observables for all molecules: the density, enthalpy of vaporization, heat capacities at constant pressure and constant volume, the surface tension, isothermal compressibility, volumetric expansion coefficient and the static dielectric constant. For details about the underlying simulations and analysis, we refer to (Caleman *et al.*, 2012).

All data are stored in a database, using the structured query langauge (SQL) to ensure portability, which is processed using a script to obtain the (at present) completely static website. The website is organized using a relatively simple directory structure with one subdirectory per compound, and within these one subdirectory per force field.

3 DISCUSSION

The molecular modeling community has a lot to gain from easy access to pre-equilibrated simulation structures and topologies that

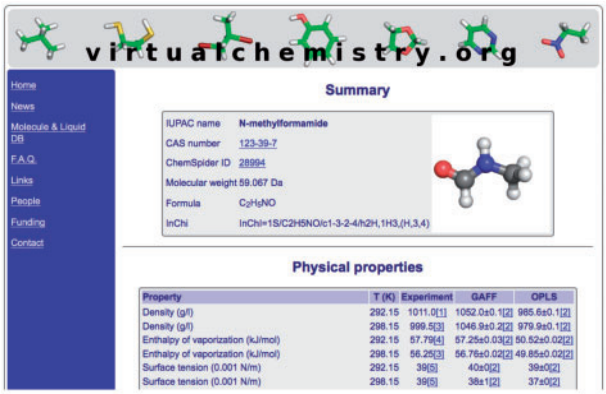

Fig. 1. Example (partial) molecule entry for *N*-methylformamide.

are validated against experimental data. The GROMACS molecule & liquid database freely provides this for a set of over a hundred molecules, an example molecule entry being shown in Figure 1. Presently, two of the most widely used force fields for organic molecules, GAFF (Wang *et al.*, 2004) and OPLS/AA (Jorgensen and Tirado-Rives, 2005) are supported, and there are plans to include models based on the CHARMM generalized force field (Vanommeslaeghe *et al.*, 2010) as well. The expectation is that this test set will help users to establish how good these force fields are not only for the molecules included, but also for similar molecules. For this reason, the test set was picked such that it would include a large variety of chemical moieties. All topologies and structures available have been checked manually and calculated physical properties along with experimental data are provided for each of the the molecules, including corresponding experimental data which is fully referenced. A similar effort providing input files for molecules based on the GROMOS96 force field (van Gunsteren *et al.*, 1996) was announced recently (Malde *et al.*, 2011).

Further molecules, simulation results and corresponding experimental data will be added to the database in the future. The ultimate goal with this work is to build up a large database, that can be used both for getting reliable topologies and liquid structures, and to provide a list of relevant physical properties for comparison to experiments. The reference values provided (Caleman *et al.*, 2012) can also serve as a benchmark for force field improvements, for instance using polarizable models (Caleman and van der Spoel, 2007; Caleman *et al.*, 2011; van der Spoel and Hess, 2011; van Maaren and van der Spoel, 2001).

ACKNOWLEDGEMENTS

We would also like to thank Minyan Hong and Marjolein van der Spoel for help with collecting experimental data.

Funding: Swedish research council (to D.v.d.S.); Helmholtz Association through the Center for Free-Electron Laser Science (to C.C). We acknowledge a grant for computer time (SNIC-022-09/10) through the national supercomputer center (NSC) in Linköping and the parallel computing center (PDC) in Stockholm, Sweden.

Conflict of Interest: none declared.

REFERENCES

Bayly,C.I. *et al.* (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges – the RESP model. *J. Phys. Chem.*, **97**, 10269–10280.

Blaudeau,J. *et al.* (1997) Extension of Gaussian-2 (G2) theory to molecules containing third-row atoms K and Ca. *J. Chem. Phys.,* **107**, 5016–5021.

Caleman,C. and van der Spoel,D. (2007) Evaporation from water clusters containing singly charged ions. *Phys. Chem. Chem. Phys.*, **9**, 5105–5111.

Caleman,C. *et al.* (2011) Atomistic simulation of ion solvation in water explains surface preference of halides. *Proc. Natl Acad. Sci. USA*, **108**, 6838–6842.

Caleman,C. et al. (2012) Force field benchmark of organic liquids: density, enthalpy of vaporization, heat capacities, surface tension, compressibility, expansion coefficient and dielectric constant. *J. Chem. Theor. Comp.*, **8**, 61–74

Frisch,M.J. *et al.* (2004) *Gaussian 03, Revision C.02*. Gaussian, Inc. Wallingford CT.

Hess, B. *et al.* (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, **4**, 435–447.

Jorgensen,W.L. and Tirado-Rives,J. (2005) Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc. Natl Acad. Sci. USA*, **102**, 6665–6670.

Krishnan,R. *et al.* (1980) Self-consistent molecular-orbital methods. 20. Basis set for correlated wave-functions. *J. Chem. Phys.,* **72**, 650–654.

Lange,O.F. *et al.* (2010) Scrutinizing molecular mechanics force fields on the submicrosecond timescale with NMR data. *Biophys. J.,* **99**, 647–655.

Malde,A.K. *et al.* (2011) An automated force field topology builder (ATB) and repository: version 1.0. *J. Chem. Theory Comput.*, **7**, 4026–4037.

Mobley,D.L. *et al.* (2006) On the use of orientational restraints and symmetry corrections in alchemical free energy calculations. *J. Chem. Phys.*, **125**, 084902.

Schaftenaar,G. and Noordik,J.H. (2000) Molden: a pre- and post-processing program for molecular and electronic structures. *J. Comput. Aid. Mol. Des.*, **14**, 123–134.

Schuchardt,K.L. *et al.* (2007) Basis set exchange: a community database for computational sciences. *J. Chem. Inf. Model.*, **47**, 1045–1052.

Schuettelkopf,A.W. and van Aalten,D.M.F. (2004) PRODRG - a tool for high-throughput crystallography of protein-ligand complexes. *Act. Crystallogr. D.*, **60**, 1355–1363.

van der Spoel,D. and Hess,B. (2011) Gromacs - the road ahead. *Comput. Mol. Sci.*, **1**, 710–715.

van Gunsteren,W.F. *et al.* (1996) *Biomolecular Simulation: The GROMOS96 Manual and User Guide*. Hochschuleverlag AG an der ETH Zürich, Zürich, Switzerland.

van Maaren,P.J. and van der Spoel,D. (2001) Molecular dynamics simulations of water with a novel shell-model potential. *J. Phys. Chem. B.*, **105**, 2618–2626.

Vanommeslaeghe,K. *et al.* (2010) Charmm general force field: a force field for drug-like molecules compatible with the charmm all-atom additive biological force fields. *J. Comput. Chem.*, **31**, 671–690.

Wang,J. *et al.* (2004) Development and testing of a general amber force field. *J. Comput. Chem.*, **25**, 1157–1174.

Wang,J. *et al.* (2005) Antechamber, an accessory software package for molecular mechanical calculations. *J. Comput. Chem.*, **25**, 1157–1174.