Short Report

MKTOP: a Program for Automatic Construction of Molecular Topologies

André A. S. T. Ribeiro,* Bruno A. C. Horta and Ricardo B. de Alencastro

Physical Organic Chemistry Group, Departamento de Química Orgânica, Instituto de Química, Universidade Federal do Rio de Janeiro, Cidade Universitária, CT, Bloco A, lab. 609, 21941-909 Rio de Janeiro-RJ, Brazil

A mecânica molecular é amplamente usada na simulação de biomoléculas. Um arquivo de topologia molecular contendo todos os tipos de átomos, que dependem do ambiente químico, tem de ser construído. MKTOP é um software gratuito capaz de construir topologias moleculares para o GROMACS, sendo assim uma ferramenta útil para seus usuários.

Molecular mechanics methods are widely-used for simulations of biomolecules. A molecular topology file containing all atom types, which depend on the chemical environment, must be constructed. MKTOP is a free-software capable of automatic atom type recognition and construction of molecular topologies for GROMACS, being a useful tool for its users.

Keywords: molecular mechanics, atom-type recognition, molecular topology, GROMACS, OPLS

Introduction

The use of molecular mechanics (MM) methods in the treatment of biomolecular systems has been increasing steadily. These methods use empirical force fields to calculate interactions between atoms. Applications range from simple geometry optimizations to estimations of binding affinity, molecular dynamics (MD) simulations, among others.

GROMACS[1] is one of the most utilized MD software packages and it has the advantage of being free software, distributed under the GNU license. It is able to perform calculations with several force fields, one of them being OPLS-AA.[2] This combination is widely-used in protein simulations.

In order to set up a MM calculation, one must generate a topology file for the system, in which the connectivities between the atoms and the atom type, which is determined by the concept of chemical function, are specified.

Chemical function determination is the major problem to address when creating a topology. Most molecular mechanics softwares have routines for generating a topology, but these are usually only effective for peptides. Although the systems of main interest in the field are proteins, possible inhibitors, co-factors, substrates and agonists may not be peptides. Therefore, automatic routines which successfully identify atom types are extremely helpful to the user. One such routine is called PRODRG[3], but it is only able to create topologies for united-atom force fields. More recently, a routine named antechamber,[4] capable of creating all-atom molecular topologies, has been included in the last AMBER[5] distribution. However, AMBER is not freely available. Therefore, a program capable of building topologies for GROMACS would be of great appeal.

In this paper we report the development of an automatic routine, named MKTOP, to create molecular topologies for GROMACS, utilizing the OPLS-AA force-field.

Computational Methods

MKTOP is a user-friendly computational routine written in Perl. A single command line is required to generate a molecular topology.

How MKTOP works

The starting point for constructing a topology is, obviously, a list of the coordinates of all the atoms present in the system; these must be provided to MKTOP in a PDB format file. The atomic charges, which are essential to perform a MM calculation, may be provided by the user in a separate file if he wants them to be written in the topology file; otherwise, the generated topology will have charges equal to zero on all atoms.

*e-mail: aastr@iq.ufrj.br

The connectivities are determined by comparison of the distance between two atoms and a selected value for each bond type. Distances within 0.3 Å of the selected value are considered bonded. This rather large value was arbitrarily chosen to eliminate the need to consider bond orders. Another reason is that *ab initio* methods are generally used to optimize molecular geometries, and different levels of calculation will yield structures with different bond lengths. This criterion can be easily modified by the user, who is presented with a list of bonded atoms in the output. The user can also utilize different criteria for bonds involving specific atoms.

The lists of angles and dihedrals are determined from the bonded atoms list, utilizing two and three combinations of bonded pairs.

Atom types are determined by utilizing several connectivity criteria based on the chemical environment. The program is able to differentiate between ketones, aldehydes, acids, anhydrides, amides, amines, aromatic rings, alcohols, ethers and various other chemical functions. Furthermore, MKTOP is able to recognize characteristic rings such as those present in pyrrole, pyridine, pyrimidine, isoxazole and others.

MKTOP specifies harmonic potentials for bond stretching and angle bending. The equilibrium values are taken from the PDB input. Force constants are taken from the force-field parameter list, for the respective atom types involved. However, the molecule may contain bond or angle terms for which there are no force-field parameters. In these cases, MKTOP tries to use the force constant of a reasonably similar bond/angle term. For example, the force constant for the bond between a benzene carbon atom and an alkene carbon atom is used to describe a pyrrole-type ring carbon - alkene carbon bond. These adaptations are specified in the program in a case-by-case basis and are thus limited. The generic values for carbon-carbon and carbon-hydrogen bonds which are used when no other adaptation is found on the list are those of simple hydrocarbons. For bonds between other atoms, a generic and arbitrary value of 200,000 kJ mol^{-1} nm^{-2} was used. This value is close to a weak bond force constant, such as a carbon-sulfur bond in a thiol. The procedure for angle terms is analogous and a generic value of 200 kJ mol^{-1} rad^{-2} was used. The output presents a list of adaptations made and recommends a parameterization for these terms.

The dihedral terms are specified as of the Ryckaert-Bellemans type,[6] which uses sums of sine functions to describe the potential. Therefore, the minimal values for the potential are determined by the coefficients of the summation and no explicit reference to equilibrium data is made in the topology. Dihedrals for which there are no available parameters are thus not treated by MKTOP; these will be presented to the user by GROMACS.

The improper dihedral list, used to force some groups to planarity, is determined by comparison of each dihedral with a list of atom types which must be restrained, as determined by the force-field.

Testing MKTOP

In order to evaluate the effectiveness of the program, a subset of 54,821 molecules was chosen from the ZINC database.[7] For each structure, PM3 charges were calculated utilizing GAMESS-US[8] (version of 24/03/2007). The charges obtained for 694 structures were not reasonable and were thus excluded. The determination of the topologies for the remaining 54,127 structures was done using MKTOP and took approximately 7 hours on a 3.0 GHz Intel Xeon Dell PowerEdge 1800 Red Hat Linux system.

Geometry optimizations utilizing the ZINC database geometry and the determined topology were carried out on GROMACS (version 3.3.1) with steepest descent algorithm and maximum force 25 kJ mol^{-1} nm^{-1}.

Results and Discussion

The final structures obtained from the energy minimizations were compared with the respective initial structure. The RMSD distributions for coordinates (Figure 1), bond lengths (Figure 2) and angles (Figure 3) were calculated. The average RMSD values obtained were 0.201 Å for coordinates; 0.039 Å for bond lengths and 2.75° for angles.

The results obtained indicate that the final geometries are reasonably similar to the initial ones, reflecting MKTOP's ability to accurately describe different chemical environments in a large set of molecules.
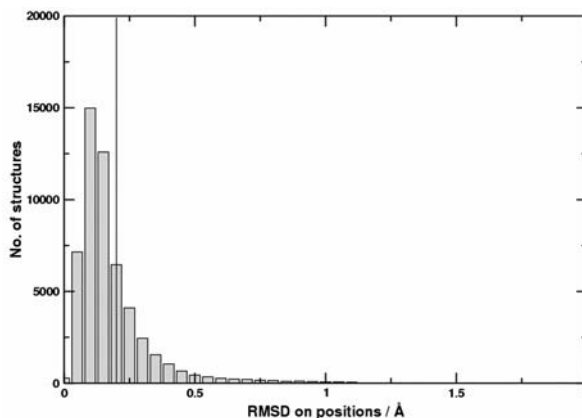

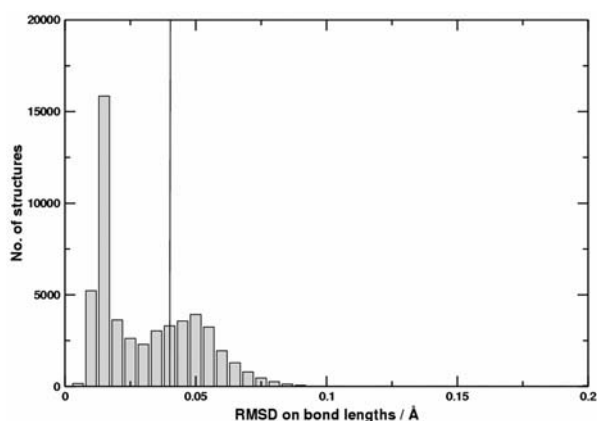**Figure 1.** RMSD distribution for coordinates. The vertical line indicates the average value.

Figure 2. RMSD distribution for bond lengths. The vertical line indicates the average value.
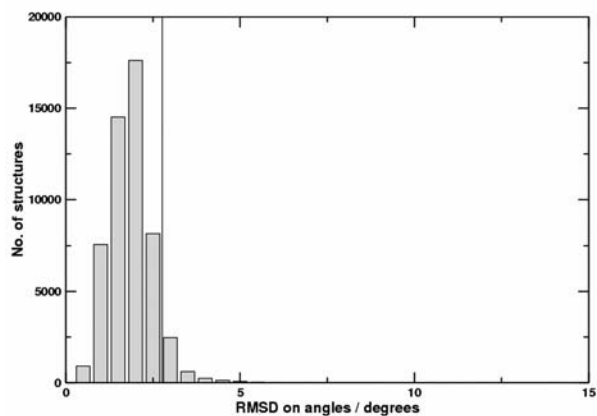

Figure 3. RMSD distribution for angles. The vertical line indicates the average value.

Conclusions

The results indicate that the topologies created are adequate and MKTOP describes successfully the atom types of several molecules, being a good choice for automatic non-peptide topology determination. The topologies thus created should be refined to achieve an extremely accurate account of the equilibrium geometry; however, MKTOP provides a reasonable and useful first approximation. Future versions of MKTOP will include support for other force-fields.

MKTOP is available for download free-of-charge at http://labmm.iq.ufrj.br/mktop/

Acknowledgments

References

1. Lindahl, E.; Hess, B.; van der Spoel, D.; *J. Mol. Mod.* **2001**, *7*, 306.

2. Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L.; *J. Phys. Chem. B* **2001**, *105*, 6474.

3. Schüttelkopf, A. W.; van Aalten, D. M. F.; *Acta Cryst. D* **2004**, *60*, 1355.

4. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A.; *J. Mol. Graph. Mod.* **2006**, *25*, 247.

5. Case, D. A.; Cheatham, T.; Darden, T.; Gohlke, H.; Luo, R.; Merz, Jr., K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R.; *J. Comput. Chem.* **2005**, *26*, 1668.

6. Ryckaert, J. P.; Bellemans, A.; *Far. Disc. Chem. Soc.* **1978**, *66*, 95.

7. Irwin, J. J.; Shoichet, B. K.; *J. Chem. Inf. Model.* **2005**, *45*, 177.

8. 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.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A.; *J. Comput. Chem.* **1993**, *14*, 1347.