OBGMX: A Web-Based Generator of GROMACS Topologies for Molecular and Periodic Systems Using the Universal Force Field

Giovanni Garberoglio* [a]

OBGMX is a web service providing topologies for the GROMACS molecular dynamics software package according to the Universal Force Field, as implemented in the Open Babel package. OBGMX can deal with molecular and periodic systems. The geometrical parameters appearing in the potential energy functions for the bonded interactions can be set to those measured in the input configuration. The performance of OBGMX in reproducing the structure of periodic systems is analyzed by calculating the root mean-squared displacements of optimized configurations of a large set of metal-organic frameworks. OBGMX is available at http://software-lisc.fbk.eu/obgmx/. © 2012 Wiley Periodicals, Inc.

DOI: 10.1002/jcc.23049

Introduction

The enormous increase in the computational power of digital processors during the past years has made it possible to perform atomistic simulations of systems made by millions of particles, or to follow the dynamics of medium-sized biomolecules for time scales approaching the microsecond. These achievements are also based on the availability of mature and free software tools that are capable of harnessing the power of modern parallel computers to perform these calculations and analyze the resulting trajectories. Programs such as the large-scale atomic/molecular massively parallel simulator (LAMMPS) [1] or DL_POLY [2] are general-purpose molecular dynamics (MD) engines that can be applied to study a considerable variety of problems in condensed matter systems. The development of software has been paralleled by that of classical force fields capable to describe the bonded and nonbonded interactions of a wide range of molecular and periodic systems. In the jargon of MD simulations, the specification of all the interactions among the force centers in a molecule is called the molecular topology.

The temporal and spatial scales of the problems that can be successfully studied with MD methods makes them suitable to investigate complex biomolecular systems, such as small proteins or DNA. Indeed, some of the most popular MD programs (such as CHARMM [3], NAMD [4], AMBER [5] or GROMACS [6]) have been developed specifically to study biopolymers and thus offer some automated way to setup molecular topologies in these cases. Nevertheless, all these programs are general-purpose in spirit and can therefore be used to investigate a variety of systems much larger than biomolecular ones. Among these programs, GROMACS is generally considered to be the fastest: its considerable performance comes from the use of single-precision floating-point arithmetic and from the fact the most time-consuming routines have been written in assembly, harnessing the single-instruction multiple-data vector instructions present in modern processors. Moreover, GROMACS comes with a quite complete set of software tools for the analysis of the MD trajectories. For these reason, we have chosen to consider this software in the following.

In the case of non-biomolecular systems, the generation of molecular topologies is a task which is left almost entirely on the user. Although topologies can be straightforwardly written from scratch in the case of the simplest molecules by carefully following the prescriptions detailed by a particular force field, this procedure becomes very lengthy and error prone when the number of atoms increases. Only a limited number of free tools are available to alleviate this burden. Many of them, such as PRODRG [7] or the Automated Force Field Topology Builder [8] are web-based services, although some programs, such as MKTOP [9] have also been developed.

However, all the freely available tools are limited to organic molecules, and cannot be used to set up topologies for crystalline structures. To the best of our knowledge, only commercial software suites are able to perform classical simulations of periodic and molecular systems with an automated setup of the molecular topology.

To overcome the current limitations of freely available topology generators, we decided to use the Universal Force Field (UFF) [10] which has been developed to provide a reasonably accurate description for systems containing all the elements of the periodic table. Similarly to other force fields, UFF is based on the concept of atom type, which corresponds to an element in a particularly chemical environment. For example, the carbon atom in methane has different properties from the carbon atoms in benzene, and therefore the two correspond to
different atom types (C_3 and C_R in the UFF nomenclature, respectively).

As a consequence, the first step in the automated creation a topology involves the assignment of types to the actual atoms present in the system. In the case of UFF, the completion of this step enables to assign the nonbonded parameters of the force field. The nonbonded interactions in UFF are of the Lennard–Jones type. The atom-type assignment is then followed by the analysis of the molecular connectivity and the application of the UFF rules to assign all the bonded interactions present in the system (that is, stretching, angular, dihedral, and inversion potentials).

In the case of OBGMX, the very efficient routines provided by the Open Babel package are used to perform these tasks. The current release of Open Babel (version 2.3.1), already comes with an implementation of UFF, which is, however, limited to molecular systems. We have extended Open Babel to deal also with periodic systems, and we have developed a code to translate Open Babel’s UFF parametrization into molecular topologies to be used as input for the GROMACS package.

The resulting program, OBGMX, can be accessed via the World Wide Web at the address http://software-lisc.fbk.eu/obgmx/. In the following sections, we will present the details of our implementation, and we will discuss the performance of UFF in describing the equilibrium structure of large library of metal-organic frameworks (MOFs). The root mean-squared displacement (RMSD) between UFF optimized configurations and X-ray structural data will be calculated and discussed.

Software Design and Methods

OBGMX produces molecular topologies for the GROMACS MD program, starting from a file with the coordinates and types of all the atoms involved. Both isolated molecules or periodic systems can be processed: in the latter case, unit-cell parameters must be present in the input file.

OBGMX uses Open Babel to interpret the input file and can therefore process a large number of chemical formats: in the case of files with unit-cell parameters, OBGMX assumes that the system is periodic and a suitably patched version of Open Babel is used in this case. The recognition of the UFF atom types and the assignment of UFF’s bonded and nonbonded parameters are also performed using Open Babel’s routines.

However, in the case of bonded interactions not all the functional forms prescribed by UFF are available in GROMACS. In this case, UFF prescriptions are approximated as detailed below.

Bond stretching potential

Both UFF and GROMACS use the same form of harmonic bond-stretching potential, and so no approximation is required in this case.

Angle bending potential

The angle-bending energy term in Open Babel’s UFF depends on the coordination number of the atom at the vertex of the angle. The various possibilities are:

linear $\rightarrow k_{\alpha}(1 + \cos \theta)$

triangular $\rightarrow \frac{2k_{\alpha}}{9}[1 + 4 \cos \theta(1 + \cos \theta)]$

tetrahedral $\rightarrow \frac{k_{\alpha}}{2 \sin^{2} \alpha} (\cos \theta - \cos \alpha)^{2}$

square planar or octahedral $\rightarrow k_{\alpha}[\cos \theta - \cos \beta](\cos \theta - \cos(2\beta))^2$

pentagonal bipyramidal $\rightarrow k_{\alpha}[(\cos \theta - \cos \beta)(\cos \theta - \cos(2\beta))^{2}$

where $\cos \alpha = -1/3$ ($\alpha \simeq 109.47^\circ$) and $\beta = 2\pi/5 = 72^\circ$. GROMACS has only two possible angular forms, that is,

harmonic $\rightarrow \frac{1}{2}k_{\alpha}^{\text{harm}}(\theta - \theta_0)^2$

G96 $\rightarrow \frac{1}{2}k_{\alpha}^{\text{G96}}(\cos \theta - \cos \theta_0)^2$

which can both be characterized by an equilibrium angle ($\theta_0$) and the second derivative of the potential at equilibrium ($k_{\alpha}^{\text{harm}}$ for the harmonic angles, $k_{\alpha}^{\text{G96}}/\sin^{2} \theta_0$ for the G96 angles). Using this principle, the UFF to GROMACS parameter correspondence becomes, in the case of the harmonic potential:

linear $\rightarrow k_{\alpha}^{\text{harm}} = k_{\alpha}, \theta_0 = \pi$

triangular $\rightarrow k_{\alpha}^{\text{harm}} = \frac{4k_{\alpha}}{3}, \theta_0 = \frac{2\pi}{3}$

tetrahedral $\rightarrow k_{\alpha}^{\text{harm}} = k_{\alpha}, \theta_0 = \alpha$

square planar or octahedral $\rightarrow k_{\alpha}^{\text{harm}} = 2k_{\alpha}, \theta_0 = \frac{\pi}{2}$

pentagonal bipyramidal $\rightarrow k_{\alpha}^{\text{harm}} = 2k_{\alpha} \sin^{2} \beta(\cos \theta - \cos(2\beta))^2$, $\theta_0 = \beta$

In the case of the G96 potential, $\theta_0$ retains its meaning as the equilibrium angle, whereas one has

$$k_{\alpha}^{\text{G96}} = \frac{k_{\alpha}^{\text{harm}}}{\sin^{2} \theta_0}.$$  (1)

Notice that when $\theta_0 = \pi$, the first nonzero term of the Taylor series of the G96 potential is $O((\theta - \theta_0)^3)$, whereas the UFF potential would be $O((\theta - \theta_0)^2)$. In this case, we will use the harmonic angular potential.

The G96 and UFF angular potential functions coincide in the case of tetrahedral coordination. As both potentials are periodic, OBGMX prefers the G96 form over the harmonic one, although the latter is available as an option for the interested user. In the case non-tetrahedral coordination, the two angle-bending energies have a different functional dependence on the angle $\theta$. However, the qualitative shape of the two functions is quite similar. As shown in Figure 1, the UFF and G96 potential curves in the very common case of triangular coordination. Recalling that typical values of $k_{\alpha}$ are of the order of 500 kJ/mol, and that $k_{\beta} \approx 2.5$ kJ/mol at room temperature, it is not expected that the difference between the G96 and UFF bending potential could lead to sensible differences in many situations.
whereas GROMACS uses the harmonic form

\[ V_{\text{GMX torsion}} = k_\phi \left[ 1 + \cos \left( n\phi_0 - \phi \right) \right] \]

Conversely, the GROMACS functional form is

\[ V_{\text{GMX torsion}} = k_\phi \left[ 1 + \cos \left( n\phi_0 - \phi \right) + \sin \left( n\phi_0 \right) \sin \left( n\phi \right) \right] \]

so that the UFF form is recovered in GROMACS by choosing:

\[ k_\phi = \frac{1}{2} V_\phi \]

(4)

\[ \phi_s = n\phi_0 - \pi. \]

(5)

Inversions

The UFF and GROMACS definition of the inversion angle \( \phi \) are the same. However, UFF prescribes a two-term Fourier expansion of the potential in the form:

\[ V_{\text{UFF}}(\phi) = K \left[ c_0 + c_1 \cos \phi + c_2 \cos(2\phi) \right], \]

(6)

whereas GROMACS uses the harmonic form

\[ V_{\text{GMX}} = \frac{1}{2} k_\phi (\phi - \phi_0)^2. \]

(7)

Using the same procedure outlined in the case of the bending potential, one finds out that the mapping between UFF and GROMACS parameters are:

\[ k = 4c_2 \left[ 1 - \left( \frac{c_1}{4c_2} \right)^2 \right] K \]

(8)

\[ \phi_0 = \pi - \arccos \left( \frac{c_1}{4c_2} \right). \]

(9)

**Dispersion potential**

In the case of nonbonded parameters, the form of the dispersion interaction used by UFF has the same Lennard–Jones dependence

\[ V(r) = 4\epsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right] \]

(10)

as that available in GROMACS, and so no further treatment is needed.

**Optional features**

OBGMX provides additional choices that can be used to generate the GROMACS topology.

The first is the choice of the combination rule. UFF prescribes that the geometrical average should be used to combine both the \( \sigma \) and \( \epsilon \) parameters of the Lennard–Jones potential for unlike atom types. The user can choose to use the Lorentz–Berthelot combination rules, resulting in arithmetic average of the \( \sigma \) parameter and geometric average for the \( \epsilon \) parameter.

The second choice is related to the weight to be assigned to the 1–4 dispersive interactions. The UFF paper implicitly assumes that the full interaction should be used for atoms separated by three bonds. OBGMX users are given the option to weigh the 1–4 interaction with a parameter \( 0 \leq \xi \leq 1 \).

OBGMX can optionally assign the geometric parameters of the force field—that is, bond lengths, average angles, and dihedrals—from those present in the input geometry. These options are meant to be used in those cases where UFF lacks a proper parametrization for specific atom types. As shown below, this option may give a better starting point for the dynamical analysis of periodic crystals.

UFF has been originally parameterized without using partial atomic charges, but the original paper suggests to use partial charges obtained with the QEq charge-equilibration scheme. This algorithm for partial-charge assignment is indeed present in the latest version of the Open Babel package, but we noticed that it tends to assign partial charges having a very large magnitude when compared with charge population studies performed using the results of \textit{ab initio} quantum chemical calculations. For this reason, OBGMX does not currently provide any method for atomic charge calculation, leaving this task to the end user.

**Performance: Structure of Metal-Organic Frameworks**

The generation of GROMACS topologies for periodic systems is a distinctive feature of OBGMX. Moreover, the performance of UFF is well known in the case of molecular systems, but—to the best of our knowledge—there is only a single and very recent study dealing with the performance of UFF for periodic systems, where a very good similarity between UFF-optimized configurations and actual X-ray data has been observed.

---

Figure 1. The angular bending energy for triangular coordination in the case of UFF (dashed line) together with the approximating G96 expression (solid line), in the case \( n = 1 \). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
To estimate the performance of UFF in describing the structure of periodic systems, we used OBGMX to perform structural optimizations of a large and representative set of MOFs. The set considered here includes the original series of Isoreticular MOFs,[16] Covalent Organic Frameworks,[17] Zeolitic Imidazolate Frameworks,[18,19] breathing MOFs[20–22] and the molecular sieve Zn-TBIP.[23]

The starting configurations for the optimizations were taken as the X-ray structure indicated in the original papers. These configurations were replicated to have simulation cells capable of accommodating a cutoff of 17 Å, and subsequently optimized using the conjugate gradient method available in GROMACS. We used OBGMX to generate two kinds of molecular topologies: the first kind was calculated according to the prescription of UFF, whereas in the second topology, the values of the equilibrium bond lengths and angles were taken from the provided geometry.

The RMSDs between the initial and final configuration were calculated, and they are shown in Figure 2. In the calculation of the RMSD between two periodic structures, we translated the positions of the atoms in the second structure so that the first atom in the configuration file had the same position as in the first structure and we did not attempt to perform any rigid rotation (as is customary in the calculation of RMSD of molecular systems).

In the majority of the cases, the RMSD between the X-ray structure and the UFF-optimized one is less than \( \sim 0.7 \) Å, although values as high as 1.5 Å were observed. Our values of the RMSD are larger than those quoted by Wilmer et al.,[15] where the typical RMSD observed was of the order of 0.1 Å. We notice, however, that we used a different definition of the RMSD, and that they performed the comparison on only four MOFs, observing an RMSD of \( \sim 0.7 \) Å in one case.

In the case of ZIF-5, the compound with the highest RMSD observed here, the origin of the discrepancy can be traced back to the presence of octahedrally coordinated indium atoms in the structure. The current implementation of UFF does not recognize this atom type, and erroneously assigns the parameter for a tetrahedrally coordinated In, resulting in a very distorted structure after minimization.

In this and other cases, the assignment of bond and angular equilibrium parameters from the provided structure usually reduces the RMSD of the optimized structure (by roughly a factor of two), as might be expected. In the case of ZIF-5, this approach correctly assigns octahedral coordination to the In atoms (with roughly 90° equilibrium angles). Notice that in this case, the remaining discrepancy between the UFF-optimized structure and the X-ray structure is due to the effect of dispersion and torsional forces.

In almost all the cases, visual comparison between the X-ray and optimized structures shows that the overall disposition of the atoms within the unit cell is not changed. The RMSD mostly comes from relatively small atomic rearrangements during the optimization process.

Finally, we checked the stability of our implementation by running a series of MD runs on a selected subset of the MOFs considered so far. The equation of motions have been integrated into the isothermal ensemble, with a fixed temperature of \( T = 300 \) K, using a timestep of 1 fs and generating 200-ps long trajectories, starting from the GROMACS optimized configurations.

We extracted from each trajectory 2000 configurations evenly separated by 0.2 ps and calculated for each of them the RMSD with respect to the MOF structure obtained using X-rays. The results, shown in Figure 3, show that the topologies generated with OBGMX oscillate steadily with a typical RMSD width of \( \sim 0.4 \) Å, showing no sign of instability.

**Conclusions**

In this article, we presented OBGMX, an automated software tool for the automatic generation of GROMACS topologies.
using the UFF. OBGMX is based on a modified version of the Open Babel library, which has been extended to deal with periodic structures.

We assessed the effectiveness of UFF to describe flexibility in MOFs. In the small but representative sample of MOF structures that have been considered in this work, the root mean-squared distance between the UFF-optimized and the actual structure—as determined by X-ray experiments—is typically less than 0.7 Å, with a maximum value close to 1.5 Å observed in one case. When structural information is used to refine the UFF parameters, the distance between the optimized and the actual structures is reduced by a factor of roughly two. This procedure also allows to overcome some of the limitations of UFF, which does not currently provide bonded parameters for all the possible coordination states of the metal atoms encountered in widely used MOFs. The topologies generated with OBGMX have been shown to produce stable MD trajectories.

Keywords: molecular topologies · GROMACS · universal force field · periodic systems · metal-organic frameworks

How to cite this article: G. Garberoglio, J. Comput. Chem. 2012, 00, 000–000. DOI: 10.1002/jcc.23049