



MDWiZ: A platform for the automated translation of molecular dynamics simulations



Victor H. Rusu^{a,b,1}, Vitor A.C. Horta^{a,c,1}, Bruno A.C. Horta^{d,e}, Roberto D. Lins^b,
Riccardo Baron^{a,*}

^a Department of Medicinal Chemistry, College of Pharmacy, and The Henry Eyring Center for Theoretical Chemistry, The University of Utah, Salt Lake City, UT 84112-5820, USA

^b Department of Fundamental Chemistry, Federal University of Pernambuco, Cidade Universitária, Recife, PE 50740-560, Brazil

^c Department of Computational Sciences, Federal University of Juiz de Fora, Juiz de Fora, Brazil

^d Department of Electrical Engineering, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, Brazil

^e Department of Biotechnology and Biological Sciences, Centro Universitário Estadual da Zona Oeste, Rio de Janeiro, Brazil

ARTICLE INFO

Article history:

Received 23 September 2013

Received in revised form

10 December 2013

Accepted 17 December 2013

Available online 25 December 2013

Keywords:

Software

Force field

Language barrier

Molecular simulation

MD Wizard

Validation

ABSTRACT

A variety of popular molecular dynamics (MD) simulation packages were independently developed in the last decades to reach diverse scientific goals. However, such non-coordinated development of software, force fields, and analysis tools for molecular simulations gave rise to an array of software formats and arbitrary conventions for routine preparation and analysis of simulation input and output data. Different formats and/or parameter definitions are used at each stage of the modeling process despite largely contain redundant information between alternative software tools. Such Babel of languages that cannot be easily and univocally translated one into another poses one of the major technical obstacles to the preparation, translation, and comparison of molecular simulation data that users face on a daily basis. Here, we present the MDWiZ platform, a freely accessed online portal designed to aid the fast and reliable preparation and conversion of file formats that allows researchers to reproduce or generate data from MD simulations using different setups, including force fields and models with different underlying potential forms. The general structure of MDWiZ is presented, the features of version 1.0 are detailed, and an extensive validation based on GROMACS to LAMMPS conversion is presented. We believe that MDWiZ will be largely useful to the molecular dynamics community. Such fast format and force field exchange for a given system allows tailoring the chosen system to a given computer platform and/or taking advantage of a specific capabilities offered by different software engines.

© 2013 The Authors. Published by Elsevier Inc. Open access under [CC BY-NC-ND license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

The use of computer simulations is increasingly relevant in various fields of molecular sciences. As a result, users from different backgrounds become interested in performing molecular simulations despite little formal training. A variety of popular molecular dynamics (MD) simulation packages such as AMBER [1], CHARMM [2], DESMOND [3], DLPOLY [4–6], GROMACS [7–9], GROMOS [10–12], LAMMPS [13], NAMD [14], NWChem [15], OpenMM [16], TINKER [17] were independently developed in the last decades

to match diverse scientific needs. A brief history of the origins of this software development was presented [18]. These software engines allow performing MD simulations with alternative underlying force fields and numerous simulations setups, which are often difficult to be compared within single simulation packages. For example, this leads to the problem of validating results from different research groups or correctly reproducing data generated with force fields that have alternative mathematical forms of the energy terms, which requires entirely new software implementations to be performed [19].

In various fields of science, researchers agreed on general format requirements for sharing and validating data of common interest to the field. For example, structural biologists agreed upon the Protein Data Bank format as a common text format to store atomic level information of molecular structures. However, such a common ground has not yet been reached in the field of chemical theory and computation. The independent and non-cooperative development of software, force fields, and analysis tools for molecular simulations gave and still gives rise to an array of equally popular software

* Corresponding author. Tel.: +1 801 585 7117.

E-mail address: r.baron@utah.edu (R. Baron).

¹ Identical author contributions.

formats to prepare the simulation input and generate the trajectories. Consequently, an increasing number of researchers spend an unreasonable amount of time to convert data that largely contains redundant type of information among different software tools (e.g. a dihedral potential term that represent a same energy function, yet is expressed following different dihedral conventions).

This Babel of computational modeling languages cannot be easily and univocally translated one into another leading to one of the major technical obstacles to the preparation, translation, comparison, and public accessibility of simulation data that users and developers face on a daily basis. Indeed, useful tools to aid the automated preparation of simulation topologies have been proposed. Yet, they either help users of specific software engines [20–22] or aid solely structural/chemical file format conversion (e.g. the Babel family of tools [23]). The latter approach is extremely valuable for a number of modeling scenarios, yet it is not sufficient to translate input files between molecular simulation software engines using alternative definitions of the underlying parameters and potentials. Therefore, a major problem in the molecular simulation field remains the translation of input files and formats and a rigorous assessment of what a force field comparison involves. This study is our first step to contribute with a practical solution to some of these issues.

Aiming to overcome this limitation, we present the MDWiZ platform, an online portal designed to perform fast and reliable preparation (e.g. input, topology, coordinate file formats) and parameter conversion (e.g. actual parameters in those files) across molecular dynamics software engines. The translation includes force fields and models with different underlying potential forms and any required format conversion (including different number of input files) for the desired software. The general structure of MDWiZ is presented and the features of version 1.0 are detailed. MDWiZ allows a user to transform in one single click a set of input files for a given combination of software engine and force field into various alternative combinations, as summarized by the workflow in Fig. 1. This release focuses on translating input files from GROMACS to LAMMPS using different force fields followed by an extensive validation. MDWiZ is freely accessible to the public in the form of an openly accessible web interface, which includes features for user support, tutorials, and examples. The source code is also available upon request to the authors.

2. Materials and methods

The diverse set of molecular systems simulated for validation of MDWiZ, ranging from methane to proteins, is summarized in Fig. 2. Initial coordinates for the methane, ethane, propane and butane were generated with GaussView (version 5) [24]. N-methylacetamide and alanine dipeptide coordinates were generated manually. Initial coordinates for the solute atoms of the Trp-cage mini-protein were taken from the structure by Neidigh et al. [25] (PDB ID: 1L2Y) and the coordinates for the Triclinic Hew Lysozyme from the structure by Walsh et al. (PDB ID: 4LZT) [26]. Each system was independently simulated under vacuum and solvated conditions. Table 1 summarizes the system set up for the simulations in solution. In all cases, standard preparation, minimization, and simulation procedures were performed using the GROMACS software (version 4.6.1; compiled in double precision) [7–9]. We note that MDWiZ is also compatible with GROMACS 4.5.X versions.

A variety of force fields and simulation setup were compared in this study, as summarized in Table 2 and Supporting Information, Tables S1 and S2. Validation of MDWiZ conversions from GROMACS to LAMMPS was carried out using three different force fields. The GROMOS parameter set 53A6 was used [27] with the SPC water

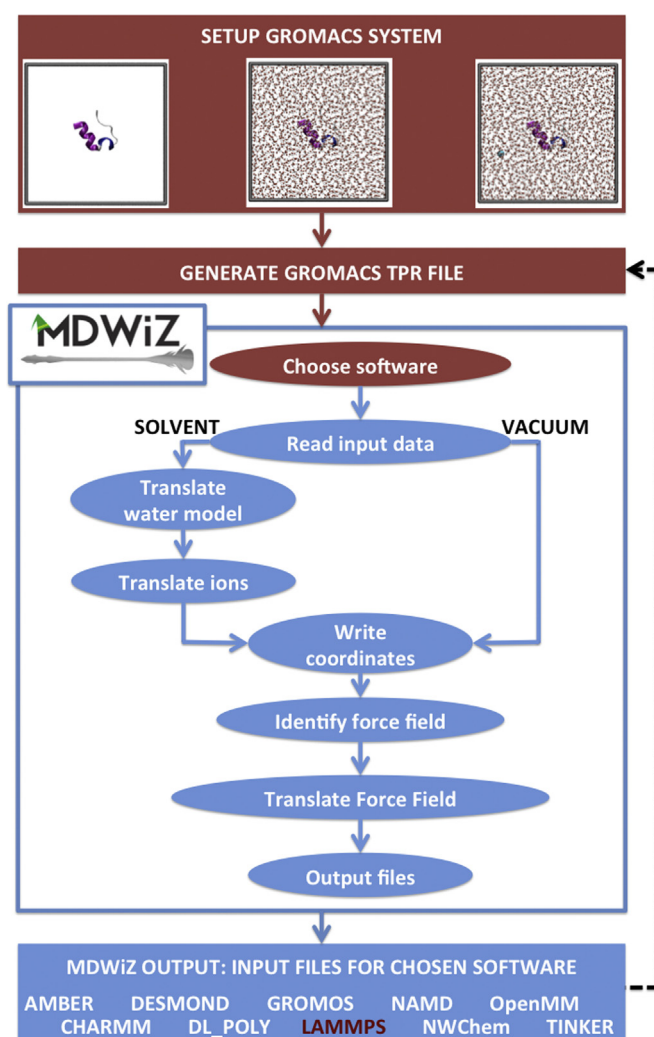


Fig. 1. Philosophy underlying the MDWiZ platform for the automated preparation and translation of molecular dynamics simulations. The workflow highlights the user-defined steps (red) and the steps performed by MDWiZ without user intervention (blue). The released MDWiZ version 1.0 uses a GROMACS input for conversion between the GROMACS and LAMMPS software engines. MDWiZ automatically converts all files needed to run a simulation with a different engine, which might include various combinations of topology, parameters, coordinates, and input control information. In the future releases MDWiZ will be extended to other software engines and to the automated back-conversion to a GROMACS input set (dashed line) starting from various simulation setup formats.

model [28] and compatible ion parameters [29]; the AMBER ff99SB [30] parameter set was chosen as representative of the force field family with the TIP3P water model [31] and compatible ion parameters [1,29]; transformation using the OPLS all atoms force field was employed [32] with the TIP3P water model [31] and compatible ion parameters [29]. The initial configurations were solvated under rectangular periodic boundary conditions in a (pre-equilibrated) water box large enough to avoid interactions between mirror images along the entire MD trajectory. A first steepest-descent energy minimization was performed in order to relax the initial solvent and ion configuration and to eliminate any residual strain. MD simulations were initialized from the energy-minimized configurations with atomic velocities taken from Maxwell–Boltzmann distributions at 300 K. All simulations were carried out in the N, V, T ensemble by separately coupling the temperature of solute and solvent degrees of freedom to a heat bath through a weakly coupled algorithm [33] (relaxation time 0.6 ps). Newton's equations of motion were integrated using the leap-frog algorithm [34] with a

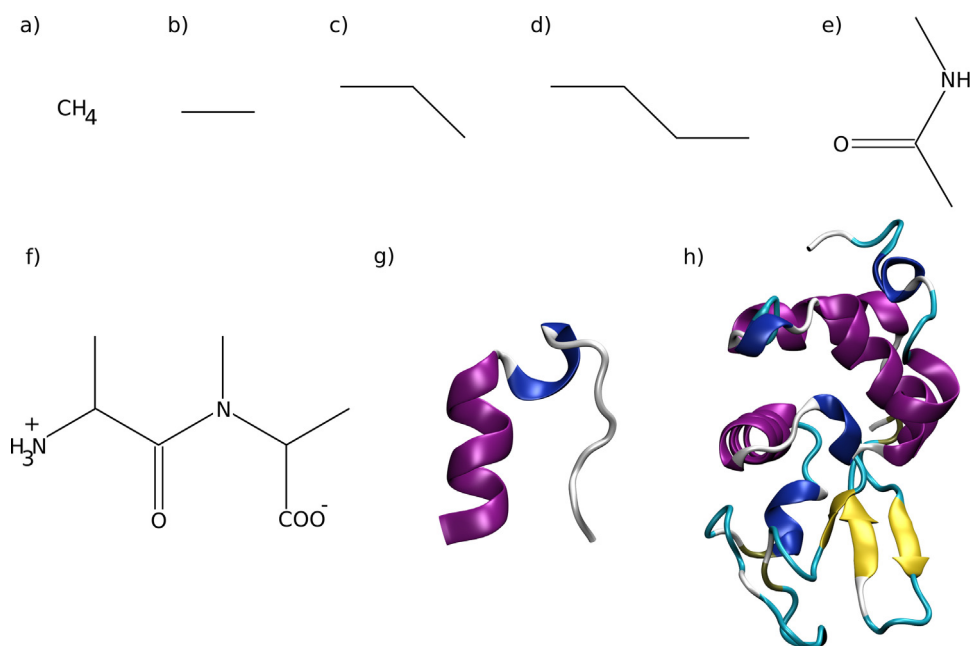


Fig. 2. Graphical representation of the molecular systems used for benchmark and validation of MDWiZ version 1.0. (a) Methane, (b) ethane, (c) propane, (d) butane, (e) N-methylacetamide (NMA), (f) alanine dipeptide, (g) the Trp-cage mini-protein (1L2Y), and (h) the Triclinic Hew Lysozyme (4LZT). Each system was simulated in both vacuum and solvated conditions. See Table 1 for a summary of the explicit solvent MD simulations.

Table 1

Summary of the explicit solvent MD simulations used for validation of MDWiZ. The values for the united-atoms GROMOS force field are given in parentheses.

System	Total number ^a of particles	Number of water molecules	Number of ions	Simulated time/ns
Methane	6533 (6529)	2176	0	10
Ethane	6533 (6527)	2175	0	10
Propane	6533 (6525)	2174	0	10
Butane	6536 (6526)	2174	0	10
Alanine dipeptide	6548 (6537)	2175 (2174)	0	10
N-methylacetamide	6537 (6534)	2175 (2176)	0	10
1L2Y (Trp-cage)	12415 (21370)	4037 (7057)	1 ^b	10
4LZT (Lysozyme)	21527 (20869)	6522 (6512)	9 ^b	10

^a Numbers in brackets refer to the united atom GROMOS force field.

^b Chloride ion(s).

1 fs time step. The water molecules were kept rigid using the SETTLE algorithm [35] in the GROMACS and using the SHAKE algorithm [36] in LAMMPS. To avoid inconsistent comparison of simulations in GROMACS and LAMMPS, because different algorithms are currently implemented for approximation of the long-range interaction algorithm types in GROMACS and in LAMMPS, no long-range interaction was taken into account during the runs. Instead, all non-bonded interactions were directly calculated within a cut-off of 1.4 nm [37]. In all cases, the non-bonded interaction list was updated at every step. MD snapshots were extracted every 1 ps from each trajectory in full precision and used for analysis.

Supplementary material related to this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jmgm.2013.12.006>.

We note that to reproduce the results presented in this study, the double precision trajectories from GROMACS (.trr files) are needed, as using compressed trajectories would lead to numerical inconsistencies. Conversion from GROMACS to LAMMPS trajectories was performed using the VMD software [38]. These double precision trajectories were used as input for LAMMPS rerun calculations. As described in Section 3, this procedure was undertaken in order to allow a direct validation of the transformation performed by MDWiZ, as identical structures are employed to calculate the energy terms of Table 2 and Supporting Information, Tables S1 and S2. Relative accuracy up to 10⁻⁹ kJ mol⁻¹ was

used for all calculations. During the LAMMPS rerun calculations the force field energies were printed every step to compare to GROMACS ones. VMD [38], inkscape (inkscape.org), and xmgrace (<http://plasma-gate.weizmann.ac.il/Grace/>) were used to prepare the graphical representations. The first release of MDWiZ (version 1.0) can be freely accessed through a web interface at: <http://barongroup.medchem.utah.edu/tools>, which includes features for user support, tutorials, and examples. MDWiZ source code is programmed in C language in order to maintain direct transferability with the GROMACS libraries, and is available upon request to the authors.

3. Results

The philosophy underlying MDWiZ is summarized in the workflow of Fig. 1. MDWiZ performs the automated translation between several input sets in terms of formats and actual parameters. MDWiZ automatically detects which of the format conversions and parameter translations are required based on the files provided by the user as a start, and the conversion of choice. In one simple user step, all input files for a given simulation set up are converted in corresponding input for a different software engine and/or force field. Parameters for force field terms with identical underlying potential forms may be expressed using different measure units or

Table 2
Summary of MDWiZ validation for simulations with GROMACS and LAMMPS engines using different force fields. Relative accuracy up to 10^{-9} kJ mol $^{-1}$ was used for all calculations; digits following 10^{-2} are not reported as they are zeros.

System	Software	Force field	Potential energy ratio (kJ mol $^{-1}$)	Bond energy ratio (kJ mol $^{-1}$)	Angle energy ratio (kJ mol $^{-1}$)	Dihedral energy ratio (kJ mol $^{-1}$)	Improper energy ratio (kJ mol $^{-1}$)	Coulomb energy ratio (kJ mol $^{-1}$)	van der Waals energy ratio (kJ mol $^{-1}$)
Methane	GROMACS/LAMMPS ^a	GROMOS	1.01	n/a	n/a	n/a	n/a	1.01	1.00
	GROMACS/LAMMPS	AMBER	1.00	1.00	1.00	n/a	n/a	1.00	1.00
	GROMACS/LAMMPS	OPLS/AA	1.00	1.00	1.00	n/a	n/a	1.00	1.00
Ethane	GROMACS/LAMMPS ^a	GROMOS	1.01	1.00	n/a	n/a	n/a	1.01	1.00
	GROMACS/LAMMPS	AMBER	1.00	1.00	1.00	1.00	n/a	1.00	1.00
	GROMACS/LAMMPS	OPLS/AA	1.00	1.00	1.00	1.00	n/a	1.00	1.00
Propane	GROMACS/LAMMPS ^a	GROMOS	1.01	1.00	1.00	n/a	n/a	1.01	1.00
	GROMACS/LAMMPS	AMBER	1.00	1.00	1.00	1.00	n/a	1.00	1.00
	GROMACS/LAMMPS	OPLS/AA	1.00	1.00	1.00	1.00	n/a	1.00	1.00
Butane	GROMACS/LAMMPS ^a	GROMOS	1.01	1.00	1.00	1.00	n/a	1.01	1.00
	GROMACS/LAMMPS	AMBER	1.00	1.00	1.00	1.00	n/a	1.00	1.00
	GROMACS/LAMMPS	OPLS/AA	1.00	1.00	1.00	1.00	n/a	1.00	1.00
Alanine dipeptide	GROMACS/LAMMPS ^a	GROMOS	1.01	1.01	1.00	1.00	1.00	1.01	1.00
	GROMACS/LAMMPS	AMBER	1.00	1.00	1.00	1.00	n/a	1.00	1.00
	GROMACS/LAMMPS	OPLS/AA	1.00	1.00	1.00	1.00	n/a	1.00	1.00
NMA	GROMACS/LAMMPS ^a	GROMOS	1.01	1.00	1.00	1.00	1.00	1.01	1.00
	GROMACS/LAMMPS	AMBER	1.00	1.00	1.00	1.00	n/a	1.00	1.00
	GROMACS/LAMMPS	OPLS/AA	1.00	1.00	1.00	1.00	n/a	1.00	1.00
1L2Y	GROMACS/LAMMPS ^a	GROMOS	1.02	1.01	1.00	1.00	1.00	1.01	0.97
	GROMACS/LAMMPS	AMBER	1.00	1.00	1.00	1.00	n/a	1.00	1.00
	GROMACS/LAMMPS	OPLS/AA	1.00	1.00	1.00	1.00	n/a	1.00	1.00
4LZT	GROMACS/LAMMPS ^a	GROMOS	1.01	1.01	1.00	1.00	1.00	1.01	0.98
	GROMACS/LAMMPS	AMBER	1.00	1.00	1.00	1.00	n/a	1.00	1.00
	GROMACS/LAMMPS	OPLS/AA	1.00	1.00	1.00	1.00	n/a	1.00	1.00

^a As discussed in the text, a 1–3% difference in the non-bonded energy terms (Coulomb and van der Waals) is due to the different cutoff and switch schemes implemented and used in GROMACS vs. LAMMPS when considering the GROMOS force field. See also Supporting Information, Tables S1 and S2, reporting the potential energies and potential energy ratios for all studied systems (in solution and vacuum, respectively).

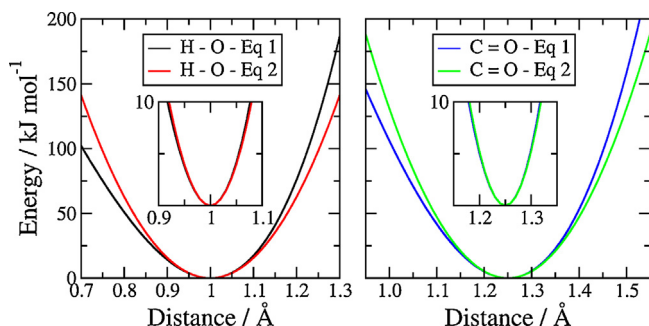


Fig. 3. MDWiZ performs the automated transformation of different covalent bond-stretching interaction forms. As example, the expressions of Eqs. (1) and (2) are compared in the case of the hydroxyl O–H bond (left panel) and of the carbonyl C=O double bond (right panel). Inset panels: energies and distance of interest during a molecular dynamics simulation, i.e. few kT values and typically less than 0.1 Å stretching distances at standard conditions. The energy terms of Eqs. (1) and (2) are virtually identical in the ranges of interest.

including conversion constants. The parameters are converted allowing the use of different simulation setups, even if different underlying potential forms are considered. We exemplify different scenarios in which MDWiZ will be a useful preparation and translation tool.

As a first example, the terms that represent the covalent bond-stretching interaction can be expressed as [11,12]

$$V^{\text{bond}} = \sum_{n=1}^{N_b} \frac{1}{4} k_{b_n} (b_n^2 - b_{0_n}^2)^2 \quad (1)$$

within the GROMOS force field typically used with GROMOS and GROMACS, where N_b is the number of covalent bonds in the topology, k_{b_n} are the force constants and b_{0_n} are the ideal bond lengths for the specific bond type. However, the force constants k_{b_n} have to be transformed for the GROMOS force field in LAMMPS through corresponding parameters consistent with the underlying harmonic potential [1,12,32,39], i.e.

$$k_{b_n}^{\text{harm}} = k_{b_n} 2b_{0_n}^2. \quad (2)$$

Fig. 3 shows the correspondence between the energy term of Eq. (1) and the corresponding harmonic energy term obtained applying Eq. (2) [12].

As a second example, alternative software engines and force fields may express the terms that describe the covalent dihedral potentials as [11,12].

$$V^{\text{dihed}} = k_\phi (1 + \cos(n\phi - \gamma)), \quad (3)$$

(e.g. GROMOS) where γ is the phase shift due to dihedral periodicity, or using Fourier functions such as the Ryckaert–Bellemans (RB) function [40]

$$V_{\text{RB}}^{\text{dihed}} = \sum_{n=0}^5 C_n (\cos \psi)^n = \frac{1}{2} [F_1 (1 + \cos \phi) + F_2 (1 - \cos 2\phi) + F_3 (1 + \cos 3\phi) + F_4 (1 - \cos 4\phi)] \quad (4)$$

with $\psi = \phi - 180^\circ$, where the dihedrals are defined using the IUPAC/IUB convention, and ψ is defined according to the polymer convention (i.e. $\psi_{\text{trans}} = 0$). RB functions are used e.g. within OPLS in GROMACS. Fig. 4 shows the correspondence between the RB function (second term of Eq. (4)) and the corresponding rightmost term of Eq. (4).

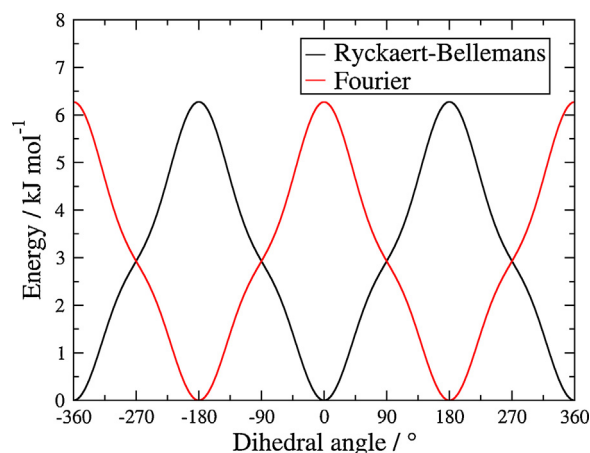


Fig. 4. MDWiZ performs the automated transformation of different dihedral forms. As example, the Ryckaert–Bellemans (second term of Eq. (4); polymer convention) and Fourier dihedral (rightmost term of Eq. (4); IUPAC/IUB convention) forms are compared in the case of the C–C–C–C dihedral for the butane molecule. MDWiZ used the system of equations in Eq. (5) to perform this conversion and takes into account also different dihedral definitions.

When needed for proper translation, MDWiZ converts one dihedral term into another by translating the RB F_n parameters in Eq. (4) as:

$$\begin{cases} F_1 = -2C_1 - \frac{3}{2}C_3 \\ F_2 = -C_4 - C_2 \\ F_3 = -\frac{1}{2}C_3 \\ F_4 = -\frac{1}{4}C_4 \end{cases}, \quad (5)$$

known the corresponding C_n parameters of Eq. (4) [12].

However, other force fields (e.g. OPLS) use a different number of coefficients C_n depending on the specific dihedral considered, and involve additional conversions from polymer-to-protein dihedral convention (i.e. $\phi_{\text{trans}} = 0$). In this case, Eq. (4) can be recovered using Eq. (5) as

$$V^{\text{dihed}} = \frac{1}{2} [F_1 (1 + \cos \phi) + F_2 (1 - \cos 2\phi) + F_3 (1 + \cos 3\phi) + F_4 (1 - \cos 4\phi)], \quad (6)$$

where we note that the sign preceding the last term changed from – to + as a consequence of the different conventions employed to define the dihedral.

In addition, different software engines implement these terms differently to achieve best performance. For example, GROMACS implements OPLS dihedral terms using the cosine relationship of Eq. (4) as it is more efficient than a corresponding Fourier counterpart implementation. Other possible complications arise from different mixing and exclusion rules for the treatment of the non-bonded interactions. In all cases, MDWiZ automatically performs all transformations required to convert input formats and parameters. For example, the force field terms needed to define and represent a simple dihedral energy term result into a total of four different implementations based on the simulation frameworks considered. It is obvious that preparation and translation of entire simulation input and set up cannot be performed on a manual basis, as this results in numerous complex transformations even for small and simple molecules.

The two illustrative examples presented show that – despite the simplicity of the information contained in a force field for a given engine software (the covalent bond terms or the dihedral angle

terms) – multiple and complex conversions are required when translating these terms between alternative simulation setup, making this conversion generally highly error-prone or impossible to be achieved within the current software implementations. Independently of the conversion required, MDWiZ prepares the input files in one user-step to ensure a proper translation of alternative force fields and simulation setup.

A few additional points are worth it mentioning: (1) only some force fields distinguish the dihedral terms between proper (actual dihedral torsions) and improper (dihedrals that are introduced to fix specific out-of-plane stereochemistry) entries (e.g. GROMOS); (2) different force fields have a different treatment of the exclusions in the list of bonded interactions. For example, 1–4 bonded interactions can be excluded in the list of non-bonded interactions to be calculated; (3) different force fields express their parameters with different measure units or multiplying pre-factor constants in the parameter sets. MDWiZ takes care of all these translation issues depending on the software engine and/or force field conversion required by the user. We point out that while the description herein was kept simple using specific examples, support documentation and examples are available on the MDWiZ web interface. At any rate, any deficiency in MDWiZ translation and conversion discussed would result in major energetic differences in the simulated systems presented as validation benchmarks.

Validation was carried out for each system under both vacuum and explicit-solvent conditions. The setup for the explicit solvent simulations is summarized in Table 1. In this release, we have chosen to focus on the conversion between GROMACS to LAMMPS simulation engines using different force fields (AMBER, GROMOS, and OPLS/aa). This choice is motivated by the fact that it includes challenging transformations in terms of the underlying energy forms, as discussed in the previous paragraphs. For validation, a variety of energetic properties of the systems were monitored. They include the total system potential energy, the total system Coulomb energy, the van der Waals energy, and the decomposed force field energy terms for the bond, angles, and dihedral energy terms. The improper dihedral energy terms are also considered for the GROMOS force field; the improper dihedral terms are instead considered within the standard dihedrals in the AMBER and OPLS/aa force fields, as the latter force fields do not make a formal distinction between these degrees of freedom. The full validation data is reported in Supporting Information, Table S1, which includes the actual output with full significant digits (up to 10^{-9} kJ mol $^{-1}$). It is important to note that the validation presented was performed on the basis of single-point calculations that run the LAMMPS *rerun* routine on GROMACS-generated MD snapshots. This procedure ensures full independence on the trajectory used for validating the individual energy terms in the force field, as calculated by independent software engines.

For each of these energy terms, the ratio of the average values between the simulation setup is considered as indicative of proper conversion within the scope of MDWiZ. A ratio value of 1 shows that the translation performed by MDWiZ can produce identical simulations to be performed with the output setup. Fig. 5 shows an example of the GROMACS/LAMMPS energy ratio time series for all force field potential energy terms in the case of the Triclinic Hew Lysozyme for simulations with GROMOS, AMBER, and OPLS force fields prepared by MDWiZ. From the validation set of Fig. 2, the Lysozyme system is the largest and expectedly found to be the one where deviations are more pronounced. Similar analyses were performed for all systems and Table 2 summarizes the validation of MDWiZ in terms of the ratio values. In most cases a perfect match (unit ratio) is found. However, a 1–3% difference in the non-bonded energy terms is found when comparing GROMACS vs. LAMMPS setup in the case of the GROMOS force field. A careful analysis of these results indicates that this difference does not arise from the

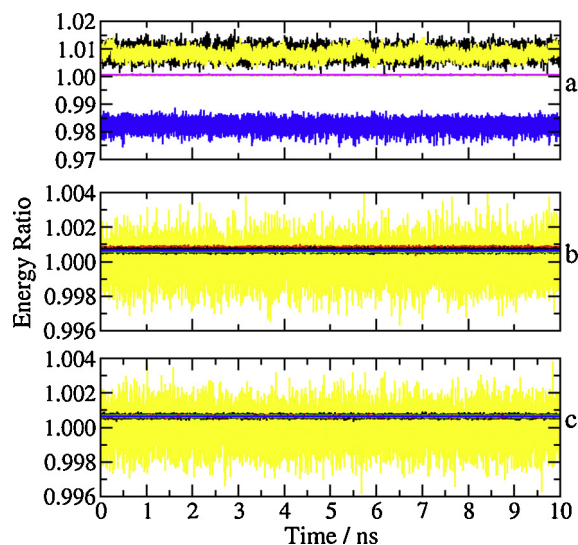


Fig. 5. Ratio of GROMACS/LAMMPS potential energy terms for the Triclinic Hew Lysozyme (PDB ID: 4LZT) computed using different force fields: (a) GROMOS; (b) AMBER; and (c) OPLS. Each color represents a different force field potential energy term. Black: bond; red: angle; green: dihedral; magenta: improper; yellow: Coulomb; and blue: van der Waals. Note that energy ratios are obtained from single-point energy calculations by post-processing identical trajectory snapshots. See also Supporting Information, Table S1, reporting the potential energies and potential energy ratios for all studied systems with full significant digits (up to 10^{-9} kJ mol $^{-1}$). For interpretation of the references to color in this text, the reader is referred to the web version of the article.

MDWiZ conversion algorithms, but it is instead due to the different cutoff and switch schemes implemented in GROMACS (described in Appendix of Ref. [41]) and LAMMPS (the “CHARMM” functional form implemented in LAMMPS was used as described in Ref. [42]). More details are provided as Supporting Information (Table S1) including the mean values of most relevant energy terms as calculated throughout the explicit solvent molecular dynamics runs. The same calculations were repeated for the systems in vacuum (see Supporting Information, Table S2) confirming the robustness of MDWiZ.

We note that MDWiZ is meant to help users, but it is out of its scope to improve or complement capabilities of the software engine involved in the desired translation. Therefore, it is important to keep in mind that if an algorithm is intrinsically missing (e.g. different cutoff and switches schemes implemented) this will lead to somewhat different treatment of the simulation conditions. When identical algorithms are not implemented in the “input” and “output” engines involved in the desired conversion minor differences are expected, such as those observed and described above for the non-bonded energy terms (1–3% difference for GROMACS vs. LAMMPS for the GROMOS force field). Nevertheless, MDWiZ will provide the “best possible” files to perform the simulation in the software of choice. It is also important noting that force fields are compared routinely including some arbitrary contribution due to the specifics of the software engine that is chosen for that comparison. In this context, MDWiZ can be easily employed to yield more rigorous comparisons as part of force field comparison studies.

Overall, we have demonstrated that the MDWiZ tool (version 1.0) enables the automated conversion of GROMACS files for simulations with GROMOS, AMBER, and OPLS/AA force fields for simulations into LAMMPS with accuracy as high as the internal accuracy provided by the engines considered. The conversion capabilities are extended to various water models. In the future, new features will be implemented in MDWiZ to extend the translation to other software engines, as summarized in Fig. 1.

Documentation and details of the progress and release of new versions of the MDWiZ tool can be accessed at: <http://barongroup.medchem.utah.edu/tools>.

4. Conclusions

We have presented the first release of the MDWiZ tool, an online platform for the fast and reliable conversion of MD simulations input files. MDWiZ allows the user to translate different MD engine setup options, including force fields and models with different underlying potential forms. MDWiZ is freely accessible to the community as a web interface. The platform allows for: (i) force field comparison across software engines; (ii) to take advantage of specific features of different MD software packages without dealing with new topology formats and the learning curve associated with new pre-processing tools; and, (iii) easy matching of code efficiency to computer architecture for a system that was initially generated with a MDWiZ supported input format. Overall, MDWiZ will be useful to the broad computational chemistry community for avoiding waste of time and resources in file and format conversion, for enhancing the reproducibility of simulation data, and for helping the accurate comparison of alternative simulation protocols. However, we believe that a broader solution and a general agreement on formats specification similarly to what done for the Protein Data Bank formats would be extremely beneficial on the longer period for the molecular simulation field. We hope this study will contribute to stimulate a general agreement in this direction.

Acknowledgments

This work was supported by startup funds from the Department of Medicinal Chemistry, The University of Utah (R.B.). We acknowledge the Center of High Performance Computing for technical support and allocation of computer time and a generous allocation at the Extreme Science and Engineering Discovery Environment (XSEDE) supercomputers (award TG-CHE120086 to R.B.). XSEDE is supported by National Science Foundation grant number OCI-1053575. Support from the Brazilian government to V.H.R. through a Nanobiotec-BR/CAPES scholarship Proc. 9614-12-5 and to V.A.C.H through a CAPES/Ciências sem Fronteiras scholarship is warmly acknowledged. R.B. warmly thanks Eveline Bruenger for a generous donation to the Baron group development fund.

References

- [1] D.A. Case, T. Darden, T. Cheatham III, C. Simmerling, J. Wang, R. Duke, et al., AMBER, 2006.
- [2] B.R. Brooks, C.L. Brooks III, A.D. Mackerell Jr., L. Nilsson, R.J. Petrella, B. Roux, et al., CHARMM: the biomolecular simulation program, *J. Comput. Chem.* 30 (2009) 1545–1614.
- [3] K.J. Bowers, E. Chow, H. Xu, R.O. Dror, M.P. Eastwood, B.A. Gregersen, et al., Scalable algorithms for molecular dynamics simulations on commodity clusters, in: *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*, ACM, Tampa, FL, 2006, p. 84.
- [4] W. Smith, T.R. Forester, DL-POLY 2.0: a general-purpose parallel molecular dynamics simulation package, *J. Mol. Graph.* 14 (1996) 136–141.
- [5] I.T. Todorov, W. Smith, DL-POLY 3: the CCP5 national UK code for molecular-dynamics simulations, *Philos. Trans. A: Math. Phys. Eng. Sci.* 362 (2004) 1835–1852.
- [6] I.T. Todorov, W. Smith, K. Trachenko, M.T. Dove, DL-POLY 3: new dimensions in molecular dynamics simulations via massive parallelism, *J. Mater. Chem.* 16 (2006) 1911–1918.
- [7] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation, *J. Chem. Theory Comput.* 4 (2008) 435–447.
- [8] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J.C. Berendsen, Gromacs: fast, flexible, and free, *J. Comput. Chem.* 26 (2005) 1701–1718.
- [9] S. Pronk, S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, et al., GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit, *Bioinformatics (Oxford, England)* 29 (2013) 845–854.
- [10] M. Christen, P.H. Hünenberger, D. Bakowies, R. Baron, R. Burgi, D.P. Geerke, et al., The GROMOS software for biomolecular simulation: GROMOS05, *J. Comput. Chem.* 26 (2005) 1719–1751.
- [11] W.R.P. Scott, P.H. Hünenberger, I.G. Tironi, A.E. Mark, S.R. Billeter, J. Fennen, et al., The GROMOS biomolecular simulation program package, *J. Phys. Chem. A* 103 (1999) 3596–3607.
- [12] W.F. van Gunsteren, S.R. Billeter, A.A. Eising, P.H. Hünenberger, P. Kruger, A.E. Mark, et al., Biomolecular Simulation: The GROMOS96 Manual and User Guide, vdf Hochschulverlag AG an der ETH Zürich and BIOMOS b.v. Groningen, Zürich, Groningen, 1996.
- [13] S. Plimpton, Fast parallel algorithms for short-range molecular dynamics, *J. Comput. Phys.* 117 (1995) 1–19.
- [14] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, et al., Scalable molecular dynamics with NAMD, *J. Comput. Chem.* 26 (2005) 1781–1802.
- [15] M. Valiev, E.J. Bylaska, N. Govind, K. Kowalski, T.P. Straatsma, H.J.J. van Dam, et al., NWChem: a comprehensive and scalable open-source solution for large scale molecular simulations, *Comput. Phys. Commun.* 181 (2010) 1477–1489.
- [16] P. Eastman, M.S. Friedrichs, J.D. Chodera, R.J. Radmer, C.M. Bruns, J.P. Ku, et al., OpenMM 4: a reusable, extensible, hardware independent library for high performance molecular simulation, *J. Chem. Theory Comput.* 9 (2012) 461–469.
- [17] J.W. Ponder, TINKER: Software Tools for Molecular Design, 2004, Washington University Medical School, 2009, Available at <http://dasher.wustl.edu/tinker/>
- [18] M. Levitt, The birth of computational structural biology, *Nat. Struct. Biol.* 8 (2001) 392–393.
- [19] P. Bjelkmar, P. Larsson, M.A. Cuendet, B. Hess, E. Lindahl, Implementation of the CHARMM force field in GROMACS: analysis of protein stability effects from correction maps, virtual interaction sites, and water models, *J. Chem. Theory Comput.* 6 (2010) 459–466.
- [20] A.K. Malde, L. Zuo, M. Breeze, M. Stroet, D. Poger, P.C. Nair, et al., An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0, *J. Chem. Theory Comput.* 7 (2011) 4026–4037.
- [21] A. Ribeiro, B.A. Horta, R.B. de Alencastro, MKTOP: a program for automatic construction of molecular topologies, *J. Braz. Chem. Soc.* 19 (2008) 1433–1435.
- [22] M.F. Crowley, M.J. Williamson, R.C.C.H.A.M.B.E.R. Walker, Comprehensive support for CHARMM force fields within the AMBER software, *Int. J. Quantum Chem.* 109 (2009) 3767–3772.
- [23] N.M. O'Boyle, M. Banck, C.A. James, C. Morley, T. Vandermeersch, G.R. Hutchison, Open Babel: an open chemical toolbox, *J. Cheminform.* 3 (2011) 33.
- [24] D. Roy, K. Todd, J. Millam, GaussView Version 5, Semichem Inc., Shawnee Mission, KS, 2009.
- [25] J.W. Neidigh, R.M. Fesinmeyer, N.H. Andersen, Designing a 20-residue protein, *Nat. Struct. Biol.* 9 (2002) 425–430.
- [26] M.A. Walsh, T.R. Schneider, L.C. Sieker, Z. Dauter, V.S. Lamzin, K.S. Wilson, Refinement of trypsin hen egg-white lysozyme at atomic resolution, *Acta Crystallogr. D: Biol. Crystallogr.* 54 (1998) 522–546.
- [27] C. Oostenbrink, A. Villa, A.E. Mark, W.F. van Gunsteren, A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6, *J. Comput. Chem.* 25 (2004) 1656–1676.
- [28] H.J.C. Berendsen, Interaction Models for Water in Relation to Protein Hydration, Dordrecht, Pullman, BE, 1981.
- [29] J. Åqvist, Ion water interaction potentials derived from free-energy perturbation simulations, *J. Phys. Chem.* 94 (1990) 8021–8024.
- [30] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, Comparison of multiple amber force fields and development of improved protein backbone parameters, *Proteins* 65 (2006) 712–725.
- [31] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.* 79 (1983) 926–935.
- [32] G.A. Kaminski, R.A. Friesner, J. Tirado-Rives, W.L. Jorgensen, Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides, *J. Phys. Chem. B* 105 (2001) 6474–6487.
- [33] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. Di Nola, J.R. Haak, Molecular-dynamics with coupling to an external bath, *J. Chem. Phys.* 81 (1984) 3684–3690.
- [34] R.W. Hockney, The potential calculation and some applications, *Methods Comput. Phys.* 9 (1970) 136–211.
- [35] S. Miyamoto, P.A. Kollman, Settle – an analytical version of the shake and rattle algorithm for rigid water models, *J. Comput. Chem.* 13 (1992) 952–962.
- [36] J.-P. Ryckaert, G. Cicotti, H.J.C. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, *J. Comput. Phys.* 23 (1977) 327–341.
- [37] W.F. van Gunsteren, H.J.C. Berendsen, Computer-simulation of molecular-dynamics – methodology, applications, and perspectives in chemistry, *Angew. Chem. Int. Ed.* 29 (1990) 992–1023.
- [38] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, *J. Mol. Graph.* 14 (33–38) (1996) 27–28.
- [39] W.L. Jorgensen, D.S. Maxwell, J. Tirado-Rives, Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids, *J. Am. Chem. Soc.* 118 (1996) 11225–11236.
- [40] J.P. Ryckaert, A. Bellemans, Molecular dynamics of liquid n-butane near its boiling point, *Chem. Phys. Lett.* 30 (1975) 123–125.
- [41] R. Baron, D. Trzesniak, A.H. de Vries, A. Elsener, S.J. Marrink, W.F. van Gunsteren, Comparison of thermodynamic properties of coarse-grained and atomic-level simulation models, *ChemPhysChem* 8 (2007) 452–461.
- [42] A.D. MacKerell, D. Bashford, R.L. Bellott Dunbrack, J.D. Evanseck, M.J. Field, et al., All-atom empirical potential for molecular modeling and dynamics studies of proteins, *J. Phys. Chem. B* 102 (1998) 3586–3616.