Bartender: Martini 3 Bonded Terms via Quantum Mechanics-Based Molecular Dynamics

Gilberto P Pereira, Riccardo Alessandri, Moisés Domínguez, Rocío Araya-Osorio, Linus Grünewald, Luís Borges-Araújo, Sangwook Wu, Siewert J Marrink, Paulo Cesar Telles de Souza, Raul Mera-Adasme

Bartender: Martini 3 Bonded Terms via Quantum Mechanics-based Molecular Dynamics

Gilberto P. Pereira[a,b], Riccardo Alessandri[c], Moisés Domínguez[d], Rocío Araya-Osorio[e], Linus Grünewald[f], Luís Borges-Araújo[a,b], Sangwook Wu [g,h], Siewert J. Marrink[f], Paulo C. T. Souza*[a,b], Raul Mera-Adasme*[e]

[a] Laboratoire de Biologie et Modélisation de la Cellule, CNRS, UMR 5239, Inserm, U1293, Université Claude Bernard Lyon 1, Ecole Normale Supérieure de Lyon, 46 Allée d'Italie, 69364, Lyon, France.

[b] Centre Blaise Pascal de Simulation et de Modélisation Numérique, Ecole Normale Supérieure de Lyon, 46 Allée d'Italie, 69364, Lyon, France.

[c] Pritzker School of Molecular Engineering, University of Chicago, Chicago, Illinois 60637, United States

[d] Departamento de Ciencias del Ambiente, Facultad de Química y Biología, Universidad de Santiago de Chile (USACH), Av. Libertador Bernardo O'Higgins 3363, 9170022 Estacion Central, Chile.

[e] Departamento de Quimica, Facultad de Ciencias, Universidad de Tarapacá, Av. Gral. Velasquez 1775, Arica, Chile.

[f] Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands

[g] PharmCADD, Busan, 48792, Republic of Korea.

[h] Department of Physics, Pukyong National University, Busan, 48513, Republic of Korea.

* Paulo Cesar Telles de Souza - email: paulocts@gmail.com, and Raul Mera-Adasme - email: rmeraa@academicos.uta.cl.

KEYWORDS . Martini force field, coarse-grained models, molecular dynamics, automatic parameterization

Coarse-grained (CG) molecular dynamics (MD) simulations have grown in applicability over the years. The recently released version of the Martini CG force field (Martini 3) has been successfully applied to simulate many processes, including protein-ligand binding. However, the current ligand parameterization scheme is manual and requires an a priori reference all-atom (AA) simulation for benchmarking. For systems with suboptimal AA parameters, which are often unknown, this translates into a CG model which does not reproduce the true dynamical behavior of the underlying molecule. Here we present Bartender, a quantum mechanics (QM)/MD-based parameterization tool written in Go. Bartender harnesses the power of QM simulations and produces reasonable bonded terms for Martini 3 CG models of small molecules in an efficient and user-friendly manner. For small, ring-like molecules, Bartender generates models whose properties are indistinguishable from the human-made models. For more complex, drug-like ligands, it is able to fit functional forms beyond simple harmonic dihedrals, and thus better captures their dynamical behavior. Bartender has the power to both increase the efficiency and the accuracy of Martini 3-based high-throughput applications by producing stable and physically realistic CG models.

1. Introduction

Molecular dynamics (MD) simulations are useful tools for probing the dynamical behavior of biomolecular systems. For example, they can be integrated into larger workflows, such as drug discovery pipelines[1,2]. Within drug discovery campaigns, these methods can be used to identify druggable pockets in proteins[3,4], to calculate the binding affinity of several ligands to a given protein[5,6] or to guide lead-optimization[7,8] efforts. Among the available simulation methods, coarse-grained (CG) approaches represent a trade-off between molecular detail and computational efficiency. One of the most popular CG methods is the Martini force field[9–11], the most recent version being Martini 3.[10] Simulations of molecular systems with Martini 3 have found wide application, such as in the study of protein-lipid interactions[12–15], green solvents[16,17], polymer systems[18–20] and protein-ligand binding[21–24] . The Martini force field philosophy rests on a building-block approach, where groups of atoms (representing chemical moieties) are grouped into simulations beads of various sizes and chemical types. The bonded parameters for these building blocks (bonds, angles, dihedrals) are parameterized with respect to reference

all-atom (AA) simulations. The bead types are assigned based on the hydrophobicity and underlying chemical groups they represent and calibrated using experimental data such as oil-to-water transfer free energies[10,25].

Currently, small molecule parameterization in Martini 3 is a manual, time-intensive, and iterative process. While inspiration may be drawn from the molecules deposited on the MAD database[26] and the Martini 3 small molecule models[25], these repositories do not cover the whole of the chemical space. Thus, a tool that could help automate the process, starting from AA to CG mapping followed by bonded parameter determination and bead assignment, would be highly desirable. Steps have been taken towards this direction in the past: in 2015, Bereau and Kremer unveiled AutoMartini[27], an automatic parameterization workflow for Martini 2, with the aim of streamlining Martini small-molecule parameterization schemes by integrating the mapping and parameterization steps into a easy-to-use pipeline. Later, Graham et al. developed PyCGTool[28], a Python pipeline to generate bonded parameters for small molecules. This pipeline was validated against two small-molecule drugs (atenolol and capsaicin), one lipid (dipalmitoylphosphatidylcholine, DPPC) and a polyalanine strand. While PyCGTool[28] achieved good accuracy in generating the bonded parameters, it has some limitations in terms of the types of potentials included within the code. For example, the authors state that in keeping with the Martini philosophy, PyCGTool uses a cos-harmonic angle potential and dihedrals are not usually defined. Recently, other approaches for bonded-parameter generation and optimization were explored. Examples are cg_params[29], a tool for generating mapping and bonded parameters covering a wide range of chemical space, and CGcompiler[30] or Swarm-CG[31], tools dedicated to the optimization of Martini models.

These tools utilize reference AA MD, generated by the user, to obtain bonded-parameter distributions, from which the Martini bonded parameters are obtained. This approach requires significant manual work from the user, as, at minimum, the AA simulation must be prepared, run and compared to CG parameters for further optimization. Moreover, manual parametrization of the AA system might be required, as some certain heavier atoms (third period or higher) and some torsions (conjugated systems) may not be accurate or even supported by some standard biomolecular force-fields and their automated topology builders. Furthermore, the procedure assumes that the force field used to generate the AA MD data describes appropriately the conformational flexibility of the small molecule to be parameterized. It is known that in some cases, AA parameters may be suboptimal, including but not restricted to imprecisions in the samplable free energy surface of dihedrals[32]. An alternative approach is to generate the parameters from quantum mechanical (QM) simulations. While post-Hartree-Fock and even Density Functional Theory (DFT) methods are too computationally expensive, fast and reasonably accurate semiempirical methods have been available for some years, which can make the problem treatable[33].

In this work we present Bartender, a tool for assisting in the determination of Martini 3 bonded parameters for small molecules. Bartender was tested on two separate datasets, the first one containing 85 ring-like small molecules from [25] and a second one composed of 10 conjugated, drug-like ligands. The first dataset was used as a benchmark, to validate Bartenders' workflow on molecular fragments. The second dataset was constructed such that it contained flexible molecules which are difficult to parameterize. It was used to compare Bartenders' parameterization capabilities against PyCGTool[28], which was shown to produce parameters of reasonable accuracy for different molecular classes. The impact of Bartender in small-molecule parameterization is showcased using thyroxine, comparing dihedral distributions arising from QM-MD and standard atomistic simulations.

By relying on modern semiempirical QM methods[34], Bartender bypasses the need for a classical atomistic MD simulation, replacing it by easy-to-use QM-MD calculations with xtb[33]. Additionally, Bartender implements a variety of functional forms for both angles and dihedrals, enabling a better fit to the reference distributions. Thus, most of the procedure can be automated, and the quality of both the reference MD trajectories and the resulting Martini 3 parameter set is improved. Importantly, Bartender allows non-experts to obtain Martini parameters for their systems of interest and thus constitutes an important step towards the routine employment of Martini's computational efficiency for the high-throughput study of large sets of molecules.

2. Computational methods

2.1 Implementation

Bartender is a Go program, employing the goChem[35] and Gonum[36] libraries for chemical and mathematical facilities, respectively. Although fully open-source, it is most easily installed from the binary distribution provided. Bartender has no runtime dependency other than the xtb program[33] (which is free and open source). The open-source xdrfile library, distributed by the GROMACS[37] developers is required only if XTC trajectory support is wanted.

Bartender was programmed in a modular fashion to simplify maintenance and the addition of new features. It is a command-line program. Its user interface comprises

several optional arguments that modify the program's behavior. For all of them, sensible defaults are provided, so, in most cases, they can be omitted.

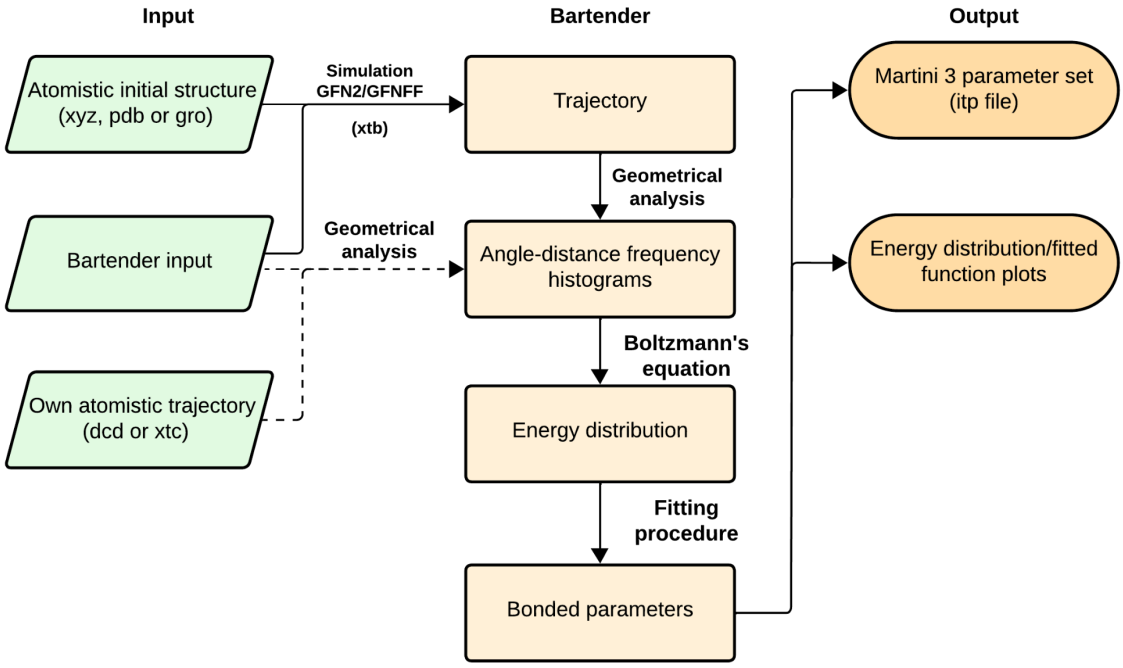

Figure 1: Flow chart of the Bartender procedure, including input files required, and output files produced by the program. Boxes represent data, while arrows represent procedures.

In addition to the optional arguments, Bartender takes two required input files. One with the atomistic structure of the system to be parametrized (in XYZ, PDB or GRO formats) and a second one, containing the mapping from the atoms to Martini beads, as well as the bonded terms to be parametrized. Note that, while Bartender requires an input file, this file is fairly simple (see the Supporting Information for an example) and contains only information that would be required regardless of the method employed for obtaining the "target" trajectory and for the fitting procedure. A script to automatically produce the file from a Gromacs topology is also provided.

From the given input, Bartender can perform an MD simulation of the system with one of the GFN-xTB semiempirical quantum-chemical methods[34,38] or with the related GFNFF force-field[39], and using a continuum model for solvent effects[40] (**Figure 1**). All these methods are of high computational efficiency, compared to most QM-based methods, and have been shown to yield reliable geometries[33]. The xtb program is employed for all the simulations. The length of the simulation can be controlled by the user. A user flag is also available for the user to provide their own trajectory file, in DCD or XTC format, bypassing the need for Bartender to perform a simulation at all.

From the MD trajectory, and for every bonded parameter in the input file, Bartender obtains a probability distribution by (1) mapping the atoms involved into Martini beads according to a center-of-geometry (COG) approach and (2) extracting the underlying dynamics of that parameter and (3) constructing a corresponding frequency histogram, where the time-dependent information is lost (**Figure 1**). The distribution is then employed to calculate energies for each numerical value of each term, by means of the Boltzmann distribution.

The values for the parameters are then obtained by fitting the energy plots for each bonded term to different potential functions available in the GROMACS package[37]. Thus, a harmonic potential is fitted for bond lengths, angles and improper dihedrals. However, bonds with a force constant greater than $2*10^4$ kJ.mol^{-1}.nm^{-2}, are modeled with constraints. For proper dihedral terms, both a simple-periodic and a Ryckaert-Bellemans (RB) function are fitted. A restricted-bending potential[41] is provided for the latter terms, so as to prevent spurious coplanarity of the atoms involved, a problem that can arise during CG simulations. A combined torsion-bending potential is also provided[42]. The native-go Gonum library[36] is employed for the fitting procedures.

The Martini 3 bonded terms often model complex underlying motions in the atomistic molecule. Thus, the potential energy surface for the terms, even bond stretchings, can be highly anharmonic or multimodal in the mapped atomistic simulations. Bartender contains several heuristics to deal with complex potential energy surfaces, which reproduce the methods employed by experts when parametrizing molecules manually. The heuristics include: (i) the elimination of high-energy points, which reflect points with very low sampling (i.e. only 1 frame in the whole trajectory); (ii) the average fitting for multimodal distributions, which can be used to fit potentials in bond and angle energy plot with two or more minima (see example in **Figure S1**); (iii) the removal of anharmonicities by a slope criterion. The behavior of the heuristics can be tuned by command line flags. It is worth noting that to ensure Martini 3 compatibility, and simulation efficiency, the potentials provided by Bartender are either harmonic or, in the case of dihedrals, cosine expansions such as the Ryckaert-Belleman potentials. However, these solutions are limited in fitting or multimodal and/or anharmonic distributions. Software packages that do not have these constraints have addressed these problematic distributions with different approaches[43].

In addition to writing the obtained parameters to a GROMACS-formatted topology file, which contains the RMSD for each fit as a comment, Bartender writes, for each term, a PNG image file with a plot of both the MD-derived values and the potential function fitted to them. Grace-formatted[44] text files with the same values are also written, so users can make their own plots. During its run, Bartender will print messages and warnings to the standard output, designed to guide corrections to the initial parameterization and alert of possible problems. Bartender's verbosity level can also be adjusted. The code of Bartender can be found at https://github.com/Martini-Force-Field-Initiative/Bartender

2.2 General Simulation Parameters

2.2.1 - Generic Martini 3

Bartenders' capabilities were evaluated on two datasets, one with 85 ring-like molecules and another containing 10 flexible drug-like ligands, totaling 95 molecules. The datasets are detailed in **Tables S1** and **S2** of the Supporting Information.

An AA structure of each molecule was first uploaded to the CGBuilder web interface[https://jbarnoud.github.io/cgbuilder/]. Within CGBuilder, the molecule was mapped to the CG resolution and the corresponding mapping and structure files were retrieved. The settings for the CG simulations adhere to the "new" set of Martini run parameters[45], which is default for Martini 3. Specifically, a time step of 20 fs is used, and the Verlet neighbor search algorithm was applied to update the neighbor list, employing a straight cutoff of 1.1 nm. Pressure and temperature were maintained using the Parrinello–Rahman barostat[46] (coupling parameter of 12.0 ps) and the velocity-rescaling thermostat[47] (coupling parameter of 1.0 ps), respectively. [45]

2.2.2 - AA Simulations

For each molecule, a SMILES string generated using openbabel[48] was supplied to the LigParGen server[49]. From the server, a OPLS-AA/1.14*CM1A-LBCC model was obtained which also included a GROMACS itp file and a coordinate file. The molecule was then inserted into a 4x4x4 nm TIP3P water box. The solvated system was then simulated for 2 microseconds at 298.15K, controlled using the Nose-Hoover thermostat[50], and 1 bar pressure, controlled using the Parrinello-Rahman barostat[46], using a 2fs integration time-step. The thermostats' coupling parameter was set at 1 ps and the barostats' coupling parameter was set at 5 ps. The Verlet algorithm was used to build the neighbor list. All simulations were run with GROMACS[37] version 2019.x or later.

2.2.3 - QM Simulations

Simulations of each molecule at the GFN2[34] level of theory were carried out using xtb[34] inside Bartender. The simulations were carried out in an implicit solvent mimicking water with a dielectric constant of 80 at 298.15K. For the Martini 3 small molecule dataset, consisting of rigid rings, simulations were limited to 1 ns. Meanwhile, for the dataset comprising flexible systems, including conjugated and drug-like small molecules, extensive 100 ns trajectories were conducted.

2.2.4 - Martini 3 small molecule dataset

Densities and enthalpies of vaporization have been computed for the models in the small molecule Martini 3 dataset as described in the work of Alessandri and co-workers.[25] Briefly, a liquid phase was modeled as an equilibrated box with dimensions of around 5×5×5 nm³, while a gas phase was represented by a single molecule occupying a simulation box of 7×7×7 nm³. Simulations for the liquid phase were conducted in the NPT ensemble at 298 K and 1 bar, while gas phase simulations were carried out in the NVT ensemble at 298 K. The enthalpy of vaporization (ΔH_{vap}) was calculated as $\Delta H_{vap} \approx U_{gas} - U_{liq} + RT$, in which U_{gas} and U_{liq} represent the total energies per mole for the gas and liquid phases, respectively, obtained from NVT simulations. Densities were acquired using the GROMACS tool gmx density. Note that in this analysis we used the actual molar mass

of the molecule rather than the standard Martini 3 masses (72 Da, 54 Da, and 36 Da for regular, small, and tiny beads, respectively).

2.2.5 - Martini 3 Conjugated systems and/or drug-like small-molecules

The CG models of the molecules were embedded into a 5x5x5 nm³ water box, and simulated for 2 microseconds at 298.15K and 1 bar pressure using a 20 fs integration time-step, with the same setup as for the ring systems described above. Simulations were carried out for three sets of models: a manually optimized model (described as "Human"), a model produced by PyCGTool (PyCGTool) and a model produced by Bartender (Bartender). All the models were parametrized following as an initial step the general guidelines of Martini 3 for mapping, bead assignment, bead positioning (center of geometry rule) and choices of bonded terms and virtual sites to be used[10,23,25]. The "Human" model has its bonded parameters optimized by an iterative trial-and-error process (as typically described on the Martini tutorials[23], aiming at achieving the best possible overlap to the reference simulations. Details about PyCGTool and Bartender optimizations are described in the section below. All simulations were run with GROMACS version 2019.x[37] or later. Stability testing was also performed by inserting 10 copies of each molecule in a 15x15x15 nm³ water box and simulating this box for 2 microseconds using the same setup as above. This dataset was selected specifically because it includes (1) realistic cases where parameterization is not straightforward, (2) atoms whose atomistic parameters may be suboptimal (Thyroxine case**)** and (3) molecules that are flexible and have several dihedrals, both hindered and freely rotatable.

2.3 Bonded-parameter extraction strategies

2.3.1 Bartender: Martini 3 small molecule dataset

Bartender was used to generate bonded parameters for each molecule of the Martini 3 small molecule dataset by providing the following two files: (1) an all-atom structure file (in pdb format) of the molecule and (2) a Bartender input file. The Bartender input file contains two sections: a top section where the mapping is defined and a bottom section where the bonded-parameters to be fit are specified. We provide a utility script, `write_bartender_inp.py`, to generate Bartender input files given a GROMACS topology file (itp format) and a GROMACS index file that defines the mapping from AA to CG. Such files are available for all the models in the Martini 3 small molecule dataset at the repo https://github.com/ricalessandri/Martini3-small-molecules . After the run, the bonded parameters generated were complemented by the bead (and virtual site) definitions taken from the original dataset topologies in order to generate the complete Bartender topologies that can be used for simulations. We also provide a utility script, `produce_bartender_sm3_itps.py`, to facilitate this last step.

2.3.2 - Bartender: Conjugated systems and/or drug-like molecules

The Bartender software suite was used to generate initial parameter sets for each of the ten molecules in the flexible molecules dataset either using a QM simulation generated by xtb or using all-atom MD simulations produced using OPLS in GROMACS. In both cases, a structure file (pdb or xyz format) and a Bartender input file (described in **2.3.1**) were fed to the pipeline. In some cases, where the fitting procedure may fail or struggle to fit the underlying distribution, some flags can be employed as described in **Section 2.1**. For the parameter sets coming from fitting all-atom MD simulations, the "-removeAnharmonic" flag was used to adjust 5 bonds and 1 improper dihedral and the "-histCutoff" flag was used to adjust 2 angles and 1 improper dihedral. Additionally, some of the angles had to be removed from analysis because they were manually changed to yield more stable models. The flags were also used to adjust some parameters arising from the QM simulation fitting procedure. In particular, the "-histCutoff" flag was used to adjust one angle and the "-improperIncrement" flag was used to adjust one improper dihedral. A Table (**Table S3**) highlighting which bonded-parameters were modified or removed from the analysis is given in the Supporting Information.

2.3.3 - PyCGTool: Conjugated systems and/or drug-like molecules

In order to compare Bartender to PyCGTool, four input files were generated per molecule: (1) A file in PyCGTool style describing the mapping of the molecule, (2) a file containing the internal coordinates of the molecule as defined by the user, (3) a coordinate file generated by LigParGen[49] by supplying the corresponding SMILES code[51], (4) and a 2 microsecond long all-atom simulation of the molecule in question (see paragraph **2.2**). These input files were supplied to PyCGTool to generate the parameters for the model. Measured distributions for the given bonds, angles, and dihedral angles were set to be printed out as well.

3. Results and Discussions

Most Martini 3 small-molecule parameterization strategies employ a comparison to AA reference simulations for the optimization of the bonded parameters, i.e., all parameters required for each bonded term (which vary depending on the potential employed). However, there are cases where the

underlying AA parameters for small molecules are inaccurate, either due to lack of appropriate benchmark data during the parameterization process or other problems. Typically, this is observed in flexible molecules, especially those with hindered torsions, although theoretically, it could impact any molecule. One way to circumvent this issue at the level of CG small-molecule model building, and bypass the "man-in-the-middle" scenario, is to use QM simulations as a reference. While this strategy is attractive, the setup of a QM simulation is not trivial and may be quite computationally expensive to carry out. By integrating the xtb software suite developed by the Grimme lab[34] in its simulation-and-fitting pipeline, Bartender enables users to access fast and easy-to-use QM simulations *via* the GFN2 method. Given the inherent differences between classical force fields and QM simulations, we investigated how different would the resulting bonded-parameter distributions be. Two sets of small molecules were investigated: the Martini 3 small molecule dataset, composed of rigid rings and a complete new set of more flexible systems composed of conjugated and drug-like small molecules.

3.1 Benchmark on the Martini 3 small molecule dataset: rigid ring molecules

We start by testing the ability of Bartender to assign and derive bonded parameters for a published dataset of Martini 3 small molecule models.[25] This dataset contains 90 models, and it is rich in models representing rigid, ring-like structures, both aliphatic and aromatic. To ensure a fair comparison with Bartender, we exclude models that were optimized beyond the standard COG-mapping scheme, leading to a subset of 85 models that we use for the comparison. Bartender properly assigned the bonds of the models as constraints or harmonic potentials in 90% of the cases, with the misassignments still resulting in reasonable and computationally stable models. The choice between constraints and harmonic potentials has no defined rule in Martini 3, being usually a choice made by the developer based on a balance between fitting of bond distance distributions, numerical stability of simulations, and model consistency with similar molecules. Instead, Bartender has a defined upper limit value of force constant to select harmonic potentials or constraints. Given the fact that all the models produced were stable at 20 fs, the different assignments here are acceptable choices. The correlation between QM-based Bartender-obtained and the reference, atomistic-based human-obtained bond lengths is shown in **Figure 2A**, highlighting a very high correlation (R^2 = 0.96) and low RMSE (0.02 nm). Similarly, **Figure 2B** shows the correlation for the improper dihedral angles (note that only 41 out of the 85 models contain at least one improper dihedral), depicting a practically perfect correlation (R^2 = 1.00) and very low RMSE (3.08 deg). In summary, Bartender can be used to derive bonded parameters for the rigid, ring-like structures of the Martini 3 small molecule dataset.

The comparison of bonded parameters serves as a useful proxy for predicting the final performance of the model when all other factors are held constant. However, the ultimate objective of these models is to exhibit properties that align with experimental observations. Hence, for all the Bartender models we computed the mass density and compared the obtained values to experiments (**Figure 2C**). This allows for a direct comparison with the COG-mapped models of the Martini 3 small molecule dataset (**Figure 2D**; data from Ref[25]). Comparing the performance of the two sets, we see how the achieved mean absolute percentage errors (MAPEs) are very similar, with Bartender leading to an overall MAPE only marginally worse than the human-fitted set (7.5% vs 7.2%). A similar comparison is shown in the Supporting Information for heat of vaporization (**Figure S2**), showing how the Bartender-fitted models are virtually indistinguishable from the COG-mapped models of Ref. [25]. In summary, Bartender generates Martini 3 models for ring-like structures whose properties are indistinguishable from the human-made models and with a very small computational cost (see **Section 3.4**), demonstrating high promise for high-throughput applications.

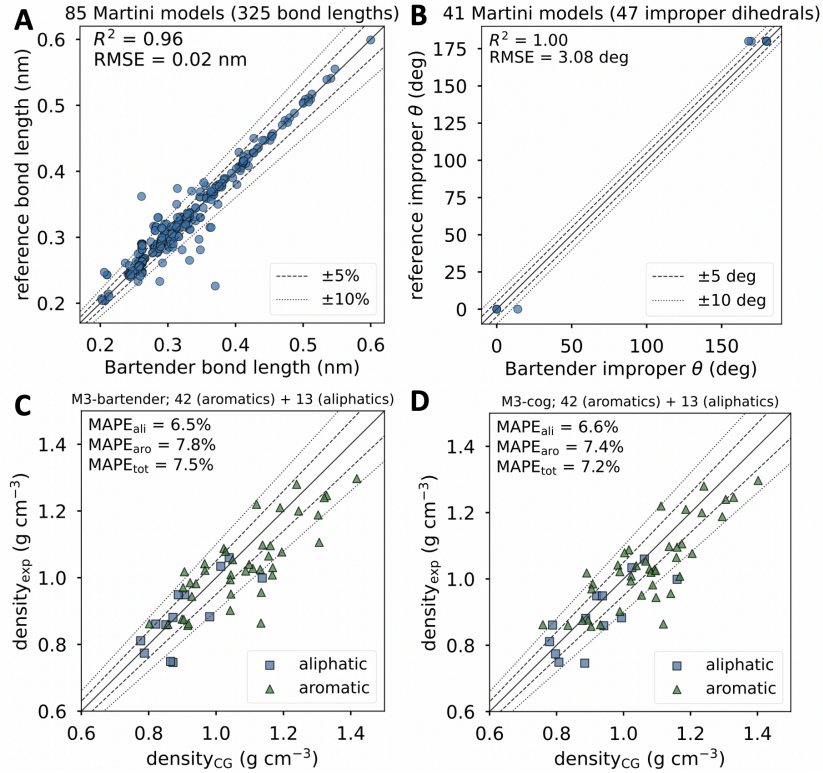

Figure 2. Bartender benchmark on the Martini 3 small molecule dataset bonded parameters. Correlation between Bartender-obtained and the reference, human-obtained (A) bond lengths and (B) improper dihedral angles for the 85 COG-mapped models of the Martini 3 small molecule dataset.[25] The dashed and dotted lines represent error percentages and absolute errors for the bond lengths and improper dihedral angles, respectively. Correlation of the experimental mass density and the computed mass density obtained with the models parameterized using (C) Bartender and (D) human experts.[25] In (C) and (D), green triangles denote data points associated with aromatic compounds, while blue squares represent those related to aliphatic compounds, while dashed and dotted lines denote threshold error lines of ±5% and 10%, respectively.

3.2 Conjugated systems and drug-like: flexible small molecules dataset

3.2.1 - Bartender enables CG parameterization via "easy-to-use" QM simulations

Here we selected a new dataset of ten flexible molecules, exclusively parametrized for this work, which contained: (1) periodic and improper dihedrals, (2) hindered dihedrals, or (3) multimodal angles. The dataset (Supplementary Information) comprised five small organic compounds, three approved drugs (Pomalidomide, Pitolisant and Thyroxine) and two plastic dimers (PET and PEF). Bartender may use different potential forms to fit the bonded parameters: Restricted bending potentials[41] for angles that are part of a dihedral, or RB potentials for dihedrals, as opposed to harmonic and simple periodic functions, respectively. Because of that, the results were analyzed in this set of molecules directly comparing the distributions of bonded terms (distances, angles, etc) of the QM/Bartender based models and atomistic/manually built ones. As observable in **Figure 3A**, there are significant differences across all of the bonded parameters, in particular for both angles and dihedrals. These differences were expected, as angle distributions will necessarily impact the dihedrals. Interestingly, it appears that Bartender-estimated equilibrium values for bonds (or constraints) and angles are systematically shorter than what is predicted from fitting the AA simulations (**Figure 3B, left plot**). In an opposite trend,the QM-derived bonded-parameters have stronger force constants than their atomistic counterparts, which may indicate that the potentials underlying AA distributions may be too soft. It is important to note that the periodic dihedrals are absent from **Figure 3B** because it was impossible for Bartender to fit a harmonic dihedral for all of them. Comparison between RB and simple, harmonic, dihedrals is not straightforward as the functional forms are different and it is difficult to estimate equilibrium values and force constants from the RB parameters.

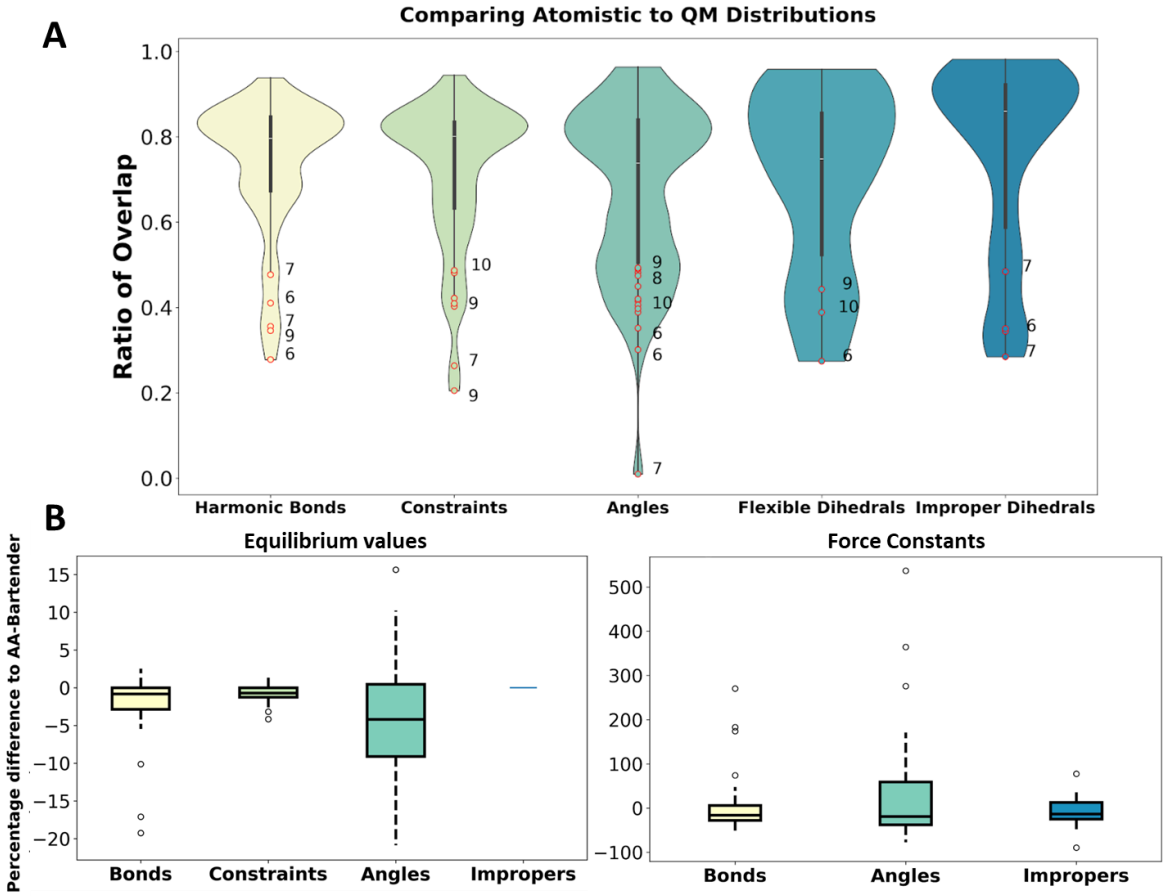

Figure 3. Comparison between parameters extracted from Bartender-fitting QM and AA MD simulations of flexible molecules. A) Martini 3 bonded parameter overlap between mapped QM and atomistic simulations, where 0 means no overlap between distributions and 1 corresponds to perfect overlap. The overlap is computed by comparing the probability distribution of each parameter from the CG-mapped QM simulations to the atomistic simulations. Models 6 to 10 corresponds to the drug-like compounds PET, PEF, Pitolisant, Pomalidomide and Thyroxine. **B)** Comparison of equilibrium and force constant values obtained from fitting the QM and AA trajectories.

Additionally, since the QM simulations are only 100 ns long, there is always the possibility that the QM simulations suffer from undersampling of the potential energy landscape, although the use of implicit solvent may speed-up the conformational dynamics of the molecule. It is interesting to note that the large majority, if not all, of the outliers here noted correspond to either the drug molecules (models 8,9,10) or the plastic dimers (models 6,7), which are clearly the harder molecules in our dataset. As an example, we detail the results obtained for Thyroxine.

3.2.2 - Challenge case: thyroxine

The importance of nuclear receptors is well-known, as they regulate important cellular processes impacting cell homeostasis and development, metabolism and reproduction.[52] Binding of small-molecule ligands leads to their modulation, affecting downstream signaling cascades[53]. Among them, isoforms of the thyroid receptor regulate growth and cell metabolism by binding thyroid hormones[54]. Thyroxine (also known as T4) is a non-steroidal and hydrophobic hormone produced in the thyroid gland. It is shielded and protected in the bloodstream by binding to several proteins, mainly to the thyroxine-binding globulin (TBG) but also to transthyretin and albumin[55]. In solution, thyroxine can explore two distinct conformations, which are interconvertible by rotating the outer ring from one side of the middle ring to the other: cisoid, where the ring on the end points upwards, and transoid, when it points downward[56]. The free energy barrier governing the cisoid-to-transoid dihedral rotation was estimated, at 185K, to be around 37 kJ/mol[56]. As our simulations were performed at a higher temperature, we calculated the barrier at both temperatures, at the r2SCAN-3c composite density functional level of theory[57–59] for geometries and thermal corrections and the SCS-MP2/def2-QZVPP[60,61] level for the electronic energy, employing the RI approximation in both cases[59,62,63], and the Turbomole[64] and Orca[65] program packages, with COSMO-RS[66] for the solvation

free energy (more details are given in the Supporting information). We find barriers of 38.5 kJ/mol at 185K, in very good agreement with the experimental data, and of 43.6 kJ/mol at 298K. The barrier is essentially unchanged with temperature, considering the over 100K of difference between the two conditions.

Experimental data shows that thyroxine binds to TBG in the cisoid conformation[67,68] but to the canonical pocket of thyroid receptors in the transoid configuration,[69,70] hinting that proper parameterization of the dihedrals within the molecule is fundamental to developing a realistic thyroxine model. In particular, the parameterization strategy must take into account that the free energy barrier associated with ring-to-ring rotation is high even at room temperature, as shown by our quantum-chemical calculations, and that both transoid and cisoid configurations are accessible but not easily interconvertible. Thus, thyroxine constitutes an interesting showcase to compare atomistic to QM-derived parameters for CG simulations. The molecule also has four iodine atoms, which can be challenging to properly parametrize in classical atomistic force fields.

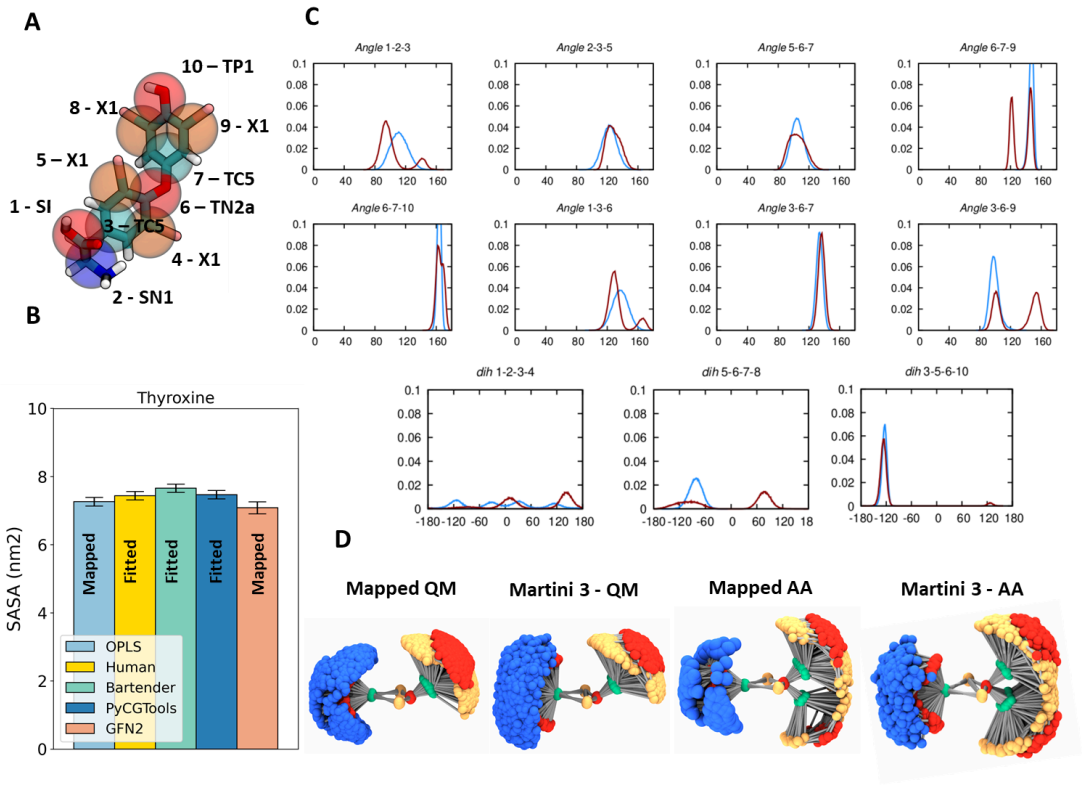

Figure 4. Thyroxine challenge case. A) AA structure with Martini 3 beads overlaid on top. The numbers identify the bead number and the letters the bead type. **B)** Average Solvent-Accessible Surface Area (SASA) estimated for each simulation (Atomistic-mapped, Human-generated CG model, Bartender model, PyCGTool model and QM-mapped, respectively). **C)** Distribution plots for angles and dihedrals of thyroxine, computed from a mapped QM simulation at 298.15K (blue), and a mapped AA simulation (red). **D)** Configurational ensembles sampled during the QM-MD simulation, the AA simulation, and the Martini 3 simulations parametrized based on either QM or AA MD.

The results from this showcase are summarized in **Figure 4,** where we compare the performance of Bartender to either human-generated or pyCGtool-based models. The overall molecular volume is consistent across the various models, as assessed via the average Solvent Accessible Surface Area (SASA) calculations (**Figure 4B**). It is interesting to note that for most angles, the equilibrium values extracted from AA MD are significantly larger than those for QM-MD. Additionally, for angles 3-6-9 and 6-7-9, AA MD samples two states whereas the QM simulations were locked in the energy basin furthest away from 180 degrees. These angles are important because they impact the inter-ring dihedral 5-6-7-8 (**Figure 4C)**. In particular, since only one of the states is accessible in the QM simulation, the ring-to-ring orientation is locked into the conformation shown at the left of **Figure 4D**. In the atomistic simulation, both basins are accessible in either of the angles mentioned, which enables the rotation around the

5-6-7-8 dihedral and leads to the molecule sampling both configurations (transoid and cisoid) but with a greatly underestimated free energy barrier (7 kJ/mol as estimated from the atomistic simulation). Since Martini 3 models are typically parameterized from AA MD, it follows that the CG simulations will naturally capture both these states, separated by a low-energy barrier. However, the evidence in the literature and our QM results indicate that the AA parameters for this dihedral rotation comprise a too low force constant, with potential downstream impact in binding-free energy estimates when studying the thyroxine-thyroid receptor system. For example, it is expected that the entropy of binding for this "too flexible" model will become too large and hamper binding into its pocket in the receptor during MD simulation. Thus, the test case highlights the limitations of automatically generated atomistic parameters, and the accuracy of the GFN2 method. However, given that the QM simulation was not able to sample the transition, undersampling of the potential energy landscape may bias the analysis.

3.3 Assessing the performance of Bartender

3.3.1. Accuracy of the fitting

To verify if the difference obtained between QM/Bartender based models and atomistic/manually build models described in the previous sections were related to the reference model (QM vs atomistic) or with fitting (semi-automatic in Bartender versus manual), we performed comparisons focusing solely on the fitting. Bartender parameters were compared to those obtained from a trial-and-error parametrization strategy (Human) and from PyCGTool. The comparison was carried out by evaluating the overlap between the bonded-parameter distribution from the same reference AA simulations for a particular parameter (bond, angle or dihedral) and the distribution obtained from the simulations using either of the strategies above for the same parameter (**Figure 5**). PyCGtool also gives an estimate of the quality of the automatic protocols, using a parametrization tool already well established in the field. In general, it appears that Bartender and PyCGTool are comparable in terms of fitting ability (overlaps of 0.74 and 0.73, respectively) and that the human-optimized parameterization strategy is slightly superior to both (overlap of 0.79). It appears Bartender is slightly better than PyCGTool in reproducing AA distributions for constraints (0.85 and 0.75 overlap, respectively), whereas PyCGTool (0.74 and 0.65 overlap) is better than Bartender (0.61 and 0.63 overlap) in terms of bonds and angles, respectively. Nonetheless, Bartender is clearly superior when it comes to fitting periodic dihedrals (0.817 overlap, with respect to 0.65 for PyCGTool), save for one dihedral which Bartender failed to fit. This improvement in terms of fitting quality is in part due to the RB potentials available in Bartender, and missing in PyCGTool. As mentioned in the Methods section, PyCGTool is only able to fit harmonic periodic dihedrals (dihedral function type 1 in Gromacs), which in many cases may fail to reproduce the atomistic-mapped reference distributions. The RB dihedrals are particularly useful to fit freely rotating dihedrals which may sample several minima with different energy barrier heights (see section 2 of Supplementary Information, molecule 7 - PEF, dihedral 9-10-11-12). It is noteworthy that the fitting of some angles, harmonic bonds and improper dihedrals distributions, seems to be suboptimal. This is because these particular parameters correspond to multi-modal bonds, angles and impropers where the best case scenario in many cases is to fit a harmonic potential in the middle of both minima with a soft force constant, such that it is possible to sample configurations on both basins. Nonetheless, this means that the maximal achievable value of overlap to the AA simulations in these cases is around 0.5 (for an example on how to compute the distribution overlap, see **Figure S1**). The overlap metric serves as a useful indicator, directing users to areas where potential model enhancements may be necessary. However, we contend that any value below 0.5 suggests a parameter that may be poorly suited and warrants investigation, while values below 0.2 likely indicate incorrect parameters.

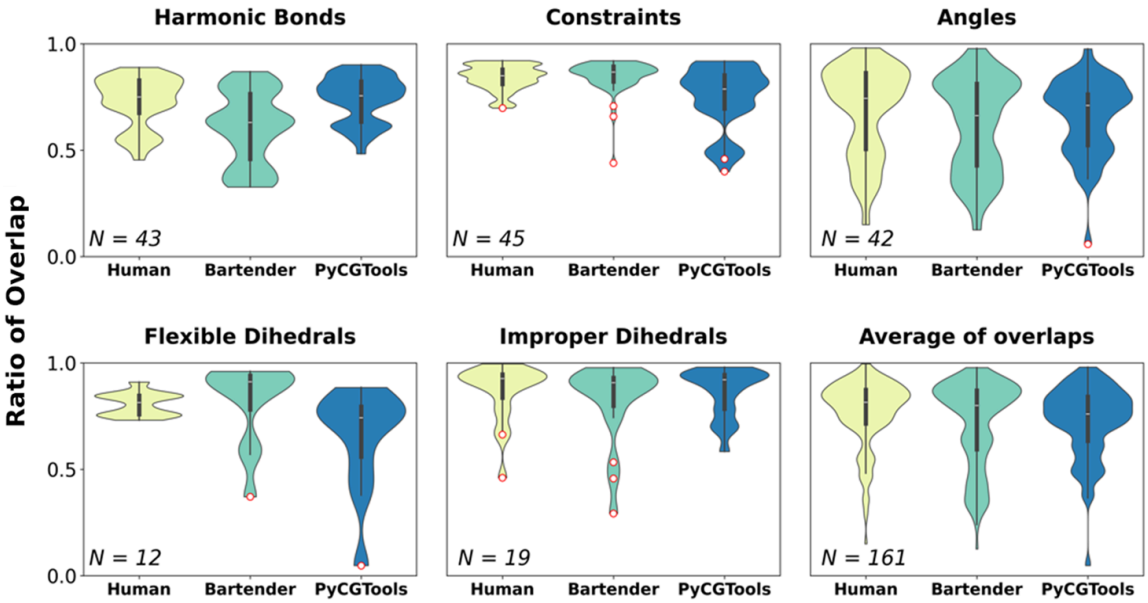

Figure 5. Bartender benchmark on conjugated systems and/or drug-like small-molecules. Comparison across Martini 3 bonded parameter distributions obtained through different parameterization strategies with respect to distributions obtained from a mapped AA reference. The explored strategies were trial-and-error (Human, yellow), Bartender fitting (green) and PyCGTool fitting (blue), each having a small number of outlier points (red). The overlap values have the same meaning as described in the caption of Figure 3.

3.3.2. Computational performance

In exploring the computational performance of Bartender with the flexible molecule dataset (**Table 1**), we find that, on average, Bartender is able to produce an average 9 ns of QM simulations per day using only 1 CPU, which is a reasonable performance given the systems tested. By moving from QM to a molecular mechanics (MM) force field in AA MD with explicit solvent in GROMACS, the performance increases 4-fold, with an average performance of 44 ns/day in one CPU and without using the GPU. The inclusion of the GPU would increase simulation efficiency significantly. In the case that a more efficient simulation approach is required, xtb provides an alternative to the QM simulation by means of a MM forcefield in implicit solvent (GFNFF), which we benchmarked here as well. It appears that GFNFF simulations are much more efficient than QM (134 ns/day and 9 ns/day, respectively), and may be an alternative for simple molecules. Further speed up may be obtained with CPU parallel computing, which is available in xtb/Bartender. Thus, Bartender is not only accurate in generating appropriate initial configuration for Martini 3 simulations, it also does so efficiently, even in the case of flexible molecules. In the case of the rigid molecule set, it appears that 1 ns is enough to sample appropriately the phase-space. However, this simulation length may not be sufficient for converging on the dynamical motions of more flexible molecules, and we believe that at 100 ns simulations, as detailed in the methodology section, represents a reasonable compromise between phase-space exploration and time.

Table 1. Comparison of computational performance across the flexible small-molecule dataset. Timings were estimated in calculations using only 1 CPU in ns/day

Molecule	Bartender GFN2	Bartender GFNFF	All-atom MD
Pomalidomide	12.6	180.0	46.5
PEF	6.1	130.9	27.7
Pitolisant	7.9	120.0	48.0
PET	6.4	120.0	27.7
Thyroxine	6.3	144.0	48.0
M21	11.2	144.0	46.5
M22	12.5	160.0	51.4
M23	7.9	110.8	48.0
M24	7.3	102.9	48.0
M25	10.3	130.9	49.7
Average (ns/day)	8.85	134.3	44.1

*All data is provided using a single cpu with two threads in an Intel Xeon Silver 4210R with 64 Gb RAM memory.

3.4 Limitations and future developments

A limitation of the current implementation of Bartender is the computational resources required to adequately

sample the conformational state of flexible molecules. Though fast, GFN2-xTB is still a QM-based method, and, as such, it is expensive when compared to classical MD. Currently, the main option for large systems, if computer power is limited, is to employ the classical variant GFNFF, and accept the loss of accuracy. GFNFF is still expected to be at least competitive with other force fields[39]. Alternative solutions, such as enhanced sampling schemes based on replica-exchange molecular dynamics[71,72] are currently under development. Nevertheless, the emergence of quantum computers[73] may make direct parametrization of CG models of large and flexible molecules using QM MD feasible in the coming years.

Since Bartenders' bonded-parameter determination relies on a fitting procedure, in some cases the fitting may fail. It is thus recommended that users check not only the produced parameters but also the corresponding plots before proceeding with the CG models. For example, in **Section 3.2**, some parameters were refined using some of the options available in Bartender. Additionally, there are cases where the angles that are required to build a dihedral within a CG model are too close to 180º or have a too-low predicted force constant. In such cases, Bartender will purposely shift the equilibrium value of that angle, assign it as a restricted-bending angle potential[41] and increase its force constant. This is done to guarantee numerical stability of the model but comes at the cost of not reproducing the underlying distribution. As such, it is important to note that Bartender works as a parameterization-aide and the quality of the fits (and of the model itself) is also dependent on the mapping and the bonded parameters deemed necessary by the user. Pipelines aiming at defining optimal mapping and bead-type assignment for Martini 3 small molecules are currently under development. Alternatively, Bartender could be integrated into tools such as AutoMartini[27] and cg_params[29], as a way to refine the bonded parameters with QM accuracy.

The accuracy of angle and bond fittings, as observed previously, is comparatively suboptimal in comparison to alternative programs. Methods for automatically scaling the force constants to achieve better fits are currently being developed. For example, the case of bi-or-multimodal angles in Martini models poses a challenge to the parameterization process with the currently implemented harmonic potentials. This challenge is seen across the tests presented in section 3.3.1, where in many of the cases the overlap percentage is below 60% due to either fitting the harmonic potential on one of the wells or fitting it in the middle of the two minima. A clear path towards improvement would be to implement more complex functional forms directly in codes as OpenMM[74], or via tabulated potentials in GROMACS.

4. Conclusions and perspectives

In conclusion, this paper introduces Bartender, a powerful tool for assisting in the generation of bonded parameters for Martini 3 CG models of small molecules. Through the use of modern semiempirical QM methods, Bartender offers a unique approach that bypasses both the limitations and the inherent setup challenges of classical AA MD simulations. The implementation of a variety of functional forms for both angles and dihedrals enhances the fitting accuracy to reference distributions. Bartender's command-line interface makes it suitable for high-throughput applications.

The presented results showcase Bartender's effectiveness in generating Martini 3 models for both rigid, ring-like structures and flexible, drug-like small molecules. The tool demonstrates high correlation and low root mean square error when compared to human-obtained bonded parameters in the Martini 3 small molecule dataset. In the case of conjugated systems and drug-like molecules, Bartender's fitting approach competes well with PyCGTool, yielding superior results for proper dihedral parameters which are typically difficult to obtain. As a tool, Bartender has an in-place system of warnings and alerts which notifies the user for potentially problematic parameters. It is also able to identify and add angles which are involved in some of the user-defined dihedrals which were not set in the input file. As such, it facilitates model optimization by both providing plots showing the bonded-parameter distributions and highlighting those that potentially need refinement. Moreover, by enabling CG parameterization through easy-to-use QM simulations, Bartender offers an alternative strategy that is particularly valuable when AA parameters may be suboptimal. Particularly, Bartender-generated parameters arising from QM simulations can be useful within future drug-discovery pipelines which harness the computational efficiency of Martini 3, due to a more faithful reproduction of the molecular motions of ligands. Additionally, QM-derived bonded parameters are also expected to be useful to make electronic predictions at CG resolutions[18,75] with Martini 3.

Supporting Information

Tables containing the degree of overlap to the reference distributions for all parameters of all molecules;

Individual figures for each of the ten drug-like compounds containing mapping, bonded-parameter distributions for simulations arising from the different parameterization strategies and corresponding average SASA values;

The rigid/cyclic molecule dataset and their alternative models obtained with Bartender are made available in https://github.com/ricalessandri/Martini3-small-molecules

A folder containing itp and other files for the flexible molecule dataset will be made available upon request from Batender's GitHub page, which is also the repository for the official release of the code: https://github.com/Martini-Force-Field-Initiative/Bartender

AUTHOR INFORMATION

Corresponding Author

Paulo C. T. Souza − Laboratoire de Biologie et Modélisation de la Cellule, CNRS, UMR 5239, Inserm, U1293, Université Claude Bernard Lyon 1, Ecole Normale Superieure de Lyon, 46 Allée d'Italie, 69364, Lyon, France; Centre Blaise Pascal de Simulation et de Modélisation Numérique, Ecole Normale Superieure de Lyon, 46 Allée d'Italie, 69364, Lyon, France
orcid.org/0000-0003-0660-1301;
Email:paulo.telles_de_souza@ens-lyon.fr

Raul Mera-Adasme − Departamento de Química, Facultad de Ciencias, Universidad de Tarapacá. Av. Gral. Velasquez 1775, Arica, Chile.
orcid.org/0000-0003-0345-6217
Email:rmeraa@academicos.uta.cl

AUTHOR CONTRIBUTIONS

The manuscript was written and revised by all authors. Approval for publication was obtained from all authors.

CONFLICT OF INTEREST

The authors declare no conflict of interest

ACKNOWLEDGMENTS

R.A. acknowledges support by the Dutch Research Council (NWO Rubicon 019.202EN.028). The authors thank SURF (www.surf.nl) for the support in using the National Supercomputer Snellius. R.M-A. and R.A-O. thank ANID-Chile for support under FONDECYT N. 1200200. G.P, L.B.A and P.C.T.S would like to thank the support of the French National Center for Scientific Research (CNRS) and the funding from research collaboration agreements with PharmCADD. This work was granted access to the HPC resources of IDRIS and TGCC under the allocations 2022-A0120713456 and 2023-A0140713456 made by GENCI. We also acknowledge the support of the Centre Blaise Pascal's IT test platform at ENS de Lyon (Lyon, France) for the computer facilities. The platform operates the SIDUS solution developed by Emmanuel Quemener[76]. S.J.M. acknowledges funding from the European Research Council with the Advanced grant "COMP-O-CELL" (101053661).

ABBREVIATIONS

QM - Quantum Mechanics, AA - All-Atom, CG - Coarse-Grained, MD - Molecular Dynamics, RB - Ryckaert-Bellemans, DFT - Density Functional Theory, TBG - Thyroxine-binding Globulin.

References

[1] D. E. Barreto Gomes, K. Galentino, M. Sisquellas, L. Monari, C. Bouysset, M. Cecchini, *J. Chem. Inf. Model.* **2023**, *63*, 407–411.
[2] A. V. Sadybekov, V. Katritch, *Nature* **2023**, *616*, 673–685.
[3] V. Le Guilloux, P. Schmidtke, P. Tuffery, *BMC Bioinformatics* **2009**, *10*, 168.
[4] V. D. Ustach, S. K. Lakkaraju, S. Jo, W. Yu, W. Jiang, A. D. MacKerell Jr, *J. Chem. Inf. Model.* **2019**, *59*, 3018–3035.
[5] V. Gapsys, L. Pérez-Benito, M. Aldeghi, D. Seeliger, H. van Vlijmen, G. Tresadern, B. L. de Groot, *Chem. Sci.* **2019**, *11*, 1140–1152.
[6] M. Papadourakis, H. Sinenka, P. Matricon, J. Hénin, G. Brannigan, L. Pérez-Benito, V. Pande, H. van Vlijmen, C. de Graaf, F. Deflorian, G. Tresadern, M. Cecchini, Z. Cournia, *J. Chem. Theory Comput.* **2023**, *19*, 7437–7458.
[7] A. Heifetz, T. James, I. Morao, M. J. Bodkin, P. C. Biggin, *Curr. Opin. Pharmacol.* **2016**, *30*, 14–21.
[8] T. Steinbrecher, in *Protein-Ligand Interactions*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, **2012**, pp. 207–236.
[9] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, A. H. de Vries, *J. Phys. Chem. B* **2007**, *111*, 7812–7824.
[10] P. C. T. Souza, R. Alessandri, J. Barnoud, S. Thallmair, I. Faustino, F. Grünewald, I. Patmanidis, H. Abdizadeh, B. M. H. Bruininks, T. A. Wassenaar, P. C. Kroon, J. Melcr, V. Nieto, V. Corradi, H. M. Khan, J. Domański, M. Javanainen, H. Martinez-Seara, N. Reuter, R. B. Best, I. Vattulainen, L. Monticelli, X. Periole, D. Peter Tieleman, A. H. de Vries, S. J. Marrink, *Nat. Methods* **2021**, 1–7.
[11] S. J. Marrink, L. Monticelli, M. N. Melo, R. Alessandri, D. P. Tieleman, P. C. T. Souza, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2023**, *13*, DOI 10.1002/wcms.1620.
[12] W. Song, R. A. Corey, T. B. Ansell, C. K. Cassidy, M. R. Horrell, A. L. Duncan, P. J. Stansfeld, M. S. P. Sansom, *J. Chem. Theory Comput.* **2022**, *18*, 1188–1201.
[13] V. Thallmair, L. Schultz, W. Zhao, S. J. Marrink, D. Oliver, S. Thallmair, *Sci Adv* **2022**, *8*, eabp9471.
[14] V. Corradi, E. Mendez-Villuendas, H. I. Ingólfsson, R.-X. Gu, I. Siuda, M. N. Melo, A. Moussatova, L. J. DeGagné, B. I. Sejdiu, G. Singh, T. A. Wassenaar, K. Delgado Magnero, S. J. Marrink, D. P. Tieleman, *ACS Cent Sci* **2018**, *4*, 709–717.
[15] L. Borges-Araújo, P. C. T. Souza, F. Fernandes, M. N. Melo, *J. Chem. Theory Comput.* **2022**, *18*, 357–373.
[16] P. Vainikka, S. Thallmair, P. C. T. Souza, S. J. Marrink, *ACS Sustainable Chem. Eng.* **2021**, *9*, 17338–17350.
[17] L. I. Vazquez-Salazar, M. Selle, A. H. de Vries, S. J.

Marrink, P. C. T. Souza, *Green Chem.* **2020**, *22*, 7376–7386.
[18] R. Alessandri, J. J. de Pablo, *Macromolecules* **2023**, *56*, 3574–3584.
[19] F. Grünewald, P. C. Kroon, P. C. T. Souza, S. J. Marrink, *Methods Mol. Biol.* **2021**, *2199*, 315–335.
[20] F. Grünewald, R. Alessandri, P. C. Kroon, L. Monticelli, P. C. T. Souza, S. J. Marrink, *Nat. Commun.* **2022**, *13*, 68.
[21] E. Diamanti, P. C. T. Souza, I. Setyawati, S. Bousis, L. M. Gómez, L. J. Y. M. Swier, A. Shams, A. Tsarenko, W. K. Stanek, M. Jäger, S. J. Marrink, D. J. Slotboom, A. K. H. Hirsch, *Commun Biol* **2023**, *6*, 1182.
[22] L. R. Kjølbye, G. P. Pereira, A. Bartocci, M. Pannuzzo, S. Albani, A. Marchetto, B. Jiménez-García, J. Martin, G. Rossetti, M. Cecchini, S. Wu, L. Monticelli, P. C. T. Souza, *QRB Discov* **2022**, *3*, e19.
[23] R. Alessandri, S. Thallmair, C. G. Herrero, R. Mera-Adasme, S. J. Marrink, P. C. T. Souza, in *A Practical Guide to Recent Advances in Multiscale Modeling and Simulation of Biomolecules*, AIP Publishing, **2023**, pp. 1–34.
[24] P. C. T. Souza, S. Thallmair, P. Conflitti, C. Ramírez-Palacios, R. Alessandri, S. Raniolo, V. Limongelli, S. J. Marrink, *Nat. Commun.* **2020**, *11*, 3714.
[25] R. Alessandri, J. Barnoud, A. S. Gertsen, I. Patmanidis, A. H. de Vries, P. C. T. Souza, S. J. Marrink, *Adv. Theory Simul.* **2022**, *5*, 2100391.
[26] C. Hilpert, L. Beranger, P. C. T. Souza, P. A. Vainikka, V. Nieto, S. J. Marrink, L. Monticelli, G. Launay, *bioRxiv* **2022**, 2022.08.03.502585.
[27] T. Bereau, K. Kremer, *J. Chem. Theory Comput.* **2015**, *11*, 2783–2791.
[28] J. A. Graham, J. W. Essex, S. Khalid, *J. Chem. Inf. Model.* **2017**, *57*, 650–656.
[29] T. D. Potter, E. L. Barrett, M. A. Miller, *J. Chem. Theory Comput.* **2021**, *17*, 5777–5791.
[30] K. S. Stroh, P. C. T. Souza, L. Monticelli, H. J. Risselada, *J. Chem. Theory Comput.* **2023**, *19*, 8384–8400.
[31] C. Empereur-Mot, L. Pesce, G. Doni, D. Bochicchio, R. Capelli, C. Perego, G. M. Pavan, *ACS Omega* **2020**, *5*, 32823–32843.
[32] S. Raniolo, V. Limongelli, *Front Mol Biosci* **2021**, *8*, 760283.
[33] C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher, S. Grimme, *WIREs Comput Mol Sci* **2020**, *140*, 18A301.
[34] C. Bannwarth, S. Ehlert, S. Grimme, *J. Chem. Theory Comput.* **2019**, *15*, 1652–1671.
[35] Dominguez, M., Jimenez, V., Savasci, G., Araya-Osorio, R., Pesonen, J, Mera-Adasme, R., **n.d.**
[36] "Gonum Numerical Packages," can be found under https://www.gonum.org, **n.d.**
[37] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl, *SoftwareX* **2015**, *1-2*, 19–25.
[38] P. Pracht, E. Caldeweyher, S. Ehlert, S. Grimme, **2019**.
[39] S. Spicher, S. Grimme, *Angew. Chem. Int. Ed Engl.* **2020**, *59*, 15665–15673.
[40] S. Ehlert, M. Stahn, S. Spicher, S. Grimme, *ChemRxiv* **2021**, DOI 10.26434/chemrxiv.14555355.v1.
[41] M. Bulacu, N. Goga, W. Zhao, G. Rossi, L. Monticelli, X. Periole, D. P. Tieleman, S. J. Marrink, *J. Chem. Theory Comput.* **2013**, *9*, 3282–3292.
[42] M. Bulacu, E. van der Giessen, *J. Chem. Phys.* **2005**, *123*, 114901.
[43] V. Rühle, C. Junghans, A. Lukyanov, K. Kremer, D. Andrienko, *J. Chem. Theory Comput.* **2009**, *5*, 3211–3223.
[44] "Grace Home," can be found under https://plasma-gate.weizmann.ac.il/Grace/, **n.d.**
[45] D. H. de Jong, S. Baoukina, H. I. Ingólfsson, S. J. Marrink, *Comput. Phys. Commun.* **2016**, *199*, 1–7.
[46] M. Parrinello, A. Rahman, *J. Appl. Phys.* **1981**, *52*, 7182–7190.
[47] G. Bussi, D. Donadio, M. Parrinello, *J. Chem. Phys.* **2007**, *126*, 014101.
[48] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, G. R. Hutchison, *J. Cheminform.* **2011**, *3*, 33.
[49] L. S. Dodda, I. Cabeza de Vaca, J. Tirado-Rives, W. L. Jorgensen, *Nucleic Acids Res.* **2017**, *45*, W331–W336.
[50] D. J. Evans, B. L. Holian, *J. Chem. Phys.* **1985**, *83*, 4069–4074.
[51] D. Weininger, A. Weininger, J. L. Weininger, *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 97–101.
[52] C. K. Glass, S. Ogawa, *Nat. Rev. Immunol.* **2006**, *6*, 44–55.
[53] R. Sever, C. K. Glass, *Cold Spring Harb. Perspect. Biol.* **2013**, *5*, a016709.
[54] T. M. Ortiga-Carvalho, A. R. Sidhaye, F. E. Wondisford, *Nat. Rev. Endocrinol.* **2014**, *10*, 582–591.
[55] G. C. Schussler, *Thyroid* **2000**, *10*, 141–149.
[56] B. M. Duggan, D. J. Craik, *J. Med. Chem.* **1997**, *40*, 2259–2265.
[57] S. Grimme, A. Hansen, S. Ehlert, J.-M. Mewes, *ChemRxiv* **2020**, DOI 10.26434/chemrxiv.13333520.v2.
[58] J. W. Furness, A. D. Kaplan, J. Ning, J. P. Perdew, J. Sun, *J. Phys. Chem. Lett.* **2020**, *11*, 8208–8215.
[59] F. Weigend, *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057–1065.
[60] S. Grimme, L. Goerigk, R. F. Fink, *WIREs Comput Mol Sci* **2012**, *2*, 886–906.
[61] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.
[62] F. Weigend, M. Häser, H. Patzelt, R. Ahlrichs, *Chem. Phys. Lett.* **1998**, *294*, 143–152.
[63] F. Neese, F. Wennmohs, A. Hansen, U. Becker, *Chem. Phys.* **2009**, *356*, 98–109.
[64] S. G. Balasubramani, G. P. Chen, S. Coriani, M. Diedenhofen, M. S. Frank, Y. J. Franzke, F. Furche, R. Grotjahn, M. E. Harding, C. Hättig, A. Hellweg, B. Helmich-Paris, C. Holzer, U. Huniar, M. Kaupp, A. Marefat Khah, S. Karbalaei Khani, T. Müller, F. Mack, B. D. Nguyen, S. M. Parker, E. Perlt, D. Rappoport, K. Reiter, S. Roy, M. Rückert, G. Schmitz, M. Sierka, E. Tapavicza, D. P. Tew, C. van Wüllen, V. K. Voora, F. Weigend, A. Wodyński, J. M. Yu, *J. Chem. Phys.* **2020**, *152*, 184107.
[65] F. Neese, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2022**, *12*, DOI 10.1002/wcms.1606.
[66] A. Klamt, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 699–709.
[67] A. Zhou, Z. Wei, R. J. Read, R. W. Carrell, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 13321–13326.
[68] J. A. Jarvis, S. L. Munro, D. J. Craik, *Protein Eng.* **1992**, *5*, 61–67.
[69] R. L. Wagner, J. W. Apriletti, M. E. McGrath, B. L. West, J. D. Baxter, R. J. Fletterick, *Nature* **1995**, *378*, 690–697.
[70] P. C. T. Souza, G. B. Barra, L. F. R. Velasco, I. C. J. Ribeiro, L. A. Simeoni, M. Togashi, P. Webb, F. A. R.

Neves, M. S. Skaf, L. Martínez, I. Polikarpov, *J. Mol. Biol.* **2011**, *412*, 882–893.
[71] Y. Sugita, Y. Okamoto, *Chem. Phys. Lett.* **1999**, *314*, 141–151.
[72] R. Qi, G. Wei, B. Ma, R. Nussinov, *Methods Mol. Biol.* **2018**, *1777*, 101–119.
[73] B. Bauer, S. Bravyi, M. Motta, G. Kin-Lic Chan, *Chem. Rev.* **2020**, *120*, 12685–12717.
[74] J. L. MacCallum, S. Hu, S. Lenz, P. C. T. Souza, V. Corradi, D. P. Tieleman, *Biophys. J.* **2023**, *122*, 2864–2870.
[75] C.-I. Wang, N. E. Jackson, *Chem. Mater.* **2023**, *35*, 1470–1486.
[76] E. Quemener, M. Corvellec, *Linux J.* **2013**, DOI 10.5555/2555789.2555792.

TOC Figure