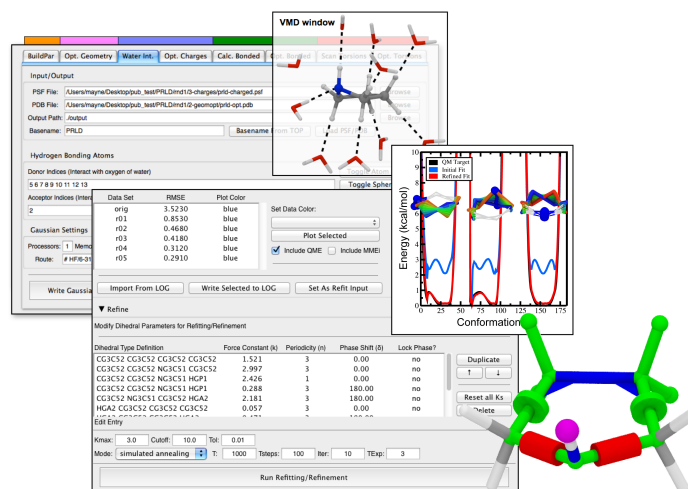Computational Structural Biology and Molecular Biophysics Group
NIH Center for Macromolecular Modeling and Bioinformatics
Beckman Institute for Advanced Science and Technology
University of Illinois at Urbana–Champaign

Parameterizing Small Molecules Using the Force Field Toolkit (ffTK)

Tutorial Written by:
Christopher G. Mayne

Current Editors:
Christopher G. Mayne
Melanie Muller
Emad Tajkhorshid

September 2015

A current version of this tutorial is available at
`http://www.ks.uiuc.edu/Training/Tutorials/`

Contents

1 Introduction

1.1 Molecular Mechanics Force Fields

Molecular dynamics (MD) simulations are a powerful addition to the biophysicist's toolbox for investigating biologically relevant systems at the molecular level. Complimentary to many experimental structural techniques (e.g., x-ray crystallography, electron microscopy, nuclear magnetic resonance), computational techniques such as MD simulations allow for in-depth studies with unparalleled spatial and temporal resolution. Classical MD simulations are founded on a mathematical description for the potential energy of the molecular system. The CHARMM force field for biopolymers[1] is a popular example of a force field typically used for MD simulations. The CHARMM potential energy function, Eq. (1), defines the functional form for computing each energetic contribution. Each specific term, however, contains one or more constants (highlighted with red text) which must be defined on a per-atom basis. As an example, for every bond between two particular atom types in the molecular system, there must be a corresponding force constant (k_b) and equilibrium distance (b_0). Collections of these values are known as parameter sets and the process by which they are derived, "parameterization", is the focus of this tutorial.

$$
\begin{aligned}
U_{\text{CHARMM}} = \quad & \sum_{\text{bonds}} k_b(b - b_0)^2 \; + \\
& \sum_{\text{angles}} k_\theta(\theta - \theta_0)^2 \; + \\
& \sum_{\text{UB}} k_{ub}(b^{1,3} - b_0^{1,3})^2 \; + \\
& \sum_{\text{dihedrals}} k_\psi(1 + cos(n\psi - \delta)) \; + \\
& \sum_{\text{impropers}} k_\omega(b - \omega_0)^2 \; + \\
& \sum_{\text{cmap}} u_{\text{cmap}}(\phi, \psi) \; + \\
& \sum_{\text{atoms } i,j} \frac{q_i q_j}{\epsilon r_{ij}} \; + \\
& \sum_{\text{atoms } i,j} \epsilon_{ij}\left[\left(\frac{r_{ij}^{min}}{r_{ij}}\right)^{12} - 2\left(\frac{r_{ij}^{min}}{r_{ij}}\right)^{6}\right]
\end{aligned}
\tag{1}
$$

$$
\text{where, } \; r_{ij}^{min} = \left(r_i^{min} + r_j^{min}\right)/2 \; \text{ and, } \; \epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}
$$

Classical MD simulations have traditionally focused on proteins, for which many different force fields have been developed (e.g., CHARMM[1], AMBER[2], OPLS[3], GROMOS[4]). Each of these force fields takes advantage of the polymeric nature of proteins (and other structures such as DNA, RNA, and carbohydrates), for which parameters are required to describe a relatively limited set of building blocks (e.g., amino acid residues) and the common linking element (e.g., peptide backbone). The inclusion of small molecules, however, is significantly more challenging and has largely prevented the application of MD technologies to fields such as drug discovery. Recently, general versions of popular biopolymer force fields, such as the CHARMM General Force Field (CGenFF)[5] and the Generalized AMBER Force Field (GAFF)[6], have been specifically developed to target small drug-like molecules while maintaining compatibility with the existing biopolymer versions. While these generalized force fields provide coverage for common elements of many biologically relevant small molecules, it is unreasonable to assume that a single parameter set can provide adequate coverage of small molecule space. To address these limitations in available parameter sets, the task of parameterizing new functional groups and novel linkages between existing groups has become an increasingly important task in force field development and biomolecular simulations at-large.

1.2   A Brief Description of the Tutorial

The goal of this tutorial is to describe the steps required to derive parameters for organic small molecules that are compatible with the CHARMM family of force fields (CHARMM, CGenFF, and specialized variants such as nucleic acid and sugar parameter sets). Achieving this goal requires developing parameters yielding an accurate description of the internal dynamics of the small molecule, while following established protocols to describe intermolecular interactions. In this manner, the newly parameterized molecule can be included in simulations of molecular systems based on established force fields without requiring any modifications to the potential energy function.

The FORCE FIELD TOOLKIT (ffTK)[7] is a plugin for the VMD molecular modeling software package designed specifically to develop CHARMM-compatible parameters for small molecules. ffTK protocols follow the established parameterization philosophies of CGenFF, and the technical details of the plugin are published elsewhere[7]. The graphical user interface (GUI) of the plugin mirrors the parameterization workflow, shown in Figure 1, in which each tab of the GUI represents a step. The contents of each tab provide convenient GUI elements, such as menus, buttons, and file dialogs, to accomplish a variety of modeling tasks. The tutorial begins by constructing the molecular system using another VMD plugin, MOLEFACTURE, followed by traversing each step of the parameterization workflow in ffTK to yield a complete set of parameters for an example molecule, ethanol.

The development of force field parameters is an advanced topic in biomolecular simulations. Ethanol was deliberately chosen as a simple, small, organic molecule because it allows for the demonstration of each step of the parame-
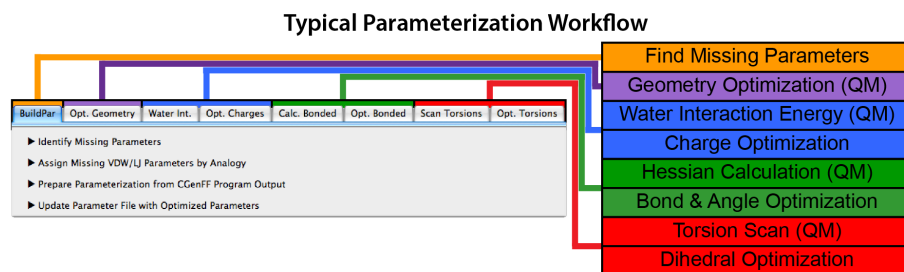
Figure 1: The parameterization workflow maps on to the tabs contained within the ffTK graphical user interface.

terization workflow, yields good fits to the target data, and requires very few computational resources. Although the tutorial traverses the entire workflow, there are many advanced features and analytical tools available within ffTK that are outside of the scope for this tutorial and will not be described. Many of these tools, such as the Charge Optimization LOG Plotter or Torsion Explorer, are indispensible when parameterizing complex molecules or troubleshooting difficult optimizations; we encourage users to seek out these tools once they are familiar with the parameterization workflow and ffTK. In addition to the detailed description for each parameterization step, a complete set of example files are provided. These files should be used to troubleshoot any of the input/output generated during the course of the tutorial, or to provide the QM target data to users who do not have access to the Gaussian09[8] software package (see below).

Organizing the Tutorial Files. Throughout the tutorial, readers will use ffTK to manipulate a large number of different file types, including structural files, input and output files for generating the QM target data, output from ffTK optimizations, and updated parameters files. The tutorial text will refer to directories and filenames as provided in the example files, and readers are encouraged to use a similar directory organization.

1.3 Software and Knowledge Prerequisites

The vast majority of the tutorial steps are performed within the VMD molecular modeling package. Version 1.9.2 of VMD (or higher) is required due to many significant improvements to ffTK and other plugin dependencies. Several of the optimization routines performed within ffTK also require that NAMD 2.9 (or higher) is locally installed, and will request the location of the `namd2` executable binary file. Each optimization requires target data computed at the quantum mechanical level of theory. Currently, ffTK only supports the use of Gaussian09 for these calculations, and it is suggested that users obtain access to this commercial software package. For the purposes of this tutorial, however, all input files and the resulting output data from the QM calculations are provided in the example files; therefore, direct access to Gaussian09 is not required to complete the tutorial.

In addition to the software requirements above, users are expected to have a solid understanding of both the theoretical underpinnings of MD simulations and practical knowledge performing such simulations using CHARMM-compatible structure files. Specifically, the tutorial will involve the creation and modification of varied file types, such as PSF, PDB, and parameter (e.g., PAR, PRM, STR) files. Users should be familiar with the structure of these files and the information contained within. For readers unfamiliar with manipulating molecular systems or performing MD simulations, we recommend the tutorials for VMD, NAMD (especially the appendices for PDB, PSF, Topology, and Parameter files), Topology File, and MOLEFACTURE, available at `http://www.ks.uiuc.edu/Training/Tutorials/`. A critical reading of the papers describing CGenFF[5] and ffTK[7] will provide an excellent background for understanding the theoretical underpinings and technical details of each parameterization task.

2   System Preparation

2.1   Constructing a Small Molecule Using Molefacture

The entry point into the parameterization workflow requires that users provide properly formatted PSF and PDB files that describe the molecule of interest. There are many routes to obtaining these files depending on the availability and familiarity of users for different software packages. Due to the importance of setting the molecule up properly, we will demonstrate how to use the MOLEFAC-TURE plugin to construct a very simple small molecule, ethanol. Important steps include creating the molecule, assigning the necessary molecular and atomic properties, and writing the required PSF and PDB files. Further information on MOLEFACTURE can be found at the resources below.

- Documentation website:
  http://www.ks.uiuc.edu/Research/vmd/plugins/molefacture/

- Tutorials website:
  http://www.ks.uiuc.edu/Training/Tutorials/

1 Launch the MOLEFACTURE GUI in VMD:
Extensions → Modeling → Molefacture
When the startup dialog pops up, leave the selection blank and click the Start Molefacture button.

Molecules are typically constructed in MOLEFACTURE starting from a base molecule by replacing one or more hydrogen atoms with the desired molecular fragments. In this manner, ethanol can be constructed from methane by replacing two of the hydrogens with a methyl and hydroxyl group, respectively.

2 Create the methane base molecule:
Build → New Molecule from Fragment → Methane
If prompted, click Yes to abandon the current molecule.

3 Select the last hydrogen atom (index = 4), and replace it with a methyl fragment: Build → Replace hydrogen with fragment → Methyl

4 Repeat this process, selecting the last hydrogen (index = 7) to add a hydroxyl group:
Build → Replace hydrogen with fragment → Hydroxyl

Next, we will modify the molecular and atomic properties from the Atoms and Molecular sections of the GUI (Fig. 2). Atomic values are set by picking an atom followed by clicking the Edit selected atom button, which brings up a dialog box where each property is set. Molecular properties are set in an analogous manner using the buttons provided in the Molecule box. Values for the atomic and molecular properties are provided in Figure 3, for which the type and charge columns are the most notable. The provided atom types were taken from the
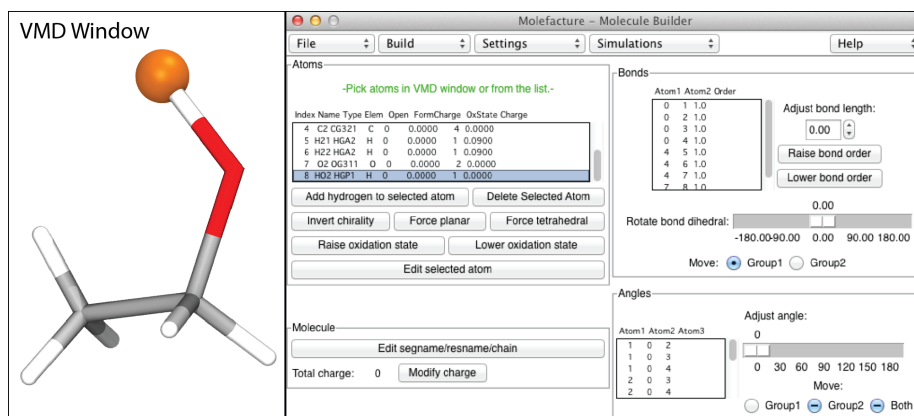
Figure 2: Molefacture GUI after constructing and editing ethanol.

CHARMM General Force Field (CGenFF), where they are routinely used for atoms comprising aliphatic hydrocarbons and alcohols. Although optimizing partial atomic charges is a critical step in the parameterization workflow, it is convention to fix the charge of non-polar hydrogens to +0.09 for CHARMM-comptible force fields. Therefore, charges for these atoms are set during the system preparation phase while the remaining charges, currently set to 0.0, will be computed at a later step. Once all of the properties are set as given in Figure 3, MOLEFACTURE can write the molecular data out to the desired PSF and PDB file types.

Index	Name	Type	Charge	Element
0	C1	CG331	0	C
1	H11	HGA3	0.09	H
2	H12	HGA3	0.09	H
3	H13	HGA3	0.09	H
4	C2	CG321	0	C
5	H21	HGA2	0.09	H
6	H22	HGA2	0.09	H
7	O2	OG311	0	O
8	HO2	HGP1	0	H

Molecular Properties:

resname	ETOH
segname	L
chain	L
resid	1

Figure 3: Molecular and atomic properties for ethanol.

5 Update the atomic and molecular properties provided in Figure 3.

6 Write out the PSF and PDB files for ethanol to the `1-sysPrep` directory using `ETOH.0` as the output prefix:
File → Write psf and pdb files

7 Quit MOLEFACTURE:

File → Quit

Since we have already written out the PSF/PDB files, there is no need to save the molecule if prompted upon exit.

Limitations to Atomic Properties in PSF and PDB File Formats. When setting properties, be aware of length limitations imposed by file formats. MOLEFACTURE and its dependencies limit atom names to 4 characters, atom types to 7 characters, resname to 4 characters, segname to 4 characters, and chain to 1 character. Violations of these limits will not always generate an error message, so it is prudent to carefully check the contents of the PSF and PDB files after generation.

---

ffTK is Sensitive to Atomic Order. ffTK is **very** sensitive to the order in which atoms are listed in the PSF and PDB files. Any modifications to the structure that alters the atomic indices will effectively scramble the methods used to connect parameterization data to the molecular structure (e.g., QM target data, ffTK LOG data). Accordingly, it is crucial that readers construct the molecular structure at the beginning of the workflow, and only modify properties using the tools provided within ffTK. For readers that plan on using the QM target data provided in the example files (e.g., don't have access to Gaussian), we suggest replacing the PSF and PDB files constructed in the previous step (System Preparation, 2.1) with those provided in the example files. This will insure that the atomic order contained in the initial structure matches that of the example files for all subsequent steps.

2.2 Creating an Initial Parameter File

The remainder of the tutorial will proceed through the parameterization workflow using the FORCE FIELD TOOLKIT (ffTK).

1 Start the ffTK GUI:

Extensions → Modeling → Force Field Toolkit

PSF files contain the topological and charge information that define all of the terms required to compute the potential energy. The parameter file (PAR), however, defines the quantitative constants (red elements in Eq. (1)) for each of the bonded and Lennard-Jones (LJ, or vdW) terms. The first step of a parameterization task is to identify which of these parameters are unknown and where they appear in the molecule. ffTK provides several tools to analyze the required parameters, cross-check them against existing parameter sets, visualize them within the context of the 3D molecular structure, and write out only the missing parameters to an initialized PAR file. As described in the introduction, this tutorial covers a full parameterization of ethanol without considering any terms that may be present in existing parameter sets.

2 Navigate to the BuildPar tab and expand the Identify Missing Parameters section.

3 Use the provided file dialogs to specify the locations to the PSF and PDB files created during system preparation, `1-sysPrep/ETOH.0.psf` and `1-sysPrep/ETOH.0.pdb`, respectively. Leave the box entitled Associated Parameter Files empty.

4 Run the analysis via the Analyze button.

> **Ethanol.** Ethanol is a common organic molecule for which parameters are provided in the standard CGenFF distribution. For the purposes of this tutorial we will ignore these existing parameters to demonstrate how to take a simple molecule through the entire parameterization workflow from start to finish. When performing partial parameterizations that use terms from existing parameter sets, the associated parameter files should be provided here so that ffTK can cross-check the parameters required for the molecule of interest against those provided.

Upon completing the parameter analysis, ffTK populates the series of boxes entitled Bonds, Angles, Dihedrals, and Nonbonded located just below the Analyze button, and loads the ethanol structure into the main VMD window, as shown in Figure 4. Each entry represents a parameter of the given type that is required to describe ethanol. Clicking on any term, or collection of terms (shift+click) in the GUI draws a colored graphic object where the selected parameter(s) appear in the molecule. Each object type (shapes representing a bond, angle, dihedral, or vdW term) is designed such that one or more parameters can be selected, including overlapping terms, without obstructing each other. Individual terms can be deselected using a control+click, or all terms can be deselected for the box in focus by pressing the "escape" key. Double clicking an entry will toggle the parameter between "active" (black text) and "inactive" (grey text). Deactivated parameters will be excluded when writing an initial parameter file.

All of the entries from the analysis above represent the full parameter set required to describe the loaded molecule (i.e., ethanol) but were not found in any of the parameters provided as "Associated Parameter Files." In order to perform MD simulations that include our molecule, we must construct a parameter file that defines each of these terms. ffTK will gather all terms and write out a properly formatted parameter file with all values initialized to 0.0. The remainder of the tutorial will refer to this file as the In-Progress PAR file, and will serve as the starting point of our parameterization.

5 Use the SaveAs button to specify where to store the initialized parameter file and name it `1-sysPrep/par_ETOH.0.par`. Click Write Initial Parameter File to write the file.

Currently, ffTK does not support the optimization of vdW terms (i.e., nonbonded LJ parameters) for new atom types. In practice, this is only a minor limitation as LJ parameters exist for most atoms found in organic small molecules,
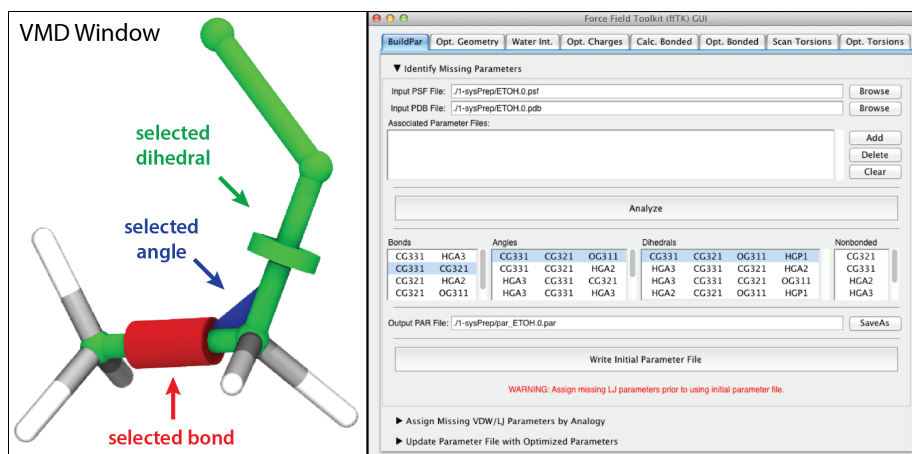
Figure 4: Missing parameters can be visualized within the context of the 3D molecular structure. Your VMD window may look slightly different depending on the default represenation settings.

or are largely transferable between highly related atoms. Since we are ignoring any pre-existing parameter sets for this tutorial, we must manually define the vdW terms required for ethanol. ffTK can facilitate making these assignments and update the In-Progress PAR file from the Assign Missing VDW/LJ Parameters by Analogy section, also found within the BuildPar tab.

6 Navigate to the BuildPar tab and expand the Assign Missing VDW/LJ Parameters by Analogy section.

The top portion of this section is used to read in the In-Progress PAR file containing the unassigned LJ parameters, while the bottom portion operates as a "parameter browser" in which one or more pre-existing parameter sets (topology + parameter files) are parsed and organized for easy browsing. Users can then scroll through existing parameter sets to find appropriate values for any unassigned LJ parameters. When the desired terms are found, the parameter values can be copied from the reference set by selecting the target and source entries from the top and bottom sections, respectively, and clicking the Set from Reference button. The atom types provided in Figure 3 are commonly used CGenFF types. Accordingly, we will load CGenFF files into the browser to search for appropriate LJ parameters.

First, load the In-Progress PAR file into the top portion of the GUI using the provided Browse and Load buttons. Upon loading, you should see the VDW/LJ Parameters box is populated with all of the atom types that appear in our ethanol molecule.

7 Use the Browse button to specify the In-Progress parameter file (labeled as "Incomplete PAR" in the GUI), `1-sysPrep/par_ETOH.0.par`, and click Load.

Second, load in the CGenFF topology and parameter files for browsing using
the buttons provided in the bottom portion of the GUI. We have provided you
with the standard CGenFF topology and parameter files in the `common` subdirec-
tory, named `top_all36_cgenff.rtf` and `par_all36_cgenff.prm`, respectively.

8 Use the Load Topology + Parameter Set button to specify the CGenFF
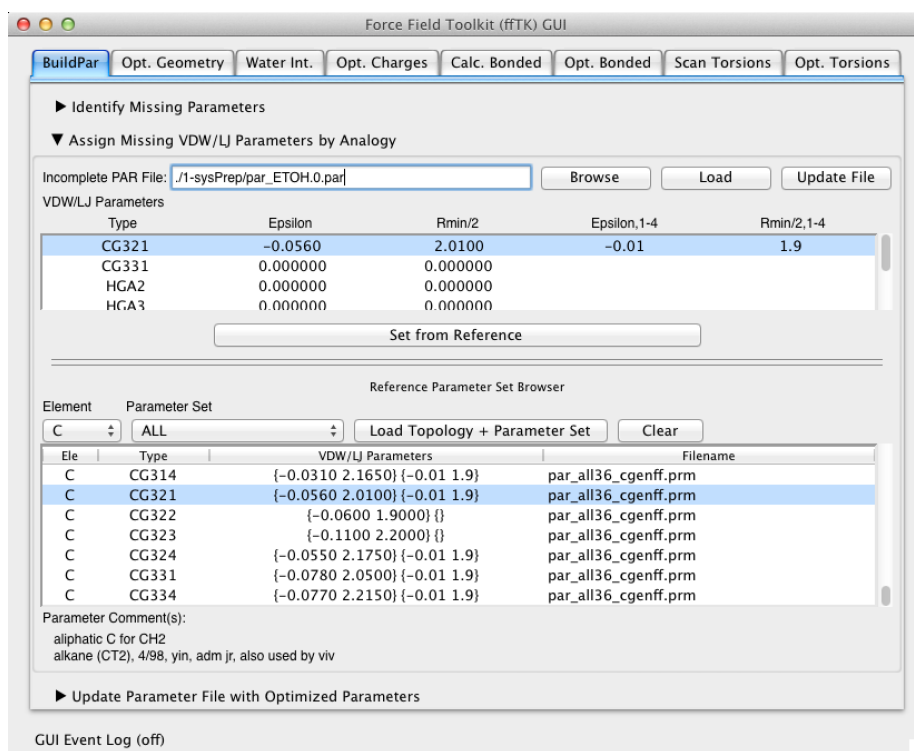topology and parameter files, in that order.

Figure 5: Unassigned Lennard-Jones parameters are easily assigned by analogy
from existing parameter sets using the parameter browser.

After both parameter set files are specified, the browser box is populated
with every atom type found in CGenFF. Note the Element and Parameter Set
drop-down menus, which are convenient for browsing only a subset of the loaded
atom types. For instance, when assigning the two different types of carbon atoms
in ethanol, selecting "C" from the Element drop-down menu will show only the
atom types for carbon in the lower browser box. Since we used standard atom
types from CGenFF when constructing our molecule of ethanol, we can simply
search for each atom type directly.

9 Select the entry for the unassigned "CG321" atom type in the upper box.

10 Select "C" from the Element drop-down to filter the reference parameter set, showing only carbon atom types. Scroll through the available atom types to find the corresponding entry for "CG321" and select it.

Upon selecting the specified entry, the Parameter Comments section of the browser, located just below the browser box, provides any comments that are associated with the selected atom type. From Figure 6, we see that CG321 is described as "aliphatic C for CH2" along with additional comment information. The information provided here is often critical for assigning LJ parameters for atom types that do not have an exact match in CGenFF and must be assigned by analogy.

11 With the corresponding entries set in the upper and lower boxes, click on the Set from Reference button to copy the parameter values.

12 Repeat this process for all unassigned parameters of ethanol listed for the In-Progress PAR file.

13 Update the In-Progress PAR file by clicking on the Update File button, located in the upper section. This action will rewrite the file with the new LJ parameters.

Throughout the remainder of this tutorial, the initialized parameter file created in the preceding step is referred to as the "In-Progress" parameter file. Together, the PSF, PDB, and PAR files form a complete set required to perform the calculations that follow. Many of the properties contained within these files will be updated as we perform optimizations of charges, bonds/angles, and dihedrals, in the sections that follow.

3 Geometry Optimization

Using Gaussian09 to Perform Quantum Mechanics Calculations. Geometry Optimization is the first section that requires performing a quantum mechanics calculation. As described in the introduction, ffTK currently only supports use of the Gaussian09 software package, which will be used to generate the target data that guides each optimization. It is assumed that the user has access to this package on their own computing resources, and is familiar with running Gaussian jobs. ffTK does not perform these calculations for you, nor can TCBG provide users with access to this commercial software package. For the purposes of this tutorial, however, we will provide examples of input and output files for the QM calculations that are required to parameterize ethanol.

The first required calculation is a geometry optimization performed at the quantum mechanical level. The resulting low energy conformation will be used as the input conformation for all subsequent calculations. The Opt. Geometry tab provides a compact graphical interface to setup the calculation, read in the optimized coordinates, and write the resulting structure to a PDB file.

1 Select the Opt. Geometry tab from the ffTK GUI.

The upper portion of the GUI is designed to setup the QM geometry optimization calculation. Users first supply the location to the initial PDB coordinates and where to write the Gaussian input file (GAU), using the file dialog buttons associated with each entry. The GUI also provides the Gaussian Settings section which allows user to modify specific details of the calculation. Users may generally want to change the number of processors and available memory to best leverage their computing resources. The charge and multiplicity should be set as appropriate for the molecule of interest. The Route section, however, should only be modified by expert users that are familiar with the function and syntax of Gaussian keywords. Only under very special cases should the default Route be changed. Finally, clicking the Write Gaussian Input File button constructs the Gaussian input file, writing it to the location supplied by the input location above. This is the only file that is required for the QM calculation, and the user is to run this calculation where they have access to Gaussian.

2 Use the provided file dialogs to specify the location of the input PDB file, `1-sysPrep/ETOH.0.pdb`, and a filename used to write the Gaussian input file (denoted as the Output GAU file in the GUI), `2-geomOpt/ETOH.geomOpt.gau`.

3 Leaving the default QM settings, write the input file by clicking the Write Gaussian Input File button.

4 Optional: Run the calculation where you have access to Gaussian. If you do not have access, use the output provided in the example files.

Upon completion of the QM geometry optimization, check the Gaussian LOG file to ensure that the calculation terminated without error, which can be done by using a text editor to read the final message printed to the LOG file, e.g., "Normal termination of Gaussian...". The lower portion of the Geom. Opt. tab can be used to read in each step of the optimization for visualization, or to write out the file coodinates to a PDB file.

5 Use the provided file dialogs to specify the following input:
PDB: `1-sysPrep/ETOH.0.pdb`
G09 LOG: `2-geomOpt/ETOH.geomOpt.log`
Output: `2-geomOpt/ETOH.1.pdb`

6 Optional: To visualize each step of the geometry optimization, click on the Load Gaussian LOG File.

7 Write the updated coordinates to file by clicking the Write Optimized Geometry to PDB button.

The PDB file containing the updated coordinates should be used for the remainder of the tutorial whenever a PDB file is requested.
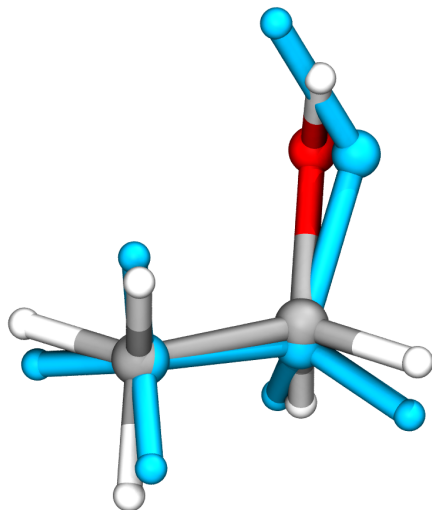

Figure 6: Initial (cyan) and final (by element) geometries.

4 Charge Optimization

According to the CHARMM philosophy[1,5], partial atomic charges are derived to reproduce interactions with explicit TIP3P[9] water molecules. This is accomplished by first characterizing optimal interactions of a water molecule with specific "interaction sites" on the ligand at the quantum mechanical (QM) level of theory. The resulting interaction profiles serve as target data for a second step whereby the partial atomic charges are modified until the interaction profile, now computed at the molecular mechanics (MM) level of theory, converges to within an acceptable fit of the QM profile. This two-step process, a common paradigm in the parameterization workflow, is split into two tabs within ffTK, Water Int. and Opt. Charges. The details of each step are described in detail below.

4.1 Characterizing Water Interaction Sites

The water interaction target data is obtained by probing optimized positions of a water molecule with respect one or more "interaction sites" on the ligand. Water molecules are oriented to adopt an ideal hydrogen bonding geometry with the interaction site, and a 2-dimensional optimization is performed on 1) the distance between the water and interaction site, and 2) the rotation of the water molecule around the interaction vector (Fig. 7). Typically, each atom that is water-accessible, i.e., the approach of a water molecule is not obstructed by adjacent bonded atoms, is defined as a water interaction site. Sites are further categorized as "hydrogen bond donors", which interact favorably with the oxygen of the water molecule, or "hydrogen bond acceptors", which interact favorably with the hydrogen of the water molecule. The ffTK GUI provides entry boxes to specify and categorize each atom manually, facilitated by the ability to toggle atom labels and spheres that denote the hydrogen bond type in the VMD window. Further, an automated routine can attempt to guess the interaction sites, albeit using relatively simple heuristics.

1 Select the Water Int. tab of the ffTK GUI.

For the case of ethanol, assigning the donors and acceptors is trivial: all hydrogens are "donors" and the oxygen is the only "acceptor". Since both carbon atoms are sp^3-hybridized with four bonded atoms, there is no clear approach for a water molecule and therefore these atoms are not explicitly probed. As with the geometry optimization performed in the previous section, the GUI provides access to the details of the QM calculation, although we will use the default values.

2 Use the provided file dialogs to specify the structural input, as well as the folder to contain the charge optimization.
PSF: 1-sysPrep/ETOH.0.psf
PDB: 2-geomOpt/ETOH.1pdb
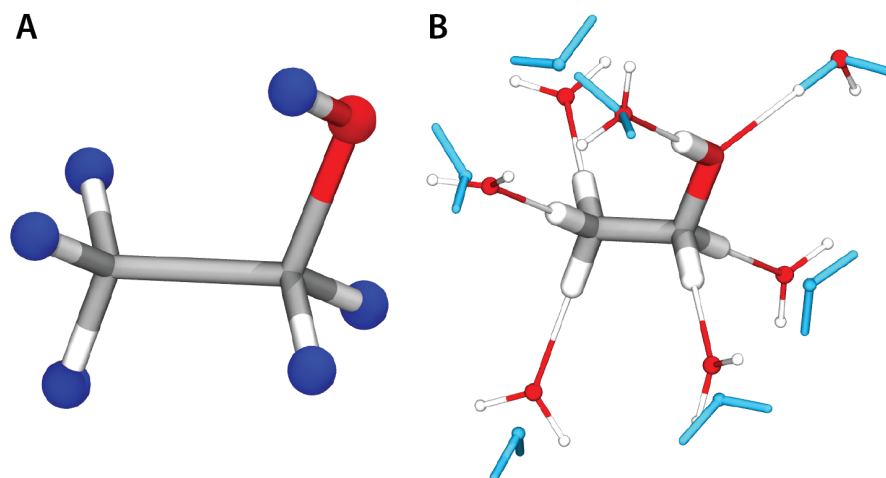Output Path: 3-chargeOpt.

Figure 7: Donor (blue) and acceptor (red) assignments for water interaction sites in ethanol (A). Initial (colored by element) and QM-optimized (cyan) positions of water (B). The dummy atoms have been excluded for clarity.

3 Set the basename by first loading the structure via the Load PSF/PDB button, followed by the Basename from TOP button.

Water interaction sites are most easily assigned for ethanol using the autodetection algorithm. After visual inspection to check the donor/acceptor assignments, all water interactions calculations should be written to file as Gaussian input files.

4 Assign donor and acceptor atoms using the AutoDetect Indices button.

5 Visually verify the donor/acceptor assignments using the Toggle Sphere Viz. button. Check that all hydrogens are represented by blue spheres, the oxygen is a red sphere, and the carbons are excluded.

6 Write the water interaction calculations to file using the Write Gaussian Input Files button.

Behind the scenes, ffTK uses a geometric algorithm to position a water molecule with respect to each interaction site, and sets up the Gaussian 2-dimensional optimization accordingly. Once complete, the water placement can be visualized by loading the generated GAU file(s) into VMD. One or more input files can be loaded simultaneously; all water interactions are shown for ethanol in Figure 7B. Note that when visualizing the input files, each water interaction will contain a dummy atom, typically shown in pink, that is required to define the water position using a z-matrix notation. Finally, ffTK generates three additional input files for computing single point energies (`ETOH-sp-HF.gau`, `ETOH-sp-MP2.gau`, and `wat-sp-HF.gau`), which will be required during the optimization step.

7 Load the water interaction input (GAU) files into VMD via the Load GAU Files button. Select `ETOH-DON-*.gau` and `ETOH-ACC-O2.gau` files using a shift+click to select multiple files. Visually inspect the initial position of each water with respect to the interaction site on ethanol.

8 Optional: If you have access to Gaussian, run each input file. If you do not have access, use the output provided in the example files.

Since the optimization is performed at a modest level of theory (HF/6-31G*) for the small ethanol molecule (three heavy atoms), the calculation should only take a few minutes on a single core. If you do not have access to Gaussian, use the output Gaussian LOGs provided in the example files. After all of the Gaussian jobs have completed, the optimized water positions can be visualized using the Load LOG File button. Visual inspection of the final optimized position is critical to make sure each water molecule finds a low energy position within a reasonable distance from the specified interaction site. Cases in which water molecules appear to settle at large distances or even "fly away", indicate the absence of the expected minimum for the given interaction site. This can be due to improperly defined interaction type (donor vs. acceptor), secondary interactions (usually steric clashes) that destabilize the water interaction, or simply an unfavorable environment due to the local electronic structure. Observations from visual inspection can be used to troubleshoot the input settings or exclude particular target data from the optimization.

9 Load the optimization output (LOG files) generated by Gaussian into VMD via the Load LOG Files button. Select `ETOH-DON-*.log` and `ETOH-ACC-O2.log` files using a shift+click to select multiple files. Visually inspect the final water positions for each interaction site to ensure a reasonable positioning was found.

4.2   Optimizing Partial Atomic Charges

With the target data for water interactions in hand, we are prepared to setup the charge optimization from the Opt. Charges tab. The charge optimization requires a more extensive setup process than any of the previous steps. To make the GUI manageable, the input settings are divided into several collapsable sections, which we will now describe.

1 Select the Opt. Charges tab of the ffTK GUI.

The Input section provides the structural data (PSF, PDB, residue name), the parameter files required to describe the molecule, including the In-Progress PAR file, and a location to write output from the optimization. These settings should be straightforward to fill out using the provided file dialogs.

2 Use the provided file dialogs to specify the input files given below. The residue name can be manually set to "ETOH" or by clicking the Resname from TOP after loading the PSF/PDB file via the Load PSF/PDB button.
PSF: `1-sysPrep/ETOH.0.psf`
PDB: `2-geomOpt/ETOH.1.pdb`
PAR: `1-sysPrep/par_ETOH.0.par`
LOG: `3-chargeOpt/chargeOpt.0.log`

The Charge Constraints section indicates which charges should be included in the optimization (i.e., all atoms or a subset of atoms), and what constraints should be applied to their values (e.g., oxygen atoms should always have a partial atomic charge < 0.0). Atoms can also be grouped together and constrained to have identical charges; this is usually applied to atoms that are identical or very similar by symmetry. Charge constraints can be set manually, or ffTK provides the Guess button. The charge constraints box should only contain atoms that are included in the optimization, while any atoms with fixed charges should be removed.

3 Use the Guess button to define the charge constraints.

Remove the entries for all non-polar hydrogens, whose charges are fixed to $+0.09$ by CHARMM convention and were assigned during the system preparation. Entries are removed by selecting the desired entry and either clicking on the Delete button, or pressing the "delete" key.

4 Delete the entries for "H11 H12 H13" and "H21 H22" by selecting each entry and clicking the Delete button or pressing the "delete" key.

Once all of the charge groups are set, we have to compute the total charge available for distribution amongst the atoms included in the optimization. The available charge is computed by subtracting the charge for all excluded atoms from the total net charge of the molecule. ffTK provides the Calculate from PSF button to perform this calculation.

The net charge is provided by the user, while charges for any excluded atoms are extracted from the input PSF file or explicitly set from the Advanced Settings section. During system preparation, all non-polar hydrogens of ethanol were assigned a charge of $+0.09$ (refer back to Fig. 3), which sum to $+0.45$. Ethanol is neutral under physiological conditions, and therefore, should be assigned a net charge of 0. Based on this input, the partial charges for all atoms included in the optimization should sum to -0.45, labeled as the "Optimized Sum". After computing the optimized sum, the resulting constraints section should match Figure 8.

5 Set the Net Charge to 0 and click the Calculate from PSF to compute the optimized sum.

Figure 8: Charge constraints are applied to each atom during the optimization.

Manipulating Fixed Charges During Charge Optimization.
When providing a partial atomic charge during the system prepara-
tion stage (e.g., non-polar hydrogens were set to $+0.09$), ffTK will
extract this value and use it as the "initial charge" for each entry.
Although we have removed all non-polar hydrogens from the opti-
mization, their partial atomic charges will still be accounted for when
computing water interaction profiles during the course of optimiza-
tion. For atoms that are not included in the optimization, atomic
charge is set based on what is provided in the PSF file, or it can be
explicitly assigned via the Override ReChargeFromPSF option found
in the Advanced Settings section. Note that the sum of all charges
being optimized plus all fixed charges (i.e., the net charge) should
equal an integer charge to avoid non-physical results. Recomputing
the optimized sum after making any modifications to the Charge
Constraints or Override ReChargeFromPSF options will ensure an
accurate Optimized Sum is used.

The following section, labeled QM Target Data, organizes the target data
computed in the section 4.1. The upper section provides the location to each
to the single point calculations, while the lower section contains all of the water
interaction data. This latter section requires additional information pertaining
to each data file. The LOG File column contains the path to the file itself, the
Atom Name column should contain the name of the atom used as the interaction
site, and the Weight column tunes how heavily that particular interaction should
be weighted by the optimizer.

6 Use the file dialogs to select the Gaussian LOG files from the single point
   calculations, as provided below.
   Cmpd LOG (HF): `3-chargeOpt/ETOH-sp-HF.log`
   Cmpd LOG (MP2): `3-chargeOpt/ETOH-sp-MP2.log`
   Water LOG: `3-chargeOpt/wat-sp.log`

7 Use the Add button to select the Gaussian LOG files (`ETOH-DON-*.log`
   and `ETOH-ACC-O2.log`) from water interaction profiles. The atom name
   is automatically parsed from the filename; leave the weight set to 1.0.

As its name suggests, the Advanced Settings section allows the user to modify advanced features of the optimization, which are described in more detail on the documentation website (`http://www.ks.uiuc.edu/Research/vmd/plugins/fftk`). For the purposes of this tutorial use the default settings.

When all of the above input is complete, the optimization is launched by clicking the Run Optimization button at the bottom of the tab. On startup, ffTK performs a brief check of the input data. If an error is detected, the routine will exit and a popup message will specify the error encountered; simply correct the errors and try again. If valid input has been provided for all options, the optimization will proceed and the "Status" label will update with the number of iterations that have been performed. While the optimization is running, the GUI will become unresponsive to user interaction.

8 Run the optimization using the Run Optimization button.

Upon completing the optimization, the GUI will become interactive again, and the Results box will be populated with the final charges for each charge group. If there were any previous results, these will be shifted to the Prev. Charge column in order to allow easy comparison between the current results and the preceding run. It is relatively easy to perform iterative optimizations by using the Set As Initial button, which replaces the initial input charges with the current final charges. When running successive optimizations, it is suggested that the user change the filename provided as the Output LOG to prevent overwriting the ffTK log file generated during the course of the optimization, preserving a record of each step.

9 Run a second iteration of the optimization. Use the output of the first iteration as the input for the next iteraction by clicking the Set As Initial button. In the Input section, change the output LOG filename from `*.0.log` to `*.1.log`.

Table 1 shows the results from each iteration of the optimization and the charges for ethanol taken from the standard CGenFF distribution. Note that the optimization algorithm is non-deterministic, and your values may differ slightly from those presented here. Neglible differences between the first and second iterations indicate that the optimizer converged rapidly to the final charge assignments. Furthermore, the final charges compare favorably to charges provided in the CGenFF parameter set. Finally, the PSF file should be updated with the optimized charges using the utility provided at the bottom of the Results section.

10 Using the provided file dialog, set the output filename to `3-chargeOpt/ETOH.1.psf`, and write the updated charges to file via the Write button.

Table 1: Charge Optimization Results

Atom Name	1st Iteration	2nd Iteration	CGenFF
C1	-0.235	-0.235	-0.27
C2	0.034	0.033	0.05
O2	-0.665	-0.665	-0.65
HO2	0.416	0.417	0.42

5 Bond and Angle Optimization

The optimization of bond and angle parameters is one of the most difficult aspects of parameterization, largely due to the technical challenges of generating appropriate target data and fitting parameters to yield a reasonable minimum energy geometry while capturing the surrounding potential energy surface. To overcome these challenges, ffTK employs a novel method leveraging QM calculations of the Hessian (the second derivative of potential energy) to reconstruct the potential energy surface associated with distortions along bonds and angles. The details of the method are not be described here; however, interested users are referred to the methods section of the paper describing ffTK[7]. Similar to the preceding charge optimization, users must first generate the QM-computed target data, followed by performing a multidimensional optimization.

5.1 Characterizing Distortions in Molecular Geometry

The target data for fitting bonds and angles involves computing the Hessian at MP2/6-31G* level of theory, in compliance with CGenFF. This calculation requires data computed as part of the geometry optimization performed in section 3. During the course of this previous calculation, Gaussian output a special file called a checkpoint file (CHK), which is required as input for the current Hessian calculation. The setup for computing the Hessian is relatively straightforward from the Calc. Bonded tab.

1 Select the Calc. Bonded tab of the ffTK GUI.

2 Use the provided file dialogs to specify the following inputs:
 PSF: `3-chargeOpt/ETOH.1.psf`
 PDB: `2-geomOpt/ETOH.1.pdb`
 CHK: `2-geomOpt/ETOH.geomOpt.chk`
 Output GAU: `4-baOpt/ETOH.hess.gau`

The bottom portion of the tab allows advanced users access to the details of the QM calculation. The system resources required to compute the Hessian for ethanol is minimal; therefore, all default values are reasonable. Upon clicking the Write Gaussian Input File button, ffTK will setup the calculation and generate a Gaussian input file and checkpoint file; both files are required by Gaussian to perform the calculation.

3 Use the Write Gaussian Input File button to generate the input required to perform the Hessian calculation.

4 Optional: Run the calculation where you have access to Gaussian. If you do not have access, use the output provided in the example files.

> **Resource Requirements of a Hessian Calculation.**   The Hessian calculation typically requires a large amount of memory, depending on system size. If the calculation exceeds the available memory and begins to use a disk swap, the wall clock time required to complete the calculation becomes prohibitively long. If calculations are taking longer than expected, this is the best option to troubleshoot.

5.2   Optimizing Bond and Angle Parameters

Bond and angle optimizations are performed from the tab labeled Opt. Bonded. The optimization setup follows a pattern similar to that described for the charge optimization, in which the input is spread over several expandable sections within the tab and iterative runs require setting the results from the previous run as the input for the next run.

1 Select the Opt. Bonded tab of the ffTK GUI.

The Input section provides file dialog buttons to locate the necessary structural input, parameter files, and the target data (Hessian LOG file). During the course of the optimzation ffTK will call on the molecular dynamics program NAMD to perform short geometry optimization calculations in the background. Accordingly, NAMD must be installed on your local system, and the location of the executable binary should be provided in the Input. Finally, a location and filename for writing the output from the optimization is required.

2 Use the provided file dialogs to specify the following input:
   PSF: `3-chargeOpt/ETOH.1.psf`
   PDB: `2-geomOpt/ETOH.1.pdb`
   Hess Log: `4-baOpt/ETOH.hess.log`
   In-Progress PAR: `1-sysPrep/par_ETOH.0.par`
   Additional Associated Parameter Files: none

3 Specify the location for the NAMD executable (NAMD Bin), `namd2` (or `namd2.exe` on Windows). On Unix-based systems (Unix, linux, or MacOSX), if NAMD is included in the PATH environment variable, the full path to the filename is not required; and `namd2` (the default value) will suffice.

4 Use the SaveAs file dialog to provide a filename for writing output from the optimization, `4-baOpt/baOpt.0.log`.

The Parameters to Optimize section is used to define and initialize the bond and angles terms under consideration during the optimization. There are three ways to populate this section using the provided buttons to the right of the data box. The Guess button constructs the input based on the bond and angle terms provided in the In-Progress PAR file, but sets the initial values (Force Constant and b_0/θ) based on data extracted from the raw target data. The Import button reads in the data provided in the In-Progress PAR file and retains any values

provided therein. Finally, users can add terms in manually via the Add button, which creates a blank entry to modify as desired. Since the In-Progress PAR file was initialized to 0.0 for all values, the Guess button is the most appropriate for our example. Figure 9 shows the populated data box.

5 Use the Guess button to populate the Parameters to Optimize data box.

Figure 9: Initial values for bond and angle parameters set from the Guess button.

The Advanced Settings section provides user access to detailed variables that control the optimization. Here we will make a few changes, namely modifying the Geom. Weight value to 2.0 and to reduce the Angles – Eq. Deviation tolerance from 10.0 to 5.0. The first modification increases the weight associated with how well the MM-optimized geometry matches the QM-optimized geometry. The Eq. Deviation value sets the threshold below which deviations no longer contribute to the objective function. Overly tight Eq. Deviation thresholds tend to yield large force constants, such that reasonable thresholds are critical to balancing the interplay between obtaining suitably close optimized geometries while reproducing the surrounding potential energy surface.

6 In the Advanced Settings section, set the Geom. Weight to 2.0 and Angles – Eq. Deviation to 5.0.

The optimization is launched by clicking on the Run Optimization button at the bottom of the tab. Similar to the charge optimization, the current routine will perform a brief check of the input data and return with an error message detailing any inappropriate input, if detected. While running, the GUI will become unresponsive until the optimization is complete. The calculations required at each step of the optimization are more rigorous than for charge optimization, and therefore, take longer to complete (up to 15 minutes depending on the processor and disk IO speed). To perform iterative rounds of optimization, use the Set As Initial button within the Results section to update the input until the Current Final Obj. Value converges to a minimum (usually 2-3 runs).

7 Perform iterations of the optimization until the "Current Final Obj. Value"
begins to increase. Launch each optimization via the Run Optimization
button. Upon completion set the output from the current optimization
as the input for the subsequent iteration by clicking on the Set as Initial.
Prior to each run, change the Output Log by increasing the incrementer in
the filename, i.e., `baOpt.0.log` is followed by `baOpt.1.log`, to preserve
the data output for each run.

Change the Output Log for Iterative Optimizations.   Remember
to change the filename provided in the Input → Output LOG entry
box with each successive run to prevent overwriting data. This is
particularly important if intend to perform iterative optimizations
while retaining the intermediate data.

5.3   Updating the In-Progress Parameter File

During the course of each parameterization step, ffTK outputs detailed infor-
mation describing the path of each optimization to a user-specified log file.
Contained within these files are the final values for the parameters under con-
sideration, which are to replace the initial values found in the In-Progress PAR
file. The BuildPar tab contains a tool to perform updates on parameter values,
located in the Update Parameter File with Optimized Parameters section. Within
this section, the user provides the parameter file to be updated (Input Parameter
File), the ffTK log file containing the optimization data (Optimization LOG File),
and a filename where a new parameter file will be written (Output Parameter
File). The Write Updated Parameter File button will then update any relevant
parameters and write a new parameter file.

1 Navigate to the BuildPar tab and expand the section labeled Update Pa-
rameter File with Optimized Parameters.

2 Update the In-Progress PAR file with the ffTK log file from the bond
and angle optimization which yielded the lowest final objective function,
writing the updated parameters to a new PAR file. Click the Write Updated
Parameter File to perform the update. Example input is provided below.

Input Parameter File: `1-sysPrep/par_ETOH.0.par`
Optimization LOG File: `4-baOpt/baOpt.1.log`
Output Parameter File: `4-baOpt/par_ETOH.1.par`

6 Dihedral Optimization

The final step in the parameterization workflow is the optimization of dihedral terms describing torsional potentials within the molecule. This task follows the familiar pattern of first computing the target data from high level QM calculations, followed by a multidimensional optimization to fit MM dihedral parameters. Unlike bonds and angles, which are described using a simple harmonic potential, a complete description of a particular dihedral can require anywhere from one to five cosine terms (referred to as the multiplicity), each containing three parameters: periodicity (n), force constant (k), and phase shift angle (δ). While the general shape of the potential, described by the multiplicity and periodicity, must be provided by the user, ffTK will optimize the values for the force constant and phase shift angles.

Computing the potential energy for a particular torsion requires summing the contributions for all dihedral angles centered on the specified torsional atoms. Teasing out the individual contributions of each dihedral to the potential energy of a torsion, however, can be difficult or impossible. This scenario highlights how dihedral fitting is an under-determined problem (many possible solutions in parameter space that satisfy the given target data) with implications on the transferability of the resulting parameters. Coupled systems, such as rings or congested scaffolds, often prevent the isolation of specific torsions, and further exacerbate the complexity of dihedral fitting. In an attempt to mitigate many of these complicating factors, ffTK employs a holistic approach in which all dihedrals are simultaneously fit to a torsional potential energy surface (PES). This approach, however, is an incomplete solution; it remains important to keep these challenges in mind during both the parameterization process and any subsequent simulations that utilize the resulting parameters.

6.1 Characterizing the Torsional Energy Landscape

The target data for fitting dihedrals requires performing a relaxed PES scan. This calculation consists of stepping through a range of angles for a given dihedral. At each step, the dihedral is fixed and a geometry optimization is performed for the rest of the molecule. The Scan Torsions tab provides a GUI to facilitate identifying which dihedrals to scan and adjusting the parameters that control the details of the scan.

1 Select the Scan Torsions tab of the ffTK GUI.

The Input/Output section contains file dialogs to specify required inputs for structural files and to select a folder where the Gaussian input files will be written. Additionally, the Basename input is used as a prefix when naming the Gaussian input file(s), and is usually set to the resname of the molecule.

2 Use the provided file dialogs and function buttons to specfiy the required input files given below:
PSF: `3-chargeOpt/ETOH.1.psf`
PDB: `2-geomOpt/ETOH.1.pdb`
Output Path: `5-dihOpt`

3 Load the structure into VMD, then set the basename input via the Load PSF/PDB and Basename from TOP buttons, sequentially.

The Dihedrals to Scan section provides two methods for specifying the details of each dihedral scan. The simplest method is to use the Read from PAR button. Using this function, a file dialog requests the In-Progress PAR file, from which ffTK will read in the dihedral section, identify all examples of each dihedral within the input molecule, remove any dihedrals that terminate in a hydrogen, filter out any redundant dihedrals with respect to a shared bond torsion, and add each remaining entry into the scan box portion of the GUI. Alternatively, dihedrals can be added manually via the Add button on the right side of the GUI, which will add a generic entry to the provided scan box. Entries can be edited by selecting the desired entry and modifying the data via the Edit Entry fields below. Clicking on the checkmark box updates the data within the scan box, while the "X" button will revert any changes to the edit fields. When entering dihedrals manually, it is convenient to turn on labels for atom indices via the Toggle Atom Labels provided in Input/Output section above.

> **(!) Filtering Out Dihedrals that Terminate in Hydrogen from Explicit Dihedral Scanning.** It is notable that dihedrals terminating in hydrogens are not typically scanned explicitly, and are excluded by default when using the Read from PAR function. This behavior sometimes excludes dihedrals that should be scanned, e.g., dihedrals terminating in polar hydrogens, which can be added manually.

When considering our example molecule, ethanol, attempting to use the Read from PAR will fail to load any dihedrals to scan. Ethanol contains only three heavy atoms, and all dihedrals terminate in hydrogens. As described above, entries terminating hydrogens are filtered out by the automated routine; accordingly, scans for the C-C and C-O torsions should be setup manually. When complete, the scan box should look similar to Figure 10.

4 Turn on atom labels via the Toggle Atom Labels.

5 Use the Add button to create a generic scan entry and select this entry from the scan box. Change the properties of the scan the properties given below, and then re-select the entry to update the visualization of the dihedral in the VMD window.
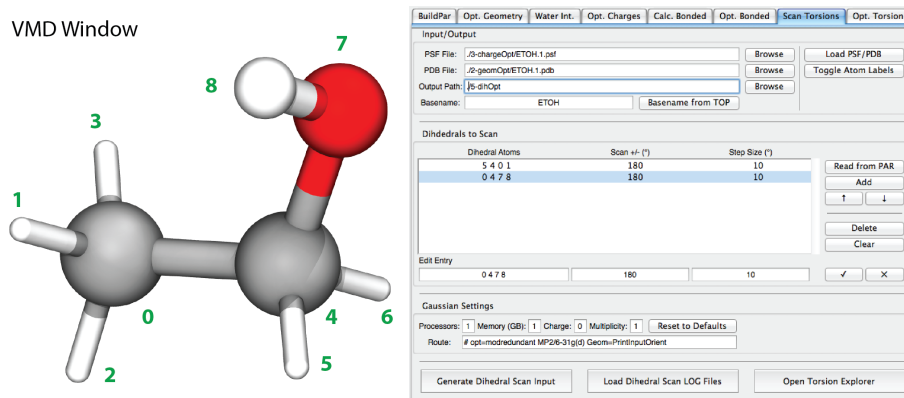Dihedral Atoms: 5 4 0 1
Scan +/-: 180
Step Size: 10

Figure 10: Input required to setup the torsion scan using the ffTK GUI. The atoms associated with the selected entry selected are shown in the VMD window using the CPK represenation.

6 Repeat the preceding steps to add a scan for the C-O torsion (indices: 0 4 7 8) using the same scan range and step size.

The last section of the tab allows advanced users access to the details of the QM calculation. The relaxed PES scan performed at the default MP2/6-31G* level of theory can require significant resources and compute time. Adjust the number of processors and available memory as appropriate for your local resources, leaving all other settings as the default values. Upon clicking the Generate Dihedral Scan Input button ffTK will write out the necessary Gaussian input files for each entry in the scan box. Note that ffTK employs a bidirectional scanning approach in which two scans are performed for each dihedral: one in the (+)-direction and the second in the (-)-direction.

7 Write out the input files for the dihedral scan via the Generate Dihedral Scan Input button.

8 Optional: Run the dihedral scanning calculations where you have access to Gaussian. If you do not have access, use the output provided in the example files.

Upon completion of all Gaussian scans, the resulting conformations can be loaded into VMD as a trajectory using the Load Dihedral Scan LOG Files button for visual inspection. An advanced analysis module, named Torsion Explorer, can also be used to link the details of the torsional PES to the visualization of the scanned conformations in VMD. While detailed analysis is not required for a simple molecule such as ethanol, the Torsion Explorer module can be extremely useful for troubleshooting regions of the PES that fail to optimize satisfactorily.

6.2   Optimizing Dihedral Parameters

The optimization of dihedrals is performed from the Opt. Torsions tab, the contents of which are organized into sections consistent with the previous optimizations. The optimization scheme for dihedrals, however, is broken into two separate, but related, steps: an initial optimization, followed by iterative refinements. The reasons for this are mostly technical, but also to simplify the input required by the user, described in detail below.

  1 Select the Opt. Torsions tab of the ffTK GUI.

The Input section follows the same patterns as seen in previous sections, requiring structural files (PSF/PDB), parameter files (PAR), the NAMD2 binary executable (`namd2` or `namd2.exe`), and an output log file (LOG). File dialogs are provided for all required input.

  2 Use the provided file dialogs to specify the following inputs, and to locate the NAMD executable.
  PSF: `3-chargeOpt/ETOH.1.psf`
  PDB: `2-geomOpt/ETOH.1.pdb`
  PAR: `4-baOpt/par_ETOH.1.par`
  Output LOG: `5-dihOpt/DihOpt.log`

The QM Target Data section is even more straightforward, requiring only the locations to the Gaussian LOG files for all scans performed in the preceding section.

  3 Use the provided file dialog to add all LOG files from the dihedral scans performed in the preceding section (i.e., `5-dihOpt/ETOH.scan.*.log`).

The Dihedral Parameter Settings section is used to define the dihedrals that will be optimized. Input can be provided via two methods. The easiest method is to read the dihedral terms directly from the In-Progress PAR file via the Read from PAR button. This function will read in all terms, including multiplicities, that are present in the provided parameter file (Fig. 11). Alternatively, dihedrals can be added manually via the Add button. Entries can be edited by selecting the entry of interested, modifying the data in the edit entry fields, and clicking the checkmark button to update the entry. A complete description of a dihedral often requires multiple terms bearing the appropriate periodicity (n) and phase shifts (δ) in addition to a force constant (k). At this stage it is not critical to provide all of the terms correctly, as multiplicities can be added, removed, and modified during the refinement stage. It is required, however, to provide at least one term for every dihedral to be optimized.

  4 Use the Read from PAR button and select the In-Progress PAR file, `4-baOpt/par_ETOH.1.par`, to add all dihedrals to the optimization. The specific details of the Force Constant, Periodicity, Phase Shift, and Lock Phase can be left as the default values.

Figure 11: Initial dihedral parameters when read from the In-Progress PAR file.

The Advanced Settings section provides access to the details of the optimization routine, including thresholds, the optimization algorithm, and output options. The default values are sufficient for our example.

Run the initial optimization by clicking on the Run Optimization at the bottom of the tab. This first pass through the optimizer may take several minutes to complete depending on how many steps were taken during the PES scans, during which the GUI will be unresponsive. Upon completing the initial optimization, the GUI will return interactivity and the final data is loaded into Visualize Results section. The Visualize Results section provides a plotting utility to compare the MM PES from multiple optimization/refinements against the QM target data (QME), as shown in Figure 12. The QME is the target PES profile and the MMEi is the MM PES computed when all dihedrals under consideration are excluded (i.e., $k = 0$), providing an idea of where the fit is starting from.

The results of the initial optimization show a poor fit to the QME, clearly observed from Figure 12A. Further refinement of the dihedral parameters is performed from the Refine section. First, the output from the initial optimization must be copied into the refinements section by selecting the desired data set (i.e., "orig") from the results box and clicking the Set As Refit Input button.

5 Select the entry corresponding to the initial optimization from the data set box, marked as "orig", and click the Set As Refit Input button.

Expanding the Refine section now shows all of the dihedral entries originally input into the Dihedral Parameter Settings section; however, the force constant and phase shift contain the data returned from the initial optimization. The most effective strategy for refining dihedrals is to alter the multiplicities and periodicity settings to get the shape of the MM PES to match the QME profile. Once the shape is correct, the optimizer becomes extremely effective at fitting the maxima and minima by altering the force constant. In order to get the shape of the PES correct for ethanol, we will need to add some multiplicities and modify several periodicities. For very simple molecular geometries such terms can be determined by chemical intuition; however, it is also a good strategy to find analogous molecules (or functional groups) in the standard CGenFF
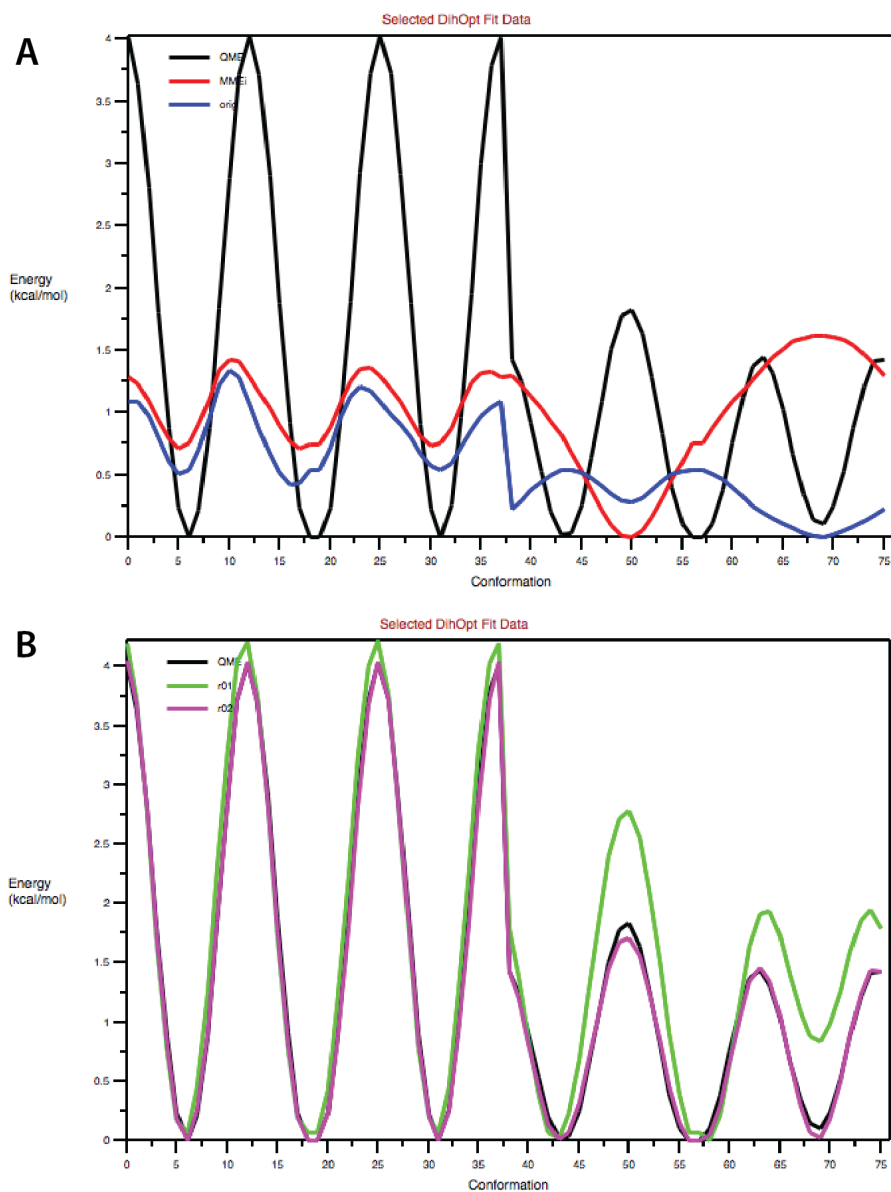
Figure 12:   The Visualize Results section provides a utility to plot the PES profile for the optimization results. The profile from the initial optimization (blue) yielded a poor fit to the target data (black) with minimal improvement over the MMEi (red) profile (Panel A). Subsequent refinements using modified dihedral terms yield substantially better profiles after the 1st (green) and 2nd (purple) iterations (Panel B).

distribution to obtain more information regarding dihedral terms for specific connectivities.
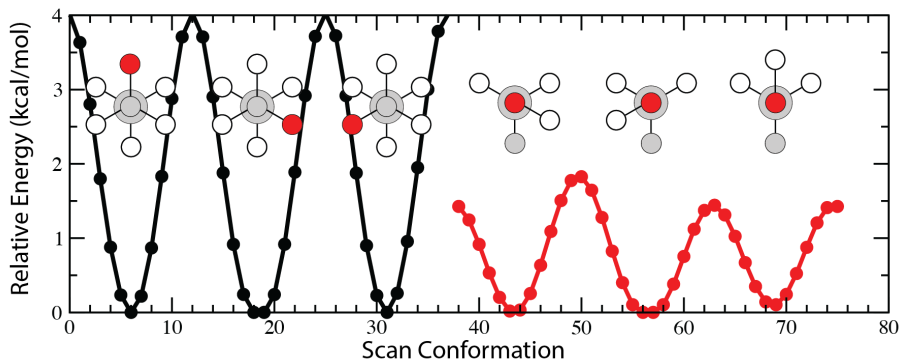

Figure 13: QM-computed torsional profile of ethanol along with Newman projections for each minimum energy conformation.

Table 2: Dihedral Terms Describing Ethanol

Dihedral Definition	Periodicity (n)	Phase Shift (θ)
CG331-CG321-OG311-HGP1	1	0.0
	2	0.0
	3	180.0
HGA2-CG321-OG311-HGP1	3	0.0
HGA3-CG331-CG321-HGA2	3	0.0
HGA3-CG333-CG321-OG311	3	0.0

The QME profile represents the energy profile for the two different torsional scans shown in Figure 13; the first set of data corresponds to the C-C scan (black), and the second to the C-O scan (red). The C-C torsional profile is characteristic of a highly symmetric bond between two sp^3-hybridized atoms (e.g., C-C), which contain three minima corresponding to the Newman projections in Figure 13. For ethanol, these three conformations are degenerate in energy and rotational barriers. Accordingly, the torsion is best described via a single dihedral term in which the periodicity (n) is set to 3 with no phase shift $(\delta = 0)$. Since the oxygen of the hydroxyl group only has two bonded atoms in addition to two electron pairs, the torsional profile is no longer symmetric, as can be seen from the Newman projections. As a result, the C-C-O-H dihedral requires a more complex description, given by a multiplicity of three for which the details are provided in Table 2. This description is common to aliphatic alcohols found in the standard CGenFF parameter set.

6 Update the dihedral parameters in the refine section as given in Table 2.

7 Run the refinement by clicking the Run Refitting/Refinement button in the Refine section. Caution: Do not accidentally click the Run Optimization button, which will rerun the original optimization from the input parameters.

The refinement procedure is significantly faster than the initial optimization, and the resulting data will be added to the data sets in the Visualize Results box, where it can be analyzed against both the target data and any other data sets. This process should be iterated until the MM PES is sufficiently close to the QME target data. The fit does not have to be exact; however, the position and general magnitude of the minima and maxima should be reproduced.

Visual assessment of the first refinement results show marked improvement over the initial optimization, particularly with respect to the shape of the PES (green profile in Fig. 12B). Switching the refinement algorithm from "Simulated annealing" to "downhill" changes the focus from a broader sampling approach to reaching the closest local minimum. Once the dihedral terms have been well-defined, the downhill algorithm is often quite effective in improving the magnitude of the rotational barriers and minimum energies.

8 Use the results from the first iteration to perform a second run by selecting the "r01" entry from the results box, followed by clicking the Set As Refit Input button.

9 Change the refinement algorithm to "downhill", reduce the Tol. to 0.0001, and launch the refinement.

The RMSE column of the results data now suggests significant improvements in the parameters, starting from 1.050 for the original optimzation followed reductions to 0.460 and 0.112 after the first and second refinements, respectively. These data are further supported by visually inspecting the improvements both the shape of the MM profiles and the reproduction of energy minima and maxima (Fig. 12B)–hallmarks of a successful parameterization.

Overfitting Dihedral Parameters. Adding additional multiplicities to any dihedral should, in principle, give a better fit by virtue of including additional terms to achieve a better mathematical fit. Adding too many terms or terms with inappropriate periodicities, however, often yields non-physical results due to over-fitting. Users should exercise caution when considering multiplicities during dihedral refinement.

6.3   Exporting Dihedral Optimization Results

While the initial optimization writes the details of the optimization out to file, specified in the Input section, it is notable that the refinement routine does not write any data to file. All of the data required related to the refinement is stored in the results box within the Visualize Results section. Once you are satisfied with the PES torsional profile, selecting the data set you wish to save,

followed by clicking the Write Selected to LOG button will write the data to file. The resulting LOG files can be used to reload the refinement data into ffTK for further work, via the Import From LOG, or to update the parameter file with the optimized dihedral parameters. The latter task is handled in the same manner as updating bond and angle parmeters, via the Update Parameter File with Optimized Parameters section of the BuildPar tab.

1 Write second refinement data to file by selecting the "r01" data set from the Visualize Results section, clicking the Write Selected to LOG button, and saving the data as `5-dihOpt/DihOptRefine.r01.log`.

2 Navigate to the Update Parameter File with Optimized Parameters section of the BuildPar tab. Use the file dialogs to specify the input given below, and click on the Write Updated Parameter File button.
Input Parameter File: `4-baOpt/par_ETOH.1.par`
Optimization LOG File: `5-dihOpt/DihOptRefine.r01.log`
Output Parameter File: `5-dihOpt/par_ETOH.2.par`

7   Conclusions

The steps detailed in this tutorial introduced many of the tools provided by the FORCE FIELD TOOLKIT for parameterizing small molecules. Readers derived parameters required by Eq. (1) for bonds, angles, dihedrals, and electrostatics to describe a small organic molecule, ethanol. The end product of the presented workflow is a complete parameter set required to perform MD simulations of ethanol. Although it is beyond the scope of this tutorial, examples of rigorous benchmarking studies for ethanol and several other small organic molecules were performed as part of the validation for ffTK's protocols, and demonstrated outstanding agreement both with established parameter sets (CGenFF) and experimentally measured values[7]. These data indicate that even the straightforward protocols described in the current tutorial can yield research-grade results.

For the purposes of the current tutorial, the example molecule was parameterized without taking any known parameter sets into account; however, this is rarely the best approach when considering real-world research problems. Often, the molecule of interest contains substructures with known parameters (e.g., phenyl rings, saturated alkyl chains) and only a subset of parameters lack adequate descriptions. Furthermore, the emergence of analogy-based tools, such as the CGenFF Program (formerly known as ParamChem)[10,11] or MATCH[12], can also provide an initial parameter set. These parameter sets often serve as a good initial starting point, requiring only minimal refinements or validation. Users will find that ffTK provides built-in tools to handle each of these scenarios without significant pertubations to workflows described in this tutorial. In this manner, ffTK represents a flexible tool for both those new to parameterizations and experts in force field development.

8   Acknowledgments

9 Bibliography

[1] A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, I. W. E. Reiher, B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, 102:3586–3616, 1998.

[2] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, 117:5179–5197, 1995.

[3] W.J. Jorgensen and J. Tirado-Rives. The OPLS potential functions for protein energy minimization for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.*, 110:1657–1666, 1988.

[4] C. Oostenbrink, A. Villa, A.E. Mark, and 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. Comp. Chem.*, 25: 1656–1676, 2004.

[5] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and A. D. MacKerell, Jr. CHARMM General Force Field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comp. Chem.*, 31(4):671–690, 2010.

[6] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case. Development and testing of a general amber force field. *J. Comp. Chem.*, 25:1157–1174, 2004.

[7] C. G. Mayne, J. Saam, K. Schulten, E. Tajkhorshid, and J.C. Gumbart. Rapid parameterization of small molecules using the Force Field Toolkit. *J. Comp. Chem.*, 34:2757–2770, 2013.

[8] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene,

J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. *Gaussian 09 Revision B.01*, 2009. Gaussian Inc. Wallingford CT 2009.

 [9] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. 79(2):926–935, 1983.

[10] K. Vanommeslaeghe and A. D. MacKerell, Jr. Automation of the CHARMM General Force Field (CGenFF) I: Bond perception and atom typing. 52(12):3144–3154, 2012.

[11] K. Vanommeslaeghe, E. P. Raman, and A. D. MacKerell, Jr. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of bonded parameters and partial atomic charges. 52(12):3155–3168, 2012.

[12] J. D. Yesselman, D. J. Price, J. L. Knight, and C. L. Brooks, III. MATCH: An atom–typing toolset for molecular mechanics force fields. *J. Comp. Chem.*, 33:189–202, 2012.