Published in final edited form as:
J Comput Chem. 2013 December 15; 34(32): . doi:10.1002/jcc.23422.

Rapid parameterization of small molecules using the Force Field Toolkit

Christopher G. Mayne[†,¶], **Jan Saam**[†], **Klaus Schulten**[†,§], **Emad Tajkhorshid**[†,‖,*], and **James C. Gumbart**[‡,*]

[†]Beckman Institute, University of Illinois at Urbana-Champaign, Urbana IL 61801

[‡]School of Physics, Georgia Institute of Technology, Atlanta, GA (30332)

[¶]Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana IL 61801

[§]Department of Physics, University of Illinois at Urbana-Champaign, Urbana IL 61801

[‖]Department of Biochemistry, University of Illinois at Urbana-Champaign, Urbana IL 61801

Abstract

The inability to rapidly generate accurate and robust parameters for novel chemical matter continues to severely limit the application of molecular dynamics (MD) simulations to many biological systems of interest, especially in fields such as drug discovery. Although the release of generalized versions of common classical force fields, e.g., GAFF and CGenFF, have posited guidelines for parameterization of small molecules, many technical challenges remain that have hampered their wide-scale extension. The Force Field Toolkit (ffTK), described herein, minimizes common barriers to ligand parameterization through algorithm and method development, automation of tedious and error-prone tasks, and graphical user interface design. Distributed as a VMD plugin, ffTK facilitates the traversal of a clear and organized workflow resulting in a complete set of CHARMM-compatible parameters. A variety of tools are provided to generate quantum mechanical target data, set up multidimensional optimization routines, and analyze parameter performance. Parameters developed for a small test set of molecules using ffTK were comparable to existing CGenFF parameters in their ability to reproduce experimentally measured values for pure-solvent properties (<15% error from experiment) and free energy of solvation (±0.5 kcal/mol from experiment).

Introduction

The advancement of molecular dynamics (MD) simulations as a method for probing biological systems requires overcoming a number of key barriers, including limitations in time scale, system size, and the accurate representation of the underlying molecular system. While the first two rely primarily on advances in hardware and algorithms, the last, accuracy of the molecular description, requires assiduous development of better force fields that adequately describe important interactions within the simulation system. The inclusion of CMAP potentials[1,2] and polarizability[3–8] are two examples illustrating the current course of force field development. An often overlooked, but fundamental and long-persisting limitation, however, is the complexity of developing missing force field parameters for novel chemical species, such as modified amino acids or small molecule ligands. These chemical entities are frequently critical components within the biological system of interest,

[*]To whom correspondence should be addressed, emad@life.illinois.edu; gumbart@physics.gatech.edu.

yet the inability to easily and accurately parameterize them greatly impedes the utility of MD technologies across many fields, including most notably drug discovery.[9,10]

Many different empirical force fields (e.g., AMBER, CHARMM, GROMOS, OPLS) have been developed for use in MD simulations.[11] Although each particular force field follows a specific philosophy based on a set of theoretical underpinnings and an implementation strategy, they share a common reliance on the concept of transferability, in which a single set of parameters for a given atom type accurately describes its behavior in a wide range of chemical (connectivity) and spatial (conformation) contexts. Biopolymers (e.g., proteins, DNA, and carbohydrates) lend themselves quite well to this concept due to their modular nature and composition from a relatively small set of independent building blocks (amino acids, nucleic acids, sugars). While developing accurate, yet transferable, parameters for these building blocks requires significant effort initially, the payoff is tremendous, and allows for the simulation of biological systems covering a wide range of compositions, scales, and functions from a modestly sized parameter set.

Small molecules, in contrast to most biomolecules, represent a vastly increased diversity of structures and chemistries. For comparison, it is estimated that the human genome encodes for ~25,000 proteins,[12] while estimates of chemical space range from 10^{18} to 10^{200}.[13] It is unreasonable to suppose a single parameter set can adequately describe such a large number of compounds. One approach to addressing this “small molecule problem” is by developing a limited set of building blocks that cover a specific class or family of molecules. This has been the underlying principle of the General Amber Force Field (GAFF)[14] and CHARMM General Force Field (CGenFF),[15] which target only “druglike molecules in a biological environment.”[15] Pharmaceutically relevant compounds (drugs, chemical probes, etc.), however, tend to be comprised of linked or fused aromatic (frequently heteroaromatic) scaffolds that are highly decorated with a great variety of engineered functional groups, with the goal being to improve potency, selectivity, ADMET properties, and to avoid intellectual property liabilities. The exotic nature of many substituents combined with the complexities of charge delocalization and conformational dynamics run counter to the principles of transferability, thereby reducing the applicability of the building-block approach.

Despite the diverse challenges faced, several tools have been developed to assign missing parameters directly from analogy to exisiting ones. These tools rely on databases of molecules and molecular fragments of previously parameterized compounds for a given force field. Examples of such tools are the ParamChem[16,17] and MATCH[18] web servers for the CHARMM force field, and the Automated Topology Builder (ATB)[19] and PRODRG[20] web servers for the GROMOS force field. The Swiss-Param[21] web server, in contrast, assigns vdW terms by analogy to existing CHARMM atom types while all other parameters (charges, bonds, angles, dihedrals, impropers) are taken by analogy from Merck Molecular Force Field (MMFF),[22] and translated into the CHARMM format.

Significantly fewer tools are available for the development of parameters directly from first principles. The most prominent of these is Antechamber,[23] which is used to generate parameters for the AMBER and associated GAFF force fields. Paratool, released as a plugin within VMD, provided limited ability to derive CHARMM parameters from quantum mechanical (QM) calculations; however, the tool never went beyond the development stage. The release of CGenFF, along with a loosely codified set of procedures for parameterization,[15] made possible the development of a comprehensive parameterization tool capable of yielding a *complete* set of CHARMM/CGenFF-compatible parameters. To our knowledge, however, no such tools have been described in the literature. To the contrary, recent software solutions (e.g., ParamChem, MATCH) have focussed on parameter assignment based on analogy only.

We recently surveyed the literature for research papers that cite the original CGenFF publication[15] (Web of Science, n=142) to better understand how the force field and associated parameterization philosophies are currently implemented by the MD community. In over three years since the publication of CGenFF, fewer than 10% of references, excluding subsequent work from the original authors, describe the development of parameters from *ab initio* calculations in accordance with the published methodologies. Furthermore, this subset of publications is limited to deriving partial atomic charges or, more frequently, dihedral parameters. We are unaware of any reports other than those associated with the MacKerell laboratory that derive bond or angle parameters according to prescribed methods. An increasing number of publications report obtaining parameters from the ParamChem webserver.[16,17] While this is an excellent resource for obtaining initial parameters based on analogy, very few of these publications discuss the penalty scores that accompany the ParamChem output, verify parameter performance through additional calculations, or refine the provided parameters within the novel chemical context. Most disconcerting, however, is the occurrence of published reports that assign or derive parameters using methodologies that are incompatible with the accepted best practices for CHARMM force fields, e.g., restrained electrostatic potential (RESP) fitting or extracting charges directly from quantum mechanical calculations.

The preceding observations strongly suggest that the published methodologies for parameterization, while enabling, are currently intractable for much of the MD community. The Force Field Toolkit (ffTK), described herein, serves as a framework to resolve the disconnect between the theoretical and practical facets of parameterization by organizing parameterization tasks into a clearly defined modular workflow (Figure 1) supported by a collection of optimization methods and algorithms. The included graphical user interface (GUI) steps through the parameterization workflow, providing tools to automate many tedious and error-prone tasks, without obscuring the underlying processes, and to assess parameter quality during development. ffTK, therefore, represents a powerful utility for developing parameters *ab initio*, refining existing parameters, and quantitive assessment of parameter performance, both for individual terms as well as a collective set of parameters.

Parameter Optimization Components of ffTK

The literature contains several publications discussing guidelines for developing CHARMM-compatible parameters,[24–29] culminating in the workflow outlined by Vanommeslaeghe et al.;[15] however, the onus of actuation remains entirely on the user. Many significant barriers remain, namely, obtaining the QM target data, both generating and parsing the data, iterating through test parameters, and scoring the fit of the resulting MM properties to the target data. The result of ffTK is the minimization of these barriers through careful GUI design, helpful support functions, and employing the optimization methods coupled to tailored scoring algorithms (referred to as “objective functions”).

A large number of support functions in ffTK serve the important task of automating tedious portions of the parameterization. A particularly useful subset of these functions that aid users during the “System Preparation” requires some clarification. Entry into the parameterization workflow requires that the user provides a PSF/PDB file pair containing molecular information such as pre-assigned atom types, atom names, and an initial molecular geometry. Using this molecular information, ffTK facilitates tasks such as identifying missing parameters, generating an initialized in-progress parameter file, and geometry optimization. ffTK also does not currently support the development of CHARMM non-bonded (Lennard-Jones) parameters, nor does it provide automated atom typing. However, it does provide a utility that parses parameter values and comment information from existing topology/parameter files to aid the user in selecting an appropriate LJ value and update the in-progress parameter file accordingly. Rather than relying on one s own intuition, the automated atom-typing functionality provided by the ParamChem webserver has also proven to be extremely accurate for CGenFF atom types.[16] Further details of ffTK s other support functions will not be discussed here (see the ffTK documentation website: http://www.ks.uiuc.edu/Research/vmd/ffTK); instead, the discussion will focus on only the salient features with respect to the core goal of computing CHARMM-compatible parameters that reproduce the target data, specifically the Charges, Bonds & Angles, and Torsions/Dihedrals parameterization stages of Figure 1, with some additional attention to assessing parameter quality.

The following sections cover the aforementioned stages of the parameterization, and are presented in an order that corresponds to the workflow given in Figure 1. The Charge Optimization section briefly reviews the derivation of partial atomic charges from water-interaction profiles, a distinctive feature of the CHARMM force field.[15,27] The subsequent section describes a new approach to deriving the bonded parameters, which focuses on fitting potential energy surfaces of small perturbations about the optimized structure. Finally, dihedral fitting is addressed, for which ffTK employs a direct adaptation of the method developed by Guvench and MacKerell.[29]

Optimization of partial atomic charges from water-interaction profiles

A key distinguishing aspect of atomic force fields is how they derive the partial charges on atoms. For example, in the AMBER force field, charges result from fitting to the electrostatic potential surrounding the molecule,[30] and OPLS charges are derived directly by fitting experimentally measured condensed phase properties.[31],[32] In contrast, the CHARMM force field emphasizes reproducing QM interactions with a TIP3P[33] water molecule. Following the CHARMM convention, in ffTK, each water-accessible atom of the compound is assigned to a list of potential hydrogen bond donors, acceptors, or both. For each atom (interaction site) in these lists, a complex between the geometry-optimized target compound and a water molecule is automatically constructed in which the water is ideally oriented for hydrogen bonding (Figure 2A). The initial position of the water is defined by the molecular geometry of the interaction site to minimize steric repulsion between the water molecule and all neighboring atoms covalently bound to the interaction site (Figure 2B). For each target-water complex generated, a corresponding Gaussian[34] input file is written. Because hydrogen bonds are almost exclusively assumed to be linear in fixed-charge force fields, only two free parameters remain, namely the distance between the water molecule and the target atom and the rotation angle of the water about the line connecting them. These two parameters are optimized quantum mechanically, with all other degrees of freedom constrained at the HF/6-31G(d) level of theory to maintain consistency with the CHARMM force field.[15,27] ffTK also automatically generates Gaussian input files to compute single-point energies for the compound and water molecules separately, which are required during the optimization.

After the Gaussian calculations are run (notably outside of VMD), the resulting output can be imported back into ffTK. The Gaussian output is processed to extract the optimized distances between each atom and its associated water molecule, as well as their optimal interaction energies; the latter is taken as the difference between the total energy of the optimized complex and the independent energies of the two molecules, loaded separately. To better approximate the bulk-phase, for uncharged, polar target compounds the QM-optimized distances are shifted by an offset of −0.2Å and the interaction energies are scaled by 1.16 (for neutral molecules only), although both parameters can be adjusted by the user.[15,31,35–37] The compound s dipole moment is also targeted for fitting, within a range of 1.2 to 1.5 times the QM value.[15,27] The full set of QM-derived interaction distances, energies, and dipole moment provide the necessary data for subsequent fitting and optimization of the atomic partial charges.

In order to optimize the charges, initial values and bounds (allowable range for a charge) for each charge group, defined as a set of chemically equivalent atoms expected to have identical charges, must first be set. While ffTK can set these constraints automatically (Figure 3), the user has the option to override or modify them easily. ffTK employs a connectivity-based fingerprinting method to detect and group equivalent atoms. Starting from each atom, the method traverses the molecular graph and records the set of pre-assigned atom types located at each step. Atoms with equivalent fingerprints are grouped within the charge constraints section, and assigned an initial value and bound based on the element. Furthermore, one can assign an overall integer charge to the target compound, accounting for those atoms with fixed charges that are excluded from the optimization but are nevertheless required to calculate water interaction energies. For example, in the CHARMM force field aliphatic hydrogen atoms are assigned a charge of +0.09 by default.[15] Although the typically large amount of QM-generated data makes the charge-fitting problem overdetermined, the additional user-defined constraints ensure that the resulting charges are physically realistic.

The typical optimization algorithm utilized for charge fitting is the Complex method (see Methods and Algorithms), a modification of the Simplex method that incorporates an additional, implicit bound on the parameters; here, the sum of the partial charges is made to match the net charge of the target compound, or of whatever subset of atoms is being fitted. For each QM-optimized target-water complex, the corresponding MM interaction energy as a function of distance is measured for a small range about the QM minimum for the trial charges. In the interest of speed, this MM interaction energy is calculated by ffTK using its own implementation of the non-bonded energy functions. Deviations in the minimum distances and energies for each interaction ($\Psi_{\text{interactions}}$, Eq. (1)), as well as the molecular dipole moment (Ψ_{dipole}, Eq. (1)), between QM and MM calculations determine the objective value for the trial set of charges. In the objective function

$$\Psi_{\text{charge}}=\Psi_{\text{interactions}}+\Psi_{\text{dipole}} \quad (1)$$

distances are scaled by 0.1Å and energies by 0.2 kcal/mol, i.e., the target accuracies, to make them dimensionless and, thus, comparable (see Methods and Algorithms Eq. (3) and Eq. (4) for a precise definition of the objective function). Relative weights between the energies, distances, and dipole moment can be set in ffTK; it is recommended to weight distances less than other factors in the fitting.[38] Similarly, each individual interaction can be weighted differently. Finally, the optimization proceeds until the change in the objective function is below a pre-set tolerance. The Charge Optimization Log Plotter (COLP) utility provides a way to visualize the convergence of the objective function (see the section on analysis tools below), helping the user to decide if further iterations of refinement are needed.

Optimization of bonds and angles from distortions along internal coordinates

The next stage of parameterization is to determine the bonded parameters, i.e., equilibrium values and force constants for all bonds and angles in the target molecule, encompassed more generally by its minimized geometry and vibrational spectrum, respectively. The CHARMM force field relies more than others on experimental data as a reference for the vibrational spectrum of the compound, although in practice obtaining such a spectrum is often impractical or impossible for the vast majority of compounds. Thus, one typically resorts to computing the spectrum quantum mechanically and using that as a reference for the subsequent parameterization.[15] However, unlike for partial charge determination, the comparison between QM and MM quantities is no longer straightforward. The difficulty lies in the different coordinate systems used for the description of vibrational spectra and for the bonded and non-bonded interactions of atoms in the force field model. Vibrational spectra are best expressed in terms of normal mode vibrations while the interactions of atoms in the CHARMM force field are described by primarily harmonic potentials for the redundant internal coordinates (ICs) defined by bonds, bond angles, dihedrals and impropers. Depending on the molecular geometry some normal modes will have only one contributing force-field coordinate, e.g., a single bond, while many other modes will reflect combinations of multiple bonds and angles.

The approach recommended for the CHARMM force field is to derive a Potential Energy Distribution (PED) from the MM Hessian calculated using the trial parameters, and compare it with the corresponding QM PED.[15] Due to the lack of a strict correlation between the force constants and the spectrum, one must manually and iteratively update the parameters until satisfactory agreement is reached. Convergence, however, can be challenging because a single parameter change can affect multiple normal modes. The hands-on nature of this approach, which also requires one to carefully define a non-unique mapping between internal coordinates and normal modes (the so-called U Matrix), causes it to be generally impractical for rapid or extensive parameterization efforts. An alternative, automated approach involves performing an eigenanalysis of two- and three-pair interaction matrices extracted from the QM-calculated Hessian, a matrix containing the second derivatives of the potential energy with respect to pairs of the input coordinates.[39] A more advanced, iterative procedure in which the trial parameters are determined directly through comparison of the MM and QM Hessian matrices has also been developed, although this procedure also relies on a translation between Cartesian coordinates and ICs.[40]

In the approach to computing bond and angle force constants developed for ffTK, rather than comparing Hessians directly, QM and MM PESs are computed and matched, working entirely with internal coordinates, rather than Cartesian coordinates or normal modes. For each IC, i.e., each bond and angle, a small distortion in two opposing directions is generated and the corresponding increase in potential energy compared to the undistorted conformation is computed (see Figure 4 and Methods and Algorithms). The resulting three energy values are a local description of the shape of the PES. While in some cases, e.g., bonds that are not part of a ring, distortion of a single IC is isolated from all others, for many other cases, the coordinates are coupled. This coupling is particularly evident for systems containing rings where, for example, distorting a single bond also distorts at least two neighboring bonds and two angles. Thus, determining the change in energy requires calculating the contribution from the targeted IC as well as from coupled coordinates.

In MM, computing the energies of different conformations is trivial, as one can simply use the force field, including the trial bond and angle parameters, evaluated for each conformation. For these evaluations ffTK currently relies on the NAMD Energy plugin in VMD. Although relatively slow due to NAMD overhead, future versions of ffTK will provide a hard-coded energy function akin to that already implemented for partial charge optimization. While the local PES could be computed in QM in a manner analogous to that for MM by carrying out a large number of time-consuming single-point energy evaluations for the distorted conformations, a simpler approach is taken. Because the Hessian describes a local harmonic approximation of the PES for the vibrational motion of a molecule about its minimized geometry, it also can be used to compute changes in energy for small distortions. Although traditionally, the Hessian is defined in Cartesian coordinates, ffTK takes advantage of Gaussian s flexibility by providing ICs in the input file for the Hessian calculation. ffTK determines the affected ICs for each distortion and sums their contributions to the energy, i.e.,

$$\Delta E=\sum_j \frac{1}{2}\frac{\partial^2 E}{\partial q_i \partial q_j}\delta q_i \delta q_j \quad (2)$$

where q_i is the bond or angle IC targeted, the sum is taken over all ICs q_j, and the derivative is the corresponding entry in the QM Hessian matrix. As prescribed for the CHARMM force field, the QM Hessian matrix is calculated at the MP2/6-31G(d) level of theory and scaled by 0.89,[15] an empirically derived factor that accounts for a systematic overestimation of vibrational frequencies.[41] Comparing distortion energies instead of force constants is of great benefit in addressing the degeneracy that naturally arises when mapping molecular vibrations onto ICs. All redundant contributions to a given molecular vibration are collapsed into one PES, the accurate reproduction of which is ultimately the goal of almost any parameterization effort.

Similar to the procedures for partial charge optimization, ffTK requires initial guesses for the equilibrium value and force constant for each bond and angle term in the force field to be parameterized. At each iteration, distortions are generated for all bonds and angles that have undetermined parameters, the corresponding MM and QM energy changes determined, and an objective measure of the fit computed. This measure relies on deviations in the energies of the distorted conformations as well as on deviations of the MM-minimized geometry from the QM one (see Methods and Algorithms Eq. (5) for a precise definition of the objective function used). Equilibrium values converge very quickly, as they are typically almost identical to the QM-minimized bond and angle lengths. Force constants, on the other hand, are much slower to converge, due to the aforementioned degeneracy in the PES. The downhill Simplex algorithm with 500–1000 iterations is usually sufficient for optimization, although multiple rounds may improve the quality of the resulting parameters.

Optimization of dihedral terms using torsion scans

The parameterization cycle concludes with the optimization of the four-bodied dihedral terms, which address rotations about bonds. Although an initial estimate for dihedral force constants and minima can be extracted from the vibrational analysis during the optimization of bonds and angles, because the functional form for dihedrals is not harmonic, an assumption implicit in the Hessian-based derivation, dihedral-parameter optimization typically requires a more elaborate treatment. Parameters are derived based on QM-determined PESs, generated by explicitly scanning the dihedrals of interest at the MP2/6-31G(d) level of theory.[15] Such scans provide the net energy for a progression of fixed dihedral values, while the remainder of the molecule is allowed to relax at each step in order to isolate the contribution of the energy associated with the dihedral of interest.

In ffTK, unparameterized dihedrals are either detected automatically or specified manually (Figure 5A), with visualization in VMD to aid selection (Figure 5B); the requisite Gaussian input files are then written for each dihedral scan. We note that ffTK uses a bidirectional scanning strategy whereby two Gaussian input files are generated for each scan—one in the positive direction and one in the negative direction (Figure 5C). The scan starts from the optimized geometry and proceeds according to the user-input step size and scan range. The benefit of running the scan in each direction separately derives primarily from the fact that attempts to traverse extremely high energy conformations, such as those encountered for dihedrals involving rings, are prone to abnormal termination. However, the resulting log file will still contain the data from the low energy starting point (equilibrium geometry) to the point of termination, which represents the relevant regime for parameterization. Upon import back into ffTK, data from the negative-direction scan is reversed in order to reconstruct the complete torsion profile, thus mimicking a single-sweep scan (Figure 5D).

For each optimized conformation and energy in the QM dihedral scan, a corresponding MM energy must be determined. However, even though the MM bond and angle parameters have previously been optimized with respect to the QM spectra (see the previous section), some differences between the QM and MM model will remain, i.e., some of the bonds and angles for each QM-determined structure will be slightly offset from their corresponding minimum in the MM force field. In order to prevent the contamination of the dihedral energy being fitted with these additional contributions from bonds and angles, the molecule s geometry is first optimized (using the MM model) before the MM energy is compared to the QM reference. To maintain the validity of the comparison for the torsional potential, those dihedrals that are being parameterized are kept fixed during the geometry optimization step.

The subsequent optimization of the dihedral parameters proceeds via a simulated annealing protocol (see Methods and Algorithms) to minimize the difference between the QM and MM PESs for all scanned dihedrals, identical to that developed by Guvench and MacKerell.[29] The MM energy excluding the contributions of each unparameterized dihedral is precomputed during the restrained minimization described above. At each step of the optimization, the additional dihedral contributions are computed via a hard-coded implementation of the CHARMM dihedral energy function and added to the initial MM-computed energy to yield the full MM PES. In the optimization, only the value of the force constant *k* in the dihedral energy term is continuously varied, leaving still two undetermined parameters, namely the periodicity *n* and phase shift angle δ. The periodicity is set by the user, with multiple values for a single dihedral allowed, each with its own value of *k*. For the majority of chemical bonds, $n = 1, 2, 3, 4$, and/or 6 are sufficient to cover the periodicity of rotation. The latter parameter, δ, in principle can take on any value in the CHARMM potential energy function. Values other than 0 or 180°, however, introduce an asymmetry that results in different energies for molecules with stereogenic centers (e.g., enantiomers, diastereomers), an undersirable trait for a generalized force field. Accordingly, the current approach restricts δ to 0 or 180°, with both possibilities considered during optimization.

As was the case for other optimization routines, ffTK offers the user control over several bounding criteria, e.g., the maximum allowed force constant (k_{max}), locking phase shift value to either 0 or 180°, and defining an energy cutoff in the PES fitting to prevent rarely visited high-energy conformations from dominating the objective function. Once the optimization satisfies the tolerance criterion, the QM PES, initial MM PES, and the final MM PES are stored in memory and can be visually analyzed using an internal plotting utility to determine if further refinement is necessary (see Figure 6 and the section on analysis tools below).

Methods and algorithms

A critical component of ffTK is the ability to quantitatively assess parameter performance on-the-fly by employing novel algorithms that score the ability of a given parameter set to reproduce the target data. When coupled to a variety of existing optimization schemes, these algorithms enable a programatic approach to parameter development, shifting the burden of tedious trial-and-error from the user to ffTK. The following sections first describe these algorithms, referred to as “objective functions”, and elaborate upon built-in features, such as implicit constraints, user-defined constraints, and the ability to tune specific terms, to ensure one arrives at physically-relevant parameters. Highlighted next are embedded analysis tools that allows the user to track the progress of the optimization for the purpose of assessing convergence and parameter performance. The final section of the Methods describes MD simulation protocols for rigorous testing of complete parameter sets against experimental data.

Multidimensional optimization of MM parameters

In the course of developing parameters multiple optimization problems are encountered that are best solved by numerical approaches, including determination of partial charges, as well as bond, angle, and dihedral force constants and equilibrium values. The ffTK plugin utilizes two methods for optimization, the downhill Simplex method and a simulated-annealing variant of the same algorithm; both are implemented through the Optimization plugin in VMD. The downhill Simplex method was developed by Nelder and Mead[42,43] and works by continually contracting a so-called simplex composed of N+ 1 function evaluations, i.e., points, in N-dimensional space, where N is the number of independent variables in the objective function being optimized. The initial simplex is generated randomly such that it is within the user-defined bounds on the variables. For constrained minimization problems, such as those encountered when fitting charges, which must sum to a fixed value, the Complex method[44] was also implemented in VMD.

The downhill Simplex method is slow, requiring a significant number of function evaluations, and can potentially miss a global minimum in favor of a local one. The simulated annealing method, on the other hand, can more effectively explore parameter space and is particularly useful for problems of high dimensionality. Both methods introduce an element of randomness, and it is often appropriate to run multiple cycles to check for consistency; additionally, following simulated annealing with downhill Simplex optimization can further enhance convergence.

For fitting of partial charges, the objective function being optimized targets three quantities measured in quantum chemical calculations, namely, the net dipole moment of the compound and the minimum energy and distance for interaction between a TIP3P[33] water and each atom tested. There does not need to be a one-to-one correlation between the atoms tested and the charges being fit and, in fact, there is not for most compounds. The objective function for charge optimization, Eq. (1), includes a target-water objective term given by:

$$\Psi_{\text{interactions}} = \sum_{\text{interactions}} w_i \left[\left(\frac{E^{\text{QM}} - E^{\text{MM}}}{E_{\text{scale}}} \right)^2 + w_{\text{dist.}} \left(\frac{d^{\text{QM}} - d^{\text{MM}}}{d_{\text{scale}}} \right)^2 \right] \quad (3)$$

where w_i and $w_{\text{dist.}}$ are the weighting factors for each water-target interaction and the distance term, respectively, and E_{scale}=0.2 kcal/mol and d_{scale}=0.1Å are the target accuracies of the interaction energies and the interaction distances, respectively. The second term, accounting for the dipole moment, is more complex and is given by:

$$\Psi_{\text{dipole}} = N_{\text{charge}} w_{\text{dip.}} \left(\begin{cases} \left(\frac{p^{\text{MM}}/p^{\text{QM}} - 1.2}{0.1} \right)^2 \text{if} p^{\text{MM}}/p^{\text{QM}} < 1.2 \\ \left(\frac{p^{\text{MM}}/p^{\text{QM}} - 1.5}{0.1} \right)^2 \text{if} p^{\text{MM}}/p^{\text{QM}} < 1.5 \end{cases} + \begin{cases} 0 & \text{if} \Delta\theta < 30^\circ \\ p^{\text{QM}} \left(\frac{\Delta\theta - 30}{5} \right)^2 & \text{if} \Delta\theta < 30^\circ \end{cases} \right) \quad (4)$$

where Δθ is the angle between dipole moment vectors $\mathbf{p}^{\text{QM}}$ and $\mathbf{p}^{\text{MM}}$, N_{charges} is the number of atoms under consideration, and $w_{\text{dip.}}$ is the relative weight for the dipole term. The target accuracies and the range of acceptable values for both the interaction and dipole terms are all taken from the protocol prescribed by Vanommeslaeghe et al. for CGenFF.[15]

For bond and angle parameter optimization, both the force constants and equilibrium values for all bonds and angles are simultaneously optimized. The objective function targets the QM-optimized geometry and the rise in energy for small distortions (0.1 Å for bonds and 5° for angles) about that minimum energy conformation. Thus, each objective function evaluation requires a full molecular mechanics (MM) energy minimization of the compound using the trial parameters. The function is given by:

$$\Psi_{\mathrm{bonded}} = \sum_{\mathrm{bonds,\ angles}} \left(\frac{q^{\mathrm{QM}} - q^{\mathrm{MM}}}{q_{\mathrm{scale}}} \right)^2 + w_{\mathrm{E}} \left(E^{\mathrm{QM}}_{\mathrm{distort}} - E^{\mathrm{MM}}_{\mathrm{distort}} \right)^2 \quad (5)$$

where q is the minimized value of each bond or angle, q_{scale} is 0.03Å for bonds and 3° for angles, and w_{E} is the relative weight for the energy component (E_{scale} is implicitly taken to be 1 kcal/mol). Distortions are generated for each bond and angle in the internal coordinate (IC) list. The process is straightforward for non-redundant coordinates, i.e., bonds and angles for which a distortion can be carried out without affecting the geometry of any other ICs (see (1) under bonds and angles in Figure 4). In the case of redundant angles at a branching point, labeled (2) in Figure 4, the distortion also affects neighboring angles, for which the contribution to the change in energy is accounted. Finally, in the case of single and multi-ring structures, in which the ICs are highly coupled, both bond and angle distortions change neighboring angles as well as bonds (labeled (2) under bonds and (3) under angles in Figure 4).

Dihedral parameter optimization in ffTK follows the approach developed by Guvench and MacK-erell.[29] Variables are the force constant and phase (fixed to 0 or 180°) for each dihedral parameter. The possible multiplicities are chosen by the user, with each contributing an additional pair of variables to the objective function. The objective function is given by the difference between potential energy surfaces (PESs) for QM and MM scans about a set of pre-defined dihedrals of interest; we emphasize here again that there is not a unique mapping between the dihedrals scanned and those being parameterized. More specifically,

$$\Psi_{\mathrm{dihed.}} = \sum_{\mathrm{conformations}} w_i \left(E^{\mathrm{QM}} - E^{\mathrm{MM}} + c \right)^2 \quad (6)$$

where c is a normalization constant set to make $\partial\psi_{\mathrm{dihed.}}/\partial c = 0$, i.e., $c = \bar{E}^{\mathrm{MM}} - \bar{E}^{\mathrm{QM}}$ (averages being weighted by w_i). The sum is taken over those discrete conformations determined in the QM scans, with energies recomputed using MM.

Embedded tools for assessing parameter performance

The overwhelming utility of ffTK is the automation of repetitive and tedious tasks; however, it has been our experience that this automation must be matched with a high attention to detail to yield high-quality parameters. Functionality is provided throughout the GUI to aid in such attention; two exemplary tools, COLP and the embedded PES plotter, are described in detail below.

The COLP utility (Figure 6, top) aides in assessing the performance of assigned partial atomic charges. During the course of the optimization, many quantitative measures that drive the objective function are written to the log file. These measures range from cumulative or bulk quantities such as the total objective function value, the cumulative energy and distance contributions, and the dipole moment contribution, to the specific energy and distance of each compound-water interaction. COLP parses these quantities from the file and organizes them for easy visualization in the embedded plot window. From the resulting plot it is straightforward to both glean insight into the overall performance of the collection of charges and to focus on specific interactions that are particularly useful in identifying problematic charge assignments. This insight can be further leveraged to modify various input settings, such as weighting factors or bounds, to refine charges in an iterative manner, or to identify molecular interactions that are difficult to adequately describe using a water interaction-based point-charge force field.

An undesirable attribute of dihedral fitting is that the optimized PES is strongly dependent on the user input. This dependence arises from the complexity of the dihedral description (with *n*, *k*, and δ parameters) compounded by the overdetermined nature of the dihedral fitting, such that inappropriate user input can yield a good fit via unphysical parameters, or prevent the optimizer from arriving at an acceptable fit altogether. While reasonable user input can often be determined *a priori* from chemical intuition, this is not always the case, and can be especially frusturating for inexperienced users. Accordingly, the dihedral optimization tools feature an embedded plotting utility to directly visualize the MM PES fit to QM target data for each refinement iteration (Figure 6, bottom), in addition to reporting the objective value. With this utility, users are encouraged to experiment with input values and optimization constraints to resolve the effects on the MM PES, e.g., shape and minima/maxima, that are not adequately captured by the scalar objective returned by the optimizer.

Simulation methods for computing condensed phase properties

In addition to the embedded analysis tools, parameters generated from ffTK were further assessed by their ability to reproduce experimentally measured properties, specifically density, enthalpy of vaporization, and free energy of solvation. The first two, pure-solvent properties required simulating a 6 × 6 × 6 grid of molecules, in which the center of mass for each molecule was positioned at a grid point with an initial random orientation. The system was then subjected to a 10,000-step conjugate gradient minimization, followed by 1.2 ns of equilibration, and a final 0.4 ns of production simulation in the NPT ensemble, all in accordance with the procedures used for CGenFF.[15] The density was computed by calculating the average periodic cell volume over the production simulation. The enthalpy of vaporization was computed via Eq. (7) after a further set of simulations in the gas phase. The final conformation of each molecule from the preceding condensed phase simulation was isolated and simulated in the gas phase for 0.1 ns. All simulations were run using NAMD[45] with T=298.15 K and a 1 fs timestep.

$$\Delta H_{vap} \approx -\frac{\langle U_{liq}\rangle}{N_{mol}}+\langle U_{gas}\rangle+RT \quad (7)$$

Solvation free energies were calculated using free-energy perturbation (FEP)[46] on each molecule in a 30-Å/side box containing approximately 1,000 water molecules. Both forward and reverse calculations were run in which the molecule was coupled and decoupled from the environment, respectively; intramolecular interactions were not perturbed, obviating the need for an additional vacuum-state calculation. The reaction coordinate was subdivided into 50 windows, each run for 200 ps, equally divided between equilibration and sampling (10 ns per simulation). Finally, the forward and reverse results were combined using the Bennett acceptance ratio.[47] Reported values are the average of five independent runs. A long-range correction accounting for the cutoff of van der Waals (vdW) interactions at 12Å was also computed. This correction was taken as the average difference between the solute-solvent interaction energy with a cutoff of 50 Å and with a cutoff of 12 Å over the last 50 ps of an 80-ps simulation in a 70-Å/side box of water.[48] The correction was typically between −0.4 and −0.2kcal/mol.

Software implementation

ffTK is distributed as a plugin for VMD,[49] a software package for the visualization of structural data with over 195,000 registered users and a longstanding history of providing tools for preparing, visualizing, and analyzing molecular dynamics simulations. Conveniently, VMD contains additional plugins that support tasks relevant to parameterization, such as structure building (Molefacture), an interface to external QM packages/software (QMTool), energy calculations (NAMDEnergy), and plotting utilities (Multiplot), among many others. ffTK is written entirely in Tcl, a powerful yet flexible scripting language embedded into VMD. A distinct advantage of coding in Tcl is that ffTK runs similarly on Linux, MacOS, and Windows machines without any platform-specific code, and is easily accessible for examination, modification, and extension by developers and users alike.

Users are expected to interact with ffTK via the provided GUI, written in Tk. Each step of the workflow is organized into tabs contained within the main window, generally proceeding from left to right (Figure 7). Tabs containing many sections, or distinct but related tasks, are divided into collapsible treeview-like elements to prevent users from becoming overwhelmed with too much information or too many settings. All settings, options, and processes are controlled via common software interaction paradigms such as buttons, menus, and file dialogs. Where possible, themed widgets (Ttk) have been used to provide a native appearance in a platform-specific manner. All tasks are completed directly within ffTK, with the exception of QM calculations which are run outside of ffTK using the Gaussian[34] software package.

Parameterizing pyrrolidine using ffTK

To demonstrate the simplicity with which ffTK can be used to generate a full parameter set for a molecule, we briefly describe the key steps in the parameterization of pyrrolidine, a small molecule with biological relevance found in both natural and pharmaceutical contexts.

Pyrrolidine contains several unusual features that complicate the development of static parameters capable of describing dynamically varied behavior. These features include: a cyclic structure in which dihedrals are highly coupled, an sp^3-hybridized nitrogen atom that can undergo inversion to yield an alternative low-energy conformation, and facial asymmetry as a result of an envelope conformation that places forcefield-equivalent hydrogens in different chemical environments. It is notable that Vanommeslaeghe et al.[15] also featured pyrrolidine as an example molecule in the initial publication describing CGenFF; familiarity with the discussion therein will greatly aid understanding of the workflow and tools presented here. Additional documentation and screencasts animating the parameterization of pyrrolidine can also be found online at http://www.ks.uiuc.edu/Research/vmd/plugins/fftk.

Entry into the workflow outlined in Figure 1 minimally requires PSF and PDB files describing the connectivity and atomic coordinates, respectively. For this example the Molefacture plugin distributed with VMD was used to construct pyrrolidine, assign atom names, atom types, and initial charges. An initialized parameter file was generated in ffTK using the `BuildPar` tab to identify all bonds, angles, dihedrals, and non-bonded terms required to describe the molecule. The non-bonded terms were then assigned by analogy to existing atom types using the parameter browser (also in `BuildPar`), while all other parameter values were left zeroed out. Finally, the molecular geometry was optimized at the MP2/6-31G(d) level of theory using the `Opt. Geometry` tab, which provides a targeted interface for generating the required Gaussian input file, visualizing the Gaussian output, and writing a new PDB file containing the final optimized coordinates.

Charge optimization started with generation of the required water interaction target data from the `Water Int.` tab. Auto-detection of water interaction sites correctly identified all hydrogens as hydrogen bond “donors” and the nitrogen as a hydrogen bond “acceptor”, followed by automated positioning of each water molecule during the generation of each corresponding Gaussian input file. Upon completion of the Gaussian water optimization calculations, the resulting log files were loaded into VMD for visual inspection to ensure reasonable interaction distances and orientations were achieved (Figure 8). Three additional single-point energy calculations were performed, two at the RHF/6-31G(d) level of theory (for pyrrolidine and water), and one at the MP2/6-31G(d) level of theory (to obtain the dipole moment of pyrrolidine). The Gaussian input files for these additional single-point calculations were automatically generated by ffTK along with the water interaction input files.

The QM water-interaction and single-point data were input into the charge optimization routine in the `Opt. Charges` tab, along with the initial PSF, optimized PDB, and initialized parameter files. Charge constraints were taken directly from the "Guess" button, with the exception of non-polar hydrogens, which are assigned a fixed charge of +0.09 and removed from the optimization. The charge optimization was run iteratively, first in simulated annealing mode, followed by additional optimizations performed in downhill mode. Convergence of the final charges was assessed using the built-in COLP utility (see the section on analysis tools above) prior to writing the updated charges to a new PSF file.

The bonded parameters were optimized against QM target data extracted from the Hessian computed at the MP2/6-31G(d) level of theory. After providing the PSF/PDB file pair and the Gaussian checkpoint file (CHK) from the initial geometry optimization, ffTK generates the Gaussian input file used to calculate the Hessian. The resulting Gaussian log file was used as input target data for the bonded parameter optimization. Initial parameters for the "Parameters to Optimize" were computed for all bonds and angles using the "Guess" button. The optimization was iteratively performed in downhill mode weighting the geometry and energy terms 2:1, respectively, until the final objective value became higher than the previous run (suggesting further improvement could not be achieved). The ffTK log file containing the lowest returned objective value was used to update the initial parameter file in the `BuildPar` tab.

Torsion scans were selected automatically using the "Read from PAR button" in the `Scan Torsions` tab, which cross-checks dihedral entries from the provided parameter file(s) against any dihedrals found in the PSF file. It is notable that this routine excludes from scanning dihedrals from the same torsion (redundancies) and any dihedrals terminating in a hydrogen (as per CGenFF protocol specifications[15]). However, relevant torsions involving polar hydrogens (e.g., hydroxyls, sulfhydryls) should be explicitly scanned by adding them manually. In the case of pyrrolidine, ffTK identifies five torsions; however, two symmetric entries can be removed, leaving three unique torsions. Recalling that nitrogen inversion yields an alternate low-energy conformation, an additional scan should be performed to sample this region of the PES. Accordingly, an additional input file was prepared manually to scan the improper angle centered on the nitrogen. This inversion scan started from the optimized geometry and proceeded through 90° in 1° increments, while imposing a constraint on the C–C–C–C dihedral angle to preserve ring shape. The resulting scans, shown in Figure 9A, clearly demonstrate a coupling between each individual scan and overall ring shape.

Using chemical intuition regarding sp^3-sp^3 connectivities, the initial dihedral parameters were set to $k = 0$, $n = 3$, and $\delta = 0$ for all dihedrals under consideration, and the first round of optimization was performed in simulated annealing mode. Visual analysis of the QM PES for pyrrolidine revealed complex fine details about each energy minimum for the three ring torsions, further highlighting the coupled nature of the cyclic structure. Additionally, the nitrogen-inversion scan identified two local minima separated by an energy barrier of approximately 8 kcal/mol. Since this is the only accessible barrier at reasonable simulation temperatures, the energy cutoff for fitting was reduced from the default of 10 kcal/mol to 8 kcal/mol to capture this barrier while reducing the impact of higher energy conformations. Only one change was made to the input dihedral parameters by adding an additional term with initial settings of $k = 0$, $n = 1$, and $\delta = 0$ for the C–C–N–H dihedral. Combining the $n = 1$ and $n = 3$ terms allows for two different energy minima representative of the axial and equatorial positions of the polar hydrogen, and is required to fit the N-inversion target data appropriately. All other results from the initial optimization were retained, and the refinement optimization was run, again in simulated annealing mode. The resulting refined parameters provided better agreement between the shape of the MM and QM PESs; however, the magnitudes of the local minima were still disparate. A final refinement was performed in downhill mode with a reduced tolerance of 0.0001, yielding an excellent fit of the MM PES to the QM target PES (Figure 9B). The final optimized dihedral parameters were written to an ffTK log file, and the parameter file was updated accordingly using the BuildPar tab.

An additional round of parameterization from start-to-finish was performed to ensure self-consistency, and is generally encouraged as standard practice. In this round, the geometry was optimized via a 1000-step conjugate gradient minimization in NAMD, in lieu of the QM geometry optimization, using the parameters developed in the first round. Next, water interactions were re-optimized using this MM-minimized geometry; however, all other QM-derived target data (Hessian and torsion scans) were taken from the initial round, significantly reducing the time required for subsequent stages. Further rounds of parameterization did not improve the parameters to any appreciable degree.

The density, enthalpy of vaporization, and free energy of solvation were then computed for pyrrolidine starting from the QM-optimized molecular geometry in which the polar hydrogen adopts the axial conformation. Condensed phase properties were also computed for a small test set of other parameterized molecules (Table 1), with the goal being to assess the performance of ffTK-derived parameters in comparison to those in the latest CGenFF release (2b7). Despite containing challenging atom types and functionalities (e.g., sulfur, fluorine, electrophilic carbons, etc.), parameters developed using ffTK yielded condensed phase properties comparable with highly tuned parameters available in CGenFF. Computed densities and enthalpies of vaporization were within 15% of experimental data. For the most rigorous benchmark, free energy of solvation, computed values were within 0.5 kcal/mol. The lone exception was the value computed for acetaldehyde, a molecule known to undergo rapid and reversible hydration reactions in dilute aqueous solutions, a phenomenon that is outside the scope of a MM force field.

Conclusions

The task of developing suitable force field parameters for novel chemical species presents a non-trivial barrier to the simulation of many biological systems, such as proteins containing modified amino acids, small molecules, or metal centers. Small molecule parameterization also contributes to the challenges associated with extending MD simulation techniques to other fields, most prominently drug discovery. One solution put forth by others is to assign parameters based on analogy to molecules, or molecular fragments, for which parameters have been reported. An alternative approach has been presented here in which parameterization was addressed head-on by constructing a set of tools that facilitate parameter development directly from first principles in accordance with the current best practices. These efforts, released as the Force Field Toolkit (ffTK), were realized through the confluence of three initiatives: algorithm and method development, automation, and GUI design, which serve to reduce both practical and theoretical barriers associated with parameterization. Rapidly developed parameters for a small test-set of organic molecules were benchmarked by computing condensed phase properties (density, enthalpy of vaporization, free energy of solvation) and compared against both existing CGenFF parameters and experimental data. All three properties were comparable to values obtained using existing CGenFF parameters and within acceptable thresholds of given experimental values with few exceptions.

While the primary goal of ffTK is to facilitate the generation of high-quality parameters for users with varying degrees of experience with parameterization, the toolkit is not limited to this goal. Specifically, the embedded analytical tools provide several functionalities targeting the experienced user. The convenience of automation does not sacrifice access to the underlying data or the ability to closely monitor parameter performance. A logical extension of this environment is to employ ffTK not just for generating parameters, but to challenge the force field with complex molecules and functional groups for the purpose of assessing *force field performance* to characterize successes, limitations, and eccentricities. When viewed within this context, ffTK represents a unique and flexible framework for testing and shaping novel methodologies in force field development.

ffTK contains all of the necessary tools to develop a complete set of parameters; however, a number of improvements and additions are currently underway. Improvements are largely technical, such as internalization of structure optimization and energy evaluation to accelerate bonded and dihedral optimization routines, as well as removing dependencies on external programs that currently handle these tasks (e.g., NAMD). Additional planned features include expanded analysis tools, support for multiple QM packages, and parameterization of improper angles. Finally, as the feature set stabilizes, ffTK capabilities will be expanded to support parameterization of other force fields, such as AMBER and GROMOS.

Acknowledgments

This work was supported in part by the National Institutes of Health (grants R01-GM101048, R01-GM086749, U54-GM087519, and P41-GM104601 to E.T. and K22-AI100927 to J.C.G.).

References

1. MacKerell AD Jr. Feig M, Brooks CL III. Improved treatment of the protein backbone in empirical force fields. J. Am. Chem. Soc. 2004; 126:698–699. [PubMed: 14733527]
2. MacKerell AD Jr. Feig M, Brooks CL III. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J. Comp. Chem. 2004; 25:1400–1415. [PubMed: 15185334]
3. Lamoureux G, Roux B. Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm. J. Chem. Phys. 2003; 119:3025–3039.
4. Patel S, Brooks CL III. CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. J. Comp. Chem. 2004; 25:1–15. [PubMed: 14634989]
5. Patel S, Mackerell AD Jr. Brooks CL III. CHARMM fluctuating charge force field for proteins: II protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. J. Comp. Chem. 2004; 25:1504–1514. [PubMed: 15224394]
6. Anisimov VM, Lamoureux G, Vorobyov IV, Huang N, Roux B, MacKerell AD Jr. Determination of electrostatic parameters for a polarizable force field based on the classical Drude oscillator. J. Chem. Theor. Comp. 2005; 1:153–168.
7. Lamoureux G, Roux B. Absolute hydration free energy scale for alkali and halide ions established from simulations with a polarizable force field. J. Phys. Chem. B. 2006; 110:3308–3322. [PubMed: 16494345]

8. Jiang W, Hardy D, Phillips J, MacKerell A, Schulten K, Roux B. High-performance scalable molecular dynamics simulations of a polarizable force field based on classical Drude oscillators in NAMD. J. Phys. Chem. Lett. 2011; 2:87–92. [PubMed: 21572567]

9. Durrant JD, McCammon JA. Molecular dynamics simulations and drug discovery. BMC Biology. 2011; 9:71. [PubMed: 22035460]

10. Borhani DW, Shaw DE. The future of molecular dynamics simulations in drug discovery. J. Comp.-Aided Mol. Design. 2012; 26:15–26.

11. Guvench O, MacKerell AD. Comparison of Protein Force Fields for Molecular Dynamics Simulations. Methods Mol. Biol. 2008; 443:63–88. [PubMed: 18446282]

12. International Human Genome Sequencing Consortium, Finishing the euchromatic sequence of the human genome. Nature. 2004; 431:931–945. [PubMed: 15496913]

13. Drew KLM, Baiman H, Khwaounjoo P, Yu B, Reynisson J. Size estimation of chemical space: How big is it? J. Pharm. Pharmacol. 2011; 64:490–495. [PubMed: 22420655]

14. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. J. Comp. Chem. 2004; 25:1157–1174. [PubMed: 15116359]

15. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, MacKerell AD 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. 2010; 31:671–690. [PubMed: 19575467]

16. Vanommeslaeghe K, MacKerell AD Jr. Automation of the CHARMM General Force Field (CGenFF) I: Bond perception and atom typing. J. Chem. Inf. Model. 2012; 52:3144–3154. [PubMed: 23146088]

17. Vanommeslaeghe K, Raman EP, MacKerell AD Jr. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of bonded parameters and partial atomic charges. J. Chem. Inf. Model. 2012; 52:3155–3168. [PubMed: 23145473]

18. Yesselman JD, Price DJ, Knight JL, Brooks CL III. MATCH: An atom-typing toolset for molecular mechanics force fields. J. Comp. Chem. 2012; 33:189–202. [PubMed: 22042689]

19. Malde AK, Zuo L, Breeze M, Stroet M, Poger D, Nair PC, Oostenbrink C, Mark AE. An automated force field topology builder (ATB) and repository: Version 1.0. J. Chem. Theor. Comp. 2011; 7:4026–4037.

20. Schüettelkopf AW, van Aalten DM. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Cryst. D. 2004; D60:1355–1363.

21. Zoete V, Cuendet MA, Grosdidier A, Michielin O. SwissParam: A fast force field generation tool for small organic molecules. J. Comp. Chem. 2011; 32:2359–2368. [PubMed: 21541964]

22. Halgren TA. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. J. Comp. Chem. 1996; 17:490–519.

23. Wang J, Wang W, Kollman PA, Case DA. Automatic atom type and bond type perception in molecular mechanical calculations. J. Mol. Graph. Model. 2006; 25:247–260. [PubMed: 16458552]

24. MacKerell AD Jr. Bashford D, Bellott M, Dunbrack JRL, Evanseck J, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Roux B, Schlenkrich M, Smith J, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. Self-consistent parameterization of biomolecules for molecular modeling and condensed phase simulations. FASEB J. 1992; 6:A143–A143.

25. MacKerell AD Jr. Wiorkiewicz-Kuczera J, Karplus M. An all-atom empirical energy function for the simulation of nucleic acids. J. Am. Chem. Soc. 1995; 117:11946–11975.

26. Schlenkrich, M.; Brickmann, J.; MacKerell, AD., Jr.; Karplus, M. Empirical Potential Energy Function for Phospholipids: Criteria for Parameter Optimization and Applications. In: Merz, KM.; Roux, B., editors. Biological Membranes: A Molecular Perspective from Computation and Experiment. Boston: Birkhauser; 1996. p. 31-81.

27. MacKerell AD Jr. Bashford D, Bellott M, Dunbrack RL Jr. Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher IWE, Roux B, Schlenkrich M, Smith J, Stote R, Straub J,

Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B. 1998; 102:3586–3616.

28. MacKerell, AD, Jr.. Atomistic Models and Force Fields. In: Becker, OM.; MacKerell, AD., Jr.; Roux, B.; Watanabe, M., editors. Computational Biochemistry and Biophysics. 1st ed.. New York: Marcel Dekker, Inc.; 2001. p. 7-38.

29. Guvench O, MacKerell AD Jr. Automated conformational energy fitting for force-field development. J. Mol. Mod. 2008; 14:667–679.

30. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr. Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J. Am. Chem. Soc. 1995; 117:5179–5197.

31. Jorgensen W, Tirado-Rives J. The OPLS potential functions for protein energy minimization for crystals of cyclic peptides and crambin. J. Am. Chem. Soc. 1988; 110:3469.

32. Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J. Am. Chem. Soc. 1996; 118:11225–11236.

33. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 1983; 79:926–935.

34. Frisch, MJ.; Trucks, GW.; Schlegel, HB.; Scuseria, GE.; Robb, MA.; Cheeseman, JR.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, GA.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, HP.; Izmaylov, AF.; Bloino, J.; Zheng, G.; Sonnenberg, JL.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, JA., Jr.; Peralta, JE.; Ogliaro, F.; Bearpark, M.; Heyd, JJ.; Brothers, E.; Kudin, KN.; Staroverov, VN.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Ren-dell, A.; Burant, JC.; Iyengar, SS.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, JM.; Klene, M.; Knox, JE.; Cross, JB.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, RE.; Yazyev, O.; Austin, AJ.; Cammi, R.; Pomelli, C.; Ochterski, JW.; Martin, RL.; Morokuma, K.; Zakrzewski, VG.; Voth, GA.; Salvador, P.; Dannenberg, JJ.; Dapprich, S.; Daniels, AD.; Farkas; Foresman, JB.; Ortiz, JV.; Cioslowski, J.; Fox, DJ. Gaussian 09 Revision B.01. Wallingford CT: Gaussian Inc; 2009. p. 2009

35. MacKerell AD Jr. Karplus M. Importance of attractive van der Waals contribution in empirical energy function models for the heat of vaporization of polar liquids. J. Phys. Chem. 1991; 95:10559–10560.

36. Reiher, W, III. Ph. D. Thesis. Cambridge, MA: Harvard University; 1985.

37. Jorgensen WL. Optimized intermolecular potential functions for liquid alcohols. J. Phys. Chem. 1986; 90:1276–1284.

38. Vanommeslaeghe K. personal communication. 2012

39. Seminario JM. Calculation of intramolecular force fields from second-derivative tensors. Int. J. Quantum Chem. 1996; 59:1271–1277.

40. Burger SK, Lacasse M, Verstraelen T, Drewry J, Gunning P, Ayers PW. Automated parametrization of AMBER force field terms from vibrational analysis with a focus on functionalizing dinuclear zinc(II) scaffolds. J. Chem. Theor. Comp. 2012; 8:554–562.

41. Scott AP, Radom L. Harmonic vibrational frequencies: An evaluation of Hartree-Fock, Møller-Plesset, quadratic configuration interaction, density functional theory, and semiempiri-cal scale factors. J. Phys. Chem. 1996; 100:16502–16513.

42. Nelder JA, Mead R. A Simplex Method for function minimization. Computer Journal. 1965; 7:308–313.

43. Press, WH.; Teukolsky, SA.; Vetterling, WT.; Flannery, BP. Numerical Recipes in C. 2nd ed.. New York: Cambridge University Press; 1992.

44. Richardson JA, Kuester JL. The Complex Method for constrained optimization. Comm. ACM. 1973; 16:487–489.

45. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable Molecular Dynamics with NAMD. J. Comp. Chem. 2005; 26:1781–1802. [PubMed: 16222654]

46. Pohorille A, Jarzynski C, Chipot C. Good practices in free-energy calculations. J. Phys. Chem. B. 2010; 114:10235–10253. [PubMed: 20701361]

47. Bennett CH. Efficient estimation of free energy differences from Monte Carlo data. J. Comp. Phys. 1976; 22:245–268.

48. Baker CM, Lopes PEM, Zhu X, Roux B, MacKerell AD Jr. Accurate calculation of hydration free energies using pair-specific Lennard-Jones parameters in the CHARMM Drude polarizable force field. J. Chem. Theor. Comp. 2010; 6:1181–1198.

49. Humphrey W, Dalke A, Schulten K. VMD – Visual Molecular Dynamics. J. Mol. Graphics. 1996; 14:33–38.

50. Lide, DR. Handbook of Chemistry and Physics. USA: CRC Press Inc; 2008.

51. Rochester CH, Symonds JR. Thermodynamic studies of fluoroalcohols. Part 1.–Vapour pressures and enthalpies of vaporization. J. Chem. Soc. – Faraday Trans. 1. 1973; 69:1267–1273.

52. Abraham MH, Whiting GS. Thermodynamics of solute transfer from water to hexadecane. J. Chem. Soc. – Perkin Trans. 1990; 2:291–300.

53. Cabani S, Gianni P, Mollica V, Lepori L. Group contributions to the thermodynamic properties of non-ionic organic solutes in dilute aqueous solution. J. Solut. Chem. 1981; 10:563–595.

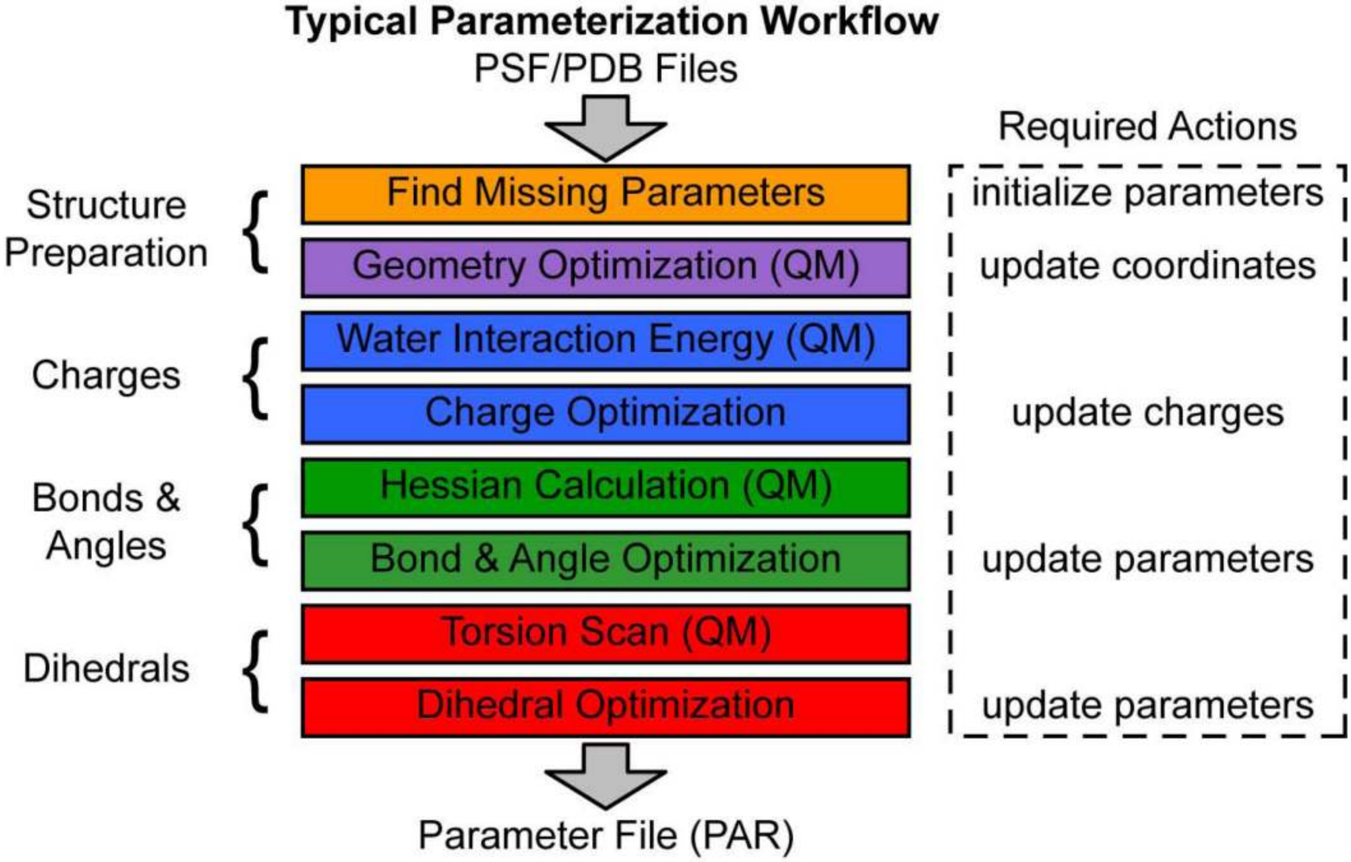


Figure 1.

A typical parameterization workflow addresses four major stages (left), each of which requires a specific set of calculations (center), and subsequent action to update a variety of file types (right). ffTK is designed as a graphical user interface that facilitates traversal of the workflow without obscuring the underlying processes or data.

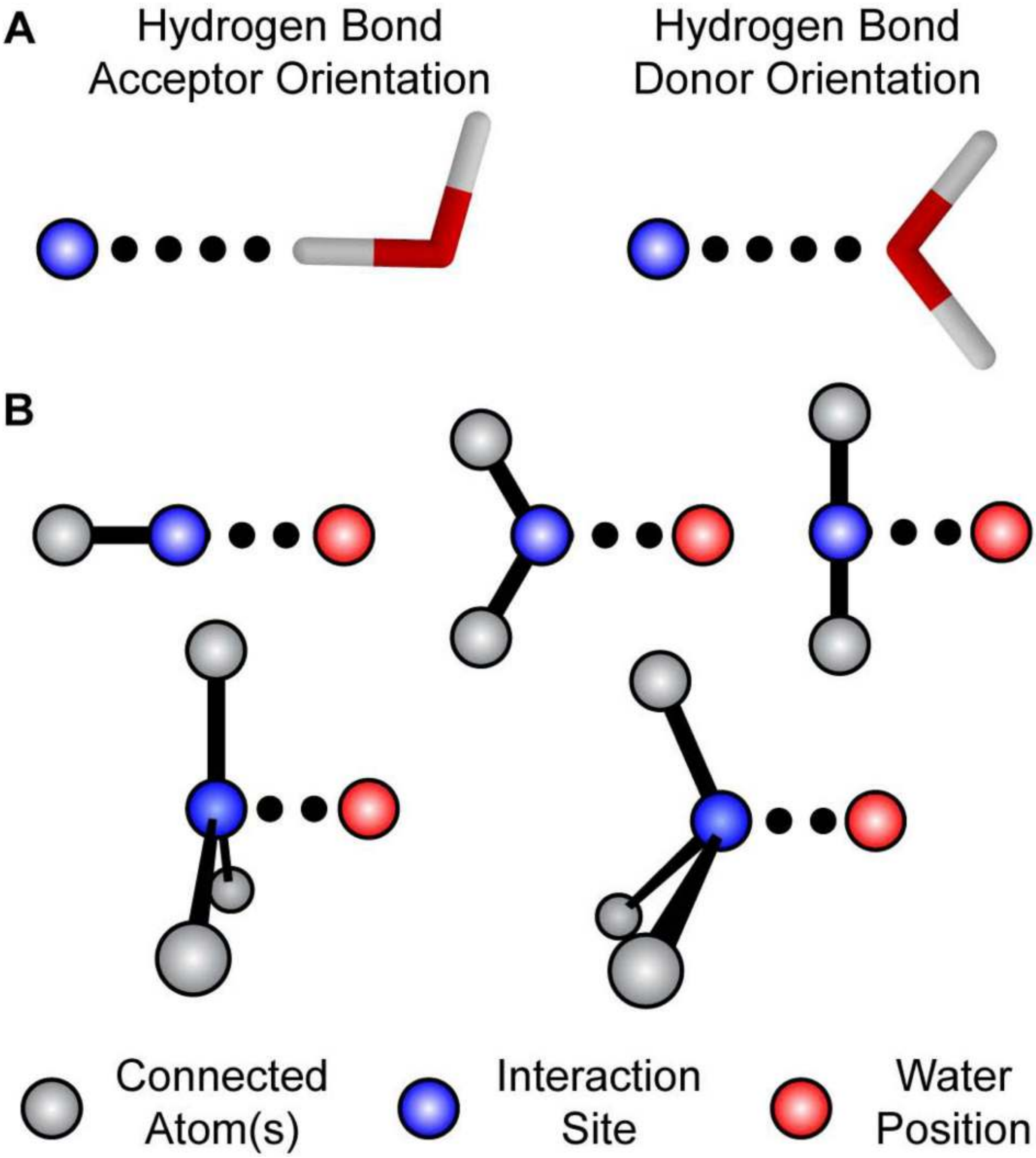
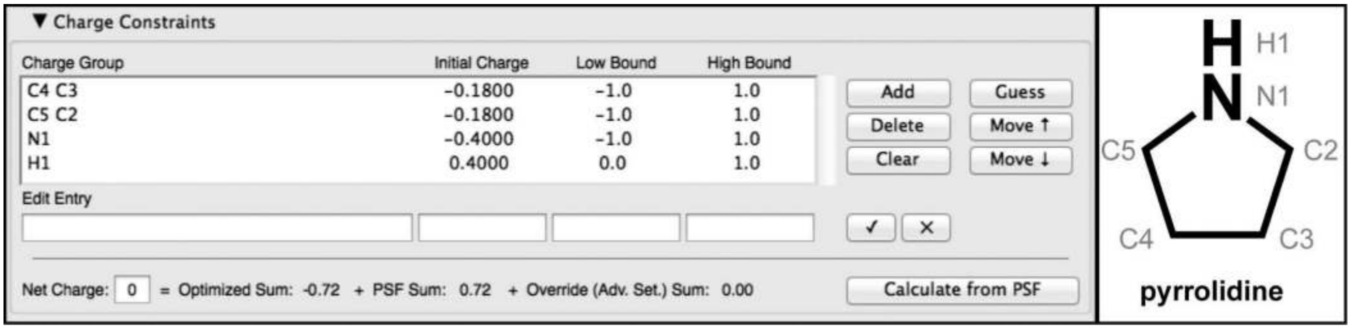

A
Hydrogen Bond Acceptor Orientation
Hydrogen Bond Donor Orientation
B
Connected Atom(s)
Interaction Site
Water Position

Figure 2.

Molecule-water interactions. (**A**) Each water interaction site (blue spheres) is identified as a hydrogen bond “acceptor” or “donor” by the orientation of the water molecule with respect to the interaction site (**B**) The position of the interacting atom of the water molecule (H or O; red spheres) is determined based on the geometry of the interaction site to reduce steric interactions with covalently bound neighbors (grey spheres).



▼ Charge Constraints

Charge Group	Initial Charge	Low Bound	High Bound
C4 C3	−0.1800	−1.0	1.0
C5 C2	−0.1800	−1.0	1.0
N1	−0.4000	−1.0	1.0
H1	0.4000	0.0	1.0

Edit Entry

Net Charge: 0 = Optimized Sum: -0.72 + PSF Sum: 0.72 + Override (Adv. Set.) Sum: 0.00

Figure 3.

The connectivity-based fingerprinting algorithm correctly identifies symmetric atoms in pyrrolidine, and provides reasonable bounds by element. The aliphatic hydrogens (8 +0 09) have been removed from the optimization; therefore, the total charge of the remaining atoms must sum to –0.72 to preserve the net neutral charge for the molecule.

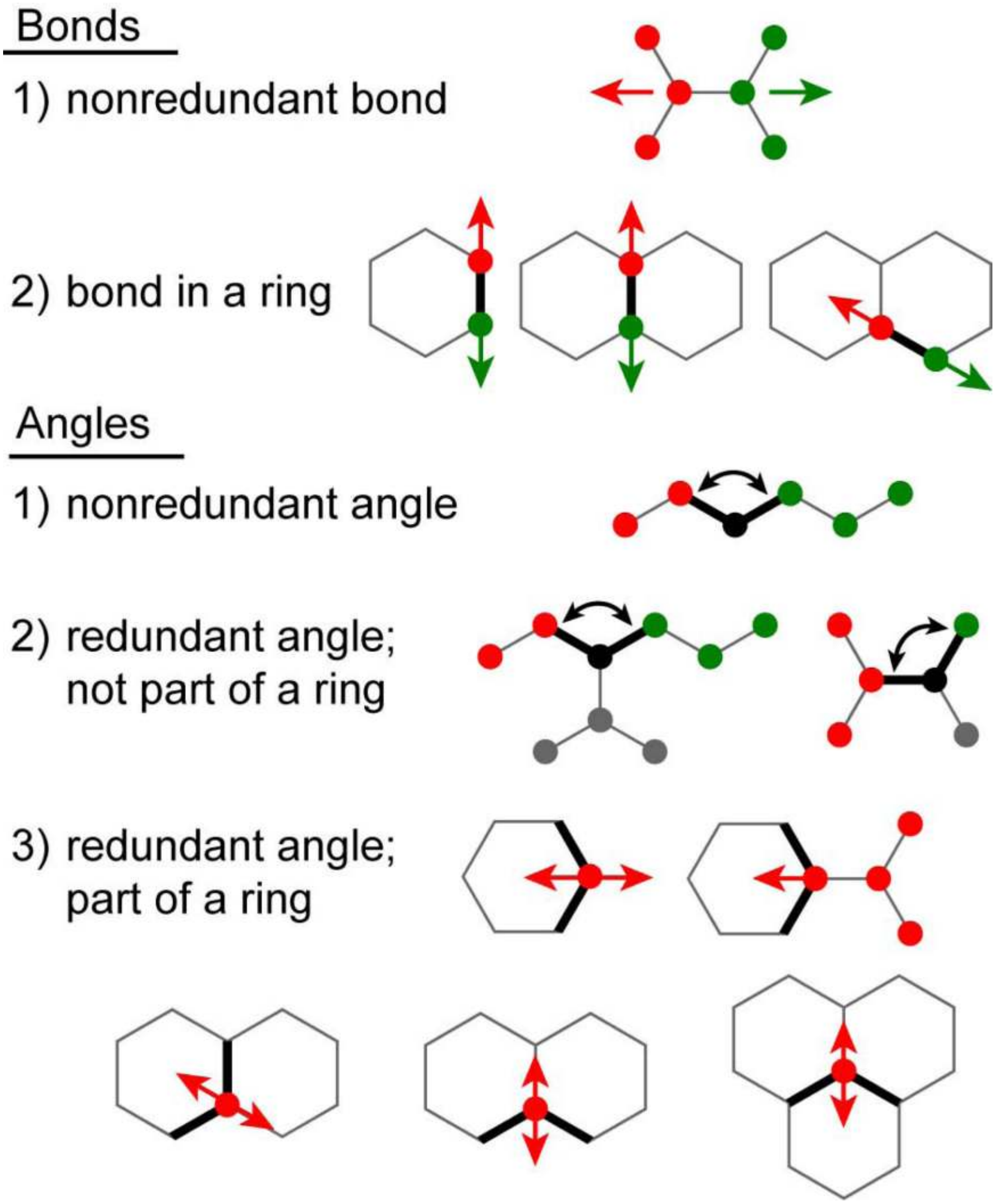
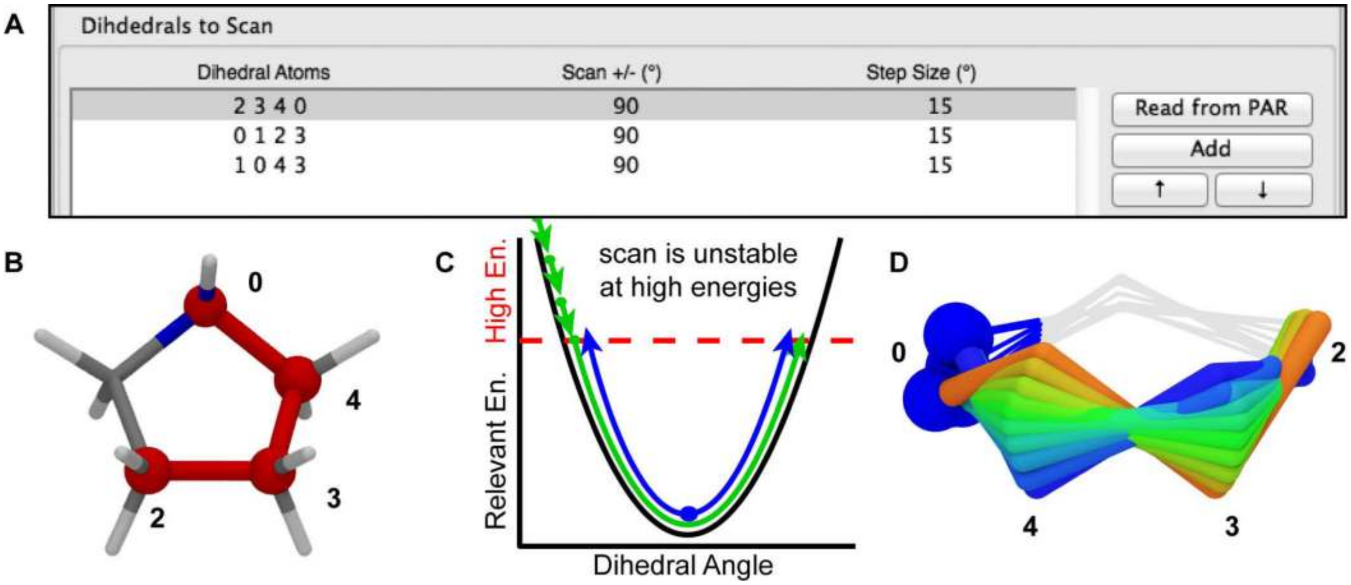

Bonds
1) nonredundant bond
2) bond in a ring
Angles
1) nonredundant angle
2) redundant angle; not part of a ring
3) redundant angle; part of a ring

Figure 4.

The MM potential energy surface used to fit bonded parameters is constructed by computing the energetic perturbation of small structural distortions from the equilibrium geometry along unparameterized internal coordinates. Red and green atoms denote distinct groups moved to effect the distortion, while black atoms remain unchanged.

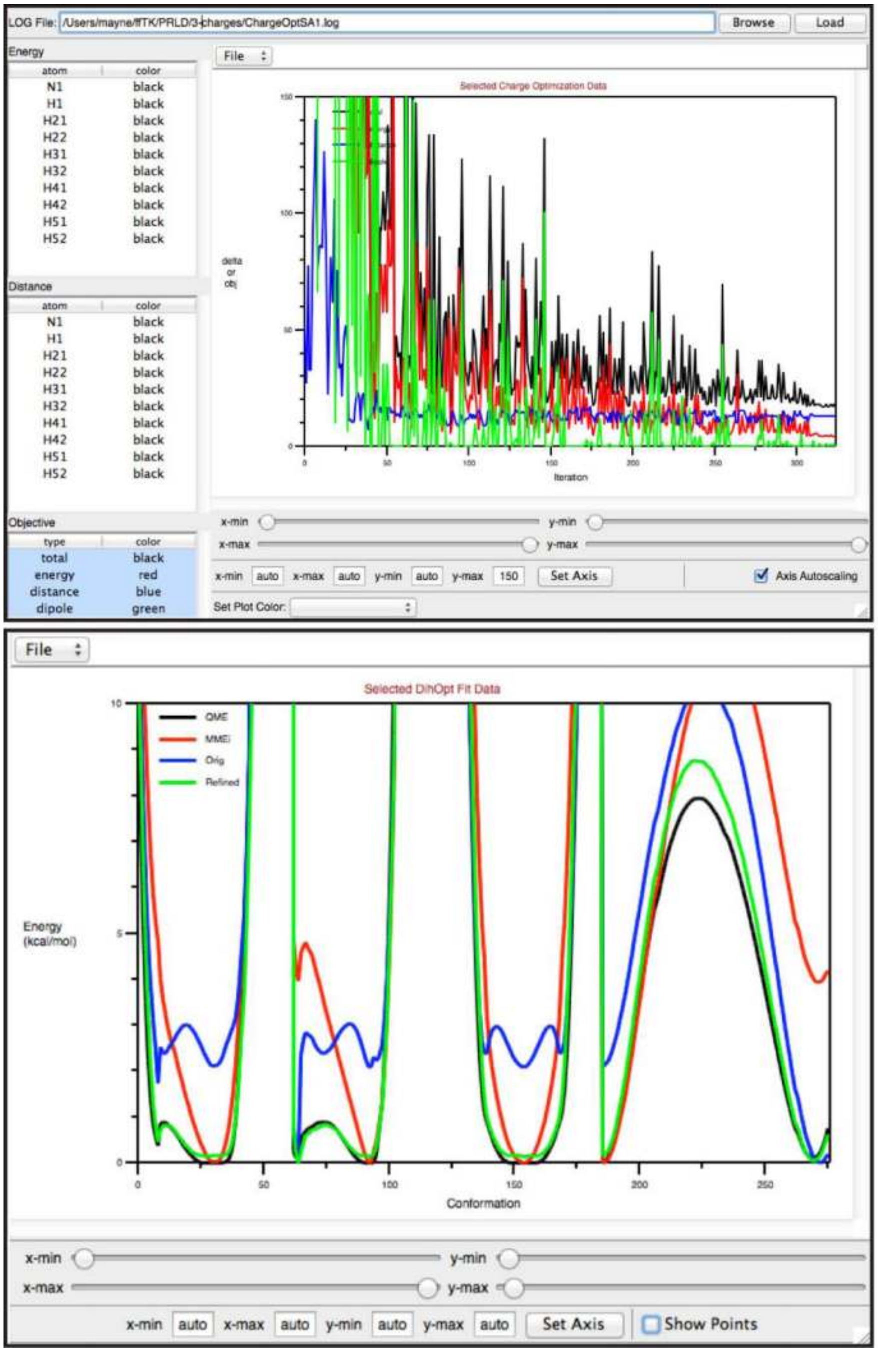


Dihdedrals to Scan

Dihedral Atoms	Scan +/- (°)	Step Size (°)
2 3 4 0	90	15
0 1 2 3	90	15
1 0 4 3	90	15

Figure 5.

Scanned dihedrals are auto-detected from a parameter file, or specified manually in the GUI (**A**). When a specific entry is selected, the corresponding dihedral is highlighted in the main VMD window (**B**). ffTK employs a bi-directional scanning technique that scans the energy regime relevant to parameterization and avoids complications of starting the scan from high-energy conformations (**C**). The result of the scan can be directly loaded into VMD for visual inspection (**D**; colored by step proceeding from blue to red).



LOG File: /Users/mayne/ffTK/PRLD/3-charges/ChargeOptSA1.log

Energy

atom	color
N1	black
H1	black
H21	black
H22	black
H31	black
H32	black
H41	black
H42	black
H51	black
H52	black

Distance

atom	color
N1	black
H1	black
H21	black
H22	black
H31	black
H32	black
H41	black
H42	black
H51	black
H52	black

Objective

type	color
total	black
energy	red
distance	blue
dipole	green

x-min auto x-max auto y-min auto y-max 150

☑ Axis Autoscaling

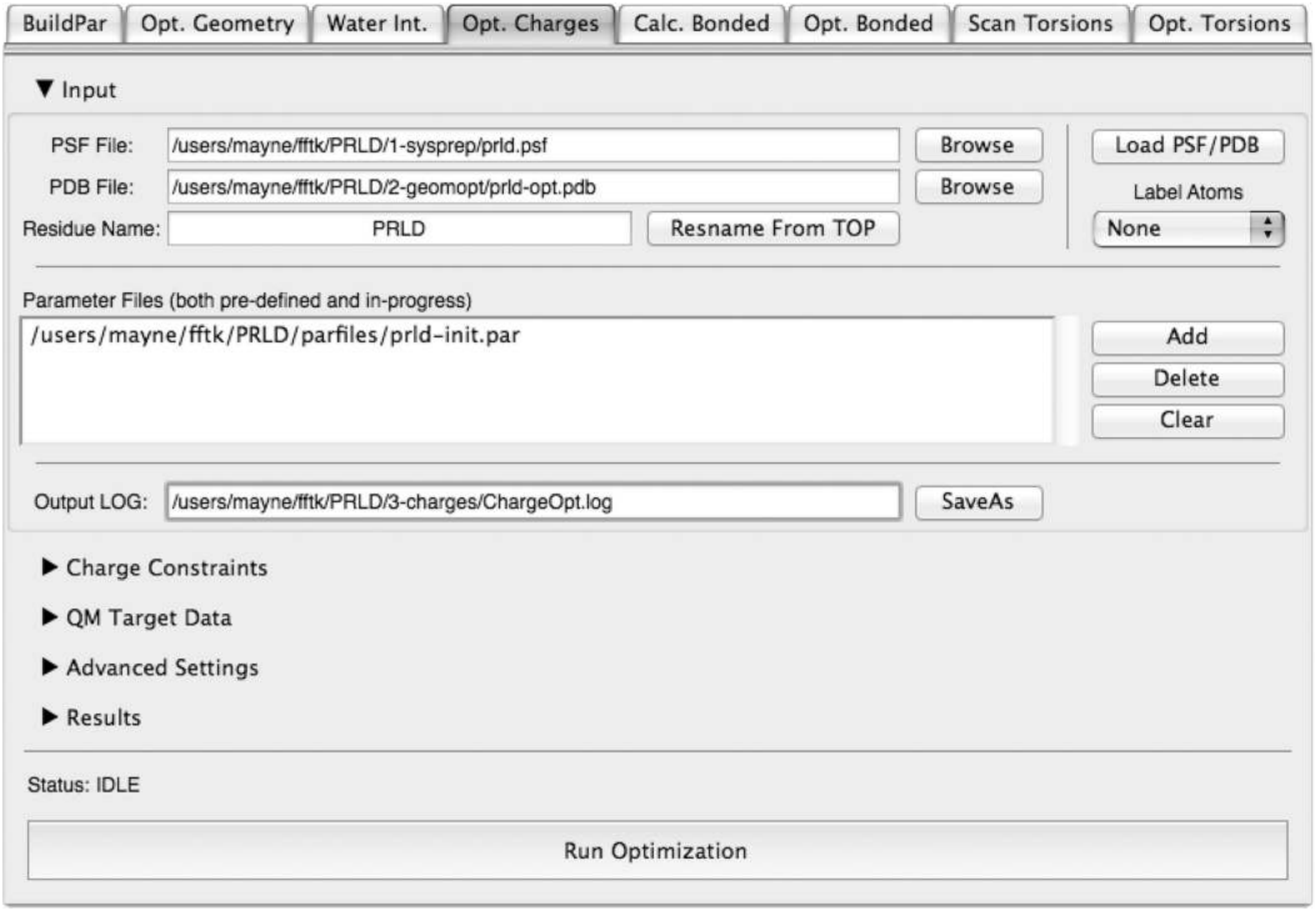
Set Plot Color:

x-min auto x-max auto y-min auto y-max auto

□ Show Points

Figure 6.

(Top) The Charge Optimization Log Plotter (COLP) extracts relevant data from the charge optimization log file for visual analysis. These data include the total objective function, the individual contributions from energy, distance, and dipole moment to the objective function (shown), as well as, the minimum interaction energy and distance for each interaction site probed during the optimization. (Bottom) During dihedral fitting, the MM potential energy surface (PES) computed for each refinement step can be visualized using an embedded plotting utility for comparison against the QM target PES. The data from both plotting utilities can be directly exported to file for import into popular plotting applications to generate publication-quality plots.



PSF File: /users/mayne/fftk/PRLD/1-sysprep/prld.psf

PDB File: /users/mayne/fftk/PRLD/2-geomopt/prld-opt.pdb

Residue Name: PRLD

Label Atoms: None

Parameter Files (both pre-defined and in-progress)

/users/mayne/fftk/PRLD/parfiles/prld-init.par

Output LOG: /users/mayne/fftk/PRLD/3-charges/ChargeOpt.log

Status: IDLE

Figure 7.

The graphical user interface (GUI) organizes each parameterization step into a separate tab containing collapsable elements to group-related settings and options. The GUI also makes use of common software interaction paradigms, such as file dialogs, buttons, and menus, that are native to each particular operating system. It also includes many tools for automatically detecting common settings, performing the required actions from Figure 1, and interacting with the VMD main window.

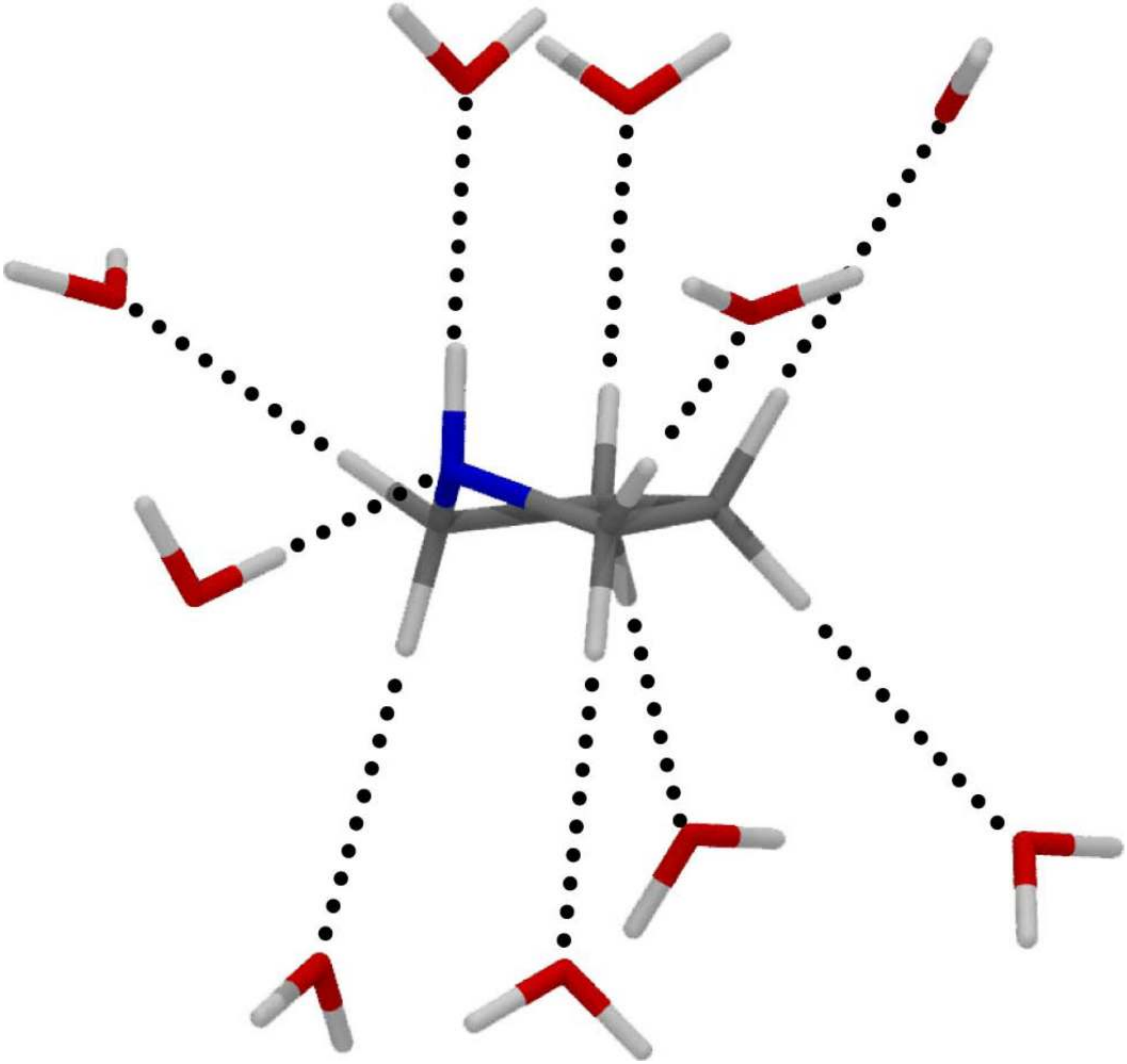


Figure 8.

Optimized water interactions with pyrrolidine were individually computed at the MP2/6-31G(d) level of theory. The calculations were run in Gaussian; however, all input files were generated from ffTK and the results visualized in VMD.

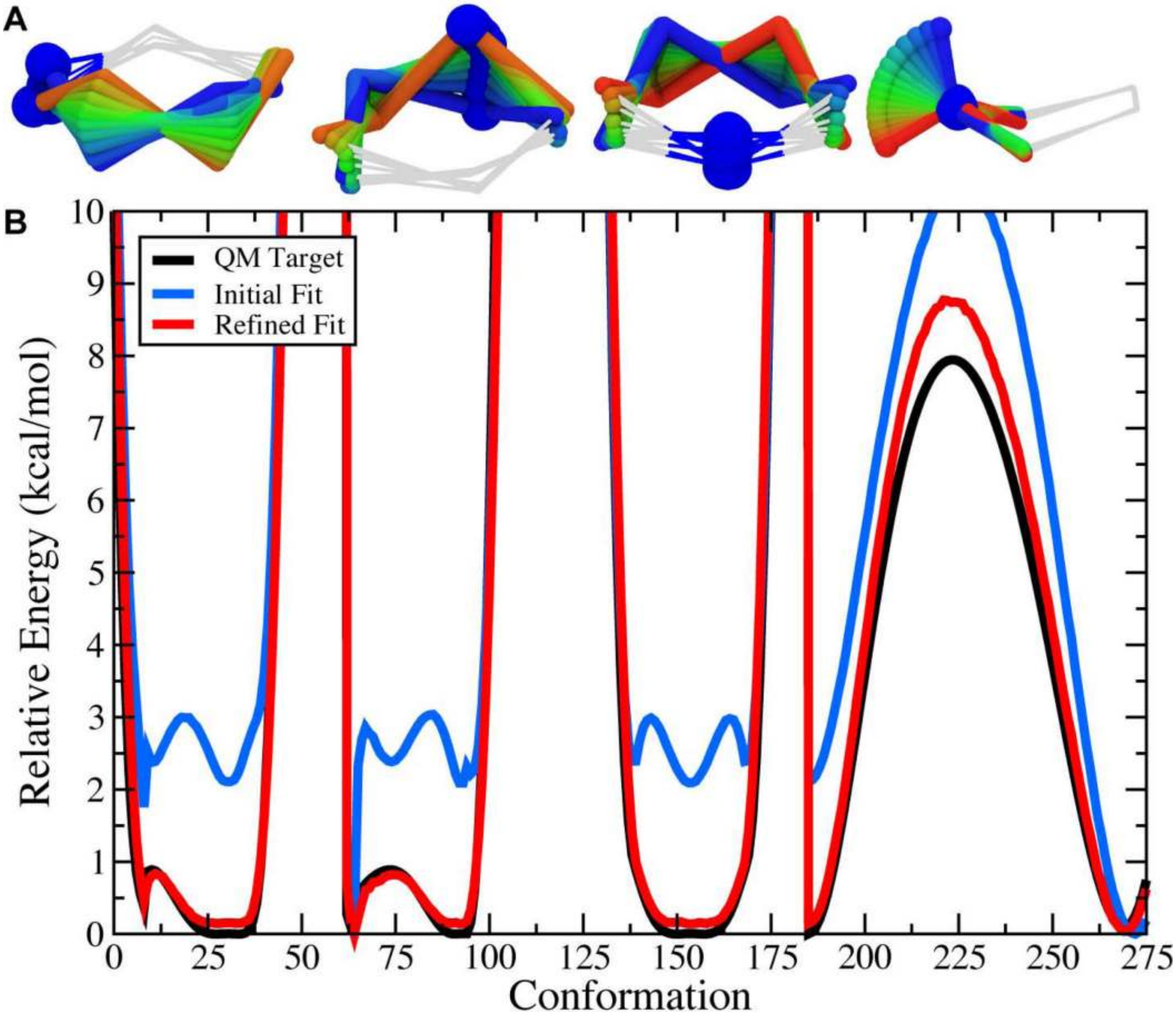


Figure 9.

(**A**) The scanned torsions for pyrrolidine were loaded into VMD and colored by iteration (blue to red) to provide a structural context for assessing the dihedral PESs. (**B**) The cyclic structure of pyrrolidine yields a complex dihedral scan profile, in which much of the fine detail lies below 1 kcal/mol; however, an accessible barrier of 8 kcal/mol exists for the N-inversion. The refined MM parameters yield a PES (red) that reproduces the QM PES (black) with excellent agreement.

Table 1

Computed[a] *vs.* Experimental Condensed Phase Properties

Molecule	ρ (g/mL)			ΔH_{vap} (kcal/mol)			ΔG_{solv} (kcal/mol)		
	ffTK	CGenFF	Exp[b,c]	ffTK	CGenFF	Exp[b,c]	ffTK	CGenFF	Exp[g]
pyrrolidine	0.89	0.88	0.86[d]	10.22	9.70	8.99	−5.79	−5.33	−5.48
butanone	0.78	0.79	0.80	7.93	8.74	8.31	−3.28	−3.34	−3.71
acetic acid	1.06	1.01	1.04	13.15	10.85	5.58	−7.16	−6.83	−6.69
acetaldehyde	0.81	0.77	0.78[e]	7.26	6.22	6.09	−4.47	−4.38	−3.50
ethanol	0.81	0.80	0.79[d]	9.88	10.29	10.11	−4.61	−4.49	−5.00
ethanethiol	0.89	0.86	0.83	7.29	6.60	6.52	−0.96	−0.99	−1.14
trifluoroethanol	1.41	1.35	1.38[d]	9.55	9.04	10.51[f]	−4.05	−3.41	−4.31[h]

[a] all simulations performed at 298.15 K as described in Methods

[b] exp. values obtained from ref.[50]

[c] exp. T=298.15 K

[d] exp. T=293.15 K

[e] exp. T=291.15 K

[f] exp. values obtained from ref.[51]

[g] exp. values obtained from ref.[52]

[h] exp. values obtained from ref.[53]